Mengmeng Xu | Curriculum Vitae

G3741, KAUST, Thuwal, Kingdom of Saudi Arabia

📱 966-570-999-413   •   ✉ xu.frost@gmail.com   •   ⭘ frostinassiky

"Interdisciplinary Engineer with Entrepreneurial Mindset"

RESEARCH INTERESTS

I am focusing on problems that arise in image and video understanding. Particularly, I am interested in the representation of long untrimmed videos via self-supervised learning with novel model architectures. I am also interested in video localization tasks such as temporal action localization and language video grounding.

SKILLS

- Python/Matlab/C/C++
- PyTorch/TensorFlow

WORK EXPERIENCE

Samsung Electronics (UK) Ltd   **Cambridge, UK**
Research Intern in Samsung AI Center (SAIC)   *2020 Sep.- 2021 Mar.*

EDUCATION

King Abdullah University of Science and Technology   **Thuwal, KSA**
MS-PhD in Electrical and Computer Engineering, GPA: 3.88/4.00   *2017 - Present*

Zhejiang University   **China**
B.S. in Opto-Elctronics Information Science and Engineering, GPA 3.92/4.00   *2013 - 2017*

SELECTED PROJECTS

Low-Fidelity End-to-End Video Encoder Pre-training for TAL   *2021 Feb - Now*
End-to-end video encoder pre-training for temporal action localization is not operable subject to the GPU memory constraints, due to the prohibitive computational cost in processing long untrimmed videos.
- We reduce the mini-batch composition so that end-to-end optimization for the video encoder becomes operable.
- LoFi-TAL favourably solves the task discrepancy problem and providing more effective feature representations.
- LoFi-TAL with lightweight ResNet18 in a RGB stream surpasses RGB+optical-flow two-stream ResNet50 models.

Boundary-sensitive Pre-training for Temporal Localization in Videos   *2020 Sep - 2021 Mar*
Most existing models for temporal localization tasks are pre-trained on video classification tasks. The domain gap between action recognition and localization can be addressed by a temporal boundary datasets.
- For the first time, we investigate pre-training for localization by introducing a novel boundary-sensitive pretext task.
- We propose to synthesize temporal boundaries in existing video classification datasets to help localize action.
- Extensive experiments show that the proposed BSP is superior and complementary to the existing action classification based pre-training counterpart, and achieves new state-of-the-art performance on **several** temporal localization tasks.

Video-Language Graph Matching Network for Video Grounding   *2020 Jun - Nov*
The solution to grounding language queries in videos demands the understanding of videos' and queries' semantic content and the fine-grained reasoning about their multi-modal interactions.
- We adopted a graph matching layer for cross-modal context modeling and multi-modal fusion via graph convolutions.
- The method achieves SOTA performance on three large-scale video benchmarks for video grounding.

Precise Temporal Action Detection via Semantic Segmentation   *2020 Feb - Jun*
Temporal action detection (TAD) is an important yet challenging task in video analysis. Most existing works draw inspiration from image object detection. Besides, owing to the 1-dimensional property of TAD, we are able convert the coarse-grained detection annotations to fine-grained semantic segmentation annotations for free.
- We use fully-supervised semantic segmentation to regularize the detection network without extra annotation efforts.
- We design a multi-scale hybrid network architecture to model snippet-snippet and proposal-proposal correlations.
- We achieve competitive performance on the large-scale datasets for temporal action localization problem.

Sub-Graph Localization for Temporal Action Detection   *2019 Jan - Nov*

Recent studies show that context can be used as a clue to help understanding action before or after the snippet. Thus, we formulated action localization task to a sub-graph detection problem, solved by graph convolutional network.

- A multi-graph convolutional layer is designed to progressively represent the video snippet by its adaptive semantics.
- The method achieves SOTA performance on two large-scale video benchmarks for human activity localization.

Object Detection Using Multiple Level Annotations   *2018 Mar - 2019 Mar*

This thesis work studied hybrid learning methods on the object detection problems. We trained an object detector from a dataset where both instance-level and image-level labels are employed. Furthermore, we built a dataset with annotation budget constraints.

- The hybrid learning gives a trade-off between fewer annotations and a more accurate object detector.
- We can achieve the performance of a strongly supervised detector while saving 12.8% of its original annotation budget.

SELECTED PUBLICATIONS/PREPRINTS

Low-Fidelity E2E Video Encoder Pre-training for Temporal Action Localization   **2021**
Mengmeng Xu, ⋯, Xiatian Zhu, Bernard Ghanem, Brais Martinez   *arXiv Preprint*

Boundary-sensitive Pre-training for Temporal Localization in Videos   **2020**
Mengmeng Xu, ⋯, Xiatian Zhu, Li Zhang, Bernard Ghanem, Tao Xiang   *arXiv Preprint*

VLG-Net: Video-Language Graph Matching Network for Video Grounding   **2020**
Sisi Qu*, Mattia Soldan*, Mengmeng Xu*, Jesper Tegner, Bernard Ghanem   *arXiv Preprint*

G-TAD: Sub-Graph Localization for Temporal Action Detection   **2020**
Mengmeng Xu, Chen Zhao, David S. Rojas, Ali Thabet, Bernard Ghanem   *CVPR*

BAOD: Budget-Aware Object Detection   **2021**
Alejandro Pardo*, Mengmeng Xu*, ⋯, Bernard Ghanem   *CVPR Workshops,* **Oral**

Semantic Part RCNN for Real-World Pedestrian Detection   **2019**
Mengmeng Xu, Yancheng Bai, Sisi Qu, B Ghanem   *CVPR Workshops*

Missing Labels in Object Detection   **2019**
Mengmeng Xu, Yancheng Bai, B Ghanem   *CVPR Workshops*

Object Detection Using Multiple Level Annotations   **2019**
Mengmeng Xu   *Master Thesis, KAUST*

ACADEMIC/PROFESSIONAL ACTIVITIES

Academic Reviewer   **2019-present**
ICCV19, AAAI20, CVPR20, ECCV20, NeurIPS20, AAAI21, CVPR21

Program Chair   **2019-present**
ActivityNet Challenge workshop in CVPR19, CVPR20, CVPR21

Teaching Assistant   **2019-2020**
CS390DD: Special Topics in Machine Learning, and AMCS211: Numerical Optimization   *KAUST*

Certificates and Nano-degrees   **2016-2020**
NLP Nanodegree, Deep Learning Specialization, Machine Learning.   *Coursera and Udacity*

AWARDS/HONORS

- **Outstanding Reviewer**, CVPR 21   2021
- **Runner-up**, HACS Temporal Action Localization Challenge, *CVPR Workshop*   2020
- **First Silver Medal**, The 3rd YouTube-8M Video Understanding Challenge, *Kaggle Competition*   2019
- **Outstanding Graduates** of Zhejiang University   2017
- **Chu Kochen Honors Program** for Advanced Engineering Education, *ZJU.*   2014-2017
- **Gold Medal**, International Genetically Engineered Machine Competition (iGEM), *Boston, USA*   2016
- **Honorable Mention** of American Mathematical Contest in Modeling, *USA*   2016
- **National Scholarship**, *Zhejiang University*   2014
- **First prizes** of Olympic Competition of Physics and Mathematics, *Henan, China*   2012