Transfer Learning

Motivation:

- **Distribution gap:** Downstream tasks may differ in image styles and text formats.
- **Objective gap:** VLMs are trained with general objectives, while downstream tasks require task-specific objectives (e.g., classification, detection).

Transfer Techniques:

- **Prompt Tuning:** Modifies input text/image with learnable prompts. Includes:
  - Text Prompt Tuning (e.g., CoOp, CoCoOp, DualCoOp, PLOT)
  - Visual Prompt Tuning (e.g., VP, RePrompt)
  - Text-Visual Prompt Tuning (e.g., UPT, MAPLE)
- **Feature Adapter:** Adds lightweight trainable layers after VLM encoders (e.g., CLIP-Adapter, Tip-Adapter, SVL-Adapter)
- **Other Methods:**
  - Direct Fine-tuning (e.g., Wise-FT)
  - Architecture Modification (e.g., MaskCLIP)
  - Cross-modal Attention (e.g., VT-CLIP, CALIP)