

絵コンテ識別のための 深層畳み込みニューラルネットワークの進化的獲得

The Evolutionary Deep Learning based on Deep Convolutional Neural Network for the Anime
Storyboard Recognition

藤野 紗耶 *1 畑中 太一 *2 森 直樹 *3 松本 啓之亮 *3
Saya Fujino Taichi Hatanaka Naoki Mori Keinosuke Matsumoto

*1大阪府立大学工学域電気電子系学類情報工学課程 *2大阪府立大学工学部知能情報工学科
Osaka Prefecture University Osaka Prefecture University

*3大阪府立大学大学院工学研究科
Osaka Prefecture University

Recently, the researches of image recognition have been developed remarkably by means of deep learning. In this study, we focused on the anime storyboards and applied deep convolutional neural networks (DCNNs) to those data for obtaining a video-conte (V-conte) in the anime. Though DCNN is a very powerful method, the problem is that it takes lots of effort to tune its hyperparameters. To solve this problem, we propose a novel method called evolutionary deep learning (evoDL) by means of genetic algorithms (GAs).

The effectiveness of evoDL is confirmed by computer simulations taking a real anime storyboard recognition problem as an example.

1. はじめに

近年のアニメブームにより、日本国内だけでも年間 100 本以上のアニメが制作されている。それにより、アニメのクオリティは進歩を遂げており 1 本あたりにかかる制作費や時間も上昇傾向にある。一方で、アニメを作る際の出資額がそれに伴っておらず、制作会社の金銭的負担が増加している。そのため、制作会社が経営破綻を起こすことや、アニメに携わる人間の低賃金化が問題になっている。そこで、現場の負担の軽減を目的に工学的な手法でアニメ作業の効率化を図り、アニメ制作会社への負担軽減が必要とされている。

アニメ制作において、絵コンテが重要な役割を果たすことが知られている。絵コンテはアニメの元となる多くの情報を含む一種のコミックであるとみなすことができる。特に、絵だけではなく自然言語や記号が重要な意味を持つという点がコミックと類似している。一方で、絵コンテは必ずしも絵に精通した人間が描くとは限らないため、線の精緻さやキャラクターの同一性の観点からはラフスケッチであるとも考えることもできる。このような性質と、絵コンテデータの入手困難性から、絵コンテの工学的な解析はほとんどなされてこなかった。そこで今回は、これまでほとんど研究がなされていない実際のアニメの絵コンテに着目し、絵コンテから V コンテを自動生成を最終目的として、絵コンテからのコマ画像の切り出しと絵コンテデータの画像認識について検討する。

アニメにおける絵コンテの識別部には、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) を導入する。DCNN は強力な手法であるが、一方で DCNN は非常に多くのパラメータを持つため、ネットワーク構造および可調整パラメータすべてを試行錯誤的に設定することは容易ではない。そこで、今回は進化的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的深層学習 (Evolutionary Deep Learning: evoDL) を適用した。evoDL では任意の進化的計算を応用することができるが、今回は代

連絡先: 藤野紗耶, 大阪府立大学工学域電気電子系学類情報工学課程 4 年, fujino@ss.cs.osakafu-u.ac.jp

表的な進化的計算 (Evolutionary Computation: EC) である遺伝的アルゴリズム (Genetic Algorithm: GA)[1] を採用し、DCNN の整数型のハイパーパラメータを中心とした最適化をした。また、日本で実際に放送されたアニメの絵コンテ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. アニメにおける絵コンテ

本章では、アニメにおける絵コンテについて述べる。

2.1 アニメの制作工程



図 1: アニメ工程の全体図

図 1 にアニメ工程の全体図を示す。アニメの製作工程には大きく分けて、絵コンテ、原画、作画監督、動画、仕上げ、撮影という工程があり、その工程をまとめる制作進行という人間がいる。それぞれの工程は以下のような作業となる。

絵コンテ シナリオを元に絵コンテを書く

原画 絵コンテを元にアニメの動きの元となる原画を描く

演出 原画に対して表情や構図の修正をする

作画監督 原画のキャラなどのディテールを整える

動画 原画の中割の作成や、線のクリンナップをする

仕上げ 作成された動画に対してデジタルペイントをする

撮影 デジタルペイントされたデータを動画にする

制作進行 上記全体のスケジュールの管理をする

2.2 絵コンテの重要性

絵コンテにはカメラがどのように動くか、どのキャラクターがどのような動きをするか、どのような会話が発生するかなどの画面を構成する情報が多く入っている。そのため、アニメの制作作業に入る前に絵コンテを解析し、物量を事前に把握することで、より効率的に作業を進めることが出来ると考えられる。本研究の最終目的である V コンテはアニメに音をつけるアフレコ工程で使用する重要な中間生成物であり、その生成には多大な手間がかかることが知られている。この部分を自動化できればアニメ制作に掛かる費用の大幅な削減が可能となる。

Creators
inPack Inc.

No	ピクチャー	内容	セリフ	秒数
1		▲(F・I) (12k) 敷(前)前(行)利 ・(ヤ)PAU - UP)		
		PAN UP ▼(0・I) (1+0)	スカーレット 「ここが、今日から 通うオズ学園か…」	(4+0)
2		(FIX) (4+0) ば(音) 敷(音) (音) (サブタイトル in)	スカーレット (off) 「地味で人見知りで 頭が固くて 要領が悪くて	
		(PAN) (4+12) キャラ (スライド上下歩き)	引くのは貧乏くじ ばかりの横が うまくやっていると だろうか。	
			心配だな……」	(9+0)
		▼(0・I) (0+12)		

図 2: 絵コンテの例 (@Poni-Pachet SY/HOBIBOX/OZMAFIA!! 製作委員会 [2])

2.3 絵コンテにおけるカメラワーク

アニメの設計図である絵コンテは、カメラワーク、セリフ、時間尺などの要素を含んでいる。V コンテの自動生成のためには絵コンテのカメラワークに注目し、解析をする必要がある。以下に代表的なカメラワークの例を 3 つ示す。

T.U トラックアップと言われるカメラワーク。カメラが被写体に向かって近づき、対象を大きく写していく。

T.B トラックバックと言われるカメラワーク。カメラが被写体から遠ざかり、対象を小さく写していく。

PAN 固定したカメラの向きを上下左右に振るカメラワーク。

3. 深層畳み込みニューラルネットワーク

本研究では近年、画像認識分野において視覚野における受容野の性質に着想を得た畳み込みニューラルネットワーク (Convolutional Neural Network: CNN)[3] を利用する。図 3 に Convolution 層と Pooling 層を用いる CNN の概念図を示す。

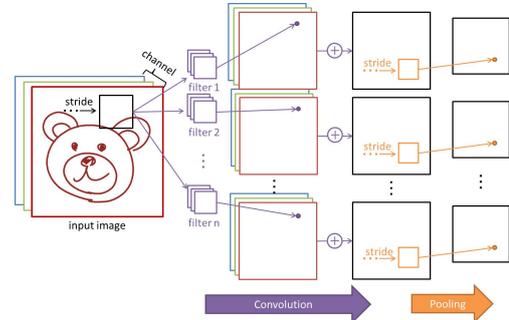


図 3: CNN 概念図

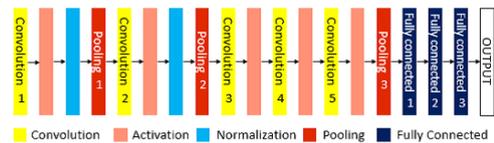


図 4: DCNN の構造

CNN は主として画像の局所的な特徴抽出をする Convolution 層と Convolution 層で得られた特徴を統合する Pooling 層からなる。図 4 に今回用いたネットワークの構造を示す。このネットワーク構造は AlexNet [4] と呼ばれる Hinton らによって提案された 深層 CNN (Deep CNN: DCNN) の代表的なモデルのひとつである。ここで、Convolution Layer は Convolution 層, Pooling Layer は Pooling 層, Fully connected Layer は隣接層間のユニットが全結合した全結合層を表す。本研究では活性化関数として以下の ReLU 関数を用いた。

$$f_{\text{ReLU}}(x) = \max(0, x) \quad (1)$$

4. 提案手法

本研究では、絵コンテからコマを切り出し、擬似的な V コンテを作成する手法と、絵コンテから情報を抽出するために、絵コンテを識別するための深層畳み込みニューラルネットワーク (Deep Convolutional Neural Network: DCNN)[5] を進化型深層学習 (Evolutionary Deep Learning: evoDL) によって獲得する手法を提案する。

4.1 絵コンテ画像におけるコマの自動抽出

画像認識により、絵コンテの枠である直線を認識し、絵コンテにおける画像を表すコマとテキストにより状態を表すコマを、順序をラベルとした形式で切り出し表示するシステムを構築した。これが V コンテの基盤となる。ただし、実際には絵コンテ中に、カメラワークを示す、PAN, 付け PAN, T.U, T.B などの指示がありその情報に基づいて V コンテを作成する必要があるが、現段階では一定の間隔で表示するのみとし

表 1: 遺伝子型における対立遺伝子

Design Variables	Allele
フィルター数 (NF)	8, 16, 32
フィルターサイズ (FS)	3, 5, 7
プーリングサイズ (PS)	3, 5, 7
全結合層 1 のノード数 (NL1)	64, 128, 256, 512
全結合層 2 のノード数 (NL2)	16, 32, 64, 128
バッチサイズ (BS)	5, 10, 15
ReLU 活性化関数 (Re)	1 (use), 0 (not use)

表 2: DCNN による絵コンテの識別結果

	Storyboard Creator		Accuracy	
	Anime1	Anime2	500 epoch	2000 epoch
episode 1	A (215)	C, K (198)	train	train
episode 2	B (155)	C, G (149)	train	train
episode 3	D (135)	A (162)	train	train
episode 4	C, B (133)	L, K (217)	train	train
episode 5	A (177)	M, B (134)	train	train
episode 6	E (114)	N, B (115)	0.99	0.98
episode 7	D (173)	O (139)	0.99	1.0
episode 8	F (161)	A (149)	0.47	0.49
episode 9	B (242)	P (169)	0.83	0.92
episode 10	G (158)	A (185)	0.43	0.48
episode 11	H, I (226)	B (170)	0.90	0.88
episode 12	J (153)	Q (127)	0.73	0.83
weighted average accuracy(E1)	-	-	0.76	0.79
shuffled data accuracy(E2)	-	-	0.96	0.97

た. また, 非表示コマもすべて表示するとした. コマの中から絵コンテのカメラワークに関する情報を抽出し, 適切なカメラワークを付与することで, 精度の高い V コンテが作成可能であると考えられる.

4.2 進化型深層学習による絵コンテ識別

深層学習を実現するためには大量の hyperparameter の調整が必要となるため, 筆者らは, 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獲得する進化型深層学習 (Evolutionary Deep Learning : evoDL) を提案してきた. 今回は evoDL による深層畳み込みニューラルネットワーク (DCNN) を用いて絵コンテ識別の可能性について検討した.

4.2.1 進化型深層学習

進化型深層学習 (evoDL) は, 基本的には任意の深層学習の構造を任意の進化型計算によって最適化可能である. しかしながら, 現段階では, DCNN のすべての構造を進化的に獲得することは極めて困難であったため, 本研究では DCNN の基本的な構造は AlexNet[5] と同一で固定とし, DCNN における離散的に表現可能なパラメータのみに焦点を当てて evoDL を適用した. また, 進化型計算には遺伝的アルゴリズム (Genetic Algorithm: GA)[1] を用いた.

4.2.2 遺伝子表現

GA において, 最適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を持つ. この染色体を作り出す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に特定の対立遺伝子の数を決定した. 表 1 に遺伝子型における対立遺伝子のとり得る値について示す. また, 図 5 に evoDL による遺伝子型の例を示す.

4.2.3 適応度関数

本研究では, GA の適応度関数 $F(s)$ として, 進化個体評価用訓練データを k 分割交差検定にかけることで得られた交

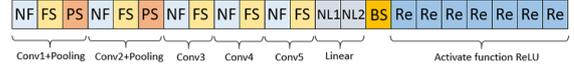


図 5: evoDL による遺伝子型の例

表 3: evoDL における GA の設定

世代数	12
個体数	16
遺伝子長	22
交叉	一様
交叉率	1.0
遺伝子座ごとの突然変異率	$\frac{1}{L}$ (L : 遺伝子長)
選択	トーナメント選択
トーナメントサイズ	3
エリート主義	あり

差エントロピーに基づく誤差の和を用いる. 式中の s はある DCNN に関わる個体を表す. この最適化問題は下の式によって表される.

$$\arg \max_s F(s) \quad (2)$$

$$F(s) = - \sum_{i=1}^k f_{\text{loss}_i}(s) \quad (3)$$

この他にも識別率の和や簡潔なネットワークを優先するようにバイアスをかける適応度関数が考えられる.

4.3 GA による DCNN 獲得の流れ

表 3 に今回用いた evoDL における GA の設定を示す. GA による DCNN 獲得の流れを以下に示す.

1. GA の個体群を初期化する.
2. GA における各個体が表現する DCNN を生成する.
3. 進化個体評価用訓練データによる k 分割交差検定により DCNN を評価し, 適応度を算出する. この際, 訓練時に用いるデータ数を一定とするため (バッチサイズ, 訓練期間) を (5,150), (10,75), (15,50) とする.
4. 算出された適応度に基づき, GA によって個体を進化させる.
5. 最終世代におけるエリート個体を進化的に獲得した DCNN のパラメータとする.
6. 進化的に獲得したパラメータに基づいて最終的な DCNN を訓練する. この場合は, k 分割交差検定ではなく, すべての進化個体評価用訓練データを利用し, 訓練期間を 2000 epoch とする.

4.4 計算時間短縮手法

DCNN の進化と訓練には多くの時間と手間を要するために, 本研究では 2 種類の計算時間短縮手法を用いた. GA 内において既に適応度を計算した個体が再び出現した場合, それぞ

れの個体毎の適応度を保存しておくことで再度適応度計算をする時間を短縮した。また進化時の DCNN の訓練において、10 epoch を終えた後、1 epoch 時から 0.01 以上識別率が上がらなかった場合、その遺伝子型での訓練を中断することで、評価学習の識別率が上がりきらない可能性の高い個体を訓練時に取り除き、計算時間を短縮した。

5. 数値実験

今回は、絵コンテの画像識別がそもそも可能なのかという点を明らかにするために、まずは異なるタイトルのアニメの絵コンテ識別について数値実験をした。

図 6 に evoDL で得られたエリート個体の遺伝子型を示す。

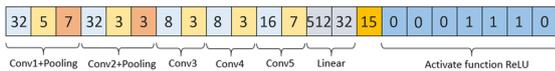


図 6: evoDL により得られたエリート個体

本実験では、2014 年～2015 年の間に日本で放映されたアニメ 2 作品に関する絵コンテデータを対象とした。どちらのアニメも 12 話構成で、絵コンテとしてサイズ 100×141 の各話約 150 画像を含んでいる。以後これら 2 アニメを Anime1 および Anime2 とし、以下の 2 実験をした。

- E1** 第 1 話～第 5 話までを訓練データとして、第 6 話～第 12 話までをテストデータとして用いて、Anime1 と Anime2 を識別する。Anime1, Anime2 の訓練データ数はそれぞれ 860, 815 である。
- E2** 各アニメそれぞれの 12 話分のデータをランダムに 5:7 の比率で分割し、5/12 のデータを訓練データとして、7/12 のデータをテストデータとして用いて、Anime1 と Anime2 を識別する。

今回は、適応度関数を算出する際、 $k = 3$ とした。表 2 に、エリート個体が示す DCNN による絵コンテの識別結果を示す。表中では、大文字のアルファベットで各話の絵コンテの担当者を、括弧内の数字で画像データ数を示している。evoDL で獲得した hyperparameter に基づいて DCNN を 2000epoch まで訓練した結果、テストデータにおける識別率は **E1** で 79%、**E2** で 97% となった。**E1** と **E2** の識別率の差より、複数の絵コンテ制作者に関する情報を均一化している **E2** の方が難易度が低い問題となることわかる。これは同一の制作者が異なるタイトルの絵コンテを担当する場合があるためである。例えば第 8, 10 話の識別率がベースラインである 0.513 ($\approx 860 / (860 + 815)$) よりも低いことがわかる。この識別率の低さの理由としては、絵コンテ担当者“A”によると考えられる。A は Anime1 の第 1, 5 話の絵コンテを担当しているが、一方で A は Anime2 の第 8, 10 話も担当している。また、訓練データである Anime2 の第 3 話も A が担当している。A は Anime1 の監督でもあるため、A は Anime2 よりも Anime1 に対する影響力が強いと考えられる。最も識別率が低かった第 10 話については、A が Anime2 の絵コンテを担当していることに加えて、Anime2 の訓練データである第 2 話を担当している G が Anime1 の第 10 話を担当しており、A と G が丁度訓練データと互い違いの関係で第 10 話を担当していたことが原因であると考えられる。

絵コンテの識別においては、アニメタイトルと同様に制作者の画風を考慮しなければならないという知見が得られた。実際に V コンテの自動生成のためには今回のようなアニメタイトルごとの絵コンテ識別の必要性は低いが、実際の絵コンテデータに対する画像識別の研究例自体がほとんどないため、今回は一定の成果が得られたと考えられる。

6. まとめ

本研究では、絵コンテ画像から V コンテ用のコマ切り出しをする手法と、evoDL により得られた深層学習による絵コンテデータの識別手法を提案し、数値実験結果を示した。今後の課題としては、以下の項目があげられる。

DCNN の進化的獲得 今回は、DCNN のハイパーパラメータの中で、整数パラメータのみを対象としたが、進化的に獲得したい内容は多岐にわたる。今後は、構造を考慮した遺伝子型の提案や CMA-ES[6] による実数値パラメータを対象とする。

深層学習による絵タッチの学習 絵コンテの担当者は、絵のエキスパートからほとんど初心者まで絵の習熟度に大きな差がある。一方で、アニメ業界では同一人物が異なるアニメの絵コンテを担当することが多い。そこで、事前にユーザの絵タッチを学習することで、絵コンテの識別精度を向上させる。

実用的なレベルの V コンテの自動生成 今回は、切り出したコマの単純な表示のみを実現したが、今後は不要なコマの削除、カメラワークの理解、表示秒の反映などを考慮した実用的なレベルの V コンテの自動生成を実現する。

なお、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。

参考文献

- [1] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [2] Creators in pack. <http://anime.ozmafia.com/>.
- [3] Patrice Y Simard, David Steinkraus, and John C Platt. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, Vol. 3, pp. 958–962, 2003.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In P. Bartlett, F.c.n. Pereira, C.j.c. Burges, L. Bottou, and K.q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pp. 1106–1114. 2012.
- [6] Nikolaus Hansen. The cma evolution strategy: a comparing review. In *Towards a new evolutionary computation*, pp. 75–102. Springer, 2006.