

Many Perception Tasks are Highly Redundant Functions of their Data

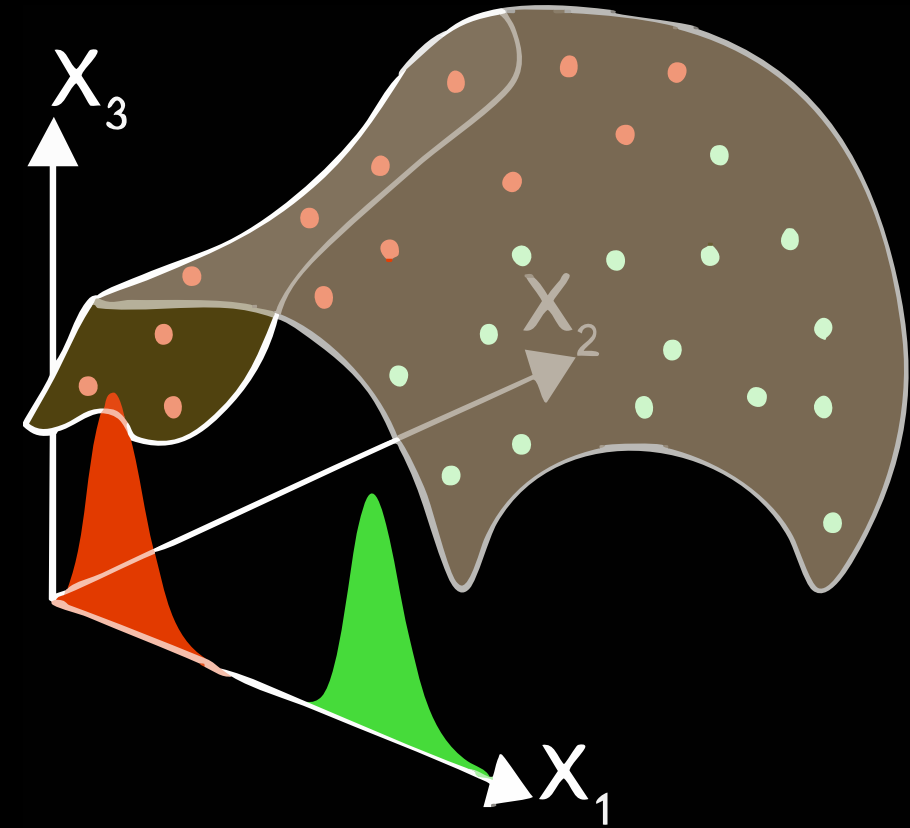
Anthony Bisulco

Self-Supervised Learning: The
Final Frontier of AI workshop

04/29/2025

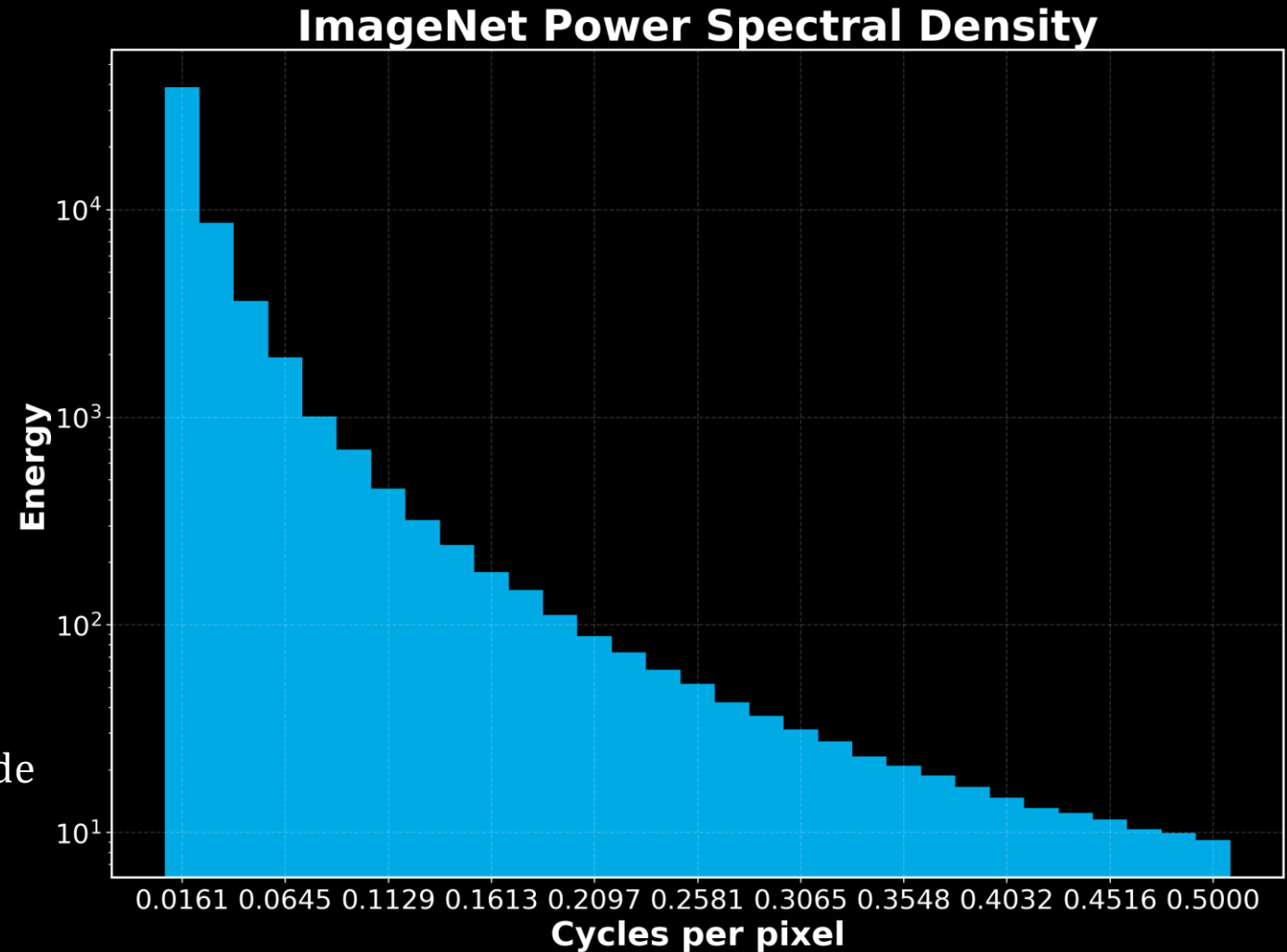
Task Information & Representation Objectives

- Two common SSL techniques:
 - **Contrastive** learning: build invariant representation to sets of augmentations
 - **Reconstruction** based: recover input signal or encoding of the input from encoded features
- **Assumption:** learned invariances or features still capture task-relevant information
- Core question: *How is task-relevant information distributed across the data?*

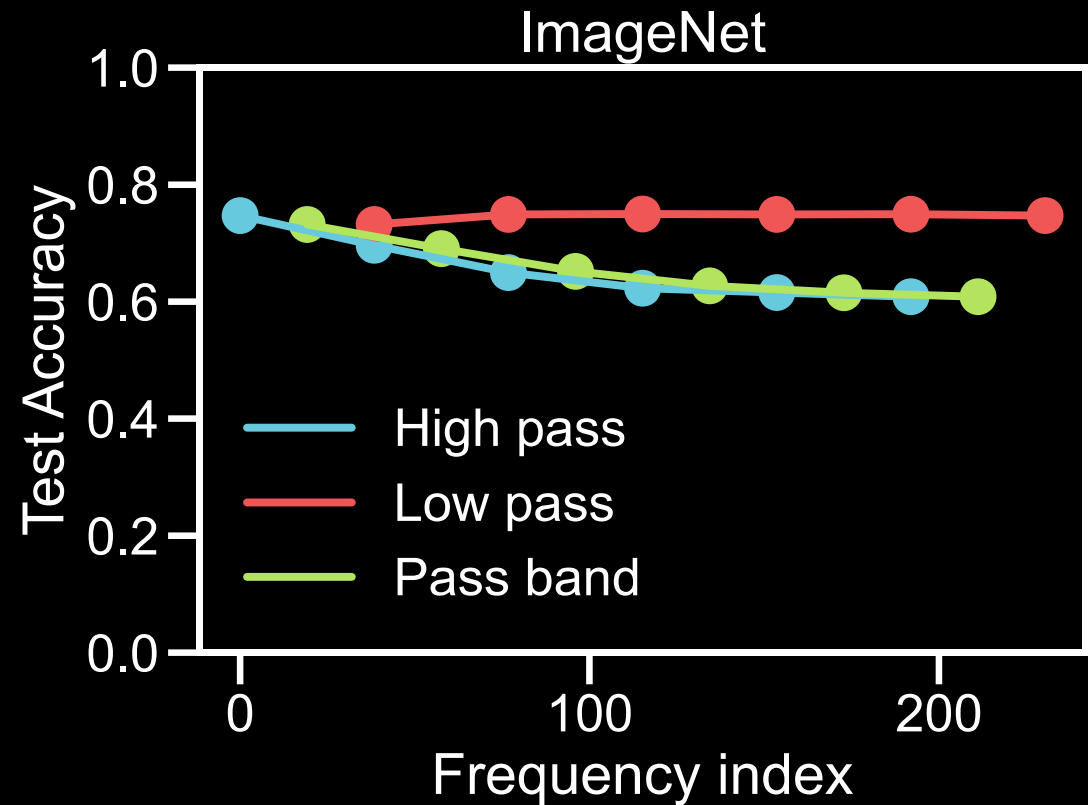
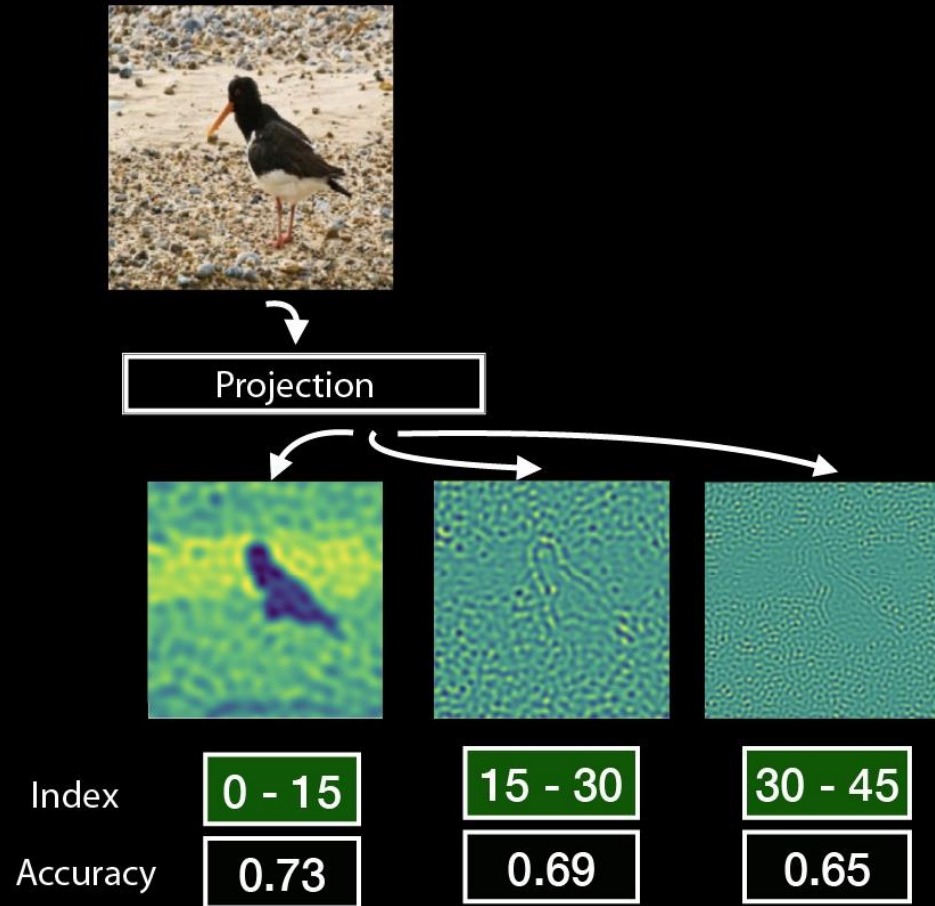


Linear Reconstruction

- Linear reconstruction approaches learn the principal components of data
 - For natural images \rightarrow sinusoids varying frequency ω
- Reconstruction power = Fourier energy (Parseval's Identity)
 - First 5 components (Excluding DC): 99.90%
 - Last 27 components: 0.10%
- ill-conditioned optimization
 - Energy distribution spans 3 orders of magnitude
- Where is the task information in these bands?

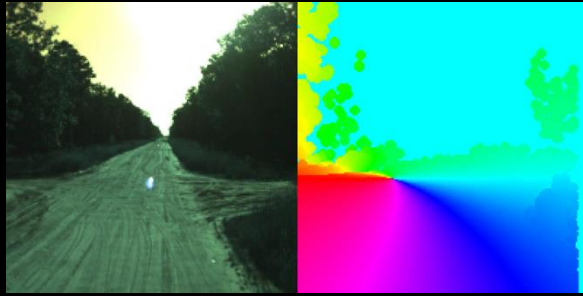


Task Information: Redundantly Distributed Across Subspaces

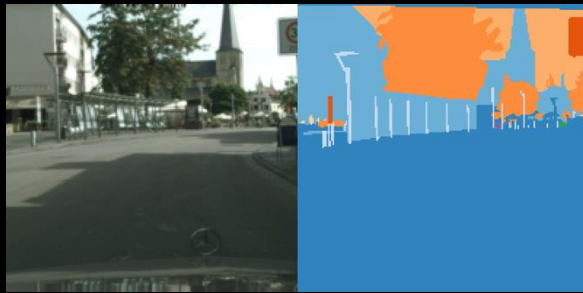


The low-frequency subspace of the input data is most predictive of the task. However, the high-frequency subspace is also remarkably predictive of the task

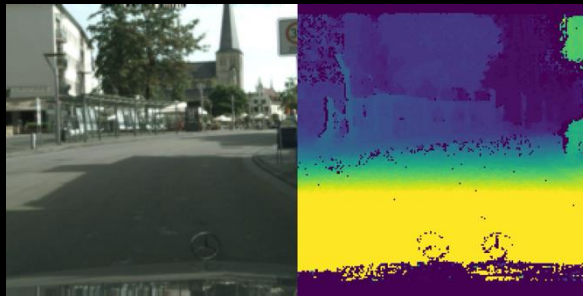
Many Perception Tasks are Redundant Functions of their Data



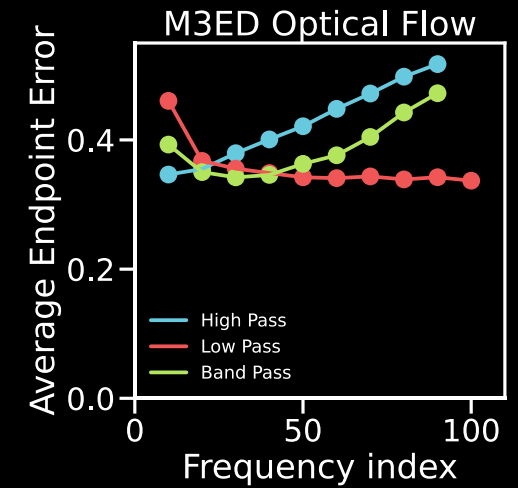
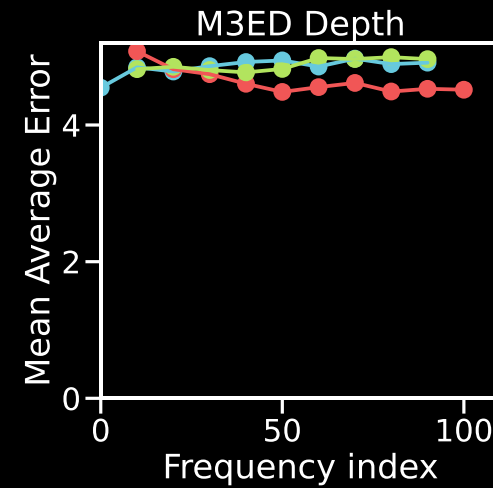
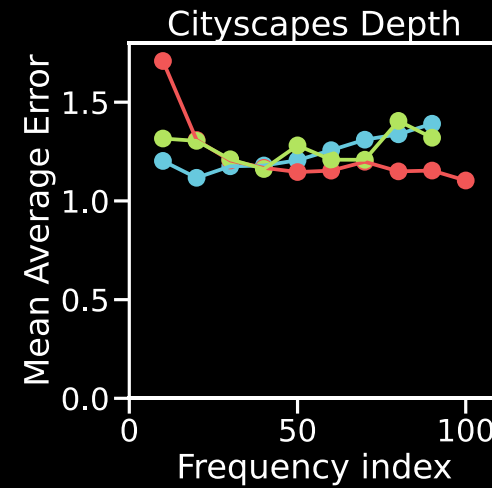
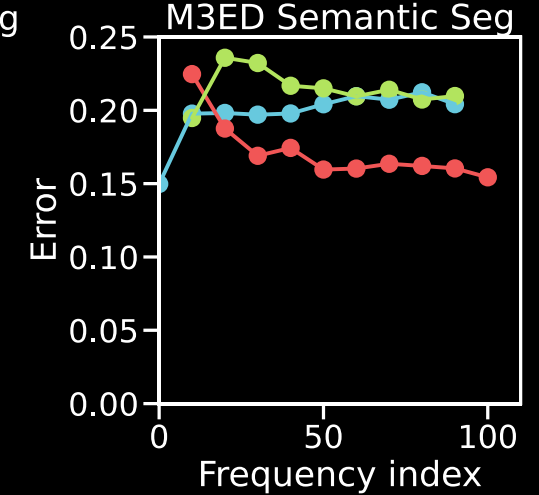
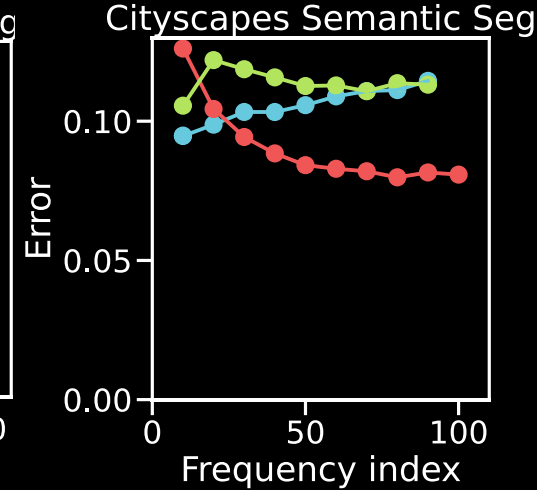
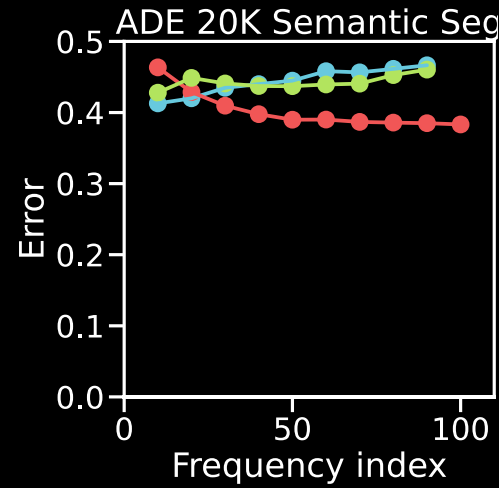
Optical Flow



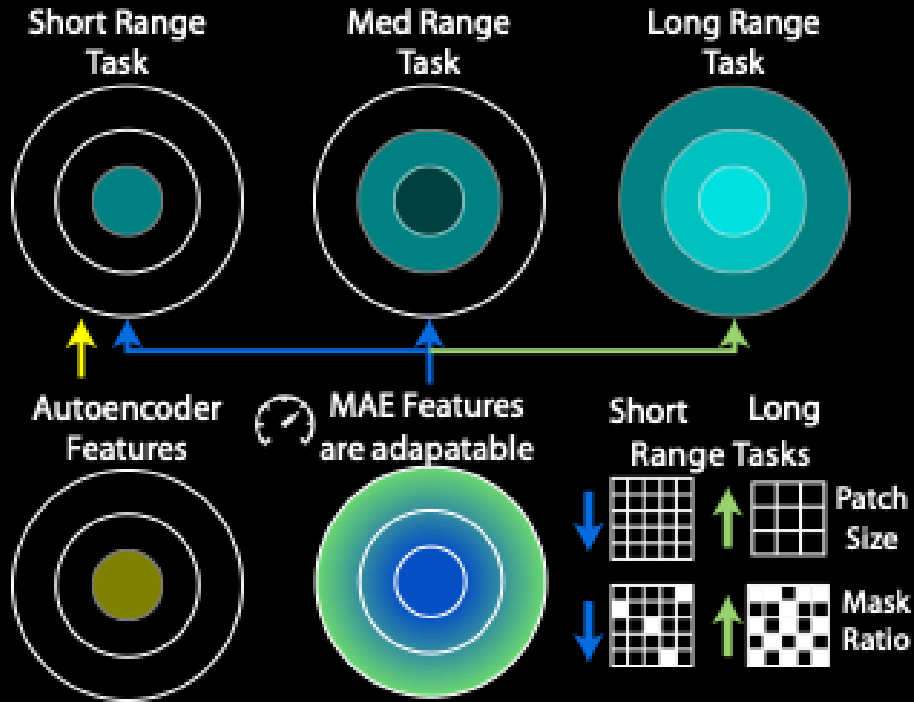
Semantic Segmentation



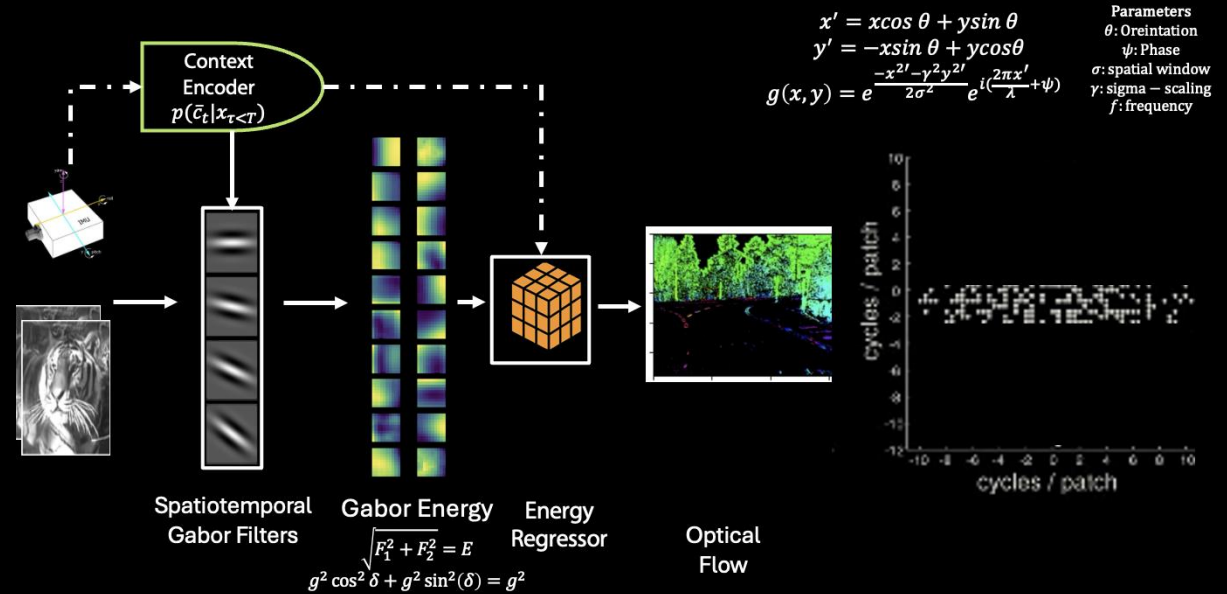
Depth



Current Research Directions

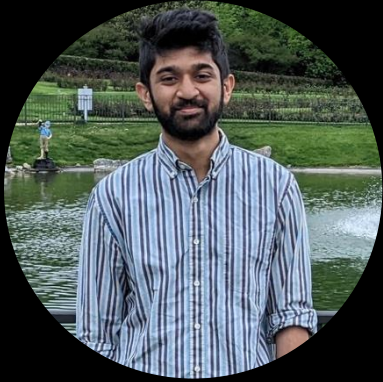


Understanding Masked Autoencoders



Context Modulated Motion Coding

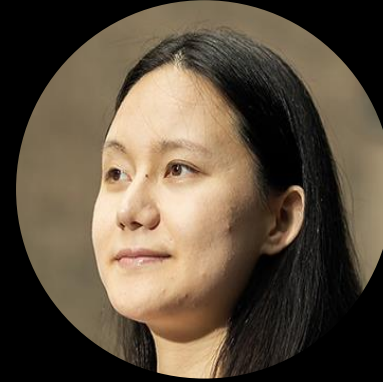
Collaborators



Rahul
Ramesh



Ronald
DiTullio



Linran
Wei



Vijay
Balasubramanian



Kostas
Daniilidis



Pratik
Chaudhari



Randall Balestriero