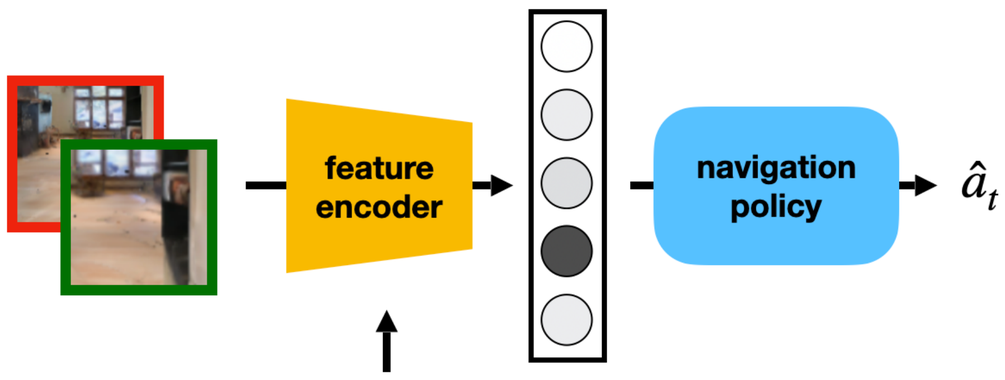
Visual Pre-training for Navigation: What Can We Learn from Noise?

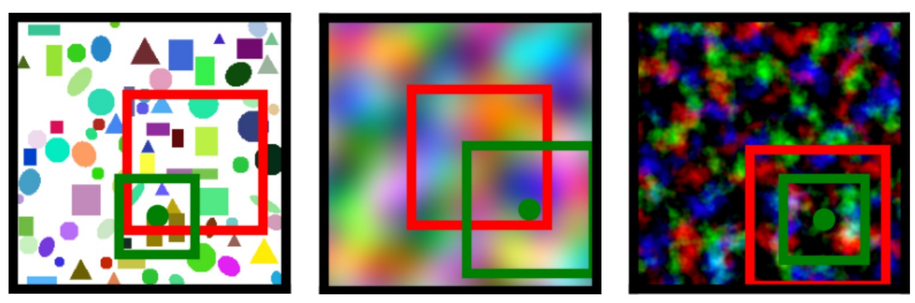
Felix Yanwei Wang, Ching-Yun Ko, Pulkit Agrawal

What makes self-supervised visual navigation hard? - Expensive data collection!



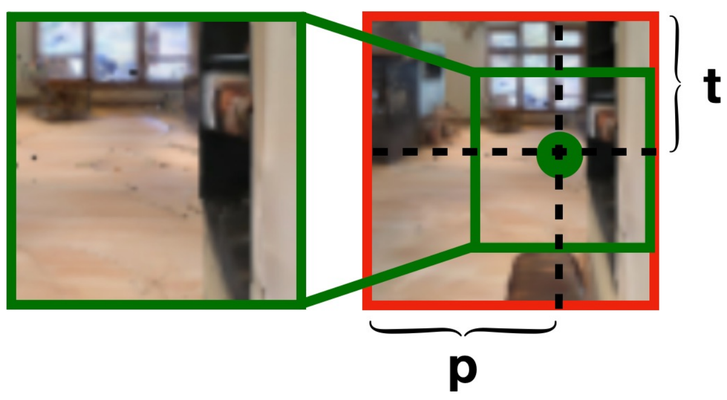
wasteful to train with interaction data

Can we learn a feature encoder from noise?

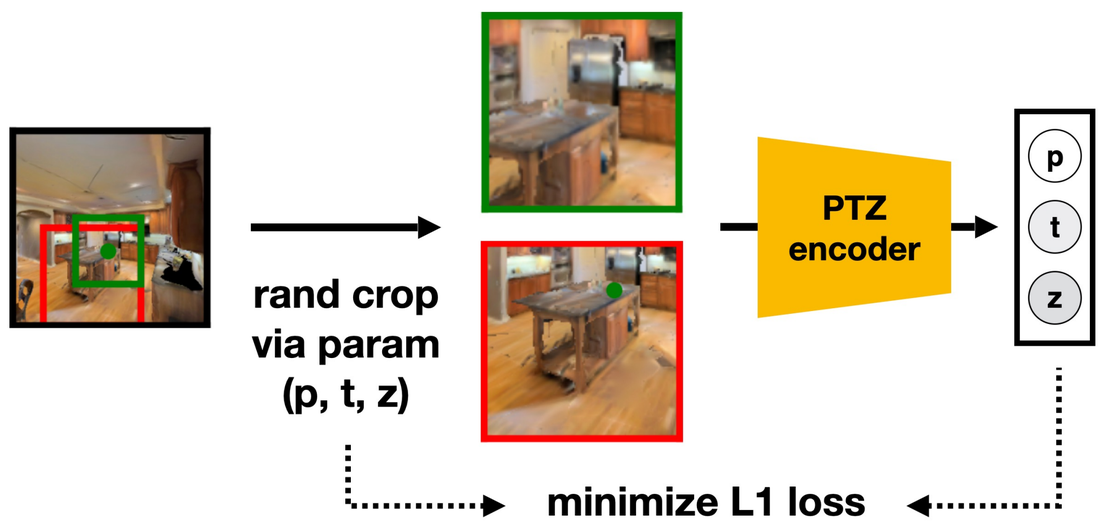


random shapes **Perlin noise** **fractal noise**

Learning PTZ representations from random crops prediction



use (**P**an, **T**ilt, **Z**oom) factors to represent the relative transformation of two views

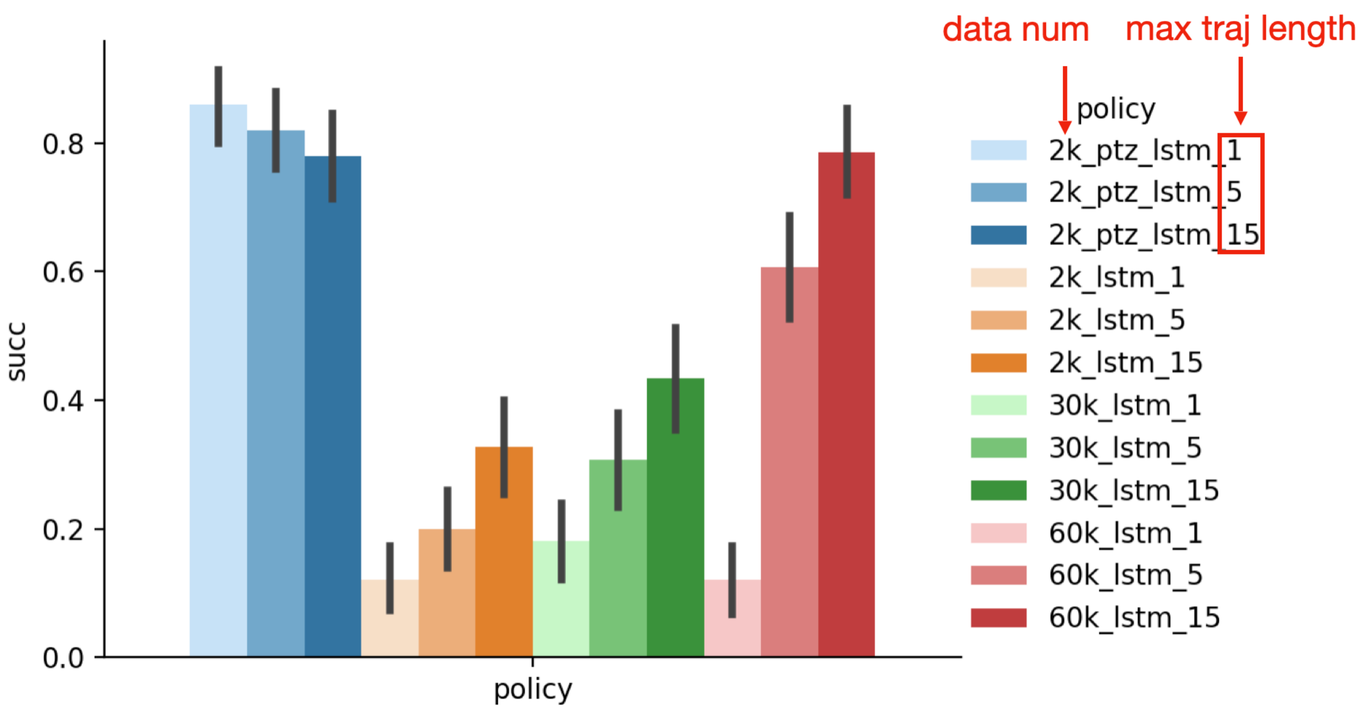


Train and test on static home images - IOU **96.4%**

Train on noises and test on home images - IOU **92.4%**

Does PTZ-encoder improve data efficiency?

Goal reaching rate in 5 environments, 30 trajs/env



PTZ-enabled navigation policy uses **30x** less interaction data

Which noise type transfers the best?

IOU on home domain (testing)

training domain	mean	std
home images	**96.4%**	**0.1%**
all noise	**92.4%**	**0.1%**
Perlin	61.6%	0.4%
fractal	**87.3%**	**0.5%**
shape	72.1%	0.4%

Training random crops prediction on synthetic noises **transfers** to real images