**Top 5 Conferences in Machine Learning**

**International Conference on Machine Learning(ICML)**

* ICML is more generic compared to NeurIPS, ICLR.
* ICML has a small bias towards machine learing theory.
* Pure application paper are low in number at ICML.
* **Analytic Marching: An Analytic Meshing Solution from Deep Implicit Surface Networks held at 2020.**

**Abstarct:**

This paper studies a problem of learning surface mesh via implicit functions in an emerging field of deep learning surface reconstruction, where implicit functions are popularly implemented as multi-layer perceptrons (MLPs) with rectified linear units (ReLU). To achieve meshing from learned implicit functions, existing methods adopt the de-facto standard algorithm of marching cubes; while promising, they suffer from loss of precision learned in the MLPs, due to the discretization nature of marching cubes. Motivated by the knowledge that a ReLU based MLP partitions its input space into a number of linear regions, we identify from these regions analytic cells and analytic faces that are associated with zero-level isosurface of the implicit function, and characterize the theoretical conditions under which the identified analytic faces are guaranteed to connect and form a closed, piecewise planar surface. Based on our theorem, we propose a naturally parallelizable algorithm of analytic marching, which marches among analytic cells to exactly recover the mesh captured by a learned MLP. Experiments on deep learning mesh reconstruction verify the advantages of our algorithm over existing ones.

* **Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations held at 2019**

**Abstract**

The key idea behind the unsupervised learning of disentangled representations is that real-world data is generated by a few explanatory factors of variation which can be recovered by unsupervised learning algorithms. In this paper, we provide a sober look at recent progress in the field and challenge some common assumptions. We first theoretically show that the unsupervised learning of disentangled representations is fundamentally impossible without inductive biases on both the models and the data. Then, we train more than 12 000 models covering most prominent methods and evaluation metrics in a reproducible large-scale experimental study on seven different data sets. We observe that while the different methods successfully enforce properties “encouraged” by the corresponding losses, well-disentangled models seemingly cannot be identified without supervision. Furthermore, increased disentanglement does not seem to lead to a decreased sample complexity of learning for downstream tasks. Our results suggest that future work on disentanglement learning should be explicit about the role of inductive biases and (implicit) supervision, investigate concrete benefits of enforcing disentanglement of the learned representations, and consider a reproducible experimental setup covering several data sets.

* **Delayed Impact of Fair Machine Learning held at 2018**

**Abstract**

Fairness in machine learning has predominantly been studied in static classification settings without concern for how decisions change the underlying population over time. Conventional wisdom suggests that fairness criteria promote the long-term well-being of those groups they aim to protect. We study how static fairness criteria interact with temporal indicators of well-being, such as long-term improvement, stagnation, and decline in a variable of interest. We demonstrate that even in a one-step feedback model, common fairness criteria in general do not promote improvement over time, and may in fact cause harm in cases where an unconstrained objective would not. We completely characterize the delayed impact of three standard criteria, contrasting the regimes in which these exhibit qualitatively different behavior. In addition, we find that a natural form of measurement error broadens the regime in which fairness criteria perform favorably. Our results highlight the importance of measurement and temporal modeling in the evaluation of fairness criteria, suggesting a range of new challenges and trade-offs.

**International Conference on Learning and Representation(ICLR)**

* Foused on Deep Learning and Learning Representations.
* More recent conference started from 2013 onwards.
* Lesser submission and higher acceptance rate.
* **Plug and Play Language Models held at 2020**

**Abstract**

Large transformer-based language models (LMs) trained on huge text corpora have shown unparalleled generation capabilities. However, controlling attributes of the generated language (e.g. switching topic or sentiment) is difficult without modifying the model architecture or fine-tuning on attribute-specific data and entailing the significant cost of retraining. We propose a simple alternative: the Plug and Play Language Model (PPLM) for controllable language generation, which combines a pretrained LM with one or more simple attribute classifiers that guide text generation without any further training of the LM. In the canonical scenario we present, the attribute models are simple classifiers consisting of a user-specified bag of words or a single learned layer with 100,000 times fewer parameters than the LM. Sampling entails a forward and backward pass in which gradients from the attribute model push the LM’s hidden activations and thus guide the generation. Model samples demonstrate control over a range of topics and sentiment styles, and extensive automated and human annotated evaluations show attribute alignment and fluency. PPLMs are flexible in that any combination of differentiable attribute models may be used to steer text generation, which will allow for diverse and creative applications beyond the examples given in this paper.

* **Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks held at 2019**

**Abstract**

Natural language is hierarchically structured: smaller units (e.g., phrases) are nested within larger units (e.g., clauses). When a larger constituent ends, all of the smaller constituents that are nested within it must also be closed. While the standard LSTM architecture allows different neurons to track information at different time scales, it does not have an explicit bias towards modeling a hierarchy of constituents. This paper proposes to add such inductive bias by ordering the neurons; a vector of master input and forget gates ensures that when a given neuron is updated, all the neurons that follow it in the ordering are also updated. Our novel recurrent architecture, ordered neurons LSTM (ON-LSTM), achieves good performance on four different tasks: language modeling, unsupervised parsing, targeted syntactic evaluation, and logical inference.

* **Deep Complex Networks held at 2018**

**Abstract**

At present, the vast majority of building blocks, techniques, and architectures for deep learning are based on real-valued operations and representations. However, recent work on recurrent neural networks and older fundamental theoretical analysis suggests that complex numbers could have a richer representational capacity and could also facilitate noise-robust memory retrieval mechanisms. Despite their attractive properties and potential for opening up entirely new neural architectures, complex-valued deep neural networks have been marginalized due to the absence of the building blocks required to design such models. In this work, we provide the key atomic components for complex-valued deep neural networks and apply them to convolutional feed-forward networks. More precisely, we rely on complex convolutions and present algorithms for complex batch-normalization, complex weight initialization strategies for complex-valued neural nets and we use them in experiments with end-to-end training schemes. We demonstrate that such complex-valued models are competitive with their real-valued counterparts. We test deep complex models on several computer vision tasks, on music transcription using the MusicNet dataset and on Speech spectrum prediction using TIMIT. We achieve state-of-the-art performance on these audio-related tasks.

**AAAI Conference on Artificial Intelligence**

* AAAI is a generic conference.
* It is an old conference since 1980.
* AAAI is quite popular with Industry.
* **WINOGRANDE: An Adversarial Winograd Schema Challenge at Scale held at 2020**

**Abstract**

The Winograd Schema Challenge (WSC) (Levesque, Davis, and Morgenstern 2011), a benchmark for commonsense reasoning, is a set of 273 expert-crafted pronoun resolution problems originally designed to be unsolvable for statistical models that rely on selectional preferences or word associations. However, recent advances in neural language models have already reached around 90% accuracy on variants of WSC. This raises an important question whether these models have truly acquired robust commonsense capabilities or whether they rely on spurious biases in the datasets that lead to an overestimation of the true capabilities of machine commonsense. To investigate this question, we introduce WINOGRANDE, a large-scale dataset of 44k problems, inspired by the original WSC design, but adjusted to improve both the scale and the hardness of the dataset. The key steps of the dataset construction consist of (1) a carefully designed crowdsourcing procedure, followed by (2) systematic bias reduction using a novel AFLITE algorithm that generalizes human-detectable word associations to machine-detectable embedding associations. The best state-of-the-art methods on WINOGRANDE achieve 59.4 – 79.1%, which are ∼15-35% (absolute) below human performance of 94.0%, depending on the amount of the training data allowed (2% – 100% respectively). Furthermore, we establish new state-of-the-art results on five related benchmarks — WSC (→ 90.1%), DPR (→ 93.1%), COPA(→ 90.6%), KnowRef (→ 85.6%), and Winogender (→ 97.1%). These results have dual implications: on one hand, they demonstrate the effectiveness of WINOGRANDE when used as a resource for transfer learning. On the other hand, they raise a concern that we are likely to be overestimating the true capabilities of machine commonsense across all these benchmarks. We emphasize the importance of algorithmic bias reduction in existing and future benchmarks to mitigate such overestimation.

* **Combine Tree-Search Methods in Reinforcement Learning held at 2019**

**Abstract**

Finite-horizon lookahead policies are abundantly used in Reinforcement Learning and demonstrate impressive empirical success. Usually, the lookahead policies are implemented with specific planning methods such as Monte Carlo Tree Search (e.g. in AlphaZero (Silver et al. 2017b)). Referring to the planning problem as tree search, a reasonable practice in these implementations is to back up the value only at the leaves while the information obtained at the root is not leveraged other than for updating the policy. Here, we question the potency of this approach. Namely, the latter procedure is non-contractive in general, and its convergence is not guaranteed. Our proposed enhancement is straightforward and simple: use the return from the optimal tree path to back up the values at the descendants of the root. This leads to a γ h -contracting procedure, where γ is the discount factor and h is the tree depth. To establish our results, we first introduce a notion called multiple-step greedy consistency. We then provide convergence rates for two algorithmic instantiations of the above enhancement in the presence of noise injected to both the tree search stage and value estimation stage.

* **Elastic-Net Attacks to Deep Neural Networks via Adversarial Examples held at 2018**

**Abstract**

Recent studies have highlighted the vulnerability of deep neural networks (DNNs) to adversarial examples — a visually indistinguishable adversarial image can easily be crafted to cause a well-trained model to misclassify. Existing methods for crafting adversarial examples are based on L2 and L∞ distortion metrics. However, despite the fact that L1 distortion accounts for the total variation and encourages sparsity in the perturbation, little has been developed for crafting L1-based adversarial examples. In this paper, we formulate the process of attacking DNNs via adversarial examples as an elastic-net regularized optimization problem. Our elastic-net attacks to DNNs (EAD) feature L1-oriented adversarial examples and include the state-of-the-art L2 attack as a special case. Experimental results on MNIST, CIFAR10 and ImageNet show that EAD can yield a distinct set of adversarial examples with small L1 distortion and attains similar attack performance to the state-of-the-art methods in different attack scenarios. More importantly, EAD leads to improved attack transferability and complements adversarial training for DNNs, suggesting novel insights on leveraging L1 distortion in adversarial machine learning and security implications of DNNs.

**International Joint Conference on Artificial Intelligence(IJCAI)**

* IJCAI is a gathering of Artificial Intelligence.
* It is an old conference since 1969.
* IJCAI is quite popular with Industry.
* **Multi-Attribute Diverse Matching held at 2020**

**Abstract**

Bipartite b-matching, where agents on one side of a market are matched to one or more agents or items on the other, is a classical model that is used in myriad application areas such as healthcare, advertising, education, and general resource allocation. Traditionally, the primary goal of such models is to maximize a linear function of the constituent matches (e.g., linear social welfare maximization) subject to some constraints. Recent work has studied a new goal of balancing whole-match diversity and economic efficiency, where the objective is instead a monotone submodular function over the matching. Basic versions of this problem are solvable in polynomial time. In this work, we prove that the problem of simultaneously maximizing diversity along several features (e.g., country of citizenship, gender, skills) is NP-hard. To address this problem, we develop the first combinatorial algorithm that constructs provably-optimal diverse b-matchings in pseudo-polynomial time. We also provide a Mixed-Integer Quadratic formulation for the same problem and show that our method guarantees optimal solutions and takes less computation time for a reviewer assignment application. The source code is made available at https://github.com/ faezahmed/diverse matching.

* **Boosting for Comparison-Based Learning held at 2019**

**Abstract**

We consider the problem of classification in a comparison-based setting: given a set of objects, we only have access to triplet comparisons of the form “object A is closer to object B than to object C.” In this paper we introduce TripletBoost, a new method that can learn a classifier just from such triplet comparisons. The main idea is to aggregate the triplets information into weak classifiers, which can subsequently be boosted to a strong classifier. Our method has two main advantages: (i) it is applicable to data from any metric space, and (ii) it can deal with large scale problems using only passively obtained and noisy triplets. We derive theoretical generalization guarantees and a lower bound on the number of necessary triplets, and we empirically show that our method is both competitive with state of the art approaches and resistant to noise.

* **A Degeneracy Framework for Graph Similarity held at 2018**

**Abstract**

The problem of accurately measuring the similarity between graphs is at the core of many applications in a variety of disciplines. Most existing methods for graph similarity focus either on local or on global properties of graphs. However, even if graphs seem very similar from a local or a global perspective, they may exhibit different structure at different scales. In this paper, we present a general framework for graph similarity which takes into account structure at multiple different scales. The proposed framework capitalizes on the wellknown k-core decomposition of graphs in order to build a hierarchy of nested subgraphs. We apply the framework to derive variants of four graph kernels, namely graphlet kernel, shortest-path kernel, Weisfeiler-Lehman subtree kernel, and pyramid match graph kernel. The framework is not limited to graph kernels, but can be applied to any graph comparison algorithm. The proposed framework is evaluated on several benchmark datasets for graph classification. In most cases, the corebased kernels achieve significant improvements in terms of classification accuracy over the base kernels, while their time complexity remains very attractive.

**Conference on Computer Vision and Pattern Recognition(CVPR)**

* 3D computer vision.
* Vision applications systems , vision for robotics and autonomous vehicles.
* Visual reasoning and logical representation.
* **Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild held at 2020**

**Abstract**

We propose a method to learn 3D deformable object categories from raw single-view images, without external supervision. The method is based on an autoencoder that factors each input image into depth, albedo, viewpoint and illumination. In order to disentangle these components without supervision, we use the fact that many object categories have, at least in principle, a symmetric structure. We show that reasoning about illumination allows us to exploit the underlying object symmetry even if the appearance is not symmetric due to shading. Furthermore, we model objects that are probably, but not certainly, symmetric by predicting a symmetry probability map, learned end-to-end with the other components of the model. Our experiments show that this method can recover very accurately the 3D shape of human faces, cat faces and cars from single-view images, without any supervision or a prior shape model. On benchmarks, we demonstrate superior accuracy compared to another method that uses supervision at the level of 2D image correspondences.

* **FEELVOS: Fast End-to-End Embedding Learning for Video Object Segmentation held at 2019**

**Abstract**

Many of the recent successful methods for video object segmentation (VOS) are overly complicated, heavily rely on fine-tuning on the first frame, and/or are slow, and are hence of limited practical use. In this work, we propose FEELVOS as a simple and fast method which does not rely on fine-tuning. In order to segment a video, for each frame FEELVOS uses a semantic pixel-wise embedding together with a global and a local matching mechanism to transfer information from the first frame and from the previous frame of the video to the current frame. In contrast to previous work, our embedding is only used as an internal guidance of a convolutional network. Our novel dynamic segmentation head allows us to train the network, including the embedding, end-to-end for the multiple object segmentation task with a cross entropy loss. We achieve a new state of the art in video object segmentation without fine-tuning with a J &F measure of 71.5% on the DAVIS 2017 validation set

* **Revisiting Oxford and Paris: Large-Scale Image Retrieval Benchmarking held at 2018**

**Abstract**

In this paper we address issues with image retrieval benchmarking on standard and popular Oxford 5k and Paris 6k datasets. In particular, annotation errors, the size of the dataset, and the level of challenge are addressed: new annotation for both datasets is created with an extra attention to the reliability of the ground truth. Three new protocols of varying difficulty are introduced. The protocols allow fair comparison between different methods, including those using a dataset pre-processing stage. For each dataset, 15 new challenging queries are introduced. Finally, a new set of 1M hard, semi-automatically cleaned distractors is selected. An extensive1 comparison of the state-of-the-art methods is performed on the new benchmark. Different types of methods are evaluated, ranging from local-feature-based to modern CNN based methods. The best results are achieved by taking the best of the two worlds. Most importantly, image retrieval appears far from being solved.