

Filtered Batch Normalization

Peter Pazmany Catholic University
Faculty of Information Technology and Bionics



András Horváth, Jalal Al-afandi



Normalization

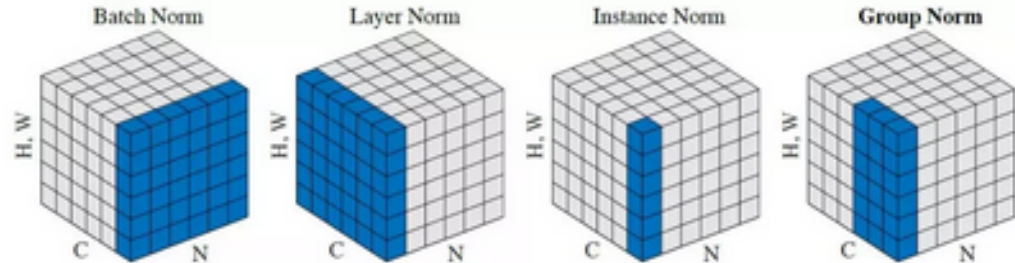
Activation normalization –
ubiquitous in neural networks

Batch normalization, group
normalization, response
normalization, etc.

$$y_i = \gamma \frac{(x_i - \mu_i)}{\sigma_i} + \beta$$

$$\mu_i = \frac{1}{m} \sum_{k \in S_i} x_k$$

$$\sigma_i = \sqrt{\frac{1}{m} \sum_{k \in S_i} (x_k - \mu_i)^2 + \epsilon}$$



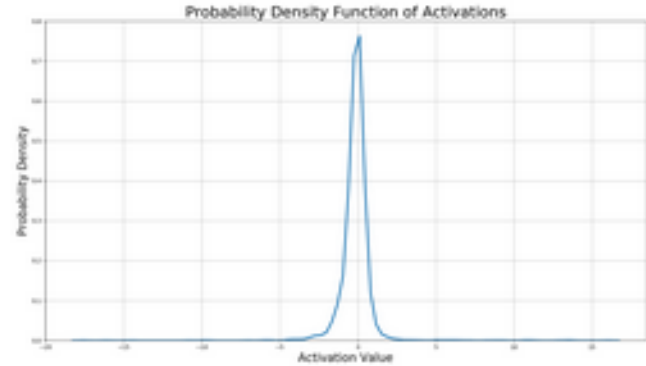


Normalization

Activation normalization –
ubiquitous in neural networks

Batch normalization, group
normalization, response
normalization, etc.

We assume that activations follow
Gaussian distribution





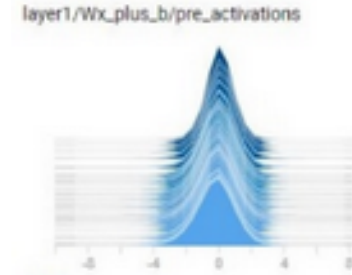
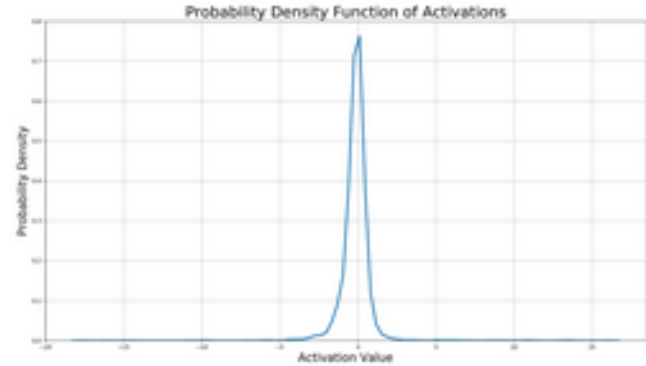
Normalization

Activation normalization –
ubiquitous in neural networks

Batch normalization, group
normalization, response
normalization, etc.

We assume that activations follow
Gaussian distribution

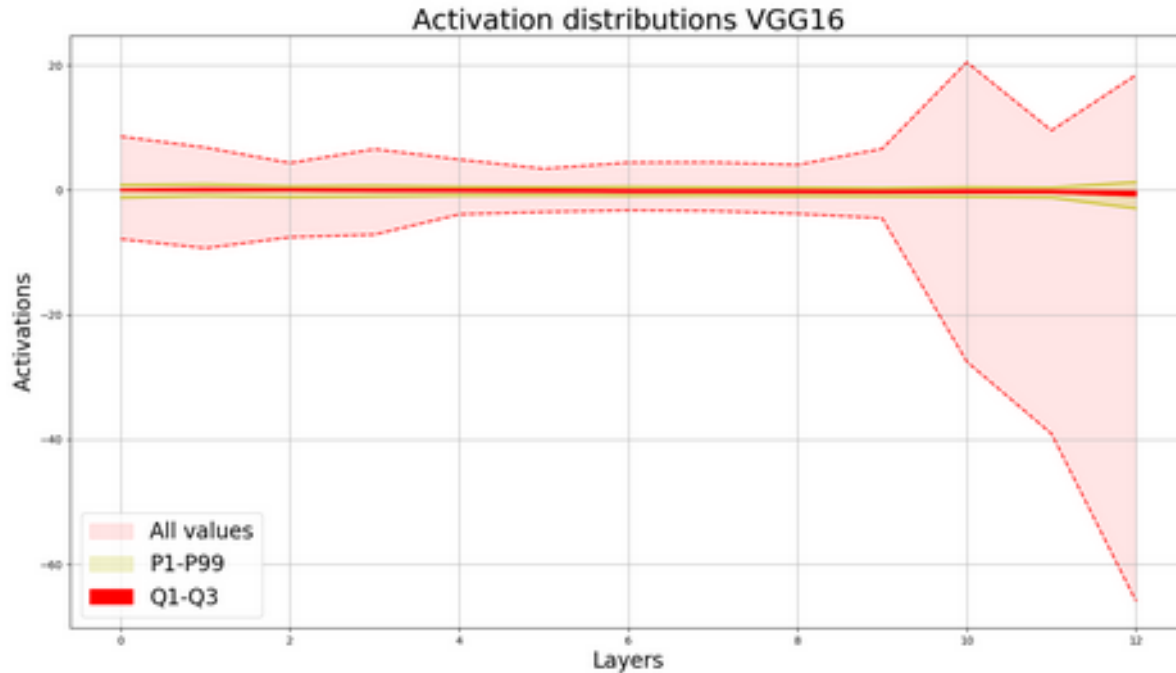
Is it true? Is it true for all layers?





Distribution of activations

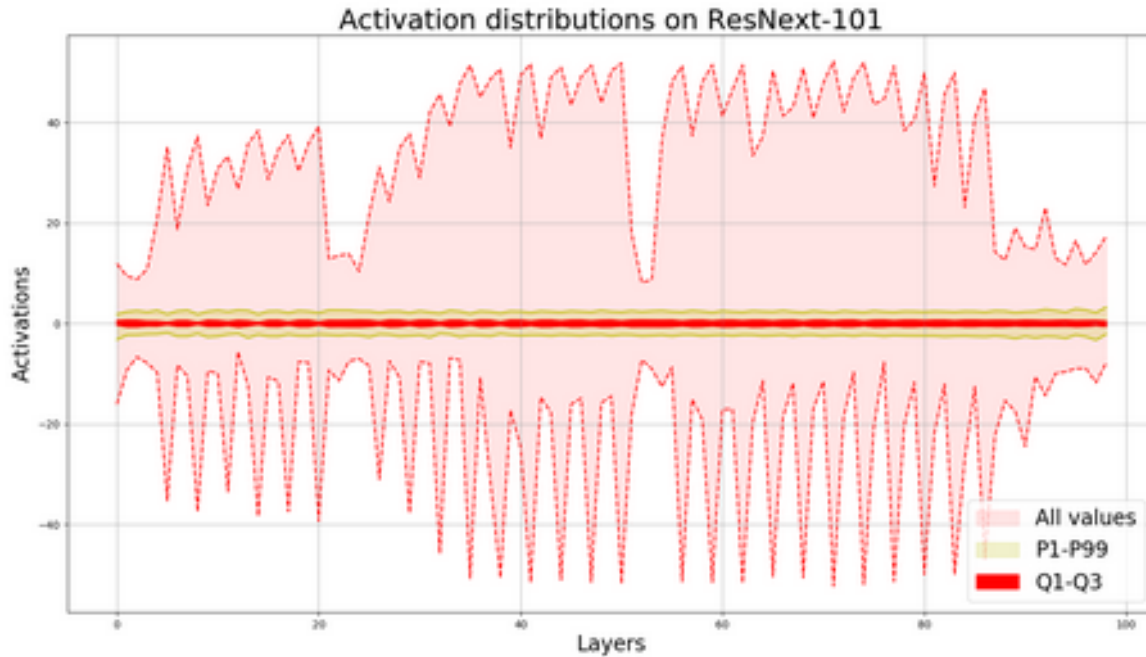
Pretrained version of VGG 16 on test set of ImageNet





Distribution of activations

Pretrained version of ResNext-101 on test set of ImageNet



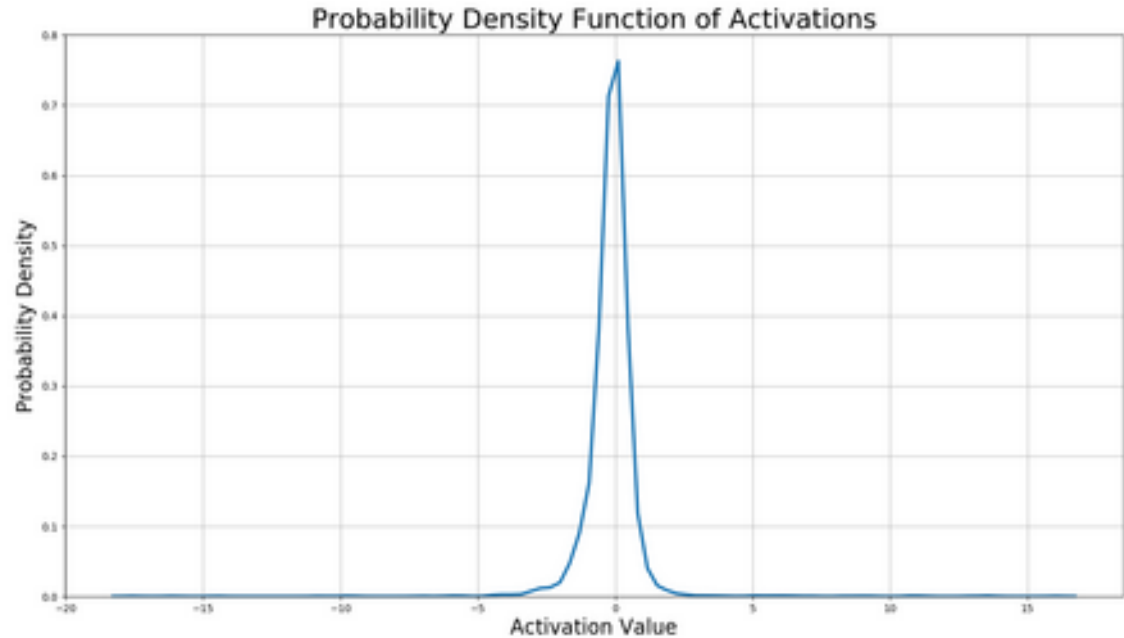


Gaussian distribution

Activations follow Gaussian distribution

With some outliers

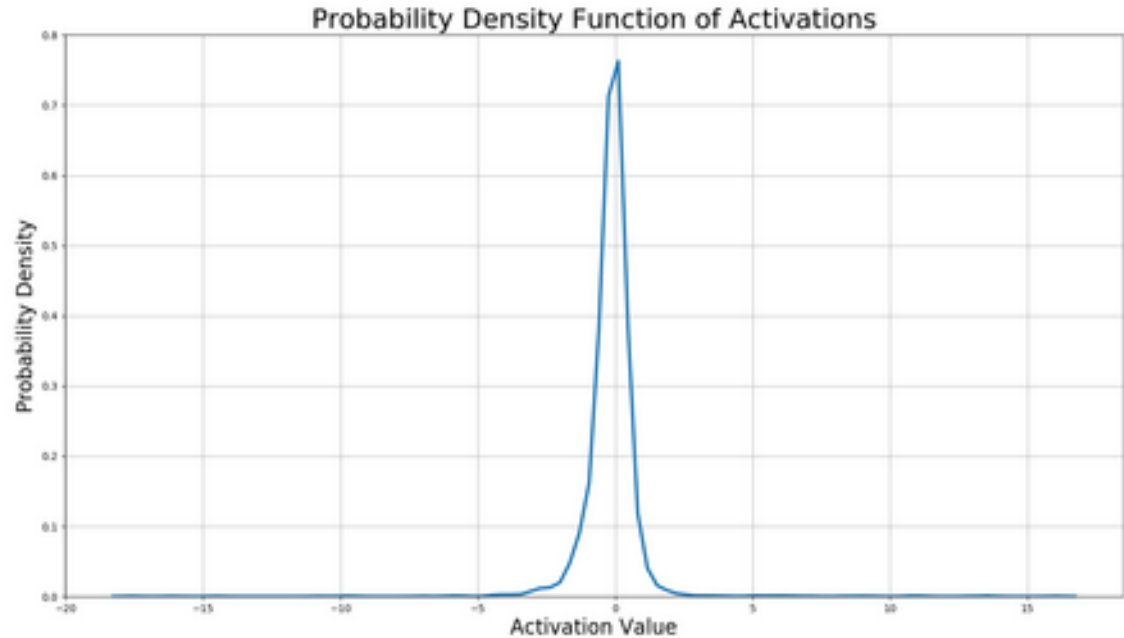
In case of Gaussian distribution samples outside 7σ has a 1:390,682,215,445 probability of appearance





Gaussian distribution

Activations follow Gaussian distribution
With some outliers



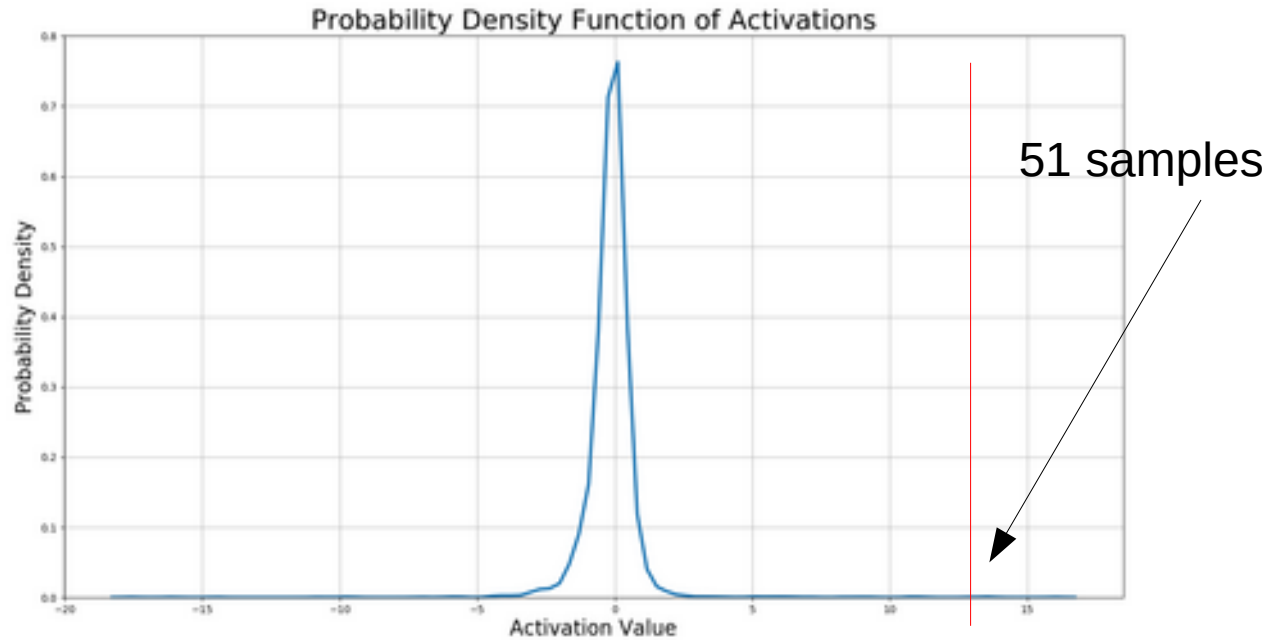


Out of distribution samples

Activations follow Gaussian distribution

With some outliers

What are these outliers?





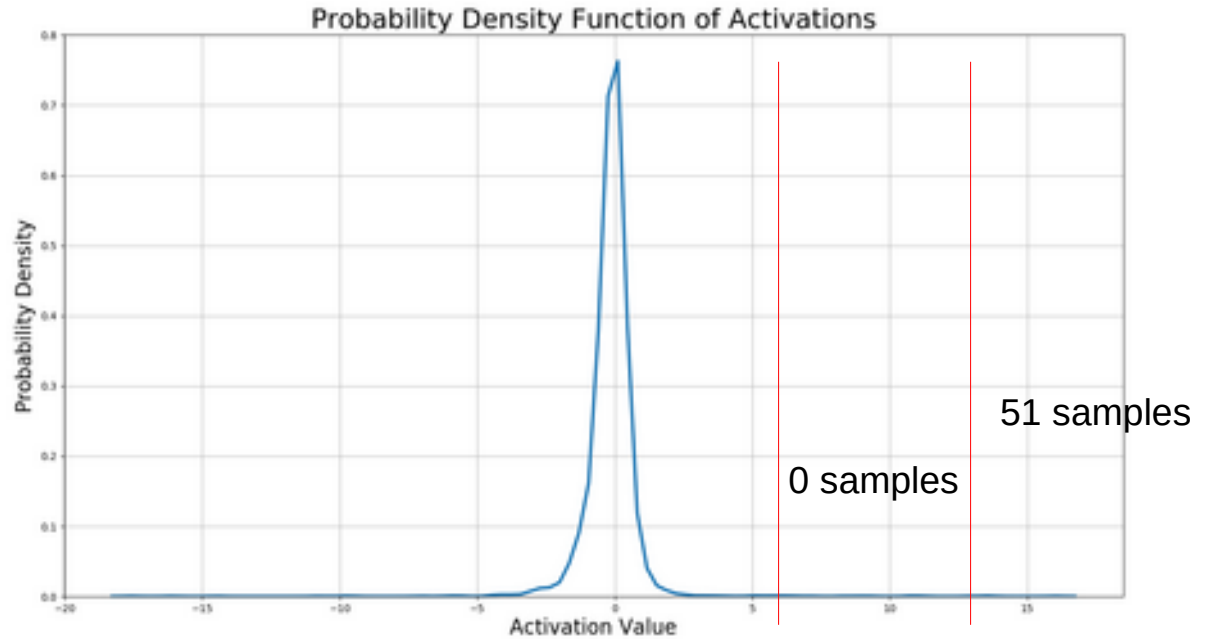
Specificity vs generality

Activations follow Gaussian distribution

With some outliers

What are these outliers?

51 out of 50000 samples



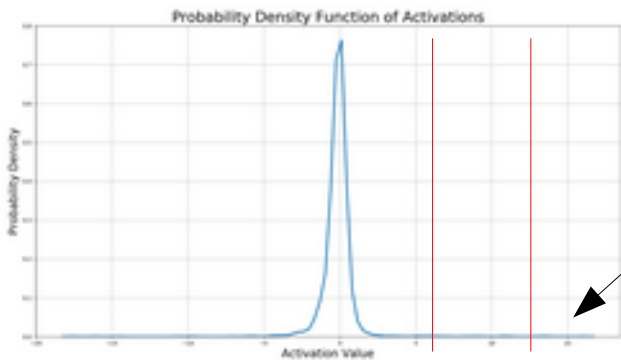


Specificity vs generality

The samples are specific (kernel 453)

Not all layers are Gaussian- logit layer

We want specific neurons (saliency maps, deepdream)





Specificity vs generality (kernel 34)





Specificity channel 206

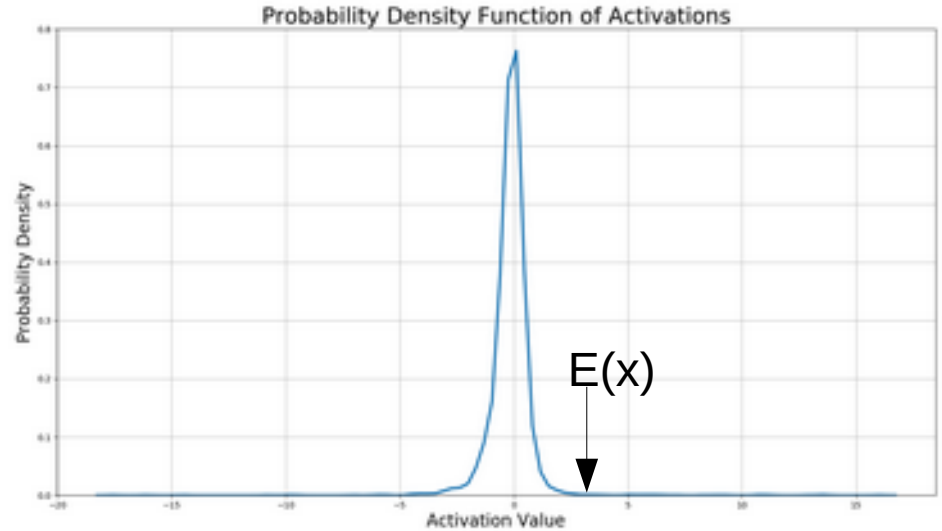




Filtered Batch Normalization

Filtered Batch Normalization:

- we normalize the samples according to batch normalization



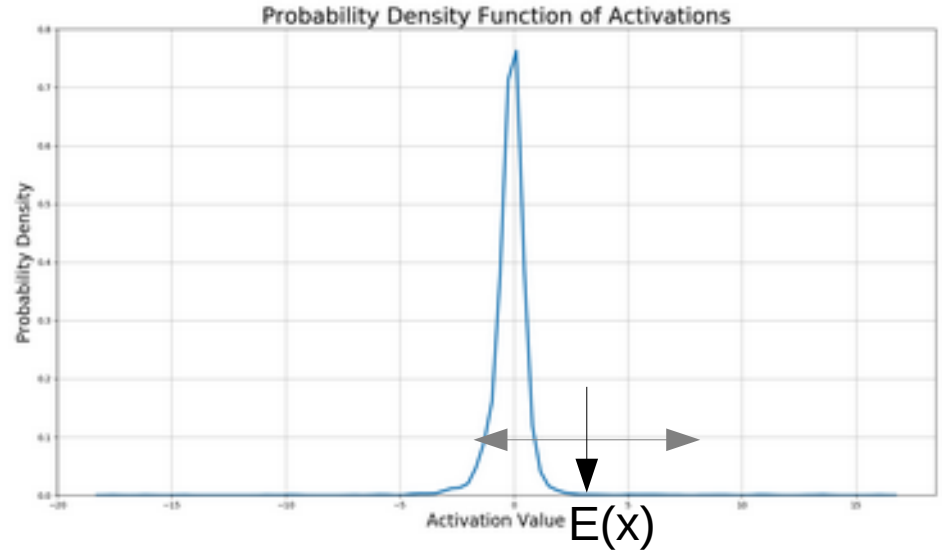


Filtered Batch Normalization

Filtered Batch Normalization:

- we normalize the samples according to batch normalization

- Select samples around the mean ($t\sigma$ distance) t is a parameter of the algorithm ($t=4$)





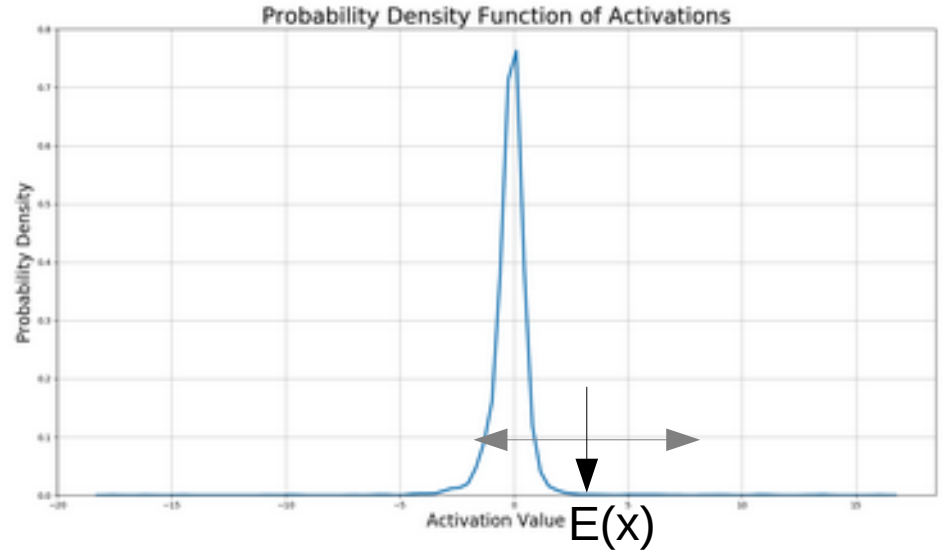
Filtered Batch Normalization

Filtered Batch Normalization:

- we normalize the samples according to batch normalization

- Select samples around the mean (in $t\sigma$ distance) t is a parameter of the algorithm ($t=4$)

- Execute batch normalization using the selected samples only

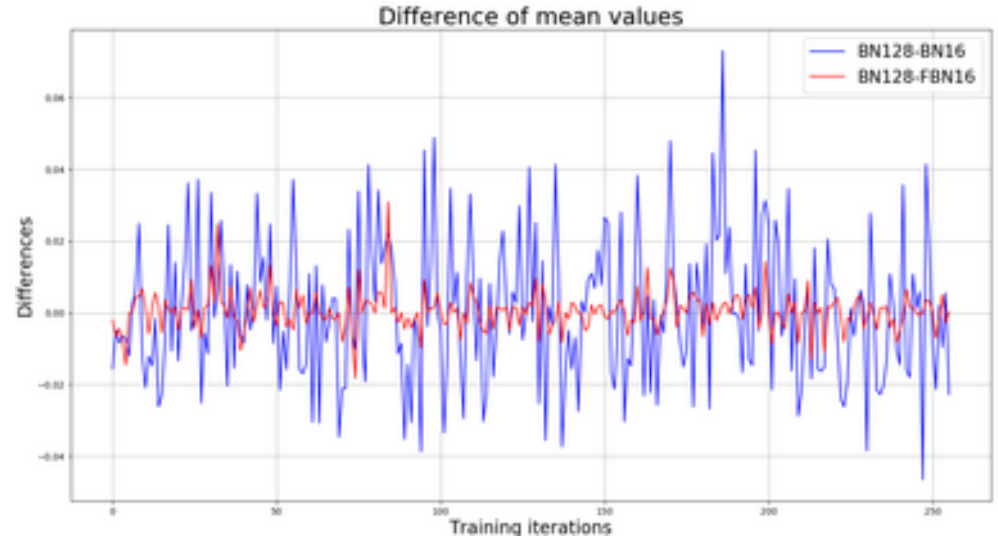




Filtered Batch Normalization

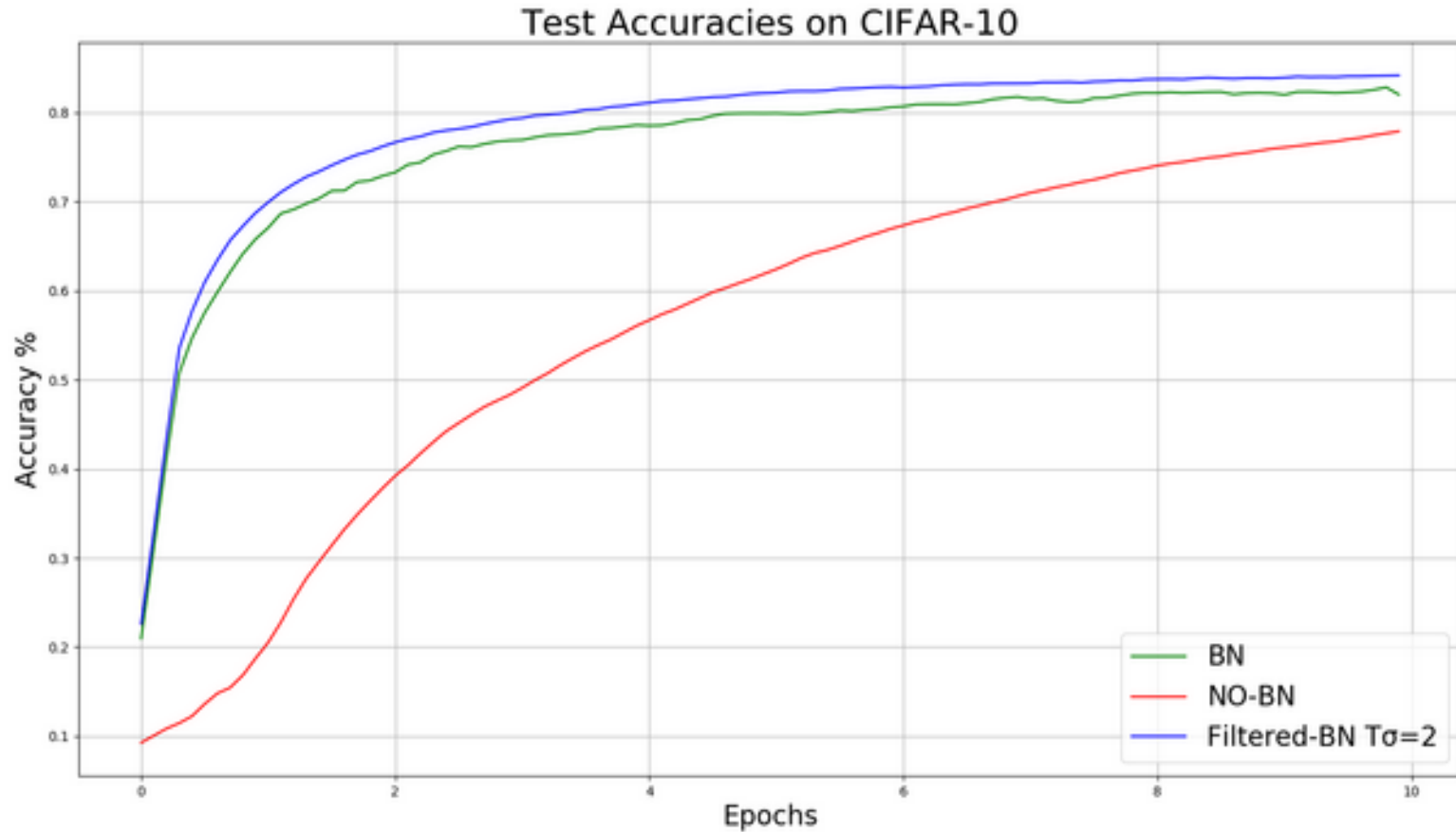
Measuring the consistency of the batches on CIFAR-10 using AlexNET

Batches of 128 as a reference
Comparing to batches of 16





Accuracy increase, faster convergence

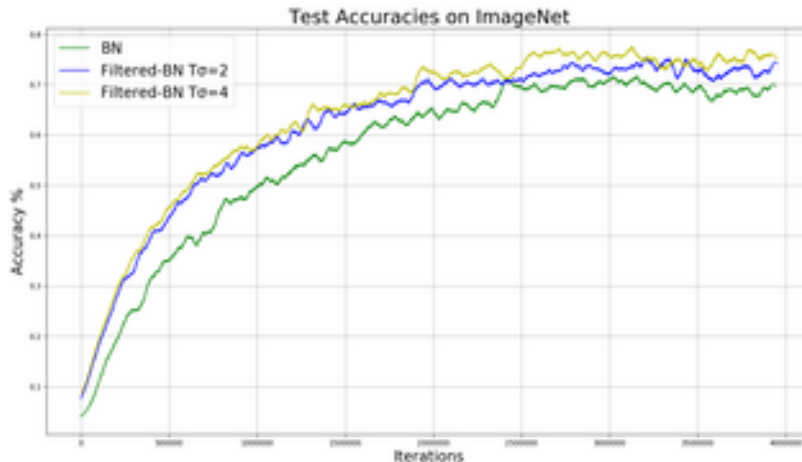




Filtered Batch Normalization

ImageNet results VGG-16

3% increase in test accuracy compared to regular batch normalization



MS-COCO results Mask R-CNN with

ResNext-101 backbone

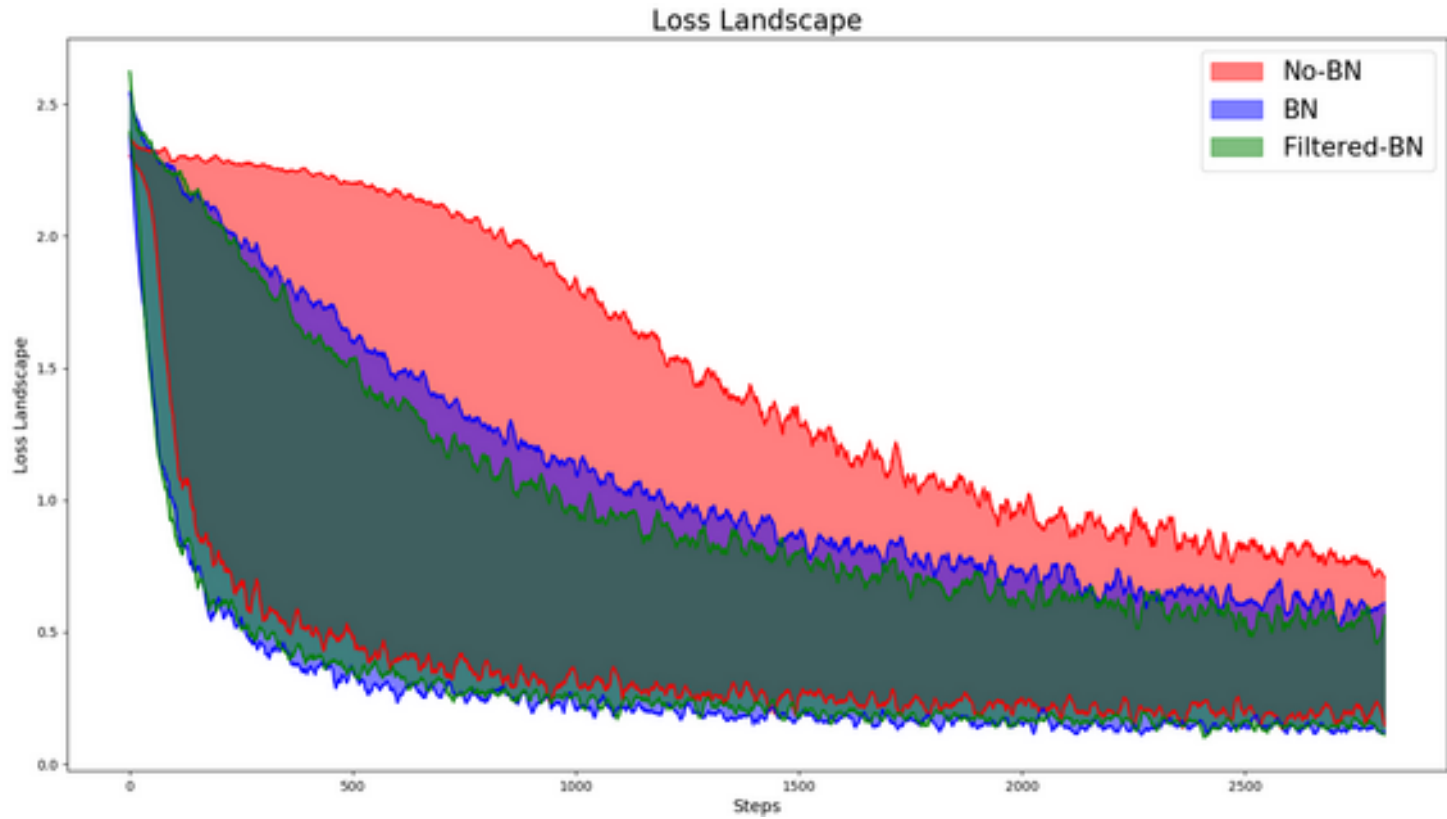
0.6 mAP increase in segmentation

1.1 mAP increase in detection

	$BNAP$	$BNAP_{50}$	$F-BNAP$	$F-BNAP_{50}$
SEG(50k)	23.87	44.56	25.41	45.42
SEG(100k)	26.86	45.55	27.34	52.40
SEG(150k)	28.66	51.80	34.15	55.43
SEG(270k)	36.47	58.07	37.06	58.92
BOX(50k)	23.63	41.90	27.86	47.84
BOX(100k)	28.16	48.43	28.74	49.65
BOX(150k)	30.53	50.79	34.24	53.13
BOX(270k)	40.01	61.32	41.12	61.71



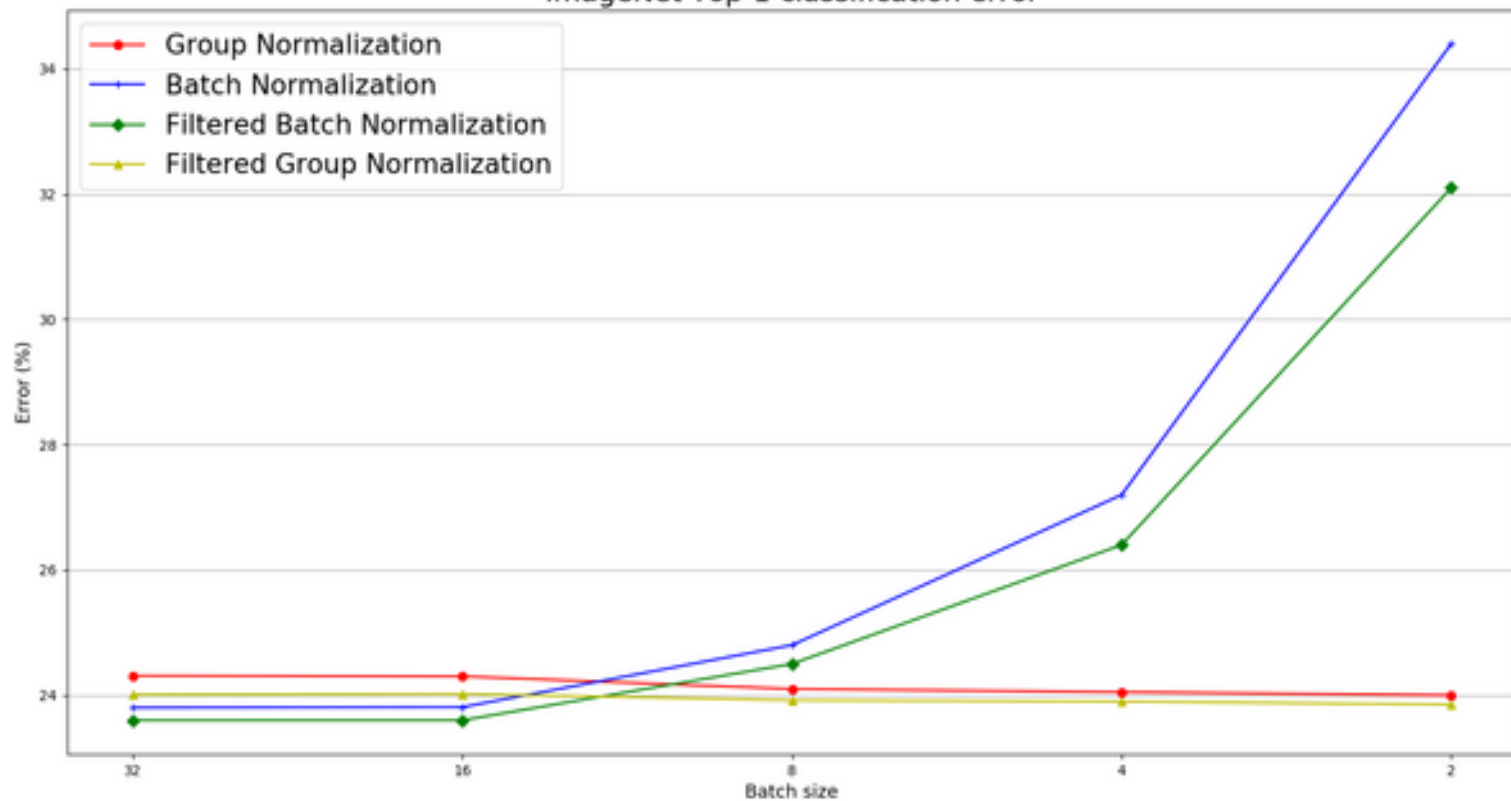
Filtered Batch Normalization – Smoother loss landscape





Group normalization

ImageNet Top-1 classification error





Conclusion

- Activaitons does not always follow Gaussian distribution, there are outliers for specificity
- These samples can disturb normalization
- We filter out out-of-distribution samples before normalization during training, no alteration in inference
- Results faster convergence and higher test accuracy

In case of questions please come to our poster or write us:
horvath.andras@itk.ppke.hu