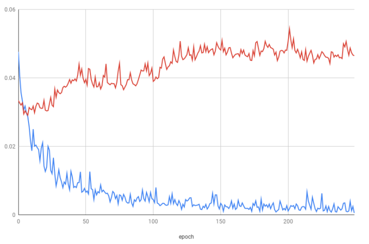
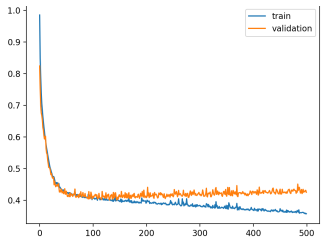
(1) Monitoring the neural network learning process

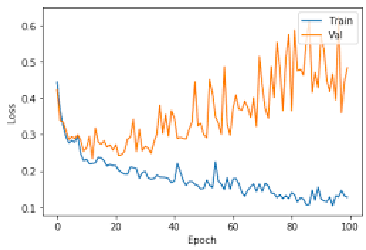
(a) You are training three neural networks and you are observing the following loss functions on the training and validation sets. What can we conclude for each of the three networks? In which training epoch would you stop training each network?



Network 1



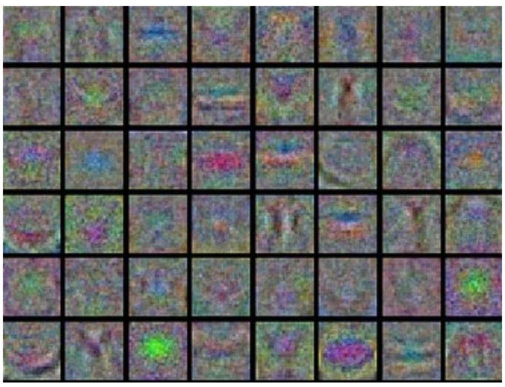
Network 2



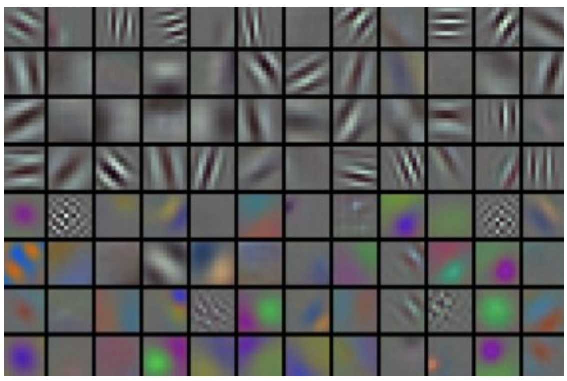
Network 3

For all networks, the training loss keeps decreasing in every epoch, indicating that the model recognizes the specific training samples for the task of interest.
For Network 1, the validation loss decreases with an increasing number of epochs, indicating that the model does not generalize well enough on the validation set (and most probably it will not generalize well on the testing set).
For Network 2, the validation loss decreases up to 150 epochs, while remaining stable or increasing beyond that. This indicates that we would stop training around 150 epochs and utilize that model for testing.
For Network 3, the validation loss decreases up to 10 epochs, while increasing beyond that. Although we can use the network trained with 10 epochs for the test data, we might need to be careful with this, especially if we think that 10 epochs are not enough for our task.

(b) You have trained two neural networks and visualized various kernels learned in the first layer. Which of the two networks would you use and why?



Network 1



Network 2

Network 1 appears to have only learned noise. Network 2 appears to have learned weights that can be interpreted in terms of 2D operations in images (e.g., detection of horizontal, diagonal, vertical edges), therefore we would use Network 2.

(2) Convolution operations

(a) Consider a 1-dimensional signal

$$F = [1 \quad 2 \quad 1 \quad 3 \quad 2 \quad 3 \quad 1 \quad 2 \quad 3 \quad 8 \quad 7 \quad 8 \quad 9 \quad 9 \quad 7 \quad 8]$$

We will define the 1-dimensional convolution of the signal F with a filter F at point i as:

$$(F \star W)[i] = \sum_{-\infty}^{\infty} F[j]W[j]$$

(a.i) Calculate the convolution of F with filter $W_1 = [1 \quad 1 \quad 1]$, i.e., $F1 = F \star W_1$. What is the operation that filter W_1 performs? *Note:* You can perform zero-padding for the edge elements.

$F_1 = F \star W_1 = [3, 4, 6, 6, 8, 6, 6, 6, 13, 18, 23, 24, 26, 25, 24, 15]$

(a.ii) Calculate the convolution of F with filter $W_2 = [1 \quad 0 \quad -1]$, i.e., $F2 = F \star W_2$. What is the operation that filter W_2 performs? *Note:* You can perform zero-padding for the edge elements.

$F_2 = F \star W_2 = [-2, 0, -1, -1, 0, 1, 1, -2, -6, -4, 0, -2, -1, -2, -1, 7]$

(b) Consider an image $I \in \mathbb{R}^{D \times D}$ and the following 2-d convolution filters.
What operations will the following filters apply to image I?

$$F_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, F_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, F_3 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

F_1 performs smoothing, or image blurring, since it replaces each element with the sum of its local values.

F_2 amplifies the horizontal edges.

F_3 sharpens the image, since it multiplies the value of a given pixel by 9 and then subtracts all the neighbors. This will result in higher values for the pixels that already have a big difference from their neighbors.

(c.i) A gray color 10×10 image is the input to a CNN. The first layer of the CNN has 64 filters of dimensionality 3×3, stride size of 1, and zero-padding. What is the output of the first hidden layer and how many parameters are learned by this layer?

We apply zero-padding and a stride of 1, therefore the output from each convolution is 10×10.
We have 64 filters, therefore the final output is $10 \times 10 \times 64$.
Each filter has $3 \times 3 + 1 = 10$ parameters (i.e., the parameters of the convolution and the bias).
We have 64 filters, therefore $64 \times 10 = 640$ parameters.

(c.ii) A color 10×10 image is the input to a CNN. The first layer of the CNN has 64 filters of dimensionality 3×3, stride size of 1, and zero-padding. What is the output of the first hidden layer and how many parameters are learned by this layer?

We apply zero-padding and a stride of 1, therefore the output from each convolution is 10×10.
We have 64 filters, therefore the final output is $10 \times 10 \times 64$.

Each filter has $3 \times 3 \times 3 + 1 = 28$ parameters (i.e., the parameters of the convolution and the bias). We have 64 filters, therefore $64 \times 28 = 1792$ parameters.

(3) Neural network hyper-parameters

(a) What are important factors that contribute to selecting the depth of a neural network?

(i) Type of neural network (eg. FNN, CNN etc)
(ii) Input data
(iii) Computation power, i.e. Hardware capabilities and software capabilities
(iv) Learning Rate

All of the above.

(b) You are fine-tuning a pre-trained neural network. What would be the best strategy for the learning rate?

(i) Keep the same learning rate as the one in the pre-trained network.
(ii) Have a larger learning rate compared to the one in the pre-trained network.
(iii) Have a smaller learning rate compared to the one in the pre-trained network.
(iv) The learning rate will not affect the fine-tuning result.

The correct answer is (iii).