Emilio Parisotto Website : http://cs.cmu.edu/~eparisot

Education

- **Carnegie Mellon University** Advisor: Ruslan Salakhutdinov
  PhD in Machine Learning *Aug. 2016 – Present*
- **University of Toronto (Transferred to CMU)** Advisor: Ruslan Salakhutdinov
  PhD in Machine Learning *Aug. 2015 – Apr. 2016*
- **University of Toronto** Trinity College
  Honours Bachelor of Science, Computer Science Specialist *Aug. 2012 – July 2015*
  GPA: 3.98/4.00

Preprints and Publications

Stabilizing Transformers for Reinforcement Learning 2019
E. Parisotto, H. F. Song, J. W. Rae, R. Pascanu, C. Gulcehre, S. M. Jayakumar, M. Jaderberg, R. L. Kaufman, A. Clark, S. Noury, M. M. Botvinick, N. Heess, R. Hadsell
Preprint.

Abstract: Owing to their ability to both effectively integrate information over long time horizons and scale to massive amounts of data, self-attention architectures have recently shown breakthrough success in natural language processing (NLP), achieving state-of-the-art results in domains such as language modeling and machine translation. Harnessing the transformer's ability to process long time horizons of information could provide a similar performance boost in partially observable reinforcement learning (RL) domains, but the large-scale transformers used in NLP have yet to be successfully applied to the RL setting. In this work we demonstrate that the standard transformer architecture is difficult to optimize, which was previously observed in the supervised learning setting but becomes especially pronounced with RL objectives. We propose architectural modifications that substantially improve the stability and learning speed of the original Transformer and XL variant. The proposed architecture, the Gated Transformer-XL (GTrXL), surpasses LSTMs on challenging memory environments and achieves state-of-the-art results on the multi-task DMLab-30 benchmark suite, exceeding the performance of an external memory architecture. We show that the GTrXL, trained using the same losses, has stability and performance that consistently matches or exceeds a competitive LSTM baseline, including on more reactive tasks where memory is less critical. GTrXL offers an easy-to-train, simple-to-implement but substantially more expressive architectural alternative to the standard multi-layer LSTM ubiquitously used for RL agents in partially observable environments.

Efficient Exploration via State Marginal Matching 2019
Lisa Lee, Benjamin Eysenbach, **Emilio Parisotto**, Eric Xing, Sergey Levine, Ruslan Salakhutdinov
Preprint.

Abstract: Reinforcement learning agents need to explore their unknown environments to solve the tasks given to them. The Bayes optimal solution to exploration is intractable for complex environments, and while several exploration methods have been proposed as approximations, it remains unclear what underlying objective is being optimized by existing exploration methods, or how they can be altered to incorporate prior knowledge about the task. Moreover, it is unclear how to acquire a single exploration strategy that will be useful for solving multiple downstream tasks. We address these shortcomings by learning a single exploration policy that can quickly solve a suite of downstream tasks in a multi-task setting, amortizing the cost of learning to explore. We recast exploration as a problem of State Marginal Matching (SMM), where we aim to learn a policy for which the state marginal distribution matches a given target state distribution, which can incorporate prior knowledge about the task. We optimize the objective by reducing it to a two-player, zero-sum game between a state density model and a parametric policy. Our theoretical analysis of this approach suggests that prior exploration methods do not learn a policy that does distribution matching, but acquire a replay buffer that performs distribution matching, an observation that potentially explains prior methods' success in single-task settings. On both simulated and real-world tasks, we demonstrate that our algorithm explores faster and adapts more quickly than prior methods.

Concurrent Episodic Meta Reinforcement Learning 2019
Emilio Parisotto, Soham Ghosh, Sai Bhargav Yalamanchi, Varsha Chinnaobireddy, Yuhuai Wu, Ruslan Salakhutdinov
Preprint.

Abstract: State-of-the-art meta reinforcement learning algorithms typically assume the setting of a single agent interacting with its environment in a sequential manner. A negative side-effect of this sequential execution paradigm is that, as the environment becomes more and more challenging, and thus requiring more interaction episodes for the meta-learner, it needs the agent to reason over longer and longer time-scales. To combat the difficulty of long time-scale credit assignment, we propose an alternative parallel framework, which we name "Concurrent Meta-Reinforcement Learning" (CMRL), that transforms the temporal credit assignment problem into a multi-agent reinforcement learning one. In this multi-agent setting, a set of parallel agents are executed in the same environment and each of these "rollout" agents are given the means to communicate with each other. The goal of the communication is to coordinate, in a collaborative manner, the most efficient exploration of the shared task the agents are currently assigned. This coordination therefore represents the meta-learning aspect of the framework, as each agent can be assigned or assign itself a particular section of the current task's state space. This framework is in contrast to standard RL methods that assume that each parallel rollout occurs independently, which can potentially waste computation if many of the rollouts end up sampling the same part of the state space. Furthermore, the parallel setting enables us to define several reward sharing functions and auxiliary losses that are non-trivial to apply in the sequential setting. We demonstrate the effectiveness of our proposed CMRL at improving over sequential methods in a variety of challenging tasks.

The Hanabi Challenge: A New Frontier for AI Research 2019
N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, **E. Parisotto**, V. Dumoulin, S. Moitra, E. Hughes, I. Dunning, S. Mourad, H. Larochelle, M. G. Bellemare, M. Bowling
In Journal of Artificial Intelligence 2020.

Abstract: From the early days of computing, games have been important testbeds for studying how well machines can do sophisticated decision making. In recent years, machine learning has made dramatic advances with artificial agents reaching superhuman performance in challenge domains like Go, Atari, and some variants of poker. As with their predecessors of chess, checkers, and backgammon, these game domains have driven research by providing sophisticated yet well-defined challenges for artificial intelligence practitioners. We continue this tradition by proposing the game of Hanabi as a new challenge domain with novel problems that arise from its combination of purely cooperative gameplay with two to five players and imperfect information. In particular, we argue that Hanabi elevates reasoning about the beliefs and intentions of other agents to the foreground. We believe developing novel techniques for such theory of mind reasoning will not only be crucial for success in Hanabi, but also in broader collaborative efforts, especially those with human partners. To facilitate future research, we introduce the open-source Hanabi Learning Environment, propose an experimental framework for the research community to evaluate algorithmic advances, and assess the performance of current state-of-the-art techniques.

Global Pose Estimation with an Attention-based Recurrent Network 2018
Emilio Parisotto*, Devendra Singh Chaplot*, Jian Zhang, Ruslan Salakhutdinov
In 1st International Workshop on Deep Learning for Visual SLAM, CVPR 2018. **Oral**. Best Student Paper Award.

Abstract: The ability for an agent to localize itself within an environment is crucial for many real-world applications. For unknown environments, Simultaneous Localization and Mapping (SLAM) enables incremental and concurrent building of and localizing within a map. We present a new, differentiable architecture, Neural Graph Optimizer, progressing towards a complete neural network solution for SLAM by designing a system composed of a local pose estimation model, a novel pose selection module, and a novel graph optimization process. The entire architecture is trained in an end-to-end fashion, enabling the network to automatically learn domain-specific features relevant to the visual odometry and avoid the involved process of feature engineering. We demonstrate the effectiveness of our system on a simulated 2D maze and the 3D ViZ-Doom environment.

Gated Path Planning Networks 2018
Lisa Lee*, **Emilio Parisotto***, Devendra Singh Chaplot, Eric Xing, Ruslan Salakhutdinov
In Proceedings of the 35th International Conference on Machine Learning (ICML 2018).

Abstract: Value Iteration Networks (VINs) are effective differentiable path planning modules that can be used by agents to perform navigation while still maintaining end-to-end differentiability of the entire architecture. Despite their effectiveness, they suffer from several disadvantages including training instability, random seed sensitivity, and other optimization problems. In this work, we reframe VINs as recurrent-convolutional networks which demonstrates that VINs couple recurrent convolutions with an unconventional max-pooling activation. From this perspective, we argue that standard gated recurrent update equations could potentially alleviate the optimization issues plaguing VIN. The resulting architecture, which we call the Gated Path Planning Network, is shown to empirically outperform VIN on a variety of metrics such as learning speed, hyperparameter sensitivity, iteration count, and even generalization. Furthermore, we show that this performance gap is consistent across different maze

transition types, maze sizes and even show success on a challenging 3D environment, where the planner is only provided with first-person RGB images.

Active Neural Localization 2018

Devendra Singh Chaplot, **Emilio Parisotto**, Ruslan Salakhutdinov

In Proceedings of the 6th International Conference on Learning Representations (ICLR 2018).

Abstract: Localization is the problem of estimating the location of an autonomous agent from an observation and a map of the environment. Traditional methods of localization, which filter the belief based on the observations, are sub-optimal in the number of steps required, as they do not decide the actions taken by the agent. We propose "Active Neural Localizer", a fully differentiable neural network that learns to localize efficiently. The proposed model incorporates ideas of traditional filtering-based localization methods, by using a structured belief of the state with multiplicative interactions to propagate belief, and combines it with a policy model to minimize the number of steps required for localization. Active Neural Localizer is trained end-to-end with reinforcement learning. We use a variety of simulation environments for our experiments which include random 2D mazes, random mazes in the Doom game engine and a photo-realistic environment in the Unreal game engine. The results on the 2D environments show the effectiveness of the learned policy in an idealistic setting while results on the 3D environments demonstrate the model's capability of learning the policy and perceptual model jointly from raw-pixel based RGB observations. We also show that a model trained on random textures in the Doom environment generalizes well to a photo-realistic office space environment in the Unreal engine.

Neural Map: Structured Memory for Deep Reinforcement Learning 2018

Emilio Parisotto, Ruslan Salakhutdinov

In Proceedings of the 6th International Conference on Learning Representations (ICLR 2018).

Abstract: A critical component to enabling intelligent reasoning in partially observable environments is memory. Despite this importance, Deep Reinforcement Learning (DRL) agents have so far used relatively simple memory architectures, with the main methods to overcome partial observability being either a temporal convolution over the past k frames or an LSTM layer. More recent work (Oh et al., 2016) has went beyond these architectures by using memory networks which can allow more sophisticated addressing schemes over the past k frames. But even these architectures are unsatisfactory due to the reason that they are limited to only remembering information from the last k frames. In this paper, we develop a memory system with an adaptable write operator that is customized to the sorts of 3D environments that DRL agents typically interact with. This architecture, called the Neural Map, uses a spatially structured 2D memory image to learn to store arbitrary information about the environment over long time lags. We demonstrate empirically that the Neural Map surpasses previous DRL memories on a set of challenging 2D and 3D maze environments and show that it is capable of generalizing to environments that were not seen during training.

Neural Program Synthesis 2017

Emilio Parisotto, Abdel-Rahman Mohamed, Rishabh Singh, Lihong Li, Dengyong Zhou, Pushmeet Kohli

In Proceedings of the 5th International Conference on Learning Representations (ICLR 2017).

Abstract: Recent years have seen the proposal of a number of neural architectures for the problem of Program Induction. Given a set of input-output examples, these architectures are able to learn mappings that generalize to new test inputs. While achieving impressive results, these approaches have a number of important limitations: (a) they are computationally expensive and hard to train, (b) a model has to be trained for each task (program) separately, and (c) it is hard to interpret or verify the correctness of the learnt mapping (as it is defined by a neural network). In this paper, we propose a novel technique, Neuro-Symbolic Program Synthesis, to overcome the above-mentioned problems. Once trained, our approach can automatically construct computer programs in a domain-specific language that are consistent with a set of input-output examples provided at test time. Our method is based on two novel neural modules. The first module, called the cross correlation I/O network, given a set of input-output examples, produces a continuous representation of the set of I/O examples. The second module, the Recursive-Reverse-Recursive Neural Network (R3NN), given the continuous representation of the examples, synthesizes a program by incrementally expanding partial programs. We demonstrate the effectiveness of our approach by applying it to the rich and complex domain of regular expression based string transformations. Experiments show that the R3NN model is not only able to construct programs from new input-output examples, but it is also able to construct new programs for tasks that it had never observed before during training.

Actor-Mimic: Deep Multitask and Transfer Reinforcement Learning 2016

Emilio Parisotto, Jimmy Lei Ba, Ruslan Salakhutdinov

In Proceedings of the 4th International Conference on Learning Representations (ICLR 2016)

Abstract: The ability to act in multiple environments and transfer previous knowledge to new situations can be considered a critical aspect of any intelligent agent. Towards this goal, we define a novel method of multitask and transfer learning that enables an autonomous agent to learn how to behave in multiple tasks simultaneously, and then generalize its knowledge to new domains. This method, termed "Actor-Mimic", exploits the use of deep reinforcement learning and model compression techniques to train a single policy network that learns how to act in a set of distinct tasks by using the guidance of several expert teachers. We then show that the representations learnt by the deep policy network are capable of generalizing to new tasks with no prior expert guidance, speeding up learning in novel environments. Although our method can in general be applied to a wide range of problems, we use Atari games as a testing environment to demonstrate these methods.

Generating Images from Captions with Attention 2016

Elman Mansimov, **Emilio Parisotto**, Jimmy Lei Ba, Ruslan Salakhutdinov

In Proceedings of the 4th International Conference on Learning Representations (ICLR 2016). **Oral**

Abstract: Motivated by the recent progress in generative models, we introduce a model that generates images from natural language descriptions. The proposed model iteratively draws patches on a canvas, while attending to the relevant words in the description. After training on Microsoft COCO, we compare our model with several baseline generative models on image generation and retrieval tasks. We demonstrate that our model produces higher quality samples than other approaches and generates images with novel scene compositions corresponding to previously unseen captions in the dataset.

Professional Experience

DeepMind *July 2019 - Nov 2019*
Research Scientist Intern *London, UK*
Advisor: Raia Hadsell

Google Brain Montreal *May 2018 - Aug 2018*
Research Intern *Montreal, QC*
Advisor: Vincent Dumoulin

Apple *May 2017 - Dec 2017*
Research Intern *Sunnyvale, CA*
Advisor: Ruslan Salakhutdinov

Microsoft *May 2016 - Aug 2016*
Research Intern *Redmond, WA*
Advisor: Pushmeet Kohli

University of Toronto *May 2014 – Aug 2014*
CSC384 Course Development *Toronto, ON*
Advisors: Sheila McIlraith, Fahiem Bacchus

Professional Activities

Reviewer NeurIPS 2017-2020, ICML 2019-2020, AAAI 2019.

Talks

Neural Map: Structured Memory for Deep Reinforcement Learning Nvidia GTC 2017

Awards and Honours

Natural Sciences and Engineering Council (NSERC) PGS-D	2016 – 2018
Natural Science and Engineering Council (NSERC) CGS-D (declined)	2016
Natural Science and Engineering Council (NSERC) CGS-M	2015 – 2016
Drew Thompson Graduation Scholarship	2015
Provost's Graduation Scholarship	2015
James Scott Scholarship	2014
Drew Thompson Scholarship	2013
Dean's List Scholar at the University of Toronto	2013 – 2015
NSERC Undergraduate Research Award (declined)	2014
Rensselaer Medalist (declined)	2010

Programming Skills

Languages: Python, C, C++, Lua, Matlab, Java, C#, Prolog, Scheme, Perl
Frameworks: Pytorch, Torch, Tensorflow, Theano, Caffe

Courses

10-725 Convex Optimization
10-705 Intermediate Statistics
10-702 Statistical Machine Learning
10-715 Advanced Introduction to Machine Learning
10-703 Deep Reinforcement Learning and Control