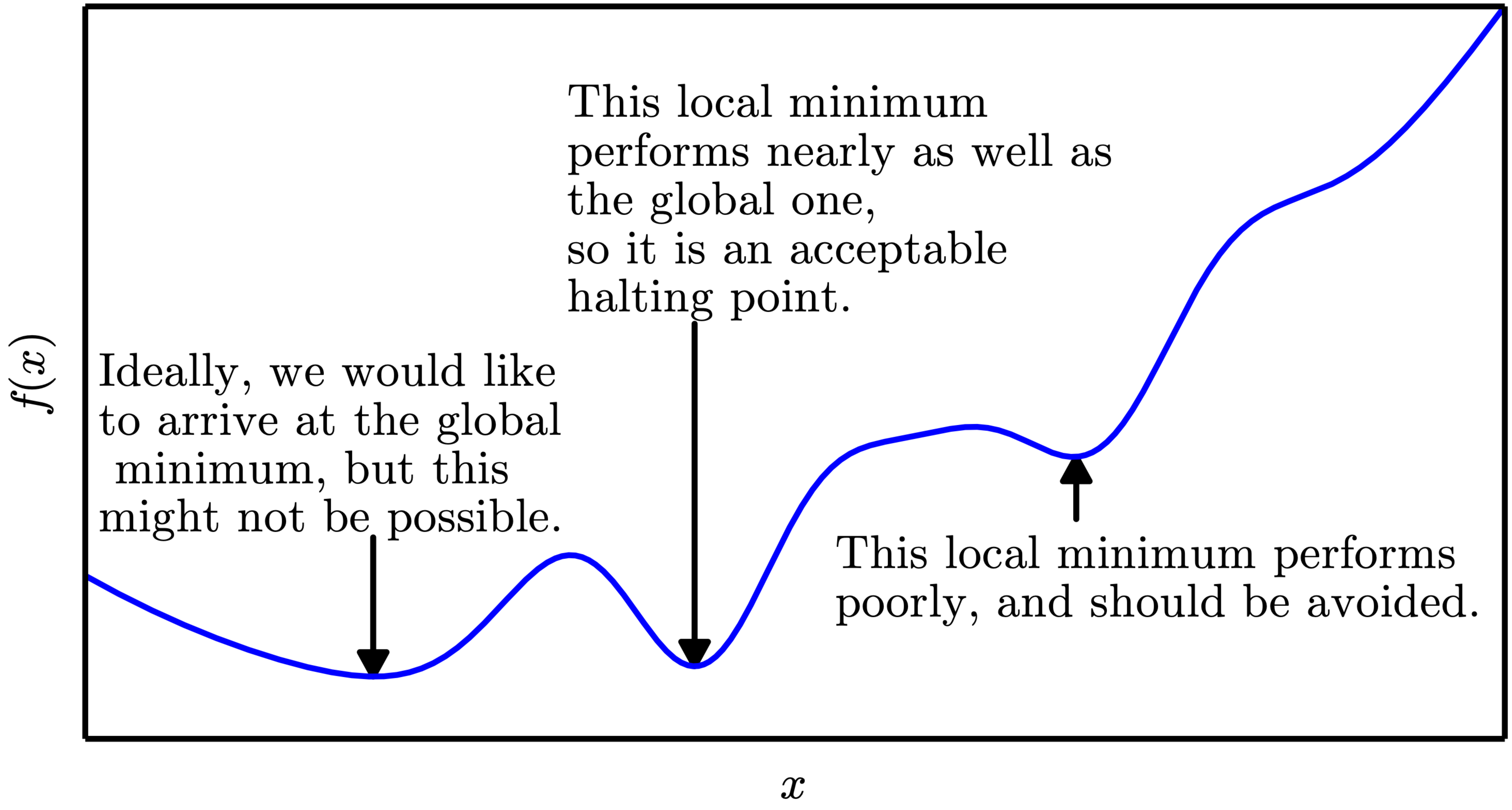
Unit 5 Optimization
Part 2: Tutorial on Optimization for Deep Networks

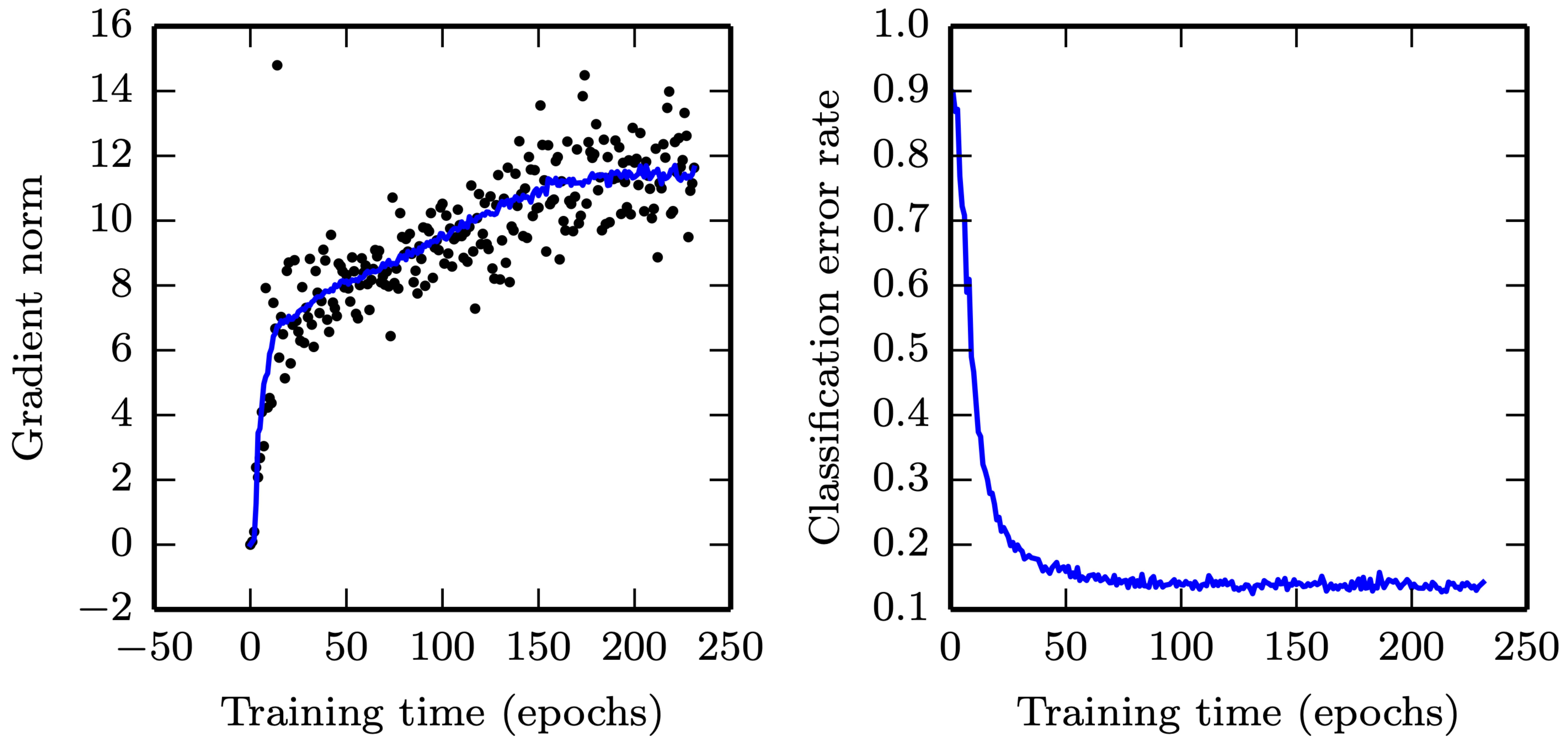
— Visualisation of Neural Network Cost Function

TFIP-AI Artificial Neural Networks and Deep Learning

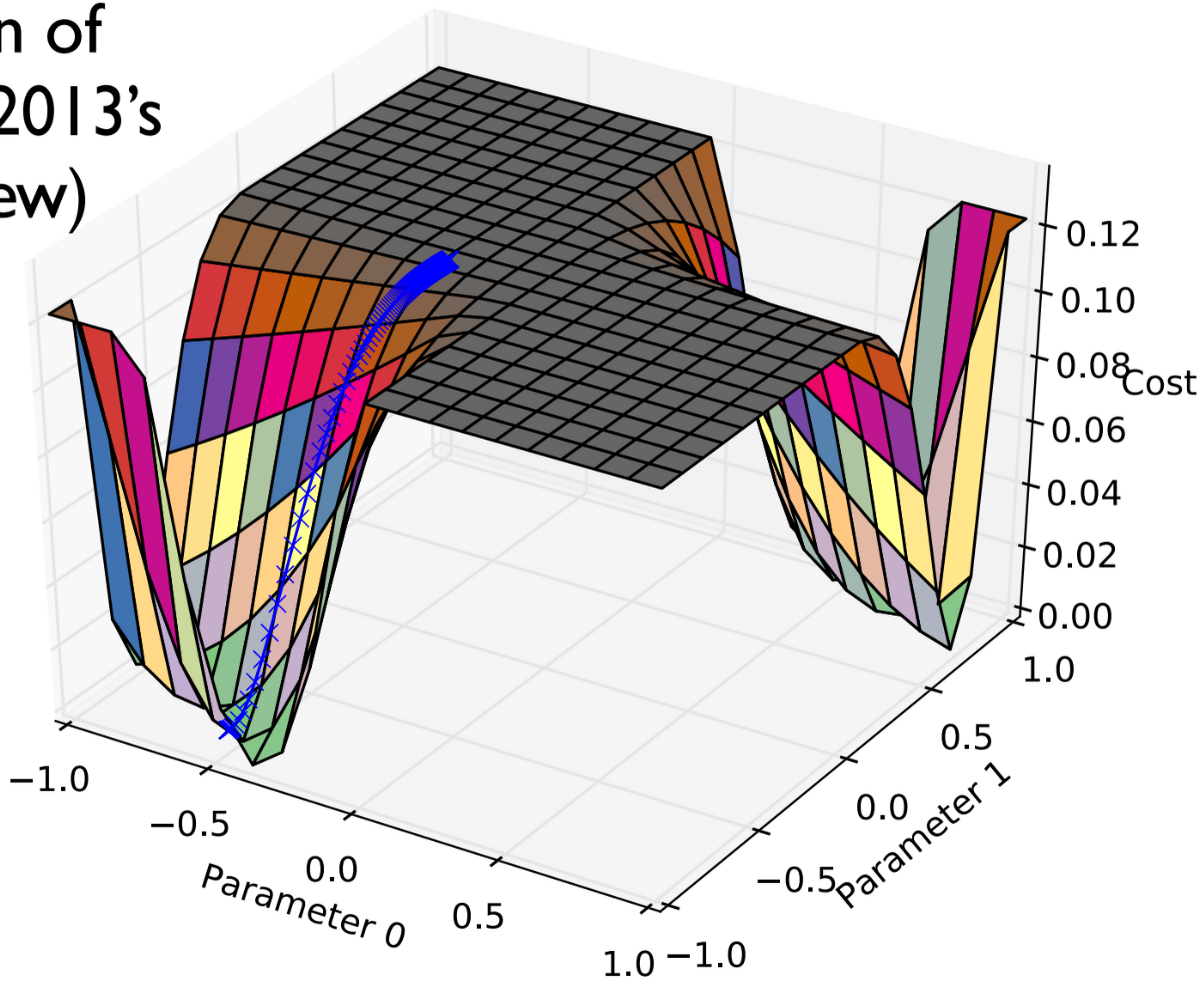
Approximate minimization

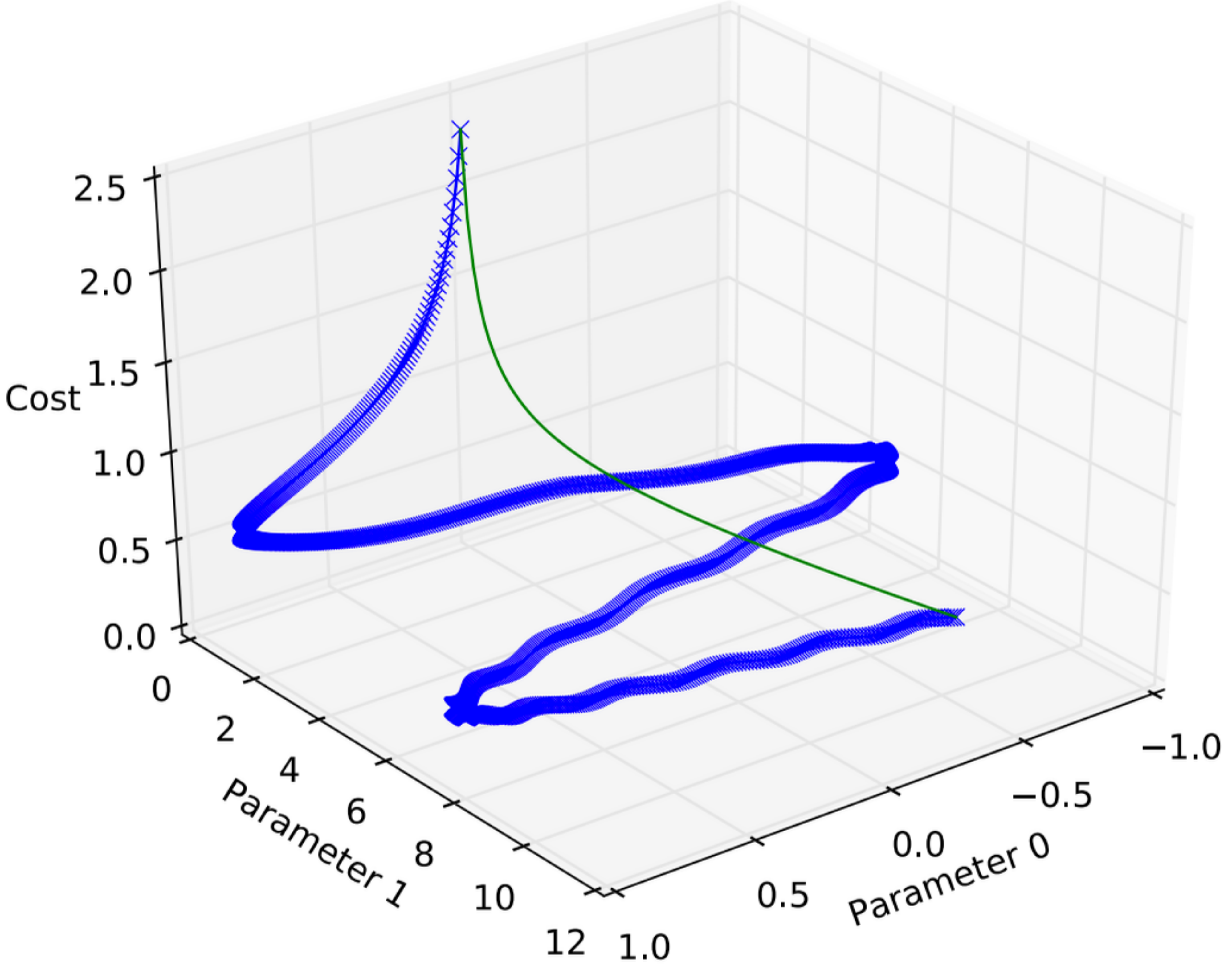


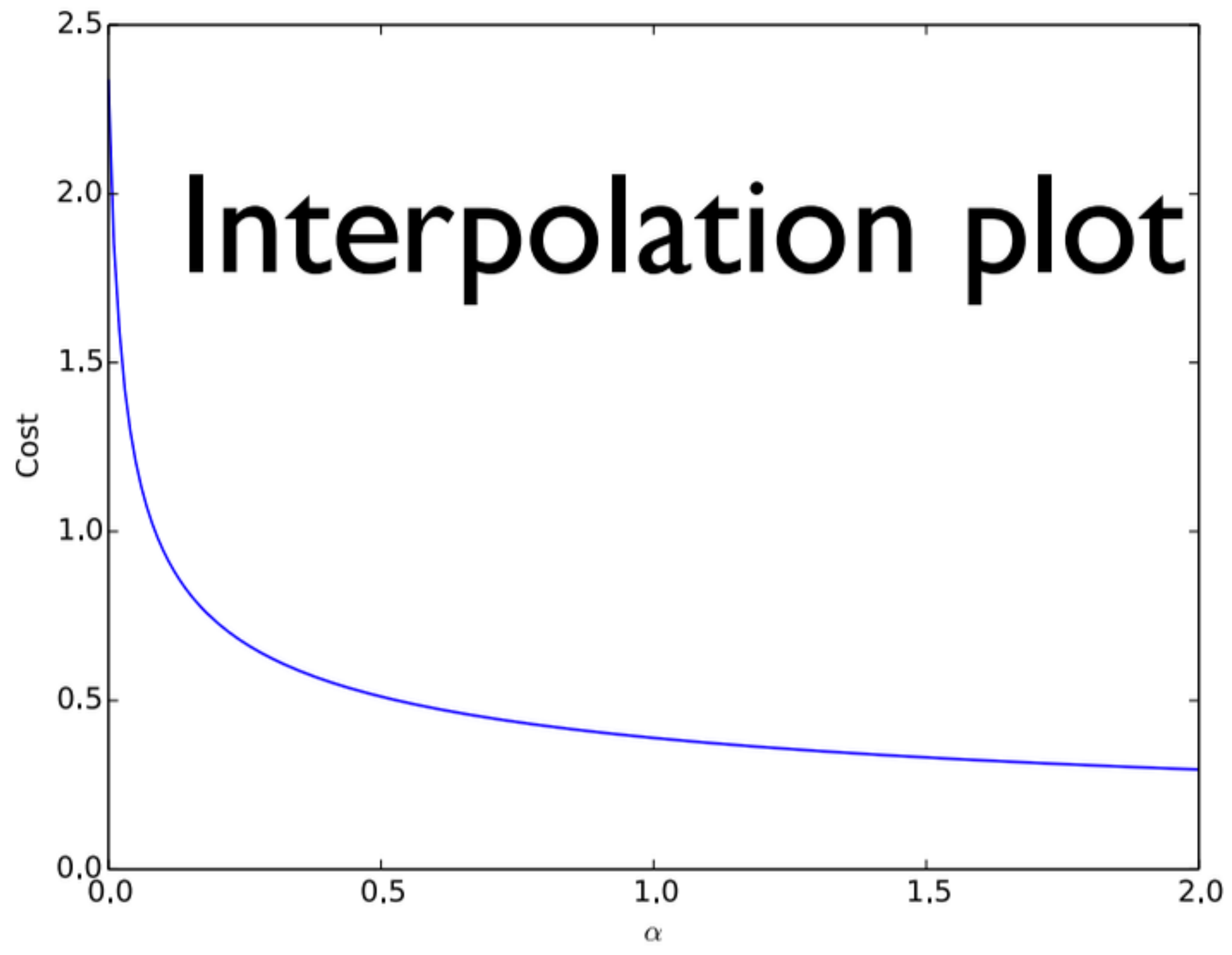
No Critical Point

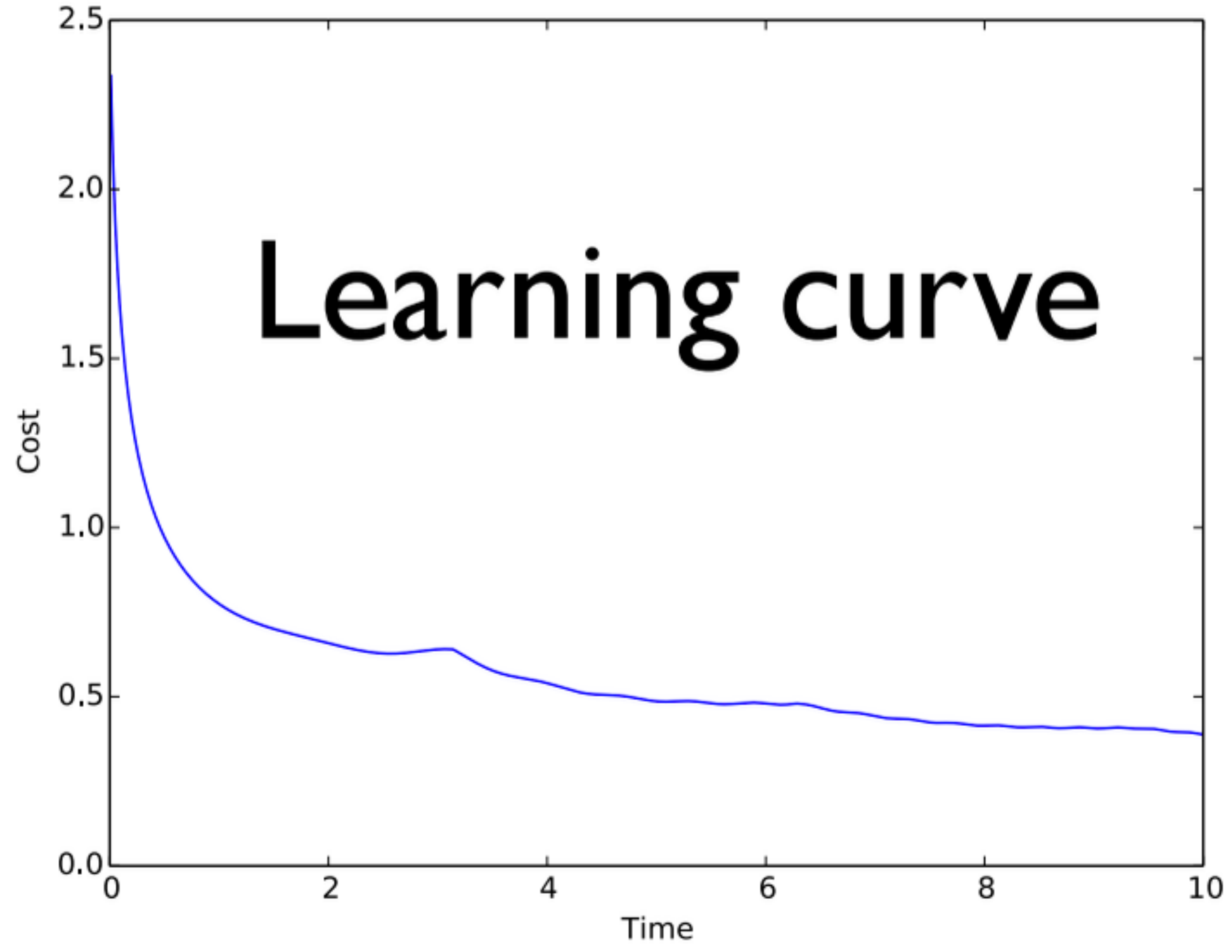


(Cartoon of Saxe et al 2013's worldview)

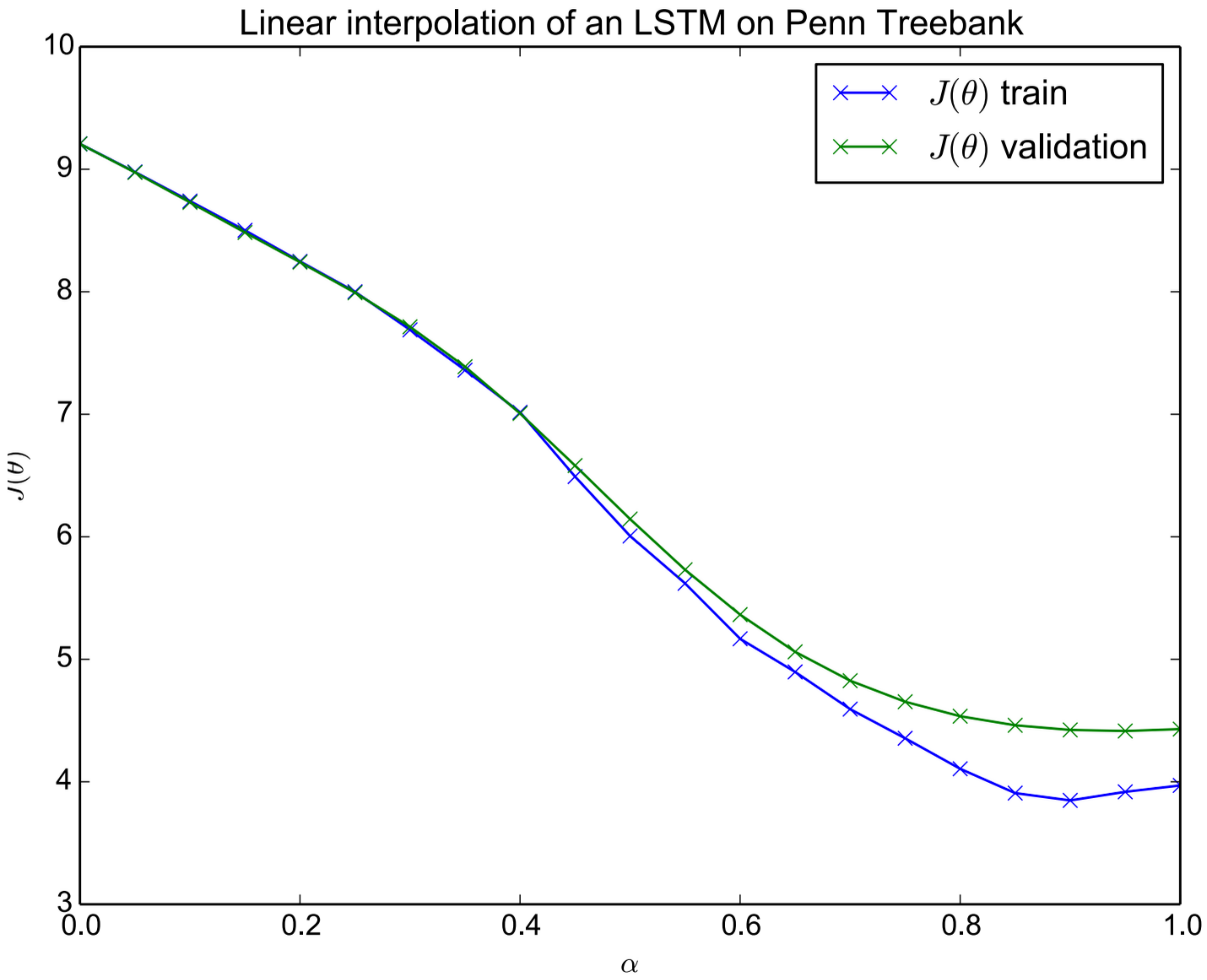




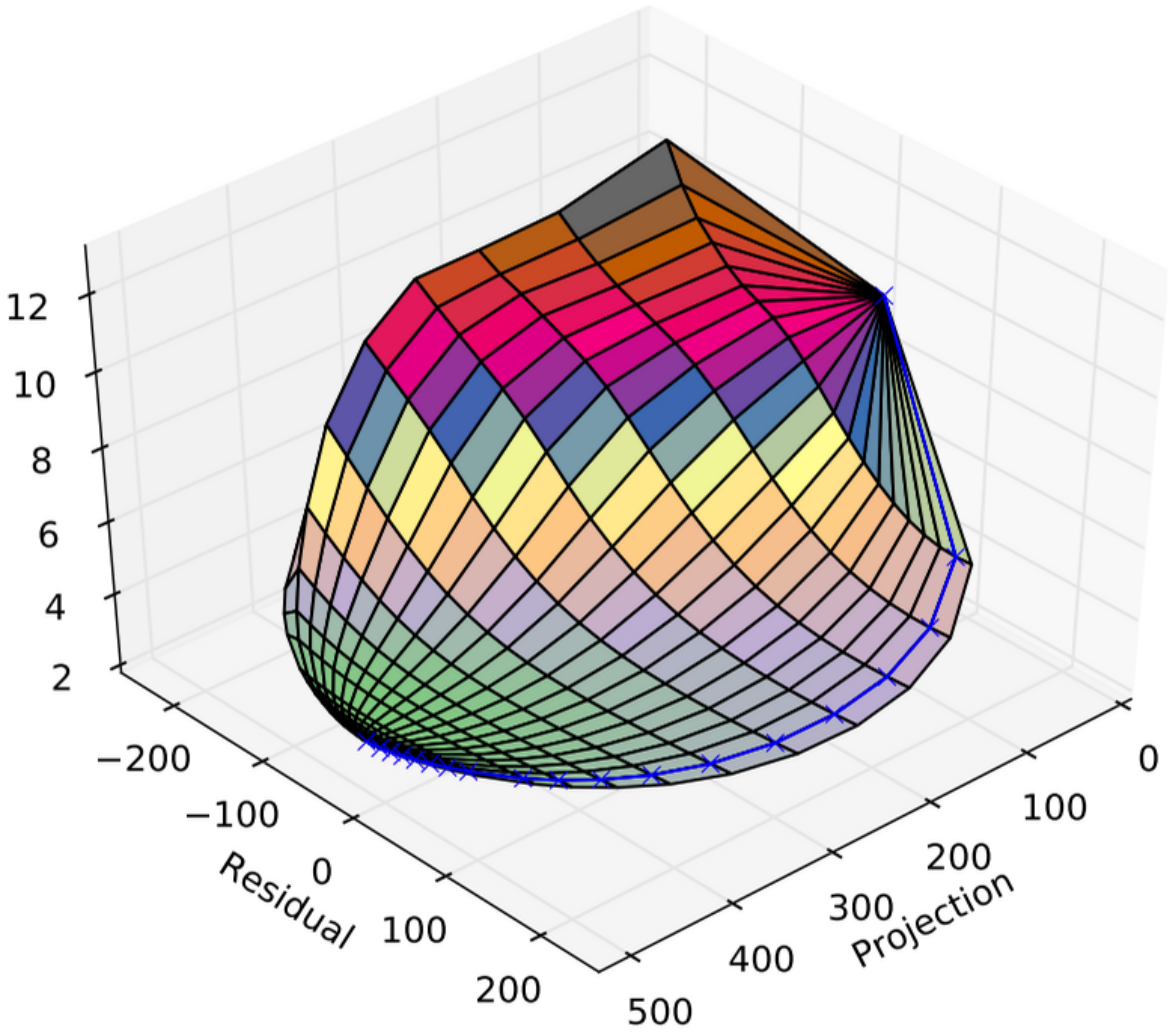




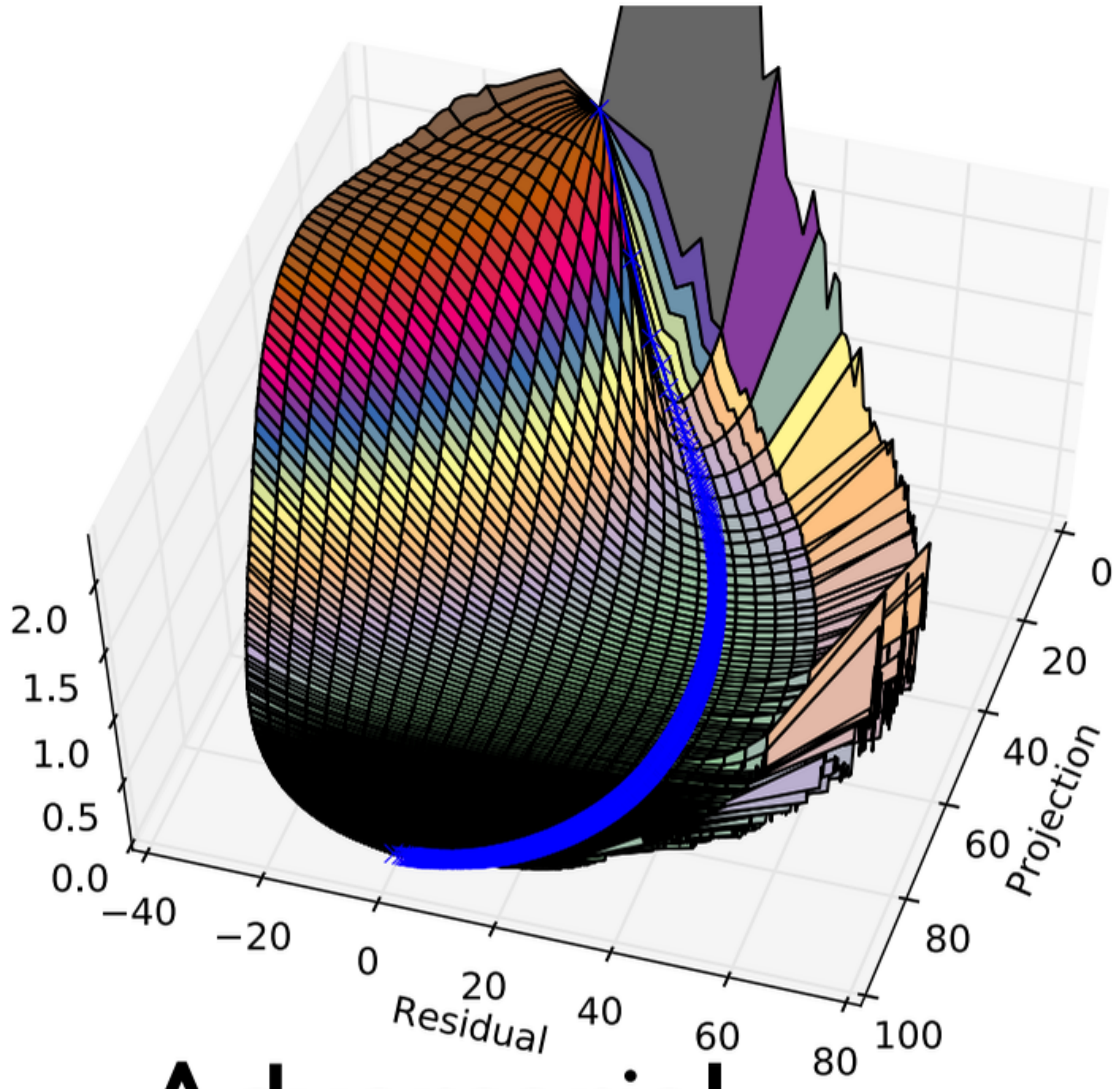
“Qualitatively Characterizing Neural Network Optimization Problems,”
Goodfellow, Vinyals and Saxe, ICLR 2015



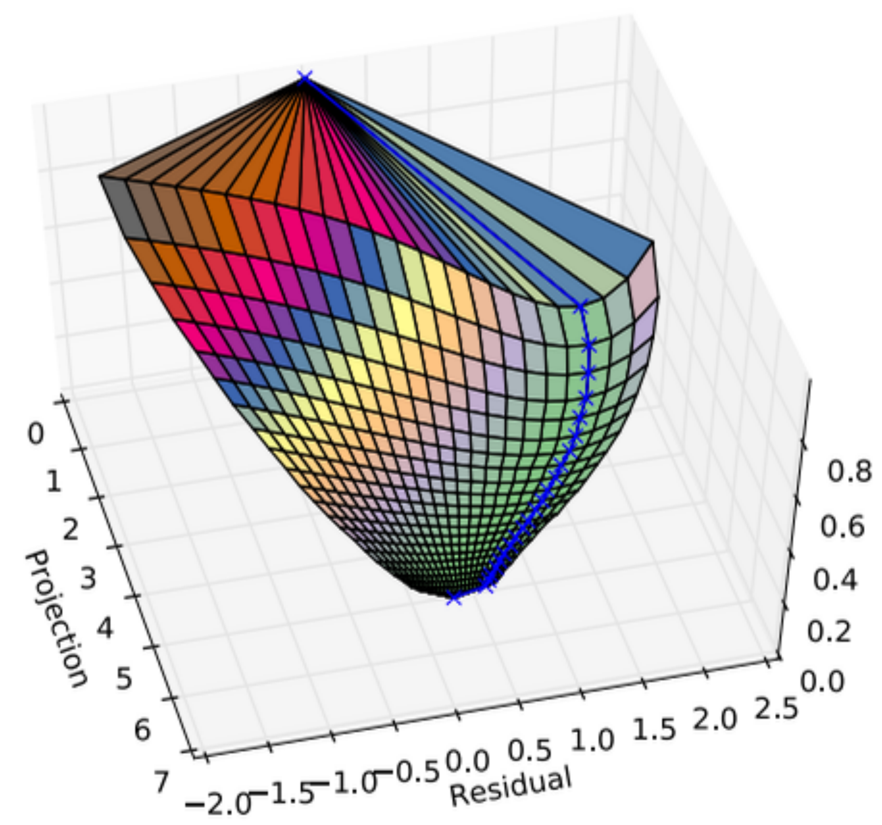
"Qualitatively Characterizing Neural Network Optimization Problems,"
Goodfellow, Vinyals and Saxe, ICLR 2015



LSTM



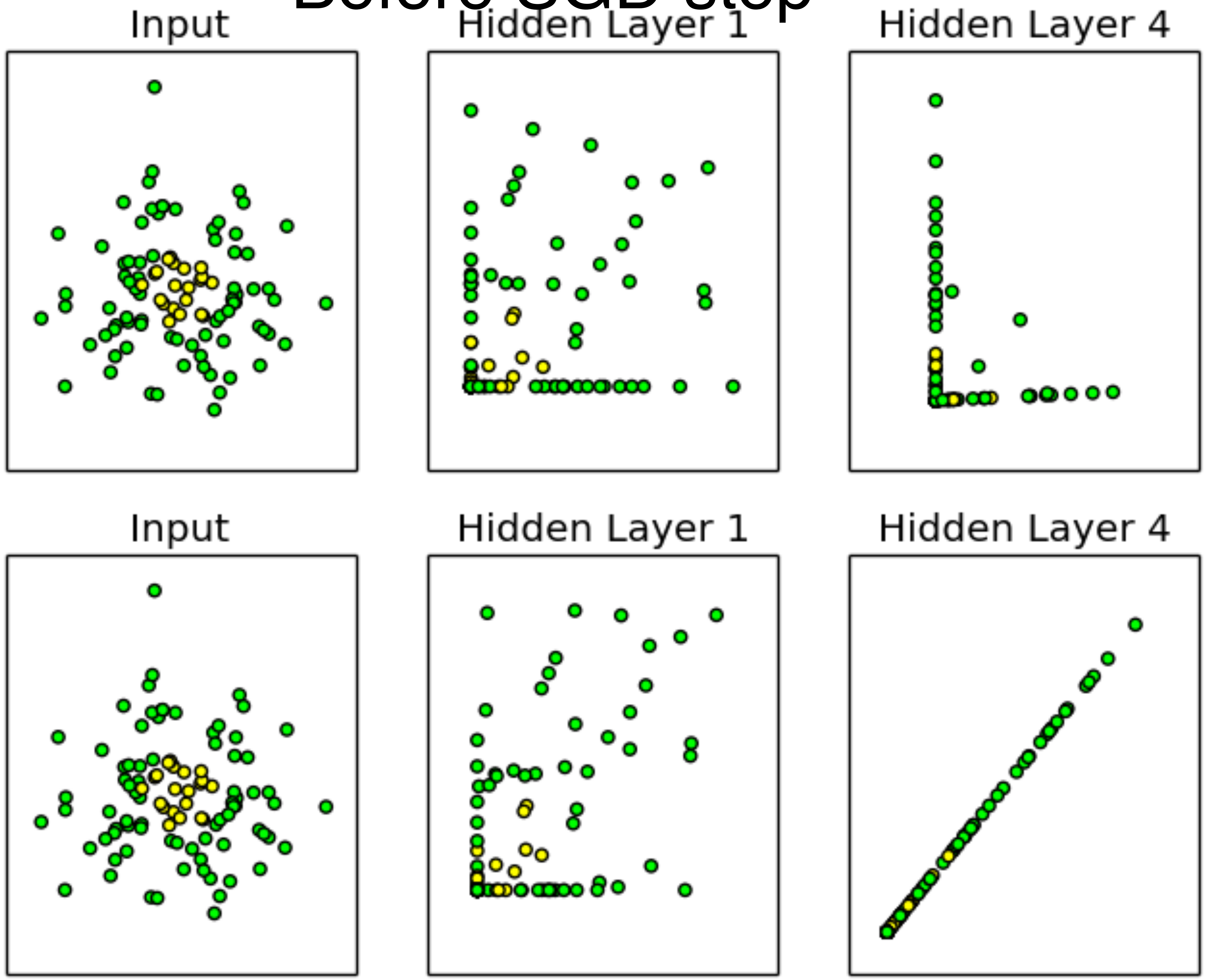
Adversarial
ReLUs



Factored Linear

"Qualitatively Characterizing Neural Network Optimization Problems,"
Goodfellow, Vinyals and Saxe, ICLR 2015

Before SGD step



After SGD step

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

Batch Normalization

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{W}$$

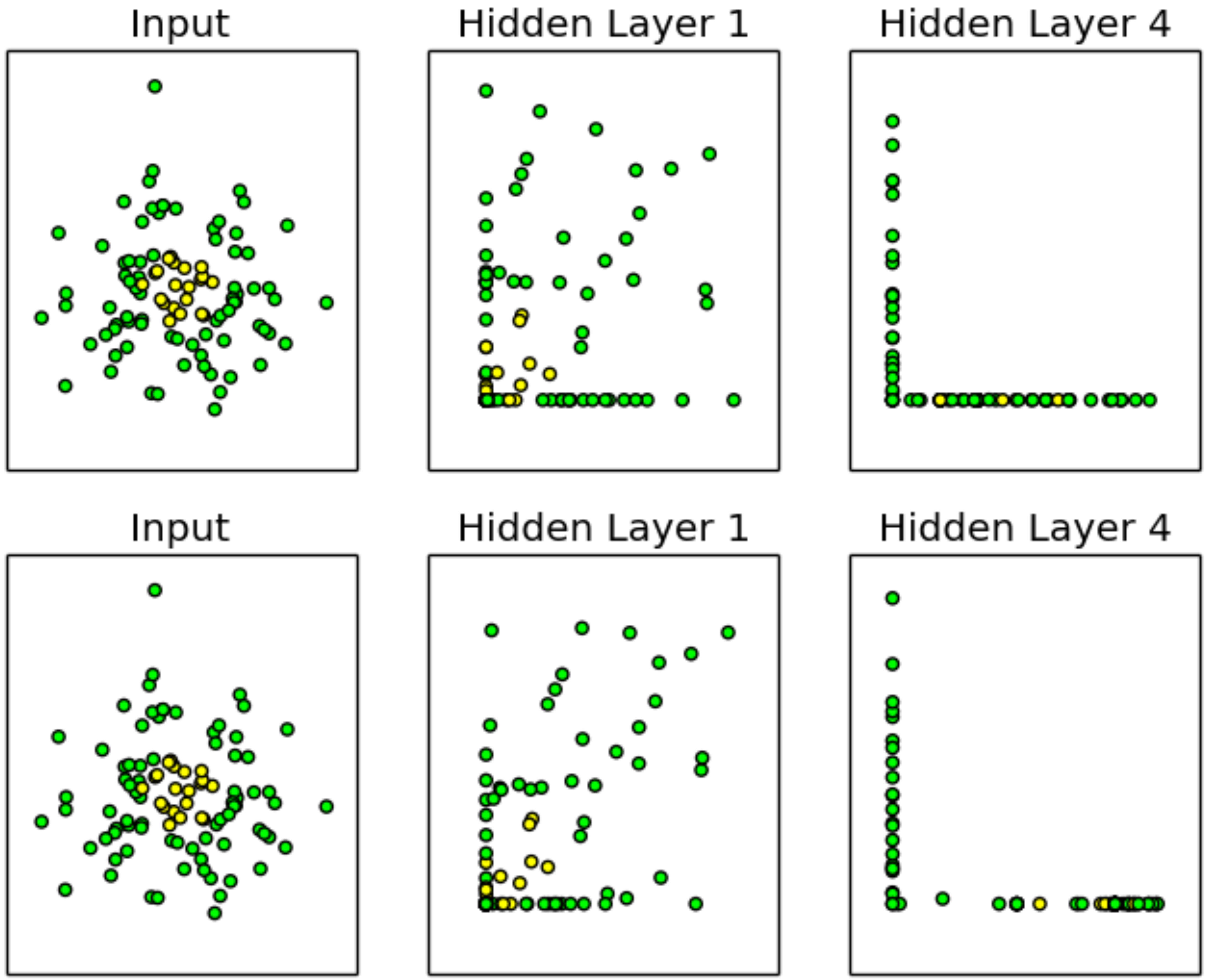
$$\tilde{\boldsymbol{Z}} = \boldsymbol{Z} - \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{Z}_{i,:}$$

$$\hat{\boldsymbol{Z}} = \frac{\tilde{\boldsymbol{Z}}}{\sqrt{\epsilon + \frac{1}{m}\sum_{i=1}^{m} \tilde{\boldsymbol{Z}}_{i,:}^{2}}}$$

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

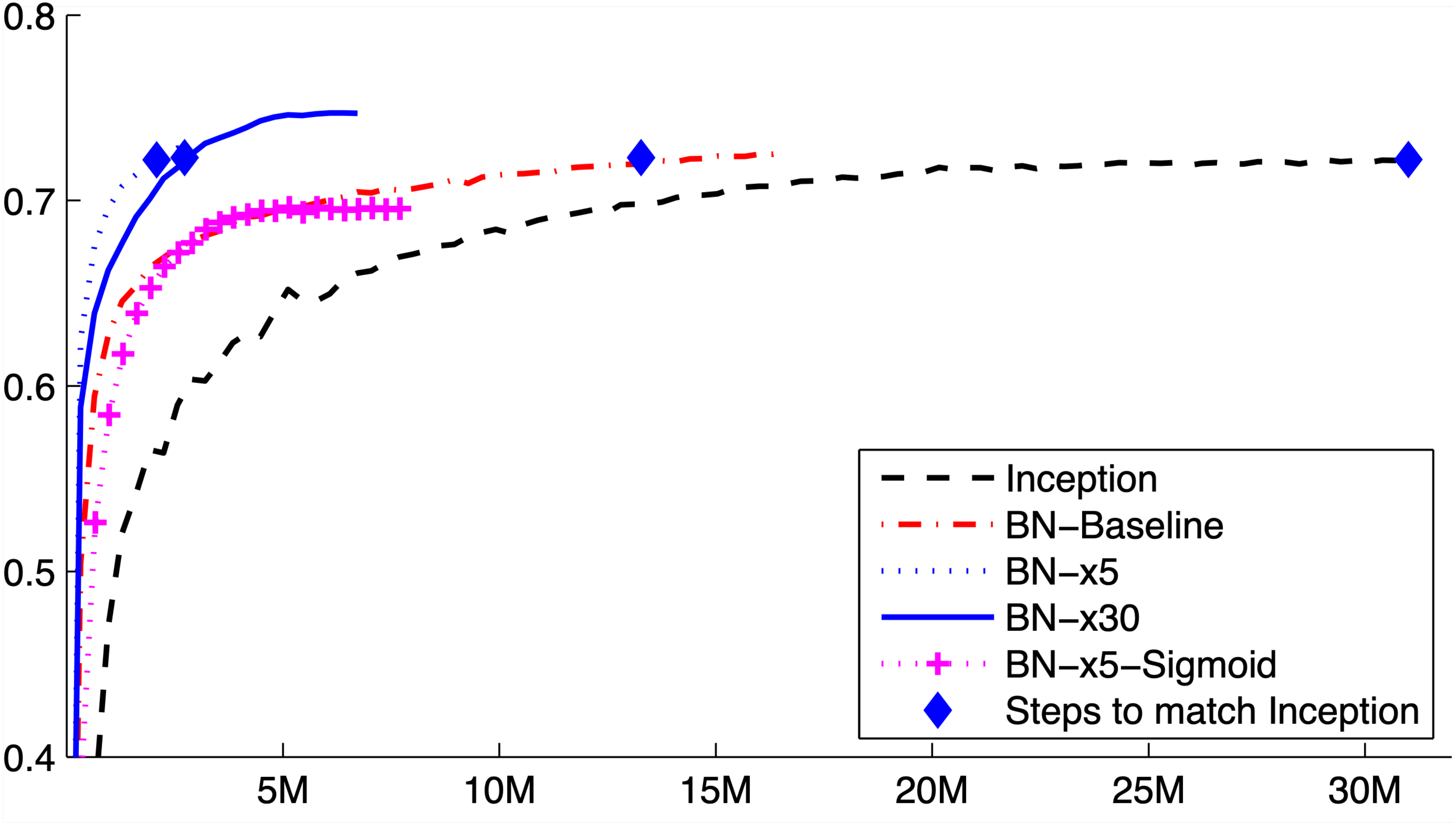
“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Ioffe and Szegedy 2015

Before SGD step

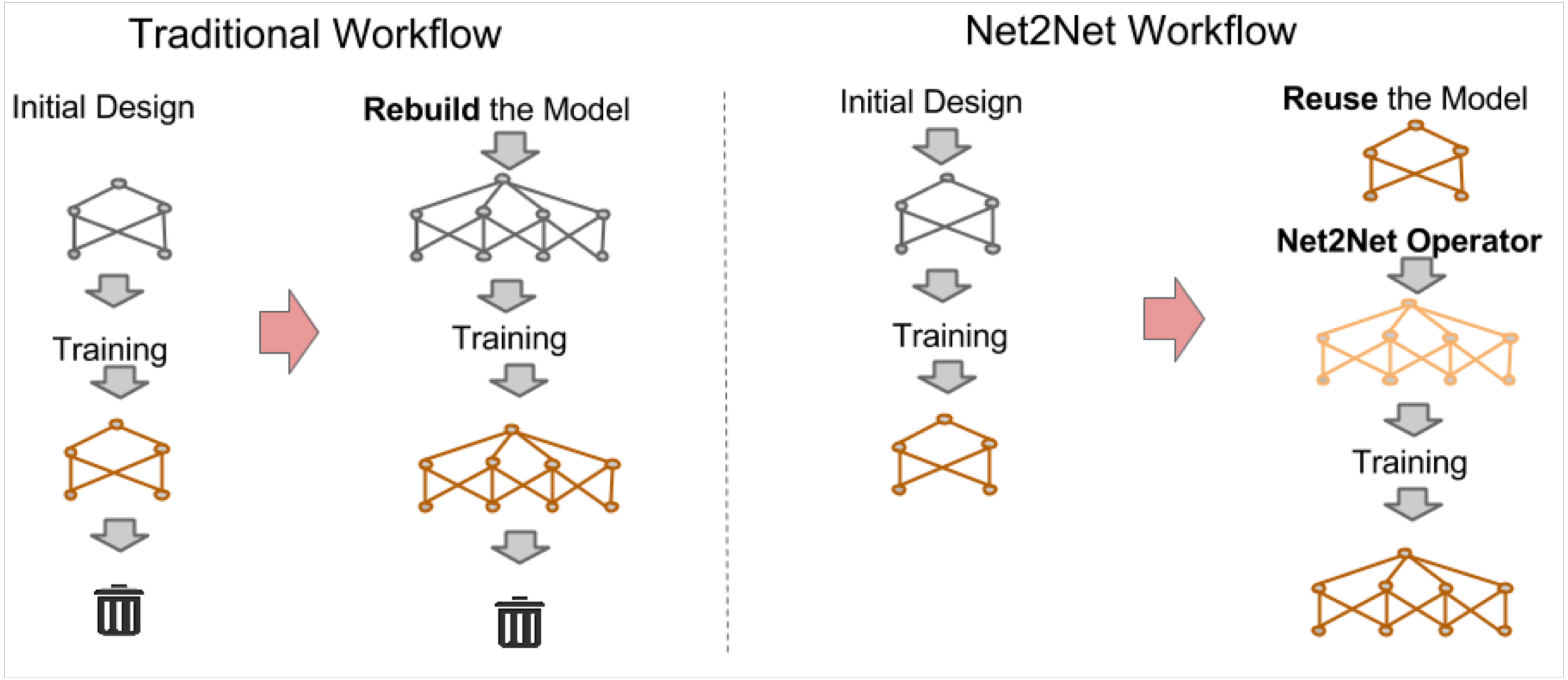


After SGD step

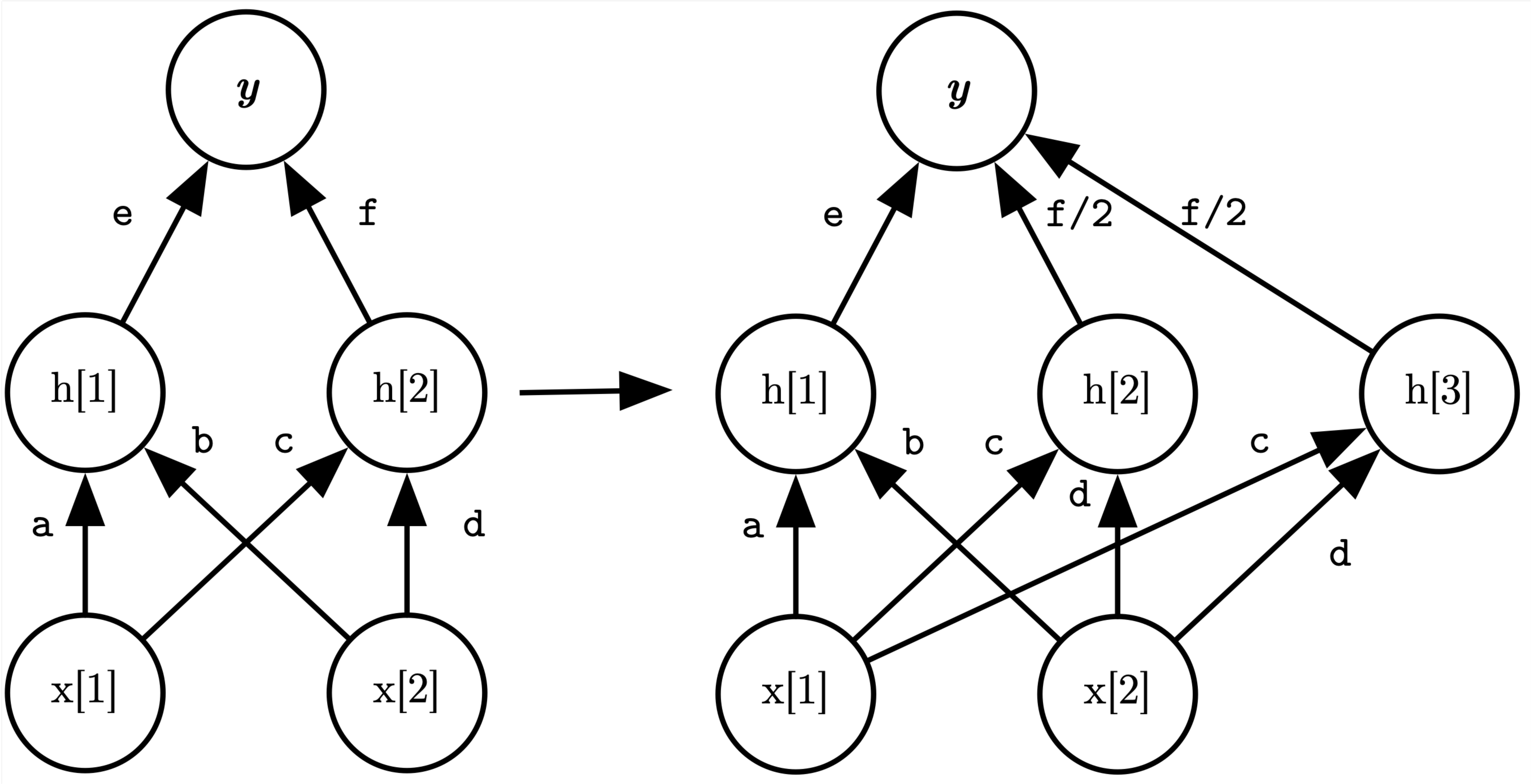
"Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," Ioffe and Szegedy 2015



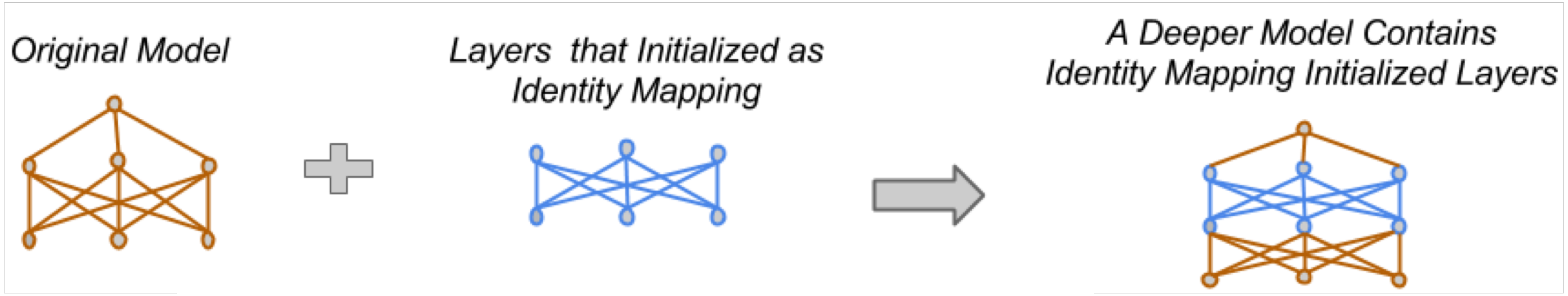
“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Ioffe and Szegedy 2015

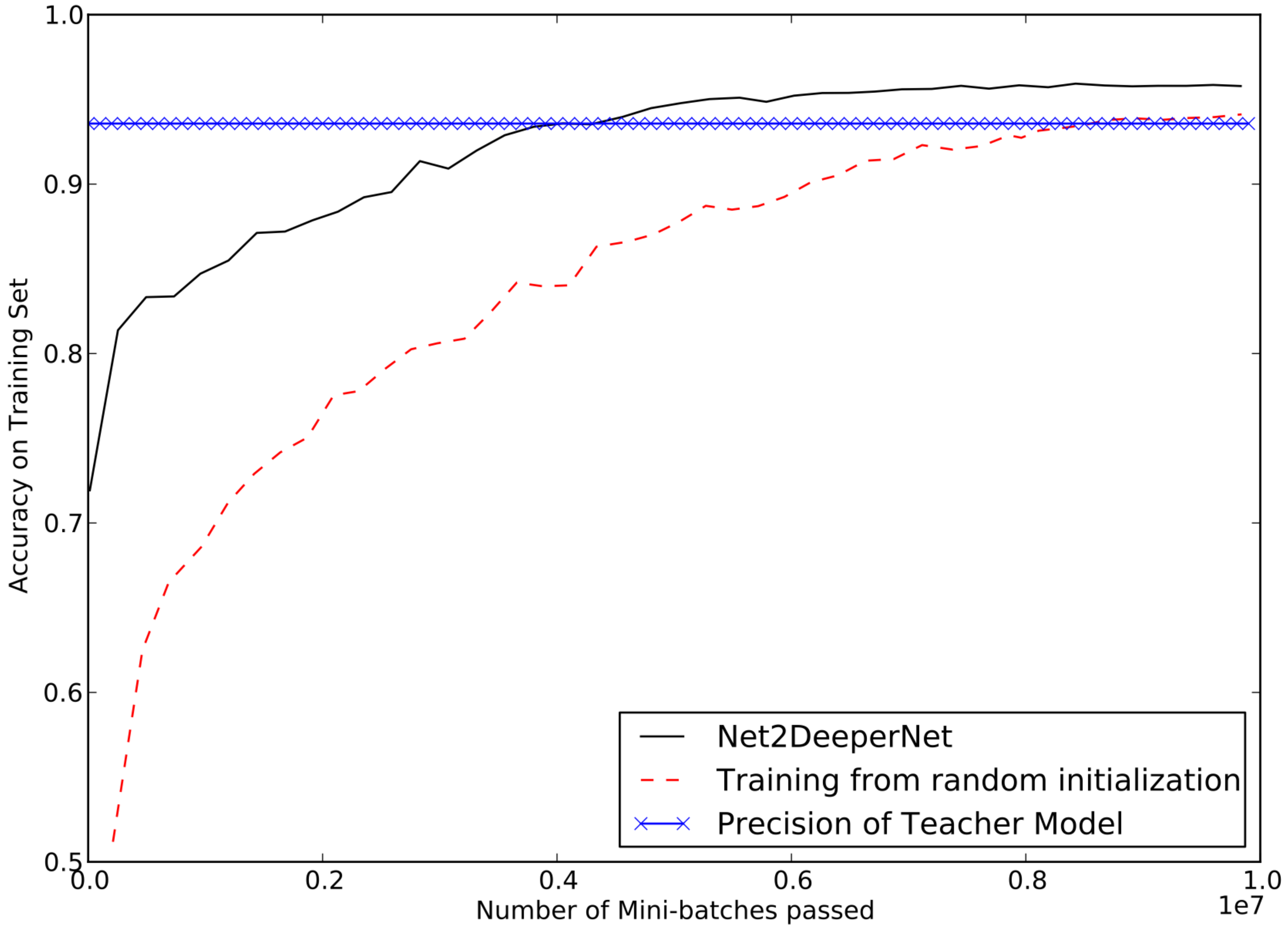


"Net2Net: Accelerating Learning via Knowledge Transfer." Chen, Goodfellow, and Shlens, submitted to ICLR 2016

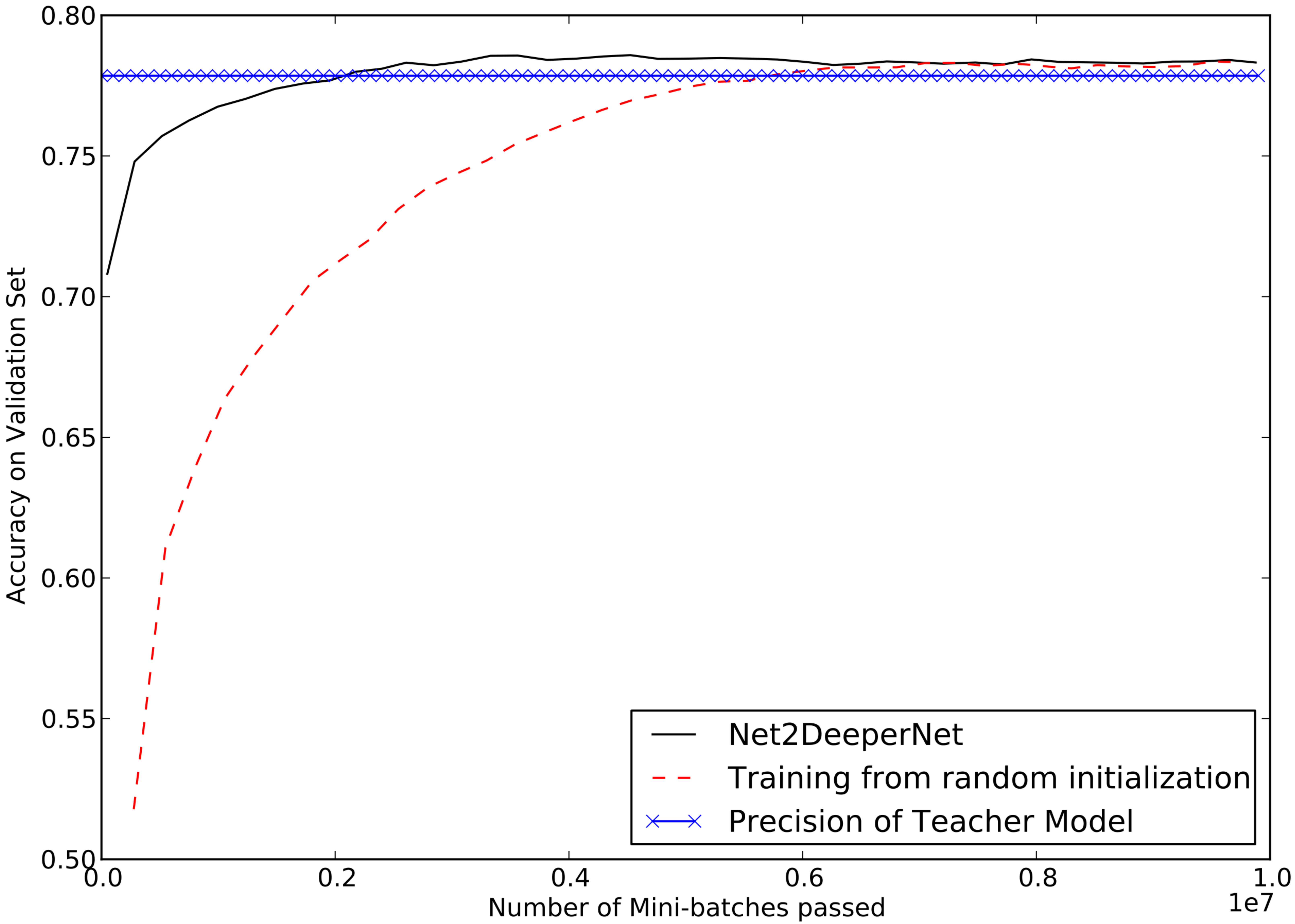


“Net2Net: Accelerating Learning via Knowledge Transfer.” Chen, Goodfellow, and Shlens, submitted to ICLR 2016





"Net2Net: Accelerating Learning via Knowledge Transfer." Chen, Goodfellow, and Shlens, submitted to ICLR 2016



“Net2Net: Accelerating Learning via Knowledge Transfer.” Chen, Goodfellow, and Shlens, submitted to ICLR 2016

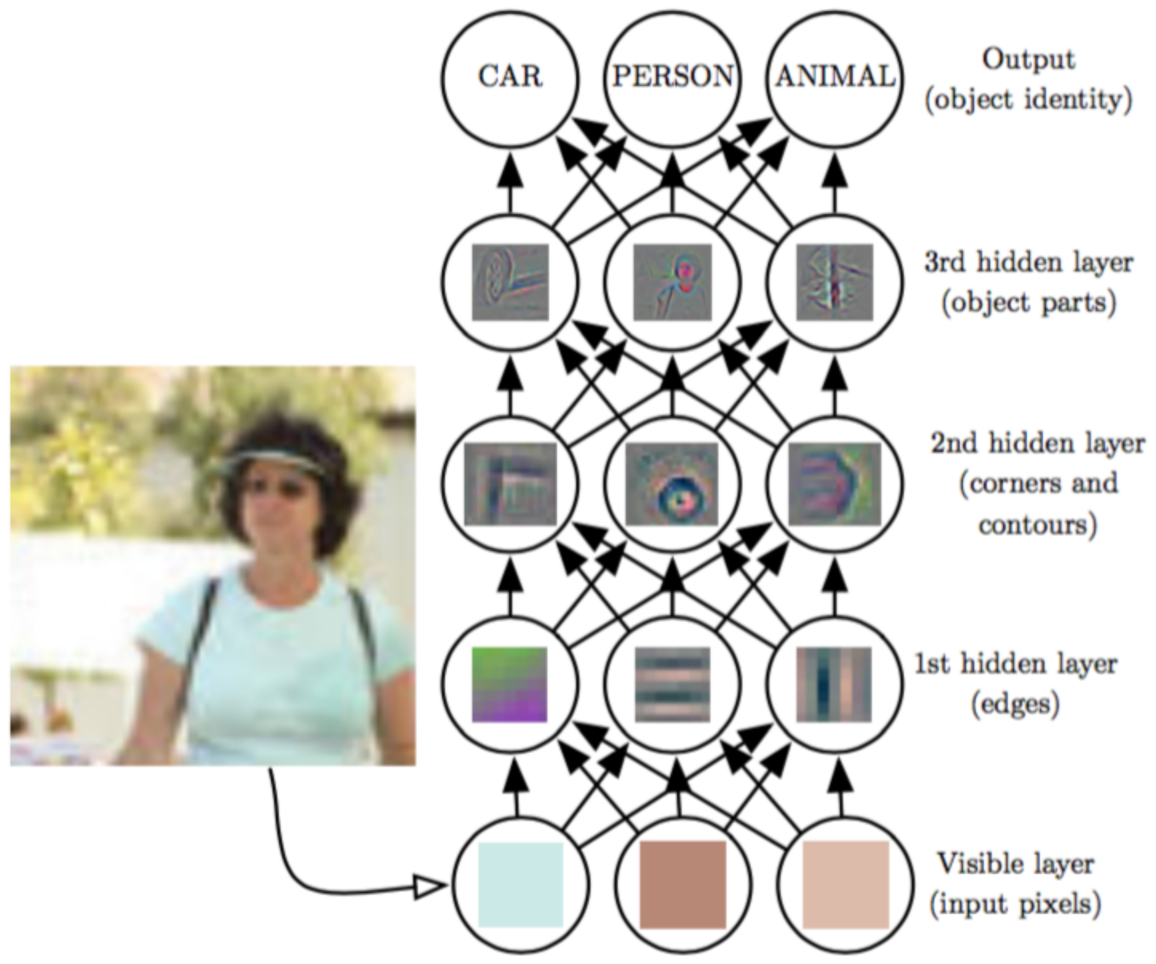


Figure 1.2: Illustration of a deep learning model. It is difficult for a computer to understand the meaning of raw sensory input data, such as this image represented as a collection of pixel values. The function mapping from a set of pixels to an object identity is very complicated. Learning or evaluating this mapping seems insurmountable if tackled directly. Deep learning resolves this difficulty by breaking the desired complicated mapping into a series of nested simple mappings, each described by a different layer of the model. The input is presented at the *visible layer*, so named because it contains the variables that we are able to observe. Then a series of *hidden layers* extracts increasingly abstract features from the image. These layers are called "hidden" because their values are not given in the data; instead the model must determine which concepts are useful for explaining the relationships in the observed data. The images here are visualizations of the kind of feature represented by each hidden unit. Given the pixels, the first layer can easily identify edges, by comparing the brightness of neighboring pixels. Given the first hidden layer's description of the edges, the second hidden layer can easily search for corners and extended contours, which are recognizable as collections of edges. Given the second hidden layer's description of the image in terms of corners and contours, the third hidden layer can detect entire parts of specific objects, by finding specific collections of contours and corners. Finally, this description of the image in terms of the object parts it contains can be used to recognize the objects present in the image. Images reproduced with permission from Zeiler and Fergus (2014).

Deep Learning

Goodfellow, Bengio,
and Courville

www.deeplearningbook.org

In preparation for MIT Press