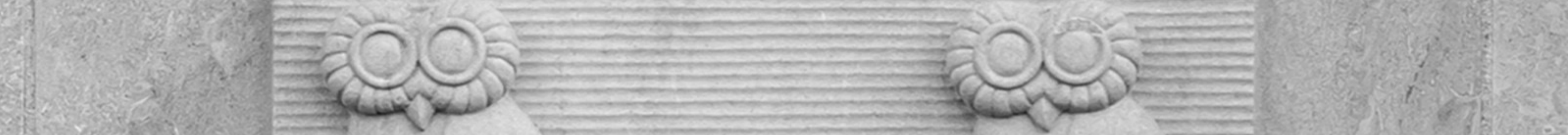
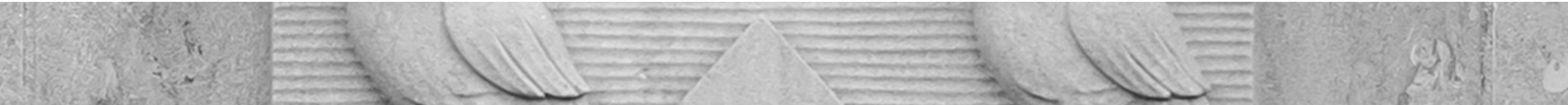


Computer Vision Seminar

Welcome and Introduction

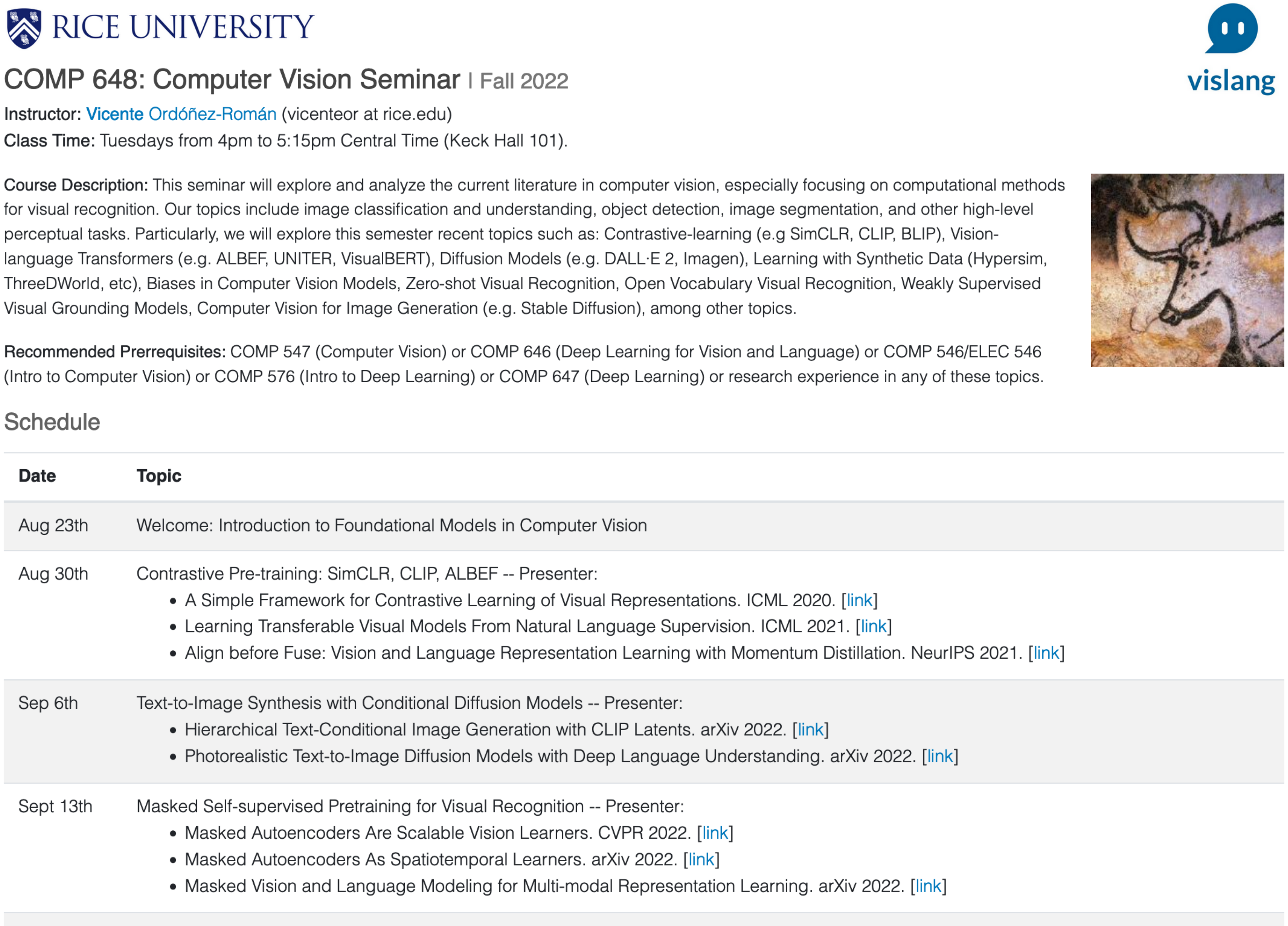
RICE UNIVERSITY



About the class

- COMP 648: Computer Vision Seminar
- Instructor: **Vicente** Ordóñez (Vicente Ordóñez Román)
- Website: https://www.cs.rice.edu/~vo9/cv-seminar
- Location: Zoom – Keck Hall 101
- Times: Tuesdays from 4pm to 5:15pm Central Time
- Office Hours: TBD
- Discussion Forum: Piazza – Sign up below

  https://piazza.com/rice/fall2022/comp648 https://www.cs.rice.edu/~vo9/cv-seminar



RICE UNIVERSITY

vislang

COMP 648: Computer Vision Seminar | Fall 2022

Instructor: Vicente Ordóñez-Román (vicenteor at rice.edu)
Class Time: Tuesdays from 4pm to 5:15pm Central Time (Keck Hall 101).

Course Description: This seminar will explore and analyze the current literature in computer vision, especially focusing on computational methods for visual recognition. Our topics include image classification and understanding, object detection, image segmentation, and other high-level perceptual tasks. Particularly, we will explore this semester recent topics such as: Contrastive-learning (e.g SimCLR, CLIP, BLIP), Vision-language Transformers (e.g. ALBEF, UNITER, VisualBERT), Diffusion Models (e.g. DALL·E 2, Imagen), Learning with Synthetic Data (Hypersim, ThreeDWorld, etc), Biases in Computer Vision Models, Zero-shot Visual Recognition, Open Vocabulary Visual Recognition, Weakly Supervised Visual Grounding Models, Computer Vision for Image Generation (e.g. Stable Diffusion), among other topics.

Recommended Prerrequisites: COMP 547 (Computer Vision) or COMP 646 (Deep Learning for Vision and Language) or COMP 546/ELEC 546 (Intro to Computer Vision) or COMP 576 (Intro to Deep Learning) or COMP 647 (Deep Learning) or research experience in any of these topics.

Schedule

Date	Topic
Aug 23th	Welcome: Introduction to Foundational Models in Computer Vision
Aug 30th	Contrastive Pre-training: SimCLR, CLIP, ALBEF -- Presenter: • A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020. [link] • Learning Transferable Visual Models From Natural Language Supervision. ICML 2021. [link] • Align before Fuse: Vision and Language Representation Learning with Momentum Distillation. NeurIPS 2021. [link]
Sep 6th	Text-to-Image Synthesis with Conditional Diffusion Models -- Presenter: • Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv 2022. [link] • Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv 2022. [link]
Sept 13th	Masked Self-supervised Pretraining for Visual Recognition -- Presenter: • Masked Autoencoders Are Scalable Vision Learners. CVPR 2022. [link] • Masked Autoencoders As Spatiotemporal Learners. arXiv 2022. [link] • Masked Vision and Language Modeling for Multi-modal Representation Learning. arXiv 2022. [link]

My Background

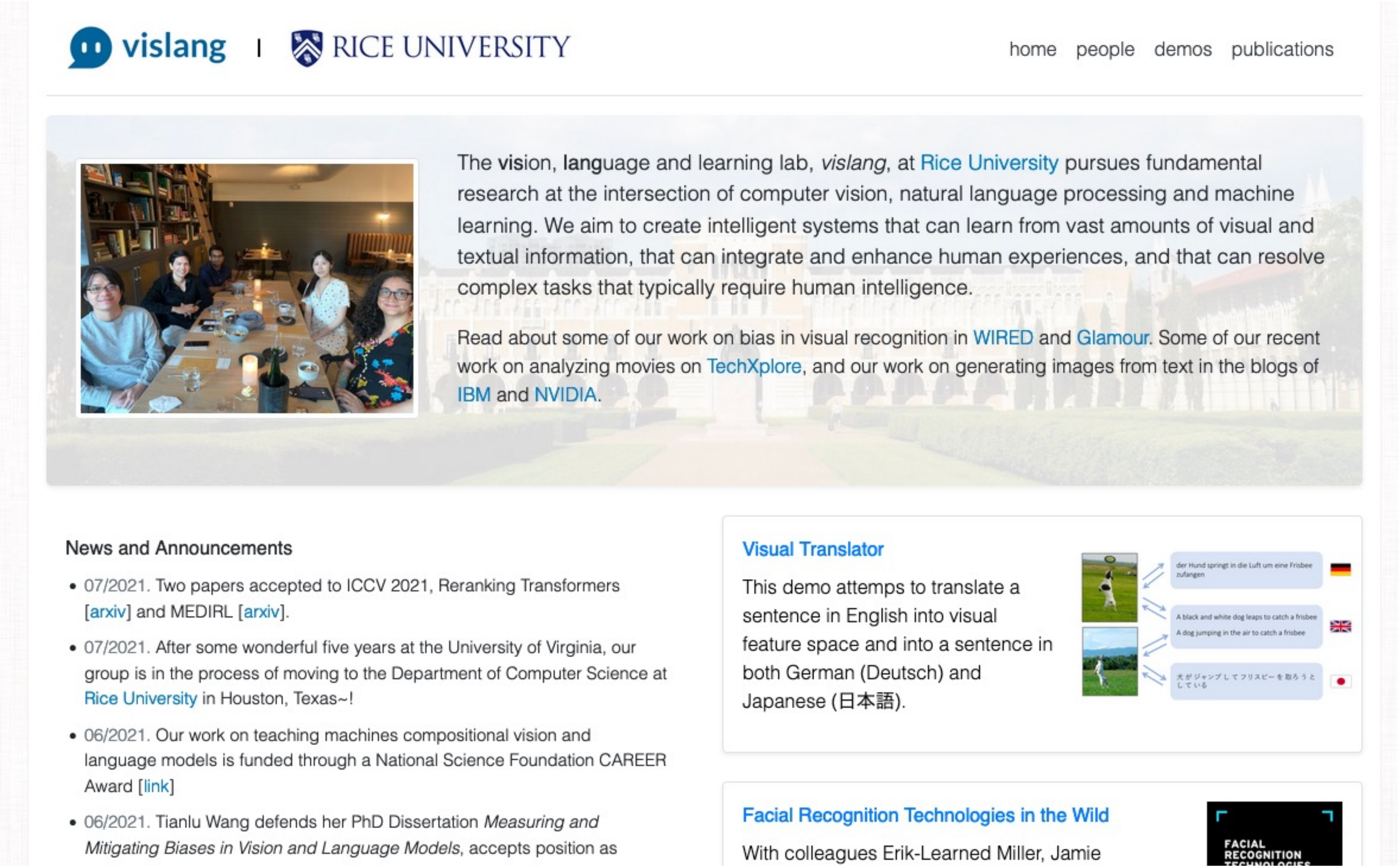
Associate Professor, 2021 - Present	RICE UNIVERSITY
Visiting Academic 2021 - Present	amazon alexa
Assistant Professor, 2016 - 2021	UNIVERSITY of VIRGINIA
Visiting Professor, 2019	Adobe Adobe Research
Visiting Researcher, 2015 - 2016	AI2 ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE
MS, PhD in CS, 2009-2015	THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL Stony Brook University

… also spent time at:

Google Microsoft ebay

vision, language and learning

http://vislang.ai



The **vision**, **lang**uage and learning lab, *vislang*, at Rice University pursues fundamental research at the intersection of computer vision, natural language processing and machine learning. We aim to create intelligent systems that can learn from vast amounts of visual and textual information, that can integrate and enhance human experiences, and that can resolve complex tasks that typically require human intelligence.

Read about some of our work on bias in visual recognition in WIRED and Glamour. Some of our recent work on analyzing movies on TechXplore, and our work on generating images from text in the blogs of IBM and NVIDIA.

News and Announcements

- 07/2021. Two papers accepted to ICCV 2021, Reranking Transformers [arxiv] and MEDIRL [arxiv].
- 07/2021. After some wonderful five years at the University of Virginia, our group is in the process of moving to the Department of Computer Science at Rice University in Houston, Texas~!
- 06/2021. Our work on teaching machines compositional vision and language models is funded through a National Science Foundation CAREER Award [link]
- 06/2021. Tianlu Wang defends her PhD Dissertation *Measuring and Mitigating Biases in Vision and Language Models*, accepts position as

Visual Translator

This demo attemps to translate a sentence in English into visual feature space and into a sentence in both German (Deutsch) and Japanese (日本語).

der Hund springt in die Luft um eine Frisbee zufangen
A black and white dog leaps to catch a frisbee
A dog jumping in the air to catch a frisbee
犬がジャンプしてフリスビーを取ろうとしている

Facial Recognition Technologies in the Wild

With colleagues Erik-Learned Miller, Jamie

Pre-requisites for this Seminar

At least introductory knowledge of Deep Learning: Convolutional Neural Networks, Recurrent Neural Networks, Transformers, Generative Adversarial Networks.

COMP 547 (Computer Vision) or COMP 646 (Deep Learning for Vision and Language) or COMP 546/ELEC 546 (Intro to Computer Vision) or COMP 576 (Intro to Deep Learning) or COMP 647 (Deep Learning) or research experience in any of these topics.

COMP 646: Deep Learning for Vision and Language

- Computer Vision: Image Processing, Image Filtering
- Deep Neural Networks, Multi-layer Perceptrons
- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)
- Transformers – Deep Muti-head Soft Attention Layers
- Generative Adversarial Networks (GANs)

- Optimization: Learning Rates, Learning Rate Schedules, Momentum, Stochastic Gradient Descent, Mini-batch SGD, Sampling, Data Augmentation, Regularization, Overfitting/underfitting.

COMP 648: Recent Advances (2020/2021/2022/2023)

- Contrastive Pretraining (e.g. CLIP and CLIP-derived models)
- Self-supervised Pretraining (e.g. Masked Modeling for Images)
- Diffusion Models (e.g. DALLE-2 – how are they replacing GANs)
- Deep Matching (e.g. SuperGLUE, Reranking Transformers)
- Pretraining for Visual Grounding/Localization (e.g. GLIP, OwL-ViT)
- Universal Deep Models (e.g. DeepMind's Flamingo, GATO)
- Bias and Fairness concerns in Deep Learning and ML
- Other Recent Topics of Interest…

How to best take advantage of this seminar

- Read the papers in advance. They will be posted on the website. Especially if not familiar with a topic.
- Ask questions in advance about papers that are going to be discussed. Use the discussion forum for this class. (Sign up here: https://piazza.com/rice/fall2022/comp648)

Grading

- Satisfactory/Unsatisfactory

- Satisfactory as long as you present a paper at least once throughout the semester and participate actively in discussions (soft attendance). E.g. aim for attending at least 10 out of the 14 sessions of the seminar. However, **stay home if you're sick** overrides any other concern. I'm doing my part today but I should also be resting.

Questions?

Next session: Ziyan will be presenting

Please review the following papers

- A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020. [link]
- Learning Transferable Visual Models From Natural Language Supervision. ICML 2021. [link]
- Align before Fuse: Vision and Language Representation Learning with Momentum Distillation. NeurIPS 2021. [link]