CLG-INet: Coupled Local-Global Interactive Network for Image Restoration

Yuqi Jiang, Chune Zhang, Shuo Jin, Jiao Liu, Jiapeng Wang*

ABSTRACT. Image restoration is an ill-posed problem due to the infinite feasible solutions for degraded images. Although CNN-based and Transformer-based approaches have been proven effective in image restoration, there are still two challenges in restoring complex degraded images: 1) local-global information extraction and fusion, and 2) computational cost overhead. To address these challenges, in this paper, we propose a lightweight image restoration network (CLG-INet) based on CNN-Transformer interaction, which can efficiently couple the local and global information. Specifically, our model is hierarchically built with a " *Sandwich-like* " structure of coupling blocks, where each block contains three layers insequence (CNN-Transformer-CNN). The Transformer layer is designed with two core modules: Dynamic Bi-Projected Attention (DBPA), which performs dual projection with large convolutions across windows to capture long-range dependencies, and Gated Non-linear Feed-Forward Network (GNFF), which reconstructs mixed feature in-formation. In addition, we introduce interactive learning, which fuses local features and global representations in different resolutions to the maximum extent. Extensive experiments demonstrate that CLG-INet significantly boosts performance on various image restoration tasks, such as deraining, deblurring, and denoising.

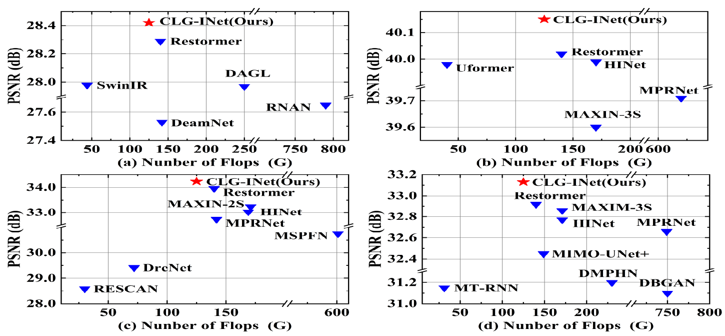

Fig1. (a) Synthetic denoising (Urban100), (b) Real denoising (SIDD), (c) Deblurring (Average), and (d) Deblurring (GoPro). Our model is lightweight and efficient.

Contributions

- We propose CLG-INet, an interactive network for multiple image restoration tasks, which is hierarchically built with multiple coupling blocks for CNN-Transformer incorporation. It can effectively aggregate the information between local details and global contexts.
- We design two effective and lightweight modules: the dynamic bi-projected attention (DBPA) to enhance the modeling of long-range dependencies, and the gated non-linear feed-forward network (GNFF) to control the forward flow of complementary information for reconstruction.
- We introduce an intra-block coupling connection and an inter-block dual hybrid unit for interactive learning, which preserves the well-embedded information flow for feature propagation.

Method Overview

We present an efficient interactive network based on a hierarchical structure, named as CLG-INet. In this section, we first introduce the overall architecture of CLG-INet. Next, we describe the important components of the designed coupling block in detail. Finally, we introduce the intra-block connection and the inter-block unit for interactive learning.

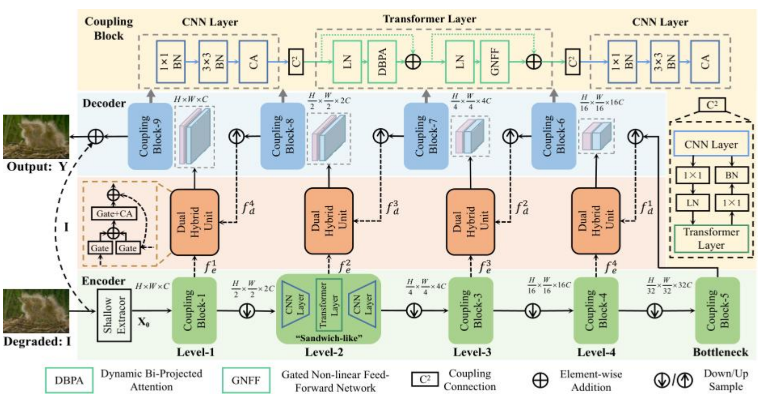

Fig2. The architecture of the proposed CLG-INet.

CNN Layer

Our CNN layer cascades two convolutional kernels and a channel attention layer to efficiently model local features with the residual learning strategy. We put the CNN layer before and after the Transformer layer to form a "*sandwich-like*" structure, where the first layer is intended to feed local features and the second one complements detailed information subsequently. This structure enables collaborative learning for CNN-Transformer complementary strengths.

Transformer Layer

To leverage the global representation, a Transformer layer is placed in the middle of the coupling block. This layer aggregates global information from local features of the first CNN layer and is transmitted to the second CNN layer for local-global interaction. In addition, to reduce the calculation complexity, we modify the self-attention module (DBPA) and feed-forward network (GNFF) of the Transformer.

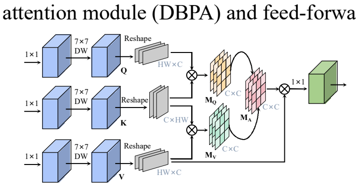

Fig3. Dynamic Bi-Projected Attention.

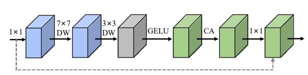

Fig4. Gated Non-linear Feed-Forward Network.

Experiments

Table 1: Quantitative results (PSNR↑/RMSE↓) with grayscale image denoising at Gaussian noise levels of 30, 50 and 70.

Methods	BSD68 [36] 30	BSD68 [36] 50	BSD68 [36] 70	Kodak24 30	Kodak24 50	Kodak24 70	Urban100 [23] 30	Urban100 [23] 50	Urban100 [23] 70
BM3D [13]	27.76/0.43	25.62/0.70	24.44/0.92	29.13/0.31	26.99/0.51	25.73/0.68	28.75/0.34	25.94/0.65	24.27/0.95
RED [35]	28.50/0.36	26.37/0.59	25.10/0.79	29.77/0.27	27.66/0.44	26.39/0.59	29.18/0.31	26.51/0.57	24.82/0.84
DnCNN [74]	28.36/0.37	26.23/0.61	24.90/0.83	29.62/0.28	27.51/0.45	26.08/0.63	28.88/0.33	26.28/0.60	24.36/0.93
MemNet [49]	28.43/0.37	26.35/0.59	25.09/0.79	29.72/0.27	27.68/0.45	26.42/0.58	29.10/0.31	26.65/0.55	25.01/0.80
IRCNN [75]	28.26/0.38	26.15/0.62	-	29.53/0.28	27.45/0.46	-	28.85/0.33	26.24/0.61	-
FFDNet [76]	28.39/0.37	26.30/0.60	25.04/0.80	29.70/0.27	27.63/0.44	26.34/0.59	29.03/0.32	26.52/0.57	24.86/0.83
RIDNet [4]	28.54/0.36	26.40/0.58	25.12/0.78	29.90/0.26	27.79/0.42	26.51/0.57	-	-	-
RDN [78]	28.56/0.36	26.41/0.58	25.10/0.79	30.00/0.26	27.85/0.42	26.54/0.57	30.01/0.25	27.40/0.46	25.64/0.70
Restormer [64]	-	26.62/ -	-	-	-	-	-	28.33/ -	-
CLG-INet	**28.79/0.35**	**26.70/0.56**	**25.39/0.76**	**30.19/0.24**	**28.01/0.41**	**26.69/0.56**	**31.03/0.21**	**28.45/0.40**	**26.62/0.65**

Table 2: Quantitative results (PSNR↑/RMSE↓) with color image denoising at Gaussian noise levels of 30, 50 and 70.

Methods	CBSD68 [36] 30	CBSD68 [36] 50	CBSD68 [36] 70	Kodak24 30	Kodak24 50	Kodak24 70	Urban100 [23] 30	Urban100 [23] 50	Urban100 [23] 70
CBM3D [12]	29.73/0.27	27.38/0.47	26.00/0.64	30.89/0.21	28.63/0.35	27.27/0.48	30.36/0.23	27.94/0.41	26.31/0.60
RED [35]	28.46/0.36	26.35/0.59	25.09/0.79	29.71/0.27	27.62/0.44	26.36/0.59	29.02/0.32	26.40/0.58	24.74/0.86
DnCNN [74]	30.40/0.23	28.01/0.40	26.56/0.56	31.39/0.19	29.16/0.31	27.64/0.44	30.28/0.24	28.16/0.39	26.17/0.62
MemNet [49]	28.39/0.37	26.33/0.59	25.08/0.79	29.67/0.28	27.65/0.44	26.40/0.58	28.93/0.33	26.53/0.57	24.93/0.82
IRCNN [75]	30.22/0.24	27.86/0.42	-	31.24/0.19	28.93/0.33	-	30.28/0.24	27.69/0.43	-
FFDNet [76]	30.31/0.24	27.96/0.41	26.53/0.57	31.39/0.19	29.10/0.31	27.68/0.44	30.53/0.23	28.05/0.40	26.39/0.59
RIDNet [4]	30.47/0.23	28.12/0.39	26.69/0.55	31.64/0.17	29.25/0.30	27.94/0.41	-	-	-
RDN [78]	30.67/0.22	28.31/0.38	26.85/0.53	31.94/0.16	29.66/0.28	28.20/0.39	31.69/0.17	29.29/0.30	27.63/0.44
Restormer [64]	-	28.59/ -	-	-	30.00/ -	-	-	30.02/ -	-
CLG-INet	**30.99/0.20**	**28.65/0.36**	**27.26/0.50**	**32.70/0.13**	**30.31/0.25**	**29.01/0.36**	**32.41/0.16**	**30.10/0.28**	**28.34/0.44**

Table 3: Real image denoising on SIDD [1] and DnD [42].

Methods	Params (M)	SIDD [1] PSNR↑	SIDD [1] SSIM↑	DnD [42] PSNR↑	DnD [42] SSIM↑	Average PSNR↑	Average SSIM↑
DnCNN [74]	0.7	23.66	0.583	32.43	0.790	28.04	0.686
CBDNet [19]	4.3	30.78	0.801	38.06	0.942	34.42	0.872
BM3D [13]	-	35.65	0.685	34.51	0.851	35.08	0.768
RIDNet [4]	1.5	38.71	0.951	39.26	0.953	38.99	0.952
VDN [63]	7.8	39.28	0.956	39.38	0.952	39.33	0.954
SADNet [8]	4.3	39.46	0.957	39.59	0.952	39.53	0.955
CycleISP [65]	-	39.52	0.957	39.56	0.956	39.54	0.957
MPRNet [67]	20.1	39.71	0.958	39.80	0.954	39.76	0.956
Uformer [55]	50.88	39.98	0.960	40.04	0.956	39.97	0.958
MAXIM-3S [52]	-	39.96	0.960	39.84	0.954	39.90	0.957
HINet [10]	-	39.99	0.958	-	-	39.99	0.958
Restormer [64]	26.13	40.02	0.960	40.03	0.956	40.03	0.958
CLG-INet	**24.7**	**40.15**	**0.963**	**40.09**	**0.957**	**40.12**	**0.960**

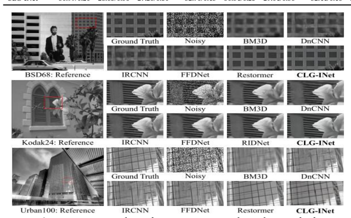

Fig5. Synthetic grayscale denoising.

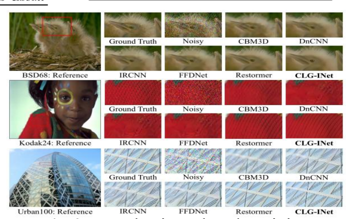

Fig6. Synthetic color denoising.

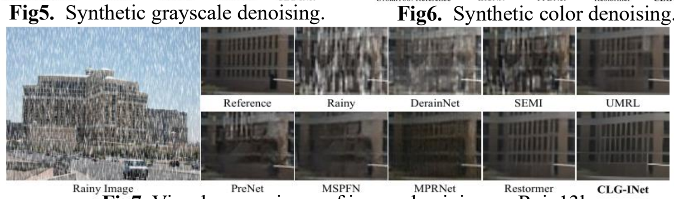

Fig7. Visual comparisons of image deraining on Rain13k.

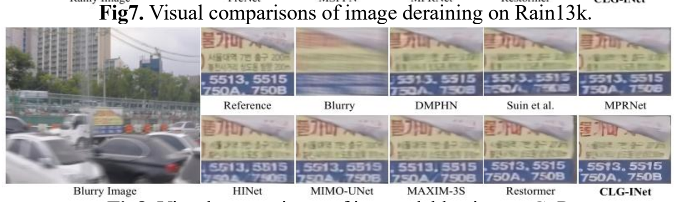

Fig8. Visual comparisons of image deblurring on GoPro.

Conclusion

In this paper, we propose CLG-INet, a lightweight and interactive image restoration network with a hierarchical structure. Specifically, we design several coupling blocks to combine CNN layers and Transformer layers using a "sandwich-like" paradigm, which effectively exploits and interacts with the local-global information. In addition, we present two important modules, DBPA and GNFF, to sufficiently model the global representation with less computational cost. In the multi-scale information propagation, we introduce an intra-block coupling connection and an inter-block hybrid gated unit for feature interaction, which simplifies the information flow and enhances the multi-scale feature representation.