H2
GROUP 51
João Moniz nº 83480
Liana Baghdasaryan nº 104708

Work Distribution
Liana Baghdasaryan: q1.1 (a,b,c), q1.2 and q2
João Moniz: q1.1(b) and q3
in Question 1 question b was made by both members, João showed that M exists that z’ = Mx’, and Liana did the general expression for element (i,j).

Question 1:

Question 1

Mo Tu We Th Fr Sa Su

Meme No.
Date / /

① ⓐ $(H-M+1)\times(W-N+1)$

ⓑ $M \in R^{H'W' \times HW}$

$H' = H-M+1$

$W' = W-N+1$

$z' = M' \cdot x'$

let's assume there is a matrix $M \in R^{H'W' \times HW}$ such that $z' = M \cdot x'$
if we assume that the elements of the matrix M are equal in each row, then each jth element of i^{th} row will be equal to $M_{ij} = \frac{z'_i}{sum(x')}$

ⓒ convolutional_network

$M \times N$

fully connected network

$$\underbrace{(H\times W\times(H-M+1)\times(W-N+1))}_{\text{first one}} + \underbrace{((H-M+1)\times(W-N+1)\times([\tfrac{H-M+1}{2}]\times[\tfrac{W-N+1}{2}]))}_{\text{second one}}$$

② $Q = x' \cdot W_Q = x'$

$K = x' \cdot W_K = x'$

$V = x' \cdot W_V = x'$

$$Att\_prob = softmax\left(\frac{Q\cdot K^T}{\sqrt{M\cdot N}}\right)$$

$$Att\_out = softmax\left(\frac{Q\cdot K^T}{\sqrt{M\cdot N}}\right)\cdot V$$

CamScanner

Q 1.1) b) continuation

collums

$$z' = M x' \Leftrightarrow \begin{bmatrix} z_1 \\ \vdots \\ z_2 \\ \vdots \end{bmatrix} = \begin{array}{c} a \\ \\ \\ \\ b \\ \\ \\ c \end{array} \begin{bmatrix} [W_1^T|0] & [W_2^T|0] & \cdots & [W_n^T|0] & 0\,0 \cdots 0 \\ 0 & [W_1^T|0] & \cdots & [W_n^T|0] & 0\,0 \cdots 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 \cdots 0 & [W_1^T|0] & \cdots & \cdots & [W_n^T|0] \\ [0\,W_1^T|0] & [0\,W_2^T|0] & \cdots & [0\,W_n^T|0] & 0 \cdots 0 \\ 0 & [0\,W_1^T|0] & \cdots & [0\,W_n^T|0] & 0 \cdots 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 \cdots 0 & [0|W_1^T] & [0|W_2^T] & \cdots & [0|W_n^T] \end{bmatrix} \begin{bmatrix} x_1 \\ | \\ x_2 \\ | \\ \cdots \end{bmatrix}$$

~~Where $[W_1^T|0]$ is~~

Where in the first line $[W_1^T|0]$ is a padded version of W_1^T with zeros to only apply to x_1 "collum" of x'. in line 2 we "moved" the filter one to the right so that $[W_1^T|0]$ now applies to x_2. We do this for every collum of the first line of the image, in "step" a. then we move the filter one line bellow, $[0\ W_1^T|0]$, in "step b" and repeat the process for the second line of the image. We do this for all lines of the image to get $\begin{bmatrix} z_1 \\ \vdots \\ z_2 \\ \vdots \end{bmatrix}$.

Question 2: 1) 2) 3)

Question 2

Memo No.
Date / /

a) A convolutional neural network has fewer parameters than a fully-connected network with the same input size and the same number of output classes because it uses shared weights and biases in the convolutional layers In a fully-connected network, each neuron in a layer is connected to every neuron in the previous layer, which requires a weight for each connection. This means that the number of weights in the convolutional layer is much smaller than in fully connected layers

b) Yes.

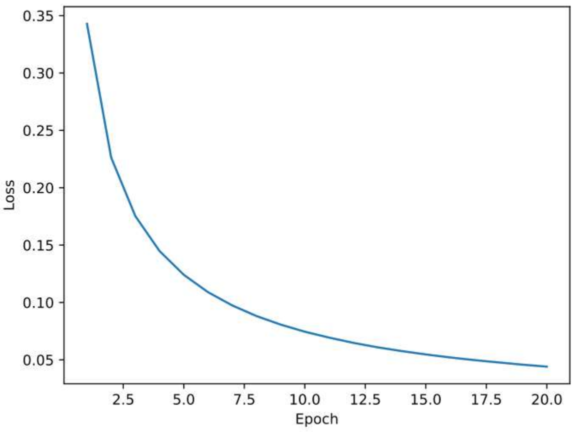
CNN's are able to learn features that are more spatially local, which is important for image classification tasks. For example, in an image of a digit, the curves and lines that make up the digit are often located in a particular part of the image, and a CNN can learn to recognize these features by looking at a small region of the input rather than the entire input.

c) In general, CNN's are expected to achieve better generalization than a fully connected network when the input is from a source with some spatial structure, such as an image. If the input is from a source with no spatial structure, such as a set of independent sensors, a CNN may not be a best choice. In this case we can use fully connected layers.

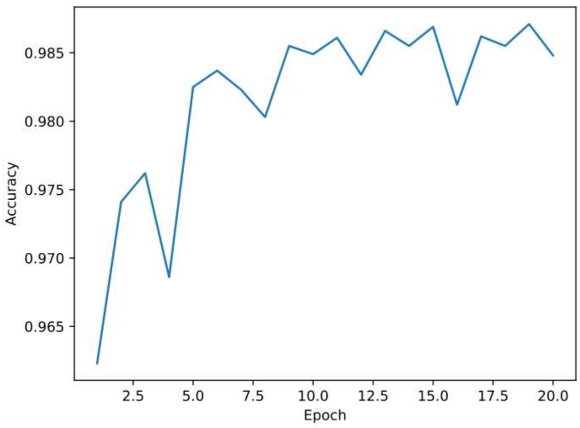
Question 2:

4) Best Learning Rate: 0.0005

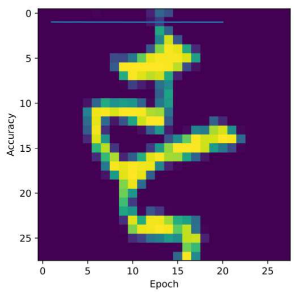
Training loss:



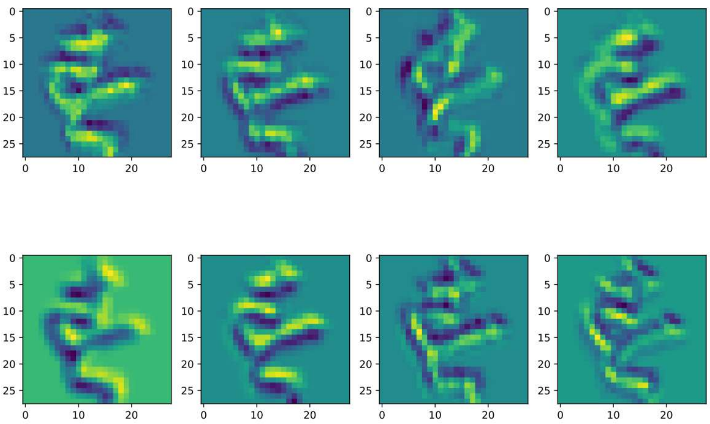
Validation accuracy:



5) Original:

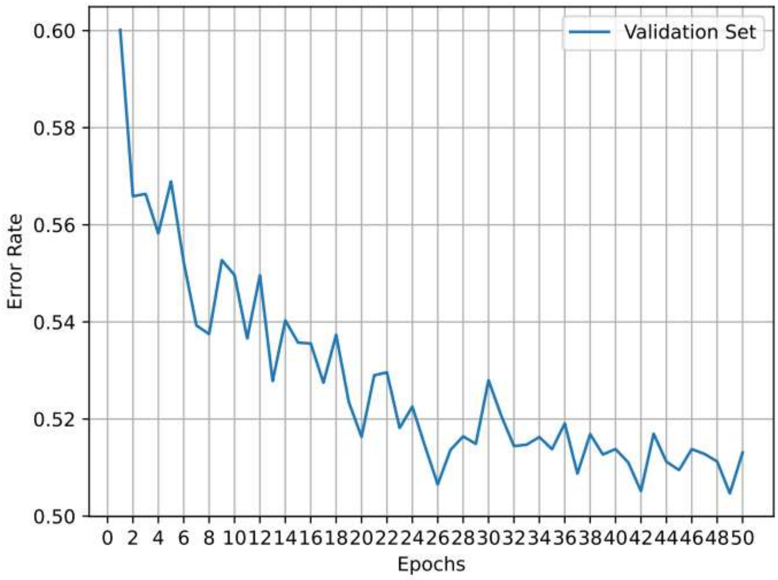


Activation maps:



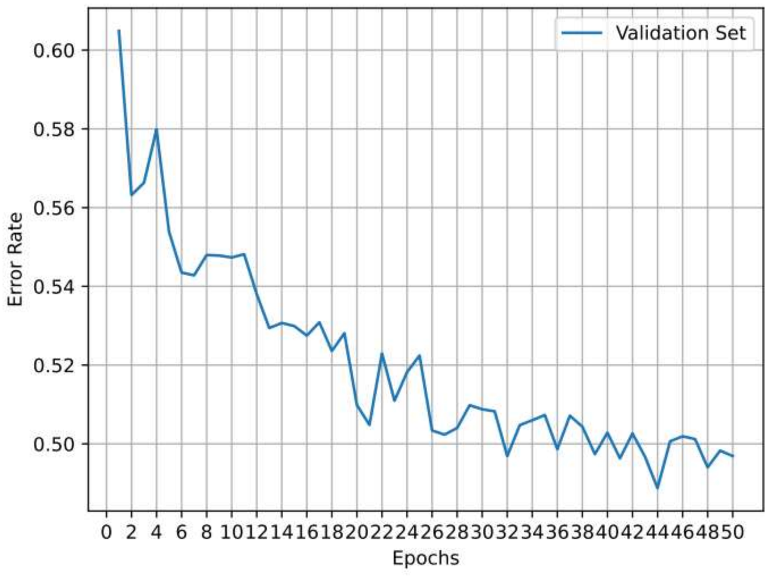
Analysis: The activation maps seem to be highlighting different borders between parts of the image, for example the second one from the top starting from the right is highlighting the upper borders of the figure.

Question 3:
a)



Final validation error rate: 0.5131
Test error rate: 0.5178

b)



Final validation error rate: 0.4969
Test error rate: 0.5046

c)It is possible to improve the results by adding beam search, by tracking K of the most probable partial translations, and not just the best one, helps preventing the model from getting stuck with one bad decision and sticking with it for the rest of the translation.