

実物画像とイラスト画像の識別器導入による画像認識 Image recognition by introducing a discriminator between real and illustration images

麻野 泰地[†] 土屋 誠司[‡] 渡部 広一[‡]
Taichi Asano Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年、画像認識技術は多くの分野で活用されている。人間が視認した物体の名称を特定できるのと同様に、物体認識モデルは、画像に写っている物体の一般的な名称を推測することができる。被写体に一般的な名称がある画像は 2 種類あると考えられ、1 つは、実在する物体をカメラで撮影した画像であり、本稿ではこれを“実物画像”という。もう 1 つは、実在する物体の特徴を捉えて人間が作成した画像であり、本稿ではこれを“イラスト画像”という。画像認識モデルには CNN がよく用いられている。一般的に、実物画像中の物体認識を実現するためには、学習データにも実物画像を使用する。しかし、実物画像のみを学習したモデルでは、イラスト画像中の物体認識において正解率は低下してしまう。世の中にはイラスト画像も多く存在するため、モデルが画像の種類によらずに認識できることは重要である。しかし、実物画像とイラスト画像の両種に対して物体を認識できるモデルはまだ精度が低い。そこで本研究では、画像が実物画像かイラスト画像かを分類できる識別器を作成し、物体認識モデルに導入する手法を提案する。本研究では、実物画像を ImageNet^[1]から収集した。

2. 関連技術

2.1 ImageNet^[1]

ImageNet とは、スタンフォード大学の研究チームによって作成された、1400 万枚以上のカラー画像を持つ大規模なデータベースである。

2.2 CNN

CNN は画像認識でよく使われる深層学習アルゴリズムの 1 つである。CNN は中間層が、畳み込み層、プーリング層、全結合層で構成されている。

2.2.1 畳み込み層

畳み込み層ではフィルターを画像に作用させることで、エッジや色の変化などの画像の局所的な特徴量を抽出することができる。

2.2.2 プーリング層

プーリング層の役割は主にデータの圧縮である。前の層の出力である特徴マップ上でカーネルをスライドさせ、局所的な空間の代表値を取得する。プーリング層によって画像の小さな変化に対して強くなり、頑健性が獲得される。

2.2.3 全結合層

全結合層は、重みを掛けて全てのノードを結合する層のことである。CNN ではよく、出力層に近い層で用いられ、特徴マップを全て結合して、1 次元の数値を得る。

[†] 同志社大学大学院理工学研究科

[‡] 同志社大学理工学部インテリジェント情報工学科

2.2.4 AlexNet^[2]

AlexNet は、物体認識の精度を競う大会である ILSVRC (ImageNet Large Scale Visual Recognition Challenge) で 2012 年に優勝したことで知られている CNN ベースの画像認識モデルである。AlexNet の構造を図 1 に示す。

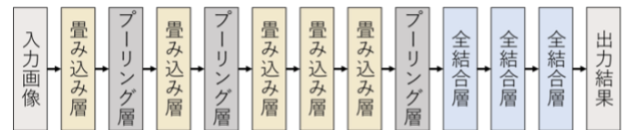


図 1 AlexNet の構造

3. 先行研究

本研究は竹中智哉氏の「CNN を用いたイラスト画像と実画像の両方を認識する学習方式^[3]」を先行研究とした。

3.1 先行研究の概要

先行研究では、実物画像とイラスト画像の両種に対応する物体認識モデルの学習方式が提案された。先行研究で提案された手法は主に 2 つある。1 つは、モデルの学習データに擬似イラスト画像のみを使用する手法である。擬似イラスト画像とは、実物画像に輪郭抽出、減色処理を行なった画像のことである。擬似イラスト画像の例を図 2 に示す。このモデルの正解率を調べる実験が行われた結果、実物画像のみ学習したモデルよりもイラスト画像に対する物体認識の正解率が向上したことが示された。もう 1 つは、学習データに実物画像と擬似イラスト画像を両方使用する手法である。この手法では物体認識モデルに実物画像と擬似イラスト画像を同じ枚数だけ学習させた。同様に正解率を調べる実験が行われた結果、テスト画像が実物画像とイラスト画像のどちらの場合であっても、一定の正解率を示した。



図 2 擬似イラスト画像の例

3.2 先行研究の問題点

先行研究で提案された両種の画像を学習したモデルは、実物画像のみを学習したモデルと比較すると、実物画像に対する物体認識の正解率が低下した結果を示した。また、擬似イラスト画像のみを学習したモデルと比較すると、イラスト画像に対する物体認識の正解率が低下した結果を示した。以上のことから、一つのモデルに両種の特徴を学習させることで、それぞれの画像に対する正解率が相対的に低くなってしまふことが、問題点として挙げられる。

4. 提案手法

先述の問題点を踏まえて、本研究では実物画像のみ学習したモデルと、擬似イラスト画像のみ学習したモデルを作成し、実物画像とイラスト画像の識別器を導入する手法を提案する。以降、単に識別器と呼ぶ。提案手法のフローを図 3 に示す。入力画像が実物画像であると分類された場合は、実物画像のみ学習したモデルを使って物体認識を行い、イラスト画像であると分類された場合は、擬似イラスト画像のみ学習したモデルを使って物体認識を行うようにする。この手法により、正解率が相対的に高いモデルを使用することができ、先行研究の手法よりも多くの画像に対して物体認識が可能になると考えた。

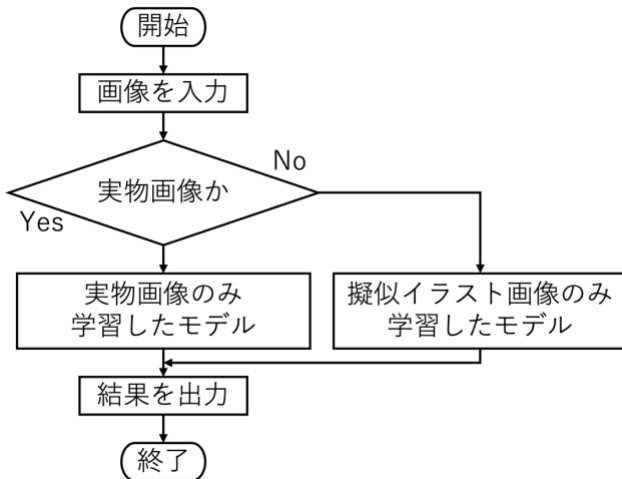


図 3 提案手法のフロー

4.1 使用したデータ

表 1 に本研究で使用したデータを示す。本研究では画像分類の対象を 30 クラスとし、物体認識用の学習画像は 30 クラス×900 枚用意した。イラスト画像 (テスト用) は収集が困難であったため、クラス数と枚数を絞って用意した。イラスト画像 (学習用) は識別器の学習用に別途用意した。

表 1 使用した画像データ

画像	枚数	用途
実物 (学習用)	30×900	物体認識モデル
擬似イラスト (学習用)	30×900	物体認識モデル
実物 (学習用)	50	識別器
イラスト (学習用)	50	識別器
実物 (テスト用)	30×100	物体認識モデル 識別器
イラスト (テスト用)	5×30	物体認識モデル 識別器

4.2 物体認識モデル

表 2 に物体認識モデルの情報を示す。モデルは AlexNet の出力層のノード数を 30 に変更した CNN 構造である。

表 2 物体認識モデルの学習枚数

モデル	学習枚数	
	実物	擬似イラスト
実物学習モデル	27,000	0
擬似イラスト学習モデル	0	27,000
両種学習モデル	27,000	27,000

4.3 識別器

表 3 に識別器の情報を示す。モデルは AlexNet の出力層のノード数を 2 に変更した CNN 構造である。実物画像 50 枚は、表 1 の実物画像 (学習用) 27,000 枚中から取得した。

表 3 識別器の学習枚数

モデル	学習枚数	
	実物	イラスト
識別器	50	50

5. 実験

本研究では、作成したモデルについて再現率と正解率を調べる実験を行なった。

5.1 実験結果

表 4 に識別器の再現率を示す。結果より、実際に実物画像であるデータに対しては、82.8%の割合で正しく予測し、実際にイラスト画像であるデータに対しては、95.3%の割合で正しく予測した。

表 4 識別器の再現率

モデル	再現率[%]	
	実物	イラスト
識別器	82.8	95.3

表 5 に物体認識モデルの正解率を示す。結果より、先行研究で提案された両種の画像を学習したモデルより、提案手法の方が多くの画像に対して正しく物体を認識したことが確認できる。しかし、正解した枚数を集計しカイ二乗検定を行なったところ、有意水準 5%で、先行研究の手法との間に有意差は示せなかった。

表 5 物体認識モデルの正解率

モデル	正解率[%]	
	実物	イラスト
実物学習モデル	73.1	46.7
擬似イラスト学習モデル	69.7	57.3
両種学習モデル	71.9	58.0
提案手法 (識別器導入)	72.7	58.7

5.2 考察

本研究では、擬似イラスト画像のみ学習したモデルがイラスト画像に対して想定より正しく認識できなかった。この原因として、擬似イラスト画像の生成方法がイラストの特徴を表現する上で不十分であることが考えられる。本実験で使用した擬似イラスト画像は単純な画像処理のみで生成しているため、デフォルメされたイラストの特徴を学習することはできない。よって、別の生成手法の検討が必要である。擬似イラスト画像がよりイラストに近い特徴を持つと、両種の画像を学習するモデルでは 1 つのモデルで多くの特徴を捉えることが必要となり学習が安定しなくなる可能性がある。この場合、本研究の提案手法がより有用な結果を示せるのではないかと考えられる。

参考文献

- [1] ImageNet, <https://image-net.org/>, 7.February.2024
- [2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [3] 竹中智也, CNN を用いたイラスト画像と実画像の両方を認識する学習方式, 信学技報, vol.118, no.492, AI2018-53, pp.1-5, 2019年3月.