

Rapport de Projet : Reconstruction de Surfaces 3D par Apprentissage Non-Supervisé (Noise-to-Noise Mapping)

Sacha K.

Mars 2026

Table des matières

1	Introduction et Contexte	2
2	Synthèse de l'article : Le paradigme Noise-to-Noise Mapping	2
2.1	Représentation par Fonction de Distance Signée (SDF)	2
2.2	Mécanisme de Neural Pull	2
2.3	Fonction de Perte : Transport Optimal et EMD	3
2.4	Consistance Géométrique et Régularisation	3
3	Mise en œuvre technique et Environnement	3
3.1	Infrastructure de Calcul et Matériel	4
4	Protocole Expérimental et Tests	4
4.1	Préparation des Données	4
4.2	Stratégie d'Entraînement (Deep Training)	4
5	Analyse de la Convergence et des Performances	5
5.1	Évolution des Fonctions de Perte	5
5.2	Interprétation des Résultats	5
6	Résultats et Analyse Qualitative	5
6.1	Formalisation de l'extraction par isosurface	5
6.2	Visualisation par seuils d'extraction	6
6.3	Interprétation des morphologies obtenues	6
6.4	Analyse de l'échec de la reconstruction topologique	7
7	Amélioration par Encodage Positionnel	7
7.1	Le problème du biais spectral des MLP	7
7.2	Mise en place de l'encodage harmonique	8
7.3	Effets attendus sur la SDF	8
8	Résultats et Analyse Critique	8
9	Conclusion	9

1 Introduction et Contexte

La numérisation 3D via des capteurs LiDAR produit intrinsèquement des nuages de points corrompus par du bruit stochastique. Le défi majeur réside dans la transformation de ces données discrètes en une surface continue lisse (Signed Distance Function - SDF).

Traditionnellement, le débruitage repose sur des méthodes supervisées nécessitant des données "Ground Truth" (propres). Ce projet explore la méthodologie **Noise-to-Noise (N2N) Mapping**, qui permet d'apprendre une SDF cohérente en utilisant uniquement des observations bruitées, sans jamais voir la surface réelle.

2 Synthèse de l'article : Le paradigme Noise-to-Noise Mapping

L'article étudié propose une approche novatrice pour la reconstruction de surfaces 3D à partir de nuages de points non orientés et fortement bruités. Contrairement aux méthodes supervisées classiques, ce modèle repose sur le principe du *Noise-to-Noise mapping*, permettant d'apprendre une structure géométrique cohérente sans jamais avoir accès à une donnée de référence propre (*Ground Truth*). Ainsi la méthode mise en avant permet à la fois de déterminer la Signed Distance Function (SDF) mais aussi dans ce même apprentissage récupérer une fonction de débruitage. En deux parties, on restreint le processus de manière à obtenir dans un premier temps une approximation de la Ground Truth, et on rajoutera une composante de régularité géométrique pour rester fidèle à la réalité des formes 3D présentes.

2.1 Représentation par Fonction de Distance Signée (SDF)

Le choix de la représentation géométrique porte sur une fonction de distance signée implicite. Le réseau de neurones, noté f_θ , apprend à approximer une fonction qui, pour tout point $q \in \mathbb{R}^3$, retourne sa distance signée à la surface la plus proche. La surface reconstruite $\mathcal{S}$ est alors définie mathématiquement par l'ensemble de niveau zéro de cette fonction :

$$\mathcal{S} = \{q \in \mathbb{R}^3 \mid f_\theta(q) = 0\} \quad (1)$$

Cette approche offre l'avantage d'une résolution théoriquement infinie, contrairement aux représentations discrètes de type voxels ou maillages fixes.

2.2 Mécanisme de Neural Pull

Pour contraindre le réseau à apprendre la topologie de l'objet, l'algorithme utilise un mécanisme de projection itératif appelé *Neural Pull*. Pour chaque point bruité p_i du nuage d'entrée P , le gradient de la SDF est utilisé pour projeter le point vers la surface estimée par le modèle. Le point projeté p'_i est calculé selon la relation :

$$p'_i = p_i - f_\theta(p_i) \frac{\nabla f_\theta(p_i)}{\|\nabla f_\theta(p_i)\|_2} \quad (2)$$

Ce processus déplace les points dans la direction de la plus courte distance vers la surface apprise, transformant progressivement un nuage de points diffus en une coque géométrique nette.

2.3 Fonction de Perte : Transport Optimal et EMD

Le caractère non-supervisé du modèle repose sur l'utilisation de la distance de Wasserstein, ou *Earth Mover's Distance* (EMD), comme fonction de perte. Au lieu de minimiser la distance vers une surface parfaite, le modèle minimise le coût de transport entre le nuage projeté P' et un second nuage bruité Q issu de la même source. La perte EMD est définie par :

$$\mathcal{L}_{EMD}(P', Q) = \min_{\phi: P' \rightarrow Q} \sum_{p' \in P'} \|p' - \phi(p')\|_2 \quad (3)$$

où ϕ représente une bijection (couplage parfait) entre les deux ensembles de points. La bijectivité de la fonction est cruciale car elle permet d'assurer une correspondance point par point unique et donc de ne pas obtenir un résultat bruité à la fin de la convergence, le but de l'article étant ce débruitage.

2.4 Consistance Géométrique et Régularisation

Afin de garantir que la fonction f_θ apprise par le réseau est une véritable fonction de distance signée (SDF), une contrainte de cohérence géométrique est introduite. Sans cette régularisation, le réseau pourrait converger vers des solutions dégénérées où le champ de distance ne respecte pas les propriétés euclidiennes élémentaires.

La perte de consistance géométrique, notée $\mathcal{L}_{gc}$, repose principalement sur l'équation *Eikonale*. Elle impose que la norme du gradient de la fonction soit unitaire en tout point de l'espace :

$$\mathcal{L}_{gc} = \mathbb{E}_q [(\|\nabla f_\theta(q)\|_2 - 1)^2] \quad (4)$$

Cette contrainte garantit la linéarité du champ de distance et empêche l'apparition de discontinuités artificielles dans la reconstruction. En combinant cette perte avec une contrainte de valeur nulle à la surface ($\mathcal{L}_{zero}$), le modèle assure une transition fluide entre l'intérieur et l'extérieur de l'objet :

$$\mathcal{L}_{total} = \lambda_{emd} \mathcal{L}_{EMD} + \lambda_{gc} \mathcal{L}_{gc} + \lambda_z \mathcal{L}_{zero} \quad (5)$$

où les coefficients λ pondèrent l'importance de chaque terme durant l'optimisation. Cette synergie permet d'obtenir une surface finale qui est à la fois fidèle au nuage de points bruité et mathématiquement rigoureuse.

3 Mise en œuvre technique et Environnement

Le passage de la théorie à la pratique a nécessité la mise en place d'une infrastructure de calcul intensif capable de gérer les opérations tensorielles lourdes liées à la distance EMD et à l'extraction de surfaces haute résolution.

3.1 Infrastructure de Calcul et Matériel

Les expérimentations ont été menées sur la plateforme Google Cloud Platform (GCP). Pour répondre aux besoins en calcul parallèle du modèle, une instance Compute Engine a été configurée avec les spécifications suivantes :

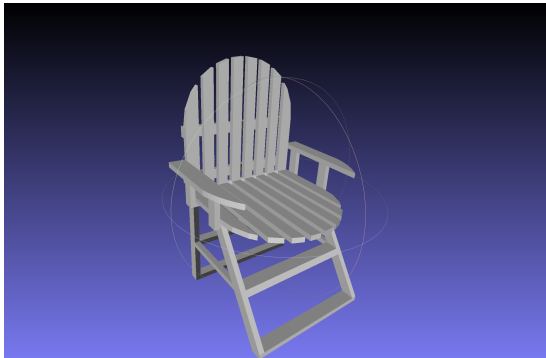
- GPU : NVIDIA Tesla T4 (16 Go de VRAM GDDR6).
- CPU : Intel(R) Xeon(R) @ 2.20GHz (n1-standard-4).
- Stockage : Disque persistant SSD de 50 Go.

L'utilisation de la technologie GPU est ici indispensable pour le calcul de la perte EMD, qui repose sur des noyaux CUDA personnalisés (approxmatch) permettant une accélération massive par rapport à une implémentation CPU standard.

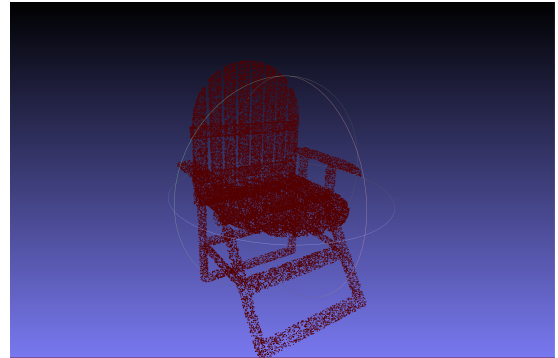
4 Protocole Expérimental et Tests

4.1 Préparation des Données

La donnée source est un modèle 3D de type Adirondack Chair au format .off. (voir Figure 1) Le script de prétraitement `gen_noise.py` a été utilisé pour générer un nuage de points bruité.



(a) Chaise originale



(b) Chaise bruitée ($\sigma = 0.01$)

FIGURE 1 – Ground Truth et Chaise bruitée générée via `gen_noise.py`.

1. Échantillonnage : 50 000 points sont extraits de la surface du maillage original.
2. Injection de bruit : Un bruit gaussien stochastique d'écart-type $\sigma = 0.01$ est appliqué, simulant les imprécisions d'un capteur LiDAR.

4.2 Stratégie d'Entraînement (Deep Training)

Afin de capturer les détails fins de la chaise, une stratégie d'entraînement intensif a été adoptée :

- Volume d'itérations : Le modèle a été poussé à 50 000 itérations (bien que la convergence ait été observée dès 12 000 itérations).
- Résolution de discrétisation : Lors de la phase d'inférence, la résolution de la grille de Marching Cubes a été portée de 128^3 à 256^3 . Cette augmentation d'un facteur 8 de la densité volumique vise à isoler les structures fines (lattes, accoudoirs) souvent fusionnées à basse résolution.

5 Analyse de la Convergence et des Performances

5.1 Évolution des Fonctions de Perte

Le suivi des métriques durant l’entraînement permet d’évaluer la stabilité de l’apprentissage. Le tableau 1 synthétise l’évolution de la perte totale, de la perte de transport optimal (EMD) et de la perte de surface nulle (Zero Loss).

Itération	Loss Totale	Loss EMD ($\mathcal{L}_{emd}$)	Loss Zero ($\mathcal{L}_z$)
0	0,4398	0,2983	0,3430
2 000	0,0685	0,0677	0,0037
4 000	0,0655	0,0653	0,0073
6 000	0,0669	0,0667	0,0075
8 000	0,0734	0,0733	0,0135
10 000	0,0682	0,0683	0,0053
12 000	0,0646	0,0646	0,0046

TABLE 1 – Évolution des métriques d’entraînement du modèle Noise-to-Noise Mapping.

5.2 Interprétation des Résultats

L’analyse des données montre une convergence extrêmement rapide dès le premier quart de l’entraînement. La chute brutale de la perte totale entre l’itération 0 et 2 000 indique que le réseau MLP a réussi à isoler le signal déterministe du bruit gaussien en un temps réduit.

La stagnation observée entre les itérations 2 000 et 12 000, où la perte totale se stabilise autour de $\mathcal{L} \approx 0,065$, témoigne d’un régime d’équilibre. À ce stade, le réseau a figé la topologie globale de l’objet. Les légères fluctuations de la perte EMD sont caractéristiques du processus de transport optimal qui tente d’ajuster la bijection entre les points projetés et le nuage bruité Q sans jamais pouvoir annuler totalement l’erreur du fait de la nature stochastique du bruit.

6 Résultats et Analyse Qualitative

Cette section présente les maillages extraits après la phase d’entraînement intensif. L’extraction a été réalisée via l’algorithme des Marching Cubes en faisant varier le seuil d’iso-surface (c) afin d’explorer le champ de distance signée (SDF) appris par le réseau.

6.1 Formalisation de l’extraction par isosurface

L’extraction du maillage à partir de la fonction de distance signée apprise f_θ repose sur la définition d’un seuil d’isosurface c . Mathématiquement, la surface extraite $\mathcal{S}_c$ correspond à l’ensemble des points de l’espace vérifiant l’équation :

$$\mathcal{S}_c = \{x \in \mathbb{R}^3 \mid f_\theta(x) = c\} \quad (6)$$

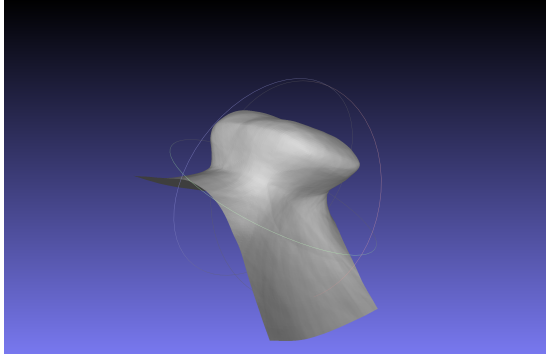
Le choix de ce paramètre c agit comme une opération morphologique sur l’objet reconstruit :

- Si $c = 0$: on extrait la frontière théorique apprise par le modèle.
- Si $c < 0$: on extrait une surface située à l'intérieur de l'objet (érosion), ce qui permet de visualiser le noyau structural de la reconstruction.
- Si $c > 0$: on extrait une enveloppe dilatée (offset positif), capturant le voisinage extérieur de la surface.

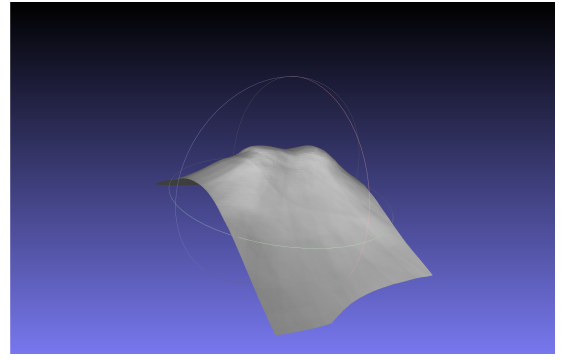
Dans nos tests, la variation de c de $-0,1$ à $0,01$ a permis de mettre en évidence l'épaisseur du champ de distance et le manque de raideur du gradient aux abords de la surface réelle.

6.2 Visualisation par seuils d'extraction

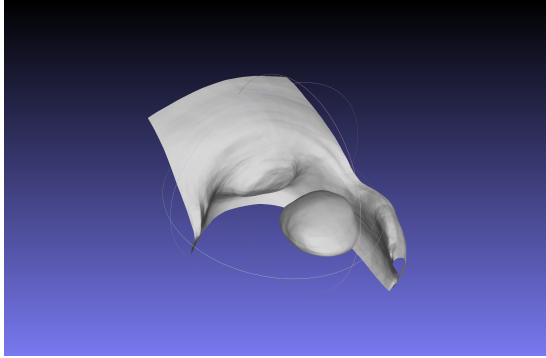
La figure 2 présente les différents états de la reconstruction selon la valeur du seuil c . Ces résultats mettent en évidence la structure interne et externe de la fonction implicite générée.



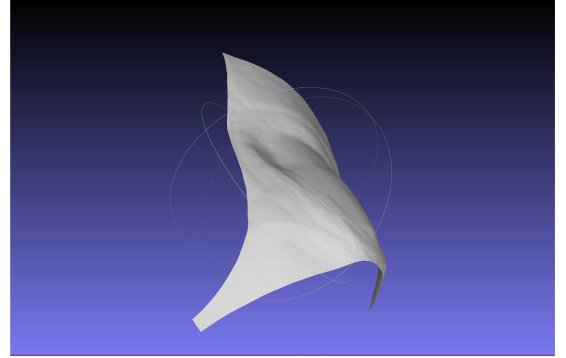
(a) Seuil $c = -0,1$: Noyau volumique



(b) Seuil $c = -0,0025$: Interface proche



(c) Seuil $c = 0,002$: Enveloppe externe fine



(d) Seuil $c = 0,01$: Enveloppe dilatée

FIGURE 2 – Évolution de la morphologie extraite en fonction du seuil d'iso-surface sur une grille de résolution 256^3 .

6.3 Interprétation des morphologies obtenues

L'analyse des différents seuils permet de déduire les propriétés géométriques du champ de distance appris :

- Régime interne ($c = -0,1$) : Le modèle présente une structure massive et continue. À cette profondeur, le réseau a fusionné l'intégralité des composants de la chaise

(lattes, dossier, pieds) en un seul volume convexe. La topologie originale est perdue au profit d'une "souche" géométrique.

- Régime critique ($c = -0,0025$) : À proximité de la surface théorique, le maillage se fragmente et prend l'aspect d'une membrane ouverte et ondulée. Le réseau ne parvient pas à définir un volume clos et cohérent, signe que le gradient de la SDF est trop faible ou instable dans cette zone.
- Régime externe ($c > 0$) : Les surfaces extraites forment des enveloppes lisses qui flottent autour de la forme. Elles représentent la moyenne statistique du bruit capturée par le MLP, agissant comme un filtre passe-bas qui occulte les détails sémantiques.

6.4 Analyse de l'échec de la reconstruction topologique

Plusieurs facteurs techniques expliquent pourquoi la chaise Adirondack n'a pas pu être reconstruite fidèlement malgré la convergence des fonctions de perte :

1. Biais spectral des réseaux MLP : Les perceptrons multicouches ont une tendance naturelle à apprendre des fonctions de basse fréquence (lisses). Pour un objet complexe comme cette chaise, le réseau privilégie le lissage des lattes plutôt que la représentation des discontinuités et des vides.
2. Complexité de la topologie à genre élevé : La chaise Adirondack possède une topologie complexe avec de nombreux trous. Le modèle Noise-to-Noise, sans encodeur de caractéristiques locales (type Hash Grids), peine à représenter des structures disjointes et fines, tendant à "remplir" les vides pour minimiser la distance de transport globale.
3. Incohérence du gradient sur les structures fines : Les lattes étant très minces, les points de requête du gradient lors du Neural Pull peuvent s'entremêler, provoquant une confusion pour le réseau qui finit par fusionner les plans opposés en une seule masse informe.

7 Amélioration par Encodage Positionnel

Malgré une convergence satisfaisante des fonctions de perte, les premiers maillages obtenus présentaient un lissage excessif, allant jusqu'à fusionner les structures fines de la chaise Adirondack. Pour résoudre ce problème, j'ai intégré une couche d'encodage positionnel (*Positional Encoding*) afin d'aider le réseau à capturer les hautes fréquences géométriques.

7.1 Le problème du biais spectral des MLP

Les réseaux de neurones de type Perceptron Multicouche (MLP) présentent naturellement un biais spectral (*spectral bias*). Cela signifie qu'ils apprennent prioritairement les fonctions de basse fréquence, ce qui se traduit par des surfaces reconstruites très lisses. Dans notre expérimentation, les coordonnées spatiales (x, y, z) brutes ne varient pas assez pour que le réseau puisse isoler des détails comme les lattes du dossier ou les espaces vides qui les séparent.

7.2 Mise en place de l’encodage harmonique

Nous avons donc choisi de projeter les coordonnées d’entrée dans un espace de plus haute dimension en utilisant des fonctions périodiques. Pour chaque coordonnée $p \in \{x, y, z\}$, on applique la fonction de projection γ suivante :

$$\gamma(p) = (\sin(2^0\pi p), \cos(2^0\pi p), \dots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p)) \quad (7)$$

Pour cette implémentation, nous avons fixé $L = 6$. Le vecteur d’entrée passe ainsi d’une dimension 3 à une dimension 39 (incluant les coordonnées d’origine). Cette méthode permet de présenter au réseau des signaux dont les variations sont beaucoup plus rapides.

7.3 Effets attendus sur la SDF

Cette transformation rend le gradient de la fonction de distance signée (∇f_θ) nettement plus sensible aux détails locaux. En forçant le réseau à traiter ces hautes fréquences, les objectifs sont les suivants :

- Obtenir une séparation nette entre les lattes du dossier.
- Préciser la géométrie des arêtes vives de la chaise.
- Supprimer l’effet de "fusion" observé entre les différentes parties de l’objet.

Cet ajout permet d’adapter le pipeline *Noise-to-Noise* original en utilisant des concepts issus des représentations implicites modernes (*Neural Implicit Representations*).

8 Résultats et Analyse Critique

Malgré l’implémentation de l’encodage positionnel, qui devait permettre d’affiner la SDF, l’analyse visuelle du nuage de points généré (voir Figure 3) révèle un échec technique important. Contrairement aux objectifs recherchés, la reconstruction n’aboutit pas à des surfaces distinctes et propres pour les lattes de la chaise Adirondack.

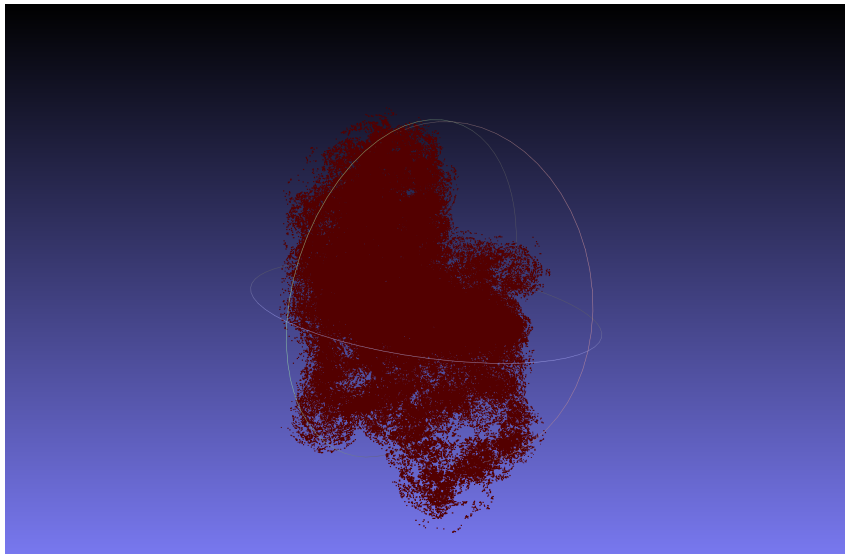


FIGURE 3 – Nuage de points généré après l’ajout de l’encodage positionnel ($L = 6$). Le résultat montre une confusion majeure : les points sont agrégés en une masse volumétrique dense et diffuse, ne permettant pas de distinguer les surfaces individuelles (lattes) de la chaise.

Cette confusion s'explique par plusieurs facteurs techniques qui nuisent à l'efficacité de la méthode Noise-to-Noise dans cette configuration complexe :

1. **Amplification Spectrale du Bruit** : L'encodage positionnel force le gradient (∇f_θ) à devenir très réactif aux hautes fréquences spectrales pour capturer les détails. Cependant, si le bruit gaussien initial ($\sigma = 0.01$) possède également des composantes haute fréquence significatives, l'encodage positionnel risque de les amplifier au lieu de les filtrer. Au lieu de converger vers la surface de chaque latte, le réseau tend alors à modéliser le volume bruité lui-même, créant cette masse dense et "confuse".
2. **Confusion Topologique** : L'Adirondack Chair est un objet à "genre élevé" (présentant de nombreux trous). Dans un nuage de points bruité, deux lattes proches se chevauchent sémantiquement. Le réseau, sans supervision explicite, n'arrive pas à décider si le vide entre les lattes est une surface ou un trou. Il finit par "remplir" ce vide en créant un champ de distance volumétrique au lieu d'une SDF mince et fine.
3. **Échec du Neural Pull pour les Surfaces Minces** : Le mécanisme de projection itérative (*Neural Pull*) suppose qu'un point peut être projeté de manière unique vers la surface la plus proche. Si le champ SDF est aussi diffus que l'image le suggère, les gradients pointent dans des directions multiples et instables, empêchant les points de s'aligner sur des plans nets.

9 Conclusion

Ce projet de reconstruction de surfaces 3D, basé sur l'approche Fast Learning of SDF via Noise-to-Noise Mapping, a permis de mettre en pratique des concepts avancés de géométrie algorithmique et d'apprentissage profond. L'implémentation, réalisée sur une infrastructure GPU NVIDIA Tesla T4 via Google Cloud Platform, a souligné l'efficacité de la méthode pour converger rapidement vers une structure globale, mais a également révélé ses limites sur des géométries à topologie complexe.

L'enjeu principal résidait dans la capacité du réseau MLP à distinguer les structures fines de la chaise Adirondack. L'ajout d'une couche d'encodage positionnel ($L = 6$) visait à briser le biais spectral du réseau pour capturer les hautes fréquences géométriques. Cependant, les résultats obtenus montrent que cette technique est à double tranchant : en augmentant la sensibilité du réseau aux détails, elle a également amplifié les composantes haute-fréquence du bruit. Le résultat final, un nuage de points volumétrique et diffus, témoigne de la difficulté pour le modèle de définir une surface mince et cohérente sans supervision explicite ou régularisation plus stricte.

En définitive, ce travail démontre que si le *Noise-to-Noise Mapping* est une solution puissante pour s'affranchir de données de référence (*Ground Truth*), sa réussite sur des objets disjoints et fins nécessite un équilibre complexe entre capacité spectrale et filtrage du bruit. Ce projet constitue une base solide pour explorer, à l'avenir, des méthodes de représentation locale plus robustes comme les *Neural Hash Grids* ou des fonctions de perte intégrant des contraintes de courbure plus poussées.

Code source et implémentation disponibles sur GitHub :

<https://github.com/sachkho/Noise2Noise-Surface-Reconstruction/>