EEEB4023/ECEB463
Artificial Intelligence and Neural-fuzzy Systems

Dr Mohd Zafri Bin Baharuddin, Dr.

Test 2, Sem 1 2021/2022

27 Dec 2021

100 marks in this Test contribute to 30% of the final grade. Instructions:

1. This is an **open book** test.
2. Do not communicate, share, or receive any help in this test.
3. Prepare your own pieces of paper, A4 or foolscap.
4. Write your full name and ID number on the first page.
5. **You have 150 MINUTES.**
6. **There are SEVEN (7) QUESTIONS. Answer all questions.**
7. Convert your answers to PDF using Adobe Scan, CamScanner or similar app or scanner.
8. Submit your PDF file to the Teams assignment page.

QUESTION 1 [MULTIPLE CHOICE QUESTIONS, 10 MARKS]

Answer on your own piece of paper. Write down your answers as the following format:

1.1) A

1.2) B and C

⋮

1.8) D

1.1 Which of the following are valid activation functions (non-linearities) for a neural network? There can be multiple answers. **[1 mark]**

A. $f(x) = \max(0.3x, 0.7x)$

B. $f(x) = \min(0, x)$

C. $f(x) = 0.8x$

D. $f(x) = \begin{cases} 1, & \text{if } x > 0.5 \\ -1, & \text{otherwise} \end{cases}$

1.2 Calculating gradients is important in the backpropagation step of model optimization. Among the following activation functions, which one can lead to vanishing gradients?

[1 mark]

A. Tanh
B. LeakyReLU
C. ReLU
D. None of the above

1.3 Object detection models may detect the same object multiple times and cause bounding boxes to overlap. Which one of the following is a method can solve this problem?

[1 mark]

A. Region-based convolutional neural network
B. Intersection over union
C. Mean average precision
D. Non-max suppression

1.4 In object detection neural network architectures, the backbone network is typically a pre-trained network that functions as a feature extractor. Which of the following is **NOT** a backbone network:

[1 mark]

A. VGG
B. ResNet
C. MobileNet
D. ImageNet

1.5 Which of the following sentences are **TRUE** about a convolutional layer in a CNN? There can be more than one answer.

[2 marks]

A. The number of learnable parameters depends on the input volume and number of filters.
B. Stacking multiple convolutional layers will result in a non-linearity.
C. Padding maintains the output size to be the same as the input.
D. Stride maintains the output size to be the same as the input.

1.6 If you have a sequence of data, such as video as input, and you want to output a sequence of text, which of the following models is suitable? There can be more than one answer.

[2 marks]

A. Convolutional network
B. Generative adversarial networks
C. Region based convolutional neural network
D. Long short term memory

1.7 After training an image recognition model, you notice a large difference between the training accuracy (100%) and the test accuracy (40%). Which of the following methods can be used to reduce this gap?

[1 mark]

A. RMSprop
B. Dropout
C. Generative Adversarial Networks
D. Sigmoid activation

1.8 Which of the following best describes transfer learning?

[1 mark]

A. Use a model that was implemented on another application
B. Use weights that were trained on another dataset
C. Use a combination of models to obtain the highest accuracy
D. Augmenting a dataset to increase data variation

QUESTION 2 [17 MARKS]

The computational graph below describes the function:

$$f(x,w) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

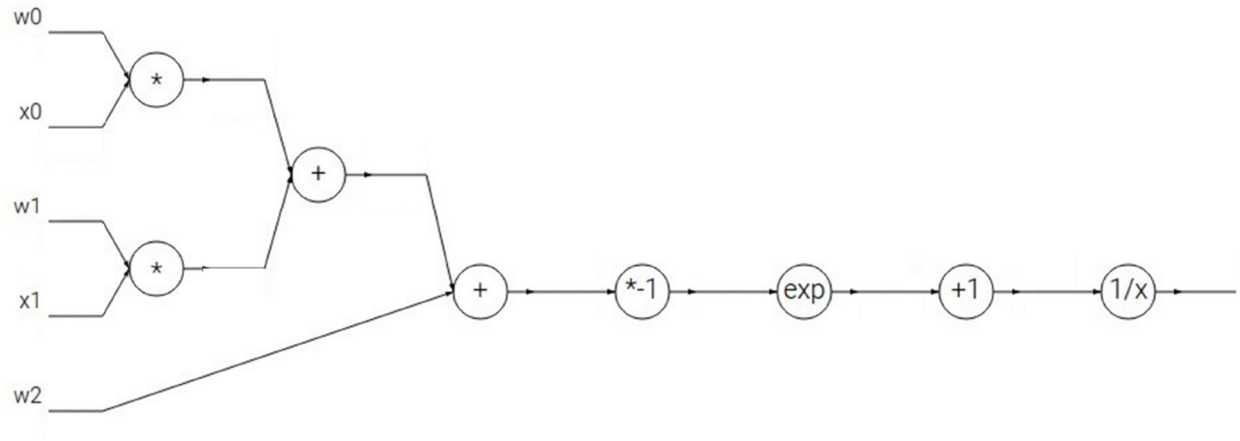


Figure 1: Computational graph for Question 2

The weights, *W*, and input, *x*, to the computational graph are:

$$W = \begin{bmatrix} w0 \\ w1 \\ w2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad x = \begin{bmatrix} x0 \\ x1 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

(a) Perform the forward pass calculations to find the score at the output node.

[5 marks]

(b) Compute the gradients for all nodes using back-propagation. You can apply your knowledge of patterns in gradient flow to reduce your calculations. The gradient of the output with respect to the loss is always 1.00.

[12 marks]

QUESTION 3 [18 MARKS]

A LeNet network was modified to accept coloured images and has a layers parameters table as shown below. Calculate the 10 values marked as (i) – (x) in the table for the conv1 and pool1 layer output channel, output width/height, memory, trainable parameters and flops. To obtain maximum marks, answers must include workings or brief explanations.

[18 marks]

	Input size		**Layer**				**Output size**				
Layer	*C*	*H/W*	*filters*	*kernel*	*stride*	*pad*	*C*	*H'/W'*	*memory (KB)*	*params (k)*	*flop (M)*
conv1	3	32	14	5	1	0	**(i) [2 marks]**	**(ii) [2 marks]**	**(iii) [2 marks]**	**(iv) [2 marks]**	**(v) [2 marks]**
pool1	14	28		4	2	0	**(vi) [2 marks]**	**(vii) [2 marks]**	**(viii) [2 marks]**	**(ix) [1 mark]**	**(x) [1 mark]**
conv2	14	13	16	5	1	0	16	9	5.0625	5.6	0.45
pool2	16	9		6	1	0	16	4	1	0	0
flatten	16	4					256		1	0	0
fc1	256		120				120		0.4688	30.72	0.03
fc2	120		84				84		0.3281	10.08	0.01
fc3	84		10				10		0.0391	0.84	0.00
Total									**60.02**	**48.29**	**1.32**

QUESTION 4 [15 MARKS]

You want to create a classification application using a convolutional neural network to sort different qualities of cucumbers at a farm in Cameron Highlands. For the dataset, you have collected thousands of pictures annotated with labels of high quality, medium quality, low quality and damaged cucumbers.

(a) Discuss TWO (2) advantages of using convolutional layers over fully connected layers.

[4 marks]

(b) Explain the importance of using non-linearities between the layers of neural networks?

[2 marks]

(c) As you train your model, you realize that you do not have enough data to classify for damaged or low quality cucumbers. Suggest THREE (3) data augmentation techniques that can be used to increase the amount of training data.

[3 marks]

(d) Briefly explain why dividing a dataset into train, validation and test sets can assist to identify overfitting?

[3 marks]

(e) Your training and validation accuracies are as in the graph below. What can you do to improve the results?

[3 marks]

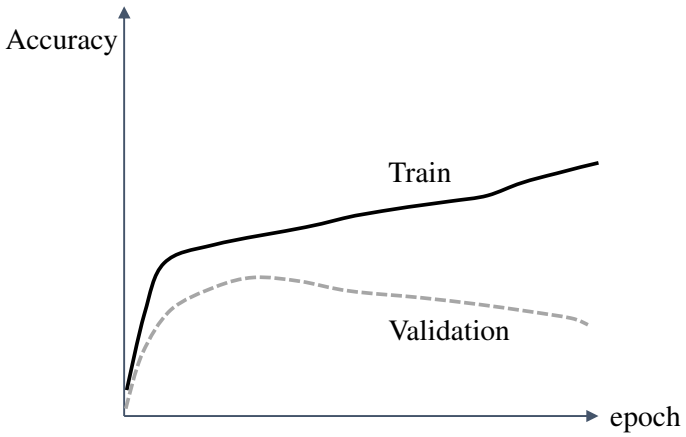


Figure 2: Train/validation curves for Question 4(e)

QUESTION 5 [10 MARKS]

On a manufacturing production line producing children's toys, an inspection station is to classify different types of toys. An image classifier needs to be trained to perform the sorting. The following two datasets are available:

D1: 100 labelled samples

D2: 1 million labelled samples

Provide brief answers for the following questions.

(a) If dataset D1 is trained from scratch, state a problem that will occur.

[1 marks]

(b) Dataset D2 is trained from scratch, and then transfer learning is used to train on D2. Explain how this method solves the problem in part (a).

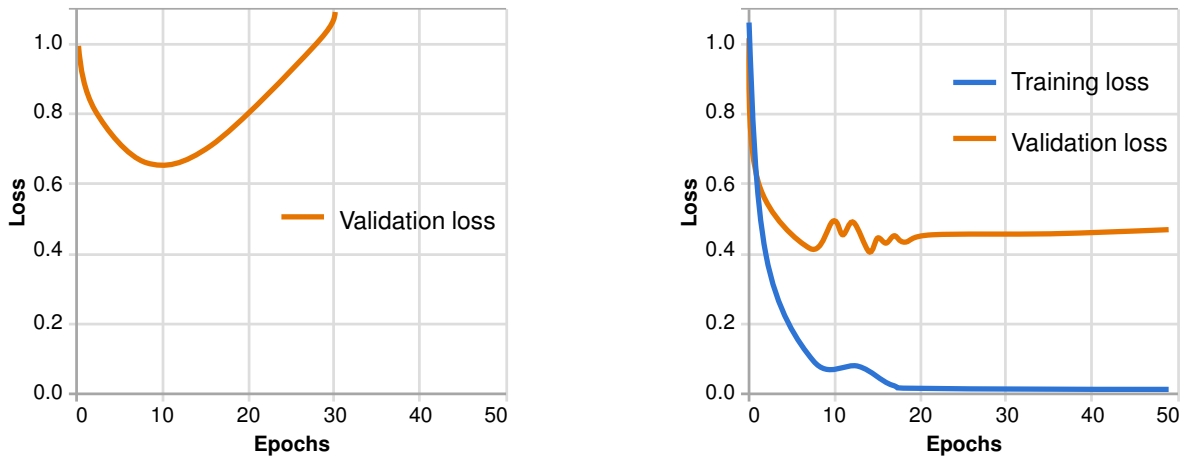
[3 marks]

(c) The model was trained and its training loss graph is shown in Figure 3(a). Explain this graph shape, then recommend a hyperparameter needs to be adjusted, and how must it be changed to improve the results?

[3 marks]

(d) After making corrections, the model was trained again. The training and validation loss graph is shown in Figure 3(b). What hyperparameter needs to be adjusted, and how must it be changed to improve the results?

[3 marks]



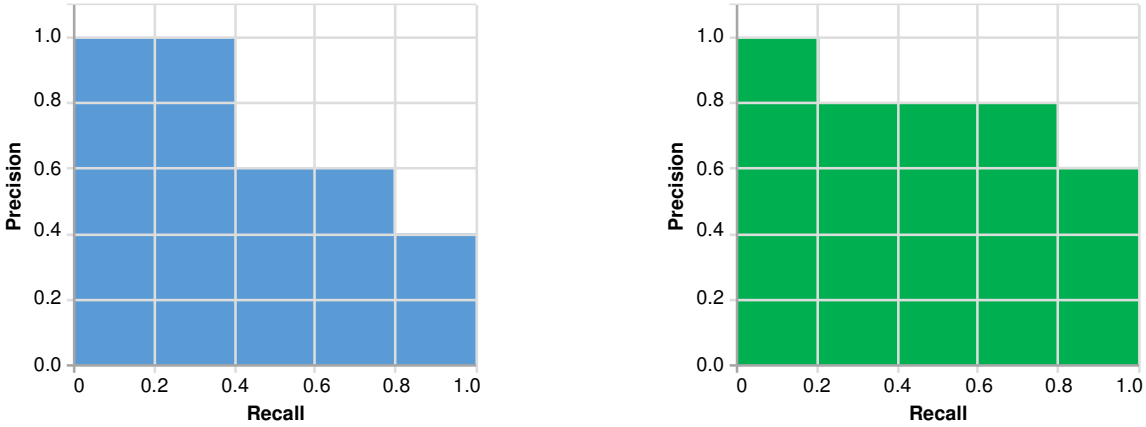
(a) (b)

Figure 3: Loss curves for Question 5

QUESTION 6 [15 MARKS]

(a) Describe at least one similarity and one difference between object classification and object detection. **[4 marks]**

(b) An object detector trained with two classes was run on a set of test images. The precision versus recall graph for each class with an IoU > 0.5 is shown in the figure below.



(a) Class 0: seal (b) Class 1: whale

Figure 4: (a) Precision-recall graphs for Question 6(b)

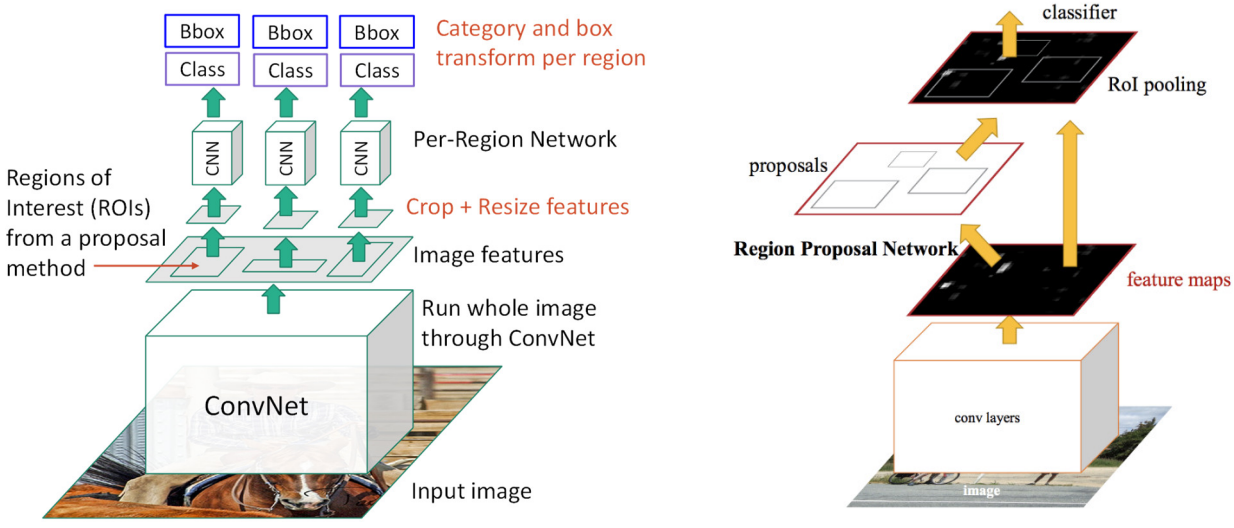
(i) Calculate the average precision (AP) of each class.

[4 marks]

(ii) Determine the mean average precision (mAP) for the object detector.

[2 marks]

(c) Briefly explain how Faster RCNN improves over RCNN and Fast RCNN. **[5 marks]**



(a) (b)

Figure 5: (a) Fast RCNN, and (b) Faster RCNN

Girshick, "Fast R-CNN", ICCV 2015. Figure copyright Ross Girshick, 2015;
Ren et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", NIPS 2015

QUESTION 7 [15 MARKS]

You are the Chief AI Engineer for an autonomous vehicle company. You want to use an object detection model for the self-driving vehicle to detect road signs, pedestrians and other vehicles.

(a) Explain your plan according to the data driven machine learning approach, which consists of three main steps: (1) dataset preparation, (2) training/optimization of the object detection model, and (3) model testing/implementation. Describe each step in detail. Recommend a hardware for training and another for inferencing (if different).

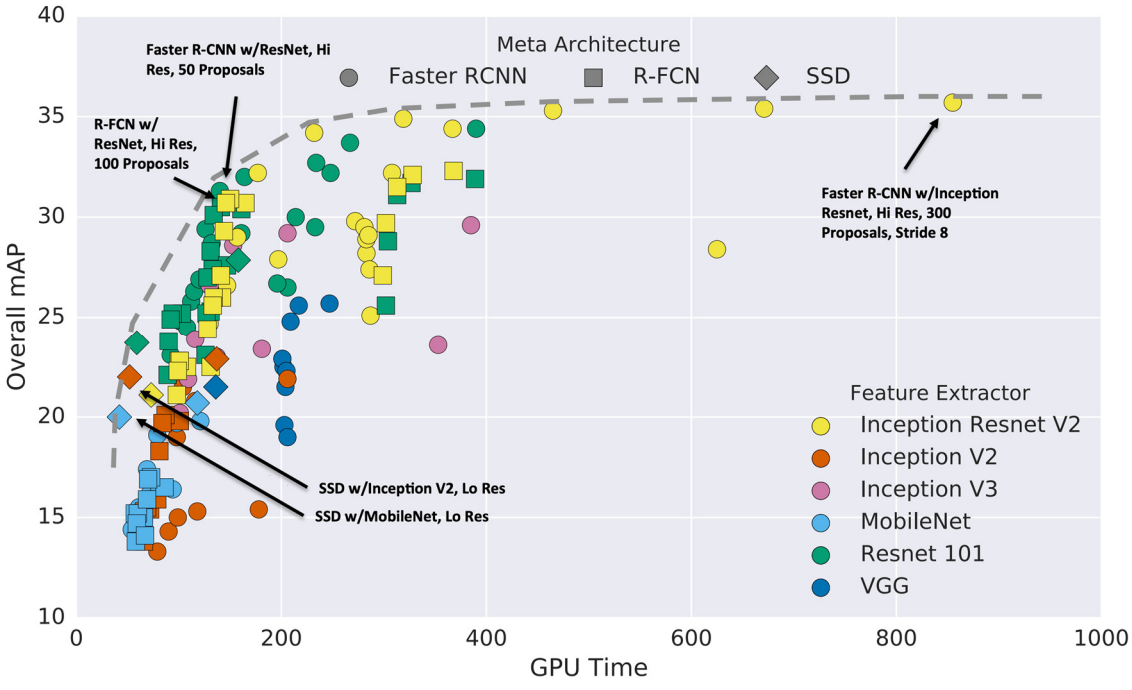
[9 marks]

(b) For this specific application, discuss which is more important: inference accuracy or speed?

[3 marks]

(c) Recommend an object detection model by selecting an architecture as well as a backbone network. You may refer to the graph below, or recommend other models. Justify and explain your selection.

[3 marks]



-END OF QUESTION PAPER-