|  |
| --- |
| 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 |

# はじめに [[1]](#footnote-1)\*【\*の文字書式「隠し文字」】

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．

# 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．

# 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, …, *in* } 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, …, *vn* }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, …, **t***m* ) を用いることもできる(図2)．ここで，各**t***x*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, …, *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，入力の画像特徴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

# 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．

## 実験環境

画像データセットして，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．

|  |  |
| --- | --- |
| クラス1  田畑 |  |
| クラス2  水源地帯 |  |
| クラス3  森林地帯 |  |
| クラス4  建物 |  |

図 4　航空写真データの例 (4分類)

## 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する．

M5：M1のAEと距離学習NN-2で構成する．M1のAEか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する．

再現率，適合率は下記の式で算出する．

再現率 = 正解数 / 正解総数 (1)

適合率 = 正解数 / 正解と判定した数 (2)

## 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 |

|  |  |
| --- | --- |
|  |  |
| M5の損失値変化 | M6の損失値変化 |

図 7　学習における損失値の比較

# 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).

1. \* 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 [↑](#footnote-ref-1)