

Caiming Xiong

Contact Information	Salesforce Inc. 172 University Ave, Palo Alto, CA, 94301, USA	cxiong@salesforce.com http://www.stat.ucla.edu/~caiming
Research Interest	Deep Learning, Computer Vision, Natural Language Processing, Speech, Recommendation, Reinforcement Learning.	
Education	The State University of New York at Buffalo Ph.D., Computer Science and Engineering, 2014, Huazhong University of Science and Technology, China M.S., Computer Science and Technology, 2007 B.S., Computer Science and Technology, 2005	
Experience	Salesforce Director of Research, AI	08/2017 to present
	Salesforce Lead Research Scientist	04/2016 to 08/2017
	<ul style="list-style-type: none">• Develop the long term vision for DL/ML in computer vision, speech and natural language processing.• Propose and implement novel AI models, and publish the work in top-tier AI conference.• Lead the several teams and build AI systems such as visual system, chatbot, speech recognition and recommendation system.	
	Metamind (<i>acquired by Salesforce Inc.</i>) Senior Researcher	09/2015 to 04/2016
	<ul style="list-style-type: none">• Developed multimodal deep learning for image/text question-answering, caption ranking• Developed new deep learning model for satellite image segmentation and object detection• Designed and implemented 3D Convolutional Neural Network for medical image diagnosis	
	University of California, Los Angeles(UCLA) Postdoctoral Researcher	06/2014 to 08/2015
	<ul style="list-style-type: none">• Video understanding and interactive robot learning• Joint human pose estimation and action recognition in video• Deep learning for visual question-answering	
	Honeywell ACS labs Research Intern	06/2012 to 08/2012
	<ul style="list-style-type: none">• Research in latent topic discovery for domain adaptation	

Publications

- (Peer Reviewed) L. Zhou, Y. Zhou, J. J. Corso, R. Socher, Caiming Xiong*. End-to-End Dense Video Captioning with Masked Transformer. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*. (* corresponding author)
- S. Min, V. Zhong, R. Socher, C. Xiong*. Efficient and Robust Question Answering from Minimal Context over Documents. *Association for Computational Linguistics 2018 Conference (ACL 2018)*. (* corresponding author)
- V. Zhong, C. Xiong, R. Socher. Global-Locally Self-Attentive Encoder for Dialogue State Tracking. *Association for Computational Linguistics 2018 Conference (ACL*

2018).

- C. Xiong, V. Zhong, R. Socher. DCN+: Mixed Objective and Deep Residual Coattention for Question Answering. *International Conference on Learning Representations (ICLR)*, 2018.
- J. Gu, J. Bradbury, C. Xiong, R. Socher. Non-Autoregressive Neural Machine Translation. *International Conference on Learning Representations (ICLR)*, 2018.
- T. Shu, C. Xiong*, R. Socher. Hierarchical and Interpretable Skill Acquisition in Multi-task Reinforcement Learning. *International Conference on Learning Representations (ICLR)*, 2018. (* corresponding author)
- A. Trott, C. Xiong*, R. Socher. Interpretable Counting for Visual Question Answering. *International Conference on Learning Representations (ICLR)*, 2018. (* corresponding author)
- R. Paulus, C. Xiong*, R. Socher. A Deep Reinforced Model for Abstractive Summarization. *International Conference on Learning Representations (ICLR)*, 2018. (* corresponding author)
- V. Zhong, C. Xiong, R. Socher. Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning. <https://arxiv.org/abs/1709.00103>
- B. McCann, J. Bradbury, C. Xiong, R. Socher. Learned in Translation: Contextualized Word Vectors. *Advances in Neural Information Processing Systems (NIPS 2017)*.
- K. Hashimoto, C. Xiong*, Y. Tsuruoka and R. Socher. A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks. *The 2017 Conference on Empirical Methods on Natural Language Processing (EMNLP 2017)*. (* corresponding author)
- C. Xiong, J. Lu, D. Parikh and R. Socher. Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*.
- C. Xiong, V. Zhong and R. Socher. Dynamic Coattention Networks For Question Answering. *International Conference on Learning Representations (ICLR)*, 2017.
- J. Bradbury, S. Merity, C. Xiong and R. Socher. Quasi-Recurrent Neural Networks. *International Conference on Learning Representations (ICLR)*, 2017.
- S. Merity, C. Xiong, James Bradbury and R. Socher. Pointer Sentinel Mixture Models. *International Conference on Learning Representations (ICLR)*, 2017.
- S. Longpre, S. Pradhan, C. Xiong and R. Socher. A Way out of the Odyssey: Analyzing and Combining Recent Insights for LSTMs. arxiv.org/abs/1611.05104.
- C. Xiong, D. M. Johnson and J. J. Corso. Active Clustering with Model-Based Uncertainty Reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2016.
- C. Xiong, S. Merity and R. Socher. Dynamic Memory Networks for Visual and Textual Question Answering. *International Conference on Machine Learning (ICML)*, 2016.
- B. Li, T.F. Wu, C. Xiong and S.C. Zhu. Car Fluent Recognition with Spatial-Temporal And-Or Models. *In Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*, 2016. (Oral Presentation)
- C. Xiong, N. Shukla, W. Xiong, and S.C. Zhu. Robot Learning with a Spatial, Temporal, and Causal And-Or Graph. *In Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2016. (Oral Presentation)

- D. M. Johnson, C. Xiong and J. J. Corso. Semi-Supervised Nonlinear Distance Metric Learning via Forests of Max-Margin Cluster Hierarchies. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, doi:10.1109/TKDE.2015.2507130
- G. Chen, Haiying Zhang, C. Xiong. Maximum Margin Dirichlet Process Mixtures for Clustering. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2016*.
- N. Shukla, C. Xiong and S.C. Zhu. A Unified Framework for Human-Robot Knowledge Transfer. *AAAI Fall Symposium on AI for Human-Robot Interaction (AI-HRI), 2015*.
- B. X. Nie, C. Xiong and S.C. Zhu. Joint Action Recognition and Pose Estimation From Video. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2015*.
- C. Xu, S.-H. Hsieh, C. Xiong, and J. J. Corso. Can humans fly? Action understanding with multiple classes of actors. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2015*.
- R. Xu, C. Xiong, W. Chen, and J. J. Corso. Jointly modeling deep video and compositional text to bridge vision and language in a unified framework. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2015*.
- G.-S. Xia, Z. Wang, C. Xiong, L. Zhang. Accurate Annotation of Remote Sensing Images via Active Spectral Clustering with Little Expert Knowledge. *Remote Sens. 2015, 7, 15014-15045*.
- A. Barbu, D. Barrett, W. Chen, N. Siddharth, C. Xiong, J. J. Corso, C. D. Fellbaum, C. Hanson, S. J. Hanson, S. Helie, E. Malaia, B. A. Pearlmutter, J. M. Siskind, T. M. Talavage, and R. B. Wilbur. Seeing is worse than believing: Reading people's minds better than computer vision methods recognize actions. In *Proceedings of European Conference on Computer Vision (ECCV), 2014*.
- C. Xiong, Scott McCloskey, Shao-Hang Hsieh and J. J. Corso. Latent Domains Modeling for Visual Domain Adaptation. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2014. (Oral Presentation)*
- W. Chen, C. Xiong and J. J. Corso. Actionness Ranking with Lattice Conditional Ordinal Random Fields. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2014*.
- C. Xiong, W. Chen, G. Chen, D. M. Johnson and J. J. Corso. Adaptive Quantization for Hashing: An Information-Based Approach to Learning Binary Codes. In *Proceedings of SIAM International Conference on Data Mining (SDM), 2014. (Oral Presentation)*
- Z. Wang, G.-S. Xia, C. Xiong and L. Zhang. Spectral Active Clustering of Remote Sensing Images. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2014*.
- C. Xiong, D. M. Johnson and J. J. Corso. Uncertainty reduction for active image clustering via a hybrid global-local uncertainty model. In *Proceedings of AAAI Conference on Artificial Intelligence (Late-Breaking Papers Track), 2013*.
- D. M. Johnson, C. Xiong and J. J. Corso. Comprehensive cross-hierarchy cluster agreement evaluation. In *Proceedings of AAAI Conference on Artificial Intelligence (Late-Breaking Papers Track), 2013*.
- C. Xiong, C. Xu and J. J. Corso. Streaming hierarchical video segmentation. In *Proceedings of European Conference on Computer Vision (ECCV), 2012. (Oral Presentation)*

- C. Xiong, D. Johnson, R. Xu and J. J. Corso. Random forests for metric learning with implicit pairwise position dependence. *In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012)*.
(Oral Presentation)
- C. Xiong and J. J. Corso. Coaction discovery: Segmentation of common actions across multiple videos. *In Proceedings of Multimedia Data Mining Workshop in Conjunction with the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (MDMKDD), 2012*.
- G. Chen, C. Xiong and J. J. Corso. Dictionary transfer for image denoising via domain adaptation. *In Proceedings of IEEE International Conference on Image Processing (ICIP), 2012*. (Oral Presentation)
- C. Xiong, D. Johnson and J. J. Corso. Efficient max-margin metric learning. *In Proceedings of European Conference on Data Mining (ECDM), 2012*.
(Oral Presentation) (**Best paper award**)
- C. Xiong, D. Johnson and J. J. Corso. Spectral active clustering via purification of the k-nearest neighbor graph. *In Proceedings of European Conference on Data Mining (ECDM), 2012*.
- D. R. Schlegel, A. Y. C. Chen, C. Xiong, J. A. Delmerico, and J. J. Corso. AirTouch: Interacting with computer systems at a distance. *In Proceedings of IEEE Winter Vision Meetings: Workshop on Applications of Computer Vision (WACV), 2011*. (Oral Presentation)
- K. Zeng, M. Zhao, C. Xiong, and S. C. Zhu. From Image Parsing to Painterly Rendering. *ACM Transactions on Graphics(TOG), 2009*.
- L. Lin, Y. Wang, Y. Liu, C. Xiong and K. Zeng. Marker-less Registration Based on Template Tracking for Augmented Reality. *Multimedia Tools and Applications (MTA), 2009*.

- Press Coverage** 2017 Our work on Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning
- **TechCrunch**: Salesforce is using AI to democratize SQL so anyone can query databases in natural language
 - **Venturebeat**: Salesforce creates AI tool for talking to databases
- 2017 Our work on a Deep Reinforced Model for Abstractive Summarization
- **Forbes**: Salesforce Announces AI Breakthrough, Reducing Information Overload
 - **MIT Technology Review**: An Algorithm Summarizes Lengthy Text Surprisingly Well
- 2016 Our work on Dynamic Memory Networks for Visual and Textual Question Answering
- **New York Times**: Taking Baby Steps Toward Software That Reasons Like Humans
 - **MIT Technology Review**: The Memory Trick Making Computers Seem Smarter

Professional Services

- Journal Reviewer:
- International Journal of Computer Vision (IJCV)
 - IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
 - IEEE Transactions on Image Processing (TIP)
 - Neurocomputing
 - Pattern Recognition

Conference Reviewer/

- IEEE International Conference on Computer Vision (ICCV), 2017
- Neural Information Processing Systems (NIPS), 2017
- IEEE Computer Vision and Pattern Recognition (CVPR), 2017
- International Conference on Learning Representations (ICLR), 2017
- Neural Information Processing Systems (NIPS), 2016
- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016
- IEEE Computer Vision and Pattern Recognition (CVPR), 2016
- IEEE International Conference on Computer Vision (ICCV), 2015
- IEEE International Conference on Robotics and Automation (ICRA), 2015
- IEEE International Conference on Semantic Computing (ICSC), 2013