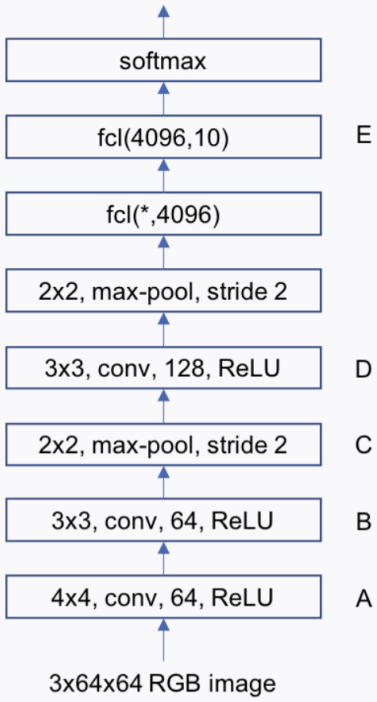
Artificial Intelligence Exam

Q1 Convolutional neural networks
11 Points

Consider the following convolutional neural network (CNN) classifier, where the layer labelled A involves 4x4 convolutions into 64 feature maps, followed by ReLU; and the layer labelled E is fully-connected from a 4096-vector on input to a 10-vector on output.



Q1.1
5 Points

How many parameters (weights + bias values) are there in each of the layers A, B, C, E? Set your answers out below, in each case showing the computation (e.g. '4x10+8'). You do not need to work out the final total for each layer.

A

(4 * 4) * 3 * 64 + 64

B

(3 * 3) * 64 * 64 + 64

C

0

E

4096 * 10 + 10

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q1.2
6 Points

What is the receptive field size in the output from each of the layers A, B, C, D, E? Set your answers out below, in each case in the form 'nxn' (e.g. '2x2').

A

4 * 4

B

6 * 6

C

7 * 7

D

9 * 9

E

64 * 64

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q2 Backpropagation

3 Points

In using a pre-trained CNN as a fixed part of a larger deep neural network, what are the implications for backpropagation in training the larger network?

Using backpropagation to adjust the pre-trained CNN's output or input weights and bias. We may regard the trained CNN as a single node with outputs. And use its output as input to the larger network.
It can make training better. Backpropagation will update each weight and bias expect the pre-trained CNN. So the backpropagation will focus more on the rest part of the network.

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q3 Word embedding

4 Points

The word2vec skip-gram method is used to learn a word-embedding from a text corpus (e.g. social media posts). Briefly explain the motivation behind the hidden layer having lower dimensionality than the input and output layers.

As for lower dimensionality than the input. The main reason is to do the dimensionality reduction process. For example, we have a 2 * 6 matrix, then multiply a 6 * 3 matrix, then we got a 2 * 3 matrix. So embedding reduces the complexity of the calculation and saves memory space.
As for lower than the outputs. The feature that the low dimensionality include may be very general, cannot find the useful word. After the dimensionality increased, some of the features may be highlighted or the general normal feature may be excluded. Also can find the similarity or some kind of relationship between words.
Words that occur in similar contexts tend to have similar meanings. And words with similar contexts are represented by similar vectors as well. The word meanings are to be embedded in the vector space.

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q4 Recurrent neural networks

5 Points

A Recurrent Neural Network (RNN) can be trained from a text corpus to generate novel character sequences.

Q4.1

1 Point

Briefly explain the way in which the input and output character sequences used in training are related.

Train on chunks of text from a corpus. The output is the same as the input, advanced by one time-step.

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q4.2
4 Points

How is the trained network used to generate a random text string?

First pre-process the data, like establish a connection between each word and a unique number. Then each character is continuously inputted into the network, and compute the output (o). Sample from the probability distribution softmax (o). Then set the next input to be this value.
Repeat the output is used as the next input, advanced by one time-step. Because it cannot loop forever, set the stop criteria can be the length of the text string or something else.

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q5 LSTMs
4 Points

In a Long Short-Term Memory (LSTM), the logistic function is used several times in connection with a 'gate'. Explain the purpose of this in relation to the operation of the gate.

Gate control flow of data, from 0 (close), to 1 (open). It controls several functions. For example, deciding what to forget and what to keep from the cell state. Deciding what new data to store in the cell state and updating the cell state. Deciding what to output from the cell state.

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q6 Motion-based image segmentation
4 Points

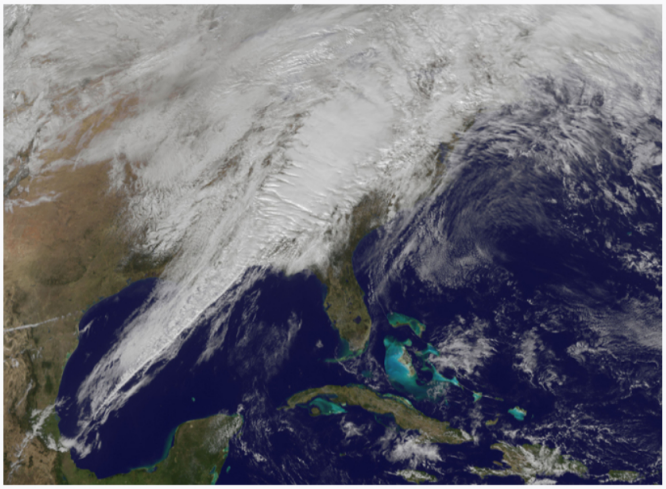
Motion-based image segmentation can be achieved by comparing the current intensity at each pixel location with a simple statistical model of the past intensity values at that location. Explain briefly how continuous variation of the background intensity is accommodated within this approach.

We need to update the background. Because of the background continuous variation, it may become out of date. We can update the background model at each time-step. Calculate and migrate the new mean for the pixel value of each location for each time period. Also, update the new standard deviation every time. Then calculate the outlier in each frame.

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q7 System design
20 Points

There is an increasing trend towards combining different kinds of data in applications of AI. You are required to outline an initial design for a system, based on a deep neural network, for estimating the probability of rainfall at a specific location, 6 hours into the future. The system will be given two kinds of input for the current time: (1) a single satellite image (such as that below) showing cloud cover over the wider region surrounding the location, and (2) the temperature and wind direction at ten ground-stations distributed over the same region.



Q7.1
12 Points

Describe the architecture of your system, including as appropriate: the precise representation of inputs and output, any component neural networks used, types and number of layers, numbers of feature maps, use of activation functions, and resizing operations. You are not expected to undertake any implementation.

If you wish to include diagram(s), upload you answer as a PDF using the link. Otherwise, you can enter your answer into the space below.

The input is a 300 by 300satellite image, through a trained CNN network, classify the image as one of the 5 classes. The class is set as the different cloud coverage in the image. Like 0 -20% cloud coverage, 21% - 40% and so on.
Then on the other side, combined one set of temperature and wind director into a string as the input of the LSTM network.
Finally, combine the two output from the previous network, through two fully connected layer output the probability of rainfall.

CURRENTLY UPLOADED FILES

AI Q7.pdf Download Remove

Please select file(s) Select file(s)

Save Answer Last saved on **Jun 02 at 10:05 AM**

Q7.2
6 Points

Give details of how you would train and test the system, including the data you would need, how the training is carried out, choice of loss function, and at least one performance measure.

First, a 300 by 300 current image shot by satellite is needed. Through a solo CNN network, it will generate the cloud coverage percentage for the current image. And a set of text strings, which contains temperature and wind direction. Using two LSTM compare with the last information, and output the result. And use cross-entropy as the loss function. The system predicts the 6 hours future weather, so if the system output there are about 60% per cent or higher possibility to rain. We can see if there is raining after 6 hours. Or the time to train the network, we want it trained as soon as possible.

Save Answer Last saved on **Jun 01 at 3:44 PM**

Q7.3
2 Points

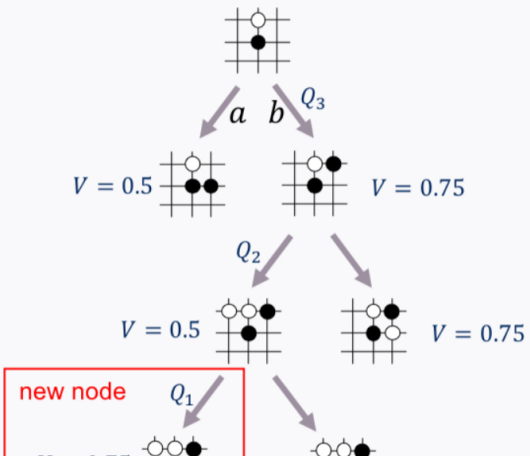
Give one strength and one weakness of your chosen design.

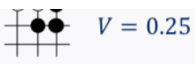
Strenght is easy to understand and implement.
Weakness is the final predicted output is depended on the image output and the data. If the output for the cloud coverage is inaccuracy. Then it will affect the final predicted result, which makes the system more fragile.

Save Answer Last saved on **Jun 01 at 3:42 PM**

Q8 AlphaGo Zero
9 Points

In AlphaGo Zero, a probability distribution π is generated for each move by performing many simulations of forward play from the current board position. These simulations form a tree structure with nodes and edges representing board positions and actions respectively. The simulations are guided by a numerical action value associated with each edge and propagated up the tree. The tree structure below shows the tree after six simulations, each ending with the discovery of a new node. The evaluation V of each board position, provided by the deep network f_θ, is shown.





Q8.1
5 Points

The most recent simulation has discovered the node at the bottom left. What are the action values Q_1, Q_2, Q_3? Show your working.

Q_1

0.75 / 1 = 0.75

Q_2

(0.75 + 0.25 + 0.5) / 3 = 0.5

Q_3

(0.75 + 0.5 + 0.75 + 0.75 + 0.25) / 5 = 0.6

Save Answer Last saved on **Jun 02 at 8:58 AM**

Q8.2
2 Points

After the most recent simulation, how many times has each of the two edges (labelled a, b), leading from the root node, been traversed?

number for a

1

number for b

5

Save Answer Last saved on **Jun 01 at 9:34 PM**

Q8.3
2 Points

Suppose the simulations are now complete for the current board position. Only two edges lead from the root node. What would be the probabilities π_a, π_b computed from the counts given in answer to Q8.2 (assuming the 'peakiness' term $\tau = 1$).

π_a

1

π_b

5 / (5 + 1) = 0.83

Save Answer Last saved on **Jun 02 at 8:58 AM**

Save All Answers Submit & View Submission ›