ESSAformer: Efficient Transformer for Hyperspectral Image Super-resolution (Supplementary Materials)

Mingjin Zhang[1], Chi Zhang[1*], Qiming Zhang[2*], Jie Guo[1], Xinbo Gao[3], Jing Zhang[2]
[1] Xidian University, China
[2] The University of Sydney, Australia
[3] Chongqing University of Posts and Telecommunications, China
mjinzhang@xidian.edu.cn, ch_zhang@stu.xidian.edu.cn, qzha2506@uni.sydney.edu.au,
jguo@mail.xidian.edu.cn, gaoxb@cqupt.edu.cn, jing.zhang1@sydney.edu.au

1. Ablation study on different attention types

To study the effectiveness of the proposed method, we use the ESSAformer structure with different attention types, i.e., conventional MHSA, inductive bias-induced SCC attention, and the proposed ESSA, for a thorough comparison. The results of PSNR, SSIM, and SAM are given in Table 1. Different from the test set settings in the paper, we crop the test image from 512 × 512 to 128 × 128 due to the huge computational and GPU footprint burden in MHSA and SCC attention. Thanks to the channel-wise inductive bias in SCC attention, it outperforms MHSA significantly on all the metrics. Meanwhile, ESSA has significantly less computation cost and performs on par with the SCC attention. The results demonstrate the effectiveness of the proposed ESSA.

Method		Dataset	PSNR	SSIM	SAM
MHSA	×4	Chikusei	41.1242	0.952	2.3693
SCC			41.3273	0.9539	2.3127
ESSA			41.4177	0.9554	2.3096
MHSA	×4	Cave	44.5341	0.969	6.9794
SCC			44.9210	0.9720	5.1792
ESSA			45.2727	0.9710	4.9596
MHSA	×4	Pavia	29.8248	0.8351	5.6746
SCC			30.0249	0.8410	5.5294
ESSA			30.1598	0.8468	5.5235

Table 1. Comparison on different attention. The experiments are conducted on three datasets with a scale factor 4×.

2. Qualitative result

In this section, we will provide more visual results to demonstrate the superiority of our method.

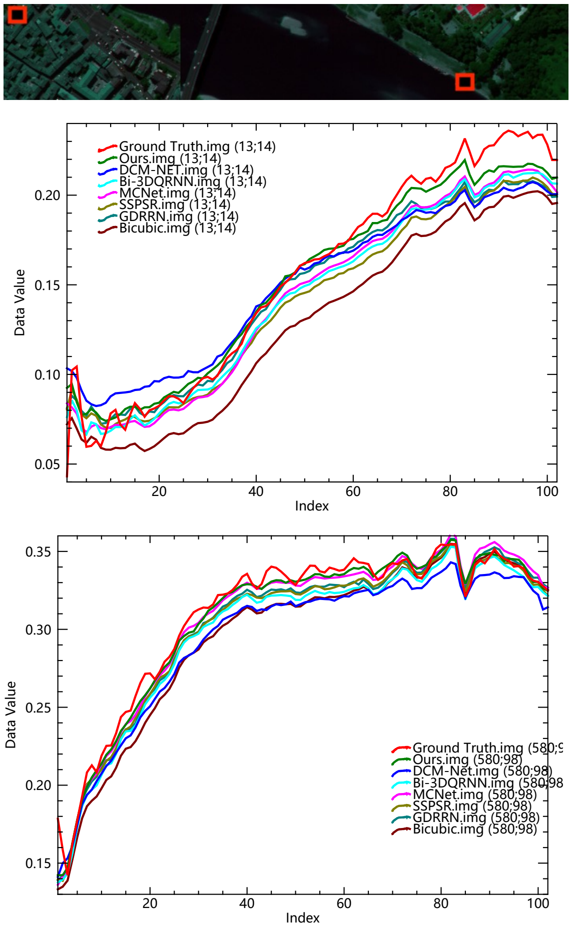


Figure 1. The spectral profiles of a test image from the Pavia dataset.

2.1. Spectral profiles

First, we select several images in the test set from Pavia, Chikusei, and Cave and plot the spectral profiles of the red

[*] Corresponding author

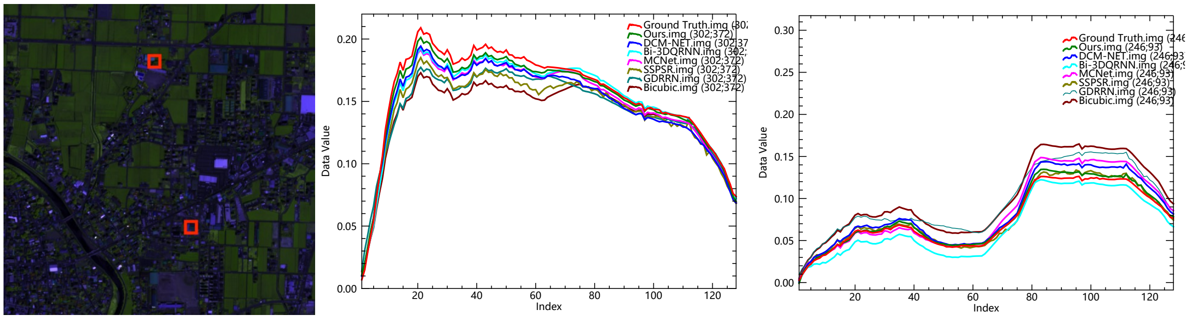


Figure 2. The spectral profiles of a test image from the Chikusei dataset.

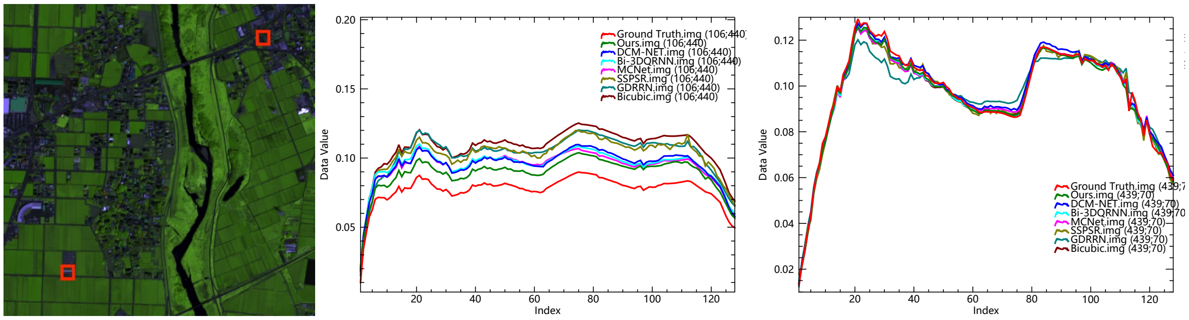


Figure 3. The spectral profiles of a test image from the Chikusei dataset.

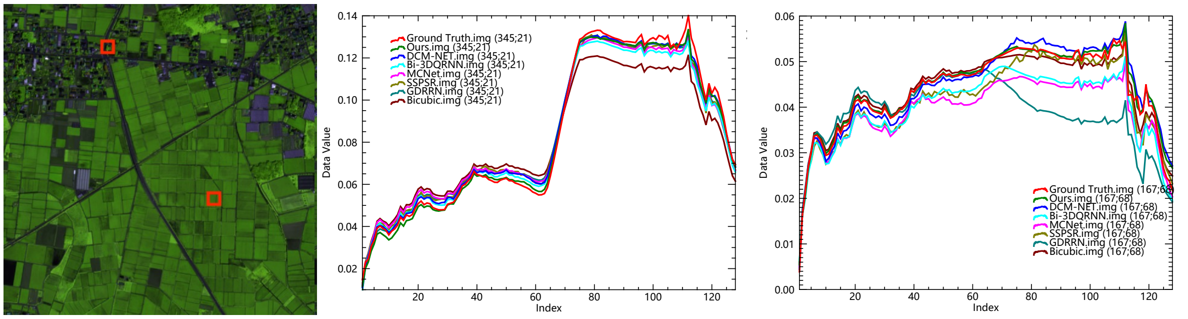


Figure 4. The spectral profiles of a test image from the Chikusei dataset.

circle regions as shown in Figure 1, 2, 3, 4, 5, and 6, respectively. It can be seen that our method, i.e., the green line, recovers the most information and is the closest to the ground truth red line in all figures at all data values. For example, the green line approaches the red most in the high-value range, i.e., the index between 60 and 100, in Figure 1, while others fail to reach the peaks compared to ESSAformer. In contrast, as shown in Figure 3, all the models tend to generate higher values in spectral profiles compared to the ground truth. Our ESSAformer, however, relieves this tendency most thanks to the introduced channel-wise inductive bias in ESSA attention.

2.2. Visualized absolute error

We visualize the absolute error maps of different methods as shown in Figure 7, 8, 9, 10, and 11. The original images after re-formatting to RGB ones are also given in each figure for reference. The pixels with dark colors denote the small error and the bright refers to having a large absolute error. From the figure, we can see that ESSAformer generates the images with the least textures and thus obtains the best performance. For example, our method produces results with the slightest difference from the ground truth, better recovering the bird's eye view image of Pavia city, as shown in Figure 7. Besides, the residual maps of other methods in Figure 8 and 9 show the 'road' clearly while ESSAformer has a strong ability to restore such edges in its outputs. When comparing the living area with abundant variance in the figures, ESSAformer also has darker results than the others, demonstrating its superiority for processing such challenging regions. Besides, as shown in the bottom leather and upper 'rectangle' regions in Figure 10, ESSAformer effectively restores the details, leading to a conclusion similar to the previous analysis.

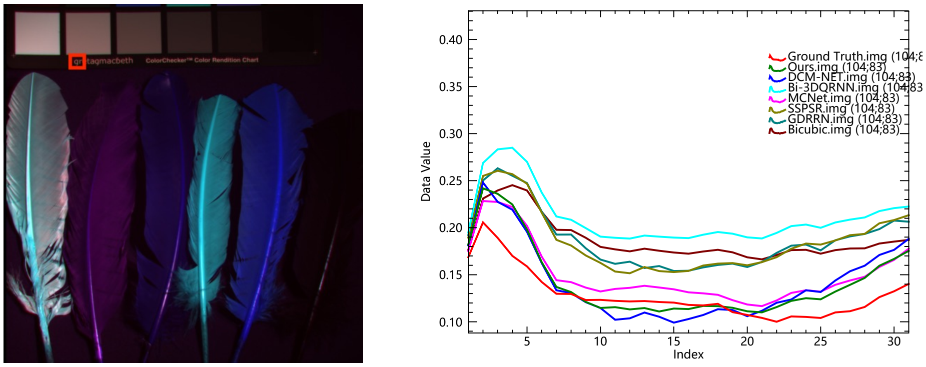


Figure 5. The spectral profiles of a test image from the Cave dataset.

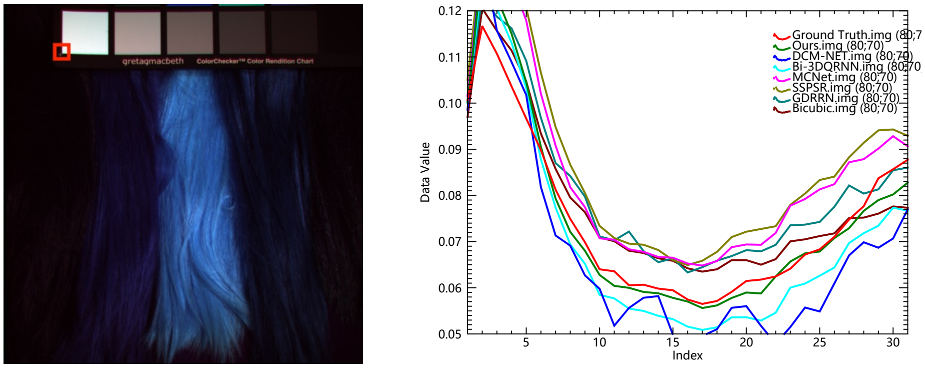
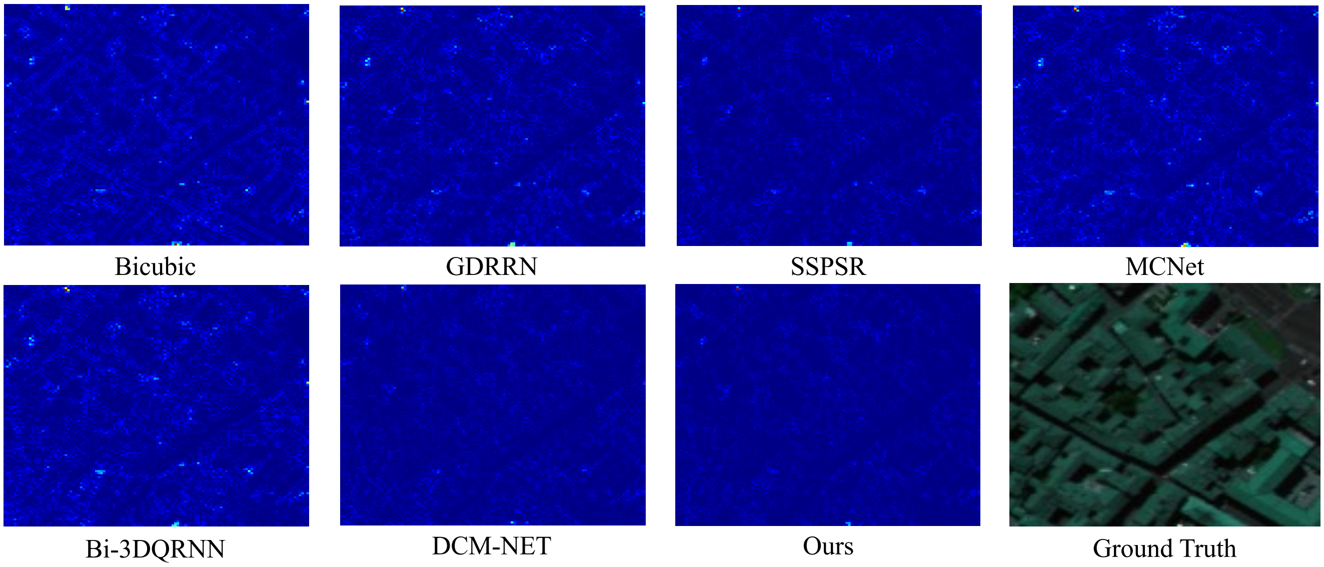


Figure 6. The spectral profiles of a test image from the Cave dataset.


Bicubic

GDRRN

SSPSR

MCNet

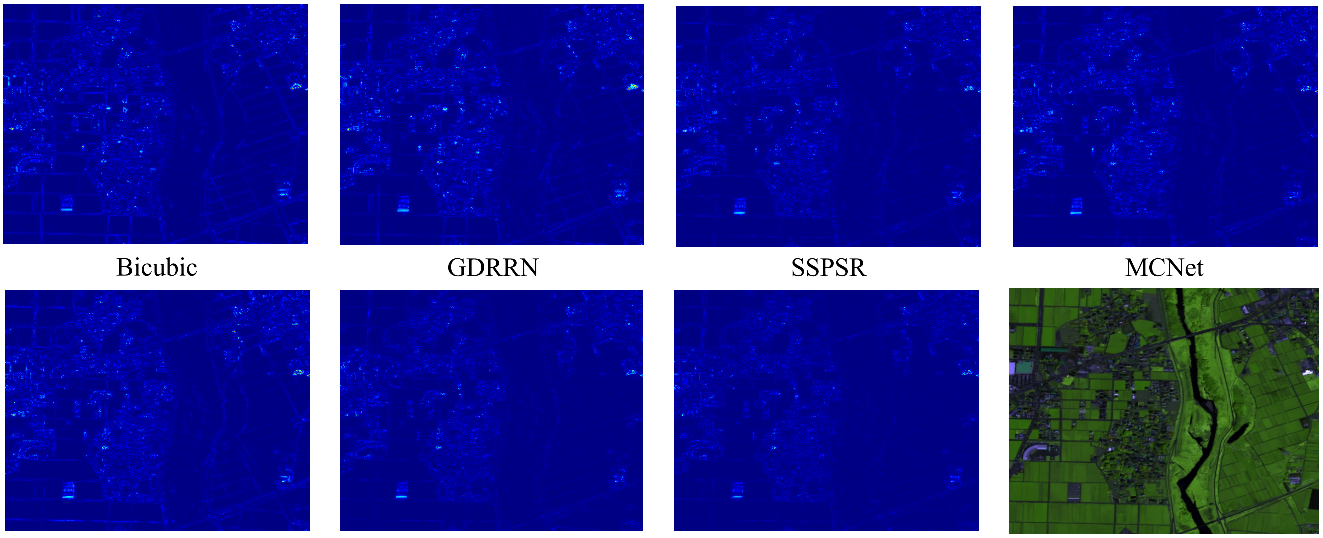
Bi-3DQRNN

DCM-NET

Ours

Ground Truth

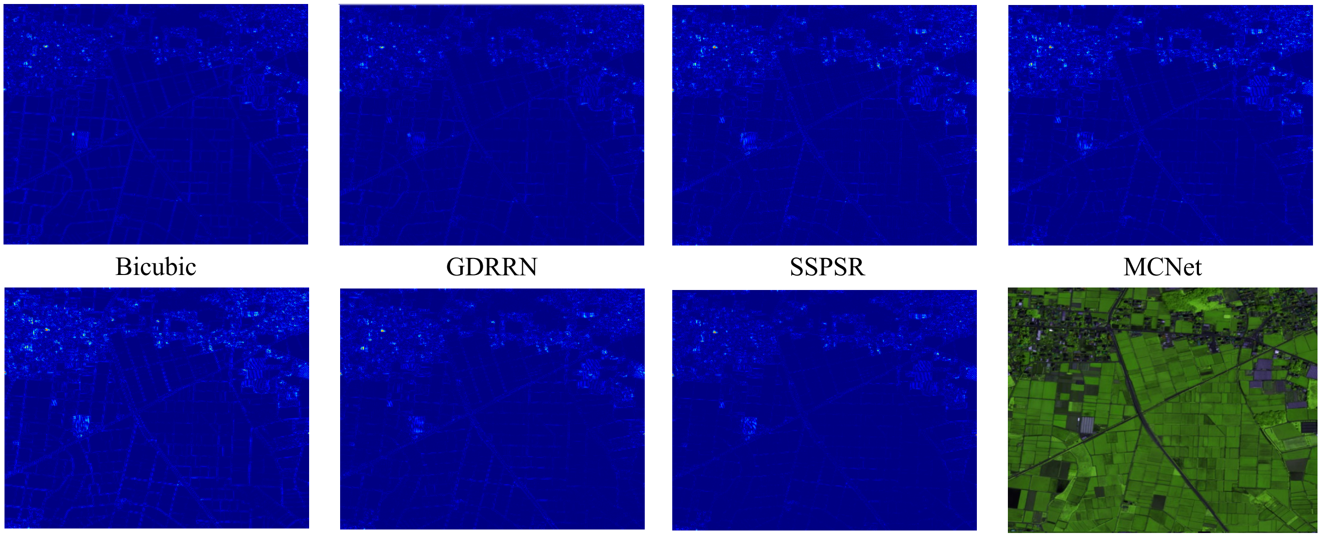
Figure 7. The absolute error map of a test image from the Pavia dataset.



Bicubic GDRRN SSPSR MCNet

Bi-3DQRNN DCM-NET Ours Ground Truth

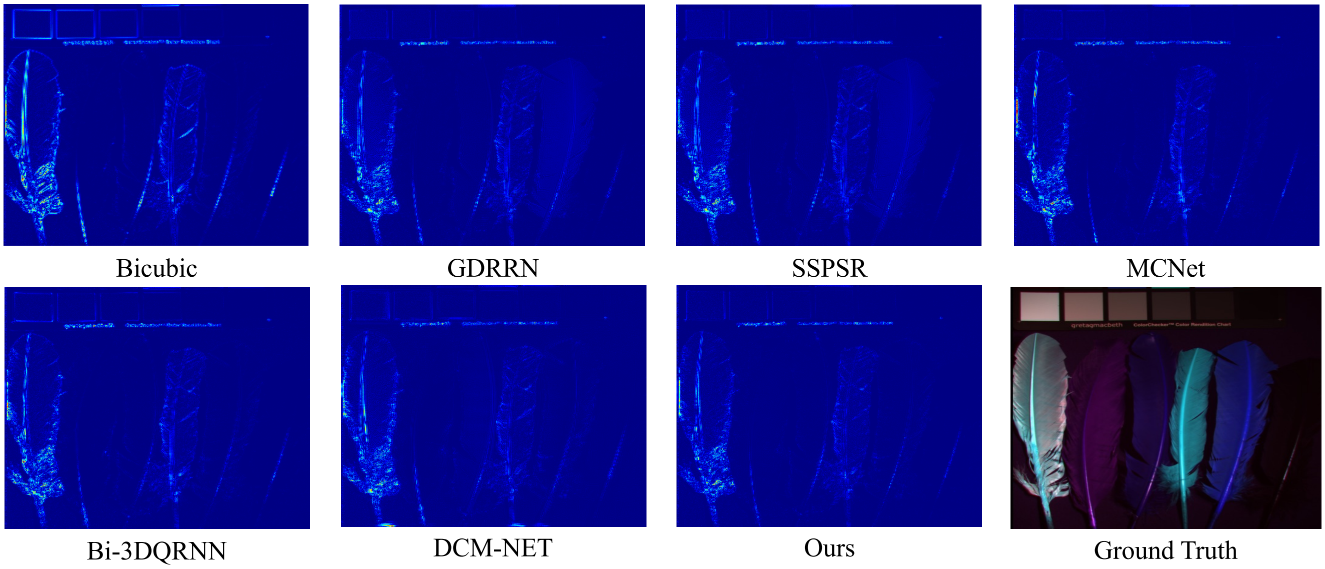
Figure 8. The absolute error map of a test image from the Chikusei dataset.



Bicubic GDRRN SSPSR MCNet

Bi-3DQRNN DCM-NET Ours Ground Truth

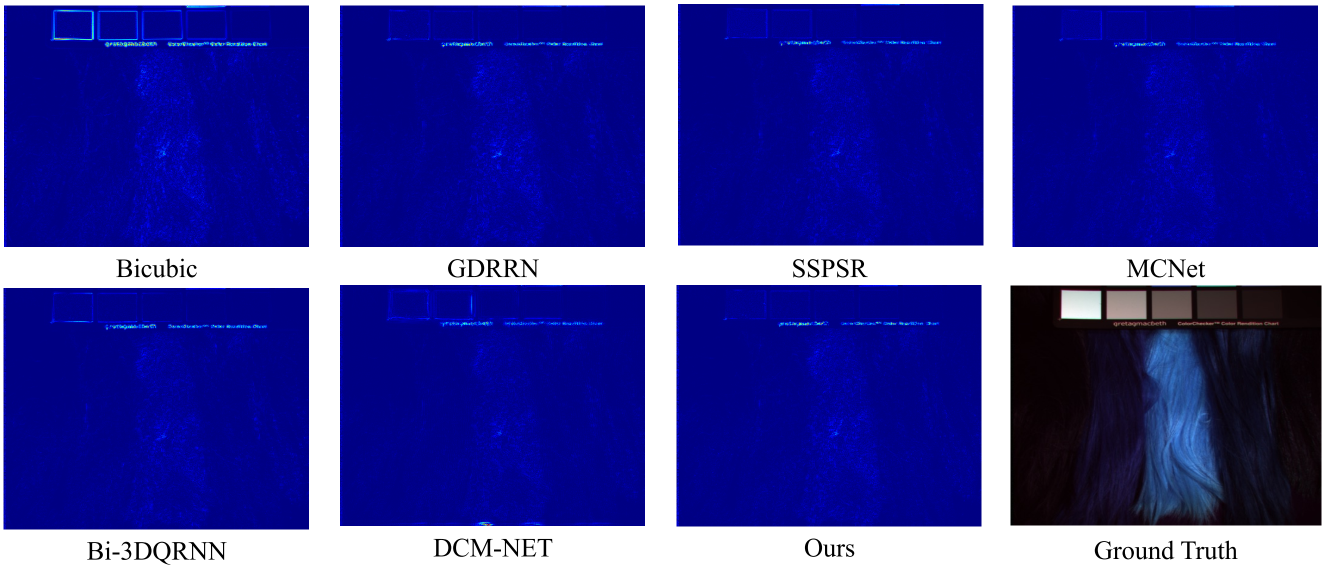
Figure 9. The absolute error map of a test image from the Chikusei dataset.



Bicubic GDRRN SSPSR MCNet

Bi-3DQRNN DCM-NET Ours Ground Truth

Figure 10. The absolute error map of a test image from the Cave dataset.



Bicubic GDRRN SSPSR MCNet

Bi-3DQRNN DCM-NET Ours Ground Truth

Figure 11. The absolute error map of a test image from the Cave dataset.