**MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE**
**IEEE COMMUNICATIONS SOCIETY**
*http://mmc.committees.comsoc.org/*

# MMTC Communications – Review

IEEE COMMUNICATIONS SOCIETY

**Vol. 14, No. 6, December 2023**

## TABLE OF CONTENTS

# Message from the Review Board Directors

Welcome to the December 2023 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises three reviews that cover multiple facets of multimedia communication research including few-shot face recognition under occlusion, intelligent reflective surface-assisted communication, and privacy-preserving object detection in connected autonomous vehicles. These reviews are briefly introduced below.

The first paper, published in IEEE Transactions on Multimedia, edited by Prof. Wenming Cao, presents a two-stream prototype learning network for few-shot face recognition under occlusions.

The second paper, published in IEEE Transactions on Communications, edited by Prof. Jinbo Xiong, investigates how many reflecting elements are needed for energy- and spectral-efficient intelligent reflecting surface-assisted communication.

The third paper, edited by Dr. Ye Liu, was published in IEEE Transactions on Intelligent Transportation Systems. The authors proposed edge-cooperative privacy-preserving object detection over random point cloud shares for connected autonomous vehicles.

All the authors, reviewers, editors, and others who contribute to the release of this issue deserve appreciation with thanks.

IEEE ComSoc MMTC Communications – Review Directors

Yao Liu
Rutgers University, USA
Email: yao.liu@rutgers.edu

Wenming Cao
Shenzhen University, China
Email: wmcao@szu.edu.cn

Dongfeng (Phoenix) Fang
California Polytechnic State University, USA
Email: dofang@calpoly.edu

Ye Liu
Macau University of Science and Technology, Macau, China Email: liuye@must.edu.mo

# Two-Stream Prototype Learning to Learn Optimal Class Prototypes for Few-Shot Face Recognition under Occlusion

*A short review for "Two-Stream Prototype Learning Network for Few-Shot Face Recognition Under Occlusions"*

Edited by Wenming Cao

*X. Yang, M. Han, Y. Luo, H. Hu and Y. Wen, "Two-Stream Prototype Learning Network for Few-Shot Face Recognition Under Occlusions," in IEEE Transactions on Multimedia, vol. 25, pp. 1555-1563, 2023, doi: 10.1109/TMM.2023.3253054.*

Face recognition under occlusion is a challenging face recognition task. In the past few years, deep learning approaches have achieved significant advances in recognizing occluded face images, mainly attributing to the following factors: large-scale and high-quality image galleries and limited variants of occluded scenes [1]. For example, the occluded face dataset is synthesized using occluders such as sunglasses and scarves. However, such occlusion patterns are quite different from those in the real world application. In practice, face images also suffer from self-occlusion (Non-frontal pose), facial accessory occlusion (masked face image) [2], extreme illumination (Part of face highlighted), and low resolution. Besides, In real-world applications, it is common that only a few labeled face images are available for a novel subject. Therefore, it is desirable to design a method for few-shot face recognition under occlusion (FSFRO).

Current face recognition approaches usually deal with the occlusion issue by extracting features that are robust to occlusion scenes. If a large number of training data or augmented data are provided, the authors can obtain the occlusion-robust face feature representations by employing advanced deep CNN architectures [3]. However, it is challenging to collect a large number of occlusion data and augmented data cannot be adapted to general diverse types of occlusion. It is extremely nontrivial to obtain features that are robust to occlusion in the few-shot setting. Some works [4],[5] try to learn robust features by employing the feature fusion strategy, but the learning for novel subjects given only a few and probably occluded face images is not investigated.

To deal with this issue, the authors can simply apply existing few-shot image classification approaches to face recognition, where a subject is considered as a category. For few-shot face recognition under occlusion (FSFRO), this paper considers the impact of occluded images on the feature space as class-level feature jitters. Since the number of labeled face images for each novel subject is small (usually from 1 to 5), the class-level feature jitter to the prototypes caused by the occlusion in one image can be significant.

Thus, the authors' major contribution is to propose a novel framework termed **T**wo-**S**tream **P**rototype **L**earning **N**etwork (**TSPLN**) to learn optimal class prototypes by simultaneously exploring both kinds of relationships. The main idea is to learn adaptive weights for different support images by considering both their quality and relevance to the query, where the weights are induced using middle-level features since they can better adapt to novel classes due to their high transferable ability.

The proposed TSPLN is based on a meta-learning paradigm and mainly consists of a support-centered stream and query-centered stream. Specifically, for the support-centered stream, the authors adopt a pre-trained transferable similarity relation network (TSRN) to obtain the similarities between support images in the same class (e.g., for a N-way K-shot task, the authors obtain N similarity matrices of size $K \times K$). Then low-quality (such as occluded) support images are given lower weights when constructing class prototypes due to their low similarities to normal support images. In regard to the query-centered stream, the authors propose a middle-level feature alignment module, which aims to match the query and support images, and larger alignment scores indicate higher relevance to the query image. In addition, only utilizing the middle-level features to calculate the score may lead to mismatches between a query and some support images belonging to different classes.

TSPLN learns two sets of class prototypes from the support centered stream and the query-centered stream, respectively. The two-stream share the same backbone and will be optimized jointly. The authors then apply the prototype network metric to classify the query images according to the respective class prototypes of the two streams. In one episode, the sampled support set $S_N$ and query set $Q_N$ are fed into TSPLN, and the pre-trained backbone extracts deep features for support and query images into common embedding space. In the support-centered stream, the authors obtain similarities between images from the same class in the support set based on a pre-trained transferable similarity relation network (TSRN). Then the intra-class similarity can be utilized as the weight of the support image for inducing the class prototype. As a result, those low-quality (such as occluded) support images are given lower weights due to their low similarity to other normal support images. The support-centered prototype can effectively mitigate the negative impact of the occluded face images. The query-centered stream feeds the extracted middle-level feature maps to the feature alignment module, which matches the middle-level features between query and support images. Then the alignment scores are used as sample weights for class prototypes in the query-centered stream.

The authors conduct extensive experiments on two popular datasets: CASIA-WebFace and RMFRD. The experimental results show that the proposed method achieves the state-of-the-art performance for occluded face recognition in the few-shot setting.

In summary, the authors propose a two-stream framework to learn optimal class prototypes for the FSFRO problem. The proposed TSPLN can learn adaptive weights for different support images by simultaneously considering their quality and correlations with query images. To achieve this goal, the authors introduce a pre-trained transferable similarity relation network to the support-centered stream to reduce the negative impact of occluded images for class prototype learning. In the future, they intend to evaluate the significance of different images in a more fine-grained manner, and apply the proposed method to more recognition problems under occlusions in addition to faces.

**References:**


[1] H. Qiu, D. Gong, Z. Li, W. Liu, and D. Tao, “End2End occluded face recognition by masking corrupted features,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 10, pp. 6939–6952, Oct. 2022.

[2] D. Zeng, R. Veldhuis, and L. Spreeuwers, “A survey of face recognition techniques under occlusion,” IET Biometrics, vol. 10, no. 6, pp. 581-606, 2021.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no.6, pp. 84–90, 2017.

[4] A. Holkar, R. Walambe, and K. Kotecha, “Few-shot learning for face recognition in the presence of image discrepancies for limited multi-class datasets,” Image Vis. Comput., vol. 120, 2022, Art. no. 104420.

[5] W. Zheng, L. Yan, F.-Y. Wang, and C. Gou, “Learning from the web: Webly supervised meta-learning for masked face recognition,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 4304–4313.


**Wenming Cao** received the M.S. degree from the System Science Institute, China Science Academy, Beijing, China, in 1991, and the Ph.D. degree from the School of Automation, Southeast University, Nanjing, China, in 2003. From 2005 to 2007, he was a Post-Doctoral Researcher with the Institute of Semi-conductors, Chinese Academy of Sciences, Beijing, China. He is currently a Professor with Shenzhen University, Shenzhen, China. He is also a foreign academician of the Russian Academy of Natural Sciences. He has authored or coauthored over 80 publications in top conferences and journals. His research interests include pattern recognition, image processing, and visual tracking.

# Formalized Analysis of Reflective Elements for Intelligent Reflective Surface-Assisted Communication Systems

*A short review for "How Many Reflecting Elements Are Needed for Energy- and Spectral-Efficient Intelligent Reflecting Surface-Assisted Communication"*

Edited by Jinbo Xiong

*D. Li, "How Many Reflecting Elements Are Needed for Energy- and Spectral-Efficient Intelligent Reflecting Surface-Assisted Communication," in IEEE Transactions on Communications, vol. 70, no. 2, pp. 1320-1331, Feb. 2022,DOI: 10.1109/TCOMM.2021.3128544.*

Backscatter communication (BackCom) is heavily dependent on interference suppression techniques for reliable signal detection to reduce signal interference [1]. End nodes suffer expensive communication costs when BackCom's externally generated carrier signals are interfered by direct link interference and reflected scattering signals. Intelligent reflective surfaces (IRS) can optimize signal transmission, suppress interference, and improve communication efficiency by using many reconfigurable reflective elements with adjustable phases to ensure battery-free transmission and encourage passive transmission [2]. Generally speaking, IRS research focuses on spectral efficiency (SE) maximization [3], energy efficiency (EE) performance [4], and the number of reflective elements [5], while SE maximization leads to a linear increase in IRS power consumption of the circuit. In order to adequately respond to the number of reflecting elements, Zappone *et al*. [6] were used to maximize the maximum number of reflecting elements in the SE/EE maximization and SE-EE trade-off maximization scenarios. However, the issue of how to achieve the realization of the minimum number of reflection elements by optimizing the beamforming and phase shift remains unresolved.

In order to solve the above issue, contrary to previous research perspectives, the authors focus on how many reflection elements are needed to ensure predefined EE and SE performance, and investigate and analyze the number of reflection elements for IRS-assisted systems with optimized beamforming and phase shift. Also, the performance constraints of joint beamforming constraints and phase-shift nonconvexity optimization are utilized to decompose the problem into two subproblems, and closed-form expressions for the minimum number of reflective elements are theoretically derived to minimize the number of reflective elements for EE and SE performance constraints of IRS-assisted systems. The specific contributions of the authors can be summarized in the following three areas.

Firstly, due to the intractability of the EE and SE performances associated with the joint optimization of beamforming and phase-shift, neither of which allows closed form expressions, the authors recast their performance bounds to reformulate the original problem and decompose it into two sub-problems that can be easily analyzed using only the EE/SE constraints. The authors propose to derive the minimum number of reflective elements using a full channel phase information (CPI) and coherent phase-shift (CPS) oriented scheme and an oriented scheme with random phase-shift (RPS) with no CPI, and compare the results of the two schemes with the advantage of obtaining the minimum number of reflective elements in closed form expressions, achieving low computational complexity and without excessive computational burden in practice.

Second, the authors study and derive the upper bound closed expressions for both CPS-oriented and RPS-oriented schemes, reveal the quantitative relationship between the subproblems for minimizing the number of reflective elements in EE-only constraints and in SE-only constraints, elucidate how to select predefined EE and SE thresholds, as well as quantify the region in which only one reflective element can be utilized, completing the placement of the IRS.

Finally, the authors generated experimental data by using the MATLAB function "fmincon" and observed that there is a precise agreement between the analytical results and the simulation results, which verified the correctness of the minimum number of reflective elements derived for both schemes. It is also shown that by properly placing the IRS, the RPS oriented scheme can be considered in practice due to the CPS oriented scheme and the PCS feedback overhead with respect to phase shift is released.

In summary, the authors' derived closed-form expressions reveal the relationship between the minimum number of reflective elements for CPS-oriented and RPS-oriented, yielding upper bounds for the EE and SE constraints in both scenarios, indicating the appropriate placement region for IRS. It guides the number of reflective elements required for IRS-assisted communication systems to ensure EE and SE performance, which reduces signal transmission cost and improves practicability at the same time.

**References:**


[1] D. Li, H. Zhang and L. Fan, "Adaptive Mode Selection for Backscatter-Assisted Communication Systems With Opportunistic SIC," in IEEE Transactions on Vehicular Technology, vol. 69, no. 2, pp. 2327-2331, Feb. 2020, DOI: 10.1109/TVT.2019.2960381.

[2] S. Gong et al., "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey," in IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2283-2314, Fourthquarter 2020, DOI: 10.1109/COMST.2020.3004197.

[3] D. Li and Y. C. Liang, "Price-Based Bandwidth Allocation for Backscatter Communication With Bandwidth Constraints," in IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5170-5180, Nov. 2019, DOI: 10.1109/TWC.2019.2934101.

[4] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah and C. Yuen, "Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication," in IEEE Transactions on Wireless Communications, vol. 18, no. 8, pp. 4157-4170, Aug. 2019, DOI: 10.1109/TWC.2019.2922609.

[5] H. Zhang, B. Di, Z. Han, H. V. Poor and L. Song, "Reconfigurable Intelligent Surface Assisted Multi-User Communications: How Many Reflective Elements Do We Need?," in IEEE Wireless Communications Letters, vol. 10, no. 5, pp. 1098-1102, May 2021, DOI: 10.1109/LWC.2021.3058637.

[6] A. Zappone, M. Di Renzo, X. Xi and M. Debbah, "On the Optimal Number of Reflecting Elements for Reconfigurable Intelligent Surfaces," in IEEE Wireless Communications Letters, vol. 10, no. 3, pp. 464-468, March 2021, DOI: 10.1109/LWC.2020.3034686.


**Jinbo Xiong**, is a full Professor and Ph.D. supervisor with the Fujian Provincial Key Laboratory of Network Security and Cryptology and the College of Computer and Cyber Security at Fujian Normal University. He received the M.S. Degree in Communication and Information Systems from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2006, and the Ph.D. degree in Computer System Architecture from Xidian University, China, in 2013. He was a visiting scholar with the Department of Computer Science and Engineering at the University of North Texas, Denton, USA, from 2019-2020. His research interests include connected and autonomous vehicle, secure deep learning, cloud data security, mobile media management, privacy protection, and Internet of Things. He has published more than 100 papers in prestigious journals such as IEEE WCM, IEEE TITS, IEEE TII, IEEE TCC, IEEE TNSE, IEEE IoT J and in major International Conferences such as IEEE ICC, IEEE GLOBECOM, IEEE TrustCom, IEEE HPCC and IEEE ICPADS. He has approved 15 patents and two monographs in these fields. He is a senior member of IEEE and CCF.

# Random Point Cloud based Privacy-Preserving Object Detection in Connected Autonomous Vehicles

*A short review for "Edge-Cooperative Privacy-Preserving Object Detection Over Random Point Cloud Shares for Connected Autonomous Vehicles"*

Edited by Ye Liu

*R. Bi, J. Xiong, Y. Tian, Q. Li and X. Liu, "Edge-Cooperative Privacy-Preserving Object Detection Over Random Point Cloud Shares for Connected Autonomous Vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 24979-24990, Dec. 2022.*

The integration of autonomous vehicles (AVs) and machine vision is a transformative development in intelligent transportation systems (ITS) [1]. AVs utilize on-board LiDAR technology to capture precise object movement information within their field of perception. Unlike traditional 2D images captured by visible light cameras, the 3D point cloud data generated by LiDAR provides a more authentic representation of objects. This point cloud data comprises position coordinates, motion angles, reflection intensities, and other detailed information. By utilizing this point cloud data, AVs can more effectively interact with their environment and make safe driving decisions.

Thanks to the continuous advancements in wireless communication technologies, such as vehicle-to-vehicle (V2V), vehicle-to-pedestrian (V2P), and vehicle-to-infrastructure (V2I), it has become feasible for vehicles to establish connections with other vehicles, pedestrians, infrastructure, and even everything, leading to the concept of connected autonomous vehicles (CAVs) [2]. By fusing point cloud data from multiple sensor sources, CAVs can enhance their perception range and accuracy. However, the training of convolutional neural network (CNN) models for accurate object detection often requires a significant amount of point cloud data, leading to increased computational, storage, and battery life overheads on CAVs. To address these limitations, edge nodes, which possess abundant computing resources and lower communication delays compared to central cloud servers, can be utilized to perform object detection model inference and training tasks, offloading the computational burden from CAVs.

Although point cloud data contains valuable information for accurate object detection, it also contains sensitive data such as vehicle profiles, landmark buildings, and traffic identifications. In the vehicle edge computing (VEC) architecture, sharing the raw point cloud data directly with edge servers without any security measures can lead to serious privacy leakage issues. If edge servers sell or exchange this point cloud information with others, the travel information of CAVs and even the home addresses of drivers can be exposed. These privacy breaches can erode public trust in CAVs, making people less likely to embrace this technology. Therefore, it is essential for CAVs to implement secure measures on point cloud data before uploading and sharing them with edge servers. This ensures that edge servers can collaborate on object detection tasks without compromising the confidentiality of the point cloud data.

To ensure the privacy of point cloud data while performing object detection tasks, several common privacy-preserving computing techniques [3] can be employed, including trusted execution environments, differential privacy, homomorphic encryption, and secret sharing. However, these techniques introduce noise that accumulates rapidly in large object detection models. Excessive noise can lead to decreased model accuracy or even prevent convergence, necessitating the allocation of additional differential privacy budget to mitigate computational errors caused by noise. It is crucial to note that the driving state and effective control of CAVs are highly dependent on the detection accuracy of on-board sensors. Erroneous identifications, such as mistaking a blank space for an obstacle, can result in sudden braking and traffic disruptions. Conversely, mistaking an obstacle for a blank space can lead to catastrophic traffic accidents and pose a serious threat to personal safety.

To bridge above gaps, this paper presents a method that employs the Additive Homomorphic Sharing (ASS) technique to randomly divide point cloud data into two additive shares. Additionally, the paper introduces an edge-cooperative privacy-preserving point cloud object detection framework called SecPCV. Unlike previous methods that focused primarily on simple images, this approach deals with larger, irregularly formatted 3D point clouds. To address the inherent challenges posed by such data, the authors developed more intricate secure protocols and privacy-preserving frameworks. SecPCV is built upon the classical PointRCNN model due to its ability to mitigate semantic loss caused by point cloud projection and voxel encoding techniques. Unlike PointRCNN, all intermediate outcomes in SecPCV are securely shared. The proposed framework utilizes random point cloud shares as input instead of the raw point cloud and relies on two edge servers to collaboratively execute multiple layers of PointRCNN, ensuring secure object detection.

In summary, the paper presents SecPCV, a system designed to support random point cloud shares in CAVs. It offloads the entire object detection process to edge servers, significantly reducing the computational burden on CAVs. By randomly splitting the raw point cloud data into two shares, SecPCV lowers the computational cost. The system also introduces secure computing protocols based on ASS, ensuring privacy while enabling edge servers to collaboratively perform end-to-end and accurate object detection. SecPCV maintains the structure of PointRCNN, eliminating the need for re-training. It maintains the same computational complexity as PointRCNN and requires only constant communication rounds between edge servers. Experimental results using the KITTI point cloud dataset demonstrate that SecPCV achieves consistent object detection accuracy with PointRCNN, while incurring only a 1.56× increase in computational cost.

**References:**


[1] S. Ge et al., "Making Standards for Smart Mining Operations: Intelligent Vehicles for Autonomous Mining Transportation," in IEEE Transactions on Intelligent Vehicles, vol. 7, no. 3, pp. 413-416, Sept. 2022.

[2] J. Guo et al., "Slim-FCP: Lightweight-Feature-Based Cooperative Perception for Connected Automated Vehicles," in IEEE Internet of Things Journal, vol. 9, no. 17, pp. 15630-15638, 1 Sept.1, 2022.

[3] J. Xiong, R. Bi, M. Zhao, J. Guo and Q. Yang, "Edge-Assisted Privacy-Preserving Raw Data Sharing Framework for Connected Autonomous Vehicles," in IEEE Wireless Communications, vol. 27, no. 3, pp. 24-30, June 2020.


**Ye Liu**, received the M.S. and Ph.D. degrees in electronic science and engineering from Southeast University, Nanjing, China, in 2013 and 2018, respectively. He was a Visiting Scholar with Montana State University, Bozeman, MT, USA from October 2014 to October 2015. He was a visiting Ph.D. Student from February 2017 to January 2018 with the Networked Embedded Systems Group, RISE Swedish Institute of Computer Science, Kista, Sweden. He is currently a Macao Young Scholar with Macau University of Science and Technology, Macau, China. He has authored or co-authored papers in several prestigious journals and conferences, such as the IEEE WCM, IEEE IEM, IEEE ComMag, IEEE Network, IEEE IoTJ, IEEE TII, ACM TECS, INFOCOM, IPSN, ICNP, and EWSN. His current research interests include wireless sensor networks, energy harvesting systems, and smart agriculture. Dr. Liu was awarded the 1st place of the EWSN Dependability Competition in 2019

**Multimedia Communications Technical Committee Officers**



MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.