

画像認識精度を保つ画像集合テンソル分解

加古 遼太郎 松井 勇佑

東京大学

1 序論

テンソル分解は、テンソルをいくつかのよりシンプルなテンソルで表現する手法の総称である。画像処理分野においては、ノイズ除去や信号復元、圧縮など、様々な応用がなされている [1]。しかし、テンソル分解した画像を機械学習で利用した場合における影響は明らかでない。

本論文では、CP 分解による低ランク近似を行った画像をニューラルネットワークを使って分類するタスクを取り上げ、圧縮率が精度にどのように影響するか議論する。

2 関連研究

n 階テンソル $\mathbf{A} \in \mathbb{R}^{a_1 \times \dots \times a_n}$ に対して、 $\mathbf{A}(i_1, i_2, \dots, i_n) \in \mathbb{R}$ は、 \mathbf{A} の (i_1, i_2, \dots, i_n) 成分であると定義する。ベクトルに対しても、1 階テンソルとみなして同様の記法を使用する。

n 階テンソル $\mathbf{A} \in \mathbb{R}^{a_1 \times \dots \times a_n}$ に対して、ある n 本のベクトルの組 $\mathbf{v}_1 \in \mathbb{R}^{a_1}, \mathbf{v}_2 \in \mathbb{R}^{a_2}, \dots, \mathbf{v}_n \in \mathbb{R}^{a_n}$ が存在して、

$$\mathbf{A}(i_1, i_2, \dots, i_n) = \mathbf{v}_1(i_1)\mathbf{v}_2(i_2)\dots\mathbf{v}_n(i_n) \quad (1)$$

を満たす時、 \mathbf{A} はランク 1 テンソルであるという [1]。この式は n 階のランク 1 テンソルが、 n 本のベクトルで表せることを示している。

n 階テンソル $\mathbf{A} \in \mathbb{R}^{a_1 \times \dots \times a_n}$ を r 個のランク 1 テンソル $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_r \in \mathbb{R}^{a_1 \times \dots \times a_n}$ を用いて、

$$\mathbf{A} = \sum_{i=1}^r \mathbf{P}_i \quad (2)$$

と分解することを CP 分解と呼ぶ [1]。これを満たすような r のうち最小の R を \mathbf{A} の CP ランク、または単にランクと言い、 \mathbf{A} はランク R テンソルであると言う。

$k \in \{1, \dots, R-1\}$ とする。 \mathbf{A} を近似するような

$$\hat{\mathbf{A}} = \sum_{i=1}^k \mathbf{Q}_i \quad (3)$$

を求めることを、CP 分解による \mathbf{A} の低ランク近似と言う。ただし、 $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_k \in \mathbb{R}^{a_1 \times \dots \times a_n}$ はランク 1 テンソルである。

3 提案手法

入力を $\mathbf{I} \in \mathbb{R}^{B \times C \times H \times W}$ とする。ただし、入力は複数の画像であり、 B はバッチの数、 $C = 3$ はカラーチャンネル (RGB) の数、 H は画像の高さ、 W は画像の幅である。この入力に対し、3 つの方法で CP 分解による低ランク近似をすることを考える。

- チャンネルごとの分解： \mathbf{I} を BC 個の $H \times W$ の 2 階テンソルの集合であると捉え、それぞれのテンソル $\mathbf{C} \in \mathbb{R}^{H \times W}$ を低ランク近似する。
- 画像ごとの分解： \mathbf{I} を B 個の $C \times H \times W$ の 3 階テンソルの集合であると捉え、それぞれのテンソル $\mathbf{S} \in \mathbb{R}^{C \times H \times W}$ を低ランク近似する。
- バッチごとの分解： \mathbf{I} をそのまま $B \times C \times H \times W$ の 4 階テンソルと捉え、低ランク近似する。

これらの低ランク近似を行ったのちに復元した画像に対し画像分類タスクを行って、その精度と画像テンソルによるメモリ消費を計測し、通常の画像データを入力とした場合と比較する。

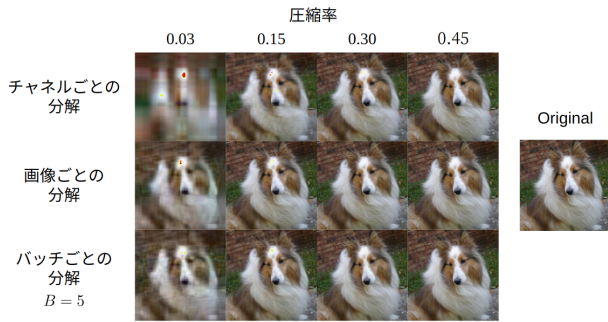


図 1: 低ランク近似した画像

4 評価

4.1 圧縮率の定義

異なるテンソルに対する低ランク近似を比較するために低ランク近似前後のパラメータの個数の変化率を表す圧縮率という概念を導入する. $\mathbf{A} \in \mathbb{R}^{a_1 \times \dots \times a_n}$ に対しランク k で低ランク近似をすることを考える. 低ランク近似後のパラメータの個数はランク 1 テンソルを複数のベクトルで表現したときのパラメータの個数であることに注意すると, このとき, 圧縮率 ρ は以下のように定義できる.

$$\rho = \frac{k(a_1 + a_2 + \dots + a_n)}{a_1 a_2 \dots a_n} \quad (4)$$

4.2 画像分類の精度

ネットワークとして ResNet-18 [2] を利用し, ILSVRC2012 の画像分類タスクをこれらの画像に対して実行した. まず, 低ランク近似した画像の例を図 1 に示す. 圧縮率 0.3 周辺から大きな画質の劣化は見られなくなる. 次に, 画像分類タスクの正解率 Acc@1 を図 2 に示す. チャンネルごとでは約 0.5, 画像ごとおよびバッチごとでは約 0.25 あたりまで精度を維持し, それ以下の圧縮率で精度が低下していくことが分かる. また, 図 1 とも比較すると, 圧縮率 0.1 付近は犬であると視認できる程度の画像であるにも関わらず, 画像認識の精度は大きく低下していることが分かる.

5 今後の課題

ある一定以上のランクでの低ランク近似は画像認識精度に大きな影響を及ぼさない一方で, 画像の

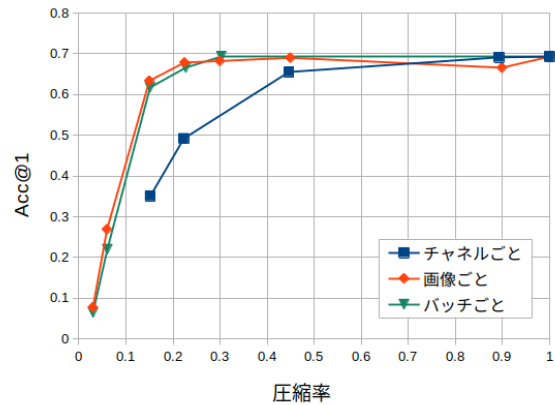


図 2: ResNet-18 による低ランク近似した画像の画像分類精度

上では大きな劣化が見られないにもかかわらず, 精度が大きく落ちるような分解があることも分かった. 人間とニューラルネットワークでは, 画像の周波数に対する分解能が異なるとする研究もあり [3], 人間が認識する上では影響しない要素が機械学習には影響していると考えられ, この点について検証を重ねる必要がある.

参考文献

- [1] Tatsuya Yokota. テンソル分解の基礎と応用. <https://speakerdeck.com/yokotatsuya/tensorufen-jie-falseji-chu-toying-yong-miru2022tiyutoriuru>, 2022. Meeting on Image Recognition and Understanding のチュートリアル講演にて発表.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [3] Ajay Subramanian, Elena Sizikova, Najib J Majaj, and Denis G Pelli. Spatial-frequency channels, shape bias, and adversarial robustness. *Advances in neural information processing systems*, 2023.