MaskConver Architecture

Input Image

Backbone

Pixel Decoder

$H_i \times W_i \times C_E$

Center Heatmap Head

$H_i \times W_i \times C_{class}$

Center Embedding Head

$H_i \times W_i \times C_{embd}$

Mask Feature Head

$H/4 \times W/4 \times C_{embd}$

Mask Embedding Generator

K Classes

Top K

Class Embedding Lookup Table

K Class Embeddings

MLP

K Centers

Center Embedding Gather

K Center Embeddings

K Mask Predictions

$K \times H/4 \times W/4$

Panoptic Mask