

図地反転図形の認識と生成に向けてのデータセット構築

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は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、計算機による多義図形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。

堀江ら [1] はこれまで人の顔と風景の多義図形の認識や、回転多義図形の認識、その判断根拠の解析を中心に研究してきた。その結果、人の顔と風景の多義図形において、92 % の精度で人工知能に多義図形とそうでない画像を識別させることに成功した。また、90 度回転によりうま・かえるに変化して見える画像について、94% 以上の精度で人工知能に両オブジェクトを認識させることができた。また、回転によってうさぎ・あひるに変化して見える画像について、アンケート調査と人工知能による認識の相関係数が 0.772 と非常に強い相関を持つことを示した。生成に関しては、人の顔と風景の多義図形を識別するモデルを使用し、風景画の中から多義図形らしい箇所を切り取る実験をしてきたが、多義図形らしくない画像が多く生成されることが課題であった。

そこで今発表では、はっきりとしたバイナリ画像で表される図地反転図形に着目し、その生成を目標とすることにした。そのために、まずは 3 辺を共有する 2 つのランダムな六角形を生成し、Ying-Miao Kuo らによる論文 [2] を元にして、外接矩形で背景領域を区切り、2 色で塗り分ける最適化を施したデータセットを構築するためのアルゴリズムを実装した。また、Convolutional Auto Encoder [3] を改変し、入力を 2 画像、出力を 1 画像にしたモデルを実装し、それを用いて深層学習で画像の合成が可能かを実験した。また、Ying-Miao Kuo らによって公開されている図地反転図形の結果より、考察用のテストデータセットを作成し、合成を試みた。



図 1: E.J. ルビンによるルビンの壺

2 先行研究

2.1 図地反転図形

反転図形とは多義図形的一种であり、同一図形でありながら 2 種の見え方が交代して現れる図形である。通常、図形を見続けていると、突然他の見え方が観察者の意志にかかわりなく生じ、以後この二通りの見え方が交互に生じる。図地反転図形とは、反転図形のうち図形の一定の領域が図に見えたり地に見えたりするような図形である。E.J. ルビンの盃と横顔の図形 (ルビンの壺 図 1) はその一例である。

ゲシュタルト心理学では、視覚野の連続する 2 つの領域で共有される境界の反対側にある手がかりの間の抑制的な競合が、図の割り当てプロセスにおいて重要な役割を果たすことが盛んに研究されている。低レベルの幾何学的特徴 (凸部、対称性、囲みなど) と高レベルの手がかり (注意、親しみやすさ、過去の経験など) の両方が競合プロセスに影響を与えることが報告されている。競争に勝つ側には図形が知覚され、反対側の領域には形のない地面が知覚される。これは、複数のオブジェクトを共有する境界線に沿って図の割り当ての競争が均等に行われる形で組み合わせることで、効果的な曖昧な図-地のイメージを作り出すメカニズムを示唆している。

2.2 形状マッチング

形状マッチングは、コンピュータビジョンにおいてよく研究されているトピックであり、輝度ベースの手法と特徴ベースの手法の 2 つに大別される。輝度ベースの手法は、各画素の強度を形状記述子として扱うもので、物

体の姿勢や画像の照明の変化に敏感である．これに対して，特徴量ベースのアプローチは，画像の幾何学的特性（輪郭，サンプルポイントなど）を用いて物体の形状を記述し，より良い性能を得ることが証明されている．しかし，このような従来の形状マッチングアルゴリズムは，オブジェクト間の形状の類似性のみを測定するものであり，有望な曖昧な図-地の画像を生成できないことが多い．

2.3 Auto Encoder

Auto Encoder（自己符号化器）とは，ニューラルネットワークの 1 つあり，エンコーダにより入力されたデータを一度圧縮し，重要な特徴量だけを残した後，デコーダにより再度もとの次元に復元処理をするアルゴリズムを意味する．このように，小さい次元に落とし込む作業を次元削減や特徴抽出と呼ぶが，オートエンコーダはそれだけでなく，生成モデルとしても用いられる．本実験では，入力層と出力層の間に，入力データの特徴量を捉える「畳み込み層」と，その特徴への依存性を減らす「プーリング層」を加えた Convolutional Auto Encoder を用いた．

3 提案手法

3.1 実験目的

本研究では，図地反転図形を深層学習で生成，分解することを最終目的としている．その為に，本実験ではまずは 2 つの 3 辺を共有する多角形で擬似的に図地反転図形を生成し，それらを分解することによってデータセットを生成した．また，それらのデータセットを後述する Convolutional Auto Encoder を改変したモデルで学習させ，擬似的な図地反転図形の生成が可能かを実験した．その後，意味を持つバイナリ輪郭画像を 2 枚モデルに入力し，図地反転図形が生成されるか確認した．

3.2 データセット作成

3.2.1 自己交差の無い 3 辺を共有する 2 つの六角形の生成

まずは，自己交差の無い 3 辺を共有する 2 つの六角形を生成する．左側の六角形を Pa ，右側の六角形を Pb ，合体後の多角形を P とすると， Pa ， Pb ， P がどれも自己交差していないことが必要である．図 2 に輪郭線のみを描画した多角形の例を示す．

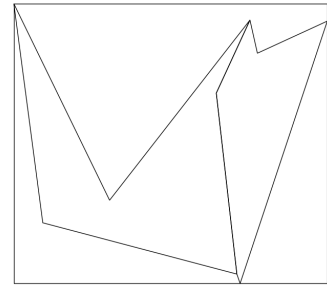


図 2: 輪郭線のみを描画した 3 辺を共有する 2 つの六角形の例



図 3: 入力画像 1 の例



図 4: 入力画像 2 の例

また，本実験では仮に左側の背景を黒色，六角形を白色，右側の背景を白色，六角形を黒色として，合成前のデータセットを作成した．図 3 に左側の六角形，図 4 に右側の六角形の合成前のデータセットの例を示す．

3.2.2 バイナリ塗り分け最適化アルゴリズム

複合物体の輪郭が $C = C_a \cup C_b$ であるとする．

バイナリ塗り分けの単純なやり方として，特定の物体の境界線を利用して画像を切り取り，輪郭の内側を黒（または白）で塗りつぶし，残りの領域を白（または黒）の背景として扱うという方法がある．このような単純な方法では，一方の物体の形状は完全に保持されるが，他方の物体の形状は大きく犠牲になり，共有される境界部分でしか認識できない．

この問題を解決するため，両方の物体の輪郭の完全性を最大化するために，画像の切り取りと 2 値化を同時に計算する共同最適化を Ying-Miao Kuo ら [2] は提案している．システムは，外接矩形， C_a ， C_b に従って，画像を



図 5: バイナリ塗り分け最適化後のデータセットの例

不連続な領域 $R = \{r_1, \dots, r_n\}$ に分割する．ここでは、 C_a と C_b で囲まれた領域をそれぞれ r_a と r_b と呼ぶことにする．次に、 V の各頂点が領域に対応する隣接グラフ $G = (V, E)$ を構築する．辺 $e_{i,j}$ は、2つの領域 r_i と r_j が互いに隣接するときに E に追加される．辺の重み $l_{i,j}$ は、 r_i と r_j の間の共有境界の長さを表す．式を最大化するような領域の2値色割り当て $B = \{b_1, \dots, b_n | b_i \in 0, 255\}$ を計算することで、最適な2値画像が生成出来るとしている．

$$B^* = \arg \max_B \left[\frac{1}{\|C_a\|} \sum_{\substack{e_{i,j} \in E \\ b_i \neq b_j}} l_{i,j} + \frac{1}{\|C_b\|} \sum_{\substack{e_{i,j} \in E \\ b_i \neq b_j}} l_{i,j} \right] \quad (1)$$

ここで、 $e_{i,j}^a$ と $e_{i,j}^b$ は、それぞれ r_a と r_b に入射するエッジを表す．本実験では、全ての塗り分けが網羅される十分な回数（今回は100回とした）ランダムに2値塗り分けし、最も (1) 式の値が大きい時の塗り分け方を採用した．

3.2.3 考察用データセットの作成

Ying-Miao Kuo ら [2] によって生成された公開されているデータセットをまとめた pdf より、考察用テストデータセットを作成した．元の画像はオブジェクトが中央に配置されているが、それらをそれぞれ左側、右側に寄せて正方形に切り取り、データセットを作成した．また、作成したデータセットをそれぞれ2値化した．

4 実験条件

図 6 に本実験で用いた Convolutional Auto Encoder を入力画像2枚、出力画像2枚に改変したモデルを、表 1 に本実験の実験条件を示す．

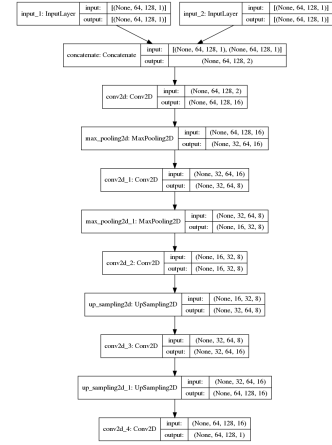


図 6: 実験に用いたモデル

表 1: 実験条件

データサイズ	256 × 256 × 1
活性化関数	ReLU
オプティマイザ	Adam
エポック	500
損失関数	binary cross entropy
バッチサイズ	32
Train 数	171 セット
Validation 数	19 セット
Test 数	20 セット

5 実験結果

図 7 に実験結果を示す．一段目が1枚目の入力、二段目が2枚目の入力、三段目が中間層の出力、四段目が出力画像、五段目がラベル画像である．これより、2つの六角形の色と形は保たれたまま合体していることがわかる一方、背景は復元されていないことがわかる．

6 考察

図 8 に学習したモデルに Train データを入力した出力結果例を示す．これより、Train 段階でも背景は復元されていないことがわかる．

図 9 に学習したモデルに考察用データセットを入力した出力結果を示す．一段目が1枚目の入力、二段目が2枚目の入力、三段目が中間層の出力、四段目が出力画像である．これより、図 7 と同じように、2つのオブジェクトの色と形は保たれたまま合体していることがわかる一方、背景は復元されていないことがわかる．

図 10 に学習したモデルに白黒反転させた考察用データセットを入力した出力結果を示す．一段目が1枚目の

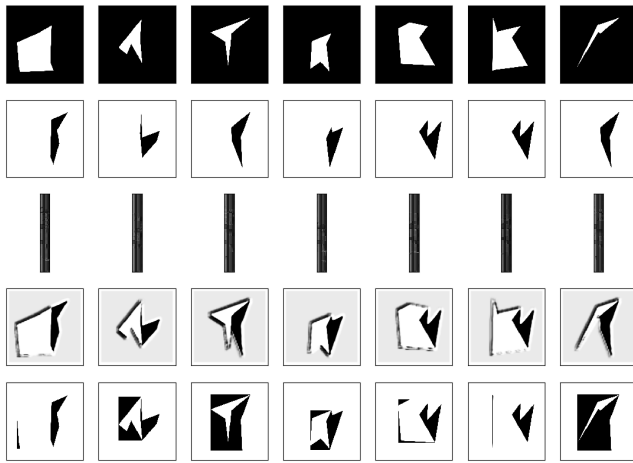


図 7: 実験結果

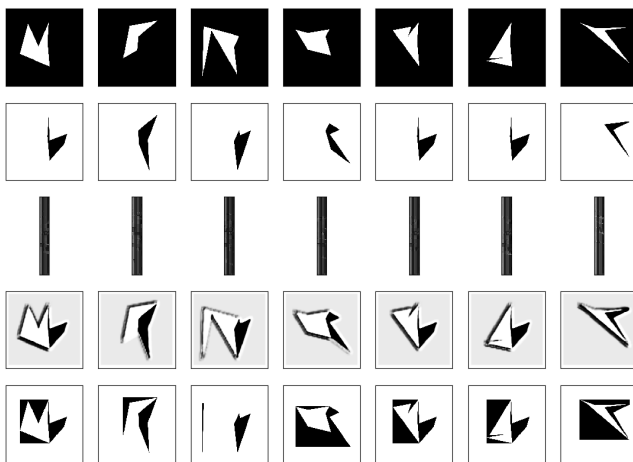


図 8: Train データを入力した出力結果例

入力，二段目が 2 枚目の入力，三段目が中間層の出力，四段目が出力画像である．これより，図 9 と同じように，2 つのオブジェクトの形と色は保たれたまま合体していることが出来ていることがわかる．よって，学習は色依存ではなく形依存でなされていることがわかる．

7 まとめと今後の課題

今発表では，図地反転図形の認識と生成に向けて，3 辺を共有する自己交差しない 2 つのランダムな六角形を生成した．また，Ying-Miao Kuo らによる論文 [2] を元にして，外接矩形で背景領域を区切り，バイナリで塗り分ける最適化を施したデータセットを構築するためのアルゴリズムを実装した．また，Convolutional Auto Encoder [3] を改変し，入力を 2 画像，出力を 1 画像にしたモデルを実装し，それを用いて深層学習で 2 つの異なるオブジェクトが描かれた画像の合成が可能かを実験した．そ

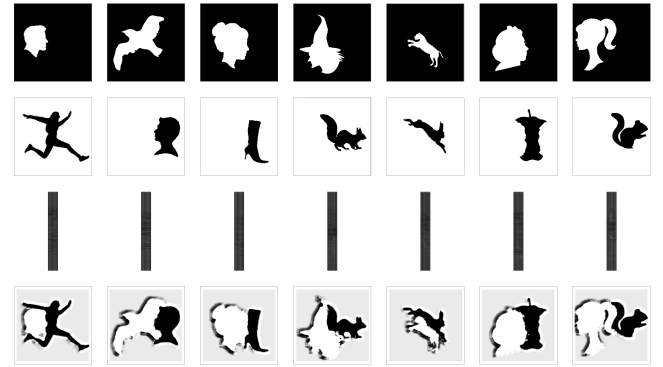


図 9: 考察用テストデータセットを用いた結果

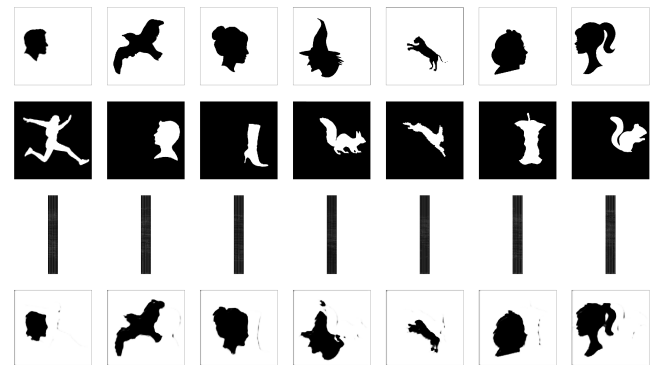


図 10: 白黒反転考察用テストデータセットを用いた結果

の結果，2 つの入力オブジェクトの色と形は復元されるが，背景は復元されないことがわかった．また，学習は色依存ではなく形依存でなされていることもわかった．

今後の課題としては，左右以外の組み合わせを学習させること，回転や位置を配慮しなくても適切な合成画像が作ることが出来るようにすること，背景色の学習方法を模索することが挙げられる．

参考文献

- [1] 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.
- [2] Ying-Miao Kuo, Hung-Kuo Chu, Ming-Te Chi, Ruen-Rone Lee, and Tong-Yee Lee. Generating ambiguous figure-ground images. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, No. 5, pp. 1534–1545, 2017.
- [3] G E Hinton and R R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504–507, July 2006.