4N-02

車載カメラ画像の分類精度向上検討

森　光輝†
東京電機大学大学院　理工学研究科

築地　立家‡
東京電機大学大学院　理工学研究科

1．はじめに

セマンティックセグメンテーション[1]は，画像内に含まれる画素全てに対してラベルやカテゴリを割り当てる深層学習のアルゴリズムである．画像単体の分類とは異なり，画像中に含まれる意味的領域を複数分類する事ができる．セマンティックセグメンテーションは自動運転分野で中心的な技術として研究されており，車載カメラから撮影された画像にディープニューラルネットワークを用いて特徴抽出し，セマンティックセグメンテーションを行う方法などがある．

「A large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes」[3]では，３D の郊外モデルを作成し，郊外でみられる様々なシーンを切り取り，大量のセマンティックセグメンテーション用のデータセットとして用いている．

この研究では，切り取った大量のシーンと既存の実世界のデータセットを組み合わせて学習させることが，実世界だけのデータセットで学習させるよりも，分類精度が高くなることが分かっている．

2．研究の背景と目的

2.1　背景

セマンティックセグメンテーション用ニューラルネットワークの分類精度を上げるためには，画像データと，画像データに適切な前処理をしたピクセルラベルデータを組み合わせたデータセットが大量に必要である．そのため，画像データに対して闇雲に前処理したデータセットを使用しても分類精度が良くならない事に加え，作成時間コストも余計にかかってしまう．

この問題の対策として，目的に応じた画像データを選択し，適切な前処理を行ったデータセットを用いる必要がある．

2.2　目的

本研究では，国内の車載カメラから撮影した映像データに対してセマンティックセグメンテーションを行い，その映像データの推定ラベルの分類精度向上を目的とした．

3．分類精度向上の検討方法

本研究では，ディープニューラルネットワークの実装及びラベル推定，分類精度の比較は MATLAB を用いて行った．

国内映像データに対応したニューラルネットワークを学習させる場合，適当な国内の車載カメラから撮影した画像データとそのピクセルラベルデータを用いて学習させる事が本目的に対して優位的なのか不明確である．加えて，現時点でそのような国内データセットが無い．そこで，オリジナルデータセットとしてドライブレコーダーから切り取った画像データからデータセットを作製した．また，既存のイギリス国内データセットである CamVid，ドイツ国内データセットの Cityscapes で学習したニューラルネットワークとの分類精度の比較をし，国内画像データの特徴を調べた．加えて，その特徴を基にそれぞれのニューラルネットワークで分類精度の高かった方のラベルピクセルを組み合わせた混合データセットを用いることで分類精度向上検証を行った．

4．実験内容

本研究ではセマンティックセグメンテーションで使用するアーキテクチャに DeepLab v3+[2]を使用した．また，学習用データセットはオリジナルと CamVid，Cityscapes 共に 200 枚使用し，評価用データには 30 枚の正解ラベル付き画像データを用いて分類精度比較を行った．さらに，その評価を参考に各データセットのうち 2 つの中から評価の高かった方のピクセルクラスを選び，新しく混合データセットとして 3 つデータセットを作製し，同様に評価を行った．

4．1　DeepLab v3+

DeepLabv3+は PSPNet[5]による空間ピラミッドプーリングモジュール及びエンコーダー・デコーダー構造の利点を組み合わせて作成されたセマンティックセグメンテーション用のアーキテクチャである．空間ピラミッドプーリングによって異なる解像度の特徴マップをプーリングし，大規模な特徴を捉えた特徴マップと小さな特徴を捉えた特徴マップを生成する事で，複雑なセグメント可能になる．

4．2　推定ラベル

本研究で推定するラベルを，Sky，Building，Pole，Pavement，Tree，Sign Symbol，Fence，Car，Pedestrian，Bicyclist とした．Sign Symbol に関しては，道路交通標識及び信号機を含むものとし，Fence に関してはガードレール及び柵を含むものとした．

† Tokyo Denki University graduate school　Science-and-engineering graduate course

‡ Tokyo Denki University graduate school　Science-and-engineering graduate course

4．3　評価指標

セマンティックセグメンテーションの評価尺度として，mean Intersection Over Union(mIOU)を用いた．mIOU は判定されるクラスをn，正解ラベルiのピクセル集合をA_i，ラベルiを推定したピクセル集合をB_iとすると，

$$mIOU = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i \cap B_i}{A_i \cup B_i}$$

で表される．

5．実験結果

オリジナル，CamVid，Cityscapes データセット単体で学習した時の評価データによる IoU 値及び，混合データセットで学習した時の評価データによる IoU 値を比較した．それらをまとめたグラフ図 1 を以下に示す．

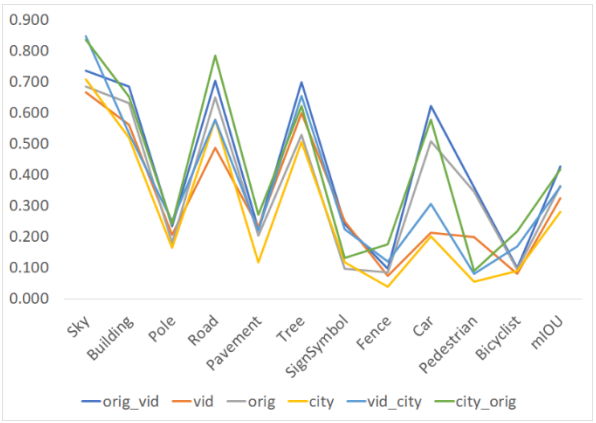

図 1　全データセットの IoU 値

また，それぞれのデータセットで学習させたネットワークで評価用データをセグメンテーションした時の分類画像図 2 を以下に示す．

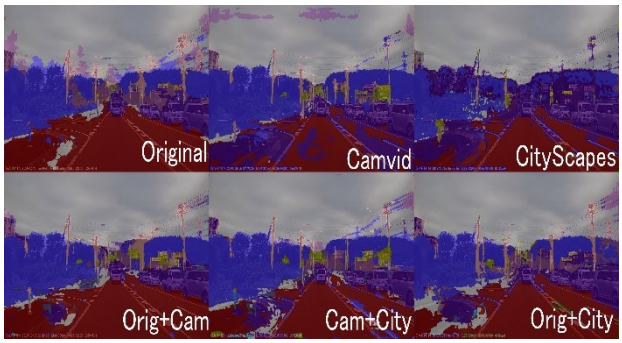

図　2　各ネットワークの評価データセグメンテーション結果

6．考察

本研究では，学習用と評価用のオリジナルデータをそれぞれ 200 枚，30 枚と作製した．3 つのデータセットを学習させ，評価データで評価してみると，それぞれのデータセットで異なる特徴を持っている事が分かった．オリジナルと CamVid の学習後の分類精度を見てみると，オリジナルデータセットに分類時における優位的なクラスは Sky，Road，Car，Pedestrian，Bicyclist であったことが分かった．CamVid では Car クラスの分類精度が著しく低くい事が分かる．これについては，検証用にデータが 30 枚程度しかなかったため，CamVid の分類ネットワークの一般性能を検証するに十分にデータが足りていないと考えられる．しかし，オリジナルと Cityscapes の分類精度の比較を見てみると，ここでも Cityscapes の Car クラスの分類精度が著しく低く表れている．このことから，オリジナルデータセット中の Car クラスにおいては特に優位性が高いことが分かった．同様に Building に関しても同じような傾向が見られた．これに関しては，建築様式などの違いにより，国内と国外データセットに差が出たのではないかと考えられる．混合データセットに関しては，混合前と比べると，全体的に精度が上がっていることが確認できた．理由として，お互いのデータセット中の分類精度の高かったピクセルクラスを使用した事により，クラス単体の精度が上がり，クラス間の分類精度が相対的に上がったのではないかと考えた．本研究では，学習用に作製したデータが 200 枚程度であるため，実際に自動運転運用システムに組み込むネットワークとしては不十分な学習データ数である．今後の展望として，少ないデータセットでも分類強度として遜色のない精度を持つようなアーキテクチャの構築ができるかどうかを検討していきたい．

7．結論

国内の車載カメラからの映像のクラス分類の精度を特定のクラスのみであれば，向上させることができた．

参考文献

[1] Krizhevsky,A.,Sutskever,I.and Hinton,G.E.:Imagnet classification with deep convolutional neural networks,Proceedings of Advance in Neural Information Processing Systems,pp.1097-1105(2012).

[2] L.Chieh Chen,Y.Xhu,G.Papandreou,F.Schroff,H.Adam
In 2018 IEEE Conference on Computer Vision and Pattern Recognition(ECCV) Aug 2018.

[3] German.Ros, Laura.Sellart, Joanna.Materzynska, David.Vazquez, Antonio M,Lopez.:A large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes, In 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).

[4] Cordts,M., Omran, M., Ramos, S., Rshfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B.: The cityscapes dataset for semantic urban scene understanding, Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3213-3223(2016).

[5] H.Zhao,J.Shi,X.Qi,X.Wang,and J.Jia. Pyramid scene parsing network. In 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),p.6230-6239,July 2017.