LMR — Supplementary Material

Lin Zhang* [1,3,4,5], Xin Li[2], Dongliang He[2], Fu Li[2], Errui Ding[2], Zhaoxiang Zhang[1,3,4,6]
[1] University of Chinese Academy of Sciences
[2] Department of Computer Vision Technology (VIS), Baidu Inc.
[3] Institute of Automation, Chinese Academy of Sciences
[4] State Key Laboratory of Multimodal Artificial Intelligence Systems
[5] School of Future Technology, UCAS
[6] Center for Artificial Intelligence and Robotics, HKISI_CAS
{zhanglin2019, zhaoxiang.zhang}@ia.ac.cn
lixin41@baidu.com, hedlcc@126.com, lifu.dlut@gmail.com, dingerrui@baidu.com

In this supplementary material, we introduce the training loss of MRefSR in Sec. 1. And we provide the full-resolution comparison (corresponding to Figure 1 in the paper) of C^2-Matching [2] and our MRefSR in Sec. 2. Then, we present more data samples of LMR in Sec. 3. At last, we show more qualitative results in Sec. 4.

1. Loss Functions

Our generative objective is formulated as

$$\mathcal{L}_G = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{per}\mathcal{L}_{per} + \lambda_{adv}\mathcal{L}_{adv}\,, \tag{1}$$

where λ_{rec}, λ_{per} and λ_{adv} are weight coefficients, assigned 1, 10-4 and 10-6.

Reconstruction loss. We adopt ℓ_1-norm to calculate loss between the ground-truth image X_{HR} and the output image X_{SR} , as

$$\mathcal{L}_{rec} = \|X_{HR} - X_{SR}\|_1\,. \tag{2}$$

Perceptual loss. Our perceptual loss is defined as

$$\mathcal{L}_{per} = \|\phi(X_{HR}) - \phi(X_{SR})\|_F\,, \tag{3}$$

where ϕ indicates the features obtained at the ReLU5_1 layer of the pretrained VGG19 model [4], and $\|\cdot\|_F$ denotes the Frobenius norm.

Adversarial loss. Our adversarial loss $\mathcal{L}_{adv}$ is expressed as

$$\mathcal{L}_{adv} = -\mathbb{E}[D(X_{SR})]\,, \tag{4}$$

where $D(x)$ is the probability of x being a real HR image, predicted by the discriminator D.

The training objective for D is

$$\begin{aligned}\mathcal{L}_D = {} & \mathbb{E}[D(X_{SR})] - \mathbb{E}[D(X_{HR})] \\ & + \lambda_p \mathbb{E}[(\|\nabla_{\hat{X}} D(\hat{X})\|_2 - 1)^2]\,,\end{aligned} \tag{5}$$

in which the last term is a penalization term of gradient norm and $\hat{X}$ is the random convex combination of X_{SR} and X_{HR}.

As is common for training the generative adversarial networks [1], we alternate one step between training our SR network and the discriminator.

2. Full-resolution visual comparison of C^2-Matching [2] and our MRefSR

Figure 1 is the full-resolution version of Figure 1 in the paper. It can be seen that our MRefSR effectively utilizes information from multiple reference images to produce visually pleasing details.

3. More Training Samples of LMR

In Figure 2, we present more examples of our LMR training dataset. It can be observed that our dataset is constructed with various scene contents and different levels of similarity.

4. More Visual Comparisons

In this section, we provide more visual results among the proposed MRefSR and the current top-performing methods, such as ESRGAN [5], MASA [3], and C^2-Matching [2]. Figures 3 and 4 show comparison results on the testing sets of CUFED5 and LMR, respectively. As shown in the figures, our model has good generalization performance in various scenes such as faces, buildings, animals, landscapes, texts, etc. Clearly, the results of our MRefSR exhibit far more realistic textures than other methods.

*Work done during an internship at Baidu Inc.

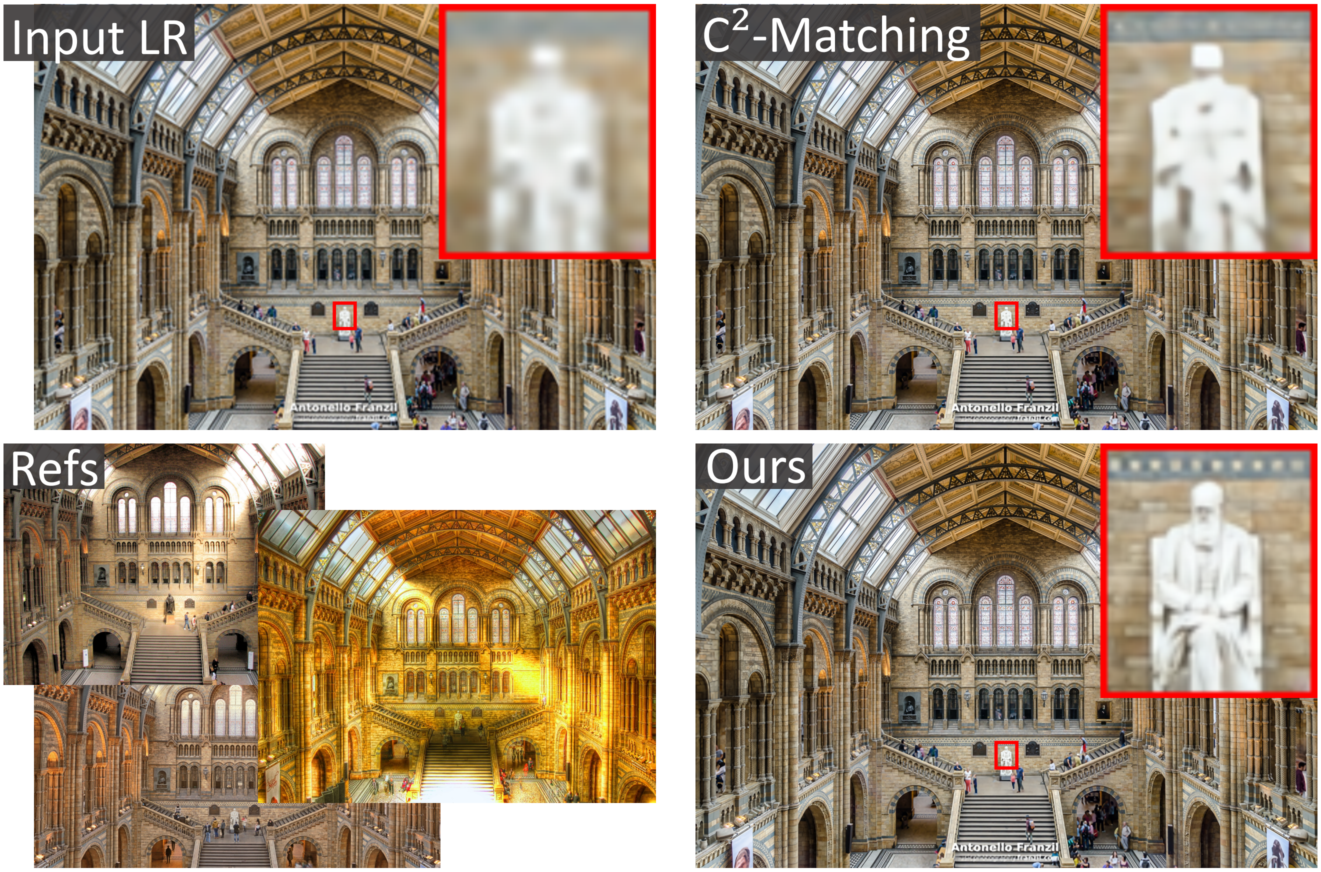

Figure 1. Visual comparison of single-reference training RefSR method C^2-Matching [2] and our multi-reference training MRefSR. Our MRefSR can more fully utilize arbitrary number of multiple reference images to achieve the best results. This figure is best viewed by zoom-in.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014. 1

[2] Yuming Jiang, Kelvin CK Chan, Xintao Wang, Chen Change Loy, and Ziwei Liu. Robust reference-based super-resolution via c2-matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1, 2

[3] Liying Lu, Wenbo Li, Xin Tao, Jiangbo Lu, and Jiaya Jia. Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. 1

[5] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018. 1

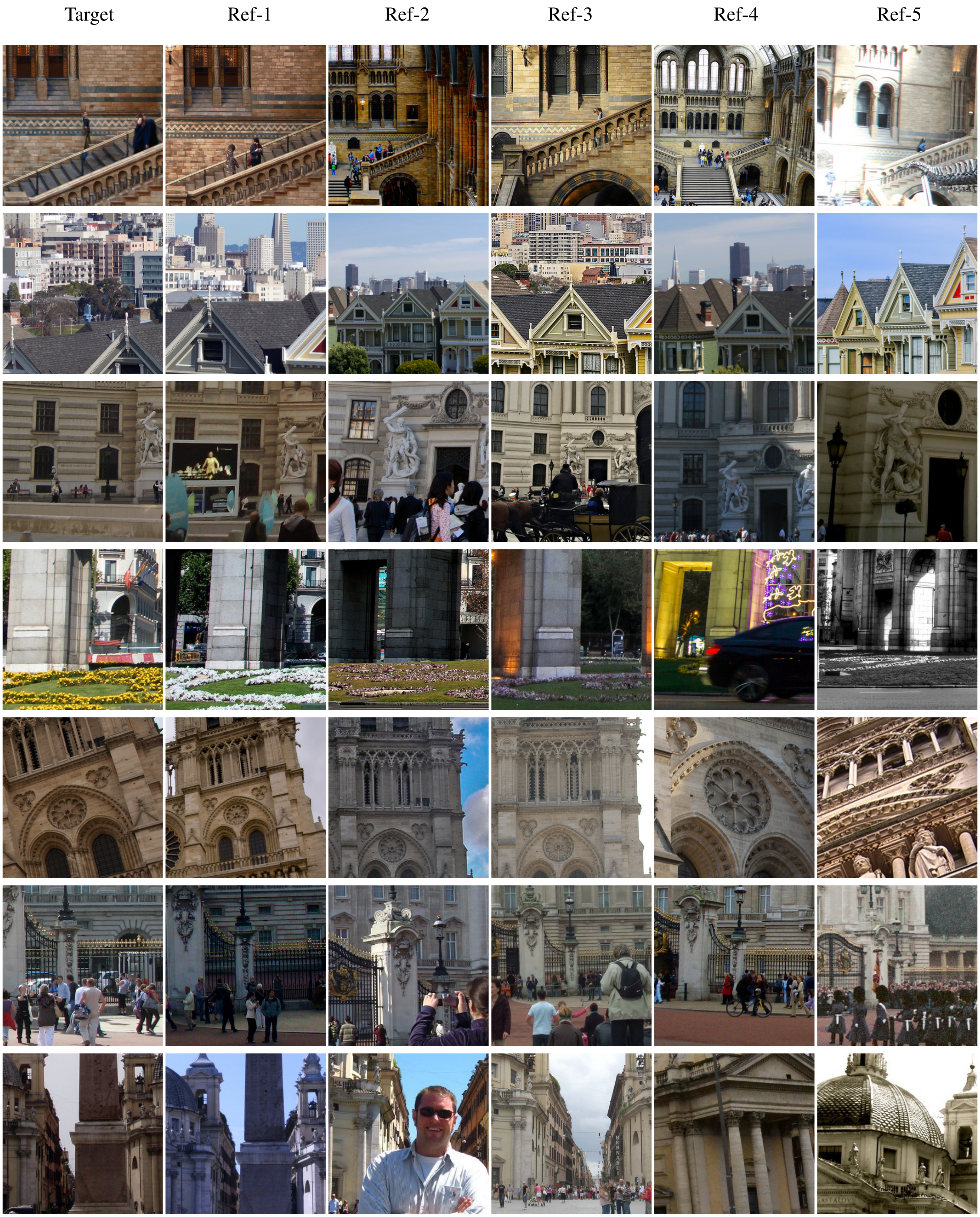

Figure 2. Examples from our LMR training dataset. From left to right, there is one target image, one high-similarity (H) reference image, two medium-similarity (M) reference images, and two low-similarity (L) reference images.

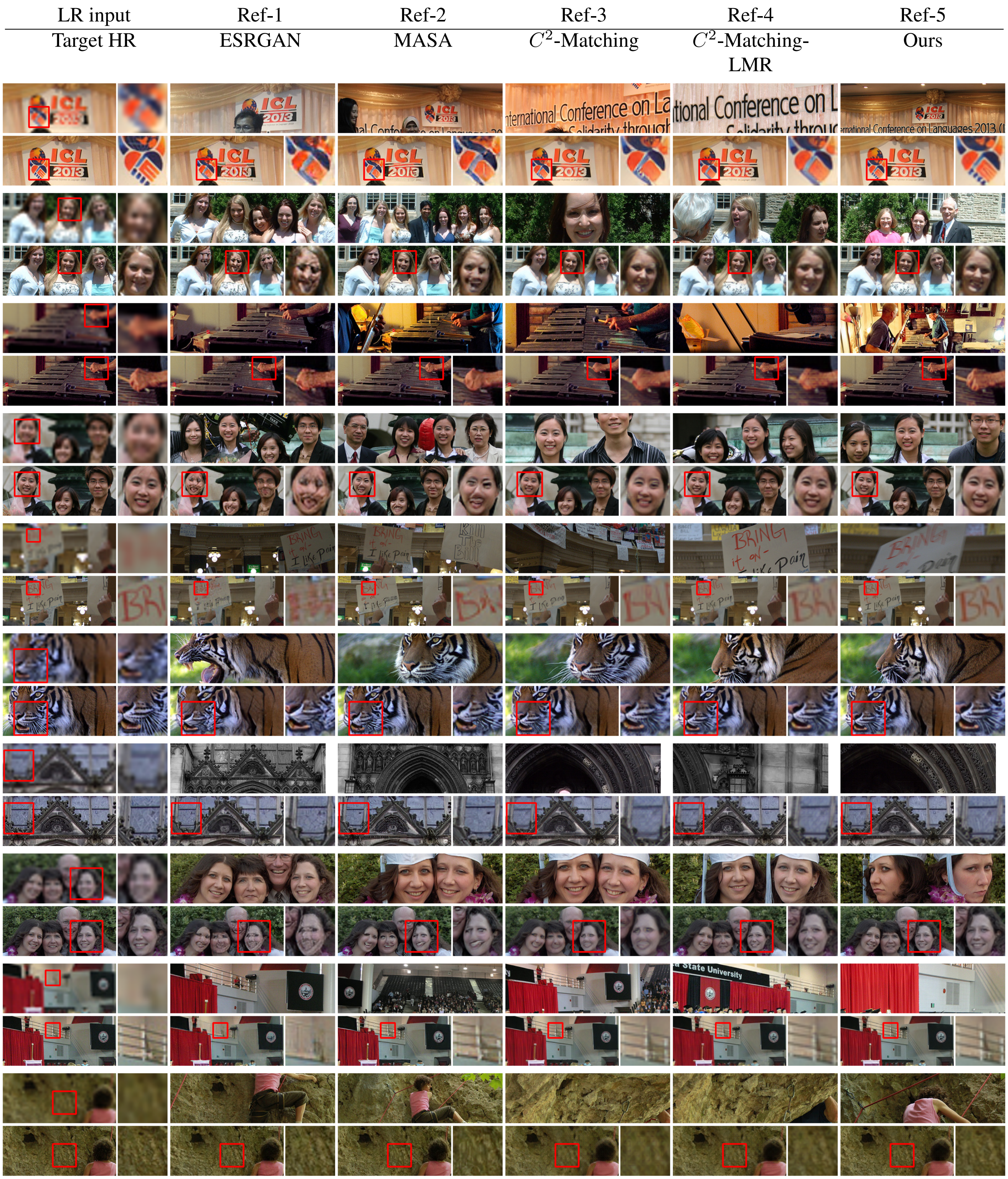

Figure 3. Qualitative comparisons on the testing set of CUFED5.

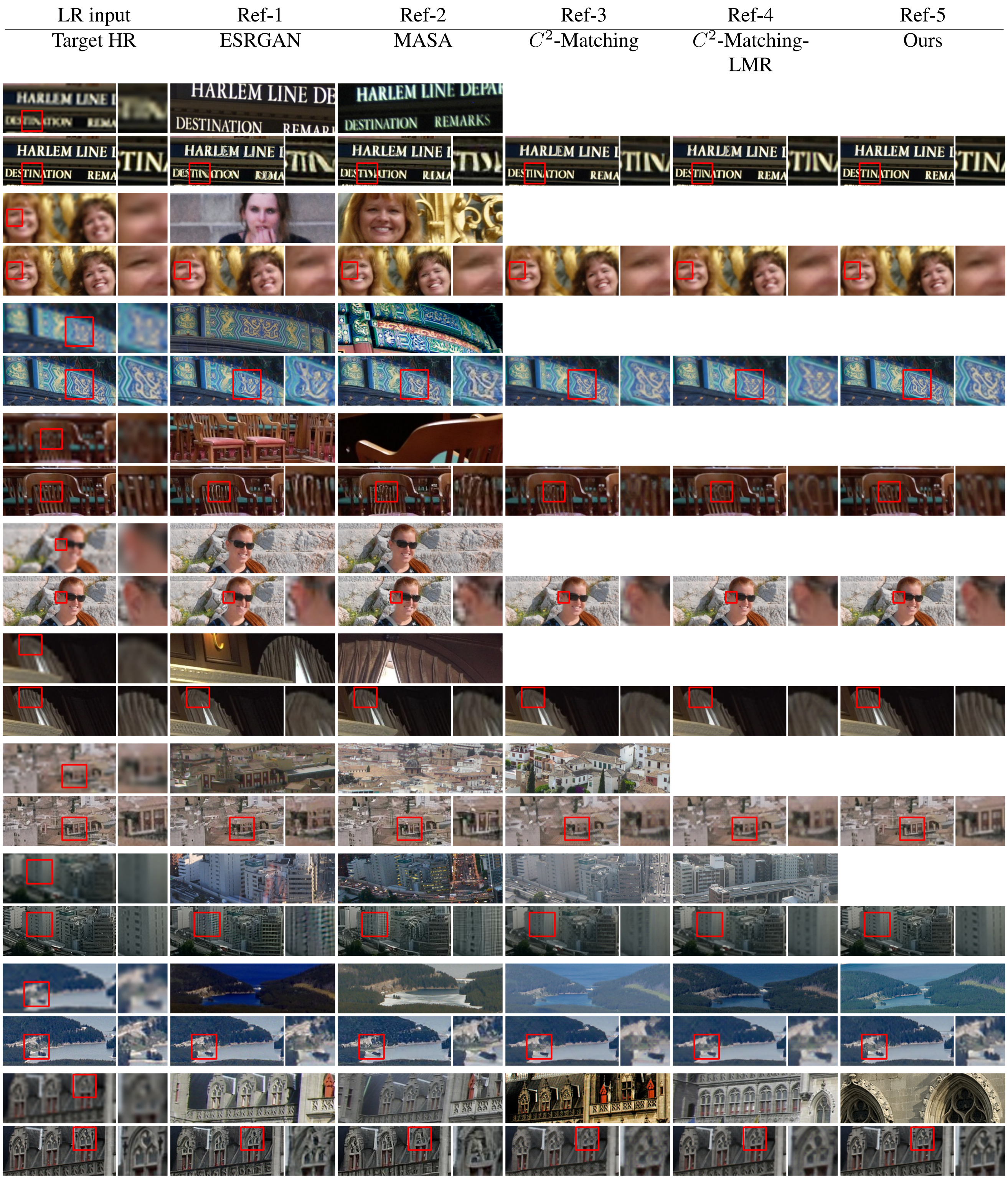

Figure 4. Qualitative comparisons on the testing set of LMR.