

1. はじめに

安全な自動運転を実現するには、走行シーンの危険度を予測し、その状況に合わせて制御することが重要である。Chanらが提案した危険度を予測する手法[1]では、画像から検出した車やバイク、トラックなどの物体の位置や大きさから危険度を予測している。しかし、人や自動車などの物体がないシーンにおいても壁などの静的物体により危険な場合がある。本研究では、前方の3次元情報から現シーンでの危険度を推定する手法を提案する。

2. 提案手法

本研究では、急ブレーキが発生するような危険なシーンを予測するため、フレーム間の速度変化を危険度とする。また、危険度の予測を実現するにあたり、3次元情報の入力方法、危険度予測を行うネットワーク構造について検討する。

2.1. ネットワーク構造

3次元情報の入力は、3次元点群または、3次元点群から作成した距離画像とする。ネットワークには、Convolutional Neural Network(CNN)、および時系列情報を考慮する構造として Long Short-Term Memory(LSTM)[3] と 3D Convolution を用いる。これらを組み合わせた以下のネットワークを提案する。

LSTM for PointCloud

3次元点群を入力とし、図1のように PointNet[2] をもにした CNN と LSTM から構成されるネットワークである。LSTM を導入することで前時刻を考慮した予測ができる。

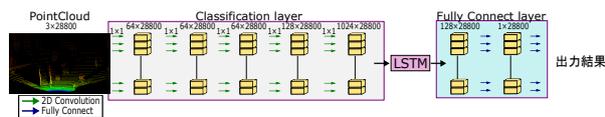


図 1: LSTM for PointCloud

LSTM for DepthMap

距離画像 1 フレームを入力とし、図2のように CNN と LSTM から構成されるネットワークである。

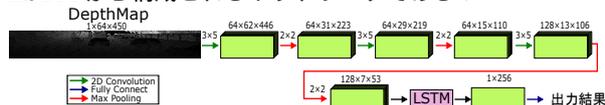


図 2: LSTM for DepthMap

CNN for DepthMap

LSTM の代わりに時系列の距離画像 3 フレームを入力とし、図3のように CNN に入力することで時間変化を考慮した危険度予測を行うネットワークである。

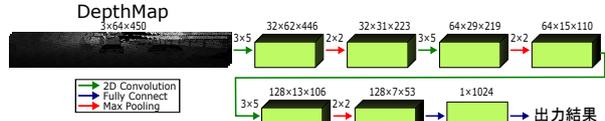


図 3: CNN for DepthMap

3DCNN for DepthMap

LSTM の代わりに時系列の距離画像 30 フレームを入力とし、図4のように CNN に入力し、3D Convolution を行うことで時間変化を考慮した危険度予測を行うネットワークである。

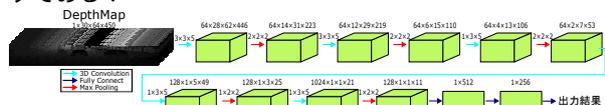


図 4: 3DCNN for DepthMap

2.2. 誤差関数

誤差関数に Mean Squared Error(MSE) と Large Difference Error(LDE) を用いる。LDE は危険度の高い部分の誤差を強調する誤差関数であり、教師ラベルを t_i 、出力を o_i 、フレーム番号を i とした場合、式 (1) に示す。

$$E_i = \begin{cases} |t_i - o_i|^2 \times (1 + |t_i|) & (t_i \geq 0.5) \\ |t_i - o_i| & (t_i < 0.5) \end{cases} \quad (1)$$

3. 評価実験

3次元情報を用いた危険度予測の有効性を確認する。

3.1. データセット

データセットは、Velodyne HDL-64e を用いて撮影した走行シーンの 3次元情報を使用する。本実験では、水平方向の前方-45度 ~ 45度のデータを使用するため、1フレームあたり 28,800点となる。使用するデータ数は学習に 53,222フレーム、テストに 5,985フレームである。

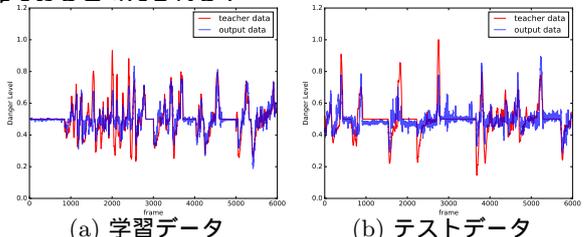
3.2. 実験結果

ネットワーク、および誤差関数の変更による評価結果を表1に示す。評価結果から誤差関数に MSE を用いた CNN for DepthMap が最も高精度であった。

表 1: 危険度予測の誤差結果

ネットワーク	誤差関数	平均誤差
LSTM for PointCloud	MSE	0.1446
	LDE	0.1023
LSTM for DepthMap	MSE	0.0781
	LDE	0.0876
CNN for DepthMap	MSE	0.0489
	LDE	0.0540
3DCNN for DepthMap	MSE	0.0706
	LDE	0.0676

CNN for DepthMap(MSE) の学習データに対する推定結果を図5(a)に、テストデータに対する推定結果を図5(b)に示す。図5(b)より、急激に上昇する危険度を予測することが出来ていない。これは、図6に示すように危険度が高いにも関わらず前方に物体がないため、危険度の予測が困難であると考えられる。



(a) 学習データ

(b) テストデータ

図 5: 評価結果

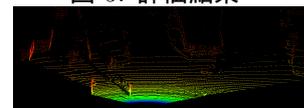
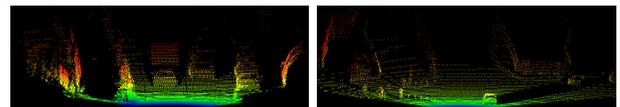


図 6: 最も危険度が高いシーン

図7(a)に危険度の高いフレームの点群、図7(b)に危険度の低いフレームの点群を示す。図7(a)では前方に車両があるため、危険度が高く、図7(b)は前方に車両がなく、危険度が低くなっていることがわかる。



(a) 危険度 (高)

(b) 危険度 (低)

図 7: フレームの可視化例

4. おわりに

本研究では、3次元情報を用いた危険度予測を行う手法を提案した。今後の予定として、データの増加・前処理の導入、またネットワーク構造の調整を検討する。

参考文献

- [1] F. Chan, *et al.*, "Anticipating Accidents in Dashcam Videos", ACCV, 2016.
- [2] C. Qi, *et al.*, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation", CVPR, 2017.
- [3] S. Hochreiter, *et al.*, "Long Short-Term Memory", Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.