

Discussion Presentation Instructions

The goal of the discussions in this course is to gain familiarity with and practice critical thinking and critique of recent, cutting-edge papers in the field. In each class session that is allocated to these discussions, ~4 papers will be covered. Assigned presenters will give a presentation explaining and providing critique of each paper, and then assigned questioners and the rest of the class will participate in Q&A and discussion.

Presenters should read and understand the paper carefully, and present each paper in 12 minutes, covering the following aspects (with grade breakdown):

- **(15%) Overview:** Summarize the key motivations, objectives, and contributions of the paper. Include a high-level understanding of the background and significance of the work.
- **(30%) Approach:** Explain the methods and techniques used in the paper. Include technical details, and highlight novel aspects of the methodology.
- **(10%) Real-world considerations:** Discuss practical aspects of the work, such as compute requirements for training or inference. Under what settings could the work be used or built upon, in academic, industry, or clinical environments?
- **(30%) Advantages and disadvantages:** Offer your perspective on the strengths and weaknesses of the work. This could include the method, evaluation, impact, or other aspects.
- **(10%) Avenues for future work:** Comment on opportunities for future research. What are the remaining limitations? How could the work be extended or improved?
- **(5%) Questions for other discussion sessions** (see next section).

Questioners should also read the paper carefully, and will be responsible for preparing questions for the discussion. These prepared questions should not be clarification or fact-check questions; instead they should be comments and questions that stimulate discussion and/or request opinions. At least two questions should be submitted in a pdf document through Gradescope by 11:59pm PT the day before class, and each questioner should ask at least one question during the discussion period for the paper in class. These questions will complement additional questions raised by the class at large.

Each student should sign up for 1 paper as presenter, and 2 papers as questioner. Signup details will be provided several weeks into the quarter, after class enrollment has stabilized.

Paper List:

Discussion: Recent Advances in Vision Representation Learners

1. [Rotary Position Embedding for Vision Transformer](#)
2. [Vision Transformers Need Registers](#)
3. [SAM 2: Segment Anything in Images and Videos](#)
4. [Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology](#)
5. (Not presented: [Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis](#))

Discussion: Recent Advances in Vision-Language Representation Learners

1. [ImageBind: One Embedding Space To Bind Them All](#)
2. [Interpreting CLIP's Image Representation via Text-Based Decomposition](#)
3. [CLIP-Driven Universal Model for Organ Segmentation and Tumor Detection](#)
4. [Visual Explanations of Image-Text Representations via Multi-Modal Information Bottleneck Attribution](#)
5. (Not presented [GLIPv2: Unifying Localization and VL Understanding](#))

Discussion: Recent Advances Vision Generative Models

1. [Adding Conditional Control to Text-to-Image Diffusion Models](#)
2. [An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion](#)
3. [Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets](#)
4. [Generation of synthetic whole-slide image tiles of tumours from RNA-sequencing data via cascaded diffusion models](#)
5. (Not presented [Adversarial Diffusion Distillation](#))

Discussion: Recent Advances in Vision-Language Generative Models

1. [VILA: On Pre-training for Visual Language Models](#)
2. [Transfusion: Predict the Next Token and Diffuse Images with One Multi-Modal Model](#)
3. [A Multimodal Generative AI Copilot for Human Pathology](#)
4. [VideoAgent: Long-form Video Understanding with Large Language Model as Agent](#)
5. (Not Presented [Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution](#))

Discussion: Recent Advances in Computing Efficiency

1. [An image is worth 1/2 tokens after layer 2: Plug-and-play inference acceleration for large vision-language models](#)

2. [FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness](#)
3. [LoRKD: Low-Rank Knowledge Decomposition for Medical Foundation Models](#)
4. [ST-Adapter: Parameter-Efficient Image-to-Video Transfer Learning](#)
5. (Not presented [DistriFusion: Distributed Parallel Inference for High-Resolution Diffusion Models](#))