深層強化学習によるモデル選択機構を導入した一貫学習ベースの自動運転

EP17042 倉橋利空　　指導教授：山下隆義

1. はじめに

一貫学習による自動運転は，車載カメラ画像を CNN に入力し制御値を出力することで，車両制御を実現している [1]．また，様々なシーンに対応した自動運転制御を実現するために，車線追従や衝突回避のタスクごとに学習した制御モデルから，運転シーンに応じてモデルをルールに基づいて選択する手法が提案されている．しかし，ルールベースの手法では，ルールから逸脱するイレギュラーな場面への対応が困難である．

本研究では，イレギュラーな場面でも適応的にモデルを選択するために，深層強化学習によるモデル選択機構を導入した手法を提案する．

2. モデル選択による自動運転制御

ルールに基づくモデル選択による自動運転制御は，個別に用意した制御モデルを予め設定したルールにより選択するアプローチである．この手法は，制御値推定を行う制御モジュールと，ルールベースによりモデル選択を行う選択機構から構成される．制御モジュールには，車線追従と衝突回避を目的とした 2 つの制御モデルを用いる．選択機構では，自動車や歩行者の位置情報をバウンディングボックス (BB) として取得し，BB の大きさが閾値以下の場合は車線追従モデル，BB の大きさが閾値以上の場合は衝突回避モデルを選択する．これにより，運転シーンに応じた車両制御を行うことができる．しかし，自動車や歩行者の検出を失敗すると，誤って車線追従モデルが選択され，前方の自動車や歩行者に衝突してしまう．モデル選択の手法は，このような物体検出が失敗した場面への対応が困難である．

3. 提案手法

本研究では，図 1 に示すような深層強化学習を用いて，最適な制御モデルを選択する手法を提案する．

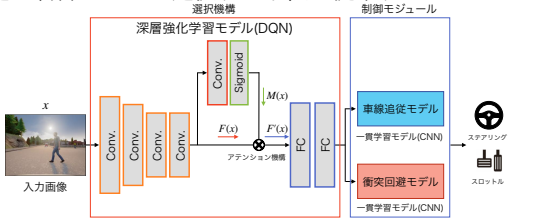

図 1：提案手法のネットワーク構造

3.1. 選択機構

深層強化学習には，Deep Q-Network(DQN)[2] を用いる．車載カメラ画像を入力とし，各モデルを選択すべき確率を出力する．報酬設計は，画像中に自動車や歩行者が一定以上の大きさで存在する場合に衝突回避モデルが選択されると+0.1，車線追従モデルが選択されると-0.1 とする．反対に，自動車や歩行者が一定以下の大きさで存在する場合に車線追従モデルが選択されると+0.1，衝突回避モデルが選択されると-0.1 とする．この報酬設計を加えることで，従来法と同様に物体の大きさを考慮した制御モデルの選択が可能となる．また，選択機構の視覚的説明を獲得するために，アテンション機構を導入する．ネットワークの注視領域を表すアテンションマップは，畳み込み層の直後に 1×1 の畳み込み層と Sigmoid 関数により獲得する．アテンション機構におけるマスク処理を式 (1) に示す．

$$F'(x) = F(x) \cdot M(x) \quad (1)$$

これにより注視領域を考慮したモデル選択が可能となる．ここで x は入力画像，$F(x)$ は畳み込み層の特徴マップ，$M(x)$ はアテンションマスク，$F'(x)$ はマスク処理後の特徴マップである．

3.2. 制御モジュール

制御モジュールは，車線追従モデルと衝突回避モデルから構成されている．車線追従モデルは自動車や歩行者が存在しないシーンで用いる．一方で，衝突回避モデルは自動車や歩行者が存在するシーンで用いる．

4. 評価実験

提案手法の有効性を評価実験により検証する．

4.1. 実験概要

提案手法の有効性を示すため，物体情報を報酬設計に導入しない選択機構 (DQN) と比較する．共通の報酬設計として，走行車線を越えるか障害物に衝突したら-1 とする．DQN の学習条件は，エピソード数を 500，エピソード終了条件を 1000 フレーム経過とする．走行実験では，CARLA の Town01，Town02 上に車 20 台，歩行者 100 人を出現させ，6 分間の走行を 20 回試行した際の自律性 ($autonomy$) を評価する．人間の介入開始から自律走行再開までの時間を 6 秒と仮定すると，自律性は式 (2) で算出できる．

$$autonomy = (1 - \frac{intervention * 6seconds}{elapsed\ time}) * 100 \quad (2)$$

ここで，人間の介入回数を $intervention$，走行時間を $elapsed\ time$ とする．

4.2. 実験結果

表 1 に評価結果を示す．提案手法では，ステアリングの精度は DQN より約 10 ポイント精度が高く，ルールベースと同程度の運転精度である．また，物体情報を 10%欠落させた環境での評価では，ルールベースはスロットルの精度が大きく低下した．一方で，提案手法のスロットル精度は低下しない．以上より，物体検出が失敗したシーンにおいて提案手法は有効である．

表 1：自律性の評価結果 [%]

環境	選択機構	ステアリング		スロットル	
		Town01	Town02	Town01	Town02
通常	ルールベース	**82.7**	67.8	89.1	**84.1**
	DQN	73.8	55.3	88.2	83.4
	提案手法	82.5	**68.2**	**89.4**	82.2
物体情報 10%欠落	ルールベース	**82.8**	68.2	78.6	75.2
	DQN	73.8	55.3	88.2	**83.4**
	提案手法	82.5	**68.4**	**89.4**	82.2

4.3. 注視領域の可視化による視覚的説明

図 2 に実際の走行シーンの可視化例を示す．図 2(a) は，車線上にアテンションが反応し，車線追従モデルを選択している．図 2(b) は，車線上に加えて前方の車両下部に反応し，衝突回避モデルを選択している．以上より，選択機構のモデルは前方の動的物体の有無から選択する制御モデルを決定していると考察できる．

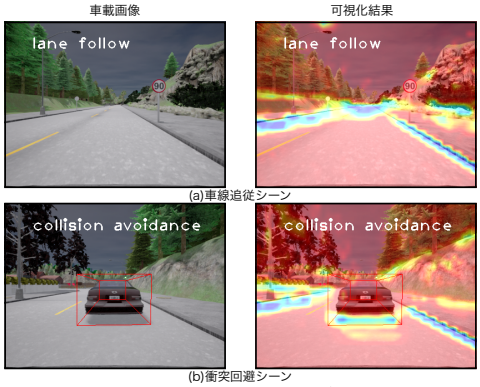

図 2：走行シーンの可視化例

5. おわりに

本研究では，選択機構に深層強化学習を導入した自動運転制御の提案と，深層強化学習のモデル判断根拠の調査を行った．今後は，新たな制御モデルを生成し，より多様なシーンに対応する自動運転の実現などに取り組む．

参考文献

[1] M. Bojarski, *et al.*, “End to End Learning for Self-Driving Cars”, arXiv preprint, arXiv:1604.07316, 2016.

[2] V. Mnih, *et al.*,“Human-level control through deep reinforcement learning”, Nature, 2015.