# Beyond the Ground Truth: Enhanced Supervision for Image Restoration

## Supplementary Material

Donghun Ryou*[2] Inju Ha*[1] Sanghyeok Chu[1] Bohyung Han[1,2]
Computer Vision Laboratory, [1]ECE & [2]IPAI, Seoul National University
{dhryou, hij1112, sanghyeok.chu, bhhan}@snu.ac.kr

## 1. Additional implementation details

### 1.1. Basis masks

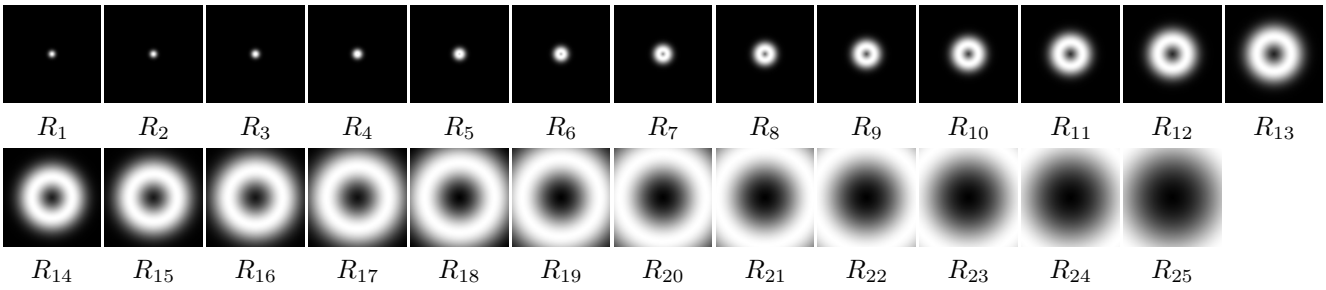



Figure S.1. Visualization of the predefined masks $R_1$-$R_{25}$. It demonstrates denser partitioning in the low-frequency domain and broader partitioning in the high-frequency domain.

In our main paper, Equation (2) defines the $b$-th ring-shaped Gaussian basis mask $R_b$ using parameters $\mu_b$ and $\sigma_b$. Here, $\mu_b$ represents the radial distance from the frequency-domain center where the mask has its peak, and $\sigma_b$ indicates the spread of the mask. We construct a total of $B = 25$ Gaussian basis masks. The first mask is centered at $\mu_1 = 0$ with a standard deviation of $\sigma_1 = 0.05$. For $b = 1, \ldots, B$, the peak positions $\mu_b$ are arranged by quadratically spacing values between $0$ and $\sqrt{H^2 + W^2}/2$, where $H$ and $W$ are height and width of the image, yielding denser coverage near the DC component and sparser placement at higher frequencies. Simultaneously, the spreads $\sigma_b$ increase quadratically from $0.05$ up to $0.55$, providing narrower rings at low frequencies and broader ones at high frequencies. This design ensures fine control around the low-frequency region and efficient coverage of the full frequency range. The all predefined masks are visualized in Figure S.1.

### 1.2. Architecture details

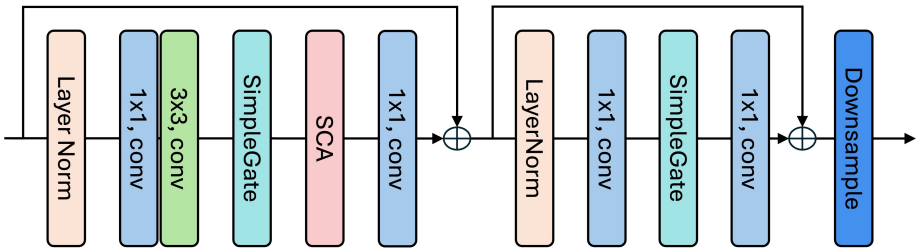



Figure S.2. The details of the NAFBlock.

Figure S.2 shows the details of the NAFBlock, utilized within the frequency mask generator illustrated in Figure 2 (b). The foundational block structures, including the Simple Gate and Simplified Channel Attention (SCA), are adopted from the NAFNet architecture [3]. An additional Downsample operation, composed of a convolution with a kernel size of $2 \times 2$ and a stride of 2, is incorporated into this NAFBlock variant. And the proposed output refinement network (ORNet) consists of the 4 encoder, 1 middle, and 4 decoder blocks, employing the NAFBlock as their fundamental building unit. The encoder block is same with figure S.2. The middle block does not incorporate last downsampling operation. The decoder block implements an

*indicates equal contribution.

GoPro GT Enhanced GT

Figure S.3. Visualization of enhanced ground truth. Our enhanced GT not only exhibits sharper text and superior perceptual quality but also maintains semantic consistency. Zoom in for better visualization.

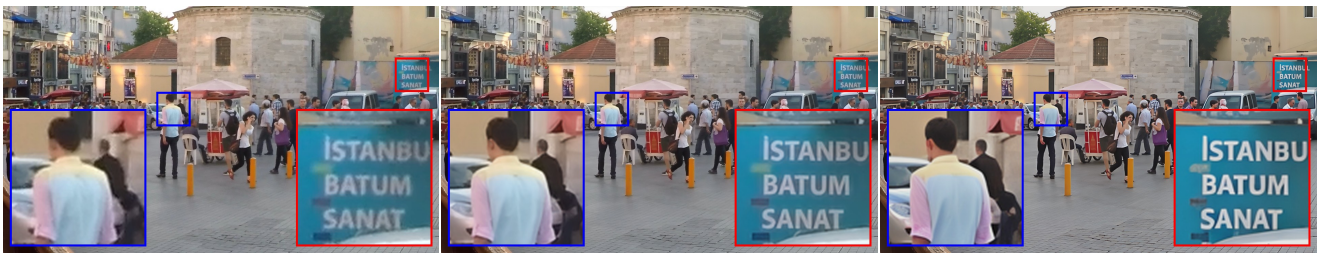
upsampling instead of downsampling: it first doubles the channel dimensionality using a $1 \times 1$ convolution, followed by a pixel shuffle module that doubles both the height and width of the feature maps, following the NAFNet architecture [3].

To control the perceptual enhancement intensity, we incorporate $\lambda$ as a conditional input for both the frequency mask generator and the output refinement network. Specifically, we spatially expand the scalar $\lambda$ into a tensor of size $H \times W$ to match the input resolution and concatenate it with the input image along the channel dimension. This conditioning allows the network to recognize the target enhancement level directly.

## 2. Additional experiments

### 2.1. Visualization of Enhanced GT

In Figure S.3, we provide additional qualitative examples of our enhanced ground truth (GT) images. Our method consistently shows superior perceptual quality by effectively removing residual degradations—such as noise and blur—from the original GTs, while preserving semantic fidelity. In Figure S.4, we compare our supervision enhancement with a simple super-resolved variant. While the super-resolved version mainly sharpens edges and increases overall brightness, our approach yields richer




Original GT | Enhanced GT (Ours) | SR Variant ($I_i^{\mathrm{GT}}$)

Figure S.4. Comparison of supervision enhancement with SR variant. Zoom in for better visualization of the semantic details and color tone.

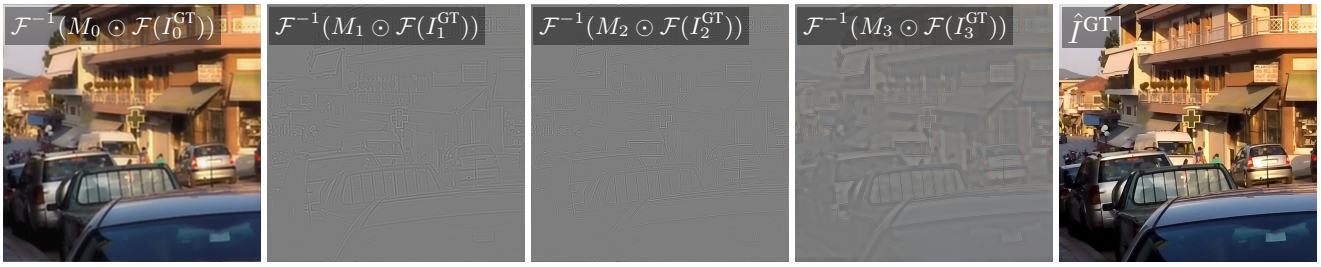



Figure S.5. Visualization of composition of our enhanced ground truth image. Super-resolved variants and the original ground truth image are blended in the frequency domain using the generated masks. Each image visualized is the inverse Fourier transform of the masked frequency components of the corresponding original or super-resolved variant ground truth image. Zoom in for better visualization.

Table S.1. The results are evaluated on the HIDE dataset [7] with the same settings as in the main paper.

| Method | *No Ref.* MUSIQ↑ | MANIQA↑ | TOPIQ↑ | LIQE↑ | *VLM-based.* VisualQuality-R1↑ | Q-Insight↑ | A-FINE↓ |
|---|---|---|---|---|---|---|---|
| AdaRevD [5] | 55.91 | 0.5882 | 0.4064 | 2.0823 | 4.3013 | 3.4613 | 33.18 |
| + ORNet ($\lambda$=0.3) | 68.48 | 0.6282 | 0.5375 | 2.6209 | 4.3341 | 3.5697 | 10.86 |
| FFTFormer [4] | 54.42 | 0.5768 | 0.3978 | 2.0440 | 4.3176 | 3.4733 | 36.35 |
| + ORNet ($\lambda$=0.3) | 68.19 | 0.6256 | 0.5295 | 2.5832 | 4.3351 | 3.5691 | 13.57 |

Table S.2. The results are evaluated on the LOL dataset [8] with the same settings as in the main paper.

| Method | *No Ref.* MUSIQ↑ | MANIQA↑ | TOPIQ↑ | LIQE↑ | *VLM-based.* VisualQuality-R1↑ | Q-Insight↑ | A-FINE↓ |
|---|---|---|---|---|---|---|---|
| Retinexformer [2] | 63.15 | 0.5870 | 0.5419 | 2.8354 | 3.4400 | 3.3153 | 49.50 |
| + ORNet ($\lambda$=0.3) | 72.80 | 0.6652 | 0.6435 | 3.9872 | 3.5200 | 3.5047 | 26.96 |
| CIDNet [9] | 69.51 | 0.6256 | 0.6288 | 3.8336 | 3.9133 | 3.6967 | 46.52 |
| + ORNet ($\lambda$=0.3) | 74.88 | 0.7035 | 0.7108 | 4.7373 | 3.9000 | 3.7613 | 20.05 |

perceptual improvements without distorting semantic structure or altering the original color tone. Lastly, Figure S.5 illustrates how the enhanced GT is composed. Our enhanced GT selectively integrates semantic details from the original GT and fine high-frequency components from the super-resolved variants, resulting in a sharper yet semantically consistent target image.

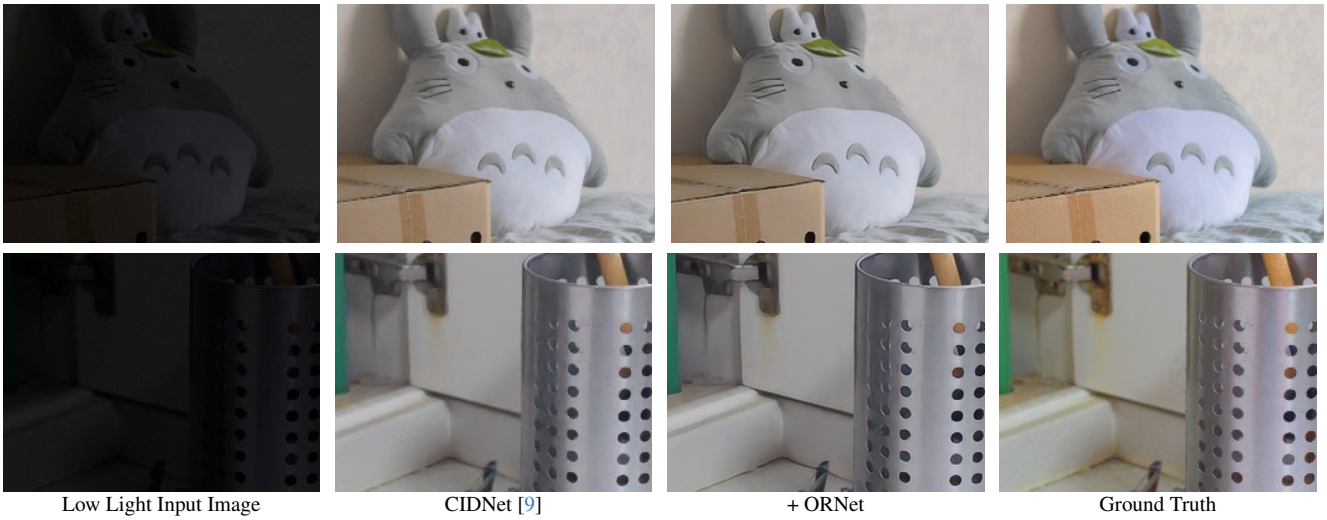
## 2.2. Generalization on out-of-distribution dataset and unseen task

**Quantitative results** Furthermore, to evaluate generalization performance, we test our method on HIDE [7] as an out-of-distribution (OOD) deblurring benchmark. We additionally evaluate it on LOL [8], a low-light enhancement benchmark, as an unseen restoration task. For evaluating low light-enhancement, we adapt our refinement network to Retinexformer [2] and CIDNet [9]. As shown in Table S.1 and Table S.2, our approach consistently enhances perceptual quality even for the unknown dataset and task, demonstrating its strong generalization capabilities.

Additionally, we evaluate our method on the GyroBlur dataset [10], which consists of both synthetic (GyroBlur-Synth) and

Table S.3. The results evaluated on the GyroBlur dataset.

| Method | *GyroBlur-Synth* | | *GyroBlur-Real* | |
|---|---|---|---|---|
| | LPIPS↓ | MUSIQ↑ | MUSIQ↑ | LIQE↑ |
| GyroDeblurNet | 0.2408 | 41.58 | 49.75 | 1.3407 |
| + ORNet (**Ours**) | **0.2310** | **56.88** | **69.75** | **3.3835** |



Low Light Input Image | CIDNet [9] | + ORNet | Ground Truth

Figure S.6. Qualitative results of our ORNet when applied to the output of CIDNet [9] on the LOL low light enhancement dataset [8]. Zoom in for better visualization.

Table S.4. Comparison of our ORNet with only using supervision enhancement trained with only ground truth variant with upscale factor 4.

| Method | MUSIQ↑ | MANIQA↑ | TOPIQ↑ | LIQE↑ |
|---|---|---|---|---|
| FFTformer [4] | 46.47 | 0.5420 | 0.3456 | 1.6131 |
| + ORNet ($\lambda$=0.1) | 48.07 | 0.5484 | 0.3537 | 1.6619 |
| + ORNet ($\lambda$=0.3) | 64.57 | 0.5949 | 0.4924 | 2.4664 |
| + ORNet ($\lambda$=0.5) | 69.18 | 0.6189 | 0.5905 | 2.9944 |
| + ORNet ($\lambda$=0.7) | 69.76 | 0.6352 | 0.6104 | 3.2444 |
| + ORNet ($\lambda$=0.9) | 69.76 | 0.6440 | 0.6198 | 3.3953 |
| + ORNet_4x ($\lambda$=0.1) | 47.86 | 0.5469 | 0.3534 | 1.6511 |
| + ORNet_4x ($\lambda$=0.3) | 62.81 | 0.5509 | 0.5069 | 2.1908 |
| + ORNet_4x ($\lambda$=0.5) | 67.53 | 0.5792 | 0.5738 | 2.6217 |
| + ORNet_4x ($\lambda$=0.7) | 68.15 | 0.5970 | 0.5775 | 2.7956 |
| + ORNet_4x ($\lambda$=0.9) | 68.20 | 0.6051 | 0.5773 | 2.8558 |

real-world (GyroBlur-Real) motion blur sequences captured with gyroscope data. As shown in Table S.3, applying ORNet to GyroDeblurNet [10] yields significant perceptual improvements on both subsets, demonstrating that our framework generalizes effectively beyond standard deblurring benchmarks.

**Qualitative results** Figure S.6 presents the qualitative results of our ORNet ($\lambda$=0.3) when applied to the output of CIDNet [9] on the LOL low light enhancement dataset [8]. Our ORNet effectively enhances the overall quality of the output, resulting in a more visually appealing image.

## 2.3. Comparison with using single GT variant

To generate an enhanced ground truth, we first employ a super-resolution model to create ground truth variants. For this purpose, we utilize upscale factors of 2, 3, and 4. By applying a frequency mixup strategy to these diverse ground truth

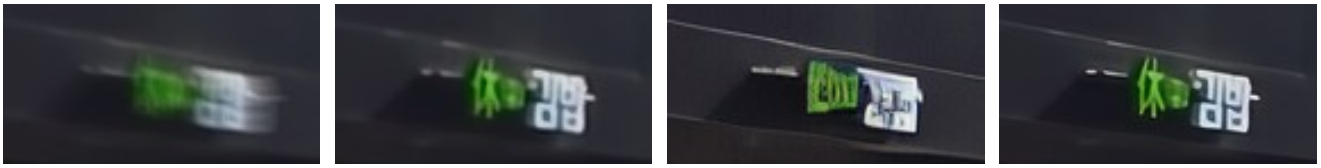



Figure S.7. Comparison of applying the ISR model (OSEDiff) directly and using our ORNet to refine the output of the restoration model (AdaRevD). The ISR model generates details that are not present in the input image, leading to unrealistic results.

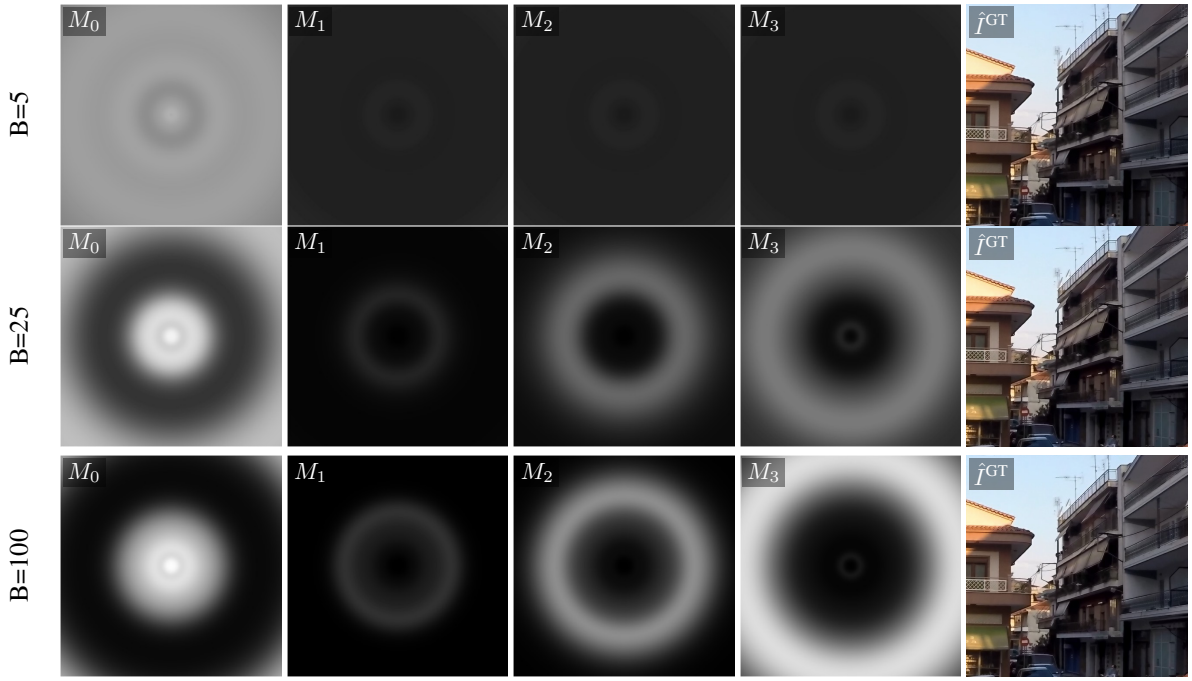



Figure S.8. Visualization of generated masks and enhanced results with different numbers of frequency bands ($B$). The rows correspond to $B = 5$, $B = 25$ (Ours), and $B = 100$, from top to bottom. $M_i$ denotes the generated masks, and $\hat{I}^{GT}$ represents the enhanced ground truth generated using these mask.

variants, we successfully construct an enhanced ground truth. As an ablation study, we conduct an experiment with different ground truth variant setting, solely using upscale factor of 4. Using these 4x variants, we follow our supervision enhancement framework and output refinement network. The results are presented in Table S.4. As shown, ORNet_4x, representing the model trained exclusively with the 4 upscale factor, achieved perceptual scores that are marginally lower than those obtained by the model trained with a mixture of variants. This indicates that the incorporation of diverse GT variants is beneficial for achieving optimal final enhancement.

## 2.4. Comparison with using ISR network instead of ORNet

Our output refinement network (ORNet) is trained to refine the output of any existing image restoration model, which is trained with enhanced supervision. To enhance the output of the restoration model, we could also directly apply the diffusion based image super-resolution (ISR) model used in our framework. However, ISR model, which is designed to enhance the perceptual quality, often hallucinates details that are not present in the input image when applied directly. Figure S.7 shows the results of applying the ISR model directly to the output of the restoration model, where the refined output destroys the textual detail.

## 2.5. Ablation study on number of Gaussian ring-shaped basis

We conduct an ablation study on the number of predefined Gaussian ring-shaped bases, denoted as $B$, to analyze its impact on frequency control. As shown in Fig. S.8, we compare the generated masks and results with $B \in \{5, 25, 100\}$. When $B$ is too small (e.g., $B = 5$), we observe that the model fails to achieve fine-grained control over the frequency components due to the limited resolution of the basis functions. However, once $B$ exceeds a certain threshold, the model exhibits a consistent

Table S.5. Full table with diverse enhancement levels. The results are evaluated on the GoPro deblurring [6] and SIDD denoising [1] test sets.

| | Method | *Perceptual Quality Metrics* MUSIQ↑ | MANIQA↑ | TOPIQ↑ | LIQE↑ | *VLM-based Metrics* VisualQuality-R1↑ | Q-Insight↑ | A-FINE↓ |
|---|---|---|---|---|---|---|---|---|
| *GoPro Deblurring* | AdaRevD [5] | 45.49 | 0.5363 | 0.3393 | 1.5656 | 4.1737 | 3.4386 | 36.19 |
| | + ORNet ($\lambda$=0.1) | 47.42 | 0.5451 | 0.3490 | 1.6256 | 4.1939 | 3.4605 | 33.33 |
| | + ORNet ($\lambda$=0.3) | 64.25 | 0.5916 | 0.4880 | 2.4291 | 4.1952 | 3.5206 | 14.38 |
| | + ORNet ($\lambda$=0.5) | 69.11 | 0.6161 | 0.5893 | 2.9711 | 4.0686 | 3.5163 | 6.80 |
| | + ORNet ($\lambda$=0.7) | 69.72 | 0.6330 | 0.6098 | 3.2308 | 4.0025 | 3.5063 | 4.77 |
| | + ORNet ($\lambda$=0.9) | 69.73 | 0.6420 | 0.6195 | 3.3870 | 3.9146 | 3.4838 | 4.24 |
| | FFTformer [4] | 46.47 | 0.5420 | 0.3456 | 1.6131 | 4.0942 | 3.4569 | 36.91 |
| | + ORNet ($\lambda$=0.1) | 48.07 | 0.5484 | 0.3537 | 1.6619 | 4.2334 | 3.4764 | 34.24 |
| | + ORNet ($\lambda$=0.3) | 64.57 | 0.5949 | 0.4924 | 2.4664 | 4.1995 | 3.5278 | 15.69 |
| | + ORNet ($\lambda$=0.5) | 69.18 | 0.6189 | 0.5905 | 2.9944 | 4.0918 | 3.5234 | 7.68 |
| | + ORNet ($\lambda$=0.7) | 69.76 | 0.6352 | 0.6104 | 3.2444 | 4.0262 | 3.5162 | 5.49 |
| | + ORNet ($\lambda$=0.9) | 69.76 | 0.6440 | 0.6198 | 3.3953 | 3.9600 | 3.4957 | 4.91 |
| *SIDD Denoising* | Xformer [11] | 22.57 | 0.3828 | 0.2472 | 1.2040 | 1.0759 | 1.6710 | 52.33 |
| | + ORNet ($\lambda$=0.1) | 26.23 | 0.3819 | 0.2738 | 1.3238 | 1.1029 | 1.7972 | 51.25 |
| | + ORNet ($\lambda$=0.3) | 35.68 | 0.4310 | 0.3710 | 1.9510 | 1.3228 | 2.1227 | 40.25 |
| | + ORNet ($\lambda$=0.5) | 37.53 | 0.4517 | 0.3908 | 2.1195 | 1.3835 | 2.1827 | 35.46 |
| | + ORNet ($\lambda$=0.7) | 37.99 | 0.4615 | 0.3968 | 2.1711 | 1.3975 | 2.2157 | 34.39 |
| | + ORNet ($\lambda$=0.9) | 38.05 | 0.4661 | 0.3989 | 2.1867 | 1.4262 | 2.2176 | 33.99 |
| | NAFNet [3] | 22.73 | 0.3937 | 0.2458 | 1.2189 | 1.0826 | 1.7060 | 51.86 |
| | + ORNet ($\lambda$=0.1) | 26.39 | 0.3917 | 0.2776 | 1.3228 | 1.1224 | 1.8217 | 50.96 |
| | + ORNet ($\lambda$=0.3) | 35.87 | 0.4380 | 0.3776 | 1.9591 | 1.3513 | 2.1584 | 40.98 |
| | + ORNet ($\lambda$=0.5) | 37.89 | 0.4605 | 0.3977 | 2.1394 | 1.4030 | 2.2269 | 35.39 |
| | + ORNet ($\lambda$=0.7) | 38.40 | 0.4709 | 0.4039 | 2.1934 | 1.4321 | 2.2498 | 34.14 |
| | + ORNet ($\lambda$=0.9) | 38.46 | 0.4758 | 0.4057 | 2.2088 | 1.4492 | 2.2501 | 33.77 |

Table S.6. Evaluation on an OOD environment, where an additional Gaussian blur ($\sigma$ = 2.5) is applied to the blurry GoPro test set images.

| Method | *Original GT* PSNR↑ | SSIM↑ | LPIPS↓ | *Enhanced GT* PSNR↑ | SSIM↑ | LPIPS↓ | *Perceptual Quality Metrics* MUSIQ↑ | MANIQA↑ | TOPIQ↑ | LIQE↑ |
|---|---|---|---|---|---|---|---|---|---|---|
| FFTFormer | 24.5688 | 0.7532 | 0.4714 | 23.6891 | 0.7224 | 0.5441 | 22.3812 | 0.2284 | 0.1832 | 1.0108 |
| +ORNet ($\lambda$=0.1) | 24.5898 | 0.7550 | 0.4592 | 23.7141 | 0.7245 | 0.5310 | 23.1592 | 0.2481 | 0.1843 | 1.0111 |
| +ORNet ($\lambda$=0.3) | 24.5774 | 0.7670 | 0.3429 | 23.8124 | 0.7405 | 0.3777 | 42.9131 | 0.2638 | 0.2646 | 1.0656 |
| +ORNet ($\lambda$=0.5) | 24.5789 | 0.7742 | 0.3107 | 23.8835 | 0.7500 | 0.3334 | 49.6099 | 0.2689 | 0.3324 | 1.2576 |
| +ORNet ($\lambda$=0.7) | 24.4442 | 0.7759 | 0.3042 | 23.8007 | 0.7524 | 0.3253 | 50.9512 | 0.2950 | 0.3487 | 1.3977 |
| +ORNet ($\lambda$=0.9) | 24.2407 | 0.7754 | 0.3036 | 23.6454 | 0.7524 | 0.3243 | 51.5943 | 0.3174 | 0.3174 | 1.0656 |

Table S.7. Evaluation on an OOD environment, where an additional white noise ($\sigma$ = 9) is applied to the blurry GoPro test set images.

| Method | *Original GT* PSNR↑ | SSIM↑ | LPIPS↓ | *Enhanced GT* PSNR↑ | SSIM↑ | LPIPS↓ | *Perceptual Quality Metrics* MUSIQ↑ | MANIQA↑ | TOPIQ↑ | LIQE↑ |
|---|---|---|---|---|---|---|---|---|---|---|
| FFTFormer | 24.3574 | 0.5867 | 0.4463 | 23.8352 | 0.5713 | 0.4751 | 30.1430 | 0.4517 | 0.2655 | 1.1819 |
| +ORNet ($\lambda$=0.1) | 24.3804 | 0.5886 | 0.4438 | 23.8616 | 0.5734 | 0.4720 | 30.5873 | 0.4528 | 0.2664 | 1.1909 |
| +ORNet ($\lambda$=0.3) | 24.4088 | 0.6179 | 0.4070 | 23.9693 | 0.6057 | 0.4233 | 41.8760 | 0.4699 | 0.3188 | 1.4321 |
| +ORNet ($\lambda$=0.5) | 24.4819 | 0.6771 | 0.3549 | 24.1483 | 0.6671 | 0.3602 | 55.2554 | 0.5204 | 0.4034 | 1.8986 |
| +ORNet ($\lambda$=0.7) | 24.3252 | 0.6943 | 0.3372 | 24.0463 | 0.6852 | 0.3397 | 58.9384 | 0.5473 | 0.4212 | 2.1779 |
| +ORNet ($\lambda$=0.9) | 24.0224 | 0.7018 | 0.3313 | 23.7722 | 0.6928 | 0.3335 | 60.1486 | 0.5618 | 0.4295 | 2.2923 |

tendency in frequency selectivity. Since we observed that $B = 25$ already provides sufficient capacity for precise frequency control, we adopt $B = 25$ as the default setting in our experiments.

## 2.6. Ablation study on enhancement level $\lambda$

**Quantitative result** Tables S.5, S.6, and S.7 present the extended results across different $\lambda$ values. We consistently observe that increasing $\lambda$ leads to better scores in perceptual quality metrics. However, an excessively high $\lambda$ exacerbates the risks discussed in the main paper, such as hallucination issues. Furthermore, in the out-of-distribution (OOD) settings shown in Tables S.6 and S.7, we observe that an excessively high $\lambda$ value can degrade reference-based performance against both the original and enhanced GTs. This appears to be because when $\lambda$ is too large, the ORNet applies changes in color tone


Figure S.9. Qualitative comparison of GT enhancement with varying $\lambda$ values. Excessively large $\lambda$ values increase the risk of hallucinations, such as color shifts and semantic deviations from the original GT. Zoom in for better visualization.

and further enhancements that go beyond removing the remaining degradations (blur, noise), resulting in a reference based performance drop. This highlights the importance of selecting an appropriate $\lambda$ to achieve a balance between enhancement and fidelity.

**Qualitative results** Figure S.9 visualizes the results of our Ground Truth (GT) enhancement with varying values of the hyperparameter $\lambda$. As $\lambda$ increases, the perceptual quality may be enhanced, but this can introduce undesirable artifacts such as altered color tones and semantic changes that deviate from the original GT. In contrast, an optimally chosen $\lambda$ effectively removes residual noise and blur, leading to a perceptually improved image while preserving the color and semantic integrity of the original.

**Mask visualization on diverse enhancement level $\lambda$** Figure S.10 shows the visualization of the generated masks with different $\lambda$ values. When $\lambda$ is small, the generated masks are mostly focused on $M_0$, dedicated for original ground truth image. As $\lambda$ increases, generated masks cover a diverse other ground truth variants.

**Additional user study** We conducted a user study to validate the perceptual quality of the output images of ORNet with different $\lambda$ values. Following the same protocol as described in the main paper, we evaluated three levels of the refinement weight $\lambda \in \{0.1, 0.3, 0.5\}$. As shown in Figure S.11, the results indicate that while both $\lambda = 0.3$ and $\lambda = 0.5$ achieved similarly high preference rates, $\lambda = 0.3$ yielded the lowest loss rate.

## 2.7. Training Stability

To assess the training stability of our modular output refinement network (ORNet), we trained the ORNet with five independent times using distinct random seeds. Following training, each refinement network was applied to the outputs of a pretrained AdaRevD [5] on the GoPro test dataset. Then, the standard deviation is calculated for each metric with $\lambda = 0.3$. The resulting standard deviations were as follows: PSNR (0.023), SSIM (0.0002), LPIPS (0.0005), DISTS (0.0005), MUSIQ (0.097), MANIQA (0.0007), TOPIQ (0.001), and LIQE (0.01). The observed standard deviations for each metric are notably low. This outcome indicates a high degree of stability in our training procedure for the output refinement network.

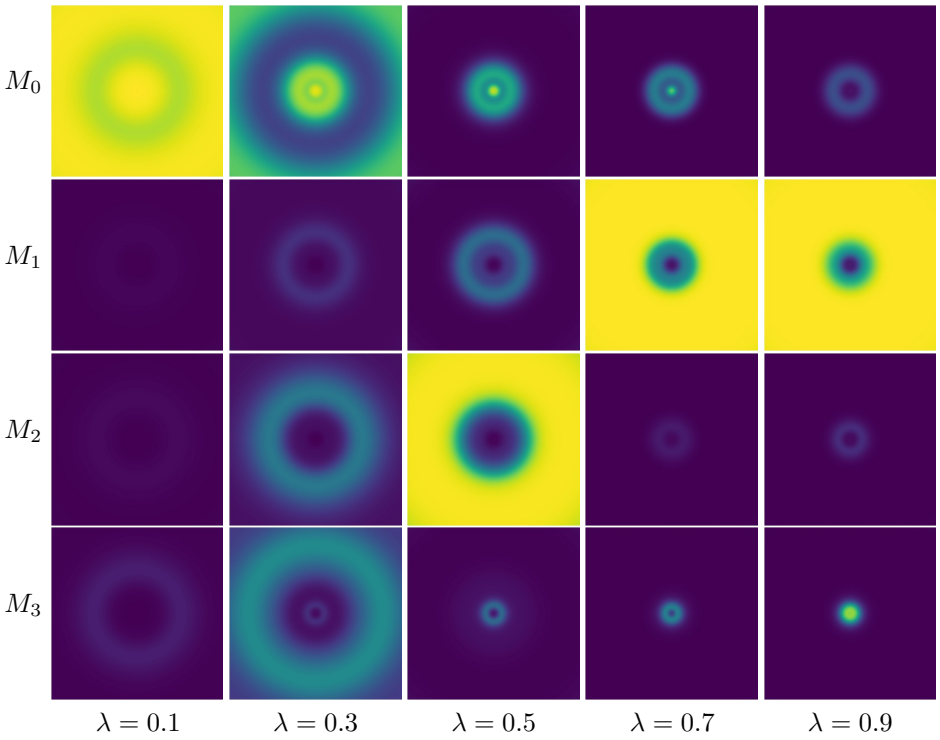



Figure S.10. Frequency masks generated by the proposed controllable frequency mask generator. The enhancement level is controlled by the input parameter $\lambda$.

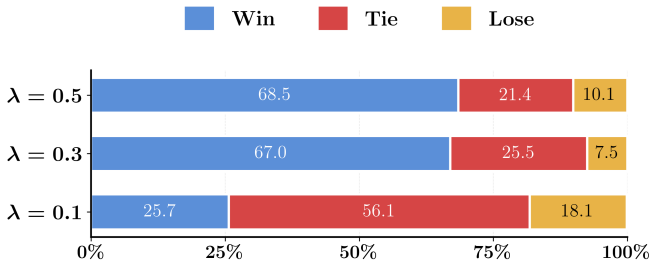



Figure S.11. User study with various $\lambda$ values. Participants consistently preferred our ORNet outputs to the baselines.

## 3. Limitations

Although our enhanced supervision framework achieves notable perceptual gains, selecting the enhancement strength $\lambda$ remains inherently challenging. Existing no-reference IQA metrics capture perceptual naturalness but provide limited insight into semantic fidelity, making it difficult to determine $\lambda$ in a fully principled manner.

To partially mitigate this issue, we evaluate fidelity at $\lambda = 0.3$ using the available surrogate metrics—CLIP embedding similarity together with object-level consistency assessments based on detection alignment with the original ground truth (see Figure 7 and Table 5 of the main paper). These analyses indicate that $\lambda = 0.3$ preserves semantic structures while providing noticeable perceptual enhancement. Nevertheless, this evaluation remains constrained by the lack of a unified metric that jointly reflects fidelity and naturalness. Developing such a metric, or an adaptive mechanism that balances the two aspects automatically, is an important direction for future work.

## References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, 2018. 6

[2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, 2023. 3

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 1, 2, 6

[4] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, 2023. 3, 4, 6

[5] Xintian Mao, Qingli Li, and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *CVPR*, 2024. 3, 6, 7

[6] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 6

[7] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *ICCV*, 2019. 3

[8] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 3, 4

[9] Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. Hvi: A new color space for low-light image enhancement. *arXiv preprint arXiv:2502.20272*, 2025. 3, 4

[10] Heemin Yang, Jaesung Rim, Seungyong Lee, Seung-Hwan Baek, and Sunghyun Cho. Gyro-based neural single image deblurring. In *CVPR*, 2025. 3, 4

[11] Jiale Zhang, Yulun Zhang, Jinjin Gu, Jiahua Dong, Linghe Kong, and Xiaokang Yang. Xformer: Hybrid x-shaped transformer for image denoising. In *ICLR*, 2024. 6