

位置情報に基づいた画像分類モデルの選定による 看板画像の分類精度向上

Improving Classification Accuracy of Signboard Images by Selecting Location-Based Image Classification Models

中道 翔大* 松下 光範
Shota Nakamichi Mitsunori Matsushita

関西大学総合情報学部
Faculty of Informatics, Kansai University

Abstract: 本研究の目的は、看板から店舗の詳細情報を取得するシステムの実現である。これまでに、携帯端末のカメラで撮影した看板画像から店舗を判別し、ARを用いてその店舗の詳細情報を表示する手法が提案されている。しかし、このシステムは識別可能な店舗数が限られており、対象店舗数を拡大する場合、分類する店舗数の増加に伴って識別精度が低下する懸念がある。そこで本稿では、対象店舗の所在地をもとに予め複数の画像分類モデルを構築し、ユーザの位置情報に基づいて使用するモデルを選定して識別を行うことで、分類精度を向上させる手法を提案する。

1 はじめに

看板とは、「屋外広告物法」により定められた条件を満たし、商店などが店名・業種・商品名などの情報を道路利用者の目に付きやすいように屋外に掲げたものであり、各店舗は街中を歩く人々に対して来店を促すために看板を掲示する [10]。人々は、目に入った看板に描かれている内容から店舗の種類を判別し、どの店に入るかななどの意思決定を行っている。加えて、スマートフォンなどの携帯端末を用いた、看板に記述されている店舗名をもとに行うテキスト入力による Web 検索や、位置情報をもとに行う付近の店舗検索によって、その店舗の商品や口コミなどの詳細情報を取得できる。

看板の多くはその店舗が位置する地域の言語で記述されている。そのため、看板の文字がユーザの母国語でない場合、言語障壁によって看板から店舗の情報を得ることが困難となる。草書体のような読みづらい書体で記述されている看板の場合でも、同様のことが言える。テキスト入力による Web 検索を行う場合でも、ユーザが文字を自力で入力できることが前提となっているため、看板の文字が理解できない場合、店舗の詳細情報の取得が困難となる。位置情報を用いた検索によって周辺にある

自分の条件に合った店舗が見つかったとしても、看板の文字が読めない場合、多くの店舗が密集している中から目的の店舗を発見するには手間がかかる。そのため、人々が目的の店舗を探す上で、その店舗の看板の文字が理解可能であるかは重要な要素であると言える。

本研究では、ユーザが周囲の文字が理解できない状況にいる場合でも、看板から目の前の店舗の情報を容易に取得できるシステムの実現を目的としている。北村らは、携帯端末のカメラで撮影した看板画像から、機械学習を用いた物体検出及び画像分類によって店舗を判別し、ARを用いて店舗の詳細情報を重畳表示するシステムを開発している [8]。しかし、北村らのシステムでは特定の地域に位置する 15 店舗のみを対象としているため、適用範囲が限られていた。このシステムの対象を拡大する場合、分類する店舗数の増加に伴って精度が低下する懸念がある。この問題を解決するため本稿では、対象店舗の所在地をもとに予め複数の画像分類モデルを構築し、ユーザの位置情報に基づいて使用するモデルを選定することで、分類精度を向上させる手法を提案する。

2 関連研究

本章では、看板に記述されている文字列の認識手法に関する研究と、画像分類の精度を向上させるための手法の 1 つである画像データセットの拡張に関する研究について述べる。

* 連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

2.1 看板の文字認識に関する研究

情景画像中の看板に記述されている文字列を認識する研究は数多く行われている。Panhwar らは、機械学習を用いて自然景観画像の中にある看板画像に記述された文字情報に対し、認識が難しいウルドゥー語と英語を区別して認識する手法を提案した [6]。CNN を導入した MSER を用いて自然景観画像から看板画像を検出した後に、記述されている文字情報を同様に検出及び抽出し、HOG 特徴量及び SVM 分類器を用いて文字認識を行った。Kavati らは、OCR を用いて携帯端末のカメラに写っている看板の文字を認識し、英語からテグル語に翻訳してユーザに提示する旅行者向けの Web アプリケーションを開発した [4]。OCR は、手書き文字や印刷された文字などが写っている画像を解析し、その中に含まれている文字に相当するパターンを検出し、書かれている内容をテキストデータとして取り出す技術である。

2.2 画像データセットの拡張に関する研究

河野らは、GAN を用いた画像データセットの拡張手法を提案した [9]。GAN は生成モデルの 1 つであり、2 つのニューラルネットワークを交互に競合させて学習を進める事で、実在しないデータの生成や、存在するデータの特徴に沿った変換ができる手法である。これを用いた学習画像の生成によるデータセットの拡張を行った結果、分類モデルの正解率が上昇した。fujita らは、進化的画像処理を用いて指定された特徴を持つ画像を生成する手法を提案し、画像分類においてデータ拡張に適用することの有効性を検証した [2]。進化的画像処理は、最適化手法の一種である進化計算法を用いて、与えられた入力画像とその理想的な処理結果の画像である目標画像の組からなる学習画像セットによって、指定された画像変換を自動構築するものである。提案手法で水増画像を生成し、SVM 分類器を用いた上で、ランダムデータを用いた 2 種の分類問題によって精度を検討した結果、多くのケースの分類問題で正解率を向上させることができた。

2.3 本研究の位置づけ

2.1 節で述べた看板の文字認識を用いて、看板の文字をそのまま翻訳したとしても、ユーザはその店舗がどのような店舗であるかを理解できるとは限らない。そのため、本研究の目的を達成する場合において、ただ文字を翻訳してユーザの母国語に直すだけでは不十分である。

さらに、看板の文字が草書体のような認識が難しい書体である場合、看板から文字を正しく認識することが難しい。そのため、本研究では看板の文字認識を行わずに、ユーザへ店舗の情報を提示する。

2.2 節で述べた先行研究では、学習用画像の生成によるデータセットの拡張によって分類精度を向上させている。しかし、GAN を適用した水増し処理を行うには多くの学習用画像が必要となり、そのために全国数多にある各店舗の看板画像を用意することは容易ではない。進化的画像処理を用いた水増し処理においても、各店舗の看板画像ごとに特徴空間中の生成したい目標特徴点を指定する画像処理を行う必要があるため、多クラス分類において数多くある各店舗の看板画像ごとに行うには多大なコストがかかる。そのため本稿では、データセットの拡張を行わずに分類精度を向上させる手法を提案する。

3 看板画像を用いた店舗情報検索システム

本章では、本研究において基礎となる、看板画像から店舗の情報を取得するシステムについて述べる。北村らは、携帯端末のカメラを用いて看板から店舗の情報を取得するシステムを開発している。ユーザが目の前にある店舗の情報を取得する際、まず始めに、携帯端末のカメラを通して店舗の看板を写す。次に、カメラに写っている画像を、機械学習を用いた看板領域の検出及び看板画像の分類モデルによって識別を行う API サーバに送信する。看板領域の検出には Tensorflow1.0[1] で実装した YOLOv2[7]、看板画像の分類には、VGG16[5] を用いて構築した分類モデルを使用しており、これらを用いてユーザから送られてきた画像に写っている看板がどの店舗のものであるかを識別する。その後、OpenStreetMap [3] に登録されている識別した店舗に紐付いた店舗情報を、AR を用いてユーザの携帯端末に重畳表示する。これらの工程を 500 ミリ秒ごとに行うことで、リアルタイムな店舗情報の提示を可能とする。看板認識の処理を行うマシンのスペックを表 1 に示す。API サーバの構築には、Python3.6.5 と WebAPI フレームワークである Falcon1.4.1*1 を用いている。このシステムでは店舗の看板を分類する際、文字列ではなく看板画像を用いるため、草書体のような認識が難しい書体で記述されている看板であっても分類が可能である [8]。

北村らのシステムでは、大阪府高槻市にある商店街の一部の店舗を対象としている。そのため、システムを適

*1 <https://falconframework.org/> (2022/2/1) 存在確認

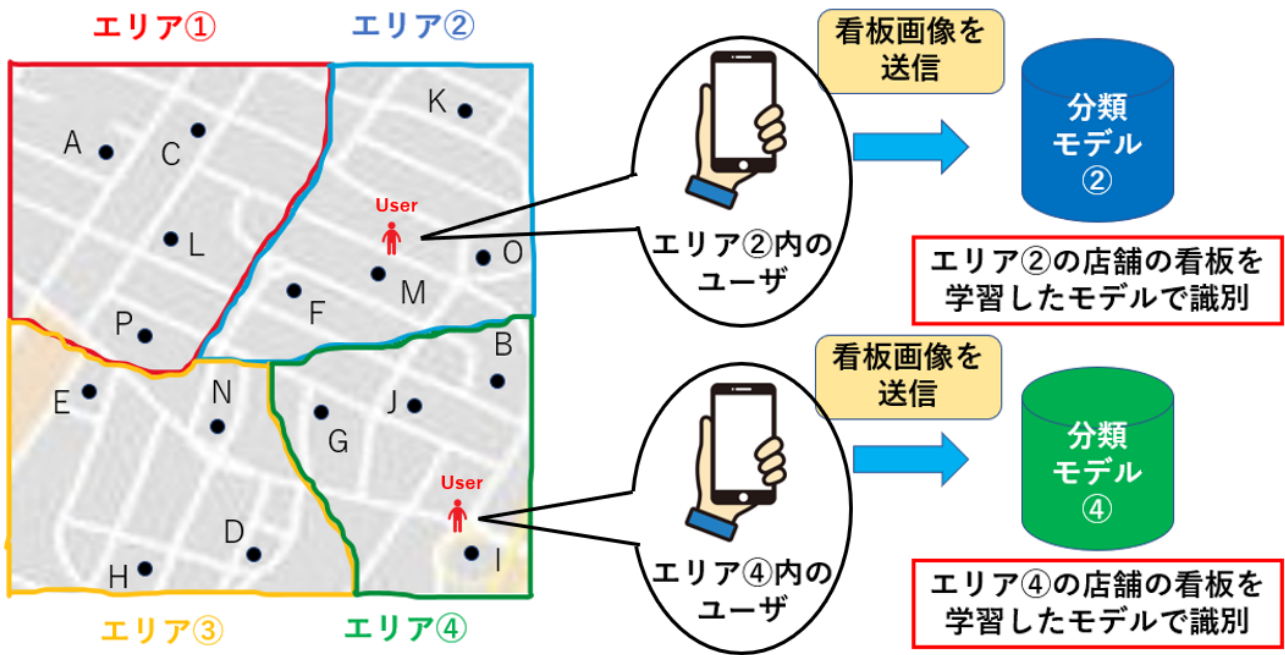


図1 ユーザの位置情報に基づく分類モデルの選定

表1 看板認識の処理を行うマシンのスペック

要素	スペック
CPU	Intel (R) Core (TM) i7-8700K @ 3.70 GHz
RAM	16 GB
GPU	NVIDIA GeForce GTX 1080 Ti
VRAM	12 GB
OS	Ubuntu 17.10

用できる店舗数が15店舗のみとなっており、特定の地域でしか使用できない。システムの実用性を上げるためには対象店舗数を増やす必要があるが、その場合は看板画像を分類する際に候補となる店舗数も同時に増加することになるため、それに伴って画像分類の精度が低下する懸念がある。そのため、システムの対象店舗を増加させることで適用範囲を拡大する場合、同時に画像分類の精度の低下を低減する必要がある。

4 提案手法

ユーザから送られてきた看板画像に対し、北村らのシステムでは対象店舗全ての看板画像を学習させた1つのモデルを使用して識別を行う。それに対して本手法では、予め特定の地域ごとに対象店舗の看板画像を学習さ

せた複数のモデルを構築し、その中からユーザが居る位置に応じて使用するモデルを選定して識別を行う。提案手法の概要を図1に示す。まず始めに、システムの対象地域を店舗間の地理的近接性に基づいて分割し、それぞれのエリア内に位置する店舗の看板を学習させた画像分類モデルを、各エリアごとに構築する。ユーザから送られてきた看板画像をもとに店舗を識別する際、ユーザの携帯端末の位置情報に基づいてどのエリアにユーザが居るかを判定し、ユーザが居るエリア内に位置する店舗の看板を学習させた画像分類モデルを用いて識別を行う。これにより、各モデルにおいて店舗を識別する際に候補となる店舗数は、全ての店舗の看板画像を学習させたモデルと比べて大きく削減できるため、精度の向上が期待できる。加えて、システムの対象地域が拡大することで識別の対象となる店舗が増えた場合でも、システムが識別すべき店舗の数はエリアに属する店舗数に限定されるため、分類精度の低下を防ぐことが可能となる。

5 実装

提案手法の有用性を確認するため、以下の手順をもとに実装を行う。

1. 画像分類モデル構築のための看板画像データセットの作成
2. 地域の分割に用いる対象店舗の位置情報データセットの作成

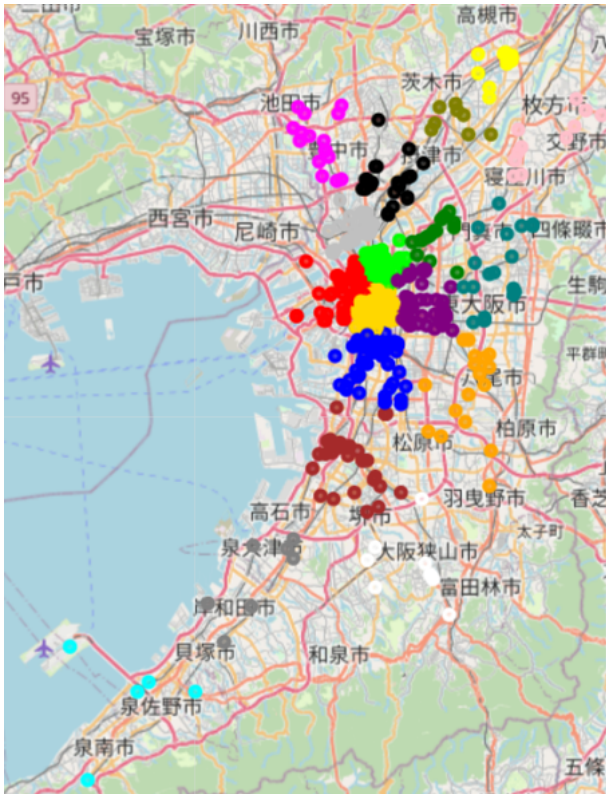


図2 対象地域の分割

3. 位置情報データセットをもとにした対象地域の分割
4. 分割した各エリアごとの画像分類モデルの構築

5.1 看板画像データセットの作成

北村らは看板画像データセットを作成する際、対象店舗の看板画像を様々な角度から撮影することで看板画像を収集している。しかし、店舗の看板を撮影して手動でデータセットを作成する場合、システムの対象店舗を増やす際に多大なコストがかかる。そのため、BeautifulSoup4によるスクレイピングを用いて、グルメサイトである食べログ*2から大阪府にある対象店舗の看板画像を自動で収集した。食べログでは地域を指定して店舗を検索することができ、大阪府は計44区域に分けられている。また、各店舗のページには、ユーザーによって投稿された店舗の外観の写真が多数掲載されている。本稿では対象地域から幅広く看板画像を収集するため、各区域ごとに人気ランキングページの上位から指定店舗数分の外観画像を全て収集した。

スクレイピングによって収集した外観画像に対し、

*2 <https://tabelog.com/> (2022/1/31 存在確認)

表2 各エリアの点の色と店舗数

エリア名	点の色	店舗数	エリア名	点の色	店舗数
エリア1	red	95	エリア10	orange	17
エリア2	brown	35	エリア11	white	9
エリア3	green	18	エリア12	pink	11
エリア4	blue	101	エリア13	silver	59
エリア5	yellow	20	エリア14	lime	195
エリア6	gray	7	エリア15	teal	18
エリア7	black	24	エリア16	aqua	5
エリア8	purple	95	エリア17	fuchsia	16
エリア9	gold	197	エリア18	olive	8

YOLOv2を用いて看板の物体検出を行った。北村らの物体検出モデルは、店舗の分類は行わず、看板の検出のみに特化しているため、様々な店舗の看板を検出できる汎用性を有する。このモデルを利用して信頼度が80%以上である看板領域を切り出し、切り出した看板画像が20枚以上存在している計930店舗を本システムの対象店舗として収集した。収集した各店舗の看板画像の60%をトレーニングデータ、20%をバリデーションデータ、20%をテストデータとしてデータセットを作成した。

5.2 対象店舗の位置情報データセットの作成

本手法では、前節でシステムの対象とした店舗同士の地理的近接性に基づいて対象地域の分割を行うため、前節で述べたスクレイピングを用いて各店舗のページに掲載されている住所を収集した。収集した住所に対し、対象店舗の居所の可視化及び電子的な処理を行うため、ジオコーディングを適用して各店舗の住所を緯度及び経度に変更した。ジオコーディングとは、住所、地名、郵便番号などが示す場所に対して地理座標を付与する処理を指す。本稿では、WebサービスであるGeocoding API*3を用いてジオコーディングを行った。このサービスでは、クエリとして住所やランドマーク名を入力して検索を行うことで、対応するGPS座標を表示できる。この変換作業をシステムが対象とする930店舗分全て手動で行う場合、多大な時間と労力を要するため、Selenium*4を用いて検索クエリの入力と検索の実行を自動で行った。その後、BeautifulSoup4を用いたスクレイピングによって検索結果から自動で対象店舗の緯度及び経度を収集し、対象店舗の位置情報データセットを作成した。

*3 <https://www.geocoding.jp/api/> (2022/1/31 存在確認)

*4 <https://www.selenium.dev/> (2022/1/31 存在確認)

表3 モデルの評価

モデル	店舗数	正解率	適合率	再現率	F 値
全店舗	930	0.799	0.714	0.625	0.667
エリア1	95	0.858	0.820	0.771	0.794
エリア2	35	0.839	0.854	0.819	0.836
エリア3	18	0.871	0.887	0.847	0.867
エリア4	101	0.821	0.826	0.772	0.798
エリア5	20	0.903	0.900	0.865	0.883
エリア6	7	0.911	0.923	0.907	0.915
エリア7	24	0.926	0.931	0.909	0.920
エリア8	95	0.825	0.826	0.785	0.805
エリア9	197	0.836	0.835	0.788	0.811
エリア10	17	0.858	0.897	0.846	0.871
エリア11	9	0.886	0.939	0.872	0.904
エリア12	11	0.94	0.959	0.919	0.939
エリア13	59	0.876	0.855	0.822	0.838
エリア14	195	0.824	0.805	0.765	0.785
エリア15	18	0.848	0.802	0.809	0.805
エリア16	5	1.0	1.0	1.0	1.0
エリア17	16	0.943	0.943	0.930	0.936
エリア18	8	0.857	0.896	0.846	0.870

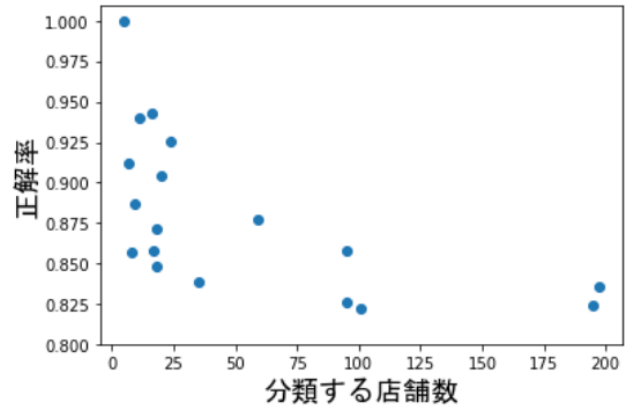


図3 分類する店舗数と正解率の関係

5.3 対象地域の分割

対象地域内に位置する対象店舗は、規則性に基づいて立地しているわけではないため、対象店舗間の地理的近接性をもとに対象地域を分割し、各エリアごとに大きさを変える必要がある。そのため、前節で作成した位置情報データセットをもとに、k平均法によるクラスタリングを行った。クラス数は各エリア内の店舗数と分類精度のトレードオフを考慮して18クラスに設定している。その後、図2のように地理的近接性に基づいて対象地域を分割した。各点は店舗の所在地を示しており、各点の色は店舗が分割したエリアのどこに位置しているかを示す。各エリアに対応する点の色と、各エリアに属する店舗の数を表2に示す。

各エリアに属する店舗数は様々であり、最も店舗数が多いエリア9は197店舗であったのに対し、最も店舗数が少ないエリア16は5店舗であった。都市部に近い店舗ほど、システムの対象とするために必要な量の看板画像を集めることができた店舗が多かったため、都市部に近いエリアであるほど多くの対象店舗が属していた。

5.4 看板の画像分類

5.1節で作成した看板画像データセットをもとに、CNNモデルであるVGG16を用いて画像分類を行う。VGG16とは、1000クラスに分類された130万枚の画像を学習させた、13層の畳み込み層と3層の全結合層の計16層からなるモデルである。本稿では北村の手法と同様に、全結合層を取り除き、新たに全結合層を追加する転移学習を行っており、学習に用いる画像が少ない店舗であって短時間で高精度なモデルを構築できる。

6 看板分類モデルの評価と考察

6.1 評価

提案手法の有用性を確認するため、対象地域にある全ての930店舗の看板画像を学習させたモデルと、対象地域を18区域に分割し、エリアごとに看板画像を学習させて構築したモデルの精度を比較した。各モデルの画像サイズを 224×224 、バッチサイズを16、学習率を 10^{-4} 、慣性を0.9とした。エポック数は、930店舗を分類するモデルは2000とし、エリアごとに看板画像を学習させて店舗を分類するモデルは、各モデルのクラス数に合わせて50-400とした。損失関数には、各モデルのクラス数の交差エントロピーを用いた。

評価指標には正解率及び適合率、再現率、F値を使用し、テストデータを用いて算出した。各モデルの分類する店舗数及び計算結果を表3に示す。全ての対象店舗の看板を学習させたモデルの正解率は0.799であったのに対し、対象地域を分割してエリアごとに構築した18区域のモデルの正解率の平均は0.879であった。また、エリアごとに構築した全てのモデルにおいて、全ての対象

店舗の看板を学習させたモデルと比較して正解率が高かった。以上の結果から、提案手法を用いることによる分類精度の向上が確認できた。

6.2 考察

対象地域を分割してエリアごとに構築した各モデルの、分類する店舗数と正解率の関係を図3に示す。分類する店舗数が多いモデルほど正解率が低い傾向を確認でき、最も正解率が高いモデルはエリア16の1.0であるのに対し、最も正解率が低いモデルはエリア4の0.821であり、エリアによってモデルの精度が大きく異なっていた。分割されたエリアの1つに店舗が集中した場合、その地域の店舗の看板画像を学習させたモデルの精度が大きく低下する懸念がある。そのため、各エリアごとの店舗数を可能な限り均一にすることで、各モデルの分類精度の差を低減するように地域を分割する必要がある。

エリアを細分化して1つのエリア辺りの店舗数を削減することで、各モデルの更なる精度向上が期待できる。しかしそれに伴って構築するモデル数も同時に増えることになり、構築するモデルが増えるほど多大なコストがかかる。そのため、1つのエリアあたりの店舗数と構築するモデル数のトレードオフを考慮して地域を分割する手法を検討する必要がある。

7 課題と展望

データセットとなる看板画像を食べログから収集した際、20枚以上収集できなかった店舗は、画像分類の精度を十分に確保できないためにシステムの対象外としている。しかし、大阪府には食べログに載っている店舗だけでも6万店舗以上あり、システムが適用可能な店舗はごく一部に留まっている。今後システムの適用範囲を広げる場合、看板画像を収集する手法や看板画像を学習させる手法を検討する必要がある。

看板画像データセットの中には、建物の壁面の一部や張り紙のような、学習に不要な画像が多く混ざっていた。これらの画像の混入がシステムの精度に影響を与えていると考えられる。そのため、更なる分類精度の向上のためにはデータセットに含まれる不要なデータを取り除くか、看板の物体検出の精度を向上させる必要がある。

8 終わりに

本稿では、看板画像を用いた店舗情報検索システムにおいて、画像分類モデルの精度を向上させる手法を提案し、その有用性を確認した。今後は、6.2節で述べた考察

をもとに、地域の分割手法の改良を行う。加えて、ユーザの位置情報に基づいて自動で識別する際に使用するモデルを選定できるようにシステムの改良を行い、実世界で看板画像の識別精度が向上するかを検証する。

参考文献

- [1] Abadi, M. et al.: TensorFlow: A System for Machine Learning, *12th USENIX Symposium on Operating Systems Design and Implementation*, pp. 265–283 (2016).
- [2] Fujita, K. et al.: Data Augmentation using Evolutionary Image Processing, *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–6 (2018).
- [3] Haklay, M. and Weber, P.: OpenStreetMap: User-Generated Street Maps, *IEEE Pervasive Computing*, pp. 12–18 (2008).
- [4] Kavati, I. et al.: Signboard text translator: A guide to tourist, *International Journal of Electrical and Computer Engineering*, Vol. 7, No. 5, pp. 2496–2501 (2017).
- [5] Liu, S. and Deng, W.: Very deep convolutional neural network based image classification using small training sample size, *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 730–734 (2015).
- [6] Panhwar, M. A. et al.: Signboard Detection and Text Recognition Using Artificial Neural Networks, *IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 16–19 (2019).
- [7] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525 (2017).
- [8] 北村茂生, 松下光範: オンサイト検索: 携帯端末を用いた看板画像からの店舗アクセス手法, *電子情報通信学会論文誌*, Vol. J103-D, No. 5, pp. 364–372 (2020).
- [9] 河野曜平, 川本一彦: GANを用いたデータ拡張, *情報処理学会研究報告*, Vol. 2017-CVIM-207, No. 14, 情報処理学会, pp. 1–5 (2017).
- [10] 野村松信ほか: 背景色情報を活用した看板内の文字列領域抽出法, *電気学会論文誌C*, Vol. 134, No. 1, pp. 121–130 (2014).