

撮影環境に左右されないランドマーク画像の識別

庄司 将孝[†] 秋岡 明香[‡]

明治大学先端数理科学研究科[†] 明治大学[‡]

1. はじめに

近年、SNSの普及により、写真を見るという機会が増えた。その写真は昔のパンフレットで見られるような情報が付加されたものだけでなく、写真単体の場合も多い。そのため、写っている物体や建物を特定することが難しい写真が増えた。他にも、意図せず物体が背景に映り込んでしまっている写真もよく見られる。物体が一部だけ映り込んでしまっているものや、ガラスや鏡などに反射して映り込んでしまっていたものなど様々である。本稿では、その中でも建物に注目し、ガラスや鏡に映り込んだ建物の識別を行う。被写体は撮影時刻などにより色が変わる場合がある。正しく識別をするためには、被写体の形状自体を特徴量とした識別を行う必要がある。

これまでに、ニューラルネットワークを用いたランドマーク識別に取り組んできたが、写真を構成する主要な色につられて識別精度が下がることや、背景がノイズとなり、本来目的とする建物自体の識別が難しくなることが分かっている[1]。そこで、物体検出アルゴリズムを用いて物体の切り抜きを行う。そしてその画像を入力とし、昼と夜のラベル情報を付加して別の学習モデルを作成することで、環境に左右されない建物識別モデルを作成し精度の向上を図る。

2. 関連研究

ランドマーク識別の研究は数多く存在する。その理由の一つに google が主催している「google Landmark Recognition Challenge」という大会がある[2]。ランドマーク識別において、大規模なデータセットが存在しないという大きな問題を解決するため Google が500万枚を超える画像と20万の個別ラベルが内包されている世界最

大規模のデータセットを公開した。このデータセットを用いたランドマーク識別と検索の精度を競う大会が定期的開催されている。参加者たちの多くは畳み込みニューラルネットワーク[3]やDELFF(Deep Local Features)[4]などの技術を用いて識別を行なっている。本研究でも、畳み込みニューラルネットワークを利用した機械学習を行い、精度の向上を図るが、Googleが公開しているデータセットは使わない。特定の状況における画像を学習画像として使用したい場合、既存のデータセットでは対応が難しい。そのため本研究では、自作のデータセットを用いる。

3. 提案手法

3.1. 提案手法の概要

これまでの研究では背景がノイズになってしまい建物自体の識別に影響を及ぼしていた。そのため、学習画像に対し、背景をなくすことで精度向上を図る。建物部分をアノテーションとして付加させ学習を行うYOLO[4,5]は物体を切り取る手段として適していると考えた。

本研究では、学習画像に対して物体検出アルゴリズムであるYOLOを用いてランドマークを検出させる。学習画像には検索エンジンでのスクレイピングを用いて対象の物体がしっかりと写っている画像を収集し用いる。写り込みの画像について今回は自身で撮影したものを用意し用いる。

4. 実験と評価

本稿では二つの実験を行う。一つ目は、東京タワー、ランドマークタワー、東京スカイツリー、東京都庁の5種類のランドマークについて物体検出を行う。それぞれのランドマークについて学習用にそれぞれ30枚、検証用にそれぞれ5枚ずつ用意する。これを物体検出アルゴリズムであるYOLOを用いて識別を行う。学習回数は300、バッチサイズは16とした。検証用に用意した25

Robust Landmark Recognition under Various Shooting Environment

[†]Masataka Shoji, Meiji University

[‡]Sayaka Akioka, Meiji University

枚のうち 16 枚に対して検出を行った結果を図 1 に表示。ほとんどの建物について信頼度の高い検出を行っていた。



図 1 実験 1 の検証用データに対する精度



図 2 写り込み画像を加えたモデル

二つ目は東京都庁の学習画像と検証画像に二割ほど写り込みの画像を追加して物体検出を行う実験である。検証用データについての結果は図 2 のようになった。図 1 のしっかりと写った画像を用いたモデルと比較すると都庁以外の画像についてほとんど変化は見られなかった。しかし、都庁について、写り込み画像が増加したことに伴い、しっかりと都庁が写っている画像に対して精度の低下が見られた。また、写り込み画像の検出については図 3 のような結果となった。この画像は都庁をガラス越しに移した写真

である。ガラスに部屋の中が反射しており肉眼でも都庁かどうか判断するのは難しい画像であったがこの学習モデルでは黄色く囲った部分を都庁と識別した。



図 3 ガラス越しの都庁の検出結果

5. おわりに

本研究では物体検出アルゴリズムである YOLO を用いてランドマークの切り抜きを行ない、予測モデルの精度向上を図った。学習画像が少ない場合でもかなり精度の良い識別を行うことができた。一方、ガラスや鏡などに映り込んでしまった画像を用いた実験では都庁自体の検出はできたが他の画像の精度の低下が見受けられた。今後は写り込み画像について、識別の精度を上げていくかが課題になると考える。

参考文献

- [1] 庄司将孝, 秋岡明香, “機械学習を用いたランドマーク識別”, 第 83 回全国大会講演論文集, 2021 年 2 月.
- [2] Google landmark recognition Challenge, <https://www.kaggle.com/c/landmark-recognition-challenge> (参照:2021-12-31)
- [3] Francois Chollet ; “Deep Learning with Python”, Manning Publications, 2017
- [4] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You only look once: Unified, real-time object detection, Computer Vision and Pattern Recognition (CVPR), 2015
- [5] Glenn Jocher, yolov5, GitHub repository, <https://github.com/ultralytics/yolov5>, 2020