COMP4211_20552827_PA2

Stud no.: 20552827

Name: Jeong Jin Hong

Q1. How does the ELU function differ from the ReLU function?

In contrast to ReLUs, ELUs have negative values which allows them to push mean unit activations closer to zero like batch normalization but with lower computational complexity.

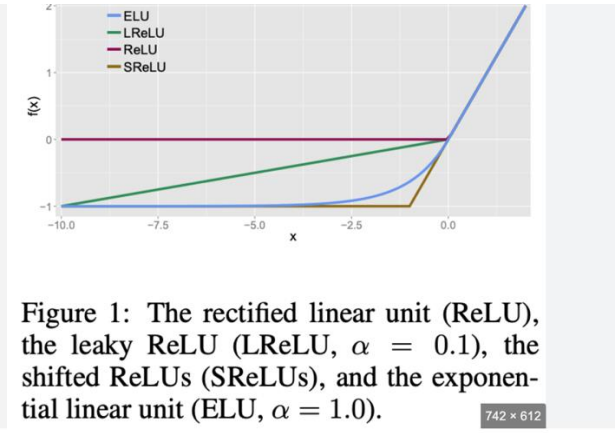


Figure 1: The rectified linear unit (ReLU), the leaky ReLU (LReLU, $\alpha = 0.1$), the shifted ReLUs (SReLUs), and the exponential linear unit (ELU, $\alpha = 1.0$).

Q2.

'valid' means no padding. same" results in padding with zeros evenly to the left/right or up/down of the input. When padding="same" and strides=1, the output has the same size as the input.

Given that padding = 'same', strides = 1, kernel size = 5, filters = $N^{filters}$, the B,H,W will remain the same but C will change to 'filter' size which was 64. Hence from input tensor shape of (32,32,32,1) will change to (32,32,32,64).

Q3.

Valid padding means that we only apply a convolutional filter to valid pixels of the input. Since only the pixels of the original image are valid, valid padding is equivalent to no padding at all.

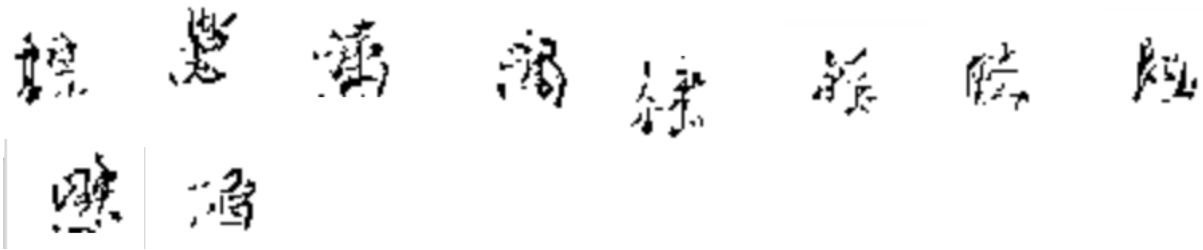
Q4

Trainable parameter is 3,571,925.

Q5.

the training process of a neural network with residual connections is empirically shown to converge much more easily than that of feedforward neural network which could cause vanishing gradient problem. Residual connections avoid these problems by having shallow networks in the "ensembles".

Q6.



Q7.
Binary crossentropy loss value was `0.1851`

‘Q8.
We set from_logit true when predicted value is logit and from_logit is false when predicted value is probabilities ranging from [0, 1]