深層学習によるだまし絵認識手法の提案および解析

1　はじめに

近年，深層学習の登場により，画像認識の分野は急速な発展を遂げている．単純な物体認識では既に人を凌駕する成果も報告されており，今後も計算機による画像識別能力は向上していくと考えられている．一方, 人間の感性や認知に関わる分野では深層学習を用いても十分な学習ができないという課題が報告されている．これは, 人間の感性や認知といった明確な正解が存在しない抽象的な概念を定量的に評価すること現在の人工知能の枠組みでは難しいためである．このため，従来の深層学習研究では, 画像分類や物体検出など, 答えが一意に定まる問題が主として扱われており, 計算機によるだまし絵の認識のような唯一解を定義しにくい問題については十分な研究がなされてこなかった．そこで，今回はだまし絵のように複数の意味解釈が可能な画像を対象とする．また，だまし絵ではないが, 意図せずそのような性質を持つ場合も考えられるため，以後本論文では複数の意味解釈が可能な画像を多義図形と定義する．

以上，本研究では, 人の視覚認知の多義性を計算機に理解させることを目的とし, 一般物体認識において高い性能を示している深層学習手法を用いて，計算機による多義図形の識別に必要な実験の枠組みおよび計算機による識別手法を提案する．また，実際のだまし絵を用いた数値実験によって提案手法の有効性を示す.

2　従来研究

2.1　多義図形

だまし絵の一種に，多義図形と呼ばれるものがある．多義図形とは，人の視覚系によって 2 通り以上に解釈される図形である．本研究では，画像中に存在する各オブジェクトのラベルが一意に定まるものを一義図形，だまし絵のようにラベルとして複数の解釈が可能なものを多義図形と定義する．認知科学の分野では，多義図形の解釈に影響を与える要因は，注視点や選択的注意とされている [1][2]. 図 1 に示す兎と鴨の両方に見える多義図形に関して，点 1 を注視しているときは 98 % 「鴨」と認識される一方，点 5 を注視している時は 94% 「兎」として解釈されることが報告されている [3]．このような多義図形の解釈について，計算機を用いたアプローチは堀江らによって報告されている [4].

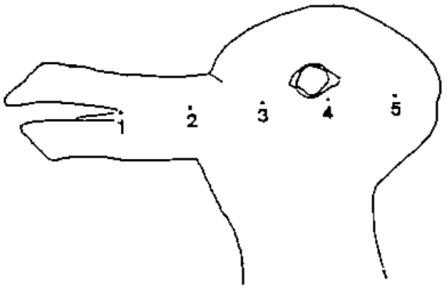

図 1: 兎鴨画像の注視点

2.2　Convolutional Neural Network

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深層学習手法として，Convolutional Neural Network (CNN) [5] が注目されている．CNN の持つ層として，主にフィルタによって画像の局所的な特徴抽出をする Convolution 層, 特徴を統合する Pooling 層, 特徴量に基づいた分類をする全結合層がある. 2014 年に発表された VGG-16 は，“ImageNet” と呼ばれる大規模画像データセットで学習された，Convolution 層 13 層と全結合層 3 層を組合わせた 16 層からなる CNN モデルである.

2.3　Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) [6] とは，CNN が分類のために注視している範囲をカラーマップで表示する CNN の判断根拠の可視化技術である．Grad-CAM は, 予測クラスに対する勾配の大きさを寄与の大きさと考え, 分類予測を行う時に重要な箇所であると判断する．寄与の計算の際には, 一般的に最終畳み込み層の予測クラスの損失値に対する勾配が用いられる．本研究では, 勾配を可視化することで, 多義図形の判断根拠の解析をしている. 本研究では Grad-CAM の図としてヒートマップを用いて勾配の大きさを可視化し, 勾配の大きい部分を赤色, 小さい部分を青色とした. 本研究では, 赤色が最も CNN の最終畳み込み層の勾配が大きい, つまり判断に大きな影響を及ぼしているとし, また, 青色が最も CNN の最終畳み込み層の勾配が小さい, つまり判断にあまり影響を及ぼしていない, となるように可視化した.

表 1: Optuna によるパラメータ調整区間

中間層のユニット数	100〜500
ドロップアウト率	0〜1
学習率	0.00001〜0.01

3 提案手法

3.1 深層学習の構成

本研究では, 既存の深層学習手法を用いて, 多義図形を理解するために必要な問題の枠組みについて提案し, 具体的な実験方法に基づいて数値実験をした.

本研究では，計算機が多義図形を理解するとはどのようなことであるかを示すために，CNN を用いた多義図形を理解する手法および実験の枠組みについて提案する．まずデータセットに対して Data Augmentation を施し，ImageNet で学習済みの VGG-16 の最終畳み込み層 3 層以降の Transfer Learning を適用し, Optuna を用いて中間層のユニット数, ドロップアウト率, 学習率の 3 パラメータを調整した．表 1 に Optuna によるパラメータ調整区間を示す. 最後に, CNN による判断根拠を Grad-CAM によって可視化し, 比較した. また, 最終全結合層の Softmax 関数の出力値を入力画像のそのクラスに属するクラスらしさと呼ぶことにする.

3.2 数値実験の構成

以下に今回の実験について示す.

実験 1 風景と人の顔の多義図形, 風景画, 肖像画の 3 クラスの識別をした．また, テスト画像中の色と特徴点が類似した多義図形, 一義図形を抽出し, その判断根拠を比較した.

実験 2 実験 1 で作成した風景と人の顔の多義図形, 風景画, 肖像画の 3 クラス識別モデルを用いて, 一義図形の中に含まれる多義図形らしい部分構造を取り出せるか実験した．今回は, 風景画の中から顔にも風景にも見える箇所を取り出すことが可能か確認した.

3.3 データセット

本研究ではインターネットから多義図形として解釈できるだまし絵画像を収集し，データセットを作成した．

実験では, 風景と人の顔をモチーフにした多義図形が多く，また多義図形と似た肖像画や風景画が多く存在することに着目した．そこで, 風景と人の顔の多義図形については，著者が風景と人の顔であると判断した画像を集めて作成したデータセットを 「多義図形」クラスとした．「多義図形」クラスの訓練データのみ, 257 枚を 10 倍に Data Augmentation し, 2570 枚にして使用した. また, 一義図形のデータセットとして, WikiArt [7] 中の “landscape”, “cityscape” ラベルの画像を「風景画」クラス, “portrait” ラベルの画像を「肖像画」クラスとして用いた.

表 2: 実験 1 実験条件

クラス	3 クラス (多義図形, 風景画, 肖像画)
エポック	22
バッチサイズ	32
訓練枚数	2570 枚/クラス
評価枚数	36 枚/クラス
テスト枚数	72 枚/クラス
データサイズ	200 × 200 × 3(RGB)
活性化関数	Softmax
最適化関数	Adam
損失関数	交差エントロピー

表 3: 実験 1 Optuna による最適化結果

ドロップアウト率	0.67640
学習率	2.2686e-05
中間層のユニット数	100

4 数値実験

4.1 実験 1

4.1.1 実験条件

表 2 に実験条件を，表 3 に Optuna によるネットワークパラメータの最適化結果を示す.

4.1.2 実験結果

CNN による風景と人の顔の多義図形, 風景画, 肖像画の 3 クラス識別の精度はベースライン 33.3% に対して, 91.2% となった. 図 4 に縦軸を真値，横軸を CNN による予測値とした混同行列を示す．また図 2 に Grad-CAM による計算機の判断根拠結果を示す. Grad-CAM の結果より，風景画については全体に満遍なく判断根拠があり，

表 4: 実験 1 混同行列

真値	多義図形	60	5	7
	肖像画	1	67	4
	風景画	0	2	70
		多義図形	肖像画	風景画
		CNN による予測値		

肖像画については人の顔の領域に判断根拠があることがわかった．多義図形についても肖像画と同様に，人の顔の領域に判断根拠が集まっていることが確認できた．

また, 実験 1 で作成した 3 クラス識別モデルが類似した多義図形と一義図形を見分けることが可能かを確認した. 色, 特徴点共に似た「多義図形」クラスの画像と「肖像画」クラスの画像, 「多義図形」クラスの画像と「風景画」クラスの画像の対をテスト画像から抽出した．ここで, 色, 特徴点共に似ているというのは, ヒストグラム類似度が 0.9 以下かつ AKAZE 類似度が 120 以上の画像の組と定義する. 図 3 に類似画像の組と, 「多義図形」クラス, 「肖像画」クラス/「風景画」クラスの判断根拠を示す. この図より, 類似画像においても, 多義図形と一義図形との判断根拠は全く違っている.

4.2 実験 2

4.2.1 実験条件

入力画像を 5 × 5 個の格子に分割し, 1 × 1, 2 × 2, 3 × 3, 4 × 4 格子の全通りの選び方 (54 通り) で切り取り, 切り取った画像を全て学習済みの 3 クラス識別器を使ってテストした. 本実験では, 「風景画」クラスの画像を上記の方法で切り出して「肖像画」クラスと識別された画像, 「多義図形」クラスと識別された画像について詳しく見る.

4.2.2 実験結果

図 4 に「風景画」クラスの画像を切り出して「肖像画」クラスと識別された画像の例, 図 5 に「風景画」クラスの画像を切り出して「多義図形」クラスと識別された画像の例を示す. WikiArt 中の “landscape”, “cityscape” ラベルの画像 17960 枚の風景画画像を 54 通りに切り分けると 969840 枚となり, その内「多義図形」クラスとして識別された部分画像は 26990 枚 (2.78%) だった. 図 4 より, 肖像画と識別された画像には, 風景画の中に描かれている人の顔領域が映り込んでいることがわかる. 一方, 図 5 より, 多義図形と識別された画像には, 風景にも顔にも見える部分が映り込んでいることがわかる. このことから, 風景画の中から「風景に映り込んだ人の顔」と「風景にも顔にも見える多義図形」を分けて切り出すことができたと考えられる.

5 まとめと今後の課題

本研究ではだまし絵に代表される多義図形を計算機に理解させるための手法を提案し，複数の数値実験を通じてその有効性および人間の認知との関係を示した．実験 1 では，計算機に多義図形と多義図形を構成する 2 要素の一義図形とを 91.2 % の精度で識別させることに成功した．そして, Grad-CAM を用いた解析により，計算機による肖像画と多義図形の注視点における類似性を確認した．また, 類似した多義図形と一義図形においても, それらを識別することができた．実験 1-2 では, 風景画より, 風景にも顔にも見える多義図形を切り出すことができた.

今後の課題としては, 画像の見せ方や周囲の環境に配慮したアンケート実験，アイトラッキング技術の活用, 計算機によるだまし絵の自動生成などが挙げられる．

いまだその多くが解明されていない脳神経系の機構に迫ることから工学分野への応用まで, 錯視研究が担う領域は実に幅広いと言われる．人は無意識が 8 割ほどで行動を行っているとされるが，この無意識を再現することは非常に難しいため，AI による学習をさせることも困難であるとされている．本研究を進めることにより，脳のまだ解明されていない機能の理解への一歩となることを期待している．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 (B) (課題番号 19H04184) の補助を得て行われたものである．

参考文献

[1] Kiyoshi Noaki Morikazu Kawabata N, Nobuo Yamagami. Visual fixation points and depth perception. 1977.

[2] Kawabara N. Attention and depth perception. 1986.

[3] 岸本充史, 川端信男. 局所的・大域的情報選択モデルによる多義図形の非あいまい化. テレビジョン学会誌, Vol. 50, No. 5, pp. 594–598, 1996.

[4] 堀江紗世, 森直樹. 人工知能による多義図形認識手法の提案及び解析. 人工知能学会全国大会論文集, Vol. JSAI2020, pp. 3D1OS22a05–3D1OS22a05, 2020.

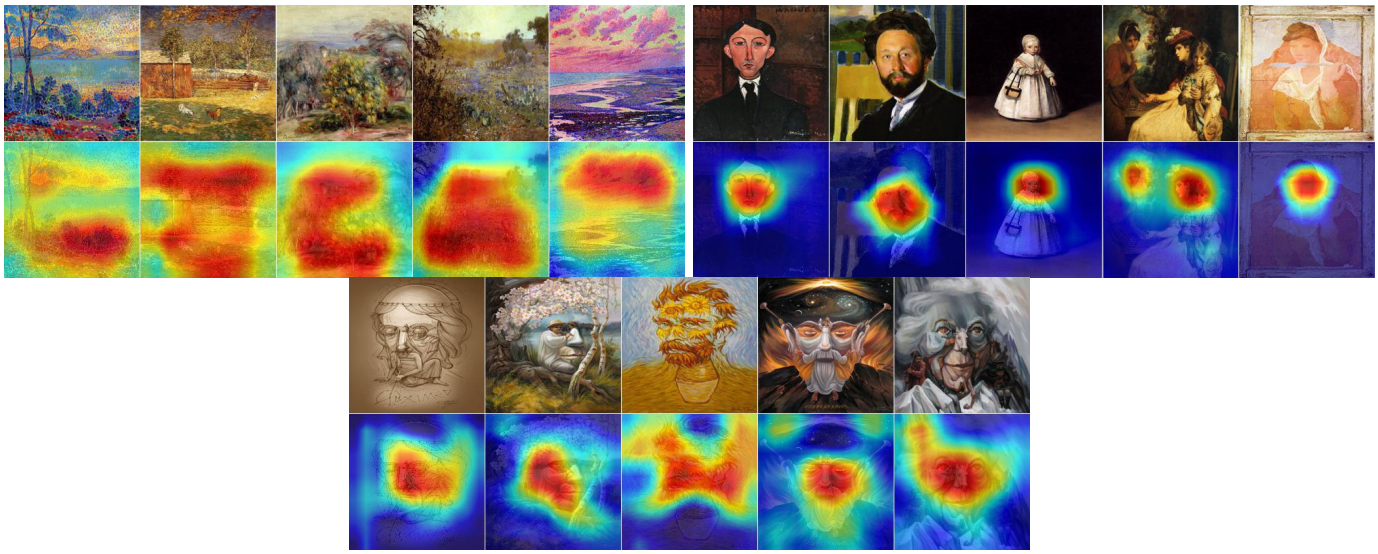

図 2: 実験 1 Grad-CAM の結果の例

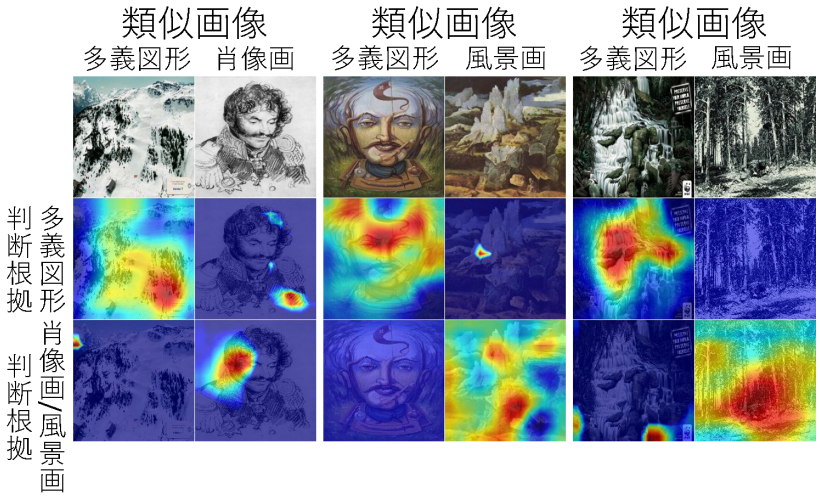

図 3: 実験 1 類似画像の識別

図 4: 実験 2 風景画を切り出して肖像画と識別された部分画像例

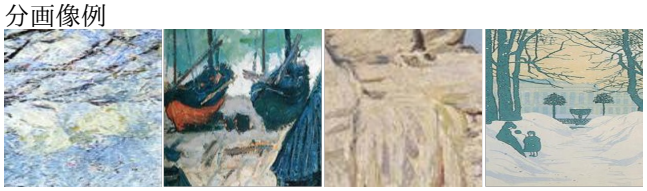

図 5: 実験 2 風景画を切り出して多義図形と識別された部分画像例

[5] Patrice Y. Simard, Dave Steinkraus, and John C. Platt. Best practices for convolutional neural networks applied to visual document analysis. 2003.

[6] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[7] WikiArt. `https://github.com/cs-chan/ArtGAN/tree/master/WikiArt%20Dataset`.