

# 深層学習を用いたゴミ収集動画からゴミ袋個数の測定

## Counting garbage bags from garbage collecting video using deep learning

三上量弘<sup>\*1</sup> 河野 慎<sup>\*1</sup> 陳 寅<sup>\*1</sup> 中澤 仁<sup>\*1</sup>  
Kazhiro Mikami Makoto Kawano Yin Chen Jin Nakazawa

<sup>\*1</sup>慶應義塾大学  
Keio University

Reducing garbage generated in our daily life is a key aspect to fulfill a sustainable human society. The lack of an effective method to obtaining information regarding the collected garbage, such as where, when and what kinds of waste are collected, is one of the major obstacles to promoting the garbage reduction in cities. In this research, we propose a system to automatically count the number of garbage bags, by using convolutional neural networks (CNN)-based technology to process the videos of garbage collecting process taken by a camera mounted at the back of garbage trucks. The proposed system is evaluated using the realistic garbage collecting videos of Fujisawa city. The experiment result shows that the proposed method achieves a high recall score of 0.90.

### 1. はじめに

持続可能な社会を実現するために、ゴミ減量化運動が日本各地で行われている。ゴミを減量するためには、人々に自分がどれだけゴミを出したのかや削減したのかなどを知らせることによって、ゴミ減量への意識を高める必要がある。しかし、現在得られるゴミの量のデータ粒度は図1のように市や区単位といったものであり、人々のゴミ減量への意識を高めるにはさらに微細粒度なゴミの量のデータが必要となる。これらのゴミ量のデータを人で集めることは可能であるが、コストの観点から現実的ではない。したがって、自動的にそれらのデータを収集可能なシステムが必要である。

そこで、深層学習による物体検出と清掃車に取り付けられたドライブレコーダーに着目した。清掃車の後部にはドライブレコーダーが取り付けられ、清掃員らがゴミを清掃車のトランクに投入する様子が全て撮影されている。これらの動画からゴミ袋の種類と数を検出できれば、清掃車が街の中でゴミを収集しながら、人間を介せずに、収集されたゴミ袋の種類と数を検出することができる。清掃車にセンサを取り付けた様々な研究が行なっている [中澤 17, Kawano 17]。図2のように清掃車は網羅的に市内を走行している [Chen 17]。これらの清掃車に提



図2: 慶應大学が開発した清掃車を用いたセンサーネットワークのカバレッジ

案するシステムを実装することによって、微細粒度のゴミの収集量を自動的に取得できると考えられる。

本研究では、清掃車に取り付けられたドライブレコーダーで撮影された動画とGPSによって取得された位置情報を用いて、地域の微細粒度なゴミの量を自動的にリアルタイムセンシングするシステムを提案する。今回は、畳み込みニューラルネットワーク (CNN) を用いて動画のフレームにおいて目標のゴミ袋を検出、トラッキングし、清掃車に投げ込まれたかを判定するアルゴリズムによりゴミ袋を数える手法を実装した。また可燃ゴミを例としてデータセットを作成し、CNNで可燃ゴミを検出できるように学習した。実際に藤沢市を走行する清掃車後部に取り付けられたドライブレコーダーで撮影された動画を用いて実験を行い、手法の精度について評価を行う。

### 2. 微細粒度な地域のゴミ收拾量測定システム

#### 2.1 システム構成

システムの全体図を図3に示す。清掃車後部に設置されたカメラでゴミが投げ込まれる動画を撮影し、動画からGPUが搭載された小型コンピューターでゴミ袋の種類と個数を検出して、各種類のゴミの個数と位置情報をデータサーバに3Gネットワークで送信する。これにより微細粒度な地域毎のゴミ量を自動的に取得することが可能になる。また、動画そのものを送る必要がないので、市民の個人情報の安全性の観点や、通信コストや維持コストにおいても優れている。

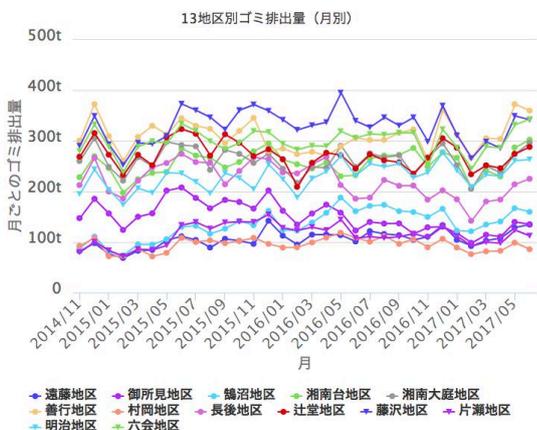


図1: 藤沢市各地区のゴミ排出量の推移 [筑波]

連絡先: 三上量弘, 慶應義塾大学, quantan@ht.sfc.keio.ac.jp



図 3: システムの全体図

## 2.2 ゴミの個数カウント手法

動画に写る物体のカウントには、物体検出、トラッキング、カウントの判定といったプロセスが必要となる。ゴミ袋の個数を清掃車のトラックに投げ込まれた個数と考え、清掃車に取り付けられた動画を用いてゴミ袋の個数をカウントするには以下の3つのプロセスが必要である。

1. 動画の各フレームに映るゴミ袋を検出
2. フレーム間において、検出されたゴミ袋を同定
3. ゴミ袋が清掃車に投げ込まれたかを判定

提案手法では、物体検出可能な畳み込みニューラルネットワーク (CNN) である Single shot multibox detector (SSD) を用いて動画のフレームにおいて目標のゴミ袋を検出、Intersection over Union (IoU) によるトラッキングアルゴリズムを用いて検出したゴミ袋をフレーム間で同定、清掃車に投げ込まれたかを判断するアルゴリズムによりゴミ袋の個数を数える。手法のフローチャートを図4に示す。

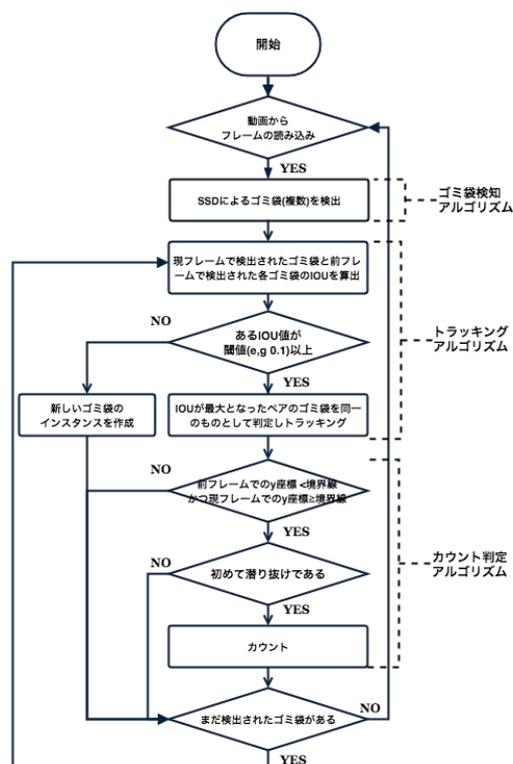


図 4: ゴミの個数カウントのフローチャート

### 2.2.1 ゴミ袋の検出

ゴミ袋の検出には物体検出可能な CNN である Single shot multibox detector (SSD) を用いる [Liu 16]. SSD はベースとなる CNN に VGG-16[Simonyan 14] を使用し、VGG-16 の fully-connected 層の代わりに新たに 6 層の畳み込み層を加え、end-to-end で物体検出をすることが可能である。物体検出可能な CNN は SSD の他にも多くあるが、その中でも SSD は検出の精度と速度において優れている。SSD は入力画像が異なる SSD-300 と SSD-500 の 2 つのタイプがある。リアルタイム性を考慮し、検出精度は劣るが速度が高い SSD-300 を今回は採用した。また SSD よりも精度と速度の両方において優れている YOLOv2[Redmon 16] が存在するが、画像に大きく写る物体の検出において SSD は検出精度が優れている。この性質が図6のように動画の手前で行われるカウントの判定において効果的と考え、ゴミ袋検出に SSD を採用した。

### 2.2.2 トラッキングアルゴリズム

動画に映る物体のトラッキングアルゴリズムは、カラーヒストグラムを利用する Mean shift[Fukunaga 75] がある。しかし、SSD によって出力されるのはゴミ袋の画像における領域情報である。そのため、トラッキングアルゴリズムには領域情報を使用する IoU を用いた。図5のように IoU は 2 つの領域を和と積で割ったものであり、重なっている割合を表す値である。SSD によって検出された現在のフレームのゴミ袋と 1 つ前のフレームのゴミ袋全ての IoU を算出し、閾値以上かつ IoU が最大となったゴミ袋のペアを同一のものとして、トラッキングする。閾値を越えなかった場合はそのゴミ袋を新しいゴミ袋と判定する。ドライブレコーダーで撮影された動画は 8pfs とフレームレートが低く、閾値を高く設定してしまうとトラッキングすることが難しかったため、今回のアルゴリズムでは閾値を 0.1 とした。

$$IoU = \frac{A \cap B}{A \cup B}$$

図 5: Intersection over Union

### 2.2.3 カウント判定アルゴリズム

清掃車後部に設置されたドライブレコーダーで撮影された動画は図6のような角度で撮影されているため、ゴミ袋は画面上方から下方へとフレーム間を移動していくように映っている。この特性を利用し、カウントは図6のように SSD で出力されたゴミ袋のボックスの中にある赤い点、前フレームでは青いライン越えておらずかつ次フレームでは越えていたときにカウントされる。下方から青いラインを超えてもカウントされない。また、トラッキングが成功している限り、1度カウントされたゴミ袋は条件判定により2度カウントされない。赤い点はボックスの幅の 1/2、高さの 6/7 に位置する点を表している。中心点をカウント判定とするとゴミがカウントされずそのまま画面外へ行ってしまい、カウントされないことが多かったため、カウント判定するゴミ袋の位置座標をこのような値に設定した。



図 6: (上): 実際に清掃車後部に取り付けられたドライブレコーダーで撮影された動画のフレーム画像. (下): ゴミ袋がフレーム間において上方から下方へ移動しトラッキング, カウントされている.

### 3. 実験

実際に藤沢市を走行する清掃車の後部に取り付けられたカメラで撮影された動画を用いて, 提案手法の精度を評価した.

#### 3.1 データセット

学習用のデータセットを作成するため, 藤沢市を走行する異なる 2 台の清掃車の後部に取り付けられたドライブレコーダーで可燃ゴミが収集される場面が撮影された動画を用いた. 今回可燃ゴミを学習対象としたのは, 可燃ゴミが回収されるゴミの量の 76.2%[?] を占め, 可燃ゴミを検出できるメリットが大きいと考えたためである. 動画は 8fps, 640 × 480 ピクセルで撮影されている. その動画をフレーム化し, その中からゴミ袋が映る画像のみを選別し, 画像内に映るゴミ袋にアノテーションを施した. 訓練データセットは 2016 年 10 月 28 日に撮影された動画から 1372 枚, 2017 年 5 月 11 日に撮影されたものから 259 枚を作成した. テストデータには 2017 年 5 月 11 日に撮影された動画から 539 枚作成した.

#### 3.2 SSD の学習

学習では可燃ゴミ袋を分類するように学習した. ベースとなる VGG-16 の重みには ILSVRC CLS-LOC で事前に学習されたものを使用した. 学習したのち, ゴミをどれだけ正確に識別できるかを示す Average Precision を PASCAL VOC の基準を用いてテストデータで評価したところ 0.53 と算出された.

#### 3.3 評価手法

評価は F 値を用いた. TP をゴミを正確にカウントできた数, FN をゴミを見逃してカウントできなかった数, FP をゴミでないものをゴミと判断してカウントした数として, 式 1 のようにそれぞれ Precision, Recall, F 値を求めた.

表 1: ゴミ袋カウントの混同行列

		実際の結果	
		True	False
アルゴリズムによる判定	True	TP	FP
	False	FN	TN

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F\text{-値} = \frac{2\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

### 3.4 実験結果

実験結果を表 2 に示す. Precision が 0.42 とあまり高くない一方で, Recall は 0.90 と高い値である. これはゴミ袋を見逃してカウントできないことが多い一方, ゴミ袋ではないものを誤ってゴミ袋としてカウントすることが少ないことを示している.

表 2: 指標と精度

指標	ゴミの個数	指標	精度
TP	184	Precision	0.42
FP	255	Recall	0.90
FN	21	F-measure	0.57

### 4. 考察

実験結果に基づく考察を以下に示す. ゴミ袋でないものを誤ってゴミ袋としてカウントする例として, 図 7 のように道路の白線のかすれや清掃員の帽子をゴミと誤認識してしまうことが多く見られた. この問題点は, 新しく帽子のクラスを作ることによって帽子を検出できるようにし, 帽子が通過してもカウントしないようにすることで解決できると考える.

また, ゴミ袋を見逃してしまいカウントできない原因として, 動画のフレームレートが 8FPS と低いことが考えられる. フレームレートが低いため, 清掃車にゴミ袋を投げ入れる速度が高いと IOU の値が低くなり, 閾値を超えることができずトラッキングできない事例が多く見られた. 図 8 が示すように, 画面に映るゴミ袋の大きさも関係があり, 小さいと IOU の値もまた小さくなりやすく, トラッキングができないことも多く見られた. また図 9 の左側のように, ゴミ袋が画面横方向に離れているとカウントできる一方で, 右側のように画面横方向に近接し直列的にゴミ袋があると, ゴミ袋が画面上方から下方へ移動するという収集作業の特性のため, 前フレームにあったゴミ袋が次のフレームで後ろにあるゴミ袋との IOU 値が高くなり同定されてしまい, カウントされないことが多く見られた. これらの問題点の解決法として, 動画のフレームレートを高める方法がある. フレームレートを高めることによって, フレーム間でのゴミの領域情報の類似性が向上し, IOU の値が高まりトラッキングをより正確に行うことができると考える.



図 7: (左): 掠れている白線をゴミ袋と誤認している. (右): 帽子を誤認識し, 誤ってカウントしてしまっている.



図 8: (左): ゴミ袋が大きく写っているためトラッキングが成功して正しくカウントできている。(右): 2つのゴミ袋が画面横方向に近いためトラッキングが失敗しカウントできていない。



図 9: (左): 2つのゴミ袋が画面横方向に離れているため正しくトラッキングできている。(右): 2つのゴミ袋が画面トラッキングができていない。

## 5. 結論と今後

本研究では、地域毎のゴミの量を自動的に推定するシステムを考案した、また、物体検出可能な CNN と清掃車後部に設置されたカメラで撮影された動画を用いて、ゴミ袋の個数を自動的にカウントする手法を実装した。実際に藤沢市を走行する清掃車後部に取り付けられたカメラで撮影された動画を用いて実験を行い、提案手法の精度を評価した。実験の結果、Precision が 0.42 とあまり高くない一方で、Recall は 0.90 と高い数値を示し、提案手法がゴミ袋ではないものをゴミ袋としてカウントすることが少ないことを明らかにした。また、提案手法の精度を高めるためには動画のフレームレートを高める必要があることがわかった。

今後の研究の発展として、SSD のアーキテクチャーを軽量化しゴミ袋を検出する速度を高めることがある。清掃車に用いられる予定の GPU が搭載された小型コンピューターで SSD によるゴミ袋の検出を試してみたところ、1 フレームあたり約 0.3 秒の処理時間を要した。8fps の動画を処理するとなると、動画 1 秒間あたり約 2.4 秒の処理時間が必要となる。本手法を実際

の清掃車に搭載しリアルタイムでゴミ袋の個数をカウントすることはできなくはないが、不便を生じさせる。そのため、SSD のアーキテクチャーの軽量化が必要となる。今回 SSD のベースネットワークに用いた VGG-16 は SSD の処理時間の 80% が費やされている。したがって、ベースネットワークに VGG-16 ではなく、より軽量なものを用いることによってリアルタイム化ができると考える。また、軽量化することによって、処理速度が高まる一方検出精度は低くなるが、ゴミ袋でないものを誤ってカウントすることが少なかったという今回の実験結果から、その影響は低いと考えられる。

## 謝辞

本研究の一部は NTT ソフトウェアイノベーションセンタに支援頂いた。

## 参考文献

- [Chen 17] Chen, Y., Yonezawa, T., and Nakazawa, J.: Towards Vehicle-Assisted Adaptive Wireless GeoMesh Network for Smarter Cities, in *VEHICULAR 2017: The Sixth International Conference on Advances in Vehicular Systems, Technologies and Applications* (2017)
- [Fukunaga 75] Fukunaga, K. and Hostetler, L. D.: The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition, *IEEE Transactions on Information Theory*, Vol. 21, No. 1, pp. 32–40 (1975)
- [Kawano 17] Kawano, M., Mikami, K., Yokoyama, S., Yonezawa, T., and Nakazawa, J.: Road Marking Blur Detection with Drive Recorder, pp. 4010–4015 (2017)
- [Liu 16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C.: Ssd: Single shot multibox detector, in *European conference on computer vision*, pp. 21–37 Springer (2016)
- [Redmon 16] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *arXiv preprint arXiv:1612.08242* (2016)
- [Simonyan 14] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *ArXiv e-prints* (2014)
- [筑波] 筑波大学: ゴミ減量 G1 グランプリ, <http://bigdata.kde.cs.tsukuba.ac.jp/g1gp/app/index.html/>
- [中澤 17] 中澤 仁, 陳 寅, 米澤 拓郎, 大越 匡, 徳田 英幸: ユニバーサルセンサネットワークと清掃車を活用した藤沢市のスマート化, *デジタルプラクティス*, Vol. 8, No. 3, pp. 244–252 (2017)