

# Filtered Batch Normalization

Peter Pazmany Catholic University Faculty of Information Technology and Bionics

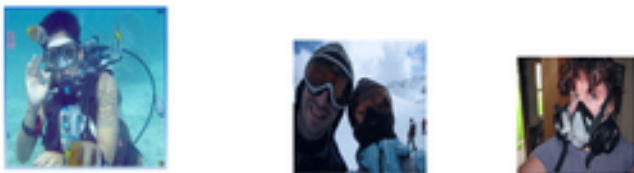
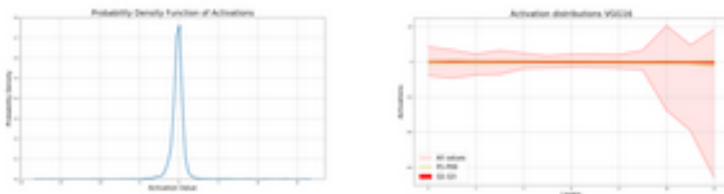
*András Horváth, Jalal Al-afandi*

## Do activations follow Gaussian distribution?

This is a common assumption in normalization techniques.

Most of the activations are Gaussian, but there are extreme outliers, which are oddly specific

Specificity is required from neurons, which goes against Gaussian distribution



## Filtered batch normalization

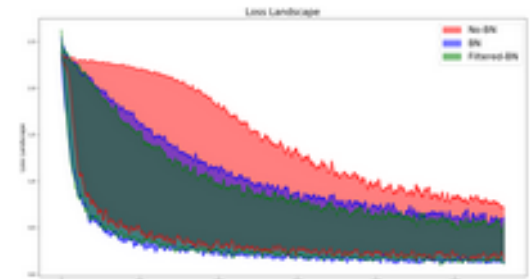
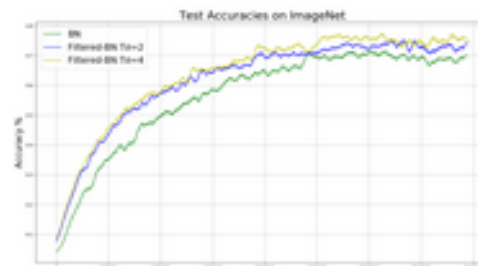
Application of a batch norm and selection of samples in  $T\sigma$  range

$$\hat{x}'_i = \frac{1}{\sigma_i} (x_i - \mu_i) \quad f(x_k) = \begin{cases} 1 & \text{if } -T\sigma \leq \hat{x}'_k \leq T\sigma \\ 0 & \text{if } \hat{x}'_k < -T\sigma \vee T\sigma < \hat{x}'_k \end{cases}$$

Using the selected samples only for normalization:

$$\mu'_i = \frac{1}{\sum_{k \in S_i} f(x_k)} \sum_{k \in S_i} f(x_k) x_k \quad \sigma'_i = \sqrt{\frac{1}{\sum_{k \in S_i} f(x_k)} \sum_{k \in S_i} f(x_k) (x_k - \mu'_i)^2 + \epsilon}$$

## Results



	BNAP	BNAP <sub>50</sub>	F-BNAP	F-BNAP <sub>50</sub>
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

Test results on MS-COCO

