

Depthwise Separable Convolution を用いた RDDCNN の計算効率の向上

日大生産工 ○大木 太飛 日大生産工 山内 ゆかり

1. まえがき

近年、画像ノイズ除去の分野では、高精度かつ効率的な方法の開発が強く求められている。従来のCNNでは畳み込みにより元のノイズ分布が変化してしまい、学習が難しくなる場合があった。Oi Zhangらは、変形可能な畳み込み、残差ブロックを組み込んだRobust Deformed Denoising Convolutional Neural Network (RDDCNN)[1]を提案し、標準データセットや実カメラ画像において高いノイズ除去性能を示す報告がされている。しかし、計算コストが高くモデルサイズも大きいため、リアルタイム処理やカラー画像対応、未知ノイズへの汎用化などの点で課題がある。

本研究では、RDDCNNにおける計算コストの高さとモデルサイズの大きさの問題の解決を目指しDepthwise Separable Convolution[2]の導入を提案する。標準ノイズ除去ベンチマークデータセットにおける計算機実験により提案手法と従来のRDDCNNを比較し、推論時間短縮、およびノイズ除去精度の維持について報告する。

2. 従来研究

Robust Deformed Denoising Convolutional Neural Network (RDDCNN) [1]は、まず12層のDeformable Block (DB) で変形可能な畳み込みを積み重ねることで周囲のピクセルの関係を柔軟に捉え、複雑なノイズ成分を特徴的に抽出する。続いて5層のEnhanced Block (EB)では拡張畳み込みを用いて広い文脈的な情報を効率的に取り込み、少ないコストでノイズに強い特徴を学習する。さらにResidual Learning (RB)によって浅い層の情報を深い層へ橋渡しすることで長期依存問題を解消し、潜在的にクリーンな画像を生成する力を高めている。このようにDB・EB・RBが連携することで、RDDCNNはノイズに頑強で高精度な復元を可能にしている。

2.1 変形可能ブロック

12層の変形可能ブロックは、変形可能な学習可能カーネルとスタックされた畳み込みを組み合わせて、より代表的なノイズ特徴量を抽出する。DBの処理を式(1)で示す。

$$O_{DB} = DB(I_n) = 11RBC(R(DC(I_n))) \quad (1)$$

ここでは入力画像 I_n に対して変形可能な畳み込み $DC(I_n)$ を適用しここでカーネルの各点の位置は学習により自由に変形される。得られた活性化関数ReLUを通し $R(DC(I_n))$ が得られる。その後、このReLUの出力に対して11層の残差ブロック付き畳み込み RBC を順に適用し、最終的にDBの出力 O_{DB} が得られる。変形可能な畳み込み(DC)の計算は次式で表される。

$$y(p_0) = \sum_{p_n \in R} w(p_n)x(p_0 + p_n + \Delta p_n) \quad (2)$$

この式は、出力画像 y のある1ピクセル p_0 の値 $y(p_0)$ を計算する方法を示している。ここで y は畳み込み処理の結果として得られる画像全体を指し、 p_0 はその出力画像上の「今計算している特定の1点の座標」を表す。つまり $y(p_0)$ とは、出力画像 y の p_0 という位置にあるピクセル値のことになる。 R は受容野を表し、具体的にはカーネルの形状そのものを意味する。例えば 3×3 カーネルの場合、 R は中心を $(0,0)$ としたときの9個の相対座標の集合である。 w はフィルターの重みであり、学習によって獲得される値である。 p_n はカーネル内の相対位置を指し、例えば $(-1, -1)$ はカーネルの左上の点を意味する。したがって $w(p_n)$ は、「フィルター w の中で p_n に対応する重み」のことになる。

2.2 強化ブロック

より深いネットワーク構造を用いることで画像のノイズ除去に必要な複雑で文脈的な特徴を学習することである。特に拡張畳み込みを導入することで広い範囲の情報を取り込み、文脈的相互作用を強化してより効果的にできるノイズ処理を式(2)に示す。

$$\begin{aligned} O_{EB} &= EB(O_{DB}) \\ &= C(RBC(RBC(RBC(R(B(DC(O_{DB}))))))) \end{aligned} \quad (3)$$

ここで O_{EB} は強化ブロックの出力を示している。入力は O_{DB} で、これは前段の変形ブロック(DB)の出力を意味する。まず $DC(O_{DB})$ はその出力に対して拡張畳み込みを適用したものであり、通

常の畳み込みより広い範囲を捉えることができる。拡張畳み込みの計算は次式で示される。

$$y(p_0) = \sum_{p_n \in R} w(p_n)x(p_0 + d * p_n) \quad (4)$$

ここで $y(p_0)$ は出力画像の特定の1点 p_0 の値、 R はフィルターの重み p_n はカーネル内の相対位置、 x は入力特徴マップ、 d は拡張率を示す。通常の畳み込みでは、 $x(p_0 + p_n)$ として隣接ピクセルから値を読み取るが、拡張畳み込みでは $d * p_n$ により間隔を空けたピクセルを参照するため、広い範囲の情報を効率的に取り込むことができる。この後に式(3)の B が適用され、これはバッチ正規化を意味し、学習を安定させる役割を持ち、続く R はReLUで、非線形性を導入する活性化関数である。その後に3回繰り返されるRBCは、通常の畳み込み、バッチ正規化、ReLUを順に適用する処理であり、第2層から第4層に対応する。これによって特徴が精錬され、ノイズ除去に特化した表現が学習される。最後に C は通常の畳み込みを意味し、第5層で適用される。この層は出力を整えて次の再構成ブロックに渡す役割を担っている。

2.3 残差ブロック

残差ブロックでは、入力された特徴量をそのまま出力へ伝える経路と、畳み込みや活性化関数を通して変換した特徴量を加算する処理を行う。これにより、ネットワークは入力をそのまま利用するか、必要な変化のみを加えるかを選択でき、深い層構造においても学習が安定しやすくなる。式(5)に残差ブロックの処理の数式を示す。

$$I_c = RB(O_{EB}) = I_n - O_{EB} \quad (5)$$

ここで I_c は残差ブロックの出力を示している。入力画像 I_n は、ノイズが含まれている画像である。RDDCNNでは、この入力に対して畳み込み層を通し、残差ブロックRBによってノイズ成分を推定する。これにより得られる出力を O_{EB} とし、これは入力画像に含まれるノイズ成分であると学習される。その後入力画像 I_n から推定されたノイズ O_{EB} を差し引くことで、ノイズ除去後のクリーンな画像 I_c を得る。

2.4 Depthwise Separable Convolution [2]

Depthwise Separable Convolutionは、従来の畳み込み処理を2段階に分解して行う手法である。各チャネルごとに独立して畳み込みを行い、空間的な特徴を抽出する。この段階ではチャネル間の情報は混ざらない。次に、Pointwise Convolutionと呼ばれる 1×1 の畳み込みを適用し、異なるチャネル間の情報を統合する。この2段階構造によって、元の畳み込みと同様に

空間的・チャネル的な特徴を学習できるが、必要な計算量とパラメータ数を大幅に削減できる。

3. 提案手法

RDDCNNのノイズ除去モデルは、通常のCNNよりも複雑なノイズパターンを柔軟に学習し、高精度な画像復元を実現することができる。しかし一方で、畳み込み演算における計算コストが大きく、学習や推論に時間がかかるという課題がある。この問題点を改善するために、本研究ではDepthwise Separable Convolutionを導入することで、計算量の削減、モデルの軽量化、推論速度の向上を図る。具体的には、RDDCNN内部の畳み込み層に対して従来の標準的な畳み込みをDepthwise Separable Convolutionに置き換えることで従来の畳み込みに比べて演算量を大幅に削減しながら、ノイズ特徴の表現力を維持することを狙う。

4. 実験および検討

実験に用いるデータセットは、学習データ50000枚、テストデータ10000枚入ったCIFAR-10を用いて行った。各画像に対しては、独自に実装したガウシアンノイズ生成関数を用いてノイズを付加しノイズ付き入力画像を生成する。平均値を0.0とし、標準偏差を任意に設定したガウシアン分布に従う乱数を各画素に加えることで実現した。また、ノイズを加えた後の画素値が0.0から1.0の範囲を超えないようにクリッピング処理を行った。これらのデータを用いて通常のCNN、従来のRDDCNN、および本研究で提案するDepthwise Separable Convolutionを用いたRDDCNNを推論時間短縮、およびノイズ除去精度の結果で比較する。

5.まとめ

本研究では、高性能なノイズ除去モデルであるRDDCNNが持つ、モデルサイズが大きく計算コストが高いという問題に対し、Depthwise Separable Convolutionを導入した軽量なRDDCNNを提案する。提案手法はノイズ除去精度を維持しつつ、モデルサイズと推論時間を削減することを目指す。

参考文献

- 1) C. Tian, X. Zhang and Z. Wang, "A robust deformed convolutional neural network (CNN) for image denoising", *Circuits, Systems and Signal Processing* 41(10), (2022) pp. 5101–5123.
- 2) F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 1251–1258.