

センサーデータ検索・分析・予測のための深層学習モデルを用いた 特徴量データベースの構築

鷹野孝典*、渡部展也**

*神奈川工科大学、**中部大学国際 GIS センター

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。

このような問題に対処するために、昨年度は、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た[1]。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。オートエンコーダは教師なしで画像等のデータを複製するように学習するニューラルネットワークであり、各種データを対象とした特徴抽出モデル[2][3][4][5]として利用できる。距離学習ニューラルネットワークは、正例や負例を用いた学習によりデータ間の距離を最適化するニューラルネットワークであり、教師なし学習を導入した手法[6]やオートエンコーダに距離学習ニューラルネットワークを組み込む手法[7]も提案されている。

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て、提案手法の実現可能性を検証する。

2. 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。提案手法では、図1と図2に示すように、画像特徴が抽出可能な学習済みモデルとして構築したオートエンコーダ AE の中間層の出力から、画像集合の画像特徴ベクトル $V = \{v_1, v_2, \dots, v_n\}$ 、畳み込み層の出力である m 階の特徴テンソル $T = (t_1, t_2, \dots, t_m)$ を抽出する。

さらに、図3に示すように、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。ここで、図3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。

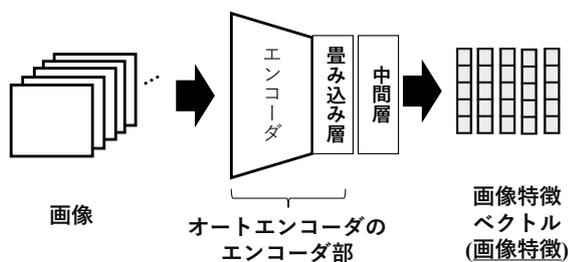


図1 中間層による画像特徴ベクトル抽出

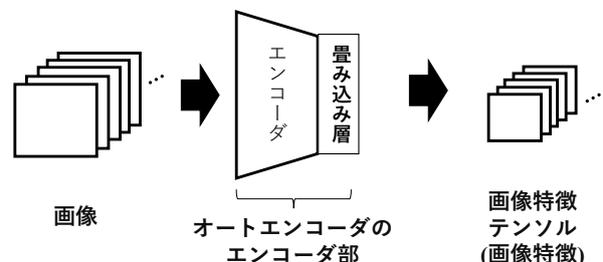


図2 畳み込み層による画像特徴テンソル抽出

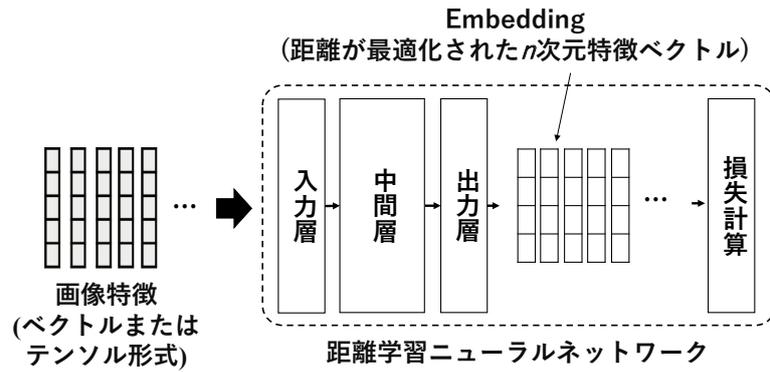


図 3 距離学習ニューラルネットワークによる画像特徴の最適化

3. 結果

学習用データとして国土地理院の航空写真データセット[8]の 576 枚を利用し、「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(図 4)。



図 4 航空写真データの例 (4 クラス)

オートエンコーダ AE は、畳み込み層を持つもの 1 種類、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 種類を用意した。これらのオートエンコーダ AE と距離学習ニューラルネットワーク NN-1、NN-2 を組み合わせることにより、下記のように 6 つのモデル M1～M6 を構築・訓練した。

- M1 : AE のみで構成する。AE のエンコーダの出力および畳み込み層の出力を、それぞれ特徴ベクトルおよび特徴テンソルとする。
- M2 : AE のみで構成する。AE のエンコーダの畳み込み層の出力をベクトル化し、特徴ベクトルとする。
- M3 : M1 の AE と NN-1 で構成する。M1 の特徴ベクトルを NN-1 で変換し、特徴ベクトルとする。
- M4 : M2 の AE と NN-1 で構成する。M2 の特徴ベクトルを NN-1 で変換し、特徴ベクトルとする。
- M5 : M1 の AE と NN-2 で構成する。M1 の特徴テンソルを NN-2 で変換し、特徴ベクトルとする。
- M6 : NN-2 のみで構成する。画像を特徴テンソルとして入力して NN-2 で変換し、特徴ベクトルとする。

実験結果として、学習した6つのモデルについて、各モデルの再現率、適合率の比較結果を図5、図6に示す。また、図7にt-SNEによるクラスタリング結果を示す。さらに、M5とM6の比較のため、表1に平均再現率と平均適合率の比較結果を示す。なお、再現率、適合率は下記の式で算出した。

$$\text{再現率} = \text{正解数} / \text{正解総数} \tag{1}$$

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

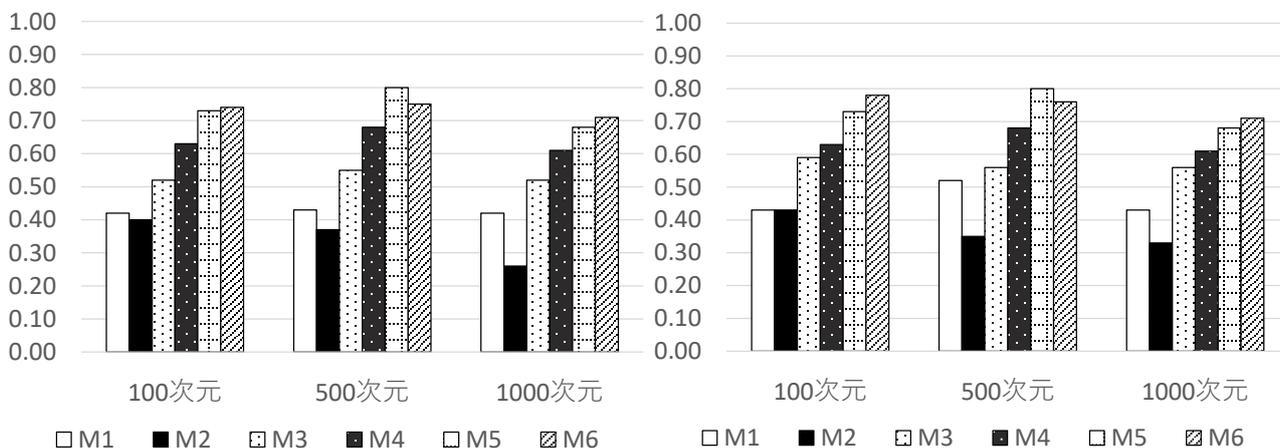


図5 各モデルの再現率の比較

図6 各モデルの適合率の比較

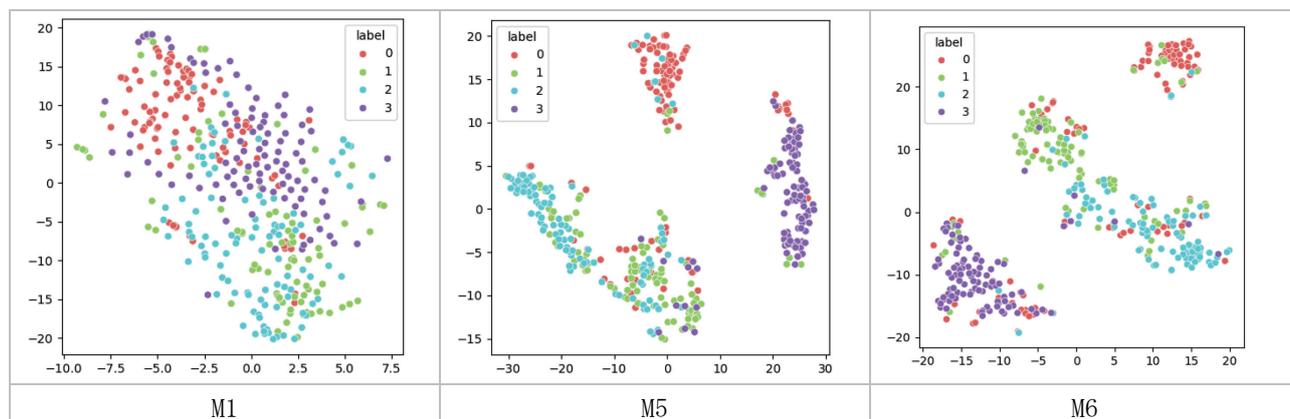


図7 t-SNEによるクラスタリング結果

表1 エポック数の比較 (M5とM6)

	M5	M6
エポック数	600	1000

4. 考察

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。図5と図6の結果より、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。

また、図 7 より、距離学習ニューラルネットワークを適用した M5 と M6 においてデータ間の距離関係が改善されたクラスタリング結果となっていることがわかる。これにより、動画中のシーン検索において、適切なランキングスコアが算出可能となることが期待される。

一方、表 1 に示すように、M5 と M6 の比較において、M5 の方が学習エポック数が少ない結果となった。これは、M5 では事前学習済みのオートエンコーダより抽出した特徴を距離学習に用いており、その結果距離学習の安定性をもたらしたためと考えられる。これは、提案手法が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ことを示している。

5. プロトタイプの実装

観測カメラの実装に Raspberry Pi を用いて試作したプロトタイプの概観を図 8 に示す。撮影画像は 5 秒ごとにサーバに送信され、サーバ上にてオートエンコーダより特徴ベクトルとして保存される。

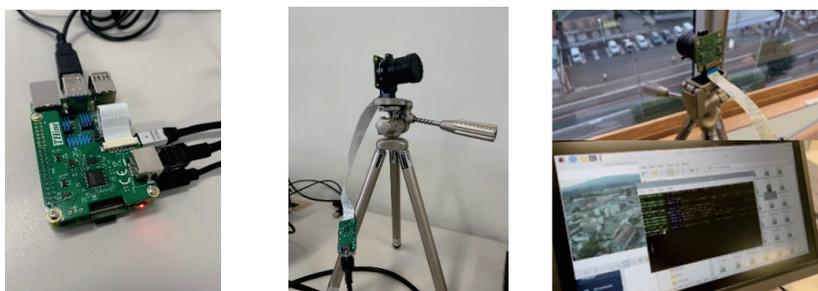


図 8 プロトタイプの概観

6. まとめ

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、提案手法のプロトタイプを完成させるとともに、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。

7. 謝辞

本研究の一部は、中部大学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 IDEAS202303、日本学術振興会科研 JP23K11120 の助成を受けたものです。

参考文献・データ

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. Giuseppina Andresini, Annalisa Appice, Donato Malerba: Autoencoder-based deep metric learning for network intrusion detection, INFORMATION SCIENCES, Volume 569, pp.706-727 (2021)
5. Ross Goroshin, Joan Bruna, Jonathan Tompson, David Eigen, Yann LeCun: Unsupervised Learning of Spatiotemporally Coherent Metrics, International Conference on Computer Vision, pp.4086-4093 (2015)
6. 松岡佑磨, 島田大樹, 彌富仁: 深層距離学習による教師データ生成, 第32回ファジィシステムシンポジウム 講演論文集, pp.91-96 (2016)
7. Haque Ishfaq, Assaf Hoogi, Daniel Rubin: TVAE: Triplet-Based Variational Autoencoder using Metric Learning, arXiv:1802.04403 [stat.ML] (2018)
8. 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)