Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 PM — 1:45 PM

#	Poster Title	Presenter and Affiliation
1	OpenKBP: An international competition for predicting cancer treatment plans	Aaron Babier, University of Toronto, Vector Postgraduate Affiliate
2	Investigating the Impact of Spatio-temporal Augmentations on Self-supervised Audiovisual Representation Learning	Haider Al-Tahan, Western University, Student of Vector Faculty Affiliate
3	Learning with Less Label in Digital Pathology via Scribble Supervision	Eu Wern Teh, University of Guelph, Student of Vector Faculty Affiliate
4	Predicting Dreissenid Mussel Abundance in Nearshore Waters using Underwater Imagery and Deep Learning	Angus Galloway, University of Guelph, Student of Vector Faculty Affiliate
5	Task-based Optimization of Facial Landmarks for Pain Prediction and Action Unit Prediction	Abhishek Moturu, University of Toronto, Student of Vector Faculty Member
6	Quantifying Static and Dynamic Biases in Spatio-temporal Models	Mennatullah Siam, York University, Postdoctoral Fellow of Vector Faculty Affiliate
7	Towards Good Practices for Efficiently Annotating Large-Scale Image Classification Datasets	Andrew Liao, University of Toronto, Student of Vector Faculty Member
8	Anomaly Detection with Adversarially Learned Perturbations of Latent Feature Space	Vahid Reza Khazaie, Western University, Student of Vector Faculty Affiliate
9	Unconstrained Scene Generation with Locally Conditioned Radiance Fields	Terrance DeVries, University of Guelph, Student of Vector Faculty Member
10	Systematic Generalization in Learning Abstract Visual Analogies via Neural Structure Mapping	Shashank Shekhar, University of Guelph, Student of Vector Faculty Member

VECTOR INSTITUTE

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 PM — 1:45 PM

#	Poster Title	Presenter and Affiliation
11	Generative Compositional Augmentations for Scene Graph Prediction	Boris Knyazev, University of Guelph, Student of Vector Faculty Affiliate
12	Contrastive Learning for Sports Video: Unsupervised Player Classification	Maria Koshkina, York University, Student of Vector Faculty Affiliate
13	Robust UGV Localization in GPS-Denied Environments Using Correlated Semantic Segmentation Features	Jordy Sehn, University of Toronto, Vector Scholarship in Artificial Intelligence Recipient
14	Applying Computer Vision and Deep Learning to Semen Morphology & Motility Analysis	Gad Gad, Lakehead University, Vector Scholarship in Artificial Intelligence Recipient
15	Improving Self-supervised Learning with Hardness-aware Dynamic Curriculum Learning: An Application to Digital Pathology	Chetan Srinidhi, University of Toronto, Postdoctoral Fellow of Vector Faculty Affiliate
16	Computer Vision Applications in Anomaly Detection and Semantic Segmentation	Elham Ahmadi, Senior Data Scientist, RBC. Jinbao Ning, Technical Specialist, System Analyst, Thales Group
17	Automated Traffic Incident Detection with Two-Stream Neural Networks	Andrew Alberts, Data Scientist, Intact. Matthew Kowal, Graduate Researcher, York University / Vector Post-Graduate Affiliate
18	Identifying Clinically Relevant Features of Interest in Cholecystectomy Procedures	Kuldeep Panjwani, Software Engineer - AI Lab, Telus. Shuja Khalid, Graduate Researcher, Vector / University of Toronto
19	Transfer Learning for Efficient Video Classification/Detection	Raghav Goyal, PhD Student, University of British Columbia
20	Demo: Video Classification Using COVID-19 Ultrasound	Gerald Shen, Associate Applied Machine Learning Specialist, Vector Institute

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 1

OpenKBP: An international competition for predicting cancer treatment plans

Presenter: Aaron Babier, University of Toronto
Collaborators: Aaron Babier (Vector Institute, University of Toronto), Binghao Zhang (University of Toronto), Rafid Mahmood (Vector Institute, University of Toronto), Kevin L. Moore (University of California, San Diego), Thomas G. Purdie (UHN Princess Margaret Cancer Centre, University of Toronto), Andrea L. McNiven (UHN Princess Margaret Cancer Centre, University of Toronto), Timothy C.Y. Chan (Vector Institute, Techna Institute for the Advancement of Technology for Health, University of Toronto)

Radiotherapy is one of the primary modalities that is used to treat cancer, however, the conventional methods for developing radiotherapy treatment plans are inefficient. As the global cancer burden continues to increase, there is a growing need to develop new approaches for generating radiotherapy treatment plans. The most promising approach is knowledge-based planning (KBP), which is a process that uses artificial intelligence to develop treatment plans without human intervention. Although KBP research is flourishing, it is largely limited to institution-specific datasets and evaluation metrics, which makes comparing competing approaches difficult. The purpose of this project is to launch a large open-source dataset with standardized metrics to advance fair and consistent comparisons of KBP approaches

Poster 2

Investigating the Impact of Spatio-temporal Augmentations on Self-supervised Audiovisual Representation Learning

Presenter: Haider Al-Tahan, Western University
Collaborators: Yalda Mohsenzadeh (Vector Institute Faculty Affiliate, Department of Computer Science; University of Western Ontario)

Representation learning is a crucial component in the wide success of deep learning algorithms by disentangling compact independent high-level factors from low-level sensory data. Recently, contrastive self-supervised learning has been successful in learning rich visual and auditory representations by leveraging the inherent structure of unlabeled data. The wide success of these contrastive methods can be attributed to the underlying augmentations used to construct the contrasting views. Although prior method predominantly focused on spatial or temporal augmentations, we introduce spatio-temporal augmentations for learning audiovisual representations from unlabeled videos. Compared to self-supervised models pre-trained on only sampling-based temporal augmentation, self-supervised models pre-trained with our temporal augmentations lead to approximately 6.5% gain on linear classifier performance on AVE dataset. Lastly, we show that despite their simplicity, our proposed transformations work well across self-supervised learning frameworks (SimSiam, MoCoV3, etc).

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 3

Learning with Less Label in Digital Pathology via Scribble Supervision
Presenter: Eu Wern Teh, University of Guelph
Collaborators: Graham Taylor (Vector Institute, University of Guelph)

Digital Pathology (DP) is a field that involves the analysis of microscopic images. Supervised learning has made progress in DP for cancer classification and segmentation tasks in recent years; however, it requires a large number of labels to be effective. Unfortunately, labels from medical experts are scarce and extremely costly. On the other hand, it is relatively cheap to obtain weak labels from a layperson in the natural image (NI) domain. We leverage cheap annotations in the NI domain to help models to perform better in DP tasks. We hypothesize that models pre-trained on weak spatial labels can help DP models achieve higher performance when compared to models pre-trained on class labels. A demonstration of our proposed approach in cross-domain transfer learning. We first train a model by using scribble labels from the natural image (NI) domain. We transfer knowledge from the NI domain to the Digital Pathology (DP) domain by initializing the DP models with the pre-trained weights. Lastly, we train these DP models using labels provided by medical experts on cancer classification tasks.

Poster 4

Predicting Dreissenid Mussel Abundance in Nearshore Waters using Underwater Imagery and Deep Learning
Presenter: Angus Galloway, University of Guelph
Collaborators: Angus Galloway (1, 2); Dominique Brunet (3); Reza Valipour (3); Megan McCusker (3); Johann Biberhofer (3); Magdalena K. Sobol (1); Medhat Moussa (1); Graham W. Taylor (1, 2, 4), 1. School of Engineering, University of Guelph, 2. Vector Institute, 3. Environment and Climate Change Canada, Science & Technology Branch, 4. CIFAR"

Accurate and cost-effective dreissenid mussel abundance maps are vital to assess their ecological roles in aquatic systems. Here, a deep neural network (DNN) modeling framework using semantic segmentation was developed to advance critical information on the abundance distribution of two invasive mussel species Zebra and Quagga. DNN models were trained using images captured by an in situ underwater colour imaging technique. The accuracy of the method was assessed relative to manual laboratory counts of harvested mussels, their dry biomass, as well as live coverage estimated from fixed-size quadrats. Assessments performed on a test set collected from 2016–2018 show that model predictions explain 80% of the variance in Scuba diver estimated live coverage, 79% for biomass, and 71% for abundance (=179). When identical images were presented to humans and the DNN, the agreement in live mussel coverage predictions increased to 85%. Models generalize well to diverse underwater illuminations, camera orientations, and resolutions, but are adversely impacted by occluding vegetation and suspended sediment.

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 5

Task-based Optimization of Facial Landmarks for Pain Prediction and Action Unit Prediction

Presenter: Abhishek Moturu, University of Toronto
Collaborators: Siavash Rezaei (UHN KITE Research Institute, University of Toronto)
Babak Taati (UHN KITE Research Institute, University of Toronto)

Facial landmark detection is a crucial step in face-based tasks such as facial animation, facial recognition, facial registration, facial tracking, facial generation and reconstruction, head gesture analysis, and facial expression analysis. Although there is a lot of research on developing better models for facial landmark detection in varying conditions, poses, and settings, to our knowledge, there is no work on optimizing the landmarks for a given downstream task. The proposed method aims to improve the downstream tasks of pain prediction and action unit prediction by optimizing the landmarks. This is achieved by pre-training a landmark detection model and then using the landmarks to pre-train a regression perceptron for pain prediction and action unit prediction. Then, training the model end-to-end from an image to pain prediction and action unit prediction, via the landmarks, modifies the landmarks in such a way that improves the prediction tasks. We also study whether pain prediction and action unit prediction can be improved with this method in older adults with dementia

Poster 6

Quantifying Static and Dynamic Biases in Spatio-temporal Models

Presenter: Mennatullah Siam, York University
Collaborators: Matthew Kowal (Vector Postgraduate Affiliate, York University), Amirul Islam (Vector Postgraduate Affiliate, York University), Mennatullah Siam (York University), Neil D. B. Bruce (Guelph University), Richard Wildes (York University), Konstantinos Derpanis (Vector Institute, York University)

Deep spatiotemporal models are widely used in different computer vision tasks. However, there is limited understanding of what these models learn in the intermediate features. The question of whether these models are biased towards the spatial or the temporal factors has been the subject of only very limited study. We tackle that question, with a novel quantifiable approach for understanding spatial versus temporal biases. In action recognition, we study numerous architectures and discover that all networks are heavily spatially biased except, for the Fast branch of SlowFast architectures. In video object segmentation, most of these models are found to be spatially biased in the sensor fusion layers, except for the early fusion layer. Insights on the effect of training datasets is discussed, with key findings that questions mainstream assumptions on Diving-48 dataset being temporally biased. Our proposed interpretability for spatiotemporal models can be seen as a means towards Explainable AI, which has been recently encouraged in the European commission proposed regulations as well as by the Law Commission of Ontario. Thus, our work has important implications for industry as it seeks regulatory approval and compact algorithms for its AI products and services, e.g., mobile applications and autonomous driving.

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 7

Towards Good Practices for Efficiently Annotating Large-Scale Image Classification Datasets
Presenter: Andrew Liao, University of Toronto
Collaborators: Yuan-Hong Liao (Vector Institute, University of Toronto), Amlan Kar (Vector Institute, University of Toronto, NVIDIA), Sanja Fidler (Vector Institute, University of Toronto, NVIDIA)

Data is the engine of modern computer vision, which necessitates collecting large-scale datasets. This is expensive, and guaranteeing the quality of the labels is a major challenge. In this paper, we investigate efficient annotation strategies for collecting multi-class classification labels for a large collection of images. While methods that exploit learnt models for labeling exist, a surprisingly prevalent approach is to query humans for a fixed number of labels per datum and aggregate them, which is expensive. Building on prior work on online joint probabilistic modelling of human annotations and machine-generated beliefs, we propose modifications and best practices aimed at minimizing human labeling effort. Specifically, we make use of advances in self-supervised learning, view annotation as a semi-supervised learning problem, identify and mitigate pitfalls and ablate several key design choices to propose effective guidelines for labeling. Simulated experiments on a 125k image subset of the ImageNet100 show that it can be annotated to 80% top-1 accuracy with 0.35 annotations per image on average, a 2.7x and 6.7x improvement over prior work and manual annotation, respectively.

Poster 8

Anomaly Detection with Adversarially Learned Perturbations of Latent Feature Space
Presenter: Vahid Reza Khazaie, Western University
Collaborators: John Taylor Jewell (Western University), Yalda Mohsenzadeh (Western University, Vector Institute)

Anomaly detection is to identify samples that do not conform to the distribution of the training set. Due to the unavailability of anomalous data, training a supervised deep neural network is a cumbersome task. As such, unsupervised or self-supervised learning methods are preferred as a common approach to solve this task. Deep autoencoders have been broadly adopted as a base of many unsupervised anomaly detection methods. However, a notable shortcoming of deep autoencoders is that they provide insufficient representations for anomaly detection by generalizing to reconstruct outliers. In this work, we have designed an adversarial framework consisting of two competing components, an Adversarial Distorter, and an Autoencoder. The Adversarial Distorter is a convolutional encoder that learns to produce effective perturbations and the autoencoder is a deep convolutional neural network that aims to reconstruct the images from the perturbed latent feature space. The networks are trained with opposing goals in which the Adversarial Distorter produces perturbations that are applied to the encoder's latent feature space to maximize the reconstruction error and the autoencoder tries to neutralize the effect of these perturbations to minimize it. When applied to anomaly detection, the proposed method learns semantically richer representations due to applying perturbations to feature space.

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 9

Unconstrained Scene Generation with Locally Conditioned Radiance Fields
Presenter: Terrance DeVries, University of Guelph
Collaborators: Miguel Angel Bautista (Apple), Nitish Srivastava (Apple), Graham W. Taylor (University of Guelph, Vector Institute), Joshua M. Susskind (Apple)

We tackle the challenge of learning a distribution over complex, realistic, indoor scenes. In this work, we introduce Generative Scene Networks (GSN), which learns to decompose scenes into a collection of many local radiance fields that can be rendered from a free moving camera. Our model can be used as a prior to generate new scenes, or to complete a scene given only sparse 2D observations. Recent work has shown that generative models of radiance fields can capture properties such as multi-view consistency and view-dependent lighting. However, these models are specialized for constrained viewing of single objects, such as cars or faces. Due to the size and complexity of realistic indoor environments, existing models lack the representational capacity to adequately capture them. Our decomposition scheme scales to larger and more complex scenes while preserving details and diversity, and the learned prior enables high quality rendering from view-points that are significantly different from observed view-points. When compared to existing models, GSN produce quantitatively higher-quality scene renderings across several different scene datasets.

Poster 10

Systematic Generalization in Learning Abstract Visual Analogies via Neural Structure Mapping
Presenter: Shashank Shekhar, University of Guelph
Collaborators: Graham Taylor (Vector Institute, University of Guelph)

Building conceptual abstractions from sensory information and then systematically reasoning about these abstractions is central to human intelligence (Lake et al., 2015). This process relies strongly on analogical reasoning, where the syntactic knowledge about an entity's features is applied across perceptual domains (Mitchell, 2021). In order to test the ability of neural networks to build and reason about analogies, Hill et al. (2019) constructed a dataset on learning analogies from Raven's Progressive Matrices (RPMs) (Raven, 2000), which is an abstract visual reasoning test of fluid intelligence. In particular, this dataset serves as a testbed for measuring systematic generalization in learning visual analogies that can generalize to novel object and attribute values (visual domains). In this work, we introduce a two-stage neural net framework to learn visual analogies from RPMs. We show that our engine is able to better learn visual analogies capable of systematic generalization across visual domains. Next, we show the importance of inferring the correct structure (relation) on the downstream performance of the engine. We also show empirically that the layouts we chose of our engine align well with the corresponding relation that they are chosen for. Finally, we show that having an adaptive engine that uses neural module nets outperforms having a static engine for analogical reasoning.

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 11

Generative Compositional Augmentations for Scene Graph Prediction
Presenter: Boris Knyazev, University of Guelph
Collaborators: Harm de Vries (Element AI), Cătălina Cangea (University of Cambridge), Graham W. Taylor (University of Guelph, Vector Institute), Aaron Courville (Université de Montréal, Mila), Eugene Belilovsky (Concordia University, Mila)

Inferring objects and their relationships from an image in the form of a scene graph is useful in many applications at the intersection of vision and language. We consider a challenging problem of compositional generalization that emerges in this task due to a long tail data distribution. Current scene graph generation models are trained on a tiny fraction of the distribution corresponding to the most frequent compositions. However, test images might contain zero- and few-shot compositions of objects and relationships. Despite each of the object categories and the predicate (e.g. 'on') being frequent in the training data, the models often fail to properly understand such unseen or rare compositions. We propose a method to synthesize rare yet plausible scene graphs by perturbing real ones. We then propose and empirically study a model based on conditional generative adversarial networks (GANs) that allows us to generate visual features of perturbed scene graphs and learn from them in a joint fashion. When evaluated on the Visual Genome dataset, our approach yields marginal, but consistent improvements in zero- and few-shot metrics. We analyze the limitations of our approach indicating promising directions for future research.

Poster 12

Contrastive Learning for Sports Video: Unsupervised Player Classification
Presenter: Maria Koshkina, York University
Collaborators: James Elder (York University), Hemanth Pidaparthy (York University)

We address the problem of unsupervised classification of players in a team sport according to their team affiliation, when jersey colours and design are not known a priori. We adopt a contrastive learning approach in which an embedding network learns to maximize the distance between representations of players on different teams relative to players on the same team, in a purely unsupervised fashion, without any labelled data. We evaluate the approach using a new hockey dataset and find that it outperforms prior unsupervised approaches by a substantial margin, particularly for real-time application when only a small number of frames are available for unsupervised learning before team assignments must be made. Remarkably, we show that our contrastive method achieves 94% accuracy after unsupervised training on only a single frame, with accuracy rising to 97% within 500 frames (17 seconds of game time). We further demonstrate how accurate team classification allows accurate team-conditional heat maps of player positioning to be computed. Source code is available at https://github.com/mkoshkina/teamId.

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 13

Robust UGV Localization in GPS-Denied Environments Using Correlated Semantic Segmentation Features

Presenter: Jordy Sehn, University of Toronto
Collaborators: Jack Collier (Defence Research and Development Canada), Simon Monckton (Defence Research and Development Canada)

Reliability is critical when examining the effectiveness and safety of autonomous navigation systems in densely populated urban environments. As many traditional commercial unmanned ground vehicle (UGV) systems require a stable global positioning system (GPS) link for state estimation, interference due to tall buildings and the emergence of cheap, publicly accessible signal jamming technologies present a major barrier to overcome before such systems become feasible. This paper presents an alternative, more dependable pipeline for robot localization in dynamic urban topographies under GPS-denied conditions using a pre-processed map with specially segmented light detection and ranging (LiDAR) data. A high altitude quad-rotor drone was used to take a single nadir image of the urban environment and is fed to a classical convolutional neural network (CNN) to segment buildings from the image. On the ground robot, a point cloud segmentation neural network based on the “LU-Net” architecture was optimized using an efficient matrix based 3D feature extraction pre-processing stage enabling real-time operation. The network was then trained to classify buildings, ground, vehicles, vegetation, and people using a labelled point cloud dataset. Isolating the building points allow us to filter out the majority of the dynamic obstacles in the scene which otherwise hinder localization performance in changing environments. Together, the map and segmented building point clouds are used in a Monte Carlo particle filter algorithm to localize the robot.

Poster 14

Applying Computer Vision and Deep Learning to Semen Morphology & Motility Analysis

Presenter: Gad Gad, Lakehead University
Collaborators: Zubair Fadlullah (Associate professor, Dept. of Computer science, Lakehead University)

Semen morphology & motility test aims to characterize male sperm in terms of count, size, speed, path, and shape then compare these results with norms published by organizations like the World Health Organization (WHO) to assess fertility. Morphology test studies the morphology (form) of the sperm from a high-resolution image of dyed sperms. On the other hand, motility test studies the count and motility (movement) of sperms from a short video clip (2 seconds) of moving sperms. In this applied project, morphology features of sperms are extracted from an image of a dyed sperm(s) and motility features are extracted from a video clip of moving sperms. Assessment of the extracted characteristics is a medical subject and out of the scope of this project. Different methods were investigated and compared to account for different parameters like annotated data availability, annotation effort, accuracy

Computer Vision Symposium

Poster Session

Thursday, October 28th, 2021
12:45 AM — 1:45 PM

Poster 15

Improving Self-supervised Learning with Hardness-aware Dynamic Curriculum Learning: An Application to Digital Pathology

Presenter: Chetan Srinidhi, University of Toronto

Collaborators: Anne L. Martel (Sunnybrook Research Institute, University of Toronto)

Self-supervised learning (SSL) has recently shown tremendous potential to learn generic visual representations useful for many image analysis tasks. Despite their notable success, the existing SSL methods fail to generalize to downstream tasks when the number of labeled training instances is small or if the domain shift between the transfer domains is significant. In this paper, we attempt to improve self-supervised pretrained representations through the lens of curriculum learning by proposing a hardness-aware dynamic curriculum learning (HaDCL) approach. We discover that by progressive stage-wise curriculum learning, the pretrained representations are significantly enhanced and adaptable to both in-domain and out- of-domain distribution data. We performed extensive validation on three histology benchmark datasets on both patch-wise and slide-level classification problems. Further, we empirically show that our approach is more generic and adaptable to any SSL methods and does not impose any additional overhead complexity. Besides, we also outline the role of patch-based versus slide-based curriculum learning in histopathology to provide practical insights into the success of curriculum based fine-tuning of SSL methods.

Poster 16 (Vector's CV Project Presenter)

Computer Vision Applications in Anomaly Detection and Semantic Segmentation

Presenters: Elham Ahmadi, Senior Data Scientist, RBC
Jinbao Ning, Technical Specialist, System Analyst, Thales Group

Poster 17 (Vector's CV Project Presenter)

Automated Traffic Incident Detection with Two-Stream Neural Networks

Presenters: Andrew Alberts, Data Scientist, Intact
Matthew Kowal, Graduate Researcher, York University / Vector Post-Graduate Affiliate

Poster 18 (Vector's CV Project Presenter)

Identifying Clinically Relevant Features of Interest in Cholecystectomy Procedures

Presenters: Kuldeep Panjwani, Software Engineer - AI Lab, Telus
Shuja Khalid, Graduate Researcher, Vector / University of Toronto

Poster 19 (Vector's CV Project Presenter)

Transfer Learning for Efficient Video Classification/Detection

Presenter: Raghav Goyal, PhD Student, University of British Columbia

Poster 20 (Demo)

Video Classification Using COVID-19 Ultrasound

Presenter: Gerald Shen, Associate Applied Machine Learning Specialist, Vector Institute

VECTOR INSTITUTE