

Feature Distillation

특징 수준 증류 (Feature-Level Distillation)

Teacher와 Student의 중간층(intermediate layers) 출력을 매칭

레이어 매칭 전략

1:1 Matching

Teacher Layer 12 → Student Layer 6

Multiple Matching

여러 레이어 동시 매칭

Adaptive Matching

학습 가능한 변환 사용

Feature Distillation Loss

$$L_{\text{feature}} = ||f_T(x) - f_S(x)||^2$$

f_T : Teacher features, f_S : Student features