Probabilistic U-Net

A Preprint

Ivan Rodin
Data Science track
Skoltech

Vladislav Shlyonskiy
Data Science track
Skoltech

October 26, 2018

Abstract

Many real-world vision problems suffer from inherent ambiguities. In clinical applications for example, it might not be clear from a CT scan alone which particular region is cancer tissue. Therefore a group of graders typically produces a set of diverse but plausible segmentations.

Keywords Segmentation · Bayesian Machine Learning

1 Introduction

The architecture of probabilistic UNet is to learn conditional density model over segmentation, conditioned on image. To do so, the latent space of dimension N is introduced. The modelitself can be viewed as a combination on standard UNet architecture and VAE for latent space modelling.

2 Architecture

2.1 Sampling

Each position in this space encodes a segmentation variant. Prior net, which is parametrized by weights ω and depicted on figure 2, estimates the probability of these variants for a given input image X. This prior probability distribution, let's name it P, is modelled as an axis-aligned Gaussian with mean $\mu_{\text{prior}}(X,\omega) \in \mathbb{R}^N$ and variance $\sigma_{\text{prior}}(X,\omega) \in \mathbb{R}^N$. To predict a set of m segmentation maps we apply the network m times to the same input image. In each iteration $i \in \overline{1,n}$, a random sample $z_i \in \mathbb{R}^N$ is drawn from P.

$$\mathbf{z}_i \sim P\big(\bullet \mid X\big) = \mathcal{N}\Big(\mu_{\text{prior}}(X;\omega), \text{diag}\big[\sigma_{\text{prior}}(X;\omega)\big]\Big) \tag{1}$$

broadcast the sample to an N-channel feature map with the same shape as the segmentation map, and concatenate this feature map to the last activation map of a U-Net parameterized by weights θ.

A function f_{comb} is composed of three subsequent 1×1 convolutions, having ψ as the set of their weights, combines the information and maps it to the desired number of classes. The output S_i is the segmentation map corresponding to point z_i in the latent space:

$$S_i = f_{\text{comb.}}\Big(f_{\text{U}-\text{Net}}(X;\theta),\ \mathbf{z}_i;\ \psi\Big) \tag{2}$$

2.2 Training

The networks are trained with the standard training procedure for conditional VAEs, i.e. by minimizing the variational lower bound. The main difference with respect to training a deterministic segmentation model, is that the training process additionally needs to find a useful embedding of the segmentation variants in the latent space. This is solved by posterior net, parametrized by weights ν, that learns to recognize a segmentation variant (given the raw image X

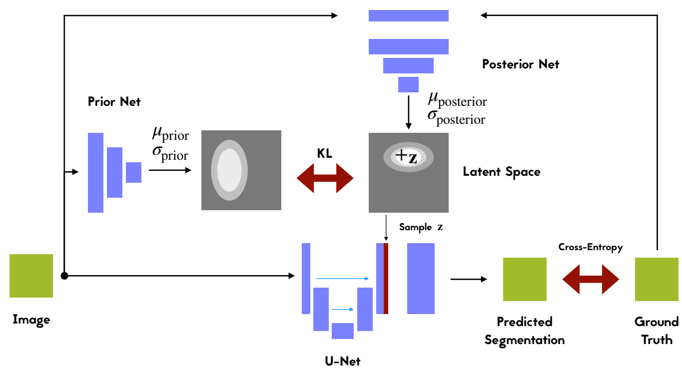

Figure 1: Training process

and the ground truth segmentation Y) and to map this to a position $\mu_{\text{posterior}}(X, Y; \nu) \in \mathbb{R}^N$ with some uncertainty $\sigma_{\text{posterior}}(X, Y; \nu) \in \mathbb{R}^N$ in the latent space. The output is denoted as posterior distribution Q. A sample $\mathbf{z}$ from this distribution:

$$\mathbf{z} \sim Q(\bullet \mid X, Y) = \mathcal{N}\Big(\mu_{\text{posterior}}(X, y; \omega), \operatorname{diag}\big[\sigma_{\text{posterior}}(X, Y; \omega)\big]\Big), \tag{3}$$

combined with the activation map of the U-Net must result in a predicted segmentation S identical to the ground truth segmentation Y provided in the training example. A cross-entropy loss penalizes differences between S and Y (the cross-entropy loss arises from treating the output S as the parameterization of a pixel-wise categorical distribution P_c). Additionally there is a Kullback-Leibler divergence $KL(Q\|P) = \mathbb{E}_{z\sim Q}[\log(Q) - \log(P)]$ which penalizes differences between the posterior distribution Q and the prior distribution P. Both losses are combined as a weighted sum with a weighting factor β.

$$\mathcal{L}(Y, X) = \mathbb{E}_{z\sim Q(\bullet|Y,X)}\Big(- \log P_c\big[Y \mid S(X, z)\big]\Big) + \beta \cdot KL\Big(Q(z \mid Y, X) \parallel P(z \mid X)\Big) \tag{4}$$

The training is done from scratch with randomly initialized weights. During training, this KL loss «pulls» the posterior distribution (which encodes a segmentation variant) and the prior distribution towards each other. On average (over multiple training examples) the prior distribution will be modified in a way such that it «covers» the space of all presented segmentation variants for a specific input image.

3 Experiments

3.1 Data

We used LITS data set, containing CT scans of abdominal area, where can be two classes besided backgorund — liver and lesion. It is worth saying, the initial dataset was big enough, so we used only some subset of the dataset.

3.2 Implementation

We set hyperparameters as it was proposed in article: latent dimensionality $N_{latent} = 6$, $\beta = 1$ (balance between losses).

3.3 Results

We implemented Probabilistic UNet proposed by authors of article. Results of sampling from latent space is presented on figures 4. Some numerical results can be found at table 1.

The probabilistic UNet was implemented with PyTorch 4.0.

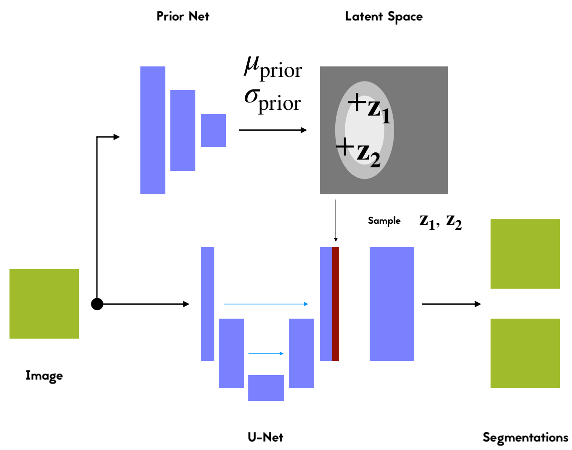

Figure 2: Sampling process

	Label 1	Label 2
Mean IoU	0.81	0.62

Table 1: Metrics of segmentation

Original UNet architecture was adapted from https://github.com/jaxony/unet-pytorch

Usual UNet shows better perfomance in terms of IoU metric: (0.88 for label 1 and 0.79 for label 2), but it is normal process, since the better perfomance could be achieved only if all generated samples would the identical to ground-truth mask.

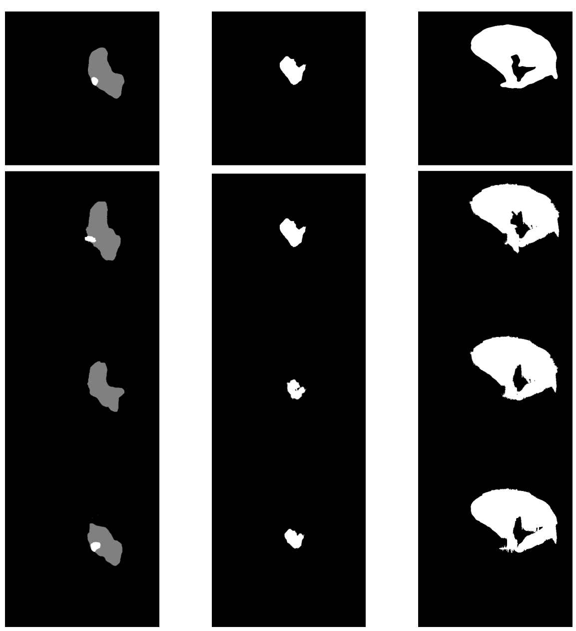

Figure 3: Sampling examples

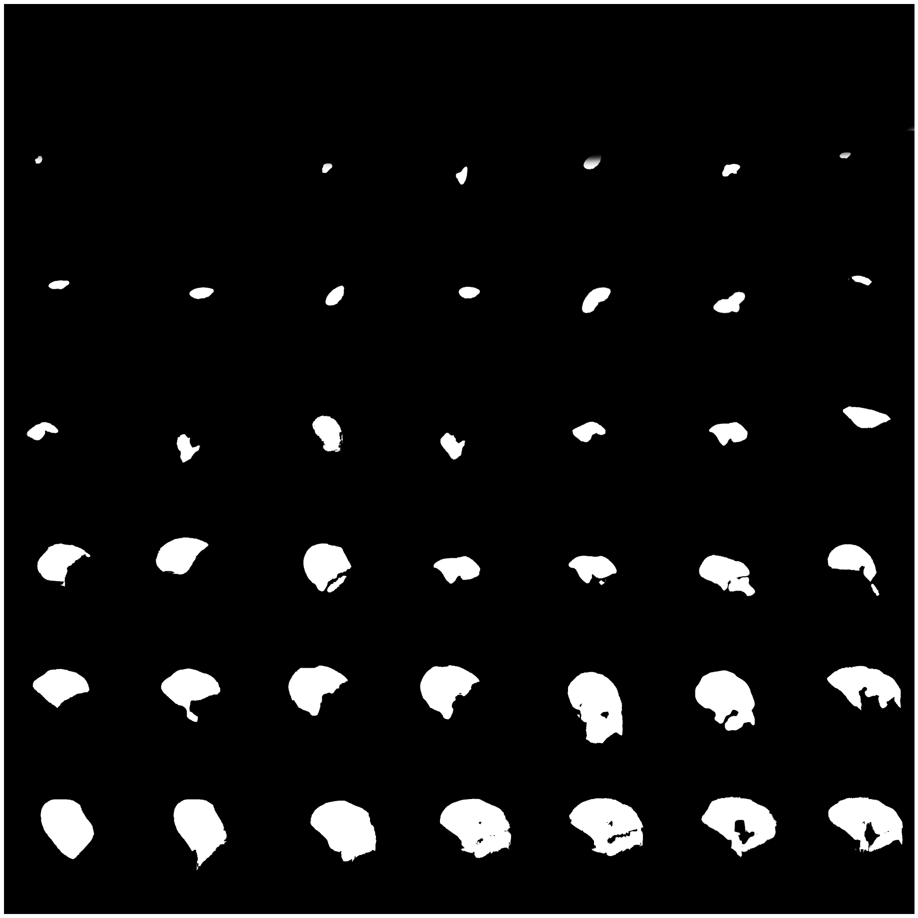

Figure 4: Latent space visualization for LITS dataset. Dimensionality of latent space = 2

4 Contributions

Ivan Rodin:

1. training procedure
2. axis aligned gaussian
3. convolutional decoder (f_{comb})

Vladislav Shlyonskiy:

1. dataloader
2. ProbUnet class
3. convolutional encoders

Both: reading the article, analysis of theory.

References

[1] S. A. A. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. H. Maier-Hein, S. M. A. Eslami, D. Jimenez Rezende, and O. Ronneberger. A Probabilistic U-Net for Segmentation of Ambiguous Images. *ArXiv e-prints*, June 2018.

[2] Liver tumor segmentation challenge.