論　文

SfM-student：SfM法を用いたデータ拡張による列車前方映像からのセマンティックセグメンテーション

振津　勇紀[†a)]　出口　大輔[†]　川西　康友[†,††]　井手　一郎[†]
村瀬　洋[†]　向嶋　宏記[†††]　長峯　望[†††]

SfM-Student: Semantic Segmentation of Train Front-View Images with SfM Data Augmentation

Yuki FURITSU[†a)], Daisuke DEGUCHI[†], Yasutomo KAWANISHI[†,††], Ichiro IDE[†], Hiroshi MURASE[†], Hiroki MUKOJIMA[†††], and Nozomi NAGAMINE[†††]

あらまし　重要な公共交通機関として広く社会に普及している鉄道の沿線には，信号機や踏切など列車の安全運行を支える多くの設備が設置されている．これらの設備の日常的な整備や設置状況などに関する情報収集業務の多くは人手により行われており，その維持管理作業の自動化・効率化を実現する技術が強く求められている．このような課題に対して，営業運転中の列車に搭載したカメラにより前方を撮影した列車前方映像のみを用い，セマンティックセグメンテーションを施すことで鉄道環境における沿線設備などを自動認識する技術に期待が寄せられている．しかし，セマンティックセグメンテーションでは画素単位でクラス情報を人手で付与した学習データが必要であり，高い性能を得るために必要な大量の学習データを用意するコストは非常に高い．そこで本論文では，教師なしデータに対するセマンティックセグメンテーション結果と Structure from Motion（SfM）法による 3 次元復元結果を組み合わせることによってラベル付き 3 次元点群を生成し，それらを画像平面に投影することで擬似的なデータ拡張を行う SfM-student 法を提案する．これにより，限られたラベルあり学習データのみからセマンティックセグメンテーションの精度向上を図る．実際の鉄道環境で撮影したデータを用いた実験を行ったところ，提案する 3 次元情報を利用した擬似的なデータ拡張手法は既存のデータ拡張手法と比べてセマンティックセグメンテーション精度を向上させることを確認した．

キーワード　セマンティックセグメンテーション，鉄道，半教師あり学習，データ拡張，SfM 法

1. まえがき

鉄道は旅客人数，速達性，信頼性などの高さから日本における重要な公共交通機関の一つであり，我々の日常生活を支える基盤となっている．鉄道における事故の防止や旅客の安全は鉄道会社の最優先課題となっており，信号機や踏切，自動列車停止装置など様々な沿線設備や技術により安全が保たれている．

このような鉄道沿線設備の日常的な整備やそれらの設置状況などに関する情報収集は，多くの鉄道会社・路線において人手により行われているのが実状である．一方，このような業務の自動化・効率化を目的として，点群計測装置を搭載した特殊な車両を使用して鉄道沿線設備の密な 3 次元点群を取得する技術が一部路線で実用化されている [1]．この技術は，Mobile Mapping System（MMS）車両を軌陸車に搭載し，線路上を走行させることで建築限界や駅ホームの形状，信号機の見通し確認などを行うことを可能にしている．しかし，専用の車両・装置の導入費用が高いことから，全ての路線での運用は現実的でない．そこで，列車の運転席前方に安価なカメラを設置することで列車前方を撮影

† 名古屋大学，名古屋市
Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464–8601 Japan
†† 理化学研究所情報統合本部ガーディアンロボットプロジェクト，京都府
RIKEN Information R&D and Strategy Headquarters GRP, Seika-cho, Soraku-gun, Kyoto-fu, 619–0288 Japan
††† 鉄道総合技術研究所，国分寺市
Railway Technical Research Institute, 2–8–38 Hikari-cho, Kokubunji-shi, 185–8540 Japan
a) E-mail: furitsuy@murase.is.i.nagoya-u.ac.jp
DOI:10.14923/transfunj.2021JAP1021

し，得られる列車前方映像から鉄道環境における詳細な物体認識を行う技術への期待が高まっている．このような技術は鉄道車両の大規模な改造や新たな地上設備の増設を必要とせず，また，乗用車におけるドライブレコーダのように運行情報を記録する役目を果たすことも可能である．このような背景から，列車前方映像から周囲に存在する物体をつぶさに認識する技術が求められている．

一方，画像中の物体を画素単位で詳細に識別する技術として，深層学習に基づくセマンティックセグメンテーションが近年注目を集めている．一般にセマンティックセグメンテーションで高い精度を得るためには，画素単位で正解が定義された大量の学習データが必要である．しかし，このような学習データの作成は人手により行われており，画像の解像度や含まれる物体の種別によっては，その作業に 1 枚あたり 1 時間以上を要することもある．そのため，認識したい環境に合わせて数千・数万枚単位の大規模なデータセットを用意することは難しい．特に鉄道を対象としたこのようなデータセットは少なく，少ないコストで学習データを大量に作成可能な技術が求められている．そこで本論文では，新たな学習データを擬似的に生成し，既存の限られた学習データに追加することで性能改善を図る手法に着目する．これまでに，フレーム間の動きベクトルを使用してデータ拡張を行う手法 [2] が提案されているが，鉄道環境においては列車の移動速度が速いため動きベクトルを用いたデータ拡張は難しい．また，学習済みセマンティックセグメンテーションモデルの推定結果を新たな教師データとして追加する手法 [3] も提案されているが，列車前方映像中の遠方領域はセグメンテーション精度が低く，教師データとして不適切であるという問題がある．以上の問題に対して，本研究では Structure from Motion（SfM）法 [4] により得られる 3 次元情報と学習済みモデルにより得られるセグメンテーション結果を組み合わせることにより，クラスラベル付き 3 次元データから擬似的に 2 次元の教師データを生成し，それを追加してセマンティックセグメンテーションのモデルを再構築する手法を提案する．これにより，鉄道環境のように十分な量の学習データを用意することが難しい対象に対して，少ないコストで高精度なセマンティックセグメンテーションモデルを構築可能な技術を実現する．

本研究の貢献は，以下のとおりである．

- SfM 法を用いた 3D-2D データ拡張手法の提案：正解ラベルが付与されていない時系列画像に対して，学習済みセマンティックセグメンテーションモデルから得られる結果と SfM 法を組み合わせることでラベル付き 3 次元点群を生成し，それを元の 2 次元画像平面に投影することで，事前に与えられたラベル付き学習データには存在しない新たなカメラ視点での擬似教師データを生成する．
- 鉄道環境における高精度なセマンティックセグメンテーションの実現：画素単位のアノテーションが付与された少数の学習データを入力とし，提案する 3D-2D データ拡張によって学習データを増強することで，高精度な鉄道環境の認識を実現する．

2. 関 連 研 究

鉄道環境の測定に Mobile Mapping System（MMS）を利用する研究 [1] では，MMS 車両を鉄道用の台車に乗せて線路上を走行することにより，鉄道沿線設備の密な 3 次元点群の取得を試みている．しかし，MMS 車両は高価であり，全ての鉄道区間への導入は非現実的である．また，密な点群の取得には台車の低速走行が必要であり，営業時間内でのデータ収集は困難である．一方，営業時間外の夜間には，物体表面のテクスチャなどの視覚的な情報を取得できないため，沿線設備の種類の判別などに必要な情報を得ることができないという問題がある．

一方，画像群から 3 次元情報を再構築する Structure from Motion（SfM）法 [4] の技術が広く研究されている．同一の対象物を異なる視点から撮影した複数枚の画像を入力することで対象物の 3 次元構造を復元する SfM 法は，高精度な自己位置推定と 3 次元構造の復元が可能な技術である．これを用いることにより，都市単位での 3 次元構造の復元 [5] や，図 1 に例示するような鉄道環境における RGB 値を保持した点群の生成も可能である．更に，この点群に画素単位のクラスラベルの推定結果を投影することで，各点がクラスラベル情報をもつような，ラベル付き 3 次元点群の生成も可能である．このような画素単位のクラスラベル推定技術として，セマンティックセグメンテーションが近年注目を集めている．Long ら [6] は，畳み込みニューラルネットワーク（Convolutional Neural Network；CNN）の最終層に全結合を採用しない，全層畳み込みネットワーク（Fully Convolutional Network；FCN）を用いたセマンティックセグメンテーション手法を提案している．FCN の登場以降，SegNet [7] や DeepLabv3+ [8]

など様々なセマンティックセグメンテーション手法が提案されている．

セマンティックセグメンテーションモデルの構築には多数の学習データが必要なことから，擬似的なデータ拡張によるデータ増強や学習手順の最適化に注目した手法も数多く提案されている．Zhu ら [2] は，正解ラベルが付与された映像中の 1 フレームを入力とし，正解ラベルが存在しない前後数フレームに擬似的なラベルを伝搬させる Joint Image-Label Propagation 法を提案している．この手法は，フレーム間の動きベクトル（Optical Flow）を用いて正解ラベルが付与されたフレームを時間方向に変形し，正解ラベルが付与されていないフレームにアノテーションを付与する．これにより，擬似的に生成した RGB 画像とその擬似正解ラベルの間の整合性を保つことができ，学習データセットの規模を 11 倍に拡張することに成功している．Chen ら [3] は，半教師あり学習の枠組みをセマンティックセグメンテーションに応用した Naïve-student 法を提案している．この手法では，人手で付与した教師ありデータで学習済みの CNN を用いて教師なしデータに擬似的にラベルを付与し，その擬似データを用いた CNN の学習（Pre-training）と人手で付与した教師ありデータでの追加学習（Fine-tuning）により精度向上が得られることを示している．

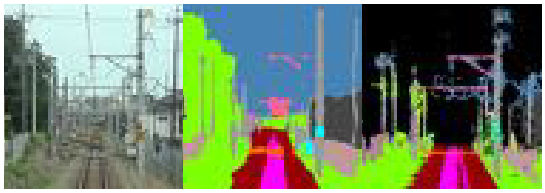
しかし，高速で移動する列車では動きベクトルによる画像変形を用いたデータ拡張は難しく，また，列車前方映像には遠くの対象も含まれることから単純に Naïve-student 法を適用するだけでは遠方で画像中に小さく写る対象への対応は難しい．

3. 提案手法：SfM-student 法

3.1 提案手法の概要

鉄道環境の多くは直線区間から構成されていることから，列車前方映像の中央部には遠方の対象が写り，それらは画像中で非常に小さいという特徴がある．そのため，事前に構築した学習済みモデルで正しい推定結果を得ることは難しく，Naïve-student 法 [3] により得られる 2D 疑似教師データに誤りが多く含まれるという問題がある（図 2 (b)）．ここで，本研究で対象とする列車前方映像においては，列車の進行に伴って遠方の対象もいつかは近くで大きく撮影される性質に着目する．近くで大きく撮影された対象であれば，事前に構築した学習済みモデルでも比較的精度良くセグメンテーション可能である．そこで列車前方映像に対して Structure from Motion（SfM）法 [4] を適用することで 3 次元再構成を行い，この 3 次元情報を活用することで，近くで撮影された対象の推定結果を遠方のカメラ視点における画像平面に投影して，新たな疑似教師データを作成する．これにより，Naïve-student 法で問題となる遠方対象の推定精度低下問題を解決する．一方，SfM 法により復元される 3 次元点群は比較的疎であることから，画像中央部と比べて画像周縁部に投影される点数は少なくなり，疑似教師データ中でクラスラベルを定めることができない領域が増える可能性がある．そこで，画像中央部においては SfM 法により生成した 3 次元情報を用いて擬似教師データを作成し，画像周縁部においては Naïve-student 法のように初期学習済みの畳み込みニューラルネットワーク（Convolutional Neural Network；CNN）の推定結果を使用する手法を提案する．本研究では，この手法を

図 1　SfM 法で出力された RGB 値をもつ 3 次元点群の例.
Fig. 1　Example of a 3D point cloud contaning RGB values, generated by the SfM method.



(a)　(b)　(c)

図 2　画像中央部の遠方物体 (a) に対し，2D 擬似教師データで不正確なラベルが生成された例 (b) と，SfM 擬似教師データで正確なラベルが生成された例 (c).
Fig. 2　Examples where for distant objects in image centers (a), inaccurate labels were generated in 2D pseudo training data (b), while accurate labels were generated in 3D pseudo training data (c).

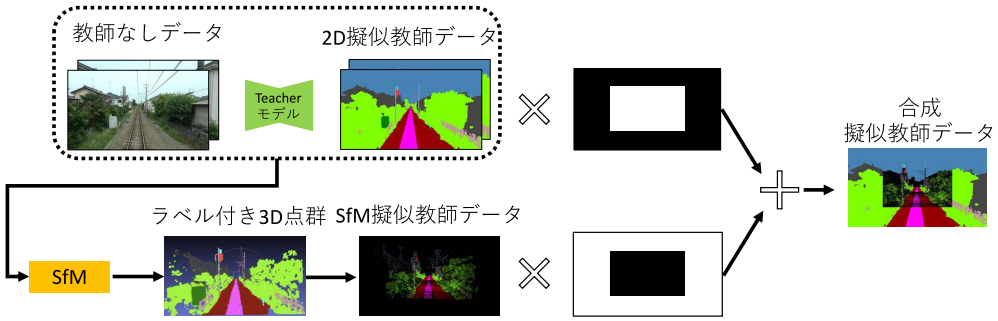


図 3　SfM-student 法による合成擬似教師データの生成手順.
Fig. 3　Generation of composite pseudo training data by the SfM-student method.

SfM-student 法と呼ぶ．SfM-student 法は，2D 擬似教師データと SfM 擬似教師データから，合成擬似教師データを生成することでデータ拡張を行う．

図 3 に提案手法における合成擬似教師データの生成手順を示す．まず，画像とその正確な画素単位でのクラスラベル（教師ありデータ）を用いて，任意の構造をもつセマンティックセグメンテーション用 CNN を学習する．これを，Teacher モデルと定義する．その後，学習済みの CNN を用いて，映像から切り出した連続フレーム画像の教師なしデータに対してセマンティックセグメンテーションを施し，擬似的なラベル画像（2D 擬似教師データ）を生成する．更に，連続フレーム画像に対して SfM 法を適用することにより 3 次元復元を行うとともに，各フレームに対応する擬似ラベルを用いてラベル付き 3 次元点群を生成する．そして，SfM 法により得られる各フレームのカメラパラメータを用いてこのラベル付き 3 次元点群を画像平面に再投影する．これにより，3 次元的な整合性を保った擬似的なラベル画像（SfM 擬似教師データ）を生成する．また，同一フレーム画像に関する 2D 擬似教師データと SfM 擬似教師データをマスク処理で統合することで，合成擬似教師データを生成する．このようにして生成した各擬似教師データを用いて任意の CNN を初期から学習し，教師ありデータを用いて Fine-tuning を行うことで新たに CNN を学習する．これを，Student モデルと定義する．これにより，セマンティックセグメンテーションの精度向上を図る．

3.2　初 期 学 習

まず，N 枚の画像 $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ と，各画像 I_N を画素単位でアノテーションしたラベル画像 y_n の組を教師ありデータとし，ラベルが付与されていない M 枚の画像 $\mathcal{J} = \{J_1, J_2, \ldots, J_M\}$ を教師なしデータとする．

教師ありデータを用いて学習したセマンティックセグメンテーションを行う CNN を関数 f とし，次式で与えられる損失関数 L を最小化するように f のパラメータ θ_t を最適化する．

$$\theta_t^* = \arg\min_{\theta_t} \frac{1}{N} \sum_{n=1}^{N} L(y_n, f(I_n; \theta_t)) \tag{1}$$

この初期学習処理により得られたモデルを Teacher モデルとする．なお，この際の損失関数 L としては，一般的にセマンティックセグメンテーションの学習で用いられる損失関数を利用することが可能であり，本論文では交差エントロピー誤差 [8] を用いる．

3.3　2D 擬似教師データの生成

Teacher モデル（**3.2** の初期学習で得られるセマンティックセグメンテーション用 CNN を表す関数 f と最適化したパラメータ θ_t^*）を用いて，教師なしデータ $\mathcal{J}$ の各画像 J_m から以下のように 2D 擬似教師データ $\widetilde{y}_{m;2D}$ を生成する．

$$\widetilde{y}_{m;2D} = f(J_m; \theta_t^*) \tag{2}$$

3.4　SfM 擬似教師データの生成

教師なしデータに対して SfM 法を適用することで復元した 3 次元点群と **3.3** までの手法で生成した 2D 擬似教師データ $\widetilde{y}_{m;2D}$ からラベル付き 3 次元点群を構築し，それを投影することで SfM 擬似教師データを生成する．

まず，単一の映像から切り出した教師なしデータ $\mathcal{J}$ に対して 2D 擬似教師データ $\widetilde{Y}_{2D} = \{\widetilde{y}_{1;2D}, \widetilde{y}_{2;2D}, \ldots, \widetilde{y}_{M;2D}\}$ を生成する．次に，SfM 法によりこの教師なしデータから復元した 3 次元点群と 2D 擬似教師データを入力とし，各教師なしデータ J_m

を撮影した際の推定カメラパラメータ P_m と，各点の 3 次元座標位置とクラスラベルをもつ K 個の点からなるラベル付き点群 $\mathcal{Z} = \{Z_k \in \mathbb{R}^3 \times \mathbb{Z} \,|\, k = 1, \ldots, K\}$ を生成する．

その後，各教師なしデータの推定カメラパラメータを用いて，以下の式でラベル付き点群を画像平面に投影し，SfM 擬似教師データ $\widetilde{y}_{m;\mathrm{SfM}}$ を生成する．

$$\widetilde{y}_{m;\mathrm{SfM}} = \pi(\mathcal{Z}; P_m) \tag{3}$$

ここで，$\pi(\cdot;\cdot)$ は SfM 法で推定したカメラの内部・外部パラメータに基づいて 3 次元点群の各点を画像平面に投影し，元のフレーム画像の範囲内に再投影された場合にその点のラベルを該当画素位置に記録する関数である．

3.5 合成擬似教師データの生成と CNN の学習

3.3 までの手法で生成した 2D 擬似教師データと **3.4** の手法で生成した SfM 擬似教師データを用いて，合成擬似教師データを生成する．

2D 擬似教師データと SfM 擬似教師データそれぞれの性質をふまえ，それらを組み合わせることで合成擬似教師データを生成する．画像中央部は遠方に対応する領域であり，SfM 擬似教師データから密な投影点が得られることから，SfM 擬似教師データを使用する．一方，画像周縁部では物体が大きく写ることから，2D 擬似教師データを使用する．具体的には，2D 擬似教師データ $\widetilde{y}_{m;\mathrm{2D}}$ と SfM 擬似教師データ $\widetilde{y}_{m;\mathrm{SfM}}$ からマスク行列 Q を用いて，以下の式で合成擬似教師データ $\widetilde{y}_{m;\mathrm{Mask}}$ を生成する．

$$\widetilde{y}_{m;\mathrm{Mask}} = Q \otimes \widetilde{y}_{m;\mathrm{2D}} + (\mathbf{1} - Q) \otimes \widetilde{y}_{m;\mathrm{SfM}} \tag{4}$$

なお，$\mathbf{1}$ は全ての要素が 1 である行列であり，$\otimes$ は Hadamard 積を計算する演算子である．

この際のマスク行列 Q の各要素 Q_ℓ は，各画素位置 $\ell \in \mathbf{L}$ について以下のように定義される．

$$Q_\ell = \begin{cases} 1 & (\ell \notin \mathbf{L}_c) \\ 0 & (\ell \in \mathbf{L}_c) \end{cases} \tag{5}$$

ここで，$\mathbf{L}_c \subset \mathbf{L}$ は，図 4 のように画像中央部に 1/4 の面積を占める領域（画像縦幅に対して 1/4 から 3/4 の範囲かつ画像横幅に対して 1/4 から 3/4 の範囲）に属する画素位置 ℓ の集合である．

式 (1) において，このように生成した合成擬似教師データ $\widetilde{y}_{m;\mathrm{Mask}}$ を I_n に置き換え，Student モデルに対応する CNN のパラメータ θ_s を θ_t と同様に学習する．

最後に，擬似教師データで学習した Student モデルに対して，教師ありデータを用いて式 (1) と同様に Fine-tuning を行う．

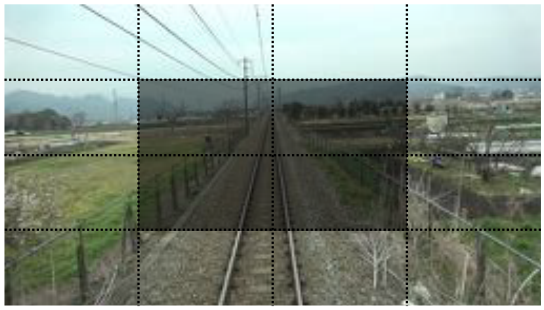


図 4 画像中央部に 1/4 の面積を占める領域．点線はそれぞれ縦横を 4 分割しており，網かけ領域を $\mathbf{L}_c$ と定義する．

Fig. 4 Region that occupies 1/4 of the overall area at the image center. The dotted lines split the height and width into four, and the shaded region is defined as $\mathbf{L}_c$.

4. 評価実験

擬似教師データを用いてセマンティックセグメンテーションの半教師あり学習を行う提案手法の有効性を確認するために，評価実験を行った．

まず **4.1** で，本実験の概要について述べた後，**4.2** で，使用したデータセットの詳細について述べる．そして **4.3** で，使用した評価指標について述べ，最後に **4.4** で評価結果について報告する．

4.1 実験の概要

本実験では，提案した合成擬似教師データを用いて，列車前方映像のセマンティックセグメンテーションを行った．

また比較手法として，教師データのみを用いて学習する通常のセマンティックセグメンテーション手法と，2D 擬似教師データを用いた Naïve-student 法 [3] の簡易実装を用意し，提案手法と精度比較を行った．

- 比較手法 1：教師ありデータのみで学習した DeepLabv3+ [8]
- 比較手法 2：2D 擬似教師データを用いて学習を行った DeepLabv3+（Naïve-student 法の簡易実装）
- 提案手法 1：SfM 擬似教師データを用いて学習を行った DeepLabv3+
- 提案手法 2：合成擬似教師データを用いて学習を行った DeepLabv3+（SfM-student 法）

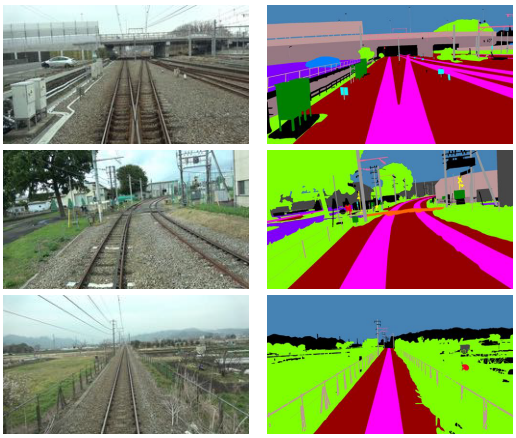


図 5　列車前方映像データセットに含まれる元画像と正解ラベル画像の例.

Fig. 5　Examples of original and annotated images included in the train front-view dataset.

各手法で畳み込みニューラルネットワーク（Convolutional Neural Network；CNN）の学習を行う際には，あらかじめ路上環境データセットで事前学習したのちに，列車前方映像データセットでの学習を行った.

4.2　データセット

本実験のために，実環境において列車の先頭車両に搭載した市販のビデオカメラ（SONY FDR-AX55）から進行方向の前方をフル HD 画質（1,920×1,080 画素），30 fps で撮影し，画素単位のクラスラベルを人手で付与した列車前方映像データセットを構築した．本実験では，撮影した映像から抽出した全 315 枚の列車前方の画像とその正解ラベル画像（Train：265 枚，Test：50 枚）を用意した．このデータセットに含まれる画像・正解ラベル画像の例を図 5 に示す．また，データセットに含まれるクラスラベルの内，鉄道環境に関連するラベルと色の詳細を図 6 に示す.

擬似教師データを生成するための教師なしデータとして，映像から 5 フレームごとに抽出した画像を用いた．また，SfM 擬似教師データを生成する際には，長い区間を Structure from Motion（SfM）法 [4] で再構築することによる処理時間の増大・誤差の蓄積を防ぐため，抽出した画像から時系列順に 20 枚ずつ選択し，それらについて推定されたラベル画像群とともに SfM 法に入力してラベル付き点群を生成した.

また，各手法で用いる CNN を事前学習するために，以下の路上環境データセットを用いた.

- Mapillary Vistas [9]：世界各国で車載カメラから前方を撮影し，画素単位でアノテーションを付与した 25,000 枚（Train：18,000 枚，Validation：2,000 枚，Test：5,000 枚）のアノテーション付き画像から構成されるデータセット.
- Cityscapes [10]：ドイツの複数の都市で車載カメラから前方を撮影し，画素単位でアノテーションを付与した 5,000 枚（Train：2,975 枚，Validation：500 枚，Test：1,525 枚）のアノテーション付き画像から構成されるデータセット.

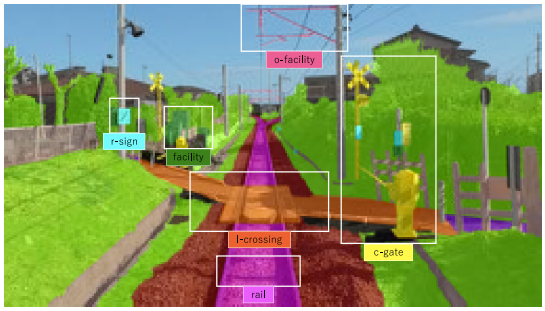


図 6　列車前方映像データセットで定義した鉄道に関連するクラスラベルと対応色.

Fig. 6　Class labels and corresponding colors defined in the train front-view dataset.

4.3　評 価 指 標

各手法の評価指標について説明する．列車前方映像データセットの Test セットにおけるセマンティックセグメンテーションの精度を評価するため，推定画像と正解ラベル画像の全ての対応する画素位置 $\ell \in \mathbf{L}$ に対して，クラス k ごとに以下のように推定画像のある画素位置におけるクラスラベルと正解ラベル画像の同画像位置におけるクラスラベルを分類し，該当個数を記録する.

- True Positive（TP_k）：正解ラベルと推定結果の双方のクラスラベルが k である画素数.
- False Positive（FP_k）：推定結果のクラスラベルが c であり，正解のクラスラベルが k でない画素数.
- False Negative（FN_k）：正解のクラスラベルが c であり，推定結果のクラスラベルが k でない画素数.
- True Negative（TN_k）：正解ラベルと推定結果の双方のクラスラベルが k でない画素数.

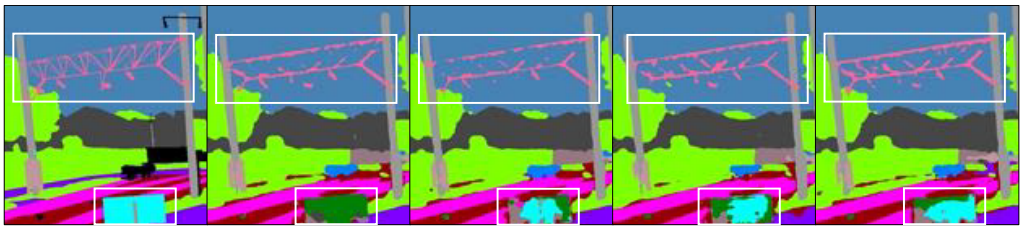
この分類結果を用いて，以下のようにクラス k ごとの Intersection over Union（cIoU_k）を計算する.

$$\mathrm{cIoU}_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FP}_k + \mathrm{FN}_k} \tag{6}$$

これをもとに，Test セットにおいて真値に含まれる

表 1　各手法における mIoU 及びに cIoU の評価結果．Rail, l-crossing, o-facility, r-sign の各ラベルの具体例については図 6 を参照．
Table 1　MIoU and cIoU of the evaluation results of each method. Refer to Fig. 6 for actual examples of labels rail, l-crossing, and r-sign.

手法	擬似教師データ	mIoU ↑	cIoU ↑				
			human	rail	l-crossing	o-facility	r-sign
比較手法 1	—	0.660	0.733	0.937	0.774	0.548	0.674
比較手法 2	2D	0.673	0.735	0.939	0.765	0.562	0.670
提案手法 1	SfM	0.668	0.719	**0.942**	0.759	0.555	0.658
提案手法 2	合成	**0.681**	**0.744**	0.934	**0.792**	**0.571**	**0.701**



(a) 真値ラベル画像　(b) 比較手法 1　(c) 比較手法 2　(d) 提案手法 1　(e) 提案手法 2

図 7　同一の入力画像のある領域における，真値ラベル画像と各手法の推定結果．手法間に特徴的な違いが現れた箇所を白線で囲んだ．
Fig. 7　Ground-truth label image and inference results of each method for a region of the same input image. White boxed regions show characteristic differences between each methods' output.

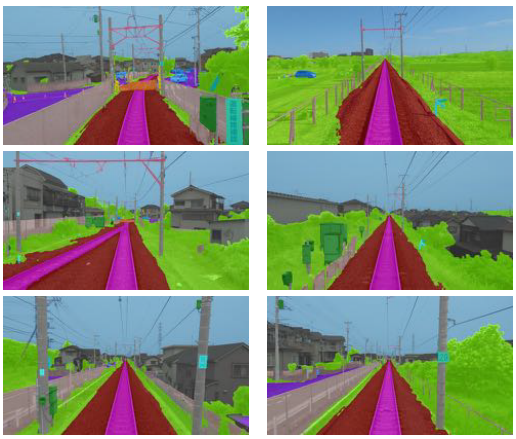


図 8　提案手法の推定結果を元画像に重ねた例．
Fig. 8　Examples of inference results overlaid on the original images.

クラスラベル集合 K に含まれる全クラスの IoU の平均値として，以下の式から mIoU を計算する．

$$\mathrm{mIoU} = \frac{1}{|K|} \sum_{k \in K} \mathrm{cIoU}_k \tag{7}$$

なお，本実験では各手法について 10 回ずつ学習・評価を行い，性能の限界を検証する目的から 10 回の試行から最も mIoU が高い結果を記録した．

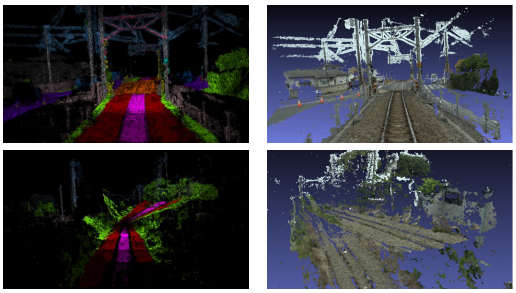
4.4　評価結果

4.1 で挙げた各手法について，**4.2** で用意したデータセットと，**4.3** で挙げた評価指標を用いて精度評価を行った．

表 1 に各手法の mIoU 及びに鉄道環境に関連するクラスの cIoU の評価結果を示す．また，図 7 に各手法における推定結果の例を，図 8 に提案手法による推定結果を元画像に重ね合わせた例を示す．

5.　考　　察

表 1 から，合成擬似教師データを用いて半教師あり学習を行った提案手法 2 が最も高い mIoU を示したことがわかる．また，多くのクラスの IoU に向上が見られ，特に鉄道標識（r-sign）や頭上施設（o-facility）など本研究で主に識別したい沿線設備の識別精度にも改善がみられた．図 7 に示すように，実際の識別結果においても，細かな頭上設備や比較的小さな鉄道標識について，提案手法 2 が最も良くセグメンテーションできる傾向が見られた．このように，画像の周縁・中央の領域において正確なクラスラベルをもつ合成擬似教師データを畳み込みニューラルネットワーク（Convolutional Neural Network；CNN）の学習に用いることで，限られた学習データから高精度なセマンティックセグメンテーションモデルを構築できること



(a) SfM 擬似教師データ　(b) 作成に使用した SfM 点群

図 9　SfM 擬似教師データと，その作成に使用した点群の例.

Fig. 9　Examples of SfM pseudo training data and corresponding base point clouds.

を確認した．また，本実験では鉄道環境におけるセマンティックセグメンテーションを対象として評価したが，Structure from Motion（SfM）[4] による 3 次元復元結果を用いて対象をインスタンス単位に分離することにより，提案手法の考え方をインスタンスセグメンテーションに対して拡張することもできると考えられる．

一方，本来高精度な 3 次元情報を用いて学習した提案手法 1 において，データ拡張を行わない比較手法 1 よりも一部の cIoU が低くなった．これは，SfM 擬似教師データの生成方法に起因したと考えられる．本実験においては，前述のとおり一つの映像から 5 フレームずつ画像を切り出し，抽出した画像から時系列順に 20 枚ごとに SfM 法による 3 次元再構築を行った．この方法では列車の走行速度・実際の移動距離を考慮しておらず，駅近辺では画像群中の列車の移動距離が短く，駅間では列車の移動距離が長くなっている．これにより，SfM 法により復元される点群の密度・精度が画像群ごとに異なってしまい，最終的な CNN の精度にも影響を与えたと考えられる．一方，図 9 にあるように，SfM 点群を正しく復元できた場合（上段）には，適切な疑似教師データを得られることが確認できるが，SfM 点群の復元に失敗した場合（下段）には，得られる疑似教師データも正しくないことが分かる．このことから，SfM 点群の復元精度が提案手法の性能に影響を与えることが分かる．これに対して，近年注目を集めている DeepMVS [11] や MVSNet [12] のような深層学習による SfM 法の研究が進むことで近い将来改善できるようになると考えられる．

6. むすび

列車の安全な運行を支える鉄道沿線設備の維持管理作業の効率化のために，安価な列車前方映像のみから鉄道環境の認識を正確に行う技術への期待が高まっている．これに対して，Structure from Motion（SfM）法 [4] による 3 次元再構築を用いることで，画像中の遠方領域でも高精度な擬似ラベルを生成し，学習データの増強によりセマンティックセグメンテーションの精度向上を図る SfM-student 法を提案した．実験により，3 次元情報をもつ擬似教師データを畳み込みニューラルネットワーク（Convolutional Neural Network；CNN）の学習に用いることで，限られた学習データから高精度に鉄道環境が認識できることを確認した．

今後の課題として，カメラの移動距離に応じた SfM 擬似教師データの生成間隔の決定や，鉄道環境以外への応用が考えられる．

謝辞　本研究の一部は JSPS 科研費（JP17H00745）の助成を受けたものである．

文　献

[1] Y. Niina, E. Oketani, H. Yokouchi, R. Honma, K. Tsuji, and K. Kondo, “Monitoring of railway structures by MMS,” J. Japan Society of Photogrammetry and Remote Sensing, vol.55, no.2, pp.95–99, Jan. 2016. DOI:10.4287/jsprs.55.95

[2] Y. Zhu, K. Sapra, F.A. Reda, K.J. Shih, S. Newsam, A. Tao, and B. Catanzaro, “Improving semantic segmentation via video propagation and label relaxation,” Proc. 2019 IEEE Conf. on Computer Vision and Pattern Recognition, pp.8856–8865, Long Beach, CA, USA, June 2019. DOI:10.1109/CVPR.2019.00906

[3] L.-C. Chen, R. Lopes, B. Cheng, and M. Collins, “Naive-student: Leveraging semi-supervised learning in video sequences for urban scene segmentation,” Proc. 16th European Conf. on Computer Vision, Part IX, pp.695–714, Aug. 2020, Online. DOI:10.1007/978-3-030-58545-7_40

[4] J.L. Schonberger and J.-M. Frahm, “Structure-from-motion revisited,” Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp.4104–4113, Las Vegas, NV, USA, June 2016. DOI:10.1109/CVPR.2016.445

[5] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, and R. Szeliski, “Building Rome in a day,” Proc. 12th IEEE Int. Conf. Computer Vision, pp.72–79, Kyoto, Japan, Sept. 2009. DOI:10.1109/ICCV.2009.5459148

[6] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition, pp.3431–3440, Boston, MA, USA, June 2015. DOI:10.1109/CVPR.2015.7298965

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.39, no.12, pp.2481–2495, Dec. 2017. DOI:10.1109/

TPAMI.2016.2644615
[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” Proc. 15th European Conf. on Computer Vision, Part VII, pp.833–851, Munich, Germany, Sept. 2018. DOI: 10.1007/978-3-030-01234-2_49
[9] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, “The Mapillary Vistas dataset for semantic understanding of street scenes,” Proc. 16th IEEE Int. Conf. Computer Vision, pp.4990–4999, Venice, Italy, Oct. 2017. DOI:10.1109/ICCV.2017.534
[10] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes dataset for semantic urban scene understanding,” Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition, pp.3213–3223, Las Vegas, NV, USA, June 2016. DOI:10.1109/CVPR.2016.350
[11] P.-H. Huang, K. Matzen, J. Kopf, N. Ahuja, and J.-B. Huang, “DeepMVS: Learning multi-view stereopsis,” Proc. 2018 IEEE Conf. Computer Vision and Pattern Recognition, pp.2821–2830, Salt Lake City, UT, USA, June 2018. DOI:10.1109/CVPR.2018.00298
[12] Y. Yao, Z. Luo, S. Li, T. Fang, and L. Quan, “MVSNet: Depth inference for unstructured multi-view stereo,” Proc. 15th European Conf. Computer Vision, Part VIII, pp.767–783, Munich, Germany, Sept. 2018. DOI:10.1007/978-3-030-01237-3_47

（2021 年 4 月 20 日受付，9 月 4 日再受付，
2022 年 1 月 7 日早期公開）

振津　勇紀　（正員）

令 1 名大・工・情報卒．令 3 同大大学院情報学研究科修了．修士（情報学）．同年よりソニーグループ株式会社に勤務．セマンティックセグメンテーションや 3 次元再構築に関する研究に従事．

出口　大輔　（正員）

平 13 名大・工・情報卒．平 18 同大大学院情報科学研究科博士後期課程了．博士（情報科学）．平 16–18 日本学術振興会特別研究員．平 18 名大大学院情報科学研究科研究員，同大大学院工学研究科研究員，平 20 同大大学院情報科学研究科助教，平 24 同大情報連携統括本部情報戦略室准教授，令 2 より同大大学院情報学研究科准教授，現在に至る．主に画像処理・パターン認識技術の開発とその ITS 及び医用応用に関する研究に従事．CARS2004 Poster Award，CADM2004 大会賞，平 18 日本医用画像工学会奨励賞，平 18 日本コンピュータ外科学会講演論文賞．IEEE 会員．

川西　康友　（正員）

平 18 京大・工・情報卒．平 24 同大大学院情報学研究科博士後期課程了．博士（情報学）．平 24 同大学術情報メディアセンター特定研究員．平 26 名大未来社会創造機構特任助教．平 27 同大情報科学研究科助教．平 29 同大情報学研究科助教．令 2 間講師．令 3 理化学研究所ガーディアンロボットプロジェクト感覚データ認識研究チームチームリーダー．現在に至る．ロボットによる周囲環境認識及び，人物追跡・属性認識・行動認識などの人物画像処理に関する研究に従事．平 23 年度 PRMU 研究奨励賞受賞．IEEE ITS Society Nagoya Chapter Young Researcher Award 受賞．IEEE 会員．

井手　一郎　（正員：シニア会員）

平 6 東大・工・電子卒．平 8 同大大学院工学系研究科情報工学専攻修士課程了．平 12 同研究科電気工学専攻博士課程了．博士（工学）．同年国立情報学研究所助手．平 16 名古屋大学大学院情報科学研究科助教授，平 19 より准教授．平 29 同大大学院情報学研究科准教授，令 2 より同大数理・データ科学教育研究センター教授，現在に至る．この間，平 14–16 総合研究大学院大数物科学研究科助手併任，平 16–22 情報・システム研究機構国立情報学研究所客員助教授・准教授兼任，平 17，18，19 フランス情報学・統計システム研究所（IRISA）招聘教授．平 22–23 オランダアムステルダム大情報学研究所上級訪問研究員．パターン認識技術の実応用や映像メディア処理全般に興味をもっている．情報処理学会，IEEE 各シニア会員，映像情報メディア学会，人工知能学会，ACM 各会員．

村瀬　洋　（正員：フェロー）

昭 53 名大・工・電気卒．昭 55 同大大学院修士課程了．同年日本電信電話公社（現 NTT）入社．平 4 から 1 年間米国コロンビア大客員研究員．平 15 名古屋大学大学院情報科学研究科教授．平 29 同大大学院情報学研究科教授．令 3 より同大名誉教授，特任教授，現在に至る．文字・図形認識，コンピュータビジョン，マルチメディア認識の研究に従事．工博．昭 60 本会学術奨励賞，平 6 IEEE-CVPR 最優秀論文賞，平 7 情報処理学会山下記念研究賞，平 8 IEEE-ICRA 最優秀ビデオ賞，平 13 高柳記念奨励賞，平 13 本会ソサエティ論文賞，平 14 本会業績賞，平 15 文部科学大臣賞，平 16 IEEE Trans. MM 論文賞，平 22 前島密賞，平成 24 紫綬褒章，他受賞．IEEE，情報処理学会各フェロー．

向嶋　宏記

平 27 名大・工・情報卒．平 29 同大大学院情報科学研究科了．修士（情報科学）．同年（公財）鉄道総合技術研究所入所．現在，同研究所信号・情報技術研究部画像・IT 研究室研究員．列車への車載カメラ映像を対象とした前方監視・物体検出，画像によるメンテナンスの補助，画像処理技術の鉄道への応用に関する研究開発に従事．電気学会会員．

長峯　　望

平 16 筑波大大学院理工学研究科理工学専攻了．同年（財）鉄道総合技術研究所入所．現在，（公財）鉄道総合技術研究所信号・情報技術研究部画像・IT 研究室主任研究員．画像処理技術の鉄道への応用，列車前方監視，状態監視，画像によるメンテナンスの自動化などに関する研究開発に従事．電気学会，情報処理学会各会員．博士（工学）．