New Work to Expand Upon

Adam Pickeral

Vision Transformers

Vision Transformers (ViT) are a novel concept in Computer Vision that use a method called “self-attention” to provide more accurate results on computer vision tasks than previous CNN based methods.

Some good intro articles to explain ViTs

First

Second

A well cited paper that describes the use of ViTs to increase accuracy.

ViTs can be implemented in Pytorch, as described by this article. This is something to consider when using our system because the nature of ViTs allow them to find strong relationships with image-class, regardless of where certain features are in a image (it has global attention), something that is not possibl with CNNs (they call CNNs “translation invariant”). This is important because desktops have many similar features, but these features aren’t always in the same area. e.g., some people use their desktop task bars on different sides of their screen, as opposed to at the bottom of the screen, which is standard.