Fact Sheet for "NTIRE 2024 Light Field Image Super-Resolution Challenge - Track 1 Fidelity Only"

Wentao Chao[1] Yiming Kan[1] Xuechun Wang[1] Fuqing Duan[1] Guanghui Wang[2]
[1] Beijing Normal University [2] Toronto Metropolitan University

1. Team Name

Team Name: BNU&TMU-AI-TRY.
CodaLab User Name: @wentaochao

2. Method

Method Name: BigEPIT.
Introduction: *Please describe the motivations, technical details (including but not limited to network architectures, data augmentation or preprocess, and training details), and the contributions of the proposed method. Note that, a figure of the overall network architecture is* ***REQUIRED****.*

In order to better solve the disparity problem of LF image super-resolution, we propose a novel method, called BigEPIT, including the network, data augmentation, and test scheme. Our network is is built on EPIT [5]. We use a Transformer network with repetitive self-attention operations to learn the spatial-angular correlation by modeling the dependencies between each pair of the epipolar plane image (EPI) pixels. Specifically, we increase the channel and width of feature maps to improve the model capability. In the existing training data, there is an imbalance in disparity and large disparity data are scarce. Therefore, we use the resampling method, which manually specifies different sampling intervals when generating the training data rather than just using the view of the central region. Meanwhile, the augmented of training data helps to alleviate the problem of overfitting the network. At present, the test scheme is carried out in a cropped strategy and 0 is padding in the surrounding region. Padding 0 destroys the disparity of the LF image, resulting in lower results. Therefore, 0 cannot be filled with the cropped scheme. We also find that the results of the model are positively correlated with the patchsize of the crop. Finally, we used a full-size test scheme to get better results if GPU memory is available.

2.1. Network Architecture

An overview of our BigEPIT is shown in Fig. 1. Our network takes an LR $\mathcal{L}_{LR} \in \mathbb{R}^{U\times V\times H\times W}$ as its input and produces an HR LF $\mathcal{L}_{HR} \in \mathbb{R}^{U\times V\times \alpha H\times \alpha W}$ where α presents the upscaling factor. Our network consists of three stages including initial feature extraction, non-local cascading block, and spatial upsampling.

Initial Feature Extraction: We follow the EPIT [5] to use three 3×3 convolutions with LeakyReLU to map each SAI to a high-dimensional feature. The initially extracted feature can be represented as $F \in \mathbb{R}^{U\times V\times H\times W\times C}$, where C denotes the channel dimension. Different from EPIT, we increase the channel of feature maps to improve the model capability ($64 \rightarrow 128$).

Non-Local Cascading Block: Different to EPIT [5], we employ ten Non-Local Cascading blocks to achieve a global perception of all angular views and follow SwinIR [4] to adopt spatial convolutions to enhance the local feature representation. The non-local cascading block consists of a horizontal EPI transformer, a vertical EPI transformer, and spatial convolutions sequentially. Note that the weights of the two basic transformer units in each block are shared, which can help teach the spatial-angular correlation better.

Spatial Upsampling: Following EPIT [5], we apply the pixel shuffling operation to increase the spatial resolution of LF features, and further employ a 3×3 convolution to obtain the super-resolved LF image work. We also use the L1 loss function to train our network, due to its robustness to outliers. We convert input images into the YCbCr color space, and only super-resolve the Y channel of images, leaving Cb and Cr channel images being bicubicly upscaled.

2.2. Training

Data Augmentation: All LFs in the released datasets have an angular resolution of 9×9. Inspired by [2], we also use the augmented data sampling strategy to extract 5×5 SAIs for training and testing, as shown in Fig. 2, including central sampling, even sampling, and uneven sampling. This strategy explicitly increases the number of images of large disparity light fields, which can improve the robustness of the model to disparity changes, but increases the training time by three times. In the training stage, we cropped each SAI into patches of size 128×128 with a stride of 32 and used the bicubic downsampling approach to generate LF patches of size 32×32. We performed random horizontal flipping, vertical flipping, and 90-degree rotation to aug-

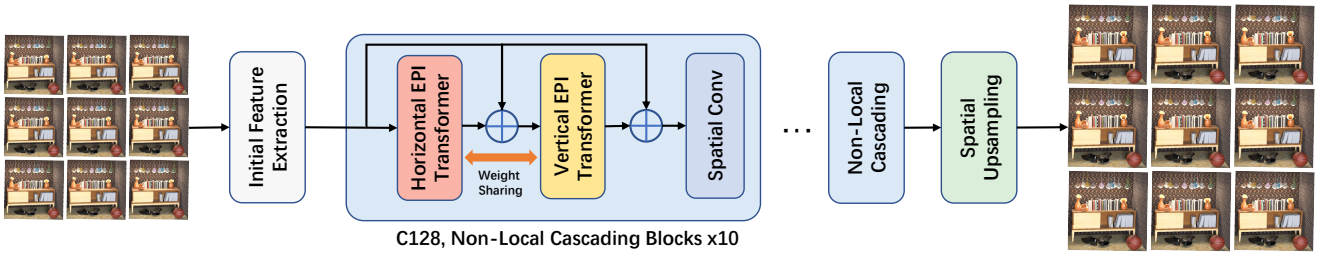

Figure 1. An overview of our BigEPIT network. Here, a 3×3 LF is used as an example for illustration.

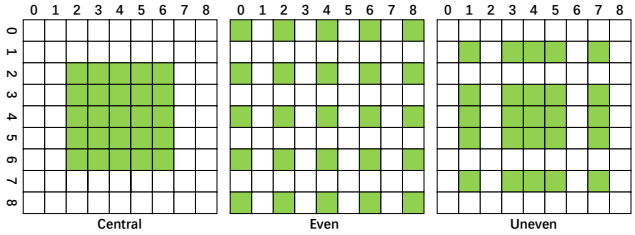

Figure 2. Illusatrtion of the augmented data sampling strategy, including central sampling, even sampling, and uneven sampling

ment the training data. Note that, the spatial and angular dimensions need to be flipped or rotated jointly to maintain LF structures.

Regularization: Our network was trained using the L1 loss and optimized using the Adam method [3] with β_1=0.9, β_2=0.999, and a batch size of 8. Our BigEPIT was implemented in the framework PyTorch-based *BasicLFSR* on a cluster with four NVIDIA A100 GPUs. The learning rate was initially set to 2×10^{-4} and decreased by a factor of 0.5 for every 15 epochs. The training was stopped after 31 epochs, we selected the best model according to the performance on the validation set.

2.3. Testing

In the testing experiments, we find that LF patches introduced by the commonly used zero-padding in the LF divide-and-integrate operation[1] can degrade the model performance since it disrupts the disparity structure of the LF image. Therefore, we use full-size images as input if GPU memory is available. We utilize 8-set spatial self-ensemble strategies [6] to improve the final results. We feed augmented input LFs independently to the network, including horizontal flip, vertical flip, and rotation, and use the average outputs as predictions. Besides data self-ensemble, we also use a conventional multi-model ensemble strategy to further improve the result, including BigEPIT (Ours), DistgEPIT_d_w [2] and RR-HLFSR [1]. DistgEPIT_d_w and RR-HLFSR and trained with augmented data sampling strategy from scratch. However, the multi-model ensemble is time-consuming and gets a little improvement.

[1] https://github.com/ZhengyuLiang24/BasicLFSR

3. Members

Wentao Chao (chaowentao@mail.bnu.edu.cn),
Yiming Kan (kanyiming@mail.bnu.edu.cn),
Xuechun Wang (wangxuechun@mail.bnu.edu.cn),
Fuqing Duan (fqduan@bnu.edu.cn), and
Guanghui Wang (wangcs@torontomu.ca).

The first member will be referred to as the captain of the team.

4. Affliation

Beijing Normal University, Toronto Metropolitan University.

5. Result Link

Please attach a download link, e.g., OneDrive, Baidu Drive or Google Drive, to the final SR results on the test set.

Link 1: https://pan.bnu.edu.cn/l/a1Sbr3

Link 2: https://pan.baidu.com/s/1HEoTQnr47AjeNGcmtSvTaw?pwd=LFSR

6. Submission File

Please create a ZIP archive containing this latex file, the original image files for all figures and the codes (with pretrained models and a readme file to clarify how to run the codes), and then submit this ZIP file to the official submission account (ntire.lfsr@outlook.com).

References

[1] V. V. Duong, T. H. Nguyen, J. Yim, and B. Jeon. Light field image super-resolution network via joint spatial-angular and epipolar information. *IEEE Transactions on Computational Imaging*, 2023. 2

[2] Kai Jin, Angulia Yang, Zeqiang Wei, Sha Guo, Mingzhi Gao, and Xiuzhuang Zhou. Distgepit: Enhanced disparity learning for light field image super-resolution. In *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1373–1383, 2023. 1, 2

[3] DiederikP Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Proceedings of the International Conference on Learning and Representation (ICLR)*, 2015. 2

[4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1

[5] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, Shilin Zhou, and Yulan Guo. Learning non-local spatial-angular correlation for light field image super-resolution. *arXiv preprint arXiv:2302.08058*, 2023. 1

[6] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1865–1873, 2016. 2