

2次元画像における面内回転不変な物体認識に向けて

森山明日佳†, 古屋貴彦†, 大淵竜太郎†

概要 : 2次元画像からの物体認識では, 物体の幾何変形(平行移動, 拡大縮小, 面内回転)に対する不変性が要求される。近年注目を集める畳み込みニューラルネットワーク(CNN)はその構造上, 物体の平行移動および拡大縮小に対する頑強性を持つが, 一方で物体の面内回転に対する頑強性の獲得が困難である。本研究では, 面内回転に頑強な物体認識をねらい, CNNのアーキテクチャと学習方法を改良する。畳み込み層とこれに続く全結合層から成る「標準的」なCNNのアーキテクチャに新たに面内回転プーリング(RP)層を追加する。学習方法には画像の回転拡張を行う。RP層は面内回転した複数の画像の処理結果である特徴マップや特徴ベクトルを, 同座標の画素間でプーリングする層である。RP層は特徴マップ(特徴ベクトル)を1つに統合することで, 画像の回転により特徴マップ(特徴ベクトル)に生じる差異に対し頑強な識別を行う役割がある。RP層は面内回転した画像の組を統合するため, オリジナルの画像を多方向に回転させる必要がある。そのために, データセットの回転拡張を行う。RP層はその性質上, 畳み込み層と全結合層の直後に挿入される。実験では, 面内の向きが不定な物体(プランクトンと銀河画像)を対象に, RP層を層のどの位置に挿入するのが良いか調査する。またRP層と回転拡張数の関係を調査する。実験の結果, 「標準的」CNNの場合, 全結合層の中ほどにRP層を挿入すると良いことが分かった。

キーワード : 情報処理学会論文誌ジャーナル, コンピュータビジョン, 機械学習, 回転, 不変性

1. はじめに

2次元画像認識は, コンピュータビジョンにおける重要な研究分野の1つである。文字認識システムを始めとした多くの2次元画像認識システムでは, 入力パターンの幾何変形, 例えば平行移動, 非等方拡大縮小, 回転, などが正しく認識出来ることが求められる。例えば図1に示した画像に代表されるように, 実世界の画像に含まれる物体の殆どは, その向きや大きさが多種多様である。2次元画像識別システムは, このような条件の下で正確な認識を行うために, 物体の幾何変形に頑強なことが必要になる。そこで, 本研究では, 2次元画像の面内回転に不変な物体認識を目指す。ここで, 面内回転とは, 画像に鉛直な軸周りの回転である。

畳み込みニューラルネットワーク(CNN)は, 画像の局所領域の畳み込み処理と, 空間プーリングによる統合処理の2つによって, 物体の平行移動と非等方拡大縮小に対して頑強な認識が可能である。しかし, 従来のCNNでは回転不変性を得るための明示的な対策が施されておらず, 結果として既存のCNNは十分な回転不変性を持たない。

本研究ではCNNを用いて面内回転不変な識別を目指す。本研究では, CNNのアーキテクチャと学習方法の2つを改良する。アーキテクチャ上の提案は, 入力画像を複数方向に回転した後に処理した結果をプーリングする面内回転プーリング層(RP層)の導入である。入力画像を複数の角度で回転した画像群をCNNのRP層以前の段で個別に処理し, その複数回転方向の処理結果である特徴マップ(RP層を畳み込み層に挿入する場合)ないし特徴ベクトル(RP層を全結合層に挿入する場合)をプーリングすることで, 回転に対する頑強性を得ることを狙う。提案手法のRP層は

Suらの手法[1]に発想を得た。彼らの手法におけるView-Poolingを2次元画像に対して行ったのがRPである。ここで, RP層をCNNにおける複数の畳み込み層, 全結合層, 等どの段階で挿入するかを検討の余地がある。また, 学習方法については, 入力画像を面内で回転して拡張するデータ拡張方法を提案する。この場合, どのような角度で回転し, どのくらいの数に拡張するか, などに検討の余地がある。本論文の研究では, 畳み込み層と全結合層からなる「標準的」でかつ比較的単純な構造のCNNの複数の位置にRP層を挿入し, その効果を比べる。また, 学習用画像の回転による拡張についてもその影響を調べる。

評価実験は, 多種のプランクトンを撮影した画像からなるプランクトン識別のデータベース, および, 多種の銀河を撮影した画像についてこれらを識別する属性が振られたデータベース, の2つを用いて行った。実験の結果, RP層は畳み込み層の後の全結合層の中ほどで行うのが良いという傾向が得られた。また, 今回のCNN構成とデータセットの場合, 実験画像の拡張枚数は14~20枚が最適だという結果が得られた。



図1. 文字, プランクトン, 車の衛星写真, 銀河画像

† 山梨大学

2. 関連研究

2次元画像認識は手作り特徴を用いる方法と、特徴抽出から認識までを一手に行う方法の2種類がある。手作り特徴を用いた手法としては、DalalらのHistograms of Oriented Gradients (Hog) [2], LowらのScale-invariant Feature Transform (SIFT) [3], 回転不変性はSIFTに劣るもののSIFTより処理が高速なBayらのSpeeded Up Robust Features (SURF) [4]などが挙げられる。HoGは回転や拡大縮小の影響を受けるが、照明変化に頑強である。特にSIFTやSURFは照明変化や回転変換、拡大縮小に対して頑強な2次元局所画像特徴として物体認識タスクにおいて広く用いられてきた。これらの統合には、Bag-of-Features法[5]やその改良版であるFisher-Vector法[6]などが用いられてきた。

近年注目を浴びているのが、多層のニューラルネットを用い、画像からの低レベル特徴抽出から物体識別までの処理を多量のデータから学習して一貫して処理する手法である。福島の新オコグニトロン[7]がその起源の一つであるが、最近では2次元画像識別の問題に対し、AlexNet [8]やVGG [9]などに代表されるCNNを適用し、従来の手作り特徴を用いた方法よりも高精度な識別が可能になった。CNNは物体の平行移動と非等方拡大縮小に対する頑強性が高い一方で、面内回転に対する頑強性が低い。面内回転不変な識別を行うには、CNNのアーキテクチャを変える方法と学習方法を変える方法の2種類が考えられる。

アーキテクチャを変えた研究としては佐藤ら[10]の回転対応型新オコグニトロンがある。新オコグニトロンは厳密にはCNNとは異なるが、現在のCNNの畳み込み層とプーリング層は、新オコグニトロンに基づいている。佐藤らは新オコグニトロンの構成を拡張し、CNNの畳み込み層に相当するS層の細胞面群に、回転角を表す次元を付与し、画像の平行移動と回転変換の両方に対応したネットワークを提案した。

CNNのアーキテクチャを変えて回転不変性を狙った例として、このほか、Fasel [11], Dieleman [12]などがある。Faselらの手法では、入力画像を少しずつ回転させた複数の画像群を作り、それらをパラメータ共有したCNNに入力した。畳み込み層で作成された特徴マップ群を、ガウシアンフィルタによってぼかしながら、元画像からの回転角が近い特徴マップ同士を繰り返し統合することで、画像あたり1つの特徴に統合した。彼らの手法はSIFTよりも回転に頑強な識別を可能にした。Dielemanらの手法は、特徴マップに対して拡張・統合処理を4種類組み込むことを提案した。彼らは特徴マップを90度ずつ4方向に回転させた特徴マップ群を作成し、畳み込み層や全結合層の直後の特徴マップ群に対し回転による拡張および統合処理を行った。評価の結果、提案する拡張・統合処理を用いることにより、識別精度が向上した。

特徴に内在的な回転不変性を求めるのではなく、SIFTやSURFなどのように画像の向き（あるいは特徴抽出器の向き）を画像に合わせて正規化することで回転不変性を得る、CNNに基づく方法もある。Yiら[13]では、SIFTの各処理ステップに相当する、顕著点検出、画像の向き検出と正規化、特徴抽出、の処理を学習ベースのCNNの機構でほぼそのまま実現し、回転不変性を持った特徴を得た。

CNNアーキテクチャについて本研究に最も関連するのがSuら[1]の3次元形状モデルの形状類似検索の研究である。この手法は、3次元モデルの見かけ（レンダリングした2次元画像）の比較で行う。2次元画像の回転は面内回転の1自由度のみだが、3次元モデルの場合はその回転の自由度が3である。Suらの手法では、彼らは3次元モデルの「自然な上向き」ベクトルの軸周りに重心に向けた多数のパーティクルカメラを配置して複数の2次元レンダリング画像を作成し、これらの画像をもとに3次元形状を比較する。軸周りの多数のカメラの画像を個別にCNNに入力し、得られた特徴マップをView-Poolingというプーリング層で統合することで、その軸周りの回転（回転の3自由度のうちの1自由度のみ）に対する不変性を得た。

回転不変性を得るためにCNNの学習方法を変えた例としては、Sermanetら[14]やDielemanら[15]がある。彼らは元の学習用画像に対して、非等方拡大縮小や、回転変換を含む様々な幾何変換を加えてデータを拡張し、拡張したデータを使ってネットワークの学習を行った。これには、物体の描画位置や大きさ、回転角度などが少しずつ異なる画像を大量に学習させて、様々な幾何変換に不変な特徴を学習させる狙いがあった。

3. 提案手法

本研究では、アーキテクチャの改良と学習方法の改良の2つを組み合わせることで面内回転に不変な識別を狙う。アーキテクチャは、Suら[1]の手法を2次元画像の面内回転に適用した処理層であるRotation Pooling層 (RP層)を導入した。RP層は、データセットを回転拡張して得た複数の回転画像を個別に処理することで得られる複数の特徴マップ（ないし特徴ベクトル）を1つの特徴マップ（ないし特徴ベクトル）にプーリングする処理である。RP層で行われるプーリングには、最大値プーリングと平均値プーリングの2通りを用いた。また、学習方法はSermanet [14]やDieleman [15]らの手法を参考に、画像を多方向に面内回転させてデータセットの拡張を行った。CNN構造におけるRP層の最適な位置を探るため、CNNの様々な層の後にRP層を挿入し、識別誤差を調べた。CNNは学習に時間がかかることが多い。本研究で用いた比較的単純なCNNと比較的小規模なデータベースでも学習に数日を要することがある。高精度を目指した複雑な構成のCNNを用いてRP層の位置を変える等のアーキテクチャ上の探索を行うと非常に時間がかかる。

そこで、本論文で述べる研究では、比較的小規模な CNN を用いることとした。

提案手法では、回転角が 0~360 度の回転行列をランダムに S 個作成し、画像中心を回転軸として入力画像を変換し、S 枚の面内回転画像を作成する。そして、得られた画像群をパラメータ共有した複数の CNN に入力する。これらの画像は RP 層までの前段 (画像ごとに処理)、RP 層による統合、RP 層以後の後段、による処理を経て、クラスラベルの尤度、属性値の尤度のベクトル、などとなって出力される。それぞれの画像は、RP 層によって統合される前の前段では、パラメータ共有された CNN で処理される。

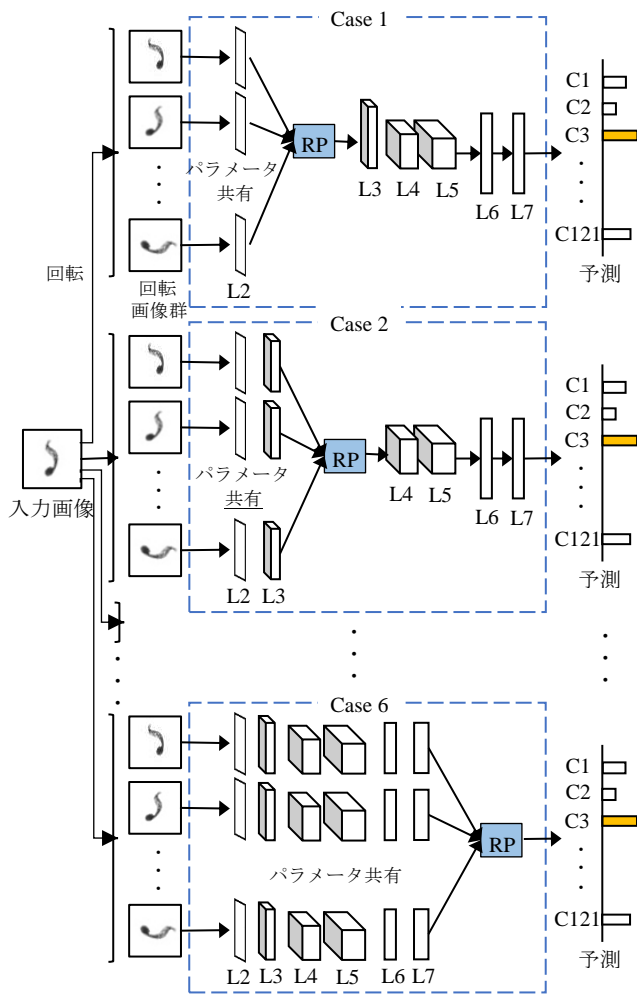


図 2. 提案手法の概要図

本研究では、図 2 のように RP 層の位置をずらしながら最も識別誤差が小さくなるような RP 層の位置を探索する。図 2 は面内回転したプランクトンの画像群が RP 層を経て統合され識別されるまでの様子、RP 層挿入位置を変えた Case 1 から Case 6 までの複数の CNN アーキテクチャについて示す。ここで L2 や L3 などの表記は CNN の処理層で、例えば L2 は第 2 層目を指す。

図 2 の Case 2 は、第 3 層の直後で RP した場合で、前段である第 3 層まではパラメータ共有した複数の CNN で処

理し、RP 層で統合された後、後段は 1 本の畳み込み層および全結合層で処理する。Case 6 の例では、第 7 層 (全結合層) を通過した後で RP を行い、RP された特徴が直接出力される。

本研究では、VGG [9] を参考した小規模な CNN (図 3, 表 1) をもとに RP 層を挿入して実験を行う。ネットワークは全部で 8 層で、畳み込み層が 4 層、全結合層が 3 層である。過学習を防ぐため、全結合層の後にドロップアウト層 [16] を追加している。空間プーリングは 3 層目と 5 層目の直後で最大値プーリングを行った。ネットワークの最適化アルゴリズムには Adam [17] を採用した。学習係数は 0.001 に設定した。これより大きい学習係数だと学習が収束しない為である。畳み込み層と全結合層の重み係数 W は標準偏差 $\sqrt{N/2}$ の切断正規分布に従う乱数で初期化した。ただし N は入力次元数である。

一般に、識別誤差等を最小化するには、多段で複雑な CNN を用いたほうが良い。しかし、複雑な CNN をもとに RP 層の位置を変えつつアーキテクチャ探索の実験を行うと学習等に時間がかかりすぎる。そこで、本論文の実験では、比較的層の数が少ない単純な構成の CNN をもとに RP 層の位置を探ることとした。

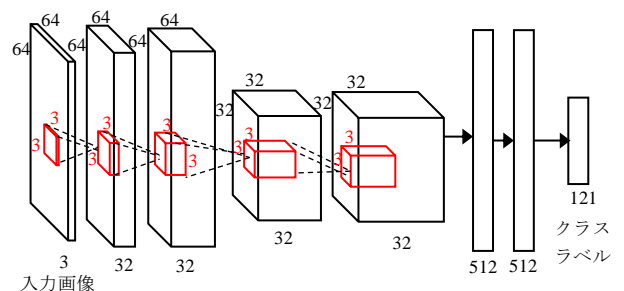


図 3. ネットワーク構造

表 1. ネットワーク構造

層	種類	特徴次元数	フィルタ
1	入力	64×64×1, 64×64×3	—
2	畳み込み	64×64×32	3×3×32
3	畳み込み	64×64×32	3×3×32
		2×2 最大値プーリング, ゼロパディング	
4	畳み込み	32×32×32	3×3×32
5	畳み込み	32×32×32	3×3×32
		2×2 最大値プーリング, ゼロパディング	
6	全結合	512	—
7	全結合	512	—
8	全結合	121, 37	—

図 4, 図 5 に RP 層による統合処理の詳細を示す。図 4 は畳み込み層後の特徴マップを RP 層で統合する場合の処理を示し、図 2 では Case 1 と Case 2 に対応する、図 5 は全結合層後の特徴ベクトルを RP 層で統合する場合で、図 2

では Case 6 に対応する。なお、説明の簡単化のために入力画像が回転方向で $S=3$ 枚に拡張された場合の例を示す。

図 4 の場合、拡張した画像群が CNN に入力され、畳み込み処理を経て、拡張された枚数だけ特徴マップが生成される。入力画像を 3 枚に拡張した場合、仮に畳み込みフィルタ数が 2 であったならば、3 枚の画像それぞれに対して異なる特徴マップが 2 枚ずつ生成される。すなわち、畳み込み層を通過した直後では計 6 枚の特徴マップが存在している。ただし、全ての CNN はパラメータ共有しているため、学習によって得られた同一の畳み込みフィルタを異なる画像に対して適用している点に注意が必要である。全特徴マップのうち、同一のフィルタから生成された特徴マップ間の同座標の画素間でプーリングすることで、画像当たり 1 枚の特徴マップに統合する。RP 層後の特徴マップの枚数は RP 層前の畳み込み層のフィルタ数と一致する。

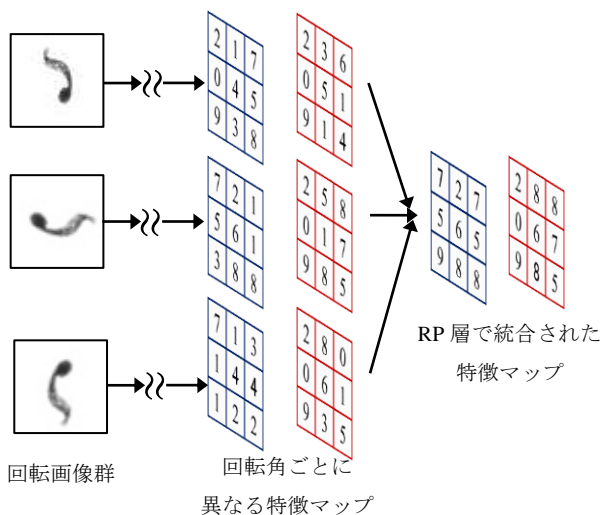


図 4. 畳み込み層後で行われる RP 層の処理

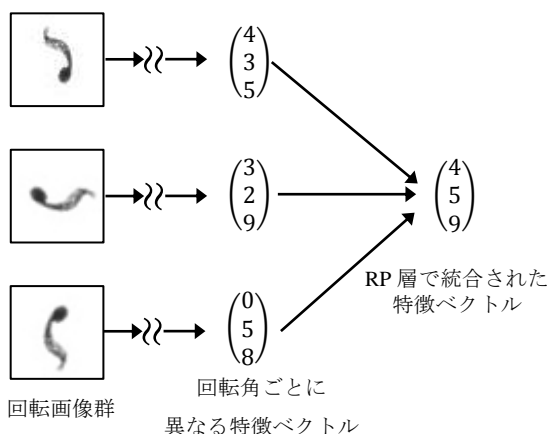


図 5. 全結合層後で行われる RP 層の処理

図 5 では畳み込み層、全結合層を経て得られた回転方向別の $S=3$ 個の特徴ベクトルが RP 層によりプーリングされる。全結合層のニューロン数が N なら、複数の N 次元特徴

ベクトル群がプーリングされる。

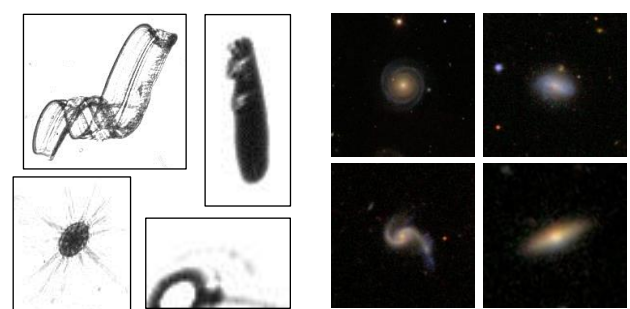
プーリングは、畳み込み層、全結合層いずれの場合も、最大値プーリングと平均値プーリングの 2 種類を試みた。図 4 と図 5 では最大値プーリングの場合を図示している。

RP 層によって統合された特徴マップや特徴ベクトルは、元の入力画像（回転していないオリジナルの向きの画像）と、ランダムに回転した画像群の両方の向き情報を含んでいる。既存の多くの 2 次元画像識別 CNN における空間プーリング層では同一特徴マップ内の隣接する画素間でプーリングを行い、特徴の平行移動に対して頑強になった。RP 層では、特徴マップ間の同座標の画素間でプーリングを行うことにより、特徴マップ間の差異、つまり回転変換に対する頑強性を得ることが期待できる。また、RP 層は特徴間のプーリングを行うだけで固有のパラメータを持たず、通常の誤差逆伝播法をそのまま用いて学習することが出来る。また、もしネットワークの浅い（入力に近い）所で RP 層による統合ができれば、その後の層における計算量を削減できることになる。

4. 評価実験

本研究の目的は面内回転に不変な画像識別であり、この目的を達成するために RP 層を持つ CNN 構成と、これを回転方向で拡張した画像で学習することを提案した。評価実験では、CNN の処理過程のおよそどのあたりに RP 層を挿入すべきかを調べる実験、および、学習画像を回転方向で拡張する場合の最適な拡張枚数を調査する。

まず、RP 層の挿入位置を変え、挿入位置と識別誤差の関係を調べる実験を行った。次いで、この実験で得た最も識別誤差の小さい RP 層の挿入位置において、拡張枚数を変え、拡張枚数と誤差の関係を調べる実験を行った。



(a) プランクトン (b) 銀河

図 6. データセットの画像例

実験には、回転を含むデータセットとして Cowen ら[18]によるプランクトンの画像のデータセット、および Willet ら[19]の銀河の画像のデータセット Galaxy Zoo2 の 2 種類を用いた。図 6 に実験で利用したデータセットの画像例を示す。これは、本提案手法を評価するに当たり、回転のあるデータセットが必要だったからである。この 2 つのデー

タセットは学習用と評価用に分かれていなかったの、実験はクロスバリデーションによって評価した。実験で本手法と識別誤差を比較する対象は Dieleman [12]である。実験には、CPU が Intel Core i7-6700 (3.4GHz)、メインメモリは 64GB、GPU は GeForce GTX 1080 (8GB)、OS は Ubuntu 14.04、プログラミング環境には Python 2.7、TensorFlow 0.12.0 rc0 のフレームワークを用いた。

(1) Plankton

Cowen ら [18]によるプランクトンデータセットは計 30,336 枚のグレースケール画像で構成される。各画像にはそれぞれクラスラベルが付与されておりクラス数は 121 である。また各クラスに所属する画像の枚数にもばらつきがある。画像の画素数や縦横比は統一されていない。そこで、以下のような処理を施して画素数を 64×64 に統一した画像を作成した。まず、プランクトンが (画像) の中心に位置するようセンタリングする。プランクトンデータセットの画像は背景が画素値=255 (白) のため、プランクトンの一部であろう画素値≠255 の領域を囲む矩形バウンディングボックスを設定する。この矩形バウンディングボックスの長辺の長さを n としたとき、1 辺が $n \times 1.6$ の画素値=255 (白) のキャンバスを用意し、その中心に元の画像を埋め込む。ここで 1.6 倍したのは、この画像を回転したときにはみ出さないためである。次に、こうして得た全画像をバイリニア補間を用いて 64×64 にリサイズした。画像サイズ 64×64 は Dieleman [12]に準ずる。回転によるデータ拡張を行う場合はこれら画像をバイリニア補間しつつ回転する。プランクトンデータセットはクラスラベルを持ち、CNN の出力はクラスラベルを示す確率値である。そこで、識別誤差にはクロスエントロピー誤差を用いた。

評価実験では全 30,336 枚の画像を訓練用に 27,302 枚、評価用に 3,034 枚になるように分割し、10 クロスバリデーションによって評価した。また、RP 層を挿入したときは画像を拡張する必要があるため、訓練用に 27,302× (拡張数) を割り当て、評価用に 3,034× (拡張数) を割り当てる。

(2) Galaxy

銀河データセットは Willet ら [19]の Galaxy Zoo2 を利用する。このデータは計 61,578 枚のカラー画像で構成される。各画像解像度は 424×424 で、これを 64×64 にリサイズして実験に用いた。リサイズ時の補間はバイリニア補完を用いた。

Galaxy Zoo2 は、銀河の形状に対する 11 項目の選択質問を複数回答することで得られる 37 個の属性によって画像を分類する。例えば、「銀河は単純な楕円形か、もしくは円盤銀河か?」のような質問や、「オリオン腕 (銀河の中心から伸びる腕のような形状部分) の兆候が見られるか?」などである。Galaxy Zoo2 はクラウドソーシングサービスによって膨大な画像群を分類している。全画像は、各属性への

投票数を確率値とみなした 37 次元ベクトルと対応している。そのため、銀河データはプランクトンデータとは異なり、クラス識別問題ではなく 37 次元ベクトルの回帰問題を解く。そこで、識別誤差には平均二乗誤差を用いた。

評価実験では、全 61,578 枚の画像を訓練用に 55,420 枚、評価用に 6,158 枚になるように分割し、プランクトンデータと同様に 10 クロスバリデーションによって評価した。また、RP 層を挿入したときは、訓練用に 55,420× (拡張数) を割り当て、評価用に 6,158× (拡張数) を割り当てる。画像の回転変換によって出来る画素の欠損には、欠損位置に黒 (R=0, G=0, B=0) の画素を補完した。

4.1 最適な RP 層の挿入層を探索

まず、RP 層を CNN の何層目に挿入するのが良いかを探索する実験を行った。この実験はプランクトンと銀河の両データそれぞれについて、RP 層で最大値プーリングと平均値プーリングの 2 種類を行った場合を比較した。図 7 と図 8、はそれぞれ、プランクトンと銀河データにおいて、入力画像を 4 倍に拡張し、更に RP 層をネットワークの各所に挿入して学習したときの評価用データの誤差を計測した図である。プランクトンと銀河の両データとも、RP 層の位置が入力層に近い 2 層目 3 層目では、識別誤差が高くなっているのが分かる。これは、ネットワークの 2 層目や 3 層目の入力層に近い位置に RP 層を挿入すると、画像の局所特徴の空間的配置が、畳み込み層で十分に学習出来なかった為と思われる。逆に 6 層目以降の全結合層では、既に局所特徴の空間的配置が失われている為、回転角度の差異に注力したプーリングをすることで、識別誤差が低下した。

また、最大値プーリングと平均値プーリングの 2 種類を比較したところ、畳み込み層の直後に RP 層を置く場合は最大値プーリングを用い、全結合層の直後に RP 層を置く場合には、平均値プーリングを用いるのが良い結果となった。この実験結果から、全結合層後に平均値プーリングの RP 層を挿入するのが一番良いと言える。

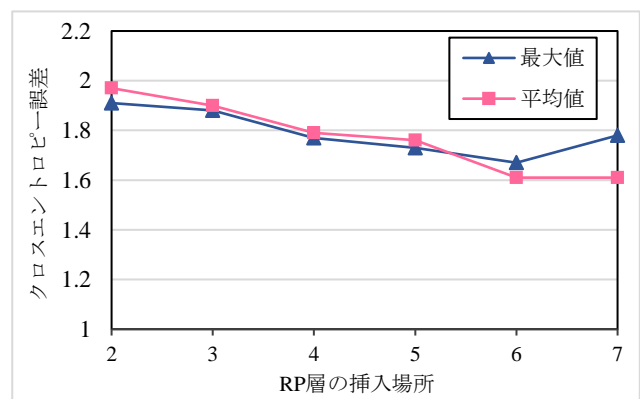


図 7. プランクトンデータにおける RP 層の挿入場所と識別誤差

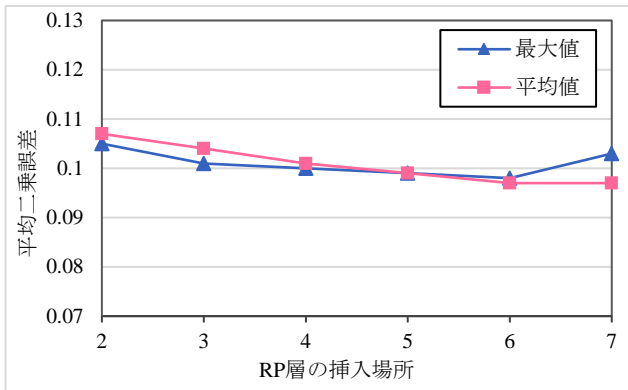


図 8. 銀河データにおける RP 層の挿入場所と識別誤差

4.2 最適な拡張枚数との探索と RP 層の効果検証

次に、入力画像の最適な拡張枚数を確認する実験を行った。図 9、図 10 に両データセットにおいてデータ拡張枚数を変えたときの、RP 層の有無による評価用データの識別誤差を示す。なお図中の等倍とは、データセットの拡張を行わない場合の評価用データの識別誤差を意味する。本研究では等倍をベースラインとして、データ拡張と RP 層の有効性を確かめた。4 倍という表記は、入力画像にランダム回転を加えて 4 倍に拡張したデータセットを用いて学習を行ったときの、評価用データの識別誤差を意味する。S 倍 RP という表記は、データセットを S 倍に拡張した上で、6 層目の直後に RP 層を挿入したときの評価用データの誤差を意味する。また、RP 層は全て 6 層目の直後に挿入し、プーリングには平均値プーリングを用いている。

図 9 では等倍のデータで学習した場合、150 回ほど学習した時点で収束した。一方 4 倍、8 倍にデータ拡張を行ったときには、70 回程度で学習が収束した。学習回数は全データを 1 度ずつミニバッチ処理したときを 1 回と数え、ミニバッチ数は 100 である。以上の結果を踏まえ、プランクトンデータでの学習回数を 70 回と決め、14 倍、20 倍に拡張したときの識別誤差を計測した。拡張枚数を増やすと学習回数が少なくなっていくが、データ拡張だけでは、等倍の誤差を超えることはなかった。

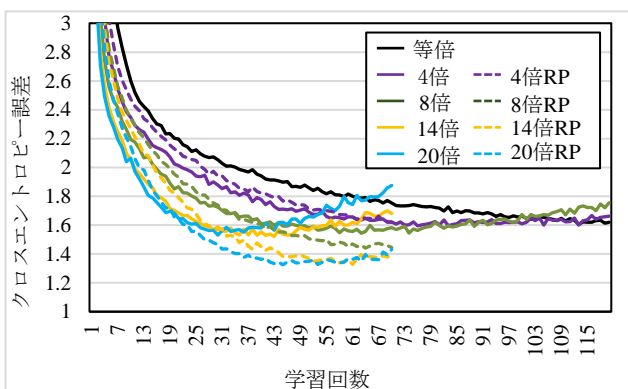


図 9. プランクトンデータセットにおける 評価用データのクロスエントロピー誤差の推移

次に、RP 層を導入することで識別誤差がどのように変化するかを調べた。4 倍拡張時には RP 層を挿入したことで識別誤差が大きくなったが、8 倍、14 倍、20 倍のときには RP 層を挿入したことにより識別誤差が小さくなった。つまり、同じ拡張枚数でも RP 層を入れることで、より過学習しにくくなっていることが分かる。最も識別誤差が小さくなったのは拡張枚数 14 枚のときに RP 層を入れた場合で、等倍のときと比べて有意に低下した。

銀河データセットでは、データ拡張だけでも効果があった。特に 20 倍に拡張したときが最も識別誤差が小さくなった。しかし、RP 層を挿入した実験では、プランクトンのときのように RP 層挿入による識別誤差の低下は見られず、反対に RP 層なしの場合と比べて若干誤差が大きくなる結果となった。また、最も識別誤差が小さくなった条件での既存研究との比較を行った。本提案手法では Dieleman ら[12]の手法には及ばない結果となったが、プランクトンデータにおいては RP 層を導入することで、面内回転に対して一定の頑強性を得ることができることを示せた。表 2 に Dieleman らとの比較結果を示す。

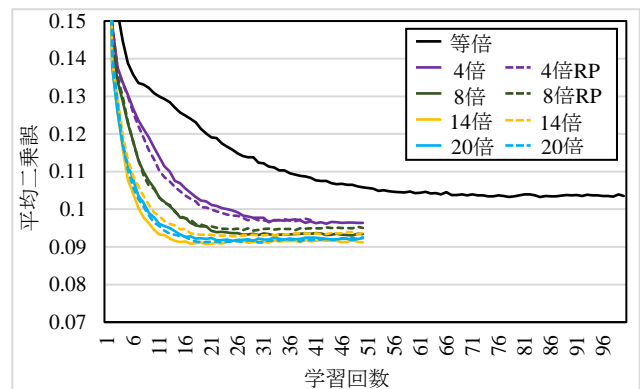


図 10. 銀河データセットにおける 評価用データの平均二乗誤差の推移

表 2 既存研究との比較

データセット	データセット	拡張枚数	誤差
Dieleman [12]	Plankton	4	0.678
提案手法	〃	14	1.398
Dieleman [12]	Galaxy	4	0.0821
提案手法	〃	20	0.0926

5. まとめと今後の課題

2次元画像からの物体認識システムは、物体の平行移動、非等方拡大縮小、回転変換に頑強である必要がある。本研究では、物体の移動、拡大縮小に頑強なフレームワークである畳み込みニューラルネットワーク (CNN) をもとに、画像の回転変換に頑強な識別を目指し、CNN に対する回転プーリング層 (RP 層) の導入と、入力画像をランダムに回転変換させるデータ拡張方法を提案した。さらに、実験的に、CNN 内においてどの位置に RP 層を挿入するのが良い

かの探索と、データの回転拡張の効果の評価を行った。プランクトンと銀河の2つのデータセットを用いた実験では、プランクトンデータではRP層を比較的後ろ、畳み込みが終わった後の全結合層のなかの第6層付近に挿入することで最も良い結果が得られ、識別誤差が最低となった。RP層の効果はデータセットに依存し、銀河データセットではRP層の効果が確かめられなかったが、データ拡張による識別誤差の低下が確認できた。また、今回のCNN構成とデータセットの場合、回転によるデータ拡張は14~20倍程度が良いという結果が得られた。

今後の課題としては、まず銀河画像の拡張時における外挿手法を改良することが挙げられる。銀河データで効果が見られなかったのは、データ拡張時にできる空白の画素位置に、黒で外挿をしたのが影響を及ぼした可能性があったためである。また今回学習時間の関係で実現できなかった、VGG [9]などの多層のDCNNを用いた実験を行うことが挙げられる。また、[11]のような、ネットワーク内で複数回に渡り、RP層を挿入する多重解像度版RPなどを実験することが挙げられる。

参考文献

- [1] Su, H., Maji, S., Kalogerakis, M., Miller, L M.. Multi-view Convolutional Neural Networks for 3D Shape Recognition. Proc. International Conference on Computer Vision (ICCV) 2015, pp. 945-953.
- [2] Dalal, N. and Triggs, B.. Histograms of oriented gradients for human detection. Proc. Int'l Conf. on Computer Vision and Pattern Recognition (CVPR) 2005, pp. 886-893.
- [3] Lowe, G D.. Object recognition from local scaleinvariant features. ICCV 1999, pp. 1150-1157.
- [4] Bay, H., Tuytenlaars, T., and Gool, V L.. SURF:Speeded Up Robust Features. Proc. European Conference on Computer Vision (ECCV) 2006, pp.404-417.
- [5] Csurka, Gabriella., Dance, R C., Fan, L., Willamowski, J., Bray, C.. Visual Categorization with Bags of Keypoints. Proc. Workshop on Statistical Learning in Computer Vision ECCV 2004, Vol. 1. pp.59-74.
- [6] Perronnin, F., Sanchez, J. and Mensink, T.. Improving the fisher kernel for large-scale image classification. Proc. ECCV 2010, pp.143-156.
- [7] Fukushima, K. and Miyake, S.. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognition, 1982 15(6), pp.455-469.
- [8] Krizhevsky, A., Sutskever, I. and Hinton, E G.. Imagenet classification with deep convolutional neural networks. Proc. Advances in Neural Information Processing Systems (NIPS) 2012, pp. 1097-1105.
- [9] Simonyan, K. and Zisserman, A.. Very deep convolutional networks for large-scale image recognition. Proc. International Conference on Learning Representations (ICLR) 2015. arXiv preprint arXiv:1409.1556
- [10] 佐藤俊治, 黒岩文介, 阿曾弘具. 回転対応型ネオコグニトロン. 電子情報通信学会論文誌. 1998, D-2, 情報・システム, 2-情報処理 J81-D-2(6), pp.1365-1374.
- [11] Fasel, B. and Gatica-Perez, D.. Rotation-Invariant Neoperceptron. Proc. International Conference on Pattern Recognition (ICPR) 2006. pp.336-339.
- [12] Dieleman, S., Fauw, D J., and Kavukcuoglu, K.. Exploiting Cyclic Symmetry in Convolutional Neural Networks. Proc International Conference on Machine Learning (ICML) 2016. pp.1889-1898.
- [13] Yi, K. M., Trulls, E., Lepetit, V. and Fua, P.. LIFT: Learned Invariant Feature Transform. Proc. ECCV 2016, pp.467-483.
- [14] Sermanet, P., and LeCun, Y.. Traffic sign recognition with multi-scale convolutional networks. Proc. International Joint Conference on Neural Networks (IJCNN) 2011, pp.2809-2813.
- [15] Dieleman, S., Willett, K. and Dambre, J.. Rotation-invariant convolutional neural networks for galaxy morphology prediction. Proc. Monthly Notices of the Royal Astronomical Society (MNRAS) 2015,450(2), pp.1441-1459.
- [16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Proc. Journal of Machine Learning Research (JMLR) 2014, 15(1), pp.1929-1958.
- [17] Kingma, D. and Ba, J.. Adam: A Method for Stochastic Optimization. Proc. ICLR 2015. arXiv preprint arXiv:1412.6980.
- [18] Cowen., Robert, K., Sponaugle, S., Robinson, K.L., and Luo, J.. Planktonset 1.0: Plankton imagery data collected from F.G. Walton Smith in Straits of Florida from 2014-06-03 to 2014-06-06 and used in the 2015 National Data Science Bowl, 2015. (NODC Accession 0127422). NOAA National Centers for Environmental Information. Dataset. doi:10.7289/V5D21VJD.
- [19] Willett, W K., Lintott, J C., Bamford, P S., Masters, L K., Simmons, D B., Casteels, RV K., Edmondson, M E., Fortson, F L., Kaviraj, S., Keel, C W., Melvin, T., Nichol C R., Raddick, J M., Schawinski, K., Simpson, J R., Skibba, A R., Smith, M A., Thomas, D.. Galaxy zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey. MNRAS 2013, pp 2835-2860.