

卒業論文概要書

Summary of Bachelor's Thesis

Date of submission: 02/02/2021 (MM/DD/YYYY)

学科名 Department	情報通信	氏名 Name	小谷晟経	指導 教員 Advisor	渡辺 裕 印
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	1W153050-0 ^{CD}		
研究題目 Title	照明変化に対して頑健なセマンティックセグメンテーション Robust Semantic Segmentation against Lighting Changes				

1. まえがき

近年、自動車やドローンをはじめとした機械技術が急速に進展する中で、自動運転に関する研究が注目を集めている。自動運転システムの実用化において、周辺環境の把握と対象物の判定を行う必要がある。従って、高精度で安定した画像認識技術が重要となる。このように、画像認識技術の一手法であるセマンティックセグメンテーションは、周辺環境の正確な把握と対象物のクラス予測の双方に対応可能である。一方、入力データを取得するセンサの性能や特性に大きく影響を受ける課題点がある。現在主流となっているセンサにはカメラが挙げられる。カメラは対象のテクスチャや色を取得可能な一方で、照明変化に対して脆弱であるという欠点がある。そこで注目されているのが、LiDAR (Lighting Detection and Ranging) である。LiDAR は短波長のレーザ光を用いたセンサであり、周辺環境のマッピングや対象物の材質の判定等に使用可能である。このセンサは、照明変化の影響を受けないという特性がある一方で、遠方の対象物の観測が困難であり、ガラスや鏡等を透過するという欠点がある。本研究では、このようなカメラと LiDAR の特性を考慮し、照明変化に対して頑健なセマンティックセグメンテーション手法を検討する。

2. 関連技術

本研究では、セマンティックセグメンテーションを行うにあたり、SalsaNext[1] (以下手法 1 と呼称) と、DeepLabV3 Plus+SDCNe-tAug[2] (以下手法 2 と呼称) を併用する。

手法 1 は、3D LiDAR 点群を用いた手法である。出力結果は、各点においてクラス毎に色分けされた、入力データと同サイズの 3D 点群データとなる。拡張畳み込みにより多くのコンテキスト情報を保持し、活用するモデルである。

手法 2 は、カメラのカラー画像を用いた手法である。出力結果は、各ピクセルにおいてクラス毎に色分けされた、入力画像と同サイズの RGB 画像となる。トレーニングでは、過去と未来の双方のフレームを用いたビデオ予測を行う。また、クラス間の境界において、一つのピクセルに対し複数のクラスを予測可能とする。以上の手法を用いることにより、トレーニング時の損失を抑制し、十分な照明条件下において、セグメンテーション精度を向上させている。

3. 提案手法

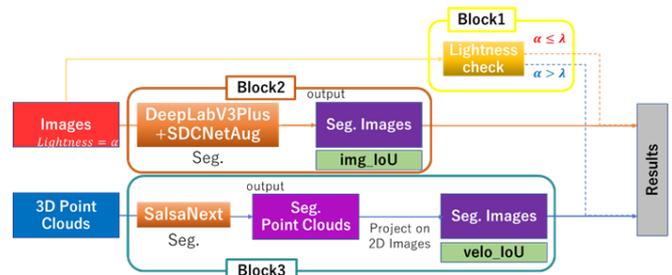


図 3.1: Proposed Method

本研究で提案する手法を図 3.1 に示す。本研究は、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーションを行うことを目的とする。本提案手法では、カラー画像に対する明度判定手法と、二つのセマンティックセグメンテーション手法により構成される。明度判定基準には、事前に各クラスに対して決定した閾値 λ を用いる。判定結果によって、セグメンテーション結果を選択する。しかし、手法 1 と手法 2 ではセグメンテーション結果の次元数が異なる。従って、評価基準が異なるという問題がある。そこで、本提案手法では、手法 1 の結果を出力した場合、3D 点群を 2D 画像に投影することにより、双方の問題に対応する。

また、本研究では KITTI Dataset[3][4]を用いた閾値 λ の決定を行う。閾値 λ は、手法 2 の精度が手法 1 の精度を下回ったときにおける入力画像の明度とした。なお本研究では、閾値 λ の決定は building クラスと car クラスに対して行う。

4. 実験

本研究では、同一情景での照明条件を変化させ、LiDAR 点群を取得することが困難であったため、入力画像の明度を調整し、撮影シーンにおける照明条件の変化を疑似的に再現することとした。入力画像には KITTI Dataset[3][4]及び SemanticKITTI Dataset [5][6]を用いることとする。

はじめに、本提案手法における明度の閾値 λ の決定を行った。まず、3D LiDAR 点群を用いた手法 1 の実行結果を、2D 画像へ投影した。投影した結果の例を図 4.1 に示す。次に、図 4.1 において、building クラス及び car クラスに分類されている領域を、それぞれペイントツールを用いて選択

した。そして、選択した領域とそれ以外の領域において画素値を二極化することで、対象クラスの領域を抽出した。Building クラスの領域を抽出した結果を図 4.2 に示す。同様にして作成した Ground Truth を図 4.3 に示す。図 4.2 と図 4.3 をピクセル毎に比較し、式 1 を用いて評価値 IoU を導出した。手法 2 については、まず KITTI Dataset に用意されている画像の明度を $\frac{1}{2}$ 倍、 $\frac{1}{4}$ 倍、 $\frac{1}{8}$ 倍、 $\frac{1}{12}$ 倍した画像をそれぞれ用意した。明度調整した画像に対し、手法 2 によるセマンティックセグメンテーションを行った。手法 2 の実行結果と図 4.3 をピクセル毎に比較し、評価値 IoU を導出した。本研究では、三枚の画像について実験を行った。手法 1 及び手法 2 を用いて導出した各画像における building クラス及び car クラスの IoU を、それぞれ表 4.1、表 4.2 に示す。なお、表 4.2 の手法 1 における括弧内の数値は、LiDAR による観測が困難である窓ガラス (図 4.1) を観測若しくは補間可能な場合における IoU である。

以上の結果を踏まえ、building クラスに対する閾値 λ は、表 4.1 より KITTI Dataset の明度の $\frac{1}{4}$ 倍と決定した。一方で、car クラスに対する閾値 λ は、上述した LiDAR の特性により手法 1 の精度が低くなったため、本実験では決定できなかった。但し、LiDAR における課題を克服する技術が開発された場合は、 $\frac{1}{4}$ 倍から $\frac{1}{8}$ 倍前後の値となることが期待される。

$$IoU = \frac{TP}{TP + FP + FN} \times 100 [\%] \quad (1)$$

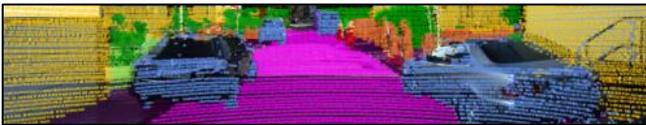


図 4.1: 手法 1 の実行結果を 2D 画像へ投影した結果



図 4.2: 図 4.1 を building クラスの領域を抽出した画像

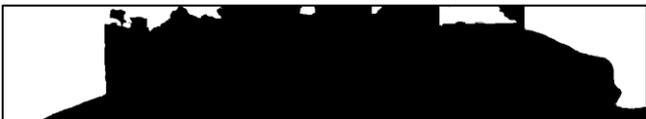


図 4.3: 作成した building クラスの Ground Truth

表 4.1: 各画像における building クラスの IoU [%]

		画像 1	画像 2	画像 3
手法 2	1/2 倍	85.17	79.05	68.31
	1/4 倍	77.15	68.98	67.16
	1/8 倍	73.86	59.69	63.85
	1/12 倍	73.74	69.76	60.40
手法 1		82.21	78.69	64.12

表 4.2: 各画像における car クラスの IoU [%]

		画像 1	画像 2	画像 3
手法 2	1/2 倍	87.63	90.95	95.99
	1/4 倍	81.52	89.82	95.96
	1/8 倍	78.33	84.78	93.29
	1/12 倍	78.09	75.24	91.77
手法 1		74.38 (80.51)	63.47 (85.65)	78.52 (88.58)

5. 結論と課題

本研究では、自動運転システムにおいて重要となる、高精度で安定した画像認識技術を実現するため、カメラと LiDAR を用いた二つのセマンティックセグメンテーションモデルに対し、明度判定手法を追加する手法を提案した。また、building クラスと car クラスに対し、それぞれ明度判定基準となる閾値 λ の決定を行った。その際、building クラスにおいて、照明変化によりカラー画像を用いたセグメンテーション精度が LiDAR 点群を用いたセグメンテーション精度よりも下回ることを示した。また、LiDAR のガラス等を検出できないという特性から、評価対象クラスによって閾値 λ が異なることを示した。以上より、本提案手法は、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法である。

今後の課題として、閾値 λ の決定を自動化且つ精緻化する必要があるであろう。また、劣悪な照明条件下における精度向上のため、LiDAR による観測が困難な領域に対し、新たなアプローチを検討する必要がある。

参考文献

- [1] T. Cortinhal, G. Tzelepis, and E. E. Aksoy, "SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving", International Symposium on Visual Computing (ISVC), pp.207-222, Oct. 2020.
- [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", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.8848-8857, Jun. 2019.
- [3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset", International Journal of Robotics Research (IJRR), 32, pp. 1229-1235, 2013.
- [4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "The KITTI Vision Benchmark Suite", <http://www.cvlibs.net/datasets/kitti/>, Mar. 2012.
- [5] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences", The IEEE International Conference on Computer Vision (ICCV), pp.9296-9306, Oct. 2019.
- [6] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences", <http://www.semantic-kitti.org/>, Jul. 2019.

2020 年度 卒業論文

照明変化に対して頑健な
セマンティックセグメンテーション

Robust Semantic Segmentation against Lighting Changes

提出日 : 2020.02.02

指導教員 : 渡辺 裕 教授 印

早稲田大学基幹理工学部 情報通信学科

1W153050-0

小谷 晟経

目次

第1章 序論.....	1
1.1 まえがき	1
1.2 研究背景.....	1
1.3 自動運転システムの関連技術における課題	1
1.3.1 画像認識技術.....	2
1.3.2 関連入出力機器.....	2
1.4 研究目的.....	3
1.5 本論文の構成.....	4
1.6 むすび.....	4
第2章 関連研究	5
2.1 まえがき	5
2.2 セグメンテーション	5
2.2.1 インスタンスセグメンテーション	6
2.2.2 セマンティックセグメンテーション	6
2.2.3 パノプティックセグメンテーション	6
2.2.4 第2.2節における総括.....	7
2.3 SalsaNext.....	7
2.4 DeepLabV3Plus+SDCNetAug.....	9
2.5 評価指標.....	11
2.6 むすび.....	12
第3章 提案手法	13
3.1 まえがき	13
3.2 提案手法.....	13
3.3 入力画像の明度の調整.....	15
3.4 評価用データセットの Ground Truth の作成	16

3.5 むすび.....	18
第4章 実験.....	19
4.1 まえがき.....	19
4.2 KITTI Dataset / SemanticKITTI Dataset	19
4.3 実験概要.....	21
4.4 予備実験.....	22
4.5 明度の閾値決定.....	26
4.5.1 SalsaNext を用いた 3D LiDAR 点群におけるIoU _{SN} の導出.....	26
4.5.2 DeepLabV3Plus+SDCNetAug を用いたカラー画像におけるIoU _{DL} の導出.....	31
4.6 むすび.....	37
第5章 結論と課題.....	38
5.1 結論.....	38
5.2 課題.....	39
謝辞.....	40
参考文献.....	41
図一覧.....	42
表一覧.....	44

第 1 章 序論

1.1 まえがき

本章では、まず、本研究における研究背景として、近年注目されている自動運転システムの必要性と、それに係る画像認識技術の重要性について論ずる。次に、画像認識技術の動向と課題について説明した後、本研究の目的について述べる。最後に、本論文の構成について簡単に示す。

1.2 研究背景

近年、機械学習を我々の生活で用いられている家電や機械などに応用することが急速に進展している。例えば、空調をはじめとした IoT 製品や、自動車における先進運転支援システム ADAS (Advanced Driver-Assistance Systems) などが挙げられる。中でも、特に自動運転システムに関する研究が注目を集めている。この自動運転システムの実用化において、周辺環境を把握し、対象物の検出・判定を正確に行う必要がある。これは、自動運転システムは人為的な事故の防止を防止することが最大の目的であり、交通事故防止の為に障害物や人間を事前に検知する必要があるためである。従って、高精度で安定した画像認識技術が重要となる。将来的には、完全自動運転システムを実現し、交通事故による死傷者数を 0 にし、安心して生活できる社会基盤の構築を目指している。そこで、完全自動運転を含めた自動運転技術に関する標準規格が、2014 年に SAE (米国自動車技術会) が規定し、2018 年に第三版として発行済みの「Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles [1]」において定義された。

1.3 自動運転システムの関連技術における課題

第 1.2 節において、画像認識技術が自動運転システムによる交通事故数 0 の実現において重要であると論じた。そこで、本節では、第 1.3.1 節において、自動運転システムの研究分野における画像認識技術の現在の動向と、課題について論ずる。なお、詳しくは後述するが、画像認識技術の課題として、入出力機器に用いるセンサの性能や特性に大きく影響を受けることが挙げられる。従って、第 1.3.2 節において、画像認識技術に用いられる関連入出力機器の特性とそれぞれの課題点について詳しく説明する。

1.3.1 画像認識技術

オブジェクト検出： オブジェクト検出とは、入力データに対して対象物の存在する領域を予測してバウンディングボックスで抽出し、クラス分類を行う手法である。実用例として、ADAS における自動ブレーキ機能などが挙げられる。また、近年では、高速道路など限定的な条件下における自動運転システムを搭載した自動車が開発され、注目を集めている。しかし、完全自動運転システムによる交通事故数 0 を達成するためには、より正確に周辺環境を把握する必要がある。また、ドローンや支援ロボットなど、より複雑な撮影シーンにおける自動運転が求められる場合も同様である。このようなケースに対し、オブジェクト検出は領域をアバウトに予測するため、対応できないという課題がある。

セグメンテーション： セグメンテーションとは、入力データの各ピクセルに対して対象物のクラス予測を行う手法である。予測結果は、付加されたクラス情報に基づき、ピクセル毎に色分けされて確認することが可能な手法である。従って、詳細に周辺環境を把握可能であるため、オブジェクト検出における課題に対応することが可能な手法である。近年では自動運転システムへの適用の他、医療分野における疾患部位の特定などに用いられている。

以上のように、二つの手法にはそれぞれ特徴があるが、共通した課題として、入出力機器に用いるセンサの性能や特性に大きく影響を受けることが挙げられる。これは、上述したように、両手法とも入力データに依存して出力結果を得ることから明らかである。

1.3.2 関連入出力機器

ミリ波レーダ： ミリ波レーダとは、ミリ波光を照射し、その反射光を観測するセンサである。照射する光の波長が長いので、天候に左右されず、指向性が強い。このような特性から、ミリ波レーダは対象物までの距離を計測可能である一方で、ミリ波光の特性から非金属の検出が困難なため、車体の検出が難しいという課題がある。

カメラ： カメラ（カメラセンサ）とは、レンズに入る RGB の光の強度を光電変換してデジタル化にすることで、周囲の情報を取得するセンサである。対象物のテクスチャや色を識別可能であり、密な画素で構成された画像を出力する。一方で、照明変化に脆弱であるという課題がある。

更に、ミリ波レーダとカメラの両センサに共通している課題として、周辺環境の把握や対象物の位置関係などを詳細に検出することが困難であることが挙げられる。以上の理由から、現在の画像認識技術では、完全自動運転化の実現は非常に困難であるという課題がある。そこで、近年注目を集めているセンサとして、全方位レーザーレーダ LiDAR (Laser Imaging Detection And Ranging) がある。

LiDAR: LiDARとは、非常に短い波長のレーザを照射し、反射光を検出するまでの時間と光の強度を計測するセンサである。具体的には、カメラや人の目が干渉しない、主に電磁スペクトルの近赤外線部分である750nm~1.5 μ mの波長の光を用いる。出力するデータは、水平方向の全視野における3D点群データとなる。LiDARは長距離測位や3次元情報の取得が可能であり、照明条件の影響を受けない。その特性から、主に周辺環境の3Dマッピングや対象物の材質の判定、自己位置推定を行うために用いられる。また、照射する波長の長さから、以前より気象衛星などに用いられている。これは、エアロゾルは直径1nmから100 μ m程度、雲粒は直径2 μ mから20 μ m程度と、LiDARの波長よりも大きいために検出可能である。近年では小惑星探査機はやぶさにも搭載され、小惑星までの測距や衛星表面の環境把握などに利用されていた。

このように、LiDARはカメラやミリ波レーダの抱える課題を解決できる一方で、距離が離れると反射光の散乱度が大きくなるため、解像度が低下し、遠方の対象物を検出できないという課題がある。また、ガラスのような透過率が高い対象物、若しくは鏡など反射率が高く、周辺環境を写し映し出すような対象物を検出することができないという欠点がある。更に、出力結果が3Dデータであるため、他のセンサと次元数が異なり、各センサの出力データを統合して利用する場合に2Dデータに変換する必要がある。

1.4 研究目的

第1.2節で論じたように、自動運転システムの実現において、高精度で安定した画像認識技術が重要である。これに対し、第1.3節で説明したように、画像認識技術は入出力機器に用いるセンサの性能や特性に大きく影響を受けるという課題がある。現在、自動運転システムに関する画像認識技術には、主にカメラが用いられている。しかし、既存のカメラ画像を用いた画像認識技術において、照明変化に脆弱であるという課題がある。また、照明条件の影響を受けないセンサであるミリ波レーダやLiDARにおいて、センサの特性により検出できない対象物があるという課題がある。従って、双方の課題に対応する画像認識手法を実現する必要がある。そこで、LiDARをカメラやミリ波レーダなどと併用して自動運転システムに組み込むことで、SAE定義の完全自動運転システムの実現が期待されている。

従って、本研究では、上述したカメラとLiDARの双方の課題に対応するため、カメラとLiDARの二つのセンサを併用することとする。その上で、両センサの特性を考慮した、照明変化に対して頑健な画像認識技術を実現することを目的とし、手法の提案を行う。提案手法については第3章で詳しく説明する。

1.5 本論文の構成

以下に本論文の構成を示す。

- 第1章 本章であり、研究背景及び目的について述べる。自動運転システムの必要性と、自動運転に用いられる画像認識技術の重要性について論ずる。また、画像認識技術の課題を示し、本研究の目的について説明する。
- 第2章 本研究で用いる関連研究について述べる。本研究では、周辺環境を詳細に把握する画像認識技術として、セグメンテーション手法を用いる。従って、本章では、セグメンテーション手法の分類、及び本研究に用いる二つのセグメンテーション手法について説明する。また、本研究で用いる評価指標について説明する。
- 第3章 本研究の提案手法について論ずる。本研究ではカメラとLiDARの特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法の提案を行う。はじめに提案手法の理論を説明する。そして、本実験に用いるGround Truthの作成について説明する。
- 第4章 提案手法の実験概要と実験結果を記述し、実験結果の考察を論ずる。本研究では、3D LiDAR点群とカラー画像の評価対象クラスに対して実験を行い、評価を行う。また、実験結果の考察を行う。なお、詳しい実験概要は第4.3節に示す。
- 第5章 本研究の結論と今後の課題について論ずる。はじめに本研究の総括を行い、続いて実験結果から得られた成果と今後の課題について説明する。

1.6 むすび

本章では、はじめに第1.2節において、自動運転システムの必要性と、自動運転システムに用いられる画像認識技術の重要性について論じた。自動車における自動運転システムは、交通事故の件数を0にし、安心して生活ができる社会基盤の構築を最大の目的としている。将来的に完全自動運転の実現を目指す上で、高精度で安定した画像認識技術が重要である。しかし、現在の画像認識技術には、実用化する上で大きな課題がある。そこで、第1.3節では、画像認識技術の動向と課題について詳細に説明した。第1.3節の内容を踏まえ、第1.4節では本研究の動機と目的について説明した。以上より、本研究ではカメラとLiDARの特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法の提案を行うこととする。

第 2 章 関連研究

2.1 まえがき

本章では、関連研究について詳しく述べる。

まず、第 2.2 節では、画像認識手法の一つであるセグメンテーション手法について詳しく説明する。セグメンテーション手法とは、入力画像の各ピクセルが属するクラスを予測し、ピクセル毎にクラス情報を付加する技術である。近年、主に自動運転システムや医療分野などに利用され始めている。セグメンテーション手法は次の三種類に大きく分類される。具体的には、インスタンスセグメンテーション、セマンティックセグメンテーション、パノプティックセグメンテーションである。インスタンスセグメンテーションは、オブジェクト検出のようにバウンディングボックスを用いて、ピクセルレベルで対象物の領域を検討する手法である。セマンティックセグメンテーションは、ピクセル毎にクラス情報を付加する手法であり、最も研究が進んでいる手法である。しかし、可算な対象物が重なっている場合、個別に認識することができないという欠点がある。パノプティックセグメンテーションは、上述したセマンティックセグメンテーションにおける課題を解決した手法である。ただし、最新の手法であり、評価指標も従来のもものと異なるため、比較や評価が困難であるという課題がある。第 2.2 節では、上述した三つの手法について詳しく説明する。

また、第 2.3 節及び第 2.4 節では、本研究で用いるセグメンテーション手法である、SalsaNext[3]及び DeepLabV3Plus+SDCNetAug[4]についてそれぞれ説明する。SalsaNext は 3D LiDAR 点群において、各点に対してクラス情報を付加する手法である。出力データはクラス毎に色分けされた 3D 点群データとなる。一方、DeepLabV3-Plus+SDCNetAug はカメラ画像において、各ピクセルに対してクラス情報を付加する手法である。出力データはクラス毎に色分けされた RGB 画像となる。第 2.5 節では、本研究において評価指標に用いる IoU の導出について説明する。

2.2 セグメンテーション

本節では、第 2.1 節で述べた、画像認識手法の一手法であるセグメンテーション手法について詳しく説明する。セグメンテーション手法は大きく三種類に分類される。以下では、三種類のセグメンテーション手法について、それぞれ詳しく説明する。

2.2.1 インスタンスセグメンテーション

インスタンスセグメンテーションは、オブジェクト検出のようにバウンディングボックスを用いた、オブジェクトの存在する領域を検出し、セグメンテーションを行う手法である。評価指標はオブジェクト検出と同様に mAP (mean Average Precision) を用いる。オブジェクト検出との相違点は、ピクセルレベルでの検出となる点である。従って、より正確に対象物の領域を検出することが可能となる。一方で、オブジェクトの存在する候補領域に対してセグメンテーションを行うため、全てのピクセルに対しクラス情報を付加するものではない。従って、ピクセル単位でクラス情報を付加する他のセグメンテーションと性質が異なる。

2.2.2 セマンティックセグメンテーション

セマンティックセグメンテーションは、入力画像上の全てのピクセルに対し、ピクセル毎にクラス情報を付加する手法である。評価指標には IoU (Intersection over Union) や mIoU (mean IoU) を用いる。この手法は、同じクラスのオブジェクトが重なっている場合、一つのクラスの領域として認識することから、オブジェクトを個別に認識することができない。しかし、空や道などの数えられないオブジェクトに対してもクラス情報を付加することが可能である。従来手法では日中に屋外で撮影された画像をトレーニングに用いている。

なお、本手法の課題としては、ピクセル単位でクラス情報を付加するため、入出力機器に用いるセンサの性能や特性に大きく影響を受けることが挙げられる。カメラを用いる場合は、カメラが照明変化に対して脆弱であるため、照明条件が劣悪になるとセグメンテーション精度が低下する。また、LiDAR を用いる場合は、LiDAR がガラスのような透過性が高い対象物や鏡など反射率の高い対象物を検出できないため、これらの対象物がある領域にブラックが生じる。例えば、自動車の窓ガラスやガラス張りの建築物などが挙げられる。以上のように、本手法は入出力機器に用いるセンサの性能や特性を十分に考慮する必要がある。

2.2.3 パノプティックセグメンテーション

パノプティックセグメンテーションは、2017年に Anurag らが Pixelwise Instance Segmentation with a Dynamically Instantiated Network[2]で提案した、新しいセグメンテーションである。この手法は、全てのピクセルに対しクラス情報とインスタンス番号を付加するものである。従って、数えられる対象物において、個別に認識することが可能であることが、最大の利点として挙げられる。評価指標には PQ (Panoptic Quality) を用いている。しかし、最新の手法であり、評価指標も従来の画像認識手法と異なる。そのため、現状では既存研究との精度の比較や、他の技術と併用する際に評価が難しいという難点がある。また、第 2.2.2 節で述べた入出力機器に用いるセンサに関する課題を有する。

2.2.4 第 2.2 節における総括

第 2.2 節では、画像認識手法の一つであるセグメンテーションについて詳しく述べた。初めにセグメンテーションについて説明し、次に三種類のセグメンテーション手法について詳しく説明した。なお本研究では、カメラのカラー画像及び LiDAR 点群に対しセグメンテーションを行い、評価を行う。従って、第 2.2 節で説明した各手法の特徴を踏まえ、両センサを用いた研究が最も進んでいる、セマンティックセグメンテーションを用いることとする。なお、第 2.2.2 節では、セマンティックセグメンテーションにおける課題として、入出力機器に用いるセンサの性能や特性に大きく影響を受けることを示した。

2.3 SalsaNext

本節では、本研究で用いる、SalsaNext (以下手法 1 と呼称) について詳しく説明する。手法 1 とは、2020 年に T. Cortinhal らが SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving[3]で提案した、3D LiDAR 点群に対するセマンティックセグメンテーションである。手法 1 では 2019 年に J. Behley らが SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences[4]において公開した SemanticKITTI Dataset[5]を用いて評価を実施した。SemanticKITTI Dataset については、第 4.2 節で詳しく説明する。SemanticKITTI Dataset で公開されているベンチマーク一覧によると、手法 1 の精度は $mIoU = 59.5[\%]$ で、現在 8 位である。但し、本研究においてモデルの実装を行った当初は、当時オープンソースとして提供されていたモデルの中では 1 位の精度であった。従って、本研究で手法 1 を用いることとした。手法 1 のネットワーク構造の概要図を図 2.1 に示す。

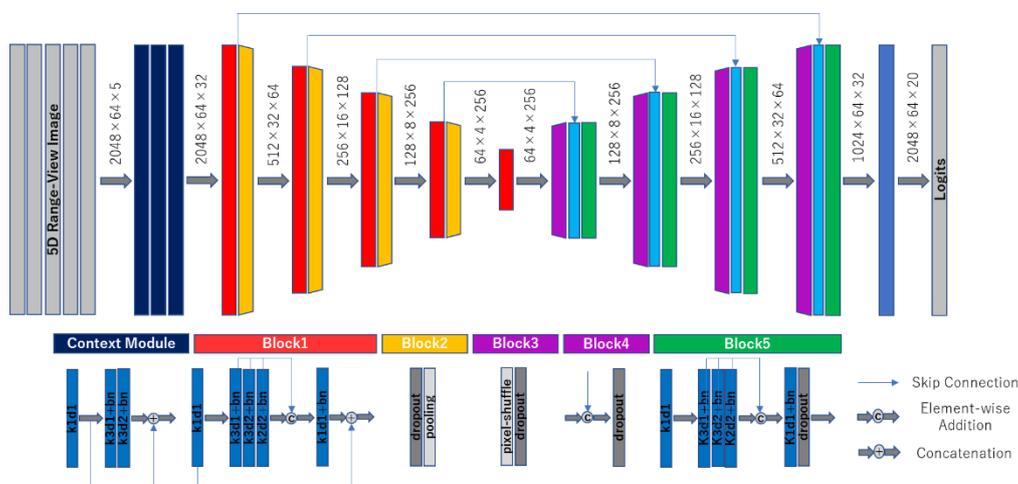


図 2.1: SalsaNext のネットワーク構造の概要図

従来のセマンティックセグメンテーションでは、ネットワーク全体でコンテキスト情報が不足していることが主な問題点の一つとして知られている。この問題点のシンプルな解決方法としてはカーネル数を拡大することであるが、同時にパラメータ数も大幅に増加してしまうといった欠点がある。

そこで、T. Cortinhal らは、図 2.1 に示すように、ネットワークの最初に 1×1 カーネルと 3×3 カーネルを追加した。また、Encoder に ResNet ブロックを含めた残留拡張畳み込みネットワークを用いている。これにより、詳細な空間情報とグローバルコンテキスト情報を取得している。残留拡張畳み込みネットワークでは、図 2.1 の Block1 に示すように、まず ResNet ブロックを 3, 5, 7 の受容野を持つ拡張畳み込みに置換する。次に、Block2 にあたる Dropout layer と Pooling layer を結合させたものを、Block1 と連結させる。これを残留拡張畳み込みネットワークの一つのスタックとする。これにより、畳み込みネットワークの各深度で取得された特徴量をより多く活用できる。

一方 Decoder では、Pixel-shuffle layer を用いることで、アップサンプリングされた特徴マップを生成する。その特徴マップに対し、スキップ接続を用いて Encoder で取得された残差ブロックの特徴量を結合させることで、畳み込みのスタックとしている。

なお、手法 1 では、3D LiDAR 点群を球面に投影することで RV 画像を生成し、畳み込み演算を可能としている。また、360 度全視野に対し、3 次元座標と反射光の強度、及び範囲インデックスをそれぞれ個別の RV 画像チャンネルに保存し、ネットワークへ入力している。出力結果は入力と同じサイズの 3D 点群データで、各点に対しその点が属するクラスのラベルが付加されたものとなる。手法 1 を用いた 3D LiDAR 点群に対するセマンティックセグメンテーションの結果の例を以下の図 2.2 に示す。なお、各図において、左の画像が入力データ、右の画像が出力結果である。

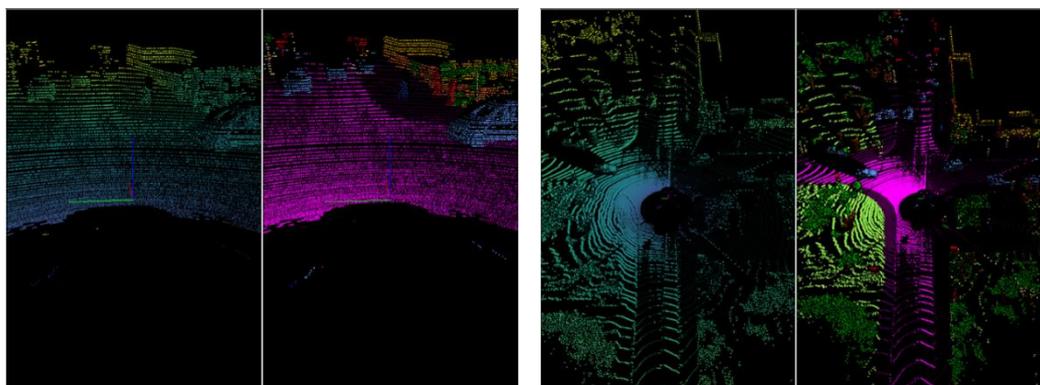


図 2.2: 手法 1 を用いたセマンティックセグメンテーション結果の例

(左: ドライバー視点, 右: 鳥瞰図)

2.4 DeepLabV3Plus+SDCNetAug

本節では、DeepLabV3Plus+SDCNetAug（以下手法2と呼称）について詳しく説明する。手法2は、2018年にY. ZhuらがImproving Semantic Segmentation via Video Propagation and Label Relaxation[6]において提案した、カラー画像に対するセマンティックセグメンテーションである。手法2は複数のデータセットにおいて評価実験が行われている。このうち、2016年にM. CordtsらがThe Cityscapes Dataset for Semantic Urban Scene Understanding[7]において発表したCityscapes Dataset [8]によると、手法2のベンチマークはmIoU=83.5[%]で、現在22位である。一方、A. GeigerらがVision meets Robotics: The KITTI Dataset [9]において公開したKITTI Dataset [10]によると、ベンチマークはmIoU=72.82[%]で、現在2位である。なお、本研究においてモデル実装時には1位であった。手法1においてKITTI Datasetを用いていることから、KITTI Datasetにおいて精度の良い手法2を用いることとした。

手法2のネットワーク構造の概要図を図2.3に示す。手法2は図2.3に示すように、Video PropagationとLabel Relaxationを適用することで、疎らにアノテーションされたビデオフレームシーケンスに対してトレーニングする手法である。図2.4において、 I_t は時刻 t におけるビデオフレームであり、 L_t は時刻 t における I_t に対応するラベルである。

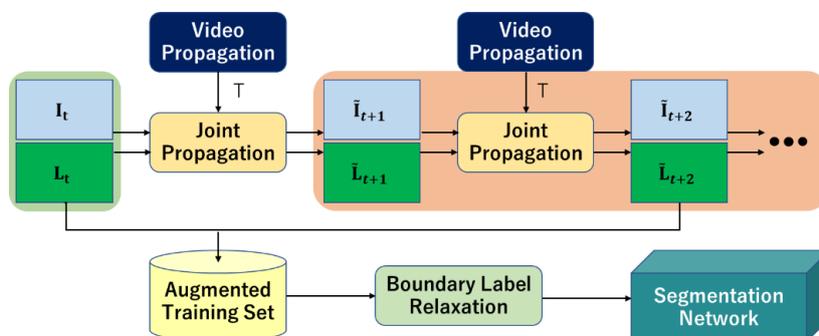


図 2.3: DeepLabV3Plus+SDCNetAug のネットワーク構造の概要図

まず、トレーニングでは、ビデオフレームに対して、現在と過去のフレームからベクトルベースで将来のフレーム \hat{I}_{t+1} とラベル \hat{L}_{t+1} を予測し生成するVideo Propagationを適用する。次に、Joint Propagationにおいてこれらの画像とラベルを結合させる。続いて、過去のフレームから将来のフレームにおける、結合させたビデオフレームとラベルのペアを用いたAugmented Training Setを生成する。これにより、フレーム間の対応が正確でない領域に対し、不整合が生じないものとしている。また、従来の過去のフレームのみを用いたビデオ予測モデルと比較して、より優れたパラメータが作成可能となる。

一方、従来のセマンティックセグメンテーション手法では、クラス間の境界に位置するピクセルに対して分類が難しいという課題がある。これは、取得されたコンテキスト情報の半分以上が異なるクラスのものである可能性がある場合、受容野の中心ピクセルを分類することが困難であるためである。更に、アノテーションがエッジに沿って完全にピクセル化されないため、より分類が難しくなる。従って、手法2では、上述の課題に対して、Label Relaxation を適用することで、クラス間の境界のピクセルに対し、複数のクラスを予測可能としている。

以上の2つの手法を用いることにより、トレーニングの損失が小さくなり、セマンティックセグメンテーション精度を向上させている。しかしながら、トレーニングセットは日中の屋外で撮影された画像を用いるため、十分な照明条件が必要となるモデルである。

出力結果は入力画像と同じサイズで、各ピクセルに対し、そのピクセルが属するクラスのラベルが付加される。図 2.4 に示す入力画像に対し、手法2を用いたセグメンテーションの結果例を以下の図 2.5 に示す。また、図 2.5 を図 2.4 に投影した図を、図 2.6 に示す。



図 2.4: 入力画像の例



図 2.5: 図 2.4 に対し手法2を用いたセマンティックセグメンテーション結果



図 2.6: 図 2.6 のセグメンテーション結果を図 2.5 に投影した結果

2.5 評価指標

本節では、本研究の評価実験で用いる評価指標の導出について詳しく説明する。本研究では、評価指標にIoU（Intersection over Union）を用いる。図 2.7 に示すように、正解領域を A、セマンティックセグメンテーションによる予測領域を B としたとき、IoUは以下の式によって導出される。また、全てのクラスもしくは画像におけるIoUの平均値はmIoU（mean IoU）で表される。

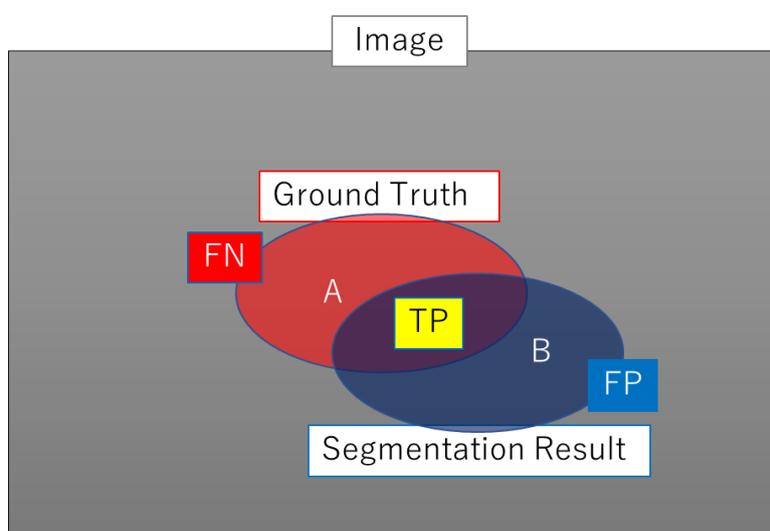


図 2.7: IoU の導出における対応イメージ

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \times 100 = \frac{A \cap B}{A \cup B} \times 100 \quad [\%] \quad (2.1)$$

$$\begin{cases} \text{TP} = A \cap B \\ \text{FP} = B - \text{TP} \\ \text{FN} = A - \text{TP} \end{cases} \quad (2.2)$$

IoUは正解領域と予測領域の一致している割合を示すため、IoUが大きいほど精度の良いモデルであると言える。

本研究では、まず Ground Truth 画像とセマンティックセグメンテーション結果において、評価対象クラスにラベル付けされた範囲をアノテーションした。次に、それぞれの画像を読み込み、アノテーションされた範囲内のピクセルを 1、範囲外のピクセルを 0 とした。そして、各ピクセルにおいて論理和と論理積を求めることで、画像全体の IoU を導出した。

2.6 むすび

本章では、本論文の関連研究について詳しく述べた。まず、第 2.2 節では画像認識技術に分類される三種類のセグメンテーション手法について詳しく説明した。なお、本研究は自動運転システムへの実用化を目的としているため、対象物の位置関係や周辺環境を詳細に把握する必要がある。また、第 1.4 節で述べたように、本研究ではカメラと LiDAR を併用した新しい手法を提案する。従って、現在最も研究が進んでおり、比較や評価を十分に行うことができる、第 2.2.2 節で述べたセマンティックセグメンテーションを用いることとする。

しかし、第 2.2.2 節で示したように、セマンティックセグメンテーションは入出力機器に用いるセンサの性能や特性に大きく影響を受ける課題がある。従って、本研究では、カメラ画像を用いたセマンティックセグメンテーションと、LiDAR 点群を用いたセマンティックセグメンテーションの二つのモデルを併用し、双方のセグメンテーションにおける課題を考慮した手法を提案する。

そこで、第 2.3 節及び第 2.4 節において、本研究で用いる二つのセマンティックセグメンテーションについてそれぞれ説明した。まず、第 2.3 節では、LiDAR 点群に対するセマンティックセグメンテーション手法である SalsaNext について詳しく説明するとともに、実行結果の例を図示した。次に、第 2.4 節では、カメラ画像に対するセマンティックセグメンテーションである DeepLabV3Plus +SDCNetAug について詳しく説明するとともに、実行結果の例を図示した。その上で、トレーニングに日中の屋外で撮影された画像を用いているため、十分な明度の撮影シーンである必要があるという課題について示した。なお、照明条件が悪化した際のセマンティックセグメンテーション結果については、第 4.4 節で説明する予備実験において示す。

最後に、第 2.5 節では本研究で評価指標に用いる IoU の概要と、本研究の評価実験における導出方法について詳しく説明した。

第3章 提案手法

3.1 まえがき

本章では、本研究の提案手法について詳細に説明をする。

第2章で論じた通り、従来のセマンティックセグメンテーションには三つの課題がある。一つ目は、カメラ画像を用いたセマンティックセグメンテーションは照明変化に対し脆弱で、劣悪な撮影環境下において画像認識精度が低下する点である。二つ目は、LiDAR はガラスのように透過性の高い対象物や、鏡のような反射率の高い対象物を検出できない特性を持つ点である。従って、自動車の窓ガラスやガラス張りの建築物などを検出できないという課題がある。そして三つ目は、カメラと LiDAR を併用する場合、出力されるデータの次元数や性質が異なるため、併用が難しいという点である。

そこで本研究では、上述した課題に対応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。第3.2節では、本提案手法について詳細に説明する。また、第3.3節では、明度判定を行う根拠と、評価実験にデータセットの明度を調整することで、夜間の撮影シーンを疑似的に再現可能である根拠を示す。第3.4節では、本研究で用いる評価用データセットの Ground Truth の作成について説明する。

3.2 提案手法

本節では、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法の提案を行う。提案手法の概要図を図3.1に示す。

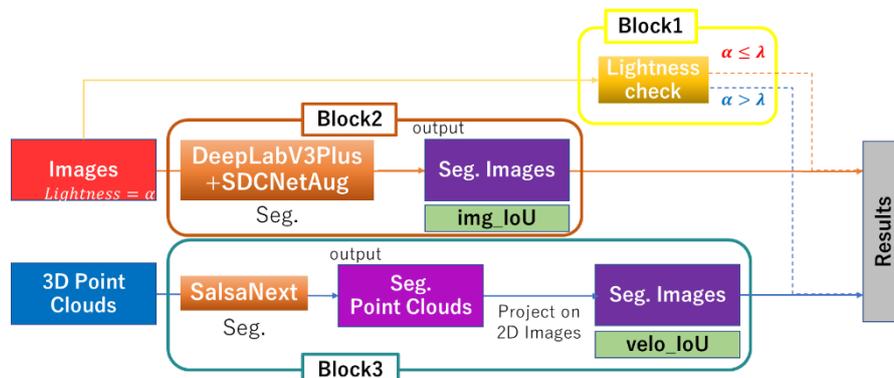


図 3.1: 提案手法の概要図

第 2.2.2 節で示したように、セマンティックセグメンテーションは、入出力機器に用いるセンサの性能や特性に大きく影響を受けるという課題がある。一般に、3D LiDAR 点群を用いたセマンティックセグメンテーションの精度は、カメラ画像を用いたその精度に劣る。しかし、カメラは照明変化に対して脆弱であるという特性から、照明条件の悪化により、手法 2 の精度が手法 1 の精度より低下する。

そこで、本提案手法では、図 3.1 に示すように、本提案手法では入力画像に対し明度判定を行う (Block1)。この明度判定手法では、入力画像の明度 α を、第 4.4 節において決定する明度の閾値 λ と比較する。閾値 λ の決定方法の概要は、以下の通りである。

1. LiDAR 点群に対し手法 1 を用いたセマンティックセグメンテーションを行い、各クラスの評価値 IoU_{SN} をそれぞれ求める。
2. 様々な明度に調整した画像に対して、それぞれ手法 2 を用いたセマンティックセグメンテーションを行い、各明度における各クラスの評価値 IoU_{DL} をそれぞれ求める。
3. 各クラスにおいて IoU_{SN} と IoU_{DL} を比較し、 $IoU_{SN} > IoU_{DL}$ となる時点における入力画像の明度を、それぞれ閾値 λ として決定する。

なお、クラス毎に閾値 λ を決定する理由は、第 1.3.2 節で述べたように、LiDAR はガラスや鏡などを検出できない特性を持つため、クラスによってセマンティックセグメンテーション精度が大幅に低下するからである。

Block2 では、手法 2 を用いた、画像に対するセマンティックセグメンテーションを行う。出力結果は各ピクセルに対しクラスラベルが付加され、クラス毎に色分けされた RGB 画像が出力される。一方 Block3 では、入力画像と同じフレームの LiDAR 点群に対し、手法 1 を用いたセマンティックセグメンテーションを行う。出力結果は、各点に対しクラス情報のラベルが付加され色分けされた、3D LiDAR 点群となる。この出力結果を対応するカラー画像に投影し、2D 画像に変換する。

Block2 及び Block3 におけるセグメンテーション結果に対し、各クラスにおいて Block1 の明度判定結果を適用する。入力画像の明度 α とクラス毎に設定した閾値 λ に対し、 $\alpha \geq \lambda$ の場合は、Block2 の結果を、 $\alpha \leq \lambda$ の場合は Block3 の結果を採用する。

この明度判定手法を追加することにより、各クラスにおいて最適なセグメンテーション結果を得ることが可能となる。また、閾値 λ を設けることにより、良好な照明条件下においてはカメラを、劣悪な照明条件下では LiDAR を用いたセグメンテーション結果を選択することが可能となる。これにより、照明条件の影響を受けない LiDAR のみを用いたセマンティックセグメンテーション以上の精度を担保することが可能となる。

以上の手法により、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法となる。

3.3 入力画像の明度の調整

図 3.2 及び図 3.4 は、それぞれ夜間に撮影したカラー画像と日中に撮影したカラー画像である。また、図 3.3 は、図 3.2 の画像を構成するピクセルのヒストグラムである。そして、図 3.5-A、図 3.5-B、図 3.5-C、図 3.5-D、図 3.5-E は、図 3.4 の明度をそれぞれ1倍、 $\frac{1}{2}$ 倍、 $\frac{1}{4}$ 倍、 $\frac{1}{8}$ 倍、 $\frac{1}{12}$ 倍に変更した画像におけるピクセルのヒストグラムである。各ヒストグラムにおいて、横軸はピクセルの画素値である。画素値とは、 $[0, 255]$ の 256 階調で表される、各画素の色の濃淡や明度を表す値である。また、縦軸は画像内に存在する、画素値におけるピクセルの個数である。

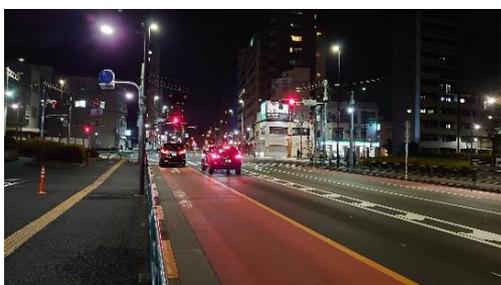


図 3.2: 夜間のカラー画像

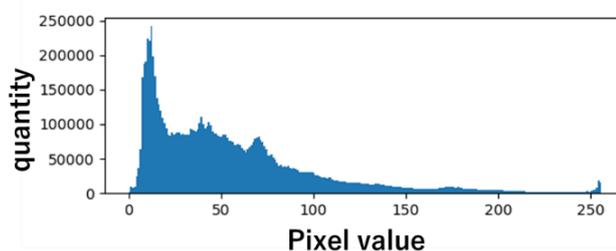


図 3.3: 夜間のカラー画像のヒストグラム



図 3.4: 日中のカラー画像

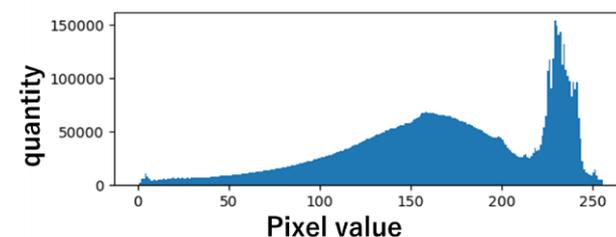


図 3.5-A: 日中のカラー画像のヒストグラム

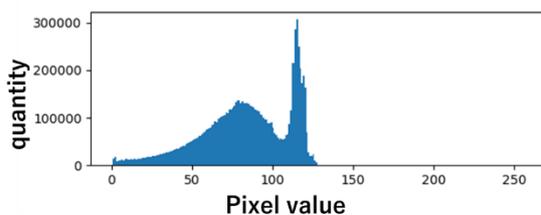


図 3.5-B: 明度 $\frac{1}{2}$ 倍時のヒストグラム

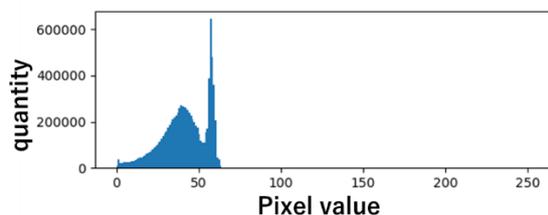


図 3.5-C: 明度 $\frac{1}{4}$ 倍時のヒストグラム

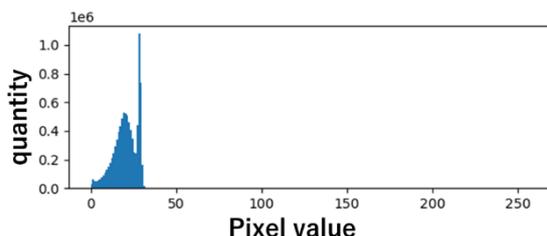


図 3.5-D: 明度 $\frac{1}{8}$ 倍時のヒストグラム

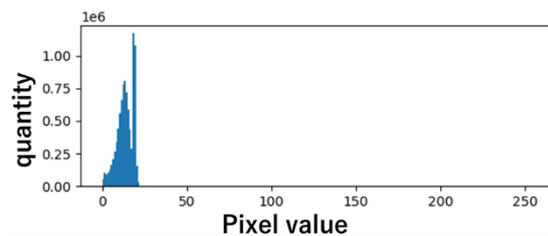


図 3.5-E: 明度 $\frac{1}{12}$ 倍時のヒストグラム

ここで、図 3.3 及び図 3.5-D、図 3.5-E を比較すると、ヒストグラムのピーク値が近いことが分かる。従って、両画像を構成する画素値は比較的近似していると言える。一方で、差異としては、高画素値帯の有無が挙げられる。これは、図 3.2 では明るく色が明瞭な領域が存在しているのに対し、図 3.4 では、画像全体の明度を小さくしていることが原因である。従って、本研究の実験は、街灯や町明かりが殆どない道や、トンネルや洞窟内といった照明条件が悪い撮影シーンに近い条件となる。以上より本研究では、実際の撮影シーンにおける照明条件の変化を、入力画像の明度を調整することで近似させて実験を行うこととする。

そこで、第 4.5 節では、本提案手法の明度決定手法において基準となる、明度の閾値 λ の決定方法について説明する。また、第 3.4 節では、閾値 λ の決定及び評価実験において使用する Ground Truth の作成について、詳細に説明する。

3.4 評価用データセットの Ground Truth の作成

本節では、本研究で使用する Ground Truth の作成について説明する。

本研究では、KITTI Dataset の Odometry シーケンスの Sequence 08 を用いて実験を行う。Odometry シーケンスは LiDAR 点群データを含むが、セマンティックセグメンテーションの Ground Truth が用意されていない。従って、Odometry シーケンスに対応した Ground Truth を用意する必要がある。なお、本研究では building クラス及び car クラスにおいて評価を行うこととする。

次に、Ground Truth の作成方法について述べる。LiDAR は x-z 軸方向（水平方向）に 360 度全視野のデータを取得できるが、KITTI Dataset において LiDAR の y 軸方向（画像の縦方向）の視野はカメラの視野よりも狭い。従って、まず、データセットのカラー画像に対し、LiDAR の視野に合わせてトリミングを行う。次に、building クラスと car クラスの正解領

域を、windows のペイントツールを用いてそれぞれアノテーションする。そして、アノテーションした画像をニッチ化する。以下の図 3.6-A に Ground Truth として用いる画像の例を示す。また、図 3.6-A から作成した building クラス及び car クラスの Ground Truth を、図 3.6-B 及び図 3.6-C にそれぞれ示す。

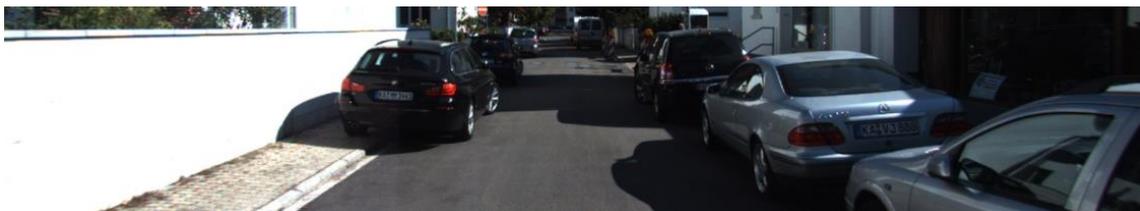


図 3.6-A: データセットのカラー画像の例



図 3.6-B: 図 3.6-A を基に作成した building クラスの Ground Truth



図 3.6-C: 図 3.6-A を基に作成した car クラスの Ground Truth

3.5 むすび

本章では、本研究の提案手法と、評価用データセットの Ground Truth の作成方法について、詳細に説明をした。

まず、第 3.2 節において、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーションの手法を提案した。本提案手法では、3D LiDAR 点群を用いたセマンティックセグメンテーションモデルである SalsaNext (手法 1) と、カメラ画像を用いたセマンティックセグメンテーションモデルである DeepLabV3Plus+SDCNet-Aug (手法 2) を併用した。手法 1 は手法 2 よりもセグメンテーション精度が悪いが、LiDAR を用いているため、照明条件に影響を受けないモデルである。一方、手法 2 はカメラを用いているため、照明変化に脆弱なモデルである。従って、照明条件が変化した際、手法 2 の精度が手法 1 の精度よりも悪化する。そこで、入力画像に対して明度判定を行う手法を追加した。明度判定を行うことにより、手法 1 以上の精度を担保する。

また、本提案手法における明度判定において、基準となる明度の閾値 λ の決定方法の概要を説明した。上述したように、手法 1 以上の精度を担保するため、閾値 λ は手法 2 の精度が手法 1 の精度を下回った時点における入力画像の明度とした。なお、LiDAR はガラスや鏡など、特定の対象物の検出ができないことから、自動車の窓ガラスやガラス張りの建築物などを検出することが困難である。従って、手法 2 では評価対象クラスによって精度が大きく異なる。このような LiDAR の特性を考慮するため、第 3.4 節において説明したように、building クラスと car クラスに対して閾値 λ を決定することとした。明度判定によってクラス毎に二つの手法からセグメンテーション結果を選択することで、両センサの課題に対応する手法となる。

なお、本提案手法では、両手法の精度の比較をする必要がある。また、実用化する上で全クラスのセグメンテーション結果を統合する必要がある。従って、手法 1 を用いた 3D LiDAR 点群に対するセグメンテーション結果を 2D 画像へ投影する手法を追加した。

以上の手法を用いることで、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法となる。

第4章 実験

4.1 まえがき

本章では、本研究で行った実験の説明を行い、考察を述べる。第4.2節では、本研究で用いる KITTI Dataset 及び SemanticKITTI Dataset の概要について説明する。第4.3節には、第4.4節、第4.5節で行う実験の一覧を掲載する。第4.4節では事前実験について述べる。第4.5節では、本提案手法における明度の閾値 λ の決定に伴う評価実験について、実験概要と手法及び結果を記す。以上により、本提案手法が照明変化に対し頑健なセマンティックセグメンテーションであることを示す。最後に第4.6節で実験の総括を行う。

なお、本研究における実験環境は以下の表4.1の通りである。

表 4.1: 実験環境

OS	Ubuntu 18.04.5 LTS 64bit
LAM	31.3 GB
CPU	Intel® Core™ i7-5820K CPU @ 3.30GHz × 12
GPU	NVIDIA GeForce GTX TITAN X/PCIe/SSE2

また、本研究において、手法1は表4.1に示した実験環境においてトレーニング及び実験を行った。一方で、手法2はGitHubに公開されている Pre-trained Model を使用して実験を行った。

4.2 KITTI Dataset / SemanticKITTI Dataset

本研究では KITTI Dataset を用いる。中でも、KITTI Dataset に用意されている Odometry シーケンスの Sequence 08 (以下 Odometry08 と呼称)、及び Semantics シーケンス (以下 Semantics と呼称) を用いる。KITTI Dataset には様々なシーケンスにおける各種データが用意されている。以下の図4.1にKITTI Datasetの構造を示す。

図4.1に示すように、LiDARの点群データは Odometry シーケンス内に velodyne laser

data として用意されている。また、それらに対応するセマンティックセグメンテーション用の Ground Truth は、SemanticKITTI Dataset 内に用意されている。図 4.2 に SemanticKITTI Dataset の構造を示す。一方、カラー画像に対応するセマンティックセグメンテーション用の Ground Truth は、Odometry シーケンスには用意されていない。従って本研究では、第 3.4 節で説明したように Odometry08 シーケンスの color images に含まれる画像を基に、セマンティックセグメンテーション用 Ground Truth を作成し、評価実験を行った。

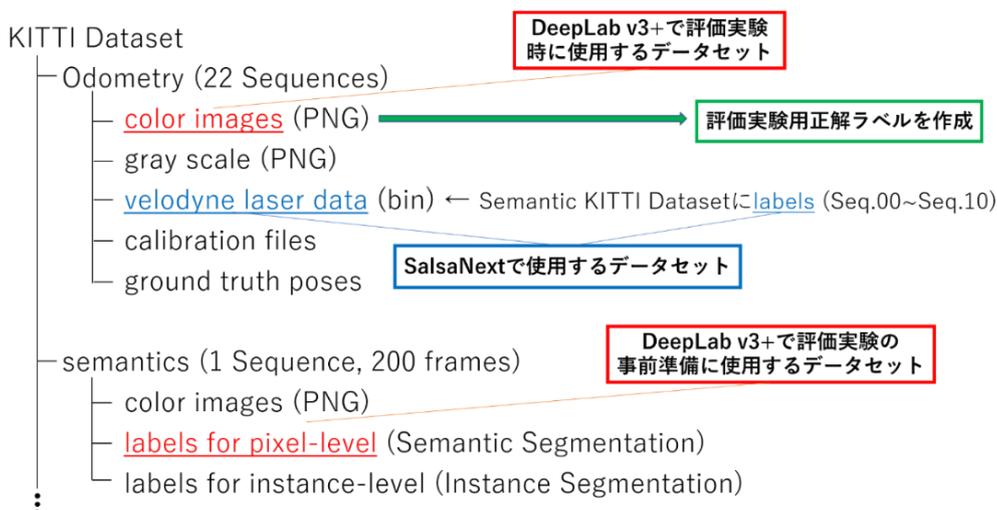


図 4.1: KITTI Dataset の構造

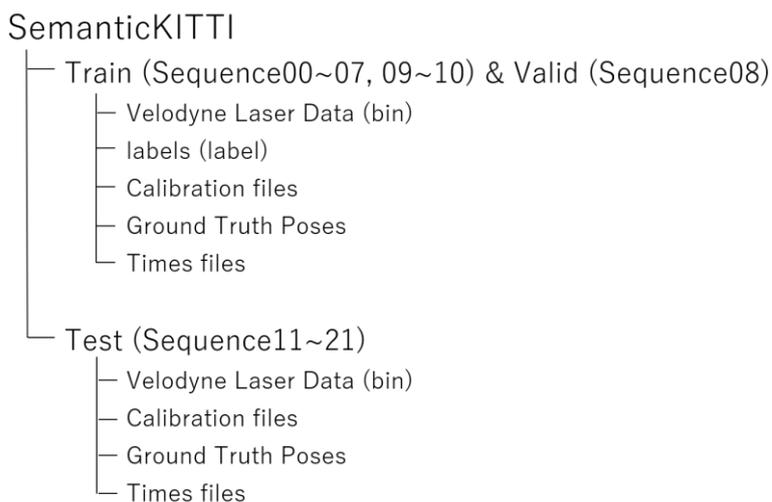


図 4.2: SemanticKITTI Dataset の構造

4.3 実験概要

本研究では、提案手法の評価実験として、以下の手順で実験を行う。

実験 1 予備実験

実験 1.1 Semantics の画像から明度調整をしたデータセットの作成

実験 1.2 実験 1.1.1 で作成した各データセットにおける $mIoU_{DL}$ の導出

実験 2 明度の閾値決定

実験 2.1 手法 1 を用いた 3D LiDAR 点群+における IoU_{SN} の導出

実験 2.1.1 Odometry08 の 3D LiDAR 点群に対するセマンティックセグメンテーションと、3次元空間上における $mIoU_{SN}$ の導出

実験 2.1.2 セマンティックセグメンテーションされた 3D 点群データの 2D カラー画像への投影

実験 2.1.3 実験 2.1.2 で生成した画像による、2次元空間上における IoU_{SN} の導出

実験 2.2 手法 2 を用いたカラー画像に対する IoU_{DL} の導出

実験 2.2.1 Odometry08 のカラー画像から明度調整をしたデータセットの作成

実験 2.2.2 実験 2.3.1 で作成した各データセットにおける IoU_{DL} の導出

実験 2.2.3 IoU_{SN} と IoU_{DL} の比較、及び閾値 λ の決定

なお、評価実験は KITTI Dataset の Odometry08 を用いる。Odometry08 には、第 4.2 節で示した通り、4071 フレームの画像や LiDAR 点群データなどが格納されている。

4.4 予備実験

本章では実験 1 について説明する。実験 1 では、入力画像の明度が低下することで、セマンティックセグメンテーションの精度が低下することを確認する。

まず、実験 1.1 について説明する。実験 1.1 では、Semantics に用意されているカラー画像の明度を1倍, $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍, $\frac{1}{16}$ 倍に調整したデータセットをそれぞれ作成した。作成したデータセットのカラー画像の例を、以下の図 4.3 に示す。



図 4.3: 実験 1.2.1 において Semantics のカラー画像の明度を調整した例
(上から順にそれぞれ明度を1倍, $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍, $\frac{1}{16}$ 倍に調整)

次に、実験 1.2 について説明する。実験 1.2 では、実験 1.1 で作成した各データセットに対し手法 2 を用いてセマンティックセグメンテーションを行い、 $mIoU_{DL}$ を導出した。それぞれの明度におけるクラス毎の $mIoU_{DL}$ を、以下の表 4.4 に示す。表 4.4 から、五つのクラスを抜粋してグラフにしたものが図 4.4 である。また、セマンティックセグメンテーション結果の RGB 画像、及びセマンティックセグメンテーション結果を入力画像に投影した結果を、それぞれ図 4.5 及び図 4.6 に示す。なお、明度はそれぞれ図 4.3 の順と同じである。

表 4.2: 実験 1.2.2 で導出した各データセットにおける $mIoU_{DL}$ [%]

Class Name	original	$\times 1/2$	$\times 1/4$	$\times 1/8$	$\times 1/12$	$\times 1/16$
road	95.40	95.23	93.17	87.99	84.69	80.67
sidewalk	86.31	85.23	78.25	63.40	54.87	44.23
building	91.30	90.88	88.37	78.96	70.11	59.22
wall	81.56	74.75	69.74	59.98	52.02	38.44
fence	68.06	65.52	56.76	46.44	39.53	27.83
pole	74.06	74.12	73.10	69.06	64.54	58.37
traffic-light	76.05	75.93	72.72	67.15	58.03	53.30
traffic-sign	81.99	81.80	79.78	56.78	50.53	44.97
vegetation	93.85	93.69	92.51	89.21	85.65	81.80
terrain	85.45	84.44	82.41	75.33	61.80	51.22
sky	96.94	96.86	94.94	88.22	78.47	58.99
person	82.11	84.23	83.83	80.50	76.90	73.91
rider	54.11	53.26	53.09	52.26	22.89	5.22
car	93.95	93.92	93.24	90.11	86.39	82.48
truck	95.29	94.25	92.90	80.19	47.00	27.61
bus	0.00	0.00	0.00	0.00	0.00	0.00
train	89.55	88.02	82.41	27.38	0.58	0.00
motorcycle	55.39	52.73	50.55	58.68	38.15	30.75
bicycle	51.50	49.96	50.12	52.67	50.16	32.04

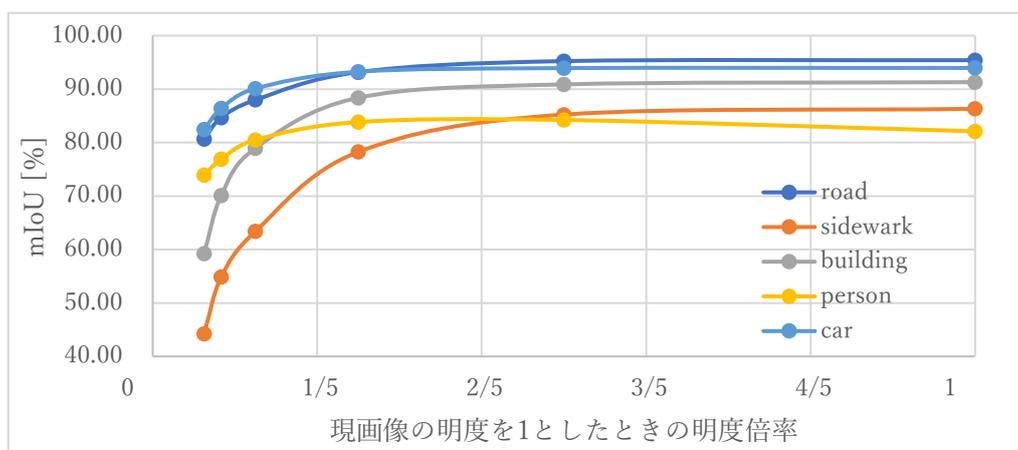


図 4.4: 明度変化による五つのクラスにおける $mIoU_{DL}$ の変化

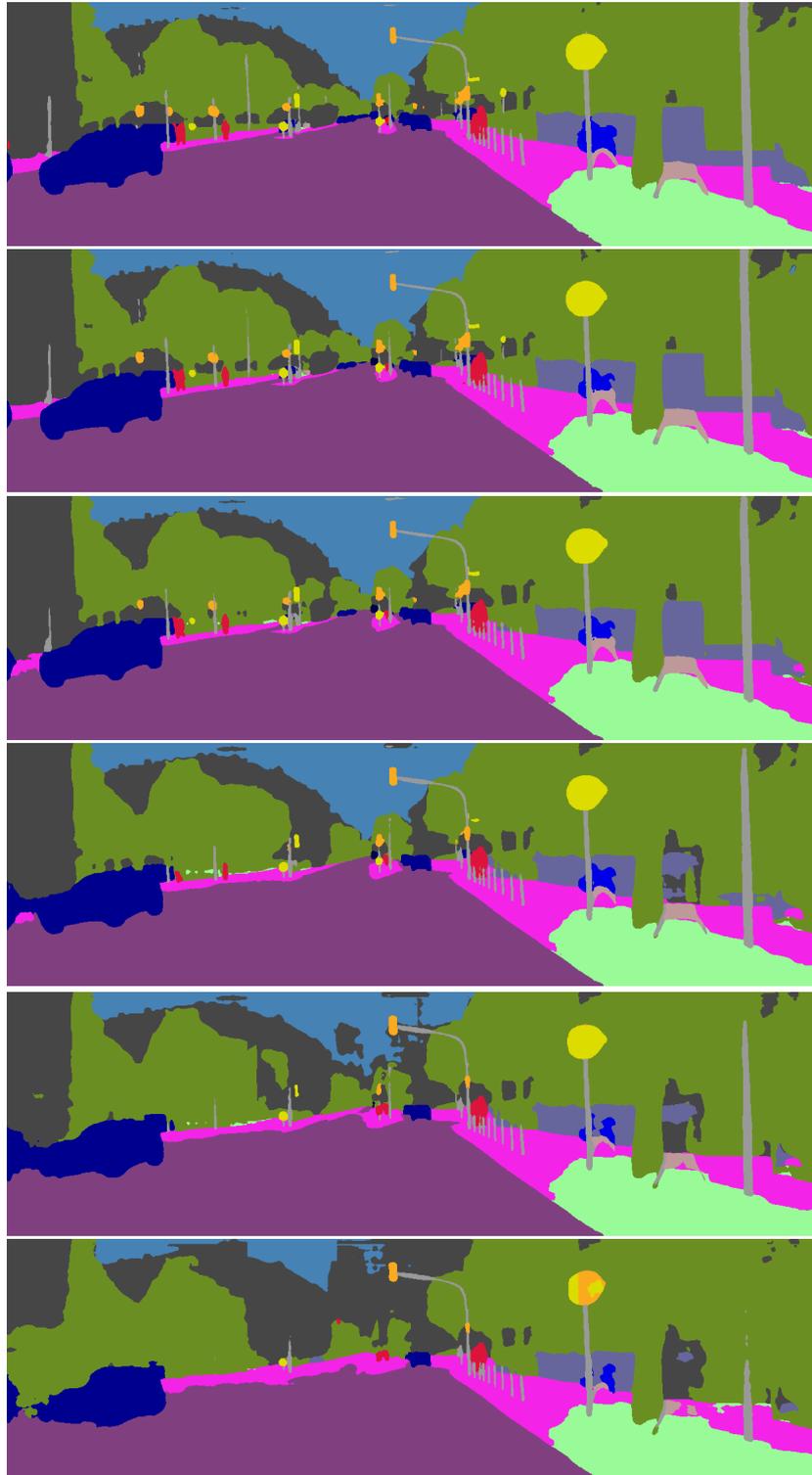


図 4.5: 実験 1.2 における手法 2 のセマンティックセグメンテーション結果
(上から順にそれぞれ明度を1倍, $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍, $\frac{1}{16}$ 倍)

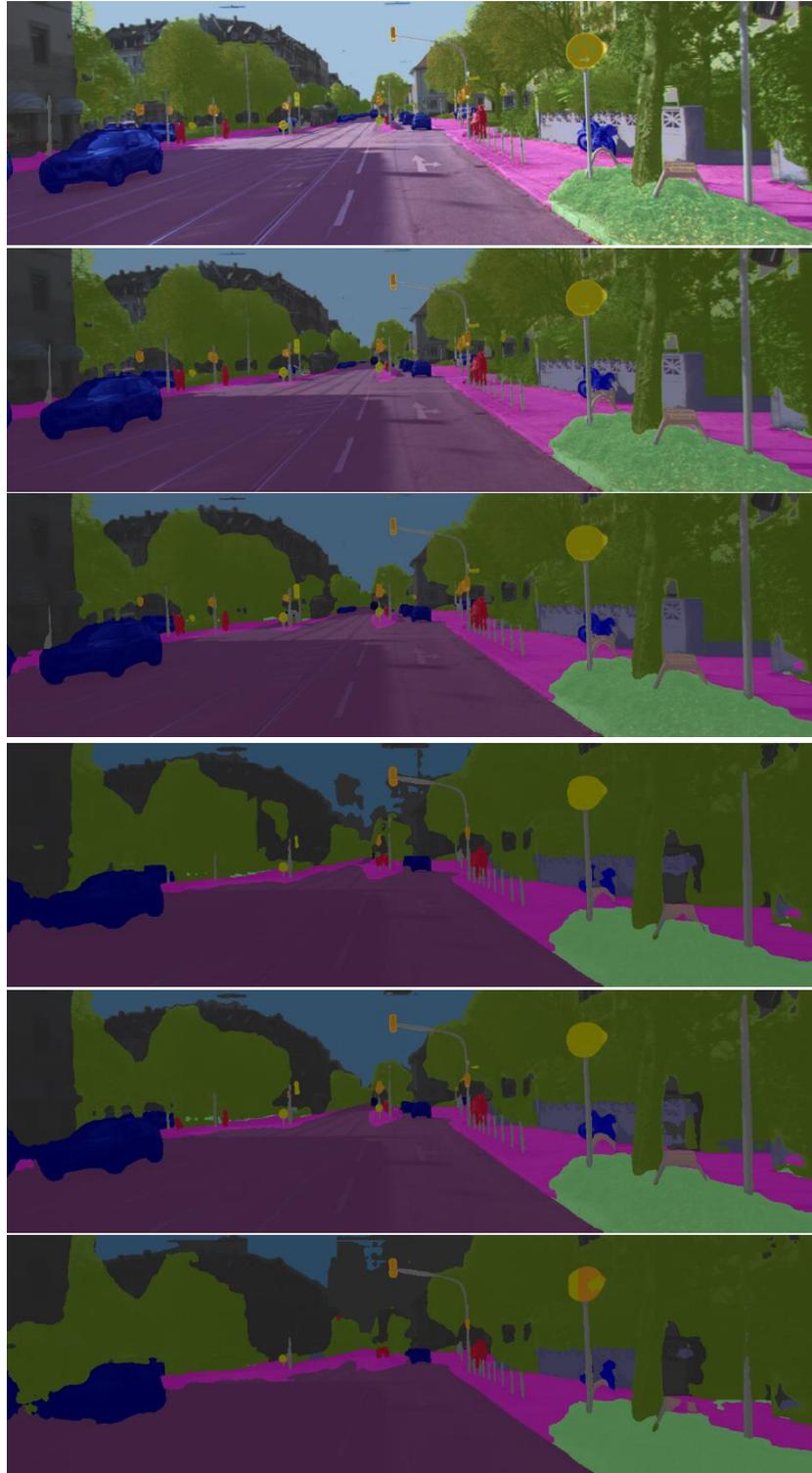


図 4.6: 図 4.4 のセグメンテーション結果を図 4.3 に投影した結果
(上から順にそれぞれ明度を1倍, $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍, $\frac{1}{16}$ 倍)

4.5 明度の閾値決定

本節では，実験 2 について説明をする．実験では，明度判定手法に用いる明度の閾値 λ を決定する．

4.5.1 手法 1 を用いた 3D LiDAR 点群における IoU_{SN} の導出

まず，実験 2.1.1 について説明をする．実験 2.1.1 では，Odometry08 の velodyne laser data に用意されている LiDAR 点群に対して，手法 1 によるセマンティックセグメンテーションと評価を行った．実験 2.1.1 の結果の例を，以下の図 4.7 に示す．また，評価実験で得られた mIoU_{SN} を以下の表 4.3 に示す．なお，ここで表 4.2 の mIoU_{SN} は 3 次元空間上での評価値である．従って，2D 画像のセマンティックセグメンテーションにおける評価値とは評価方法が異なるため，両者を比較し評価することは出来ない点に注意する．

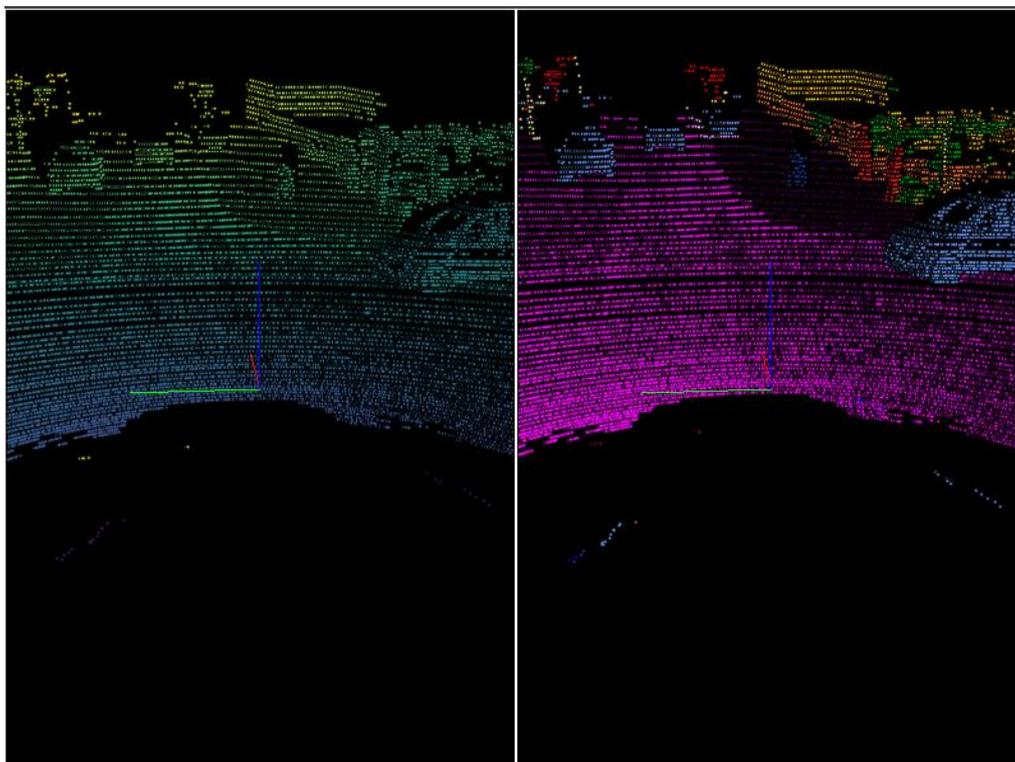


図 4.7: 手法 1 を用いたセマンティックセグメンテーション結果の例

表 4.3: 実験 2.1.1 で導出した $mIoU_{SN'}$

Class Number	Class Name	IoU [%]
1	Car	91.1
2	Bicycle	43.3
3	Motorcycle	48.3
4	Truck	51.0
5	Other-vehicle	38.7
6	Person	66.9
7	Bicyclist	80.8
8	Motorcyclist	0.0
9	Road	94.2
10	Parking	41.8
11	Sidewalk	80.1
12	Other-ground	01.8
13	Building	83.4
14	Fence	40.2
15	Vegetation	82.0
16	Trunk	62.4
17	Terrain	63.3
18	Pole	56.5
19	Traffic-sign	44.7
Average (mIoU)		56.3

次に、実験 2.1.2 について説明する。実験 2.1.2 では、実験 2.1.1 の結果を Odometry08 の color images に用意されている 2D カラー画像へ投影した。手法 1 のセマンティックセグメンテーション結果の例を図 4.8-A 及び図 4.9-A に、対応するフレームのカラー画像を図 4.8-B 及び図 4.9-B に示す。また、図 4.8-A 及び図 4.9-A を図 4.8-B 及び図 4.9-B にそれぞれ投影した結果の例を、図 4.8-C 及び図 4.9-C にそれぞれ示す。なお、LiDAR は x-z 軸方向（水平方向）に 360 度全視野のデータを取得できるが、KITTI Dataset において LiDAR の y 軸方向（画像の縦方向）の視野はカメラの視野よりも狭い。そのため、図 4.8-C 及び図 4.9-C において投影されている範囲が狭くなっている。

従って、実験 2.1.3 及び実験 2.2.3 で評価を行う際は、比較対象の画像も含め、LiDAR 点群が投影されている範囲をトリミングした上で評価を行うものとする。

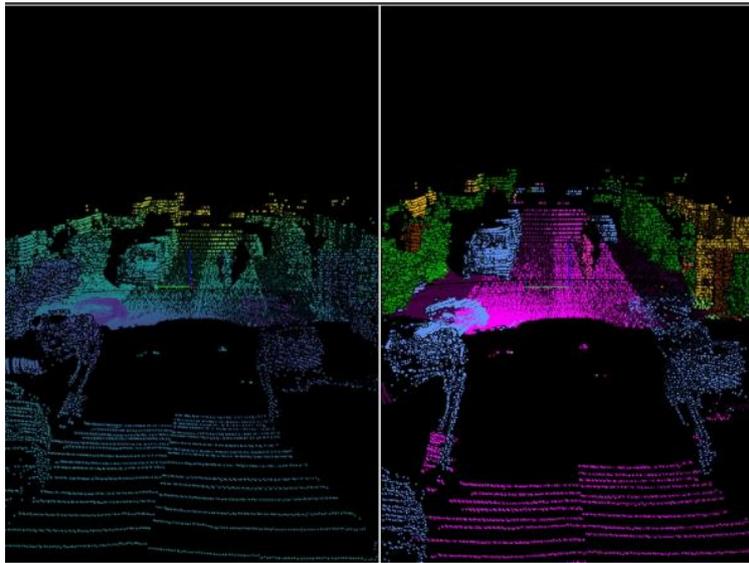


図 4.8-A: 手法 1 による 3D LiDAR 点群のセマンティックセグメンテーション結果の例 1



図 4.8-B: 図 4.8-A に対応するフレームのカラー画像

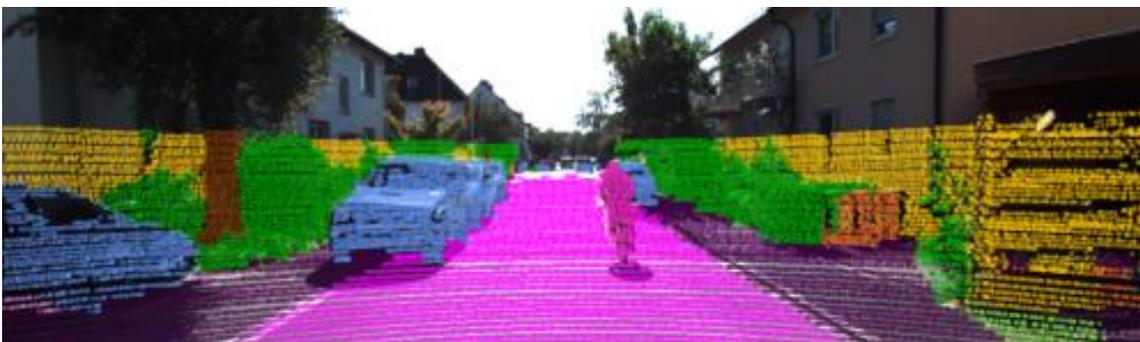


図 4.8-C: 図 4.8-A のセマンティックセグメンテーション結果を図 4.8-B に投影した結果

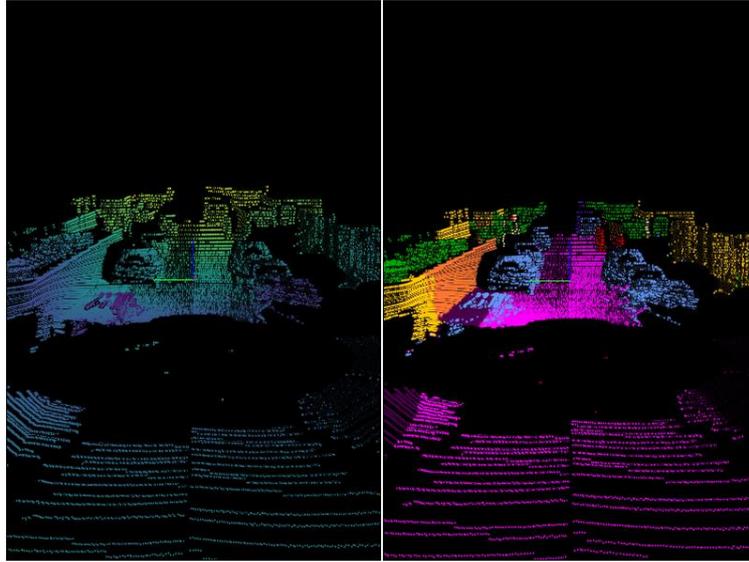


図 4.9-A: 手法 1 による 3D LiDAR 点群のセマンティックセグメンテーション結果の例 2



図 4.9-B: 図 4.8-A に対応するフレームのカラー画像

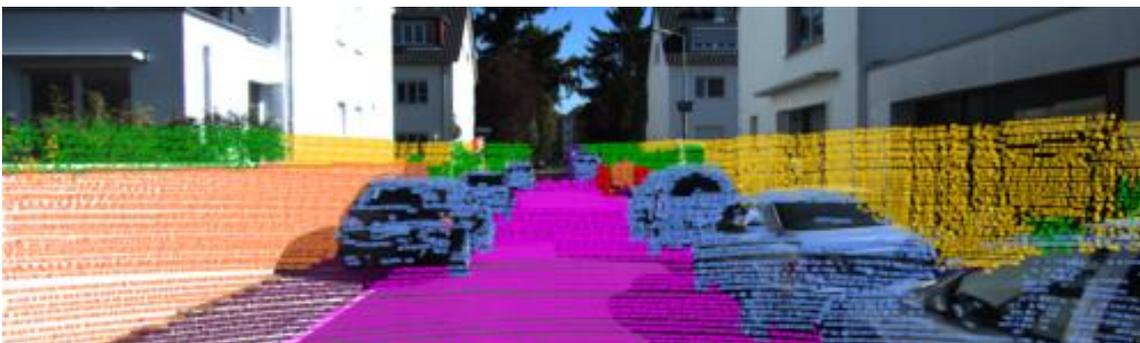


図 4.9-C: 図 4.9-A セマンティックセグメンテーション結果を図 4.9-B に投影した結果

最後に、実験 2.1.3 について説明する。実験 2.1.3 では、実験 2.1.2 の結果から building クラスと car クラスにおける 2 次元空間上での IoU_{SN} 導出を行う。導出の事前準備として、まず図 4.8-C 及び図 4.9-C を LiDAR 点群が投影されている領域においてトリミングした。次に、building クラスと car クラスに分類されている領域を、windows のペイントツールを用いてそれぞれアノテーションした。そして、アノテーションした画像をニッチ化した。図 4.8-C において building クラスと car クラスに対しそれぞれニッチ化した画像を、図 4.10-A 及び図 4.10-B に示す。また、図 4.9-C において building クラスと car クラスに対しニッチ化した画像を、図 4.11-A 及び図 4.11-B に示す。

実験 2.1.2 で生成したニッチ化画像と Ground Truth を比較し、 IoU_{SN} を導出した。同様の実験を三枚の画像において導出した IoU_{SN} を以下の表 4.4 に示す。この IoU_{SN} は、2 次元空間上での評価値であるため、2D カラー画像に対するセマンティックセグメンテーションの結果と比較することが可能となる。



図 4.10-A: 図 4.8-C において building クラスに対しニッチ化した画像



図 4.10-B: 図 4.8-C において car クラスに対しニッチ化した画像



図 4.11-A: 図 4.9-C において building クラスに対しニッチ化した画像



図 4.11-B: 図 4.9-C において car クラスに対しニッチ化した画像

表 4.4: 実験 2.1.3 で導出した三枚の画像における各クラスのIoU_{SN} [%]

	画像 1	画像 2	画像 3
building	82.21	78.69	64.12
car	74.38	63.47	78.52

4.5.2 手法 2 を用いたカラー画像におけるIoU_{DL}の導出

本節では、実験 2.2 について説明する。第 4.2 節で述べた通り、セマンティックセグメンテーション用に用意されている LiDAR 点群のシーケンスと、セマンティックセグメンテーション用に用意されているカラー画像のシーケンスは共通していない。従って、手法 1 と手法 2 のセマンティックセグメンテーション結果を比較・評価を行うためには、Odometry08 に含まれるカラー画像を用いて実験・評価を行う必要がある。

まず、実験 2.2.1 について説明する。実験 2.2.1 では、Odometry08 に用意されているカラー画像の明度を $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍に調整したデータセットをそれぞれ作成した。作成したデータセットのカラー画像の例を、図 4.12 及び図 4.13 に示す。

次に、実験 2.2.2 について説明する。実験 2.2.2 では、実験 2.2.1 で作成した各データセットに対し、手法 2 を用いてセマンティックセグメンテーションを行った。また、第 3.4 節で作成した Ground Truth と比較して、評価対象クラスにおけるIoU_{DL}を導出した。図 4.12 に対しセマンティックセグメンテーションを行った結果の RGB 画像を図 4.13 に示す。ま

た, 図 4.14 に対しセマンティックセグメンテーションを行った結果の RGB 画像を図 4.15 に示す. そして, 図 4.14 及び図 4.15 を Ground Truth と比較してIoU_{DL}を導出する. 同様の実験を複数枚の画像において行った結果を表 4.5 に示す.

最後に, 実験 2.2.3 について説明する. 実験 2.2.3 では, 実験 2.2.2 で導出した評価対象クラスのIoU_{DL}と, 実験 2.1.3 で導出したIoU_{SN}の比較を行った. building クラスと car クラスにおける比較結果を, 表 4.5 及び表 4.6 に示す. なお, 図 4.10-B 及び図 4.11-B に示すように, 手法 1 では車の窓ガラスが検出できていない. そこで, 表 4.5 では, 手法 1 において窓ガラスの検出が可能と仮定した場合の評価値を併せて括弧内に記す.



図 4.12: 実験 2.2.1 において Odometry08 のカラー画像の明度を調整した例 1

(上から順にそれぞれ明度を $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍)

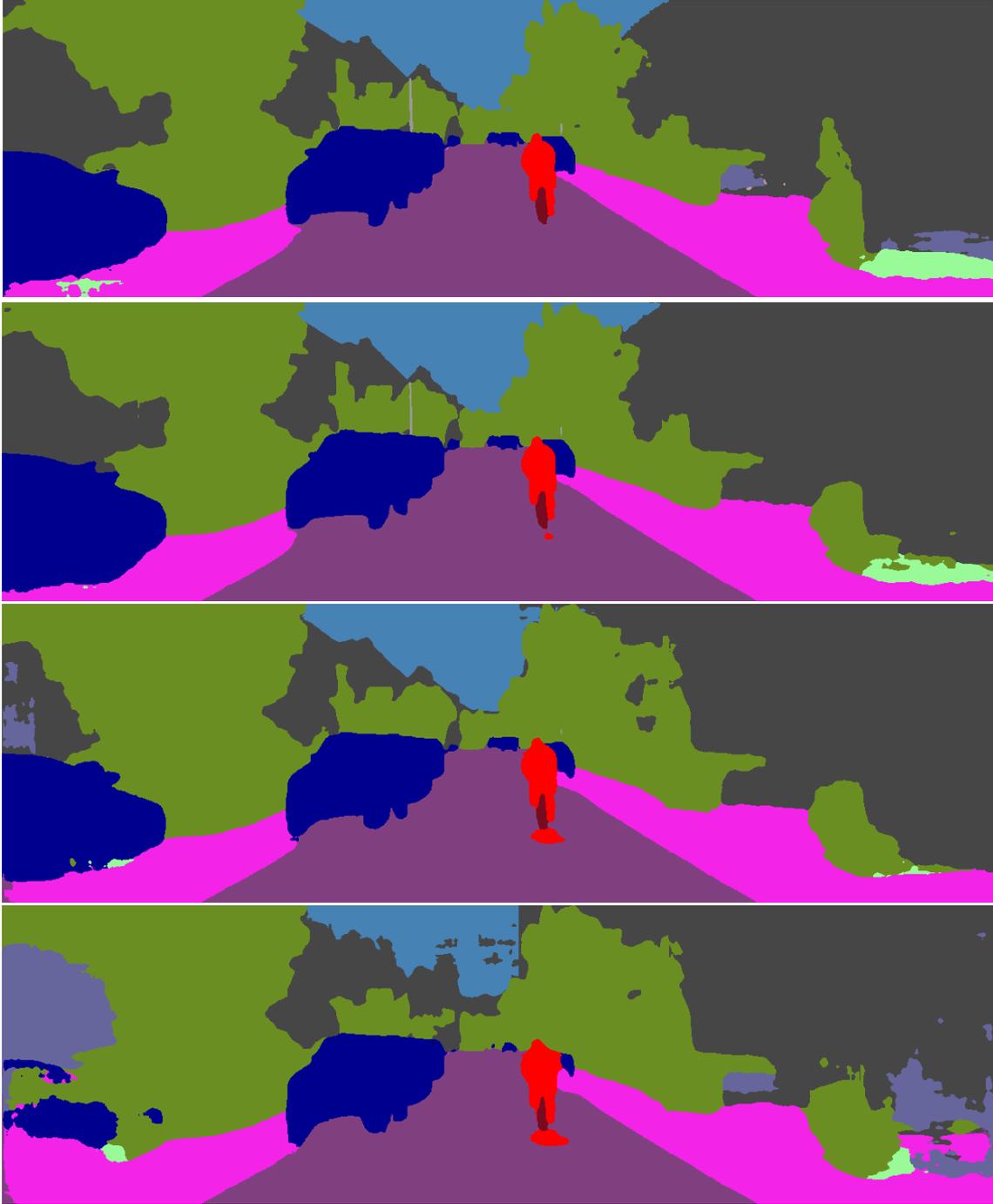


図 4.13: 図 4.12 の各カラー画像に対するセマンティックセグメンテーション結果
(上から順にそれぞれ明度を $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍)



図 4.14: 実験 2.2.1 において Odometry08 のカラー画像の明度を調整した例 2

(上から順にそれぞれ明度を $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍)

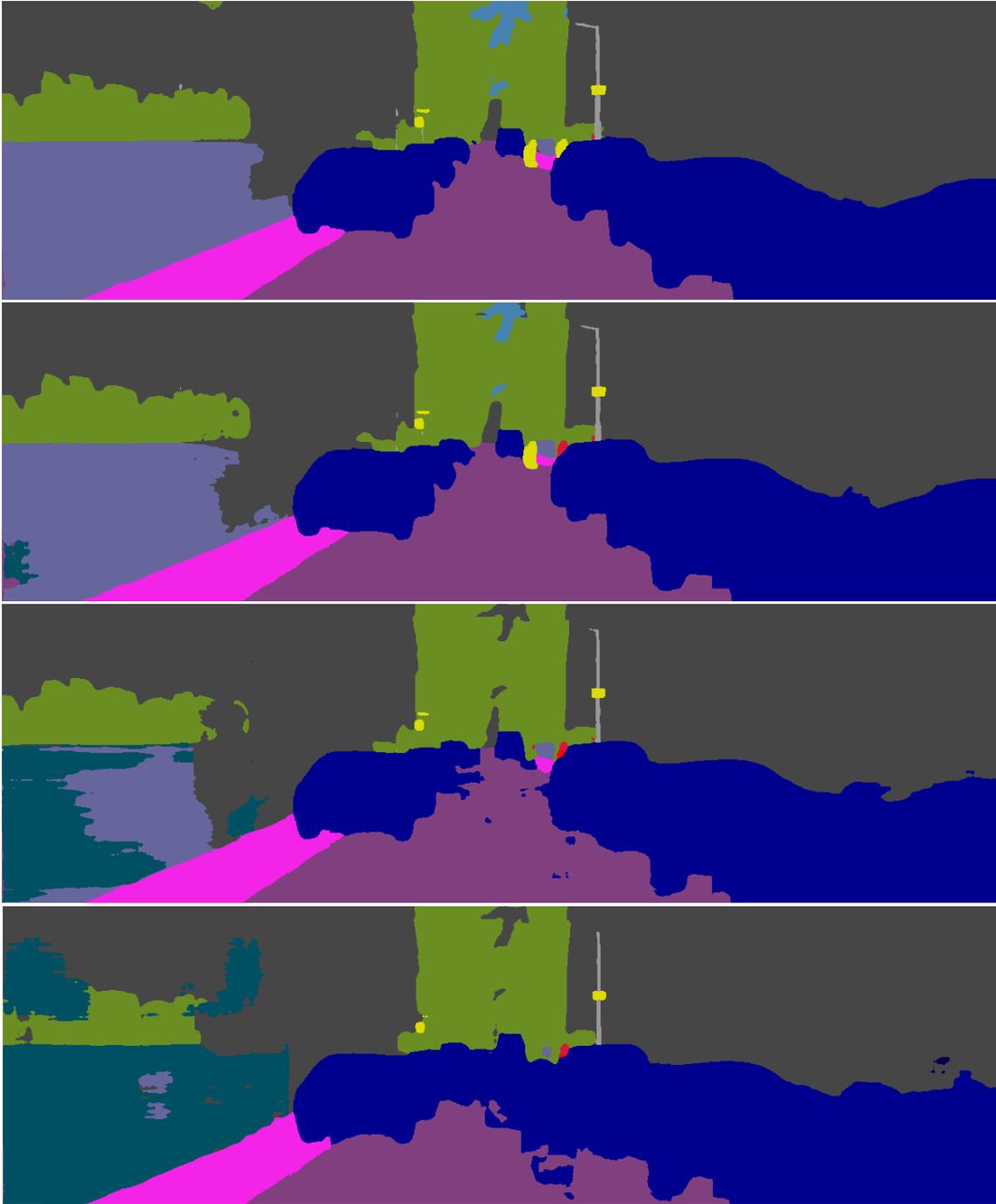


図 4.15: 図 4.14 の各カラー画像に対するセマンティックセグメンテーション結果
(上から順にそれぞれ明度を $\frac{1}{2}$ 倍, $\frac{1}{4}$ 倍, $\frac{1}{8}$ 倍, $\frac{1}{12}$ 倍)

表 4.5: building クラスにおけるのIoU_{SN}とIoU_{DL}の比較 [%]

		画像 1	画像 2	画像 3
手法 2	1/2 倍	85.17	79.05	68.31
	1/4 倍	77.15	68.98	67.16
	1/8 倍	73.86	59.69	63.85
	1/12 倍	73.74	69.76	60.40
手法 1		82.21	78.69	64.12

表 4.6: car クラスにおけるのIoU_{SN}とIoU_{DL}の比較 [%]

		画像 1	画像 2	画像 3
手法 2	1/2 倍	87.63	90.95	95.99
	1/4 倍	81.52	89.82	95.96
	1/8 倍	78.33	84.78	93.29
	1/12 倍	78.09	75.24	91.77
手法 1		74.38 (80.51)	63.47 (85.65)	78.52 (88.58)

4.6 むすび

本研究では、3D LiDAR 点群を用いたセマンティックセグメンテーションモデルである SalsaNext (手法 1) と、カラー画像を用いたセマンティックセグメンテーションモデルである DeepLabV3Plus+SDCNetAug (手法 2) を用いて実験を行った。また、実験には KITTI Dataset 及び SemanticKITTI Dataset を使用した。両データセットは、自動車に各センサを装着し、日中に様々なシーンを走行することで取得したデータを基に作成されたものである。データセットには、カメラ画像や 3D LiDAR 点群データ、セマンティックセグメンテーション用の Ground Truth などが用意されている。ただし、3D LiDAR 点群データとカメラ画像を用いたセマンティックセグメンテーション用のシーケンスは対応していない。従って、第 3.4 節において、3D LiDAR 点群データに対応したカメラのカラー画像を基に、事前にセマンティックセグメンテーション用の Ground Truth を作成した。

第 4.4 節では予備実験を行った。予備実験では、明度を変更した KITTI Dataset のカラー画像に対し、手法 2 を用いてセマンティックセグメンテーションを行い、各明度の画像における IoU の比較を行った。その結果、入力画像の明度が低下すると、IoU も低下することを確認した。なお、第 3.3 節において、夜間に撮影した画像における画素値のヒストグラムと、日中に撮影した画像の明度を低下させた場合における画素値のヒストグラムを比較した。その結果、ヒストグラムが近似していたことから、入力画像の明度を変更することにより、照明変化が生じるケースを疑似的に再現可能であることを示した。

以上より、予備実験において、照明条件が悪化した場合、カメラを用いたセマンティックセグメンテーション精度が低下するということを示した。

予備実験の結果を受け、第 4.5 節では、本提案手法の明度決定手法に用いる明度の閾値 λ の決定を行った。具体的には、まず、3D LiDAR 点群を用いた手法 1 によるセマンティックセグメンテーションを行い、結果を 2D カラー画像に投影した。続いて、building クラスと car クラスにおいてそれぞれ 2 次元上において評価値 IoU_{SN} を導出した。次に、カラー画像の明度を $\frac{1}{2}$ 倍、 $\frac{1}{4}$ 倍、 $\frac{1}{8}$ 倍、 $\frac{1}{12}$ 倍したデータセットを用いて、手法 2 によるセマンティックセグメンテーションを行った。続いて、building クラスと car クラスにおいてそれぞれ IoU_{DL} を導出した。以上により導出した IoU_{SN} と IoU_{DL} を比較し、 $\text{IoU}_{\text{SN}} > \text{IoU}_{\text{DL}}$ となる画像の明度を、閾値 λ として決定した。以上の実験における結果を踏まえ、building クラスに対する閾値 λ は、表 4.1 より KITTI Dataset の明度の $\frac{1}{4}$ 倍と決定した。一方、car クラスでは、本実験では $\text{IoU}_{\text{SN}} > \text{IoU}_{\text{DL}}$ となる結果が得られなかった。これは、LiDAR が自動車の窓ガラスを検出できないためである。なお、表 4.2 の括弧内に示すように、窓ガラスが検出できたと仮定して IoU_{DL} を導出した場合には、 $\text{IoU}_{\text{SN}} > \text{IoU}_{\text{DL}}$ となる結果を得られた。従って、窓ガラスの領域を補間する手法が開発された場合は、閾値 λ を決定することができると期待される。

第5章 結論と課題

5.1 結論

本研究は、完全自動運転システムを実現する上で重要となる、高精度で安定した画像認識技術を提案することを目的としている。画像認識技術は、入出力機器に用いるセンサの性能や特性に大きく影響を受けるという課題がある。この画像認識技術に用いられるセンサとしては、カメラや LiDAR が挙げられる。カメラは照明変化に対して脆弱であるという課題を持ち、LiDAR はガラスや鏡など、一部の対象物を検出できないという課題を持つ。

そこで、本研究では、カメラと LiDAR を併用し、更に明度判定手法を追加したセマンティックセグメンテーション手法を提案した。本提案手法では、二種類のセマンティックセグメンテーションモデルを使用した。一つ目は 3D LiDAR 点群を用いた SalsaNext であり、二つ目はカラー画像を用いた DeepLabV3Plus+SDCNetAug である。一般に、3D LiDAR 点群を用いたセマンティックセグメンテーションの精度は、カメラ画像を用いたその精度に劣る。そこで、明度判定手法を追加する。明度判定手法では、まず入力画像の明度をクラス毎に設定した閾値 λ と比較する。その後、出力された双方のセグメンテーション結果のうち、判定結果によってセグメンテーション結果を選択する。この明度判定手法を追加することにより、各クラスにおいて最適なセグメンテーション結果を得ることが可能となる。

続いて、第 4.5 節において、building クラスと car クラスに対し、それぞれ明度判定基準となる閾値 λ の決定をする実験を行った。その際、照明変化により、building クラスにおいてカラー画像を用いたセグメンテーション精度が、LiDAR 点群を用いたセグメンテーション精度よりも低くなることを示した。また、精度の優劣が逆転したときの入力画像の明度を、閾値 λ として決定した。閾値 λ を設けることにより、良好な照明条件下においてはカメラを、劣悪な照明条件下では LiDAR を用いたセグメンテーション結果を選択することが可能となる。これにより、照明条件の影響を受けない LiDAR のみを用いたセマンティックセグメンテーション以上の精度を担保することが可能となった。従って、本提案手法が照明変化に対して高精度で安定した手法であることが示された。

また、第 4.6 節で論じたように、building クラスの結果と car クラスの結果を比較し、評価対象のクラスによって閾値 λ の値が異なることを示した。これは、LiDAR が持つ窓ガラスや鏡などを検出できないという特性によるものである。従って、LiDAR の特性を考慮するため、クラス毎に明度判定の閾値を設定する必要があることが示された。

以上より、本提案手法は、カメラと LiDAR の特性を考慮した、照明変化に対して頑健なセマンティックセグメンテーション手法である。

5.2 課題

今後の課題として、2点挙げられる。

まず閾値 λ の決定の自動化且つ精緻化を行う必要がある。閾値 λ は入出力機器に用いるセンサが取得したデータを基に決定される。そこで、自動運転システムの実用化に際し、使用するセンサによって異なるパラメータの影響を受けることが想定される。また、目的に応じて扱うクラスが異なるが、第4.5節で示したようにクラス毎に閾値 λ が異なる。従って、閾値 λ の値を自動化し、本研究では対象としていない他のクラスについても閾値 λ を決定することが必要となるであろう。

また、本研究ではLiDAR観測を用いることで、劣悪な照明条件下において精度を維持し、照明変化に対して頑健なセマンティックセグメンテーションを実現している。しかし、完全自動運転システムの実用化に向けて、様々な撮影シーンにおいて、より精度の高い手法を実現する必要がある。従って、LiDARによる観測が困難な領域に対し、新たなアプローチを検討することが必要となる。

謝辞

本研究を行うにあたり，研究方針や手法に関して丁寧にご指導くださり，研究や論文執筆に必要な環境を整えてくださった渡辺裕教授に，心より感謝申し上げます。

チームゼミにおいて研究方針や手法に関して丁寧にご指導くださった，早稲田大学情報通信研究センター招聘研究員の原潤一博士に深謝申し上げます。

ゼミや研究室においてアドバイスやコメントをしてくださった研究室の皆様には感謝いたします。特に，日頃よりチームゼミや研究室において研究方針や手法に関してアドバイスやコメントをくださり，時には実験のサポートもしてくださった，加藤裕也氏，堀隼也氏，浅見莉絵子氏，山野井佑介氏に心より御礼申し上げます。

所属する研究室や大学は異なりますが，研究方針をはじめ，諸々の相談に親身に乗ってくださり，時には遊びに誘ってくださった友人の皆様には，厚く感謝申し上げます。

様々な面で多大なサポートをし続けてくださった家族に深謝いたします。

最後に，コロナ禍で世界的に混乱が続く大変な情勢下にもかかわらず，支えてくださった全ての皆様に，心より御礼申し上げます。

参考文献

- [1] Society of Automotive Engineers International: , “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles”, SAE J3016_JUN 2018, pp.19-33, Jun. 2018.
- [2] A. Arnab, and P. H. S. Torr, ‘Pixelwise Instance Segmentation with a Dynamically Instantiated Network’, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.879-888, Jul. 2017.
- [3] T. Cortinhal, G. Tzelepis, and E. E. Aksoy, “SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving”, International Symposium on Visual Computing (ISVC), pp.207-222, Oct. 2020.
- [4] 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”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.8848-8857, Jun. 2019.
- [5] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, “SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences”, The IEEE International Conference on Computer Vision (ICCV), pp.9296-9306, Oct. 2019.
- [6] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, “SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences”, <http://www.semantic-kitti.org/>, Jul. 2019.
- [7] 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”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3213-3223, Jun. 2016.
- [8] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “CityScapes Dataset: Semantic Understanding of Urban Street Scenes”, <https://www.cityscapes-dataset.com/>, Feb. 2016.
- [9] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets Robotics: The KITTI Dataset”, International Journal of Robotics Research (IJRR), 32, pp. 1229-1235, 2013.
- [10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “The KITTI Vision Benchmark Suite”, <http://www.cvlibs.net/datasets/kitti/>, Mar. 2012.

図一覧

図 2.1: SalsaNext のネットワーク構造の概要図	7
図 2.2: 手法 1 を用いたセマンティックセグメンテーション結果の例	8
図 2.3: DeepLabV3Plus+SDCNetAug のネットワーク構造の概要図	9
図 2.4: 入力画像の例	10
図 2.5: 図 2.4 に対し手法 2 を用いたセマンティックセグメンテーション結果	10
図 2.6: 図 2.5 のセグメンテーション結果を図 2.4 に投影した結果	10
図 2.7: IoU の導出における対応イメージ	11
図 3.1: 提案手法の概要図	13
図 3.2: 夜間のカラー画像	15
図 3.3: 夜間のカラー画像のヒストグラム	15
図 3.4: 日中のカラー画像	15
図 3.5-A: 日中のカラー画像のヒストグラム	15
図 3.5-B: 明度 1/2 倍時のヒストグラム	15
図 3.5-C: 明度 1/4 倍時のヒストグラム	15
図 3.5-D: 明度 1/8 倍時のヒストグラム	16
図 3.5-E: 明度 1/12 倍時のヒストグラム	16
図 3.6-A: データセットのカラー画像の例	17
図 3.6-B: 図 3.6-A を基に作成した building クラスの Ground Truth	17
図 3.6-C: 図 3.6-A を基に作成した car クラスの Ground Truth	17
図 4.1: KITTI Dataset の構造	20
図 4.2: SemanticKITTI Dataset の構造	20
図 4.3: 実験 1.2.1 において Semantics のカラー画像の明度を調整した例	22
図 4.4: 明度変化による五つのクラスにおける $mIoU_{DL}$ の変化	23
図 4.5: 実験 1.2 における手法 2 のセマンティックセグメンテーション結果	24
図 4.6: 図 4.4 のセグメンテーション結果を図 4.3 に投影した結果	25
図 4.7: 手法 1 を用いたセマンティックセグメンテーション結果の例	26
図 4.8-A: 手法 1 による 3D LiDAR 点群のセマンティックセグメンテーション結果の例 1	28
図 4.8-B: 図 4.8-A に対応するフレームのカラー画像	28
図 4.8-C: 図 4.8-A のセマンティックセグメンテーション結果を図 4.8-B に投影した結果	28

図 4.9-A: 手法 1 による 3D LiDAR 点群のセマンティックセグメンテーション結果の例 2	29
図 4.9-B: 図 4.9-A に対応するフレームのカラー画像	29
図 4.9-C: 図 4.9-A のセマンティックセグメンテーション結果を図 4.9-B に投影した結果	29
図 4.10-A: 図 4.8-C において building クラスに対しニッチ化した画像	30
図 4.10-B: 図 4.8-C において car クラスに対しニッチ化した画像.....	30
図 4.11-A: 図 4.9-C において building クラスに対しニッチ化した画像	31
図 4.11-B: 図 4.9-C において car クラスに対しニッチ化した画像.....	31
図 4.12: 実験 2.2.1 において Odometry08 のカラー画像の明度を調整した例 1	32
図 4.13: 図 4.12 の各カラー画像に対するセマンティックセグメンテーション結果	33
図 4.14: 実験 2.2.1 において Odometry08 のカラー画像の明度を調整した例 2	34
図 4.15: 図 4.14 の各カラー画像に対するセマンティックセグメンテーション結果	35

表一覧

表 4.1: 実験環境	18
表 4.2: 実験 1.2.2 で導出した各データセットにおける $mIoU_{DL}'$	23
表 4.3: 実験 2.1.1 で導出した $mIoU_{SN}'$	27
表 4.4: 実験 2.1.3 で導出した IoU_{SN}	31
表 4.5: building クラスにおけるの IoU_{SN} と IoU_{DL} の比較 [%]	36
表 4.6: car クラスにおけるの IoU_{SN} と IoU_{DL} の比較 [%]	36