Feature Distillation

Feature-Level Distillation

Matching intermediate layers outputs of Teacher and Student

Layer Matching Strategy

1:1 Matching	**Multiple Matching**	**Adaptive Matching**
Teacher Layer 12 → Student Layer 6	Matching multiple layers simultaneously	Using learnable transformations

Feature Distillation Loss

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

f_T: Teacher features, f_S: Student features