

顧客の満足度推定のための改良型 Conv-RNNs を用いた動画からの表情認識

中京大学大学院工学研究科 ○三好 遼, 橋本 学 関西学院大学理工学部 長田典子

概要

本研究では, Convolutional Recurrent Neural Networks (Conv-RNNs) に時間および空間方向のスキップ接続を, また時間方向に新たなゲートを追加することによって改良した Conv-RNNs さらに, これを用いた動画からの表情認識手法を提案する. Conv-RNNs は, 過去のタイムステップへ勾配を伝搬することが困難である. 一方, 表情認識においては, 隣接フレーム間での変化が乏しいため, 古い情報を参照することが有効であると考えられる. そこで本研究では, Conv-RNNs に時空間方向のスキップ接続と時間方向に新たなゲートを加えることにより, 時間および空間方向における勾配消失を防ぎ, より古い有効な情報を利用することができるように改良し, この問題を改善した. 実験では, 改良型 Conv-RNNs (提案手法) と Conv-RNNs (従来手法) を比較した. また, 実験では, ConvRNNs の一般的モデルである Convolutional Gated Recurrent Unit (ConvGRU) と Convolutional Long Short-Term Memory (ConvLSTM) を改良した. 改良型 ConvGRU および改良型 ConvLSTM の認識精度はそれぞれ, 88.99%, 90.52%であり, ConvGRU および ConvLSTM に比べてそれぞれ, 2.45%, 4.59%の精度向上を確認した. 今後の方針として, 店舗における顧客の感情推定への応用を検討している.

1. はじめに

顔表情は, コミュニケーションにおいて感情や意図を伝達するための重要な非言語的情報の一つであり, ヘルスケアやモビリティなどの幅広い分野で活用されている.

顔表情は, オンセット, ピーク, オフセットの3つフェーズに分けることができる. オンセットは表情の開始, ピークは表情が最も強く表れている瞬間, オフセットは表情が消える瞬間を表す. そして, 顔表情は, オンセットからオフセットまでの変化によって表現される. これらの動的情報は, 人が微細な表情を認識する際に活用されていることが心理学分野において明らかになっている[1]. そのため, 表情認識システムにおいても顔の動的情報を利用することが有効であると考えられる.

近年, 動画からの顔表情認識では, Deep Neural Networks (DNNs) を用いた手法が高い認識精度を示している. DNNs 手法は 3 つに大別することができる. 1 つ目は, Convolutional Neural Networks (CNNs) と Recurrent Neural Networks (RNNs) を組み合わせた手法である. この手法では, 空間的特徴と時間的特徴が別のモジュールによって抽出されているため, 時空間の関係を同時に表現した特徴を得ることができない. 2 つ目は, 3D CNNs を用いた手法である. 3D CNNs は大規模なデータセットの公開を背景に行

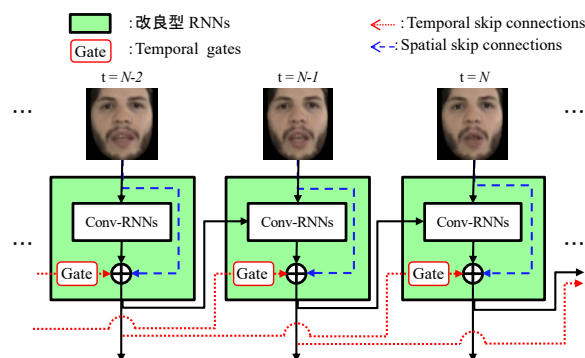


図1 改良型 Conv-RNNs の概要

動認識において有効性が示されている. しかし, 表情認識においては大規模なデータセットがないため, 3D CNNs の大量のパラメータを学習することが難しい. 3 つ目は, Convolutional RNNs (Conv-RNNs) を用いた手法である. この手法は, 時空間的特徴を得ることができ, 3D CNNs よりパラメータ数が少ない. しかし, RNNs は時間方向に展開されるため, その方向において勾配消失が起きやすい. また, 直前の情報の影響を強く受けるため, 過去の情報を保持することが難しいという問題がある.

本研究では, 心理学分野で示唆されている表情変化の特性の利用と Conv-RNNs の問題を改善した, 改良型 Conv-RNNs (図 1) および, それを用いた表情認識手法 (図 2) を提案する.

2. 改良型 Conv-RNNs ・提案する表情認識手法

2.1 改良型 Conv-RNNs

図 1 に改良型 Conv-RNNs の概要を示す. 改良型 Conv-RNNs は, 表情の変化をとらえるためのゲート (Temporal gates) と勾配消失を抑制する skip connection (Temporal skip connections, Spatial skip connections) を従来の Conv-RNNs に加えることによって改良された. Temporal gates は, タイムステップ $t-2$ の情報を直接的に参照し, その情報から有効な情報のみを抽出することによって表情変化をとらえる. Temporal skip connections, Spatial skip connections は, 時間方向, 空間方向に勾配を伝搬するパスである.

2.2 提案する表情認識手法

図 2 に提案する表情認識手法の概要を示す. 提案する表情認識手法は, 2 つの改良型 Conv-RNNs ストリームと 3 層の全結合層から構成される. 各ストリームには, RGB およびオプティカルフローの画像シーケンスが入力される. 改良型 Conv-RNNs ストリームでは, 多層の改良型 Conv-RNNs によって特徴を抽出する. そうすることによって, 時空間の関係を同時に表現した特徴を抽出することができる.

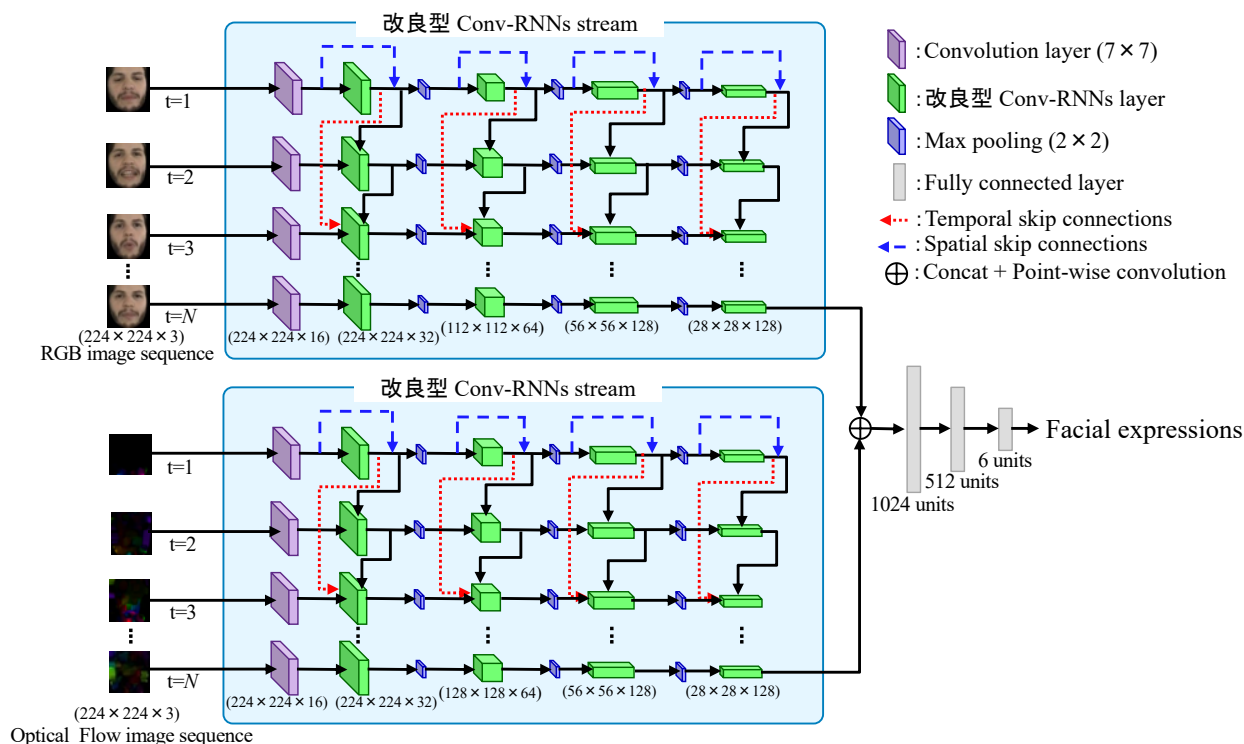


図2 提案する表情認識手法の概要

3. 実験

3.1 実験条件

本実験では、CK+[2]を用いて実験をおこなった。CK+は、123名の被験者から得られた593本の画像シーケンスを含んだデータセットである。また、それらの画像シーケンスのうち、118名から得られた327本の画像シーケンスにanger, contempt, disgust, fear, happiness, sadness, surpriseの7種類のラベルのどれかが教師信号として付与されている。本実験では、提案手法および従来手法を10-fold person-independence cross-validationによって評価した。

本実験では、ConvRNNsの一般的モデルであるConvolutional Gated Recurrent Unit (ConvGRU) [3]およびConvolutional Long Short-Term Memory (ConvLSTM) [4]とそれらにTemporal gatesおよびTemporal skip connections, Spatial skip connectionsを加えて改良した、改良型ConvGRUおよび、改良型ConvLSTMを比較する。

3.2 実験結果

実験結果を表1に示す。従来のConvGRUの認識精度が86.54%であるのに対して、改良型ConvGRUは、88.99%であり、2.45%の精度向上を確認した。また、従来のConvLSTMの認識精度が85.93%であるのに対して、改良型ConvLSTMは、90.52%であり、4.59%の精度向上を確認した。改良型ConvGRUより改良型ConvLSTMの方が認識率の向上が大きい。この要因として、ConvLSTMはConvGRUよりパラメータ数が多いため学習が難しいが、提案するアイデアを加えることによって、勾配が消失せずに伝搬でき、より良いパラメータの更新ができたためである。

表1 提案手法と従来手法の認識精度

Method	Accuracy
ConvGRU	86.54%
ConvLSTM	85.93%
改良型 ConvGRU	88.99%
改良型 ConvLSTM	90.52%

ると考えられる。

4. まとめ

本研究では、従来のConv-RNNsに表情の変化をとらえるためのゲートと勾配消失を抑制する2つのSkip connectionsを加えた改良型Conv-RNNsとそれを用いた表情認識手法を提案した。実験では、ConvRNNsの一般的モデルであるConvGRUとConvLSTMを改良した。改良型ConvGRUおよび改良型ConvLSTMの認識精度はそれぞれ、88.99%、90.52%であり、ConvGRUおよびConvLSTMに比べてそれぞれ、2.45%、4.59%の精度向上を確認した。

謝辞

JST 研究成果展開事業 COI プログラム「感性とデジタル製造を直結し、生活者の創造性を拡張するファブ地球社会創造拠点」の支援によっておこなわれた。

参考文献

- [1] Ambadar et al., "Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions", Psychological science, 2005.
- [2] Kanade et al., "Comprehensive database for facial expression analysis", FG, 2000.
- [3] J. Wang et al., "Gated Recurrent Neural Networks for OCR", NIPS, 2017.
- [4] X. Shi et al. "Convolutional LSTM Network: A Machine Learning Approach for Prediction Nowcasting", NIPS, 2015.