

## Grad-CAM を用いた着目箇所制御方法と漫画作品の著者推定への応用 Grad-CAM based interest region control and its application to author estimation of manga works

竹内 悠太<sup>†</sup>      黒木 修隆<sup>†</sup>      沼 昌宏<sup>†</sup>  
Yuta Takeuchi   Nobutaka Kuroki   Masahiro Numa

### 1. はじめに

畳み込みニューラルネットワーク(CNN)を用いた画像認識において、画像中の一部の特徴に CNN の関心が集中し、過学習に陥る場合がある。そこで、本稿ではより高精度な画像分類器の実現を目的とし、CNN の着目箇所を Grad-CAM [1] を用いて制御する手法を提案する。

### 2. CNN の着目箇所の制御手法

#### 2.1 CNN による画像分類の利点と問題点

CNN をベースとする画像分類器は優れた性能を発揮する一方で、画像中の一部の特徴に関心が集中し、過学習に陥りやすい問題がある。一例として、CNN を用いて漫画作品の著者推定を行った結果を図 1 に示す。これら 3 つの作品はいずれも正しく著者が推定された。その際の CNN の着目箇所を Grad-CAM を用いて可視化すると、CNN はコマ割りの線などの幾何学的な情報を判断根拠としている。このように CNN は人が気付かないような特徴を見つけ出し、それを手掛かりとすることがある。この現象は新たな発見のきっかけとなる一方で、逆に人間の感覚との乖離から、思わぬ判断ミスを引き起こす可能性もある。これを防ぐためには、CNN の判断根拠が一部の特征に偏らないように、注目領域を分散させる必要がある。

#### 2.2 先行研究

CNN の注目領域を分散させる手法として、Zhong らに提案された Random Erasing [2] がある。Zhong らは訓練用画像にランダムにノイズをかけることにより、画像分類タスクにおいて精度向上が見込まれることを示した。しかし、Random Erasing によって生成されるノイズの発生箇所は乱数により与えられるため、必ずしも画像の重要な特徴をマスキングしているとは限らない。



図 1 CNN の着目箇所の可視化例

### 3. 提案手法

提案手法は、Random Erasing を発展させたものである。その概要を図 2 に示す。

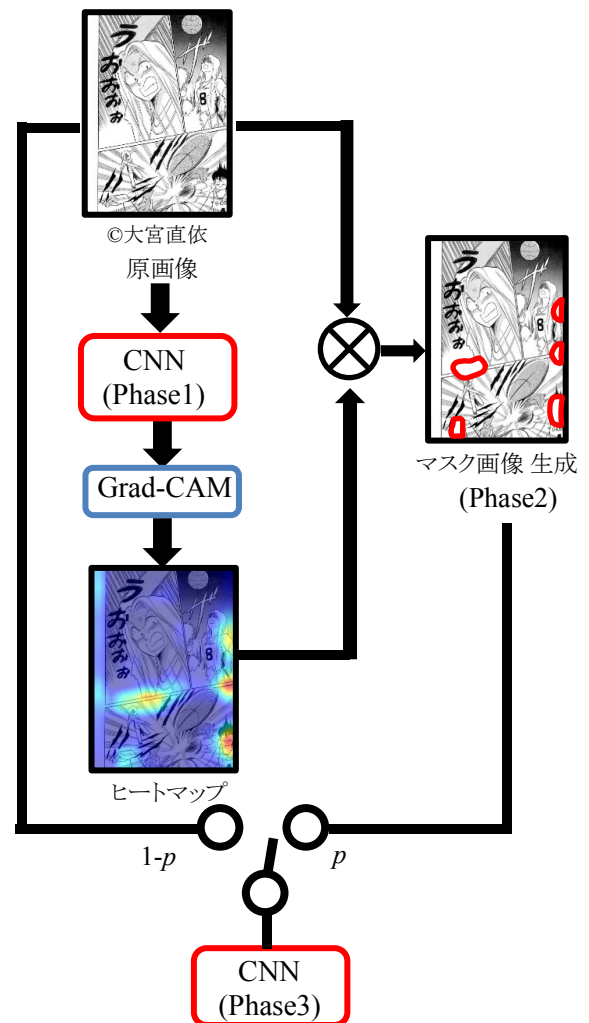


図 2 提案手法の概要

提案手法では、Grad-CAMによってCNNの着目箇所を探り、それらの領域をマスキングすることで、強制的に関心を分散させる。これによってCNNがより多角的に被写体の特徴を捉えるように誘導する。提案手法を用いたCNNの学習は、3段階に分けて行う。以下にその詳細を述べる。

<sup>†</sup> 神戸大学大学院工学研究科,  
Graduate School of Engineering, Kobe University

### Phase1: 原画像で学習

原画像を訓練用画像として、CNN の汎化性能の向上が見られなくなるまで学習を行う。具体的には、検証用データに対する認識率が向上しなくなるサンプリングポイントが得られた時点で、Phase1 を終了する。

### Phase2: マスク画像生成

Phase1 で得られた CNN の着目箇所を可視化するため、Grad-CAM を用いてヒートマップ画像を作成する。ヒートマップ中の赤い領域は、CNN がクラス分類の判断根拠にしている部分と考えられる。この領域を原画像から消去したものを、新たにマスク画像として生成する。

### Phase3: 原画像とマスク画像で学習

原画像とマスク画像を訓練用画像とし、CNN の再学習を行う。マスク画像の入力は、CNN に汎化性能の向上が見られなくなった段階で入力していく。この時点で、原画像による学習では過学習が発生するが、マスク画像を一定の確率  $p$  で入力することでアンサンブル効果が生まれ、分類精度の向上が期待できる。

Phase1-3 のように段階的に学習を行うことで、原画像の学習時とは異なる箇所を重点的に学習し、CNN の着目箇所が変化すると考えられる。

## 4. 評価実験

### 4.1 実験内容

本実験では提案手法を漫画作品の著者推定に応用した。提案手法の有用性を検証するため、以下の 3 つの手法を比較評価した。

- 手法 A : AlexNet
- 手法 B : AlexNet + Random Erasing
- 提案手法 : AlexNet + マスキング

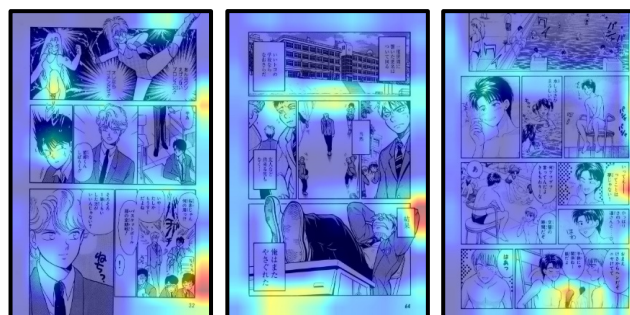
訓練用画像は 13,942 枚、検証用画像は 2,020 枚、評価用画像は 3,030 枚、学習回数は 60epoch、バッチサイズは 4 とする。マスク画像の入力確率  $p$  は Random Erasing と同様  $p = 0.5$  とした。実験に用いるデータセットには Manga109 [3] [4] を用いた。本実験ではその中から、巻数の異なる同一作品及び画像枚数の極端に少ない 8 作品を除いた、101 人の著者による 101 作品を用いた。また、画像サイズは 414×585pixel とした。

### 4.2 実験結果と考察

表 1 に各手法の認識率を示す。認識率は AlexNet のみを用いた手法 A より提案手法の方が 2.6pt 高くなった。また、Random Erasing を用いた手法 B より 1.6pt 高くなることを確認した。また、図 3 にヒートマップ画像の一例を示す。このように、手法 A ではコマの枠線付近に高い関心が表れているが、提案手法ではキャラクターの顔や体にも関心が強く表れている。すなわち、提案手法では CNN の着目箇所が広がり、漫画画像の持つ特徴を幅広く判定根拠に用いていると考えられる。以上より、提案するマスキング手法の有用性を確認できた。

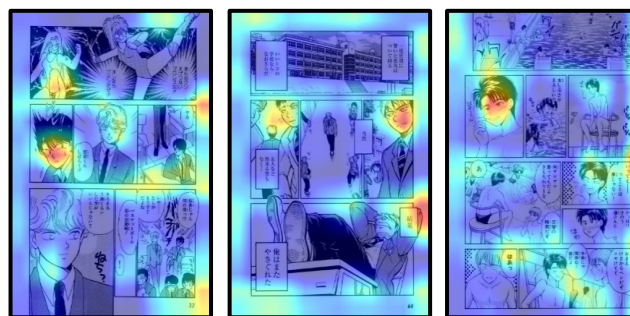
表 1 各手法の認識率

評価手法	認識率 [%]
手法 A : AlexNet	85.2
手法 B : AlexNet + Random Erasing	86.2
提案手法 : AlexNet + マスキング	87.8



©大宮直依                      ©竹山裕右                      ©河方かおる

(i) 手法 A の画像例



©大宮直依                      ©竹山裕右                      ©河方かおる

(ii) 提案手法の画像例

図 3 各手法のヒートマップ例

## 5. まとめ

本稿では、より高精度な分類機の実現を目的とし、CNN の着目箇所を Grad-CAM を用いて制御する手法を提案した。評価実験において本手法を漫画作品の著者推定に応用した所、着目箇所が広がり、また、それによって推定精度が向上することを確認した。今後は、漫画画像以外にも本手法を応用する予定である。

### 参考文献

- [1] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *The IEEE International Conference on Computer Vision (ICCV)*, pp. 618-626, 2017.
- [2] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, "Random Erasing Data Augmentation," arXiv:1708.04896, 2017.
- [3] T. Ogawa, A. Otsubo, R. Narita, Y. Matsui, T. Yamasaki, K. Aizawa, "Object Detection for Comics using Manga109 Annotations," arXiv, 2018.
- [4] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, K. Aizawa, "Sketch-based Manga Retrieval using Manga109 Dataset," *Multimedia Tools and Applications*, Springer, 2016.