Progress

- RepNet baseline training
- Differentiable patch selection
- Experimenting with optical flow
RepNet

$V \xrightarrow{\phi} \tilde{X} \xrightarrow{\phi} X$  
$X = \phi(V)$  
$\tilde{X}$  
512  
64  
Per Frame Embeddings  
Temporal Self-similarity Matrix  
$S$  
35  
32  
5x3x1  
32 filters  
Transformer  
Per frame Multi-headed Self-attention  
515  
64  
Period Length Predictor  
2 FC layers  
Per Frame Period Length  
Periodicity Predictor  
2 FC layers  
Per Frame Periodicity Prediction  
Period Predictor  

Video Frames  
Video  
Encoder
Countix

- Collected Youtube clips of repetitive actions
- 8757 videos total
- Labeled with repetition count
- Largest non-synthetic action counting dataset currently
Patch Selection

- Sparsely sample frames throughout the video
- Linearly interpolate between important patches from sampled frames
Differentiable Patch Selection

- Cordonnier, Mahendran, Dosovitskiy, Weissenborn, Uszkoreit, Unterthiner
- Argmax alone is not differentiable
- Adds Gaussian noise to the inputs and empirically calculates expected value of argmax

\[ Y_\sigma = \mathbb{E}_Z \left[ \arg \max_{Y \in C} \langle Y, \eta + \sigma Z \rangle \right] \]
Optical Flow
Next Steps

- Finish running RepNet baseline
- Calculate optical flow for Countix videos
- Brainstorm more potential patch selection ideas