Supplementary Material

FHDe²Net: Full High Definition Demoireing Network

Bin He[1], Ce Wang[1], Boxin Shi[1,2,3], and Ling-Yu Duan[1,3*]

[1]NELVT, Department of CS, Peking University, Beijing, China
[2]Institute for Artificial Intelligence, Peking University, Beijing, China
[3]The Peng Cheng Laboratory, Shenzhen, China

{cshebin, wce, shiboxin, lingyu}@pku.edu.cn

Contents

Capture Settings

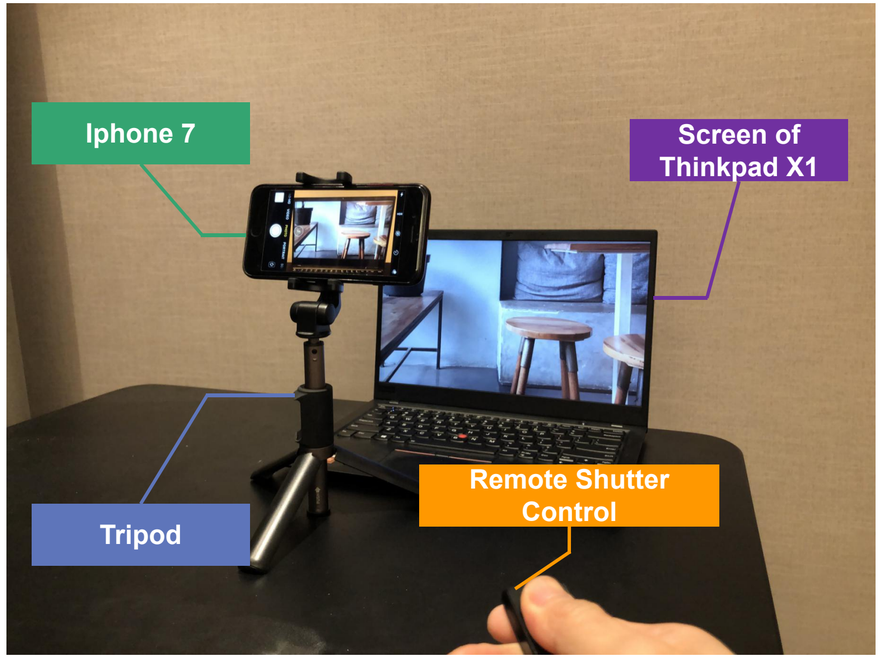


Figure S1: Moire pattern capture setting

As demonstrated in Figure S1, the clean images are shown in full-screen size on the monitor screen, with the camera phone fixed on a tripod. The distance and viewing angle between the screen and the tripod are adjusted to guarantee the clarity of the moire patterns captured and the whole screen within the viewfinder, and the capture is conducted after the parameters are fixed according to current imaging condition, including exposure and focal length.

The viewpoint and camera parameters are changed to diversify the data, the viewpoints are slightly moved every five shots within the depth of field of the camera, while exposure and focal length are changed every 100 shots.

Capture Settings

Table S1: Phone model specifications

Manufacturer	Model	Camera
Apple	Iphone 7	12 MP
SAMSUNG	Galaxy S10	16 MP
Huawei	Mate 20	24 MP

Table S2: Screen model specifications

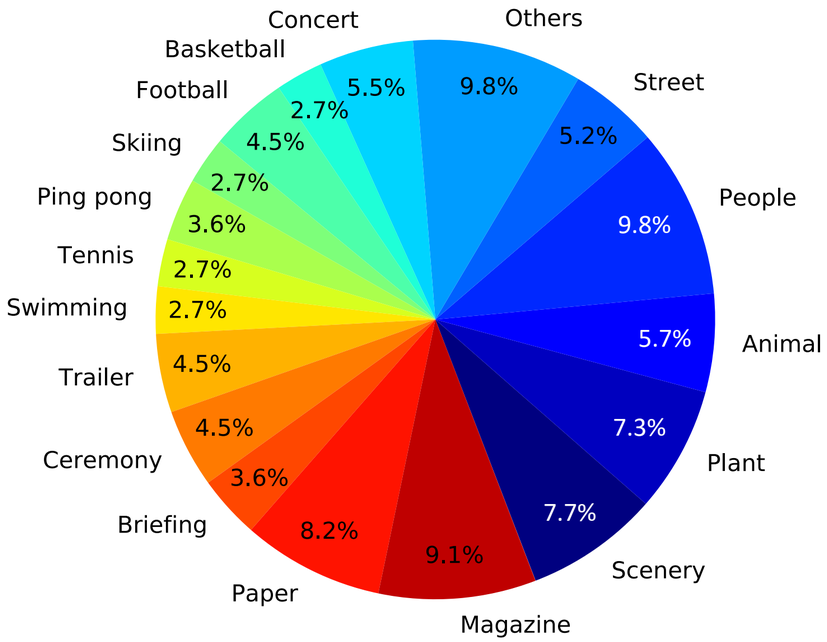
Manufacturer	Model	Resolution	Size
LENOVO	Thinkpad X1	1920 x 1080	14"
DELL	P2412H	2048 x 1152	23"

We adopt 6 combinations of different camera phones and display screens for the diversity of the intrinsic distributions of images. The detailed information is shown in Table S1 and Table S2.

After being captured, the moire-contaminated screen image is calibrated and aligned towards the clean source image, to alleviate the negative influence of pixel shifts in the supervision for learning. Although dense and absolutely accurate calibration for camera distortion is infeasible, we employ the widely-adopted calibration method [1] to estimate camera intrinsics every time the camera parameters are modified, and use the estimated parameter values to mitigate the nonlinear distortions through matlab plug-in.

The calibrated moire image is then aligned to the source image by calculating projection matrix based on SIFT descriptors [2]. The alignment for full-screen image pairs differs from that of cropped pairs in TIP18 [3], we rely on the matching of SIFT features within image contents, instead of additional markers in four corners of the display image. The final alignment matrix is computed based on RANSAC algorithm using the coordinates of matched keypoint pairs.

Dataset Details



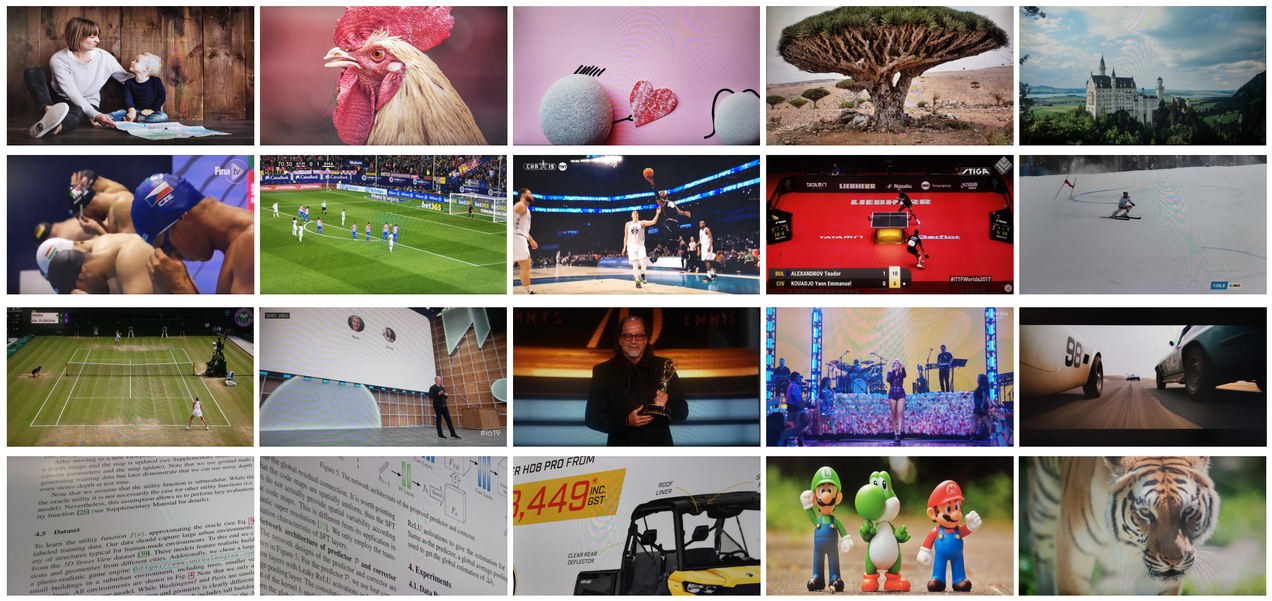
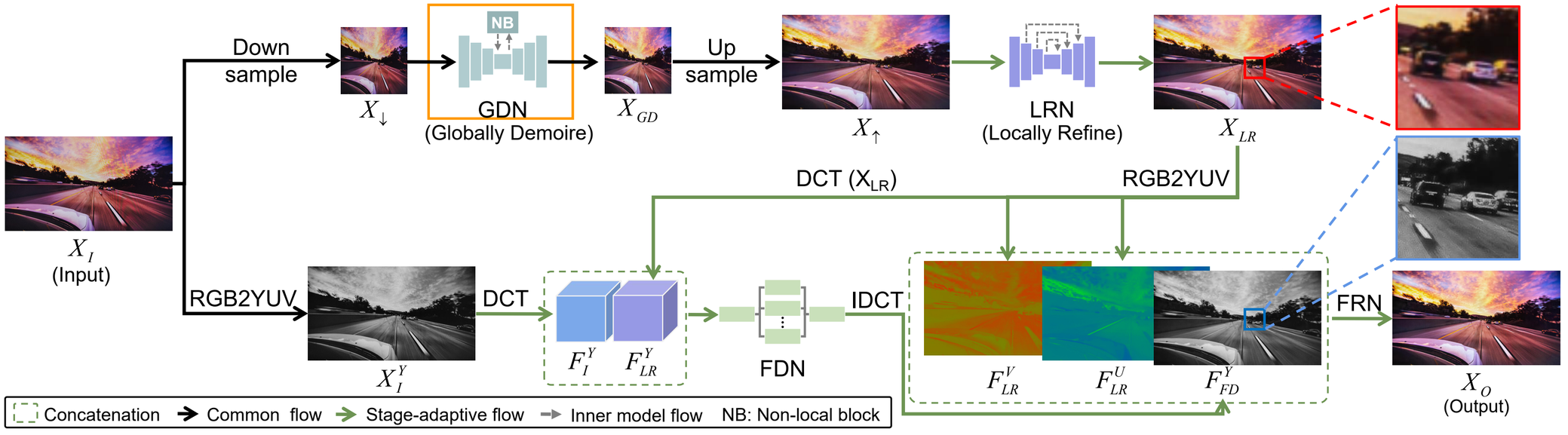


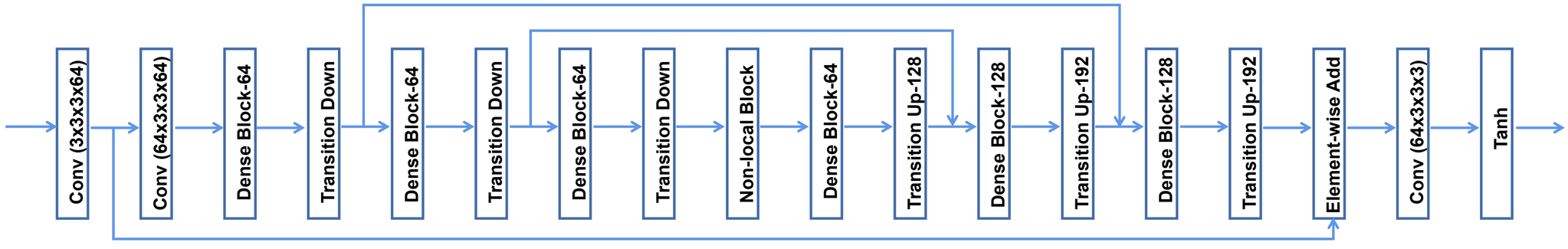
Figure S2: Dataset detail (left) and typical instances (right)

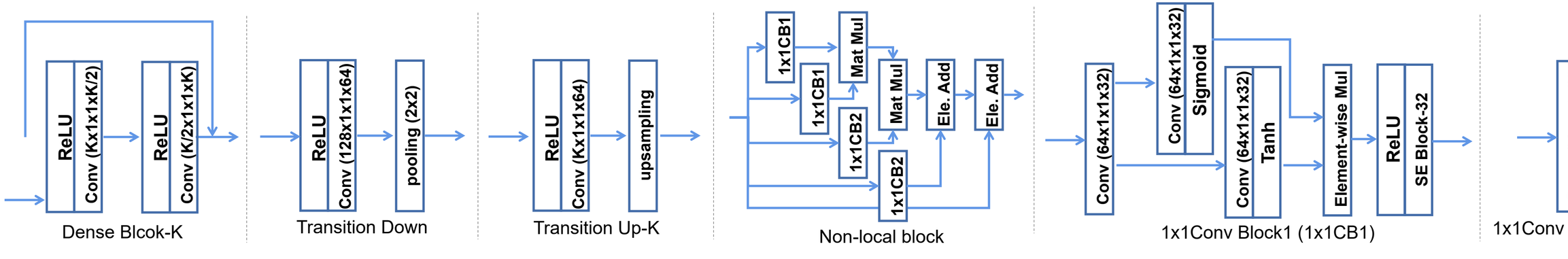
The distribution of categories of images in FHDMi and proportions of each category are illustrated in Figure S2 (left). The images cover major scenarios that are highly likely to be captured on screens: static images (including streetview, animal, scenery, etc.), Internet videos (various sports games, ceremony, film trailer, and concert), and documents (paper and magazine). The data are collected from high-resolution Internet resources on websites including Pexel, Youtube, and magazinelib. Typical instances are shown in Figure S2 (right).

Network Architecture

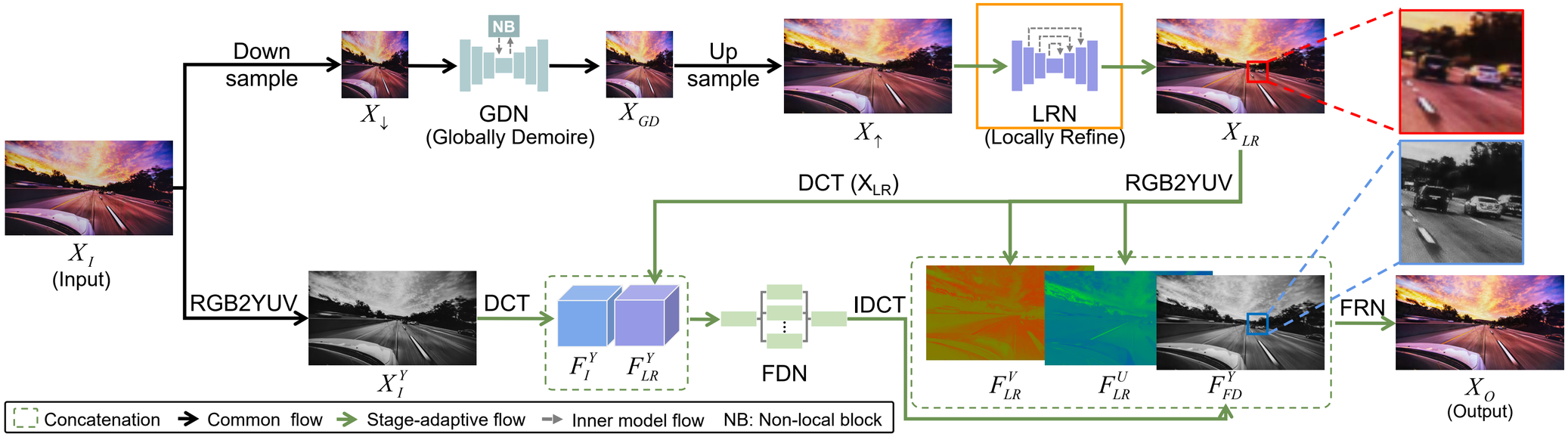


GDN:

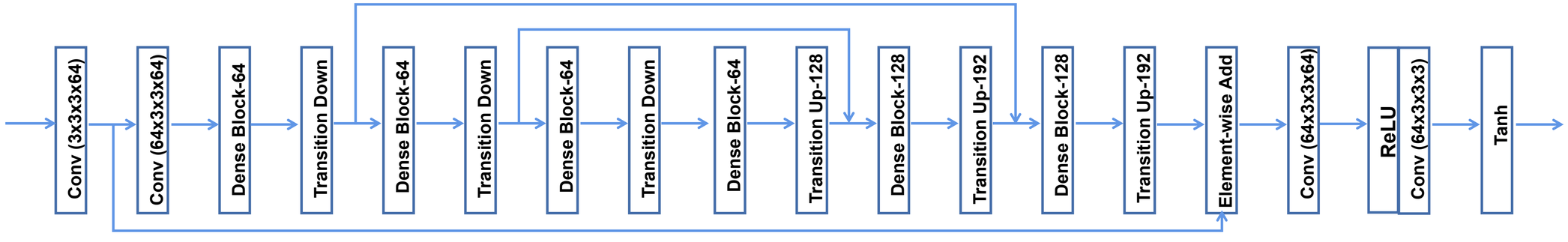


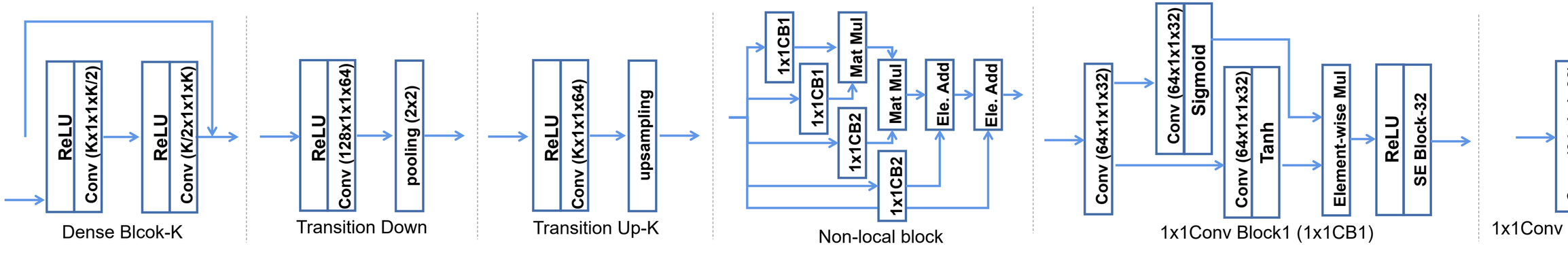


Network Architecture

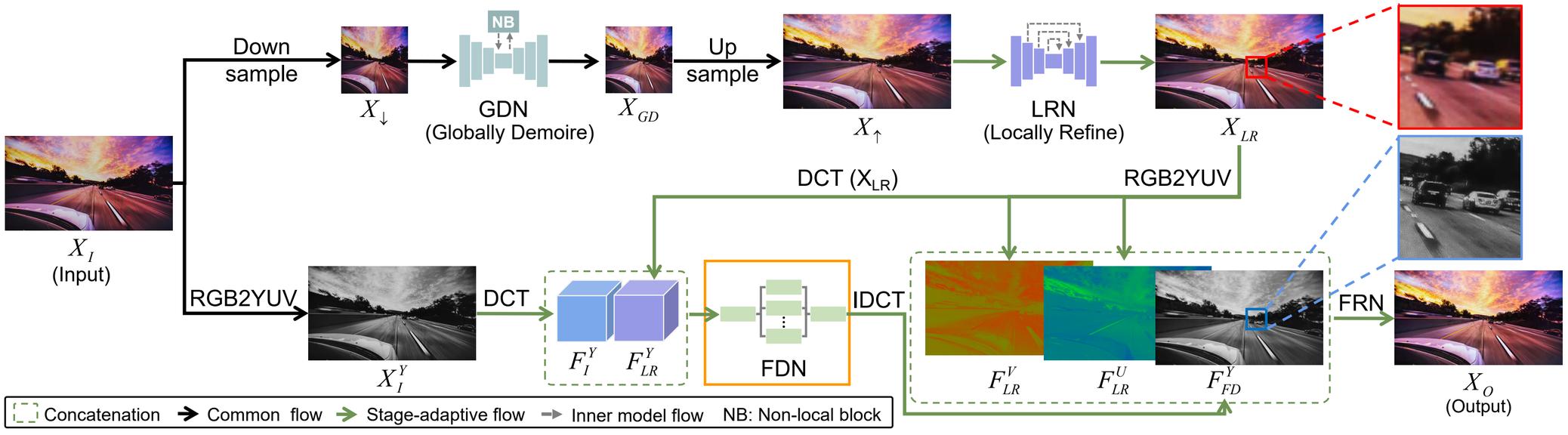


LRN:

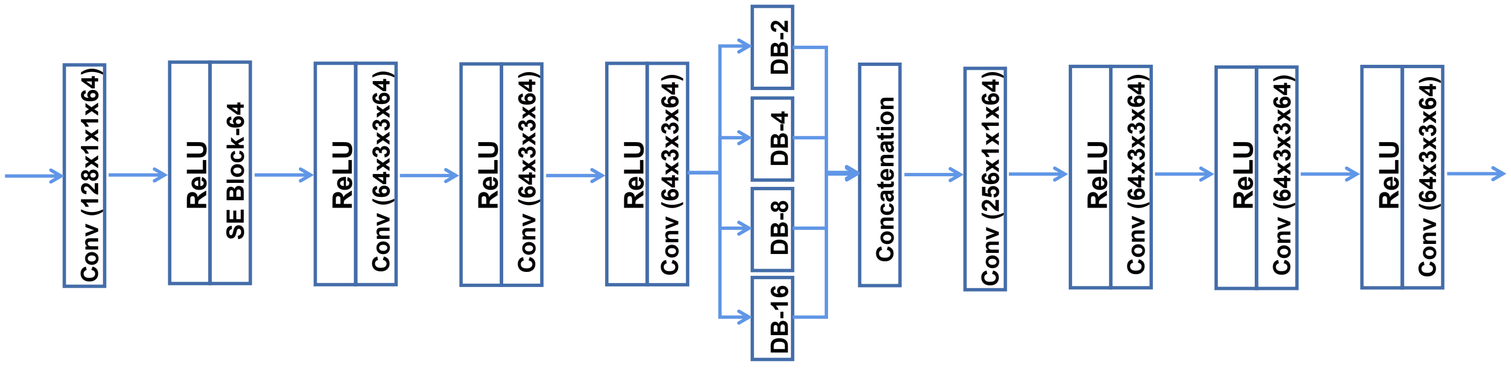


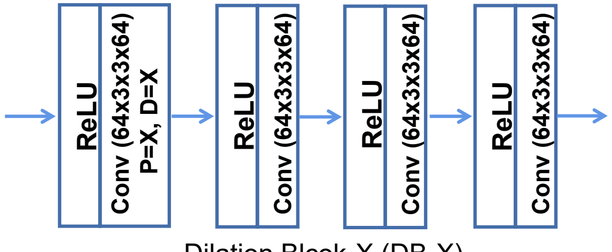


Network Architecture

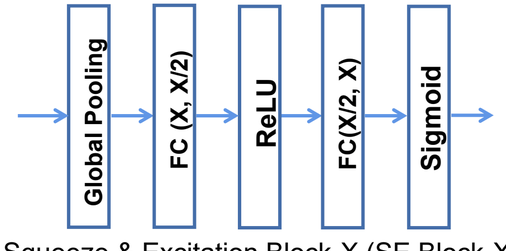


FDN:

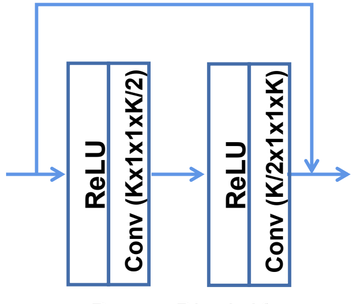




Dilation Blcok-X (DB-X)

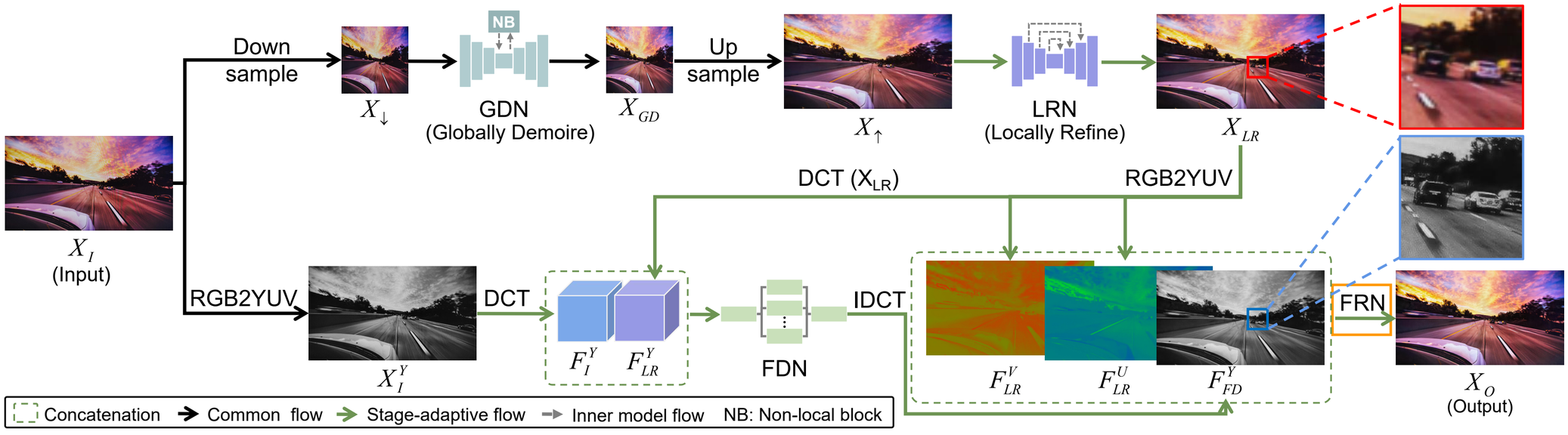


Squeeze & Excitation Block-X (SE Block-X)

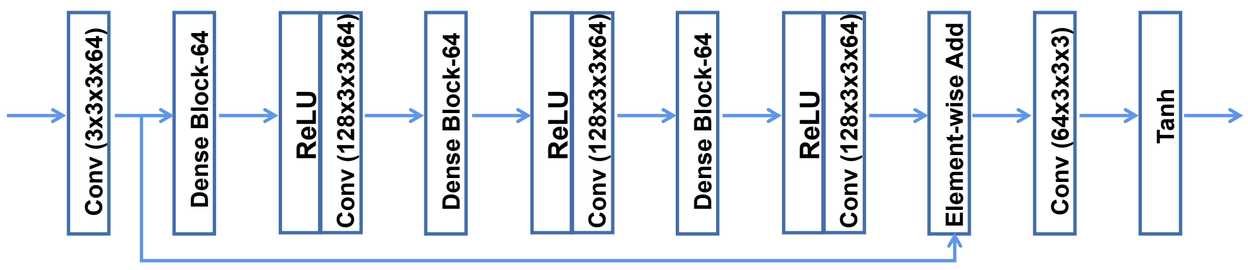


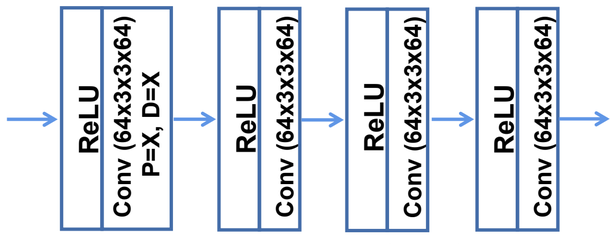
Dense Blcok-K

Network Architecture

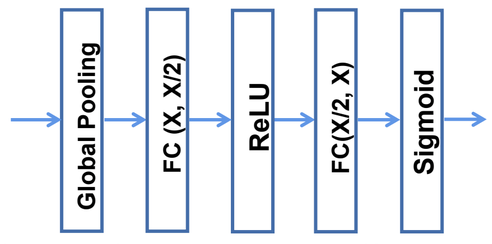


FRN:

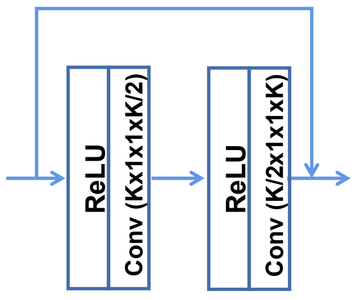




Dilation Blcok-X (DB-X)



Squeeze & Excitation Block-X (SE Block-X)



Dense Blcok-K

Training Set Distillation

Given demoired result X_{GD} from GDN, and corresponding groung truth image X_{GT}, we can distill a better training set for LRN as following:

We obtain the edge maps E_{GD}, E_{GT} by convoloving X_{GD} and X_{GT} with Sobel kernel. To narrow down the potential moire-sensitive regions, we then adopt OTSU [5], to get the binary mask based on the ground truth image. After that, we can focus on the edge information on regions according to the binary mask: X_M

$$E_{GD}^{M} = E_{GD} * X_{M}$$
$$E_{GT}^{M} = E_{GT} * X_{M} \ ,$$

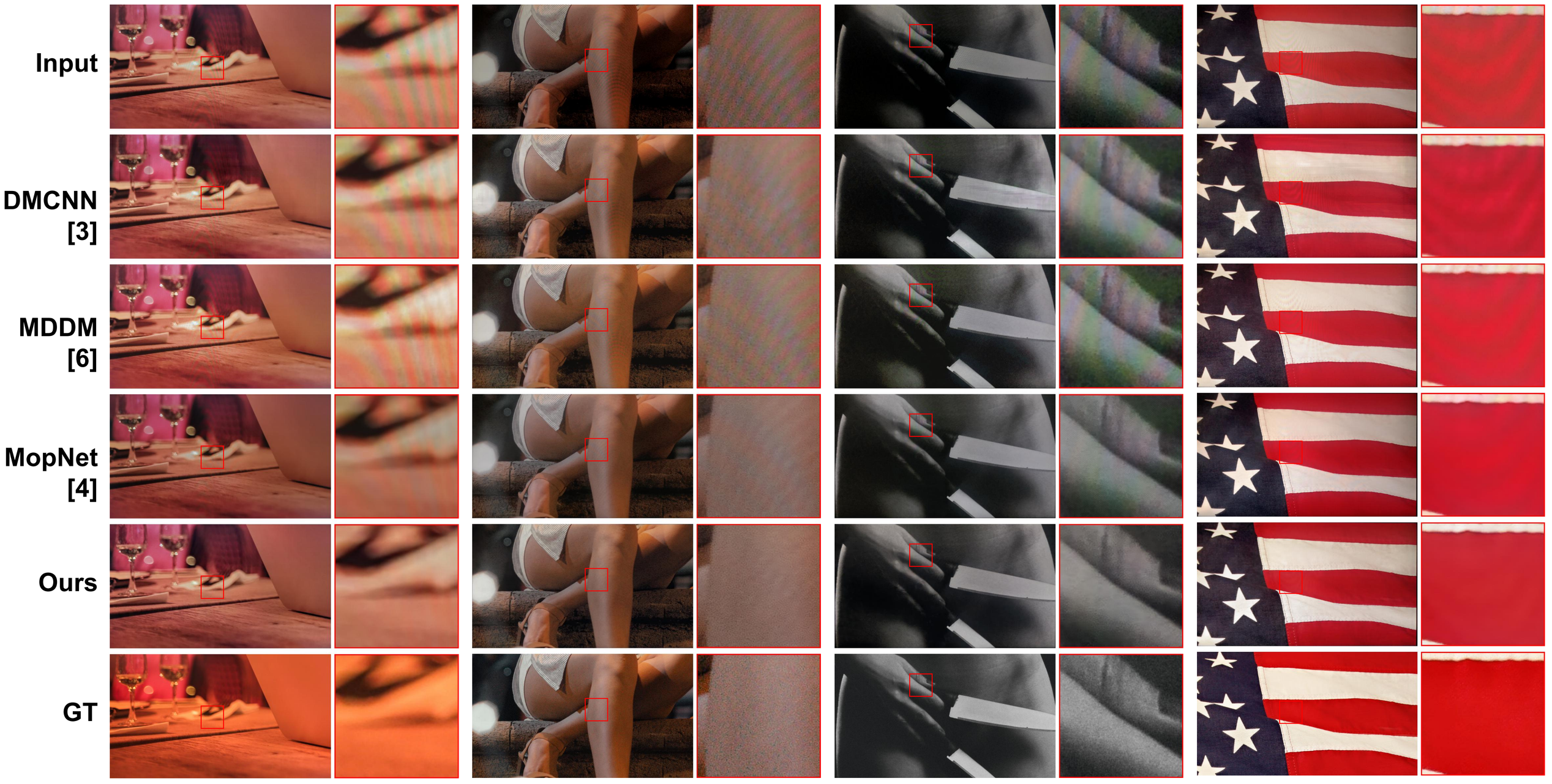
where $*$ stands for element-wise multiplication, E_{GD}^{M} and E_{GT}^{M} for masked edge maps. Then we can calculate the distance d between the edge maps in regions we are interested in as:

$$d = D(E_{GD}^{M}, E_{GT}^{M}) \ ,$$

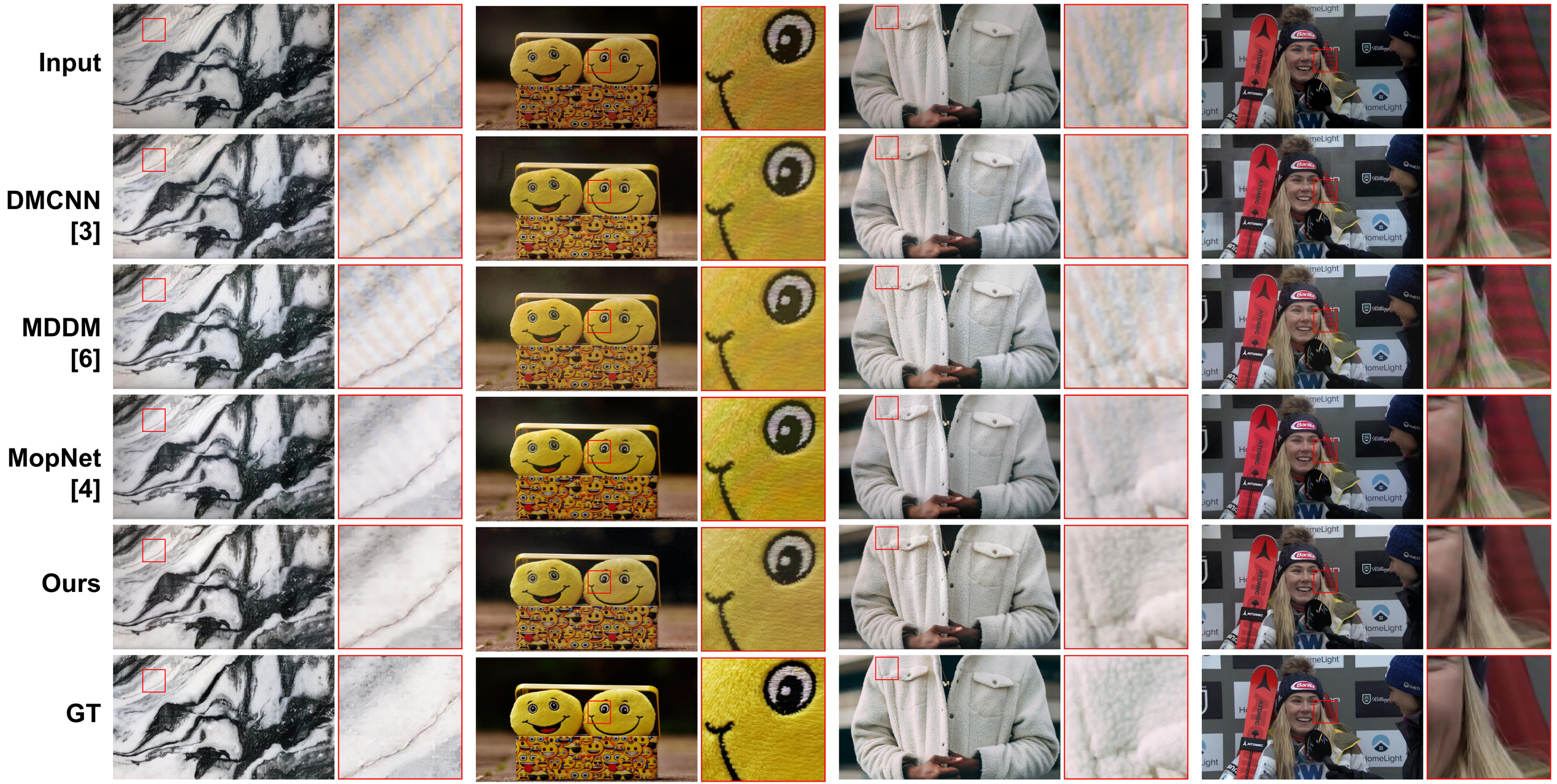
where D denotes distance metric. Finally, we sort the images from GDN in descending order based on the distance d calculated above and select top K images to build the distilled training set for LRN.

In our experiments, we choose K as 1000. Considering the mild misalignment between demoired and gound truth images, we adopt PLIPS [7] to meansure the distance.

More Comparison Results



More Comparison Results



More Comparison Results

Input

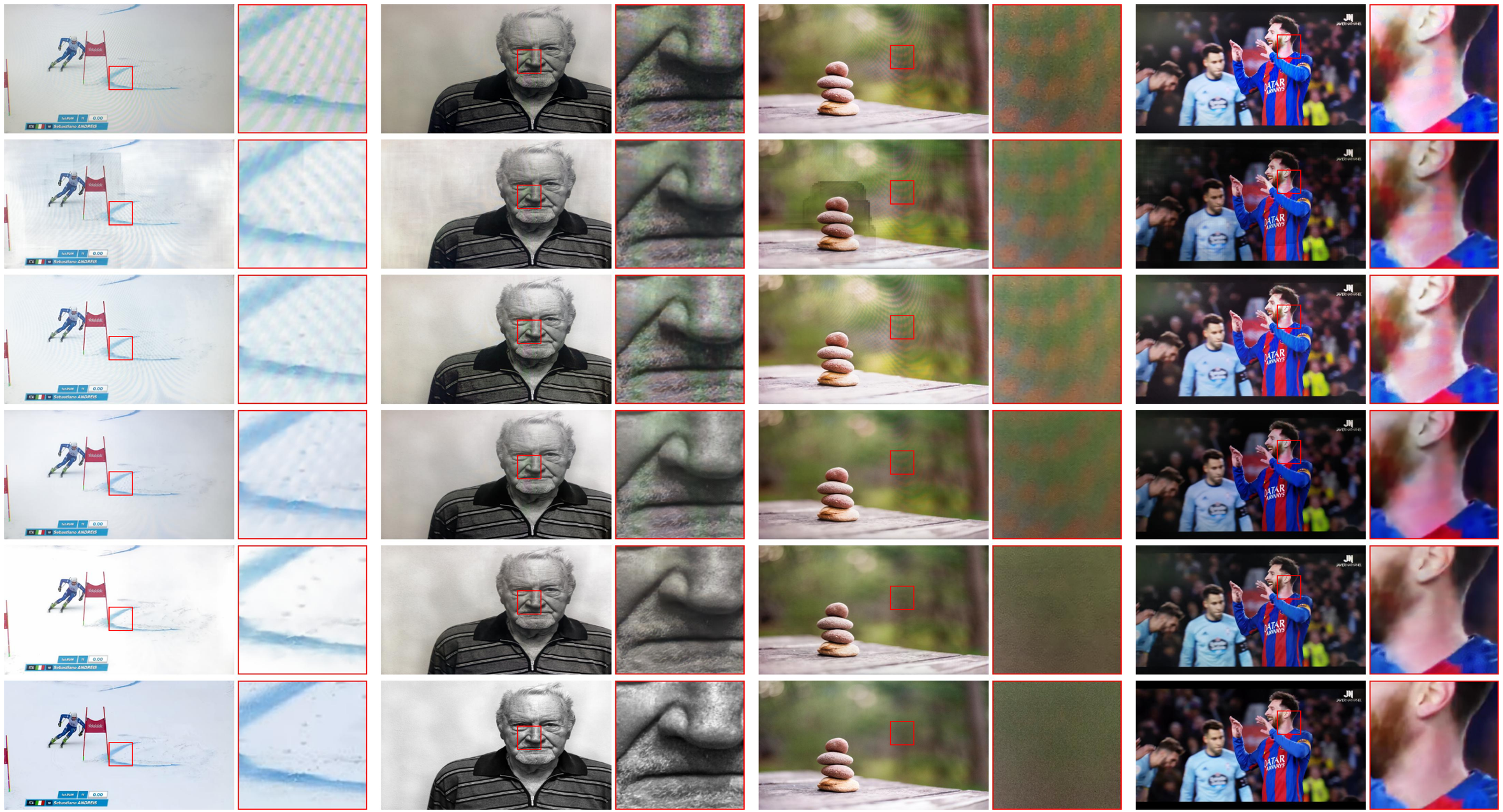
DMCNN [3]

MDDM [6]

MopNet [4]

Ours

GT



More Comparison Results

More Comparison Results

More Comparison Results

Input

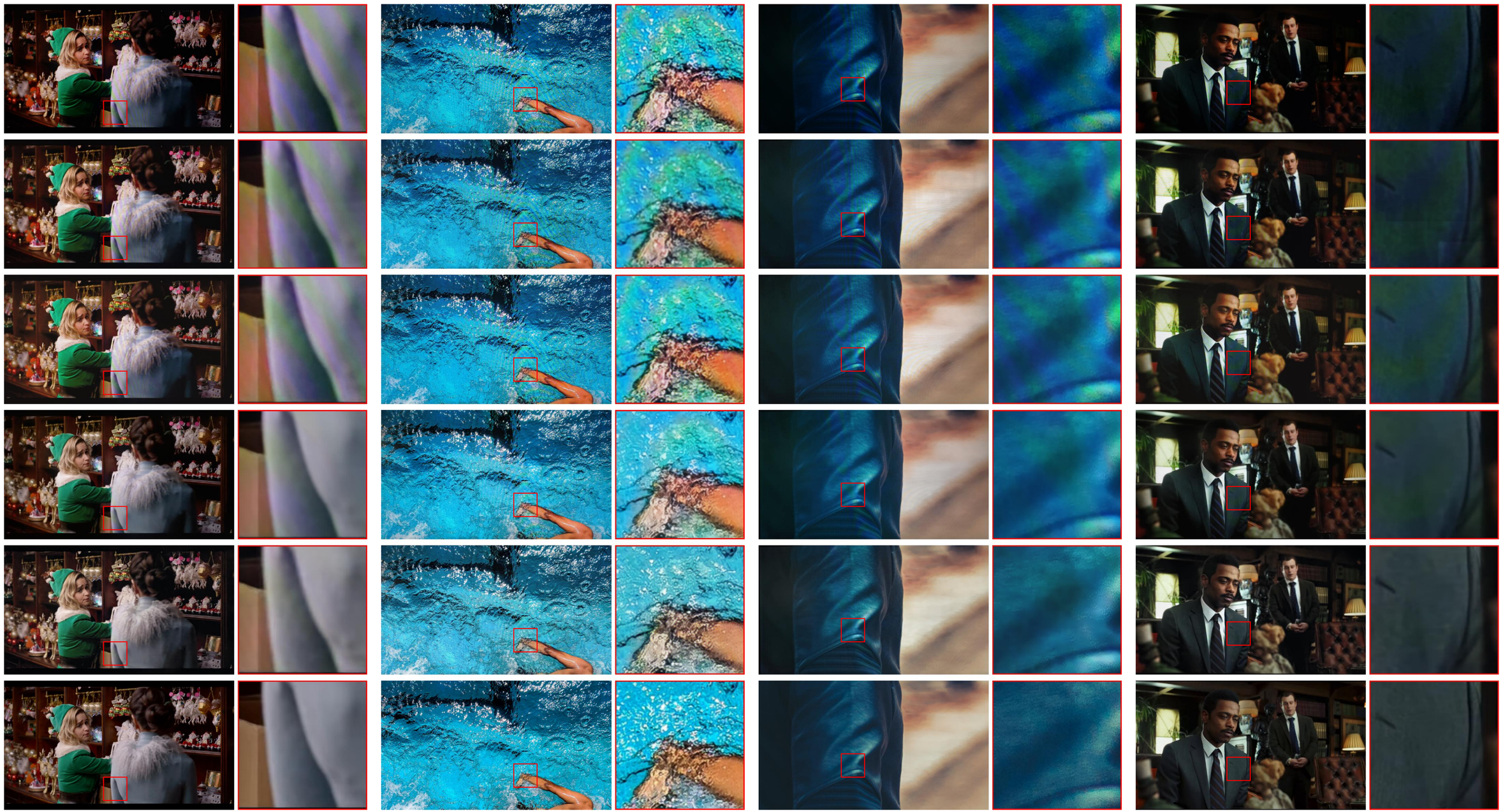
DMCNN [3]

MDDM [6]

MopNet [4]

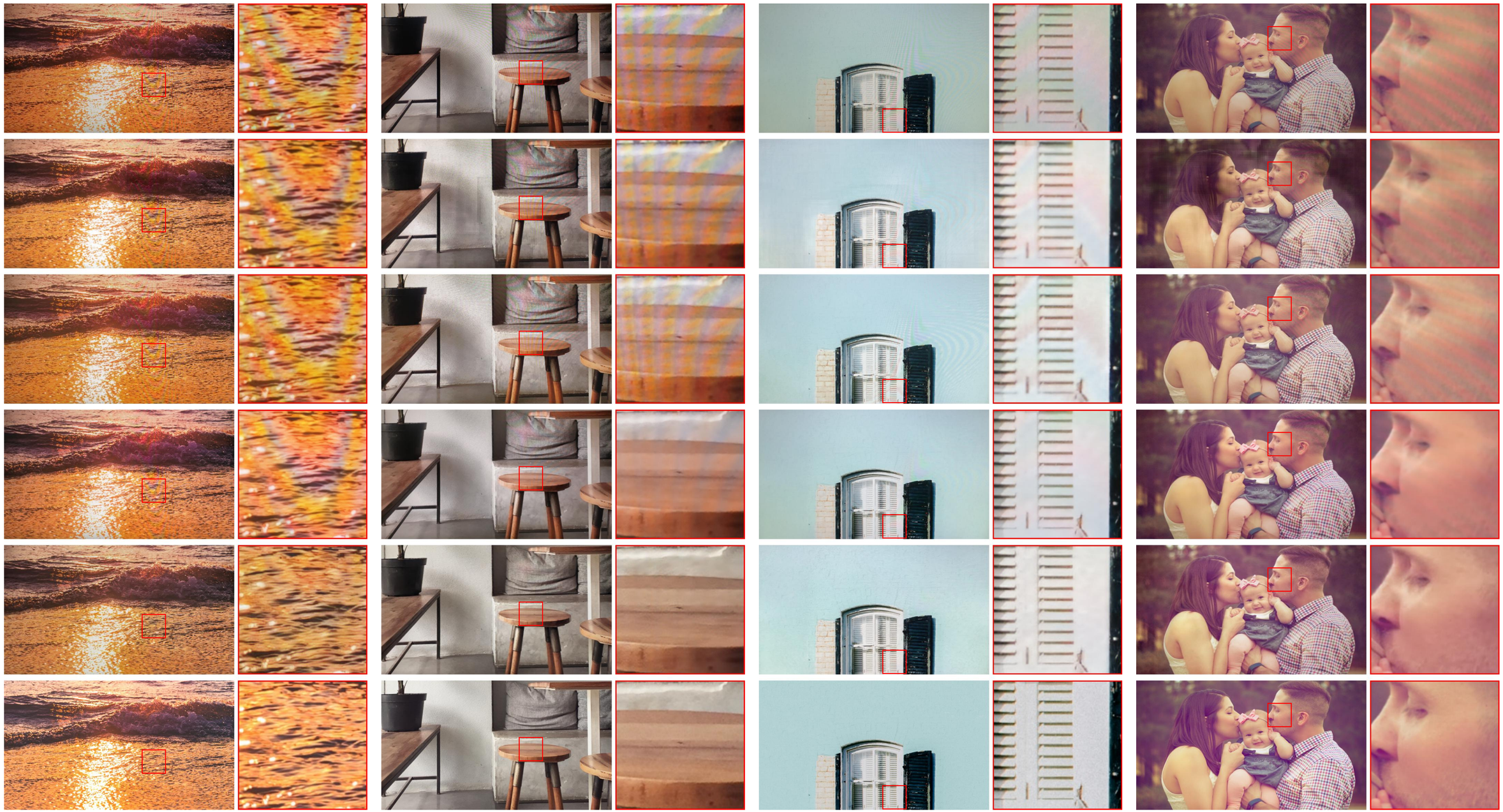
Ours

GT

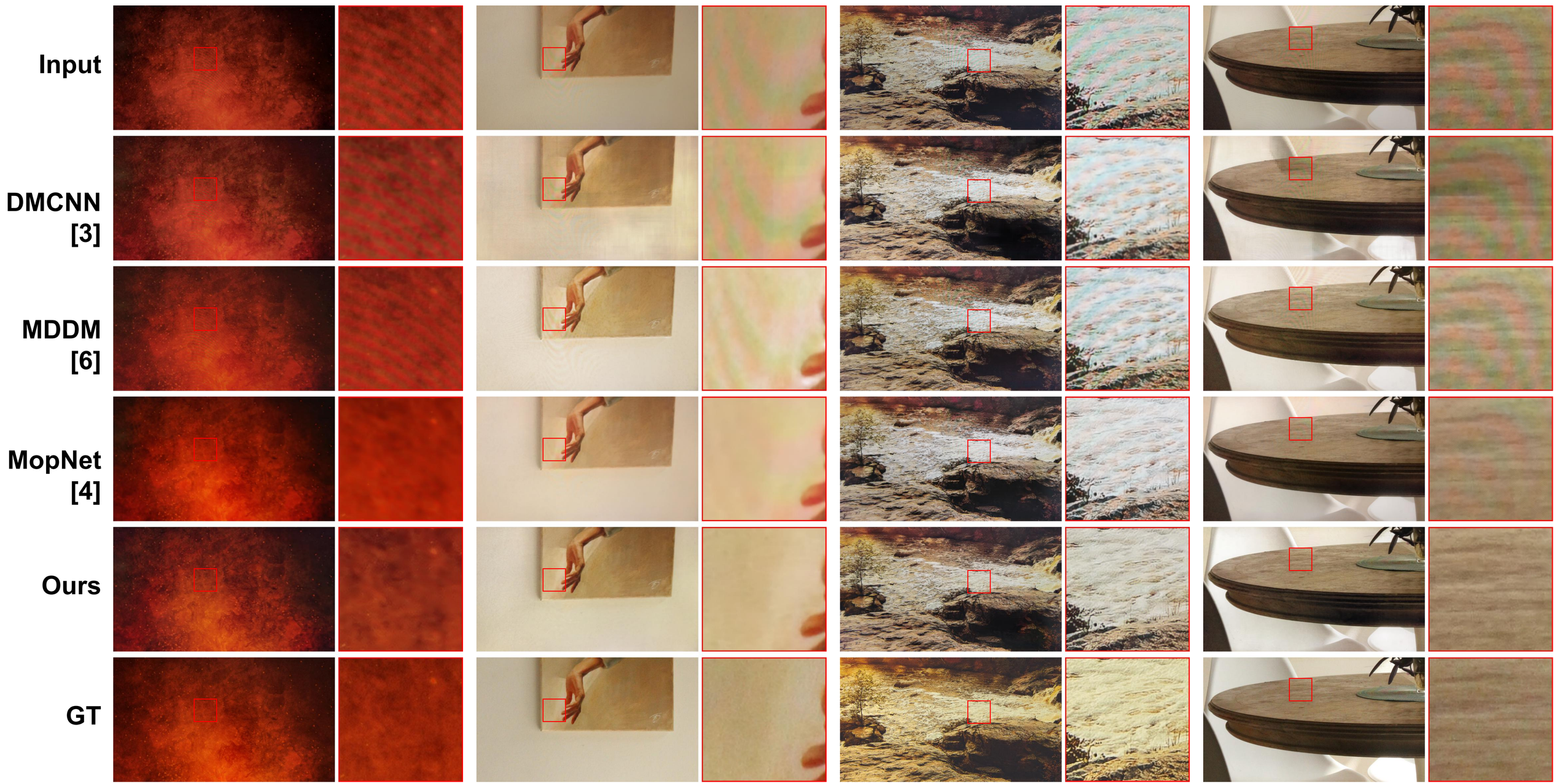


More Comparison Results

More Comparison Results

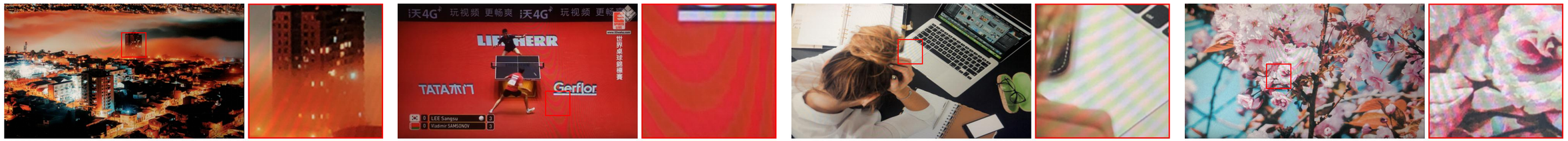


More Comparison Results

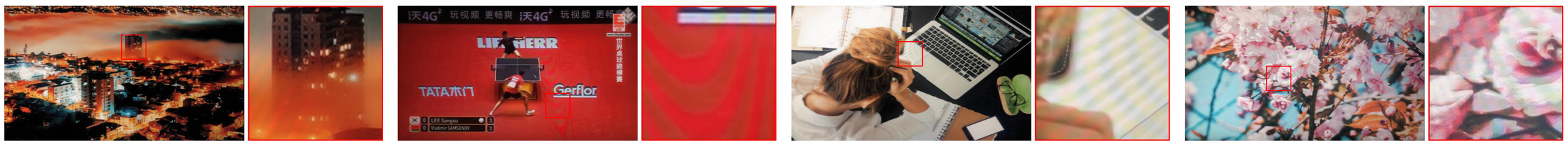


More Ablation Results

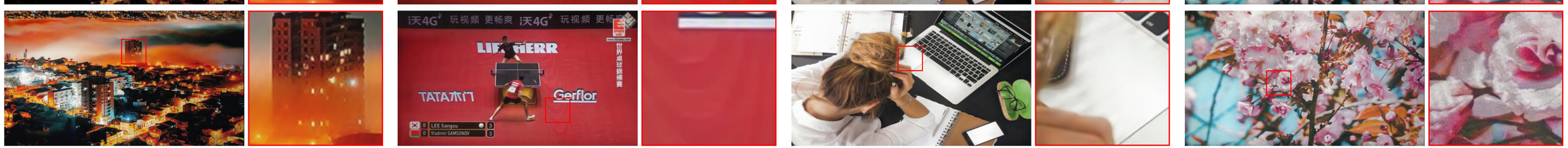
Input



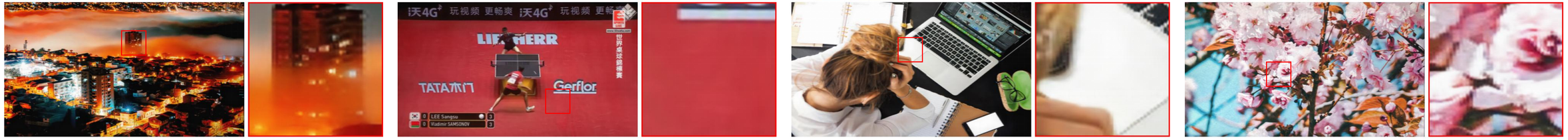
w/o GDN



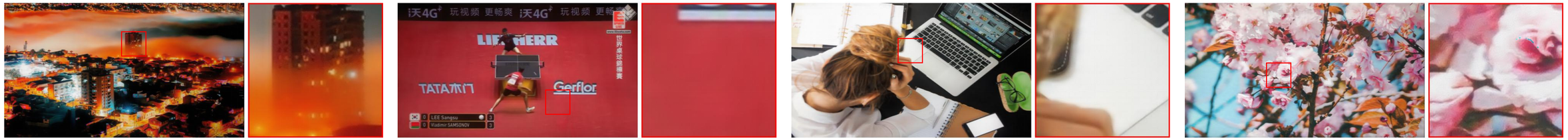
w/o LRN



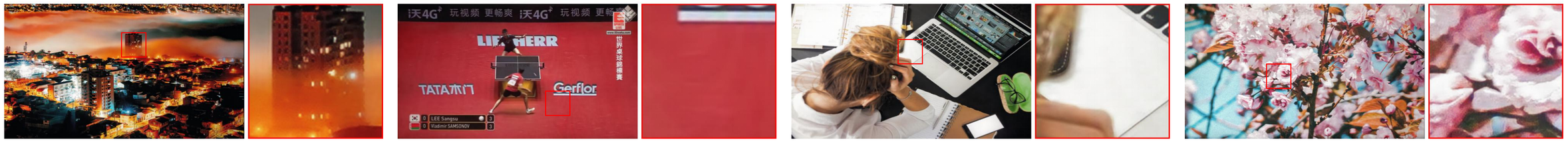
w/o FDN



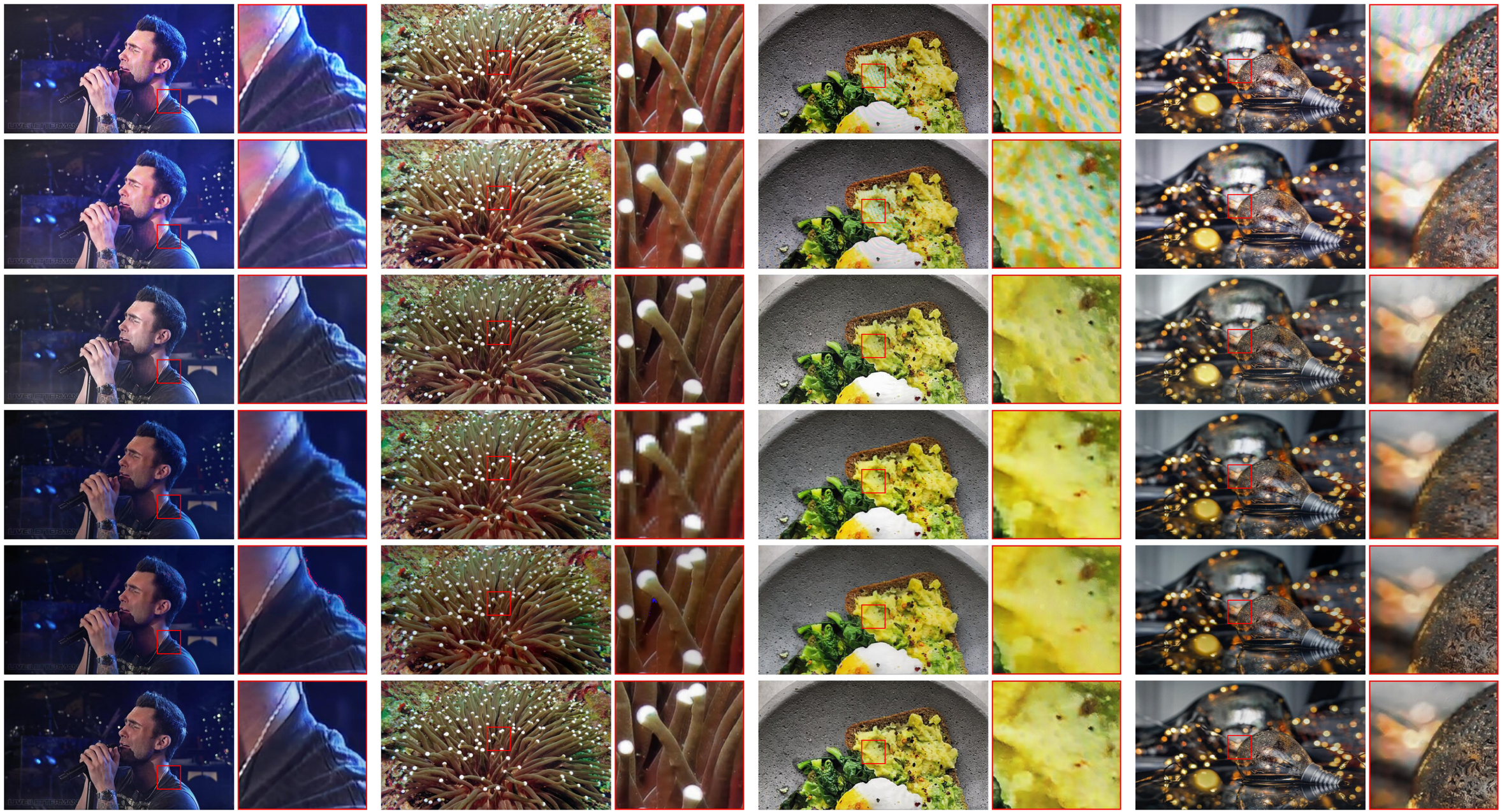
w/o FRN



Full

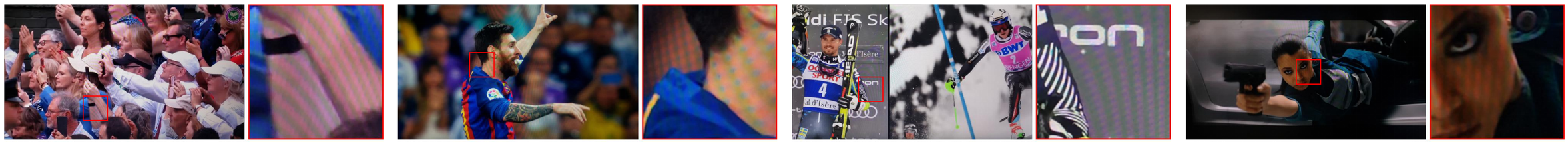


More Ablation Results

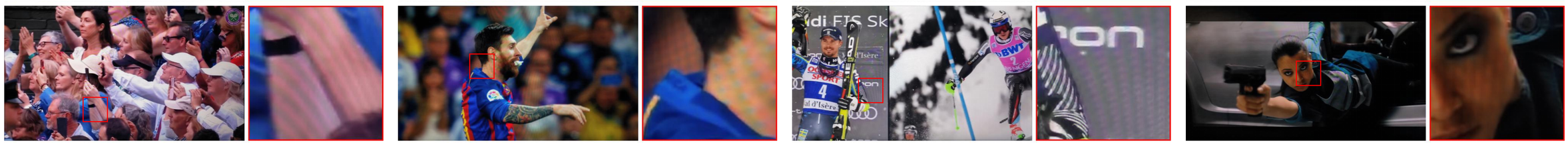


More Ablation Results

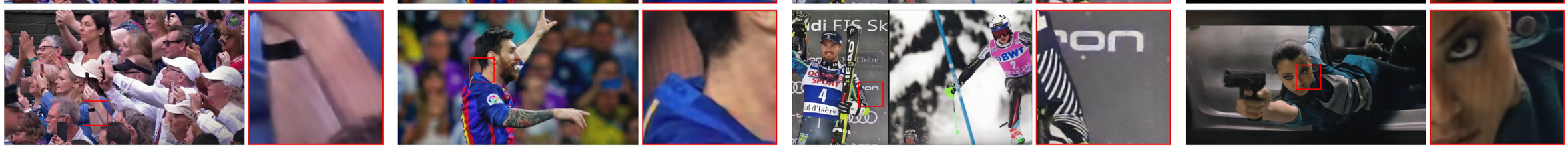
Input



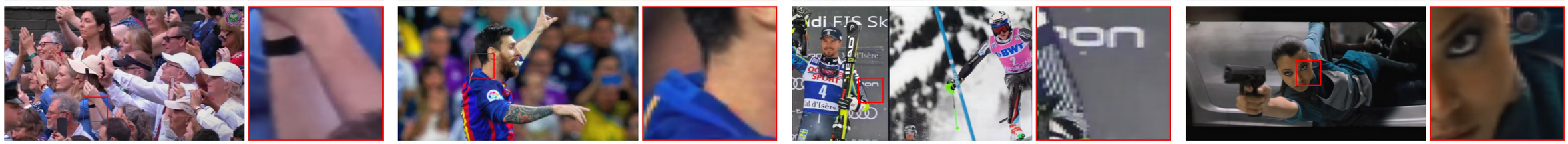
w/o GDN



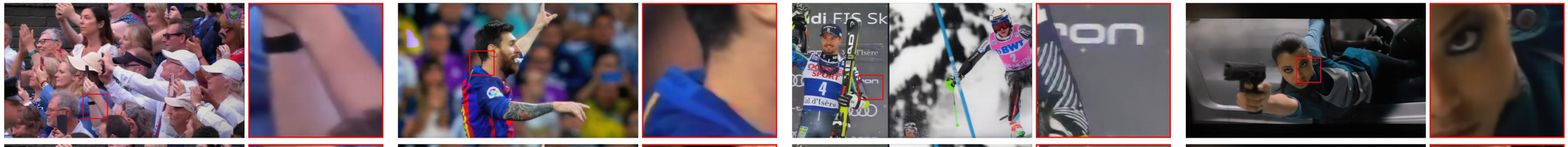
w/o LRN



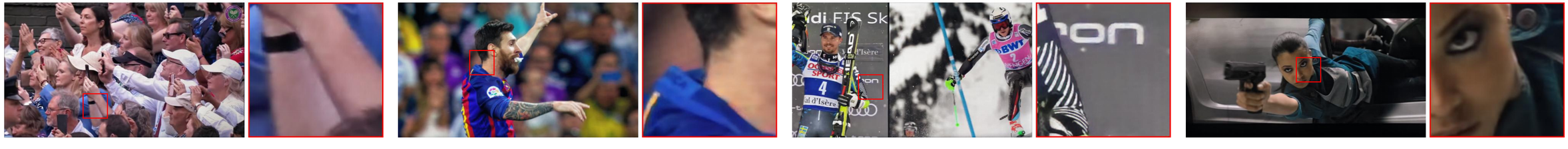
w/o FDN



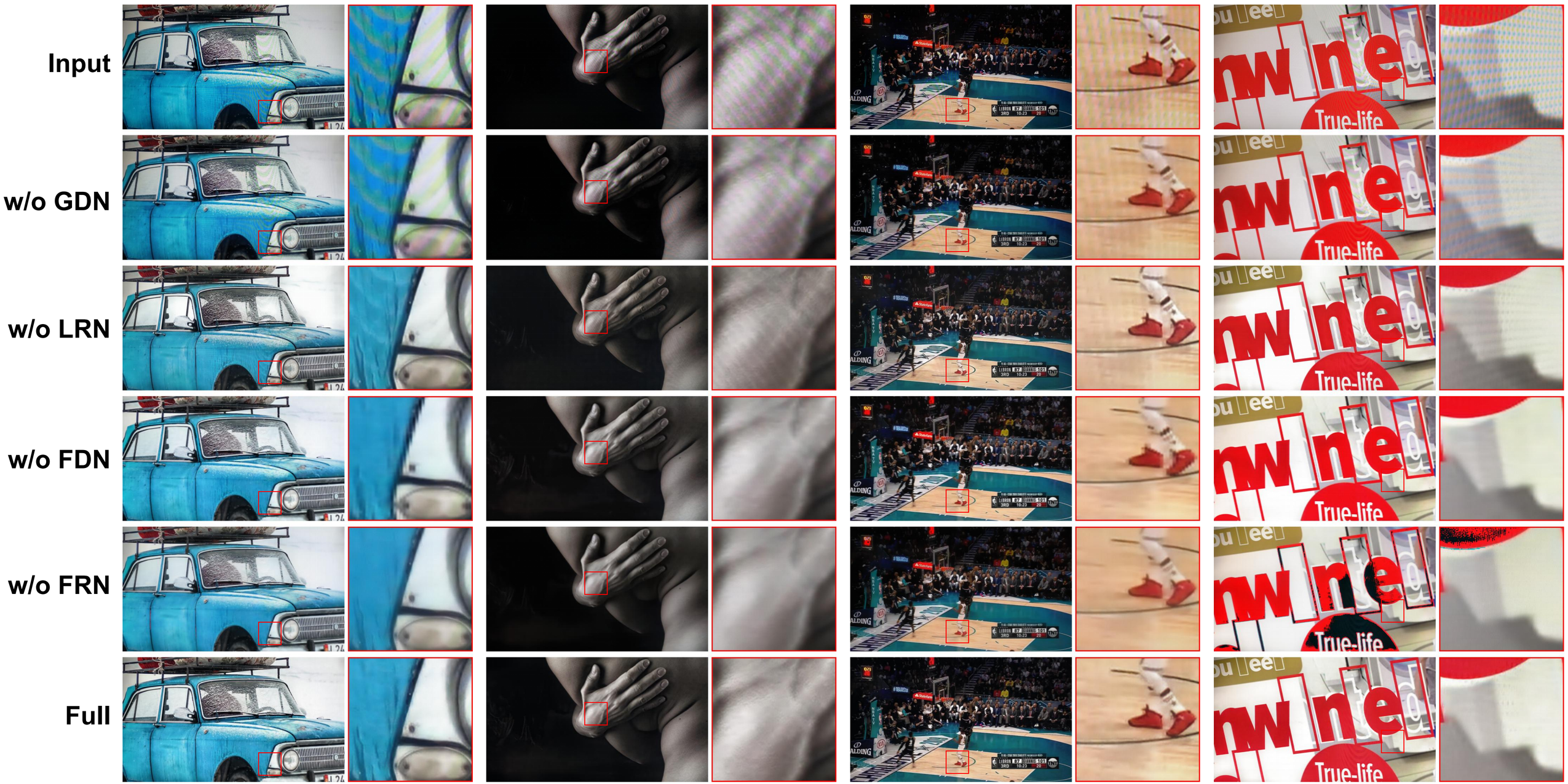
w/o FRN



Full



More Ablation Results



Run Time and Network Parameters

Table S3: Run time and network parameters

	DMCNN [3]	MDDM [6]	MopNet [4]	Ours
Run Time (s)	0.015	0.1	0.47	0.04
Parameter (MB)	5.93	30.80	231.00	51.92

We can see that the performance of demoireing model seems not to show direct relations to the model size (parameter amount) since our method achieves generally better performances than MopNet with significantly fewer network parameters. We are working on combining model compression with image restoration to maintain demoireing performance with fewer model parameters towards mobile deployment.

As for the run time among different models, the proposed FHDe^{2}Net is the efficient enough to handle moire pattern removal on images with high resolution.

Limitation

Input **MopNet** **Ours**

Figure S3: Challenging cases of 4K images

As can be observed in Figure S3, FHDe^{2}Net suffers from undesired local residues (yellow boxes) in low texture regions due to the sparsity of keypoint features. However it still outperforms state-of-the-art MopNet [4] overall, which cannot effectively remove large-scale patterns (red boxes). We will explore better learning constraints with less dependence on matchable features to improve such challenging cases in our future work.

Reference

1. Zhang, Z.: A flexible new technique for camera calibration. IEEE transactions on Pattern Analysis and Machine Intelligence 22(11), 1330-1334 (2000)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91-110 (2004)
3. Sun, Y., Yu, Y., Wang, W.: Moire photo restoration using multiresolution convolutional neural networks. IEEE Transactions on Image Processing 27(8), 4160-4172 (2018)
4. He, B., Wang, C., Shi, B., Duan, L.Y.: Mop moire patterns using mopnet. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2424-2432 (2019)
5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1), 62-66 (1979)
6. Cheng, X., Fu, Z., Yang, J.: Multi-scale dynamic feature encoding network for image demoireing. arXiv preprint arXiv:1909.11947 (2019)
7. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586-595 (2018)