

Développement d'approches basées sur les réseaux de neurones convolutifs (CNN) pour prédire la granulométrie des sols à partir d'images synthétiques, de laboratoire et de terrain

par

Harold SIMO TENEKAM

THÈSE PAR ARTICLES PRÉSENTÉE À L'ÉCOLE DE TECHNOLOGIE
SUPÉRIEURE COMME EXIGENCE PARTIELLE À L'OBTENTION
DU DOCTORAT EN GÉNIE
PH. D.

MONTRÉAL, LE 17 JUIN 2026

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

©Tous droits réservés, Harold Simo Tenekam, 2026

©Tous droits réservés

Cette licence signifie qu'il est interdit de reproduire, d'enregistrer ou de diffuser en tout ou en partie, le présent document. Le lecteur qui désire imprimer ou conserver sur un autre media une partie importante de ce document, doit obligatoirement en demander l'autorisation à l'auteur.

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. François Duhaime, directeur de thèse
Département du génie de la construction, École de technologie supérieure

M. Jean-Sébastien Dubé, codirecteur de thèse
Département du génie de la construction, École de technologie supérieure

M. Matthew F. Toews, codirecteur de thèse
Département du génie de la construction, École de technologie supérieure

M. Philippe Bocher, président du jury
Département de génie mécanique, École de technologie supérieure

M. Jonathan Aubertin, membre du jury
Département du génie de la construction, École de technologie supérieure

M. Richard Brunet, examinateur externe
Département d'innovation, WSP Canada Inc.

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 27 MAI 2026

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à mon directeur de thèse, M. François Duhaime, ainsi qu'à mes co-directeurs, M. Jean-Sébastien Dubé et M. Matthew F. Toews, pour leur encadrement, leurs conseils avisés et leur soutien constant tout au long de cette thèse. Je remercie également M. Jonathan Aubertin, M. Philippe Bocher, et M. Richard Brunet pour le temps consacré à l'évaluation de ce travail.

Je souhaite rendre hommage à mon père Tenekam Joseph, décédé le 12 octobre 2025, qui a toujours été là pour moi dans les moments les plus difficiles, ainsi qu'à ma mère Zebaze Bixille Alexise, pour son amour et son soutien indéfectible. Je remercie ma grande sœur Tankeu Tenekam Russelle Beriette, ainsi que mes frères, Tenekam Bill Hurell et Fomethe Tenekam Darell pour leur encouragement et leur présence tout au long de mon parcours.

Enfin, je remercie tout particulièrement ma conjointe Ouambo Kwekam Ashley pour son soutien inconditionnel sa patience, sa compréhension et surtout sa présence à chaque étape de cette aventure.

Développement d'approches basées sur les réseaux de neurones convolutifs (CNN) pour prédire la granulométrie des sols à partir d'images synthétiques, de laboratoire et de terrain

Harold SIMO TENEKAM

RÉSUMÉ

L'analyse granulométrique est l'un des essais les plus demandés dans les laboratoires d'analyse géotechnique. Cet essai permet de déterminer la distribution des tailles de particules d'un sol (granulométrie). Traditionnellement, le tamisage et la sédimentométrie sont respectivement utilisés pour l'analyse granulométrique des sols granulaires et des sols fins. Bien que fiables, ces techniques présentent l'inconvénient d'être chronophages. Les outils d'analyse d'image existants comme Wipfrag réduisent ce temps d'analyse. Néanmoins, ils sont limités aux particules de grande taille et utilisent des approches de segmentation sensibles à la résolution des images.

Cette thèse propose une approche fondée sur l'apprentissage profond pour estimer la granulométrie à partir de photographies afin de surmonter les limites des méthodes traditionnelles et des logiciels d'analyse d'image actuels. À cette fin, un jeu de données d'images synthétiques et segmentées a été généré à l'aide du moteur Unity game engine afin de reproduire des textures de sol réalistes. L'entropie conditionnelle ainsi qu'un modèle de segmentation d'objets (GrainSegNet) ont été utilisés pour évaluer leur réalisme visuel. Par la suite, un réseau de neurones convolutionnels (GrainSizeNet) a été développé pour analyser l'influence des types d'images d'entrée (vues dessus/dessous, avec ou sans segmentation) ainsi que l'apport des images synthétiques générées comparativement aux méthodes traditionnelles d'augmentation de données. Enfin, GrainSizeNet a été adapté à l'analyse des photographies de cuillères fendues (CF) prélevées lors de l'essai de pénétration standard afin de distinguer cinq catégories de sols du système USCS (Unified Soil Classification System) et de prédire la proportion massique de particules fines (diamètre inférieur à 0,075 mm), en comparant l'apprentissage direct et hiérarchique.

L'entropie conditionnelle révèle une similarité entre les images synthétiques et réelles, et l'évaluation de GrainSegNet sur des images réelles montre des performances satisfaisantes sur des photographies de laboratoire (IoU moyen de 0,60). En revanche, elles diminuent dans des contextes plus complexes (IoU moyen de 0,26), notamment pour les matériaux à forte densité de particules ou plus fins. Ces résultats soulignent la pertinence des images synthétiques tout en montrant la nécessité d'adaptations supplémentaires pour des environnements très différents de ceux simulés.

Les variantes de GrainSizeNet, évaluées avec différentes configurations de données d'entrée, présentent des performances similaires, avec des MAE (Mean Absolute Error) de 2,90% à 3,66% calculées sur les pourcentages massiques retenus aux 11 tamis de contrôle de la norme BNQ 2501-025, indiquant qu'une seule image du dessus suffit pour assurer une estimation fiable de la granulométrie. Par ailleurs, l'intégration d'images synthétiques améliore la

généralisation du modèle : la MAE a été réduite de 11,54 % (jeu de données composé uniquement d'images réelles) à 6,19 %.

Appliqué aux données de terrain, le modèle obtient des résultats prometteurs. Les stratégies hiérarchiques offrent les meilleures performances pour la classification (précision = 69 %) et pour l'estimation de la proportion de fines (MAE = 15,3 %). Malgré ces résultats encourageants, GrainSizeNet demeure dépendant de la quantité et la qualité des données disponibles pour son entraînement. En outre, ses performances s'améliorent à mesure qu'il est alimenté par des ensembles de données plus volumineux et diversifiés.

Mots-clés : Analyse granulométrique, cuillère fendue, GrainSizeNet, images synthétiques, réseaux de neurones convolutionnels, segmentation, Unity game engine, USCS.

Development of Convolutional Neural Network (CNN)-Based Approaches for Predicting Soil Particle Size Distribution from Synthetic, Laboratory, and Field Images

Harold SIMO TENEKAM

ABSTRACT

Particle size analysis is one of the most frequently requested tests in geotechnical analysis laboratories. This test determines the distribution of particle sizes in a soil (particle size distribution). Traditionally, sieving and sedimentometry are used for particle size analysis of granular soils and fine soils, respectively. Although reliable, these techniques have the disadvantage of being time-consuming. Existing image analysis tools such as Wipfrag reduce this analysis time. However, they are limited to large particles and use segmentation approaches that are sensitive to image resolution.

This thesis proposes a deep learning-based approach to estimating particle size from photographs in order to overcome the limitations of traditional methods and current image analysis software. To this end, a dataset of synthetic, segmented images was generated using the Unity game engine to reproduce realistic soil textures. Conditional entropy and an object segmentation model (GrainSegNet) were used to evaluate their visual realism. Subsequently, a convolutional neural network (GrainSizeNet) was developed to analyze the influence of input image types (top/bottom views, with or without segmentation) and the contribution of the generated synthetic images compared to traditional data augmentation methods. Finally, GrainSizeNet was adapted to analyze photographs of split-spoon samples (SS) taken during the standard penetration test in order to distinguish five soil categories from the USCS (Unified Soil Classification System) and predict the mass fraction of fine particles (diameter less than 0.075 mm) by comparing direct and hierarchical learning.

Conditional entropy reveals a similarity between synthetic and real images, and GrainSegNet's evaluation on real images highlighting satisfactory performance on laboratory photographs (average IoU of 0.60). However, performance decreases in more complex contexts (average IoU of 0.26), particularly for materials with high particle density or finer particles. These results highlight the relevance of synthetic images while showing the need for further adaptations for environments that are very different from those simulated.

The GrainSizeNet variants, evaluated with different input data configurations, show similar performance, with Mean Absolute Errors (MAE) ranging from 2.90% to 3.66% calculated on the mass percentages retained on the 11 control sieves of the BNQ 2501-025 standard, indicating that a single top view image is sufficient to ensure a reliable estimate of particle size distribution. Furthermore, the integration of synthetic images improves the generalization of the model: the MAE was reduced from 11.54% (dataset composed solely of real images) to 6.19%.

When applied to field data, the model produces promising results. Hierarchical strategies offer the best performance for classification (precision = 69%) and for estimating the proportion of

finer (MAE = 15.3%). Despite these encouraging results, GrainSizeNet remains dependent on the quantity and quality of the data available for training. Furthermore, its performance improves as it is fed with larger and more diverse datasets.

Keywords: Particle size analysis, split-spoon, GrainSizeNet, synthetic images, convolutional neural networks, segmentation, Unity game engine, USCS.

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE DE LITTÉRATURE	6
1.1 Analyse granulométrique	6
1.2 Tamisage et sédimentométrie	7
1.2.1 Tamisage	7
1.2.2 Sédimentométrie	9
1.3 Logiciels commerciaux d'analyse granulométrique	10
1.4 Analyse granulométrique par techniques d'apprentissage supervisé.....	13
1.4.1 Revue des méthodes d'apprentissage supervisé	13
1.4.2 Contraintes des modèles d'apprentissage supervisé	22
1.4.3 Solutions générales pour surmonter les contraintes des modèles d'apprentissage supervisé	22
1.4.4 Revue des modèles d'apprentissage les plus populaires pour la caractérisation des sols.....	26
1.4.5 Stratégies palliatives aux contraintes des modèles d'apprentissage supervisé pour l'analyse granulométrique des sols.....	31
1.5 Synthèse critique et justification de l'approche proposée.....	34
CHAPITRE 2 SEGMENTATION OF GRANULAR MATERIAL IMAGES USING SYNTHETIC DATA AND DEEP LEARNING	37
2.1 ABSTRACT.....	37
2.2 Introduction.....	38
2.3 Methodology	41
2.3.1 Scripts for synthetic image preparation	41
2.3.2 Data annotation	45
2.3.3 Validation and calibration of synthetic images.....	47
2.3.4 GrainSegNet training and testing.....	51
2.4 Results and discussion	54
2.4.1 Modeling and annotations of synthetic images.....	54
2.4.2 Validation and calibration of computer-generated images using real photographs.....	55
2.4.3 GrainSegNet Performance	61
2.5 Conclusions.....	66
CHAPITRE 3 DEEP LEARNING APPROACH TO CLASSIFY GRANULAR MATERIAL USING SYNTHETIC AND REAL IMAGES	68
3.1 Abstract.....	68
3.2 Introduction.....	69
3.3 Methodology	72
3.3.1 Image types and datasets.....	72

3.3.2	GrainSizeNet model.....	80
3.3.2.1	GrainSizeNet architecture.....	80
3.3.2.2	Loss function and optimization.....	82
3.3.2.3	Performance metrics	83
3.4	Results and discussion	86
3.4.1	Hold-out validation	86
3.4.1.1	Performance comparison across model variants (D1, D2, D3)	86
3.4.1.2	MTQ 1101 classification performance	87
3.4.1.3	Comparison of predicted vs. ground truth grain size distributions.....	89
3.4.1.4	Influence of traditional data augmentation and synthetic data augmentation	92
3.4.1.5	Assessing the effectiveness of GrainSizeNet in light of previous studies.....	95
3.4.2	Cross-validation performance	96
3.5	Conclusions.....	100

CHAPITRE 4 CONVOLUTIONAL NEURAL NETWORKS FOR SOIL CLASSIFICATION BASED ON PHOTOGRAPHS OF SPLIT-SPOON SAMPLER

4.1	ABSTRACT.....	102
4.2	Introduction.....	104
4.3	Materials and methods	107
4.3.1	SPT dataset and image preprocessing.....	107
4.3.2	Classification datasets.....	110
4.3.3	Regression datasets	113
4.3.4	Model architectures and training	115
4.3.5	Metrics for evaluating classification and regression models	121
4.4	Results.....	122
4.4.1	Training and validation performance.....	122
4.4.2	Classification results	123
4.4.3	Regression results	125
4.5	Discussion	126
4.6	Conclusions.....	130

CONCLUSION.....	132
-----------------	-----

RECOMMANDATIONS	135
-----------------------	-----

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....	137
---	-----

LISTE DES TABLEAUX

	Page
Table 2.1 Global configuration of GrainSegNet.....	54
Table 3.1 Characteristics of the four image types.....	73
Table 3.2 Characteristics of the six datasets	78
Table 3.3 GrainSizeNet model configuration	82
Table 3.4 Number of samples per MTQ 1101 class (Fines, SP, SM, GP, GW) for all real and synthetic samples.....	86
Table 3.5 Mean absolute error on mean grain diameter (d_m) for GrainSizeNet – type 1 images.....	98
Table 4.1 Number of images in the training and validation datasets for classification before data augmentation	111
Table 4.2 Training and validation subdatasets for the classification tasks	112
Table 4.3 Training and validation subdatasets for the regression tasks.....	114
Table 4.4 Model architecture summary for each subdataset.....	119
Table 4.5 Performance of the classification models on the test dataset.....	124
Table 4.6 Comparison of the performance metrics for the SediNet classification task (Buscombe 2020) and the texturally similar GM, SP, and SM classification.....	129

LISTE DES FIGURES

	Page
Figure 1.1	Exemple de courbe granulométrique selon l'USCS..... 7
Figure 1.2	Dispositif expérimental du tamisage pour l'analyse granulométrique (Tian, 2024) 9
Figure 1.3	Processus d'analyse granulométrique par traitement d'images, 1-image du milieu granulaire, 2-mise à l'échelle de l'image, 3-délimitation des contours des particules, 4-obtention de la courbe granulométrique Tiré de Bouchard (2016)..... 11
Figure 1.4	Segmentation de l'image originale (a) comprenant l'extraction des bords (b), la conversion en image binaire (c), la segmentation (d), la détermination de la région occupée par l'échantillon (e) et la détection du matériau fin (f) Tiré de Hamzeloo et al. (2014)..... 12
Figure 1.5	Flux de travail des modèles d'apprentissage supervisé..... 14
Figure 1.6	Arbre des différents algorithmes des méthodes d'apprentissage machine Tiré de Mahesh (2019)..... 14
Figure 1.7	Principe de fonctionnement des couches convolutives 16
Figure 1.8	Principe de la convolution standard (a) vs convolution séparables en profondeur (b) Tiré de Sarkar (2021) 18
Figure 1.9	Architecture U-Net (exemple pour des images de 32x32 pixels à la plus basse résolution). Chaque boîte bleue correspond à une carte de caractéristiques multi-canaux. Le nombre de canaux est indiqué au-dessus de la boîte. La taille en x et y est précisée dans le coin inférieur gauche de la boîte. Les boîtes blanches représentent des cartes de caractéristiques copiées. Les flèches indiquent les différentes opérations Tiré de Ronneberger et al. (2015)..... 19
Figure 1.10	Résultat sur le défi de suivi cellulaire ISBI, (a) Une partie d'une image d'entrée du jeu de données "PhC-U373", (b) Résultat de la segmentation (masque cyan) avec la vérité terrain manuelle (bordure jaune), (c) Image d'entrée du jeu de données "DIC-HeLa", (d) Résultat de la segmentation (masques colorés aléatoirement) avec la vérité terrain manuelle (bordure jaune) Tiré de Ronneberger et al. (2015)..... 19

Figure 1.11	Quelques exemples d'images synthétiques avant et après leur transformation par le CycleGAN démontrant l'amélioration du réalisme visuel des images de synthèse après traitement vers le domaine réel et réciproquement Tiré de Barth et al. (2020).....	25
Figure 1.12	Illustration des deux produits finaux générés avec GRAINet sur le Rhône. À gauche : carte de la distribution spatiale des granulométries caractéristiques (D_m). À droite : courbe de granulométrie pour l'ensemble des bancs de graviers Tiré de Lang et al. (2021)	27
Figure 1.13	Exemple de tuiles d'images (1,25 m×0,5 m) avec une distance d'échantillonnage au sol de 0,25 cm. Chacune des 25 tuiles d'exemple provient d'un banc de gravier différent Tiré de Lang et al. (2021)	27
Figure 1.14	Architecture du modèle GRAINet Tiré de Lang et al. (2021).....	28
Figure 1.15	Exemple d'images utilisées pour l'entraînement de SediNet Tiré de Buscombe (2020)	29
Figure 1.16	Architecture simplifiée du modèle SediNet (Buscombe, 2020).....	30
Figure 1.17	a) et b) représentent respectivement des exemples d'images de synthèse de matériaux granulaires développés avec YADE et Unity game engine Tiré de Pirnia et al. (2018)	33
Figure 2.1	Examples and characteristics of the 3D particle models	42
Figure 2.2	Container and cameras in the Unity 3D model	43
Figure 2.3	An example of image annotation and labeling during the generation of synthetic samples using Perception.....	47
Figure 2.4	PSD for the validation images and PSD range for the segmentation dataset	48
Figure 2.5	Algorithm for synthetic image validation based on local entropy, a) Original images, b) V-channel images thresholded using the Otsu method to focus on the particle contours, c) Images entropy calculated for a neighborhood with a 2-pixel radius, d) Histogram of conditional entropy for the images	51
Figure 2.6	Simplified architecture of GrainSegNet	52

Figure 2.7	a) Sample synthetic image from the validation dataset for GrainSegNet and b) its segmentation mask, c) synthetic validation image corresponding to sample 0002 and d) its segmentation mask	55
Figure 2.8	Example of synthetic image of granular material (a) and its Perception annotations including bbox, mask and label for each particle (b)	55
Figure 2.9	Top and bottom views of real photographs and their corresponding synthetic images used for calibration	56
Figure 2.10	Histograms of the HSV color channels for the real (a), synthetic (c), and Yade-generated (e) images, corresponding to subsamples 0001 and 0002. The x-axis represents the pixel intensity values of each channel (hue: 0–179°; saturation and value: 0–255), while the y-axis indicates the number of pixels for each intensity level.....	58
Figure 2.11	V-channel images thresholded using Otsu’s method for samples 0001 (a) and 0002 (b). The top and bottom views of the samples are shown on the left and right, respectively. Images are shown for real photographs, Unity, and Yade	59
Figure 2.12	Comparison of the mean (a, c) and standard deviation (b, d) of conditional entropy for real photographs and synthetic images (Yade and Unity). The first row corresponds to the top view of sample 0001 (a, b) and the second row to the top view of sample 0002 (c, d)	59
Figure 2.13	Comparison of the mean and standard deviation of conditional entropy for real images (a), synthetic images generated with Unity (b), and synthetic images generated with Yade (c), for the top and bottom views of sample 0002	60
Figure 2.14	GrainSegNet training losses (a-f) and validation losses (g-l). Total losses are shown in subfigures a) and g), RCNN bounding box regression losses in subfigures b) and h), RPN bounding box regression losses in subfigures c) and i), RCNN classification losses in subfigures d) and j), RPN classification losses in subfigures e) and k) and mask losses in subfigures f) and l)	62
Figure 2.15	Synthetic images for the first test dataset (a) corresponding to the highest (top row) and lowest (bottom row) mean IoU scores. Images b) shows the ground truth segmentation masks generated using the Perception Toolkit and images c) shows the predicted segmentation masks	63

Figure 2.16	Distribution of IoU scores per object across the entire test dataset composed of synthetic images.....	64
Figure 2.17	Evaluation of GrainSegNet on real soil photographs. (a) original images; (b) manual segmentation using MakeSense.AI; (c) predicted segmentation generated by GrainSegNet.....	65
Figure 2.18	Histograms illustrating the distribution of Intersection over Union (IoU) scores for each segmented object (left) and for each image (right) in the test dataset with real photographs	65
Figure 3.1	Particle size distributions of a) artificial material photographs of image type 1, b) synthetic images prepared with the Unity game engine for image types 2 and 3 and c) real material photographs in image type 4.....	73
Figure 3.2	Image samples for type 1 (a and b), type 2 (c and d), type 3 (e and f), and type 4 (g and h). In each pair, the left and right images show the top-down and bottom-up views, respectively	74
Figure 3.3	Photo setup in the laboratory (a) and Unity game engine (b).....	76
Figure 3.4	Variation in the texture of two samples (a and b) from image type 4 as a function of water content.....	76
Figure 3.5	Particle size distributions (PSDs) for training and test datasets a) D4, b) D5, and c) D6. In each plot, the red curves represent the test sample PSDs, while the black curves correspond to the training and validation samples	79
Figure 3.6	Transformations applied to the training and validation images used in datasets D4 and D5. Row a) shows examples of augmented images for artificial materials (type 1), while row b) presents transformations applied to synthetic images (type 2)	79
Figure 3.7	GrainSizeNet architecture a) with two input images (top and bottom) and b) with one input image (top).....	81
Figure 3.8	Detailed flow chart for the classification of coarse-grained soils with more than 50% retained on 0.080 mm sieve with the MTQ 1101 standard.....	86
Figure 3.9	Learning curves for training on datasets a) D1, b) D2, c) D3, d) D4, e) D5, and f) D6	87
Figure 3.10	Confusion matrices for classification using GrainSizeNet trained on a) D1, b) D2, and c) D3, evaluated on the same test dataset.....	88

Figure 3.11	Confusion matrices for classification using GrainSizeNet trained on D1 (a), D2 (b), and D3 (c), evaluated on the same test dataset	89
Figure 3.12	Representative test samples and predicted vs. ground truth granulometric curves for GrainSizeNet trained on D1, D2, and D3 datasets. SubFigures (a)–(d) correspond to real samples (types 1 and 2), while (e) and (f) represent synthetic samples (types 3 and 4). For each sample, the top-view image is shown alongside the predicted and reference granulometric curves. Note that for models trained on D1, both top- and bottom-view images were used during inference, even though only the top-view image is displayed here for visualization	91
Figure 3.13	Comparison of the actual and predicted cumulative percentages passing for all sieves of the test samples using GrainSizeNet trained on a) D1, b) D2, and c) D3	92
Figure 3.14	(a, e) Example test-sample images with actual and predicted particle size curves using GrainSizeNet trained on (b, f) D4, (c, g) D5, and (d, h) D6.....	94
Figure 3.15	Confusion matrices for classification using GrainSizeNet trained on D4 (a), D5 (b), and D6 (c), evaluated on the same test dataset, (d) and (e) present the extended confusion matrices for the models trained on D5 and D6, respectively, including all classes	94
Figure 3.16	Evolution of mean absolute error by group and training folds.....	97
Figure 3.17	Confusion matrices illustrating the classification performance of GrainSizeNet with respect to MTQ 1101 classes, evaluated across the different test folds used during cross-validation.....	99
Figure 4.1	USCS class distribution for the full dataset.....	108
Figure 4.2	Example of split-spoon photograph subsampling and resizing with a) original image, and b) and c) selected subsamples cropped and resized to 299×299 pixels	109
Figure 4.3	Example of the different augmentation steps applied to a training image. The original images were modified by horizontal and vertical flipping (left-right and up-down), color, brightness and contrast enhancement, rotation, Gaussian blur and hue shift to enrich the diversity of the dataset	110

Figure 4.4	Image samples taken from subdataset C1. Each column corresponds to a USCS class, with two representative images per class. The axes are labeled in pixels111
Figure 4.5	Examples of images from the R1 dataset, accompanied by the proportion of fine particles114
Figure 4.6	Particle size distributions of a) training/validation dataset R1 and b) the corresponding test dataset. The particle size distributions only show the sieving results115
Figure 4.7	Network architecture for a) GrainSizeNet (Simo Tenekam et al., 2025) and b) the networks trained in this paper. Red letters indicate parameters that were set following the hyperparameter study, W is the kernel size for the convolutional layers, X indicates the SpatialDropout2D probability, Y ₁ , Y ₂ , Y ₃ and Y ₄ are the number of neurons in each fully connected layer and Z the dropout probability for the fully connected layers117
Figure 4.8	Evolution of training and validation losses during model learning for a) one-step classification with subdataset C1, and hierarchical models corresponding to subdatasets b) C2, c) C3, d) C4, and e) C5123
Figure 4.9	Training and validation loss for the regression models predicting the proportion of fine particles for a) all materials using subdataset R1, b) fine-grained materials using subdataset R2, and c) coarse grained materials using subdataset R3123
Figure 4.10	Confusion matrices illustrating the classification results for the test dataset according to a) the single-step and b) hierarchical approaches125
Figure 4.11	Comparison of predicted and actual mass proportions of particles passing the 75 μm sieve for a) the single-step model trained on R1 and b) the hierarchical model combining classification with C2 and regression with R2 for fine soils and R3 for coarse soils126
Figure 4.12	Sample images from the regression test dataset with the actual and predicted mass proportions of fine particles for the single-step and hierarchical regression models. The images in the first and second columns correspond to fine and coarse materials, respectively127
Figure 4.13	Normalized confusion matrices for a) C1 and b) the hierarchical models tested on the GM, SM, and SP classes, compared with c) Buscombe's (2020) SediNet model129

Figure 4.14	Typical appearances of fine-grained soils after separating the split-spoon for a) CH and b) CL, and after manual handling by the technician for field classification for c) CH and (d) CL.....	130
-------------	--	-----

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

CCE	Categorical Cross Entropy
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
IoU	Intersection over Union
MAE	Mean Absolute Error
mAP	Mean Average Precision
MSE	Mean Squared Error
PSD	Particle Size Distribution
ReLU	Rectified Linear Units
RMSE	Root Mean Squared Error
SIFT	Scale-Invariant Feature Transform
SPT	Standard Penetration Test
USCS	Unified Soil Classification System

LISTE DES SYMBOLES ET UNITÉS DE MESURE

Cc	Curvature coefficient
ρ	Density (kg/m ³)
$\Delta\rho$	Difference in density between particle and fluid (g/cm ³)
μ	Dynamic Viscosity (g/cm.s)
D ₁₀	Grain sizes corresponding to percentages passing of 10% (mm)
D ₃₀	Grain sizes corresponding to percentages passing of 30% (mm)
D ₆₀	Grain sizes corresponding to percentages passing of 60% (mm)
D _m	Mean particle diameter (mm)
g	Gravity acceleration (m/s ²)
D	Particle diameter (cm)
K	Permeability coefficient (m/s)
n	Porosity
V _s	Sedimentation velocity (cm/s)
Cu	Uniformity coefficient

INTRODUCTION

La compréhension des propriétés mécaniques du sol est primordiale durant la phase de conception d'un projet pour garantir la sécurité d'une infrastructure. L'analyse granulométrique est fondamentale pour caractériser un sol afin de déterminer les proportions qu'occupent les différentes classes de grosseurs de particules qui le composent (la granulométrie). Ce paramètre permet d'estimer certains comportements tels que la compacité ou même la perméabilité des matériaux. Conventionnellement, les analyses granulométriques sont réalisées au laboratoire par des méthodes manuelles à l'exemple du tamisage et de la sédimentométrie (ASTM International, 2017; ASTM International, 2021). Bien qu'elles soient reconnues pour leur précision, elles demeurent laborieuses, chronophages et suffisamment coûteuses.

Dans un contexte industriel où l'efficacité et la réduction des coûts deviennent des priorités, l'émergence des technologies d'analyse d'image est une perspective pour optimiser ces méthodes traditionnelles (Duhaime et al., 2021). Il existe de nombreux outils commerciaux pour déterminer la granulométrie à partir d'images segmentées (Maerz et al., 1996; Raina et al., 2002; Souza, 2018). Ces outils reposent sur des algorithmes de segmentation, comme ceux basés sur la méthode des bassins versants pour délimiter les contours des fragments visibles de l'image. Les surfaces des contours segmentés sont ensuite analysées pour estimer la composition granulaire du milieu photographié. Néanmoins, la détection précise des particules fines (diamètre inférieur à 0.08 mm) demeure un défi en raison des limitations de performance des algorithmes de détection des contours associées à des enjeux tels que la taille minimale de particule détectable (WipWare Inc., n.d.).

Les techniques d'apprentissage en profondeur ont démontré une grande efficacité à effectuer diverses tâches de vision par ordinateur (Alom et al., 2019; LeCun et al., 2015). L'adoption des architectures de réseaux de neurones convolutionnels (CNN) est adéquate à la caractérisation des particules fines (Shrivastava et al., 2022). Contrairement aux méthodes de segmentation, l'analyse granulométrique à l'aide de CNN ne nécessite pas une identification visuelle claire et distincte des particules dans l'image. En effet, les CNN sont capables de se

concentrer sur l'apparence globale de l'image, en s'abstenant de dépendre des motifs texturaux de petite échelle (Basha et al., 2020). Cette approche est palliative au problème de détection des particules fines rencontré par les outils d'analyse granulométrique par segmentation d'images.

L'application des techniques d'apprentissage en profondeur repose sur la disponibilité de vastes ensembles de données. Pour cette raison, quatre types d'images ont été utilisés dans cette thèse. Un jeu de données constitué de milliers de matériaux synthétiques préparés avec Unity game engine. Ces matériaux ont pu être segmentés et annotés automatiquement à l'aide de Perception Toolkit, donnant lieu à un second ensemble de données d'images segmentées. En complément des données de synthèse, des photographies ont été réalisées au laboratoire et sur le terrain afin de créer des jeux de données de matériaux artificiels et réels. Les travaux de Duhaime et al. (2021) et Nowruzi et al. (2019) ont démontré qu'une combinaison contrôlée de données réelles et de données synthétiques réalistes pourrait améliorer significativement les performances des modèles d'apprentissage supervisé.

Les systèmes de classification normalisée des sols facilitent la communication entre les professionnels du génie, mais leur dépendance aux analyses en laboratoire limite l'identification directe des sols sur le chantier. Le système unifié de classification des sols (USCS) se distingue en s'appuyant principalement sur la granulométrie et les limites d'Atterberg (ASTM International, 2020). L'intégration des méthodes novatrices d'analyse d'image sur le terrain permettrait la classification et la description directe d'échantillons de sol, notamment ceux prélevés à l'aide de cuillères fendues (CF) lors de l'essai de pénétration standard (SPT). Cette approche représente une opportunité de liaison entre le développement des méthodes novatrices et leur application sur les chantiers d'investigation géotechnique. Les méthodes d'apprentissage par transfert offrent un moyen de transposer les modèles entraînés avec des données issues de conditions contrôlées en laboratoire à des données brutes directement collectées au chantier (Yosinski et al., 2014).

Le principal objectif de cette thèse était de développer une série de modèles pour déterminer la granulométrie des sols avec les techniques d'apprentissage profond supervisé. Cet objectif a été subdivisé en trois objectifs secondaires. Le premier visait à constituer une banque de données d'images de sols associés à leurs granulométries respectives. Ceci permettrait de disposer de données complètes pour l'entraînement et la validation des modèles. Le second portait sur le développement d'une architecture de CNN capable de prédire la granulométrie en utilisant les images de synthèse et les photographies réelles. Deux variantes de ce modèle ont été proposées : la première intégrant des structures adaptées pour exploiter à la fois les images prises au-dessus et en dessous des échantillons et la seconde se limitant exclusivement à l'utilisation des images capturées au-dessus des échantillons. Enfin, le dernier objectif visait à adapter le modèle pour la caractérisation de sols directement à partir de photographies d'échantillons obtenus dans des CF malgré les contraintes inhérentes aux conditions de terrain et aux variations de qualité des images.

Pour répondre à ces objectifs, cette thèse adopte une approche méthodologique subdivisée en plusieurs chapitres. Le chapitre 1 est dédié à une revue de la littérature qui aborde divers aspects liés au sujet de l'étude. La première section contextualise l'importance de l'analyse granulométrique en passant en revue les méthodes traditionnelles et novatrices récemment proposées dans la littérature. Une attention particulière est portée sur l'essor de l'apprentissage profond dans le domaine des sciences des matériaux et spécifiquement sur des tâches de classification et de régression appliquées aux sols. Une section supplémentaire examine la plus-value des données synthétiques au vu des difficultés de collecte des données dans divers domaines de vision par ordinateur aussi bien qu'en géotechnique (p. ex. l'indisponibilité des données et le coût important d'acquisition qui y est associé). Un regard est également porté sur les architectures de CNN afin de surmonter les contraintes liées à la faible variabilité texturale des images de matériaux granulaires. Après avoir cerné les enjeux théoriques, une discussion est proposée sur l'adaptation des CNN pour le contexte de terrain à l'aide de méthodes comme l'apprentissage par transfert ainsi que les limitations de ces méthodes.

Les chapitres 2 à 4 correspondent chacun à un article scientifique soumis pour publication dans des revues à comité de lecture. Le chapitre 2 (article soumis à *Applied Computing and Geosciences*) décrit la méthodologie utilisée pour la création d'images synthétiques de matériaux granulaires. Le moteur de jeu vidéo Unity game engine a été choisi au vu du vaste champ de possibilités qu'il propose en termes de modélisation 3D. Plusieurs paramètres comme la gravité, la masse, les déplacements, les collisions et les coefficients de frottement (statique et dynamique) sont directement contrôlables sur le moteur. Afin de gérer le dynamisme de l'environnement 3D, des algorithmes en C# ont été utilisés, en particulier pour simuler le déversement d'échantillons de matériaux dans un contenant, les photographies des échantillons et le nettoyage de la scène de modélisation. Une boîte à outils native de Unity (le Perception Toolkit) a été exploitée pour automatiser le processus d'annotation et étiquetage des images. L'annotation permet de sauvegarder pour chaque image, les contours des particules visibles par la caméra à des fins de segmentation. L'étiquetage consiste quant à lui à attribuer une classification spécifique à chaque modèle 3D. Par exemple, en désignant une particule comme appartenant à une classe correspondante (sable, gravier, etc.).

Le chapitre 3 (article soumis à *Acta Geotechnica*) récapitule les phases de conception et d'entraînement de GrainSizeNet, une architecture de CNN proposée pour l'analyse granulométrique à partir de données collectées dans des environnements contrôlés. L'ensemble des images provient à la fois du moteur Unity game engine, de matériaux préfabriqués selon des granulométries ciblées au laboratoire, ainsi que certains matériaux réels issus de projets d'investigation géotechnique. Une fois les échantillons photographiés, l'analyse granulométrique a été faite par tamisage en utilisant 11 tamis de contrôle référencés à la norme BNQ 2501-025 (2013) avec des ouvertures de mailles comprises entre 0.80 mm et 31.5 mm. Deux architectures ont été proposées pour évaluer l'impact de la ségrégation des matériaux (Bashir & Jumba, 2020; Deleuil, 1993; Ottino & Khakhar, 2000) lors de la prise de photographies et la plus-value des données synthétiques.

Le chapitre 4 (article soumis à *Canadian Geotechnical Journal*) évalue les performances de GrainSizeNet sur les photographies de sols contenus dans des cuillères fendues utilisées lors

des essais SPT. Un jeu de données de 1206 échantillons a été exploité pour adapter le modèle aux données brutes collectées sur le terrain en recourant à un ajustement de l'architecture du modèle et lorsque pertinent à l'apprentissage par transfert. GrainSizeNet a été entraîné pour répondre à la fois à des tâches de classification et de régression. Cinq classes de matériaux du système USCS ont été utilisées pour la classification. Ces classes (CL, CH, SM, GM et SP) ont été sélectionnées dans une banque de données d'entreprise, reflétant également les classes de matériaux les plus courantes au Québec. Par la suite le modèle a été entraîné pour estimer la proportion de particules fines présentes sur une image. La proportion de particules fines dans un sol soulève généralement des enjeux environnementaux face à la réutilisation des sols durant les travaux de construction.

Le chapitre 5 conclut cette thèse par une discussion des résultats obtenus et propose des recommandations pour les travaux futurs.

CHAPITRE 1

REVUE DE LITTÉRATURE

Ce chapitre présente une revue du concept d'analyse granulométrique des sols, en détaillant les méthodes conventionnelles (tamisage et sédimentométrie) utilisées pour sa détermination. Les limitations liées à ces méthodes traditionnelles sont abordées, suivies d'une présentation d'approches et de logiciels commerciaux qui reposent sur un processus de traitement d'images de sols conçues pour pallier ces contraintes. Une section particulière se concentre sur l'exploitation des méthodes d'apprentissage supervisé pour des tâches de vision par ordinateur, plus spécifiquement pour l'analyse des tailles de particules en passant en revue les méthodes existantes, les contraintes associées, ainsi que les solutions et stratégies développées pour améliorer leur applicabilité et leur robustesse. La dernière section conclut par une synthèse critique justifiant l'approche proposée dans cette thèse.

1.1 Analyse granulométrique

L'analyse granulométrique d'un sol permet de déterminer les pourcentages massiques de différentes fractions de tailles de particule normalisées qui le composent. C'est une méthode à l'issue de laquelle une courbe granulométrique est obtenue, illustrant les pourcentages de tamisats cumulés en fonction des tailles des particules (Figure 1.1). Elle renferme une source d'informations pour estimer les propriétés hydrauliques (p. ex. perméabilité) ou mécaniques (p. ex. la compacité) du sol.

L'utilisation qualitative de la courbe granulométrique permet d'évaluer la gradation du sol. Un sol est considéré comme bien gradué s'il couvre une large gamme de tailles de particules favorisant ainsi une meilleure compacité. À l'inverse, les sols à granulométrie uniforme ou mal gradués caractérisés par la prédominance d'une fraction spécifique sont le plus souvent difficiles à compacter efficacement. La distribution granulométrique d'un sol peut être caractérisée à partir des coefficients d'uniformité (C_u) et de courbure (C_c) qui décrivent respectivement l'étalement et la continuité des tailles de grains et sont définis comme suit :

$$C_u = \frac{D_{60}}{D_{10}}, C_c = \frac{D_{30}^2}{D_{10} \cdot D_{60}} \quad (1.1)$$

où D_{30} et D_{60} représentent respectivement les tailles de particules pour lesquelles 30% et 60% de la masse totale de l'échantillon ont des grains de taille inférieure ou égale. La gradation des sols sera associée à certaines conditions que doivent respecter les coefficients C_u et C_c (p. ex. un sable sera bien gradué si $C_u > 6$ et C_c compris entre 1 et 3)

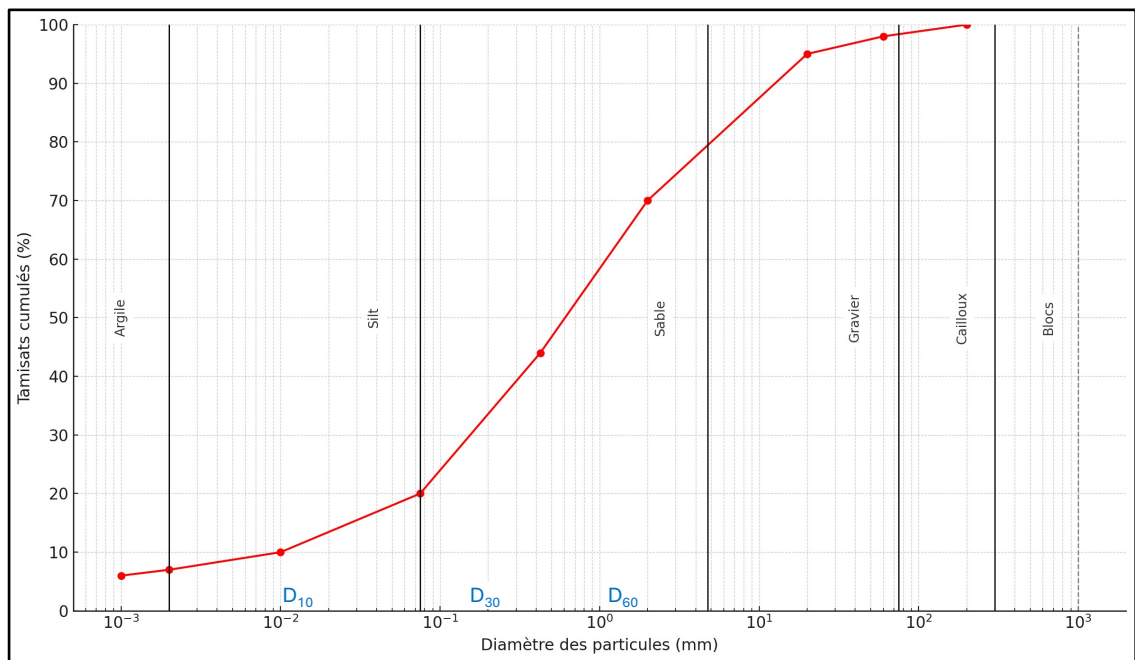


Figure 1.1 Exemple de courbe granulométrique selon l'USCS

1.2 Tamisage et sédimentométrie

1.2.1 Tamisage

Parmi les différentes méthodes d'analyse granulométrique, le tamisage est certainement la méthode la plus utilisée dans les laboratoires d'analyse géotechnique. Cette méthode repose sur la détermination des quantités de sol contenues dans différentes fractions de taille. Les particules sont triées à l'aide de tamis équipés de mailles d'ouverture standardisée. La taille de

maille est relative au choix de la norme utilisée, parmi lesquelles on peut citer les normes ASTM International (2021), BNQ 2501-025 (2013) et ISO 17892-4 (2016).

Le dispositif expérimental du tamisage comprend plusieurs tamis empilés en série du plus fin à celui avec l'ouverture de maille la plus grossière (Figure 1.2). Ce système est monté et fixé sur un vibreur mécanique qui permet de favoriser le passage des particules à travers les mailles en effectuant des secousses à fréquence contrôlée. Cette configuration minimise potentiellement l'erreur humaine associée à l'agitation manuelle des tamis tout en garantissant une répartition homogène des particules et en améliorant la reproductibilité des résultats.

Cette méthode présente certaines limitations. Elle peut vite devenir chronophage pour les échantillons volumineux qui nécessitent une réduction de masse préliminaire par des techniques d'échantillonnage comme le quartage, le séparateur à fentes etc. (Head, 1980). De plus, Holtz et al. (1981) ont souligné le biais relatif aux particules de densité différentes. Les particules légères mais volumineuses peuvent avoir des propriétés mécaniques différentes des particules denses de même taille. Puisque le tamisage repose uniquement sur un principe de rapport de masse pour chaque fraction, la densité des particules devrait être prise en compte pour assurer une bonne caractérisation granulométrique.

Le tamisage devient inefficace pour les particules plus petites que 0.075 mm (Head, 1980). Les sols cohésifs en raison de leur plasticité (p. ex. argile ou silt) ont une forte tendance à former des agrégats qui ne se séparent pas facilement par tamisage même avec le dispositif du vibrant mécanique (Terzaghi et al., 1996). Les particules fines agglomérées peuvent se fixer à tort sur les tamis de tailles plus grossières que leur taille réelle faussant ainsi la répartition granulaire mesurée. Ces limitations rendent le tamisage inapproprié pour les sols fins nécessitant ainsi des méthodes alternatives comme la sédimentométrie.



Figure 1.2 Dispositif expérimental du tamisage pour l'analyse granulométrique
Tiré de Tian (2024)

1.2.2 Sédimentométrie

La sédimentométrie est une méthode d'analyse granulométrique complémentaire au tamisage. Son but est de déterminer la distribution granulaire des particules fines en mesurant leur vitesse de sédimentation. C'est un essai standardisé dont le processus de réalisation est spécifié selon différentes normes (AASHTO T88, 2022; ASTM International, 2021; BNQ 2501-025, 2013; ISO 17892-4, 2016). La loi de Stokes (Stokes, 1851) permet de lier la vitesse de sédimentation des matériaux en fonction du diamètre des grains :

$$V_s = \frac{D^2 \Delta \rho \cdot g}{18\mu} \quad (1.2)$$

où D représente le diamètre des grains en cm, g l'accélération gravitationnelle en cm/s^2 , μ la viscosité du fluide ($\text{g}\cdot\text{cm}^{-1}\cdot\text{s}^{-1}$) et $\Delta\rho$ est la différence entre les masses volumiques des particules et du fluide ($\text{g}\cdot\text{cm}^{-3}$).

Une quantité de matériau sec comprise entre 50 g et 100 g est disposée dans un agent dispersant comme l'hexa-métaphosphate de sodium afin de désagréger les particules agglomérées. Le mélange est ensuite versé dans une éprouvette contenant de l'eau déminéralisée. La mesure de la densité est prise à l'aide d'un hydromètre (densimètre) pendant des intervalles de temps spécifiques. La température du mélange, le temps de sédimentation ainsi que les lectures hydrométriques permettent de déduire les différentes tailles de particules fines qui composent le matériau.

En dépit de ses avantages pour la caractérisation des sols fins, cette méthode présente des contraintes dues à la complexité des mesures des particules d'argile. La non-dispersion des particules de très petite taille (diamètre équivalent inférieur à 0.002 mm) de haute plasticité est un problème fréquent malgré le recours à l'hexa-métaphosphate de sodium (ASTM International, 2021). De plus, la sédimentométrie nécessite une bonne stabilité du dispositif expérimental pour éviter les turbulences susceptibles d'interférer avec les mesures de la vitesse de sédimentation. Bien qu'elle permette d'estimer la distribution granulométrique des particules fines, la plasticité influence davantage les propriétés géotechniques que la seule taille des particules. Ces deux paramètres ne sont pas systématiquement corrélés dans les matériaux fins. Ainsi, l'utilisation des limites d'Atterberg est souvent privilégiée pour caractériser ces sols, permettant une meilleure estimation de leur comportement mécanique (ASTM International, 2018; ASTM International, 2020).

1.3 Logiciels commerciaux d'analyse granulométrique

La granulométrie des matériaux grossiers est de plus en plus déterminée à partir de la segmentation des photographies collectées sur le terrain (Bouchard, 2016; Nanda & Pal, 2020) (Figure 1.3). La segmentation d'images est une méthode de regroupement de pixels qui permet

de classifier tous les pixels d'une image en y affectant à chacun une étiquette. La distribution statistique des zones segmentées peut être reliée à la distribution de la taille des particules (Yaghoobi et al., 2019). Cette méthode est couramment utilisée dans l'industrie minière pour la mesure de la fragmentation des roches (l'estimation de la distribution granulométrique des fragments) (Maerz et al., 1996). C'est le principal paramètre utilisé pour évaluer les performances du dynamitage.

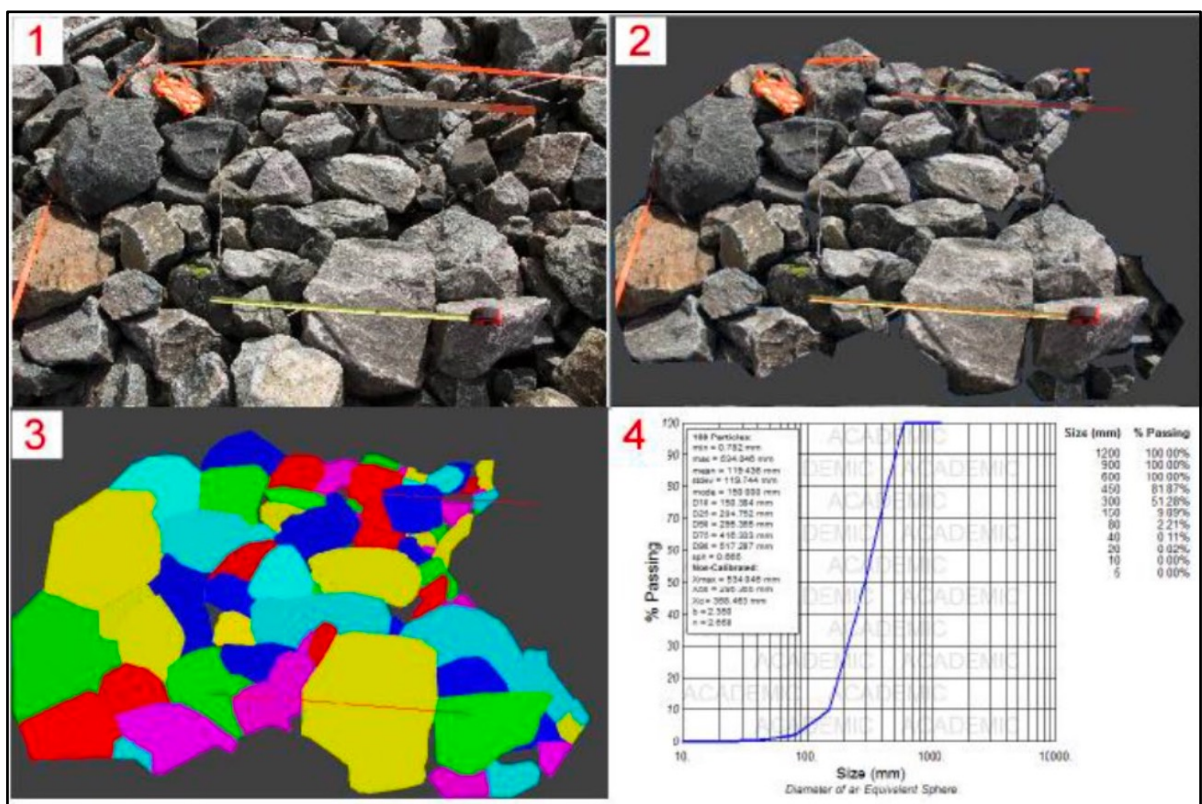


Figure 1.3 Processus d'analyse granulométrique par traitement d'images, 1-image du milieu granulaire, 2-mise à l'échelle de l'image, 3- délimitation des contours des particules, 4- obtention de la courbe granulométrique

Tiré de Bouchard (2016)

Wipfrag est un logiciel d'analyse d'image qui fournit la mesure de la fragmentation à partir d'images numériques collectées par drones, ou sur un milieu granulaire après une explosion (Maerz et al., 2018). Des algorithmes de traitement d'image permettent de différencier les particules rocheuses des zones fines qui ne sont pas prises en compte pour la caractérisation.

Un ajustement manuel des contours délimités par le logiciel peut se faire en cas de présence de contours erronés sur l'image segmentée (Figure 1.4). Les surfaces délimitées par ces contours sont transposées par des outils statistiques pour estimer la granulométrie. Cependant, l'utilisation des images présentent certaines limitations dues à la faible résolution, au problème de distorsion de perspective et la détection efficace des zones fines (Noy, 2012; Sanchidrián et al., 2009).

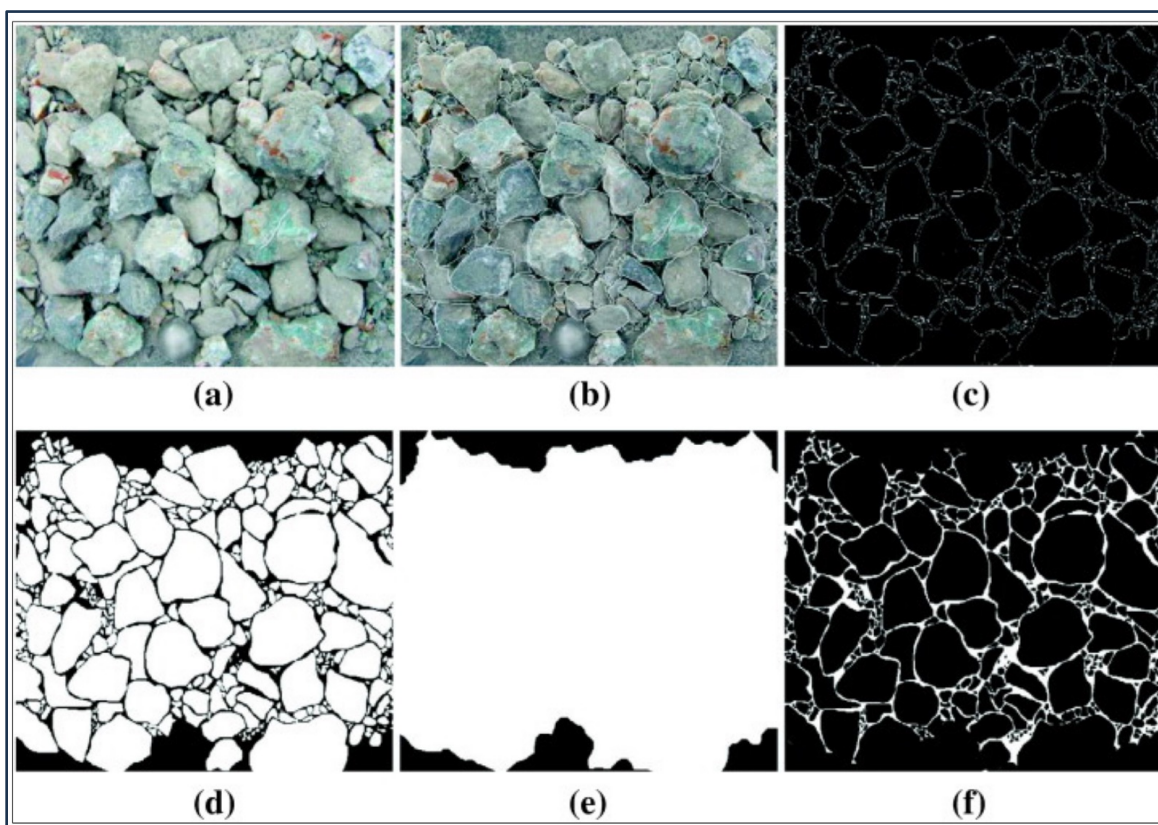


Figure 1.4 Segmentation de l'image originale (a) comprenant l'extraction des bords (b), la conversion en image binaire (c), la segmentation (d), la détermination de la région occupée par l'échantillon (e) et la détection du matériau fin (f)

Tiré de Hamzeloo et al. (2014)

Split Online et Split Desktop sont des logiciels similaires adaptés respectivement au terrain et au laboratoire. Ils utilisent des technologies de caméra à grande vitesse pour la mesure de la fragmentation. La granulométrie est estimée en analysant la forme et la répartition des particules (BoBo, 2001). Une étude comparative menée par Ouchterlony et al. (2007) a révélé

que les deux modèles présentent des résultats similaires pour la détermination des D_{50} et D_{80} . La version de terrain peut être influencée par les conditions d'éclairage pouvant affecter jusqu'à 50% les prédictions de particules fines. En revanche, les images de Split Desktop nécessitent préalablement 20 minutes pour ajuster manuellement les contours des particules.

Il existe de nombreux autres logiciels commerciaux d'analyse de fragmentation tels que Fragalyst ou Fragscan qui permettent directement d'analyser les matériaux fragmentés en mouvement sur des convoyeurs (Raina et al., 2002; Schleifer & Tessier, 2018). Toutes ces méthodes présentent l'avantage d'être adaptées au traitement rapide de grandes quantités de matériaux. Liu & Tran (1996) ont comparé différents systèmes de mesure de fragmentation mettant en lumière leurs avantages et inconvénients en soulignant que si certains systèmes offrent une bonne précision pour la caractérisation des fragments moyens et gros, aucun ne permettait de mesurer efficacement les particules fines. Dans cette perspective, Campbell & Thurley (2017) ont proposé une nouvelle approche basée sur le scanner laser 3D pour surmonter les limitations de distorsion et d'éclairage des outils commerciaux existants utilisant les images 2D. Cependant, bien que cette technique permette de détecter et regrouper les fines, elle ne résout pas entièrement la problématique de leur caractérisation précise, ce qui a motivé plusieurs chercheurs à explorer des méthodes plus novatrices et complètes.

1.4 Analyse granulométrique par techniques d'apprentissage supervisé

1.4.1 Revue des méthodes d'apprentissage supervisé

L'essor des méthodes d'apprentissage supervisé dans les domaines des sciences ne cesse de révolutionner les méthodes traditionnelles. Bishop (2006) définit l'apprentissage supervisé comme un cadre d'apprentissage qui repose sur des paires de données entrée-sortie dans le but de permettre à un modèle de généraliser pour faire des prédictions sur de nouvelles données inédites (Figure 1.5). Le modèle apprend sur des données étiquetées pour ajuster ses paramètres afin de faire des prédictions sur de nouvelles données non étiquetées (Zhou, 2021). L'apprentissage supervisé est généralement utilisé pour réaliser des tâches de classification et

de régression. Il représente l'une des principales catégories d'algorithmes de l'apprentissage machine parmi plusieurs comme illustré à la Figure 1.6 (Mahesh, 2019).

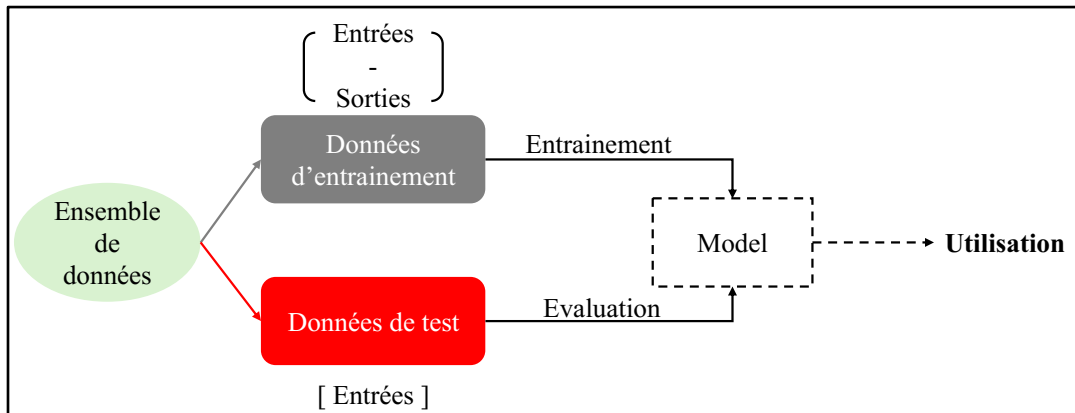


Figure 1.5 Flux de travail des modèles d'apprentissage supervisé

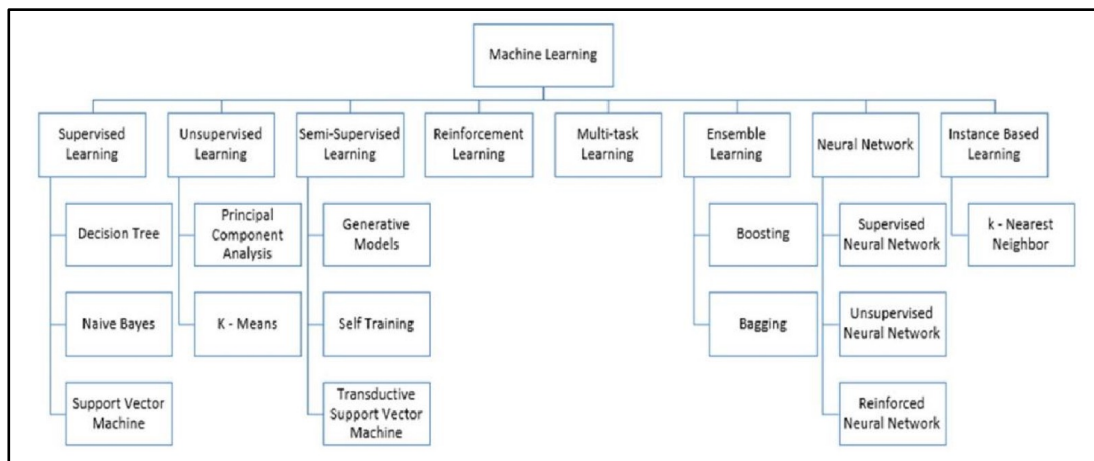


Figure 1.6 Arbre des différents algorithmes des méthodes d'apprentissage machine
Tiré de Mahesh (2019)

L'importance de la préparation des données est essentielle pour optimiser les performances des modèles d'apprentissage supervisé (I. Goodfellow et al., 2016). Cette étape permet aux modèles d'apprendre sur des motifs pertinents plutôt que de s'adapter aux artéfacts et aux bruits présents dans l'information utilisée. Les données mal étiquetées ou ayant de faibles corrélations entre les caractéristiques des données d'entrée et la cible à prédire peuvent induire des limitations sévères au processus d'entraînement des modèles (Budach et al., 2022). Pour

une meilleure généralisation des modèles sur de nouvelles données inédites, il est important de s'assurer que les données soient adaptées à la tâche ciblée (Anouar, 2023).

Les images brutes contiennent généralement beaucoup d'informations (pixels) dont une minorité est généralement nécessaire pour les tâches de classification ou de régression. Les méthodes d'extraction de caractéristiques permettent de concentrer l'analyse d'image sur les pixels d'intérêt selon l'objectif du modèle. Traditionnellement, l'extraction des caractéristiques d'une image se faisait par des descripteurs de texture. Les paramètres texturaux comme le SIFT (Scale-Invariant Feature Transform), l'histogramme des niveaux de gris, les ondelettes de Haar et les textures de Haralick étaient privilégiées pour définir quelles caractéristiques à extraire de l'image (Haar, 1909; Haralick, 1979; Lowe, 2004; Shannon & Weaver, 1949). Toutefois, le choix de ces paramètres avait un impact direct sur les performances des modèles d'apprentissage supervisé au vu de la corrélation étroite existante entre les données d'entrée et les performances des modèles (Temimi, 2020).

Les réseaux de neurones convolutionnels (CNN) sont dotés de couches spéciales qui leur confèrent une particularité d'automatiser l'extraction des caractéristiques pertinentes de l'image (Figure 1.7). Cette autonomie permet de surpasser les approches classiques d'extraction de caractéristiques (Krizhevsky et al., 2012). En outre, les CNN sont moins sensibles aux variations non pertinentes des données d'entrée comme les variations d'éclairage, d'orientation ou même le bruit visuel (Goodfellow et al., 2016). Le traitement de l'image se fait à différents niveaux de la chaîne de convolution du modèle permettant à chaque fois de se concentrer sur différents marqueurs d'intérêt de l'image. Toutefois, une sélection adéquate de la profondeur des couches de convolution ainsi que la taille de leurs filtres est nécessaire pour améliorer les performances des modèles selon la tâche ciblée.

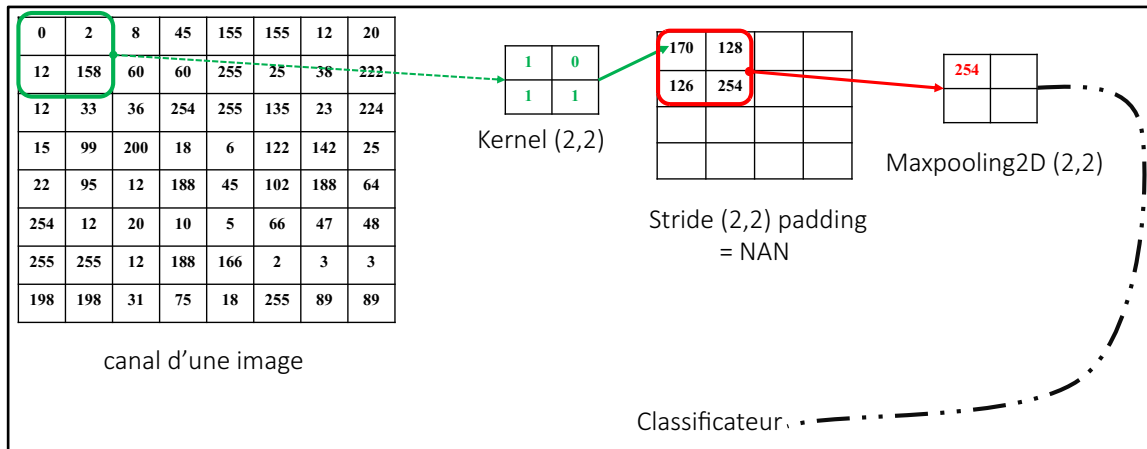


Figure 1.7 Principe de fonctionnement des couches convolutives

De nombreux modèles d'apprentissage automatique se sont distingués par l'innovation architecturale qu'ils apportent ainsi que leurs performances sur des ensembles de données standardisés comme ImageNet. Les modèles d'apprentissage en profondeur comme VGG ont démontré que l'empilement de plusieurs couches convolutionnelles permet d'extraire des caractéristiques plus complexes à partir des images en réduisant les erreurs de classification (Simonyan & Zisserman, 2015). VGG se distingue par l'utilisation des couches de convolution homogènes, appliquant systématiquement des filtres de taille 3×3 . L'objectif était de simplifier le traitement de l'information en rendant le modèle adaptable facilement à d'autres jeux de données (Qawaqneh et al., 2017; Tammina, 2019; S. Zhou et al., 2018).

He et al. (2016) ont introduit le concept de connexion résiduelle (encore appelé bloc résiduel) avec le modèle ResNet qui a joué un rôle significatif face au problème de dégradation de performances des modèles profonds. Un bloc résiduel repose sur l'addition d'une série de transformations convolutionnelles $F(x)$ à une version de son entrée x , transmise soit directement soit après une transformation linéaire légère (par exemple une convolution 1×1) afin d'assurer la compatibilité dimensionnelle. Ce mécanisme permet au réseau d'apprendre d'une fonction résiduelle (p. ex. $F(x)+x$) plutôt qu'une transformation directe. Ce mécanisme facilite la propagation du gradient lors de l'entraînement des modèles profonds et permet l'empilement de réseaux plus profonds sans dégradation des performances ni perte d'information associée au gradient évanescent.

Le modèle Inception proposé par Szegedy et al. (2015) se caractérise par une architecture à analyse multi-échelle en renfermant des blocs composés de plusieurs couches convolutionnelles ayant des tailles de filtres variées. Cette approche permet d'optimiser la représentation de l'image par les modèles tout en limitant en même temps le nombre de paramètres. Une évolution de ce modèle a été proposée par Chollet (2017) en introduisant le concept de convolutions séparables en profondeur, soit un système de décomposition qui simplifie la convolution classique en 2 opérations distinctes, notamment une convolution indépendante appliquée sur chaque canal et une convolution 1x1 pour fusionner les canaux de l'image (Figure 1.8). Ce mécanisme favorise la réduction du nombre de paramètres et réduit par la même occasion le temps de calcul requis.

La littérature rapporte de nombreuses adaptations de réseaux de neurones convolutionnels pour le regroupement de pixels basé sur certains critères spécifiques comme la couleur et la texture (Liu et al., 2021; Minaee et al., 2020; Picon et al., 2022). Ces modèles sont applicables pour des tâches de classification des pixels par catégorie (segmentation sémantique), de distinction d'objets individuels (segmentation par instances) ou même de détection d'objets visant à localiser et identifier les objets présents dans une image (Hafiz & Bhat, 2020; Salaudeen, 2023; Wołk & Tatara, 2024). La segmentation d'images a connu des avancées significatives au cours des dernières décennies avec l'introduction de modèles d'apprentissage profond. Parmi eux, des modèles comme U-net proposé par Ronneberger et al. (2015) et doté d'une structure encodeur-décodeur (Figure 1.9) ont révolutionné la segmentation d'images biomédicales. Son efficacité a été démontrée par un IoU (Intersection over Union) à 92% qui reflète un fort chevauchement entre les segmentations prédites et réelles (Figure 1.10).

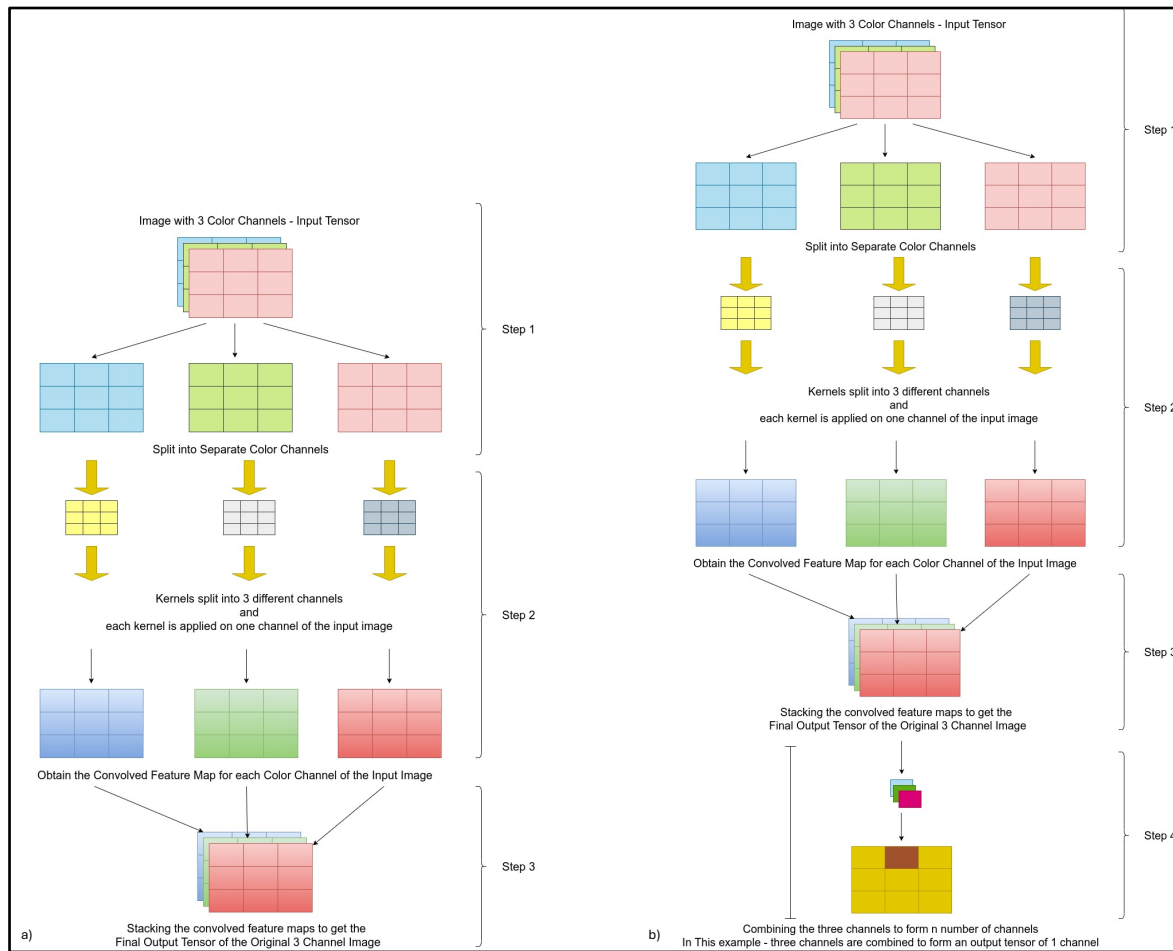


Figure 1.8 Principe de la convolution standard (a) vs convolution séparables en profondeur (b)
Tiré de Sarkar (2021)

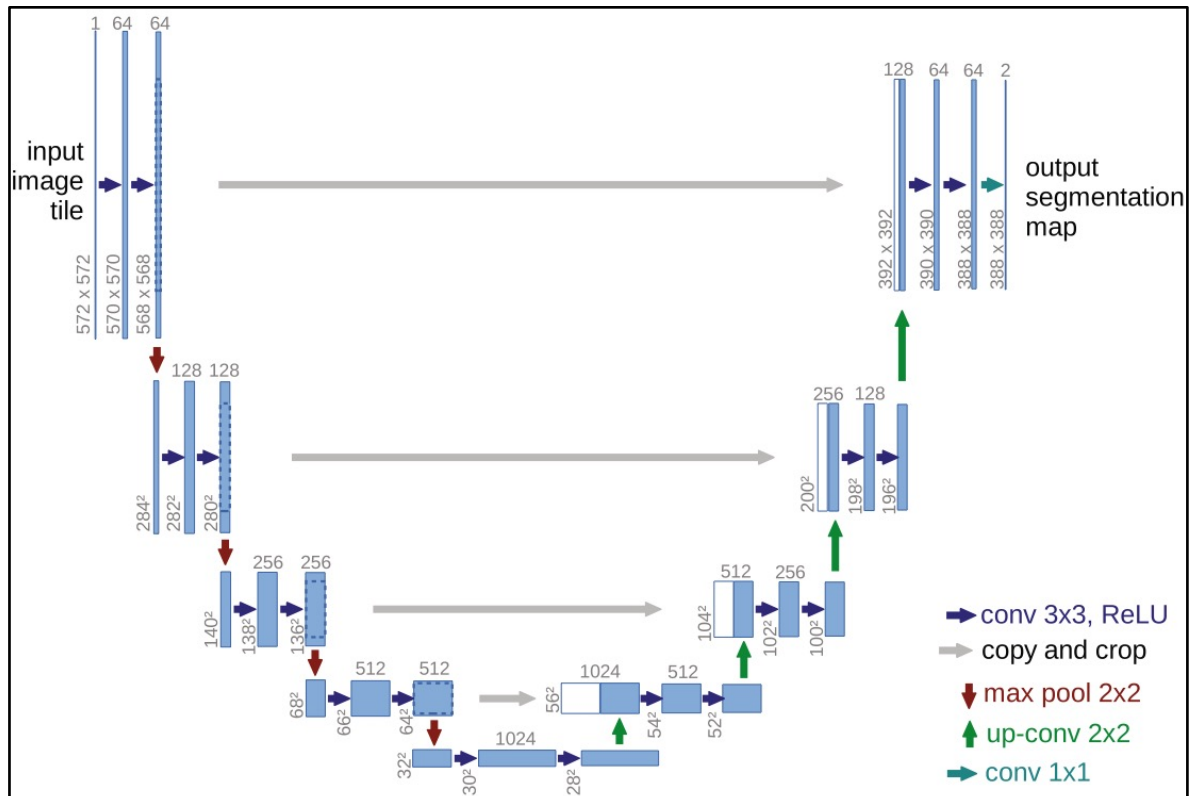


Figure 1.9 Architecture U-Net (exemple pour des images de 32×32 pixels à la plus basse résolution). Chaque boîte bleue correspond à une carte de caractéristiques multi-canaux. Le nombre de canaux est indiqué au-dessus de la boîte. La taille en x et y est précisée dans le coin inférieur gauche de la boîte. Les boîtes blanches représentent des cartes de caractéristiques copiées. Les flèches indiquent les différentes opérations
Tiré de Ronneberger et al (2015)

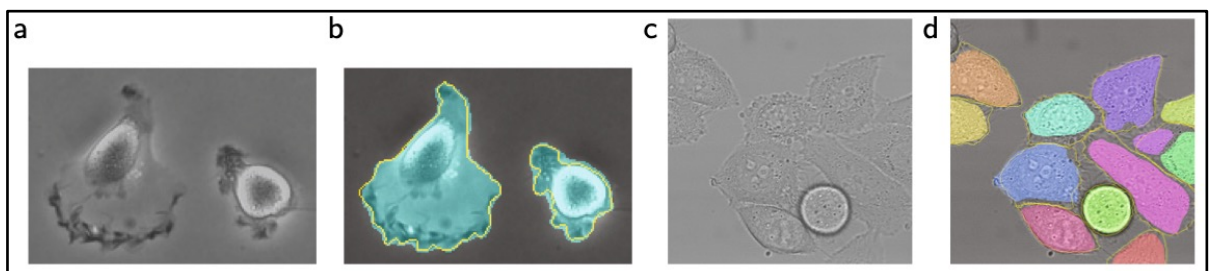


Figure 1.10 Résultat sur le défi de suivi cellulaire ISBI, (a) Une partie d'une image d'entrée du jeu de données "PhC-U373", (b) Résultat de la segmentation (masque cyan) avec la vérité terrain manuelle (bordure jaune), (c) Image d'entrée du jeu de données "DIC-HeLa", (d) Résultat de la segmentation (masques colorés aléatoirement) avec la vérité terrain manuelle (bordure jaune)

Tiré de Ronneberger et al. (2015)

Les méthodes de détection d'objets ont considérablement évolué avec l'émergence des modèles de réseaux de neurones convolutionnels. Particulièrement les architectures R-CNN développées par Girshick et al. (2014) qui reposent sur une architecture en deux étapes : les régions d'intérêt (RoIs) sont détectées et le modèle effectue une classification sur chaque région suivie d'une régression pour définir la position de la boîte englobante de chaque objet dans l'image. Cependant, ce modèle était chronophage vu le nombre de CNN à exécuter sur chaque RoI. Pour surmonter cette contrainte, Fast R-CNN et Faster R-CNN ont respectivement été développés par Girshick (2015) et Ren et al. (2015). La première approche améliore la méthode R-CNN en permettant de traiter une image entière une seule fois pour extraire des caractéristiques, ce qui est plus efficace que de calculer les caractéristiques pour chaque région de l'image. Le Faster R-CNN quant à lui introduit les réseaux de propositions de régions (RPN) qui prédisent les RoIs à partir des cartes de caractéristiques extraites par le réseau.

L'évolution de la segmentation d'instances a été marquée par l'introduction du Mask R-CNN (He et al., 2018) qui ajoute une branche de segmentation par masque binaire à l'architecture de Faster R-CNN. Vasam & Nayak (2019) ont proposé des modifications supplémentaires pour améliorer encore la précision et la vitesse des modèles de détection en temps réel en utilisant des réseaux de neurones plus légers pour optimiser les processus de calcul. Ces avancées ont permis d'élargir leur champ d'application sur des tâches qui nécessitent une réactivité instantanée (p. ex. la surveillance vidéo en direct et la conduite autonome).

La segmentation d'images est particulièrement nécessaire pour automatiser le processus d'annotation et d'étiquetage de photographies de sols ou de minéraux. Elle permet d'extraire les informations texturales pertinentes à des fins de caractérisation comme la granulométrie ou l'identification des formes de particules (Tariel, 2008; Zafari et al., 2019). Le Mask R-CNN a déjà été utilisé pour différencier les sols photographiés dans des environnements naturels de terrain (Y. Chen et al., 2023). Les auteurs atteignent une précision moyenne (mAP) de 0,8804 sur l'ensemble de validation et un IoU de 50%. Ces résultats ont permis de montrer la

transférabilité des méthodes de segmentation d'instance et d'apprentissage en profondeur vers le domaine de la caractérisation des sols.

L'évaluation des performances des modèles d'apprentissage supervisé est primordiale avant leur utilisation. Il existe plusieurs métriques qui permettent de quantifier l'efficacité d'un modèle. Le score F1, le rappel et la précision sont les métriques couramment utilisées pour les tâches de classification (Anouar, 2023). La précision mesure la proportion de prédictions positives correctes parmi toutes les prédictions positives effectuées par le modèle, ce qui est essentiel lorsque le coût des faux positifs est élevé (Powers, 2011). Le rappel permet d'évaluer la proportion de prédictions positives correctes par rapport au nombre total d'instances positives. Le score F1 quant à lui est un compromis entre les 2 métriques en calculant la moyenne harmonique entre les deux valeurs :

$$Précision = \frac{VP}{VP + FP} \quad (1.3)$$

$$Rappel = \frac{VP}{VP + FN} \quad (1.4)$$

$$F1\ score = 2 \frac{Précision \cdot Rappel}{Précision + Rappel} \quad (1.5)$$

où VP représente les vrais positifs, FN représente les faux négatifs et FP les faux positifs.

Les modèles de régression adoptent des métriques plus quantitatives pour l'interprétation des performances. L'erreur quadratique moyenne (MSE) mesure la moyenne des carrés des écarts entre les données réelles (y_i) et les données prédites ($\bar{y}_i$) :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (1.6)$$

Ce paramètre est plus sensible aux erreurs aberrantes contrairement à l'erreur absolue moyenne :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i| \quad (1.7)$$

Toutefois, le choix de la métrique appropriée pour un problème de régression est étroitement lié aux données d'entraînement (Lewinson, 2023).

1.4.2 Contraintes des modèles d'apprentissage supervisé

L'indisponibilité des données est l'un des principaux goulots d'étranglement du processus d'entraînement des modèles d'apprentissage supervisé (Nowruzi et al., 2019). Les structures de CNN ne font pas exception à cette contrainte. De grands volumes de données sont souvent exigés durant le processus de formation du modèle. Cette exigence est relativement liée à la complexité structurelle des modèles souvent dotés de millions de paramètres entraînaibles. Disposer d'une quantité importante de données devient essentielle pour garantir un entraînement efficace en minimisant les problèmes de sur-apprentissage ou de sous-apprentissage (Krizhevsky et al., 2012). Une étude menée par Sun et al. (2017) avait confirmé l'existence d'une progression logarithmique entre les performances des modèles de CNN et l'augmentation des données d'entraînement.

Les méthodes d'apprentissage supervisé sont souvent confrontées à des contraintes computationnelles, notamment la dépendance à des ressources importantes de calcul (p. ex. GPU, TPU, CPU) (I. Goodfellow et al., 2016) et la gestion des hyperparamètres (ensemble de paramètres prédéfinis qui contrôlent le processus d'apprentissage des modèles sans pour autant être appris à partir des données d'entraînement). Parmi ces hyperparamètres, le taux d'apprentissage qui détermine la vitesse d'entraînement d'un modèle, le nombre d'itérations du processus d'entraînement et la profondeur du réseau sont généralement déterminés de manière empirique et requiert fréquemment des essais répétés pour identifier la configuration optimale à la tâche ciblée.

1.4.3 Solutions générales pour surmonter les contraintes des modèles d'apprentissage supervisé

Tremblay et al. (2018) montrent que l'intégration des données synthétiques durant l'entraînement d'un modèle peut suffisamment combler le fossé entre l'indisponibilité des données d'entraînement et les besoins de données pour l'apprentissage profond. Présentant une augmentation de performance de l'ordre de 15% entre le modèle entraîné uniquement avec les

données réelles et celui incorporant les données synthétiques durant son entraînement. L'utilisation des données synthétiques est généralement moins coûteuse que la collecte des grandes quantités de données réelles. De plus, les images de synthèse offrent la possibilité de contrôler les conditions d'éclairage, de disposition d'objets, d'angle ou même de bruit qui peuvent aider les modèles face aux variations rencontrées sur les données réelles (Ros et al., 2016; Tobin et al., 2017)

Pour répondre aux limitations de quantité et de qualité de données, les GANs (réseaux adverses génératifs) ont été proposés par Goodfellow et al. (2014) pour générer des données synthétiques réalistes en partant d'un vecteur de bruit latent. Ces données générées ont contribué à l'amélioration des performances de nombreux modèles d'apprentissage supervisé (Abdusalomov et al., 2023; Z. Chen et al., 2024; Lamba et al., 2022; Waheed et al., 2020). Les GANs sont entraînés sur un processus de jeu à somme nulle à l'aide de deux structures. Le générateur tente de créer des données synthétiques suffisamment ressemblantes aux données réelles, à l'inverse du discriminateur qui évalue si une donnée est réelle ou synthétique. En outre, le générateur essaie de tromper la vigilance du discriminateur en créant des données assimilables aux échantillons réels.

Dans une optique d'amélioration des performances des modèles de segmentation à l'aide de données synthétiques plus réalistes, une approche intéressante a été proposée par Barth et al. (2020) visant à exploiter les CycleGANs (Cycle-Consistent Generative Adversarial Network) pour la traduction d'images agricoles (Barth et al., 2020) du domaine synthétique au domaine réel (Figure 1.11). En exploitant les images synthétiques dont la segmentation avait été faite manuellement sur PhotoShop ou automatiquement par un modèle de segmentation sémantique, il était possible d'optimiser le réalisme de ces images tout en conservant le masque de segmentation initial. Le but était de disposer de nouvelles données sans avoir à répliquer le processus d'annotation manuel fastidieux et chronophage de l'ordre de 30 min par image. En utilisant ces images transformées vers le domaine réel pour l'entraînement d'un modèle de segmentation, Barth et al. (2020) ont abouti à un IoU moyen de 0.31, soit une augmentation de

55 % par rapport aux méthodes précédentes utilisant uniquement les simples images synthétiques.

Les modèles de diffusion offrent une alternative intéressante et prometteuse aux GANs qui peuvent parfois restreindre la diversité des échantillons créés en raison de leur dépendance à un seul générateur. Récemment introduits dans les domaines de l'intelligence artificielle, les modèles de diffusion ont démontré un potentiel remarquable pour surmonter le problème de limitation de données couramment rencontré par les modèles d'apprentissage supervisé (Croitoru et al., 2023). Ces modèles reposent sur un processus de bruitage progressif qui consiste à ajouter du bruit sur les données d'entrée jusqu'à obtenir un bruit pur et exploiter les algorithmes de reconstruction adéquats pour générer de nouvelles données à partir de ce bruit (Ho et al., 2020; Song et al., 2022). Dhariwal & Nichol (2021) ont montré que les modèles de diffusion surpassent les GANs pour la création d'images synthétiques de haute qualité en préservant les détails et les structures complexes des données. Le FID (Frechet Inception Distance) qui mesure la distance entre les distributions de caractéristiques extraites des images réelles et générées, a atteint 2.97 pour les modèles de diffusion contrairement à 6.02 pour les BigGAN-deep 8 sur les images de la base de données d'ImageNet 128x128 pixels. Le FID plus bas indique que les modèles de diffusion génèrent des images dont les caractéristiques sont plus similaires à celles des images réelles par rapport aux modèles GANs.

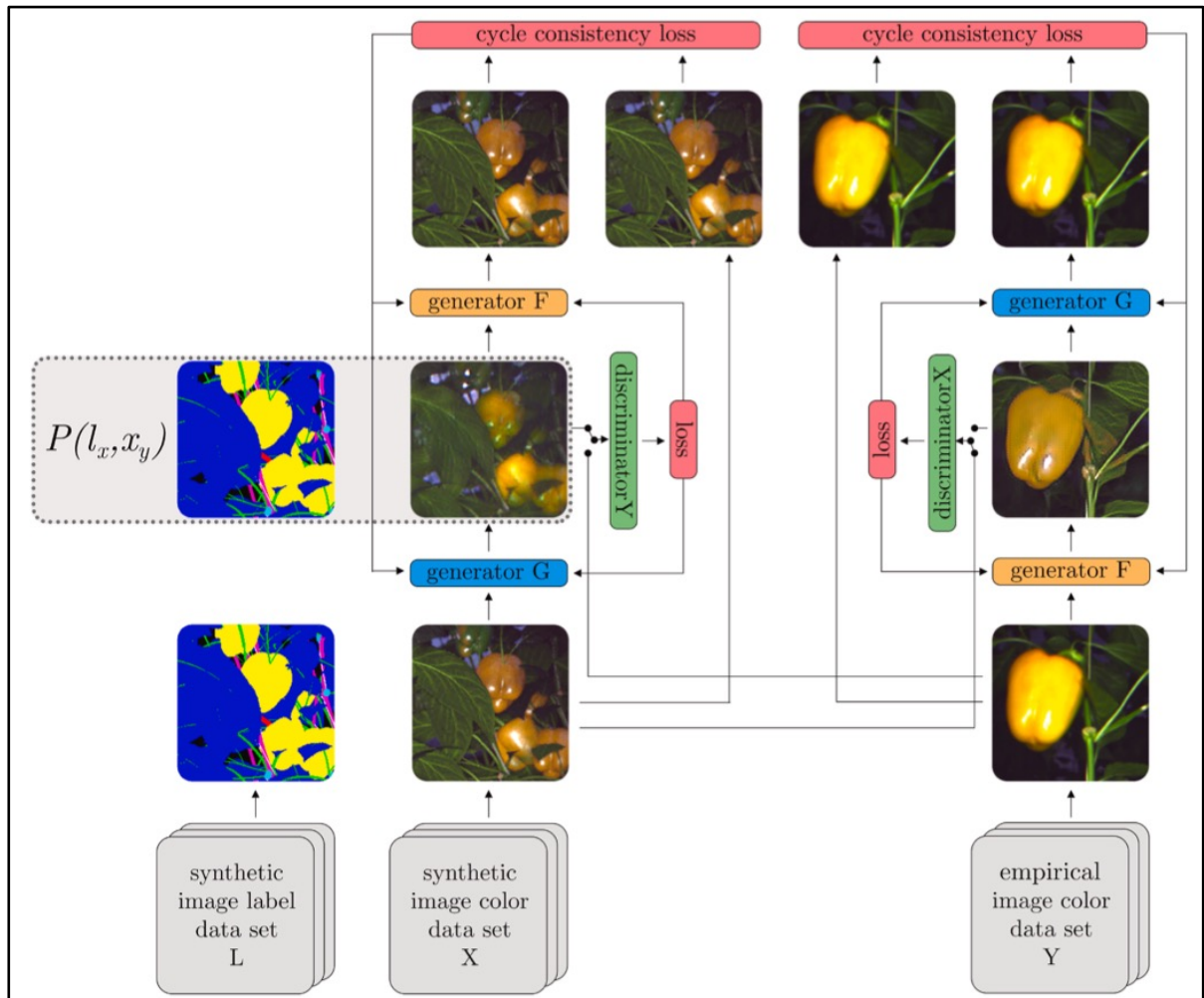


Figure 1.11 Quelques exemples d'images synthétiques avant et après leur transformation par le CycleGAN démontrant l'amélioration du réalisme visuel des images de synthèse après traitement vers le domaine réel et réciproquement
Tiré de Barth et al. (2020)

Il existe de nombreuses approches documentées dans la littérature pour surmonter la dépendance aux hyperparamètres inhérente aux modèles d'apprentissage supervisés. L'optimisation bayésienne (Moćkus, 1975) est une méthode automatique pour le réglage des hyperparamètres grâce à des algorithmes avancés de modèles probabilistes. À cette étape les processus gaussiens permettent de trouver un optimum de la fonction objective afin de guider efficacement l'exploration des hyperparamètres (Joy et al., 2016). Également, plusieurs approches modernes parmi lesquelles Hyperopt (Bergstra et al., 2013), SMAC (Hutter et al., 2011), Ray Tune (Liaw et al., 2018) et Optuna (Akiba et al., 2019), intègrent différentes

approches (optimisation bayésienne, estimateurs de type ‘Tree-structured Parzen Estimator’, les forêts aléatoires et les méthodes d’allocation adaptative des ressources et d’arrêt anticipé) visant à transformer le réglage d’hyperparamètres en un problème d’optimisation systématique et reproductible au lieu d’une procédure heuristique fortement dépendante des essais empiriques successifs.

1.4.4 Revue des modèles d’apprentissage les plus populaires pour la caractérisation des sols

Les réseaux de neurones convolutionnels sont de plus en plus utilisés pour l’analyse granulométrique. Ils offrent des moyens précis et automatiques pour déterminer la distribution des tailles de particules à partir d’images ou des indicateurs comme les D_{10} , D_{30} , D_{50} etc. Plusieurs modèles basés sur les CNN ont été développés ces dernières années pour prédire la taille des particules à partir d’images (Baraian et al., 2022; Iwashita et al., 2022; Manashti et al., 2022; Olivier et al., 2020; Soranzo et al., 2025; Xiao et al., 2022). Parmi ces modèles, deux CNN se distinguent particulièrement dans la littérature. Le modèle GRAINet a été développé pour prédire des caractéristiques granulométriques de matériaux grossiers (bancs de gravier) à partir d’images optiques issues de photographies aériennes haute résolution. En parallèle, le modèle SediNet, introduit par Buscombe (2020), est conçu pour l’analyse granulométrique de granulats plus fins tels que des sables et des sédiments littoraux à partir d’images optiques.

GRAINet a démontré sa capacité à surpasser certains experts dans l’estimation visuelle du diamètre moyen (D_m) de divers types de sol (Figure 1.12), avec une erreur quadratique moyenne de 24 % (Lang et al., 2021). Un jeu de données de 1419 photographies de 25 bancs de gravier de tailles de grains comprises entre 0.5 cm et 40 cm avait été utilisé dans l’exercice d’apprentissage de ce modèle (Figure 1.13).

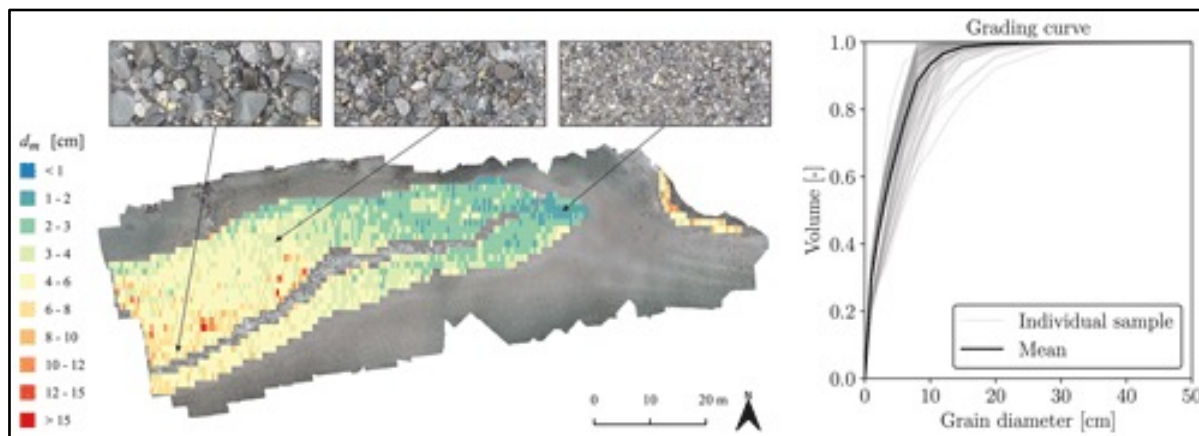


Figure 1.12 Illustration des deux produits finaux générés avec GRAINet sur le Rhône. À gauche : carte de la distribution spatiale des granulométries caractéristiques (D_m). À droite : courbe de granulométrie pour l'ensemble des bancs de graviers
Tiré de Lang et al. (2021)



Figure 1.13 Exemple de tuiles d'images (1,25 m×0,5 m) avec une distance d'échantillonnage au sol de 0,25 cm. Chacune des 25 tuiles d'exemple provient d'un banc de gravier différent
Tiré de Lang et al. (2021)

GRAINet adopte une structure inspirée du modèle ResNet en intégrant plusieurs blocs résiduels qui jouent un rôle de stabilisateur durant l'apprentissage du modèle (Figure 1.14). L'optimisation de ce modèle est facilitée d'une part avec l'utilisation des couches de

normalisation par lot et d'autre part, par l'exploitation de la fonction d'activation ReLU (Rectified Linear Unit). La ReLU définie par :

$$ReLU(x) = \max(0, x) \quad (1.8)$$

renvoie la valeur d'entrée si elle est positive et 0 dans le cas échéant. Cette activation non linéaire simplifie la propagation du gradient entre les couches et accélère ainsi la convergence du modèle. Elle permet aux CNN de mieux modéliser les relations complexes. La normalisation par lot quant à elle est une technique de standardisation des activations intermédiaires qui limite les dépendances liées à l'initialisation des poids. Elle joue un rôle clé durant l'apprentissage du modèle tout en contrôlant les variations excessives du gradient d'une couche à l'autre (Ioffe & Szegedy, 2015). Cette approche améliore la robustesse et la généralisation du modèle tout en permettant l'utilisation de taux d'apprentissage (vitesse d'apprentissage) plus élevés.

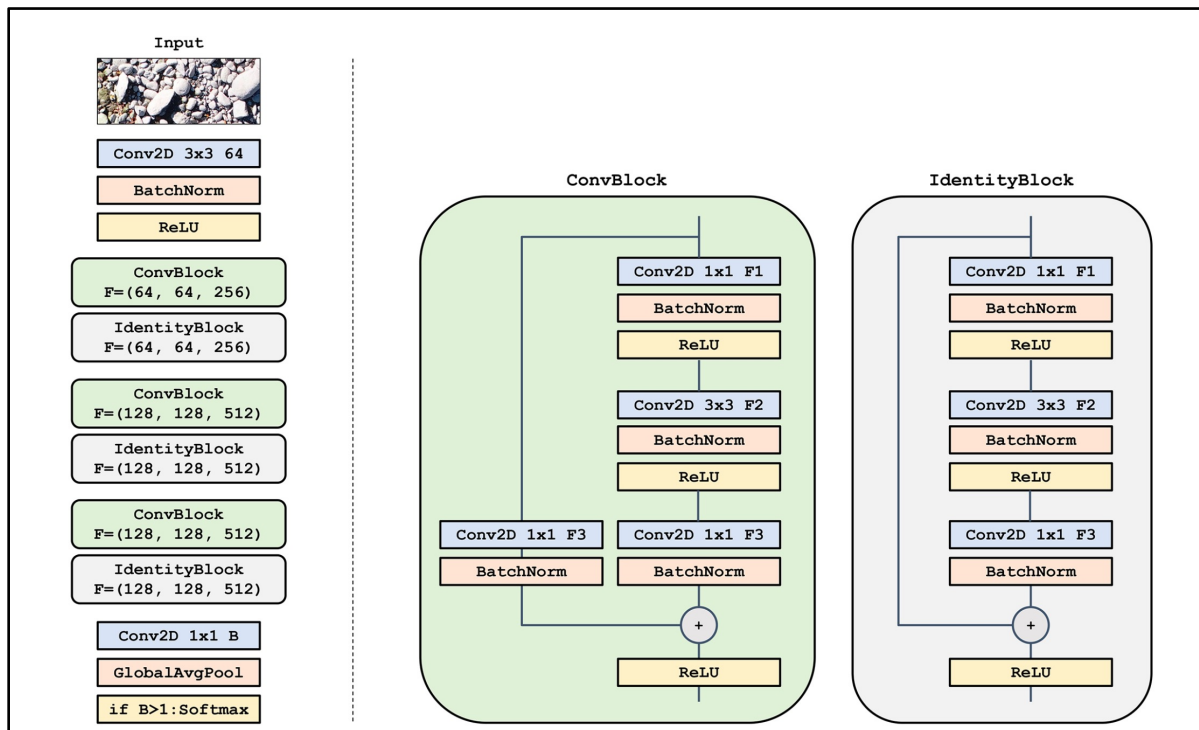


Figure 1.14 Architecture du modèle GRAINet
Tiré de Lang et al. (2021)

SediNet est un autre cadre d'apprentissage automatique configurable préparé pour l'analyse de la distribution granulométrique des sols et leur classification selon la forme et la taille des

particules (Buscombe, 2020). Un total de 409 images a été utilisé pour entraîner le modèle parmi lesquelles certaines sont présentées à la Figure 1.15. La première estimation du modèle pour prédire 9 percentiles de la distribution cumulative (à savoir les 5^e, 10^e, 16^e, 25^e, 50^e, 75^e, 84^e, 90^e et 95^e percentiles) a donné des erreurs (racine carrée de l'erreur quadratique moyenne normalisée) oscillantes entre 24% à 45% sur l'ensemble de données de test. La seconde estimation du modèle SediNet reposait sur la même tâche que la première implémentation (percentiles granulométriques) mais sur des matériaux beaucoup plus homogènes incluant uniquement 31 images de sable de plage, réparties en 16 pour le test et 15 pour l'entraînement. L'erreur moyenne des données de test était comprise entre 16% et 29%. La troisième implémentation de SediNet a été employée pour estimer la taille réelle des grains (implicitement l'échelle de l'image). Dans cette étape, une sous-estimation jusqu'à 22% de la valeur réelle de la taille des grains a été mentionnée.

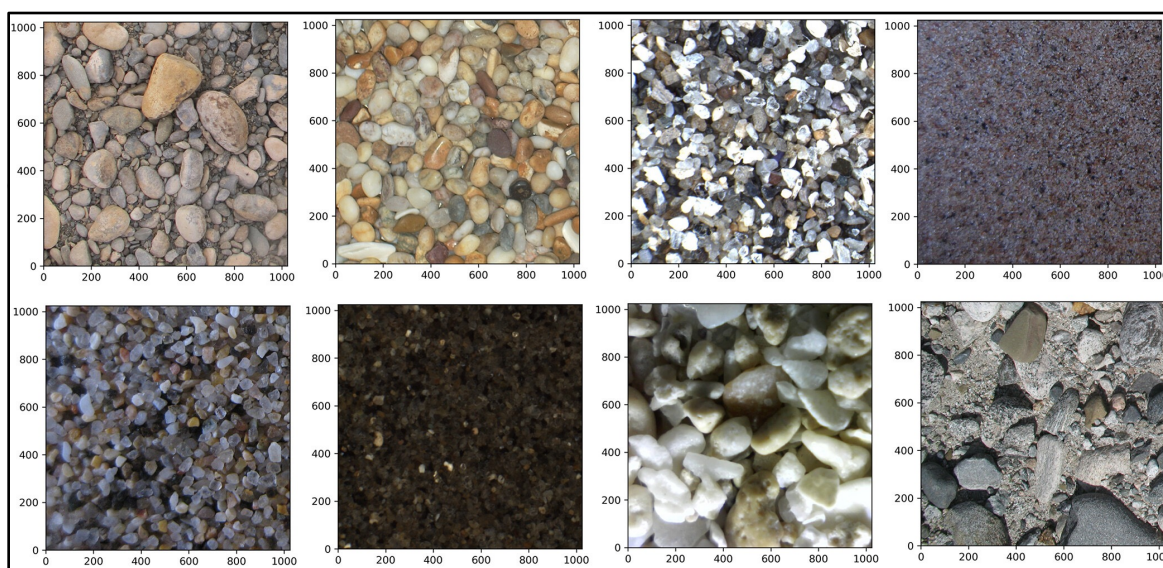


Figure 1.15 Exemple d'images utilisées pour l'entraînement de SediNet
Tiré de Buscombe (2020)

SediNet repose donc sur une architecture comparable à celle du modèle VGG (Qawaqneh et al., 2017), caractérisée par une structure séquentielle où les couches sont empilées de manière linéaire et ordonnée. À l'instar de VGG, ses convolutions sont systématiquement faites avec des filtres homogènes (typiquement dont la taille du noyau est de 3x3). Néanmoins, il se

distingue principalement du modèle de base par l'utilisation des couches de convolution séparables en profondeur introduites par Chollet (2017). SediNet adopte ainsi l'approche des modèles séquentiels en incorporant des optimisations inspirées des architectures comme Xception et MobileNet (Sarkar, 2021). L'architecture globale simplifiée du modèle SediNet est illustrée à la Figure 1.16.

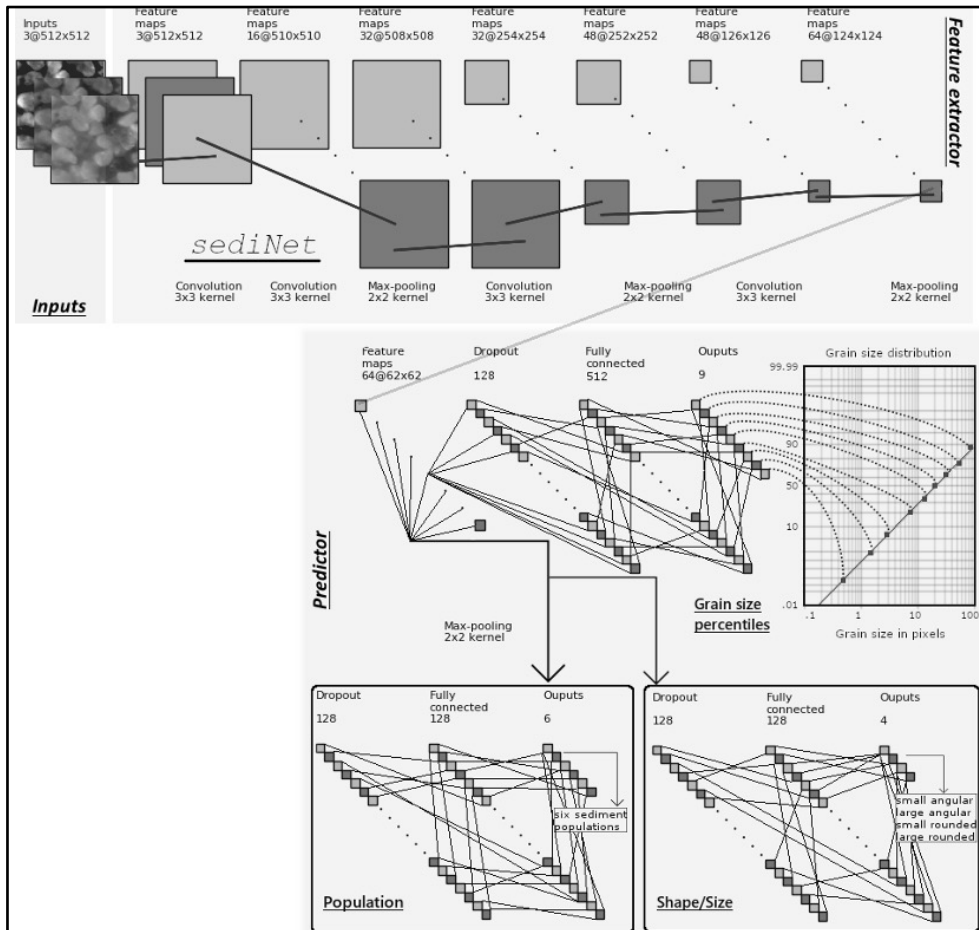


Figure 1.16 Architecture simplifiée du modèle SediNet
Tiré de Buscombe (2020)

Malgré les performances encourageantes rapportées dans ces contextes spécifiques, l'exploitation des méthodes d'apprentissage supervisé basée sur les CNN est confrontée à de nombreuses limitations. L'indisponibilité des données associée au coût élevé de leur acquisition est couramment mentionné dans de nombreux travaux (Iwashita et al., 2022; Manashti et al.,

2022; Olivier et al., 2020). Les résultats affichés par Lang et al. (2021) laissent observer que les modèles montrent une efficacité notable pour des matériaux grossiers caractérisés par des particules clairement discernables à l'échelle de l'image, mais restent difficilement transposables à des sols plus fins. De manière similaire, les différentes implémentations de SediNet mettent en évidence une dégradation significative des performances lorsque la variabilité sédimentologique augmente ou lorsque le nombre d'images disponibles pour l'apprentissage est limité.

1.4.5 Stratégies palliatives aux contraintes des modèles d'apprentissage supervisé pour l'analyse granulométrique des sols

La question d'indisponibilité des données a conduit à deux familles d'approches palliatives. D'un côté, l'apprentissage par transfert et l'autre côté, la génération de nouvelles images par transformations photométriques ou création de données synthétiques. Au-delà des limitations liées aux données, la complexité intrinsèque de la prédiction granulométrique constitue une contrainte supplémentaire surtout pour la prédiction des pourcentages de particules fines. Ceci justifie le choix de certains auteurs de restreindre ce problème à la prédiction de paramètres granulométriques spécifiques comme les D_m , D_{10} , D_{30} , etc. ou uniquement à la caractérisation des matériaux granulaires (Choi et al., 2025; Jung et al., 2022; Lang et al., 2021; Soranzo et al., 2025).

L'apprentissage par transfert est une méthode où un modèle préalablement entraîné sur une tâche est adapté et réutilisé pour résoudre une nouvelle tâche. Ce concept mis en avant par Yosinski et al. (2014) permet de tirer parti des connaissances acquises sur des données, en particulier lorsqu'une quantité limitée de données est disponible pour la tâche cible. Plusieurs auteurs ont eu recours à l'apprentissage par transfert pour mesurer la distribution des particules dans des cas de données insuffisantes (Olivier & Aldrich, 2021; Wei et al., 2024; Xiao et al., 2022). Par exemple, dans une étude récente menée par Soranzo, Guardiani et Wu (2025), un réseau de neurones convolutionnels pré-entraîné nommé MobileNet, a été adapté pour prédire la distribution granulométrique des sols (notamment le paramètre d'échelle b et le paramètre de forme c de la distribution de Weibull) photographié dans un cadre contrôlé. Le modèle a

atteint un $R^2 \approx 0,995$ pour le paramètre d'échelle b et $R^2 \approx 0,84$ pour le paramètre de forme c , avec des erreurs moyennes respectives d'environ 5 % et 10,7 %, montrant que l'apprentissage par transfert permet de réduire la charge de données nécessaires tout en garantissant une prédiction robuste des caractéristiques granulométriques des sols à partir d'images.

L'amélioration de la qualité des images mais aussi l'augmentation de la quantité d'images à partir des méthodes classiques de transformation d'images (p. ex. la rotation, le recadrage et les changements de luminosité) permettent d'enrichir les données d'entraînement sans nécessiter la collecte de nouveaux échantillons (Shorten & Khoshgoftaar, 2019). Wei et al. (2024) ont récemment démontré l'efficacité des techniques de retournement horizontal et vertical associées aux modifications de couleurs. Un RMSE normalisé de 11.38% avait été obtenu en estimant la distribution granulométrique des sols composés de particules supérieures à 2 mm. Par ailleurs, lorsque le jeu de données initial contient des images de qualité insuffisante, des méthodes de super-résolution (GAN) comme celle proposée par Gong et al. (2024) permettent de synthétiser les images de haute résolution à partir des images de basse résolution par le biais des GAN avant leur traitement dans les CNN.

Le développement d'images synthétiques de matériaux granulaires a été exploré comme un outil pour compléter la couverture (le fuseau granulométrique) de l'espace d'apprentissage des modèles de CNN. Pirnia et al. (2018) et Temimi (2020) ont proposé des images synthétiques de sol (Figure 1.17) développés respectivement à l'aide du logiciel YADE (code d'éléments discrets) et du moteur de jeu Unity. Leurs jeux de données ont permis de prédire les pourcentages de passant pour sept tailles de tamis, allant de 0,075 mm à 1,180 mm, avec respectivement une erreur quadratique moyenne de 6,94 % et 8,5 %.

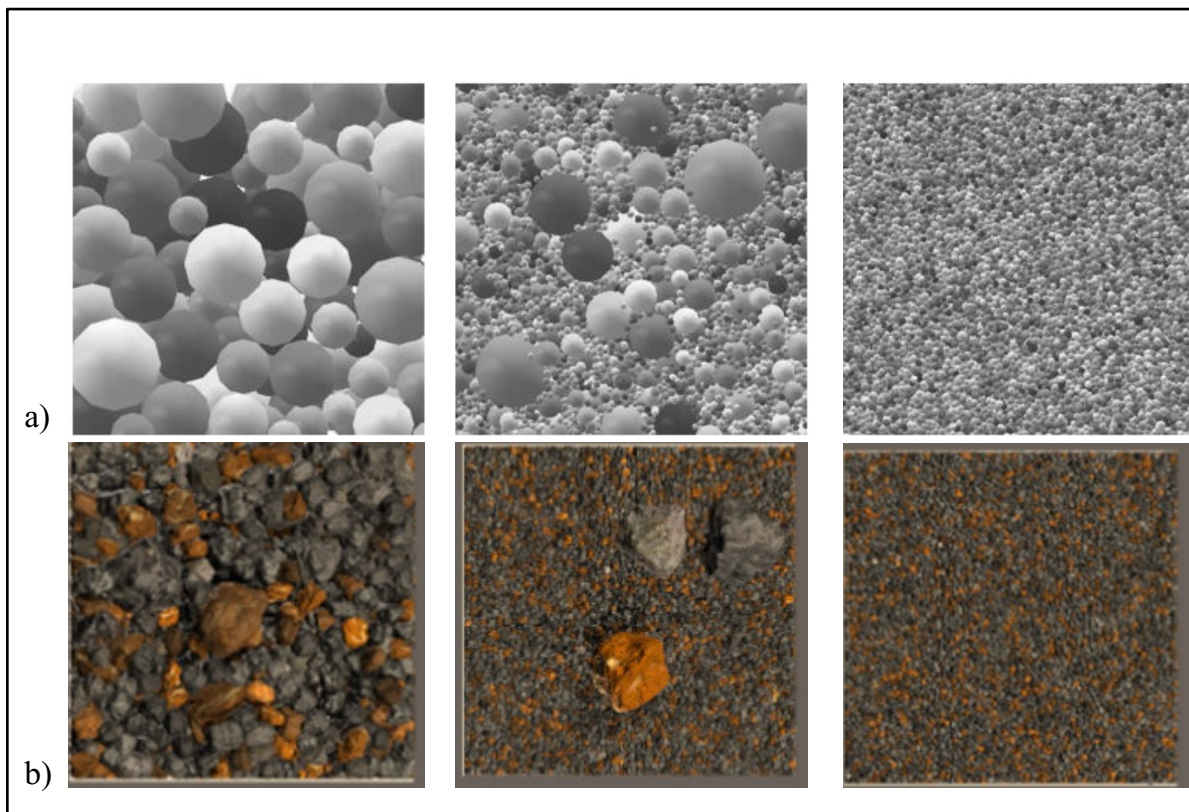


Figure 1.17 a) et b) représentent respectivement des exemples d'images de synthèse de matériaux granulaires développés avec YADE et Unity game engine
Tiré de Pirnia et al. (2018)

La structure hiérarchique des systèmes de classification des sols peut être exploitée comme stratégie pour améliorer la précision de l'analyse granulométrique. Plutôt que de prédire directement toute la courbe granulométrique en une seule étape, une approche hiérarchique peut être envisagée pour premièrement différencier les sols selon différentes catégories comme celles du système USCS. Par exemple, un premier réseau distingue les sols fins des sols grossiers, puis un second raffine la sous-catégorie (sable vs gravier, argile vs limon, etc.) et enfin un troisième prédit les proportions massiques des classes de matériaux spécifiques. Ces méthodes de décomposition en sous tâches plus faciles à résoudre ont montré de meilleurs résultats que les approches en une étape, en réduisant la confusion entre classes proches et en affinant les frontières entre catégories (Silla & Freitas, 2011 ; Xiao et al., 2014 ; Yan et al., 2015).

1.5 Synthèse critique et justification de l'approche proposée

Les approches automatisées intégrant l'exploitation des modèles d'apprentissage profond pour la prédiction des courbes granulométriques peinent à offrir une reproductibilité et une adaptabilité suffisante pour remplacer les méthodes d'analyse conventionnelle. Les modèles récemment développés tels que SoilGrids250m, GRAINet, SediNet et PSDnet présentent l'avantage de caractériser les sols (p. ex. analyse granulométrique, diamètre moyen, percentiles caractéristiques de la distribution granulométrique) à partir d'images de haute et basse résolution (Buscombe, 2020; Hengl et al., 2017; Lang et al., 2021; J. Manashti et al., 2023). Toutefois, bien qu'ils soient conçus pour donner une caractérisation globale du spectre granulométrique, l'estimation des proportions de particules fines reste un défi majeur attribué d'une part à l'insuffisance de données réelles de sols fins et d'autre part à l'incapacité de discernement des particules fines dont la taille projetée est souvent inférieure à un pixel dans l'image.

La plus-value des images de synthèse est une problématique qui sépare la littérature en deux courants distincts. D'une part, certains auteurs soutiennent que leur intégration améliore systématiquement les performances des modèles tandis que d'autres expriment des réserves quant à leur utilisation surtout de manière non contrôlée. Cependant, ces méthodes présentent souvent des limitations inhérentes au concept de photoréalisme, tel que discuté par Nikolenko (2021). Les images synthétiques doivent être capables de reproduire fidèlement les caractéristiques visuelles des photographies réelles pour une meilleure application (Goodfellow et al., 2014). Une utilisation non contrôlée ou inappropriée des images de synthèse est susceptible de dégrader les performances des modèles. Ce comportement est généralement observé lorsque les images générées ne parviennent pas à capturer la complexité des détails des environnements réels ou lorsqu'elles ne disposent pas d'artéfacts caractéristiques des données réelles (Kothai et al., 2024; Liu et al., 2023).

Bien que les modèles de diffusion et les réseaux antagonistes génératifs soient reconnus pour leur capacité à produire des images réalistes, leur application à l'analyse granulométrique pose des limites spécifiques. En effet, il demeure particulièrement complexe pour ces modèles de générer à partir d'une courbe granulométrique donnée, une image synthétique dont les particules respecteraient fidèlement la distribution de tailles prescrite et la ségrégation naturelle des particules. Par ailleurs, les approches de simulation développées pour pallier ces limitations en utilisant des outils comme YADE ou Unity game engine présentent un réalisme visuel encore insuffisant bien qu'elles permettent de générer de nouvelles images couvrant des plages granulométriques absentes des ensembles de données réelles disponibles pour l'entraînement. Il devient essentiel d'étudier avec précision les modalités de contrôle de la qualité et de la quantité de ces images de synthèse, afin d'en garantir l'efficacité dans ce contexte spécifique.

Les modèles d'apprentissage automatique généralement développés pour l'analyse granulométrique sont entraînés dans la grande majorité des cas à partir d'images acquises dans des conditions contrôlées et idéalisées en laboratoire ou sur des sites spécifiques. Ces environnements garantissent une homogénéité en termes d'éclairage, de fond, de distance focale et d'angle de prise de vue, ce qui facilite l'apprentissage des modèles. Toutefois, cette approche limite considérablement leur capacité de généralisation à d'autres conditions d'échantillonnage par photographie. Sur les chantiers de construction, les variations d'éclairage dépendent de facteurs environnementaux incontrôlables tels que l'ensoleillement, l'humidité, la présence d'ombres et les angles de prise de vue peuvent être contraints par la configuration des lieux ou des équipements. Certaines contraintes sont difficilement évitables dans la pratique. Il apparaît dès lors nécessaire de développer des méthodes capables d'exploiter efficacement les images de sols collectées en conditions réelles, en tenant compte de ces variations inévitables, afin de garantir des analyses granulométriques fiables et robustes, indépendamment des aléas du contexte d'acquisition.

Malgré les récents progrès mentionnés, la littérature actuelle ne règle pas plusieurs verrous de la caractérisation granulométrique par vision par ordinateur notamment les lacunes de précision suffisante (au sens des tolérances acceptables pour la prise de décision géotechnique)

sur l'ensemble du spectre granulométrique, les difficultés liées à la caractérisation des particules fines et l'applicabilité limitée de ces modèles en conditions réelles de chantier d'investigation géotechnique. Cette thèse répond à ces lacunes en exploitant un triptyque méthodologique combinant premièrement la création d'une vaste base de données diversifiée constituée de matériaux fins et grossiers enrichie par des données synthétiques réalistes pour améliorer l'apprentissage des relations entre texture et distribution granulométrique. Par la suite, la création de modèles d'apprentissage profond entraînés sur un corpus hétérogène (photographies de sols réels, de matériaux artificiels, images de sols synthétiques et images segmentées de sols) afin de limiter les artefacts spécifiques au contexte d'acquisition de données et accroître leur potentiel de généralisation. Enfin, l'adaptation et la validation de ces modèles en conditions réelles de chantier afin de démontrer leur capacité à caractériser automatiquement les sols dans un contexte opérationnel. Cette approche vise à améliorer la capacité de généralisation des prédicteurs granulométriques et à réduire la dépendance aux méthodes de laboratoire chronophages, tout en fournissant une bonne précision pour des applications géotechniques opérationnelles.

CHAPITRE 2

SEGMENTATION OF GRANULAR MATERIAL IMAGES USING SYNTHETIC DATA AND DEEP LEARNING

Harold Simo Tenekam ^a, François Duhaime ^b, Jean Sébastien Dubé ^c, Matthew F. Toews ^d

^{a, b, c} Département de Génie de la construction, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

^d Département de Génie des systèmes, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Article soumis pour publication, novembre 2025

2.1 ABSTRACT

A primary challenge in applying modern deep learning methods in specialized fields such as geology and geotechnical engineering is the limited availability of labeled training data. This problem is especially relevant in the analysis of granular material images, where deep learning methods can be used to segment individual particles, predict the particle size distribution (PSD), and classify materials. Creating a dataset with real photographs of granular materials, their PSD, and a manual segmentation of each image is extremely laborious and potentially imprecise. To address this limitation, a script was developed to generate synthetic training data from realistic physical simulations performed in the Unity game engine. The script models a set of three-dimensional soil particles with known PSD and mass. The script produces images of the granular material and a segmentation mask from top and bottom viewpoints. As an application example, GrainSegNet, a deep neural network based on mask R-CNN, was trained for particle segmentation based on a synthetic dataset. The script presented in this paper overcomes the limitations inherent in small training datasets by providing synthetic data specifically designed to reflect the textural characteristics of real photographs. The synthetic images improve the robustness of models applied to soil analysis and also establishes a scalable methodological framework applicable to other fields requiring large annotated databases.

2.2 Introduction

Particle size analysis is a ubiquitous task in civil engineering. It consists in determining the statistical distribution of particle sizes in a material, which is often represented as a curve of cumulative mass percentages of particles smaller than a given size. In geotechnical engineering, the particle size distribution (PSD) can be determined according to various methods, such as sieving or the hydrometer test (ASTM International, 2017; ASTM International, 2021). Particle segmentation, deep neural networks and textural features have also been used for predicting particle size distributions based on photographs (Duhaime et al., 2021; Hryciw et al., 2013; Lang et al., 2021; Parent et al., 2021; Xiao et al., 2022; Yaghoobi et al., 2019).

Image segmentation is widely used for solving computer vision tasks, such as object detection or classification. It involves grouping image pixels into regions sharing similar characteristics (X. Liu et al., 2021). Segmentation may seek to identify multiple instances of an object category (Hafiz & Bhat, 2020) or a semantic category (e.g. crowd, plastic objects). In mining and rock mechanics, segmented images are used to determine the PSD of coarse-grained material resulting from blasting (Bamford et al., 2021; Campbell & Thurley, 2017).

Wipfrag is a particle size analysis software based on image segmentation (Bouchard et al., 2013; Maerz et al., 1996). It is one of the most widely used software to characterize blasting efficiency in mining operations. It applies an automatic watershed segmentation filter to the blasted material photographs to define the particle contours, which often need to be adjusted manually by the operator to improve delineation (Bouchard et al. 2013). This is often one of the main bottlenecks when using software programs based on segmentation. After segmentation, volume and polygon approximations are used to predict the PSD from the size of the segmented regions. Segmentation methods based on 3D laser scanning have also been developed (Campbell & Thurley 2017).

Chen et al. (2021) and Yamada & Di Santo (2022) explored the automation of instance segmentation of granular material images using convolutional neural networks (CNN). CNNs are a type of neural network that can learn features directly from images using filters that perform convolution operations on the image. Region-based CNN masks (mask R-CNN) proposed by He et al. (2018) perform instance segmentation tasks by assigning each pixel in an image to an object (Bharati & Pramanik, 2020). Yamada & Di Santo (2022) presented a neural model based on ResNet-101, a CNN model consisting of 101 layers, to detect contours and segment the particles present in a granular material image. A significant drawback of automatic segmentation with CNN is the requirement for very large training datasets including annotated and labeled images of tens of thousands of particles. The performance of CNN generally increases with the size of the training dataset. Yamada & Di Santo (2022) used 261 images with 20 295 grains for their training dataset. The images were segmented semi-automatically using watershed transforms and clustering followed by manual fine-tuning. As a comparison, mask R-CNN was trained on the COCO dataset which contains 333 000 images including 80 categories of objects (He et al., 2018).

Several authors have shown that synthetic training datasets can be used to train deep neural networks (Man & Chahl, 2022; Picon et al., 2022; Sagues-Tanco et al., 2020; Wong et al., 2019). Preparing synthetic data is often a useful solution to create large datasets of annotated and labeled data. Access to real data can be limited for several reasons including confidentiality, availability, cost, or the time required to acquire the dataset. For particle size analyses, Pirnia et al. (2018) developed synthetic images of granular materials using Yade, a discrete element code (Smilauer et al., 2021). The images in this dataset were created to correspond to a range of PSD. The main objective of Pirnia et al. (2018) was to generate data to train CNN and neural networks fed with textural features, such as the mean and standard deviation of the conditional entropy of intensity (Duhaime et al., 2021; Shannon, 1948), to predict the PSD of granular materials without segmentation. Video games like Grand Theft Auto V (Richter et al., 2016) and game editing engines like Unity (Gaidon et al., 2016; Temimi, 2020) have also been used to generate synthetic images for different applications. Synthetic datasets can also be created by combining real images. For example, Zhang et al. (2022) created

images of granular materials by segmenting individual grains and combining the segmented grains in a new image.

Synthetic images necessarily differ to some extent from real photographs. For example, the granular material images created by Pirnia et al. (2018) were composed of perfect spheres with random colors, whereas real granular material photographs show angular grains with varying surface texture, and colors that are related to grain mineralogy. CNN trained with images that differ from the test (or final application) images often perform poorly (Barth et al., 2020). Domain adaptation refers to the different approach that can be used to adapt a network or a dataset to different target images, in other words to a different image domain (Patel et al., 2015). For example, Generative Adversarial Networks (GANs) can be trained to improve the photorealism of synthetic images (Shrivastava et al. 2016). GANs trained for that purpose are composed of generator and discriminator networks that are trained to modify the images, and to distinguish real and synthetic images, respectively. The generator learns to improve the synthetic images to deceive the discriminator. Several metrics can also be used to compare real and synthetic images, such as textural features (Barth et al., 2020; Haralick et al., 1973) and features extracted from a pretrained CNN (Heusel et al., 2017).

This paper presents a novel image generation program specifically designed to create highly realistic synthetic images of granular materials conforming to predefined PSDs. One of the main advantages of this work lies in the integration of perfect particle segmentation with photorealistic rendering, enabling the generation of large sets of labeled data that are difficult to obtain in the real world. To guarantee the realism of the generated data, a new validation approach based on conditional entropy, a textural feature closely related to PSD, is proposed and applied to compare synthetic and real-world images. To demonstrate the method's potential, a convolutional neural network (GrainSegNet), built from a mask R-CNN, was trained using a dataset of 20,500 synthetic images. This result is all the more remarkable as the mask R-CNN, originally trained on 80 categories excluding soil particles, has been successfully adapted to the segmentation of granular materials using transfer learning. This contribution not only provides a powerful tool for training segmentation networks but also

opens up new avenues for automated analysis in geotechnical engineering and geosciences, such as PSD estimation and beyond.

2.3 Methodology

The proposed method to generate synthetic images of segmented granular materials consists of two steps. The first step was to create and take screenshots of a virtual sample of 3D particle models in a container. The second step was to annotate and label each particle in the screenshots. These steps result in synthetic images of granular materials with known PSDs and exact segmentation masks. The synthetic images were calibrated and validated by comparing their conditional entropy with real images of granular materials.

2.3.1 Scripts for synthetic image preparation

Unity is a video game development platform that allows two- and three-dimensional worlds to be created as well as interactive simulations. Scenes form the basis of simulations in Unity. They include several objects that are manipulated to produce a simulation. Scenes can be 2D or 3D depending on the application. The physics of the 3D objects in the scenes (e.g. gravity, velocity, response to collisions) can be controlled. The other main aspects to take into consideration are the graphics (camera and lighting management), the script (C Sharp program to control the dynamics in the scene), and finally the animations (particle movement). Realism depends on the objects that are used and how they are handled. A total of 188 3D scans of real particles were utilized (Figure 2.1) to simulate various particles shapes (e.g., rounded, semi-rounded, angular, semi-angular). The particles were standardized in Blender, a 3D computer graphics editor, in order to initially have the same scale. For a particle P_i of dimension x_i, y_i, z_i , the intermediate dimension was used as a scaling factor for the other dimensions to preserve the shape of the model and to make sure that the models had the same scale:

$$\text{For } y_i \in [x_i, z_i], \quad P_i(x_i, y_i, z_i) \rightarrow P_i\left(\frac{x_i}{y_i}, 1, \frac{z_i}{y_i}\right) \quad (2.1)$$

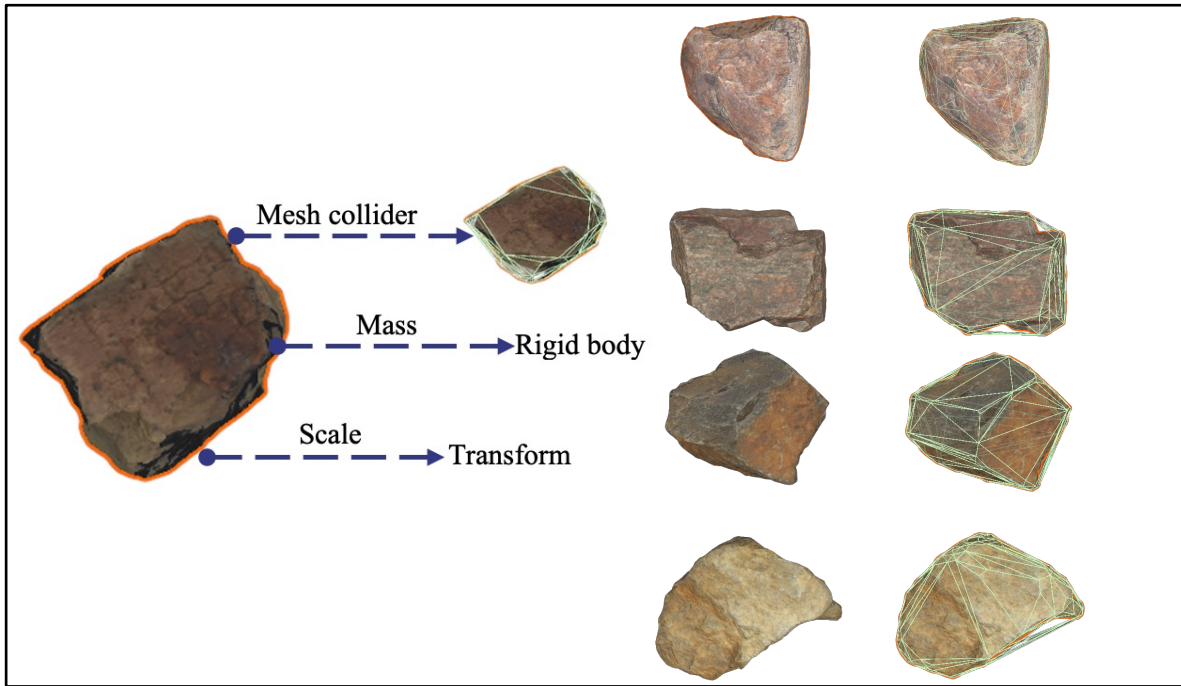


Figure 2.1 Examples and characteristics of the 3D particle models

A transparent container (180 mm wide x 180 mm long x 100 mm high) was also created in Blender for holding the synthetic samples. Subsequently, all objects were imported into the Unity game engine in the FBX format before being manipulated in the scene. The environment modeled to generate the samples comprises two cameras that allow images to be taken from above and below the transparent container (Figure 2.2). Since particles do not form a single layer, some of them can be blocked by other particles. Particle overlap combined with segregation results in different apparent PSD for images taken from above and below. Acquiring images from two separate perspectives can provide valuable information on the spatial distribution of grains, an essential element for applications requiring detailed characterization of the granular structure of a soil sample.

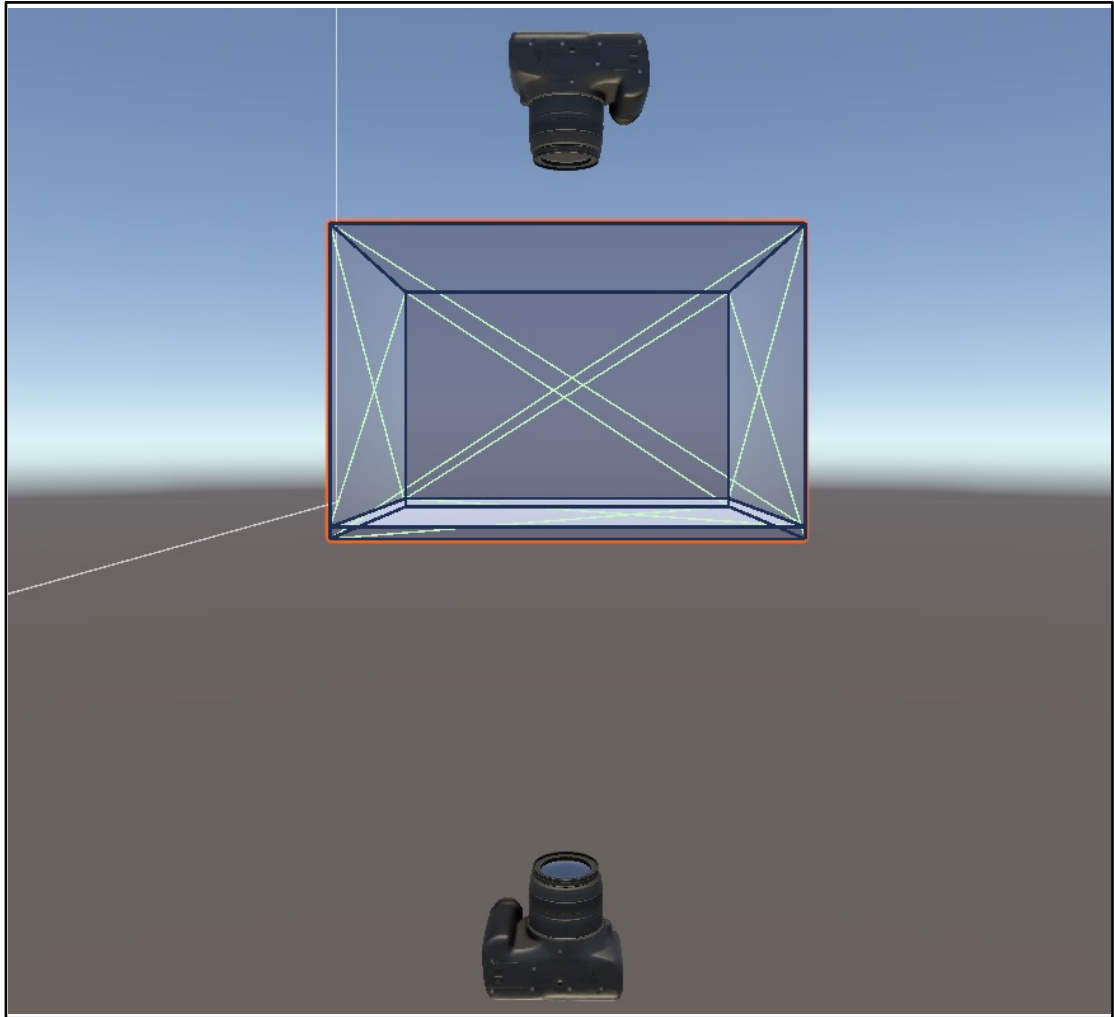


Figure 2.2 Container and cameras in the Unity 3D model

The PSDs are provided via a text file. The file contains the mass percentage of particles retained on sieves of 2.5, 5, 10, 14, 20 and 31.5 mm. These sizes correspond to the standard sieve series in standard BNQ 2501-025. For each PSD, particles are generated randomly based on the mass proportion retained between each sieve. For example, for a sample in which 30% of the mass passes through the 5 mm sieve and none passes through the 2.5 mm sieve, the program generates an equivalent number of particles with randomly selected sizes in the interval $[2.5 \text{ mm}, 5 \text{ mm}]$, so that the mass of all particles generated corresponds to 30% of the total mass of the sample m_T (in g). The number of particles generated in this interval, N_i is calculated based on the total mass and the average mass of the particles in the interval of

interest. Specifically, the mass proportion of the fraction, x_i is set to 0.3 in this example. The average mass of a particle in this interval is estimated using its volume v_i (in cm^3), computed from a diameter equal to the mean of the minimum ($d_{min} = 2.5$ mm) and maximum ($d_{max} = 5$ mm) diameters of the interval, and assuming spherical particles. The volume is then multiplied by the solid density ρ_i fixed to 2.7 g/cm^3 for all particles (Equations 2.2 et 2.3):

$$N_i = \frac{m_T x_i}{v_i \rho_i} \quad (2.2)$$

$$v_i = \frac{4}{3} \pi \left(\frac{d_{min} + d_{max}}{4} \right)^3 \quad (2.3)$$

Each particle is assigned a mass and a mesh collider composed of 255 polygons, which surrounds the 3D objects and is used for collision detection. The size of each particle is assigned at the time of its generation. The particles are generated at random positions above the transparent container. The friction between the particles in the model is controlled using two parameters: the dynamic and static friction coefficients. They represent the amount of surface friction experienced by a particle if it is moving or at rest, respectively. These coefficients range from 0 (frictionless) to 1 (very high friction). In this case, the dynamic, and static friction coefficients were set to 0.7 based on the general appearance of the virtual samples. Changing the friction coefficients has a direct impact on the distribution and mobility of particles in the container. A lower friction coefficient leads to a more pronounced segregation of particles with respect to their size.

The duration of each frame during program execution was set to 0.05 s. This duration was required for stability and to allow the processor to capture the thousands of collisions that occurred while the particles settled in the container.

The time required to generate an image depends on the number of particles. Generating a sample with a large number of particles requires more time. For example, the code required 10 minutes to generate a sample of 15 000 particles with a 2.8 GHz Core i7 x8 processor with 16 GB of RAM, and a 4 GB graphics card. The main bottleneck that limits the number of

particles that can be included in a scene is the time required to evaluate all collisions in the main container as well as the particle displacements and rotations. Generating scenes with particles of 2.5 mm or less is challenging. For instance, a sample composed exclusively of particles with a diameter of 2.5 mm and a solid density of 2.7 g/cm^3 with a total mass of 500 g requires 61 146 particles.

The resulting images are generated in PNG format, their size being adjustable according to the specific requirements of the program. In this study, two resolutions were selected: 4016×4016 pixels for high-resolution images (corresponding a scale of $45 \text{ }\mu\text{m/pixel}$) and 299×299 pixels for segmented images. The latter resolution was adopted to reduce the surface area of the segmented regions, thus minimizing computing time and resource requirements. Once the images are generated, annotations are created using the Perception library applied on the top-view camera to detect and segment each particle on the images.

2.3.2 Data annotation

Data annotation involves adding labels or metadata to a dataset in order to use it for supervised learning. Several data formats can be used for segmentation tasks. In this case, the COCO (common objects in context) format was used (Lin et al., 2015). Each annotation contains information on the image ID, the class of the annotated object, the size of the object region (in pixels), and the coordinates of the bounding box (bbox). In this case, the segmented regions belong to two classes: the particle class and the background class for pixels in the images that are not covered by particles. The annotations were used to train GrainSegNet to segment the particles that are present in the images.

Unity provides a toolbox called Perception which allows large-scale annotated training and validation datasets to be created (Unity Technologies, 2020). Perception can extract the characteristics of an images generated on Unity (e.g. bounding box, labels, segmentation). It takes the form of a script that is applied to the cameras which generate the images in the 3D scenes. Annotations are saved in a file. For the images that are generated in this paper, all

particles have a single label (Figure 2.3). This allows Perception to visualize all the objects to be segmented in the container. For each synthetic image, the annotation file contains information on the position of each particle, the global contour of each particle in the form of a set of polygons whose points are encoded in an optimal RLE (Run-Length Encoding segmentation) data format to reduce the size of the data. These annotations are generated in the SOLO (synthetic optimized labeled objects) format, which is specific to Unity.

After storing the images and annotations in a folder, the data in SOLO format are converted to COCO format using Pysolotools, a Python package developed by Unity Technologies. The data is subsequently converted to an eXtensible Markup Language (XML) structure to attach an XML file containing the data for each image. This data format allows annotations to be stored and organized hierarchically. In this work, a python script was developed to convert the annotation file to XML format while preserving the data structure of the COCO annotations.

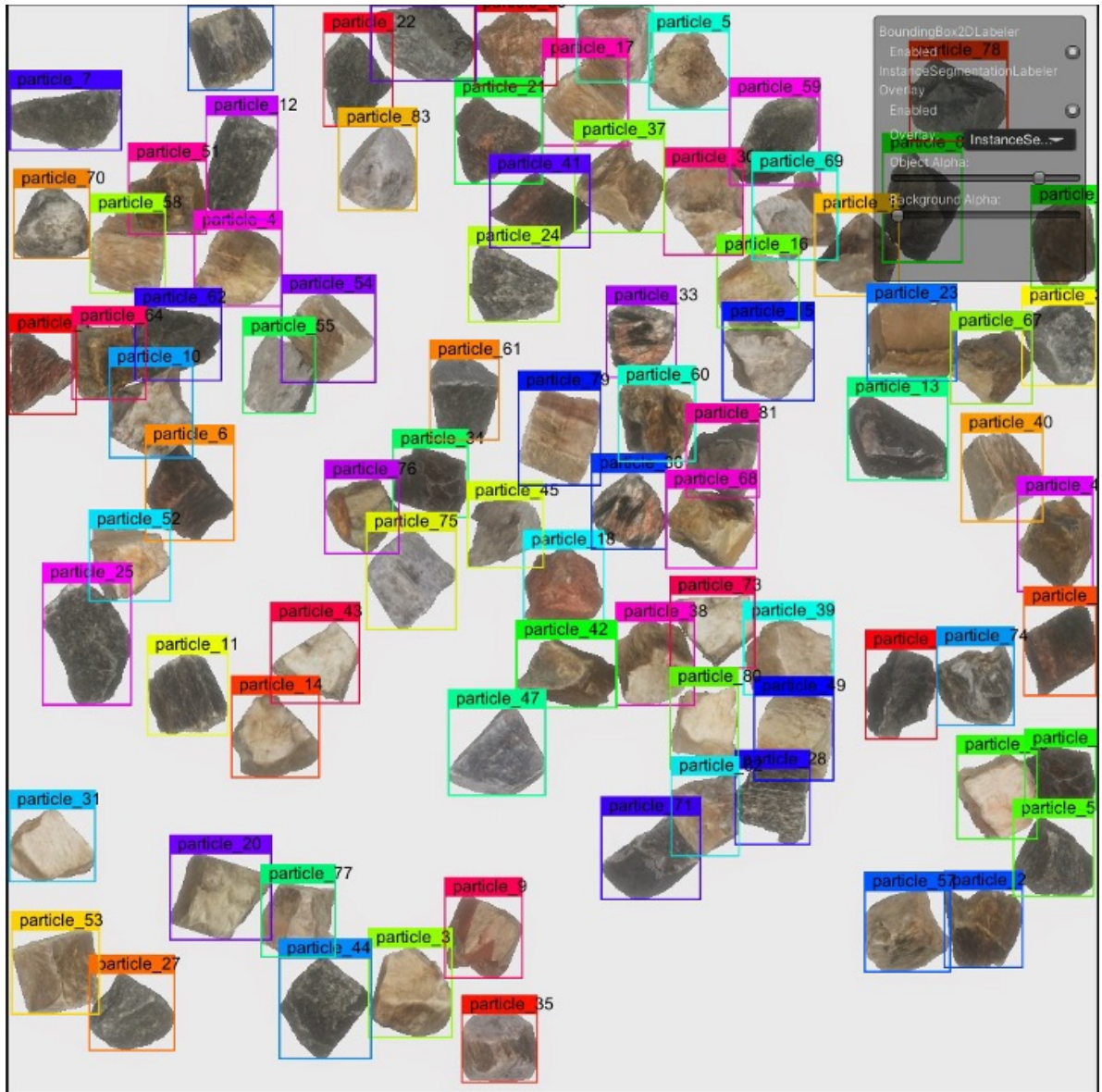


Figure 2.3 An example of image annotation and labeling during the generation of synthetic samples using Perception

2.3.3 Validation and calibration of synthetic images

The visual similarity of the synthetic images and real photographs of granular materials was verified by comparing synthetic images and real photographs with matching PSD, mass, and scale. The conditional entropy of the two sets of images was compared. As mentioned previously, conditional entropy, a textural feature, can be used to predict the PSD of a material.

It can also be used for image segmentation (Kittaneh et al., 2016; Zhang et al., 2003). Two images with the same PSD are expected to have similar conditional entropy.

Two physical samples of granular material were prepared in the laboratory. Their PSDs are given in Figure 2.4. The mass of samples 0001 and 0002 were 1061.40 g and 1220.33 g, respectively. The two samples were placed in a clear plastic container with the same dimensions as the virtual container used to create the synthetic images. Photographs of the samples were taken from the top and below using two Nikon Z6II cameras with 50 mm objectives. Each sample was mixed and photographed three times.

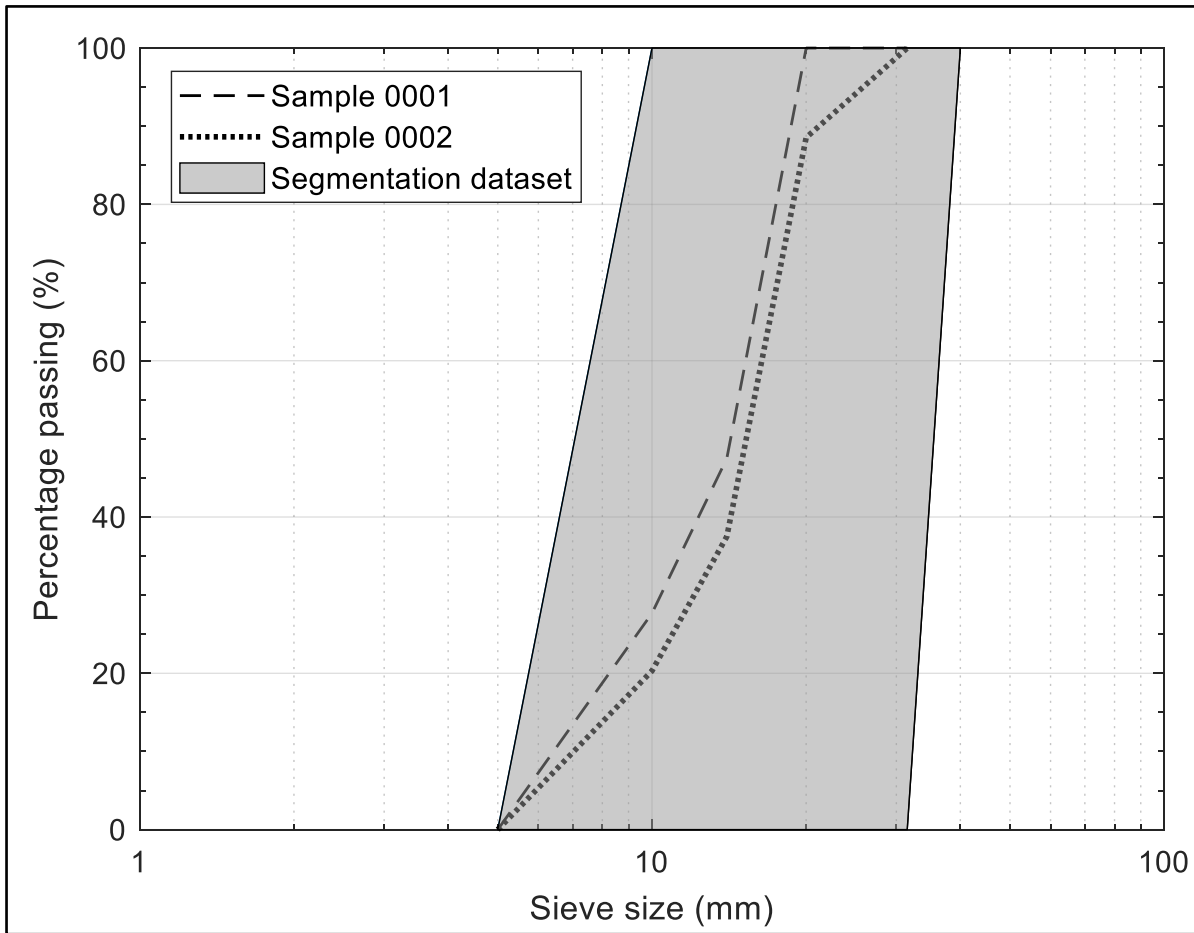


Figure 2.4 PSD for the validation images and PSD range for the segmentation dataset

The script described in the previous sections was used to generate six synthetic samples corresponding to the same PSD, mass, and scale as the real photographs. Screenshots were taken from the top and below resulting in 12 synthetic images matching the real photographs.

As a comparison, synthetic images were also generated with the discrete element code Yade based on the method presented by Pirnia et al. (2018). Screenshots were obtained from the top and below to generate 12 synthetic images with spherical particles. The images were generated to have the same PSD, mass, and scale as the real photographs.

Conditional entropy (Shannon, 1948) was used to compare the texture of the 36 real, synthetic, and Yade images. Conditional entropy H is a measure of the variability of gray levels for the pixels of a neighborhood X (Wu et al., 2013). Let $I = \{I_1, I_2, \dots, I_{N_g}\}$ be a discrete variable of image intensity, let X define a region of pixels in the image and let $p(I_i|X)$ represent the distribution of intensities I_i within region X . Let N_g be the number of distinct gray levels considered in the analysis (in this case 2). The conditional entropy is then defined as:

$$H(X) = - \sum_{i=1}^{N_g} p(I_i/X) \log_2[p(I_i/X)] \quad (2.4)$$

A neighborhood or an image with a constant gray level will have zero entropy while higher entropy indicates greater local variation in pixel intensity. Conditional entropy can be calculated for neighborhoods centered on each pixel of an image to generate a map. The mean and standard deviation of the Conditional entropy map can be calculated to describe the overall entropy of an image and its variability. A low mean entropy value indicates relatively constant gray level at the neighborhood scale or a coarse-grained texture. In contrast, a high mean entropy value signifies a fine-grained texture (Gonzalez & Woods, 2018).

Before calculating the entropy, each synthetic and real image was resized to 299 x 299 pixels and converted to the HSV color space (Hue, Saturation, and Value) (Figure 2.5a). Emphasis was placed on the value channel because it represents the brightness or intensity of the image, which is less sensitive to color variations and lighting conditions than the hue and saturation

channels. It is therefore particularly useful for analyzing textures without the influence of lighting variations or color deviations (Gonzalez & Woods, 2018).

A binary thresholding method was then applied to the V-channel, using Otsu's (1979) method. This is an automatic thresholding technique that seeks to minimize the variance between classes, thus maximizing the separation between the foreground (in this case the particles) and the background of the image (in this case, the particle contours). By applying Otsu's method, particle contours are effectively isolated, enabling a more targeted comparison of their textures (Figure 2.5b). The aim of these pre-processing steps was to eliminate potential interferences from illumination or color variations to make synthetic and real domain images better suited to the comparison of granular structures than grayscale images.

Circular neighborhoods with radii of 2, 5, 10, 14 and 20 pixels were used to calculate conditional entropy maps. These radii are arbitrary, but of the same order as those used by Duhaime et al. (2021) to predict the PSD of granular materials from the conditional entropy. The means and standard deviations of the conditional entropy was calculated for each image and each entropy map to compare the texture of the real and synthetic images.

The calibration process for the synthetic images was designed to optimize their textural similarities with the soil photographs from the real domain. It mainly involved modifying the lighting intensity in the 3D scene and adjusting the background and the color of the transparent glass supporting each sample. These parameters were adjusted in an iterative cycle until a significant match was obtained between the means and standard deviations distributions of the entropies of the real and synthetic images (Figure 2.5).

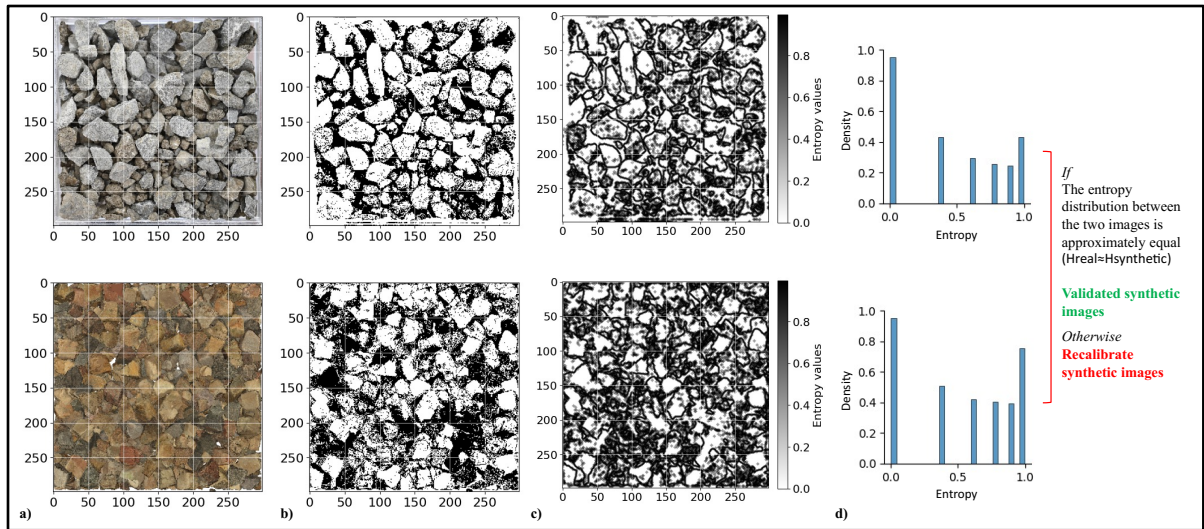


Figure 2.5 Algorithm for synthetic image validation based on local entropy, a) Original images, b) V-channel images thresholded using the Otsu method to focus on the particle contours, c) Images entropy calculated for a neighborhood with a 2-pixel radius, d) Histogram of conditional entropy for the images

2.3.4 GrainSegNet training and testing

This section presents an application example for the synthetic images of granular material. In this example, a new segmentation neural network, GrainSegNet, was trained using synthetic images generated in Unity with the method presented in this article. GrainSegNet was then tested with real photographs of granular material and rockfill to verify if a synthetic dataset could be used to train a network for the segmentation of real images.

A Python script was developed to create a set of 20 500 random PSD curves corresponding to the PSD range shown in Figure 2.4. For each PSD, a virtual sample corresponding to a total mass of 500 g was generated. This mass represents between 10 and 1000 particles per virtual sample, with some overlapping particles. An image and a segmentation mask were generated for each PSD curve. The 20 500 images were generated in 72 hours with a 2.8 GHz Core i7 x8 processor with 16.0 GB of RAM, and a 4 GB graphics card. The 20 500 images were split between a training dataset comprising 20 000 images and a validation dataset comprising 500 images. Each image used for training had a size of 299 x 299 pixels. This size was chosen so

that the smallest particles with a size of 5 mm have a size of 8 pixels in the image for a scale of 1.66 pixels per mm.

The overall structure of GrainSegNet is presented in Figure 2.6. Transfer learning creates a neural network by adapting and retraining an existing pre-trained neural network. The implementation is based on the open-source Mask R-CNN framework developed by Abdulla (2017), which provides a Keras and TensorFlow-based reference for instance segmentation tasks. GrainSegNet encompasses the basic structure of R-CNN Mask (He et al., 2018), which was adapted for the two classes featured in the synthetic images (particles and background). R-CNN mask performs classification, detection, and segmentation tasks. This model consists of three complementary modules. A CNN is responsible for image recognition through visual feature extraction. These features are integrated into a region proposal network (RPN), which identifies candidates likely to contain objects in the image. Bounding boxes (bboxes) and confidence scores are generated to focus attention on relevant areas of the image. A region of interest alignment (ROI align) is applied to each region proposed by the RPN as a quantizer that samples each sub-region for a better appreciation of the contours of the objects identified. The aligned regions are sent to classification and regression branches to label each object and refine the size of the bounding box.

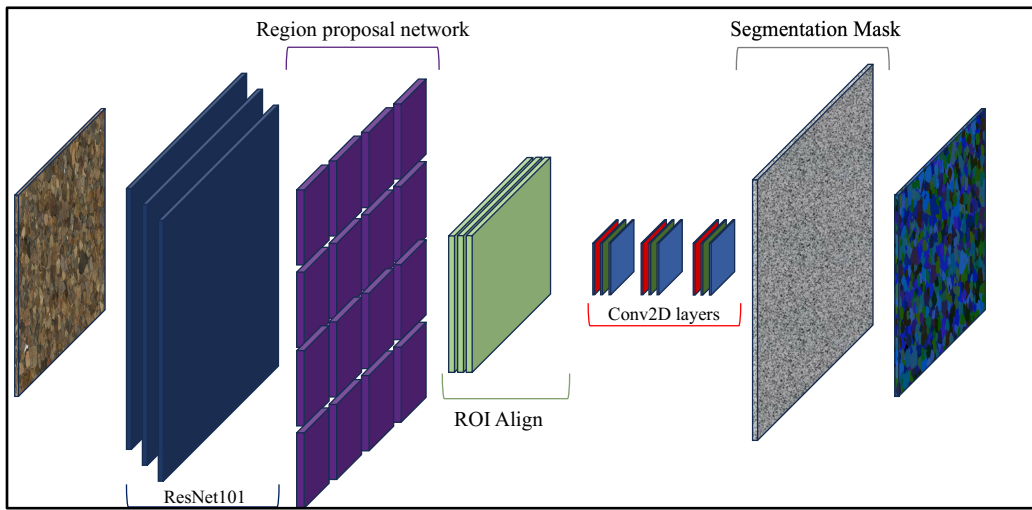


Figure 2.6 Simplified architecture of GrainSegNet

Table 2.1 summarizes the general configuration of the model. The initializer (backbone) used to extract the image features was the ResNet101 model (He et al., 2016). The possibility of replacing the backbone with ResNet50 was investigated, yielding results that were similar to those obtained with ResNet101. All layers remained trainable throughout the training.

Training a neural network involves minimizing a loss function which compares predicted and target values. GrainSegNet was trained using the sum of two loss functions for bounding box localization ($L_{box|RPN}$ and $L_{box|RCNN}$), two loss functions for classifying the object instances present in the image ($L_{classification|RPN}$ and $L_{classification|RCNN}$), and a loss function for the segmentation mask (L_{mask}) (He et al., 2018):

$$Loss = (L_{classification|RPN} + L_{classification|RCNN}) + (L_{box|RPN} + L_{box|RCNN}) + L_{mask} \quad (2.5)$$

$$L_{classification} = Cross\ entropy = - \sum y_{pred} \log(y_{target}) \quad (2.6)$$

$$L_{bbox} = Mean\ squared\ error = \frac{1}{N} \sum (y_{pred} - y_{target})^2 \quad (2.7)$$

$$L_{mask} = -\frac{1}{m^2} \sum_{i,j \in [1,m]} (y_{ij,target} \log(y_{ij,pred}) + (1 - y_{ij,target}) \log(1 - y_{ij,pred})) \quad (2.8)$$

where y_{pred} stands for the predicted values and y_{target} stands for the actual (target) values, m^2 is the total number of pixels, i is the instance of the object category, and j is the pixel under consideration.

GrainSegNet was tested with two test datasets. The main objective was to evaluate the performance of a model trained with synthetic images for other synthetic images and real photographs. The first test dataset was composed of 50 synthetic images prepared with Unity. The PSD of the synthetic images were in the same range as the training and validation images. The second test dataset was composed of real photographs. It included images of rockfill taken in the field and images of real soils taken in the laboratory. Six images were manually annotated in COCO format using the MakeSense.AI platform, including both bounding boxes and segmentation masks. The annotation process required approximately 15 minutes per image

on average. The photographs taken in the laboratory were obtained using a test set-up with a scale similar to the scene created in Unity (Figure 2.2).

Table 2.1 Global configuration of GrainSegNet

Parameter	Value
STEPS_PER_EPOCH	30
VALIDATION_STEPS	11
DETECTION_MIN_CONFIDENCE	0.7
DETECTION_MAX_INSTANCES	2 000
MAX_GT_INSTANCES	2 000
IMAGES_PER_GPU	2
BATCH_SIZE	2
LEARNING_RATE	0.0003
EPOCHS	14

The evaluation was carried out using the standard Intersection over Union (IoU) metric. The IoU quantifies segmentation accuracy by measuring the overlap between the regions predicted by the model and the actual segmented areas:

$$IoU = \frac{|Prediction \cap Ground\ truth|}{|Prediction \cup Ground\ truth|} \quad (2.9)$$

where a IoU value close to 1 indicates strong agreement between prediction and ground truth, while a value close to 0 indicates poor segmentation performance.

2.4 Results and discussion

2.4.1 Modeling and annotations of synthetic images

Figure 2.7 shows examples of synthetic images generated using the Unity game engine with their respective segmentation mask. Figure 2.7a shows an image of the validation dataset used to evaluate the performance of the GrainSegNet model corresponding to a soil sample with a

total mass of 500 g. Figure 2.7c shows a synthetic soil sample that simulates the PSD and mass of real sample 0002. An example of complete annotation using the Perception Toolkit is illustrated in Figure 2.8.

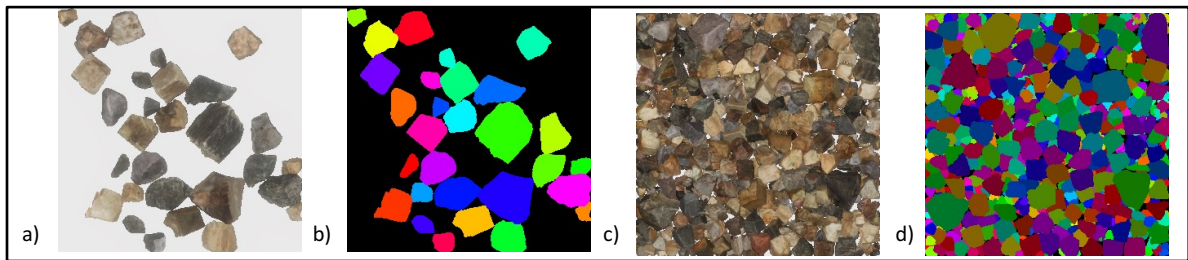


Figure 2.7 a) Sample synthetic image from the validation dataset for GrainSegNet and b) its segmentation mask, c) synthetic validation image corresponding to sample 0002 and d) its segmentation mask

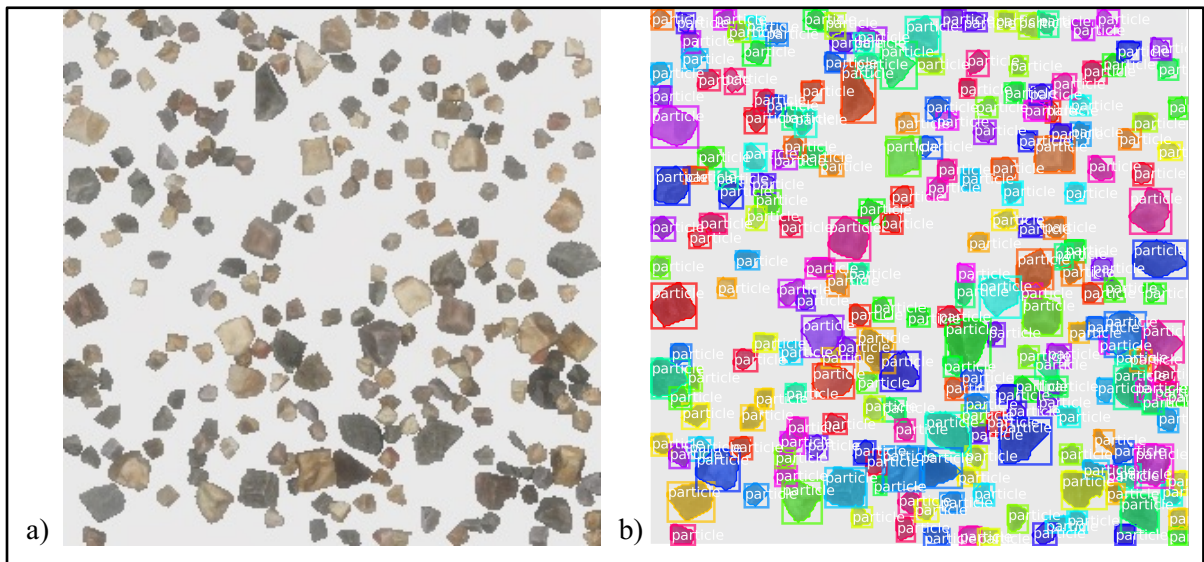


Figure 2.8 Example of synthetic image of granular material (a) and its Perception annotations including bbox, mask and label for each particle (b)

2.4.2 Validation and calibration of computer-generated images using real photographs

Figures 2.9 and 2.10 compare the real photographs for samples 0001 and 0002, with the synthetic images that were obtained with Unity after calibration. The synthetic images clearly

reproduce the overall image texture and the finer-grained aspect of the images taken from below. Figure 10 compares the color histogram in the HSV format, where hue (H) represents the dominant color (ranging from 0 to 179°), saturation (S) measures the color intensity (ranging from 0 to 255), and value (V) indicates the brightness level (also ranging from 0 to 255). Synthetic images of perfect spheres of random colors generated using the discrete element code Yade (Pirnia et al. 2018) are also shown as a comparison. Examination of the color histograms for the real photographs and synthetic images reveals clear differences in the distribution of colors. Both the Unity images and the real photographs show distinct peaks on the H curve near $20\text{--}30^\circ$, which corresponds to brown and yellow hues. Compared to the Unity images, the real photographs show a larger number of distinct peaks on the hue curves. Another difference between the real photographs and Unity images resides in the saturation, which is lower for the real photographs because of their gray tone. The Yade images show flatter H , S , V curves because they include many different colors.

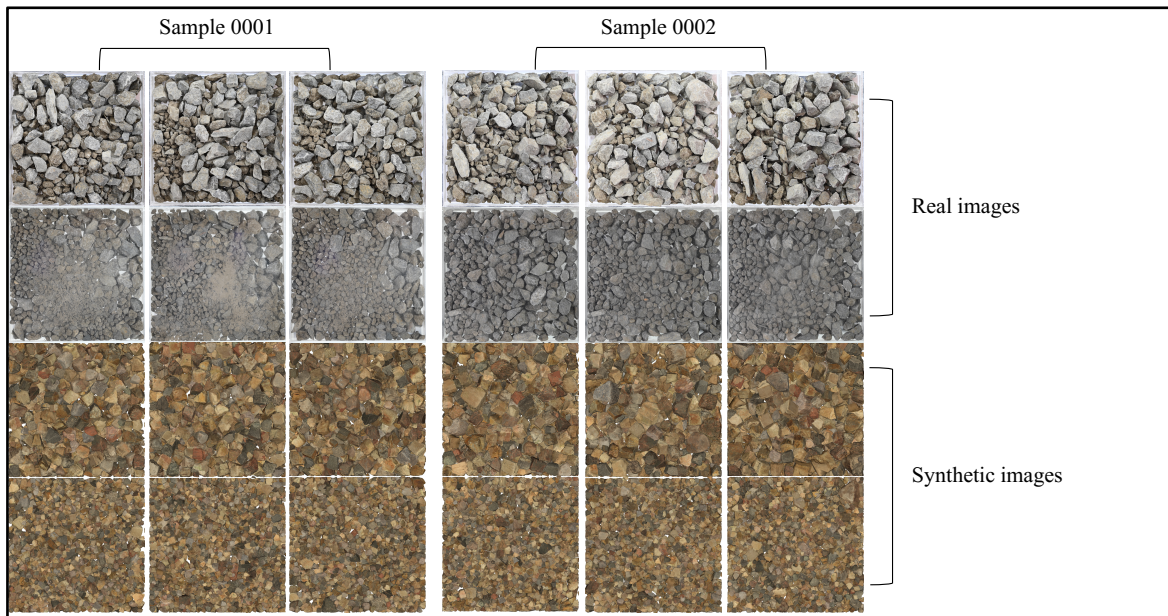


Figure 2.9 Top and bottom views of real photographs and their corresponding synthetic images used for calibration

Figure 2.11 shows the V channel after binary thresholding with Otsu's method for the real photographs and synthetic images. For the real photographs, the binary thresholding

distinguishes the particles (white pixels) from the particle contours (black pixels). Thresholding of the Unity images also segments the particles and particle contours, but some particles are segmented with their contours and appear black in Figure 11. The thresholding for the Yade images is very different. It shows a strong influence of the lighting conditions with the shadow side of each particle (bottom side of each particle) shown in black. The thresholding shows that the particle contours are more contrasted in the real photographs compared to the synthetic images, but that the lighting of the Unity images is more similar to the real photographs compared to the Yade images.

Figure 2.12 compares the standard deviation and mean value of conditional entropy for real photographs and synthetic images obtained with Unity and Yade from the top viewpoint. Conditional entropy describes the image texture and the variability of pixel intensity at the neighborhood scale (Duhaime et al. 2021). For larger neighborhoods (radii of 10, 14, and 20 pixels), the synthetic images obtained with Yade and Unity yield mean values of conditional entropy that are similar to those of real photographs with the same PSD. This is consistent with the overall appearance of the images presented in Figures 2.9 and 2.10. For smaller neighborhoods (radii of 2 and 5 pixels), the mean conditional entropy differs based on image type. The Yade images have a lower mean entropy value compared to the real photographs and Unity images. This can be explained by the smoother surface of the individual spheres in the Yade images compared to real particles. This smooth surface leads to a lower variability of pixel intensity at a small scale. On the other hand, the Unity images show a higher mean conditional entropy at a small scale compared to the real photographs. This is probably due to the relative uniformity of the particle surface texture and colour in the real photographs. The standard deviation of conditional entropy is generally higher for the real photographs, especially for larger neighborhoods. This is likely related to the more contrasting particle contours in the real photographs and the slightly more heterogeneous appearance of real photographs. For example, for the top image of sample 0002 in Figure 2.10, the lower half of the image is clearly finer-grained compared to the top half. The Unity and Yade images are generally more homogeneous with a less obvious influence of segregation on horizontal planes.

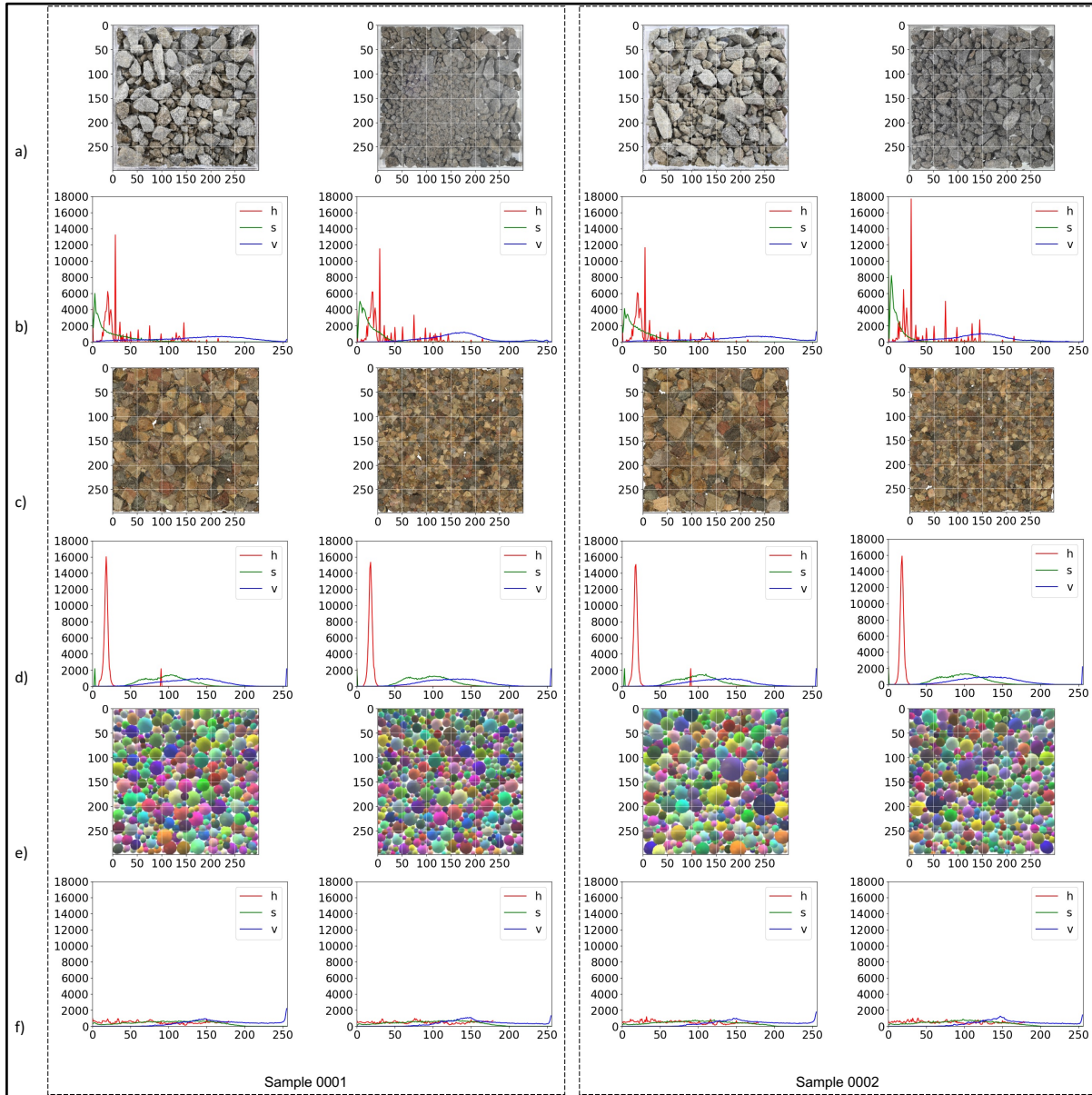


Figure 2.10 Histograms of the HSV color channels for the real (a), synthetic (c), and Yade-generated (e) images, corresponding to subsamples 0001 and 0002. The x-axis represents the pixel intensity values of each channel (hue: 0–179°; saturation and value: 0–255), while the y-axis indicates the number of pixels for each intensity level

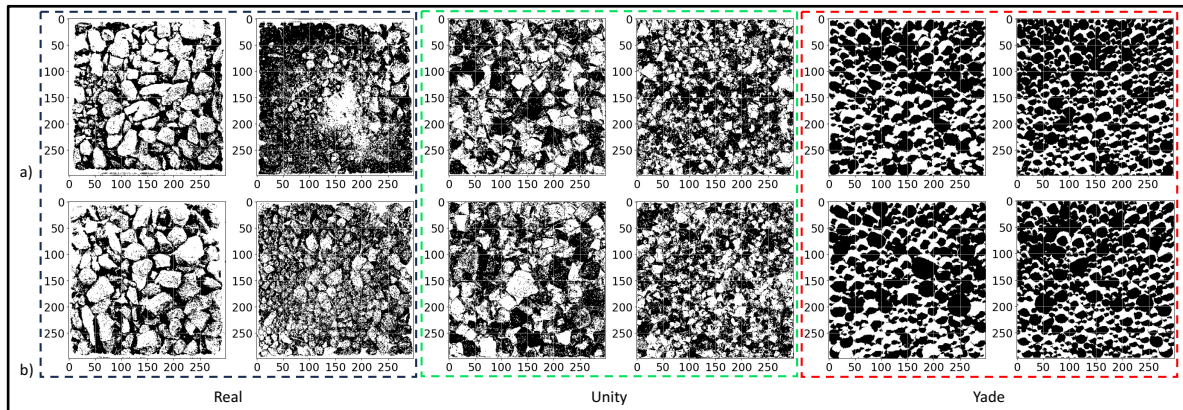


Figure 2.11 V-channel images thresholded using Otsu's method for samples 0001 (a) and 0002 (b). The top and bottom views of the samples are shown on the left and right, respectively. Images are shown for real photographs, Unity, and Yade

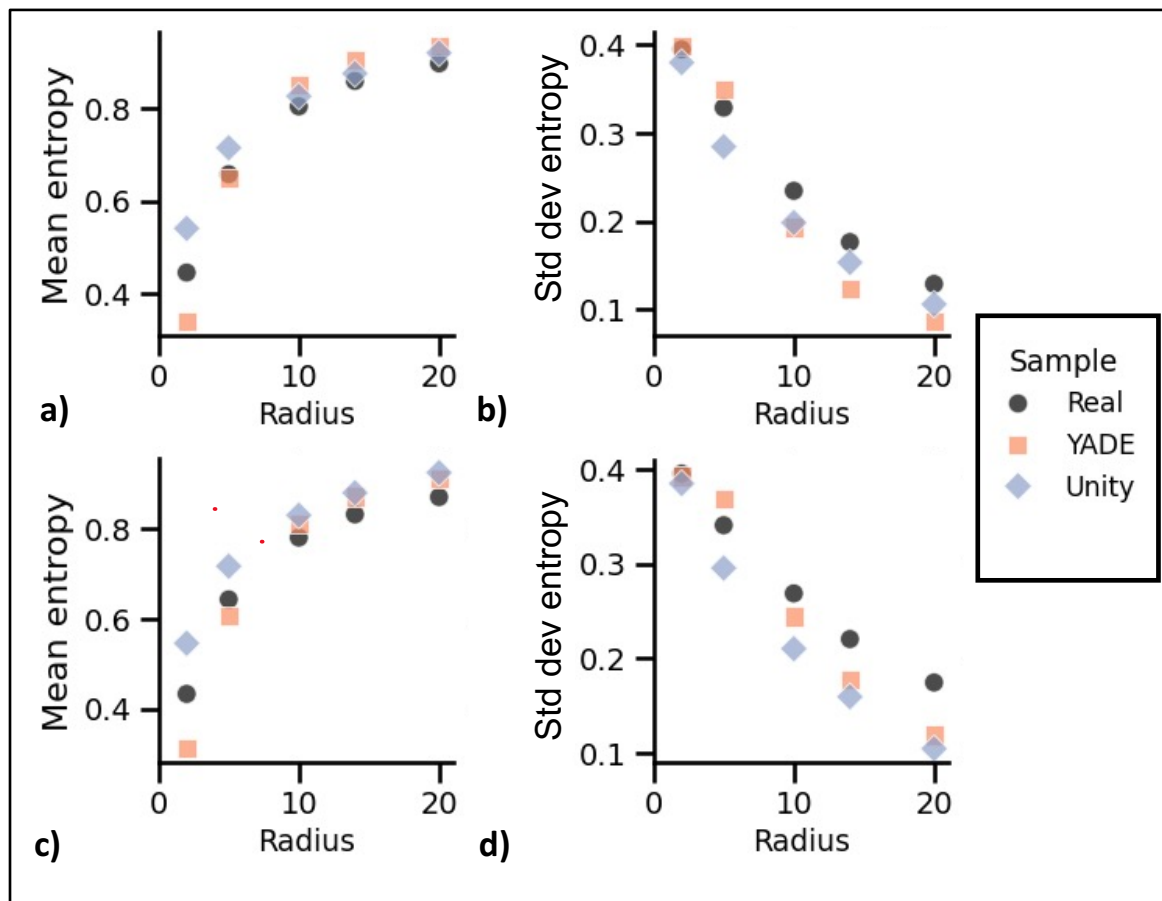


Figure 2.12 Comparison of the mean (a, c) and standard deviation (b, d) of conditional entropy for real photographs and synthetic images (Yade and Unity). The first row corresponds to the top view of sample 0001 (a, b) and the second row to the top view of sample 0002 (c, d)

As shown by Duhaime et al. (2021), conditional entropy is related to particle size: images containing larger particles exhibit less small-scale texture variation and thus lower entropy, particularly when the neighborhood size is smaller than the particle size. This trend is apparent in the validation and calibration dataset when comparing the mean and standard deviation of conditional entropy between images taken from the top and bottom viewpoints. Figure 2.13 illustrates this comparison for sample 0002. The contrast in conditional entropy between top and bottom images reflects the segregation observed in the photographs (Figure 2.9). Segregation is more pronounced in the real photographs. As a consequence, the difference in conditional entropy is also greater. In contrast, the Yade images—which show minimal segregation—exhibit nearly identical mean entropy values for both viewpoints. The Unity images show more segregation than the Yade images, but still less than the real photographs.

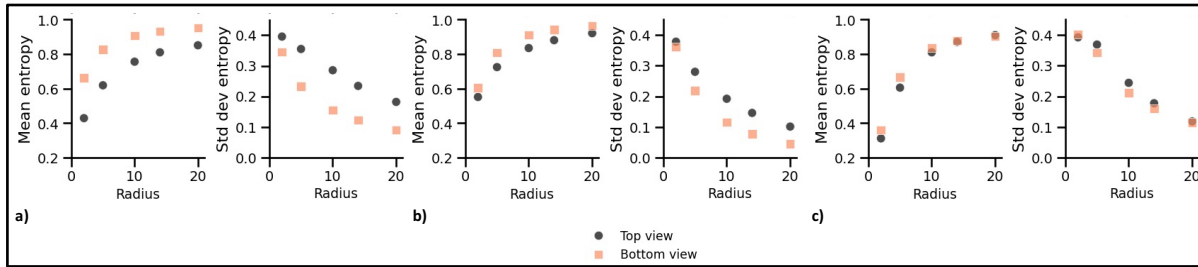


Figure 2.13 Comparison of the mean and standard deviation of conditional entropy for real images (a), synthetic images generated with Unity (b), and synthetic images generated with Yade (c), for the top and bottom views of sample 0002

The similarity of small-scale texture for the Unity images and real photographs could potentially be improved by choosing other particle models or a subset of the particle models that were used in the current script. Small scale texture could have an influence on the performances of CNN models trained with synthetic images. CNN models use filter sizes of 3x3 or 5x5 pixels in their first layers to extract local features from images. If synthetic images prepared with the Unity game engine have very different small-scale texture compared to real photographs, the first convolutions are likely to learn features that are not representative of real images and thus impair the model's ability to generalize.

2.4.3 GrainSegNet Performance

Figure 2.14 shows the training and validation loss for GrainSegNet. After training, a main loss of 0.99 was achieved for the validation dataset. The cumulative bounding box loss, cumulative classification loss, and the segmentation mask loss for the validation dataset were 0.52, 0.31, and 0.16, respectively.

The performance of GrainSegNet on the first test dataset consisting of 50 synthetic images shows excellent segmentation robustness. A mean IoU of 0.88 was obtained. The segmentation masks that were obtained for the images with the highest (0.9385) and lowest (0.8247) mean IoU are presented in Figure 2.15. A histogram of the IoU scores for the 2641 particles is presented in Figure 2.16. The IoU is above 0.8 for most particles.

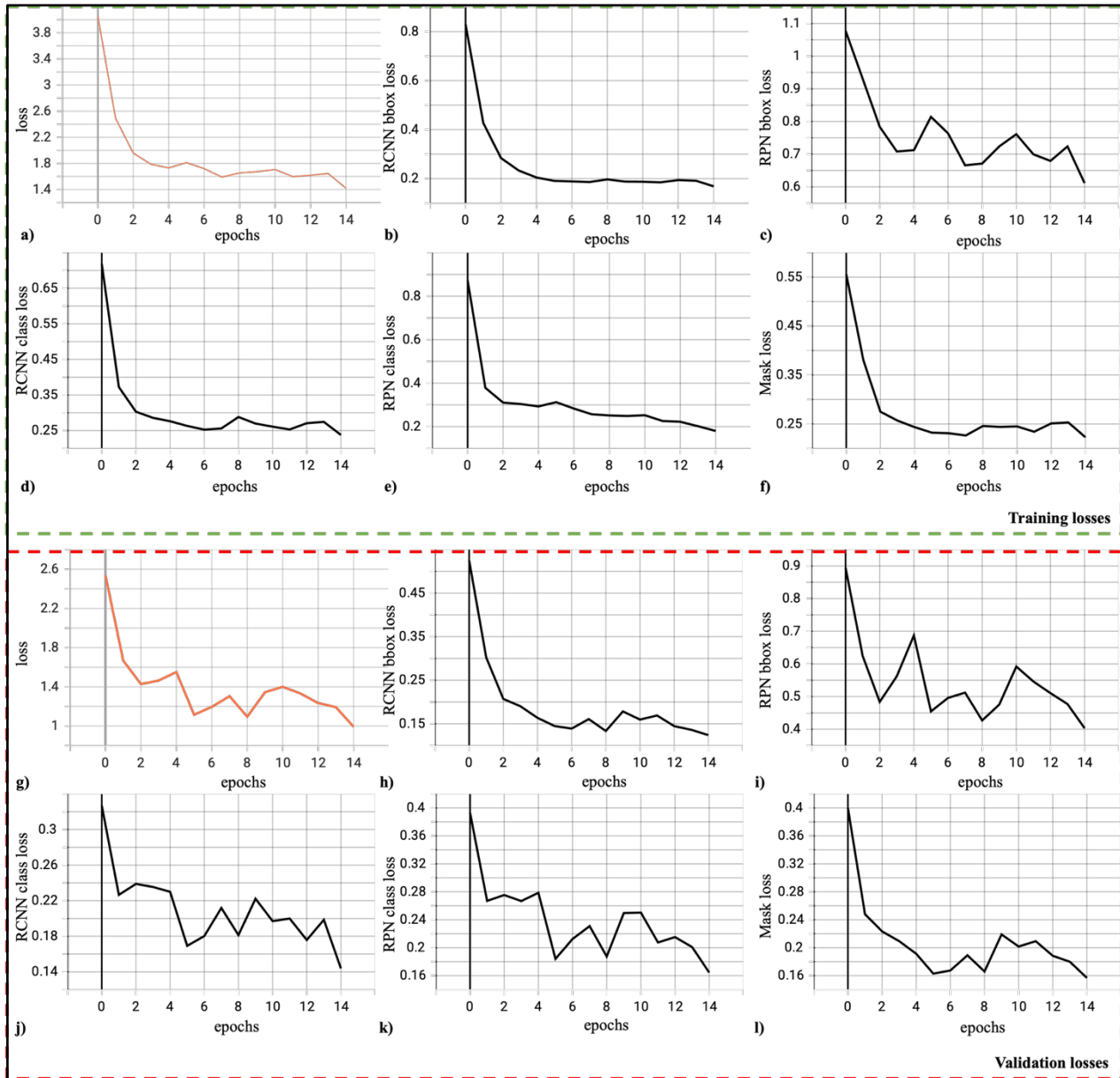


Figure 2.14 GrainSegNet training losses (a-f) and validation losses (g-l). Total losses are shown in subfigures a) and g), RCNN bounding box regression losses in subfigures b) and h), RPN bounding box regression losses in subfigures c) and i), RCNN classification losses in subfigures d) and j), RPN classification losses in subfigures e) and k) and mask losses in subfigures f) and l)

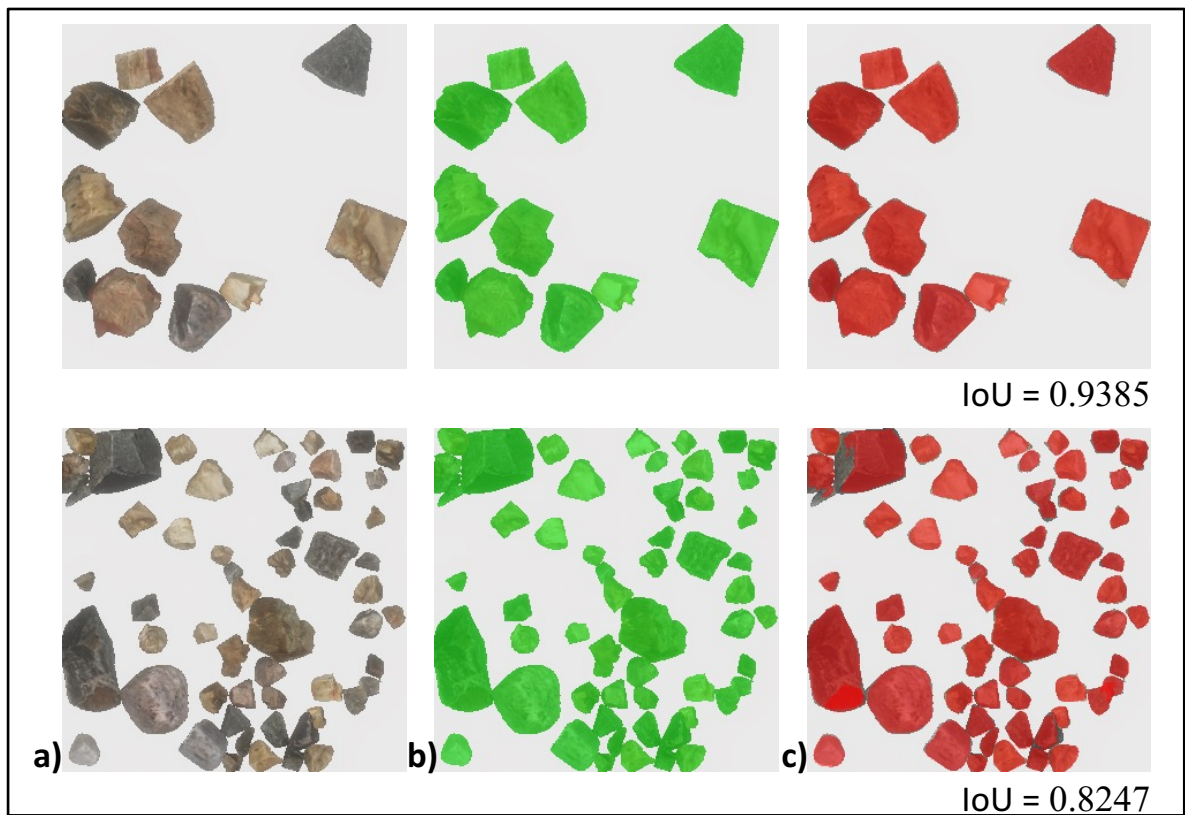


Figure 2.15 Synthetic images for the first test dataset (a) corresponding to the highest (top row) and lowest (bottom row) mean IoU scores. Images b) shows the ground truth segmentation masks generated using the Perception Toolkit and images c) shows the predicted segmentation masks

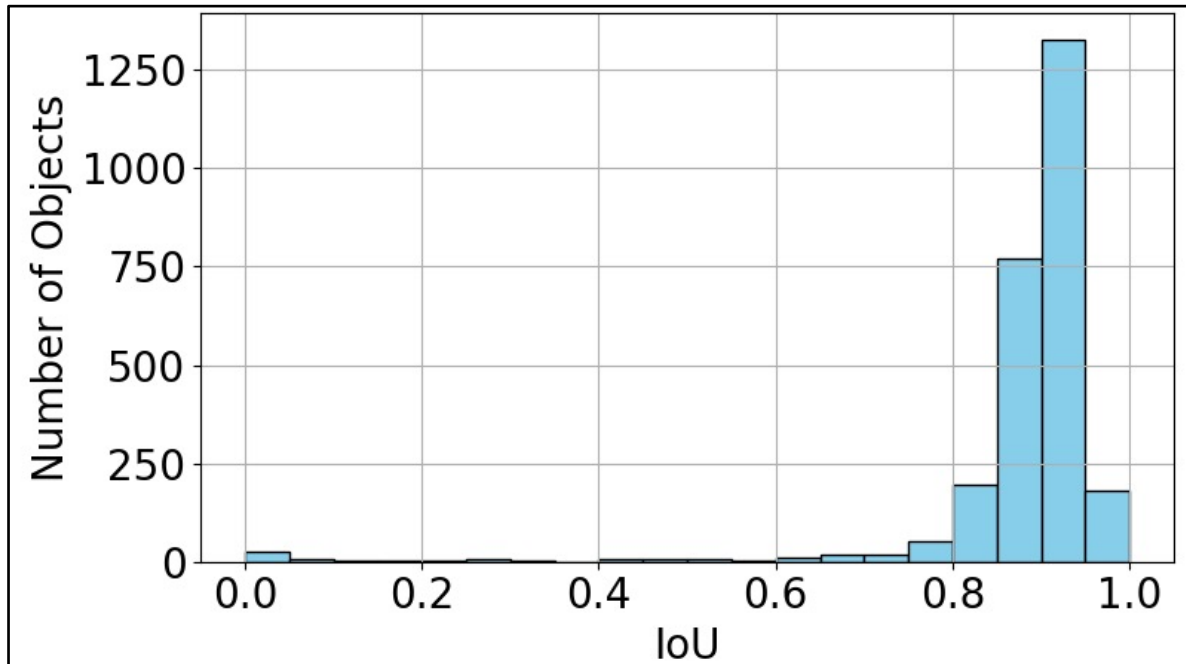


Figure 2.16 Distribution of IoU scores per object across the entire test dataset composed of synthetic images

The second test dataset was composed of real photographs of granular material and rockfill taken in the laboratory and in the field. Figure 2.17 compares the manual segmentation masks with the GrainSegNet predictions. Figure 2.18 shows the IoU histogram for all objects in the six photographs and the mean IoU for each photograph. Although the model was trained exclusively with synthetic data, the predicted segmentation masks for these photographs are generally good with a mean IoU of 0.41 and a standard deviation of 0.35. The high standard deviation indicates significant variability in GrainSegNet performance. Results were better for the laboratory photographs (photographs 1-2), which correspond to a set-up similar to the scene in Unity. For the field photographs, GrainSegNet succeeded in identifying most particles present in an environment where the particle number was low and the particle size similar to the training images (photograph 3). On the opposite, lower mean IoU scores were obtained for rockfill with small particles (photograph 4) and rockfill characterized by a higher number of particles (photographs 5-6). These results indicate that GrainSegNet performed well on relatively simple laboratory or site data, and that further adjustments are required for environments that differ markedly from the scene in Unity.

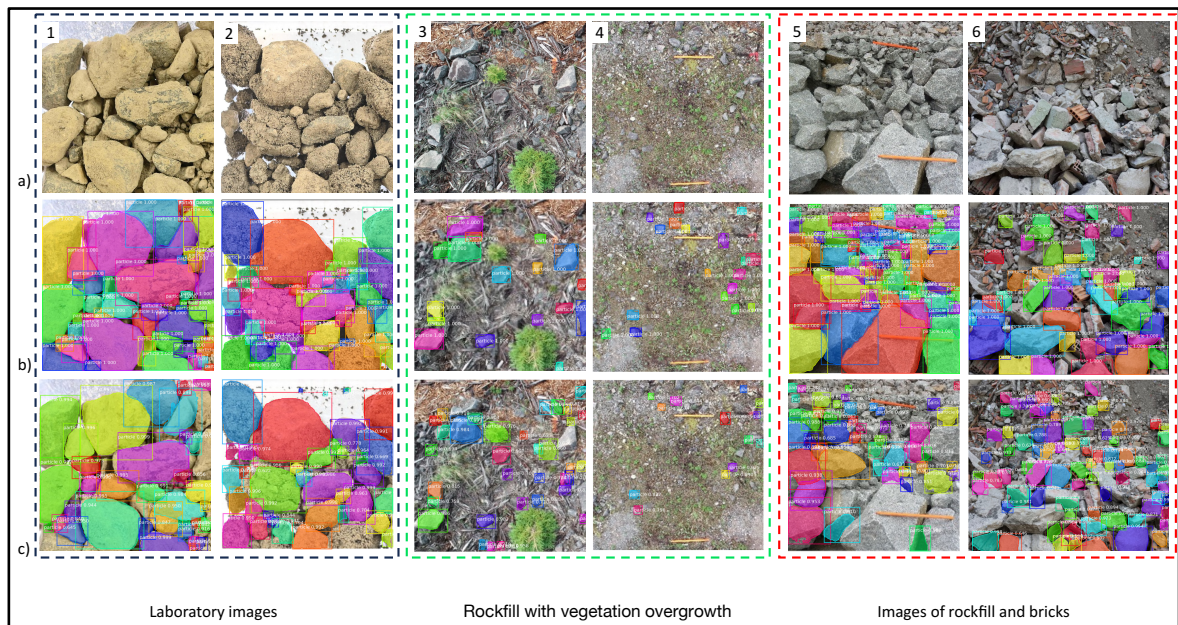


Figure 2.17 Evaluation of GrainSegNet on real soil photographs. (a) original images; (b) manual segmentation using MakeSense.AI; (c) predicted segmentation generated by GrainSegNet

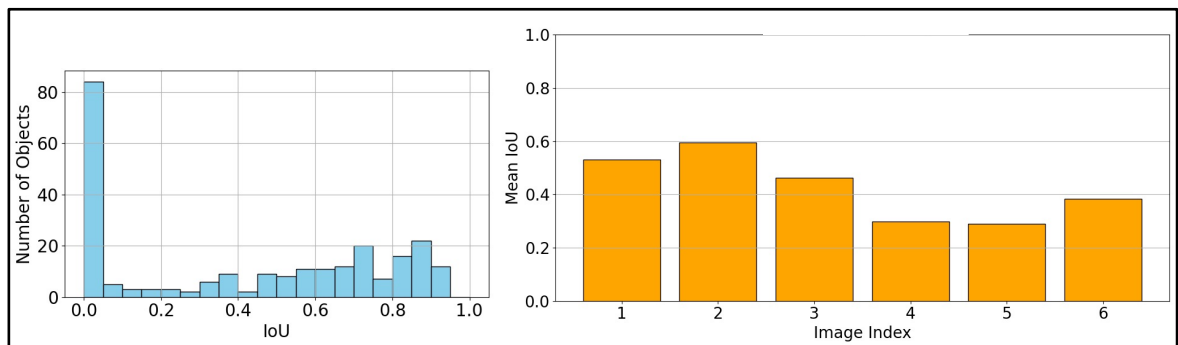


Figure 2.18 Histograms illustrating the distribution of Intersection over Union (IoU) scores for each segmented object (left) and for each image (right) in the test dataset with real photographs

The results obtained in this study can be compared with those reported in previous work on segmentation from synthetic images and segmentation of granular material. Barth et al. (2020) proposed the use of CycleGAN to improve the visual realism of synthetic images created in the context of automated bell pepper harvest. Their best configuration, combining translated

images and real photographs, achieved a maximum IoU of 0.52 for segmentation via a neural network developed in their work. Training based exclusively on synthetic images led to an average IoU of 0.31. When evaluated on real soil photographs, GrainSegNet which was trained exclusively on synthetic images achieved an average IoU of 0.88 on its own synthetic test set, without using domain translation techniques or real images. When evaluated on real soil photographs, GrainSegNet recorded an IoU of 0.41, outperforming the purely synthetic configuration of Barth et al. (2020) but remaining below their result on combined data. This performance gain underlines the ability of our synthetic data pipeline to capture the repetitive, well-defined morphology of granular particles, enabling more accurate segmentation than methods applied to highly variable natural objects (plants).

The study of Yamada & Di Santo (2022) was also aimed at segmenting granular materials, but based on training with real photographs. They do not provide IoU values, but conceptually similar recall values that exclude false positive from the denominator in Equation 8. Their recall score was 0.70 compared to 0.49 for the real photograph dataset in this study. The similarity of the images in the training and test dataset in Yamada & Di Santo (2022) probably explains their better performances, unlike in this paper's synthetic-to-real transfer scenario, which lacks such domain consistency for evaluation.

2.5 Conclusions

This article introduced a new method for creating synthetic images of granular material for the training of neural networks. Preparing the synthetic images with Unity requires a PSD and the total sample mass. Conditional entropy was used to compare the texture of synthetic images and real photographs of granular materials. A comparison of the mean and standard deviation of conditional entropy for different neighborhood sizes demonstrated that the texture of the synthetic images was similar to real photographs, albeit with some differences in segregation patterns and small-scale image texture.

As an application example, GrainSegNet, a mask R-CNN model, was trained for the automatic segmentation of granular material images. The model was trained from a database consisting of 20,500 synthetic soil images with segmentation masks generated using the Perception toolbox in Unity. A cumulative loss on the validation dataset (500 samples) of 0.99 including bounding box loss, mask loss, and classification loss, was obtained after training the model. GrainSegNet was also tested on real photographs, which demonstrated that the performance loss associated with the change in image domain is influenced by the characteristics of the real photographs (number of particles, particle size).

Other applications can be envisioned for the synthetic images that were presented in this paper. For example, the script could be used to generate images to train a CNN to predict the PSD of a granular material directly from the images, without segmenting the images (e.g. Manashti et al. 2022). The scene in Unity could also be modified to better fit a specific type of real photographs, for example by adding vegetation between the particles to resemble the field photographs in Figure 2.17.

CHAPITRE 3

DEEP LEARNING APPROACH TO CLASSIFY GRANULAR MATERIAL USING SYNTHETIC AND REAL IMAGES

Harold Simo Tenekam ^a, François Duhaime ^b, Jean Sébastien Dubé ^c, Matthew F. Toews ^d,
Laurent Pilon ^e

^{a, b, c} Département de Génie de la construction, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

^d Département de Génie des systèmes, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

^e Département de recherche et développement, Solnor Environnement Inc.
10636 Ave L'Archevêque, Montréal-Nord, Quebec H1H 3A2

Article soumis pour publication, février 2026

3.1 Abstract

Grain size analyses are conventionally performed using time-consuming sieving methods for coarse materials in geotechnical engineering. This paper proposes an innovative approach to conduct particle size distribution (PSD) analyses using soil photographs. To this end, GrainSizeNet, a convolutional neural network with a residual architecture, was trained on six datasets combining synthetic images created with the Unity game engine and images of artificial and real materials. Datasets D1, D2, and D3 were designed to assess the added value of images taken from the top and bottom, as well as the contribution of segmented synthetic images in which photorealistic textures are removed, and particle geometry is emphasized. Datasets D4, D5, and D6 were designed to evaluate the relative contribution of synthetic data and traditional data augmentation methods for improving generalization. On a common test set of 50 samples, D3 achieved the best performance with a mean absolute error (MAE) of 2.90%, followed by D2 (3.47%) and D1 (3.66%). In the second test set on 34 gravelly samples underrepresented in the real training set, classical augmentation produced an MAE of 11.54% (D4), while adding synthetic images reduced the MAE to 6.19% and 6.44% (D5 and D6, respectively). The results indicate that photographs taken from the top are sufficient for accurate PSD predictions, and that realistic synthetic images can significantly improve

generalization for underrepresented particle size classes. Finally, the model's predictions led to minimal variations in soil classification according to MTQ-1101. This stability compared with traditional methods validates the robustness of the approach.

3.2 Introduction

Particle size analysis is a key test to determine the particle size distribution (PSD) of a soil and to gain insights into its geotechnical properties (Polakowski et al., 2021; Tong et al., 2018). Conventional tests for particle size analyses include sieving, sedimentation analysis, and laser diffraction. Soil classification systems, such as the Unified Soil Classification System (USCS), are often based on the PSD. The USCS differentiates soil particles based on their size into three main categories: fine particles (clay and silt), sand, and gravel. Additional attributes, such as gradation and plasticity, are also characterized within the USCS. Soil classification provides a basis for predicting geotechnical parameters such as bearing capacity or soil permeability.

Image analysis methods are increasingly used to determine the PSD of granular materials (Chang & Chung, 2012; Iwashita et al., 2022; Purinton & Bookhagen, 2019). These methods usually entail the segmentation of each particle in the image. The segmented areas and particle contours are used to reconstruct the particle size distribution. These techniques are commonly applied in commercial software programs including Wipfrag, Fragalyst and Split-Online (Maerz et al., 2018; Raina et al., 2002; Souza, 2018). While these software programs are very popular, they are also not without their shortcomings. The emphasis on particle segmentation necessitates focusing solely on particle contours. Consequently, the agglomeration and superposition of particles present significant sources of error in particle size analysis methods using image segmentation (Frei & Kruis, 2020). A potential remedial approach involves analyzing the overall texture of the image. Soil images exhibit a variety of texture patterns, which impart a distinct identity to the overall image composition. Computer vision techniques can effectively characterize or quantify particle size distribution by leveraging these texture variations.

Deep learning is an advanced data processing method in artificial intelligence, relying on complex neural network architectures to perform regression or classification tasks. In particular, Convolutional Neural Networks (CNN) are a type of deep learning model particularly effective in image processing and computer vision tasks (Olivier et al., 2020). These architectures exploit convolution layers that apply filters to the input image, detecting features such as contours, texture patterns and spatial relationships between pixels. In addition, CNNs typically incorporate pooling and fully connected layers that progressively aggregate spatial information and transform local image features into higher-level representations.

Existing CNN-based approaches for image-based granulometry differ primarily in the type of visual cues available in the input images rather than in strictly defined architectural choices. In some studies, CNNs are applied to images where individual particle boundaries remain visually distinguishable (e.g., coarse granular materials), while in others the models operate on fine-grained soils for which particle contours cannot be resolved and prediction relies instead on texture-like patterns arising from particle assemblies (e.g., Iwashita et al., 2022). Importantly, even in approaches historically motivated by contour-based image analysis, such as Buscombe (2020), the CNN itself does not explicitly extract particle boundaries but learns implicit features from the overall image appearance.

Data augmentation remains the cornerstone of improving the performance of CNN models, especially when real-world datasets are limited. Traditional techniques such as rotation, flip, scaling or brightness and contrast allow the training datasets to be densified, but by integrating only visual variants of existing images. They still make it possible to increase the classification performance of CNN models by 2 to 6% accuracy (Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). However, their contribution in the context of particle size analysis remains to be appraised since they do not provide a new PSDs but different images with the same PSD and grain arrangement.

In particle size analysis using computer vision, different types of images provide distinct information for model training. Synthetic images offer an interesting alternative by allowing the creation of new and representative data (de la Fuente et al., 2024; Keivan Ekbatani et al., 2017; Maktabdar Oghaz et al., 2019), while segmented images tend to simplify visual information, enhancing the model's ability to analyze and interpret the images (Csurka et al., 2022). Nikolenko, (2021) specified that in contexts where real data have more complex textures, synthetic images closer to photorealism (the degree to which a synthetic image faithfully reproduces the visual characteristics of a real photograph) improved model results in classification, regression and detection tasks. It is noteworthy that inadequately calibrated or improperly utilized synthetic data can negatively impact model performance. For this reason, several authors recommend rigorous control of the quality, quantity and representativeness of synthetic data during the model training phase (Nowruzi et al., 2019; Simo Tenekam et al., 2025).

Several previous works have exploited synthetic images for PSD prediction. Pirnia et al. (2018) generated 53,130 top/bottom image pairs via YADE and obtained a root mean square error (RMSE) of 6.9% on the mass percentages passing for 5 sieve sizes between 0.106 mm and 0.71 mm. Temimi (2020) developed 3003 synthetic images with the Unity game engine to train different models to reach a RMSE of 7.5% on passing percentages of materials. Both Temimi (2020) and Pirnia et al. (2018) did not evaluate the performance of their models with real soil images.

Temimi (2020) and Manashti (2022) created both bottom-up and top-down images of granular materials in transparent containers. Manashti (2022) observed a decrease in passing percentage RMSE when using both top-down and bottom-up views as network input. However, his analysis was limited to synthetic data and did not extend to real material images. Despite these encouraging results, most studies have not explicitly compared the contribution of single-view versus dual-view configurations with real soil images.

This work focuses on the evaluation of synthetic data integration in training datasets for PSD prediction, and the combined use of images from top-down and bottom-up views. Six datasets were developed to train a series of CNN. The datasets combines real soils, artificial materials created by combining different particle sizes, and synthetic images including segmented images. Experiments were designed to verify if synthetic images can be used to fill particle size ranges that are less represented in the real image data or to focus the network attention on particle contours, an image feature that should be closely related to particle size. Soil classification based on the MTQ 1101 standard, a system similar to the USCS, was used to judge the ability of the models to correctly discriminate between granulometric classes.

3.3 Methodology

GrainSizeNet, a CNN model, was trained using different datasets to predict the PSD of granular materials for particle sizes between 0.080 mm and 31.5 mm. This section describes the datasets and the GrainSizeNet architecture.

3.3.1 Image types and datasets

The datasets were created from four image types including photographs of artificial mixtures combining multiple particle sizes, two types of synthetic images generated using the Unity game engine, and photographs of real soils. Combining different image types serves both to limit overfitting to dataset-specific features and to ensure a more uniform coverage of the PSD space, particularly in regions that are sparsely represented. Table 3.1 summarizes the characteristics of the four image types, and Figure 3.1 presents the corresponding PSD ranges. Representative images are shown in Figure 3.2.

Table 3.1 Characteristics of the four image types

Image type	Number of samples	Number of images per view
Type 1 - Artificial materials obtained by mixing and sieving	30	294
Type 2 - Synthetic images	5000	5000
Type 3 - Segmented synthetic images	2500	2500
Type 4 - Real soils	9	181

Image type 1 consists of 588 images of artificial materials, including 294 top-down and 294 bottom-up views. The samples were prepared by first mixing materials obtained during several geotechnical investigations. The mixture was then separated into seven particle size fractions using sieves in accordance with the BNQ 250-025 (2013) standard, with mesh sizes ranging from 0.08 mm to 5 mm. After separation, the fractions were recombined to form 2 kg samples, with gradations selected to provide a broad coverage of the PSD range (Figure 3.1a). These samples were then photographed using the setup shown in Figure 3.3a. The sample holder included a transparent glass plate at the bottom, enabling the acquisition of bottom-up images.

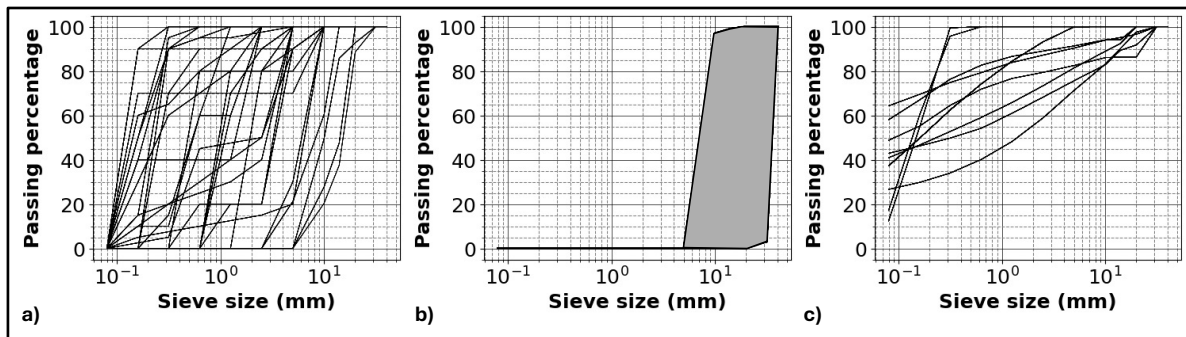


Figure 3.1 Particle size distributions of a) artificial material photographs of image type 1, b) synthetic images prepared with the Unity game engine for image types 2 and 3 and c) real material photographs in image type 4

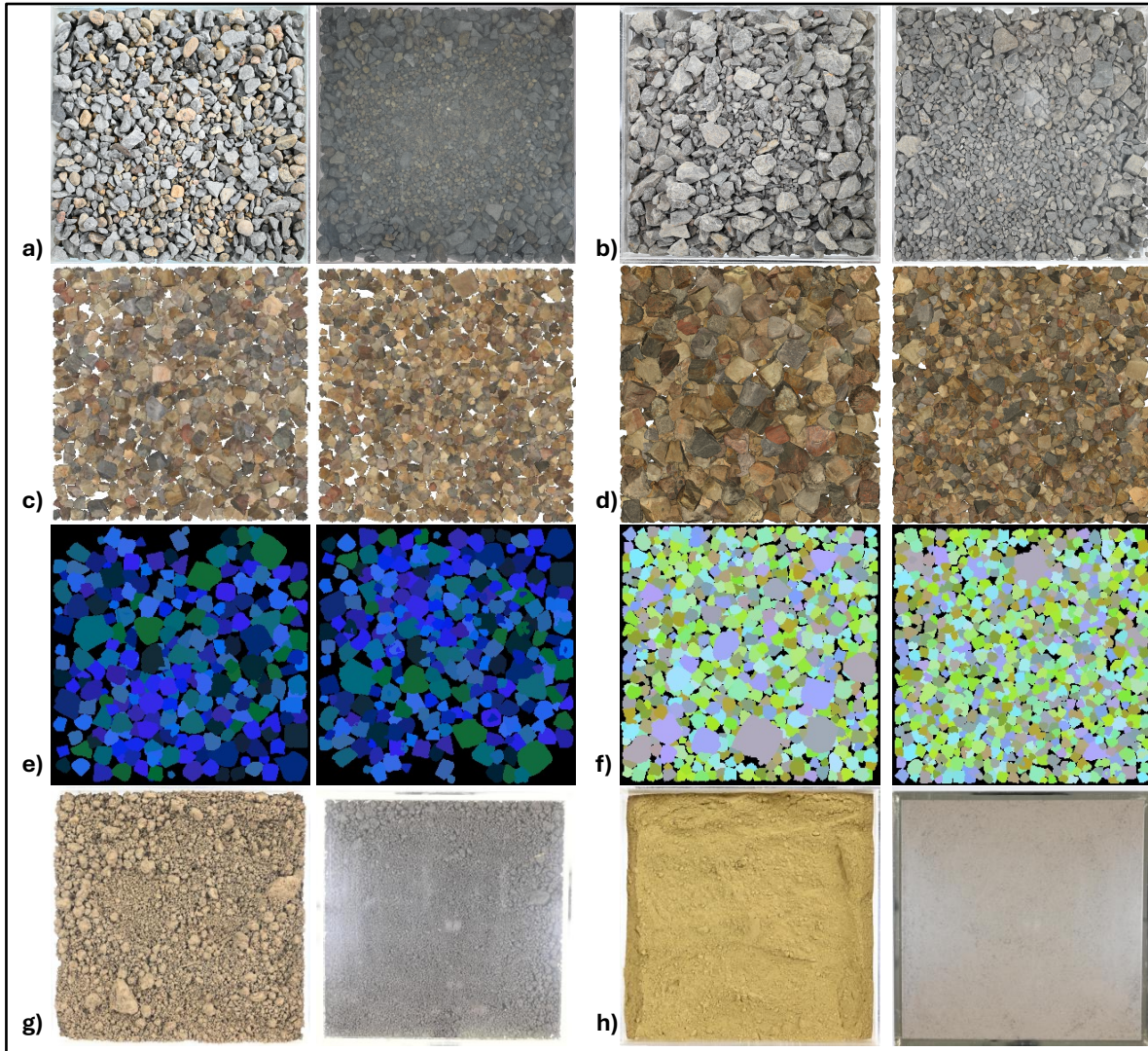


Figure 3.2 Image samples for type 1 (a and b), type 2 (c and d), type 3 (e and f), and type 4 (g and h). In each pair, the left and right images show the top-down and bottom-up views, respectively

Type 2 images were generated using a Unity script developed by Simo Tenekam et al. [20]. A total of 2500 synthetic samples were created, each consisting of 2 kg of virtual particles. The PSD for each sample was generated randomly, with particle sizes ranging from 5 mm to 31.5 mm. The range of PSD is shown in Figure 3.1b. Images were obtained for each synthetic sample from top-down and bottom-up views of the virtual sample holder (Figure 3.3b), resulting in a total of 5000 images.

Type 3 images were obtained by segmenting the type 2 images using the native segmentation mask in Unity. The segmented regions were assigned random colors to generate simplified images that preserve particle geometry and spatial arrangement while removing photorealistic texture information.

Image type 4 was created using nine real soil samples. It comprises 362 photographs acquired with the test setup shown in Figure 3.3a, including 181 top-down views and 181 bottom-up views. The samples were obtained during geotechnical investigations in Quebec. The ground truth PSD was obtained by sieving in accordance with the BNQ 2501-025 (2013) standard. The water content, defined as the mass of water divided by the mass of solids, was successively adjusted to 0 %, 3 %, 6 %, and 9 % for each sample to capture photographs under varying moisture conditions. For some samples, photographs were also taken at the natural water content. This variation in water content influences particle segregation and the overall material texture (Figure 3.4). Higher water contents can produce aggregates, which influence image texture. Sample mass ranged from 380 g to 485 g. When sufficient material was available, the mass was selected to fully cover the bottom of the transparent sample holder.

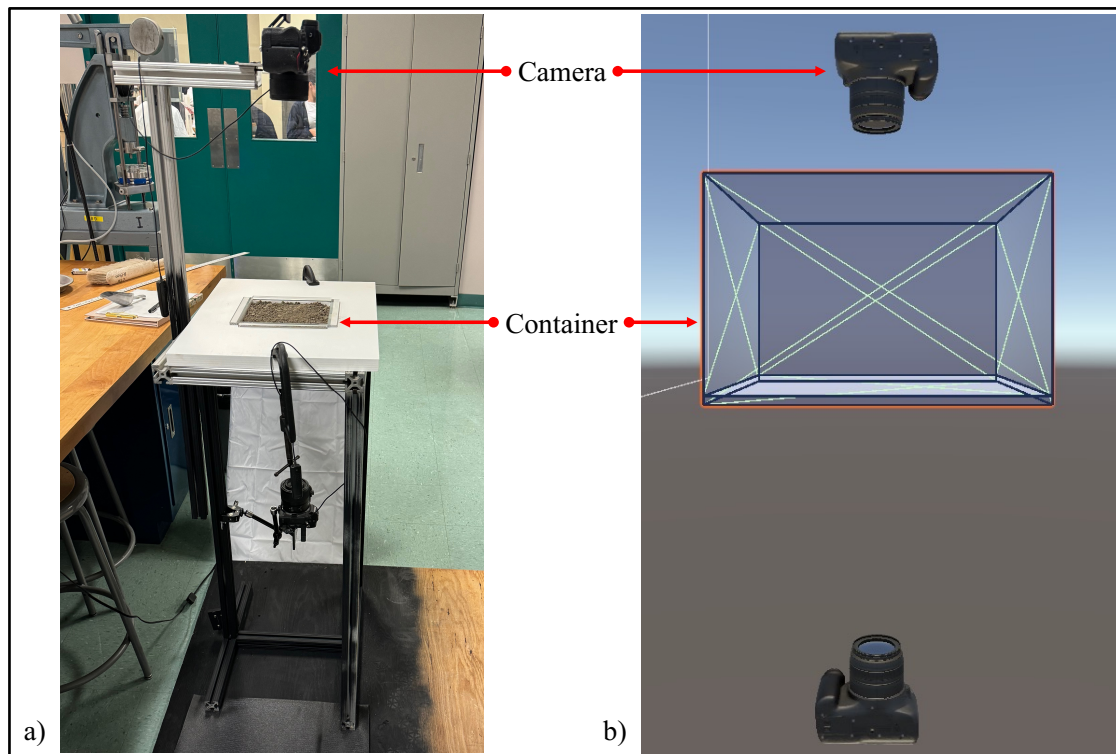


Figure 3.3 Photo setup in the laboratory (a) and Unity game engine (b)

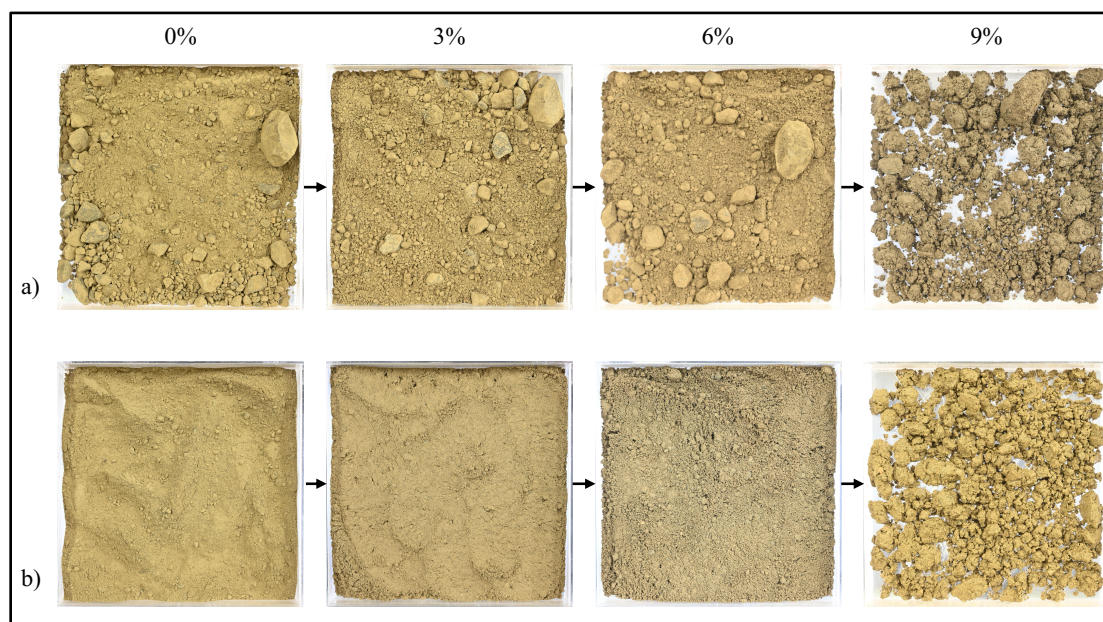


Figure 3.4 Variation in the texture of two samples (a and b) from image type 4 as a function of water content

All image types had a resolution of 299×299 pixels and a scale of 0.6 mm/pixel during model training and inference. The virtual and real sample holders had the same size (180×180 mm). Image types 1 and 4 were obtained with a Nikon Z6 II camera and a 50 mm lens with an initial resolution of 4016×4016 pixels. Bilinear interpolation was employed to resize these images to 299×299 pixels. The synthetic images (types 2 and 3) were generated with a resolution of 299×299 pixels. Low-resolution images were used to facilitate model training. However, using images with a lower resolution has a direct impact on prediction accuracy for particles smaller than 1 pixel. As demonstrated by Lang et al. (2021), a CNN can predict the PSD for subpixel particles using the overall image texture, albeit with a lower accuracy.

Six training datasets (D1 to D6; Table 3.2) were assembled from the four image types to address distinct experimental objectives. Datasets D1 and D2 were designed to evaluate the effect of the number of input images per sample (using only the top-down view versus combining top-down and bottom-up views). Datasets D3, D4, D5, and D6 were used to assess the contribution of synthetic images and segmented synthetic images in training datasets relative to conventional data augmentation methods, such as rotation, flips, and contrast or brightness adjustments.

Dataset D1 includes the bottom-up and top-down images for image types 1, 3, and 4, as well as 2,500 type 2 image pairs (total of 5,475 image pairs). Dataset D2 includes only the top-down views for the same image, with 5,475 individual images. Dataset D3 includes the top-down views for types 1, 2, and 4 (also 5,475 images). It excludes the segmented images (type 3) in order to assess their influence. The main dataset (D1), which contains paired top-down and bottom-up images, was divided into 4,500 image pairs for training, 1,850 image pairs for validation, and 50 image pairs for testing. The same split was used for datasets D2 and D3, with only single-view images. For consistency and comparability, the same 50 samples were randomly selected from the full dataset (image types 1, 2, 3, and 4) to constitute a fixed test set, which was used to evaluate the three models under identical conditions.

D4 is based exclusively on photographs of real and artificial materials. From the 475 type 1 and 4 images, 34 images of materials composed mostly of gravel were isolated to form a test dataset with a PSD range similar to the type 2 and 3 synthetic images (Figure 3.5a). The remaining 441 images were used for training and augmented using conventional techniques to introduce variability in image appearance while preserving particle arrangement. Each training image first underwent a random horizontal or vertical flip. Then, photometric transformations were randomly applied among the following: an increase in saturation by a factor of 3 in HSV space; an increase in brightness, or an increase in contrast (Figure 3.6). These operations expanded the training set to 5,280 images.

Table 3.2 Characteristics of the six datasets

Dataset	Image types	View	Data augmentation	Training images	Validation images	Test images
D1	1, 2, 3, 4	Top + Bottom	No	4,500 pairs	925 pairs	50 pairs
D2	1, 2, 3, 4	Top	No	4,500	925	50 (same D1, only top view)
D3	1, 2, 4	Top	No	4,500	925	50 (same as D1, only top view)
D4	1, 4	Top	Yes (flip + photometric)	4,000	1,280	34 gravel samples
D5	1, 2, 4	Top	Yes (flip + photometric)	4,000	1,280	34 gravel samples
D6	1, 2, 3, 4	Top	No	4,000	1,441	34 gravel samples

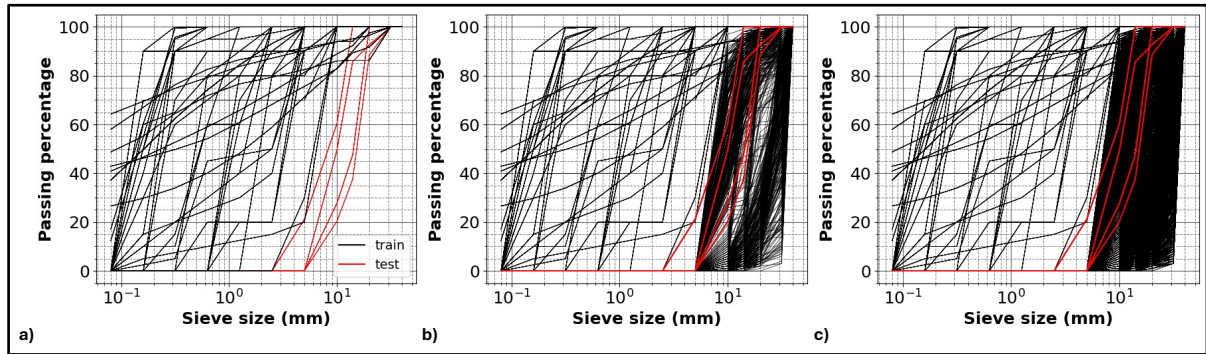


Figure 3.5 Particle size distributions (PSDs) for training and test datasets a) D4, b) D5, and c) D6. In each plot, the red curves represent the test sample PSDs, while the black curves correspond to the training and validation samples

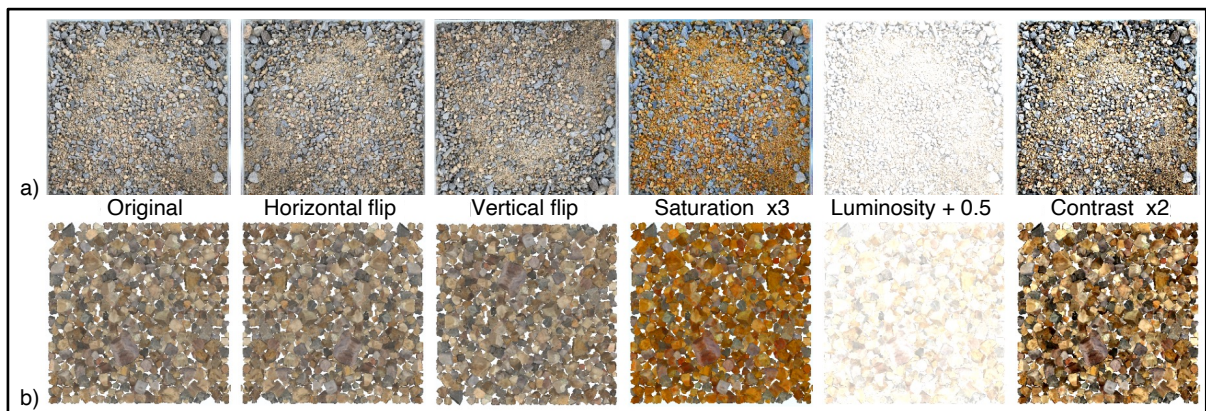


Figure 3.6 Transformations applied to the training and validation images used in datasets D4 and D5. Row a) shows examples of augmented images for artificial materials (type 1), while row b) presents transformations applied to synthetic images (type 2)

In addition to the 475 samples in dataset D4, 475 synthetic type 2 images chosen randomly were introduced to form dataset D5 (Figure 3.5b). It should be noted that the synthetic images have PSDs that are similar to the real images in the test dataset. The training set of 441 real and artificial soils and 475 synthetic soils was also increased to 5,280 images by the same data augmentation process. The objective of this dataset was to combine synthetic data with conventional augmentation applied methods.

Finally, dataset D6 retains the original 475 real images, but adds 5,000 synthetic images (2,500 type 2 images and 2,500 type 3 images) without applying any form of augmentation. The same 34 gravel samples are retained for testing (Figure 5c). The objective of this dataset was to replace conventional augmentation by synthetic data.

3.3.2 GrainSizeNet model

3.3.2.1 GrainSizeNet architecture

GrainSizeNet was trained on images of 299 by 299 pixels. GrainSizeNet is a residual neural network model that is trained in an end-to-end manner to predict the PSD from images. End-to-end learning involves training a model directly from raw data to final output, without any human intervention or intermediate preprocessing steps (Goodfellow et al., 2016). The model comprises 44 layers, with its overall architecture illustrated in Figure 3.7. Networks with residual architecture integrate skip connections between their intermediate convolutional layers. These skip connections enable the network to bypass multiple layers, facilitating direct information transfer to downstream layers. This approach, initially proposed by He et al. (2016) was used in this paper to create an architecture adapted to the complexity associated with texture patterns in images of granular materials. Given the high texture similarity among soil photographs, the use of residual blocks mitigates issues of degradation and vanishing gradient, enabling improved information propagation throughout the network.

The model that receives top and bottom images is trained using a tuple of input images (Figure 3.7-a). The feature maps that are produced by the convolutional layers are fed to fully connected layers paired with dropout layers to mitigate the risk of underfitting or overfitting. The model outputs a vector corresponding to the mass proportions passing the 0.080 mm sieve and the mass proportion retained on 11 control sieves from the BNQ 2501-025 (2013) standard (0.080, 0.160, 0.315, 0.63, 1.25, 2.5, 5, 10, 14, 20, 31.5 mm). A derivative of GrainSizeNet that accepts a single image as input enables the use of D2 to D6 with the same training parameters (Figure 3.7b).

Using the mass proportion retained on each sieve results in an output vector with a sum of one. To explicitly enforce this mass conservation constraint in the model output, a softmax activation function was applied to the final layer. For the sieve i output, the softmax function can be defined as follows:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (3.1)$$

where z_i and z_j represents the input of the softmax function for sieves i and j , respectively, and C is the total number of sieves.

GrainSizeNet comprises 1,315,308 trainable parameters for its architecture based on paired images (D1), and 678,748 parameters for the single image architecture (D2, D3, D4, D5, and D6). In all fully connected layers, a dropout layer is applied with a probability of 0.2, systematically deactivating outputs for random neurons.

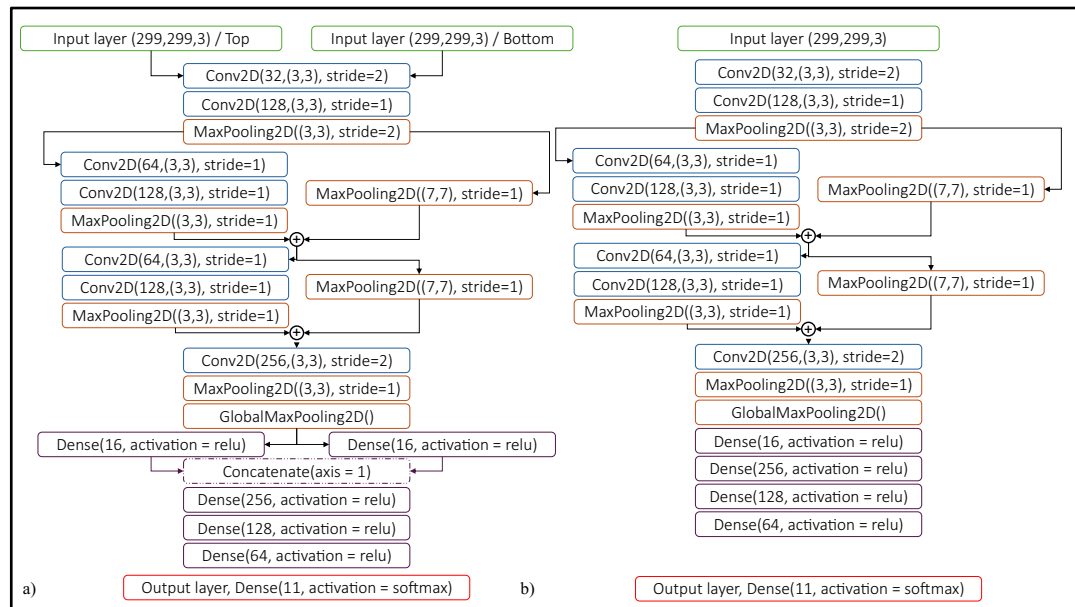


Figure 3.7 GrainSizeNet architecture a) with two input images (top and bottom) and b) with one input image (top)

3.3.2.2 Loss function and optimization

The model was trained using gradient backpropagation which applies successive partial derivatives to update the model parameters (weight and bias) with the objective of minimizing the loss function. The model is tasked with predicting the mass proportion for each particle size class relevant to the image, employing categorical cross entropy (CCE) as the loss function:

$$CCE = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_{c,n} \log \widehat{y}_{c,n} \quad (3.2)$$

where $y_{c,n}$ and $\widehat{y}_{c,n}$ designate respectively the real and model-predicted mass proportions retained on sieve c for image n , C is the number of sieves, and N is the number of images. Categorical cross-entropy is commonly used in multi-class classification tasks. It measures the difference in probability distribution between the model predictions and the ground truth. When determining the PSD, the model predicts continuous proportions of material across multiple particle size classes rather than discrete labels. Categorical cross-entropy (CCE) is used because it effectively trains the model to estimate the probability of belonging to each class, directly corresponding to the material proportions within each grain-size range. For regression tasks, the gradient backpropagation minimizes the loss function by adjusting weights and biases to emphasize the most prominent classes in the image, without exclusively maximizing a single class. Other hyperparameters used during model training are presented in Table 3.3.

Table 3.3 GrainSizeNet model configuration

Hyperparameter	Value
Learning rate	0.001
Optimizer	Adam
Loss	Categorical cross entropy
Batch size	32
Epochs	30

Subsequently, a structured data pipeline was established, integrating images with PSD. This pipeline was subdivided into several data subsets of 32 images each, to optimize model generalization by balancing computational efficiency and memory management.

During the initial training phase of GrainSizeNet, a conventional hold-out validation strategy was applied across all datasets uniformly on D1 to D3 and D4 to D6, using fixed training, validation, and test partitions as previously described. This baseline approach enabled consistent evaluation across datasets and served as a reference to compare the effects of more advanced training strategies.

To complement model evaluation based on the same test dataset and assess the model's ability to generalize, a cross-validation strategy was later implemented specifically on dataset D2. This technique provides a more robust estimate of performance by averaging results over multiple train-test splits. This procedure was applied to three distinct groups. The first group comprised 250 of the 294 real soil samples (type 1), the second consisted of 250 synthetic soil samples (types 2 and 3) and the third included 150 images of the 181 types 4 images. Each of these groups was divided into folds of 50 images for cross-validation: 5 folds for groups 1 and 2, then 3 folds for group 3. In all experiments, the dataset used for cross-validation contained the 5,475 samples equally divided into 4,500 training samples associated with 925 samples for internal validation and 50 test samples per fold for consistent evaluation between folds and dataset types. In summary, during the cross-validation procedure for dataset D2, the test set varies with each fold.

3.3.2.3 Performance metrics

To evaluate model performance, the mean absolute error (MAE) on the mass proportion retained on each sieve was selected as the evaluation metric:

$$MAE = \frac{1}{C} \sum_{c=1}^C |y_c - \hat{y}_c| \quad (3.3)$$

where C is the number of sieves, y_c and $\hat{y}_c$ represent the actual and the predicted mass proportion retained by the model for sieve c , respectively. The MAE is less sensitive to large deviations or outliers in predictions compared to metrics such as the RMSE. For comparison with GRAINet [25], the mean diameter weighted by the fraction of particles retained on each sieve size (D_m) was calculated for each fold according to the equation 4:

$$D_m = \sum_{c=1}^C y_c D_c \quad (3.4)$$

where D_c is the corresponding sieve size.

In addition to the MAE, the coefficient of determination R^2 was computed to assess how well the predicted distributions fit the measured data:

$$R^2 = 1 - \frac{\sum_{n=1}^N \sum_{c=1}^C (z_{c,n} - \hat{z}_{c,n})^2}{\sum_{n=1}^N \sum_{c=1}^C (z_{c,n} - \overline{z}_{c,n})^2} \quad (3.5)$$

$$\overline{z}_{c,n} = \frac{1}{N \cdot C} \sum_{n=1}^N \sum_{c=1}^C z_{c,n} \quad (3.6)$$

where $z_{c,n}$ and $\hat{z}_{c,n}$ designate respectively the real and model-predicted mass proportions passing on sieve c for image n . $\overline{z}_{c,n}$ represents the global mean of the observed passing percentages computed across all sieves and samples. An R^2 value close to 1 indicates that the model captures most of the variance in the measured PSDs. Using both MAE and R^2 allows evaluation of the model's average prediction error and the degree to which predicted distributions match the observed PSDs.

The predicted PSD allows the soil to be classified according to various classification systems, providing a concrete means to evaluate the practical impact of prediction errors beyond numerical metrics such as the MAE. In this case, the MTQ 1101 system was applied (Figure

3.8). This system is identical to the USCS system for particle sizes < 80 mm, except for the BNQ 2501-025 sieves. The PSD enables the calculation of sand and gravel proportions, as well as the uniformity coefficients (Cu) and curvature coefficients (Cc) when applicable for gradation assessment:

$$Cu = \frac{d_{60}}{d_{10}} \quad (3.7)$$

$$Cc = \frac{d_{30}^2}{d_{60}d_{10}} \quad (3.8)$$

where d_{60} , d_{30} , and d_{10} are the particle sizes corresponding respectively to cumulative percentages passing of 60%, 30%, and 10% from the particle size curve of the fraction passing the 80 mm sieve.

The marked imbalance between the numbers of different MTQ 1101 classes (Fines, SP, SM, GP, GW) in the data set combining images of types 1, 2, 3 and 4 (Table 3.4) with a notable predominance of the GP class. Consequently, the classification task was initially simplified to a primary ternary classification distinguishing fines, sand, and gravel, corresponding to the first logical step of the MTQ 1101 system. This reformulation enables a robust assessment of the models' ability to distinguish fines from granular materials, and subsequently differentiate gravel from sand, before extending the analysis to a finer multi-class classification. Performance on this task was evaluated via a confusion matrix, then a second matrix was constructed to evaluate classification errors on the selected MTQ 1101 classes (Fines, SP, SW, SM, GP, GW). Confusion matrices used as performance metrics include the fines class for D1 to D3 test dataset, whereas for D4 to D6, which contain only gravelly samples, the matrices were computed using only the classes represented in the test data.

Table 3.4 Number of samples per MTQ 1101 class (Fines, SP, SM, GP, GW)
for all real and synthetic samples

Image types	Fines	SP	SM	GP	GW
Type 1	0	240	0	54	0
Type 2	0	0	0	2495	5
Type 3	0	0	0	2495	5
Type 4	30	0	151	0	0

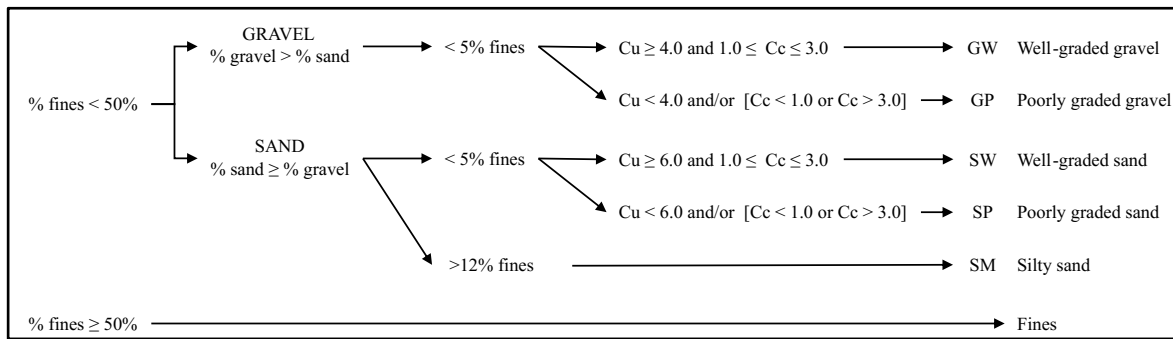


Figure 3.8 Detailed flow chart for the classification of coarse-grained soils with more than 50% retained on 0.080 mm sieve with the MTQ 1101 standard

3.4 Results and discussion

3.4.1 Hold-out validation

3.4.1.1 Performance comparison across model variants (D1, D2, D3)

The categorical cross-entropy loss function yielded values of 1.3395 for the model trained with the image tuple from dataset D1, 1.3273 for the D2 and 1.2764 for dataset D3 on the test dataset. The MAE on the mass proportion retained for the test dataset reached 3.7 %, 3.5 % and 2.9 % for the D1, D2 and D3 respectively. Figure 3.9 shows the training progression of the different models after 30 epochs.

Although the bottom view provides additional information to the neural network trained with D1, its performance remained comparable to models trained exclusively on top-view images (D2 and D3). A plausible explanation for these results is that the images captured from the bottom were ultimately similar due to the systematic deposition of fine particles at the base of the sample irrespective of the PSD. In addition, occasional artifacts such as reflections observed on most of the real images probably reduced their usefulness (soil types 1 and 4).

These results indicate that top-view images alone provide sufficient textural information to capture PSD characteristics, which has practical implications by simplifying acquisition protocols and reducing imaging requirements.

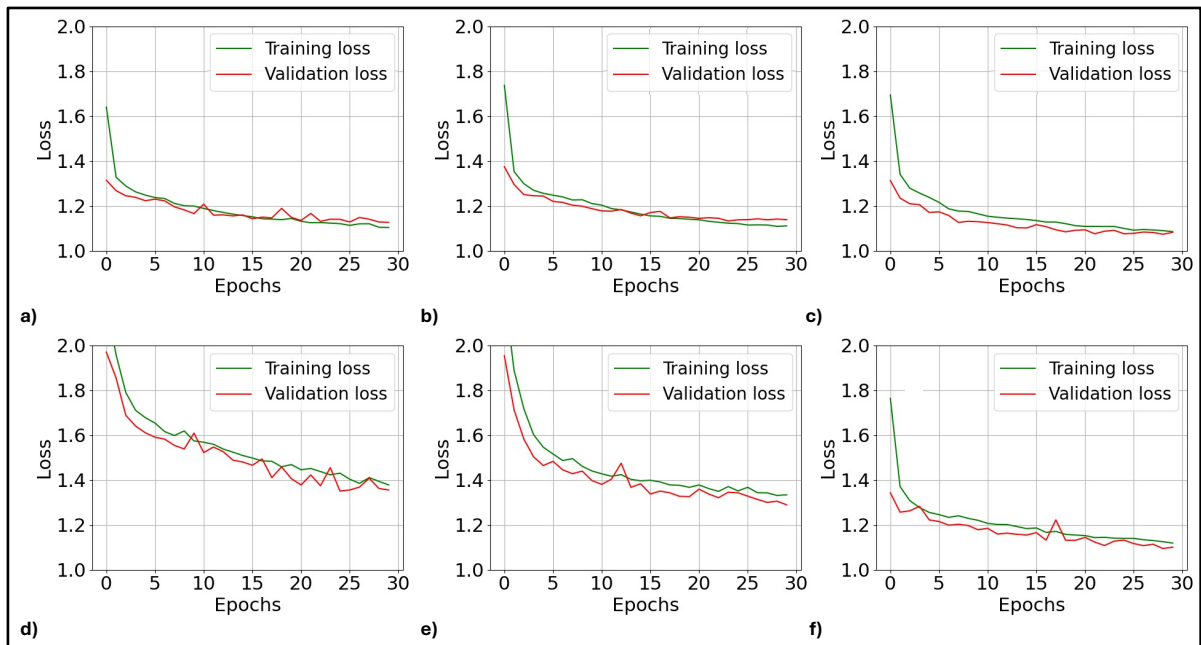


Figure 3.9 Learning curves for training on datasets a) D1, b) D2, c) D3, d) D4, e) D5, and f) D6

3.4.1.2 MTQ 1101 classification performance

While the categorical cross-entropy loss and MAE provide useful indicators of overall predictive accuracy on continuous PSD values, they may not fully reflect the model's effectiveness in practical classification contexts. As a first step, the model's ability to distinguish the primary granulometric classes was assessed using a ternary classification (fines, sand, and gravel).

The confusion matrices associated with the models trained on datasets D1, D2, and D3 (Figure 3.10) reveal an overall classification accuracy defined as the ratio between the sum of diagonal elements and the total number of predictions of 96% for all three cases. Classification errors were limited to two samples belonging to the fine category, which had been classified as sand. These samples correspond to borderline grain sizes containing both fine and sandy fractions, while all sand and gravel samples were correctly classified.

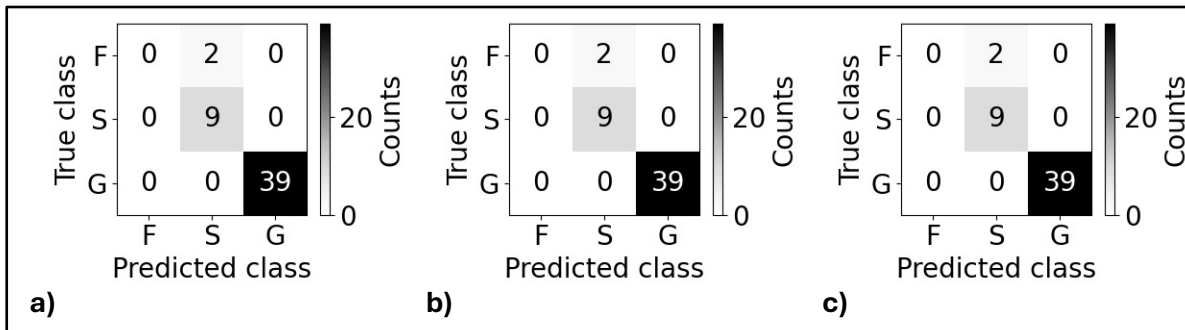


Figure 3.10 Confusion matrices for classification using GrainSizeNet trained on a) D1, b) D2, and c) D3, evaluated on the same test dataset

A more detailed evaluation was then carried out using a multi-class classification based on MTQ classes (Fines, SW, SP, SW, SM, GP, and GW (Figure 3.8). The resulting overall classification accuracies were 94% for D1, 96% for D2 and 92% for D3 (Figure 3.11). Notably, model D2 achieved the highest classification accuracy despite presenting a slightly higher MAE than model D3. This result highlights that classification reliability does not necessarily correlate directly with PSD prediction errors.

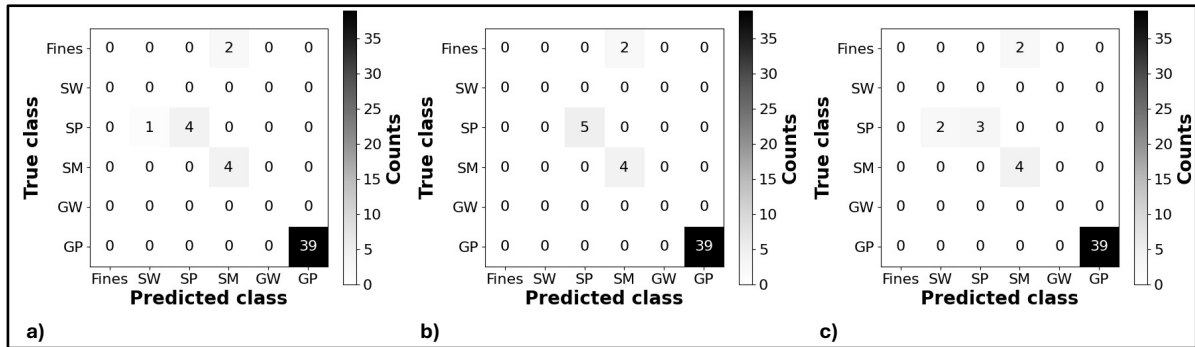


Figure 3.11 Confusion matrices for classification using GrainSizeNet trained on D1 (a), D2 (b), and D3 (c), evaluated on the same test dataset

From a geotechnical engineering perspective, accurate soil classification is often a more relevant performance indicator than individual PSD prediction errors. The results indicate that GrainSizeNet can accurately classify the samples using photographs of real samples, as well as segmented and non-segmented synthetic samples, although a minor discrepancy in the classification of the real samples was observed. The slightly lower classification performance observed for real soils is likely related to their limited representation within the training dataset.

Overall, these findings suggest that GrainSizeNet model can achieve soil classification results comparable to traditional sieve analysis, supporting their potential use in practical geotechnical characterization workflows.

3.4.1.3 Comparison of predicted vs. ground truth grain size distributions

Figure 3.12 shows representative test samples evaluated by GrainSizeNet after training on datasets D1, D2 and D3, respectively. This Figure includes a selection of real, artificial, synthetic and segmented soil samples, illustrating the diversity of the test dataset. For each example, the top-view image is displayed along with the predicted particle size and ground truth distribution curves, enabling a direct visual comparison of model performance across the three training scenarios.

Although only the top-view image is shown for illustrative purposes, it is important to remember that for the models trained on D1, both the top-view and bottom-view images were provided as input data during inference. Despite these differences in input modalities, the predicted particle size distribution curves obtained from the different training configurations show strong agreement with the corresponding ground truth curves, with slight variations observable between datasets.

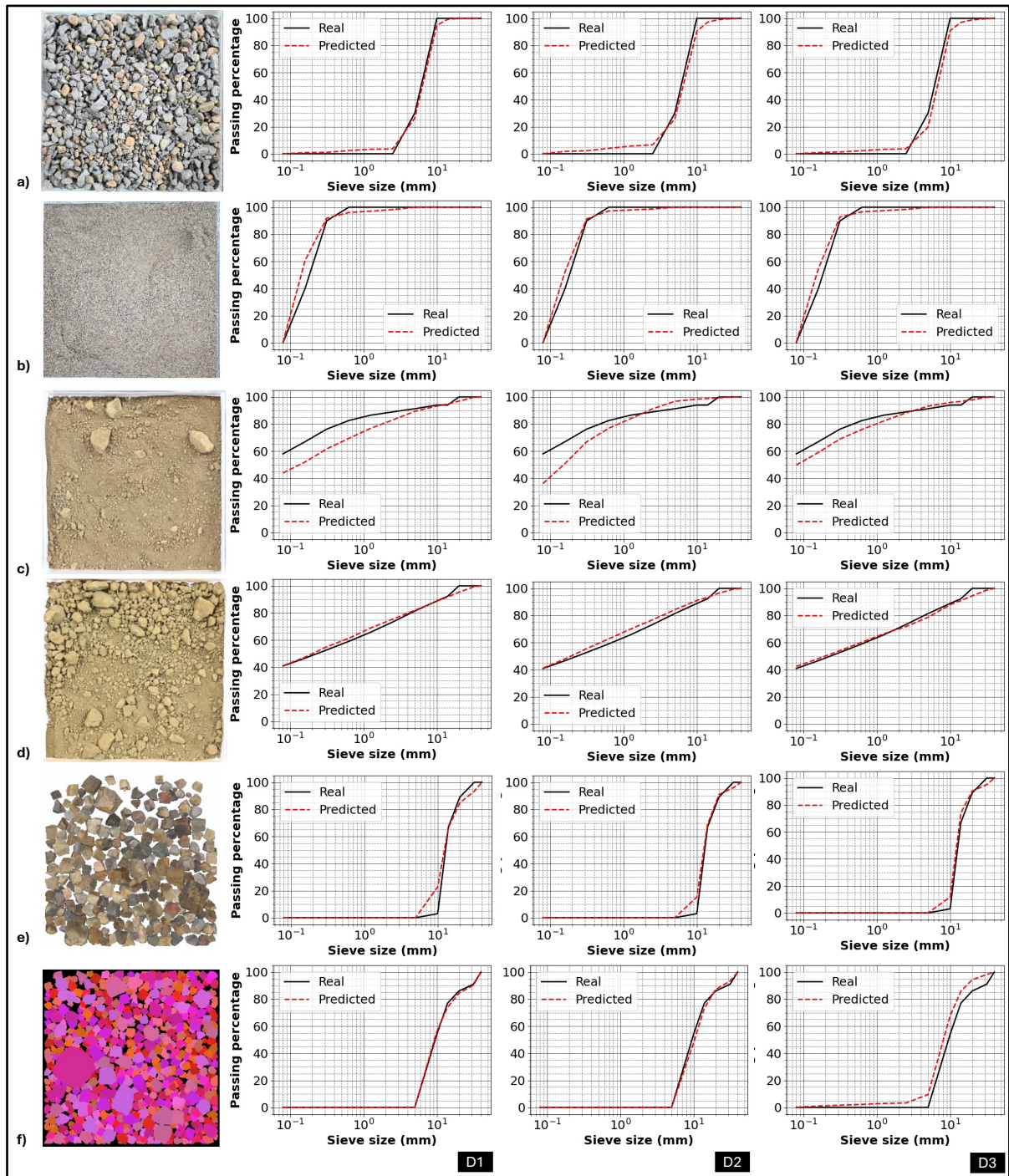


Figure 3.12 Representative test samples and predicted vs. ground truth granulometric curves for GrainSizeNet trained on D1, D2, and D3 datasets. SubFigures (a)–(d) correspond to real samples (types 1 and 2), while (e) and (f) represent synthetic samples (types 3 and 4). For each sample, the top-view image is shown alongside the predicted and reference granulometric curves. Note that for models trained on D1, both top- and bottom-view images

were used during inference, even though only the top-view image is displayed here for visualization

Figure 3.13 compares the predicted and actual cumulative passing percentages for the three data sets. Despite the high classification accuracy, some outliers can be observed. The models achieved high linear correlation with the ground truth, as reflected by the coefficients of determination: $R^2 = 0.967$ for D1, $R^2 = 0.972$ for D2, and $R^2 = 0.978$ for D3. Some outliers are visible in all cases. This reflects that the bottom images and segmented images did not lead to any noticeable improvement in regression performance. Overall, the results are very similar across all the datasets.

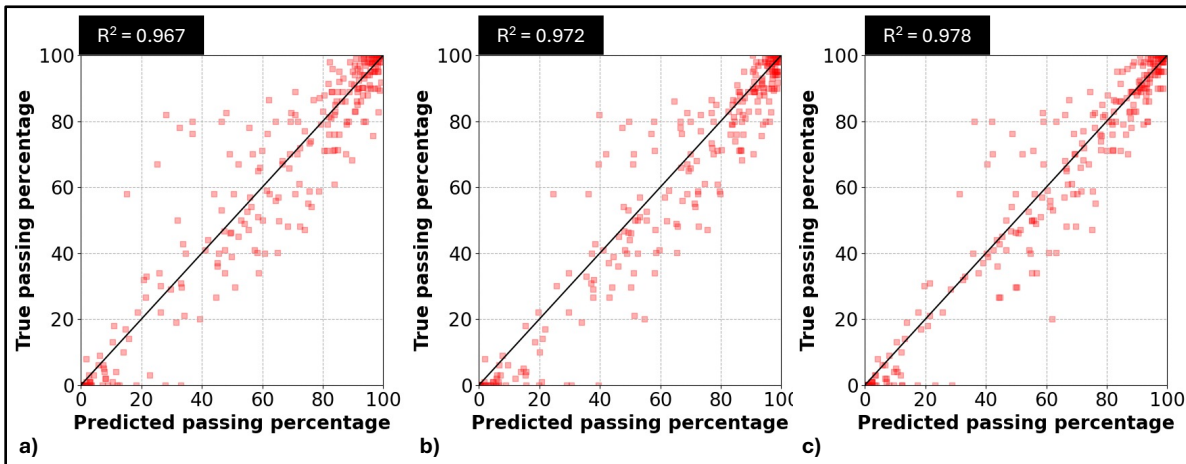


Figure 3.13 Comparison of the actual and predicted cumulative percentages passing for all sieves of the test samples using GrainSizeNet trained on a) D1, b) D2, and c) D3

3.4.1.4 Influence of traditional data augmentation and synthetic data augmentation

The approach adopted in this study also departs from traditional data augmentation techniques such as flipping or rotation, which do not introduce new particle size diversity. Instead, the dataset was expanded by incorporating additional synthetic samples characterized by distinct particle size distributions. This strategy aimed to increase the representation of PSD variability in the training dataset.

The model trained on D4 based solely on augmented real data performed the worst, with a loss in categorical cross-entropy of 2.7916 and a MAE of 11.5% on the test set. The targeted addition of synthetic data in D5 results in a significant improvement, with a loss reduced to 1.7441 and an MAE of 6.2%. Finally, the extensive integration of synthetic data in D6, without any form of augmentation, leads to the best loss performance, achieving a loss of 1.626 and a MAE of 6.4%.

These results, obtained on the same test set composed exclusively of gravelly samples, underline the effectiveness of synthetic data, particularly when produced in large quantities, in improving model generalization on particle size classes under-represented in real data (Figure 3.14).

In addition to the quantitative evaluation based on MAE, a ternary classification analysis was carried out, as illustrated in Figure 3.15 (a, b and c). When considering this distinction between correctly classified and misclassified samples, the model trained solely on real data with augmentation (D4) achieved an accuracy classification of 5.9%. In contrast, both models incorporating synthetic data (D5 and D6) correctly classified all 34 samples, demonstrating 100% accuracy under this simplified scheme.

A more detailed MTQ 1101 classification analysis revealed differences between synthetic data strategies. The confusion matrices for D5 and D6, presented in Figure 3.14d and 3.14e, show differences in class assignment. The model trained on D5 achieved a classification accuracy of 100%, with all 34 samples assigned to the GP class. The model trained on D6 achieved a classification accuracy of 91.2%, with 31 samples classified as GP, one sample as GW, and two samples labeled as Unclassified (any sample falling outside these classes labeled).

These values demonstrate that increasing synthetic data significantly improves regression and classification performance but an unbalanced increase in synthetic data can lead to performance losses in both case. This aligns with conclusions reported by Nowruzi et al.

(2019), where the balance between real and synthetic datasets is shown to be critical for maintaining robust and reliable classification behavior.

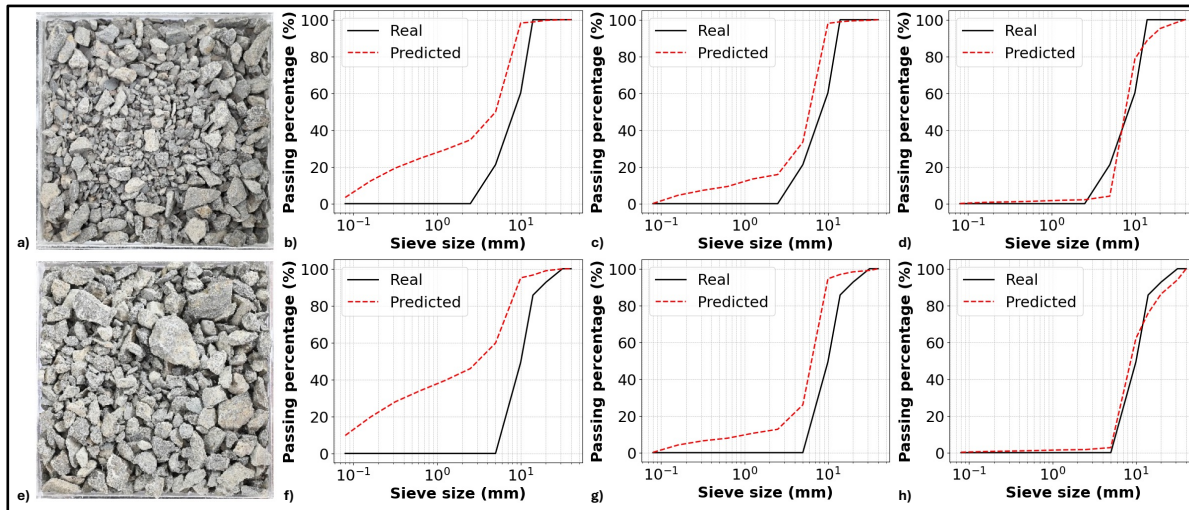


Figure 3.14 (a, e) Example test-sample images with actual and predicted particle size curves using GrainSizeNet trained on (b, f) D4, (c, g) D5, and (d, h) D6

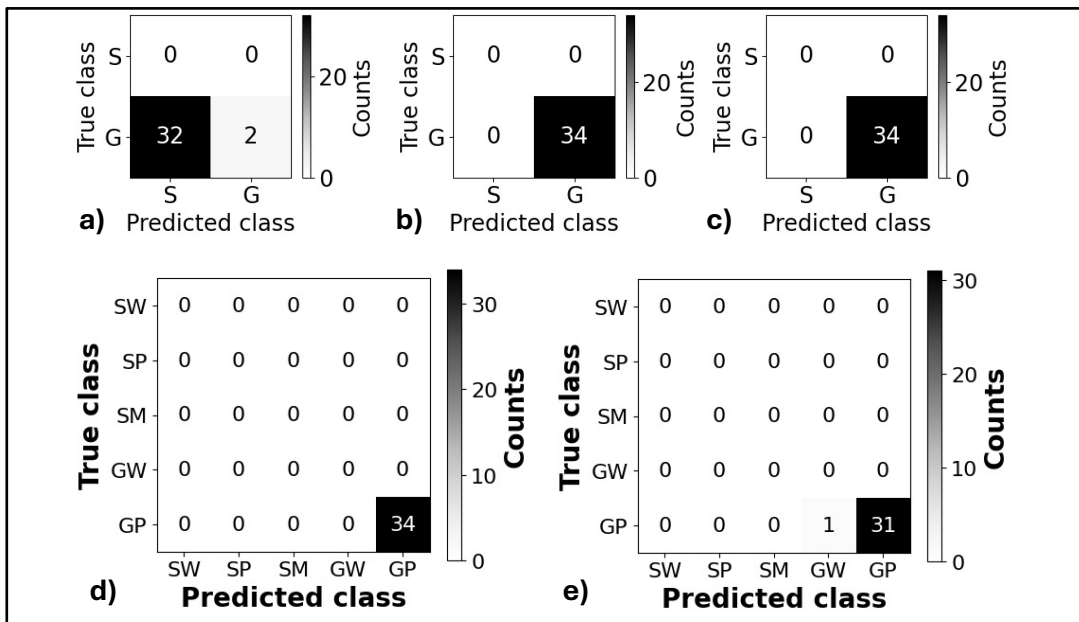


Figure 3.15 Confusion matrices for classification using GrainSizeNet trained on D4 (a), D5 (b), and D6 (c), evaluated on the same test dataset, (d) and (e) present the extended confusion matrices for the models trained on D5 and D6, respectively, including all classes

3.4.1.5 Assessing the effectiveness of GrainSizeNet in light of previous studies

CNNs usually have an architecture based on an input image to which you assign an output target value. This is done on the assumption that the image provides sufficient information for the problem at hand. Pirnia et al. (2018) also assumed that the combined images of the bottom and top could better represent the particle size. A large dataset consisting only of synthetic samples (53,130 images with PSD) was generated using discrete element method (YADE) to train a CNN to predict the percentages passing in mass over 5 sieves. Their model achieved a root mean square error (RMSE) of 6.94%.

In comparison, the present study uses a dataset of 10,950 images, almost five times smaller, but more heterogeneous as it includes both real and synthetic samples. In addition, GrainSizeNet predicts more than 11 sieve sizes plus passing 0.080 mm, a relatively more complex task. Despite these constraints and the difference in size of the data sets, models trained on datasets D1, D2 and D3 achieved RMSEs of 8.40%, 8.02% and 7.09%, respectively. Although slightly higher than the results of Pirnia et al. (2018) these values remain of the same order of magnitude, thereby reinforcing the effectiveness of GrainSizeNet model.

In parallel, Manashti (2022) developed PSDNet, a convolutional neural network trained on synthetic and real soil images with varying resolutions and color modalities. a moderate performance improvement in RMSE was observed when using combined top and bottom images compared to top images alone, with values decreasing from approximately 5.8% to 3.1% for color images at 160 pixels resolution. This difference, while notable, remains relatively modest and comparable to the small performance gain observed in the present study between datasets incorporating both views (D1) and those using only top views (D2 and D3). These results suggest that although incorporating multiple views can enhance granulometric prediction accuracy, the benefit may be limited depending on dataset characteristics and task complexity.

These comparisons indicate that while multi-view imaging can improve PSD prediction accuracy, its benefit may depend strongly on dataset characteristics and prediction complexity. The results therefore support the robustness and competitiveness of GrainSizeNet relative to existing approaches.

3.4.2 Cross-validation performance

The model's performance on synthetic soils (types 2 and 3) reveals a MAE over the entire granulometric distribution ranging from 2.32% to 3.44%, i.e. an average MAE of 2.75% over the five folds. The results obtained for artificial soils (type 1) range from 5.96% to 11.58%, for an average MAE of 8.02% and real soils (type 4) range from 4.59% to 6.38%, for an average MAE of 5.27% (Figure 3.16). Considering both type 1 and type 4 as real soil images, the overall average MAE on real soils reaches 6.96%.

Overall, the cross-validation protocol yields higher MAE values compared with the classic random holdout (mixing all 5,475 images and testing on 50), which reported an average MAE below 4%. This indicates that the method provides a realistic assessment of model performance and identifies specific samples or granulometric classes where predictions are more challenging. In this study, difficulties are observed primarily on type 4 soils, the most under-represented class, where variations in water content can produce features that resemble other granulometric classes or aggregation patterns.

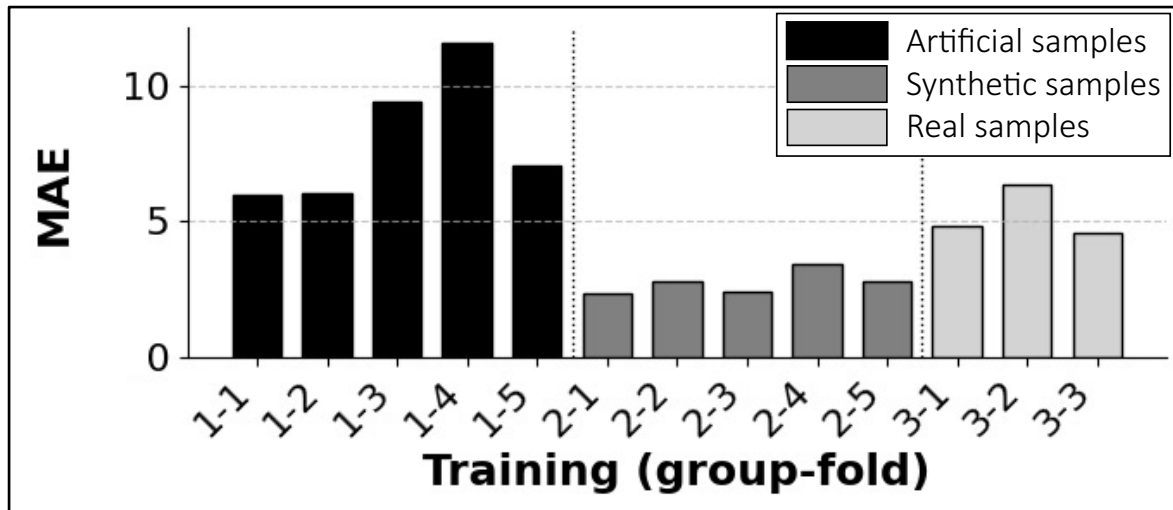


Figure 3.16 Evolution of mean absolute error by group and training folds

To enable a relevant comparison with the GRAINet model (Lang et al., 2021), the mean absolute error on the mean grain diameter was calculated for the different real soil fold groups, focusing specifically on Type 1 samples (Table 3.5). This choice is motivated by the fact that Type 1 includes both coarse sand and gravel, making it more comparable to the gravel bars analyzed in GRAINet, whereas Type 4 consists of sand and fines class. Among these folds, fold 1 contained the highest number of gravel samples (34 out of 50), while only fold 3 included any additional gravel (10 samples). The remaining folds consisted primarily of coarse to fine sands, which are typically more prone to introducing error due to their visual homogeneity and finer textures.

Focusing on fold 1, which is most representative of the coarse-grained material analyzed by GRAINet, the GrainSizeNet model achieves a MAE of 1.05 mm for a mean D_m of 6.25 mm, corresponding to a relative error (RE) of 16.8% defined as the ratio between MAE and D_m . This is notably lower than the 18% relative error reported for GRAINet, which had a MAE of 11 mm on gravel bars with an average D_m of 62 mm. These results support the conclusion that GrainSizeNet offers more accurate D_m predictions even in samples containing fine materials and confirm that the method remains robust when applied to coarser sediment types similar to those used in the GRAINet study.

Table 3.5 Mean absolute error on mean grain diameter (d_m) for GrainSizeNet – type 1 images

type 1	MAE_ d_m (mm)	Mean d_m	RE (%)
Fold 1	1.05	6.25	16.80
Fold 2	0.13	0.41	31.70
Fold 3	0.56	1.88	29.28
Fold 4	0.22	1.01	21.78
Fold 5	0.30	1.12	26.79

The MTQ 1101 soil classification algorithm was used to systematically evaluate k-folds during cross-validation on real and synthetic data (Figure 3.17). For soil types 1, 2, and 3, GrainSizeNet achieved a consistent classification accuracy of 100% across all folds, indicating a robust and stable prediction capability for these soil types. In contrast, a higher variability was observed for type 4 soils, with classification accuracies of 64%, 90%, and 100% for the three folds, respectively.

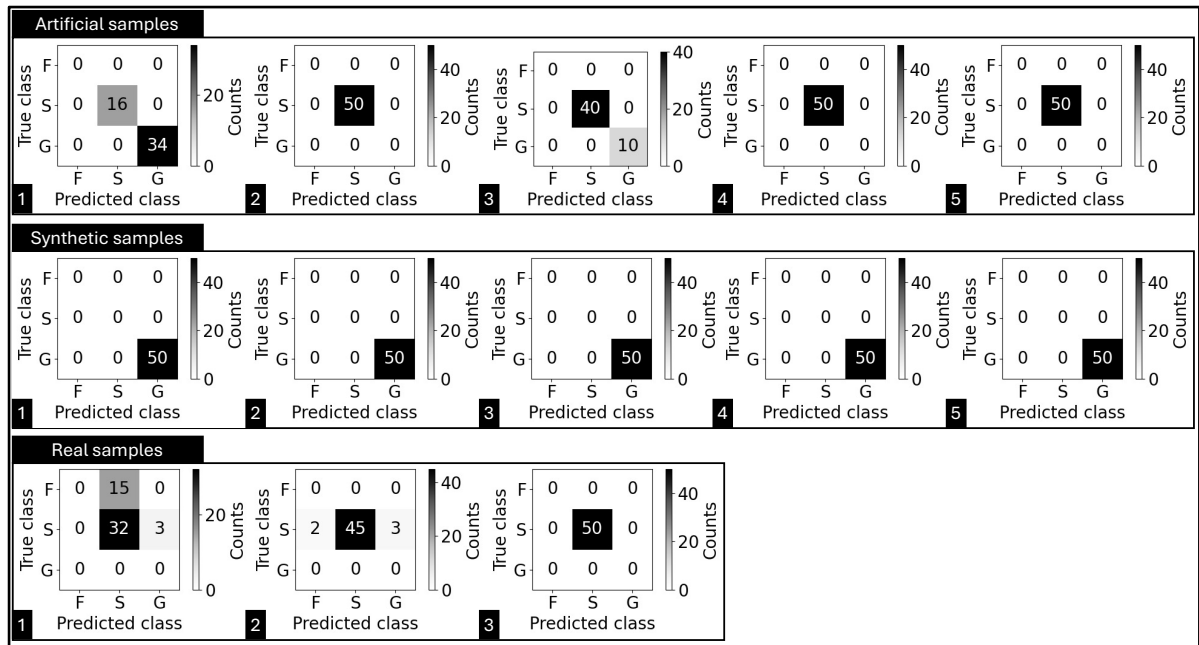


Figure 3.17 Confusion matrices illustrating the classification performance of GrainSizeNet with respect to MTQ 1101 classes, evaluated across the different test folds used during cross-validation

A detailed examination of the misclassified samples indicates that these specimens correspond to fine soils with fines contents ranging between 55% and 65%, while the remaining fraction is predominantly sandy. Although classified as fine soils according to the MTQ 1101 criteria, these samples retain a significant sandy character, which may explain their systematic misclassification as sand-dominated soils by the model. This observation suggests that GrainSizeNet may be more sensitive to dominant coarse-grained textural features when the fines content lies close to the classification threshold. However, the image dataset corresponding to type 1 soils does not include samples with higher fines contents, which limits the ability to further evaluate or confirm this hypothesis.

Manashti (2022) reported that cross-validation strategies involving the exclusion of specific materials can lead to a decline in predictive performance when the distributions of the training data differ from those of the test fold. In this context, the misclassification of marginal fine soils reflects a realistic generalization challenge rather than an isolated model failure. More

broadly, these results reinforce the importance of adequately covering the particle size distribution space during training.

3.5 Conclusions

In this paper, we present a new approach to particle size analysis based on deep learning techniques. The proposed model, GrainSizeNet, was developed from a dataset of real and synthetic images, totaling 5,475 samples, each associated with a particle size distribution. This convolutional neural network model, based on a residual architecture, demonstrated significant accuracy in terms of soil classification according to the MTQ 1101, by comparing the classification from the actual versus predicted particle sizes.

The results demonstrate that the integration of synthetic data improve model generalization, particularly in under-represented granulometric ranges. The model trained on augmented real data alone (D4) exhibited the lowest performance, while the progressive integration of synthetic samples in D5 and D6 yielded significantly better results reducing both the categorical cross-entropy loss and the MAE, and achieving good accuracy in soil classification on gravel-rich test sets. These findings confirm the key role of targeted synthetic data in bridging the gaps left by real datasets and in addressing granulometric class imbalance.

Additionally, the dual-view (top/bottom) configuration was tested to assess its contribution. Results showed that while the bottom view provides additional information, it does not significantly improve the model's performance compared to using the top view alone. This is likely due to the lack of visual diversity and systematic accumulation of fine particles in the lower layers of real samples, as well as occasional visual artifacts in the bottom images. The top view alone proved sufficient to capture the essential textural cues required for accurate particle size estimation.

Nevertheless, this study is a first step towards an innovative method of particle size determination of soils by photographic analysis nested with CNNs. Beyond PSD estimation,

the ability of convolutional neural networks to extract textural information from soil images suggests broader potential applications in geotechnical characterization. However, such extensions require dedicated investigations and validation using targeted datasets.

Transfer learning is a concept that is becoming more prevalent in the machine learning community. This offers the possibility of improving the prediction capacity and especially the generalization of models by exploiting a pre-trained model on a similar dataset (Yosinski et al., 2014). Using GrainSizeNet as a basic model could therefore prove relevant to carry out grain size analyses directly on-site photographs, offering substantial time savings in geotechnical or geoenvironmental investigation projects.

CHAPITRE 4

CONVOLUTIONAL NEURAL NETWORKS FOR SOIL CLASSIFICATION BASED ON PHOTOGRAPHS OF SPLIT-SPOON SAMPLER

Harold Simo Tenekam ^a, François Duhaime ^b, Jean Sébastien Dubé ^c, Matthew F. Toews ^d,
Simon Grenier ^e

^{a, b, c} Département de Génie de la construction, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

^d Département de Génie des systèmes, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

^e Département de science de la terre et en environnement
1600, boulevard René-Lévesque Ouest, Montréal, QC, Canada

Article soumis pour publication, janvier 2026

4.1 ABSTRACT

This paper presents a series of convolutional neural networks designed to analyze photographs of split-spoon samples from SPT field test in boreholes. GrainSizeNet, a network previously trained to predict the particle size distribution (PSD) of laboratory samples from images, was adapted and fine-tuned to classify materials into five categories of the United Soil Classification System (USCS) and to estimate the mass proportion of fine particles ($<75\ \mu\text{m}$) with regressions. The training datasets were composed of 1,206 split-spoon soil sample photographs paired with corresponding PSD curves and Atterberg limits for classification and 786 photographs with corresponding mass proportions of fine particles for regression. Two strategies were compared for both tasks: single-step models and hierarchical approaches combining specialized classifiers and regression models. The hierarchical models performed best for both tasks when tested with unseen images with 69 % of the samples correctly classified and a mean absolute error of 0.153 on the proportion of fine particles. These results demonstrate the potential of using image analysis for automated soil classification in the field pending a larger dataset.

Keywords: Convolutional neural network, Split-spoon, Transfer learning, United Soil Classification System.

4.2 Introduction

Soil sampling is a fundamental aspect of geotechnical investigation (AASHTO, 2022; Hunt, 2005; Parker, 1996). Different types of samplers can be used in boreholes. The split-spoon sampler is commonly used for obtaining remolded soil samples and measure the number of blows (N value in terms of blows per 300 mm) to drive the sampler into the soil. It is opened in the field to record information about soil texture, water content, and contamination on borehole logs. The split-spoon samples are also systematically photographed.

Laboratory tests can also be conducted on split-spoon samples to obtain the soil classification. Classification is most often performed using the Unified Soil Classification System (USCS) (ASTM International, 2020), which relies on the particle size distribution (PSD) for coarse soils (sands and gravels) and Atterberg limits for fine soils (silts and clays). The PSD is obtained based on standard methods such as ASTM D6913 (ASTM International, 2021). This method requires at least two cycles of overnight drying to obtain the dry mass before and after washing the soil on a 75 μm sieve. This implies that a single test requires few days. Although relatively inexpensive for a single sample, particle size analyses can be costly because of the large number of tests that are typically involved in a given project.

Soil classification is directly related to soil appearance (Holtz et al., 1981). Experienced field technicians are able to classify split-spoon samples in the field based on visual inspection. It is hypothesized that artificial intelligence can be used to classify soils based on the texture of photographs of split-spoon samples acquired in the field, thereby bridging field observations and laboratory-based soil classification within a single workflow.

The extraction of soil texture from image can be done either manually by using conventional textural descriptors, such as conditional entropy (Duhaime et al., 2021; C. Shannon & Weaver, 1949) and Haralick features (Haralick, 1979), or automatically using convolutional neural networks (CNNs) (LeCun et al. 1998). Convolutional layers, the main building block of CNN, act as learned filter banks that capture relevant textural patterns in an image. Convolutional

layers are typically followed by nonlinear activation functions and pooling layers. The activation functions introduce nonlinearity by retaining positive activations while pooling layers retain the dominant features and reduce spatial resolution. The resulting feature maps are then passed to one or more fully connected network for classification.

Buscombe (2020) was the first to train a CNN to predict the PSD of granular materials. SediNet was trained to predict nine PSD percentiles from 409 sandy sediment images. Despite the small size of the training dataset, it achieved normalized root mean square errors (RMSE) on PSD percentiles ranging from 16 to 29%.

Several authors have trained CNNs to predict the PSD of granular materials from laboratory photographs. PSDnet, a CNN trained with both real soil photographs and synthetic images created with the discrete element code YADE (Pirnia et al., 2018; Smilauer et al., 2021), predicted percentages passing for nine sieve sizes with a RMSE of 4% (Duhaime et al., 2021; Manashti, 2022). Simo Tenekam et al. (2025) trained GrainSizeNet on more than 10,000 soil images to predict the mass percentages retained on 11 control sieves with sizes ranging from 0.080 to 31.5 mm, including the fraction passing the 0.080 mm sieve (BNQ 2501-025, 2013) with a mean absolute error of 2.9 %. These results are encouraging, but they were obtained with laboratory photographs taken under controlled conditions. Photographs of split-spoon samples taken in the field are expected to be noisier with variable image quality and lighting conditions.

Lang et al., (2021) presented GRAINet, the first CNN trained specifically with field data. Their database comprised aerial photographs of gravel banks with mean diameters ranging approximately from 1.3 cm to 29.3 cm, corresponding to coarse gravel material. Their CNN was trained to predict different PSD percentiles. It was able to predict the mean diameter with a mean absolute error of 1.1 cm. The real PSD were obtained based on image segmentation, i.e. particle contours, not from the standard tests used for soil classification in geotechnical engineering. The sediments photographed by Lang et al. (2021) were relatively clean gravels,

a material that differs significantly from the fine-grained soils often encountered in geotechnical studies.

Transfer learning allows existing networks, such as GrainSizeNet, to be adapted and retrained for different tasks (Yosinski et al., 2014). Pre-trained networks are typically used to extract features from images. These features are then fed into a smaller network trained with fewer parameters to accomplish the new task. Transfer learning is particularly beneficial when limited data are available to train the network for the new task. Several authors have used transfer learning to predict the particle distribution of granular materials based on small training datasets (J. Olivier & Aldrich, 2021; Wei et al., 2024; Xiao et al., 2022). For example, an adaptation of GoogleNet (Szegedy et al., 2015) was proposed by Olivier & Aldrich (2021) to predict the particle size from hydrocyclone flow images. In general, transfer learning yields better performance when the new model task is closely related to the original task (Yosinski et al., 2014).

Soil classification can be inferred from predicted PSD and Atterberg limits, or it can be obtained directly through a classification network. Most CNN use a flat or single-step classification, where all final nodes in the decision tree are treated the same. For instance, for soil classification, the final layer of the network would have to choose between all possible USCS symbols. Other approaches can consider the hierarchical nature of classification systems. For instance, for soil classification, a first network would determine if the soil is fine- or coarse-grained. For coarse-grained soils, a second network would determine if the soil is sand or gravel, and so on. Hierarchical approaches that involve local classification usually perform better than single-step (or flat) classification (Silla and Freitas 2011). Hierarchical CNN strategies have been proposed in computer vision to progressively refine class boundaries and reduce inter-class confusion (e.g., Xiao et al., 2014).

This article proposes an adaptation of GrainSizeNet for automatic USCS classification and estimation of the proportions of fine particles based on photographs of split-spoon samples. This study explores the potential of CNNs to automatically extract textural information from

photographs taken in the field, even though these images have variable resolution and are often affected by environmental conditions (dust, humidity, wind, variable lighting conditions). To improve the robustness of the model, a hierarchical strategy is implemented to break down the overall classification into specialized subtasks. This structure aims to reduce confusion between visually similar soil types and improve the accuracy of grain size estimation, particularly in complex or ambiguous cases. To the authors best knowledge, this paper is the first to introduce a split-spoon photograph dataset and to leverage CNNs for automatic soil classification in geotechnical engineering.

4.3 Materials and methods

4.3.1 SPT dataset and image preprocessing

The split-spoon sample photographs used in this study were obtained during various field investigations conducted by WSP across Canada. The full dataset includes 6,103 split-spoon photographs, each accompanied by the tests results required to obtain their USCS classification (sieving and/or Atterberg limits). The samples in the full dataset were obtained with B-, N-, H- and P-caliber split-spoons. Sieving was conducted based on different standards depending on project requirements (ASTM International, 2015; BNQ 2501-025, 2013; LC 21-040, 2024). These standards use slightly different sieve sizes. For example, the fine-particle content corresponds to the material passing the 75 μm sieve under ASTM, and the 80 μm sieve under BNQ/LC. This difference was considered negligible for the purpose of this study. Figure 4.1 shows the overall distribution of the USCS classification results. As can be noticed, sample classifications are not uniformly distributed. Silty sands (SM) associated with glacial deposits are over-represented in the database.

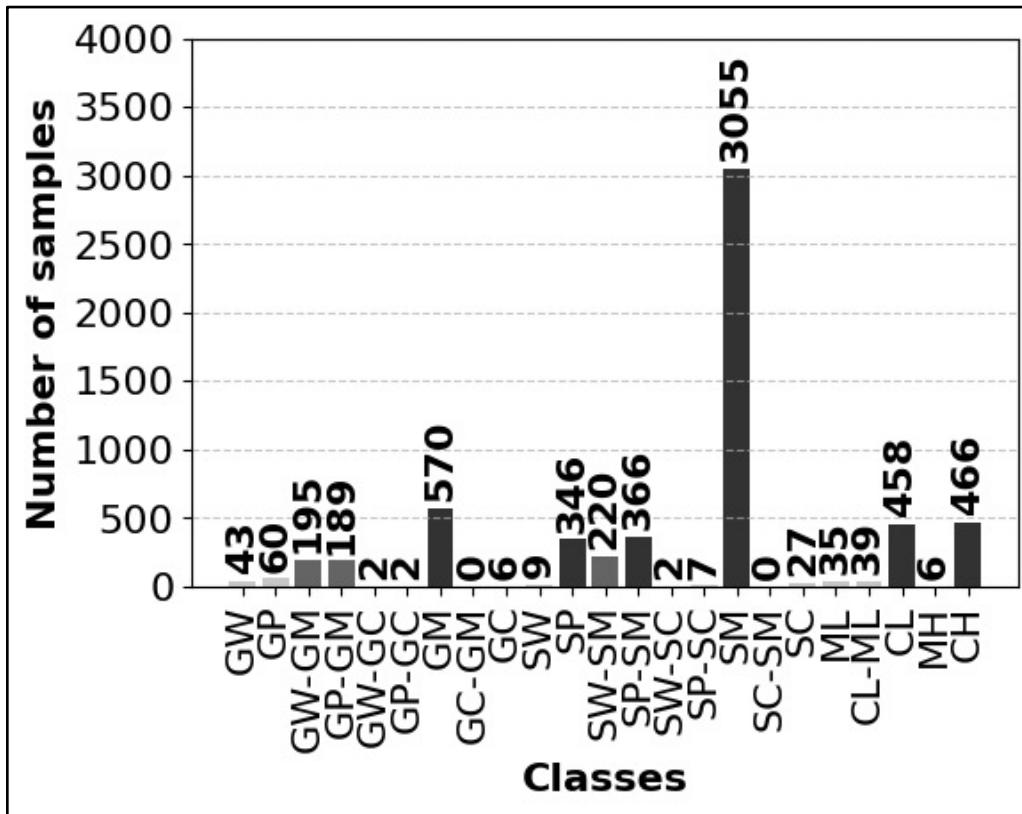


Figure 4.1 USCS class distribution for the full dataset

Only photographs of B-caliber split-spoons were utilized to maintain a constant image scale. The split-spoon photographs were subsampled and resized. An example is shown in Figure 4.2. For each image, up to three square subsamples were selected based on the particle size analysis carried out in the laboratory and based on sample appearance. In the case of heterogenous split-spoon samples, the square areas were selected based on our judgment. The subsampling excludes the background, which contains unnecessary information for the CNN. Using square subsamples with sides corresponding to the inside diameter of the B-caliber sampler (38 mm) allows the split-spoon to be used as a scale reference. The images were resized to 299×299 pixels after cropping. This allows all images to have the same resolution (8 pixels per mm), independently of the device used to take the photograph and the distance between the device and the sample.



Figure 4.2 Example of split-spoon photograph subsampling and resizing with a) original image, and b) and c) selected subsamples cropped and resized to 299×299 pixels

Data augmentation techniques were implemented to densify some datasets. The aim was to enrich the data without introducing transformations that might alter the interpretation of image texture. The transformations included horizontal and vertical flipping, 90° and 180° rotations, brightness and contrast adjustment, color enhancement, Gaussian blur, and a hue shift (shifting the hue channel by 30° in HSV space while keeping saturation and brightness unchanged). Figure 4.3 illustrates the various data enhancement steps that can be applied to a specific image.

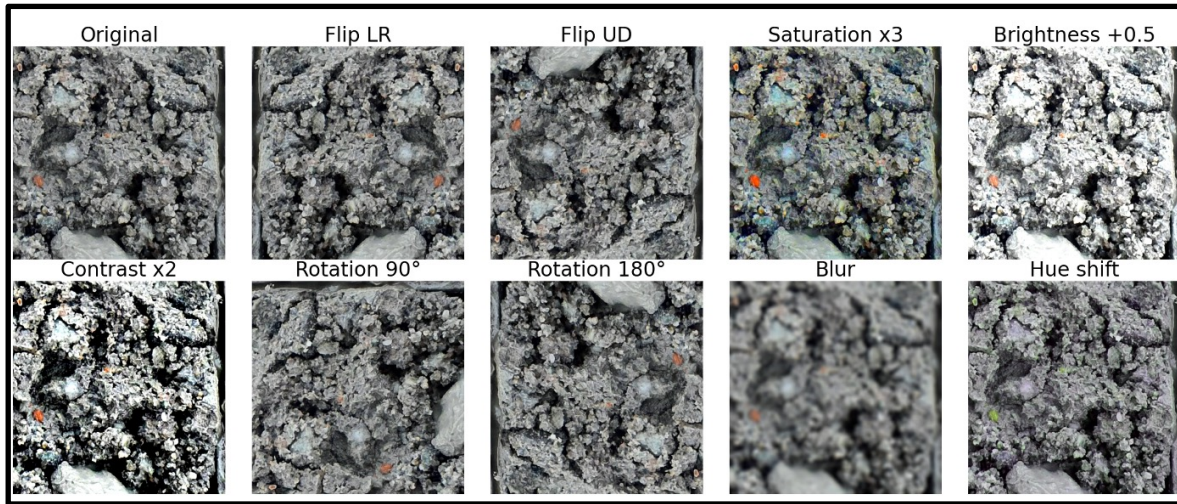


Figure 4.3 Example of the different augmentation steps applied to a training image. The original images were modified by horizontal and vertical flipping (left-right and up-down), color, brightness and contrast enhancement, rotation, Gaussian blur and hue shift to enrich the diversity of the dataset

4.3.2 Classification datasets

Based on image quality and USCS classification, 1,206 images taken between 2022 and 2024 were selected to construct the training and validation datasets for the USCS classification. 200 images acquired in 2025 were used to prepare the classification test datasets.

Compared to other applications (e.g. Yosinski et al. 2014), the size of the CNN training dataset presented in this paper is small. To simplify the classification task, only five USCS classes were selected for training the classification models: CL, CH, SP, SM, and GM. CL and CH correspond to clays with low and high plasticity, respectively, while SP denotes poorly graded sand, SM silty sand, and GM silty gravel. To further simplify the classification, the GW-GM and GP-GM classes were grouped into the GM class, while the SW-SM and SP-SM classes were integrated into the SM class. Table 4.1 gives the number of images in the classification dataset for the five USCS classes.

Table 4.1 Number of images in the training and validation datasets for classification before data augmentation

USCS class	Number of images
CH	213
CL	235
GM	236
SM	348
SP	174

Five subdatasets were created for the USCS classification tasks from the 1,206 images in Table 1. Each subdataset was used to train a separate CNN (Table 4.2).

Subdataset C1 was used to train a CNN to predict the soil classification in a single step. The C1 subdataset includes 10,000 images, or 2,000 images per class. For each class, 1,750 images were used for training and 250 for model validation. Data augmentation was applied as needed to reach the desired class size based on the images in Table 4.1. Figure 4.4 shows sample images from the C1 subdataset.

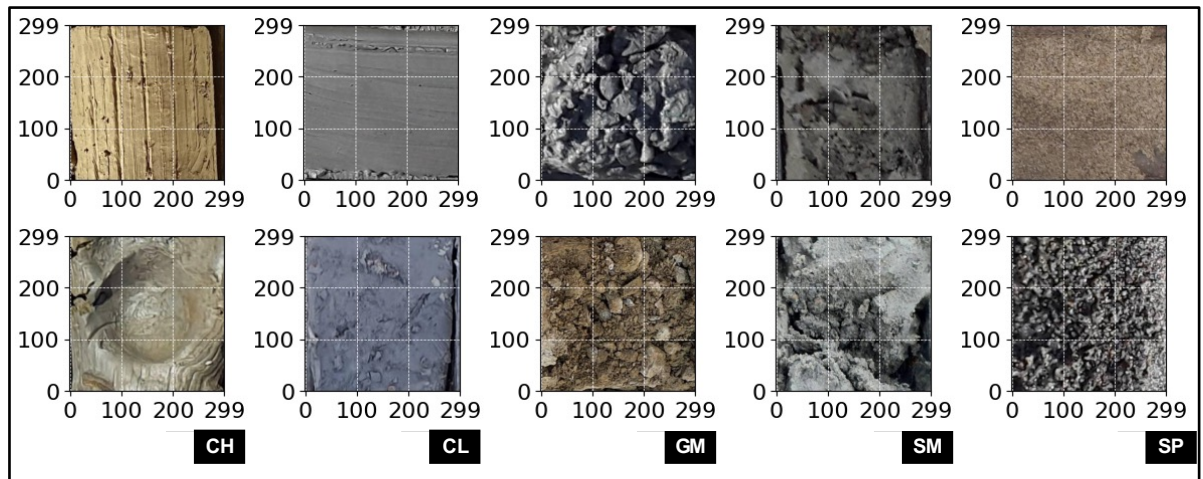


Figure 4.4 Image samples taken from subdataset C1. Each column corresponds to a USCS class, with two representative images per class. The axes are labeled in pixels

Subdatasets C2, C3, C4, and C5 were used to decompose the overall classification problem into a sequence of binary classification steps. The first model, trained on C2, separated fine-grained soils (more than 50% passing the 75 μm sieve) from coarse-grained soils. Fine-grained samples were then classified as CH or CL using a model trained on C3. For coarse-grained samples, a classifier trained on C4 distinguished gravel (GM) from sand (SP+SM). Sand images were further processed by a final model trained on C5 to separate SP from SM. Each subdataset contained 2,000 images per class (or class group); for example, C2 included 2,000 fine-grained and 2,000 coarse-grained samples. One eighth of the images in each subdataset was reserved for validation and excluded from training.

Table 4.2 Training and validation subdatasets for the classification tasks

Subdataset	Classification type	Classes or groups of classes	Number of images	
			Training	Validation
C1	Single-step classification	CL CH SP SM GM	8,750	1,250
C2	Fine-grained vs. coarse-grained soils	<u>Fine:</u> CL, CH <u>Coarse:</u> GM, SP, SM	3,500	500
C3	CH vs. CL	CH CL	3,500	500
C4	Sand vs. gravel	<u>Sand:</u> SP, SM <u>Gravel:</u> GM	3,500	500
C5	SP vs. SM	SP SM	3,500	500

A classification test dataset was created from projects completed after January 2025. It consisted of 200 images (40 per class). The test images were not subjected to data augmentation. The model trained on subdataset C1 for single step classification was tested with the entire test dataset. The other models – those trained for successive binary classification – were first tested with the test images relevant to their subtask. For example, the model trained with C3 was first evaluated on the 40 CL and 40 CH test images. The full hierarchical classification workflow was also tested by propagating classification errors through the chain of models for the 200 test images. For example, a GM sample erroneously classified as fine-grained by the first classifier was passed to the network trained to distinguish CH and CL clays, as would occur during real-world inference.

4.3.3 Regression datasets

Three separate subdatasets were created to train a series of networks to predict the proportion of fine particles (mass proportion passing the 75 μm sieve) based on split-spoon photographs. These datasets were created from 786 images obtained using the method presented in section 4.3.1. The data augmentation methods shown in Figure 4.3 were applied to increase the total number of images to 7860. Subdataset R1 includes all 7860 images. It was used to train a network to predict the proportion of fine particles for both fine-grained and coarse-grained samples. Examples of images in R1 with the corresponding proportion of fine particles are shown in Figure 4.5. The PSDs for subdataset R1 are shown in Figure 4.6a. As with classification, regression was also divided in subtasks. Subdatasets R2 and R3 allowed the percentage of fine particles to be determined for finer (i.e. more than 50 % mass passing the 75 μm sieve) and coarser materials (i.e. less than 50 % mass passing the 75 μm sieve), respectively. These networks were meant to be used after the first classification step (network trained with C2). The three subdatasets were divided randomly between training (90 % of images) and validation (10 %) data (Table 4.3).

The test dataset for the regression tasks comprises 126 unseen images divided into 32 images corresponding to fine samples and 94 coarse samples. Their PSD is given in Figure 4.6b. Data

augmentation was not applied to the test dataset. All images were used to evaluate the performance of the network trained with R1. Similarly to the classification tasks, two strategies were compared to evaluate the performance for R2 and R3. The networks were first tested only with relevant images (e.g. fines samples for R2). Secondly, all images were used to test R2 and R3 after a first classification by C2 to distinguish fine and coarse samples. This methodology allowed the hierarchical decomposition and direct approach to be compared.

Table 4.3 Training and validation subdatasets for the regression tasks

Subdataset	Regression type	Number of images	
		Training	Validation
R1	Percentage of fine particles for all materials	7,074	786
R2	Percentage of fine particles for fine materials (CH, CL)	2,232	248
R3	Percentage of fine particles for coarse materials (GM, SM, SP)	4,842	538

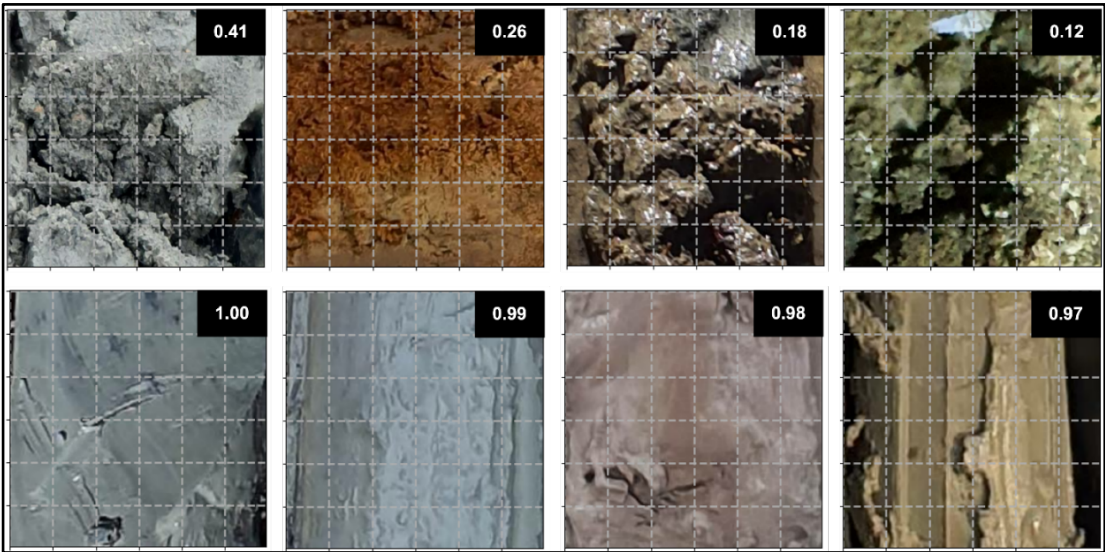


Figure 4.5 Examples of images from the R1 dataset, accompanied by the proportion of fine particles

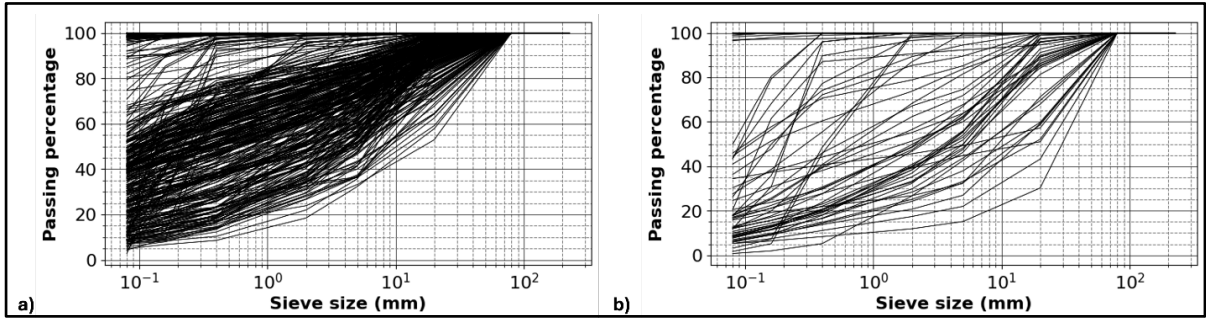


Figure 4.6 Particle size distributions of a) training/validation dataset R1 and b) the corresponding test dataset. The particle size distributions only show the sieving results

4.3.4 Model architectures and training

The CNN architectures used in this study were derived from the GrainSizeNet model (Simo Tenekam et al., 2025). The GrainSizeNet architecture is illustrated in Figure 4.7a. This model was originally trained to predict the mass proportion retained on a series of 11 sieves and the mass proportion passing the smallest sieve, effectively estimating the particle-size distribution (PSD) of granular materials from digital photographs. The convolutional component of the model consists of several filter layers, each defined by the number of convolutional filters, the size of the filter kernels, and the stride (the displacement between successive convolution windows). For example, the first convolutional layer, Conv2D (32, (3,3), stride = 2), applies 32 filters with a spatial size of 3×3 and a depth equal to the number of input channels (three for an RGB image).

Two residual blocks are included in the network. Each residual connection adds the feature maps produced by a convolutional block to those entering the same block. This helps preserve spatial information and stabilizes the learning process (He et al. 2015). Max pooling layers are used after each block to downsample the feature maps based on local maximum values to retain the strongest local responses. This downsampling reduces computational cost and helps the network focus on dominant textural patterns. The final convolutional block is followed by a global max pooling layer that condenses the remaining spatial information into a single feature vector, which contains to the maximum value for each feature map.

The feature vector is passed to a fully connected head composed of four dense layers with varying numbers of neurons (e.g., a Dense(256) layer contains 256 neurons). A ReLU activation function is used throughout the convolutional and dense layers to introduce non-linearity by setting negative outputs to zero. The final layer uses a softmax activation to ensure that the predicted mass proportions sum to 1.

The GrainSizeNet architecture was adapted for each subdataset using a hyperparameter study in which different network architectures were tested. Figure 4.7b summarizes the parameters that were varied.

Two kernel sizes were compared for the convolutional layers (3×3 and 7×7 , W in Figure 4.7b). Transfer learning was used for the 3×3 kernels. Their weights were initialized based on the GrainSizeNet kernels. Two transfer learning approaches were tested: freezing the inherited 3×3 kernels or fine-tuning their weights during training.

Different strategies were compared for the pooling layers that followed the last convolutional layers. A max pooling layer was used for some networks. The max pooling layer or the last convolutional layer was followed by either a global max pooling or a global average pooling layer. These layers take the maximum and average feature value for each feature map, respectively.

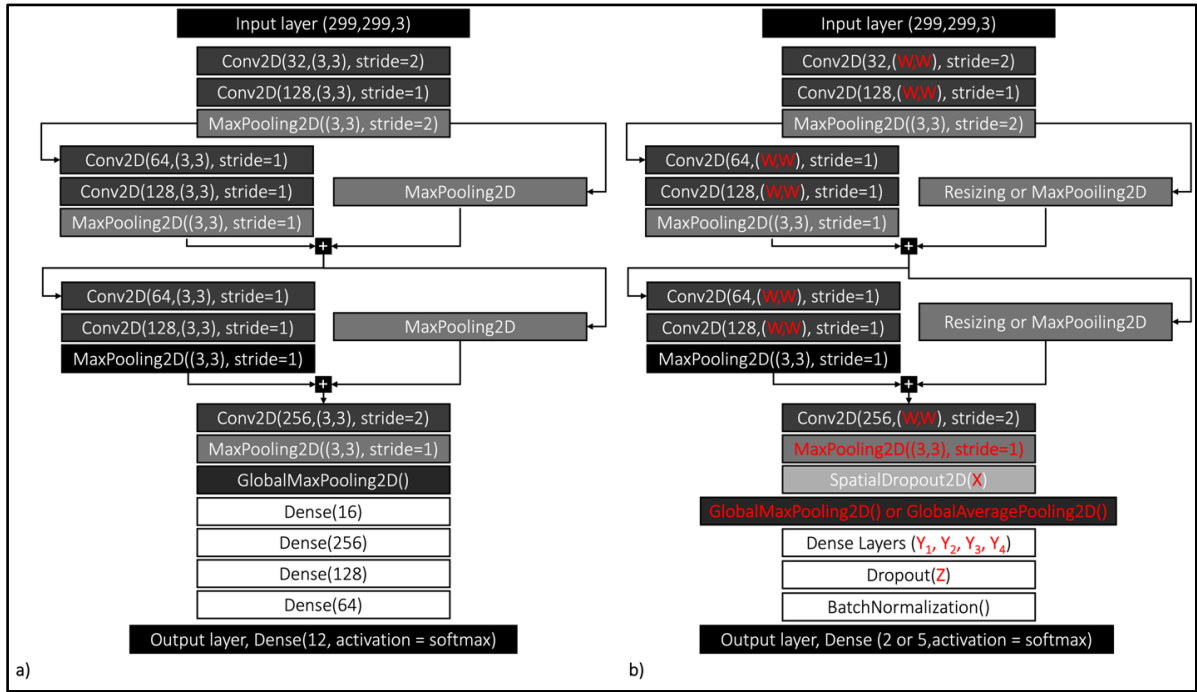


Figure 4.7 Network architecture for a) GrainSizeNet (Simo Tenekam et al., 2025) and b) the networks trained in this paper. Red letters indicate parameters that were set following the hyperparameter study, W is the kernel size for the convolutional layers, X indicates the SpatialDropout2D probability, Y_1 , Y_2 , Y_3 and Y_4 are the number of neurons in each fully connected layer and Z the dropout probability for the fully connected layers

The influence of dropout connections was investigated for both the convolutional layers (X) and fully connected layers (Y). Dropout connections for fully connected layers involve deactivating a fixed proportion of neurons randomly during training to prevent overfitting, a significant problem with small datasets. An overfitting network learns the training dataset and loses the capacity to generalize when tested with a different dataset. For convolutional layers, spatial dropouts involve deactivating a fixed proportion of filters randomly. The proportion of dropout connections and filters (Z) was varied between 0.1 and 0.5.

The regularization parameters and the number of neurons and layers in the fully connected layers were varied. Up to four layers were used with Y_1 , Y_2 , Y_3 , and Y_4 neurons. L_2 regularization was used to penalize large neuron weights in the fully connected layer. This method adds a penalty to the loss function proportional to the squared magnitude of the network weights,

discouraging large parameter values and reducing model variance. This penalty has the effect of smoothing the learned mapping and improving generalization, particularly for small or noisy datasets (Hastie et al., 2009; Goodfellow et al., 2016). The best architecture was chosen based on classification accuracy and mean absolute errors as defined in the next section. The architectures with the best performance are summarized in Table 4.4.

Categorical cross-entropy (CCE) was adopted as the loss function for classification models to assess differences between real and model-predicted probability distributions:

$$CCE = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_{c,n} \log \widehat{y}_{c,n} \quad (4.1)$$

where $y_{c,n}$ et $\widehat{y}_{c,n}$ designate respectively the real and model-predicted probabilities for class c and image n , C is the number of classes and N is the number of images. The CCE is maximum when the model predicts a uniform probability for all classes, i.e. when it is uncertain of its prediction. On the other hand, it is minimal when the model predicts a probability close to 1 for the correct class.

The mean absolute error (MAE) was used to define the training loss of the regression models:

$$MAE = \frac{1}{N} \sum_{n=1}^N |z_n - \widehat{z}_n| \quad (4.2)$$

where z_n and $\widehat{z}_n$ represent the actual and model-predicted mass proportions passing the 75 μm sieve for image n , respectively.

Table 4.4 Model architecture summary for each subdataset

Network dataset	Type	Convolutional layers				Fully connected layers		
		Conv. filter size ($W \times W$)	Transfer learning	Spatial dropout (X)	Final Pooling layers	Neurons (Y_1, Y_2, Y_3, Y_4)	Dropout (Z)	Reg.
C1	Single-step classification	7×7	No	0.1	Global avg. pool.	256	No	No
C2	Fine vs. coarse soils	3×3	Yes (all layers trainable)	No	Global avg. pool.	256	0.5	L2, 0.0001
C3	CH vs. CL	7×7	No	0.1	Max pool., Global avg. pool.	64	No	L2, 0.001
C4	Sand vs. gravel	7×7	No	0.2	Max pool., Global avg. pool.	32	0.5	L2, 0.01
C5	SP vs. SM	7×7	No	0.1	Max pool., Global avg. pool.	64	0.5	L2, 0.001
R1	Fine particles - All materials	7×7	No	0.2	Max pool., Global max. pool.	512, 256, 256, 128	0.2	L2, 0.001
R2	Fine particles - fine materials only	3×3	Yes, (first 4 conv. layers frozen)	No	Global avg. pool.	256	0.2	L2, 0.0001

Table 4.4 Model architecture summary for each subdataset (continued)

Network dataset	Type	Convolutional layers				Fully connected layers		
		Conv. filter size ($W \times W$)	Transfer learning	Spatial dropout (X)	Final Pooling layers	Neurons (Y_1, Y_2, Y_3, Y_4)	Dropout (Z)	Reg.
R3	Fine particles - coarse materials only	7×7	No	0.2	Max pool., Global max. pool.	512, 256, 256, 128	0.2	L2, 0.001

The ADAM (Adaptive Moment Estimation) optimizer was employed for training all models. This method updates each weight in the network based on a stochastic gradient descent method that incorporates the average of past gradients and their variability, resulting in adaptive update magnitudes for each weight (Kingma and Ba 2014). ADAM is well suited for noisy or heterogeneous datasets, as it stabilizes training.

The learning rate – the hyperparameter that controls the magnitude of each weight update was adjusted dynamically during training using the ReduceLROnPlateau callback provided in Keras (Keras Team, 2025). Whenever the validation loss failed to improve two epochs (i.e. two complete iterations over the training set), the learning rate was automatically reduced. This widely used strategy helps the optimizer converge toward a better minimum of the loss function (Bengio, 2012).

Training was pursued for a maximum of 30 epochs. Early stopping was employed to terminate training when the validation loss was no longer improving and to avoid overfitting. Patience, the number of consecutive epochs during which validation losses can increase without early stopping, was set to four epochs.

4.3.5 Metrics for evaluating classification and regression models

Several metrics were used alongside confusion matrices to analyse the performance of classification models. Precision indicates the proportion of correct positive predictions among all samples predicted as class c :

$$P_c = \frac{TP}{TP + FP} \quad (4.3)$$

where TP and FP represent true positives and false positives, respectively. Recall measures the proportion of samples correctly assigned to class c among all samples belonging to class c :

$$R_c = \frac{TP}{TP + FN} \quad (4.4)$$

where FN represents false negatives. The $F1$ score combines precision and recall into a harmonic mean to balance the two criteria:

$$F1_c = \frac{2TP}{2TP + FP + FN} \quad (4.5)$$

For each network, the average values across all classes of precision, recall and $F1$ score denoted as $\bar{P}$, $\bar{R}$, and $\bar{F1}$ were used as performance metrics.

The evaluation of the regression models, which predict the proportions of fine particles, is based on two key metrics: the MAE, which was also used as the loss function, and the coefficient of determination R^2 :

$$R^2 = 1 - \frac{\sum_{n=1}^N (z_n - \widehat{z}_n)^2}{\sum_{n=1}^N (z_n - \bar{z})^2} \quad (4.6)$$

Where $\bar{z}$ is the actual average proportion of fine particles. An R^2 value close to 1 implies that the model accounts for most of the variance in the measured data. The combined use of MAE and R^2 therefore makes it possible to evaluate both the average prediction error and the degree to which predictions fit the data distribution.

4.4 Results

4.4.1 Training and validation performance

Loss curves for network training and validation are shown in Figure 4.8 for all classification models. The single-step classification (C1) yielded categorical cross-entropy losses of 0.8095 on the training dataset, 0.9086 on the validation dataset and 0.8391 on the test dataset.

Hierarchical models generally converged more quickly because each binary model was trained on a specific subtask, making training easier. The hierarchical sub-models trained on C2 and C3 stopped before the maximum number of epochs (30) because of the early stopping criterion. The lowest loss was obtained for subdataset C2, which was trained to distinguish fine-grained and coarse-grained material. This is consistent with the relative simplicity of this subtask compared to the other binary classifications, i.e. distinguishing gravel vs. sand, SP vs. SM sands, and CH vs. CL clays, which are more complex.

The high loss of the model trained on C3, whose purpose was to differentiate high and low plasticity clays, reflects the difficulty of this task. The distinction between low and high plasticity clays based on visual observation is very subtle and almost impossible to establish reliably without touching the material. In practice, simple tests can be used to classify fine-grained materials in the field, such as the dilatancy and toughness tests (ASTM International 2018), but these tests involve manipulating the material. The gloss of the sample surface can also be used to distinguish between CH and CL clays. However, this criterion is strongly influenced by image quality and ambient lighting. This introduces uncertainty into the prediction.

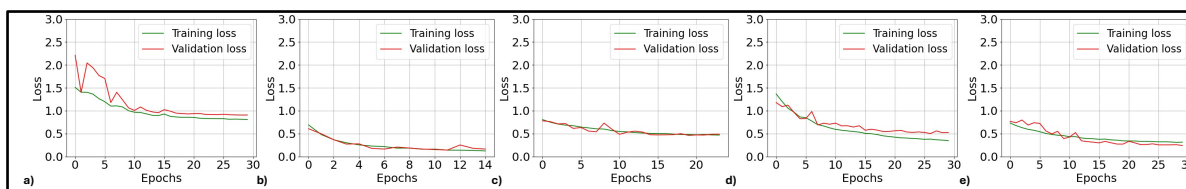


Figure 4.8 Evolution of training and validation losses during model learning for a) one-step classification with subdataset C1, and hierarchical models corresponding to subdatasets b) C2, c) C3, d) C4, and e) C5

The loss curves for the regression models based on the mean absolute error (MAE) are shown in Figure 4.9. The models trained with R1 and R3 to predict the proportion of fine particles for both fine- and coarse-grained materials, and only coarse-grained materials, respectively, show very similar loss curves during training. In contrast, the model trained with R2 (fine-grained materials) shows a rapid decrease in loss within the first five epochs. The final losses for the models trained on R1, R2, and R3 based on the relevant test dataset are 0.241, 0.035, and 0.238, respectively.

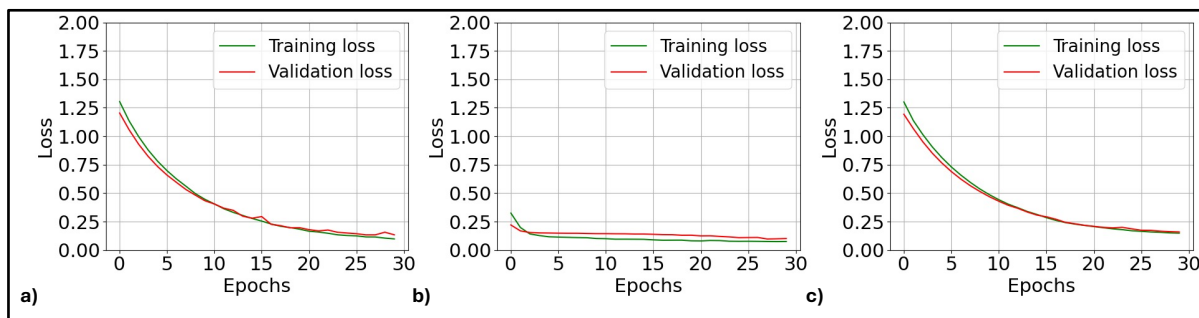


Figure 4.9 Training and validation loss for the regression models predicting the proportion of fine particles for a) all materials using subdataset R1, b) fine-grained materials using subdataset R2, and c) coarse-grained materials using subdataset R3

4.4.2 Classification results

The performance metrics for the classification networks are given in Table 4.5. Results are given for both single-step classification (network trained with subdataset C1) and hierarchical

classification (networks trained with subdatasets C2, C3, C4, and C5). For the hierarchical classification, performance metrics are given for each binary classification step when considering only the relevant images and for the full chain of models when propagating the classification errors.

The model trained on C1 achieves a $\bar{R}$ of 0.625, implying that 62.5 % of the samples were classified correctly. The F1 score is more sensitive to differences in performance between classes. A slightly lower $\overline{F1}$ score indicates uneven classification performance across classes when using the single-step approach.

Hierarchical classification performed better than single step classification, but the performance metrics clearly depends on the task complexity. The chain of models led to an improved overall $\bar{R}$ of 0.690. A near-perfect $\bar{R}$ of 0.995 was obtained for the network trained with C2 indicating that 99.5 % of the samples were correctly identified as fine-grained or coarse-grained. On the other hand, the model trained to differentiate between the CH and CL (C3) had a $\bar{R}$ of 0.600, which is only slightly better than a random choice between CH or CL that would give a $\bar{R}$ of approximately 0.500. Models C4 (SP vs. SM) and C5 (SP-SM vs. GM) performed well, with $\bar{R}$ of 0.825 and 0.863, respectively.

Table 4.5 Performance of the classification models on the test dataset

Network subdataset	Type	Performance metrics		
		$\bar{P}$	$\bar{R}$	$\overline{F1}$
C1	Single-step classification	0.625	0.625	0.589
C2, C3, C4, C5	Classification with chain of models	0.708	0.690	0.678
C2	Fine vs. coarse soils	0.995	0.995	0.995
C3	CH vs. CL	0.633	0.600	0.573
C4	Sand vs. gravel	0.825	0.825	0.829
C5	SP vs. SM	0.863	0.863	0.862

The confusion matrices for the single-step and hierarchical models are shown in Figure 4.10. The matrices clearly show that both approaches were able to distinguish fine- and coarse-grained soils relatively easily with almost all samples in either the CL-CH or GM-SM-SP squares of the confusion matrices. Single-step classification performed very poorly with fine-grained soils. Most of them were classified as CH, which resulted in a P_c of only 0.527 for class CH and a R_c of only 0.125 for class CL.

4.4.3 Regression results

Regressions were used to predict the mass proportion of particles passing the 75 μm sieve from split-spoon photographs. Two methods were compared: a single-step approach using a model trained on R1 and a hierarchical approach combining networks trained on C2 for fine- and coarse material classification, R2 for regression on fine materials, and R3 for regression on coarse materials.

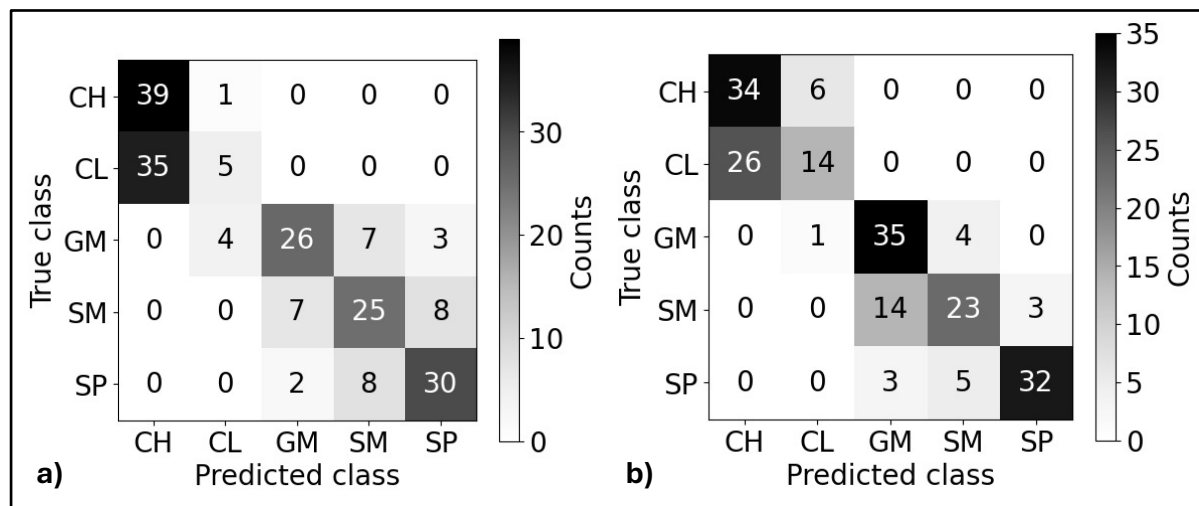


Figure 4.10 Confusion matrices illustrating the classification results for the test dataset according to a) the single-step and b) hierarchical approaches

Figure 4.11a compares the actual and predicted mass proportions of particles passing the 75 μm sieve for the single step approach. The network trained on R1 achieved a MAE on the proportion fine particles of 0.21. The correlation is relatively poor with a R^2 of 0.525.

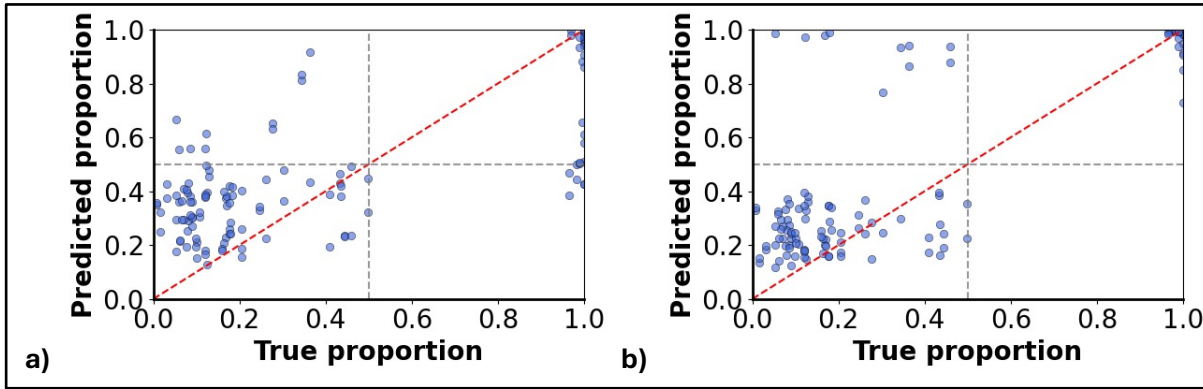


Figure 4.11 Comparison of predicted and actual mass proportions of particles passing the 75 μm sieve for a) the single-step model trained on R1 and b) the hierarchical model combining classification with C2 and regression with R2 for fine soils and R3 for coarse soils

Figure 4.11b compares the actual and predicted mass proportions of fine particles for the hierarchical approach. A $\bar{R}$ of 0.947 was obtained for the C2 classifier with 10 coarse materials incorrectly labeled as fine-grained and all fine materials labeled correctly. This very good performance is consistent with the results obtained with the other test dataset in the classification results. When using the model trained on R2 for fine materials and the model trained on R3 for coarse material, as determined by the C2 classifier, the MAE on the mass proportion of particles passing the 75 μm sieve was reduced to 0.153 and the R^2 was increased to 0.611 compared to single-step regression.

4.5 Discussion

Figure 4.12 shows representative images from the regression test dataset, accompanied by the mass proportions of fine particles predicted using the single-step and hierarchical models. The very good performance obtained for the C2 classifier (fine- versus coarse-grained materials)

can be readily explained by visual inspection of the images in Figure 4.12, as fine- and coarse-grained materials exhibit markedly different surface textures.

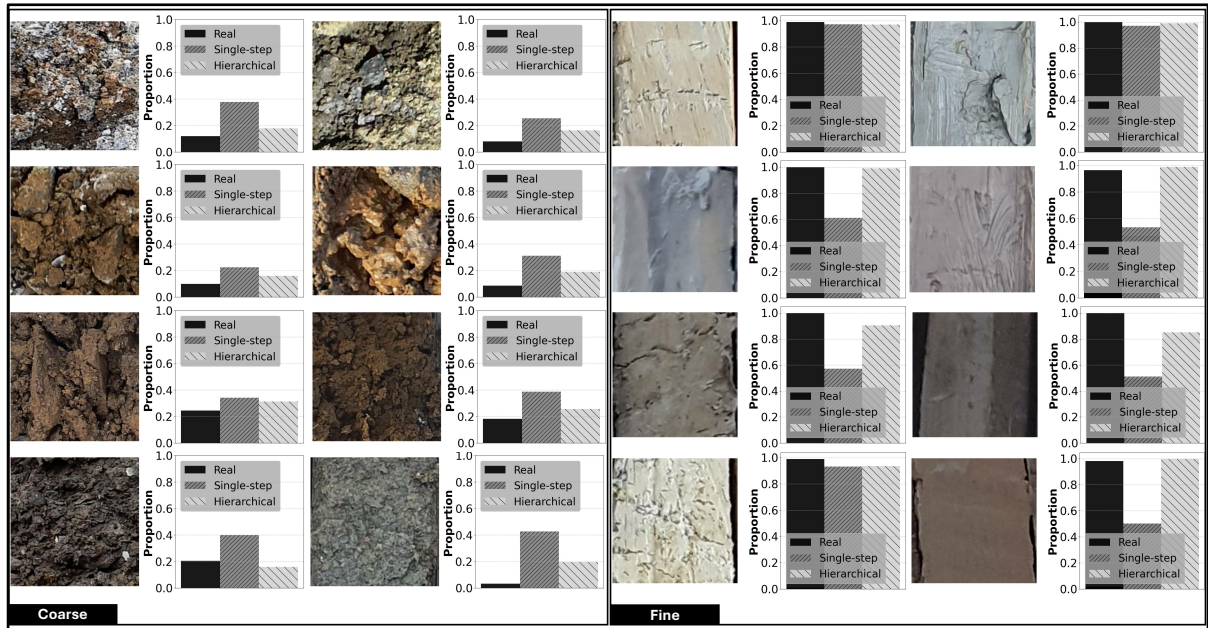


Figure 4.12 Sample images from the regression test dataset with the actual and predicted mass proportions of fine particles for the single-step and hierarchical regression models. The images in the first and second columns correspond to fine and coarse materials, respectively

The distribution of fine particle proportions in the test dataset also strongly favours this separation. As shown in Figure 4.11, the proportion of particles passing the 75 μm sieve is either below 0.50 or above 0.96. A similar distribution is observed in the training dataset, where only 15 % of the samples exhibit fine particle proportions between 0.50 and 0.96. This pronounced gap in fine particle content likely contributes to the ease with which fine- and coarse-grained materials can be distinguished. The strong concentration of fine particle proportions above 0.96 for fine-grained materials also explains the rapid decrease of the training loss in Figure 9b as it reduces the true proportion variability.

The hierarchical approach led to better performances for all tasks. This is consistent with the results reported in the literature for other classification applications (Silla and Freitas 2011).

Performance metrics were poorer for the downstream tasks. The worst metrics were obtained for the C3 classifier and the fine particle regression with R2 and R3. The C3 classifier performed only slightly better than choosing randomly between the CL and CH. Taken separately, the relationship between actual and predicted proportions for the networks trained on datasets R2 and R3 do not show any correlation. Downstream classification and regression tasks would benefit from larger training subdatasets. The performance metrics could be improved by focusing on these subdatasets. It should be noted that the training datasets (total of 1992 images) are very small compared to typical CNN applications that often involve millions of images (e.g. Russakovsky et al. 2015).

The performance metrics that were obtained in this work are typical of networks trained with small dataset for similar tasks. For instance, the performance of the classification models can be compared with the results presented by Buscombe (2020) for its classification task. Six classes were considered by Buscombe (2020): well-sorted gravel, well-sorted sand and shell debris captured by underwater cameras, relatively poorly sorted sand, well-sorted volcanoclastic sand, well-sorted terrigenous sand and poorly sorted terrigenous sand. To compare material with similar soil textures, only The GM, SM, and SP grain classes were included in the comparison dataset. Table 4.6 presents the performance metrics of the different models evaluated on their respective test data. The normalized confusion matrices presented in Figure 4.13 allow for a qualitative assessment of the classification errors specific to each model. The performance metrics are slightly better for the Buscombe classification task with $\bar{R} = 0.837$ compared to $\bar{R} = 0.757$ for C4 and C5. Considering that the split-spoon dataset is highly heterogeneous and that it combines low-resolution images (299x299) obtained from different angles for different projects, the performance metrics obtained with C4 and C5 appear to be consistent with those obtained by Buscombe (2020).

Table 4.6 Comparison of the performance metrics for the SediNet classification task (Buscombe 2020) and the texturally similar GM, SP, and SM classification

Model	$\bar{P}$	$\bar{R}$	$\bar{F1}$
SediNet	0.885	0.837	0.843
C1	0.702	0.697	0.699
C4-C5	0.772	0.757	0.755

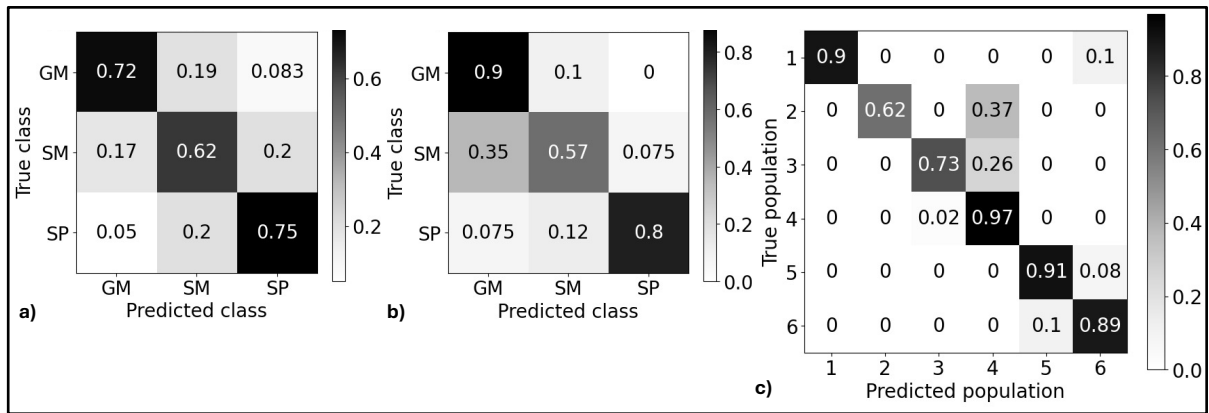


Figure 4.13 Normalized confusion matrices for a) C1 and b) the hierarchical models tested on the GM, SM, and SP classes, compared with c) Buscombe's (2020) SediNet model

The handling of the sample by the technician before taking the photograph can introduce biases affecting data quality. In some cases, for cohesive soils such as highly plastic clays, this handling can result in a texture that appears more granular than lustrous or fine (Figure 4.14), a bias that is difficult to control and may compromise model performance. Furthermore, the particle overlap in the sample can introduce errors, affecting the interpretation of fine and coarse particle percentages. In addition, factors such as brightness, contrast, shadows and image quality can adversely affect texture recognition. Drilling operations often take place at night on highways, which can alter the quality of photographs and influence texture interpretation due to artificial lighting and the shadows created. Standardizing the imaging procedure is therefore essential to minimize these biases and improve model reliability.

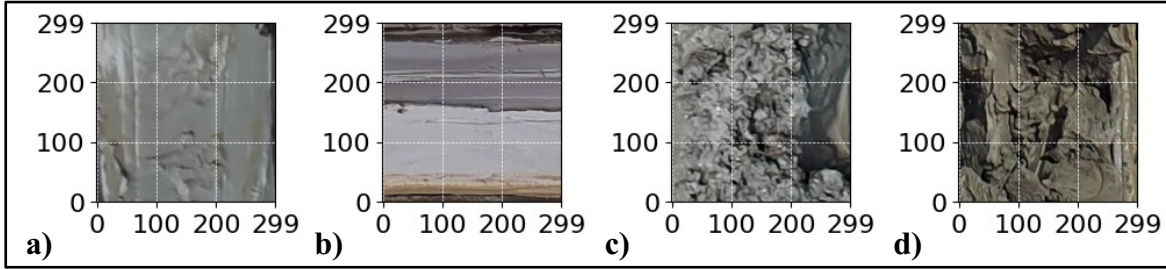


Figure 4.14 Typical appearances of fine-grained soils after separating the split-spoon for a) CH and b) CL, and after manual handling by the technician for field classification for c) CH and d) CL

4.6 Conclusions

This article explores the adaptation of GrainSizeNet for the automatic classification of soils according to the USCS and the estimation of the mass fraction of fines (diameter $< 75 \mu\text{m}$) from photographs taken using split-spoon samplers. The total training dataset comprised 1992 images with subsets annotated either with USCS classes (classification) or with particle size distributions (regression). Evaluations were conducted on the independent test datasets comprising 200 images for classification and 132 images for regression.

The single-step C1 model correctly classified 62.5% of the 200 test samples ($\bar{R} = 0.625$), while the hierarchical chain improved the overall recall to 0.690. Inspection of the hierarchical approach models reveals that some tasks are handled very robustly (e.g., C2 for the fine vs. coarse distinction: $\bar{R} = 0.995$, and C4, which returns an $\bar{R}$ of 0.829 for the differentiation of sand vs. gravel), while others remain difficult (e.g., C3, which is responsible for distinguishing between the CH and CL classes, has an $\bar{R}$ of 0.60). For regression, the single-step approach gives an $\text{MAE} = 0.21$ and $R^2 = 0.525$, while the hierarchical approach reduces the MAE to 0.153 and increases the R^2 to 0.611 on the 132 images tested. The strong performance of the hierarchical approach indicates that dividing the problem into targeted subtasks helps to reduce confusion between soils with similar textures.

Although hierarchical models offer better results, the performance of both approaches remains limited by common constraints related to data acquisition. Handling a sample before taking photographs can alter its surface, especially for highly cohesive soils with a plastic behavior, their texture may then appear artificially granular. The overlapping of particles also makes it difficult to estimate the fine particle proportions. Finally, field conditions such as brightness, shadows, or night photography directly affect the quality of textural information in the images.

Future progress will depend on expanding and diversifying training datasets. This is especially important for downstream subtasks where current performance is limited. Greater attention should be paid to standardizing field photography protocols to reduce variability related to site conditions and improve model robustness. These concerns need to be resolved before applying the proposed approach to other USCS classes and ultimately strengthening its applicability as a practical tool for image-based soil characterization in geotechnical investigations.

CONCLUSION

Cette thèse avait pour but d'explorer les approches novatrices d'intelligence artificielle pour surmonter les contraintes de temps et de coûts inhérents aux méthodes traditionnelles d'analyse granulométrique. À cette fin, en combinant les méthodes d'apprentissage automatique avec les données granulométriques récoltées sur cinq années de projets d'investigation géotechnique, un modèle de réseaux de neurones convolutionnels nommé GrainSizeNet a permis de répondre à plusieurs défis spécifiques à l'analyse des sols. Les principales conclusions tirées de cette thèse sont présentées dans cette section.

Quatre ensembles de données ont été exploités combinant des données réelles et synthétiques. Les données réelles provenaient d'une part d'un ensemble de matériaux artificiels préfabriqués en laboratoire avec des granulométries prédéfinies. D'autre part d'une série de campagnes d'investigation géotechnique, au cours desquelles les échantillons ont été photographiés avant d'être analysés en laboratoire. Les images de synthèse ont été modélisées avec le moteur de jeu vidéo Unity game engine. Un programme dédié a été développé pour simuler le déversement des matériaux granulaires dans un contenant cubique avec une distribution granulométrique ciblée. Les échantillons simulés ont été photographiés et segmentés automatiquement à l'aide de l'outil natif Perception de Unity game engine. Cette approche a permis de diversifier les données granulométriques résultantes de l'analyse des échantillons réels sans nécessiter la collecte de nouveaux échantillons.

Les images segmentées sont couramment utilisées par les logiciels d'analyse de fragmentation des roches dans les opérations minières pour évaluer l'efficacité du dynamitage. Leur incorporation dans l'étape d'apprentissage du modèle a démontré que bien qu'elles mettent en évidence les contours des particules en simplifiant l'information contenue dans l'image, elles n'apportent pas de valeur ajoutée significative à l'estimation de la granulométrie par traitement d'image. Cette observation s'explique principalement parce que même si la granulométrie est considérée comme une distribution de taille de particules, la superposition naturelle des particules est un biais considérable sur l'échantillon photographié à un instant précis. Ce

phénomène réduit l'intérêt des contours individuels des particules superposées. En revanche, l'apparence globale de l'échantillon s'avère plus représentative de sa granulométrie réelle.

L'ensemble de données final initialement constitué de plusieurs images de tailles variées a été normalisé en redimensionnant toutes les images à une taille de 299 par 299 pixels. Bien que cette résolution soit relativement faible, ce choix a permis d'accélérer considérablement le temps de traitement des images durant l'entraînement du modèle. Par ailleurs, GrainSizeNet a été formé pour se concentrer sur l'aspect global de l'image plutôt que de se focaliser sur les primitives tonales de textures individuelles. La réduction de dimension n'a pas altéré l'aspect global des images utilisées et a permis une analyse et interprétation précises des textures granulaires et fines. Finalement, cette étude a démontré que les particules de sol n'ont pas nécessairement besoin d'être discernées visuellement les unes des autres pour pouvoir être caractérisées.

Nous présentons GrainSizeNet, un réseau neuronal convolutif à blocs résiduels conçu pour l'analyse granulométrique automatisée à partir de photographies de sols. GrainSizeNet a été entraîné sur 5 475 images de sols (réelles, synthétiques et segmentées), chacune couplée à sa courbe granulométrique. Deux architectures avaient été comparées dont l'une avait une configuration à deux entrées (vues du dessus et du dessous) et l'autre à une entrée (image du dessus). L'architecture à une seule vue, sans recours aux images segmentées, a atteint la valeur de MAE la plus faible soit 2,90 %. Bien que les écarts de MAE observés par rapport aux autres modèles demeurent modérés, ces résultats suggèrent d'une part que la vue du dessus capture l'essentiel des variations texturales et d'autre part que l'utilisation exclusive d'images réelles et synthétiques réalistes est plus efficace que les combinaisons incluant des images segmentées. La simplification de l'information texturale des images segmentées peut nuire à la représentation fidèle des textures des sols. Par ailleurs, l'introduction progressive de données synthétiques a réduit significativement l'erreur de prédiction des proportions de tailles de particules (MAE passant de 11,54% à 6,19%) sur les classes granulométriques sous-représentées dans les données d'entraînement, confirmant ainsi leur apport pour la généralisation du modèle. Enfin, la correspondance entre classifications MTQ 1101 réelles et

prédites s'avère très élevée (plus de 91% des matériaux testés ont été bien classés), validant la robustesse de GrainSizeNet comme alternative rapide et fiable aux méthodes de tamisage classiques.

Une adaptation du modèle de base de GrainSizeNet a été faite afin de transférer son application. Initialement entraîné sur des images obtenues dans des conditions idéales de laboratoire, ce modèle a été appliqué sur des images capturées directement sur les sites d'investigation géotechnique. L'approche d'apprentissage par transfert introduite par Yosinski et al. (2014) a permis d'outrepasser les contraintes inhérentes aux faibles résolutions des photographies de terrain et aux conditions environnementales défavorables. Deux adaptations ont été proposées, la première destinée à la classification des sols en cinq catégories référées par le système USCS (CH, CL, GM, SM, SP). Pour un jeu de données de 1206 échantillons associé à 200 autres échantillons distincts réservés exclusivement pour l'évaluation du modèle, GrainSizeNet a atteint une précision globale de 69% malgré les similarités quasiment indissociables à l'œil humain entre les matériaux de classe CL et CH. La seconde variante de régression destinée à prédire les proportions de matériaux de sol fins a atteint une erreur absolue moyenne (MAE) de 15,32% sur un total de 126 échantillons du jeu de test. Malgré des conditions environnementales difficiles et une faible résolution des images, GrainSizeNet démontre un bon potentiel pour l'analyse d'images de sols directement au chantier.

RECOMMANDATIONS

À la lumière des résultats obtenus et des limitations observées au cours de cette étude, plusieurs pistes d'amélioration et de développement futur peuvent être envisagées. Notamment l'enrichissement des jeux de données d'entraînement, l'optimisation du processus de prétraitement des données d'apprentissage et de sélection d'hyperparamètres des modèles, l'adaptation plus poussée de modèles de réseaux de neurones profonds pré-entraînés sur de larges bases de données génériques afin de mieux prendre en compte les spécificités du problème granulométrique.

L'augmentation du volume de données d'entraînement est fondamentale pour améliorer les performances des modèles d'apprentissage en profondeur. L'intégration tant qualitative que quantitative d'échantillons de provenance de sites aux conditions environnementales et géographiques variées renforcerait la robustesse et la capacité de généralisation du modèle de réseau de neurones. Bien que les images synthétiques soient généralement palliatives au problème d'insuffisance de données durant les étapes d'entraînement, elles sont souvent dépourvues d'artéfacts texturaux caractéristiques des images réelles. Cette disparité pourrait à l'opposé affecter négativement les performances des modèles à généraliser efficacement sur les données réelles spécifiques.

Une exploration plus approfondie des méthodes d'échantillonnage des données d'entraînement issues du terrain serait pertinente pour minimiser les biais d'interprétation des modèles. Dans cette étude, le prétraitement des images de chantier repose sur la sélection d'une région d'intérêt limitée au sein de la cuillère d'échantillonnage, choix guidé principalement par l'expérience et l'appréciation visuelle du caractère vraisemblable de la zone sélectionnée par rapport aux résultats d'analyse du laboratoire. L'utilisation de la cuillère fendue dans son intégralité permettrait de mieux capturer la variabilité spatiale réelle de l'échantillon, améliorant ainsi la représentativité globale des données. De plus, il serait important de filtrer les images des échantillons de sol à sélectionner en fonction de leur teneur en eau ou condition

d'humidité. Les sols généralement prélevés sur site ont des conditions manifestées par les états secs, humides et saturés. Les matériaux saturés pour lesquels une proportion importante des particules est immergée, sont susceptibles de rendre la caractérisation granulométrique plus difficile et pourraient biaiser l'apprentissage du modèle

Par ailleurs, au-delà des aspects d'enrichissement et de diversification des bases de données, il serait intéressant d'automatiser le processus de sélection d'hyperparamètres à l'aide d'algorithmes bayésiens qui reposent sur des connaissances préalablement établies pour inférer sur le choix optimal des hyperparamètres durant l'apprentissage du modèle. Parmi ceux-ci, Optuna offre un cadre d'optimisation efficace permettant d'explorer efficacement l'espace des hyperparamètres à partir des stratégies d'échantillonnages adaptatives. L'utilisation de tels outils constituerait une piste prometteuse pour améliorer les performances des modèles de caractérisation de sols tant pour les tâches de classification que de régression.

Une approche intéressante serait d'explorer davantage les possibilités d'adaptation des modèles de réseaux de neurones profonds reconnus pour leur robustesse et leurs hautes performances tels que ceux pré-entraînés sur de vastes bases de données génériques en vision par ordinateur (p. ex. ImageNet comportant plus de 1000 classes). Plusieurs modèles existants comme Xception, Resnet101, VGG16 et EfficientNet-B4, bien que ces modèles n'aient pas été entraînés sur des images de sols, ils détiennent une forte capacité à extraire des représentations visuelles robustes, transférables et exploitables efficacement dans une grande variété de tâches de vision par ordinateur. Il est donc possible de s'appuyer sur le développement de fonctions de perte dédiées intégrant des contraintes physiques, statistiques et géotechniques pour encadrer et contraindre l'adaptation de ces modèles dans la caractérisation des sols.

Ces recommandations visent à orienter les futures recherches et promouvoir les applications pratiques dans le domaine de l'analyse granulométrique par techniques d'analyse d'images et apprentissage en profondeur.

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- AASHTO. (2022). *Standard method of test for particle size analysis of soils (AASHTO T88-22)*. American Association of State Highway and Transportation Officials.
- Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. GitHub. https://github.com/matterport/Mask_RCNN
- Abdusalomov, A. B., Nasimov, R., Nasimova, N., Muminov, B., & Whangbo, T. K. (2023). Evaluating synthetic medical images using artificial intelligence with the GAN algorithm. *Sensors*, 23(7), 3440. <https://doi.org/10.3390/s23073440>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2623–2631). Association for Computing Machinery. <https://doi.org/10.1145/3292500.3330701>
- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292. <https://doi.org/10.3390/electronics8030292>
- American Association of State Highway and Transportation Officials. (2022). Manual on subsurface investigations (2nd ed.). <https://store.transportation.org/Item/CollectionDetail?ID=235>
- Anouar, M. (2023). L'importance de la préparation des données pour l'apprentissage automatique. Astera. <https://www.astera.com/fr/type/blog/data-preparation-for-machine-learning/>
- ASTM International. (2015). ASTM C136-06: *Standard test method for sieve analysis of fine and coarse aggregates*. ASTM International. <https://www.astm.org/c0136-06.html>
- ASTM International. (2017). ASTM D6913-04(2009)e1: Standard test methods for particle-size distribution (gradation) of soils using sieve analysis. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D6913-04R09E01>
- ASTM International. (2018). ASTM D2488-17e1: *Standard practice for description and identification of soils (visual-manual procedures)*. ASTM International. <https://www.astm.org/d2488-17e01.html>
- ASTM International. (2018). *ASTM D4318-17e1: Standard test methods for liquid limit, plastic limit, and plasticity index of soils*. ASTM International.

- ASTM International. (2020). ASTM D2487-17e1: *Standard practice for classification of soils for engineering purposes (Unified Soil Classification System)*. ASTM International. <https://www.astm.org/d2487-17e01.html>
- ASTM International. (2021). ASTM D6913/D6913M-17: *Standard test methods for particle-size distribution (gradation) of soils using sieve analysis*. ASTM International. https://www.astm.org/d6913_d6913m-17.html
- ASTM International. (2021). ASTM D7928-21e1: *Standard test method for particle-size distribution (gradation) of fine-grained soils using the sedimentation (hydrometer) analysis*. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D7928-21E01>
- ASTM International. (2021). *ASTM E112-13 Standard test methods for determining average grain size*. <https://www.astm.org/e0112-13r21.html>
- ASTM International. (2022). ASTM D1586/D1586M-18: *Standard test method for standard penetration test (SPT) and split-barrel sampling of soils*. ASTM International. https://store.astm.org/d1586_d1586m-18e01.html
- ASTM International. ASTM C136/C136M-06 (2015). *Standard Test Method for Sieve Analysis of Fine and Coarse Aggregates*.
- Bamford, T., Esmaili, K., & Schoellig, A. P. (2021). A deep learning approach for rock fragmentation analysis. *International Journal of Rock Mechanics and Mining Sciences*, 145, 104839. <https://doi.org/10.1016/j.ijrmms.2021.104839>
- Baraian, A., Kellokumpu, V., Paaso, J., Koresaar, L., & Kaartinen, J. (2022). Computing particle size distribution of mineral rocks using deep learning-based instance segmentation. In *2022 10th European Workshop on Visual Information Processing (EUVIP)* (pp. 1–6). IEEE. <https://doi.org/10.1109/EUVIP53989.2022.9922748>
- Barth, R., Hemming, J., & Van Henten, E. J. (2020). Optimising realism of synthetic images using cycle generative adversarial networks for improved part segmentation. *Computers and Electronics in Agriculture*, 173, 105378. <https://doi.org/10.1016/j.compag.2020.105378>
- Basha, S. H. S., Dubey, S. R., Pulabaigari, V., & Mukherjee, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378, 112–119. <https://doi.org/10.1016/j.neucom.2019.10.008>
- Bashir, S. D., & Jumba, A. (2020). Experimental investigation of segregation of granular materials using the vibrational phenomenon. *International Journal of Advances in*

Scientific Research and Engineering, 6(5), 193–207. <https://doi.org/10.31695/IJASRE.2019.33673>

- Bengio, Y. (2012). Practical Recommendations for Gradient-Based Training of Deep Architectures. In *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_26
- Bergstra, J., Yamins, D., & Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)* (Vol. 28, pp. 115–123). PMLR.
- Bharati, P., & Pramanik, A. (2020). Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey. In A. K. Das, J. Nayak, B. Naik, S. K. Pati, & D. Pelusi (Eds.), *Computational Intelligence in Pattern Recognition* (pp. 657–668). Springer. https://doi.org/10.1007/978-981-13-9042-5_56
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- BNQ 2501-025. (2013). *Sols—Analyse granulométrique des sols inorganiques*. Bureau de normalisation du Québec. <https://bnq.qc.ca/fr/normalisation/genie-civil-et-infrastructures-urbaines/sols/sols-analyse-granulometrique-des-sols-inorganiques.html>
- BoBo, T. (2001). *What's new with the digital image analysis software Split-Desktop®?* Split Engineering.
- Bouchard, J. (2016). *Caractérisation de l'enrochement du barrage Romaine-2* [Mémoire de maîtrise, Université Laval].
- Bouchard, J., Smith, M., & Konrad, J. (2013). Rockfill size evaluation using digital image analysis. In *GeoMontréal 2013* (Montréal, Canada). Canadian Geotechnical Society.
- Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Harmouch, H., & Naumann, F. (2022). The effects of data quality on machine learning model performance. *arXiv*. <https://doi.org/10.48550/arXiv.2207.14529>
- Bureau de normalisation du Québec. (2019). *Sols : Analyse granulométrique des sols inorganiques (BNQ 2501-025)*. Québec, Canada : Bureau de normalisation du Québec.
- Buscombe, D. (2020). SediNet: A configurable deep learning model for mixed qualitative and quantitative optical granulometry. *Earth Surface Processes and Landforms*, 45(3), 638–651. <https://doi.org/10.1002/esp.4760>

- Buscombe, D. (2020). SediNet: A configurable deep learning model for mixed qualitative and quantitative optical granulometry. *Earth Surface Processes and Landforms*, 45(3), 638–651. <https://doi.org/10.1002/esp.4760>
- Cai, Z., & Vasconcelos, N. (2018). Cascade R-CNN: Delving into High Quality Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), 6154–6162. <https://doi.org/10.1109/CVPR.2018.00644>
- Campbell, A., & Thurley, M. (2017). Application of laser scanning to measure fragmentation in underground mines. *Mining Technology*, 126(1), 1–8. <https://doi.org/10.1080/14749009.2017.1296668>
- Chang, F.-J., & Chung, C.-H. (2012). Estimation of riverbed grain-size distribution using image-processing techniques. *Journal of Hydrology*, 440–441, 102–112. <https://doi.org/10.1016/j.jhydrol.2012.03.032>
- Chen, M., Li, M., Li, Y., & Yi, W. (2021). Rock particle motion information detection based on video instance segmentation. *Sensors*, 21(12), 4108. <https://doi.org/10.3390/s21124108>
- Chen, Y., Liu, K., Xin, Y., & Zhao, X. (2023). Soil image segmentation based on Mask R-CNN. In *2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 507–510). IEEE. <https://doi.org/10.1109/ICCECE58074.2023.10135317>
- Chen, Z., Chen, C., Shao, C., Cai, C., Song, X., Xiang, Y., Liu, R., & Xuan, Q. (2024). MITNet: GAN-enhanced magnetic induction tomography based on complex convolutional neural networks. *IEEE Sensors Journal*, 24(20), 33387–33397. <https://doi.org/10.1109/JSEN.2024.3390200>
- Choi, J., Kim, S., Jin, J., Kim, J., Chang, S., & Kim, I. (2025). Estimation of sediment grain size distribution using optical image-based spatial feature representation learning with data augmentation. *Journal of Marine Science and Engineering*, 13(6), 1108. <https://doi.org/10.3390/jmse13061108>
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1251–1258). IEEE.
- Croitoru, F.-A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9), 10850–10869. <https://doi.org/10.1109/TPAMI.2023.3261988>

- Csurka, G., Volpi, R., & Chidlovskii, B. (2022). Semantic image segmentation: Two decades of research. *Foundations and Trends® in Computer Graphics and Vision*, 14(1–2), 1–162. <https://doi.org/10.1561/06000000095>
- De la Fuente, N., Majó, M., Luzko, I., Córdova, H., Fernández-Esparrach, G., & Bernal, J. (2024). Enhancing image classification in small and unbalanced datasets through synthetic data augmentation. In K. Drechsler, C. Oyarzun Laura, M. Freiman, Y. Chen, S. Wesarg, & M. Erdt (Eds.), *Clinical image-based procedures* (pp. 11–21). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73083-2_2
- Deleuil, M. (1993). Échantillonnage des solides divisés. Homogénéisation, ségrégation. *Sciences géologiques. Bulletins et mémoires*, 46(1), 5–12. <https://doi.org/10.3406/sgeol.1993.1894>
- Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *arXiv*. <https://doi.org/10.48550/arXiv.2105.05233>
- Duhaime, F., Pirnia, P., Manashti, M., Temimi, M., Dubé, J.-S., & Toews, M. (2021). Particle size distribution from photographs: Comparison of synthetic and real granular material images. In GeoNiagara 2021 (Niagara Falls, Canada) Canadian Geotechnical Society.
- Frei, M., & Kruis, F. E. (2020). Image-based size analysis of agglomerated and partially sintered particles via convolutional neural networks. *Powder Technology*, 360, 324–336. <https://doi.org/10.1016/j.powtec.2019.10.020>
- Gaidon, A., Wang, Q., Cabon, Y., & Vig, E. (2016). Virtual worlds as proxy for multi-object tracking analysis (arXiv:1605.06457). *arXiv*. <https://doi.org/10.48550/arXiv.1605.06457>
- Girshick, R. (2015). Fast R-CNN. *arXiv*. <https://doi.org/10.48550/arXiv.1504.08083>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10), 580–587. <https://doi.org/10.1109/TPAMI.2014.2372895>
- Gong, J., Liu, Z., Zhao, K., Xu, H., Zheng, Y., Jiang, J., & Ou, X. (2024). Quantification of particle size and shape of sands based on the combination of GAN and CNN. *Powder Technology*, 447, 120122. <https://doi.org/10.1016/j.powtec.2024.120122>
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson. pp. 545–547. <https://dl.icdst.org/pdfs/files4/01c56e081202b62bd7d3b4f8545775fb.pdf>

- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *arXiv*. <https://doi.org/10.48550/arXiv.1406.2661>
- Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep learning*. MIT press, Cambridge, MA.
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in Information Retrieval* (Vol. 3408, pp. 345–359). Springer. https://doi.org/10.1007/978-3-540-31865-1_25
- Haar, A. (1909). *Zur Theorie der orthogonalen Funktionensysteme* [Thèse de doctorat, Georg-August-Universität Göttingen].
- Hafiz, A. M., & Bhat, G. M. (2020). A survey on instance segmentation: State of the art. *International Journal of Multimedia Information Retrieval*, 9(3), 171–189. <https://doi.org/10.1007/s13735-020-00195-x>
- Hamzeloo, E., Massinaei, M., & Mehrshad, N. (2014). Estimation of particle size distribution on an industrial conveyor belt using image analysis and neural networks. *Powder Technology*, 261, 185–190. <https://doi.org/10.1016/j.powtec.2014.04.038>
- Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5), 786–804. <https://doi.org/10.1109/PROC.1979.11328>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 6, 610–621.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning (2nd edition)*. Springer, New York, NY.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN (arXiv:1703.06870). *arXiv*. <http://arxiv.org/abs/1703.06870>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv:1512.03385*. <https://doi.org/10.48550/arXiv.1512.03385>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)* (pp. 770–778). IEEE. <https://doi.org/10.1109/CVPR.2016.90>
- Head, K. H. (1980). *Manual of soil laboratory testing* (Vol. 1). Pentech Press.
- Hengl, T., de Jesus, J. S. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotic, A., Wei, S., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A.,

- Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E. D. C., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium (arXiv:1706.08500). *arXiv*. <https://doi.org/10.48550/arXiv.1706.08500>
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *arXiv*. <https://doi.org/10.48550/arXiv.2006.11239>
- Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., & Salimans, T. (2021). Cascaded diffusion models for high fidelity image generation. *arXiv*. <https://doi.org/10.48550/arXiv.2106.15282>
- Holtz, R. D., Kovacs, W. D., & Sheahan, T. C. (1981). *An introduction to geotechnical engineering*. Prentice Hall.
- Holtz, R. D., Kovacs, W. D., & Sheahan, T. C. (2011). *An introduction to geotechnical engineering* (2nd edition). Pearson, Upper Saddle River, NJ.
- Hryciw, R., Ohm, H., & Zhou, J. (2013). Theoretical basis for optical granulometry by wavelet transformation. *Journal of Computing in Civil Engineering*, 29, 04014050. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000345](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000345)
- Hunt, R. E. (2005). *Geotechnical Engineering Investigation* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781420039153>.
- Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In *Proceedings of the 5th International Conference on Learning and Intelligent Optimization (LION 5)* (pp. 507–523). Springer. https://doi.org/10.1007/978-3-642-25566-3_40
- International Organization for Standardization. (2016). *ISO 17892-4:2016 — Geotechnical investigation and testing — Laboratory testing of soil — Part 4: Determination of particle size distribution*. ISO.
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*. <https://doi.org/10.48550/arXiv.1502.03167>
- Iwashita, M., Otsuka, Y., & Katoh, M. (2022). Predicting the particle size distribution of fine-grained and sandy soils using deep learning for classifying recovered soils separated from tsunami deposits. *Journal of Material Cycles and Waste Management*, 24(4), 1304–1316. <https://doi.org/10.1007/s10163-022-01410-4>

- Joy, T. T., Rana, S., Gupta, S., & Venkatesh, S. (2016). Hyperparameter tuning for big data using Bayesian optimisation. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 2574–2579). IEEE. <https://doi.org/10.1109/ICPR.2016.7900023>
- Jung, J.-H., Lee, S.-J., & Kim, H.-S. (2022). Estimation of average grain size from microstructure image using a convolutional neural network. *Materials*, 15(19), 6954. <https://doi.org/10.3390/ma15196954>
- Keivan Ekbatani, H., Pujol, O., & Segui, S. (2017). Synthetic data generation for deep learning in counting pedestrians. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods* (pp. 318–323). <https://doi.org/10.5220/0006119203180323>
- Keras Team. (2025). ReduceLROnPlateau [Documentation]. Keras. https://keras.io/api/callbacks/reduce_lr_on_plateau/
- Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv Preprint arXiv:1412.6980*.
- Kittaneh, O. A., Khan, M. A. U., Akbar, M., & Bayoud, H. A. (2016). General Average Entropy: A New Uncertainty Measure with Application to Image Segmentation. *The American Statistician*, 70(1), 18–24. <https://doi.org/10.1080/00031305.2015.1089788>
- Kothai, R., Prabakaran, N., Srinivasa Murthy, Y. V., Reddy Cenkeramaddi, L., & Kakani, V. (2024). Pavement distress detection, classification, and analysis using machine learning algorithms: A survey. *IEEE Access*, 12, 126943–126960. <https://doi.org/10.1109/ACCESS.2024.3455093>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25). Curran Associates, Inc.
- Lamba, S., Baliyan, A., & Kukreja, V. (2022). GAN-based image augmentation for increased CNN performance in paddy leaf disease classification. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2054–2059). IEEE. <https://doi.org/10.1109/ICACITE53722.2022.9823799>
- Lang, N., Irniger, A., Rozniak, A., Hunziker, R., Wegner, J. D., & Schindler, K. (2021). GRAINet: Mapping grain size distributions in river beds from UAV images with convolutional neural networks. *Hydrology and Earth System Sciences*, 25(5), 2567–2597. <https://doi.org/10.5194/hess-25-2567-2021>

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323. <https://doi.org/10.1109/5.726791>
- Lewinson, E. (2023). A comprehensive overview of regression evaluation metrics. *NVIDIA Technical Blog*.
- Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., & Stoica, I. (2018). Tune: A research platform for distributed model selection and training. *arXiv*. <https://doi.org/10.48550/arXiv.1807.05118>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: Common objects in context (arXiv:1405.0312). *arXiv*. <https://doi.org/10.48550/arXiv.1405.0312>
- Liu, H., Yang, C., Li, A., Huang, S., Feng, X., Ruan, Z., & Ge, Y. (2023). Deep domain adaptation for pavement crack detection. *arXiv*. <https://doi.org/10.48550/arXiv.2111.10101>
- Liu, Q., & Tran, H. (1996). Comparing systems: Validation of FragScan, WipFrag, and Split. In A. Franklin & T. Katsabanis (Eds.), *Measurement of blast fragmentation* (pp. 315–330). Routledge. <https://doi.org/10.1201/9780203747919-23>
- Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3), 1224. <https://doi.org/10.3390/su13031224>
- Loshchilov, I., & Hutter, F. (2016). CMA-ES for hyperparameter optimization of deep neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.1604.07269>
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Maerz, N. H. (1996). Image sampling techniques and requirements for automated image analysis of rock fragmentation. In *Proceedings of the FRAGBLAST 5 Workshop on Measurement of Blast Fragmentation* (pp. 115–120). Montréal, Canada.
- Maerz, N. H., Palangio, T. C., & Franklin, J. A. (1996). WipFrag image-based granulometry system. In A. Franklin & T. Katsabanis (Eds.), *Measurement of blast fragmentation* (pp. 63–76). Routledge.

- Maerz, N. H., Palangio, T. C., & Franklin, J. A. (1996). WipFrag image based granulometry system. Proceedings of the FRAGBLAST 5 Workshop on *Measurement of Blast Fragmentation*, Montreal, Quebec, Canada, 23-24 Aug. 1996, pp. 91-99
- Maerz, N. H., Palangio, T. C., & Franklin, J. A. (2018). WipFrag image based granulometry system. In *Measurement of blast fragmentation* (pp. 91–99). Routledge.
- Mahesh, B. (2019). Machine learning algorithms—A review. *International Journal of Science and Research*, 9(1). <https://doi.org/10.21275/ART20203995>
- Maktabdar Oghaz, M., Argyriou, V., & Remagnino, P. (2019). Learning how to analyse crowd behaviour using synthetic data. <https://doi.org/10.1145/3328756.3328773>
- Man, K., & Chahl, J. (2022). A review of synthetic image data and its use in computer vision. *Journal of Imaging*, 8(11), 310. <https://doi.org/10.3390/jimaging8110310>
- Manashti, J., Duhaime, F., Toews, M., & Pirnia, P. (2022). Predicting the particle size distributions of spherical particle sets from synthetic images: A comparison of nine classic image features. *OSF Preprints*. <https://doi.org/10.31224/osf.io/phjkm>
- Manashti, J., Duhaime, F., Toews, M., & Pirnia, P. (2022). Predicting the particle size distributions of spherical particle sets from synthetic images: A comparison of 9 classic image features [Preprint]. <https://doi.org/10.31224/osf.io/phjkm>
- Manashti, J., Pirnia, P., Manashty, A., Ujan, S., Toews, M., & Duhaime, F. (2023). PSDNet: Determination of particle size distributions using synthetic soil images and convolutional neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.2303.04269>
- Manashti, M. (2022). *Grain size analyses of granular material based on image analysis techniques and machine learning* [Ph.D thesis, École de technologie supérieure] https://espace.etsmtl.ca/id/eprint/3076/1/MANASHTI_Mohammadjavad.pdf
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2020). Image segmentation using deep learning: A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2001.05566>
- Ministère des Transports et de la Mobilité durable. (2024). Méthode d'essai LC 21-040 : Analyse granulométrique. *Recueil des méthodes d'essai LC*, Publication du Québec, Québec, QC.
- Močkus, J. (1975). On Bayesian methods for seeking the extremum. In G. I. Marchuk (Ed.), *Optimization Techniques IFIP Technical Conference, Novosibirsk, July 1–7, 1974* (pp. 400–404). Springer. https://doi.org/10.1007/3-540-07165-2_55

- Nanda, S., & Pal, B. K. (2020). Analysis of blast fragmentation using WipFrag. *International Journal of Innovative Science and Research Technology*, 5(6), 1561–1566. <https://doi.org/10.38124/IJISRT20JUN1086>
- Nikolenko, S. I. (2021). *Synthetic data for deep learning* (Vol. 174). Springer. <https://doi.org/10.1007/978-3-030-75178-4>
- Northcutt, C. G., Jiang, L., & Chuang, I. L. (2022). Confident learning: Estimating uncertainty in dataset labels. *arXiv*. <https://doi.org/10.48550/arXiv.1911.00068>
- Nowruzi, F. E., Kapoor, P., Kolhatkar, D., Hassanat, F. A., Laganriere, R., & Rebut, J. (2019). How much real data do we actually need: Analyzing object detection performance using synthetic and real data. *arXiv*. <https://arxiv.org/abs/1907.07061>
- Noy, M. J. (2012). Automated rock fragmentation measurement with close-range digital photogrammetry. In *Measurement and analysis of blast fragmentation*. CRC Press.
- Olivier, J., & Aldrich, C. (2021). Underflow particle size estimation of hydrocyclones by use of transfer learning with convolutional neural networks. *IFAC-PapersOnLine*, 54(11), 85–90. <https://doi.org/10.1016/j.ifacol.2021.10.055>
- Olivier, L. E., Maritz, M. G., & Craig, I. K. (2020). Estimating ore particle size distribution using a deep convolutional neural network. *IFAC-PapersOnLine*, 53(2), 12038–12043. <https://doi.org/10.1016/j.ifacol.2020.12.740>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Ottino, J. M., & Khakhar, D. V. (2000). Mixing and segregation of granular materials. *Annual Review of Fluid Mechanics*, 32, 55–91. <https://doi.org/10.1146/annurev.fluid.32.1.55>
- Ouchterlony, F., Nyberg, U., Bergman, P., & Esen, S. (2007). Monitoring the blast fragmentation at Boliden Mineral's Aitik copper mine. In P. Moser et al. (Eds.), *Proceedings of the EFEE World Conference on Explosives and Blasting* (pp. 47–62). European Federation of Explosives Engineers.
- Parent, E. J., Parent, S.-É., & Parent, L. E. (2021). Determining soil particle-size distribution from infrared spectra using machine learning predictions: Methodology and modeling. *PLOS ONE*, 16(7), e0233242. <https://doi.org/10.1371/journal.pone.0233242>
- Parker, H. W. (1996). Geotechnical investigations. In J. O. Bickel, T. R. Kuesel, & E. H. King (Eds.), *Tunnel Engineering Handbook* (pp. 46–79). Springer US. https://doi.org/10.1007/978-1-4613-0449-4_4

- Patel, V. M., Gopalan, R., Li, R., & Chellappa, R. (2015). Visual domain adaptation: A survey of recent advances. *IEEE Signal Processing Magazine*, 32(3), 53–69. <https://doi.org/10.1109/MSP.2014.2347059>
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621. <https://arxiv.org/abs/1712.04621>
- Picon, A., San-Emeterio, M. G., Bereciartua-Perez, A., Klukas, C., Eggers, T., & Navarra-Mestre, R. (2022). Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. *Computers and Electronics in Agriculture*, 194, 106719. <https://doi.org/10.1016/j.compag.2022.106719>
- Picon, A., San-Emeterio, M. G., Bereciartua-Perez, A., Klukas, C., Eggers, T., & Navarra-Mestre, R. (2022). Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. *Computers and Electronics in Agriculture*, 194, 106719. <https://doi.org/10.1016/j.compag.2022.106719>
- Pirnia, P., Duhaime, F., & Manashti, J. (2018). Machine learning algorithms for applications in geotechnical engineering. In *Proceedings of GeoEdmonton 2018*. https://www.researchgate.net/publication/326305480_Machine_learning_algorithms_for_applications_in_geotechnical_engineering
- Polakowski, C., Ryżak, M., Sochan, A., Beczek, M., Mazur, R., & Bieganski, A. (2021). Particle size distribution of various soil materials measured by laser diffraction—The problem of reproducibility. *Minerals*, 11(5), Article 5. <https://doi.org/10.3390/min11050465>
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63. <https://doi.org/10.9735/2229-3981>
- Purinton, B., & Bookhagen, B. (2019). Introducing PebbleCounts: A grain-sizing tool for photo surveys of dynamic gravel-bed rivers. *Earth Surface Dynamics*, 7, 859–877. <https://doi.org/10.5194/esurf-7-859-2019>
- Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Deep convolutional neural network for age estimation based on VGG-Face model. *arXiv*. <https://doi.org/10.48550/arXiv.1709.01664>
- Raina, A. K., Choudhary, P. B., Ramulu, M., Chakraborty, A. K., Dudhankar, A. S., Udpikar, V., Ghatpande, N., & Misra, D. D. (2002). FRAGALYST—An indigenous digital image analysis system for grain size measurements in mines. *Journal of the Geological Society of India*, 59(6). <https://indianjournalofmarketing.com/index.php/jgsi/article/view/83742>

- Reddy, B. H., Vickram, A. S., & Karthikeyan, P. R. (2024). Classification of fire and smoke images using random forest algorithm in comparison with decision tree to measure accuracy, precision, recall, F-score. *AIP Conference Proceedings*, 2853(1). <https://pubs.aip.org/aip/acp/article-abstract/2853/1/020077/3290172>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Richter, S. R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for data: Ground truth from computer games (arXiv:1608.02192). *arXiv*. <https://doi.org/10.48550/arXiv.1608.02192>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *arXiv*. <https://doi.org/10.48550/arXiv.1505.04597>
- Ros, G., Sellart, L., Materzynska, J., Vazquez, D., & Lopez, A. M. (2016). The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3234–3243). <https://doi.org/10.1109/CVPR.2016.352>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sagues-Tanco, R., Benages-Pardo, L., López-Nicolás, G., & Llorente, S. (2020). Fast synthetic dataset for kitchen object segmentation in deep learning. *IEEE Access*, 8, 220496–220506. <https://doi.org/10.1109/ACCESS.2020.3043256>
- Salaudeen, T. (2023). *Détection et classification d'objets sur l'imagerie satellite: Utilisation de la vision par ordinateur et de l'IA*. Éditions Notre Savoir.
- Sanchidrián, J. A., Segarra, P., Ouchterlony, F., & López, L. M. (2009). On the accuracy of fragment size measurement by image analysis in combination with some distribution functions. *Rock Mechanics and Rock Engineering*, 42(1), 95–116. <https://doi.org/10.1007/s00603-007-0161-8>
- Sarkar, A. (2021, 25 juin). Understanding depthwise separable convolutions and the efficiency of MobileNets. *Towards Data Science*. <https://medium.com/towards-data-science/understanding-depthwise-separable-convolutions-and-the-efficiency-of-mobilenets-6de3d6b62503>

- Schleifer, J., & Tessier, B. (2018). FRAGSCAN: A tool to measure fragmentation of blasted rock. In *Proceedings of the 15th International Symposium on Rock Fragmentation by Blasting* (pp. 73–78). CRC Press. <https://doi.org/10.1201/9780203747919-12>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication* (p. 53). Urbana, IL: The University of Illinois Press.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., & Webb, R. (2016). Learning from simulated and unsupervised images through adversarial training. arXiv preprint arXiv:1612.07828.
- Shrivastava, S., Deb, D., & Bhattacharjee, S. (2022). Prediction of particle size distribution curves of dump materials using convolutional neural networks. *Rock Mechanics and Rock Engineering*, 55(1), 471–479. <https://doi.org/10.1007/s00603-021-02657-y>
- Silla, C. N., & Freitas, A. A. (2011). A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22(1–2), 31–72. <https://doi.org/10.1007/s10618-010-0175-9>
- Simo Tenekam, H., Duhaime, F., Dubé, J.-S., & Toews, M. F. (2025). *Segmentation of granular material images using synthetic data and deep learning*. Manuscript submitted for publication to *Applied Computing & Geosciences*.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv*. <https://doi.org/10.48550/arXiv.1409.1556>
- Smilauer, V., Angelidakis, V., Catalano, E., Caulk, R., Chareyre, B., Chèvremont, W., Dorofeenko, S., Duriez, J., Dyck, N., Elias, J., Er, B., Eulitz, A., Gladky, A., Guo, N., Jakob, C., Francois Kneib, & Kozicki, J., ... Yuan, C. (2021). Yade documentation (3rd Version) [Logiciel]. *Zenodo*. <https://doi.org/10.5281/zenodo.5705394>
- Song, J., Meng, C., & Ermon, S. (2022). Denoising diffusion implicit models. *arXiv*. <https://doi.org/10.48550/arXiv.2010.02502>
- Soranzo, E., Guardiani, C., & Wu, W. (2025). Convolutional neural network prediction of the particle size distribution of soil from close range images. *Soils and Foundations*, 65(1), 101575. <https://doi.org/10.1016/j.sandf.2025.101575>

- Souza, J. C. de, Silva, A. C. S. da, & Rocha, S. S. (2018). Analysis of blasting rocks prediction and rock fragmentation results using Split-Desktop software. *Tecnologia em Metalurgia, Materiais e Mineração*. <https://tecnologiammm.com.br/article/10.4322/2176-1523.1234/pdf/tmm-15-1-22.pdf>
- Stokes, G. G. (1851). On the effect of the internal friction of fluids on the motion of pendulums. Wellcome Collection. <https://wellcomecollection.org/works/hcy5wu4>
- Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. *arXiv*. <https://doi.org/10.48550/arXiv.1707.02968>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html
- Tammina, S. (2019). Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications*, 9(10), p9420. <https://doi.org/10.29322/IJSRP.9.10.2019.p9420>
- Tariel, V. (2008). Conceptualization of seeded region growing by pixels aggregation. Part 4: Simple, generic and robust extraction of grains in granular materials obtained by X-ray tomography. *arXiv*. <https://doi.org/10.48550/arXiv.0806.3939>
- Temimi, M. (2020). *Création d'images de matériaux granulaires avec la plateforme de développement de jeux Unity et prédiction de la granulométrie à l'aide de textures et de réseaux de neurones* [Mémoire de maîtrise, École de technologie supérieure]. <https://espace.etsmtl.ca/id/eprint/2681/>
- Terzaghi, K., Peck, R. B., & Mesri, G. (1996). *Soil mechanics in engineering practice* (3rd ed., pp. 1–37). John Wiley & Sons.
- Tian, L. (2024). What is a soil sieve test? *LinkedIn*. <https://www.linkedin.com/pulse/what-soil-sieve-test-lan-tian-libcc/>
- Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. *arXiv*. <https://doi.org/10.48550/arXiv.1703.06907>
- Tong, C.-X., Burton, G. J., Zhang, S., & Sheng, D. (2018). A simple particle-size distribution model for granular materials. *Canadian Geotechnical Journal*, 55(2), 246–257. <https://doi.org/10.1139/cgj-2017-0098>

- Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., & Birchfield, S. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *CVPR 2018 Workshop on Autonomous Driving* (pp. 969–977). <https://doi.org/10.48550/arXiv.1804.06516>
- Unity Technologies. (2020). Unity-Technologies/com.unity.perception [C#]. Unity Technologies. <https://github.com/Unity-Technologies/com.unity.perception>
- Vasam, R., & Nayak, P. (2019). Instance segmentation on real-time object detection using Mask R-CNN. *International Journal of Engineering and Advanced Technology*, 9, 2932–2935. <https://doi.org/10.35940/ijeat.A1265.109119>
- Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). CovidGAN: Data augmentation using auxiliary classifier GAN for improved COVID-19 detection. *IEEE Access*, 8, 91916–91923. <https://ieeexplore.ieee.org/abstract/document/9093842/>
- Wei, W., Xu, X., Hu, G., Shao, Y., & Wang, Q. (2024). Deep learning and histogram-based grain size analysis of images. *Sensors*, 24(15), 4923. <https://doi.org/10.3390/s24154923>
- WipWare Inc. (n.d.). WipFrag fragmentation analysis software. <https://wipware.com/products/wipfrag/>
- Wolk, K., & Tatara, M. S. (2024). A review of semantic segmentation and instance segmentation techniques in forestry using LiDAR and imagery data. *Electronics*, 13(20), 4139. <https://doi.org/10.3390/electronics13204139>
- Wong, M. Z., Kunii, K., Baylis, M., Ong, W. H., Kroupa, P., & Koller, S. (2019). Synthetic dataset generation for object-to-model deep learning in industrial applications. *PeerJ Computer Science*, 5, e222. <https://doi.org/10.7717/peerj-cs.222>
- Wu, Y., Zhou, Y., Saveriades, G., Agaian, S., Noonan, J. P., & Natarajan, P. (2013). Local Shannon entropy measure with statistical tests for image randomness. *Information Sciences*, 222, 323–342. <https://doi.org/10.1016/j.ins.2012.07.049>
- Xiao, B., Miao, S., & Gao, Q. (2022). Quantifying particle size and size distribution of mine tailings through deep learning approach of autoencoders. *Powder Technology*, 397, 117088. <https://doi.org/10.1016/j.powtec.2021.117088>
- Xiao, T., Zhang, J., Yang, K., Peng, Y., & Zhang, Z. (2014). Error-driven incremental learning in deep convolutional neural network for large-scale image classification. In *Proceedings of the 2014 ACM Conference on Multimedia* (pp. 177–186). <https://doi.org/10.1145/2647868.2654926>

- Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. In *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems* (pp. 79–91). <https://doi.org/10.18653/v1/2020.eval4nlp-1.9>
- Yaghoobi, H., Mansouri, H., Ebrahimi Farsangi, M. A., & Nezamabadi-Pour, H. (2019). Determining the fragmented rock size distribution using textural feature extraction of images. *Powder Technology*, 342, 630–641. <https://doi.org/10.1016/j.powtec.2018.10.006>
- Yamada, T., & Di Santo, S. (2022). Instance segmentation of piled rock particles based on Mask R-CNN. In *IGARSS 2022 – 2022 IEEE International Geoscience and Remote Sensing Symposium*. <https://doi.org/10.1109/IGARSS46834.2022.9883183>
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, 27. <https://doi.org/10.48550/arXiv.1402.1869>
- Zafari, S., Murashkina, M., Eerola, T., Sampo, J., Kälviäinen, H., & Haario, H. (2019). Resolving overlapping convex objects in silhouette images by concavity analysis and Gaussian process. *arXiv*. <https://doi.org/10.48550/arXiv.1906.01049>
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. *arXiv*. <https://doi.org/10.48550/arXiv.1611.03530>
- Zhang, H., Fritts, J. E., & Goldman, S. A. (2003). Entropy-based objective evaluation method for image segmentation. In *Proceedings of SPIE*, Vol. 5307, pp. 38–49. <https://doi.org/10.1117/12.527167>.
- Zhang, Z., Yin, X., & Yan, Z. (2022). Rapid data annotation for sand-like granular instance segmentation using mask-RCNN. *Automation in Construction*, 133, 103994. <https://doi.org/10.1016/j.autcon.2021.103994>
- Zhou, S., Liang, W., Li, J., & Kim, J.-U. (2018). Improved VGG model for road traffic sign recognition. *Computers, Materials & Continua*, 57(1), 11–24. https://cdn.techscience.cn/files/cmc/2018/v57n1/20181019033013_23617.pdf
- Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.