Residual Pixel Attention Network for Spectral Reconstruction from RGB Images

Hao Peng[1,2], Xiaomei Chen[1,2,*], Jie Zhao[1,2],
1 School of Optics and Photonics, Beijing Institute of Technology, Beijing, China
2 Key Laboratory of Photoelectronic Imaging Technology and System, Beijing Institute of Technology, Ministry of Education of China, Beijing, China

Abstract

In recent years, hyperspectral reconstruction based on RGB imaging has made significant progress of deep learning, which greatly improves the accuracy of the reconstructed hyperspectral images. In this paper, we proposed a convolution neural network of the hyperspectral reconstruction from a single RGB image, called Residual Pixel Attention Network (RPAN). Specifically, we proposed a Pixel Attention (PA) module, which was applied to each pixel of all feature maps, to adaptively rescale pixel-wise features in all feature maps. The RPAN was trained on the hyperspectral dataset provided by NTIRE 2020 Spectral Reconstruction Challenge and compared with previous state-of-the-art method HSCNN+. The results showed our RPAN network had achieved superior performance in terms of MRAE and RMSE.

1. Introduction

Hyperspectral imaging technology can obtain detailed spectral information from a target or a scene. Hyperspectral images can have hundreds or thousands of spectral bands. Different objects have different spectral characteristics. By comparing the spectral characteristics, different targets can be distinguished. Therefore, hyperspectral images are widely used in various computer vision tasks, such as target tracking [29, 32], detection [14, 9], and classification [35, 7].

Hyperspectral images are acquired by hyperspectral cameras. Although there are many advantages of hyperspectral images, the cost of hyperspectral cameras is often very high, which hinders many low-cost applications. To obtain high spectral resolution, the spatial or temporal resolution is inevitably reduced, which severely limits the application range of hyperspectral images. Therefore, obtaining hyperspectral image with high spatial resolution at low cost has become an important direction.

Reconstructing hyperspectral images from RGB images is a way to solve the problem. RGB images can be obtained by low-cost RGB cameras with high spatial resolution. Reconstructing hyperspectral images from RGB images is a severely ill-posed problem, so some methods [4, 13, 33, 1] introduce the inherent statistics of hyperspectral images into the model as a prior. The prior models are often based on experience, so the generalization ability is limited. With the excellent performance of deep learning methods on many computer vision tasks, convolutional neural network(CNN) methods have also been introduced into spectral reconstruction tasks, these CNN-based methods greatly improve the quality of the reconstructed hyperspectral images [10, 2, 27, 26, 21, 18].

Recently, the effect of the attention mechanism has been proved on many computer vision tasks [12, 6, 31], which can further improves the performance of CNN. The attention mechanism makes the output of the neural network more affected by the most content components and reduce the influence of useless information, which is the main reason that the attention mechanism works. However, few methods based on CNN methods apply the attention mechanism to the spectral reconstruction task.

In the single-image super-resolution task, the effect of the attention mechanism has also been proved in RCAN [36], which effectively improves the super-resolution performance of the RCAN model through the channel attention module. Both image super-resolution and spectral reconstruction tasks are serious ill-posed problems, and spectral reconstruction can be regarded as super-resolution in the spectral dimension. It can be seen that there are some similarities between the spectral reconstruction and image super-resolution. Therefore, we try to introduce the attention mechanism into the spectral reconstruction network.

Channel attention mechanism can adaptively rescale each channel-wise feature by learning the interdependency between the channels, which will make CNNs pay more attention to the more important features. However, the dif-

*Corresponding author: cxiaomei@bit.edu.cn

ferent positions of a channel also have different degrees of importance. The channel attention used by RCAN cannot simultaneously adaptively rescale the channel-wise features and the features of different positions in a channel with different scale factors.

In this paper, we propose a novel attention module called pixel attention (PA) to solve the above problems, and apply it to the hyperspectral image reconstruction task. Specifically, we designed the residual pixel attention network(RPAN) for spectral reconstruction from RGB image, in which the PA can generates a rescale factor cube for each pixel of all feature maps by considering the interdependency in feature cube.

Our main contributions include:

(1) We propose a residual pixel attention network (RPAN) for spectral reconstruction from RGB. (2) Present a novel module called pixel attention block (PA), which can adaptively rescale each pixel-wise weights of all input feature maps. (3) Experimental results have shown our RPAN has a significantly improvement over previous state-of-the-art HSCNN+ [26] on NTIRE 2020 spectral reconstruction challenge dataset.

2. Related Work

Hyperspectral images are generally obtained by scanning, including whiskbroom [24] and pushbroom [25]. Scanning methods have inherent deficiencies, and their low temporal resolution makes them unsuitable for quickly acquiring hyperspectral images and capturing dynamic scenes.

In order to improve the time resolution and quickly capture hyperspectral images and dynamic scenes, various snapshot hyperspectral imaging systems have been developed.

Based on the theory of compressed sensing, the Coded Aperture Snapshot Spectral Imaging (CASSI) [30] can quickly capture hyperspectral images.The detector receives a two-dimensional image with the coded spectrum and spatial information mixed, which can be restored to a hyperspectral image through a restoration algorithm. The computer tomography imaging spectrometer (CTIS) [8] records the projection results of the three-dimensional hyperspectral information on the two-dimensional plane at different angles, and then reconstructs the hyperspectral image through the algorithm. However, these imaging systems are quite complex in hardware and reconstruction algorithms.

RGB images can be obtained at low cost by RGB cameras, and RGB images can have high spatial and temporal resolution, so reconstructing hyperspectral images through RGB images has aroused widespread interest.

Oh *et al*. [22] proposed a framework for reconstructing hyperspectral images by using multiple consumer-level digital cameras, which reconstructs hyperspectral images through multiple RGB pictures taken by cameras with different sensitivity functions.

Nguyen *et al*. [20] use a radial basis function network to learn the mapping relationship between RGB and scene spectral reflectance, and then reconstructs the scene spectral reflectance image. Arad *et al*. [4] built a large hyperspectral data set of natural scenes, and used this data set to build an over-complete dictionary to reconstruct hyperspectral images from RGB images. Aeschbacher *et al*. [33] introduced the A+ [28] algorithm to improve the performance of the Arad method. With the excellent performance of the CNN method on many tasks, the CNN method began to be introduced into the spectral reconstruction task.

Galliani *et al*. [10] introduces the CNN originally used for semantic segmentation, realizes the reconstruction of hyperspectral images from RGB images, and achieves excellent performance. AlvarezGila *et al*. [2] proposed a method based on generative adversarial network (GAN) to achieve hyperspectral image reconstruction.

In the NTIRE 2018 Spectral Reconstruction Competition, participants developed a variety of novel CNN methods to achieve spectral reconstruction [3]. Among them, Shi *et al*. [26] improved the HSCNN [34] network, proposed the HSCNN + network and won the first and second places in the NTIRE 2018 Spectral Reconstruction Competition.

In the above CNN method, almost all methods treat each position of each feature map equally, but the features of different positions and different channels are not the same in importance. The features of some positions and some channels are more helpful for the spectral reconstruction. In order to make better use of features, we proposed a novel attention module pixel attention (PA), which can properly adjust each feature in all channels and make the network pay more attention to more important features.

3. Proposed Network

3.1. Network Design

The proposed Residual pixel attention network (RPAN) details is shown in Figure 1.

The RPAN was inspired by the RCAN, so it should be noted that the proposed RPAN network and RCAN network have some similarities in composition. The RCAN is decomposed of the residual group (RG) and residual channel attention block (RCAB), RPAN is decomposed of the residual attention group (RAG) and residual pixel attention block (RPAB).

RCAB uses the channel attention module, while RPAB uses the new attention module pixel attention (PA) block. In the overall structure of the network, we have also concatenate feature maps generated by RAGs to take full advantage of the different levels of feature maps generated by RAGs. The experimental results in Sec 4.3 show that the proposed

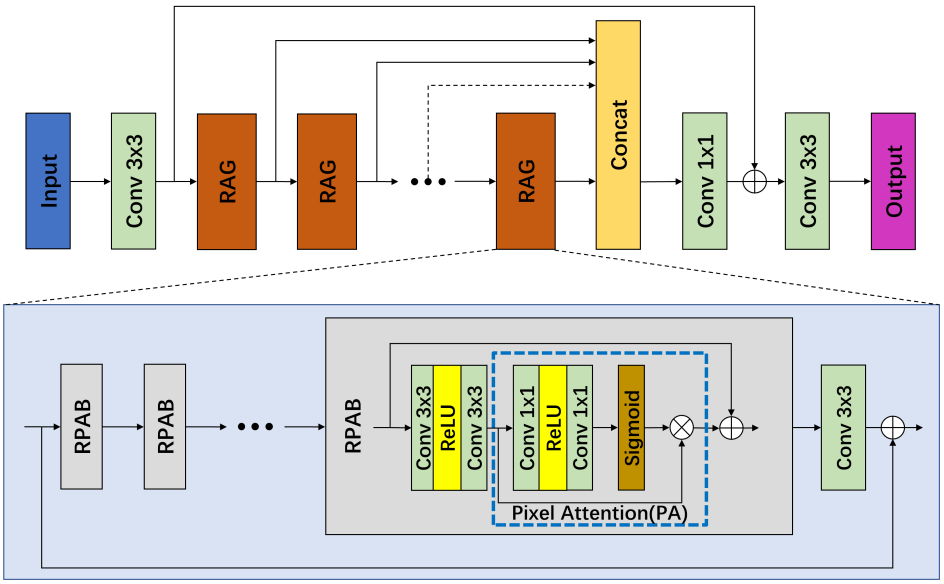

Figure 1. RPAN(up) and a RAG block(down).

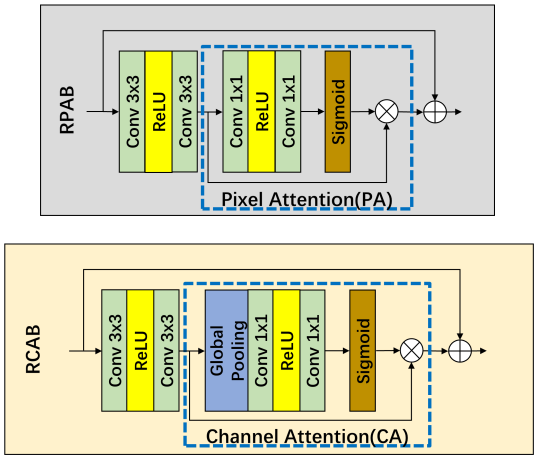

Figure 2. RPAB block(up) and a RCAB block (down).

method have significantly improved the effect of hyperspectral image reconstruction.

3.2. Overall Network Structure

Our proposed network uses a global residual network structure like VDSR [15], it can speed up convergence. The network first goes through a 3×3 convolution layer which have 64 filters, then the main network is constructed by stacking 4 RAG blocks followed by 1×1 convolution layer in order to better use these features. By skip connection, it adds 64 feature maps from the first 3×3 convolution layer by global skip connection. Then output 31 channel spectral image through the last 3×3 convolution layer.

Each RAG is composed of 8 residual pixel attention blocks(RPAB) and a 3×3 convolution layer. The RAG also use residual learning to speed up convergence. We integrate PA and RB [11] together, and proposed the RPAB block.

When not stacking RAG modules, input RGB picture, and the output of the network containing m RAG modules can be expressed as:

$$y = g(f(x) + F_m(\cdots F_2(F_1(f(x))))) \tag{1}$$

where x denotes the input RGB picture and y represents the reconstructed hyperspectral image. f denotes the function of the first convolutional layer of the network and g represents the last convolutional layer of the network and F_m denotes the function of the m-th RAG.

3.3. Residual Pixel Attention Block (RPAB) and Pixel Attention (PA)

Previous CNN-based spectral reconstruction methods treat pixel-wise features almost equally without attention mechanism. However, the influence of the features in different positions and channels on the reconstruction results of hyperspectral images are likely to be different. The features of some locations and some channels are more important, and some features are less important. For example, the pixels with sharp spectrum changes, each band has a stronger correlation with the adjacent band, which requires the network to pay more attention to the adjacent channels; the pixels at the boundary between the foreground and back-

ground also need to pay more attention to the surrounding pixels to better reconstruct this part of the area.

In order to solve the above problems, inspired by RCAN [36] and SE-net [12], the attention mechanism is introduced into the RPAB module which contains the novel attention module PA.

Both RCAN and SENet used the channel attention (CA) module. The channel attention generates different attention for each channel. In order to adaptively rescale the weight of each channel, the CA module needs to learn the nonlinear relationship between the channels.

As shown in Figure 2, the channel attention module first obtains a $1 \times 1 \times C$ size vector through global pooling, and each component represents the statistical information of the features in a channel, and then passes through $1 \times 1\times$ convolution layer(squeeze Conv layers), ReLU activation function, $1 \times 1\times$ Convolution(expand Conv layer), learn the nonlinear relationship of each channel during the training process, and finally normalize the scale factor to the range [0,1] through the sigmoid function. In this process, the vector of $1 \times 1 \times C$ will first reduce the dimension to $1/r$ of the input, and then after ReLU activation, it will rise back to the original dimension, which can greatly reduce the calculation complexity.

Since not only different channels but also different positions have different importance, different positions of a channel require different scale factors. The channel attention mechanism cannot simultaneously adaptively rescale the channel-wise features and the features of different positions in a channel with different scale factors. The new attention module PA is proposed to solve the problem.

Unlike RCAB and SENet block, PA is not CA, and it removes the global pooling layer and the rest is the same as CA. In PA, the input with the dimension of $H \times W$ that contains C feature maps , it will becomes $H \times W \times C/r$, and then back to $H \times W \times C$, through training to learn the relationship between each position of all feature maps, and finally normalize the scale factor to the range [0,1] through the sigmoid function. Through the PA module, we obtain a 3D scale factors data cube with the same size as the input feature, so that each position of the 3D feature cube can be adaptively rescaled to better reconstruct the hyperspectral image from RGB images.

PA can be formulated as

$$PA = s(W_E(\delta(W_S(x)))) \tag{2}$$

where s and δ denote the sigmoid function and ReLU [19] function. W_s is the weight of the squeeze Conv layer, which act channel-downscaling with reduction ratio r. We is the weight of the expend Conv layer, which act channel-upscaling with reduction ratio r. x denotes the input feature with dimension $H \times W \times C$.

So we can rescale the feature maps K_{in} by:

$$K_{out} = K_{in} + K_{middle} \cdot PA \tag{3}$$

where K_{in} is the input feature maps to RPAB block, K_{out} is the output from RPAB block which is a rescaled feature maps by PA block. K_{middle} is the input feature maps to PA block.

3.4. Feature Fusion

As the network depth increases, each RAG in the proposed network will extract different levels of features from the input. Taking full advantage of these different levels of features will improve the quality of the reconstructed hyperspectral image.

According to MSRN [17], in order to make full use of the different level of feature maps, we concatenated the features generated by each RAG through skip connection to form the Concat layer, as shown in Figure 1.

When reconstructing hyperspectral images, stacking RAGs allows complex parts to be reconstructed through deeper paths and simple parts to be reconstructed through shallower paths, which can effectively improve the performance of spectral reconstruction. These feature maps contain different levels of information, but there is also a lot of redundant information. In order to make better use of these features and reduce the computational complexity, we use 1×1 convolution layer to reduce the feature-map dimension.Input RGB image, the output obtained by the proposed network can be expressed as

$$y = g(f(x) + C(F_1(f(x)), \cdots, F_m(\cdots F_1(f(x))))) \tag{4}$$

where C denotes the 1×1 convolution layer behind the Concat layer

4. Experiment

4.1. Dataset and Metrics

Dataset In this study, we used the data set provided by the NTIRE 2020 Spectral Reconstruction Challenge to train the proposed RPAN network.

The NTIRE 2020 Spectral Reconstruction Challenge dataset [5] (Arad Hyperspectral Database, ARAD HSDB) provides 450 hyperspectral images as the training set, 10 hyperspectral images as the validation set, and 10 hyperspectral images as the test set. In order to meet the needs of the challenge, a large number of hyperspectral images have been newly collected, and these hyperspectral images contain various scenes with very different contents.

The challenge consists of two parts: (1) The RGB projection function of the track Clean is known and no noise or compression is applied to the RGB images. (2) The RGB projection function of the track Real World is unknown, the JPG compression and noise are applied to the RGB images.

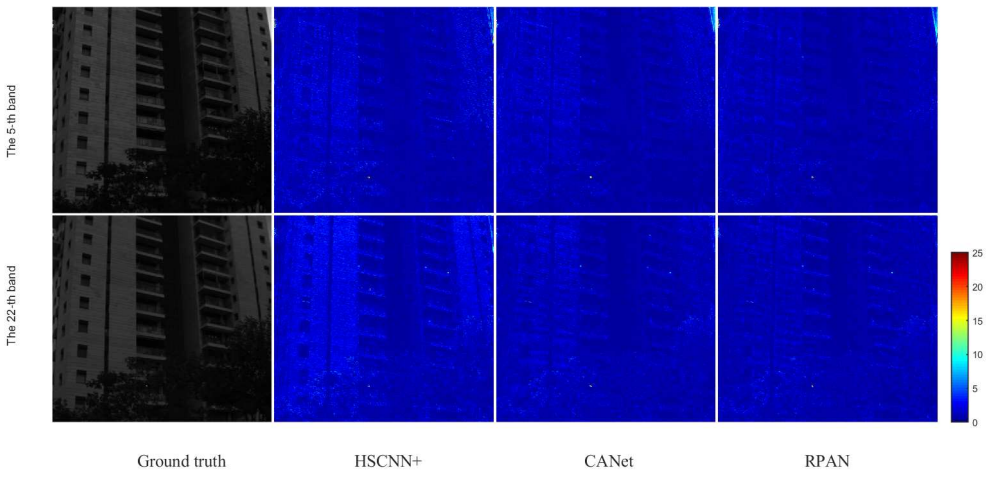

Figure 3. Visual error map of two selected bands(the 5-th and 22-th band) from track Clean images.

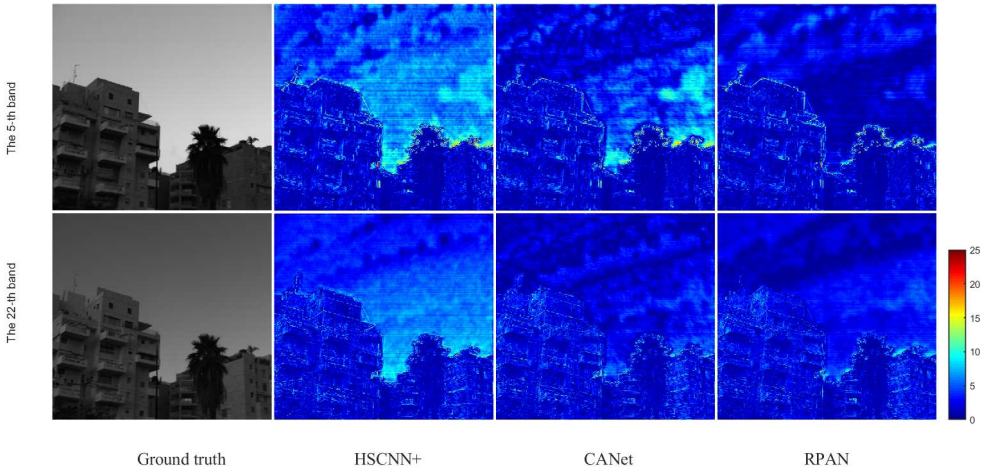

Figure 4. Visual error map of two selected bands(the 5-th and 22-th band) from track Real World images.

The RGB image and the corresponding hyperspectral image have the same resolution of 482×512. Each Hyperspectral image consists of 31 successive spectral bands. In the validation set, each track provides 10 RGB images and corresponding hyperspectral images. The test set only provides RGB images, so we can only use the validation set to evaluate the proposed method. The test result is used by challenge organizers to evaluate the results of each participants.

Metrics We followed the instructions of the NTIRE 2020 Spectral Reconstruction Challenge and evaluated the performance of our proposed method on two tracks ("Clean" and "Real World"). We use the scoring script provided by the challenge, and Mean Relative Absolute Error (MRAE) and Root Mean Square Error (RMSE) as the evaluation method. In general, a smaller RMSE or MRAE represent higher quality. MRAE and RMSE are computed as follows:

$$MRAE = \frac{1}{N}\sum_{i=1}^{N}(\frac{|P_{RE} - P_G|}{P_G}) \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((P_{RE} - P_G)^2)} \tag{6}$$

Where N is the total number of pixels, P_{RE} and P_G represent the corresponding pixels of the reconstructed and ground truth hyperspectral images respectively.

4.2. Training Detail

In RPAN model, the number of RAGs is set to 4, each RAG includes 8 RPAB blocks. In PA block, the reduction

track	Resnet-based HSCNN+ MRAE/RMSE	CANet MRAE/RMSE	RPAN(ours) MRAE/RMSE
Clean	0.03999/4.490	**0.03739**/4.422	0.03756/**4.301**
RealWorld	0.07174/5.517	0.06877/5.081	**0.06787/4.984**

Table 1. The quantitative results of different methods. The best results are in bold.

ratio r is 16. Except the last convolutional layer of the network, there are 31 filters to output 31 spectral band images and the first 1×1 convolutional layer in the PA module contains 4 filters, the remaining convolutional layers all are 64 filters. Therefore, the Concat layer which stacked 4 RAGs contains 256 channel feature maps, and then returns to 64 channel feature maps through a 1×1 convolutional layer. All 3×3 convolutional layers use a zero-padding to keep the feature map size unchanged.

The network settings used for each track are slightly different. In track Clean, because no noise or compression is applied to the RGB images, we remove the biases in the convolutional layer. In track Real World, due to the noise and JPG compression is applied to the RGB images, we retain the biases in the convolutional layer to compensate for the noise and JPG compression.

During the training, the ground truth hyperspectral images and ground-truth RGB images are cropped into small pieces of 64×64 from the training dataset and the batchsize is set to 16. Adam optimizer [16] is used for optimizing the proposed RPAN network with $\beta_1 = 0.9$, $\beta_2 = 0.999$,$\epsilon = 10^{-8}$ and the weight decay was set to 10^{-6}. The initial learning rate is set to 8×10^{-5}, the learning rate decays by 0.8 after every 5 epochs, the network is ended the optimization at the 50-th epoch. MRAE loss function is used for training the RPAN network.

When reconstructing the hyperspectral image of the validation set, we no longer segment the RGB image into small blocks, but instead use the complete RGB image as input to get a complete spectral image. The proposed RPAN is trained by Pytorch [23] platform on a single NVIDIA Titan Xp GPU. It takes about 13 hours to train the RPAN network for each track.

4.3. Comparison with State-of-the-Art

Comparison Methods We compare RPAN with the previous state-of-the-art method HSCNN+ to illustrate that our proposed method has superior performance in spectral reconstruction tasks. HSCNN + has two methods based on Resnet and Densenet, the Resnet-based HSCNN+ is used to compare with our RPAN method in this paper.

To illustrate the advantages of the PA module over CA, we added a comparison method CANet. In CANet, we use the CA block to replace the PA block in RPAN, and the rest is exactly the same as RPAN. Through this comparison, it will be shown whether the PA block is more robust than the CA block, and whether PA makes the network have better performance.

Comparison Results

Under the same training settings, the three methods were trained on the NTIRE 2020 Spectral reconstruction Challenge dataset. We use the official validation set to compare the three methods on both Clean and Real World tracks. The quantitative results are reported in Table 1.

It can be seen that on the two tracks, the proposed RPAN and CANet are noticeably better than Resnet-based HSCNN+ method on NTIRE 2020 validation set. When compared with Resnet-based HSCNN+, RPAN can significantly reduce the MRAE/RMSE results by as much as 0.00243/0.189 in the track Clean. In the track Real World, RPAN also significantly reduce the MRAE/RMSE results by as much as 0.00387/0.533. The results prove that our proposed RPAN method and PA module greatly improve the effect of spectral reconstruction compared with Resnet-based HSCNN+.

In the track Clean, RPAN achieves the better results on the RMSE, which decrease 0.121 than CANet. Although the results of CANet is reduced by 0.00017 compared to the RPAN method on the MRAE, the proposed RPAN is more robust than CANet, because the PA module can adaptively adjust pixel-wise features in all channels, while CA can only adaptively adjust channel-wise features.

To illustrate that RPAN is more robust than CANet, we conducted noise interference experiment in the track Clean. Adding different degrees of White Gaussian Noise to the clean RGB image, and then used these RGB with noise for hyperspectral image reconstruction. The results are reported in Table 2.

track Clean	CANet MRAE/RMSE	RPAN(ours) MRAE/RMSE
without noise	**0.03739**/4.422	0.03756/**4.301**
SNR 40dB	0.04983/4.757	**0.04115/4.327**
SNR 35dB	0.07487/5.871	**0.05161/4.732**
SNR 30dB	0.11950/7.029	**0.083051/5.670**
SNR 25dB	0.20315/9.537	**0.15532/7.952**

Table 2. The quantitative results of CANet and RPAN at different noise levels in track Clean, i.e., SNR=25, 30, 35, 40dB.The best results are in bold.

In Table 2, the proposed RPAN method is noticeably better than CANet method When different degrees noises are

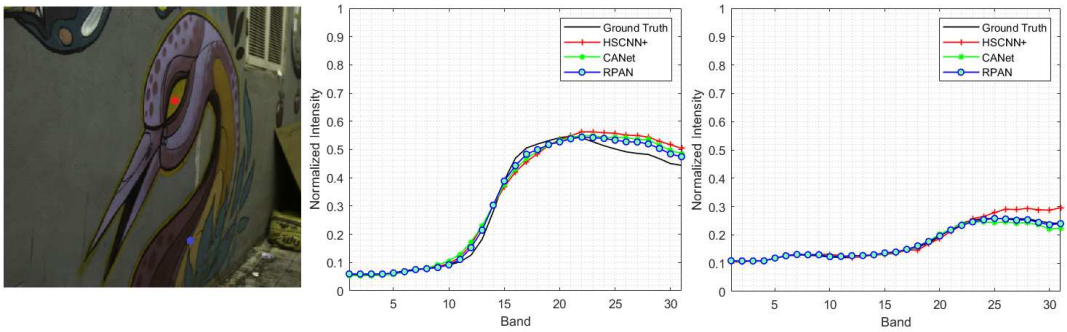

Figure 5. Recovered spectral curve in track Clean. (a) is the spectral curve of the red dot, (b) is the spectral curve of the blue dot.

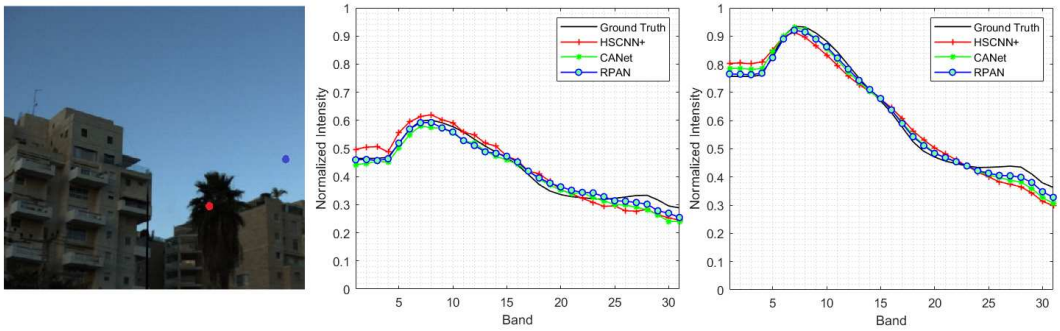

Figure 6. Recovered spectral curve in track Real World. (a) is the spectral curve of the red dot, (b) is the spectral curve of the blue dot.

added to the RGB image. The results prove that RPAN is more robust than CANet. Despite using pure RGB for spectral reconstruction training, RPAN still exhibits good performance when using RGB with noise for spectral reconstruction. CANet has been more affected, and the results of spectral reconstruction have dropped significantly.

In the track Real World, the RGB projection function of the track Real World is unknown and the noise and JPG compression is applied to the RGB picture. The proposed RPAN method is significantly better than CANet. The RPAN achieves the better results on the MRAE and RMSE, which decrease 0.0009 and 0.097 than CANet, as shown in Table 1. The result further illustrate that RPAN method is more suitable for spectral reconstruction tasks from poor quality RGB.

Beside numerical results, we have also drawn two example of reconstructed spectral image error maps to better illustrate the results. Figure 3 and Figure 4 correspond to the track Clean and the track Real World respectively. The error of the RPAN method is smaller than the other methods on both tracks. We also plotted the reconstructed spectrum curves of these methods in Figure 5 and Figure 6, we can see that the spectral curve reconstructed by RPAN is more closer to the ground truth.

Track	RPAN (without Concat layer) MRAE/RMSE	RPAN MRAE/RMSE
Clean	0.03806/4.353	**0.03756/4.301**
Real World	0.0705/5.200	**0.06787/4.984**

Table 3. Effect of the Concat layer. The best results are in bold.

4.4. Other Discuss

In this part, we conducted ablation studies on the NTIRE 2020 dataset to verify the effect of the Concat layer on hyperspectral reconstruction. The quantitative results on the validation set are shown in Table 3. Compared with RPAN without Concat layer, RPAN with Concat layer performs better. This demonstrate that the Concat layer really helps to make better use of different levels of features and improve the accuracy of hyperspectral reconstruction.

5. Conclusion

In this paper, we propose a residual pixel attention network (RPAN) for the NTIRE 2020 Spectral Reconstruction Challenge. It uses a novel attention module PA, which can adaptively rescale the pixel-wise features in all feature maps, enhance the weight of more important features, sup-

press unimportant information. It significantly improves the performance of the network. By comparing with the state-of-art method HSCNN+ on the NTIRE 2020 spectral reconstruction dataset, it proved that the RPAN method significantly improves the accuracy of hyperspectral reconstruction from a single RGB image.

References

[1] Naveed Akhtar and Ajmal Mian. Hyperspectral recovery from rgb images using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):100–113, 2020. 1

[2] Aitor Alvarezgila, Joost Van De Weijer, and Estibaliz Garrote. Adversarial networks for spatial context-aware spectral image reconstruction from rgb. pages 480–490, 2017. 1, 2

[3] Boaz Arad, Ohad Ben-Shahar, and Radu Timofte. Ntire 2018 challenge on spectral reconstruction from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 929–938, 2018. 2

[4] Boaz Arad and Ohad Benshahar. Sparse recovery of hyperspectral signal from natural rgb images. pages 19–34, 2016. 1, 2

[5] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, Graham Finlayson, et al. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 4

[6] Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, and Tat-Seng Chua. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5659–5667, 2017. 1

[7] Yi Chen, Nasser M Nasrabadi, and Trac D Tran. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3973–3985, 2011. 1

[8] Michael R Descour and Eustace L Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 34(22):4817–4826, 1995. 2

[9] Michael T Eismann, Joseph Meola, and Russell C Hardie. Hyperspectral change detection in the presenceof diurnal and seasonal variations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1):237–249, 2008. 1

[10] Silvano Galliani, Charis Lanaras, Dimitrios Marmanis, Emmanuel P Baltsavias, and Konrad Schindler. Learned spectral super-resolution. *arXiv: Computer Vision and Pattern Recognition*, 2017. 1, 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. pages 770–778, 2016. 3

[12] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2019. 1, 4

[13] Yan Jia, Yinqiang Zheng, Lin Gu, Art Subpaasa, Antony Lam, Yoichi Sato, and Imari Sato. From rgb to spectrum for natural scenes via manifold-based mapping. pages 4715–4723, 2017. 1

[14] Zohaib Khan, Faisal Shafait, and Ajmal Mian. Automatic ink mismatch detection for forensic document analysis. *Pattern Recognition*, 48(11):3615–3626, 2015. 1

[15] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. pages 1646–1654, 2016. 3

[16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2015. 6

[17] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. pages 527–542, 2018. 4

[18] Kin Gwn Lore, Kishore Reddy, Michael Giering, and Edgar A Bernal. Generative adversarial networks for spectral super-resolution and bidirectional rgb-to-multispectral mapping. pages 0–0, 2019. 1

[19] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. pages 807–814, 2010. 4

[20] Rang M H Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. pages 186–201, 2014. 2

[21] Shijie Nie, Lin Gu, Yinqiang Zheng, Antony Lam, Nobutaka Ono, and Imari Sato. Deeply learned filter response functions for hyperspectral reconstruction. pages 4767–4776, 2018. 1

[22] Seoung Wug Oh, Michael S Brown, Marc Pollefeys, and Seon Joo Kim. Do it yourself hyperspectral imaging with everyday digital cameras. pages 2461–2469, 2016. 2

[23] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 6

[24] Wallace M. Porter and Harry T. Enmark. A System Overview Of The Airborne Visible/Infrared Imaging Spectrometer (Aviris). In Gregg Vane, editor, *Imaging Spectroscopy II*, volume 0834, pages 22 – 31. International Society for Optics and Photonics, SPIE, 1987. 2

[25] R Glenn Sellar and Glenn D Boreman. Classification of imaging spectrometers for remote sensing applications. *Optical Engineering*, 44(1):013602, 2005. 2

[26] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. pages 939–947, 2018. 1, 2

[27] Tarek Stiebei, Simon Koppers, Philipp Seltsam, and Dorit Merhof. Reconstructing spectral images from rgb-images using a convolutional neural network. pages 948–953, 2018. 1

[28] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*, pages 111–126. Springer, 2014. 2

[29] Hien Van Nguyen, Amit Banerjee, and Rama Chellappa. Tracking via object reflectance using a hyperspectral video camera. pages 44–51, 2010. 1

[30] Ashwin A Wagadarikar, Renu John, Rebecca Willett, and David J Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10), 2008. 2

[31] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2017. 1

[32] Tao Wang, Zhigang Zhu, and E Blasch. Bio-inspired adaptive hyperspectral imaging for real-time target tracking. *IEEE Sensors Journal*, 10(3):647–654, 2010. 1

[33] Jiqing Wu, Jonas Aeschbacher, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. pages 471–479, 2017. 1, 2

[34] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, and Dong Liu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. pages 518–525, 10 2017. 2

[35] Xiang Xu, Jun Li, Xin Huang, Mauro Dalla Mura, and Antonio Plaza. Multiple morphological component analysis based decomposition for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):3083–3102, 2016. 1

[36] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. pages 294–310, 2018. 1, 4