Questions for the Block-Seminar on Deep Learning WS23-24

Please send your answers for each paper to the corresponding advisor BEFORE the beginning of the seminar. You do not need to answer the questions for the paper that you are presenting.

1 LoRA: Low-Rank Adaption of Large Language-Models (Advisor: Leonhard Sommer)

Email to: sommerl@cs.uni-freiburg.de

Question 1 Using LoRA with GPT3, what are the VRAM memory reductions at training and at deployment? (Assuming a rank of 4 and only adapting the query and value projection matrices) (3 sentences)

Question 2 Using LoRA, compared to the Adapter H, where lies the major advantage in memory savings at training or at deployment? (2 sentences)

Question 3 Using LoRA with GPT2, what is the additional inference latency? (1 sentence)

2 The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision) (Advisor: Jelena Bratulic)

Email to: bratulic@cs.uni-freiburg.de

Question 1 What input modes are supported by GPT4V? For each mode name 2 possible tasks which could use such mode. (3 sentences)

Question 2 Explain the prompt design for the default working mode (zero-shot) and in-context few-shot learning. (2 sentences)

Question 3 Name some failure cases demonstrated in the paper. Did different prompting design make it work? (3 sentences)

3 Large Language Models Cannot Self-Correct Reasoning Yet (Advisor: David Hoffmann)

Email to: hoffmann@cs.uni-freiburg.de

Question 1 What are the tested prompting strategies studied? Explain each of them in a short sentence. (4 sentences)

Question 2 How is intrinsic self-correction different from self-correction with oracle feedback? (1-2 sentences)

Question 3 What's the difference between multi-agent debate and self-consistency? (1-2 sentences)

4 Hiera: A Hierarchical Vision Transformer without the Bells-and-Whistles (Advisor: Artur Jesslen)

Email to: jesslen@cs.uni-freiburg.de

Question 1 Why and how is MAE used in the paper? ($\sim$ 3 sentences)

Question 2 Why is the proposed method faster (both training and inference) while reaching higher performances ? ($\sim$ 3 sentences)

Question 3 The authors refer to their method as a *hierarchical* Vision Transformer, explain why it is hierarchical. ($\sim$ 2 sentences)

5 Sigmoid Loss for Language Image Pre-Training (Advisor: Simon Ging)

Email to: gings@cs.uni-freiburg.de

Question 1 What are the advantages when using sigmoid loss instead of contrastive loss? ($\sim$ 3 sentences)

Question 2 What new hyperparameter is introduced when switching from contrastive loss to sigmoid loss, and why is it needed? ($\sim$ 2 sentences)

Question 3 What is the difference between SigLiT and SigLIP? ($\sim$ 2 sentences)

6 Towards In-context Scene Understanding (Advisor: Sudhanshu Mittal)

Email to: mittal@cs.uni-freiburg.de

Question 1 What is purpose of using a memory bank in the model?

Question 2 What are the advantages of the proposed method over previous self-supervised methods like MoCo or DINO? (List 2)

Question 3 What are the major drawbacks of the proposed method? (List 2)

7 Leveraging Unpaired Data for Vision-Language Generative Models via Cycle Consistency (Advisor: Silvio Galesso)

Email to: galessos@cs.uni-freiburg.de

Question 1 What is the purpose of the "stop gradient" operations visible in the T2I2T and I2T2I pipelines in Figure 3? ($\sim$ 1 sentence)

Question 2 The training objective for the Text-to-Image pipeline is categorical (i.e. Cross Entropy). Does this choice seem natural to you? What do you think is the reason for it? ($\sim$ 3 sentences)

Question 3 What are the main characteristics and differences of the datasets used for training the model? ($\sim$ 3 sentences)

8 Single-Stage Diffusion NeRF: A Unified Approach to 3D Generation and Reconstruction (Advisor: Philipp Schröppel)

Email to: schroepp@cs.uni-freiburg.de

Question 1 Previous approaches for diffusion-based 3D generation are trained in two stages. What are these two stages and what problems arise from the two-stage training? (3 sentences)

Question 2 How does the proposed approach achieve single-stage training? (2-3 sentences)

Question 3 The proposed approach trains an unconditional diffusion model. How can this model be used for both unconditional generation and reconstruction? (2-3 sentences)

9 DynIBaR: Neural Dynamic Image-Based Rendering (Advisor: Philipp Schröppel)

Email to: schroepp@cs.uni-freiburg.de

Question 1 The DynIBaR paper builds on IRBNet, which in turn is related to NeRF. Explain the difference between IRBNet and NeRF in terms of network architecture and the usage of given images for rendering/optimization. ($\sim$ 3 sentences)

Question 2 Explain on a high level why IRBNet can not deal with dynamic objects and how this is resolved in DynIBaR. (2-3 sentences)

Question 3 Give a short overview of the different steps in the training pipeline of DynIBaR. ($\sim$ 3 sentences)

10 TAPIR: Tracking Any Point with per-frame Initialization and temporal Refinement (Advisor: Johannes Dienert)

Email to: dienertj@cs.uni-freiburg.de

Question 1 What are the three core desigin decisions that define TAPIR? ($\sim$ 3 sentences)

Question 2 Given a RGB video with 50 frames and a resolution of 512x256 pixels. What dimensions will its feature map F have?

Question 3 How is thresholding used to suppress spurious matches? ($\sim$ 2 sentences)