The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

3 分岐型畳み込みニューラルネットワークによる 4 コマ漫画の順序識別

Recognizing the Order of Four-scene Comics by Three-Path Convolutional Neural Networks

藤野 紗耶 *1　森 直樹 *2　松本 啓之亮 *2
Saya Fujino　Naoki Mori　Keinosuke Matsumoto

*2大阪府立大学工学研究科
Graduate School of Engineering, Osaka Prefecture University

Recently, the comic analysis has become an attractive research topic in the artificial intelligence fields as comic engineering. In this study, we focused on the four-scene comics and applied deep convolutional neural networks (DCNNs) to those data for understanding the order structure. We proposed the novel approach for that problem by three-path DCNN with special input data formats. The hyperparameters of three-path DCNN are obtained by evolutionary deep learning (evoDL). The effectiveness of the proposed method is confirmed by computer simulations taking a real four-scene comics structure recognition problem as an example.

1. はじめに

近年，マンガを工学的に解析しようと試みるコミック工学 [松下 13] が機械学習の急速な発展を背景として注目されている．同分野では多様な研究可能性が示唆されているが，現状では画像に関する研究が圧倒的多数を占めている．この背景には，近年その優秀な性能により注目を浴びている深層畳み込みニューラルネットワーク (Deep Convolutional Neural Network: DCNN)[Krizhevsky 12] の存在がある．

本研究では，4 コマ漫画を対象として，DCNN によるコマの順序識別手法を提案する．コマの順序識別は，一般物体認識のカテゴリに属するキャラクター識別などとは異なり，人間の認知に直接関わるメタレベルがより高い高難度の課題であると考えることができる．一般的には，我々は画像部分と台詞部分を総合的に判断して，4 コマ漫画の識別をしているが，本研究ではこれが画像部分のみで識別可能かを示すために，コマごとに分割した画像を入力として 1〜2 コマからなる前半部分と 3〜4 コマからなる後半部分の識別を目的とする．DCNN は強力な手法であるが，一方で DCNN は非常に多くの hyperparameter を持つため，そのすべてを試行錯誤的に設定することは容易ではないという問題点もある．そこで，今回は進化型計算により優れた深層ニューラルネットワークの構造を進化的に獲得する進化型深層学習 (Evolutionary Deep Learning: evoDL)[Fujino 17] を適用し 3 分岐型 DCNN の構造を最適化した．また，実際の 4 コマ漫画の画像データを用いた数値実験によって，提案手法の有効性を示した．

2. 従来研究および本研究の位置付け

2.1 コミック工学

これまでのコミック工学は，いわゆる「マンガ」を対象としており，データセットに関する研究 [Matsui 15] や既存の画像認識手法によるコミック内パーツの識別研究を中心に発展してきた．これらの研究は，コミック工学という分野を広げるために大きく貢献している．また，ユーザの視点に立ったコミック検索システムなどについても積極的な研究がされている [山下 17].

連絡先: 藤野 紗耶，大阪府立大学，fujino@ss.cs.osakafu-u.ac.jp

一方で，狭義のコミックに収まらない対象についても研究の展開が始まっている．例えば，厳密にはコミックには分類されないアニメの絵コンテや絵本に関する研究 [Fujino 17] やストーリー生成系の研究 [Fukuda 17] などがその範疇に当てはまる．

本研究は，実際に市販されている 4 コマ漫画を対象としている点で狭義なコミック工学に属するが，一方で識別対象が具体的なパーツではなく，順序識別という点に大きな特徴がある．これはマンガリテラシー [中澤 04] を背景とした人間の認知に関する問題であり，そもそも画像認識ベースの機械学習のみで解決可能なのかすら明確ではない難解な課題であるといえる．

2.2 4 コマ漫画に関する研究

4 コマ漫画は日本における代表的な漫画形式のひとつであり，4 つのコマ (齣) によって完結した短い話を表現する．人間がストーリーを感じることができるコンパクトな表現形式として，古くから新聞や雑誌の一部に掲載されてきた．現在では，4 コマ漫画だけを扱う専門誌も数多く存在している．

4 コマ漫画の基本的なストーリー展開は，各コマを最初から順に起承転結に対応させ，結に相当する最終コマをオチとする場合が多い．それ以外にも 3 コマを序破急に対応させる場合や，2 段オチ, オチを必ずしも必要としないストーリー 4 コマなど現在では多様な形式が存在する．

従来研究としては，4 コマにおける感情識別に関する研究 [上野 16] やストーリー理解過程の解析研究 [上野 17] が報告されている．また 4 コマ漫画の自動生成に関する研究 [M. Ueno 14, Ueno 16a] や遺伝的アルゴリズムに基づく感性解析に 4 コマ漫画を用いた研究 [野村 17] もなされている．

本研究で対象とする 4 コマ漫画の順序識別に関しては，深層学習を用いた手法 [Ueno 16b] が提案されている．しかしながら従来研究では 1〜2 コマと 3〜4 コマの識別精度が 67% と十分な結果は得られていなかった．そこで，本研究では人工知能によるより高いメタレベルでの 4 コマ漫画理解を目指し，3 分岐型 DCNN による 4 コマ漫画の順序識別手法を提案する．

本研究では，ストーリーの導入に当たる 1〜2 コマとオチに関わる 3〜4 コマには画像に関する構造上の差異があることを仮定している．心理学的にこの問題を扱った研究 [Cohn 14] が報告されており，この仮定には一定の妥当性が認められるが，工学的な検証はほとんどなされていない．

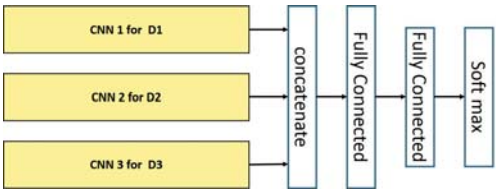

図 1: 3 分岐型 DCNN のモデル

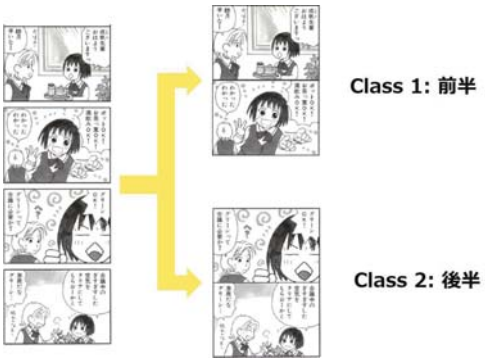

図 2: 実験におけるクラス

3. 提案手法

3.1 3 分岐型 DCNN

4 コマ漫画の順序識別に関する従来研究 [Ueno 16b] では AlexNet に類似した分岐のない深層畳み込みニューラルネットワーク (DCNN) を用いていたが，本研究では 3 分岐型 DCNN を適用する．図 1 に用いた 3 分岐型 DCNN の構造を示す．近年分岐型のモデルは多く報告されているが，これをコミック工学に応用した研究はまだ報告されていない．このモデルではそれぞれの分岐で異なる入力データおよび hyperparameter を設定することで，相補的な特徴量の抽出が期待できる．

なお，今回は 4 コマ漫画の画像のみを対象とするが，台詞の自然言語部分の消去はせずにそのまま画像として扱った点に注意を要する．

3.2 入力データ

今回は 4 コマ漫画の前半 (1〜2 コマ) と後半 (3〜4 コマ) の識別という特殊な問題を対象としたため，入力データに関して考察が必要である．一般に，人間はコマごとの情報は当然として，コマ間の連続性にも着目してストーリーを理解していると考えられる．例えば，導入に当たる前半の 1〜2 コマには大きな変化がなく，オチに関わる後半の 3〜4 コマの方が大きな変化が現れると予想される．これ以外にも，何らかの潜在的な差異が前半と後半に存在すれば，同一作家の同一作品であっても識別は可能である．以下では，本研究における 4 コマ漫画の順序識別で用いる入力データについて示す．

4 コマ漫画におけるいずれかのコマに対応する 2 画像を A, B とする．今回は A, B をサイズが 227x157，255 階調のグレイスケール画像としたが，RGB や二値画像等を用いることも可能である．ここで A, B を画素値に基づいて数値化したテンソルをそれぞれ A, B とおく．今回は A, B は行列となる．この A, B から 3 種類の入力データを作成した．

まずは従来手法と同様に A と B を縦方向に単純に連結したサイズ 227x314 のデータである．これは画像としては通常の 4 コマ漫画のコマ状態を示す．この他に，今回は $\mathcal{X}$ を A, B の各要素に作用する二項演算子として $A\mathcal{X}B$ からなるサイズ 227x157 のデータも用いる．これは人間がコマの連結を判断するときに，コマ間の差異や類似点のような相関を認識しているという仮説に基づく．$\mathcal{X}$ としては複数の演算が考えられるが，今回は最も自然な二項演算子として，加法 (A+B) と減法 (A-B) を利用した．以下では A+B によるデータを D2, A-B によるデータを D3 とする．図 2 に本実験で用いるクラスを示す．以下では，4 コマ漫画の前半を class 1，後半を class 2 とする．

表 1: GA における遺伝子表現

Design Variables	Allele
フィルター数 (NF)	32, 48, 64
フィルターサイズ (FS)	3, 5, 7
プーリングサイズ (PS)	3, 5, 7
プーリングタイプ (PT)	0(not use),1(max), 2(average)
全結合層 1 のノード数	512,1024
活性化関数 (Ac)	0(not use), 1(ReLU), 2(leaky ReLU)

4. 進化型深層学習

3 分岐の DCNN を用いるため大量の hyperparameter 調整が必要となる．そこで，進化型計算により優れた深層ニューラルネットワークの構造を進化的に獲得する進化型深層学習 (Evolutionary Deep Learning : evoDL)[Fujino 17] を利用した．

4.1 ネットワーク構造における制約

計算機資源の観点から，性能が同等の場合はよりサイズの小さいネットワークが優れていると評価するため以下の制約を導入した．

フィルター総数: 畳み込み層全体でのフィルター数の総和を 352 に固定した．evoDL によって畳み込み層 1, 2, 4, 5 のフィルター数を探索し，畳み込み層 3 のフィルター数はその他のフィルター数の和を 352 から減算することで一意に定める．

全結合層のノード総数: 今回は，出力層の直前に 3 つの全結合層を用いた．最終全結合層のノード数はクラス数に固定される．全結合層 1 および 2 のノード数を それぞれ N_1 ,N_2 と定義した場合, $N_1 \times N_2 = 2^{18}$ とする． この制約は全結合層 1 と 2 の間の重みの総数を固定することに相当する．

プーリング層 1, 2: DCNN ではプーリングが重要な役割を示すことが知られている．しかしながら進化型計算では確率的にプーリングをまったく用いない個体も生成されてしまう．このため，プーリングを持たない遺伝子型を致死個体とみなし，第 1，2 畳み込み層の後では必ずプーリングを適用するとした．ただし，この際に最大プーリングと平均プーリングのどちらを用いるかは進化的に獲得した．

4.2 個体表現

GA において，解は個体としてあらわす必要があり，それぞれの個体は染色体で表現される．この染色体を構成する記号は遺伝子と呼ばれる．さらに染色体内における遺伝子の位置を

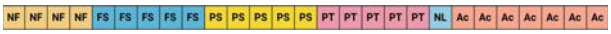

図 3: 遺伝子表現の例

遺伝子座，変更可能な遺伝子集合のことを対立遺伝子と呼ぶ．本研究において，それぞれの遺伝子座ごとに対立遺伝子の数を決定した．表 1 に設定した対立遺伝子を示す．また，図 3 に evoDL による遺伝子型の例を示す．“Ac” で表現される遺伝子は，各畳み込み層および全結合層の直後に置かれる活性化関数の種類を示す．“Ac” が 0 の場合は活性化関数を適用せず，1, 2 の場合にはそれぞれ ReLU 関数，leaky ReLU 関数を用いる．また “PT ” の遺伝子はプーリングタイプを示す．PT が 1, 2 の時それぞれ最大プーリング，平均プーリングが使われる．また，PT が 0 の時にはプーリングを適用しない．

4.3 適応度関数

GA の適応度関数 $F(s)$ として，訓練時における k 分割交差検定の識別率の総和を用いた．

5. 数値実験

evoDL により進化的に得られた 3 分岐型 DCNN を用いて数値実験をした．

5.1 問題および evoDL の設定

本実験では，全 188 話で構成される 4 コマ漫画である “こんぺいと！ 1”[ふじの 07] をデータセットとして用意した．画像サイズは **D1** を 227 × 314 pixel とし，**D2** を 227 × 157 pixel とした．本実験において，4 コマ漫画の前半 (class 1) と後半 (class 2) の間にある傾向が存在するという仮説に従い，DCNN のモデルは class 1 と class 2 の識別タスクによる適応度によって進化したモデルを用いた．

evoDL による訓練過程では 264 枚の画像を用いて，4.3 に示す適応度関数で進化させた．今回は $k = 2$ とした．それぞれの個体は DCNN モデルとして表現型に変換され，訓練がなされる．訓練の後に識別率と誤差を得るが，2 分割交差検定のため適応度関数はこれを 2 回繰り返して得ることができる．評価データとしては 112 枚の画像を用意した．このデータは最終的なエリート個体によって得られた DCNN モデルで評価される．今回は，各分岐を統合する場合の訓練に関して以下に示す手法 1, 2 による比較実験をする．

5.1.1 手法 1

3 分岐の DCNN におけるそれぞれの分岐に相当する DCNN の hyperparameter を evoDL によって進化的に設定する．また，**D1** ∼ **D3** に基づく DCNN の分岐は独立に訓練をした．evoDL によって得られた hyperparameter による各分岐の DCNN の訓練終了後，それぞれの DCNN における全結合層への入力に相当する最終畳み込み層の出力を結合し，ランダムな重みで初期化された新しい全結合層へと接続する．次に，畳み込み層のフィルターに関する重みは固定し，全結合層のみを 3 分岐の入力に基づいて訓練をする．

5.1.2 手法 2

evoDL によって 3 種の DCNN の構造を得るまでは手法 1 と同一である．手法 1 では，3 分岐のそれぞれの DCNN を独立に訓練したのちそれぞれの全結合層の結果を廃棄し，新たに 3 つの最終層を統合して全結合層の訓練をしたが，手法 2 ではこの課程を省略し最初から 3 つの DCNN を結合して分岐型 DCNN を構築した．訓練時には全結合層と畳み込み層を同時に一度だけ学習させることになる．手法 1 に比べて訓練時間が大幅に削減可能という利点がある．

表 2: evoDL における GA の設定

世代数	20
個体数	20
遺伝子長	27
交差	一様
交差率	1.0
突然変異率	$\frac{1}{L}$ (L: chromosome length)
選択	トーナメント選択
トーナメントサイズ	2
Elitism	true

5.2 実験条件

表 2 に GA の設定を示す．バッチサイズは 20 に固定し，最適化手法には Adam (学習率:2×10^{-6}) を用いた．DCNN による進化の evoDL の過程は以下に示す．

1. 最終世代で得られたエリート個体からそれぞれの DCNN の最適な hyperparameter を獲得する．
2. 得られた hyperparameter に基づく DCNN を訓練する．ここで最大訓練エポック数は 2000 と設定し，交差検定を用いないためすべての訓練データを用いた．

5.3 実験結果および考察

図 4 ∼ 6 は evoDL によって探索されたエリート個体を示す．これら 3 種のエリート個体を統合して最終的な 3 分岐型 DCNN を生成した．図 4 ∼ 6 から，畳み込み層 1, 2 のフィルターサイズはそれぞれ 7 および 3 と一致した．また畳み込み層 1 のプーリングタイプ，畳み込み層 1, 2 の直後に配置される活性化関数もそれぞれの個体で同一の遺伝子を示した．特に，それぞれの DCNN の第一活性化関数はすべて leaky ReLU が用いられていることが分かる．これは入力データが多くの負値を持つことが原因として挙げられる．オリジナルの画像は同一のものであるため，3 つの DCNN への入力データの基本的な特徴は同一なものである．以上より，evoDL によって得られた DCNN 構造が妥当であることが示された．

また， 3 分岐型 DCNN において全結合層 1, 2 のノード数はそれぞれ 1024, 256 に設定した．表 3 に それぞれのエリート個体による DCNN での 2000epoch の訓練後の識別率を示す．表 3 から **D3** を用いた際の識別率が低いことがわかる． **D3** は 2 シーンの差分であるため連続性に関する情報を含むが， class1 と class2 には連続性という観点で大きな差異が見られず，差分による情報量の減少が原因であると考えられる．表 4 に 500, 1000, 1500, 2000 epoch の訓練後の 3 分岐型 DCNN における識別率を示す．**D3** のみを用いたときの識別率は低かったが，手法 1 による 3 分岐型 DCNN が最もよい識別率を得ていることから **D3** は **D1** および **D2** の下位互換的なデータではなく，独自の重要な情報を持っていると考えられる．それぞれの分岐に相当する DCNN が異なる特徴の抽出に成功しているため， 3 分岐型 DCNN では単一の DCNN より良い識別率を得ることができたと考えられる．さらに，手法 1 の評価データの識別率は 先行研究 [Ueno 16b] の 0.67 より 0.79 と約 12% 上昇した．一方，表 4 より手法 2 の最大識別率は 2000 epoch における 0.70 で従来研究よりは改善が見られたが，手法 1 と比べると低く，また，表 3 における **D2** のみを用いた場合よりも低くなった．これは，すべての分岐を

図 4: **D1** におけるエリート個体

図 5: **D2** におけるエリート個体

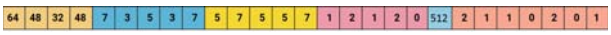

図 6: **D3** におけるエリート個体

表 3: 各 DCNN による評価識別率

モデル名	識別率
DCNN for **D1**	0.70
DCNN for **D2**	0.73
DCNN for **D3**	0.64

表 4: 手法 1, 2 による評価識別率

epoch	手法 1	手法 2
500	0.77	0.63
1000	**0.79**	0.67
1500	0.78	0.63
2000	0.77	0.70

同時並列的に学習したことで，各分岐の特徴が失われ相補的な役割を果たすことができず性能が劣化したと考えられる．**D2** と **D3** を併用した場合は， 重ね合わせデータ (**D2**) によって得られたシーンの類似性と差分データ (**D3**) によって得られたシーンの連続性をより正確に判断できたと考えられる．これらの結果より， DCNN による 4 コマ漫画の順序識別が十分に可能であることがわかった．

6. まとめと今後の課題

本研究では，機械学習による 4 コマ漫画の順序理解について検討した．evoDL によって DCNN の最適な hyperparameter を獲得し，3 分岐型 DCNN による 4 コマ漫画の前半，後半の識別をして，数値実験結果を示した．今後の課題としては提案した 3 分岐型 DCNN を用いて他の 4 コマ漫画やアニメおよび絵本などコミックに限定されないコンテンツデータによる実験が挙げられる．本研究は一部, 株式会社リバネスおよび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．また，本研究を進めるにあたり，貴重なデータを提供して頂いたふじのはるか様に多大なご協力を頂いた．この場を借りて感謝の意を示す．

参考文献

[Cohn 14] Cohn, N.: You ’re a good structure, Charlie Brown: The distribution of narrative categories in comic strips, *Cognitive Science*, Vol. 38, pp. 1317–1359 (2014)

[Fujino 17] Fujino, S., Hatanaka, T., Mori, N., and Matsumoto, K.: The Evolutionary Deep Learning based on Deep Convolutional Neural Network for the Anime Storyboard Recognition, in *Distributed Computing and Artificial Intelligence, 14th International Conference, DCAI 2017, Porto, Portugal, 21-23 June, 2017*, pp. 278–285 (2017)

[Fukuda 17] Fukuda, K., Fujino, S., Mori, N., and Matsumoto, K.: *Semi-automatic Picture Book Generation Based on Story Model and Agent-Based Simulation*, pp. 117–132, Springer International Publishing, Cham (2017)

[Krizhevsky 12] Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in Bartlett, P., Pereira, F., Burges, C., Bottou, L., and Weinberger, K. eds., *Advances in Neural Information Processing Systems 25*, pp. 1106–1114 (2012)

[M. Ueno 14] M. Ueno, K. M., N. Mori: 2-Scene Comic Creating System Based on the Distribution of Picture State Transition, in *Advances in Intelligent Systems and Computing*, Vol. 290, pp. 459–467 (2014)

[Matsui 15] Matsui, Y., Ito, K., Aramaki, Y., Yamasaki, T., and Aizawa, K.: Sketch-based Manga Retrieval using Manga109 Dataset, *CoRR*, Vol. abs/1510.04389, (2015)

[Ueno 16a] Ueno, M.: Computational Interpretation of Comic Scenes, in *Advances in Intelligent Systems and Computing*, Vol. 474, pp. 387–393 (2016)

[Ueno 16b] Ueno, M., Mori, N., Suenaga, T., and Isahara, H.: Estimation of structure of four-scene comics by convolutional neural networks, in *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding*, p. 9ACM (2016)

[ふじの 07] ふじの はるか：こんぺいと! 1, 芳文社 (2007)

[山下 17] 山下 諒, 朴炳宣, 松下 光範：コミックの内容情報に基づいた探索的な情報アクセスの支援, 人工知能学会論文誌, Vol. 32, No. 1, pp. WII–D_1–11 (2017)

[松下 13] 松下 光範：コミック工学：マンガを計算可能にする試み, 日本知能情報ファジィ学会ファジィ システム シンポジウム 講演論文集, Vol. 29, pp. 199–199 (2013)

[上野 16] 上野 未貴, 森 直樹, 松本 啓之亮：漫画内の特徴的要素が与えるストーリーの印象についての検討, 第 30 回人工知能学会全国大会発表論文集, pp. 2J5–OS–08b–4in2 (2016)

[上野 17] 上野 未貴, 末長 寿規, 井佐原 均：漫画中の表現獲得方法に基づくストーリー理解過程の解析, 第 31 回人工知能学会全国大会発表論文集, pp. 4F1–5in2 (2017)

[中澤 04] 中澤 潤：マンガ読解力の規定因としてのマンガの読みリテラシー, マンガ研究, Vol. 5, pp. 6–25 (2004)

[野村 17] 野村 俊太, 荒井 幸代：進化計算を用いた人間の感性理解のための遺伝子解析法, 第 31 回人工知能学会全国大会発表論文集, pp. 3H2–OS–04b–2 (2017)