

Teacher-Student Framework



Teacher Model

Size: Large (millions to billions of parameters)

Performance: Highest accuracy

Purpose: Knowledge provision

Examples: ResNet-152, BERT-Large



Student Model

Size: Small (10-100x smaller)

Performance: Approaches teacher

Purpose: Practical deployment

Examples: MobileNet, DistilBERT

Knowledge Transfer



Transferred Knowledge



Soft Targets
(Probability distributions)

Feature Maps
(Intermediate layer outputs)

Attention Maps
(Focus regions)