

FlexIR: Towards Flexible and Manipulable Image Restoration – Supplementary Materials –

Anonymous Authors

1 TRAINING DETAILS

For the tasks of color image denoising and JPEG compression artifact reduction, we utilized random crops from a diverse dataset amalgamation including 900 images from DIV2K [1], 2650 images from Flickr2K [10], 400 images from BSD [2], and 4744 images from WED [7]. Inspired by the methodology outlined in SwinIR [6], we adopted patch sizes of 128x128 (with a window size of 8x8) and 126x126 (with a window size of 7x7) for color image denoising and JPEG artifact reduction, respectively. To simulate real-world conditions, we introduced additive white Gaussian noise to images at varying levels (15, 25, 50) and applied JPEG compression at quality levels (10, 20, 30, 40) using the opencv-python JPEG encoder.

Our initialization strategy involved employing pre-trained embedding layers, the final RSTB, and reconstruction layers from SwinIR to bootstrap all B-RSTBs. Subsequently, we fine-tuned FlexIR, achieving satisfactory performance across all B-RSTBs. The training regimen employed the Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$), spanning 200 epochs with a batch size of 1. We employed an initial learning rate of $1e^{-5}$, adopting a linear warm-up strategy over the initial 0.1 of epochs, followed by a stable or decaying learning rate, depending on the phase of training, as indicated in Equation (1):

$$\begin{aligned} g_t &= \text{float}(\text{epoch} + 1) / \text{epochs} \\ l_r &= \text{base_lr} * w_l \\ w_l &= \begin{cases} g_t / 0.1 & \text{if } g_t < 0.1 \\ 1 - g_t, & \text{otherwise} \end{cases} \end{aligned} \quad (1)$$

For real-world image super-resolution (SR), we mirrored SwinIR's degradation model (BSRGAN [11]) and trained FlexIR on the DIV2K dataset with 900 images, employing patch sizes of 64x64, a window size of 8x8, and scales of 2 and 4 for SR tasks (x2 and x4). Training parameters for image SR mirrored those used in color image denoising and JPEG compression artifact reduction.

2 QUALITATIVE RESULTS ON IMAGE DENOISING

Extending the qualitative evaluation presented in our main manuscript, we delve deeper into the efficacy of our proposed method for color image denoising across a spectrum of noise levels (15, 25, and 50), employing renowned benchmark datasets: CBSD68 [8], Kodak24 [4], McMaster [13], and Urban100 [5]. As depicted in Fig. 1, Fig. 2, and Fig. 3, we showcase the robust performance of our model, revealing only minimal visual differences across outputs from distinct branches. Remarkably, even as noise levels intensify, these discrepancies remain nearly indistinguishable from the original images. A case in point is the enhanced preservation of detail in green areas, such as those observed above the parrot's eye in the second row of Fig. 1, where our model, FlexIR, outperforms the ground truth in maintaining textural integrity. This observation

underscores the potential of downsizing FlexIR without compromising its throughput capabilities, thereby enhancing its applicability within practical service contexts.

3 QUALITATIVE RESULTS ON JPEG COMPRESSION ARTIFACT REDUCTION

We present qualitative results illustrating the prowess of our approach in reducing JPEG compression artifacts across different quality factors: 10, 20, 30, and 40. The benchmark datasets Classic5 [3] and LIVE1 [9] are employed for evaluation. As depicted in Fig. 4 and Fig. 5, akin to the observations in color image denoising, we discern no substantial visual gap among the images generated by various B-RSTBs. Intriguingly, FlexIR-generated images exhibit enhanced visual smoothness compared to the ground truth images. While a full-size FlexIR theoretically yields optimal performance given ample computing resources, practical considerations prompt the exploration of a smaller FlexIR size, offering comparable visual outcomes tailored to real-world scenarios.

4 QUALITATIVE RESULTS ON REAL-WORLD IMAGE SUPER-RESOLUTION

In addition to the results highlighted in the main manuscript, we provide further visualizations for real-world image super-resolution at scales of $\times 2$ and $\times 4$, as showcased in Fig. 6 and Fig. 7. These comparisons leverage images sourced from [12]. Unlike the scenarios of image denoising and JPEG compression artifact reduction, discernible improvements in visual quality emerge with increasing scaling factor, revealing a clearer visual distinction between outputs from full-size FlexIR and compact-size FlexIR. However, upon meticulous examination, for instance, the animation cat's ear in the first row of Fig. 7, we observe that outputs utilizing $P = 4$ closely approximate those derived from the full-size FlexIR. This suggests that real-world image super-resolution, being inherently challenging, can benefit from a minimum FlexIR size of $P = 3$ for application services. Notably, this configuration remains adaptable to match varying computational resources in practical contexts.

REFERENCES

- [1] Agustsson et al. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [2] Pablo Arbeláez, Michael Maire, Charles Fowlkes, and Jitendra Malik. 2011. Contour Detection and Hierarchical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 5 (2011), 898–916. <https://doi.org/10.1109/TPAMI.2010.161>
- [3] Foi et al. 2007. Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images. *IEEE Transactions on Image Processing* 16, 5 (2007), 1395–1411. <https://doi.org/10.1109/TIP.2007.891788>
- [4] Rich Franzen. 1999. Kodak lossless true color image suite. source: <http://r0k.us/graphics/kodak> 4, 2 (1999), 9.
- [5] Huang et al. 2015. Single Image Super-Resolution From Transformed Self-Exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

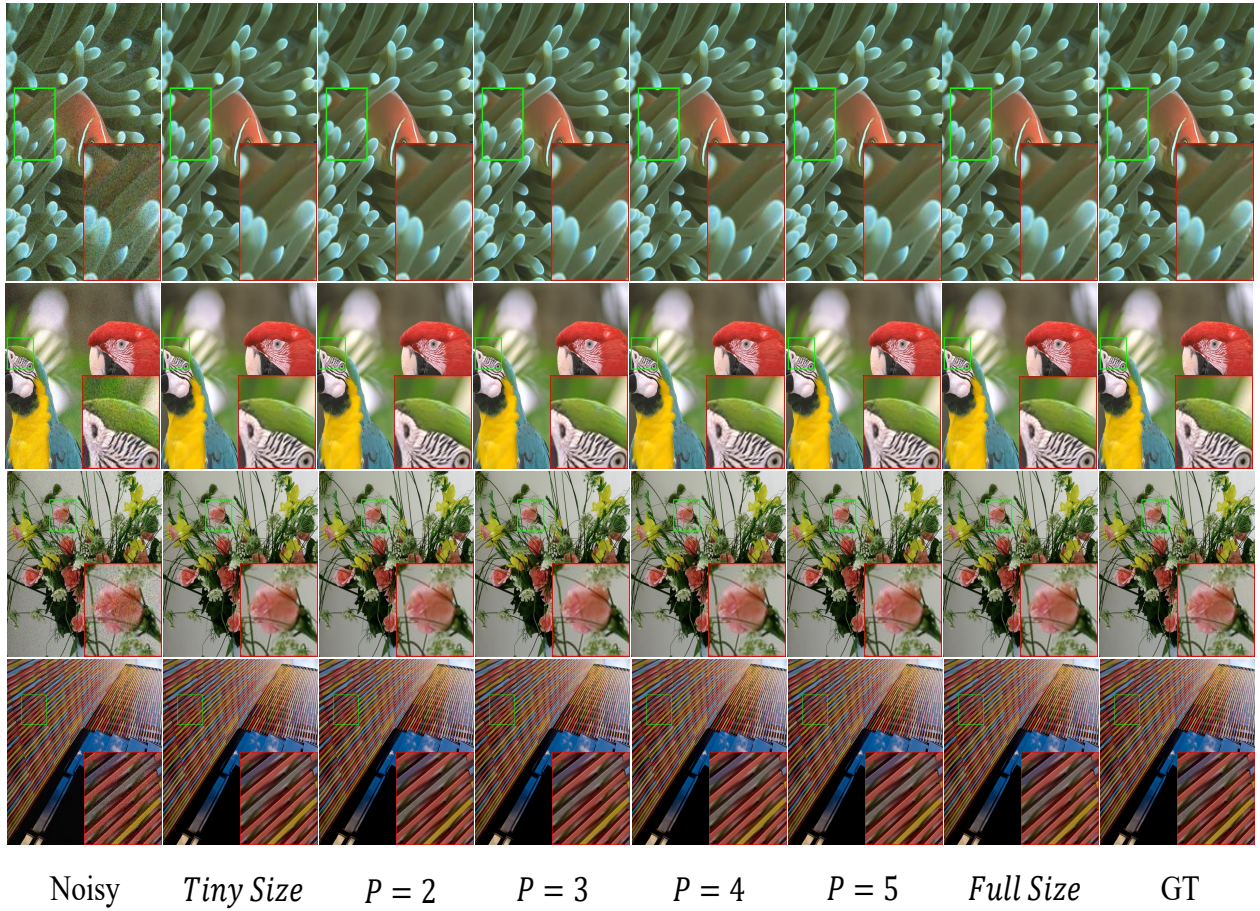


Figure 1: Visual comparison of color image denoising (noise level 15) on benchmark datasets. From first row to fourth row, sample images are selected from CBSD68, Kodak24, McMaster and Urban100 respectively. Pointer P indicates the number of activated B-RSTB while only one B-RSTB is activated in "Tiny Size" FlexIR and all B-RSTBs are activated in "Full Size" FlexIR.

- [6] Jingyun Liang, Jie Zhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1833–1844.
- [7] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. 2017. Waterloo Exploration Database: New Challenges for Image Quality Assessment Models. *IEEE Transactions on Image Processing* 26, 2 (2017), 1004–1016. <https://doi.org/10.1109/TIP.2016.2631888>
- [8] D. Martin, C. Fowlkes, D. Tal, and J. Malik. 2001. 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. 416–423 vol.2. <https://doi.org/10.1109/ICCV.2001.937655>
- [9] H Sheikh. 2005. LIVE image quality assessment database release 2. <http://live.ece.utexas.edu/research/quality> (2005).
- [10] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [11] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. 2021. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4791–4800.
- [12] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. 2021. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 4791–4800.
- [13] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. 2011. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *J. Electronic Imaging* 20 (2011), 023016.

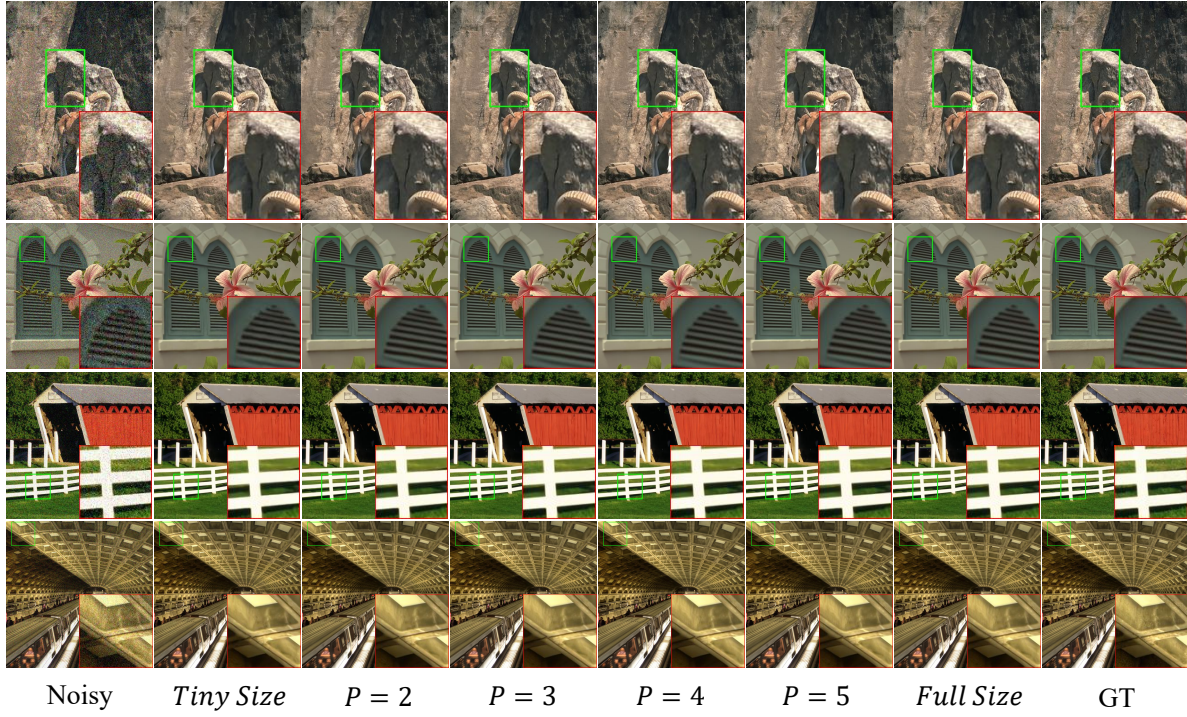


Figure 2: Visual comparison of color image denoising (noise level 25) on benchmark datasets. From first row to fourth row, sample images are selected from CBSD68, Kodak24, McMaster and Urban100 respectively. Pointer P indicates the number of activated B-RSTB while only one B-RSTB is activated in "Tiny Size" FlexIR and all B-RSTBs are activated in "Full Size" FlexIR.

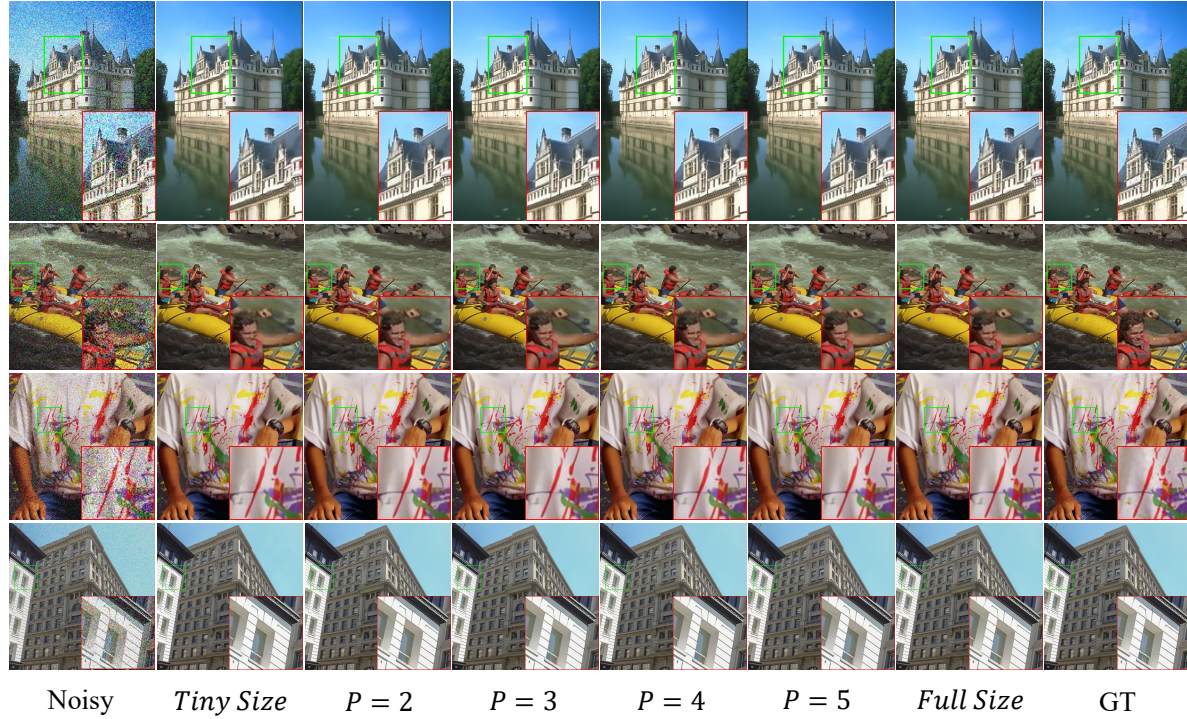


Figure 3: Visual comparison of color image denoising (noise level 50) on benchmark datasets. From first row to fourth row, sample images are selected from CBSD68, Kodak24, McMaster and Urban100 respectively. Pointer P indicates the number of activated B-RSTB while only one B-RSTB is activated in "Tiny Size" FlexIR and all B-RSTBs are activated in "Full Size" FlexIR.

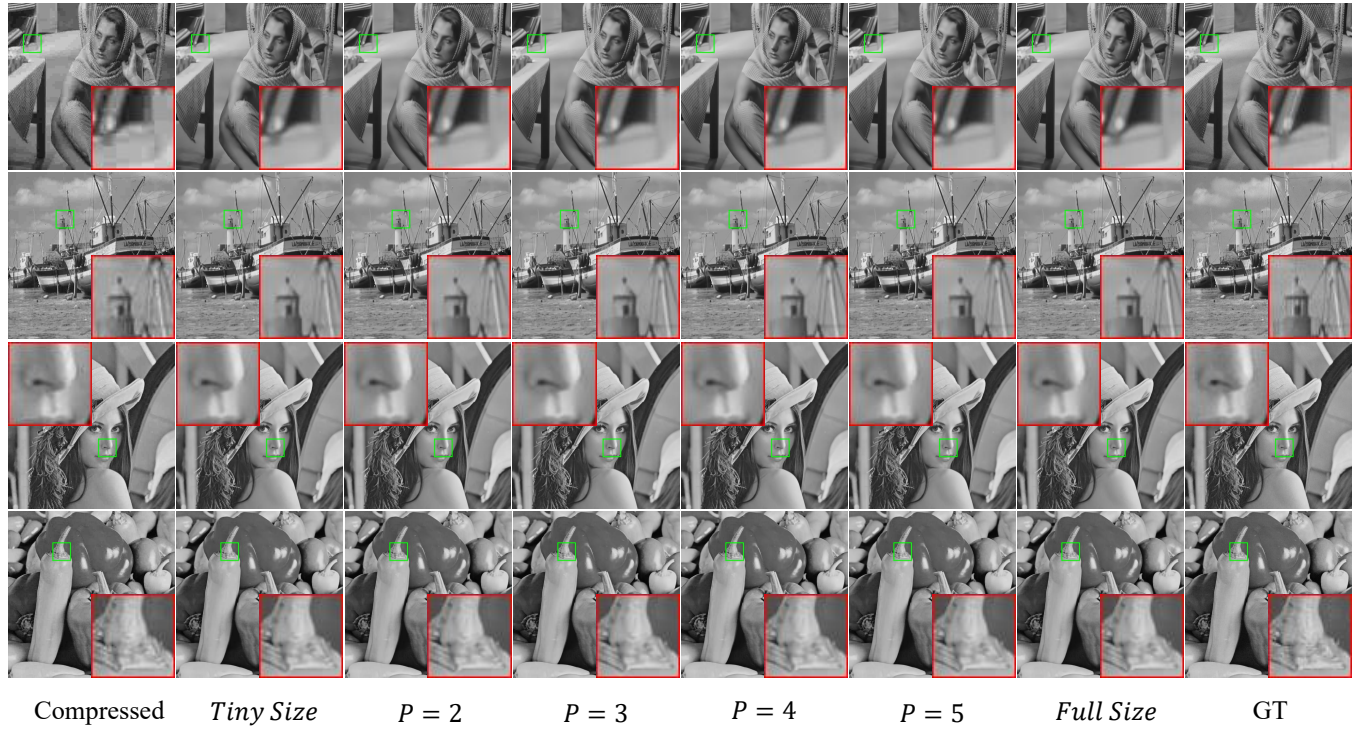


Figure 4: Visual comparison of JPEG compression artifact reduction on classic5 dataset. From first row to fourth row, the quality factor is 10, 20, 30, 40 respectively. Pointer P indicates the number of activated B-RSTB while only one B-RSTB is activated in "Tiny Size" FlexIR and all B-RSTBs are activated in "Full Size" FlexIR.

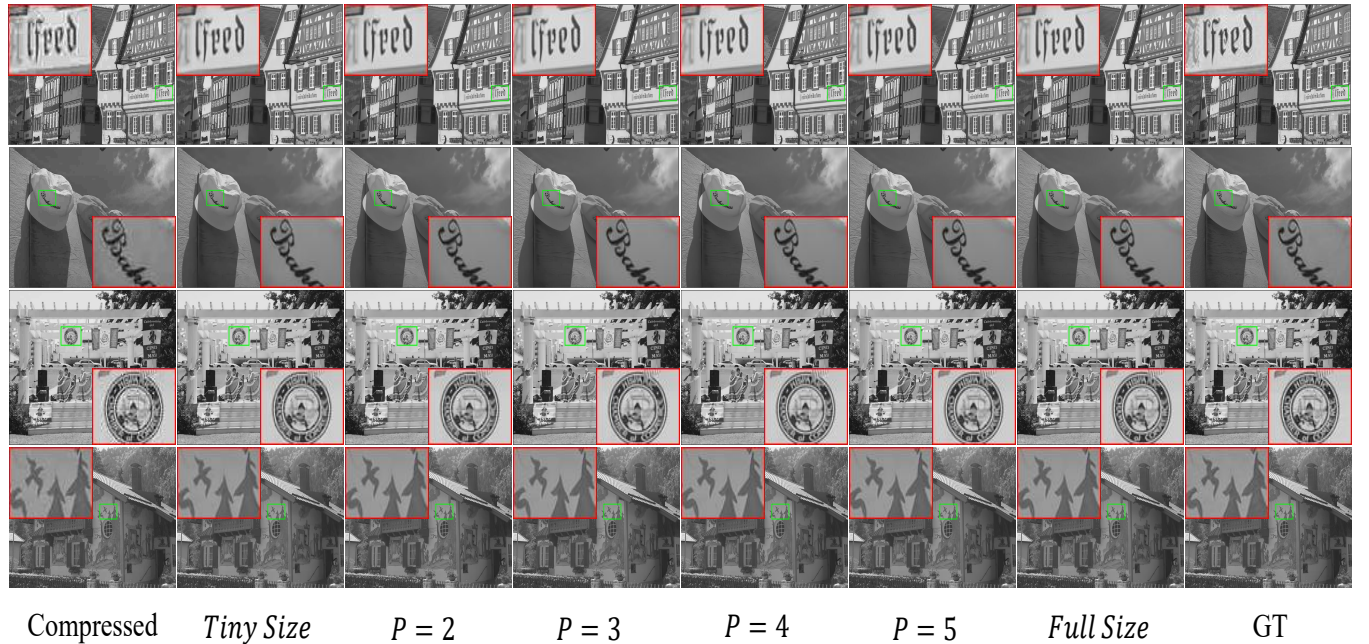


Figure 5: Visual comparison of JPEG compression artifact reduction on LIVE1 dataset. From first row to fourth row, the quality factor is 10, 20, 30, 40 respectively. Pointer P indicates the number of activated B-RSTB while only one B-RSTB is activated in "Tiny Size" FlexIR and all B-RSTBs are activated in "Full Size" FlexIR.

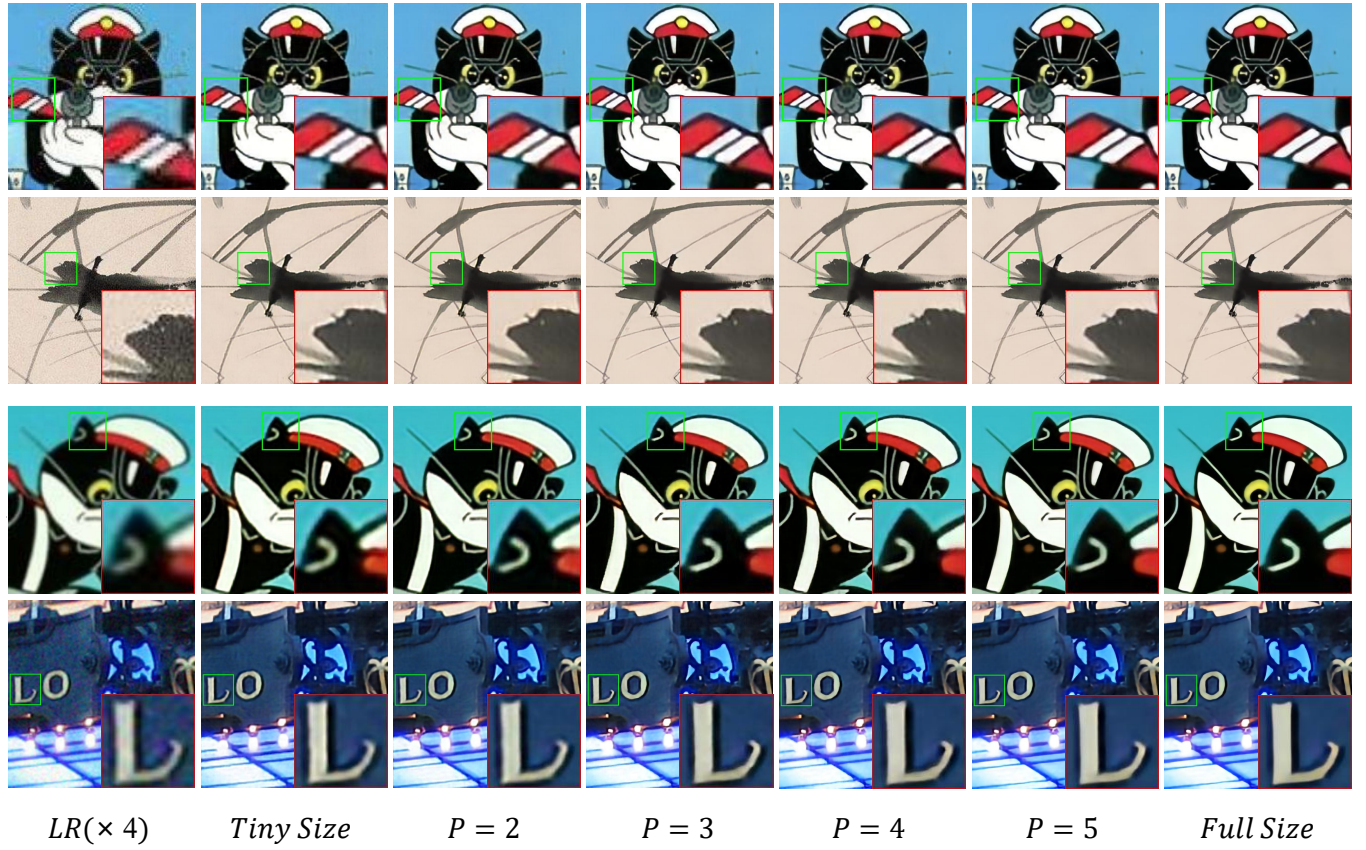


Figure 7: Visual comparison of real-world image SR ($\times 4$) on RealSRSet. Pointer P indicates the number of activated B-RSTB while only one B-RSTB is activated in "Tiny Size" FlexIR and all B-RSTBs are activated in "Full Size" FlexIR.