ICRI-CI Retreat May 24-25, 2016 – Agenda + Abstracts (updated May 5, 2016)

Start	Duration	End	Session	Speaker	Title/Chair
Day 1, Tuesday, 24-May-2016					
8:15	0:45	9:00	**Registration, Gathering, Breakfast**		
9:00	0:20	9:20	**Opening**		
				Ronny + Shalom	**Opening Notes**
9:20	3:55	13:15	**Deep Learning**		**Shai Fine**
9:20	0:10	9:30		**Shai Fine**	**Deep Learning - Capstone Overview**
9:30	0:25	9:55		**Shie Mannor**	**Graying the Black Box of Deep Q Network** In recent years there is a growing interest in using deep representations for reinforcement learning. We present a methodology and tools to analyze Deep Q-networks (DQNs) in a non-blind matter. Using our tools we reveal that the features learned by DQNs aggregate the state space in a hierarchical fashion, explaining their success. Moreover, we understand and describe the policies learned by DQNs for several Atari2600 games and suggest ways to interpret, debug, and optimize of deep neural networks in reinforcement learning.
9:55	0:25	10:20		**Naftali Tishby**	**Irreducible Representations of Symmetries and Solvable Deep Learning Models** TBD
10:20	0:25	10:45		**Shai Shalev Shwartz**	**Minimizing the Maximal Loss: Why and How** We argue that in some situations, minimizing the maximal loss over the training set is essential for achieving good performance on unseen examples. We present a stochastic algorithm for solving the resulted optimization problem.
10:45	0:30	11:15	**Break + Posters**		
11:15	0:25	11:40		**Amnon Shashua**	**On the Quest for a Deeper Understanding of Deep Learning** Despite of its unprecedented success, and the vast attention it is receiving, our formal understanding of deep learning is limited. In this talk we survey some of the fundamental theoretical questions and challenges the field is facing. Subsequently, we present a line of works addressing the question of expressiveness with convolutional networks. We first show that convolutional arithmetic circuits exhibit complete depth efficiency, meaning that besides a zero measure set, all functions realized by a deep network cannot be efficiently realized by shallow networks. Second, we find the expressiveness of the popular convolutional rectifier networks to be inferior, in the sense that there is positive measure to the set of functions realized by a deep network that can also be efficiently realized by shallow networks. Third, we analyze functions in terms of the correlations they model between input variables, showing that shallow networks are incapable of modeling strong correlations, while deep networks bring forth the ability to model strong correlations between input regions that are interleaved.

11:40	0:25	12:05		**Ohad Shamir**	**Geometric Structure of Multilayer Neural Networks** Deep learning, in the form of multilayer neural networks, has achieved remarkable success in recent years, for a variety of difficult machine learning applications. However, we have little understanding of the type of functions expressed by these networks, and training them involves optimizing a highly non-convex objective. In this talk, I'll describe some recent results on the geometry of these objects, focusing on how and why larger networks can be more powerful and easier to train in practice.
12:05	0:25	12:30		**Nati Srebro**	**Geometry, Optimization and Regularization in Deep Learning** TBD
12:30	0:25	12:55		**Lior Wolf**	**RNN Fisher Vectors for Action Recognition and Image Annotation** Recurrent Neural Networks (RNNs) have had considerable success in classifying and predicting sequences. We demonstrate that RNNs can be effectively used in order to encode sequences and provide effective representations. The methodology we use is based on Fisher Vectors, where the RNNs are the generative probabilistic models and the partial derivatives are computed using backpropagation. State of the art results are obtained in two central but distant tasks, which both rely on sequences: video action recognition and image annotation. We also show a surprising transfer learning result from the task of image annotation to the task of video action recognition.
12:55	0:20	13:15		**Shai Fine**	**Arch/ML Brainstorm Report**
13:15	1:00	14:15	**Lunch + Posters**		
14:15	1:00	15:15	**Arch/ML Brainstorm Breakout session**		**Shai Fine**
15:15	0:30	15:45	**Break + Posters**		
15:45	2:00	17:45	**Architecture**		**Debbie Marr**
15:45	0:20	16:05		**Debbie Marr**	**Architecture - Capstone Overview**
16:05	0:25	16:30		**Uri Weiser**	**To Process-in-Storage, or not to Process-in-Storage? That is the question** Memory hierarchy in modern computing systems perform well for workloads that exhibit temporal data locality. Data that is accessed frequently is brought to DRAM and caches that are close to the computing cores, allowing fast data access, high data bandwidth, and reduced transmission energy. However, many Big Data applications, demonstrate “Read once” data access from memory that exhibit Non-Temporal locality Memory Access patterns. When running these applications on modern computing systems, large amount of data that is read once is nevertheless transmitted and copied to memory hierarchy levels, leading to energy waste and bandwidth pressure. In this talk we’ll ask the question: “To Process-in-Storage, or not to Process-in-Storage?” i.e. when target is performance and energy savings, in which-case/where should we perform the Big Data computation. We will revisit the “Funnel” function and provide some analytical analysis on the usage and the impact.

16:30	0:25	16:55		**Ran Ginosar**	**From Processing-in-Memory to Processing-in-Storage** Processing In-Storage is an evolution of Processing in Memory. Processing moves one level further in the storage hierarchy, from the main memory to storage. Some of its benefits are obvious: significant reduction in latency and energy by eliminating or significantly shortening data transfer. There are two phenomena unique to big data processing. First, the workloads (10^{12}B-10^{15}B) don't fit in main memory (which is limited to 10^{11}B-10^{12}B). Second, in many data analytics and machine learning tasks, majority of raw data is used only once. Processing In-Storage eliminates transfer of the used-once-data across the entire memory hierarchy by processing it in-situ. Processing In-Storage also enables in-situ processing of workloads that do not fit in main memory. However, to make it truly efficient, a new high performance computing architecture, which integrates mass storage and massively parallel processing is of great value. One such architecture is the award-winning* "Resistive Associative Processor", a massively parallel SIDM machine built upon Resistive Content Addressable Memory. ** one of four Best of CAL winners for 2015*
16:55	0:25	17:20		**Yuval Cassuto**	**Similarity Calculations in Resistive Memories** In data-intensive applications, memory accesses are becoming the bottleneck for performance and power consumption. This motivates moving significant parts of the computation into the memory. In the talk we will explore one such in-memory computing paradigm. We consider a memory device that stores a large number of binary vectors. Each 0/1 vector element is mapped to a primitive resistive element that is in high/low resistance state, respectively. Each vector is mapped to a memory row. The problem we consider is how to use the resistive physical medium to compute similarity between pairs of vectors within the memory array. In particular, we show how coding the data vectors can significantly improve performance.
17:20	0:25	17:45		**Yoav Etsion**	**Using Neural Networks for Memory Prefetching: The Good the Bad and the Ugly...** Semantic locality describes the use of inherent program semantics to characterize memory access relations, and how it can capture the relationship between data elements in a manner agnostic to the actual data layout. By capturing the relationship between data elements, semantic locality transcends spatio-temporal locality and can be used for memory prefetching. Our work explores the potential of a neural memory prefetcher, which uses neural networks to approximate semantic locality and guide prefetching decisions. In this talk I will discuss the effectiveness of neural networks in dynamically inferring access patterns from machine and code attributes. I will further discuss some limitations that, we believe, make neural networks the wrong machine learning tool for this job.
17:45	1:15	19:00	**Posters + Reception**		

Day 2, Wednesday, 25-May-2016					
8:15	0:45	9:00	**Registration, Gathering, Breakfast**		
9:00	1:40	10:40	**Conversational Understanding**		**Moshe Wasserblat**
9:00	0:25	9:25		**Moshe Wasserblat**	**Conversational Understanding - Capstone Overview**
9:25	0:25	9:50		**Ido Dagan**	**Integrated Pattern-Based and Distributional Learning of Lexical Relations with Recurrent Neural Networks** Recognizing lexical inference relations between terms is an essential component in semantic tasks. In question answering, for instance, identifying that 'broadcast' and 'air' are synonymous enables answering the question "When was 'Friends' first aired?" given the text "'Friends' was first broadcast in 1994". Semantic relations such as synonymy (tall, high) and hypernymy (cat, pet) are used to infer the meaning of one term from another, in order to overcome lexical variability. In this talk we describe a novel corpus-based method for detecting lexical inference relations, based on a recurrent neural network architecture. Our approach integrates distributional similarity signals, whose supervised variants are the current best performers, and pattern-based methods, which received less research attention in recent years. We suggest an improved path-based algorithm, in which syntactic dependency paths are encoded using a recurrent neural network, and achieve results comparable to distributional methods on the hypernymy detection task. We then extend the approach to integrate both path-based and distributional signals, significantly improving the state-of-the-art on this task. Our current work extends this approach to recognize simultaneously multiple types of lexical semantic relations.
9:50	0:25	10:15		**Moshe Koppel**	**Improving Speech Recognition with Discriminative Fluency Classification** Speech recognition systems have been improving greatly in recent years, but are still far from being perfect. It is very common to find a generative language model as a core component in such systems, estimating the fluency of possible outputs for a given input signal. We propose a complementary approach for fluency estimation, by introducing the Discriminative Fluency Classifier - a classifier that discriminates the systems outputs from general, human generated text. We show that there is a strong negative correlation between the classifier accuracy and the speech output quality, which enables to use it as an evaluation metric for speech recognition or as a re-ranking component when choosing the best output for a given input. This is a natural step from our previous work which showed similar results for machine translation.
10:15	0:25	10:40		**Amir Globerson**	**Semisupervised learning with variational probabilities** Predicting the label Y of an object X is a core task in machine learning. From a probabilistic perspective, this involves reasoning about conditional probabilities p(y|x). However, it is hard to obtain reliable estimates for these probabilities. Here we provide a variational approach for bounding p(y|x) given statistical information, and show how it can be used within various learning setups.
10:40	0:30	11:10	**Break + Posters**		

11:10	1:50	13:00	**Visual**		**Ronny Ronen**
11:10	0:10	11:20		**Ronny Ronen**	**Session Overview**
11:20	0:25	11:45		**Sarit Kraus**	**Strategical Argumentative Agent for Human Persuasion** Automated agents should be able to persuade people in the same way people persuade each other - via dialogs. Today, automated persuasion modeling and research use unnatural assumptions regarding the persuasive interaction which creates doubt regarding their applicability for real-world deployment with people. In this work we present a novel methodology for persuading people through argumentative dialogs. Our methodology combines theoretical argumentation modeling, machine learning and Markovian optimization techniques that together result in an innovative agent named SPA. Two extensive field experiments, with more than 100 human subjects, show that SPA is able to persuade people significantly more often than a baseline agent and no worse than people are able to persuade each other.
11:45	0:25	12:10		**Lior Wolf**	**PatchBatch: a Batch Augmented Loss for Optical Flow** We propose a new pipeline for optical flow computation, based on Deep Learning techniques. We suggest using a Siamese CNN to independently, and in parallel, compute the descriptors of both images. The learned descriptors are then compared efficiently using the L2 norm and do not require network processing of patch pairs. The success of the method is based on an innovative loss function that computes higher moments of the loss distributions for each training batch. Combined with an Approximate Nearest Neighbor patch matching method and a flow interpolation technique, state of the art performance is obtained on the most challenging and competitive optical flow benchmarks.
12:10	0:25	12:35		**Yair Weiss**	**Statistics of RGBD Images** Cameras that can measure the depth of each pixel in addition to its color have become easily available and are used in many consumer products worldwide. Often the depth channel is captured at lower quality compared to the RGB channels and different algorithms have been proposed to improve the quality of the D channel given the RGB channels. Typically these approaches work by assuming that edges in RGB are correlated with edges in D. In this research we approach this problem from the standpoint of natural image statistics. We obtain examples of high quality RGBD images from a computer graphics generated movie (MPI-Sintel) and we use these examples to compare different probabilistic generative models of RGBD image patches. We then use the generative models together with a degradation model and obtain a Bayes Least Squares (BLS) estimator of the D channel given the RGB channels. Our results show that learned generative models improve the state-of-the-art in improving the quality of depth channels given the color channels in natural images even when training is performed on artificially generated images.

12:35	0:25	13:00		**Shmuel Peleg**	**Calibration of Multiple Cameras using Scene Dynamics** We address multiple videos of a single event, captured from significantly different viewpoints. E.g. Two mobile phones videos viewing an accident from opposite sides of the road. We would like to find the relationship between the videos, such as: (i) Are the videos watching the same scene. (ii) What is the relative geometry of the videos ("camera calibration"). Appearance-based methods fail in such cases, as appearances change with large changes of viewpoints. We therefore introduce a temporal feature, “motion barcode”, which records the existence/non-existence of motion as a function of time. For some purposes motion barcode can be an image-point feature, and for other purposes it can be a feature along an image line.
13:00	1:00	14:00	**Poster Session + Lunch**		