The supplementary materials are organized as follows:

- Section A gives the implementation details including network configurations and training strategies.
- Section B presents more visual results.
- Section C discusses the limitations of our work.
- Section D gives the broader impacts of our work.

A IMPLEMENTATION DETAILS

We elaborate on the implementation details in terms of network configurations and training strategies. For convenience, we summarize the implementation details in Table 6. *It is worth noting that since Uformer and Restormerare equipped with different network configurations and training strategies, we train Uformer171, and Restormer1311 with their respective settings.* For better illustration, we give the U-Net architectures in Figure 5. All the experiments are conducted on NVIDIA-SMI A100 GPUs.

Network Configurations. The input projection and output projection are implemented as 3×3 convolutions, as shown in Figure 5.

For **Uformer171**, we adopt a 4-level U-Net. The channel dimension is set to 32, and the channel expansion factor in Channel MLP is set to 4 as those of Uformer Wang et al. (2022). DownSample is implemented using a 4×4 convolution with stride 2, and UpSample is implemented using a 2×2 transposed convolution with stride 2. For the depths of encoder, bottleneck, and decoder, Uformer is set to [1, 2, 8, 8], 2, and [8, 8, 2, 1], respectively. To keep the comparable parameters and complexity, we adjust the depths of encoder, bottleneck, and decoder of Uformer171 to [1, 2, 8, 10], 3, and [10, 8, 2, 1], respectively.

For **Restormer1311**, we adopt a 3-level U-Net. The channel dimension is set to 48, and the channel expansion factor in Channel MLP is set to 2.66 as those of Restormer Zamir et al. (2022). DownSample is implemented using a 3×3 convolution with stride 1 and the PixelUnshuffle operation, and UpSample is implemented using
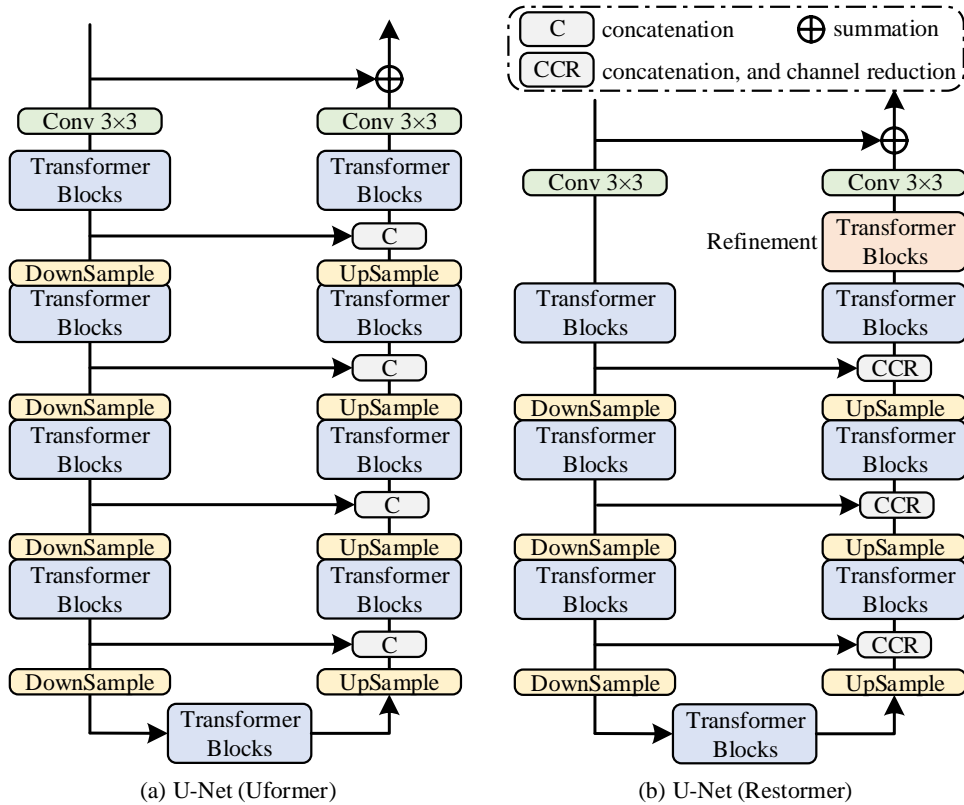

Figure 5: U-Net architectures for representative transformer-based image restoration methods. (a) Uformer Wang et al. (2022), (b) Restormer Zamir et al. (2022).

Table 6: Details of network configurations and training strategies for different models.

		Uformer Wang et al. (2022)	Uformer171	Restormer Zamir et al. (2022)	Restormer1311
Network configurations	Input Projection	convolution (k=3, s=1)	convolution (k=3, s=1)	convolution (k=3, s=1)	convolution (k=3, s=1)
	Output Projection	convolution (k=3, s=1)	convolution (k=3, s=1)	convolution (k=3, s=1)	convolution (k=3, s=1)
	Levels of U-Net	4	4	3	3
	Depths of Encoder	[1, 2, 8, 8]	[1, 2, 8, 9]	[4, 6, 6]	[4, 6, 8]
	Depths of Bottleneck	2	3	8	9
	Depths of Decoder	[8, 8, 2, 1]	[9, 8, 2, 1]	[6, 6, 4]	[8, 6, 4]
	Channel Dimension	32	32	48	48
	DownSample	convolution (k=4, s=2)	convolution (k=4, s=2)	convolution (k=3, s=1), PixelUnshuffle	convolution (k=3, s=1), PixelUnshuffle
	UpSample	transposed convolution (k=2, s=2)	transposed convolution (k=2, s=2)	convolution (k=3, s=1), PixelShuffle	convolution (k=3, s=1), PixelShuffle
	Channel Expansion Factor in Channel MLP	4	4	2.66	2.66
Training strategies	Training Patch Size	256×256 (deblurring), 128×128 (denoising)	256×256 (deblurring), 128×128 (denoising)	128, 160, 192, 256, 320, 384 (Progressive)	128, 160, 192, 256, 320, 384 (Progressive)
	Batch Size	8 (deblurring), 32 (denoising)	8 (deblurring), 32 (denoising)	64, 40, 32, 16, 8, 8 (Progressive)	64, 40, 32, 16, 8, 8 (Progressive)
	Epochs	3000 (deblurring), 250 (denoising)	3000 (deblurring), 250 (denoising)		
	Iterations			300K	300K
	Optimizer	AdamW ((0.9, 0.999), 0.02)	AdamW ((0.9, 0.999), 0.02)	AdamW ((0.9, 0.999), 10^{-4})	AdamW ((0.9, 0.999), 10^{-4})
	Initial Learning Rate	2×10^{-4}	2×10^{-4}	3×10^{-4}	3×10^{-4}
	Final Learning Rate	1×10^{-6}	1×10^{-6}	1×10^{-6}	1×10^{-7}
	Decay Strategy	Cosine	Cosine	Cosine	Cosine
	Data Augmentation	horizontal flipping or rotation (90°, 180°, or 270°)	horizontal flipping or rotation (90°, 180°, or 270°)	horizontal or vertical flipping	horizontal or vertical flipping
	Loss functions	Charbonnier loss	Charbonnier loss	L1 loss	L1 loss

a 3×3 convolution with stride 1 and the PixelShuffle operation. For the depths of encoder, bottleneck, and decoder, Restormer is set to [4, 6, 6], 8, and [6, 6, 4], respectively. To keep the comparable parameters and complexity, we adjust the depths of encoder, bottleneck, and decoder of Restormer1311 to [4, 6, 8], 9, and [8, 6, 4], respectively. In the refinement stage, we keep the number of blocks to 4 as those of Restormer.

Training Strategies. For fair comparison, we train Uformer171, Restormer1311, and Baseline131/NAFNet131 with the same train strategies as those of Uformer Wang et al. (2022), Restormer Zamir et al. (2022), and Baseline/NAFNet Chen et al. (2022a), respectively.

Specifically, for **Uformer171**, we randomly crop patches of size 256×256 from the original training image pairs for training. We train Uformer171 with a batch size of 8 for deblurring and 32 for denoising by 3K epochs. The AdamW with weight decay 0.02 and momentum terms (0.9, 0.999) is used for optimization. The learning rate declines from 2×10^{-4} to 1×10^{-6} with the cosine decay strategy Loshchilov & Hutter (2016). The training data is randomly augmented with horizontal flipping or rotation (90°, 180°, or 270°). For Uformer171, we use the Charbonnier loss for training as those of Uformer Wang et al. (2022).

For **Restormer1311**, we adopt the progressive learning strategy as those of Restormer. Specifically, we crop patches of size 512×512 with step size 256 from the training data. We train Restormer1311 for 300K iterations. The AdamW with weight decay 10^{-4} and momentum terms (0.9, 0.999) is used for optimization. We first train Restormer1311 with the patch size 128×128, batch size 64, and learning rate 3×10^{-4} for 92K. Then, the patch size and batch size pairs are updated to [(160^2, 40), (192^2, 32), (256^2, 16), (320^2, 8), (384^2, 8)] at iterations [92K, 156K, 204K, 240K, 276K], and the learning rate correspondingly declines from 3×10^{-4} to 1×10^{-6} with the cosine decay strategy. The training data is randomly augmented with horizontal or vertical flipping. For Restormer1311, we use the L1 loss for training as those of Restormer Zamir et al. (2022).

B MORE VISUAL RESULTS

Figure 6 presents more visual results on GoPro Nah et al. (2017). Figure 7 presents a visual example on SIDD Abdelhamed et al. (2018). Figure 8 presents a visual example on GoPro Nah et al. (2017). Figure 9 presents a visual example on SIDD Abdelhamed et al. (2018). Figure 10 presents a visual example on Rain14000 Fu et al. (2017b).
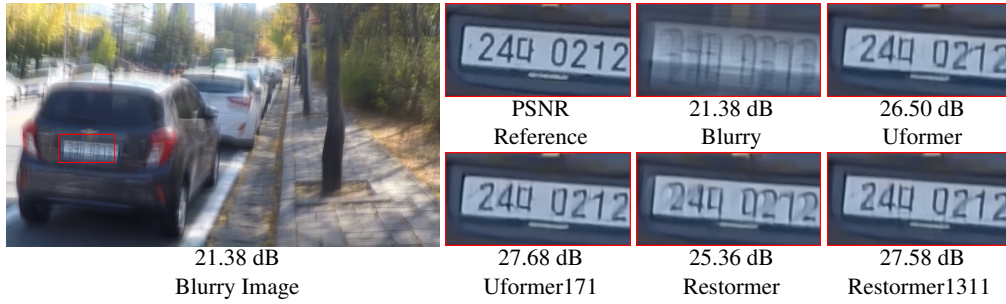
Figure 6: **Single image motion deblurring** on the GoPro dataset Nah et al. (2017). Compared to Uformer Wang et al. (2022), and Restormer Zamir et al. (2022), Uformer171 and Restormer1311 generate sharper and visually more faithful results.
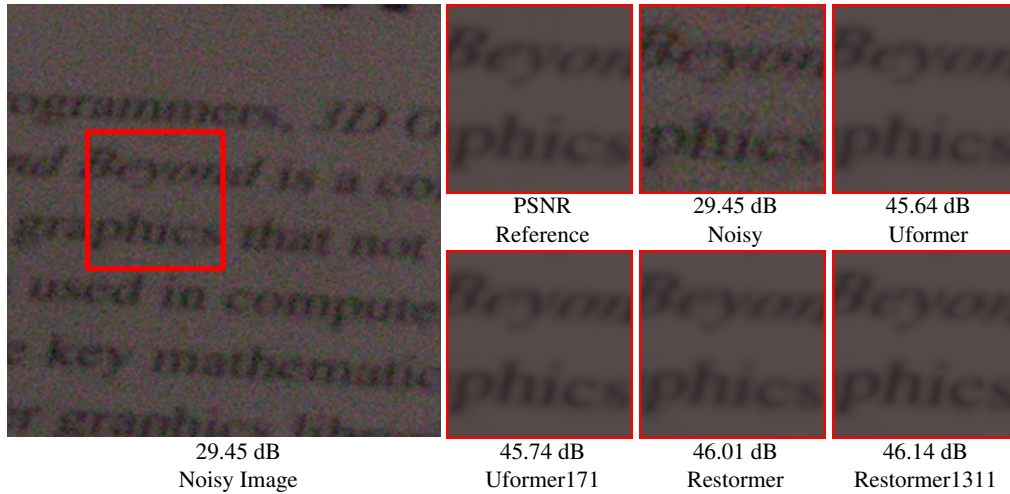

Figure 7: **Real image denoising** on the SIDD dataset Abdelhamed et al. (2018). Compared to Uformer Wang et al. (2022), and Restormer Zamir et al. (2022), Uformer171 and Restormer1311 generate cleaner and visually more faithful results.
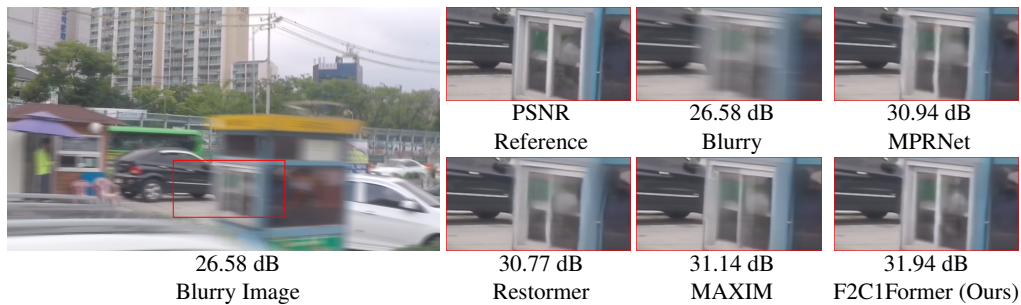

Figure 8: **Single image motion deblurring** on the GoPro dataset Nah et al. (2017). Compared to MPRNet Zamir et al. (2021), Restormer Zamir et al. (2022), and MAXIM Tu et al. (2022), F2C1Former(Ours) generates sharper and visually more faithful results.
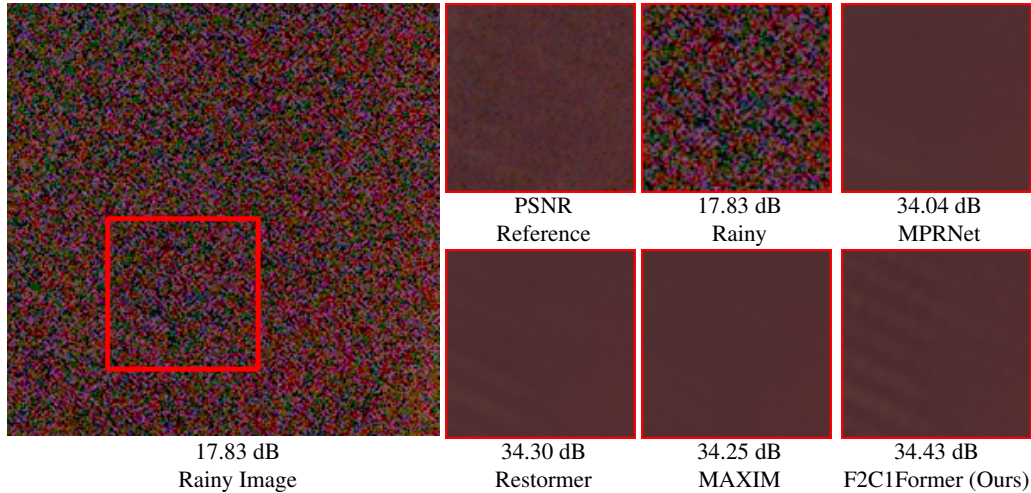
Figure 9: **Real image denoising** on the SIDD dataset Abdelhamed et al. (2018). Compared to MPRNet Zamir et al. (2021), Restormer Zamir et al. (2022) and MAXIM Tu et al. (2022), F2C1Former(Ours) generates cleaner and visually more faithful results.
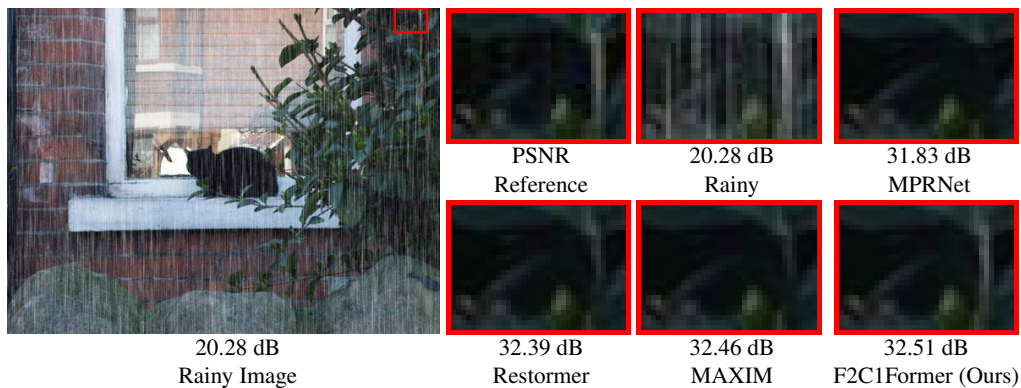

Figure 10: **Single image deraining** on the Rain14000 dataset Fu et al. (2017b). Compared to MPRNet Zamir et al. (2021), Restormer Zamir et al. (2022), and MAXIM Tu et al. (2022), F2C1Former(Ours) generates sharper and visually more faithful results.

C LIMITATIONS AND DISCUSSIONS

We have demonstrated the effectiveness of F2C1Former on image restoration tasks including image deblurring, image denoising and deraining. More experiments could be conducted on other image restoration tasks, such as dehazing, desnowing, *etc*.

Notwithstanding the extensive and promising results, there are still open questions remaining for transformer-based image restoration methods.

Is there a more effective transformer-like structure for image restoration? The success of MetaFormer demonstrates that the structure LN + FeatureMixing + LN + ChannelMLP matters for image restoration, even without attention modules. This reminds us that an effective structure is also vital for image restoration. Since MetaFormer is abstracted from transformers proposed first for NLP tasks, exploring structures customized for image restoration is also a meaningful task.

D   BROADER IMPACTS

Nowadays, image acquisition systems inevitably suffer from various degradations due to camera shakes, bad weather conditions, *etc*. Image restoration solves the problem well, and has important research and application value. Our proposed MetaFormer, such as Uformer171, Restormer131, and F2C1Former, achieve competitive restoration results. However, there are also potential negative societal impacts. For example, image restoration technology is prone to privacy leakage.

REFERENCES

Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1692–1700, 2018.

Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3155–3164, 2019. doi: 10.1109/ICCV.2019.00325.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European Conference on Computer Vision*, pp. 213–229. Springer, 2020.

Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE transactions on Image Processing*, 7(3):370–375, 1998.

Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *European Conference on Computer Vision*, pp. 171–187. Springer, 2020.

Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12294–12305, 2021a. doi: 10.1109/CVPR46437.2021.01212.

Liang Chen, Faming Fang, Tingting Wang, and Guixu Zhang. Blind image deblurring with local maximum gradient prior. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1742–1750, 2019. doi: 10.1109/CVPR.2019.00184.

Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 182–192, 2021b. doi: 10.1109/CVPRW53098.2021.00027.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pp. 17–33. Springer, 2022a.

Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Cross aggregation transformer for image restoration. *Advances in Neural Information Processing Systems*, 35:25478–25490, 2022b.

Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. Nbnet: Noise basis learning for image denoising with subspace projection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4894–4904, 2021. doi: 10.1109/CVPR46437.2021.00486.

Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4641–4650, 2021.

X Chu, L Chen, C Chen, and X Lu. Improving image restoration by revisiting global information aggregation. In *ECCV*, 2022.

Yuning Cui, Wenqi Ren, Sining Yang, Xiaochun Cao, and Alois Knoll. Irnext: Rethinking convolutional network design for image restoration. In *International Conference on Machine Learning*, 2023a.

Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The Eleventh International Conference on Learning Representations*, 2023b.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. doi: 10.1109/TIP.2007.901238.

Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017a. doi: 10.1109/TIP.2017.2691802.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1715–1723, 2017b. doi: 10.1109/CVPR.2017.186.

Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3843–3851, 2019. doi: 10.1109/CVPR.2019.00397.

Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1712–1722, 2019. doi: 10.1109/CVPR.2019.00181.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8343–8352, 2020. doi: 10.1109/CVPR42600.2020.00837.

Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3479–3489, 2020. doi: 10.1109/CVPR42600.2020.00354.

Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8183–8192, 2018. doi: 10.1109/CVPR.2018.00854.

Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8877–8886, 2019. doi: 10.1109/ICCV.2019.00897.

Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 254–269, 2018.

Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. *arXiv preprint arXiv:2303.00748*, 2023.

Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2736–2744, 2016.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

Chong Mou, Jian Zhang, and Zhuoyuan Wu. Dynamic attentive graph learning for image restoration. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4308–4317, 2021. doi: 10.1109/ICCV48922.2021.00429.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 257–265, 2017. doi: 10.1109/CVPR.2017.35.

Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. l_0 -regularized intensity and gradient prior for deblurring text images and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (2):342–355, 2017. doi: 10.1109/TPAMI.2016.2551244.

Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *European Conference on Computer Vision*, pp. 327–343. Springer, 2020.

Kuldeep Purohit, Maitreya Suin, A. N. Rajagopalan, and Vishnu Naresh Boddeti. Spatially-adaptive image restoration using distortion-guided networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2289–2299, 2021. doi: 10.1109/ICCV48922.2021.00231.

Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8592–8602, 2021. doi: 10.1109/CVPR46437.2021.00849.

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3932–3941, 2019. doi: 10.1109/CVPR.2019.00406.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Maitreya Suin et al. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3603–3612, 2020. doi: 10.1109/CVPR42600.2020.00366.

Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8174–8182, 2018. doi: 10.1109/CVPR.2018.00853.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning*, pp. 10347–10357. PMLR, 2021.

Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, pp. 146–162. Springer, 2022.

Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5769–5780, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17683–17693, 2022.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3872–3881, 2019. doi: 10.1109/CVPR.2019.00400.

Jie Xiao, Xueyang Fu, Feng Wu, and Zheng-Jun Zha. Stochastic window transformer for image restoration. *Advances in Neural Information Processing Systems*, 35:9315–9329, 2022.

Jie Xiao, Xueyang Fu, Man Zhou, Hongjian Liu, and Zheng-Jun Zha. Random shuffle transformer for image restoration. In *International Conference on Machine Learning*, pp. 38039–38058. PMLR, 2023.

Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6978–6986, 2017. doi: 10.1109/CVPR.2017.738.

Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1685–1694, 2017. doi: 10.1109/CVPR.2017.183.

Rajeev Yasarla and Vishal M. Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8397–8406, 2019. doi: 10.1109/CVPR.2019.00860.

Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. *Advances in neural information processing systems*, 32, 2019.

Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *European Conference on Computer Vision*, pp. 41–58. Springer, 2020.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2693–2702, 2020a. doi: 10.1109/CVPR42600.2020.00277.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, pp. 492–511. Springer, 2020b.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14821–14831, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 695–704, 2018. doi: 10.1109/CVPR.2018.00079.

He Zhang, Vishwanath Sindagi, and Vishal M. Patel. Image de-raining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3943–3956, 2020a. doi: 10.1109/TCSVT.2019.2920407.

Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5971–5979, 2019a. doi: 10.1109/CVPR.2019.00613.

Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson W.H. Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2521–2529, 2018a. doi: 10.1109/CVPR.2018.00267.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017a. doi: 10.1109/TIP.2017.2662206.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017b. doi: 10.1109/TIP.2017.2662206.

Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Björn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2734–2743, 2020b. doi: 10.1109/CVPR42600.2020.00281.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 286–301, 2018b.

Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In *arXiv preprint arXiv:1903.10082*, 2019b.

Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495, 2021. doi: 10.1109/TPAMI.2020.2968521.

Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: An efficient global modeling paradigm for image restoration. In *International Conference on Machine Learning*, pp. 42589–42601. PMLR, 2023.