

Posters

Click link to see abstracts and links to papers

Poster Room 1

[Analytically tractable inference in neural networks - An alternative to backpropagation](#)

James-A. Goulet (Polytechnique Montreal)*; Luong-Ha Nguyen (Polytechnique Montreal)

[Q-Layer: Latent Space Constraints for Robust Convolutional Neural Network](#)

Sirui Song (McGill University)*; Derek C Yang (UC San Diego); Xi Chen (McGill University); Xue Liu (McGill University); Tong Zhang (Hong Kong University of Science and Technology)

[Adaptive Whitening and Coloring Style Injection in GANs](#)

Antoine Dufour (Université Laval); Jean-Francois Lalonde (Université Laval)*

[Mixture-based Feature Space Learning for Few-shot Image Classification](#)

Arman Afrasiyabi (Université Laval)*; Jean-Francois Lalonde (Université Laval); Christian Gagné (Université Laval)

[Precision neurostimulation medicine enabled by autonomous learning algorithms](#)

Marco Bonizzato (Université de Montréal)*; Sandrine Côté (Université de Montréal); Elena Massai (Université de Montréal); Rose Guay Hottin (Université de Montréal); Samuel Laferriere (MILA); Stephan Quessy (Université de Montréal); Guillaume Lajoie (Mila, Université de Montréal); Marina Martinez (Université de Montréal); Numa Dancause (Université de Montréal)

[VPTR: Video prediction Transformer](#)

Xi Ye (Polytechnique Montreal)*; Guillaume-Alexandre Bilodeau (Polytechnique Montréal)

[Fairness-aware Item Recommendation in Multi-stakeholder Marketplaces via Multi-Objective Optimization](#)

Haolun Wu (McGill University)*; Chen Ma (City University of Hong Kong); Bhaskar Mitra (Microsoft); Fernando Diaz (Google); Xue Liu (McGill University)

[OptInter: Learning Optimal Feature Interaction Methods for CTR Prediction](#)

Fuyuan Lyu (McGill University)*; Xing Tang (Huawei Noah's Ark Lab); Huifeng Guo (Huawei Noah's Ark Lab); Ruiming Tang (Huawei Noah's Ark Lab); Xiuqiang He (Huawei Noah's Ark Lab); Rui Zhang (Tsinghua University Graduate School at Shenzhen); Xue Liu (McGill University)

[Aggregating From Multiple Target-Shifted Sources](#)

Changjian Shui (Université Laval)*; Zijian Li (Guangdong University of Technology); Jiaqi Li (University of Western Ontario); Charles X Ling (Western University); Christian Gagné (Université Laval); Boyu Wang (University of Western Ontario)

[Prediction of Individual Subject's Fiber Response in Vagus Nerve Stimulation with Meta Learning](#)

Ximeng Mao (Mila, UdeM)*; Yao-Chuan Chang (Feinstein Institutes); Theodoros Zanos (Feinstein Institutes); Stavros Zanos (Feinstein Institutes); Guillaume Lajoie (Mila, Université de Montréal)

[AFB: Improving Communication Load Forecasting Accuracy with Adaptive Feature Boosting](#)

Chengming Hu (Samsung AI Center)*; Xi Chen (Samsung); Ju Wang (Samsung AI Center); Xue Liu (Samsung AI Center); Gregory Dudek (McGill University)

[Multi-Object Tracking and Segmentation with a Space-Time Memory Network](#)

Mehdi Miah (Polytechnique Montréal)*; Guillaume-Alexandre Bilodeau (Polytechnique Montréal); Nicolas Saunier (Polytechnique Montréal)

[Deep Reinforcement Learning at the Edge of the Statistical Precipice](#)

Rishabh Agarwal (Google Research, Brain Team)*; Max Schwarzer (Mila); Pablo Samuel Castro (Google); Aaron Courville (MILA, Université de Montréal); Marc G. Bellemare (Google Brain)

[Pretraining Representations for Data-Efficient Reinforcement Learning](#)

Max Schwarzer (Mila, Université de Montréal)*; Nitarshan Rajkumar (Mila, Université de Montréal); Michael Noukhovitch (Mila, Université de Montréal); Ankesh Anand (Mila, Université de Montréal); Laurent Charlin (HEC Montreal and Mila); R Devon Hjelm (Microsoft Research); Philip Bachman (Microsoft Research); Aaron Courville (MILA, Université de Montréal)

[MLCR: Meta-Learning Communication Rules\ for Multi-Agent Reinforcement Learning](#)

Jikun Kang (McGill University)*; Jie Fu (University of Montreal, Mila); Xue Liu (McGill University)

[Epidemiological compartmental models and machine learning: importance of the model's structure](#)

Vincent Dandenault (Université de Montréal)*; Sofia Alfonso (CHU Sainte-Justine Research Center); Alexandre Simard (CHU Sainte-Justine Research Center); Masoumeh Sajedi (CHU Sainte-Justine Research Center); Benoît Mâsse (School of Public Health - Université de Montréal); Marie-Claude Boily (School of Public Health - Imperial College London); Simon de Montigny (School of Public Health - Université de Montréal)

[Data augmentation instead of explicit regularization](#)

Alex Hernandez-Garcia (Mila - Quebec AI Institute)*

[CARL: Conditional value-at-risk Adversarial Reinforcement Learning](#)

Mathieu Godbout (Université Laval)*; Maxime Heuillet (Universite Laval); Audrey Durand (Université Laval)

[Learning Aggregations of Binary Activated Neural Networks with Probabilities over Representations](#)

Louis Fortier-Dubois (Université Laval)*; Benjamain Leblanc (Université Laval); Gaël Letarte (Université Laval); Pascal Germain (Université Laval); François Laviolette (Université Laval)

[Sequential Pipeline Optimization: Bandits-driven Exploration using a Collaborative Filtering Representation](#)

Maxime Heuillet (Universite Laval)*; Audrey Durand (Université Laval); Benoit Debaque (Thales Digital Solutions)

[Quantum machine learning for text classification](#)

Jean Frederic Laprade (Universite de Sherbrooke)*; Sarah Blanchette (Universite de Sherbrooke); Zachary Zanussi (Statistics Canada); Omar Chikhar (Statistics Canada); Vladimir Skavysh (Bank of Canada)

[RandomSCM: interpretable ensembles of sparse classifiers tailored for omics data](#)

Thibaud Godon (Laval University)*; Pier-Luc Plante (Laval University); Élina Francovic-Fontaine (Université Laval); Baptiste Bauvin (Université Laval); Alexandre Drouin (ServiceNow); Jacques Corbeil (Université Laval); François Laviolette (Université Laval)

[Learning Function From Structure In Neuromorphic Networks](#)

Laura E Suarez (McGill University)*; Bratislav Misic (McGill University); Guillaume Lajoie (Mila, Université de Montréal); Blake Richards (McGill University)

[A deep learning end-to-end system for vision-based fault recognition in thermal and civil inspection of the underground distribution network](#)

Francois Miralles (Hydro-Québec); Luc Cauchon (Hydro-Québec); Arnaud Zinflou (Hydro-Québec)*; Mouhamadou Makhtar Dione (Hydro-Québec); Gregoire Francois (Hydro-Québec); Magnan Marc-Andre (Hydro-Québec); Jinga Livius (Hydro-Québec); Dafali Salwa (École polytechnique de Montréal,)

[Information Sharing in mHealth](#)

Antoine Gariépy (Université Laval)*; Audrey Durand (Université Laval)

[Curating the Twitter Election Integrity Datasets for Better Online Troll Characterization](#)

Albert M Orozco Camacho (Mila - Québec AI Institute / McGill University)*

Poster Room 2

[Multimodal Transformer for Emotion Recognition and Sentiment Analysis](#)

Rihab Hajlaoui (Polytechnique Montréal)*; Guillaume-Alexandre Bilodeau (Polytechnique Montréal); Jan Rockemann (Airudi)

[Exploring the loss landscape of neural networks with manifold learning and topological data analysis](#)

Stefan Horoi (Université de Montréal)*; Jessie Huang (Yale University); Bastian A Rieck (Institute of AI for Health, Helmholtz Centre Munich); Guillaume Lajoie (Mila, Université de Montréal); Guy Wolf (Université de Montréal); Smita Krishnaswamy (Yale University)

[Typing assumptions improve identification in causal discovery](#)

Philippe Brouillard (Mila, Université de Montréal)*; Perouz Taslakian (Element AI); Alexandre Lacoste (ServiceNow); Sebastien Lachapelle (Mila, Université de Montréal); Alexandre Drouin (ServiceNow)

[Task-Assisted GAN for resolution enhancement, quality improvement and modality translation in fluorescence microscopy](#)

Catherine Bouchard (Université Laval)*; Theresa Wiesner (Université Laval); Andréanne Deschênes (Université Laval); Flavie Lavoie-Cardinal (Université Laval); Christian Gagné (Université Laval)

[A Vision-Based Automatic Transcription of Guitar Music From RGBD Videos](#)

Mark Asmar (Polytechnique Montreal)*; Lama Seoud (Polytechnique Montreal); Guillaume-Alexandre Bilodeau (Polytechnique Montréal)

[Parallel Gaussian Process-Based Bayesian Optimization for Multiple Cortical Array Neuroprosthetic Control](#)

Julien Rimok (Université de Montréal)*; Numa Dancause (Université de Montréal); Guillaume Lajoie (Mila, Université de Montréal)

[Learning A Hierarchy of Behaviours though A Unified Reward](#)

Hanqing Zhao (McGill University)*; Gregory Dudek (McGill University); Xue Liu (McGill University)

[Gradient Boosted Tree Classification vs. Regression for Highly Imbalanced Classification Problems](#)

Aristides T Milios (McGill)*

[Machine Learning for Optimising Patient Follow-up for Cancer Rehabilitation \(A CanRehab Study\)](#)

Rupali Bhati (Université Laval)*; Jennifer Jones (University of Toronto); Audrey Durand (Université Laval)

[HOLISTIC GUIDANCE FOR OCCLUDED PERSON RE-IDENTIFICATION](#)

Madhu Kiran (ETS Montreal)*; Gnana Praveen Rajasekhar (Ecole Technologie Superieure); Le Thanh Nguyen-Meidine (ETS Montreal); Soufiane Belharbi (ÉTS Montreal); Louis-Antoine Blais-Morin (Genetec Inc.); Eric Granger (ETS Montreal)

[On the Limits of Algorithmic Counting and Out-of-Distribution Generalization](#)

Martin Weiss (Mila)*

[Cogment: Open Source Framework For Distributed Multi-actor Training, Deployment And Operations](#)

Sai Krishna Gottipati (AI-Redefined)*; Sagar Kurandwad (AI-r); Jonathan M Fisher (AI Redefined); Vincent Robert (AI-r); Gregory Szriftgiser (AI Redefined); Clodéric Mars (AI Redefined)

[Dynamic Consolidation for Continual Learning](#)

Hang Li (McGill University)*; Xue Liu (McGill University)

[Deliberation gated by opportunity cost adapts to context with urgency](#)

Maximilian Puelma Touzel (Mila)*; Paul Cisek (UdeM); Guillaume Lajoie (Mila, Université de Montréal)

[Sample Boosting Algorithm - SamBA - Greedy local sparse votes and their applications to bioinformatics](#)

Baptiste Bauvin (Université Laval)*; Cécile Capponi (Aix-Marseille University); François Laviolette (Université Laval)

[A Thompson sampling approach to learn the Whittle index in restless bandit problems](#)

Nima Akbarzadeh (McGill University)*; Aditya Mahajan (McGill University)

[Learning Bike Origins and Destinations from Counts and Trajectories](#)

Ouassim Manout (Polytechnique Montreal / ENTPE); Qingwu Liu (Polytechnique Montreal); Francesco Ciari (Polytechnique Montreal); Lijun Sun (McGill University); Catherine Morency (Polytechnique Montreal); Martin Trepanier (École Polytechnique de Montréal); Nicolas Saunier (Polytechnique Montreal)*

[Towards an inclusive AI: Research on the revitalization and preservation of indigenous languages](#)

Tan Le (Universite du Quebec a Montreal)*; fatiha sadat (UQAM)

[SESNO: Sample Efficient Social Navigation from Observation](#)

Bobak H Baghi (Samsung Electronics)*

[Learn to Perturb: A Self-Learning Hyper-Heuristic for the Simultaneous Stochastic Optimization of Mining Complexes](#)

Yassine Yaakoubi (Mila, McGill University)*; Roussos Dimitrakopoulos (McGill University)

[Can we trust the Wizard 🧙? Interpreting Knowledge-Grounded Dialogue Systems.](#)

Yuvraj Prabhakar (Concordia University)*; Mehar Bhatia (NeuralSpace Research)

[Improving Fairness in Heterogeneous Federated Learning](#)

Shaoxiang Qin (McGill University)*; jiachen lei (Zhejiang University); Hillary Tao (McGill University); Xi Chen (McGill University); Qiao Xiang (Yale University); Zhongjie Ba (Zhejiang University); Xue Liu (McGill University)

[Learning To Optimize Iterative Optimizers](#)

Hugo S Gomes (Laval University)*; Benjamin Léger (Université Laval, Institut Intelligence et Données); Christian Gagné (Université Laval)

[Object-centric language emergence with Transformers](#)

Tom Bosc (Mila)*; Pascal Vincent (MILA, FAIR)

[Improving Continuous Normalizing Flows using a Multi-Resolution Framework](#)

Vikram Voleti (Mila, University of Montreal)*

Poster Room 1

Analytically tractable inference in neural networks - An alternative to backpropagation

James-A. Goulet (Polytechnique Montreal)*; Luong-Ha Nguyen (Polytechnique Montreal)

Location: Poster Room 1

ABSTRACT: Until to now, neural networks have been predominantly relying on backpropagation and gradient descent as the inference engine in order to learn the neural network's parameters. This is primarily because closed-form Bayesian inference for neural networks has been considered to be intractable. This presentation will outline a new analytical method for performing tractable approximate Gaussian inference (TAGI) in BNNs. The method enables the analytical inference of the posterior mean vector and diagonal covariance matrix for weights and biases. One key aspect is that the method matches or exceeds the state-of-the-art performance while having the same computational complexity as current methods relying on the gradient backpropagation.

Performing Bayesian inference in neural networks enables several key features, such as the quantification of epistemic uncertainty associated with model parameters, the online estimation of parameters, and a reduction in the number of hyperparameters due to the absence of gradient-based optimization. Moreover, the analytical framework proposed also enables unprecedented features such as the propagation of uncertainty from the input of a network up to its output, and it allows inferring the value of hidden states, inputs, as well as latent variables. The first part of the presentation will cover the theoretical foundation and working principles of the analytically tractable uncertainty propagation in neural networks, as well as the parameter and hidden state inference. Then, the second part will go through several benchmarks demonstrating the superiority of the approach on feedforward, convolutional, recurrent, and generative neural network architectures. In addition, the presentation will showcase how TAGI can be applied to reinforcement learning problems such as the Atari game environment. Finally, the last part will present how we can leverage the analytic inference capabilities to enable novel applications of neural networks.

[PDF of paper](#)

[Return to top](#)

Q-Layer: Latent Space Constraints for Robust Convolutional Neural Network

Sirui Song (McGill University)*; Derek C Yang (UC San Diego); Xi Chen (McGill University);
Xue Liu (McGill University); Tong Zhang (Hong Kong University of Science and Technology)

Location: Poster Room 1

ABSTRACT: Carefully crafted adversarial attacks pose a major challenge to the robustness of Convolutional Neural Networks (CNNs), and yet have little impact on humans. This inspires us to imitate human's cognitive process, where common semantic attributes among objects are extracted and memorized for recognition. Specifically, we explore the potential of robust CNN with two new latent-space constraints. A soft constraint is applied to let data points in the latent space to form concentrated semantic clusters, making it hard for attacks to distort. A quantization, or hard constraint is also enforced on the latent space, where data points are represented by their cluster centroids, nullifying most small-scale attacks. We further design a multi-head quantization method to compensate for the information loss during quantization. Moreover, a Quantization Layer (Q-Layer) is implemented to combine all the above new designs, and can be seamlessly integrated into any neural network structures. Extensive experiments on different datasets and CNN architectures show that Q-Layer largely improves the robustness of CNN against different kinds of adversarial attacks.

[PDF of paper](#)

[Return to top](#)

Adaptive Whitening and Coloring Style Injection in GANs

Antoine Dufour (Université Laval); Jean-Francois Lalonde (Université Laval)*

Location: Poster Room 1

ABSTRACT: In this work, we replace AdaIN by an explicit operation of whitening and coloring for style injection in GANs. This operation, dubbed "Adaptive Whitening and Coloring Style Injection" (AWCSI) can be viewed as a generalization of the existing AdaIN method. More specifically, we introduce a module that can be used as a replacement for the AdaIN blocks in existing popular GAN architectures and we present its impact on generation tasks. Indeed, in the conditional image generation tasks, where the latent space is intended to represent the style of the images, we find that whitening helps ensure that the space encodes only stylistic information which allows the content of the input image to be more visible.

[PDF of paper](#)

[Return to top](#)

Mixture-based Feature Space Learning for Few-shot Image Classification

Arman Afrasiyabi (Université Laval)*; Jean-Francois Lalonde (Université Laval); Christian Gagné (Université Laval)

Location: Poster Room 1

ABSTRACT: We introduce Mixture-based Feature Space Learning (MixtFSL) for obtaining a rich and robust feature representation in the context of few-shot image classification. Our MixtFSL aims to learn a multimodal representation for the base classes using a mixture of trainable components that are iteratively refined during training. The idea is to learn both the representation and the mixture model jointly in an online manner, which effectively unites these two tasks by allowing the gradient to flow between them. Previous works have proposed to model each base class either with a single point or with a mixture model by relying on offline clustering algorithms. In contrast, we propose to model base classes with mixture models by simultaneously training the network and learning the mixture model parameters in an online manner. This results in a richer and more discriminative feature space that can be employed to classify novel examples from very few samples. Two main stages are proposed to train the MixtFSL model. First, the multimodal mixtures for each base class and the feature extractor parameters are learned using a combination of two loss functions. Second, the resulting network and mixture models are progressively refined by a leader-follower procedure, which uses the current estimate as a "target" network. This target network is used to make a consistent assignment of instances to mixture components, which increases performance and stabilizes training.

[PDF of paper](#)

[Return to top](#)

Precision neurostimulation medicine enabled by autonomous learning algorithms

Marco Bonizzato (Université de Montréal)*; Sandrine Côté (Université de Montréal); Elena Massai (Université de Montréal); Rose Guay Hottin (Université de Montréal); Samuel Laferriere (MILA); Stephan Quessy (Université de Montréal); Guillaume Lajoie (Mila, Université de Montréal); Marina Martinez (Université de Montréal); Numa Dancause (Université de Montréal)

Location: Poster Room 1

ABSTRACT: Neural stimulation is an established clinical intervention for a variety of indications. For instance, deep brain stimulation alleviates debilitating symptoms of Parkinson's Disease and spinal or nerve stimulation provides relief from chronic pain. Novel applications of brain and spinal stimulation have shown great promise in reversing paralysis and sensory deficits.

Development of next-generation neurostimulation medicine, featuring multipronged interventions and high-density electrode arrays, will require an algorithmic framework to handle generating complex neuroprosthetic stimulation patterns, which is currently lacking. A class of algorithms, Gaussian-Process (GP)-based Bayesian Optimization (BO), is ideally suited to solve this problem.

We here demonstrate real-time AI-based control of neurostimulation. Across diverse interventions and models, targeting brain and spinal cord, rat and non-human primates (NHP), we show that GP-BO quickly optimizes multi-channel neurostimulation. GP-BO rapidly determined stimulation strategies and improved walking after spinal cord injury (SCI). It allowed mapping the input parameters' space, while testing only a fraction of its combinations.

Furthermore, BO is suitable to naturally embed 'prior' expert/clinical knowledge and to transparently make 'posterior' learnt knowledge available to the experimenter, which are enticing features for a clinical decision support system.

Our framework could be extended to other neurostimulation treatments to maximize their therapeutic benefits. These results support broad establishment of learning agents in research and clinical scenarios. Our work represents a first important step towards the goal of therapeutic personalization and maximization of therapy effectiveness, with great promises for better treatment of neurological disorders. We propose that the introduction of automation as a structural part of neuroprosthetic design is translationally mature and clinically viable.

[PDF of paper](#)

[Return to top](#)

VPTR: Video prediction Transformer

Xi Ye (Polytechnique Montreal)*; Guillaume-Alexandre Bilodeau (Polytechnique Montréal)

Location: Poster Room 1

ABSTRACT: The state-of-the-art models for video future frames prediction are based on convolutional long short-term memory neural networks (Conv-LSTM). In this paper, we propose to apply Transformer to video

future frames prediction, which is promising to overcome some well-known drawbacks of recurrent neural networks and thus promote the performance of video future frames prediction. The proposed model and expected contributions of our method are briefly described.

[PDF of paper](#)

[Return to top](#)

Fairness-aware Item Recommendation in Multi-stakeholder Marketplaces via Multi-Objective Optimization

Haolun Wu (McGill University)*; Chen Ma (City University of Hong Kong); Bhaskar Mitra (Microsoft); Fernando Diaz (Google); Xue Liu (McGill University)

Location: Poster Room 1

ABSTRACT: Multi-stakeholder marketplaces are important components of many existing online services such as LinkedIn, Amazon, and Yelp, which have both consumers (i.e. users who seek content) and producers (i.e. users who provide content). Conventional recommendation systems in these platforms mainly focus on maximizing consumers' satisfaction by recommending the most relevant items based on the learned user preference, which may lead to unfair exposure of items, thus jeopardizing the benefits of producers. To tackle this problem, we propose a fairness-aware recommendation framework, Multi-FR, by using multi-objective optimization to adaptively balance the accuracy and fairness in a multi-stakeholder recommendation scenario. Our model adopts the multi-gradient descent to generate a Pareto set of solutions, where the most appropriate one is selected from the Pareto set. In addition, four fairness constraints are applied to make the recommendation results fair for each stakeholder. We extensively evaluate our model on three real-world datasets, comparing with SOTA fair recommendation methods and using various performance metrics. The experimental results demonstrate that Multi-FR can largely improve the recommendation fairness on both the consumer and producer side with little drop in recommendation quality, outperforming several state-of-the-art fairness-aware recommendation approaches.

[PDF of paper](#)

[Return to top](#)

OptInter: Learning Optimal Feature Interaction Methods for CTR Prediction

Fuyuan Lyu (McGill University)*; Xing Tang (Huawei Noah's Ark Lab); Huifeng Guo (Huawei Noah's Ark Lab); Ruiming Tang (Huawei Noah's Ark Lab); Xiuqiang He (Huawei Noah's Ark Lab); Rui Zhang (Tsinghua University Graduate School at Shenzhen); Xue Liu (McGill University)

Location: Poster Room 1

ABSTRACT: Click-through rate prediction is one of the core tasks in commercial recommender systems. It aims to predict the probability of a user clicking a particular item given user and item features. As feature interactions bring in non-linearity, they are widely adopted to improve the performance of CTR prediction models. Therefore,

effectively modelling feature interactions has attracted much attention in both the research and industry field. The current approaches can generally be categorized into three classes: (1) naïve methods, which do not model feature interactions and only use original features; (2) memorized methods, which memorize feature interactions by explicitly viewing them as new features and assigning trainable embeddings; (3) factorized methods, which learn latent vectors for original features and implicitly model feature interactions through factorization functions. Studies have shown that modelling feature interactions by one of these methods alone are suboptimal due to the unique characteristics of different feature interactions. To address this issue, we first propose a general framework called OptInter which finds the most suitable modelling method for each feature interaction. Different state-of-the-art deep CTR models can be viewed as instances of OptInter. To realize the functionality of OptInter, we also introduce a learning algorithm that automatically searches for the optimal modelling method. We conduct extensive experiments on three large datasets. Experimental results demonstrate the effectiveness of OptInter. Because our OptInter finds the optimal modelling method for each feature interaction, our experiments show that OptInter improves the best performed state-of-the-art baseline deep CTR models by up to 2.21%. Compared to the memorized method OptInter-M, which also outperforms baselines, we reduce up to 91% parameters.

[PDF of paper](#)

[Return to top](#)

Aggregating From Multiple Target-Shifted Sources

Changjian Shui (Université Laval)*; Zijian Li (Guangdong University of Technology); Jiaqi Li (University of Western Ontario); Charles X Ling (Western University); Christian Gagné (Université Laval); Boyu Wang (University of Western Ontario)

Location: Poster Room 1

ABSTRACT: Multi-source domain adaptation aims at leveraging the knowledge from multiple tasks for predicting a related target domain. A crucial aspect is to properly combine different sources based on their relations. In this paper, we analyzed the problem for aggregating source domains with different label distributions, where most recent source selection approaches fail. Our proposed algorithm differs from previous approaches in two key ways: the model aggregates multiple sources mainly through the similarity of semantic conditional distribution rather than marginal distribution; the model proposes a unified framework to select relevant sources for three popular scenarios, i.e., domain adaptation with limited label on target domain, unsupervised domain adaptation and label partial unsupervised domain adaptation. We evaluate the proposed method through extensive experiments. The empirical results significantly outperform the baselines.

[PDF of paper](#)

[Return to top](#)

Prediction of Individual Subject's Fiber Response in Vagus Nerve Stimulation with Meta Learning

Ximeng Mao (Mila, UdeM)*; Yao-Chuan Chang (Feinstein Institutes); Theodoros Zanos (Feinstein Institutes); Stavros Zanos (Feinstein Institutes); Guillaume Lajoie (Mila, Université de Montréal)

Location: Poster Room 1

ABSTRACT: Vagus nerve stimulation (VNS) is an emerging treatment for a multitude of disorders. It is crucial to acquire quantifiable fiber responses during VNS on a single subject basis. Previous work demonstrates the feasibility of estimating fiber responses from biological markers, but its requirement on data normalization makes it difficult to apply on unseen subjects. In this work, we propose to directly learn the unnormalized individual model for each subject, with the learned population model as prior. We formulate the problem as few-shot learning and present a meta learning pipeline as the solution.

In the proposed pipeline, the population model is trained a priori with the normalized data. For each subject, an inference model infers its scaling factors from the sampled few-shots, then we have the subject's individual model by rescaling the population fit by the inferred normalization. Individual model's prediction error on few-shots and the rest within the subject are called inner-loop and outer-loop error respectively. We adopt Bayesian optimization as inference model and consider 4 variations using the pipeline, varying on 1) regression or Gaussian process (GP) population model and 2) if using inner-loop and outer-loop error. Inner-loop error is used in one-step gradient update on individual model and outer-loop error to update inner-loop's learning rate

We apply this novel method on a rat VNS experiment dataset, which contains neurophysiological data of 11 subjects on 3 regression tasks. Compared with closed-form regression baseline currently adopted in the literature, the results show that our proposed meta learning pipeline performs as good in easier task and outperforms in the harder ones. The results show also the compatibility between GP population model and the pipeline, which is crucial when no closed-form expression can be derived from data. These preliminary results reveal an encouraging trend of meta learning in estimating subjects' fiber responses.

[PDF of paper](#)

[Return to top](#)

AFB: Improving Communication Load Forecasting Accuracy with Adaptive Feature Boosting

Chengming Hu (Samsung AI Center)*; Xi Chen (Samsung); Ju Wang (Samsung AI Center); Xue Liu (Samsung AI Center); Gregory Dudek (McGill University)

Location: Poster Room 1

ABSTRACT: Prediction of key system characteristics, such as the communication load, is required to overcome the delays in wireless communication systems. State-of-the-art approaches mostly apply existing Neural Network (NN) structures, and extract latent features purely based on their sensitivity to the forecasting accuracy. This way of feature extraction may neglect some non-obvious yet informative dimensions in the model input, leading to inaccurate forecasting results. In this work, we present an Adaptive Feature Boosting (AFB) approach, which integrates multiple AutoEncoders (AEs) to automatically extract robust and comprehensive latent features for communication load forecasting. The recurrent and residual connections among the AEs make sure that the

extracted latent features are representative for all input dimensions. With more comprehensive information extracted from the history, the forecasting accuracy is thus improved. We evaluate AFB against existing approaches on a real-world dataset that contains Call Detail Records (CDRs) in the city of Milan over a period of two months. The evaluation shows that our AFB-based approach achieves 35.2% more accurate load forecasting results than the state-of-the-art deep approaches.

[PDF of paper](#)

[Return to top](#)

Multi-Object Tracking and Segmentation with a Space-Time Memory Network

Mehdi Miah (Polytechnique Montréal)*; Guillaume-Alexandre Bilodeau (Polytechnique Montréal); Nicolas Saunier (Polytechnique Montréal)

Location: Poster Room 1

ABSTRACT: We propose a method for multi-object tracking and segmentation that does not require fine-tuning or per benchmark hyper-parameter selection. The proposed tracker focuses on the data association problem. Indeed, the recently introduced HOTA metric, which has a better alignment with the human visual assessment by evenly balancing detections and associations quality, has shown that improvements are still needed for data association. After creating tracklets using instance segmentation and optical flow, the proposed method relies on a space-time memory network developed for one-shot video object segmentation to improve the association of tracklets with temporal gaps. We evaluated our tracker on KITTIMOTS and MOTSCChallenge and show the benefit of our data association strategy with the HOTA metric.

[PDF of paper](#)

[Return to top](#)

Deep Reinforcement Learning at the Edge of the Statistical Precipice

Rishabh Agarwal (Google Research, Brain Team)*; Max Schwarzer (Mila); Pablo Samuel Castro (Google); Aaron Courville (MILA, Université de Montréal); Marc G. Bellemare (Google Brain)

Location: Poster Room 1

ABSTRACT: Deep reinforcement learning (RL) algorithms are predominantly evaluated by comparing their relative performance on a large suite of tasks. Most published results on deep RL benchmarks compare point estimates of aggregate performance such as mean and median scores across tasks, ignoring the statistical uncertainty implied by the use of a finite number of training runs. Beginning with the Arcade Learning Environment (ALE), the shift towards computationally-demanding benchmarks has led to the practice of evaluating only a small number of runs per task, exacerbating the statistical uncertainty in point estimates. In this paper, we argue that reliable evaluation in the few run deep RL regime cannot ignore the uncertainty in results without running the risk of slowing down progress in the field. We illustrate this point using a case study on the Atari 100k benchmark, where we find substantial discrepancies between conclusions drawn from point estimates alone versus a more thorough statistical analysis. With the aim of increasing the field's confidence in reported

results with a handful of runs, we advocate for reporting interval estimates of aggregate performance and propose performance profiles to account for the variability in results, as well as present more robust and efficient aggregate metrics, such as interquartile mean scores, to achieve small uncertainty in results. Using such statistical tools, we scrutinize performance evaluations of existing algorithms on other widely used RL benchmarks including the ALE, Procgen, and the DeepMind Control Suite, again revealing discrepancies in prior comparisons. Our findings call for a change in how we evaluate performance in deep RL, for which we present a more rigorous evaluation methodology, accompanied with an open-source library reliable, to prevent unreliable results from stagnating the field.

[PDF of paper](#)

[Return to top](#)

Pretraining Representations for Data-Efficient Reinforcement Learning

Max Schwarzer (Mila, Université de Montréal)*; Nitarshan Rajkumar (Mila, Université de Montréal); Michael Noukhovitch (Mila, Université de Montréal); Ankesh Anand (Mila, Université de Montréal); Laurent Charlin (HEC Montreal and Mila); R Devon Hjelm (Microsoft Research); Philip Bachman (Microsoft Research); Aaron Courville (MILA, Université de Montréal)

Location: Poster Room 1

ABSTRACT: Data efficiency is a key challenge for deep reinforcement learning. We address this problem by using unlabeled data to pretrain an encoder which is then finetuned on a small amount of task-specific data. To encourage learning representations which capture diverse aspects of the underlying MDP, we employ a combination of latent dynamics modelling and unsupervised goal-conditioned RL. When limited to 100k steps of interaction on Atari games (equivalent to two hours of human experience), our approach significantly surpasses prior work combining offline representation pretraining with task-specific finetuning, and compares favourably with other pretraining methods that require orders of magnitude more data. Our approach shows particular promise when combined with larger models as well as more diverse, task-aligned observational data – approaching human-level performance and data-efficiency on Atari in our best setting.

[PDF of paper](#)

[Return to top](#)

MLCR: Meta-Learning Communication Rules for Multi-Agent Reinforcement Learning

Jikun Kang (McGill University)*; Jie Fu (University of Montreal, Mila); Xue Liu (McGill University)

Location: Poster Room 1

ABSTRACT: Learning how to communicate is important for the efficient and effective operation of a real-world multi-agent system. Specifically, agents need to learn to understand: 1) how to extract latent features from raw messages, and 2) how to automatically learning generalized communication rules that can be transferred among

tasks. Most existing rules are manually designed, and they are either (i) limited by the message processing ability, which only utilizes the messages in centralized training, or (ii) designed for a single task, which lacks the ability to learn a more generalized communication protocol. In this paper, we propose a meta-learning approach, called Meta-Learning Communication Rules (MLCR), which generates communication rules to facilitate agent-to-agent collaboration automatically from the interaction between the agent and environment. Furthermore, our approach is orthogonal to many multi-agent methods and settings and thus can be seamlessly incorporated into existing actor-critic algorithms. Extensive evaluations on several multi-agent cooperation benchmark environments demonstrate that our proposed approach outperforms the state-of-the-art approaches by 9.7% and 74.5% on single tasks and transferred tasks, respectively.

[PDF of paper](#)

[Return to top](#)

Epidemiological compartmental models and machine learning: importance of the model's structure

Vincent Dandenault (Université de Montréal)*; Sofia Alfonso (CHU Sainte-Justine Research Center); Alexandre Simard (CHU Sainte-Justine Research Center); Masoumeh Sajedi (CHU Sainte-Justine Research Center); Benoît Mâsse (School of Public Health - Université de Montréal); Marie-Claude Boily (School of Public Health - Imperial College London); Simon de Montigny (School of Public Health - Université de Montréal)

Location: Poster Room 1

ABSTRACT: Compartmental epidemiological models have been embedded into machine learning systems during the COVID-19 pandemic for the purpose of assimilating daily data produced by public health officials to make forecasts and evaluate interventions over short time periods [1, 2].

This data assimilation approach is known to be sensitive to the model's structure and specifications [3, 4, 5, 6]. In the case of COVID-19, the emergence of variants (having distinct epidemiological features such as a higher probability of transmission per contact, or higher hospitalizations and death rates) and the progression of vaccination campaigns are key mechanisms that can be captured into compartmental models. Since the pandemic and the public health response are continuously changing, using a fixed model structure with time-dependent parameters that are adapted to real-time data does not constitute a robust method to generate forecasts. Ideally, the model structure should be adapted to account for dynamic changes [5, 6, 7, 8]. For example, the emergence of new variants would need to be parameterized into the existing model, which is very time-demanding. There is thus a crucial need for new methods and tools that will help specify, construct, and calibrate compartmental models automatically, with the end goal of having interpretable and machine-learnable model structures [9].

To tackle this open problem, we developed a compositional modeling framework for the automatic building of generalized versions of the S-I-R (Susceptible, Infected, Recovered) model, based on a flexible representation of the natural history of the disease. The framework is capable of adding several disease variants over time. Models are expressed with human-editable configuration files such that users can test different structures without having to explicitly program them. This framework will be extended to handle different types of model extensions to segment compartments by age, by region and by vaccination status.

[PDF of paper](#)

[Return to top](#)

Data augmentation instead of explicit regularization

Alex Hernandez-Garcia (Mila - Quebec AI Institute)*

Location: Poster Room 1

ABSTRACT: Contrary to most machine learning models, modern deep artificial neural networks typically include multiple components that contribute to regularization. Despite the fact that some (explicit) regularization techniques, such as weight decay and dropout, require costly fine-tuning of sensitive hyperparameters, the interplay between them and other elements that provide implicit regularization is not well understood yet. Shedding light upon these interactions is key to efficiently using computational resources and may contribute to solving the puzzle of generalization in deep learning. Here, we first provide formal definitions of explicit and implicit regularization that help understand essential differences between techniques. Second, we contrast data augmentation with weight decay and dropout. Our results show that visual object categorization models trained with data augmentation alone achieve the same performance or higher than models trained also with weight decay and dropout, as is common practice. We conclude that the contribution on generalization of weight decay and dropout is not only superfluous when sufficient implicit regularization is provided, but also such techniques can dramatically deteriorate the performance if the hyperparameters are not carefully tuned for the architecture and data set. In contrast, data augmentation systematically provides large generalization gains and does not require hyperparameter re-tuning. In view of our results, we suggest to optimize neural networks without weight decay and dropout to save computational resources, hence carbon emissions, and focus more on data augmentation and other inductive biases to improve performance and robustness.

[PDF of paper](#)

[Return to top](#)

CARL: Conditional value-at-risk Adversarial Reinforcement Learning

Mathieu Godbout (Université Laval)*; Maxime Heuillet (Université Laval); Audrey Durand (Université Laval)

ABSTRACT: In this paper we present a risk-averse reinforcement learning (RL) method called Conditional value-at-risk Adversarial Reinforcement Learning (CARL). To the best of our knowledge, CARL is the first game formulation for Conditional Value-at-Risk (CVaR) RL. The game takes place between a policy player and an adversary that perturbs the policy player's state transitions given a finite budget. We prove that, at the maximin equilibrium point, the learned policy is CVaR optimal with a risk tolerance explicitly related to the adversary's budget. We provide a gradient-based training procedure to solve CARL by formulating it as a zero-sum Stackelberg Game, enabling the use of deep reinforcement learning architectures and training algorithms. Finally, we show that solving the CARL game does lead to risk-averse behaviour in a toy grid environment, also confirming that an increased adversary produces increasingly cautious policies.

[PDF of paper](#)

[Return to top](#)

Learning Aggregations of Binary Activated Neural Networks with Probabilities over Representations

Louis Fortier-Dubois (Université Laval)*; Benjamain Leblanc (Université Laval); Gaël Letarte (Université Laval); Pascal Germain (Université Laval); François Laviolette (Université Laval)

ABSTRACT: We present Aggregation of Binary-activated Networks (ABNet), a learning algorithm that considers a probability distribution over neural network parameters using a binary activation function. By propagating whole probability distributions over a layer representation, ABNet circumvents the non-differentiability issue of such networks. It is also well suited for analysis based on the PAC-Bayesian theory, an active search field for analyzing generalization abilities of deep neural networks. Complementing a recent idea of a practical, approximative algorithm called PBGNet, our strategy focuses on the theoretical properties of considering both binary activations and randomized weights in neural networks.

[PDF of paper](#)

[Return to top](#)

Sequential Pipeline Optimization: Bandits-driven Exploration using a Collaborative Filtering Representation

Maxime Heuillet (Université Laval)*; Audrey Durand (Université Laval); Benoit Debaque (Thales Digital Solutions)

Location: Poster Room 1

ABSTRACT: The Pipeline Optimization problem consists in recommending automatically a pipeline efficient at conducting a selected predictive task on a given dataset. Collaborative Filtering (CF) methods have shown great success in this task. In this work, we frame the pipeline optimization problem under a sequential setting where datasets arrive one at a time. On each dataset, an agent can try a small number of pipelines (exploration) before recommending a pipeline for this dataset (recommendation). The goal is to maximize the performance of the recommended pipelines over the sequence of datasets. More specifically, we focus on the exploration policy used for selecting the pipelines to explore before making the recommendation. We propose an approach based on the LinUCB bandit algorithm that leverages the latent representations extracted from matrix factorization (MF). We show that the exploration policy impacts the recommendation performance and that MF-based latent representations are more useful for exploration than for recommendation.

[PDF of paper](#)

[Return to top](#)

Quantum machine learning for text classification

Jean Frederic Laprade (Universite de Sherbrooke)*; Sarah Blanchette (Universite de Sherbrooke); Zachary Zanussi (Statistics Canada); Omar Chikhar (Statistics Canada); Vladimir Skavysh (Bank of Canada)

Location: Poster Room 1

ABSTRACT: The availability of near-term intermediate-scale quantum devices provides an opportunity for data scientists to begin developing quantum methods for a variety of machine learning tasks. With this objective in mind, we explore three different approaches in supervised quantum machine learning for text classification, namely variational quantum classifier, quantum kernel methods and hybrid classical-quantum transfer learning. We develop our methods using the IBM Qiskit and Xanadu PennyLane quantum software libraries and implement our methods on real quantum hardware devices.

[PDF of paper](#)

[Return to top](#)

RandomSCM: interpretable ensembles of sparse classifiers tailored for omics data

Thibaud Godon (Laval University)*; Pier-Luc Plante (Laval University); Élina Francovic-Fontaine (Université Laval); Baptiste Bauvin (Université Laval); Alexandre Drouin (ServiceNow); Jacques Corbeil (Université Laval); François Laviolette (Université Laval)

ABSTRACT:

Background :

- Understanding the relationship between the Omics and the phenotype is a central problem in precision medicine
- The high dimensionality of metabolomics data challenges learning algorithms in terms of scalability and generalization
- Most learning algorithms do not produce interpretable models

Method :

- We propose an ensemble learning algorithm based on conjunctions or disjunctions of decision rules

Results :

- Applications on metabolomics data shows that it produces models that achieves high predictive performances
- The interpretability of the models makes them useful for biomarker discovery and patterns discovery in high dimensional data

[PDF of paper](#)

[Return to top](#)

Learning Function From Structure In Neuromorphic Networks

Laura E Suarez (McGill University)*; Bratislav Misic (McGill University); Guillaume Lajoie (Mila, Université de Montréal); Blake Richards (McGill University)

Location: Poster Room 1

ABSTRACT: Recent imaging technologies allow the reconstruction of the human connectome, a network model of the macroscale wiring patterns of the brain. Descriptive analyses of the connectome have found evidence of non-random architectural features, characteristic of complex systems, that theoretically shape the computational capacity of the brain. Features include: high clustering and short-path length, specialized segregated communities, and heavy-tailed degree distributions. Still, how network organization supports information-processing remains unknown. Broadly, we address the structure-function relationship in the brain, but with a focus on computation. To do so, we combine connectomics and reservoir computing to investigate the link between macroscale connectivity, and the computational properties that emerge from network dynamics in the human connectome. Specifically, we construct artificial neural networks endowed with biologically realistic connection patterns derived from MRI. We train these connectome-informed reservoirs to perform a memory task. To evaluate how performance depends on network structure and dynamics, we parametrically drive the network to transition between stable, critical and chaotic states. Throughout, we assess computational performance of empirically-derived connectomes against two null network models. We find that: i) the underlying macroscale topology and mesoscale modular organization of the brain enhances computational performance in the context of critical dynamics; ii) the modular organization of the brain in functional systems constitutes a computationally relevant feature of the human connectome; iii) the connectome's network topology optimizes the trade-off between computational capacity and metabolic costs. Remarkably, we observe a prominent interaction between network structure and dynamics, such that the same underlying architecture can support a wide range of learning capacities across dynamical regimes.

[PDF of paper](#)

[Return to top](#)

A deep learning end-to-end system for vision-based fault recognition in thermal and civil inspection of the underground distribution network

Francois Miralles (Hydro-Québec); Luc Cauchon (Hydro-Québec); Arnaud Zinflou (Hydro-Québec)*; Mouhamadou Makhtar Dione (Hydro-Québec); Gregoire Francois (Hydro-Québec); Magnan Marc-Andre (Hydro-Québec); Jinga Livius (Hydro-Québec); Dafali Salwa (École polytechnique de Montréal,)

ABSTRACT: In past years, deep learning techniques like convolutional neural networks have achieved impressive results in numerous computer vision tasks, such as classification, segmentation, localization object detection, semantic segmentation or instance segmentation, as demonstrated on ImageNet or Coco object detection in academic field. Companies across various industries from healthcare to utilities are finding multiple use-cases for this groundbreaking technology. However, there are some unique practical challenges remain for real-world applications, e.g., limited labeled data samples, imbalanced data distributions and the detection of small components and defects. In this work, we present an end-to-end vision-based deep learning system developed at Hydro-Québec to overcome these challenges and improve the efficiency of the maintenance of the underground distribution network.

[PDF of paper](#)

[Return to top](#)

Information Sharing in mHealth

Antoine Gariépy (Université Laval)*; Audrey Durand (Université Laval)

ABSTRACT: In this work we present a contextual bandit setting for information sharing in mobile health (mHealth). We consider a mHealth setting where patients are clustered into predefined clinical groups before the study and investigate if we can leverage information from these groups. Our main contribution is the application of a statistical test used to tackle routine bandits in order to decide whether to aggregate information about treatments from different groups. Preliminary experiments show promising results.

[PDF of paper](#)

[Return to top](#)

Curating the Twitter Election Integrity Datasets for Better Online Troll Characterization

Albert M Orozco Camacho (Mila - Québec AI Institute / McGill University)*

Location: Poster Room 1

ABSTRACT: In modern days, the transmission of novel information has strengthened its dependency to the World Wide Web. Social media platforms provide accessible channels for the interaction and immediate reflection upon the most important events happening around the world. Nevertheless, as large communities of users interact with each other, important network design bottlenecks often permit unreliable information to get spread, especially, in automatic fashions. In this paper, we present a curated set of datasets whose origin stems from the Twitter's Information Operations2efforts. Most activity comes from state-linked accounts, which has been identified to target specific (political) goals. More notably, these accounts, which have been already suspended, provide a notion of how state-backed human trolls operate. We present detailed analyses of how this activity varies over time, and motivate its use and abstraction in the context of deep representation learning: for instance, to learn and, potentially track, troll behaviour. We present baselines for such tasks and highlight the differences there may exist, especially from several structural ablation studies, directly on the proposed graphs. In particular we focus on learning differences between the reported troll user activity and that created from active normal users. To the best of our knowledge, there is no previous work who addresses the problem of modelling and distinguishing troll activity on actual verified state-backed fake accounts, in the same way we are proposing.

[PDF of paper](#)

[Return to top](#)

Poster Room 2

Multimodal Transformer for Emotion Recognition and Sentiment Analysis

Rihab Hajlaoui (Polytechnique Montréal)*; Guillaume-Alexandre Bilodeau (Polytechnique Montréal); Jan Rockemann (Airudi)

Location: Poster Room 2

ABSTRACT: Implementing an end-to-end deep learning model for emotion recognition or sentiment analysis has become an emerging research area in natural language and computer vision. Based on the work of Tsai et al. we propose a multimodal transformer architecture. Our approach relies on cross-modal attention to encode the features of a single modality based on the information sent by other modalities. We also propose to add a GRN (Gated Residual Network) that allows the network to remove irrelevant information received from other modalities.

[PDF of paper](#)

[Return to top](#)

Exploring the loss landscape of neural networks with manifold learning and topological data analysis

Stefan Horoi (Université de Montréal)*; Jessie Huang (Yale University); Bastian A Rieck (Institute of AI for Health, Helmholtz Centre Munich); Guillaume Lajoie (Mila, Université de Montréal); Guy Wolf (Université de Montréal); Smita Krishnaswamy (Yale University)

Location: Poster Room 2

ABSTRACT: Recent work has established clear links between the generalization performance of trained neural networks and the geometry of their loss landscape near the local minima to which they converge. This suggests that qualitative and quantitative examination of the loss landscape geometry could yield insights about neural network generalization performance during training. To this end, researchers have proposed visualizing the loss landscape through the use of simple dimensionality reduction techniques. However, such visualization methods have been limited by their linear nature and only capture features in one or two dimensions, thus restricting sampling of the loss landscape to lines or planes. Here, we expand and improve upon these in three ways. First, we present a novel "jump and retrain" procedure for sampling relevant portions of the loss landscape. We show that the resulting sampled data holds more meaningful information about the network's ability to generalize. Next, we show that non-linear dimensionality reduction of the jump and retrain trajectories via PHATE, a trajectory and manifold-preserving method, allows us to visualize differences between networks that are generalizing well vs poorly. Finally, we combine the PHATE trajectories with a computational homology characterization to quantify the differences in trajectories.

[PDF of paper](#)

[Return to top](#)

Typing assumptions improve identification in causal discovery

Philippe Brouillard (Mila, Université de Montréal)*; Perouz Taslakian (Element AI); Alexandre Lacoste (ServiceNow); Sebastien Lachapelle (Mila, Université de Montréal); Alexandre Drouin (ServiceNow)

Location: Poster Room 2

ABSTRACT: Causal discovery from observational data is a challenging task to which an exact solution cannot always be identified. Under assumptions about the data-generative process, the causal graph can often be identified up to an equivalence class. Proposing new realistic assumptions to circumscribe such equivalence classes is an active field of research. In this work, we propose a new set of assumptions that constrain possible causal relationships based on the nature of the variables. We thus introduce typed directed acyclic graphs, in which variable types are used to determine the validity of causal relationships. We demonstrate, both theoretically and empirically, that the proposed assumptions can result in significant gains in the identification of the causal graph.

[PDF of paper](#)

[Return to top](#)

Task-Assisted GAN for resolution enhancement, quality improvement and modality translation in fluorescence microscopy

Catherine Bouchard (Université Laval)*; Theresa Wiesner (Université Laval); Andréanne Deschênes (Université Laval); Flavie Lavoie-Cardinal (Université Laval); Christian Gagné (Université Laval)

Location: Poster Room 2

ABSTRACT: The power of GANs for translating images between domains is no secret to the deep learning community. However, their applicability to biology is questioned by life science experts. To analyze biological samples, the data acquired has to be extremely precise down to the nanoscopic scale, which is not guaranteed by black-box like generative approaches. We developed a GAN-based model which is specifically trained to accurately decode and translate the information relevant for a biological question of interest. The GAN is trained to translate an input image from a low-information domain into an output domain of higher resolution or quality. An additional network is trained parallelly to the generator to compel it to not only translate the low-resolution (LR) features into realistic high-resolution (HR) features, but also to generate accurately the unresolved details of interest. This network uses the ground truth HR images to learn a task that is impossible to perform on the LR images and applies corrections on the generator to make this task solvable on the generated images.

The task-assisted GAN's (TA-GAN) flexibility and functional simplicity makes it applicable to an extensive number of tasks, modalities and translation objectives. We first demonstrate its applicability for single-image

super-resolution by training it on multiple datasets of paired diffraction-limited confocal and super-resolved stimulated emission depletion (STED) images. We show that the distribution of nanoscopic structures is maintained through the translation, leading to the same biological conclusions using the ground truth STED images than using those generated from confocal images. Our TA-GAN model increases the resolution of microscopy images while minimizing the hallucination of artifacts and enables the recycling of complete datasets and annotations to train networks on new imaging modalities.

[PDF of paper](#)

[Return to top](#)

A Vision-Based Automatic Transcription of Guitar Music From RGBD Videos

Mark Asmar (Polytechnique Montreal)*; Lama Seoud (Polytechnique Montreal);
Guillaume-Alexandre Bilodeau (Polytechnique Montréal)

Location: Poster Room 2

ABSTRACT: Guitarists play music by listening or by relying on transcriptions. These transcriptions can use different types of notations (e.g. common music notation, tablature, etc.). Tablature (or "tabs" for short) is a form of notation that is adapted by guitarists and that tell the guitarists the frets to press and on what string. Nowadays, transcription is still manual which makes it a slow process and vulnerable to mistakes. To address this issue, automatic transcription for guitar music videos can be applied; it can be done through audio, vision or a combination of both. However, having the same note in multiple positions on a guitar fretboard makes it more ambiguous for audio, hence it is believed that vision suffices. Unlike previous methods that rely on image processing algorithms and detecting fingertips to process a guitar playing video and extract tablatures, our approach uses an RGBD camera, semantic segmentation, hand pose estimation and optical flow. The RGBD camera captures RGB and depth frames, the segmentation is used for fretboard localization in a frame; the hand pose estimation estimates the location of joints and their 3D coordinates to recognize the different pressings and their places while the optical flow serves as a detector of string picking/plucking.

[PDF of paper](#)

[Return to top](#)

Parallel Gaussian Process-Based Bayesian Optimization for Multiple Cortical Array Neuroprosthetic Control

Julien Rimok (Université de Montréal)*; Numa Dancause (Université de Montréal); Guillaume Lajoie (Mila, Université de Montréal)

Location: Poster Room 2

ABSTRACT: A parallel Gaussian process-based Bayesian Optimization algorithm was implemented for neuroprosthetic control of multiple cortical arrays implanted in a NHP brain. It was compared to both random search and fitting all the data with a single Gaussian process and was found to be better performing.

[PDF of paper](#)

[Return to top](#)

Learning A Hierarchy of Behaviours though A Unified Reward

Hanqing Zhao (McGill University)*; Gregory Dudek (McGill University); Xue Liu (McGill University)

Location: Poster Room 2

ABSTRACT: We will present new approaches of learning human-like rhythmic behaviour reward from few-shot human driving demonstrations, the learnt reward is combined with other rewards so as to form a unified reward that represents an hierarchy of behaviours, while minimizing the adversary impact on high-level discrete behaviour learning results.

From a classical perspective, enabling a hierarchy of behaviours involves a hierarchy of controllers that correspond to a set of behaviours with different abstraction levels, these controllers are either manually designed or learnt from data. The difficulty of designing such a hierarchy of independent controllers increases with the number and complexity of behaviors that need to be learned. Instead of designing a hierarchy of controllers, recent works often learn an end-to-end controller that exhibits multiple behaviours with different abstractions through a unified reward, for example, enables an amphibious robot to demonstrate drive-to-goal behaviour and its obstacle avoidance sub-behaviour simultaneously, Xie et al.\cite{xie2017towards} assumes human drivers always prefer driving straight and fast, and built a vehicle that learns to avoid obstacles while keeping straight and fast driving rhythm through a unique reward. In these works, a reward function that represents multiple behaviours is constructed as a sum of the reward function of each behavior with different scales. For examples for obstacle avoidance behaviour and its straight driving sub-behaviour.

However, combing the high-level behaviour reward with the one for low-level behaviour may adversely affect the learning result of both behaviours, For a better multi-behaviour learning from unified reward, we investigate the learning of reward functions for each individual behaviour with respect to two following principles: better behaviour description, and lower cross-behaviour interference.

[PDF of paper](#)

[Return to top](#)

Gradient Boosted Tree Classification vs. Regression for Highly Imbalanced Classification Problems

Aristides T Milios (McGill)*

Location: Poster Room 2

ABSTRACT: Detecting anomalous health events early from time series data is of critical importance for improving healthcare outcomes in intensive care units (ICUs) of hospitals. By predicting these events far enough in advance, nurses are able to mitigate the worst effects, or even prevent the event entirely. Two common tasks in this category are detecting acute hypotensive episodes (AHEs) and detecting episodes of anomalously high heart rate. The datasets for these tasks are typically highly imbalanced, as these events are relatively infrequent compared to the total time patients spend in ICUs. Most commonly, early detection scenarios for time series are modelled as a binary classification problem, with sliding observation windows over time, where the predictive label is whether or not the event happened in a time frame X minutes into the future. An alternative way to model such tasks is as a regression problem. In this case, instead of predicting a binary label, the goal is to predict a value between 0 and 1. This value represents, as a percentage, the duration of time the anomalous condition is true in the future time frame as a proportion of its total duration (e.g., for an episode of length 15 mins and a future time frame of 30 mins, this would be 0.5). For experimentation, the well-known MIMIC patient dataset is used for the above two tasks, and XGBoost and LightGBM are used as learners. We demonstrate that this alternative setup consistently leads to a greater area under the precision-recall curve, which has been demonstrated previously to be a more effective metric in cases of highly imbalanced datasets than area under ROC. We hypothesize that this increase in area under PR-curve is due to the reduced data imbalance in the regression case, compared to the binary classification case, as the regression labels are "spread out" over a continuous spectrum and thus allow the regression model to learn more effectively than the corresponding binary classification model.

[PDF of paper](#)

[Return to top](#)

Machine Learning for Optimising Patient Follow-up for Cancer Rehabilitation (A CanRehab Study)

Rupali Bhati (Universite Laval)*; Jennifer Jones (University of Toronto); Audrey Durand (Université Laval)

Location: Poster Room 2

ABSTRACT: Cancer survivors face unique challenges during and after treatment. There is a need for tailored models of care. The Canadian Cancer Rehabilitation (CanRehab) Team aims to improve the systematic identification of the adverse effects of cancer and its treatments using innovative e-health solutions. The aim of this study is to predict a patient's future symptoms based on past data. The analysis of this study needs to be

comprehensible not only by machine learning experts but also by domain experts. Therefore, we use interpretable ML models like Decision Trees and XGBoost. There is a class imbalance issue in the dataset which we resolve using the SMOTE technique. The preliminary results show that the Decision Trees Classification algorithm obtains the lowest weight mean average error of 0.86 indicating that the weighted mean deviation from the true pain level is 0.86. Moreover, it is observed that as number of samples in a class decreases, the MAE increases indicating a possible correlation between the two quantities. In the future, we want to form this problem as a contextual bandit problem such that the model can act and at the same time learn from new data in real time.

[PDF of paper](#)

[Return to top](#)

HOLISTIC GUIDANCE FOR OCCLUDED PERSON RE-IDENTIFICATION

Madhu Kiran (ETS Montreal)*; Gnana Praveen Rajasekhar (Ecole Technologie Superieure); Le Thanh Nguyen-Meidine (ETS Montreal); Soufiane Belharbi (ÉTS Montreal); Louis-Antoine Blais-Morin (Genetec Inc.); Eric Granger (ETS Montreal)

Location: Poster Room 2

ABSTRACT: In real-world video surveillance applications, person re-identification (ReID) suffers from the effects of occlusions and detection errors. Despite recent advances, occlusions continue to corrupt the features extracted by state-of-art CNN backbones, and thereby deteriorate the accuracy of ReID systems. To address this issue, methods in the literature use an additional costly process such as pose estimation, where pose maps provide supervision to exclude occluded regions. In contrast, we introduce a novel Holistic Guidance (HG) method that relies only on person identity labels, and on the distribution of pairwise matching distances of datasets to alleviate the problem of occlusion, without requiring additional supervision. Hence, our proposed student-teacher framework is trained to address the occlusion problem by matching the distributions of between- and within-class distances (DCDs) of occluded samples with that of holistic (non-occluded) samples, thereby using the latter as a soft labeled reference to learn well separated DCDs. This approach is supported by our empirical study where the distribution of between- and within-class distances between images have more overlap in occluded than holistic datasets. In particular, features extracted from both datasets are jointly learned using the student model to produce an attention map that allows separating visible regions from occluded ones. In addition to this, a joint generative-discriminative backbone is trained with a denoising autoencoder, allowing the system to self-recover from occlusions. Extensive experiments on several challenging public datasets indicate that the proposed approach can outperform state-of-the-art methods on both occluded and holistic datasets.

[PDF of paper](#)

[Return to top](#)

On the Limits of Algorithmic Counting and Out-of-Distribution Generalization

Martin Weiss (Mila)*

Location: Poster Room 2

ABSTRACT: This work makes two contributions. First, we analyze the ability of a recurrent neural network to extrapolate when trained to recognize the counting language $a^n b^n$, and then second, we prove a bound on the maximum value that a fixed, finite-size neural network with layer normalization can count to by simulating the counting algorithmic used by K-Counter Machines.

[PDF of paper](#)

[Return to top](#)

Cogment: Open Source Framework For Distributed Multi-actor Training, Deployment And Operations

Sai Krishna Gottipati (AI-Redefined)*; Sagar Kurandwad (AI-r); Jonathan M Fisher (AI Redefined); Vincent Robert (AI-r); Gregory Szriftgiser (AI Redefined); Clodéric Mars (AI Redefined)

Location: Poster Room 2

ABSTRACT: Involving humans directly for the benefit of AI agents' training is getting traction thanks to several advances in reinforcement learning and human-in-the-loop learning. Humans can provide rewards to the agent, demonstrate tasks, design a curriculum, or act in the environment, but these benefits also come with architectural, functional design and engineering complexities. We present Cogment, a unifying open-source framework that introduces an actor formalism to support a variety of humans-agents collaboration typologies and training approaches. It is also scalable out of the box thanks to a distributed micro service architecture, and offers solutions to the aforementioned complexities.

[PDF of paper](#)

[Return to top](#)

Dynamic Consolidation for Continual Learning

Hang Li (McGill University)*; Xue Liu (McGill University)

Location: Poster Room 2

ABSTRACT: Learning models from a stream of non-stationary data is a critical problem to be solved to achieve general artificial intelligence. As a promising solution, Continual Learning (CL) technique aims to build intelligent systems that have the plasticity to learn from new information without forgetting the previously obtained knowledge. Unfortunately, existing CL methods face two non-trivial limitations: (1) When updating a model with new data, existing CL methods usually constrain the model parameters within the vicinity of the parameters optimized for old tasks, limiting the exploration ability of the model; (2) The important strength of each parameter (used to consolidate the previously learned knowledge) is usually fixed and thus is sub-optimal for the dynamic parameter updates. To address these limitations, we first relax the vicinity constraints with a global definition of the important strength, which allows us to explore the whole parameter space. Specifically, we define the important strength as the sensitivity of the global loss function to the model parameters. Moreover, we propose adjusting the important strength adaptively to align it with the dynamic parameter updates.

[PDF of paper](#)

[Return to top](#)

Deliberation gated by opportunity cost adapts to context with urgency

Maximilian Puelma Touzel (Mila)*; Paul Cisek (UdeM); Guillaume Lajoie (Mila, Université de Montréal)

Location: Poster Room 2

ABSTRACT: The value we place on our time impacts what we decide to do with it. Value it too little, and we obsess over all details. Value it too much, and we rush carelessly to move on. How to strike this often context-specific balance is a challenging decision-making problem. Average-reward, putatively encoded in the brain by tonic dopamine, serves in existing reinforcement learning theory as the stationary opportunity cost of time. In the challenging, real-world setting of continuing environments, this average-reward formulation of RL has many advantages over the more popular discount-reward formulation. However, environmental context and the cost of deliberation therein often varies in time and is hard to infer and predict. Here, we consider continuing tasks of successive episodes and define a non-stationary opportunity cost of deliberation arising from performance variation on multiple timescales. Estimated from reward history, this cost readily adapts to reward-relevant changes in context and suggests a generalization of average-reward reinforcement learning (AR-RL) to account for non-stationary contextual factors. We demonstrate the utility of this deliberation cost in a simple decision-making heuristic called Performance-Gated Deliberation, which is an explicit approximation to AR-RL and is consistent with empirical results in both cognitive and systems decision-making neuroscience. We use behaviour and neural recordings from non-human primates in a non-stationary continuing episodic task involving random walk prediction to support the brain's use of PGD over AR-RL and discount-reward RL

solutions, suggesting it may confer an advantage in this setting. As a relatively simple and nimble strategy, PGD makes for an attractive candidate when acknowledging that a combination of knowledge and resource limitations over task, development, and evolutionary timescales have shaped decision-making in non-stationary environments.

[PDF of paper](#)

[Return to top](#)

Sample Boosting Algorithm - SamBA - Greedy local sparse votes and their applications to bioinformatics

Baptiste Bauvin (Université Laval)*; Cécile Capponi (Aix-Marseille University); François Laviolette (Université Laval)

Location: Poster Room 2

ABSTRACT:

Background :

- Ensemble methods combine classifiers to output robust models
- The vote is usually done using a linear combination of the classifiers, with scalar weights

Method :

- We propose an ensemble method that uses local properties of the learning set to output a decision function with variable voting weights

Results :

- A sparse ensemble method performing as well as the state of the art
- Interpretable decision function allowing the application to real-world bio-informatics problems

[PDF of paper](#)

[Return to top](#)

A Thompson sampling approach to learn the Whittle index in restless bandit problems

Nima Akbarzadeh (McGill University)*; Aditya Mahajan (McGill University)

Location: Poster Room 2

ABSTRACT: A well-known class of scheduling problems is restless multi-armed bandit (RMAB). In a RMAB problem, there are certain number of processes where each is a Markov process. We refer to each process as an arm and our action is to choose a subset of them at a time. The state evolution of each arm depends on

whether the arm is selected or not. The state space of such problems is enormously large, i.e., it is exponential in the number of arms. As a result, finding the optimal policy becomes intractable when the number of arms is large. A widespread heuristic method which is being used in this setup is Whittle index policy by which an index is computed for each state of the arms and at each time, the arm with the highest index is chosen. Whittle index policy is asymptotically optimal as the number of arms approaches infinity and it has been shown that it performs close-to-optimal and outperforms state-of-the-art methods if a technical condition known as indexability is satisfied.

We consider an instance of RMAB where the transition probabilities under either of the actions are unknown for all arms and propose an algorithm to learn the transition probabilities by Thompson sampling and estimate the Whittle index policy. We measure the performance of our algorithm by regret, i.e., the difference in the performance of the learning algorithm with the Whittle index policy, and we prove an upper bound on the regret. Let n be the number of arms, K be the state space size of each arm, m be the number of arms to be selected at each time and T be the time horizon. Then, using the results in the literature for using Thompson sampling to learn a Markov decision process results in an upper-bound on the regret as $O(n K^n \sqrt{\binom{n}{m} T})$ for a RMAB problem. While we prove that an upper bound on the regret as $O(n^2 K \sqrt{n T \log T})$ for a RMAB problem. Our upper bound on the regret significantly improves the existing upper bound for this class of problems.

[PDF of paper](#)

[Return to top](#)

Learning Bike Origins and Destinations from Counts and Trajectories

Ouassim Manout (Polytechnique Montreal / ENTPE); Qingwu Liu (Polytechnique Montreal); Francesco Ciari (Polytechnique Montreal); Lijun Sun (McGill University); Catherine Morency (Polytechnique Montreal); Martin Trepanier (École Polytechnique de Montréal); Nicolas Saunier (Polytechnique Montreal)*

ABSTRACT: In various metropolitan areas, cycling has the potential to meet the mobility needs of many urban dwellers and reduce the adverse effects of individual motorized vehicles. However, despite its evident contribution to urban mobility, bike use is still marginal, especially in North America. One barrier to its widespread use is the development of safe and dedicated bike infrastructure. To plan such infrastructure, planning authorities are often in need of data on bike demand, i.e. bike flows between origins and destinations or bike Origin Destination (OD) matrix. In this paper, we propose a machine learning method to estimate the bike OD matrix using count and trajectory data. The method is validated using the shared bike system of Montreal, Canada (BIXI).

[PDF of paper](#)

[Return to top](#)

Towards an inclusive AI: Research on the revitalization and preservation of indigenous languages

Tan Le (Universite du Quebec a Montreal)*; fatiha sadat (UQAM)

ABSTRACT: Indigenous languages have been very challenging when dealing with several NLP tasks and applications. These languages are recognized as low-resource polysynthetic but also endangered languages; having morphological complexity and dialectal variations with different spelling (Littell et al., 2018). Therefore, the NLP international community has a growing research interest towards Indigenous languages, realities and challenges for their revitalization preservation. This research focuses on an endangered Indigenous language, Inuinnaqtun, one of the polysynthetic languages spoken in Northern Canada, in the Inuit language family (Littell et al., 2018; Kudlak and Compton, 2018). Herein, we propose an unsupervised morphological segmentation based on Adaptor Grammars approach for Inuit language family and an empirical study of Inuinnaqtun. Experiments and evaluations on the Inuinnaqtun word segmentation task showed promising results and performance improvements of the proposed system.

[PDF of paper](#)

[Return to top](#)

SESNO: Sample Efficient Social Navigation from Observation

Bobak H Baghi (Samsung Electronics)*

Location: Poster Room 2

ABSTRACT: As mobile robots come to inhabit and traffic social spaces, they must account for social cues and behave in a socially compliant manner that is consistent with the norms of acceptable human interaction. In this paper, we present the Sample Efficient Social Navigation From Observation (SESNO) algorithm that learns socially-compliant navigation policies from observations of human trajectories efficiently. SESNO is an inverse reinforcement learning-based algorithm which learns from human trajectory observations without knowing their specific actions. We increase the sample-efficiency of our approach over previous methods by leveraging the notion of a replay buffer (found in many off-policy reinforcement learning methods) to avoid the additional sample complexity normally associated with inverse reinforcement learning. SESNO is evaluated using publicly available pedestrian motion data sets and we compare its performance to related baseline methods in the literature. We show that SESNO yields superior performance than existing baseline approaches while also decreasing training time and sample complexity.

[PDF of paper](#)

[Return to top](#)

Learn to Perturb: A Self-Learning Hyper-Heuristic for the Simultaneous Stochastic Optimization of Mining Complexes

Yassine Yaakoubi (Mila, McGill University)*; Roussos Dimitrakopoulos (McGill University)

Location: Poster Room 2

ABSTRACT: A new self-learning hyper-heuristic is proposed to address the need for self-managed solution approaches that can tackle large-scale problems such as the simultaneous stochastic optimization of mining complexes (SSOMC). The proposed method is a multi-neighborhood simulated annealing algorithm used in conjunction with reinforcement learning (RL), where the selection of a perturbation (low-level heuristic) is made in self-adaptive learning. Several state-of-the-art agents have been implemented and incorporated into the RL framework. Results show the RL agent's ability to generalize and the method's effectiveness on real-sized mining complexes reducing the execution time by 30-55%.

[PDF of paper](#)

[Return to top](#)

Can we trust the Wizard 🧙? Interpreting Knowledge-Grounded Dialogue Systems.

Yuvraj Prabhakar (Concordia University)*; Mehar Bhatia (NeuralSpace Research)

Location: Poster Room 2

ABSTRACT: Knowledge-grounded dialogue systems tend to generate responses based on information provided in grounding corpora. Though there has been recent progress in training end-to-end informative systems that mimic human language at the linguistic level, yet there are no controls available that ensure they are truthful. Everyday, such systems help educate users about a particular topic through conversational multi-turn interaction. This research focuses on interpreting these generative neural dialogue models, and to ensure that responses stay 'faithful' to information from a text document in a conversation.

[PDF of paper](#)

[Return to top](#)

Improving Fairness in Heterogeneous Federated Learning

Shaoxiang Qin (McGill University)*; Jiachen lei (Zhejiang University); Hillary Tao (McGill University); Xi Chen (McGill University); Qiao Xiang (Yale University); Zhongjie Ba (Zhejiang University); Xue Liu (McGill University)

Location: Poster Room 2

ABSTRACT: Federated learning acquires knowledge from multiple clients' locally collected data. These data are usually class imbalanced and non-identically distributed over clients in real-world systems. Models trained under such data distribution can be biased over classes, resulting in unfairness. To address this issue, we propose a bilevel optimization algorithm FedFair to adaptively adjust the weights of local parameter updates from different clients, so as to enhance the fairness of the aggregated model. FedFair requires no extra information from clients other than that of the standard federated learning. The core idea of FedFair is to iteratively 1) increase the model accuracy and 2) preserve inter-class fairness. We perform experiments on class imbalanced heterogeneous federated datasets and show that our algorithm achieves improved fairness compared to existing methods.

[PDF of paper](#)

[Return to top](#)

Learning To Optimize Iterative Optimizers

Hugo S Gomes (Laval University)*; Benjamin Léger (Université Laval, Institut Intelligence et Données); Christian Gagné (Université Laval)

Location: Poster Room 2

ABSTRACT: Learning-to-optimize is a powerful paradigm that can be applied to a wide range of optimization and machine learning problems. For that purpose, we are proposing a general Learning-to-Optimize POMDP framework that allows the learning of meta-optimizers flexible enough to adapt to a wide range of problems of interest while achieving state-of-the-art performance in low-budget settings. Furthermore, the proposal employs a shared structure of most interactive black-box optimization heuristics jointly with a meta-loss function based on the performance of black-box stochastic algorithms across a variety of related optimization tasks. To evaluate its performance, we compare the learned optimizer's performance across many black-box numerical optimization tasks and practical hyperparameter tuning of machine learning models. Experimental evidence supports that the proposed meta-loss function allows a learned optimizer to adapt its search behavior in order to fit easily into a new context, resulting in a better generalization and higher sample efficiency than existing state-of-the-art methods.

[PDF of paper](#)

[Return to top](#)

Object-centric language emergence with Transformers

Tom Bosc (Mila)*; Pascal Vincent (MILA, FAIR)

Location: Poster Room 2

ABSTRACT: Computational approaches to the study of language emergence use optimisation techniques to evolve artificial languages. These artificial languages are then compared to natural languages. The deep learning-based language emergence literature focuses on the referential function of language, i.e. how artificial languages can pick out elements in a semantic space. For instance, researchers study the similarities between artificial utterances and noun phrases like "blue triangle". Here, we focus on the predication function of language, i.e. the expression of relations between entities. Predication is typically realized via sentences, for example "The blue triangle intersects the red circle."

To study predication, we task agents with communicating a relation and a set of entities along with their roles in the relation. The statistical dependencies between the meaning representation and the words are very sparse. For example, the same noun phrase denoting an entity ("John") can be used to communicate about the same entity in very different contexts ("John loves Mary" or "Andrew talks to John"). This sparsity seems to hold cross-linguistically, which enables us to propose metrics to measure naturalness of sentence structure.

Previous works have highlighted the importance of input representations and architectures on the properties of the artificial languages. Since natural language is used by humans, cognitively-plausible architectures and representations should help agents learn more natural languages. In particular, we focus on the notion of objectness. Children have access to object-centric representations at the time language is acquired. We compare the use of object-centric representations -- implemented using attention over whole object representations -- to unstructured representations and show that the former yield more natural languages.

[PDF of paper](#)

[Return to top](#)

Improving Continuous Normalizing Flows using a Multi-Resolution Framework

Vikram Voleti (Mila, University of Montreal)*

Location: Poster Room 2

ABSTRACT: Recent work has shown that Neural Ordinary Differential Equations (ODEs) can serve as generative models of images using the perspective of Continuous Normalizing Flows (CNFs). Such models offer exact likelihood calculation, and invertible generation/density estimation. In this work we introduce a Multi-Resolution variant of such models, by characterizing the conditional distribution over the additional

information required to generate a fine image that is consistent with the coarse image. We introduce a transformation between resolutions that allows for no change in the log likelihood. We show that this approach yields comparable likelihood values for various image datasets, with improved performance at higher resolutions, with fewer parameters, using only 1 GPU.

[PDF of paper](#)

[Return to top](#)