Modeling clinical assessor intervariability using deep hypersphere auto-encoders

Joost van der Putten[1] J.A.V.D.PUTTEN@TUE.NL
Fons van der Sommen[1] FVDSOMMEN@TUE.NL
Jeroen de Groof[2] A.J.DEGROOF@AMC.UVA.NL
Maarten Struyvenberg[2] M.R.STRUYVENBERG@AMC.UVA.NL
Svitlana Zinger[1] S.ZINGER@TUE.NL
Wouter Curvers[3] WOUTER.CURVERS@CATHARINAZIEKENHUIS.NL
Erik Schoon[3] ERIK.SCHOON@CATHARINAZIEKENHUIS.NL
Jacques Bergman[2] J.J.BERGMAN@AMC.UVA.NL
Peter H.N. de With[1] P.H.N.DE.WITH@TUE.NL

[1] *VCA Group, Eindhoven University of Technology, Eindhoven, the Netherlands*

[2] *Amsterdam University Medical Center, Amsterdam, the Netherlands*

[3] *Catherina Hospital, Eindhoven, the Netherlands*

Editors: Under Review for MIDL 2019

1. Introduction

A long standing challenge in machine learning is the usage of data sets that lack a gold-standard ground truth. Especially in medical imaging applications, many data set segmentations exhibit a large inter-observer variability (Sørensen et al., 1993) (Williamson et al., 2018). In this work, we propose the Hypersphere Auto-Encoder (HAE), a novel deep learning architecture capable of modeling inherent data ambiguity, based on multi-assessor input.

2. Methods

2.1. VLE data set

Volumetric Laser Endomicroscopy (VLE) is a second-generation Optical Coherence Tomography (OCT) imaging modality, capable of making a full circumferential scan of the esophageal wall up to a tissue depth of 3 mm (Gonzalo et al., 2010). However, the lower boundary of relevant tissue is ambiguous (van der Putten et al., 2019). Since the lower boundaries are not clearly defined, multiple proposals can greatly benefit the user. The VLE data set was acquired in a prospective single-center study in which 23 Barrett's patients with and without early neoplasia were included. In total, 131 Regions Of Interest (ROI) were extracted from the patients, where an ROI is defined as a selected section of the entire scan where histology was proven. The original ROIs have a resolution of $1,342 \times 1,024$ pixels, which were resized to 256×256 pixels for computational efficiency. An example of a VLE image are shown in Figure 1a. For the remainder of this work, union of the expert annotations is referred to as the softspot, while the intersection of all annotations is referred to as the sweetspot.

© 2019 J.v.d. Putten *et al.*

2.2. Hypersphere auto-encoder

Model definition The proposed HAE architecture uses U-net (Ronneberger et al., 2015) as the base model, but is extended with an additional feature mapping layer. In this section, model $M_\theta(Y|X,\rho)$ is discussed, which produces segmentation mask Y conditioned on input image X and parameter ρ. Model $M_\theta(Y|X,\rho)$ is comprised of encoder network $q_\theta(\mathbf{z}|X,\rho)$ which maps the input to latent variable vector $\mathbf{z} \in \mathbb{R}^N$, and decoder network $p_\theta(Y|\mathbf{z},\rho)$ which produces a segmentation mask depending on latent variable $\mathbf{z}$ and size factor ρ. Additionally, the original skip connections from standard U-net are replaced with fully residual connections to reduce the amount of learnable parameters in the decoder. Finally, we propose dynamic leaky ReLUs (DLReLU) which can dynamically be changed for each input. DLReLUs are added before each skip connection in the auto-encoder, thereby propagating size information to the decoder.

Mask interpolation Since the goal of the proposed algorithm is to condition the output prediction size on ρ, one of the most important learning aspects is determining a ground truth that reflects this property. In order to obtain intermittent ground-truth masks, a linear interpolation is generated between the softspot and the sweetspot by using the distance transform (Maurer et al., 2003).

Hypersphere loss In order to condition the output prediction on ρ, the HyperSphere Loss (HSL) is introduced and specified by:

$$HSL = \frac{1}{n}\sum_{i=1}^{n}(||\mathbf{z}_i||_2 - 1)^2, \tag{1}$$

where n is the batch size, i is the batch index and $||\mathbf{z}_i||_2$ is the dot product of $\mathbf{z}_i$ with itself. The resulting HSL is minimal when the latent vector $\mathbf{z} \in \mathbb{R}^N$ is a point on an N-dimensional hypersphere $\mathcal{S}$ with unity radius. Consequently, $\mathbf{z}$ is scaled such that $\mathbf{z}_{scaled} = \mathbf{z}(\chi + \rho)$ with $0 \leq \rho \leq 1$ prior to being used as input to the decoder. The offset $\chi = 0.5$ is included to prevent multiplication by 0, which otherwise would produce a null feature vector while the output is a non-zero mask. The normalized vector $||\mathbf{z}_{scaled}||$ is now directly correlated with the desired output annotation size, where we define that $\rho = 0$ corresponds to the sweetspot and $\rho = 1$ corresponds to the softspot. The total loss of the network is defined by $Loss = DICE + 0.5 \cdot HSL$

Dynamic leaky ReLU A second way to condition the output on ρ is by introducing the Dynamic Leaky ReLU (DLReLU). The DLReLU is very similar to the original Leaky Rectified Linear Unit (LReLU) first introduced in (Maas et al., 2013). For a DLReLU, instead of fixed constant sloped, the negative slope is dependent on size factor ρ It should be noted that the value of ρ can be different per pass through the network. Hence, this dynamically changes the network based on ρ. Since $0 \leq \rho \leq 1$, the slope for negative inputs will be 0 when the output annotation should correspond to the sweetspot and 0.5 when it corresponds to the softspot. In this way, the negative values after activation correspond to the size of the output segmentation.

(a) (b) (c) (d) (e)

Figure 1: (a) VLE inference image. (b) Corresponding softspot. (c) Corresponding sweetspot. (d) Heatmap generated from inference results with 6 levels. (e) Standard U-net prediction.

3. Results

Figure 1 shows visual results of the hypersphere auto-encoder architecture on VLE data. As can be seen in Figure 1e, a standard U-net prediction contains no information about the assessor uncertainty, whereas the prediction heatmap in Figure 1d clearly shows a decrease in prediction certainty for deeper layers in the tissue where the signal-to-noise ratio degrades. Additionally, similar to the softspot (Figure 1b) and the sweetspot (Figure 1c), the predicted heatmap shows no ambiguity near the top layer of the tissue where the boundary is clearly defined and there is hardly any inter-observer variability. These results highlight that the network does not simply increase the general size of the prediction depending on the value of ρ. Instead, the network learns to increase the size of the prediction based on the modeled uncertainty. Figure 1d clearly shows the decreasing confidence in the heatmap at lower tissue layers. In clinical practice, this model can assist the physician with informed decision making based on the modeled multiple expert opinions. In this way, the methodology facilitates explainable AI to the clinical users.

4. Discussion and Conclusion

Many medical imaging data sets contain ground-truth segmentations with a large inter-observer variability. Most state-of-the-art segmentation models do not take this drawback into account and are fully deterministic in nature. To this end, we have proposed the Hypersphere Auto-Encoder (HAE), which represents an architecture for explicitly incorporating multi-assessor intervariability for multiple segmentation proposals into the segmentation model. We have shown that conditioning this model on parameter ρ effectively facilitates the generation of probability heatmaps that visualize model uncertainty along segmentation boundaries. Additionally, we provide a proof of concept on a medical data set with ambiguous ground truth as well as provide improved interpretability of the results.

The output of $M_\theta(Y|X, \rho)$ was conditioned on ρ in two ways. First, we proposed dynamic leaky ReLUs as a way of propagating size information through the negative activations of the network. Second, a feature-mapping layer was added which scales the length of the low-dimensional feature vector proportional to the value of ρ. In future work, variational methods such as in (Kingma and Welling, 2013) can be explored to sample e.g. coarseness in addition to size. This facilitates the generation of infinite unique region proposals for even better heatmap generation.

Acknowledgments

We thank Nvidia for supplying a Titan X for this research through their GPU grant program. Additionally, we thank KWF for funding this research.

References

Nieves Gonzalo, Guillermo J Tearney, Patrick W Serruys, Gijs van Soest, Takayuki Okamura, Héctor M García-García, Robert Jan van Geuns, Martin van der Ent, Jurgen Ligthart, Brett E Boum, et al. Second-generation optical coherence tomography in clinical practice. high-speed data acquisition is highly reproducible in patients undergoing percutaneous coronary intervention. *Revista Española de Cardiología (English Edition)*, 63(8):893–903, 2010.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.

Calvin R Maurer, Rensheng Qi, and Vijay Raghavan. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270, 2003.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

JB Sørensen, M Klee, T Palshof, and HH Hansen. Performance status assessment in cancer patients. an inter-observer variability study. *British journal of cancer*, 67(4):773, 1993.

Joost van der Putten, Fons van der Sommen, Maarten Struyvenberg, Jeroen de Groof, Wouter Curvers, Erik Schoon, Jaqcues Bergman, and Peter de With. Tissue segmentation in volumetric laser endomicroscopy data using fusionnet and a domain-specific loss function. *SPIE Medical Imaging*, 2019.

Sean R Williamson, Priya Rao, Ondrej Hes, Jonathan I Epstein, Steven C Smith, Maria M Picken, Ming Zhou, Maria S Tretiakova, Satish K Tickoo, Ying-Bei Chen, et al. Challenges in pathologic staging of renal cell carcinoma: A study of interobserver variability among urologic pathologists. *The American journal of surgical pathology*, 2018.