Toward Real-World Light Field Super-Resolution
—— Supplementary Material ——

Zeyu Xiao* Ruisheng Gao* Yutong Liu Yueyi Zhang Zhiwei Xiong†
University of Science and Technology of China
{zeyuxiao,grsmc4180,ustclyt}@mail.ustc.edu.cn {zhyuey,zwxiong}@ustc.edu.cn

Overview

This supplementary document is organized as follows:
Section 1 provides detailed structures of frequency scale-up and scale-down blocks.
Section 2 provides detailed structures of the frequency up-projection unit and the frequency down-projection unit.
Section 3 provides more visual results to demonstrate the effectiveness of the frequency decomposition in OFPNet.
Section 4 provides an overview of the content from our collected LytroZoom dataset.
Section 5 provides more visual comparison results.

1. Details of the Frequency Scale-Up/-Down Blocks

The frequency scale-up (FSU) block aims at projecting the extracted frequency feature to corresponding high-resolution (HR) representation, and the frequency scale-down (FSD) block can back-project the HR representation back to the low-resolution (LR) one. Detailed structures of both blocks are shown in Figure 1.

Without losing the generality, we omit the superscript and subscript here to explain how the FSU block works. Given the extracted frequency feature $\mathcal{F} \in \mathbb{R}^{(U\times V)\times H\times W\times C}$, where $U \times V$, $H \times W$, and C denote the angular dimension, spatial dimension and channel dimension, respectively, we first reshape it to obtain $U \times V$ feature with the dimension of $H \times W \times C$. We then concatenate the reshaped feature $\mathcal{F}^r$ along the channel dimension and feed it to a convolutional layer followed by a residual block (CR block) to obtain $\boldsymbol{F}$, which contains the multi-view information in a light field

$$\boldsymbol{F} = \text{CR}([\mathcal{F}^r]), \tag{1}$$

where $\text{CR}(\cdot)$ denotes the convolutional layer followed by a residual block, and $[\cdot]$ is the concatenation operation. $\mathcal{F}$ is then replicated and concatenates with $\mathcal{F}^r$ along the channel dimension, followed by the CR block to obtained the fused frequency feature. The fused frequency feature and $\mathcal{F}^r$ are concatenated and fed to the CR block, followed by the residual addition to obtain $\mathcal{F}^{fuse}$

$$\mathcal{F}^{fuse} = \text{CR}([\text{CR}([\boldsymbol{F}, \mathcal{F}^r]), \mathcal{F}^r]) + \mathcal{F}^r \tag{2}$$

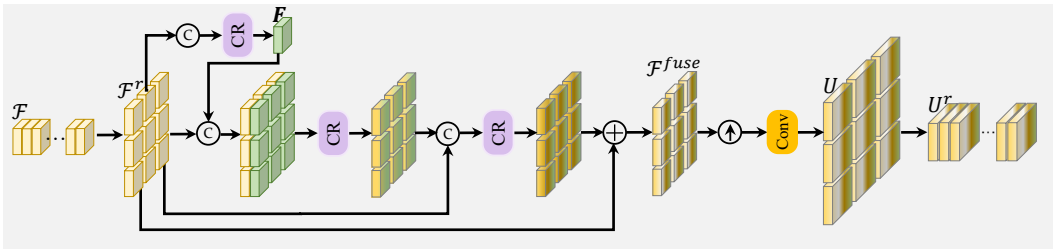
To obtain the HR representation U, $\mathcal{F}^{fuse}$ is upsampled by the bilinear interpolation operation followed by a 1×1 convolutional layer

$$U = \text{conv}(\text{Bi}(\mathcal{F}^{fuse})), \tag{3}$$

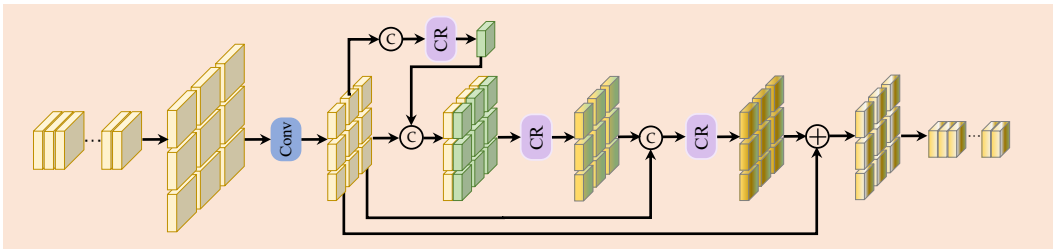
where $\text{Bi}(\cdot)$ is the bilinear interpolation operation and $\text{conv}(\cdot)$ is the 1×1 convolutional layer. U is finally reshaped to U^r, which is the final output of the FSU block.

As is shown in Figure 1(b), we won't go into depth of the structure of the FSD block because there are numerous similarities between the FSU block and the FSD block.

* These authors contribute equally to this work.
† Corresponding author.



(a) Frequency Scale-Up Block

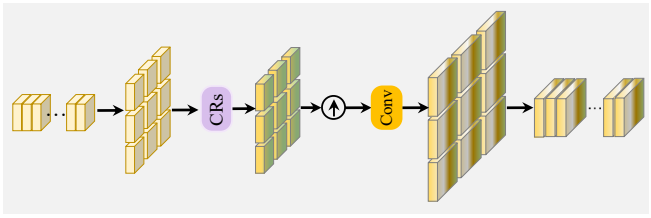


(b) Frequency Scale-Down Block

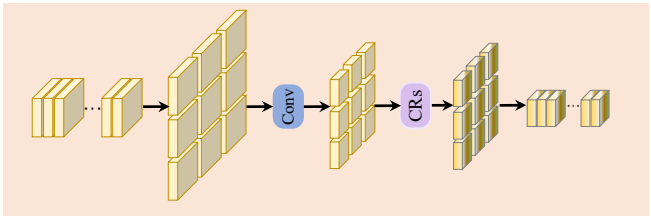
CR Convolutional layer + Residual block Conv 1x1 convolutional layer Conv Convolutional layer with stride=2

Figure 1. The structure of the frequency scale-up block and the frequency scale-down block.

To analyze the effectiveness of the FSU block and the FDU block, we conduct the following experiment on LytroZoom. We design the variants shown in Figure 2, and the results are shown in Table 1. Results in the table demonstrate the effectiveness of the FSU and FSD blocks.



(a) Frequency Scale-Up Block Variant



(b) Frequency Scale-Down Block Variant

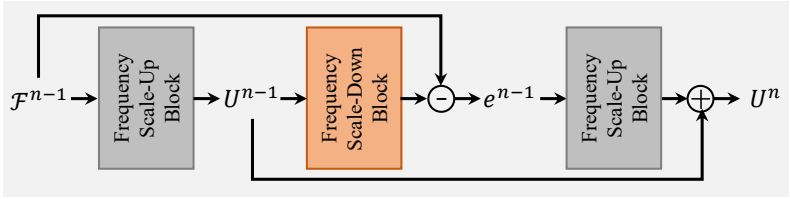
Figure 2. Structure of the frequency scale-up block variant and the frequency scale-down block variant.

Table 1. Analysis of the FSU block and the FSD block.

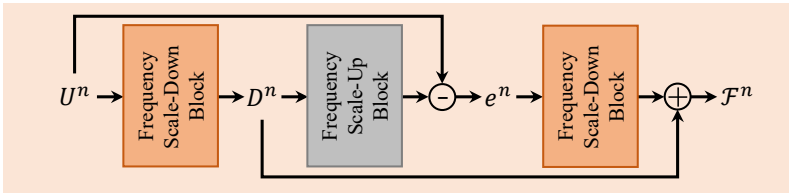
Original	30.11
Variant	29.90

2. Details of the Frequency Up-Projection Unit and the Frequency Down-Projection Unit

Detailed structures of the frequency up-projection unit and the frequency down-projection unit are shown in Figure 3.



(a) Frequency up-projection unit



(b) Frequency down-projection unit

Figure 3. Structures of (a) frequency up-projection unit (FUPU) and (b) frequency down-projection unit (FDPU). For a more straightforward depiction, we omit the subscript from the figure that represents the low, medium, and high frequency features.

3. Investigation of the Frequency Decomposition in OFPNet

To study the influence of the decomposed frequency components in the OFPNet, we visualize the features for three frequency components in Figure 4.

We first visualize the features for three frequency components in the first line of Figure 4. We see that the feature in the high-frequency branch contains more details and texture information, and the feature in the low-frequency branch contains the least details. After we feed the extracted frequency feature components to the frequency projection (FP) operations, we obtain the enhanced feature representations, which are shown in the second line of Figure 4. It can be seen that more pixels are activated, indicating that FP operations can effectively enhance the feature representations.

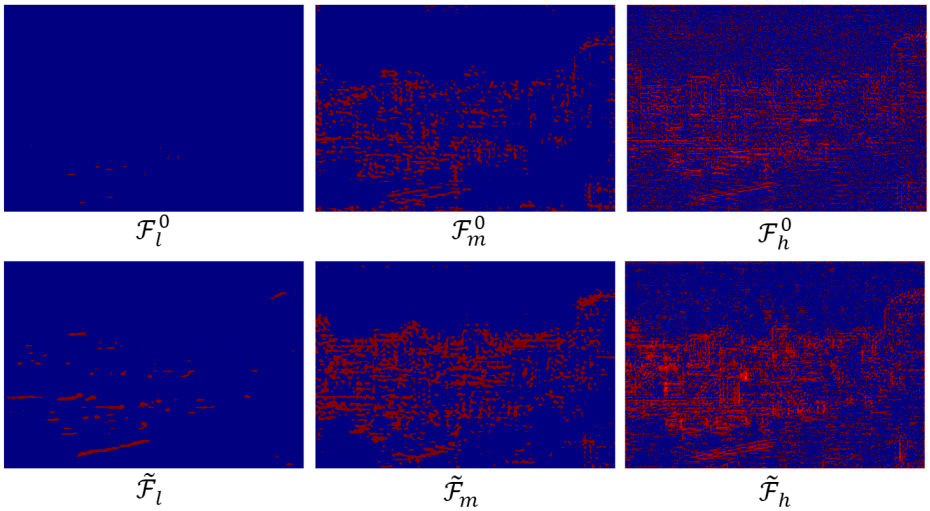


Figure 4. Visualized frequency feature components of the scene of *Bangkok2*.

4. Overview of LytroZoom

Here we show an overview of the content from our LytroZoom dataset. 94 city scenes printed on postcards and 65 outdoor static objects are adopted as the subjects. As shown in Figure 5 and Figure 6, images from LytroZoom-P and LytroZoom-O have diverse colors and contents which facilitate leaning-based light field super-resolution.

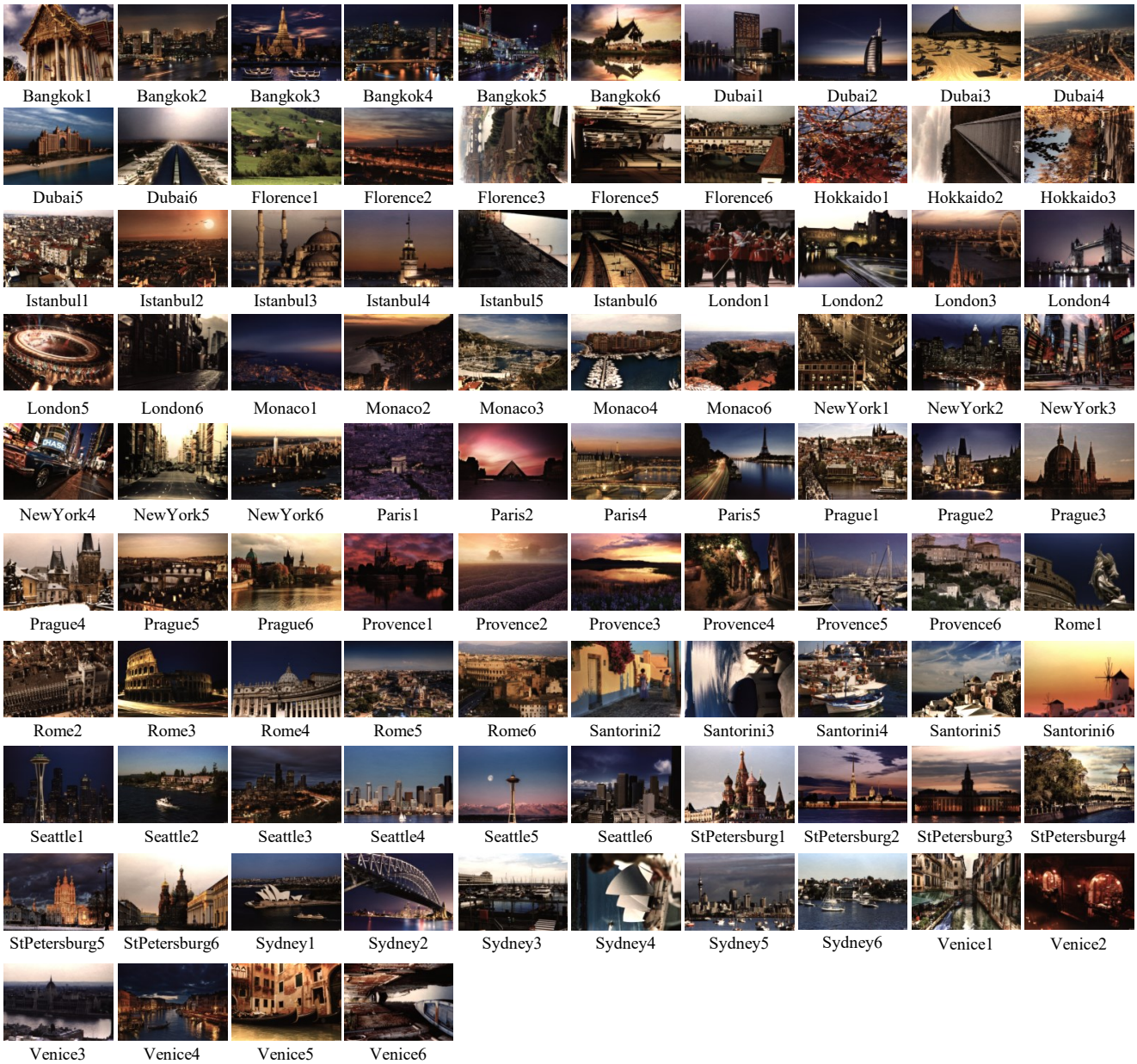


Figure 5. The overview of LytroZoom-P content.

In Table 2, we provide the details of the validation set and the testing set.

Table 2. The validation set and the testing set in LytroZoom-P.

Validation					Testing				
Bangkok1	Dubai5	Hokkaido1	Hokkaido2	Istanbul6	Bangkok2	Florence6	Istanbul2	London1	London4
London3	London6	Monaco6	NewYork5	Paris5	Monaco2	NewYork4	Paris1	Prague1	Provence4
Prague5	Rome4	Santorini5	Seattle3	Sydney1	Santorini2	Seattle2	Seattle6	Sydney6	Venice5
Venice4	Venice6								

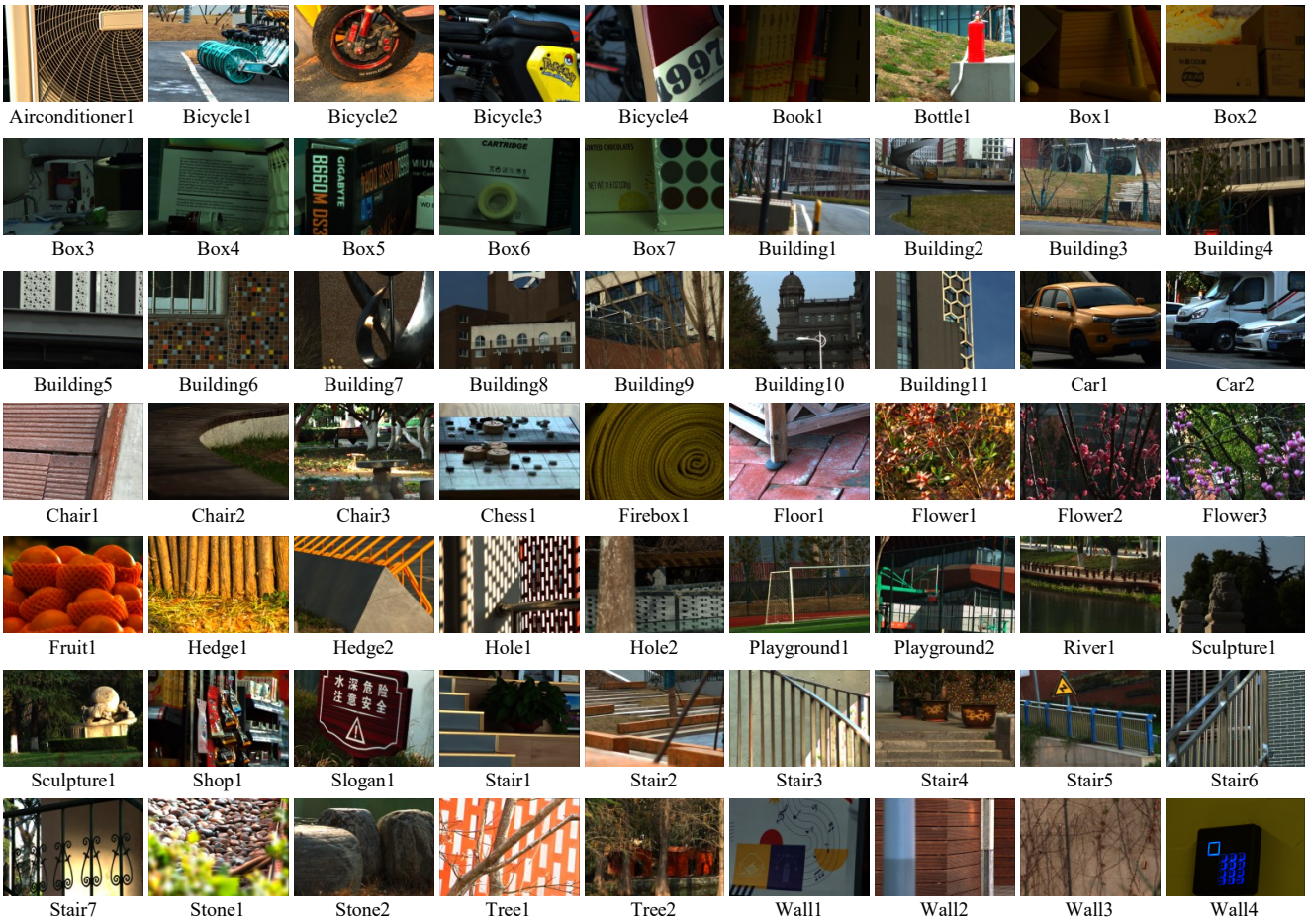


Figure 6. The overview of LytroZoom-O content.

5. More Experimental Results

In this section, we provide more visual comparison results on InterNet [3], IINet [1], DPT [2] and our proposed OFPNet in terms of $\times 2$ and $\times 4$ light field SR.

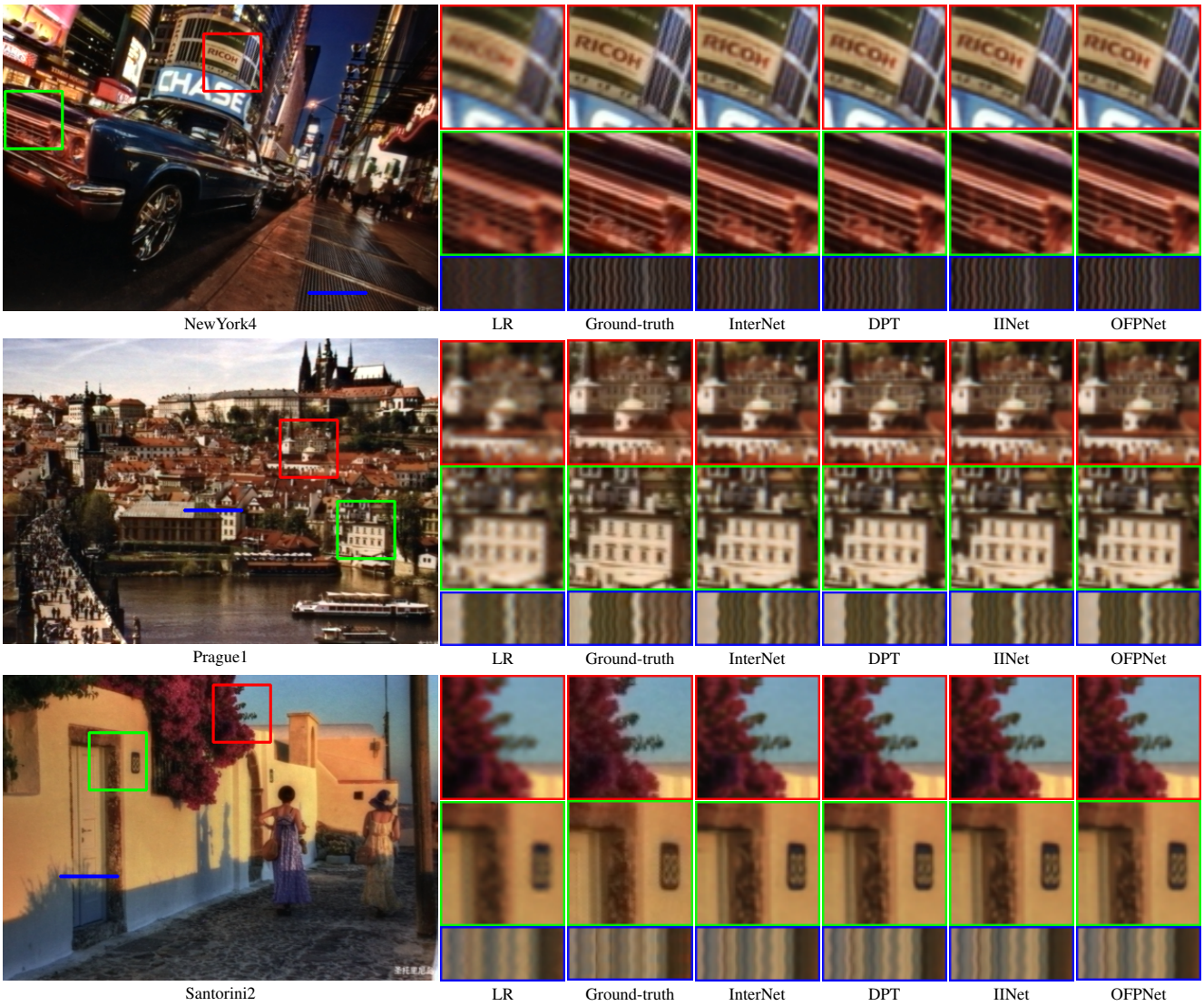


Figure 7. Visual comparisons ($\times 2$ SR) of different models on the LytroZoom-P testset.

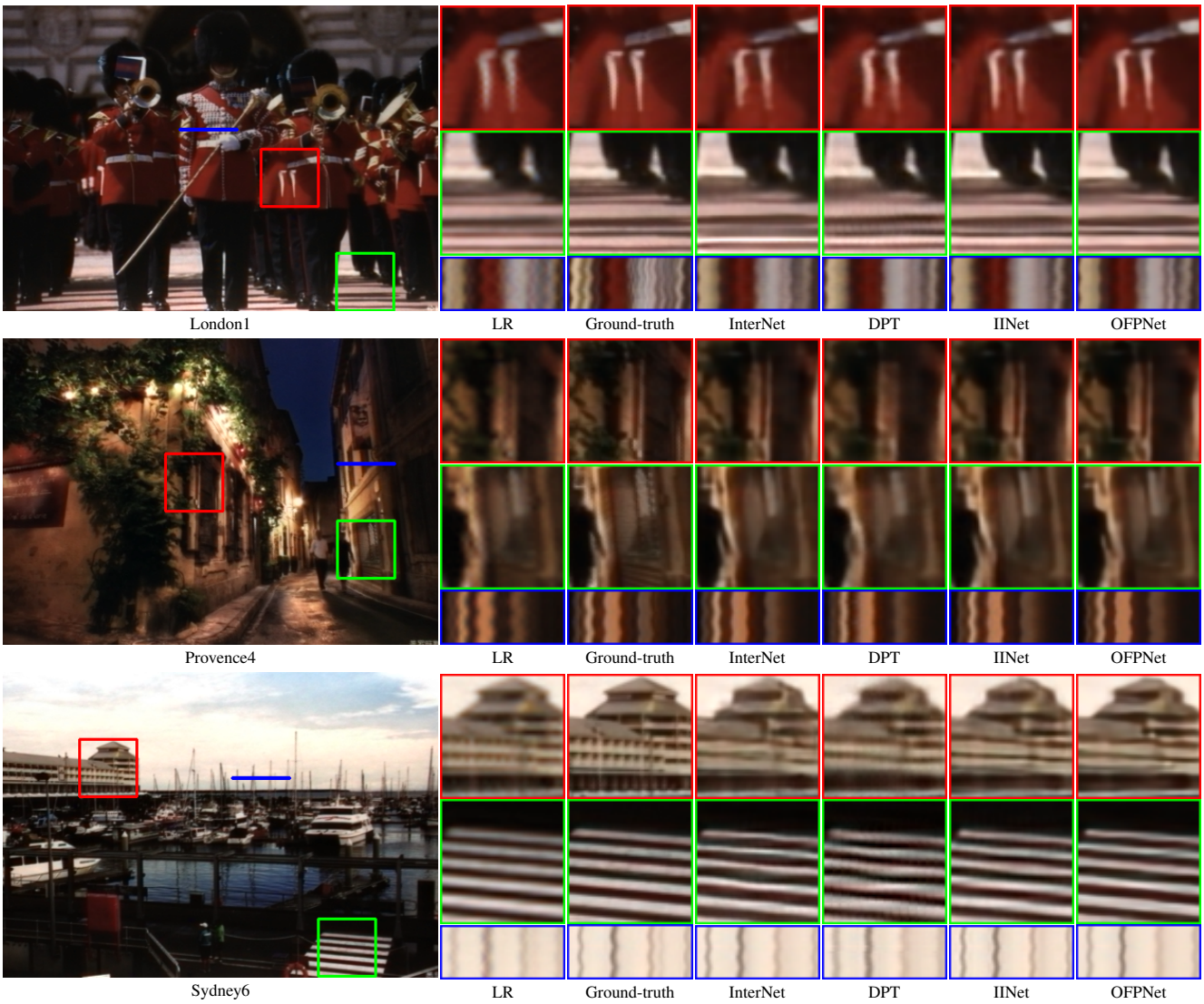


Figure 8. Visual comparisons ($\times 4$ SR) of different models on the LytroZoom-P testset.

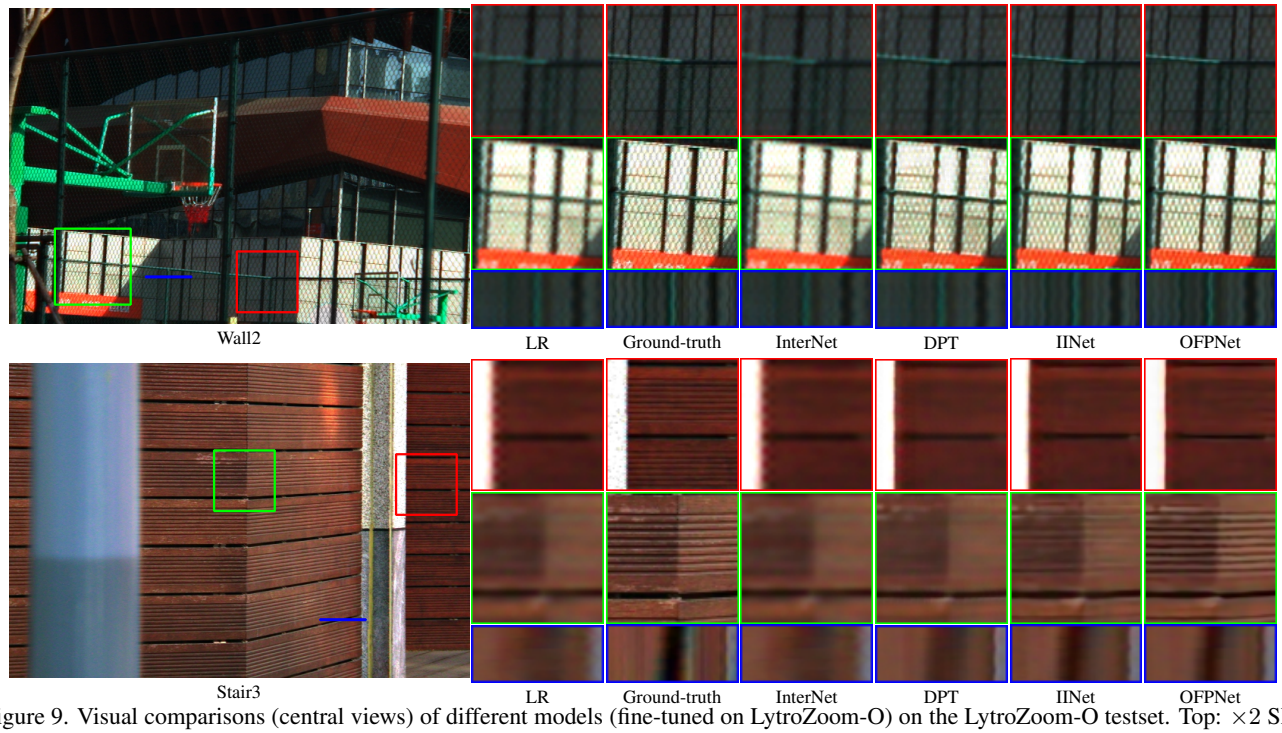


Figure 9. Visual comparisons (central views) of different models (fine-tuned on LytroZoom-O) on the LytroZoom-O testset. Top: $\times 2$ SR. Bottom: $\times 4$ SR. Please zoom in for better visualization and best viewed on the screen.

References

[1] Gaosheng Liu, Huanjing Yue, Jiamin Wu, and Jingyu Yang. Intra-inter view interaction network for light field image super-resolution. *IEEE Transactions on Multimedia*, 2021. 7

[2] Shunzhou Wang, Tianfei Zhou, Yao Lu, and Huijun Di. Detail preserving transformer for light field image super-resolution. In *AAAI*, 2022. 7

[3] Yingqian Wang, Longguang Wang, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Spatial-angular interaction for light field image super-resolution. In *ECCV*, 2020. 7