Détection de défauts dans les scans sectoriels avec des architectures d'apprentissage profond avancées

par

Éloi LOMBARD

MÉMOIRE PAR ARTICLE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE AVEC MÉMOIRE EN GÉNIE MÉCANIQUE
M. Sc. A.

MONTRÉAL, LE 28 DÉCEMBRE 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE:

M. Pierre Bélanger, directeur de mémoire
Département de génie mécanique à l'École de technologie supérieure

M. Martin Viens, président du jury
Département de génie mécanique à l'École de technologie supérieure

M. Matthew Toews, examinateur externe
Département de génie des systèmes à l'École de technologie supérieure

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

LE 10 DÉCEMBRE 2025

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma profonde gratitude envers mon directeur de recherche, le professeur Pierre Bélanger, pour son encadrement rigoureux, sa disponibilité et ses précieux conseils tout au long de ce projet. Sa vision scientifique et son expertise en contrôle non destructif ont été déterminantes pour mener à bien cette recherche.

Je remercie chaleureusement l'ensemble des membres du laboratoire pour leur soutien, leurs encouragements et les nombreux échanges qui ont enrichi ce travail. L'environnement collaboratif et stimulant qu'ils ont su créer a grandement contribué à l'avancement de mes travaux.

Je souhaite également remercier l'entreprise partenaire pour sa collaboration et pour avoir mis à disposition les données expérimentales essentielles à cette étude. Leur expertise industrielle et leur confiance ont permis de donner à cette recherche une dimension pratique et concrète.

Mes remerciements vont également aux membres du jury pour avoir accepté d'évaluer ce mémoire et pour le temps consacré à sa lecture.

Enfin, je tiens à remercier ma famille et mes proches pour leur soutien inconditionnel, leur patience et leurs encouragements constants qui m'ont accompagné tout au long de ce parcours académique.

Détection de défauts dans les scans sectoriels avec des architectures d'apprentissage profond avancées

Éloi LOMBARD

RÉSUMÉ

Le contrôle par ultrasons en multiéléments (PAUT) s'est largement imposé en contrôle non destructif en raison de sa précision et de son efficacité. Cependant, l'interprétation des données PAUT demeure fortement dépendante de l'expertise de l'inspecteur, et la présence de caractéristiques géométriques dans les matériaux génère des artefacts qui compliquent l'analyse et augmentent le temps d'interprétation. Bien que l'intelligence artificielle ait démontré de très bonnes performances en détection d'objets dans divers domaines, son application aux ultrasons reste relativement limitée en raison des contraintes de confidentialité des données.

Cette étude explore l'application de l'apprentissage profond pour la détection automatisée de défauts dans des cordons de soudures à l'aide de données PAUT. Deux architectures sont comparées : un Faster R-CNN amélioré et les dernières versions de YOLO (v5 et v8). Pour explorer l'influence de l'information contextuelle sur les performances des modèles, trois configurations de données distinctes sont analysées : des images PAUT brutes servant de référence, des images PAUTs avec une superposition schématique de la géométrie (overlay) indiquant la géométrie de la soudure, et des images PAUTs prétraitées par une soustraction de la médiane de la géométrie pour supprimer les artefacts structurels et mettre en évidence les régions défectueuses.

L'entraînement et la validation se portent sur des spécimens de soudure industriels contenant divers types de défauts incluant des manques de fusion, de la porosité et des fissures. L'intégration de composants tels que ROI Align, EfficientNet-B5 et un réseau pyramidal de caractéristiques, combinée à une optimisation des boîtes d'ancrage par K-means, a permis au Faster R-CNN d'atteindre les meilleures performances avec une précision moyenne (mAP) de 58.05% lorsqu'elle est combinée à la soustraction de géométrie médiane, surpassant les modèles YOLO (avec 57,03% au maximum). La soustraction de géométrie médiane permet une amélioration du mAP de 5 à 7 points de pourcentage tout en réduisant les faux positifs. À l'inverse, l'ajout de l'overlay n'améliore pas les performances et entrave parfois la détection, malgré l'alignement avec les pratiques des inspecteurs.

Mots-clés: Contrôle par ultrasons en multi-éléments (PAUT), Réseaux de neurones convolutifs (CNN), Faster R-CNN, YOLO, Détection de défauts, Classification

Classification and characterisation of defects in PAUT using CNN

Éloi LOMBARD

ABSTRACT

Phased Array Ultrasonic Testing (PAUT) has become a widely adopted technique in non-destructive testing (NDT) due to its precision and efficiency. However, the interpretation of PAUT data remains highly dependent on the inspector's expertise, and the presence of geometric features within materials generates artifacts that complicate analysis and increase interpretation time. Although artificial intelligence has demonstrated excellent performance in object detection across various fields, its application to ultrasonic data remains relatively limited, mainly due to data confidentiality constraints.

This study investigates the use of deep learning for the automated detection of weld defects using PAUT data. Two architectures are compared : an enhanced Faster R-CNN and the latest versions of YOLO (v5 and v8). To explore the influence of contextual information on model performance, three distinct data configurations are analyzed : raw PAUT images used as a reference, PAUT images with a schematic overlay indicating weld geometry, and PAUT images preprocessed by median geometry subtraction to remove structural artifacts and highlight defect regions.

Training and validation are conducted on industrial weld specimens containing various defect types, including lack of fusion, porosity, and cracks. The integration of components such as ROI Align, EfficientNet-B5, and a feature pyramid network, combined with K-means–based anchor box optimization, enables the enhanced Faster R-CNN to achieve the best performance, reaching a mean Average Precision (mAP) of 58.05% when combined with median geometry subtraction, outperforming YOLO models (up to 57.03%). Median geometry subtraction improves mAP by 5 to 7 percentage points while reducing false positives. Conversely, the addition of the overlay does not improve performance and sometimes hinders detection, despite its alignment with standard inspection practices.

Keywords: Phased Array Ultrasonic Testing (PAUT), Convolutional Neural Networks (CNN), Faster R-CNN, YOLO, Flaw Detection, Classification

TABLE DES MATIÈRES

Page

LISTE DES FIGURES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

CND/NDT	Contrôle Non Destructif (*Non-Destructive Testing*)
PAUT	Contrôle par ultrasons multiéléments (*Phased Array Ultrasonic Testing*)
FMC	Capture de matrice complète (*Full Matrix Capture*)
TFM	Méthode de focalisation en tout point (*Total Focusing Method*)
IA / AI	Intelligence Artificielle (*Artificial Intelligence*)
ML	Apprentissage automatique (*Machine Learning*)
CNN	Réseau de neurone convolutif (*Convolutional Neural Network*)
ReLU	Unité de Rectificaton linéaire (*Rectified Linear Unit*)
SGD	Descente de gradient stochastique (*Stochastic Gradient Descent*)
ADAM	Estimation adaptive du moment (*adaptive moment estimation*)
AdaGrad	Descente du gradient adaptative (*Adaptive gradient descent*)
ROI	Région d'intérêt (*Region Of Interest*)
IoU	Intersection sur l'Union (*Intersection over Union*)
mAP	Précision Moyenne (*mean Average Precision*)
LoF	Manque de fusion (*Lack of Fusion*)

LISTE DES SYMBOLES ET UNITÉS DE MESURE

m	mètre
mm	millimètre
u	Déplacement de la particule (m)
t	Temps (s)
V	Vitesse de propagation de l'onde (m/s)
c	Vitesse de propagation de l'onde (m/s)
ρ	Densité massique (kg/m^3)
A	Amplitude du signal
T	Tenseur de contrainte
S	Tenseur de déformation
c_{ijkl}	Composantes du tenseur de rigidité
δ	Symbole de Kronecker
E	Module de Young (Pa)
ν	Coefficient de Poisson
∇	Nabla
I	Intensité acoustique (W/m^2)
α	Coefficient d'atténuation du matériau
f	Fréquence (Hz)
k	Nombre d'onde (1/m)
Z	Impédance acoustique
σ	Écart-type
Hz	Hertz
dB	Décibel

INTRODUCTION

Le contrôle non destructif (CND) vise à inspecter l'intégrité des matériaux sans compromettre leur utilisation future. Ces techniques sont essentielles dans des secteurs critiques tels que l'énergie nucléaire, l'aérospatiale et l'industrie pétrolière et gazière, où la défaillance d'une structure peut avoir d'importantes conséquences. Parmi les différentes méthodes disponibles, le contrôle par ultrasons occupe une place prépondérante en raison de son efficacité, de sa rapidité et de sa facilité d'utilisation.

Le contrôle par ultrasons multiéléments ou Phased Array Ultrasonic Testing en anglais (PAUT) s'est imposé dans l'industrie, particulièrement pour l'inspection des soudures. Contrairement aux méthodes conventionnelles à élément unique, le PAUT utilise des sondes composées de multiples éléments piézoélectriques (ici 32 ou 64) dont les émissions sont précisément synchronisées. Cette technique permet d'orienter et de focaliser le faisceau ultrasonore électroniquement, offrant ainsi une imagerie haute résolution des défauts tout en réduisant considérablement le temps d'inspection. La capacité à inspecter rapidement des volumes complexes et à générer des images sectorielles détaillées en fait un outil particulièrement adapté à l'inspection sur le terrain.

Toutefois, si l'acquisition d'images PAUT se fait sans difficulté, leur interprétation nécessite une expertise considérable. En effet, pour comprendre les images ultrasonores, il faut connaître la réflexion, la diffraction et la réfraction de l'onde. La présence de caractéristiques géométriques liées, dans le cas de cette étude, aux cordons de soudure introduit des artefacts qui se superposent aux signatures des défauts, rendant leur identification particulièrement ardue. Une mauvaise compréhension de ces phénomènes et de la manière dont ils se manifestent dans l'image peuvent aboutir à des interprétations erronées. En conséquence, les résultats de l'analyse sont dépendants de nombreux paramètres : l'expertise, la fatigue de l'inspecteur, les conditions de l'analyse (sur le terrain ou non), etc. Cette variabilité inter-inspecteurs représente un enjeu majeur pour la fiabilité et la reproductibilité des résultats d'inspection (Kang *et al.* (2023)).

Au cours de la dernière décennie, les réseaux de neurones convolutifs (CNN) ont révolutionné le domaine de la vision par ordinateur, démontrant des capacités exceptionnelles dans la détection et la classification d'objets. L'introduction d'architectures avancées telles que ResNet, EfficientNet, ainsi que des détecteurs d'objets comme Faster R-CNN et YOLO, a considérablement amélioré la précision et l'efficacité de la reconnaissance d'images. Ces progrès ouvrent des perspectives prometteuses pour l'automatisation de l'analyse des images PAUT. Néanmoins, l'application de l'intelligence artificielle au contrôle par ultrasons demeure relativement peu explorée, principalement en raison de la confidentialité des données industrielles et du manque de jeux de données annotés suffisamment volumineux pour entraîner des modèles robustes.

L'objectif de cette étude est de développer un système de détection automatisée de défauts PAUT pour assister les inspecteurs dans leur analyse. Ce travail s'intéresse à des architectures de réseau de neurones complexes et explore l'impact de stratégies de prétraitement contextuel sur la capacité des modèles à distinguer les défauts présents dans des spécimens de cordons de soudure.

CHAPITRE 1

REVUE DE LITTÉRATURE

1.1 Introduction aux ondes ultrasonores

Les ondes ultrasonores sont des ondes mécaniques de longueur d'onde inférieure au domaine de l'audible soit au-dessus de 20kHz. Découvertes durant le XIX^{e} siècle, le développement de leur contexte mathématique a permis leur utilisation dans de nombreux domaines (comme avec les ouvrages de Shull (2002), Cheeke (2002)...).

Spécifiquement, elles ont ouvert une nouvelle voie dans le contrôle non destructif (CND). Les ultrasons peuvent en effet obtenir les informations telles que l'épaisseur d'un matériau mais également les défauts présents dans celui-ci, et donc permettre la vérification de l'état d'un matériau inspecté.

Afin de comprendre plus en détail le fonctionnement des techniques d'imagerie utilisées en CND, il faut d'abord revenir aux équations qui régissent les ondes.

1.1.1 Équations de propagation

Si l'on se place selon une dimension en considérant une onde se propageant selon une direction x (dans une corde par exemple), on peut obtenir une équation la décrivant en combinant la deuxième loi de Newton avec la loi de Hook. Il faut pour cela considérer que la corde est un ensemble de plusieurs systèmes de masse-ressort (Cheeke (2002)). On obtient alors l'équation suivante, appelée équation de d'Alembert ou équation des ondes :

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2} \tag{1.1}$$

Où u est le déplacement de particules et c est la célérité de l'onde . Cette équation a pour solution générale l'expression suivante :

$$u(x,t) = F(x - ct) + G(x + ct) \tag{1.2}$$

Avec F et G deux fonctions arbitraires, correspondant à la propagation de l'onde dans le sens des x positifs (pour F) et celle dans le sens des x négatifs (pour G). Si l'on prend une excitation sinusoïdale selon les x positifs, on peut obtenir la solution suivante :

$$u(x,t) = Acos(\omega t - kx) \tag{1.3}$$

Avec A l'amplitude de l'excitation, ω sa vitesse angulaire et k son nombre d'onde angulaire ($k = \frac{2\pi}{\lambda_u}$ avec λ_u la longueur d'onde de la perturbation).

Si l'on considère maintenant un solide homogène et isotrope, la propagation d'une onde en une dimension peut se généraliser dans l'espace en trois dimensions. On obtient l'équation suivante avec la deuxième loi de Newton (Cheeke (2002)) :

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \sum_j \frac{\partial T_{ij}}{\partial x_j} \tag{1.4}$$

Avec T le tenseur de contrainte et ρ la masse volumique du solide.
On utilise alors la loi de Hooke généralisée :

$$T_{ij} = \sum_{kl} c_{ijkl} S_{kl} \tag{1.5}$$

Avec c_{ijkl} les composantes du tenseur de rigidité (ou d'élasticité). On peut également écrire l'équation comme suit (le solide étant isotrope) :

$$T_{ij} = (c_{11} - c_{44}) \sum_k S_{kk} \delta_{ij} + c_{44} S_{ij} \tag{1.6}$$

Où δ_{ij} est le symbole de Kronecker, S est le tenseur de déformation dont les coefficients peuvent s'obtenir via la formule :

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{1.7}$$

Et

$$\begin{cases} c_{11} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \\ c_{44} = \frac{E}{2(1+\nu)} \end{cases} \tag{1.8}$$

Avec E le module d'Young et ν le coefficient de Poisson du matériau

On peut donc réinjecter la formule des coefficients du tenseur des contraintes (1.6) dans celle obtenue avec la deuxième loi de Newton (1.4) :

$$\rho\frac{\partial^2 u_i}{\partial t^2} = (c_{11} - c_{44})\frac{\partial}{\partial x_i}\left(\sum_k \frac{\partial u_k}{\partial x_k}\right) + \frac{c_{44}}{2}\left(\sum_j \frac{\partial^2 u_i}{\partial x_j^2} + \frac{\partial}{\partial x_i}\left(\sum_j \frac{\partial u_j}{\partial x_j}\right)\right) \tag{1.9}$$

Ce que l'on peut réécrire sous forme vectorielle par :

$$\rho\frac{\partial^2 u}{\partial t^2} = \left(c_{11} - \frac{c_{44}}{2}\right)\vec{\nabla}\left(\vec{\nabla}\cdot\vec{u}\right) + \frac{c_{44}}{2}\nabla^2\vec{u} \tag{1.10}$$

Avec

$$\vec{\nabla} = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) \tag{1.11}$$

En écrivant que chaque vecteur peut s'écrire comme somme du rotationel d'un vecteur Ψ et du gradient d'un scalaire ϕ :

$$\vec{u} = \vec{\nabla}\phi + \vec{\nabla}\times\vec{\Psi} \tag{1.12}$$

On obtient après substitution dans l'équation 1.9 :

$$\rho\frac{\partial^2\phi}{\partial t^2} + \rho\frac{\partial^2\left(\vec{\nabla}\times\vec{\Psi}\right)}{\partial t^2} = \left(c_{11} - \frac{c_{44}}{2}\right)\vec{\nabla}\left(\nabla^2\phi\right) + \frac{c_{44}}{2}\left(\Delta(\vec{\nabla}\phi) + \Delta(\vec{\nabla}\times\Psi\right) \tag{1.13}$$

On utilise l'identité d'Helmholtz et on obtient :

$$\vec{\nabla}\left(\rho\frac{\partial^2\phi}{\partial t^2} - c_{11}\nabla^2\phi\right) + \vec{\nabla}\times\left(\rho\frac{\partial^2\Psi}{\partial t^2} - \frac{c_{44}}{2}\nabla^2\Psi\right) = 0 \tag{1.14}$$

Ainsi, comme les deux termes sont respectivement un scalaire et un vecteur, ils doivent tous les deux être nuls :

$$\begin{cases} \rho\frac{\partial^2\phi}{\partial t^2} = c_{11}\nabla^2\phi \\ \rho\frac{\partial^2\Psi}{\partial t^2} = \frac{c_{44}}{2}\nabla^2\Psi \end{cases} \tag{1.15}$$

On a donc deux équations de propagation et on peut identifier les deux vitesses qui sont respectivement longitudinale et transversale comme suit :

$$\begin{cases} V_L = \sqrt{\frac{c_{11}}{\rho}} \\ V_T = \sqrt{\frac{c_{44}}{2\rho}} \end{cases} \tag{1.16}$$

En remplaçant, dans l'équation 1.16, c11 et c44 par leur définition donnée dans l'équation 1.8, on obtient :

$$\begin{cases} V_L = \sqrt{\frac{E(1-\nu)}{\rho(1+\nu)(1-2\nu)}} \\ V_T = \sqrt{\frac{E}{2\rho(1+\nu)}} \end{cases} \tag{1.17}$$

Dans les solides homogènes et isotropes, deux types d'ondes peuvent donc se propager : les ondes longitudinales et les ondes transversales. Comme montré à la figure 1.1, les ultrasons sont le résultat d'une combinaison de ces deux modes.

1.1.2 Phénomène d'atténuation

Le phénomène d'atténuation apparait lors de la propagation des ondes dans un matériau. Ce phénomène est dû à deux facteurs principaux. D'une part, l'absorption de l'onde par le solide est la transformation de son énergie en chaleur, qui dépend de la fréquence de l'onde et des propriétés intrinsèques du matériau. D'autre part, la dispersion résultante des irrégularités du

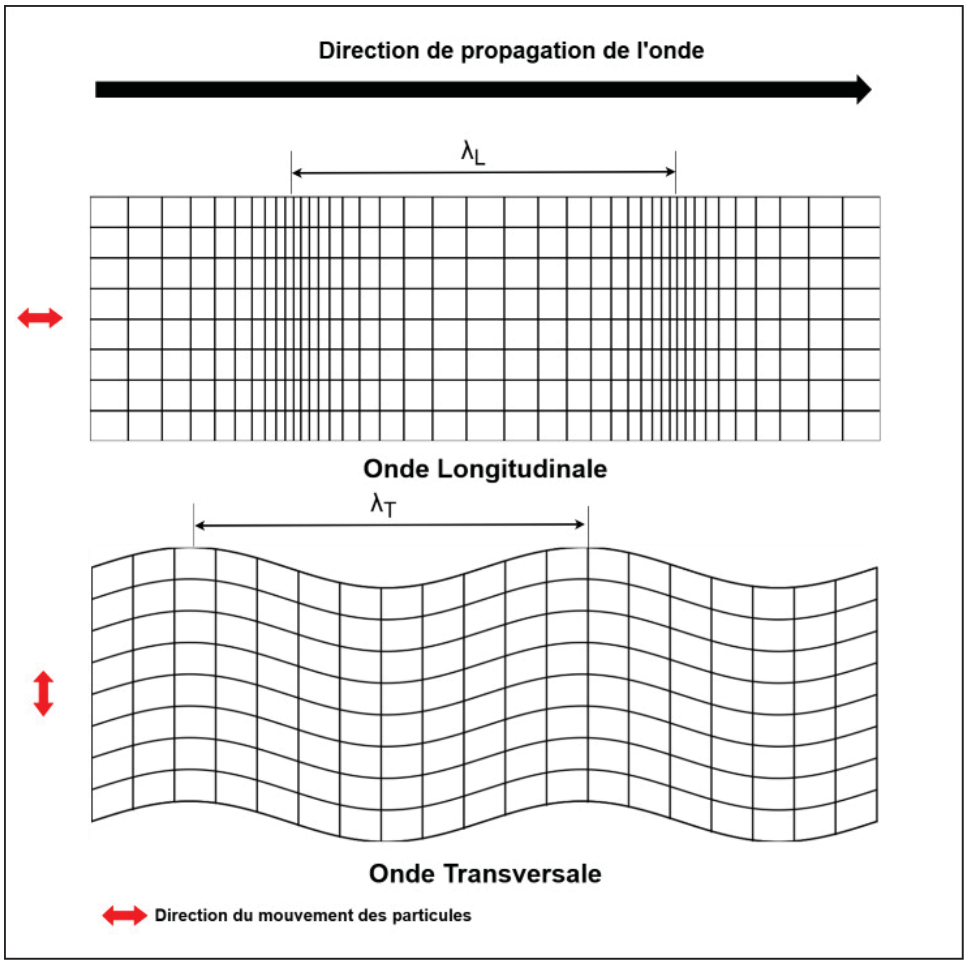


Figure 1.1 Propagation d'onde longitudinale et transversale

matériau, comme les impuretés, les joints de grains ou les fissures qui perturbent la propagation de l'onde et réduisent son intensité.

On peut exprimer l'atténuation grâce à l'intensité acoustique de l'onde, qui correspond au flux moyen d'énergie acoustique par unité de surface par unité de temps :

$$I(x) = I_0 e^{-2\alpha x} \tag{1.18}$$

Où I_0 est l'intensité initiale et α le coefficient d'atténuation du matériau. Dans le cas de l'absorption, à température ambiante et pour un matériau homogène et isotrope α est proportionnel

au carré de la fréquence f^2 (Cheeke (2002)). Ainsi, l'inspection de matériau avec une épaisseur élevée sera préférablement faite avec une sonde de fréquence plus faible et l'utilisation d'une sonde de fréquence plus élevée accentuera l'atténuation et donc donnera une pénétration plus faible mais une meilleure précision.

1.1.3 Comportement des ondes ultrasonores aux interfaces

Lors de la propagation de l'onde dans un solide, cette dernière peut rencontrer des défauts comme des fissures ou des porosités et également le fond de la pièce. Ces contacts avec un changement de milieu induisent des réflexions, des transmissions et des réfractions de l'onde. Cette section s'intéresse au comportement des ondes à ces interfaces.

1.1.3.1 Transmission et réflexion en incidence normale

Dans le cas simple où une onde arrive à la surface séparant deux matériaux avec un angle de 90°, il n'y a pas de réfraction et l'onde se transmet et se réfléchie (Fig 1.2). On peut associer à chaque milieu une impédance acoustique Z (en analogie avec les impédances électriques). Cette impédance est définie par :

$$Z = \rho V \tag{1.19}$$

Où ρ est la masse volumique du matériau et V la vitesse de propagation de l'onde.

On peut démontrer, avec les conditions de vitesse et de pression à l'interface (Shull (2002)) et avec la conservation de l'énergie, qu'avec T, la part d'énergie transmise, et R, la part d'énergie réfléchie, on a :

$$\begin{cases} T + R = 1 \\ T = \frac{4Z_1Z_2}{(Z_1+Z_2)^2} \\ R = (\frac{Z_2-Z_1}{Z_2+Z_1})^2 \end{cases} \tag{1.20}$$

Il est important de distinguer les coefficients de transmission et de réflexion selon qu'ils expriment des rapports d'amplitude (pression acoustique) ou d'énergie (intensité acoustique).

Les coefficients T et R définis à l'équation 1.20 représentent les rapports d'intensité acoustique, c'est-à-dire la fraction d'énergie transmise et réfléchie respectivement.

En incidence normale, on peut également définir les coefficients de transmission et de réflexion en amplitude, notés t et r, qui sont liés aux coefficients en intensité par les relations suivantes :

$$\begin{cases} t = \sqrt{T\frac{Z_2}{Z_1}} = \frac{2Z_2}{(Z_1+Z_2)} \\ r = \sqrt{R} = \frac{Z_2-Z_1}{Z_2+Z_1} \end{cases} \tag{1.21}$$

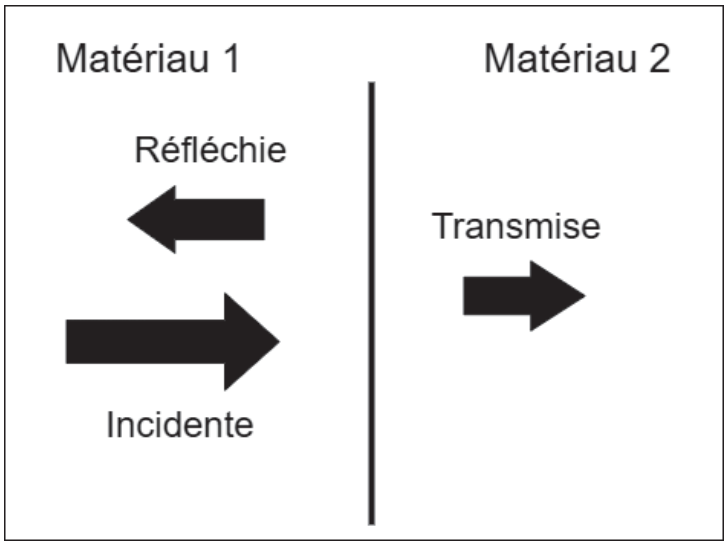


Figure 1.2 Transmission et réflexion d'une onde incidente normale

Si les deux milieux ont des impédances proches ($Z_2 = Z_1$), $r = R = 0$ et $t = T = 1$. L'onde est donc totalement transmise. À l'inverse, si $Z_1 \gg Z_2$, on a $r = -1$ et $t = 0$, l'onde est totalement réfléchie.

1.1.3.2 Transmission et réflexion en incidence oblique

Maintenant que l'onde n'est plus en incidence normale, il faut prendre en compte l'angle d'incidence de l'onde ultrasonore. Ce changement implique, en plus des angles de réflexion et de transmission, une conversion de mode d'un type d'onde à un autre (voir la figure 1.3). Afin d'avoir les directions de propagations de ces ondes, on utilise la Loi de Snell Descartes :

$$\frac{sin(\theta_i)}{v_i} = \frac{sin(\theta_{t_T})}{v_{t_T}} = \frac{sin(\theta_{t_L})}{v_{t_L}} = \frac{sin(\theta_{r_L})}{v_{r_L}} = \frac{sin(\theta_{r_T})}{v_{r_T}} \tag{1.22}$$

Avec v la vitesse de l'onde, θ son angle et les indices i, t et r indiquant respectivement l'onde incidente, l'onde transmise et l'onde réfléchie. Cette équation met en lumière la possibilité d'angle critique. En effet, si l'on considère un exemple où l'onde incidente est longitudinale et que sa vitesse est moins élevée que la vitesse longitudinale dans le milieu transmis, il existe un angle θ_i tel que $\theta_{t_L} > 90°$. Ainsi, l'onde transmise sera uniquement transversale (l'onde longitudinale transmise ne se propagera plus (Shull (2002)))

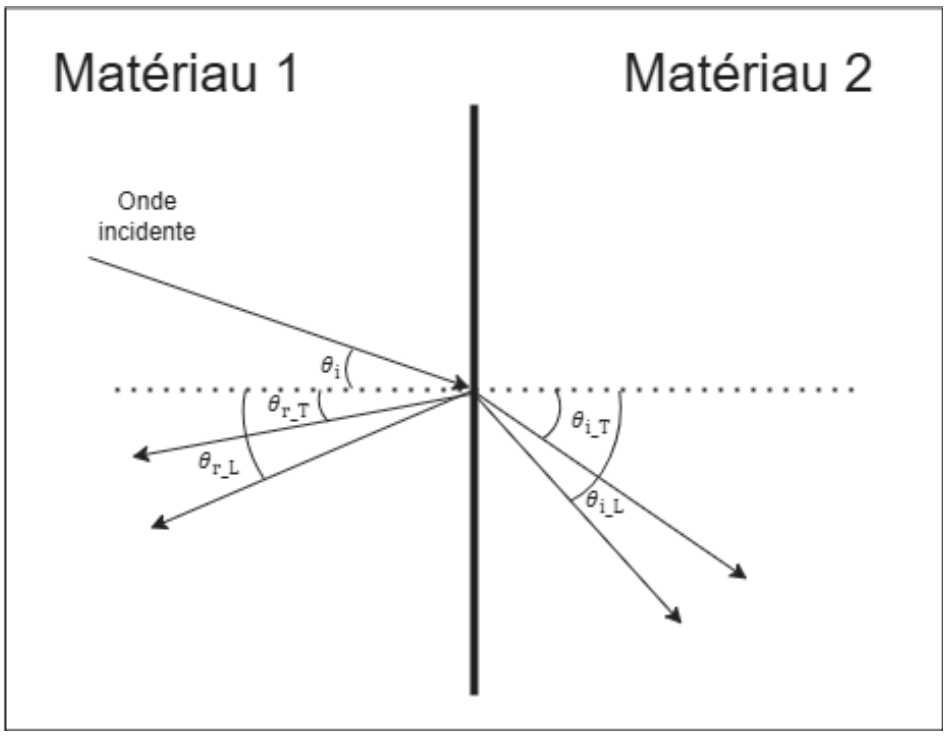


Figure 1.3 Transmission et réflexion d'une onde incidente oblique

1.2 Imagerie ultrasonore

La partie précédente a montré les équations permettant d'obtenir la propagation des ondes dans un milieu, puis a abordé l'absorption et la transmission de celles-ci dans un milieu. Cela permet donc d'obtenir des informations sur le milieu dans lequel se propage l'onde comme par exemple

la présence de défaut ou l'épaisseur de la pièce. La partie suivante s'intéresse aux différentes méthodes utilisées pour l'analyse des ondes ultrasonores.

1.2.1 A-scan

Le A-scan permet de représenter l'amplitude d'un signal en fonction du temps à un endroit donné. Une onde ultrasonore est émise grâce à une sonde et est réfléchie lorsqu'elle rencontre un changement de milieu ce qui permet d'obtenir grâce à la vitesse de l'onde dans le matériau la profondeur d'un défaut ou l'épaisseur d'une pièce (R A Mountford & P N T Wells (1972)). Comme sur la figure Fig. 1.4, après l'émission de l'onde donnant le premier pic, l'onde émise se réfléchit contre le fond de la pièce et revient vers la sonde générant le deuxième pic. En présence de défaut, d'autres échos se forment et ainsi d'autres pics apparaissent. Ces réflexions se répètent à mesure que l'onde se réfléchit entre les interfaces, mais l'atténuation du matériau réduit progressivement l'amplitude des pics successifs. Ce type de représentation constitue la base de l'imagerie ultrasonore. Il est notamment utilisé pour la mesure d'épaisseur.

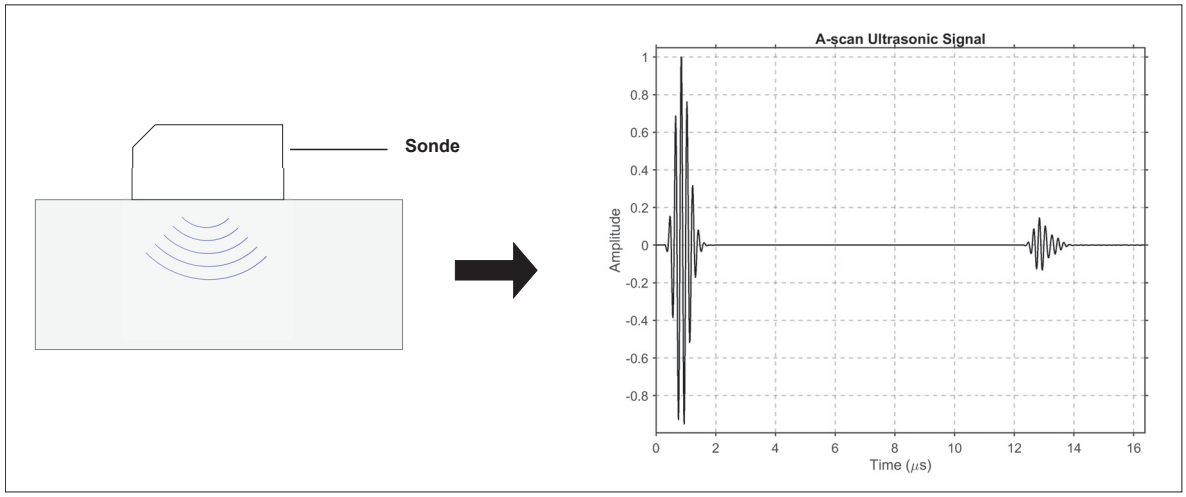


Figure 1.4 Acquisition d'un A-scan

1.2.2 B-scan

L'utilisation d'une sonde multi-élément permet de répéter l'opération du A-scan en différents points de la pièce. On obtient alors des A-scans successifs constituant un plan bidimensionnel de la pièce. Cette méthode est appelé B-scan et permet une meilleure observation des défauts (taille, position) et de la pièce.

Ainsi, contrairement au A-scan, qui fournit une information ponctuelle, fournit une vue en coupe bidimensionnelle de la pièce dans la direction du balayage, facilitant grandement l'interprétation pour les opérateurs et les algorithmes de traitement d'image. Il constitue l'un des formats d'imagerie les plus utilisés dans les inspections industrielles, notamment pour l'évaluation des soudures.

1.2.3 S-Scan

Les méthodes précédentes utilisaient de manière indépendante chaque élément de la sonde. Cependant, il est possible, en donnant à chaque élément un certain retard, de focaliser les ondes ultrasonores en une certaine direction. En effet, on peut faire superposer les fronts d'ondes grâce au principe des interférences constructives ce qui permet d'amplifier le signal selon la direction voulue mais on peut également faire converger les fronts d'ondes vers un point focal spécifique (Figure 1.5).

En faisant varier les retards, il est possible de balayer de manière angulaire la pièce et donc d'obtenir une large zone depuis une position de la sonde. Les différents A-scans obtenus à chaque angle sont utilisés pour reconstruire une image sectorielle complète du matériau : c'est le principe du S-scan.

1.2.4 Capture matricielle complète

La méthode de Capture Matricielle Complète (CMC) ou "Full Matrice Capture"(FMC) en anglais permet d'utiliser la totalité des éléments de la sonde. En effet, contrairement aux méthodes

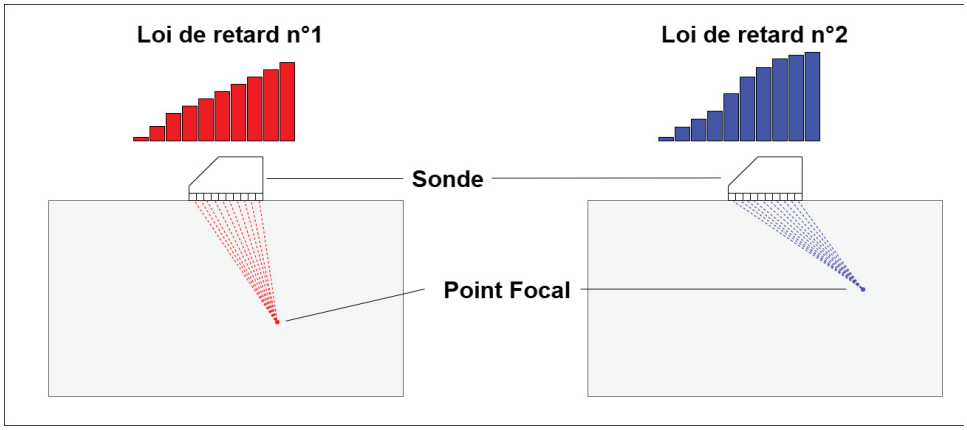


Figure 1.5 Exemple de deux S-scan différents

précédentes, elle consiste à enregistrer l'ensemble des signaux reçus pour chaque signal émis pour chaque élément de la sonde : un élément émet une impulsion ultrasonore et l'entièreté des éléments enregistrent les signaux réflechis puis c'est la même chose avec l'élément suivant (Fig 1.6). Ainsi, pour une sonde comportant N éléments, on obtient un total de N^2 A-scans, chacun correspondant à une combinaison émetteur-récepteur.

L'un des grands intérêts de la FMC réside dans la richesse et la flexibilité des données collectées. Elle permet, en effet, de reconstruire numériquement différentes représentations d'image ultrasonore (telles que les B-scans ou S-scans), en appliquant des lois de retard artificielles pour simuler des configurations d'émission focalisées, notamment dans le cadre du balayage sectoriel.

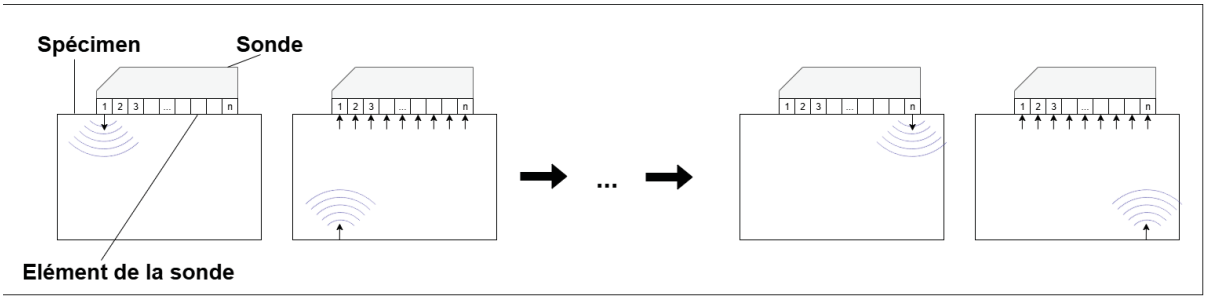


Figure 1.6 Principe d'une FMC

1.2.5 Méthode de focalisation totale

La richesse de l'information contenue dans les acquisitions FMC peut être pleinement exploitée par la Méthode de Focalisation Totale ou "Total Focusing Method"(TFM) (Holmes, Drinkwater & Wilcox (2005)). Celle-ci utilise le principe de focalisation numérique de l'onde ultrasonore sur chaque point de l'image. Ainsi, pour chaque point (x, y) de l'image, on peut calculer le temps de propagation de chaque combinaison émetteur-récepteur (i, j) vers ce point. On regroupe ensuite chacune des contributions ce qui donne :

$$I_{TFM}(x, y) = \sum_i \sum_j V_{i,j}(T_i(x, y) + T_j(x, y)) \tag{1.23}$$

Où les indices i et j indiquent les éléments émetteurs et récepteurs, respectivement. V est alors l'amplitude de l'onde qui se propage entre ces deux éléments et T est le temps de vol entre un élément et le point (x, y) de l'image, temps calculé grâce à la vitesse de l'onde dans le matériau.

Cette focalisation en tout point permet d'obtenir une très bonne résolution (Fig. 1.7) et donc une localisation précise des défauts de toute taille et du fond de la pièce. C'est la raison pour laquelle cette méthode représente une référence en matière de contrôle non destructif haute résolution, notamment dans les domaines du nucléaire, de l'aéronautique ou de l'énergie... Cependant, ces avantages s'accompagnent également d'un problème majeur : la taille importante des fichiers de la FMC et le temps de calcul important lié à l'ensemble des sommations faites et aux nombres de points de l'image.

1.3 Réseaux neuronaux

Cette étude utilise l'apprentissage supervisé afin d'aider les inspecteurs à identifier des défauts dans des images PAUT. Ce dernier constitue l'approche la plus couramment utilisée en intelligence artificielle pour les tâches de classification et de détection. Cette technique repose sur l'utilisation d'un ensemble de données annotées, où chaque entrée est associée à une sortie attendue (ou label). Le modèle apprend à établir une relation entre entrées et sorties en ajustant ses paramètres

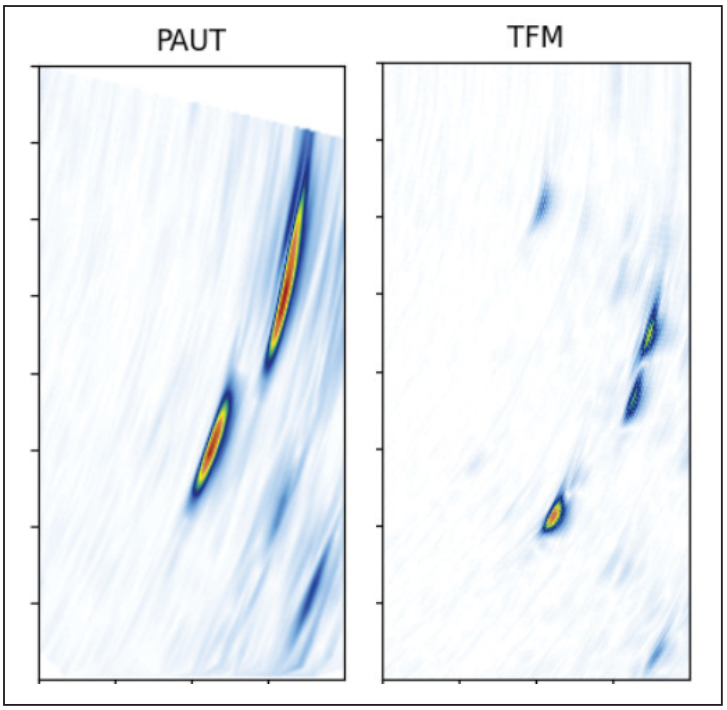


Figure 1.7 Exemple de PAUT et de TFM avec un manque de fusion

(ou poids) de manière itérative, afin de minimiser une fonction de coût. Une fois entraîné, il est capable de prédire la sortie pour de nouvelles données non vues, en généralisant les relations apprises (Murphy (2014)). Cette partie s'intéresse donc aux différentes techniques utilisées en intelligence artificielle et plus particulièrement en apprentissage supervisé

1.3.1 Réseau de neurones profonds

Les réseaux de neurones profonds sont une des architectures de base de l'apprentissage supervisé. Ils se composent de plusieurs couches, la première étant la couche d'entrée et la dernière la couche de sortie. Dans chacune d'entre elles, chaque neurone est relié à tous ceux de la couche précédente (Fig 1.8). Ces neurones calculent une combinaison linéaire pondérée de ses entrées, suivie d'une fonction d'activation non linéaire, permettant au réseau de modéliser des relations

complexes (développé par Rosenblatt (1958)). Ce qui donne donc :

$$y_j = f_j(\sum_{i}^{n}(x_i * w_{ij}) + b_j) \tag{1.24}$$

Avec y_j la sortie pour le neurone j, les x_i correspondant aux entrées, w_{ij} les poids associés aux entrées x_i pour le neurone j, b_j le biais du neurone et f_j sa fonction d'activation.

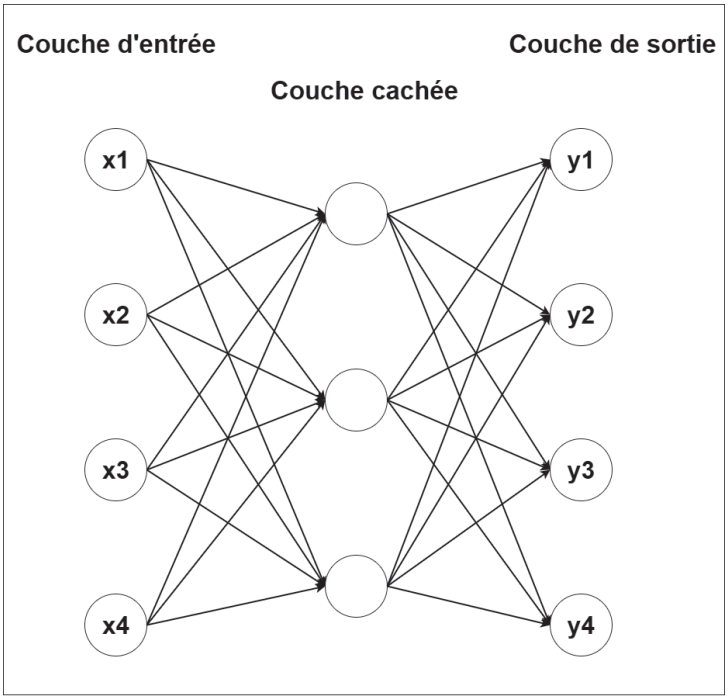


Figure 1.8 Schéma d'un réseau de neurones profonds

1.3.2 Fonctions d'activations

Les fonctions d'activation jouent un rôle essentiel dans les réseaux de neurones, car elles introduisent de la non-linéarité dans le modèle. Sans elles, un réseau de neurones, même profond, se réduirait à une simple combinaison linéaire, incapable de représenter des relations complexes entre les données.

La fonction sigmoïde est définie par :

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Elle contraint la sortie dans l'intervalle $[0, 1]$, ce qui la rend adaptée aux probabilités. Cependant, elle présente le problème de la disparition de gradient (vanishing gradient en anglais) dans les réseaux profonds.

La tangente hyperbolique, ou *tanh*, est donnée par :

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

Elle produit des valeurs comprises entre $[-1, 1]$. Souvent plus performante que la sigmoïde, elle reste toutefois sujette au même problème de disparition des gradients.

La fonction Rectified Linear Unit (ReLU) (proposée dans les réseaux neuronaux par Glorot, Bordes & Bengio (2011)) est définie par :

$$\mathrm{ReLU}(x) = \max(0, x)$$

Elle est aujourd'hui la plus répandue. Elle permet une convergence rapide et une propagation de gradients plus stable. Néanmoins, elle peut souffrir du problème des neurones morts, lorsque certaines unités cessent de s'activer.

Pour pallier ses limites, plusieurs variantes ont été proposées, telles que le Leaky ReLU, le Parametric ReLU (PReLU) ou l'Exponential Linear Unit (ELU). Ces fonctions conservent les avantages de la fonction ReLU tout en réduisant les risques de neurones inactifs (Xu, Wang, Chen & Li (2015)).

Le choix de la fonction d'activation influence directement la vitesse d'apprentissage, la stabilité du modèle et ses performances finales. Dans les réseaux modernes, la ReLU et ses variantes sont généralement privilégiées, notamment dans les architectures CNN. Dans le cadre de cette

étude portant sur la détection de défauts dans les images PAUT, la fonction ReLU a été adoptée dans l'ensemble des couches convolutives du Faster R-CNN et d'EfficientNet-B5, permettant une convergence rapide tout en maintenant des gradients stables lors de la rétropropagation à travers les multiples couches du réseau.

1.3.3 Réseau de neurones convolutifs

Les réseaux de neurones convolutifs (ou Convolutional Neural Networks (CNN) en anglais) représentent une classe particulière de réseaux de neurones artificiels qui a été spécifiquement conçue pour traiter des données structurées en grilles, telles que les images. L'utilisation des CNN pour l'analyse d'images ultrasonores présente une cohérence fondamentale avec la physique sous-jacente : les images PAUT elles-mêmes résultent d'une convolution entre l'onde ultrasonore émise et les réflecteurs présents dans le matériau. Ainsi, l'application d'opérations de convolution par les CNN pour extraire des caractéristiques pertinentes s'aligne naturellement avec le processus physique de formation de l'image, rendant cette architecture particulièrement adaptée à la détection de défauts en contrôle non destructif.

Proposés initialement par Lecun, Bottou, Bengio & Haffner (1998) avec l'architecture LeNet-5, puis popularisés par Krizhevsky, Sutskever & Hinton (2012) avec AlexNet, les CNN ont profondément transformé le domaine de la vision par ordinateur et constituent aujourd'hui l'architecture de référence pour l'analyse d'images.

Contrairement aux réseaux de neurones entièrement connectés, qui traitent toutes les entrées de manière uniforme, les CNN exploitent la structure spatiale des données. Une image peut être vue comme une matrice de valeurs (en niveaux de gris ou en canaux RGB), et les opérations de convolution permettent d'extraire automatiquement des motifs locaux tels que des contours, des textures ou encore des formes complexes. Une couche de convolution applique un ensemble de filtres qui balayent l'image en effectuant un produit scalaire local. Chaque filtre agit alors comme un détecteur de caractéristiques, et les activations produites forment ce que l'on appelle des cartes de caractéristiques (ou feature maps en anglais) (Fig 1.9).

Un réseau convolutif est généralement constitué de plusieurs couches qui remplissent des rôles complémentaires (Goodfellow, Bengio & Courville (2016)). Les couches de convolution assurent l'extraction hiérarchique des caractéristiques, en apprenant d'abord des motifs simples comme les bords ou les couleurs, puis des structures de plus en plus complexes dans les couches profondes. Après chaque convolution, une fonction d'activation non linéaire, généralement la fonction ReLU, est appliquée afin d'introduire de la non-linéarité et de permettre au réseau de modéliser des relations complexes. Des couches de pooling viennent ensuite réduire la dimension spatiale des cartes de caractéristiques tout en préservant les informations essentielles. Cette opération rend le modèle plus robuste aux petites translations et réduit le coût computationnel. Enfin, dans les dernières couches du réseau, des couches entièrement connectées combinent l'ensemble des caractéristiques extraites pour produire la sortie finale, qu'il s'agisse d'une classification ou d'une régression.

L'un des principaux avantages des CNN réside dans le partage des poids : un même filtre est appliqué à l'ensemble de l'image, ce qui réduit drastiquement le nombre de paramètres par rapport à un réseau dense classique. De plus, la combinaison de la convolution et du pooling confère au modèle une invariance locale aux transformations, tout en permettant une hiérarchie de représentations de plus en plus abstraites.

Les réseaux convolutifs ont ainsi trouvé des applications dans un large éventail de tâches en vision par ordinateur, notamment la classification d'images, la détection et la reconnaissance d'objets, la segmentation sémantique, la reconnaissance faciale ou encore l'analyse médicale d'images. Leur efficacité et leur capacité à apprendre directement des données brutes expliquent leur rôle central dans les avancées récentes de l'intelligence artificielle.

1.3.4 Optimisation

L'entraînement d'un réseau de neurones convolutif repose sur la recherche d'un ensemble de paramètres, tels que les poids et les biais, qui minimise une fonction de coût donnée. Cette étape, appelée l'optimisation, constitue l'un des aspects les plus critiques de l'apprentissage profond.

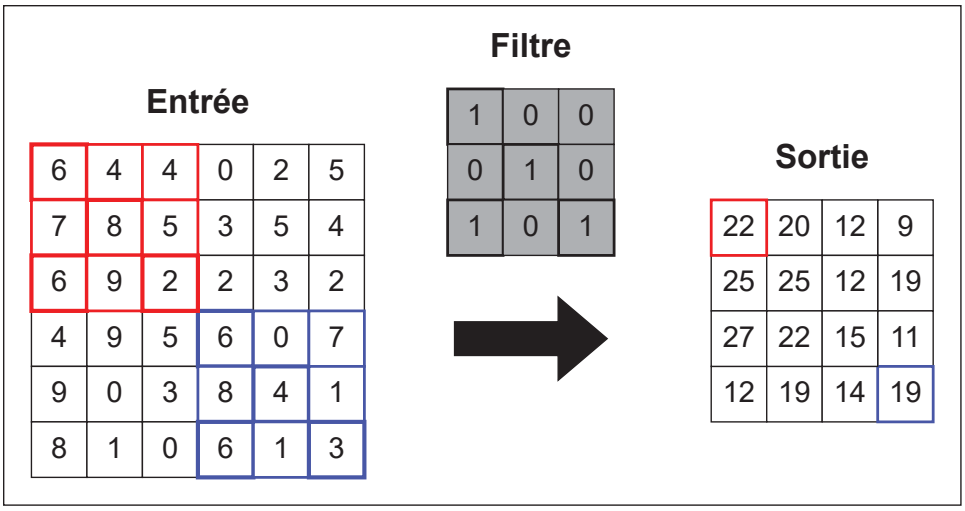


Figure 1.9 Exemple de fonctionnement d'un filtre de convolution

L'objectif est de réduire l'erreur entre les prédictions du modèle et les valeurs attendues, tout en garantissant une bonne capacité de généralisation aux nouvelles données.

L'optimisation repose principalement sur l'algorithme de descente de gradient. La descente de gradient consiste à calculer la dérivée de la fonction de coût par rapport aux paramètres du réseau, puis à mettre à jour ces derniers afin de minimiser l'erreur (Goodfellow *et al.* (2016)).

Différentes variantes de la descente de gradient ont été proposées afin d'accélérer la convergence et d'éviter les minima locaux ou les plateaux. Parmi les plus couramment utilisées, on retrouve la Stochastic Gradient Descent (SGD) (Robbins Herbert (1951)), qui met à jour les poids à partir d'un sous-échantillon aléatoire des données, ainsi que des méthodes plus avancées comme ADAM (Kingma & Ba (2017)) ou Adagrad, qui adaptent dynamiquement le taux d'apprentissage en fonction des gradients observés.

Outre le choix de l'optimiseur, plusieurs stratégies influencent la qualité de l'optimisation. Le réglage du taux d'apprentissage est crucial : une valeur trop élevée peut empêcher la convergence, tandis qu'une valeur trop faible entraîne un apprentissage lent. Des techniques comme le learning rate scheduling, qui consiste à diminuer progressivement le taux d'apprentissage au cours de l'entraînement, se révèlent particulièrement efficaces. De plus, l'ajout de termes de régularisation,

tels que la pénalisation $L2$ ou le dropout, permet de limiter le surapprentissage en contraignant la complexité du modèle.

Enfin, l'optimisation des modèles ne se limite pas à l'ajustement des poids par l'algorithme d'entraînement. Elle inclut également le choix des hyperparamètres, tels que la profondeur du réseau, le nombre de filtres par couche, ou encore la taille des lots.

En somme, l'optimisation des modèles constitue une étape déterminante dans la performance des réseaux convolutifs. Elle combine des techniques algorithmiques, des stratégies de régularisation et des choix d'architecture, afin de produire un modèle à la fois précis, robuste et capable de généraliser efficacement.

1.4 Conclusion de la revue de littérature

L'étude des phénomènes liés à la propagation des ondes ultrasonores, combinée aux avancées récentes en apprentissage profond, met en évidence l'importance d'une approche intégrée associant physique et intelligence artificielle. D'une part, la compréhension des mécanismes tels que l'atténuation ou la réflexion aux interfaces est essentielle pour interpréter correctement les signaux acquis et en extraire des informations pertinentes sur les matériaux analysés. D'autre part, les réseaux de neurones convolutifs se révèlent particulièrement efficaces pour traiter ces données complexes, grâce à leur capacité à extraire automatiquement des caractéristiques hiérarchiques et à modéliser des relations non linéaires. Enfin, l'optimisation des modèles, qu'il s'agisse du choix des fonctions de coût, des algorithmes de mise à jour ou des techniques de régularisation, constitue une étape décisive pour garantir à la fois la performance et la robustesse des réseaux.

1.5 Objectifs et méthodologie

L'objectif principal de ce projet de recherche est de développer un système de détection automatisée de défauts dans les images PAUT capable d'assister les inspecteurs en contrôle non destructif. Plus spécifiquement, ce travail vise à :

- Comparer les performances de deux familles d'architectures de détection d'objets de pointe : les détecteurs à deux étapes basés sur Faster R-CNN et les détecteurs à une étape de la famille YOLO (v5 et v8) ;
- Évaluer l'impact de différentes stratégies d'enrichissement contextuel des données sur les performances de détection et de classification des défauts ;
- Déterminer si l'intégration d'informations contextuelles sur la géométrie du spécimen peut améliorer la capacité du modèle à distinguer les défauts réels des artefacts structurels.

Pour atteindre ces objectifs, une méthodologie en plusieurs étapes a été mise en place. Premièrement, un vaste jeu de données expérimental a été constitué à partir de spécimens de soudure industriels contenant divers types de défauts (manques de fusion, porosité, fissures). Afin de contourner les difficultés d'obtention de données expérimentales et les contraintes de confidentialité inhérentes au domaine industriel, des acquisitions par capture matricielle complète (FMC) ont été réalisées. Cette approche permet de générer, à partir d'un nombre limité d'acquisitions physiques, une grande quantité d'images sectorielles avec différentes configurations d'imagerie (focalisations, ouvertures, arrangements de sous-ouvertures).

Deuxièmement, trois bases de données distinctes ont été créées pour explorer l'influence de l'information contextuelle : des images PAUT brutes servant de référence, des images enrichies par l'ajout d'un schéma (overlay) indiquant la géométrie de la soudure, et des images prétraitées par soustraction de la médiane de la géométrie afin de réduire les artefacts structurels et mettre en évidence les régions défectueuses. Cette dernière méthode nécessite plusieurs acquisitions avec et sans défauts, ce qui s'aligne avec les pratiques d'inspection où les inspecteurs effectuent généralement plusieurs passages sur les pièces.

Troisièmement, l'architecture Faster R-CNN a été progressivement améliorée par l'intégration de composants avancés : remplacement du backbone ResNet par EfficientNet-B5 avec réseau pyramidal de caractéristiques (FPN), adoption de ROI Align pour un alignement spatial précis, et optimisation des boîtes d'ancrage par K-means pour mieux s'adapter aux caractéristiques

géométriques des défauts PAUT. Une étude d'ablation systématique a permis de quantifier l'impact individuel de chaque amélioration architecturale.

Enfin, les performances des différentes architectures ont été évaluées de manière rigoureuse sur un jeu de validation provenant d'un fabricant différent de celui utilisé pour l'entraînement, permettant ainsi de mesurer la capacité de généralisation des modèles dans des conditions réalistes d'inspection industrielle. Les métriques utilisées incluent la précision moyenne (mAP), le rappel, la précision de classification, et l'analyse des faux positifs et faux négatifs.

CHAPITRE 2

ENHANCED FLAW DETECTION IN PHASED ARRAY ULTRASONIC TESTING USING ADVANCED DEEP LEARNING ARCHITECTURES

Eloi Lombard , Sevan Bouchy , Hugo Hervé-Côte , Pierre Bélanger

Département de Génie Mécanique, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Article soumis à la revue « Engineering Applications of Artificial Intelligence » en Novembre 2025.

2.1 Abstract

Phased Array Ultrasonic Testing (PAUT) is one of the most widely used techniques in nondestructive testing (NDT), for identifying and characterizing defects within materials, particularly in welds. However, interpretation of PAUT data significantly depends on the inspector's expertise. Moreover, presence of geometric features associated with the weld often introduces artefacts, complicating the analysis and increasing interpretation time. While artificial intelligence (AI) has demonstrated outstanding capabilities in object detection across various fields, its application to ultrasonic testing has also advanced significantly, despite the limited data availability due to confidentiality constraints. This work investigates the integration of AI into PAUT analysis by comparing two state-of-the-art deep learning architectures for automatic flaw detection and characterization : 1) a two-step detector based on the Faster R-CNN, and 2) the latest version of YOLO. To explore the impact of contextual information on model performance, three types of data preprocessing were analyzed : raw PAUT data used as a baseline, data enhanced with overlay annotations to indicate the weld geometry, and data preprocessed by subtracting echoes associated with the geometry, aiming to suppress structural artefacts and highlight defect regions. Experimental results show that the enhanced Faster R-CNN architecture, combined with median geometry subtraction, achieved the highest performance with a mean Average Precision (mAP) of 58.05%, outperforming YOLOv5 and YOLOv8 across all dataset configurations. These findings demonstrate the potential of deep learning for improving PAUT

flaw detection while highlighting the importance of contextual preprocessing for robust and accurate defect characterization.

Keywords : Phased Array Ultrasonic Testing(PAUT), Convolutional Neural Network(CNN), Faster R-CNN, YOLO, Flaw, Classification

2.2 Introduction

Non-destructive testing (NDT) plays a crucial role in maintaining safety and cost efficiency in critical fields such as nuclear energy (Norris, Naus & Graves (1999)), aerospace, and oil and gas. Among the various techniques available, ultrasonic imaging stands out due to its versatility and practical implementation. Phased Array Ultrasonic Testing (PAUT) has become a widely adopted method due to its precision and efficiency. Multi-element probes with precisely timed emission delays allow beam steering and focusing on specific Regions of Interest (RoIs). This technological approach allows for high-resolution imaging while significantly reducing inspection time compared to conventional methods. However, despite these advantages, inspection quality still depends heavily on the inspector's skill and experience, as interpreting complex signals and images is challenging and can lead to significant variability between inspectors.

To address these challenges of labor-intensive analysis and result variability, the implementation of machine learning (ML) has grown in interest. Image recognition, in particular, has shown an important shift by the advancements of Convolutional Neural Networks (CNNs) and their ability to learn complex representations through multiple layers. The introduction of advanced architectures such as ResNet (He, Zhang, Ren & Sun (2015)), DenseNet (Huang, Liu, Maaten & Weinberger (2019)), and EfficientNet (Tan & Le (2020)) has greatly enhanced the capabilities of CNNs which now makes them suitable for PAUT image analysis. These models have addressed critical challenges in training deep networks, leading to improved accuracy in image recognition tasks. In the context of ultrasonic NDT, these advancements have facilitated more reliable and efficient defect detection, thereby enhancing the safety and integrity of inspected structures and components. Existing research has demonstrated promising results in this domain

(Chen *et al.* (2024), Zhou, Lu, Wang & Huang (2019), Provencal & Laperrière (2021)). Zhang et al. used CNNs to classify various types of defects in X-ray images, demonstrating the effectiveness of deep learning models for weld defect classification (Zhang, Chen, Zhang, Xi & Le (2019)). Zhu et al. obtained over 95% of success rate with a CNN coupled with the random forest algorithm for weld surface defect classification (Zhu, Ge & Liu (2019)). For defect characterization and sizing, Latête et al. employed Faster R-CNN to detect and size flat-bottom holes in ultrasonic images, achieving 70% detection accuracy (Latête, Gauthier & Belanger (2021)). Chen et al. achieved results exceeding 93% of precision on S-scan images using an enhanced Faster R-CNN architecture (Chen, Wang & Huang (2023)). From a single-stage detection perspective, Posilovic et al. (Posilovic *et al.* (2019)) and Medak et al. (Medak, Posilovic, Subasic, Budimir & Loncaric (2021)) demonstrated the effectiveness of YOLO and EfficientNet architectures for defect recognition on B-scan ultrasonic datasets of varying sizes. While these studies show promise, greater difficulty is observed in defect classification and sizing in the presence of weld geometry (Hervé-Côte, Dupont-Marillia & Bélanger (2024)), where image interpretation becomes considerably more difficult. Another problem arises from the lack of data due to confidentiality concern.

This study addresses the challenge of detecting welding defects in complex inspection environments using a large experimental dataset derived from Full Matrix Capture (FMC) acquisitions converted into PAUT images. Two state-of-the-art deep learning architectures are compared : Faster R-CNN and YOLO. Strategies to enrich contextual information are incorporated. To assess the impact of data representation, three dataset configurations we explored : (1) raw PAUT images as a baseline, (2) images augmented with manual geometry overlays to provide structural context, and (3) images preprocessed through median geometry subtraction to reduce artifacts and emphasize defect regions.

The main contribution of this work lies in providing the first systematic comparison of Faster R-CNN and YOLO for PAUT flaw detection, combined with an in-depth evaluation of contextual preprocessing techniques. Through extensive experiments and a comprehensive ablation study, we demonstrate that both architectural choices and data preparation strategies, including median

subtraction, play a critical role in optimizing performance, offering practical insights for deploying AI-driven solutions in industrial nondestructive testing.

The remainder of this paper is organized as follows : Section 2 describes the dataset, preprocessing strategies, and model architectures. Section 3 presents the experimental results and ablation studies. Section 4 discusses the findings, limitations, and implications for industrial applications. Finally, Section 5 concludes the paper and outlines directions for future research in industrial nondestructive testing.

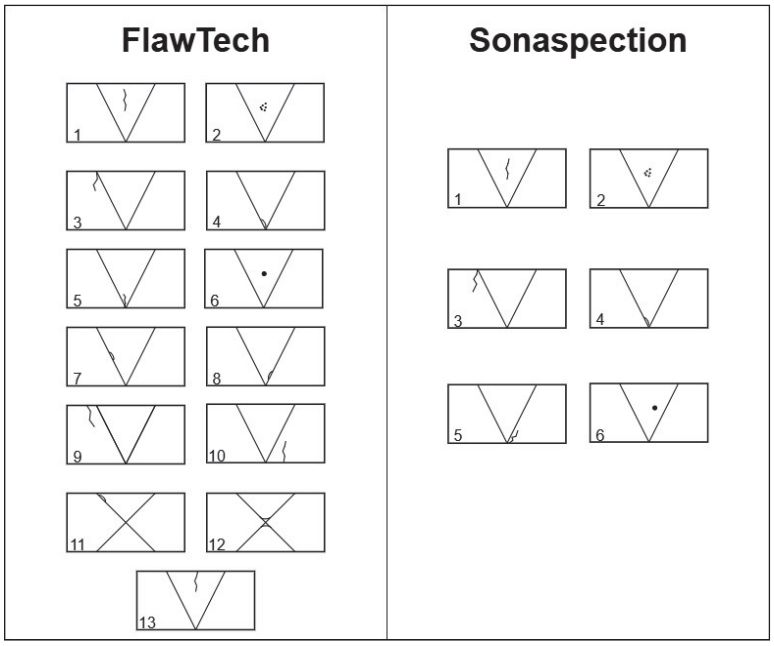
2.3 Materials and methods

2.3.1 Dataset

Labeled PAUT data from industrial inspections are highly challenging to access as they are typically proprietary and confidential. However, the development of artificial intelligence models require large, well-annotated training datasets to achieve reliable performance. To address this limitation, this study used training specimens from Sonaspection and FlawTech which are widely used for certification and inspector training. The specimens, as schematically presented in Figure 2.1, consisted of seven 100 mm-long blocks (FlawTech samples) and two 250 mm-long blocks (Sonaspection samples), all featuring V or double V weld profiles with different thicknesses (9.5, 12 and 16 mm). Each block contains two to three distinct flaws – such as lacks of fusion, porosities, and various types of cracks – providing a diverse set of realistic case scenarios that reflect those encountered during industrial phased array inspections.

2.3.1.1 Converting FMC acquisitions into multiple PAUT images

To maximize data exhaustivity, Full Matrix Capture (FMC) acquisitions were used to generate PAUT images. This acquisition method records A-scans for every possible transmitter–receiver pair within the array, providing a comprehensive dataset suitable for post-processing. The specimens were scanned multiple times, with varying probe positions to enhance image diversity,



N°	Flaw
1	Centerline Crack
2	Porosity
3	Toe Crack
4	Lack of Fusion
5	Root Crack
6	Slag Inclusion
7	Lack of Fusion
8	Lack of Fusion
9	Base metal Crack
10	Base metal Crack
11	Lack of Fusion
12	Incomplete Penetration
13	Longitudinal Crack

Figure 2.1 Schematics of specimens used for training and evaluation. Data on 13 different flaws were acquired and saved into a database.

including inspections from both sides of the weld and at different lateral offsets from it. The scanning path followed the weld along the index axis in 1 mm increments, yielding approximately 90 FMCs per configuration for the 100 mm samples and 250 for the 250 mm samples.

Only FMC acquisitions containing defects were retained. This choice is technically justified by Faster R-CNN architecture : it learns to classify the majority of its proposals (around 90% per image) as background, while the detection head includes a dedicated background class. Combined with each PAUT image containing around 85% of background, this ensures robust negative sample learning without requiring dedicated defect-free images. The flaws in these specimens spans approximately 10 to 14 mm. To focus the training process on the most informative regions while reducing redundancy, mitigating overfitting, and limiting the time required for manual annotation and training, only three FMC acquisitions per flaw – corresponding to the two edges and the central position – were selected for training across each acquisition configuration. This selection yields three FMCs per flaw for each acquisition configuration, resulting in a total of

576 FMCs for training. For validation and testing, the full spatial extent of each defect was preserved, producing 240 FMCs acquisitions for the validation set.

Following the work of Holmes et al. (Holmes *et al.* (2005)), each FMC was post-processed into an S-scan by applying delay laws that synthetically focus the array at angle θ and distance r. The focused contribution is obtained by coherently summing the appropriately time-shifted signals :

$$I(r,\theta) = |\sum h_{ij}(\frac{2r + x_i sin(\theta) + x_j sin(\theta)}{c})| \tag{2.1}$$

where h_{ij} is the complex Hilbert transform of the signal recorded from transmitter i to receiver j. c corresponds to the velocity of the wave. Building upon this, Hervé-Côte et al. (Hervé-Côte *et al.* (2024)) proposed a set of imaging configurations designed to reflect realistic inspection scenarios. These augmentations, described in Figure 2.2 involve variations in focal depth, aperture size, and sub-aperture arrangements simulating 32- and 64-element probes. This approach, substancially increases the variability and representativeness of the dataset.

As a result, for the training dataset, approximately 9000 PAUT images were generated from 13 defects identified and characterized by a level 3 inspector. For validation and test, 4000 PAUT images were obtained from 6 different defects.

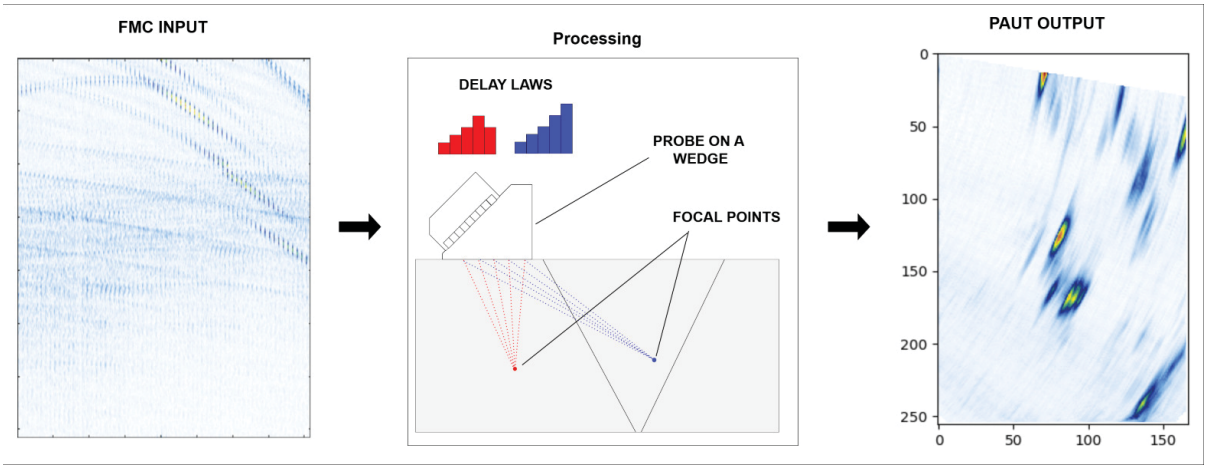


Figure 2.2 Processing of PAUT using FMC input

2.3.1.2 Labeling

Each phased array was manually labeled to accurately identify defects and create a reliable dataset for training the AI models. This labor-intensive task cannot be fully automated, as the size of the defects varies between images, and incorrect labeling would lead to suboptimal results. Defect sizing was performed manually using a straightforward labeling tool, where bounding boxes were drawn around each defect and a corresponding defect type was assigned.

It is important to note that, in practice, inspectors rely on multiple PAUT configurations to localize, size, and characterize a defect before making a final assessment. In contrast, conventional object detection architectures process each image independently, without access to additional contextual views. As a result, defects may appear only partially in certain images, limiting the information available to the model.

To address this limitation, a partially biased labeling strategy was adopted. Instead of annotating the entire extent of the defect, labels were placed on the intensity peaks, which generally correspond to the defect extremities. This approach mimics the way inspectors interpret PAUT data—by focusing on regions of interest with high signal response—and ensures that the model is guided by the most informative features. In addition to improving interpretability, this labeling method also helps prevent overfitting by avoiding excessive focus on uncertain or noisy areas.

2.3.1.3 Adding contextual informations to the dataset

To address the challenges associated with the image-by-image visualization and inherent complexity of PAUT images, this study explored the integration of a global contextual strategy. This approach provides per-image information about the specimen's geometry, thus facilitating a more straightforward comprehension of the data. Two distinct approaches were examined to reduce the complexity of interpreting the raw image dataset.

The first approach consists in adding an overlay to the PAUT images, mimicking the way human inspectors incorporate geometric context during evaluation. This overlay encodes the geometry

of the specimen and the weld (Fig 2.3a)), thereby guiding the model by providing spatial cues that help distinguish defect signatures. While effective in simplifying classification, this method requires precise alignment of the overlay with each PAUT acquisition, introducing additional operational effort.

The second method adresses the challenge of separating the constant geometry of the specimen which is present across all images in a 100-millimeter-long block from the defects, which typically span around ten millimeters. By calculating the median along the specimen's length (rather than the mean, which would be biased by defects due to their intensity), it is possible to generate an image representing the baseline noise and artifacts generated by the weld. Subtracting this median geometry from PAUT images containing defects yields a simplified dataset (Fig 2.3b)). This method assumes, however, that inspectors must acquire a sufficient scan length before the image median can reliably capture the geometry and exclude defects. Specifically, the approach requires more defect-free PAUT acquisitions than with defects to ensure the median accurately represents the baseline geometry. However, complete block scanning is not necessary, typically 60 to 70 acquisitions spaced by 1 mm are sufficient for a functional dataset when dealing with two defects of approximately 14mm length. Fortunately, this aligns with standard inspection practices, where multiple passes are typically performed before delivering a final evaluation.

2.3.1.4 Data Augmentation

To enhance the robustness and generalization capability of the model, a data augmentation strategy was implemented, inspired by the techniques employed in YOLO (Varghese & M. (2024)). Several transformations were applied probabilistically, including horizontal flips, brightness and contrast adjustments, as well as shifts, scaling, and small rotations. These augmentations simulate real-world variability ; for instance, acquiring data from both sides of the weld can be effectively reproduced through horizontal flipping, thereby enriching the diversity of the training dataset and reducing the risk of overfitting. All images were resized to a fixed resolution of

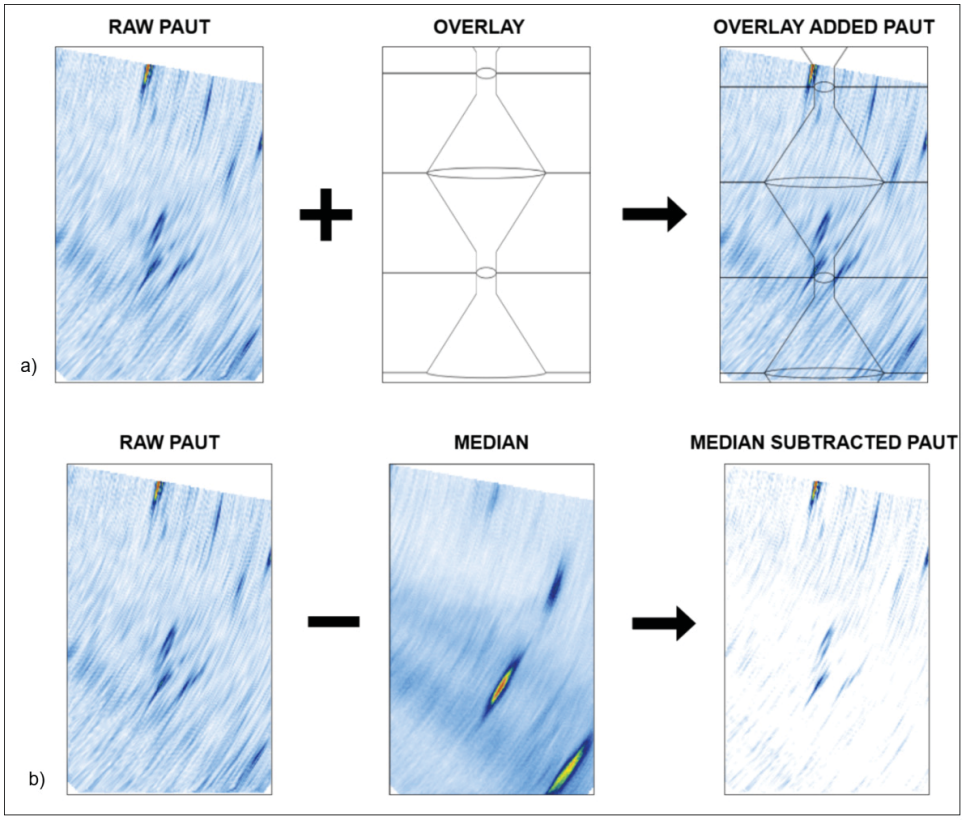


Figure 2.3 Preprocessing methods : a) adding an overlay and b) subtracting the median.

300×300 pixels and normalized. Bounding box annotations were adjusted consistently after each transformation to ensure label validity and accurate training.

2.3.1.5 Training and Validation

To train each model (see Section 2.3.2) on the three different dataset variants, a validation protocol was employed to assess robust performances and good generalization.

Training was performed on 13 labeled defects (9000 PAUT images) from the FlawTech specimens. To evaluate cross-source generalization, two evaluation datasets were created from 6 defects originating from the other manufacturer Sonaspection : a validation set of 288 PAUT images (3 FMCs per defect) for model selection, and a test set of 4000 PAUT images covering the full span of each flaw to assess detection robustness (Fig 2.1).

This strategy enabled a detailed comparison of the model's ability to detect and localize defects across different dataset types (raw, overlay, and geometry-subtracted) and deep learning architectures, providing insight into how contextual information influences detection robustness.

Models were trained for 80 epochs with a batch size of 2 using a constant learning rate of $\eta = 10^{-4}$. The Adam optimizer (Kingma & Ba (2017)) was used with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-}8$. Training was performed on NVIDIA Quadro P6000 GPUs with 24 GB of memory.

2.3.2 Architecture

The previous section explored dataset augmentation strategies designed to enhance defect classification and sizing by incorporating global contextual information. Whether through explicit overlays or geometry subtraction, these methods aim to provide the CNN with additional information to distinguish defects from the surrounding structure. However, the effectiveness of these techniques is closely tied to the choice of detection architecture.

Image recognition algorithms can generally be categorized into two main types : single-stage detectors, which prioritize speed by directly predicting object locations and classifications, and two-stage detectors, which trade off speed for enhanced accuracy by incorporating an additional region proposal step. While the former, with architectures like YOLO (You Only Look Once) (Redmon, Divvala, Girshick & Farhadi (2016)), offer real-time performance, the latter, represented by the R-CNN family, achieve superior detection performance.

This work focused on the second category of object detection models (two-stage detectors) and more specifically on Faster R-CNN, which has demonstrated strong performance in various object detection tasks. In addition, a comparative analysis was conducted with YOLOv5 and YOLOv8 (Varghese & M. (2024)).

2.3.2.1 Faster R-CNN

The Faster R-CNN framework adopted in this work operates in two stages : first, it extracts rich feature representations from input images and generates candidate regions ; second, it refines these proposals through precise pooling and a classification/regression head. The key components are detailed below.

A batch of PAUT images is processed by a deep convolutional backbone (originally VGG (Simonyan & Zisserman (2015)) implemented here with ResNet-50 (He *et al.* (2015)) and subsquently replaced by EfficientNet (Tan & Le (2020))) to produce multi-scale feature maps. In our implementation, we further enhanced these maps with a Feature Pyramid Network (FPN)Lin *et al.* (2017), which fuses high-level semantic information with fine-grained spatial details, thereby enabling robust detection of defects across a range of sizes.

From the backbone feature maps, the Region Proposal Network (RPN) generates a set of anchor boxes at each spatial location. These anchors, varying in scale and aspect ratio, serve as initial guesses for defect locations. To optimize these anchors, the K-means clustering algorithm is applied to the dataset's ground-truth bounding boxes. By clustering these bounding boxes based on their width and height, K-means determines a set of anchor sizes that best represent the actual defect distribution in the dataset. This data-driven approach improves detection performance by ensuring that the proposed regions align more closely with real-world defect shapes, reducing the reliance on arbitrary anchor settings.

Each anchor is scored by the RPN as "object" or "background." A non-maximum suppression (NMS) filters overlapping proposals, retaining a manageable number of high-confidence candidate RoIs.

The RoIs are then projected back onto the appropriate level of the FPN feature hierarchy and processed by RoI Align. Unlike max-pooling-based methods, RoI Align uses bilinear interpolation to extract fixed-size feature tensors for each RoI, preserving exact spatial relationships. This

precise alignment is critical for accurately localizing small or irregularly shaped defects in PAUT images.

Each RoI feature tensor is forwarded through fully connected layers that branch into two sibling heads : a classification Head which assigns each RoI to a defect class or background and a regression Head which refines the anchor coordinates by predicting offsets, minimizing a smooth L1 loss between the refined and ground-truth boxes. The entire network is trained end-to-end with a multi-task loss that jointly optimizes classification accuracy and bounding-box localization.

In the end, for each input image, the model outputs a set of refined bounding boxes, each defined by its center coordinates, width, and height, alongside classification scores. These outputs constitute the final RoIs used for defect characterization and sizing.

Fig 2.4 shows the structure of this enhanced Faster R-CNN.

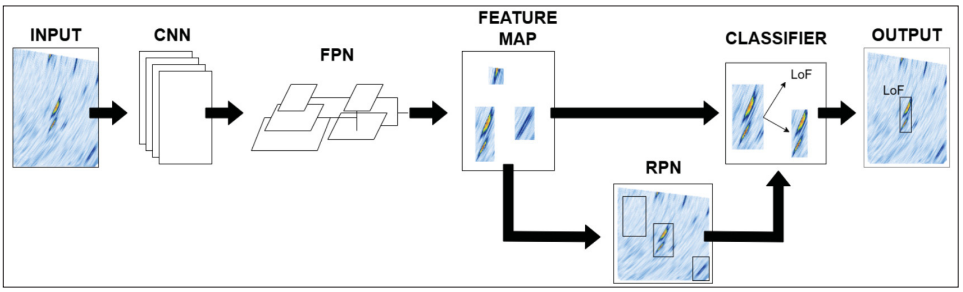


Figure 2.4 Architecture of Faster R-CNN

2.3.2.2 Evaluation metrics

To assess the performance of the neural network for defect detection and localization, the following metrics were employed.

Intersection over Union (IoU) is used to determine whether a predicted bounding box is considered a true positive. It is computed using the following formula :

$$IoU = \frac{A_{overlap}}{A_{union}} \tag{2.2}$$

where $A_{overlap}$ represents the overlapping area between the predicted and ground truth bounding boxes, and A_{union} is the total area covered by both boxes (Fig 2.5). A threshold of 50% is applied (as commonly used in PASCAL VOCEveringham, Van Gool, Williams, Winn & Zisserman (2010)), meaning that any predicted box with an IoU greater than 50% is considered a true positive.

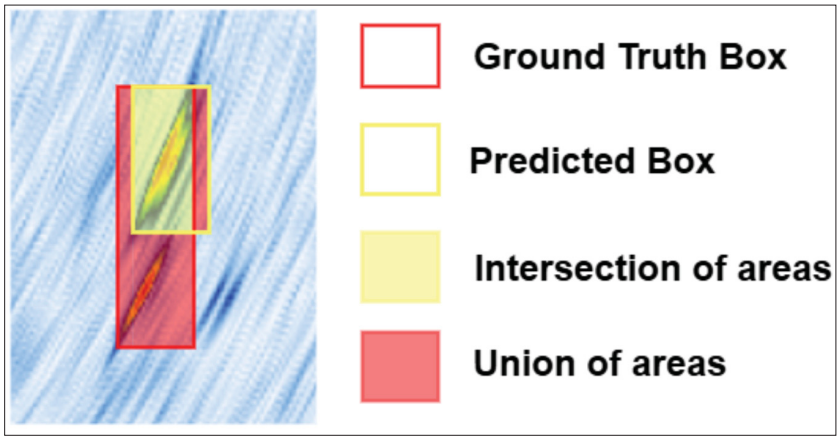


Figure 2.5 Example of the IoU metric. In this example, the IoU is about 40.2%

Following this, the precision metric is computed as follows :

$$P = \frac{N^{tp}}{N^{tp} + N^{fp}} \tag{2.3}$$

In this work, precision and recall are computed differently than in standard object detection scenarios. In fact, the datasets used contain exactly one defect per image. Although multiple reflections may appear within the same PAUT image, resulting in several predicted bounding boxes, these all correspond to the same physical flaw.

Therefore, with P which represents the precision, N^{tp} is the number of images with at least 1 bounding box correctly matching the defect, and N^{fp} is the number of bounding boxes predicted as defects but not corresponding to actual defects. This precision metric provides a robust evaluation of the detection accuracy, ensuring that the model does not generate an excessive number of false positives.

Recall measures the proportion of correctly identified positive instances among all actual positive samples. It is thus calculated on an image-level basis, as the primary objective is to correctly identify the presence of the defect rather than detecting all its individual reflections. It is defined as :

$$R = \frac{N^{tp}}{N^{tp} + N^{fn}} \tag{2.4}$$

where N^{fn} is the number of defect-containing images with no valid detections.

To provide a balanced assessment of both precision and recall performance, the F1-score is computed as follow :

$$F1 = \frac{2PR}{P + R} \tag{2.5}$$

The F1-score provides a harmonic mean of precision and recall, offering a single metric that balances the trade-off between detection accuracy and completeness.

Finally, the Average Precision (AP) is computed as the area under the Precision–Recall curve for a given IoU threshold (50% in this paper) :

$$AP = \int_0^1 P(R)\, dR \tag{2.6}$$

where $P(R)$ denotes precision as a function of recall.

The mean Average Precision (mAP) is then obtained by averaging AP scores across all defect classes :

$$mAP = \frac{1}{n} \sum_{k=1}^{n} AP_k \tag{2.7}$$

This adapted formulation ensures that the mAP remains aligned with the practical objective of defect localization in PAUT images.

A complementary classification accuracy metric is computed to assess classification performance independently of localization quality (as this work aims to provide additional support to inspectors). This metric uses the trace of the confusion matrix, comparing the number of bounding boxes with correct class labels to the total number of proposed bounding boxes, regardless of their IoU values.

2.4 Results and discussions

This section presents and analyzes the results obtained with the Faster R-CNN architecture and the impact of incorporating additional information into the PAUT images. Table 2.1 compares the performance of the Faster R-CNN proposed in this paper with YOLO v5 and 8 for the median subtracted dataset. A first observation can be drawn from those results. Specifically, two defect types, the centerline crack and the slag inclusion, did not achieve satisfactory performance on both models. Although these defects were detected, the corresponding bounding box annotations were inaccurate.

This limitation can be attributed primarily to the nature of the data used. Variations in defect positioning as well as the number of visible reflections further contributed to the inconsistency. In addition, for slag inclusions, their location and shape often resemble the specimen's geometry, leading to confusion by the model and resulting in a lower detection rate, or in some cases, a complete failure to detect them.

Regarding the remaining defect types, significant performance differences emerged between architectures. YOLO models showed limited capability in detecting Root Cracks features, achieving only 6.30% of classification accuracy for root cracks compared to 53.86% for our proposed Faster R-CNN model. This disparity reflects the fundamental trade-off between these architectures : speed-optimized single-stage detection versus precision-focused two-stage approaches. Nevertheless, even our enhanced model faced challenges in distinguishing Root

Tableau 2.1 Performance Evaluation of our proposed Faster R-CNN and YOLOv5 for different defects on the median subtracted dataset. Total column shows weighted average performance.

Model	Metric	Defect Types						Total
		Centerline Crack	Porosity	Root Crack	Slag Inclusion	Outside Crack	Lack of Fusion	
Faster R-CNN (Ours)	*AP* (%)	49.46	59.02	66.38	18.92	79.73	70.08	58.05
	Classification (%)	2.05	76.87	53.86	7.64	81.05	44.58	48.42
YOLOv5	*AP* (%)	49.78	81.28	66.99	7.73	78.29	58.13	57.03
	Classification (%)	1.03	96.05	6.30	20.31	89.15	15.11	36.62
YOLOv8	*AP* (%)	55.18	58.35	73.93	28.91	59.05	62.81	56.37
	Classification (%)	0.42	85.54	3.92	32.08	91.94	58.61	41.69

Cracks from Lack of Fusion defects, both being root-located flaws, resulting in moderate classification performance.

2.4.1 Ablation study and impact of global information on detection performance

2.4.1.1 Ablation study

To systematically evaluate the contribution of each architectural component, an ablation study was conducted (Table 2.2). This analysis aimed to quantify the individual impact of key design choices on the overall detection performance. A progressive approach was followed in the study, where components were incrementally added to establish their individual contributions. Each configuration was evaluated on the three different datasets.

Data augmentation was the first step in our approach, addressing the limited diversity in PAUT datasets. This approach enhanced the robustness of our dataset, using small rotations mimicking probe jitter and horizontal flips simulating bilateral acquisition. Consequently, the accuracy improved from 52.10% to 53.11% for the median subtracted dataset.

The integration of ROIAlign was a logical enhancement over the standard ROIPooling mechanism. ROIAlign addresses the quantization errors inherent in ROIPooling by using bilinear interpolation, which is particularly important for PAUT images where defect boundaries are often subtle and require precise localization. It yielded modest improvements for raw images (from 31.94% to 32.98% mAP), though with increased variance (1.37 and 3.24), suggesting sensitivity to specific defect configurations. However, for median-subtracted data, ROIAlign provided consistent gains (+1.93%) with stable variance (around 2.5%), indicating more reliable performance when structural artifacts are suppressed.

The replacement of the traditional ResNet backbone with EfficientNet B5 coupled with Feature Pyramid Network (FPN) represents a significant architectural upgrade. EfficientNet B5 employs a compound scaling approach that simultaneously optimizes network depth, width, and resolution, while FPN enhances multi-scale feature extraction by leveraging information from different feature maps with reduced computational overhead. This combination delivers substantial performance improvements of +9.63% mAP (44.43% versus 34.8%) and +9.72% in classification accuracy for raw images. The compound scaling ensures that the network can capture both fine-grained local features and broader contextual information essential for PAUT analysis and therefore gives better precision and classification results.

The optimization of anchor boxes using K-means clustering on the training dataset showed dataset-dependent effects : Raw and Overlay-added images showed an improvement of about 3% in precision, while the mAP of Median-subtracted images increased by 0.91%.

The modest improvement on median-subtracted data, despite lower overall variance, suggests the simplified defect signatures after geometry subtraction may already align well with default anchors, reducing the benefit of custom optimization.

K-means optimization adapts anchor shapes to PAUT-specific defect geometries (elongated cracks, irregular porosities). This data-driven customization proved most beneficial for complex backgrounds (raw : +3.03%, overlay : +3.44%), where precise region proposals are critical, but

offered less results when defect signatures are already simplified through median subtraction (+0.91%).

Tableau 2.2 Ablation Study : Impact of Global Information on Detection Performance

Architecture	**Metric**	**Datasets**			**Params (M)**	**FLOPs (G)**
		Raw Images	**Overlay Added**	**Median Subtracted**		
Faster R-CNN (baseline)	*mAP* (%)	31.94±1.37	35.23±2.02	52.10±2.48	165.22	103.39
	Classification (%)	23.38±4.77	20.68±2.93	33.48±3.44		
Baseline + Data Aug.	*mAP* (%)	32.36±3.10	35.00±3.43	53.11±3.09	165.22	103.39
	Classification (%)	26.58±1.95	23.17±0.99	32.51±2.42		
+ ROIAlign	*mAP* (%)	32.98±3.24	33.35±2.16	54.03±2.47	165.22	121.26
	Classification (%)	26.62±3.10	21.57±1.64	36.49±2.28		
+ EfficientNet B5 (FPN)	*mAP* (%)	43.3±3.46	32.48±2.91	57.14±2.90	45.04	98.47
	Classification (%)	**37.21**±4.72	27.02±1.85	44.27±3.19		
+ K-means	*mAP* (%)	46.33±3.68	**35.92**±2.37	**58.05**±0.87	45.04	98.47
	Classification (%)	35.82±2.86	**29.61**±2.45	**48.42**±1.94		
YOLOv8	*mAP* (%)	**47.44**	29.07	56.37	25.90	9.89
	Classification (%)	36.90	21.22	41.69		
YOLOv5	*mAP* (%)	31.89	28.32	57.03	25.11	8.05
	Classification (%)	25.79	24.20	36.62		

2.4.1.2 Impact of global information

Experiments were also carried out using the different previously generated datasets (Fig 2.6). The first observation is that adding overlays generally did not improve the model's performance independently of the chosen model. For example, YOLOv5 showed a 3.57% reduction in average precision and 1.59% in classification accuracy when overlays were used compared to the raw dataset. Meanwhile, our Faster R-CNN decreased in precision by 10.41% and 6.21% in classification accuracy.

This performance drop can be explained by the intrinsic limitations of overlays. Their presence tends to cause the model to overfit. Although artificial data augmentation was applied through FMC-based reconstructions, it was insufficient to compensate for the limited diversity in the training dataset. As defects remained at the same relative positions with respect to the overlay,

with only their intensity and shape varying, the model tended to associate specific image locations with defects. Consequently, when evaluated on a distinct validation dataset, its performance deteriorated, revealing a lack of generalization capability. Moreover, overlays masked small defects such as porosities, further reducing detection performance.

Nevertheless, the subtraction of the specimen's median geometry gave more effective results than overlays, as it produced smoother input data and facilitated feature extraction by CNNs. Using this method, mAP improved by 12 percentage points without compromising classification accuracy (also improved by around 12 percentage points) compared to raw data. Furthermore, noise reduction through median subtraction significantly decreased the number of false positives, resulting in overall better performance. While the median subtraction method demonstrated superior performance, it is important to acknowledge the practical constraints of both contextual approaches. Median subtraction requires multiple acquisitions for effective geometry estimation, and overlay integration demands manual annotation of each image, adding operational overhead. However, these requirements are consistent with established inspection workflows, where inspectors routinely perform multiple scans and incorporate geometric context during evaluation.

2.5 Conclusion

This study investigated the integration of deep learning architectures for automated flaw detection and characterization in PAUT, addressing the critical challenge of reducing interpretation variability in non-destructive testing. Through comprehensive experimentation with Faster R-CNN and YOLOv5 and v8 architectures across three distinct dataset configurations, several key findings emerge.

The enhanced Faster R-CNN architecture, incorporating the EfficientNet B5 backbone with a Feature Pyramid Network and optimized anchor boxes, demonstrated superior performance with a mean Average Precision (mAP) of 58.05% on median-subtracted data. The systematic ablation study revealed that architectural enhancements, particularly the compound scaling approach of

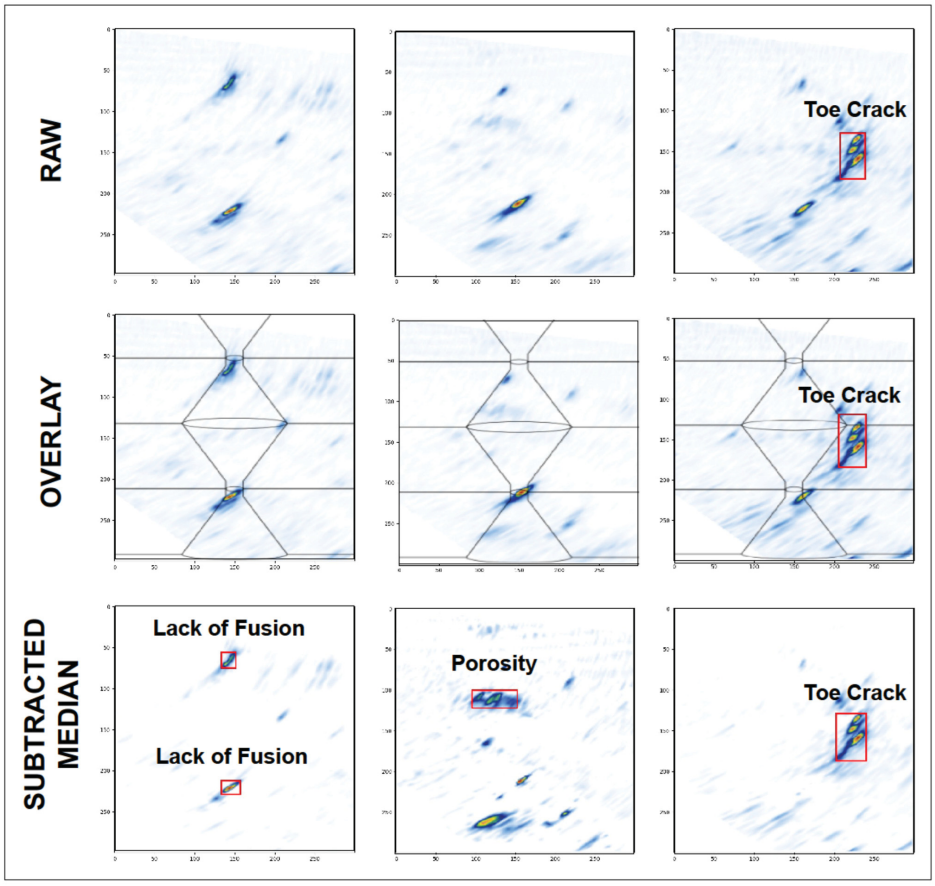


Figure 2.6 Comparative detection results across preprocessing strategies for the same defect instance. Each column represents identical PAUT acquisitions under different preprocessing. Only the median-subtracted configuration achieved consistent perfect detection and classification across all test cases.

EfficientNet B5 combined with FPN, provided the most significant performance gains (up to +10% mAP), highlighting the importance of multi-scale feature extraction for PAUT analysis.

Among the three data preparation strategies evaluated, median geometry subtraction proved most effective, improving mAP by 10 percentage points without compromising classification accuracy compared to raw data. This preprocessing technique successfully reduced structural artifacts while preserving defect signatures, facilitating better feature extraction by CNNs. Conversely, overlay annotations, despite providing contextual information similar to human inspector workflow, failed to improve performance and hindered detection of small defects,

suggesting that explicit geometric overlays may not be the optimal approach for AI-based PAUT analysis.

The study also revealed specific challenges in detecting certain flaw types, particularly centerline cracks and slag inclusions, which achieved lower detection rates due to their similarity to specimen geometry and variations between training and validation datasets. This highlights the critical importance of dataset consistency and the need for more diverse training data to achieve robust generalization across different inspection scenarios.

Future work should focus on the following key areas : First, expanding the dataset diversity through additional specimen types and acquisition conditions to improve model generalization. Second, looking for better architectures for detection such as Vision Transformer (ViT) (Dosovitskiy *et al.* (2021)), which could potentially capture long-range dependencies such as the artefacts associated with the geometry in PAUT images more effectively than CNN-based approaches. Finally, exploring multi-view fusion approaches (Beery, Wu, Rathod, Votel & Huang (2020)) that can leverage multiple PAUT configurations simultaneously, mimicking the comprehensive evaluation process used by human inspectors.

CONCLUSION ET RECOMMANDATIONS

Ce travail de recherche visait à développer un système de détection automatisée de défauts dans les images PAUT pour assister les inspecteurs en contrôle non destructif. L'objectif était de démontrer que l'intégration d'architectures avancées d'apprentissage profond, combinée à des stratégies de prétraitement contextuel, pouvait améliorer significativement la détection et la caractérisation des défauts dans les soudures, tout en réduisant la variabilité d'interprétation liée à l'expertise humaine.

Une architecture Faster R-CNN optimisée a été développée, intégrant un réseau d'extraction de caractéristiques EfficientNet-B5 avec réseau pyramidal de caractéristiques (FPN), l'alignement précis des régions d'intérêt (ROI Align) et l'optimisation des boîtes d'ancrage par K-means. Ce modèle a montré des performances supérieures par rapport aux architectures YOLO (v5 et v8), atteignant une précision moyenne (mAP) de 58.05 % lorsque combiné avec la soustraction médiane de géométrie. Cette approche surpasse les modèles YOLO testés, qui obtiennent respectivement 57.03 % (YOLOv5) et 56.37 % (YOLOv8) sur le même jeu de données.

L'étude d'ablation systématique a révélé que l'amélioration architecturale la plus déterminante provient de l'adoption d'EfficientNet-B5 combiné au FPN, apportant un gain de +5.04 % en mAP et +10.79% en classification par rapport à l'architecture de base avec ResNet-50 avec la soustraction de la médiane et +10.32% en mAP pour les images sans preprocessing. Cette approche de mise à l'échelle composée permet une extraction efficace de caractéristiques multi-échelles, essentielle pour détecter des défauts de tailles et de formes variées dans les images PAUT.

L'analyse de trois configurations de données distinctes a permis d'évaluer l'influence de l'information contextuelle sur les performances de détection. La soustraction de la géométrie médiane s'est révélée être la stratégie la plus efficace, améliorant le mAP et la classification de 10 points de pourcentage tout en réduisant significativement les faux positifs. Cette méthode, qui

supprime les artefacts structurels en préservant les signatures des défauts, facilite l'extraction de caractéristiques par les CNN sans compromettre la précision et la classification.

À l'inverse, l'ajout d'overlays schématiques, bien qu'aligné avec les pratiques des inspecteurs, n'a pas amélioré les performances et a même parfois nui à la détection. Cette observation contre-intuitive suggère que l'intégration explicite d'informations géométriques peut induire un surapprentissage lorsque la diversité du jeu de données est limitée, le modèle associant alors des positions spécifiques de l'image à la présence de défauts plutôt que d'apprendre des caractéristiques généralisables.

Certains types de défauts, notamment les fissures au centre de la soudure, ont présenté des taux de détection significativement plus faibles. Ces difficultés s'expliquent principalement par leur similarité avec la géométrie du spécimen et par les variations entre les jeux de données d'entraînement (FlawTech) et de validation (Sonaspection). Cette observation souligne l'importance cruciale de la cohérence et de la diversité des données pour assurer une généralisation robuste du modèle.

De plus, la distinction entre certains types de défauts partageant des localisations similaires, comme les fissures à la racine de la soudure et les manques de fusion, demeure un défi. Cette limitation rappelle que même les architectures avancées nécessitent des données d'entraînement suffisamment représentatives pour capturer la subtilité des différences entre classes de défauts.

Des perspectives d'amélioration existent encore, notamment en explorant d'autres architectures avancées et en affinant les stratégies de prétraitement. Parmi les axes de recherche prometteurs :

L'exploration des Vision Transformers (ViT) pourrait offrir de nouvelles perspectives pour capturer les dépendances à longue portée dans les images PAUT, potentiellement mieux que les approches CNN traditionnelles. Ces architectures basées sur l'attention pourraient

particulièrement exceller dans la différenciation entre défauts et artefacts géométriques, tout en offrant une meilleure interprétabilité des décisions du modèle.

Le développement d'approches de fusion multi-vues, permettant d'exploiter simultanément plusieurs configurations PAUT (différents angles, focalisations et positions de sonde), représente une direction intéressante. Cette stratégie imiterait le processus d'évaluation complet utilisé par les inspecteurs humains, qui s'appuient sur plusieurs vues complémentaires avant de formuler un diagnostic définitif. L'utilisation des acquisitions FMC pourrait faciliter cette approche en fournissant toutes les informations nécessaires pour reconstruire diverses configurations d'imagerie. Cependant, un plus grand nombre de défaut et de blocs est nécessaire pour obtenir des résultats utilisables.

L'expansion de la diversité du jeu de données, tant en termes de types de spécimens (différentes épaisseurs, géométries de soudure, matériaux) que de configurations d'acquisition, permettrait d'améliorer significativement la capacité de généralisation des modèles. L'intégration de données provenant de multiples fabricants et incluant une plus grande variété de types de défauts renforcerait la robustesse du système.

BIBLIOGRAPHIE

Beery, S., Wu, G., Rathod, V., Votel, R. & Huang, J. *(2020)*. Context R-CNN : Long Term Temporal Context for Per-Camera Object Detection. arXiv. Repéré le 2025-05-08 à http://arxiv.org/abs/1912.03538.

Cegla, F. & Allin, J. (2015). Ultrasonic Monitoring of Pipeline Wall Thickness with Autonomous, Wireless Sensor Networks. Dans Revie, R. W. (Éd.), *Oil and Gas Pipelines* (éd. 1, pp. 571–578). Wiley. doi : 10.1002/9781119019213.ch39.

Cheeke, J. (2002). *Fundamentals and Applications of Ultrasonic Waves*.

Chen, C., Wang, S. & Huang, S. (2023). An improved faster RCNN-based weld ultrasonic atlas defect detection method. 56(3), 832–843. doi : 10.1177/00202940221092030.

Chen, Y., He, D., He, S., Jin, Z., Miao, J., Shan, S. & Chen, Y. (2024). Welding defect detection based on phased array images and two-stage segmentation strategy. 62, 102879. doi : 10.1016/j.aei.2024.102879.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. *(2021)*. An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale. arXiv. Repéré le 2025-04-15 à http://arxiv.org/abs/2010.11929.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. 88(2), 303–338. doi : 10.1007/s11263-009-0275-4.

Girshick, R. *(2015)*. Fast R-CNN. arXiv. Repéré le 2025-01-30 à http://arxiv.org/abs/1504.08083.

Glorot, X., Bordes, A. & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press.

He, K., Zhang, X., Ren, S. & Sun, J. *(2015)*. Deep Residual Learning for Image Recognition. arXiv. Repéré le 2024-12-02 à http://arxiv.org/abs/1512.03385.

Hervé-Côte, H. (2023). Détection automatique des défauts dans les balayages sectoriels à l'aide de l'apprentissage automatique. 87.

Hervé-Côte, H., Dupont-Marillia, F. & Bélanger, P. (2024). Automatic flaw detection in sectoral scans using machine learning. 141, 107316. doi : 10.1016/j.ultras.2024.107316.

Holmes, C., Drinkwater, B. W. & Wilcox, P. D. (2005). Post-processing of the full matrix of ultrasonic transmit–receive array data for non-destructive evaluation. 38(8), 701–711. doi : 10.1016/j.ndteint.2005.04.002.

Huang, G., Liu, Z., Maaten, L. v. d. & Weinberger, K. Q. *(2019)*. Densely Connected Convolutional Networks. arXiv. Repéré le 2025-02-03 à http://arxiv.org/abs/1608.06993.

Ioffe, S. & Szegedy, C. *(2015)*. Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv. Repéré le 2024-12-02 à http://arxiv.org/abs/1502.03167.

Jocher, G. (2020). Ultralytics YOLOv5 (Version 7.0). Repéré à https://github.com/ultralytics/yolov5.

Jocher, G., Chaurasia, A. & Qiu, J. (2023). Ultralytics YOLOv8 (Version 8.0.0). Repéré à https://github.com/ultralytics/ultralytics.

Kang, D., Choi, Y. M., Lee, D. M., Kim, J. B., Kim, Y. K., Park, T. S. & Park, I. K. (2023). Reliability Analysis of PAUT Based on the Round-Robin Test for Pipe Welds with Thermal Fatigue Cracks. 16(21), 6908. doi : 10.3390/ma16216908.

Khanam, R. & Hussain, M. *(2024)*. What is YOLOv5 : A deep look into the internal features of the popular object detector. arXiv. Repéré le 2025-09-03 à http://arxiv.org/abs/2407.20892.

Kingma, D. P. & Ba, J. *(2017)*. Adam : A Method for Stochastic Optimization. arXiv. Repéré le 2025-09-11 à http://arxiv.org/abs/1412.6980.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25. Repéré à https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Latête, T., Gauthier, B. & Belanger, P. (2021). Towards using convolutional neural network to locate, identify and size defects in phased array ultrasonic testing. 115, 106436. doi : 10.1016/j.ultras.2021.106436.

Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition. 86(11), 2278–2324. doi : 10.1109/5.726791.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. & Belongie, S. *(2017)*. Feature Pyramid Networks for Object Detection. arXiv. Repéré le 2025-01-09 à http://arxiv.org/abs/1612.03144.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. & Berg, A. C. (2016). SSD : Single Shot MultiBox Detector. Dans Leibe, B., Matas, J., Sebe, N. & Welling, M. (Éds.), *Computer Vision – ECCV 2016* (vol. 9905, pp. 21–37). Springer International Publishing. doi : 10.1007/978-3-319-46448-0_2.

Medak, D., Posilovic, L., Subasic, M., Budimir, M. & Loncaric, S. (2021). Automated Defect Detection From Ultrasonic Images Using Deep Learning. 68(10), 3126–3134. doi : 10.1109/TUFFC.2021.3081750.

Munir, N., Kim, H.-J., Park, J., Song, S.-J. & Kang, S.-S. (2019). Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions. 94, 74–81. doi : 10.1016/j.ultras.2018.12.001.

Murphy, K. P. (2014). *Machine Learning - A Probabilistic Perspective*. MIT Press.

Norris, W., Naus, D. & Graves, H. (1999). Inspection of nuclear power plant containment structures. 192(2), 303–329. doi : 10.1016/S0029-5493(99)00125-9.

Posilovic, L., Medak, D., Subasic, M., Petkovic, T., Budimir, M. & Loncaric, S. (2019). Flaw Detection from Ultrasonic Images using YOLO and SSD. *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 163–168. doi : 10.1109/ISPA.2019.8868929.

Provencal, E. & Laperrière, L. (2021). Identification of weld geometry from ultrasound scan data using deep learning. 104, 122–127. doi : 10.1016/j.procir.2021.11.021.

R A Mountford & P N T Wells. (1972). Ultrasonic liver scanning : the A-scan in the normal and cirrhosis. 17(2), 261–269. doi : 10.1088/0031-9155/17/2/012.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. *(2016)*. You Only Look Once : Unified, Real-Time Object Detection. arXiv. Repéré le 2025-02-05 à http://arxiv.org/abs/1506.02640.

Ren, S., He, K., Girshick, R. & Sun, J. *(2016)*. Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks. arXiv. Repéré le 2024-03-21 à http://arxiv.org/abs/1506.01497.

Robbins Herbert. (1951). A Stochastic Approximation Method.

Rosenblatt, F. (1958). The perceptron : A probabilistic model for information storage and organization in the brain. 65(6), 386–408. doi : 10.1037/h0042519.

Shull, P. J. *(2002)*. Nondestructive evaluation : theory, techniques, and applications. M. Dekker.

Simonyan, K. & Zisserman, A. *(2015).* Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. Repéré le 2025-02-05 à http://arxiv.org/abs/1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout : A Simple Way to Prevent Neural Networks from Overfitting.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. *(2014).* Going Deeper with Convolutions. arXiv. Repéré le 2025-02-03 à http://arxiv.org/abs/1409.4842.

Tan, M. & Le, Q. V. *(2020).* EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks. arXiv. Repéré le 2024-12-02 à http://arxiv.org/abs/1905.11946.

Varghese, R. & M., S. (2024). YOLOv8 : A Novel Object Detection Algorithm with Enhanced Performance and Robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pp. 1–6. doi : 10.1109/ADICS58448.2024.10533619.

Xu, B., Wang, N., Chen, T. & Li, M. *(2015).* Empirical Evaluation of Rectified Activations in Convolutional Network. arXiv. Repéré le 2025-09-09 à http://arxiv.org/abs/1505.00853.

Zhang, H., Chen, Z., Zhang, C., Xi, J. & Le, X. (2019). Weld Defect Detection Based on Deep Learning Method. *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pp. 1574–1579. doi : 10.1109/COASE.2019.8842998.

Zhang, K. & Shen, H. (2021). Solder Joint Defect Detection in the Connectors Using Improved Faster-RCNN Algorithm. 11(2), 576. doi : 10.3390/app11020576.

Zhou, Z., Lu, Q., Wang, Z. & Huang, H. (2019). Detection of Micro-Defects on Irregular Reflective Surfaces Based on Improved Faster R-CNN. 19(22), 5000. doi : 10.3390/s19225000.

Zhu, Ge & Liu. (2019). Deep Learning-Based Classification of Weld Surface Defects. 9(16), 3312. doi : 10.3390/app9163312.