The Impact Label Noise and Choice of Threshold has on Cross-Entropy and Soft-Dice in Image Segmentation

Marcus Nordström [1,2] Atsuto Maki [1] Henrik Hult [1]

[1] KTH Royal Institute of Technology, Stockholm, Sweden

[2] RaySearch Laboratories, Stockholm, Sweden

Introduction

Background

- Two of the most widely used loss functions in medical image segmentation are cross-Entropy and soft-Dice.
- Cross-entropy is known to have better stability properties and soft-Dice has been shown to perform better experimentally with respect to the Dice metric.

This work

- Show how noise affects the optimal segmentation
- Show how optimal solutions to soft-Dice can be recovered by thresholding optimal cross-entropy solutions with a particular a priori unknown but efficiently computable threshold

Definitions

- Let $\Omega = [0,1]^n$ be a continuous image domain and λ be the associated Lebesgue measure.
- Let $\mathcal{M}$ be the space of soft segmentations $\Omega \to [0,1]$ and $\mathcal{S}$ the hard segmentations $\Omega \to \{0,1\}$.
- Let $I_{[a,b]}(\cdot)$ be the indicator function with respect to an interval $[a, b]$.
- Let cross-Entropy and soft-Dice be denoted by:

$$\mathrm{CE}_m(c) \doteq -\int_\Omega [m(\omega)\log(c(\omega)) + (1-m(\omega))\log(1-c(\omega))]\lambda(d\omega), \quad c \in \mathcal{M}, \tag{1}$$

$$\mathrm{SD}_m(c) \doteq 1 - \frac{2\int_\Omega c(\omega)m(\omega)\lambda(d\omega)}{\|c\|_1 + \|m\|_1}, \quad c \in \mathcal{M}. \tag{2}$$

Noise Model

To be able to study how noise affects delineation in a controlled setting, a noise model is introduced. The model formally treats a noisy segmentation as a particular kind of randomly deformed deterministic reference segmentation. This allows for relatively realistic noise, that incorporates complicated correlations found among border pixels.

- Let $\Omega = [0,1]^n$ be a continuous image domain.
- Let $I : \Omega \to \{0,1\}$ be a deterministic reference segmentation.
- Let $X \doteq (X_1, \dots, X_n)^T$ be a vector of $\{X_i\}_{i=1}^n$ independent Gaussian fields.
- Let L be a random deformation of the reference segmentation:

$$L(\omega) \doteq \begin{cases} I(\omega + X(\omega)) & \text{if } \omega + X(\omega) \in \Omega, \\ 0 & \text{if } \omega + X(\omega) \notin \Omega. \end{cases}, \quad \omega \in \Omega.$$

Propositions

Reference	Statement	Interpretation
Proposition 1	Let l be a reference segmentation and L an associated noisy segmentation, then $$\mathbb{E}[L(\omega)] = \int_{\Omega} l(\omega')p_{a^2}(\omega-\omega')\lambda(d\omega'), \quad \omega \in \Omega.$$	Explicit formula for marginal probabilities.
Proposition 2	Let l be a reference segmentation and L an associated noisy segmentation, then $$\mathbb{E}[\|L\|_1] = \|l\|_1 - \xi,$$ where $$\xi = \int_{\Omega}\int_{\mathbb{R}^n\setminus\Omega} l(\omega')p_{a^2}(\omega-\omega')\lambda(d\omega)\lambda(d\omega').$$	Noise is approximately volume preserving.
Proposition 3	Let X be a vector of independent Gaussian fields and $W_1, \dots, W_n$ be independent copies of independently scattered Gaussian measures on $\mathbb{R}^n$ with control measure λ, then it follows that $$X(\omega) \stackrel{d}{=} \begin{pmatrix} a(2\pi b^2)^{n/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega-\omega')W_1(dw') \\ \vdots \\ a(2\pi b^2)^{n/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega-\omega')W_n(dw') \end{pmatrix}.$$	Explicit formula for efficient sampling.

[1] The isotropic normal density with variance σ^2 is denoted by p_{σ^2}

Optimal Segmentation Under Label Noise

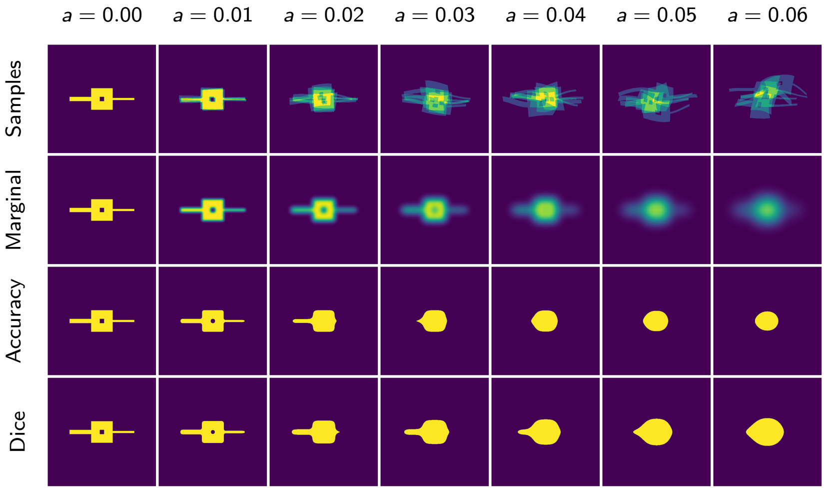

Figure: Illustration of the noise model and its effects on a two dimensional segmentation problem, with the columns showing the effect of different noise strengths. The first row shows the pixel-wise average obtained from five samples. The second row shows the marginal function. The third row shows optimal segmentations associated with Accuracy, which in this and most natural cases is unique. The forth row shows optimal segmentations associated with Dice, which in this and most natural cases is unique.

Alternative Threshold

By using an alternative, a priori unknown but efficiently computable threshold:

$$t_m \doteq \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2. \tag{3}$$

That is, by thresholding an optimizer to cross-entropy with this alternative recursively defined threshold t_m, it follows that

$$\mathrm{CE}_m(c) = \inf_{c' \in \mathcal{M}} \mathrm{CE}(c') \implies \mathrm{SD}_m(I_{[t_c,1]} \circ c) = \inf_{c' \in \mathcal{M}} \mathrm{SD}(c'). \tag{4}$$

This type of thresholding has not been studied in the medical image segmentation community but has been studied in the statistical learning literature [2].

[1] [1] Z. C. Lipton, C. Elkan, and B. Naryanaswamy. Optimal Thresholding of Classifiers to Maximize F1 Measure.

Experiments

- Standard U-net setup with data augmentation and regularization.
- Three organs are selected from TotalSegmentor data set [1].
- Each image in dataset is randomly deformed with the noise model.
- Models are trained with cross-entropy and soft-Dice
- Evaluation is done for
  - Cross-entropy with 0.5-threshold
  - Soft-Dice with 0.5-threshold
  - Cross-entropy with the alternative threshold
- Four different noise levels
- Five folds

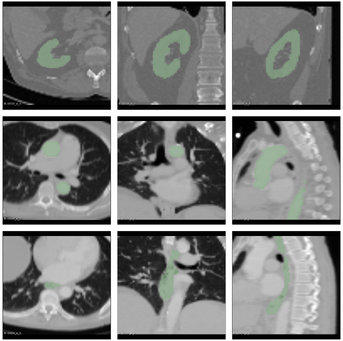

Figure: Illustration of the data from the TotalSegmentor data [1] used in the experiments. The first depicts a kidney, the second row an aorta and the third row an esophagus. The columns correspond to axial, coronal and sagittal views respectively.

[1] J. Wasserthal, M. Meyer, H.-C. Breit, J. Cyriac, S. Yang, and M. Segeroth. Totalsegmentator: robust segmentation of 104 anatomical structures in ct images.

Results and Conclusion

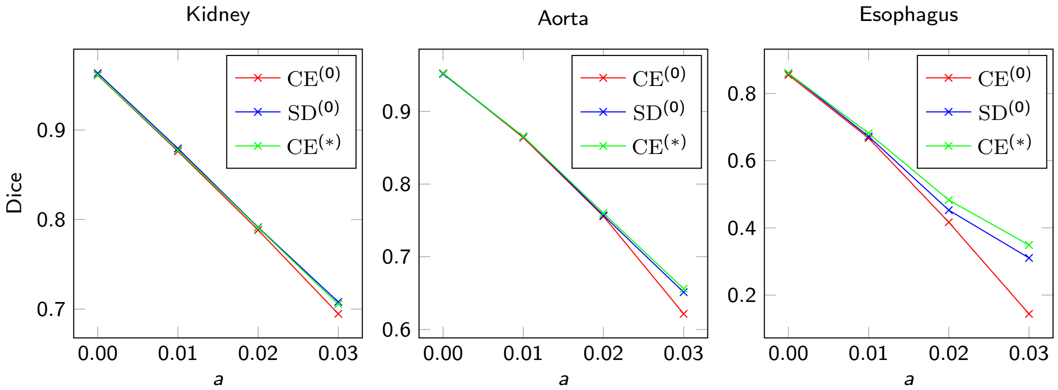

Figure: In each of the figures, the average Dice value obtained from five fold experiments is plotted as a function of the noise level a. The legend $\mathrm{CE}^{(\mathbf{0})}$ indicate that cross-entropy with the $1/2$-threshold has been used, $\mathrm{SD}^{(\mathbf{0})}$ indicate that soft-Dice with the $1/2$-threshold has been used and $\mathrm{CE}^{(*)}$ indicate that cross-entropy with the alternative threshold.

Thanks!