

Vers des CNN réduits pour débruiter les images de phase corrompues par du bruit speckle

MARIE TAHON
SILVIO MONTRESOR
PASCAL PICART

Table des matières

I. Présentation	3
II. Cours	4
1. Bases de données.....	8
1.1. Base de données HOLODEEP.....	8
1.2. Base de données DATAEVAL.....	9
1.3. Base de données NATURAL.....	12
2. Approches de base.....	13
2.1. Approches de traitement du signal pour le débruitage des taches.....	14
2.2. Approche d'apprentissage profond pour le débruitage des taches.....	14
3. Protocoles expérimentaux.....	16
3.1. Prétraitement et mise en œuvre des données.....	17
3.2. Architecture réseau.....	17
3.3. Profondeur et architecture du réseau d'évaluation.....	19
3.4. Surajustement et influence du nombre de couches du CNN.....	19
3.5. Évaluation d'un réseau pré-formé.....	20
3.6. Protocole.....	20
4. Résultats et discussion.....	21
4.1. Profondeur et architecture du réseau.....	21
4.2. Pré-formation.....	22
4.3. Évaluation sur images cibles.....	25
III. Conclusion	26
IV. Remerciements	27
Ressources annexes	28
Bibliographie	30
Webographie	32

I.Présentation

Module :

Vers des CNN réduits pour débruiter les images de phase corrompues par du bruit speckle

Auteur(s) :

Marie Tahon - LIUM- Le Mans Université

Silvio Montresor - LAUM - Graduate School (IA-GS) - Le Mans Université

Pascal Picart - LAUM - Graduate School (IA-GS) - Le Mans Université

Résumé :

L'holographie numérique est une technique très efficace pour l'imagerie 3D et la caractérisation des changements à la surface de tout objet. Cependant, lors du processus d'interférométrie holographique, les images de phase reconstruites souffrent de bruit de speckle. Dans cet article, le débruitage est abordé avec des images de phase corrompues par du bruit speckle. Pour ce faire, des réseaux résiduels DnCNN de différentes profondeurs ont été construits et entraînés avec diverses données de phase holographiques bruyantes. La possibilité d'utiliser un réseau pré-entraîné sur des images naturelles avec bruit gaussien est également étudiée. Tous les modèles sont évalués en termes d'erreur de phase avec les données de référence HOLODEEP et avec 3 images inédites correspondant à différentes conditions expérimentales. Les meilleurs résultats sont obtenus en utilisant un réseau avec seulement 4 blocs convolutifs et formés avec une large gamme de modèles de phases bruités.

Mots-clés :

Holographie numérique, débruitage d'image, deep learning, DnCNN, mise au point

Pré-requis :

-

Objectif(s) pédagogique(s) :

-

Plan du cours :

- Introduction
- Bases de données
- Approches de base
- Protocoles expérimentaux
- Résultats et discussion
- Conclusion

Conception & production :

PRN - Le Mans Université

Licence :

Creative Commons

II.Cours

L'holographie numérique et les méthodes associées basées sur le speckle sont des techniques très efficaces pour la mesure des champs de déplacement et de la forme des surfaces [1] [Digital Holography]. Grâce aux mesures sans contact, la caractérisation des objets peut être obtenue avec une très bonne précision grâce à des motifs de taches. La rétro-propagation numérique permet la reconstruction des images d'amplitude et de phase de l'objet. Bien que ce motif de speckle soit très utile pour le codage, son inconvénient est que l'image d'amplitude reconstruite souffre de bruit de speckle. Le bruit de speckle dans les données de phase holographique est très particulier car il a des statistiques non gaussiennes et présente une propriété non stationnaire, alors que généralement dans les images d'amplitude, ce bruit est considéré comme un bruit multiplicatif.

La technique d'holographie numérique est un processus optique basé sur le laser. Il permet de mesurer les champs de déplacement et la forme de la surface. Grâce aux mesures sans contact, la caractérisation des surfaces rugueuses des objets peut être obtenue avec une très bonne précision.

L'holographie numérique repose sur un mélange cohérent d'une onde de référence et d'une onde objet qui résulte de la diffraction de la lumière provenant de n'importe quel objet. Lorsque la surface de l'objet est rugueuse, des taches sont incluses dans l'hologramme numérique. Dans le cas de la microscopie holographique numérique, les objets sont généralement transparents et il n'y a donc aucune tache dans les images de phase. Dans cet article, le cas d'une surface d'objet rugueuse produisant des taches en phase extraites d'hologrammes est considéré.

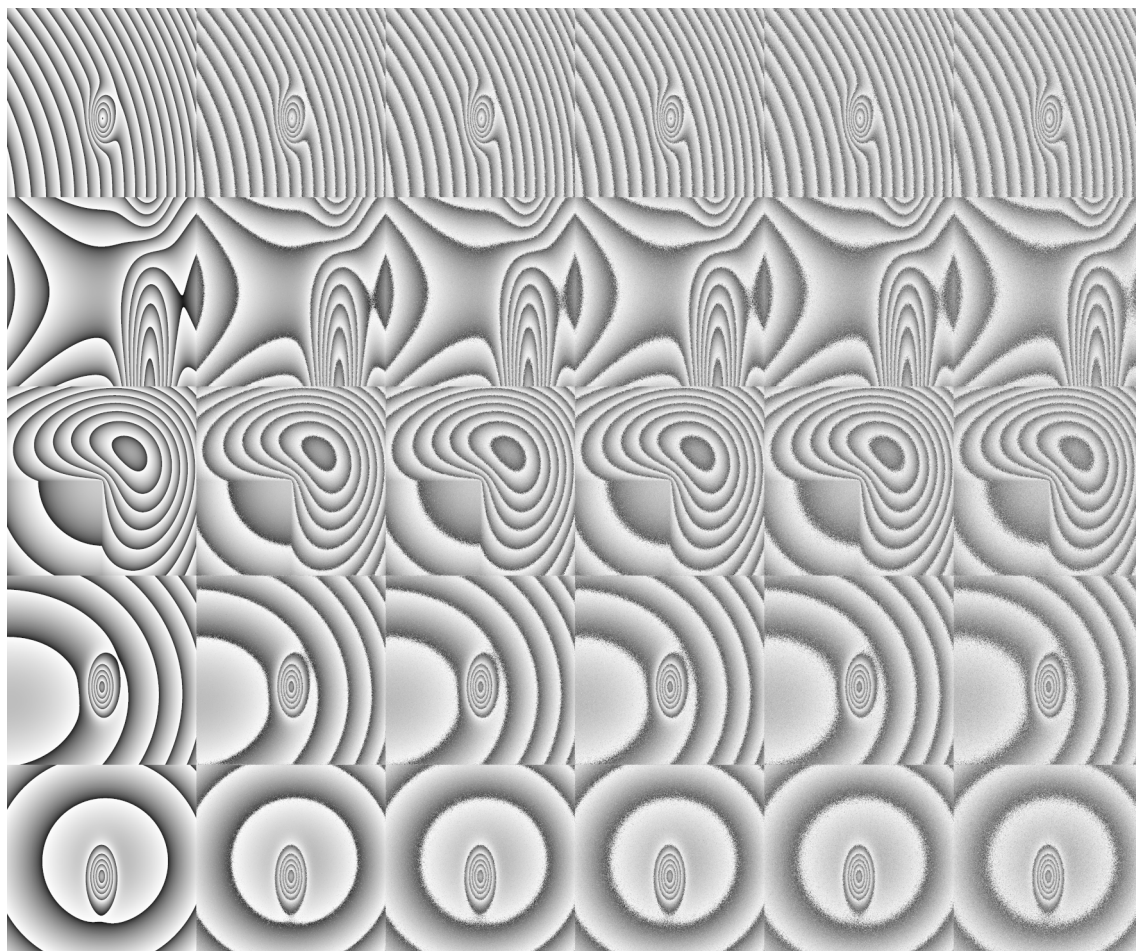


Figure 1 : Un capteur d'image (CMOS ou CCD) enregistre les motifs de franges délivrés par le mélange cohérent.

La figure 1 présente le principe de l'holographie numérique.

Si l'hologramme est enregistré $\sim \text{cite}\{A3\}$ conformément aux conditions de Shannon, nous obtenons une image du front d'onde de l'objet à partir de la reconstruction numérique de l'hologramme. Le champ complexe de l'objet est calculé par des algorithmes de transformation de Fourier rapide simple (S-FFT) ou de transformation de Fourier rapide double (D-FFT) utilisant un grossissement réglable $\sim \text{cite}\{B1\}$. Le calcul numérique du champ diffracté donne une amplitude complexe $O_0(x, y)$ échantillonné sur une grille correspondant au nombre de points de reconstruction de l'algorithmme. L'image d'amplitude (module), A_r , et image de phase, ψ_r (argument du champ) sont issus du calcul numérique du champ diffracté par l'objet lors de l'enregistrement de l'hologramme.

à savoir :

$$A_r(x, y) = |O_0(x, y)|, \quad (1)$$

$$\psi_r(x, y) = \text{atan2}\left\{\frac{\Im m[O_0(x, y)]}{\Re e[O_0(x, y)]}\right\} \bmod(2\pi). \quad (1)$$

La phase du champ est calculée par une fonction arctangente, et par conséquent le résultat est enveloppé dans l'intervalle $[-\pi, +\pi]$, module 2π .

Par conséquent au calcul au moyen de la fonction arctangente, ψ_r est enveloppé dans l'intervalle $[-\pi, +\pi]$, module 2π .

Dans la plupart des cas [2] [Deep residual learning for image recognition], le caractère aléatoire de la phase, qui est reconstruite à partir du champ diffracté de l'objet, est dû à la rugosité de l'objet. L'objet reconstruit est alors soumis au phénomène de speckle noise.

L'estimation de phase du champ optique reconstruit est un enjeu clé pour un grand nombre d'applications de l'holographie numérique.

Les applications de métrologie nécessitent une phase optique, c'est pourquoi cet article se concentre sur les changements de phase au fil du temps. La grandeur d'intérêt est une différence de phase entre deux instants, permettant de suivre l'évolution d'un phénomène dans le temps. Compte tenu de l'effet Doppler, la différence de phase est proportionnelle au champ de déplacement de l'objet entre les deux instants. Comme la phase optique est calculée à partir de la fonction arctangente, elle est ensuite enveloppée. La phase doit être déroulée afin d'accéder aux grandeurs cinématiques physiques de l'objet [3] [Two-dimensional phase unwrapping: theory, algorithms, and software]. Par exemple, l'holographie numérique permet d'étudier des phénomènes acoustiques complexes en utilisant la méthode d'holographie numérique ultra-rapide avec un taux d'échantillonnage allant jusqu'à 100 kHz [4] [High-speed holographic metrology: principle, limitations, and application to vibroacoustics of structures].

Concernant le débruitage des images, les algorithmes sont généralement conçus avec l'hypothèse d'un bruit gaussien additif et il existe un réel besoin de nouvelles approches de débruitage capables de gérer le bruit de speckle et les motifs de franges complexes. Depuis une décennie, les algorithmes de référence sont liés à des méthodes basées sur des patches non locaux comme BM3D [3] [Two-dimensional phase unwrapping: theory, algorithms, and software], méthodes basées sur les ondelettes telles que DTDWT [5] [Experimental and theoretical investigation of the pixel saturation effect in digital holography], et des algorithmes de transformée de Fourier à court terme comme le WFT2F [6] [Windowed Fourier transform for fringe pattern analysis: theoretical analyse].

Les algorithmes d'apprentissage automatique manifestent un intérêt croissant pour le traitement du signal et des images depuis les 10 dernières années. Les réseaux neuronaux sont notamment capables d'apprendre des fonctions très complexes à partir de bases de données.

Au cœur de ces algorithmes, les réseaux convolutifs (CNN) ont été largement utilisés à la place des approches traditionnelles de traitement d'images..

Contrairement à ces approches traditionnelles, les solutions basées sur l'apprentissage automatique telles que les réseaux de neurones convolutifs (CNN) utilisent des exemples d'ensembles de données et sont capables d'apprendre à inverser des fonctions de dégradation très complexes [7] [Natural image denoising with convolutional networks].

Ils ont été utilisés pour simuler des ondelettes et des analyses multirésolutions, des algorithmes de réduction et de seuillage, des représentations clairsemées, la correspondance de blocs et l'apprentissage de dictionnaires [8] [Computational image speckle suppression using block matching and machine learning].

De nombreuses architectures neuronales ont été développées pour le bruit gaussien, comme l'apprentissage résiduel pour la reconnaissance d'images [2] [Deep residual learning for image recognition], réseaux contradictoires génératifs (GANs) [9] [Generative adversarial nets].

A noter que dans le domaine de l'holographie numérique et de la microscopie holographique numérique, plusieurs articles liés aux applications de CNN ont été publiés [10] [On the use of deep learning for computational imaging].

Actuellement, les systèmes de débruitage d'image de pointe sont dominés par DnCNN [8] [Computational image speckle suppression using block matching and machine learning] et ses modifications récentes telles que l'apprentissage résiduel hiérarchique HRLNet [11] [Beyond a Gaussian denoiser: Residual learning of Deep CNN for image denoising].

Les réseaux résiduels apprennent à prédire l'image résiduelle entre les entrées propres et bruyantes. Il comprend des connexions sautées qui consistent en un mappage d'identité placé entre deux couches non adjacentes et permet d'éviter le problème de gradient de disparition lorsque la profondeur du réseau est élevée [2] [Deep residual learning for image recognition]. Grâce à l'apprentissage résiduel, des réseaux très profonds peuvent être facilement formés et une précision améliorée a été obtenue pour la classification des images et la détection d'objets.

Plusieurs approches ont été proposées en tomographie par cohérence optique [12] [On the use of self-supervised pre-trained acoustic and linguistic features for continuous speech emotion

recognition], en imagerie hyperspectrale ou en utilisant des décompositions multi-échelles [10] [On the use of deep learning for computational imaging]. Le problème de la décorrélation des speckles a également été abordé avec des réseaux d'apprentissage profond avec des GANs [13] [Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks].

Alors que la quantité et la diversité des images naturelles sont énormes et permettent ainsi de former des réseaux profonds avec de nombreux paramètres, lors du passage au traitement des données en phase en holographie numérique, la quantité et la diversité sont clairement réduites.

En effet, il n'existe actuellement aucun moyen d'obtenir des données de phase expérimentale avec du bruit speckle avec sa version propre. C'est la raison pour laquelle des données simulées sont nécessaires.

Une telle simulation peut favoriser la diversité des motifs et des fréquences de franges.

Des images nettes de vérité terrain anti-taches ont été générées à partir des sorties de scanners de tomographie optique cohérente commerciaux [13] [Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks].

Dans [14] [Quantitative appraisal for noise reduction in digital holographic phase imaging], une base de données comprenant 25 motifs de franges répartis en 5 motifs et 5 rapports signal/bruit différents a été générée avec un simulateur de bruit réaliste [15] [Error analysis for noise reduction in 3D deformation measurement with digital color holography] pour favoriser la diversité des modèles de franges de phase.

Pour améliorer les performances de débruitage, une solution consiste à approfondir, *i.e.* pour ajouter plus de couches au réseau. Cependant, avec une capacité plus élevée, deux problèmes apparaissent : le surajustement et les gradients qui disparaissent ou explosent.

Ces derniers peuvent être contrôlés par la normalisation par lots et l'utilisation de connexions sautées comme dans les réseaux résiduels.

Le surajustement peut être contrôlé par régularisation.

Cependant, la quantité de données est cruciale pour éviter le surajustement, même avec des techniques de régularisation. L'utilisation de l'augmentation des données permet généralement d'augmenter artificiellement la quantité de données d'entraînement [16] [The effectiveness of data augmentation in image classification using deep learning].

Cependant, une solution peut aussi être de ne pas aller plus loin.

https://www2.cs.duke.edu/courses/fall20/compsci371d/slides/s_10_CNNGeneralization.pdf¹

S'il est connu qu'il existe une relation entre la profondeur du réseau et la taille des filtres convolutifs (et par conséquent le champ récepteur) [17] [Very deep convolutional networks for large-scale image recognition], la question de la nécessité de la profondeur n'a pas été beaucoup étudiée. Dans [18] [Depth selection for deep ReLU nets in feature extraction and generalization], les auteurs proposent de quantifier la correspondance entre les fonctionnalités apprises par le réseau et sa profondeur.

DnCNN [2] [Deep residual learning for image recognition] a été conçu selon cette approche.

Suivant cette approche, pour les images naturelles avec débruitage gaussien avec un certain niveau de bruit, le champ récepteur DnCNN a été réglé sur 35×35 avec 17 couches convolutives correspondantes [2] [Deep residual learning for image recognition].

Une étude récente a montré que les différentes couches entraînées d'un réseau sont relativement peu corrélées, ce qui pourrait expliquer pourquoi les réseaux neuronaux résistent au surajustement. Cette étude conclut également que les grands réseaux formés sur des ensembles de données avec un petit nombre d'échantillons de formation se généralisent également bien et évitent le surajustement.

Le pouvoir de généralisation des algorithmes d'apprentissage automatique réside dans la « capacité à bien fonctionner sur des entrées jusqu'alors inobservées » [19] [Deep Learning].

1 - To improve de-noising performances, one solution is to go deeper, \textit{i.e.} to add more layers to the network. However, with higher capacity two problems emerge: overfitting and vanishing or exploding gradients.

Pour ce faire, les données sont généralement divisées en ensembles de formation, de développement et de test, le rappel étant constitué des entrées non observées.

Dans des travaux antérieurs, les auteurs ont formé un DnCNN pour les données de phase holographique avec débruitage des taches [20] [Computational de-noising based on deep learning for phase data in digital holographic interferometry]. Ce réseau atteint de bonnes performances avec les données de référence par rapport à d'autres techniques de débruitage telles que BM3D ou WFT2F sur la plupart des images de phase évaluées.

Dans le présent article, les réseaux sont évalués en termes d'erreurs de phase et de pouvoir de généralisation défini comme la « capacité à bien fonctionner sur des entrées jusqu'alors inobservées » [19] [Deep Learning]. L'objectif est de réduire le temps d'entraînement tout en atteignant des performances similaires. Pour ce faire, les bases de données à développer et à valider sont présentées dans la section Bases de données. Les algorithmes de débruitage de base et les résultats sont résumés dans la section Approches de base. Les protocoles de formation incluent des réseaux de différentes profondeurs, sur des données d'images de différentes phases (voir section Protocoles Expérimentaux). Avec l'avantage d'un réglage fin avec des données de phase corrompues par du bruit speckle, un réseau préalablement formé sur des images bruitées naturelles est également étudié. Les résultats expérimentaux sont discutés dans la section Résultats et discussion.

1. Bases de données

1.1. Base de données HOLODEEP

Cette base de données se compose de 5 types différents de modèles de franges de phase sans bruit et est utilisée pour entraîner les modèles et à des fins de développement.

Chaque motif est dégradé avec un bruit de décorrélation de speckle réaliste avec des statistiques décrites dans [14] [Quantitative appraisal for noise reduction in digital holographic phase imaging]. A partir de chaque motif de franges sans bruit, cinq motifs de franges bruyantes contrôlés par un paramètre, à savoir Δ , sont générés avec le simulateur présenté [14] [Quantitative appraisal for noise reduction in digital holographic phase imaging], correspondant à différents rapports signal/bruit (SNR) dans la plage [3dB-12dB].

Le paramètre Δ est utilisé pour imiter les données de la phase expérimentale fortement dégradées. *Plus le Δ est supérieur, plus le SNR est petit.* En conditions réelles, il existe plusieurs sources de dégradation qui pourraient induire plus de bruit de décorrélation que prévu si tout était parfait. À titre d'exemple, la reconstruction de données holographiques peut ne pas être parfaitement mise au point [21] [Refocus criterion based on maximization of the coherence factor in digital three-wavelength holographic interferometry], les pixels pourraient avoir une grande surface active [4] [High-speed holographic metrology: principle, limitations, and application to vibroacoustics of structures], l'enregistrement peut avoir un faible nombre de pixels ou des pixels saturés [5] [Experimental and theoretical investigation of the pixel saturation effect in digital holography], le nombre de bits de quantification utiles pourrait être insuffisant [22] [Quality assessment of combined quantization-shot-noise-induced decorrelation noise in high-speed digital holographic metrology], ou il pourrait également y avoir des changements de longueur d'onde entre les expositions [23] [Improvement of accuracy in digital holography by use of multiple holograms]. Toutes ces sources de dégradation ont pour conséquence l'augmentation de la décorrélation du speckle puis l'augmentation du bruit. Ainsi, en utilisant Δ est un moyen utile d'obtenir des données avec plus de bruit afin d'imiter les conditions expérimentales possibles. Dans le simulateur décrit [14] [Quantitative appraisal for noise reduction in digital holographic phase imaging], Δ correspond à de petits changements de longueur d'onde entre les deux expositions. Donc ajuster Δ est utile pour augmenter la décorrélation des taches et ainsi diminuer le SNR dans les données de phase. Les images simulées, dimensionnées 1024×1024 pixels, ont été générés à

l'aide de Matlab et sont disponibles dans Matlab `mat` format ou comme `tiff` images. Les 25 images utilisées pour entraîner les modèles sont présentées dans la figure 2.

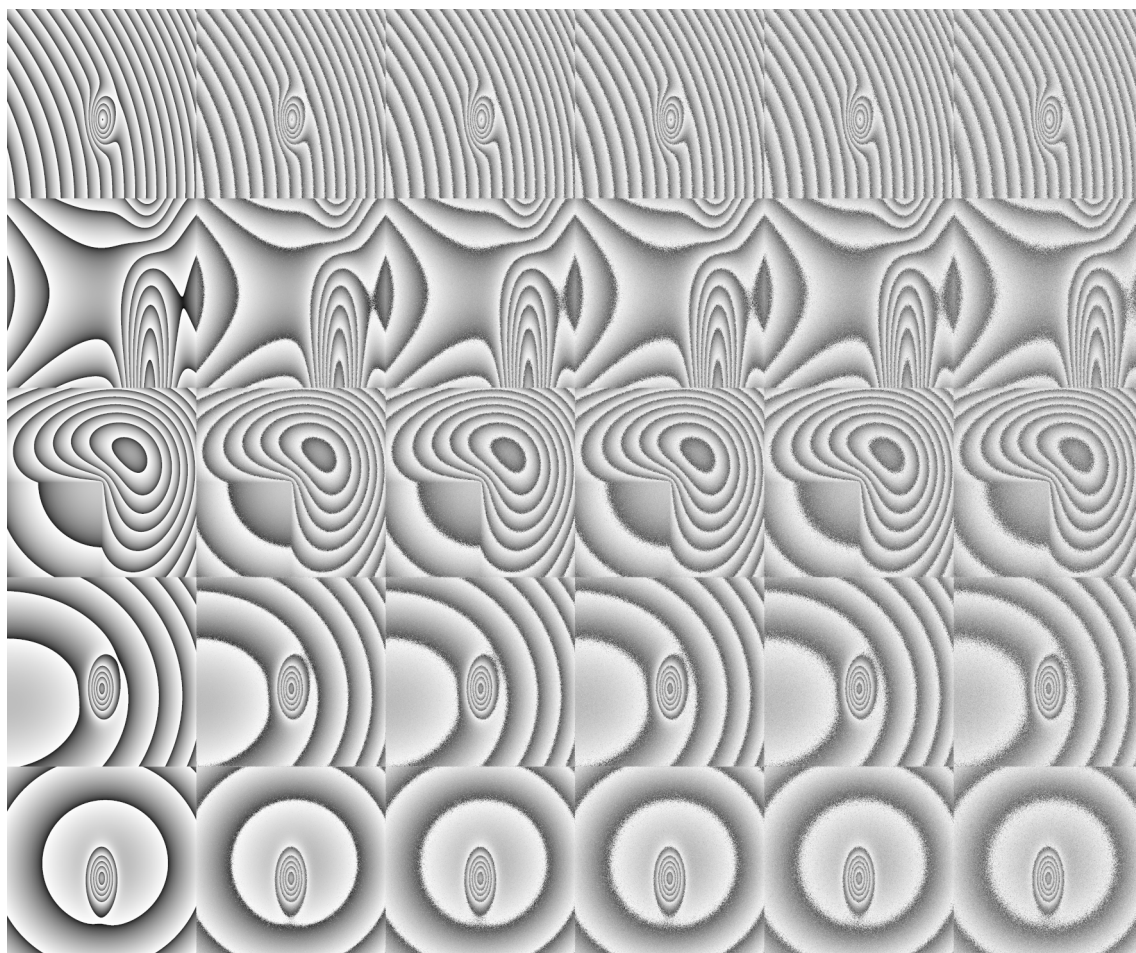


Figure 2 : Images de la phase d'entraînement HOLODEEP : 5 modèles (en lignes) avec bruit de speckle simulé avec cinq valeurs de Δ (en colonnes)

1.2. Base de données DATAEVAL

Cette base de données de validation est constituée de 3 images utilisées pour tester le modèle avec des images qui n'ont pas été vues lors du processus de formation ou de développement. Images en deux phases, à savoir `Test1` et `Test2`, sont simulés à l'aide du simulateur de référence [16] [The effectiveness of data augmentation in image classification using deep learning], de la même manière que pour la simulation de la base de données HOLODEEP. Les SNR des deux phases sont respectivement 3.05 dB (voir Figure 3) et 1.26 dB (voir Figure 4). Ces cartes de phases ne sont pas incluses dans la base de données HOLODEEP. La dernière phase est une phase expérimentale bruyante provenant des vibrations mesurées à $17\,512 \text{ Hz}$, nommée `Test3` avec une $SNR = 2.52 \text{ dB}$. La phase propre est indiquée dans la Figure 5, la phase bruyante dans la Figure 6 et la phase bruyante est obtenue dans la Figure 7. Le dispositif expérimental et la méthodologie pour obtenir de telles images de phase sont décrits dans les références ([4] [High-speed holographic metrology: principle, limitations, and application to vibroacoustics of structures], [2 [Visualization of travelling waves propagating in a plate equipped with {2D ABH} using wide-field holographic vibrometry]4] [Visualization of travelling waves propagating in a plate equipped with {2D ABH} using wide-field holographic vibrometry]). Le lecteur est invité à consulter les articles pour plus de détails .

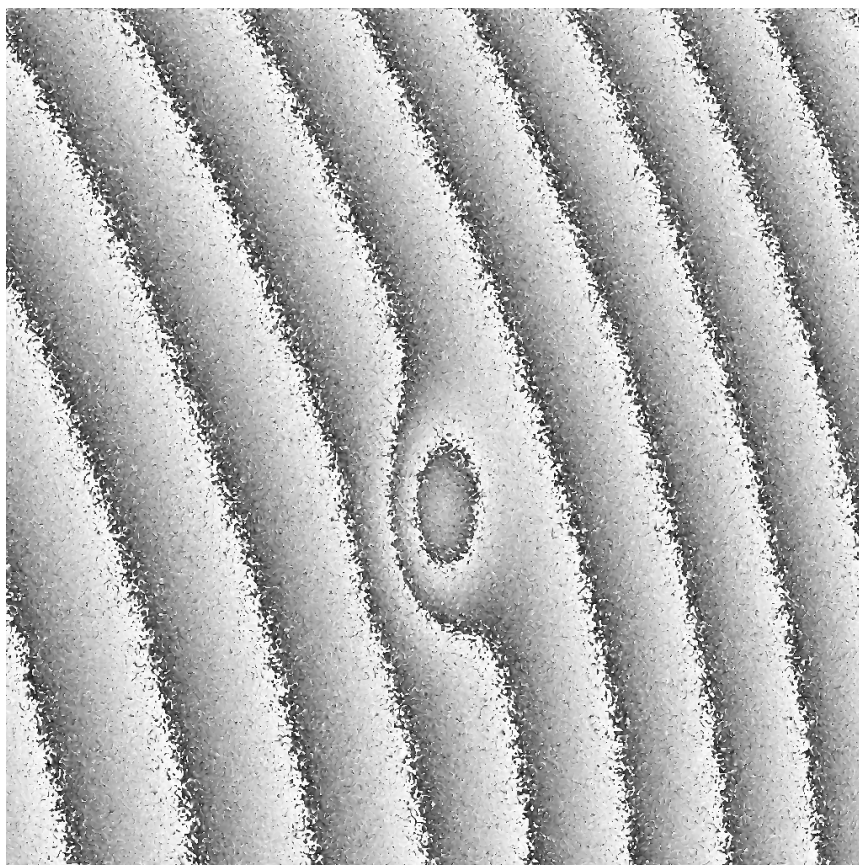


Figure 3 : Phase Test 1 bruyant

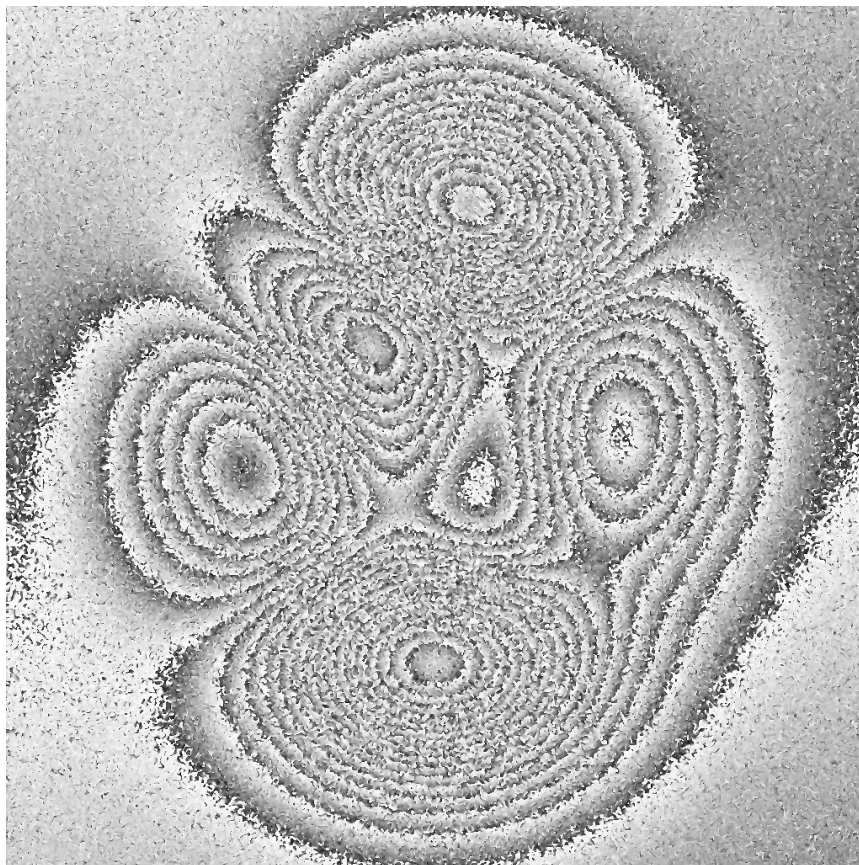


Figure 4 : Phase Test 2 bruyant

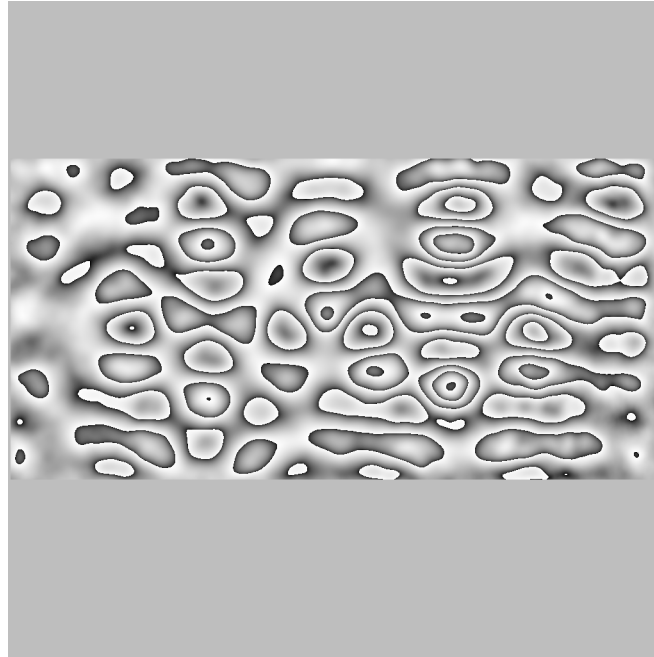


Figure 5 : Phase Test 3 sans-bruit

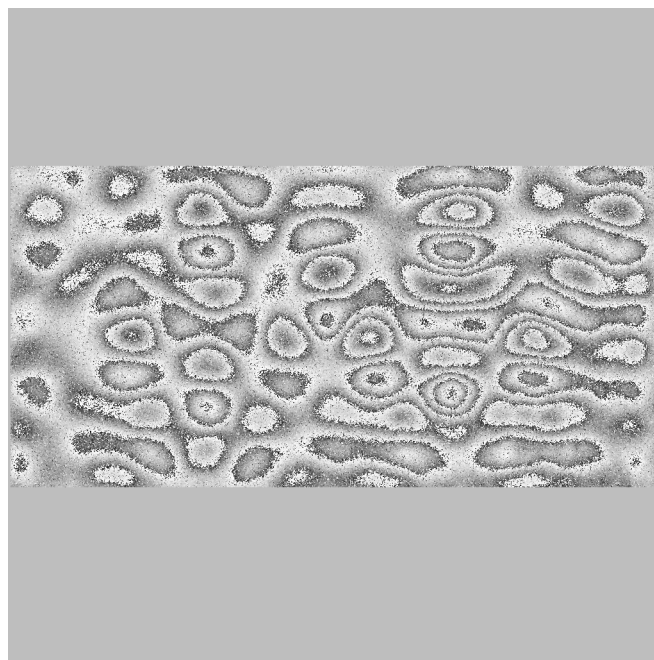


Figure 6 : Phase Test 3 bruyant

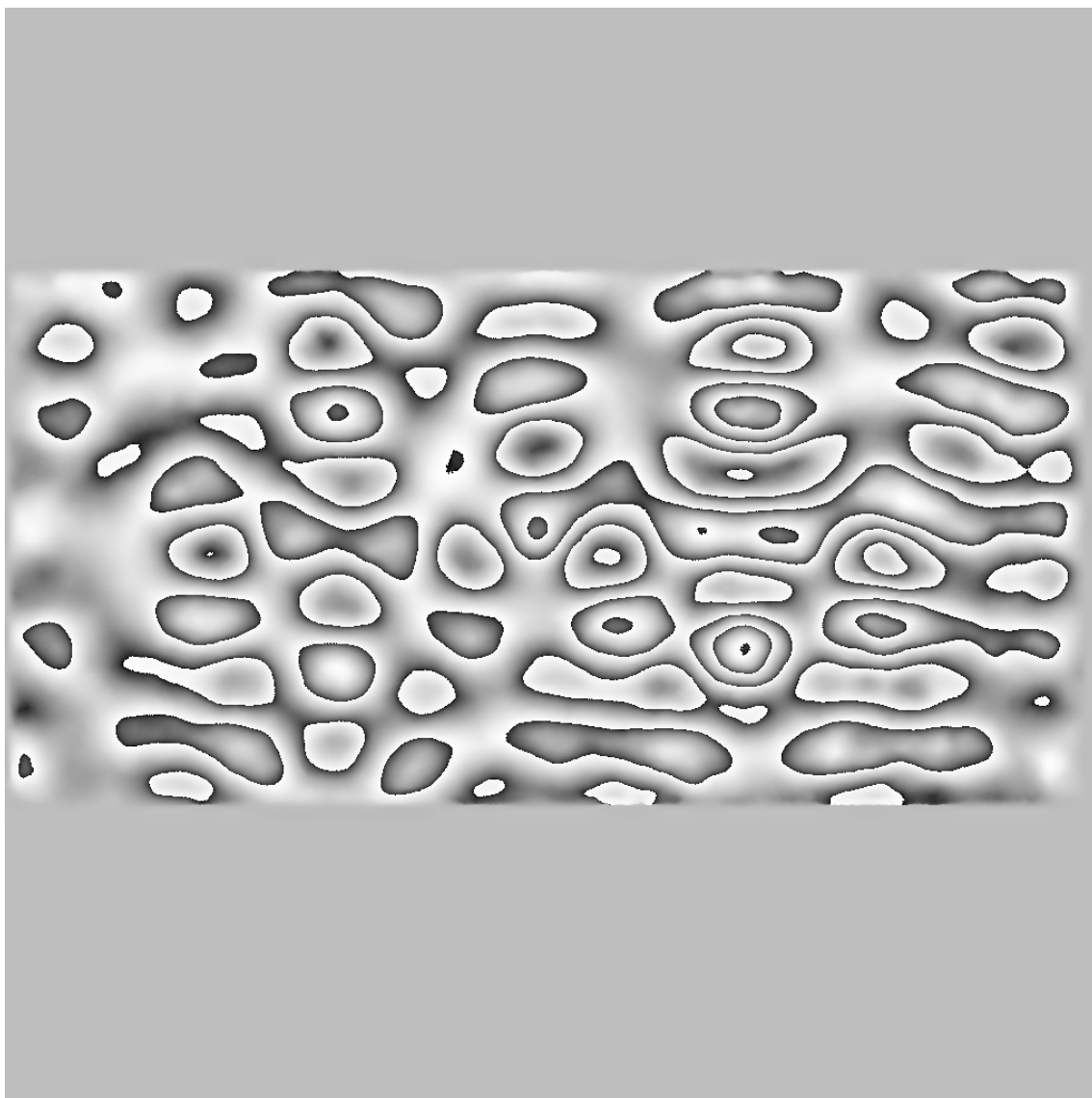


Figure 7 : Phase 3 Test Dé-bruit

1.3. Base de données NATURAL

Cette base de données est généralement utilisée pour le débruitage gaussien d'images en niveaux de gris naturels. Il consiste en 400 images de taille 180×180 . Les images RVB sont disponibles sur le lien ². Les images bruyantes sont obtenues en ajoutant du bruit gaussien avec différentes valeurs SNR (plus de 13 dB) directement aux images propres.

2 - http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/BSR/BSR_bsds500.tgz

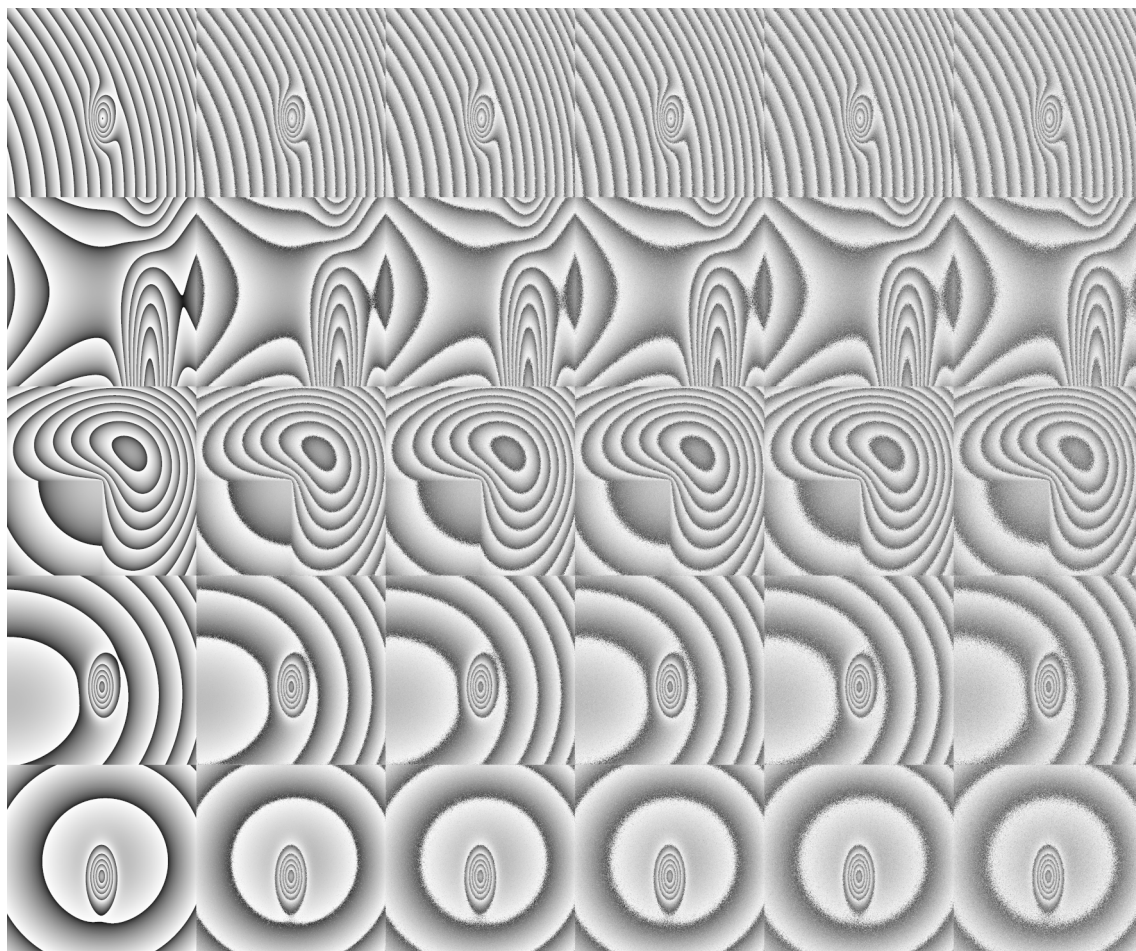


Figure 2 : Images de la phase d'entraînement HOLODEEP : 5 modèles (en lignes) avec bruit de speckle simulé avec cinq valeurs de Δ (en colonnes)

2. Approches de base

Dans cette section, les résultats de base de l'état de l'art sont présentés dans la Table 4. L'erreur de phase en radians a été obtenue à partir de la base de données de référence HOLODEEP et des images DATAEVAL.

En suivant le protocole décrit dans [15] [Error analysis for noise reduction in 3D deformation measurement with digital color holography], trois algorithmes issus du traitement du signal ont été testés : WFT2F, BM3D et DtDWT.

Un premier CNN nommé DL-3 [25] [An iterative scheme based on deep learning combined with input noise estimator for phase data processing in digital holographic interferometry], ont été formés sur des images HOLODEEP avec un petit bruit de speckle *i.e.* $\Delta\lambda = 0$.

Les résultats de validation sont donnés en termes de moyenne $\Delta\Phi$ sur l'ensemble de la base de validation HOLODEEP (soit 25 images de taille 1034×1024), et les résultats des tests sont donnés sur les trois images de DATEVAL et en moyenne. Tous ces résultats sont résumés dans la Table 1 en termes d'erreur de phase.

2.1. Approches de traitement du signal pour le débruitage des taches

En suivant le protocole décrit [14] [Quantitative appraisal for noise reduction in digital holographic phase imaging], trois algorithmes issus du traitement du signal ont été testés : WFT2F, BM3D et et DtDWT.

Les résultats sont donnés en termes d'écart type $\Delta\phi$ de la phase d'erreur e_{ij} définis dans l'équation 1, où N est le nombre total de pixels et $e_{ij} = \phi_{\text{débruité}}(i, j) - \phi_{\text{sans bruit}}(i, j)$ is the difference between the de-noised phase $\phi_{\text{débruité}}$ and the noise-free phase $\phi_{\text{sans bruit}}$ au pixel (i, j) ,

$$\Delta\phi^2 = \frac{1}{N} \sum_{i,j} (e_{ij} - m_e)^2, \quad (1)$$

1

où m_e est la moyenne de $e(i, j)$ sur l'ensemble des pixels. Notez que depuis $\phi_{\text{débruité}}$ et $\phi_{\text{sans bruit}}$ sont calculés modulo 2π , il y a une différence e_{ij} doit également être calculé modulo 2π selon $e_{ij} = \arg[\exp(i e_{ij})]$.

Les résultats de base sont donnés en termes de moyenne de $\Delta\phi$ sur l'ensemble de la base de données HOLODEEP (soit 25 images de taille 1024×1024), et avec les 3 images de la base de données DATAEVAL. Les résultats pour l'erreur de phase $\Delta\phi$ sont résumés dans la Table 1.

Le numéro d'itération correspond au nombre de fois que l'image bruitée a été traitée par le débruiteur. Depuis la Table 1 on peut observer qu'une seule itération est nécessaire avec WFT2F pour obtenir la meilleure erreur à $\Delta\phi = 0.026 \text{ rad}$ avec HOLODEEP, car WFT2F utilise un seuil sur les formes d'onde de décomposition 2D et le processus se termine après une itération. Même avec 3 itérations, les deux autres méthodes n'atteignent que $\Delta\phi = 0.046 \text{ rad}$ (DtDWT) et $\Delta\phi = 0.068 \text{ rad}$ (BM3D), confirmant ainsi les meilleures performances pour WFT2F.

Method	# iter	HOLODEEP 25 Images	DATAEVAL		
			Test1	Test2	Test3
WFT2F	1	0.026	0.044	0.164	0.105
DtDWT	1-3	0.046	0.078	0.519	0.214
BM3D	1-3	0.068	0.113	0.580	0.094
ine DL-3	1	0.041	0.107	0.585	0.105
DL-3	3	0.031	0.078	0.559	0.077

Tableau 1. Écart type de base des erreurs de phase (Df en rad) obtenues sur les 25 images de la base de données HOLODEEP (en moyenne) et images individuelles de DATAEVAL. Iter est le nombre de fois que l'image passe à travers le débruiteur

2.2. Approche d'apprentissage profond pour le débruitage des taches

a) Augmentation des données

Étant donné que la base de données de formation n'est peut-être pas suffisamment étendue, le traitement du signal est utilisé pour l'augmenter. Pour chaque image de phase originale, ses versions cosinus et sinus ($\times 2$) sont considérés avec leur version transposée et déphasée ($\pi/4$ déphasage). Cette opération permet d'augmenter par 8 le nombre d'images originales.

b) Mise en œuvre de base

Le réseau de départ considéré dans cette section est [11] [Beyond a Gaussian denoiser: Residual learning of Deep CNN for image denoising], nommé DnCNN.

Il comprend 59 couches organisées sur 6 blocs résiduels (ResBlocks) avec fonction d'activation ReLu, une couche d'entrée, une couche de sortie. Le principe du réseau résiduel implique que le réseau se comporte comme un estimateur de bruit.

Une première couche d'entrée (3×3 couche convolutive et unités linéaires rectifiées ReLU), **16** blocs convolutifs intermédiaires (ConvBlocks : $3 \times 3 \times 64$ couche convolutive, normalisation par lots et ReLU), et une couche de sortie ($3 \times 3 \times 64$ couche convolutive) qui est utilisé pour reconstruire le bruit de sortie. L'image débruitée est la soustraction de l'image bruitée et du bruit de sortie. Différentes tailles pour le nombre de couches D ont été formés.

Le réseau a été préalablement entraîné sur des images naturelles corrompues par du bruit gaussien, puis entraîné sur des images de phase holographique.

La fonction de perte est une perte L2 entre la référence et les valeurs de pixel prédites.

Le réseau pré-entraîné a été téléchargé directement depuis la boîte à outils Matlab Deep Learning : <https://www.mathworks.com/help/images/ref/dncnnlayers.html>³.

Les paramètres du processus de formation sont résumés dans la Table 2 (cf. Table 2. Paramètres utilisés pour former les réseaux. Δ se trouve pour le bruit de speckle simulé p 28).

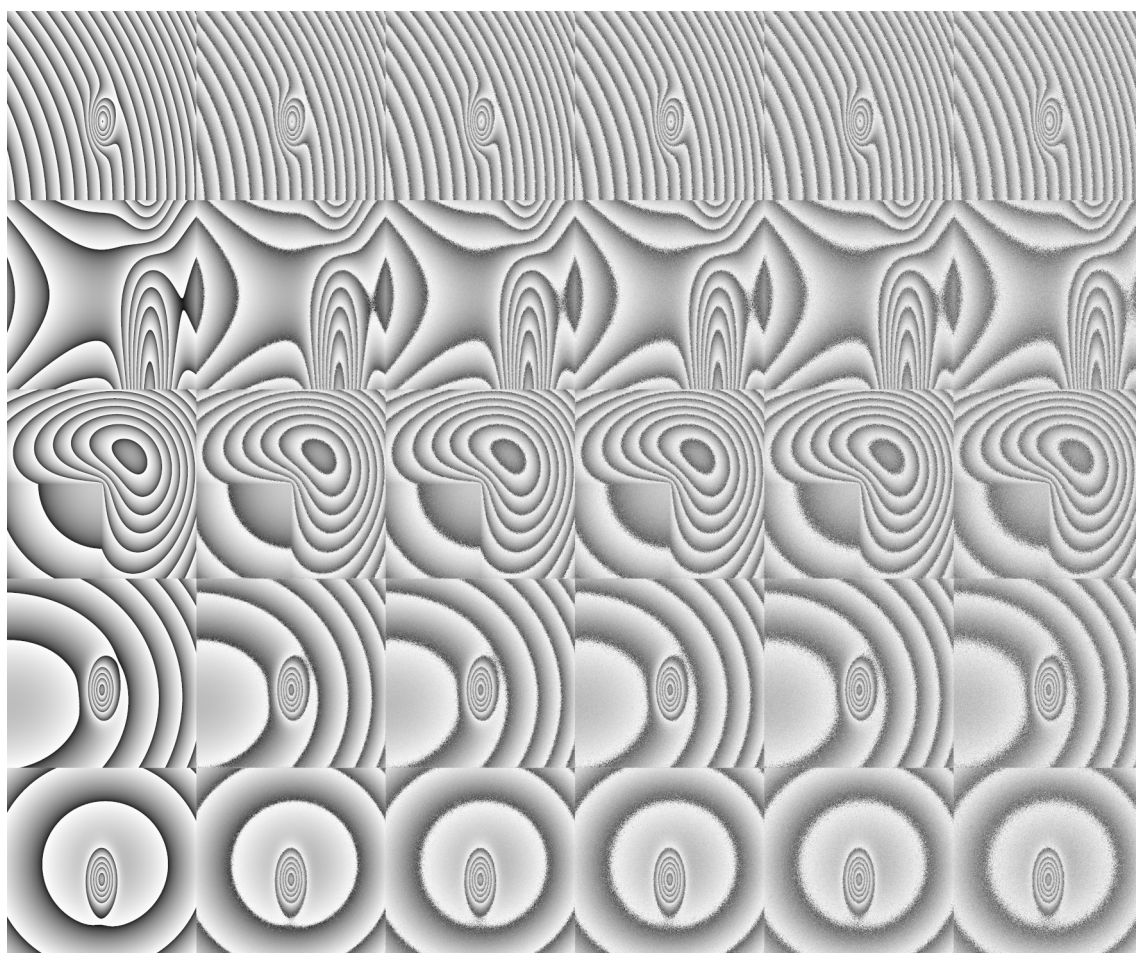


Figure 2 : Images de la phase d'entraînement HOLODEEP : 5 modèles (en lignes) avec bruit de speckle simulé avec cinq valeurs de Δ (en colonnes)

DnCNN network[11] [Beyond a Gaussian denoiser: Residual learning of Deep CNN for image denoising] a été pré-entraîné avec 400 images grises naturelles dimensionnées 80×80 issu de la base de données NATURAL et optimisé avec l'algorithme Adam. Le débruiteur gaussien aveugle a été entraîné avec un large ensemble de niveaux de bruit, ($\sigma \in [0; 55]$ est l'écart type du bruit supplémentaire, la moyenne est 0) et une taille de patch de 50×50 . Sur la fin 128×3000 patches ont été rognés pour entraîner le modèle.

3 - <https://www.mathworks.com/help/images/ref/dncnnlayers.html>

128 étant le nombre de patches par lot utilisé dans la mise en œuvre (taille du lot).

DL-3[20] [Computational de-noising based on deep learning for phase data in digital holographic interferometry] utilise un réseau préformé (<https://www.mathworks.com/help/images/ref/dncnnlayers.html>), qui est ensuite affinée avec des données provenant des 5 motifs de franges, et un niveau de bruit fixé à 2 pixels par grain de speckle dans le simulateur ($\Delta = 0$). Le modèle a été optimisé à l'aide de l'algorithme de descente stochastique en gradient (SGD). Cette situation correspond à des conditions réalistes d'enregistrement holographique numérique en ligne. Chaque image de phase est ensuite augmentée 8 fois, soit un total de 40 images dimensionnées 1024×1024 sont utilisés pour adapter le modèle.

c) Résultats de base

Les résultats de la base de données de validation sont présentés dans la table 3 . Ils sont donnés en termes de moyenne $\Delta\Phi$ sur l'ensemble de la base de données de validation, soit 25 images de taille 1034×1024 .

Method	# iter	HOLODEEP #25
25 images		
DL-3	0	
DL-3	1	.0376
DL-3	2	.0308
DL-3	3	.0292
DL-3	4	.0292

Table 3. Résultats en $\Delta\phi$ (rad), obtenus avec 4 itérations avec le modèle DL-3 adapté au bruit de speckle de motif de franges. Résultats de (Montresor, 2020)

Les résultats obtenus avec DL-3 sont rapportés dans la Table 1. The aforementioned deep learning model is compared to the signal processing approaches.

The results show that DL-3 model slightly under-performs WFT2F on HOLODEEP with 3 iterations, however the computation time is more interesting in the case of deep learning [25] [An iterative scheme based on deep learning combined with input noise estimator for phase data processing in digital holographic interferometry]. The addition of a noise estimator can further improve the performances.

To be comparable with the baseline of de-noising algorithms, only one iteration is taken into account in the following experiments.

From Table 1, with DL-3 and 3 iterations the results are in the range of those from DtDWT and better than BM3D for phase maps **Test1** and **Test2** (speckle size at 4 pixels per grain). DL-3 was trained with only speckle grain at size 2, so this shows that the neural network can generalize with phase maps which do not exactly correspond to the same trained speckle size.

Method	# iter	HOLODEEP		DATAEVAL			
		Time	25 images	DATA1	DATA20	VibMap	AVG
WFT2F*	1	241	.025	.0443	0.1643	.1051	.1046
DL-3	1						
DL-3	3	210	0.41	.0778	.5590	.0737	.2368

Table 4. Comparaison des performances de débruitage (en $\Delta\phi$ (rad)) entre les modèles DL et WFT2F

3. Protocoles expérimentaux

4 - <https://www.mathworks.com/help/images/ref/dncnnlayers.html>

Le cadre global est présenté dans la Figure 8 , où la base de données HOLODEEP est utilisée pour former les réseaux. La métrique d'évaluation est l'erreur de phase $\Delta\phi$ calculé entre l'image sans bruit prédite et la référence sans bruit (se référer à l'équation 1).

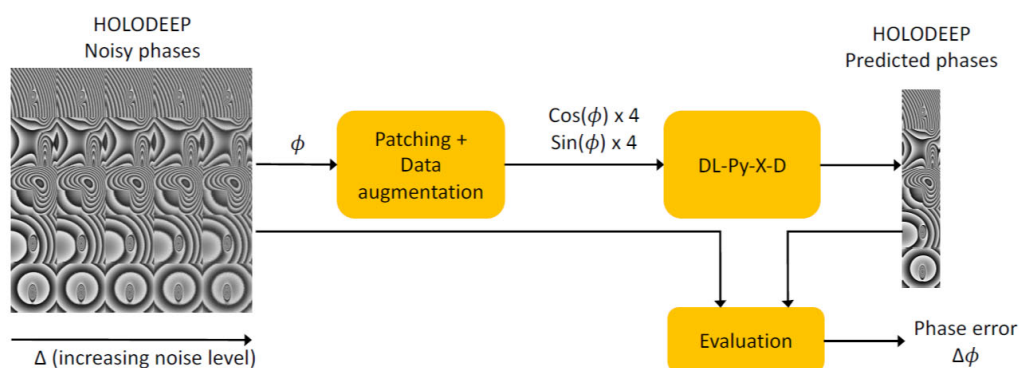


Figure 8 : Aperçu global de la phase de formation du système

3.1. Prétraitement et mise en œuvre des données

Les expériences suivantes considèrent deux paramètres indépendants : le type de motif de phase (5 modèles dans la base de données HOLODEEP) et le niveau de bruit speckle.

qui est donné dans $\Delta\lambda$ (de 0 à 2.5). Pour chaque taille d'image originale 1024×1024 , les correctifs candidats sont extraits. Ces patches sont dimensionnés 50×50 sans aucun chevauchement. Entre chaque patch (fouée = 50) et toutes les images sont considérées (étape = 0).

Par conséquent, un total de 400 patches sont extraits par image.

Une sélection aléatoire vise à extraire 384 patches par image. La graine est fixée une fois pour toutes les expériences afin d'avoir une sélection de patch reproductible. L'ensemble des correctifs est ensuite mélangé afin de supprimer leur dépendance à une image spécifique.

Plus précisément, nous générons une permutation aléatoire de tous les index de patches, ce qui nous permet de garder le lien entre les images propres (cible) et bruitées (source).

Les patches d'entrée cosinus et sinus sont normalisés entre 0 et 1.

A partir de cette permutation, tous les lots configurés pour la formation réseau sont construits et une matrice numpy finale dimensionnée ($\#batches \times 50 \times 50 \times 1$) est enregistré sur le serveur pour les expériences.

3.2. Architecture réseau

Nous traitons le format matriciel Matlab en entrée, chacune d'elles est convertie au format numpy.

```

def dncnn(input, D, is_training=True, output_channels=1): #D: number of
    ConvBlocks.
    with tf.variable_scope('input'):
        output = tf.layers.conv2d(input, 64, 3, padding='same', activation=tf.nn.relu)
    for layers in range(2, D + 1):
        with tf.variable_scope('block%d' % layers):
            output = tf.layers.conv2d(output, 64, 3, padding='same', name='conv%d' % layers,
                use_bias=False)
            output = tf.nn.relu(tf.layers.batch_normalization(output, training=is_training))
        with tf.variable_scope('output'):
            output = tf.layers.conv2d(output, output_channels, 3, padding='same')
    return input-output

```

Figure 9 : Code Python avec le framework tensorflow (comme tf) qui définit l'architecture du modèle.

Une implémentation Tensorflow a été utilisée comme point de départ (<https://github.com/wbhu/DnCNN-tensorflow>⁵) et adapté avec les matrices Matlab comme entrées (<https://git-lium.univ-lemans.fr/tahon/dncnn-tensorflow-holography/>⁶).

DL-Py est l'implémentation Python utilisée dans cet article. L'architecture est décrite dans Figure 9, où tf désigne la bibliothèque tensorflow et D est le nombre de ConvBlocks.

Lors de l'étape d'entraînement, la convergence est très rapide dans les 10 premières époques, puis la fonction de perte diminue continuellement et lentement.

Le nombre maximum d'époques a été fixé à 200 car les performances n'augmentent pas de manière significative avec le nombre d'époques. Cependant, en raison des contraintes d'utilisation du cluster, la formation doit être arrêtée avant que le temps de calcul ne dépasse une limite de 20 jours. Cette limite a été définie en raison de l'utilisation du cluster.

Le nombre d'époques correspondant à la meilleure erreur de phase est inclus dans Table 6 (cf. Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patchs utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28).

Le modèle final est celui qui obtient les meilleurs résultats avec l'ensemble de développement.

Tous les modèles sont formés sur un serveur de cluster avec des GPU. Des expériences préliminaires ont montré que l'entraînement direct de notre réseau sur une image de phase normalisée (entre $[-\pi; \pi]$) a considérablement diminué les performances comme le montre Table 5.

Les images d'entrée utilisées pour la formation sont des données de phase au format Matlab, leur transformation en cosinus ou en sinus est calculée avec leur transposition, augmentant ainsi la base de données quatre fois. Les images d'entrée cosinus et sinus sont donc comprises entre -1 et 1. Elles sont toutes normalisées en $[0;1]$ pour être cohérentes avec la fonction sigmoïde de sortie.

Vérifiez si c'est le nombre de patchs pour l'entraînement ou l'ajout de bruit en entrée qui apporte de meilleures performances.

Ajoutez plus d'époques.

5 - <https://github.com/wbhu/DnCNN-tensorflow>

6 - <https://git-lium.univ-lemans.fr/tahon/dncnn-tensorflow-holography/>

3.3. Profondeur et architecture du réseau d'évaluation

L'architecture du réseau diffère légèrement de celle proposée dans la section précédente. Le modèle peut être entraîné avec différents niveaux de bruit (depuis $\Delta = 0$ to 2.5), différents modèles de franges de phase sans bruit (de 1 à 5) et différentes profondeurs i.e. nombre différent de ConvBlocks ($D = 4$ ou 16).

Les expériences suivantes visent à évaluer l'influence de ces facteurs sur les performances de débruitage des modèles d'apprentissage profond.

Le nombre de données et de paramètres utilisés pour la formation et l'évaluation des réseaux DL-Py sont indiqués dans Table 2 (cf. Table 2. Paramètres utilisés pour former les réseaux. Δ se trouve pour le bruit de speckle simulé p 28). Le taux d'apprentissage est fixé à $LR = 0.001$, car il a été démontré que ce paramètre a un impact important sur la durée de l'entraînement et les résultats, avec un optimiseur Adam.

Depth of the network: En raison de la grande spécificité des images de phase, l'objectif est de garantir que le réseau ne sera pas surajusté aux données d'entraînement. Pour ce faire, deux réseaux différents sont formés, l'un avec les 16 ConvBlocks d'origine, l'autre avec seulement 4 ConvBlocks. Avec le choix de 4 ConvBlocks comme petit modèle, la formation peut être réalisée rapidement tout en conservant un certain niveau de complexité.

Noise level for training: De plus, le réseau est censé être capable de débruiter les images présentant une large gamme de niveaux de bruit. Par conséquent, l'inclusion de différents niveaux de bruit dans les données de formation pourrait aider le réseau à y parvenir. Pour ce faire, 3 réseaux sont formés sur différentes gammes de bruit.

Learning rate: Par choix, un seul hyper paramètre est considéré, à savoir le taux d'apprentissage. Ce choix est motivé par la comparaison avec des travaux antérieurs. La formation se fait avec $LR = 0.0005$ (comme travail précédent) ou $LR = 0.001$, car il a été démontré que ce paramètre a un impact important sur la durée de l'entraînement.

Method	Train noise $\Delta \lambda$						AVG
	0	1	1,5	2	2,5		
cos/sin (best 100 epochs)	0	0.06	0.08	0.10	0.12	0.15	0.10
phase (best 100 epochs)	0	0.21	0.25	0.28	0.32	0.36	0.28
cos/sin (best 100 epochs)	0, 1, 1.5	0.04	0.05	0.06	0.07	0.09	0.06
phase (best 92 epochs)	0, 1, 1.5	0.17	0.19	0.21	0.23	0.26	0.21
original size	1024 × 1024						
patch size	50 × 50						
# patches per image	384						
batch size	128						
# epochs	< 200						
learning rate	0.001 - 0.0005						
train noise $\Delta \lambda$	0	0 - 1,5	0 - 2,5				
# train original images	4	12	20				
# train images + augment.	16	48	80				
# batches	60	80	300				
# patches	7\,680	23\,040	38\,400				
# eval images	5 patterns × 5 noise levels						
# test images	data1	data20	VibMap				

Table 5. Paramètres utilisés pour entraîner nos modèles DL-Py.

Pour le reste des évaluations, seuls les transformés cosinus/sinus normalisés sont utilisés pour entraîner les réseaux. Comme l'ajout de données de phase corrompues par du bruit speckle semble donner de bonnes performances, nous conserverons différents niveaux de bruit dans notre ensemble de données d'entraînement.

3.4. Surajustement et influence du nombre de couches du CNN.

L'architecture a d'abord été conçue pour le débruitage naturel des images, où la diversité des contours de l'image est assez grande. Dans le domaine du débruitage des images de phase, la tâche est probablement moins diversifiée. On peut également noter que la quantité d'images naturelles est près de 20 fois supérieure à la quantité de patchs utilisés lors des premières expériences dans la Table 2 (cf. Table 2. Paramètres utilisés pour former les réseaux. Δ se trouve pour le bruit de speckle simulé p 28) réalisé sur des images de phase corrompues par

du bruit speckle. La faible quantité de données pour l'entraînement et la diversité réduite des images traitées peuvent conduire à des réseaux formés très spécialisés. Si le modèle est trop spécialisé pour traiter des images proches de celles de l'ensemble d'apprentissage, les résultats sur de telles images peuvent être très impressionnants, tandis que les résultats sur des images invisibles chuteront considérablement en performances.

Nous avons émis l'hypothèse que la réduction du nombre de couches du réseau pourrait aider à ne pas surajuster les données et aider à la formation de modèles moins spécialisés. Le réseau de référence contient une première couche convolutive (3×3 et unités linéaires rectifiées ReLU), $D = 16$ couches convolutives intermédiaires ($3 \times 3 \times 64$ + normalisation par lots et ReLU), et une dernière couche convolutive ($3 \times 3 \times 64$ est utilisé pour reconstruire le bruit de sortie. Nous avons décidé de former des tailles différentes pour le nombre de couches D .

3.5. Évaluation d'un réseau pré-formé

Dans un deuxième temps, nous évaluerons l'amélioration du réseau pré-entraîné sur des images naturelles avec un bruit gaussien supplémentaire. Ensuite, adaptez-le aux images de phase holographiques, ou à l'utilisation directe d'un réseau entièrement formé avec des images de phase holographiques.

En effet, les images de phase sont très spécifiques (motifs de franges) par rapport aux images naturelles, et le bruit de speckle est différent du bruit gaussien, principalement parce qu'il dépend de la fréquence et ne peut être modélisé avec du bruit supplémentaire.

Dans cette section, nous présentons les résultats obtenus en pré-entraînant le réseau sur des images naturelles, puis en l'affinant sur des images de phase holographique.

3.6. Protocole

400 images de la base de données NATURAL sont utilisées pour pré-entraîner le réseau avec la meilleure architecture obtenue dans la section précédente, *i.e.* 4 ConvBlocks (voir section "Résultats et discussion"). Une fois le réseau pré-entraîné, une deuxième étape de réglage fin est réalisée à l'aide d'images holographiques suivant le protocole précité. Le modèle DL-nat-pt correspond au modèle entraîné avec des images naturelles pendant 75 époques, ce qui semble raisonnable au regard des 50 époques utilisées pour entraîner le DnCNN original [8] [Computational image speckle suppression using block matching and machine learning]. Sans mise au point, ce modèle atteint $\Delta\phi = 0.380 \text{ rad}$ avec le kit de développement, qui ne convient pas du tout aux images holographiques. Les résultats du réglage fin sont présentés dans la section suivante.

Cependant, son adaptation aux images holographiques améliore significativement les performances de ce modèle préliminaire. De la Table 6 (cf. Table 6. Erreurs de phase ($\Delta\phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patches utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28) on voit également que, pour des époques similaires, les performances atteintes par les modèles pré-entraînés sont comparables à celles obtenues sans pré-entraînement.

4. Résultats et discussion

4.1. Profondeur et architecture du réseau

Les résultats obtenus avec HOLODEEP sont résumés dans la Table 6 (cf. Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patches utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28). Pour aider le lecteur, les noms de modèles explicitent les différents paramètres : DL-Py-X-D-z avec X étant le maximum Δ dans les données d'entraînement, D représente la profondeur du modèle ($D = 4$ ou $D = 16$) et le z facultatif indique si le modèle a été préalablement entraîné sur des images naturelles (pt) ou non.

Lorsque le bruit d'entraînement est $\Delta = 0$, les meilleurs résultats sont obtenus avec un réseau complexe (DL-Py-0-16, $\Delta \phi = 0.057 \text{ rad}$). Cependant, globalement, les meilleurs résultats sont obtenus avec seulement 4 ConvBlocks et une large gamme de bruit d'entraînement (DL-Py-2.5-4, $\Delta \phi = 0.035 \text{ rad}$).

Nous concluons que la complexité du réseau n'est pas bénéfique lorsque le modèle est appris uniquement sur de petits bruits, probablement parce que les réseaux ne se généralisent pas bien à d'autres niveaux de bruit.

L'introduction de la diversité des niveaux de bruit permet de réduire considérablement l'erreur de phase moyenne pour toutes les configurations. Surtout la meilleure configuration ($D = 4$ ConvBlocks) abaisse $\Delta \phi$ de 0.058 rad ($\Delta = 0$) à 0.035 rad ($\Delta = 0 - 2.5$).

Cela suggère qu'un réseau réduit formé avec une grande diversité est probablement plus à même de généraliser qu'un réseau profond formé avec très peu de données. Un point reste incertain, on ne sait pas si l'amélioration observée sur le débruitage est due à la diversité du bruit, ou à la plus grande quantité de données utilisées pour entraîner le réseau.

L'avantage d'utiliser un plus petit nombre de couches est que le temps de calcul est plus de deux fois inférieur.

Presque 1 jour avec $D = 4$ et 1,5 jours avec $D = 16$.

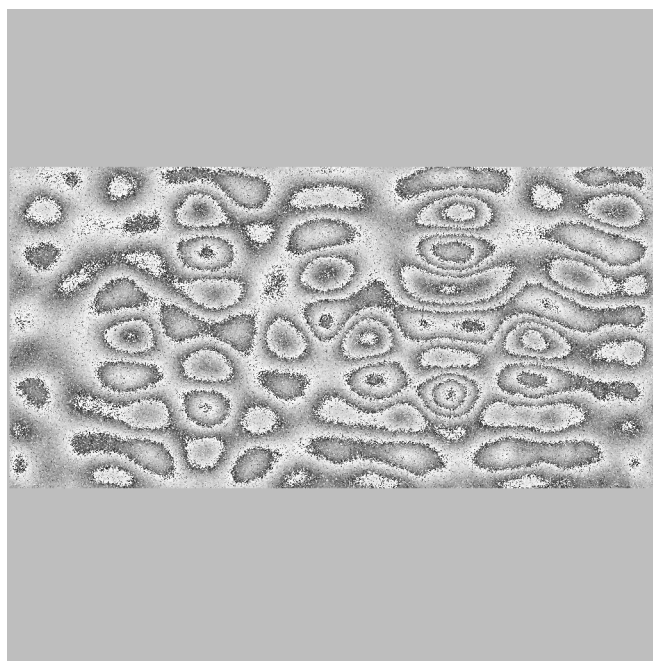


Figure 6 : Phase Test 3 bruyant

L'analyse des résultats en fonction du niveau de bruit speckle dans les images HOLODEEP confirme que plus le niveau de bruit est élevé, plus l'erreur dans la carte de phase restaurée est élevée.

Table 6 (cf. Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patches utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28) détaille les valeurs obtenues lors de l'évaluation sur HOLODEEP, en fonction de leur niveau de bruit (paramètre Δ) avec les trois meilleurs modèles DL-Py-0-4 (série de niveau de bruit $\Delta = 0$), DL-Py-1.5-4 (série de niveau de bruit $\Delta = 0 - 1.5$) and DL-Py-2.5-4 (série de niveau de bruit $\Delta = 0 - 2.5$).

Comme mentionné ci-dessus, le DL-Py-2.5-4 est meilleur en moyenne que le DL-Py-0-4 sur HOLODEEP. Mais les expériences complémentaires montrent que cette amélioration des performances est significativement plus importante sur les images à fort niveau de bruit (-49% de réduction relative avec $\Delta = 2.5$) qu'avec des images à faible bruit (-31% with $\Delta = 0$). Ces résultats soulignent la pertinence d'introduire une grande diversité de motifs et de niveaux de bruit lors de l'étape d'apprentissage si les images applicatives à traiter présentent également des niveaux de bruit élevés.

4.2. Pré-formation

Afin d'évaluer le réseau, nous utilisons la base de données DATAEVAL qui contient 3 images qui n'ont pas été vues lors des étapes de formation et de validation. Il existe une petite différence entre les résultats présentés dans la Table 1 pour le set de développement (25 images) et ceux présentés dans la Table 6 (cf. Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patches utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28). Cela est dû au fait que dans la Table 6 (cf. Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patches utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28) le débruitage est appliqué en utilisant le modèle extrait à la meilleure époque, tandis que dans la Table 1 le débruitage est appliqué en utilisant le modèle extrait à la dernière époque. Avec le cadre actuel, nous ne sommes pas en mesure de sélectionner un point de contrôle spécifique. C'est la raison pour laquelle nous aimerions passer au framework PyTorch à l'avenir.

La Table 6 (cf. Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patches utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle. p 28) montre que le modèle pré-entraîné surpasse les modèles initiaux uniquement lorsqu'un faible niveau de bruit ($\Delta = 0$) est utilisé pour le réglage fin. Cela conduit à la conclusion que la pré-entraînement du réseau sur des images naturelles permet de compenser le manque de diversité des données d'entraînement spécifiques ainsi que la quantité relativement faible de données d'entraînement. Ces résultats confirment l'avantage d'utiliser des modèles pré-entraînés lorsque la quantité de données cibles spécifiques est faible [12] [On the use of self-supervised pre-trained acoustic and linguistic features for continuous speech emotion recognition].

Deux hypothèses peuvent expliquer les mauvaises performances atteintes par le modèle pré-entraîné.

Les bases de données NATURAL et HOLODEEP diffèrent sur de nombreux points : bruit de speckle additif gaussien ou multiplicatif et images de phase naturelle ou enveloppée. Une telle différence de données pourrait expliquer les mauvaises performances obtenues avec le pré-entraînement : entraîner un réseau avec des images de phase à l'aide d'une initialisation obtenue sur la base de données NATURAL ne semble pas intéressant dans le cas présent.

Par conséquent, la formation d'un réseau avec des données de phase corrompues par du bruit speckle nécessite une étude plus approfondie.

La deuxième hypothèse concerne les performances du modèle formé sur des données NATURELLES. En raison des contraintes d'utilisation du cluster, le nombre total d'époques pour entraîner ce modèle est de 75 époques. Il vise à obtenir un modèle performant sur des images naturelles. Cependant, ce nombre est supérieur aux 50 époques utilisées pour entraîner le modèle DnCNN original mentionné dans [11] [Beyond a Gaussian denoiser: Residual learning of Deep CNN for image denoising]) et le modèle pourrait être trop spécifique pour les images naturelles. Comme la formation de tels modèles nécessite beaucoup de ressources, nous n'avons pas eu la possibilité de nous entraîner sur un plus grand nombre d'époques. Cependant, cet aspect mérite d'être pris en compte.

a) Effet des itérations

Dans le domaine du débruitage des images speckle, les images sont généralement décodées plusieurs fois. Par conséquent, l'image bruitée est transmise plusieurs fois à travers le réseau, comme le montre la Table 1 où 6 itérations ont été utilisées pour débruiter les images de test avec les algorithmes BM3D et DtDWT.

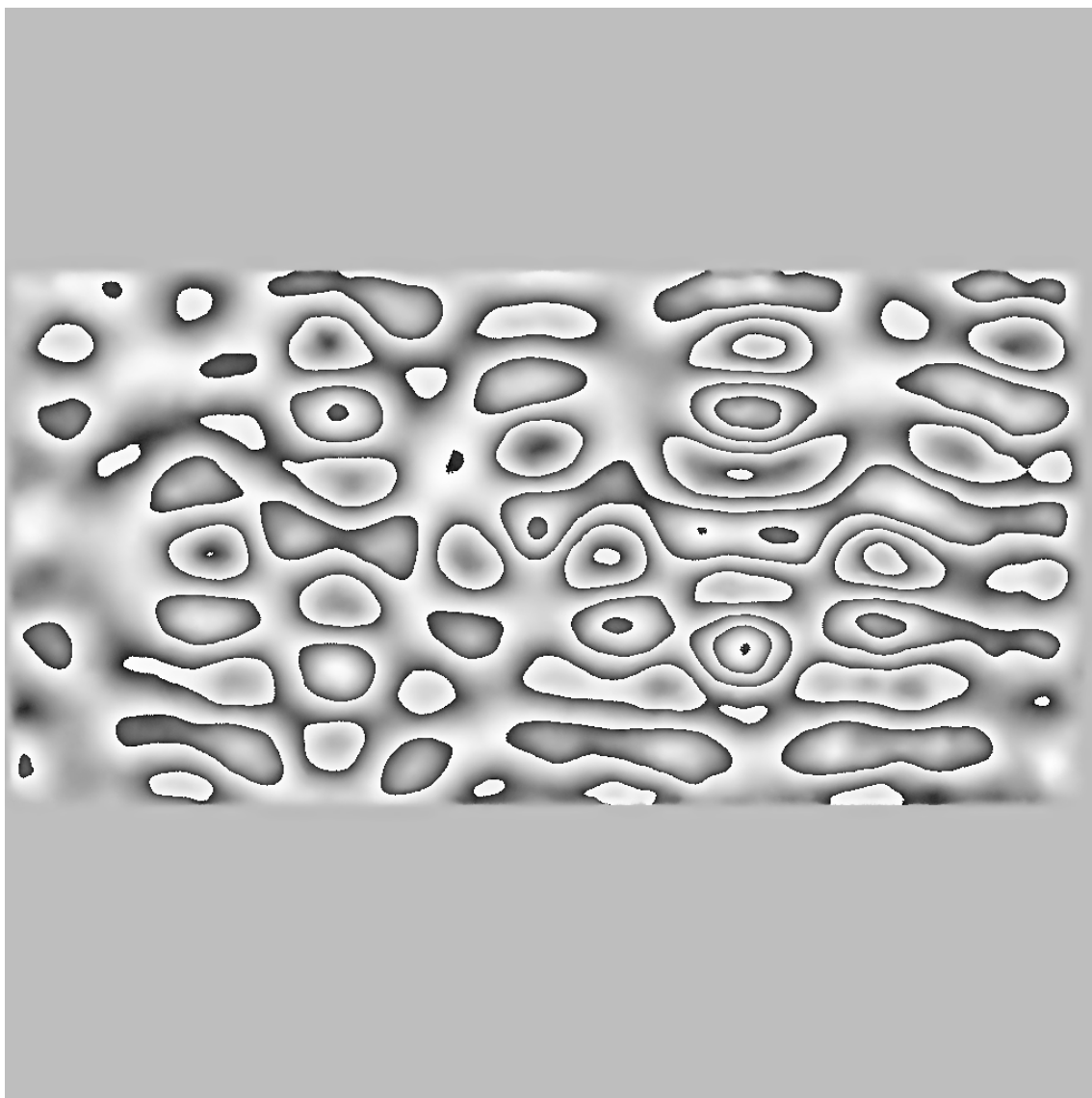


Figure 7 : Phase 3 Test Dé-bruit

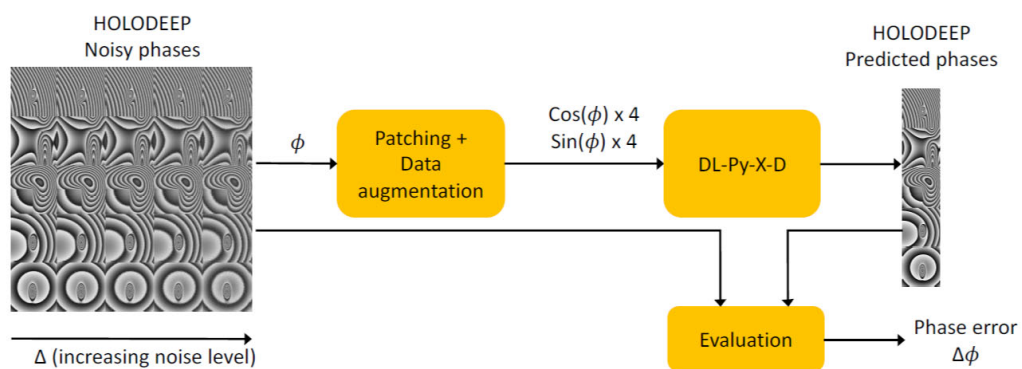


Figure 8 : Aperçu global de la phase de formation du système

4.3. Évaluation sur images cibles

Method	HOLODEEP	DATAEVAL		
	25 Images	Test1	Test2	Test3
WFT2F	0.026	0.044	0.163	0.105
DL-3	0.041	0.107	0.585	0.105
DL-Py-0-4	0.058	0.142	0.629	0.117
DL-Py-0-4-pt	0.055	0.146	0.629	0.105
DL-Py-1.5-4	0.040	0.095	0.593	0.103
DL-Py-1.5-4-pt	0.045	0.112	0.609	0.111
DL-Py-2.5-4	0.035	0.072	0.597	0.109
DL-Py-2.5-4-pt	0.048	0.097	0.660	0.134

Tableau 9. $\Delta\phi$ (rad) obtenu sur la base de données HOLODEEP (en moyenne) et images individuelles de DATAEVAL avec une itération. Meilleures époques pour les modèles pré-entraînés et entraînés sur la base de données de validation HOLODEEP

La Table 9 résume les performances obtenues avec les images de développement et de validation.

DL-Py-2.5-4 fonctionne mieux sur les données d'entraînement HOLODEEP ($\Delta\phi = 0.035$ rad) et sur Test1 ($\Delta\phi = 0.072$ rad).

Cependant, les performances sont dégradées lors des tests avec Test2 qui a un niveau de bruit élevé et avec Test3, qui est l'image de phase issue des expériences de vibration. Aucune réponse claire ne peut être donnée ici. Le modèle DL-Py-2.5-4 est entraîné sur un grand nombre de données et de bruit, il devrait donc être capable de gérer un niveau de bruit élevé.

Cependant, par construction de la base de données HOLODEEP, il y a peu de redondances dans les images de phase, et Test1 semble relativement similaire à ceux de HOLODEEP, tandis que Test2 et Test3 ne sont pas.

Par conséquent, le modèle pourrait ne pas être en mesure de se généraliser facilement à des images invisibles.

Une autre hypothèse est que la structure du modèle implique un bruit additif, ce qui pourrait être pertinent pour un faible SNR, mais pas pour un SNR élevé où le bruit de speckle est clairement multiplicatif.

On peut remarquer que si le modèle entraîné sur des images de phase avec une grande diversité de bruit (DL-Py-2.5-4) atteint les meilleures performances sur l'ensemble de développement ($\Delta\phi = 0.035$ rad), cela ne se généralise pas bien sur les images inédites de DATAEVAL test1 l'image est très similaire à celles contenues dans HOLODEEP.

Le modèle qui généralise le mieux sur le test2 et Test3 est celui formé sur une plage moyenne de bruit speckle (DL-Py-1.5-4). Ce modèle est même capable de surpasser le WFT2F de base sur la carte de vibration expérimentale Test3 image de phase. La figure 10 (cf. Figure : Noise-free (left), noisy (middle) and de-noised (right) phase images from DATAEVAL. De-noising is performed with DL-Py-1.5-4 mode. p 29) montre comment ces images de DATAEVAL sont débruitées par le meilleur modèle. Ainsi, les réseaux proposés sont capables d'atteindre des performances intéressantes par rapport à WFT2F, notamment pour certaines images expérimentales spécifiques. Ces réseaux ont l'avantage d'être plus rapides à former que le réseau DL-3 car ils ne contiennent que 4 ConvBlocks.

Concernant les modèles pré-entraînés, il semble qu'ils ne soient pas capables de généraliser sur des images inédites sauf DL-Py-0-4-pt qui obtient $\Delta\phi = 0.105$ rad avec Test3. Des expériences supplémentaires montrent que les modèles entraînés avec plus d'époques peuvent améliorer les performances sur Test1, mais se dégrade en Test2 et Test3.

III. Conclusion

Cet article traite du débruitage des images de phase holographiques et présente une approche alternative spécifique au bruit de speckle. Les résultats montrent qu'un modèle pré-entraîné n'est utile que si la quantité et la diversité des données simulées sont faibles. Dans ce cas, la pré-formation compense le manque de données. Les expériences démontrent également que l'utilisation de réseaux très profonds n'est pas nécessaire et que l'utilisation de 4 ConvBlocks donne des performances fiables par rapport à WFT2F. Les réseaux réduits ont également l'avantage d'être plus rapides à former.

Cette étude aborde également la question du pouvoir de généralisation des réseaux. Il apparaît que WFT2F reste le meilleur algorithme pour les images de phase à fort niveau de bruit

(**Test2**). Cependant, le meilleur modèle est capable de surpasser la référence de WFT2F avec des données expérimentales (**Test3**).

Les mauvaises performances des modèles DL-Py avec image de phase avec un niveau de bruit élevé peuvent être liées à l'hypothèse additive mise en œuvre dans le réseau lui-même. Un modèle multiplicatif sera étudié dans le futur.

Des travaux ultérieurs visent à améliorer le débruitage du speckle en combinant les avantages des deux approches suite à des travaux préliminaires sur l'ajout d'un estimateur de bruit. [25] [An iterative scheme based on deep learning combined with input noise estimator for phase data processing in digital holographic interferometry].

D'autres fonctions d'augmentation des données seront mises en œuvre afin d'augmenter la quantité de données d'entraînement.

De plus, la construction d'une nouvelle base de données avec une diversité accrue d'images de franges serait intéressante pour former des réseaux avec une grande diversité de motifs.

IV.Remerciements

Nous remercions le LIUM pour l'accès au cluster.

Les travaux de recherche ne bénéficient d'aucun financement externe.

Nous remercions le LIUM pour l'autorisation d'accès au cluster GPU.

Les auteurs ne déclarent aucun conflit d'intérêts.

Ressources annexes

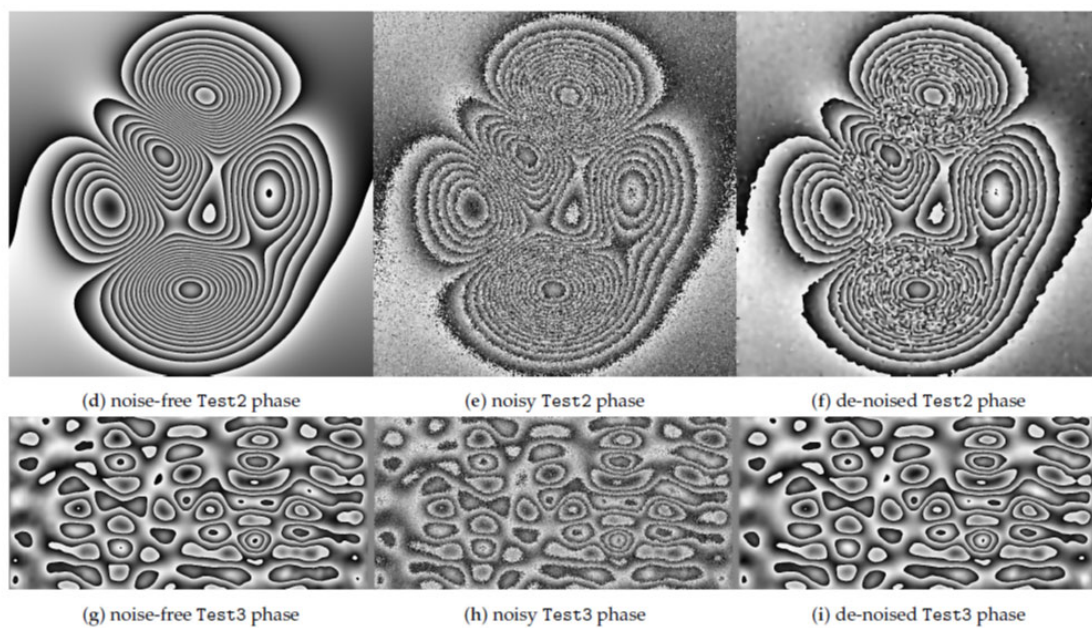
- Table 2. Paramètres utilisés pour former les réseaux. Δ se trouve pour le bruit de speckle simulé

	DnCNN [11]	DL-3 [20]	DL-Py		
original size	180 × 180	1024 × 1024	1024 × 1024		
patch size	50 × 50	50 × 50	50 × 50		
batch size	128	128	128		
learning rate	0.1 to 0.001	0.0006	0.001 ; 0.0005		
# epochs	50	1920	<200		
noise type	Gaussian	Gauss+speckle	speckle		
noise	$\sigma \in [0; 55], \mu = 0$	$\Delta = 0$	$\Delta = 0$	$\Delta = 0 - 1,5$	$\Delta = 0 - 2,5$
SNR (dB) range	>13	7.32 - 11.46	7.32 - 11.46	5.08 - 11.46	3.10 - 11.46
# train images	400	5 × 8 = 40	5 × 8 = 40	5 × 3 × 8 = 120	5 × 5 × 8 = 200
# patches	128 × 3000 = 384K	384 × 40 = 15.3k	384 × 40 = 15.3k	384 × 120 = 46.1k	384 × 200 = 76.8k

- Table 6. Erreurs de phase ($\Delta \phi$ en rad), obtenues avec une itération sur HOLODEEP. Meilleures configurations en caractères gras. Trois ensembles d'entraînement sont utilisés, chacun correspondant à une plus grande diversité de bruit, le nombre de patchs utilisés pour entraîner le modèle dans chaque cas est donné. Les noms de modèles sont donnés pour chaque configuration. La meilleure époque est donnée par rapport au nombre total d'époques utilisées pour entraîner le modèle.

Δ (#patch)	<i>D</i>	Trained on HOLODEEP		Pre-Trained
		16	4	4
0 (15.3k)	model	DL-Py-0-16	DL-Py-0-4	DL-Py-0-4-pt
	BestEpoch/Max	195/200	200/200	190/200
	$\Delta \phi$	0.057	0.058	0.055
ine 0-1.5 (46.1k)	model	DL-Py-1.5-16	DL-Py-1.5-4	DL-Py-1.5-4-pt
	BestEpoch/Max	70/70	140/150	85/95
	$\Delta \phi$	0.042	0.040	0.045
ine 0-2.5 (76.8k)	model	DL-Py-2.5-16	DL-Py-2.5-4	DL-Py-2.5-4-pt
	BestEpoch/Max	40/50	90/95	50/55
	$\Delta \phi$	0.038	0.035	0.048

- Figure : Noise-free (left), noisy (middle) and de-noised (right) phase images from DATAEVAL. De-noising is performed with DL-Py-1.5-4 mode.



Bibliographie

[Beyond a Gaussian denoiser: Residual learning of Deep CNN for image denoising] K. ZHANG AND W. ZUO AND Y. CHEN AND D. MENG AND L. ZHANG , *Beyond a Gaussian denoiser: Residual learning of Deep CNN for image denoising, IEEE Transactions on Image Processing*, 2017

[Computational de-noising based on deep learning for phase data in digital holographic interferometry] MONTRESOR,S. AND TAHON,M. AND LAURENT,A. AND PICART,P., *Computational de-noising based on deep learning for phase data in digital holographic interferometry, APL Photonics*, 2020

[Deep residual learning for image recognition] K. HE AND X. ZHANG AND S. REN AND J. SUN, *Deep residual learning for image recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016

[Depth selection for deep ReLU nets in feature extraction and generalization] HAN, ZHI AND YU, SIQUAN AND LIN, SHAO-BO AND ZHOU, DING-XUAN, *Depth selection for deep ReLU nets in feature extraction and generalization, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE*, 2020

[Digital Holography] P. PICART AND J.C. LI, John Wiley & Sons, Ltd, *Digital holography*, 2012

[Generative adversarial nets] GOODFELLOW, I. J. AND POUGET-ABADIE, J. AND MIRZA, M. AND XU, B. AND WARDE-FARLEY, D. AND OZAIR, S. AND COURVILLE, A. AND BENGIO, Y., *Generative adversarial nets, Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS) - Volume 2, MIT Press*, 2014

[High-speed holographic metrology: principle, limitations, and application to vibroacoustics of structures] POITTEVIN, JULIEN AND GAUTIER, FRANÇOIS AND PÉZERAT, CHARLES AND PICART, PASCAL, *High-speed holographic metrology: principle, limitations, and application to vibroacoustics of structures, Optical Engineering, International Society for Optics and Photonics*, 2016

[Improvement of accuracy in digital holography by use of multiple holograms] BAUMBACH, TORSTEN AND KOLENOVIC, ERVIN AND KEBBEL, VOLKER AND JÜPTNER, WERNER, *Improvement of accuracy in digital holography by use of multiple holograms, Applied Optics*, 2006.

[Natural image denoising with convolutional networks] JAIN, V. AND SEUNG, S., *Natural image denoising with convolutional networks, Advances in Neural Information Processing Systems*, D. Koller and D. Schuurmans and Y. Bengio and L. Bottou, 2008

[On the use of deep learning for computational imaging] BARBASTATHIS, G. AND OZCAN, A. AND SITU, G., *On the use of deep learning for computational imaging, Optica*, 2019

[On the use of self-supervised pre-trained acoustic and linguistic features for continuous speech emotion recognition] MACARY, M.; TAHON, M.; ESTÈVE, Y.; ROUSSEAU, A., *On the use of self-supervised pre-trained acoustic and linguistic features for continuous speech emotion recognition in 2021 IEEE Spoken Language Technology Workshop (SLT)*, 2021

[The effectiveness of data augmentation in image classification using deep learning] LUIS PEREZ AND JASON WANG, *The effectiveness of data augmentation in image classification using deep learning, arXiv:1712.04621*, 2017

[Two-dimensional phase unwrapping: theory, algorithms, and software] GHIGLIA, DENNIS C. AND PRITT, MARK D, *Two-dimensional phase unwrapping: theory, algorithms, and software, Wiley*, 1998

[Very deep convolutional networks for large-scale image recognition] K. SIMONYAN AND A. ZISSERMAN, *Very deep convolutional networks for large-scale image recognition, 3rd International Conference on Learning Representations, ICLR 2015, 2015*

[Visualization of travelling waves propagating in a plate equipped with {2D ABH} using wide-field holographic vibrometry] LAGNY, L. AND SECAIL-GERAUD, M. AND LE MEUR, J. AND MONTRESOR, S. AND HEGGARTY, K. AND PEZERAT, C. AND PICART, P., *Visualization of travelling waves propagating in a plate equipped with {2D ABH} using wide-field holographic vibrometry, Journal of Sound and Vibration, 2019*

[Windowed Fourier transform for fringe pattern analysis: theoretical analyse] Q. KEMAO AND H. WANG AND W. GAO, *Windowed Fourier transform for fringe pattern analysis: theoretical analyse, Applied Optics, OSA, 2008*

Webographie

[An iterative scheme based on deep learning combined with input noise estimator for phase data processing in digital holographic interferometry] S. MONTRESOR AND M. TAHON AND A. LAURENT AND P. PICART, *An iterative scheme based on deep learning combined with input noise estimator for phase data processing in digital holographic interferometry, Imaging and Applied Optics Congress, Optical Society of America, 2020,*

[Computational image speckle suppression using block matching and machine learning] T. ZENG AND H. K.-H. SO AND E. Y. LAM, *Computational image speckle suppression using block matching and machine learning, Applied Optics, OSA, 2019,*

[Deep Learning] I. GOODFELLOW AND Y. BENGIO AND A. COURVILLE, *Deep Learning, MIT Press, 2016,*

[Error analysis for noise reduction in 3D deformation measurement with digital color holography] S. MONTRESOR AND P. PICART AND O. SAKHARUK AND L. MURAVSKY, *Error analysis for noise reduction in 3D deformation measurement with digital color holography, Journal of the Optical Society of America B, OSA, 2017,*

[Experimental and theoretical investigation of the pixel saturation effect in digital holography] PICART, PASCAL AND TANKAM, PATRICE AND SONG, QINGHE, *Experimental and theoretical investigation of the pixel saturation effect in digital holography, Journal of the Optical Society of America A, 2011*

[Quality assessment of combined quantization-shot-noise-induced decorrelation noise in high-speed digital holographic metrology] POITTEVIN, JULIEN AND PICART, PASCAL AND GAUTIER, FRANÇOIS AND PEZERAT, CHARLES, *Quality assessment of combined quantization-shot-noise-induced decorrelation noise in high-speed digital holographic metrology, Optical Express, Optical Society of America, 2015.*

[Quantitative appraisal for noise reduction in digital holographic phase imaging] S. MONTRÉSOR AND P. PICART, *Quantitative appraisal for noise reduction in digital holographic phase imaging, Optics Express, OSA, 2016*

[Refocus criterion based on maximization of the coherence factor in digital three-wavelength holographic interferometry] PICART, PASCAL AND MONTRESOR, SILVIO AND SAKHARUK, OLEKSANDR AND MURAVSKY, LEONID, *Refocus criterion based on maximization of the coherence factor in digital three-wavelength holographic interferometry, Optics Letters, Optical Society of America, 2017*

[Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks] W. JEON AND W. JEONG AND K. SON AND H. YANG, *Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks, Opt. Lett., OSA, 2018,*