Apprentissage de méthodes itératives pour l'imagerie de contraste de phase des rayons X

Kannara Mom, Max Langer, Bruno Sixou

Apprentissage de méthodes itératives pour l'imagerie de contraste de phase des rayons X

Kannara MOM[1], Max LANGER[2], Bruno SIXOU[1]

[1]Laboratoire CREATIS
21 avenue Jean Capelle, 69621, Villeurbanne cedex, France

[2]Laboratoire TIMC
Rond-Point de la Croix de Vie, Domaine de la Merci, 38706 La Tronche cedex, France

kannara.mom@creatis.insa-lyon.fr, bruno.sixou@insa-lyon.fr
max.langer@univ-grenoble-alpes.fr

Résumé – L'imagerie par contraste de phase peut être obtenue avec des rayons X si le faisceau est suffisamment cohérent. Cependant, la récupération de la phase et de l'absorption à partir d'une ou plusieurs de ces images est un problème inverse non linéaire et mal posé. Nous proposons ici une méthode d'apprentissage profond pour résoudre ce double problème à partir d'une seule mesure d'intensité. Notre méthode est basée sur l'approche unrolling, qui consiste à dérouler un algorithme itératif et à remplacer une partie des itérations avec un réseau de neurones convolutifs. Dans notre travail, nous n'avons pas à choisir de régularisation, seul le gradient du terme d'attache aux données est fourni à chaque itération, et le réseau est entraîné à apprendre une itération complète à partir de cette information et de l'itération en cours. Nous comparons un schéma de descente de gradient avec pour régularisation la variation totale lissée (GD-TV$^\epsilon$) et sa version unrolling (DUGD). Nous montrons que la version unrolling permet d'obtenir de meilleurs résultats quantitativement et qualitativement, et que le temps de calcul nécessaire est beaucoup plus court que la version itérative classique.

Abstract – In-line phase contrast imaging can be obtained with X-rays if the beam is sufficiently coherent. Retrieving the phase and absorption from one or several of these images is a non-linear, ill-posed inverse problem, however. Here, we propose a deep learning method to solve this twofold problem from a single intensity measurement. It is based on the unrolling approach, in which an iterative algorithm is "unrolled" into a sequence, and some part of the iterations is replaced with a convolutional neural network. An advantage is that no regularization is needed, only the gradient of the data fidelity term is calculated at each iteration. The network is trained to learn a complete iteration from this information and the current iteration. We apply the method to a gradient descent scheme using smooth total variation regularization (GD-TV$^\epsilon$) and compare the standard and the unrolling version (DUGD). We show that the unrolled version gives better results quantitatively and qualitatively, and that the computation time is much shorter than the classical iterative version.

1 Introduction

L'imagerie par contraste de phase a permis d'augmenter la sensibilité en tomographie à l'échelle microscopique et nanoscopique [1], elle est maintenant largement utilisée en science des matériaux et dans l'imagerie biomédicale [2]. La relation non-linéaire entre l'absorption et le décalage de phase induit par l'objet et les intensités mesurées repose sur la théorie de diffraction de Fresnel et traduit un problème inverse non-linéaire et mal posé. L'estimation de ce décalage (et éventuellement de l'absorption) est un processus appelé récupération de phase. Malgré les efforts récents, la récupération de phase reste un problème difficile, les principales difficultés résident dans : **(i)** l'obtention d'une haute résolution spatiale des images reconstruites, **(ii)** être capable de reconstruire simultanément l'absorption et la phase à partir d'une ou plusieurs mesures d'intensité, **(iii)** limitation des bruits/artefacts basses fréquences présents dans les reconstructions et **(iv)** réduire le temps de calcul ainsi que la mémoire nécessaire. Il y a deux principales catégories de méthodes existantes, l'inversion directe et les méthodes itératives. La première se base sur la linéarisation du problème qui permet d'avoir une solution analytique, donc une reconstruction rapide mais qui n'est valide que sous certaines hypothèses. On retrouve notamment la méthode Paganin qui repose sur l'équation de transport d'intensité (TIE), ou encore la méthode Contrast Transfer Function (CTF). La seconde catégorie est celle des méthodes itératives, elles consistent à minimiser de manière itérative le terme d'attache aux données. Ces approches ne sont pas limitées aux contraintes que peuvent avoir les méthodes analytiques. Parmi elles se trouvent les méthodes qui projettent sur des contraintes ou les approches variationnelles qui impliquent un terme de régularisation. Ces méthodes permettent d'avoir de bonnes reconstructions avec moins d'artefacts mais aussi de prendre en compte la non linéarité du problème. Cependant, le temps de calculs et la mémoire requise sont importants, de plus, le choix d'une régularisation appropriée reste un problème difficile.

Les méthodes d'apprentissage profond ont beaucoup évolué ces dernières années pour des tâches de traitement d'image et de signal. Des approches récentes ont montré des résultats prometteurs pour la reconstruction dans plusieurs problèmes

inverses [3]. Plusieurs architectures ont été proposées pour répondre à ce problème de récupération de phase, notamment MS-D Net [4] et PhaseGAN [5], qui entraînent un réseau à approcher l'opérateur inverse, afin d'avoir une construction directe. D'autres approches sont basées sur les méthodes itératives et incorporent des réseaux de neurones convolutifs en déroulant ces algorithmes. Cette approche, dite *unrolling*, permet entre autre d'avoir les avantages des solutions itératives, mais également de meilleures reconstructions et un temps d'exécution plus court. Notre travail ici est de développer une méthode unrolling pour le problème de récupération de phase et d'absorption.

2 Formation de l'image

Pour des objets minces et une propagation rectiligne du faisceau, l'interaction du faisceau de rayons X cohérents et parallèles avec la matière peut être décrite par une fonction de transmittance T. Dans le cadre de la théorie de la diffraction de Fresnel, le fait de laisser le faisceau se propager dans l'espace libre sur une distance D relativement courte après interaction avec l'objet peut être décrit comme une convolution 2D de la transmittance et du propagateur de Fresnel. Le modèle direct décrivant la relation non linéaire entre l'absorption B et le déphasage φ induits par un échantillon et l'intensité détectée peut s'écrire de la façon suivante

$$F_D\left(B,\varphi\right) = |\mathrm{e}^{-B+i\varphi} * P_D|^2 \tag{1}$$

où le propagateur de Fresnel est donné par

$$P_D(\mathbf{x}) = \frac{1}{i\lambda D}\exp\left(i\frac{\pi}{\lambda D}|\mathbf{x}|^2\right) \tag{2}$$

L'estimation du déphasage (resp. de l'absorption) à partir d'une ou plusieurs de ces mesures d'intensités est appelée récupération de phase (resp. d'absorption). Notre objectif est d'estimer à la fois la phase et l'absorption à partir d'une seule mesure d'intensité.

3 Dérouler une méthode itérative

Ici, nous considérons les couples (B,φ) de reconstructions qui sont des solutions du problème d'optimisation suivant :

$$\min_{B,\varphi}\left\{\mathcal{J}\left[(B,\varphi),I_D^{\mathrm{obs}}\right] := d\left[F_D(B,\varphi),I_D^{\mathrm{obs}}\right] + R(B,\varphi)\right\} \tag{3}$$

où I_D^{obs} est une intensité (bruitée) mesurée à une certaine distance D, $d\left[F_D(B,\varphi),I^{\mathrm{obs}}\right]$ est un terme d'attache aux données et $R(B,\varphi)$ un terme de régularisation. Le terme d'attache aux données dépend d'une fonction de coût d et oblige le couple recherché à s'ajuster aux données observées. Le terme de régularisation force la solution à satisfaire des informations a priori.

3.1 Un algorithme de descente (GD-TV$^\epsilon$)

Reprenons le problème (3), et prenons pour d la norme L^2 :

$$d\left[F_D(B,\varphi),I_D^{\mathrm{obs}}\right] = \frac{1}{2}\left|\left|F_D(B,\varphi) - I_D^{\mathrm{obs}}\right|\right|_2^2 \tag{4}$$

et pour régularisation R la variation totale lissée par un paramètre ϵ : $\mathrm{TV}^\epsilon(u) := \int_\Omega \sqrt{\epsilon^2 + |\nabla u(x)|^2}\mathrm{d}x$ où $\Omega \subset \mathbb{R}^2$ est le domaine de l'image. On est donc ramené à minimiser la fonctionnelle

$$\mathcal{J}_\epsilon(B,\varphi) = \frac{1}{2}\left|\left|F_D(B,\varphi) - I_D^{\mathrm{obs}}\right|\right|_2^2 + \eta\mathrm{TV}^\epsilon(B) + \mu\mathrm{TV}^\epsilon(\varphi) \tag{5}$$

où $\eta,\nu > 0$ sont des paramètres de régularisation. Ceci peut être fait avec un algorithme de descente de gradient projeté, afin de contraindre les solutions recherchées à être positives, on appellera cette approche (GD-TV$^\epsilon$). Le gradient du terme d'attache aux données étant donné par

$$\nabla d\left[F_D(B,\varphi),I_D^{\mathrm{obs}}\right] = \left[F_D'(B,\varphi)\right]^*\left(F_D(B,\varphi) - I_D^{\mathrm{obs}}\right) \tag{6}$$

où $[F_D'(B,\varphi)]^*$ est l'adjoint de la dérivée de Fréchet de l'opérateur direct au point (B,φ), dont on a une formule analytique [8].

3.2 Apprentissage d'une itération

L'approche dite *unrolling* consiste à remplacer une itération par un réseau de neurones convolutionnel ce qui permet de généraliser de nombreux algorithmes itératifs [3]. Elle a trouvé de nombreuses applications dans divers problèmes en traitement d'image [6]. Formellement, en supposant que l'application d de (3) est différentiable, une descente de gradient que nous arrêtons après N itérations peut s'écrire, pour $k = 1,\ldots,N$:

$$\left\{(B_{k+1},\varphi_{k+1}) = (B_k,\varphi_k) - \theta_k\nabla\mathcal{J}\left[(B_k,\varphi_k),I_D^{\mathrm{obs}}\right]\right. \tag{7}$$

Dans le schéma de descente de gradient défini par (7)

$$(B_N,\varphi_N) = \left(\Gamma_{\theta_N}\circ\cdots\circ\Gamma_{\theta_1}\right)(B_0,\varphi_0) \tag{8}$$

où $\Gamma_{\theta_k} := \mathrm{Id} - \theta_k\nabla\mathcal{J}(\cdot,g)$, pour $k = 1,\ldots,N$.

En déroulant cette itération, on peut alors considérer que $\Lambda_\Theta = \Gamma_{\theta_N}\circ\cdots\circ\Gamma_{\theta_1}$ est un réseau de neurones convolutifs représentant N itérations et que $\Theta = (\theta_1,\ldots,\theta_N)$ représente les paramètres de ce réseau.

3.3 Deep Unrolling Gradient Descent (DUGD)

Considérons le cas où l'on a la même opération à chaque itération, i.e. $\theta_1 = \cdots = \theta_N = \theta$, dans ce cas Λ_Θ peut être vu comme un *réseau de neurones récurrent*. Au lieu de faire le choix d'une régularisation, nous proposons d'apprendre directement l'itération, ce qui, pour $k = 1,\ldots,N$, correspond à :

$$(B_{k+1},\varphi_{k+1}) = \Gamma_\theta\left\{(B_k,\varphi_k),\nabla d\left[F_D(B_k,\varphi_k),I_D^{\mathrm{obs}}\right]\right\} \tag{9}$$

Dans ce cas, le terme de régularisation est implicitement appris pendant l'entraînement du réseau. L'architecture du réseau Γ_θ utilisée ici est inspirée de [7], elle est représentée dans la Fig. 1. Elle a été adaptée à notre problème, et la fonction d'activation ReLU a été remplacée par la fonction Leaky ReLU, définie par $\mathrm{LReLU}_\alpha(x) = \max(x,\alpha x)$, $\alpha > 0$. Elle est maintenue simple pour plusieurs raisons : tout d'abord nous cherchons à apprendre

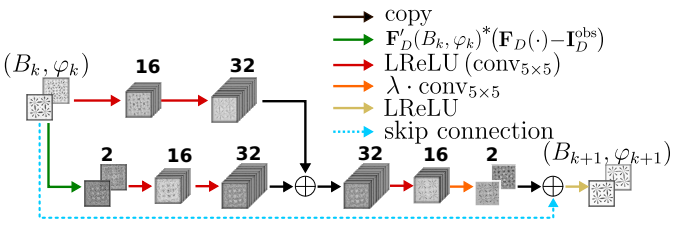

FIG. 1: Schéma d'un réseau de neurones convolutifs Γ_θ représentant une itération de la méthode DUGD.

une combinaison entre l'itération actuelle et l'information du gradient afin d'imiter une étape de descente et non à faire du post-traitement avec un large réseau, mais aussi pour minimiser la mémoire requise pour l'entraînement. Cette structure de réseau, similaire à un algorithme de descente de gradient, est appelé *Deep Unrolling Gradient Descent* (DUGD).

4 Expériences

4.1 Simulation des données

Afin de comparer les performances des différents algorithmes, nous avons généré des images synthétiques de contraste de phase. L'énergie des rayons X a été fixée à 13 keV pour une longueur d'onde de $\lambda = 0{,}095$ nm, et la taille des pixels à 24 nm. Nous avons créé une base de données de projections d'objets 3D créés à partir de combinaisons aléatoires d'une à dix formes d'ellipsoïdes et de paraboloïdes, constituées de trois matériaux différents (Or, Palladium et Zinc) pour créer des objets hétérogènes. Ensuite, des projections tomographiques 2D des parties réelle et imaginaire de l'indice de réfraction, correspondant respectivement à la phase et à l'absorption, ont été obtenues à partir des objets 3D pour une taille d'image de 2048×2048 pixels. Les objets et les projections ont été générés à l'aide du logiciel *TomoPhantom*. Les images de contraste de phase ont été générées à partir des images de projection selon (1) à une distance de propagation $D = 20{,}3$ mm et sous-échantillonnées à 512×512 pour éviter les effets d'*aliasing* dans le calcul des images d'intensités. Une base de données a été générée en utilisant différents niveaux aléatoires de bruit blanc gaussien pour obtenir un rapport signal/bruit crête à crête (PPSNR) compris entre 10 et 100 dB. A partir de celle-ci, 10 000 images ont été utilisées pour l'entraînement, 1 000 pour la validation pendant l'entraînement.Enfin, trois bases de données avec différents niveaux de bruits ont été générées pour l'évaluation : **(i)** PPSNR aléatoire compris entre 10 et 100 dB, comme la base d'entraînement et de validation, **(ii)** PPSNR égal à 50 dB et **(iii)** PPSNE égal à 100 dB.

4.2 Détails d'implémentation

Les conditions de convergence pour la descente du gradient n'ont pas été analysées en détail, nous avons pu constater que le choix d'un pas fixe suffisamment petit $\tau = 0.01$ et un nombre d'itérations égal à 1 000 étaient suffisants pour obtenir la convergence en pratique. Les paramètres de régularisation ont été choisis empiriquement, ce qui est sous-optimal. Nous avons choisi $\eta = 10^{-1}$, $\mu = 10^{-3}$ et le facteur de lissage ϵ a été fixé à 10^{-3} pour GD-TV$^\epsilon$.

Pour la méthode DUGD, le nombre d'itérations considéré est $N = 10$, ce qui veut dire que l'opérateur F_D et l'adjoint de sa dérivée de Fréchet $[F'(B,\varphi)]^*$ sont évalués 10 fois par le réseau. Le nombre de paramètres appris pour une itération Γ_θ, qui correspond au nombre total de paramètres appris par le réseau Λ_Θ est approximativement de 3 000. Nous sommes dans le cas d'un réseau profond mais avec peu de paramètres. Le réseau est entraîné sur 100 époques, avec une taille de batch de 10, un optimiseur d'Adam, un pas d'apprentissage de 10^{-3} avec une décroissance en cosinus et la fonction coût considérée est l'erreur quadratique. La valeur du paramètre de la fonction d'activation LReLU a été laissé par défaut à $\alpha = 0.3$. L'entraînement a duré environ 18 heures. Les résultats présentés ici sont obtenus en considérant des initialisations nulles $(B_0,\varphi_0) = (0,0)$.

4.3 Résultats

Les différentes méthodes ont été évaluées sur différentes bases de données contenant chacune 1 000 données tests, nous les avons comparées en utilisant l'erreur quadratique moyenne normalisée (NMSE) définie par $\text{NMSE}(x,x_{\text{true}}) = \frac{||x-x_{\text{true}}||_2}{||x_{\text{true}}||_2}$. Les résultats sont affichés dans le Tableau 2. On peut observer que le niveau de bruit n'a pas trop d'influence sur la qualité de reconstruction de la méthode DUGD, ceci montre que l'approche unrolling est robuste et qu'elle est peu dépendante du niveau de bruit sur lequel elle a été entraînée. Pour la méthode GD-TV, on voit qu'elle est robuste pour la récupération de la phase, mais la qualité de reconstruction de l'absorption se détériore lorsque le bruit est plus important. Globalement, la méthode DUGD a donné de meilleurs résultats que la méthode classsique GD-TV$^\epsilon$, obtenant en moyenne une meilleure reconstruction de l'absorption et de la phase, de plus le temps de calcul pour la reconstruction d'un cas est presque 40 fois plus rapide pour DUGD. Nous pouvons remarquer que pour l'approche unrolling, la qualité de la phase est meilleure que celle de l'absorption. L'évolution des moyennes NMSE, ainsi que l'écart-type, sont représentés dans la Fig. 4, nous pouvons voir que l'amélioration la plus significative pour unrolling est réalisée lors de la

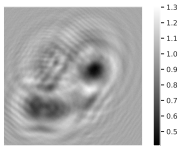

FIG. 2: Intensité simulée

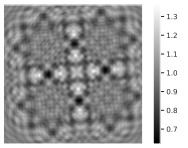

FIG. 3: Intensité expérimentale

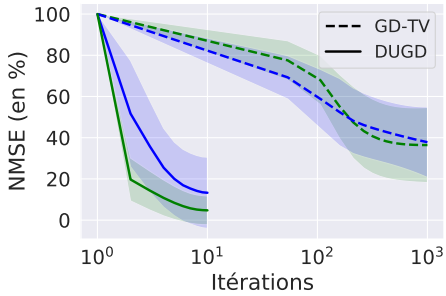

FIG. 4: Évolution des NMSE moyennes (en %) pour la base d'évaluation [10,100] dB. L'absorption est représentée en bleu et la phase en vert. Les zones transparentes représentent l'écart-type.

TAB. 1: Moyennes NMSE et écart-type (en %) pour les bases d'évaluation ainsi que le temps de calcul nécessaire pour un cas.

Bruit	[10,100] dB		50 dB		10 dB		Temps (en s)
Méthode	Absorption	Phase	Absorption	Phase	Absorption	Phase	
GD-TV$^\epsilon$	37.5 (17.4)	36.4 (18.2)	36.8 (17.9)	37.6 (17.8)	48.7 (25.5)	38.3 (17.6)	209
DUGD	**13.2 (17.2)**	**4.74 (6.99)**	**13.1 (16.5)**	**4.55 (6.69)**	**15.7 (19.5)**	**5.77 (7.25)**	**5**

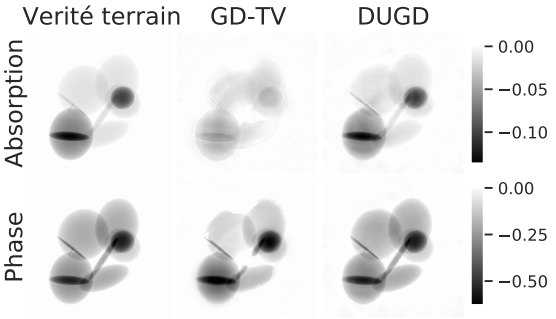

FIG. 5: Reconstructions obtenues sur donnée simulée.

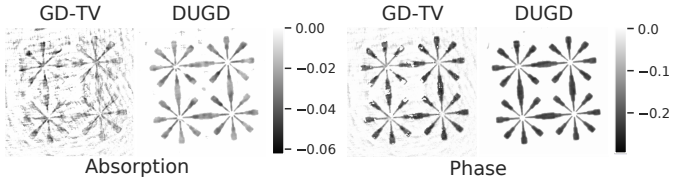

FIG. 6: Reconstructions obtenues sur donnée expérimentale.

première itération. Tandis que pour la descente de gradient, la reconstruction est améliorée le long des itérations.

Pour une évaluation qualitative, un exemple de reconstructions de phase et d'absorption de la mesure d'intensité (Fig. 2) sont affichées dans la Fig. 5 (avec contraste négatif). L'approche GD-TV$^\epsilon$ efface quelques morceaux du contour de l'objet, surtout pour l'absorption, ce qui peut être dû à l'approximation de la variation totale ainsi que la projection sur les valeurs positives. Au contraire, l'approche basée sur l'apprentissage effectue une reconstruction de très bonne qualité. Dans ce cas le terme de régularisation est implicitement appris à partir de la base d'exemples et permet de mieux reconstruire les bords, elle semble donc être mieux adaptée que la variation totale.

Enfin, l'approche proposée a été appliquée sur une donnée expérimentale acquise au synchrotron MAX IV (Lund, Suède). L'intensité détectée qui est affichée sur la Fig. 3 a été obtenue avec une taille de pixel effective de 12 nm et une distance de défocalisation $D = 20$ mm. L'énergie des rayons X est de 13 keV pour une longueur d'onde de $\lambda = 0.095$ nm. Les résultats obtenus (Fig. 6) montre que la descente de gradient est moins robuste que sa version déroulée. En effet, la méthode GD-TV laisse beaucoup de franges apparentes tandis que DUGD récupère l'objet avec très peu d'artefacts. Et bien que les données simulées ne contiennent pas explicitement ce genre de formes, l'approche unrolling a été capable de reconstruire l'objet et ne semble donc pas dépendant des formes utilisées dans la base d'entraînement.

5 Discussion et conclusion

Ce travail a présenté l'approche dite unrolling appliquée à une descente de gradient, ce qui a permis de tirer avantage de l'interprétabilité des algorithmes itératifs et de la puissance croissante des réseaux de neurones. Cette approche s'est avérée être efficace pour la résolution du problème inverse non linéaire qu'est la récupération de phase (et d'absorption). En déroulant une simple descente de gradient, nous avons pu notamment améliorer la qualité des reconstructions et raccourcir le temps nécessaire à de telles reconstructions. Cette approche peut être généralisée à des algorithmes itératifs plus complexes, telle que la méthode de Gauss-Newton, ou encore une méthode primal-dual. Une extension directe du travail proposé serait d'appliquer une telle méthode à la tomographie de contraste de phase. Mais comme toute méthode basée sur l'apprentissage, cette dernière est fortement dépendante de la base d'entraînement utilisée. Des recherches futures pourraient impliquer l'inclusion de paramètres physiques tels que l'énergie ou la distance de propagation dans de tels schémas.

Références

[1] M. Langer et F. Peyrin. *3D X-ray ultra-microscopy of bone tissue*. Osteoporosis International, 2016.

[2] M. Langer. *In-Line X-Ray Phase Tomography of Bone and Biomaterials for Regenerative Medicine In: Giuliani A., Cedola A. (eds) Advanced High-Resolution Tomography in Regenerative Medicine. Fundamental Biomedical Technologies*. Springer, Cham., 2018.

[3] S. Arridge, P. Maass, O. Öktem, et C-B. Schonlieb. *Solving inverse problems using data-driven models*. Acta Numerica, 2019.

[4] K. Mom, B. Sixou et M. Langer. *Mixed scale dense convolutional networks for x-ray phase-contrast imaging*. Applied Optics, 2022.

[5] Y. Zhang, M. Andreas Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, et P. Villanueva-Perez *PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets*. Optics Express, 2021.

[6] V. Monga, Y. Li et Y.C. Eldar. *Algorithm Unrolling : Interpretable, efficiend deep Learning for signal and image processing*. IEEE Signal Processing Magazine, 2019.

[7] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin et S. Arridge. *Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography*. IEEE Transactions on Medical Imaging, 2018.

[8] V. Davidoiu, B. Sixou, M. Langer et F. Peyrin. *Non-linear iterative phase retrieval based on Frechet derivative*. Opt. Express, 2011.

TAB. 2: Moyennes NMSE et écart-type (en %) pour les bases d'évaluation ainsi que le temps de calcul nécessaire pour un cas.

Méthode	GD-TV$^{\epsilon}$		DUGD	
Bruit	Absorption	Phase	Absorption	Phase
[10,100] dB	37.5 (17.4)	36.4 (18.2)	**13.2 (17.2)**	**4.74 (6.99)**
50 dB	36.8 (17.9)	37.6 (17.8)	**13.1 (16.5)**	**4.55 (6.69)**
100 dB	48.7 (25.5)	38.3 (17.6)	**15.7 (19.5)**	**5.77 (7.25)**