MIDDLE EAST TECHNICAL UNIVERSITY

SEMESTER I EXAMINATION 2024-2025

CENG 403 – Deep Learning - CNN Visualization & Classic Architectures

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. Class Activation Maps and Global Average Pooling (25 marks)

Based on the professor's explanation: "Each channel actually specializes in capturing one part or maybe the whole part of an object... it functions as a confidence map over pixels."

(a) The professor explained that global average pooling forces each channel to specialize. Explain how class activation maps exploit this property to visualize what the network is "paying attention to." Include the mathematical process of weighting channels. (10 marks)

(b) Compare class activation maps with Grad-CAM as described by the professor. Explain why Grad-CAM replaces learned weights with "summation of gradients with respect to the channel" and how this makes the visualization "input dependent." (10 marks)

(c) The professor mentioned that "if you do this with a CNN that has fully connected layers you are not going to get such activation maps." Explain why global average pooling is essential for class activation maps and why fully connected layers prevent this visualization. (5 marks)

Question 2. Feature Inversion and Network Understanding (22 marks)
The professor discussed: "Given the feature vector for X we can try to generate an image whose feature vector is similar to that image."

(a) Formulate the feature inversion optimization problem as the professor described. Explain why this is useful for understanding "what the CNN is doing" and "what level of detail it has kept and captured at different layers." (8 marks)

(b) The professor showed reconstructions from different CNN layers. Explain his observation: "In the earlier layers information is redundant and we are able to capture the low-level details... when you go up in the network low-level details are lost." (8 marks)

(c) The professor mentioned regularization terms to make generated images "visually pleasing." List and explain the regularization terms he discussed for controlling "smoothness" and ensuring "neighboring pixels have similar intensities." (6 marks)

Question 3. Adversarial Attacks (23 marks)

Based on the professor's explanation: "We can generate a noise image that's called R with a very small magnitude such that when you add this noise image to the original image human eye is not able to visually distinguish this from the original."

(a) Formulate the adversarial attack optimization problem as described by the professor. Distinguish between "targeted attack" (where "we have a target class in mind") and "untargeted attack" (where "as long as the correct class is not predicted it doesn't matter"). (8 marks)

(b) Explain the difference between "white box approach" and "black box approach" for generating adversarial attacks. Describe the professor's explanation of using a "proxy network that mimics the predictions of the blackbox model." (8 marks)

(c) The professor discussed serious implications: autonomous cars could "misdetect a traffic sign" or "miss a crossing." Explain why adversarial robustness is challenging, including the professor's point about decision boundaries and the "trade-off between robustness and accuracy." (7 marks)

Question 4. AlexNet Architecture and Design (25 marks)
The professor described AlexNet as "the 2012 ImageNet winner" with "60 million parameters" that had to be "split into two branches" because "they were not able to fit the whole network into a single GPU."

(a) Analyze the AlexNet architecture as described by the professor. Explain why they used "a stride of four" in the first layer and why this "huge reduction in dimensionality" worked in early layers but "would lose performance if tried in the top of the network." (10 marks)

(b) The professor found it "striking" that AlexNet filters showed specialization: one GPU learned "more or less color insensitive" filters while the other learned "more color selective filters." Explain this phenomenon and what it suggests about "group convolution." (8 marks)

(c) List the key techniques AlexNet used to "mitigate overfitting" as mentioned by the professor, including their approach to "manually" decaying learning rate when "the loss was not improving." (7 marks)

Question 5. GoogleNet and Inception Modules (30 marks)
The professor explained GoogleNet's contributions: reducing parameters "from 60 million in AlexNet to 4 million" while achieving better performance.

(a) Explain the motivation for inception modules as described by the professor: "information can exist at multiple scales" and "if you use a fixed receptive field size at a layer actually you are restricting the type of information you can extract to only a fixed scale." (8 marks)

(b) The professor described the "scalability problem" with naive inception modules: "the number of channels explodes." Draw the improved inception module and explain how "one by one convolution" solves this by ensuring "if the input layer has C channels the output layer after concatenation has C many channels as well." (12 marks)

Draw improved inception module with 1x1 convolutions

Show how channel count is controlled

(c) Explain GoogleNet's second contribution: "adding fully connected layers into the intermediate layers of the network." Discuss both benefits (gradient flow, multi-scale recognition) and drawbacks (reproducibility issues, loss of generalization) as mentioned by the professor. (10 marks)

Question 6. ResNet and Skip Connections (25 marks)
The professor explained the key insight: "After a certain number of layers performance doesn't improve... it actually becomes even worse" and ResNet's solution using skip connections.

(a) Analyze the professor's explanation of why deeper networks perform worse: "when you multiply all those gradients through all of these layers you don't get useful gradients to the earlier parts of the network." Explain why this isn't simply overfitting. (8 marks)

(b) Draw a ResNet block and explain the professor's description: "in forward pass we have a function of the input as well as the identity being propagated" and "in backward pass gradient will flow through two paths." (10 marks)

Draw ResNet block showing skip connections

Show both forward pass (identity + function) and gradient flow

(c) The professor mentioned two interpretations of why ResNet works: the "smoother loss function" hypothesis and viewing ResNet as "an ensemble

of multiple sub networks." Explain both interpretations and discuss why "there is no theorem showing this." (7 marks)

END OF PAPER