

MIDDLE EAST TECHNICAL UNIVERSITY

SEMESTER I EXAMINATION 2024-2025

CENG 403 – Deep Learning - CNN Architectures & RNN  
Introduction

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. ResNet and Skip Connections**

(25 marks)

Based on the professor's explanation: "The identity function doesn't have to be learned as part of the solution, we are giving identity as part of the solution."

- (a) The professor mentioned that ResNet researchers "noticed that up to a certain number of layers actually performance degrades." Explain why this was counterintuitive and how skip connections solve this problem. Include the professor's explanation of why networks have difficulty learning the identity function. (8 marks)

- (b) Draw a ResNet block as described by the professor, showing how "in forward pass the identity can be implemented as part of the solution" and "in backward pass through these residual connections gradient can flow without any vanishing issues." (10 marks)

**Draw ResNet block showing identity path**

**Show how gradient flows through skip connections**

- (c) The professor explained that ResNet "smooths the loss surface" and provides "robustness to slight changes in the weights." Explain this concept using the professor's reasoning about how identity transform provides robustness when weights change slightly. (7 marks)

**Question 2. ResNeXt and Multiple Pathways** (20 marks)

The professor described ResNeXt as extending ResNet "by utilizing multiple paths acting on the same input layer."

- (a) Compare ResNeXt with the inception module from GoogleNet as the professor discussed. Explain the key difference: "in inception module the filter sizes are different in different paths, whereas here in ResNeXt the sizes are the same." (8 marks)
- (b) The professor noted that ResNeXt "can provide better performance than ResNet however because of the multiple paths it is slower." Explain this trade-off and when you would choose ResNeXt over ResNet according to the professor's guidance. (7 marks)
- (c) Draw a ResNeXt block showing multiple functions  $f_1(x), f_2(x), \dots, f_{32}(x)$  acting on the same input as the professor illustrated, resulting in  $x + f_1(x) + f_2(x) + \dots + f_{32}(x)$ . (5 marks)

**Question 3. DenseNet Architecture**

(22 marks)

Based on the professor's explanation: "We have residual connections not only skipping the next block but skipping to and making connections to all following blocks."

- (a) Explain DenseNet's approach to skip connections as described by the professor. Why does having "skip connections for all following layers" make it "a very dense network" and potentially "more difficult to implement"? (8 marks)
  
  
  
  
  
  
  
  
  
  
- (b) The professor mentioned that DenseNet "can provide better performance compared to ResNet" but wasn't sure "why it didn't become as popular as ResNet." Analyze the potential reasons for this based on implementation complexity and computational requirements. (8 marks)
  
  
  
  
  
  
  
  
  
  
- (c) Draw a DenseNet block showing how "from this layer we have skip connections to all following layers" as the professor described. Show at least 4 layers with their dense connections. (6 marks)

**Draw DenseNet dense connections**

**Show connections to ALL following layers**

**Question 4. Highway Networks and Gating** (23 marks)

The professor explained highway networks as providing "more flexibility to the network where we can modulate the skip connection."

- (a) Explain the highway network concept as described by the professor: "multiply the skip connection by a learnable function" and "multiply this by another function." Write the mathematical formulation the professor presented. (10 marks)
  
  
  
  
  
  
  
  
  
  
- (b) The professor showed how highway networks use gating: "t controls the information coming through here and with this we are also controlling how much x is propagated to the next layer." Explain this gating mechanism and how it differs from simple residual connections. (8 marks)
  
  
  
  
  
  
  
  
  
  
- (c) The professor noted that "highway networks were not as common as ResNets but the idea of gating is actually utilized in many different architectures." Give examples of where this gating concept is used and why it's important. (5 marks)

**Question 5. Binary Networks and Efficiency** (25 marks)

Based on the professor's discussion: "In binary networks people are trying to replace real valued weights with binary weights and this would get rid of multiplication."

- (a) Explain the motivation for binary networks as the professor described: the need to "run these networks on low resource devices edge devices" where "it will be very difficult to run and execute deep architectures requiring a lot of memory." (8 marks)
  
- (b) The professor mentioned specific trade-offs: "we get 30 times around 30 times gain in terms of memory but we lose some accuracy." Calculate the memory savings for a network with 10 million 32-bit parameters when converted to binary, and explain when this trade-off is acceptable. (10 marks)
  
- (c) The professor explained that "if you change the input to binary and work only with binary values throughout the network then actually in addition to obtaining significant memory gain you can actually get very large gain in running time." Explain why binary operations are faster and when this approach is practical. (7 marks)



**Question 6. Introduction to Sequence Problems** (35 marks)

The professor introduced sequence modeling: "We have many problems where we have a sequence we need to process that sequence sequentially."

(a) Classify the following problems according to the professor's framework (one-to-one, one-to-many, many-to-one, many-to-many) and explain your reasoning: (15 marks)

- Image captioning (which the professor said is "a very good example" of one-to-many)
- Spam detection (professor's example of many-to-one)
- Language translation (professor's example with "Turkish" to "English")
- Online speech recognition ("live speech recognition")
- Character recognition in a sequence

(b) The professor emphasized that in sequence problems "the information at a certain point might depend on the information we have seen so far." Explain why this context dependency makes traditional CNNs and MLPs insufficient for sequence modeling. (10 marks)

(c) Compare feedforward networks with RNNs as the professor explained: "In feedforward network we don't have that option but the network has

seen for the previous input we don't know." Draw both architectures showing how RNNs provide "memory capacity" through recurrent connections. (10 marks)

**Compare Feedforward vs RNN architectures**

**Show how RNN maintains memory through recurrent connections**

**END OF PAPER**