

# Intro Deep Learning Homework 6

Jaskin Kabir

Student Id: 801186717

GitHub:

[https://github.com/jaskinkabir/Intro\\_Deep\\_Learning/tree/master/HM6](https://github.com/jaskinkabir/Intro_Deep_Learning/tree/master/HM6)

April 2025

# 1 Problem 1: ViT From Scratch

## 1.1 Model Architectures

This section focuses on training four variations of a vision transformer based network for image classification on the CIFAR-100 Dataset. The four models shared the same parameters other than the patch size and number of attention layers.

The shared parameters are as follows:

- Embedding Size: 192
- MLP Hidden Size: 384
- Number of Attention Heads: 4

The four variations are combinations of a patch size of 4 or 8 and either 4 or 8 attention layers. A summary of the model complexities can be seen in Table 1 below.

Model	Parameters	MACs
<b>Patch4 Attn4</b>	1,377,508	756,480
<b>Patch4 Attn8</b>	2,565,604	756,480
<b>Patch8 Attn4</b>	1,395,940	756,480
<b>Patch8 Attn8</b>	2,584,036	756,480

Table 1: Model Complexities