WeightAlign: Normalizing Activations by Weight Alignment

XiangWei Shi*, Yunqiang Li*, Xin Liu* and Jan van Gemert
Computer Vision Lab, TU Delft, Netherlands
ICPR2020

Motivation

- BatchNorm (BN) suffers from performance degradation on small batch-size that is statistically unstable.

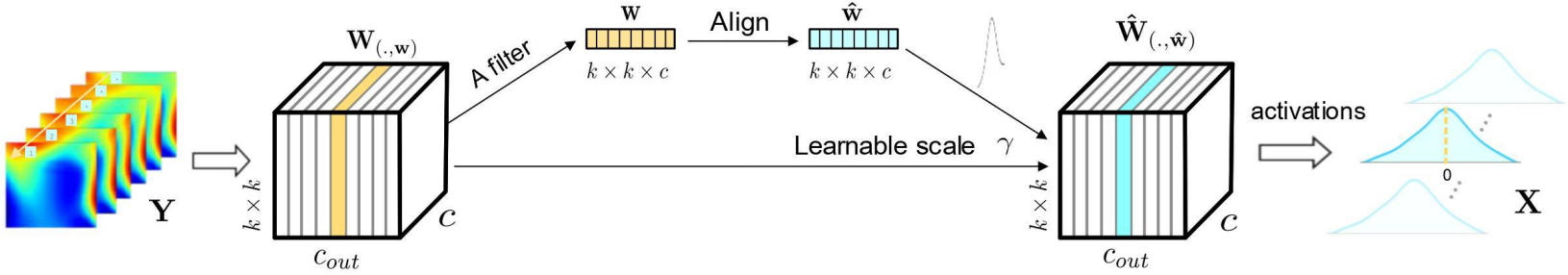
Motivation

- BatchNorm (BN) suffers from performance degradation on small batch-size that is statistically unstable.
- Proposed WeightAlign: normalizing activations without using sample statistics.

Motivation

- BatchNorm (BN) suffers from performance degradation on small batch-size that is statistically unstable.
- Proposed WeightAlign: normalizing activations without using sample statistics.
- Our method re-parameterize the weights within a filter to reach correctly normalized activations.

Method Overview



Proposed Method

BN: $$\hat{\mathbf{x}} = \frac{\mathbf{x} - \mu_\beta}{\sigma_\beta}, \quad \mathbf{r} = \gamma\hat{\mathbf{x}} + \beta, \qquad (1)$$

Proposed Method

BN: $$\hat{\mathbf{x}} = \frac{\mathbf{x} - \mu_\beta}{\sigma_\beta}, \quad \mathbf{r} = \gamma\hat{\mathbf{x}} + \beta, \qquad (1)$$

Expressing statistics by weights:

$$\mu_\beta = E[x] = nE[w]E[Y], \qquad (2)$$

$$\sigma_\beta^2 = \mathrm{Var}[x] = n\left(E[w^2]E[Y^2] - E^2[w]E^2[Y]\right) \qquad (3)$$

Proposed Method

BN: $$\hat{\mathbf{x}} = \frac{\mathbf{x} - \mu_\beta}{\sigma_\beta}, \quad \mathbf{r} = \gamma\hat{\mathbf{x}} + \beta, \qquad (1)$$

Expressing statistics by weights:

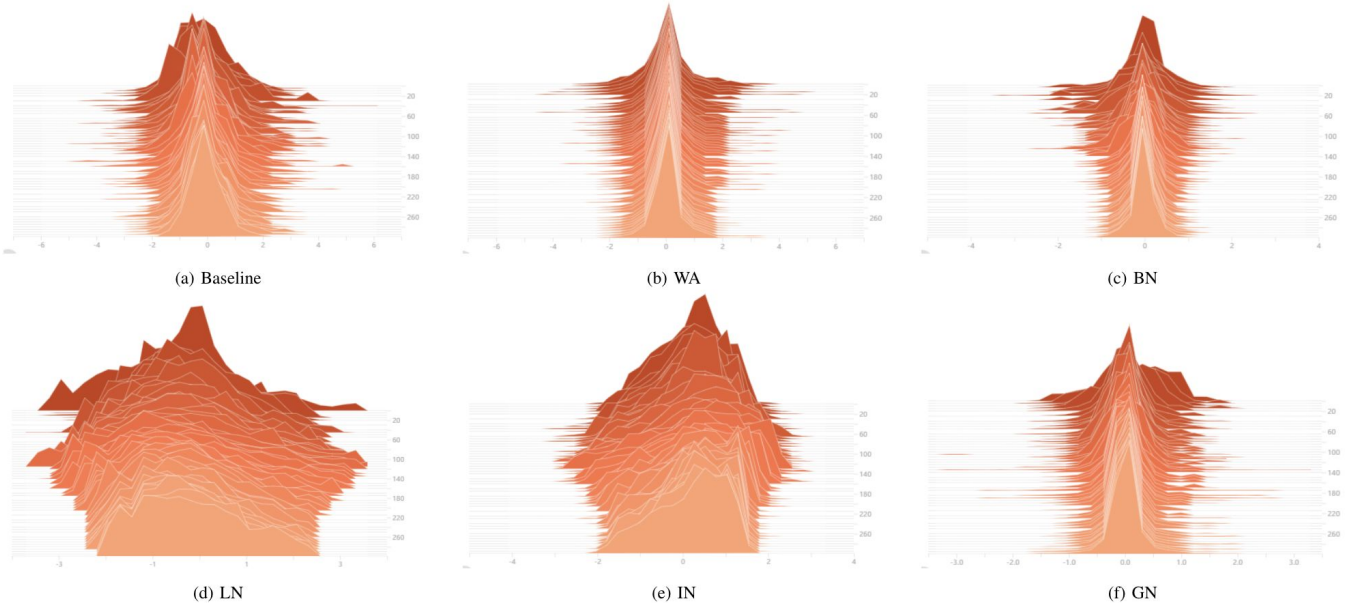
$$\mu_\beta = E[x] = nE[w]E[Y], \qquad (2)$$

$$\sigma_\beta^2 = \mathrm{Var}[x] = n\left(E[w^2]E[Y^2] - E^2[w]E^2[Y]\right) \qquad (3)$$

WeightAlign (WA): $$E[w] = 0, \quad \frac{1}{2}n\mathrm{Var}[w] = 1. \qquad (4)$$

$$\hat{w} = \gamma \frac{w - E[w]}{\sqrt{n/2 \cdot \mathrm{Var}[w]}}, \qquad (5)$$

Experiment results



(a) Baseline (b) WA (c) BN

(d) LN (e) IN (f) GN

visualization of a single channel activation in training

Experiment results

CIFAR-10							
Batch size 64				Batch size 1			
Method	Error	Method	Error	Method	Error	Method	Error
Baseline [36]	6.46	WA	6.21	Baseline	7.27	WA	6.61
BN [23]	4.30	BN+WA	4.29	BN	-	BN+WA	-
IN [20]	6.49	IN+WA	6.42	IN	6.91	IN+WA	6.50
LN [19]	5.02	LN+WA	5.12	LN	6.82	LN+WA	5.76
GN [21]	4.96	GN+WA	4.60	GN	5.79	GN+WA	5.51

Error rate of ResNet50 on CIFAR-10 for classification

Experiment results

model	Top-1 (%) Error	Top-5 (%) Error
VGG16* (Baseline)	31.30	11.19
VGG16 (BN)*	29.58	10.16
VGG16 (WA)	29.78	10.23
VGG16 (BN+WA)	**27.07**	**8.78**
ResNet50 (Baseline)† [36]	27.60	-
ResNet50 (BN)*	24.89	7.71
ResNet50 (WA)	26.62	8.91
ResNet50 (BN+WA)	**24.04**	**7.12**

Image classification on ImageNet

Thank you