Maskpan: Mask Prior Guided Network for Pansharpening

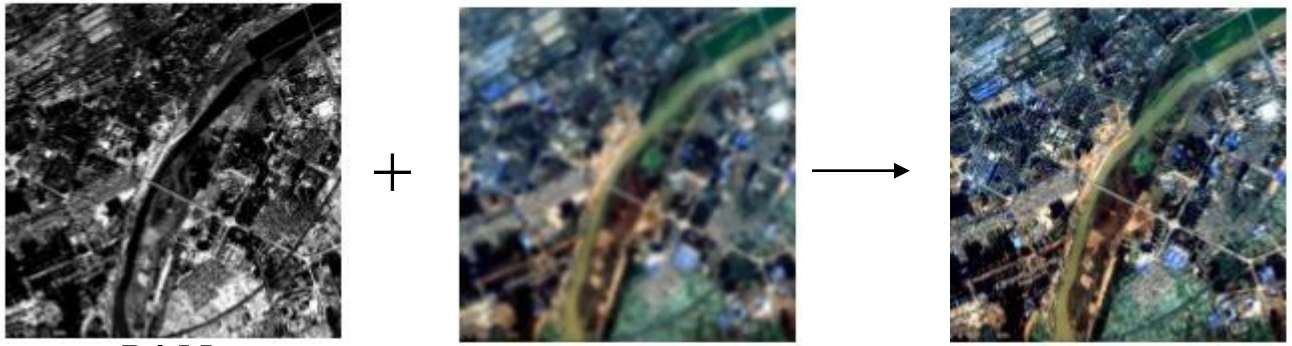
Xue Rui[1], Yang Cao[3,*], Yu Kang[2,3], Weiguo Song[1], Rui Ba[1]

[1]State Key Laboratory of Fire Science, [2]Institute of Advanced Technology,
[3]Department of Automation, School of Information Science and Technology,
University of Science and Technology of China

Introduction&Problem

University of Science and Technology of China

➢ *Pansharpening*



PAN LRMS HRMS [1]

➢ *Previous works： visual pleasing*

- Traditional image processing methods
- Deep Learning based pansharpening methods

➢ *Limitation：visual discrimination*

- Pansharpened images may be limited in providing discriminablity for visual tasks
- The applications of satellite images are greatly affected by image visual discrimination.

[1]PAN(panchromatic images), LRMS(low spatial resolution multispectral images), HRMS(high resolution multispectral images)
Liu X , Wang Y , Liu Q . PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening[J]. 2018.

Motivation

中国科学技术大学 University of Science and Technology of China

- High-level semantic feature has the guiding role to low-level image task contribution
- Existing pansharpening methods are limited in providing discriminability for visual tasks

Contribution

- The CNN-based method that utilizes high semantic features (mask prior) to guide pansharpening task.
- The features of MS and PAN images are fused with prior information in feature domain by an attention mechanism.
- Semantic segmentation is adopted as an evaluation metric for the visual discrimination.

Architecture

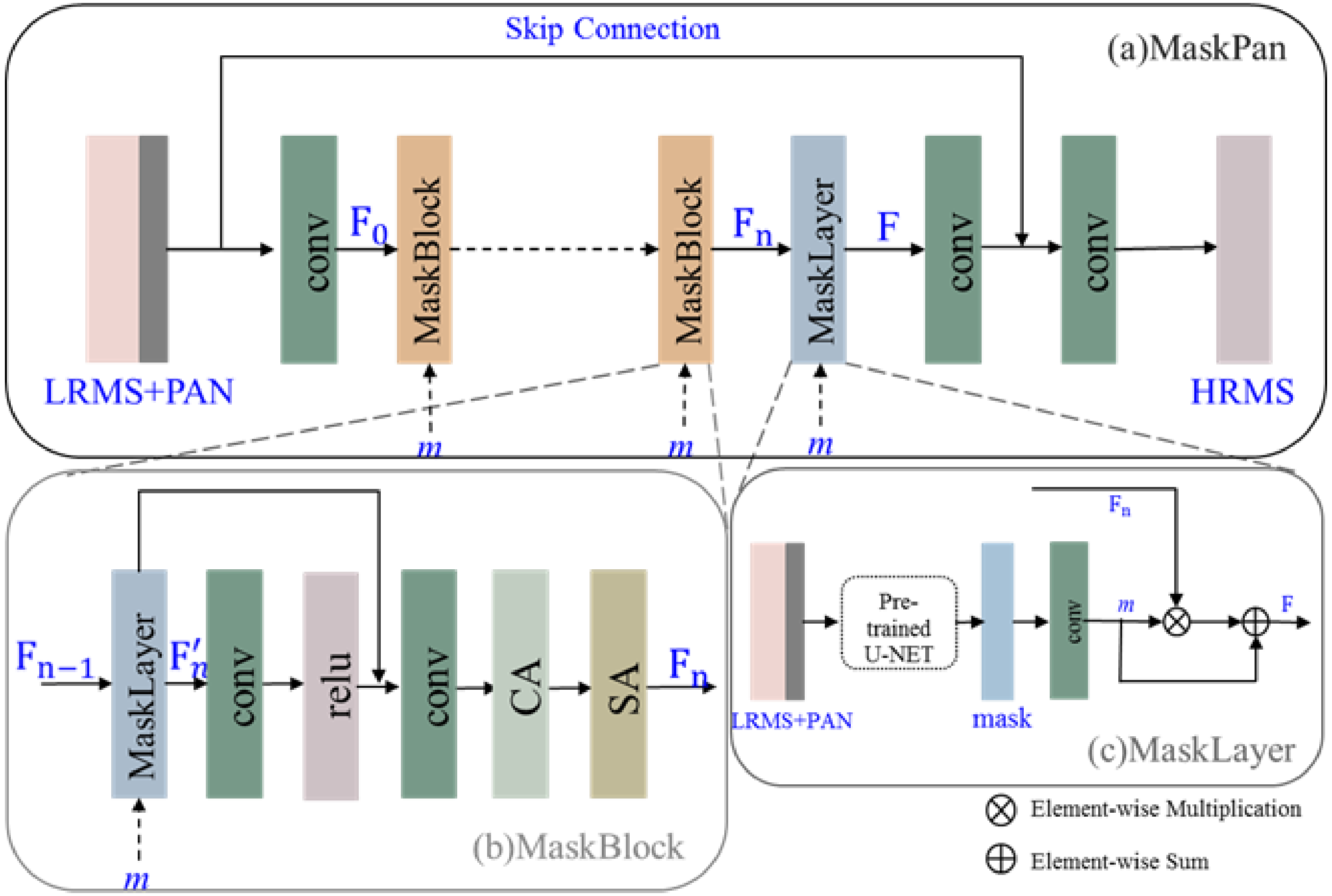


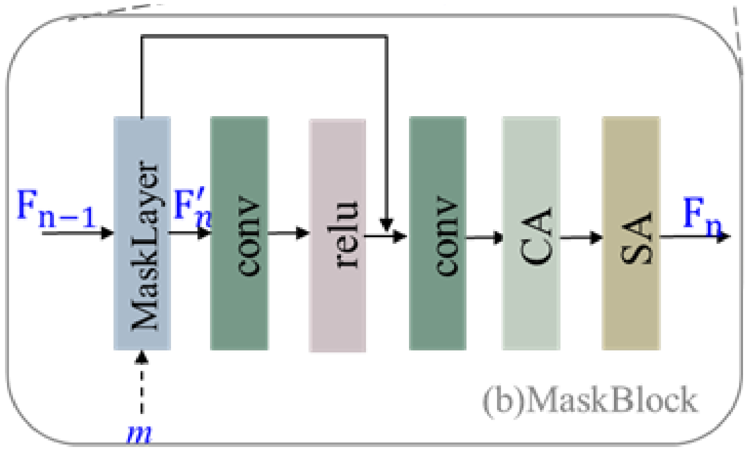
Figure 1. Structure illustration of the proposed MaskPan. (a) network architecture of MaskPan. (b) MaskBlock: integrating the MaskLayer and attention mechanism, where CA and SA refer to the (channel-wise and spatial attention). (c) MaskLayer:showing the combination way of mask prior, all MaskLayers share the same mask input.

Architecture

- *MaskBlock*

Inspired by the CBAM of [2]

MaskLayer / channel& spatial attention mechanism



$$
\begin{aligned}
F_n &= G_{MB,n}(F_{n-1}, m) \\
&= G_{MB,n}\Big(G_{MB,n-1}\big(\cdots\big(G_{MB,1}(F_0, m)\big)\cdots\big)\Big)
\end{aligned}
$$

$G_{MB,n}$:the operations of the n-th MB

m :estimated mask prior information

[2]CA(channel attention), SA(spatial attention)

S. Woo, J. Park, J. Lee, I. Kweon, "CBAM: Convolutional Block Attention Module," European Conference on Computer Vision (ECCV), 2018

Architecture

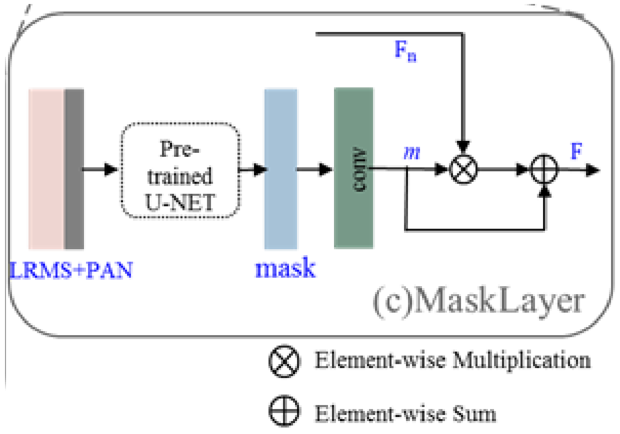
- *MaskLayer*

Designed based on the the structure of SFT Layer [3]

n-th MaskLayer in the n-th MaskBlock

$$m' = u(x, p; \alpha)$$

$$F_n' = mF_{n-1} + m$$



m':mask generated from U-Net
m: processed mask prior
F_{n-1} :features from $(n-1)$-th Mask Block
F_n' :features to be fed into the attention layer

[3]X. Wang, K. Yu, C. Dong, C. Loy, "Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform"The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 606-615,2018

Experiment

Dataset:World-View3 satellite image dataset of the *kaggle* DSTL competition[4]

➢ *Ablation study*: Comparison of different kinds of mask prior

- Mask_*C*: directly concatenates the mask with the stacked up-sampled MS and PAN images
- Mask_*M*: ten gray-level prior maps

 (multiplies the mask with stacked up-sampled MS images and PAN images)
concatenated with the stacked up-sampled MS images and PAN images

- Mask_*EX*: only MS and PAN images

Table 1.Comparison of the methods with different types of mask prior.

Method	PSNR (↑)	SSIM (↑)	SCC (↑)	SAM (↓)	ERGAS (↓)
Mask_*C*	27.605	0.982	0.998	0.016	0.590
Mask_*M*	27.694	0.982	0.998	0.016	0.590
Mask_*EX*	27.714	0.982	0.998	0.016	0.576
MaskPan	**28.401**	**0.983**	**0.999**	**0.015**	**0.535**

[4] Dstl Satellite Imagery Feature Detection, https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection

Experiment

- Comparison of MaskPan with other models

1.Image quality evaluation indices: PSNR/SSIM/SCC/SAM/ERGAS

Table 3. Comparison of image quality evaluation indices.

Method	PSNR (↑)	SSIM (↑)	SCC (↑)	SAM (↓)	ERGAS (↓)
BT[25]	19.855	0.795	0.901	1.452	3.578
IHS[26]	22.094	0.946	0.924	1.570	2.006
PNN[2]	20.343	0.878	0.969	0.086	2.553
TFNET [8]	26.799	0.960	0.995	0.034	0.792
Mask_*EX*	27.714	0.982	0.998	0.016	0.576
MaskPan	**28.401**	**0.983**	**0.999**	**0.015**	**0.535**

Experiment

- Comparison of MaskPan with other models

2.Evaluation metric for visual discrimination of pansharpened images: IoU

Table 2.Comparison of segmentation results for different classes in terms of IoU.

Method	IoU-mean	IoU-buildings	IoU -Misc.	IoU -road	IoU -track	IoU -trees	IoU -crops	IoU -waterway	IoU -standing water
PNN[2]	0.419	0.691	0.104	0.656	0.343	0.602	0.765	0.801	0.132
TFNET[8]	0.448	0.706	0.170	0.732	0.369	0.615	0.827	0.824	0.134
Mask_*EX*	0.452	0.710	0.170	0.734	0.391	0.614	0.842	0.829	0.135
MaskPan	**0.465**	**0.728**	**0.173**	**0.739**	**0.436**	**0.634**	**0.849**	**0.841**	**0.143**

Experiment

- Comparison of MaskPan with other models

3.Visual comparison

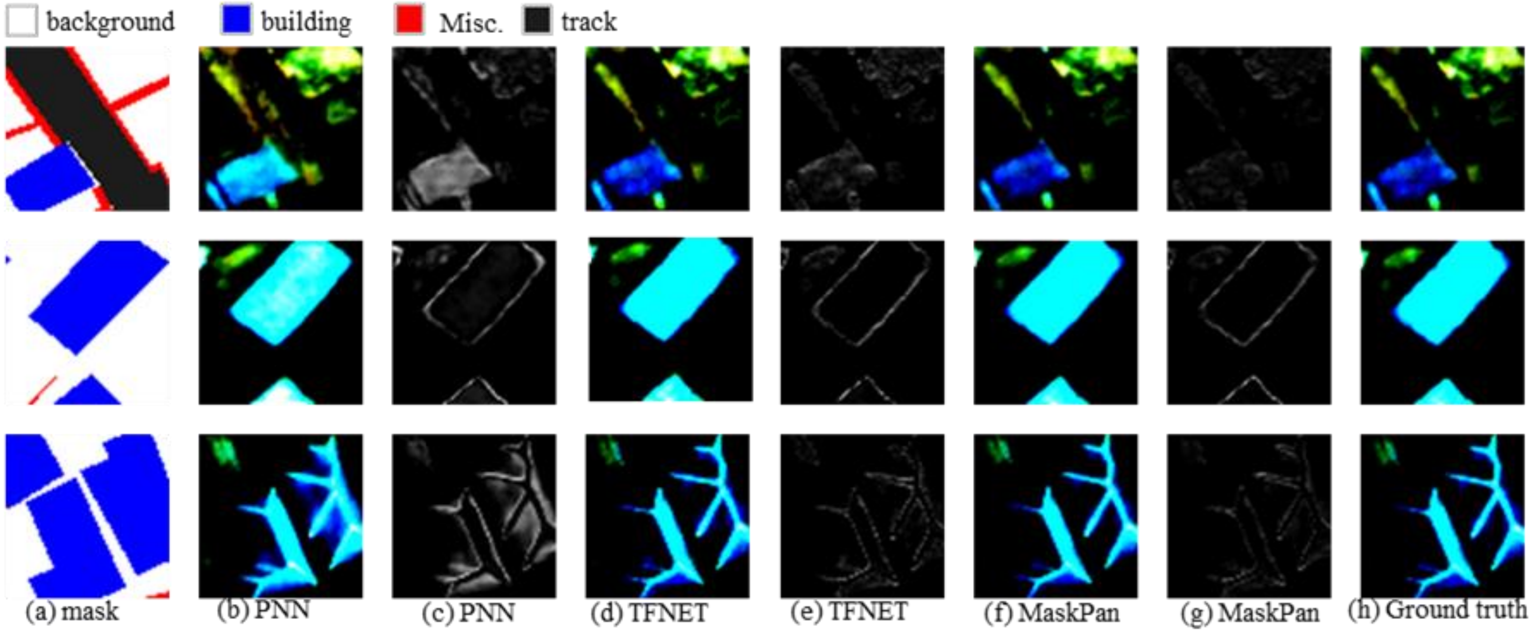


Figure 2. Comparison of pansharpening results for WordView-3 images. (b), (d) and (f) are the pansharpening results of the network PNN, TFNET and our MaskPan, while (c), (e) and (g) are the corresponding residual images; (a) and (h) are the ma-sk and ground truth of the images, respectively.

Thank You!

中国科学技术大学
University of Science and Technology of China