

Foreground Prompt

cat, dog, horse, wolf ...

a clear origami
a clear origami {*}

ground, tree, sky ...

Background Prompt

Input



CLIP Text Encoder

CLIP Image Encoder

Transformer Block

Transformer Block

Transformer Block

Transformer Block

Softmax

GradCAM



AugCAM

EWM



Refined CAM



🔥 : Trained ❄️ : Frozen

Ⓢ : Compute Distance

Ⓡ : CAM Refinement

⊕ ⊖ : Sum and Difference

⊗ : Multiplication Operation

WT : Wavelet Transform

EWM : Edge Weight Mapping

IWT : Inverse Wavelet Transform

📦 : Feature Map

No Gradient
Backpropagated

🔥 WT-Decoder

Predict