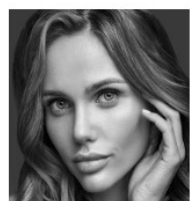


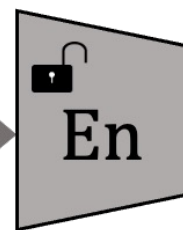
## Uni-modal Pre-training



...



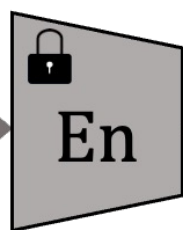
Data in the wild



Pretext Tasks

## Downstream-specific fine-tuning

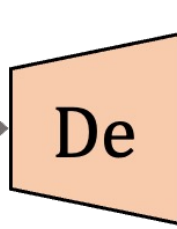
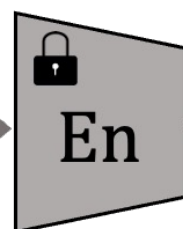
Domain  
Dataset 1



*Task 1*

Image Classification

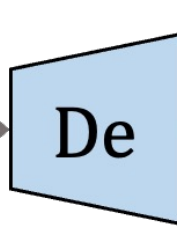
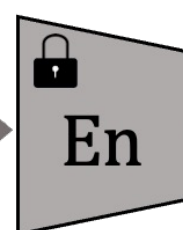
Domain  
Dataset 2



*Task 2*

Object Detection

Domain  
Dataset 3



*Task T*

Semantic Segmantion

## Prediction on multiple tasks

Dataset 1

Fine-tuned model 1

*Task 1*

⋮

⋮

⋮

Dataset T

Fine-tuned model T

*Task T*