Rethinking Coarse-to-Fine Approach in Single Image Deblurring

Group Meeting

Presenter: Keng Seng Ho

Date: 2025/05/06

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Outline

- Introduction
- Related work
- Proposed method
- Result

Outline

- Introduction
- Related work
- Proposed method
- Result

Introduction

Image Blurring model as follow

$$B = I * K + N$$

B: Blurry image, I: Sharp image, K: Blur kernel, N: Noise, and $*$: Convolution.

Problem type:

- K(known): Non-Blind Deblurring
- K(unknown): Blind Deblurring

[3] Zhang, K., Ren, W., Luo, W., Lai, W. S., Stenger, B., Yang, M. H., & Li, H. (2022). Deep image deblurring: A survey. International Journal of Computer Vision, 130(9), 2103-2130.

Conventional (Blind motion deblurring)

- **Workflow:**

1. Blur kernel estimation
2. Apply non-blind deblurring
   - Inverse filter
   - Wiener filter
   - Deconvolution methods

- **Problems:**

1. Motion blur kernel (non-uniform)
2. Blur kernel estimation is inaccurate

(a) (b) (c)

Figure 2. (a) Ground truth sharp image. (b) Blurry image generated by convolving a uniform blur kernel. (c) Blurry image by averaging sharp frames. In this case, blur is mostly caused by person motion, leaving the background as it is. The blur kernel is non-uniform, complex shaped. However, when the blurry image is synthesized by convolution with a uniform kernel, the background also gets blurred as if blur was caused by camera shake. To model dynamic scene blur, our kernel-free method is required.

[2]

[2] S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring,” Jul. 2017.

Outline

- Introduction
- Related work
- Proposed method
- Result

DeepDeblur

- Deep learning
- End-to-end
- Coarse-to-fine strategy

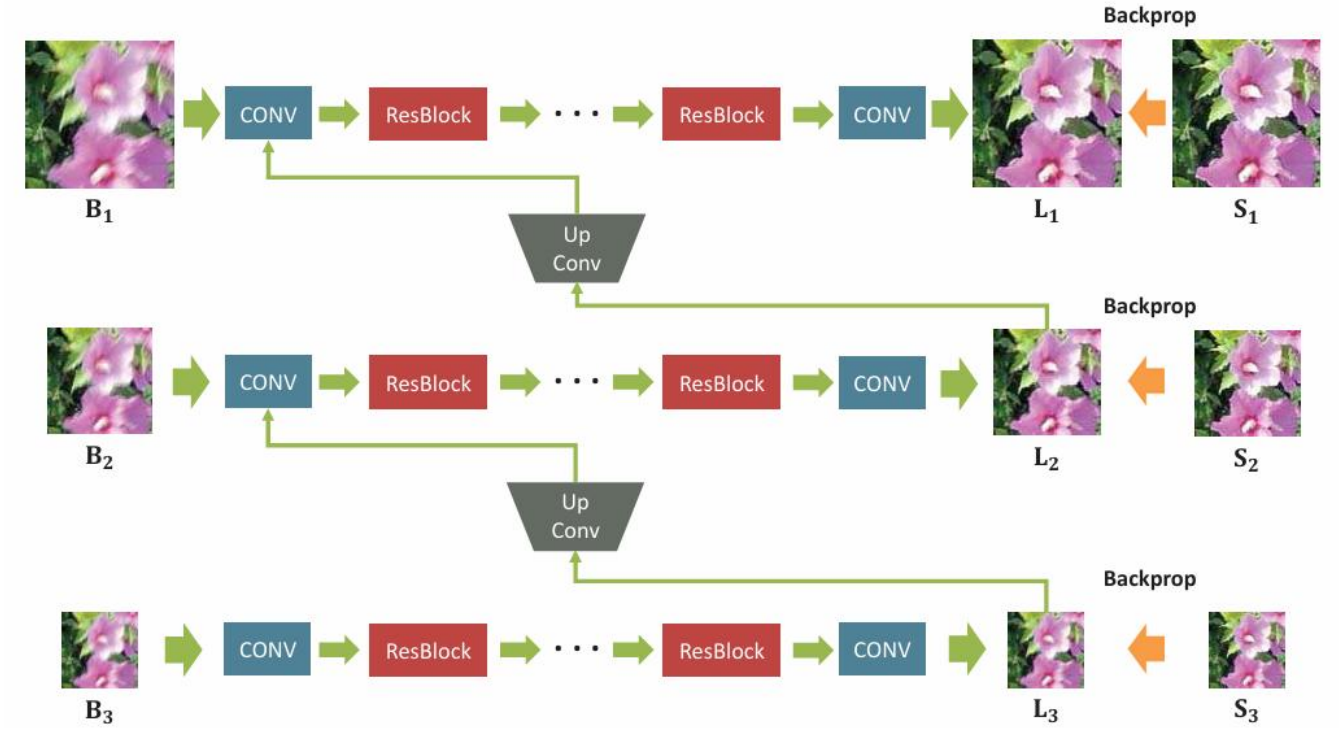


Figure 4. Multi-scale network architecture. B_k, L_k, S_k denote blurry and latent, and ground truth sharp images, respectively. Subscript k denotes k-th scale level in the Gaussian pyramid, which is downsampled to $1/2^k$ scale. Our model takes a blurry image pyramid as the input and outputs an estimated latent image pyramid. Every intermediate scale output is trained to be sharp. At test time, original scale image is chosen as the final result.

[2]

[2] S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring,” Jul. 2017.

DeepDeblur

- Dataset (GoPro)

A blurry image is generated when the camera sensor accumulates light during the exposure time. The blurry image can be approximated by accumulating signals from high-speed video frames [2].

$$B = g\left(\frac{1}{T}\int_{t=0}^{T} S(t)dt\right) \cong g\left(\frac{1}{M}\sum_{i=0}^{M-1} S[i]\right)$$

T: Exposure time.

$S(t)$: Sharp image at time t.

M: Sampled frames.

$S[i]$: i-th sharp frame.

$g(.)$: Camera Response Function(CRF).

Blurry image by averaging sharp frames [2]

[2] S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring,” Jul. 2017.

Outline

- Introduction
- Related work
- Proposed method
- Result

Proposed network (MIMO-UNet)

- Apply coarse-to-fine strategy in Unet.
- The architecture of MIMO-UNet is divided into three parts:
  - Multi-input Single Encoder (MISE)
  - Asymmetric Feature Fusion (AFF)
  - Multi-output Single Decoder (MOSD)
- Main components:
  - Shallow convolutional module (SCM)
  - Feature attention module (FAM)
  - Asymmetric feature fusion (AFF)

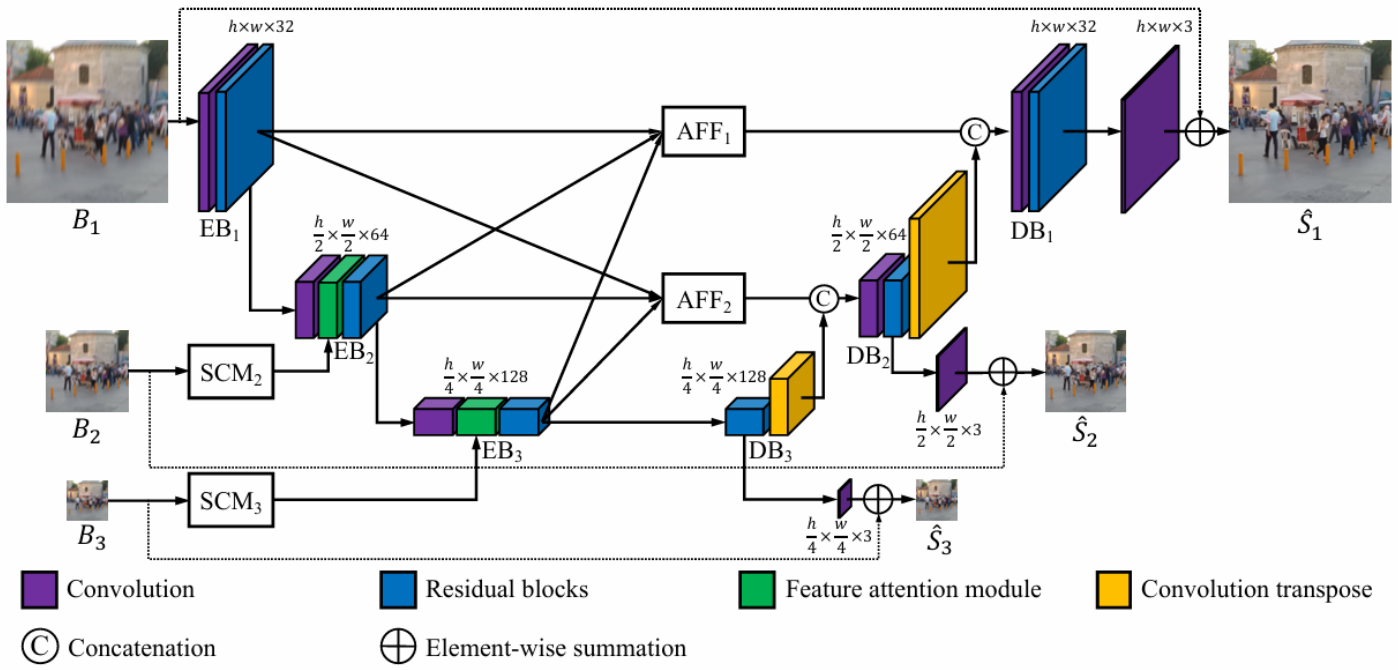


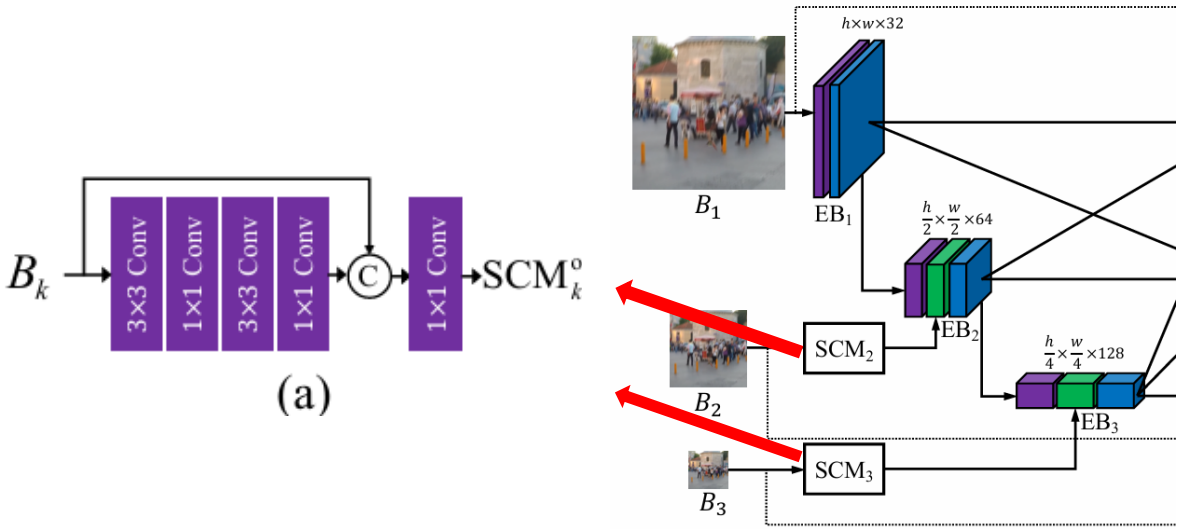
Figure 3. The architecture of the proposed network.

[1]

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Multi-input Single Encoder

- **Shallow convolutional module (SCM)**
- Using SCM to extract downsampled image features.

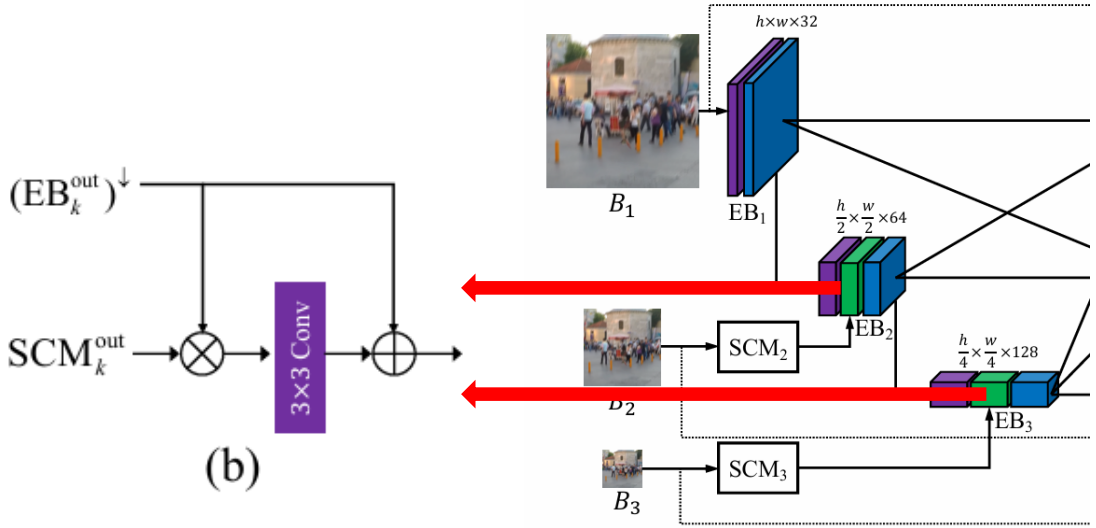


[1]

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Multi-input Single Encoder

- **Feature attention module (FAM)**
- Using FAM to emphasize or suppress the features from the previous scale.
- Learning the spatial/channel importance of the features from SCM.

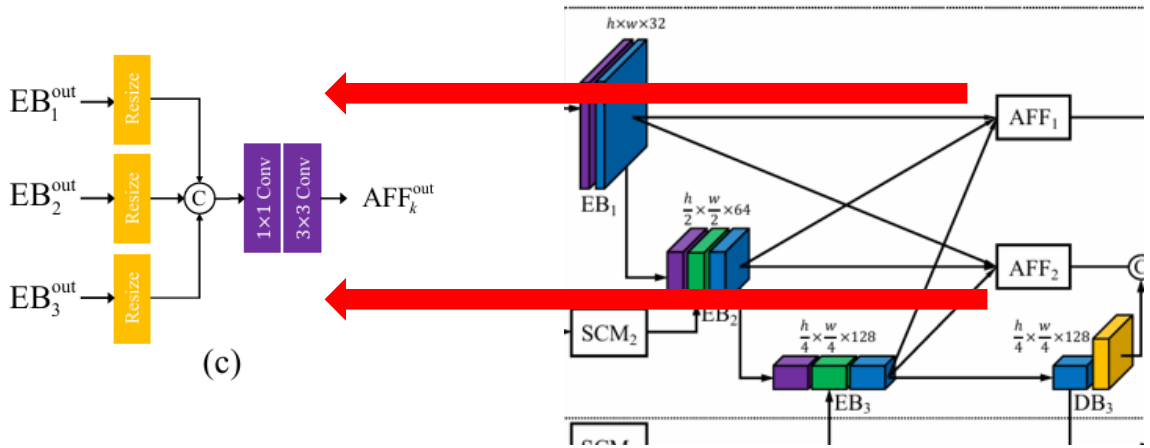


[1]

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Asymmetric feature fusion (AFF)

- **Asymmetric feature fusion (AFF)**
- AFF enables effective information flow across different scales.



[1]

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Multi-output Single Decoder

- Similar to UNet decoder, but with an additional convolution layer at each level to map feature maps to deblurred sharp images.

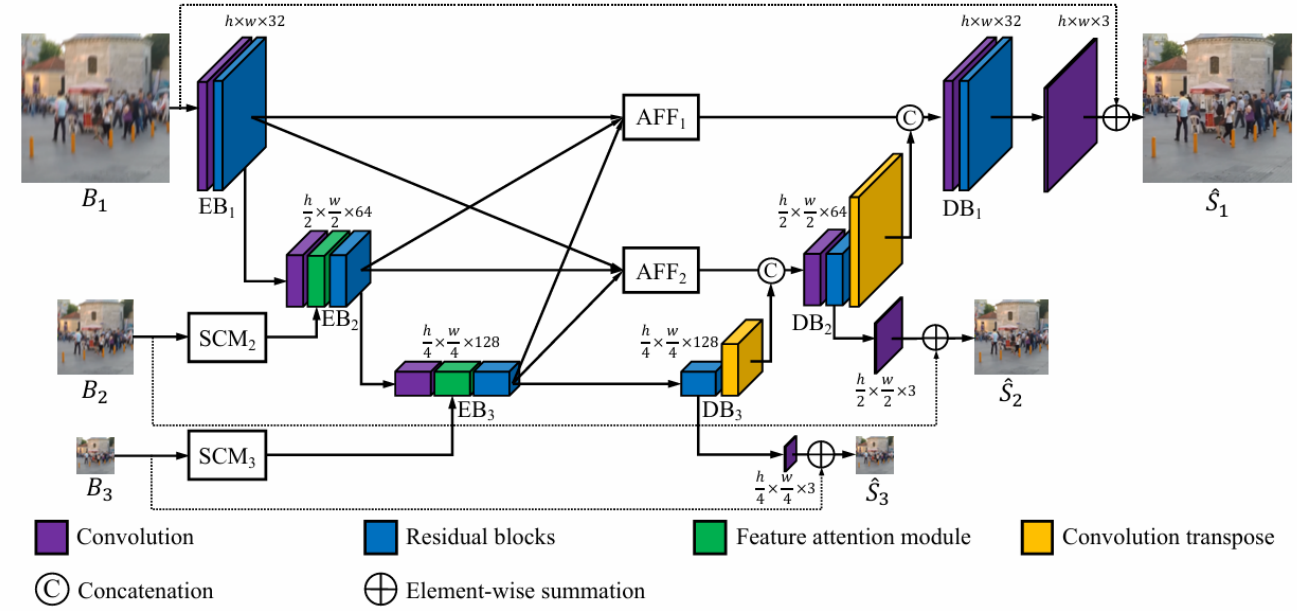


Figure 3. The architecture of the proposed network.

[1]

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Loss function

- **Overall loss:**

$$L_{total} = L_{cont} + \lambda L_{MSFR}$$

- **Content loss:**

$$L_{cont} = \sum_{k=1}^{K} \frac{1}{t_k} \left\| \hat{S}_k - S_k \right\|_1$$

- **Multi-scale frequency reconstruction (MSFR) loss:**

$$L_{MSFR} = \sum_{k=1}^{K} \frac{1}{t_k} \left\| F(\hat{S}_k) - F(S_k) \right\|_1$$

λ: Hyperparameter.

K: The number of levels.

t_k: Number of total elements for normalization.

$\hat{S}_k$: Deblurring images.

S_k: Ground truth.

$F(.)$: FFT.

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Outline

- Introduction
- Related work
- Proposed method
- Result

Result (GoPro/RealBlur)

Synthesis dataset

Model	PSNR	SSIM	Runtime		Params.
DeepDeblur [20]	29.23	0.916	N/A	4.33	11.7
SRN [31]	30.26	0.934	0.342	1.87	6.8
PSS-NSC [5]	30.92	0.942	0.985	1.6	2.84
DMPHN [35]	31.20	0.945	1.061	0.424	21.7
SAPHN† [29]	31.85	0.948	N/A	0.34	N/A
SAPHN‡ [29]	32.02	0.953	N/A	0.77	N/A
MT-RNN [22]	31.15	0.945	0.063	0.07	**2.6**
RADN [23]	31.76	0.953	N/A	0.038	N/A
SVDN [33]	29.81	0.937	N/A	0.01	N/A
MPRNet [34]	32.66	**0.959**	0.162	0.18	20.1
MIMO-UNet	31.73	0.951	**0.008**		6.8
MIMO-UNet+	32.45	0.957	0.017		16.1
MIMO-UNet++	**32.68**	**0.959**	0.040		16.1

Table 1. The average PSNR and SSIM on the GoPro test dataset. The SAPHNs with † and ‡ denote the models with and without offsets, respectively. We employ stacked(4) version for DMPHN. The runtime and parameters are expressed in seconds and millions.

Real dataset

Model	PSNR	SSIM
DeblurGAN-v2 [15]	29.69	0.870
SRN [31]	31.38	0.909
MPRNet [34]	31.76	**0.922**
MIMO-UNet+	31.92	0.919
MIMO-UNet++	**32.05**	0.921

Table 2. The average PSNR and SSIM on the RealBlur test dataset [25].

- MIMO-UNet: 8 residual blocks.
- MIMO-UNet+: 20 residual blocks.
- MIMO-UNet++: MIMO-UNet+ with geometric self-ensemble.

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Result (GoPro)

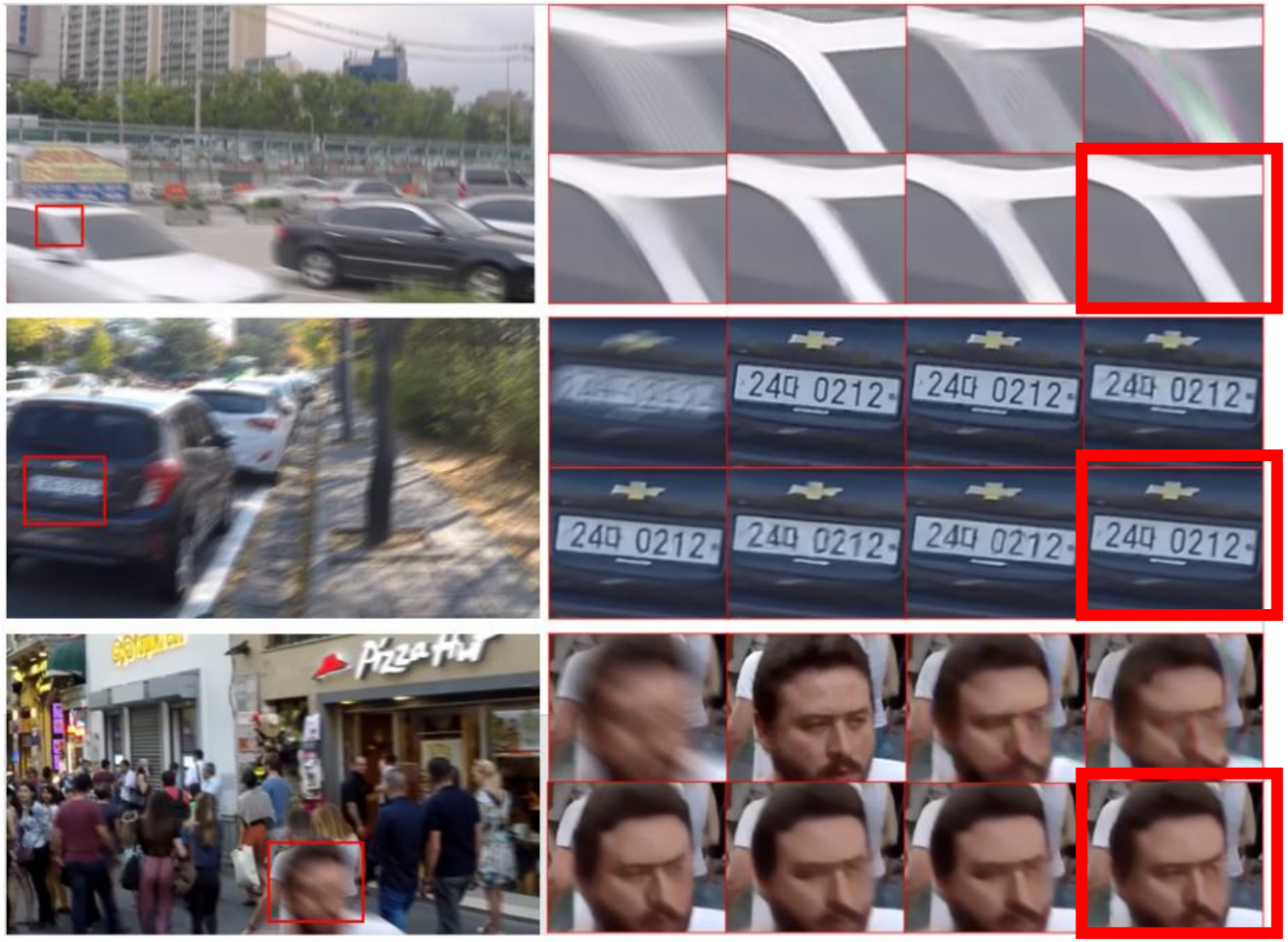


Figure 5. Several examples on the GoPro test dataset. For clarity, the magnified parts of the resultant images are displayed. From left-top to right-bottom: Blurry images, ground-truth images, and the resultant images obtained by SRN, PSS-NSC, DMPHN, MT-RNN, MPRNet, and MIMO-UNet++, respectively.

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Result (RealBlur)

Figure 6. Several examples on the RealBlur test dataset. For clarity, the magnified parts of the resultant images are displayed. From left to right: Blurry images, ground-truth images, and the resultant images obtained by DeblurGAN-v2, SRN, and MIMO-UNet++, respectively.

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

Reference

[1] Cho, S. J., Ji, S. W., Hong, J. P., Jung, S. W., & Ko, S. J. (2021). Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4641-4650).

[2] S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring,” Jul. 2017.

[3] Zhang, K., Ren, W., Luo, W., Lai, W. S., Stenger, B., Yang, M. H., & Li, H. (2022). Deep image deblurring: A survey. International Journal of Computer Vision, 130(9), 2103-2130.