Unit 5 Optimization Part 3: Batch Normalization

TFIP-Al Artificial Neural Networks and Deep Learning

Batch Normalization

$$Z = XW$$

$$ilde{oldsymbol{Z}} = oldsymbol{Z} - rac{1}{m} \sum_{i=1}^m oldsymbol{Z}_{i,:}$$

$$\hat{oldsymbol{Z}} = rac{oldsymbol{Z}}{\sqrt{\epsilon + rac{1}{m} \sum_{i=1}^{m} ilde{oldsymbol{Z}_{i,:}^2}}}$$

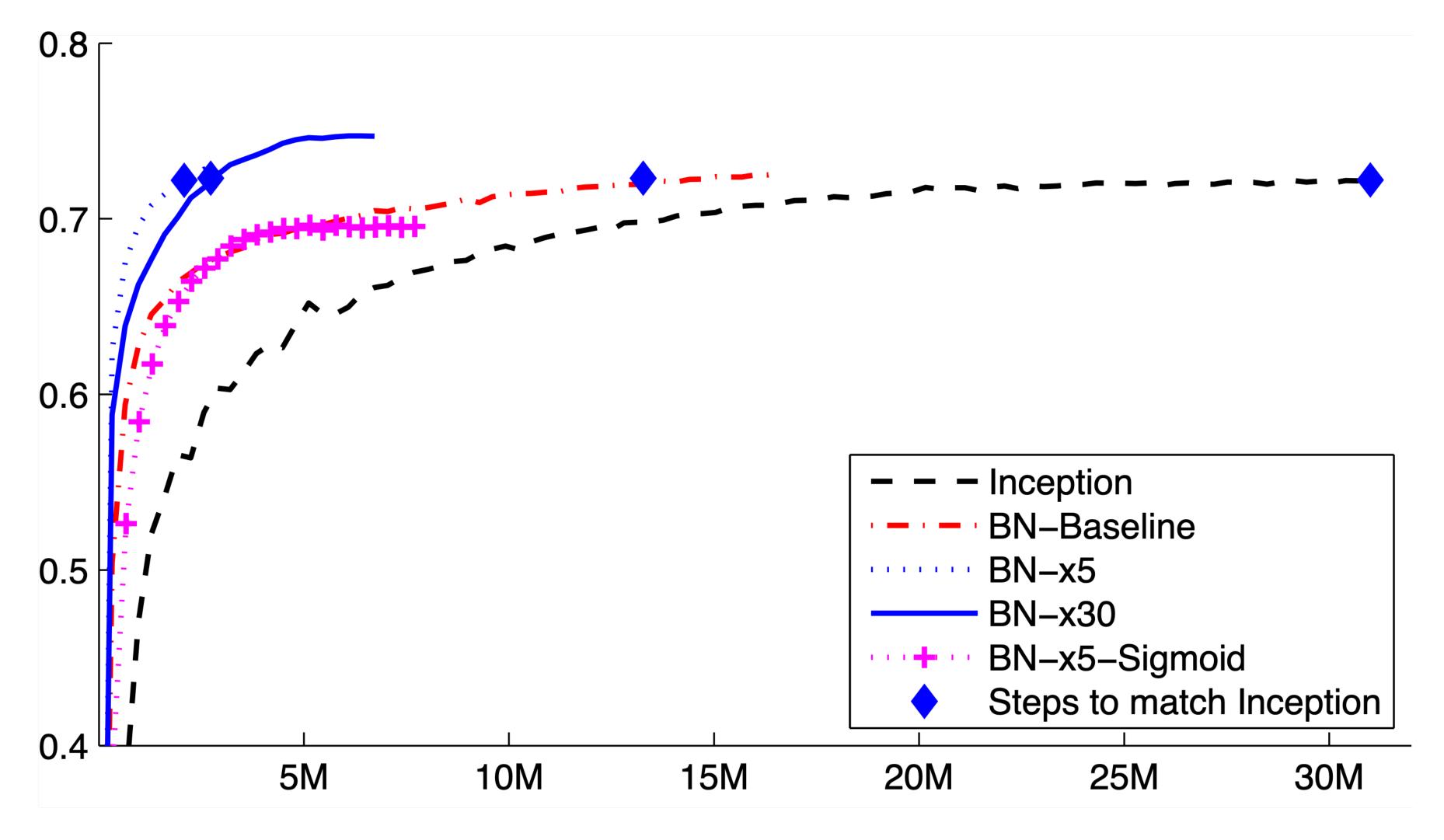
$$\boldsymbol{H} = \max\{0, \boldsymbol{\gamma}\hat{Z} + \boldsymbol{\beta}\}$$

"Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," Ioffe and Szegedy 2015

Before SGD step Hidden Layer 1 Hidden Layer 4 Input g as assessesses o asses Hidden Layer 4 Input Hidden Layer 1

After SGD step

"Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," Ioffe and Szegedy 2015



"Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," Ioffe and Szegedy 2015