

MIDDLE EAST TECHNICAL UNIVERSITY

SEMESTER I EXAMINATION 2024-2025

**CENG 403 – Deep Learning - CNN Fundamentals & Convolution
Types**

January 2025

TIME ALLOWED: 3 HOURS

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **SIX (6)** questions and comprises **EIGHT (8)** printed pages.
2. Answer all questions. The marks for each question are indicated at the beginning of each question.
3. Answer each question beginning on a **FRESH** page of the answer book.
4. This **IS NOT an OPEN BOOK** exam.
5. Show clear reasoning for your answers, especially intuitive explanations.
6. For architectural diagrams, draw clear components and explain design choices.
7. Connect concepts to examples discussed in lectures where relevant.
8. Explain the practical implications of design decisions.

Question 1. CNN Motivation and Limitations of Multi-Layer Perceptrons (25 marks)

Based on the professor's discussion: "If you want to work with real life large scale problems let's say object recognition with high resolution...you want to recognize objects in such high resolution images what we do is we vectorize...then we use multiperceptron."

- (a) The professor calculated that for a 1000×1000 input image with 10,000 neurons in the first hidden layer, "just this layer is going to introduce 10 billion parameters." Explain why this parameter explosion is problematic and analyze the professor's statement that "more parameters means more data" with the quadratic relationship. (8 marks)
- (b) Analyze the professor's examples of equivariance and invariance problems. For image segmentation, explain why we need "if you transform the input...we expect the output to be transformed by the same transformation." For object recognition, explain why we need the prediction to "be independent of that transformation." (10 marks)
- (c) The professor showed how shifting an image by one pixel in a multi-layer perceptron causes "all of the activations...they are going to change." Explain why this is a "very severe limitation" and how CNNs address this through architectural design. (7 marks)

Question 2. Neuroscience Inspiration and CNN Fundamentals (22 marks)

The professor discussed Hubel and Wiesel's Nobel Prize-winning findings: "They did recordings real recordings from the brains of cats...they showed that...connectivity in biological neurons are not fully connected...neurons have receptive fields."

- (a) Explain the two key findings from Hubel and Wiesel that inspired CNN design: restricted connectivity and columnar structure. Describe how these translate to "receptive fields" and specialized neurons for "recognizing certain patterns at different scales." (10 marks)

- (b) Compare the evolution from Neocognitron (1979) to CNNs. Explain how Fukushima's "simple cells" and "complex cells" correspond to "convolution" and "pooling" in modern CNNs, and why gradient descent training was crucial for practical success. (8 marks)

- (c) The professor mentioned that in CNNs we use "fixed receptive field size for every neuron" unlike the brain where "in the central parts...we have smaller receptive field...in the peripheries...we have higher receptive field." Discuss the implications of this design choice. (4 marks)

Question 3. CNN Architecture and Parameter Sharing (20 marks)

The professor emphasized: "The first critical bit is that connectivity is restricted...the second important change...is that the weights are shared...we have the same parameters shared and used by every receptive field."

- (a) Explain how restricted connectivity and parameter sharing address the dimensionality problem. The professor stated: "if this is 1 million...the number of parameters here it doesn't depend on that actually...we just have three parameters W_1 W_2 W_3 ." (8 marks)
- (b) Analyze the trade-off the professor discussed: "if you restrict connectivity...a neuron receives information just from a restricted part of the input" versus the solution of increasing depth so "neurons in the following layers...will have the chance to integrate information across the whole input." (8 marks)
- (c) The professor showed that parameter sharing provides "equivariance to translation...if you shifted the input...the pattern will shift to the next receptive field...accordingly the activations will shift as well." Explain this mathematical property and why it doesn't extend to scale and rotation. (4 marks)

Question 4. Convolution Operation and Hyperparameters (25 marks)

The professor explained: "We have stride...padding...receptive field size...we need to be careful when we are choosing filter size padding and stride because this needs to be an integer."

- (a) Derive and apply the formula the professor used: "size of the next layer = $(W - F + 2 \times \text{padding}) / \text{stride} + 1$." For AlexNet's first layer with input 227×227 , filter size 11, stride 4, padding 0, verify the output size of 55×55 . (8 marks)
- (b) Analyze the professor's AlexNet example where "stride is four...we are reducing dimensionality by a factor of four that is a huge reduction." Explain why this large stride works in early layers but would be problematic in later layers, citing his explanation about information redundancy. (10 marks)
- (c) The professor discussed the channel dimension: "when we say a two-dimensional filter actually it might have a third dimension...that spans the channels of its input layer." For 96 filters of size $11 \times 11 \times 3$, calculate the total parameters and explain how "different filters...learned to extract different types of information." (7 marks)

Question 5. Alternative Convolution Types (28 marks)

The professor introduced multiple convolution variants: "We have different ways to actually restrict connectivity and share parameters in a layer."

- (a) Compare unshared convolution, dilated convolution, and transposed convolution. For each, explain the professor's rationale: unshared for problems without equivariance needs, dilated for increasing "effective coverage...without increasing number of parameters," and transposed for "upsampling." (12 marks)

- (b) Analyze separable convolution as the professor explained: "we can write a 3×3 matrix as a multiplication of two vectors...we can reduce the number of parameters...from nine parameters...to six parameters." Explain depthwise separable convolution and its efficiency benefits. (10 marks)

- (c) The professor described 1×1 convolution as controlling "the number of channels...without using any...without combining any information from a neighborhood." Explain how this operation works and its role in efficient architectures like GoogleNet that will be discussed later. (6 marks)

Question 6. Deformable Convolution and Advanced Concepts (30 marks)

The professor introduced deformable convolution: "What if we learn directly from the data...a better positioning of the filter...for each position what if I estimate an offset along X and Y that would be more meaningful."

- (a) Explain the concept of deformable convolution as "dynamic receptive field." Describe how "for each parameter...we need to have two offsets for X and Y δX and δY " and why this makes the operation "very expensive." (12 marks)

- (b) The professor showed the benefit: "receptive fields are adjusted automatically...in such a way that receptive field actually pays attention to the most relevant content." Analyze the sheep example and explain why this improves upon vanilla convolution's "rigid...regular" placement. (8 marks)

- (c) Solve the backpropagation challenge the professor discussed: "indices don't enter the calculations...we cannot do that because it is not part of the calculation." Explain how bilinear interpolation makes "everything differentiable" by incorporating offsets into the computation. (10 marks)

END OF PAPER