

# R-CNNを用いた LAMAT分布把握のための基礎的研究

金野 貴紘<sup>1</sup>・荒井 勇輝<sup>2</sup>・屋井 鉄雄<sup>3</sup>

<sup>1</sup>正会員 東京地下鉄株式会社

(〒110-8614 東京都台東区東上野三丁目19番6号)

E-mail: takahiro.konno0719@gmail.com

<sup>2</sup>学生会員 東京工業大学 環境・社会理工学院 土木・環境工学系 都市・環境学コース

(〒226-8502 横浜市緑区長津田町4259)

E-mail: arai.y.ar@m.titech.ac.jp

<sup>3</sup>正会員 東京工業大学 環境・社会理工学院 土木・環境工学系 都市・環境学コース

(〒226-8502 横浜市緑区長津田町4259)

E-mail: tyai@enveng.titech.ac.jp

東南アジアには特有の交通モードであるLAMATが存在する。これらは現地住民の足として根付いているが既定の路線や停留所を持たず、また現地政府による登録や管理も不十分であることからその実態については不明瞭な点が多く存在する。そこで本研究ではアジア途上国における交通計画策定のための一助としてLAMATの都市内分布を明らかにすべく、一般物体認識・画像判別の手法であるR-CNN (Region with CNN)の方法論を基にした車体検出ツールの構築を行い、開発したツールの精度検証及び、実務に用いる上でのロバスト性の検討を行う事で、LAMATを検出・判別する上で寄与する要因について把握した。

**Key Words :** LAMAT, Developing countries, Image recognition, Vehicle detection, Deep learning

## 1. 序論

近年、東南アジア各国では急激な経済成長などを背景にモータリゼーションが進行し、それに伴い交通渋滞や交通事故等の交通問題が深刻化している。これら諸問題の解決として公共交通の導入が検討されているが、その導入に際しては現地の文化や雇用の保全・保存の必要性の観点より、既存の交通モードと調和した計画づくりが求められている。

東南アジア特有の交通状況として LAMAT<sup>1)</sup> (Locally Adapted, Modified, and Advanced Transport) の存在が挙げられる。LAMAT の定義は、自家用車と典型的な大量旅客輸送公共交通の間に位置する、“自転車タクシーから定員 25 名のミニバスまでわたる交通モード”である。LAMAT は細街路でも走行可能な点や安価な運賃、乗降の自由度の高さ等から、住民の足として深く根付いている。しかしこれらは既定の路線や停留所を持たないこともある上

に、現地政府による登録や管理も不十分であり、その実態については不明瞭な点が数多く存在する。

交通計画策定にてプロジェクト評価や需要予測を行うためには交通量を含む各種統計データが必要であり、またその把握には大規模調査や、トラフィックカウンターによる手法が存在するが、途上国においては各種統計データやインフラが未整備であることから、それらの手法を検討することは容易ではない。

そこで本研究では、アジア途上国における交通計画策定のための一助として、現状の LAMAT の運行台数・分布等で示される市場特性の把握を最終的な目標とし、衛星写真等を想定した画像データから複数の車両を検出・判別するツールを開発し、その実用可能性について検討することを目的とした。

そのため、近年画像認識分野において注目されているディープラーニングによる一般物体認識の手法の適用を試みた。具体的には、画像上の物体を検出し、検出され

た候補に対し CNN (ディープラーニングの一手法である畳み込みニューラルネットワーク) による判別を行う R-CNN<sup>2)</sup> (Region with CNN) の方法を基にした車体検出システム (以下ツール) の構築を試みる。

さらに、フィリピン国の LAMAT に属する tricycle を対象にして、学習用の画像データを現地観測調査より独自に収集し、開発したツールの精度検証及び、実務に用いる上でのロバスト性の検討を行うことで、LAMAT を検出・判別する上で寄与する要因等について把握する。

## 2. 既往研究と本研究の位置づけ







LAMAT に関する研究は LAMAT 利用者の意識や行動に着目した研究 (需要者特性) や、シミュレーション等を用いたオペレーションに関する研究 (供給者特性) は比較的数量多くあるものの、車両特性、および市場特性といった LAMAT そのものの情報について調査している研究は少ない。そのなかでも市場特性については Phun ら<sup>1)</sup>はカンボジアにおける LAMAT のコスト構造の分析などの供給者特性や LAMAT メーカー、エンジンサイズなどの車両特性の整理を行っている。また、一般組合、ルート組合といった LAMAT の組合に対するインタビューからその概要を示している。WICAKSONO ら<sup>3)</sup>はカンボジア、インドネシア、日本、フィリピン、タイ、ベトナムの 6 か国の都市人口増加に伴う交通需要を満足する政策について法的条件とインターモーダルに関する問題に焦点を当てて考察している。

しかしながら各研究におけるデータ入手方法はその多くが現地調査によるものであるが、調査の効率化や他地域でも応用可能な調査手法に言及したものはないため、LAMAT の都市内存在分布を明らかにする手法を検討する意義は大きいと言える。

またディープラーニングによる一般物体認識において車両を対象物と扱っている研究数も数多く存在している。Sebastien<sup>4)</sup>らはディープラーニングを用い航空写真からの車種判別として、car, van, truck 等車種の判別を行い、将来的な監視システムへの適用について提案している。Matthias<sup>5)</sup>らは路上の監視カメラの映像を用い、昼・夜間、雨天、雪等悪条件の映像からの車体検知を行い運転支援システムへの適用について言及している。

しかし判別対象として、車両の色や形が多様な LAMAT を取り上げているものはなく、また交通計画策定において有益な情報である車両の分布や台数推計を目

表-1 CNN モデルデータセット (GoogleEarth 画像)

名称	bike	bus	car	jeepney	tricycle	truck
画像						
画像枚数	学習用 80 テスト用 20	80 20	80 20	80 20	80 20	80 20

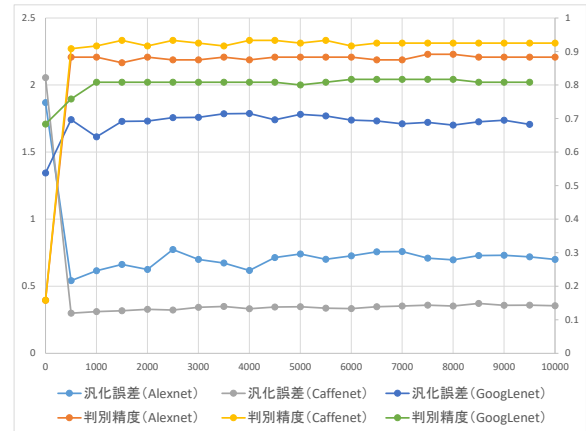


図-1 モデル別の GoogleEarth 画像判別結果

的としている研究も見当たらない。そこで本研究ではツールの開発過程において LAMAT を含む車両判別精度に寄与する要因も含めた分析を行い、ツールとしての適用性の検証を行った。

## 3. CNN による画像判別の手順

CNN による画像判別は判別対象の特徴量の抽出を多層化したニューラルネットワークにより行われる。各画像の  $n$  ピクセルの画素値を入力情報とし (1) 式に定義される、関数  $Y$  に変換し複数個のニューロンより特徴量を抽出する。

$$Y = \sum_{i=1}^n W_i X_i - \theta \quad (1)$$

結果の出力では、活性化関数(ReLU)にて各種のニューロンの持つ関数  $Y$  を結合結果として確率の値で出力し、最も確率が高いものを判別結果として返す。判別結果と理想的な出力との乖離を (2) 式に示す交差エントロピー誤差関数により分類誤差  $C$  を求め、誤差逆伝搬法により  $C$  が最小となるように  $W_i$ ,  $\theta$  を調整する。この調整が一回行われることを学習が一回行われたと定義される。

$$C = - \sum_{j=1}^n d_j \log p_j \quad (2)$$

また CNN はネットワーク中間層において畳み込み、プーリングといった処理を行うことで特徴量を抽出する領域の範囲を増大させ、より抽象的な情報にて物体の特徴を判断する性質を持つ。

本研究ではディープラーニング用ライブラリ Caffe を

用い、車種を判別する CNN モデルの構築を行った。学習 Imagenet の画像群約 120 万枚にて事前学習済みの CNN モデルを用い、Fine-tuning（転移学習）を行った。Fine-tuning とは汎用的な教師データを基に学習済モデルと、そのネットワークを利用して再度学習させる手法であり、一からモデル構築を行うよりも、学習を高速で行うことが出来、かつ教師データの量が十分に無くても性能の良いモデルの学習が可能という特徴がある。

#### 4. CNN による判別モデルの構築

##### (1) GoogleEarth 画像を用いた LAMAT 判別モデル

CNN による画像判別モデルの構築には学習データとして、判別したい対象物につき数百～数千枚単位での画像データが必要となる。本研究のツールでは、Google 画像等のオープンソースデータ利用を実用上想定しているため、まず GoogleEarth を用い LAMAT の画像収集を行い、CNN モデルによる画像判別を試行した。判別対象は、ツールの利用目的より、実在の地域内を走行している車両群の判別が求められることから、フィリピン国を対象地域として、観光ガイドや GoogleStreetView 等より走行が確認できた bicycle, motorcycle, car, jeepney, taxi, tricycle, truck の 7 種を今回の判別対象と設定した。内 tricycle, jeepney はフィリピン特有の LAMAT に属する交通モードである。tricycle は二輪車とサイドカーより構成されるタクシーであり、通常のタクシー同様に道路上で拾い希望の場所まで移動することが出来る。jeepney はフィリピンの人々が日常的に使う乗り合いバスであり決められた路線を有している。交通量・分布等の情報を得るには jeepney はその路線やターミナルを把握すればよいが、tricycle は個人による運営が主体であり決められた走行ルートを持たないため、本研究においてツールを構築する際の主対象は tricycle と設定した。なお GoogleEarth 上より画像収集を行う際に car と taxi および bike と bicycle については Google earth 上では見分けが困難だったため、それぞれ car と bike に統合し計 6 種の車両にて画像判別を試行した。データセットとして各車両 400 枚づつ画像を用意し、内 8 割を学習用、2 割をテスト用とした (表-1)

学習済ネットワークには、オープンソースとして公開されている Alexnet, CaffeNet, GoogLeNet の 3 つの CNN モデルに対し同様の画像データセットを与え、学習は最大

表-2 現地調査概要

調査日時	2016年11月28日～12月1日
観測者	4名(東工大3名, フィリピン大学1名)
調査方法	高所地, および地上部からの定点観測
観測対象	LAMATを含む各種現地交通モード
調査地	マカティアベニュー(フィリピン, マカティ市)
調査項目	①衛星画像に相当する高所地からの撮影データ ②高所地と一致した撮影タイミングによる地上撮影データ ③天候等, 自然環境条件が異なるデータ

表-3 CNN モデルデータセット (日向画像)




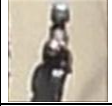





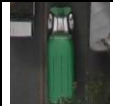
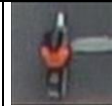



車種名	bicycle	car	jeepney	motorcycle
画像				
学習用	75	75	75	75
テスト用	25	25	25	25
車種名	taxi	tricycle	truck	
画像				
学習用	75	75	75	
テスト用	25	25	25	

表-4 CNN モデルデータセット (日陰画像)

車種名	bicycle	car	jeepney	motorcycle
画像				
学習用	150	150	150	150
テスト用	50	50	50	50
車種名	taxi	tricycle	truck	
画像				
学習用	150	150	150	
テスト用	50	50	50	

習回数 10000 回, 基本学習率 0.001 にて実行しその汎化誤差と判別精度について比較したところ図-1 のようになった。CaffeNet という畳み込み層 5 つ, プーリング層 3 つ, 全層結合層 3 つからなるニューラルネットワークのアーキテクチャが最も汎化誤差の収束する値が小さく, 判別精度も 92.5% と最も高かったことから, CaffeNet の構造を本ツールの CNN 構造として採択した。

##### (2) 現地調査による画像データ収集

GoogleEarth による画像収集は, 解像度の粗さにより正解ラベルの指定が正確に行えないことや, 実務において重要な日影や天候, 地理特性といった自然環境条件を含むロバスト性の検討が困難であることから, 複数の環境条件化における LAMAT の走行風景画像を高解像度の画像情報として取得すべく, 現地観測調査を行った (表-2)。

調査地はフィリピン国マカティ市における主要道路の一つであるマカティアベニューを選定し, ビデオカメ

表-5 判別結果の混同行列

		推計後のカテゴリ							計
		bicycle	car	jeepney	motorcycle	taxi	tricycle	truck	
本来の カテゴリ	bicycle	48 (96%)	0	0	1 (2%)	0	1 (2%)	0	50
	car	0	46 (92%)	0	0	3 (6%)	0	1 (2%)	50
	jeepney	0	0	43 (86%)	0	0	0	7 (14%)	50
	motorcycle	10 (20%)	0	0	40 (80%)	0	0	0	50
	taxi	0	0	0	0	50 (100%)	0	0	50
	tricycle	2 (4%)	0	0	0	0	48 (96%)	0	50
	truck	0	0	0	0	1 (2%)	0	49 (98%)	50

ラによる定点観測を行い、得られた動画よりキャプチャーを行うことで画像データを収集した。

データ収集には撮影環境条件を考慮し観測時に路面が日射により照らされている画像（以下日向画像）および日陰になっている画像（以下日陰画像）に分けて画像を作成した。日向画像は各車両 100 枚ずつ、日陰画像は各車両 200 枚ずつ取得した（表-3、表-4）。

(3) 現地撮影画像を用いた LAMAT 判別モデル

モデル構造には Caffenet を用い、学習は最大学習回数 10000 回、基本学習率 0.001 にて実行し、学習データ、精度検証を行うテスト用画像ともに日陰画像のみを用いた。その理由は、LAMAT を含む車両群を CNN モデルで判別可能であるかを検証するには、影等の影響が少なく車体の外形が読み取りやすい画像を用いる必要があるためである。構築したモデルの各車種の判別精度は 80~90%以上であり（表-5），また誤判別している例も二輪車同士での誤判別等、人間の目視での誤判別に似通っていることが分かった。すなわち LAMAT という他の車両と比べ特徴的な外形を有する車両を、新たなカテゴリとして CNN モデルで学習させれば、他の車両と同程度の精度にて判別を行えることが明らかとなった。

(3) 環境条件変更に対するロバスト性の検証分析

前節では、Caffe を用い LAMAT 判別モデルの構築を行いその精度について評価したが、ツールとして活用し実風景から車体を判別する際には、撮影環境を含む各種条件等によりその物体を構成する画像情報は大きく変動することが予想される。そこで日影による影響と、判別精度のばらつきについて分析を行い、モデルのロバスト性について検証した。

日影による影響を評価すべく、日向画像のみを学習デ

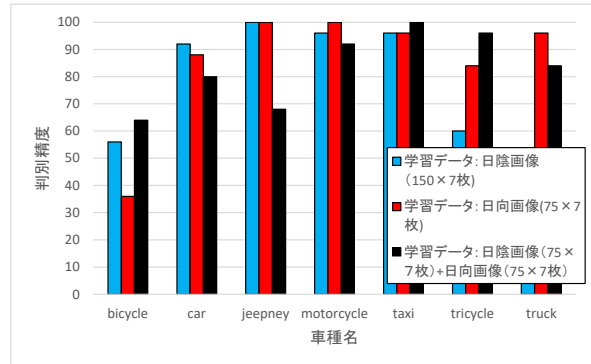


図-2 日向画像に対する各モデルの判別精度

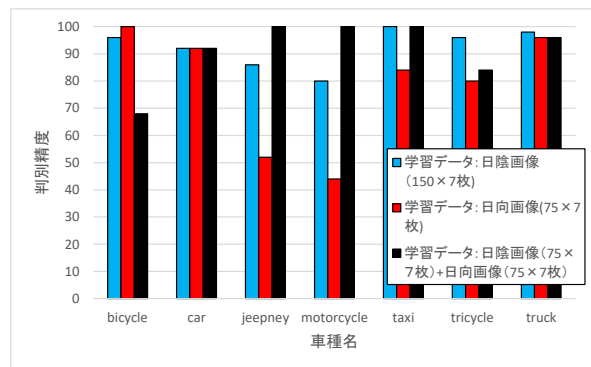


図-3 日陰画像に対する各モデルの判別精度

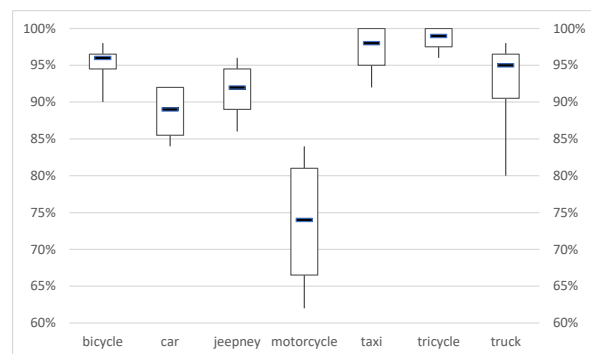


図-4 四分分割差検定によるモデルの判別精度検証結果

ータとしたモデル、及び日陰画像のみを学習データとしたモデル、日向と日陰画像を 1:1 の割合で混合させて学習させたモデルの 3 つのモデルを構築し、それぞれのモ

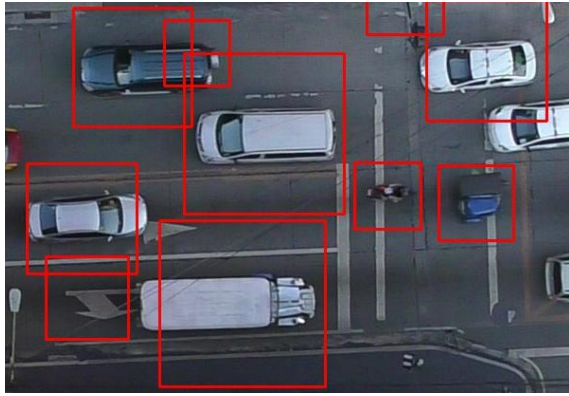


図-5 カスケード分類器による検出結果（現地撮影画像）



図-6 カスケード分類器による検出結果（GoogleEarth 画像）

デルにおいて日向画像，日陰画像に対する判別精度を比較した（図-2，図-3）。

その結果，双方の画像に対し同一のモデルを用いているにもかかわらず判別精度に差が生じていることから，日影による影響が判別精度に寄与することが把握できた。

相対的に低い結果が得られた原因として，車体に対する影の割合が大きいことで車体の正確な外形を把握できていない可能性が考えられるため，観測時における影の向きや大きさについては留意する必要がある。また日向と日陰の画像を混合した場合でも，双方の特徴量を得ることにより必ずしもロバスト性の高い判別精度が得られるとは限らないことが分かった。つまり実務においてツールを用いた検出を行う際には，あらかじめ日影条件を検討しておく必要があると言える。

次にモデルの判別精度のばらつきを評価すべく，日陰画像 200 枚を用いて四分分割交差検定によりモデルの判別精度の検証を行ったところ taxi, tricycle は常に 90% を維持している反面，motorcycle は 62%～84% と最も判別精度のばらつきが大きい結果となった（図-4）。これは今回用いた画像データでは taxi, tricycle はほぼすべてのデータにおいて車体の色が統一されていたが，motorcycle は車体及び搭乗者の服装等，色彩の情報がデータにより大きく異なっていたことが原因と考えられる。したがって既存の画像に対し色変更等の処理を行い，多様な色彩に対応できる学習データを追加することで精度の向上及びばらつきの収束が期待できると考える。

### 5. カスケード分類器による車体検出

CNN による画像判別は，一枚の画像につき一台の画像が投影されている物体に対してのみ行うことが出来るが，実際にツールを用いた分布把握を行う際に，画像内に複

表-6 カスケード分類器の学習データセット

正例(車体)画像			負例(非車体)画像		
1400枚(200×7種)			3000枚		

表-7 分類器の車体検出精度評価表

	GoogleEarth画像	現地撮影画像
車体一台を構成する平均画素数	1026	16800
検出窓の車体面積への被覆割合(平均)	91.75%	50.61%
負例(非車体)検出率	0.00%	32.25%
車両検出率	46.83%	97.33%

数台の車両が撮影されていると考えられ，その中から車体を検出する工程が必要となる。今回は車体検出の方法として，Selective Search 等との比較を行った上で，CNN 同様に機械学習のプロセスを持つカスケード分類器を適用することにした。カスケード分類器とは Viola ら<sup>6)</sup>が提案した AdaBoost ベースの分類器の応用であり，大量の正例画像と負例画像による学習により対象物を検出する手法である。OpenCV2.1 の traincascade を用い，画像の Haar-like 特徴量より学習を行った。

学習に用いた正例画像は CNN モデルに用いた日陰画像である各車両 200 枚ずつ計 1400 枚の車体画像を用い，負例画像は現地撮影画像より車体が投影されていない，道路表面，道路施設，歩行者，ビル壁面等の画像を計 3000 枚用いた(表-6)。日陰画像を用いた理由は影等の影響が少なく検出が容易であると考え，基礎的な検出を行う上で適していると考えたためである。

学習した分類器に複数台の車両が撮影されている GoogleEarth 画像 10 枚，現地撮影画像 50 枚を与え，その検出精度の比較評価を行った。検出精度の算出には，図

-5 に示すように車体の一部のみを検出する場合（以下欠損検出）や一台の車体に対し複数の検出窓による検出（以下重複検出）が行われるケースが発生することを考慮し、式(3)~(5)に定義した検出窓の車体面積への被覆割合、非車体検出率、車両検出率を用いることで評価を行った。

$$\text{検出窓の車体面積への被覆割合} = \frac{\text{検出窓内車体面積}}{\text{車体面積}} \quad (3)$$

$$\text{非車体検出率} = \frac{\text{非車体検出数}}{\text{総検出窓数}} \quad (4)$$

$$\text{車両検出率} = 1 - \frac{\text{未検出車両数}}{\text{画像内車両数}} \quad (5)$$

検出窓の車体面積への被覆割合とは、欠損検出が行われず車体の外形を正確に収めた状態で検出が行われているかを評価する指標である。目視にて観測した、検出窓内車体面積を対象の車体面積で除した際の値を用いており、値の算出に用いたサンプルは車体を検出した事例のみを扱い、非車体を検出している検出窓は除外している。非車体検出率とは、画像内に存在する非車体検出数を総検出窓数で除した値を用いた。非車体検出数とは全く車体に被覆していない検出窓の数と定義した。車両検出率は、まず画像内の未検出車体（検出窓による被覆が確認されなかった車体）数を求め、画像内車両数にて除して未検出車両数を算出したのち、その余事象として車両検出率を算出した。これは重複検出が行われた場合に、画像内の車両検出率を過大評価してしまうことを防ぐためである。

GoogleEarth 画像に対する検出結果では、検出窓の車体面積への被覆割合が 91.75%であり欠損検出が行われにくい傾向が把握できた。また非車体の検出の発生は見受けられなかったが、未検出の車両も約半数ほど生じる結果となった。現地撮影画像に対する検出結果では、車体検出率は 97.33%だが、非車体も多く検出しており、また欠損検出が行われる場合も多いことが確認できた(表-7)。

分類器に用いた学習画像は現地撮影画像を用いたが GoogleEarth 画像のように車体を構成する画像の解像度に乖離が存在している画像に対しても、画像内に存在している車両の内、半数ほどの車両の検出が可能であることが明らかになった。また検出窓の車体面積への被覆割合、非車体検出率、車両検出率には対象画像の解像度においてトレードオフの関係があることが示唆された。そのため現地撮影画像の解像度を段階的に下げていき、最

表-8 変更を行った各種パラメータ

	変更前	変更後
scaleFactor	1.1	1.04
minNeighbors	2	18
minSize	10 × 10	50 × 50
maxSize	—	200 × 200



図-7 現地撮影画像の車体検出結果(パラメータ変更前)



図-8 現地撮影画像の車体検出結果(パラメータ変更後)

も検出に適した解像度の条件を検討することで検出精度の向上が期待できると考えられる。

また現地撮影画像に対して、カスケード分類機において変更可能なパラメータである、scaleFactor, minNeighbors, minSize, maxSize を変更し、検出の変化を確認した(表-8)。scaleFactor とは対象画像上において物体候補領域を検出する上での探索窓の大きさを示す値であり minNeighbors は物体と認識する上でどれだけの検出窓により検出されたかを示す閾値である。また minSize, maxSize は画像内において物体が取り得る最小・最大サイズを示しており、これより外れる物体は検出から除外される。図-7に対し図-8は、目視により判別より非車体検出数が減少していることが確認できた。一方検出窓の車体面積への被覆割合については必ずしも向上するとは限らないという結果が得られた。パラメータの変更により検出の傾向が変化することが確認できたため、検出精度の向上に向けた取り組みとして、各種パラメータを段階的に変更し、最も車体検出において適したパラメータを検討することが有意義であると考えられる。

## 6. R-CNN による LAMAT 検出モデルの構築

### (1) 負例画像を用いた、CNN モデルの構築

前節において、カスケード分類器による車体の検出において負例や欠損検出といった、CNN の学習時に用いた車体画像とは異なる物体が候補として検出されることが

表-9 負例カテゴリを追加した CNN モデルの判別結果

		推計後のカテゴリ								計
		bicycle	car	jeepney	motorcycle	taxi	tricycle	truck	neg	
本来の カテゴリ	bicycle	47(94%)	0	0	3(6%)	0	0	0	0	50
	car	0	46(92%)	0	0	3(6%)	0	1(2%)	0	50
	jeepney	0	0	45(90%)	0	0	0	5(10%)	0	50
	motorcycle	7(14%)	0	0	43(86%)	0	0	0	0	50
	taxi	0	0	0	0	50(100%)	0	0	0	50
	tricycle	1(2%)	0	0	0	0	49(98%)	0	0	50
	truck	0	0	0	0	1(2%)	0	49(98%)	0	50
	neg	0	0	0	0	0	0	0	50(100%)	50

表-10 カスケード分類器の検出部位に対する CNN モデルの判別結果

		推計後のカテゴリ								計
		bicycle	car	jeepney	motorcycle	taxi	tricycle	truck	neg	
本来の カテゴリ	bicycle	4 (100%)	0	0	0	0	0	0	0	4
	car	28 (15%)	78 (40%)	1 (1%)	10 (5%)	10 (5%)	40 (21%)	12 (6%)	14 (7%)	193
	jeepney	1 (4%)	0	9 (39%)	0	0	4 (17%)	2 (9%)	7 (30%)	23
	motorcycle	3 (9%)	0	0	31 (89%)	0	0	0	1 (3%)	35
	taxi	0	0	0	0	31 (89%)	0	0	4 (11%)	35
	tricycle	0	0	0	0	0	34 (100%)	0	0	34
	truck	0	1 (7%)	0	0	0	1 (7%)	12 (80%)	1 (7%)	15
	neg	38 (38%)	0	0	6 (6%)	0	8 (8%)	1 (1%)	46 (46%)	99

把握された。したがって検出後の物体に対して行う CNN による判別は、誤検出された負例画像に対しても判別が可能である必要があるため、表-4 の判別結果を持つモデルに負例画像を新たな判別対象“neg”として追加し全8種の判別を行うモデルを構築した。neg として新たに学習させた画像は表-6 に示すカスケード分類器において負例として学習させた 3000 枚と同様の画像を用いたところ結果は表-9 のようになった。車両7種の各カテゴリにおける精度は表-5 の結果と同程度であり、負例画像を追加したことによる、車両の誤判別への影響はないと考えられる。また負例画像自身も全て誤判別なく判別できていることから、本モデルは車両7種および背景に相当する画像を判別する上で適用可能であることが示唆される。

(2) R-CNN による LAMAT の検出・判別

現地撮影画像 50 枚を対象にカスケード分類器による検出を行い、物体候補として検出された計 438 枚の画像に対し前節にて得られた CNN モデルによる判別を行った結果を表-10 示す。表-9 に示したテスト用の画像を用いた結果より相対的に判別精度が低下する結果となったが、これは検出された車体にはズレや欠損が生じていることが原因であると考えられる。

目視より 438 枚の検出結果のうち neg を除く 339 枚の検出車体において、検出窓内に車体の面積の約 8 割が含まれていれば欠損が生じていないと定義し、欠損のあり

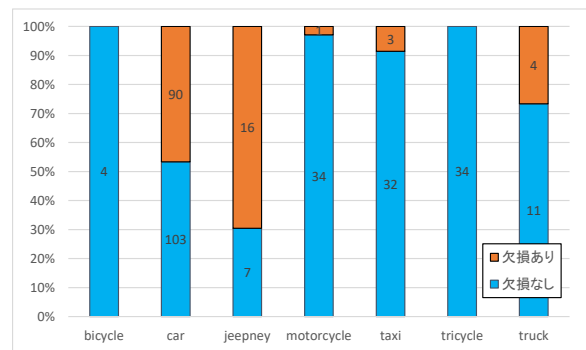


図-9 車種別の欠損検出発生割合

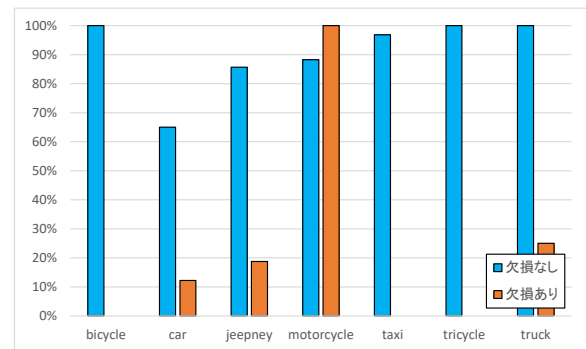


図-10 欠損の有無による判別精度の比較

なしを確認したところ図-9 のようになった。検出時の欠損が生じているのは四輪車が大半であり、car, jeepney については総検出数の約半分が、欠損が生じた状態で検出されている結果となった。次にサンプルを欠損ありなしに二分し別途判別精度を確認したところ図-10 のようになった。欠損が生じていない画像に対しては表-9 の結果

表-11 負例画像枚数を追加した CNN モデルによる分類器の検出部位への判別結果

		推計後のカテゴリ								計
		bicycle	car	jeepney	motorcycle	taxi	tricycle	truck	neg	
本来の カテゴリ	bicycle	1 (25%)	0	0	2 (50%)	0	0	0	1 (25%)	4
	car	0	59 (31%)	0	3 (2%)	10 (5%)	0	10 (5%)	111 (58%)	193
	jeepney	0	0	8 (35%)	0	0	0	1 (4%)	14 (61%)	23
	motorcycle	1 (3%)	0	0	28 (80%)	0	0	0	6 (17%)	35
	taxi	0	0	0	0	28 (80%)	0	0	7 (20%)	35
	tricycle	0	0	0	0	0	29 (85%)	0	5 (15%)	34
	truck	0	0	0	0	0	0	11 (73%)	4 (27%)	15
	neg	1 (1%)	0	0	0	0	0	0	98 (99%)	99

と同等の判別精度にて車種の判別が行えていることが分かる、一方欠損が生じている状態で検出された画像に対しては、car, jeepney, taxi, truck の判別精度は欠損なしの画像と比較して、著しく低下する結果となった。したがって車体が欠損して検出され学習時の車体の外形と異なるような場合に対しては正確な判別が行えないことが示唆される。すなわち分類器によっていかに欠損を生じずに車体候補を検出できるかがその後の CNN による車種判別の精度に大きく寄与する要因であると考えられる。



図-11 負例画像として追加した欠損検出画像の例

### (3) CNN の学習画像の追加変更による判別精度向上への検討

表-10 における結果では本研究の主対象とした tricycle に各車種が誤判別される傾向が散見された。この傾向については実務において本ツールを活用し tricycle の交通量を求めた際、過大に評価する可能性があることから CNN の学習画像の変更による判別結果の改善を検討した。

分類器による検出において tricycle は欠損が生じていないことと、欠損検出画像の欠損の生じ方は車両により大きく異なることから、今回は tricycle の判別を優先し、欠損検出画像を正カテゴリに判別させる改善ではなく、欠損検出画像を全て“非車体”と判別する対応策を検討した。図-11 に示すような車体の欠損画像を新たに“neg”のカテゴリに 1800 枚追加し再度 CNN モデルを構築した。得られたモデルに対し、前節にて R-CNN の精度検証に用いた現地撮影画像 50 枚と同様の画像を用いて判別を行った結果を表-11 に示す。

判別結果は tricycle へ誤判別されるケースを解消することが出来た一方で、各車両の判別精度も微減する結果となった。これは欠損画像であっても車体を構成する特徴量が含まれており、それを非車体と定義づけたこと

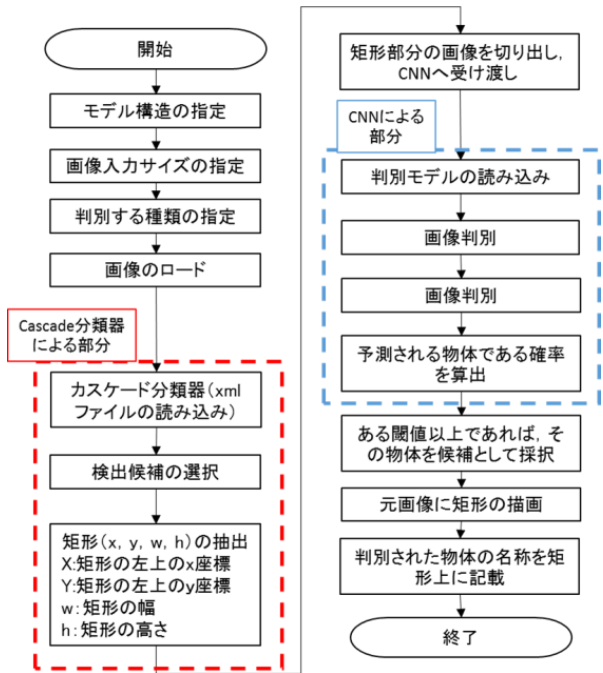


図-12 カスケード分類器を用いた R-CNN のアルゴリズムによる学習阻害が生じた可能性が考えられる。

以上の手順で行った R-CNN のフローチャートを改めて図-12 に示しておく。

## 7. まとめと今後の課題

本研究では、R-CNN の方法論を基に衛星写真等に相当



表-11 CNN モデルの判別精度に寄与する要因候補

条件	検討事項	予想される要因	今回検討した要素
自然環境 条件	日向	影による物体外形への影響	○
	日陰	影による物体外形への影響	○
	雨天	画像内の映り込み、ノイズの発生	
	雪天	画像内の映り込み、ノイズの発生	
撮影条件	夜間	影による物体外形の読み取りへの影響	
	車体の向き	対象の写る角度による情報変化	
	走行によるブレ	走行速度による車体のブレの影響	
CNNの 条件	撮影者によるブレ	撮影画面自体のブレによる影響	
	精度のばらつき	入力データが変化した際の精度誤差の変化	○
	最大学習回数	学習の仕方による影響	
	学習車	学習の仕方による影響	
	層構造	ネットワーク構造の変化による影響	
	入力画像のサイズ	入力情報の変化による影響	
データ作成 の条件	学習画像の枚数	学習させる画像の枚数の変化による影響	
	元画像の解像度	入力情報の変化による影響	○
	トリミングの仕方	画像の切り抜き方の変化による影響	
	リサイズの仕方	画像加工のアルゴリズム変化による影響	

する、車両を上部から捉えた画像により、LAMAT を含む車両の検出・判別を行うツールの開発を行い、開発において必要となる条件の幾つかを分析し、幾つの特徴や性能等を明らかにした(表-12)。CNN による車種判別ではLAMATを含むフィリピン国マカティ市を走行する7種の交通モードを判別するモデルを構築し、画像判別精度に寄与する要因を洗い出し、日影による画像情報の変化による影響や学習データ次第で変わる判別精度のばらつき等について把握した。

本研究で構築したツールにより、カスケード分類器とCNNの双方を用いることで複数台の車体が撮影されている実際の道路上のデータから、目的とする車体を検出し、検出物体の車種を判別することを可能とした。またtricycleを含む車両7種及び、背景に相当するnegativeを判別することが可能となり、tricycleを検出し判別する上では、他車両からの誤判別を抑えた上で85%の精度を確

保することができた。

今後の課題としては、本研究で開発したツールが現地撮影画像からの検出・判別を行えるものの、GoogleEarth画像等の粗い解像度に対して十分に対応可能とする必要がある。またtricycleの分布把握を行う上で、より広範囲の分布情報を取得するため、画像を連続して読み込ませる等のツールとしての工夫が求められる。

## 参考文献

- 1). Veng Kheang PHUN, Tetuso YAI (2015), State of the Art of Paratransit Literatures in Asian Developing Countries, EASTS 11th Conference, pp. 1169
- 2). Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014
- 3). 1). Achmad WICAKSONO, Iv LIM, Yasunori MUROMACHI, Karl N. VERGEL, Kasem CHOOCHARUKUL, Van Hong TAN, Kiyohisa TERAJ, Daisuke FUKUDA, Tetsuo YAI (2015), Road-based Urban Public Transport and Paratransit in Six Asian Countries: Legal Conditions and Intermodal Issues, Journal of the Eastern Asia Society for Transportation Studies, Vol. 11, pp. 227-242
- 4). Sebastien Razakarivony, Frederic Jurieb (2016), Vehicle detection in aerial imagery : A small target detection benchmark , Journal of Visual Communication and Image Representation Volume 34, January 2016, Pages 187–203
- 5). Matthias Limmer, Julian Forster, Dennis Baudach, Florian Schüle, Roland Schweiger, Hendrik P.A. Lensch (2016), Robust Deep-Learning-Based Road-Prediction for Augmented Reality Navigation Systems, <http://arxiv.org/abs/1605.09533>
- 6). P. Viola and M. Jones, "Robust real-time object detection," Proc. of IEEE Workshop on Statistical and Computational Theories in Computer Vision, pp.1–25

## A Basic Study on Observing Distribution of LAMAT using R-CNN

Takahiro KONNO, Yuki ARAI, Tetsuo YAI

In Southeast Asia, there are unique transport modes: LAMAT which stands for locally adopted, modified and advanced transport. Although these are rooted as the legs of local residents, they often have no prescribed route and stop. Due to insufficient registration and management by local governments, LAMAT's actual operation and distribution are not clear in some countries.

We developed a tool to detect LAMAT's car body from the satellite image data such as Googleearth and to estimate their distribution in the city using the method of R-CNN (Region with CNN). We verified the accuracy and the robustness of the developed tool by applying the tool to tricycle and jeepney along the actual street in Makati, the Philippines. We could grasp some factors that contribute to accurate detection and classification of LAMAT.