

Education

Purdue University

Ph.D., Artificial Intelligence

adviser: Jeffrey M. Siskind

research areas: Computer Vision and Natural Language Processing

WEST LAFAYETTE, IN, USA

Aug '11 – May '16

Peking University

B.S., Computer Science

focus: Image Processing, Computer Vision, and Multimedia

BEIJING, CHINA

Sep '07 – Jun '11

Work Experience

Baidu Research, Institute of Deep Learning (<http://research.baidu.com/>)

SUNNYVALE, CA, USA

Research Scientist

Jun '16 – present

- *Deep learning and reinforcement learning.* We investigate the problem of how a robot can interact with a teacher (either a program or a real human being) through conversation, and learn linguistic concepts in a simulated environment, similar to the case in which a child seeks help and/or follows instructions from its parents when exploring and learning concepts of the world.

CCCP, Purdue (<http://upplysingaoflun.ecn.purdue.edu/~qobi/cccp/>)

WEST LAFAYETTE, IN, USA

Research Assistant

Aug '11 – May '16

- *Grounded language learning from video paired with sentences.* We successfully learned word meanings that are comparable to human knowledge in a weakly supervised fashion of video-sentence pairs, on a dataset of 61 videos and 15 lexical entries over 6 parts of speech. This work won the best paper award at ACL2013.
- *Grounding natural language in robotic path planning and self description.* We learned word meanings from robot navigation paths paired with human instructions. The learned words were used to generate description given a new navigation path. The generated sentences achieved a correctness of 94.6% and a completeness of 85.6% on a dataset of 10 floorplans.
- *Sentence-directed video object co-detection.* We used sentences to provide semantic constraints for the video object codetection problem. A significant improvement (23.54% in the IoU score) was obtained over methods without sentence semantics, on a dataset of 150 video-sentence pairs.

Baidu Research, Institute of Deep Learning (<http://research.baidu.com/>)

SUNNYVALE, CA, USA

Research Intern

May '15 – Aug '15

- *Video paragraph captioning with hierarchical RNN.* We applied hierarchical Recurrent Neural Network to model the dependency among sentences in a paragraph. The learned dependency can be exploited to generate a paragraph for a long video. Our approach is the new state of the art on two large-scale benchmark datasets: YouTube2Text and TACoS-MultiLevel. We achieved improvements of 10.15% and 4.20% in BLEU-4 scores on the two datasets, respectively.

NELVT, Peking University (<http://idm.pku.edu.cn/>)

BEIJING, CHINA

Research Intern (part-time)

May '09 – May '11

- *Object segmentation with complementary saliency maps.* Our interaction-free method achieved an F1 score of 0.89 on a segmentation dataset containing 1,000 images.
 - *User-targeted video advertisement system.* Implemented a video-streaming system that enabled the user to select regions and segment objects in the video and retrieve their information from Amazon.
-

Selected Publications

Barrett, D.P., Bronikowski, S.A., Yu, H., and Siskind, J.M.

Driving Under the Influence (of Language)

IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2017.

<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/tnnls2017.pdf>

Barrett, D.P., Xu, R., Yu, H., and Siskind, J.M.

Collecting and Annotating the Large Continuous Action Dataset

Machine Vision and Applications (MVA), 2016.

<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/mva2016.pdf>

Yu, H., Wang, J., Huang, Z., Yang, Y., and Xu, W.
Video Paragraph Captioning using Hierarchical Recurrent Neural Networks
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 (oral).
<http://arxiv.org/pdf/1510.07712v2.pdf>

Yu, H., Siddharth, N., Barbu, A., and Siskind, J.M.
A Compositional Framework for Grounding Language Inference, Generation, and Acquisition in Video
Journal of Artificial Intelligence Research (JAIR), 2015 (accepted in Award-Winning Paper Track).
<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/jair2015.pdf>

Yu, H. and Siskind, J.M.
Learning to Describe Video with Weak Supervision by Exploiting Negative Sentential Information
AAAI Conference on Artificial Intelligence (AAAI), 2015 (oral).
<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/aaai2015.pdf>

Yu, H. and Siskind, J.M.
Grounded Language Learning from Video Described with Sentences
Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL), pp. 53–63, 2013 (oral, **Best Paper Award**).
<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/acl2013.pdf>

Cao, Y., Barrett, D.P., Barbu, A., Siddharth, N., Yu, H., Michaux, A., Lin, Y., Dickinson, S., Siskind, J.M., and Wang, S.
Recognizing Human Activities from Partially Observed Videos
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2658–2665, 2013.
<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/cvpr2013.pdf>

Yu, H., Li, J., Tian, Y., and Huang, T.
Automatic interesting object extraction from images using complementary saliency maps
ACM Multimedia (ACMMM), pp. 891–894, 2010.
<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/acmmm2010.pdf>

Patents Awarded

Siskind, J.M., Barbu, A., Siddharth, N., and Yu, H.

Correlating Videos and Sentences

US patent 9183466, 10 November 2015.

<https://patents.google.com/patent/US9183466B2/en>

Li, J., Gao, Y., Yu, H., Zhang, J., Tian, Y., and Yan, J.

Method and System for Personalized Advertisement Push Based on User Interest Learning

US patent 8750602, 10 June 2014.

<https://patents.google.com/patent/US8750602B2/en>

Tian, Y., Yu, H., Li, J., Gao, Y., Zhang, J., and Yan, J.

Method and system for extraction and association of object of interest in video

US patent 20130101209, 25 April 2013.

<https://patents.google.com/patent/US20130101209A1/en>

Honors

Best Paper Award, the Annual Meeting of the Association for Computational Linguistics 2013

Yang Fuqing & Wang Yangyuan Academician Scholarship 2010

Silver Medal, the 33rd ACM International Collegiate Programming Contest, Asia Regional 2008

Teaching

Teaching Assistant (informal)

Instructor: Jeffrey Mark Siskind

Artificial Intelligence ECE47300/ECE57000, spring/fall, 2013-2015.

Professional Activities

Reviewer, *ACM Multimedia*, 2017

Reviewer, *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 2016, 2017

Reviewer, *Artificial Intelligence Journal (AIJ)*, 2016

Reviewer, *Journal of Artificial Intelligence Research (JAIR)*, 2016

Reviewer, *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, 2016

Reviewer, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013

Skills

Technical specialties: Natural Language Processing, Deep Learning, and Reinforcement Learning

Software systems: Linux, Emacs, Git, L^AT_EX, CMake, PaddlePaddle

Programming languages (ranked by proficiency): C/C++, Scheme, Python, Bash, HTML, and Javascript

Spoken languages: Mandarin (*mother tongue*) and English (*professional proficiency*)