



Multi-scale Backbone
(e.g., ResNet, SwinT)

Clustered Pixels