

Convolutional-Recurrent Neural Network における 2 段階学習の効果

EP12038 川口俊樹

指導教授：藤吉弘亘，山下隆義

1.はじめに

Convolutional-Recurrent Neural Network(C-RNN)[1]は、DCNN[2]による特徴抽出と RNN[3]による時系列対応により高精度な時系列パターンの学習が可能である。C-RNN の学習は、DCNN の特徴抽出部と RNN の識別部を End-to-End で誤差逆伝播法により各パラメータの更新を行う。しかし、後段の RNN の誤差を前段の DCNN まで伝播できないと十分なパラメータ更新ができず、性能を発揮できない場合がある。そこで本研究では、DCNN と RNN の学習を別々に行う 2 段階学習による学習効果について調査する。

2.C-RNN の 2 段階学習と識別処理

本研究では、C-RNN の学習方法として DCNN と RNN の学習を別々に行う 2 段階学習法を提案し、その効果を調査する。2 段階学習は、特徴抽出を行う DCNN と、時系列情報を用いて学習を行う RNN を別々に学習することで、識別性能の向上を目的としている。

2.1. 2 段階学習の流れ

図 1 に C-RNN の 2 段階学習の流れを示す。まず Step1 では、DCNN の学習を通常の DCNN と同様に誤差逆伝播法を用いて行う。その後、Step2 で学習した DCNN の畳み込み層及びブーリング層を転移させて、RNN と結合する。入力する動画像の各フレームから DCNN により得られる特徴ベクトルを時系列に従い RNN に入力する。そして、通常の RNN と同様に誤差勾配を用いて RNN のパラメータを更新する。

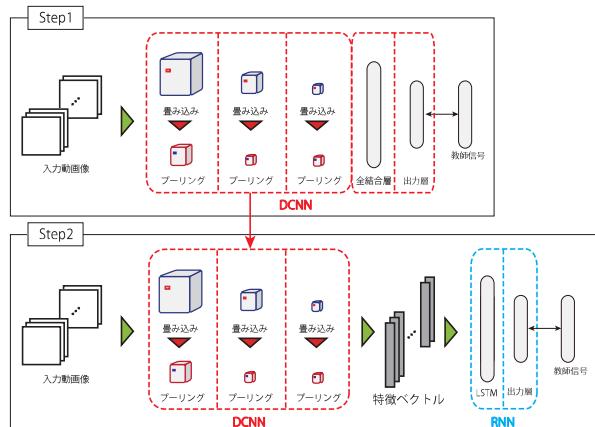


図 1 : C-RNN の学習の流れ

2.2. 識別処理の流れ

動画像を対象とした C-RNN による識別の流れを図 2 に示す。

Step1：方向ベクトルの画像化

動画像中の連続する 2 フレームから、オプティカルフローにより方向ベクトルを算出する。方向ベクトルを算出することで、背景の動きを表現することができるため、局所的な時系列情報を表すことができる。算出した方向ベクトルから、方向ベクトルの向きを色、強度を明るさとした可視化画像を生成する。方向ベクトルを画像化することで、DCNN の畳み込みに適した入力となる。

Step2：C-RNN の処理

Step1 で生成した可視化画像を入力とし、C-RNN により識別する。まず、学習済み DCNN の畳み込み層及びブーリング層により可視化画像から特徴ベクトルを生成する。その後、生成した特徴ベクトルを時系列に従い RNN へ入力し識別を行う。RNN の識別結果を最終的な C-RNN の識別結果とする。

3.評価実験

C-RNN の 2 段階学習による効果を調査するために評価実験を行う。

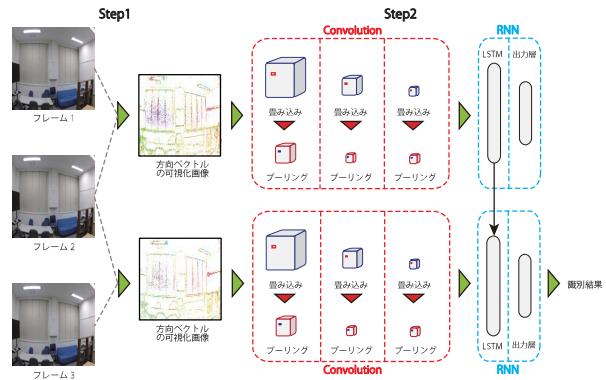


図 2 : C-RNN による識別の流れ

3.1. 実験概要

本実験では、1 人称視点のカメラで撮影された動画像から、自己運動を分類する。対象とする自己運動は、前進、後退、右回転、左回転の 4 パターンである。学習、評価には通常のカラー画像と、方向ベクトルの可視化画像を用いる。比較する手法は、2 段階学習の C-RNN と End-to-End で学習する C-RNN, DCNN, RNN の 4 種類である。

3.2. 実験結果

1 人称映像における各手法の認識率を表 1 に示す。表 1 より、可視化画像と 2 段階学習を用いた C-RNN の認識精度が最も良いことがわかる。

表 1 : 自己運動推定における各手法の認識率 [%]

手法	カラー画像	可視化画像	カラー+可視化
C-RNN(2 段階)	30.71	84.52	61.00
C-RNN	49.66	82.66	73.57
DCNN	28.47	81.52	48.90
RNN	35.71	71.71	67.57

表 1 において、DCNN と RNN の認識率を比較すると、カラー画像では RNN が高く、可視化画像では DCNN が高い。また、2 種類の C-RNN の認識率を比較すると、カラー画像では End-to-End の学習が高く、可視化画像では 2 段階学習が高い。これは、RNN が適した問題は End-to-End の学習が良く、DCNN が適した問題であれば 2 段階学習が良いということである。カラー画像+可視化画像においても同様であり、RNN の認識率が DCNN よりも高いため End-to-End の学習が 2 段階学習より認識率が高い。図 3 に、提案手法における識別結果の例を示す。



図 3 : 識別結果の例

4.おわりに

本研究では C-RNN に用いる DCNN と RNN の学習方法として、2 段階学習を導入することで得られる効果について調査した。RNN による時系列対応が効果的であれば End-to-End の学習が適しており、DCNN の特徴抽出のほうが効果的であれば 2 段階学習が適していることが判明した。今後は、C-RNN の学習時に注視点情報を与えることで、効率的な学習を検討する。

参考文献

- [1] Z.Zuo, et al, "Convolutional Recurrent Neural Networks: Learning Spatial Dependencies for Image Representation ", IN CVPR2015, 2015.
- [2] Y.LeCun, et al, "Gradient-Based Learning Applied to Document Recognition ", Proceedings of the IEEE, pp. 2278-2324, 1998.
- [3] J.Elmans, "Finding Structure in Time", Cognitive Science 14 (2): 179-211, 1990