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

MMTC Communications – Review

Vol. 14, No. 4, August 2023

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

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

This issue comprises four reviews that cover multiple facets of multimedia communication research including Multi-Modal Alignment, Video Codec, few-shot hyperspectral image classification. These reviews are briefly introduced below.

The first paper, published in IEEE Transactions on Multimedia and edited by Dr. Qin Wang, develops an effective multi-modal news and propose cross-modal interaction method.

The second paper is published in IEEE Transactions on Pattern Analysis and Machine Intelligence and edited by Dr. Liu Qifan. It proposes contrastive Bayesian analysis to address FSL.

The third paper, published in IEEE Transactions on Broadcasting and edited by Dr. Takuya Fujihashi, measures the performance of all codecs on a computer equipped with an AMD Ryzen 5 1600X .and evaluation results.

The fourth paper, published in IEEE Transactions on Image Processing, and edited by Prof. Cao Wenming, proposes a novel FSL framework with a class-covariance metric (CMFSL) for accurate HSI classification.

The fifth paper, published in IEEE Transactions on Knowledge and Data Engineering, and edited by Prof. Cao Guitao, adopts a meta-learning framework to optimize the reweighting of pseudo-labeled target samples, and it can integrate with any adversarial-based UDA methods.

All the authors, nominators, 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

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

Multi-Modal Alignment and Fusion Network for Fake News Detection

A short review for "Entity-Oriented Multi-Modal Alignment and Fusion Network for Fake News Detection" Edited by Qin Wang

P. Li, X. Sun, H. Yu, et al., " Entity-Oriented Multi-Modal Alignment and Fusion Network for Fake News Detection," in IEEE Transactions on Multimedia, vol. 24, pp. 3455-3468, 2022.

The rapid development of social media enables news to be expressed in a multi-modal form. Multi-modal news which consists of images and videos engages more readers and provides them with a better reading experience [1]. However, the multi-modal representation of news also fosters various forms of fake news, which brings harmful social impacts and raises public concern. To reduce the harm caused by the dissemination of fake news, automatic multi-modal fake news detection was proposed to identify the authenticity of the news by verifying its multi-modal content. Therefore, the automatic multi-modal fake news detection has become an important topic in the field of news communication.

Thus far, it is effective to develop an effective multi-modal fake news detection system, including exploiting textual information to verify the truthfulness of news [2], introducing RNN and attention mechanism [3]. Traditional multi-modal fake news detection consists of three approaches with cross-modal interaction and fusion, including the method of combining the visual and social information into textual features by means of the attention mechanism, the method of introducing auxiliary tasks to capture correlations across various modalities and presenting the current state-of-the-art method named SpotFake, which consists of powerful pre-trained models to extract text features and image features.

The content of multi-modal news is narrated around entities. the attributes of entities are essential for predicting the truthfulness of the sample. One can identify fake news by comparing the context of the same entity in different modalities, regardless of whether the news may contain aligned text and image. Prior methods have proven to be effective in detecting common types of fake news, but they underperform with respect to handling samples that require entity-centric comparisons. Authors divide the prior methods into fine-grained methods and coarse-grained methods according to the granularities of cross-modal interaction objects. The fine-grained methods perform cross-modal interaction at the word level and pixel level [4] and cannot maintain the semantic integrity of the entities, especially for visual entities. While the coarse-grained methods perform cross-modal interaction at sentence-level and image-level features, they may neglect the relationships among the objects within the same modality.

Authors focus on the entities that appear in multi-modal news and propose an entity-oriented cross-modal interaction method, which achieves a compromise between the coarse-grained and fine-grained methods. Specifically, authors adopt a two-stage approach to promote the interaction and fusion of multi-modal features and propose the Entity-oriented Multi-modal Alignment and Fusion network (EMAF). The EMAF revolves around multi-modal entities and mainly includes cross-modal Alignment and Fusion modules. The former aims at obtaining the representations of the same entity in different modalities, and the latter intends to capture the consistencies and differences among the aligned entities.

First, considering the advantages of the Capsule Network in encapsulation and aggregation of entities, the Alignment module employs two group capsules to encapsulate visual and textual entities. Then, unlike the original Capsule Network that abstracts the primary capsule to generate digital capsules, Authors improve its core dynamic routing algorithm and take the visual entities and textual entities as primary capsules and digital capsules, respectively. In this manner, authors perform interaction between the entities in the visual and textual contexts and align the visual entities to the textual entities.

To capture the consistencies and differences among entities in different modalities, authors design a cross-modal Fusion module. The Fusion module is composed of the Attend, Compare, and Aggregate steps to perform second-order interaction for aligned entities and source textual entities. In each step, apart from the entity-level features, authors also supplement the semantic relationships among the entities from the same modality that are missed

within the alignment process. This operation collects the features that reflect the authenticity of the sample.
Authors conduct extensive experiments on three public datasets collected from Reddit, Weibo, and Twitter to evaluate the effectiveness of the proposed model. In view of the marked variations between the datasets, authors adopt different methods to pre-process the multi-modal samples. Following the text pre-processing approach, for the Fakeddit and Twitter dataset, authors first translate the non-English textual sentences into English to keep the data coherent and employ a pre-trained BERT model to map each textual sentence into an embedding sequence. Authors utilize the NLTK toolkits to extract the part-of-speech (POS) tags for each sentence. Then, authors extract the BERT representations of noun words as the entity representations. For the Weibo dataset, in which sentences are written in Chinese, authors use LTP [34] toolkits to extract POS tags and take the BERT representations of noun words as entity representations. and then authors provide the implementation details of the proposed model. Next, authors present the previous fake news detection approaches and compare their performance with that of our proposed EMAF. Finally, authors conduct the ablation study and necessary analytical experiments to explore the utility of the core modules in the proposed EMAF. EMAF method has following advantages: 1) At the level of information utilization, EMAF integrates both textual and visual information and focuses on the entities that appear in each modality. 2) EMAF utilizes powerful pre-trained models, including VGG and BERT, to extract multi-modal features.3) Our EMAF adopts an entity-oriented cross-modal alignment and fusion method, alleviating the disadvantage of fine-granted or coarse-granted cross-modal interaction.

In summary, authors propose a method named Entity-oriented Multi-modal Alignment and Fusion network (EMAF) that focuses on the entities that appear in multi-modal news. Specifically, EMAF adopts a two-stage approach, that is, cross-modal Alignment and Fusion, to combine multi-modal features. Experimental results on multiple datasets reveal the superiority of our proposed method, and further analysis illustrates the effectiveness of our proposed cross-modal Alignment and Fusion modules. For future work, in addition to exploring methods for fusing multi-modal information, authors will focus on how to integrate prior world knowledge and related information to assist with the verification of specific news.

References:

[1] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, “Novel visual and statistical image features for Microblogs news verification,” IEEE Trans. Multime-dia, vol. 19, no. 3, pp. 598–608, Mar. 2017.
[2] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter,” in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 675–684.
[3] T. Chen, X. Li, H. Yin, and J. Zhang, “Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection,” in Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining, 2018, pp. 40–52.
[4] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal fusion with recurrent neural networks for rumor detection on Microblogs,” in Proc. 25th ACM Int. Conf. Multimedia, 2017, pp. 795–816.

Qin Wang, Ph.D, is an Associate Professor at Nanjing University of Posts and Telecommunications (NJUPT), China. She received B.S. and Ph.D. degrees from NJUPT, in 2011 and 2016. Prior to joining NJUPT, she was with the New York Institute of Technology (NYIT) between Feb. 2017 and Aug. 2020. From July 2018 to June 2020, she was a Postdoctoral Research Fellow at NJUPT. From 2015 to 2016, she was a visiting scholar at San Diego State University, USA. Her research interests include multimedia communications, multimedia pricing, resource allocation in 6G, and Internet of Things. She has published papers in prestigious journals such as IEEE Transactions on Vehicular Technology and IEEE Communications Magazine, in prestigious conferences such as IEEE INFOCOM SDP Workshop.

Deep Metric Learning with Contrastive Bayesian Analysis

A short review for "Contrastive Bayesian Analysis for Deep Metric Learning"

Edited by Qifan Liu

S. Kan, Z. He, Y. Cen, Y. Li, V. Mladenovic and Z. He, "Contrastive Bayesian Analysis for Deep Metric Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7220-7238, 1 June 2023, doi: 10.1109/TPAMI.2022.3221486.

Research on deep metric learning or feature embedding has achieved remarkable progress in image retrieval, fine-grained object classification and matching, person re-identification and vehicle re-identification [1] [2]. Existing state-of-the-art methods have been focusing on learning deep neural networks with carefully designed loss functions to generate discriminative features with the goal to minimize intra-class sample distance and maximize inter-class sample distance. More recent deep metric learning methods, for example, lifted structured loss, proxy loss and ranked list loss [3], further extend these loss functions by considering richer sample structure information.

However, these loss functions heavily depend on how the positive and negative samples are selected, which directly affects their metric learning performance and algorithm convergence rate [4]. During metric learning, minimizing the feature distance between samples from the same class or maximizing their similarity does not necessarily guarantee that these samples can obtain similar representations. What's more, in many deep metric learning settings, the test classes are totally different from the training classes. how to make sure the features learned on the training classes can generalize well onto novel test classes is also an important research problem.

In this paper, the authors propose contrastive Bayesian analysis to address these two important issues. Specifically, the authors propose to analyze and model the inherent relationship between metric learning at the intermediate feature layer and their semantic labels at the final output layer based on a Bayesian conditional probability analysis. The authors develop this Bayesian analysis in a contrastive learning setting for positive and negative pairs and formulate a metric learning process. This new contrastive Bayesian analysis bridges the gap between the learned features of images and their class labels, resulting in a new loss function for deep metric learning. Because the gap between the learned features of images and their class labels is bridged, the new loss function based on contrastive Bayesian analysis can easily overfit the training set which can result in performance degradation on novel classes. To improve the generalization capability of the proposed method onto new classes, the authors further extend the contrastive Bayesian loss with a metric variance constraint. Moreover, the authors couple this contrastive Bayesian analysis with clustering-based pseudo label generation in an iterative manner to achieve improved performance for pseudo-supervised deep metric learning.

This work is related to Bayesian analysis, which has been studied in metric learning. Different from previous works [5] [6], the authors proposed to analyze and model the inherent relationship between sample labels and their similarity scores using a Bayesian conditional probability analysis approach for image retrieval. The authors also derive this new Bayesian analysis in a contrastive learning setting

Compared to existing work, the major contributions of this work can be summarized as follows. (1) Existing methods on deep metric learning have been focusing on the contrastive metric loss function design at the intermediate feature layer. This work addresses the important limitation in existing approaches and bridges the semantic gap between features and image labels. The authors derive the contrastive Bayesian analysis to estimate the posterior probability of labels conditioned by their feature metric in a contrastive learning setting, which leads to a new loss function for deep metric learning. (2) The second major contribution is that the authors extend the contrastive Bayesian analysis by considering the metric variance constraint and improve the generalization capability of the proposed method. (3) The new method based on contrastive Bayesian analysis has improved the performance of deep metric learning, outperforming existing state-of-the-art methods by a large margin.

Extensive experimental results and ablation studies demonstrate that the proposed contrastive Bayesian metric learning method significantly improves the performance of deep metric learning in both supervised and pseudo-supervised scenarios.

In summary, the authors have developed a contrastive Bayesian analysis to bridge the semantic gap between features at intermediate feature layers and class label decision at the final output layer. Based on this analysis, we are able to model and predict the posterior probabilities of image labels conditioned by their features similarity in a contrastive learning setting. This contrastive Bayesian analysis leads to a new loss function for deep metric learning. To improve the generalization capability of the proposed method onto new classes, the authors further extend the contrastive Bayesian loss with a metric variance constraint.

References:

[1] B. Chen and W. Deng, “Hybrid-attention based decoupled metric learning for zero-shot image retrieval,” in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pp. 2750–2759.

[2] S. Kan, Y. Cen, Z. He, Z. Zhang, L. Zhang, and Y. Wang, “Supervised deep feature embedding with handcrafted feature,” IEEE Trans. Image Processing, vol. 28, no. 12, 2019, pp. 5809–5823.

[3] X. Wang, Y. Hua, E. Kodirov, G. Hu, R. Garnier, and N. M. Robertson, “Ranked list loss for deep metric learning,” in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pp. 5207–5216.

[4] Y. Movshovitz-Attias, A. Toshev, T. K. Leung, S. Ioffe, and S. Singh, “No fuss distance metric learning using proxies,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22- 29, 2017, 2017, pp. 360–368.

[5] V. E. Liong, J. Lu, and Y. Ge, “Regularized bayesian metric learning for person re-identification,” in Computer Vision - ECCV 2014 Workshops Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III, 2014, pp. 209–224.

[6] T. Xiao, J. Ren, Z. Meng, H. Sun, and S. Liang, “Dynamic Bayesian metric learning for personalized product search,” in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019, 2019, pp. 1693–1702.

Qifan Liu, received the M.S. and Ph.D. degrees from the College of Electronics and Information Engineering, Shenzhen University, Guangdong, China, in 2020, and 2023, respectively. His research interests include metric learning, few-shot learning, image classification, and deep learning.

A Measurement Study of Volumetric Video Codecs

A short review for "A Comparative Measurement Study of Point Cloud-Based Volumetric Video Codecs"

Edited by Takuya Fujihashi

M. Yang. Z. Luo, M. Hu, M. Chen, D. Wu, "A Comparative Measurement Study of Point Cloud-Based Volumetric Video Codecs," IEEE Transactions on Broadcasting, 2023 (early access).

In contrast to the traditional 2D and omnidirectional video content, volumetric video enables users to watch an object and a scene from freely switchable angles and positions. The volumetric video is significantly expected for various fields such as medicine, education, entertainment, and so on. The typical formats of the volumetric video are point cloud or polygon mesh, and the point cloud has been a primary format in recent years because of the acquisition cost and the easy conversion to the polygon mesh.

The naïve delivery of point cloud video over the Internet requires massive bandwidth and computation resources. Specifically, each point cloud frame consists of non-uniformly distributed 3D points with attributes of 3D coordinates and color components. The representation of each 3D point typically takes 15 bytes, 4 bytes for each attribute of the 3D coordinates, and 1 byte for each color component. For example, a point cloud for entertainment contains over one million points per frame. If not compressed, a bandwidth of 3.6Gbps is required to achieve a playback at 30 frames per second. It means a deep compression is required to deliver the point cloud over the Internet.

The codecs in the existing point cloud-based volumetric video streaming systems can be divided into two main categories: conventional codecs [1-4] and neural-based codecs [5, 6].
The conventional codecs use the data projection technique to represent a point cloud with a 2D or 3D-tree structure. V-PCC maps the 3D points onto a 2D space and compresses the 2D-projected 3D points using 2D video compression methods, such as H.264/Advanced Video Coding and H.266/Versatile Video Coding. In contrast, Point Cloud Library, Draco, and G-PCC use 3D-tree structures, such as octree and kd-tree, to compress the 3D points. The neural-based codecs used the deep learning-based super-resolution method for the point cloud to compress the 3D points.

The above codecs have been proposed for volumetric video streaming. However, there needs to be more work to study the efficiency and practicality of codecs. This paper investigates essential questions. The first question is to understand the appropriate application scenarios for each codec. For example, a codec may require a large computation time for encoding and decoding. In this case, the codec is inadequate for live and real-time volumetric video streaming. For this purpose, this paper measures the performance of the codecs in terms of compression ratio, speed, and video quality. This paper considers chamfer distance, peak signal-to-noise ratio (PSNR), color-PSNR, and hybrid PSNR.
The second question is the impact of the point cloud features (e.g., video quality, texture, and geometric complexity) on the codec performance and delivery efficiency. This paper uses six volumetric video datasets (Pose, Office, Haggling, Pizza, Longdress, and Soldier) with different characteristics to investigate the effect of the characteristics.
The third question is the effect of the user's viewing behavior on the codec performance. Since users can watch volumetric videos from free angles and positions based on the user's viewing angle and position, the behavior of the viewing angle and position may significantly affect the codec and delivery efficiency. This paper collects viewing traces of eight users by recording viewport positions and orientations for each frame to study the effect of the viewing behavior.

This paper measures the performance of all codecs on a computer equipped with an AMD Ryzen 5 1600X 6-core 3.60GHz CPU and an NVIDIA GeForce GTX 1080Ti GPU. Here, the authors implemented the codecs of Draco, G-PCC, V-PCC, PU-GCN+, and MPU+.From evaluation results, this paper highlights the investigation results as follows:

A simple implementation of all existing codecs can hardly meet the delay requirement of live volumetric video streaming. For on-demand volumetric video streaming, G-PCC is more suitable for delivering high-quality point clouds, while Draco outperforms other codecs when the size of a point cloud is small or there is sufficient bandwidth.

The texture and scene complexity and the noise level of point clouds significantly affect the compression performance of codecs. On the contrary, the geometric complexity has little impact.

A user's viewing behavior significantly affects the delivery performance of volumetric videos. Especially for videos with multiplayer scenes, it is hard to maintain high QoE during viewing. For videos with single-person scenes, G-PCC and Draco outperform others.

References:

[1] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," APSIPA Trans. Signal Inf. Process., vol. 9, p. e13, Apr. 2020.

[2] Draco 3D Data Compression. https://google.github.io/draco/.

[3] Point Cloud Library (PCL). http://pointclouds.org/

[4] B. Han, Y. Liu, and F. Qian, "ViVo: Visibility-aware mobile volumetric video streaming," in Proc. 26th Annual International Conference on Mobile Computing and Networking, Sep. 2020, pp. 1–13.

[5] Y. Wang, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung, "Patch-based progressive 3D point set upsampling," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 5951–5960.

[6] D. Chen, Y.-J. Chiang, and N. Memon. Lossless Compression of Point-Based 3D Models. In Proceedings of the 13th Pacific Conference on Computer Graphics and Applications, 2005.

[7] Qian, A. Abualshour, G. Li, A. K. Thabet, and B. Ghanem, "PU-GCN: Point cloud upsampling using graph convolutional networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2021, pp. 11683–11692

Takuya Fujihashi received the B.E. degree in 2012 and the M.S. degree in 2013 from Shizuoka University, Japan. In 2016, he received Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan. He is currently an assistant professor at the Graduate School of Information Science and Technology, Osaka University since April, 2019. He was an assistant professor at the Graduate School of Science and Engineering, Ehime University, Japan from Jan. 2017 to Mar. 2019. He was research fellow (PD) of Japan Society for the Promotion of Science in 2016. From 2014 to 2016, he was research fellow (DC1) of Japan Society for the Promotion of Science. From 2014 to 2015, he was an intern at Mitsubishi Electric Research Labs. (MERL) working with the Electronics and Communications group. He selected one of the Best Paper candidates in IEEE ICME (International Conference on Multimedia and Expo) 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming.

Class-covariance metric network design for few-shot hyperspectral image classification

A short review for "Few-Shot Learning With Class-Covariance Metric for Hyperspectral Image Classification"

Edited by Wenming Cao

B. Xi, J. Li, Y. Li, R. Song, D. Hong and J. Chanussot, "Few-Shot Learning With Class-Covariance Metric for Hyperspectral Image Classification," in IEEE Transactions on Image Processing, vol. 31, pp. 5079-5092, 2022, doi: 10.1109/TIP.2022.3192712.

Hyperspectral image (HSI) is captured through dozens to hundreds of narrow and contiguous bands. Benefited from the refined spectral information, it has been investigated in a variety of tasks, such as target/abnormal detection, fusion, etc [1]. In HSI, each pixel is represented by a spectral curve containing wealthy discriminative information, which can be utilized to identify its owned category precisely [2]. Thus, HSI classification (HSIC) has been applied to various applications, such as medical care, environmental protection, and urban planning, to name a few[3]. However, due to the high-dimensional spectral signal versus limited annotated training samples, HSIC with high accuracy is still a challenging problem and deserves further study [4].

To settle this issue, researchers have devoted their efforts to various learning strategies and feature engineering. The former explores different learning schemes to maximize the prior information embodied in the limited labeled samples, and even involve the unlabeled data and the data from other fields. The latter seeks to reduce the feature dimension of the hyperspectral data while preserving and enhancing the contained discriminative information. Particularly, few-shot learning (FSL) is one of the most effective learning paradigms, which can rapidly generalize to new tasks with prior knowledge and limited supervisory experience.

Leveraging meta-learning (learning to learn)[5], many FSL approaches based on embedding and metric learning methods are developed and introduced to HSIC tasks [6]. Generally, these methods aim to learn a transformation function by reorganized meta-tasks (i.e., episodes including support and query samples), such that when projected into the embedding space, novel-class samples are easy to distinguish through using a linear classifier based on the distance measure.

To further enhance the performance with few labeled samples, the authors in this paper propose a novel FSL framework for HSIC with a class-covariance metric (CMFSL). Overall, the CMFSL learns global class representations for each training episode by interactively using training samples from the base and novel classes, and a synthesis strategy is employed on the novel classes to avoid overfitting. During the meta-training and meta-testing, the class labels are determined directly using the Mahalanobis distance measurement rather than an extra classifier. Benefiting from the task-adapted class-covariance estimations, the CMFSL can construct more flexible decision boundaries than the commonly used Euclidean metric. Additionally, a lightweight cross-scale convolutional network (LXConvNet) consisting of 3D and 2D convolutions is designed to thoroughly exploit the spectral-spatial information in the high-frequency and low-frequency scales with low computational complexity. Furthermore, the authors devise a spectral-prior-based refinement module (SPRM) in the initial stage of feature extraction, which cannot only force the network to emphasize the most informative bands while suppressing the useless ones, but also alleviate the effects of the domain shift between the base and novel categories to learn a collaborative embedding mapping.

The main contributions of the proposed CMFSL are as follows: (1) A novel FSL framework for HSIC is proposed, which can achieve meta task adapted classification and obtain state-of-the-art performance with few-shot labeled samples by learning a class-covariance-based metric space. (2) Inspired by the efficient Octave convolutions, the authors propose an LXConvNet, which is able to exploit the spectral-spatial features across high-frequency and low-frequency scales by joint 3DCONV and 2DCONVs, and transform the 3D patch samples into the discriminative embedding space with comparatively low computational complexity. (3) The authors also propose a new spectral-prior-based refinement module (SPRM). Apart from adaptively refining the band-wise information of the 3D patch samples, SPRM is supposed to reduce the domain shift between the source and target data sets in the initial stage of the feature extraction.

The authors conduct extensive experiments on four benchmark datasets, showing that the proposed CMFSL can outperform state-of-the-art methods with few annotated samples. In addition, the authors also analyze the computational complexity of the proposed CMFSL. They compare the number of parameters and inference time of the CMFSL with the FSL-related HSIC methods.

In summary, the authors propose a novel FSL framework with a class-covariance metric (CMFSL) for accurate HSI classification. First of all, to exploit the information of the few-shot annotated labels, the network is interactively trained between the base and novel classes, aiming to learn a collaborative discriminative embedding space for them. Secondly, to emphasize the most informative bands and narrow the domain shift between the source and target classes, an SPRM is designed in the initial phase of the feature extraction, which reasonably utilizes the spectral prior information. Thirdly, the authors propose an LXConvNet consisting of 3D and 2D convolutions to investigate the spectral-spatial features across the high-frequency and low-frequency scales with comparatively low computational complexity. The experiments demonstrate that the CMFSL can achieve superior results than other methods with few-shot labeled samples. Moreover, the used Mahalanobis distance with task adapted class-covariance estimations is more competitive in modeling the decision boundaries than the Euclidean and Cosine metrics in the proposed frameworks.

References:

[1] Ghamisi P, Yokoya N, Li J, et al. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art[J]. IEEE Geoscience and Remote Sensing Magazine, 2017, 5(4): 37-78.

[2] Xi B, Li J, Li Y, et al. Multiscale context-aware ensemble deep KELM for efficient hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(6): 5114-5130.

[3] Paoletti M E, Haut J M, Plaza J, et al. Deep learning classifiers for hyperspectral imaging: A review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 158: 279-317.

[4] He L, Li J, Liu C, et al. Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(3): 1579-1597.

[5] Huisman M, Van Rijn J N, Plaat A. A survey of deep meta-learning[J]. Artificial Intelligence Review, 2021, 54(6): 4483-4541.

[6] Liu B, Yu X, Yu A, et al. Deep few-shot learning for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(4): 2290-2304.

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 Semiconductors, Chinese Academy of Sciences, Beijing, China. He is currently a Professor with Shenzhen University, Shen zhen, China. He is also a foreign academician of the Russian Academy of Natural Sciences. He has authored or coauthored over 80 publications in toptier conferences and journals. His research interests include pattern recognition, image processing, and visual tracking.

Meta Learning for Sample Reweighting in Unsupervised Domain Adaptation

A short review for "Meta-Reweighted Regularization for Unsupervised Domain Adaptation"

Edited by Guitao Cao

S. Li, W. Ma, J. Zhang, C. H. Liu, J. Liang and G. Wang, "Meta-Reweighted Regularization for Unsupervised Domain Adaptation," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 3, pp. 2781-2795, 1 March 2023.

Unsupervised Domain Adaptation (UDA) [1] is a pivotal challenge in the field of machine learning, addressing the issue of effectively transferring knowledge gained from a labeled source domain to an unlabeled target domain. Traditional UDA methods endeavor to bridge the domain gap by minimizing the distribution discrepancy between the source and target domains [2]. Adversarial learning techniques [3] have gained prominence by introducing domain classifiers that encourage the extraction of domain-invariant features. Despite their effectiveness, challenges such as model instability and residual domain bias persist.

Recent studies have explored self-training [4] as a means to bolster UDA efforts. Self-training refines classification decision boundaries using target domain samples and their pseudo labels. However, the presence of noisy pseudo labels introduces a substantial hurdle, prompting the design of intricate target selection strategies or optimization goals to counteract their adverse effects.

In this paper, the authors adopt a meta-learning [5] framework to optimize the reweighting of pseudo-labeled target samples, and it can integrate with any adversarial-based UDA methods. Their central intuition is that an ideally trained target classifier should exhibit minimal classification errors on source samples resembling the target domain. Consequently, the paper design a meta reweighting objective to identify the importance of each pseudo-labeled target sample in the current training stage, which is effectively addressed using a simplified approximation technique.

The paper explicitly defines the meta reweighting problem, which seeks to find optimal weights for various pseudo-labeled target samples by minimizing the classification loss on a validation set. This validation set comprises source samples that are both class-balanced and most similar to target samples. Technically, the authors utilize the domain discriminator to select source samples with a domain score being about 0.5, indicating that the data located around the domain decision boundary are actually domain confused samples, which are domain informative and more robust to the unexpected noise.

Initially, instance weights are set uniformly across target domain samples. This initializes the adaptation process and allows for gradual refinement. A meta-learner is introduced to guide the instance weight updates. The meta-learner learns to provide informative guidance based on the model's performance on the source and target domains. Using the guidance from the meta-learner, the instance weights for target domain samples are updated. Samples that contribute positively to reducing the domain gap receive higher weights, while those hindering adaptation receive lower weights. This dynamic reweighting allows the model to adapt selectively to different target samples.

A meta-reweighted regularization loss is introduced to enforce the instance weight adaptation and maintain consistency across domains. This loss encourages the model to make better use of samples with higher weights while minimizing the negative impact of noisy or irrelevant samples. The regularization loss is added to the primary task loss during training. By integrating the optimized weights into the training process, the algorithm guides the model to adapt more effectively to the target domain.

The paper employs a simplified approximation technique to tackle the optimization challenge, which ensures that the optimization process remains computationally tractable and can be efficiently solved. The process of meta-reweighted regularization and instance weight adaptation is iterative. The model is trained for multiple epochs, and in each epoch, the instance weights are updated based on the guidance from the meta-learner. This iterative approach ensures progressive alignment between the source and target domains.

To validate the effectiveness of the proposed method, the authors conducted extensive experiments on benchmark datasets commonly used in UDA research. Quantitative metrics such as accuracy, precision, recall, and F1-score are used to measure performance improvements over existing methods. Ablation studies are conducted to analyze the impact of different components of the approach, providing insights into their contributions.

In summary, the paper's contributions are significant, particularly in introducing meta-reweighted regularization to UDA. This approach provides a fresh perspective on addressing domain shift by dynamically adjusting instance weights. The concept of meta-learning enhances the adaptability of the model and allows for better exploitation of shared information between domains. The innovative approach enriches the domain adaptation landscape and paves the way for future research in this area. Further exploration could involve investigating the method's performance on more diverse datasets, understanding its behavior in scenarios with extreme domain shifts, and devising techniques to mitigate potential challenges related to the approach's complexity.

References:

[1] S. J. Pan and Q. Yang, "A survey on transfer learning", IEEE Trans. Knowl. Data Eng. (TKDE), vol. 22, no. 10, pp. 1345-1359, 2010.

[2] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf and A. J. Smola, "A kernel method for the two-sample-problem", Proc. Int. Conf. Neural Inf. Process. Syst, pp. 513-520, 2007.

[3] Y. Ganin et al., "Domain-adversarial training of neural networks", J. Mach. Learn. Res., vol. 17, no. 1, pp. 2030-2096, May 2015.

[4] M. Long, J. Wang, G. Ding, J. Sun and P. S. Yu, "Transfer feature learning with joint distribution adaptation", Proc. IEEE Int. Conf. Comput. Vis., pp. 2200-2207, 2013.

[5] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, “Meta learning in neural networks: A survey,” 2020, arXiv: 2004.05439.

Prof. Guitao Cao obtained her Ph.D. in 2006 from Shanghai Jiao Tong University with a focus on pattern recognition. She is currently a professor of Software Engineering Institute, East China Normal University (ECNU), Shanghai. ECNU is the top tier university in China with a high rank (Level A) in Software Engineering in China. She was also a visiting researcher with University of Missouri Columbia. She has published decades of peer reviewed papers in top venues including IEEE Transactions on Cybernetics, IEEE Transactions on Multimedia, and IEEE Transactions on Biomedical Engineering. Prof. Cao is also the Principal Investigator for many research funding with major sponsors including the National Science Foundation of China, Ministry of Industry and Information Technology of the People's Republic of China, and Science Foundation of Shanghai. Her research interests include AIoT, embedded systems, federated learning, pattern recognition and machine learning.

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.