

# 自己注意によるコンテキスト抽出を用いた画像ノイズ除去手法

19C3020 有働 和矢

指導教員 宮田 高道 教授

## 1 背景

画像ノイズ除去はノイズが発生した画像から、ノイズを含まない原画像を推定することを目的としている。現在、画像のノイズ除去においては深層畳み込みニューラルネットワーク (CNN) を用いる手法が一般的であるが、このような手法の問題点として、畳み込み層の受容野が比較的狭く、長距離の画素間依存性を認識できないことが挙げられる。このため、既存の CNN を用いた画像のノイズ除去手法では、画像の持つコンテキスト (文脈情報) を判断することが困難であり、除去すべきノイズと残すべき細部の特徴とを区別できないことが知られている。

このような問題を解決するために、Transformer[1] に代表される自己注意機構の利用が画像処理の分野で広がっている。自己注意機構は自然言語処理の分野で開発された長距離間の依存関係を認識できる手法であり、ノイズ除去に応用することで画像のコンテキストを把握しながらノイズ除去を行うことが可能となり、高いノイズ除去性能が実現できることが明らかになっている。一方で、コンテキスト抽出とノイズ除去の二つのタスクを異なる二種類の CNNで行うことで効率的にノイズ除去を行える GTCNN とよばれる手法が提案されており [2], 画像ノイズ除去タスクにおいてコンテキストの把握とノイズ除去処理のすべてを自己注意機構のみで行うことは、CNN と同様に効率的でない可能性がある。

前述のように CNN におけるノイズ除去は二種類の CNN を用いることで効率的に行うことができる。このうちコンテキスト抽出タスクを行う CNN を自己注意機構に置き換えることで、既存手法と比較してより高いノイズ除去性能または効率的なノイズ除去性能を得ることを目的とする。

## 2 関連研究

### 2.1 Gated context CNN

Gated context CNN (以下 GTCNN) は、図 1 に示すように入力層、出力層、中間層である  $L$  個の GCBR 層 (Gated CBR Layer) からなる。GCBR 層はノイズ除去のみを行う CNN (CBR) とコンテキスト抽出のみを行う CNN (GTL) とに分離されている。GTL が抽出したコンテキストはゲート機構を介して CBR に渡されノイズ除去を制御する。このように機能分離を行うことで、機能分離を行わない既存手法と比較して高いノイズ除去性能を獲得でき、またパラメータ数の削減に成功した。GTCNN

の GTL はマルチスケール構造の U-Net をベースに設計されている。

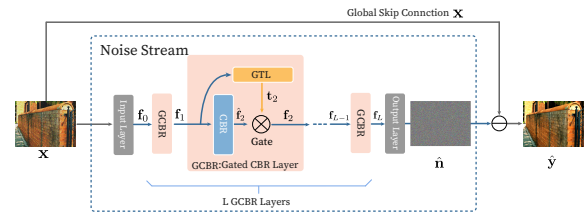


図 1: GTCNN のアーキテクチャ全体図。文献 [2] より引用。(C) 2020 Springer

## 2.2 自己注意機構

自己注意は式 (1) のように表される [1]。 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  は入力にそれぞれ異なる埋め込み処理をしたものである。

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

式 (1) では  $\mathbf{Q}$  と  $\mathbf{K}$  から得られる全データ間の類似度をもとに、 $\mathbf{V}$  を処理することを表しており、このことから、局所的な処理をする CNN とは異なり、自己注意は大域的な処理を行っているといえる。Transformer を画像認識に応用した ViT (Vision Transformer) ではこの特性のため、広域なコンテキストの取得を可能とした。しかし自己注意機構には、入力データ長の二乗に比例し計算量が増えてしまう課題が存在する。Restormer[3] はこの問題を解決するために U-net に類似した構造を持つネットワーク構造と、自己注意を画素位置ではなく特徴量のチャンネルに対して適用する転置自己注意を提案したが、ノイズ除去のすべてを自己注意で行うためにパラメータ数が増大することが課題であった。

## 3 提案手法

既存手法である GTCNN において、コンテキスト抽出を担う GTL は U-Net をベースに設計されている。本論文では GTL の各層に CNN 以上にコンテキスト抽出に Transformer を組み込んだ手法、SAGTCNN (Self-Attention GTCNN) を提案する。Transformer を挿入する場所を変更したいくつかのバージョンを提案し、それらの性能を比較する。実験を通して、Transformer のヘッドの数は 1 としている。提案手法は GTL を変更した以外は従来の GTCNN のアーキテクチャを踏襲している。

### 3.1 SAGTCNN-S4

SAGTCNN-S4 は図 2 のような構造になっている。青い枠の部分が従来手法に変更を加えたところである。従来手法では全ての層が CNN Block で形成されていたが SAGTCNN-S4 では 4 段目を SA Block に変更しているところに特徴がある。さらに、プーリングによるダウンサンプリングを畳み込みで行うようにし、バイリニアでのアップサンプリングを畳み込みで行うように変更した。SA Block は畳み込みを 2 度行う D-conv 層と Attention 層の 2 層からなっている。Transformer は画像のスケールが大きくなるほど計算量が増大するため、スケールの小さい 4 段目で Transformer を用いることで計算量の増大を防いでいる。

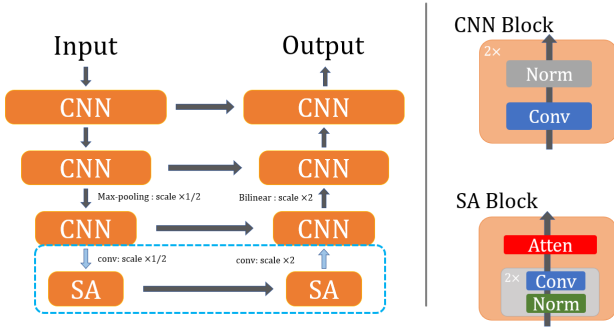


図 2: SAGTCNN-S4 の GTL のアーキテクチャ

## 4 実験

提案手法の有効性を調べるため、既存手法と提案手法のノイズ除去性能を比べた。その際、生成ノイズを加算したグレースケール画像に対するノイズ除去性能を比較した。本実験では、既存手法 GTCNN の CBR 層の数を 1 つに固定した。提案手法である自己注意機構を追加した GTCNN (Self-Attention GTCNN: 以後 SAGTCNN) は GTL の 4 層目の CNN を自己注意に置き換えた SAGTCNN-S4、ミドル層に自己注意を追加した SAGTCNN-M、そのどちらの変更を加えた SAGTCNN-S4M の 3 種類のパターンで比較を行った。GTCNN および SAGTCNN の GTL は共に 4 段で構成した。

データセットには DIV2K を用いる。DIV2K の全画像から  $192 \times 192$  画素のパッチをストライド 192 画素で切り出したものを学習画像とする。学習のエポック数は 600 とする。

既存手法と同様に、ノイズ除去評価手法において一般的に使われるガウシアンノイズを原画像に加算した画像 (生成ノイズ画像) から原画像を推定する。ノイズ除去結果の評価は PSNR を用いる。原画像には Set12, BSD68, Urban100 のデータセットを使用した。従来手法である

GTCNN の評価値は、論文に記載されている値を引用した。

表 1 に、ノイズの分散が  $\sigma = 30$  のときの SAGTCNN-S4 の結果を示す (これ以外の手法ならびに  $\sigma$  のときの結果は、卒業論文に記載する)。この結果より、 $\sigma = 30$  では提案する SAGTCNN-S4 が従来手法である GTCNN をわずかに上回る結果となった。一方で、これ以外の手法や  $\sigma$  では、GTCNN の性能を上回ることが無いことも示された。

表 1: 生成ノイズ  $\sigma = 30$  における GTCNN と SAGTCNN の性能比較

手法	Set12	BSD68	Urban100
GTCNN	29.80	<b>28.53</b>	29.43
SAGTCNN-S4	<b>29.81</b>	28.52	<b>29.46</b>

## 5 結論

本論文では、GTCNN のコンテキスト抽出を行う GTL に自己注意機構を組み込んだ手法、SAGTCNN を提案した。自己注意機構を挿入する場所を変更した SAGTCNN-S4, SAGTCNN-M, SAGTCNN-S4M の 3 種類のパターンで実験を行った。実験では GTCNN に対し SAGTCNN は一部上回ったが、明確な優位性を示すことができなかった。ノイズ除去性能の向上が見られなかった理由はいくつか考察できる。まず一般的に自己注意を用いる手法では大規模なデータセットを用いており、今回の実験で利用した学習用のデータセットでは学習画像枚数が少ない可能性が考えられる。また、自己注意は広域のコンテキストを捉えることを得意とするが、今回 GTL の 4 段目という画像サイズの小さい領域でしか自己注意を用いておらず、本来の性能が活かせなかったことが考えられる。最後に画像に自己注意を用いる際に ViT などで行われている位置埋め込みを行っていないため、コンテキストの抽出が上手く行えなかった可能性が考えられる。

今後の課題としては、位置埋め込みの実装やリアルノイズでの性能比較、自己注意を GTL の上層にて使用した際の有効性の検討などが挙げられる。

## 参考文献

- [1] Ashish Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, p. 5998 – 6008, 2017.
- [2] Kaito Imai et al. Gated texture cnn for efficient and configurable image denoising. In *European Conference on Computer Vision AIM Workshop*, pp. 665–681, 2020.
- [3] Syed Waqas Zamir et al. Restormer: Efficient transformer for high-resolution image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.