

MLP-Mixer のオーバーラップ処理

日大生産工 ○高橋 恭平 日大生産工 山内 ゆかり

1. まえがき

MLP-Mixer[1]とは2021年5月にGoogleの研究チームが発表した手法である。畳み込みニューラルネットワークやAttentionを使わず、多層パーセプトロンのみを用いた画像認識モデルであり、SOTAと比較しても見劣りしない精度を誇る。特にスループットに関しては優秀な性能をたたき出した。しかし、SOTAを上回る性能ではなく、CNNやTransformerなどのものと比較すると、性能や学習速度、パラメータ数が及ばないという問題がある。

本研究では、MLP-Mixerの性能面での問題においてパッチの範囲を拡大する手法を提案する。この提案手法を画像認識における計算機実験によりMLP-Mixerと比較し、報告する。

2. 従来研究

2.1 MLP-Mixer

Figure1に、Mixer-Layerの全体像を示す。

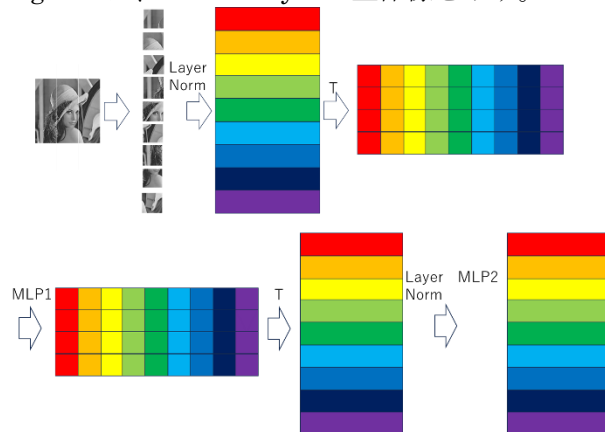


Figure1 Mixer-Layerの全体像

MLP-Mixerは、まず画像を $P \times P$ のパッチに分割する。それぞれベクトルにEmbedding、取得したベクトルを複数回Mixer Layerで処理し、Mixer Layerで処理されたベクトルをクラス分類処理するという3ステップで画像認識を行う。

MLP-Blockの構造をFigure2に示す。Blockの中では入力値をfcに通して活性化関数GELU、そしてもう一度fcに通して出力するといった構造になっている。

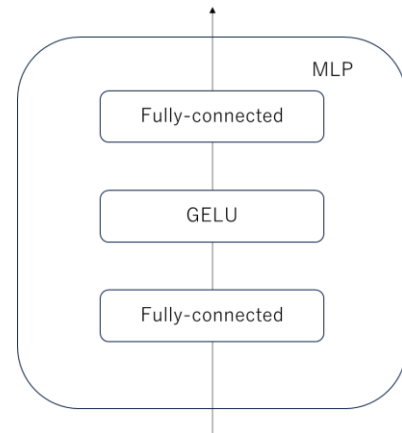


Figure2 MLP-Blockの構造

Mixer-Layerはtoken-mixingとchannel-mixingの二部分によって構成されている。式(1)にtoken-mixing、式(2)にchannel-mixingの式を示す。

$$U_{*,i} = X_{*,i} + W_{2\sigma}(W_1 \text{LayerNorm}(X)_{*,i}) \quad (1)$$

for: $i=1 \dots C$

$$Y_{j,*} = U_{j,*} + W_{4\sigma}(W_3 \text{LayerNorm}(X)_{j,*}) \quad (2)$$

for: $j=1 \dots S$

$X_{n,m}$ は画像を分割して得られたパッチテンソルの習合を $n \times m$ の二次元コテンソルに射影したもの、 n, m はそれぞれパッチの空間方向、チャンネル方向を表す軸、 $U_{n,m}, Y_{n,m}$ はそれぞれtoken-mixing、channel-mixingのMLPブロックからの出力、 W_i はそれぞれMLPブロック内の i 番目の全結合層、 σ はMLPブロック内のGELU関数を表す。

2.2 token-mixing

token-mixingは画像パッチのEmbeddingテンソル X の空間方向に対して全結合ネット層 W_1 による変換、活性化関数 σ の適用、全結合ネット層 W_2 による変換という処理を行い、各パッチの空間成分の情報を混合することが目的である。また、Figure1で行われている転置処理は、一回目は X の列方向に対してMLP-Blockの処理を実施するため、二回目は一回目に行った転置を元に戻すためである。

2.3 channel-mixing

channel-mixingはtoken-mixingからの出力 U のチャンネル方向に対して全結合ネット層 W_3 による変換、活性化関数 σ の適用、全結合ネ

ット層 W_2 による変換という処理を行いチャンネル成分の情報の混合することが目的である。

3. 提案手法

本研究では従来のものより、分割範囲を拡大して画像認識を行う。例として 30×30 の画像の場合、従来手法では 10×10 のパッチで均等に9分割するが、提案手法では 12×12 のパッチで9分割し、パッチ同士が一部分重なるようにする。従来研究のイメージを Figure3 に、提案手法のイメージを Figure4 に示す。



Figure3 従来研究の画像分割例

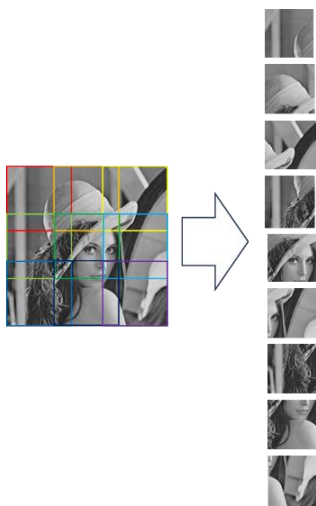


Figure4 提案手法の画像分割例

上記のように分割することで、断面部分をより正確に認識出来るようになる。

4. 実験および検討

MNIST 参考文献の画像を用いて、MLP-Mixer と提案手法の精度を比較する。提案手法では MNIST の画像を 24×24 とし、 9×9 パッチで 9 つに分割して実験を行う。

5. まとめ

本研究では、MLP-Mixer の問題点である、最先端のモデルに性能が劣るという問題を解決するため、オーバーラップ手法を用いて、MLP-Mixer の性能向上を目指す。提案手法を行うことで、断面部分をより正確に認識させ、性能を向上させることが出来ると考えられる。

参考文献

[1] Tolstikhin, Ilya O., et al. "Mlp-mixer: An all-mlp architecture for vision." Advances in neural information processing systems 34 (2021): 24261-24272.