# Supplementary Material: 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

## 1. The details of HRNet architecture

The proposed HRNet contains 4-level. The exact specification of each level is shown in the Table 1 where the parameter $l$ indicates the level (top level equals to 0 while bottom level is 3). We utilize the PixelUnShuffle layers [7] to downsample the input with scale of 2, 4 and 8 for three lower levels. Therefore, the input sizes for each level (from top to bottom) are $3\times256\times256$, $12\times128\times128$, $48\times64\times64$, and $192\times32\times32$. The output feature maps of lower level are pixel-shuffled and concatenated to a convolutional layer in the superior level.

For each level, except the inter-level integration, we use residual dense block (ResDB) [2, 4] for artifacts reduction and residual global block (ResGB) [2, 3] for global feature extraction. The input and output channels are related to current level, as shown in Table 1. Each residual global block contains 5 dense-connected convolutional layers and a residual. While each residual global block contains 2 convolutional layers and 3 MLP layers with a residual.

## 2. Experiment on NIR image colorization

The proposed Hierarchical Regression Network (HRNet) is initially utilized for spectral reconstruction. As the reverse task, the colorization for NIR image is significant for near infrared data visualization. To further demonstrate the advance of HRNet, we perform an experiment on KAIST multispectral pedestrian detection dataset [5]. The spectral images from the dataset contains the near infrared information but the channel equals to 1. It implies the information of different spectrum is compacted into one channel.

We randomly divide the training and validation images from KAIST dataset to form 19:1 ratio. The other experimental settings are unchanged compared with spectral reconstruction. However, the input and output data for each algorithm is replaced by NIR data and RGB images, respectively. The U-ResNet [1] NIR data colorization method and proposed HRNet (original size and each channel reduced to half) are included in this experiment.

Some colorized results are shown in Figure 1 and quantitative comparison results on validation set is summarized in Table 2. The PSNR and SSIM [8] are utilized for evaluation. The HRNet can still achieve fair performance across all the methods since it effectively extracts different scales of features by proposed 4-level architecture, residual dense block, and residual global block. We believe these designs will improve the NIR image colorization performance.

## 3. Additional results of spectral reconstruction from RGB images by HRNet

To better visualize the results from different methods, we show more generated samples. The Figure 2 and 3 illustrate all generated spectral bands for both tracks. Each band is represented by pesudo-color map. For different bands, the color characteristics are obviously distinct. Since the input data of track 2 is noisy, the generated results from real world RGB images are less sharper than track 1.

While the Figure 4, 5, 6, 7, 8 and 9 show the additional comparison results generated by U-Net [6], U-ResNet [6, 2], and proposed HRNet. The readers are encouraged to compare the texture information and background details. The proposed HRNet produces spectral images with more similar visual quality to ground truth than other two methods.

## References

[1] Amanda Berg, Jorgen Ahlberg, and Michael Felsberg. Generating visible spectrum images from thermal infrared. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1143–1152, 2018.

*Corresponding author: yzzhao2-c@my.cityu.edu.hk

| Stage | Number | Convolution | Input Feature Map Size | Output Feature Map Size |
|---|---|---|---|---|
| Input Conv | 1 | k3s1p1 | $(3 \times 4^{l}) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ |
| Concatenation | 1 | / | $(64 \times (2^{l} + 2^{l+1}/4) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ |
| RDB | 1-4 | k3s1p1 | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ |
| RGB | 1 | k3s1p1 | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ |
| TM Conv | 1 | k1s1p0 | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ |
| Output Conv | 1 | k3s1p1 | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ or 31 |
| PixelShuffle | 1 | / | $(64 \times 2^{l}) \times (h/2^{l}) \times (w/2^{l})$ | $(64 \times 2^{l}/4) \times (h/2^{l-1}) \times (w/2^{l-1})$ |

Table 1. Specification for each level. The tone mapping convolution ($TM\ Conv$) only exists in bottom level. If it is the top level, the final output channels should be 31 to match the number of bands. Otherwise, the output needs to be pixel-shuffled. For instance, the 3 × 256 × 256 represents channels × height × width. The $k3s1p1$ represents a convolutional layer with kernel size, stride, and padding number equal to 3, 1, and 1, respectively.

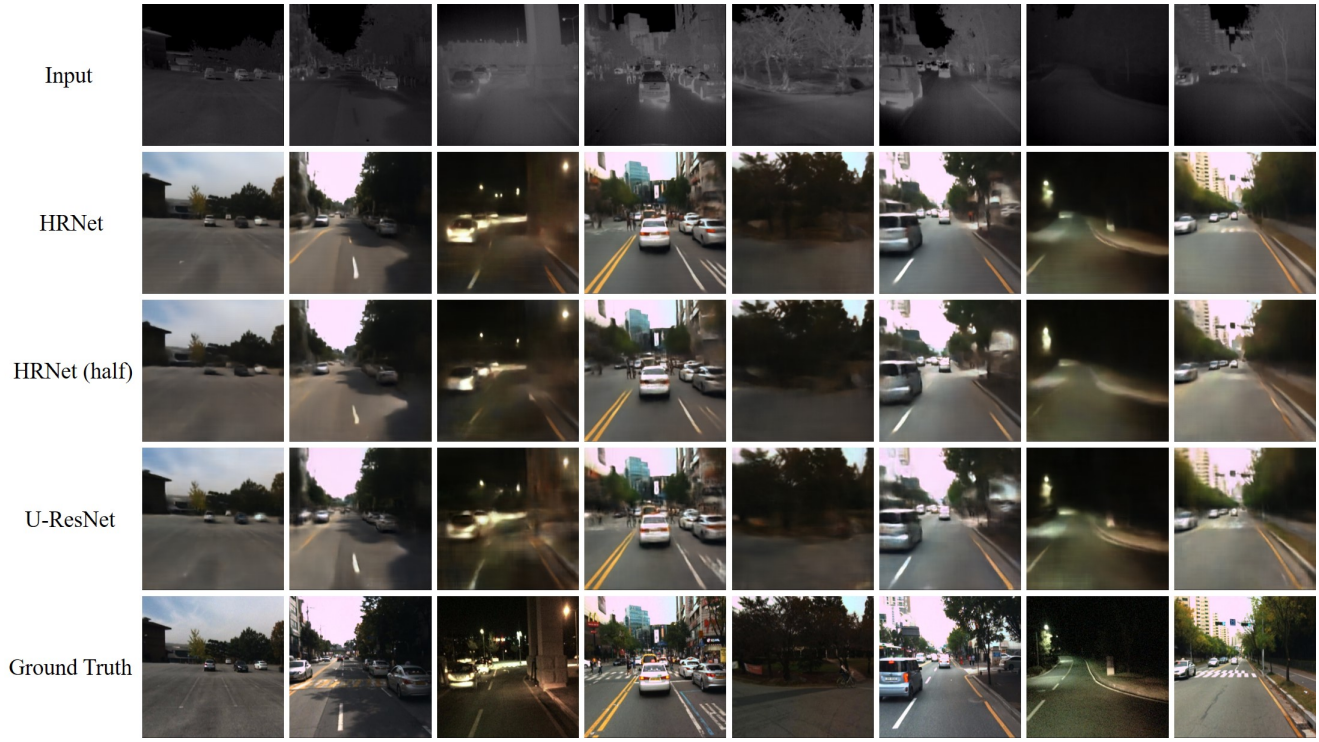


Figure 1. Visual comparison of the results generated by different architectures on KAIST multispectral pedestrian detection validation dataset.

| Method | HRNet | HRNet (half) | U-ResNet |
|---|---|---|---|
| PSNR | **23.28** | 21.88 | 23.12 |
| SSIM | **0.8177** | 0.7876 | 0.8126 |

Table 2. The comparison results of different architectures on KAIST multispectral pedestrian detection validation dataset.


[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

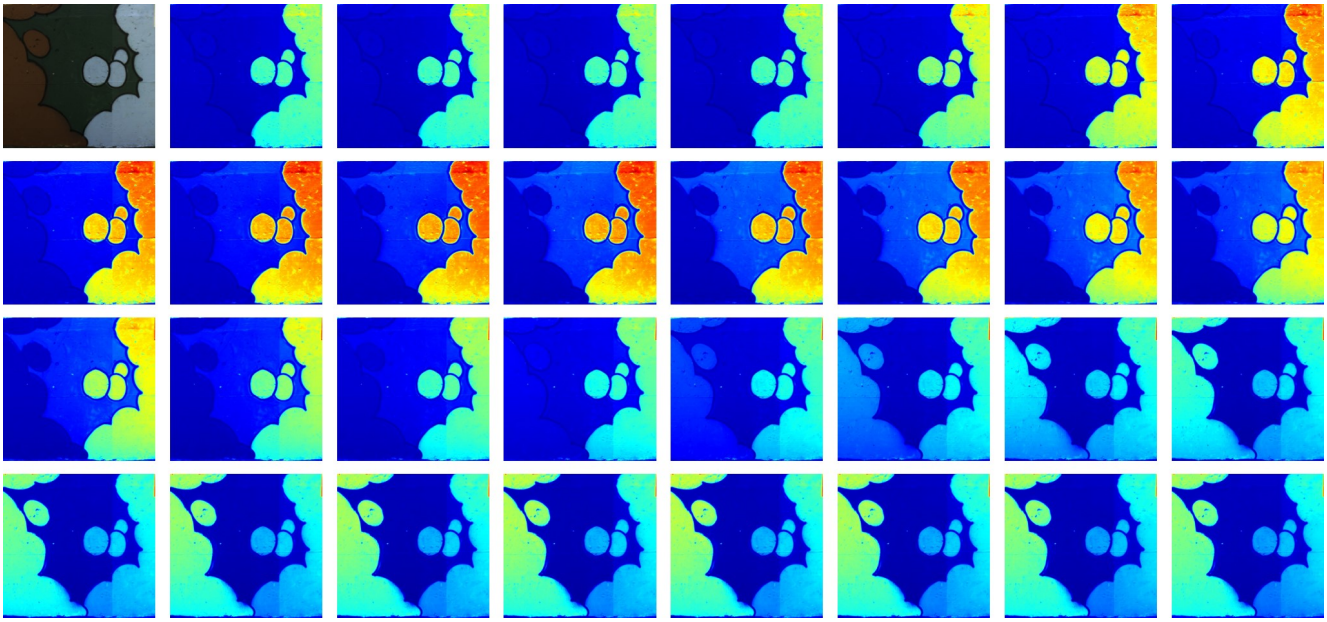

Figure 2. Visualization of input image from track 1 and all generated bands by HRNet. From left upper image to right bottom image, the results represent input clean RGB image and spectral image with bands from 400nm - 410 nm to 690nm - 700 nm.

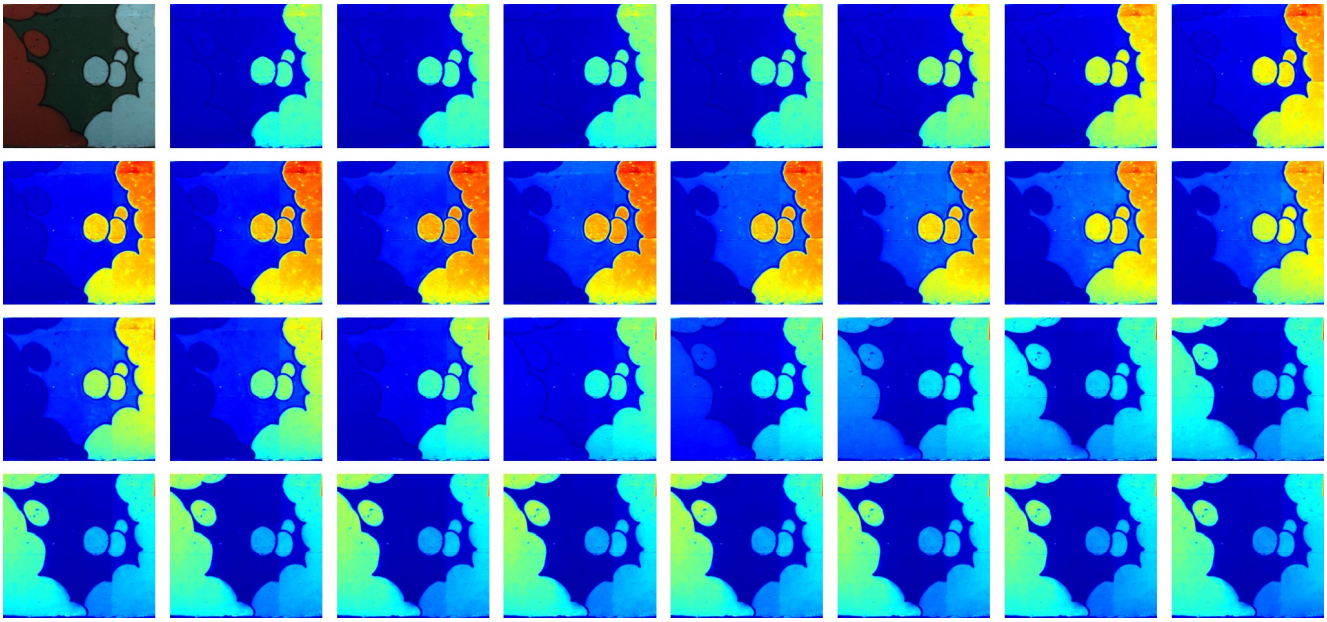

Figure 3. Visualization of input image from track 2 and all generated bands by HRNet. From left upper image to right bottom image, the results represent input real world RGB image and spectral image with bands from 400nm - 410 nm to 690nm - 700 nm.

Springer, 2015.

[7] 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.

[8] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

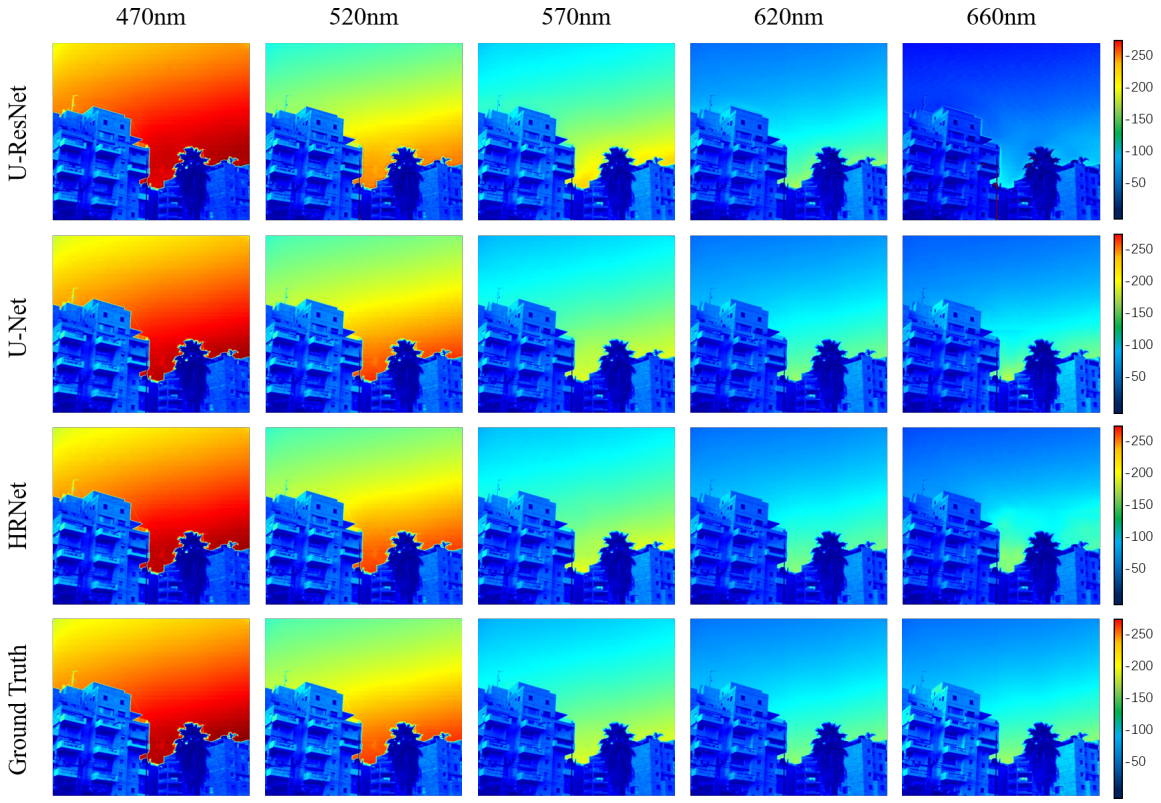


Figure 4. Visualization of additional generated result (1) from U-ResNet, U-Net, and proposed HRNet on track 1.

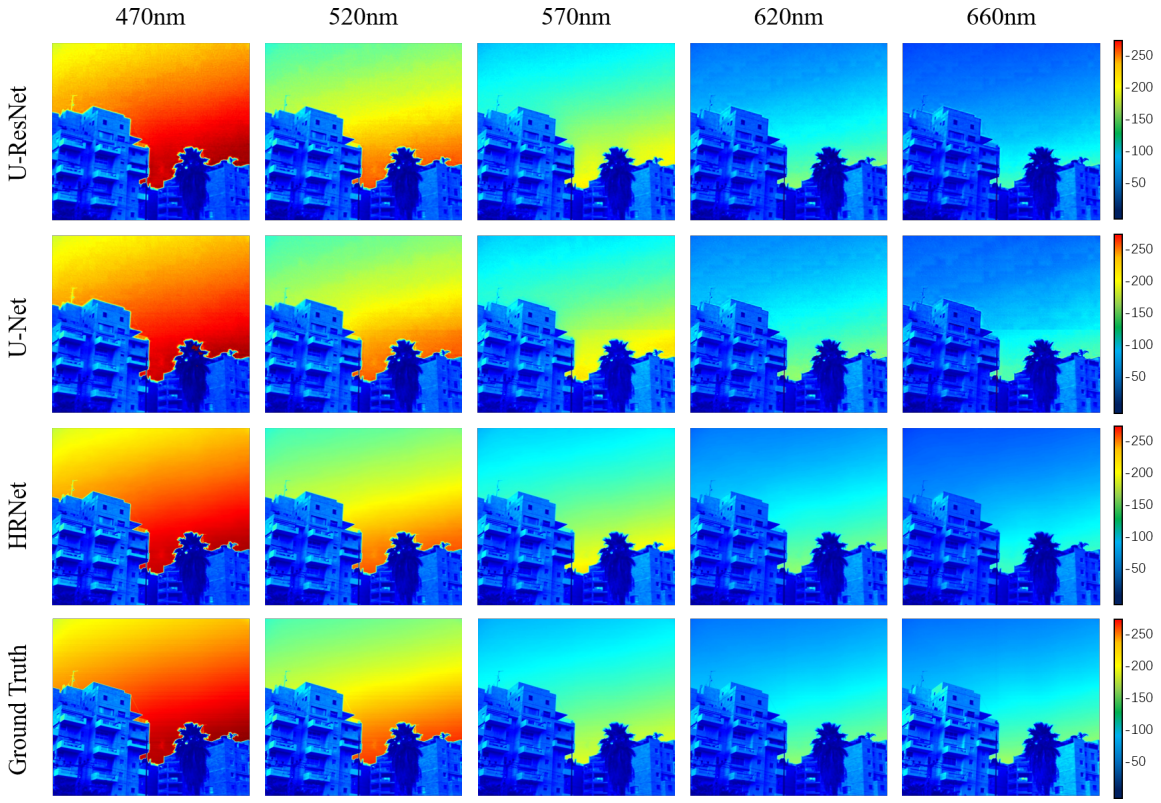


Figure 5. Visualization of additional generated result (1) from U-ResNet, U-Net, and proposed HRNet on track 2.

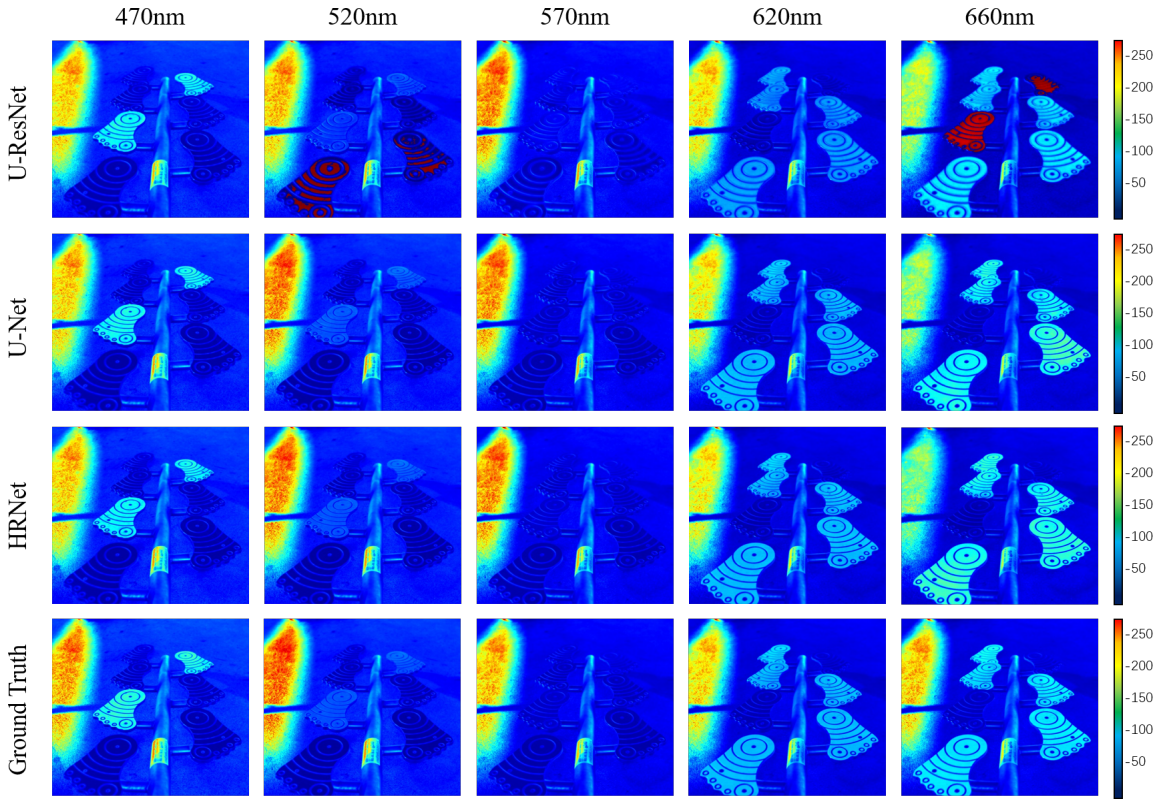


Figure 6. Visualization of additional generated result (2) from U-ResNet, U-Net, and proposed HRNet on track 1.

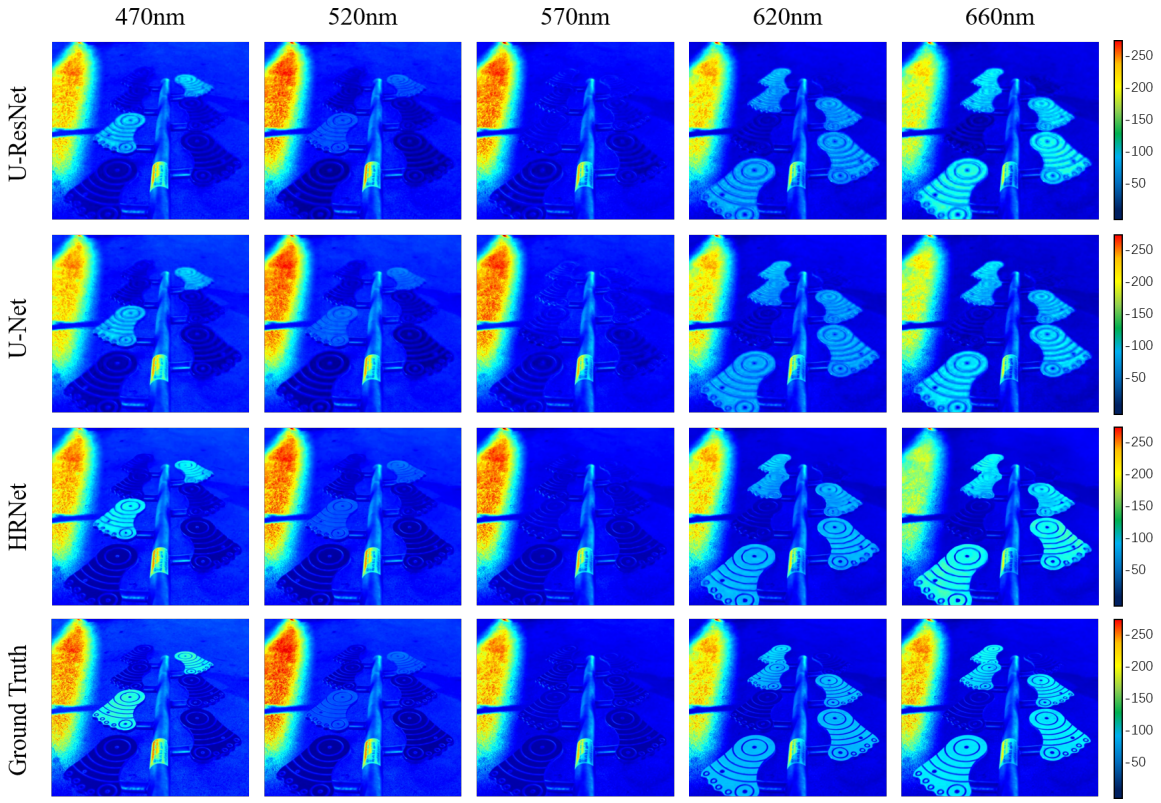


Figure 7. Visualization of additional generated result (2) from U-ResNet, U-Net, and proposed HRNet on track 2.

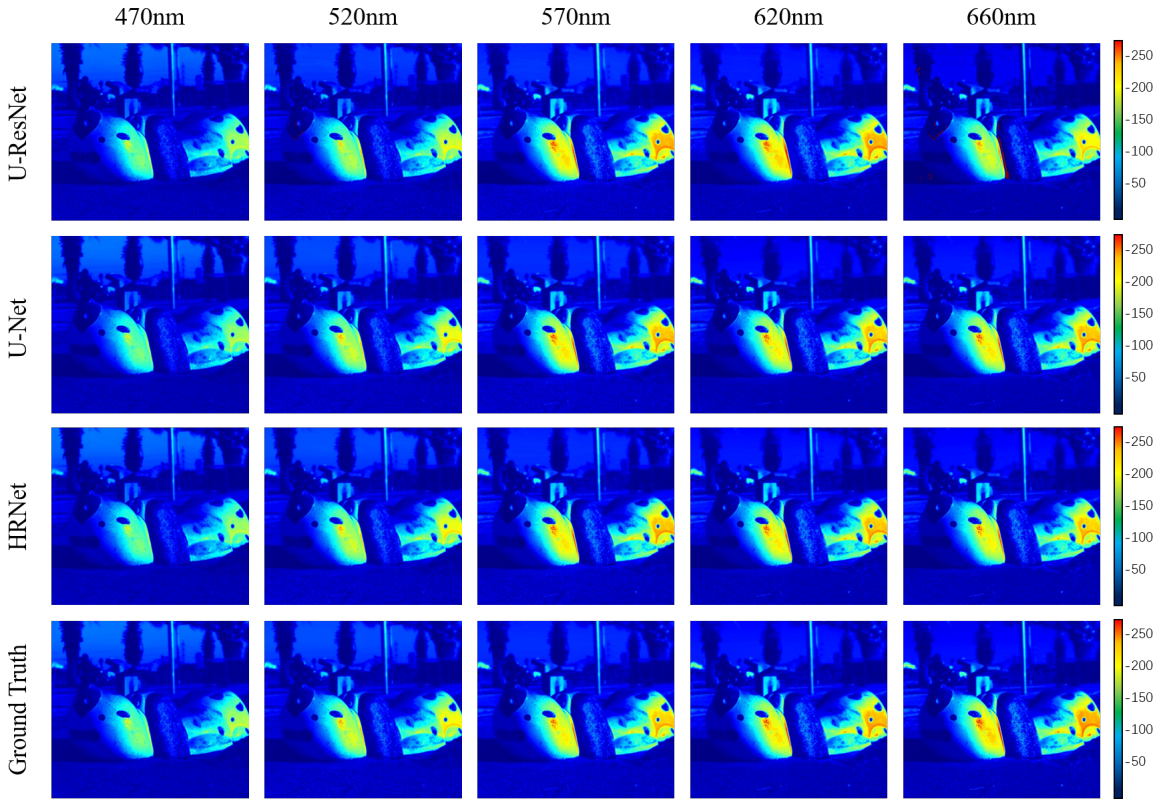


Figure 8. Visualization of additional generated result (3) from U-ResNet, U-Net, and proposed HRNet on track 1.

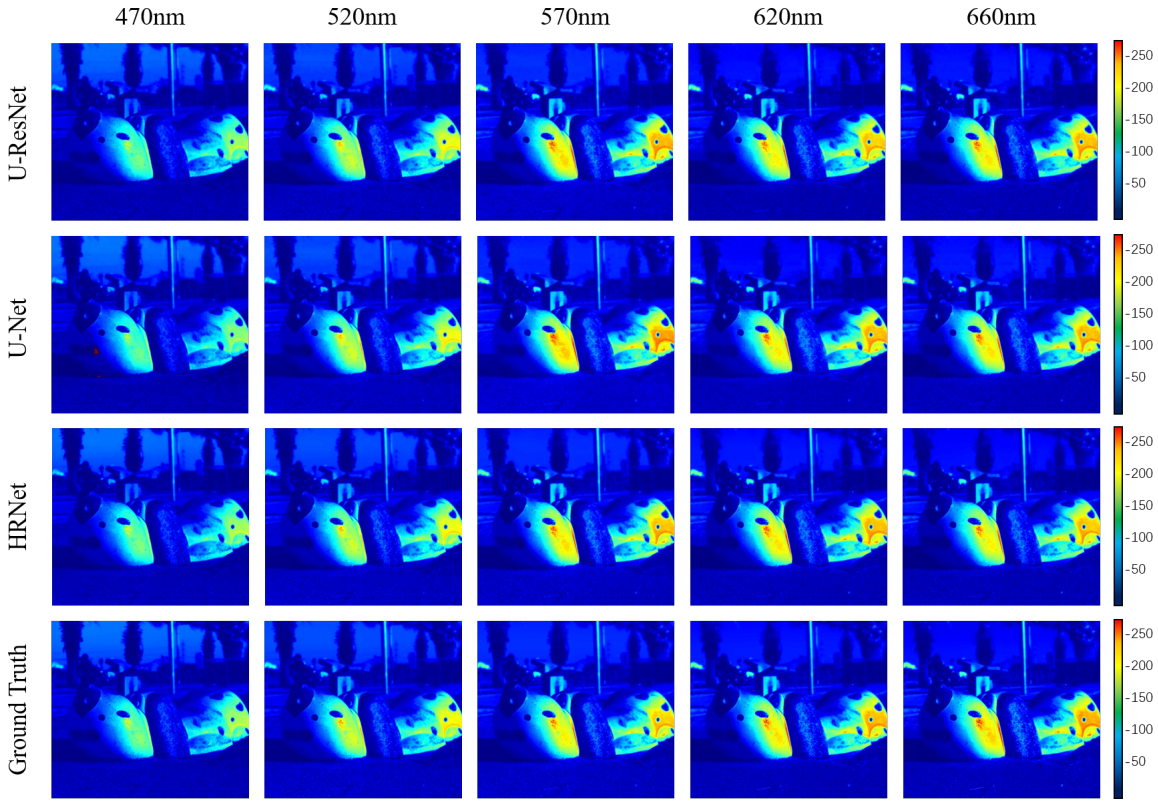


Figure 9. Visualization of additional generated result (3) from U-ResNet, U-Net, and proposed HRNet on track 2.