

Haonan Yu

+1 (765) 400-0678
haonanu@gmail.com
<http://haonanyu.com>
<https://github.com/you239>
June 06, 2018

My research focus is learning to ground language in vision and control, with deep reinforcement learning as the primary tool, under weak or no supervision. My previous work (during PhD) was learning to ground language in realistic video data, but with much more expensive supervision. I am also generally interested in writing machine learning code.

Research interests: Language grounding in vision and control, Deep learning, Reinforcement learning
Open-source projects: PaddlePaddle, XWorld, PARL
Programming languages: C/C++, Python, Bash, Scheme, HTML, and Javascript
Softwares: Linux, Emacs, Git, L^AT_EX, CMake, PaddlePaddle

Education

Purdue University WEST LAFAYETTE, IN, USA
Ph.D., Artificial Intelligence Aug '11 – May '16
Adviser: Jeffrey M. Siskind
Research areas: computer vision and natural language processing

Peking University BEIJING, CHINA
B.S., Computer Science Sep '07 – Jun '11

Work Experience

Baidu Research SUNNYVALE, CA, USA
Research Scientist Jun '16 – present

Grounding language in vision and control with deep reinforcement learning.

We investigate the problem of how a virtual agent can learn to ground language in perception and control, via interacting with a virtual teacher through conversation and rewards.

- Grounded zero-shot instruction execution. [C6]
- Interactive one-shot visual concept learning. [C7]
- A neural language grounding module for embodied agents. [C8]

Purdue University WEST LAFAYETTE, IN, USA
Research Assistant Aug '11 – May '16

Grounding language in video with supervised learning. We study the problem of learning to ground language in video from annotated video-sentence pairs.

- Grounded language learning from video paired with sentences. [C2,C3,J0]
- Grounding natural language in robotic path planning and control. [J2]
- Sentence-directed video object co-detection. [J3]

Baidu Research SUNNYVALE, CA, USA
Research Intern May '15 – Aug '15

- Video paragraph captioning with hierarchical RNNs. [C4]

Peking University BEIJING, CHINA
Research Intern May '09 – May '11

- Object segmentation with complementary saliency maps. [C0]
-

Publications

- [C8]
Yu, H., Lian, X., Zhang, H., and Xu, W.
Guided Feature Transformation (GFT): A Neural Language Grounding Module for Embodied Agents
arXiv preprint, 2018.
<https://arxiv.org/abs/1805.08329.pdf>
- [C7]
Zhang, H., Yu, H., and Xu, W.
Interactive Language Acquisition with One-shot Visual Concept Learning Through a Conversational Game
Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2018.
<https://arxiv.org/pdf/1805.00462.pdf>
- [C6]
Yu, H., Zhang, H., and Xu, W.
Interactive Grounded Language Acquisition and Generalization in a 2D World
International Conference on Learning Representations (ICLR), 2018.
<https://openreview.net/forum?id=H1U0m4gA->
- [C5]
Zhang, H., Yu, H., and Xu, W.
Listen, Interact and Talk: Learning to Speak via Interaction
NIPS workshop (Visually grounded interaction and language), 2017.
http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/Speak_interaction.pdf
- [J3]
Yu, H. and Siskind, J.M.
Sentence Directed Video Object Codiscovery
International Journal of Computer Vision (IJCV), 2017.
<http://upplysingaoflun.ecn.purdue.edu/~yu239/papers/ijcv2017.pdf>
- [J2]
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>
- [J1]
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>
- [C4]
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>
- [J0]
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>

[C3]

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>

[C2]

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>

[C1]

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>

[C0]

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>

Reviewing Activities

ACM Multimedia (ACMMM)	2017, 2018
IEEE Transactions on Neural Networks and Learning Systems (TNNLS)	2016, 2017, 2018
Artificial Intelligence Journal (AIJ)	2016, 2018
Journal of Artificial Intelligence Research (JAIR)	2015, 2016, 2018
ACM Transactions on Intelligent Systems and Technology (TIST)	2018
IEEE International Conference on Robotics and Automation (ICRA)	2015, 2016
IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2013

References

References available upon request.