Supplementary materials: Wavelength-Embedding-guided Filter-Array Transformer for Spectral Demosaicing

Haijin Zeng[1(✉)], Hiep Luong[1], and Wilfried Philips[1]

IMEC-Ghent University, Ghent 9000, Belgium
haijin.zeng@imec.be

As detailed in the main body of our manuscript, this supplementary material showcases supplementary spectral demosaicing results across more than 100 scenes. These results are accompanied by both quantitative and visual comparisons for comprehensive evaluation.

Implementation details. Our network was trained using a combination of the L_1 Charbonnier loss and the focal frequency loss (FFL), with the FFL assigned a weight of 0.1. The FFL serves to enable our model to dynamically prioritize frequency components, particularly high-frequency details, which are challenging to reconstruct, by reducing the emphasis on easier components. The hyperspectral cubes utilized for both training and testing are standardized to a range of 0 to 1. During the training phase, random cropping is applied to generate patches of size $128 \times 128 \times$ while preserving the original spectral bands. For the wavelength embedding, similar to the timestep embedding in denoising diffusion models (DDPM), we randomly select a wavelength setting by shuffling the original ARAD wavelength layout for each training sample. The learning rate is 10^{-4}, employing the Adam optimizer. Furthermore, we evaluated two variants of our model: **WeFAT-S**, featuring 4 GAT blocks, and **WeFAT-M**, comprising 6 GAT blocks, varying in scale.

Ablation study w/ and w/o band order shuffling. We conducted an ablation study to assess the impact of shuffling the original ARAD band order on the model's transferability. It is important to note that the shuffling operation does not directly affect the Multispectral Filter Array (MSFA); rather, it rearranges the band order of the original hyperspectral image (HSI) cube, similar to the classic channel shuffle in deep learning. This strategy is employed to maximize the utility of limited HSI datasets by synthetically generating diverse MSFA patterns with varying wavelength orders. Consequently, our MSFA-attention mechanism and wavelength embedding become accustomed to a variety of MSFA and wavelength orders during training, thereby enhancing the model's transferability and generalization. To evaluate this, we included an ablation study on the effect of shuffling. We retrained our model from scratch without band shuffling and tested it on the ARAD test dataset from the CVPR NTIRE 2022 Spectral Demosaicing Challenge. The quantitative results are presented in Table 3, showing that shuffling improves the PSNR by 0.11 dB.

Table 1: Quantitative spectral demosaicing comparison using the ICVL dataset featuring 119 scenes, highlighting superior results in bold. Here, our primary focus is on comparing WeFAT with state-of-the-art (SOTA) learning-based methods.

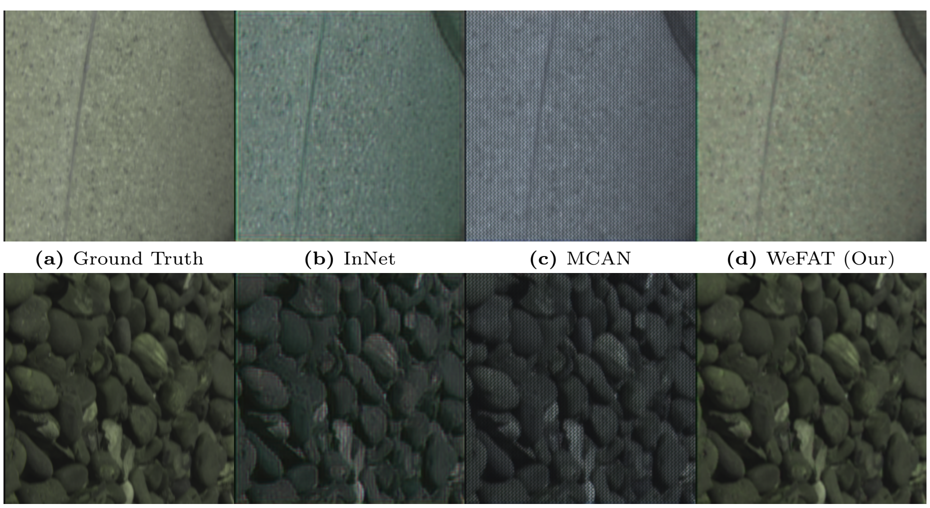
Dataset	Method	InNet	MCAN	**WeFAT**(Ours)
ICVL 119 scenes averaged	PSNR (↑)	33.549	30.647	**47.329**
	SSIM (↑)	0.931	0.578	**0.988**
	SAM (↓)	0.141	0.264	**0.057**

Table 2: Quantitative comparison on CAVE dataset.

Dataset	Image size	Method	MCAN	InNet	**WeFAT** (Ours)
103 cubes	$256 \times 256 \times 16$	PSNR (↑)	37.09	37.56	**45.27**
		SSIM (↑)	0.890	0.936	**0.974**

Table 3: Ablation study on band (channel) shuffling.

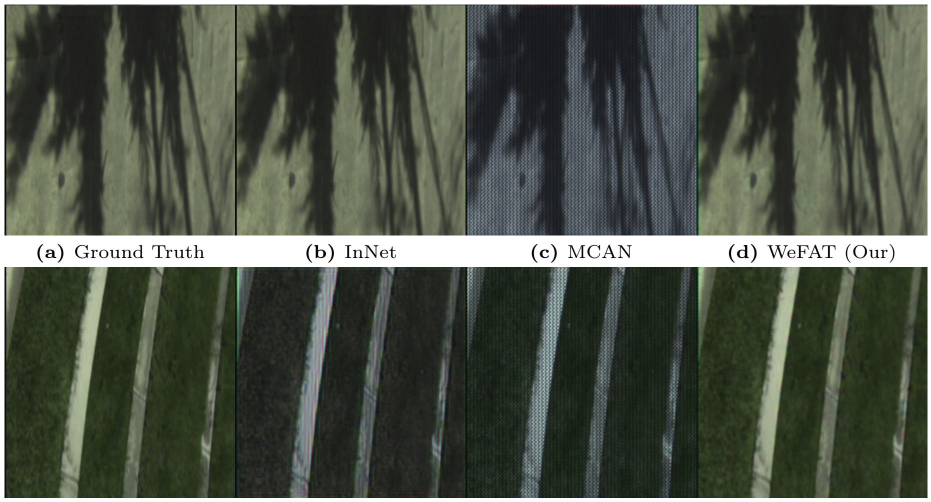
Dataset	Image Size	Band Shuffling	PSNR	SSIM
ARAD 50 scenes	$320 \times 320 \times 16$	w/o	47.92	0.993
		w/	48.03	0.994



(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

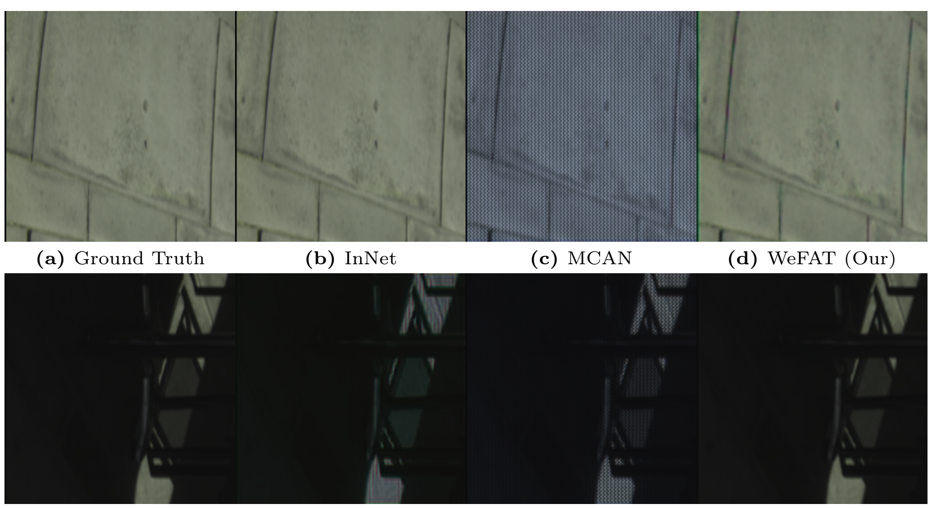
Fig. 1: Visual comparison (false color with spatial size 256×256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.



(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

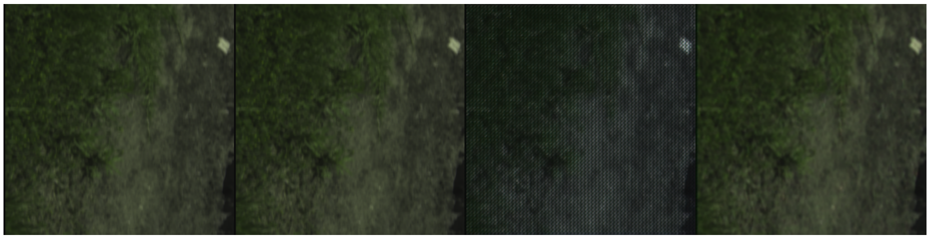
Fig. 2: Visual comparison (false color with spatial size 256×256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.



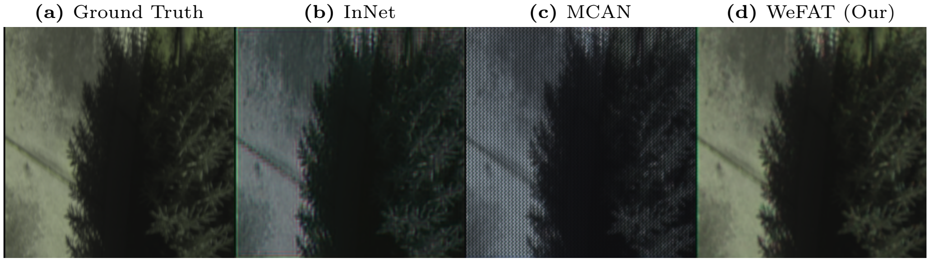
(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

Fig. 3: Visual comparison (false color with spatial size 256×256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.



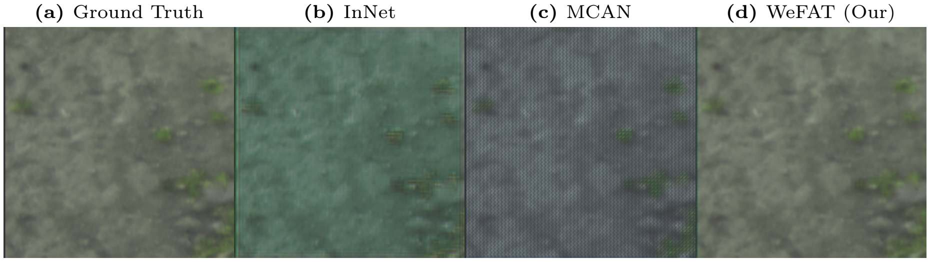
(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)



(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our) (Our)

Fig. 4: Visual comparison (false color with spatial size 256 × 256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.

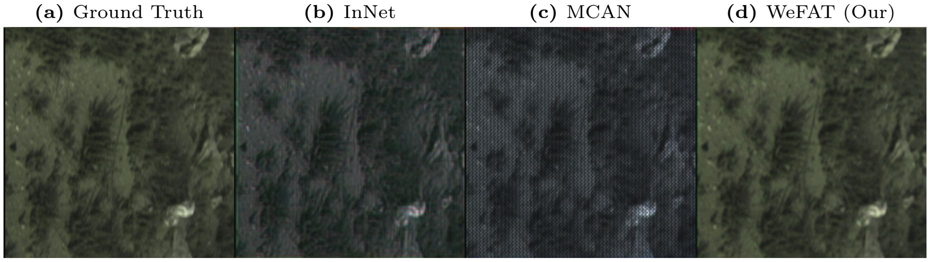
(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)



(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our) (Our)

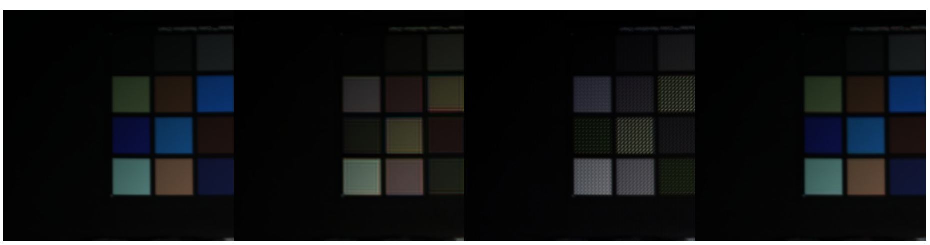
Fig. 5: Visual comparison (false color with spatial size 256 × 256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.

(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

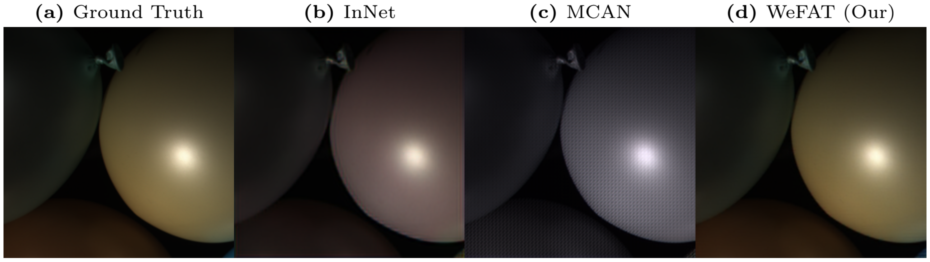


(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our) (Our)

Fig. 6: Visual comparison (false color with spatial size 256 × 256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.

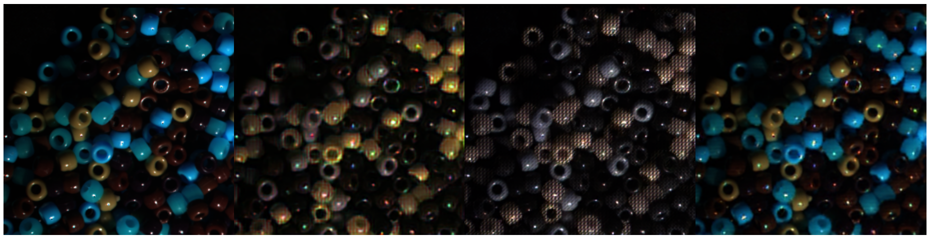


(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

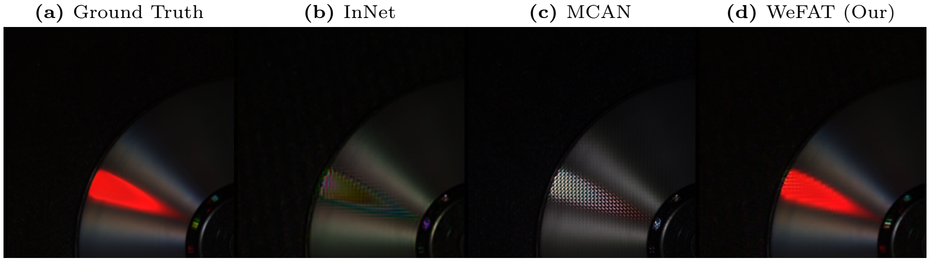


(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

Fig. 7: Visual comparison (false color with spatial size 256 × 256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.

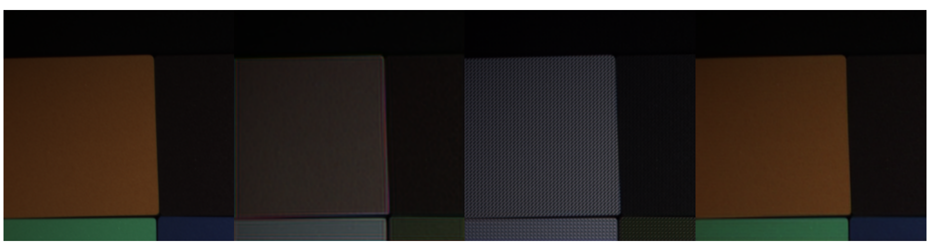


(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

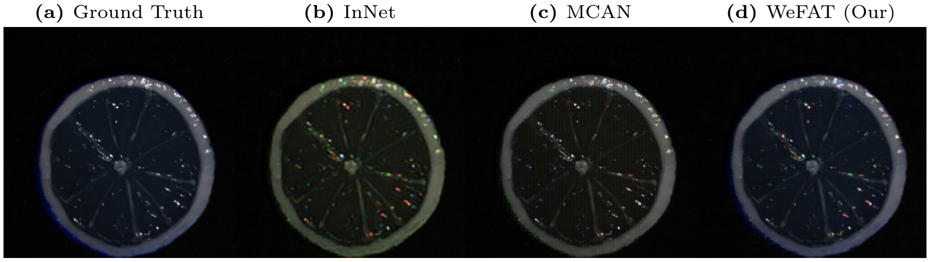


(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

Fig. 8: Visual comparison (false color with spatial size 256×256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.

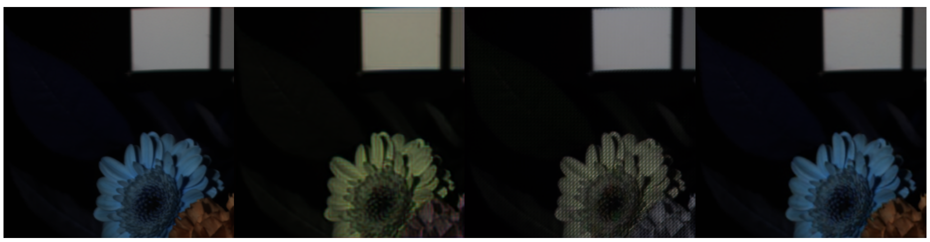


(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)

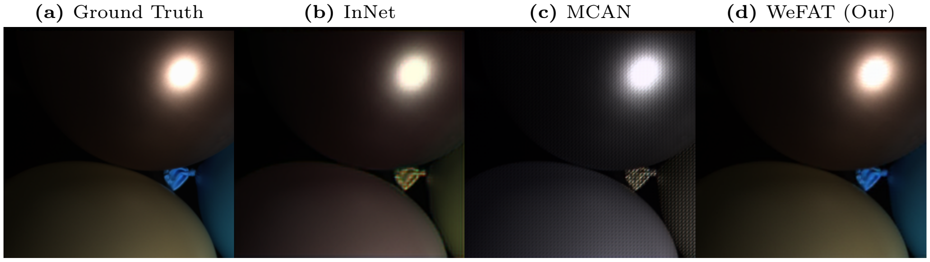


(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

Fig. 9: Visual comparison (false color with spatial size 256×256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.



(a) Ground Truth **(b)** InNet **(c)** MCAN **(d)** WeFAT (Our)



(e) Ground Truth **(f)** InNet **(g)** MCAN **(h)** WeFAT (Our)

Fig. 10: Visual comparison (false color with spatial size 256×256) of the proposed WeFAT and SOTA spectral demosaicing methods. The MSFA pttern used here is 4×4.