H-054

シーンから得た多重解像度特徴群を attention 機構で選択的に統合する semantic segmentation
Semantic segmentation that selectively integrates multi-resolution features obtained from a scene with an attention mechanism

金子 周平[1)]　荒井 秀一[1)]
Shuhei Kaneko　Shuichi Arai

1　はじめに

セマンティックセグメンテーション [1], [2] は, 画像内の各ピクセルを特定のクラスに分類する重要なタスクであり, 自動運転や医療画像解析など多くの分野で利用が期待されている. 近年, ディープラーニングの発展により, この分野における精度は飛躍的に向上した. しかし, 小さなオブジェクトのセグメンテーションや, 多クラスの大きなオブジェクトの認識においては依然として課題が残っている.

我々の以前の研究 [3] では, High Resolution Network (HRNet)[4] に Squeeze-and-Attention（SA）ブロック [5] を導入し, 解像度間の動的な特徴選択と統合を行うことで, セグメンテーションの精度を向上させる手法を提案した. 本研究では, 前回の研究を基に, SA ブロックの導入方法をさらに進化させ, 3 種類の SA ブロックの導入手法を提案する. これらの手法は, それぞれ異なる方法で特徴を統合し, セグメンテーションの精度をさらに向上させる.

2　従来手法

画像セグメンテーションにおいて, ディープラーニングに基づく方法は多くの進展を遂げている. 特に, 畳み込みニューラルネットワーク（CNN）[?] を用いた方法は, その高い特徴抽出能力により広く採用されている. HRNet V2 は, 高解像度の特徴をネットワーク内で最後まで維持し, 複数解像度の並列畳み込みと情報交換を繰り返すことによりセグメンテーション性能の向上を目指している. 図 1 は, HRNet V2 のの構造を示している. これは, 各解像度で畳み込まれた特徴マップを Exchange ユニットを使用して統合し, 複数の解像度間で情報を交換するものである.

しかし, HRNet V2 の情報交換を担う Exchange ユニットは, 異なる解像度の特徴を統合する際に, アップサンプリングやダウンサンプリングの操作によりサイズを揃えたマップを単純に加算することでそれぞれの特徴がぼやけてしまい, 高解像度特徴マップの利点を十分に活用できないという欠点がある.

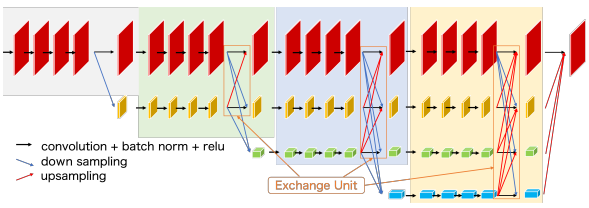


図 1: HRNet V2

3　提案手法

我々の以前の研究では, HRNet V2 を基盤として, SA ブロックを導入することで, 解像度間の動的な特徴選択と統合を可能にする手法を提案した.

SA ブロックは, 解像度間の特徴選択と統合を動的に行う. 具体的には, 各解像度の特徴を抽出し, それらを一旦圧縮した後, 畳み込みによって空間方向とチャンネル方向の両方に重要な特徴を強調する. これにより, 従来の単純なアップサンプリングやダウンサンプリングでは得られなかった識別精度をの向上を達成する. 本研究では, HRNet V2 を基礎とし, SA ブロックを統合する三つの異なる手法を提案する. これらの手法は, それぞれ異なる方法で特徴を統合し, セグメンテーションの精度を向上させることを目指している. master: 出力解像度の特徴に焦点を当て, 加算する他の解像度の特徴を重み付けする. slave: 加算前に他の解像度からの特徴自信を重み付けする. dual: 二つの解像度からの特徴を結合したマップによって重み付けする.

Exchange Unit は複数解像度の特徴マップの統合を各解像度ごとに行う. 例えば 4 つの解像度の特徴マップを統合して 1/2 解像度の特徴マップを出力する場合, 1/1 解像度の特徴マップは 3x3 ストライド畳み込みによるダウンサンプリング, 1/4 と 1/8 解像度の特徴マップは 1x1 畳込みとバイリニア補完によるアップサンプリングでそれぞれの解像度, チャンネル数を 1/2 解像度のマップサイズに揃え, 4 つを加算する.

3.1　master

master では, 出力解像度の特徴に焦点を当て, 他の解像度の特徴マップに向けそれぞれ重みを生成する. これにより, 加算前に出力解像度の情報から持ち込む情報を選択できる.

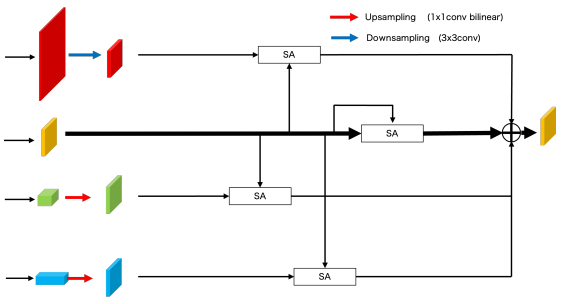


図 2: master 手法

3.2　slave

slave では, 統合前に各解像度の特徴マップを自分自身で重み付けする. これにより, 異なる解像度間での情報交換が促進される.

1)　東京都市大学 Tokyo City University

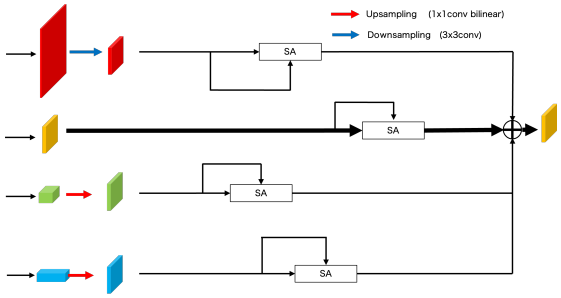


図 3: slave 手法

3.3 dual

dual では, 二つの解像度からの特徴を結合したマップによって重み付けする. これにより, 各解像度の利点を活用しつつ, 出力先の解像度に合わせた情報を取捨選択する.

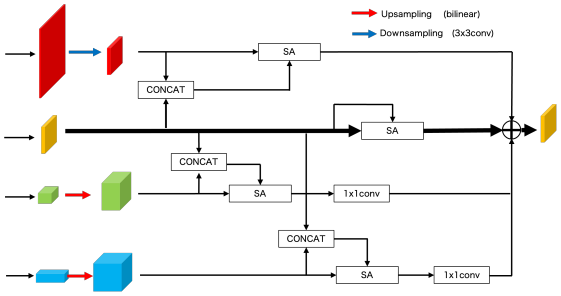


図 4: dual 手法

4 実験及び実験結果

4.1 実験方法

提案手法の有効性を検証するために, Cityscapes データセット [6] を用いて実験を行った. Cityscapes データセットは, 都市環境におけるセマンティックセグメンテーションの評価に広く用いられているデータセットであり, 評価指標として mIoU（mean Intersection over Union）を使用した.

4.2 実験結果

表 1 は, 提案手法が従来の方法よりも優れた性能を示している. 特に, dual 手法を用いた場合, mIoU スコアで 4.86 ポイントの向上が見られた.

表 1: mIoU による比較

Method	mIoU (%)
HRNet V2	70.25
Master	73.35 (+3.1)
Slave	72.28 (+2.03)
Dual	**75.11 (+4.86)**

図 5 は各クラスの IoU の比較である. すべてのクラスで提案手法が従来手法を上回っていることがわかる.

図 6 のセグメンテーション結果の比較例であり、画像のうち, (c) は HRNet の結果, (d) は我々の手法の結果を示している. HRNet では正しくセグメントできない画像中央の車とトラックの境界が, 我々の手法では正しくセグメントされている. これらの結果から, 提案手法がセマンティックセグメンテーションの精度を向上させる有効な手段であることが確認された.

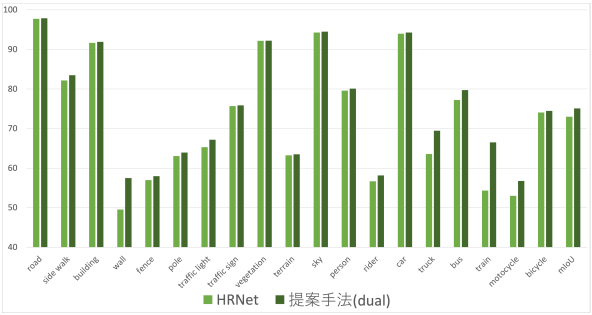
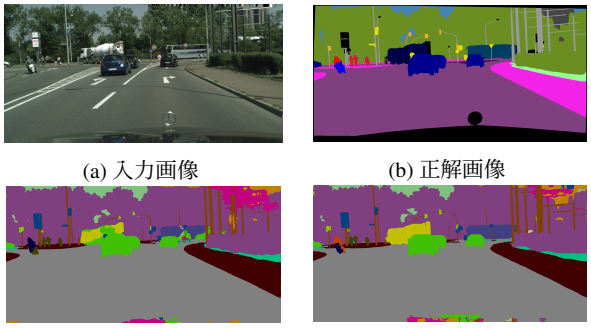


図 5: 各クラスの IoU の比較



(a) 入力画像　(b) 正解画像

(c) HRNet　(d) 提案手法 (dual)

図 6: セグメンテーション画像

5 おわりに

本研究では, HRNet V2 を基礎とし, SA ブロックを統合した新しいセマンティックセグメンテーション手法を提案した. 提案手法は, 異なる解像度間の動的な特徴選択と統合を可能にし, セグメンテーション精度を向上させることを目指している. Cityscapes データセットでの実験結果から, 提案手法が従来の方法よりも優れた性能を示していることが確認された.

今後の研究では, さらに大規模なデータセットを用いた評価や, 他のセグメンテーションタスクへの応用を検討する.

参考文献

[1] J. Shotton et al., “Semantic texton forests for image categorization and segmentation,” in Proc. Computer Vision and Pattern Recognition (CVPR 2008), 2008, pp. 1–8.

[2] J. Shotton et al., “Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context,” International Journal of Computer Vision, pp. 2–23, 2009.

[3] S. Kaneko et al., “Hrnet integrating features across multiple resolutions using an attention mechanism for semantic segmentation,” in Proceedings of the 2023 International Symposium on Nonlinear Theory and Its Applications (NOLTA), 2023, pp. 521–524.

[4] J. Wang et al., “Deep high-resolution representation learning for visual recognition,” TPAMI, 2019.

[5] Z. Zhong et al., “Squeeze-and-attention networks for semantic segmentation,” 2020.

[6] M. Cordts et al., “The cityscapes dataset for semantic urban scene understanding,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.