Adapting Probabilistic U-Net for Midline Shift Detection

Alim Bukharaev[1,2], Maxim Pisov[1,3], and Mikhail Belyaev[1,3]

[1] Institute for Information Transmission Problems
[2] Moscow Institute of Physics and Technology
[3] Skolkovo Institute of Science and Technology
bukharaev.an@phystech.edu

Abstract. Probabilistic U-Net is a novel deep-learning approach for ambiguous image segmentation tasks, proposed by [3]. In this paper we adapt it to a totally different problem - midline shift detection, which consists in drawing the curve that separates the brain hemispheres on a given MRI scan. We compare the Probabilistic U-Net with a plain U-Net and evaluate its ability to learn a meaningful latent space.

Keywords: neural networks, u-net, variational autoencoder, midline shift

1 Introduction

In many supervised machine learning tasks researches have no access to ground truth data, but rather to a number of annotations that are merely expert *opinions*. This is especially true for problems with potentially high inter-expert variability such as segmentation.

Currently the state-of-the-art approach for segmentation with ambiguous annotations is Probabilistic U-Net, proposed in [3]. Inspired by variational autoencoders [2], the authors proposed a similar generative neural network, that predicts a *distribution* of segmentations for each input image.

Among other important medical imaging problems is object delineation. For example the importance of brain midline shift (Fig. 1) estimation and the need for its automation was recently highlighted by The American College of Radiology Data Science Institute [5].

In this article we adapt the approach from [3] for the task of brain midline detection, and compare its performance with a simple U-Net [8] with the same architecture.

2 Problem

Given a series of axial MRI scans of the human brain, estimate the curve that separates the hemispheres - the brain midline (Fig. 1).

We suppose that apart from the input series, there is a set of images annotated by a number of experts - a training set.

Note that the number of annotations for each image may vary significantly, and that the training dataset contains dubious cases with high inter-expert variability.

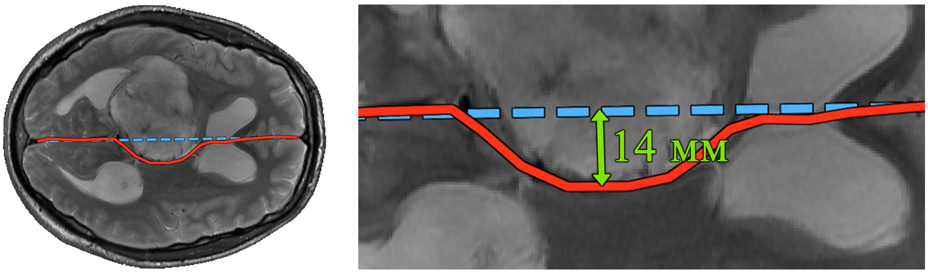

Fig. 1: An axial slice from an MRI series. The brain midline is denoted by a red curve. On the left image the green arrow shows the midline shift - the deviation from a hypothetical normal midline (blue dashed line).

3 Related work

In recent years, methods capable of producing a limited number of hypotheses were introduced such as M-Heads [4] and U-Net Ensemble [1]. In short, first one is a U-Net with M "heads" branching off the last layer, each producing an output variant. U-Net ensemble is a set of U-Nets each responsible for its own answer. Both methods are capable of generating only a predefined number of predictions, which is not applicable for our task as a number of annotations may vary from image to image. On the other hand, Probabilistic U-Net is able to generate a non-constant number of output variants, so we chose this method for our experiments.

4 Method

Both U-Net and Probabilistic U-Net were designed for segmentation tasks. However, we can reduce the task of midline detection to a segmentation problem using the method described in [7]. In a standard setting (with the sigmoid activation function and binary cross entropy loss) the output can be understood as independent probabilities of each pixel to be on the midline. Instead, we applied softmax along the x axis of the output of the U-Net and then calculated expectation along the same axis.

As in [3], we defined the loss function as a sum of mean squared error and Kullback-Leibler divergence (KL) multiplied by a hyperparameter β:

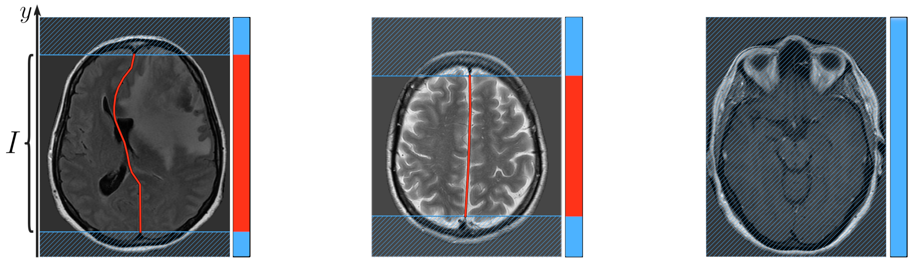

Fig. 2: The binary masks of the regions where the midline is defined (red). Note the rightmost image, for which the midline is undefined everywhere.

$$Loss = \frac{1}{|I|}\sum_{y \in I}(\hat{x}_y - x_y)^2 + \beta \cdot KL(\mu_{post}, \sigma_{post}||\mu_{prior}, \sigma_{prior}) \tag{1}$$

where I is the set of coordinates for which the midline is defined (Fig.2).

The training and inference process is shown in Fig.3 and Fig.4 respectively. For a detailed architecture of the used model please refer to the Appendix. It is worth mentioning that inspired by [6] we used residual blocks instead of simple convolutions in our implementation of U-Net.

5 Experimental Setup

In order to prevent numeric overflow in the loss function, we had to keep β small enough ($\beta = 10^{-4}$ in all tests). After a considerable number of experiments we observed that the value of beta, if kept reasonably low, does not affect the training process. Learning rate was also set to 10^{-4} in all the experiments.

Also, we noticed that increase in number of U-Net output channels enhances model's performance. However, because of computational limitations we had to keep it to 16. We also grid searched the best number of latent space dimensions (from 2 to 18, some steps omitted, see Fig.5). Surprisingly, we got the same 6 dimensions as in [3]. The results were compared with a vanilla U-Net, 5-fold cross-validation was applied.

Adam optimizer was used in all the experiments.

6 Dataset

The dataset consists of 352 MRI series that belong to patients with brain tumors. The dataset was labeled by an experienced radiologist (E1) and three specialists with limited experience in neuroradiology (E2-E4).

The series contain axial slices with various voxel spacial sizes ranging from $0.2\times0.2\times1$mm to $1\times1\times5$mm, and modalities: T1 (25%), T2 (68%) and FLAIR (7%). The images were collected using GE and Siemens scanners.

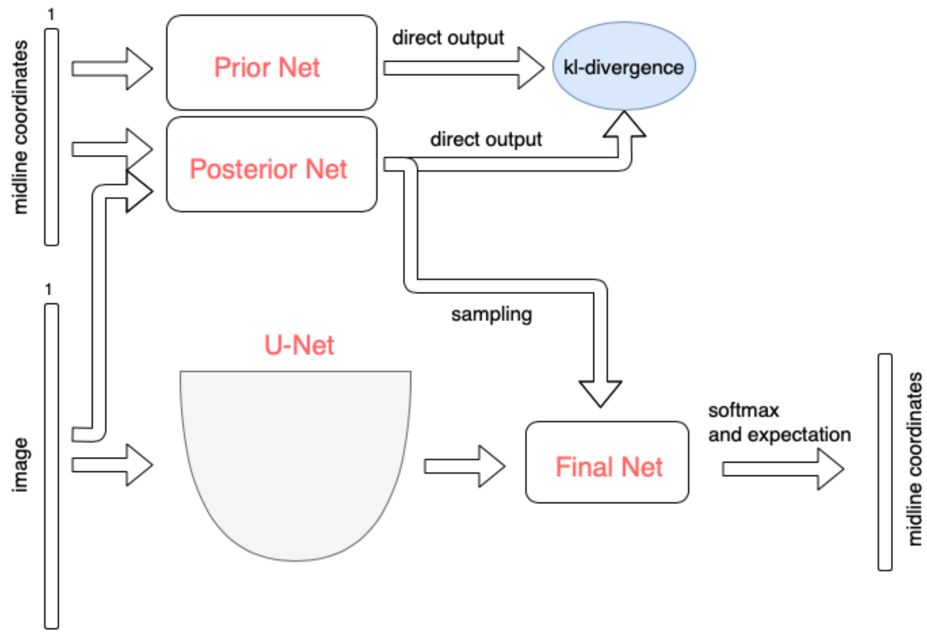

Fig. 3: An illustration of the training step.

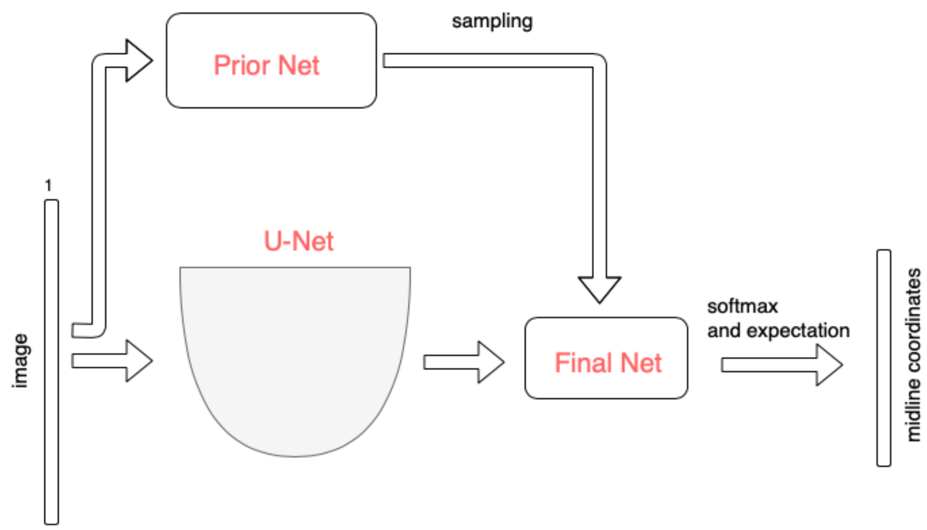

Fig. 4: An illustration of the inference step.

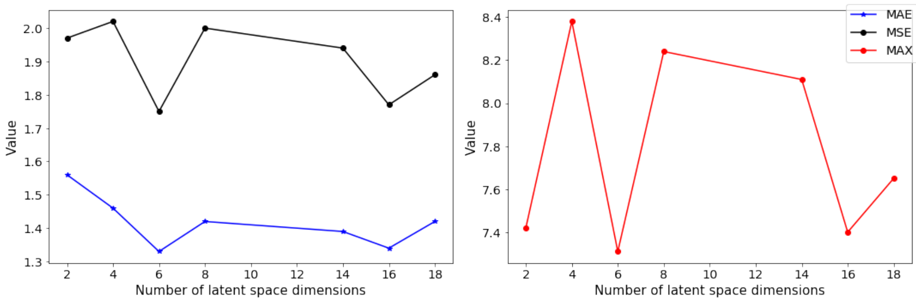

Fig. 5: Dependence of the metric value on the dimension of the latent space. The 6-dimensional space turned out to be the best.

7 Results

As it was aforementioned, we compared results of plain and Probabilistic U-Nets using 5-fold cross-validation. It is clear from Tab.1 that the difference in performance of the two models is insignificant. However, what we wanted to know much more is whether the model is capable of generating a representative latent space. Unfortunately, in this it failed, as even very distant points in latent space correspond to almost identical curves. This is illustrated perfectly in the picture bellow.

As for the metrics, we used mean absolute error (MAE), root-mean-square error (RMSE) and maximal error (MAX):

$$MAE(\hat{x}, x) = \frac{1}{I} \sum_{y \in I} |\hat{x}_y - x_y|$$

$$RMSE(\hat{x}, x) = \sqrt{\frac{1}{I} \sum_{y \in I} |\hat{x}_y - x_y|^2}$$

$$MAX(\hat{x}, x) = \max_{y \in I} |\hat{x}_y - x_y|$$

Table 1: Midline estimation metrics (± std) calculated on 5-fold cross-validation.

	MAE	RMSE	MAX
Plain U-Net	0.87 ± 0.68	1.24 ± 0.95	5.49 ± 3.45
Probabilistic U-Net	0.87 ± 0.61	1.25 ± 0.88	5.48± 3.37

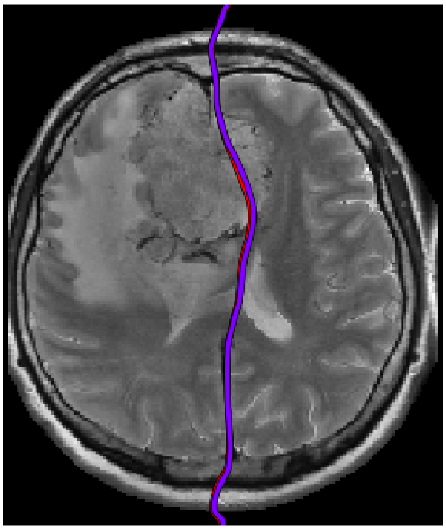

Fig. 6: Two midlines are drawn in this picture. The purple one corresponds to the median of the latent space of the image. The red one deviates from it by 20 standard deviations. These curves, that should be totally different from each other, yet almost coinciding, illustrate that our model failed to learn variational features.

8 Discussion

The problem with the latent space could be explained by the relatively low variability of the original data. Nevertheless, we are planning to continue our research. In most of the successful applications of variational autoencoders output of the "encoding" net is transmitted right to the bottle-neck of the convolutional net, not to its last layers. This is why we feel that earlier mixing is really worth trying and could potentially teach the model to generate the proper latent space.

It is also possible that simple concatenation of the outputs of the Prior Net and the U-Net is not applicable for this task and a more complicated method should be used.

References

1. Guzmán-rivera, A., Batra, D., Kohli, P.: Multiple choice learning: Learning to produce multiple structured outputs pp. 1799–1807 (2012), http://papers.nips.cc/paper/4549-multiple-choice-learning-learning-to-produce-multiple-structured-outputs.pdf
2. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2013), http://arxiv.org/abs/1312.6114, cite arxiv:1312.6114

3. Kohl, S.A.A., Romera-Paredes, B., Meyer, C., Fauw, J.D., Ledsam, J.R., Maier-Hein, K.H., Eslami, S.M.A., Rezende, D.J., Ronneberger, O.: A probabilistic u-net for segmentation of ambiguous images. CoRR abs/1806.05034 (2018), http://arxiv.org/abs/1806.05034
4. Lee, S., Purushwalkam, S., Cogswell, M., Crandall, D.J., Batra, D.: Why M heads are better than one: Training a diverse ensemble of deep networks. CoRR abs/1511.06314 (2015), http://arxiv.org/abs/1511.06314
5. McGinty, G.B., Allen, B.: The acr data science institute and ai advisory group: harnessing the power of artificial intelligence to improve patient care. Journal of the American College of Radiology 15(3), 577–579 (2018)
6. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. CoRR abs/1606.04797 (2016), http://arxiv.org/abs/1606.04797
7. Pisov, M., et al.: Incorporating task-specific structural knowledge into cnns for brain midline shift detection. Preprint (2019)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015), http://arxiv.org/abs/1505.04597

A Network architectures

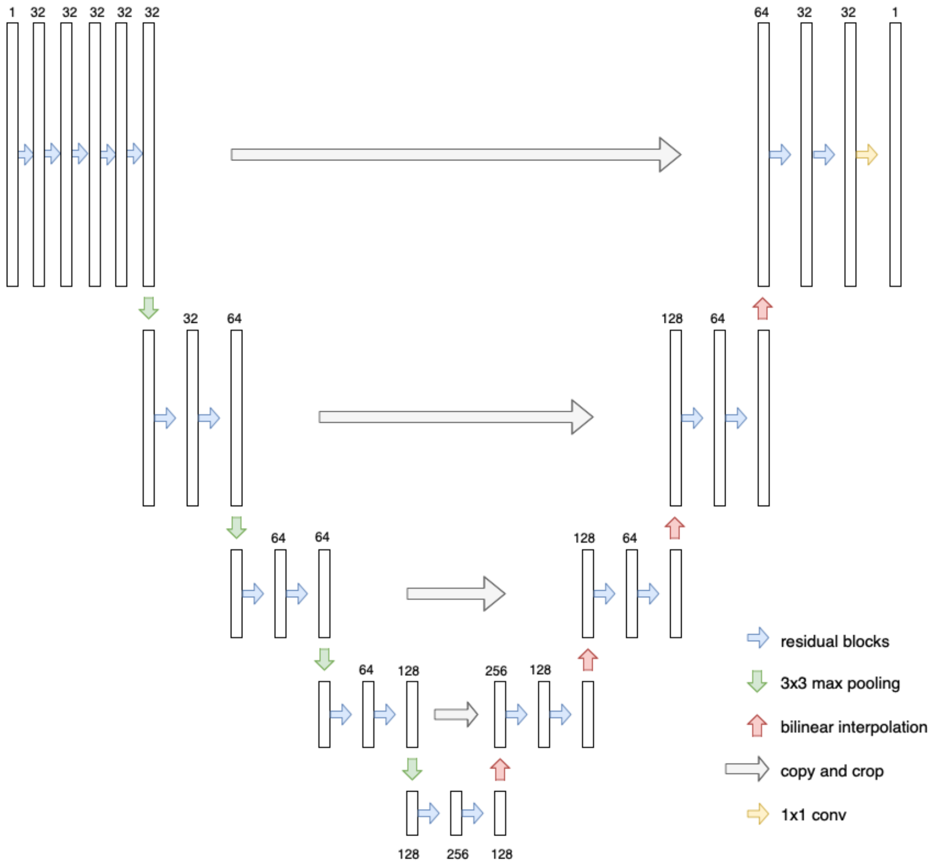

Fig. 7: U-Net architecture.

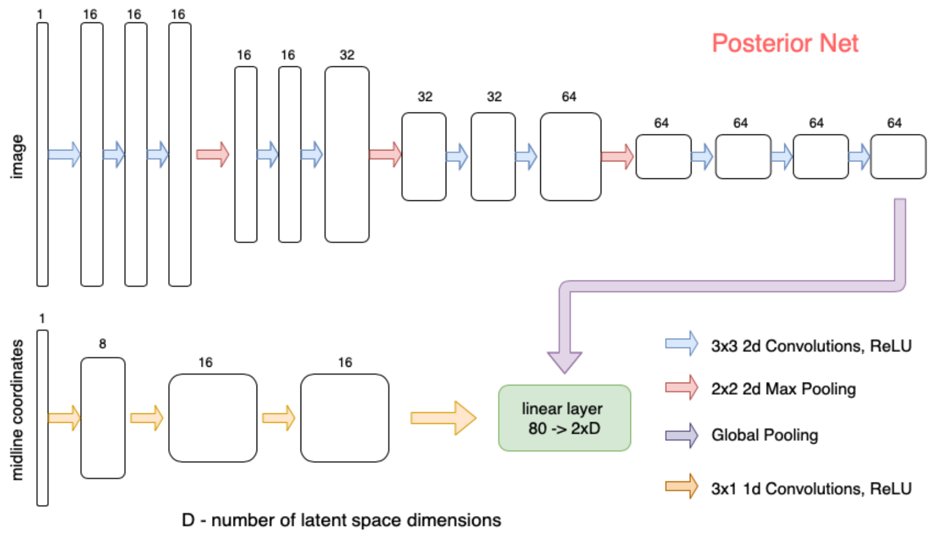

Fig. 8: Posterior Net architecture.

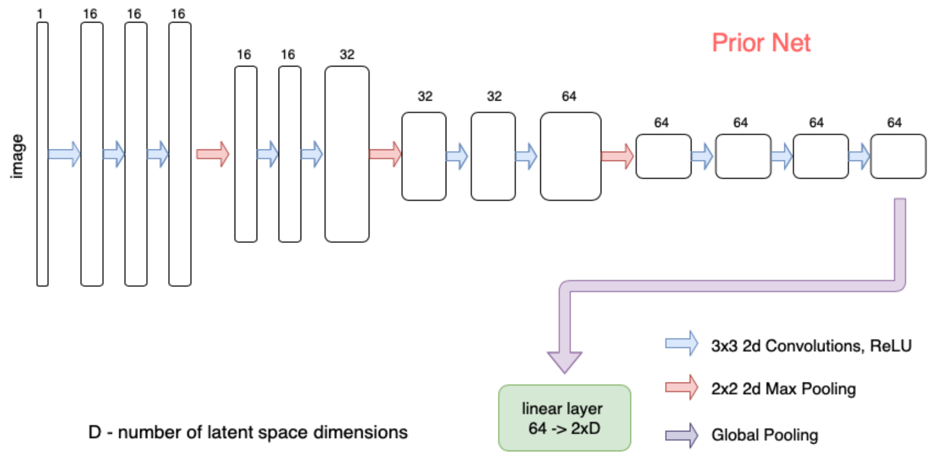

Fig. 9: Prior Net architecture.

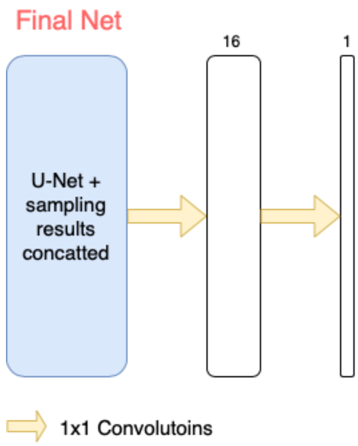

Fig. 10: Final Net architecture.