

自然環境映像データを対象とした距離学習 ニューラルネットワークによる画像特徴の分類手法

瀬尾幸斗¹ RAFLY ARIEF KANZA² 渡部展也³ KIN FUN LI⁴ 鷹野孝典¹

概要：本研究では、オートエンコーダから抽出した画像特徴ベクトルを対象として距離学習ニューラルネットワークを適用した画像特徴の分類手法を提案する。ニューラルネットワークから抽出した画像特徴ベクトルを利用した類似画像検索手法に関する研究が盛んである。画像分類性能が十分でない場合は、さらに距離学習ニューラルネットワークを適用することで、適切なランキング結果を得られるような画像特徴ベクトルに変換し、画像分類性能を向上することができる。提案手法では、オートエンコーダと距離学習ニューラルネットワークを連結した構成を持つモデルを構築することにより、オートエンコーダから抽出した画像特徴の再利用性を保つとともに、オートエンコーダと距離学習ニューラルネットワークを組み合わせた柔軟なモデル構築を実現することができる。実験では、国土地理院の航空写真データセットを用いた画像分類精度に関する評価を行い、提案手法の実現可能性を示す。

キーワード：オートエンコーダ、画像検索、検索高速化、自然環境観測、特徴ベクトル、特徴テンソル

A Method for Classifying Image Feature using Metric Learning Neural Network for Natural Environment Video Data

YUKITO SEO^{†1} RAFLY ARIEF KANZA^{†2}
NOBUYA WATANABE^{†3} KIN FUN LI^{†4} KOSUKE TAKANO^{†1}

1. はじめに

地球温暖化による気候の変化、海面の上昇、砂漠化、生態系の変化など、地球規模での自然環境問題の解決が緊急課題となっている。そのために、自然環境を常時観測し、動画や画像などの視覚データとして記録しておくことで、自然現象の時系列的な把握や類似する自然現象の抽出などによるデータ解析が重要な役割を担う。しかし、世界各地で自然環境を対象とした観測動画データを記録することを想定した場合、データ量が非常に膨大となるため、データ分析に要する処理コストも非常に高くなることが想定される。我々はこれまで、このような課題を解決するために、画像を入力としてニューラルネットワークにより抽出した低次元の画像特徴ベクトルを用いた画像シーン検索手法を提案してきた[1]。しかし、VGG 畳み込みニューラルネットワークなどの中間層を多く持つニューラルネットワークを用いて特徴ベクトルを抽出する際に、中間層によって得られる特徴ベクトルの性質にばらつきが生じるなどの課題があった。

本研究では、この手法を拡張し、画像特徴を抽出するネットワークとしてオートエンコーダ、および、距離学習ニ

ューラルネットワークを適用した画像特徴の分類手法を提案する。オートエンコーダを適用する理由は、ネットワーク構造が複雑ではなく実装しやすい点、エンコーダの出力として得られる画像特徴ベクトルの次元数を調節しやすい点などにある。しかし、画像特徴ベクトルをそのまま利用しても画像分類精度が高いとは限らないため、距離学習ニューラルネットワークを適用することで、正例、負例の教師データを用いた学習を実施する。これにより、画像分類精度を向上することが可能となる。提案手法では、オートエンコーダと距離学習ニューラルネットワークを連結した構成を持つモデルを構築することにより、オートエンコーダから抽出した画像特徴の再利用性を保つとともに、オートエンコーダと距離学習ニューラルネットワークを組み合わせた柔軟なモデル構築を実現することができる。

実験では、国土地理院の航空写真データセット[6]を用いて、提案手法の実現可能性を検証する。

2. 関連研究

オートエンコーダを画像特徴抽出器として適用し、抽出した画像特徴ベクトルを画像検索に適用する研究がなされている。

文献[1]では、自然環境観測カメラで取得される動画データを対象として、検索効率を向上するために、代表となる画像特徴ベクトルをインデックスとして設定する手法を提案している。文献[2]では、花押検索を目的として、花押という筆者署名の機能を持つ記号から、畳み込みオートエン

1 神奈川工科大学 情報学部 情報工学科
Department of Information and Computer Sciences, Kanagawa Institute of Technology
2 スラバヤ電子工学ポリテクニク
Politeknik Elektronika Negeri Surabaya
3 中部大学 中部高等学術研究所
Chubu Institute for Advanced Studies, Chubu University
4 ビクトリア大学 工学部 電子・コンピュータ工学科
Department of Electrical and Computer Engineering, Faculty of Engineering, University of Victoria

コーダを用いて字形特徴を抽出する手法を提案し、分類精度を評価している。文献[3]では、筆跡鑑定を目的として、字種ラベルという条件を付加した条件付きオートエンコーダを提案している。デコーダの画像生成部分のみに字種ラベルを付加して学習することで、エンコーダ部分は、筆跡画像から字種ラベルに依存しない筆者固有の書き癖を特徴として抽出する。文献[4]では、変分オートエンコーダとトリプレット損失を組み合わせたニューラルネットワークを構築し、類似性の概念を抽出するための埋め込み表現を学習するオートエンコーダモデルを提案している。

オートエンコーダは画像以外にも、様々なデータの特徴抽出器として利用されており、オートエンコーダとトリプレット損失を組み合わせたニューラルネットワークを構築し、ネットワークのトラフィックから攻撃や侵入を検知する手法[5]などが提案されている。

3. 提案手法

提案手法では、オートエンコーダを用いて圧縮された画像特徴を抽出し、距離学習ニューラルネットワークを適応することで画像特徴を分類可能な特徴ベクトルに変換する。変換された特徴ベクトルは、分類やランキングに適した特徴ベクトルとなっている。

下記に、提案手法の実行手順について述べる。

Step-1: 画像集合 $I = \{i_1, i_2, \dots, i_n\}$ を対象としてオートエンコーダを適用して学習させ、画像特徴が抽出可能な学習済みモデル AE を作成する。

Step-2: Step-1 で構築した AE のエンコーダの中間層の出力から、画像集合の画像特徴ベクトル $V = \{v_1, v_2, \dots, v_n\}$ を抽出する(図 1)。ここで、画像特徴として、畳み込み層の出力である m 階のテンソル $T = (t_1, t_2, \dots, t_m)$ を用いることもできる(図 2)。ここで、各 t_k は多次元ベクトルである。本研究では、この m 階のテンソルを特徴テンソルと呼ぶことにする。

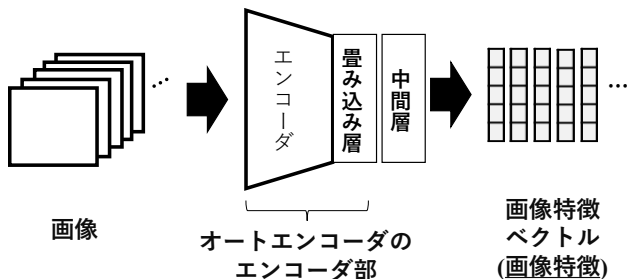


図 1 中間層による画像特徴ベクトル抽出

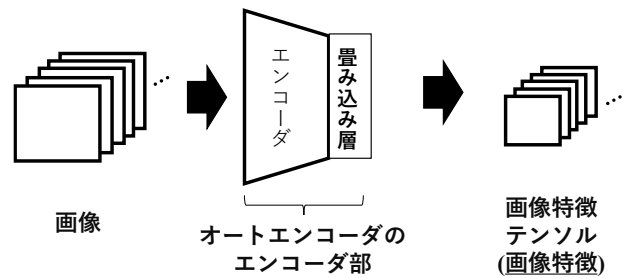


図 2 畳み込み層による画像特徴テンソル抽出

Step-3: 距離学習ニューラルネットワーク $MLNN$ を用いて、画像特徴ベクトル V または画像特徴テンソル T をランキングや分類に適した特徴ベクトル $V' = \{v'_1, v'_2, \dots, v'_n\}$ に変換する。図 5 中の Embedding は、類似する特徴ベクトルの距離関係が最適化された特徴ベクトルである。また、入力層は、入力の特徴ベクトルのデータ形式に合わせる。例えば、画像特徴ベクトル形式であれば全結合層、画像特徴テンソル形式であれば畳み込み層などを適用する。

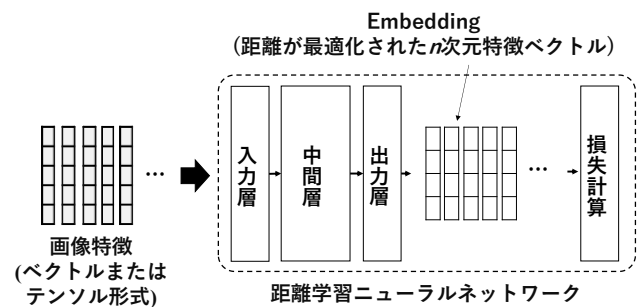


図 3 距離学習ニューラルネットワークによる画像特徴の最適化

4. 実験

提案手法により抽出した特徴ベクトルを用いた画像分類精度から、提案手法の実現可能性を確認する。

4.1 実験環境

画像データセットとして、国土地理院の航空写真データセット[6]を利用する。航空写真は 576 枚からなり、「田畑」、「水源地帯」、「森林」、「建物」の 4 つクラスに 144 枚ずつ分類した。画像データの例を図 4 に示す。

オートエンコーダ AE は、畳み込み層を持つもの 1 種類を用意する。また、距離学習ニューラルネットワークは畳み込み層を持たない距離学習 $NN-1$ 、畳み込み層を持つ距離学習 $NN-2$ の 2 種類を用いる。オートエンコーダ AE の学習には、航空写真 576 枚を用いて訓練する。距離学習ニューラルネットワークの学習は、オートエンコーダ AE から抽出された画像特徴(画像ベクトルまたは画像テンソル)または画像データを用いて教師付きで訓練する。

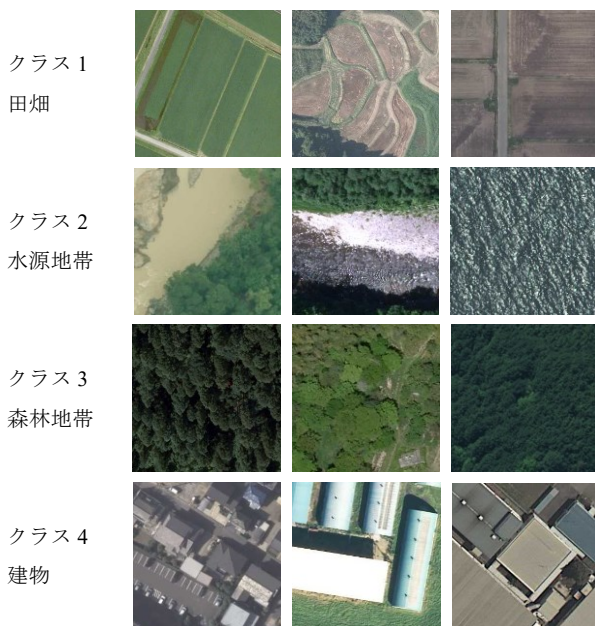


図 4 航空写真データの例 (4 分類)

4.2 実験方法

下記の 6 つのモデルについて k -means を適用して分類した精度を算出し、結果を比較する。それぞれのニューラルネットワークにおいてオートエンコーダの中間層および距離学習ニューラルネットワークの出力層の次元を、1000 次元、500 次元、100 次元で変化させる。

- M1 : AE のみで構成する。AE のエンコーダからの出力を特徴ベクトルとする。
- M2 : AE のみで構成する。AE のエンコーダ内にある畳み込み層の出力をベクトル化したものを特徴ベクトル (32*32*64 次元) とする。
- M3 : M1 の AE と距離学習 NN-1 で構成する。M1 の AE から抽出した特徴ベクトルを、距離学習 NN-1 で変換したものを特徴ベクトルとする。
- M4 : M2 の AE と距離学習 NN-1 で構成する。M1 の AE から抽出した特徴ベクトルを、距離学習 NN-1 で変換したものを特徴ベクトルとする。
- M5 : M1 の AE と距離学習 NN-2 で構成する。M1 の AE からエンコーダ内にある畳み込み層から抽出される特徴テンソル (32 次元, 32 次元, 64 次元) を、距離学習 NN-2 で変換したものを特徴ベクトルとする。
- M6 : 距離学習 NN-2 のみで構成する。画像を特徴テンソル (256 次元, 256 次元, 3 次元) として入力とし、距離学習 NN-2 で変換したものを特徴ベクトルとする。

再現率、適合率は下記の式で算出する。

$$\text{再現率} = \text{正解数} / \text{正解総数} \quad (1)$$

$$\text{適合率} = \text{正解数} / \text{正解と判定した数} \quad (2)$$

4.3 実験結果

実験結果を図 5～図 7、表 1 に示す。図 5、図 6 は、各モデルの再現率、適合率の比較結果を示している。モデル M1 と M2 はオートエンコーダのみで構成され、モデル M3 と M4 はオートエンコーダと距離学習ニューラルネットワークから構成されるモデルである。モデル M3 と M4 が再現率、適合率ともに向上しているため、距離学習ニューラルネットワークの適用により、画像分類精度を向上することが可能であることが確認できる。

また、モデル M5 は、M3、M4 と同様にオートエンコーダと距離学習ニューラルネットワークから構成されるが、距離学習ニューラルネットワークに畳み込み層をもつモデルである。一方、モデル M3 と M4 は距離学習ニューラルネットワークに畳み込み層ではなく、全結合層で構成している。モデル M5 の再現率、適合率が向上していることから、画像分類においては、距離学習ニューラルネットワークに畳み込み層を用いて構成した方が分類精度向上の効果が高いことがわかる。

モデル M6 はオートエンコーダを使用せずに、距離学習ニューラルネットワークのみで構成され、入力には抽出した画像特徴ではなく、画像データそのものを用いる。オートエンコーダの中間層および距離学習ニューラルネットワークの出力層を 100 次元と 1000 次元とした場合に、M6 の再現率、適合率が向上する結果となった。

ここで、表 1 および図 7 は、モデル M5 と M6 における平均再現率と平均適合率および学習時(100 次元)の損失変化の比較結果を示している。表 1 より、100 次元、500 次元、1000 次元の平均再現率は M5 の方が高く、平均適合率は M6 の方が若干高いが、平均適合率の差分は 0.014 と小さいことがわかる。また、図 7 の結果から、M5 の方が、損失値の変動が小さく学習が安定していることが確認できる。また、学習時において損失値がともに 1.2 付近から開始しており、オートエンコーダから抽出した特徴テンソルを用いた場合および画像データを用いた場合ともに、学習開始時点では距離の関係が十分に適切でないことがわかる。なお、M5 の学習エポック数が小さいのは、オートエンコーダで学習したことにより、距離学習に適切な画像特徴が抽出されているためと考えられる。

以上、モデル M5 と M6 の比較結果から、オートエンコーダと距離学習ネットワークを分離して構成することは、分類精度および学習の安定性の観点から一定の効果があると考えられる。

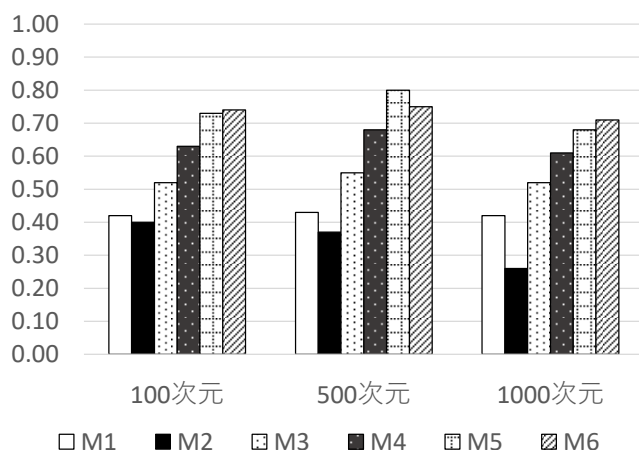


図 5 各モデルの再現率の比較

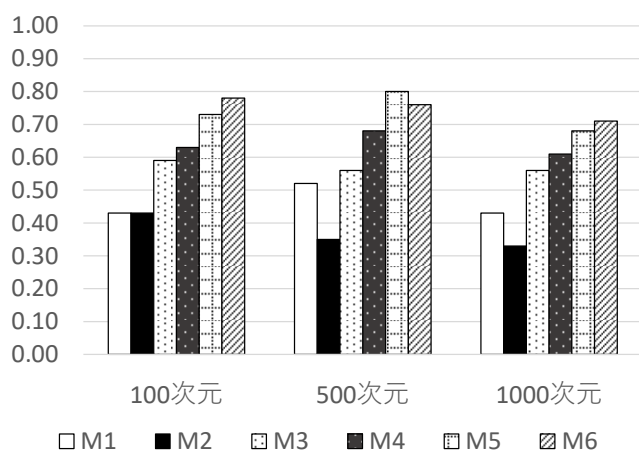


図 6 各モデルの適合率の比較

表 1 平均再現率と平均適合率の比較 (M5 と M6)

	M5	M6
エポック数	600	1000
平均再現率	0.736	0.733
平均適合率	0.736	0.750

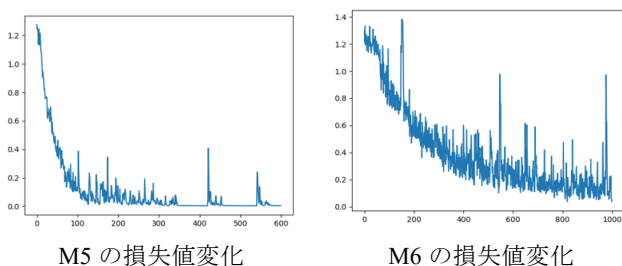


図 7 学習における損失値の比較

5. おわりに

本研究では、画像特徴を抽出するネットワークとしてオートエンコーダ、および距離学習ニューラルネットワーク

を適用した画像分類手法を提案した。航空写真データを用いた実験において、提案手法を用いて、距離学習ニューラルネットワークを適用することにより、画像特徴の分類精度を向上させることができることを示し、提案手法の実現可能性を確認した。

今後の予定として、提案手法を適用して、自然環境観測において異なる複数地点に設置された観測カメラから記録・蓄積される映像データを対象とした映像シーン分析システムの構築に応用していくことを検討している。

謝辞 本研究は問題複合体を対象とするデジタルアース共同利用・共同研究拠点 2023 年度共同研究, および JSPS 科研費 23K11120 の助成を受けたものです。

参考文献

- [1] Hiroki Mimura, Masaya Tahara, Kosuke Takano, Nobuya Watanabe, Kin Fun Li: Video Indexing for Live Nature Camera on Digital Earth, International Conference on Advanced Information Networking and Applications, pp.660-667 (2023).
- [2] 鬼塚洋輔, 大山航, 山田太造, 井上聡, 内田誠一: 花押類似検索のための畳み込みオートエンコーダによる画像特徴抽出, じんこんもん 2018 論文集, pp.257-262 (2018).
- [3] 細江麻梨子, 山田智輝, 加藤邦人, 山本智一: 条件付き AutoEncoder による書き癖抽出手法の提案, 情報処理学会第 80 回全国大会, 2C-06, No. 2, pp. 37-38 (2018).
- [4] Haque Ishfaq, Assaf Hoogi, Daniel Rubin: TVAE: Triplet-Based Variational Autoencoder using Metric Learning, arXiv:1802.04403 [stat.ML], (2018).
- [5] Giuseppina Andresini, Annalisa Appice, Donato Malerba: Autoencoder-based deep metric learning for network intrusion detection, INFORMATION SCIENCES, Volume 569, pp. 706-727 (2021).
- [6] 国土地理院: CNN による水田抽出のための教師画像データ, 国土地理院技術資料 H1-No.26 (2023).