

Summary of Bachelor's Thesis
2015年3月修了卒業論文概要書

Name 氏名	井手和磨	ID number 学籍番号	1W110042-4
題目 Title (日本語の場合は英文題目も記入)	Supervisor 指導教員 渡辺 裕		印
日本語 Japanese コミックにおける背景画像によるジャンル分類について 英語 English On Genre Classification by Comic Background Images			

概要 Summary

現在、電子コミックにおけるメタデータ抽出のための研究として、コマの認識やキャラクター同定、吹き出し検出などが行われている。本論文では、電子コミックにおける新たな利便性付与の可能性として、コミックの背景画像によるジャンル分類の自動化の実現性を検討する。コミックの整理や索引のしやすさの向上につながると考えられる。

本論文では、まず背景画像のパターンを手がかりとしたコミックのジャンル分類の実現性に向けて、ジャンルごとにどの背景が多く使われているか手動で集計し、評価した。

その結果“スポーツマンガでは集中線や効果線が多い”や“少女マンガには点描が使われている”などの背景画像のパターンごとに異なる特徴を持つことが確認できた。

次に畳み込みニューラルネットワークを用いた背景画像の分類を検討した。畳み込みニューラルネットワークとは、音声認識や画像認識によく用いられ、未知データに対しても高い識別率を持つ手法である。

実験では、切り出したコミックの背景画像を畳み込みニューラルネットワークで学習させ、入力画像を背景画像のパターンごとに分類できるかについて検証した。

その結果、全体の認識率は66.7%、エラー率は19.3%であった。また、背景画像が風景、フラッシュ、スクリーントーンの場合に認識しやすく、空白と点描の場合に認識しにくいということが分かった。

今後、分類精度を向上することが出来れば、コミックのジャンル分類は実現に一步近づくと考えられる。

基幹理工学部情報理工学科

Bachelor's Thesis
卒業論文

Title
論文題目

コミックにおける背景画像によるジャンル分類について

On Genre Classification by Comic Background Images

Student ID 学籍番号	1W110042-4
Name 氏名	Kazuma Ide 井手 和磨

Supervisor 指導教員	Hiroshi Watanabe 渡辺 裕	印
--------------------	--------------------------	---

2015年3月15日

目次

第1章	はじめに	1
	1.1 研究の背景-----	1
	1.2 本研究の目的-----	1
	1.3 論文の構成-----	1
第2章	コミックにおけるジャンルごとの背景の割合集計	2
	2.1 コマの背景パターン-----	2
	2.2 ジャンルの種類-----	4
	2.3 ジャンルごとの背景の割合集計-----	5
	2.4 ジャンルごとの背景の特徴-----	8
	2.5 むすび-----	9
第3章	畳み込みニューラルネットワークを用いた背景分類	10
	3.1 まえがき-----	10
	3.2 ニューラルネットワーク-----	11
	3.2.1 人間の脳のシステム-----	11
	3.2.2 ニューロンのモデル化-----	12
	3.2.3 単純パーセプトロン-----	13
	3.2.4 多層パーセプトロン-----	15
	3.2.5 誤差逆伝播法-----	16
	3.3 ディープラーニング-----	17
	3.3.1 畳み込みニューラルネットワークの構造-----	17
	3.3.2 畳み込み層-----	18
	3.3.3 プーリング層-----	20
	3.3.4 整流層-----	21
	3.3.5 正規化層-----	21
	3.4 むすび-----	21
第4章	実験概要と結果, 考察	22
	4.1 まえがき-----	22
	4.2 実験概要-----	22
	4.3 実験条件-----	22
	4.4 実験結果-----	24
	4.5 考察-----	26
第5章	結論と今後の課題	27
	謝辞	28
	参考文献	29

圖一覽	31
表一覽	32

第1章 はじめに

1.1 研究の背景

日本の電子書籍は多くのユーザに利用され、その市場は今後も成長を続けると予測されている。2013年度の電子書籍市場規模は936億円と推計され、2012年度の729億円から、207億円も増加している。2014年現在は、スマートフォンやタブレット、電子書籍専用端末といった新たなプラットフォーム向け電子書籍市場が急速に拡大しており、本格的な拡大期である。そのため2014年度以降の日本の電子書籍市場は今後も成長を続け、2018年度には2013年度の2.9倍の2790億円にも上ると予測されている[1]。

その中でも電子コミックは最も成長率が高いと見なされている。電子コミックとは、一般のコミック本をスキャンしたものを携帯電話やタブレット端末で読むことが一般に想定される。しかし、電子コミック化している作品を一般のコミックとして印刷するという逆の流れも存在している。このように、現在の電子コミックは紙媒体の一般のコミックの消費に影響を与えるほど普及し始めている。特に2013年にスタートした“comico”スマートフォン向け無料コミック配信サービスにおける“ReLIFE”が典型的な一例である[2][3]。

現在では、今後さらに普及していくであろう電子コミックについて新たな可能性を探り、様々な研究が行われている。従来の主だった研究には、濃度勾配の方向またはコミックのコマは矩形であることが多いという特徴を利用したコマの認識、HOG特徴量を手がかりとして予め作成した顔画像とのデータベースマッチングによる登場キャラクターの同定、AdaBoostによって文字領域を特定する吹き出しの分類などの研究がある[4]。

これらの研究が進められている中で、新たな可能性としてコミックのジャンル分類の自動化の可能性を探りたいと考えた。コミックのジャンル分類の自動化が実現することで、コミックの整理や索引のしやすさの向上につながると考えられる。また、ユーザと作者の両者にとって電子コミックにおける利便性が向上するはずである。

1.2 本研究の目的

本研究では、背景に述べた点を踏まえ、コミックのジャンル分類の自動化の実現性を評価することを目的とする。ジャンル分類を行う手がかりとして、コマごとの背景の割合を検討する。言い換えると、ジャンル単位でコマごとに多く使われている背景は何かについて、まず手動で集計する。さらに、ジャンルごとに背景の割合を特徴として、その特徴によってジャンル分類を行えるのではないかと仮説を立て、その検証を行う。

1.3 論文の構成

本論文では、第2章ではコミックにおけるジャンルごとの背景の割合集計について述べる。第3章では畳み込みニューラルネットワークを用いた背景分類について述べる。第4章では実験概要と結果、考察について述べる。第5章では、本論文の結論と今後の課題を述べる。

第2章 コミックにおけるジャンルごとの背景の割合集計

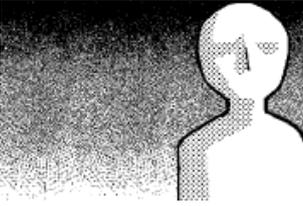
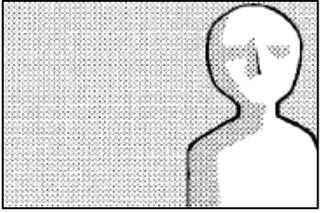
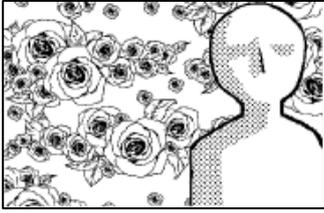
本章では背景パターンによるコミックのジャンル分類の実現の可能性を評価する。

2.1 コマの背景パターン

現代のコミックにおける背景のパターンを表 2.1 に示す[5]。

表 2.1 コミックにおける背景のパターンとその特徴

1. 風景を用いた背景	2. 空白を用いた背景
 <p data-bbox="228 853 778 1025">切り取られた一瞬をきめ細かに表す。写真のように静止した状況を表示し、他の背景と組み合わせることも可能で、さまざまな可能性がある。</p>	 <p data-bbox="810 853 1361 1025">背景以外の要素（登場人物や文字）に注目を集め、それを目立たせることに使われる。時間の変化や登場人物の心理描写など特別な意味を持つ場合もある。</p>
3. 集中線を用いた背景	4. 効果線を用いた背景
 <p data-bbox="228 1339 778 1467">注目をある一点に向ける効果を持ち、スピード感や臨場感を表すときに用いられる。激しい動きや驚きを表現する。</p>	 <p data-bbox="810 1339 1361 1512">集中線と似ており、スピード感や高揚感を表すときに用いられる。集中線よりも決まった方向へのスピード感を出す際に用いる。</p>
5. フラッシュを用いた背景	6. ベタを用いた背景
 <p data-bbox="228 1827 778 1955">集中線とベタを組み合わせたような形状で、閃きやスピード感、独特の雰囲気醸し出すのに用いる。</p>	 <p data-bbox="810 1827 1361 1955">コマ全体を塗りつぶしたもの。おどろおどろしい雰囲気を醸し出し、シリアスな雰囲気を持つ。アクセントとしての効果も持つ。</p>

7. 点描を用いた背景	8. カケアミを用いた表現
 <p data-bbox="228 566 786 696">ふんわり感や愛情表現などの微妙な表現を表すのに用いられる。小さい点の集まりで帯状や円状に配置されるものが多い。</p>	 <p data-bbox="810 566 1369 741">網目模様を作り、いくつも隙間なく並べるようなもの。ベタよりもおどろおどろしい雰囲気醸し出し、読者をゾワゾワさせたい時などに用いる。</p>
9. スクリーントーンを用いた背景	10. 模様を用いた背景
 <p data-bbox="228 1061 786 1236">規則正しく円状の点が並んでいるもの。点の大きさはいくつも種類があり、点の大きさにより様々な印象を与える。空白やベタよりも何気ない感じを醸し出す。</p>	 <p data-bbox="810 1061 1369 1236">同じようなパターンの模様を敷き詰めたり、バラバラに配置したりして、軽い感じやギャグっぽさを表す。主に雰囲気や感情を表す。</p>

コミックのコマの背景は大別して上記の 10 パターンに分けられる。これらの 10 パターンの他に背景がないコマなども存在する。

2.2 ジャンルの種類

コミックにおけるジャンルの種類とその例を以下の表 2.2 に示す[6].

表 2.2 コミックのジャンルの種類とその例

コミックのジャンル	例
バトル	ドラゴンボール, ワンピース, マギ
ギャグ	天才バカボン, ピューと吹くジャガー
スポーツ	キャプテン翼, スラムダンク
格闘	グラップラー刃牙, キン肉マン
推理サスペンス	名探偵コナン, 金田一少年の事件簿
料理	おいしんぼ, 食劇のソーマ
日常物	こちら亀有区亀有公園前派出所, クレヨンしんちゃん
少女	僕等がいた, 君に届け
ホラー	ゲゲゲの鬼太郎, 洗礼
ファンタジー	ベルセルク
ハードボイルド	ゴルゴ 13, シティーハンター
車	頭文字 D, 湾岸ミッドナイト

ジャンルを正確に定義付けするのは困難である. しかし, 知られているだけでも表 2.2 ほど存在する.

2.3 節では上記のジャンルから, “バトル”, “スポーツ”, “推理サスペンス”, “日常物”, “少女” の 5 ジャンルを取り上げ, 2.1 節による 10 個の背景パターンとの関係性を手動で集計する.

2.3 ジャンルごとの背景の割合集計

“バトル”，“スポーツ”，“推理サスペンス”，“日常物”，“少女”の5ジャンルを各3作品ずつ，1冊単位で各コマの背景の10パターンにおける割合を集計する．集計した作品は以下の表2.3の作品である．

表 2.3 集計を行ったジャンルと作品名[7]

ジャンル	作品名
バトル	ワンピース，ドラゴンボール，マギ
スポーツ	キャプテン翼，オーバードライブ，メジャー
少女	僕等がいた，僕は妹に恋をする，僕の初恋を君に捧ぐ
日常物	こちら亀有区亀有公園前派出所，クレヨンしんちゃん，釣りバカ日誌
推理サスペンス	名探偵コナン，金第一少年の事件簿

集計結果を次の表2.4に表す．

表 2.4 コミックのジャンルの背景のパターンの集計結果

	バトル	スポーツ	推理サスペ ンス	日常物	少女
1.風景	34.05%	33.57%	31.45%	55.03%	22.85%
2.空白	30.18%	26.86%	13.73%	30.24%	47.18%
3.集中線	20.01%	34.3%	12.5%	5.06%	1.44%
4.効果線	9.64%	16.27%	7.99%	6.09%	2.21%
5.フラッシュ	0%	3.59%	0.61%	0.22%	0.68%
6.ベタ	1.29%	4.35%	11.01%	0.15%	1.95%
7.点描	0%	0.06%	0.05%	0%	14.56%
8.カケアミ	0.04%	2.23%	6.3%	0%	0.13%
9.スクリーン トーン	3.11%	9.08%	15.88%	2.26%	4.37%
10.模様	0%	0%	0%	0.22%	1.7%
11.その他	0%	2.4%	0.46%	0.63%	2.55%
12.なし	1.67%	2.21%	0%	0.1%	0.38%

その他とは，10パターンはどこに属するか判断の難しい背景を持つコマ，なしとは背景自体がないコマである．

表 2.4 を参照し、各ジャンルの背景の割合における特徴を円グラフで表す。(図 2.1～図 2.5)

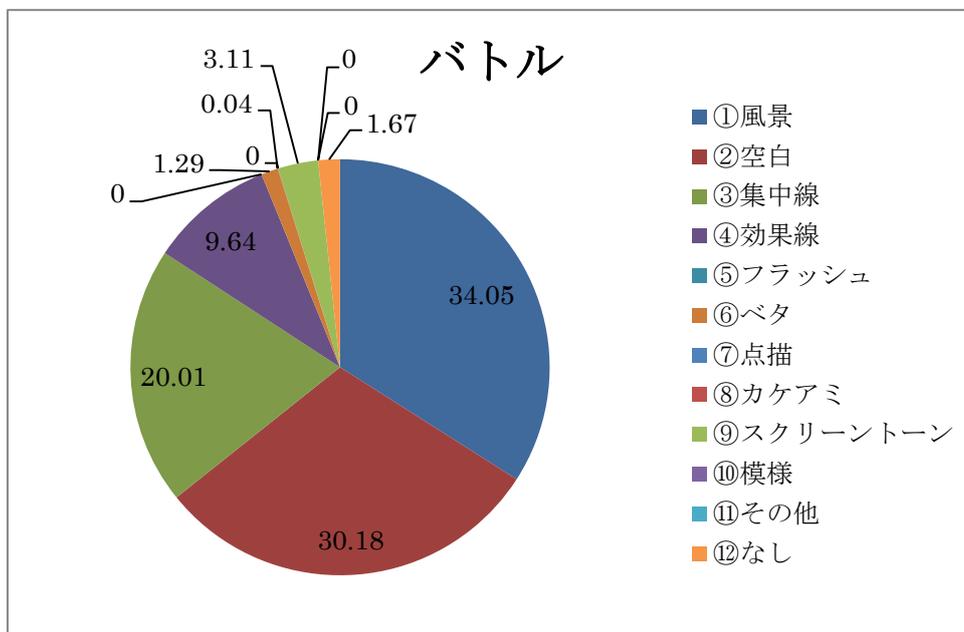


図 2.1 バトル漫画における背景の割合集計

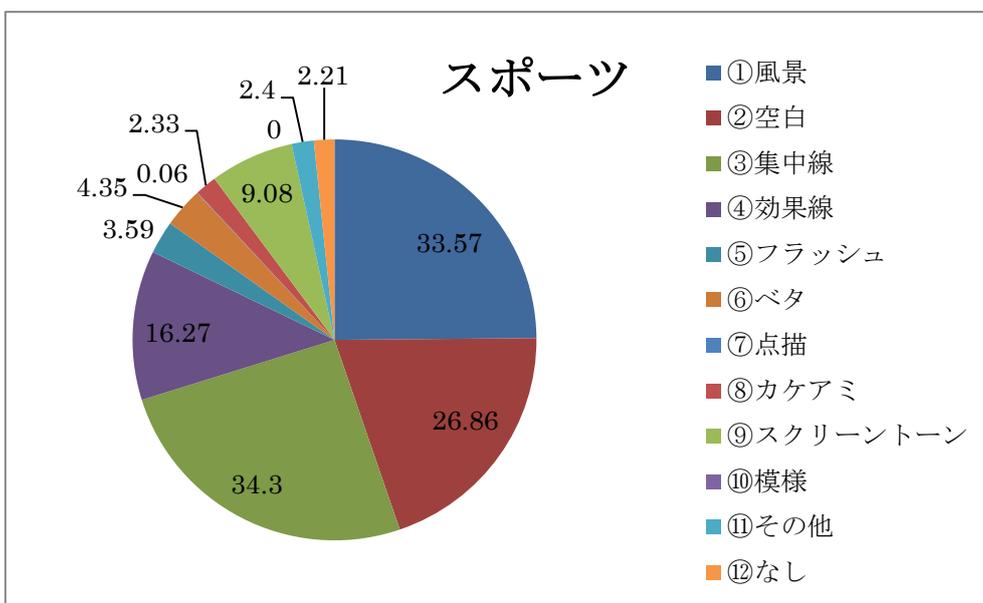


図 2.2 スポーツ漫画と背景の割合集計

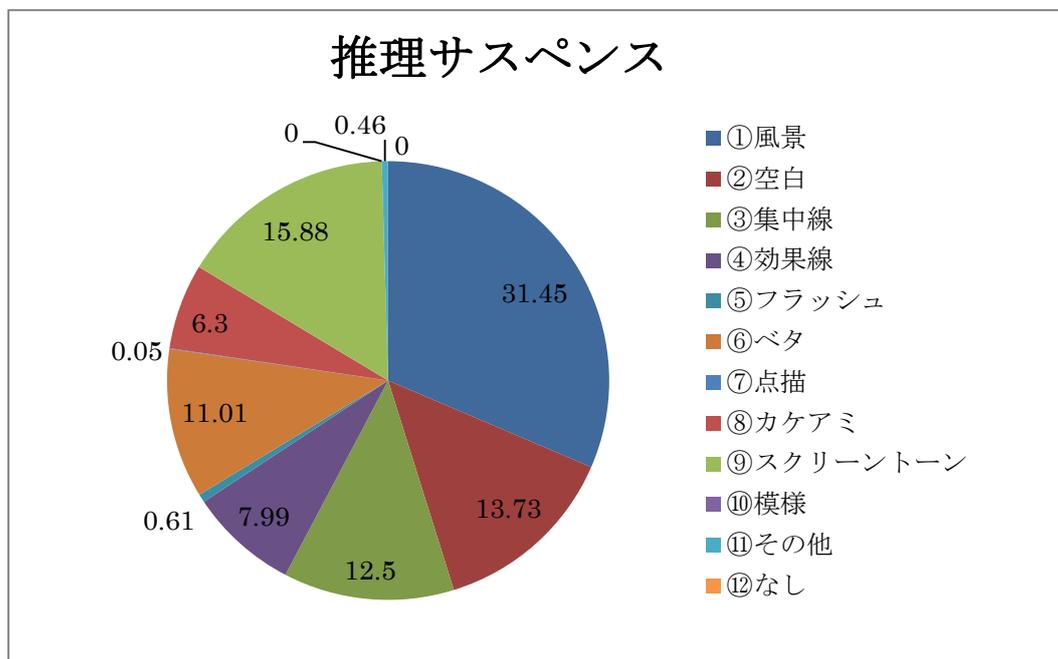


図 2.3 推理サスペンス漫画と背景の割合集計

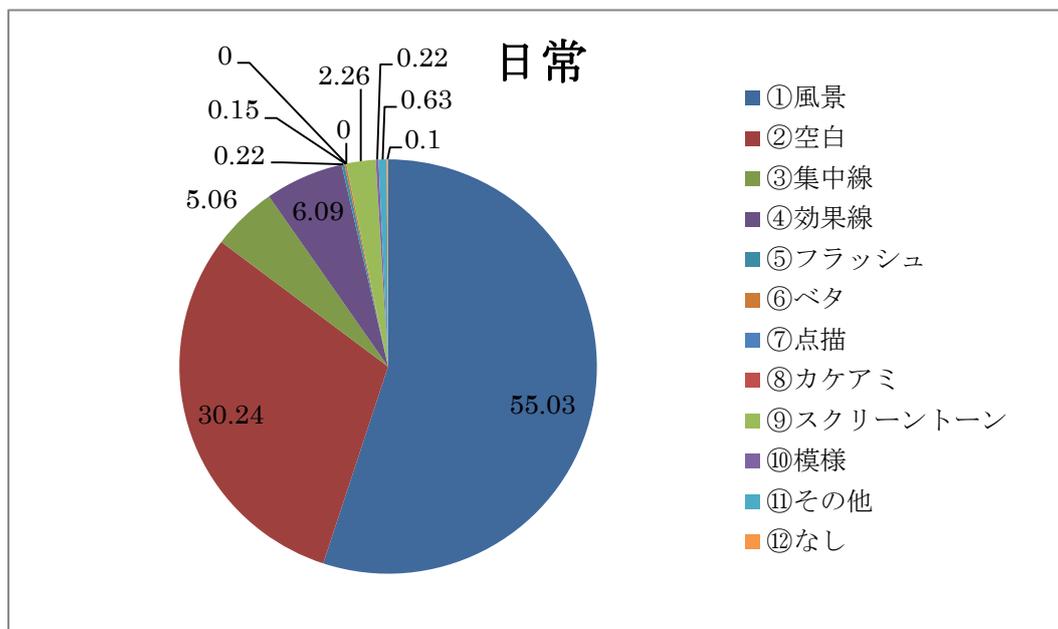


図 2.4 日常漫画と背景の割合集計

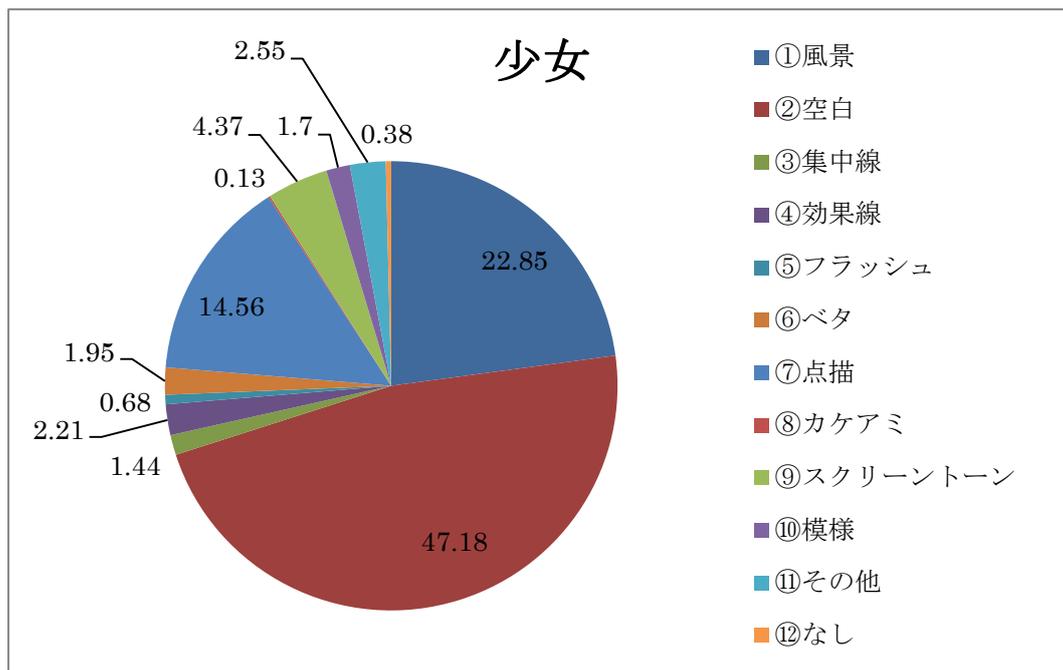


図 2.5 少女漫画と背景の割合集計

2.4 ジャンルごとの背景の特徴

集計結果である表 2.4 から読み取れる特徴をまとめ、表 2.5 に示す。

表 2.5 ジャンルと特徴

ジャンル	特徴
バトル	平均的
スポーツ	他のジャンルと比較して、集中線や効果線の割合が高い。スピード感や臨場感を出す場面が多いためであると考えられる。
推理サスペンス	他のジャンルと比較して、ベタやカケアミの割合が高い。シリアスな場面が多く、おどろおどろしい様子を表すためと考えられる。
日常物	背景もしくは空白が背景に使われている割合が全体の 85%以上を占めている。また、全体の総コマ数も比較的多い。
少女	他のジャンルでほとんどなかった点描が表れており、空白が背景に使われている割合も高い。

表 2.4, 表 2.5, 図 2.1～図 2.5 より, 背景によって, ジャンルによって異なる特徴を持つことが確認できた. そのため, 背景のパターンを手がかりとしたコミックのジャンル分類の実現の可能性があると考えられる.

2.5 むすび

本章では, コミックにおけるジャンルごとの背景の割合集計について述べた. まず, コミックにおける背景のパターンを示し, その後, コミックのジャンルとその例を挙げた. 次に, 割合集計を行った対象作品について述べ, 最後に, ジャンルごとの背景の割合集計の結果を示した. その結果ジャンルごとに異なる特徴を持つことが確認できた.

第3章 畳み込みニューラルネットワークを用いた背景分類

本章では、背景のパターン分類をする際に利用する畳み込みニューラルネットワークの概念や用語の定義について説明を行う。

3.1 まえがき

畳み込みニューラルネットワークとはディープラーニングの一つである。ディープラーニングとは、ニューラルネットワークを多層に積み上げた機械学習の方法論であり、深く大規模な構造を備え、特徴量を自動で作ることが最大の特徴である。従来の画像認識問題では、図 3.1(a)のように画像から HOG などの特徴を取り出し、その特徴量からサポートベクターマシンなどにより分類するという 2 段階構成を取る。

ディープラーニングは、特徴そのものを学習可能にし、従来の方法の弱点であった過学習を防ぎ、汎化性能の向上を実現した。その結果、未知データに対しても高い識別率を持つ利点がある。そのため現在では音声認識や画像認識、自然言語処理といったあらゆる分野で研究されている[8]。

畳み込みニューラルネットワークを説明するにあたり必要なニューラルネットワークの概念を先に述べる。

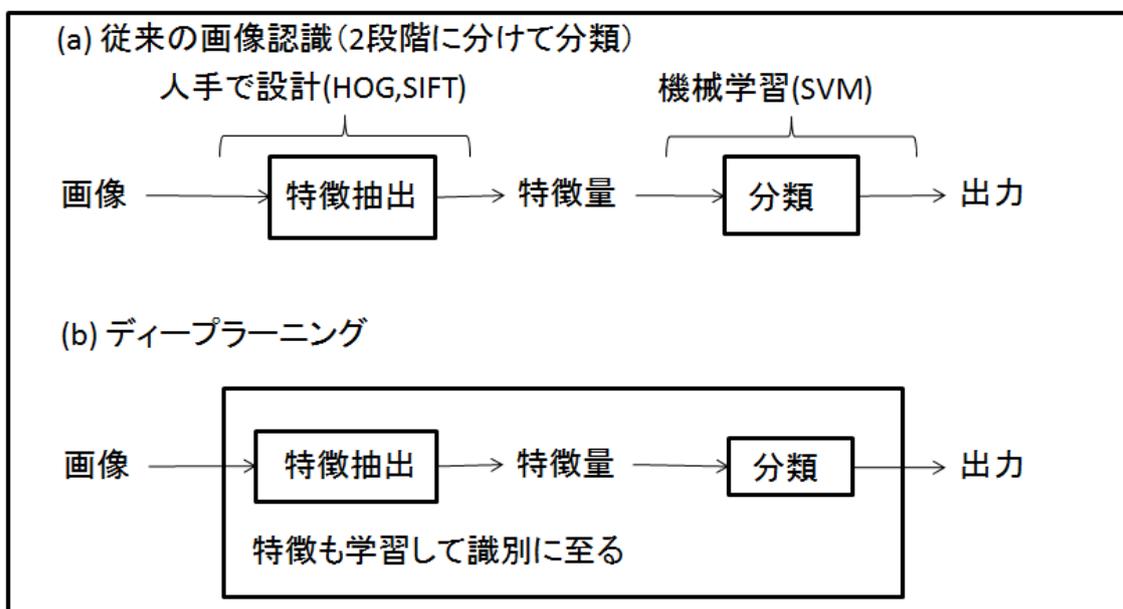


図 3.1 画像認識における従来の方法とディープラーニングの違い

3.2 ニューラルネットワーク

ニューラルネットワークとは，人間の脳神経をモデルとした情報処理システムである．コンピュータは計算など単純な処理を高速に行うことに優れている一方，人間にとっては容易い物体認識などは苦手な分野とされている．このような問題に対し，人間の脳で物体認識の際に，行われている処理メカニズムをコンピュータ上で実現しようとしたものである．

3.2.1 人間の脳のシステム

ニューロンという神経細胞が集まり脳は構成されている．ニューロン同士が互いに結合されて巨大なネットワークを作ることによって，活動している．またニューロンの構成要素と役割の関係は

- | | | |
|-----|------|----------------|
| (1) | 細胞体 | ニューロンの本体 |
| (2) | 樹状突起 | 他のニューロンからの情報入力 |
| (3) | 軸索 | 他のニューロンへの情報出力 |
| (4) | シナプス | 脳の神経細胞の接続 |

である．その構成を図 3.2 に示す[9]．

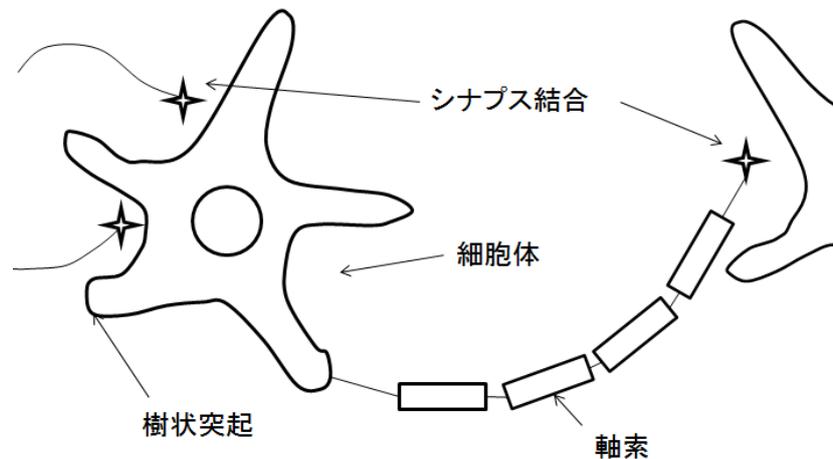


図 3.2 神経細胞の構造

3.2.2 ニューロンのモデル化

ニューロンの基本的な働きは，入力信号を受け取り一定の閾値を超えたときに出力信号を出すことである．図 3.2 に基本的な構造を示す．

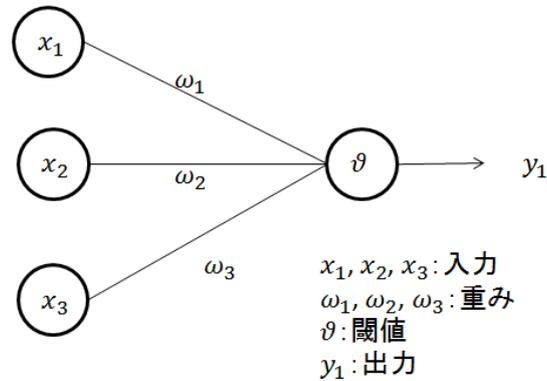


図 3.2 ニューロンの基本的構造

この動作について説明すると 1~ i 番目の入力信号を x_i ($i=1,2,3,\dots,n$)，それぞれの重みを ω_i ($i=1,2,3,\dots,n$) としたとき，他のニューロンからの入力信号の総和は

$$\sum_{i=1}^n x_i \omega_i \quad (3.1)$$

となる．受け取った信号が閾値 ϑ を超えていれば他のニューロンに対して信号を出力するので出力 y は

$$y = f\left(\sum_{i=1}^n x_i \omega_i - \vartheta\right) \quad (3.2)$$

となる．ここで関数 f はステップ関数である．つまり受け取った信号が閾値 ϑ を超えていれば 1 を出力し，超えていなければ 0 を出力する．

ステップ関数を図 3.3 に示す．

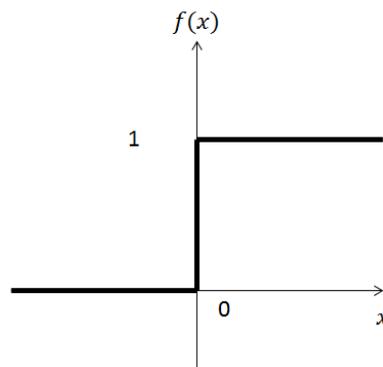


図 3.3 ステップ関数

3.2.3 単純パーセプトロン

単純パーセプトロンの構造を図 3.4 に示す。単純パーセプトロンは入力層，中間層，出力層の 3 層で構成され， n 層目の出力が， $n+1$ 層目の入力となるように信号が連結される。

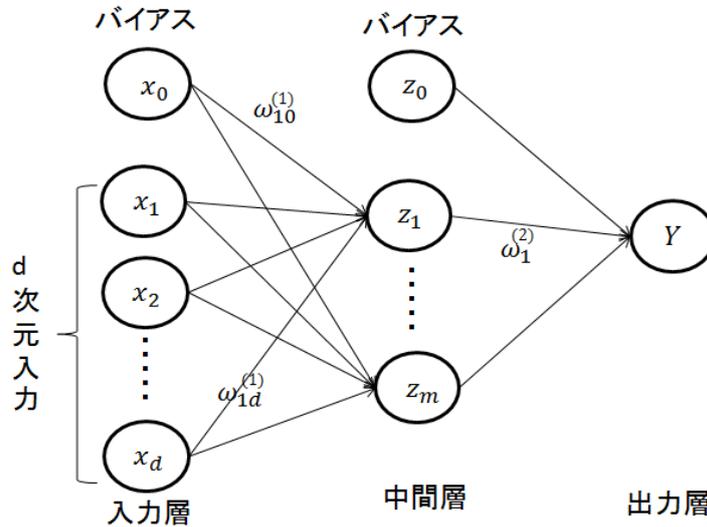


図 3.4 単純パーセプトロンの例

ここで， d 次元の情報が入力されたとき，初期値としてバイアス $x_0 = 1$ を設定する。バイアスとは関数の切片であり，誤差逆伝播法により，重みやバイアスを更新して，より精度の高いものにする効果を持つ。

このとき $z_j (j = 0, 1, \dots, m)$ への入力情報の総和は式(3.1)より

$$\sum_{i=0}^d x_i \omega_{ji}^{(1)} \quad (3.3)$$

となる。このとき肩付き数字 (1) は 1 層目を意味する。さらに隠れ層における出力 z_j は

$$z_j = h \left(\sum_{i=0}^d x_i \omega_{ji}^{(1)} \right) \quad (3.4)$$

となる。式(3.2)ではステップ関数を用いたが，ここでは非線形関数であるジグモイド関数 h を利用する。

この理由は，ジグモイド関数は連続関数であるので，パラメータに関して微分可能であり高速な学習が可能となるためである。また，微分したときの計算が簡単になる。ジグモイド関数は式(3.5)に，その微分式を式(3.6)に，ジグモイド関数を図 3.5 に示す。

$$h(a) = \frac{1}{1 + e^{-a}} \quad (3.5)$$

$$\frac{dh}{da} = h(a)(1 - h(a)) \quad (3.6)$$

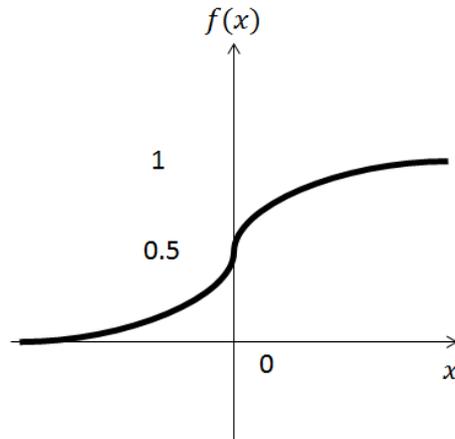


図 3.5 ジグモイド関数

また，出力層における y は

$$Y = h\left(\sum_{j=0}^m z_j \omega_j^{(2)}\right) \quad (3.7)$$

$$= h\left(\sum_{j=0}^m \omega_j^{(2)} h\left(\sum_{i=0}^d x_i \omega_{ji}^{(1)}\right)\right) \quad (3.8)$$

となる．

このとき，出力 Y と教師信号 T を用いて，式(3.9)と式(3.10)により，重みとバイアスを更新する．

$$\omega_{t+1} = \omega_t + \eta(T - Y)z_j \quad (3.9)$$

$$\vartheta_{t+1} = \vartheta_t + \eta(T - Y) \quad (3.10)$$

η は学習係数であり $0 \leq \eta \leq 1$ で定義する． t は更新回数を示しており，学習の終了条件を満たすまで学習を繰り返す．

3.2.4 多層パーセプトロン

多層パーセプトロンは、非線形の高クラス識別器である[10]。その構造を図 3.6 に示す。

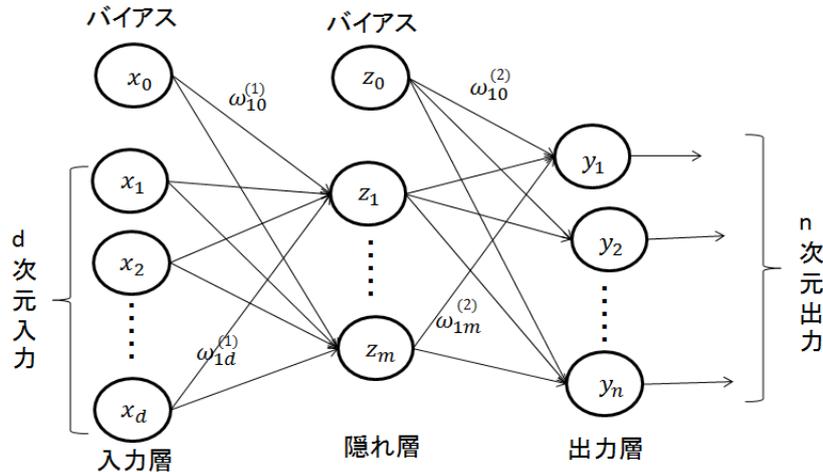


図 3.6 多層パーセプトロンの例

単純パーセプトロンと多層パーセプトロンの主な違いは次の 2 点である。

- 多層パーセプトロンではすべての重みに対して更新を行う。
- 多層パーセプトロンでは多クラスの識別を行うため出力層のユニット数はクラス数の分だけ用意する。

前者において、単純パーセプトロンは中間層と出力層のみ更新を行っていたが、多層パーセプトロンは、隠れ層に隠れているすべての層で更新を行う。

また後者において、最後の出力にはクラス分類に用いるクラス数と同じ数だけのユニットを配置し、これらに対しての入力にジグモイド関数ではなくソフトマックス関数を用いて出力とする。ソフトマックス関数を式(3.9)に示す。(j = 1, 2, ..., n)

$$p_j = \frac{e^{x_j}}{\sum_{k=1}^n e^{x_k}} \quad (3.11)$$

3.2.5 誤差逆伝播法(back propagation)

誤差逆伝播法は、出力層から入力層にかけて誤差の勾配を逆伝播させ、重みとバイアスを更新する学習アルゴリズムである。

教師信号を用意して、出力信号と教師信号の誤差から、より高い精度をもつ学習器にしようとする目的で用いる。

p 番目の教師信号を t_p ($p=1,2,\dots,n$) としたとき出力信号と教師信号の近さを表す尺度 (エラー率) として

$$E = \frac{1}{2} \sum_i^{L_n} (t_{pi} - y_i^N)^2 \quad (3.12)$$

が成り立つ。

ここで、 L_n は第 n 層のニューロンの数、 y_i^N とは第 N 層の i 番目の出力である。式(3.12)の E が 0 に近づくほど、出力信号が教師信号に近づいていることを表している。

そのため誤差逆伝播法では、 E を 0 に近づけるために重みやバイアス値を変更する。具体的にどのように誤差信号を求め、重みを更新するのか、その手順を述べる[11]。

(手順 1) 初期値として、すべての重みを乱数によって-0.1~0.1 程度の範囲の小さな値に設定する。学習率 η ($0 \leq \eta \leq 1$)を設定する。

(手順 2) 入力信号 x_{pi} ($1 \leq i \leq L_1$) をネットワークに入力する。

(手順 3) 入力層から出力層に向けて、各ニューロンの計算をする。

(手順 4) N 層における出力 y_j^N と教師信号 t_{pj}^N の誤差から、以下の式 (3.13) 式を使い誤差信号 δ_j^N を計算する。

$$\delta_j^N = -(t_{pj}^N - y_j^N)y_j^N(1 - y_j^N) \quad (3.13)$$

(手順 5) 式 (3.13) により求められた誤差信号 δ_j^N を使い、中間層の誤差信号 δ_j^n ($n < N$)を計算する。

$$\delta_j^n = \left\{ \sum_{k=1}^{L_{n+1}} \delta_k^{n+1} \omega_{k,j}^{n+1,n} \right\} y_j^n (1 - y_j^n) \quad (3.14)$$

により求まる。

(手順 6) 手順 4, 手順 5 で求めた δ_j^n を使い、重みを更新する。

$$\Delta \omega_{ji}^{n,n-1} = -\eta \delta_j^n y_j^{n-1} \quad (3.15)$$

(手順 7) すべてのパターンに対する二乗誤差 E が十分に小さくなった時点で、学習は終了する。またそうでなければ手順 2~6 を繰り返す。

3.3 ディープラーニング

第 3 章の冒頭でも述べたように，ディープラーニングは未知データに対しても高い識別率を持つ．ここからはディープラーニングの一つであり実験に用いる畳み込みニューラルネットワークの基本的な構造を述べる．

3.3.1 畳み込みニューラルネットワークの構造

畳み込みニューラルネットワークは畳み込み層とプーリング層と呼ばれる 2 層が交互に積み重ねられ，その後通常ニューラルネットワークつまり全結合層を配置する基本構造を持つ (図 3.6) [8][12]．

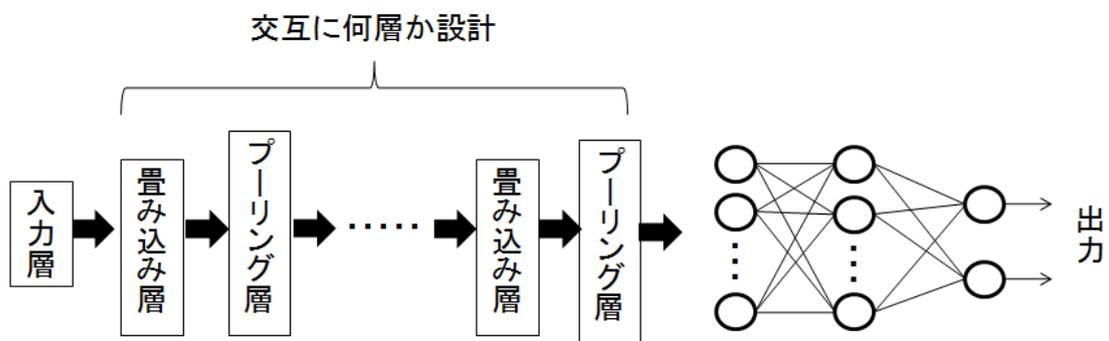


図 3.6 畳み込みニューラルネットワークの基本構造

畳み込み層とプーリング層の構造を除けば，通常ニューラルネットワークと同様の構造となる．各層の組み合わせや階層の数を変えることで，画像認識の精度が変わる．畳み込み層とプーリング層だけでなく，整流層や正規化層をオプションとして組み合わせることで，対象にあった識別器を設計することが出来る．

また，畳み込みネットワークも通常ニューラルネットワークと同様に 3.1.5 で述べた誤差逆伝播法による学習が可能である．

通常ニューラルネットワークでは，隠れ層のすべてのユニットに誤差が伝播され，ユニット間の接続が多いため 1 ユニットの出力がネットワーク全体に影響を及ぼしてしまうという問題がある．つまり微小な誤差でも処理効率が悪く，柔軟性や汎化性に欠けるといふ欠点がある．

3.3.2 畳み込み層

畳み込みの役割は以下の 2 つである。

- ・画像をぼかす（平滑化フィルタ）
- ・エッジ検出を行う（特徴抽出フィルタ）

畳み込みの具体的な手順を図 3.7 に示す。図 3.7 では 2 つの領域の計算しか示していないが、このような計算を全ての領域で行う。また緑の線部はフィルタの範囲から外れてしまう領域もあるため、適切な処理方法を決める必要がある。

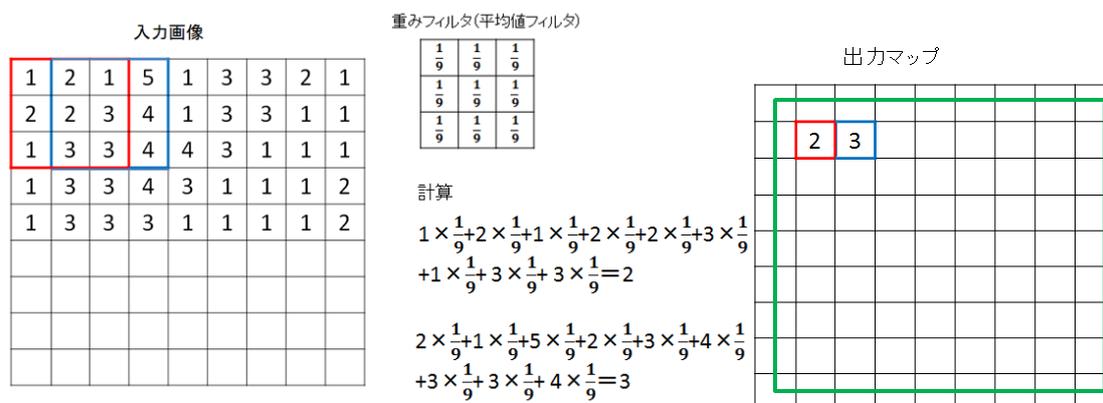


図 3.7 畳み込みの手順

図 3.7 の重みフィルタを誤差逆伝播法により更新することで、エラー率が低下し、識別器の精度が向上する。また、フィルタを変えることで、さまざまな効果がある。フィルタには以下のようなものがある[13]。

- ・平滑化フィルタ

(1) 平均値フィルタ

注目している画素とその近傍の画素の濃度値の平均値を、注目している画素の新しい濃度値とする方法。(図 3.7 のフィルタ)

(2) メディアンフィルタ

領域内の濃度の中央値、例えば 3×3 の領域であれば、9 個の濃度値を低い順番に並べ、5 番目の濃度値をその中心の新しい濃度値とする方法。

- ・エッジ検出フィルタ

(1) MAX-MIN フィルタ

MAX-MIN フィルタとは、 3×3 の近傍領域の中で、濃度値の最大値と最小値を取り出して、その差を中心の新しい濃度値とする方法。

次に畳み込み層において重要となる用語について述べる[14].

- ・特徴マップ

図 3.7 で示す出力マップを特徴マップと言う. この出力されたマップを入力として, プーリング処理などを行ったとき, また新たな特徴マップを得ることが出来る.

- ・局所的受容野

誤差がすべてのユニットに伝播することで, 柔軟性や汎化性にかける問題があるため, ある層のユニットへ接続する前層のユニットの数を制限する.

- ・共有重み

画像が平行移動した場合にも同様の認識できるようにするため, 局所受容野における重みを他の局所受容野と共有する.

3.3.3 プーリング層

プーリング層の役割は、以下の2つである。

- ・ユニット数を減らし、調整するパラメータを減らす。
- ・局所的な平行移動性を確保し、位置に関して微小な揺らぎがあった場合でも安定した検出が可能である。

画像認識ではマックスプーリングが最も一般的である。これは入力データを一定サイズの矩形領域で区切り、この分割領域は情報の重複取り込みとなるためオーバーラップさせないで、各領域の最大値を出力する。解像度を下げる効果を持つ。(図 3.8) (図 3.9)

$$\text{出力} = \max(0, \text{入力}) \quad (3.16)$$

図 3.8 では 4×4 のユニットを 2×2 のユニットに調整している。

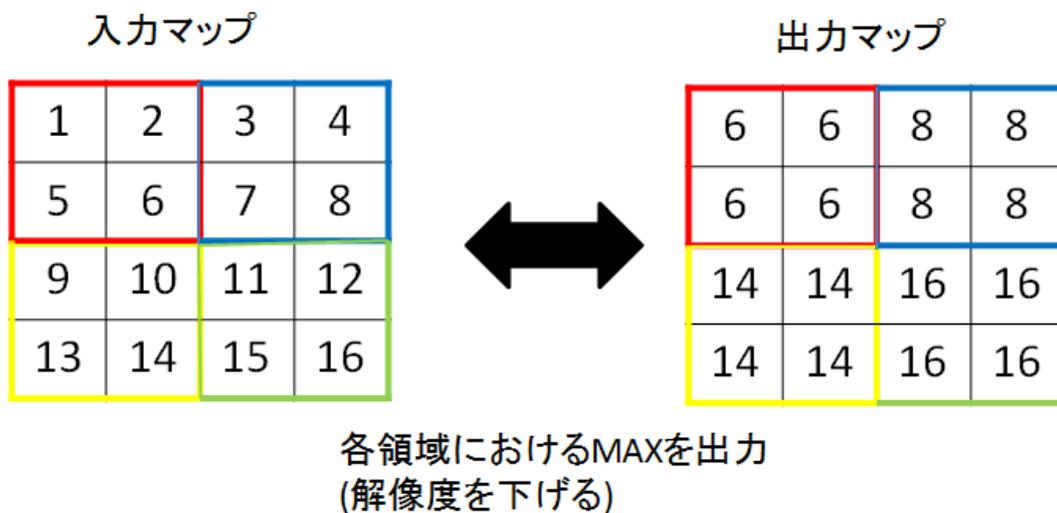


図 3.8 マックスプーリング処理の流れ

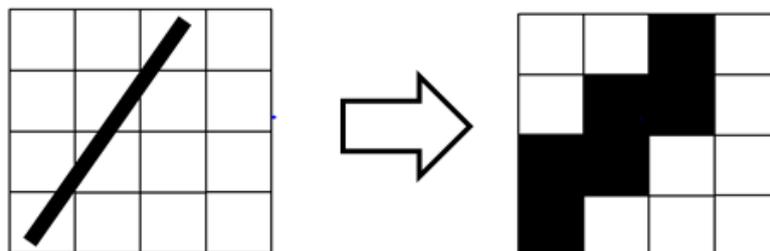
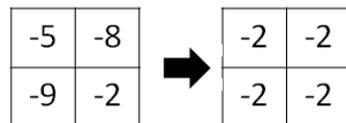


図 3.9 プーリング処理の流れ 2

3.3.4 整流層

畳み込み層の出力に対し、負数を除去する関数を持つ層である。畳み込み層とプーリング層の間に整流層を挟むことで、以下の図のような小領域に対しても負数を出力せず、0を出力する処理となる。

畳み込み層→プーリング層



畳み込み層→整流層→プーリング層

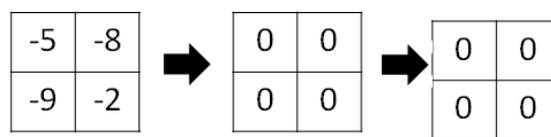


図 3.10 整流層の有無による違い

3.3.5 正規化層

正規化とはある一定の基準に従い、入力データを変形し汎化させる処理である。同一特徴マップにおける局所領域内で正規化する方法や、同一位置における異なる特徴マップ間で正規化する方法などがある。

3.4 むすび

本章では、畳み込みニューラルネットワークを用いた背景分類について述べた。まず、ニューラルネットワークにおける基本的な概念や用語について説明した。次に、畳み込みニューラルネットワークの特徴である畳み込み層、プーリング層、整流層、正規化層の役割について説明した。

第 4 章では、本章の内容に基づき、畳み込みニューラルネットワークを用いた背景分類の評価実験を行う。

第4章 実験概要と結果, 考察

4.1 まえがき

本章では手動で切り出したコミックの背景を畳み込みニューラルネットワークで学習させた場合, 入力画像を背景のパターン別に分類できるかどうか評価実験を行う.

4.2 実験概要

コミックにおけるコマごとの背景画像を抽出し, 学習を行う. その後, 完成した識別器を用いて, 学習サンプルに使用していない未知画像に対して, 正しく分類ができるか検証を行う. 学習画像は 5000 枚、評価サンプルは 2000 枚を使用した.

・学習画像

第2章で述べた背景パターン 10 種類を各 100 枚ずつ用意し, さらにその各画像から 5 枚ずつ切り出して学習データを得た. 合計で $10 \times 100 \times 5 = 5000$ 枚のデータを用いた.

4.3 実験条件

学習画像の条件は以下の 2 つを満たしているものとする.

1. 10 種類のうち必ずどれかに該当するもの (わかりにくいものを除く)
2. 背景があるもの

学習対象の例を図 4.1 に, 学習対象でない例を図 4.2 に示す.



図 4.1 学習対象の例

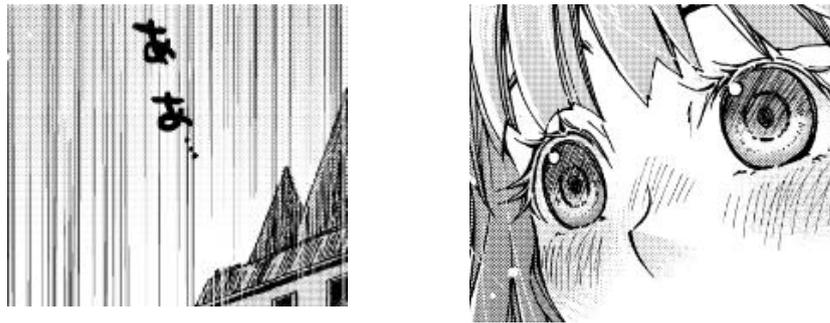


図 4.2 学習対象でない例

(左 効果線か雨かの区別が容易ではない場合, 右 背景がない場合)

次に実験で用いる畳み込みニューラルネットワークの構造を図 4.3 に示す.

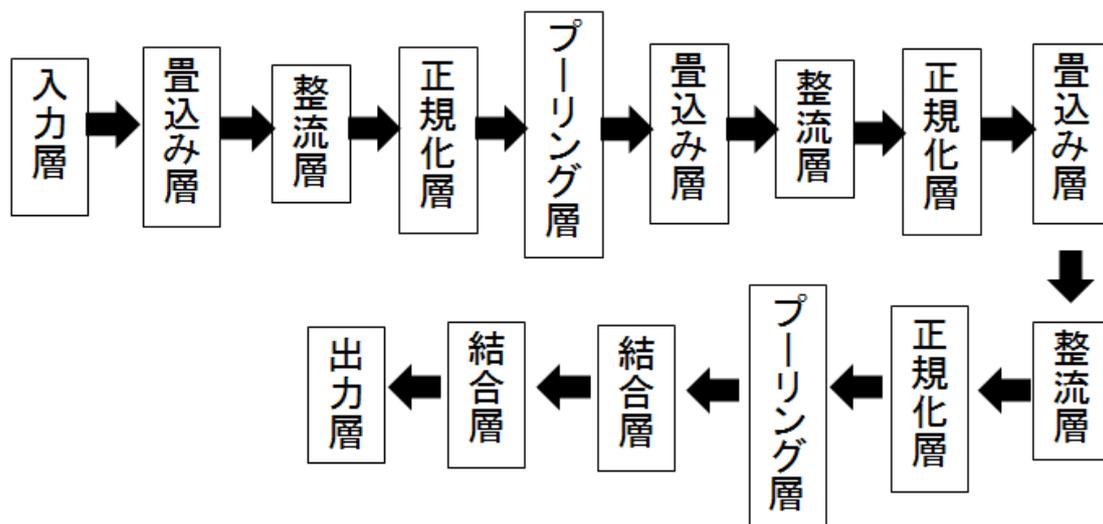


図 4.3 実験で用いる畳み込みニューラルネットワークの構造

4.4 実験結果

全体の認識率とエラー率を表 4.1 に、パターンごとの認識率とエラー率を表 4.2 に、実際の実行結果のスクリーンコピーを図 4.3 に示す。

認識率とエラー率を導出する式をそれぞれ式 (4.1)、式 (4.2) に示す。

$$\text{認識率} = \frac{\text{正解画像数}}{\text{全画像数}} \quad (4.1)$$

$$\text{エラー率} = \frac{\text{全画像数} - (\text{正解画像数} + \text{分類できない画像数})}{\text{全画像数}} \quad (4.2)$$

全体の認識率は 66.7%，エラー率は 19.3%とある程度の精度にとどまった。背景のパターンによって認識率とエラー率にばらつきがあることが分かった。また、背景が空白と点描の場合に際立って、スクリーントーンと誤認識されることが多かった。

表 4.1 全体の認識率とエラー率

認識率	エラー率
66.7%	19.3%

表 4.2 パターンごとの認識率とエラー率

背景	認識率	エラー率
0.風景	85.6%	0.065%
1.空白	32.0%	57.5%
2.集中線	73.0%	14.5%
3.効果線	56.4%	27.5%
4.フラッシュ	85.6%	0.07%
5.ベタ	69.0%	14.5%
6.点描	33.0%	40.0%
7.カケアミ	67.5%	0.09%
8.スクリーントーン	90.5%	0.045%
9.模様	74.8%	12.0%

```

100%
*****
threshold : 0
  0      0      1      2      3      4      5      6      7      8      9
0      171     0      1      1      0      2      0      2      0      5
1      0      64     2      2      0      1      2      0      1      0
2      2      5      143    3      1      2      0      2      1      0
3      1      1      3      114   0      0      3      0      0      3
4      0      0      0      0      169  0      0      0      0      0
5      0      0      0      0      0      138  1      1      0      0
6      5      7      2      1      0      2      66     1      2      1
7      0      0      0      1      1      0      0      139    2      1
8      5      102    20     45     12    20     73    12    181    14
9      0      0      1      2      0      2      1      0      3      148
10     16     21     24     33     15     33     54     49     10     26
RR    0.855000  0.320000  0.729592  0.564356  0.853535  0.690000  0.330000  0.674757  0.905000  0.747475
ER    0.065000  0.575000  0.145000  0.275000  0.070000  0.145000  0.400000  0.090000  0.045000  0.120000
total recognition rate: 0.6665
total error rate: 0.193
*****

```

図 4.4 実行結果のスクリーンコピー

4.5 考察

3章の3.3.3節における実験結果では、全体の識別率が66.7%、エラー率が19.3%となった。全体では高い識別率、低いエラー率を得ることができたとは言えないが、背景のパターンによっては高い認識率と低いエラー率を得ることができている部分を確認された。

具体的には背景が風景の場合、フラッシュの場合、スクリーントーンの場合の3パターンでは、認識率高くエラー率が低い。この3パターンの分類がうまくいった要因として、エッジ検出の対象となるエッジが多いことが考えられる。入力層の次の畳み込み層で白と黒の境目となるエッジを検出した際に、エッジとなる輪郭が多いため、多くの特徴が抽出できたのではないかと考えられる。図4.5にエッジ検出の例を示す。

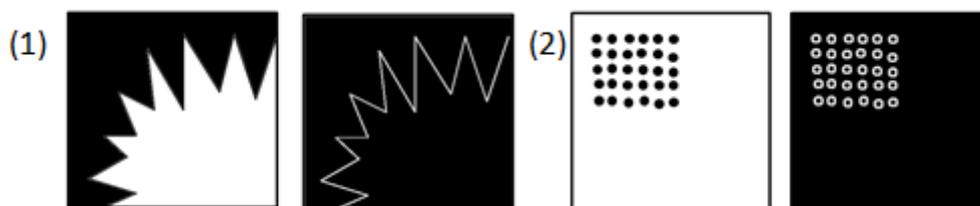


図 4.5 画像とエッジ検出（背景が(1)フラッシュの場合、(2)スクリーントーンの場合）

逆に、背景が空白の場合と点描の場合の識別率が低くエラー率が高いという結果になった。この2パターンがうまくいかなかった要因として、エッジとなる輪郭が少ないため、特徴の抽出が少なくなってしまうと考えられる。図4.6に特徴抽出が困難なエッジ抽出の例を示す。

また、この際、実験結果の図4.4から分かるように、背景が空白の場合と点描の場合においてスクリーントーンと誤認識される場合が多かった。これは、背景が空白の場合にノイズが発生し、スクリーントーンの点だと誤認識されたと考えられる。また点描の場合、点描という表現は多くの点が集まって一つの模様となっているため、その点を単体で見たとき、スクリーントーンと誤認識される傾向があると考えられる。



図 4.6 画像とエッジ検出（背景が空白）

第5章 結論と今後の課題

5.1 結論

畳み込みニューラルネットワークを使ったコミック画像の背景の認識率は 66.7%，エラー率は 19.3%で行えることが分かった。

今後、分類精度がさらに向上させることができれば、コミックのジャンル分類の自動化の実現に一步近づくと考えられる。

5.2 今後の課題

今後の課題として以下の項目がある。

・分類精度向上の検討

本研究では、背景画像の認識率が 66.7%，エラー率が 19.3%とある程度の認識率にとどまった。この精度では第 2 章の統計結果と照らし合わせてジャンルを決定するのに十分ではない。そのため精度向上のため検討が必要となる。具体的には学習データを増やすことや畳み込みニューラルネットワークの構造の再検討などがある。

畳み込みニューラルネットワークは全体の層数やオプションとなる層をどこに組み込むかによって、精度は著しく異なるため背景分類に適した構造を検討する必要がある。

・作者やコミックの巻数における背景割合の分散の考慮

本研究では、ジャンルごとに背景の集計を行ったがジャンルごとに 3 冊ずつしか集計しておらず、作者の違いにおける背景パターンの分散を考慮していない。また、ランダムに巻数を選んでおり、巻数における分散も考慮していない。そのため作者や巻数における分散の考慮が必要となる。

・背景自動抽出の検討

本研究では、背景画像の切り出しを手動で行い、機械学習させた。コミックのジャンル分類の自動化を実現するには、コマごとの背景を自動で選別できなくてはならない。

現在、電子コミックにおけるコマの認識、キャラクター検出、吹き出し検出などの研究が行われている[1]。それを利用し、各コマからキャラクターと吹き出しを除いたものが背景となるなど背景抽出のため新たな発想が必要である。

謝辞

本研究を進めるにあたり，丁寧かつ熱心なご指導を頂いた渡辺教授に感謝いたします。

日頃からアドバイスや議論をしていただいた研究室の皆様感謝いたします。特に多くの知識や示唆を頂いた張傑先輩に感謝いたします。

最後に私をここまで育ててくださった家族に感謝いたします。

参考文献

- [1] インプレスビジネスメディア, “電子書籍ビジネス調査報告書 2014” 株式会社インプレス
ビジネスメディア, 2014
- [2] 中野晴行 “電子コミックの未来はどこに”,
http://www.dotbook.jp/magazine-k/2014/10/15/future_of_e-comics/, 2014
- [3] 夜宵草, 「ReLIFE」, (comico,2013)
- [4] 松下光範, “コミック工学の可能性”, 第 2 回 ARGWEB インテリジェンスとインタラク
ション研究会, pp. 63-68, 2013.
- [5] 高林 未央, “漫画の技法「背景」を用いた制作と鑑賞の授業に関する研究”, 美術教育学 :
美術科教育学会誌 30, pp. 217-228, 2009.
- [6] yahoo, “Yahoo の知恵袋漫画のジャンル分けについて”,
http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1447326859, 2010
- [7] 尾田栄一郎, 「ワンピース」第 1 巻, (集英社, 1997)
大高忍, 「マギ」第 8 巻, (少年サンデーコミックス, 2011)
鳥山明, 「ドラゴンボール」第 3 巻, (集英社, 1986)
高橋陽一, 「キャプテン翼 W ユース編」第 1 巻, (集英社, 1994)
安田剛士, 「オーバードライブ」第 17 巻, (講談社, 2008)
満田拓也, 「メジャー」第 65 巻, (講談社, 2007)
小畑友紀, 「僕等がいた」第 1 巻, (フラワーコミックス, 2002)
青木琴美, 「僕は妹に恋をする」第 1 巻, (フラワーコミックス, 2003)
青木琴美, 「僕の初恋を君に捧ぐ」第 1 巻, (フラワーコミックス, 2005)
秋本治, 「こちら亀有区亀有公園前派出所」第 61 巻, (集英社, 1989)
臼井義人, 「クレヨンしんちゃん みんなの笑顔が金メダル! オラリンピック編」, (双葉社,
2012)
やまさき十三, 北見けんいち 「釣りバカ日誌」第 41 巻, (小学館, 1996)
高山みなみ, 「名探偵コナン」第 1 巻, (少年サンデーコミックス, 1994)
さとうふみや, 天樹征丸, 「金田一少年の事件簿 錬金術殺人事件 上」, (講談社, 2010)
- [8] 岡谷貴之, “ディープラーニング (技術解説)”, 映像情報メディア学会誌 : 映像情報メデ
ィア 68(6), 466-471, 2014-06-01
- [9] “神経細胞について”,
<http://www.tuat.ac.jp/~tuatmcc/contents/monthly/200208/index.xml>, 2002
- [10] 福井宏, “Convolutional Neural Network の特長抽出過程における不変性の獲得の調査”
中部大学工学部情報工学科卒業論文, 2014
- [11] 村上, 泉田研究室 “誤差伝播法について” ,
<http://ipr20.cs.ehime-u.ac.jp/column/neural/chapter6.html>, 2001
- [12] 関野雅則, 木村俊一, 越裕, 富士ゼロックス株式会社, “視覚情報処理モデルに基づいて

改良した畳み込みニューラルネットワーク文字認識”, The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013

[13] 村上, 泉田研究室, “空間フィルタリング処理”,

http://ipr20.cs.ehime-u.ac.jp/column/gazo_syori/chapter5.html, 2001

[14] “Convolutional Neural Network”,

<http://ceromondo.blogspot.jp/2012/09/convolutional-neural-network.html> , 2012

図一覧

図 2.1	バトル漫画における背景の割合集計	6
図 2.2	スポーツ漫画と背景の割合集計	6
図 2.3	推理サスペンス漫画と背景の割合集計	7
図 2.4	日常漫画と背景の割合集計	7
図 2.5	少女漫画と背景の割合集計	8
図 3.1	画像認識における従来の方法とディープラーニングの違い	10
図 3.2	神経細胞の構造	11
図 3.3	ニューロンの基本的構造	12
図 3.4	ステップ関数	12
図 3.5	単純パーセプトロンの例	13
図 3.6	ジグモイド関数	14
図 3.7	多層パーセプトロンの例	15
図 3.8	畳み込みニューラルネットワークの基本構造	17
図 3.9	畳み込みの手順	18
図 3.10	マックスプーリングの処理の流れ	20
図 3.11	マックスプーリング処理の流れ 2	20
図 3.12	整流層の有無による違い	21
図 4.1	学習対象の例	22
図 4.2	学習対象でない例	23
図 4.3	実験で使う畳み込みニューラルネットワークの構造	23
図 4.4	実行結果のスクリーンコピー	25
図 4.5	画像とエッジ検出 (背景が(1)フラッシュ、(2)スクリーントーン)	26
図 4.6	画像とエッジ検出 (背景が空白)	27

表一覧

表 2.1	コミックにおける背景のパターンとその特徴-----	2
表 2.2	コミックのジャンルの種類とその例-----	4
表 2.3	集計を行ったジャンルと作品名-----	5
表 2.4	コミックのジャンルと背景のパターンの集計結果-----	5
表 2.5	ジャンルと特徴-----	8
表 4.1	全体の認識率とエラー率-----	24
表 4.2	パターンごとの認識率とエラー率-----	24