

DRAMNet: Depth-initialized Region-Adaptive Map network for Single-Image Deblurring. Supplementary Materials

Anonymous submission

Limitations

We made several key design decisions in our approach, each with its own trade-offs.

First, we focus on deblurring single images rather than using a video-based model. We chose this simpler single-image approach to avoid the extra complexity of processing multiple video frames over time, especially since large-scale datasets of blurry videos are not widely available. The trade-off is that our model cannot leverage helpful information from neighboring video frames, so it might miss improvements that a dedicated video deblurring method could achieve. This is also the reason behind us not employing any optical flow in the model. However, extending our approach to the video domain remains an important direction for future work, and we plan to explore temporal consistency mechanisms and motion-aware modules to adapt DRAMNet to video deblurring settings.

Second, our model may still behave unexpectedly in certain real-world conditions or across different demographics due to domain bias. We have not extensively tested its performance on every possible group of people or type of scene (for example, very low-light settings or subjects with diverse skin tones and ages). As a result, the model could show some bias or errors when faced with images that differ significantly from our training data. We plan further evaluation on more diverse, representative datasets and making any necessary refinements or adding safeguards before deploying the model in real-world applications.

Third, we incorporate a VGG-based perceptual loss during training and report the LPIPS metric, both of which rely on features from a pretrained VGG network. Because LPIPS is computed using the same backbone that guides our perceptual loss, our optimization may be implicitly tuned to that metric. Although this coupling can boost quantitative LPIPS scores, it reflects a common practice in perceptual image restoration and is not inherently detrimental.

Finally, in our experiments on RealBlur we only use the JPEG track (RealBlur-J). DRAMNet operates on standard sRGB images and relies on 8-bit gamma-corrected inputs, whereas the raw-sensor data in the RealBlur-R track requires demosaicing and color-space conversion steps that are beyond the scope of this work. Extending DRAMNet to operate directly on raw images would necessitate integrating a full ISP pipeline and is left for future research. Future work

could investigate complementary or human-aligned perceptual criteria to provide a broader evaluation of visual quality.

Visual Comparison on Different Sets

Figure 1 compares the outputs of three state-of-the-art deblurring methods: DRAMNet, AdaRevD-L, and NAFNet64 on a variety of RSBlur test images. Across all examples, DRAMNet consistently delivers the most visually coherent results, effectively restoring fine structures and edge definitions while avoiding common artifacts.

In contrast, AdaRevD-L often leaves behind residual blur in regions of complex motion or texture. Although it reduces the overall blur, close inspection reveals that some mildly blurred areas remain underprocessed, leading to a slight softness compared to DRAMNet’s outputs.

NAFNet64, despite its strong quantitative scores, exhibits noticeable visual inconsistencies. In particular, one can observe subtle banding and spurious high-frequency noise in areas that were originally smooth. These artifacts are not reflected in PSNR or SSIM metrics, highlighting the gap between numerical performance and perceptual quality.

In general, these comparisons demonstrate that DRAMNet’s depth-aware priors and region-adaptive processing yield superior, artifact-free restorations across diverse real-world blur conditions, whereas competing methods may still suffer from under- or over-processing despite competitive metric values.

Blur Map

Several works have explored the idea of predicting blur maps that estimate the spatial distribution and intensity of blur across an image. For example, (Ma et al. 2018) propose end-to-end deep networks that produce dense binary blur maps using fully convolutional architectures. These models typically leverage high-level semantic features to distinguish between naturally smooth regions and genuinely blurred content. Others, such as (Zhang et al. 2018), incorporate attention mechanisms or multi-branch designs that jointly estimate the extent of blur and its perceptual desirability.

These works propose interesting ideas for extracting blur maps; however, they suffer from two key limitations. First, there is a lack of evidence regarding the correlation of the predicted blur maps with human perceptual judgment of blur



Figure 1: The comparison between several usecases from the ablation section of the main article

severity. Second, the blur annotations used for supervision are based on heuristic assumptions and are inherently unstable, introducing inconsistency and bias into the training process.

In this work, we use a different approach to learning blur estimation by supervising the network with a physically meaningful target. Specifically, we define the training loss using the absolute difference between the Laplacian responses of the input (blurred) image and its corresponding sharp ground truth. This target encourages the model to focus on the loss of high-frequency details caused by blur, and provides a deterministic, interpretable signal for supervision.

Since the ground truth is not available at inference time, we train a lightweight network composed of repeated Blur-Map Estimation (BME) blocks to predict the blur map from the input image alone. This architecture allows us to decouple the interpretability of the supervision signal from the flexibility of the learned model. Compared to prior methods that regress human-annotated dense blur maps, our approach benefits from a well-defined and physically grounded loss and avoids reliance on subjective human annotations.

Among various candidates for defining blur supervision targets, we choose the Laplacian operator because of its strong theoretical and practical alignment with blur perception. As a second-order derivative filter, the Laplacian is highly sensitive to high-frequency content such as edges and fine textures, which is precisely the information that is most attenuated by blur. Unlike more complex metrics that require frequency transforms, structural templates, or learned components, the Laplacian is simple, computationally efficient, and fully interpretable. Empirically, it demonstrates consistently high correlation with human-perceived blur across multiple benchmarks (Alutis et al. 2023). Furthermore, its linearity ensures a stable and convex loss surface when used as a regression target, making it particularly well-suited for training lightweight estimation networks. These properties make the Laplacian a robust and principled choice for constructing physically grounded blur maps.

Statistical Tests

We applied the one-sided Wilcoxon signed rank test to assess the statistical significance of comparisons between various parameters of our methods. This test is applicable because it is non-parametric and suited for paired samples without assuming normality. This test will show whether one parameter setup statistically outperforms another in terms of PSNR and SSIM. The results are provided in Table 1 and Table 2. D stands for depth pre-training and B stands for the blur-map branch, which uses the DRAMNet Block instead of the Ada Block.

To ensure reliability, we applied the Bonferroni correction, which controls the family-wise error rate across comparisons. This conservative adjustment minimizes false positives, reinforcing the significance of the results.

We used a significance level of $\alpha = 0.05$ for all statistical tests. The reported p -values correspond to one-sided Wilcoxon signed-rank tests. Comparisons with $p < \alpha$ are considered statistically significant. To account for multiple

DB	✓✓	✗✓	✓✗	✗✗
✓✓	-	0.0000	0.0000	0.0000
✗✓	1.0000	-	0.2499	0.0539
✓✗	1.0000	0.7501	-	0.0962
✗✗	1.0000	0.9461	0.9038	-

Table 1: P-values for pairwise comparisons (PSNR)

DB	✓✓	✗✓	✓✗	✗✗
✓✓	-	0.0000	0.0000	0.0000
✗✓	1.0000	-	0.0397	0.0353
✓✗	1.0000	0.9603	-	0.3981
✗✗	1.0000	0.9647	0.6019	-

Table 2: P-values for pairwise comparisons (SSIM)

comparisons, we applied the Bonferroni correction, adjusting the threshold for significance to control the family-wise error rate. This correction ensures that the observed significance is not due to chance and confirms the robustness of our findings.

After correction, the comparisons involving the full model (D=✓, B=✓) remained statistically significant across both PSNR and SSIM metrics, indicating that the combination of depth pretraining and blur-awareness contributes meaningfully to performance gains.

References

- Alutis, N.; Chistov, E.; Dremine, M.; and Vatolin, D. 2023. BASED: Benchmarking, Analysis, and Structural Estimation of Deblurring. In *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 1–5. IEEE.
- Ma, K.; Fu, H.; Liu, T.; Wang, Z.; and Tao, D. 2018. Deep blur mapping: Exploiting high-level semantics by deep neural networks. *IEEE Transactions on Image Processing*, 27(10): 5155–5166.
- Zhang, S.; Shen, X.; Lin, Z.; Měch, R.; Costeira, J. P.; and Moura, J. M. 2018. Learning to understand image blur. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6586–6595.