MA Draft

Josef Lorenz Rumberger

Humboldt University Berlin

1 Contribution

To help janelia research scientist with their segmentation of *drosophila melanogaster* neurons, an instance segmentation algorithm is applied. In [13], the authors describe a desirable segmentation pipeline for the processing of neuron images: first a fast instance segmentation algorithm should be applied. Which is followed by an error detection mechanism, that queries automatic correction via e.g. flood-filling networks or a manual correction by an annotator. Instead of error detection, I propose a probabilistic model as the backbone for the instance segmentation step. The probabilistic model should output local uncertainty estimates. This spatial information can then be used for the error correction step.

2 Data/Preprocessing

The dataset consists of several raw 3D volumes of microscopic images of the *drosophila melanogaster* central nervous system with dimensions ranging from $349 \times 680 \times 680$ to $401 \times 982 \times 895$ with voxel intensities measured in three channels. The ground-truth consists of volumes which assign every voxel that is part of a neuron the associated neuron ID. Therefore the dataset poses an instance segmentation problem. The preprocessing is carried out as proposed in [8].

3 Instance Segmentation Algorithm

To reduce the spatial information that is lost when splitting the volumes into smaller subvolumes, a segmentation algorithm that is parameter efficient is necessary. Therefore several instance segmentation algorithms are reviewed: region of interest (ROI) based algorithms like Faster R-CNN[18], Mask R-CNN[7] and Retina Net[14][9] all propose bounding boxes for instance segmentation and combine these with prediction masks obtained by semantic segmentation algorithms. These algorithms often have problems with thin and long structures [4]. Since neurons often build intertwined structures and are typically thin and long, bounding boxes will likely fail on our dataset.
Metric learning methods are introduced to overcome this limitation [4][2]: The general idea is to learn an embedding space, where pixels that correspond to the

same object instance are near each other and pixels that correspond to other object instances are far. This is implemented as a CNN that gets a slice of the volume as input and predicts for every voxel a vector that marks a point in the embedding space. The loss function of this embedding is based on some distance measure (e.g. euclidean distance), that accounts for the distance between each pair of voxels. To construct a segmentation, seed voxels are chosen based on some heuristic[4] or queried via user inputs[2] and all voxels that are sufficiently near to the seed in the embedding space are segmented as a single instance. Another idea is to use a clustering algorithm on the embeddings and interpret each cluster as an instance, [3] use a variant of the mean shift algorithm to do so.

I will review more literature and try to dig deeper into both, metric learning for instance segmentations. I will choose one instance segmentation model and compare the different probabilistic backbones for it.

4   Probabilistic Models and Uncertainty

The above described segmentation models, based on metric learning, all use semantic segmentation algorithms as a backbone. A branch of semantic segmentation models incorporate ideas from probabilistic modelling, where models predict a distribution over labels and thus are capable of quantifying the local uncertainty of the model parameters. Furthermore, probabilistic computer vision algorithms are for some benchmark datasets the current state of the art [17]. Kendall and Gal [10] mention that predictive uncertainty consists of two kinds of uncertainties. Epistemic (or model) uncertainty, that arises from the uncertainty about the model parameters. Aleatoric uncertainty arises from perturbations in the data and can therefore be modelled as a function of the data. They stress that epistemic uncertainty is important for safety critical applications and training on small datasets. Therefore I'll focus on models that capture epistemic uncertainty but also try to model aleatoric uncertainty.

4.1   Modelling Epistemic Uncertainty

A number of candidate probabilistic backbones which can model aleatoric uncertainty are reviewed:

Bayes by Backprop: Blundel et al. [1] propose a model where each parameter w is a random variable. By employing the reparametrization trick (here for a gaussian random variable):

$$w \sim \mathcal{N}(\mu, \sigma) = \mu + \sigma \cdot \epsilon, \text{ with } \epsilon \sim \mathcal{N}(0, 1) \tag{1}$$

the model can be trained via gradient descent based optimization techniques. Unfortunately the model shows slightly lower performance than comparable models and the number of parameters doubles due to the center and scale parameters for every parameter, which makes it impractical for the task at hand. Nevertheless it is a good starting point to get into the topic of probabilistic models.

Deep Ensembles: Lakshminarayanan et al. [12] propose the use of ensembles of random initialized and independently trained frequentist models to gather multiple predictions. The empirical mean and standard deviation of these models is then interpreted as their prediction and the associated uncertainty. I could incorporate the idea and form an ensemble of multiple probabilistic models and analyze their accuracy and uncertainty in comparison to the baseline models.

MC Dropout Gal and Ghahramani [5] show that neural networks with dropout can be seen as approximate bayesian neural networks with Bernoulli prior. They use dropout at test time to gather a number of varying predictions for a sample and then interpret their mean and variance as prediction and uncertainty. It is shown that this does not yield good results for semantic segmentation tasks [11], but there exists a variety of other bayesian approximations through regularization [15]. I will choose one of them for comparison.

Probabilistic U-Net Kohl et al. [11] propose a U-Net architecture that injects multivariate gaussian noise parameterized by a conditional variational autoencoder into the last layer of a U-Net and compare this to other probabilistic versions of U-Net. In order to use this technique on unambiguous tasks, pseudo labels have to be introduced that split labels into sublabels. As far as I understood, this technique was just used to check if the empirical frequency of occurences of the sublabels in the predictions approximately matches their theoretical frequency in the dataset. From my perspective the model is probably not suitable for the task at hand, but this needs further investigation which could be part of the thesis.

M-Heads U-Net Kohl et al.[11] also describe a M-Heads U-Net model, that augments a U-Net by placing M parallel output layers on the network, which yield M different prediction maps. These predictions can then be used to calculate the empirical mean and variance and lead to good segmentation performance [11]. The model could be further augmented by using the Shake-Shake regularizer [6] on the M parallel heads, which is known for increasing performance for image recognition multi-branch networks. This regularizer draws M random numbers from a categorical distribution with equal probabilities, multiplies the predictions of the M branches with these random numbers and sums them up to a final prediction. New random numbers are drawn for the back-propagation step, so that each head gets a noisy gradient, leading to more decorrelated feature maps [6]. As in [5], the regularizer could be used at test time to draw multiple different predictions. According to the central limit theorem, these predictions should be normally distributed. The shake-shake regularizer will be tested for this model and eventually included in the thesis.

4.2   Modelling Aleatoric Uncertainty

Aleatoric Uncertainty can be modelled as a function of the input data. Therefore models that predict the parameters of a distribution can be used for that task. The essential idea is to modify the objective function, such that the segmentation error gets weighted by the inverse of the predicted variance, while also penalizing high variances via a regularization term [10]. I have not found a paper, that uses it on instance segmentation tasks and have to do further research into this area. I will review more literature, among others the following paper and look for ways to modify the objective function, so that I can infer the aleatoric uncertainty.

Laplace U-Net for Microscopic Image Restoration Weigert et al. [19] propose a U-Net architecture that predicts the center and scale parameters of a Laplace distribution for a given input image and show good performance on image restoration tasks. To implement the model, only the shape of the output layer of the U-Net has to be changed from one to two channels and the voxel predictions have to be drawn from a Laplace distribution.

4.3   Uncertainty Evaluation

The uncertainty estimates of the models are analyzed and compared according to [16].

5   Discussion

The annotation of the aforementioned *drosophila melanogaster* volumes requires trained experts and is therefore expensive to obtain, while more unlabeled raw data is already available. Therefore a semi-supervised approach, that is capable of incorporating knowledge from unlabeled training data is desirable and should therefore be part of the discussion. Another part that could be in the discussion is about active learning strategies that rank unlabeled data based in their informativeness and therefore could help to prioritize the work of manual annotators.

References

1. Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D.: Weight uncertainty in neural networks. arXiv preprint arXiv:1505.05424 (2015)
2. Chen, Y., Pont-Tuset, J., Montes, A., Van Gool, L.: Blazingly fast video object segmentation with pixel-wise metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1189–1198 (2018)
3. De Brabandere, B., Neven, D., Van Gool, L.: Semantic instance segmentation with a discriminative loss function. arXiv preprint arXiv:1708.02551 (2017)
4. Fathi, A., Wojna, Z., Rathod, V., Wang, P., Song, H.O., Guadarrama, S., Murphy, K.P.: Semantic instance segmentation via deep metric learning. arXiv preprint arXiv:1703.10277 (2017)

5. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059 (2016)
6. Gastaldi, X.: Shake-shake regularization. arXiv preprint arXiv:1705.07485 (2017)
7. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
8. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
9. Jaeger, P.F., Kohl, S.A., Bickelhaupt, S., Isensee, F., Kuder, T.A., Schlemmer, H.P., Maier-Hein, K.H.: Retina u-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. arXiv preprint arXiv:1811.08661 (2018)
10. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in neural information processing systems. pp. 5574–5584 (2017)
11. Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S.A., Rezende, D.J., Ronneberger, O.: A probabilistic u-net for segmentation of ambiguous images. In: Advances in Neural Information Processing Systems. pp. 6965–6975 (2018)
12. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in Neural Information Processing Systems. pp. 6402–6413 (2017)
13. Lee, K., Zung, J., Li, P., Jain, V., Seung, H.S.: Superhuman accuracy on the snemi3d connectomics challenge. arXiv preprint arXiv:1706.00120 (2017)
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
15. McClure, P., Kriegeskorte, N.: Robustly representing uncertainty through sampling in deep neural networks. arXiv preprint arXiv:1611.01639 (2016)
16. Mukhoti, J., Gal, Y.: Evaluating bayesian deep learning methods for semantic segmentation. arXiv preprint arXiv:1811.12709 (2018)
17. Rasmus, A., Berglund, M., Honkala, M., Valpola, H., Raiko, T.: Semi-supervised learning with ladder networks. In: Advances in neural information processing systems. pp. 3546–3554 (2015)
18. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
19. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., et al.: Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature methods **15**(12), 1090 (2018)