Two-Stage Multi-Modal KD System Architecture

Datasets & Dataloaders (data/dataset.py, trainer/engine.py)

- dataset_type (medpix / wound)
  - if wound → WoundDataset (metadata_*.csv, images/)
  - if medpix → MedPixDataset (data_*.jsonl, descriptions_*.jsonl, images/)
- ViTImageProcessor (vit-base, resize+norm)
- get_num_classes() (modality, location)
- Teacher & Student Tokenizers (AutoTokenizer from backbones)
- DataLoader(dev)
- DataLoader(test)
- DataLoader(train)

Config (YAML → cfg dict)

- data:
  - type (medpix/wound)
  - root, batch_size, num_workers
  - task1_label, task2_label
- device:
  - cuda:X or cpu
- teacher:
  - vision, text
  - fusion_dim, fusion_layers
  - fusion_heads, dropout
- loss:
  - type ∈ {vanilla, combined, crd, rkd, mmd}
  - alpha, beta, gamma, T, ...
- fusion:
  - type
  - hidden_mult, p_img, p_txt, dropout, ... (fusion_params)
- student:
  - vision, text
  - fusion_dim, fusion_layers
  - fusion_heads, dropout

Teacher Model (models/teacher.py)

- Vision Backbone (get_vision_backbone)
- Text Backbone (get_text_backbone)
- Projection Layers proj_vis, proj_txt → fusion_dim
- Fusion Module (Simple, ConcatMLP, CrossAttention, Gated, TransformerConcat, ModalityDropout, FiLM, EnergyAwareAdaptive, SHoMR)
- Classification Heads (modality, location)
- Teacher outputs dict:
  - logits_modality
  - logits_location
  - img_raw/txt_raw
  - img_proj/txt_proj

Student Model (models/student.py)

- Text Backbone (smaller)
- Vision Backbone (smaller)
- Projection Layers proj_vis, proj_txt → fusion_dim
- Fusion Module (same interface, own params)
- Classification Heads (modality, location)
- Student outputs dict:
  - logits_modality
  - logits_location
  - img_raw/txt_raw
  - img_proj/txt_proj

KD / Loss Modules (losses/*)

- Supervised CE (teacher or student logits)
- DistillationLoss (vanilla: CE + KL + MSE)
- MedKDCombinedLoss (CE + KL + MSE + CRD)
- CRDLoss
- RKDLoss
- MMDLoss

Trainer (trainer/engine.py)

- main(cfg)
  - parse config
  - build datasets/models
  - select loss
  - run phases
- Phase 1: train_teacher() (supervised teacher training)
- Phase 2: train_student() (student KD + supervised)
- evaluate_detailed() (dev/test metrics)
- MetricsLogger / ResultsLogger

batches: (pv, ids_teacher, mask_teacher, labels)

teacher(pv, ids_t, mask_t) (no grad)

student(pv, ids_s, mask_s)

KD loss backprop