

Estimation du déphasage et réduction du repliement par apprentissage profond pour l'imagerie ultrasonore

Julia PUIG Fabien MILLIOZ Damien GARCIA Denis FRIBOULET

Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294
prenom.nom@creatis.insa-lyon.fr

Résumé – L'échocardiographie Doppler couleur permet de visualiser le mouvement du sang dans le cœur à partir de la différence de phase de signaux obtenus lors d'acquisitions successives. Le nombre d'acquisitions détermine la cadence d'images maximum atteignable. Nous proposons une méthode basée sur l'apprentissage profond pour l'estimation du déphasage et la réduction du repliement de signaux ultrasonores à partir d'un nombre réduit d'acquisitions successives. Nous avons entraîné puis testé notre approche sur des simulations d'acquisitions échocardiographiques.

Abstract – Color Doppler echocardiography allows visualizing intracardiac blood movement from the phase shift computed over a number of successive signal acquisitions. The number of acquisitions determines the maximum frame rate we can achieve. We propose a deep learning method for phase shift estimation and dealiasing of ultrasound signals from a reduced number of successive acquisitions. We trained and tested our approach on a dataset of simulated echocardiographic acquisitions.

1 Introduction

L'échocardiographie Doppler couleur permet de visualiser simultanément les tissus cardiaques et le déplacement du sang. En clinique, ce mode d'imagerie est utilisé de manière qualitative pour le diagnostic de problèmes cardiaques liés au remplissage du cœur. Cependant, elle ne permet pas d'avoir une cadence temporelle suffisante pour l'évaluation de la vitesse du sang durant un cycle cardiaque. L'augmentation de la résolution temporelle du Doppler couleur ouvrirait donc une nouvelle possibilité de diagnostic, notamment lors de l'évaluation de la fonction diastolique [1].

L'acquisition d'une image Doppler couleur comprend l'échantillonnage des signaux (appelé temps rapide) et la répétition de cette acquisition temporellement (appelé temps lent). Le nombre d'acquisitions successives en temps lent est appelé *packet size* et permet d'estimer le déphasage du signal pour un point donné, et donc la vitesse Doppler associée. L'autocorrélation dans le temps lent est classiquement utilisée pour estimer le déphasage des signaux complexes démodulés en bande de base, nommés signaux I/Q (en phase/en quadrature) [2]. En clinique, le *packet size* est généralement fixé à 8 et une diminution de ce nombre réduit de manière significative la qualité de l'estimation du déphasage par autocorrélation.

Ces dernières années, l'apprentissage profond a été utilisé dans de nombreux travaux en imagerie ultrasonore. Dans [7], les auteurs entraînent un encodeur-décodeur pour l'estimation du Doppler tissulaire à partir de signaux I/Q pour remplacer la méthode classique d'autocorrélation. À partir de données provenant d'un modèle porcin, ils obtiennent une information Doppler comprimée qui permet d'obtenir des estimations de qualité standard. Dans [4], un U-Net a été formé pour identifier et segmenter les régions qui présentent du repliement dans des images Doppler couleur *in vivo* à partir de leurs informations de fréquence Doppler, de puissance et de largeur de bande. La réduction du repliement a ensuite été effectuée par dépliement de phase. Dans [3], un réseau de neurones complexe est en-

traîné pour obtenir une image en mode B de bonne qualité à partir d'un nombre de signaux I/Q réduit. Pour l'estimation du Doppler sanguin, le seul travail antérieur, à notre connaissance, a été présenté à la conférence IUS en 2021, sans acte publié. Dans ce travail, les auteurs entraînent un réseau de neurones convolutionnel de type U-Net pour l'estimation du déphasage à partir de signaux réels non démodulés simulés avec un grand *packet size* de 16 comme référence.

Dans cet article, nous avons remplacé l'autocorrélateur par un réseau de neurones convolutionnel pour obtenir des estimations de déphasage de qualité à partir d'un nombre d'acquisitions réduit ($n = 2$), augmentant ainsi la cadence d'images. Le réseau a été entraîné et évalué sur des simulations de signaux échocardiographiques basées sur des patients.

2 Méthodes

2.1 Base de données d'entraînement

L'entraînement de modèles d'apprentissage profond supervisé nécessite de nombreuses données annotées. Dans le contexte du Doppler couleur, il est impossible d'obtenir des échantillons de référence de haute qualité auprès des patients. Il est alors nécessaire de s'appuyer sur des simulations pour construire un ensemble de données d'entraînement. Pour cela, nous avons utilisé un simulateur d'échocardiographie duplex basé sur le patient [6]. Cette approche utilise les informations du mode B et du Doppler couleur de patients pour modéliser les mouvements du tissu cardiaque et du sang dans le ventricule gauche. Des acquisitions ultrasonores sont alors simulées pour obtenir des signaux ultrasonores. Contrairement au travail présenté à IUS en 2021, nous travaillons donc avec des références théoriques.

Pour notre approche, nous avons modélisé uniquement le mouvement du sang pour obtenir des signaux équivalents aux signaux échocardiographiques obtenus après filtrage du signal du tissu (filtre de *clutter*). Nous avons obtenu des signaux ul-

trasonores I/Q $S \in \mathbb{C}^{180 \times 40 \times n}$ avec un *packet size* de $n = 8$ ainsi que des références de vitesses $D \in \mathbb{R}^{180 \times 40}$, proportionnelles au déphasage. Nous avons ensuite rajouté du bruit gaussien (avec des SNR entre 10 et 25 dB) aux signaux I/Q pour renforcer le réalisme des signaux. Nous avons obtenu 2 576 paires d’entraînement à partir des données échocardiographiques de 37 patients le long de leur cycle cardiaque et pour différents niveaux de bruit. Nous appelons ces données la base de données « patient ».

Pour obtenir la base données d’entraînement pour un *packet size* réduit de $n = 2$, nous avons sélectionné deux acquisitions successives des signaux obtenus avec $n = 8$.

2.2 Apprentissage

Modèle. Nous avons utilisé un réseau de neurones convolucional avec une architecture de type U-Net [5]. L’entrée du réseau est de la forme $S \in \mathbb{R}^{180 \times 40 \times 2n}$, où les parties réelles et imaginaires des signaux I/Q ont été concaténées pour obtenir des représentations réelles. Le réseau comprend 6 couches et un total de 1,5 millions de paramètres.

Entraînement. Pour augmenter notre base de données, dont le nombre d’échantillons est limité par le nombre d’échocardiographies cliniques en notre possession, nous avons utilisé de l’augmentation de données, à savoir des retournements verticaux et horizontaux de nos données. Nous avons entraîné le réseau en validation croisée avec 5 plis. Pour plus de robustesse, les échantillons provenant d’un patient donné étaient dans un même pli. Une fois le pli de *test* choisi, la division des échantillons en *train* et *validation* était de 90/10. Les poids du réseau ont été initialisés avec la distribution Xavier. Nous avons utilisé comme fonction de coût la minimisation de l’erreur quadratique moyenne dans le ventricule (*i.e.*, là où il y a du flux). L’optimisation a été gérée par AdamW. La taille des *batch* a été fixée à 16. La valeur initiale du *learning rate* a été trouvée de manière automatique et il a été diminué dès que la fonction de coût ne diminuait pas pendant 10 époques sur l’ensemble de validation.

Métriques. Pour l’évaluation de l’estimation du déphasage, nous avons utilisé la racine de l’erreur quadratique moyenne ainsi que l’écart-type, que nous appelons respectivement « RMSE (*Root Mean Squared Error*) » et « dispersion » dans les sections suivantes.

3 Résultats

Les performances d’estimation du déphasage de notre méthode et de l’autocorrélateur sont affichées dans le tableau 1 pour les *packet size* standard ($n = 8$) et réduit ($n = 2$). Pour le U-Net, l’écart-type des métriques est donné en fonction du pli.

Notre approche a permis de mieux estimer le déphasage des signaux I/Q, pour $n = 8$ ainsi que pour $n = 2$. Le perte de performance entre $n = 8$ et $n = 2$ pour l’autocorrélateur s’explique en grande partie par la présence de bruit dans les signaux, qu’il n’est pas parvenu pas à éliminer. Au contraire, notre réseau est parvenu à débruiter le signal, même quand $n = 2$, comme nous pouvons le voir sur la figure 1A.

Nous observons également dans le tableau 1 que notre méthode est robuste par la petite variabilité des résultats entre différents plis. Néanmoins, il existe une forte variabilité des

Méthode	Packet size	RMSE	Dispersion
Auto-corrélateur	$n = 8$	9.5	8.6
	$n = 2$	14	10
U-Net	$n = 8$	3.0 ± 0.3	3.0 ± 2.4
	$n = 2$	5.5 ± 0.4	3.3 ± 1.7

TABLE 1 : Résultats d’estimation du déphasage pour l’autocorrélateur et pour notre réseau de neurones entraîné en validation croisée avec 5 plis. Le RMSE et la dispersion sont en $\text{rad} \times 10^{-2}$ et sont calculés uniquement dans le ventricule.

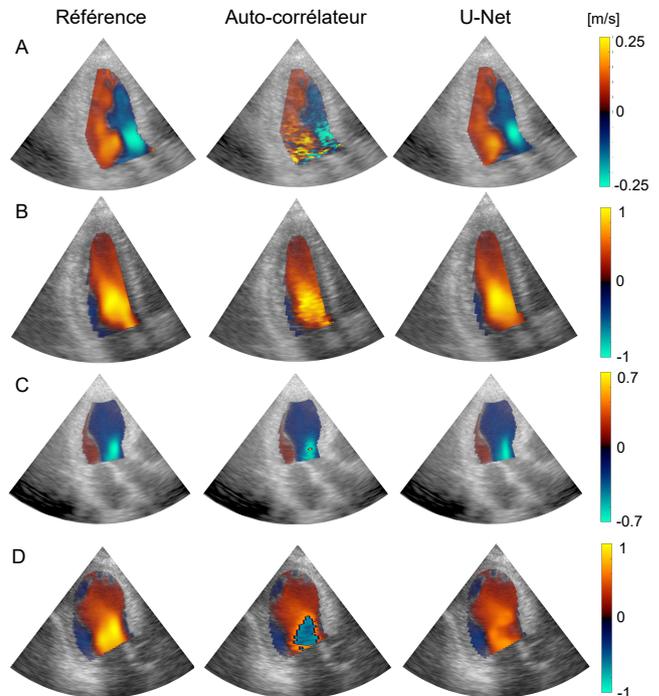


FIGURE 1 : Exemples de références de vitesses Doppler et des vitesses estimées par l’autocorrélateur et le U-Net pour des *packet size* de $n = 2$. Les champs de vitesses sont superposés sur les images B-mode simulées correspondantes. (A) Exemple de cas avec un bruit gaussien de 25 dB. (B) Exemple de cas de surlissage du U-Net. (C) Exemple de cas avec une petite région avec du repliement. (D) Exemple de cas avec une grande région avec du repliement.

performances entre échantillons, aussi bien pour l’autocorrélateur que pour notre méthode. Pour notre méthode, la cause principale de cette dispersion est le surlissage des estimations pour certains échantillons. Nous retrouvons ce phénomène dans deux types de cas : dans les bords inférieurs de nos estimations (*i.e.*, près des valves d’entrée et sortie du sang du ventricule) (voir figure 1B); et en présence de repliement. Dans le cas où le repliement était présent sur une petite région, ce surlissage a eu pour effet de le faire disparaître (voir figure 1C). Néanmoins, dans le cas où le repliement était étendu sur une plus grande région, le surlissage a produit un résultat erroné (voir figure 1D).

4 Gestion du repliement

Les résultats ainsi obtenus indiquent que notre méthode ne permet pas de réduire le repliement les signaux. La mauvaise

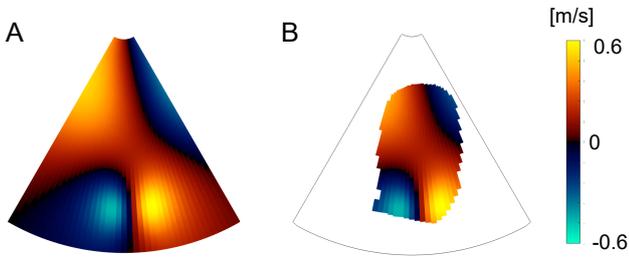


FIGURE 2 : Exemple de référence de la base de données « vortex » avant (A) et après (B) application d'un masque provenant d'un échantillon de la base de données « patient ».

gestion du repliement par le réseau peut être expliquée par le fait que seulement 24 des 2 576 échantillons de la base d'entraînement présentent du repliement. Pour permettre au réseau d'apprendre à gérer le repliement, nous avons généré une deuxième base de données constituées de champs de vitesses avec du repliement.

4.1 Génération de données avec du repliement

Nous avons constitué une base de données avec du repliement en modélisant l'écoulement sanguin par des modèles physiques de mouvement de vortex. Nous avons ensuite simulé les signaux ultrasonores Doppler comme pour la première base de données et avec un *packet size* de $n = 8$. Nous avons ainsi obtenu 3 000 paires constituées de champs de vitesses Doppler de référence et de signaux I/Q Doppler pour différents niveaux de repliement et de bruit gaussien. Nous appelons ces données la base de données « vortex ». Nous avons également considéré un sous-ensemble de 1 000 paires de cette base, où tous les échantillons contiennent du repliement. Nous avons entraîné et testé un réseau sur la base de données « vortex », et avons observé qu'il était capable de réduire le repliement sur ces mêmes données. Nous avons donc cherché à acquérir cette capacité dans le cas des données « patient ».

4.2 Stratégies d'apprentissage

Nous avons deux bases de données d'entraînement de différentes distributions. Les données vortex présentent des vitesses plus grandes en moyenne, et des motifs plus étalés et plus lisses que les données patients.

Entraîner un réseau avec l'union des deux bases de données faisait chuter les performances obtenues sur les données patients car le phénomène de surlissage était accentué. Pour pallier ce problème, nous avons envisagé deux solutions. Premièrement, nous avons appliqué des masques de ventricule gauche provenant des données simulées. Ainsi, nous avons obtenu des motifs moins étendus et pouvions espérer éviter le surlissage. Un exemple des vitesses de référence pour un échantillon de la base « vortex » avant et après l'application d'un masque « patient » est présenté dans la figure 2. Deuxièmement, nous avons envisagé le *transfer learning*. Cette technique d'entraînement est principalement utilisée lorsque la base de données d'intérêt est petite et quand il existe une deuxième base de données plus grande et *suffisamment similaire* à la première sur laquelle nous pouvons entraîner un réseau dont les poids sont ensuite utilisés comme initialisation de l'entraînement sur

Entraînement	<i>Packet size</i>	RMSE	Dispersion
patient + vortex	$n = 8$	3.1 ± 0.3	1.5 ± 0.4
	$n = 2$	5.4 ± 0.5	2.3 ± 0.4
vortex \rightarrow patient	$n = 8$	3.2 ± 0.3	1.8 ± 0.5
	$n = 2$	5.7 ± 0.5	2.6 ± 0.4
patient \rightarrow vortex	$n = 8$	7.4 ± 0.8	4.5 ± 1.3
	$n = 2$	10 ± 1.1	3.8 ± 1.0

TABLE 2 : Résultats d'estimation du déphasage pour notre réseau de neurones entraîné en validation croisée avec 5 plis et pour les trois stratégies d'entraînement. Le RMSE et la dispersion sont en $\text{rad} \times 10^{-2}$ et sont calculés uniquement dans le ventricule.

la base de données d'intérêt. Dans notre cas, nous avons utilisé le *transfer learning* pour travailler en deux temps sur nos deux bases de données de différente distribution.

Pour entraîner un nouveau réseau sur les deux bases de données constituées, nous avons donc considéré trois approches :

1. « patient + vortex ». Nous avons entraîné un réseau sur l'union de la base « patient » et de la base « vortex » de 1 000 échantillons. Pour cette approche, les conditions d'entraînement étaient les mêmes que celles définies dans la section 2.2.
2. « vortex \rightarrow patient ». Nous avons utilisé le *transfer learning* de la base de données « vortex » de 3 000 échantillons à la base de données « patient ». Nous voulions à la fois de bonnes performances sur les données « patient » et conserver les capacités de réduction du repliement obtenues par le premier entraînement. Nous avons donc utilisé un *learning rate* initial petit de 10^{-4} pour rester proches des poids initiaux du réseau.
3. « patient \rightarrow vortex ». Nous utilisons le *transfer learning* de la base de données « patient » à la base de données « vortex » de 1 000 échantillons. Dans un premier temps nous avons donné au réseau les exemples sans repliement, et donc a priori plus simples, puis dans un deuxième temps les exemples qui contiennent du repliement, et donc plus compliqués. Pour ne pas perdre le niveau de détail obtenu sur les données patient, nous avons utilisé un *learning rate* initial de 10^{-5} .

4.3 Résultats

Les résultats des trois approches d'entraînement visant à acquérir la capacité de réduire le repliement sont présentés table 2. Pour les trois approches, nous nous sommes intéressés à la performance sur les données « patient » et sur les régions avec du repliement des données « patient ». Pour les approches « patient + vortex » et « vortex \rightarrow patient », nous avons obtenu des performances équivalentes à celles obtenues dans la section 3 en termes de RMSE. De plus, la dispersion a été réduite de manière significative pour $n = 8$ et $n = 2$. De plus, pour $n = 8/n = 2$, les RMSE des échantillons contenant du repliement sur une grande région de l'image étaient de 0.13/0.14 rad, 0.14/0.17 rad et 0.22/0.23 rad respectivement pour les trois approches, et de 0.85/0.74 rad pour l'approche « patient ». La capacité des nouveaux réseaux à réduire le repliement est visible sur la figure 3D. Néanmoins, pour la deuxième approche « vortex \rightarrow patient », nous avons observé sur les exemples

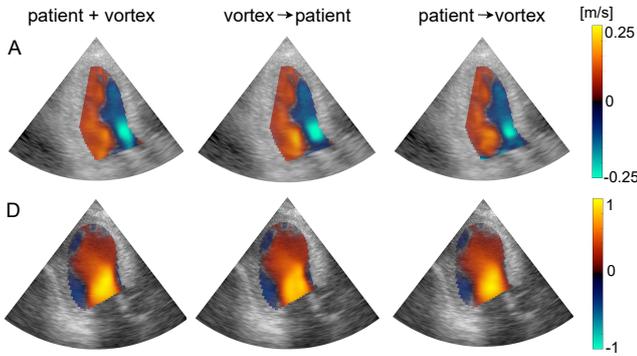


FIGURE 3 : Exemples de vitesses estimées par le U-Net pour des *packet size* de $n = 2$ pour les trois stratégies d’entraînement étudiées. Les champs de vitesses sont superposées sur les images B-mode simulées correspondantes. Les exemples (A) et (D) sont les mêmes que ceux présentés dans la figure 1.

de la figure 3 des difficultés à reconstruire les petits motifs (exemple A). Finalement, pour l’approche « patient \rightarrow vortex » nous avons observé que le phénomène de surlissage était accru lorsque la deuxième tâche d’entraînement était effectué sur les données « vortex » (exemple A). Les performances sur les données « patient » se sont retrouvées très pénalisées.

5 Discussion

Nous avons entraîné un réseau de neurones pour l’estimation de déphasage Doppler pour un *packet size* réduit. Cette approche a obtenu de meilleures performances que la méthode classique (*i.e.*, l’autocorrélateur) sur les données simulées. Le lissage intrinsèque des réseaux de neurones convolutionnels a permis en plus de débruiter naturellement les signaux.

Nous avons ensuite identifié les cas où notre méthode présentait une forte variabilité, et avons examiné le cas où la cause est la présence de repliement. Nous avons donc créé de nouvelles données dans le but d’entraîner un réseau avec la capacité de réduire le repliement. Pour cela, nous avons testé un entraînement classique sur l’union des deux bases de données et le *transfer learning*. L’entraînement classique s’est révélé suffisant pour notre problème grâce à l’introduction des données « vortex ». Pour le *transfer learning*, nous avons perdu en performance sur les données « patient ». Une approche alternative pour le *transfer learning* pourrait être l’utilisation de *learning rates* dit *différentiels* spécifiques par couches.

Finalement, les réseaux convolutionnels à valeurs complexes ont fait leur preuves dans la reconstruction d’images échocardiographiques mode B à partir des signaux I/Q en obtenant de meilleures reconstructions que le réseau analogue à valeurs réelles [3]. Dans notre cas, nous avons testé la version à valeurs complexes de notre réseau mais n’avons pas observé d’améliorations significatives des résultats.

6 Conclusions et perspectives

Nous avons montré que l’apprentissage profond est plus performant que la méthode d’estimation de vitesses classique, même pour un *packet size* réduit. Notre méthode est capable d’estimer la vitesse sanguine et de réduire le repliement en

entraînant un réseau sur des données simulées réalistes et sur des vortex avec beaucoup de repliement, à partir de seulement deux acquisitions successives.

Dans les travaux présentés dans ce document le réseau n’a été testé que sur des simulations. Afin de tester la capacité de généralisation du réseau pour l’estimation de vitesse et la réduction du repliement, nous allons mettre en place une procédure expérimentale à l’aide d’un fantôme de disque tournant. Nous allons ensuite compléter notre approche en travaillant directement sur les signaux I/Q avant filtrage des signaux du tissu. Nous pourrions ensuite tester notre méthode sur des signaux I/Q acquis *in vivo* sur des patients.

Remerciements

Ces travaux ont bénéficié d’un accès aux moyens de calcul de l’IDRIS au travers de l’allocation de ressources 2022-AD011013679 attribuée par GENCI.

Références

- [1] J. FAURIE, M. BAUDET, J. POREE, G. CLOUTIER, F. TOURNOUX et D. GARCIA : Coupling myocardium and vortex dynamics in diverging-wave echocardiography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 66(3):425–432, 2019.
- [2] T. LOUPAS, J.T. POWERS et R.W. GILL : An axial velocity estimator for ultrasound blood flow imaging, based on a full evaluation of the doppler equation by means of a two-dimensional autocorrelation approach. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 42(4):672–688, 1995.
- [3] J. LU, F. MILLIOZ, D. GARCIA, S. SALLES, D. YE et D. FRIBOULET : Complex convolutional neural networks for ultrafast ultrasound imaging reconstruction from in-phase/quadrature signal. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(2):592–603, 2022.
- [4] H. NAHAS, J. S. AU, T. ISHII, B. Y. S. YIU, A. J. Y. CHEE et A. C. H. YU : A deep learning approach to resolve aliasing artifacts in ultrasound color flow imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(12):2615–2628, 2020.
- [5] O. RONNEBERGER, P. FISCHER et T. BROX : U-net : Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015.
- [6] Y. SUN, F. VIXEGE, K. FARAZ, S. MENDEZ, F. NICOD, D. GARCIA et O. BERNARD : A pipeline for the generation of synthetic cardiac color doppler. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(3):932–941, 2022.
- [7] R. J.G. VAN SLOUN, H. BELT, K. JANSE et M. MISCHI : Learning doppler with deep neural networks and its application to intra-cardiac echography. *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2018.