Computer Vision

ASSIGNMENT:03

Due Date: 16th-Nov- 2024

Instructions:

- This assignment is a individual project.
- Submit your assignment by **16th-Nov-2024**. Late submissions will incur a **penalty**. No submissions will be accepted after **Due Time** without prior approval.
- Your work must be your own. You may discuss ideas with peers but do not share code. Plagiarism will result in serious consequences, including a failing grade.
- Ensure your code is clean, readable, and well-organized. Use meaningful variable names and comments. Include a **README** explaining how to run your code.
- Test your code thoroughly. Handle edge cases and report any known issues in the README.
- Submit through **GCR**. Include your code, README, and other required deliverables in a **.zip** or **.tar.gz** file.
- If you have questions, reach out during office hours or help sessions, but don't wait until the last minute.
- Grading will be based on correctness, code quality, efficiency, and adherence to the instructions.

• Task: Human Action Recognition

Human action recognition in videos involves identifying and categorizing different activities or movements displayed in a sequence of frames. The Vision Transformer (ViT) model, traditionally used for image classification, will be fine-tuned to recognize actions from short video clips. By learning patterns across frames, the model will distinguish among multiple action classes.

• HMDB Dataset

The **HMDB (Human Motion Database)** dataset is a well-known benchmark for human action recognition, featuring over 6,800 video clips across 51 action categories like running, eating, and waving. This dataset includes real-world, dynamic actions from various sources, providing diverse and challenging scenarios for testing action recognition models.

HMDB dataset

1. Dataset Preprocessing

- Extract frames from each video in the HMDB dataset.
- Resize the frames to the input size expected by Vision Transformer (ViT).
- Apply data augmentation techniques to improve generalization, such as cropping, flipping, and normalization.

2. Loading the Vision Transformer Model

- Load a pretrained ViT model suitable for image-based tasks.
- Modify the model’s final layers if necessary to match the number of classes in the HMDB dataset.
- You are allowed to use high level libraries such as hugging face etc.

3. Setting Up Training Configurations

- Choose appropriate batch size and number of epochs for effective training.
- Set a suitable learning rate for fine-tuning the model on this dataset.

4. Check pointing and Early Stopping

- Use check pointing to save the best-performing model during training.
- Implement early stopping based on validation performance to avoid overfitting.

5. Model Evaluation

- Evaluate the model’s accuracy on the test set.
- Aim for high accuracy at least 90% by fine-tuning parameters and iteratively improving the model’s performance.