

1. はじめに

建設現場における重機の自動運転の実現には、LiDARを用いた周囲環境の認識が重要となる。LiDARは、RGB画像では撮影しにくい夜間等の暗い環境でも3次元点群を取得できる。建設現場における認識対象には、建設機械や人のような移動物体だけでなく、盛り土や路面落下物のような高さや広がりが多様な物体も含まれる。そのため、一般的な剛体を対象とする物体検出では、認識が困難となることが想定される。そこで、本研究では3次元点群の各点に対してクラス分類を行うセマンティックセグメンテーションに注目する。3次元点群は近距離領域の点群が密であり、遠距離領域の点群が疎となることから、遠距離における精度が低下しやすい距離依存性がある。この点群密度の違いは学習プロセスにも影響すると考えられる。そこで本研究では、点群データにおける距離依存性と学習プロセスへの影響を考慮した損失関数を提案する。

2. セマンティックセグメンテーション

3次元点群データを疑似画像に変換してセマンティックセグメンテーションを行う手法として、SalsaNext[1]がある。SalsaNextのネットワーク構造を図1に示す。SalsaNextは、SalsaNetをベースに、Context Module, Pixel-Shuffle Layersを導入し、高い識別精度とリアルタイム性を実現している。

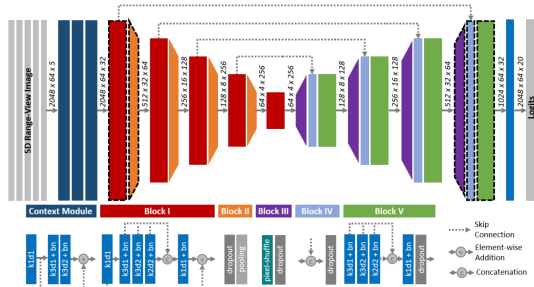


図1：SalsaNextのネットワーク構造

3. 点群データの学習プロセス

SalsaNextによる点群データの学習について、近距離(0-10m)と遠距離(60-70m)におけるエポック毎の精度を図2に示す。近距離は学習初期から精度が高く、学習中期で十分な精度に到達している。一方で、遠距離は学習初期の精度は低く、学習後期でも精度が十分に高くなっていない。このように、点群データの学習プロセスには、近距離と遠距離で異なる特性がある。

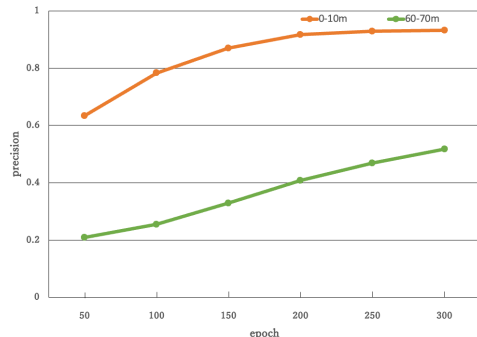


図2：点群データの学習傾向

4. 提案手法

本研究では、点群データの距離依存性と学習プロセスに応じた影響を考慮した損失関数である Distance Weighted Focal Loss(DWFL)を提案する。DWFLは、学習プロセスに応じて徐々に遠距離データを重視するように重み付けをすることで、近距離と遠距離間の学習バランスを最適化する。このアプローチにより、近距離での検出精度を低下させることなく、遠距離の検出精度を高める効果が期待できる。

4.1. Distance Weighted Focal Loss

従来の Focal Loss [2] は、不均衡なクラスが存在する分類問題に対処するための損失関数である。DWFLでは、式(1)に示すように、Focal Lossに動的距離係数 W を導入する。動的距離係数 W では2つの重み付けのアプローチを行う。1つ目のアプローチは、距離依存性を考慮した動的距離係数 W_d による重み付けを行う。2つ目のアプローチは、学習プロセスと距離依存性を考慮した動的距離係数 W_s による重み付けを行う。ここで、 P_t は正解クラス t の予測確率、 γ は難しいサンプルを強調するためのパラメータである。

$$DWFL(P_t) = -W \cdot ((1 - P_t)^\gamma \cdot \log(P_t)) \quad (1)$$

4.1.1. 動的距離係数 W_d の導入

動的距離係数 W_d は、学習が困難な遠距離の検出精度の高精度化を目的とする。 W_d は、式(2)のように、LiDARの原点から点の距離 d をデータセット内の最大距離 d_{max} で正規化することで導出する。これによって、遠距離であるほど W_d が大きくなり距離依存性が考慮される。

$$W_d = 1 + \frac{d}{d_{max}} \quad (2)$$

4.1.2. 動的距離係数 W_s の導入

学習プロセスと距離依存性を考慮した動的距離係数 W_s は、近距離への検出精度を低下させることなく、遠距離の検出精度の高精度化を目的とする。そのため、学習プロセスに応じて徐々に遠距離データの重みを大きくするように重み付けする。そこで、学習プロセスに応じて W_s を変化させるために、時間的スケール $s(n)$ を導入する。式(4)に示すように、 $s(n)$ は学習プロセスに応じて変化する値であり、学習が進むごとに値を大きくする。ここで、 s_{init} は時間的スケールの初期値であり、 s_{final} は時間的スケールの最終値である。式(3)に示すように、 $s(n)$ を W_s に導入することで、学習のエポック毎に遠距離の点群に対する重みを徐々に大きくすることが可能となる。

$$W_s = 1 + s(n) \cdot \frac{d}{d_{max}} \quad (3)$$

$$s(n) = s_{init} + \frac{n}{N-1} \cdot (s_{final} - s_{init}) \quad (4)$$

時間的スケールの有無による、動的距離係数の値の変化量の例を図3に示す。時間的スケールによって動的に重みが増加し、学習プロセスに応じて近距離と遠距離の重みが増加する。損失関数に点群の学習プロセスの傾向に基づいた動的な重み付けを行うことで、精度向上を図る。

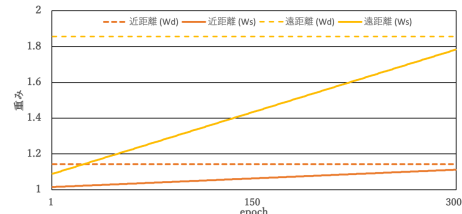


図3：近距離と遠距離のエポック毎の重み

5. 評価実験

本研究では、SalsaNextの損失関数にFocal Lossと提案手法のDWFLを用いて検出精度を比較する。また、DWFLにおける動的距離係数 W_d と W_s が精度に与える影響を調査する。さらに、セグメンテーション結果から生成した外接矩形を用いた距離別評価を行い、近距離から遠距離での精度を確認する。

5.1. データセット

本実験では、工事現場で取得した点群データセットを用いる。このデータセットは、建機に搭載したLiDARによって取得した実環境データである。クラスはSandpile, Vehicle, Humanの3つである。データセットは、学習用に350フレーム、評価用に233フレームのデータで構成されている。

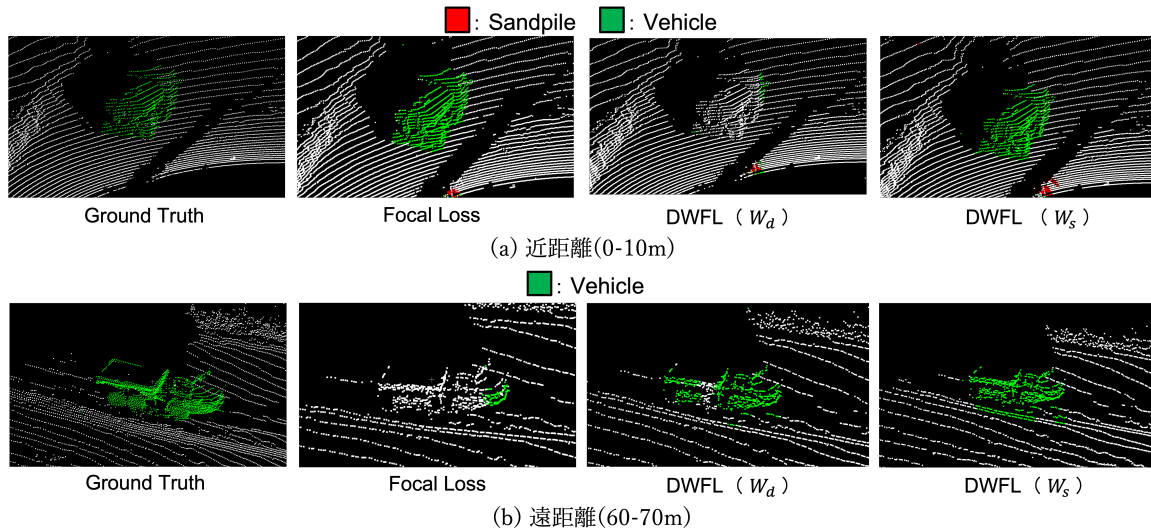


図 4：近距離および遠距離における定性的評価の比較

表 1：外接矩形を用いた距離別の精度比較

Loss	距離係数	時間的係数	0-10m	10-20m	20-30m	30-40m	40-50m	50-60m	60-70m	Average
Focal Loss			0.740	0.468	0.562	0.697	0.580	0.479	0.327	0.550
DWFL (W_d)	✓		0.629	0.419	0.512	0.685	0.596	0.491	0.390	0.620
DWFL (W_s)	✓	✓	0.731	0.438	0.570	0.695	0.596	0.486	0.371	0.648

5.2. 実験概要

本実験では学習時のネットワークに SalsaNext, 学習回数は 300 エポック, 最適化手法は Adam, バッチサイズは 4, 初期学習率は 1.0×1.0^{-4} , 時間的スケージングの初期値は 0.1, 最終値は 1 として学習する. 評価指標には IoU を用いる.

5.3. セグメンテーション結果の比較

表 2 に, Focal Loss, DWFL(W_d), DWFL(W_s) の各手法を用いた場合の IoU を示す. 表 2 から, DWFL(W_s) は Focal Loss と比較して全クラスで精度が向上している.

表 2：IoU による定量的評価

Loss	距離係数	時間的係数	Sandpile	Vehicle	Human	Average
Focal Loss			19.0	27.5	14.3	20.3
DWFL (W_d)	✓		23.6	26.2	14.6	21.5
DWFL (W_s)	✓	✓	23.5	27.7	14.8	22.0

図 4 に, Focal Loss, DWFL (W_d), DWFL (W_s) を用いた近距離および遠距離でのセマンティックセグメンテーション結果を示す. 遠距離では, DWFL(W_d) と DWFL(W_s) のどちらの場合も, Focal Loss よりも高い検出精度である. DWFL (W_d) は学習初期から遠距離の重みが大きいため高精度である. 一方で, DWFL (W_s) は学習プロセスに応じて遠距離の重みを徐々に大きくするため, 最終的には DWFL(W_d) と同精度となっている. 近距離では, DWFL (W_s) は Focal Loss と同等の精度を維持できているのに対し, DWFL (W_d) は学習初期から遠距離の重みを大きくするため, 近距離が十分に重視されず精度が低下している. 以上の結果から, DWFL (W_s) は時間的スケージングを用いることで, 学習プロセス考慮した動的重み付けが行われ, 全体的な認識精度のバランスを最適化できている.

外接矩形を用いた Precision の距離別の評価結果を表 1 に示す. DWFL(W_d) は, 近距離での精度が Focal Loss よりも低下している. これは, DWFL(W_d) の場合, 近距離が学習過程で十分に学習されていないためと考えられる. 一方で, DWFL(W_s) の場合, Focal Loss と同等の精度であり, 近距離での検出精度が高い. これは, 学習後期で近距離の精度が高くなった段階で遠距離に重み付けしているため, 近距離の精度が維持できていると考える. 次に, 遠距離では DWFL(W_s) が Focal Loss よりも高い精度を示している. これは, DWFL(W_s) は時間的スケージングを用いることで, 学習プロセスを考慮し, 学習後期に遠距離の重みを大きくしたためである. DWFL (W_d) は学習初

期から遠距離に重み付けするため DWFL(W_s) の方が精度が低い.

6. 学習エポック毎の精度比較

Focal Loss と DWFL(W_s) の学習エポック毎における精度を図 5 に示す. 近距離 (0-10m) の DWFL(W_s) は, Focal Loss より緩やかに精度向上し, 最終的には同等の精度である. 一方, 遠距離 (60-70m) では, 学習中期から後期にかけて DWFL の精度が Focal Loss よりも大幅に向上した. 以上の結果から, 近距離と遠距離間の学習バランスが最適化されたことを確認した.

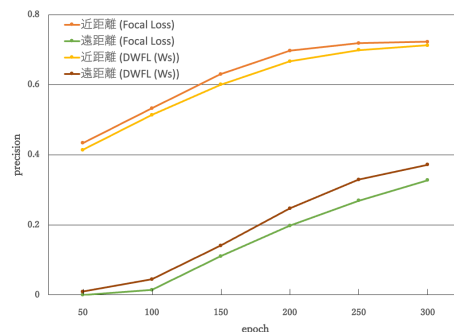


図 5：学習エポック毎の精度

7. おわりに

本研究では, 点群データを用いたセマンティックセグメンテーションにおいて, 距離依存性と学習プロセスを考慮した損失関数を提案した. 提案手法の評価実験では, 近距離の精度を維持しつつ, 遠距離での精度向上を確認した. 今後は, 他の損失関数を用いて実験を行い, 提案手法が幅広く活用できることを検証する.

参考文献

- [1] Cortinhal *et al.*, “SalsaNext: Fast, Uncertainty-Aware Semantic Segmentation of LiDAR Point Clouds”, ACM, 2020.
- [2] Lin *et al.*, “Focal Loss for Dense Object Detection”, ICCV, 2017.

研究業績

- [1] 中村智也 等, “点群データに対するセマンティックセグメンテーションにおける CG データによる事前学習の有効性検証”, 土木学会, 2022.