

## 1. はじめに

一貫学習による自動運転制御では、Convolutional Neural Network (CNN) に走行画像を入力し、ステアリング値を出力することで、車両制御を実現している [1]。その一方で、CNN が何を根拠に出力を決定したか不明確という問題がある。そこで、本研究では一貫学習による自動運転制御において、出力の判断根拠を解釈するため、Weighted Global Pooling を導入した Attention Branch Network (ABN) を提案し、ネットワークが注視する領域の可視化を行う。

## 2. Attention Branch Network

Attention Branch Network (ABN)[2] は、画像認識において CNN が着目している領域を示す Attention map を可視化しつつ、識別に利用する手法である。ABN は、入力画像から特徴を抽出する Feature extractor, Attention map を獲得する Attention branch, 得られた特徴マップと Attention map から最終的な認識結果を出力する Perception branch の 3 つのモジュールから構成されている。Attention branch の出力層には Global Average Pooling (GAP) を使用している。注視領域は、出力層手前の畳み込み層から得た特徴マップを平均することで Attention map を獲得する。得られた Attention map を Perception branch に入力することで、認識に有効な特徴のみに着目できる。

## 3. ABN の自動運転モデルへの応用

本研究では、一貫学習における自動運転制御モデル [3] に ABN の Attention branch 機構を導入し、判断根拠の解釈を試みる。ベースとなるネットワークモデルは、全結合層に車両速度を入力することで、車体速度を考慮したステアリング値とスロットル値の回帰を行う。

本モデルに ABN の Attention branch 機構を導入した構造を図 1 に示す。従来の Attention branch の出力層には、GAP が用いられている。GAP は特徴マップ全体の平均を求めるため、特徴マップ内の着目すべき位置情報が欠落する。そこで、式 (1) のように CNN のカーネルを用いた重み付けによりプーリングする Weighted Global Pooling (WGP) を提案する。

$$v^c = \tanh \left( \sum_{i=1}^h \sum_{j=1}^w \alpha_{j,i}^c \cdot f_{j,i}^c(x) \right) \quad (1)$$

WGP は、畳み込み層で得られた特徴マップ  $f^c(x)$  の各要素に対して重み値  $\alpha^c$  を乗算し合計を回帰値として出力する。これにより、注視領域に対して高い重み値になるよう学習することができる。

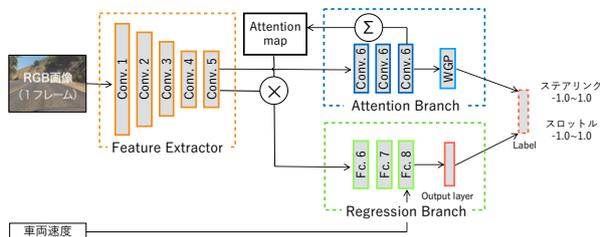


図 1: 提案する自動運転制御のネットワーク構造

## 4. 評価実験

提案手法による一貫学習の自動運転制御の有効性を評価するために、精度評価と Attention map による視覚的説明を行う。

### 4.1. 実験概要

データセットには、コンピュータゲームの Grand Theft Auto V (GTAV) をシミュレータとして使用する。ゲーム上の車両を人間が運転し、フロントカメラ映像を每秒 10 フレーム分取得する。本実験では、学習に 30,000 フレーム、評価に 16,556 フレーム使用する。

### 4.2. ステアリングおよびスロットルの精度比較

Attention branch 有りとなしきの自動運転制御モデルによるステアリングおよびスロットル値を平均二乗誤差により比較する。表 1 より、ステアリング、スロットル共にその差は 0.001 未満であり、Attention branch の導入による性能劣化への影響がないことを確認した。

表 1: 平均二乗誤差による精度比較

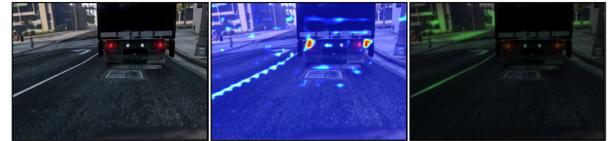
	ステアリング	スロットル
CNN のみ	0.014810	0.019124
Attention branch あり	0.015480	0.019331

### 4.3. Attention map の可視化による視覚的説明

図 2 に入力に対する可視化結果を示す。図 2(a) は、ステアリングが正の値の場合、右へ舵を切っている。中央図の Attention map より、左側の車線上に反応しており、ステアリング値は車線の位置に影響されると解釈できる。図 2(b) は、停止状態であり、スロットル値が 0.00 で、加速も減速もされていない状態である。アテンションは前方車両のブレーキランプに強く反応しており、スロットルの判断では前方車両の状態を重要視すると解釈できる。一方、右図の比較手法 [1] の注視マップでは、反対側の車線や前方車両以外を注目しており、判断根拠の解釈が困難である。



(a) 右へ舵を切る (ステアリング:0.18, スロットル:0.07)



(b) 停止 (ステアリング:0.08, スロットル:0.00)

図 2: 可視化した Attention map の例

Attention map と領域毎にキャプションを生成する手法 [4] を組み合わせることで、出力の判断根拠に適したキャプションを選択できる。図 3 の例では、スロットルの値が 0.05 から -0.15 に変化し、車両が停止しようとしている。ここでは図 3(b) の Attention map に対して、図 3(c) のように “man on a red motorcycle.” が選択された。

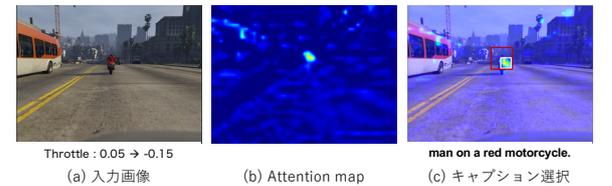


図 3: 選択されたキャプションの例

## 5. おわりに

本研究では、一貫学習における自動運転制御モデルへ ABN を応用し、Attention map から判断根拠の視覚的説明の可能性を示した。今後は、提案したアプローチをベースに、判断根拠の言語的説明を行う手法を構築する。

### 参考文献

- [1] M. Bojarski, *et al.*, “Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car,” arXiv preprint, arXiv:1704.07911, 2017.
- [2] H. Fukui, *et al.*, “Attention Branch Network: Learning of Attention Mechanism for Visual Explanation,” arXiv preprint, arXiv:1812.10025, 2018.
- [3] 村瀬卓也 等, “自己状態を付与した CNN による自動運転制御の高精度化,” パターン認識・メディア理解研究会, 2017.
- [4] J. Johnson *et al.*, “DenseCap: Fully Convolutional Localization Networks for Dense Captioning,” CVPR, 2016.