2025 年 3 月 15 日

＜研究課題＞　拡散モデルと画像セグメンテーションを用いた転移学習用画像生成による特定小型原動機付自転車の不適切な歩道走行の検出の実現

代表研究者　埼玉大学大学院理工学研究科　助教　間邊　哲也

【抄録】
本研究課題では，特定小型原動機付自転車の不適切な歩道走行を検出するため，拡散モデルと画像セグメンテーションを用いた転移学習用画像の効率的の生成方法の提案と評価を行っている．具体的には，車道走行時に得られた画像を視点変換することで学習用画像を生成していた従来手法に対して，①誤識別につながる車両オブジェクトの削除，②背景領域の画像補完に拡散モデルをそれぞれ用いている．①は，従来手法と比べて 1 枚あたりの学習データ作成時間は増加したが，視点変換アルゴリズムに車両オブジェクトの画像補完を追加することで走行環境の判定性能が向上した．②は，従来手法よりも判定性能は高く，学習データ作成時間は短くなった．以上から，拡散モデルと画像セグメンテーションを用いた転移学習用画像生成による特定小型原動機付自転車の不適切な歩道走行の検出の実現に資する知見を獲得している．

１．研究の目的

本研究課題の目的は，特定小型原動機付自転車による交通事故を防ぐ安全運転支援システムの実現に向けて，不適切な歩道走行検出に用いる画像ベースの走行環境識別に必要な転移学習用画像生成の効率化を図ることである．

不適切な歩道走行を検出するためには車両が歩道と車道どちらを走行しているか判定する必要がある．研究代表者はこれまで，ドライブレコーダーのようなカメラを車両に取り付けることを前提とした走行環境判定に関する研究を行ってきた (例えば[1])．この研究では，車道走行画像と歩道視点画像を転移学習することで高い判定性能を得ているが，学習に必要なアノテーション画像を手作業で作成しているため時間が掛かっていた．

２．研究方法と経過

2-1 画像生成を用いた学習データ作成方法

文献[2]では車道走行画像を視点変換して歩道視点画像を作成することで学習データ作成時間を削減していたが，視点変換による歪みが走行環境の判定性能の低下の要因となっていた．画像生成技術を用いて画像の欠損や画像の望まない部分を修復・補完する技術を画像補完と呼び，本研究では画像補完によって視点変換により生じる歪みを軽減することで走行環境の判定性能の向上を図る．車両オブジェクトの画像補完と背景領域の画像補完の 2 種の方法で視点変換による歪みを軽減する．

2-2 車両オブジェクトの画像補完

文献[3]では，歩道セグメンテーションは車両の存在に敏感であると述べられており，車両オブジェクトの歪みの影響が高いと考えられる．画像補完により車両オブジェクトの映っていない車道走行画像を視点変換することで歪みを軽減する．

画像補完方法は，図 1(a)の車道走行画像の車道上に存在する車両オブジェクトを抽出して図 1(b)のようにマスク画像を作成する．マスクとして抽出した領域を画像補完して図 1(c)のような車両オブジェクトが映っていない画像を生成する．補完した画像の良否は目視で判定し，アノテーション画像の画像補完した領域はラベルなしとした．画像補完ツールとして Stable Diffusion Web UI [4]の inpainting 機能を使用し，LAION-5B [5]データセットを LDM (Latent Diffusion Models) [6]でトレーニングした画像生成モデルに LaMa (Large Mask Inpainting) [7]で転移学習したモデルを使用した．

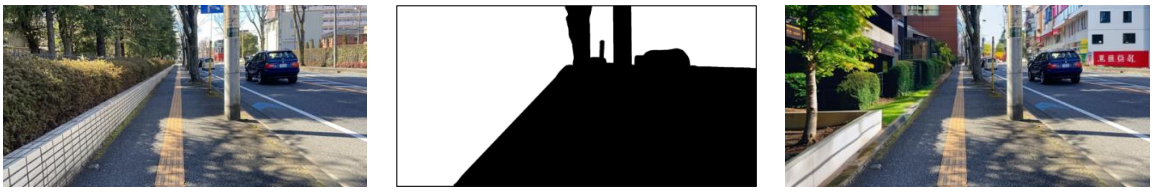
視点変換は，図 2 のように視点変換に必要な消失点，左右の道路領域と背景領域の境界面の画像端，車道と歩道の境界面の画像端の 4 点を環境ごとに求める．求めた 4 点を用いて歩道側に消失点を移動させて視点変換することで図 3 のような歩道視点画像を作成する．このとき

(a) 元画像　(b) マスク画像　(c) 画像補完画像

図 1 画像補完のためのマスク領域とマスク領域を作成した画像補完画像

図 2 視点変換のための 4 点

図 3 画像補完＋視点変換により作成した歩道視点画像



(a) 元画像　(b) マスク画像　(c) 背景領域を生成した画像

図 4 画像補完による背景領域の画像生成

歩道側に移動させる割合は求めた歩道の中心を 1 としたときの移動割合rで表す. そして, 元画像とアノテーション画像を同じパラメータで視点変換することで学習データを作成する.

2-3 背景領域の画像補完

視点変換画像には車両オブジェクト以外にも歪みが存在する. これらのオブジェクトは背景領域にも存在するため抽出が困難である. オブジェクト検出では背景領域を画像補完することで学習データを拡張する研究[8]がある. 今回の対象は非オブジェクトである歩道や車道であり, これらのセグメンテーションは周囲のコンテキスト情報が重要である. そのため, 歩道や車道以外もアノテーションが必要である. 背景に存在するオブジェクトの既存学習モデルのセグメンテーションの性能は道路領域に比べて高い. 背景のアノテーションには既存学習モデルを使用する. しかし, 歩道視点画像や車道走行画像には車両などのオブジェクトによりオクルージョンが発生するため, 道路領域上に存在するオブジェクトを画像補完したとき, 道路領域が生成される場合がある. 特に歩道視点画像の場合, 歩道領域を車道領域と誤識別する可能性が高いため, 道路領域上のオブジェクトを残し, それ以外の背景領域を生成する.

背景領域の画像補完は, はじめに図 4(a)の元画像に対して道路領域と道路領域上に存在するオブジェクトをアノテーションする. 次にアノテーションした領域以外を背景領域として抽出して図 4(b)のようにマスク画像を作成し, 抽出した領域を画像補完して図 4(c)のような背景領域が異なる画像を生成する. 最後に生成した画像を画像セグメンテーションし, 道路領域と道路領域上に存在するオブジェクトをアノテーションした情報に変更する. 画像補完は車両オブジェクトの画像補完と同じツールを用いる.

3. 研究の成果

3-1 画像生成を用いた走行環境判定の評価

実験では視点変換のみで作成した学習データ, 画像補完＋視点変換で作成した学習データ, 背景領域の画像補完により作成した学習データ, 手作業で作成した学習データの 4 種の方法で作成した学習データを用いて比較した. 画像セグメンテーションの性能比較, 走行環境の判定性能の比較, 学習データ作成時間の比較の 3 種の比較により評価した.

学習データは 300 枚 (車道走行画像 : 歩道視点画像=1 : 4) を使用した. 視点変換を用いた 2 種の方法で作成した学習データは埼玉県さいたま市の埼大前交差点～南与野駅西口エリアと芸術劇場エリアの 2 つのエリアの車道で撮影した. 自転車の前にカメラを取り付けて車道を走行して動画を撮影し, 撮影した動画から画像を切り出した. 背景領域の画像補完と手作業で作成した学習データは埼玉県さいたま市の埼大前交差点～南与野駅西口エリアを主とする複数地域の車道と歩道で撮影した. 車道は視点変換を用いた学習デ

ータと同様の方法で撮影し，歩道はスマートフォンを用いて徒歩で撮影した．

視点変換での消失点を歩道側に移動させる割合rは基礎実験より視点変換のみが 0.50，画像補完＋視点変換が 0.83 を用いた．また，元画像 1 枚に対する背景領域を画像補完して生成する枚数は実験の結果から 10 枚とした．

3-2 評価方法

画像セグメンテーションの性能比較では 4 種の方法で学習データを作成し，作成した学習データを用いて転移学習した．トレーニングデータと検証データは 8 : 2 として分けた．このトレーニングデータと検証データの画像の組み合わせを変えた 10 種類の学習データを用意して評価する交差検証をした．既存学習モデルには Cityscapes データセット [9] を OCR (Object-Contextual Representations)+HRNet (High-Resolution Network) [10]で学習したモデルを使用した．転移学習にはツールとして MMsegmentation [11]を用い，学習モデルは OCR+HRNet を用いた．学習条件はバッチサイズを 2 とし，イテレーション回数は 40000 回とした．転移学習した学習モデルを用いて評価画像を画像セグメンテーションした．評価画像の車道と歩道をアノテーションして真値とし，それぞれの推論値と真値の重なりを求めて比較した．評価指標にはIoU (Intersection over Union)を用い，IoUは式(1)のように真値∪推論値の領域を基準としたときの真値∩推論値の領域の割合を表す．

$$IoU = \frac{\text{真値} \cap \text{推論値}}{\text{真値} \cup \text{推論値}} \times 100\ [\%] \qquad (1)$$

評価データは埼玉県さいたま市の埼大通り(南与野駅入口交差点～北浦和駅入口交差点) と与野中央通り，大泉院通りの 3 種類のルートの車道と歩道で撮影した．埼大通りは学習データに近い環境であり，与野中央通りと大泉院通りは学習データと異なる環境である．車道は 3-1 の学習データと同様の方法で撮影した．歩道は自転車の前後にカメラを取り付けて車道を走行して動画を撮影し，撮影した動画から画像を切り出した．3 種類のルートで撮影した画像をそれぞれ 40 枚(車道 : 20 枚，歩道 : 20 枚) の計 120 枚使用した．

走行環境の判定性能の比較では画像セグメンテーションの性能比較と同様の方法で転移学習して転移学習したモデルで画像セグメンテーションした．セグメンテーションした画像に四角形の領域である ROI (Region of Interest)を設け，ROI の中で車道と歩道のクラスの最頻値により走行環境を判定した．ROI 内に車道と歩道の領域がない場合は判定不能とした．ROI は横幅を画像の中央から左右に画像の横幅の 16 分の 1 とし，縦幅はトリミング前の画像の中心からトリミング後の画像の下部 6 分の 1 までとした．評価データは画像セグメンテーションの性能比較と同様に撮影し，ルートごとの枚数を表 1 に示す．評価指標には式 (2) の全評価画像に対する TP ($True\ Positive$)の割合R_{TP}を用いた．TPは正しい走行環境を判定した枚数である．

表 1 走行環境の判定性能の比較で使用した各ルートの評価データの枚数

ルート名	車道走行画像	歩道視点画像
埼大通り	204 枚	307 枚
与野中央通り	220 枚	316 枚
大泉院通り	208 枚	292 枚

$$R_{TP} = \frac{TP}{\text{全評価画像の枚数}} \qquad (2)$$

学習データ作成時間の比較では車道走行画像と歩道視点画像が同じ枚数必要であると仮定したときの 1 枚あたりの学習データ作成時間で比較した．

3-3 実験結果

作成した学習データごとのIoUの結果を表 2 に示す．画像補完＋視点変換で作成した学習データは視点変換のみで作成した学習データよりも車道が 1.1%ポイント，歩道が 2.0%ポイント高くなった．また，背景領域の画像補完で作成した学習データは画像補完＋視点変換で作成した学習データよりも歩道のIoUは 4.4%ポイント高くなり，視点変換を用いた 2 種の方法で作成した学習データよりも歩道のIoUは高くなった．しかし，車道のIoUは画像補完＋視点変換で作成した学習データに比べて 2.9%ポイント低くなり，視点変換を用いた 2 種の方法で作成した学習データよりも低くなった．視点変換を用いた学習データの車道走行画像は手を加えていない画像をしているが，背景領域の画像補完で作成した学習データは車道走行画像も画像補完している．車道走行画像での車道のIoUが低下したことで全体のIoUも低下したと考えられる．

作成した学習データごとのR_{TP}の結果を図 5 に示す．画像補完＋視点変換で作成した学習データは視点変換のみで作成した学習データよりもR_{TP}は全体で 0.02 増加した．背景領域の画像補完で作成し

表 2 作成した学習データごとのIoUの結果

	視点変換のみ	画像補完＋視点変換	背景領域の画像補完	手作業
車道	70.1%	71.2%	68.3%	79.3%
歩道	56.0%	58.0%	62.4%	74.2%

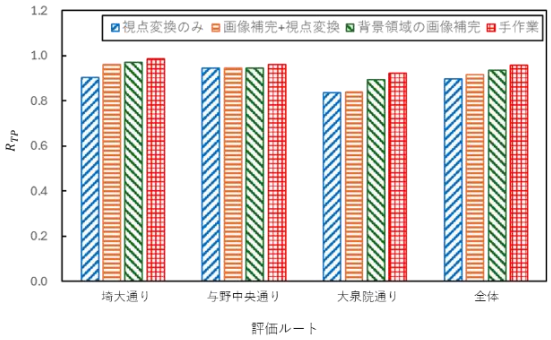


図 5 作成した学習データごとのR_{TP}の結果

た学習データは画像補完＋視点変換で作成した学習データよりも全体で 0.01 高くなり，比較した 3 種の中では手作業で作成した学習データに走行環境の判定性能が近づいた．

1 枚あたりの学習データ作成時間は視点変換のみが 373 秒，画像補完＋視点変換が 448 秒，背景領域の画像補完が 136 秒だった．画像補完＋視点変換で作成した学習データは手作業で作成した学習データに比べて約 39%削減したが，視点変換のみで作成した学習データに比べて約 20%増加した．画像補完＋視点変換で作成した学習データ画像生成技術の向上や生成画像の良否の判定の自動化により学習データ作成時間はさらに削減できると考えられる．背景領域の画像補完で作成した学習データは視点変換のみで作成した学習データよりも約 64%削減し，比較した学習データの中では最も短くなった．しかし，視点変換を用いた方法はストリートビュー画像を使用して画像の収集時間を削減したり，車道走行画像のセグメンテーションの性能の向上によってアノテーション時間を短縮したりすることで今後さらに学習データ作成時間が短くなる可能性がある．

４．今後の課題

生成画像の良否の判定の自動化や実環境での性能評価などが挙げられる．

５．研究成果の公表方法

[A]片山竣介，新井宏映，間邊哲也，“視点変換と画像セグメンテーションによる低速車両の歩道走行検出の一検討，” 電子情報通信学会技術研究報告，ITS2024-4，pp.6‐11，June 2024.（口頭発表，愛知淑徳大学（長久手市））

[B]片山竣介，間邊哲也，“画像生成と視点変換を組み合わせた画像セグメンテーションによる特定小型原動機付自転車の歩道走行検出の一検討，” 電子情報通信学会技術研究報告，ITS2024-42，pp.169‐174，Dec. 2024.（口頭発表，宝山ホール（鹿児島市））

[C]片山竣介，間邊哲也，“画像セグメンテーションによる特定小型原動機付自転車の歩道走行検出の ROI に対する性能評価，” 情報処理学会研究報告，2024-ITS-100(14)，Mar. 2025.（口頭発表，同志社大学（京都市））

参考文献

[1] T. Manabe et al, “Bicycle riding environment identification for detecting traffic violation in a riding safety support information system,” IATSS Res., vol.48, no.3, pp.357–366, Oct. 2024.

[2] 新井宏映，“自転車の走行環境判定に関する研究，” 2022 年度埼玉大学大学院理工学研究科博士前期課程学位論文，Feb. 2023.

[3] R. Shetty et al, “Not using the car to see the sidewalk –Quantifying and controlling the effects of context in classification and segmentation,” Proc. IEEE/CVF CVPR, Long Beach, CA, USA, pp.8218–8226, June 2019.

[4] https://github.com/AUTOMATIC1111/stable-diffusion-webui

[5] C. Schuhmann et al, “LAION-5B: an open large-scale dataset for training next generation image-text models,” Proc. NIPS, New Orleans, LA, USA, pp.25278–25294, Dec. 2022.

[6] R. Rombach et al, “High-resolution image synthesis with latent diffusion models,” Proc. IEEE/CVF CVPR, New Orleans, LA, USA, pp.10684–10695, June 2022.

[7] R. Suvorov et al, “Resolution-robust large mask inpainting with fourier convolutions,” Proc. WACV, Waikoloa, HI, pp.2149–2159, Jan. 2022.

[8] Y. Li et al, “A simple background augmentation method for object detection with diffusion model,” Proc. ECCV, Milano, Italy, pp.462–479, Nov. 2024.

[9] M. Cordts et al, “The cityscapes dataset for semantic urban scene understanding,” Proc. IEEE CVPR, Las Vegas, NV, USA, pp.3213–3223, June 2016.

[10] Y. Yuhui et al, “Object-contextual representations for semantic segmentation,” Proc. ECCV, Glasgow, UK, pp.173–190, Aug. 2020.

[11] https://github.com/open-mmlab/mmsegmentation

以上

Realization of detection of illegal sidewalk-riding by specific small motorized bicycles using generated images for transfer learning with diffusion models and image segmentation

Primary Researcher: Tetsuya Manabe
Assistant Professor, Saitama University

In this research project, we propose and evaluate an efficient training data generation method using a diffusion model and image segmentation in order to suppress illegal sidewalk-riding by specific small motorized bicycles. Specifically, the proposed method uses a diffusion model to (1) remove vehicle objects that lead to incorrect identification and (2) supplement images of background areas, respectively, in contrast to the conventional method that generates training images by viewpoint conversion images obtained from roadway driving. In (1), the time required to generate training data per image increased compared to the conventional method, but the addition of the vehicle object image completion to the viewpoint conversion algorithm improved the identification performance of the riding environment. In (2), the training data generation time was shorter than that of the conventional method. These results contribute to the realization of detecting illegal sidewalk-riding by specific small motorized bicycles using generated images for tranfer learning with diffusion models and image segmentation.