

# 深層学習に基づいた屋外大型駐車場の空きスペースの検出

三木 守<sup>\*1</sup> 姜 有宣<sup>\*2</sup>

## Empty Space Detection in Spread Parking Lots Using Deep Learning

Mamoru Miki<sup>\*1</sup> Yousun Kang<sup>\*2</sup>

**Abstract** Parking management is a hard work for many parking agents in outdoor parking lots. It is difficult to have overall information about vacant and occupied spot in outdoor parking area. An effective detection method of empty space would reduce physical labor and increase the efficiency of utilizing the parking spot information. In this paper, we implemented a system to detect empty space in a spread parking lot with a deep learning algorithm. The detection method was accomplished with a convolutional neural net (CNN) based on the You Only Look Once (YOLO) model architectures. Using a DJI camera drone, we make an image dataset to perform train and test for YOLO, which is a powerful object detection model. Results of our experiments using our dataset showed that the proposed system is effective for implementing an empty space detection in spread parking lots.

### 1. はじめに

大型ショッピングモールや遊園地に設置されている広い屋外駐車場の管理には多大な労働力とコストを要する。警察庁の調査データによると駐車場警備員の平均年齢はア高くなっており、年齢層で見ると 30 代未満の若年層はわずか 10.1%である[1]。戸外での警備勤務は身体的、精神的に負担が大きいため、効率的な管理方法の提案が求められている。

一方、人的補助なしに、機械的に多くのデータからディープニューラルネットワーク (DNN: Deep Neural Network) を用いて特定の物体の特徴を抽出するディープ ラーニング (深層学習) を用いた研究が様々な分野、特に物体認識や言語処理の分野で展開され成果を上げている。深層学習には十分なデータが必要であるが、データ採取方法として自律移動型ロボット (ドローン) 利用が有効である。ドローンとは無線による遠隔操縦あるいは搭載コンピュータにプログラムされたパターンに従い自律飛行をする無人機を指す。元来軍事用に開発され、RQ-1 プレデターや無人航空機 (UAV) が偵察や救援物資の運搬に利用されていたことは有名である。現在では民間でも利用されるようになった。例えば災害時の報道や救助、生態調査や発電所の点検などで活躍している。これは深層学習技術に応用したものである。そのほかにも AED を搭載したドローンは、緊急を要する事故現場救助要請対応や、赤外線カメラ搭載ドローンによる夜間監視など、医療分野およびセキュリティ分野でも活躍し様々な分野で注目されている。ドローン飛行は航空法、小型無人機等飛行禁止法、道路交通法 (第 77 条)、民法 (第 207 条)、電波法の縛りがある。市街地を飛行させるには無人航空機飛行許可証取得の必要がある。

しかし許可証取得は容易である。さらに昨今は低価格化が進み、個人での入手も容易になった。

本研究では、大型駐車場管理における多大な労働負担を軽減する目的で駐車場と道路上の駐車スペースを自動的に検出するシステム開発を意図して、ドローンによるデータ採取と深層学習モデル作成の最適化を行った。具体的には東京工芸大学厚木キャンパス内の駐車場を観測対象とし、ドローンを用いて画像データを収集する。ついで画像データをアノテーションツールにてラベリングし、データセットを構築する。さらに、深層学習法の中から YOLO を基本とする物体検出法を用いて駐車場空きスペースを学習させる学習モデルを作成した後、改訂 YOLO で物体検出を行う。

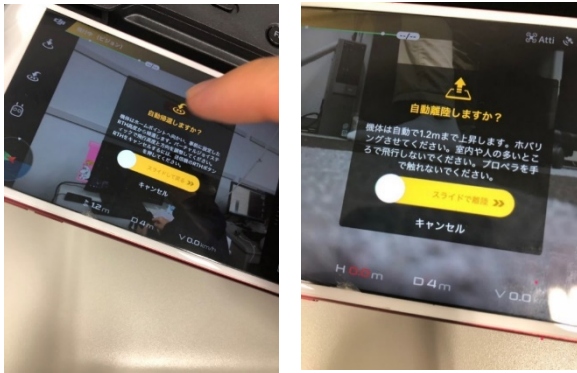
### 2. ドローンの操作と調整

ドローンは DJI 製 MavicPro2 を用いた。操作は、DJI のアプリを使用し、飛行前にキャリブレーションを行った。GPS 情報をあらかじめ取得して、着陸位置に自動帰還・離陸できるように調整した。図 1 に実際に自動帰還・自動離陸を試した時の画面を示した。DJI 製のドローンには障害物回避センサーが内蔵されており、進路に障害物がいた際、それを感知して通知し、さらに機体によってはその場で静止する機能が備わっている。その他にも被写体を自動で認識し、その物体や移動速度に合わせて追尾するアクティブトラックという機能がある。DJI 製のドローンにはそれぞれ障害物センサーが内蔵されており障害物に接触する 1m 手前で前進を停止する。Spark の場合は前方にしか障害物回避センサーを搭載していないため、この機能を使う際は注意が必要になる。

<sup>\*1</sup> 東京工芸大学工学部コンピュータ応用学科 2019 年度卒業生

<sup>\*2</sup> 東京工芸大学工学部コンピュータ応用学科教授

2020 年 3 月 25 日 受理



(a)

(b)

図1 DJIの自動帰還 (a) 自動離陸 (b) 実行画面

### 3. 学習モデルの作成

#### 3.1 深層学習とYOLO学習機

本研究では物体検出のためYOLO (You Only Look Once) を使い学習モデルを作成した。ディープラーニングには様々なアルゴリズムがあり、それぞれ得意分野はアルゴリズムに依存する。物体認識に関わってくるものにはCNN (Convolutional Neural Network, 畳み込みニューラルネットワーク) がある[3]。これは、局所的な情報の抽象化及び位置普遍性をもたせた順伝播型ニューラルネットワークを利用したアルゴリズムで画像や動画認識に広く使われている。近年、CNNを用いた様々な物体検出手法が装着されている。CNNを利用すると入力画像全体から特徴量を抽出し物体検出ができる。一方、R-CNN(Regions with CNN)手法は物体に写っている領域の候補を抽出し、リサイズした領域に対してCNNで特徴量を推論し何が写っているのか分類する[4]。R-CNNは領域候補を抽出した後分類するが、YOLOは入力画像をグリッドに分割してから各グリッドのクラス分類と領域候補の抽出を同時に行う。そのため、R-CNNより精度は低い早い処理速度を有する。YOLOは今までYOLO, YOLOv2, YOLO9000, YOLOv3など色々なバージョンが発表されている。

最初のバージョンであるYOLOは22層の畳み込み(convolution)層と2層の全結合層(分類や物体領域の座標修正)で学習を行う。通常のCNNでは、最終層に全結合層を導入し、ソフトマックス(softmax)関数など処理して、画像の分類を行うが、FCN(Fully Convolutional Networks)では、最初から最後まですべてが畳み込み層で構築され、特徴マップの明確な位置情報を保持したまま最終層まで伝播される。YOLOv2ではこれを導入したことで大幅な精度の向上を実現した。また、YOLOv3はYOLOv2の19層のモデルに対して53層のニューラルネットワークを使用している。YOLOv2と比較すると若干速度が落ちたが、精度が向上した。従ってYOLOv3を実装し学習モデルを作成した。

YOLOv3は学習モデルを作成する時、画像ごとに異なるスケールに対応するため異なるサイズの検出レイヤーを利用する。それを3D tensorと呼び各レイヤーから検出したい物体を予測する。そのためYOLOv3では事前情報として、データセットにどのくらいのサイズの画像があるかという情報を使っている。その情報は正解ラベルのバウンディングボックス幅をk-meansによって9つのクラスタリングを実行して得ている。この9つを3つずつ各サイズの予測テンソルに割り当てる。対応する三つのスケールはそれぞれ $52 \times 52$ ,  $26 \times 26$ ,  $13 \times 13$ である。

実際ディープラーニングのライブラリであるTensorflowやChainer, Kerasなどを実装しYOLOv3学習モデルを作成した。構築した学習モデルを用いて検出結果を見ると画像に検出した物体を四角で囲み、その左上に検出率を表示してくれる。例えば0.8なら80%の確率でcarだと認識していることになる。



図2 YOLOv3の物体検出例

#### 3.2 データセットの構築

YOLOは自前のデータを学習させることも可能である。事前にデータを学習させるには学習させるための画像データと画像にある物体の位置をラベリングしなければならない。そこでドローンを使い収集したデータをVoTT (Visual Object Tagging Tool) [5]というアノテーションでラベリングを行った。

VoTTとは、Microsoftが提供する動画・画像に対してアノテーションをおこなう無償のアノテーションツールである。Windows, Mac, Linuxを問わずクロスプラットフォームで利用でき様々なアルゴリズムで使用できる形式で出力できる機能が備わっている。アノテーションとは、撮影したデータにラベリングすることである。図3のようにGUI操作で簡単にアノテーションを行える。クラスを「space」と設定しアノテーションを行い、YOLOv3で利用できるPascal VOCという形式で出力した。アノテーション作業を行った出力データはフォルダごとに保存した。図4はデータセットの構築のため格納された代表的入力画像である。

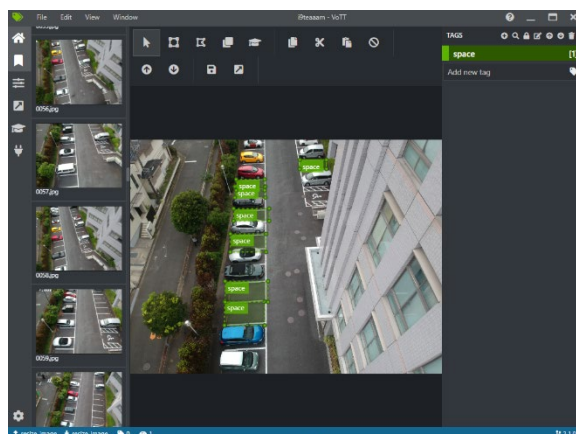


図3 アノテーション VoTT 実行画面例

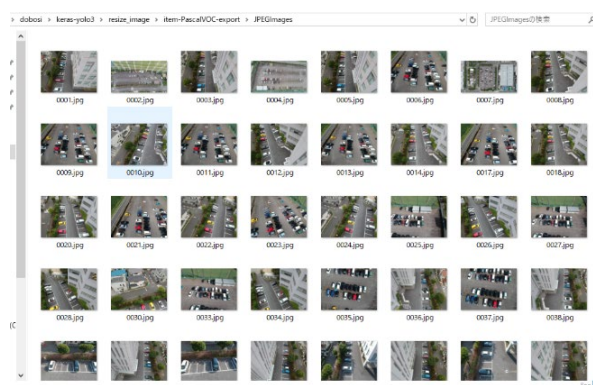


図4 アノテーション処理入力画像例

## 4. 実験結果

### 4.1 実験環境

実際に学習させた流れは次の通りである。

- ① ドローンを使い空撮
- ② 画像のリサイズ
- ③ 駐車スペースを「space」としてアノテーション
- ④ 学習モデルの作成
- ⑤ 学習モデルの評価

その後、収集したデータサイズが大きいため 128×128 にリサイズを行った。VoTT を使い駐車スペースを「space」と検出できるようにして学習モデルの作成、評価を行った。

撮影する際は、検出に差が出ないようにするため高さを 20m に統一し撮影した。ドローン二機を使い学内の駐車スペースを学習用に 500 枚、学習モデルの評価（テスト）を行うために 100 枚の計 600 枚収集した。ドローンで撮影したデータは大きいため画像のリサイズを行った。学習させる際、画像サイズは 32 の倍数となるようにした。サイズが大きすぎると学習に時間がかかってしまうからである。学習で使われるサイズは一般に 128 ピクセルとしたので撮影した画像もすべて 128 ピクセルにリサイズを行った。

最終的には学習モデルの作成のため 500 枚の画像を使い残りの 100 枚はテストに使用した。学習は 500 枚のデータ一周が 1 Epoch にして、45～50 秒ほど時間がかかっている。これを 100 Epoch 学習させた。

### 4.2 学習モデルの評価

作成した学習モデルを使用し、テストデータ 100 枚の検出を行った結果、正常に検出された結果と誤った場所を正解と検出してしまった 2 パターン見られた。誤って正解と検出されてしまったデータの調査を行った。

作成した学習モデルでテストデータ 100 枚の検出を行った結果のなかで、正常に検出された「space」は 568 個検出された。図 5 に示すように「space」として認識ができていない。動画で検出を行った結果、実際に撮影した動画の速度が 20 fps だったが学習モデルを使い検出している動画は 4 fps の速度が出た。

図 5 の結果画像は動画で検出した結果である。検出した結果画像から誤認識の結果が得られた。図 5 左上の結果画像のように奥が「space」として認識されないものがあった。また、図 5 右上の結果画像のように車の止まっている場所を「space」として認識してしまうものがあった。図 5 の下の結果画像は誤認識がない結果画像である。特に使っている学習機 YOLO は小さい物の検出能が低いため、画像データから空きスペースが細かくなってしまう場所の検出は誤認識する場合が多かった。

学習モデルの評価のため混合行列[6]を使用した。まずモデルの評価を行うためテスト画像 100 枚の結果を TP (True Positive), FP (False Positive), FN (False Negative) の三つに分類した。この実験で TP は空きスペースを正しく「space」と認識している場合を示し、FP は車の止まっている場所を「space」と認識している場合、また FN は空きスペースなのに「space」として認識できない場合である。テストデータから分類した結果、TP が 568, FP が 16, FN が 140 という結果になった。混合行列を用いて評価を行うと正解率、適合率、再現率、F 値の 4 種類の数値が計算できた。正解率は 0.78, 適合率は 0.97, 再現率は 0.80 だった。これを用いて F 値は以下の式に従い 0.88 と算出できた。

$$F \text{ 値} = \frac{2\text{Recall} + \text{Precision}}{\text{Recall} + \text{Precision}} = 0.88$$

## 5. 結論

最適化 YOLO, YOLOv3 を実装した深層学習用学習機を用いて DJI 制 Spark, MavicPro2 種のドローンにて撮影した画像データから広い屋外駐車場と道路上の駐車スペースの物体検出を行った。ドローンから撮影した結果画像データ 600 枚のうち、学習用に 500 枚を、テスト用に 100 枚を使用した。





図5 空きスペースの代表的検出画像

結果画像データは VoTT を用いてアノテーションを行い、学習モデルの作成およびテストデータの評価を行った。学習機の性能を定量化するために混合行列評価を適用して F 値を算出した。適合率、再現率はそれぞれ 0.97, 0.80 と高い値が得られた。一般的に混合行列評価大法で両者とも高い値となることは少ない。撮影対象を大学キャンパス内に限定してデータ収集をしたことが偏った値に反映したと考えられる。

本研究で構築した学習モデルを駐車スペース管理に応用展開するには、精度向上が必要である。そのためには第一に画像データ収集対象地域を市街地や郊外の大型ショッピングモールや遊園地などの様々な広大な屋外駐車場に拡大し、学習およびテストを行う、第二にホールドアウト法やクロスバリデーションなどの多様な評価方法を適用した精度計算を施すことが必要と考えられる。実用可能なシステム提供のためには、いくつかの改善が必要であるが、ドローンと深層学習の組み合わせが駐車場管理システムに有効であることを明示することができた。

## 参考文献

- 1) 平成 29 年における警備業の概況 - 警察庁  
<https://www.npa.go.jp/safetylife/seianki/statistics/29keibi.pdf>
- 2) コトバンク  
<https://kotobank.jp/word/ドローン-107015>
- 3) R. Collobert, J. Weston, “A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning”, Proceedings of the 25th International Conference on Machine Learning, ICML '08, New York, pp. 160–167, 2008
- 4) A. Tsantekidis, N. Passalis, A. Tefas, J. Kannianen, M. Gabbouj, A. Iosifidis, “Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks”, IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, Greece, 2017
- 5) VoTT: Visual Object Tagging Tool,  
<https://github.com/microsoft/VoTT>

## 謝辞

本研究は東京工芸大学「学生による工・芸共同研究」の助成を受けたものです。