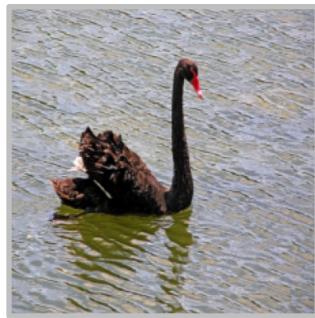


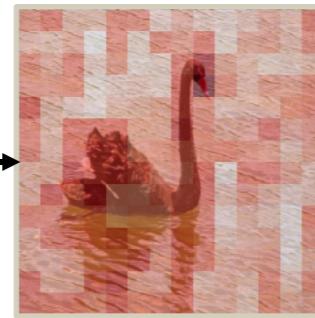
■ Pretraining

Large image dataset



Base Vision Transformer (BVT)

Unified Representation



Vision Expert Library

Teacher-Specific Router

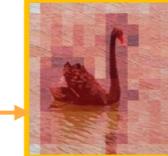


Experts



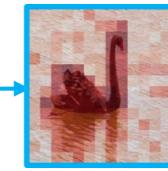
$\times N$

Student Representations

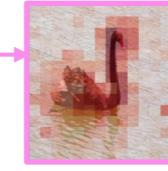


Distillation

DINOv2



ViT



CLIP

Trainable

Frozen

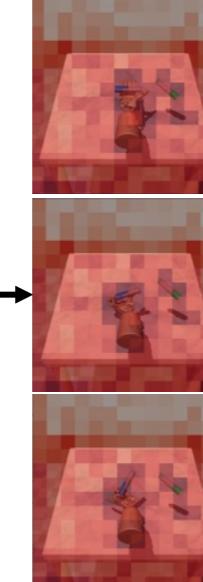
■ Downstream Robot Tasks

Previous images in robot dataset



BVT

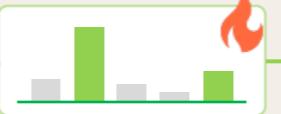
Unified Representation



Vision Expert Library

Frozen Experts

Trainable Robot Router
<0.4% Params



Structured Representation



Policy Head



Actions

Inactive vision experts
Active vision experts