説明可能なAIを活用した CNN型交通計測モデルの戦略的高度化

芳賀 柚希[1]・柳沼 秀樹[2]・寺部　慎太郎[3]・海野　遥香[4]・鈴木　雄[5]

[1]学生非会員　東京理科大学大学院　理工学研究科土木工学専攻（〒 278-8510 千葉県野田市山崎 2641）
E-mail: 7622531@ed.tus.ac.jp
[2]正会員　東京理科大学准教授　理工学部土木工学科（〒 278-8510 千葉県野田市山崎 2641）
E-mail: yaginuma@rs.tus.ac.jp
[3]正会員　東京理科大学教授　理工学部土木工学科（〒 278-8510 千葉県野田市山崎 2641）
E-mail: terabe@rs.tus.ac.jp
[4]正会員　東京理科大学助教　理工学部土木工学科（〒 278-8510 千葉県野田市山崎 2641）
E-mail: unoharuka@rs.tus.ac.jp
[5]正会員　東京理科大学助教　理工学部土木工学科（〒 278-8510 千葉県野田市山崎 2641）
E-mail: yusuzuki@rs.tus.ac.jp

昨今，交通量計測の実務では，人手観測から AI 画像解析を援用した自動観測に転換しつつある．畳み込みニューラルネットワーク (CNN) を基本とする交通量計測 AI により，交通量が精度良く常時観測可能となっている．しかし，夜間や逆光などの特定条件下では著しく計測精度が低下しており，実務的な利用には改善の余地がある．本研究では，様々な条件における交通量の計測精度向上を目的とする．具体的には，CNN に対応した説明可能な AI（XAI）手法である Grad-CAM から得られる判断根拠をベースに戦略的なアノテーションと転移学習を実施する．これにより，CNN が獲得出来ていない特徴量を重点的に学習させることが可能となる．この手法を道路上の CCTV カメラ画像に適用した結果，最大約 30%の交通量計測精度の向上が確認された．

Key Words: *Convolutional Neural Network, Explainable AI, Grad-CAM, Layer-wise Relevance Propagation, Traffic volume survey*

1. はじめに

国土交通省が実施している一般交通量調査は道路の計画，建設，管理等に必要不可欠な基礎資料である．しかしながら，5 年に 1 度の調査では時空間的に細やかな道路サービスに結びつかないという課題がある．また，平成 27 年度調査では，交通量調査員による手動計測が主たる計測方法であり，時間的・人的コスト面での問題の解消が緊要であった．そこで近年では，解決策の一つとして AI カメラによる調査が実施されつつある，しかしながら，AI カメラの計測精度が特定条件下で著しく悪化することが課題となっている[1)]．特定条件は主に 3 つに分類され，1 つ目に，逆光や夜間などの日照条件の変化．2 つ目は，雨や雪などの気象条件の変化．最後に，カメラの画角や遮蔽物などの設置条件による精度低下である．本研究では，このうち日照条件を中心に改善を図る．

改善方法として，説明可能な AI(Explainable AI，以下 XAI) と呼ばれる手法を用い，解釈性を持たせた上で AI 解析の特性を分析した．具体的には，Grad-CAM(Gradient-weighted Class Activation Mapping) および LRP(Layer-wise relevance propagation) を用いることで，CNN が画像のどの部分を重要視しているかをヒートマップの出力によって判別することが可能となる．分析に基づき，事前学習モデルが得ている特徴量の不足部分を補うようなデータセットを作成する戦略的アノテーションを行った．戦略的アノテーションは，通常の量的にデータを投入する学習方法の上，投入するデータの質を中心とする学習の実施を目的とする．最後に，ネットワーク構造を一部固定した状態で事前学習モデルを更新する手法である転移学習を実施した上で，精度検証と再度の XAI による分析を行い，転移学習が与えた影響を考察する．

2. 本研究で使用する AI モデル

(1) CNN

a) CNN の基本的構造

CNN は，NN(Neural Network) をベースに開発され，画像処理タスクに広く使われる手法である．NN では入力層，中間層，出力層の 3 つが最も基本的な構造となる．CNN では，中間層を畳み込み層とプーリング層で構成することで，画像内の位置情報を損なわず特徴を抽出することができる．畳み込み層では，入力画像に

対してフィルタを適用し特徴を検出する．畳み込み処理は RGB チャンネル毎に行われるものが一般的で，その上でバイアス加算や活性化関数による処理が加えられる．その後，プーリング層において画像各領域を代表する値を並べることで，位置に関する情報を落とし，位置変化へのロバスト性を確保する役割となっている．これらの処理を繰り返したのち，抽出した特徴量は全結合層でまとめられ，活性化関数を通して検知した物体が各クラスに分類しうる確率として出力される．

基本的にこれらの精度は適合率 (Precision) と再現率 (Recall) より求められる mAP(mean Average Precision) によって評価されることが多い．適合率は予測した全てのバウンディングボックスのうち，正解であるものを表す値で，再現率は正解であったバウンディングボックスのうち正しくクラス分類したものといえる．これらは 1 式，2 式によって求められる．

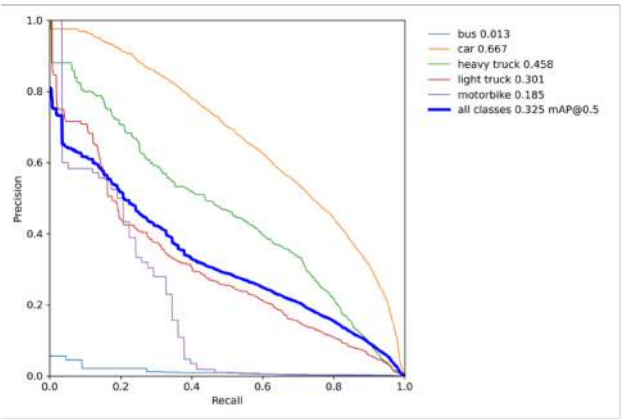
表–1 TP，FP，FN の定義

	正解	不正解
検知された	TP	FP
検知されない	TN	FN

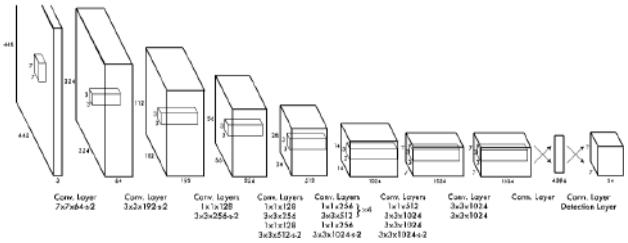
$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

なお，**表–1** にて，TN は通常定義されない．**図–1** を例に示すような，Precision を縦軸に，Recall を横軸にとったグラフを PR 曲線とし，曲線より下部分の面積を AP(Average Precision) と呼ぶ．それらのクラス全ての平均が mAP と定義される．



図–1 PR 曲線



図–2 YOLO アーキテクチャ

b) YOLOv5 モデルの概要

CNN によるリアルタイム物体検出手法の一つとして，YOLO シリーズによるものが挙げられる．YOLO は 2015 年，One-Stage 型の検出手法として Joseph[2] らにより提案された手法である．One-Stage 型検出器は，検出対象を囲うバウンディングボックスを回帰として予測し，その後クラス分類を行う．したがって，FasterR-CNN のような Two-Stage 型検出器と比較して高速な推論が可能である．現時点では開発元を変えながら YOLOv7 に至るまで複数のバージョンが公開されている．YOLOv5 は ultralytics によって 2020 年に公開されたバージョンであり，以前のバージョンに比べて計算速度が高速化されていることが特徴である．YOLOv5 を採用する理由として，推論時間の高速さや XAI への拡張性を加味した結果である．拡張機能として alubumentation のようなデータオーグメンテーションや，ハイパーパラメータの自動調整機能などが公開されている．また本研究では，Microsoft COCO Dataset によって事前に学習された軽量な YOLOv5s モデルをベースに転移学習を進める．

(2) XAI の概要

一般的に，AI は中間層でどのような処理が行われ，特徴量を選択しているかを解釈することは困難とされている．近年では，AI の説明性や透明性を向上させるという目的で，様々な XAI(Explainable AI) の手法が開発されている．画像分野においては，ネットワークの判断根拠をヒートマップとして可視化する手法が考案されており，一定の定性的な解釈が可能である．本節では，CNN モデルに対して適用可能である Grad-CAM と LRP について述べる．

a) Grad-CAM

Grad-CAM[3] は CAM(Class Activation Mapping) をベースに開発された，CNN の判断根拠を可視化する手法の一つである．CAM では CNN が保持している位置情報 A^k_{xy} と，特徴量マップの画素平均をまとめる処理を行う GAP(Global Average Pooking) 層における重み w^k_c の積和を取ることで，値の大きさによりクラス分類への影響力を算出することができる．結果をヒートマップとして出力することで，画像のどの部分を重要視し

て物体を判断しているか，視覚的に解釈可能となる．クラス c における Grad-CAM は

$$L^c_{Grad-CAM} = ReLU(\sum_k \alpha^c_k A^k) \tag{3}$$

のように出力される．なお，重み α^c は，

$$\alpha^c = \frac{1}{Z}\sum_{i,j}\frac{\partial Y^c}{\partial A^k_{ij}} \tag{4}$$

A_{ij}：特徴量マップ
Z：正規化定数
Y^c：クラス c に分類される確率
と表される．CAM では GAP 層が存在しないモデルに対しては適用できないため，Grad-CAM では重みの代わりに出力に対する勾配を出力することによって，任意の CNN モデルで利用可能となる．また，ヒートマップを出力する際に，ReLU 関数を通すことにより，判断に正に寄与する部分のみを可視化する．Grad-CAM による車両の分析は，Lee[4] らが行った蒸留と呼ばれる教師モデルから生徒モデルへのモデルの圧縮処理の結果検証で用いられた例がある．Web スクレイピングにより取得した車両画像の判断根拠可視化を行っているため，日本の車両形態や CCTV 特有の画角や画質を考慮された判断は反映されていない．

b) LRP

LRP(Layer-wised Relevance Propagation)[5] は CNN の判断根拠を可視化する手法である．レイヤー毎における各入力要素の貢献度の総和は変わらないという特性をもとに設計された．各層のニューロンにこの関連度を出力層から入力層に向かって逆伝播させることで，各予測クラスに対する入力層の貢献度をヒートマップとして出力することが可能である．

$$R_j = R_k \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} \tag{5}$$

R_j：入力層側の層の関連性 (Relevance)
R_k：出力層側の層の関連性
a_j：ニューロンの活性化
w_{jk}：ニューロン間の重み
5 式は，LRP の最も基本的な式であり，層の深度によってノイズ除去のための処理を加える必要がある．逐次的に計算することで，最終的な出力に対するニューロンの影響の合計を辿ることが可能である．また，計算されるニューロンより下層部分の貢献度の総和で除することで，上述した貢献度の保存性を確保する．YOLOv5 では，Karasmanoglou[6] らにより実装がなされており，バウンディングボックス内で LRP が可視化されるような調整が行われている．

表–2 主な地点別の計測精度と課題点

地点名	精度 (%)	課題点
箱根新道 01	going:56.4 coming:53.0	誤分類
田野倉	going:84.6 coming:81.7	日陰 低画質
神宮橋潮来側	going:56.5 coming:63.4	レンズフレア 誤分類
上江橋 (下り)	going:93.5 coming:77.9	日陰 遮蔽
仲町二丁目	going:86.5 coming:91.8	遮蔽
湯舟橋	going:63.4 coming:56.9	低画質

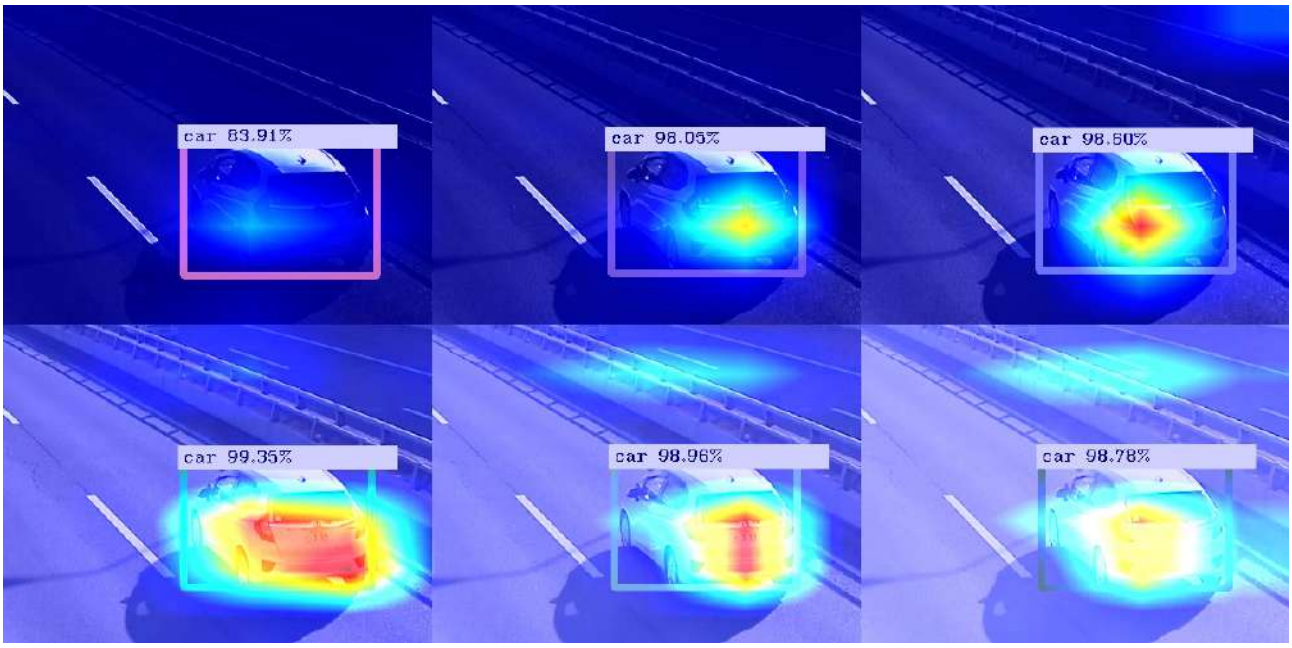
3. XAI を用いた判断根拠の抽出

はじめに，YOLOv5 における Grad-CAM を用いて解釈を行った．しかしながら，解釈性が低いため．事前分析においては YOLOv3 バージョンの Grad-CAM を用いる．YOLOv5 アーキテクチャは YOLOv3 に高速推論を行う機能を追加したものであるため，基本的な構造は変わらないものとした．同時に LRP による分析を行う．LRP に関しては転移学習後の分析も実施する．

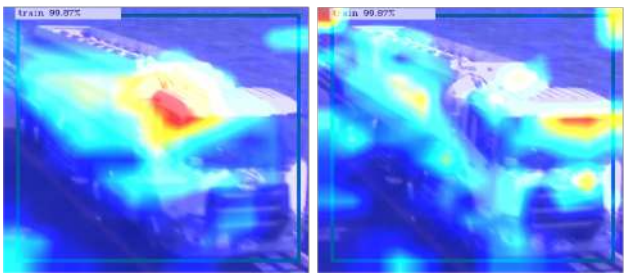
(1) 分析対象データ

分析に使用したデータは，国土交通省関東地方整備局より提供された CCTV カメラ画像である．**表–2** では，撮影全時間帯における精度と定性的に分析した課題を示す．以下より，動画中車両が奥から手前に向かう車線を coming 方向，手前から奥へ向かう車線を going 方向と呼称する．精度が 80%を下回るような地点に対し共通する課題として，主に 3 点の課題が挙げられる．1 点目に適切なクラス分類が行われないことが挙げられる．特に，正解が truck である車両を，train として分類が行われるケースが多く存在した．また，車両に反射するフレアやヘッドライト部分を person や motorbike と分類して適切な検知が行われないケースが多い．2 点目に，車両が日陰部分を通過するような地点では，車両の検知率が低下する．3 点目に，夜間時，車両のヘッドライトのフレアによる遮蔽や画像の検知が不可能となる点がある．ヘッドライトのフレアは低画質のカメラで特に悪化しやすく，目視による分類が困難な場合も存在する．

図–3 車両のマスキングによる比較



図–4 ガンマ値の変化による比較



図–5 truck と train の判断根拠可視化

(2) XAI による判断根拠の可視化と抽出

以上の課題を踏まえた上で，Grad-CAM と LRP による判断根拠の可視化を実施した．検知される場合の全体的な傾向として，車両を立体的に取るような輪郭部分に強く反応する傾向がある．これらは Grad-CAM と LRP 同様な反応を示した．また，車両全体に対して反応する．検知自体が行われない場合，反応を示さないあるいは全体的に弱いケースが多い．特に，車体側面や窓部分に帯状に反応が分布するものが多く，タイヤを含む車体下部や上部は優先されない傾向がある．クラス分類が誤っている場合，大型貨物車やバスのミラーや窓など，普通乗用車と比較して特徴的な部分に対して強い反応を示す．これらは中型貨物車等の似た部分を持つ車両と混同してクラス分類が行われることが考えられる．**図–3** では，車両をマスキングした上で可視化を実施したところ，車両側面のみを根拠として物体検知がなされていることがわかり，画角により正解率に差が生じることが考察される．車両が日陰部分を通過するケースでは，車両の明度の違いにより判断が異なるという仮説のもと分析を実施した．**図–4** に示す通り，画像のガンマ値を変化させた比較では，背景と車両の明度が近い場合に反応が弱くなる傾向が見られる．**図–5** のように，車両上部や二台部分を中心に反応しており，車と鉄道の相違点である車輪部分の考慮はほとんどされていないことを読み取ることができる．これ

らはクラス分類確率の差としても表れる．これらの分析よりアノテーションの対象は，
(1) 車両の一面のみしか見えないような画角を持つ地点
(2) 影がかかっているまたは夜間
(3) 貨物車が多く通行する地点
3 点を軸に選定を行った．

4. 判断根拠に基づく戦略的なモデル学習

(1) 転移学習の概要

転移学習は，事前に学習された CNN ネットワークの構造を一部固定させることで，比較的少量のデータセットからモデルの持つ知識を転用するように学習する方法である．データセットの作成は，対象の物体にラベル付けを行うアノテーションと呼ばれる作業を必要とする．アノテーション作業は複数の人員と時間を必要とする．Microsoft COCO や PASCAL VOC といった数十万枚規模の大規模データセットに匹敵するデータセットを作成する必要はなく，新たな学習モデル作成の時間短縮と省力化が可能となる．

(2) 転移学習用データセット

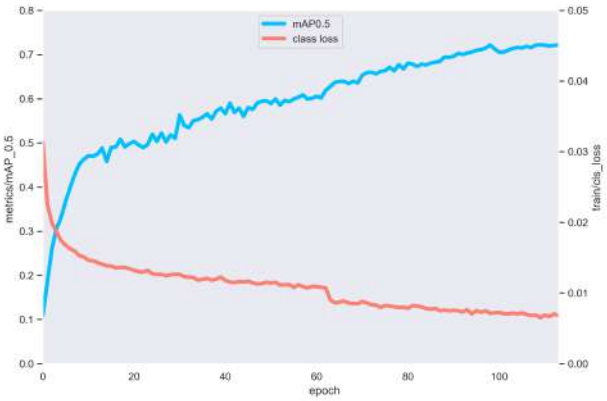
転移学習用データセットは Grad-CAM で分析した地点を含む 5 地点から中心に設定される．車種分類は，国土交通省の一般交通量調査の 9 分類に準拠するが，本研究では簡単のため歩行者や自転車を除く 5 種類に分類した．分析に基づき，車両の 1 側面のみが映り込むような地点，日陰もしくはフレアが強く反映されるような地点，truck と train のクラス誤分類が発生しやすい地点を主なアノテーション対象とする．アノテーションされたラベル数は**表–3** に示す通りとする．また，画像を加工することで 1 枚の画像を複数の画像として実質的に水増しするデータオーグメンテーションを行う．オーグメンテーション学習時の主なパラメータおよびオーグメンテーションの内容は**表–4** の通りに設定した．

表–3 アノテーション内容

ラベル名	ラベル数
car	28,261
light truck	6,221
heavy truck	3,552
bus	408
motorbike	143
合計	38,585

表–4 主な学習パラメータ設定

batch size	16
image size	416 × 416
epoch	113
augmentation	Blur, MedianBlur, RandomShadow Verticalflip, Shear, ToGray, CLAHE, RandombrightnessContrast, RandomGamma, ImageCompression, RandomRain, RandomSnow



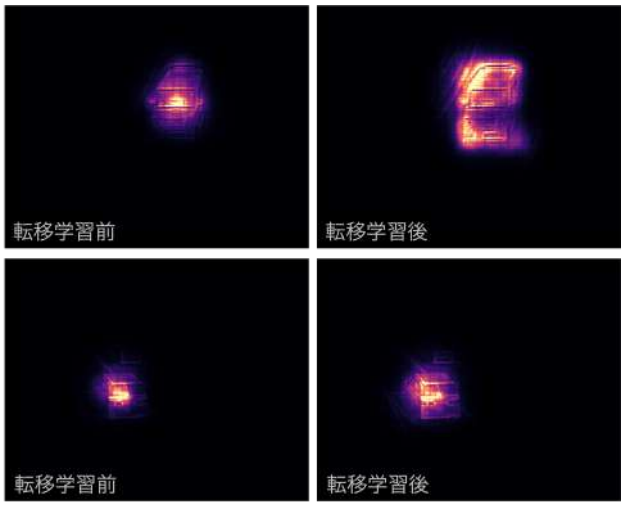
図–6 class loss と mAP0.5 の推移

5. 学習結果と精度検証

モデルの性能を表す指標 mAP0.5 は 72.2 を記録した．またオーバーフィッティングと見られるような挙動は示していない．精度検証は特に精度が低いとされた**表–5** 中の 3 地点 6 時間分で行った．なお，本研究での精度は人による目視計測と AI 計測によるカウントの比とする．このうち，全対象で精度の向上が認められる．精度が向上した要因が 2 点考えられる．1 点目に，解像度が低くインターレース方式による録画となっていた箱根新道や湯舟橋では 30%以上の精度向上となっている．推察される要因として，そのため，インターレース方式は動画送受信法の一つであり，画像が縞状に描画され，残像が残ることがある．本転移学習ではインターレース解除などの事前処理をせず実施し，そのまま特徴を取得したため，低画質のまま推論を行った場合でも精度が向上したと推察される．2 点目に，適切なクラス分類が行われたことによる．事前学習モデルは 80 クラス分の分類がされるようになっているが，本転移学習では車両のみに知識を限定しため，train などの余分な知識を排除した状態で推論することが可能である．

転移学習後の判断根拠可視化は LRP を用いて行った．夜間の判断根拠可視化を実施したところ，**図–7** では車

図–7 夜間における LRP の反応



図–8 LRP による転移学習前後の反応の可視化

図–9 転移学習前後のクラス分類の変化

両本体よりもヘッドライトフレア自体の特徴量が取得されていることが示唆される．このように条件が大きく変化すると全く別の特徴を得るため，時間帯や条件ごとでモデルを変更し精度を維持するという方法を考慮する余地がある．全体的な変化として，**図–8** に示す通り，転移学習前の車両と比較し，転移学習後は車両の反応が顕著となっていることがわかる．フロントガラスの部分的な反応にとどまらず，車両上面も含めた輪郭をとるような反応に変化していることがわかる．クラス分類は転移学習時に 5 クラスに絞った結果，適切な把握が可能となった．また，誤分類の原因となっていた person や train は検出されなくなったため，その分の車両が上乗せされる形で計測精度が向上したと考えられる．

表–5 転移学習前後の精度

地点	時刻	精度 (%)	
		学習前	学習後
箱根新道 01	7 時	64.1	83.4
	14 時	62.1	83.1
神宮橋潮来側	12 時	68.4	86.5
	13 時	72.5	89.6
湯舟橋	8 時	48.9	81.7
	11 時	60.9	87.8

6. おわりに

Grad-CAM や LRP を用いて YOLOv5 モデルの判断根拠を示し，それに基づいた転移学習を実施した．XAI を用いた戦略的アノテーションでは，中間層における判断根拠を定性的評価することで，アノテーション方針の明確化につながった．転移学習後では対象全地点で精度向上となり，最大で 26.9%の精度向上が見られた．また，夜間や低画質など，推論動画の事前処理が予想された対象においても，転移学習のみでの精度向上が可能であることが示された．

XAI に関する問題として，あくまで定性的な解釈に留まり，考察の範囲も XAI の判断根拠の解像度に依存する．また学習させたい特徴量を実施者側が選択することも困難である．全体における課題では，歩行者や自

転車等のモード別検知や，雨や雪などの悪天候時における検証が不十分であることが挙げられる．また，検知後の追跡プロセスでは，車両 ID 付与の最適化や画角変化への対応といった面で改善の余地がある．引き続き検知・追跡モデル部分での改善を行うとともに，アノテーション方法や転移学習の効率的かつ適切なフローを構築していく必要がある．

参考文献

1) 国土交通省道路局企画課道路経済調査室 and 国土交通省国土技術政策総合研究所道路研究室: Cctv カメラ (ai 解析) の精度に関する報告, 2021, https://www.mlit.go.jp/road/ir/ir-council/ict/pdf05/02.pdf.
2) Redmon, J., Divvala, S. K., Girshick, R. B., and Farhadi, A.: You only look once: Unified, real-time object detection, *CoRR*, Vol.abs/1506.02640, 2015.
3) Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D.: Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization, *CoRR*, Vol.abs/1610.02391, 2016.
4) Lee, Y., Ahn, N., Heo, J. H., Jo, S. Y., and Kang, S.-J.: Teaching where to see: Knowledge distillation-based attentive information transfer in vehicle maker classification, *IEEE Access*, Vol.7, pp.86412–86420, 2019.
5) Lapuschkin, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W.: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, *PLoS ONE*, Vol.10, pp.e0130140, 07 2015.
6) Karasmanoglou, A., Antonakakis, M., and Zervakis, M.: Heatmap-based explanation of yolov5 object detection with layer-wise relevance propagation, pp. 1–6, 2022.

(2022.9.30. 受付)

Strategic advancement of CNN-based traffic measurement models using Explainable AI

Yuzuki HAGA, Hideki YAGINUMA, Shintaro TERABE, Haruka UNO, Yu SUZUKI

Recently, the practice of traffic volume measurement is shifting from manual observation to automatic observation with the help of AI image analysis. Traffic volume measurement AI based on Convolutional Neural Network (CNN) enables accurate and constant observation of traffic volume. However, under certain conditions, such as nighttime and backlight, the measurement accuracy is significantly degraded, leaving room for improvement for practical use. This study aims to improve the accuracy of traffic volume measurement under various conditions. Specifically, strategic annotation and transfer learning are performed based on the decision basis obtained from Grad-CAM, an explainable AI (XAI) method corresponding to CNN. This makes it possible to focus learning on features that have not been acquired by CNN. The results of applying this method to CCTV camera images on roads show an improvement of up to approximately 30% in the accuracy of traffic volume measurement.