

1. はじめに

セマンティックセグメンテーションは、ピクセル単位でクラスを識別する問題であり、走行領域の把握等の自動運転支援において重要である。車載映像を対象とした大規模なデータセットが公開され、多くの手法が提案されている[1][2]。これらのデータセットは前方のみを対象としており、車載環境における一部の状況しか理解することができない。自動運転支援には、左右と後方を含めた 360 度全周囲を対象とする必要がある。そこで本研究では、360 度全周囲に対応した情報伝搬ネットワークを提案する。提案手法では、時間方向と空間方向に情報を伝搬することで、時間的な変化や空間的な変化を捉えた識別が可能となる。また、画像間の整合性を保つことで画像の端領域におけるセグメンテーション精度の向上を図る。

2. セマンティックセグメンテーション

代表的なセマンティックセグメンテーション手法として Fully Convolutional Network (FCN)[1] が提案されている。FCN は、畳み込み層とプーリング層で構成されたネットワークであり、全結合層をフィルタサイズ 1×1 の畳み込み層に置き換えることで、異なる入力サイズに対応できる。また、CNN と RNN を組み合わせた Encoder-Decoder 構造のセグメンテーション手法として、DAG-RNN [2] が提案されている。Encoder で得られた特徴マップを DAG-RNN に入力し、周辺の画素の情報を考慮した特徴マップを生成する。生成された特徴マップを Decoder に入力し、セグメンテーション結果を得る。

これらの手法は、一般的な車載シーンデータセットで高い精度を達成しているが、歪みのある全周囲画像を対象とした場合、同一物体でもバリエーションが増加しセグメンテーションが困難となる問題がある。また、全周囲画像を分割して学習すると、画像の端付近では周囲のコンテキスト情報を獲得できず、セグメンテーション精度が低下する問題もある。

3. 提案手法

本研究では、物体の見え方の変化に対応するため、時間方向と空間方向に情報を伝搬させる情報伝搬ネットワークを提案する。提案手法のネットワーク構造を図 1 に示す。情報伝搬ネットワークは、Encoder-Decoder 構造のネットワークである。時間方向の情報伝搬を導入することで、前方のカメラで撮影された物体が左右、後方のカメラで撮影されるまでの時間による変化を捉えることが可能となる。空間方向の情報伝搬には、DAG-RNN with LSTM (DAG-LSTM) を用いる。DAG-LSTM は、広範囲の画素情報を考慮した特徴マップを生成できる。また、画像端のセグメンテーション精度の低下を防ぐため、Image Consistency loss を導入し、画像間の整合性を保つようにする。

3.1 時間方向の情報伝搬

時間方向に情報を伝搬させるため、Encoder と Decoder の間に RNN を導入する。現時刻を t とした時、 $t-4$ から t までの 5 フレーム間の情報を RNN により伝搬させる。伝搬させる情報には、Encoder で出力した特徴を使用する。時間方向に情報を伝搬させることで、画像中の物体の動きや自車や他車の進行方向といった時間による変化を学習することが可能となる。

3.2 空間方向の情報伝搬

空間方向に情報を伝搬させるため、Encoder と Decoder の間に DAG-LSTM を導入する。従来の DAG-RNN は、Elman 型の RNN を利用している。RNN は数時刻分程度の入力しか出力に反映されず、長期的な記憶ができないという問題がある。本研究では、RNN の代替として LSTM を導入した DAG-LSTM を提案する。DAG-LSTM を導入することで、短期・長期的な情報を記憶することができ、広範囲の空間情報を考慮した特徴マップを生成することが

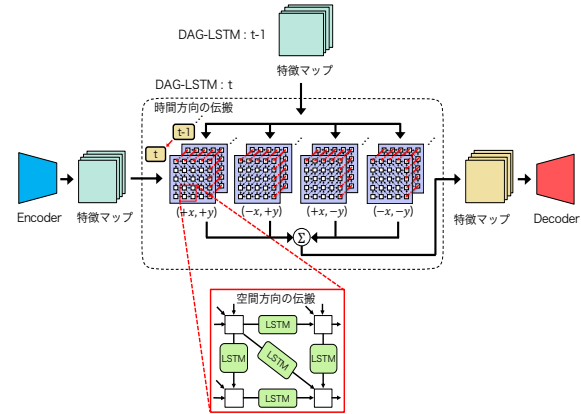


図 1: 情報伝搬ネットワークの構造

できる。これにより、DAG-RNN と比べて情報伝搬の効果を高めることができる。

時間方向における空間変化を考慮し情報伝搬の効果を高めるため、Dilated Convolution を導入する。Dilated Convolution は一定間隔離れた要素を畳み込むことで、特徴マップのサイズを縮小せずに受容野を拡大できる。これにより、大局的な特徴を捉えることができる。したがって、自車や他車が動いた場合でも周辺の情報を広範囲に捉えることが可能である。

3.3 画像間の整合性

全周囲画像は、前方・後方・左右のカメラで撮影された画像を連結している。このとき、各カメラで撮影された画像の端部分におけるセグメンテーション精度が低下する問題がある。そこで図 2 のように、画像間で整合性を保つようにする。はじめに、全周囲画像を分割する。分割する際、分割後の画像に隣接するカメラで撮影された画像の端が含まれるようにする。各画像間の端領域において、隣接する画像の端領域に近づくように学習をする Image Consistency loss を導入する。Image Consistency loss には、Jensen-Shannon divergence (JSD) を用いる。Image Consistency loss は式 (1) のように表すことができる。

$$L = L_{SCE} + JSD(S_{CL} || S_{L1}) + JSD(S_{CR} || S_{R1}) \quad (1)$$

ここで、 L_{SCE} は各ラベルとの Softmax Cross Entropy, $S_{CL}, S_{CR}, S_{L1}, S_{R1}$ は各画像の端領域である。

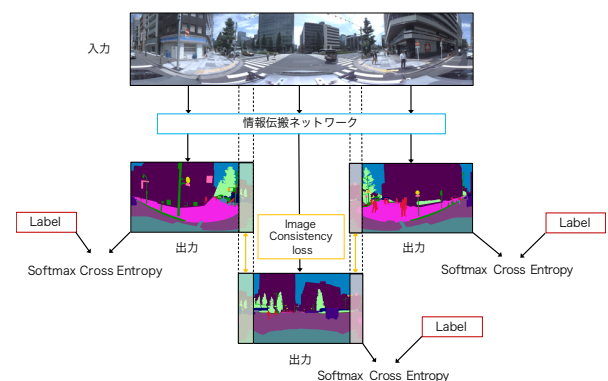


図 2: 画像間の整合性の導入

4. 評価実験

提案手法の有効性を示すために、車載カメラで撮影された全周囲画像を対象としたセマンティックセグメンテーション用のデータセットを用いて、評価実験を行う。

4.1 実験概要

評価データに用いるデータセットは、車両の周囲を 4 つのカメラで撮影した全周囲画像から構成される。本データ

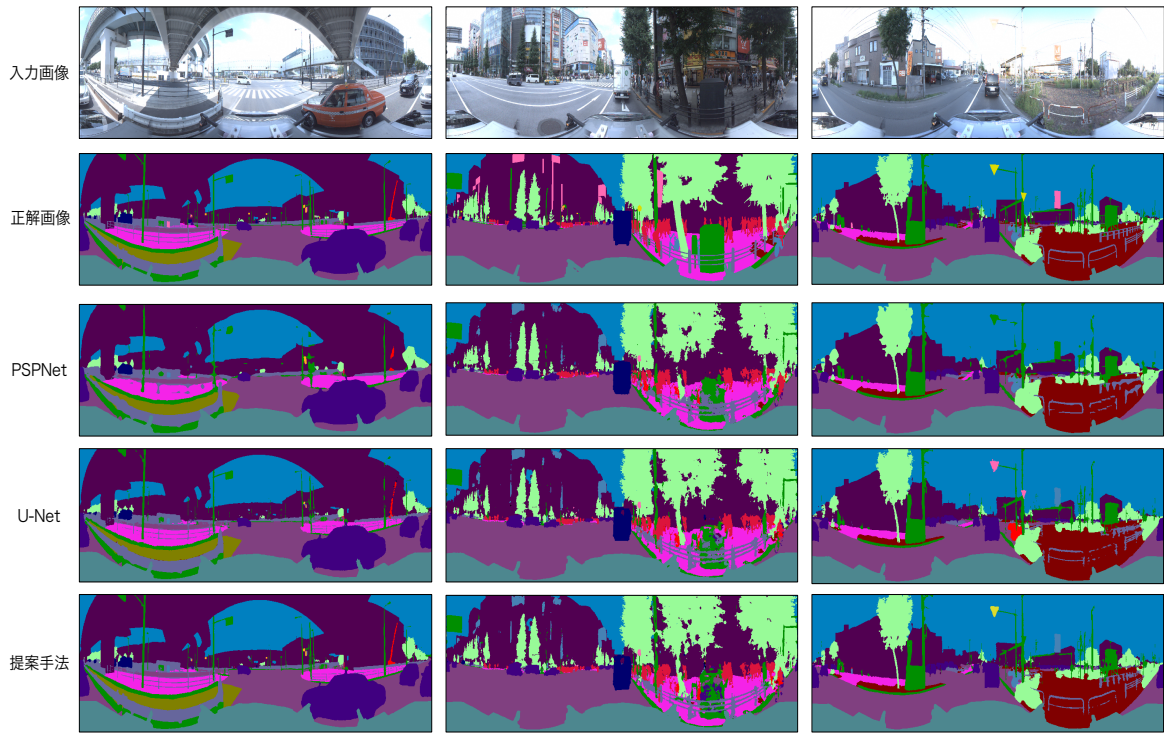


図 3：360 度データセットにおけるセグメンテーション結果例

表 1：従来手法との比較結果 [%]

手法	全体精度	クラス精度	平均 IoU
FCN	92.5	58.5	52.1
SegNet	92.4	55.2	48.9
U-Net	93.8	60.2	54.4
PSPNet	93.5	61.6	55.1
提案手法	94.3	62.4	56.2

表 2：情報伝搬処理の有無による比較結果 [%]

情報伝搬処理			全体精度	クラス精度	平均 IoU
時間	空間	画像			
✓			92.4	60.3	52.8
	✓		92.0	60.9	54.9
✓	✓		93.8	61.2	55.1
✓	✓	✓	94.3	62.4	56.2

セットは日本国内の複数地域で撮影されており、クラス数は 19 クラスである。また、画像サイズは、 5171×1160 である。本実験では、画像サイズを 2880×640 にリサイズして使用する。学習には 1,466 枚、評価用には 104 枚を用いる。学習の際は幾何変換、ノイズ付加等の Data Augmentation を行い、バリエーションを増加させる。

4.2 実験結果

定量的評価指標として、全体精度、クラス精度、平均 IoU を用いる。従来手法と提案手法の比較結果を表 1 に示す。表 1 より、提案手法は、情報伝搬処理の導入により、従来手法より精度が向上していることが分かる。また、情報伝搬処理の有無による識別精度の比較を表 2 に示す。表 2 より、いずれの伝搬処理も精度向上に寄与しているといえる。

図 3 に各手法のセグメンテーション結果を示す。図 3 より、道路や空といった大きなクラスのセグメンテーションができていることが分かる。また、従来手法と比較して提案手法は、クレーン車や柵といった小さな領域のクラスに対しても高精度なセグメンテーションができていることが分かる。画像間の整合性の有無による比較結果を図 4 に示す。図 4 より、画像間の整合性を保つことで、車や歩道クラスのセグメンテーションが改善している。

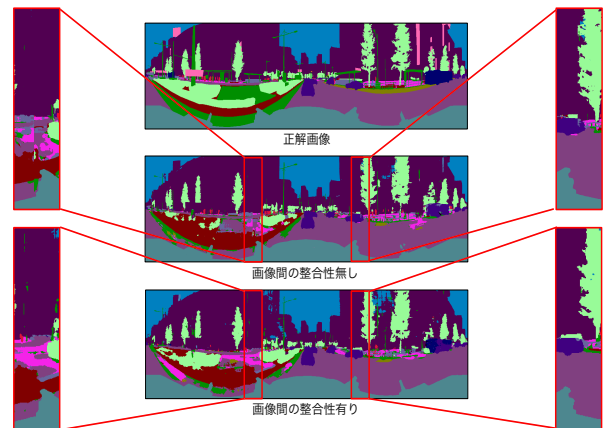


図 4：画像間の整合性の有無による比較結果

5. おわりに

本研究では、時間方向及び空間方向に情報を伝搬させる情報伝搬ネットワークを提案した。情報を伝搬させることで、360 度全周囲の画像において高精度なセマンティックセグメンテーションを実現した。また、画像間の整合性を保つことで、画像端のセグメンテーション精度を改善した。今後は、標識や信号機等の識別が困難なクラスの精度向上を検討する。

参考文献

- [1] J. Long, *et al.*, “Fully convolutional networks for semantic segmentation”, CVPR, 2015.
- [2] B. Shuai, *et al.*, “DAG-Recurrent Neural Networks For Scene Labeling”, CVPR, 2015.
- [3] F. Yu, *et al.*, “Multi-scale context aggregation by dilated convolutions”, ICLR, 2016.

研究業績

- [1] 後藤圭汰 等, “情報伝搬ネットワークによる全周囲のセマンティックセグメンテーション”, 画像センシングシンポジウム, 2018.
- [2] 後藤圭汰 等, “カメラ間の整合性を考慮した全周囲画像のセグメンテーション”, ビジョン技術の実利用ワークショップ, 2019.