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

MMTC Communications – Review

Vol. 13, No. 5, December 2022

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the December 2022 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises four reviews that cover multiple facets of multimedia communication research including remote visual monitoring, multimodal emotion analysis, ambient backscatter communication, and vision transformer model compression. These reviews are briefly introduced below.

The first paper, published in IEEE Transactions on Multimedia and edited by Dr. Dong Li, proposes a multi-template update strategy for effective visual monitoring of multimedia environment. It is motivated by the three stages of human visual memory model: matching memory, confidence memory, and cognitive memory. This strategy allows more accurate target monitoring while achieving real-time monitoring performance.

The second paper, edited by Dr. Qin Wang, was also published in IEEE Transactions on Multimedia. This paper proposes a multimodal emotion analysis model that includes three stages: feature mapping, interactive learning, and feature fusion. Results show that the proposed model can outperform baseline models by a large margin.

The third paper, edited by Dr. Ye Liu, was published in IEEE Transactions on Communications. It proposes a novel opportunistic ambient backscatter-assisted decode-and-forward relaying scheme. Results show that the proposed scheme can achieve significant performance gain compared to the traditional ambient backscatter scheme and the traditional decode-and-forward scheme.

The fourth paper, published in the 2022 IEEE International Conference on Multimedia and Expo (ICME 2022), and edited by Dr. Tiesong Zhao, proposes a vision transformer (ViT) model compression method that can jointly prunes a pre-trained ViT model via attention head, neuron, and sequence dimensions. Results show that the proposed approach can effectively reduce the computational cost of various models and outperforms pervious state-of-the-art methods.

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

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

Exploring Multi-Layer Template Updating for Remote Visual Monitoring

A short review for "Human Memory Update Strategy: A Multi-Layer Template Update Mechanism for Remote Visual Monitoring"

Edited by Dong Li

S. Liu, S. Wang, X. Liu, A. H. Gandomi, M. Daneshmand, K. Muhammad, and V. H.C. Albuquerque, "Human Memory Update Strategy: A Multi-Layer Template Update Mechanism for Remote Visual Monitoring," in IEEE Transactions on Multimedia, vol. 23, pp. 2188-2198, 2021.

Recently, computer vision (CV) has received much attention both in itself and as a technological catalyst for spurring innovation in the field of artificial intelligence (AI) [1], which is able to replace the human eye with cameras and computers, enabling the computer to achieve the functions of segmenting, classifying, identifying, tracking, and discriminating decisions like the human visual system [2]. In particular, target monitoring is one of the vital fields of CV, which has been widely applied in military reconnaissance, 3-D transmission, fire scene analysis, battlefield assessment, and security monitoring [3-4].

However, most existing works on target monitoring are not always possible to cope with the complex environmental characteristics under the current multimedia background, such as illumination variation, occlusion, fast motion and so on (such as, [5-6]). The main reason is that traditional target monitoring algorithms often adopt static templates, which may lead to target monitoring failure and insufficient robustness and effectiveness.

In order to solve this issue, this paper proposes a multi-layer template update strategy into the correlation filtering-based monitoring algorithm, where the human memory update strategy is considered to achieve effective monitoring in the multimedia environment. In a nutshell, the main contributions of this paper are twofold, which are summarized as follows:

First, the authors propose a visual memory update strategy with a multi-layer for effective monitoring in a multimedia environment, in which three kinds of memories are involved: confident memory, matching memory, and cognitive memory. Specifically, the confident memory is used to store the weighted template of the high-confidence matching memory. The cognitive memory is used to store the unweighted template of the low-confident matching template in real-timely. When the real-time cognitive memory is reliable, both the confident and matching memory are updated; otherwise, only the matching memory is updated.

Second, an alternate template selection-based scheme is proposed, where the matching template is applied to monitor when the current frame is well-tracked while the template with higher reliability in confident and cognitive memories is adopted to monitor when the current frame is hard to track. The target is not lost even if it is not tracked for a few frames because the template is performed alternately.

For the performance evaluation, extensive simulation results are conducted to compare the proposed scheme with the existing schemes in terms of accuracy and success rate. Not surprisingly, the proposed scheme results are significantly better than those of the compared schemes. Besides, to further prove the effectiveness of the proposed scheme, some benchmarks are compared to show its superiority.

In summary, this paper presents a novel solution and solid theoretical contributions for target monitoring at dynamic and complex environmental characteristics.

References:

[1] L. Deng, "Artificial intelligence in the rising wave of deep learning: The historical path and future outlook," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 180–177, Jan. 2018.

[2] D. K. Shetty, U. Dinesh Acharya, N. Malarout, R. Gopakumar, and P. P. J., "A review of application of computer-vision for quality grading of food products," in *Proc. Int. Conf. Automat., Comput. Technol. Manage*. London, United Kingdom, pp. 297–303, 2019.

[3] M. Asad *et al.*, "A split target detection and tracking algorithm for ballistic missile tracking during the re-entry phase," *Defence Technol.*, vol. 16, no. 6, pp. 1142–1150, 2019.

[4] Y. Li *et al.*, "Robust estimation of similarity transformation for visual object tracking," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 8666–8673, 2019.

[5] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., pp. 4800–4808, 2017.

[6] Y. Yang, J. Yang, L. Liu, and N. Wu, "High-speed target tracking system based on a hierarchical parallel vision processor and gray-level LBP algorithm," *IEEE Trans. Syst.,Man, Cybern.Syst.*, vol. 47, no. 6, pp. 950–964, Jun. 2017.

Dong Li received the Ph.D. degree in electronics and communication engineering from Sun Yat-Sen University, Guangzhou, China, in 2010. Since 2010, he has been with the School of Computer Science and Engineering (formally, Faculty of Information Technology), Macau University of Science and Technology, Macau, China, where he is currently an Associate Professor. He held a visiting position with the Institute for Infocomm Research, Singapore, in 2012. His current research interests focus on Backscatter Communications, Intelligent Reflecting Surface-Assisted Communications, and Machine Learning for Communications. He was a recipient of the Bank of China (BoC) Excellent Research Award by Macau University of Science and Technology in 2011, 2016, 2019 and 2021, respectively, and was listed among World's Top 2% Scientists recognized by Stanford University in 2020 and 2021. He is currently an Executive Board Member of IEEE Macau Section, and an editor for IEEE MMTC Review.

Image-text Multimodal Emotion Classification via Multi-view Attentional Network

A short review for "Image-text Multimodal Emotion Classification via Multi-view Attentional Network"

Edited by Qin Wang

X. Yang, S. Feng, D. Wang, Y. Zhang, "Image-text Multimodal Emotion Classification via Multi-view Attentional Network," in IEEE Transactions on Multimedia, vol. 23, pp. 4014-4026, 2020.

The increased use of mobile Internet and smartphones has provided researchers with massive archives of multimodal user-generated content on diverse topics and entities. For the task of extracting and analyzing the emotions contained in these data, although the existing research has achieved good results, most of the literature focuses on tasks using single modal data, such as text emotion polarity classification [1] and image emotion recognition [2], while ignoring the vivid and complementary emotional information in multimodal data.

Early studies required handcrafted feature engineering for each modality, which is a potentially biased and labor-intensive method [3]. Although some outstanding deep learning models are available for multimodal sentiment analysis, most existing methods treat the representation learning process of each modality separately and fuse the learned multimodal features at a higher level of the neural network. Furthermore, the cross-modality interactions between different modalities, such as images and text, have received relatively little attention. In this paper, the authors focus on multimodal emotion analysis for image-text pairs in social media posts.

Through the study of traditional and deep learning-based methods, there are three observations in multimodal sentiment analysis. First, the emotions are not isolated to a single data modality; on the contrary, the emotions in the texts and images are complementary and express the users' sentiments and feelings. However, in the literature, each modal feature is usually modeled separately, and the cross-modal interactions between image and text are ignored. Second, when looking at an image, people usually focus on the part of the image in which they are interested, rather than considering the entire image content equally. Although some methods have successfully utilized object or scene features for sentiment analysis, none have considered multi-view features in a unified framework. Third, the multimodal emotion analysis task is much more difficult than the multimodal sentiment polarity analysis task, not only because there are more emotion categories, and the existing models have defects but also because of the lack of large training datasets for multimodal deep learning models.

To tackle these challenges, in this paper, the authors propose a novel Multi-view Attention Network (MVAN) to achieve robust and accurate multimodal emotion analysis. MVAN consists of three stages, i.e., feature mapping, interactive learning, and feature fusion, and it explores cross-modal interactions and considers the mutual reinforcement between text and image.

By observing the image from multiple views or different feature subsets [4, 5], e.g., the image object view and the scene view, the authors can capture various beneficial features for multimodal emotion analysis. In the feature mapping stage, local object features and scene features are extracted from the image to obtain deep semantic features from a multi-view perspective. In the interactive learning stage, the authors adopt the image-text interactive learning mechanism. Specifically, the text features are learned from the self-influence of the text under the guidance of image object and scene features. Similarly, the text features help the model learn both the image object features and the scene features. In the feature fusion stage, the four learned features are concatenated and then, to improve the accuracy and F1-score of multimodal emotion analysis, the features are deeply fused through a multilayer perceptron and a stacking-pooling module.

Because no publicly available dataset exists for multimodal emotion analysis, the authors crawled Tumblr to obtain a large-scale dataset of text-image pairs and used the distant supervision method to label the obtained data. The result is a multimodal emotion analysis dataset named TumEmo. The experimental results on the publicly available MVSA-Single, and MVSA-Multiple datasets [6] and on the TumEmo dataset show that the methods proposed in this paper perform satisfactorily on the different multimodal classification tasks.

The authors conducted the experiments with the proposed models and the baselines on three datasets. The analysis is as followed.

Results of the Baseline Methods and the proposed model indicates that the authors' model (MVAN) outperforms the other models in terms of accuracy and F1-score. And the multimodal sentiment analysis models perform better than do most of the single-modal sentiment analysis models on all three datasets.

The authors conduct ablation experiments on the MVAN-M model to verify the effectiveness of different modules. The results indicate that these two modules are effective for multimodal sentiment analysis. This result is achieved mainly because the attention memory network causes both the text and image data to participate in auxiliary learning, which promotes learning during sentiment analysis. The deep fusion module enables the learned text and image features to be effectively fused in a high-dimensional space.

The authors also conduct experiments under different settings of the hyperparameter HOP. The results show that when HOP=3, the accuracy and F1-score reach their maximum values on MSV A-Single, MSV A-Multiple and TumEmo.

In summary, the authors proposed a novel multimodal emotion analysis model based on a multi-view attention network, which interactively learns text and image features through an attention memory network module. The experimental results on two publicly available datasets and the built dataset demonstrated that the authors' proposed model outperforms strongly competitive baseline models by a large margin.

References:

[1] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, "Deep learning-based sentiment classification of evaluative text based on multi-feature fusion," Information Processing & Management, vol. 56, no. 4, pp. 1245– 1259, 2019.

[2] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," Neurocomputing, vol. 333, pp. 429–439, 2019.

[3] F. Wu, Y. Huang, Y. Song, and S. Liu, "Towards building a high-quality microblog-specific chinese sentiment lexicon," Decision Support Systems, vol. 87, pp. 39–49, 2016.

[4] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," arXiv preprint arXiv:1304.5634, 2013.

[5] S. Sun, L. Mao, Z. Dong, and L. Wu, Multiview machine learning. Springer, 2019.

[6] T. Niu, S. Zhu, L. Pang, and A. El Saddik, "Sentiment analysis on multi-view social data," in International Conference on Multimedia Modeling, pp. 15–27, Springer, 2016.

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.

Opportunistic Ambient Backscattering through Cooperative Communication

A short review for "Two Birds With One Stone: Exploiting Decode-and-Forward Relaying for Opportunistic Ambient Backscattering"

Edited by Ye Liu

D. Li, "Two Birds With One Stone: Exploiting Decode-and-Forward Relaying for Opportunistic Ambient Backscattering," in IEEE Transactions on Communications, vol. 68, no. 3, pp. 1405-1416, Mar. 2020.

The Internet of Things (IoT) has become a new infrastructure in our society. With the capability of collaborative sensing, computing, intelligence, and all-stack cyber security and privacy protection, the developed various IoT systems have supported our smart city, smart home, smart health, Industry 4.0, and Agriculture 4.0. Currently, the IoT presents four trends [1], namely, (i) from simple environment monitoring to high-throughput sensing, (ii) from laboratory research to field experiment, (iii) from small-scale sparse observation to large-scale precision measurement, and (iv) from coarse-grained sampling to fine-grained recording.

The above trends pose many challenges facing engineers and academic researchers. First, high-throughput sensing usually requires power-hungry sensors for image recording, sound monitoring, or even video streams. Second, it heavily brings the maintenance overhead due to the massive IoT nodes deployed in the field. Finally, dense networks and high-duty cycles are needed to achieve large-scale fine-rained monitoring. To fill these gaps, zero-power IoT [2] pays excellent attention to industry and academia. It is a blueprint that the IoT nodes are able to harvest energy from their surroundings and operate with minimal energy consumption.

Backscatter communication [3,4] is a crucial technology towards zero-power IoT because it can perform ultra-low-power wireless communication. This is achieved by exploring ambient wireless signals instead of generating carrier waves on their own. However, although backscatter communication is very promising, it is still in its infancy. Many fundamental issues are to be solved, including system design in various scenarios, network framework, medium access control, multi-hop relay, parallel decoding, and so on.

In this paper, the author proposed a novel opportunistic ambient backscatter approach, which introduces a concept of decode-and-forward relaying in ambient backscatter communication with a cooperative relay. Moreover, the modeling of such an opportunistic backscatter system was established for the first time. In addition, the proposed approach was compared with the traditional ambient communication method and traditional decode-and-forward scheme through extensive numerical analysis. The results show a significant improvement in the proposed system.

A basic system model in the proposed opportunistic decode-and-forward ambient backscatter communication consists of three units: one source node, one relay node, and one destination. Especially a tag is embedded in the relay for an assistant. In a transmission process, the time period is divided into two slots, where the first time slot is used for signal decoding and forwarding. The uniqueness is that the relay sends not only the received data but also the information from the tag simultaneously. After that, the second time slot is responsible for data relay as in traditional decode-and-forward relaying.

The core idea of the proposed system is to explore the extra power between the source and the relay to transmit the data generated from the passive tag with nothing affection to the decode-and-forward relaying. Power splitting and hybrid signal backscattering/transmission are then designed carefully with detailed theoretical formulation to achieve this goal. Interestingly, an in-depth discussion of the practical considerations for the

new system is also presented, making this idea come true in a real system. Furthermore, the ergodic capacity analysis [5] of the proposed approach is analyzed.

The standard ambient backscatter communication system and standard decode-and-forward relaying are chosen as benchmark schemes for the performance evaluation. These two systems and the proposed one are different regarding total transmit powers, number of time slots, duplex mode, and received signal. Extensive simulations are conducted to analyze the ergodic capacity of the three systems in conditions with different emitting power, communication distance, and the power of the additive white Gaussian noise. The results show that the proposed approach is outstanding.

In summary, this paper presents a novel opportunistic ambient backscatter-assisted decode-and-forward relaying scheme. It is believed that this work will provide an important contribution to the field of backscatter communication and zero-power IoT.

References:

[1] Y. Liu, D. Li, B. Du, L. Shu, and G. Han, "Rethinking Sustainable Sensing in Agricultural Internet of Things: From Power Supply Perspective," in IEEE Wireless Communications, vol. 29, no. 4, pp. 102-109, 2022.

[2] Y. Liu, D. Li, H. Dai, C. Li, and R. Zhang, "Understanding the Impact of Environmental Conditions on Zero-Power Internet of Things: An Experimental Evaluation," in IEEE Wireless Communications, 2022.

[3] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient Backscatter Communications: A Contemporary Survey," in IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2889-2922, Fourthquarter 2018.

[4] W. U. Khan, A. Ihsan, T. N. Nguyen, Z. Ali, and M. A. Javed, "NOMA-Enabled Backscatter Communications for Green Transportation in Automotive-Industry 5.0," in IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 7862-7874, Nov. 2022

[5] I. Krikidis, H. A. Suraweera, P. J. Smith, and C. Yuen, "Full-Duplex Relay Selection for Amplify-and-Forward Cooperative Networks," in IEEE Transactions on Wireless Communications, vol. 11, no. 12, pp. 4381-4393, December 2012.

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.

An Efficient Multi-Dimensional ViT Pruning Paradigm

A short review for "Multi-Dimensional Model Compression of Vision Transformer"

Edited by Tiesong Zhao

Zejiang Hou and Sun-Yuan Kung. "Multi-dimensional model compression of vision transformer," IEEE International Conference on Multimedia and Expo (ICME'22), Taipei, Taiwan, Jul. 18-22, 2022.

Recently, image segmentation and recognition have been greatly benefited from the development of vision transformer (ViT) [1, 2]. Nevertheless, when implementing ViT in resource-limited tasks such as real-time processing, a large challenge still exists: its high computational cost. Until now, there are few works that accelerate transformers in vision tasks, e.g. [3, 4, 5, 6, 7]. Among them, [3, 4] applies structured neuron pruning or unstructured weight pruning, [5, 6] applies dynamic or static token sparsification, [7] uses post-training quantization to reduce the ViT model size. To further reduce the computational cost, excessive pruning of single sequence might lead to unacceptable accuracy loss, as shown in this paper. Inspired by this, the authors improve ViT pruning by a joint optimization of multiple network modules.

It is also noted that network pruning is not a new topic in convolutional neural networks (CNNs). To effectively implement CNNs in real-time tasks or chips, the network pruning methods, including multi-dimensional pruning, have been widely studied in the past decade [8, 9, 10]. Despite of these great efforts, the authors argue that their effectiveness for ViT is not immediately clear and thus it is still imperative to study the multi-dimensional compression for ViT.

In this paper, the authors propose a multi-dimensional compression that works on three dimensions: the number of neurons in feed-forward network (FFN), attention heads in multi-head self-attention (MHSA), and the sequence length. First, they propose a pruning criterion by calculating the statistical dependency between the model features and the output predictions. Mathematically, it is calculated as Hilbert-Schimdt norm of the cross-variance operator. With such a criterion, they are able to identify the unimportant features with least contributions to the output predictions. Second, they formulate an optimization problem of multi-dimensional compression, which tries to find the optimal pruning ratios of all dimensions, in order to maximize the model accuracy under a target complexity. Unfortunately, this optimization problem does not have a closed-form solution. Third, they propose to use Gaussian process (GP) search with expected improvement (EI) to estimate the model accuracy for different pruning ratios, thereby transform the abovementioned optimization to a non-linear programming task. In particular, they apply weight sharing [11] to efficiently evaluate the model accuracy with different pruning ratios.

To evaluate the proposed method, the authors implemented it on representative ViT models and tested them on ImageNet, DeiT and T2T-ViT datasets. Experimental results showed that the proposed method outperformed its peers with superior accuracy under the same FLOPs reductions. For DeiT and T2T-ViT models, the proposed method reduced 40% PLOPs without top-1 accuracy loss. Ablation studies also showed the effectiveness of method design, including multi-dimensional compression, pruning criterion and GP search.

Nowadays, the network pruning has been a popular topic in intelligent multimedia processing and computer vision tasks. Many researchers have devoted their efforts to deep neural networks, especially for the CNNs. Although ViT has attracted attentions of computer vision society, its practical usage is still limited due to its large

computational cost. This problem is even imperative considering its potential applications in multimedia computing and communication. To address this problem, this paper proposes the first multi-dimensional compression paradigm for ViT, with solid theoretical derivation and promising experimental results. From this point, this work may attract more attentions of researchers to the practical application of ViT.

References:

[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR,* 2021.

[2] H. Touvron, M. Cord, M. Douze, F. Massa, et al., "Training data-efficient image transformers & distillation through attention," in *ICML,* 2021.

[3] M. Zhu, K. Han, Y. Tang and Y. Wang, "Visual transformer pruning," in KDD, 2021.

[4] T. Chen, Y. Cheng, Z. Gan, L. Yuan, et al., "Chasing sparsity in vision transformers: An end-to-end exploration," in NeurIPS, 2021.

[5] Y. Rao, W. Zhao, B. Liu, J. Lu, et al., "Dynamicvit: Efficient vision transformers with dynamic token sparsification," in NeurIPS, 2021.

[6] B. Pan, R. Panda, Y. Jiang, Z. Wang, et al., "Ia-red^2: Interpretability-aware redundancy reduction for vision transformers," in NeurIPS, 2021.

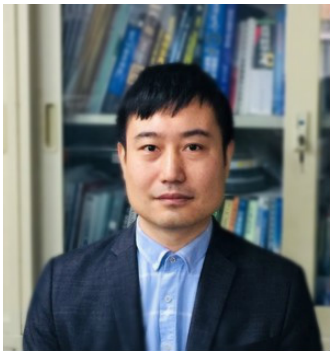
[7] Z. Liu, Y. Wang, K. Han, W. Zhang, et al., "Post-training quantization for vision transformer," in NeurIPS, 2021.

[8] S. Lin, R. Ji, C. Yan, B. Zhang, et al., "Towards optimal structured cnn pruning via generative adversarial learning," in CVPR, 2019.

[9] J. Guo, W. Ouyang and D. Xu, "Multi-dimensional pruning: A unified framework for model compression," in CVPR, 2020.

[10] W. Wang, M. Chen, S. Zhao, et al., "Accelerate cnns from three dimensions: A comprehensive pruning framework," in ICML, 2021.

[11] Z. Guo, X. Zhang, H. Mu, W. Heng, et al., "Single path one-shot neural architecture search with uniform sampling," in ECCV, 2020.



Tiesong Zhao, Ph.D, is a Minjiang Distinguished Professor in Fuzhou University, Fujian, China. He received the B. S. and PhD degree from the University of Science and Technology of China and City University of Hong Kong, in 2006 and 2011, respectively.

His research interests include multimedia signal processing, coding, quality assessment and transmission. Due to his contributions in video coding and transmission, he received the Fujian Science and Technology Award for Young Scholars in 2017. He has also been serving as an Associate Editor of IET Electronics Letters since 2019.

MMTC Communications – Review Editorial Board

Multimedia Communications Technical Committee Officers

Chair: Chonggang Wang, InterDigital, USA
Steering Committee Chairs: Shaoen Wu, Illinois State University, USA
Abderrahim Benslimane, University of Avignon, France
Vice Chair – America: Wei Wang, San Diego State University, USA
Vice Chair – Asia: Liang Zhou, Nanjing University of Post and Telecommunications, China
Vice Chair – Europe: Reza Malekian, Malmö University, Sweden
Letters & Member Communications: Qing Yang, University of North Texas, USA
Secretary: Han Hu, Beijing Institute of Technology, China
Standard Liaison: Weiyi Zhang, AT&T Research, USA

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.