ソフト順序制約付きラベル境界緩和法に基づく列車前方映像のセマンティックセグメンテーション

振津 勇紀[1,a)]　出口 大輔[1]　川西 康友[1]　井手 一郎[1]　村瀬 洋[1]　向嶋 宏記[2]　長峯 望[2]

概要

鉄道における沿線設備の位置・種類の把握や管理のためには，鉄道環境の詳細な情報が必要不可欠である．これに対して本発表では，連続フレーム画像に対する正解ラベルの伝播によるデータ拡張の際に発生する物体境界領域の歪みに対応した境界緩和手法を用いた，列車前方映像に対するセマンティックセグメンテーション手法を提案する．また，生成した画素単位のラベルと Structure from Motion による 3 次元再構築による，鉄道環境の正確なラベル付き 3 次元地図の自動構築手法についても述べる．

1. はじめに

鉄道は旅客人数，運搬速度，信頼性，などの高さから日本の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．このようなことから，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，様々な技術がその支援に利用されている．一方，架線柱や電源設備などの鉄道沿線設備は日々整備が必要であり，それらの位置・大きさ・種類などの情報の収集と管理は人手による日常的な点検作業を通して行なわれている．これらの作業コストは非常に高く，鉄道沿線設備の情報を自動的に収集し，それらを整備に活用できる技術の開発が期待されている．

鉄道環境の 3 次元的な分析調査として，道路用の Mobile Mapping System（MMS）を鉄道に利用した研究がある [7]．この研究では，実際に MMS 車両を鉄道用の台車に乗せて走行することにより，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 車両は非常に高価で，また，低速での走行が必要であるため営業時間内でのデータ収集はできない．このことから，営業時間外の夜間にデータ収集した場合は物体のテクスチャなどの視覚的な情報が取得できず，沿線設備の整備に必要な情報を得ることはできない．

一方，画素単位でクラス分類を行う 2 次元セマンティックセグメンテーションが近年盛んに研究されており，鉄道環境理解への応用が期待されている．Fully Convolutional Network（FCN）[5] が登場して以降，畳み込みニューラルネットワーク（CNN）を利用した手法が広く用いられており，近年注目を集めている SegNet [1] や Deeplabv3+ [2] などの手法が数多く提案されている．

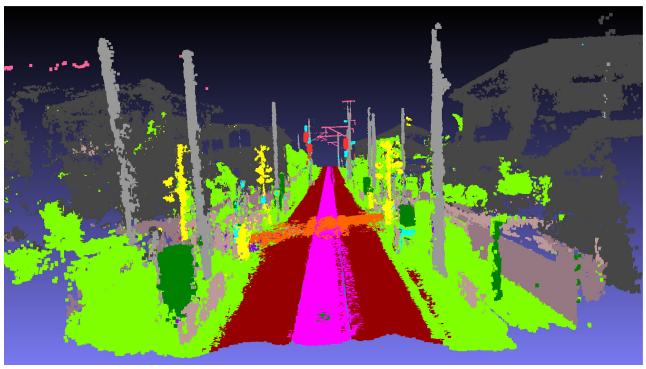


図 1　提案手法により生成した鉄道環境のラベル付き 3 次元マップ．

セマンティックセグメンテーションの性能向上において，大規模なデータセットの構築は重要な課題である．一般に，人手による正解ラベル画像作成（画素単位でのアノテーション付与）は画像 1 枚あたり 1 時間以上必要 [3] であり，データベース構築コストの低減が求められている．また，現在広く公開されているデータセットは物体単位 [4] のもの，路上環境 [3], [6] を対象としたものが多く，鉄道環境のセグメンテーションに直接利用することはできない．

上述の課題に対して，Zhu ら [8] は画像・ラベル共伝播（Joint Image-Label Propagation）を用いて限られたアノテーションデータを連続フレームに伝播することで擬似的に学習データを増強し，ラベル境界緩和法（Boundary Label Relaxation）によって擬似データのラベル境界歪みに対応する手法を提案している．画像・ラベル共伝播では，映像中のある画像とその正解ラベル画像を入力とし，密なオプティカルフローを用いて隣接するフレームへと画像及び正解ラベル画像を変形する．この変形の際，複数の画素が同じ位置に対応付く可能性がある．その場合，複数ラベルが同一画素に対応付くことから，ラベルに曖昧性が生じるという問題が発生する．このようなクラスの曖昧性を考

[1] 名古屋大学
[2] 鉄道総合技術研究所
[a)] furitsuy@murase.is.i.nagoya-u.ac.jp

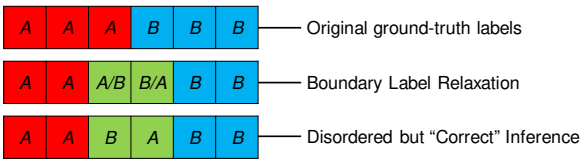


図 2 1 次元でのラベル境界緩和法の例．クラス境界の正解ラベルは A/B のいずれでも正しいとして扱われる．

慮して学習する手法がラベル境界緩和法である．彼らはこの手法を用いて，Cityscapes データセット [3] のラベル付きデータを 11 倍の規模に拡大し，高精度なモデルの学習に成功している．しかし，以下の問題は解決できていない：

(i) ラベル伝播を正確に行なうためには，フレーム間のカメラの動きが小さい必要がある．一般に鉄道の運行速度は速いため，列車前方映像が高フレームレートで撮影されたとしてもフレーム間のカメラの動きは大きい．そのため，変形後の 1 画素が対応付く元画像の範囲が広くなり，結果としてラベルの曖昧性が増加する．しかし，Zhu らの手法は境界から距離 1 の範囲の画素しか考慮していない．

(ii) ラベル境界緩和法では，ある画素が複数の正解ラベルを持つことを許すように，単純に真値を拡張している．しかし，ラベル順序が間違っている場合も正しいものと扱われてしまう（図 2）．

列車前方映像を対象とした画像・ラベル共伝播を考えた場合，ラベル境界の歪みは距離 1 画素以上の範囲で発生すると考えられる．そこで本発表では，より広い範囲を対象としたソフト順序制約付きラベル境界緩和法を提案する．具体的には，Zhu らの手法のような単純なラベル境界緩和では境界におけるラベル順序の不整合問題を解消できないことから，ラベル境界からの距離に応じて損失を与えるように境界を緩和する新たな損失関数を導入する．そして，鉄道環境のように大規模な学習データセットの準備が難しい対象に対しても高精度なセマンティックセグメンテーション手法を実現する．

更に，列車前方映像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）を組み合わせることにより，図 1 に示すような鉄道環境のラベル付き 3 次元マップを構築する手法についても述べる．

2. 提案手法

2.1 提案手法の概要

画像・ラベル共伝播法を用いて列車前方映像データを拡張した場合，車速の影響によりラベル境界の変形量と生じる歪みは大きく，変形後の 1 画素が対応付く元画像の範囲が広くなる．従来のラベル境界緩和法ではこのような複数画素にわたる歪みに対応できず，また学習時にラベル順序の整合性を保つこともできない．本発表では，ラベル境界での歪みに対して柔軟に対応可能であり，学習時にラベル順序の整合性を維持できる手法を提案する．

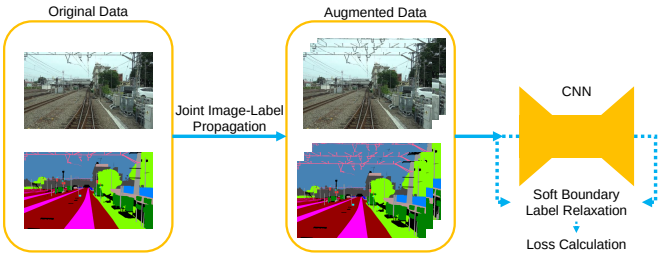


図 3 提案手法の処理手順．

図 3 に提案手法の処理手順を示す．まず画像・ラベル共伝播を用いて学習データの規模を拡大する．次に，提案するソフト順序制約付きラベル境界緩和法を用いて学習時の損失を計算し，これを逆伝播することでセグメンテーション用ネットワークを学習する．

2.2 ソフト順序制約付きラベル境界緩和法

ソフト順序制約付きラベル境界緩和法は，セマンティックセグメンテーションの学習時に行なうクラス境界の損失計算において，隣接するクラスの順序を極力保ちつつクラス境界の曖昧性の影響低減を図る技術である．以下でその詳細を述べる．

まず，入力画像の各画素に対応する教師信号が $\mathbf{t}(\mathbf{x})$（各次元がそれぞれのクラスに対応するベクトル）として与えられるとする．ここで，$\mathbf{g}(\mathbf{x})$ を正解ラベルの one-hot ベクトル（正解ラベルの次元に対応する値のみ 1 で残りはすべて 0）とすると，$\mathbf{t}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$ の場合は各画素 $\mathbf{x}$ の推定結果 $\mathbf{f}(\mathbf{x})$ を用いて cross-entropy loss は次式で計算される．

$$L = -\mathbf{t}(\mathbf{x})^{\mathrm{T}} \log(\mathbf{f}(\mathbf{x})) \tag{1}$$

次に，$d(\mathbf{x}, \mathbf{y})$ を 2 画素間の距離を与える関数とすると，画素 $\mathbf{x}$ から距離 D 以内に存在する画素集合は，

$$\mathcal{N}(\mathbf{x}) = \{\mathbf{y} \mid d(\mathbf{x}, \mathbf{y}) < D\} \tag{2}$$

と表すことができる．ここで，式 (1) の $\mathbf{t}(\mathbf{x})$ を次式で与えられる multi-hot ベクトルに置き換えることで，境界から距離 D の範囲へとラベル境界緩和法を拡張する．

$$\mathbf{t}(\mathbf{x}) = \min\left\{\mathbf{1}, \sum_{\tilde{\mathbf{x}} \in \mathcal{N}(\mathbf{x})} \mathbf{g}(\tilde{\mathbf{x}})\right\} \tag{3}$$

しかし式 (2) の単純な拡張のみでは，D の増加と共にラベル境界の不整合（ラベルの順序が保たれない）が顕著になる．以下，説明の簡単化のため，図 4 のようなクラス A と B の境界を例に，この問題を解決する手法を説明する．

まず，クラス境界からの距離に従い不正解クラスの尤度が下がるという仮定に基づき，図 4 の画素 $\mathbf{x}_2$ のクラス尤度について以下が成り立つとする．

$$P(A \,|\, \mathbf{x}_2) > P(B \,|\, \mathbf{x}_2) \tag{4}$$

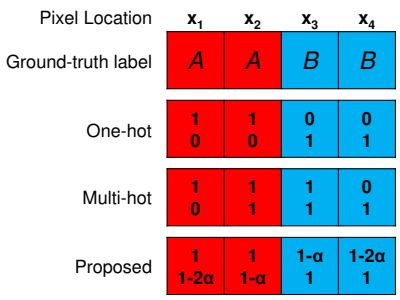


図 4 1 次元でのラベル空間の例．画素インデックスとその位置での正解ラベル，各手法における $\mathbf{t}(\mathbf{x})$ の表現方法を示す．

また画素 $\mathbf{x}_1$ の真値も A であり，かつ，クラス B との境界は画素 $\mathbf{x}_2$ より遠いことから，理想的には，

$$P(A\,|\,\mathbf{x}_1) - P(B\,|\,\mathbf{x}_1) > P(A\,|\,\mathbf{x}_2) - P(B\,|\,\mathbf{x}_2). \quad (5)$$

も成り立つと仮定できる．ここで，クラス A と B を表す one-hot ベクトル表現を $\mathbf{a}$ と $\mathbf{b}$ とすると，$P(A\,|\,\mathbf{x}_2) = \mathbf{t}(\mathbf{x}_2)^{\mathrm{T}}\mathbf{a}$ と書けることから，式 (5) は次式で表すことができる．

$$\mathbf{t}(\mathbf{x}_2)^{\mathrm{T}}\mathbf{b} - \mathbf{t}(\mathbf{x}_1)^{\mathrm{T}}\mathbf{b} > \mathbf{t}(\mathbf{x}_2)^{\mathrm{T}}\mathbf{a} - \mathbf{t}(\mathbf{x}_1)^{\mathrm{T}}\mathbf{a}. \quad (6)$$

ここで，$\mathbf{x}_1$ と $\mathbf{x}_2$ の真値は共にクラス A であるため，$\mathbf{t}(\mathbf{x}_2)^{\mathrm{T}}\mathbf{a}$ と $\mathbf{t}(\mathbf{x}_1)^{\mathrm{T}}\mathbf{a}$ の値は等しくなるはずである．そこで両者の値を 1 と置き，左辺の差が定数 α であると仮定すると，

$$\mathbf{t}(\mathbf{x}_2)^{\mathrm{T}}\mathbf{b} - \mathbf{t}(\mathbf{x}_1)^{\mathrm{T}}\mathbf{b} = \alpha > 0. \quad (7)$$

これは，クラス境界からの距離に比例して α ずつクラス B の尤度が下がることを表す．

以上をふまえ，式 (2) をソフト制約付きラベル境界緩和法へと拡張する．まず，画素 $\mathbf{x}$ がクラス λ であることに対する重みを与える関数を，以下のように定義する．

$$h(\mathbf{x}, \lambda) = \max_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} \left\{ \big(1 - \alpha \times d(\mathbf{x}, \mathbf{y})\big)\big(\mathbf{g}(\mathbf{y})^{\mathrm{T}}\lambda\big) \right\} \quad (8)$$

ここで，λ はクラス λ に対応する one-hot ベクトル表現である．全てのクラスの one-hot ベクトル表現の集合を Λ とすると，$\mathbf{t}(\mathbf{x})$ は次式により求められる．

$$\mathbf{t}(\mathbf{x}) = \sum_{\lambda\in\Lambda} \lambda \cdot h(\mathbf{x}, \lambda). \quad (9)$$

図 4 の最下段は，提案手法の $\mathbf{t}(\mathbf{x})$ の表現を示している．式 (9) により得られる $\mathbf{t}(\mathbf{x})$ を式 (1) に代入することにより，クラス境界からの距離に応じて緩やかにペナルティが生じるように cross-entropy loss を変更することができる．

3. 評価実験

列車の先頭車両に設置したカメラを用いて前方を撮影した動画像を撮影し，その一部フレームに対して画素単位でアノテーションを付与したデータセットを用いて実験を行なった．構築したデータセットは全 116 枚の列車前方画像から構成され，これを学習用 66 枚と評価用 50 枚に分割した．また，学習用データは，事前に文献 [8] の画像・ラベル共伝播法により 3 倍に規模を拡張した．CNN の事前学習には Cityscapes と Mapillary Vistas [6] の二つのデータセットを用いた．

CNN の構造としては，DeepLabv3+ [2] を使用した．学習時のエポック数は 100，クロップサイズを 896×896 とし，式 (8) の $\alpha = 0.2$ とした．その他のパラメータに関しては文献 [8] に従った．

以下の 4 つの手法について比較をした．

- **ベースライン手法（DL）**
  DeepLabv3+を画像・ラベル共伝播後の拡張データセットで学習する手法．ラベル境界緩和法は用いない．
- **既存手法（BLR-1）**
  DL に領域幅 1 画素のラベル境界緩和法を追加した手法 [8].
- **提案手法（BLR-N）**
  DL に領域幅 N 画素（$N > 1$）のラベル境界緩和法を追加した手法.
- **提案手法（SBLR-N）**
  BLR-N にラベル順序の整合性を考慮した損失関数を導入し，領域幅を N 画素（$N \geq 1$）に拡張したラベル境界緩和法を用いた手法.

評価指標として，mean Intersection over Union（mIoU）と class Intersection over Union（class IoU）を用いた．各手法を 10 回ずつ学習・評価した平均を表 1 に示す．また，各手法の出力結果例を図 5 に示す．

表 1 各手法の mIoU と鉄道関連クラスの class IoU.

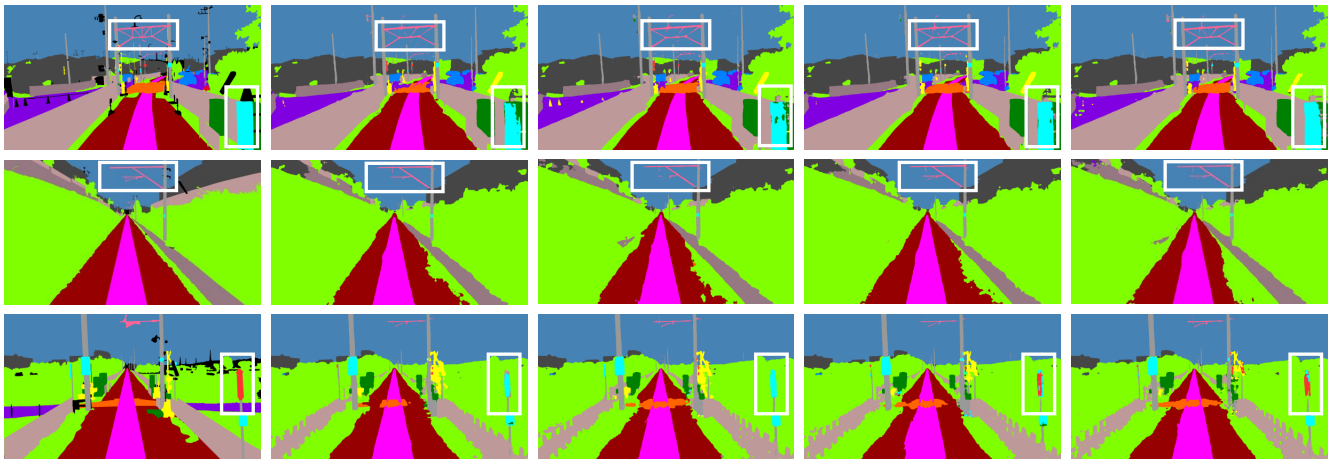
手法	mIoU	人	軌条	頭上施設
ベースライン（DL）[2]	0.581	0.617	0.880	0.504
既存（BLR-1）[8]	0.587	0.610	0.880	0.504
提案（BLR-2）	0.599	0.613	0.880	0.504
提案（BLR-3）	0.596	0.615	0.880	0.504
提案（SBLR-1）	0.591	0.620	**0.881**	0.506
提案（SBLR-2）	0.599	0.629	**0.881**	0.507
提案（SBLR-3）	**0.604**	**0.637**	**0.881**	**0.508**

4. 考察

4.1 実験結果の分析

表 1 より，提案手法は比較手法と比較して高精度に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なえることを確認した．また，DeepLabv3+に対する Zhu らの手法 [8] の mIoU の改善は約 0.6% であるのに対して，クラス境界の緩和幅を単純に拡張する提案手法（BLR-2）で約 1.2%，ソフト順序制約付きラベル境界緩和を用いた提案手法（SBLR-3）により約 2.2% の mIoU の向上を確認した．

クラス毎の結果に着目すると，提案手法では人や頭上施設といった比較的小さく細長い物体の検出精度の改善が顕



(a) 真値　(b) ベースライン（DL）　(c) 既存（BLR-1）　(d) 提案（BLR-3）　(e) 提案（SBLR-3）

図 5　各手法の出力例と真値との比較.

著であった．一般的なセマンティックセグメンテーション手法では，特に学習初期段階においてクラス境界の誤りが損失関数の計算に大きく影響する．そのため，特に検出が難しい小さな物体などは，従来の損失関数ではうまく扱うことができない．これに対して，提案するソフト順序制約付きラベル境界緩和法では，クラスラベルの空間的な整合性を保ったままクラス境界に近い誤分類の影響を軽減することができる．これにより，小さく細い物体の分類精度の向上に繋がったと考える．

4.2　鉄道環境理解への応用

列車前方映像に対して SfM を適用することにより，鉄道環境の 3 次元再構築を行なえる．SfM でマッチングした各フレームにおける特徴点に，その画素位置での 2 次元ラベルの情報を統合することで，図 1 のようなセグメンテーションラベル付き 3 次元地図の生成を可能とした．この応用例では，詳細な 3 次元情報の復元に，用意することが難しい 3 次元点群の教師ラベルが不要であるのが利点である．計算資源の関係上，現時点では図 1 に示すような約 100 m 相当の短い区間のみ生成可能であるが，今後，効率的な計算方法及びデータの格納方法を模索して，より長い区間の生成に挑戦したい．

5.　まとめ

本発表では，鉄道における沿線設備の位置・種類の把握や管理の効率化を目的として，ソフト制約付きラベル境界緩和を導入した列車前方映像のセマンティックセグメンテーション手法を提案し，従来のラベル境界緩和法よりも性能向上が得られることを確認した．また，列車前方映像に対する SfM の結果と提案手法による 2 次元セマンティックセグメンテーション結果を統合することにより，鉄道環境の 3 次元地図が構築可能であることを示した．

今後の課題として，SfM による 3 次元再構築精度の向上や学習データの規模の拡大方法の改良などが挙げられる．

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．

参考文献

[1] Badrinarayanan, V., Kendall, A. and Cipolla, R.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intelli.*, Vol. 39, No. 12, pp. 2481–2495 (2017).

[2] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation, *Proc. 15th European Conf. Comput. Vis. (Part VII)*, pp. 833–851 (2018).

[3] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B.: The Cityscapes dataset for semantic urban scene understanding, *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3213–3223 (2016).

[4] Everingham, M., Van Gool, L., Williams, C. K., Winn, J. and Zisserman, A.: The PASCAL visual object classes (VOC) challenge, *Int. J. Comput. Vis.*, Vol. 88, No. 2, pp. 303–338 (2010).

[5] Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation, *Proc. 2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3431–3440 (2015).

[6] Neuhold, G., Ollmann, T., Rota Bulo, S. and Kontschieder, P.: The Mapillary Vistas dataset for semantic understanding of street scenes, *Proc. 2017 IEEE Int. Conf. Comput. Vis.*, pp. 4990–4999 (2017).

[7] Niina, Y., Oketani, E., Yokouchi, H., Honma, R., Tsuji, K. and Kondo, K.: Monitoring of railway structures by MMS, *J. Jpn. Soc. Photogramm. Remote Sens.*, Vol. 55, No. 2, pp. 95–99 (2016).

[8] Zhu, Y., Sapra, K., Reda, F. A., Shih, K. J., Newsam, S., Tao, A. and Catanzaro, B.: Improving semantic segmentation via video propagation and label relaxation, *Proc. 2019 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 8856–8865 (2019).