

Huijuan Xu

Boston University
Department of Computer Science
111 Cummington Mall
Boston, MA 02215

hxu@bu.edu
<http://cs-people.bu.edu/hxu/>
Skype: hui.juan.

Research Interests

Computer vision and natural language processing, video activity detection

Education

Boston University Aug 2018

Ph.D., Computer Science

Advisor: Prof. Kate Saenko

Dissertation title: Vision and language understanding with localized evidence

Graduate University of Chinese Academy of Sciences Jul 2012

M.S., Computer Science

Advisor: Prof. Hua Yu

Thesis title: Bayesian network and probability inference

Hefei University of Technology Jul 2009

B.S., Computer Science

Recipient of department scholarship for three consecutive years.

Research

Project: Natural Language Localization in Videos Nov 2017 - present

Department of Computer Science, Boston University

- Propose the proposal and retrieval pipeline for solving the natural language localization task in videos;
- Explore various multi-modal feature fusion methods for building multi-modal retrieval model;
- Incorporate captioning training to benefit the multi-modal retrieval model.

Project: Dense Video Captioning Jun 2017 - Nov 2017

Disney Research Pittsburgh (research internship) and Boston University

- Design an end-to-end Joint Event Detection and Description Network (JEDDi-Net) for the dense video captioning task;
- Our JEDDi-Net contains a hierarchical captioning module with visual and language context modeling.

Project: Temporal Activity Detection Sep 2016 - present

Department of Computer Science, Boston University

- Propose Region Convolutional 3D Network (R-C3D) for temporal activity detection;
- Apply the image object detection pipelines (Faster RCNN) to solve the temporal activity detection in videos, use C3D as video feature extractor, design the RPN and ROI pooling for video data;
- Extend R-C3D model with optical flow information and hard negative mining;
- Extend R-C3D model with scale and context design.
- Our R-C3D model won the Most Innovative Award in ActivityNet Challenge 2017.
- Deploy R-C3D model for temporal activity detection of public surveillance videos in Deep Intermodal Video Analytics (DIVA) project.

Project: Visual Question Answering (VQA) Jun 2015 - Sep 2016

Department of Computer Science, University of Massachusetts-Lowell

- Design one deep neural network model called SMem-VQA based on spatial attention to solve the VQA problem, show the inference process of SMem-VQA model by visualizing the attention weights, and achieve good result on VQA challenge.
- Propose Dual Attention Network (DAN) to investigate the idea of attention on both question and image for the VQA problem, and get improved result on VQA challenge.

Project: Natural Language Generation for Youtube Videos Jun 2014 - Jun 2015

Department of Computer Science, University of Massachusetts-Lowell

- Run LSDA object detection (Large Scale Detection Through Adaptation detection algorithm) on youtube video frames, extract LSDA features for detected bounding boxes, and form video descriptor using the LSDA features of all the bounding boxes in one video. Based on video descriptor, use SVM to classify subject, verb and object, and generate sentence using template methods.
- Design a multi-scale fully convolutional architecture, use multiple instance learning mechanism (MIL) to fine-tune the network on youtube data, and integrate this multi-scale framework with LSTM to realize end-to-end video description generation.

Project: Bioterrorism Emergency Decision Support Program Sep 2009 - Jul 2012

Graduate University of Chinese Academy of Sciences

- Supported by National Natural Science Foundation of P. R. China, mainly studies the key technologies for dealing with biochemical terrorism attack emergency, including poison identification in biochemical terrorism attacks, hospital rescue capacity analysis, the wounded transportation, and evacuation simulation. I am Responsible for poison identification, namely using the preliminary symptoms to diagnose poisoning person.
- Investigate several Bayesian network structure learning algorithms and data processing methods, do comparison experiments, improve algorithms, and apply the algorithms into practical problems. Gibbs sampling, K2/Hill-Climbing, Independent Component Analysis(ICA), Markov Equivalent Class space, Greedy Equivalent Search algorithm(GES), etc.

Teaching

Department of Computer Science, University of Massachusetts-Lowell

Teaching Assistant, Machine Learning , Spring 2015

Lab Instructor, Computing I Lab, Spring 2014

Lab Instructor, Computing I Lab, Fall 2013

Honors & Awards

Most Innovative Award in ActivityNet Challenge 2017 at CVPR2017

Honolulu, USA, 2017

Outstanding Graduate of Anhui Province

Anhui, P. R. China, 2009

National Scholarship of P. R. China

P. R. China, 2007

Professional Activities

Reviewer of International Journal of Computer Vision (IJCV), ACM Transactions on Multimedia Computing Communications and Applications, ICCV2017, CVPR2018, WiCV2018, ACCV2018

Journal Publications

- [1]. **Huijuan Xu**, Abir Das, and Kate Saenko. Two-Stream Region Convolutional 3D Network for Temporal Activity Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2018. (submitted)

Conference Publications

- [1]. **Huijuan Xu**, Kun He, Leonid Sigal, Stan Sclaroff and Kate Saenko. Text-to-Clip Video Retrieval with Early Fusion and Re-Captioning. arXiv preprint arXiv:1804.05113. (submitted)
- [2]. **Huijuan Xu**, Boyang Li, Vasili Ramanishka, Leonid Sigal and Kate Saenko. Joint Event Detection and Description in Continuous Video Streams. arXiv preprint arXiv:1802.10250. (submitted)
- [3]. Bai, Yancheng, **Huijuan Xu**, Kate Saenko and Bernard Ghanem. Contextual Multi-Scale Region Convolutional 3D Network for Activity Detection. arXiv preprint arXiv:1801.09184. (submitted)
- [4]. **Huijuan Xu**, Abir Das, and Kate Saenko. R-C3D: Region Convolutional 3D Network for Temporal Activity Detection. *International Conference on Computer Vision (ICCV)*, 2017.
- [5]. **Huijuan Xu** and Kate Saenko. Ask, Attend and Answer: Exploring Question-Guided Spatial Attention for Visual Question Answering. *European Conference on Computer Vision (ECCV)*, 2016.
- [6]. Subhashini Venugopalan, **Huijuan Xu**, Jeff Donahue, Marcus Rohrbach, Raymond Mooney and Kate Saenko. Translating Videos to Natural Language Using Deep Recurrent Neural Networks. *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2015.
- [7]. Jinke Jiang, Juyun Wang, Hua Yu, **Huijuan Xu**. Poison Identification Based on Bayesian Network: A Novel Improvement on K2 Algorithm via Markov Blanket. *International Conference of Advances in Swarm Intelligence (ICSI)*, 2013.
- [8]. **Huijuan Xu**, Hua Yu, Juyun Wang and Jinke Jiang. Exploration of Greedy Hill-climbing Search in Markov Equivalent Class Space. *International Conference on Artificial Intelligence (ICAI)*, 2012.
- [9]. **Huijuan Xu**, Hua Yu and Juyun Wang. Poison Identification Based on Bayesian Method in Biochemical Terrorism Attacks. *Advanced Science Letters (SCIE, ISSN: 1936-6612)*, Vol. 5, 1-5, 2012.

Workshop Publications

- [1]. **Huijuan Xu** and Kate Saenko. Dual Attention Network for Visual Question Answering. *ECCV2016 2nd Workshop on Storytelling with Images and Videos (VisStory)*, 2016.
- [2]. **Huijuan Xu** and Kate Saenko. Ask, Attend and Answer: Exploring Question-Guided Spatial Attention for Visual Question Answering. *VQA Challenge Workshop at CVPR2016*.
- [3]. **Huijuan Xu**, Subhashini Venugopalan, Vasili Ramanishka, Marcus Rohrbach and Kate Saenko. A Multi-scale Multiple Instance Video Description Network. *ICCV15 workshop on Closing the Loop Between Vision and Language*, 2015.