F-040

CNN 分類器と自己符号化器の再構成特徴に基づくドメイン別特徴抽出の検証

Domain-Specific Feature Extraction Evaluation Based on Reconstruction Features of CNN Classifiers and Autoencoder

代　美月[1)]　小端 千佳[2)]　神野 健哉[1)2)]
Mizuki Dai　Chika Obata　Kenya Jin'no

概要

本研究では，教師あり学習を施した畳み込みニューラルネットワーク（CNN）と，教師なし学習に基づく自己符号化器（AE）を用いて，ドメイン特性の異なるデータセットにおける特徴抽出挙動を比較検証する。具体的には，AE の再構成出力に対して分類器を適用し，その分類性能を評価する。さらに，t-SNE などの可視化手法を駆使して，得られた特徴空間上の分布構造を詳細に解析する。

1　まえがき

近年，深層学習に基づく特徴抽出は画像分類や異常検知，生成モデルなど様々なタスクにおいて中心的な役割を担っている．特に，教師あり学習による畳み込みニューラルネットワーク（CNN）は，高い分類性能を発揮するだけでなく，データから有用な特徴表現を自動的に獲得する能力を有している．一方で，教師なし学習を用いる自己符号化器（AutoEncoder; AE）も，データの潜在表現を抽出し，データの構造把握や次元削減，異常検知等に広く応用されている．

これらの手法は学習の目的が異なることから，抽出される特徴表現の性質も大きく異なる可能性がある．しかし，同一のネットワーク構造を用いた場合に，教師あり学習と教師なし学習がどのような特徴抽出の違いを生むのかについては，十分に体系的な検証が行われていない．

本研究の目的は，教師あり学習による CNN 分類器と，教師なし学習による自己符号化器に着目し，両者の特徴抽出挙動を比較検証することである．両モデルは共通の Encoder 構造を持つことで，学習目的以外の影響を排除し，学習タスクの違いによる特徴抽出の差異を明確化する．

2　先行研究と本研究の位置付け

先行研究である Cavallari ら [1] は、AE が抽出する特徴表現の判別性能について体系的に検証を行っている。彼らは MNIST や Fashion-MNIST といったグレースケールデータセットを用いて、様々な AE アーキテクチャで得られた潜在特徴を SVM による分類精度で評価し、事前学習済みの CNN 特徴と同程度の判別性能を達成できることを示した。これにより、AE は教師なしであっても十分に判別的な特徴を抽出可能であることが示唆されている。

しかし、Cavallari らの研究では色情報を含まないデータセットが対象であり、主に形状の違いが分類に寄与する状況下での検証となっている。一方、現実の画像認識タスクでは、形状以外にも色やテクスチャといった多様な属性が存在し、教師あり・教師なしの学習がそれらの抽出に与える影響は必ずしも明らかではない。

本研究では、Cavallari らの主張を補完・拡張するものであり、分類性能が高いことが必ずしも特徴抽出内容の同一性を意味しないこと、特に複数の属性が混在する場合には学習目的によって抽出される特徴の性質が大きく異なるという仮説に基づいて検証を行う。

3　実験方法

本研究では [2] で使用されたノイズ付き Colored-MNIST データセットの一部を使用し、色と形状が組み合わさった 4 クラス（Blue2, Blue6, Green2, Green6）を対象とした（図 1）。

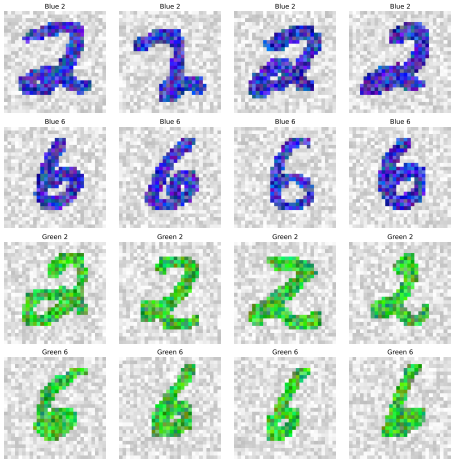

図 1　Colored-MNIST[2] の例

s

モデルは教師あり CNN 分類器と教師なし自己符号化器（AE）を構築し、両者の Encoder 部は共通構造とし

1）　東京都市大学大学院 総合理工学研究科 情報専攻
2）　東京都市大学 情報工学部 知能情報工学科

た。AE は同一の Encoder に Decoder を接続し、自己再構成学習を行った。

学習は Adam により行い、CNN はクロスエントロピー損失、AE は平均二乗誤差損失を最小化した。学習後、AE の潜在特徴を抽出し、t-SNE による可視化および分類器による分類精度評価を行った。

4　実験結果

4.1　特徴空間の可視化

本研究では、[3] より t-SNE の距離関数にコサイン類似度を用いた。図 2 は AE による潜在特徴の t-SNE 可視化結果である。形状（2 と 6）に基づくクラスタリングが確認される一方、色（青・緑）による分離は明瞭ではなかった。これは AE が主に形状情報を抽出していることを示唆する。

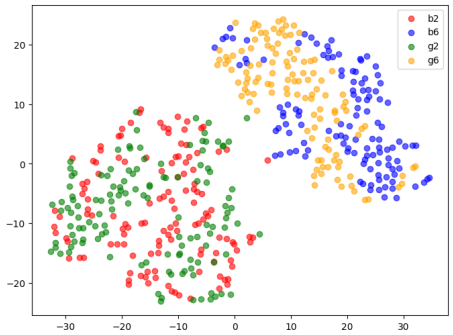

図 2　t-SNE で可視化した AE の潜在空間

図 3 は CNN 分類器の特徴空間の t-SNE 可視化結果である。教師あり学習では色と形の双方が識別に用いられ、色による分離も顕著に現れている。

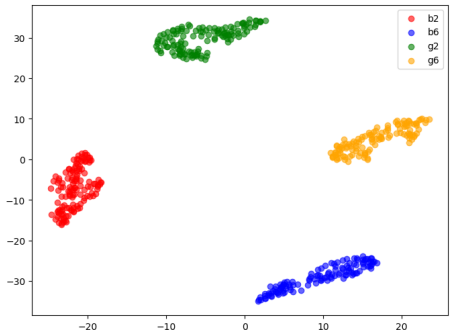

図 3　t-SNE で可視化した CNN の潜在空間

4.2　分類性能と混同行列

AE の潜在特徴を用いて分類器を学習させた結果の混同行列を図 4 に示す。色を跨ぐ誤分類は少なく、主に形状（2 と 6）の混同が観測された。色情報は分類には有効に利用されているが、可視化では明瞭に現れていなかったことがわかる。

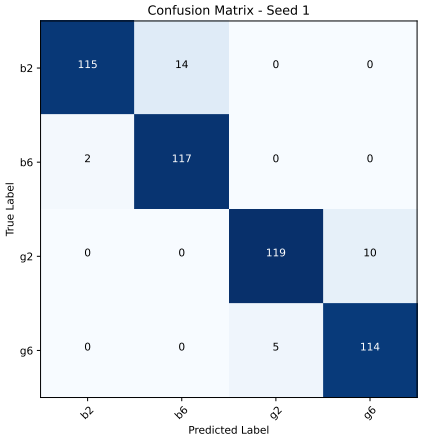

図 4　AE 特徴に基づく分類器の混同行列

以上の結果から、教師あり学習ではクラスラベルが色情報に依存するため、色の違いを積極的に識別特徴として利用する傾向があることが示された。一方、AE は再構成学習により形状特徴を主に抽出するが、色情報も一定程度保持しており、分類タスクでは色情報も活用されている可能性がある。ただし、t-SNE による低次元可視化では形状成分がより顕著に現れていたと考えられる。

5　まとめ

本研究では教師あり CNN と教師なし AE の特徴抽出挙動を比較検証した．Colored-MNIST のサブセットを用いた実験により，教師あり学習は色情報を強く抽出する傾向を，教師なし学習は形状情報を優先する傾向を確認した．分類精度の高さが必ずしも特徴抽出内容の一致を意味しないことを示し，学習目的による抽出特徴の性質の違いを明らかにした．今後はより複雑なドメインでの検証を進め，特徴抽出挙動の普遍性を検証していく．

謝辞

本研究は，科研費 JP23K11266，JP23K26077，JP24K15115，東北大学電気通信研究所共同プロジェクト研究【R06/B14】「深層学習における表現学習に関する研究」の助成によるものです．

参考文献

[1] G. B. Cavallari, L. S. F. Ribeiro, and M. A. Ponti, ”Unsupervised representation learning using convolutional and stacked auto-encoders: a domain and cross-domain feature space analysis,” *arXiv:1811.00473*, 2018.

[2] 若狭春輝, 神野健哉, ”対照学習による色と形の認識,” 電子情報通信学会 NOLTA ソサイエティ大会講演論文集, NLS-33, 2023 年 6 月.

[3] 代美月, 神野健哉, ”t-SNE を用いた中間層に対する分布の可視化,” 電子情報通信学会 2025 年総合大会講演論文集, N-1-20, 2025 年 3 月.