Knowledge Distillation

Motivation:

- **Architecture Flexibility:** Distill VLM knowledge into task-specific models without retaining the VLM structure.
- **Representation Gap:** VLMs offer image-level features, while downstream tasks need region/pixel-level understanding.

For Object Detection:

- **Embedding Alignment:** Align detector and VLM features (e.g., ViLD, HierKD, RKD)
- **Prompt-based Distillation:** Learn detection-specific prompts (e.g., DetPro, PromptDet)
- **Pseudo-label Supervision:** Use VLM-generated pseudo boxes/masks (e.g., PB-OVD, XPM, P3OVD)
- **Region Bag Distillation:** Aggregate multiple region embeddings (e.g., BARON, RO-ViT)