Lab5 – Knowledge Distillation

Advisor：Tsai, Chia-Chi

TA：柯秉鈞

Contact：course.aislab@gmail.com

Outline

1. Knowledge Distillation
   - Introduction
   - Types of Knowledge
   - Modes of Distillation
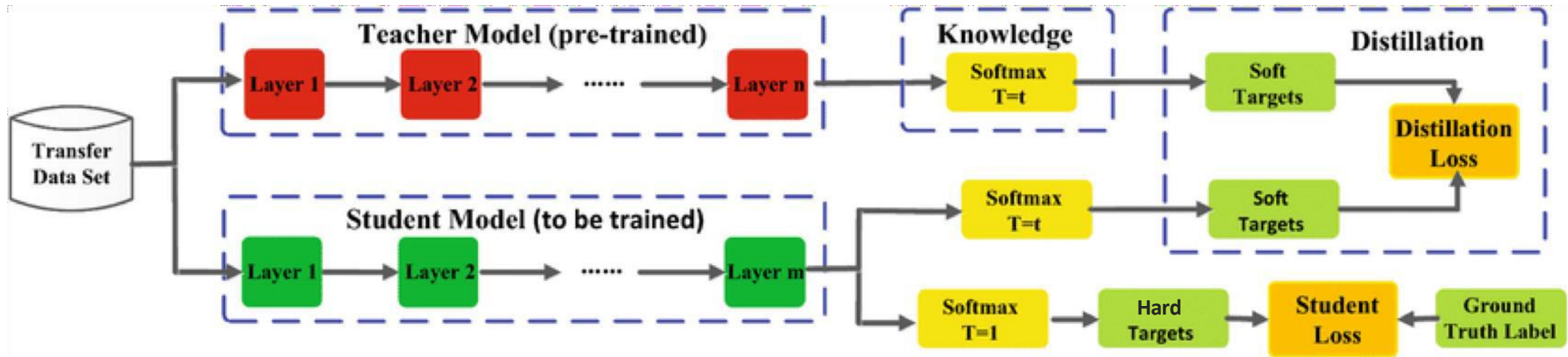
2. Lab5 task

Outline

1. Knowledge Distillation
   - Introduction
   - Types of Knowledge
   - Modes of Distillation

2. Lab5 task

Knowledge Distillation

- A small model (student) is trained to mimic a large pre-trained model (teacher)

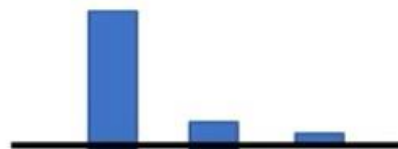$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}$$

Temperature

Knowledge Distillation

- Temperature for softmax

$$y_i' = \frac{exp(y_i)}{\sum_j exp(y_j)} \quad \Longrightarrow \quad y_i' = \frac{exp(y_i/T)}{\sum_j exp(y_j/T)}$$
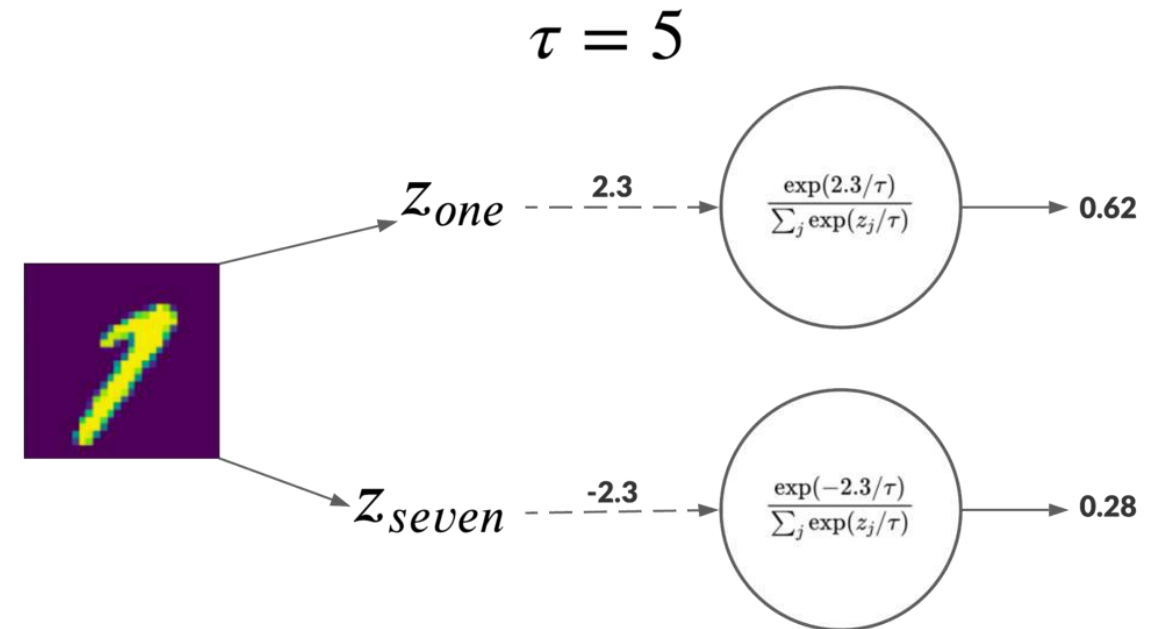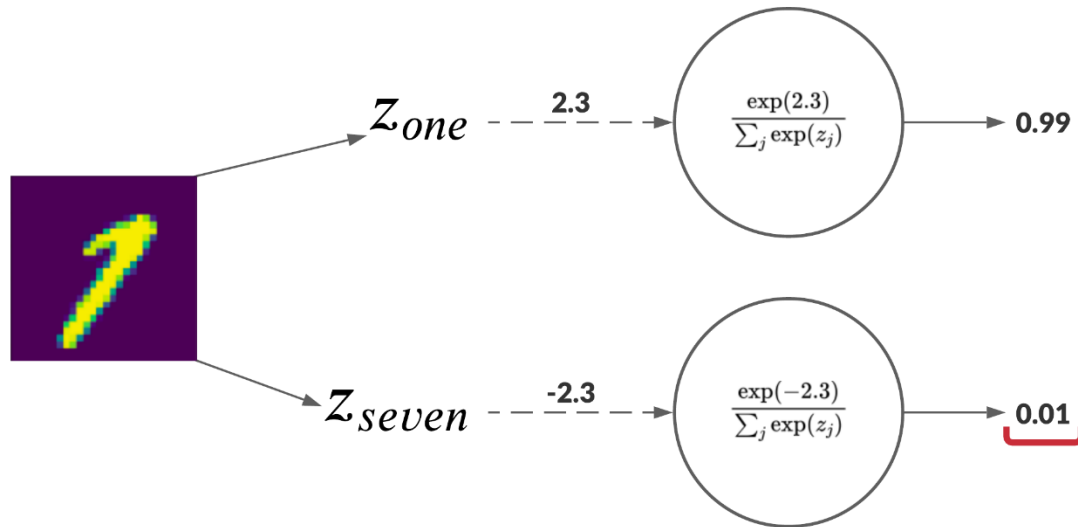
$$T = 100$$

$$y_1 = 100 \qquad y_1' = 1$$
$$y_2 = 10 \qquad y_2' \approx 0$$
$$y_3 = 1 \qquad y_3' \approx 0$$

$$y_1/T = 1 \qquad y_1' = 0.56$$
$$y_2/T = 0.1 \qquad y_2' = 0.23$$
$$y_3/T = 0.01 \qquad y_3' = 0.21$$

Soft Target

- Model prediction softened by T

$$\frac{\exp(2.3)}{\sum_j \exp(z_j)} \rightarrow 0.99$$

$$\frac{\exp(-2.3)}{\sum_j \exp(z_j)} \rightarrow 0.01$$

$$\tau = 5$$

$$\frac{\exp(2.3/\tau)}{\sum_j \exp(z_j/\tau)} \rightarrow 0.62$$

$$\frac{\exp(-2.3/\tau)}{\sum_j \exp(z_j/\tau)} \rightarrow 0.28$$
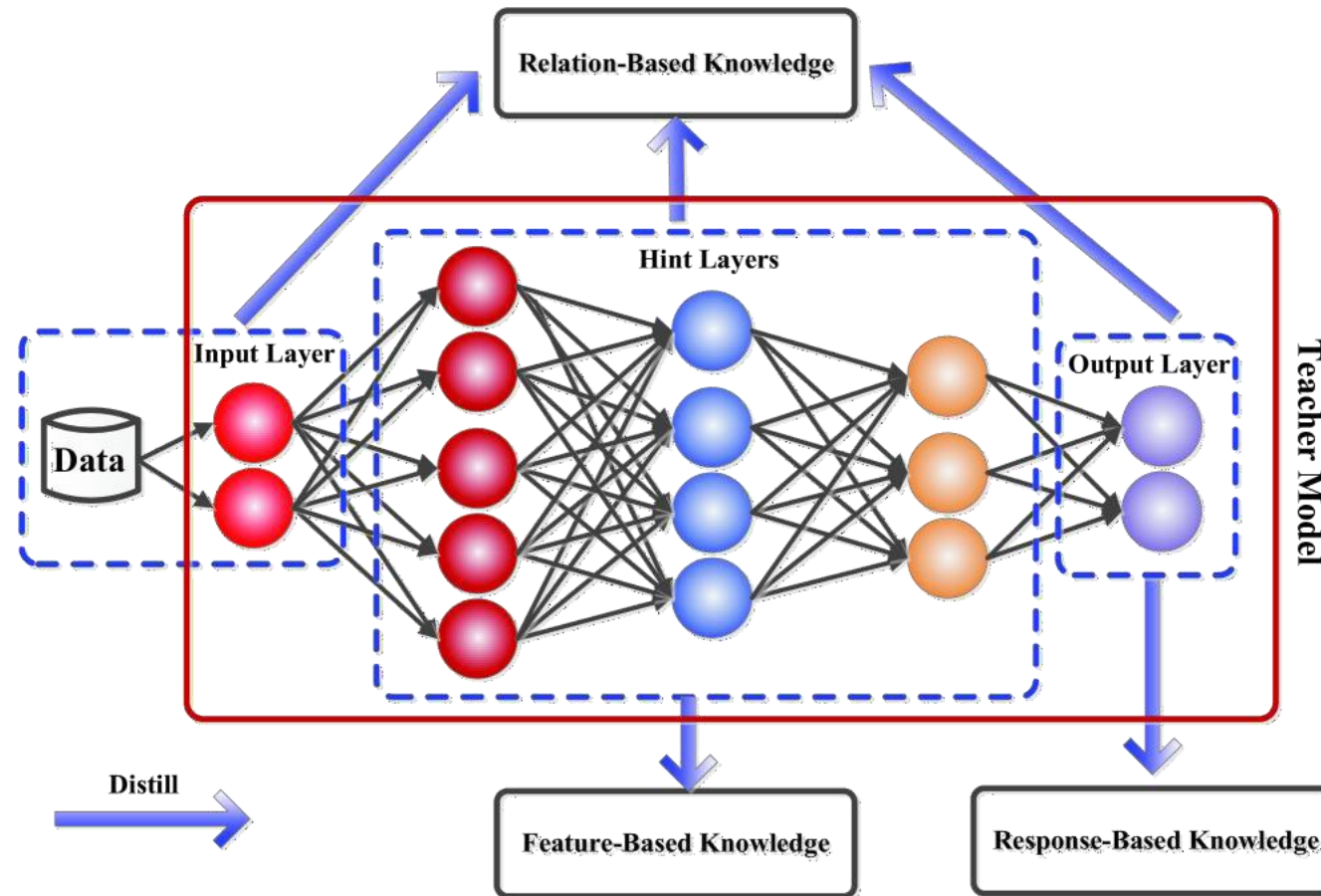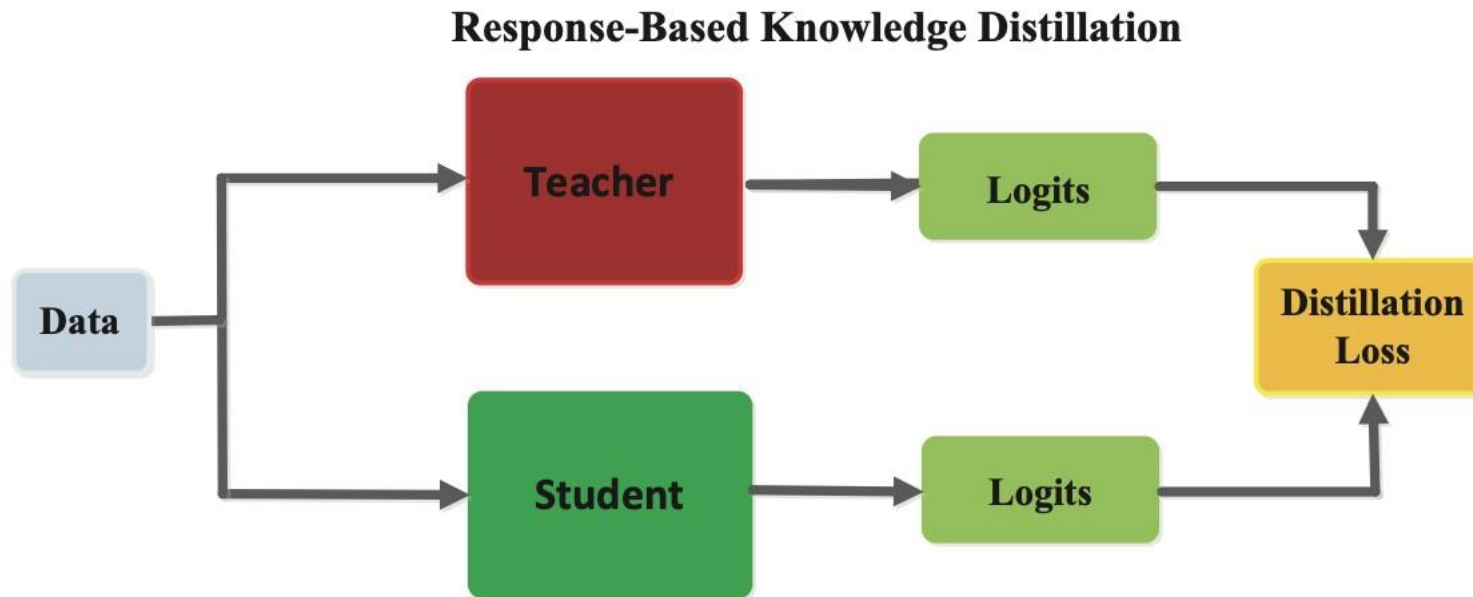
Outline

1. Knowledge Distillation
   - Introduction
   - **Types of Knowledge**
   - Modes of Distillation

2. Lab5 task

Types of Knowledge

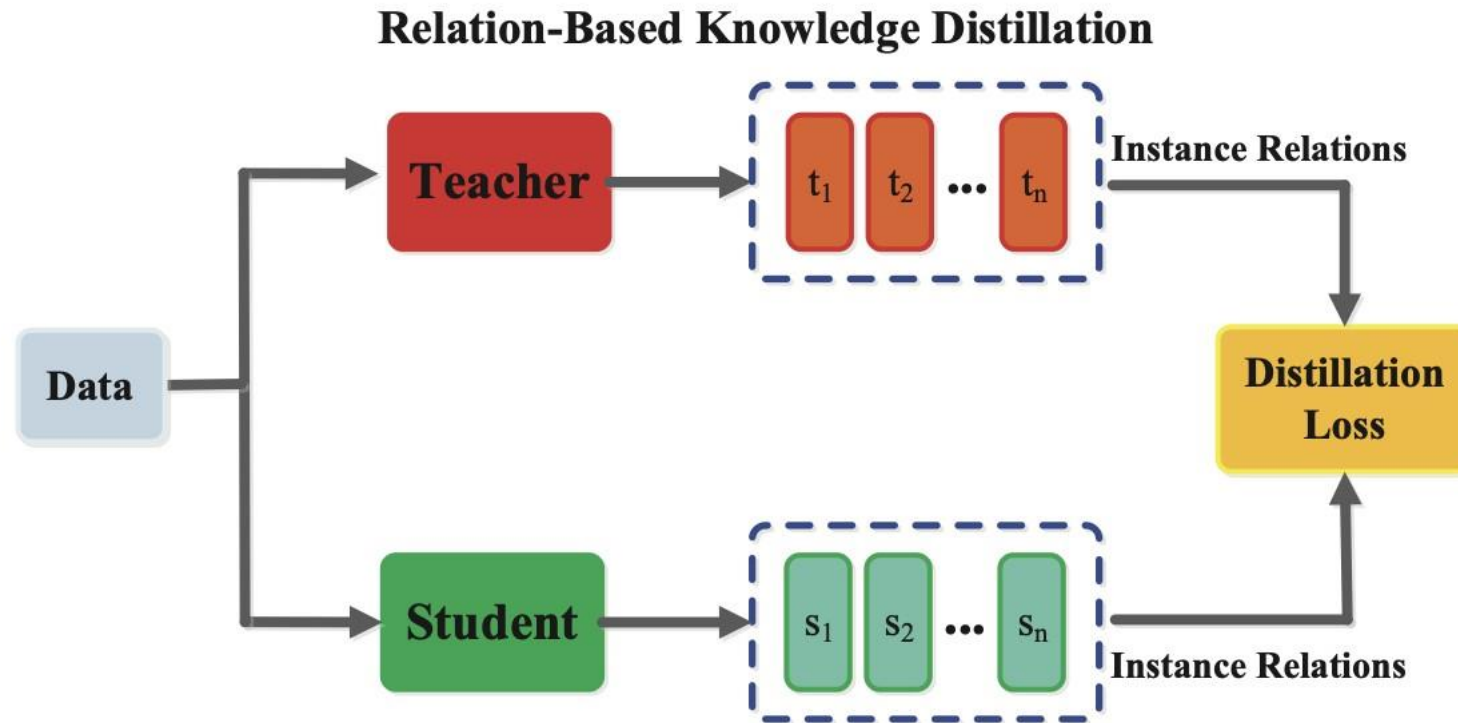# Response-Based Knowledge

Response-Based Knowledge Distillation

Feature-Based Knowledge

Feature-Based Knowledge Distillation
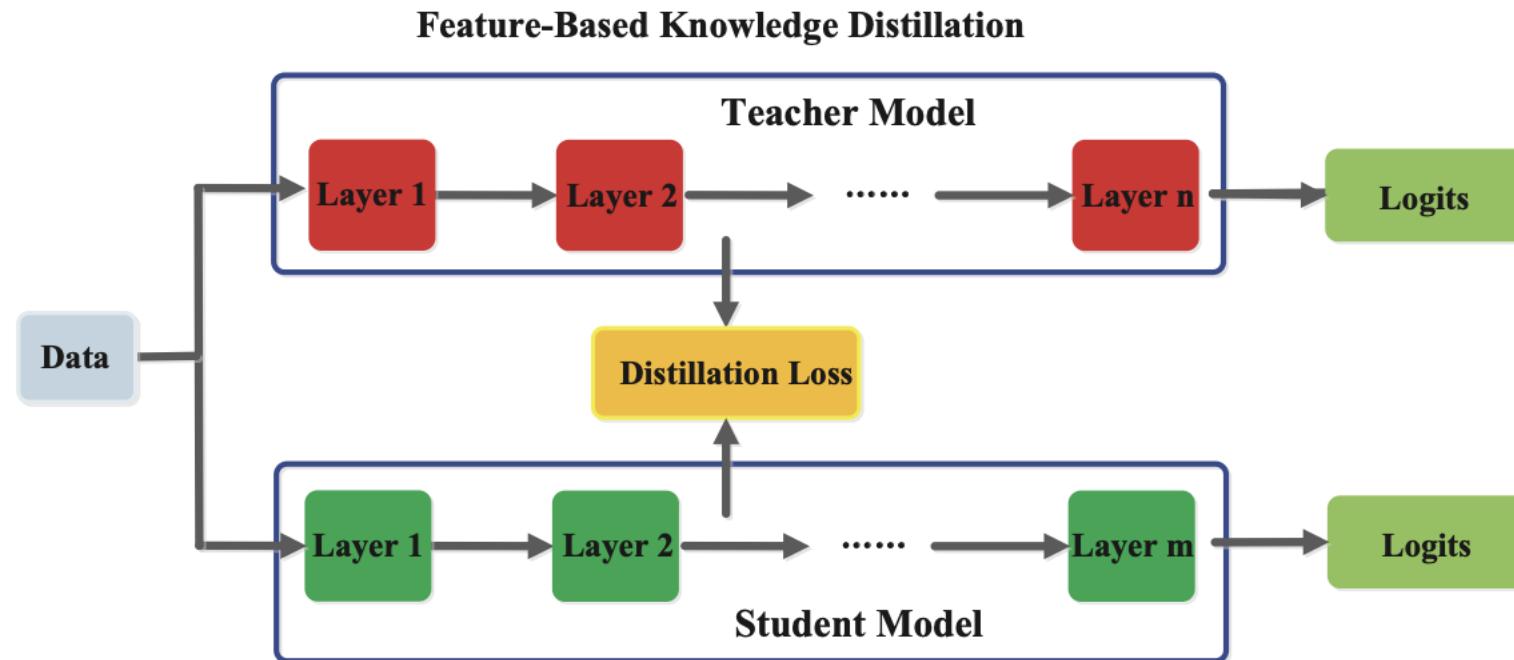
Relation-Based Knowledge

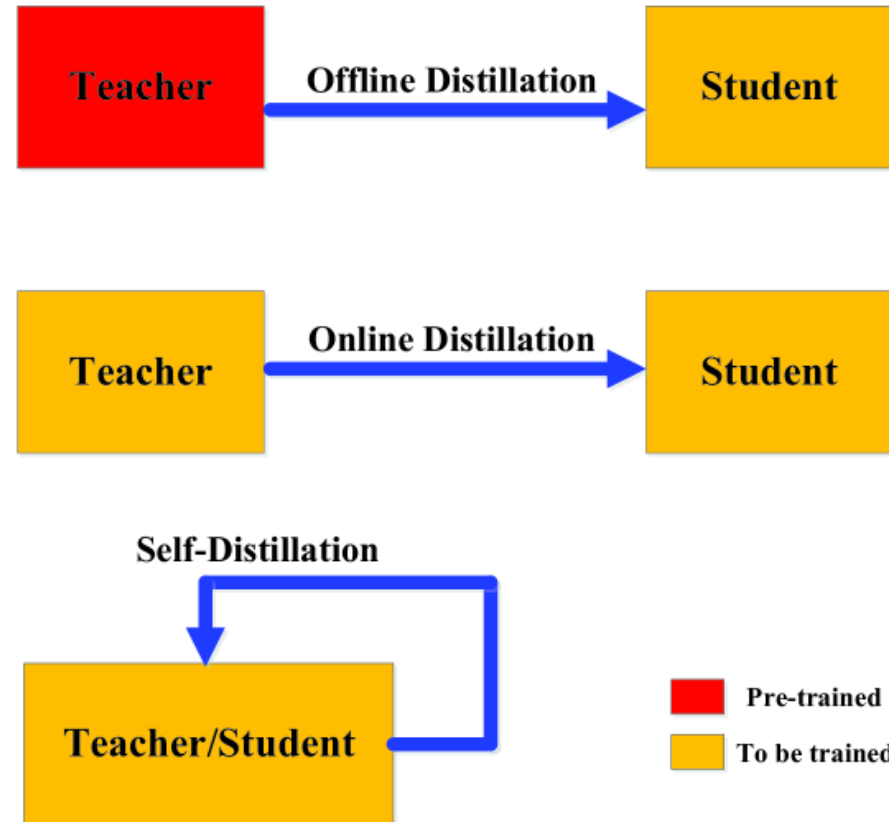**Relation-Based Knowledge Distillation**

Outline

1. Knowledge Distillation
   - Introduction
   - Types of Knowledge
   - **Modes of Distillation**
2. Lab5 task

Modes of Distillation

Outline

1. Knowledge Distillation
   - Introduction
   - Types of Knowledge
   - Modes of Distillation

2. Lab5 task

Lab5 task

- Teacher: ResNet34

- Student: ResNet18

- Dataset: Cifar10
  - Response-based distillation (35%, accuracy≥0.8 got 30%, the more improvement the higher score **compare with student from scratch**)
  - Feature-based distillation (35%, same as above)
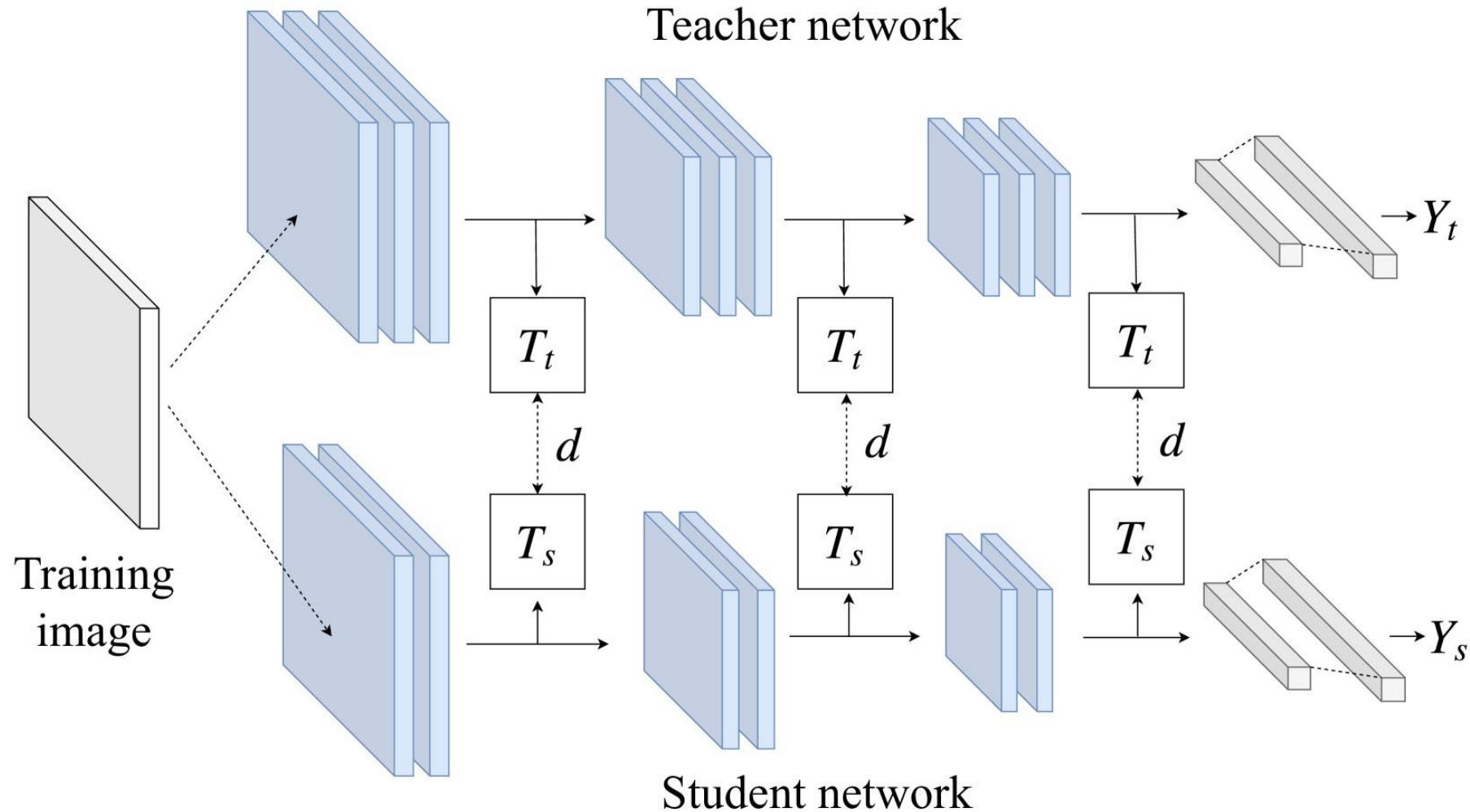 - Report (30%)
- Colab link
- **StudentID_lab5.ipynb**

Response-Based Knowledge
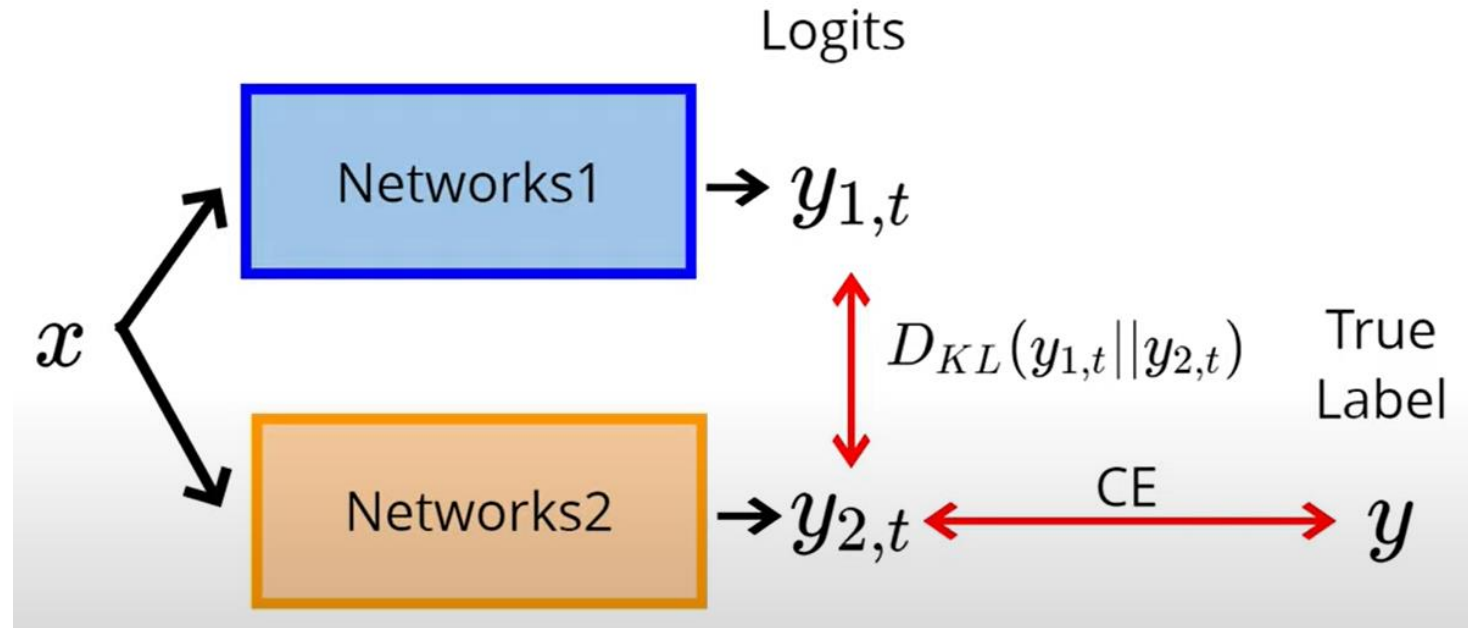
$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \qquad L = \alpha T^2 L^{(soft)} + (1-\alpha)L^{(hard)}$$

Feature-Based Knowledge

Reference

- [2006.05525] Knowledge Distillation: A Survey (arxiv.org)
- [1904.01866] A Comprehensive Overhaul of Feature Distillation (arxiv.org)
- [1503.02531] Distilling the Knowledge in a Neural Network(arxiv.org)

Thanks for listening