Réseaux neuronaux convolutifs complexes pour l'imagerie échographique rapide par ondes divergentes

Jingfeng LU[1,2], Fabien MILLIOZ[1], Damien GARCIA[1], Sébastien SALLES[1], Dong YE[2], Denis FRIBOULET[1]

[1]Université de Lyon, CREATIS, CNRS UMR 5220, INSERM U1294, Université Claude Bernard Lyon 1, INSA-Lyon
7 Avenue Jean Capelle, 69621 Villeurbanne Cedex, France

[2]School of Instrumentation Science and Engineering, Harbin Institute of Technology, Harbin, China

[1]prénom.nom@creatis.insa-lyon.fr

Résumé – L'imagerie par Onde Divergente (OD) produit des images ultrasonores à une fréquence d'acquisition élevée (ultrarapide) mais de faible qualité. L'imagerie OD conventionnelle de haute qualité repose sur la composition cohérente de plusieurs émissions consécutives, ce qui réduit le gain en fréquence d'acquisition. La reconstruction d'image par réseaux de neurones convolutifs permet d'obtenir une image de haute qualité avec moins d'émissions que les méthodes traditionnelles. Nous avons récemment décrit un tel réseau, appelée ID-Net. Nous dérivons dans ce travail l'équivalent complexe de ce réseau, fonctionnant sur des données en phase/quadrature (IQ) sous-échantillonnées plutôt que les données réelles radiofréquences (RF). Nous démontrons expérimentalement qu'un réseau traitant des données complexes est plus efficace s'il tient compte explicitement des relations entre les parties réelles et complexes des données. Un tel réseau produit des résultats équivalents au réseau entraîné avec les données RF, avec un temps d'entraînemet réduit d'un facteur 5.

Abstract – Diverging wave (DW) imaging produces ultrasound images at high frame rate (ultrafast) but of low quality. Conventional high-quality DW imaging relies on the coherent compounding of multiple consecutive steered emissions, reducing the gain in frame rate. Reconstruction using neural networks leads to high-quality images with less emissions than usual techniques. We recently described such a network, called ID-Net. We derive here a complex equivalent of this network, using subsampled In Phase/Quadrature data (IQ) instead of real radiofrequency data (RF). We experimentally show that a network using complex data is more efficient if it explicitly takes into account the relations between real and imaginary parts of the data. Such a network is as good as the network trained on RF data, with a training time 5 time shorter.

1 Introduction

Dans le domaine de l'imagerie médicale, l'imagerie ultrasonore se démarque par le fait qu'elle n'utilise pas de rayonnements ionisants, qu'elle a un faible coût et permet une haute cadence d'images. Pour ce faire, plutôt qu'utiliser des ondes focalisées, qui nécessitent un nombre d'émissions et de réceptions de signaux proportionnel à la largeur de l'image voulue, il est possible d'utiliser des ondes divergentes (OD) [1, 2], insonifiant tout le milieu à imager simultanément. La cadence d'image peut atteindre plusieurs milliers d'images par seconde, alors que la méthode focalisée n'en obtient que quelques dizaines. Toutefois, ceci se fait au détriment de la qualité de l'image, dont le rapport signal à bruit diminue drastiquement. Un compromis est alors trouvé en insonifiant le milieu selon différents angles d'émissions, et en combinant de manière cohérente (*coherent compounding*) les images obtenues après formation de voies. Une trentaine d'angles d'émission suffisent à obtenir une qualité équivalente à l'image classique.

Récemment, l'utilisation des techniques d'apprentissage profond pour améliorer l'imagerie ultrasonore a suscité un intérêt croissant. La plupart des études existantes utilisent des signaux radiofréquence (RF) [3, 4, 5], alors que les systèmes échographiques modernes permettent l'utilisation des signaux complexes en phase/quadrature (IQ), obtenus en démodulant dans la bande de base les signaux RF. Ces derniers permettent de sous-échantillonner grandement les signaux captés.

L'apprentissage automatique complexe pour l'imagerie ultrasonore n'a été développé que récemment [6]. Cet article en est une présentation brève. Dans de précédents travaux [7], nous avons décrit le réseau ID-Net, permettant d'obtenir une image de bonne qualité à partir de 3 images OD d'entrées, en signaux RF. Dans ce papier, nous modifions ce réseau pour travailler avec les signaux complexes IQ, sous le nom de CID-Net, en tenant compte explicitement des relations entre les parties réelles et imaginaires des signaux, suivant [8, 9]. Un troisième réseau, 2BID-Net, traite les deux branches partie réelle et partie imaginaire séparément afin de mettre en valeur l'intérêt d'expliciter la relation complexe.

La section 2 présente les spécificités du réseau lié à son caractère complexe, ainsi que son architecture. La section 3 présente les données d'entraînement, détaille la méthode d'entraînement des réseaux, et présente les métriques de comparaisons utilisées. Les résultats sont présentés en section 4, suivis par une conclusion en section 5.

2 Méthodes

Soit $X \in \mathbb{C}^{m\times w\times h}$ un tenseur qui représente un faible nombre m d'images provenant d'acquisitions d'OD après formation de voies, chacune composée de w signaux IQ de longueur h. La reconstruction d'image OD consiste à estimer une image IQ de haute qualité $\hat{Y} \in \mathbb{C}^{w\times h}$ à partir de la basse qualité X. Nous proposons d'utiliser un réseau de neurones convolutif CID-net afin d'obtenir l'opérateur de reconstruction optimal $f(\Theta) : \mathbb{C}^{m\times w\times h} \mapsto \mathbb{C}^{w\times h}$, en entraînant les paramètres Θ complexes pour obtenir une image cible $\hat{Y} \in \mathbb{C}^{w\times h}$ obtenue par combinaison cohérente de $n \gg m$ acquisitions OD.

2.1 Convolutions complexes

Nous avons utilisé des matrices à valeurs réelles pour représenter les composantes réelles et imaginaires de l'image, et nous avons effectué les convolutions complexes en utilisant l'arithmétique à valeurs réelles. En notant $X = X_r + jX_i$ une image complexe, où $j = \sqrt{-1}$ est l'unité imaginaire, X_r et X_i respectivement les parties réelles et imaginaires de X, et de la même manière $W = W_r + jW_i$ un noyau de convolution, la convolution de X par W s'écrit :

$$\begin{aligned} Z &= W * X \\ &= (W_r * X_r - W_i * X_i) + j(W_r * X_i + W_i * X_r) \end{aligned} \quad (1)$$

Il est ainsi possible de choisir une structure de réseau qui impose la relation de convolution complexe entre les poids appris et les données d'entrée.

2.2 Fonction d'activation

La généralisation au domaine complexe de la fonction d'activation la plus plus courante, l'unité linéaire rectifiée (ReLU), n'est pas triviale : dans [9], trois types d'activations complexes basées sur ReLU ont été étudiées. Comme l'objectif du travail est de comparer ID-Net [7] avec sa version complexe, nous nous sommes concentrés sur la conception d'une version complexe de la fonction d'activation *MaxOut*. Une fonction *MaxOut* habituelle prend en entrée plusieurs valeurs réelles, et renvoie la plus grande de ces valeurs en sortie. L'opération de maximum doit être définie pour le cas complexe. Afin de garder l'interaction souhaitée entre les parties réelles et imaginaires de nos données, nous définissons le *MaxOut* d'Amplitude (MOA) qui renvoie en sortie l'entrée de plus grand module.

2.3 Architectures des réseaux

CID-Net[1] est une version complexe de ID-Net; ainsi, il comporte trois couches de convolutions avec des noyaux de tailles croissantes, activées par des MOA. Sa particularité est d'avoir une couche d'inception en tant que quatrième couche, qui consiste en plusieurs convolutions en parallèle, avec des noyaux de tailles différentes. Nous avons montré [7] qu'une couche d'inception suivie par un *MaxOut* permet de s'adapter à la géométrie non-isotrope des images OD. Les détails de l'architecture sont données dans le tableau 1.

Table 1: Architectures de CID-Net et 2BID-Net.

Couche	Taille des données	nombre de noyaux	taille des noyaux
Entrée	$3 \times h \times w$	–	–
Convolution	$64 \times h \times w$	256	3×3
Convolution	$32 \times h \times w$	128	5×5
Convolution	$16 \times h \times w$	64	11×9
Inception	$8 \times h \times w$	8	15×11
		8	17×13
		8	19×15
		8	21×17
Convolution	$1 \times h \times w$	4	1×1

Nous comparons ce réseau avec un modèle ID-Net en utilisant des données réelles RF, ainsi qu'un modèle ID-Net à deux branches (2BID-Net) où, comme dans [3, 5], chaque branche a été entraînée séparément sur les parties réelles et imaginaires des données IQ. Pour une comparaison équitable, les trois réseaux ont le même nombre de convolutions dans chaque couche. 2BID-Net a les mêmes tailles de noyaux CID-Net; pour obtenir la même taille de champ réceptif physique sur les images RF, ID-Net a utilisé des tailles plus grandes dans la dimension axiale, respectivement 9×3, 17×5, 33×9, pour les convolutions, et 41×11, 49×13, 57×15 et 67×17 pour la couche inception. Les activations sont toutes des *MaxOut*, d'amplitude pour 2BID-Net.

3 Expérimentation

3.1 Acquisition des données

Nous avons effectué des acquisitions OD selon plusieurs angles en utilisant un scanner de recherche Verasonics (Vantage 256) équipé d'une sonde ATL P4-2 (bande passante : 2-4 MHz, fréquence centrale : 3 MHz). La sonde a été déplacée manuellement, sur des surfaces in-vitro et in-vivo, afin de générer une large gamme d'images significativement différentes. Chaque image a été obtenue à l'aide de 31 OD orientées selon des angles allant de $-30°$ à $+30°$ par pas de $2°$. Les données brutes reçues ont été démodulées en IQ, puis sous-échantillonnées par un facteur de 3. La formation de voies se fait par un système retard et somme.

L'entrée X du CID-Net est constituée de $m = 3$ images IQ, obtenues pour les angles $-20°$, $0°$ et $+20°$, tandis que la cible Y est obtenue par combinaison cohérente des images issues des 31 angles.

7500 couples (X, Y) ont été obtenus : 1500 acquis depuis

[1]Le modèle CID-Net au format textitOpen Neural Network Exchange (ONNX), avec les poids entraînés, et un script de démonstration sont disponibles sur https://github.com/Jingfeng-LU/CID-Net/

trois sujets sains (muscle de la cuisse, phalange du doigts, régions du foie), et 6000 acquis à partir de deux fantômes (Gammex 410SCG et CIRS 054SG). 5000 de ces couples ont servi à l'entraînement du réseau, 1250 à sa validation, et les 1250 restants pour son test.

3.2 Entraînement des réseaux

L'entraînement a été mis en œuvre avec la bibliothèque Pytorch, sur un GPU NVIDIA Tesla V100 avec 32 Go de mémoire. L'erreur quadratique moyenne a été choisie comme fonction de coût; les poids ont été entraînés par descente de gradient avec l'optimiseur Adam avec un taux d'apprentissage initial de 10^{-4} par groupes (*batch*) de 16, et initialisés par la méthode Xavier.

Pendant l'entraînement, le taux d'apprentissage a été divisé par deux si aucune diminution du coût de validation n'a eu lieu pendant 20 époques, et 40 époques consécutives sans réduction de ce coût met fin à l'entraînement, qui a nécessité deux jours d'entraînement.

3.3 Métriques d'évaluation

Les métriques d'évaluation dans cet article ont pour but de comparer les similarités entre l'image $\hat{Y}$ obtenue par le réseau et l'image cible Y de haute qualité.

Le PSNR (*Peak Signal to Noise Ratio*) mesure à l'échelle du pixel la ressemblance entre deux images. Il est défini comme étant :

$$\text{PSNR} = 10\log_{10}\frac{MAX_y^2}{1/N\|\hat{Y}-Y\|_2^2}, \tag{2}$$

avec MAX_y la valeur maximale de l'image Y, N est son nombre de pixels et $\|\cdot\|_2$ est la norme l_2

L'indice de similarité structurelle (SSIM) mesure la similarité des structures entre deux images. Il est défini par :

$$\text{SSIM} = \frac{(2\mu_{\hat{Y}}\mu_Y + C_1)(2\sigma_{\hat{Y}Y} + C_2)}{(\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1)(\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2)}, \tag{3}$$

avec $\mu_{\hat{Y}}$ et μ_Y ($\sigma_{\hat{Y}}^2$ et σ_Y^2) respectivement les moyennes (variances) de $\hat{Y}$ et Y, $\sigma_{\hat{Y}Y}$ la covariance entre $\hat{Y}$ et Y, C_1 et C_2 deux constantes pour stabiliser la division quand le dénominateur est très petit.

L'information mutuelle (MI) mesure la dépendance statistique entre les deux images. Elle est définie par :

$$\text{MI} = \sum_{\hat{y},y} p_{\hat{Y}Y}(\hat{y},y)\log\frac{p_{\hat{Y}Y}(\hat{y},y)}{p_{\hat{Y}}(\hat{y})p_Y(y)}, \tag{4}$$

avec $p_{\hat{Y}Y}(\hat{y},y)$ la loi de probabilité jointe de $\hat{Y}$ et Y, $p_{\hat{Y}}(\hat{y})$ et $p_Y(y)$ respectivement les lois de probabilité marginales de $\hat{Y}$ et Y.

D'autres métriques indépendantes de la référence (taux de contraste, rapport contraste à bruit, rapport contraste à bruit généralisé et résolution latérale), utilisées comme références en imagerie ultrasonore, sont présentées dans [6].

Table 2: PSNR, SSIM, et MI obtenus sur les données de test

Réseau	PSNR [dB]	SSIM	MI
ID-Net (3 RF)	$\mathbf{31.57 \pm 1.25}$	$\mathbf{0.92 \pm 0.06}$	$\mathbf{0.81 \pm 0.23}$
CID-Net (3 IQ)	31.56 ± 1.39	$\mathbf{0.92 \pm 0.06}$	$\mathbf{0.81 \pm 0.24}$
2BID-Net (3 IQ)	30.47 ± 1.35	0.86 ± 0.11	0.73 ± 0.25

Table 3: Nombre d'échantillons axiaux Nb_A, nombre de paramètres Nb_P, nombre d'opérations par seconde FLOPs, durée d'entraînement T_{ent} et temps pour l'inférence sur CPU T_{inf} pour les réseaux considérés.

Réseau	Nb_A	Nb_P	FLOPs	T_{ent}	T_{inf}
ID-Net	1013	1.7e6	23.8e9	5.5 jours	2.85 s
CID-Net	**341**	**1.1e6**	7.0e9	**1 jour**	1.56 s
2BID-Net	**341**	**1.1e6**	**3.5e9**	**1 jour**	**0.80 s**

3.4 Résultats

Le tableau 2 présente les valeurs moyennes et les écarts-types des PSNR, SSIM et MI obtenus sur les données de test. Les réseaux CID-Net et ID-Net ont des résultats comparables, ce qui implique que les données RF et IQ mènent à des résultats similaires. Le réseau 2BID-Net est celui qui donne de moins bons résultats : forcer un réseau complexe à prendre en compte les interactions entre parties réelle et imaginaire des données est donc utile.

La figure 1 illustre ces résultats via des images de cœur en B-mode, avec leurs métriques associées. Visuellement, les images reconstruites par CID-Net et par ID-Net sont proches de la référence, contrairement à l'image obtenue par combinaison cohérente des images d'entrée des réseaux, et à l'image obtenue par 2BID-Net.

3.5 Temps de calcul

Les résultats obtenus par CID-Net et ID-Net sont équivalents; toutefois, le fait de pouvoir sous-échantillonner les données IQ par rapport aux données RF permet de réduire grandement le nombre d'échantillons Nb_A dans la direction axiale de l'image. Cette réduction permet également de réduire la taille des noyaux de convolution du réseau, ce qui diminue le nombre de paramètres Nb_P du réseau.

Ainsi, le temps d'entraînement T_{ent} du réseau ID-Net, de 5.5 jours est réduit à un jour pour CID-Net et 2BID-Net.

De même, le temps de calcul d'une image de sortie T_{inf} passe de 2.85 s sur CPU à 1.56 s pour CID-Net, et à 0.8 s pour 2BID-Net, ce dernier faisant moins de calcul en ignorant les dépendances entre les parties réelles et imaginaires des données. La parallélisation des calculs possible sur GPU atténue grandement ces différences, les temps de calculs deviennent alors respectivement de 0.76 ms, 0.70 ms et 0.65 ms.

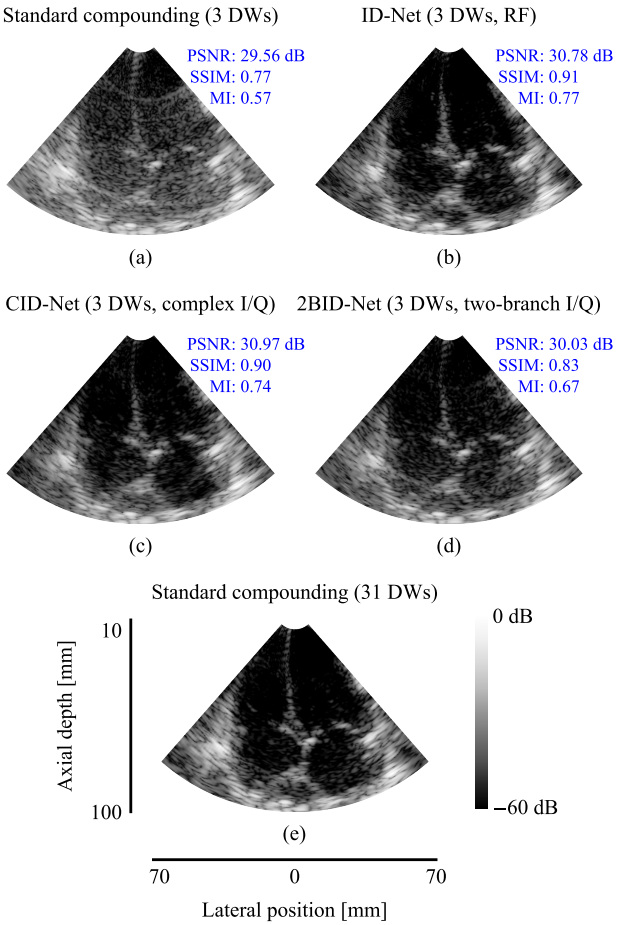

Figure 1: Images cardiaques obtenues : (a) par combinaison cohérente de 3 OD; (b) par ID-Net; (c) par CID-Net; (d) par 2BID-Net (d). L'image de référence est en bas.

4 Conclusion et perspectives

Nous avons proposé un réseau de neurones convolutifs complexe, CID-Net, pour l'imagerie rapide et de haute qualité par ondes divergentes (OD). Le réseau a été entraîné pour reconstruire des images de haute qualité, obtenues par combinaison cohérente de 31 OD, à partir de 3 OD seulement.

Une comparaison avec un réseau travaillant avec les données réelles RF, ID-Net, et un réseau travaillant avec les données complexes IQ mais ignorant les interactions entre les parties réelles et imaginaires des données, 2BID-Net, a été menée.

Les résultats expérimentaux montrent que le CID-Net donne d'aussi bons résultats que ID-Net, contrairement au 2BID-Net, ce qui encourage fortement à explicitement prendre en compte les relations complexes dans un réseau.

De plus, grâce au sous-échantillonnage axial permis par le passage des données RF aux données IQ, le temps d'entraînement est réduit d'un facteur 5 en gardant des résultats similaires, ce qui implique une nette préférence pour travailler avec les signaux IQ.

La suite de ce travail concernera la prise en compte de l'aspect temporel de l'imagerie ultrasonore.

References

[1] H. Hasegawa and H. Kanai, "High-frame-rate echocardiography using diverging transmit beams and parallel receive beamforming," *Journal of medical ultrasonics*, vol. 38, no. 3, pp. 129–140, Jul. 2011.

[2] J. Porée, D. Posada, A. Hodzic, F. Tournoux, G. Cloutier, and D. Garcia, "High-frame-rate echocardiography using coherent compounding with doppler-based motion-compensation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 7, pp. 1647–1657, 2016.

[3] O. Senouf, S. Vedula, G. Zurakhov, A. Bronstein, M. Zibulevsky, O. Michailovich, D. Adam, and D. Blondheim, "High frame-rate cardiac ultrasound imaging with deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 126–134.

[4] S. Vedula, G. Senouf, A. Bronstein, O. Michailovich, and M. Zibulevsky, "Learning beamforming in ultrasound imaging," in *International Conference on Medical Imaging with Deep Learning*, 2019, pp. 493–511.

[5] S. Vedula, O. Senouf, G. Zurakhov, A. Bronstein, M. Zibulevsky, O. Michailovich, D. Adam, and D. Gaitini, "High quality ultrasonic multi-line transmission through deep learning," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2018, pp. 147–155.

[6] J. Lu, F. Millioz, D. Garcia, S. Salles, D. Ye, and D. Friboulet, "Complex convolutional neural networks for ultrafast ultrasound imaging reconstruction from in-phase/quadrature signal," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 69, no. 2, pp. 592–603, Nov. 2021.

[7] J. Lu, F. Millioz, D. Garcia, S. Salles, W. Liu, and D. Friboulet, "Reconstruction for diverging-wave imaging using deep convolutional neural networks," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 12, pp. 2481–2492, Dec. 2020.

[8] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Transactions on Neural Networks and learning systems*, vol. 23, no. 4, pp. 541–551, 2012.

[9] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," in *International Conference on Learning Representations*, 2018.