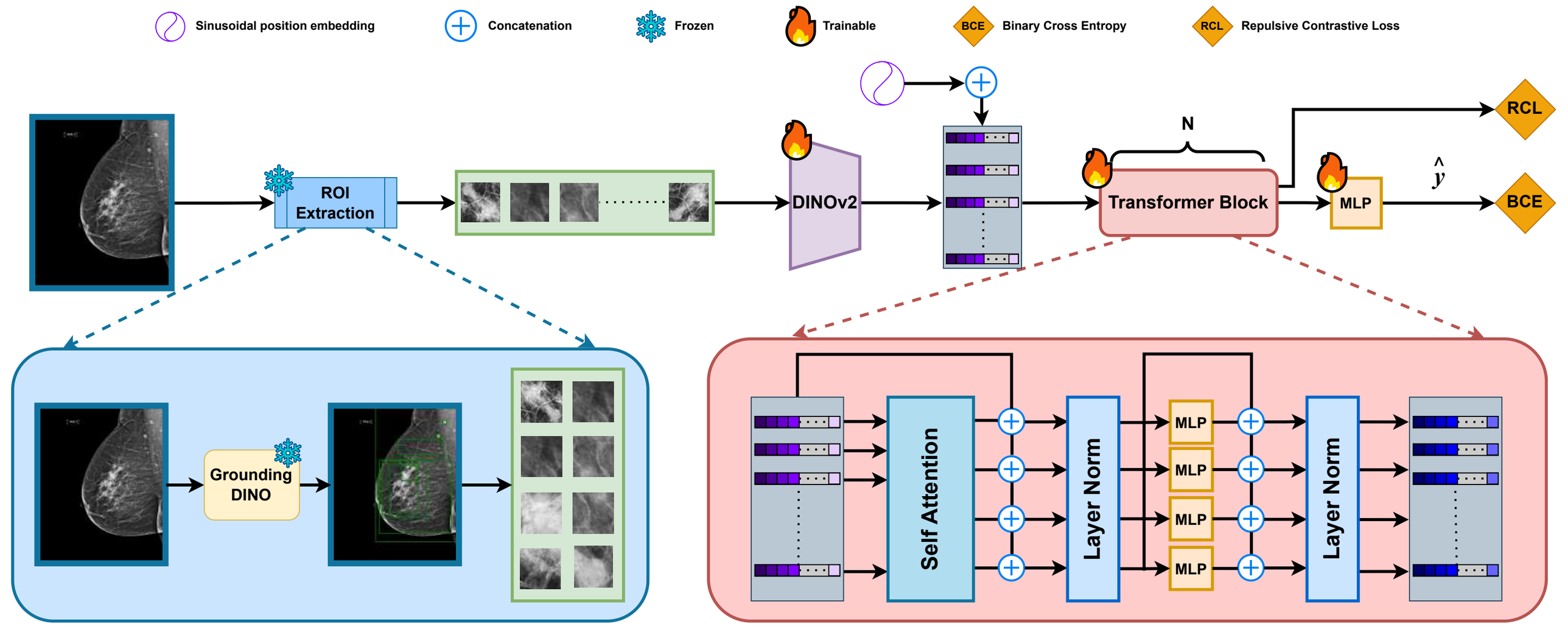
Sinusoidal position embedding
Concatenation
Frozen
Trainable
BCE
Binary Cross Entropy
RCL
Repulsive Contrastive Loss
ROI Extraction
DINOv2
N
Transformer Block
MLP
RCL
BCE
Grounding DINO
Self Attention
Layer Norm
MLP
MLP
MLP
Layer Norm