Weekly Report – Week 2

Ahmedabad University

Project Topic:

Investigation of Attention Mechanism for Multimodal Transformers

Course Name: CSE623 Machine Learning Theory and Practice
Professor Name: Prof Mehul Raval
University: Ahmedabad University

Team Name: FrameTrackers

Team Members:

1. **Shlok Shelat** - AU2240025
2. **Shrey Salvi** - AU2240033
3. **Rushi Moliya** - AU2240020

Summary of Work Done This Week: Our Group has performed a literature survey on the external knowledge-guided clustering together with its applications for multimodal transformers during this week. The main idea/focus was centered around increasing understanding of the Image Clustering with External Guidance paper and its proposed Text-Aided Clustering (TAC) method. Our research analyzed TAC's reasons for development and its textual semantics methods for image clustering and enhancements regarding classical clustering operations.

Literature Survey:

- We studied clustering methods starting from classic clustering techniques and following with deep clustering techniques and self-supervised clustering techniques.
- We investigated both the clustering weaknesses of inherent supervision signs and the need for external information in their work.
- We also focused on TAC because this method applies WordNet semantic data to improve clustering outcomes.
- TAC reveals a process of sharing neighborhood information between different modalities to maximize clustering precision rates.

Understanding the TAC Implementation:

- A detailed study of the TAC official paper alongside an evaluation of its code repository took place.
- The process for extracting external knowledge through WordNet allowed the construction of a text space that served for clustering operations.
- The paper evaluated a process that used mutual distillation as a method for accomplishing cross-modal collaboration.
- The research determined key elements that affected cluster performance results.

Challenges Identified:

- The study focuses on how clustering processes combine external text-based information.
- The replication process of the research methodology requires the utilization of the provided computer code.
- The efficient management of big data sets alongside TAC presented its own set of experimental challenges.

Plan for Next Week: Next week, we will focus on reproducing the results from the *Image Clustering with External Guidance* paper by implementing and running the TAC model on benchmark datasets.

Implementation Goals:

- Set up and run the TAC code on the provided datasets.
- Validate the performance of the model and compare it with the reported results.

Performance Analysis:

- Conduct initial experiments on TAC's ability to enhance clustering using external guidance.
- Evaluate computational efficiency and feasibility of using TAC for multimodal applications.

Conclusion: This week, we strengthened our understanding of clustering methodologies supported by external inputs along with TAC procedures. We have conducted an implementation of TAC while duplicating its original output values. The implementation will help us broaden the understanding of the same.