

## ニューラルネットワークを用いた自己運動の認識 Recognition of self-movement by neural network

青木 拓馬<sup>†</sup>  
Takuma Aoki

伊藤 浩<sup>†</sup>  
Hiroshi Ito

### 1. はじめに

近年、車やロボット、ドローンなどの多様な移動体が普及しつつある。移動体の動きを制御するためには自己運動を認識することが有効であり、加速度は移動体が受ける力を表すため、特に重要な要素である。従来、車の自動運転のため、センサデータを使用して加速度を推定する研究がある [1]。しかし、ヒトや動物は自己運動の大部分を目の情報から認識していると考えられる。

そこで本文では、一人称視点で撮影した動画からニューラルネットワークを用いて加速度を推定する方法を提案する。データセットは自動運転の評価に用いられる Waymo Open Dataset [2] を用いた。モデルとして、CNN+LSTM モデルと 3DCNN モデルを設計し、それぞれを比較評価した。

### 2. モデル

運動している移動体の加速度を一人称視点の画像 1 枚から推定することは困難である。そこで、まずシーン中の移動体の速度を抽出するために、オプティカルフローの手法を用いて各画素の方向ベクトルを計算する。図 1 は前進している車から取得した画像とそのオプティカルフローを可視化した画像である。この画像では、左向きのベクトルを緑、右向きのベクトルを赤色として表示している。前進している場合、オプティカルフローはこの図のように、画像の左半分は左向き、右半分は右向きの方向ベクトルが計算される。



図 1: Optical flow

#### 2.1 CNN+LSTM モデル

自己運動の認識には、特徴抽出と時系列処理の 2 段階に分けて行うことが有効である [3]。図 2 にモデルの構造を示す。このモデルはまず、オプティカルフローから速度の特徴量を抽出する CNN (ResNet 18) を持つ。ResNet の最終層のニューロンを 128 個に変更し、その出力を速度の特徴ベクトルとした。これに全結合層を付加して出力層を 1 つのニューロンとし、速度の値を出力する。また、特徴抽出器で得られた大きさ 128 のベクトルを LSTM によって時系列的に処理することで、フレーム毎の加速度の推定を行う。

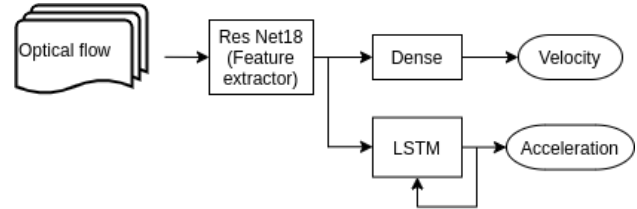


図 2: CNN+LSTM model

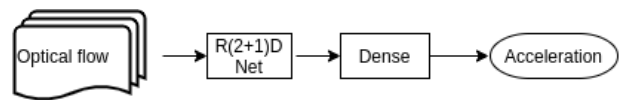


図 3: 3DCNN model

#### 2.2 3DCNN モデル

このモデルは、動画認識に用いられる CNN を 3 次元に拡張した 3DCNN を加速度の推定に用いたものである。R(2+1)D [4] は、空間方向に畳み込みを行った後、時間方向に畳み込みを行うモデルで、ここでは、出力層のニューロンを 1 つに変更した。複数フレームのオプティカルフローに対して 3 次元畳み込み処理を行い、最後のフレームの加速度の推定を行う。この処理を 1 フレームずつずらし、フレーム毎の加速度を推定する。図 3 にこのモデルを示す。

### 3. 実験

Waymo Open Dataset を用いて車両の加速度の推定を行った。データセットには、グローバル座標での速度の各成分の値がフレーム毎に与えられており、ここでは、そのノルムを計算してスカラー量を求め、これを進行方向の速度として用いる。加速度は、この速度を経過した時間で割った値とした。学習用データセットは走行中の約 20 秒のシーンが 798 シーンあり、総フレーム数は 157,283 である。テストデータは 150 シーン、総フレーム数は 29,497 である。入力データは、縦横 2 チャネルのオプティカルフローを  $144 \times 256$  にリサイズしたものを使用した。

#### 3.1 CNN+LSTM モデル

CNN で車両の速度を推定するよう学習した。学習するパラメータ数は 11,239,169、学習条件として、損失関数を MSE、最適化アルゴリズムを Adam、ミニバッチのサイズを 128 とする。

図 4 は停止している状態から前方に加速するテストデータの 1 シーンの速度の推定結果である。テストデータ全体での MSE は 0.831 となり、高い精度で推定を行うことができた。

図 5 は、速度の特徴ベクトルをフレーム毎に横に並べ

<sup>†</sup> 日本大学, Nihon University

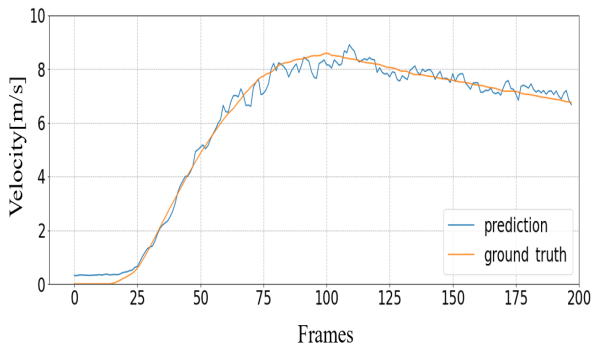


図 4: Prediction of velocity by CNN

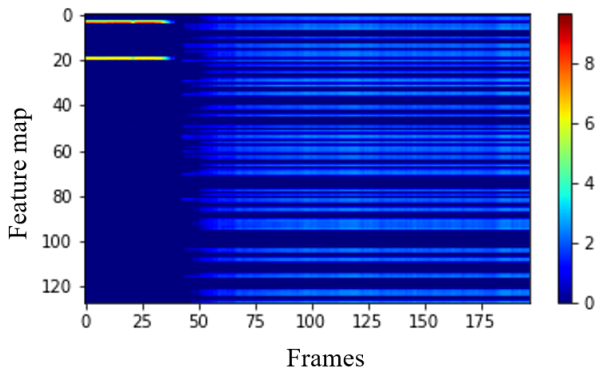


図 5: Feature map of velocity

た図である。横軸は図4と対応しており、停止している状態のとき、3番目と20番目のニューロンが反応し、動き出すとともに他のニューロンが全体的に反応していることがわかる。以上のことから、速度に関する特徴が抽出できていることが確かめられた。

次に、LSTMで車両の加速度を推定するよう学習した。学習するパラメータ数は264,321、学習条件として、損失関数をMAE(Mean Absolute Error)、最適化アルゴリズムをAdam、ミニバッチのサイズを4とする。学習において、LSTMの内部状態が不安定である最初の20フレームのバックプロパゲーションは行っていない。図6の青のグラフは、CNN+LSTMによって、加速度を推定した結果である。停止している状態から加速する時の立ち上がりが遅いが、加速度の概形が捉えられていることがわかる。

### 3.2 3DCNNモデル

入力するオプティカルフローのフレーム数を16フレームとして3DCNNを学習した。学習するパラメータ数は314,990、学習条件として、損失関数をMSE、最適化アルゴリズムをAdam、ミニバッチのサイズを16とする。図6の緑のグラフは3DCNNで推定した結果である。LSTMで生じた立ち上がりの遅れは生じず、加速度の概形が捉えられていることがわかる。

### 3.3 比較

CNN+LSTMと3DCNNのテストデータ全体に対する推定誤差(MAE)は表1のようになった。どちらのモ

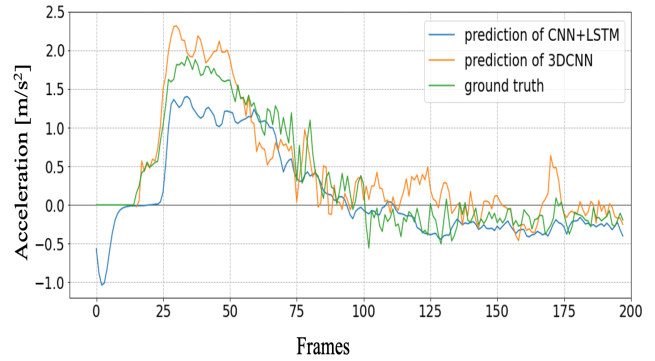


図 6: Prediction of acceleration

デルも同程度の推定精度となった。CNN+LSTMモデルは過去の情報を記憶できるため、フレーム毎の乗加算回数は抑えられる。一方、3DCNNモデルは毎回複数フレーム分の乗加算が必要となる。しかし、3DCNNモデルは逐次的に加速度を計算しているため、CNN+LSTMモデルと比較して静止している状態から動き出す時の加速度の変化に追従しやすいと考えられる。

表 1: MAE of test data

Models	MAE
CNN + LSTM	0.308
3DCNN	0.303

## 4. まとめ

一人称視点の動画のみを用いて移動する車両の加速度を推定する手法を提案した。実験の結果、オプティカルフローを入力として、CNN + LSTMモデルと3DCNNモデルを用いた自己の加速度の推定が有効であることが示せた。

今回は、進行方向の加速度のスカラー量のみを推定を行った。今後は、回転運動である角速度の推定を行う。また、今回は入力にオプティカルフローを与えたが、RGB画像を入力にした方法も比較する。

## 参考文献

- [1] Z. Gu, et al., "An LSTM-Based Autonomous Driving Model Using Waymo Open Dataset," arXiv:2002.05878, 2020.
- [2] Waymo-Team, Waymo Open Dataset. <https://waymo.com/open/>. Accessed: 2020-12-15.
- [3] 神谷 他, "Convolutional-Recurrent Neural Networkによる自己運動識別," 画像センシングシンポジウム, 2016.
- [4] D. Tran, et al., "A Closer Look at Spatiotemporal Convolutions for Action Recognition," arXiv:1711.11248, 2018.