Structure-Coherent Deep Feature Learning for Robust Face Alignment

Chunze Lin*, Beier Zhu*, Quan Wang, Renjie Liao, Chen Qian, Jiwen Lu, *Senior Member, IEEE,* and Jie Zhou, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a structure-coherent deep feature learning method for face alignment. Unlike most existing face alignment methods which overlook the facial structure cues, we explicitly exploit the relation among facial landmarks to make the detector robust to hard cases such as occlusion and large pose. Specifically, we leverage a landmark-graph relational network to enforce the structural relationships among landmarks. We consider the facial landmarks as structural graph nodes and carefully design the neighborhood to passing features among the most related nodes. Our method dynamically adapts the weights of node neighborhood to eliminate distracted information from noisy nodes, such as occluded landmark point. Moreover, different from most previous works which only tend to penalize the landmarks absolute position during the training, we propose a relative location loss to enhance the information of relative location of landmarks. This relative location supervision further regularizes the facial structure. Our approach considers the interactions among facial landmarks and can be easily implemented on top of any convolutional backbone to boost the performance. Extensive experiments on three popular benchmarks, including WFLW, COFW and 300W, demonstrate the effectiveness of the proposed method. In particular, due to explicit structure modeling, our approach is especially robust to challenging cases resulting in impressive low failure rate on COFW and WFLW datasets. The model and code are publicly available at https://github.com/BeierZhu/Sturcture-Coherency-Face-Alignment**

I. Introduction

Face alignment, also known as facial landmark detection is an important topic in computer vision and has attracted much attention over past few years [1], [2], [3], [4], [5], [6], [7]. As a fundamental step for face image analysis, face alignment plays a key role in many face applications such as face recognition [8], expression analysis [9] and face editing [10].

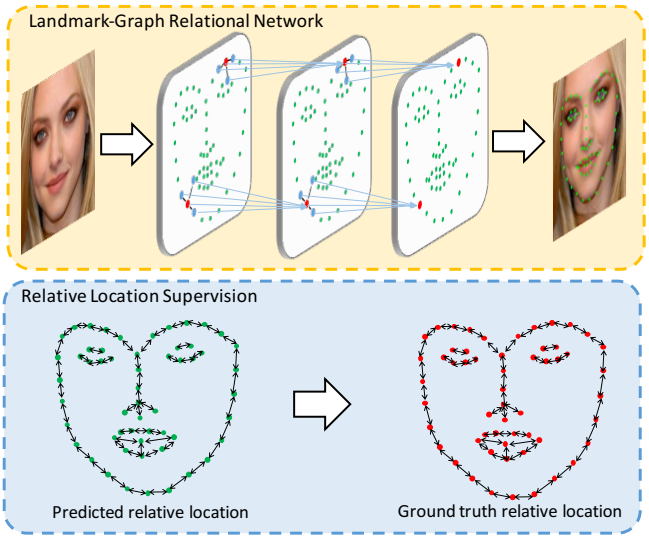

Fig. 1. The proposed structure-coherent deep feature learning method leverages a landmark-graph relational network which provides graph-based inferences among facial landmarks, exploiting facial key points relations to constrain landmarks. Most related landmarks are grouped and convolved together through graph convolutional layers to infer the facial landmarks. In addition, unlike most existing deep learning based methods which focus on penalizing the model to minimize the absolute location of landmarks, we propose a relative location loss to further enhance the facial structure coherency. With this relative location supervision signal, the model is also constrained to minimize the relative location errors of the predicted landmarks.

Although significant progress has been made, face alignment is still a challenging problem due to issues like occlusion, large head pose and complicated expression.

With the success of deep learning in several computer vision tasks such as image classification and object detection, many convolutional neural networks (CNN) based face alignment methods have been proposed. Existing CNN-based face alignment methods can mainly be divided into two categories: heatmap regression based ones [11], [3], [12] and coordinate regression based [13], [2], [1]. Heatmap regression based methods commonly produce higher precise localization for its translation equivariant property [14]. As keeping the high spatial resolution of feature maps and heatmap is essential for high accuracy, heatmap regression based methods commonly utilize stacked hourglass shape networks [15]. However, it leads to computationally heavy models which are impractical for deployment in real-world applications. Coordinate regression based methods are relatively simpler and can be built on lighter convolutional networks. Therefore, in this work,

* means equal contribution.

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFA0700802, in part by the National Natural Science Foundation of China under Grant 61822603, Grant U1813218, Grant U1713214, and Grant 61672306, in part by a grant from the Institute for Guo Qiang, Tsinghua University, and in part by the Shenzhen Fundamental Research Fund (Subject Arrangement) under Grant JCYJ20170412170602564, in part by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-PhD-2021-01-002)

Chunze Lin, Beier Zhu, Quan Wang and Chen Qian are with the Sensetime Research, Beijing, 100084, China. Email: linchunze@sensetime.com; zhubeier@sensetime.com; wangquan@sensetime.com; qianchen@sensetime.com.

Renjie Liao is with the Department of Computer Science, Toronto University, Toronto, Canada. Email: rjliao@cs.toronto.edu

Jiwen Lu and Jie Zhou are with the Department of Automation, Tsinghua University, State Key Lab of Intelligent Technologies and Systems, and Beijing National Research Center for Information Science and Technology, Beijing, 100084, China. Email: lujiwen@tsinghua.edu.cn; jzhou@tsinghua.edu.cn.