Knowledge Distillation

For Semantic Segmentation:

- **Two-stage Pipeline:** Segment-then-classify approach (e.g., ZegFormer, ZSSeg)
- **Direct Pixel-level Distillation:** Match VLM with pixel-wise features (e.g., CLIPSeg, LSeg, MaskCLIP+)
- **Prompt/Descriptor Learning:** Enhance generalization beyond base classes (e.g., ZegCLIP, OVSeg)
- **Weak Supervision with VLM:** Refine CAMs using CLIP (e.g., CLIP-ES, CLIMS)

Goal: Transfer general VLM knowledge to dense prediction tasks while enabling open-vocabulary capability.