Shugao Ma

EDUCATION

Boston University, Boston, MA, USA
Ph.D, Computer Science, May 2016

Chinese Academy of Sciences, Beijing, China
Master of Engineering, Computer Applied Technology, July 2009

Fudan University, Shanghai, China
Bachelor of Engineering , Software Engineering, July 2006

ACADEMIC EXPERIENCE

Boston University, Boston, MA, USA
Research Assistant September, 2010 - May, 2016
The focus of the research is on automatic human action / activity detection and recognition in realistic videos. Two major approaches are proposed: the first approach learns discriminative space-time structures of human actions in a data driven manner, and recognize and localize the actions by detecting these structures in videos; the other approach trains deep learning model that learns spatial visual features of human activities using CNN (Convolutional Neural Network) and, on top of the learned CNN features, learns temporal activity progression using LSTM (Long Short Term Memory), a specific type of RNN (Recurrent Neural Network). The trained deep model shows promising activity detection and early detection performances on large scale video dataset ActivityNet. Other research projects include visual attribute learning, object tracking and salient object subitizing.

Chinese Academy of Sciences, Beijing, China
Research Assistant September, 2006 - July, 2009
The theme of the research is in exploring local invariant visual features in image and video content analytic tasks including image scene classification, camera movement estimation and fighting scene spotting in action movies.

INDUSTRIAL RESEARCH EXPERIENCE

Facebook, Pittsburgh, PA, USA
Research Scientist June, 2016 - present
Researching on computer vision and machine learning technologies for continuously improving social experience in Virtual Reality.

Disney Research, Pittsburgh, PA, USA
Research & Development Lab Associate June, 2015 - Sep., 2015
Designed and implemented a deep learning model for activity detection that combines the Convolutional Neural Network with the Long Short Term Memory (*i.e.* a specific type of Recurrent Neural Network). Our method achieved very promising activity detection and early detection performances on a large scale dataset ActivityNet. This work was presented in CVPR 2016.

Disney Research, Pittsburgh, PA, USA
Research & Development Lab Associate February, 2014 - June, 2014
Proposed a new effective approach for automatic human action and interaction recognition in sports videos and TV program videos. In this approach, spatial, temporal and hierarchical structures in human actions are learned from video data and utilized to construct action classifiers. This work is published in CVPR 2015 for oral presentation.

Google, Mountain View, CA, USA
Intern May, 2013 - August, 2013

Designed and implemented a system for predicting users video advertising intention on YouTube using large scale machine learning methods. The main challenge addressed is finding relevant data for the prediction task by analyzing large amount of YouTube data, as well as constructing an efficient large scale machine learning system that achieves good prediction accuracy.

Microsoft, Beijing, China
Intern May, 2007 - August, 2007
Designed and implemented an automatic video perception quality assessment system. The main challenge is to study what visual and acoustic features are most relevant to the perception quality of videos. Based on these features, a effective machine learning system is designed and implemented to automatically score the videos.

SAP Labs, Shanghai, China
Intern April, 2006 - July, 2006
Designed and implemented a prototype of Assumption based Truth Maintenance System (ATMS). ATMS is an artificial intelligent system that, given a set of potentially conflicting rules, automatically finds the smallest changes to the rules so that all conflicts are removed.

REFEREED PUBLICATIONS

Shugao Ma, Jianming Zhang, Nazli Ikizler-Cinbis , Leonid Sigal, Stan Sclaroff. *Space-time tree ensemble for action recognition.* International Journal of Computer Vision, 2017. doi:10.1007/s11263-016-0980-8.

Jianming Zhang, Shugao Ma, Mehrnoosh Sameki, Stan Sclaroff, Margrit Betke, Zhe L. Lin, Xiaohui Shen, Brian L. Price, Radomr Mech. *Salient object subitizing.* To appear in International Journal of Computer Vision, 2017.

Shugao Ma, Sarah Adel Bargal, Jianming Zhang, Leonid Sigal, Stan Sclaroff. *Do less and achieve more: training CNNs for action recognition utilizing action images from the web.* Pattern Recognition, February 2017, ISSN 0031-3203.

Shugao Ma, Leonid Sigal, Stan Sclaroff. *Learning activity progression in lstms for activity detection and early detection.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. June 26, 2016. Pages 1942-1950.

Shugao Ma, Leonid Sigal, Stan Sclaroff. *Space-time tree ensemble for action recognition.* In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. June 7, 2015. Pages 5024-5032.

Jianming Zhang, Shugao Ma, Mehrnoosh Sameki, Stan Sclaroff, Margrit Betke, Zhe L. Lin, Xiaohui Shen, Brian L. Price, Radomr Mech. *Salient object subitizing.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. June 7, 2015. Pages 4045-4054.

Jianming Zhang, Shugao Ma, Stan Sclaroff. *MEEM: robust tracking via multiple experts using entropy minimization.* In Proceedings of the European Conference on Computer Vision. September 6, 2014. Pages 188-203.

Svebor Karaman, Lorenzo Seidenari, Shugao Ma, Alberto Del Bimbo, Stan Sclaroff. *Adaptive structured pooling for action recognition.* In the British Machine Vision Conference. September 1, 2014.

Shugao Ma, Jianming Zhang, Nazli Ikizler-Cinbis, Stan Sclaroff. *Action recognition and localization by hierarchical space-time segments.* In Proceedings of the IEEE International Conference on Computer Vision. December 1, 2013. Pages 2744-2751.

Shugao Ma, Stan Sclaroff, Nazli Ikizler-Cinbis. *Unsupervised learning of discriminative relative visual attributes.* In Proceedings Part III of the European Conference on Computer Vision 2012, Workshops and Demonstrations. October 7, 2012. Pages 61-70.

Shugao Ma, Wei-Qiang Wang. *Effectively discriminating fighting shots in action movies.* Journal of Computer Science and Technology. January 1, 2011. Volume 26. Pages 187-194.

Shugao Ma, Weiqiang Wang. *Effective camera motion analysis approach.* In Proceedings of the IEEE International Conference on Networking, Sensing and Control. April 10, 2010. Pages 111-116.

Shugao Ma, Weiqiang Wang, Qingming Huang, Shuqiang Jiang, Wen Gao. *Effective scene matching with local feature representatives.* In Proceedings of the IEEE 19th International Conference on Pattern Recognition. December 8, 2008. Pages 1-4.

US PATENT

Systems and methods for identifying activities in media contents based on prediction confidences. *Application submitted on July 13, 2016.*

ACADEMIC PRESENTATION

Carnegie Mellon University Computer Vision Club Talk Nov 3, 2016

When computer vision met with virtual reality. *invited speaker*

Carnegie Mellon University Computer Vision Course 16-720 Nov 9, 2016

Recurrent neural network and generative adversarial network. *guest speaker*

New England Machine Learning Day May 6, 2016

Learning activity progression in lstms for activity detection and early detection. *poster presentation*

Conference on Computer Vision and Pattern Recognition June 28, 2016

Learning activity progression in lstms for activity detection and early detection. *poster presentation*

Conference on Computer Vision and Pattern Recognition June 10, 2016

Space-time tree ensemble for action recognition. *oral presentation*

Conference on Computer Vision and Pattern Recognition June 9, 2016

Salient object subitizing. *poster presentation*

International Conference on Computer Vision Dec. 5, 2013

Action recognition and localization by hierarchical space-time segments. *poster presentation*

Second International Workshop on Parts and Attributes July 9, 2012

(in conjunction with ECCV 2012)
Unsupervised learning of discriminative relative visual attributes. *oral presentation*

MEDIA COVERAGE

Eurekalert June, 2016

http://www.eurekalert.org/pub_releases/2016-06/dr-dmd062316.php

Phys.org June, 2016

http://phys.org/news/2016-06-method-human-videos-earlier-accurately.html

Science Magazine June 2016

http://scienmag.com/disney-method-detects-human-activity-in-videos-earlier-and-more-accurately/

Eurekalert June 2015

http://www.eurekalert.org/pub_releases/2015-06/dr-dri060815.php

TechTalks.tv June, 2015

http://techtalks.tv/talks/space-time-tree-ensemble-for-action-recognition/61620/

Communications of the ACM June, 2015

http://cacm.acm.org/news/188152-researchers-improve-automated-recognition-of-human-body-movements-in-videos/fulltext

Phys.org June 2015

http://phys.org/news/2015-06-automated-recognition-human-body-movements.html

JOURNAL & Conference PEER-REVIEW

IEEE Transaction on Pattern Analysis and Machine Intelligence: 2016

Computer Vision and Image Understanding Journal: 2016, 2015, 2014

IEEE Transactions on Multimedia: 2016

Image and Vision Computing Journal: 2016

Neurocomputing Journal: 2016, 2015

Machine Vision and Application: 2016

IEEE Signal Processing Letters: 2016

Imaging Science Journal: 2016, 2015

IET Computer Vision: 2017, 2016, 2015, 2014

European Conference on Computer Vision: 2016

IEEE Conference on Computer Vision and Pattern Recognition: 2012, 2013, 2016, 2017

IEEE International Conference on Computer Vision: 2011

Asian Conference on Computer Vision: 2012

International Joint Conference on Artificial Intelligence: 2016

IEEE International Conference on Automatic Face and Gesture Recognition: 2017

TEACHING EXPERIENCE

Boston University September, 2014 - December, 2014
Introduction to Computers, *Teaching Fellow*

Boston University September, 2010 - December, 2010
Combinatoric Structures, *Teaching Fellow*

University of Chinese Academy of Sciences March, 2009 - July, 2009
Advanced Technologies in Human/Computer Interface, *Teaching Assistant*

University of Chinese Academy of Sciences September, 2008 - January, 2009
Data Mining, *Teaching Assistant*

TECHNIQUE EXPERTIES

Computer Vision
Familiar with many computer vision techniques, particularly experienced in human action recognition and detection, automatic video content analysis, visual attribute learning, object tracking.

Machine Learning
Familiar with many machine learning techniques, experienced in Support Vector Machines; Deep Learning techniques such as Convolutional Neural Network and Long Short Term Memory with over a year of experiences in Caffe and Torch.

Programming
Working experiences in C/C++, Java, C#, Python, Matlab and Lua.