IEEE - ISBI 2021
International Symposium on Biomedical Imaging

Uncertainty-Based Method for Improving Poorly Labeled Segmentation Datasets

E. REDEKOP[1], A. CHERNYAVSKIY [1]

1 Philips AI Research, Moscow Russia

PHILIPS

INTRODUCTION

- Modern segmentation CNNs require very high-quality ground truth masks
- Obtaining large datasets consisting of accurately labeled medical images is expensive and time-consuming task
- Low quality GT masks lead to poor performance of segmentation CNNs trained on them
- Our **goal** is to enhance the available ground truth segmentation data, and to improve the robustness of segmentation when noisy data are present

APPROACH

- Perform uncertainty estimation during training
- Relabel background/foreground in the ground truth according to uncertainty value
- This has the effect of the CNN not overfitting to noisy data and a good final performance on the clean validation dataset

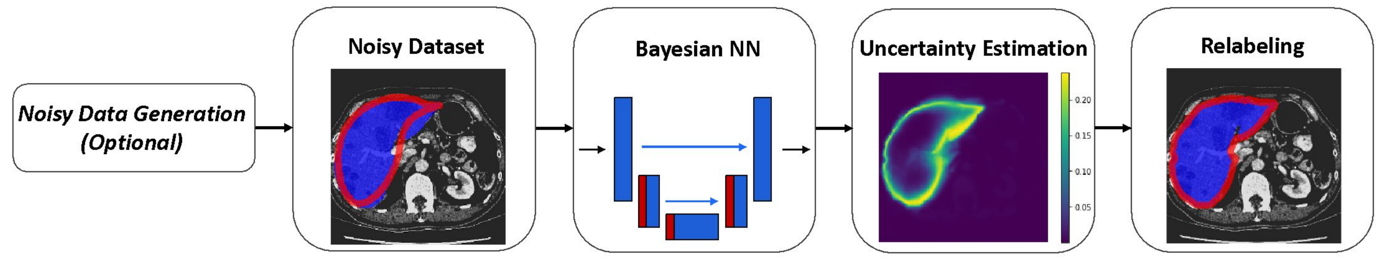

Fig. 1. Detection and relabeling of noisy ground truth labels for binary image segmentation.

NOISY DATA GENERATION

- Segmentation masks given in the training set are inherently noisy
- Otherwise clean ground truth masks should be artificially deteriorated

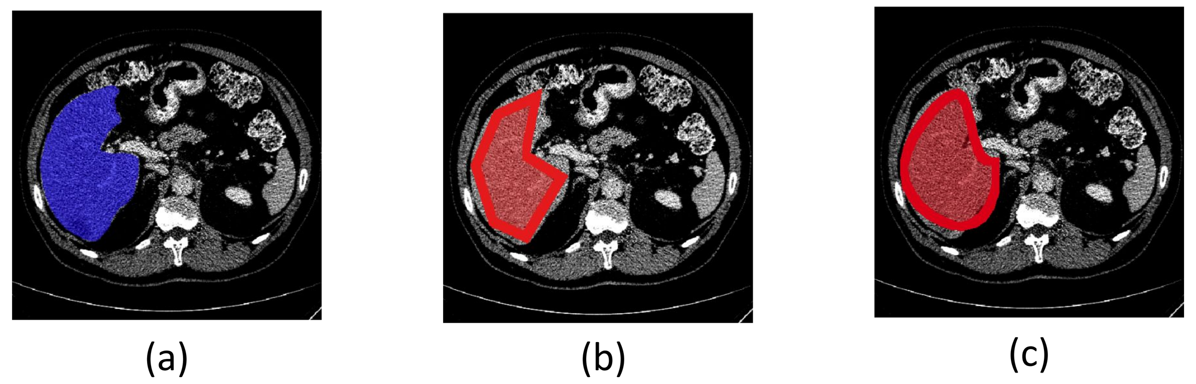

Fig. 3. (a) Ground truth mask, (b) low-vertex polygon approximation, (c) smooth curve approximation.

UNCERTAINTY-BASED RELABELING

Uncertainty-Estimation Methods:

- Monte Carlo Dropout (MCDO) [1]- run the trained model N times with activated dropout layers
- Deep Ensembles (DE) [2] - ensemble of N trained models
- Test-Time Augmentations (TTA) [3] – augment image using N different transformations

Uncertainty Types:

- Epistemic (model) uncertainty can be explained away given enough training data
- Aleatoric (data) uncertainty depends strictly on the noise in the input images

Uncertainty Calculation:

- Image of size $H \times W$
- Pixel-level predictions $\hat{p}$ of shape $N \times H \times W$
- According to [4]: $U_p = \frac{1}{N}\sum_{n=1}^{N} diag(\hat{p}_n) - \hat{p}_n \hat{p}_n^T$

Relabeling Rule:

- $p_{old} \in \{0,1\}$ – label of the pixel in the ground truth mask
- In areas of **high** uncertainty ($U_p > \delta$): $p_{new} = 1 - p_{old}$
- δ – predefined threshold

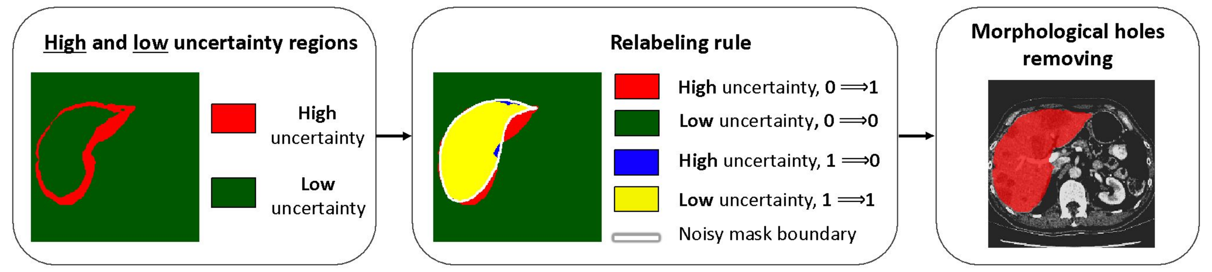

Fig. 2. Details of the relabeling step.

EXPERIMENTS

Datasets:

- Dermatoscopy [5] (2D images, 2000 training and 150 validation)
- Liver CT [6] (3D images, 100 training and 30 validation)

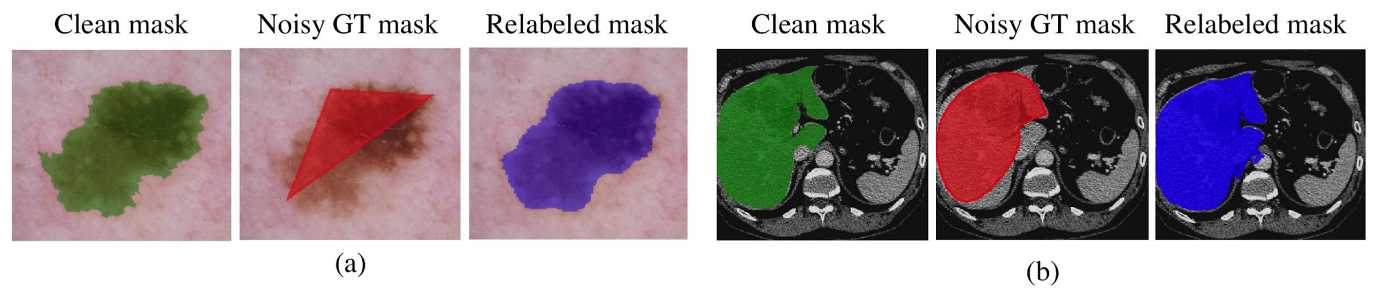

Fig. 4. Relabeling results for (a) Dermatoscopy dataset, (b) Liver CT dataset

		Dermatoscopy dataset	
Clean GT		82.5	
	3 vertices	7 vertices	smooth
No noise correction	57.9 / 65.2	78.3 / 80.2	78.7 / 80.6
Reweighting [7]	79.5 / 55.4	80.7 / 79.8	80.3 / 79.3
Relabeling (ours)	79.4 / 49.3	80.5 / 79.7	**81.5 / 80.4**
		Liver CT dataset	
Clean GT		82.5	
	3 vertices	7 vertices	smooth
No noise correction	61.5 / 69.5	85.3 / 86.2	85.1 / 86.3
Reweighting [7]	84.6 / 62.1	87.2 / 85.9	87.2 / 86.1
Relabeling (ours)	72.2 / 63.3	**88.0 / 87.1**	87.6 / 85.6

Table 1. Dice score, in percent, with respect to clean / noisy masks

RELABELING STRATEGIES

- During training, it is essential to start the relabeling process before the network **overfits** to noisy data

Dermatoscopy data:

Start at 2nd epoch as:

- Dice w.r.t. clean g.t. is maximum
- Relative change of mean cumulative uncertainty $\Delta\Sigma U_p$ is minimum

Liver CT data:

Start at 1st epoch as:

- Dice w.r.t. clean g.t. is maximum
- Relative change of mean cumulative uncertainty $\Delta\Sigma U_p$ is minimum

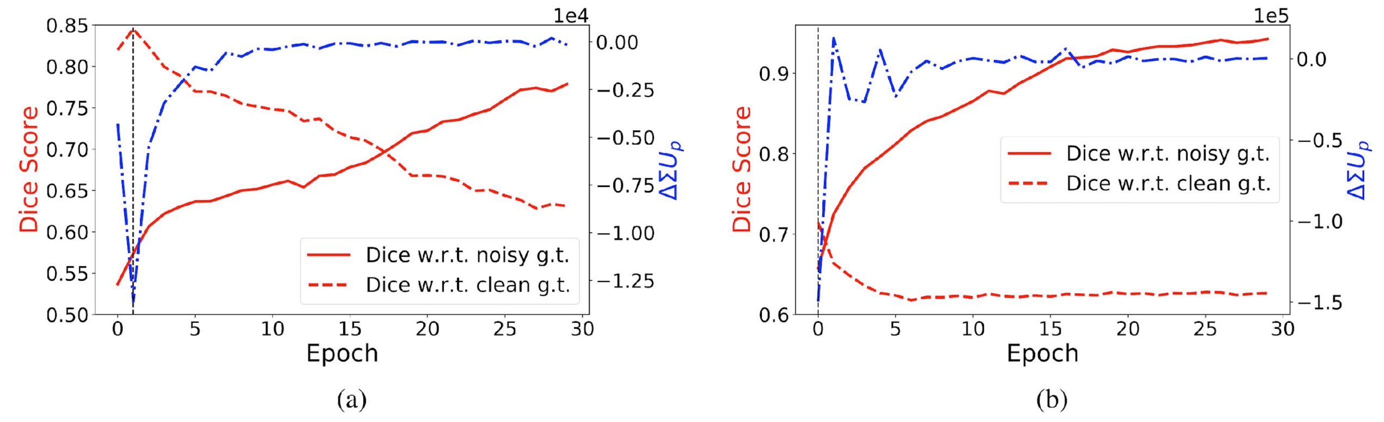

Fig. 5. Dice score of segmentations produced by the DCNN computed with respect to noisy and clean masks, and relative change of mean cumulative uncertainty for: (a) Dermatoscopy and (b) Liver CT datasets

CONTACT INFORMATION

alexey.chernyavskiy@philips.com

redekop.ep@gmail.com

CONCLUSIONS

- We have demonstrated that predictive uncertainty can be used for detecting noisy pixel-level labels in inaccurately annotated ground truth segmentation masks
- We did so by applying several uncertainty estimation methods and verified the relabeling of two image segmentation datasets
- Automatic relabeling improves segmentation quality without overfitting to data in-accuracies.
- The proposed algorithm can be used to generate cleaner datasets for training other deep learning algorithms without having to consider the impact of noisy labels.

REFERENCES

[1] Yongchan Kwon, Joong-Ho Won, Beom Joon Kim, andMyunghee Cho Paik, “Uncertainty quantification usingbayesian neural networks in classification: Applicationto ischemic stroke lesion segmentation,”MIDL, 2018.

[2] Balaji Lakshminarayanan, Alexander Pritzel, andCharles Blundell, “Simple and scalable predictive un-certainty estimation using deep ensembles,”inAd-vances in neural information processing systems, 2017,pp. 6402–6413.

[3] Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest,S´ebastien Ourselin, and Tom Vercauteren, “Aleatoricuncertainty estimation with test-time augmentation formedical image segmentation with convolutional neuralnetworks,”Neurocomputing, vol. 338, pp. 34–45, 2019.

[4] Yongchan Kwon, Joong-Ho Won, Beom Joon Kim, and Myunghee Cho Paik, “Uncertainty quantification using bayesian neural networks in classification: Application to ischemic stroke lesion segmentation”, MIDL, 2018.

[5] Noel CF Codella, David Gutman, M Emre Celebi, BrianHelba, Michael A Marchetti, Stephen W Dusza, AadiKalloo, Konstantinos Liopyris, Nabin Mishra, HaraldKittler, et al., “Skin lesion analysis toward melanomadetection: A challenge at the 2017 international sym-posium on biomedical imaging (isbi), hosted by theinternational skin imaging collaboration (isic),”in2018 IEEE 15th International Symposium on Biomed-ical Imaging (ISBI 2018). IEEE, 2018, pp. 168–172

[6] Amber L Simpson, Michela Antonelli, Spyridon Bakas,Michel Bilello, Keyvan Farahani, Bram Van Ginneken,Annette Kopp-Schneider, Bennett A Landman, GeertLitjens, Bjoern Menze, et al.,“A large annotatedmedical image dataset for the development and eval-uation of segmentation algorithms,”arXiv preprintarXiv:1902.09063, 2019.

[7] Zahra Mirikharaji, Yiqi Yan, and Ghassan Hamarneh,“Learning to segment skin lesions from noisy anno-tations,”inDomain Adaptation and RepresentationTransfer and Medical Image Learning with Less Labelsand Imperfect Data, pp. 207–215. Springer, 2019.