QWD-GAN: Quality-aware Wavelet-driven GAN for Unsupervised Medical Microscopy Images Denoising

Qijun Yang*,Yating Huang*,Lintao Xiang, Hujun Yin†
Department of Electrical Electronic Engineering, University of Manchester, UK

Introduction & Motivation

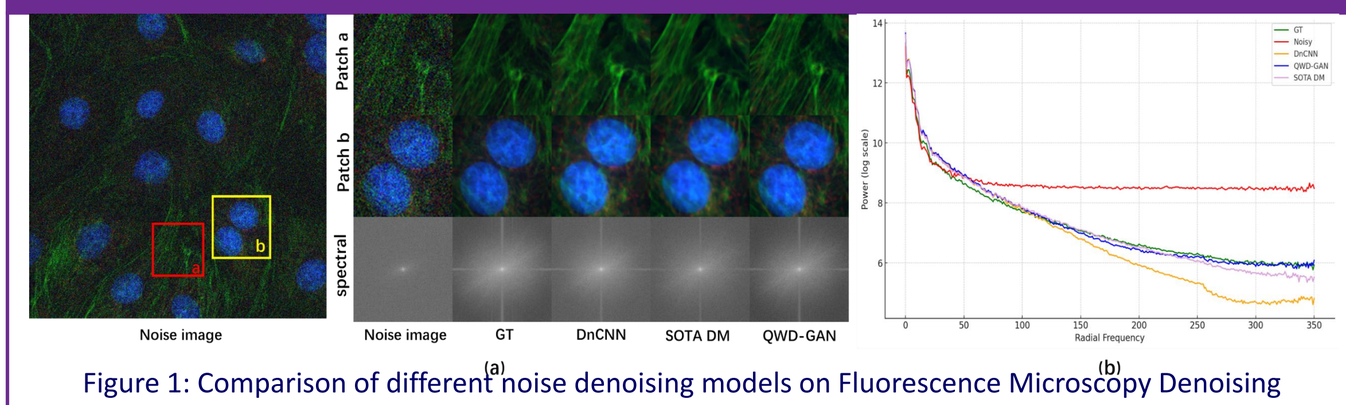


Figure 1: Comparison of different noise denoising models on Fluorescence Microscopy Denoising (FMD) dataset . Comparison of denoising of QWD-GAN in spatial domain and frequency domain (a), Radial frequency and power distribution curve (b).

➢ Challenges

- **Biomedical microscopy noise problem**
- **Limitations of supervised and diffusion models**
- **Research gap and motivation**

➢ Key questions

- **How to retain fine details?**
- **How to combine spatial & frequency features?**
- **How to ensure perceptual reliability?**

➢ Goal

We propose an unsupervised, detail-preserving, and quality-aware GAN framework for microscopy image denoising — QWD-GAN.

Proposed Method: QWD-GAN

➢ Wavelet-Guided Multi-Scale Denoising Generator (WG-MSDG)

The generator integrates both spatial and frequency cues:

- A wavelet-convolutional group (WCG) decomposes features into multi-scale subbands via Haar wavelets.
- Each subband is convolved with different kernel sizes to balance global and local representation.
- A Frequency–Spatial Feature Fusion (FS-FF) module adaptively merges both domains using a discrepancy-aware attention mechanism.

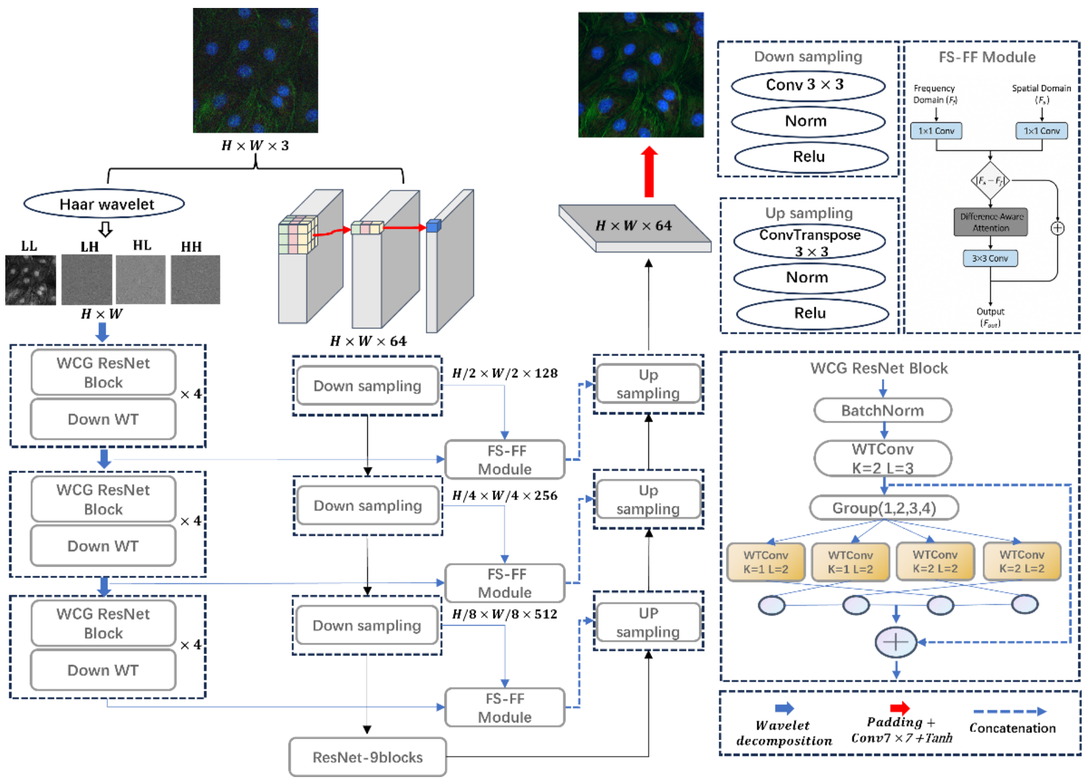


Figure 2: Structure of Wavelet-Guided Multi-Scale Denoising Generator

➢ Key modules

New Metric – High-Frequency Retention Ratio (HFRR)

1. Apply the 2D discrete Fourier transform to both the denoised image I_d and the clean image I_c, and shift the zero-frequency component to the center:

$$F_d = \text{fftshift}(\text{fft2}(I_d)), \quad F_c = \text{fftshift}(\text{fft2}(I_c)) \tag{12}$$

2. Compute the magnitude spectra:

$$M_d = |F_d| = \sqrt{\Re(F_d)^2 + \Im(F_d)^2}, \quad M_c = |F_c| = \sqrt{\Re(F_c)^2 + \Im(F_c)^2} \tag{13}$$

3. Define a binary high-frequency mask H that excludes low-frequency components around the center. The mask is constructed based on a radial threshold r:

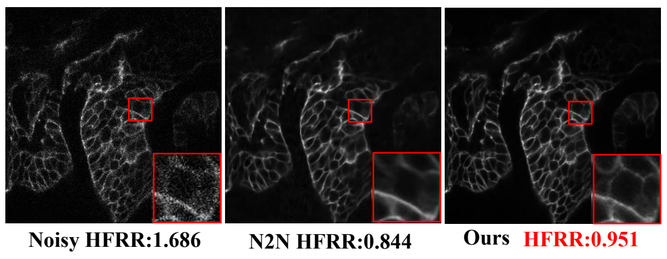
$$H(i,j) = \begin{cases} 1, & \text{if } \sqrt{(i-h/2)^2 + (j-w/2)^2} > r \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

4. Extract the high-frequency components:

$$M_d^{\text{high}} = M_d \cdot H, \quad M_c^{\text{high}} = M_c \cdot H \tag{15}$$

5. Finally, compute the HFRR:

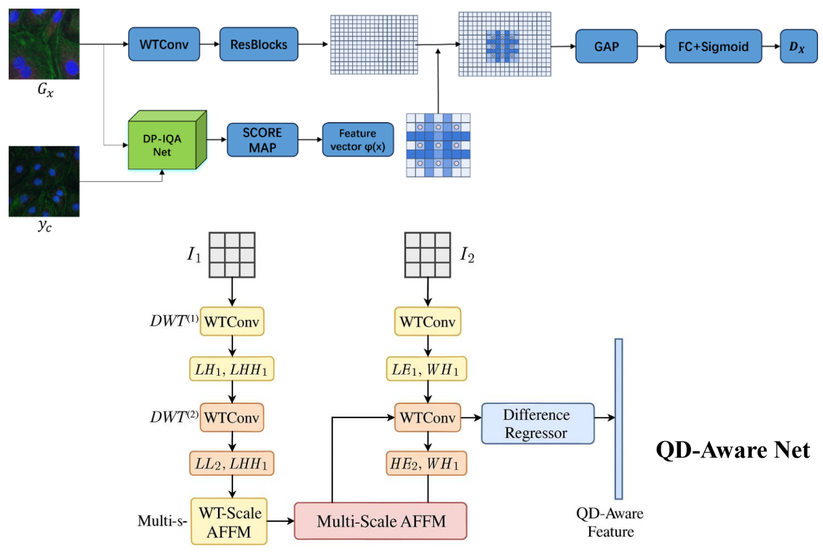
$$\text{HFRR} = \frac{\|M_d^{\text{high}}\|_2}{\|M_c^{\text{high}}\|_2} \tag{16}$$



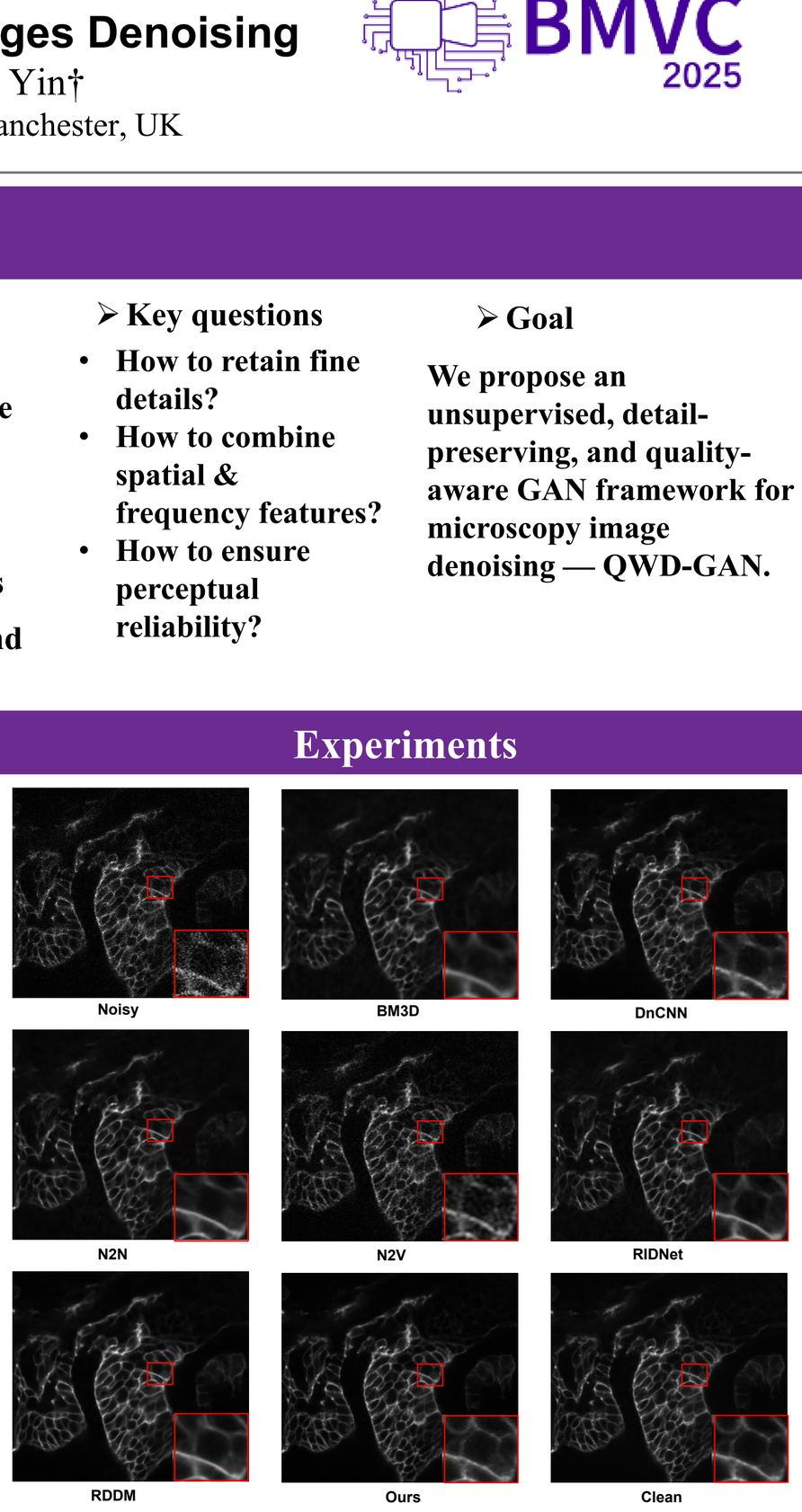
Quality-Aware Dual-Branch Discriminator

The discriminator introduces perceptual guidance through the Difference-Perception IQA (DP-IQA) network:

- A main path judging spatial realism
- An auxiliary IQA path predicting perceptual difference and structural fidelity.
- The fused features enforce quality-aware adversarial learning, aligning with human visual perception.

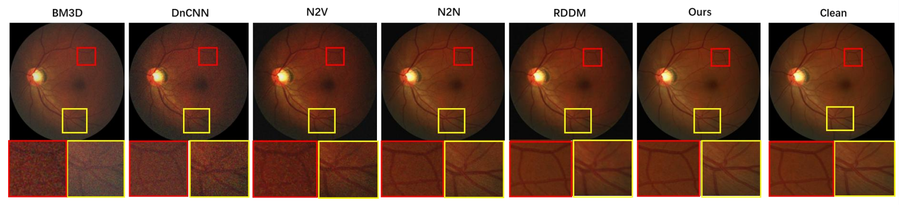


Experiments



Methods	SSIM ↑				Avg	HFRR (→ 1)				Avg	PSNR ↑ (dB)				Avg
	1	4	8	16		1	4	8	16		1	4	8	16	
BM3D[52]	0.808	0.883	0.907	0.929	0.876	1.198	1.176	1.159	1.134	1.167	24.423	29.564	33.985	36.924	31.224
DnCNN[56]	0.906	0.926	0.946	0.972	0.937	0.766	0.794	0.815	0.829	0.801	30.531	33.533	36.212	38.268	34.636
RIDNet[2]	0.896	0.905	0.926	0.942	0.917	0.709	0.754	0.795	0.809	0.767	30.321	33.343	36.014	38.568	34.561
N2N[22]	0.919	0.923	0.948	0.960	0.937	0.849	0.863	0.886	0.901	0.874	32.405	36.404	37.596	38.435	36.210
N2V[19]	0.920	0.933	0.941	0.956	0.937	0.821	0.856	0.923	0.942	0.885	32.156	35.281	37.359	40.986	36.695
RDDM[23]	**0.938**	**0.949**	0.953	0.961	0.950	0.893	0.903	0.915	0.919	0.907	**33.452**	36.223	38.059	40.556	37.072
Ours	0.936	0.945	**0.964**	**0.978**	**0.956**	**0.951**	**0.962**	**0.971**	**0.986**	**0.969**	33.311	**36.749**	**38.145**	**41.561**	**37.191**

Table 1: Quantitative comparisons with other methods in terms of the SSIM, HFRR and PSNR on the FMD dataset[58].



Methods	BM3D[52]	DnCNN[56]	N2N[22]	N2V[19]	RDDM[23]	Ours
SSIM ↑	0.837	0.842	0.876	0.872	**0.915**	0.902
HFRR (→ 1)	1.224	0.833	0.856	0.861	0.884	**0.946**
PSNR ↑ (dB)	27.924	29.269	32.454	32.986	36.556	**36.561**

Table 3: Average SSIM, HFRR and PSNR results of different methods on the 300 randomly selected from the Fundus dataset REFUGE [31].

Comparison with Other SOTA Denoising Methods Across Datasets

Category	Method	SIDD Validation [1]	SIDD Benchmark [1]	FMD [58]	W2S [62]	REFUGE [31]
Zero-shot	WNNM [13]	26.05/0.592	25.78/0.809	-	-	-
	DIP [45]	32.10/0.740	31.97/0.854	31.72/0.912	25.61/0.812	35.61/0.912
	Self2Self [36]	29.46/0.595	29.51/0.651	30.76/0.695	27.52/0.826	35.63/0.932
	PD-denoising [61]	31.87/0.820	33.61/0.890	32.86/**0.913**	27.62/0.852	36.62/0.944
	NN+denoising [59]	33.18/**0.899**	33.10/0.895	32.21/0.872	27.66/0.860	36.52/0.943
	APBSN-single [21]	30.90/0.818	30.76/0.815	31.47/0.891	26.67/0.852	35.22/0.931
	ScoreDVI [6]	34.75/0.856	34.70/0.890	33.10/0.896	27.77/**0.867**	37.09/0.845
	MASH [7]	35.06/0.851	34.89/0.920	33.01/0.888	27.71/0.853	36.87/0.935
	blind-spot denoising [38]	**35.31**/0.868	**35.05/0.922**	**33.95**/0.885	**27.88**/0.859	**37.20/0.948**
Self-supervised	NBR2NBR [15]	27.94/0.604	27.90/0.679	-	-	-
	CVF-SID [30]	34.81/0.944	34.71/0.917	32.73/0.843	29.86/0.907	33.29/0.913
	LUD-VAE [60]	34.91/0.944	34.70/0.920	33.59/0.915	30.99/0.905	35.48/0.925
	R2R [33]	35.04/0.844	34.78/0.898	-	-	-
	APBSN [21]	**36.73**/0.878	36.69/0.874	-	-	-
	PUCA [16]	-	37.54/0.936	-	-	-
	SelfFormer [37]	-	**37.69**/0.937	-	-	-
	Ours	36.31/**0.956**	37.65/**0.942**	**33.95/0.936**	**32.30/0.934**	**36.56/0.946**

Table 4: Comparison of different methods (PSNR(dB)/SSIM) on SIDD Validation, SIDD Benchmark, FMD,W2S, and REFUGE datasets. Best results are in **bold**.

Ablation & Analysis

Method	PSNR ↑	SSIM ↑	Wavelet MAE ↓
Full WG-MSDG (Ours)	**32.35**	**0.934**	**0.0213**
w/o FS-FF	31.96	0.881	0.0311
w/o Wavelet Branch	31.02	0.876	0.0455
w/o WCG Block	28.54	0.850	0.0387
w/o $\mathcal{L}_{\text{wavelet}}$	29.63	0.853	0.0420
Naive Fusion	29.71	0.856	0.0342

Table 5: Ablation study on key components of the proposed WG-MSDG model on W2S dateset by averaging 1 noisy frames.

Method	IQA	Fusion	$\mathcal{L}_{\text{IQA}}$	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Baseline (PatchGAN)	✗	–	✗	31.84	0.882	0.188	38.5
A: +IQA (no fusion)	✓	–	✗	32.12	0.895	0.165	35.4
B: +IQA + Channel	✓	Chan.	✗	32.58	0.802	0.142	32.7
C: +IQA + Spatial	✓	Spatial	✗	32.66	0.909	0.137	31.2
D: B + $\mathcal{L}_{\text{IQA}}$	✓	Chan.	✓	33.05	0.917	0.121	28.4
E: C + $\mathcal{L}_{\text{IQA}}$	✓	Spatial	✓	**33.31**	**0.938**	**0.109**	**26.8**

Table 6: Ablation study on the quality-aware discriminator. We investigate the contributions of the IQA module, fusion strategy, and IQA loss on the FMD dataset by averaging 1 noisy frames. Best results are **bolded**.

- Removing WCG or FS-FF sharply reduces PSNR/SSIM → proves importance of wavelet–spatial fusion.
- Quality-Aware Discriminator (QAD + LIQA) improves perceptual realism (↓LPIPS, ↓FID).

Takeaway: Each module contributes significantly to structural fidelity and perceptual quality.

Conclusion

- QWD-GAN bridges spatial and frequency domains for unsupervised microscopy denoising.
- It achieves high fidelity, perceptual realism, and superior efficiency.
- The model generalizes well across biomedical modalities — paving the way for interpretable, AI-driven microscopy.

University Of Manchester
qijun.yang@manchester.ac.uk