Multimedia Content Understanding in Harsh Environments

Zheng Wang
Wuhan University
wangzwhu@whu.edu.cn

Dan Xu
Hong Kong University of Science and Technology
danxu@cse.ust.hk

Zhedong Zheng
National University of Singapore
zdzheng@nus.edu.sg

Kui Jiang
Huawei Cloud & AI
kuijiang_1994@163.com

ABSTRACT

Multimedia content understanding methods often encounter a severe performance degradation under harsh environments. This tutorial covers several important components of multimedia content understanding in harsh environments. It introduces some multimedia enhancement methods, presents recent advances in 2D and 3D visual scene understanding, shows strategies to estimate the prediction uncertainty, provides a brief summary, and shows some typical applications.

ACM Reference Format:
Zheng Wang, Dan Xu, Zhedong Zheng, and Kui Jiang. 2022. Multimedia Content Understanding in Harsh Environments. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3503161.3546969

1 MOTIVATION

With the rapid development of multimedia technology, a large amount of multimedia data has rapidly emerged. The large amount of multimedia data facilitates new and innovative approaches, for example, multimedia content understanding. Multimedia content understanding is a key application for effective and efficient search, retrieval, delivery, management and sharing of multimedia content. Existing work shows that media understanding performs well in excellent environments, *i.e.*, in good light, good weather, and with sufficient training samples. Harsh environments (*e.g.*, fog, rain, snow, dark, low light, glare, blur, and low resolution) introduce challenges in visibility, analysis and understanding of visual data for real applications, such as autonomous cars and video surveillance systems. Despite the development of computing power and deep learning algorithms, the performance of current multimedia content understanding algorithms is still mainly benchmarked under high-quality environments (good weather, favorable lighting). Therefore, state-of-the-art methods often encounter a severe performance degradation under harsh environments. In this tutorial, we introduce some key directions in the field of multimedia content understanding under harsh environments. This tutorial would be useful for the multimedia community, especially for multimedia content understanding task for the practical and open-set domain.

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10.
https://doi.org/10.1145/3503161.3546969

2 TUTORIAL DESCRIPTION

This tutorial covers several important components of multimedia content understanding in harsh environments. First, we will introduce some multimedia enhancement methods, including image deraining, dehazing and low-light enhancement, and demonstrate their performances in down-stream vision tasks, such as object detection and segmentation [2–5, 20, 21]. Second, we will present recent advances on 2D and 3D visual scene understanding, and describe how deep learning and visual big data are significantly driving research and development in this domain [10–14, 17–19, 22]. Third, we will introduce strategies to estimate the prediction uncertainty during training to rectify the pseudo label learning for unsupervised semantic segmentation adaptation [23–26]. Finally, we will give a brief summary and show some typical applications and some trends in this task [1, 6–9, 15, 16]. List of topics covered in this tutorials is:

- Image Enhancement
- 2D and 3D Scene Understanding
- Domain Adaptation
- Understanding and Detection in Harsh Environments

Topic	Duration	Speaker
An opening of the tutorial	5 min	Zheng
Image enhancement: Disentanglement	40 min	Kui
2D and 3D Scene Understanding	40 min	Dan
Domain Adaptation: Consistency and Uncertainty	40 min	Zhedong
Understanding and Detection in Harsh Environments	40 min	Zheng

This tutorial is appropriate and timely for ACM MM, graduate students, researchers and industry practitioners working in the field of multimedia content retrieval, multimedia content analysis, as well as multimedia system. The course materials are mainly from recent publications in this area. The speakers will use slides to introduce their work. The materials will be publicly available after the tutorial.

3 PRESENTER INFORMATION

Zheng Wang (https://wangzwhu.github.io/home/) currently a professor at Wuhan University, China. He has been an assistant professor of the RIISE institute and Computer Vision and Media Lab at The University of Tokyo, Japan, and a JSPS Fellowship Researcher of the National Institute of Informatics, Japan, from 2017 to 2021. He received his B.E., M.S., and Ph.D. degrees from Wuhan University China. His current research interest includes multimedia

analysis and retrieval. He has served as a tutorial co-chair for the IEEE MIPR 2021 and co-organized the IEEE ICME 2020/2021 special session, and the ACM ICMR 2020 special session. He also organized the ACM MM 2020 Tutorial "Effective and Efficient: Toward Open-world Instance Re-identification" and the CVPR 2020 Tutorial "Image Retrieval in the Wild". He won the Best Paper Award at Pacific-Rim Conference on Multimedia 2014, and ACM Wuhan Doctoral Dissertation Award 2018.

Dan Xu (https://www.danxurgb.net/) is an Assistant Professor in the Department of Computer Science and Engineering (CSE), Hong Kong University of Science and Technology (HKUST). Before joining HKUST, he was a Postdoctoral researcher in the Visual Geometry Group (VGG) at the University of Oxford, working with Prof. Andrea Vedaldi and Prof. Andrew Zisserman. He received his Ph.D. in Computer Science from the University of Trento in 2018, under the supervision of Prof. Nicu Sebe. He was also a visiting Ph.D. student in the MMLab at the Chinese University of Hong Kong (CUHK) under the supervision of Prof. Xiaogang Wang. His research mainly focuses on computer vision, multimedia, and deep learning. Specifically, he is interested in multi-modal and structured representation learning, statistical modelling within deep learning, as well as their applications in 2D/3D scene understanding. He served as Senior Programme Committee (SPC) / Area Chair (AC) at multiple international conferences including AAAI, ACM Multimedia, and WACV. He received the Best Scientific Paper award at ICPR 2016 and a Best Paper Nominee at ACM Multimedia 2018.

Zhedong Zheng (https://zdzheng.xyz) is currently a postdoctoral researcher at NExT++, School of Computing, National University of Singapore. He received the Ph.D. degree from the University of Technology Sydney, Australia, in 2021 and the B.S. degree from Fudan University, China, in 2016. He was an intern at Nvidia Research (2018) and Baidu Research (2020). His research interests include robust learning for image retrieval, generative learning for data augmentation, and unsupervised domain adaptation. He served as reviewer at several top conferences, including CVPR, ICCV, ECCV, IJCAI, AAAI and ACM Multimedia.

Kui Jiang (https://github.com/kuijiang94/home/) received the Ph.D. and M.S. degree from the Wuhan University, China, in 2022 and 2019. He is currently the senior engineer of artificial intelligence with Huawei Cloud & AI. His research interests include image/video processing and computer vision. He served as reviewer at several top conferences, including CVPR, ICCV, ECCV, IJCAI, AAAI and ACM Multimedia.

4 RESOURCE AND MATERIALS

Materials are available at https://wangzwhu.github.io/home/acmmm2022_tutorial_mmhe.html.

ACKNOWLEDGMENTS

This tutorial and related research were supported by National Key R&D Project (2021YFC3320301) and National Natural Science Foundation of China (62171325).

REFERENCES

[1] Mengshun Hu, Kui Jiang, Liang Liao, Jing Xiao, Junjun Jiang, and Zheng Wang. 2022. Spatial-Temporal Space Hand-in-Hand: Spatial-Temporal Video Super-Resolution via Cycle-Projected Mutual Learning. In *CVPR*. 3574–3583.

[2] Kui Jiang, Zhongyuan Wang, Zheng Wang, Chen Chen, Peng Yi, Tao Lu, and Chia-Wen Lin. 2022. Degrade is upgrade: Learning degradation for low-light image enhancement. In *AAAI*. 1078–1086.

[3] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. 2020. Multi-scale progressive fusion network for single image deraining. In *CVPR*. 8346–8355.

[4] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zheng Wang, Xiao Wang, Junjun Jiang, and Chia-Wen Lin. 2021. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE TIP* 30 (2021), 7404–7418.

[5] Kui Jiang, Zhongyuan Wang, Peng Yi, and Junjun Jiang. 2020. Hierarchical dense recursive network for image super-resolution. *PR* 107 (2020), 107475.

[6] Liang Liao, Jing Xiao, Zheng Wang, Chia-Wen Lin, and Shin'ichi Satoh. 2020. Guidance and evaluation: Semantic-aware image inpainting for mixed scenes. In *ECCV*. 683–700.

[7] Liang Liao, Jing Xiao, Zheng Wang, Chia-Wen Lin, and Shin'ichi Satoh. 2021. Image inpainting guided by coherence priors of semantics and textures. In *CVPR*. 6539–6548.

[8] Yuting Liu, Zheng Wang, Miaojing Shi, Shin'ichi Satoh, Qijun Zhao, and Hongyu Yang. 2020. Towards unsupervised crowd counting via regression-detection bi-knowledge transfer. In *ACM MM*. 129–137.

[9] Xianzheng Ma, Zhixiang Wang, Yacheng Zhan, Yinqiang Zheng, Zheng Wang, Dengxin Dai, and Chia-Wen Lin. 2022. Both style and fog matter: Cumulative domain adaptation for semantic foggy scene understanding. In *CVPR*. 18922–18931.

[10] Zhenxing Mi, Chang Di, and Dan Xu. 2022. Generalized Binary Search Network for Highly-Efficient Multi-View Stereo. In *CVPR*. 12991–13000.

[11] Hao Tang, Dan Xu, Gaowen Liu, Wei Wang, Nicu Sebe, and Yan Yan. 2019. Cycle in cycle generative adversarial networks for keypoint-guided image generation. In *ACM MM*. 2052–2060.

[12] Hao Tang, Dan Xu, Yan Yan, Philip HS Torr, and Nicu Sebe. 2020. Local class-specific and global image-level generative adversarial networks for semantic-guided scene generation. In *CVPR*. 7870–7879.

[13] Jiapeng Tang, Jiabao Lei, Dan Xu, Feiying Ma, Kui Jia, and Lei Zhang. 2021. Sa-convonet: Sign-agnostic optimization of convolutional occupancy networks. In *ICCV*. 6504–6513.

[14] Jiapeng Tang, Dan Xu, Kui Jia, and Lei Zhang. 2021. Learning parallel dense correspondence from spatio-temporal descriptors for efficient and robust 4d reconstruction. In *CVPR*. 6022–6031.

[15] Xiao Wang, Zheng Wang, Wu Liu, Xin Xu, Jing Chen, and Chia-Wen Lin. 2021. Consistency-constancy bi-knowledge learning for pedestrian detection in night surveillance. In *ACM MM*. 4463–4471.

[16] Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin'ichi Satoh. 2019. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In *CVPR*. 618–626.

[17] Dan Xu, Wanli Ouyang, Xiaogang Wang, and Nicu Sebe. 2018. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In *CVPR*. 675–684.

[18] Dan Xu, Wei Wang, Hao Tang, Hong Liu, Nicu Sebe, and Elisa Ricci. 2018. Structured attention guided convolutional neural fields for monocular depth estimation. In *CVPR*. 3917–3925.

[19] Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, Ferdous Sohel, and Dan Xu. 2021. Leveraging auxiliary tasks with affinity learning for weakly supervised semantic segmentation. In *ICCV*. 6984–6993.

[20] Peng Yi, Zhongyuan Wang, Kui Jiang, Junjun Jiang, Tao Lu, and Jiayi Ma. 2020. A progressive fusion generative adversarial network for realistic and consistent video super-resolution. *IEEE TPAMI* (2020).

[21] Peng Yi, Zhongyuan Wang, Kui Jiang, Junjun Jiang, Tao Lu, Xin Tian, and Jiayi Ma. 2021. Omniscient video super-resolution. In *ICCV*. 4429–4438.

[22] Xuanmeng Zhang, Zhedong Zheng, Daiheng Gao, Bang Zhang, Pan Pan, and Yi Yang. 2022. Multi-View Consistent Generative Adversarial Networks for 3D-aware Image Synthesis. In *CVPR*. 18450–18459.

[23] Zhedong Zheng, Yunchao Wei, and Yi Yang. 2020. University-1652: A multi-view multi-source benchmark for drone-based geo-localization. In *ACM MM*. 1395–1403.

[24] Zhedong Zheng and Yi Yang. 2021. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *IJCV* 129, 4 (2021), 1106–1120.

[25] Zhedong Zheng and Yi Yang. 2021. Unsupervised scene adaptation with memory regularization in vivo. In *IJCAI*. 1076–1082.

[26] Zhedong Zheng, Liang Zheng, Michael Garrett, Yi Yang, Mingliang Xu, and Yi-Dong Shen. 2020. Dual-path convolutional image-text embeddings with instance loss. *ACM TOMM* 16, 2 (2020), 1–23.