



DISSERTATION DEFENSE



Mohamed El Banani

Learning Visual
Representations from Cross-
Modal Correspondence

Monday, January 22, 2024

1:30pm – 3:30pm

3725 Beyster

Hybrid – [Zoom](#)

Passcode: 375535

ABSTRACT: One of the goals of computer vision is to develop visual agents that can learn without human annotation. This is typically done by learning from images and their augmentations. In contrast, humans learn from dynamic and multi-sensory environments without requiring such explicit supervision. My dissertation delves into this contrast, exploring how models can learn visual representations directly from their environments. My core observation is that such environments, despite their complexity, present consistent patterns across modalities. These cross-modal patterns offer a rich training signal as we can leverage similarity in one modality for learning generalizable representations in another without requiring additional supervision.

In this dissertation, I argue that cross-modal correspondence provides a rich signal for learning visual representations and a useful tool for analyzing them. I first discuss how models can learn visual representations by finding 3D correspondence in RGB-D videos. Through estimating geometrically consistent correspondences between video frames, models can learn representations that rival supervised models. I then discuss how the notion of correspondence could be applied to language. I propose language-guided self-supervised learning where language models are used to find image pairs that depict similar concepts. I show that using language guidance outperforms self-supervised and language-supervised models; further showcasing the utility of learning from correspondence. Finally, I explore how correspondence can also be used to analyze the 3D awareness and consistency of visual representations learned by large-scale vision models. My analysis suggests that while current approaches yield good models for semantics and localization, their 3D awareness remains limited.

CHAIR: Prof. Justin Johnson