Multi-Temporal Recurrent Neural Networks For Progressive Non-Uniform Single Image Deblurring With Incremental Temporal Training

Dongwon Park*, Dong Un Kang*, Jisoo Kim, and Se Young Chun

Department of Electrical Engineering, UNIST, Republic of Korea
{dong1,qkrtnskfk23,rlawltn1053,sychun}@unist.ac.kr

In this supplemental document, we present:

1. more details on generating our Temporal GoPro dataset (Sec. 1);
2. simple validations on the conjecture for our proposed multi-temporal (MT) approach (Sec. 2);
3. toy example of temporal data augmentation (Sec. 3);
4. further analysis on recurrent feature map (RFM) (Sec. 4);
5. detailed ablation studies to design our MT-RNN for different network structures (Sec. 5);
6. run time of our MT-RNN (Sec. 6);
7. quantitative results on Table 3-4 of our main paper (Sec. 7);
8. and more illustrative figures for qualitative results on Tables 2-4 of our main paper (Sec. 8).

1 More Details on Generating Temporal GoPro Dataset

Our Temporal GoPro dataset was generated as follows:

Algorithm 1: Generating Temporal GoPro dataset

Input: K consecutive sharp images $I_1, ..., I_n$
for $k = 1$ *to* n **do**
  $\hat{I}_k \leftarrow I_k^{\gamma}$
end
$\hat{B} \leftarrow \mathbf{Mean}(\hat{I}_1, ..., \hat{I}_n)$
$B \leftarrow \hat{B}^{1/\gamma}$

Using a set of consecutive sharp images $I_1, ..., I_n$, a blurred image B is synthesized through the above algorithm where n denotes the temporal level, TLn. The original GoPro dataste [31] approximated a non-linear CRF (camera response function) as a gamma curve with $\gamma = 2.2$. Our Temporal GoPro dataset was simply generated by using a linear CRF with $\gamma = 1$.

* Equal contribution

2 Validation of the Conjecture for MT Approach

We quickly validated our conjecture for MT approach: will it be easier to estimate TL1 from TL7 than to estimate TL1 from TL5 or TL3? Table S1 shows the performance of U-Net [37] that was trained only with one TL images. As TL increases, PSNR clearly decreases. Thus, our conjecture for MT approach seems reasonable.

Table 1. PSNR (dB) for single image deblurring using U-Net with input images with TL 3-13.

TL	3	5	7	9	11	13
PSNR (dB)	37.8	34.4	32.3	30.5	29.1	27.8

3 Toy Example of Temporal Data Augmentation

The non-uniform deblurring dataset is well described in Section 3.1 of our main paper. For better understanding of the generation of blurred images (TLn), we provide a toy example of our temporal data augmentation in the Fig. S1.

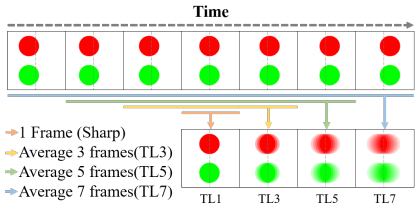


Fig. 1. An illustration for our temporal data augmentation. Blurred images (TLn) are generated by averaging n frames. Averaging more frames results in more severe blurs.

4 Further Analysis on Recurrent Feature Map

The recurrent feature map (RFM) is well explained and illustrated in Section 4.2 and Fig. 4 (right) of our main paper. For further analyzing our proposed MT-RNN, we visualized the averaged recurrent feature map for each iteration t as illustrated in Fig. S2.

Because of global residual skip connection, our MT-RNN is trained to estimate the blur region that should be deblurred. From Fig. S2, we observed that the network pays attention around edges for estimating blur regions and thus, the recurrent feature map contains more non-zero values as iteration t increases for representing more strong effect around edges for deblurring.

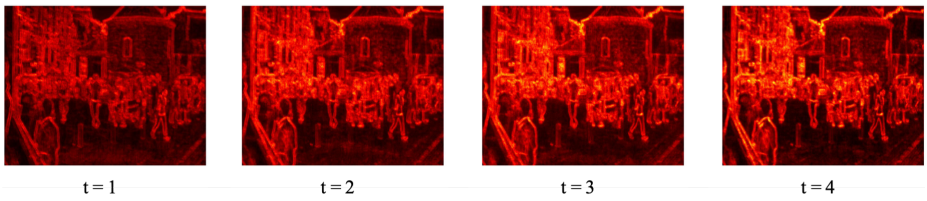


Fig. 2. Visualization of the averaged recurrent feature map (F_1^t) at different iterations t in the proposed MT-RNN.

5 Detailed Ablation Studies on Network Structures

We conducted three ablation studies to select the best network structure for our MT-RNN.

Firstly, we performed ablation studies from the baseline model (Tao [41]) by adding our proposed components such as residual learning (ResL) and kernel size (KerS) as shown in Table S2 (a),(b) and (c). Changing kernel size from 5×5 to 3×3 and using residual learning resulted in improved performances and substantially decreased overall parameter sizes.

Table 2. The ablation study from the base model with recurrent feature map on the Temporal GoPro dataset. The components of ablation study are kernel size (KerS), residual learning (ResL), and recurrent feature components (RFC) in 92×10^3 iterations. LSTM and GRU indicate Conv-LSTM and Conv-GRU, respectively, and RFM indicates our recurrent feature map.

Approach	KerS	ResL	RFC	PSNR (dB)	SSIM	Param (M)
(a)MS (Tao)	5	x	x	29.93	0.905	6.88
(b)MS	3	x	x	30.10	0.906	2.58
(c)MS	3	o	x	30.25	0.908	2.58
(d)MT	3	o	LSTM	29.08	0.883	2.99
(e)MT	3	o	GRU	30.53	0.912	2.99
(f)**MT (Ours)**	**3**	**o**	**RFM**	**30.82**	**0.917**	**2.64**

Secondly, we studied RNN models with MT approach as shown in Table S2 (d),(e) and (f). Previously, Conv-LSTM and Conv-GRU were utilized to prevent vanishing gradient in recurrent structures. However, using Conv-LSTM (d) and Conv-GRU (e) resulted in lower performance than our recurrent feature map (f) in the case of our temporal training.

Lastly, we investigated the effect of parameter size for performances as illustrated in Table S3. The number of parameters is proportional to performance with the cost of increased computation. While two times larger parameters in (h) did not seem to improve performance much over (i), its computation time and memory were substantially increased. Using half the parameter size in (g)

did degrade performance substantially while computation speed of (g) is similar to (i). Thus, we selected (i) as a proper number of parameters for our proposed MT-RNN.

Table 3. Investigation on the performances for different network parameters with our Temporal GoPro dataset in 92×10^3 iterations.

Approach	Param (M)	PSNR (dB)	SSIM	Time (sec)
(g)MT	1.46	30.21	0.908	0.060
(h)MT	5.35	30.84	0.918	0.290
(i)**MT**	**2.63**	**30.82**	**0.917**	**0.073**

6 Run Time of MT-RNN

As illustrated in Fig. S3, we observed that the run time of MT-RNN dramatically increases after the 6th iteration. While the first 6 iterations took only 0.07 seconds, the run time exponentially increases from the 7th iteration by about 0.2 seconds or more. This phenomenon seems to be related to GPU issues potentially and further investigation seems necessary. We selected 6 as the maximum iteration for reasonable performances and for fast computations.

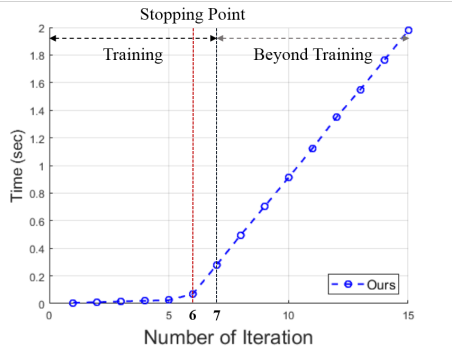


Fig. 3. Iteration vs. Run time (sec) for our proposed MT-RNN. The experiments are performed on the GoPro datasets [31].

7 Quantitative results of multi-temporal approach

To provide more solid information of Multi-Temporal (MT) approach on Table 3-4 of our main paper, we measured PSNR and SSIM on the GoPro test dataset

(1,111 images) [31] with respect to iterations and the evaluation results are shown in Tables S4 and S5. We investigated MT approach versions of Kupyn [24], Zhang [47] and Gao [12] on Table 3 of our main paper. Furthermore, "Ours" on the Table 4 of our main paper was also evaluated.

Table 4. Iteration vs. PSNR (dB) for MT version of Kupyn [24], Zhang [47], Gao [12] and Ours on the Table 3-4 of our main paper.

Test dataset	Iteration						
Method	1	2	3	4	5	6	7
Kupyn*	26.15	26.83	27.39	27.63	27.67	27.70	27.68
Zhang*	26.46	27.58	28.8	29.8	30.11	30.21	30.17
Gao*	26.35	27.46	28.74	29.83	30.18	30.32	30.29
Ours	26.38	27.48	28.80	30.18	31.03	31.15	31.14

Table 5. Iteration vs. SSIM for MT version of Kupyn [24], Zhang [47], Gao [12] and Ours on the Table 3-4 of our main paper.

Test dataset	Iteration						
Method	1	2	3	4	5	6	7
Kupyn*	0.813	0.840	0.853	0.859	0.860	0.860	0.860
Zhang*	0.831	0.856	0.881	0.901	0.909	0.910	0.910
Gao*	0.829	0.854	0.881	0.902	0.912	0.915	0.915
Ours	0.828	0.861	0.894	0.923	0.938	0.945	0.945

8 More Figures for Qualitative Results

More figures for qualitative results are presented in this Section. Fig. S4 illustrates our progressive deblurring results when using our MT approach with incremental temporal training with more examples.

Fig. S5 shows visual comparisons among One-Stage (OS), Stacking-Version (SV), Multi-Scale (MS) and our proposed Multi-Temporal (MT) approach with training 92×10^3 iterations on the GoPro dataset [31]. Our proposed method yielded deblurred images that are visually better than the results of others approaches for fine details.

Fig. S6 illustrates some powerful examples of our proposed MT approach over conventional approaches such as OS or MS that have been used in state-of-the-art methods: Kupyn [24], Zhang [47] and Gao [12]. Their corresponding modified methods using our MT approach yielded better visual performances than the original state-of-the-art methods on both GoPro dataset [31] for (a) and Su dataset [39] for (b) after training 92×10^3 iterations.

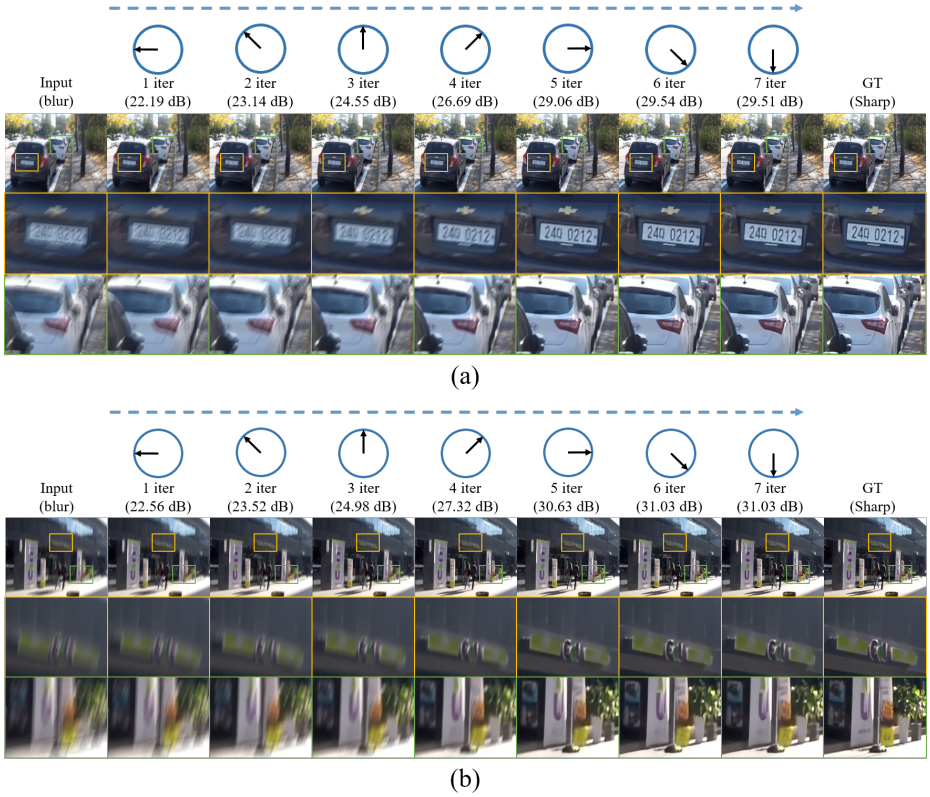
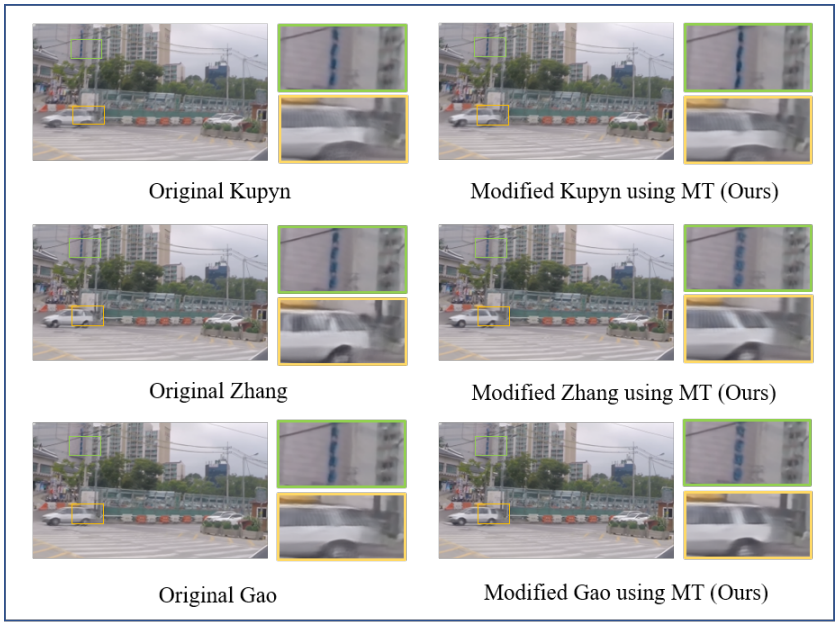


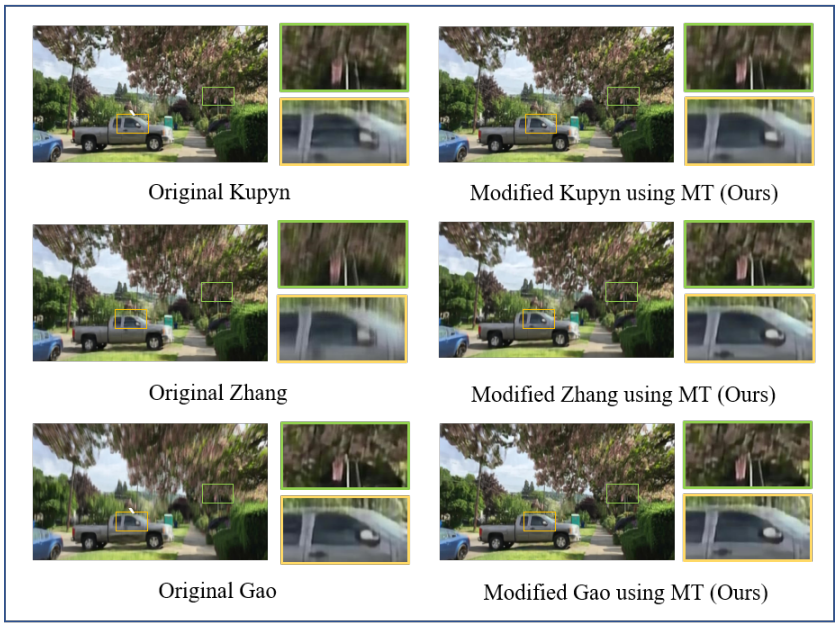
Fig. 4. Progressively deblurred images over iterations using our proposed MT-RNN after incremental temporal training.

Fig. S7 presents the visual comparisons among state-of-the-art methods (Zhang [47], Tao [41]) and our proposed MT-RNN on the GoPro benchmark dataset [31]. Our proposed method yielded deblurred images that are visually better than the results of other approaches for fine details.

Fig. 5. Visual comparisons among OS, SV, MS and our MT approaches on GoPro dataset [31]. Our MT method yielded better fine details than others.



(a) GoPro dataset [32]



(b) Su dataset [41]

Fig. 6. Visual comparisons between original state-of-the-art methods and their corresponding modified methods using our MT approach, yielding better performances than the original approaches such as OS or MS.

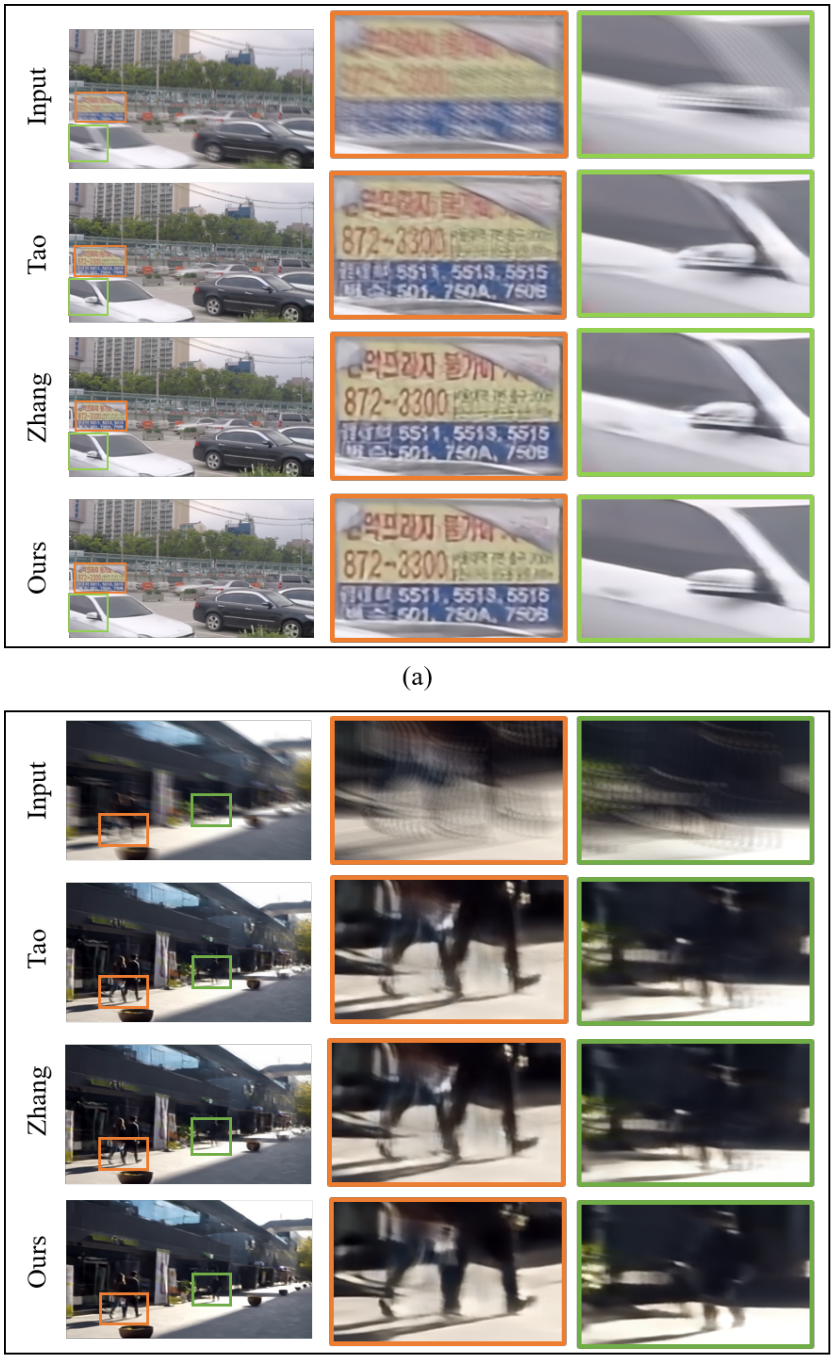


Fig. 7. Visual comparisons among state-of-the-art methods (Zhang [47], Tao [41]) and our proposed MT-RNN on GoPro dataset [31]. Our proposed method yielded visually better images than others for fine details.