



Editorial: Special Issue on Deep Learning for Face Analysis

Chen Change Loy¹ · Xiaoming Liu² · Tae-Kyun Kim³ · Fernando De la Torre⁴ · Rama Chellappa⁵

Received: 24 April 2019 / Accepted: 24 April 2019 / Published online: 8 May 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

This special issue of the International Journal of Computer Vision (IJCV) is a compilation of papers that present new deep learning approaches for various applications on face analysis. While substantial progress has been achieved in face analysis with deep learning, many issues still remain and new problems have emerged. This special issue presents papers that have gone through the rigorous process of multiple rounds of reviewing and thus present a comprehensive account of the state-of-the art in face analysis. Submissions to the special issue closed in Spring of 2018. In total, we received 60 submissions to this special issue, of which 21 are accepted after passing the thorough reviewing process of IJCV. All submissions were reviewed by at least three reviewers.

This special issue presents a snapshot of some of the best work in face analysis by deep learning, with tasks focusing on face detection, facial landmark detection, face recognition, face restoration, face synthesis, and facial expression analysis. Below we provide a brief summary of the accepted papers by categorising them into a few key topics.

Face detection remains a very active topic in the community. Thanks to the release of the WIDER Face benchmark in the year of 2016, efforts are mainly devoted to detecting faces with a wide range of scales and appearance variations in

crowded scenes. Shifeng Zhang, Longyin Wen, Hailin Shi, Zhen Lei, Siwei Lyu, and Stan Z. Li present a single-shot scale-aware convolutional neural network (CNN) based face detector, which can effectively handle faces of various scales. This is achieved through a new anchor matching strategy with scale compensation, an IoU-aware weighting scheme, and a max-out background strategy. Shuzhe Wu, Meina Kan, Shiguang Shan, and Xilin Chen introduce a hierarchical attention mechanism to better represent features of different parts of a face proposal. Specifically, given a face proposal, part-specific attention is modelled as learnable Gaussian kernels to search for proper positions and scales of local regions to extract consistent and informative features of facial parts. Then face-specific attention predicted with LSTM is introduced to model the relations among the local parts and adjust their contributions to the detection task.

After face detection, one needs to perform face tracking, facial landmark detection or face parsing to extract facial structure descriptors to facilitate further analysis. To achieve more accurate and fast tracking of head pose, Stephen Ackland, Francisco Chiclana, Howell Istance, and Simon Coupland propose a 2.5D Constrained Local Model, which combines a deformable 3D shape point model with 2D texture information to estimate the head pose parameters. Jiankang Deng, Anastasios Roussos, Grigorios Chrysos, Evangelos Ververas, Irene Kotsia, Jie Shen, and Stefanos Zafeiriou focus on the task of face alignment, which is a challenging problem due to confounding factors such as variation in pose, illumination, facial expression, and occlusions. The problem is exacerbated by the lack of representative training data. The authors address this problem by presenting two new benchmark datasets for multi-pose 2D and 3D facial landmark localisation and tracking - Menpo 2D and 3D benchmark datasets. The new benchmark datasets are more challenging and representative than previous benchmarks such as 300W and 300VW. Furthermore, they were used in a facial landmark competition at CVPR 2017. The authors summarise the algorithm submissions and present the competition results. The authors further describe an effective approach for automated facial landmark annotation. Yujiang

✉ Chen Change Loy
ccloy@ntu.edu.sg

Xiaoming Liu
liuxm@cse.msu.edu

Tae-Kyun Kim
tk.kim@imperial.ac.uk

Fernando De la Torre
ftorre@cs.cmu.edu

Rama Chellappa
rama@umiacs.umd.edu

¹ Nanyang Technological University, Singapore, Singapore

² Michigan State University, East Lansing, USA

³ Imperial College London, London, UK

⁴ Facebook, Menlo Park, USA

⁵ University of Maryland, College Park, USA

Wang, Bingnan Luo, Jie Shen, and Maja Pantic introduce a convolutional-recurrent network that can extract segmentation masks of individual facial components. They show the effectiveness of the proposed model on video-based face mask extraction.

Face recognition is the topic that receives the most attention from the research community. This special issue presents a few studies that cover this topic from various aspects, including data augmentation, loss function design, and model design to combat against adversarial attack. Specifically, Iacopo Masi, Anh Tuấn Trần, Tal Hassner, Gozde Sahin, and Gérard Medioni describe a new and efficient data augmentation method to enrich training data with face-specific appearance variations, as well as to synthesise novel views of faces to reduce the effects of appearance nuisances during test time. Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao addresses the open-set characteristic of the face recognition problem by introducing the center loss, which simultaneously learns a center for each class and penalises the distances between the deep features of face images and their corresponding class centers. The loss has shown appealing performance in encouraging inter-class separability and intra-class compactness. In the same vein, Xiangyu Zhu, Hao Liu, Zhen Lei, Hailin Shi, Fan Yang, Dong Yi, Guojun Qi, and Stan Z. Li address the problem of large intra-class variation in the task of cross-domain recognition (identity photos and probe photos captured on spot) through a novel bisample learning method. Yongming Rao, Jiwen Lu, and Jie Zhou present a method based on metric learning and adversarial learning to aggregate information across video frames before feature extraction. The method is shown to be effective in handling large pose and viewpoint variations on the tasks of video-based face recognition and person re-identification. Finally, Gaurav Goswami, Akshay Agarwal, Nalini Ratha, Richa Singh, and Mayank Vatsa present a comprehensive study to evaluate adversarial attacks towards off-the-shelf deep learning based face recognition algorithms. This study shows that such attacks can be automatically detected by characterising the abnormal filter response from hidden layers of deep networks. The authors further propose a new technique of selective dropout in the deep network to mitigate the effect of these adversarial attacks.

A closely related topic to face recognition is the disentanglement of unobserved factors, including pose, illumination, and deformation, from visual appearance. Mengjiao Wang, Zhixin Shu, Shiyang Cheng, Yannis Panagakis, Dimitris Samaras, and Stefanos Zafeiriou present a pseudo-supervised deep learning method for disentangling multiple latent factors of variation in face images captured in-the-wild. The proposed method is capable of modelling multiplicative interactions of multiple latent factors of variation by means of multilinear (tensor) structure. They show the applicability of their approach on various applications including face

editing, as well as 3D face reconstruction and classification of facial expression, identity and pose.

Face restoration is another important topic in face analysis. Papers accepted in this special issue mainly aim to restore a high-resolution or clear face image from a low-resolution and blurred input. Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan found that CNN based methods tend to produce over-smoothed outputs. To overcome this issue, they propose a wavelet-domain Generative Adversarial Network (GAN) that predicts wavelet information of high-resolution face images from its corresponding low-resolution input. In the work by Yibing Song, Jiawei Zhang, Lijun Gong, Shengfeng He, Linchao Bao, Jinshan Pan, Qingxiang Yang, and Ming-Hsuan Yang, they present a GAN-based method that can upscale and deblur low-resolution face images simultaneously. The restoration is guided by facial components discovered from the input image as well as fine-grained facial structures obtained from the high-resolution exemplar. Grigorios G. Chrysos, Paolo Favaro, and Stefanos Zafeiriou focus on the face deblurring problem. They develop a deep motion deblurring model consisting of two sub-networks, where the first one is used to restore the low-frequency details, and the second one is a conditional GAN to restore the high-frequency details. A new dataset with more videos and identities was built and a new way to simulate motion blur from videos is proposed. Extensive experiments were conducted on facial landmark detection and face verification.

Most of the aforementioned methods employ Generative Adversarial Network (GAN), which is also used by the following work for face synthesis and image-to-image translation. Linh Tran, Jean Kossaifi, Yannis Panagakis, and Maja Pantic propose to incorporate geometric information about the shape of faces into deep generative models. The generator is conditioned on a statistical shape prior with differentiable canonical shape normalisation. This enables the generation of images with realistic texture and shape. He Zhang, Benjamin S. Riggan, Shuowen Hu, Nathaniel J. Short, and Vishal M. Patel show the possibility of using GAN for synthesizing photo-realistic visible face images from polarimetric thermal images. The transformation allows off-the-shelf face recognition networks to be deployed on polarimetric thermal images. Through adversarial learning, Fatemeh Shiri, Xin Yu, Fatih Porikli, Richard Hartley, and Piotr Koniusz introduce a deep network that is capable of recovering the latent photorealistic face from a given artistic portrait, while preserving the identity of the face. The method is shown to be effective on unseen stylized portraits, artistic paintings, and hand-drawn sketches.

Accurate facial attribute recognition, e.g., facial expression recognition, affect recognition, and age recognition, remain challenging in the field of computer vision despite the many years of research devoted to these tasks. This special issue contains four papers that address this challenge by

contributing comprehensive benchmarks and deep learning methods. Shan Li and Weihong Deng present a novel multi-label facial expression database, RAF-ML, which contain 1.2 million labels from 315 participants. A deep manifold learning network is also proposed to capture discriminative features from multi-label expressions by jointly preserving the local affinity of deep features and the manifold structures of emotion labels. Dimitrios Kollias, Panagiotis Tzirakis, Mihalis A. Nicolaou, Athanasios Papaioannou, Guoying Zhao, Björn Schuller, Irene Kotsia, and Stefanos Zafeiriou introduce the AffWild benchmark for training and evaluating affect recognition algorithms. The authors also present a convolutional and recurrent neural network that is capable of predicting continuous emotion dimensions based on visual cues. Feng-Ju Chang, Anh Tuan Tran, Tal Hassner, Iacopo Masi, Ram Nevatia, and Gérard Medioni present a method for modelling 3D face shape, viewpoint, and expression from a single unconstrained photo. Interestingly, different from the practice of existing approaches, this method does not require facial landmark detection at test time. The authors also show the applicability of their method on face recognition. Chi Nhan Duong, Kha Gia Quach, Khoa Luu, T. Hoang Ngan Le, Marios Savvides, and Tien D. Bui address the age progression problem through a framework that considers subject-dependent aging path. Unlike previous methods that only accepts a single input, the proposed method is capable of taking multiple images of a subject of different ages to produce an optimal aging path.

Overall, the papers in this issue offer diverse perspectives for solving some of the most important problems in the face analysis task, and each offers novel technical contributions towards that underlying goal. The special issue also presents a few papers that contribute baselines and benchmark datasets. We believe these papers would create high impacts in the field. We wish to thank the reviewers who took time to read several versions of the submitted manuscripts, and the editorial staff at Springer, who provided us with enormous help in preparing this special issue.

Accepted Papers

1. **Single-Shot Scale-Aware Network for Real-Time Face Detection**
Shifeng Zhang, Longyin Wen, Hailin Shi, Zhen Lei, Siwei Lyu, and Stan Z. Li
2. **Hierarchical Attention for Part-Aware Face Detection**
Shuzhe Wu, Meina Kan, Shiguang Shan, and Xilin Chen
3. **Real-Time 3D Head Pose Tracking Through 2.5D Constrained Local Models with Local Neural Fields**

Stephen Ackland, Francisco Chiclana, Howell Istance, and Simon Coupland

4. **The Menpo Benchmark for Multi-pose 2D and 3D Facial Landmark Localisation and Tracking**

Jiankang Deng, Anastasios Roussos, Grigorios Chrysos, Evangelos Ververas, Irene Kotsia, Jie Shen, and Stefanos Zafeiriou

5. **Face Mask Extraction in Video Sequence**

Yujiang Wang, Bingnan Luo, Jie Shen, and Maja Pantic

6. **Face-Specific Data Augmentation for Unconstrained Face Recognition**

Iacopo Masi, Anh Tuấn Trần, Tal Hassner, Gozde Sahin, and Gérard Medioni

7. **A Comprehensive Study on Center Loss for Deep Face Recognition**

Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao

8. **Large-scale Bisample Learning on ID vs. Spot Face Recognition**

Xiangyu Zhu, Hao Liu, Zhen Lei, Hailin Shi, Fan Yang, Dong Yi, Guojun Qi, and Stan Z. Li

9. **Learning Discriminative Aggregation Network for Video-based Face Recognition and Person Re-identification**

Yongming Rao, Jiwen Lu, and Jie Zhou

10. **Detecting and Mitigating Adversarial Perturbations for Robust Face Recognition**

Gaurav Goswami, Akshay Agarwal, Nalini Ratha, Richa Singh, and Mayank Vatsa

11. **An Adversarial Neuro-Tensorial Approach For Learning Disentangled Representations**

Mengjiao Wang, Zhixin Shu, Shiyang Cheng, Yannis Panagakis, Dimitris Samaras, and Stefanos Zafeiriou

12. **Wavelet Domain Generative Adversarial Network for Multi-scale Face Hallucination**

Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan

13. **Joint Face Hallucination and Deblurring via Facial Structure Generation and Detail Enhancement**

Yibing Song, Jiawei Zhang, Lijun Gong, Shengfeng He, Linchao Bao, Jinshan Pan, Qingxiong Yang, and Ming-Hsuan Yang

14. Motion Deblurring of Faces

Grigorios G. Chrysos, Paolo Favaro, and Stefanos Zafeiriou

15. Disentangling Geometry and Appearance with Geometry-Aware Generative Adversarial Network

Linh Tran, Jean Kossaifi, Yannis Panagakis, and Maja Pantic

16. Synthesis of High-Quality Visible Faces from Polarimetric Thermal Faces using Generative Adversarial Networks

He Zhang, Benjamin S. Riggan, Shuowen Hu, Nathaniel J. Short, and Vishal M. Patel

17. Face Recovery from Stylized Portraits

Fatemeh Shiri, Xin Yu, Fatih Porikli, Richard Hartley, and Piotr Koniusz

18. Blended Emotion in-the-Wild: Multi-label Facial Expression Recognition Using Crowdsourced Annotations and Deep Locality Feature Learning

Shan Li and Weihong Deng

19. Deep Affect Prediction in-the-wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond

Dimitrios Kollias, Panagiotis Tzirakis, Mihalis A. Nicolaou, Athanasios Papaioannou, Guoying Zhao, Björn Schuller, Irene Kotsia, and Stefanos Zafeiriou

20. Deep, Landmark-Free FAME: Face Alignment, Modeling, and Expression Estimation

Feng-Ju Chang, Anh Tuan Tran, Tal Hassner, Iacopo Masi, Ram Nevatia, and Gerard Medioni

21. Learning from Longitudinal Face Demonstration - Where Tractable Deep Modeling Meets Inverse Reinforcement Learning

Chi Nhan Duong, Kha Gia Quach, Khoa Luu, T. Hoang Ngan Le, Marios Savvides, and Tien D. Bui

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.