

### 3-1 End-to-end 自動運転のためのマルチタスク学習と注意機構付き深層学習 [Ishihara 2021]

**End-to-end 自動運転**：従来の自動運転システムでは、環境認識、経路計画、車両制御の3つのモジュールを個別に構成し、それらを繋げることによって自動運転を実現していた。近年の深層学習の発展に伴い、カメラやLiDAR（レーザ距離センサ）などのセンサから得たデータを用いて車線、他車、歩行者などを認識する手法が格段に進歩している[Natan 2022]。これに対し、End-to-end 自動運転では、センサ入力から直接車両制御コマンド（アクセル開度、ブレーキ強度、ハンドル角）を出力するためのモデルを、深層学習を用いて獲得する。

**マルチタスク学習と注意機構**：単純な深層学習では、物体検出、セマンティックセグメンテーション（画像のピクセルごとの物体クラス分類）、車両制御コマンド生成など、入力に対し一つのタスクを行が、マルチタスク学習では、同じ入力に対し複数のタスクを行う。これにより、複数タスク間に共通して有効な特徴を効率的に学習することや、単一タスクよりも高い汎化性能（多様な状況に対応する能力）を得ることができる。一方、注意機構はセンサ入力のうち、タスクに有効な部分をより重要とみなすように学習を行うためのものである。

**提案するネットワーク構造**：図3-1-1に提案するネットワークの概要を示す。画像から深度画像、セマンティックセグメンテーション画像、信号機の状態を推定するタスクを行い、メインのタスクとして画像と現在の車両速度から制御コマンドを生成する。図中 Type 1, Type 2 と示されているモジュールは注意機構であり、CBAM [Woo 2018]と呼ばれるモデルをベースに構築されている。

**データセット**：データセットはCARLA 都市環境運転シミュレータ [Dosovitskiy 2017]を用いて作成した。2つの都市データが提供されており、Town01のデータで学習し、Town02の環境でテストを行った。図3-1-2に生成された画像の例を示す。画像の生成と同時にマルチタスク学習の出力値を正解値として取得しており、それらを用いてネットワークの学習を行う。

**実験結果**：2つのベンチマークデータで評価した。CoRL2017 ベンチマークは直進走行、交差点での右左折を一つ含む、複数の右左折を含む長距離の走行（他の自動車や歩行者あり、なし）の4種の課題からなる。NoCrash はさらに交通量や天候の変化を考慮するとともに、衝突のペナルティを重く評価している。提案手法では、走行の成功率では従来の手法と同等かそれ以上を示し、さらに赤信号無視の回数を大幅に減らすことができている。このことはマルチタスク学習と注意機構の組合せが End-to-end 自動運転に有効であることを示している。図3-1-3は Grad-CAM [Selvaraju 2017]と呼ばれる手法を用いて、提案手法が、注意機構によって重要と考えた場所を可視化したものである。前方の自動車、車線、赤信号など画像中の重要な場所に注意が向けられていることがわかる。

[Ishihara 2021] K. Ishihara, A. Kanervisto, J. Miura, V. Hautamäki, "Multi-Task Learning with Attention for End-to-end Autonomous Driving", Proc. CVPR2021 Workshop on Autonomous Driving, Jun. 2021.

[Natan 2022] O. Natan and J. Miura, "Towards Compact Autonomous Driving Perception with Balanced Learning and Multi-sensor Fusion", IEEE Trans. on Intelligent Transportation Systems, 2022 (to appear).

[Woo 2018] S. Woo, J. Park, J.-Y. Lee, I.-S. Kweon, "CBAM: Convolutional Block Attention Network", Proc. ECCV2018.

[Dosovitskiy 2017] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, V. Koltun, "CARLA: An Open Urban Driving Simulator", Proc. 1<sup>st</sup> Annual Conf. on Robot Learning, 2017.

[Selvaraju 2017] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", Proc. ICCV-2017.

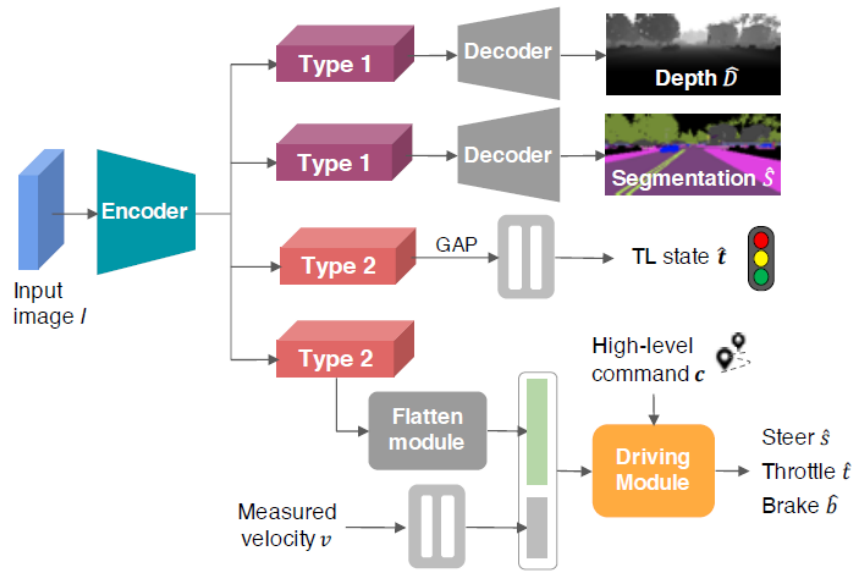
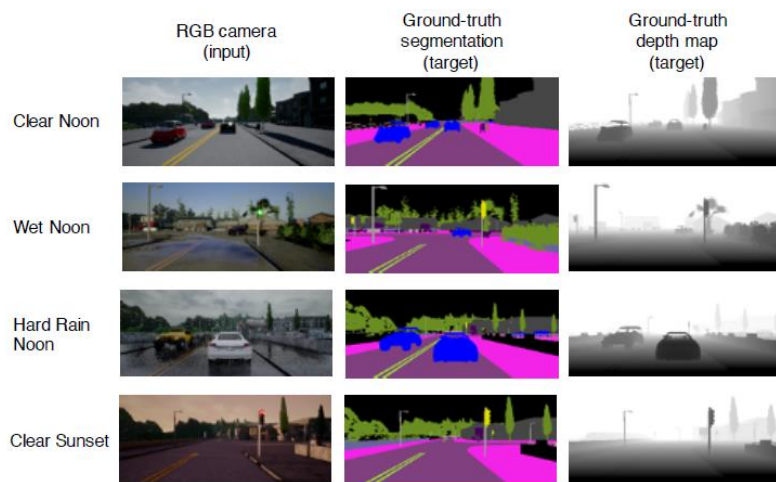
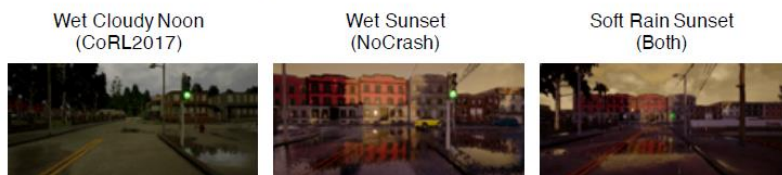


図 3-1-1: 提案するネットワーク構造



(a) Training weather conditions in Town01



(b) New weather conditions in Town02

図 3-1-2: CARLA で生成した画像の例



図 3-1-3: 注意箇所の視覚化