Hierarchical Regression Network for Spectral Reconstruction from RGB Images

Yuzhi Zhao *[1], Lai-Man Po[1], Qiong Yan[2], Wei Liu[2,3] , Tingyu Lin[1]

[1]City University of Hong Kong, Hong Kong SAR, China
[2]SenseTime Research
[3]Harbin Institute of Technology, China

Abstract

Capturing visual image with a hyperspectral camera has been successfully applied to many areas due to its narrow-band imaging technology. Hyperspectral reconstruction from RGB images denotes a reverse process of hyperspectral imaging by discovering an inverse response function. Current works mainly map RGB images directly to corresponding spectrum but do not consider context information explicitly. Moreover, the use of encoder-decoder pair in current algorithms leads to loss of information. To address these problems, we propose a 4-level Hierarchical Regression Network (HRNet) with PixelShuffle layer as inter-level interaction. Furthermore, we adopt a residual dense block to remove artifacts of real world RGB images and a residual global block to build attention mechanism for enlarging perceptive field. We evaluate proposed HRNet with other architectures and techniques by participating in NTIRE 2020 Challenge on Spectral Reconstruction from RGB Images. The HRNet is the winning method of track 2 - real world images and ranks 3rd on track 1 - clean images.

1. Introduction

Hyperspectral (HS) imaging technology refers to the spectral signature is densely sampled to many narrow bands. It combines imaging technology with spectral technology to detect the two-dimensional geometric space and one-dimensional spectral information of the target to obtain continuous, narrow-band images with high spectral resolution. Normally, most of the civil cameras capture only three primary colors. However, HS spectrometers can obtain the spectrum of each pixel in the scene and collect the information into a set of images. To visualize HS images, a response function is adopted to transform HS images into RGB format. Conversely, we can acquire HS images from the visible format by learning the inverse function. In this paper, we propose a general hierarchical regression network (HRNet) for spectral reconstruction from RGB images.

HS imaging technology has many advantages and particular characteristics. There have been many applications based on HS imaging technology, e.g, remote sensing technology [25], pedestrian detection [17, 23], food processing [29], medical imaging [2]. However, in recent years, the development of HS imaging has encountered a bottleneck since it mainly depends on spectrometers. The traditional spectrometers saves images with huge volume and need long operation time, which restricts HS imaging technology applied to portable platforms and high-speed moving scenes [28]. Although researchers have continuously optimized the traditional pipeline [7, 35], these hardware devices are still expensive and of high complexity. Thus, we present a low cost and automate approach only based on RGB cameras. To address the problem, we propose a HRNet that learns the process of RGB images to corresponding HS projections.

In general, spectral reconstruction is an ill-posed problem. Moreover, there is unknown noise in environment leading to degraded RGB images. However, there is dense correspondence between RGB images and HS images, making it possible to exploit the correlation from many RGB-HS pairs. Since the information of RGB image is much less than HS image, there may be many reasonable HS image combinations corresponding to a same RGB image. The algorithm needs to learn a reasonable mapping function that produces high-quality HS images. With the development of deep convolutional neural network (CNN), it is eligible to learn the blind mapping for spectral construction.

The previous methods [32, 21, 33, 6, 36] mainly utilize an auto-encoder structure with residual blocks [14]. The network often performs convolution at low spatial resolution since the features are more compact and the computation is more efficient. However, as the network goes deeper, it fails to remain the original pixel information due to per-

*Corresponding author: yzzhao2-c@my.cityu.edu.hk forming down-sampling by convolutions. To address this problem, we introduce a lossless and learnable sampling operator PixelShuffle [31]. To further boost the quality of generated images, we propose a hierarchical architecture that extracts the features of different scales. At each level, the input is obtained by the reverse PixelShuffle (PixelUnShuffle) that no pixel is lost. Moreover, we propose to use residual dense block and residual global block in HRNet for removing artifacts and noise and modelling remote pixel correlation, respectively.

In general, there are three main contributions of this paper:

(1) We propose a HRNet that utilizes PixelUnShuffle and Pixelshuffle layers for downsampling and upsampling without information loss. We also propose residual dense block with residual global block to enlarge perceptive field and boost generation quality;

(2) We propose a 8-setting ensemble strategy to further enhance the generalization of HRNet;

(3) We evaluate proposed HRNet on NTIRE 2020 HS dataset. The HRNet is winning method of track 2 - real world images and ranks 3rd on track 1 - clean images.

2. Related work

Hyperspectral image acquisition. Conventional methods for hyperspectral image acquisition often adopt spectrograph with spatial scanning or spectral scanning technology. There are several types of scanner utilized for capturing images including pushbroom scanner, whiskbroom scanner, and band sequential scanner. They have been widely used to many applications such as detector, environmental monitoring and remote sensor for decades. For instance, pushbroom scanner and whiskbroom scanner are used for photogrammetric and remote sensing by satellite sensors [28, 5]. However, those devices need to capture the spectral information of single points or bands separately, then scan the whole scene to get a fully HS image, which is difficult to capture scenes with moving objects. In addition, they are too large physically and not suitable for portable platforms. In order to address the problems, many kinds of non-scanning spectrometers have been developed to adapt the application of dynamic scenes [10, 7, 35].

Hyperspectral image reconstruction from RGB images. Since the traditional methods for hyperspectral image acquisition are not portable or time-consuming for many applications, current methods attempt to reconstruct hyperspectral image from RGB image. By learning the mapping from RGB images to hyperspectral images on a big RGB-HS dataset, it is more convenient to obtain many HS images. Recent years have witnessed various studies including sparse coding and deep learning. In 2008, Parmar et al. [27] proposed a data sparsity expanding method to recover the spatial spectral data cube. Arad et al. [3] first leveraged HS prior in order to create a sparse dictionary of HS signatures and their corresponding RGB projections. While Aeschbacher et al. [1] pushed the performance of Arad et al.'s method for better accuracy and runtime based on A+ framework [34].

Beyond the dataset provided by Arad et al. [3], many approaches proposed their own dataset. For instance, Yasuma et al. [37] utilized a CCD camera (Apogee Alta U260) to captured 31-band multispectral images (400–700 nm, at 10 nm intervals) of several static scenes. Nguyen et al. [26] captured a dataset by Specim's PFD-CL-65-V10E (400 nm to 1000 nm) spectral camera and there were total 64 images. Chakrabarti et al. [8] explored a statistical model based on 55 HS images of indoor and outdoor scenes. With the improvement of the scale and resolution of natural HS dataset, the training of deep learning method becomes more feasible, a number of algorithms based on convolutional neural network were proposed [21, 33]. Simon et al. [20] proposed a fully convolutional densely connected "Tiramisu" network with one hundred layers for semantic segmentation. Galliani et al. [11] enhanced it for spectral image super-resolution. Can et al. [6] improved it to avoid overfitting to the training data and obtain faster inference speed. Moreover, Xiong et al. [36] proposed a unified HSCNN framework for hyperspectral recovery from both RGB and compressive measurements. To boost the performance, they developed a deep residual network named HSCNN-R, and another distinct architecture that replaces the residual block by the dense block with a novel fusion scheme, named HSCNN-D, collectively called HSCNN+ [32].

Convolutional neural networks. The convolutional neural networks have been successfully applied in many low-level vision tasks, e.g. colorization [39, 19], inpainting [18, 38], deblurring [22], denoising [13, 9], and demosaicking [9, 40]. Hyperspectral reconstruction, as one of low-level task, has gained great improvement of performance recently by deep convolutional neural networks. In order to facilitate convergence and extract features effectively, many well-known basic blocks are utilized in those frameworks such as residual block and dense block. He et al. [14] proposed a residual network initially for image classification. It improves the accuracy obviously compared with traditional cascade convolutional structure. Then, the residual block has been widely used in image enhancement region for maintaining low-level features by the short connection. It was enhanced by densenet proposed by Huang et al. [16] to improve the feature fusion ability. Moreover, Hu et al. [15] strengthened them by a squeeze-and-excitation network including a feature attention mechanism. It was implemented by MLP layers for modelling connections of pixels in different spatial location. In general, our HRNet combines the advantages of above methods and provides a more effective and accurate solution for HS reconstruction.

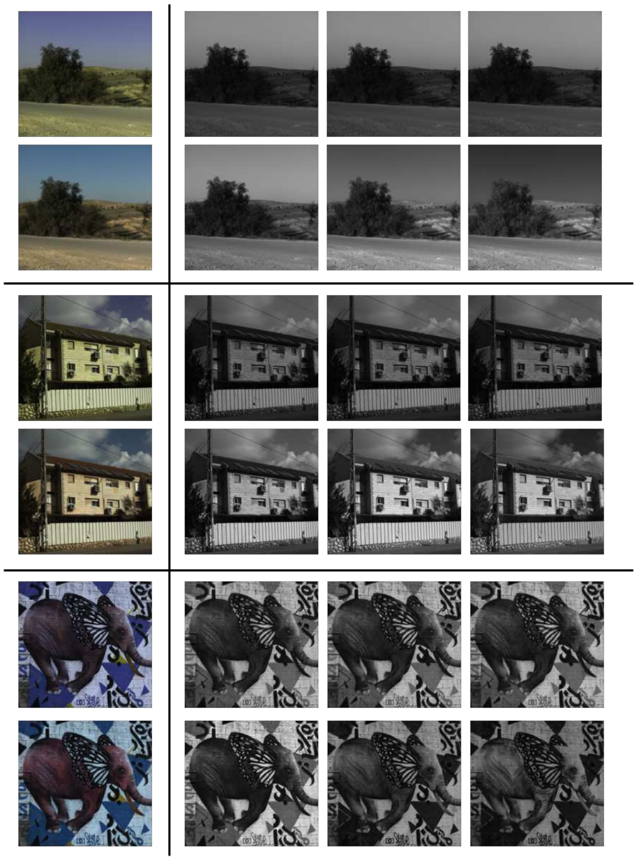

Figure 1. Visualization of NTIRE 2020 HS dataset. For each group, from top to bottom and left to right, they represent clean RGB images, real world RGB images, HS images with 400 nm, 410 nm, 420 nm, 500 nm, 600 nm, and 700 nm channels, respectively.

3. Methodology

3.1. Dataset

We train our approach on the HS dataset provided by NTIRE Challenge 2020. This dataset consists of three parts: spectral images, clean RGB images (for track 1) and real world RGB images (for track 2). There are overall 450 RGB-HS pairs in training for both tracks involving different scenes. Each spectral image has the information of 31 bands in range of 400 nm to 700 nm. It is of 482×512 spatial resolution. To generate its corresponding RGB image, there is a fixed response function applied to HS bands. The rendering process can be defined as:

$$RGB = HS \times ResponseFunc. \tag{1}$$

The RGB images and HS images include 3 and 31 channels, respectively. The $ResponseFunc$ maps each HS band to visible channel R, G, and B by 93 parameters. For clean RGB images, they are constructed by a known response function and saved as uncompressed format. However, the real world RGB images are acquired by unknown response function with additional blind noise and demosaicking operation. Some examples are illustrated in Figure 1 (e.g. 1st band approximately covers the 395-405 nm range).

3.2. HRNet architecture

Generally, we propose a 4-level network architecture for high-quality spectral reconstruction from RGB images, as shown in Figure 2. The PixelUnShuffle layers [31] are utilized to downsample the input to each level without adding parameters. Therefore, the number of pixels of input is fixed while the spatial resolution decreases. Conversely, the learnable PixelShuffle layers are adopted to upsample feature maps and reduce channels for inter-level connection. The PixelShuffle only reshapes feature maps and does not introduces interpolation like bilinear upsampling. It allows the network to learn upsampling operation adaptively.

For each level, the process is decomposed to inter-level integration, artifacts reduction, and global feature extraction. For inter-level learning, the output features of subordinate level are pixel shuffled, then concatenated to current level, finally processed by an additional convolutional layer to unify channel number. In order to effectively reduce artifacts, we adopt residual dense block [14, 16], containing 5 dense-connected convolutional layers and a residual. Moreover, the residual global block [14, 15] with short-cut connection of input is used to extract attention for every remote pixels by MLP layers.

Since the features are most compact in bottom level, there is a 1×1 convolutional layer attached to the last of bottom level in order to enhance tone mapping by weighting all channels. The two mid levels process features at different scales. Moreover, the top level uses the most blocks to effectively integrate features and reduce artifacts thus produce high-quality spectral images. The illustration of these blocks are in Figure 3.

3.3. Implementation details

We only use L1 loss in the training process, which is a PSNR-oriented optimization for the system. The L1 loss is defined as:

$$L_1 = \mathbb{E}[||G(x) - y||_1], \tag{2}$$

where x and y are input and output, respectively. The $G(*)$ is the proposed HRNet. Note that, we utilize the local patches for efficient training. The input RGB image and output spectral images are cropped in same spatial region.

For network architecture, all the layers are LeakyReLU [24] activated except output layer. We do not use any normalization in HRNet to maintain the data distribution. The reflect padding is adopted for each convolutional layer in order to reduce border effect. The weights of VCGAN are initialized by Xavier algorithm [12].

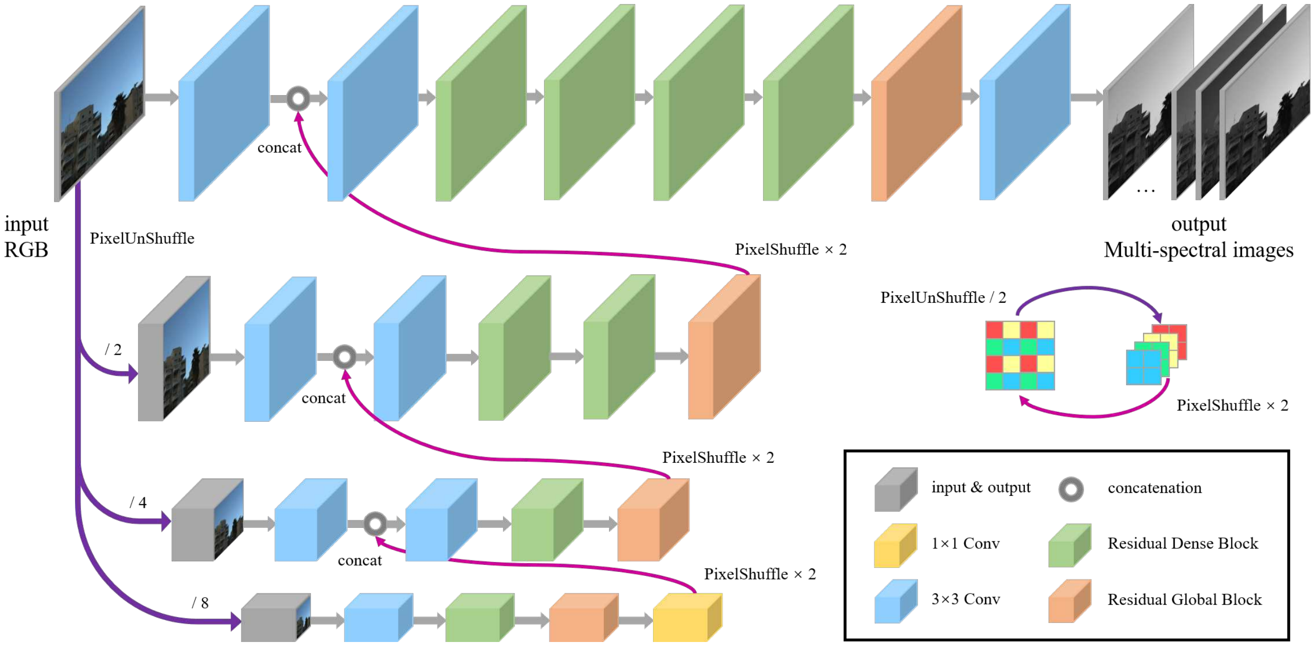

Figure 2. Illustration of the architecture of HRNet. Please visit the project web page `https://github.com/zhaoyuzhi/Hierarchical-Regression-Network-for-Spectral-Reconstruction-from-RGB-Images` to try our codes and pre-trained models.

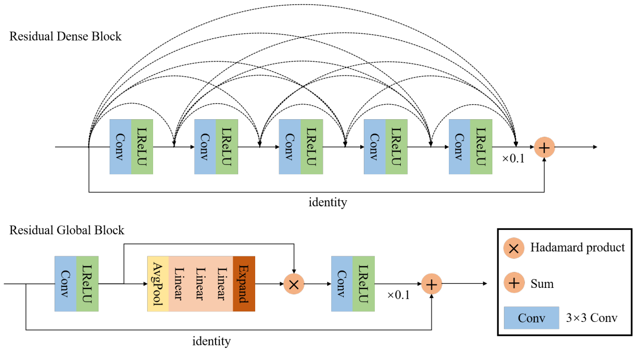

Figure 3. Illustration of the architecture of residual dense block (ResDB) and residual global block (ResGB).

For training details, we use the entire NTIRE 2020 HS dataset (450 HS-RGB pairs for both tracks) at training. The whole HRNet is trained for 10000 epochs overall. The initial learning rate is 1×10^{-4} and halved every 3000 epochs. For optimization, we use Adam optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and batch size equals to 8. The image pairs are randomly cropped to 256×256 region and normalized to range [0, 1]. All the experiments are implemented using 2 NVIDIA Titan Xp GPUs. It takes approximately 7 days for whole training process.

3.4. Ensemble strategy

Since the solution space of spectral reconstruction is often large, there may be multiple settings that achieve same performance on the training set. Therefore, a single network may lead to poor generalization performance since it tends to fall into local minima. However, we can minimize this risk by combining multiple network settings to enhance generalization and fuse the knowledge. In order to perform ensemble strategy, we use 4 other hyper-parameter settings and train HRNet from scratch for both tracks. These settings can be summarized as:

- Re-train the HRNet using baseline training setting.
- Exchange the position of residual dense block and residual global block in HRNet, and use baseline training setting.
- Train the network with different batch size (2 or 4) and keep other hyper-parameter settings, network architecture.
- Train the network with different cropping patch size (320×320 or 384×384) and keep other hyper-parameter settings, network architecture.

Therefore, there are 8 kinds of training methods. All the methods used for ensemble are trained for 10000 epochs. We record the MRAE (mean absolute value between all bands of generated spectral images $G(x)$ and ground truth y) every 1000 epochs, as shown in Table 1 and Figure 4. Finally, we utilize the epoch with best MRAE value of 8 methods for computing average.

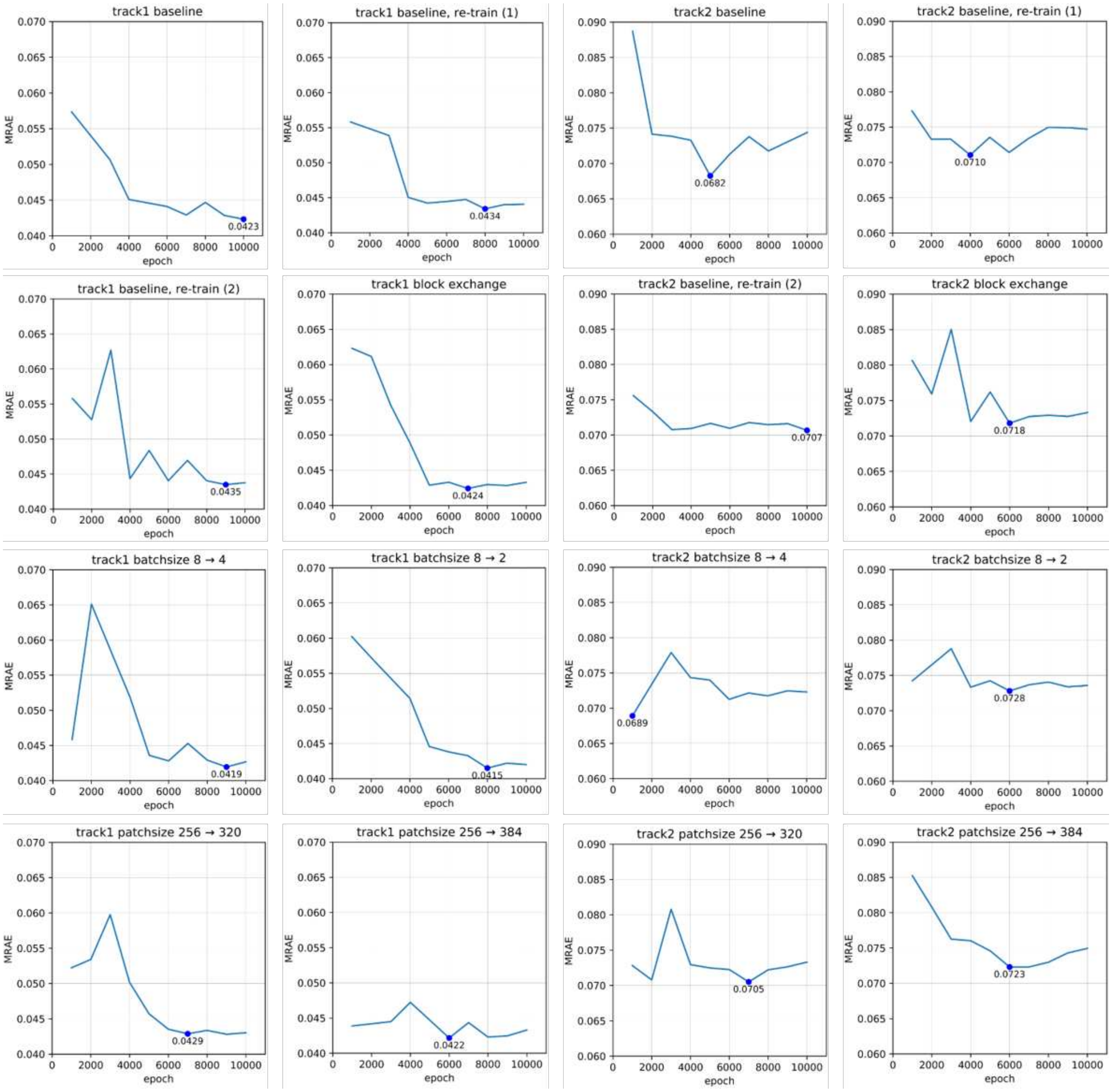

Figure 4. The MRAE between ground truth spectral images and the generated images of different hyper-parameter settings for ensemble.

4. Experiment

4.1. Experimental settings

We evaluate proposed HRNet by comparing with other network architectures and conducting ablation study on NTIRE 2020 HS dataset. For each track, there are 10 validation RGB images. The evaluation metrics are defined as:

- **MRAE.** It computes the pixel-wise disparity (mean absolute value) between all bands of generated spectral images $G(x)$ and ground truth y. It explicitly represents the construction quality of network. It is defined as:

$$MRAE = \frac{1}{N}\sum_{i=1}^{N}\frac{|G(x)^i - y^i|}{y^i}, \tag{3}$$

where N denotes the overall pixels of spectral images.

- **RMSE.** It computes the root mean square error be-

Setting	track 1	track 2
Baseline	0.042328	0.068245
Re-train baseline (1st)	0.043408	0.071044
Re-train baseline (2nd)	0.043487	0.070668
Exchange position of blocks	0.042418	0.071798
Change batch size 8 to 4	0.041936	0.071259
Change batch size 8 to 2	0.041507	0.072797
Change patch size 256 to 320	0.042810	0.070502
Change patch size 256 to 384	0.042166	0.072313
Ensemble	**0.039893**	**0.068081**

Table 1. The best MRAE value of both tracks for HRNet settings used for ensemble.

Method		U-Net	U-ResNet	HRNet
MRAE	track 1	0.047507	0.045242	**0.042328**
	track 2	0.074230	0.078892	**0.068245**
RMSE	track 1	0.014154	0.013927	**0.013537**
	track 2	0.018647	0.020630	**0.017859**
BPMRAE	track 1	0.007926	0.007171	**0.006064**
	track 2	0.044966	0.055876	**0.042105**

Table 2. The quantitative comparison results of different architectures and HRNet on NTIRE 2020 HS validation set.

tween the generated and ground truth spectral images with 31 bands. It is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(G(x)^i - y^i)^2}. \tag{4}$$

- **Back Projection MRAE (BPMRAE).** It evaluates the colorimetric accuracy of recovered RGB images from the generated and ground truth spectral images by a fixed camera response function. It is defined as:

$$BPMRAE = \frac{1}{N}\sum_{i=1}^{N}\frac{|(R \times G(x))^i - (R \times y)^i|}{y^i}, \tag{5}$$

where R denotes the function $ResponseFunc$.

4.2. Comparison with other architectures

We utilize two common network architectures for comparison: U-Net [30] and U-ResNet [30, 14]. Both of them have been widely used in many previous low-level tasks [19, 18, 38, 22, 9, 40]. The first convolutional layer and last convolutional layer utilize 7×7 convolution without changing spatial resolution. The training scheme for all methods are same. Other details are concluded as: (1) U-Net. The encoder layers perform convolution with stride of 2. The spatial resolution of bottom feature map equals to 1×1.

Method	w/o ResDB	w/o ResGB	w/o both	HRNet
MRAE	0.042448	0.042565	0.048033	**0.042328**
RMSE	0.014216	0.014092	0.015740	**0.013537**
BPMRAE	0.009507	0.007669	0.015502	**0.006064**

Table 3. The comparison results of ablation study on NTIRE 2020 HS validation set track 1 - clean images.

Method	HRNet ($\frac{1}{2}$)	HRNet ($\frac{1}{4}$)	HRNet ($\frac{1}{8}$)
MACs (G)	46.413	12.017	3.212
Params (Mb)	8.185	2.176	0.6088
Weights (Mb)	32.006	8.532	2.410
MRAE	0.042457	0.046424	0.048443
RMSE	0.015147	0.015459	0.015659
BPMRAE	0.006886	0.007806	0.009891

Table 4. The comparison results of compressed HRNet model (the number of channels decreased to $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{8}$ of the original) on NTIRE 2020 HS validation set track 1 - clean images.

There are short concatenations between each encoder layer and decoder layer with same resolution; (2) U-ResNet. The total number of encoder layers and decoder layers are half of U-Net. Instead, there are 4 residual blocks attached to the last layer of encoder. The concatenations are reserved.

We train both networks using same hyper-parameters of HRNet until convergence. There is no ensemble strategy used. We generate the reconstructed spectral images using the best epoch of them. The results are summarized in Table 2. We also visualize each method in Figure 5 and 6 by pseudo-color map. The first three rows show the data distribution of 3 methods and last row indicates ground truth. We recommend readers to compare textures of background.

There are two reasons that proposed HRNet outperforms other two methods. The first is that HRNet utilizes PixelShuffle to connect each level. Traditional nearest or bilinear upsampling will introduce redundancy information to features, which is unnecessary for feature extraction. However, by the combination of PixelUnShuffle and PixelShuffle, HRNet could process high-level features more efficiently. The second is that HRNet adopts two residual-based blocks, which facilitate convergence and assist each level to exploit different scales of features. Moreover, the blocks with residual learning helps remove artifacts. The residual global block enhances context information since it models correlation for every two pixels.

4.3. Ablation study

In order to demonstrate the effectiveness of both residual dense block (ResDB) and residual global block (ResGB), we replace them by plain convolution layers with similar FLOPs. The results in track 1 - clean images is shown in Table 3. The baseline of HRNet is shown in Table 1, which

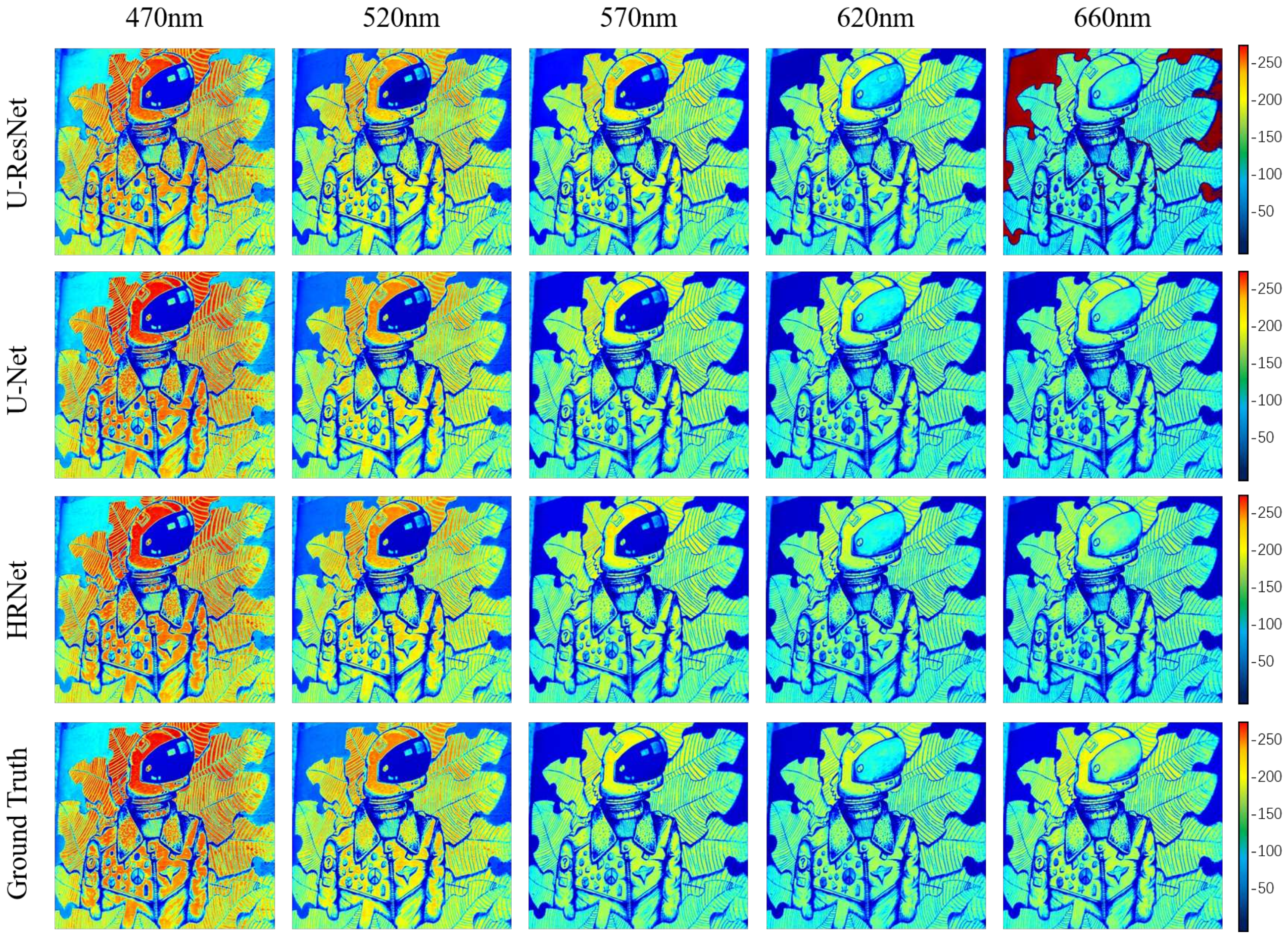

Figure 5. Visualization of generated results from U-ResNet, U-Net, and proposed HRNet on NTIRE 2020 HS validation set track 1.

Team	MRAE	Runtime / Image (seconds)	Compute Platform
Deep-imagelab	0.03010476377	0.56	2×NVIDIA 2080Ti
ppplang	0.03075687151	16	NVIDIA 1080Ti
HRNet	**0.03231183605**	3.748	2×NVIDIA Titan Xp
ZHU_ zy	0.03475963089	1	*Unknown*
sunnyvick	0.03516495956	0.7	Tesla K80 12GB

Table 5. The final testing results of NTIRE 2020 Spectral Reconstruction from RGB Images Challenge track 1 - clean images.

has better performance comparing with all ablation settings. If we delete all ResDB or ResGB in HRNet, the MRAE decreases the most, which demonstrates the combination of both blocks is significant for spectral reconstruction.

We conduct another experiment that shrinks the HRNet model size by decreasing channels of each convolutional layer to half, one fourth, and one eighth of original numbers. It will compress model size greatly by sacrificing pixel fidelity. To better compare these settings, we conclude the multiply–accumulate operation (MACs), total network parameters (Params), model size saved on machine (Weights) and 3 quantitative metrics results in Table 4. The MACs, Params, and Weights of baseline HRNet are 182.347 Gb, 31.705 Mb, and 123.879 Mb, respectively. Users can choose high-quality HRNet to obtain high pixel fidelity of spectral images (MRAE = 0.042328) or high-efficiency HRNet with small size (Weights = 2.410 Mb).

4.4. Testing result on NTIRE 2020 challenge

The proposed HRNet ranks 3rd and 1st on track 1 and track 2, respectively, of NTIRE 2020 Spectral Reconstruction from RGB Images Challenge [4]. The comparison results on testing set are summarized in Table 5 and 6. Moreover, the HRNet has better performance on track 2

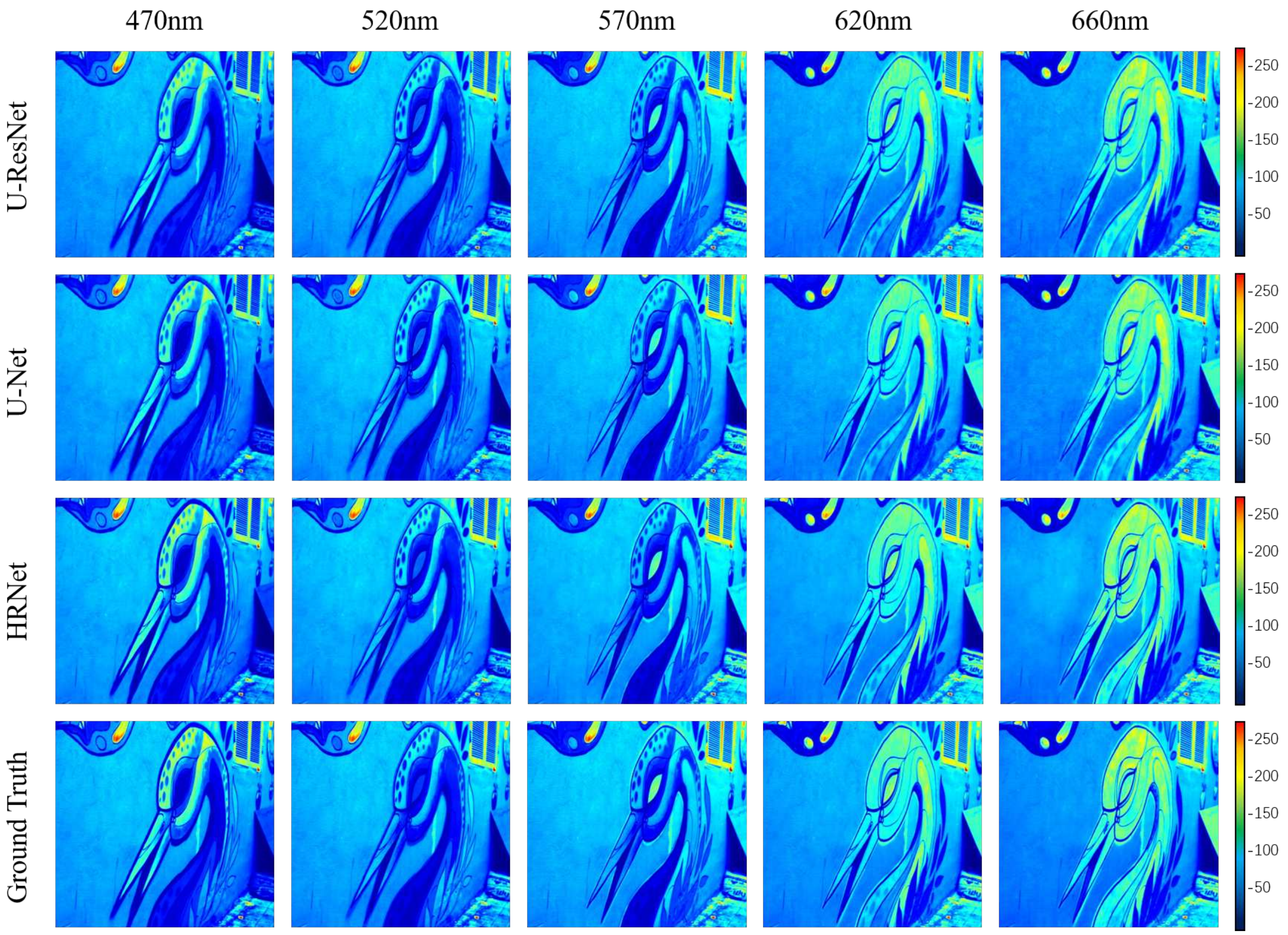

Figure 6. Visualization of generated results from U-ResNet, U-Net, and proposed HRNet on NTIRE 2020 HS validation set track 2.

Team	MRAE	Runtime / Image (seconds)	Compute Platform
HRNet	**0.06200744887**	3.748	2×NVIDIA Titan Xp
ppplang	0.06212710705	16	NVIDIA 1080Ti
Deep-imagelab	0.06216655487	0.56	2×NVIDIA 2080Ti
PARASITE	0.06514769779	30	NVIDIA Titan Xp
Tasti	0.06732598306	*Unknown*	NVIDIA 2080Ti

Table 6. The final testing results of NTIRE 2020 Spectral Reconstruction from RGB Images Challenge track 2 - real world images.

since it adopts two effective blocks for removing artifacts while utilizes learnable PixelShuffle upsampling operator. The ensemble strategy works obviously on both tracks that improves the MRAE from 0.042328 to 0.039893 since it avoids the HRNet to fall into local minima. In conclusion, both HRNet architecture and ensemble strategy contribute to spectral reconstruction performance.

5. Conclusion

In this paper, we presented a 4-level HRNet for automatically generating spectrum from RGB images. For each level, it adopts both residual dense block and residual global block for effectively extracting features. While the PixelShuffle is utilized for inter-level connection. Then, we proposed a novel 8-setting ensemble strategy to further enhance the quality of predicted spectral images. Finally, we validated the HRNet outperforms the well-known low-level vision frameworks such as U-Net and U-ResNet on NTIRE 2020 HS dataset. Furthermore, we presented 3 types of compressed HRNets and analyzed their reconstruction performance and computing efficiency. The proposed HRNet is the winning method of track 2 - real world images and ranks 3rd on track 1 - clean images.

References

[1] Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 471–479, 2017.

[2] Pstefan Andersson, Sune Montan, and Sune Svanberg. Multispectral system for medical fluorescence imaging. *IEEE Journal of Quantum Electronics*, 23(10):1798–1805, 1987.

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Proceedings of the European Conference on Computer Vision*, pages 19–34. Springer, 2016.

[4] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020.

[5] Michael Breuer and Jörg Albertz. Geometric correction of airborne whiskbroom scanner imagery using hybrid auxiliary data. *International Archives of Photogrammetry and Remote Sensing*, 33(B3/1; PART 3):93–100, 2000.

[6] Yigit Baran Can and Radu Timofte. An efficient cnn for spectral reconstruction from rgb images. *arXiv preprint arXiv:1804.04647*, 2018.

[7] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prism-mask system for multispectral video acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2423–2435, 2011.

[8] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–200. IEEE, 2011.

[9] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

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

[11] Silvano Galliani, Charis Lanaras, Dimitrios Marmanis, Emmanuel Baltsavias, and Konrad Schindler. Learned spectral super-resolution. *arXiv preprint arXiv:1703.09470*, 2017.

[12] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[13] Shuhang Gu, Yawei Li, Luc Van Gool, and Timofte Radu. Self-guided network for fast image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2511–2520, 2019.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[15] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[16] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[17] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1037–1045, 2015.

[18] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017.

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

[20] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 11–19, 2017.

[21] Sriharsha Koundinya, Himanshu Sharma, Manoj Sharma, Avinash Upadhyay, Raunak Manekar, Rudrabha Mukhopadhyay, Abhijit Karmakar, and Santanu Chaudhury. 2d-3d cnn based architectures for spectral reconstruction from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 844–851, 2018.

[22] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.

[23] Jingjing Liu, Shaoting Zhang, Shu Wang, and Dimitris N Metaxas. Multispectral deep neural networks for pedestrian detection. *arXiv preprint arXiv:1611.02644*, 2016.

[24] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the International Conference on Machine Learning*, volume 30, page 3, 2013.

[25] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, 2004.

[26] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *Proceedings of the European Conference on Computer Vision*, pages 186–201. Springer, 2014.

[27] Manu Parmar, Steven Lansel, and Brian A Wandell. Spatio-spectral reconstruction of the multispectral datacube using sparse recovery. In *IEEE International Conference on Image Processing*, pages 473–476. IEEE, 2008.

[28] Daniela Poli and Thierry Toutin. Review of developments in geometric modelling for high resolution satellite pushbroom sensors. *The Photogrammetric Record*, 27(137):58–73, 2012.

[29] Jianwei Qin, Kuanglin Chao, Moon S Kim, Renfu Lu, and Thomas F Burks. Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, 118(2):157–171, 2013.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241. Springer, 2015.

[31] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016.

[32] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018.

[33] Tarek Stiebel, Simon Koppers, Philipp Seltsam, and Dorit Merhof. Reconstructing spectral images from rgb-images using a convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 948–953, 2018.

[34] 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.

[35] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2104–2111, 2016.

[36] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 518–525, 2017.

[37] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010.

[38] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4471–4480, 2019.

[39] Richard Zhang, Isola Phillip, and Efros Alexei A. Colorful image colorization. In *Proceedings of the European Conference on Computer Vision*, pages 649–666. Springer, 2016.

[40] Yuzhi Zhao, Lai-Man Po, Tiantian Zhang, Zongbang Liao, Xiang Shi, Yujia Zhang, Weifeng Ou, Pengfei Xian, Jingjing Xiong, Chang Zhou, et al. Saliency map-aided generative adversarial network for raw to rgb mapping. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3449–3457. IEEE, 2019.