

## 概要

本研究では～

## Abstract

In the thesis, ～

## 謝辞

皆様，本当にありがとうございました．

# 目次

<b>1</b>	<b>はじめに</b>	<b>1</b>
1.1	研究背景	1
1.2	研究目的	1
<b>2</b>	<b>ニューラルネットワーク</b>	<b>2</b>
2.1	畳み込みニューラルネットワーク ( Convolution Neural Network ) の概要	2
2.1.1	畳み込み層	2
2.1.2	プーリング層	3
2.1.3	全結合層	4
2.1.4	CNN の学習方法	4
2.2	RNN ( Recurrent Neural Network ) の概要	4
2.3	LSTM ( Long Short - Term Memory ) の概要	6
2.4	CNN を用いた表情認識実験	6
2.4.1	CNN の構造	6
2.4.2	表情認識に用いる表情画像	6
2.4.3	実験方法	6
2.4.4	結果	6
2.4.5	考察	6
<b>3</b>	<b>chapter3</b>	<b>8</b>
<b>4</b>	<b>chapter4</b>	<b>9</b>
<b>5</b>	<b>考察</b>	<b>10</b>
<b>6</b>	<b>まとめと今後の課題</b>	<b>11</b>
6.1	まとめ	11
6.2	今後の課題	11

## 1 はじめに

### 1.1 研究背景

### 1.2 研究目的

はじめに [1]

## 2 ニューラルネットワーク

本章では、深層学習に用いられる CNN ( Convolution Neural Network ) と RNN ( Recurrent Neural Network ) , LSTM ( Long Short - Term Memory ) についての概要と、CNN を用いた静画像による表情認識実験について述べる。

### 2.1 畳み込みニューラルネットワーク ( Convolution Neural Network ) の概要

ニューラルネットワークとは、人間の脳の仕組みから着想を得たものであり、神経回路網をコンピュータ上で表現しようと作られた数理的モデルである。人間の脳はニューロンと呼ばれる神経細胞の結びつきで情報伝達や記憶の定着が行われており、その人間の脳神経系の持つ強力な学習能力を数学的に応用することにより、画像認識や音声認識などへの利用が期待され研究が行われている。

通常ニューラルネットワークは、入力層、出力層、隠れ層から構成されており、これらの層と層の間にニューロン同士のつながりを示す  $W$  があり、これを重みと呼ぶ。ニューラルネットワークの隠れ層は多層にすることができ、隠れ層が多数存在する多層構造の

ニューラルネットワークのことをディープラーニングと呼ぶ。そのディープラーニングの主力となっているものの一つに畳み込みニューラルネットワークがあり、畳み込み層とプーリング層を一つのペアとして、それらが複数回重ね合わせて構成される順方向性ニューラルネットワークである。

畳み込みニューラルネットワークの基本構造を図 2 . 1 に示す。

#### 2.1.1 畳み込み層

畳み込みニューラルネットワークに入力としての画像が与えられたとき、その画像の特徴量を抽出する操作のことを畳み込みと呼ぶ。

畳み込み層は単純型細胞をモデルに考えられたものであり、元の画像にフィルタをかけて元の画像よりも少し小さい特徴マップを出力する、またこのフィルタは画像全体をスライドするので特徴がどこにあってても抽出することができ、その際にフィルタの数だけ特徴マップが出力される。

ここで例として、対象とする画像を  $X \times Y$  Pixels の RGB の階調値とし、 $k$  番目の階調の素子 (  $i, j$  ) の画素値を  $I_{ij}^k$  とする。ただし、 $k = 1$  が R,  $k = 2$  が G,  $k = 3$  が B とする。最初の層の畳み込み層の  $a$  番目のフィルターの (  $i, j$  ) 番目の素子の内部状態  $y_{ij}^{(1)(a)}$

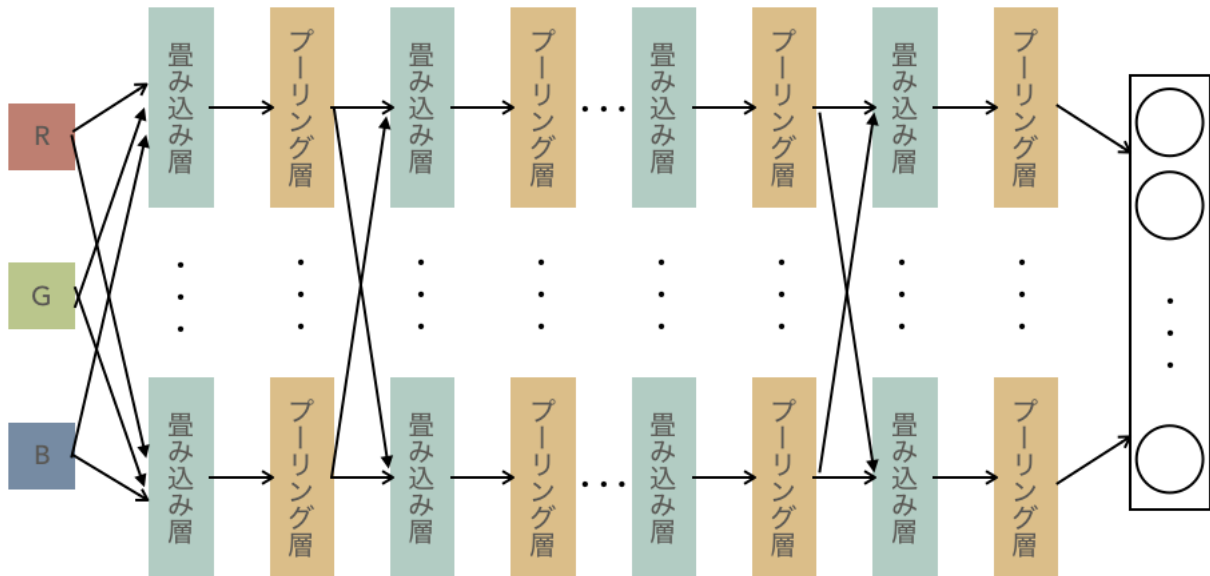


図 1: 畳み込みニューラルネットワークの基本構造

は,

$$y_{ij}^{(1)(a)} = \sum_{k=1}^3 \left( \sum_{x \in W} \sum_{y \in W} w_{ij}^{(1)(a)(k)} I_{i+x, j+y}^{(k)} + b_{ij}^{(1)(a)(k)} \right) \quad (1)$$

その出力  $\tilde{y}_{ij}^{(1)(a)}$  は,

$$\tilde{y}_{ij}^{(1)(a)} = \max \left( y_{ij}^{(1)(a)}, 0 \right) \quad (2)$$

$\ell$  番目の層の畳み込み層の内部状態  $y_{ij}^{(\ell)(a)}$  は,

$$y_{ij}^{(\ell)(a)} = \sum_{\alpha=1}^{N(\ell-1)} \sum_{x \in W} \sum_{y \in W} w_{ij}^{(\ell)(a, \alpha)} z_{i+x, j+y}^{(\ell-1)(\alpha)} + b_{ij}^{(\ell)(a)} \quad (3)$$

で表される．ここで,  $w_{ij}^{(1)(a)(k)}$  は入力層と畳み込み層間のシナプス結合加重,  $W$  は各素子が入力を受ける範囲を与える配置集合 (受容野),  $b_{ij}^{(1)(a)(k)}$  は閾値である．最終層 ( $L$ ) の内部状態  $y_k^{(L)}$  は, 前層のプーリング層の出力  $z_{ij}^{(L-1)(a)}$  との全結合として, 以下のように与えられる．

$$y_k^{(L)} = \sum_{\alpha=1}^{N(L-1)} \sum_i \sum_j w_{kij}^{(L)(\alpha)} z_{ij}^{(L-1)(\alpha)} + b^{(L)k} \quad (4)$$

これらの内部状態は活性化関数と呼ばれる非線形関数を経て, 畳み込み層の出力としてプーリング層へ伝播する．活性化関数の例として, ステップ関数, シグモイド関数, ソフトマックス関数, 恒等関数, ReLU などが挙げられる．ステップ関数は, 階段関数とも呼ばれる閾値を境にして出力が切り替わるようにする関数である．シグモイド関数は, 式で表すと以下のように表される．

$$h(x) = \frac{1}{(1 + e^{-x})} \quad (5)$$

入力した値が大きいほど 1 に近づき, 入力した値が小さいほど 0 に近づく関数である．ソフトマックス関数は, 式で表すと以下のように表される．

$$y_i = \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}} \quad (6)$$

恒等関数は, 受け取った値をそのまま出力したい場合に用いる関数である．ReLU は 0 以下の値は, 0 にして 0 以上の値の場合にそのままの値を出力するという関数である．ReLU を式で表すと以下ようになる．

$$f = \max(0, x) \quad (7)$$

畳み込み層の模式図を図 1 に示す．

### 2.1.2 プーリング層

畳み込み層は単純型細胞をモデルに考えられたものであるが, プーリング層は複雑型細胞をモデルにして作られたモデルである．

画像から重要な情報を残して縮小する方法であり, 特徴マップからとある一部分の領域を取り出し, そのうちの最大値を取り出す最大プーリング (max pooling) や, 特徴マップからとある一部分の領域を取り出し, その領域の平均値を取る平均プーリング (average pooling) などがあるが, 画像認識の分野では最大プーリングが使われていることが多い．また, プーリング層の出力を  $z_{ij}^{(1)(a)}$  とすると以下のように表される．

$$z_{ij}^{(1)(a)} = \max_{x \in W, y \in W} \tilde{y}_{i+x, j+y}^{(1)(a)} \quad (8)$$

最大プーリングと平均プーリングの例をそれぞれ図 2 と図 3 に示す．

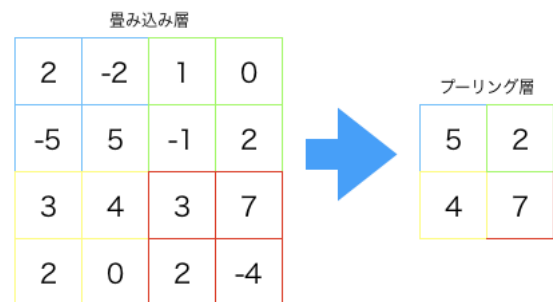


図 2: 最大プーリングの例

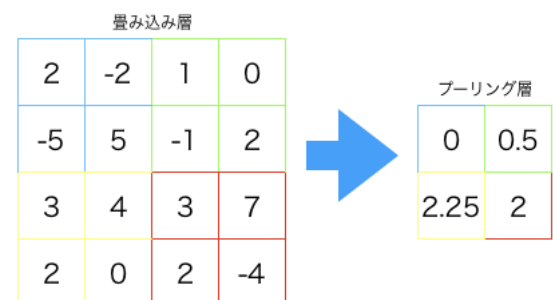


図 3: 平均プーリングの例

畳み込み層やプーリング層で行われる手法にパディングというものがある．パディングとは入力として与え

られた画像のピクセルの周りにピクセルを配置し囲むという方法である。パディングを行うことによるメリットは、畳み込み層やプーリング層を経ることによって入力画像のサイズが小さくなっていくので、パディングを行うことにより画像サイズを小さくしなくて済むという点と、畳み込み層やプーリング層を経ることにより外側から縮小されていく画像にパディングを行うことで、画像の端のピクセルも畳み込みの回数を増やすことができるので精度を高めることができるという点がある。パディングの手法の一つにゼロパディングと呼ばれるピクセルの周りをゼロのピクセルで囲むというものがある、このゼロパディングの例を図 4 に示す。

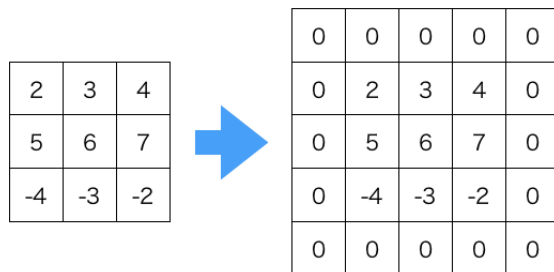


図 4: ゼロパディングの例

### 2.1.3 全結合層

畳み込み層、プーリング層を経ても入力として与えられた画像の配列のままであるが、これらの層を経て特徴データが取り出された入力として与えられた画像データを一つのノードに結合し、活性化関数によって変換された値を出力する層である。

### 2.1.4 CNN の学習方法

機械学習には教師あり学習、教師なし学習、強化学習の三つの学習法がある。教師あり学習とは、学習データに正解ラベルをつけて学習する方法である。分類や予測などは教師あり学習のクラスである。それに対して教師なし学習は、学習するデータに正解ラベルをつけずに学習する方法である。こちらの教師なし学習は、

様々な入力を読み取っていくうちに自律的に対象を認識できるようになっていく学習方法であり、教師あり学習に比べ実装難度が高い。クラスタリングや次元削減などは教師なし学習のクラスである。強化学習は価値を最大化するような行動を学習する方法である。囲碁や将棋で目先の取れるコマがあったとしてもそれが敗北につながるような手だとしたら、取らずに勝利に近づく手を打つことが強化学習の目的である。強化学習のアルゴリズムには、Q 学習、Sarsa、モンテカルロ法などがある。

Google DeepMind の開発した AlphaGo はコンピュータ囲碁プログラムであり、2015 年に人間のプロ囲碁棋士をハンディキャップ無しで破った初のコンピュータ囲碁プログラムである。膨大な数の対局データを教師あり学習として学習し、さらに強化学習を行い勝つためのパターンを特徴量として学習させており、その後コンピュータ同士を対局させ競わせていくことにより強力なコンピュータ囲碁プログラムが誕生した。

## 2.2 RNN ( Recurrent Neural Network ) の概要

ニューラルネットワークの出力を別のニューラルネットワークの入力として利用するような再帰的構造を持ったニューラルネットワークのことを RNN ( Recurrent Neural Network ) と呼ぶ。

時間を追って得られたデータを時系列データと呼び、画像などのデータが一つのベクトル  $x_n$  で表されるのに対し、 $(x(1), \dots, x(t), \dots, x(T))$  という  $T$  個のデータが一つの入力データ群このデータを用いて更に次に得られるであろうデータを予測する方法の一つである。時系列データには、気温の推移や降水量の変化などの気象データ、株価や売上の推移などが挙げられ、これらのデータは一般的なデータの集まりではなく、データの並び自体に意味を持つ。

時刻  $t$  での入力を  $x(t)$ 、入力層と隠れ層の間の重みを  $U$ 、隠れ層と出力層の間の重みを  $W$  とすると隠れ層の値  $s(t)$  とネットワークの出力  $y(t)$  は、以下のように表すことができる。

$$s(t) = f(Ux(t) + Ws(t-1)) \quad (9)$$

$$y(t) = g(Vh(t)) \quad (10)$$

また RNN の基本構造の例を図 5 に示す。

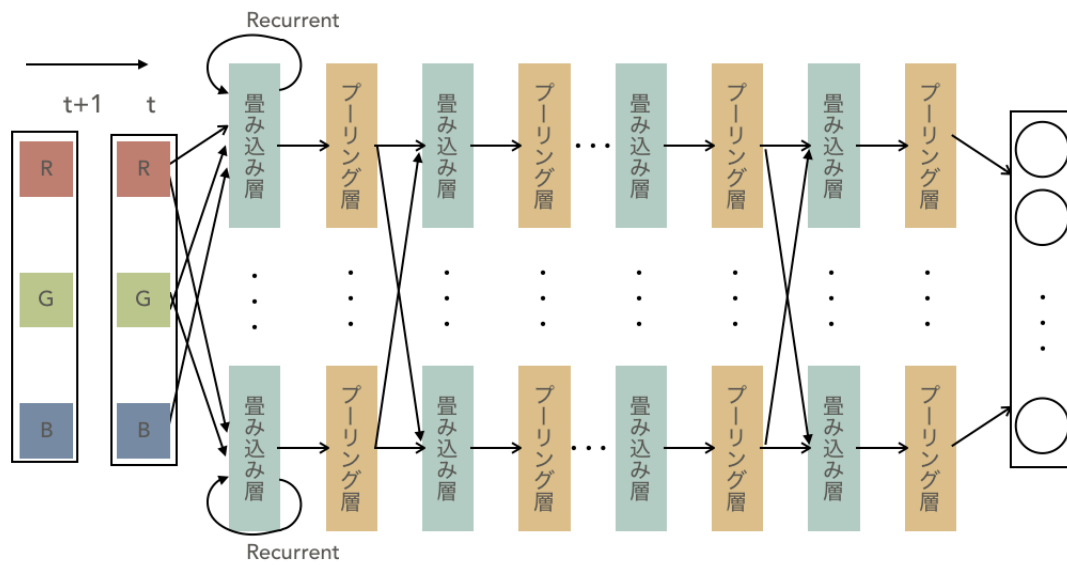


図 5: RNN の基本構造

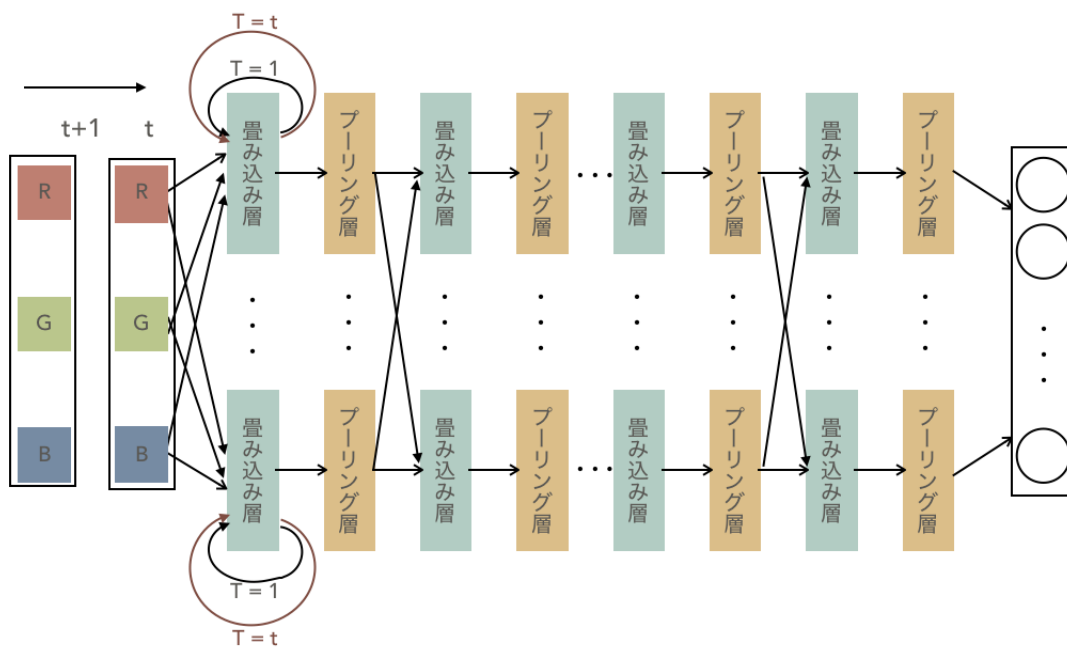


図 6: LSTM の基本構造



### 2.3 LSTM ( Long Short - Term Memory ) の概要

LSTM ( Long Short - Term Memory ) は 1995 年に RNN の拡張として 1995 年に登場した時系列データを扱うモデルである。Long Term Memory ( 長期記憶 ) と Short Term Memory ( 短期記憶 ) という神経科学における用語から取られており、従来の RNN で学習できなかった長期依存を学習できるという特徴を持つ。LSTM は RNN の中間層のユニットを LSTM block と呼ばれるメモリと三つのゲートを持つブロックに置き換えることで実現されている。畳み込み層の内部状態を以下のように表す。

$$y_{ij}^{(\ell)(a)}(t) = y_{ij}^{(\ell)(a)} + \sum_{\tau=1}^T \sum_{\alpha=1}^{N(\ell)} \sum_{x \in W} \sum_{y \in W} v_{ij}^{(\ell)(a, \alpha)} y_{i+x, j+y}^{(\ell)(\alpha)}(t-\tau) \quad (11)$$

また、LSTM の基本構造の例を図 6 に示す。

### 2.4 CNN を用いた表情認識実験

#### 2.4.1 CNN の構造

#### 2.4.2 表情認識に用いる表情画像

この実験に用いたデータセットは、イスタンブールのマルマラ大学が公開しているデータセットで、被験者 30 人に様々な画像や動画を見せたときの反応を 10 秒から 20 秒の動画にまとめたものである。表情は、Anger , Boredom , Bothered , Concentrating , Contempt , Disgust , Fear , Happiness , Neutral , Sadness , Surprise , Thinking , Unsure の 13 種類である。本実験ではこの中から、Bothered , Happiness , Neutral , Sadness の 4 種類の表情を用いて実験を行った。データセット内の人物画像の例と実験に用いた表情画像の例を図 7 と図 8 に示す。図が示す表情についてはデータセットに付属する excel シートから確認し、訓練データ、検査データを作成した。また excel データの一部を図 9 に示す。本実験では訓練データに各表情 40 データ、検査データに各表情 15 データとし学習後の識別率を確認した。

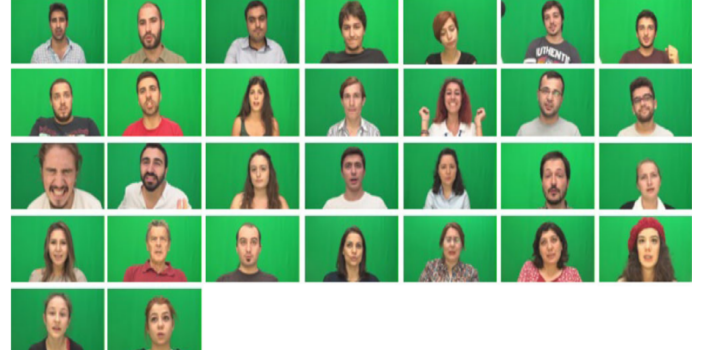


図 7: 人物画像の例



図 8: 表情画像の例

#### 2.4.3 実験方法

#### 2.4.4 結果

#### 2.4.5 考察

Number	Subject Label Clip		Clip Name	Emotion	Emotion Code	Gender	Age
1	1	5	S001_005	Unsure	13	M	23
2	1	6	S001_006	Thinking	12	M	23
3	1	7	S001_007	Concentrating	4	M	23
4	1	8	S001_008	Concentrating	4	M	23
5	1	9	S001_009	Unsure	13	M	23
6	1	10	S001_010	Happiness	8	M	23
7	1	11	S001_011	Happiness	8	M	23
8	1	12	S001_012	Concentrating	4	M	23
9	1	13	S001_013	Concentrating	4	M	23
10	1	14	S001_014	sadness	10	M	23
11	1	15	S001_015	Anger	1	M	23
12	2	1	S002_001	Neutral	9	M	24
13	2	10	S002_010	Unsure	13	M	24
14	2	11	S002_011	Unsure	13	M	24
15	2	12	S002_012	Concentrating	4	M	24
16	2	13	S002_013	Neutral	9	M	24
17	2	14	S002_014	Concentrating	4	M	24
18	2	15	S002_015	Happiness	8	M	24
19	2	16	S002_016	Unsure	13	M	24
20	2	17	S002_017	Happiness	8	M	24
21	2	18	S002_018	Happiness	8	M	24
22	2	19	S002_019	Neutral	9	M	24
23	2	20	S002_020	Happiness	8	M	24
24	2	21	S002_021	Unsure	13	M	24
25	2	22	S002_022	Happiness	8	M	24

図 9: excel シートの例

## **3 chapter3**

chapter3

## 4 chapter4

## 5 考察

考察

## 6 まとめと今後の課題

### 6.1 まとめ

まとめ

### 6.2 今後の課題

課題

## 参考文献

- [1] 筆者名. タイトル. 出版社.

- [1] 波多野寛人, 黒岩丈介, 小高知宏, 諏訪いずみ, 白井治彦, 平成 25 年度電気関係学会北陸支部連合大会 発表, F2-55 階層的な重なりを有するサイクル記憶を埋め込まれたカオスニューラルネットワークのカオス的遍歴現象 (2013.9)