2018

SL N.O	CONFERENCE	PAPER	ABSTRACT
1	International Conference on Machine Learning (ICML)	Obfuscated Gradients Give a False Sense of Security	We identify obfuscated gradients, a kind of gradient masking, as a phenomenon that leads to a false sense of security in defenses against adversarial examples. While defenses that cause obfuscated gradients appear to defeat iterative optimization-based attacks, we find defenses relying on this effect can be circumvented. We describe characteristic behaviors of defenses exhibiting the effect, and for each of the three types of obfuscated gradients we discover, we develop attack techniques to overcome it. In a case study, examining non-certified white-box-secure defenses at ICLR 2018, we find obfuscated gradients are a common occurrence, with 7 of 9 defenses relying on obfuscated gradients. Our new attacks successfully circumvent 6 completely, and 1 partially, in the original threat model each paper considers.
2	International Conference on Machine Learning (ICML)	Delayed Impact of Fair Machine Learning	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.
3	International	Near Optimal Frequent	Given a large matrix $A \in R^{n \times d}$, we consider the

	Conference on Machine Learning (ICML)	Directions for Sketching Dense and Sparse Matrices	problem of computing a sketch matrix B ∈ R `×d which is significantly smaller than but still well approximates A. We are interested in minimizing the covariance error kA T A − B T Bk2. We consider the problems in the streaming model, where the algorithm can only make one pass over the input with limited working space. The popular Frequent Directions algorithm of (Liberty, 2013) and its variants achieve optimal space-error tradeoff. However, whether the running time can be improved remains an unanswered question. In this paper, we almost settle the time complexity of this problem. In particular, we provide new space-optimal algorithms with faster running times. Moreover, we also show that the running times of our algorithms are near-optimal unless the state-of-the-art running time of matrix multiplication can be improved significantly
4	International Conference on Machine Learning (ICML)	Fairness Without Demographics in Repeated Loss Minimization	Machine learning models (e.g., speech recognizers) are usually trained to minimize average loss, which results in representation disparity---minority groups (e.g., non-native speakers) contribute less to the training objective and thus tend to suffer higher loss. Worse, as model accuracy affects user retention, a minority group can shrink over time. In this paper, we first show that the status quo of empirical risk minimization (ERM) amplifies representation disparity over time, which can even make initially fair models unfair. To mitigate this, we develop an approach based on distributionally robust optimization (DRO), which minimizes the worst case risk over all distributions close to the empirical distribution. We prove that this approach controls the risk of the minority group at each time step, in the spirit of Rawlsian distributive justice, while remaining oblivious to the identity of the groups. We demonstrate that DRO prevents disparity amplification on examples where ERM fails, and show improvements in minority group user satisfaction in a real-world text autocomplete task.
5	ECVV	Group Normalization	Batch Normalization (BN) is a milestone technique in the development of deep learning, enabling various networks to train. However, normalizing along the batch dimension introduces problems — BN’s error increases rapidly when the batch size becomes smaller, caused by inaccurate batch statistics estimation. This limits BN’s usage for training larger models and transferring features to computer vision tasks including detection, segmentation, and video, which require small batches constrained by memory consumption. In this paper, we present Group Normalization (GN) as a simple alternative to BN. GN divides the channels into groups and computes within each group the mean and variance for normalization. GN’s computation is independent of batch sizes, and its accuracy is stable in a wide range of batch sizes. On ResNet-50 trained in ImageNet, GN has 10.6% lower

			error than its BN counterpart when using a batch size of 2; when using typical batch sizes, GN is comparably good with BN and outperforms other normalization variants. Moreover, GN can be naturally transferred from pre-training to fine-tuning. GN can outperform its BN-based counterparts for object detection and segmentation in COCO, and for video classification in Kinetics, showing that GN can effectively replace the powerful BN in a variety of tasks. GN can be easily implemented by a few lines of code.
6	ECVV	Acquisition of Localization Confidence for Accurate Object Detection	Modern CNN-based object detectors rely on bounding box regression and non-maximum suppression to localize objects. While the probabilities for class labels naturally reflect classification confidence, localization confidence is absent. This makes properly localized bounding boxes degenerate during iterative regression or even suppressed during NMS. In the paper we propose IoU-Net learning to predict the IoU between each detected bounding box and the matched ground-truth. The network acquires this confidence of localization, which improves the NMS procedure by preserving accurately localized bounding boxes. Furthermore, an optimization-based bounding box refinement method is proposed, where the predicted IoU is formulated as the objective. Extensive experiments on the MS-COCO dataset show the effectiveness of IoU-Net, as well as its compatibility with and adaptivity to several state-of-the-art object detectors.
7	ICLR	SPHERICAL CNNS	Convolutional Neural Networks (CNNs) have become the method of choice for learning problems involving 2D planar images. However, a number of problems of recent interest have created a demand for models that can analyze spherical images. Examples include omnidirectional vision for drones, robots, and autonomous cars, molecular regression problems, and global weather and climate modelling. A naive application of convolutional networks to a planar projection of the spherical signal is destined to fail, because the space-varying distortions introduced by such a projection will make translational weight sharing ineffective. In this paper we introduce the building blocks for constructing spherical CNNs. We propose a definition for the spherical cross-correlation that is both expressive and rotation-equivariant. The spherical correlation satisfies a generalized Fourier theorem, which allows us to compute it efficiently using a generalized (non-commutative) Fast Fourier Transform (FFT) algorithm. We demonstrate the computational efficiency, numerical accuracy, and effectiveness of spherical CNNs applied to 3D model recognition and atomization energy regression.