**Weakly Supervised Learning for Image and Video Understanding**

**Aim & Purpose:**

With the goal of addressing fine-level image and video understanding tasks by learning from coarse-level human annotations, weakly supervised learning (WSL) is of particular importance in such a big data era as it can dramatically alleviate the human labor for annotating each of the structured visual/multimedia data and thus enables machines to learn from much larger-scaled data but with the equal annotation cost of the conventional fully supervised learning methods. More importantly, when dealing with the data from real-world application scenarios, fine-level manual annotations are very limited and difficult to obtain. Under these circumstances, the WSL-based learning frameworks would bring great benefits. Unfortunately, designing effective WSL systems is challenging and principled solutions are still under-studied. Nowadays, with the rapid development of advanced machine learning techniques, such as the Graph Convolutional Networks, Capsule Networks, Transformers, Generative Adversarial Networks, and Deep Reinforcement Learning models, new opportunities have emerged for solving the problems in WSL and applying WSL to richer vision and multimedia tasks.

This special issue aims at promoting cutting-edge research along this direction, including machine learning, computer vision and pattern recognition, and encourages novel theories and advanced techniques of weakly supervised learning for image and video understanding problems, and offers a timely collection of works to benefit researchers and practitioners. We welcome high-quality original submissions addressing both novel theoretical and practical aspects related to WSL, as well as the real-world applications based on WSL approaches.

Relevant topics include, but are not limited to:

* Multi-modality weakly supervised learning theory and framework;
* Multi-task weakly supervised learning theory and framework;
* Robust learning theory and framework;
* Audio-visual learning under weak supervision;
* Weakly supervised spatial/temporal feature learning;
* Self-supervised learning frameworks and applications;
* Graph Convolutional Networks/Graph Neural Networks-based weakly supervised learning frameworks;
* Deep Reinforcement Learning for weakly supervised learning;
* Emerging vision tasks with limited supervision.

**Session Chairs:**

1. Xiao-Yu Zhang, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China ([zhangxiaoyu@iie.ac.cn](mailto:zhangxiaoyu@iie.ac.cn))
2. Xiaobin Zhu, Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing, China ([zhuxiaobin@ustb.edu.cn](mailto:zhuxiaobin@ustb.edu.cn))
3. Changsheng Li, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China ([lcs@bit.edu.cn](mailto:lcs@bit.edu.cn))