Sections 7.3 and 7.4: Network in Network (NiN) and Networks with Parallel Concatenations (GoogLeNet)

Jake Helton

Brief Background Information

We initially wanted to work with image data. To deal with these images, we simply discarded the spatial structure of each image by flattening them into one-dimensional vectors. Preferably, we would leverage our prior knowledge that nearby pixels are typically related to each other. We would then use this information to build efficient models for learning from image data.

Modern Convolutional Neural Networks

1. Deep Convolutional Neural Networks (AlexNet).
2. Blocked Networks (VGG).

Last Lecture

3. Network in Network (NiN).
4. Parallel Concatenation Networks (GoogLeNet).

This Lecture

5. Residual Networks (ResNet).
6. Densely Connected Networks (DenseNet).

Next Lecture

Modern Convolutional Neural Networks

1. AlexNet (Krizhevsky et al. 2012) has 105,636 citations.
2. VGG (Simonyan & Zisserman 2014) has 76,686 citations.
3. NiN (Lin et al. 2013) has 6,579 citations.
4. GoogLeNet (Szegedy et al. 2015) has 38,708 citations.
5. ResNet (He et al. 2016) has 113,042 citations.
6. DenseNet (Huang et al. 2019) has 23,921 citations.

NiN (Lin et al. 2013)

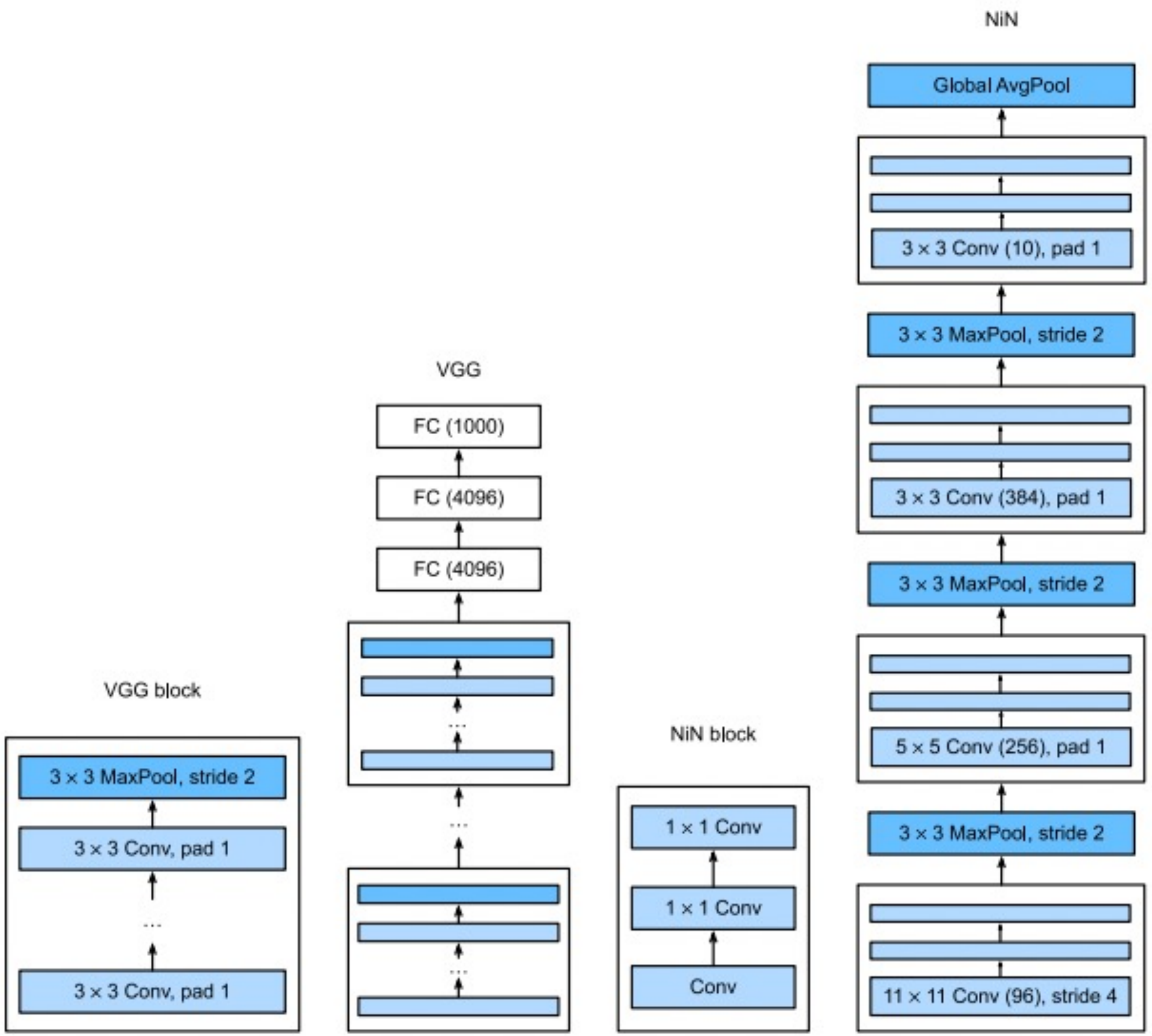


Fig. 7.3.1: Comparing architectures of VGG and NiN, and their blocks.

1 2022MNRAS.tmp..779Z 2022/03

Extracting Photometric Redshift from Galaxy Flux and Image Data using Neural Networks in the CSST Survey

Zhou, Xingchen; Gong, Yan; Meng, Xian-Min *and 6 more*

2 2022MNRAS.510.4504T 2022/03

Radio Galaxy Zoo: giant radio galaxy classification using multidomain deep learning

Tang, H.; Scaife, A. M. M.; Wong, O. I. *and 1 more*

3 2022MLS&T...3aLT03M 2022/03 cited: 4

Inferring dark matter substructure with astrometric lensing beyond the power spectrum

Mishra-Sharma, Siddharth

4 2021M&PS...56.1890J 2021/10

Automatic detection of impact craters on Al foils from the Stardust interstellar dust collector using convolutional neural networks

Jaeger, Logan; Butterworth, Anna L.; Gainsforth, Zack *and 11 more*

5 2021MNRAS.506.3313G 2021/09 cited: 1

Predicting bulge to total luminosity ratio of galaxies using deep learning

Grover, Harsh; Bait, Omkar; Wadadekar, Yogesh *and 1 more*

Summary of NiN (Lin et al. 2013)

1. NiN uses blocks consisting of a convolutional layer and multiple 1×1 convolutional layers.
2. NiN removes the fully-connected layers and replaces them with global average pooling (i.e., summing over all locations) after reducing the number of channels to the desired number of outputs.
3. Removing the fully-connected layers reduces overfitting. NiN has dramatically fewer parameters.
4. The NiN design influenced many subsequent CNN designs.

GoogLeNet (Szegedy et al. 2015)

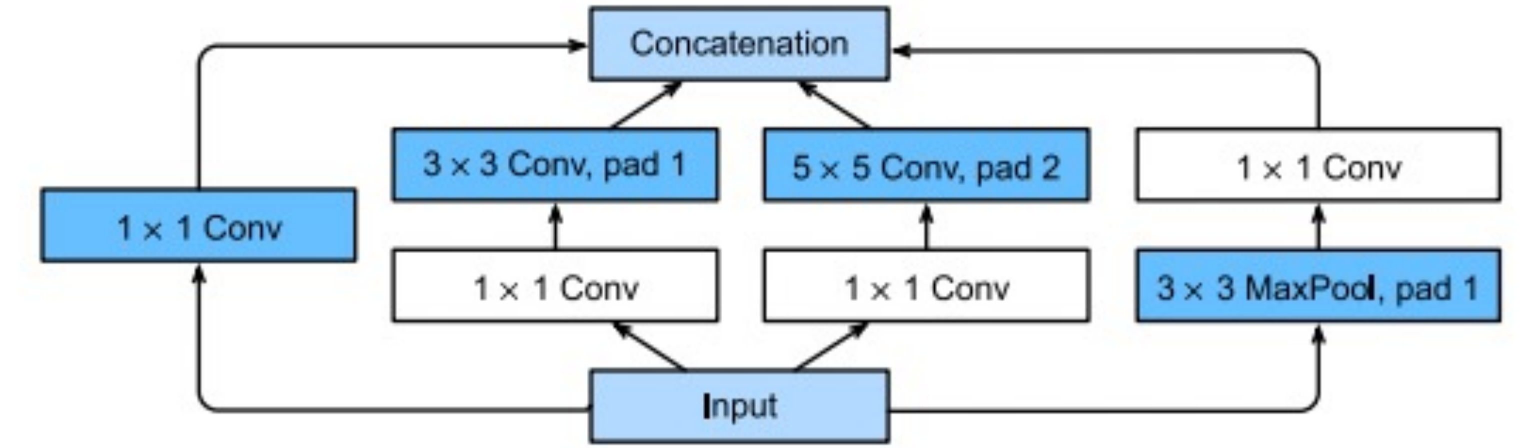


Fig. 7.4.1: Structure of the Inception block.

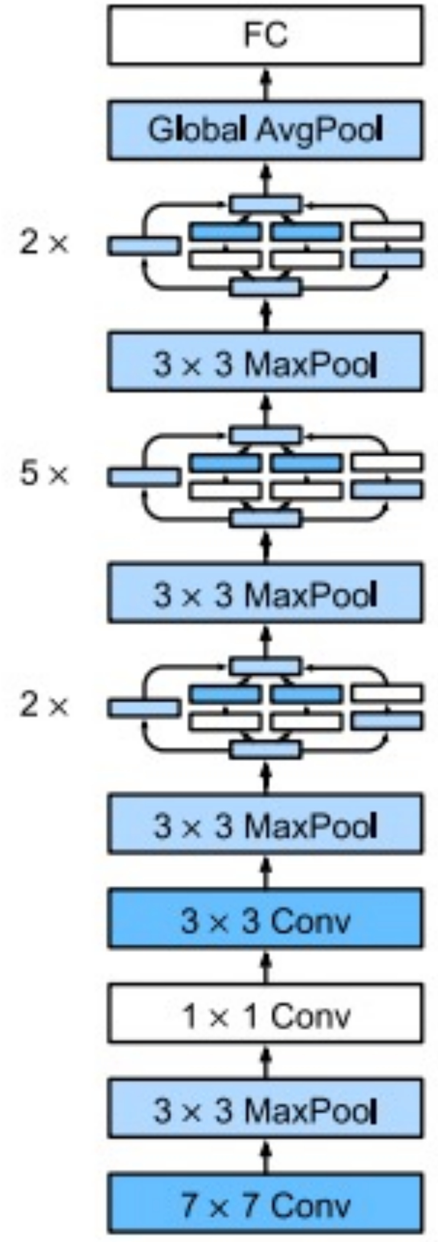


Fig. 7.4.2: The GoogLeNet architecture.

1 2022MNRAS.tmp..779Z 2022/03

Extracting Photometric Redshift from Galaxy Flux and Image Data using Neural Networks in the CSST Survey

Zhou, Xingchen; Gong, Yan; Meng, Xian-Min *and 6 more*

2 2022MNRAS.tmp..751L 2022/03

Estimating cluster masses from SDSS multi-band images with transfer learning

Lin, Sheng-Chieh; Su, Yuanyuan; Liang, Gongbo *and 3 more*

3 2022MNRAS.tmp..534W 2022/02 cited: 1

Practical galaxy morphology tools from deep supervised representation learning

Walmsley, Mike; Scaife, Anna M. M.; Lintott, Chris *and 9 more*

4 2022A&A...658A.142I 2022/02 cited: 1

Multi-scale deep learning for estimating horizontal velocity fields on the solar surface

Ishikawa, Ryohtaroh T.; Nakata, Motoki; Katsukawa, Yukio *and 2 more*

5 2021ApJ...922..232D 2021/12

Fine-grained Solar Flare Forecasting Based on the Hybrid Convolutional Neural Networks

Deng, Zheng; Wang, Feng; Deng, Hui *and 3 more*

Summary of GoogLeNet (Szegedy et al. 2015)

1. The Inception block is equivalent to a subnetwork with four paths. It extracts information in parallel through convolutional layers of different window shapes and maximum pooling layers. 1×1 convolutions reduce channel dimensionality. Maximum pooling reduces resolution.
2. GoogLeNet connects multiple well-designed Inception blocks with other layers in series.
3. GoogLeNet was one of the most efficient models on ImageNet, providing similar test accuracy with lower computational complexity.

Applications (Zhou et al. 2022)

Extracting Photometric Redshift from Galaxy Flux and Image Data using Neural Networks in the CSST Survey

Xingchen Zhou[1,2], Yan Gong[1,3★], Xian-Min Meng [1], Ye Cao [1,2], Xuelei Chen[4,2,5], Zhu Chen[6], Wei Du[6], Liping Fu[6], Zhijian Luo[6]

[1] *Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China.*

[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

[3] *Science Center for China Space Station Telescope, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China*

[4] *Key Laboratory for Computational Astrophysics, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China*

[5] *Centre for High Energy Physics, Peking University, Beijing 100871, China*

[6] *Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, China*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The accuracy of galaxy photometric redshift (photo-z) can significantly affect the analysis of weak gravitational lensing measurements, especially for future high-precision surveys. In this work, we try to extract photo-z information from both galaxy flux and image data expected to be obtained by China Space Station Telescope (CSST) using neural networks. We generate mock galaxy images based on the observational images from the Advanced Camera for Surveys of Hubble Space Telescope (HST-ACS) and COSMOS catalogs, considering the CSST instrumental effects. Galaxy flux data are then measured directly from these images by aperture photometry. The Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) are constructed to predict photo-z from fluxes and images, respectively. We also propose to use an efficient hybrid network, which combines the MLP and CNN, by employing the transfer learning techniques to investigate the improvement of the result with both flux and image data included. We find that the photo-z accuracy and outlier fraction can achieve $\sigma_{\rm NMAD} = 0.023$ and $\eta = 1.43\%$ for the MLP using flux data only, and $\sigma_{\rm NMAD} = 0.025$ and $\eta = 1.21\%$ for the CNN using image data only. The result can be further improved in high efficiency as $\sigma_{\rm NMAD} = 0.020$ and $\eta = 0.90\%$ for the hybrid transfer network. These approaches result in similar galaxy median and mean redshifts 0.8 and 0.9, respectively, for the redshift range from 0 to 4. This indicates that our networks can effectively and properly extract photo-z information from the CSST galaxy flux and image data.

Key words: cosmology – photometric redshift – large-scale structure

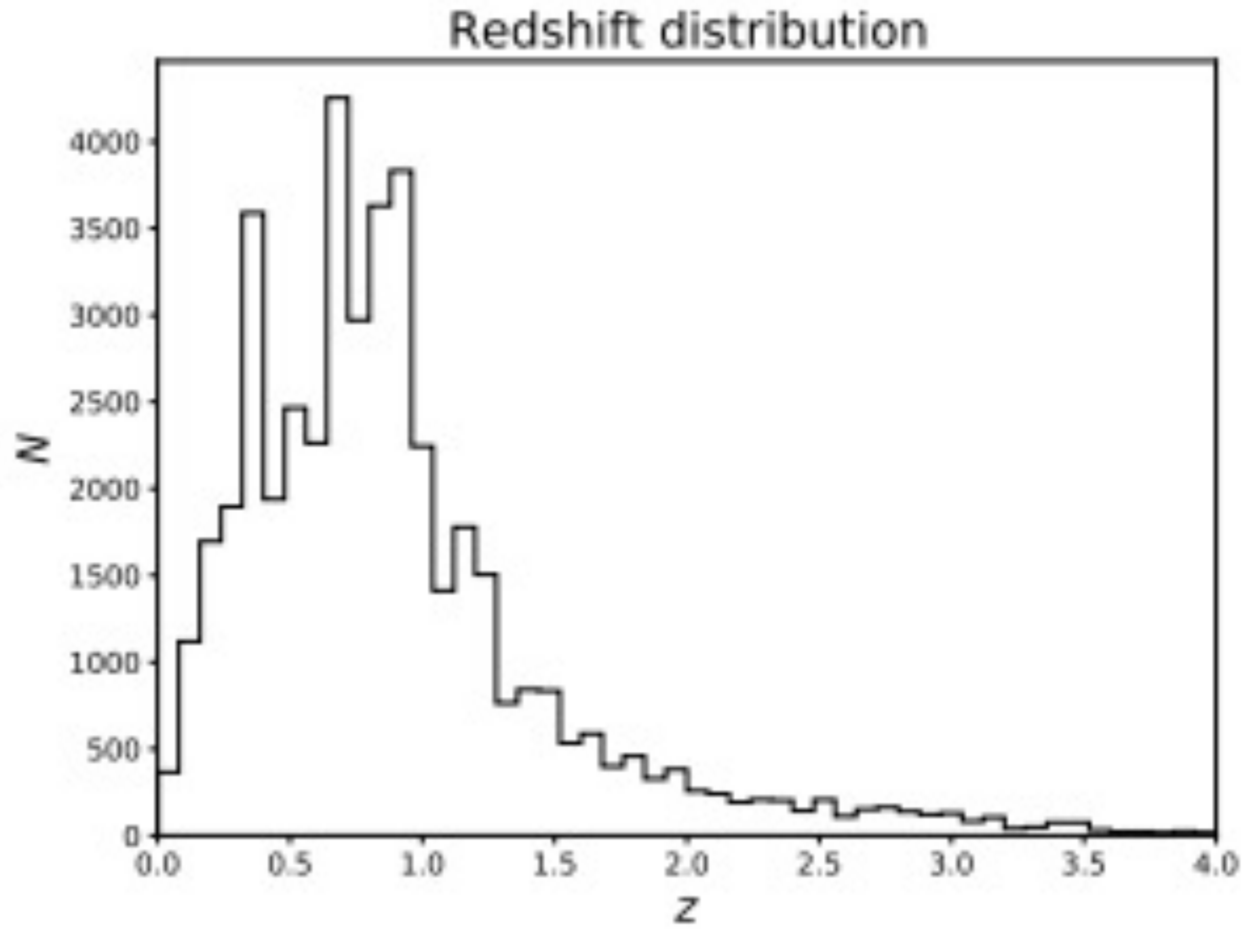


Figure 2. Galaxy redshift distribution of the sample selected from the COSMOS catalog used in the neural networks. These sources have been selected with the SNR greater than 10 in the g or i bands. The distribution has a peak around $z = 0.7$, and can extend to $z \sim 4$, which is the same as the CSST photometric galaxy sample.

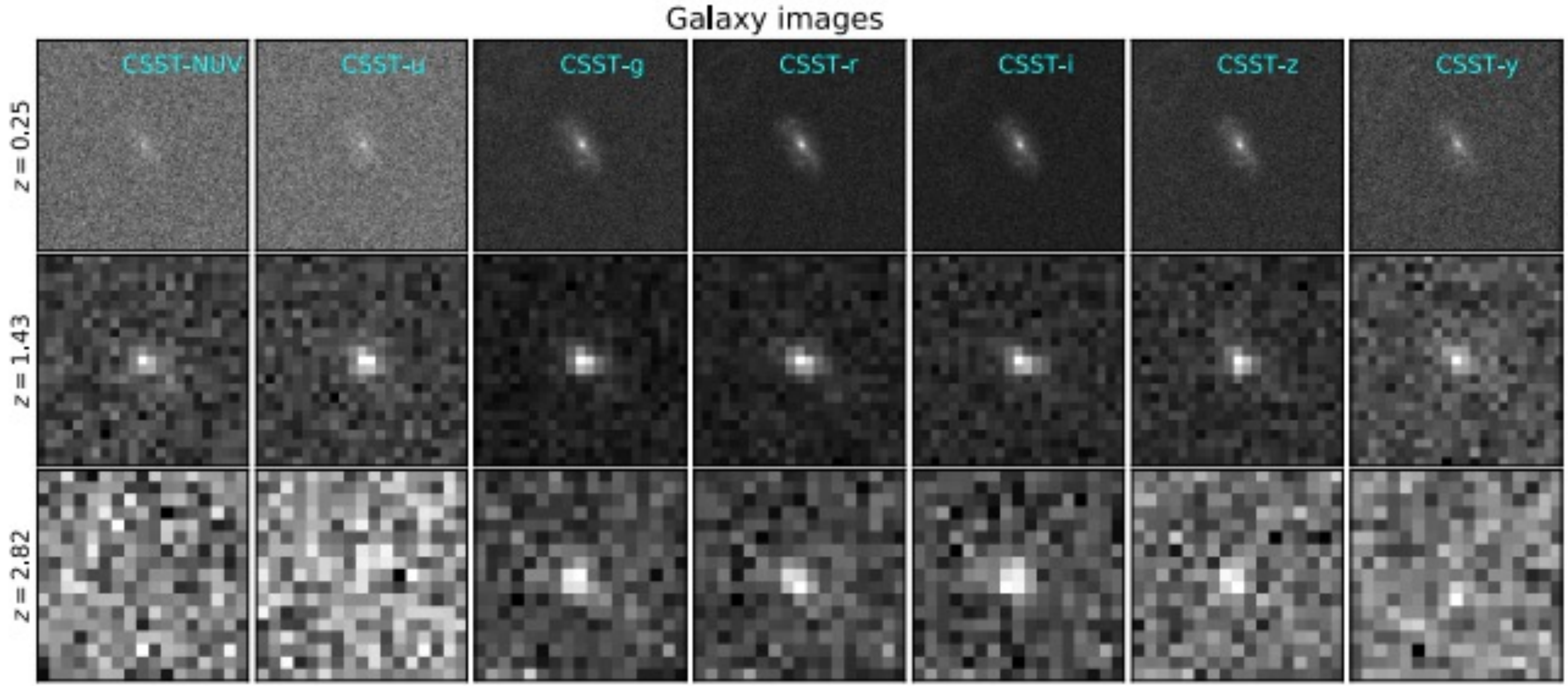


Figure 3. Examples of simulated galaxy images in the seven CSST photometric bands. we notice that noises on the NUV, u and y bands are larger than others, since these bands have lower transmissions. Many sources, especially at high redshifts, are overwhelmed by the background noises in some bands, which may indicate that neural networks are necessarily needed to extract information from these sources.

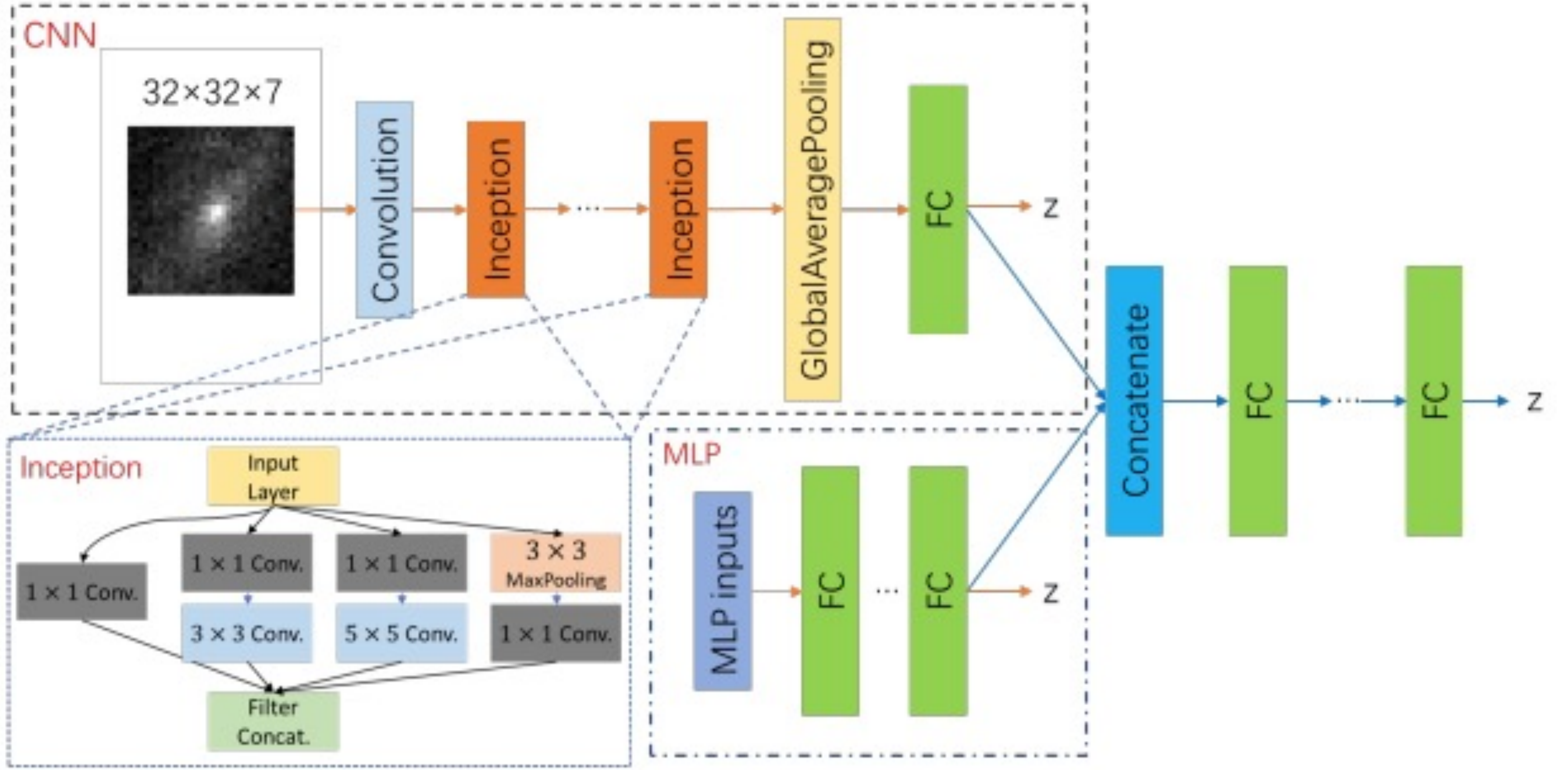
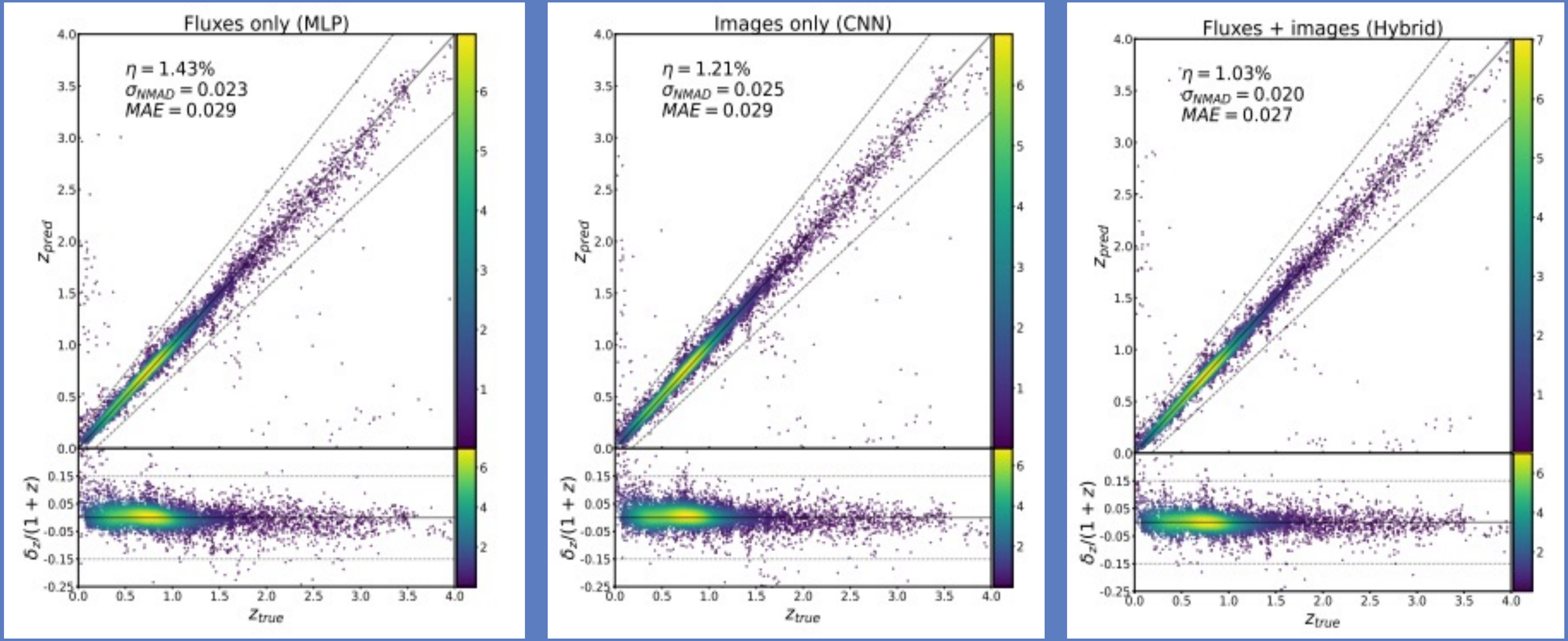


Figure 5. Architecture of the MLP, CNN and Hybrid networks. The MLP structure is shown in blue dash-dotted box. The inputs are rescaled fluxes, colors and errors, and totally 6 hidden layers are structured. The CNN structure is displayed in the dashed black box. The inputs are $(32 \times 32 \times 7)$ images, convolved and downsampled by the convolutional layer. Then three inception blocks are structured to obtain the features of size 2, which is flattened by global average pooling to a vector of size 72. Next the fully-connected layer with 40 units is applied, and then the photo-z can be obtained. The inception blocks are illustrated in the dashed blue box. We use (3×3) and (5×5) kernels to extract features parallelly, and (1×1) kernels are adopted to reduce the number of features for increasing efficiency of computation. The hybrid network are constructed by concatenating the vectors extracted by the MLP and CNN from its fully-connected layer. Totally 6 fully-connected layers with 80 units are structured, and after each layer the BatchNormalization and ReLU activation function are applied.



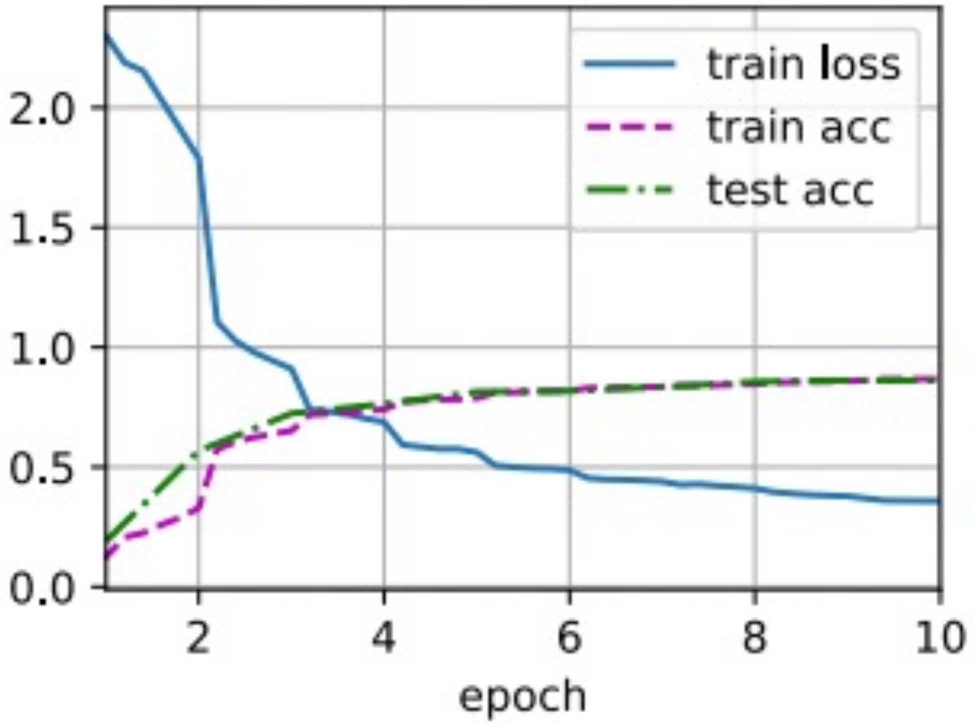
Fluxes only (MLP)
$\eta = 1.43\%$
$\sigma_{NMAD} = 0.023$
$MAE = 0.029$
Images only (CNN)
$\eta = 1.21\%$
$\sigma_{NMAD} = 0.025$
$MAE = 0.029$
Fluxes + images (Hybrid)
$\eta = 1.03\%$
$\sigma_{NMAD} = 0.020$
$MAE = 0.027$
z_{pred}
$\delta_z/(1+z)$
z_{true}

Summary (Zhou et al. 2022)

1. They try to extract photometric information from galaxy flux and image data expected to be obtained by the CSST using neural networks.
2. For their data, they generate mock galaxy images.
3. For their neural networks, they construct an MLP, a CNN, and a hybrid network of the two.
4. They find that their networks effectively and properly extract photometric redshifts from the simulated CSST data.

```
lr, num_epochs, batch_size = 0.1, 10, 128
train_iter, test_iter = d2l.load_data_fashion_mnist(batch_size, resize=224)
d2l.train_ch6(net, train_iter, test_iter, num_epochs, lr, d2l.try_gpu())
```

```
loss 0.358, train acc 0.867, test acc 0.861
3064.7 examples/sec on cuda:0
```



```
lr, num_epochs, batch_size = 0.1, 10, 128
train_iter, test_iter = d2l.load_data_fashion_mnist(batch_size, resize=96)
d2l.train_ch6(net, train_iter, test_iter, num_epochs, lr, d2l.try_gpu())
```

```
loss 0.257, train acc 0.902, test acc 0.892
3537.4 examples/sec on cuda:0
```