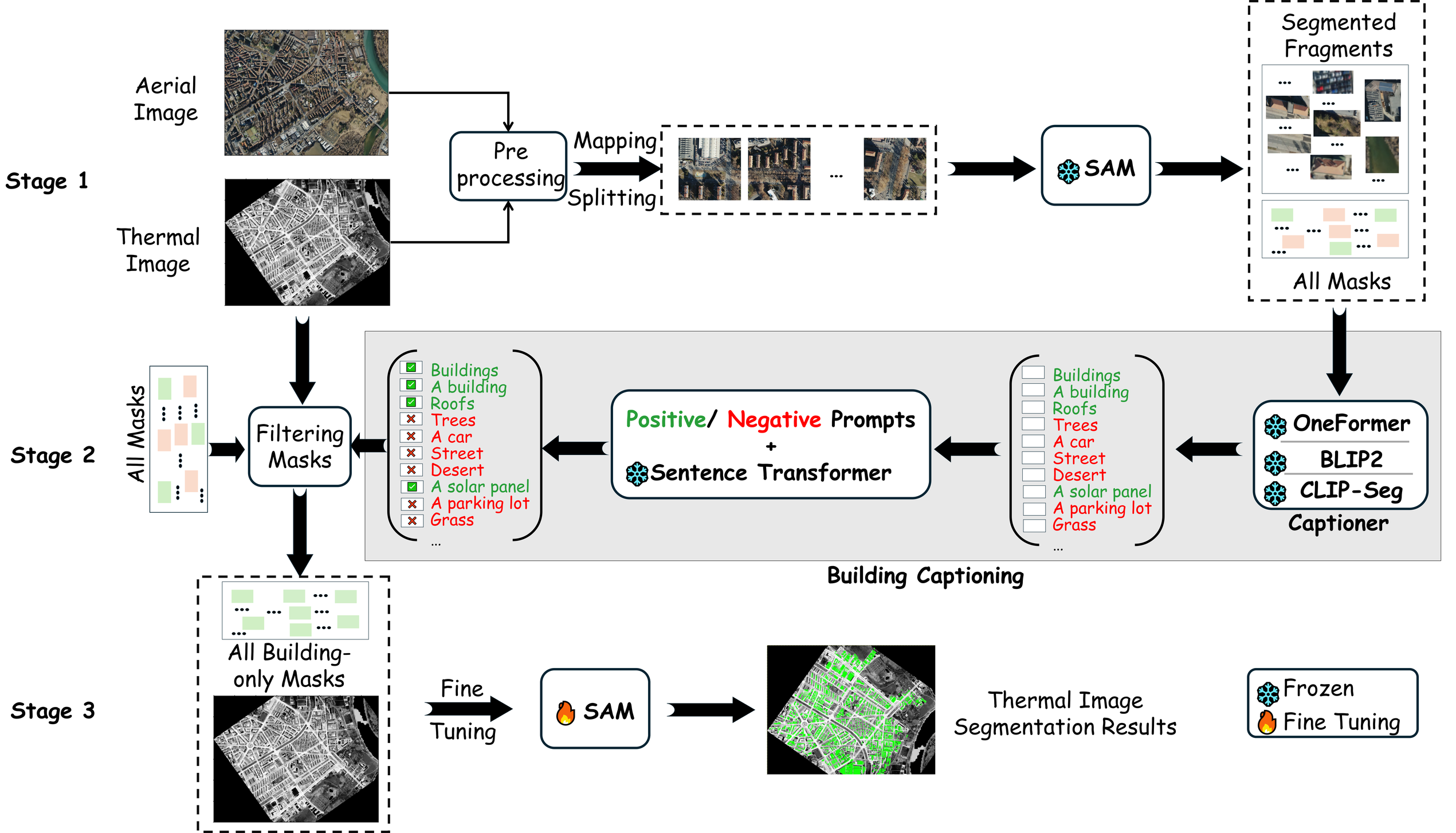
Stage 1
Aerial Image
Thermal Image
Pre processing
Mapping
Splitting
SAM
Segmented Fragments
All Masks
Stage 2
All Masks
Filtering Masks
Buildings
A building
Roofs
Trees
A car
Street
Desert
A solar panel
A parking lot
Grass
...
Positive/ Negative Prompts
+
Sentence Transformer
Buildings
A building
Roofs
Trees
A car
Street
Desert
A solar panel
A parking lot
Grass
...
OneFormer
BLIP2
CLIP-Seg
Captioner
Building Captioning
Stage 3
All Building-only Masks
Fine Tuning
SAM
Thermal Image Segmentation Results
Frozen
Fine Tuning