

Дистилляция в глубоких сетях и выравнивание структур

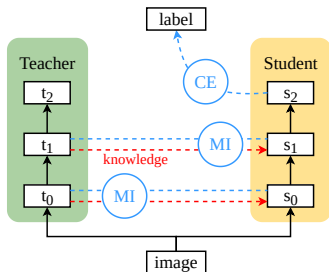


Рис.: Базовый метод

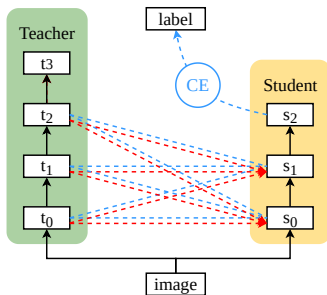


Рис.: Предлагаемый метод

$$\mathcal{L} = \beta \mathcal{L}_{\text{CE}} + (1 - \beta) \sum_{i,j=1}^{T,S} \lambda_{i,j} l(t_i, s_j) \quad (1)$$

$$\forall i \hookrightarrow \sum_{j=1}^S \lambda_{i,j} = 1 \quad (2)$$

T, S - количество слоёв учителя и ученика,
 $l(t_i, s_j)$ - взаимная информация,
 β и $\lambda_{i,j}$ — гиперпараметры.