

深層学習を用いた一般物体認識アルゴリズムによるマーカー認識 Marker Recognition by Object Recognition Algorithm Using Deep Learning

曾 彩華[†] 花沢 明俊[†]
Ayaka So Akitoshi Hanazawa

1. はじめに

生活の様々な場面で QR コード[1]や AR マーカー[2]などのマーカーが利用されている(図 1). QR コードは大容量の情報を表現できる 2 次元コードとして広く利用されている. AR マーカー, カラーバーコード[3]などは QR コードと比較すると表現できる情報量は少ないが, 情報の伝達だけでなく, 画像からマーカーを読み取ることで物体標識や位置検出にも利用されている. これらのマーカー認識では, 読み取りに専用のライブラリやアプリを必要とするものが多く, 使用者が環境や用途に応じてマーカーのデザインを自由に変えることは困難である. さらに, 既存のマーカーは切り出しシンボルと呼ばれるものをマーカー中に含め, それによってマーカーの検出を行っているものが多く, この切り出しシンボルが写っていない画像からの検出は難しいという問題点も挙げられる.

そこで, 本研究では物体標識や位置検出に特化したマーカーをデザインし, 従来のマーカーを物体標識や位置検出に使用する場合の問題点を解決することを目的とする. マーカーの学習には, 一般物体認識アルゴリズムである SSD(Single Shot Multibox Detector) [4][5]を用いた. 本研究の特徴として, 検出に特定のライブラリを必要としないこと, マーカーのデザインやバリエーションを用途に応じて変更できること, 遮蔽に強いマーカーをデザインできることが挙げられる.



図 1 既存のマーカー

2. 物体検出手法

SSD の概要, 本研究の物体検出手法について説明する.

2.1 SSD(Single Shot Multibox Detector)

SSD(Single Shot Multibox Detector) は, マルチタスク学習を用いた一般物体検出アルゴリズムであり, 対象画像から高スピードで多種類の物体を検知することができるのが大きな特徴である. 特定の物体を覚えさせ, その特定の物体を検知することも可能である. SSD には, 入力画像の大きさを 300×300 ピクセルにリサイズして入力する SSD300 と, 512×512 ピクセルで処理をする SSD512 の 2 パターンがある. 本研究では SSD300 を用いて学習を行う.

また, 画像から物体のバウンディングボックスを求める際に, バウンディングボックスの情報を出力させるのではなく, 「デフォルトボックス」と呼ばれる定型的な長方形をあらかじめ用意しておき, この長方形のデフォルトボ

ックスをどのように変形させればバウンディングボックスになるのかという情報を出力させる. このデフォルトボックスを変形させる情報をオフセット情報と呼ぶ. SSD による物体検出結果の例を図 2 に示し, 以下に物体検出の 6 ステップの流れを示す.

- ① 画像を 300×300 ピクセルの大きさにリサイズする.
- ② 画像に対して様々な大きさと縦横比のデフォルトボックスを用意する. SSD300 の場合は, 8732 個のデフォルトボックスを用意する.この時, デフォルトボックスは入力画像に関係なく, どの画像に対しても全く同じものを用意する.
- ③ 画像を SSD のネットワークに入力し, 8732 個のデフォルトボックスそれぞれに対して, デフォルトボックスをバウンディングボックスへと修正するオフセット情報とデフォルトボックスが持つ各クラスの物体である信頼度に関する情報を出力する.
- ④ 8732 個のデフォルトボックスのうち, 信頼度が高いものを上位から 200 個取り出す. デフォルトボックスには各クラスの信頼度が付随しているが, デフォルトボックスに対応するラベルはの中で信頼度が最も高いクラスになる.
- ⑤ オフセット情報を使用し, デフォルトボックスをバウンディングボックスへと変形する. ここで, ステップ 4 で取り出した 200 個のデフォルトボックスのうち, バウンディングボックスの重なりが大きいもの(同じ物体を検出していると思われるもの)が複数ある場合は, 最も信頼度が高いバウンディングボックスだけを残す.
- ⑥ 最終的なバウンディングボックスとそのラベルを出力する. ステップ 6 では信頼度の閾値を決め, 閾値以上の信頼度を持つバウンディングボックスのみを最終出力とする. 本研究では閾値を 0.5 とした.

また, 本研究は少ないマーカータグの訓練用データから精度の良い SSD のモデルを作成するために, 学習済みの VGG16 モデルを用いて転移学習を行なった SSD を用いて検出実験を行った.



図 2 SSD を用いた物体検出結果の例

(Liu et al., SSD: Single Shot MultiBox Detector, European conference on computer vision, pp. 21-37 (2016) [4] より図 5 の一部を転載)

2.2 本研究で使った物体検出手法

本研究では, 2 つの物体検出手法を用いて, マーカータグの検出を行い, その精度を比較することで, 最も適したマーカータグ検出手法の検討を行った.

[†]九州工業大学 工学府

Graduate School of Engineering, Kyusyu Institute of Technology

① 1 段階検出

この方法では SSD が検出器・分類器のどちらも担っている。

② 2 段階検出

SSD を検出器，ファインチューニングを用いて設計した深層学習ネットワークを分類器として検出を行った。本研究では，学習済みモデルである ResNet34 と VGG16 を用いた。

3. 実験準備

この章では，本研究で使ったマーカータグ，データセット，実験内容の説明を行う。

3.1 マーカータグ

本研究では，様々な状況で使用することができる自由度の高いマーカータグを自作した。作成したマーカータグを図 3 に示す。マーカー部分は，外径 28mm，内径 24mm となっている。作成したマーカータグには以下に示す特徴がある。

- ・赤外線カメラなど色の識別が行えないようなカメラからも検出を行う場合を想定し，マーカータグは白と黒の 2 色で作成した。マーカー部は 5 つに分かれている。
- ・マーカータグの上部には，取り付けの際に使用することができるように突起部を作った。この突起部もマーカーの一部に含めることで，マーカータグの上下を識別する際にも役立つのではないかと考える。
- ・背景とマーカーの境界を分かりやすくするために，マーカーの周りを黒で縁取りした。
- ・左右の認識精度向上のために，をマーカータグの右下に切り欠きを作った。

5 つそれぞれの部分における色の組み合わせを変えることで全 32 種類のマーカータグを作成した(図 4)。



図 3 作成したマーカータグの一例

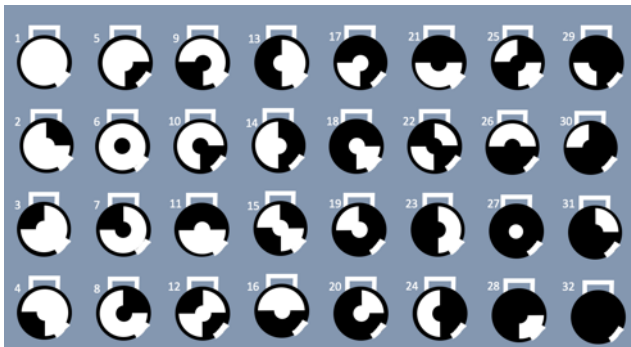


図 4 本研究で使ったマーカータグ一覧

3.2 データセット

本研究ではマーカータグをぬいぐるみに付け，32 種類全てのマーカータグについて 50 枚ずつ撮影した合計 1600 枚の画像を用いてデータセットを作成した。撮影にはスマートフォンのカメラを使用した。撮影した全ての画像に対して，labelImg[6]を使用し，アノテーションした結果を含む xml ファイルを作成した。labelImg とは物体検出のデータを作成するためのアノテーションツールである。アノテーションの様子の一例を図 5 に示す。

本研究では，撮影した 1600 枚の画像を基本とし，これらを 90 度，180 度，270 度回転させた画像を加えることで，データセットのデータ拡張(Data Augmentation)を行った。さらに，解像度を 2 段階に分けて落とした画像を使用して，さらにデータ拡張を施したデータセットも準備した。このようにデータ拡張したデータも含め，SSD を用いた学習では，実験に応じて図 6 に示す 3 種類のデータセットを使用する。さらに，2 段階検出を行う際の分類器作成用に撮影した画像からマーカータグの部分のみを切り出したデータセットも作成した(図 7)。データセットに含まれる画像の種類，枚数は SSD の学習のために作成したデータセットと同様である。表 1 には，作成したデータセットの詳細を記載する。

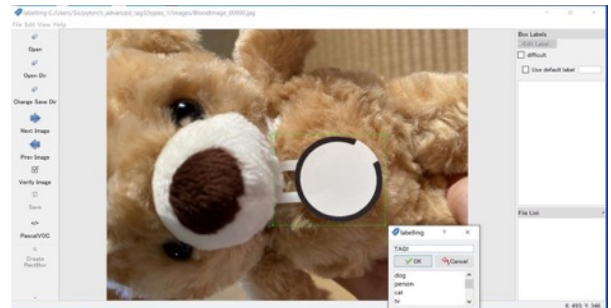


図 5 labelImg を用いたアノテーション



図 6 作成したデータセットに含まれる画像の一部上から順に撮影した元画像，回転させた画像，1 段階解像度を落とした画像，2 段階解像度を落とした画像を示す。これらの画像はデータセット①から③に含まれている。

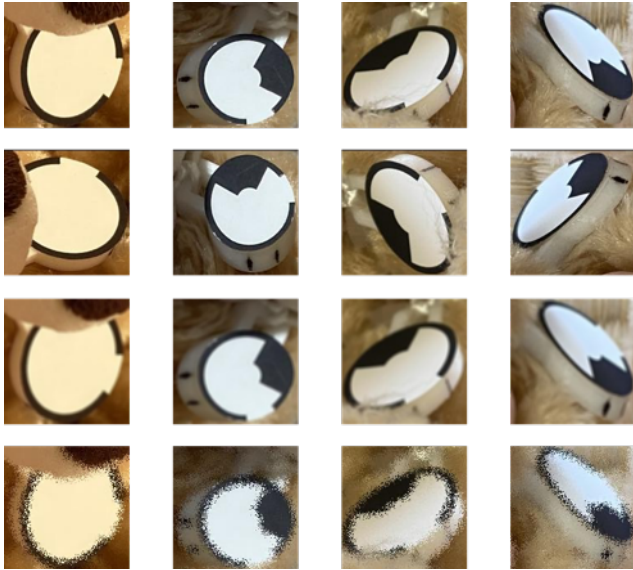


図 7 マーカータグを切り出して作成したデータセットの一部

表 1 作成したデータセット

データセット①から③は撮影した画像をデータ拡張した画像を含み、データセット④から⑥は撮影した画像からマーカータグのみを切り出した画像を含んでいる。

データセット①	訓練用データ：1296 枚 検証用データ：144 枚 テスト用データ：160 枚 合計 1600 枚 (50 枚×32 種類)
データセット②	訓練用データ：5184 枚 検証用データ：576 枚 テスト用データ：640 枚 合計 6400 枚 (200 枚×32 種類)
データセット③	訓練用データ：15552 枚 検証用データ：1728 枚 テスト用データ：1920 枚 合計 19200 枚 (600 枚×32 種類)
データセット④	訓練用データ：1280 枚 検証用データ：160 枚 テスト用データ：160 枚 合計 1600 枚 (50 枚×32 種類)
データセット⑤	訓練用データ：5120 枚 検証用データ：640 枚 テスト用データ：640 枚 合計 6400 枚 (200 枚×32 種類)
データセット⑥	訓練用データ：15360 枚 検証用データ：1920 枚 テスト用データ：1920 枚 合計 19200 枚 (600 枚×32 種類)

3.3 実験内容

実験の精度評価には 10 分割交差検証を用いた。また、どのデータセットを用いた学習が本研究に最も適しているか検討するために t 検定を行った。本研究では有意水準を 0.05 とした。

3.3.1 実験 I

2.2 の①で述べた SSD のみを用いた 1 段階検出を行う。

3.3.2 実験 II

2.2 の②で述べた SSD を検出器、ファインチューニングを用いて設計したネットワークを分類器として用いる 2 段階検出を行う。1 段階目の SSD では、マーカータグの種類の識別まで行わず、検出だけを行うことでマーカーが写っているか否かの 2 分類とすることができ、学習データを増やすことができるため、より高精度な検出が行えるのではないかと考える。なお、1 段階目の SSD の学習にはデータセット①を用いた。

4 実験結果

4.1.1 では実験 I の結果を、4.1.2 では実験 II の結果を述べ、4.2 では t 検定を行った結果を述べる。

4.1.1 実験 I

検出を行った結果、図 8 に示す結果が得られた。また、それぞれのデータセットを用いた時の正解率とその平均は表 2 に示す。

表 2 正解率とその平均

表上部の数字は 10 分割交差検証の何回目の試行であることを示している。表中の数値は各試行における正解率を示している。

	1	2	3	4	5	6	7	8	9	10	平均
データセット①	0.950	0.871	0.773	0.830	0.811	0.926	0.873	0.807	0.859	0.894	0.859
データセット②	0.865	0.909	0.908	0.861	0.917	0.923	0.906	0.890	0.883	0.863	0.892
データセット③	0.954	0.890	0.879	0.866	0.869	0.928	0.922	0.916	0.891	0.880	0.900

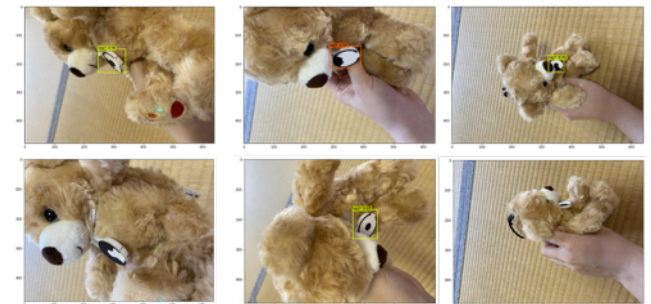


図 8 検出結果の例

上段は正しく検出できた結果を示し、下段はマーカータグの未検出や誤検出など検出に失敗した例を示す。

4.1.2 実験 II

1 段階目の SSD を用いた検出結果を表 3 に示す。その後、マーカータグの種類の識別を行った結果、表 4 に示す結果が得られた。

表 3 1 段階目の正解率とその平均

	1	2	3	4	5	6	7	8	9	10	平均
データセット①	0.9938	0.9756	0.9816	0.9755	1.0000	0.9938	0.9877	1.0000	1.0000	0.9938	0.9902

表 4 2 段階目を合わせた時の正解率とその平均

		1	2	3	4	5	6	7	8	9	10	平均
ResNet34	データセット④	0.4907	0.4329	0.4172	0.4049	0.4563	0.4793	0.4321	0.4188	0.4591	0.4286	0.4419
	データセット⑤	0.8944	0.7805	0.7853	0.7239	0.7125	0.9255	0.8951	0.8125	0.8050	0.8385	0.8173
	データセット⑥	0.9690	0.8719	0.8712	0.9080	0.9438	0.9814	0.9692	0.9625	0.9497	0.9690	0.9396
VGG16	データセット④	0.6211	0.4024	0.4724	0.4233	0.5563	0.5714	0.6729	0.3688	0.6352	0.4969	0.5221
	データセット⑤	0.9441	0.8232	0.8528	0.7546	0.8438	0.9379	0.9568	0.8563	0.9182	0.8385	0.8726
	データセット⑥	0.9752	0.8719	0.8834	0.8651	0.9188	0.9690	0.9754	0.9563	0.9434	0.9503	0.9309

4.2 t 検定

t 検定を行った結果、表 5 に示す結果が得られた。

表 5 t 検定の結果

表を横に見た時、縦軸に記載されている手法の方が精度が良い場合は青、横軸に記載されている手法の方が精度が良い場合はオレンジで数字を記載している。それ以外の箇所は有意差が見られなかったところである。

	SSD①	SSD②	SSD③	ResNet④	VGG④	ResNet⑤	VGG⑤	ResNet⑥	VGG⑥
SSD①		0.094	0.054	0.000	0.000	0.168	0.610	0.002	0.005
SSD②	0.094		0.554	0.000	0.000	0.006	0.376	0.006	0.023
SSD③	0.054	0.554		0.000	0.000	0.004	0.246	0.022	0.073
ResNet④	0.000	0.000	0.000		0.032	0.000	0.000	0.000	0.000
VGG④	0.000	0.000	0.000	0.032		0.000	0.000	0.000	0.000
ResNet⑤	0.168	0.006	0.004	0.000	0.000		0.087	0.000	0.000
VGG⑤	0.610	0.376	0.246	0.000	0.000	0.087		0.013	0.029
ResNet⑥	0.002	0.006	0.022	0.000	0.000	0.000	0.013		0.651
VGG⑥	0.005	0.023	0.073	0.000	0.000	0.000	0.029	0.651	

5 考察

5.1 実験 I

データセット①とデータセット③を用いて行った結果を比較すると正解率が約 4% 向上しているという結果からデータ拡張が正解率の向上に貢献しているということが分かる。データセット①を利用した時も高い正解率を得ることができていることから、SSD を用いた学習はマーカータグの検出に有効であると言える。また、データセット③を用いて作成したマーカータグの検出器を用いて、近距離・高画質な検出がしやすい理想的な環境での検出を行ったところ、図 9 に示す結果が得られ、約 9.7% の認識精度を得ることができた。

また、誤検出や未検出が起きたマーカータグのデザインを確認すると、誤検出が起きやすいマーカータグには左右・上下対称のものが多いという特徴があるということが分かった。このことから、誤検出が起きやすい組み合わせに関しては、実際に使用する場合はどちらかのみを利用するなど工夫する必要があると考えられる。



図 9 検出結果の例

5.2 実験 II

1 段階目のマーカータグの検出は精度が約 9.9% であることから、ほとんど全てのマーカータグを検出できていることが分かる。実験 I で行った SSD のみを用いたマーカータグの検出時は検出できていなかったマーカータグを検出できるようになっていることから、より様々な角度で撮影されたマーカータグの検出に適応できていると考えられる。

2 段階目のマーカータグの分類では、ResNet34 と VGG16 の認識精度を比較した時、訓練データが少ない場合は、わずかに VGG16 の方が精度が良いが、データセット⑥を使用した時は同等の精度が得られた。訓練データが少ない場合に VGG16 の方が精度が良かった理由として、VGG16 の方が抽出する特徴量が多いということが挙げられる。

5.3 実験 I・実験 II の比較

最後に t 検定を行った結果の考察を行う。実験 I の SSD を用いた検出は、正解率に若干の差はあるが、t 検定の結果から有意差は見られなかった。実験 II で行った 2 段階検出では、データセット④を使用した時のみ有意差が見られ

たが、その他のデータセットを使用した時は、学習済みデータの違いによる有意差は見られなかった。表 4 から正解率が最も高いのはデータセット⑥と ResNet34 を組み合わせた 2 段階検出であり、次いでデータセット⑥と VGG16 を組み合わせた 2 段階検出の正解率が高い。表 5 から、この 2 つの学習条件の結果に有意差は見られないが、データセット⑥と ResNet34 を用いたファインチューニングがデータセット⑥と VGG16 を用いたファインチューニングと比べ、他の組み合わせに対して有意差を持っていることが読み取れる。このことから、本研究で最も高い精度でマーカータグの検出を行えた手法とデータセットの組み合わせは、データセット⑥と ResNet34 を組み合わせた 2 段階検出であると言える。

6 おわりに

自作のマーカーを SSD を用いて学習し、約 9.7% の認識精度を得ることができた。このことから、用途や状況に応じてデザインを変更できる自由度の高いマーカー利用の方法論を構築できたとと言える。また、マーカータグの一部が写っていない画像からも正しく検出が行えたことから、本マーカーは遮蔽にも対応できていると考えられる。SSD を用いた 1 段階での検出実験では約 9.0% の認識精度が得られ、2 段階での検出実験では約 9.3% の認識精度が得られた。また、理想的な環境で撮影した画像を用いた検出実験では約 9.7% の認識精度が得られたことから、本研究で作成したマーカータグの検出に SSD は有効であると言える。さらに t 検定の結果から、本研究において最も認識精度が高いと言えるデータセットと手法の組み合わせは、データセット⑥(訓練用データ 15360 枚、検証用データ 1920 枚、テスト用データ 1920 枚)を使用し、ResNet34 を学習済みモデルとしてファインチューニングを行った 2 段階検出であると分かった。SSD のみを使用した 1 段階検出でも高い精度が得られていることから、実際に本マーカー検出システムを使用する場合は、実装のしやすさなどを考慮し、使用環境に応じて検出手法を選ぶ必要があると考える。本マーカー検出システムの活用例として、動物の行動観察、保育園や幼稚園などでの児童の安全確認、工場での部品や在庫の管理などが考えられる。今後はマーカータグのパターンの精査を行うことによってさらに精度を高めていきたい。

参考文献

- [1] DENSO WAVE, “QR コードとは”, 2022, <https://www.denso-wave.com/ja/adcd/fundamental/2dcode/qrc/index.html>, (参照 2022-06-20)
- [2] 富士通, “AR マーカーについて”, 2014, <https://software.fujitsu.com/jp/manual/manualfiles/m150001/b1ws1106/04z200/aropec/topics/c-iar-ope-2-2.html>, (参照 2022-06-20)
- [3] 株式会社 SHIFT “カラーバーコード「カメレオンコード」とは”, 2011, <https://www.shift-2005.co.jp/sdk.php>, (参照 2022-06-20)
- [4] Wei Liu, Dragomir Angueloy, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, “SSD: Single Shot MultiBox Detector”, European conference on computer vision, pp. 21-37 (2016)
- [5] 小川雄太郎, “Pytorch による発展ディープラーニング”, マイナビ出版 (2019)
- [6] Tzatalin, “LabelImg”, 2015, <https://github.com/tzatalin/labelImg>, (参照 2022-06-20)
- [7] チーム・カルボ, “物体・画像認識と時系列データ処理入門”, 秀和システム (2021)