

Structured Pruning

구조적 가지치기 (Structured Pruning)

전체 채널, 필터, 또는 레이어를 제거하여 구조적으로 압축

구조적 가지치기 유형

Channel Pruning

[■][■][□][■]
채널 단위 제거

Filter Pruning

Filter 1: ■■■■
Filter 2: □□□□
Filter 3: ■■■■

Layer Pruning

Layer 1 → ■
Layer 2 → □
Layer 3 → ■

구조적 가지치기의 장점

✓ 하드웨어 친화적: 특수 하드웨어 불필요

✓ 실제 속도 향상: dense matrix 연산 유지

✓ 메모리 감소: 실제 파라미터 수 감소

의료 응용: X-ray 분류 모델에서 40% 채널 제거 → 2배 빠른 추론, 정확도 98.5% → 97.8% (1% 미만 감소)

