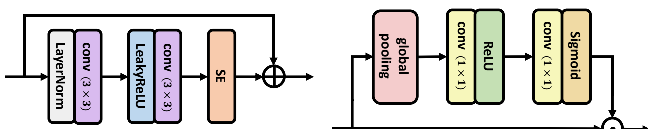
Supplemental Materials

I. Details of RCAB

See Fig. 1 for the structure of the Residual Channel Attention Block (RCAB [1]). The input feature is sequentially fed into the layer normalization (LayerNorm), a 3×3 convolution, an activation function (LeakyReLU), a 3×3 convolution and a squeeze-and-excitation (SE [2]) module, The output feature is combined with the pristine input feature via a skip connection.



(a) RCAB (b) SE

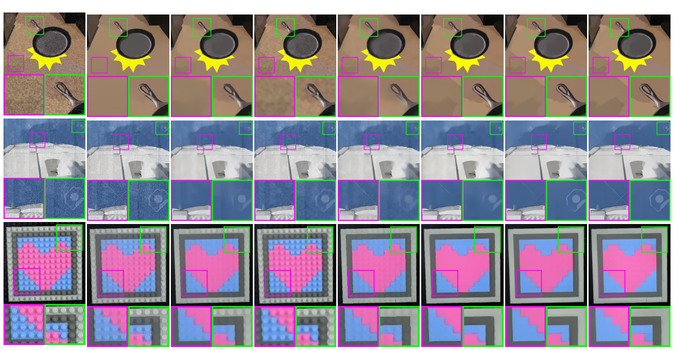
Fig. 1: Architecture of RCAB. (a) The structure of RCAB. (b) The architecture of SE.

The SE applies the channel attention mechanism. Let $\boldsymbol{Z} \in \mathbb{R}^{W\times H\times C}$ denote the input feature of SE. SE makes use of a global average pooling operation on $\boldsymbol{Z}$ to generate channel-wise statistics $\boldsymbol{S} \in \mathbb{R}^{1\times 1\times C}$. Formally, the processing in the SE can be expressed as:

$$\hat{\boldsymbol{Z}} = \boldsymbol{Z} \odot \mathrm{F}(\boldsymbol{S}), \tag{1}$$

where F consists of a stack of 1× 1 convolution and activation function, generating the excitation of each channel.

II. More Visual Comparisons



Input L0GM RTV RGF SDF JESSNet SASNet MGP-Net

Fig. 2: Visual comparison of results on synthesized images (1st row, from NKS dataset and 2nd row, from SPS dataset) and natural images (3rd row, from RTV).

III. More Analysis on Network Design

We experiment with varying the network depth by adding extra EBs and DBs, examining the impact on performance. The results in Table I show only a marginal improvement as MGPNet is deepened, primarily attributed to that the scale becomes very coarse with additional EBs and DBs. Conversely, reducing the number of EBs and DBs significantly degrades the performance, underscoring the crucial role of multi-scale processing in image smoothing.

TABLE I: Impact of network depth on performance of MGPNet.

depth	PSNR(dB)	SSIM	#Parameters(M)
2	33.42	0.9440	1.19
3	34.26	0.9477	3.72
4	34.45	0.9487	6.32

We conduct a more in-depth analysis of the network design by varying the number of GPBs. Table II presents the results of this study on the SPS test set. More (less) GPBs leads to certain performance increase (decrease) in PSNR. In terms of SSIM, the performance does not change noticeably when reducing the number of GPBs a bit, but becomes worse when more GPBs are added, possibly due to overfitting.

TABLE II: Influence of number of GPBs on performance of MGPNet.

#GPBs	PSNR(dB)	SSIM	#Parameters(M)
1	33.88	0.9477	1.94
2	34.26	0.9477	3.72
3	34.30	0.9472	5.50

IV. Comparison with Transformer-Based Methods

We include Uformer [3], a state-of-the-art transformer-based model for general image processing for performance comparison, which is trained on the same data as our MGPNet. See Table III for the result. The MGP-Net outperforms Uformer-T in terms of both PSNR and SSIM with significantly fewer parameters.

TABLE III: Quantitative comparison on transformer-based model and our MGPNet.

Method	SPS		NKS		#Params
	PSNR	SSIM	PSNR	SSIM	
Uformer-T	33.55	0.9420	35.29	0.9359	5.23
MGP-Net	34.26	0.9477	35.55	0.9467	3.72

References

[1] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European conference on computer vision*, 2018, pp. 3–19.

[2] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[3] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, “Uformer: A general u-shaped transformer for image restoration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 683–17 693.