Text-Guided Video Object Segmentation

Designing Systems for Machine Learning

TEAM MEMBERS:

Amitabh (2023EEZ0025)	Model Design and Implementation
Komal (2024EEZ0013)	Model Design and Implementation
Anshika (2021CSB1069)	System, Interface Design, and Implementation

PROBLEM STATEMENT

In this project, we aim to develop a system to segment text-referred objects from a video input using a multimodal tracking transformer (MTTR). We aim to develop a user interface where we can give video and text (referring to the object to be segmented) as input and the segmented object as the output. The overall system design will be modular and carefully built based on best design practices.

EXPECTED RESULTS

1. Developing a working system that can segment the video based on text .
2. Iteratively improve the system based on shortcomings and difficulties that may arise during the development.

REFERENCES

1. Miao, Bo, et al. "Spectrum-guided multi-granularity referring video object segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.
2. Wu, Jiannan, et al. "Language as queries for referring video object segmentation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
3. Botach, Adam, Evgenii Zheltonozhskii, and Chaim Baskin. "End-to-end referring video object segmentation with multimodal transformers." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.