Student Name: _____

Student Number: _____

Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.

Question	Points	Score
1	10	
2	15	
3	20	
4	15	
5	20	
6	10	
7	10	
8	10	
9	10	
Total:	120	

Graduate students must answer the GR question. The EC question is for extra credit.
Undergraduate students may answer the GR or the EC question for extra-credit.
Total possible points are 110/100 for undergraduate and 120/110 for graduate students.

[10 Pts] 1. **True/False.** Answer the following true/false questions.

(a) ____ Semantic Segmentation labels pixels belonging to each unique object.

(b) ____ In the bed-of-nails unpooling, we place the value for all pixels in the upsampled neighborhood.

(c) ____ Object detection using a deep neural network requires both classification and regression as operators.

(d) ____ We can consider pose estimation as a regression problem addressed using deep neural networks.

(e) ____ Generative Adversarial Networks (GAN)s are examples of explicit probability estimation.

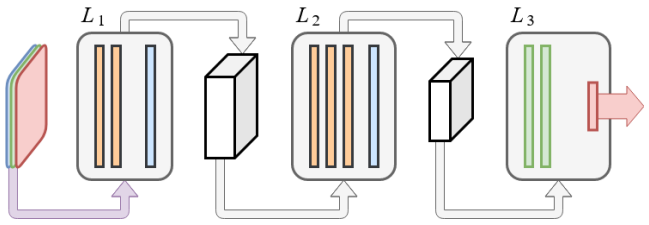
[15 Pts] 2. Answer the following questions about Transpose Convolution

(a) Explain, with equations and graphs, how transpose convolution works.

(b) Assume a 2×2 input $\mathbf{X}$ with a 3×3 mask $\mathbf{M}$ of the following values and with detailed steps show the resultant up sampled matrix.

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \mathbf{M} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix} \tag{1}$$

[20 Pts] 3. We have designed the following CNN to classify among 10 different object classes present in 64×64 RGB images.



The specification of each layer module is as follows:

Layer 1: Two (2) convolution layers, followed by a max-pool layer.

- **L1-Conv-1:** 32, 7×7 filters with appropriate padding to keep output spatial dimensions the same as input.
- **L1-Conv-2:** 64, 3×3 filters with appropriate padding to keep output spatial dimensions the same as input.
- **L1-Pool:** 3×3 max-pool layer to half the spatial dimensions.

Layer 2: Three (3) convolution layers, followed by a max-pool layer.

- **L2-Conv-1:** 128, 3×3 filters with appropriate padding to keep output spatial dimensions the same as input.
- **L2-Conv-2:** 64, 3×3 filters with appropriate padding to keep output spatial dimensions the same as input.
- **L2-Conv-3:** 16, 3×3 filters with padding to keep output spatial dimensions the same as input.
- **L2-Pool:** 3×3 max-pool layer to half the spatial dimensions.

Layer 3: Two (2) fully connected layers with 1000 nodes each, followed by softmax layer.

- **L3-FC-1:** 1000 nodes.
- **L3-FC-2:** 1000 nodes.
- **L3-Softmax:** Appropriate number of nodes to classify all classes.

Fill in the following table:

Layer	Input Size	Kernel Size	Output Size	Pad	No. of Params	No. of Ops
L-1-Conv-1	× ×	× ×	× ×			
L-1-Conv-2	× ×	× ×	× ×			
L-1-Pool	× ×	× ×	× ×			
L2-Conv-1	× ×	× ×	× ×			
L2-Conv-2	× ×	× ×	× ×			
L2-Conv-3	× ×	× ×	× ×			
L-2-Pool	× ×	× ×	× ×			
L3-FC-1		NA				
L3-FC-2		NA				
L3-Softmax		NA				

[15 Pts] 4. We have designed a word auto-complete model using a Recurrent Neural Network (RNN). If we encode each letter with 32 bits (i.e., the input dimensionality of 32) and use 64 hidden nodes in the network, answer the following questions:

(a) Draw the network diagram and clearly identify the input layer, hidden, and output layers.

(b) Clearly identify the network parameter matrices with appropriate names.

(c) Calculate the dimensionality of the input, hidden, and output units.

(d) Calculate the dimensionality of the network parameter matrices.

(e) Write down the network equations.

[20 Pts] 5. Draw the diagram of a Short Long Term Memory (LSTM) and explain how it works. Write all equations and draw additional diagrams necessary.

[10 Pts] 6. **GoogLeNet:**

(a) Draw the Inception module employed in GoogLeNet.

(b) Explain the main difference between this module and the convolution layers of the VGGNet.

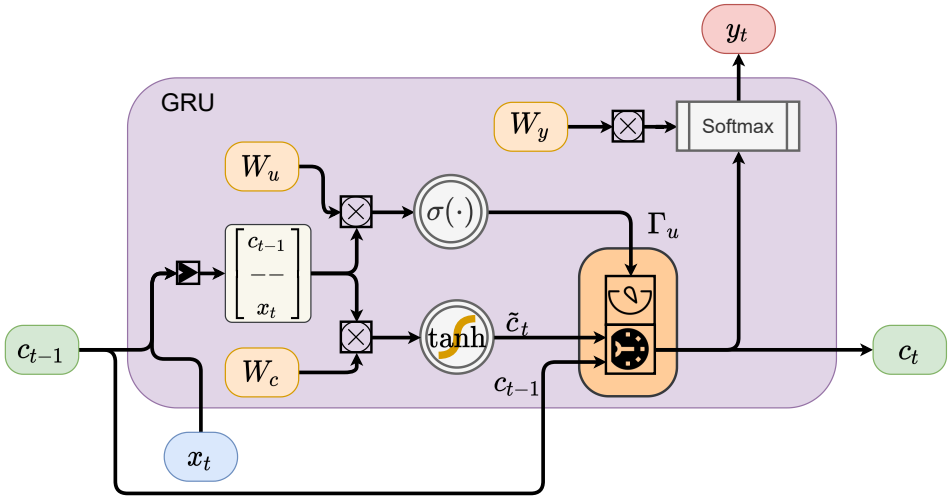
(c) Explain what bottleneck layers are in GoogLeNet and what is the main reason for utilizing them?

[10 Pts] 7. **ResNet:**

(a) Explain why stacking many layers of VGGNet or GoogLeNet decreases the network accuracy.

(b) Explain how ResNet solves this problem and allows for much deeper networks.

[10 **EC**] 8. Write down the gradient of the cell memory (c_t) with respect to network parameters W_c and W_u for the following diagram Gated Recurrent Unit (GRU). Hint: Utilize the network equations discussed in class in conjunction with the network graph below.



[10 **GR**] 9. In the paper Unsupervised GANs (download from link below or from Canvas), answer the following questions: https://www.dropbox.com/s/1ukcgg7l4zyvjhk/UnsupervisedGANS.pdf?dl=0

(a) How does the DCGAN differ from a normal GAN architecture?

(b) What use cases can cause the proposed architecture to fail?