

## 用于训练二值神经网络的去噪监督学习方法

## Training Binary Neural Networks through Learning with Noisy Supervision

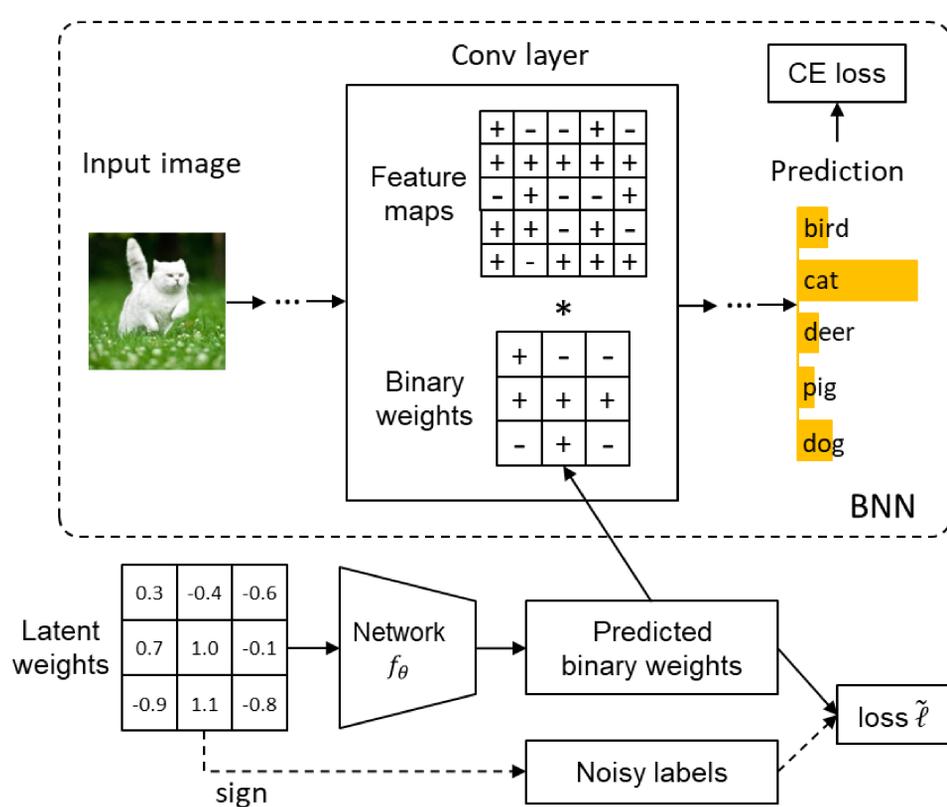
韩凯 王云鹤 许奕星 吴恩华

International Conference on Machine Learning (ICML 2020)

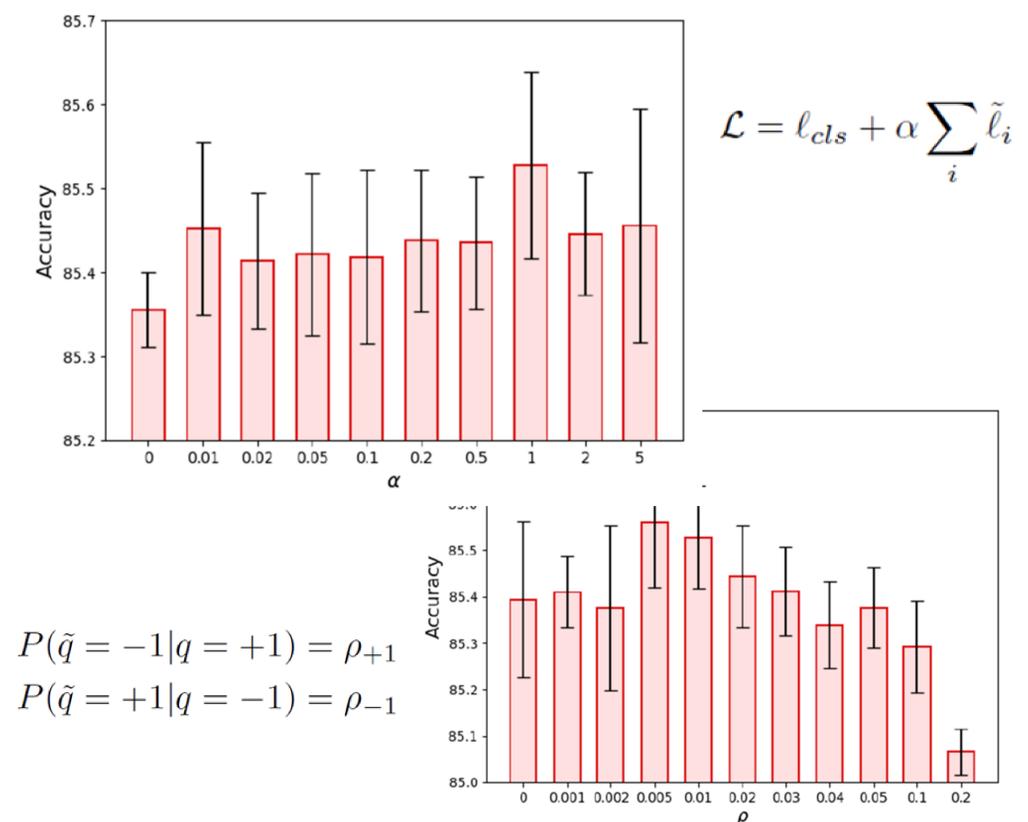
卷积神经网络已经在计算机视觉的诸多领域被证明是一种非常有效的方法。卷积神经网络需要耗费大量的计算量来完成对输入数据的处理。二值神经网络使用二值化的值来表示卷积权重和激活值，使得卷积操作可通过快速的逻辑计算完成，从而极大地加速神经网络的推理过程。

本文从学习的角度对二值神经网络上的二值化运算进行研究。与经典的手工规则（例如Sign函数）将神经元二值化相比，我们提出映射模型（Mapping model）来学习从全精度神经元到二值神经元的映射。这里，每个权值不是独立二值化，而是将权值张量作为一个整体来完成二值化，充分考虑权值之间的关联性。为了帮助训练二值化映射模型，我们将传统Sign函数量化的神经元视为一些辅助监督信号，其虽然有噪声但仍具有指导意义。因此，我们引入了无偏估计器以减轻噪声的影响。在基准数据集上的实验结果表明，所提出的二值化技术具有广泛的有效性。

## 算法框架



## 超参研究



## ImageNet大规模图像分类结果

Method	W	A	Memory	FLOPs	Top-1	Top-5
ResNet-18 (He et al., 2016)	32	32	374 Mbit	1810 M	69.6%	89.2%
BWN (Rastegari et al., 2016)	1	32	34 Mbit	975 M	60.8%	83.0%
HWGQ (Cai et al., 2017)	1	2	34 Mbit	193 M	59.6%	82.2%
TBN (Wan et al., 2018)	1	2	34 Mbit	193 M	55.6%	79.0%
BinaryNet (Hubara et al., 2016)	1	1	28 Mbit	149 M	42.2%	67.1%
Dorefa-Net (Zhou et al., 2016) <sup>†</sup>	1	1	34 Mbit	163 M	52.5%	76.7%
XNOR-Net (Rastegari et al., 2016)	1	1	34 Mbit	167 M	51.2%	73.2%
Bireal-Net (Liu et al., 2018)	1	1	34 Mbit	163 M	56.4%	79.5%
Bireal-Net (Liu et al., 2018)+PReLU (Baseline)	1	1	34 Mbit	163 M	59.0%	81.3%
PCNN ( $J=1$ ) (Gu et al., 2019)	1	1	34 Mbit	167 M	57.3%	80.0%
Quantization networks (Yang et al., 2019)	1	1	34 Mbit	163 M	53.6%	75.3%
Bop (Helwegen et al., 2019)	1	1	34 Mbit	163 M	54.2%	77.2%
GBCN (Liu et al., 2019)	1	1	34 Mbit	167 M	57.8%	80.9%
IR-Net (Qin et al., 2020)	1	1	34 Mbit	163 M	58.1%	80.0%
Ours	1	1	34 Mbit	163 M	59.4%	81.7%