Guest Editors' Introduction to the Special Section on Compact and Efficient Feature Representation and Learning in Computer Vision

Li Liu, *Senior Member, IEEE*, Matti Pietikäinen, *Fellow, IEEE*, Jie Chen, *Member, IEEE*, Guoying Zhao, *Senior Member, IEEE*, Xiaogang Wang, *Senior Member, IEEE*, and Rama Chellappa, *Fellow, IEEE*

Feature representation and learning is at the core of many computer vision problems such as image classification, object recognition, action recognition, object tracking, image search, biometrics and many others. In the past two decades remarkable progress has been witnessed in feature representation and learning, which mainly consist of two important development stages. In the first stage from 1995 to 2012 (i.e., the predeep learning era), the field was dominated by milestone handcrafted feature descriptors such as SIFT, SURF, HOG, LBP, Bag of Visual Words, Fisher Vector, etc. The second stage, i.e., the deep learning era, starts from 2012 when a team led by Hinton won the prestigious ImageNet Challenge using deep learning techniques rather than traditional handcrafted features. The second stage is featured by deep learning based representations especially Deep Convolutional Neural Networks (DeepCNNs) which can learn powerful feature representations with multiple levels of abstraction directly from data.

Deep learning techniques have attracted enormous attention and have brought about considerable breakthroughs for many problems in computer vision. Increased computational power, deeper and more complicated deep neural networks, and the availability of large scale datasets are fueling computer vision systems. Despite the great success, the known deficiencies of deep neural networks have not been fully addressed, such as data hungry, energy hungry, lack of theoretical interpretability, etc.

Nowadays, intelligence is moving towards edge devices. Running machine learning systems on the end devices (e.g., smartphones, automobiles, wearable devices or Internet of Things devices) versus in the cloud has various benefits such as immediate response, enhanced reliability, increased privacy, and efficient use of network bandwidth. However, many realtime applications such as online learning, incremental learning, mobile, embedded, or wearable devices with limited resources and tight power budgets, or real time systems in which constraints are imposed by a limited economical budget, expose the inadequacies of existing algorithms, and require feature representations that are computationally and memory efficient. In addition, those applications where only limited amounts of annotated training data can be gathered (such as with many visual inspection or medical diagnostics tasks) pose great challenges for applying state of the art deep neural networks. Therefore, despite the great strides, especially over recent years, there is continued need for vigorous research in this area to solve many challenging problems, by developing compact, efficient feature representations from three aspects: computationally efficient, label efficient, and sample efficient.

Since 2017, we have organized four international workshops associated with top conferences (ICCV2017, ECCV2018, CVPR2019 and ICCV2019), explicitly devoted to the topic "Compact and Efficient Feature Representation and Learning in Computer Vision". This is a clear sign of the growing interest in computer vision around these themes. The goal of this special section has been to solicit and publish high quality papers that bring a clear picture of the state of the art along this direction, and identify future promising research directions. As guest editors of this special section, we were happy to receive 25 submissions to our special section. After a careful review process, we accepted ten papers for publication. We thank the reviewers who provided detailed, insightful, and timely reviews, leading to the high quality of accepted papers. We also thank TPAMI EIC Sven Dickinson and Associate EICs for recognizing the widespread interest in this field, which warrants this special section. The accepted 10 papers in this special section can be grouped into five different main categories:

1) New Efficient networks and network compression;
2) Sparse dictionary coding;
3) Label efficient;
4) Efficient feature fusion;
5) Hashing codes;

We describe these next.

1 New Efficient Networks and Network Compression

The paper "*Runtime Network Routing for Efficient Image Classification*" by Yongming Rao, Jiwen Lu, Ji Lin, and Jie

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2019.2935426

Zhou presents in detail the development and evaluation of a generic Runtime Network Routing (RNR) framework for deep neural network compression. Motivated by the intuition that easy samples can be recognized by simpler and faster networks and only hard samples require complicated networks, RNR aims at selecting an optimal path inside the network by routing between different paths of the neural network dynamically according to the input image. Ideally, a faster path of the deep network can be selected when the image is an easy sample. To achieve such a goal, the network compression problem is modeled as a bottom up, layer by layer Markov decision process, and reinforcement learning is introduced to train the network. RNR is claimed to preserve the full ability of the original large scale deep network, and can achieve flexible balances according to the available computing resources. Image classification experiments on both the multipath residual network and the incremental convolutional channel pruning show that RNR outperforms static methods at the same computation complexity on both the CIFAR and ImageNet datasets.

The paper "*Denoising Prior Driven Deep Neural Network for Image Restoration*" by Weisheng Dong, Peiyao Wang, Wotao Yin, Guangming Shi, Fangfang Wu, and Xiaotong Lu proposes an efficient deep neural network by leveraging the prior of the observation model that characterizes the underlying image degradation process for image restoration tasks. Traditional denoising based image restoration methods allow the use of a more complex image prior by decoupling the optimization problem into two subproblems and optimize them iteratively, which limits their speed. While most deep neural network based ones do not take the the observation models characterizing the image degradation processes into consideration. Therefore, the authors first propose an efficient denoising based image restoration algorithm, and then unfold it by replacing the iterative optimization process with a multistage DeepCNNs. Their proposed approach is evaluated on several image restoration tasks including image denoising, deblurring and super-resolution and gives results competitive to state of the art with reduced computational complexity.

The paper "*Two Stream Region Convolutional 3D Network for Temporal Activity Detection*" by Huijuan Xu, Abir Das, and Kate Saenko addresses the problem of temporal activity detection in continuous, untrimmed videos and proposes a two stream Region Convolutional 3D (RC3D) network which incorporates spatial and temporal networks. RC3D is developed by using 3D convolutional neural networks, 3D RoI pooling to generate activity proposals with variable length, and temporal Region Proposal Network (RPN), extending the objection detection framework Faster RCNN to activity detection in 3D videos. Online hard example mining strategy is also incorporated to address the extreme foreground background imbalance problem. The authors evaluate their approach on three benchmark datasets for activity detection and show that the proposed method is effective and efficient.

The paper "*Tattoo Image Search at Scale: Joint Detection and Compact Representation Learning*" by Hu Han, Jie Li, Anil Jain, Shiguang Shan, and Xilin Chen describes a joint detection and compact representation learning approach for tattoo image search at scale. Authors notice that current tattoo image search methods mainly focus on matching cropped tattoos rather than searching in the wild, and also do not learn a compact representation for efficient tattoo search at scale. To overcome these limitations, the authors present an efficient tattoo search approach that is able to learn tattoo detection and compact representation jointly in a single DeepCNN via multitask learning. The authors evaluate their approach on tattoo detection, tattoo search at scale, sketch tattoo search and cross database search, and show that their approach outperforms the state of the art tattoo retrieval algorithms in terms of performance, computational cost and generalization ability.

2 Sparse Dictionary Coding

The paper "*Side Information for Face Completion: A Robust PCA Approach*" by Niannan Xue, Jiankang Deng, Shiyang Cheng, Yannis Panagakis, and Stefanos Zafeiriou studies the problem of Robust Principal Component Analysis (RPCA) with side information (i.e., the domain dependent prior knowledge) in the presence of missing values. RPCA is a fundamental approach for learning compact feature representations of visual data, however it fails to produce satisfactory results when the data has significant amount of incomplete and missing values (e.g., noise corruption and occlusions). To address such challenges, the authors propose to exploit side information for RPCA and propose two models Principle Component Pursuit using Side information with Missing values (PCPSM) and Principle Component Pursuit using Side information with Features and Missing values (PCPSFM). Their approaches are applied to facial UV completion which is widely used in pose invariant face recognition. In the case of UV completion, they propose the use of a Generative Adversarial Network (GAN) to extract side information and subspaces. They show the applicability and effectiveness of their approaches with facial image denoising, facial UV completion and pose invariant face recognition experiments.

The paper "*Robust Kronecker Component Analysis*" by Mehdi Bahri, Yannis Panagakis, and Stefanos Zafeiriou proposes a novel Kronecker decomposable component analysis model called Robust Kronecker Component Analysis (RKCA) for the unsupervised learning of compact representations of tensor data. Dictionary learning and component analysis methods are two fundamental paradigms to learn compact representations for a given task. The authors realize that sparse dictionary learning approaches like KSVD suffers from limited scalability to visual data of big dimensionality and limited robustness to outliers. In contrast, robust component analysis methods such as RPCA are able to recover low-complexity representations from data corrupted with noise, but do not provide a dictionary that respects the inherent structure of the data, and also suffer from large computational complexity. To combine the best of both worlds, the authors propose RKCA to learn separable dictionary, by using robust tensor factorizations to learn the dictionary and sparse representations simultaneously. The authors claim that RKCA is robust to noise and can be utilized for low-rank modeling. They demonstrate the effectiveness of their approach on background subtraction, and image denoising and completion.

3 Label Efficient

The paper "*3D Aided Dual Agent GANs for Unconstrained Face Recognition*" by Jian Zhao, Lin Xiong, Jianshu Li, Junliang Xing, Shuicheng Yan, and Jiashi Feng intends to address the face recognition problem under extreme poses via a "recognition via synthesis" approach—learning from synthesized profile view faces by using Generative Adversarial Networks (GANs) to deal with the long tail problem of training data and reduce the burden of costly data labeling. In order to narrow the domain discrepancy between synthetic and real face images, the authors propose a novel Dual Agent GAN (DAGAN) to generate visually realistic profile view faces. The authors train DAGAN with Boundary Equilibrium regularization, and incorporate additional loss terms to preserve pose, texture and identity of the generated images. The synthesized data is shown to improve the performance of large scale unconstrained face recognition.

The paper "*MinEntropy Latent Model for Weakly Supervised Object Detection*" by Fang Wan, Pengxu Wei, Zhenjun Han, Jianbin Jiao, and Qixiang Ye focuses on the problem of Weakly Supervised Object Detection (WSOD) with only image level labels available, in order to reduce significantly the cost of laborious and expensive bounding box annotations in object detection. Most existing approaches to tackle this problem resort to latent variable learning or multiple instance learning by using redundant object proposals as inputs. Due to the unavailability of bounding box annotations, the inconsistency between the weak supervision and learning objectives introduces considerable randomness to object locations and ambiguity to detectors. To address these problems, the authors propose a novel MinEntropy Latent Model (MELM) which consists of three components: proposal clique partition, object clique discovery and object localization. MELM effectively reduces the search space and alleviates the local minima problem in the optimization of the model. The authors provide detailed experiments to show the effectiveness of their method for weakly supervised object detection and localization.

4 Efficient Feature Fusion

The paper "*Late Fusion Incomplete Multiview Clustering*" by Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Ying, Dinggang Shen, Huaimin Wang, and Wen Gao investigates the problem of multiview clustering with incomplete views (e.g., one type of descriptor is considered as a view in computer version). The authors realize that currently one stage Incomplete MultiView Clustering (IMVC) algorithms, which unify imputation and clustering into a single optimization procedure and achieve excellent clustering performance, directly impute multiple incomplete similarity matrices and have high computational complexity. To improve on these methods, the authors approach the MVC as a task of information fusion, and propose an effectively and efficiently Late Fusion Incomplete MultiView Clustering (LFIMVC) algorithm to solve it. Their proposed LFIMVC jointly learns a consensus clustering matrix, imputes each incomplete base matrix, and optimizes the corresponding permutation matrices. The authors provide extensive experiments to show their algorithm significantly and consistently outperforms some state of the art algorithms with much less running time and memory.

5 Hashing Codes

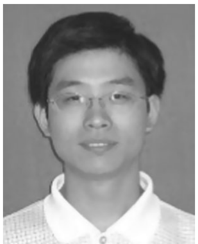
The paper "*Hashing with Mutual Information*" by Fatih Cakir, Kun He, Sarah Bargal, and Stan Sclaroff proposes a novel supervised hashing method called MIHash for nearest neighbor retrieval, based on mutual information to optimize the neighborhood ambiguity in the learned Hamming space. The authors introduce the well studied mutual information measure from information theory to capture the neighborhood structure to address binary codes learning problem and integrate it with deep neural networks to generate effective hash codes. To solve the end to end optimization problem and binary constraint problem, the authors elaborate on the differentiable histogram binning process and continuous relaxation, respectively. Experiments on four image retrieval benchmarks including ImageNet prove that MIHash is efficient to learn learning high-quality binary codes for nearest neighbor retrieval and gives state of the art performance.

6 Research Outlook

This special section focuses on new theories and algorithms for learning compact and efficient feature representations as well as applications. It presents interesting work that push the state of the art in related problems and opens new research directions as well as areas of applications. But we expect many new contributions in developing compact and efficient feature representations to meet real world requirements such as computationally efficient, label efficient, and sample efficient. We believe this special section will offer a timely collection of work to benefit the researchers working the research fields of computer vision, pattern recognition and machine learning.

Li Liu (SM'19) received the BSc, MSc and PhD degrees from the National University of Defense Technology (NUDT), China, in 2003, 2005, and 2012, respectively. She joined the faculty with NUDT in 2012, where she is currently an associate professor with the College of System Engineering. From 2008.1 to 2010.3, she visited the University of Waterloo, Canada. From 2015.3 to 2016.1, she visited the Multimedia Laboratory, Chinese University of Hong Kong. From 2016.11 to 2018.11, she worked with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. She was a cochair of seven International Workshops at CVPR, ICCV and ECCV. She was the leading guest editor of special issues of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal of Computer Vision*. Her current research interests include facial behavior analysis, image and video descriptors, object detection, and recognition. Her papers have about 2,000 citations in Google Scholar. L. Liu is with the College of System Engineering, National University of Defense Technology, China, and also with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. She is a senior member of the IEEE.

Matti Pietikäinen received the doctor of science degree in technology from the University of Oulu, Finland. He is an emeritus professor with the Center for Machine Vision and Signal Analysis, University of Oulu. From 1980 to 1981 and from 1984 to 1985, he visited the Computer Vision Laboratory, University of Maryland. He has made fundamental contributions, e.g., to Local Binary Pattern (LBP) methodology, texture based image and video analysis, and facial image analysis. He has authored more than 350 refereed papers in international journals, books, and conferences. His papers have about 57,000 citations in Google Scholar (hindex 82). He was associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, the *Pattern Recognition*, the *IEEE Transactions on Forensics and Security*, and the *Image and Vision Computing* journals. Currently, he serves as associate editor of the *IEEE Transactions on Biometrics, Behavior and Identity Science*, and guest editor for special issues of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal of Computer Vision*. He was president of the Pattern Recognition Society of Finland from 1989 to 1992, and was named its honorary member in 2014. From 1989 to 2007, he served as member of the Governing Board of International Association for Pattern Recognition (IAPR), and became one of the founding fellows of the IAPR in 1994. In 2014, his research on LBP-based face description was awarded the Koenderink Prize for fundamental contributions in computer vision. He was the recipient of the IAPR King-Sun Fu Prize 2018 for fundamental contributions to texture analysis and facial image analysis. In 2018, he was named a highly cited researcher by Clarivate Analytics, by producing multiple highly cited papers in 2006-2016 that rank in the top 1 percent by citation for his field in web of science. He is a fellow of the IEEE for contributions to texture and facial image analysis for machine vision.



Jie Chen received the MSc and PhD degrees from the Harbin Institute of Technology, China, in 2002 and 2007, respectively. He joined the faculty with the Graduate School in Shenzhen, Peking University, in 2019, where he is currently an associate professor with the School of Electronic and Computer Engineering. Since 2018, he has been working with the Peng Cheng Laboratory, China. From 2007 to 2018, he worked as a senior researcher with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. In 2012 and 2015, he visited the Computer Vision Laboratory, University of Maryland and School of Electrical and Computer Engineering, Duke University respectively. His research interests include pattern recognition, computer vision, machine learning, deep learning, and medical image analysis. He is an associate editor of the *Visual Computer*. He is a member of the IEEE.

Guoying Zhao (SM'12) received the PhD degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. Then she worked as senior researcher since 2005 and an associate professor since 2014 with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. She is currently a full professor with the University of Oulu, Finland and a visiting professor with Northwest University, China. In 2011, she was selected to the highly competitive Academy research fellow position. She was Nokia visiting professor in 2016. She has authored or co-authored more than 200 papers in journals and conferences. Her papers have currently more than 9,700 citations in Google Scholar (h-index 43). She was co-publicity chair for FG2018, has served as area chairs for several conferences and is associate editor of the *Pattern Recognition*, the *IEEE Transactions on Circuits and Systems for Video Technology*, and the *Image and Vision Computing Journals*. She has lectured tutorials at ICPR 2006, ICCV 2009, SCIA 2013 and FG 2018, authored/edited three books and eight special issues in journals. She was a co-chair of many International Workshops at ECCV, ICCV, CVPR, ACCV and BMVC. Her current research interests include image and video descriptors, facial-expression and micro-expression recognition, gait analysis, dynamic-texture recognition, human motion analysis, and person identification. Her research has been reported by Finnish TV programs, newspapers, and MIT Technology Review. She is a senior member of the IEEE.

Xiaogang Wang received the BSc degree in electronic engineering and information science from the Special Class of Gifted Young, University of Science and Technology of China, the MSc degree in information engineering from the Chinese University of Hong Kong, and the PhD degree in computer science from the Massachusetts Institute of Technology. He is an associate professor with the Department of Electronic Engineering, Chinese University of Hong Kong since 2009. He was the area chairs of CVPR 2017, ICCV 2011 and 2015, ECCV 2014 and 2016, ACCV 2014 and 2016. He received the Outstanding Young Researcher in Automatic Human Behavior Analysis Award in 2011, Hong Kong RGC Early Career Award in 2012, CUHK Young Researcher Award 2012, and the Honorable Mention of PAMI Young Researcher Award 2016. His research interests include deep learning and computer vision. Currently, he serves as associate editor of the *International Journal of Computer Vision*. His papers have currently more than 30,000 citations in Google Scholar (h-index 88). He is a senior member of the IEEE.

Rama Chellappa received the PhD degree from Purdue University, in 1981. He is a distinguished University professor, and a Minta Martin professor of the Engineering of Electrical and Computer Engineering Department, University of Maryland (UMD), College Park. His current research interests include computer vision, machine intelligence, artificial intelligence, and image processing. He has received numerous research, teaching, innovation, and service awards from UMD, IEEE, IAPR, and IBM. He is a fellow of the AAAI, AAAS, ACM, IAPR, IEEE, and OSA.