南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Adaptive Early-Learning Correction for Segmentation from Noisy Annotations

Sheng Liu[*1] Kangning Liu[*1] Weicheng Zhu[1] Yiqiu Shen[1] Carlos Fernandez-Granda[1,2]
[1] NYU Center for Data Science [2] NYU Courant Institute of Mathematical Sciences

Liu S, Liu K, Zhu W, et al. Adaptive early-learning correction for segmentation from noisy annotations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2606-2616.

Inspired points

1. Observe a phenomenon that has been previously reported in the context of classification: the networks tend to first fit the clean annotations during an "early-learning" phase, before eventually memorizing the false annotations, thus jeopardizing generalization performance.

2. However, this phenomenon in semantic segmentation differs significantly from its counterpart in classification in the following way: In semantic segmentation, early learning and memorization do not occur simultaneously for all semantic categories due to pixel-wise imbalanced labels.

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Contributions

Propose a novel approach (ADELE, ADaptive Early-Learning corrEction) to perform semantic segmentation with noisy pixel-level annotations, which exploits early learning by adaptively correcting the annotations using the model output.

1. This method detects the beginning of the memorization phase by monitoring the **Intersection over Union (IoU) curve** for each category during training. This allows it to adaptively correct the noisy annotations in order to exploit early-learning for individual classes.

2. Incorporate a regularization term to promote spatial consistency, which further improves the robustness of segmentation networks to annotation noise.

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

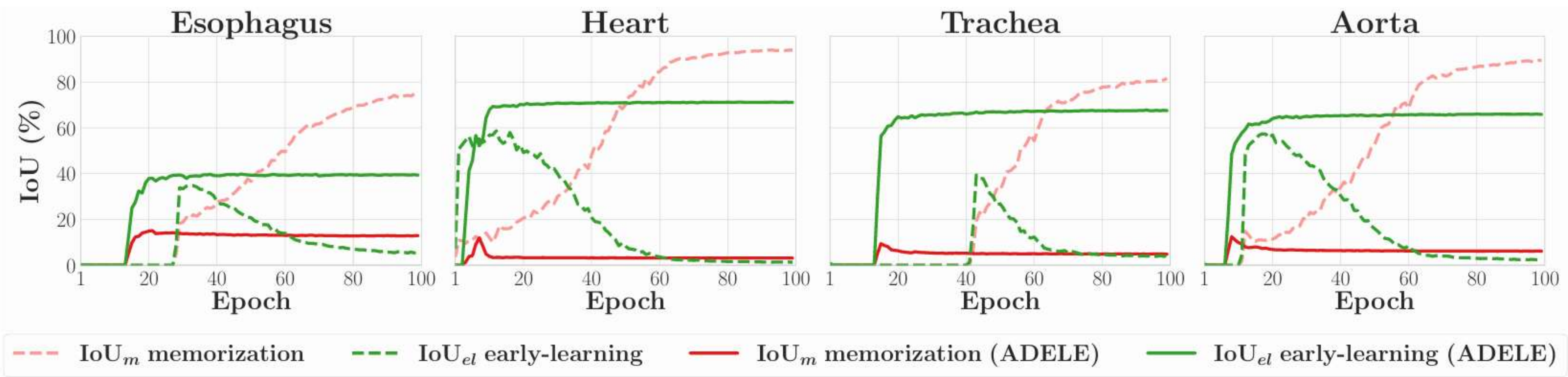
Methodology

1. Early learning and memorization in segmentation from noisy annotations

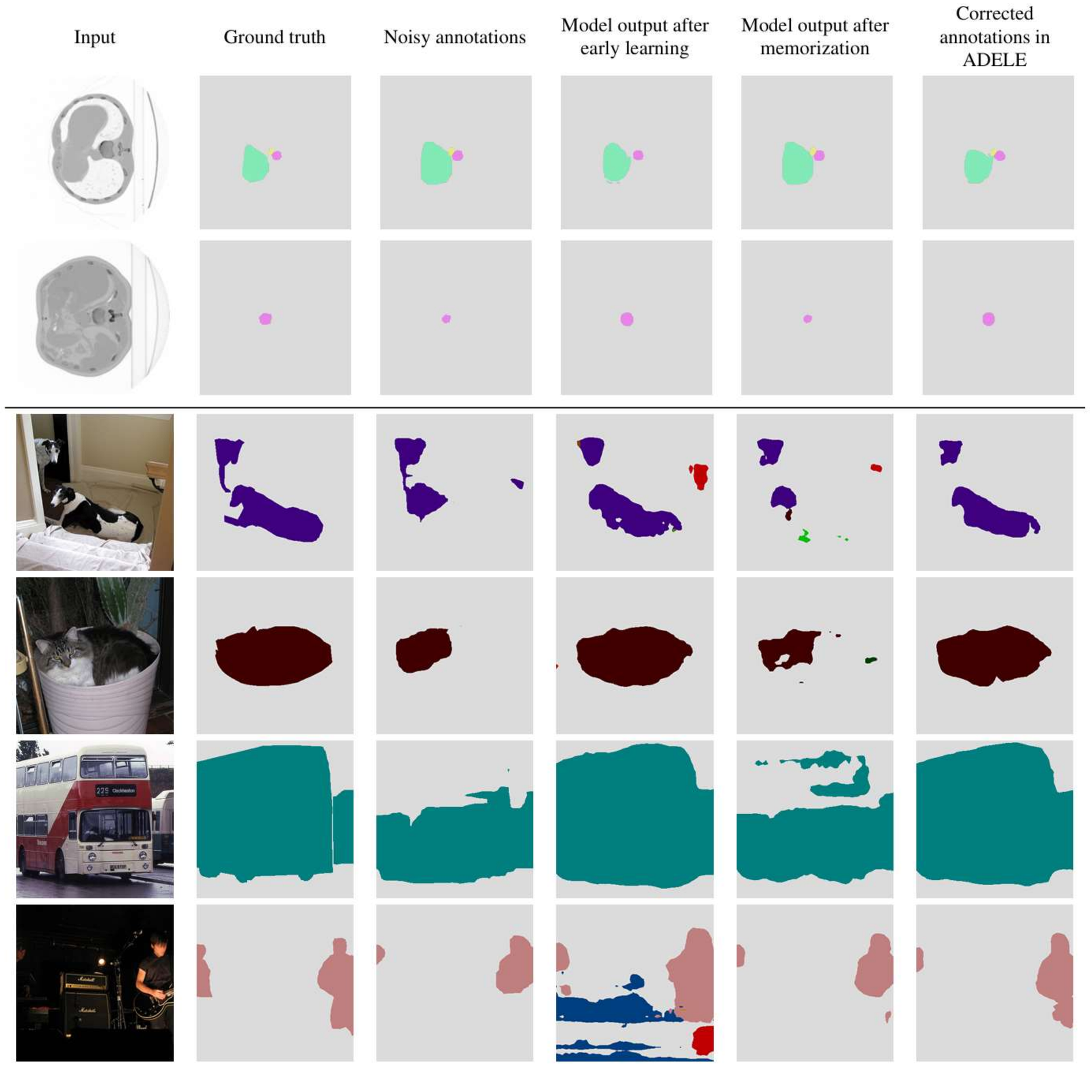
Given noisy annotations for which we know the ground truth, we can quantify the early-learning and memorization phenomena by analyzing the model output on the pixels that are incorrectly labeled:

early learning IoU_{el}: We quantify early learning using the overlap (measured in terms of the Intersection over Union (IoU) metric) between the outputs and the **corresponding ground truth label** on the pixels that are incorrectly labeled, denoted by IoU_{el}.

memorization IoU_m : We quantify memorization using the overlap (measured in IoU) between the outputs and the **incorrect labels**, denoted by IoU_m .



南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



Illustrates the effect of early learning and memorization on the model output.

noisy annotations (third column) are synthesized to resemble human annotation errors which either miss or encompass the ground truth regions (compare to second column).

在早期学习之后，这些区域被分割模型识别出来(第四列)，但是在记忆之后，模型过度拟合到不正确的注释，忘记了如何正确分割这些区域(第五列)

通过修改分割模型在噪声注释上的训练，以防止记忆。

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

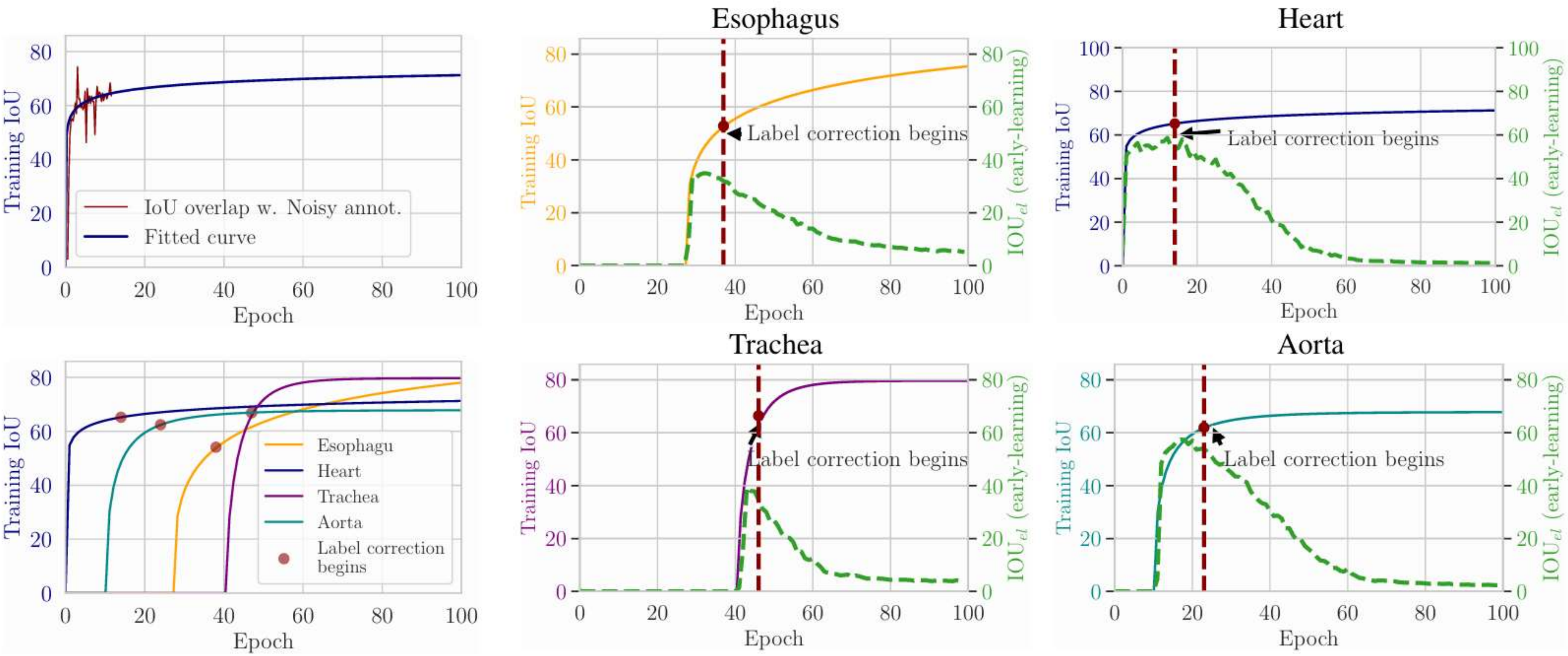
2. Adaptive label correction based on early learning

Determining when to correct the pixel-level annotations using the model output is challenging for two reasons:

• Correcting all classes at the same time can be sub-optimal.

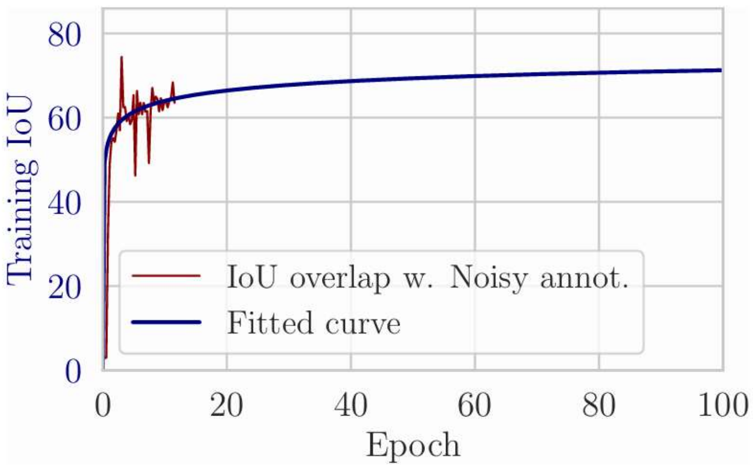
• During training, we do not have access to the performance of the model on ground-truth annotations.

Propose to update the annotations corresponding to different categories at different times by **detecting** when memorization is about to begin **using the training performance of the model**.



南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

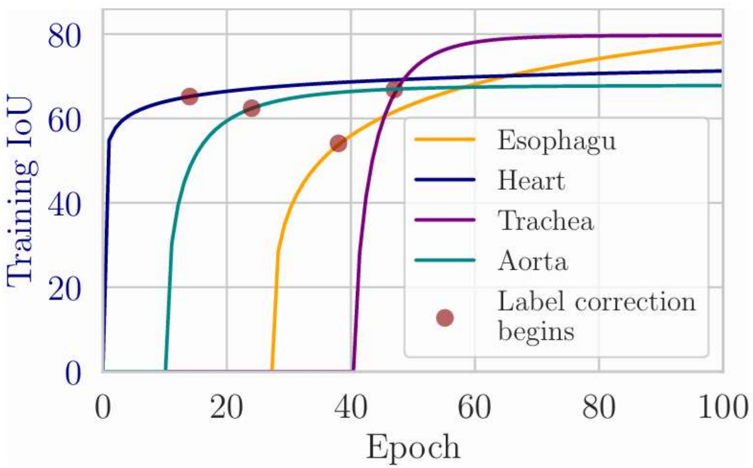
2. Adaptive label correction based on early learning



To estimate the deceleration we first f it the following exponential parametric model to the training IoU using least squares:

$$f(t) = a\left(1 - e^{-b\cdot t^{c}}\right),$$

Then we compute the derivative $f'(t)$ of the parametric model with respect to t at t =1 and at the current iteration.

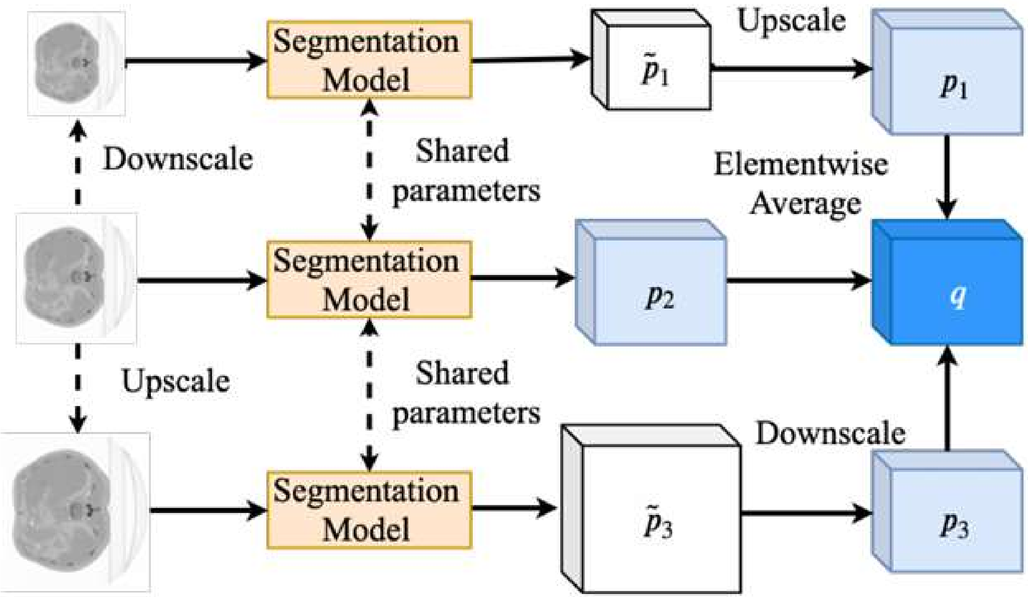


For each semantic category, the annotations are corrected when the relative change in derivative is above a certain threshold r

$$\frac{|f'(1) - f'(t)|}{|f'(1)|} > r,$$

only correct annotations for which the model output has confidence above a certain threshold τ.

3.Multiscale consistency



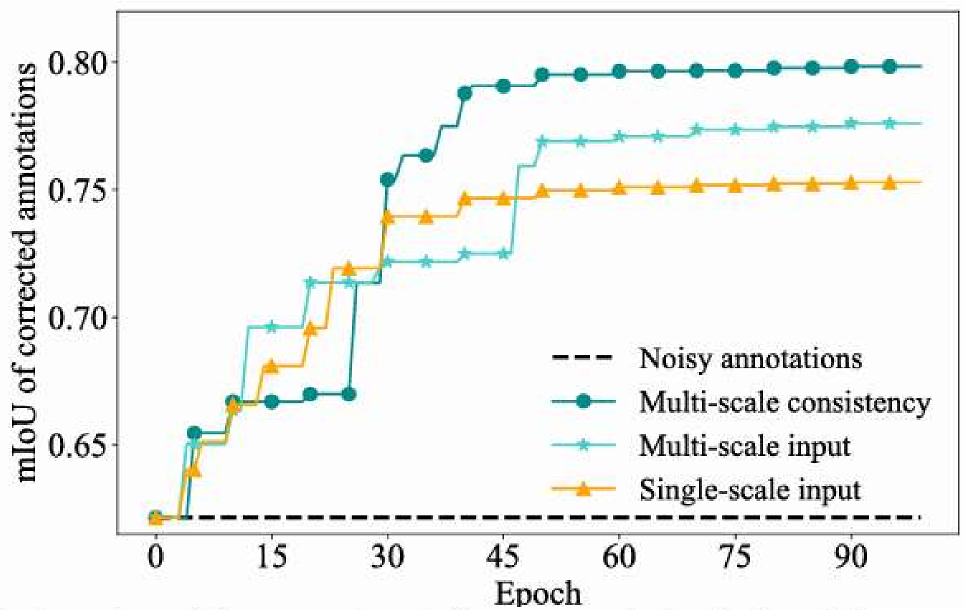


Figure 6. **Left:** In the proposed multiscale-consistency regularization, rescaled copies of the same input (here upscaled $\times 1.5$ and downscaled $\times 0.7$) are fed into the segmentation model. The outputs ($\tilde{p}_1$, p_2 and $\tilde{p}_3$) are rescaled to have the same dimensionality (p_1, p_2 and p_3). Regularization promotes consistency between these rescaled outputs and their elementwise average q. **Right:** Multi-scale consistency regularization leads to more accurate corrected annotations (results on SegThor, results for VOC 2012 can be found in Figure 12).

$$\mathcal{L}_{\text{Multiscale}}(x) = -\frac{1}{s}\sum_{k=1}^{s} \mathrm{KL}\left(p_k(x) \parallel q(x)\right),$$

$$q(x) = \frac{1}{s}\sum_{k=1}^{s} p_k(x)$$

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Experiments

	Baseline	ADELE w/o class adaptive	ADELE
Best val	62.6±2.3	40.7±2.5	**71.1±0.7**
Max test	63.3±2.0	40.7±2.4	**71.2±0.6**
Last Epoch	59.1±1.3	40.5±2.3	**70.8±0.7**

Table 1. The mIoU (%) comparison of the baseline and ADELE with or without class-adaptively correcting labels, on the test set of SegTHOR [27]. We report the test mIoU of the model that performs best on the validation set (**Best Val**), the test mIoU at the last epoch (**Last Epoch**), and the highest test performance during training (**Max Test**). We report mean and standard deviation after training the model with five realizations of noisy annotations.

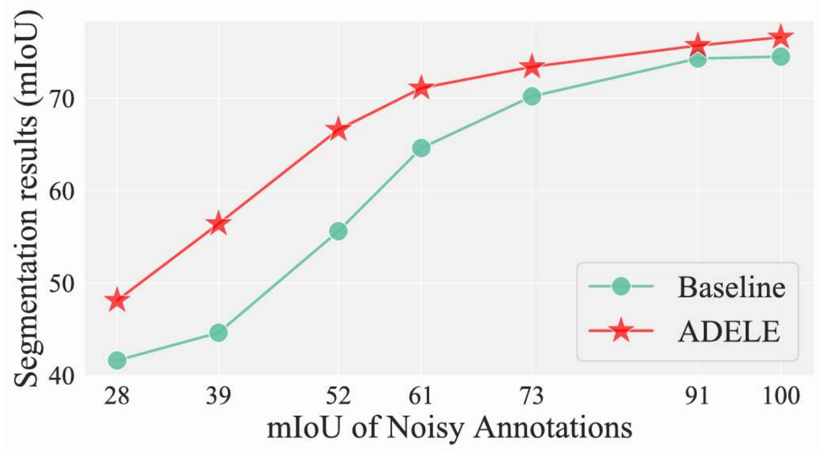


Figure 7. The performance comparison of the baseline and ADELE on the test set of SegTHOR [27]. The model is trained on noisy annotations with various levels of corruption (measured in mIoU with the clean ground truth annotations). ADELE is able to improve the model performance across a wide range of corruption levels.

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Ablation study for each part of ADELE.

	SegTHOR			PASCAL VOC 2012		
Label correction	Single scale	Multiscale input augmentation	Multiscale consistency regularization	Single scale	Multiscale input augmentation	Multiscale consistency regularization
✗	58.8	60.7	62.5	64.5	65.5	66.7
✓	65.2	69.8	**72.2**	65.6	67.3	**69.3**

Table 2. Ablation study for ADELE on SegTHOR [27] and PASCAL VOC 2012 [13]. We report the mIoU achieved at the last epoch on the validation set for both dataset. Class-adaptive label correction mechanism achieves the best performance when combined with multi-scale consistency regularization.

结　　束