Deep Learning

Instructor: Dr. Davoodabadi Farahani
Head TA: Ali Momen

Semester: Fall 2025

Homework 2:

Convolutional Neural Networks (CNNs)

Designed By:

Mohammand Haghighat
`@Haghighat_Mohammad`

Fatemeh Bagheri
`@FatemehBagherii`

Deadline: 7 Aban

Solution Release & Presentation Day: 15-16 Aban

Preface

In this homework, we aim to develop a comprehensive understanding of Convolutional Neural Networks (CNNs), a class of deep learning models that have become the cornerstone of modern computer vision. We move beyond a surface-level view to dissect the core principles that enable CNNs to efficiently process and interpret visual data.
The assignment begins by exploring fundamental concepts such as parameter sharing, which dramatically reduces model complexity, and the versatile applications of CNNs across different data modalities. We then delve into key architectural components like the transposed convolution layer, essential for tasks requiring spatial upsampling such as image generation and segmentation. Subsequently, we investigate the critical properties of invariance and equivariance, analyzing how they influence a model's ability to recognize objects regardless of their position and the trade-offs involved in various computer vision tasks. The homework then confronts a central challenge in network design: the expansion of the receptive field. We examine both traditional methods and their inherent limitations before analyzing modern, efficient solutions like the Inception module and dilated convolutions, which capture multi-scale features without prohibitive computational costs. Finally, we bridge theory and practice by performing a quantitative analysis of a complete network architecture, calculating the parameters, computational load, and effective receptive field for each layer to solidify our understanding of how design choices impact model performance.

Notes and Honor Code

This homework is part of the Deep Learning course at IUST offered in Fall 2025 by Dr.Davoodabadi. Please read all instructions carefully before starting.

- Collaboration is encouraged, but each student must submit their own work.
- Submitting other students' work or copying solutions from another student would result in a 0 score on the assignments
- Typesetting in LaTeX is strongly recommended.
- Clearly mention collaborators, if any.

If you have any further questions regarding the submission policies,course policies e.t.c. feel free to contact the Head Teaching Assistant of the course, Ali Momen, via Telegram.

Submission

The deadline for this homework is 7 Aban 23:59 PM.
Please submit your work by following the instructions below:

* Place your solution alongside the Jupyter notebook(s). Your written solution must be a single PDF file named **HW2_Solution.pdf**.

* Zip all the files together with the following naming format:
DL_HW2_[StudentNumber]_[FullName].zip.
Replace `[FullName]` and `[StudentNumber]` with your full name and student number, respectively. Your `[FullName]` must be in CamelCase with no spaces.

* Submit the zip file through Quera in the appropriate section.

1 Theoretical Problems

Problem 1.

- A) What is the concept of parameter sharing in Convolutional Neural Networks (CNNs), and how does it affect the model training process? (5 points)
- B) Explain whether Convolutional Neural Networks are suitable for each of the following scenarios: (10 points)
  - Monitoring a specific species of wolf in the wild using a drone.
  - Extracting text from audio.
  - Identifying the action performed within a video.
- C) Explain the operation of a Transposed Convolution layer. What is its primary purpose, and in what applications, such as semantic segmentation or image generation, is it commonly used? (5 points)

Problem 2. (5 points)

- What is the difference between invariance and equivariance?
- Which layer is most responsible for creating local translation invariance in a CNN?
- Are standard CNNs naturally invariant to image rotation?
- What is the primary motivation for having translation invariance in an image classification model?
- Name 3 computer vision task where strong translation invariance would be a disadvantage.

Problem 3.

Introduction

For a Convolutional Neural Network (CNN) to recognize large objects or complex concepts within an image (e.g., identifying a "cat" rather than just its "ear" or "eye"), neurons in its deeper layers must be influenced by a large portion of the input image. This region of the input image that affects the activation of a particular neuron is called its Receptive Field. A primary goal in designing deep architectures is to effectively increase the receptive field of the final neurons. (10 points)

Part A: Traditional Methods and Their Costs

1. In a simple, sequential CNN (e.g., a VGG-style architecture), what are the two primary strategies for increasing the receptive field of neurons in successive layers? Briefly explain each.
2. Why is using very large filter sizes (e.g., `15x15`) in a single convolutional layer considered an inefficient and costly approach for rapidly increasing the receptive field? (Comment on the number of parameters and computational cost).
3. Despite being more parameter-efficient, why can stacking a very large number of layers with small filters (e.g., `3x3`) lead to other significant problems during the training process? (Refer to the optimization challenges in very deep networks).

Part B: Clever Solutions in Modern Architectures

Modern architectures have devised clever solutions to the "receptive field dilemma" by expanding it without incurring prohibitive costs. Analyze two such techniques:

1. **The Inception Architecture (GoogLeNet):** How does the Inception module attempt to extract features at "multiple scales" simultaneously? Explain how this module achieves varied receptive fields using parallel paths with different filter sizes, and how it manages the computational cost using 1x1 convolutions.

2. **Dilated (or Atrous) Convolutions:** Explain the concept of a "dilated convolution" and the "dilation rate" parameter. How does this type of convolution manage to exponentially increase the receptive field without increasing the number of parameters or the computational load compared to a standard convolution with the same filter size? (A simple example with a 3x3 filter and a dilation rate of 2 would be helpful in your explanation).

Problem 4. For the neural network defined below, calculate the following for each layer, assuming a 256x256 color image is used as input:

- The number of parameters.
- The number of Multiply-Accumulate operations (MACs).
- The effective receptive field.

Note: Please round down all decimal values to the nearest integer. (15 points)

Network Architecture

- **Layer1**: `nn.Conv2d(in_channels=3, out_channels=32, kernel_size=(7,7), stride=1, padding='same')`
- **bn1**: `nn.BatchNorm2d(32)`
- **Layer2**: `nn.Conv2d(in_channels=32, out_channels=64, kernel_size=(5,5), stride=2, padding='valid')`
- **bn2**: `nn.BatchNorm2d(64)`
- **Layer3**: `nn.AvgPool2d(kernel_size=(2,2), stride=2)`
- **Layer4**: `nn.Conv2d(in_channels=64, out_channels=128, kernel_size=(3,3), stride=1, dilation=2, padding='valid')`
- **bn3**: `nn.BatchNorm2d(128)`
- **Layer5**: `nn.Conv2d(in_channels=128, out_channels=128, kernel_size=(3,3), stride=1, dilation=1, padding='valid')`
- **bn4**: `nn.BatchNorm2d(128)`
- **Layer6**: `nn.AvgPool2d(kernel_size=(2,2), stride=2)`
- **Layer7**: `nn.Conv2d(in_channels=128, out_channels=256, kernel_size=(3,3), stride=1, padding='valid')`
- **bn5**: `nn.BatchNorm2d(256)`

- **Layer8**: `nn.AvgPool2d(kernel_size=(2,2), stride=2)`
- **fc1**: `nn.Linear(in_features=43264, out_features=1024)`
- **fc2**: `nn.Linear(in_features=1024, out_features=1024)`
- **dropout**: `nn.Dropout(p=0.5)`
- **fc3**: `nn.Linear(in_features=1024, out_features=10)`

2 Coding challenges

2.1 CNN Notebook (50 points)

In this notebook, you will learn the basics of Convolutional Neural Networks!!! You will compare the performance of different optimizers like Adam when training a simple CNN, build your first CNN to classify images from the Fashion-MNIST dataset and build a ResNet model from scratch!!!!

Complete all the **TODO** sections in the notebook to obtain the correct results and achieve a full score. The deliverable for this part is **one completed notebook with results**.

Good Luck!!!

References

[1] Christopher M. Bishop and Hugh Bishop. *Deep Learning: Foundations and Concepts*. Springer, 2024.

[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. Book in preparation for MIT Press.