

4 コマ漫画ストーリーデータセットにおける コマ画像の分散表現に基づく解析

Analysis based on Distributed Representations of Koma Images
in Four-scene Comics Story Dataset

寺内 光^{*1} 森 直樹^{*1} 上野 未貴^{*2}
Akira Terauchi Naoki Mori Miki Ueno

^{*1}大阪府立大学 ^{*2}豊橋技術科学大学
Osaka Prefecture University Toyohashi University of Technology

Understanding the creation of human by artificial intelligence (AI) is increasing; however, those are still known as one of the most difficult tasks. In this research, we are challenging for the understanding of four-scene comics by AI. To aim at this challenge, we use a novel dataset what is called “Four-scene Comics Story Dataset”, which is the first dataset made by researchers and cartoonists to develop AI creations. We focused on illustration touches of comics which is determined by cartoonists. First, we applied autoencoder (AE) models to this dataset to get distributed representations, then applied classifiers to that and predict a touch. The prediction offers an indirect measure of the distributed representations. The effectiveness of the proposed method is confirmed by computer simulations taking data of various pattern of removing parts in koma images of the four-scene comics story dataset structure as an example.

1. はじめに

近年、深層学習をはじめとする機械学習技術の発展を背景に、人工知能による小説やイラストの自動生成など計算機による創作物理解が注目されている。しかしながら、創作は高次の知的活動であるため、現段階では表層的な模倣以上の成果は得られていない。

近年、コミック工学を中心に、人の創作物の中でも特に漫画を対象とする分野が発展している。同分野では従来研究として言語情報を用いて漫画のコマの順序を識別する手法や、画像情報を用いて登場人物の位置を推定する手法などが報告されている。しかしながら、データセットとして同一のストーリーを描いた物語は存在しないため、より創作者に近い感性に関わる表現方法やタッチの識別については十分な研究がなされていなかった。そこで本研究では、上野によって作られたデータセット [上野 18] を用いることでこの問題を解消し、より創作者に近い視点に基づく計算機による漫画の意味理解を実現するために、画像処理の分野から作者ごとのタッチ識別に焦点を当てる。

画像のタッチ識別をする単純な方法に畳み込みニューラルネットワークを用いた識別が考えられる。しかしながら、本研究では特定のタスクに依らない汎用性をもつ画像の分散表現を得ることを目的とし、その分散表現をタッチ識別タスクの識別率により評価する。画像のもつ特徴を分散表現として得るための手法は Illustration2Vec[Saito 15] や畳み込み自己符号化器などが提案されている。本研究ではこれまでにない新たな実験として、コマ画像から顔のパーツを抜いた画像に対するタッチ識別について検討した。これは従来のデータセットでは実現できない内容であるため、本研究においては創作と人工知能に関する新たな知見を得ることを目的とする。

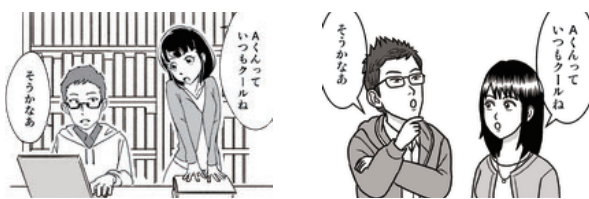
2. 研究用漫画データ

4 コマ漫画を対象としたデータセットとしては Manga109[Matsui 17][Ogawa 18] が知られているが、本研究では上野によって作られたデータセット (以下、4 コマ漫画ストーリーデータセットとする) を用いる。このデータセットは人工知能と創作物に関する研究発展のために研究者が一から開発に関わった世界初の研究用データセットであり、いくつかの特徴がある。まず、Manga109 のように市販の漫画を用いて作成されたデータセットとは異なり、4 コマ漫画ストーリーデータセットは本データセットのために幾人かの漫画家によって描き下ろされた 4 コマ漫画から構成されている。市販の漫画を用いた場合、著作権などの問題に加え、計算機上で扱うための著者の感性に関する情報が少なく、漫画の意味理解を目的とした研究には適さないという問題がある。例えば漫画に登場するキャラクターの感情は明示されていないため、読者によるアノテートによってラベルを付与する必要があるが、アノテートされたラベルが漫画家の意図とは異なる可能性を否定できない。また、オリジナリティの観点から同一プロットを複数の漫画家が描くことは稀有なため、漫画における表現の多様性に関する研究は困難である。4 コマ漫画ストーリーデータセットは上述の問題点を解決するために作られたデータセットであり、人工知能研究において大きな利点がある。

4 コマ漫画ストーリーデータセットは各コマがレイヤの重なりによって作成されており、レイヤ例としては目レイヤ、鼻レイヤ、口レイヤ、輪郭レイヤ、身体レイヤなどが存在する。これらのレイヤはそれぞれ別々に取り出すことができるように設計されている。これは必要な情報が容易に取り出せることを意味し、漫画の画像解析のために用いることができる情報が大量にあるという点で重要である。

また、上野は異なる作者によって描かれた 4 コマ漫画を、そのタッチを基にギャグタッチ、少女漫画タッチ、少年漫画タッチ、青年漫画タッチ、萌えタッチの 5 タッチに分類した。本研究では計算機による少年漫画タッチ、青年漫画タッチ、萌えタッチの 3 タッチを識別することを目的とする。図 1 に 3 タッチのコマ画像の例を示す。

連絡先: 寺内 光, 大阪府立大学 工学研究科, 大阪府堺市中区
学園町 1-1, terauchi@ss.cs.osakafu-u.ac.jp



(a) 少年漫画タッチ

(b) 青年漫画タッチ



(c) 萌えタッチ

図 1: 各タッチ例

(c) 作画: 鈴木市規 (シナリオ: (株) スポマ 播村早紀/豊橋技術科学大学 上野末貴)

(c) 作画: 湯沢としひと (シナリオ: (株) スポマ 播村早紀/豊橋技術科学大学 上野末貴)

(c) 作画: 棟田ウメコ (シナリオ: (株) スポマ 播村早紀/豊橋技術科学大学 上野末貴)

3. 要素技術

3.1 自己符号化器

自己符号化器 (AutoEncoder: AE) とはニューラルネットワークの一種で、情報量を落とさずにより低次元の特徴表現を獲得するための次元圧縮手法である。AE は出力データが入力データと一致するように損失関数を定め、学習を進めるという点に特徴がある。

3.2 畳み込み自己符号化器

畳み込み自己符号化器 (Convolutional AutoEncoder: CAE) は 3.1 節で説明した AE に畳み込み層を追加したものである。CAE ではエンコーダに畳み込みと MaxPooling を用いており、デコーダに畳み込みと UpSampling を用いている。CAE は畳み込み層を有しているため、特に画像処理に対して有効である。

4. 提案手法

本研究では画像を分散表現化し、識別タスクを解くことを目的とする。しかしながらコマ画像は自然言語とは異なり、周辺のコマ画像からその特徴を抽出することが困難である。また、単語のように同じ画像が何度も出現するという事もない。そこで、本研究では他のコマの情報ではなく自身を教師ラベルのようにして学習する AE, CAE を用いて、その中間層の出力を分散表現として用いる。

本研究で使用する手法は以下の 2 通りに分かれる。

手法 1 AE を用いた分散表現による識別

手法 2 CAE を用いた分散表現による識別

本研究では AE, CAE ともに分散表現を 26626 次元から 3332 次元まで圧縮するモデルを keras を用いて構築した。

表 1: AE のモデル構造

レイヤー名	出力次元
Input	26656
Fully Connected 1	5000
BatchNormalization	5000
Fully Connected 2	3332
Dropout	3332
Fully Connected 3	5000
Fully Connected 4	26656

表 2: CAE のモデル構造

レイヤー名	出力サイズ
Input	(136, 196, 1)
Convolution 1	(136, 196, 4)
BatchNormalization	(136, 196, 4)
MaxPooling 1	(68, 98, 4)
Dropout	(68, 98, 4)
Convolution 2	(68, 98, 2)
MaxPooling 2	(34, 49, 2)
Convolution 3	(34, 49, 2)
UpSampling 1	(68, 98, 2)
Convolution 4	(68, 98, 4)
UpSampling 2	(136, 196, 4)
Convolution 5	(136, 196, 1)

4.1 AE を用いた分散表現による識別

AE を用いた分散表現による識別について説明する。表 1 に本実験で用いた AE モデルを示す。ただし、全結合層の活性化関数には最終層に sigmoid を、それ以外は ReLU を用いた。また、dropout rate は 0.5 とした。

手法 1 では Fully Connected 2 の出力を分散表現として用い、Support Vector Machine (SVM), Random Forest, および 3 層 Multi-Layer Perceptron (MLP) でコマ画像のタッチ識別をする。

4.2 CAE を用いた分散表現による識別

CAE を用いた分散表現による識別について説明する。表 2 に本実験で用いた CAE モデルを示す。ただし、畳み込み層の活性化関数には最終層に sigmoid を、それ以外は ReLU を用いた。また、dropout rate は 0.25 とした。

手法 2 では MaxPooling 2 の出力を平滑化したものを分散表現として用い、SVM, Random Forest, 3 層 MLP でコマ画像のタッチ識別をする。

5. 数値実験

本実験では、実験 1 においてはすべてのレイヤを用いて生成したコマ画像、実験 2 においては目レイヤ、口レイヤを抜いた画像を対象に、青年漫画タッチ、少年漫画タッチ、萌えタッチのいずれであるかを識別した。SVM, Random Forest での識別には、それぞれ 5 分割交差検定および grid search で得た最良のパラメータを用いた。表 3 に実験条件を示す。

本研究では AE, CAE で得た分散表現を用いてタッチ識別タスクを解く。その識別率が、得られた分散表現の性能に関する一指標になるという仮定を基に実験を進めた。

表 3: 実験で用いたパラメータ

-	AE, CAE	3 層 MLP
最適化手法	Adam	Adam
学習率	0.001	0.001
エポック数	500	100
バッチサイズ	32	32
中間層	3332	500

表 4: 実験 1 識別率

識別器	AE	CAE
SVM	0.6042	0.7083
Random Forest	0.7083	0.8125
3 層 MLP	0.5208	0.5833
ベースライン	0.3333	0.3333

表 5: 3 タッチ識別の精度, 再現率, F 値

タッチ	精度	再現率	F 値
少年漫画タッチ	0.790	0.938	0.857
青年漫画タッチ	0.700	0.750	0.727
萌えタッチ	1.000	0.750	0.857



(a) 入力画像

(b) 出力画像

図 2: AE の入力画像と出力画像



(a) 入力画像

(b) 出力画像

図 3: CAE の入力画像と出力画像

5.1 実験 1

手法 1, 2 について実験した。実験には上述の 4 コマ漫画ストーリーデータセットを用いた。画像はすべてのレイヤを用いて生成され、1 タッチ 80 枚に対して青年漫画タッチ, 少年漫画タッチ, 萌えタッチの 3 タッチより計 240 枚を用いた。この 240 枚の画像に対して AE, CAE を用いて分散表現を獲得し、対応するタッチを正解ラベルとして付与した。そして AE, CAE で得られた分散表現を SVM, Random Forest, および 3 層 MLP を用いてタッチ識別をした。

5.1.1 実験 1 の結果

図 2 に AE, 図 3 に CAE の入力画像と出力画像の例を示す。図 2 より、AE においては入力画像とほぼ同一の画像が復元されていることがわかるが、図 3 より CAE においては入力画像と出力画像を比べた時に、出力画像が少しぼやけているのがわかる。これは MaxPooling 層によって情報の欠落が起きているため、より抽象的な特徴量が抽出できていると考察できる。

表 4 に実験 1 のテスト識別率を示す。表 4 より、最も識別率が高かったのは CAE と Random Forest を用いた実験であり、その識別率は 0.8125 であった。これは CAE を用いて得た特徴量を Random Forest がうまく分離できているということを示している。

表 5 に 3 タッチ識別の精度, 再現率, F 値を示す。表 5 より、3 タッチの中で最も F 値の低かったものは青年漫画タッチであることがわかる。これは、青年漫画タッチは萌えタッチとはトーンの使い方が似ており、少年漫画タッチとは髪の描き方が似ていたため、識別が他よりも難しかったからだと考えられた。また、実験 1, 2 で誤識別した画像を調べた結果、全体的に背景やオノマトベに黒色のべた塗りが使われている傾向があることが判明した。図 4, 図 5 に実験 1 で得た分散表現を t-SNE により可視化したものを示す。各軸は次元圧縮後の各要素を示している。図 4 においては AE で得られた分散表現はおおよそのクラスが視覚的に確認できるが、図 5 における CAE で得られた分散表現にはそのような明確なクラスは存在しない。これは CAE が AE よりも高次の複雑な多様体を分散空間上で形成しているためであると考えられる。

り、3 タッチの中で最も F 値の低かったものは青年漫画タッチであることがわかる。これは、青年漫画タッチは萌えタッチとはトーンの使い方が似ており、少年漫画タッチとは髪の描き方が似ていたため、識別が他よりも難しかったからだと考えられた。また、実験 1, 2 で誤識別した画像を調べた結果、全体的に背景やオノマトベに黒色のべた塗りが使われている傾向があることが判明した。図 4, 図 5 に実験 1 で得た分散表現を t-SNE により可視化したものを示す。各軸は次元圧縮後の各要素を示している。図 4 においては AE で得られた分散表現はおおよそのクラスが視覚的に確認できるが、図 5 における CAE で得られた分散表現にはそのような明確なクラスは存在しない。これは CAE が AE よりも高次の複雑な多様体を分散空間上で形成しているためであると考えられる。

5.2 実験 2

コマ画像から 1 つのレイヤを抜いたデータセットに対して実験 1 で最も識別率の高かった CAE と Random Forest の組み合わせによりタッチを識別し、どのレイヤが識別タスクに大きく影響を与えているか調べた。今回用いたレイヤ抜き設定は顔のレイヤのうち、目レイヤを抜いた「目抜き」と口レイヤを抜いた「口抜き」の 2 種類である。これらのレイヤ抜き画像に対する分散表現を用いたタッチ識別タスクを解き、すべてのレイヤを用いた画像の識別タスクと識別率の比較を行う。

5.2.1 実験 2 の結果

表 6 に各レイヤ抜き設定に対するテスト識別率を示す。表 6 より、最も識別率の高かったオリジナル画像から最も識別率の下がったものは口抜きの画像で、すべてのレイヤを用いた識別率より約 25 % も低下したことがわかる。これは口というパーツがコマ画像のタッチ識別に大きく寄与していることを意味する。漫画としては誇張されやすい目よりも口を除いた時のほうが識別率が下がったのは、4 コマ漫画ストーリーデータセットにおいて登場人物のひとりがメガネをかけており、目を除いた影響が比較的少なかったからだと考えられる。また、表 7 にすべてのレイヤを用いた画像, 表 8 に目抜き画像, 表 9 に口抜き画像を用いた識別タスクの混同行列を示す。表 7 から表 9 より、すべてのレイヤを用いた画像の識別では、萌えタッチに対する精度が高く、目抜き画像の識別では少年漫画タッチに対する再現率が高いことがわかる。また、全体的に少年漫画タッチと識別した個数が多いのに対し、口抜き画像の識別では、青年漫画タッチと予測した個数が多かったのも特徴的な点であった。今回の識別実験においては萌えタッチと少年漫画タッチ間の誤識別はあまり見られず、この 2 タッチは図 4 で確認できるように分散空間上で離れた位置にクラスを形成しているのではないかと考察できる。

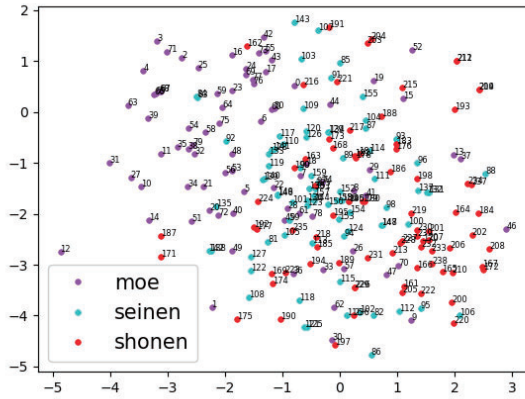


図 4: AE で得た分散表現

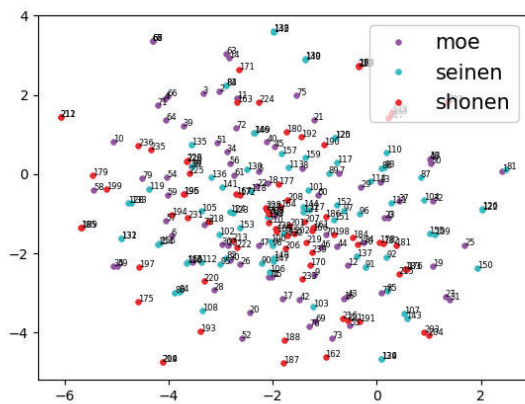


図 5: CAE で得た分散表現

表 6: 実験 2 識別率

レイヤ抜き設定	識別率
すべて使用	0.8125
目抜き	0.7917
口抜き	0.6667
ベースライン	0.3333

6. まとめと今後の課題

本研究では, AE, CAE で得た分散表現を用いて 4 コマ漫画のタッチ識別において識別率 8 割程度となる分散表現を獲得できた. また, 全体的に AE よりも CAE を用いたほうが識別率が高く, 漫画のコマ画像に対しても CAE が有効であることが確認できた. さらに, 創作視点に立つ新しいデータセットを用いることによって, これまで研究されてこなかったレイヤ抜き画像に対する識別率の変化を観察し, 顔のパーツがタッチ識別に寄与していることを確認できた. 今後の課題としては, 以下のものがあげられる.

- より効率的な分散表現の獲得
- 他のレイヤ抜き画像に対する識別率実験
- 分散表現と創作者との定量的な解析

表 7: すべてのレイヤを用いた画像識別の混同行列

		予測値		
		萌え	青年	少年
真値	萌え	12	4	0
	青年	0	12	4
	少年	0	1	15

表 8: 目抜き画像識別の混同行列

		予測値		
		萌え	青年	少年
真値	萌え	12	3	1
	青年	3	10	3
	少年	0	0	16

表 9: 口抜き画像識別の混同行列

		予測値		
		萌え	青年	少年
真値	萌え	11	4	1
	青年	1	10	5
	少年	0	5	11

謝辞

本研究は一部, 日本学術振興会科学研究補助金基盤研究 (C) (課題番号 26330282) の補助を得て行われたものである.

本研究は一部, JST, ACT-I(グラント番号:JPMJPR17U4) の支援を受けたものである.

4 コマ漫画ストーリーデータセットの制作に対し, ご協力いただいた漫画家の方々, 株式会社スポマへ謝意を示す.

参考文献

- [Matsui 17] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K.: Sketch-based Manga Retrieval using Manga109 Dataset, *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21811–21838 (2017)
- [Ogawa 18] Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T., and Aizawa, K.: Object Detection for Comics using Manga109 Annotations, *CoRR*, Vol. abs/1803.08670, (2018)
- [Saito 15] Saito, M. and Matsui, Y.: Illustration2Vec: a semantic vector representation of illustrations, pp. 1–4 (2015)
- [上野 18] 上野 未貴: 創作者と人工知能: 共作実現に向けた創作過程とメタデータ付与 4 コマ漫画ストーリーデータセット構築, 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 4Pin116–4Pin116 (2018)