広島工業大学紀要研究編
第54巻（2020）91-99
論　　文

CNNによる漫画作者の画像的特徴の抽出

上田　智之*・小池　正記*

（令和元年10月30日受付）

Image feature extraction of comic artist by CNN

Tomoyuki UEDA and Masaki KOIKE

（Received Oct. 30, 2019）

Abstract

This paper reports the extraction of image features of comic artist by convolutional neural network. In recent years, illegal uploading of comics has become a social issue. The comic artist wants to find his own illegally uploaded comic. A convolutional neural network extract image features representing a class from an image. Comics has the artist's style. The artist's style can be used as an image features that represents a class in the comic image classification problem. Therefore, convolutional neural networks can solve the artist classification problem for comic images. If the classification problem can be solved, it can be applied to image retrieval. Visualize the area of interest when convolutional neural networks predict artists. There is an element that can predict the artist in the attention area. The experimental results suggest that the comic characters face is important for convolutional neural network to predict the artist from the comic images.

Key Words: convolutional neural network, deep learning, image recognition, comic computing

1．はじめに

2018年、海賊版漫画ビューアーサイト「漫画村」が社会問題となった。漫画村は漫画などの書籍数万冊が登録不要且つ完全無料で読めるという漫画ビューアーサイトである。しかし、漫画村にアップロードされている書籍は著作権者の許諾を得ずに無断で違法コピーされた海賊版であり、漫画村の存在は出版社や作者に大きな被害を与えた。これほど巨大な海賊版サイトですら、インターネット上に存在する違法にアップロードされた漫画作品全体からみればごく一部に過ぎない。出版社や作者など作品の著作権者たちが違法にアップロードされた自分たちの漫画作品をインターネット上から探し出し、削除申請するなどの対応も活発になった。しかし、削除される以上に多くの漫画作品が新たにアップロードされるため、削除が追いついていないのが現状である。もし、インターネット上という膨大なデータの集合から、著作権者が自身の漫画作品のみを自動で探し出すことができれば、削除の手助けとなる。

そこで著作権者が自身の漫画作品を探し出す方法として畳み込みニューラルネットワーク（CNN）による画像検索技術を利用することを考えた。CNNは視覚野の特徴抽出の仕組みをモデル化した画像認識を得意とするニューラルネットワークで、特に画像を入力とした分類問題ではよく利用されている。CNNでは入力画像から人工知能アルゴリズムを使って特徴量を抽出する。この特徴をもとに入力画像の分類を行うにはクラスを代表とする特徴を掴むことが重要になるため、学習データはクラスごとに異なる画像的特長を持っている必要がある。絵には作者の特色や傾

* 広島工業大学大学院工学系研究科電気電子工学専攻

向があらわれており、これは画風と呼ばれる。漫画も絵の集まりで構成された作品であるため、同一の作者が描いた漫画であるならば画風という作者ごとの画像的特徴が存在するであろう。作者というクラスごとに共通した画像的特徴が存在するのであれば、CNN による分類問題として解くことが可能である。これを応用すれば、未知の画像データ群の中から特定の漫画作者に関係した画像のみを見つけ出すことが可能になる。

CNN を利用するためには画像の何がクラスを代表する特徴として利用できるのかを分析することが重要になる。本研究では、漫画内のどういった部分が漫画作品における作者ごとの画像的特長として利用できるのかを検討する。

2．判別実験

2-1　データセット

本研究の学習データには、学術利用可能な漫画データセットとして公開されている Manga109［5］［6］を用いた。幅広いジャンルを網羅するため、異なる雑誌で連載していた10人の漫画作者の作品を対象とした。

表 1　各作品のデータ

作者	タイトル	巻数	ページ数	連載雑誌
石岡 ショウエイ	ベルモンド Le VisiteuR	1	198	週刊少年ジャンプ
記伊 孝	犯罪交渉人 峰岸英太郎	1	200	月刊ヤングマガジン
猪熊 しのぶ	SALAD DAYS	1	182	週刊少年サンデー
猪原 大介	学園ノイズ	1	197	月刊ZERO-SUM
うえだ 美貴	爆烈！かんふー娘	1	193	ちゃお/ちゃおDX
霧賀 ユキ	てんしのはねと アクマのシッポ	1	181	月刊Gファンタジー
水上 悟志	サイコスタッフ	1	182	まんがタイム きららフォワード
能田 達規	がらくた屋まん太	2	205	月刊アスキーコミック
桜野 みねね	ひなぎく見参！ 一本桜花町編	1	178	コミックブレイド MASAMUNE
出口 竜正	ドールガン	2	191	週刊少年チャンピオン

Manga109の画像データは見開き 2 ページ分を 1 枚の画像としている（図 1 ）。クラスごとのデータ数を増やすため、1 ページずつに分割した。また、入力のサイズを正方形とするため、画素数を縦方向の1170pixel に合わせてパディングを行った（図 2 ）。

© 能田達規

図 1　Manga109の元画像（1654×1170pixel）

© 能田達規

図 2　加工後の画像（1170×1170pixel）

加工したデータをネットワークの学習に用いる訓練用データ、過学習を監視するための検証用データ、最終的なモデルの性能を評価するための評価用データに分割する。本研究ではクラスごとに評価用データ50枚、検証用データ50枚を確保し、残りを訓練用データとした。

2-2　CNN モデルの構築

訓練に用いる画像データの枚数が少ない場合、別の学習データによる学習が済んでいるモデルを利用することが有効である。CNN では入力に近い浅い層から出力層へと深くなるにつれ、データセットに特化した複雑な形状に反応するフィルタが出来上がることが知られている［9］。このため、高い精度を持った分類モデルの前半の層を利用し、後半の層のみを目的に応じて再学習させることで、少ない枚数の画像データからでも目的に特化したモデルを作り出すことができる。本研究では、ImageNet［11］で事前学習した VGG16［10］を用いて、目的に特化したモデルの再学習を行った。VGG16は 5 つの畳み込みブロックと 1 つの全結合ブロックで構成される16層の CNN である。VGG16の 4 番目の畳み込みブロックまでを ImageNet で事前学習された重みで固定し、それ以降の重みを再学習する。VGG16は224×224の入力画像を畳み込むネットワークである。しかし1170×1170pixel の画像を224×224pixel まで圧縮すると小さなコマや文字が潰れてしまうため、今回は入力画像のサイズを585×585pixel として畳み込みブロックを 1 つ追加した。

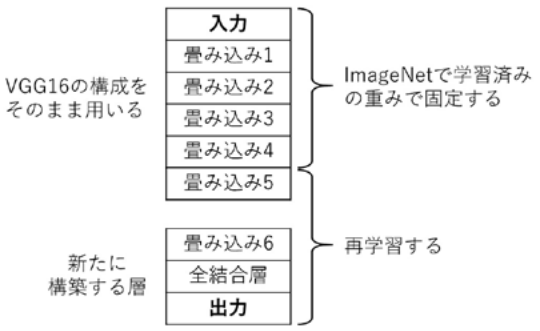


図 3　CNN の構成

2-3　学習結果

構築した CNN で訓練用データ計877枚、検証用データ計500枚を用いて学習回数300回の学習を行った結果が図４、５である。図４は学習回数ごとの訓練用データと検証用データに対するモデルの正解率、図５は学習回数ごとの訓練用データと検証用データに対するモデルの損失を示している。

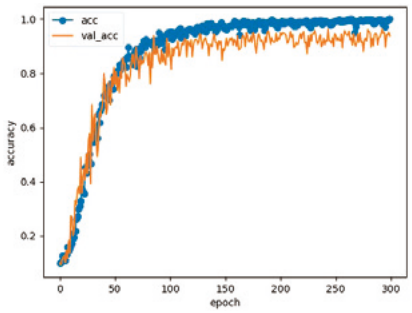


図４　正解率の推移

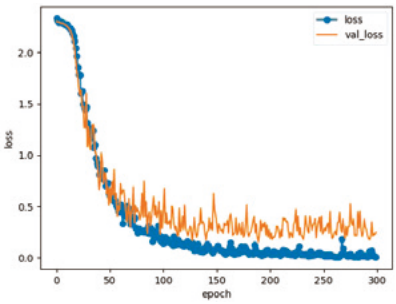


図５　損失の推移

学習回数とともに訓練データ、検証データに対する正解率が上昇し（図４）、損失が減少している（図５）ことから学習は正常に行われている。ただし、学習回数300回の中で検証データに対する損失の値が最も小さかったのは学習回数190回目である。それ以降は訓練データに対する損失である loss のグラフは減少し続けているが、検証データに対する val_loss のグラフは減少していない。過学習の傾向が確認できたことからこのモデルの学習は設定した学習回数300回を終える前に収束している。この場合、学習回数300回の中で最も損失の値が小さかった学習回数190回目のネットワークの重みが最もデータセットの中から画像的特長を抽出している。

2-4　評価結果

学習回数190回目のネットワークの重みを読み込んだ CNN モデルを用いて学習に使用していない評価用画像50枚の作者予測を行った。モデル全体の評価結果が表２、クラスごとの評価結果が表３である。

表２　モデル全体の評価結果

再現率	0.9520
精度	0.9541
F 値	0.9517

モデル全体としては F 値が95% を超え、最も予測性能の低いクラスでも50枚中40枚以上の画像で作者の予測に成功している。評価用画像が同数の10クラス分類問題の場合、ベースラインの正答率は10% となる。モデルの判別能力がベースラインの正答率を上回っていることから、CNN は判断基準を持たずに作者を予測しているのではなく、漫画からクラスごとの画像的特長を学習し、それを判断基準として作者を予測していることが推察できる。このことから漫画には画像から作者を予測できるクラスごとの画像的特長が存在していることが示唆された。

3．可視化実験

3-1　CNN の問題

漫画画像から作者を判別できたという実験結果から、CNN は漫画画像の中にある作者ごとの画像的特徴を学習していることが示唆された。しかし、それが漫画のどういった要素に現れているのかは不明なままである。そこで Grad-CAM++［８］というヒートマップにより視覚的に判断根拠を確認する方法を利用し、解析を行った。CNN が漫画画像から作者を判別する際、どの個所に注目して判別を行っているのかを可視化し傾向を見つけることができれば、そこが画像を見て作者を予測できる要素であり、画像から作者を予測できるクラスごとの画像的特長が強くあらわれている要素であると推察できる。

表３　混同行列

		予測結果										
		うえだ美貴（爆烈!かんふー娘）	記伊孝（犯罪交渉人 峰岸英太郎）	桜野みねね（ひなぎく見参！一本桜花町編）	出口竜正（ドールガン）	水上悟志（サイコスタッフ）	石岡ショウエイ（ベルモンド Le VisiteuR）	猪熊しのぶ（サラダデイズ）	猪原大介（学園ノイズ）	能田達規（がらくた屋まん太）	霧賀ユキ（てんしのはねとアクマのシッポ）	再現率 Recall
入力	うえだ美貴（爆烈!かんふー娘）	50	0	0	0	0	0	0	0	0	0	1.0000
	記伊孝（犯罪交渉人 峰岸英太郎）	0	46	0	1	0	0	2	0	1	0	0.9200
	桜野みねね（ひなぎく見参！一本桜花町編）	2	0	47	0	0	0	0	0	0	1	0.9400
	出口竜正（ドールガン）	0	0	0	46	0	0	2	0	2	0	0.9200
	水上悟志（サイコスタッフ）	0	1	1	0	42	2	0	3	0	1	0.8400
	石岡ショウエイ（ベルモンド Le VisiteuR）	0	0	0	0	0	49	1	0	0	0	0.9800
	猪熊しのぶ（サラダデイズ）	0	0	0	0	0	0	49	0	1	0	0.9800
	猪原大介（学園ノイズ）	0	0	0	0	0	0	0	49	1	0	0.9800
	能田達規（がらくた屋まん太）	0	0	0	0	0	0	0	1	49	0	0.9800
	霧賀ユキ（てんしのはねとアクマのシッポ）	1	0	0	0	0	0	0	0	0	49	0.9800
	精度 Precision	0.9434	0.9787	0.9792	0.9787	1.0000	0.9608	0.9074	0.9245	0.9074	0.9608	
	F値 F-Measures	0.9709	0.9485	0.9592	0.9485	0.9130	0.9703	0.9423	0.9515	0.9423	0.9703	

3-2　漫画の構成要素

本研究では漫画を構成する要素を４つに分類した（図６）。１つ目が漫画の登場人物であるキャラクターという絵、２つ目がキャラクターの台詞や地の文などを表現するための写植されたデジタル文字、３つ目が擬音やキャラクターの心の声などを表現するために作者が直接漫画内に描き込む文字、４つ目が背景や道具などキャラクター以外の絵である。評価用画像について Grad-CAM++ による注目領域の可視化を行い、作者を予測するために重要な画像的特長が強く現れる要素がこの４つのどれであるか検討する。

© 能田達規

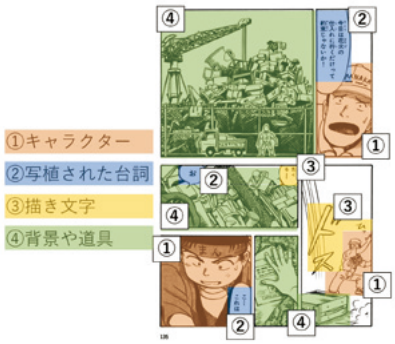


図６　漫画の構成要素

3-3　Grad-CAM++

Class Activation Mapping（CAM）はネットワークがどこに着目したかを可視化する手法である。Grad-CAM では、特徴マップのある位置に勾配の変化を加え、そのときに生じる出力の変化の大きさからクラス判定にとって重要な位置を特定できる。クラス c の k 番目のフィルタに関する重み係数 w は

$$w_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial Y^c}{\partial A_{ij}^k}$$

重み係数が大きいほど特徴マップ k がクラス c にとって重要である。重み係数 w から加重平均を計算し、ReLU の出力値を求めると画像中のピクセル位置（i , j）がその画像がクラス C であると判断されたときの重要度となる。

$$L_{ij}^c = ReLU\left(\sum_k w_k^c A^k\right)$$

Grad-CAM++ では、Grad-CAM の平均を 1 /Z をかけることで求めていたピクセル平均を 2 階微分、３階微分で重みづけして求める。これにより、Grad-CAM では確認できなかった小さな特徴量まで細かく可視化できる。Grad-CAM++ の重み係数 w は

$$w_k^c = \sum_i\sum_j \frac{\frac{\partial^2 Y^c}{\left(\partial A_{ij}^k\right)^2}}{2\frac{\partial^2 Y^c}{\left(\partial A_{ij}^k\right)^2} + \sum_a\sum_b A_{ab}^k\left\{\frac{\partial^3 Y^c}{\left(\partial A_{ij}^k\right)^3}\right\}} \cdot relu\left(\frac{\partial Y^c}{\partial A_{ij}^k}\right)$$

活性化関数が ReLU であるならヒートマップへの出力値 L は Grad-CAM と同様である。

© 桜野みねね

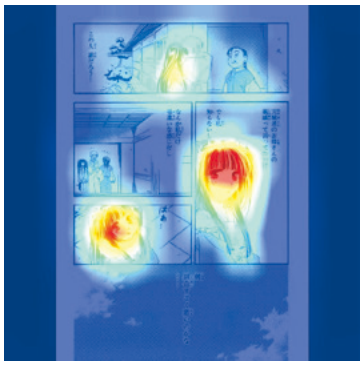
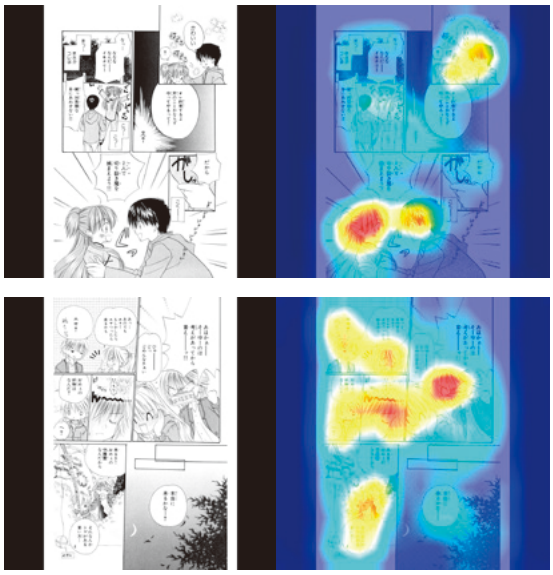


図７　注目領域の可視化例

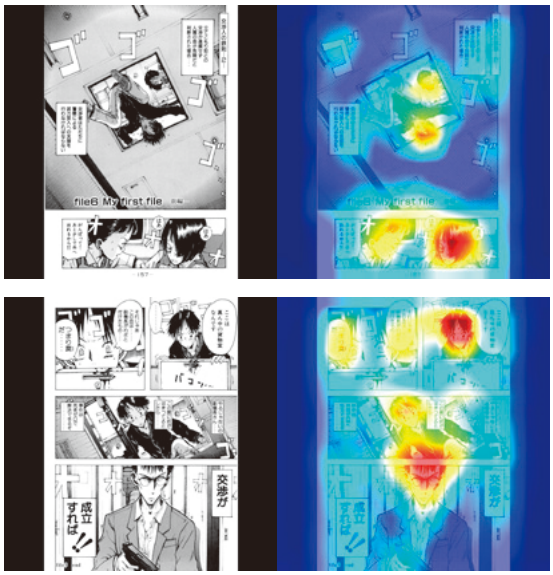
3-4　評価用画像の可視化と考察

Grad-CAM++ を用いて、作成した CNN モデルの評価用画像に対する注目領域を可視化した。作者の予測に成功した評価用画像の注目領域を可視化したところ、キャラクターに強く注目していることが確認できた（図８）。

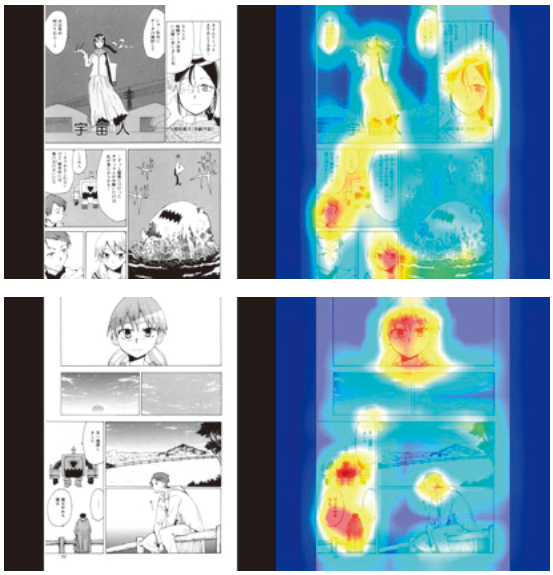
© うえだ美貴



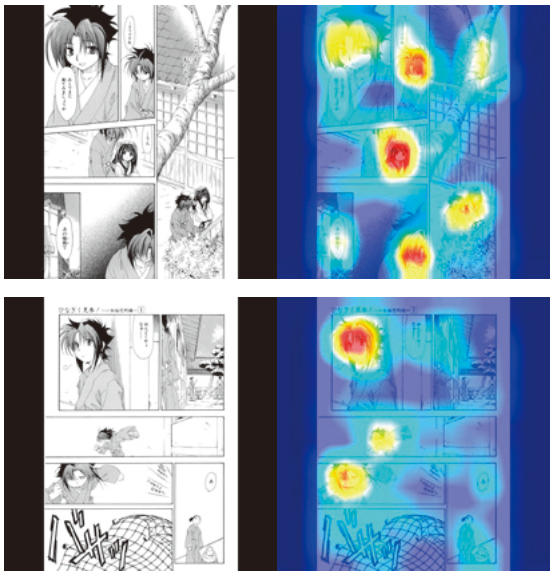
© 記伊孝



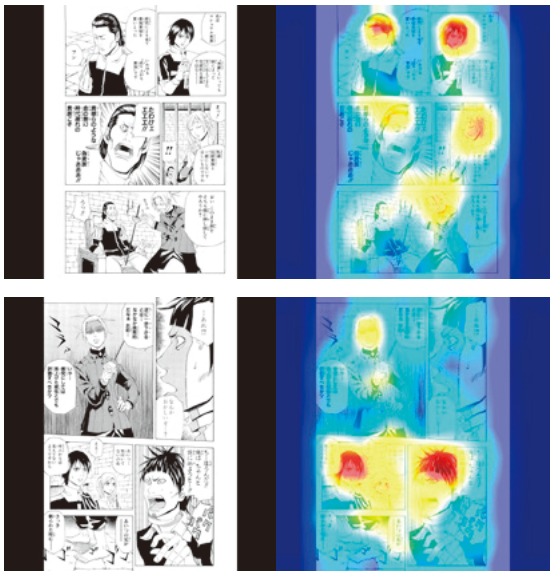
© 水上悟志



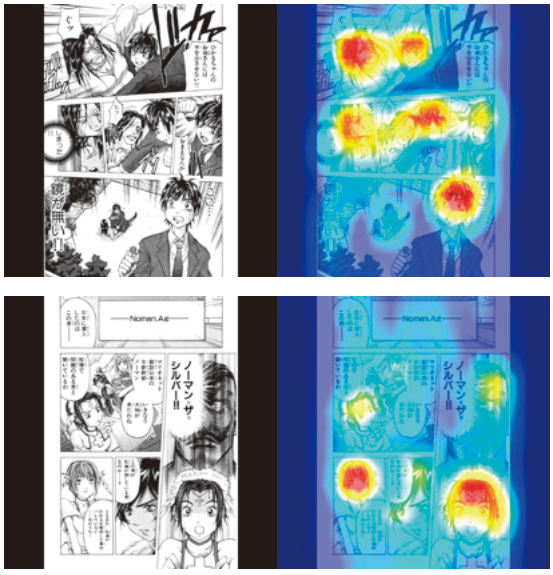
© 桜野みねね



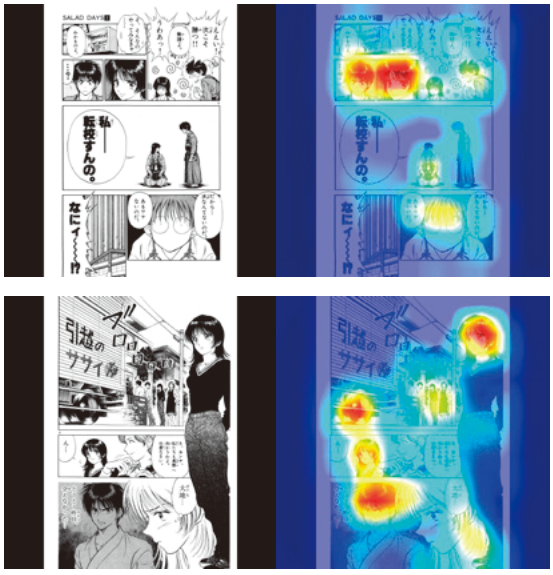
© 石岡ショウエイ



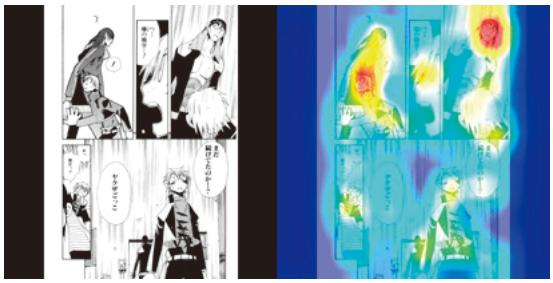
© 出口竜正

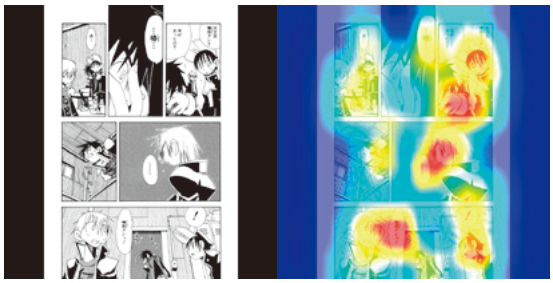


© 猪熊しのぶ

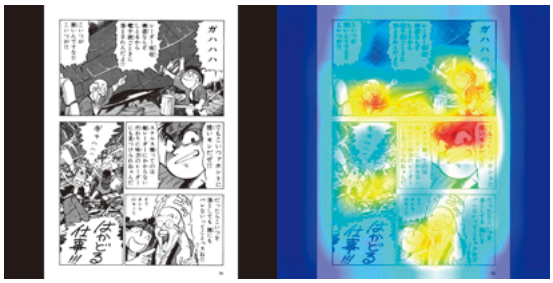


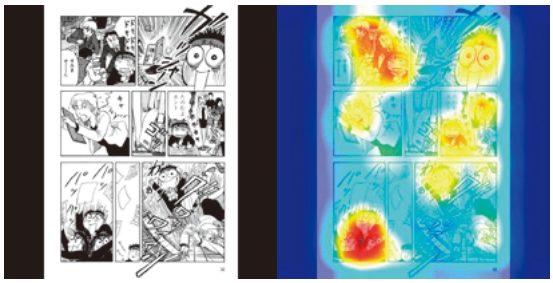
© 猪原大介



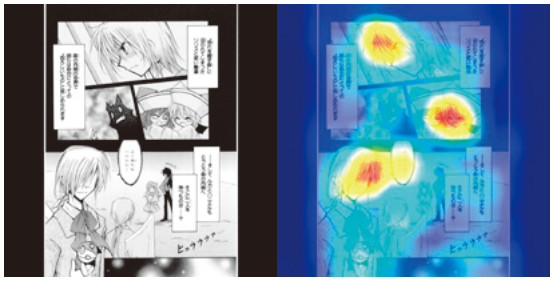


© 能田達規





© 霧賀ユキ



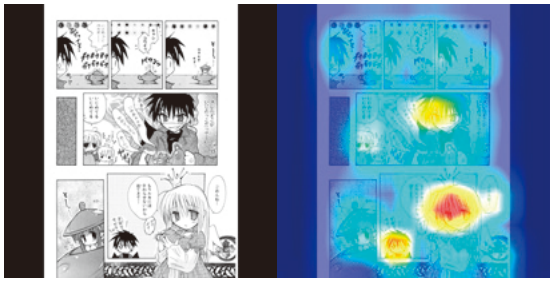


図 8　注目領域の可視化（右：元画像　左：注目領域）

また、CNN がキャラクターに注目している際、キャラクターの顔部分に特に強く注目する傾向も確認できた。画像から作者を判別する際、特に強く顔に注目しているということはその部分に作者を予測できる画像的特長が強く現れていると CNN が判断したことが示唆された。

キャラクターの次に CNN が注目しているのが漫画内の文字であった。漫画内の文字には写植されたデジタル文字と作者が直接描き込む描き文字の 2 種類ある。評価用画像の可視化結果よりデジタル文字と描き文字が混在している場合、デジタル文字が優先して注目される傾向にあった。このことから CNN は漫画画像から作者を判別するうえでデジタル文字が重要な要素で、描き文字は重要な要素ではないと判断したことが示唆された。しかし、デジタル文字はフォントによる違いがあるだけで、文字を書いた人間の癖があらわれるものではない（図10）。漫画内のデジタル文字と漫画作者を結び付ける特徴が存在するかという疑問が残った。

図10　各漫画内のデジタル文字の一部

注目領域はキャラクターと台詞などのデジタル文字に偏っており、CNN は画像から作者を予測する際に背景などキャラクター以外の絵を重視していない。画像内に自動車や建物などキャラクター以外の絵という要素が多く存在している場合もそれを避けるようにキャラクターや台詞に注目していた（図11）。

© 出口竜正

© 能田達規

© 桜野みねね

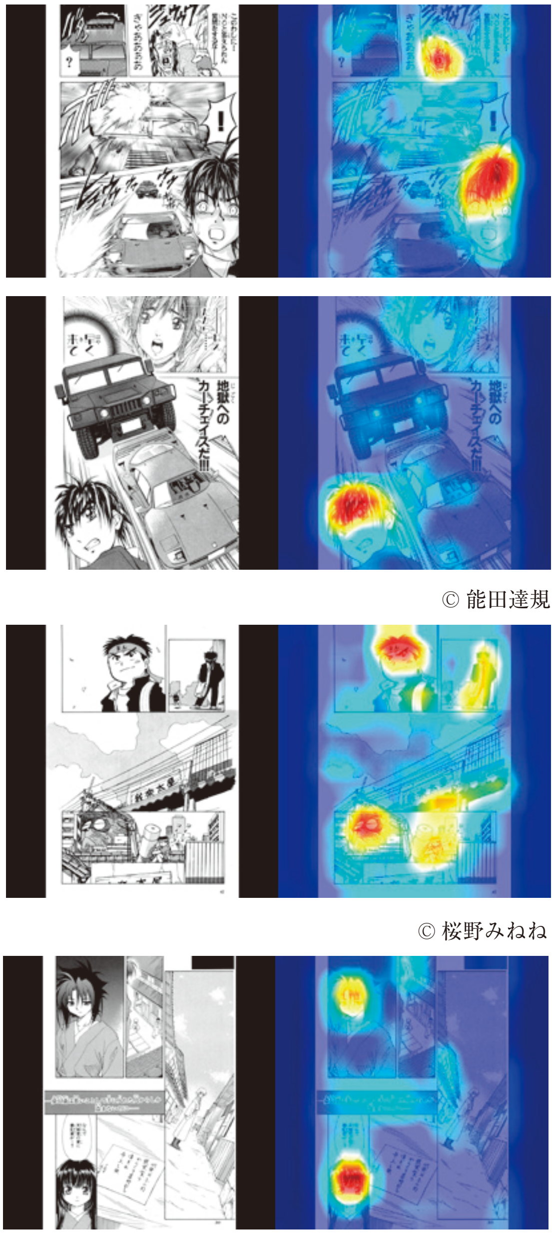


図11　注目領域の可視化（右：元画像　左：注目領域）

作者の画像的特徴を学習した CNN が背景や道具などキャラクター以外の絵に注目しないということは、そこに作者を予測できる画像的特徴が存在しないと判断したことが示唆された。

注目領域の可視化結果より、作者を予測できる画像的特徴が最も強い要素はキャラクターであり、その中でもキャラクターの顔部分は特に強く作者を予測できる画像的特徴があらわれていること、キャラクターの次に作者を予測できる画像的特徴が強くあらわれている要素はデジタル文字であることが推察された。描き文字と背景や道具などキャラクター以外の絵については CNN が注目しなかったことから作者を予測できる画像的特徴があらわれにくい要素であると推察された。

4．漫画作者の画像的特徴とデジタル文字

4-1　デジタル文字による作者判別

作者の判別に成功した評価用画像の注目領域を可視化した結果、幾つかの評価用画像において CNN はデジタル文字に注目していることが確認できた。この結果から、デジタル文字にも作者を予測できる画像的特長が存在することが推測された。しかしデジタル文字はフォントごとに定まった形状をしており、そこに作者固有の画像的特長があることは考えにくい。そこで各漫画からデジタル文字のみを抜き出し、各作者の使用したデジタル文字について学習を行った判別モデルを作成し、漫画内のデジタル文字から作者を予測することが可能か実験を行った。

4-2　判別実験

表１の10人の漫画作者の作品から台詞などの写植されたデジタル文字のみを抜き出す（図12）。

© 能田達規

図12　デジタル文字の抽出

作品ごとにデジタル文字を200個程度集め、検証用画像50枚、評価用画像50枚、残りを訓練用画像として使用し、10作品をデジタル文字から判別する CNN モデルを作成するための学習を行った。図13は学習回数ごとの訓練用データと検証用データに対するモデルの正解率、図14は学習回数ごとの訓練用データと検証用データに対するモデルの損失を示している。検証データに対する正解率は40% 程度で収束した（図13）。学習に使用しなかった CNN にとって未知のデータである500枚のデジタル文字画像から作者予測を行なったモデル全体の評価結果を表４に示し、クラスごとのデジタル文字に対する作者予測結果の混同行列を表５に示す。漫画画像から作者予測を行っていたときと比較すると予測精度は大きく低下した。CNN がデジタル文字を見て作者を予測できないということは、デジタル文字には作者を予測できる特徴がないと推察された。

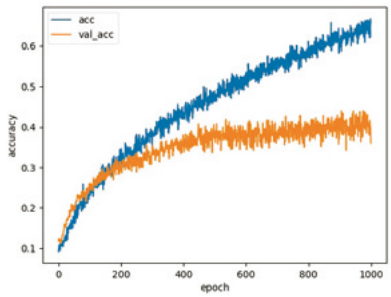


図13　正解率の推移

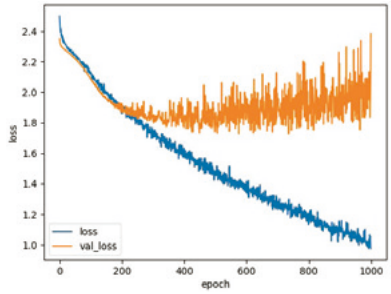


図14　損失の推移

表 4　モデル全体の評価結果

再現率	0.3760
精度	0.3901
F 値	0.3738

5．まとめ

本研究では、漫画の構成要素を 4 つに分類し、どの要素が漫画作品における作者ごとの画像的特長として利用できるのかを検討した。実験結果より以下の結論が推察された。

(1)　漫画画像から作者を予測するうえで最も重要なのはキャラクターの顔である。

(2)　背景などキャラクター以外の絵と描き文字はキャラクターと比較すると画像から作者を予測する際に重要な要素ではない。

(3)　デジタル文字には作者を予測できる画像的特長は存在せず、CNN が誤った注目を起こす原因となるため、漫画画像から作者を予測する学習を行ううえで不要な情報である。

文　献

[1] 上田智之、小池正記、“深層学習による漫画作品の判別”電子情報通信学会総合大会、D-12-6（2019）.

[2] 上田智之、小池正記、“CNN による漫画作品の特徴抽出”第 9 回電子デバイス・回路・照明・システム関連教育・研究ワークショップ（2019）.

[3] 寺前裕司、Interface44巻12号、CQ 出版社（2018）.

[4] 山下隆義、イラストで学ぶディープラーニング改訂第 2 版、講談社（2018）.

[5] Toru Ogawa, Atsushi Otsubo, Rei Narita, Yusuke Matsui, Toshihiko Yamasaki, Kiyoharu Aizawa, Object Detection for Comics using Manga109 Annotations, arXiv, (2018).

[6] 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).

[7] Ramprasath R.Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-basedLocalization.” IEEE International Conference on Computer Vision (ICCV) (2017).

[8] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, Vineeth N Balasubramanian, “Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks” IEEE Winter Conference on Applications of Computer Vision (WACV) IEEE (2018).

[9] M. D. Zeiler and R. Fergus. “Visualizing and Understanding Convolutional Networks.” European Conference on Computer Vision (ECCV) (2014).

[10] K. Simonyan and A. Zisserman. “Very Deep

表 5　混同行列

		予測結果										
		うえだ美貴（爆烈!かんふー娘）	記伊孝（犯罪交渉人 峰岸英太郎）	桜野みねね（ひなぎく見参!一本桜花町編）	出口竜正（ドールガン）	水上悟志（サイコスタッフ）	石岡ショウエイ（ベルモンド Le VisiteuR）	猪熊しのぶ（サラダデイズ）	猪原大介（学園ノイズ）	能田達規（がらくた屋まん太）	霧賀ユキ（てんしのはねとアクマのシッポ）	再現率 Recall
入力	うえだ美貴（爆烈!かんふー娘）	15	0	0	9	4	11	4	1	2	4	0.3000
	記伊孝（犯罪交渉人 峰岸英太郎）	1	34	0	4	4	0	1	1	1	4	0.6800
	桜野みねね（ひなぎく見参!一本桜花町編）	7	1	6	7	2	7	1	2	10	7	0.1200
	出口竜正（ドールガン）	1	2	2	23	4	4	6	2	4	2	0.4600
	水上悟志（サイコスタッフ）	4	1	4	6	19	2	3	1	8	2	0.3800
	石岡ショウエイ（ベルモンド Le VisiteuR）	3	0	2	8	4	20	2	0	3	8	0.4000
	猪熊しのぶ（サラダデイズ）	4	1	0	5	2	1	24	4	8	1	0.4800
	猪原大介（学園ノイズ）	3	3	3	6	9	3	2	15	3	3	0.3000
	能田達規（がらくた屋まん太）	6	0	1	16	0	3	9	1	11	3	0.2200
	霧賀ユキ（てんしのはねとアクマのシッポ）	3	0	2	2	6	6	2	5	3	21	0.4200
	精度 Precision	0.3191	0.8095	0.3000	0.2674	0.3519	0.3509	0.4444	0.4688	0.2075	0.3818	
	F値 F-Measures	0.3093	0.7391	0.1714	0.3382	0.3654	0.3738	0.4615	0.3659	0.2136	0.4000	

Convolutional Networks for Large-Scale Image Recoginition." International Conference on Learning Representations (ICLR) (2015).
[11] http://www.image-net.org/