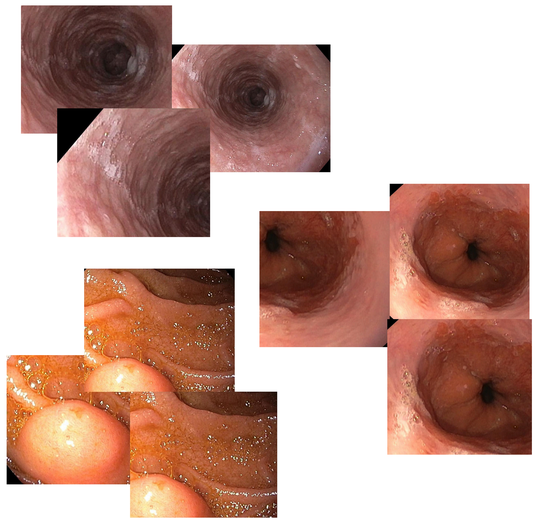
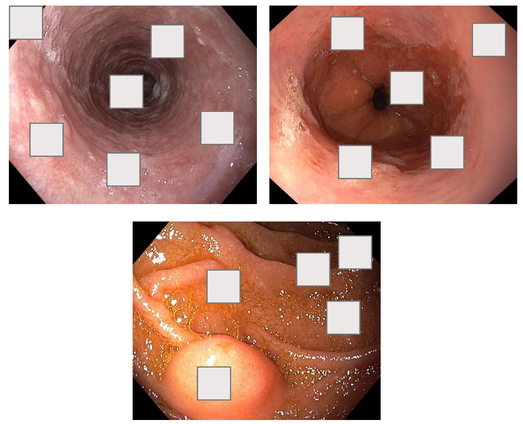
Stage 1: Self-Supervised Pre-training

UpperGI-400K
(unlabeled)



Multicrops
(image-level objective)



MIM
(patch-level objective)

ViT Image Encoder

DINO/iBOT Head

Student

EMA

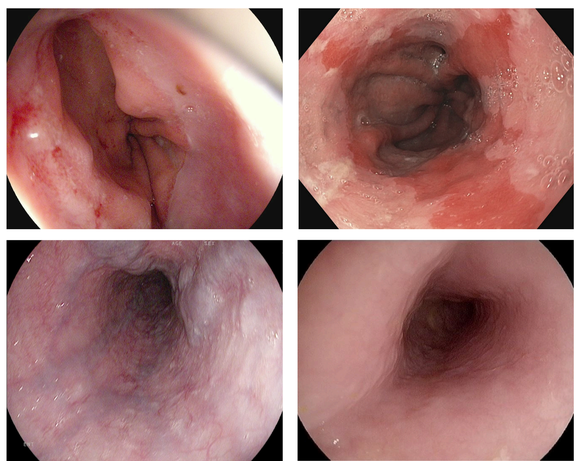
ViT Image Encoder

DINO/iBOT Head

Teacher

Stage 2: Supervised Downstream-training

Esophagitis
Dataset (labeled)



Scaling

Flipping

Rotation

Oversampling

ViT Image Encoder

Linear Head

Esophagitis

Other