On-Device AI 실습: Knowledge Distillation

Dongkun Shin
Intelligent Embedded Systems Lab.
Sungkyunkwan University

Overview

1. **Baseline 학습 (Cross-Entropy Loss)**
   - Teacher 모델과 Student 모델을 각각 Cross-Entropy Loss만으로 학습시켜 정확도를 비교합니다.

2. **Knowledge Distillation (Soft Targets)**
   - Teacher의 softmax 출력을 활용한 Knowledge Distillation을 적용하고, temperature 및 loss weight에 따른 영향을 분석합니다.

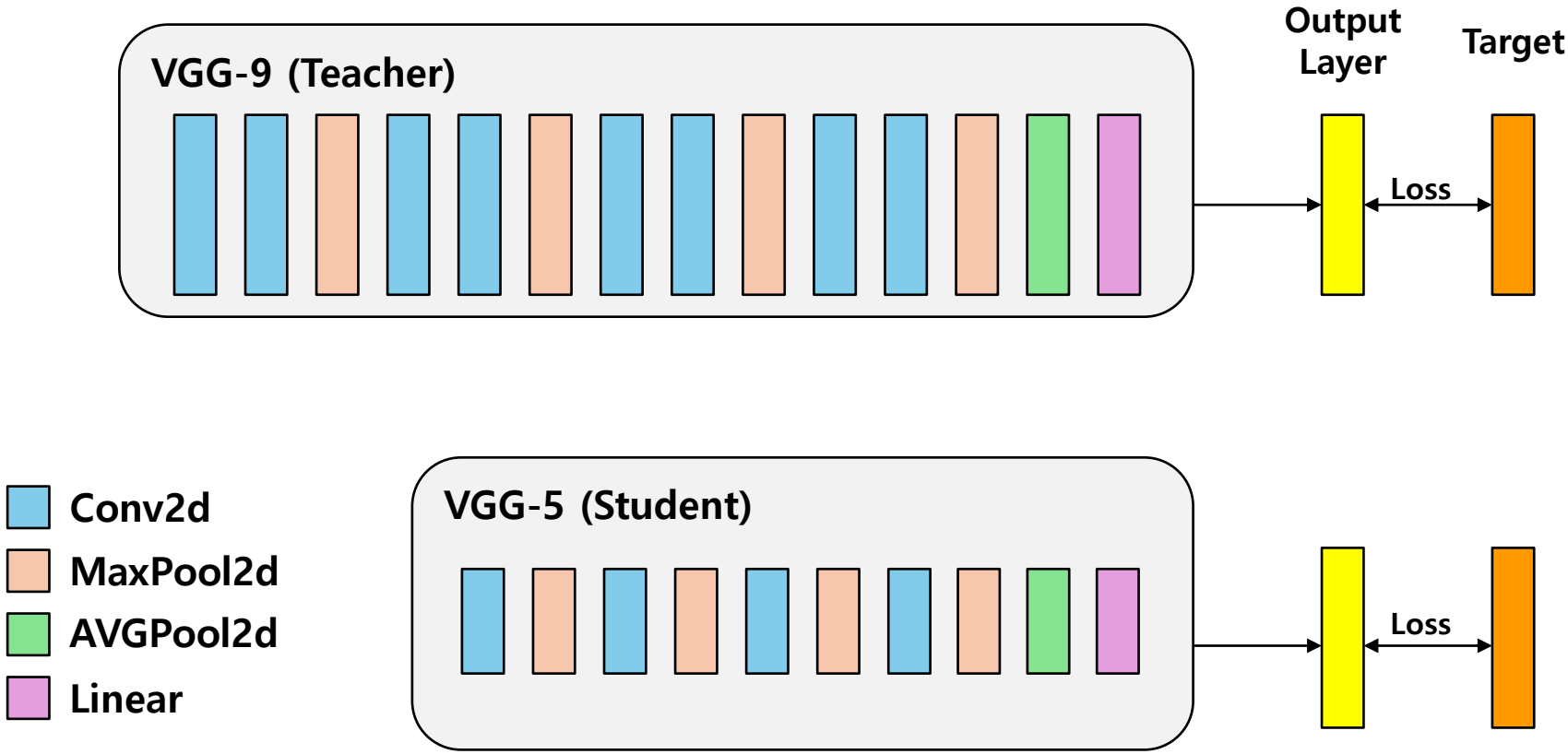
3. **Cosine Loss Minimization (Cosine Loss)**
   - Teacher와 Student의 convolutional feature를 추출하여, CosineEmbeddingLoss를 적용해 내부 표현 유사도를 증가시키는 방식으로 학습합니다.

4. **Intermediate Regressor (Regressor + MSE)**
   - Teacher의 feature map과 Student의 regressed feature map을 MSE로 정렬하며, 중간 표현을 직접 학습합니다.

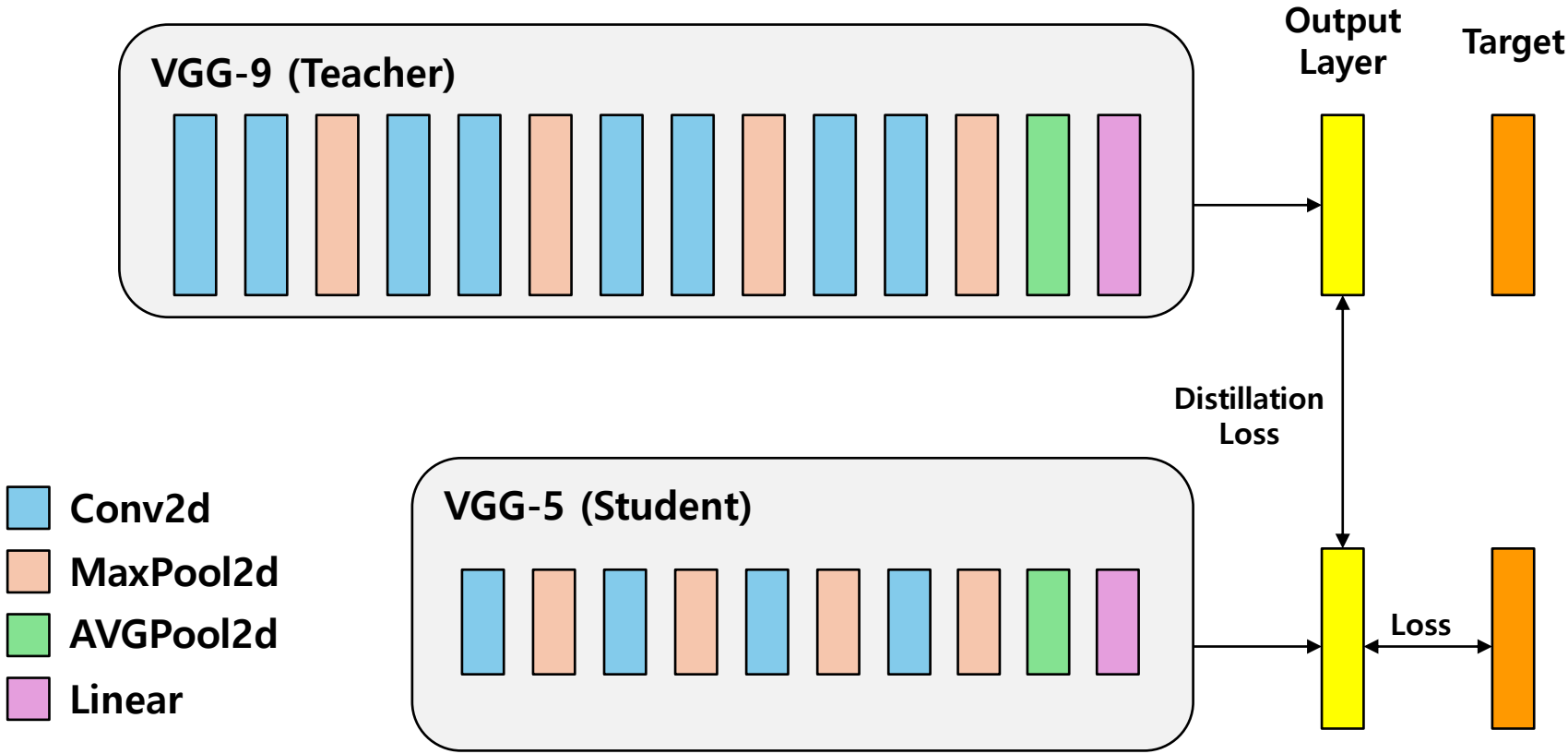
Setup

- 실습 자료 “Knowledge Distillation.ipnyb”을 colab에서 실행해주세요
- Colab 런타임을 GPU(T4)로 설정해 주세요
- Setup 코드 셀을 실행해 필요한 패키지를 설치해주세요

Baseline 학습 (Cross-Entropy Loss)



Knowledge Distillation (Soft Targets)



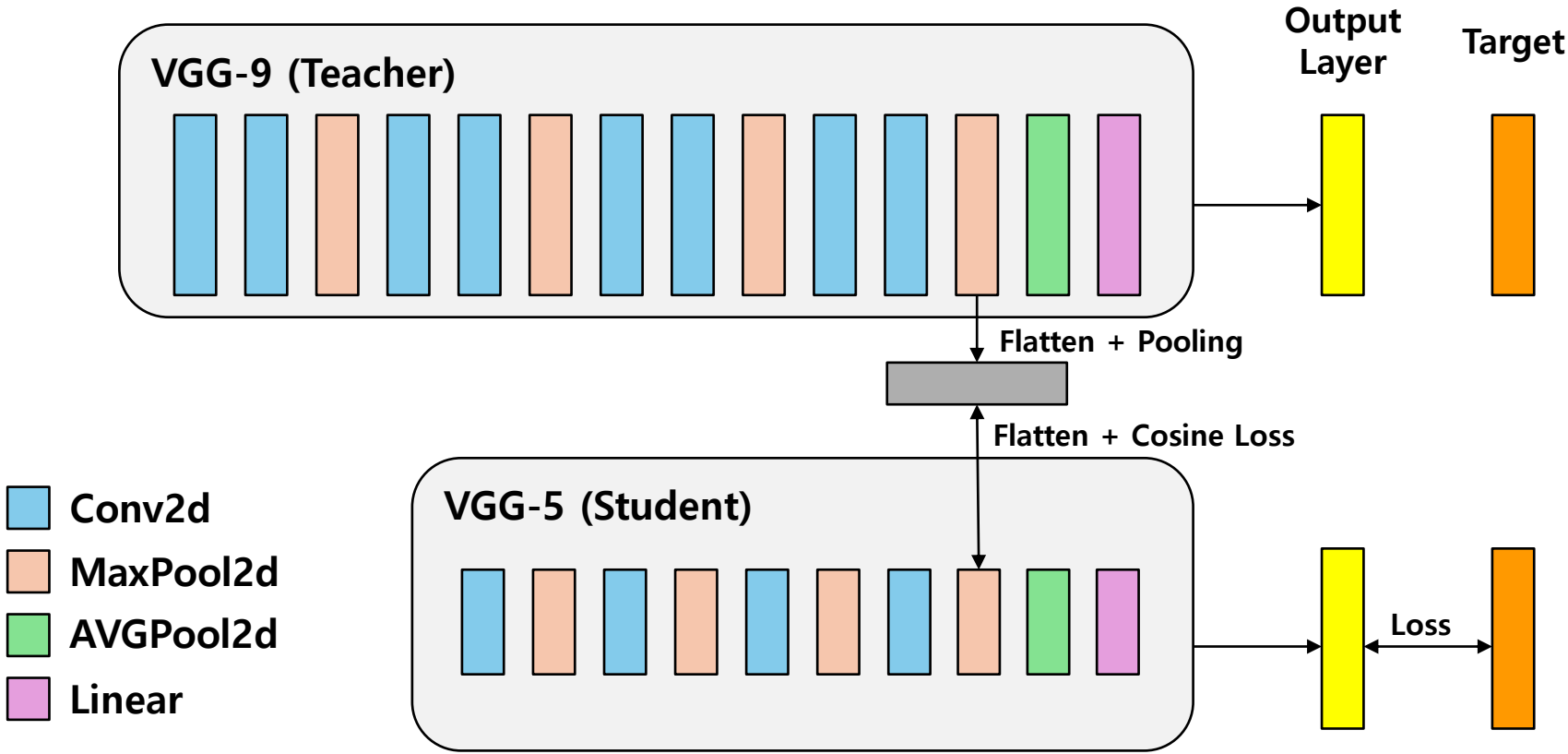
Distillation Loss 계산

- **KL divergence:**

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$$

- **Distillation Loss:**
  - $\frac{D_{\mathrm{KL}}(P \| Q)}{N} \times T^2$
    - N: batch size
    - T: temperature (loss의 scale이 T^2에 반비례하므로 T^2을 곱해 loss를 보정)

Cosine Loss Minimization (Cosine Loss)



Intermediate Regressor (Regressor + MSE)