

未知ゴミ識別のための判定手法の比較と実用化に向けた検討

立花 巧樹^{1,2} 呉 健朗² 松田 裕貴^{1,3,4} 諏訪 博彦^{1,4} 安本 慶一^{1,4}

概要：ゴミのポイ捨ては野生動物の殺傷に繋がるなど深刻な環境問題に発展している。ポイ捨てを防止する策を講じるために必要な、ポイ捨てゴミの実態やゴミ拾い活動の実績は情報として整理されていない。我々は、カメラ装着型トングを用いて、ゴミを拾うだけで、ゴミの種別・位置情報を網羅的に収集するシステムを提案してきた。しかし、現在のシステムは画像認識モデルのため、学習データに含まれていないゴミが拾われた際に、誤認識が不可避である。この問題を解決するため本稿では、未知のゴミをその他として識別する手法を列挙し、比較検討する。実験結果より、多クラス分類モデルを用いることで、F 値 77%の精度を達成した。

A Method of Littering Garbage Type Recognition using Camera-mounted Tongs with Object Detection Model

KOKI TACHIBANA^{1,2} KENRO GO² YUKI MATSUDA^{1,3,4} HIROHIKO SUWA^{1,4}
KEIICHI YASUMOTO^{1,4}

1. はじめに

ゴミのポイ捨ては社会問題に発展している。Maria によると、世界では、年間 6 兆本ものタバコが消費されているが、そのうち 4.5 兆本はポイ捨てされている [1]。また、世界経済フォーラムは、2015 年にタバコを含む 910 万トンのプラスチックゴミが流出したという調査結果を発表した [2]。これにより、海洋汚染や海洋生物の殺傷などの問題が生じている。このような問題を防ぐために、ポイ捨てが多く行われる場所を特定し、効率的にゴミ箱を配置するような試みが行われている。例えば、ゴミを拾った人が、ゴミを撮影して位置情報と共にアップロードすることで、ゴミの種別・位置情報を収集できるピリカというサービスが存在する [3]。我々も、ゴミを拾うだけで、ゴミの種別・位置情報を網羅的に収集可能なトングを提案してきた [4], [5]。このシステムでは、トングの先端に内視鏡カメラが取り付けられており、トングでゴミを掴む動作を行って

いるかの判定を常に行っている。トングでゴミを掴む動作を行うと、内視鏡カメラの映像を通し、トングで何の種類ゴミを掴んだかを画像認識で推定する。この推定結果と位置情報を結びつけ、ゴミを拾うだけで、ゴミの種別・位置情報を網羅的に収集できるようにしている。我々は研究の社会実装を目指し、ボランティアの協力の下、数多くのイベントを開催している [6]。しかし、ゴミ拾い活動においては、缶、ペットボトル、タバコ、紙類等の推定対象外のゴミも多数存在し、既存研究で提案した CNN ベースの画像認識モデル [4], [5] では、ビンやつまようじなど推定対象外のゴミも、推定対象のカテゴリに誤分類される問題がある。このため、収集したデータの信頼性が損なわれ、アノテーションの再作業など追加の労働コストが発生する。本稿では、この問題に対処するために、未知のゴミをその他として識別する手法を比較する。比較を行い、現状における最も実用的なその他判定システムを検討することを本稿の目的とする。実験結果より、多クラス分類モデルを用いることで、F 値 77%の精度を達成した。

2. 関連研究

ゴミの種別・位置情報を手動で収集する事例について 2.1 節、自動で収集する事例について 2.2 節にて述べる。

¹ 奈良先端科学技術大学院大学 〒 630-0192 奈良県生駒市高山町 8916-5

² 株式会社 mica

³ 岡山大学 Okayama University

⁴ 理化学研究所 革新知能統合研究センター

RIKEN Center for Advanced Intelligence Project

2.1 ゴミの種別・位置情報を手動で収集する事例

ゴミの種別・位置情報を手動で収集する事例として [7], [8] がある。井上ら [7] は、釜口水門から横川川合流点までの天竜川上流 11.5 km の河岸において、漂着ゴミの量、種類を手動で収集しながら調査し、漂着ゴミの分布および岸形状との関係を考察した。高橋ら [8] は、計測者が手動でタバコの位置情報を地図上にプロットし、タバコの種別・位置情報の収集を手動で行っている。その結果、タバコの多い場所に対して、灰皿を設置してポイ捨てを削減を狙うなどのタバコのポイ捨ての防止対策に関する具体的方策を提案している。

しかし、これらの手法では、ゴミ拾いをしている人が、一つ一つのゴミに対し手動で収集するため、手間がかかるという問題が存在する。したがって、ゴミの種別・位置情報を自動で検出することが必要である。

2.2 ゴミの種別・位置情報を自動で収集する事例

ここでは、ゴミの種別・位置情報を自動で収集する事例 [3], [9], [10] を紹介する。

植田ら [9] は、定点ビデオカメラによるポイ捨て位置の情報やポイ捨てしている人の特徴を収集する方法を提案している。しかし、この手法は、カメラが設置してある場所でしかゴミ情報を収集できないため、網羅的に情報収集することができないという問題が存在する。また、多くの場所にカメラを設置している環境があったとしても、高いコストがかかることが考えられる。ピリカ社は、スマホを用いた「ポイ捨てゴミ」や「歩きたばこ」の分布や深刻さを調査するサービスである、タカノメ [10] を提供している。このサービスでは、計測者がスマホで道に落ちているゴミを撮影すると、システムは、動画に写ったゴミの種類や数量、位置情報などを解析し、地図上にプロットするという仕組みが導入されている。しかし、この方法では、データ計測者がスマホで道路を撮影しながら歩くという動作が必要であることと、それ故に調査範囲に応じた費用が生じることから、データ収集の持続可能性が低いと考えられる。これに対し、ゴミを拾う人からデータを収集する「ユーザ参加型センシング [11], [12]」のアプローチであれば、センサなどの設置コストがかかることなく、ゴミの種別・位置情報を網羅的に収集できる可能性がある。

2.3 本稿の立ち位置

ゴミのポイ捨ては、海洋汚染や海洋生物の殺傷などにつながる行為であり、社会問題に発展している。このような問題を防ぐために、ポイ捨てが多く行われる場所を特定し、効率的にゴミ箱を配置するような試みが行われている。しかし、どのようなゴミがどのような場所でポイ捨てされているのかを把握することは人的コストが高く、世界中のすべてのゴミのポイ捨て情報を網羅的に収集するこ

とは困難である。このことから、我々はカメラ搭載型トングでゴミを拾うだけで、ゴミの種別と位置情報を推定するシステムを提案している [4], [5]。また、本システムの社会実装に向け、イベントを開催し、情報収集活動を行っている [6]。しかし、ゴミ拾い活動においては、缶、ペットボトル、タバコ、紙類等の推定対象外のゴミも多数存在し、既存研究で提案した CNN ベースの画像認識モデル [4], [5] では、ビンやつまようじなど推定対象外のゴミも、推定対象のカテゴリに誤分類される問題がある。このため、収集したデータの信頼性が損なわれ、アノテーションの再作業など追加の労働コストが発生する。本稿では、この問題に対処するために、未知のゴミをその他として識別する手法を比較する。比較を行い、現状における最も実用的なその他判定システムを検討することを本稿の目的とする。将来的には、その他に分類されたゴミに対して適切にアノテーションを行い、それを学習データとして活用することで、多様なゴミの種別を認識できるモデルの構築を目指す。これにより、ゴミの分類精度をさらに向上させ、より幅広い種類のゴミを正確に識別できる。

3. 比較手法・実験

推定対象外のゴミをその他と判定可能なシステム構築方法を 4 つ提案し、精度の比較実験を行う。

3.1 手法 1: 多クラス分類を用いたその他判定システム

既存手法で構築した MobileNet [4] および ResNet50 [5] の画像認識モデルの識別対象にその他ラベルを追加する多クラス分類モデルを構築することで、推定対象のゴミは従来通りに分類され、推定対象外のゴミはその他として分類されることを期待する。以降、MobileNet で構築したモデルを手法 1-1 とし、ResNet50 で構築したモデルを手法 1-2 とする。

データセットについて説明する。はじめに、事前学習モデル構築のために画像のオープンソースサイトである Roboflow [13] から、缶、ペットボトル、タバコ、紙類の 4 種類のデータを収集した。収集したデータ画像を目視で確認し、誤ってアノテーションされている画像を除外した。結果として、缶 2263 枚、ペットボトル 2047 枚、タバコ 2717 枚、紙類 3265 枚の画像を収集した (図 1)。次にポイ捨てゴミ認識モデル構築のために、実際にイベントで収集したゴミのデータに対しアノテーションを行った (図 2)。結果として、ゴミの種別ごとに 300 枚の画像データを用意した (表 1)。

モデル構築方法について説明する。まず、ImageNet で学習済みの MobileNet および ResNet50 の重みを用いて、Roboflow で収集した画像を用いてファインチューニングを行い、事前学習モデルを構築した。次に、事前学習モデルに対し、最終層を 4 値 (缶・ペットボトル・タバコ・紙

類)分類から5値分類(缶・ペットボトル・タバコ・紙類・その他)に変換し、イベントで収集した画像データセットを用いてファインチューニングを行った。具体的には、訓練データ、検証データ、テストデータをゴミの種別ごとに240:30:30に分割して学習し、最終的にテストデータの精度を算出した。以降、各手法において使用するテストデータは、すべて同一のデータセットを用いて精度を算出した。

3.2 手法 2:2 値分類モデルを組み合わせたその他判定システム

既存手法で構築した MobileNet [4] および ResNet50 [5] を用いて、缶、ペットボトル、タバコ、紙類の4種類に対する2値分類モデルを構築する。具体的には、「缶である」・「缶でない」、「ペットボトルである」・「ペットボトルでない」、「タバコである」・「タバコでない」、「紙類である」・「紙類でない」という2値分類モデルを4つ構築する。以降、MobileNet で構築したモデルを手法 2-1 とし、ResNet50 で構築したモデルを手法 2-2 とする。

事前学習モデルに対し、最終層を4値分類(缶・ペットボトル・タバコ・紙類)から2値分類(缶・缶でない...)に変換し、ファインチューニングを行った。具体的には缶の場合、3.1 節と同様に缶のデータセットを訓練データ、検証データ、テストデータを240:30:30で分け、缶でないデータセットをランダムに抽出し、240:30:30で分けてモデル構築した。その他の判別手法について説明する。4つの画像認識モデルにおいて、「○○でない」という推定結果が出た場合、その物体はその他として判定する。このアプローチにより、既知のカテゴリに該当しないゴミを正確に識別し、その他として判定することが可能になると考える。本手法を用いて、テストデータの精度を算出した。

3.3 手法 3: 物体検出モデルを用いたその他判定システム

その他判定システムの新たなアプローチとして、物体検出モデルを用いた手法を検討する。物体検出モデルのアーキテクチャには Yolo [14] を採用した。Yolo は、物体検出タスクのためのディープラーニングベースのアルゴリズムであり、画像全体を一度に解析することで物体の位置とクラスを同時に予測するという特徴がある。このアプローチにより、リアルタイム性に優れ、精度の高い物体検出が可能である。本稿では、Yolo の最新モデルである Yolov8 [15] を採用した。

物体検出モデルの構築手順を以下に述べる。まず、事前学習モデルを構築するために、Roboflow [13] という Yolo 形式でアノテーションされたデータセットがオープンソースとして公開されているサイトから、缶、ペットボトル、タバコ、紙類の4種類のデータを収集した。収集したデータ画像を目視で確認し、誤ってアノテーションされている画像を除外した。その結果、缶 2263 枚、ペットボトル

2047 枚、タバコ 2717 枚、紙類 3265 枚の画像を収集した(表 1)。収集した画像データを学習データと検証データを9:1の比率で分割して学習を行い、缶、ペットボトル、タバコ、紙類を検出する事前学習モデルを構築した。ファインチューニングには、3.1 節と同様のデータに対し馬雲ディボックスでアノテーションを行い、画像データセットを構築した。その結果、缶 300 枚、ペットボトル 300 枚、タバコ 300 枚、紙類 300 枚の画像を収集した(表 1)。収集した画像データを学習データと検証データに240:30:30の比率で分割して学習を行い、物体検出モデルを構築した。

その他判定システムについて説明する。入力画像に対して、画像内に確信度0.5以上の缶・ペットボトル・タバコ・紙類が検出できなかった場合、その他と判定する。本アプローチを用いることで、ゴミの認識を行いつつ学習していないゴミのデータをその他として識別することができると考えた。本手法を用いて、テストデータの精度を算出した。

3.4 手法 4:LLM を用いたその他判定システム

ChatGPT (Conversational Generative Pre-trained Transformer) [16] は、OpenAI によって開発された自然言語処理モデルである。GPT-4 アーキテクチャに基づいており、大規模なテキストデータセットを用いて事前学習されている。このモデルは、Transformer というディープラーニングの手法を活用しており、特に自己注意機構(self-attention mechanism)を用いた文脈理解に優れている。ChatGPT の主な機能は、入力されたテキストに対して自然で適切な応答を生成することである。これにより、ユーザーとのインタラクションを実現し、様々な質問に答えることが可能となっている。モデルの学習過程には、教師なし学習(unsupervised learning)を用いており、大量のインターネットデータを基にして、言語パターンや文脈を理解し、予測する能力を持っている。

本稿では、ChatGPT-4o を利用してその他判定システムを構築した。具体的には「The photo shows tongs grabbing garbage with the top and bottom. What type of garbage are you grabbing? Please select from 'can', 'PET bottle', 'tobacco', 'paper' and 'plastic'. Answer only the name of the type of garbage. If none of these apply to you, please select 'other'. (写真は上下のトングでゴミを掴んでいます、どのような種類のゴミを拾っていますか? 「缶」・「ペットボトル」・「タバコ」・「紙類」から選択してください。ゴミの種類名のみ選択してください。どれにも当てはまらない場合はその他を選択してください。)」というテキストデータと、画像を送信し、レスポンスを取得する。その結果を正解データとして精度の算出を行った。

3.5 実験結果

全体の実験結果を表 2 に示す。以降、混同行列(図 3-

表 1 学習時に利用したゴミの種別ごとのデータ数

Table 1 Number of data for each type of garbage used during learning.

手法	缶	ペットボトル	タバコ	紙類	その他
事前学習に用いたオープンソースのデータセット	2263	2047	2717	3265	-
ゴミ拾いイベントで収集したデータセット	300	300	300	300	300



図 1 オープンデータセットで収集したデータサンプル



図 2 ゴミ拾いイベントで収集したデータサンプル

図 8) のラベルは、「缶」を「can」, 「ペットボトル」を「pet.bottle」, 「タバコ」を「cigarettes」, 「紙類」を「paper」, その他を「others」として示す。

はじめに手法 1-1, 手法 1-2 について考察する (図 3, 図 4)。手法 1-1 の Accuracy, Precision, Recall, F 値はそれぞれ 0.63, 0.63, 0.63, 0.62 となり, 手法 1-2 の Accuracy, Precision, Recall, F 値はそれぞれ 0.77, 0.78, 0.77, 0.77 となった。ここから, MobileNet に比べ, ResNet50 の方が精度が高いことがわかる。また, 手法 1-2 の混同行列図 4 に注目すると, 紙類以外において, Recall が 80% 以上の値となっていることがわかる。特に, その他においては, 表 3 にある通り, Recall においては手法 4 と同率で他手法と比較して最も精度が高く, Precision, F 値において他手法と比較して最も精度が高い。紙類の精度が低い原因に関しては, 図 2 からわかる通り, マスクからティッシュ, ダンボールなど様々な種類が多くあるため, 誤認識してしまったと考えられる。現状 240 枚のみしか学習データとして用いていないため, 今後は, より多くのデータを学習させることで精度の向上が期待される。

次に手法 2-1, 手法 2-2 について考察する (図 5, 図 6)。手法 2-1 の Accuracy, Precision, Recall, F 値はそれぞれ 0.83, 0.85, 0.83, 0.82 となり, 手法 2-2 の Accuracy, Precision, Recall, F 値はそれぞれ 0.85, 0.87, 0.85, 0.84 となった。ここから, 手法 1 と同様に MobileNet に比べ, ResNet50 の方が精度が高いことがわかる。また, 表 2 から Accuracy, Precision, Recall, F 値において多手法と比較して最も値が高いことがわかった。手法 2-2 の混同行列 (図 6) に注目すると, 缶・ペットボトルの Recall は 100% となっており, 紙類の Recall は 90%, 以上, タバコの Recall は 80% 以上の精度を算出している。しかし, その他の Recall は 50% に留まっている。これは, 実際にその他のゴミのうち, 50% しかその他と推定できないことを示している。つまり, その他のゴミの半分が缶・ペットボトル・タバコ・紙類のどれかに分類されてしまうため, 実用的ではない。今後の改善として, 現状の閾値は 0.5 としているが, 閾値のチューニングを行うことで, 精度の向上が期待される。

手法 3 について考察する (図 7)。手法 3 の Accuracy, Precision, Recall, F 値はそれぞれ 0.72, 0.74, 0.72, 0.72 となった。手法 3 の混同行列 (図 7) に注目すると, 80% 以上の Recall は缶のみとなった。精度が他手法と比較して低下している理由を検討するために, 缶・タバコ・紙類の誤認識の多くがその他となっていることに注目する。手法 3 では, Yolo で検出できなかった場合, その他と判定するシステムとなっている。そのため今後は, 閾値を低く設定することで, ゴミの種別を検知しやすくすることで, 精度の向上を見込める可能性がある。しかし本アプローチは, ゴミの検出率低下にも繋がる可能性が高く, 慎重に行う必要がある。したがって, 現状だと Accuracy, Precision, Recall, F 値が 70% 程度しかないため, 手法 1-2 に比べると実用的ではない。

最後に手法 4 について考察する (図 8)。手法 4 の Accuracy, Precision, Recall, F 値はそれぞれ 0.81, 0.87, 0.81, 0.79 となった。手法 4 の混同行列 (図 7) に注目すると, 缶の Recall は 100% となっており, ペットボトル・紙類の Recall は 90%, 以上, その他の Recall は 80% 以上の精度を算出している。しかし, タバコの Recall は 27% に留まっている。これは, 図 2 の通り, タバコは画像内でもかなり小さいので ChatGPT-4o で認識できなかった可能性が高い。タバコの認識精度が著しく低いいため, 全体的にゴミの

表 2 実験結果 (全体)

	Accuracy	Precision	Recall	F 値
手法 1-1	0.63	0.63	0.63	0.62
手法 1-2	0.77	0.78	0.77	0.77
手法 2-1	0.83	0.85	0.83	0.82
手法 2-2	0.85	0.87	0.85	0.84
手法 3	0.72	0.74	0.72	0.72
手法 4	0.81	0.87	0.81	0.79

表 3 実験結果 (その他)

	Precision	Recall	F 値
手法 1-1	0.61	0.77	0.68
手法 1-2	0.69	0.83	0.76
手法 2-1	0.40	0.50	0.45
手法 2-2	0.40	0.50	0.45
手法 3	0.51	0.63	0.57
手法 4	0.52	0.83	0.64

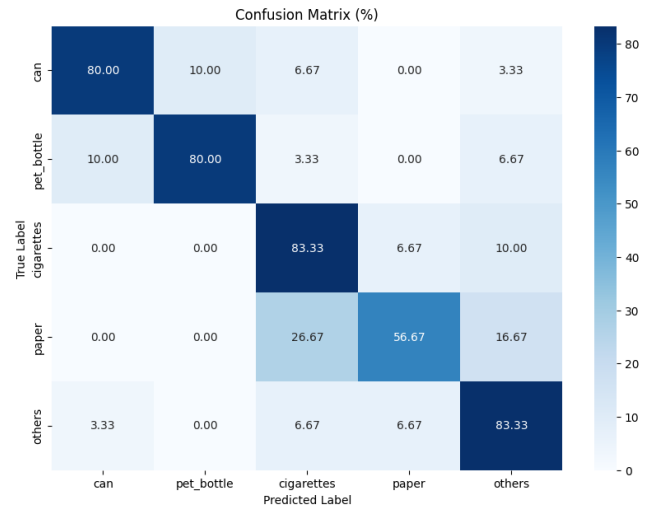


図 4 手法 1-2 の混同行列

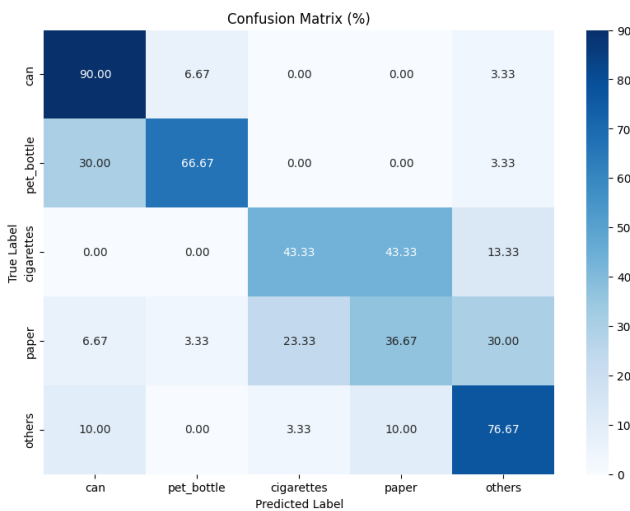


図 3 手法 1-1 の混同行列

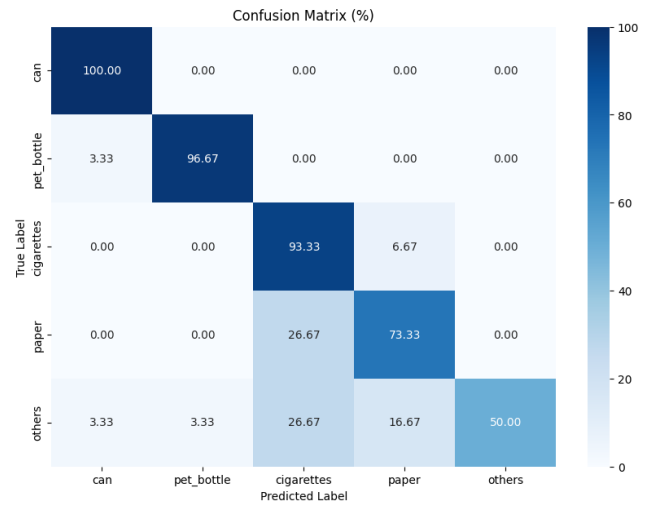


図 5 手法 2-1 の混同行列

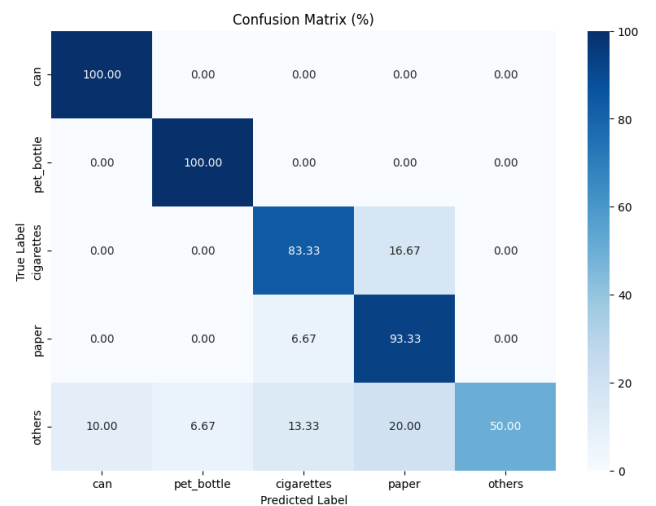


図 6 手法 2-2 の混同行列

種別を推定可能な手法 1-2 と比較すると実用的ではないと考える。今後は、小さなタバコでも検出できるようにプロンプトの調整を行い、精度向上を行う。

以上より、紙類以外全て Recall が 80%以上を算出しており、紙類に関しても画像数を増やすことで精度向上も見込める手法 1-2 が最も実用的であると判断した。今後は手法 1-2 で構築したモデルを実用的に利用する。さらに、その他と推定したゴミの画像をアノテーションすることで再学習し、モデルの精度向上を行う方式を取ることで認識モデルの精度向上を引き続き行う。

4. おわりに

本稿では、この問題に対処するために、未知のゴミをその他として識別する手法として多クラス分類モデルを利用したその他判定システム、2 値分類モデルを利用したその他判定システム、物体検出モデルを利用したその他判定システム、LLM を利用したその他判定システムを検

討し比較した。結果として、推定対象にその他を追加し、Resnet50 で構築した多クラス分類モデルを用いることで F 値 0.77 を達成することができた。今後は、紙類の様々

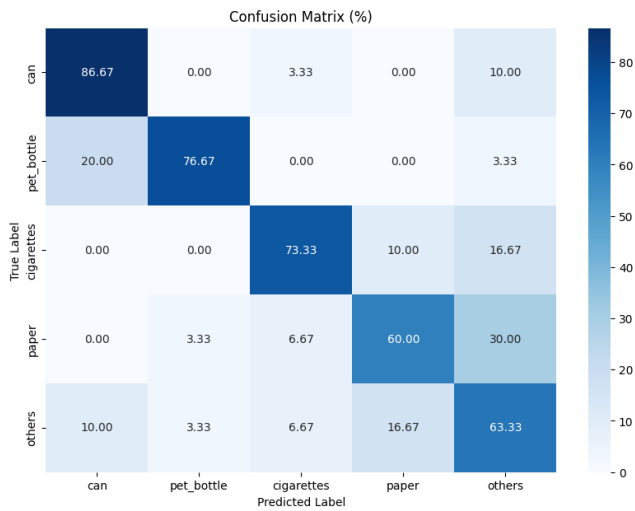


図 7 手法 3 の混同行列

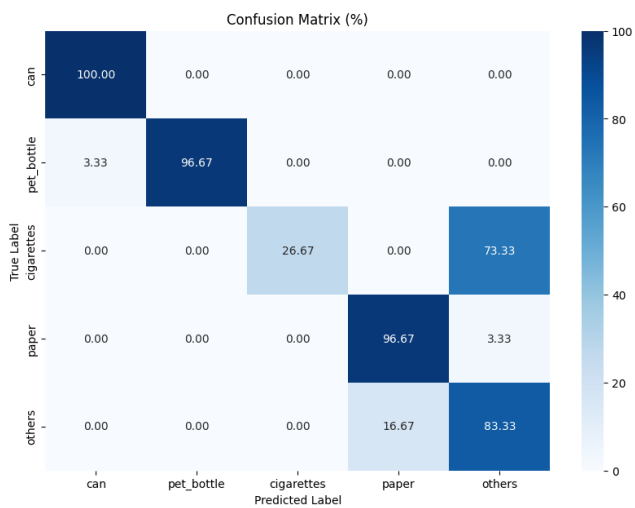


図 8 手法 4 の混同行列

なデータセットを学習し、精度向上を図る予定である。また、Zero-shot-learning の手法を用いてその他を推定するための判定システムの構築を行う予定である。

参考文献

- [1] Maria Araújo and Monica Costa. A Critical Review of the Issue of Cigarette Butt Pollution in Coastal Environments. *Environmental Research*, Vol. 172, pp. 137–149, 2019.
- [2] The World Economic Forum. The New Plastics Economy Rethinking the future of plastics. http://www3.weforum.org/docs/WEF_The_New_Plastics_Economy.pdf, 2016. (Accessed: 2023-03-19).
- [3] 株式会社ピリカ. ごみ拾い SNS ピリカ—いつでもどこでも気軽にボランティア. <https://sns.pirika.org/>. (Accessed: 2020-07-27).
- [4] 立花巧樹, 中村優吾, 松田裕貴, 諏訪博彦, 安本慶一. スマートウォッチの音響センサを用いたポイ捨てごみの種別認識手法の提案と評価. マルチメディア, 分散, 協調とモバイル (DICOMO 2021) シンポジウム論文集, pp. 325–332, 2021.
- [5] 立花巧樹, 呉健朗, 松田裕貴, 諏訪博彦, 安本慶一. ポイ捨てごみ情報収集システムにおける参加モチベーション向

上に向けた音フィードバックの比較. マルチメディア, 分散, 協調とモバイル (DICOMO 2023) シンポジウム論文集, pp. 1304–1309, 2023.

- [6] 奈良先端科学技術大学院大学. センサ装着型トングを用いたポイ捨てごみの種別・位置情報収集システムの運用に向け、奈良先端科学技術大学院大学と JT が「IoT トングでひろえば街が好きになる運動 with Rethink PROJECT」を生駒市で開催. <https://www.naist.jp/pressrelease/files/20220222.pdf>. (Accessed: 2023-03-19).
- [7] 井上芳樹, 戸田任重. 諏訪湖・天竜川上流における漂着ゴミ. *環境科学会誌*, Vol. 16, No. 3, pp. 167–178, 2003.
- [8] 高橋祐平, 石坂公一, 小地沢将之. タバコのポイ捨てポテンシャルの分布構造—仙台市中心部のアーケード街を対象として—. *日本建築学会技術報告集*, 第 15 巻, pp. 257–260, 2009.
- [9] 植田憲, 高野維斗, 神崎広史, 宮崎清. ごみの「ポイ捨て」の未然防止に関する調査・研究: 千葉市・駅周辺地域におけるごみ捨て行為の実態調査に基づいて. *日本デザイン学会研究発表大会概要集*, 第 53 巻, p. 188, 2006.
- [10] 株式会社ピリカ. ポイ捨てごみ調査サービス「タカノメ」—スマホで簡単、ポイ捨て分布調査. <https://research.pirika.org/>. (Accessed: 2020-07-27).
- [11] J. A Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. *Center for Embedded Network Sensing*, pp. 1–5, 2006.
- [12] Yuki Matsuda, Shogo Kawanaka, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. ParmoSense: Scenario-based Participatory Mobile Urban Sensing Platform with User Motivation Engine. *Sensors and Materials*, Vol. 34, No. 8, pp. 3063–3091, 2022.
- [13] Roboflow. Roboflow: Give your software the power to see objects in images and video. <https://roboflow.com/>. (Accessed: 2023-10-26).
- [14] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 779–788, 2016.
- [15] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. Real-time flying object detection with yolov8, 2023.
- [16] openai. ChatGPT. <https://chatgpt.com/>. (Accessed: 2024-05-21).