Regularized Binary Network Training

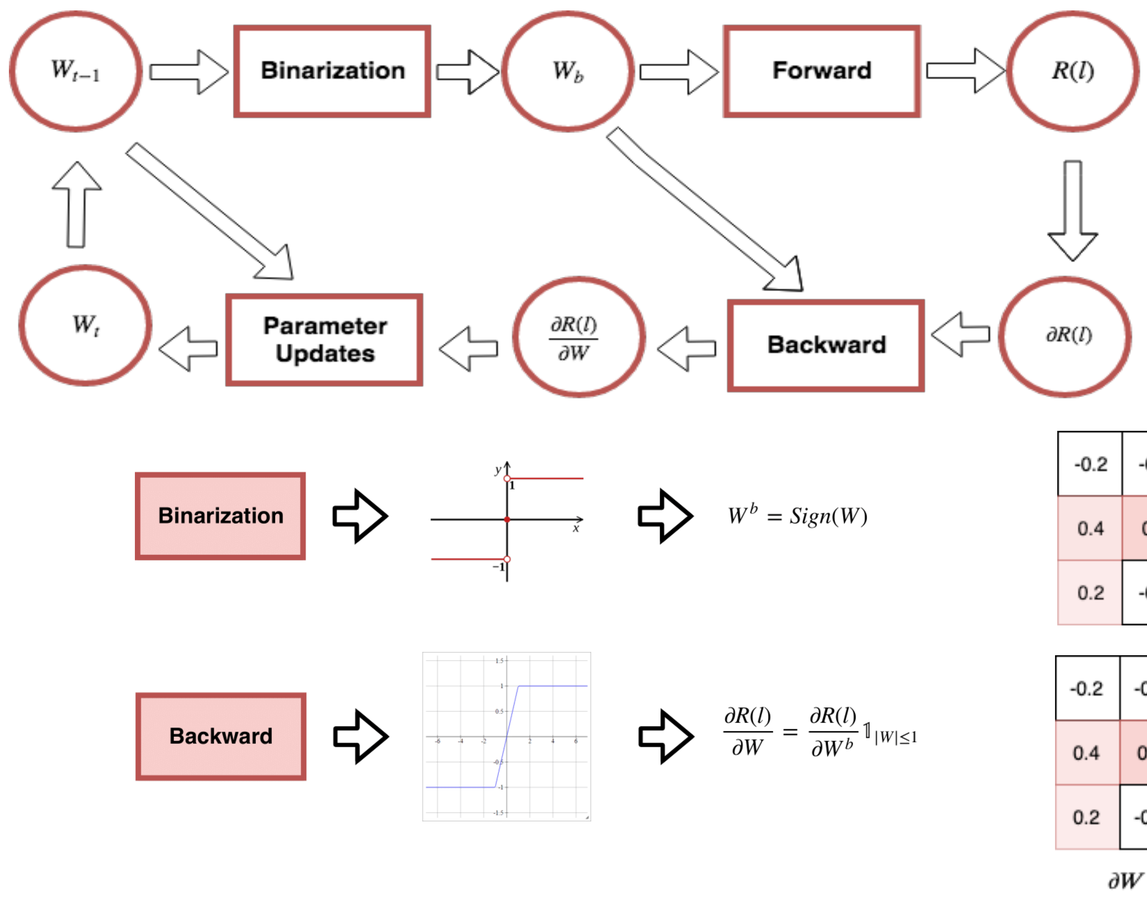
Sajad Darabi[1], Mouloud Belbahri[2], Matthieu Courbarieaux[3], Vahid Partovi Nia[4]

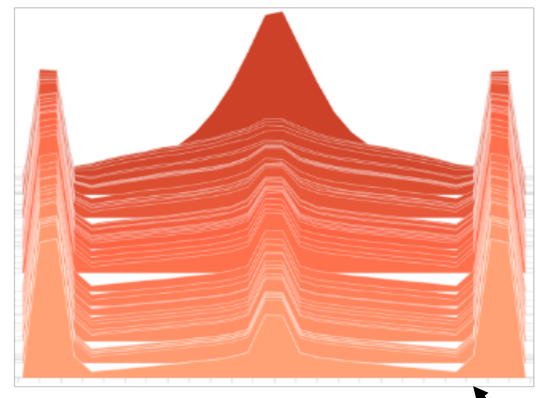
[1]University of California Los Angeles, [2]University of Montreal,[3] MILA, [4]Huawei Noah's Ark Lab
Work done as an internship at Noah's Ark Lab

BNN Overview

Binary networks are a quantized version of full precision neural networks, where the weights and activations are limited to binary values, for example {-1, 1}. At training time, as binary operations are discrete and not differentiable approximations are used in the backward pass. Naturally, there is a discrepancy between the quantized network and the full-precision counterpart encouraging research in reducing this gap.

We present slight modifications to the BNN training method, by introducing trainable scaling factors, along with regularization terms to help train them through out the network. Additionally, a modified backward straight through estimator SwishSign is used.

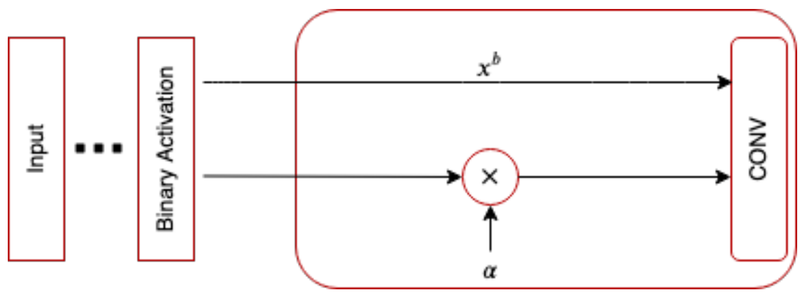




BNN weight progression of a convolutional layer in Alexnet.

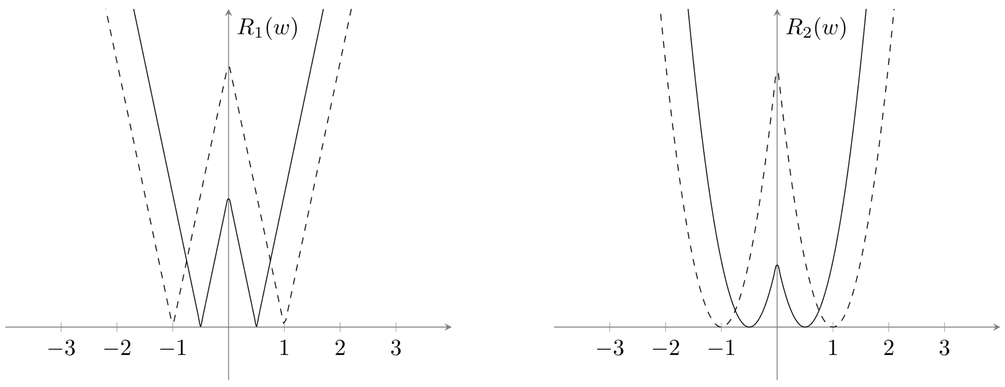
Regularized Training

Adding trainable scaling factors to Conv layers



$$J(W,b) = L(W,b) + \lambda_t \sum_h R(W_h, \alpha_h)$$

* Enforce regularization through out training or gradually start increasing with order ln/linear after t epochs

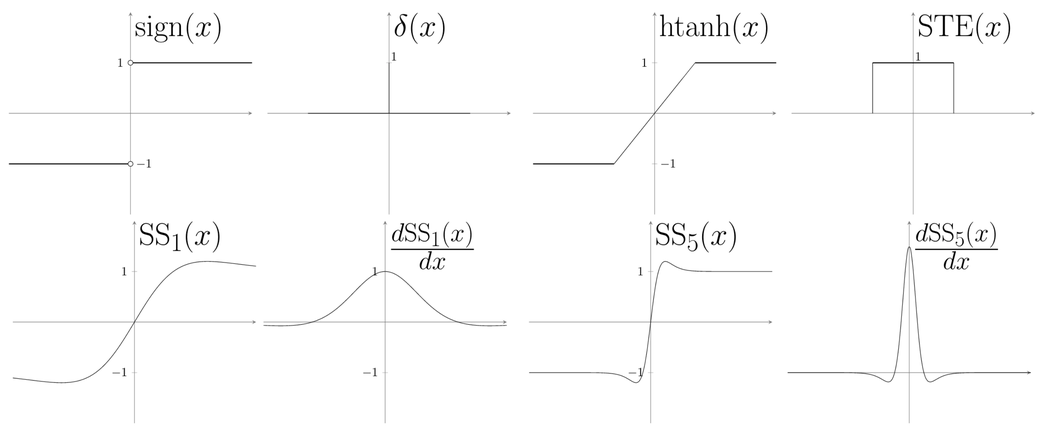


$$R_1(w) = |\alpha - |w||, \quad R_2(w) = (\alpha - |w|)^2$$

Backward Approximation

SwishSign $\quad \mathrm{SS}_\beta(x) = 2\sigma(\beta x)\left[1 + \beta x\{1 - \sigma(\beta x)\}\right] - 1.$

Backward STE $\quad \frac{d\mathrm{SS}_\beta(x)}{dx} = \frac{\beta\{2 - \beta x \tanh(\frac{\beta x}{2})\}}{1 + \cosh(\beta x)}$

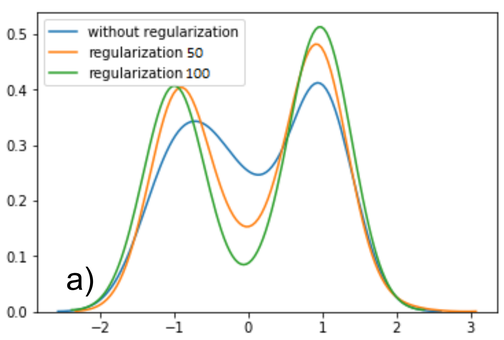


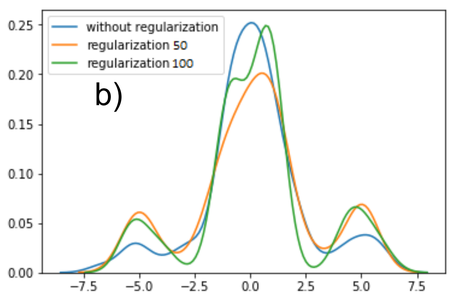
GitHub

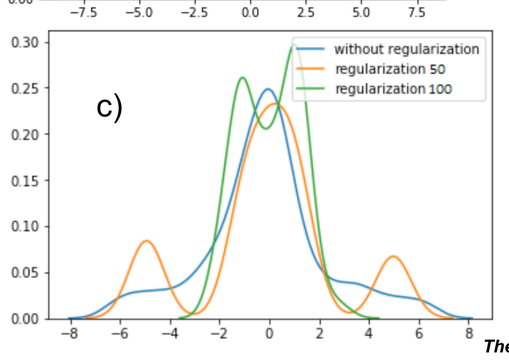
CIFAR10 Ablation

Scale	STE	AlexNet	DoReFa-Net[2]
		Top-1	Top-1
R_1	SS_t	87.16	83.92
	SS_{10}	86.80	83.21
	SS_5	86.85	83.72
	tanh	87.06	82.79
	bireal	87.04	82.64
	htanh	86.90	82.78
R_2	SS_t	87.30	83.52
	SS_{10}	86.39	82.43
	SS_5	86.80	83.58
	tanh	87.13	82.33
	bireal	87.39	82.93
	htanh	87.20	82.33
XNOR[3]	SS_t	84.94	82.58
	SS_{10}	83.64	83.12
	SS_5	85.74	82.54
	tanh	80.30	81.68
	bireal	83.84	82.58
	htanh	83.87	82.08
None	SS_t	86.24	83.00
	SS_{10}	84.70	82.98
	SS_5	84.85	83.14
	tanh	86.12	82.33
	bireal	86.24	82.56
	htanh	85.97	82.51

Effect of Regularized Training







Effect of proposed regularization on NN trained on simulated data
a) True weights {-1, 1}
b) True weights {-5, -1, +1, -1, -5} clipped regularizer
c) True weights {-5, -1, +1, -1, -5} non-clipped regularizer

Imagenet

Reg.	Activation	AlexNet		Resnet-18	
		Top-1	Top-5	Top-1	Top-5
R_1	SS_5	46.11	75.70	52.64	72.98
	SS_{10}	46.08	75.75	51.13	74.94
	htanh	41.58	69.90	50.72	73.48
R_2	SS_5	45.62	70.13	53.01	72.55
	SS_{10}	45.79	75.06	49.06	70.25
	htanh	40.68	68.88	48.13	72.72
None	SS_5	45.25	75.30	43.23	68.51
	SS_{10}	45.60	75.30	44.50	64.54
	htanh	39.18	69.88	42.46	67.56

The 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing @ NeurIPS 2019 on the bottom of your poster