

再現度の低い合成データを用いた検出器構築のための 同類化変換についての検討 ～麻雀牌検出を例題として～

鈴木 涼介^{1,a)} 大園 忠親^{1,b)} 新谷 虎松^{1,c)}

概要: 機械学習による物体検出器の構築には大規模データセットが求められる。物体検出の応用は多岐にわたり、各分野において物体検出器を構築することは、データセット構築の観点において非常に困難である。一般的に大規模データセットの構築はコストが高く、データが希少であるほどそのコストは高まる。解決手法の1つに、CGを用いた合成データ生成がある。しかし、CGで現実と酷似したデータを再現することは困難である。そこで本研究では、合成データと検出対象データの双方に同類化変換を適用することで、CGから効果的に検出器を構築することを目的とする。物体検出の応用領域として、麻雀牌検出に取り組む。本稿では変換した合成データを利用した麻雀牌検出器構築手法について述べる。

キーワード: 合成データ, 検出器, 同類化, 麻雀

On Homologous Transformation for constructing detector based on Synthesized Unrealistic Data

Abstract: Constructing an object detector by machine learning needs a large-scale data set. Object detection has many applications, and it is very difficult to construct an object detector in each field. The construction of large datasets is more expensive as data is rarer. One of the solutions is to generate synthetic data using CG. However, it is difficult to synthesize data similar to real data. Therefore, the purpose of this research is to construct a detector effectively from CG by applying the homologous transformation to both the synthesized data and the detection target real data. We engage in constructing a mahjong tile detectors as an application area of object Detection. This paper describes a method for constructing a mahjong tile detector using the converted composite data.

Keywords: Synthesized data, Detector, Homologous, Mahjong

1. はじめに

物体検出器は、画像中の物体の位置と種類を取得するためのシステムであり、多くの応用領域を持つ重要な技術である。全ての応用領域で利用できる汎用的な検出器を構築することは、検出器の複雑性が増すため現実的でない。応用領域ごとに適した検出器を構築することが望ましい。検

出器を構築する際の多様なパラメータ設定を簡略化するため、機械学習による自動的なパラメータ探索が利用できる。機械学習による物体検出器の構築には大規模データセットが必要となる。データセットの構築は、データの収集と、収集したデータへのラベル付け作業に分けることができる。データは、検出対象が含まれる画像であり、ラベルは画像中の検出対象の位置と、そのクラスである。大量のデータの蓄積は一般的に容易でなく、またラベル付け作業は煩雑である。双方にかかるコストは非常に大きく、物体検出器の構築における大きな課題である。

大規模データセットを構築するために、コンピュータグラフィックス (CG) によるラベル付きの人工的な画像デー

¹ 名古屋工業大学大学院情報工学専攻
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Japan

a) suzuri@toralab.org
b) ozono@tolarab.org
c) tora@toralab.org

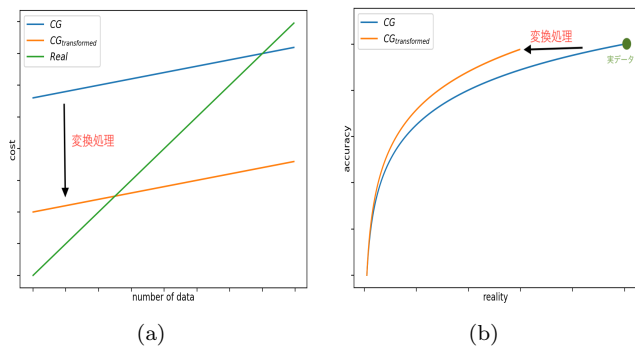


図 1 データの性質に依存した特徴の比較

タ（合成データ）生成手法が知られている。本研究では合成データを利用し、簡易に検出器を構築する手法について研究している [1]。しかし、CG データで現実のデータ（実データ）を再現することは困難な問題であり、粗悪なデータは物体検出器の精度に悪影響を及ぼす。CG の再現度を高めるのではなく、検出対象である実データと、学習データである合成データの双方に事前処理として同類化変換を適用することを検討する。本研究における同類化変換とは、性質の違う学習データと検出データのそれぞれを、学習器にとって理想的な状態に変換することを指す。実データと合成データの間には、幾何学的な違いは見られないことに注目する。データ間の性質の差を埋めるには、ノイズの有無や照明の違いなど、ピクセル毎の細かな違いを修正する事が重要である。これらの違いは、画像全体にノイズの除去や色調補正などのフィルタを適用する事で解決できる。そこで、双方のデータを似たような絵柄の画像に変換する事ができれば、CG から効果的に物体検出器を構築可能になると考えた。これは、特徴量の標準化処理であると考えられる。

本稿では、麻雀牌検出を題材とし、合成データを学習した検出器を用いて実データを検出する。合成データと及び実データには同類化処理を適用し、検出器の精度の向上を図る。本稿の以降の構成を示す。2. では、実データと合成データそれぞれのデータセット構築にかかるコストを比較し、合成データの利点を示す。3. では、合成データの生成手法及び合成データの利用に関わるドメイン適応について概観する。4. では、本研究で取り組む麻雀牌検出タスクに関して説明する。5. では麻雀を題材とした合成データ生成の手順について詳細に説明する。6. では、合成データおよびそれらに適用する同類化の実用性を評価し、考察する。最後に 7. で本稿をまとめる。

2. データセット構築にかかるコスト

合成データ、変換した合成データ、実データを比較した時、データセット構築にかかるコストの性質が異なる。それらの関係を簡潔に表したグラフを 1 に示す。(a) はデータの数に対するコスト、(b) はデータの再現度に対する検

出器の精度を示す。合成データは、データ数に対して初期コストが大きい。初期コストには現実環境をシミュレーションするための素材の収集や、各種設定が挙げられる。初期コストは、合成データの再現度に比例して大きくなる。データセット構築環境を整えてしまえば、1つのラベル付きデータ生成にかかる時間は実データに比べて短い。時間だけでなく、人的コストをかけることなく、自動で生成できる。更に、同類化処理により再現度の低いデータで効率的に検出器を構築できるため、初期コストが低くなる。一方実データでは、合成データと比べ初期コストが低い。しかし、1つあたりのデータにかかるコストは大きい。データ収集に関しては、実環境で希少なデータほどコストが大きい。ラベル付けに関しては、画像中の検出したい物体数が多く、かつその物体のクラスが判断し難いほどコストが大きい。作業負担軽減のため、labelImg^{*1}のようなラベル付け補助ツールがある。しかし尚ラベル付けは単調かつ負荷が高い作業であるためコストが高い。負荷が高いことから、人手で付与されたラベルは信頼性が問われることもある。また、実データの収集に関して、データの偏り及び網羅率が問題となる。偏りは、頻出する事例ほどデータが蓄積され、希少な事例ほどデータが少ないため生じる。網羅率は、データ収集の過程で得られない事例の数だけ低くなる。データの偏りが大きく、網羅率が低いほど質の悪いデータセットとなることが予想される。一方合成データは、任意の事例のデータを任意の数だけ生成することができる。データの偏りを無くし、また予測し得る事例を生成することで網羅率を高めることができる。

本研究では、CG の合成データを利用してデータセットを構築することを目的としている。先に述べたとおり、課題は初期コストが大きいことである。そこで、1. で述べた通り、CG の再現度を高めるのではなく、双方のデータを同類化し、検出器にとって都合のいいデータに変換する。CG の再現にかかるコストは少なくなるため、合成データ生成にかかる初期コストは減少する。

3. 関連研究

3.1 シミュレーションによる画像生成

シミュレーションによる画像生成は主に画像を切り貼りする手法と、CG を用いる手法の 2 つがある。多様な状況の実環境をシミュレーションするために、それぞれの手法は異なるパラメータの制御を行う必要がある。

画像を切り貼りする手法は、検出対象である前景画像を、検出対象が存在する背景画像に貼り付けて画像を生成する。貼り付ける位置や境界線の処理により写実度が左右される。CG と違い、オブジェクトのボーンやテクスチャ、オブジェクトが存在する環境の様々な要素を用意する必要

*1 <https://github.com/tzutalin/labelImg>

は無い。Dwibediら[2]は、様々な視点やスケールの前景を背景画像に貼り付け、多様な合成で画像を生成した。また、前景を貼り付ける際に周辺画素との不一致を解消した。画像の切り貼りをするには、様々な角度や大きさの前景が含まれる画像と背景画像を収集しなければならない。

CGを用いる手法は、検出対象が存在する環境をCGモデルによりシミュレーションし、レンダリングにより画像を生成する。Shrivastava[3]らは、実データを用いて、合成データを生成する際に必要なパラメータを自動調整することで写実度の高い画像を生成する生成モデルを開発した。しかし、生成モデルを学習させるには大量の実データが必要となる。また、画像の写実度はオブジェクトのパラメータに限らず、テクスチャやボーンの設定により大きく異なる。本研究では麻雀という限定された状況である。検出対象である麻雀牌や、麻雀牌が存在する環境においてオブジェクトやテクスチャが予期できない多様性を持つことはない。例えば、麻雀牌は表面に描かれた文字や図形により34種類に判別できるが、それらをテクスチャとして収集することは容易である。従って、麻雀牌や雀卓などのオブジェクトを作成し、麻雀をプレイする環境をシミュレーションすることができる。

本稿では、前景・背景のデータ収集にコストがかかる画像の切り貼りはせず、CGを用いて麻雀のデータ収集を行う。麻雀のデータ収集においては、前述したように、CGを用いる手法がより簡便であると考えられる。CGを、麻雀牌や雀卓といった簡易なオブジェクトを生成し、実環境をシミュレーションする。

3.2 ドメイン適応

豊富な学習データが存在するドメインで学習した学習器を、学習データが少ないドメインに対して用いることがある。似通ったドメインで事前学習させることで、未学習のドメインに対して高い精度を得ることができる。この手法はドメイン適応と呼ばれる。ドメイン適応は転移学習の一種である。ドメイン適応を行う際、ドメイン間の特徴量の差により性質の差異が生じるため、学習器の精度低下が生じることがある。差が生じる原因の一つとして、環境光や外観の変動などによる特徴の差異が挙げられる。例えば、日中の状況を学習した検出器は、夜や霧がかった状況の検出に対して未学習である。この時、学習データと検出対象の特徴量が一致しないために性質の差異が生じる。差異を定量的に扱うために、統計的手法を用いたアプローチが知られている[4]。学習データと検出対象の統計量の分布の差を最小化することで、ドメイン間の性質の差異を近似できると考えられている。本研究では、あるドメインを異なるドメインに完全に適応させるのは困難と考えた。そこで、両方のドメインを検出器にとって都合のいい理想の状態(ドメイン)に変換する手法を検討する。



図2 距離による検出率の変化

4. 麻雀牌検出

本稿では、実環境の麻雀牌検出をタスクとして取り上げる。先行研究として、麻雀における得点計算を支援するため、麻雀牌の認識を行うシステムを開発している[5]。並んだ麻雀牌を切り抜く矩形検出器と、切り抜いた画像を分類する分類器をカスケード接続することで軽量の麻雀牌認識を実現した。しかし、矩形検出精度及び分類精度が撮影時の環境により低下することが欠点として挙げられる。特に矩形検出器が環境に頑健でない。矩形検出器は、検出時の特徴量として麻雀牌の輪郭を使用している。輪郭の中でも特に麻雀牌の境目の輪郭は撮影時の角度や環境光によって見た目が変わるため、検出が不安定であった。そこで本稿では、表面に描かれた文字や図形も特徴量として用いる。それらの特徴量は環境に左右されにくく、環境に頑健な検出器を構築することができると期待できる。検出器として、機械学習を用いた検出器であるSingle Shot Multibox Detector(SSD)[6]を用いる。機械学習には、多様な状況を網羅した大量の学習データが必要とされる。麻雀牌検出のデータセットは存在しないため、データセットを構築する必要がある。実データデータの収集やラベル付けにはコストがかかる。簡易にデータセットを構築するため、合成データの生成を行う。データ生成の手法として、CGを用いる。CGを用いて様々な環境をシミュレーションすることにより、環境に頑健な検出器を構築する。また、このシステムはオフライン環境下におけるモバイル端末での実装を目的としている。そのため、複雑性が低く、処理が軽量の検出器が必要である。画像入力を大きくするに従い検出器は複雑性を増し計算量が増加するため、小さな画像で検出できることが望ましい。モバイル端末で撮影した画像を小さな画像にダウンサンプリングすることを想定する。小さな画像において、麻雀牌の識別は困難なタスクである。麻雀牌の萬子は表面に描かれた2文字のうち、1文字が異なるだけであり似通っている。また、字牌は画数が多く、遠くから撮影した場合には文字が潰れてしまう。他の種類の牌にも見た目が似通っているものや、潰れて見えないものがある。このような特徴を持っている麻雀牌が隣接して密集するため、麻雀牌を検出することは困難なタスクであると予想される。予備実験として画像中の麻雀牌の解像度(撮影時の麻雀牌との距離)による検出率の変化を調査した。結果の一例を図2に示す。撮影時のカメラと画像の中

心の被写体との距離は左の画像で 34cm, 右の画像で 29cm である。それぞれの画像には 14 個の麻雀牌が並んでいる。麻雀牌の輪郭を囲っているカラフルな線が、検出器が麻雀牌として検出した領域である。検出した矩形にはそれぞれ矩形の信頼度の数字がラベルとして付与されている。2 つの画像は、並んでいる麻雀牌の種類は異なるが、検出率の違いは麻雀牌ごとの解像度から生じている。対象物が低解像度になることで検出に影響しないように、撮影画像をダウンサンプリングする前に分割する事で検出率を高める。

5. データセット構築

5.1 CG を用いた合成データ生成

はじめに、本研究における CG を用いた合成データの生成手法について述べる。レンダリングエンジンに Blender, レンダラーに EeVee を用いる。EeVee は Blender の他のレンダラーである Cycle と比較して高速だが、実データへの再現度は劣る。大量のデータを生成することを想定して、EeVee を選択する。Blender はユーザが Python スクリプトを用いて操作できるため、画像データ及びラベルデータ生成を自動化できる。生成するオブジェクトは麻雀牌 ($tile_i$)136 個, 雀卓 ($table$), 光源 ($light_i$)9 個, カメラ ($camera$) である。保有するパラメータを次に定義する。

- $tile_i = \langle \text{大きさ固定, 位置可変, 向き可変} \rangle$
- $table = \langle \text{大きさ固定, 位置固定, 向き固定} \rangle$
- $light_i = \langle \text{エネルギー 500W, 位置固定, 向き固定} \rangle$
- $camera = \langle \text{位置可変, 向き可変, 焦点距離 30.83mm} \rangle$

麻雀牌は全部で 34 種類 (萬子, 筒子, 索子 9 種類ずつ, 字牌 7 種類) あり, それぞれ 4 つずつ作成する。直方体を面取りし, 表面にそれぞれの種類に応じたテクスチャを貼り付ける。雀卓は大きさ $H0.65m \times W0.65m$ の平面で, 表面にテクスチャを貼り付ける。貼り付けるテクスチャは, 所有する麻雀セットをカメラで撮影したものである。麻雀牌や雀卓のテクスチャは実物を撮影し, 画像として貼り付けた簡易なものである。光源は上方 2m の位置に 1 つ, その光源を中心として 1 辺 4m の正方形を形作るように等間隔で 8 つの光源が並ぶ。カメラの焦点距離は実データを撮影するカメラと同様の焦点距離としている。データを生成する際, オブジェクトの位置や方向が重要となる。実データに存在しない環境を設定した合成データは, 再現度が低くなる。先に定義した麻雀牌とカメラの可変パラメータに関して, 経験的に値を設定する。例えば, カメラを雀卓の下に配置することができるが, そのデータには何も情報が無いため除外する。得点計算などのために手牌付近を撮影することが想定されるため, 位置や向きを制限することができる。麻雀牌はレンダリングされる度に再配置されるため, 様々な状態の画像を生成することができる。プレイヤーが綺麗に麻雀牌を並べるとは限らないため, 位置や向きにノイズを加えるカメラのパラメータによって, レンダリング範

囲に雀卓の外が映る場合がある。本研究では, デフォルト設定のまま, 灰色の背景を描画するようにしている。背景の再現は様々なパターンが考えられ, シミュレーションすることが困難である。先行研究 [7] において, 背景に実環境とは関係のない単色のテクスチャを学習データに使った場合, 検出するバウンディングボックスが背景を含まないよう少し小さくなるのが分かっている。雀卓の表面や, その他の背景のシミュレーションは精度に大きな影響を与えないと考えられる。

設定したパラメータを元に, ラベル情報を生成する。Blender でレンダリングする際に内部で行なっている, 物体の表示・非表示の判定が取得できない。レンダリング範囲内の麻雀牌の表面が表示・非表示か判定する必要がある。麻雀牌の表面の中心座標と, 外側向きの法線ベクトルを取得する。カメラから表面の中心へのベクトルと, 法線ベクトルの交差する角度を θ とする。この時, 面が見えるのは $\frac{\pi}{2} < \theta < \frac{3}{2}\pi$ である。厳密には θ が $\frac{\pi}{2}$ 付近では表面の情報を識別することは難しいため, $\frac{5}{9}\pi < \theta < \frac{13}{9}\pi$ とする。2 つのベクトルの内積から計算し, それぞれの麻雀牌の表面の表示・非表示判定を得る。レンダリング範囲内かつ表示判定の麻雀牌の表面座標を, クラス名と共に xml 形式でデータ化する。データ構造は大規模画像データセット PASCAL_VOC と同様とする。生成した CG は再現度が低く, 実データとは性質の違いが生じる。差異を減少させるため, 次章ではデータに適用する同類化手法について述べる。

5.2 同類化変換

本章では, 合成データと実データのドメイン間の差を減少させるための同類化処理について説明する。ここで述べたように, 本研究における同類化変換とは, 性質の違う学習データと検出データのそれぞれを, 学習器にとって理想的な状態に変換することを指す。

合成データと実データの双方に同類化を適用し, 変換したデータを用いて検出器を学習させ, 検出する。合成データと実データには多数のオブジェクトが含まれる。これらの幾何学的な差は, CG の位置パラメータや方向パラメータを変化させることで再現する事ができる。つまりデータ間の差は環境光やノイズなどピクセル毎の細かな違いから生じている。同類化の手法としては, 何らかのフィルタ処理を適用が適切だと考えられる。同類化の具体的な手法は, 視覚情報処理を考慮して選択する。視覚情報は低次である輪郭や色などの情報から, 高次である意味情報を含んだ物体の情報などに分けることができる。ニューラルネットワークでも同様に, 入力層に近い部分は低次, 出力層に近い部分は高次の役割を担うように設計されている。低次の部分の処理である輪郭の認識, また低次と高次の間では低次で選択された視覚情報の特徴量を元に, 実環境をイラ

ストのように認識しているのではないかと仮定する。特徴量の選択を明示的に行なうことで、どのような処理が精度に影響するのかを調査する。

Dodge ら [8] は、ニューラルネットワークの入力に対して事前にぼかし・ノイズの付与・画質の劣化・コントラスト調整を加え、どの操作が精度に大きく影響を与えるかについて実験した。結果として、ぼかし及びノイズの付与が精度に大きな影響を与えるということが分かっている。データは 300px にリサイズしているため、麻雀牌は荒く表示される。検出にノイズとなる特徴量を削減するため、輪郭検出や 2 値化といった、ノイズを削減する処理が必要である。フィルタ処理としてウェーブレット分解及び eXtended Difference-of-Gaussians (xDoG) [9] の 2 種類の輪郭抽出手法を用いる。ウェーブレット分解は画素を周波数ごとに分解する。高周波数を取り出すことで、有用な特徴量と考えられる輪郭のみを抽出できる。輪郭検出は人間の視覚処理における低次の役割である。ニューラルネットワークにおける画像認識では、入力層に近い部分で輪郭抽出のような低次元の処理が行われているとされている。それらの処理を明示的に事前処理として適用することで、どのような処理（特徴りょう）が精度改善に起因したのかを調査する事ができる。xDoG はガウシアンフィルタの差分から輪郭を抽出し、更に画像をインクイラスト化する。ハイパーパラメータは麻雀牌やその表面の文字や図形を視認しやすいと感じる値に設定している。合成データ、実データ、それらに各フィルタを適用した画像を図 3 に示す。それぞれの画像の種類は、表 1 に示す通りである。フィルタ適用前は、データ間に環境光や質感の違いが顕著に表れている。ウェーブレット分解では、輪郭が強調されているように見えるが、質感やノイズに差が見られる。xDoG では、麻雀卓上にはノイズの差があるが、麻雀牌はどちらもインクイラスト化の影響により差異が減少しているように観察できる。変換した合成データを学習させ、検出器を構築する。次章で生成した合成データと適用した同類化手法の有効性を検証する。

6. 評価実験

同環境の合成データと実データ、それらに対し同類化変換のためのフィルタを適用したデータを学習させ、精度を評価する。意図的に同環境のデータを学習した場合、実データを学習した方が精度が良くなるはずである。この時、どの程度の差が生じるのかを確認する。また、フィルタを適用することによる精度の変化から、同類化に関するフィルタの性能を評価する。

6.1 実験設定

実データは手牌の検出を想定した、14 個の麻雀牌が並んだ画像である。学習データは合成データと実データのそれ

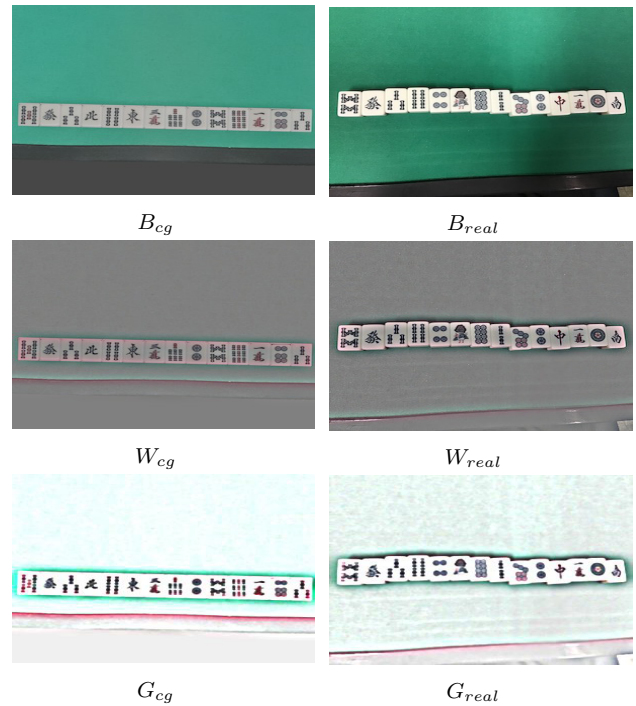


図 3 CG 画像と現実の画像の比較

表 1 データの種類

B	W	G
変換前の画像	ウェーブレット分解適用画像	xDoG 適用画像

表 2 データ生成に使用したマシン

OS	macOS Mojave 10.14.6
CPU	3 GHz Intel Core i5
GPU	Intel UHD Graphics 630 1536MB
メモリ	16GB

ぞれを用意する。麻雀牌の解像度が小さい事で検出率が低下するため、麻雀牌とカメラの距離は予備実験で示した通り、30cm 以下とする。

実データを大量に用意することは困難であることから、それぞれ 100 枚に制限する。実データ 1 枚数に対して撮影 20 秒、アノテーション 60 秒の計 80 秒を要した。合成データは 1000 枚生成した。低解像度でレンダリングした画像と高解像度でレンダリングしダウンサンプリングした画像では、後者の方が画質が良い。そのため、後者の手法で合成データを生成する。3264px×2448px のラベル付き画像 1 枚の生成に 20 秒要した。生成したデータおよびテストデータに対して同様のフィルタ処理を行った。フィルタ処理に関して、同様の解像度の画像に対してウェーブレット分解は 2 秒、xDoG は 0.02 秒要した。作業環境を表 2 に示す。手作業に比べ、データ生成は本環境下において手作業の 3.6 倍の速度である。また、データ生成中の作業への負担は無く、またラベルの信頼性は高い。大規模なデータセットを構築することを考えた場合、合成データの生成が有利である。合成データを生成する際、データ中に含む麻

雀牌のクラスに偏りを生じさせないように、一様分布を用いてランダムに選択した。

実データ 100 枚、合成データ 100 枚を 50 枚から順に 10 枚ずつ増やしそれぞれ検出器に学習させた。加えて、合成データのみ 500 枚学習させた。学習に用いるデータを、学習データ:検証データを 8:2 の割合で分けた。50epoch、重み減衰は 0.9 で学習した。重みは ImageNet で事前学習されたものを用いた。

評価指標として、ラベル付された情報と検出した矩形に対する F 値を採用する。F 値を画像毎に算出し、平均をとったものが評価値である。検出した矩形とラベル付けされた矩形の iou が 0.6 以上であれば検出できたものとする。評価値を比較することで、矩形領域がどのように変化しているのかを調べる。検出器が検出する矩形領域の信頼度に対して域値を設ける。矩形領域の域値に関して、SSD がデフォルトで採用している 0.6 を採用する。また、SSD は検出する矩形の数を指定し、その数になるまで検出した矩形を統合する。意図的な精度の上昇や再現率の低下を防ぐため、十分に大きい値である 30 を設定した。

6.2 実験結果

結果を図 4 に示す。データ数が少ない場合、合成データに比べ実データの評価指標の値が低い。学習データを調べたところ、一部クラスの物体がほとんど含まれていないことが分かった。クラス毎のデータ量の偏りは実際に起こり得るケースである。合成データはクラス毎のデータ量に偏りを生じないようにしている。データ数が 90 以上の場合、合成データよりも実データの精度が上回っている。学習データと検出対象データの性質の違いは、精度に影響することが確認できる。また、同様にデータ数が 90 以上の場合、フィルタを適用したことによる精度の明らかな改善は見られず、合成データに関しては悪化している。フィルタを適用したことで、有用な特徴量を削った、又はノイズを付加したことになる。輪郭検出、インクイラスト化は共にピクセル毎の細かな違いは無視し、目立った特徴を強調する。その際に無視された画素が有用な特徴量であり、強調された特徴がノイズとなったと考えられる。精度を改善するには、これらの特徴を保持するフィルタ処理を選択する必要がある。

7. おわりに

本稿では、機械学習による物体検出器の構築に必要なデータセット構築のため、CG を利用して合成データを生成した。実データと合成データの性質の差を減少させるため、同類化変換を試みた。手法として、ピクセル毎の細かな違いに着目し、ウェーブレット分解による輪郭検出及び xDoG を採用した。応用タスクとして麻雀牌検出を設定し、データセット構築及びデータへの事前処理を行った。評価

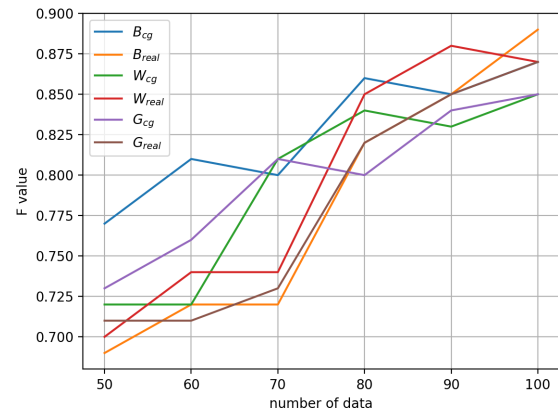


図 4 実データの検出における F 値

実験として、合成データ及び実データへの事前処理の有効性を検証した。実験結果から、今回採用した変換手法は精度の向上には繋がらず、悪化させる場合も見られた。今後の展望として、今回採用したフィルタにより削られた特徴量を分析し、それらを残すフィルタ手法の採用を検討する。また、今回は合成データの網羅率に関する優位性を示していないため、今後検証する。

謝辞 本研究の一部は JSPS 科研費 JP15K00422, JP16K00420, 19K12097, 19K12266 の助成を受けたものです。

参考文献

- [1] 鈴木涼介, 大園忠親, 新谷虎松. “抽象化された学習データを利用した麻雀牌検出器構築手法について”. 信学技報, vol.119, No.317, pp.7-12, 2019.
- [2] D. Dwibedi, I. Misra, and M. Hebert. “Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection”. ICCV 2017, pp.1301-1310, 2017.
- [3] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. “Learning from Simulated and Unsupervised Images through Adversarial Training”. CVPR, pp.2107-2116, 2017.
- [4] B. Sun, J. Feng, and K. Saenko. “Correlation Alignment for Unsupervised Domain Adaptation”. Domain Adaptation in Computer Vision Applications, pp.153-171, 2017.
- [5] R. Suzuki, T. Ozono, T. Shintani. “An Offline Mahjong Support System Based on Augmented Reality with Context-aware Image Recognition”. ESKM2019, IEEE, pp.127-132, 2019.
- [6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg. “SSD: Single Shot MultiBox Detector”. ECCV, pp.21-37, 2016.
- [7] 鈴木涼介, 大園忠親, 新谷虎松. “AR 環境における記号理解のための学習データ生成手法の検討”. 信学技報, Vol.119, No.139, AI2019-6, pp.83-88, 2019.
- [8] S. Dodge and L. Karam. “Understanding How Image Quality Affects Deep Neural Networks”. 2016 Eighth Int. Conf. QoMEX, pp.1-6, 2016.
- [9] H. Winnemoller, J. E. Kyprianidis, and S. C. Olsen. “Xdog: an extended difference-of-gaussians compendium including advanced image stylization”. Comput. Graphics, vol.36, no.6, pp.740-753, 2012.