Xiaoshan Yang

Professor · Multimedia Content Analysis, Computer Vision

1311, Intelligence Building, 95 Zhongguancun East Road, Beijing, China

xiaoshan.yang@nlpr.ia.ac.cn, yang.xiaoshan3@gmail.com | yangxs.ac.cn | userID=3bntCuQAAAAJ

Professional Experience

2024-	**Professor**, State Key Laboratory of Multimodal Artificial Intelligence Systems (MAIS), Institute of Automation, Chinese Academy of Sciences (CASIA)
2022-2024	**Associate Professor**, State Key Laboratory of Multimodal Artificial Intelligence Systems (MAIS), Institute of Automation, Chinese Academy of Sciences (CASIA)
2018-2022	**Associate Professor**, National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA)
2016-2018	**Assistant Professor**, National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA)

Education

NLPR, Institute of Automation, Chinese Academy of Sciences *Beijing, China*

PhD Pattern Recognition and Intelligent Systems *2012 - 2016*

- Advisor: Prof. Changsheng Xu

Beijing Institute of Technology *Beijing, China*

MS Computer Science *2009 - 2012*

- Advisor: Prof. Jianbing Shen, Prof. Yunde Jia

North China University of Technology *Beijing, China*

BS Mathematics and Applied Mathematics *2005 - 2009*

- Thesis advisor: Prof. Hu'an Li

Publications (Selected)

Zibo Shao, Baochen Xiong, **Xiaoshan Yang**, Yaguang Song, Qimeng Zhang, Haifeng Chen, and Changsheng Xu: PivotMerge: Bridging Heterogeneous Multimodal Pre-training via Post-Alignment Model Merging, IEEE Transactions on Multimedia (**TMM**), 2026

Kaichen Li, **Xiaoshan Yang**, Yaguang Song, Baochen Xiong, Yaowei Wang, and Changsheng Xu: Task Interference Mitigation via Conffdence-Stratiffed Classiffer Calibration for Prompt-based Continual Learning, IEEE Transactions on Image Processing (**TIP**), 2026

Baochen Xiong, **Xiaoshan Yang**, Yaguang Song, Yaowei Wang, and Changsheng Xu: A Generalized Framework for Modality-Incomplete Federated Learning, International Journal of Computer Vision (**IJCV**), 2026

Baochen Xiong, Yifan Xu, Yaguang Song, **Xiaoshan Yang**: FedPivot: Bridging Heterogeneous Multimodal Pre-training via Federated Post-Alignment, ACM International Conference on Multimedia (**MM**), 2026

Yuyang Wanyan, **Xiaoshan Yang**, Weiming Dong, Changsheng Xu: Modality-Collaborative Low-Rank Decomposers for Few-Shot Video Domain Adaptation, IEEE Transactions on Multimedia (**TMM**), 2026

Han Jiang, **Xiaoshan Yang**, Hongtao Xie, Chaoxu Mu, Yongdong Zhang, and Changsheng Xu, Active Concept-Collaborative Online Test-Time Adaptation for Compositional Zero-Shot Learning, IEEE Transactions on Multimedia (**TMM**), 2026

Fang Peng, **Xiaoshan Yang**, Yaowei Wang, Changsheng Xu: Group-Relative Visual Discrimination Enhancement for Unlocking Intrinsic Capability of MLLMs, IEEE Transactions on Circuits and Systems for Video Technology (**TCSVT**), 2026

Linhui Xiao, **Xiaoshan Yang**, Xiangyuan Lan, Yaowei Wang, Changsheng Xu: Towards Visual Grounding: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2025

Yuyang Wanyan, Xi Zhang, Haiyang Xu, Haowei Liu, Junyang Wang, Jiabo Ye, Yutong Kou, Ming Yan, Fei Huang, **Xiaoshan Yang**, Weiming Dong, Changsheng Xu: Look Before You Leap: A GUI-Critic-R1 Model for Pre-Operative Error Diagnosis in GUI Automation, Neural Information Processing Systems (**NeurIPS**), 2025

Yuyang Wanyan, **Xiaoshan Yang**, Weiming Dong, Changsheng Xu: A Comprehensive Review of Few-Shot Action Recognition, International Journal of Computer Vision (**IJCV**), 2025

Han Jiang, **Xiaoshan Yang**, Chaofan Chen, and Changsheng Xu: SPDQ: Synergetic Prompts as Disentanglement Queries for Compositional Zero-Shot Learning, IEEE Transactions on Multimedia (**TMM**), 2025

Fang Peng, **Xiaoshan Yang**, Linhui Xiao, Yaowei Wang and Changsheng Xu, Progressive Learning of Instance-Level Proxy Semantics for Few-Shot Action Recognition, IEEE Transactions on Multimedia (**TMM**), 2025

Yiming Li, **Xiaoshan Yang**, Bing-Kun Bao, Changsheng Xu: Graph Prompts: Adapting Video Graph for Video Question Answering, International Joint Conference on Artificial Intelligence(**IJCAI**), 2025

Baochen Xiong, **Xiaoshan Yang**, Yaguang Song, Yaowei Wang, Changsheng Xu: Pilot: Building the Federated Multimodal Instruction Tuning Framework, AAAI Conference on Artificial Intelligence (**AAAI**), 2025

Chaofan Chen, **Xiaoshan Yang**, Changsheng Xu: Pseudo Informative Episode Construction for Few-Shot Class-Incremental Learning, AAAI Conference on Artificial Intelligence (**AAAI**), 2025

Baochen Xiong, **Xiaoshan Yang**, Yaguang Song, Yaowei Wang, Changsheng Xu: Modality-Collaborative Test-Time Adaptation for Action Recognition, IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2024

Yifan Xu, **Xiaoshan Yang**, Yaguang Song, Changsheng Xu: Libra: Building Decoupled Vision System on Large Language Models, International Conference on Machine Learning (**ICML**), 2024

Linhui Xiao, **Xiaoshan Yang**, Fang Peng, Yaowei Wang, Changsheng Xu: HiVG: Hierarchical Multimodal Fine-grained Modulation for Visual Grounding, ACM Multimedia (**MM**), 2024

Linhui Xiao, **Xiaoshan Yang**, Fang Peng, Yaowei Wang, Changsheng Xu: OneRef: Unified One-tower Expression Grounding and Segmentation with Mask Referring Modeling, Advances in Neural Information Processing Systems (**NeurIPS**), 2024

Xiaoshan Yang, Yi Huang, Changsheng Xu: Cross-Modal Federated Human Activity Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2024

Yifan Xu, Mengdan Zhang, **Xiaoshan Yang**, Changsheng Xu: Exploring Multi-Modal Contextual Knowledge for Open-Vocabulary Object Detection, IEEE Transactions on Image Processing (**TIP**), 2024

Han Jiang, Chaofan Chen, **Xiaoshan Yang**, and Changsheng Xu: Compact Latent Primitive Space Learning for Compositional Zero-Shot Learning, IEEE Transactions on Multimedia (**TMM**), 2024

Fan Qi, Huaiwen Zhang, **Xiaoshan Yang**, Changsheng Xu: A Versatile Multimodal Learning Framework For Zero-shot Emotion Recognition, IEEE Transactions on Circuits and Systems for Video Technology (**TCSVT**), 2024

Jiabo Ye, Junfeng Tian, Ming Yan, Haiyang Xu, Qinghao Ye, Yaya Shi, **Xiaoshan Yang**, Xuwu Wang, Ji Zhang, Liang He, Xin Lin: UniQRNet: Unifying Referring Expression Grounding and Segmentation with QRNet, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2024

Baochen Xiong, **Xiaoshan Yang**, Yaguang Song, Yaowei Wang, Changsheng Xu: Client-Adaptive Cross-Model Reconstruction Network for Modality-Incomplete Multimodal Federated Learning, ACM Multimedia (**MM**), 2023

Yiming Li, **Xiaoshan Yang**, Changsheng Xu: Iterative Learning with Extra and Inner Knowledge for Long-tail Dynamic Scene Graph Generation, ACM Multimedia (**MM**), 2023

Qinghao Ye, Haiyang Xu, Ming Yan, Chenlin Zhao, Junyang Wang, **Xiaoshan Yang**, Ji Zhang, Fei Huang, Jitao Sang, Changsheng Xu: mPLUG-Octopus: The Versatile Assistant Empowered by A Modularized End-to-End Multimodal LLM, ACM Multimedia (**MM**), 2023

Yifan Xu, Mengdan Zhang, Chaoyou Fu, Peixian Chen, **Xiaoshan Yang**, Ke Li, Changsheng Xu: Multi-modal queried object detection in the wild, Advances in Neural Information Processing Systems (**NeurIPS**), 2023

Linhui Xiao, **Xiaoshan Yang**, Fang Peng, Ming Yan, Yaowei Wang, Changsheng Xu: CLIP-VG: Self-paced Curriculum Adapting of CLIP for Visual Grounding, IEEE Transactions on Multimedia (**TMM**), 2023

Fang Peng, **Xiaoshan Yang**, Linhui Xiao, Yaowei Wang, Changsheng Xu: Sgva-clip: Semantic-guided visual adapting of vision-language models for few-shot image classification, IEEE Transactions on Multimedia (**TMM**), 2023

Hao Liu, **Xiaoshan Yang**, Changsheng Xu: Counterfactual scenario-relevant knowledge-enriched multi-modal emotion reasoning, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2023

Yagong Song, **Xiaoshan Yang**, Yaowei Wang, Changsheng Xu: Recovering generalization via pre-training-like knowledge distillation for out-of-distribution visual question answering, IEEE Transactions on Multimedia (**TMM**), 2023

Yaguang Song, **Xiaoshan Yang**, Changsheng Xu: Self-supervised calorie-aware heterogeneous graph networks for food recommendation, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2023

Yuyang Wanyan, **Xiaoshan Yang**, Chaofan Chen, Changsheng Xu: Active Exploration of Multimodal Complementarity for Few-Shot Action Recognition. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023

Chaofan Chen, **Xiaoshan Yang**, Jinpeng Zhang, Bo Dong, Changsheng Xu: Category Knowledge-guided Parameter Calibration for Few-shot Object Detection. IEEE Transactions on Image Processing (**TIP**), 2023

Xuan Ma, **Xiaoshan Yang**, Changsheng Xu: Multi-Source Knowledge Reasoning Graph Network for Multi-modal Commonsense Inference. ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2022

Yuyang Wanyan, **Xiaoshan Yang**, Xuan Ma, Changsheng Xu: Dual Scene Graph Convolutional Network for Motivation Prediction. ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2022

Yi Huang, **Xiaoshan Yang**, Ji Zhang, Changsheng Xu: Relative Alignment Network for Source-Free Multimodal Video Domain Adaptation. ACM Multimedia (**MM**), 2022

Chaofan Chen, **Xiaoshan Yang**, Ming Yan, Changsheng Xu: Attribute-guided Dynamic Routing Graph Network for Transductive Few-shot Learning. ACM Multimedia (**MM**), 2022

Fan Qi, Zixin Zhang, **Xiaoshan Yang**, Huaiwen Zhang, Changsheng Xu: Feeling Without Sharing: A Federated Video Emotion Recognition Framework Via Privacy-Agnostic Hybrid Aggregation. ACM Multimedia (**MM**), 2022

Yiming Li, **Xiaoshan Yang**, Changsheng Xu: Dynamic Scene Graph Generation via Anticipatory Pre-training. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022

Jiabo Ye, Junfeng Tian, Ming Yan, **Xiaoshan Yang**, Xuwu Wang, Ji Zhang, Liang He, Xin Lin: Shifting More Attention to Visual Backbone: Query-modulated Refinement Networks for End-to-End Visual Grounding. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022

Xiaoshan Yang, Baochen Xiong, Yi Huang, Changsheng Xu: Cross-Modal Federated Human Activity Recognition via Modality-Agnostic and Modality-Specific Representation Learning. AAAI Conference on Artificial Intelligence (**AAAI**), 2022, Oral

Xinhong Ma, **Xiaoshan Yang**, Junyu Gao, Changsheng Xu: The Model May Fit You: User-Generalized Cross-Modal Retrieval. IEEE Transactions on Multimedia (**TMM**), 2022

Yiming Li, **Xiaoshan Yang**, Xuhui Huang, Zhe Ma, and Changsheng Xu: Zero-shot Predicate Prediction for Scene Graph Parsing. IEEE Transactions on Multimedia (**TMM**), 2022

Yaguang Song, **Xiaoshan Yang**, Changsheng Xu: Self-supervised Calorie-aware Heterogeneous Graph Networks for Food Recommendation, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2022

Yi Huang, **Xiaoshan Yang**, Changsheng Xu: Multimodal Global Relation Knowledge Distillation for Egocentric Action Anticipation, ACM Multimedia (**MM**), 2021, Oral

Fan Qi, **Xiaoshan Yang**, Changsheng Xu: Zero-shot Video Emotion Recognition via Multimodal Protagonist-aware Transformer Network, ACM Multimedia (**MM**), 2021, Oral

Yaguang Song, Junyu Gao, **Xiaoshan Yang**, Changsheng Xu: Learning Hierarchical Video Graph Networks for One-Stop Video Delivery, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2021

Yi Huang, **Xiaoshan Yang**, Junyu Gao, Changsheng Xu: Holographic Feature Learning of Egocentric-Exocentric Videos for Multi-Domain Action Recognition, IEEE Transactions on Multimedia (**TMM**), 2021

Chaofan Chen, **Xiaoshan Yang**, Changsheng Xu, Xuhui Huang, Zhe Ma: ECKPN: Explicit Class Knowledge Propagation Network for Transductive Few-shot Learning, IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2021

Xuan Ma, **Xiaoshan Yang**, Junyu Gao, Changsheng Xu, Health Status Prediction with Local-Global Heterogeneous Behavior Graph, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2021

Yi Huang, **Xiaoshan Yang**, Junyu Gao, Jitao Sang, Changsheng Xu, Knowledge-driven Egocentric Multimodal Activity Recognition, ACM Transactions on Multimedia Computing, Communications, and Applications (**TOMM**), 2020

Fan Qi, **Xiaoshan Yang**, Changsheng Xu, Emotion Knowledge Driven Video Highlight Detection, IEEE Transactions on Multimedia (**TMM**), 2020

Junyu Gao, **Xiaoshan Yang**, Yingying Zhang, Changsheng Xu, Unsupervised Video Summarization via Relation-aware Assignment Learning, IEEE Transactions on Multimedia (**TMM**), 2020

Wei Wang, Junyu Gao, **Xiaoshan Yang**, Changsheng Xu: Learning Coarse-to-Fine Graph Neural Networks for Video-Text Retrieval, IEEE Transactions on Multimedia (**TMM**), 2020

Yingying Zhang, Junyu Gao, **Xiaoshan Yang**, Chang Liu, Yan Li, Changsheng Xu: Find Objects and Focus on Highlights: Mining Object Semantics for Video Highlight Detection via Graph Neural Networks, Proceedings of the AAAI Conference on Artificial Intelligence (**AAAI**), 2020

Xiaoshan Yang, Changsheng Xu: Image Captioning by Asking Questions. ACM Transactions on Multimedia Computing, Communications and Applications (**TOMM**), 2019

Junyu Gao, Tianzhu Zhang, **Xiaoshan Yang**, Changsheng Xu: P2T: Part-to-Target Tracking via Deep Regression Learning. IEEE Transactions on Image Processing (**TIP**), 2018

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu: Deep-Structured Event Modeling for User-Generated Photos. IEEE Transactions on Multimedia (**TMM**), 2018

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu: Text2Video: An End-to-end Learning Framework for Expressing Text With Videos. IEEE Transactions on Multimedia (**TMM**), 2018

Yifan Jiao, Zhetao Li, Shucheng Huang, **Xiaoshan Yang**, Bin Liu, Tianzhu Zhang: Three-Dimensional Attention-Based Deep Ranking Model for Video Highlight Detection. IEEE Transactions on Multimedia (**TMM**), 2018

Fan Qi, **Xiaoshan Yang**, Changsheng Xu: A Unified Framework for Multimodal Domain Adaptation. ACM Multimedia (**MM**), 2018

Junyu Gao, Tianzhu Zhang, **Xiaoshan Yang**, Changsheng Xu: Deep Relative Tracking. IEEE Transactions on Image Processing (**TIP**), 2017

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu, Shuicheng Yan, M. Shamim Hossain, Ahmed Ghoneim: Deep Relative Attributes. IEEE Transactions on Multimedia (**TMM**), 2016

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu: Semantic Feature Mining for Video Event Understanding. ACM Transactions on Multimedia Computing, Communications and Applications (**TOMM**), 2016

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu: Abnormal Event Discovery in User Generated Photos. ACM Multimedia (**MM**), 2016

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu: Cross-Domain Feature Learning in Multimedia. IEEE Transactions on Multimedia (**TMM**), 2015

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu, M. Shamim Hossain: Automatic Visual Concept Learning for Social Event Understanding. IEEE Transactions on Multimedia (**TMM**), 2015

Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu, Ming-Hsuan Yang: Boosted Multifeature Learning for Cross-Domain Transfer. ACM Transactions on Multimedia Computing, Communications and Applications (**TOMM**), 2015

Jianbing Shen, **Xiaoshan Yang**, Xuelong Li, Yunde Jia: Intrinsic Image Decomposition Using Optimization and User Scribbles. IEEE Transactions on Cybernetics (**TCYB**), 2013

Jianbing Shen, **Xiaoshan Yang**, Yunde Jia, Xuelong Li: Intrinsic images using optimization. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2011

Awards

2024 **Best Paper Award**, MMGR Workshop at ACM Multimedia Conference
2018 **Tencent Rhino-Bird Innovation Award (Best Patent Award)**, CCF and Tencent
2017 **Best Paper Award**, China Multimedia Conference
2017 **Outstanding Doctoral Dissertation Award**, Chinese Academy of Sciences
2016 **President's Award for Excellence**, Chinese Academy of Sciences
2016 **Rokid Enterprise Award**, Rokid Inc.

Research Grants

2026	**Research on Federated Learning Technologies for Training Multimodal Large Language Models**, National Natural Science Foundation of China (Key Program of Joint Funds)	*¥2,540,000*
2024	**Multimedia Content Understanding in Open Environments**, National Natural Science Foundation of China (Excellent Young Scientists Fund)	*¥2,000,000*
2022	**Unified Representation and Cross-domain Transfer for Multimodal Few-shot Data**, Key Project of the Science and Technology Commission	*¥5,400,000*
2021	**Cross-modal Explainable Individual Behavior Analysis and Health Management**, National Natural Science Foundation of China (General Program)	*¥580,000*
2018	**Knowledge-driven Video Caption Generation**, National Natural Science Foundation of China (Youth Program)	*¥290,000*
2018	**Cloud Platform for Fine-grained Video Tagging**, China Mobile Migu Co., Ltd.	*¥1,220,000*
2018	**Video Captioning and Highlight Detection**, CCF–Tencent Rhino-Bird Research Fund	*¥150,000*

Presentations

May 2025. *Open-world Multimodal Large Model Research and Applications*. Taiyuan University of Technology.

Dec. 2024. *Open-world Multimodal Large Model Research and Applications*. Beijing Institute of Petrochemical Technology.

Dec. 2024. *Open-world Multimodal Large Model Research and Applications*. Suzhou Artificial Intelligence Conference.

Oct. 2021. *Multimodal Few-shot and Zero-shot Learning*. Tutorial@ACM MM: Few-shot Learning for Multi-Modality Tasks.

Teaching Experience

Fall 2024 **Undergraduate Seminar**, Lead Instructor

Services & Memberships

- Reviewer of TPAMI, TIP, TMM, TNNLS, PR, IJCV, TKDE, TCSVT, CVPR, ICCV, NeurIPS, ICML, ICLR, MM, AAAI, IJCAI, etc.
- Area Chair of MM, IJCAI, ICPR, MM Asia, etc.
- Guest Editor of the Multimedia System Journal
- Publication Chair of the China Multimedia Conference (ChinaMM 2026)
- Workshop Chair of the Vision and Learning Seminar (VALSE 2025)
- Workshop Chair of the China Multimedia Conference (ChinaMM 2024)
- Member of the Institute of Electrical and Electronics Engineers (IEEE)
- Senior Member of the China Computer Federation (CCF)
- Member of the CCF Technical Committee on Multimedia
- Deputy Secretary-General of the Technical Committee on Multimedia, China Society of Image and Graphics (CSIG)
- Member of the CSIG Technical Committee on Image Intelligent Edge Computing