NeurIPS 2019 Competition and Demonstration Track Revised selected papers

Hugo Jair Escalante* HUGOJAIR@INAOEP.MX
Computer Science Department,
Instituto Nacional de Astrofísica, Óptica y Electrónuca,
Puebla, Mexico
Raia Hadsell RAIA@GOOGLE.COM
DeepMind,
UK

Editors: Hugo Jair Escalante and Raia Hadsell

Machine learning competitions have grown in popularity and impact over the last decade, emerging as an effective means to advance the state of the art by posing well-structured, relevant, and challenging problems to the community at large. Motivated by a reward or merely the satisfaction of seeing their machine learning algorithm reach the top of a leaderboard, practitioners innovate, improve, and tune their approach before evaluating on a held-out dataset or environment. The competition track of NeurIPS has matured in 2019, its third year, with a considerable increase in both the number of challenges and the diversity of domains and topics. Demonstrations offer a complementary dimension to the competitions, focusing on areas of machine learning which are either human interactive or demonstrable in some way, for instance robotics applications or generative models.

This volume is a compilation of selected papers associated with the NeurIPS 2019 Demonstration and Competition Track. The scope of the volume includes the design of the competitions, analysis of the results, novel methodologies developed to respond to the competitions' challenges, and the design and development of creative demonstrations. A total of 16 competitions and 28 demonstrations were accepted to be part of the program at NeurIPS 2019. Both tracks were subject to a strict reviewing process to ensure the quality of the accepted events. The accepted competitions are briefly described in Appendix A, and the accepted demonstrations are listed in Appendix B.

Two types of competitions were part of the track: regular and live ones. Regular competitions were those that were run during a period of time before the competition session at NeurIPS 2019 took place. Most of those posed a machine learning problem and released associated resources at the beginning of the challenge, then participants should provide solutions to the problems via a challenge platform (e.g., CodaLab[1] or AICrowd[2]). The best solutions, as determined by the organizers, were eligible to win prizes for the different competitions. Live competitions, on the other hand, were run on-site during

* Part of this work was done while H.J. Escalante was invited professor at the Computer Science Department at CINVESTAV-IPN, Zacatenco, Ciudad de México, Mexico

1. https://codalab.lri.fr/
2. https://www.aicrowd.com/

the competition session at the NeurIPS conference. Preliminary qualification phases were common to select a limited number of participants that could participate in the challenges.

Demonstrations offer a complementary dimension to the competitions. Whereas competitions are defined by the notion of controlled and consistent evaluation of a method against a strict success metric, demonstrations provide insight and understanding into the proposed approach, and invite evaluation on the basis of interactive or subjective experiences. Research areas which can be presented successfully via demonstration include robotics, interactive domains such as collaborative music or dialogue, generative models, visualization and understanding of machine learning models, competitive game-playing agents, and neuromorphic hardware. All of these categories were represented in the 28 accepted demonstrations at NeurIPS 2019.

The competition track attracted more than 4,000 participants and received a total of around 15,000 submissions for the different challenges. Demonstrations were very well attended during the poster sessions of the NeurIPS conference. The success of the track has motivated this compilation that aims at providing the reader with a glimpse of outstanding contributions from organizers and participants of events that were part of the track.

For compiling this volume we launched a call for papers targeting organizers and participants of the Demonstration and Competition track. All of the submissions were subject to a strict reviewing process that resulted in the acceptance of 22 papers covering a wide diversity of topics. Table 1 provides an overview of the submissions associated to this volume. There is a good balance among the submission types: three papers associated to demonstrations were accepted; ten papers from participants reporting their solutions to a competition are part of the volume; nine papers describing the design and analysis of challenges were accepted; and, last but not least, there were fourteen papers associated to regular competitions and five accepted submissions associated to live ones.

The compilation is a snapshot of what happened in terms of Competitions and Demonstrations at NeurIPS. We are certain this volume will push further research in different fronts, inspiring researchers and practitioners to approach the problems associated to challenges in novel ways, as reference for proposing and organizing competitions and for knowing what makes a successful demonstration. We hope you enjoy this volume and that you consider joining the Competitions and Demonstration track at NeurIPS in future editions.

Acknowledgments

We are grateful with the program committee for their valuable support. Adrian P Lopez-Monroy (Centro de Investigacion en Matematicas AC); Zhengying Liu (Inria); Katja Hofmann (Microsoft Research); Pablo Hernandez-Leal (Borealis AI); Luis Pellegrin (UABC); Steven Bohez (DeepMind); Jose Francisco Martínez Trinidad (INAOE); Mauricio Gonzalez-Soto (INAOE); José Martínez-Carranza (INAOE); Cinjon Resnick (NYU); Lino Rodriguez Coayahuitl (INAOE); Heriberto Cuayahuitl (University of Lincoln); Laura Montoya (Accel AI); Martina Zambelli (Google DeepMind); Emilio Parisotto (CMU); Piotr Mirowski (DeepMind); Jorge De la Calleja (Universidad Politécnica de Puebla); Meysam Madadi (CVC); Sergio Escalera (CVC and University of Barcelona); Reinier Oves Garcia (INAOE); Esaú Villatoro-Tello (IDIAP); Jake Bruce (DeepMind);

Table 1: Overview of submissions to the volume.

Ref.	**S.T.**	**Associated event**	**Type**
(Kim et al., 2020b)	P	MicroNet challenge	R
(Weichwald et al., 2020)	P	Causality for Climate (C4C)	R
(Kim et al., 2020a)	P	Game of Drones	L
(Herruzo and Larriba-Pey, 2020)	P	Traffic4cast	R
(Kanervisto et al., 2020)	P	MineRL	R
(Scheller et al., 2020)	P	MineRL	R
(O'Kelly et al., 2020)	D	Demonstration	NA
(Ölsner and Milz, 2020)	P	Game of Drones	L
(Shin et al., 2020)	P	Game of Drones	L
(Runge et al., 2020)	O	Causality for Climate (C4C)	R
(Gardner et al., 2020)	O	Reconnaissance Blind Chess	R
(Yalçin et al., 2020)	D	Demonstration	NA
(Cartoni et al., 2020)	O	Robot open-Ended Autonomous Learning (REAL)	R
(Martin et al., 2020)	P	Traffic4cast	R
(Crosby et al., 2020)	O	The Animal-AI Olympics	R
(Madaan et al., 2020)	O	Game of Drones	L
(Herrmann, 2020)	D	Demonstration	NA
(Milani et al., 2020)	O	MineRL	R
(Yan et al., 2020)	P	MicroNet challenge	R
(Kreil et al., 2020)	O	Traffic4cast	R
(Liu et al., 2020)	O	AutoDL	R
(Remy and Ben, 2020)	O	Live Malaria Challenge	L

References

Emilio Cartoni, Francesco Mannella, Vieri Giuliano Santucci, Jochen Triesch, Elmar Rueckert, and Gianluca Baldassarre. Real-2019: Robot open-ended autonomous learning competition. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 142–152, 2020.

Matthew Crosby, Benjamin Beyret, Murray Shanahan, José Hernández-Orallo, Lucy Cheke, and Marta Halina. The animal-ai testbed and competition. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 164–176, 2020.

Ryan W. Gardner, Corey Lowman, Casey Richardson, Ashley J. Llorens, Jared Markowitz, Nathan Drenkow, Andrew Newman, Gregory Clark, Gino Perrotta, Robert Perrotta, Timothy Highley, Vlad Shcherbina, William Bernadoni, Mark Jordan, and Asen Asenov. The first international competition in machine reconnaissance blind chess. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 121–130, 2020.

Vincent Herrmann. Visualizing and sonifying how an artificial ear hears music. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 192–202, 2020.

Pedro Herruzo and Josep L. Larriba-Pey. Recurrent autoencoder with skip connections and exogenous variables for traffic forecasting. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 47–55, 2020.

Anssi Kanervisto, Janne Karttunen, and Ville Hautamäki. Playing minecraft with behavioural cloning. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 56–66, 2020.

Donghwi Kim, Hyunjee Ryu, Jedsadakorn Yonchorhor, and David Hyunchul Shim. A deep-learning-aided automatic vision-based control approach for autonomous drone racing in game of drones competition. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 37–46, 2020a.

Taehyeon Kim, Jonghyup Kim, and Seyoung Yun. Efficient model for image classification with regularization tricks. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 13–26, 2020b.

David P Kreil, Michael K Kopp, David Jonietz, Moritz Neun, Aleksandra Gruca, Pedro Herruzo, Henry Martin, Ali Soleymani, and Sepp Hochreiter. The surprising efficiency of framing geo-spatial time series forecasting as a video prediction task – insights from the iarai $\widehat{4}$c competition at neurips 2019. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 232–241, 2020.

Zhengying Liu, Zhen Xu, Shangeth Rajaa, Meysam Madadi, Julio C. S. Jacques Junior, Sergio Escalera, Adrien Pavao, Sebastien Treguer, Wei-Wei Tu, and Isabelle Guyon. Towards automated deep learning: Analysis of the autodl challenge series 2019. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 242–252, 2020.

Ratnesh Madaan, Nicholas Gyde, Sai Vemprala, Matthew Brown, Keiko Nagami, Tim Taubner, Eric Cristofalo, Davide Scaramuzza, Mac Schwager, and Ashish Kapoor. Airsim drone racing lab. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 177–191, 2020.

Henry Martin, Dominik Bucher, Ye Hong, René Buffat, Christian Rupprecht, and Martin Rauba. Graph-resnets for short-term traffic forecasts in almost unknown cities. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 153–163, 2020.

Stephanie Milani, Nicholay Topin, Brandon Houghton, William H. Guss, Sharada P. Mohanty, Keisuke Nakata, Oriol Vinyals, and Noboru Sean Kuno. Retrospective analysis of the 2019 minerl competition on sample efficient reinforcement learning. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 203–214, 2020.

Matthew O'Kelly, Hongrui Zheng, Dhruv Karthik, and Rahul Mangharam. textscF1TENTH: An open-source evaluation environment for continuous control and reinforcement learning. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 77–89, 2020.

Florian Ölsner and Stefan Milz. Catch me, if you can! a mediated perception approach towards fully autonomous drone racing. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 90–99, 2020.

Sekou L. Remy and Oliver Ben. A global health gym environment for rl applications. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 253–261, 2020.

Jakob Runge, Xavier-Andoni Tibau, Matthias Bruhns, Jordi Muñoz-Marí, and Gustau Camps-Valls. The causality for climate competition. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 110–120, 2020.

Christian Scheller, Yanick Schraner, and Manfred Vogel. Sample efficient reinforcement learning through learning from demonstrations in minecraft. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 67–76, 2020.

Sangyun Shin, Yongwon Kang, and Yong-Guk Kim. Evolution algorithm and online learning for racing drone. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 100–109, 2020.

Sebastian Weichwald, Martin E Jakobsen, Phillip B Mogensen, Lasse Petersen, Nikolaj Thams, and Gherardo Varando. Causal structure learning from time series: Large regression coefficients may predict causal links better in practice than small p-values. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 27–36, 2020.

Özge Nilay Yalçin, Nouf Abukhodair, and Steve DiPaola. Empathic ai painter: A computational creativity system with embodied conversational interaction. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 131–141, 2020.

Zhongxia Yan, Hanrui Wang, Demi Guo, and Song Han. Micronet for efficient language modeling. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track*, pages 215–231, 2020.

Appendix A. Accepted competitions

This section describes the competitions accepted for the NeurIPS 2019 track. Tables 2, 3 and 4 show the accepted regular competitions. While Table 5 describes the accepted live competitions.

Table 2: Regular competitions, part. 1

Competition	Brief summary
Causality for Climate (C4C)	This competition comprised a number of multivariate time series datasets featuring major challenges of climate data from time delays and nonlinearity to nonstationarity and selection bias. The competition aimed to open up new interdisciplinary research pathways by improving our scientific understanding of Earth's climate, while also driving method development and benchmarking in the computer science community.
Reconnaissance Blind Chess	Build the best AI bot to play reconnaissance blind chess, a challenge for making optimal decisions in the face of uncertainty. Reconnaissance blind chess is like chess except a player does not know where her opponent's pieces are a priori. Rather, she can covertly sense a chosen 3x3 square of the board each turn and also learn partial information from captures.
Automated Deep Learning (AutoDL)	The AutoDL challenge aimed taking the automate the design of deep learning (DL) methods to solve generic tasks. This was a challenge with "code submission": machine learning algorithms were trained and tested on a challenge platform on data invisible to the participants. Targetted applications were speech, image, video, and text, for which DL methods have had great success recently, to drive the community to work on automating the design of DL models. Raw data was be provided, formatted in a uniform tensor manner, to encourage participants to submit generic algorithms. Restrictions were imposed on training time and resources to push the state-of-the-art further.
3D Object Detection over HD Maps for Autonomous Cars	Autonomous cars are expected to dramatically redefine the future of transportation. The 3D Perception system of the autonomous car is a critical keystone upon which high level autonomy functions depend. This competition is designed to help advance the state of the art in 3D object detection by focusing research on this topic in the context of autonomous cars, specifically by sharing the full modality of sensor data available to typical autonomous cars, and by providing access to a high fidelity HD map.

Table 3: Regular competitions, part. 2

Competition	Brief summary
MicroNet: Large-Scale Model Compression Competition	Contestants competed to build the most efficient models that solve ImageNet classification, CIFAR-100 classification, or WikiText-103 language modeling. The competition was focused on efficient inference, and used a theoretical metric rather than measured inference speed to score entries.
The Animal-AI Olympics	The Animal-AI Olympics translates tasks from animal cognition into tests for AI. Organizers provided a fully configurable 3D environment in which they built 100 hidden food-retrieval tasks inspired by work in comparative cognition. Participants only knew ten categories under testing along with the objects included in the tests and they had to submit an agent capable of robust food-retrieval behaviour from pixel inputs alone. Part of the challenge was developing sensible environment configurations as well as mimicking animal-like cognitive abilities.
Traffic4cast – Traffic Map Movie Forecasting	Predict high resolution traffic flow volume, heading, and speed on a whole city map looking 15 minutes into the future! Kicking off a series of annual competitions, this year's data was based on 100 billion probe points from 3 cities mapped in 5 minute intervals, showing trends across weekdays and seasonal effects. Improved traffic predictions are of great social, environmental, and economic value, while also advancing our general ability to capture the simple implicit rules underlying a complex system and model its future states.
Robot open-Ended Autonomous Learning (REAL)	Open-ended learning aims to build learning machines and robots that are able to acquire skills and knowledge in an incremental fashion in a certain environment. This competition addressed autonomous open-ended learning with a focus on simulated robot systems that: (a) acquire a sensorimotor competence that allows them to interact with objects and physical environments; (b) learn in a fully autonomous way, i.e. with no human intervention (e.g., no tasks or reward functions) on the basis of mechanisms such as curiosity, intrinsic motivations, task-free reinforcement learning, self-generated goals, and any other mechanism that might support autonomous learning.
Learn to Move: Walk Around	A competition to develop human-level versatile locomotion controllers, which is a grand challenge in biomechanics, neuroscience, and robotics. The main task is to develop a controller for a 3D human musculoskeletal simulation model to walk or run following velocity commands.

Table 4: Regular competitions, part. 3

Competition	Brief summary
CellSignal: Disentangling biological signal from experimental noise in cellular images	The task was to correctly classify images of populations of human cells as exhibiting one of 1,108 different treatments in the dataset RxRx1. RxRx1 consists of 125,514 high-resolution 512x512 6-channel fluorescence microscopy images of cells under these treatments across 4 cell types and 51 different runs of the same experimental design. These images exhibit technical nuisance factors specific to each experiment called batch effects that confound the classification task. The data was split into training and test sets by experiment, so good classifiers needed to separate relevant biological factors from the batch effect factors in order to generalize outside of the training data.
Advancing State-of-the-art Learning Approaches for Disentangled Representations	Learning deep representations in which different semantic aspects of data are structurally disentangled is of central importance for advancing artificial intelligence. Despite the clear necessity and benefits of disentangled representations, recent benchmarks on simulated datasets have exposed severe limitations of state-of-the-art approaches. In this challenge the following will be addressed: (i) unsupervised learning with unsupervised model selection (ii) role of supervision and (iii) impact and application of disentanglement on real-world images.
MineRL Sample Efficient RL from Human Priors	A challenge requiring participants to develop sample efficient reinforcment and imitation learning algorithms to solve a complex task in Minecraft, a rich open-world environment featuring sparse-rewards, embodied multi-agent interactions, long-term planning, vision, navigation, and explicit and implicit sub-task hierarchies. The competition featured two components: 1) the ObtainDiamond task, a sequential decision making environment requiring an agent to collect a necessary set of requisite items, explore, and mine diamonds only using observations from its first-person perspective; and (2) the MineRL-v0 dataset, a large-scale collection of over 60 million state-action pairs of human demonstrations that can be resimulated into embodied agent trajectories with arbitrary modifications to game state and visuals. Participants had to compete to develop systems which solve the ObtainDiamond task with a limited number of samples (4-days worth) from a new Minecraft simulator called MineRLEnv which modifies the Malmo simulator to be synchronous and extremely efficient. Submissions were evaluated by being trained and then run from scratch by the competition organizers in a fixed cloud-computing in environment to ensure that truly sample-efficient algorithms are developed.

Table 5: Brief description of live competitions

Competition	Brief summary
Game of Drones	Game of Drones is a multi-drone racing tournament conducted in the high-fidelity simulation environment AirSim. Participants will have the choice of three tiers: Planning only, Perception only, or Full Autonomous Racing. The aim is to combine challenges from adversarial planning and real-time perception and to encourage fusing learning- and model-based approaches.
Live Malaria Challenge	In the NeurIPS Live Malaria Challenge participants are required to apply machine learning tools to determine novel solutions which could impact malaria policy in Sub Saharan Africa.
Pommerman Year 2: Radio	Pommerman: Train a team of communicative agents to play Bomberman in a partially observed setting. Compete against other teams.
AI Driving Olympics 3	The third edition of the AI Driving Olympics was designed to probe the state of the art in all areas of autonomous vehicles. There were three associated events: 1) Urban - based on the duckietown platform, 2) Racing - inspired by the AWS DeepRacer platform, and 3) Advanced sensing - using the nuScenes dataset.

Appendix B. Accepted demonstrations

This section lists (in Tables 6 and 7) all 28 demonstrations accepted for the NeurIPS 2019 track.

Table 6: Demonstrations

Demonstration Title and Authors	**Category**
exBERT: A Visual Analysis Tool to Explain BERT's Learned Representations Ben Hoover · Hendrik Strobelt · Sebastian Gehrmann	Visualization
Streamlit, a new app framework for machine learning tools Adrien Treuille · Amanda Kelly	Tools for ML
"How can this Paper get in?" - A game to advise researchers when writing for a top AI conference Aabhas Sharma · Narendra Nath Joshi · Michael Muller · Casey Dugan	Human-in-the-loop learning
Toronto Annotation Suite Amlan Kar · Sanja Fidler · Jun Gao · Seung Wook Kim · Huan Ling	Tools for ML
Robot-Assisted Hair-Brushing Eura Shin · Hejia Zhang · Rey J Pocius · Nathaniel Dennler · Heather Culbertson · Naghmeh Zamani · Stefanos Nikolaidis	Robotics
Learning Machines can Curl - Adaptive Deep Reinforcement Learning enables the robot Curly to win against human players in an icy world Dong-Ok Won · Sang-Hoon Lee · Klaus-Robert Müller · Seong-Whan Lee	Robotics
Human Gesture Recognition using Spiking Input on Akida Neuromorphic Platform Sounak Dey · Arijit Mukherjee · Gilles BEZARD · Douglas McLelland	Neuromorphic HW
GENO – Optimization for Classical Machine Learning Made Fast and Easy Soeren Laue · Matthias Mitterreiter · Joachim Giesen	Tools for ML
SCC: Deep Reinforcement Learning Agent plays StarCraft II at competitive human level XJ Wang · Peng Peng	Reinforcement Learning
AI in Two-sided Ride-sharing Marketplace Zhiwei Qin · Shikai Luo · lingyu zhang · yan jiao · Xiaocheng Tang · Lulu Zhang · hongtu zhu · Jieping Ye	Reinforcement Learning
NNgen: A Model-Specific Hardware Synthesis Compiler for Deep Neural Network Shinya Takamaeda-Yamazaki · Shinya Fujisawa · Shuichi Fujisaki	HW for ML
Realtime Modeling and Anomaly Detection in Multivariate Data Streams Christopher Hannon · Andrey Lokhov · Deep Deka	Data Modeling and Analysis
Empathy based Affective Portrait Painter Steve DiPaola · Ozge Nilay YALCIN · Nouf Abukhodair	Generative models
Melody Slot Machine Masatoshi Hamanaka	Music generation

Table 7: Demonstrations, continued

Demonstration Title and Authors	**Category**
Smart Home Appliances: Chat with your Fridge Denis Gudovskiy · Alec Hodgkinson · Stefano Alletto · Luca Rigazio	Classification
Project BB: Bringing AI to the Command Line Tathagata Chakraborti · Mayank Agarwal	Tools for ML
Immersions - How Does Music Sound to Artificial Ears? Vincent Herrmann	Generative models
The Option Keyboard: Combining Skills in Reinforcement Learning Daniel Toyama · Shaobo Hou · Gheorghe Comanici · Andre Barreto · Doina Precup · Shibl Mourad · Eser Aygün · Philippe Hamel	Reinforcement learning
AllenNLP Interpret: Explaining Predictions of NLP Models Jens Tuyls · Eric Wallace · Matt Gardner · Junlin Wang · Sameer Singh · Sanjay Subramanian	Analysis, visualization
Passcode: A cooperative word guessing game between a human and AI agent Katy Gero · Maria Ruis · Zahra Ashktorab · J Johnson · Sadhana Kumaravel · Wei Zhang · Qian Pan · Murray Campbell · Casey Dugan · David Millen · Sarah Miller · Werner Geyer	Human-in-the-loop learning
Deep Space-Time Prior for Realtime Mobile Novel View Synthesis Zainul Shah	Computer vision
AIDEme: An active learning based system for interactive exploration of large datasets Enhui Huang · Luciano Di Palma · Laurent Cetinsoy · Yanlei Diao · Anna Liu	Data visualization
BIM-GAN: a sketch to layout, 3D, and VR tool for architectural floor plan design Chin-Yi Cheng	Generative models
One-on-one fitness training with an AI avatar Roland Memisevic · Guillaume Berger · Tippi Puar · David Greenberg	Interactive assistant
Real Time CFD simulations with 3D Mesh Convolutional Networks Pierre Baque · Pascal Fua · François Fleuret	Modeling and visualization
F1/10: An open-source 1/10th scale platform for autonomous racing and reinforcement learning Matthew O'Kelly · Dhruv Karthik · Hongrui Zheng · Joseph Auckley · Siddharth Singh · Shashank D Prasad · Kim Luong · Matthew R Lebermann · Rahul Mangharam	Robotics