Supplementary Materials to "High-Resolution Photorealistic Image Translation in Real-Time: A Laplacian Pyramid Translation Network"

Jie Liang[1,2]*, **Hui Zeng**[1,2]* and **Lei Zhang**[1,2]
[1]The HongKong Polytechnic University, [2]DAMO Academy, Alibaba Group
{*csjliang, cshzeng, cslzhang*}*@comp.polyu.edu.hk*

In this supplementary file, we provide the statistics of the high-resolution datasets, the training and user study details. We also provide more visual results to demonstrate the effectiveness and efficiency of our Laplacian Pyramid Translation Network (LPTN).

1. Statistics of the High-Resolution Datasets

We collect two high-resolution image-to-image (I2I) translation datasets in this paper. The detailed statistics of the datasets can be found in Table 1, and the examples of the training images are shown in Figure 1 and Figure 2.

2. Training Details

We set the batch size to be 1 during training in all experiments. We use the Adam optimizer [1] and set the exponential decay rate for the first-moment estimates as $\beta_1 = 0.5$ and the exponential decay rate for the second-moment estimates as $\beta_2 = 0.999$. The initialized learning rate is $1e-4$, which decays by 0.9 for every 10 epochs. We employ two losses including a self-reconstruction loss $\mathcal{L}_{recons}$ and an adversarial loss $\mathcal{L}_{adv}$. The total loss is a weighted sum of these two losses where we have $\lambda_{recons} = 10$ and $\lambda_{adv} = 1$. In the training stage, we employ a set of data augmentation techniques including random cropping and random flipping. We train 200 epochs on day→night and summer→winter tasks for LPTN and all compared methods and train 50 epochs for these methods on the photo retouching task.

3. User Study Details

As illustrated in the main paper, we conduct a user study with 20 participants (ph.d. and master students with diverse majors). Twenty images are randomly selected from the high-resolution photorealistic day→night translation dataset (thus 400 choices for each study). There are two questions for each image on its photorealism and aesthetic, respectively. We develop an interactive user interface as shown in Figure 3 to collect the preferences of each participant when facing the given compared images.

Specifically, given each image, we present the results of LPTN and the three comparative methods including CycleGAN, UNIT and MUNIT. For MUNIT, we randomly generate 30 translated results for each input and select the one with the highest inception score as input of the user study [2]. For each question (about photorealism and visual aesthetic), the participants are required, in a short time period, to select one image out of four which suits the question most. The order of the four results is randomly changed in each question to ensure a fair comparison.

4. More Visual Results of LPTN

We provide more results produced by the proposed LPTN method. Figure 4 shows the high-resolution day→night translation results and Figure 5 shows the high-resolution summer→winter translation results. Furthermore, Figure 6, Figure 7 and Figure 8 provide more high-resolution photo retouching results.

References

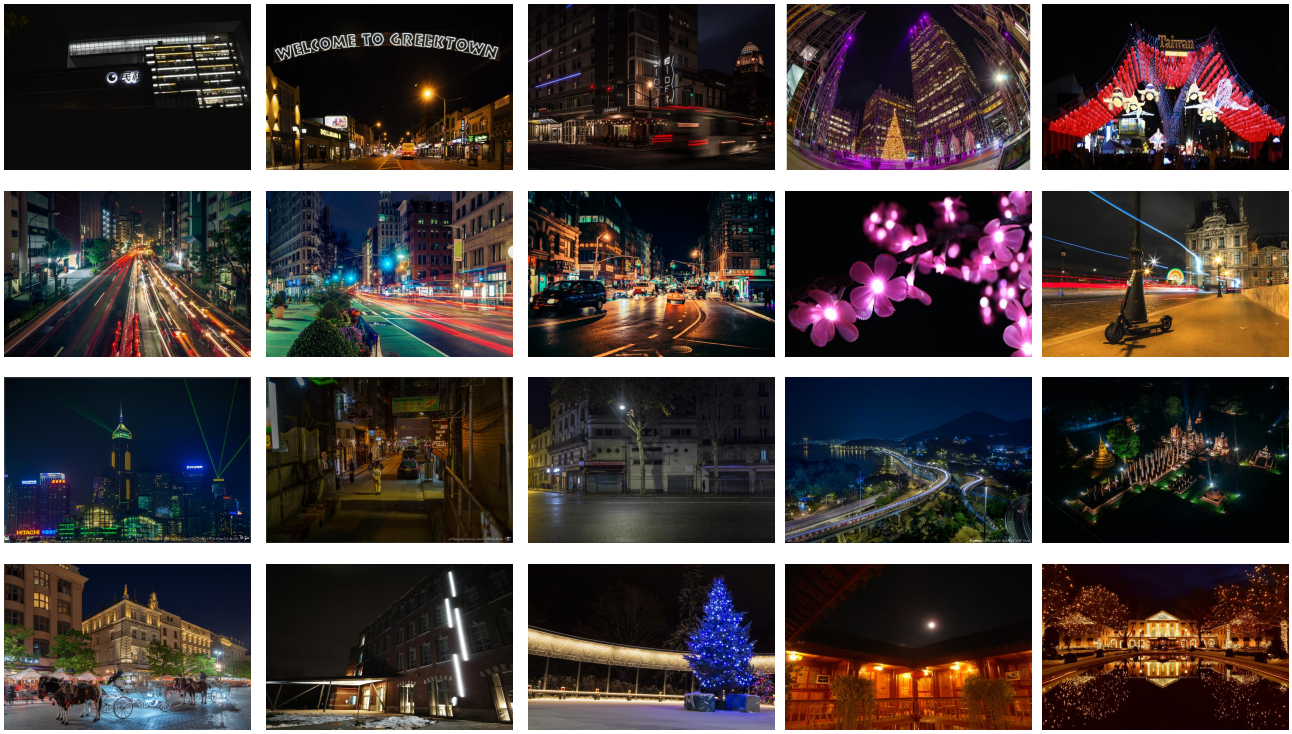
[1] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 1

Table 1. Statistics of the high-resolution I2I translation datasets collected in this paper. The training sets are of random resolutions ranging from 1080P (1920×1080) to 4K (3840×2160). For the day→night task, "A" denotes the set of day photos while "B" denotes the set of night photos. For the summer→winter task, "A" denotes the set of summer photos while "B" denotes the set of winter photos.

Datasets	Train A	Train B	Test A	Test B
Day→Night	1035	862	100	100
Summer→Winter	1173	1020	100	100

[2] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *NIPS*, 2016. 1

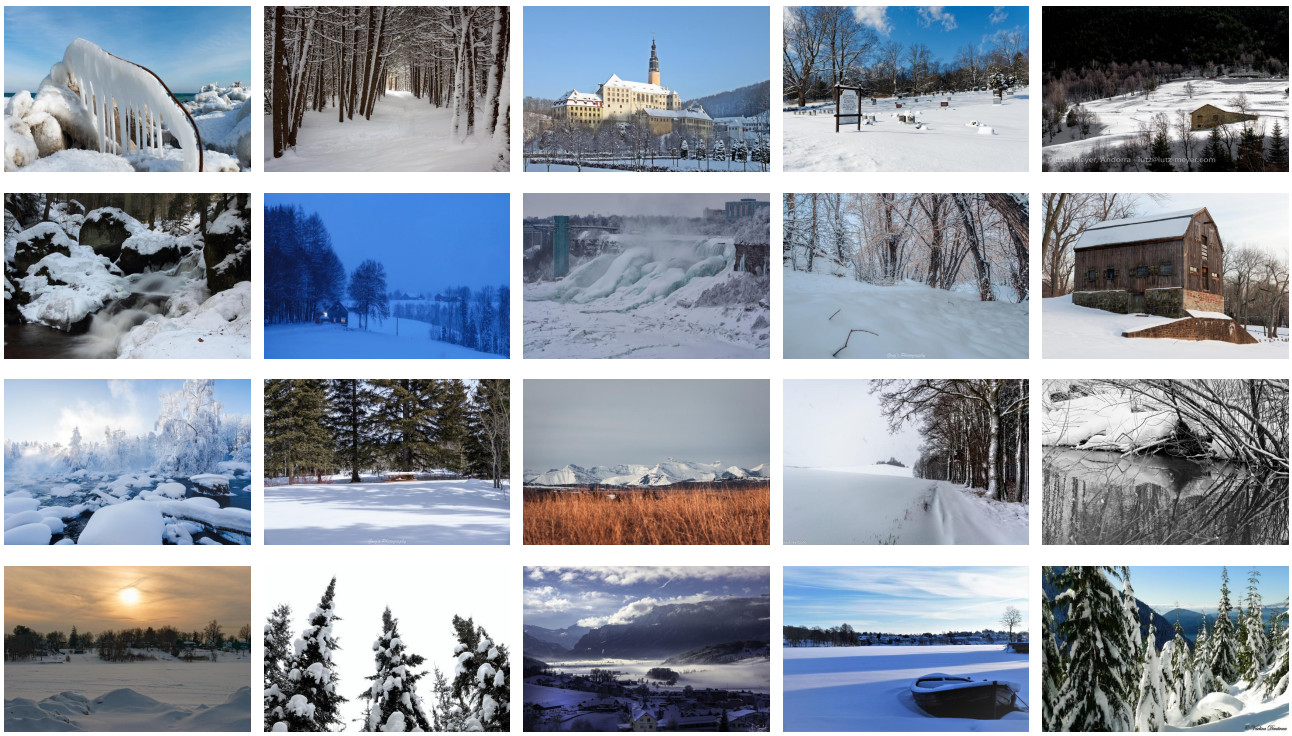
Day Images



Night Images

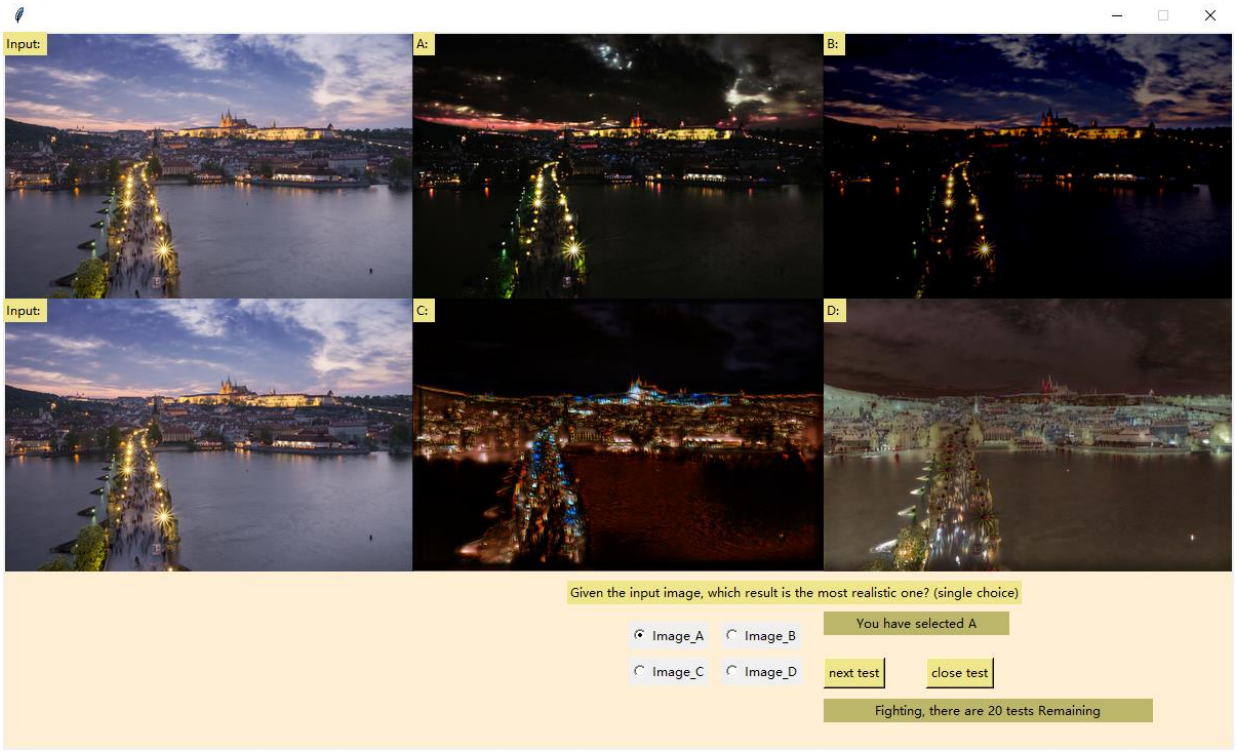
Figure 1. Example training images of the high-resolution photorealistic day→night translation task. The resolution of these images range from 1080P (1920 × 1080) to 4K (3840 × 2160). All images are collected from the Flickr website and are resized in this figure for better visulization.

Summer Images

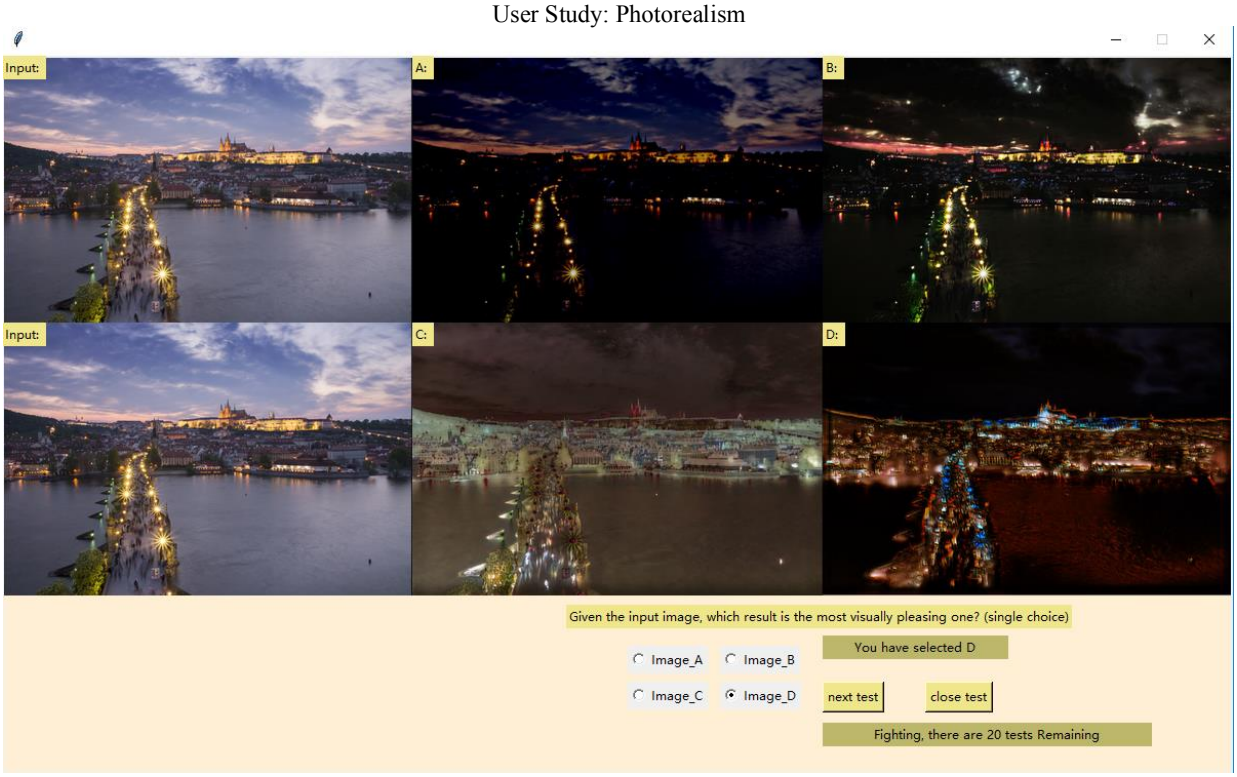


Winter Images

Figure 2. Example training images of the high-resolution photorealistic summer→winter translation task. The resolution of these images range from 1080P (1920×1080) to 4K (3840×2160). All images are collected from the Flickr website and are resized in this figure for better visulization.



User Study: Photorealism



User Study: Aesthetic

Figure 3. Interactive interface for our user study. The top figure shows the study on photorealism and the bottom one shows the study on aesthetic. For each study, the left column shows the input image (as a reference), while the middle and the right column shows the translation results of the four methods. As shown in the figure, the order of these candidates is randomly changed to make a fair comparison. All inputs and results are resized for better visulization.

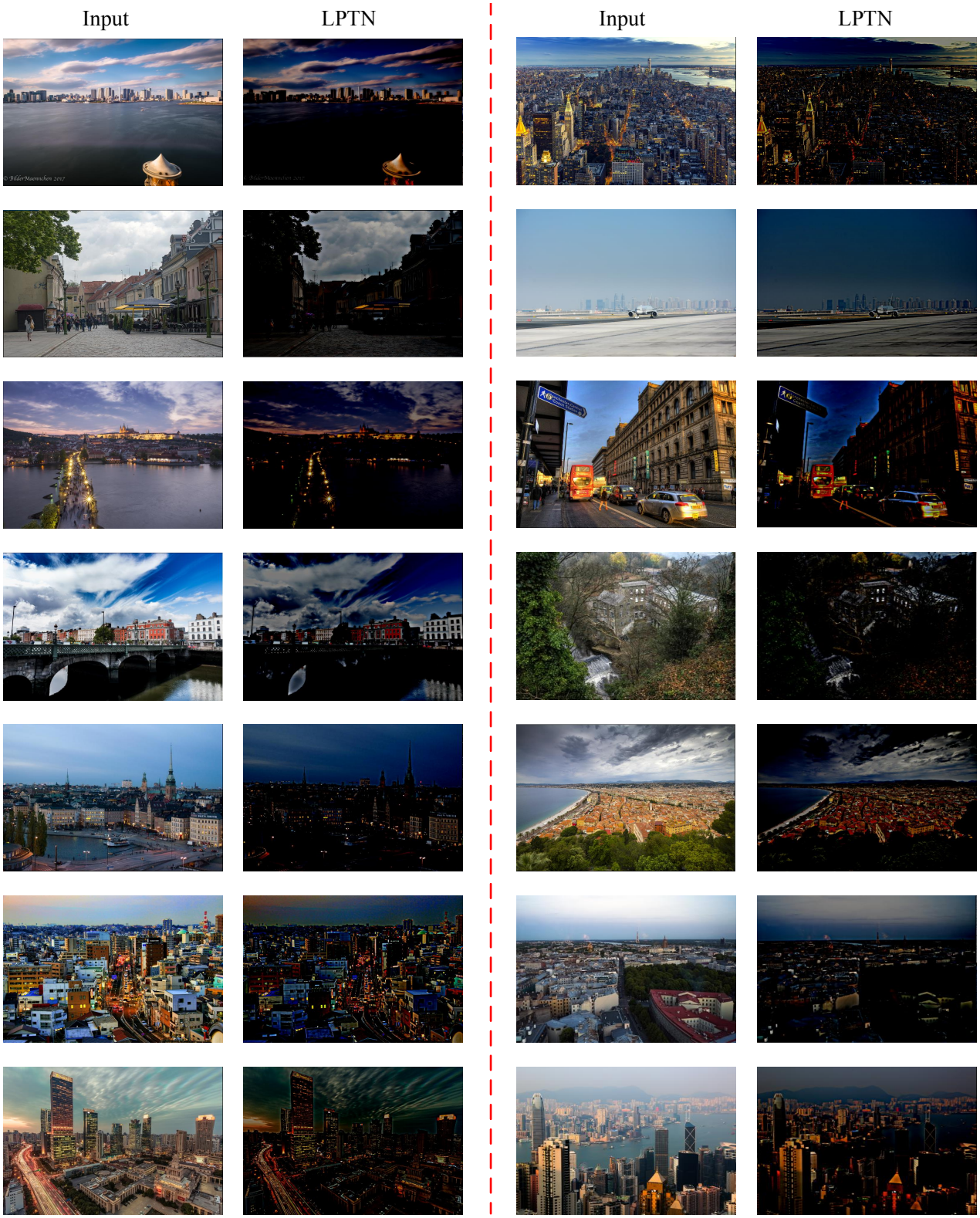


Figure 4. Translation results of the proposed LPTN method on high-resolution day→night translation task. The input images are of 4K (3840×2160) resolution and the LPTN takes about 0.03s to translate each image when $L = 4$. All inputs and results are resized in this figure for better visulization.

Figure 5. Translation results of the proposed LPTN method on high-resolution summer→winter translation task. The input images are of 4K (3840×2160) resolution and the LPTN takes about 0.03s to translate each image when $L = 4$. All inputs and results are resized in this figure for better visulization.

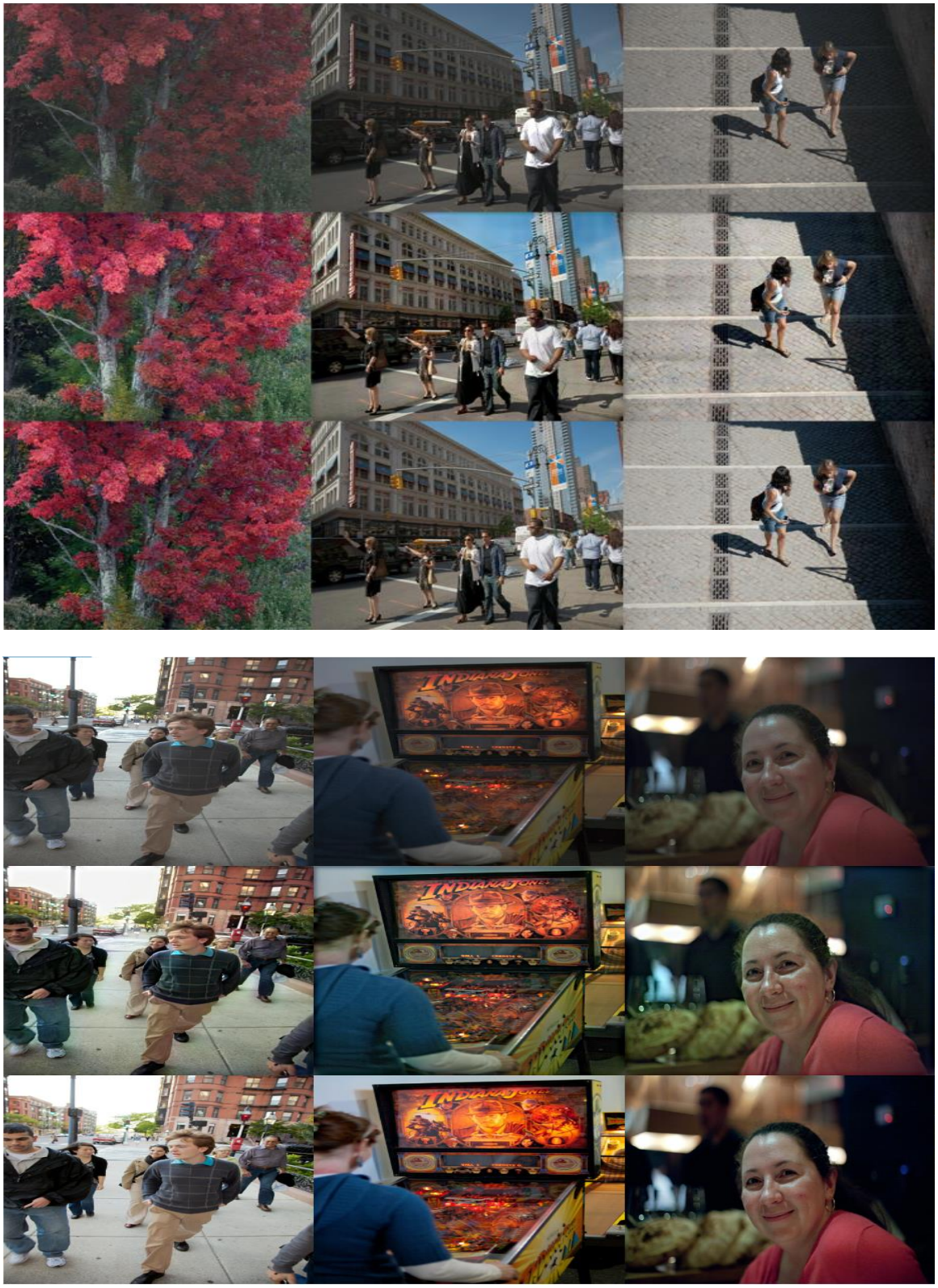


Figure 6. Translation results of the proposed LPTN method on high-resolution photo retouching task. For each column, the images from top to bottom are the input unaesthetic image, the results of the LPTN and the retouched results of the Expert C in the dataset. All inputs and results are resized in this figure for better visulization.

Figure 7. Translation results of the proposed LPTN method on high-resolution photo retouching task. For each column, the images from top to bottom are the input unaesthetic image, the results of the LPTN and the retouched results of the Expert C in the dataset. All inputs and results are resized in this figure for better visulization.

Figure 8. Translation results of the proposed LPTN method on high-resolution photo retouching task. For each column, the images from top to bottom are the input unaesthetic image, the results of the LPTN and the retouched results of the Expert C in the dataset. All inputs and results are resized in this figure for better visulization.