Keynote Speakers

Leonidas J. Guibas

Professor, Stanford University

Short Biography

Leonidas Guibas is the Paul Pigott Professor of Computer Science (and by courtesy, Electrical Engineering) at Stanford University, where he heads the Geometric Computation group. Prof. Guibas obtained his Ph.D. from Stanford University under the supervision of Donald Knuth. His main subsequent employers were Xerox PARC, DEC/SRC, MIT, and Stanford. He is a member and past acting director of the Stanford Artificial Intelligence Laboratory and a member of the Computer Graphics Laboratory, the Institute for Computational and Mathematical Engineering (iCME) and the Bio-X program. Professor Guibas has been elected to the US National Academy of Engineering and the American Academy of Arts and Sciences, and is an ACM Fellow, an IEEE Fellow and winner of the ACM Allen Newell award and the ICCV Helmholtz prize.

Title and Abstracts

Title: Shape Differences and Variability

Abstract: The world of both natural and man-made objects provides us with an abundance of geometric forms obtained through evolution or design. The shapes of humans, animals, vehicles, furniture, clothes, and organs has been forged by a myriad of forces or considerations related to their functionality, structure, aesthetics, tradition, or materials. Computational disciplines that deal with models of 3D geometry, including computer graphics, computer vision, and robotics have studied shape representations and used them to develop various notions of shape similarity, as needed by particular applications, such as shape search. So far, though, relatively little effort has been spent on trying to quantify differences between the shapes of 3D forms, beyond various notions of distance or "dissimilarity". This makes it difficult to navigate the shape space around a given shape -- as, for example, when one wants to refine a search in specific ways (e.g., "shoes like these, but with higher heels"). It is especially interesting to try to develop low-dimensional parametrizations of shape differences and variability that reflect the underlying semantics of the shapes.

This talk will present some mathematical and algorithmic efforts in this direction, trying to address the many challenges present due to the vast diversity of ways that shapes can be compared in both their geometry and their structure. For example, shape differences can be continuous (e.g., changes in size) or discrete (e.g., addition/removal of parts). In fact, beyond comparing just two shapes, one may want to understand variability across an entire shape collection, or to compare two collections. It can be important to separate different kinds of variability, for some may be nuisance and others may matter (e.g., normal human organ variability vs. various pathologies). Furthermore, variability of form is interesting not only across shapes but even within a single shape, to capture articulations of deformations, often related to shape function. Finally, the correlation between geometry and language when it comes to describing variability is an interesting object of study on its own. The talk, through small vignettes, will aim to throw a bit of light on these topics.

Presentation Abstracts SANYA.AI

Keynote Speakers

Hsiao-Wuen Hon

Corporate vice president of Microsoft

Director of Microsoft Research Asia

Short Biography

Dr. Hsiao-Wuen Hon is corporate vice president of Microsoft, chairman of Microsoft's Asia-Pacific R&D Group, and managing director of Microsoft Research Asia. He drives Microsoft's strategy for research and development activities in the Asia-Pacific region, as well as collaborations with academia.

Dr. Hon received a PhD in Computer Science from Carnegie Mellon University. He has been with Microsoft since 1995. He joined Microsoft Research Asia in 2004 as deputy managing director, stepping into the role of managing director in 2007. An IEEE Fellow and a distinguished scientist of Microsoft, Dr. Hon is an internationally recognized expert in speech technology. Dr. Hon has published more than 100 technical papers in international journals and at conferences. He co-authored a book, Spoken Language Processing, which is a graduate-level textbook and reference book in the area of speech technology used in universities around the world.

Title and Abstracts

Title: A Brief History of Intelligence

Abstract: Intelligence is the deciding factor of how human beings become the most dominant life forms on earth. Throughout history, human beings have developed tools and technologies which help civilizations evolve and grow. Computers, and by extension, artificial intelligence (AI), has played important roles in that continuum of technologies. Recently artificial intelligence has garnered much interest and discussion. As artificial intelligence are tools that can enhance human capability, a sound understanding of what the technology can and cannot do is also necessary to ensure their appropriate use. While developing artificial intelligence, we also found out the definition and understanding of our own human intelligence continue evolving. The debates of the race between human and artificial intelligence have been ever growing. In this talk, I will describe the history of both artificial intelligence and human intelligence (HI). From the great insights of the such historical perspectives, I would like to illustrate how AI and HI will co-evolve with each other and project the future of AI and HI.

Keynote Speakers

Eric P. Xing

Professor, Carnegie Mellon University

Founder, CEO and Chief Scientist of Petuum Inc.

Short Biography

Eric P. Xing is a Professor of Computer Science at Carnegie Mellon University, and the Founder, CEO and Chief Scientist of Petuum Inc., a 2018 World Economic Forum Technology Pioneer company that builds standardized artificial intelligence development platform and operating system for broad and general industrial AI applications. He completed his undergraduate study at Tsinghua University, and holds a PhD in Molecular Biology and Biochemistry from the State University of New Jersey, and a PhD in Computer Science from the University of California, Berkeley. His main research interests are the development of machine learning and statistical methodology, and large-scale computational system and architectures, for solving problems involving automated learning, reasoning, and decision-making in high-dimensional, multimodal, and dynamic possible worlds in artificial, biological, and social systems.

Prof. Xing currently serves or has served the following roles: associate editor of the Journal of the American Statistical Association (JASA), Annals of Applied Statistics (AOAS), IEEE Journal of Pattern Analysis and Machine Intelligence (PAMI) and the PLoS Journal of Computational Biology; action editor of the Machine Learning Journal (MLJ) and Journal of Machine Learning Research (JMLR); member of the United States Department of Defense Advanced Research Projects Agency (DARPA) Information Science and Technology (ISAT) advisory group. He is a recipient of the National Science Foundation (NSF) Career Award, the Alfred P. Sloan Research Fellowship in Computer Science, the United States Air Force Office of Scientific Research Young Investigator Award, the IBM Open Collaborative Research Faculty Award, as well as several best paper awards. Prof Xing is a board member of the International Machine Learning Society; he has served as the Program Chair (2014) and General Chair (2019) of the International Conference of Machine Learning (ICML); he is also the Associate Department Head of the Machine Learning Department, founding director of the Center for Machine Learning and Health at Carnegie Mellon University.

Title and Abstracts

Title: SysML: On System and Algorithm co-design for Practical Machine Learning

ABSTRACT: The rise of Big Data and AI computing has led to new demands for Machine Learning systems to learn complex models with millions to billions of parameters that promise adequate capacity to digest massive datasets and offer powerful and real-time predictive analytics thereupon. In this talk, I discuss a recent trend toward building new distributed frameworks for AI at massive scale known as “system and ML algorithm co-design”, or SysML -- system designs are tailored to the unique properties of ML algorithms, and algorithms are re-designed to better fit into the system architecture. I show how one can explore the underlying statistical and algorithmic characteristics unique to ML programs but not typical in traditional computer programs in designing the system architecture to achieve significant, universal, and theoretically sound power-up of ML program across the board. I also present a briefly introduction of the Petuum system based on such interdisciplinary innovations, which intends to dramatically improve adoption of AI solutions by lowering the barrier of entry to AI technologies via Automatic Machine Learning through Petuum. I show how, through automatable, product-grade, hardware-agnostic, standardized building blocks that can be assembled and customized, AI users can liberate themselves from the demanding experience of algorithm programming and system tuning, and easily experiment with different AI methods, parameters, and speed/resource trade-offs by themselves or automatically.

To put this in a broader context, recent discussions about AI in both research community, and the general public have been championing a novelistic view of AI, that AI can mimic, surpass, threaten, or even destroy mankind. And such discussions are fueled by mainly recent advances in deep learning experimentations and applications, which are however often plagued by its craftiness, un-interpretability, and poor generalizability. I will discuss a different view of AI as a rigorous engineering discipline and as a commodity, where standardization, modularity, repeatability, reusability, and transparency are commonly expected, just as in civil engineering where builders apply principles and techniques from all sciences to build reliable constructions. I will discuss how such a view sets different focus, approach, metric, and expectation for AI research and engineering, which we practiced in our SysML work.

Keynote Speakers

Zhengyou Zhang

Director of Tencent Robotics X & AI Labs

Short Biography

Zhengyou Zhang is the Director of Tencent AI Lab and Tencent Robotics X Lab since March 2018. He is an ACM Fellow and an IEEE Fellow. He was a Partner Research Manager with Microsoft Research, Redmond, WA, USA, for 20 years.. Before joining Microsoft Research in March 1998, he was a Senior Research Scientist with INRIA (French National Institute for Research in Computer Science and Control), France, for 11 years. In 1996-1997, he spent a one-year sabbatical as an Invited Researcher with the Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan. He received the IEEE Helmholtz Test of Time Award in 2013 for his work published in 1999 on camera calibration, now known as Zhang's method.

Title and Abstracts

Title: Neural Network: Deep Supervision and Sensitivity Analysis

Abstract: In this talk, we describe two pieces of work: Deep supervision for training a neural network, and sensitivity analysis for using a neural network. Our proposed Deeply-Supervised Nets (DSN) intends to boost the classification performance by studying a new formulation in deep networks. Our method simultaneously minimizes classification error while making the learning process of hidden layers direct and transparent. We introduce “companion objective” to the individual hidden layers, in addition to the overall objective at the output layer (a different strategy to layer-wise pre-training). The advantage of our method is evident and our experimental result on benchmark datasets shows significant performance gain over existing methods. We also conduct sensitivity analysis of a trained network to study how important a feature or a scale is to the task in hand. This is done by cumulating the gradients of the objective function with respect to the feature of interest by applying the chain rule. The sensitivity analysis reveals many insights, and has been applied to facial expression recognition.

Presentation Abstracts SANYA.AI

Keynote Speakers

Songchun Zhu

Professor,Statistics and Computer Science

Director, Center for Vision, Cognition, Learning and Autonomy

University of California, Los Angeles

Short Biography

Songchun Zhu received his Ph.D. degree from Harvard University. He is currently professor of Statistics and Computer Science at UCLA, and heads The Center for Vision, Cognition, Learning, and Autonomy (VCLA). His research interests include vision, statistical modeling, learning, cognition, situated dialogues,robot autonomy and AI. He is in Editorial board of IJCV, and was associate editr of IEEE PAMI, general Chair of CVPR 2012 and CVPR 2019. He received a number of honors, including the Helmholtz Test-of-time award in ICCV 2013, the Aggarwal prize from the IAPR in 2008, the David Marr Prize in 2003 with Z. Tu et al. for image parsing, twice Marr Prize honorary nominations with Y. Wu et al. in 1999 for texture modeling and 2007 for object modeling respectively.He received the Sloan Fellowship in 2001, a US NSF Career Award in 2001, and an US ONR Young Investigator Award in 2001. He is a Fellow of IEEE since 2011.

Title and Abstracts

Title: Explainable AI: How Machines Gain Justified Trust from Humans

Abstract: The recent progresses in computer vision, machine learning, and AI in general have produced machines for a broad range of applications, however, some key underlying representations, especially neural networks, remain opaque or black boxes. This generates renewed interest in studying representations and algorithms that are interpretable and developing systems that can explain their behaviors and decisions to human users. In this talk, I will introduce our work on explainable AI: how machines gain justified trust from humans. The objective is to let human users understand how an AI system works, when it will succeed and fail for what reasons. Thus human and machine can collaborate more effectively in various tasks. We propose a framework called X-ToM: Explanation with Theory of Minds, which poses explanation as an iterative dialogue process between human and AI system. In this process, human and machine learn the mental representations of each other to establish better understanding. Our experiments on human subject shows X-ToM gains justified trust and reliance from users over time.

Session Speaker

1. Bo Chen
Xidian University

Title: Deep Autoencoding Topic Model

Abstract: To build a flexible and interpretable model for document analysis, in this talk we introduce our recent work, deep autoencoding topic model (DATM) that uses a hierarchy of gamma distributions to construct its multi-stochastic-layer generative network. In order to provide scalable posterior inference for the parameters of the generative network and efficiently infer the local latent representations via the inference network, we develop a scalable hybrid Bayesian inference method consisting of topic-layer-adaptive stochastic gradient Riemannian MCMC and a Weibull upward-downward variational encoder. Based on this fundamental deep model, we also show its potential extensions on different tasks, such as supervised topic modeling, deep dynamic modeling, image-text joint modeling and its convolutional version for documents with word order.

2. Gregory Chirikjian
Johns Hopkins University and National University of Singapore

Title: Learning and Lie Groups

Abstract: Machine learning methods are mostly based on calculus and probability and statistics on Euclidean spaces. However, many interesting problems can be articulated as learning in lower dimensional embedded manifolds and on Lie groups. This talk reviews how learning and Lie groups fit together, and how the machine learning community can benefit from modern mathematical developments. The topics include: Introduction to Calculus on Lie Groups (Differential Operators, Integration) Probability on Lie Groups (Convolution, Fourier Analysis, Diffusion Equations) Application 1: Workspace Generation and Inverse Kinematics of Highly Articulated Robotic Manipulators Application 2: Pose Distributions for Mobile Robots Application 2: Lie-Theoretic Invariances in Image Processing and Computer Vision Application 3: Coset-Spaces of Lie Groups by Discrete Subgroups in Crystallography Prospects for the Future

3. Peng Cui,
Tsinghua University

Title: Towards Stable Prediction across Unknown Environments

Abstract: Predicting unknown outcome values based on their observed features using a model estimated on a training data set is a common statistical problem. Many machine learning and data mining methods have been proposed and shown to be successful when the test data and training data come from the same distribution. However, the best-performing models for a given distribution of training data typically exploit subtle statistical relationships among features, making them potentially more prone to prediction error when applied to test data sets where, for example, the joint distribution of features differs from that in the training data. Therefore, it can be useful to develop predictive algorithms that are robust to shifts in the environment, particularly in application areas where models cannot be retrained as quickly as the environment changes. In this talk, I will describe a view of this problem from correlation versus causality perspective, and describe a number of our recent efforts in this direction.

4. Liang Lin
Sun Yat-Sen University

Title: Representation Learning Meets Knowledge Reasoning: New Trends in AI

Abstract: With the development of deep representation learning methods, great progress has been made in many traditional perceptual AI tasks (e.g., face detection and speech recognition). These methods usually make prediction (e.g., classification or regression) by abstracting feature representations directly from the input data and however suffer from several problems such as the requirement of large amount of training data and the incapability on handling many higher-level applications (e.g., task planning and situational dialogue). Recently, much attention has been attracted for incorporating symbolic knowledge and commonsense inference with the feature representation learning, and a few pioneering works have demonstrated very promising results on harnessing the above-mentioned problems, which is the focus of this talk. And I will also introduce several relevant works on this new direction from my group, such as the graph reasoning for improving large-scale object detection and knowledge-routed relational dialogue system.

5. Yugang Jiang
Fudan University

Title: Video Content Recognition: Datasets, Algorithms and Applications

Abstract: Nowadays people produce a huge number of videos. There is a strong need to develop automatic solutions for recognizing the contents of these videos. Potential applications of such techniques include effective video content management and retrieval, smart advertising, open-source intelligence analysis, etc. In this talk, I will introduce some recent progresses on video content analysis. I will start by introducing a few representative Internet video datasets. After that I will introduce several recent approaches developed in my group, followed by some real-world applications.

6. Jian Li
Tsinghua University

Title: On Generalization Error Bounds of Noisy Gradient Methods for Non-Convex Learning

Abstract: Generalization error (also known as the out-of-sample error) measures how well the hypothesis obtained from the training data can generalize to previously unseen data. Obtaining tight generalization error bounds is central to statistical learning theory. In this paper, we study the generalization error bound in learning general non-convex objectives, which has attracted significant attention in recent years. In particular, we study the (algorithm-dependent) generalization bounds of various iterative gradient based methods.
(1) We present a very simple and elementary proof of a recent result for stochastic gradient Langevin dynamics (SGLD), due to Mou et al. (2018). Our proof can be easily extended to obtain similar generalization bounds for several other variants of SGLD (e.g., with postprocessing, momentum, mini-batch, acceleration, and more general noises), and improves upon the recent results in Pensia et al. (2018).
(2) By incorporating ideas from the PAC-Bayesian theory into the stability framework, we obtain tighter distribution-dependent (or data-dependent) generalization bounds. Our bounds provide an intuitive explanation for the phenomenon reported in Zhang et al. (ICLR 2017).

(3) We also study the setting where the total loss is the sum of a bounded loss and an additional `l2 regularization term. We obtain new generalization bounds for the continuous Langevin dynamic in this setting by leveraging the tool of Log-Sobolev inequality. Our new bounds are more desirable when the noisy level of the process is not small, and do not grow when T approaches to infinity.

7. Shiguang Shan
Institute of Computing Technology, CAS

Title: Visual Recognition across Heterogeneous Patterns

Abstract: In computer vision, we often need to recognize objects represented in different views, modalities or domains. For instance, real-world applications need to recognize a person in non-frontal view while he/she was enrolled only a face photo in frontal view. Similarly, it is also demanding to recognize faces captured under near-infrared lighting by matching against face images taken under visible lighting. Furthermore, a recognition model learned in one domain sometimes needs to be adapted to work in another domain providing only some unsupervised data in the novel domain. Even further, the so-called zero-shot object recognition needs to match between the image of an object and its textual description. All the above recognition scenarios require to learning representations that bridging heterogeneous visual patterns. In this talk, I will discuss the general principle solving these problems and introduce some of our recent practices.

8. Jingkuan Song
University of Electronic Science and Technology of China

Title: Generative Adversarial Networks with Its Applications

Abstract: GANs consist of one generator and one discriminator. The generator is designed to generate as realistic images as possible to fool the discriminator, and the discriminator is optimized to distinguish between fake generated images and real ground-truth images. Recently, GANs have been applied to various applications, especially in simulating complicated data distributions, like images, music, texts and videos. More recently, GANs have been used to generate photo-realistic fine-grained images conditioned on semantic texts. This task demands the network to learn a precise mapping from semantic text distribution to visual image distribution. Meanwhile, the network, as a generative model, is requested to synthesize various and natural images that match text descriptions, not just like the ground-truth images in pixel space. In this talk, we start from the basic ideas of GANs, and then introduce three typical applications, i.e., image retrieval, face aging and text-to-image synthesis.

9. Jinhui Tang
Nanjing University of Technology

Title: Learning Visual Semantics from Noise Corrupted Social Media with Tag Refinement

Abstract: How to improve the visual understanding performance by utilizing the machine learning techniques has attracted much attention in the multimedia and computer vision research communities. The effectiveness of these machine learning methods heavily relies on the availability of a sufficiently large set of training samples. Manual labeling the training data is very time-consuming and labor-intensive, even impossible for the big multimedia data. This talk presents some research efforts in utilizing the social images crawled from Internet with noisy tags as training data to train the visual understanding models as well as the deep transfer network. In this kind of approaches, the key problem is how to handle the tag noise. I will also introduce our work on social tag refinement by tensor completion.

10. Liang Wang
Institute of Automation, CAS

Title: Deep Cognitive Networks and Their Visual Applications

Abstract: Deep neural networks have achieved great success in a wide range of applications. However, there still exists a huge performance gap between the best deep model and human cognitive system. Unlike human cognitive (e.g., visual) system, the current best deep model still cannot reliably guide blind people across the street. Although the current deep models can effectively implement the nonlinear mappings from information perception to primary decision, they ignore to model the higher-level cognitive mechanisms, which are usually regarded to play essential roles during information processing. This talk will introduce our recent work on deep modeling of key cognitive mechanisms such as attention and memory from neuroscience, called deep cognitive networks, as well as their applications in terms of multimodal learning, person re-id and semantic segmentation.

11. Liwei Wang
Peking University

Title: Towards Understanding Learning Representation: To What Extent Do Two Networks Learn the Same Representation

Abstract: It is widely believed that learning good representations is one of the main reasons for the success of deep neural networks. Although highly intuitive, there is a lack of theory and systematic approach quantitatively characterizing what representations do deep neural networks learn. In this work, we move a tiny step towards a theory and better understanding of the representations. Specifically, we study a simpler problem: How similar are the representations learned by two networks with identical architecture but trained from different initializations. We develop a rigorous theory based on the neuron activation subspace match model. The theory gives a complete characterization of the structure of neuron activation subspace matches, where the core concepts are maximum match and simple match which describe the overall and the finest similarity between sets of neurons in two networks respectively. We also propose efficient algorithms to find the maximum match and simple matches. Finally, we conduct extensive experiments using our algorithms. Experimental results suggest that, surprisingly, representations learned by the same convolutional layers of networks trained from different initializations are not as similar as prevalently expected, at least in terms of subspace match.

12. Mingsheng Long
Tsinghua University

Title: Transfer Learning: Generalizing Deep Learning across Domains and Tasks

Abstract: Transfer learning is a subfield of machine learning that enables adaptation of learning machines to novel domains. While recent studies reveal that deep neural networks can learn transferable features to bridge different domains, the transferability drops significantly by increasing domain discrepancy. In this talk, I will introduce several recent advances made in bridging deep learning and transfer learning to empower disentangled and transferable representations that generalize across different domains and tasks. This talk will cover the principled approaches to distribution comparisons of marginal and joint distributions, from non-parametric/semi-parametric statistical distances to multiple/conditional adversarial discriminators, as well as some theoretical insights underlying these approaches.

13. Kai Yu
Shanghai Jiaotong University

Title: Dialogue Policy Adaptation with Structured Deep Reinforcement Learning

Abstract: Dialogue management is the core of conversational AI agent. It is a sequence decision problem and there has been a long lasting interest to introduce data driven approach to it. Policy optimization is the core part of statistical dialogue management. Deep reinforcement learning has been successfully used for dialogue policy optimization for a static pre-defined domain. However, when the domain changes dynamically, e.g. a new previously unseen concept (or slot) which can be then used as a database search constraint is added, or the policy for one domain is transferred to another domain, the dialogue state space and action sets both will change. This makes dialogue policy adaptation/transfer very challenging and attractive. In this talk, basic concepts of spoken dialogue system, especially dialogue management will be introduced. Then a new structured deep reinforcement learning algorithm is proposed to address the policy adaptation problem, referred to as a multi-agent dialogue policy (MADP). Simulation experiments showed that MADP can significantly speed up the policy learning and facilitate policy adaptation.

14. Yang Yu
Nanjing University

Title: Towards Read-World Decision Making via Reinforcement Learning

Abstract: Optimal decision making is a central topic in AI. Recently, reinforcement learning has shown its extraordinary, or even super-human ability of decision making in playing games, such as in the AlphaGo for the game of Go and the AlphaStar for StarCraft II. However, the successes were happened majorly in digital worlds instead of real-world scenarios, due to the unbearable cost of real-world trial-and-errors. We are exploring ways of real-world reinforcement learning. In this talk, we will report our recent progress on environment virtualization and related efforts to make reinforcement learning applicable and robust. We believe that the realization of real-world reinforcement learning will unleash a strong world-changing power of AI.

15. Hanwang Zhang
Nanyang Technological University

Title: Towards X Visual Reasoning

Abstract: For decades, we are interested in detecting objects and classifying them into a fixed vocabulary of lexicon. With the maturity of these "low-level" vision solutions, we are hunger for a "higher-level" representation of the visual data, so as to extract visual knowledge rather than merely bags of visual entities, allowing machines to reason about human-level decision-making. In particular, we wish an ``X" reasoning, where X means eXplainable and eXplicit. In this talk, I will first review a brief history of today's prevailing visual detection methods that underpin most today's multimedia frameworks. Then, I will introduce more challenging vision-language tasks such as visual relationship detections, natural language grounding, and image captioning. Finally, I will share my recent vision in visual reasoning by using visual knowledge graph to break current black-box deep learning models for the above multimodal tasks, especially those require reasoning such as visual Q&A and visual dialog. Through my talk, I hope that we are more ambitious to advance to the new era of multimedia research: knowledge representation + reasoning = AI.