

A Recurrent Encoder-Decoder Network for Sequential Face Alignment

Xi Peng

Dept. of Computer Science

12/22/2016 at 02:00 pm
CBIM 22

Abstract

We propose a novel recurrent encoder-decoder network model for real-time video-based face alignment. Our proposed model predicts 2D facial point maps regularized by a regression loss, while uniquely exploiting recurrent learning at both spatial and temporal dimensions. At the spatial level, we add a feedback loop connection between the combined output response map and the input, in order to enable iterative coarse-to-fine face alignment using a single network model. At the temporal level, we first decouple the features in the bottleneck of the network into temporal-variant factors, such as pose and expression, and temporal- invariant factors, such as identity information. Temporal recurrent learning is then applied to the decoupled temporal-variant features, yielding better generalization and significantly more accurate results at test time. We perform a comprehensive experimental analysis, showing the importance of each component of our proposed model, as well as superior results over the state-of-the-art in standard datasets.

Examination Committee: Prof. Dimitris Metaxas (Chair), Prof. Apostolos Gerasoulis, Prof. Konstantinos Michmizos and Prof. Amelie Marian