

Multi-scale Channel Transformer Network を用いた単一画像の雨除去 Single Image Deraining using Multi-scale Channel Transformer Network

難波 優駿[†] 韓 先花[†]
Yuto Namba Xian-Hua Han

1. はじめに

深層畳み込みニューラルネットワークは、大規模データから汎用的な事前知識を学習することに優れており、雨除去などの画像復元タスクに広く応用されている。近年、自然言語処理の分野で著しい性能を示している Transformer が、画像復元の分野でも注目を集めている。Transformer は、CNN の欠点である局所的な受容野や入力画像ごとの適応能力の低さを解決している。しかし、一般的な Transformer では特徴の空間的な Self-attention を求めるため、計算量やメモリの使用は画像の解像度に対して二次関数的に増加してしまうという問題点がある。本研究では、離れた画素間の相互作用を捉える multi-scale のネットワーク構造を加え、空間的な attention ではなく Channel 方向の依存関係を自動的にモデリングする Channel Transformer Network を提案し、単一画像から雨除去の有効性を検証する。

2. 提案手法

本研究では雨画像からクリアな雨除去画像を復元するために、シングルスケールのメインサブネットとマルチスケールの補助 Transformer サブネットを組み合わせた Multi-scale Network を提案する。さらに、空間的な Self-attention を求める一般的な Transformer の代わりに、高解像度の画像に適用可能なチャンネル方向の Self-attention を求める Channel Transformer を提案する。Multi-scale Network に Channel Transformer を加えた Multi-scale Channel Transformer Network の概念図を図 1 に示す。

2.1 Multi-scale Network

提案する Multi-scale Network には Encoder・Decoder の補助サブネットを取り入れられている。これにより入力画像の解像度の雨情報のみでなく、マルチスケールな雨の情報が学習可能となる。Encoder ではレベルごとに画像サイズ H と W に対してそれぞれ 1/2 倍、チャンネル数はレベルに対してそれぞれ 2 倍、4 倍、8 倍になるようにダウンサンプルし、Decoder では Encoder と逆にアップサンプルを行い元の画像サイズに再構成を行う。これらの処理により、レベルが深くなるにつれて多くのコンテキストを学習できる。また、U-Net[2]のように Skip-connection を行うことにより、情報の損失を最小限に抑える。さらに、各ブロックには畳み込み層の代わりに、提案する Channel Transformer を取り入れており、レベルが深くなるにつれてブロック数を増やすことで、より高次元の情報を学習可能である。

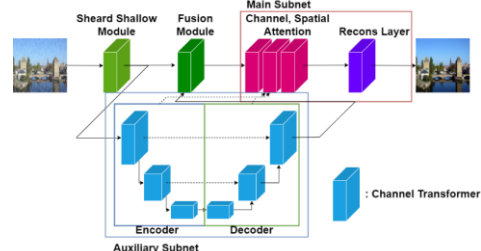


図 1 Multi-scale Channel Transformer Network

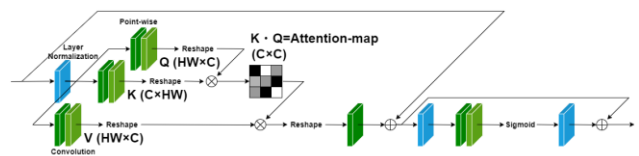


図 2 Channel Transformer

2.2 Channel Transformer

一般的な Transformer は畳み込み層の受容野の狭さや推論時の重みが固定で汎用性が低い点を解決しているが、入力の画像サイズ $H \times W$ に対して $O(H^2W^2)$ の Self-attention を求めるため、高解像度の画像を用いた画像復元には適用できない。そこで、提案する Channel Transformer(図 2)は、一般的な Transformer の query ($C \times HW$) と key ($HW \times C$)、value ($C \times HW$) の配列に対して転置を行い、転置した key ($C \times HW$) と query ($HW \times C$) の内積により Attention-map ($C \times C$) を求める。その後、転置した value ($HW \times C$) と Attention-map ($C \times C$) の内積を行う。これにより、得られる Self-attention が空間方向 $O(H^2W^2)$ の代わりにチャンネル方向 $O(C^2)$ となり、高解像度の画像にも適用可能である。また、Encoder で強化したコンテキストをさらに強調させることにより、多層に及ぶ雨の情報を学習できる。Point-wise convolution 層の後に Convolution 層を加えることで、空間的に詳細な情報も捉えることができる。

3. 実験評価

本研究では、Rain100L[3]、Rain100H[3]、Test100[4]、Test1200[5]、Test2800[6]のデータセットを用いて実験を行った。また、ネットワークモデルの有効性を示すための評価指標として PSNR[7]と SSIM[7]を用いて従来手法との比較を行った。

Rain100H を用いた提案手法と従来手法の雨除去画像を図 3 に示す。また、従来手法との定量評価値の比較を表 1 に示す。

[†] 山口大学 Yamaguchi University

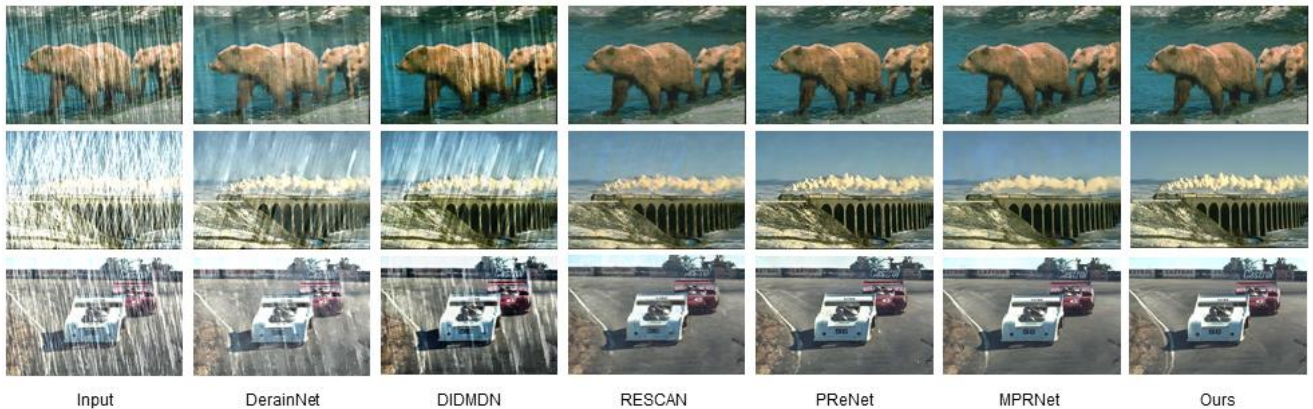


図 3 結果画像

表 1 従来法と提案手法の比較

Method	Rain100L[3]		Rain100H[3]		Test100[4]		Test1200[5]		Test2800[6]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet[8]	27.03	0.884	14.92	0.592	22.77	0.810	23.38	0.835	24.31	0.861
SEMI[9]	25.03	0.842	16.56	0.486	22.35	0.788	26.05	0.822	24.43	0.782
DIDMDN[10]	25.23	0.741	17.35	0.524	22.56	0.818	29.65	0.901	28.13	0.867
UMRL[11]	29.18	0.923	26.01	0.832	24.41	0.829	30.55	0.910	29.97	0.905
RESCAN[12]	29.80	0.881	26.36	0.786	25.00	0.835	30.51	0.882	31.29	0.904
PreNet[13]	32.44	0.950	26.77	0.858	24.81	0.851	31.36	0.911	31.75	0.916
MSPFN[14]	32.40	0.933	28.66	0.860	27.50	0.876	32.39	0.916	32.82	0.930
MPRNet[15]	36.40	0.965	30.41	0.890	30.27	0.897	32.91	0.916	33.64	0.938
SPAIR[16]	36.93	0.969	30.95	0.892	30.35	0.909	33.04	0.922	33.34	0.936
Ours	38.86	0.979	31.61	0.907	31.25	0.914	32.76	0.911	33.72	0.937

表 1 より、従来法と比べて PSNR と SSIM の両方で同等またはそれより良い結果を得られることが分かった。

4. 結論

本研究ではマルチスケールな雨情報を捉える Multi-scale Network に高解像度の画像に適用可能な Channel Transformer を加えた Multi-scale Channel Transformer Network による単一画像の雨除去法を提案した。

提案手法は従来手法と比較して PSNR と SSIM の両方でよい結果が得られることが確認された。

参考文献

- [1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. "An image is worth 16x16 words: Transformers for image recognition at scale", In ICLR, 2021.
- [2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation" In MICCAI, 2015
- [3] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, Shuicheng Yan, "Deep Joint Rain Detection and Removal from a Single Image", In CVPR, 2017.
- [4] He Zhang, Vishwanath Sindagi, and Vishal M Patel, "Image de-raining using a conditional generative adversarial network", TCSVT, 2019.
- [5] He Zhang and Vishal M Patel, "Density-aware single image de-raining using a multi-stream dense network", In CVPR, 2018.
- [6] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley, "Removing rain from single images via a deep detail network", In CVPR, 2017.
- [7] Z. 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, 2004.
- [8] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley, "Clearing the skies: A deep network architecture for single-image rain removal", TIP, 2017.
- [9] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu, "Semi-supervised transfer learning for image rain removal", In CVPR, 2019.
- [10] He Zhang and Vishal M Patel, "Density-aware single image de-raining using a multi-stream dense network", In CVPR, 2018.
- [11] Rajeev Yasarla and Vishal M Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining", In CVPR, 2019.
- [12] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining", In ECCV, 2018.
- [13] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng, "Progressive image deraining networks: A better and simpler baseline", In CVPR, 2019.
- [14] Kui Jiang, Zhongyuan Wang, Peng Yi, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang, "Multi-scale progressive fusion network for single image deraining", In CVPR, 2020.
- [15] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Multi-stage progressive image restoration", In CVPR, 2021.
- [16] Kuldeep Purohit, Maitreya Suin, AN Rajagopalan, and Vishnu Naresh Boddeti, "Spatially-adaptive image restoration using distortion-guided networks", In ICCV, 2021.