Inter IIT tech meet preparation

Computer Vision

Task 3

Goal – Part 1: Research Report on Vision Transformer

Participants must write a **report** summarizing and analyzing the Vision Transformer architecture.

- Report Requirements:
- **Overview of the Vision Transformer architecture**
- **Core components**:
  - Patch embedding
  - Positional encoding
  - Transformer encoder (multi-head self-attention, feedforward layers)
  - Classification head
- **Comparison with CNN-based models (like ResNet/SEResNet)**
- **Why ViT performs well or struggles in vision tasks**
- **Insights from the original paper (e.g., "An Image is Worth 16x16 Words")**
- **Challenges in training ViTs vs. CNNs**
- **Your observations or opinions on practical use cases**

Goal – Part 2: Implement ViT from Scratch in PyTorch

Participants must implement the **Vision Transformer (ViT)** architecture from scratch using **PyTorch**, and apply it to the **CIFAR-10 image classification dataset**.
Task Requirements:

- **Use the CIFAR-10 dataset** (available via `torchvision.datasets`)
- Implement:
- Image patching (e.g., 4x4 or 8x8 patches)
- Linear patch embedding
- Positional encoding
- Transformer encoder layers:
  - Multi-head self-attention
  - Feed-forward block

- Classification head (e.g., CLS token + MLP)
- Train on CIFAR-10 and report **final accuracy on the test set**
- **Aim for ≥ 75% accuracy for full credits**

Submission Requirements (ZIP File)

Your submission must include:

- vit.py: Well-commented source code (modular, clean)
- `train.py`: Code to train and evaluate the model
- ViT.ipynb: A notebook having complete code with all evaluation metrics on testing data.
- `report_part1.pdf`: ViT research report

Guidelines

- **Language**: Python with PyTorch is mandatory for Part 2
- **No plagiarism**: All content and code must be original
- **Code quality matters**: Proper structure, readability, comments
- **Reproducibility**: Include instructions or scripts to re-run your training and evaluation
- **Deadline :** 8 July 2025

Evaluation:

- Accuracy – 50
- Code quality and clarity – 10
- Report clarity and insights - 40