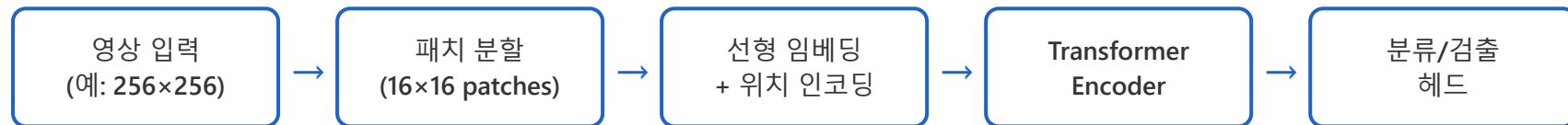


Vision Transformer (ViT) for Medical Imaging

패치 임베딩과 위치 인코딩으로 의료 영상을 처리, **Self-attention**으로 전역 관계 포착



Self-Attention 메커니즘

전체 영상의 장거리 의존성 학습

- 모든 패치 간 관계 계산
- Query, Key, Value 변환
- 병변 간 공간적 관계 포착
- Multi-head attention

의료 영상 적용

ViT의 의료 도메인 특화

- 고해상도 영상 처리
- 가변 크기 입력 지원
- 3D 볼륨 확장 (3D-ViT)
- Medical-specific 사전학습

장점

ViT의 강점

- 전역 컨텍스트 이해
- 확장 가능한 구조
- 대규모 사전학습 효과
- 다양한 다운스트림 태스크

의료 ViT 모델

대표적인 의료 특화 ViT

- MedViT: 의료 영상 사전학습
- TransUNet: 세분화 특화
- Swin Transformer: 계층적
- CoTr: CT/MRI 복원