

Cross-Modal Alignment: 모달 간 정렬

공통 임베딩 공간에서 서로 다른 모달리티를 의미적으로 가까운 위치에 배치하여 상호 이해 및 검색 가능



Projection Layers

각 모달리티를 동일 차원의 공통 공간으로 투영하는 선형/비선형 변환

- 선형 레이어로 차원 매칭
- MLP로 비선형 변환
- L2 정규화로 단위 구 상에 배치

Contrastive Loss

매칭 쌍은 가깝게, 비매칭 쌍은 멀게 학습하여 정렬 품질 향상

- InfoNCE Loss 사용
- Temperature scaling 적용
- Hard negative mining

Triplet Loss

Anchor, Positive, Negative 샘플로 상대적 거리 학습

Cross-Attention

Transformer 기반 어텐션으로 모달 간 상호작용 강화

- 동일 환자 데이터: Positive
- 다른 환자 데이터: Negative
- 마진 기반 거리 최적화

- Query-Key-Value 메커니즘
- 모달 간 정보 교환
- 동적 가중치 학습

Zero-shot 분류

텍스트→영상 검색

영상→텍스트 검색