

## 1. はじめに

歩行者検出は、現時点における歩行者の位置を求める問題であり、自動運転の安全性を高める上で重要な技術である。一方で、観測したフレーム時刻の歩行者の位置だけでなく、未来フレームでの歩行者の経路を予測することが、高度な自動運転の実現に求められている。そこで本研究では、車載カメラ映像から歩行者の目的地を予測する手法を提案する。

## 2. 提案手法

ニューラルネットワークを用いて確率分布を出力する手法として、Mixture Density Network (MDN)[1]が提案されている。本研究では、人の動きを考慮するために再帰的な構造を加えた Recurrent Mixture Density Network (RMDN)を導入し、求めた確率分布から目的地の予測を実現する。

### 2.1. Recurrent Mixture Density Network

提案手法で用いる RMDN のネットワーク構造を図 1 に示す。まず、車載カメラ映像のフレームを Convolutional Neural Network (CNN) へ入力し、特徴量を抽出する。抽出した特徴量と世界座標系での歩行者位置を Long Short-Term Memory (LSTM) へ入力する。LSTM を用いることで、時系列変化に頑健な予測を可能とする。RMDN に各フレームの画像を逐次入力すると、10 フレーム後に対象が到達する目的地の確率分布を求める。

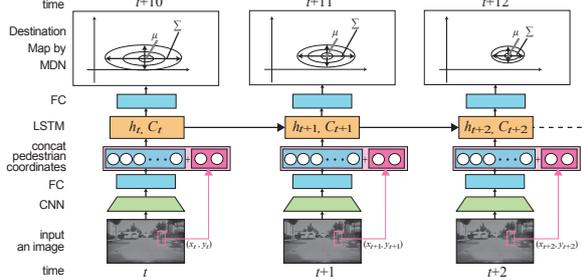


図 1：提案手法のネットワーク構造

世界座標系での歩行者の位置を  $\mathbf{x}$ 、歩行者の進行方向を  $\psi$  としたとき、目的地の確率分布は式 (1) で表すことができる。

$$p(\mathbf{x}, \psi) = \sum_{i=1}^N \pi_i \frac{1}{\sqrt{|\Sigma_i|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right) \frac{1}{2\pi I_0(\kappa_i)} \exp(\kappa_i \cos(\psi - \gamma_i)) \quad (1)$$

ここで  $\boldsymbol{\mu}_i$  及び  $\Sigma_i$  はそれぞれ 2 変量正規分布の平均と共分散行列である。また、 $\gamma_i$  及び  $\kappa_i$  はそれぞれフォン・ミーゼス分布の平均の向きと尖度である。予測する目的地は 1 点であるため、1つの正規分布により目的地を予測する。よって、重ね合わせ数は  $N = 1$ 、混合する際の重みを  $\pi_i$  とする。これらのパラメータは、ネットワークの出力ユニットの値  $\{m_x, m_y, s_x, s_y, r, p, k, g\}_i$  ( $i = 1, \dots, N$ ) を用いて次のように求められる。

$$\begin{aligned} \boldsymbol{\mu}_i &= (m_{x,i}, m_{y,i})^T \\ \Sigma_i &= \begin{bmatrix} \exp(s_{x,i})^2 & \tanh(r_i) \exp(s_{x,i}) \exp(s_{y,i}) \\ \tanh(r_i) \exp(s_{x,i}) \exp(s_{y,i}) & \exp(s_{y,i})^2 \end{bmatrix} \\ \pi_i &= \frac{\exp(p_i)}{\sum_{j=1}^N \exp(p_j)} \\ \gamma_i &= g_i \\ \kappa_i &= \exp(k_i) \quad (\kappa_i > 0) \end{aligned}$$

### 2.2. 歩行者に着目した事前学習モデルの活用

歩行者に注目した特徴量を抽出することで予測精度の向上が期待できる。本研究では、Zhe ら [2] が提案する姿勢推定手法の特徴抽出部 (VGG19) を事前学習モデルとして提案手法の CNN 部分に活用する。

## 3. 評価実験

提案手法の有効性を Daimler Pedestrian Benchmark Data Sets を用いて評価する。

### 3.1. 実験概要

本実験では、68 本の動画を 10 フレームごとに分割したものを 1 サンプルとする。学習サンプルは 7,997 サンプル、テストは 3,552 サンプルである。

再帰的な構造の有無による効果を確認するために、MDN と RMDN の比較をする。また、事前学習モデルの効果を確認するために、RMDN の CNN として畳み込み層 3 層と、事前学習モデルを導入しない VGG19 を用いる手法との比較を行う。評価指標は、世界座標系での予測分布の平均と真値の誤差 [m] である。

### 3.2. 実験結果

予測精度の比較結果を図 2 に示す。MDN は RMDN と比較して、誤差が大きいことがわかる。また、RMDN において、畳み込み層 3 層を用いる手法は、VGG19 および事前学習モデルを導入する VGG19 と比べて誤差が大きい。事前学習モデルを導入すると、事前学習モデルを導入しない VGG19 に比べ、予測誤差を減少させることができた。

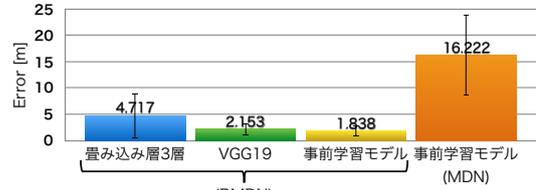


図 2：評価結果

図 3 に RMDN と MDN の予測精度の比較を示す。RMDN は、過去の経路情報を考慮することができるため、確率分布の分布が小さく、予測精度が良いことがわかる。

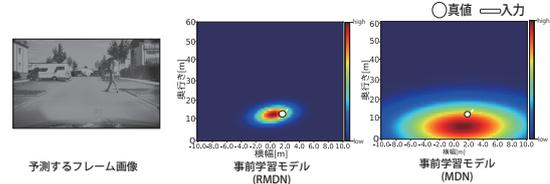


図 3：RMDN と MDN の比較

図 4 に RMDN による目的地の確率分布例を示す。畳み込み層 3 層では、真値と予測分布の平均が大きくずれている。一方、事前学習モデルは、VGG19 と比較して真値に近い目的地を予測できている。

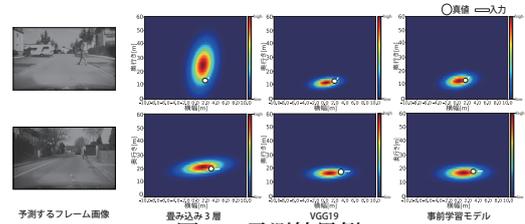


図 4：予測結果例

## 4. おわりに

本研究では、RMDN を導入した歩行者の目的地予測手法を提案した。歩行者に着目した特徴量を抽出するために姿勢推定手法の事前学習モデルを用いることで、より高精度な予測を実現した。今後は、複数の経路に対応した歩行者の目的地予測について検討する。

## 参考文献

- [1] C. M. Bishop, "Mixture Density Network", Technical Report, Aston University, 1994.
- [2] Z. Cao, *et al.*, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", CVPR, pp. 7291-7299, 2017.