

Deep Attention ConvLSTM-UNet を用いた画像の降雨ノイズ除去 Rain Removal from a Single Image Using Deep Attention ConvLSTM-UNet

山道 航平[†] 韓 先花[†]
Kohei Yamamichi Xian-Hua HAN

1. はじめに

単一画像の降雨除去では Ill-posed (不良設定) 問題であるため、コンピュータビジョン分野において重要且つ困難なタスクである。近年、深層畳み込みネットワーク (CNN) のエンドツーエンド学習を用いた手法は、このタスクに大きな進歩をもたらした。しかし、従来の降雨除去における深層学習法では、主により深く且つ複雑なネットワーク構造を構築することで精度向上を図ったが、長時間の学習と大規模なトレーニングサンプルが必要という問題点がある。本研究では、シンプルな UNet 構造に基づく新たな Deep Attention ConvLSTM UNet 構造を提案した。提案したネットワーク構造では Encoder の特徴マップを Decoder の特徴へ単方向に連結するのではなく対応する Encoder と Decoder の特徴マップの時系列を見直し、逆順の ConvLSTM を用いて融合させる。さらに、SE Block を用いて学習された特徴マップの重要な領域を強調する Attention 機構を取り込むことで表現力を強化する Deep Attention ネットワークを提案した。二種類の降雨画像データベースを用いて検証実験を行い、提案手法の有効性を示した。

2. 提案手法

本研究では雨画像からクリアな画像を復元するために、U-Net 構造に基づいた新たなネットワーク構造の提案に焦点を当てる。U-Net は様々な CNN タスクに幅広く用いられているが、対応する Encoder と Decoder の特徴マップの融合は単純に Skip-Connection で連結されている。本研究は単方向に連結されている特徴マップの時系列を見直し、逆順の ConvLSTM を用いて融合させ、さらに特徴マップに重み付けをする Squeeze-and-Excitation block (SE-Block) [3] を採用することで、大域部と細部の特徴を効率的に統合する Deep Attention ConvLSTM-UNet を提案する。

2.1 Deep Attention ConvLSTM-UNet 構造

従来の U-Net 構造と同様に、提案する Deep Attention ConvLSTM-UNet (図 1) は Encoder と Decoder の 2 つの Path で構成され、それぞれの Path は 5 つのブロックに分けられる。各ブロックは kernel が 3*3 である 2 つの畳み込み層、それに繋がる kernel が 2*2 の Max Pooling 層と ReLU 活性化関数、SE-Block で構成されている。Encoder-Path において特徴マップの横や縦のサイズは各ブロックで 1/2 倍、その特徴マップの数は 2 倍となり、段階的に画像表現を抽出するため、ブロックの深さにつれてより多くのコンテキスト情報を考慮できる。これらの処理により、Encoder-Path の最終層には高度なセマンティック情報を持つ高次元の画像表現を生成することが可能となる。

Decoder-Path の各ブロックでは、まず前のブロックの出力に対して Up-sampling を行う。従来の U-Net では Encoder-Path での対応する特徴マップを Decoder-Path へコピーし、

Up-sampling された特徴マップと連結 (Skip-connection) することで Encoder と Decoder の特徴を統合させる。U-Net の Skip-connection による統合を以下の式に示す。

$$\mathcal{F}_{U-Net}(X_L^s, X_H^s) = \text{Concat}(X_L^s, X_H^s) \quad (1)$$

X_L^s, X_H^s は Encoder の s 番目のブロックの出力と Decoder の s' 番目ブロックの入力である。提案した Deep Attention ConvLSTM-UNet は単方向に連結するのではなく、対応している Encoder と Decoder の特徴マップの時系列を見直し、逆順の ConvLSTM ユニットを用いて効率的に融合させる。ConvLSTM ユニットによる統合法は以下の式に示す。

$$\mathcal{F}_{\text{ConvLSTM-UNet}}(X_L^s, X_H^{s'}) = \text{ConvLSTM}[(X_L^s, X_H^{s'})] \quad (2)$$

これにより、大域な特徴と局所的な特徴を双方考慮したネットワークを構築することができる。

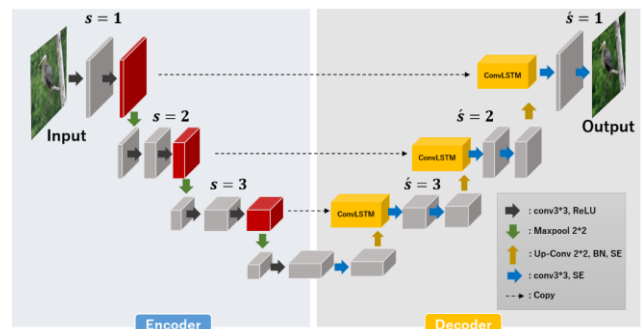


図 1. Deep Attention ConvLSTM-UNet のネットワーク構造

2.1.1 Attention ConvLSTM UNet

本研究では、Deep Attention ConvLSTM-UNet 構造に特徴間の空間的な相関をより効率的に捉えるために Attention Module として SE-Block (図 2) を取り入れている。このモジュールでは特徴マップの各チャンネルに重みを割り当てることで特徴マップの重要な部分を強調させる働きを持つ。実際には SE-block に入力された特徴マップに対し、Global Average Pooling (GAP) と Fully Connected Layer (FC) により圧縮され、最終的に $1 \times 1 \times F$ の形状で出力される。

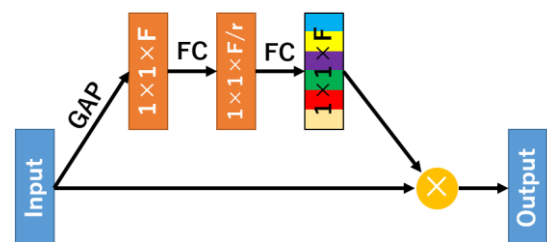


図 2. Squeeze-and-Excitation block (SE-block)

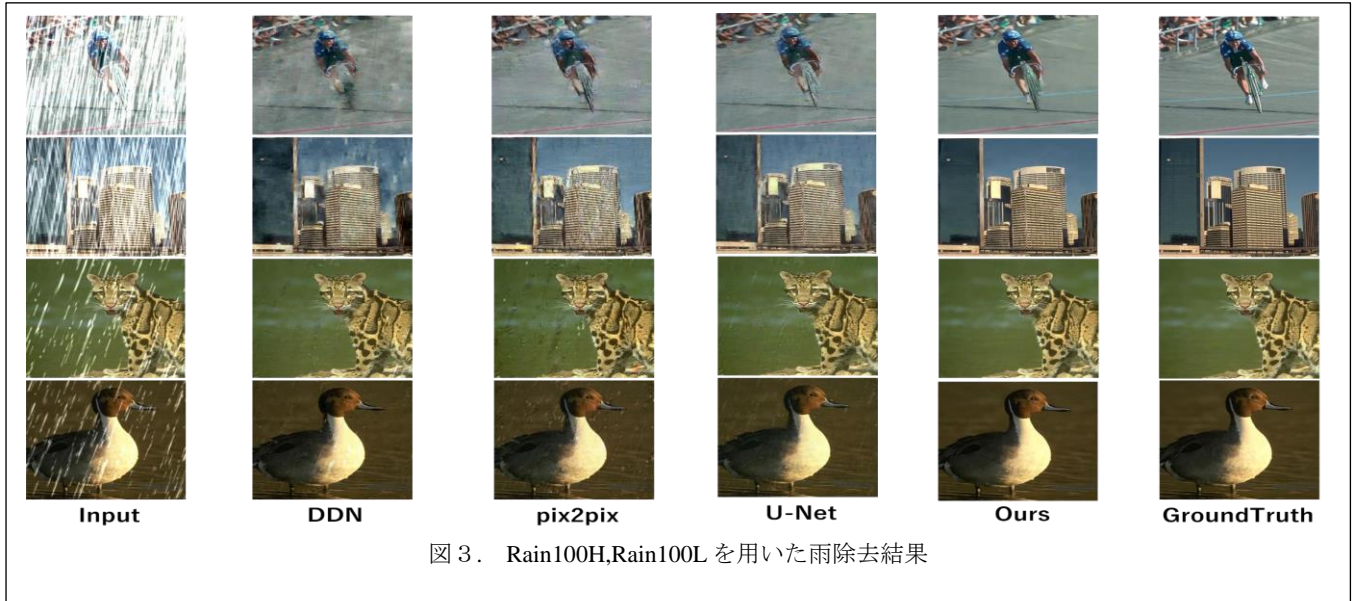


図 3. Rain100H,Rain100L を用いた雨除去結果

3. 実験評価

3.1.1 評価手法

本研究では公開されているデータセットである Rain100H[4]と Rain100L[4]の2つを評価に用いた. それぞれトレーニングデータに1800枚, テストデータに200枚を使用し, 定量評価に PSNR, SSIM[5]を用いて従来手法と比較を行った.

3.1.2 Deep Attention ConvLSTM-UNet による性能評価

提案した Deep Attention ConvLSTM-UNet の性能を, 実験を行い従来法と比較した. その結果を表1に示す. 表1から提案したネットワーク構造はベースの U-Net より PSNR と SSIM の両者とも大幅に向上され, その他の深層学習に基づく手法 (DDN[6],pix2pix[7],JORDER[8]) と比べても同等またはそれ以上であることが確認できた. また, 図3に Rain100H, Rain100L を用いた提案法と従来法の雨除去結果を示す.

表 1. 提案法と従来法との比較結果

| データセット | Rain100H | | Rain100L | |
|------------|----------|-------|----------|-------|
| | PSNR | SSIM | PSNR | SSIM |
| Methods | | | | |
| DDN[6] | 22.08 | 0.788 | 31.12 | 0.953 |
| pix2pix[7] | 21.96 | 0.679 | 29.20 | 0.886 |
| JORDER[8] | 26.08 | 0.821 | 35.23 | 0.968 |
| U-Net[9] | 23.27 | 0.831 | 24.61 | 0.859 |
| Ours | 27.06 | 0.848 | 35.23 | 0.962 |

4. 結論

本研究では単一画像から雨を除去するため, 比較的シンプルなネットワーク構造である U-Net に基づいた新たな Deep Attention ConvLSTM-UNet 構造を提案した. このネットワーク構造は従来手法と比較して, 高速且つ高精度の結果を得られ, 軽量のモデルであるため実用的である. 精度の面からも Rain100H, Rain100L の両者のデータセットにおいて向上していることが確認された.

参考文献

- [1] Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MIT Press MIT pp.326 - 366,2016
- [2] O. Ronneberger, P. Fischer, and T. Brox "U-net: Convolutional networks for biomedical image segmentation" In MICCAI,2015
- [3] Hu, J., Shen, L. and Sun, G. Squeeze-and-excitation networks. In CVPR 2018
- [4] 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 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1357-1366, 2017
- [5] 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, 13(4):600–612, 2004.
- [6] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding and J. Paisley. "Removing Rain from Single Images via a Deep Detail Network" In CVPR, 2017. pp. 1715-1723
- [7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros "Imagetoimage translation with conditional adversarial networks. I" In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [8] 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 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1357-1366, 2017
- [9] O. Ronneberger, P. Fischer, and T. Brox "U-net: Convolutional networks for biomedical image segmentation" In MICCAI,2015

† 山口大学創成科学研究科 Yamaguchi University
Graduate School of Sciences and technology for Inovation