

深層 Residual Attention ネットワークを用いた スナップショット撮影からハイパースペクトル画像の復元

Hyperspectral Image Reconstruction Using Deep Residual Attention Network from Snapshot Images

寄元 康平 *
Kohei Yorimoto

韓 先花 *
Xian Hua

1 はじめに

ハイパースペクトル画像 (以下 HSI) は RGB 画像のような 3 チャンネルで表現される画像データとは異なり, より多くの波長を計測した 3 次元テンソル型の画像データである. この画像データは色情報や特定の波長に反応する物質の分析が可能であり, リモートセンシング, 農業, 医療など様々な分野にて用いられている.

HSI を撮影するためには, 撮影するシーンの空間走査, 各ピクセルにおける波長の走査が必要となり時間が掛かるため, 動く物体の撮影や動画撮影が困難である. 近年では, 圧縮センシング技術をを用いた HSI の高速なスナップショット撮影方法が提案されている.

これまでの圧縮スナップショット画像の復元手法として, 画像処理技術を用いた数理モデルによる手法が提案されている [1]. しかし, 数理モデルによる復元手法では, 復元能力が低く, また復元までの時間が掛かるため実用的ではないとされている.

本論文では, 深層 Residual Attention を用いた HSI 復元ネットワークモデルを提案する. 提案するネットワークモデルは CNN から得られる特徴マップの空間と波長それぞれに対して Attention マップを学習させ, 残差成分をもとの特徴マップに足し合わせる深層 Residual Attention モジュール [2] を構築し, 高精度の復元能力と高速な復元処理が可能とする. 従来の手法と提案手法に対して 2 種類のデータセットを用いて実験を行うことで提案したモデルの有用性を示す.

2 深層 Residual Attention HSI 復元ネットワークモデル

本論文では, 深層 Residual Attention Network を複数回用いた HSI 復元ネットワークモデルを提案する. 図 1 に HSI 復元ネットワークモデルの構造を示す. 図 1 は, 圧縮されたスナップショット画像を入力, 復元した HSI を出力として扱う end-to-end 学習のネットワークモデルである. このネットワークモデルでは, 入力を畳み込み層を用いて復元するスペクトル数までにチャンネルを増やす. そののちに深層 Residual Attention モジュールと, 最初の畳み込み層で得られた特徴マップとの Residual 構造を用いることで入力画像の残差を復元させるように学習していく.

2.1 深層 Residual Attention モジュール

図 2 に図 1 における Residual Attention モジュールの具体的なネットワークモデルの構造を示す. このネットワークモデルは, Spatial Conv 層, Attention Map 層, Spectral Conv 層, Residual Conv 層から構成される.

Spatial Conv 層では, 入力された特徴マップ \mathbf{x}_{in} を畳み込み層と ReLU 関数を用いて空間方向への復元処理を行う.

Attention Maps 層では, Spatial Conv 層から得られた特徴マップに対して空間方向に対する Attention 特徴マップとチャンネル方向に対する Attention 特徴マップを算出し, 復元すべき画像領域, スペクトルを抽出する. 空間方向に対しては, 通常の畳み込みではなく Depthwise Convolution を用いる. チャンネル方向に対しては, 各チャンネルの分散を算出する Global Variance Pooling (GVP) と全結合層を用いてチャンネル間の相互関係を抽出する (式 (2), 式 (3)).

$$\boldsymbol{\mu}_{\mathbf{x}} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{x}_{\text{spa}}(i, j) \quad (1)$$

$$\mathbf{x}_{\text{gvp}} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (\mathbf{x}_{\text{spa}}(i, j) - \boldsymbol{\mu}_{\mathbf{x}}^{(i, j)}) \quad (2)$$

$$\mathbf{x}_{\text{spe_attn}} = \mathbf{W}_2 \mathbf{W}_1 \mathbf{x}_{\text{gvp}} \quad (3)$$

これらの特徴マップを足し合わせたのち, Sigmoid 関数を用いて注目すべき画像領域, スペクトルを算出し, Spatial Conv 層の特徴マップと要素ごとに関係を合わせることでより高度な空間の復元処理, 復元すべきスペクトルの抽出が可能となる. その後, 入力となる特徴マップ \mathbf{x}_{in} との Residual 構造を用いることで, Attention 特徴マップは空間, チャンネルにおける残差に対して学習させることが可能となる.

Spectral Conv 層では, Spatial Conv 層と Attention Map 層で得られた特徴マップを基に空間方向の復元処理を行う.

Residual Conv 層では, \mathbf{x}_{in} と Spectral Conv 層から得られた特徴マップとの Residual 構造を用いることで, \mathbf{x}_{in} の残差を復元させる.

3 実験

3.1 実験環境

本実験にて用いるデータセットは CAVE, Harvard の 2 種類である. CAVE は 32 枚の室内で撮影された $512 \times 512 \times 31$ サイズの HSI で構成されている. 20 枚を訓練データ, 12 枚をテストデータに分割する. Harvard は 50 枚の室内や外のシーンを撮影

* 山口大学 Yamaguchi University

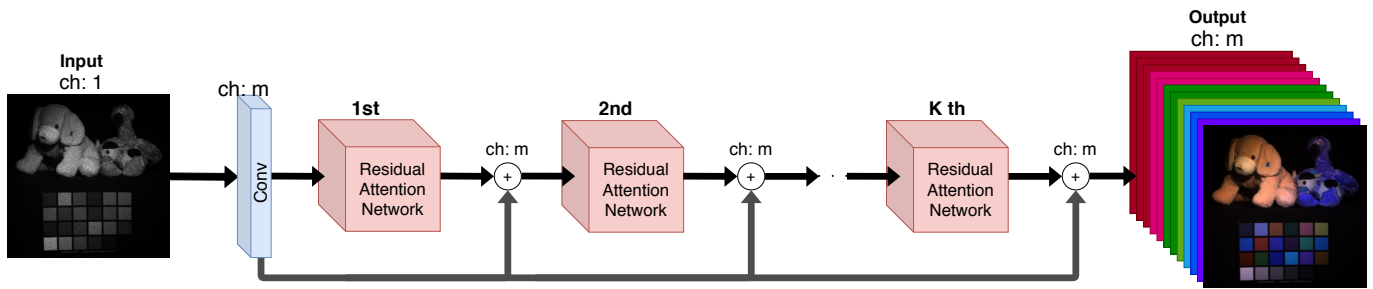


図1 HSI 復元ネットワークモデルの構造

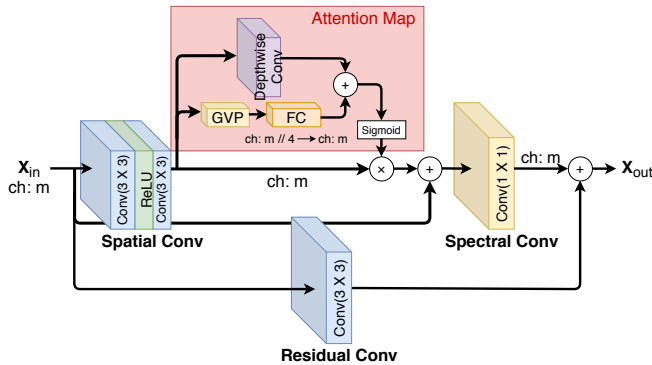


図2 Residual Attention モジュールの構造

表1 各データセット復元の評価値. 表中の(-)は評価値が不明となっている.

Dataset	Metrics	TwIST	HSCNN	DSSP	提案モデル
CAVE	PSNR	-	24.19	24.94	25.89
	SSIM	-	0.75	0.804	0.825
	SAM	-	0.219	0.146	0.109
Harvard	PSNR	27.16	35.77	36.32	36.59
	SSIM	0.924	0.93	0.939	0.939
	SAM	0.119	0.069	0.069	0.068

4 結論

本論文では、圧縮されたスナップショット画像をハイパースペクトル画像へ復元するために、深層 Residual Attention HSI 復元ネットワークモデルを提案した. 2 種類のデータセットを用いて従来の復元手法と比較実験を行い、同程度、もしくは高い復元能力を持つことを確認した.

参考文献

- [1] G. Arce, D. Brady, L. Carin, H. Arguello, and D. Kittle, "Compressive coded aperture spectral imaging: An introduction", IEEE Signal Processing Magazine, vol. 31, no. 1, pp.105-115, 2014
- [2] Jun-Hyuk Kim, Junho Choi, Manri cheon, and Jong-Seok Lee, "RAM: Residual Attention Module for Single Image Super-Resolution", arxiv.1811.12043, 2018
- [3] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multi-frame Image estimation for coded aperture snapshot spectral imagers", OSA Applied Optics, vol. 49, no. 36, pp. 6824-6833, 2010
- [4] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu, "HSCNN: CNN-Based Hyperspectral Image Recovery from Spectrally Undersampled Projections", IEEE International Conference on Computer Vision Workshops, pp.518-525, 2017
- [5] Laibao Wang, Chen Sun, Ying Fu, Min Hoi Kim, and Hua Huang, "Hyperspectral Image Reconstruction Using a Deep Spatial-Spectral Prior", IEEE Conference on Computer Vision and Pattern Recognition, pp.8024-8033, 2019

した $1040 \times 1392 \times 31$ サイズの HSI で構成されている. 40 枚を訓練データ, 10 枚をテストデータに分割する. これらのトレーニングデータに対して 64×64 にパッチ分割し画素値を $[0, 1]$ に正規化する. スナップショット画像に変換する際のマスクは, $p = 0.5$ のベルヌーイ分布に基づく 3 次元テンソルデータを各データセットごとに生成させる.

学習条件として, バッチサイズを 64, 150 エポックまで学習させる. 損失関数として MSE, Optimizer として Adam を使用する. 学習率の初期値を 1.0×10^{-3} とし, 50 エポックごとに 0.1 倍させる.

比較する生成手法は, 数理モデルのものとして TwIST[3], 深層学習を用いたものとして HSCNN[4], Deep Spatial-Spectral Prior(以下 DSSP)[5]を用いる.

生成された画像の評価方法は PSNR, SSIM, SAM を用いる. PSNR と SSIM は 1 チャンネルごとにおける空間方向の復元能力のを評価する. これらの評価値は大きいほど復元能力が高いと評価される. SAM は 1 ピクセルごとにおけるスペクトルの復元性能を評価する. この評価値は小さいほど復元能力が高いと評価される.

3.2 結果

表 1 に提案モデルと比較手法の各データセットの復元画像の評価値を示す. この表より, CAVE では空間, スペクトルでの復元能力が大幅に向上されていることがわかる. また, Harvard では数理モデルによる手法に対しては復元能力が大幅に向上されており, 深層学習による手法とは同程度の復元能力であることがわかる.