

画像認識を用いた看板広告の検出に適した 状況および媒体の調査

本木 悠介^{1,a)} 中山 誠^{2,b)} 近藤 俊輔^{2,c)} 石川 えり^{2,d)} 神野 さくら^{2,e)} 中澤 仁^{1,f)}

概要：看板やデジタルサイネージなど、公共空間に設置される看板広告の内容を理解するためには看板広告内に含まれる物体情報や文字情報を基にする必要がある。しかし、看板広告を含む一般的な街中の風景を映した画像には、広告部以外の文字や物体の情報が多く含まれており、これらの情報は広告内からの情報を抽出するにおいて必要のない情報となる。本研究は、画像内から看板広告のみを対象として物体検出する。その際により優れた物体検出モデルを作成するために状況ごとや媒体ごとに物体検出に適した看板広告を調査することを目的とする。

Appropriate Situation and Media Survey for Billboard Advertisement Detection Using Image Recognition

YUSUKE MOTOKI^{1,a)} MAKOTO NAKAYAMA^{2,b)} SHUNSUKE KONDO^{2,c)} ERI ISHIKAWA^{2,d)}
SAKURA JINNO^{2,e)} JIN NAKAZAWA^{1,f)}

Abstract: In order to understand the contents of billboard advertisements installed in a public space such as a signboard or digital signage, it is necessary to base on object and text information included in billboard advertisements. However, images of general street scenes including billboard advertisements, contain a lot of information on characters and objects other than the advertisement section, and such information isn't necessary for extracting information from billboard advertisements. In this study, the object detection is performed only from the signboard advertisement in the image. At that time, in order to create a better object detection model, we aim to investigate billboard advertisements appropriate for object detection in each situation and medium.

1. はじめに

日本国内における広告媒体を用いた企業による自社製品やサービスのプロモーションは盛んに行われており、総広告費は増加傾向にある [1]。また [1] によると、2019 年の日本国内における広告媒体は大きく 3 つに分けることができる。1 つ目は新聞、雑誌、ラジオ、テレビメディアを含む

「マスコミ四媒体」で全体の 37.6% を占めている。2 つ目はインターネット広告費で 30.3% を占めている。3 つ目は屋外広告、交通広告、折込、ダイレクト・メールなどを含むプロモーションメディアで 32.1% を占めている。プロモーションメディア広告の中でも屋外広告と交通広告の総広告費は、全体の 7.6% を占めている [2]。屋外広告は建物の壁面や屋上に設置している物やデジタルサイネージ、看板広告自体が自立している物などが挙げられる。この分野はインターネット広告やテレビメディアでのコマーシャルとは異なり、決まった場所に一定期間掲載し続けることでプロモーションを行う物である。交通広告は広告を電車やバスなどの交通機関（駅舎や空港を含む）に掲載し、特定の範囲内で広告効果をもたらす物である。本研究では、この 2 つの広告分野を物体検出の対象とし、看板広告の意味理解

¹ 慶應義塾大学
Keio University
² 株式会社ケシオン
Kesion Co., Ltd.
a) mokky@ht.sfc.keio.ac.jp
b) nakayama@kesion.co.jp
c) kondo@kesion.co.jp
d) ishikawa@kesion.co.jp
e) s_jinno@kesion.co.jp
f) jin@sfc.keio.ac.jp

につなげるために状況ごとや媒体ごとに物体検出に適した看板広告を調査する。

看板広告の意味理解をすると、看板広告の適切な設置場所の提案や消費者への看板広告を通した適切な情報提供などに応用することができる。例えば、屋外広告は設置した場所から移動することができないという特性上、設置場所に広告している内容のターゲット層が存在しているかを強く意識していると考えられる。このことから街中にある様々な情報を基にし、看板広告の最適な設置場所を提案することは非常に重要である。例として、街上の情報を基にし、看板広告に還元しようとする研究がある。Liu 氏らの [3] は大規模な GPS 軌跡データを使用して、看板広告の配置について迅速に比較する問題の解決に取り組んでいる。このような看板広告の適切な配置場所を知ったうえで、その場に適した内容の看板広告を設置することで高い広告効果を生むことができると考える。

また、看板広告ごとに製品名やジャンル、価格帯などの情報を通して看板広告の意味を理解することができるようにになれば、それらの情報を組み合わせて消費者が望む情報を適切に提供することができると思う。看板広告を使った消費者への情報提供を行う研究は盛んにおこなわれている。高橋氏らの [4] は看板広告や大型ディスプレイなどによる不特定多数の人に向けたブロードキャスト情報に対し、モバイル端末から近距離無線でリクエストを送信し、ブロードキャスト情報に関する詳細な情報を取得するユビキタス情報提供システムを提案している。池田氏らの [5] は Kinect で取得した消費者の位置や視線により意思を忖度し、消費者の注視している点に応じて表示内容を変更したり、パラメトリックスピーカによる個別の音声広告を行う広告システムを製作し、その活用方法を提案している。Stephen 氏らの [6] は道路近くにあるディスプレイ型看板広告に近づいてくる 1 台以上の車両の乗員の人口統計情報、車両位置情報を受け取る。その情報を基に、車両の乗員に関連する広告を看板広告上に選択的に表示するものである。これらの研究は消費者の意思や視線、位置などの情報を基にし、消費者毎に適切な広告内の情報を提供することを目的としている。

看板広告の意味を理解するためには看板広告内に含まれる物体情報や文字情報を基にする必要がある。しかし、看板広告を含む一般的な街中の風景を映した画像には、広告部以外の文字や物体の情報が多く含まれており、これらの情報は広告内からの情報を抽出することにおいて必要のない情報となる。このことから本研究では、看板広告内に含まれる物体情報や文字情報を抽出するために画像内から看板広告のみを対象として物体検出する。その際により優れた物体検出モデルの作成するために状況ごとや媒体ごとに物体検出に適した看板広告を調査することを目的とする。

2. 関連研究

2.1 物体検出の動向

物体検出はコンピュータビジョンにおける代表的なタスクの 1 つであり、昨今の深層学習技術により急速な進歩を遂げている。[7] は物体検出を objectness detection(OD), salient object detection(SOD), category-specific object detection(COD) の 3 つに分類している。本研究で行う看板広告の検出は COD に含まれるタスクである。COD とは特定の各画像から複数の定義済みオブジェクトクラスを検出することを目的とするタスクであり、提案ベースと回帰ベースの 2 つの主要カテゴリに分けられる。また、COD の主な課題はクラス内の外観の変化とクラス間の外観の類似性をどのように処理するかにある。他にも、[8] は物体検出における検出フレームワークを 2 つの主要カテゴリに分けている。1 つ目は、オブジェクト候補の提案を生成するための前処理ステップを含む 2 段階の検出フレームワークである。2 段階の検出フレームワークで広く使用されているフレームワークは Faster RCNN[9], RFCN[10], Mask RCNN[11] が挙げられる。2 つ目は、オブジェクト候補の提案のプロセスを分離しない単一の提案された方法を持つ、1 段階の検出フレームワーク（領域提案のないフレームワーク）である。1 段階の検出フレームワークは軽量かつ高速なモデルを作成できることが特徴である。例として SSD[12], YOLO[13] が挙げられる。計算コストが大きい場合、2 段階の検出フレームワークの方が 1 段階の検出フレームワークよりも高い検出精度を出している。本研究では、回帰ベースかつ 1 段階の検出フレームワークで、YOLO の派生モデルである YOLOv3[14] を用いる。

2.2 看板広告を対象とした物体検出

深層学習を用いた看板広告を対象とした物体検出を行う研究がある。Rahmat 氏らの [15] は AlexNet の Deep Convolutional Neural Network(DCNN)[16] を画像分類用の 1000 カテゴリの事前トレーニング済みニューラルネットワークを用いて、リアルタイムで広告看板を検出し、ジオタグを付ける手法を提案している。[15] は看板広告を検出する際にスコアボード、モニター、テレビ、映画、および Web サイトの 6 つのクラスに分類しているが、本研究ではクラスを看板広告の媒体ごとに細分化することで、媒体ごとに検出のしやすさを調査する。Liu 氏らの [17] は物体検出アルゴリズムである R-CNN[18] は一般的な物体検出には高いパフォーマンスを達成しているが、看板広告の検出においては十分に機能しないとしている。そこで [17] では R-CNN を基にした attention-based multi-scale feature fusion region proposal network(AM-RPN) を提案している。[17] は看板広告の媒体ごとの精度向上ではなく 60 個の一般的なブランド (例:KFC, Nike, Starbuckcs など) の精



図 1 各広告媒体の例
上段左から plane, sleeve, wall, roof, banner
下段左から standing, signage, vehicle, train, pillar

度向上を目指している一方で、本研究は看板広告の内容に依らない広い範囲での看板広告の検出精度の向上のために調査する。

看板広告を対象とした物体検出を行う研究には、看板広告が法規制に則ったものかを判定するために行っている物がある。Liu 氏らの [19] は看板広告上の文字情報から、違法な看板広告を検出するものである。[19] は看板広告上の文字を Optical character recognition(OCR) により検出し、Devlin 氏らの提案した言語モデルである Bidirectional Encoder Representations from Transformers(BERT)[20] によってトレーニングされた複数の分類ラベルで構成される広告情報抽出モジュールを用いることで違法な看板広告を検出している。また、Rahmat 氏らの [21] は看板広告が写った画像に対してグレースケーリングやエッジ検出などの処理を行い看板広告の 4 隅を検出し、寸法を測定する。次いで、GPS 座標を用いてカメラと看板広告の 2 点間距離を算出する、高速道路看板に対する税計算用の測定システムを提案している。[19], [21] は看板広告の検出を行っているが、対象としている媒体が平面上の看板であり、看板広告以外の物体が写っていない画像を用いている。本研究は様々な広告媒体や看板広告以外の物体が写った画像を対象とし、その状況下での物体検出を試みる。

3. 提案手法

本章では、最初に看板広告の定義や広告媒体に付与する分類ラベルの説明を行う。次いで、物体検出モデルを作成する際に状況ごとや媒体ごとに物体検出に適した看板広告を調査するための条件について説明する。

3.1 看板広告の定義

本節では、本研究内で対象とする看板広告の定義について説明する。1 章で記したように、本研究では広告媒体の中でもプロモーションメディアに含まれる屋外広告と交通広告を対象とする。また、対象とする看板広告の内容は 1.

表 1 各分類ラベルの概要

ラベル	媒体	補足
plane	室内ポスター, 額縁等	室内の平面媒体
sleeve	袖看板	-
wall	壁面広告	壁面看板, 懸垂幕, 壁面ラッピング
roof	屋上広告	屋上看板, 屋上広告塔
banner	バナー, フラッグ	-
standing	自立式看板	ロードサイン, 建植看板
signage	デジタル広告	街頭ビジョン, デジタルサイネージ
vehicle	車載広告	ラッピングバス, アドトラック
subway	電車壁面広告	電車外装にされたラッピング広告
pillar	柱巻き広告	柱巻き広告

製品、サービスのプロモーション目的のもの（ラッピングを含む）、2. 店舗への道案内目的のもの、3. 店先看板、店名看板である。道路標識や電車内のモニターに映し出される乗り換え案内、建物名などの媒体は含まない。

3.2 分類ラベル

本節では、画像内の看板広告に付与する分類ラベルについて説明する。今回は 10 個の分類ラベルを用意した。各ラベルの概要を表 1 に示す。また、図 1 に各ラベルの広告媒体の例を示す。なお、画像中の赤く塗りつぶされている部分は広告部（デザイン部）である。赤く塗りつぶしているのは広告の意匠権を保護するためである。分類ラベルは看板広告に対してシングルラベルで付与し、マルチラベルが付与されることはない。

plane は駅構内や建物内の平面に設置されている広告媒体である。なお、天井や床、階段の側面に設置されている広告媒体は含まない。sleeve は建物の壁から支点が伸び、広告部が設置されている広告媒体である。sleeve には図 1 に示した 1 つの媒体に 1 つの広告が付随するものだけでなく、1 つの媒体にいくつかの広告が付随している物がある。wall は室外の平面に設置されている広告媒体である。

plane とは異なり、比較的大きい看板広告が多いが、視覚的に似通った点がある。roof は建物の屋上に設置されている広告媒体である。banner は何らかの物体を支点として、そこから吊り下げられるように設置されている広告媒体である。standing は広告部から本体を支えるための脚部を持つ、自立している広告媒体である。なお、roof, banner, standing について本研究では広告部を支える脚部 (banner に関しては吊り下げ部) までを広告媒体の範囲としている。signage はディスプレイに表示されているデジタルサイネージ型の広告媒体である。vehicle は車の側面や後部に広告がラッピングされている広告媒体である。subway は電車の車体に広告がラッピングされている広告媒体である。なお、電車の車内にある広告は subway の対象ではない。pillar は円柱に巻きつけるように張られている広告媒体である。

3.3 学習時のラベル選択

本節では、物体検出モデルを作成する際に物体検出に適した広告媒体を調査するために学習時のラベルに条件を付け、3つのモデルを作成する。作成する各モデルの内容は以下のようになっている。なお、各モデルは後述するデータ増強を行っている。

- 全ラベルを1つのラベルに統合して学習させたモデル (ONE)

本モデルは画像中から広告媒体による区別が精度に影響を与えるか確かめるために、今回用意した10個のラベルを1つのラベル billboard に統合し、学習させたものである。

- 一般的な広告媒体を学習させたモデル (MAIN)

本モデルは収集したデータのなかで、一般的な広告媒体である plane, sleeve, wall, banner, roof, signage の6種類を対象に学習させたものである。

- 全ラベルを学習させたモデル (ALL)

本モデルは今回用意した10個のラベルすべてを学習させたものである。

3.4 学習データに対する画像処理

本節では、3.3節で作成したモデルの中で ALL モデルを対象とし、高精度のモデルを作成できる状況を調査するために学習データに複数の画像処理を加えたモデルを作成する。学習データに画像処理を行ったモデルの内容は以下のようになっている。

- 画像処理を一切行わないモデル (NON)

本モデルは学習データに画像処理を一切行わない状態で、どれくらいの精度が出るかを確かめるベースラインとして使用する。

- データ増強を行ったモデル (AUGMENTATION)

本モデルは学習データにデータ増強を行ったモデル

である。施した増強は画像の回転 (± 15 degree), 移動 (± 0.01 fraction), 拡大縮小 (± 0.2 gain), セン断 (± 10 degree) である。これらの増強を定めた範囲内で画像ごとにランダムに施すことでデータ増強を行っている。なお、本モデルは3.3節で記した ALL モデルと同義である。

- グレースケール処理しデータ増強を行ったモデル (GRAYSCALE)

本モデルは学習データにグレースケール処理を施したものである。なお、AUGMENTATION モデルと同様のデータ増強を行っている。また、テストデータにも同様にグレースケール処理を加える。

- シャープネスを強調し、データ増強を行ったモデル (SHARPNESS)

本モデルは学習データにシャープネスの強調処理を施したものである。なお、AUGMENTATION モデルと同様のデータ増強を行っている。また、テストデータにも同様にシャープネスの強調処理を加える。

3.5 検出しやすい看板広告の状況

本節では、3.3節で作成したモデルの中で ONE モデルを対象とし、検出しやすい看板広告の状況を調査するための画像パターンの選定を行う。本研究では、以下の項目ごとに看板広告の状況をパターン分けする。

- 1~複数個の広告が写っている

本状況は撮影した際に正面に広告が1つのみ写っているシンプルな画像から複数個の広告が写っている画像を対象とし、広告個数による精度の変化を調査する。図2, 3に対象となる状況の例を示す。

- 広告媒体から離れて撮影している

本状況は広告が撮影地点から離れて写っている画像を対象とすることで、看板広告の物体検出に距離がどれほど影響を与えるかを調査する。図4に対象となる状況の例を示す。

- 広告を撮影した際に角度がついている

本状況は広告を撮影した際に角度がついている画像を対象とすることで、看板広告を正面に捉えていない画像にもモデルが適応できているかを調査する。図5に対象となる状況の例を示す。

- 暗所で撮影している

本状況は夜中や照明のない場所など暗所で撮影している画像を対象とすることで、昼間や夜間などの天候状況が看板広告の物体検出にどれほど影響を与えるかを調査する。図6に対象となる状況の例を示す。

4. 実験

本章では、3章で挙げた条件下での実験を行い、各実験の考察を行う。最初に学習時のラベル選択により作成し

表 3 学習時のラベル選択（実験）の結果

	mAP	plane	sleeve	wall	roof	banner	signage	standing	vehicle	subway	pillar
ONE	49.1	-	-	-	-	-	-	-	-	-	-
MAIN	38.7	0.39	0.22	0.17	0.57	0.44	0.53	-	-	-	-
ALL	39.8	0.37	0.18	0.22	0.44	0.55	0.43	0.46	0.54	0.15	0.64



図 2 正面に広告が写っている例

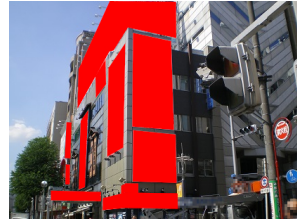


図 3 複数の広告が写っている例



図 4 広告媒体から離れて撮影している例

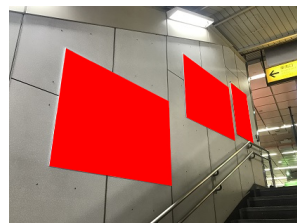


図 5 広告媒体を撮影した際に角度がついている例

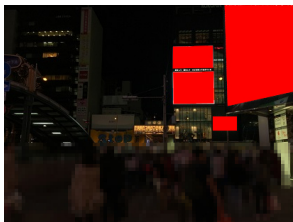


図 6 暗所で撮影している例

たモデルのテスト結果から考察する。次いで、学習データに対する画像処理を加えて作成したモデルのテスト結果から考察する。最後に、検出しやすい看板広告の状況についての考察を行う。なお、各ラベルの評価には Average Precision(AP) を用い、モデルの総合評価には mean Average Precision(mAP) を用いる。またミスレート評価には Log-average miss rate(lamr) を用いる。各評価指標の計算式を数式 1, 2, 3 に示す。

$$AP = \frac{1}{N} \sum_{j=1}^I Precision \quad (1)$$

N:j の時点でモデルが正しく認識した
正解ラベル (物体) の総数

$$mAP = \frac{1}{M} \sum_{i=1}^M AP \quad (2)$$

M:画像の総数

$$lamr = \exp \left[\frac{1}{n} \sum_{i=1}^n \ln a_i \right] \quad (3)$$

n:検出物体の総数

4.1 データ概要

本節では、実験で用いるデータの概要について説明する。本研究で用いた画像の総数は 9,070 枚である。また、各ラベルの状況は表 2 のようになっている。今回作成する全てのモデル内でこれらのデータを学習データとテストデータに 8 対 2 の割合で分割している。ALL モデルとその派生モデル、ONE モデルについては学習、テストそれぞれに使用する画像は一緒である。MAIN モデルについては使用しないラベルがある関係上、割合は同じだが、ほかのモデルとは異なる画像分配をしている。

表 2 各アノテーションラベルの取得状況

plane	sleeve	wall	roof	banner	standing	signage
4,971	5,264	11,283	3,246	3,921	1,323	3,818

vehicle	subway	pillar
954	165	2,440

4.2 YOLOv3

実験において作成する各モデルには、2.1 節で記したようにリアルタイムオブジェクト検出システムである YOLOv3 を用いる。YOLOv3 は物体検出を回帰問題としてモデル化している。画像を S×S グリッドに分割し、グリッドセルごとにロジスティック回帰を用いて、Bounding Box の confidence を予測する。予測した Bounding Box が他の Bounding Box よりも ground truth object と重なる場合、予測したものを 1 とする。予測した Bounding Box が最適ではないが、ground truth object が閾値を超えてオーバーラップしている場合は、[9] に従って予測を無視する。続いて、クラス C の確率を予測する。各ボックスはマルチラベル分類を使用し、Bounding Box に含まれるクラスを予測する。なお、softmax は使用せず、代わりに independent logistic classifier を使用する。また、トレーニング中のクラス予測のために binary cross-entropy loss を使用している。(一部、[13] より引用し和訳)。YOLO モデルの概略図を図 7 に示す。本研究では YOLOv3 を機械学習ライブラリである PyTorch を用いて Ultralytics 社が実装した PyTorch YOLOv3[22] を利用する。

表 4 学習データに対する画像処理（実験）の結果

	mAP	plane	sleeve	wall	roof	banner	signage	standing	vehicle	subway	pillar
NON	31.4	0.30	0.14	0.18	0.41	0.41	0.37	0.29	0.46	0.02	0.57
AUGMENTATION(ALL)	39.8	0.37	0.18	0.22	0.44	0.55	0.43	0.46	0.54	0.15	0.64
GRAYSCALE	38.1	0.34	0.14	0.21	0.41	0.51	0.36	0.47	0.52	0.22	0.62
SHARPNESS	38.5	0.35	0.16	0.19	0.45	0.54	0.38	0.44	0.51	0.23	0.59

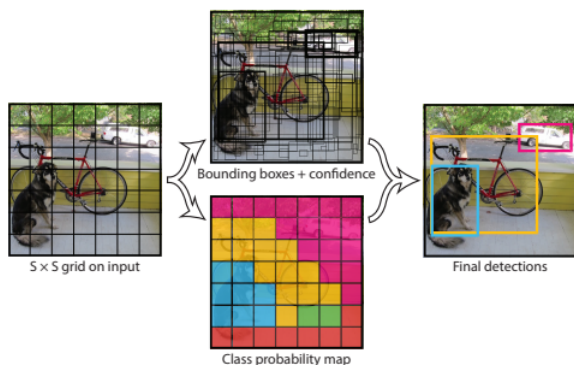


図 7 YOLO モデルの概略図 ([13] より引用)

4.3 学習時のラベル選択（実験）

3.3 節で挙げた ONE, MAIN, ALL モデルを作成し、テストデータを用いて評価した結果を表 3 に示す。表 3 から subway の mAP が低いのは表 2 でも示したように学習データが非常に少ないことが理由だと考える。しかし、同様にデータ数が少ない standing や vehicle, pillar の AP は高いことが分かる。これはこれら 3 種の広告媒体が非常に特徴的な広告媒体であるからだと考える。standing は広告部を支えるための脚部があるという形状的特徴がある。vehicle は車にのみ付随しているという位置的特徴がある。pillar は柱にのみ巻き付いているという位置的特徴がある。これはデータ数の多い roof, banner, signage にも共通すると考える。roof は建物の屋上に設置されているという位置的特徴を持っており、banner にも支点から吊り下がっているという広告媒体独自の特徴がある。signage は他の広告媒体とは異なり、媒体自体が発光しているという視覚的特徴があること AP が高くなっていると考え。

しかし、同様にデータ数が多い sleeve や wall, plane といったシンプルな平面媒体の AP が低いことが分かる。これらの広告媒体を比較すると、視覚的特徴が似通っており、位置的特徴も見られない。従って plane, sleeve, wall の AP が低いのは広告媒体としての特徴が似ており、広告媒体独自の特徴を持っていないことから、ラベル予測時に相互に誤認していることが原因であると考え。また、ラベル予測を行う必要のない ONE モデルの mAP が高いこともこの考えの理由の 1 つである。

以上のことから、看板広告の検出はデータの数以上に媒体独自の特徴が重要になってくると考える。また plane, sleeve, wall といったラベルの定義は再考の余地がある。

4.4 学習データに対する画像処理（実験）

3.4 節で挙げた NON, AUGMENTATION, GRAYSCALE, SHARPNESS モデルを作成し、テストデータを用いて評価した結果を表 4 に示す。表 4 の NON, AUGMENTATION モデルの結果から、データ増強が有効に作用していることが分かる。全てのラベルにおいて AP が向上しており、特に学習データの少ない standing, vehicle, subway, pillar では非常に有効である。一方で学習データが十分にあった plane, sleeve, wall では AP が向上はしているが、伸び幅は前述の媒体よりは小さい。

AUGMENTATION, GRAYSCALE, SHARPNESS モデルの結果を比較すると、基本的なデータ増強のみを行った AUGMENTATION モデルが最も良い mAP を記録していることが分かる。基本的に各媒体で AUGMENTATION モデルが高い mAP を記録しているが、subway に関しては他の 2 モデルの方が良い結果を残している。また、roof や standing に関してもわずかなではあるが、AUGMENTATION モデルより良い結果を残している。このことから、一部の媒体に関してはデータ増強だけでなく特定の画像処理を加えることが有効に作用することが分かる。しかし、広告媒体全般を包括するモデルを作成する際はデータ増強のみを行ったモデルが望ましいと考える。

4.5 検出しやすい看板広告の状況（実験）

3.5 節でパターン分けした各条件下で ONE モデルを使い、テストした結果を以下に記す。

- 1～複数個の広告が写っている

画像内の広告の個数の変化が精度に影響するかを調査するために、テストデータを画像内の広告個数で分割する。分割したデータをそれぞれ ONE モデルにテストさせ、結果から考察する。テストした結果を表 5 に示す。表から画像内の広告個数が 1 個の時の mAP が非常に高いことが分かる。また、全体の結果から基本的に画像内の広告個数が増えるほど、mAP が下がることが分かる。特に広告個数が 9 個を超えるものに関しては、引きで撮影したものや雑多な街中を撮影したものが多くことから mAP が低くなっていると考え。

- 広告媒体から離れて撮影している

撮影地点と広告媒体との距離が精度に影響が出るかを調査するために、広告媒体が撮影地点から離れている画像を対象としてテストさせ、結果から考察する。テ

表 5 広告個数による精度の変化

広告個数	画像枚数	mAP	lamr
1	508	92.9	0.11
2	231	62.5	0.53
3	252	64.4	0.53
4	172	55.1	0.66
5	126	56.0	0.67
6	107	48.3	0.78
7	84	45.9	0.77
8	78	48.2	0.77
9	56	38.3	0.84
≥ 10	206	37.2	0.90

ストした結果を表 6 に示す。表 6 から表 3 に挙げた ONE モデルよりも mAP が低くなっていることが分かる。また、表 5 にある画像内の広告個数が少ないものは撮影地点と広告媒体との距離が近いものも多く、mAP が高く出ている。このことから撮影地点と広告媒体との距離は看板広告の物体検出に大きく影響があると考えられる。

表 6 広告媒体から離れて撮影している画像の結果

画像枚数	mAP	lamr
213	45.7	0.82

- 広告を撮影した際に角度がついている

看板広告を正面に捉えていない状況が精度に影響を与えるかを調査するために、広告を撮影した際に角度がついている画像を対象としてテストさせ、結果から考察する。テストした結果を表 7 に示す。表 7 から表 3 に挙げた ONE モデルよりも mAP が高くなっていることが分かる。このことから看板広告の物体検出において、看板広告を正面にとらえていない状況でもある程度の mAP を記録することができると考える。

表 7 広告を撮影した際の角度がついている画像の結果

画像枚数	mAP	lamr
295	57.7	0.58

- 暗所で撮影している

夜中や照明のない場所など暗所で撮影している状況が精度に影響を与えるかを調査するために、暗所で撮影されている画像を対象としてテストさせ、結果から考察する。テストした結果を表 8 に示す。表 8 から表 3 に挙げた ONE モデルよりも mAP が低くなっていることが分かる。これは看板広告に陽の光が当たらない暗所という特殊な状況が原因であると考えられる。一部の wall や roof 全般は広告媒体の上部または下部に広告のための照明がついている。また、signage は内部から発光している。このような広告媒体に関しては正

しく検出することができている。しかし、照明などで照らされていない媒体に関しては、極端に検出精度が低下している。このことから、本状況において看板広告を正しく検出することができるかは、広告に光が当たっているまたは、広告自体が発光しているかが非常に重要であると考えられる。

表 8 暗所で撮影している画像の結果

画像枚数	mAP	lamr
59	44.7	0.75

5. 結論

本研究では、看板広告内に含まれる物体情報や文字情報を抽出するために画像内から看板広告のみを対象として物体検出する。その際により優れた物体検出モデルの作成するために状況ごとや媒体ごとに物体検出に適した看板広告を調査することを目的とした。最初に分類ラベルを選択した 3 つのモデルを用意し、物体検出に適した広告媒体を調査した。この調査の結果、ラベル分類を行わないシンプルな看板広告の検出では mAP が 49.1 を記録し、用意したモデル内で最も高い精度であった。また、ラベル分類を行う看板広告の検出にはデータ数以上に広告媒体独自の特徴を持っているかが重要であり、本研究で設定した分類ラベルを再考する必要があることが分かった。次に高精度のモデルを作成できる状況を調査するために、学習データに対して画像処理を行ったモデルを 3 つ用意し、画像処理を行わなかったモデルと精度比較をした。結果として基本的なデータ増強を行ったモデルが最も高い mAP を記録した。一方で、一部の媒体に関してはデータ増強だけでなく特定の画像処理を加えることが有効に作用することが分かった。最後に検出しやすい看板広告の状況を調査するために 4 つの状況下で画像を分類し、モデルに評価させた。結果として広告媒体が多数存在し、撮影地点と広告媒体が離れた状況での物体検出では mAP が低くなることが分かった。また、看板広告を正面にとらえていない状況でもある程度の精度を記録することができた。暗所での看板広告の検出では対象となる広告媒体が照明に照らされているまたは、自発光していることが重要であることが分かった。

今後の展望は看板広告に付与する分類ラベル設定の再考を行う。また、本研究を基に看板広告を対象としたより優れた物体検出モデルを作成することで、看板広告の内容を理解することに必要な物体情報や文字情報を検出しやすい状況を作る。そこで得た情報を使い、消費者への看板広告を通した適切な情報提供を行うことができるシステムの実装を行う。

謝辞

本研究は株式会社ケシオンとの共同研究により実施されたものである。

参考文献

- [1] dentsu : 2019 年 日本の総広告費 (online), 入手先 <https://www.dentsu.co.jp/knowledge/ad.cost/2019/> (参照 2020-04-12).
- [2] dentsu : 2019 年 日本の広告費 | プロモーションメディア (online), 入手先 <https://www.dentsu.co.jp/knowledge/ad.cost/2019/media4.html> (参照 2020-04-12).
- [3] Dongyu Liu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu : SmartAdP: Visual Analytics of Large-Scale Taxi Trajectories for Selecting Billboard Locations, IEEE Transactions on Visualization and Computer Graphics, Vol. 23, No. 1, pp. 1-10, 2017.
- [4] 高橋 三恵, 中尾 敏康 : コピキタス情報提供システムの検討と試作, 情報処理学会研究報告, No. 94, pp. 47-54(2002).
- [5] 池田 輝政, 遠藤 正隆, 中嶋 裕一, 三浦 哲郎, 菱田 隆彰 : 顧客の意思を付度するデジタルサイネージ広告システム, エンタテインメントコンピューティングシンポジウム 2017 論文集, p439-440, 2017.
- [6] James Stephen MILLER, Patrick Graf and Frank RIGGI : BILLBOARD DISPLAY AND METHOD FOR SELECTIVELY DISPLAYING ADVERTISEMENTS BY SENSING DEMOGRAPHIC INFORMATION OF OCCUPANTS OF VEHICLES, 2018.
- [7] Junwei Han, Dingwen Zhang, Gong Cheng, Nian Liu and Dong Xu : Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey, IEEE Signal Processing Magazine, Vol. 35, No. 1, pp. 84-100, 2018.
- [8] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu and Matti Pietikäinen : Deep Learning for Generic Object Detection: A Survey, International Journal of Computer Vision, Vol. 128, No. 2, pp. 261-318, 2020.
- [9] Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun : Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, 2017.
- [10] Jifeng Dai, Yi Li, Kaiming He and Jian Sun : RFCN: object detection via region based fully convolutional networks, Neural Information Processing Systems (NIPS), pp. 379-387, 2016.
- [11] Kaiming He, Georgia Gkioxari, Piotr Dollár and Ross Girshick : Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV), 2017
- [12] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu and Alexander C. Berg : SSD: Single Shot MultiBox Detector, European Conference on Computer Vision (ECCV), pp. 21-37, 2016.
- [13] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi : You Only Look Once: Unified, Real-Time Object Detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.
- [14] Joseph Redmon and Ali Farhadi : YOLOv3: An Incremental Improvement, arXiv, 2018.
- [15] Romi Fadillah Rahmat, Dennis, Opim Salim Sitompul, Sarah Purnamawati and Rahmat Budiarto : Advertisement billboard detection and geotagging system with inductive transfer learning in deep convolutional neural network, TELKOMNIKA Telecommunication Computing Electronics and Control, Vol. 17, No. 5, pp. 2659-2666, 2019.
- [16] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton : ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing System (NIPS), Vol. 1, pp. 1097-1105, 2012.
- [17] Gang Liu, Chuyi Wang and Yanzhong Hu : RPN with the Attention-based Multi-Scale Method and the Adaptive Non-Maximum Suppression for Billboard Detection, 2018 IEEE 4th International Conference on Computer and Communications (ICCC), pp. 1541-1545, 2018.
- [18] Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik : Rich feature hierarchies for accurate object detection and semantic segmentation, 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.
- [19] Huxiao Liu, Lianhai Wang, Weinan Zhang and Wei Wang : An Illegal Billboard Advertisement Detection Framework Based on Machine Learning, ICBTD2019: Proceedings of the 2nd International Conference on Big Data Technologies, pp. 159-164, 2019.
- [20] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv, 2018.
- [21] Romi Fadillah Rahmat, Sarah Purnamawati, Handra Saito, Muhammad Choirul Ichwan and Tri Murti Lubis : Android-based automatic detection and measurement system of highway billboard for tax calculation in indonesia, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 14, No. 2, pp. 877-886, 2019.
- [22] Ultralytics LLC : PyTorch YOLOv3 (online), 入手先 <https://github.com/ultralytics/yolov3> (参照 2020-04-10).