情景画像におけるピクトグラムの検出に関する検討

岩崎　涼真　　濱村　達哉　　木村　誠聡　　辻　裕之
神奈川工科大学情報学部情報工学科

1. はじめに

近年，日本を訪れる外国人観光客が増加し，ピクトグラムのように言語に依らない標識や案内の役割がますます大きくなりつつある．さらに 2020 年の東京オリンピックでは，新たなピクトグラムが追加されるとともに，その設置個所も増加する見込みである．しかし一方で，同じデザインでも海外と日本で解釈が違っていたり，同じ意味のピクトグラムが様々なデザインで表記されていたりといった問題がある．これらを解決するために，情景画像内のピクトグラムをすべて検出し，その意味を提示するアプリケーションを構築することが本研究の目標とする．これを実現するには，検出と識別の 2 つのタスクを実現することが必要であるが，本研究では検出を主として検討する．

2. YOLO を用いたピクトグラム検出

従来，高速な物体検出を可能とする YOLOv3[1]を学習させ，情景画像からのピクトグラム検出を行っていたが，十分な検出精度を実現できていなかった．その理由として，訓練画像に一般の情景を背景としないピクトグラムが多数入っていたこと，データ拡張の方法が不十分であったこと等が考えられる．本研究では，その学習方法の改善を試みることとした．

3. 研究データの改善

従来の訓練データには，図 1 のような背景白地にピクトグラムのみが写った画像が含まれていたため，まずはこれらが学習に適切であるかを検証した．当該画像を含む訓練データとこれらを予め抜いた訓練データを用いて比較を行ったところ，当該画像を抜いた訓練データの方が約 10%高い検出率を実現できることが分かった．このため，以降の実験ではピクトグラムのみが写った画像を訓練データに含めないこととした．

図 1.　背景白地と情景画像のピクトグラム

4. 検証実験

訓練データとしてピクトグラムを含む情景画像を 153 枚使用した．これらの画像に対し，ランダムな回転変換(最大±30 度)およびランダムな射影変換を行った画像をそれぞれ追加して検出率の比較を行った．データ拡張はいずれも 1 枚の元画像につき 5 枚の変換データを追加した上で 4000 回の学習を行った．拡張を行っていない学習データと比較した結果を図 2 に示す．ここで，評価尺度には AP(Average Precision)を用いた．図 2 より，射影変換でデータ拡張を行った場合にのみ改善効果がみられることが分かった．

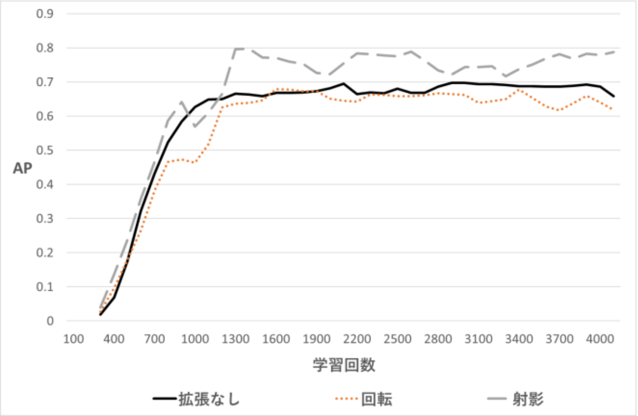

図 2.　拡張方法毎の比較

次に射影変換を用いたデータ拡張において，元画像 1 枚につき 5 枚追加する場合と 9 枚追加する場合とでピクトグラム検出率の比較を行った．学習回数はいずれも 4000 回とした．なお，先の予備実験とは異なり，今回は 28 枚の情景画像をテストデータとして新たに用意し，検出を行った．いずれの場合も検出率が最大 60%程度となり，射影変換によるデータ拡張を 5 枚から 9 枚に増やしても検出率に有意な差は見られなかった．以下に検出が成功した事例と失敗した事例を示す．

(a) 成功例　　　(b) 失敗例

図 3.　ピクトグラムの検出結果

5. まとめ

現時点でのピクトグラム検出率は実応用を想定した場合，まだ十分とはいえない．引き続き更なるアルゴリズムの改善が必要である．

参考文献

[1] Joseph Redmon and Ali Farhadi, “YOLOv3: An Incremental Improvement”, arXiv:1804.02767, 2018.

Copyright © 2020 IEICE