F-006

教師なし学習を用いた書類記入領域自動抽出手法の提案
A method of extracting entry areas from application form images using unsupervised learning

片岡 えり †　　横地 洋 †
Eri Kataoka　　Hiroshi Yokochi

1　概要

我々は、金融機関や役所等が取り扱う申請書等の紙書類（以下、帳票）を効率的に処理するためのシステムを開発している。本システムは、帳票内の顧客が記入する領域（以下、記入領域）を予め定義しなければならないが、定義作業に時間がかかるという課題がある。本課題を解決すべく、学習済みの畳み込みニューラルネットワーク（以下、CNN）から帳票の特徴箇所を可視化し、記入領域を自動で抽出する手法を開発した [1]。本稿では、上記手法を拡張し、教師なしの CNN に対して特徴箇所の可視化を適用する記入領域自動抽出手法を提案する。教師なしの CNN は帳票の種類が追加されてもモデルの再構築が不要なモデルであり、従来手法と比較して短い学習時間で記入領域を抽出できる。

2　背景と課題

2.1　教師なし学習モデル

教師なし学習は、自己符号化器（オートエンコーダ）を模した構成で実現する。学習方法の概要を図 1 に示す。図 1 内の CNN①は畳み込み層とプーリング層から成る CNN で、モデル内に全結合層を持たず、画像を入力して特徴量を出力する。CNN②は特徴量（CNN①の出力値）を画像に戻す CNN である。教師なし学習は、学習によって入力画像と出力画像が近づくようなパラメータを獲得することで、モデルを構築する。なお、学習に用いる画像は分類対象の画像と同様の性質をもつ画像が望ましいが、分類する画像そのものを学習する必要はない。

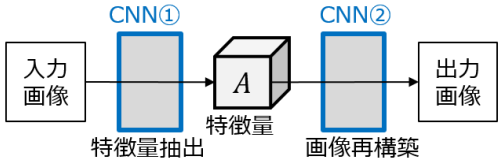


図 1. 学習方法

対象画像の分類には、学習済みの CNN①を用いる。CNN①は画像を入力すると特徴量が得られるモデルで、クラスの情報を持たないため分類はできない。そこで、クラスの情報をモデル外に持つ構成をとる。具体的には、分類対象の画像を複数枚、学習画像として CNN①に入力し、それぞれ特徴量を求める。求めた特徴量はベクトル化（平滑化）して分類クラス毎に平均を求める。これをクラス毎の「基準」として保持する。これにより、対象画像を CNN①に入力して特徴量を求め、平滑化し、クラスごとの基準と比較して類似度を求めることができるようになり、最も類似度が高いクラスを分類結果として導出できる。

2.2　Grad-CAM の適用と課題

Grad-CAM[2] は、学習済みの CNN モデルから得られる特徴量を可視化する手法である。この手法は、あるクラスの分類確率（類似度）に対する特徴量の勾配を重みとして、特徴量に付与した上で空間平均を出力することで、クラスの特性を反映した特徴量可視化を実現する。

Grad-CAM を 2.1 節で述べた教師なしの CNN（CNN①）に適用することで、モデルの再構築が不要となるため、学習時間の短縮が期待される。しかし、CNN①はモデル内に分類のための情報を保持しない構造であるため、書類の類似度を CNN①のみで導出することはできない。よって、CNN①に類似度と特徴量の勾配を使用する Grad-CAM をそのまま適用することはできない。

3　提案手法

2.2 節で述べた課題を解決するため、教師なし学習に Grad-CAM を適用する手法を提案する。提案手法の概要を図 2 に示す。本手法では、CNN①が出力する特徴量と「基準」を入力として、類似度のベクトルを出力とするモデルを新たに「類似度計算モデル」として定義する。類似度計算モデルの内部で特徴量から類似度への変換を完結させることで、類似度と特徴量の勾配を求めることができるようになる。よって、分類クラス毎の重みを特徴量に反映でき、従来の記入領域抽出手法 [1] と同様に Grad-CAM を適用して帳票のレイアウトを導出できる。さらに分類クラスが増えた場合でも、基準を作成するのみで記入領域を抽出できるため再学習が不要となり、学習時間を大幅に短縮できる。

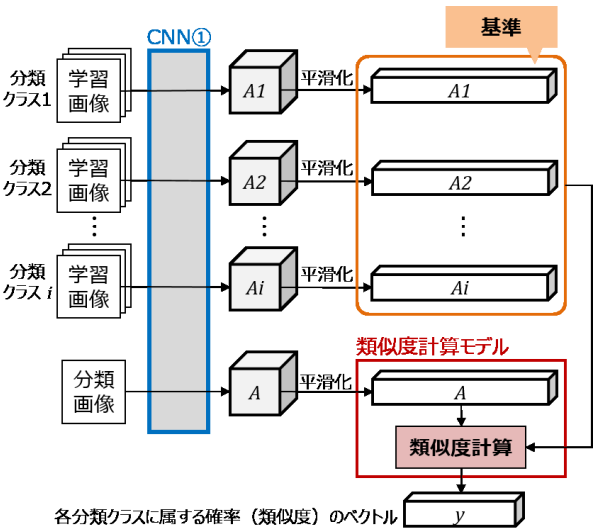


図 2. 提案手法概要

4　評価

4.1　評価方法

特徴量可視化結果および記入領域抽出精度について、従来手法と提案手法の比較評価を行った。本評価では、某金融機関の実際の A4 サイズの記入済み帳票 10 種類、各 50 枚の画像を用いた。またそれぞれの種類の帳票画像について、50 枚のうち 25 枚を基準の作成、残り 25 枚を抽出精度の評価に用いた。帳票画像の大きさは 2048×2896 ピクセル

† 三菱電機 (株) 情報技術総合研究所

とし、モデルの入力サイズに合わせてリサイズし、グレースケールに変換したものを用いた。学習モデルの構築、学習、および Grad-CAM の実装には TensorFlow[3] を用いた。従来手法と提案手法で使用した CNN モデルの詳細は図 3 および図 4 に示すとおりで、畳み込みブロックの最終層に対してそれぞれ Grad-CAM を適用して評価した。

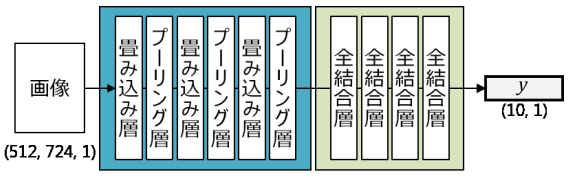


図 3. 従来手法の CNN モデル

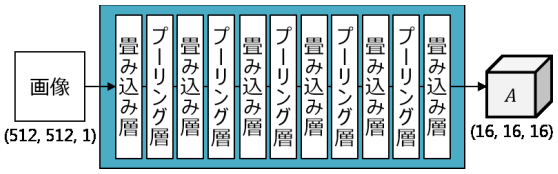
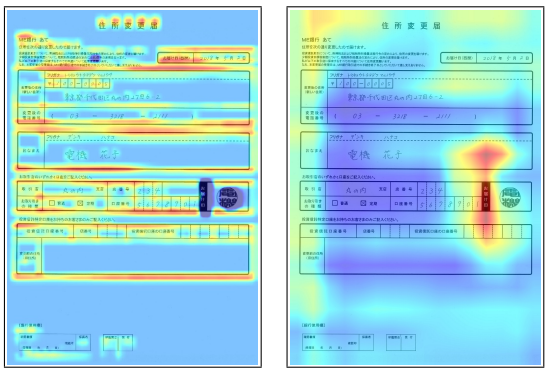


図 4. 提案手法の CNN モデル（CNN①）

4.2 特徴量可視化結果

従来手法と提案手法について、同じ帳票画像に対して Grad-CAM を適用し、特徴量を可視化した。結果は図 5 に示すとおりである。図 5 より、従来手法では記入領域の枠上に細かな特徴が強く出た。一方、提案手法では、帳票画像全体の特徴を捉えているものの、従来手法に比べて不鮮明な結果となることがわかった。



(a) 従来手法　　(b) 提案手法

図 5. 特徴量可視化例

要因として、1) モデルの違い、2) 解像度の違いが考えられる。まず 1) モデルの違いについて、提案手法は教師なし学習であるため、対象となる帳票の情報をモデル内に持たない構造である。そのため、従来手法に比べて当該帳票の特徴の細部を表現できておらず、不鮮明になったと考えられる。次に 2) 解像度の違いについて、本評価では、Grad-CAM を適用する最終層の出力形状が手法間で異なるため、可視化結果のサイズが異なる。従来手法では、64×90 のサイズで結果が得られるのに対し、提案手法は 16×16 のサイズで結果が出力される。よって、可視化結果を元画像のサイズ（2048×2896）に引き伸ばして重畳した際に、提案手法の方が情報量が少なくなり、より不鮮明な結果になったと考えられる。なお、解像度の違いは、Grad-CAM を適用する畳み込み層の形状を従来手法と同等にして解像度を上げ、モデルを再構築することで解決できると考えられる。

4.3 記入領域抽出精度

評価対象とした 10 種類の帳票画像に対して記入領域の抽出精度を求め、比較評価した。本評価では、1 つの正解領域に対して抽出領域が重ならない場合を抽出失敗、それ以外を抽出成功とみなし、ある帳票に含まれる正解領域のうち、抽出に成功した領域の割合を抽出精度として求めた。記入領域は領域を囲む矩形として抽出し、抽出数の上限は 40 個とし、上限数を越える場合は隣接する領域を結合した。評価結果は表 1 に示すとおりである。

表 1. 記入領域抽出結果

帳票	従来手法	提案手法
(1)	67%	78%
(2)	34%	72%
(3)	50%	98%
(4)	82%	88%
(5)	27%	25%
(6)	100%	100%
(7)	59%	90%
(8)	86%	76%
(9)	100%	93%
(10)	50%	37%

表 1 より、手法間で精度が大きく異なる場合があることがわかった。例えばレイアウトが単純な帳票 (7) の場合は提案手法の方が精度が高く、レイアウトが複雑な帳票 (8) の場合は従来手法の方が精度が高い。本結果は、提案手法の CNN モデルが複雑なレイアウトを捉えられず、記入領域の概要のみを可視化してしまうことが原因であると考えられる。また提案手法では、レイアウトが単純な場合、従来手法のように 1 つの正解記入領域から複数の候補を出してしまうことが少なく全体の精度が上がるが、レイアウトが複雑化するに応じて詳細な記入領域を分割できなくなり、精度が落ちてしまう。

5 まとめと今後の課題

本稿では、教師なしの CNN に Grad-CAM を適用する手法について提案した。これにより、帳票の種類が追加された場合であっても CNN モデルの再構築が不要となり、短時間で記入領域を抽出できる見込みを得た。

帳票の記入領域抽出精度について提案手法と従来手法を比較評価した結果、10 種類中 5 種類の帳票で精度が改善した。一方、帳票内のレイアウトが複雑化すると、従来手法より精度が低下する傾向にあることがわかった。以上の結果より、CNN の構造やモデルの学習を再考し、提案手法での記入領域抽出精度の向上を目指す。

参考文献

[1] 片岡えり, 松本光弘, 白木宏明. 書類記入領域自動抽出手法の提案. 第 17 回情報科学技術フォーラム (FIT), 2018.

[2] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-CAM: visual explanations from deep networks via gradient-based localization. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pp. 618–626, 2017.

[3] Martín Abadi, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.