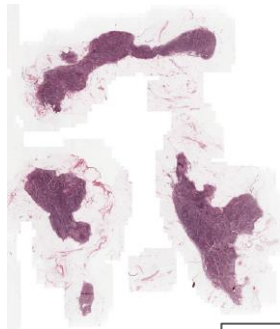
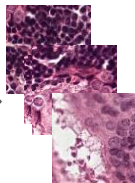


Processing of Whole Slide Image

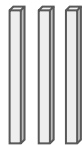


PATCH
EXTRACTION



FEATURE
EXTRACTION

$[N, \text{dim}]$



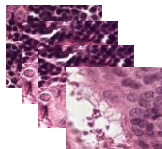
TOKEN
MIL



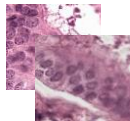
PREDICTION

$224 \times 224 \times 3$

FEATURE EXTRACTOR



RESNET-18



CONTRASTIVE LEARNING

FEATURES



$[N, \text{dim}]$

MAX POOLING

LAYERNORM

TOKEN
ATTENTION

TOKEN
GENERATION

$[N, K = 128]$

$[K, \text{dim}]$

TOKEN MIL

\mathcal{L}_{ins}

\mathcal{L}_{bag}