

# fiche de lecture: NOISE2VOID

nael klein

December 2025

## 1 Informations générales

### 1.1 Titre

Noise2Void: Learning Denoising from Single Noisy Images

### 1.2 Auteurs

Alexander Krull, Tim-Oliver Buchholz, Florian Jug

### 1.3 Année

2019 — IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

### 1.4 Source / DOI / PDF

PDF : Krull\_Noise2Void\_CVPR2019.pdf

DOI : 10.1109/CVPR.2019.00258

## 2 Résumé de l'article

Noise2Void (N2V) est une méthode d'entraînement auto-supervisée qui permet d'entraîner un réseau de débruitage à partir d'images *uniquement bruitées*, sans clean targets ni paires d'images bruit-bruit. Cette approche supprime la principale contrainte des méthodes supervisées (nécessité d'un ground truth propre) et même de Noise2Noise (besoin de deux acquisitions indépendantes d'une même scène).

L'idée centrale repose sur un réseau *blind-spot*, c'est-à-dire un CNN dont le réceptif exclut le pixel à prédire. Le réseau ne peut donc pas apprendre l'identité de l'image et doit estimer la valeur du pixel uniquement à partir de son contexte, en exploitant les corrélations du signal. Le bruit étant supposé pixel-wise indépendant (donné le signal), le réseau apprend alors à reconstruire la composante structurée de l'image.

Les auteurs démontrent que N2V obtient des résultats comparables à Noise2Noise, légèrement inférieurs aux méthodes supervisées, mais supérieurs aux méthodes classiques non apprises comme BM3D ou Non-Local Means pour plusieurs jeux de données, notamment les images microscopiques.

### 3 Objectifs

- Proposer une méthode de débruitage totalement auto-supervisée (aucun ground truth, aucune paire d'images requise).
- Permettre l'entraînement directement sur les données à traiter, ce qui est crucial en microscopie biologique où les images propres sont généralement inaccessibles.
- Démontrer la faisabilité du débruitage par apprentissage en supprimant les dépendances au centre du patch (architecture blind-spot).
- Comparer la performance à Noise2Noise, aux méthodes supervisées et aux approches classiques fondées sur les statistiques internes (ex: BM3D).

### 4 Méthodologie

#### 4.1 1. Modèle de formation d'image

Le modèle suppose que :

- le signal  $s$  présente des corrélations spatiales (*voir Eq. (2), p. 2*),
- le bruit  $n$  est conditionnellement indépendant pixel par pixel donné  $s$  (*Eq. (3)*),
- le bruit est centré ( $E[n_i] = 0$ ), ce qui garantit  $E[x] = s$ .

Ces hypothèses sont essentielles pour que l'apprentissage auto-supervisé fonctionne.

#### 4.2 2. Architecture blind-spot

La Figure 2 (p. 3) montre clairement la différence :

- un réseau standard utiliserait le pixel central → risque d'apprendre l'identité,
- un réseau blind-spot exclut ce pixel → le réseau doit prédire la valeur à partir du voisinage.

### 4.3 3. Masquage des pixels (implémentation pratique)

Plutôt que de modifier l'architecture, N2V procède à un masquage sélectif :

- des patches de  $64 \times 64$  pixels sont extraits,
- un ensemble de  $N = 64$  pixels est choisis dans chaque patch,
- pour chacun, la valeur réelle est remplacée par une valeur voisine (Figure 3, p. 5),
- la valeur d'origine sert de cible.

Le réseau apprend ainsi à prédire le pixel masqué à partir de son contexte uniquement.

### 4.4 4. Entraînement du réseau

Les auteurs utilisent un U-Net avec batch normalization :

- profondeur variable selon les expériences (2 niveaux pour BSD68),
- loss MSE,
- traitement uniquement des pixels masqués.

### 4.5 5. Expériences

Trois catégories de données :

- **BSD68 naturel** (Figure 4, p. 6) : comparaison avec N2N, supervision, BM3D.
- **Microscopie simulée** : ground truth synthétique → PSNR comparable à N2N.
- **Microscopie réelle** : cryo-TEM, CTC-MSC, CTC-N2DH (Figure 4, p. 6–7), où seul N2V et BM3D sont applicables.

## 5 Résultats principaux

- Sur BSD68 : PSNR N2V 27.71 dB, inférieur à supervision (29.06 dB) et N2N (28.86 dB), mais proche des méthodes apprises.
- Sur données simulées : N2V performances presque identiques à N2N et à la supervision.
- Sur cryo-TEM (p. 7) : N2V préserve les structures fines sensibles (ex : protofilaments tubulaires à 4 nm).

- Sur CTC-MSC et CTC-N2DH : N2V offre un rendu plus propre et plus homogène que BM3D, avec un temps de calcul plus faible.
- Robustesse pratique : fonctionne même quand aucun autre schéma d'apprentissage n'est possible.

## 6 Forces

- Ne nécessite aucune forme de ground truth : idéal en microscopie.
- Entraînement réalisable directement sur les données du projet.
- Implémentation simple basée sur un masque aléatoire.
- Résultats compétitifs par rapport aux méthodes supervisées.
- Applicable à différents types d'imagerie (naturelle, fluorescence, cryo-TEM).

## 7 Faiblesses / limites

- Ne peut pas traiter les bruits structurés (Figure 6, p. 8) → viol de l'indépendance du bruit.
- Perte des détails très irréguliers et pixels isolés imprévisibles (Figure 5, p. 7).
- Performances inférieures en moyenne au modèle supervisé lorsque du ground truth est disponible.
- La prédiction repose entièrement sur la corrélation spatiale → inefficace si la structure est totalement imprévisible.

## 8 Pertinence pour le challenge Fuse My Cells

Cet article est directement pertinent pour mon travail dans le cadre du challenge *Fuse My Cells*, car il aborde des problématiques qui correspondent exactement aux contraintes de notre jeu de données :

- **Absence totale de ground truth** : N2V propose une solution complète pour entraîner un modèle de restauration à partir des seules images disponibles.
- **Données très bruitées et anisotropes** : N2V apprend à extraire la composante structurée même lorsque le signal est fortement noyé dans le bruit.
- **Besoin de préserver les structures fines** (noyaux, membranes) : les résultats sur cryo-TEM (p. 7) montrent que N2V préserve des détails subcellulaires de très petite taille.

- **Amélioration des métriques N-SSIM et N-IOU** : un meilleur débruitage augmente mécaniquement la netteté des structures segmentables.
- **Méthode compatible avec un pipeline moderne** : N2V peut servir de prétraitement indispensable avant une architecture patch-based ou 3D.

Ainsi, N2V constitue une alternative pertinente lorsque la génération de données d'entraînement propres ou doubles est impossible, ce qui est précisément le cas pour Fuse My Cells.

## 9 Notions importantes

- Réseaux blind-spot
- Masquage aléatoire de pixels
- Hypothèse de bruit indépendant pixel par pixel
- Apprentissage auto-supervisé
- Comparaison supervision / N2N / N2V

## 10 Référence BibTeX

```
@inproceedings{Krull2019Noise2Void,
  title={Noise2Void - Learning Denoising from Single Noisy Images},
  author={Krull, Alexander and Buchholz, Tim-Oliver and Jug, Florian},
  booktitle={IEEE/CVF Conference on Computer Vision and Pattern
    Recognition (CVPR)},
  year={2019},
  doi={10.1109/CVPR.2019.00258}
}
```