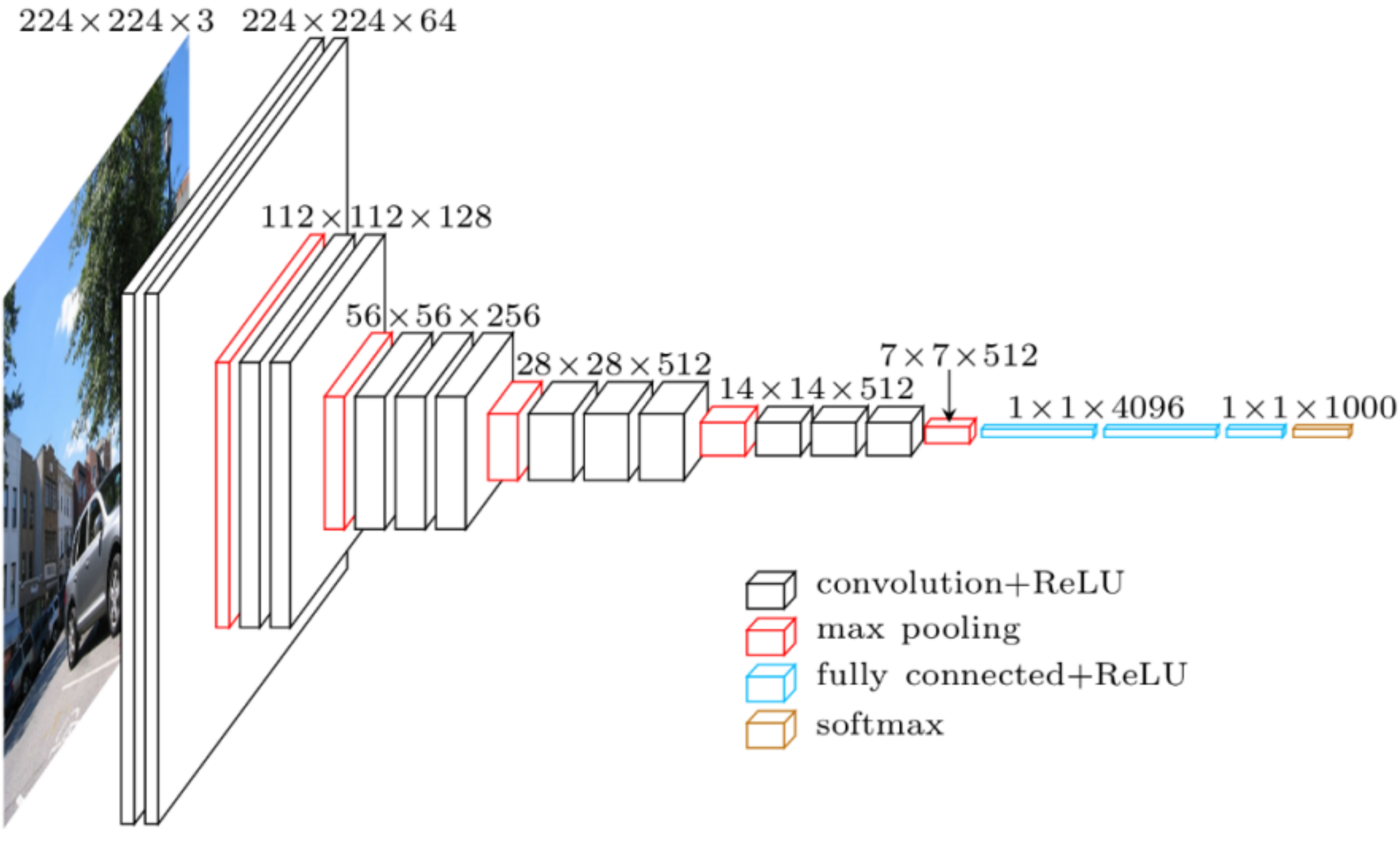
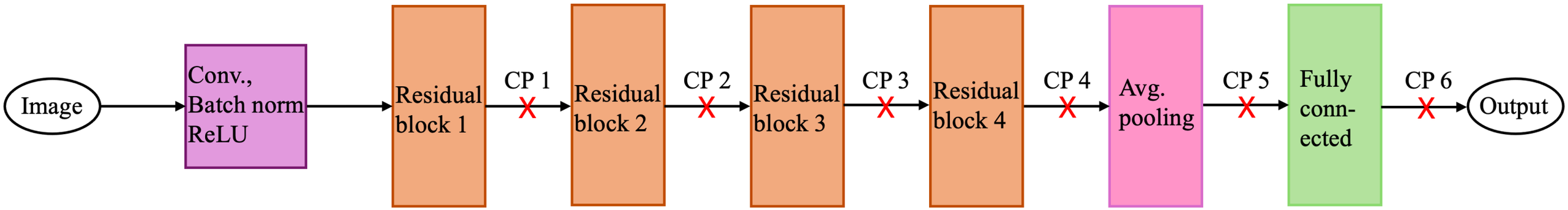
Basic Structure of CNN

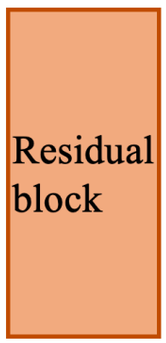


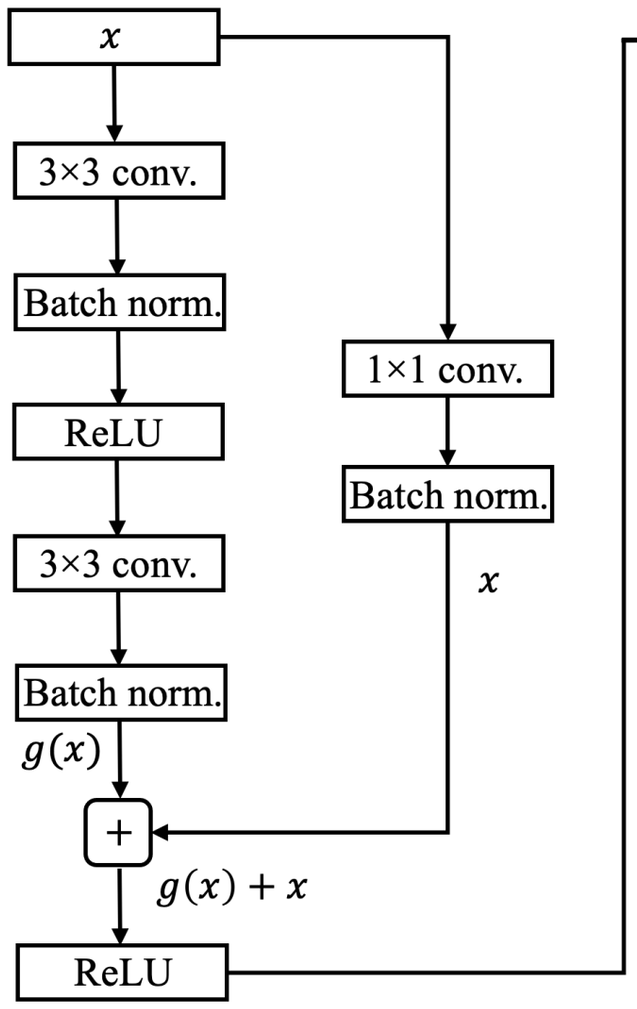
Some well known CNN architectures in the field

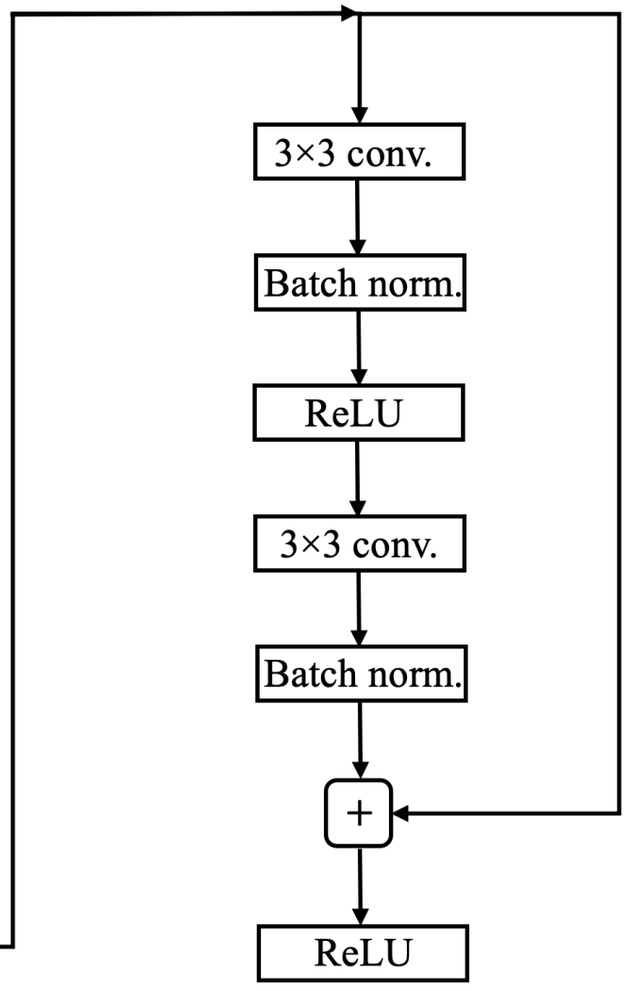
- **LeNet**. The first successful applications of Convolutional Networks were developed by Yann LeCun in 1990's. Of these, the best known is the LeNet architecture that was used to read zip codes, digits, etc.
- **AlexNet**. The first work that popularized Convolutional Networks in Computer Vision was the AlexNet, developed by Alex Krizhevsky, Ilya Sutskever and Geoff Hinton. The AlexNet was submitted to the ImageNet ILSVRC challenge in 2012 and significantly outperformed the second runner-up (top 5 error of 16% compared to runner-up with 26% error). The Network had a very similar architecture to LeNet, but was deeper, bigger, and featured Convolutional Layers stacked on top of each other (previously it was common to only have a single CONV layer always immediately followed by a POOL layer).
- **ZF Net**. The ILSVRC 2013 winner was a Convolutional Network from Matthew Zeiler and Rob Fergus. It became known as the ZFNet (short for Zeiler & Fergus Net). It was an improvement on AlexNet by tweaking the architecture hyperparameters, in particular by expanding the size of the middle convolutional layers and making the stride and filter size on the first layer smaller.
- **GoogLeNet**. The ILSVRC 2014 winner was a Convolutional Network from Szegedy et al. from Google. Its main contribution was the development of an *Inception Module* that dramatically reduced the number of parameters in the network (4M, compared to AlexNet with 60M). Additionally, this paper uses Average Pooling instead of Fully Connected layers at the top of the ConvNet, eliminating a large amount of parameters that do not seem to matter much. There are also several followup versions to the GoogLeNet, most recently Inception-v4.
- **VGGNet**. The runner-up in ILSVRC 2014 was the network from Karen Simonyan and Andrew Zisserman that became known as the VGGNet. Its main contribution was in showing that the depth of the network is a critical component for good performance. Their final best network contains 16 CONV/FC layers and, appealingly, features an extremely homogeneous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end. Their pretrained model is available for plug and play use in Caffe. A downside of the VGGNet is that it is more expensive to evaluate and uses a lot more memory and parameters (140M). Most of these parameters are in the first fully connected layer, and it was since found that these FC layers can be removed with no performance downgrade, significantly reducing the number of necessary parameters.
- **ResNet**. Residual Network developed by Kaiming He et al. was the winner of ILSVRC 2015. It features special *skip connections* and a heavy use of batch normalization. The architecture is also missing fully connected layers at the end of the network. The reader is also referred to Kaiming's presentation (video, slides), and some recent experiments that reproduce these networks in Torch. ResNets are currently by far state of the art Convolutional Neural Network models and are the default choice for using ConvNets in practice (as of May 10, 2016). In particular, also see more recent developments that tweak the original architecture from Kaiming He et al. Identity Mappings in Deep Residual Networks (published March 2016).

ResNet-18



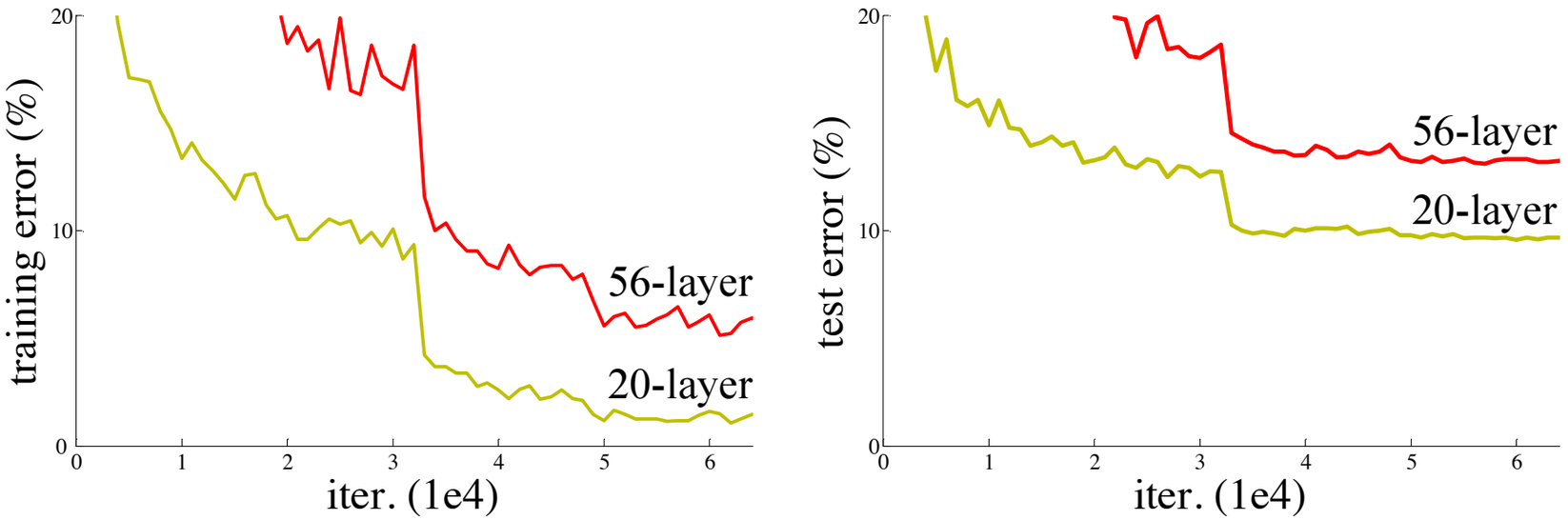
Each  represents





	Dimensionality
Checkpoint 1	$64 \times 32 \times 32$
Checkpoint 2	$128 \times 16 \times 16$
Checkpoint 3	$256 \times 8 \times 8$
Checkpoint 4	$512 \times 4 \times 4$
Checkpoint 5	$512 \times 1 \times 1$
Checkpoint 6	$10 \times 1 \times 1$

Degradation Problem and Residual Blocks



He *et al*. “Deep residual learning for image recognition”, 2016

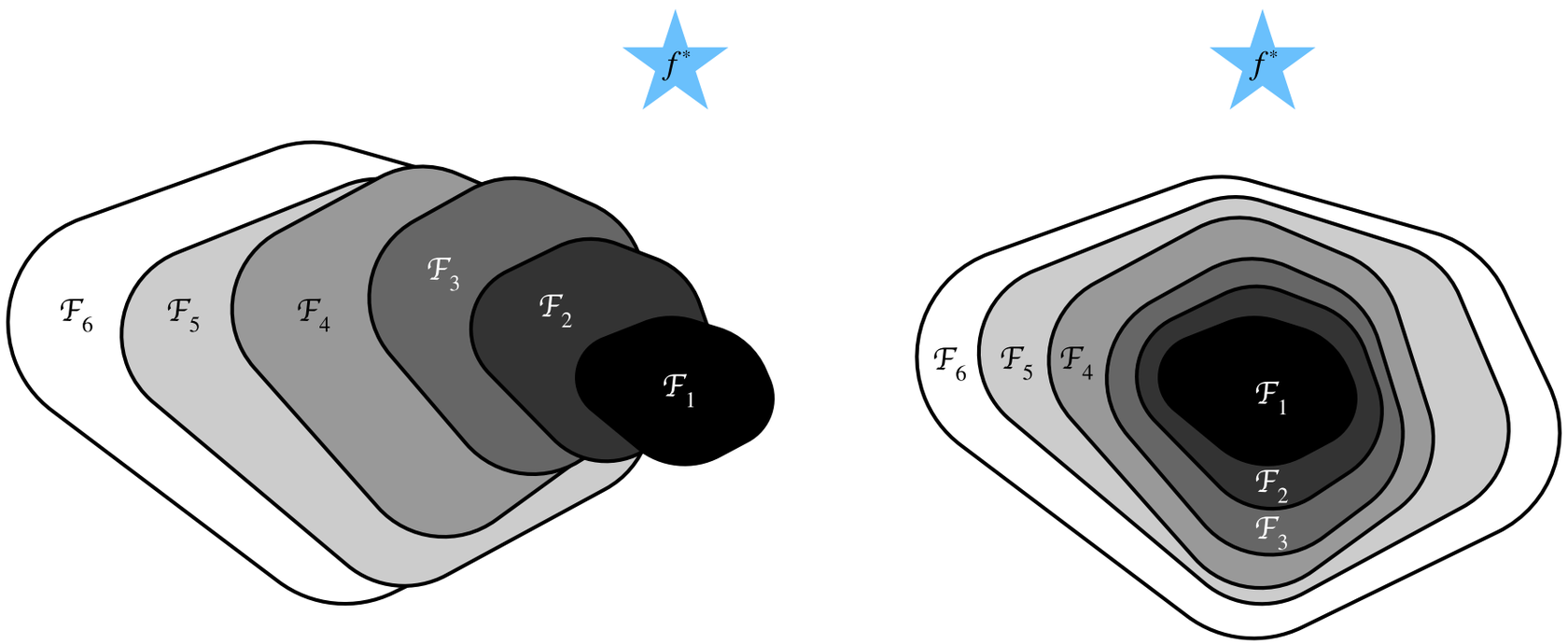


Fig. 8.6.1 (Dive into DL)

Various Normalizations

$$Normalize(x) = \gamma \odot \frac{x - \hat{\mu}}{\hat{\sigma}} + \beta$$

- $\hat{\mu}$ and $\hat{\sigma}$ are estimated by the BLUE parts for different types of normalizations
- γ and β are trainable parameters

Height **W**idth # of **C**hannels **N:** (Mini)-batch size

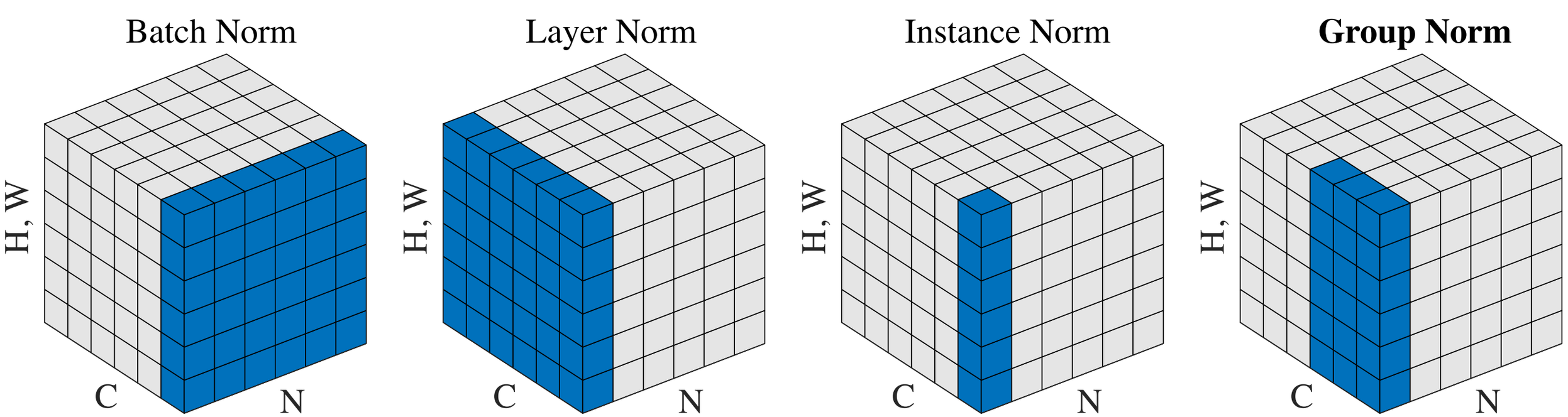


Fig. taken from https://arxiv.org/pdf/1803.08494.pdf

CIFAR-10 dataset

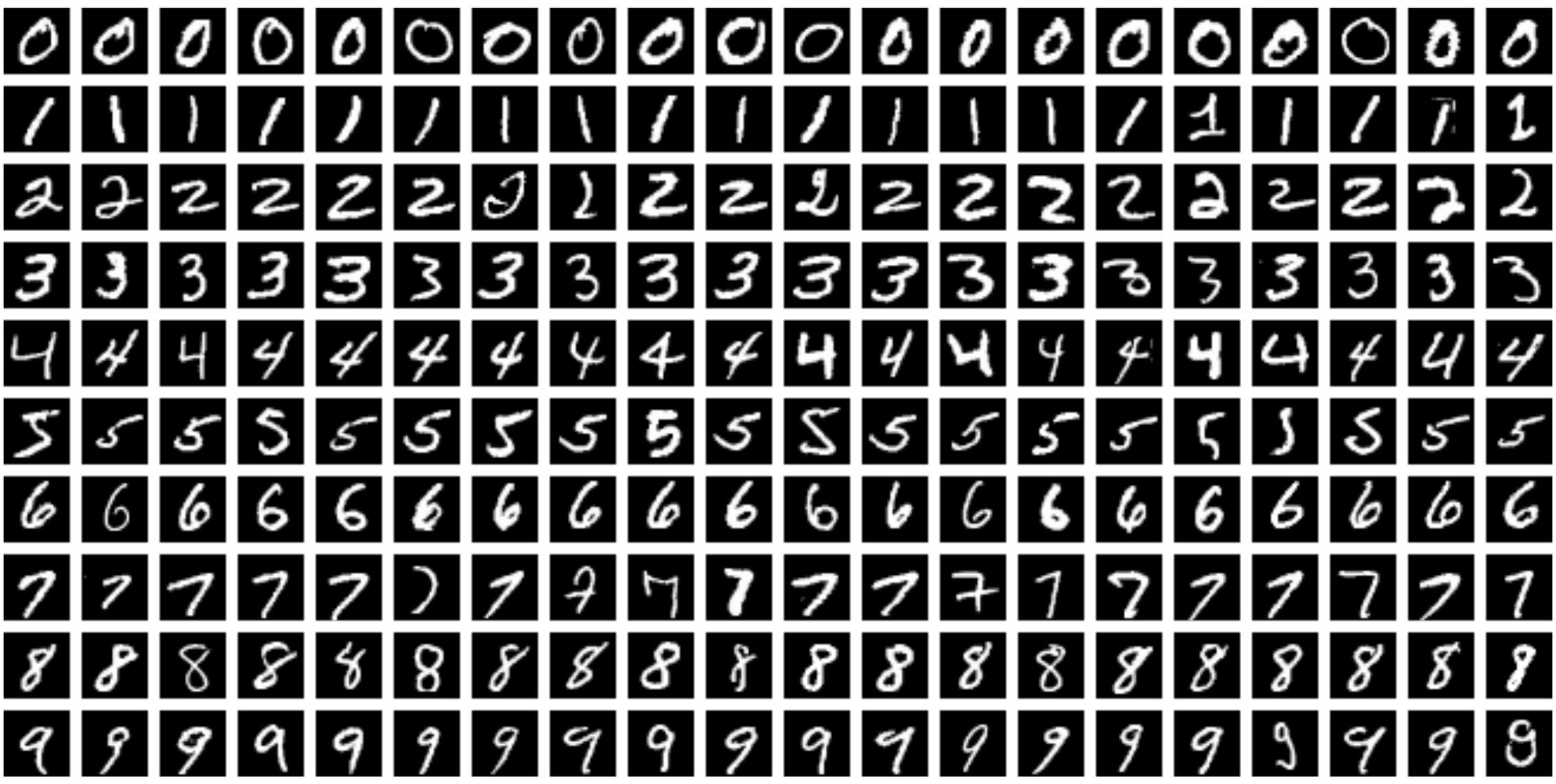
CIFAR: Canadian Institute for Advanced Research

Data: 32x32x3, 60000 Images

ResNet-18: ~ 11 millions parameters

MNIST dataset

MNIST: Modified National Institute of Standards and Technology

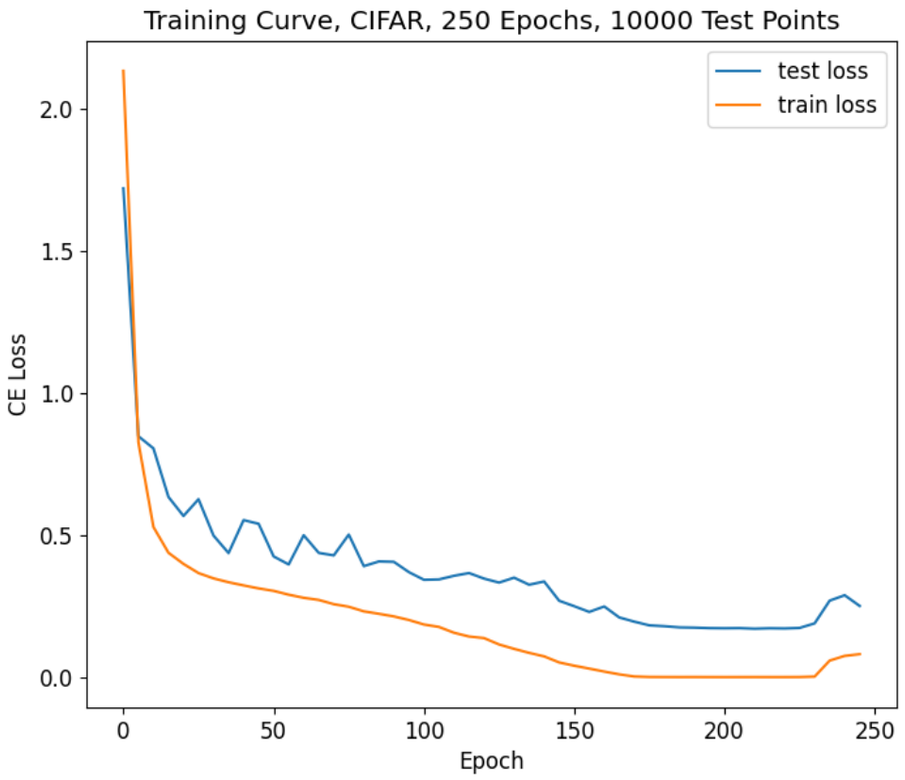


Data: 60000 Images

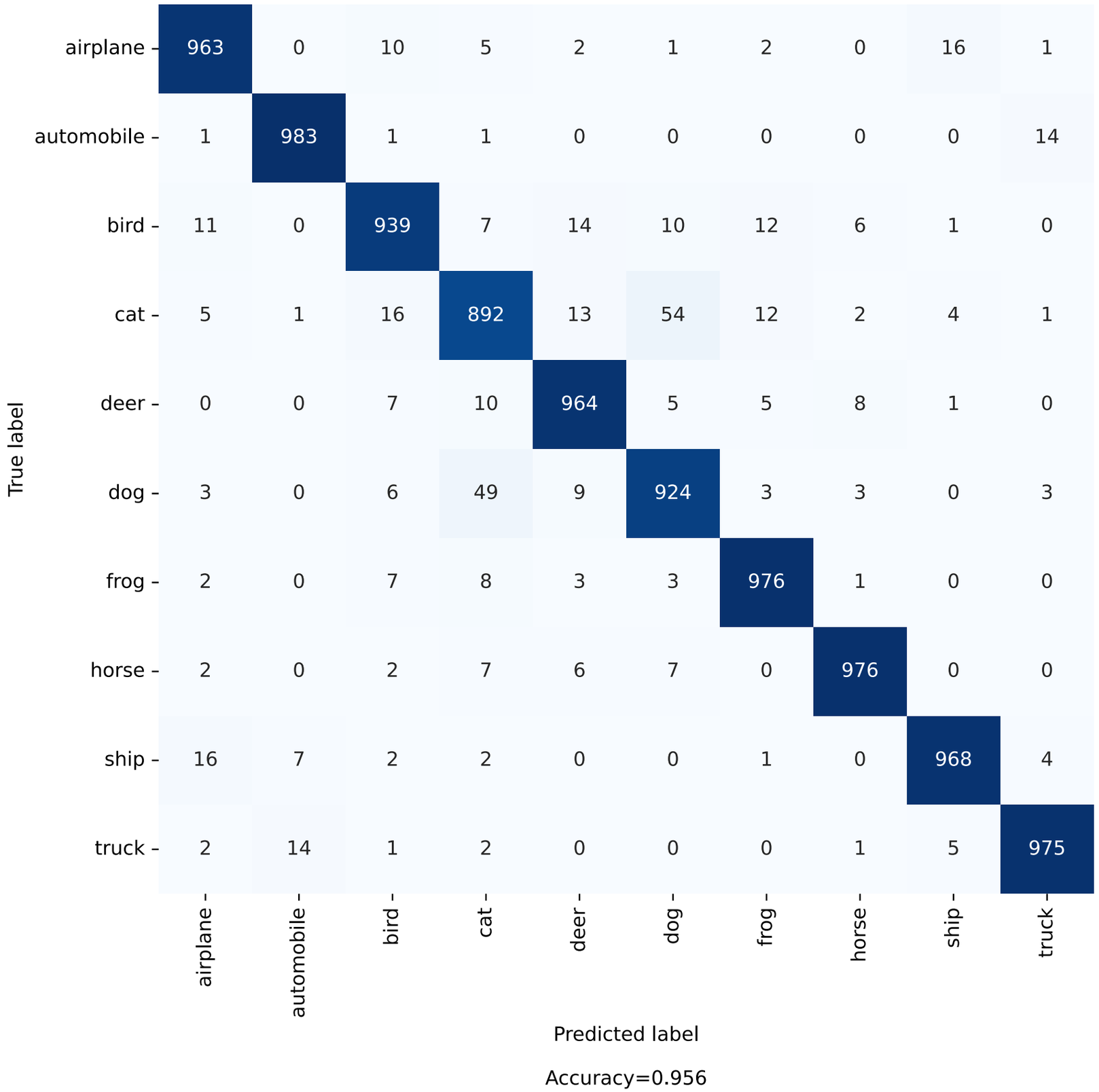
Look Into The Blackbox: CIFAR-10

Nik Tavakolian

Training/Validation Curves



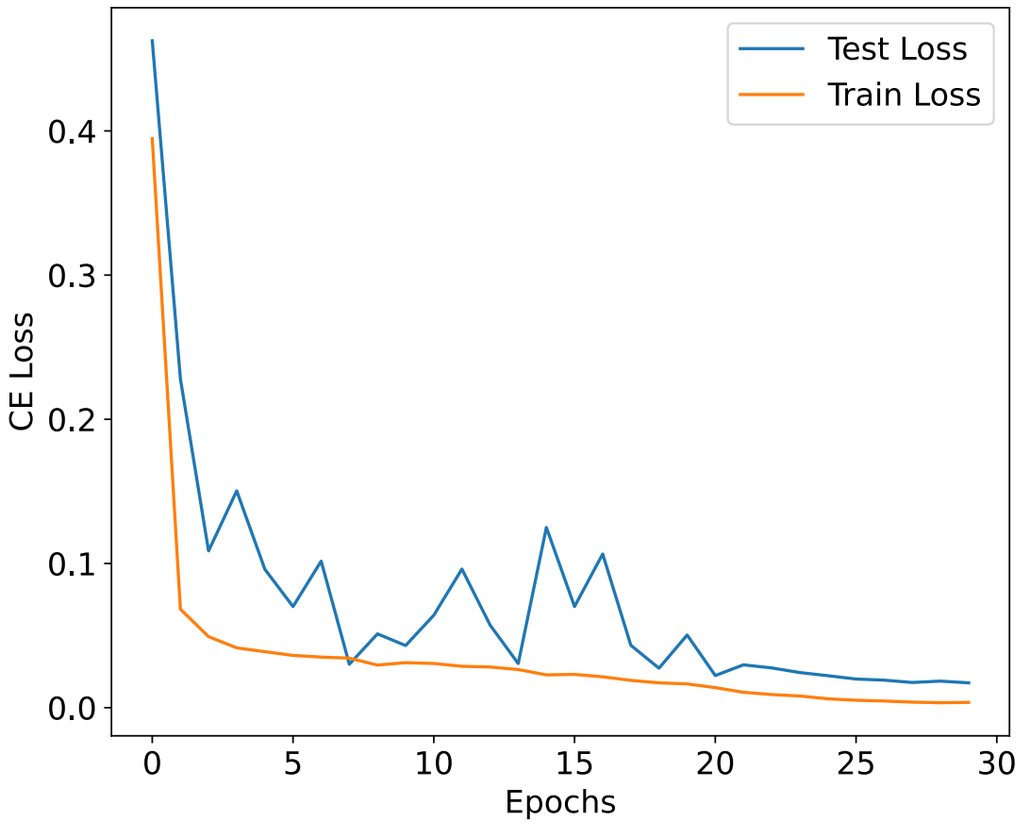
Confusion Matrix



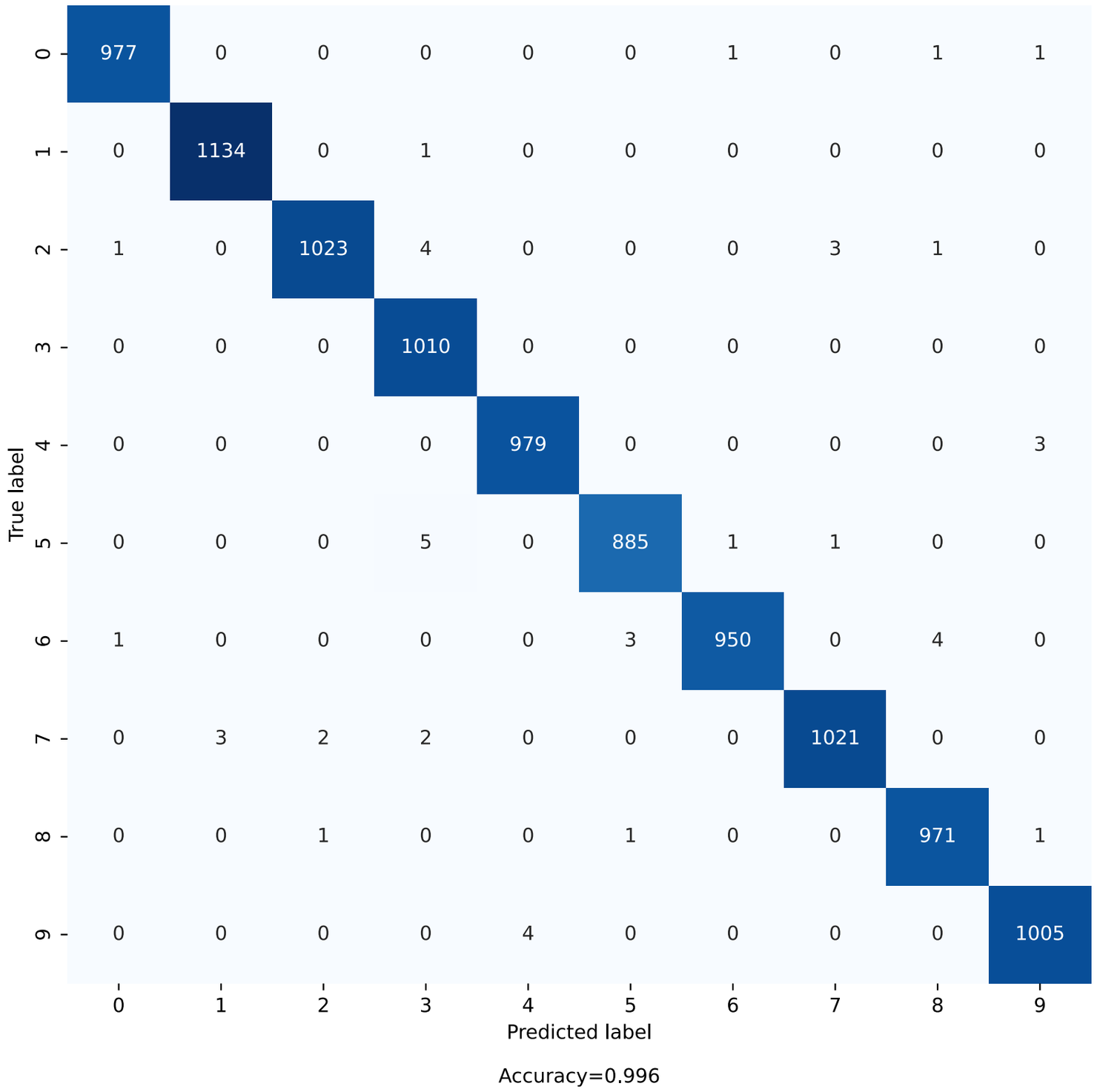
Look Into The Blackbox: MNIST

Nik Tavakolian

Training/Validation Curves



Confusion Matrix



Robust Data Visualization (SASNE)

Wängberg *et al. BMC Bioinformatics* (2022) 23:477
https://doi.org/10.1186/s12859-022-05028-8

BMC Bioinformatics

RESEARCH Open Access

Shape-aware stochastic neighbor embedding for robust data visualisations

Tobias Wängberg, Joanna Tyrcha* and Chun-Biu Li*

*Correspondence: joanna@math.su.se; cbli@math.su.se

Department of Mathematics, Stockholm University, Stockholm, Sweden

Abstract

Background: The t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm has emerged as one of the leading methods for visualising high-dimensional (HD) data in a wide variety of fields, especially for revealing cluster structure in HD single-cell transcriptomics data. However, t-SNE often fails to correctly represent hierarchical relationships between clusters and creates spurious patterns in the embedding. In this work we generalised t-SNE using shape-aware graph distances to mitigate some of the limitations of the t-SNE. Although many methods have been recently proposed to circumvent the shortcomings of t-SNE, notably Uniform manifold approximation (UMAP) and Potential of heat diffusion for affinity-based transition embedding (PHATE), we see a clear advantage of the proposed graph-based method.

Results: The superior performance of the proposed method is first demonstrated on simulated data, where a significant improvement compared to t-SNE, UMAP and PHATE, based on quantitative validation indices, is observed when visualising imbalanced, nonlinear, continuous and hierarchically structured data. Thereafter the ability of the proposed method compared to the competing methods to create faithfully low-dimensional embeddings is shown on two real-world data sets, the single-cell transcriptomics data and the MNIST image data. In addition, the only hyper-parameter of the method can be automatically chosen in a data-driven way, which is consistently optimal across all test cases in this study.

Conclusions: In this work we show that the proposed shape-aware stochastic neighbor embedding method creates low-dimensional visualisations that robustly and accurately reveal key structures of high-dimensional data.

Keywords: Data visualisation, Dimensionality reduction, Graph distance, Dimensionality reduction validation

Transpose Convolution

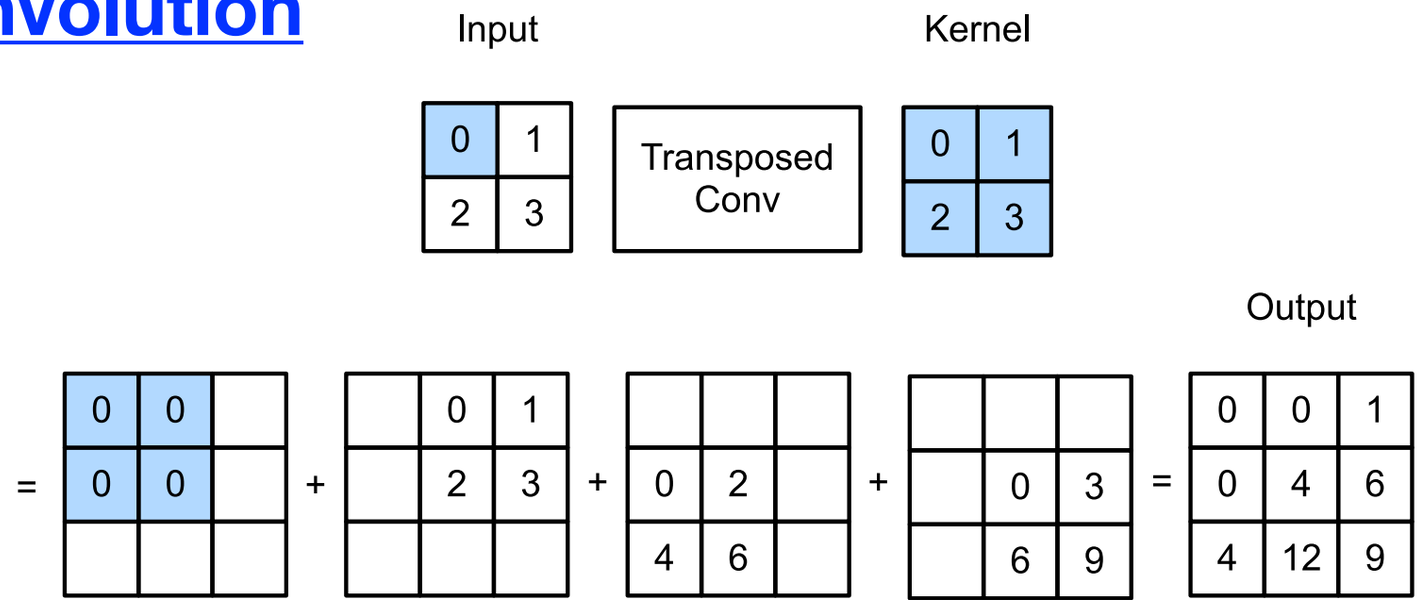


Fig. 14.10.1 (Dive in DL) 2x2 kernel, stride of 1

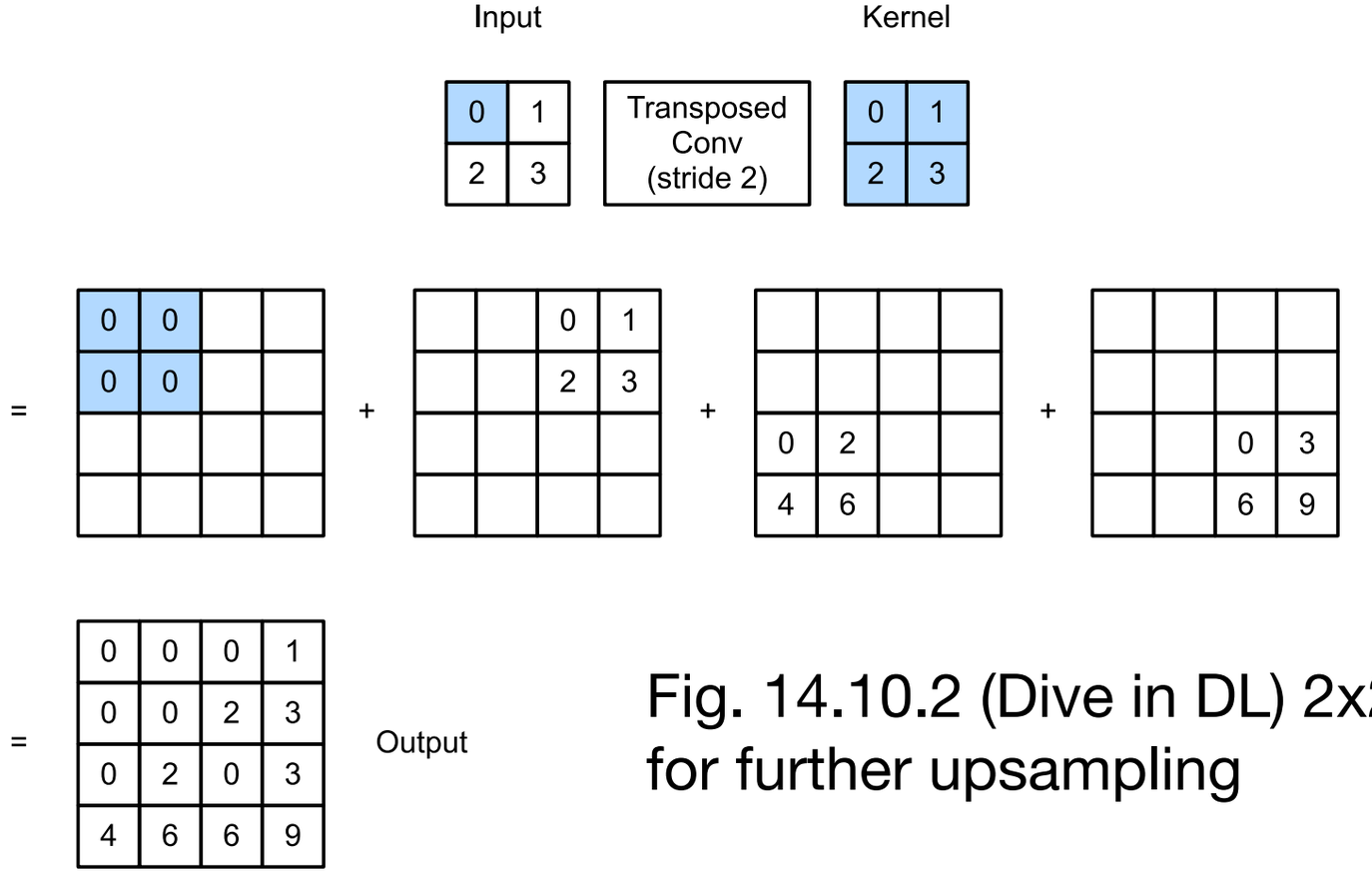


Fig. 14.10.2 (Dive in DL) 2x2 kernel, stride of 2 for further upsampling