# DSR: Towards Drone Image Super-Resolution Supplementary Material

Xiaoyu Lin, Baran Ozaydin, Vidit Vidit, Majed El Helou, and Sabine Süsstrunk

School of Computer and Communication Sciences, EPFL, Switzerland
{xiaoyu.lin,baran.ozaydin,vidit.vidit,majed.elhelou,sabine.susstrunk}@epfl.ch

**Abstract.** We present in this supplementary material further experiments and information pointed out in the main manuscript. Namely, the experiments on resizing the matched field of view (FOV), the color correction methods, and the presentation of our training details.

## 1 Resizing the matched FOV

To resize the matched FOV, we try different interpolation methods (see the third to the fifth row in Fig 1 for more details). We calculate the normalized cross correlation between resized LR images and downsampled HR images and find the results are similar among all interpolation methods. This is because the new size of the LR image is very close to the original one, as described in the paper, with only a few pixels difference. Thus, different interpolation methods achieve similar results. Finally, we choose the nearest-neighbor interpolation method in this step, as this is the simplest one and changes the pixel values in the original image the least.

## 2 Color correction

We observe the color and luminance difference between HR and LR pairs (see Fig. 2 row (a) and (b) for some examples). We try different color correction methods separately and also combine them in a different order. We evaluate those methods on our full dataset by calculating the PSNR and SSIM between downsampled HR and the corrected LR counterparts. The numerical results are shown in Table 1 and some samples are shown in Fig. 2 row (c)-(f). Based on our results in Table 1, we choose color transfer and histogram matching applied in a sequential order as it achieves the best performance.

## 3 Training details

### 3.1 SwinIR

The architecture of the SwinIR network is the same as that in [3] for the real-world super-resolution task, except that the input LR image size and window

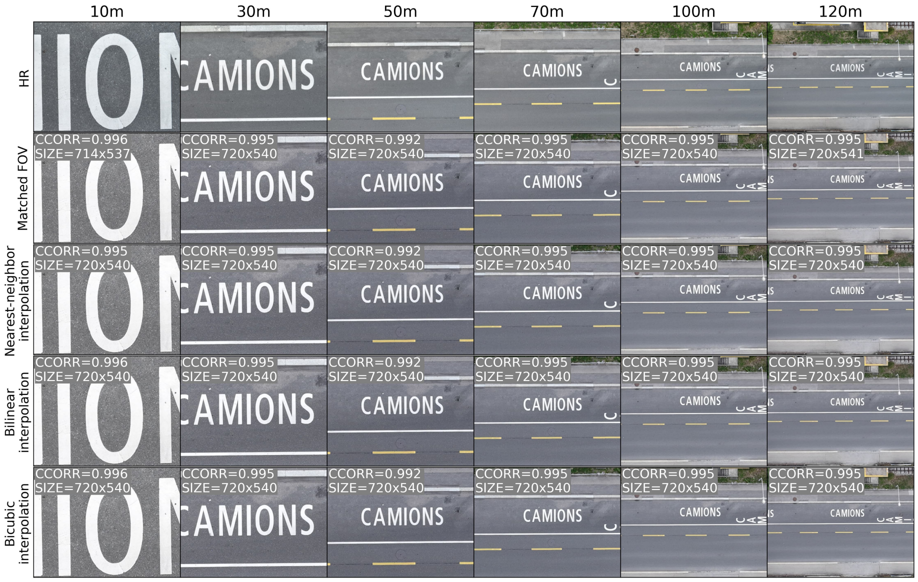

Fig. 1: FOV matching examples: the HR image is the full image captured by the Tele camera (4000×3000s resolution). The second row shows the matched FOV directly obtained from feature matching and homography. The rest row shows the resized FOV after different interpolation methods. The size of the matched FOV is shown in the top-left corner of the image. The normalized cross-correlation is calculated between matched FOV and downsampled HR image.

size are changed to $54 \times 54$ and 9 respectively to fit our scale factor. We first pretrain the network on the pubic Div2K [1] dataset with synthetic LR and HR pairs. We randomly crop the HR image into patches of size $300 \times 300$ (i.e. LR size of $54 \times 54$) with a batch size of 16 for training, and data augmentation performed through random rotation and flipping. We start with a learning rate of $1 \times 10^{-4}$, train 2000 epochs in total, and half the learning rate at [1000, 1600, 1800, 1900]. Following [4], we use the Adam optimizer [3] with $\beta_1 = 0.9$ and $\beta_2 = 0.99$.

For fine-tuning, we train the network on the DSR training set. We also randomly crop the HR image into patches of size $300 \times 300$ with a batch size of 16 for fine-tuning and also choose random rotation and flipping for data augmentation. The initial learning rate is $1 \times 10^{-5}$. We fine-tune 200 epochs and half the learning rate every 40 epochs. The remaining hyper-parameters are the same as those in the pretrained setup.

We use the same pretraining and fine-tuning strategies for our altitude-aware SwinIR.

| Method | $10m$ | $20m$ | $30m$ | $40m$ | $50m$ |
|---|---|---|---|---|---|
| Origin LR | 2.12/.7365 | 2.90/.7588 | 21.17/.7733 | 21.44/.7890 | 21.60/.7939 |
| Histogram Match(HM) | 23.74/.7598 | 24.59/.7814 | 25.04/.7959 | 25.35/.8079 | 25.39/.8109 |
| Color Transfer(CT) | 23.65/.7601 | 24.52/.7813 | 24.91/.7950 | 25.17/.8067 | 25.22/.8102 |
| HM + CT | 23.80/.7599 | 24.63/.7809 | 25.06/.7949 | 25.36/.8071 | 25.42/.8103 |
| CT + HM | 23.77/.7606 | 24.63/.7814 | 25.06/.7954 | 25.36/.8075 | 25.42/.8107 |

| Method | $70m$ | $80m$ | $100m$ | $120m$ | $140m$ |
|---|---|---|---|---|---|
| Origin LR | 21.52/.8004 | 21.54/.8025 | 21.67/.8114 | 21.99/.8233 | 22.14/.8269 |
| Histogram Match(HM) | 25.63/.8176 | 25.66/.8192 | 25.87/.8287 | 26.03/.8392 | 26.07/.8420 |
| Color Transfer(CT) | 25.47/.8169 | 25.50/.8184 | 25.66/.8276 | 25.79/.8370 | 25.83/.8401 |
| HM + CT | 25.67/.8171 | 25.70/.8185 | 25.92/.8285 | 26.08/.8386 | 26.12/.8417 |
| CT + HM | 25.68/.8174 | 25.71/.8186 | 25.93/.8290 | 26.09/.8387 | 26.13/.8420 |

Table 1: PSNR/SSIM between LR patches with different color correction methods and corresponding downsampled HR patches in our dataset. All results are calculated on the Y channel in the transformed YCbCr space. We highlight in red the best result.

### 3.2 FCNN

We also pretrain FCNN on the Div2K [1] dataset and apply the same image crop and data augmentation strategies as SwinIR. We start with a learning rate of $1 \times 10^{-4}$, train 1000 epochs in total, and half the learning rate every 200 epochs. We also use the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Next, we fine-tune 100 epochs in total and half the learning rate every 20 epochs. The remaining hyper-parameters are the same as those for pretraining. Finally, we also apply the same configuration to our altitude-aware FCNN.

## Acknowledgement

This paper is based on a master thesis [5] completed by Xiaoyu Lin, supervised by Dr. Majed El Helou and Prof. Sabine Süsstrunk and co-supervised by Baran Ozaydin and Vidit Vidit.

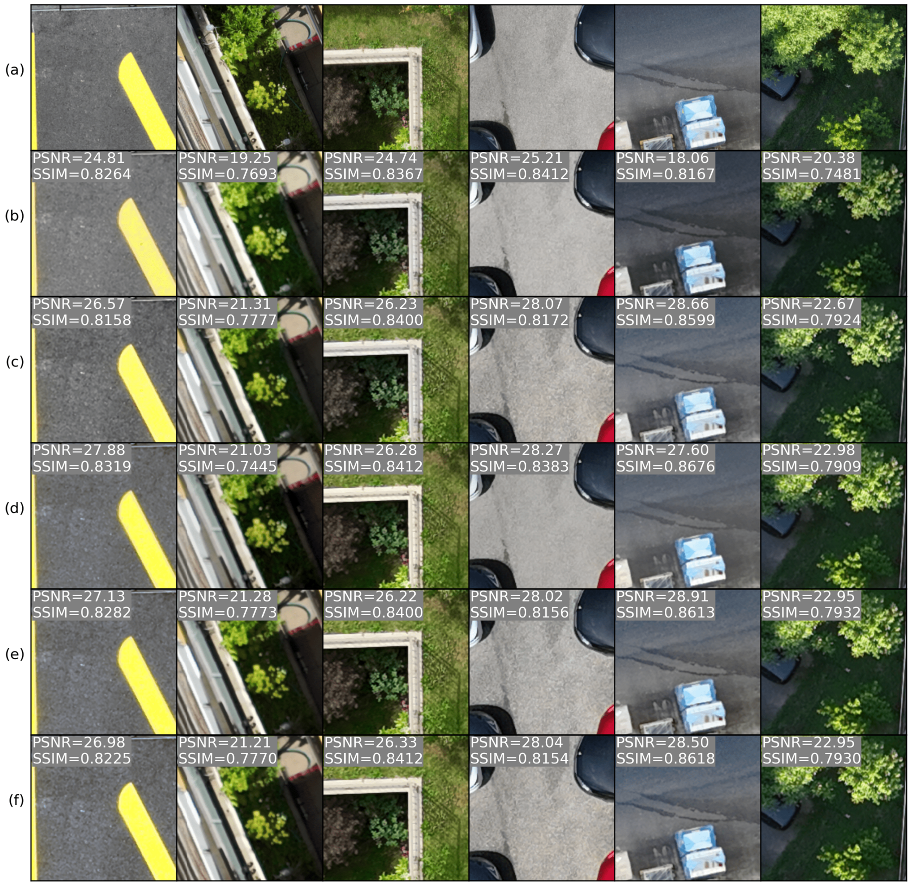

Fig. 2: Examples showing the performance of different methods for color correction. (a) Bicubic downsampled HR images, (b) LR images, (c) Histogram matching, (d) Color transform [2], (e) Histogram matching+Color transform, (f) Color transform+Histogram matching. All results are calculated on the Y channel in the transformed YCbCr space.

## References

1. Agustsson, E., Timofte, R.: NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 126–135 (2017) 2, 3
2. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: Proceedings of the IEEE International Conference on Computer Vision (2019) 4
3. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 1, 2
4. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: Image restoration using Swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1833–1844 (2021) 2
5. Lin, X.: Towards robust drone vision in the wild. arXiv preprint arXiv:2208.12655 (2022) 3