Towards Content-Independent Multi-Reference Super-Resolution: Adaptive Pattern Matching and Feature Aggregation

Supplementary Material

Xu Yan[1,⋆], Weibing Zhao[1,⋆], Kun Yuan[1,2], Ruimao Zhang[3], Zhen Li[1,†], and Shuguang Cui[1]

[1] Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, [2]University of Ottawa, [3]SenseTime Research
{xuyan1@link., weibingzhao@link., lizhen@}cuhk.edu.cn

A Overview

In this supplementary material, we first provide implementation details and more ablation studies about LFE and RP design in Sec. B and Sec. C, respectively. In Sec. D, the user study is conducted to further evaluate the visual quality of the generated HR image. At last, we provide visualizations of feature searching, the element in the reference pool, and the SR reconstruction results (4×) obtained by the SRCNN [1], MDSR [4], SRGAN [3], SRNTT [9] and our proposed CIMR-SR with a 4 upscaling factor in Sec. E.

B Implementation Details

B.1 Farthest Point Sampling Algorithm

Algorithm 1: Farthest Point Sampling Algorithm

Input: Feature point set $F^q \in \mathbb{R}^{N \times D^q}$; sampling number $N'(N' < N)$;
Output: Subset $\hat{F}^q \subsetneqq F^q$ ($\hat{F}^q \in \mathbb{R}^{N' \times D^q}$);

Randomly select a feature point $f_i^q \in F^q$ and $\hat{F}^q \leftarrow \{f_i^q\}$;
for $\ell = 1, \ldots, N' - 1$ **do**
  $\mathrm{j} \leftarrow \arg\max_j \sum_{f_k^q \in \hat{F}^q} ||f_j^q - f_k^q||_2, \forall f_j^q \in F^q \backslash \hat{F}^q$;
  $\hat{F}^q \leftarrow \hat{F}^q \cup \{f_j^q\}$;
end
Return $\hat{F}^q$;

B.2 Training Protocol

Following [9], data augmentation is performed on all training images and their corresponding indices, which are randomly rotated 90, 180, 270 degrees and flipped horizontally. In each training iteration, 8 LR color patches with the size

⋆ Equal first authorship. † Corresponding author.

of 40×40 (down-sampled from original 160×160 image) form one mini-batch. Our model is trained by ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. The learning rate is initialized as 10^{-4} and then halved every 5 epochs. All of the experiments are conducted under the Torch framework with the help of NVIDIA Titan XP GPUs.

C Ablation Study

C.1 Design of LFE Module

To verify the effectiveness of various components in the LFE module, we conduct ablation studies by removing or replacing specific components, i.e., number of key features in feature searching, Resblock, maxPool or average pool and similarity weights. As shown in Tab. 1, we firstly set two baseline models: A and G, which use the MDSR and the RCAN as the backbone respectively. In model B, if the model aggregates key features using 3 neighbours through average pool, only a tiny improvement can be achieved over baseline model A. Model C and model D show that if we add additional Resblock for feature enhancement, a significant improvement on both PSNR and SSIM can be observed. Furthermore, average pooling is a better choice compared with max pooling, since it can retain more information of key features. At last, if we further adopt similarity weights (model E) in LFE module, the proposed CIMR-SR achieves the best result presented in our paper. It is worth note that a larger K (model F with $K = 5$) cannot improve the result any more, which may result from noisy information existed in irrelevant neighbours.

Besides, as shown in the bottom part of Tab. 1, our LFE module can also significantly improve the performance against using the RCAN with high PSNR and SSIM as the backbone, which demonstrates the superior generalization capacity of our CIMR-SR model for any SR backbone.

Table 1. Ablation studies of using different components for the LFE module. PSNR/SSIM are calculated on the test set of Urban100 with RP size of 300. The number of LFE modules used in the model is 3.

Model	Backbone	K	Resblock	MaxPool	AveragePool	Weights	PSNR/SSIM
A	MDSR	-					25.51/.783
B	MDSR	3			✓		25.52/.783
C	MDSR	3	✓	✓			25.57/.785
D	MDSR	3	✓		✓		25.73/.790
E	MDSR	3	✓		✓	✓	**25.77/.792**
F	MDSR	5	✓		✓	✓	25.68/.788
G	RCAN	-					25.65/.785
H	RCAN	3	✓		✓	✓	**25.83/.794**

C.2 Efficiency of Farthest Point Sampling

To construct a universal reference pool, FPS is obligately applied to reduce the burden of computation in practice. As shown in Tab. 2, FPS can greatly reduce the time cost in feature searching while maintaining the performance.

Table 2. Feature searching (FS) time (time unit is s/image) and PSNR/SSIM on Sun80 dataset with RP size of 300. Here we present the results using different FPS ratio in the CIMR-SR architecture.

FPS Ratio	0	2	4	8	16	32
FS Time	28.48	22.12	18.21	11.16	9.32	**8.57**
PSNR/SSIM	**30.17/.816**	30.17/.816	30.15/.815	30.15/.814	30.07/.813	29.74/.809

C.3 Effect of RP size and FPS

In this section, we investigate the critical influence induced by the size of the reference pool on Urban100 benchmark. Here we create five different sizes of RPs which contain 20, 50, 100, 300, and 500 HR-Ref patches respectively. As shown in Tab. 3, with the increase of RefHRs in RP, the results of CIMR-SR will be progressively improved, achieving the best when RP size around 300. At the same time, it shows FPS can keep the diversity of RP compared with random sampling.

Table 3. PSNR with different sizes of RPs and sampling ratio on the Urban100 benchmark (denoted as Size/Ratio at the first row), where we only use an additional residual block without feature aggregation when RP size is 0. FPS and RS denote farthest point sampling and random sampling, respectively.

Sampling	0/0	20/2	50/4	100/8	300/16	800/32
FPS	25.54	25.60	25.65	25.74	**25.77**	25.75
RS	25.54	25.59	25.61	25.67	25.67	25.69

D User Study

To investigate the subjective assessment of reconstruction images, we further conduct a user study on benchmark datasets. In particular, 15 people are invited to evaluate image quality. All of them are asked to assign a score from 1 (i.e., Excellent) to 6 (i.e., Bad) for the reconstruction of the 4×bicubic down-scaling versions of images. For the validation set, 30 images from Sun80, Urban100 and CUFED5 test set are randomly selected. We compare our CIMR-SR against previous state-of-the-art methods (i.e., SRCNN [2], EDSR [4], SRGAN [3] and SRNTT [9]). For the fair comparison, SRNTT and CIMR both use GAN-based model in this experiment. Specifically, SRNTT and CIMR both exploit content-independent reference for Sun80 and Urban100 dataset, and leverage similar reference for CUFED5 test set. As shown in Fig. 1, compared with previous state-of-the-art methods, our proposed CIMR-SR achieves the highest subjective score (i.e., 18.4% rank-1 and 70.2% rank-2). It is worth note that the subjective score of EDSR is the lowest compared with other perceptual-based methods (i.e., SRGAN, SRNTT and ours). Actually, as a PSNR-oriented method, EDSR can achieve high metrics but suffer from generated unrealistic textures.

E Visualization

E.1 Visualization of Feature Searching

One of the critical parts in the proposed CIMR-SR is the Feature Searching strategy in the LFE module. Our method enables feature searching strategy to select

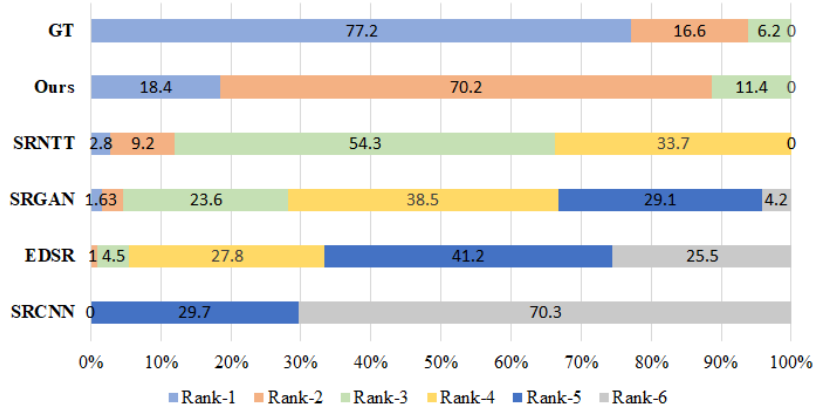


Fig. 1. The user study results, where our CIMR-SR achieves the best result among many appealing SR methods.

the most related local patterns according to the representation of query features. As shown in Fig. 2, we choose features from higher layer (e.g., `conv3_2`) as query features in feature searching to accelerate search efficiency. Then the auxiliary information is extracted from high-level representations and then applied for the subsequent feature aggregation by exploiting corresponding low-level key features in shallow layers.

Taking the first image in Fig. 2 as an example, the cloud region in LR image is usually matched with other sky clouds regions through feature searching. By taking advantage of the diversity-insurance sampling strategy, the patches corresponding to the building, sea and animal are also selected based on their high-level features extracted from `conv3_2` query features. In this way, the feature aggregation can promote the representation ability in different levels, and guarantee to generate more realistic HR images.

E.2 Visualization of Reference Pool

In this paper, multi-scale RPs are constructed by a certain number of 128×128 patches (i.e., from 20 to 800) selected from Outdoor Scene (OST) dataset [7] . We show a part of HR patches of our RP in Fig. 3, which contains diverse high resolution texture information. It guarantees that our model can reconstruct more realistic HR images after feature searching and LFE module.

E.3 Additional Visualization Results

We present more visualization results in this section. In Fig. 4 to Fig. 7, we illustrate the results of using similar reference images on CUFED5 dataset, in which `original` denotes ground-truth high resolution images, `Reference` indicates provided reference images and others are reconstructions from various methods. Moreover, reconstruction results of exploiting content-independent references are shown in Fig. 8 to Fig. 10, which are selected from Sun80, Set14 [8], B100 [5,6] dataset respectively.

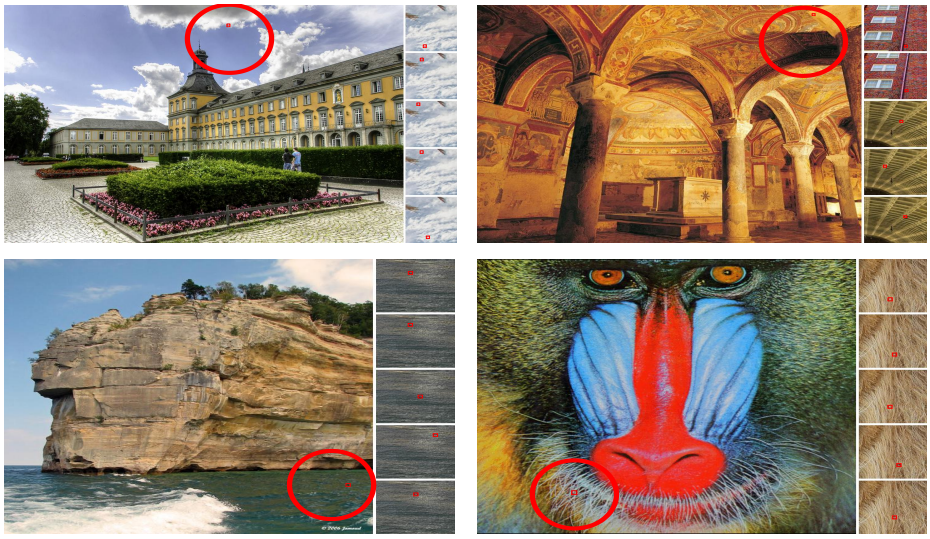


Fig. 2. Visualization of feature searching strategy with RP of size 20.

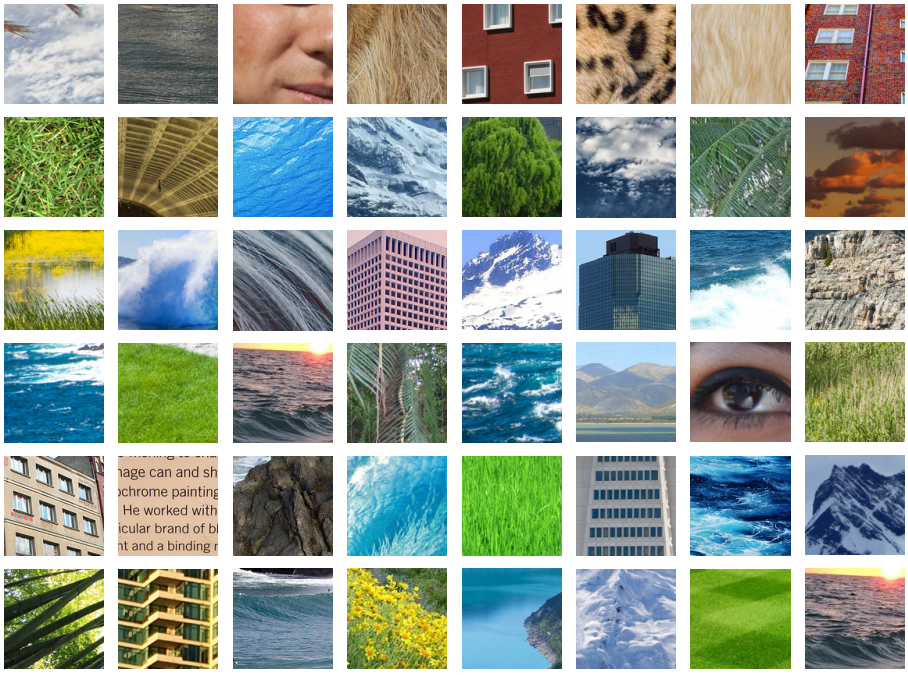
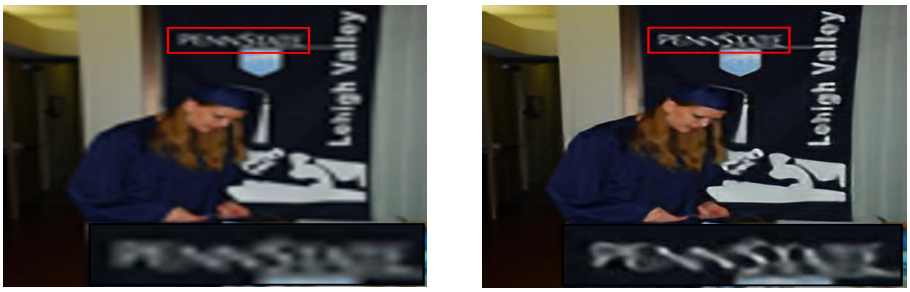


Fig. 3. Selected patches in Reference Pool (RP). The RP is constructed by a certain number of 128×128 patches cropped from Outdoor Scene (OST) dataset [7] containing 7 categories (i.e., animal, sky, grass, plant, mountain, water, building) with rich textures.

(a) Original (b) Reference



(c) Bicubic (d) SRCNN

(e) MDSR (f) SRGAN

(g) SRNTT (h) Ours

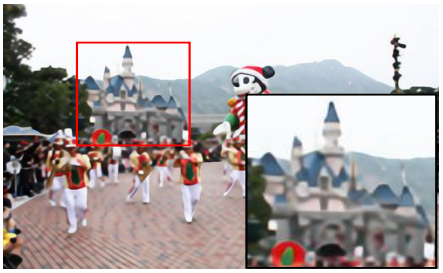
Fig. 4. Comparison of different SR algorithms on CUFED5 dataset (055).

(a) Original

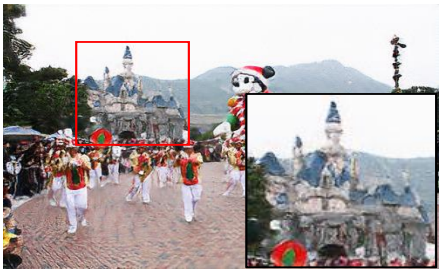
(b) Reference

(c) Bicubic

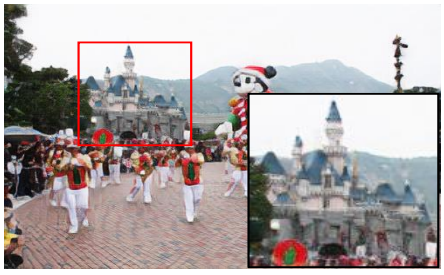
(d) SRCNN



(e) MDSR



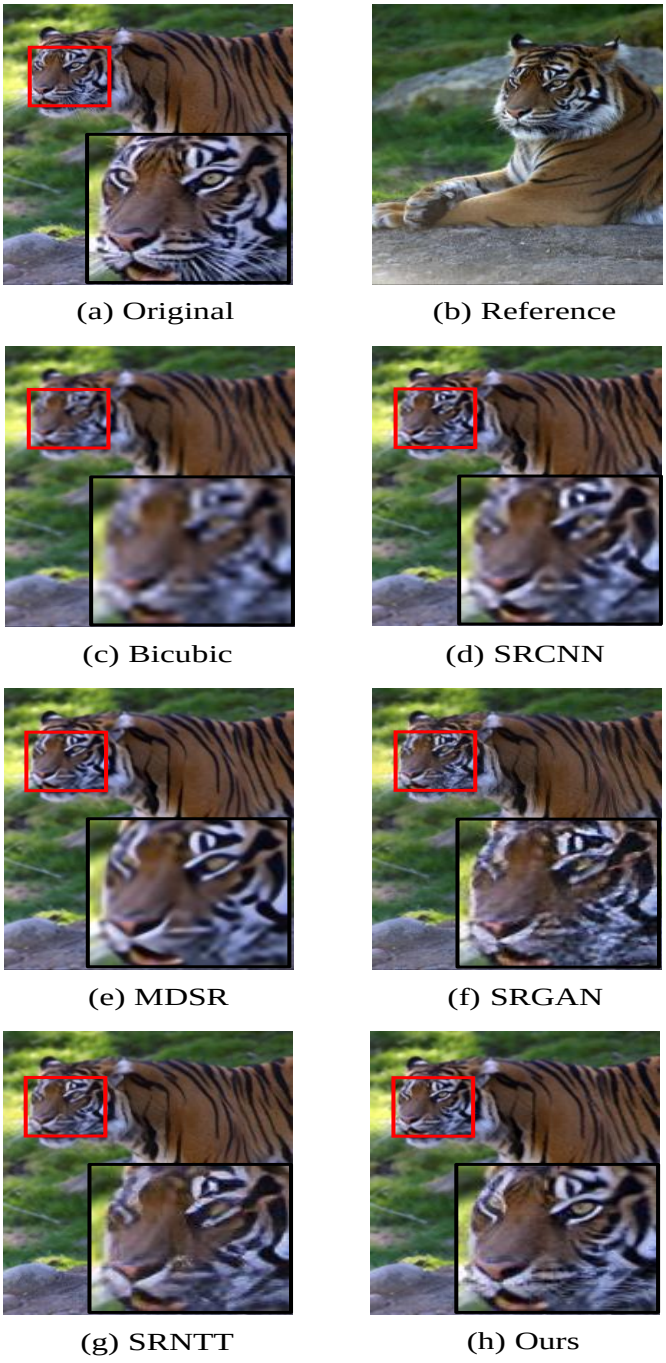
(f) SRGAN



(g) SRNTT

(h) Ours

Fig. 5. Comparison of different SR algorithms on CUFED5 dataset (065).



(a) Original (b) Reference

(c) Bicubic (d) SRCNN

(e) MDSR (f) SRGAN

(g) SRNTT (h) Ours

Fig. 6. Comparison of different SR algorithms on CUFED5 dataset (121).

Fig. 7. The samples are selected from Sun80, where CIMR and SRNTT use content-independent references.

Fig. 8. The samples are selected from Set14 [8], where CIMR and SRNTT use content-independent references.

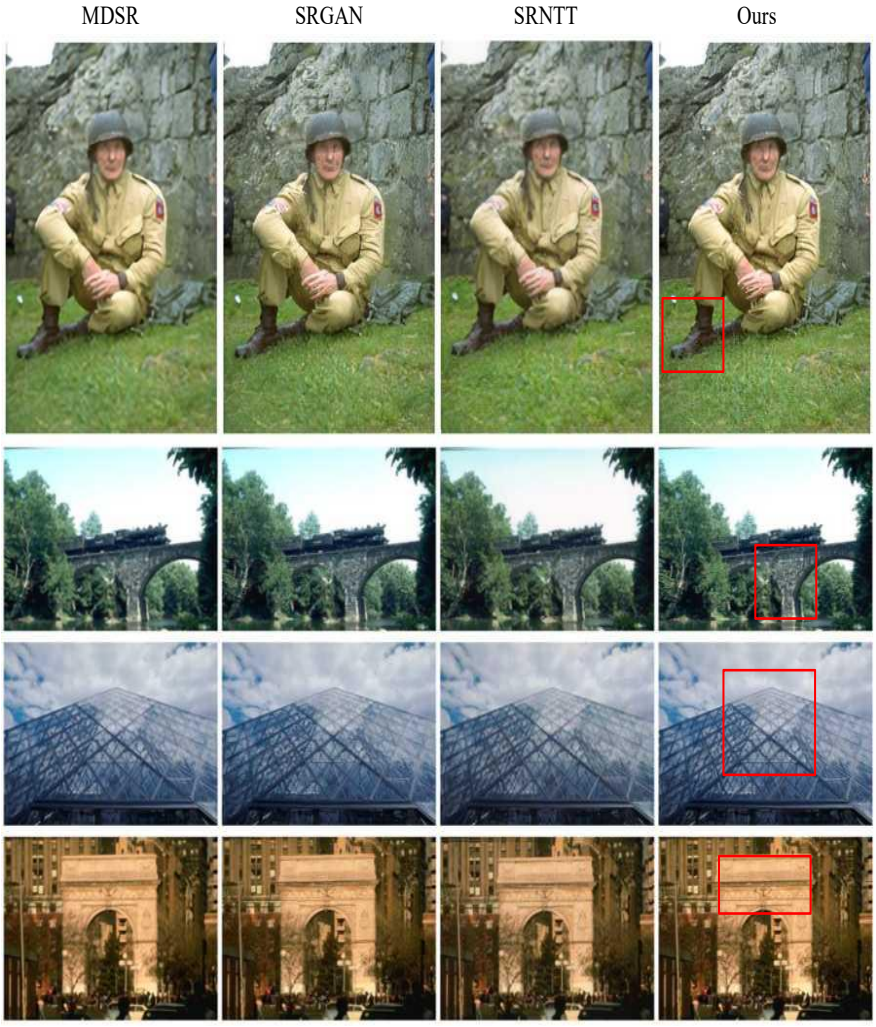


Fig. 9. The samples are selected from B100 [5,6], where CIMR and SRNTT use content-independent references.

References

1. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: European conference on computer vision. pp. 184–199. Springer (2014)
2. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38**(2), 295–307 (2015)
3. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4681–4690 (2017)
4. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
5. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. vol. 2, pp. 416–423 (July 2001)
6. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2, pp. 416–423. IEEE (2001)
7. Wang, X., Yu, K., Dong, C., Change Loy, C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 606–615 (2018)
8. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: International conference on curves and surfaces. pp. 711–730. Springer (2010)
9. Zhang, Z., Wang, Z., Lin, Z., Qi, H.: Image super-resolution by neural texture transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7982–7991 (2019)