Uncertainty Driven Multi-loss Fully Convolutional Networks for Histopathology

Aïcha BenTaieb[(✉)] and Ghassan Hamarneh

Medical Image Analysis Lab, School of Computing Science,
Simon Fraser University, Burnaby, Canada
{abentaie,hamarneh}@sfu.ca

Abstract. Different works have shown that the combination of multiple loss functions is beneficial when training deep neural networks for a variety of prediction tasks. Generally, such multi-loss approaches are implemented via a weighted multi-loss objective function in which each term encodes a different desired inference criterion. The importance of each term is often set using empirically tuned hyper-parameters. In this work, we analyze the importance of the relative weighting between the different terms of a multi-loss function and propose to leverage the model's uncertainty with respect to each loss as an automatically learned weighting parameter. We consider the application of colon gland analysis from histopathology images for which various multi-loss functions have been proposed. We show improvements in classification and segmentation accuracy when using the proposed uncertainty driven multi-loss function.

1 Introduction

Although deep learning models have shown remarkable results on a variety of prediction tasks, recent works applied to medical image analysis have demonstrated improved performance by incorporating additional domain-specific information [1]. In fact, medical image analysis datasets are typically not large enough for learning robust features, however, there exist a variety of expert knowledge that can be leveraged to guide the underlying learning model. Such knowledge or cues are generally considered as a set of auxiliary losses that serve to improve or guide the learning of a primary task (e.g. image classification or segmentation). Specifically, these cues are incorporated in the training of deep convolutional networks using a multi-loss objective function combining a variety of objectives learned from a shared image representation. The combination of multiple loss functions can be interpreted as a form of regularization as it constrains search space for possible candidate solutions for the primary task.

Different types of cues can be combined in a multi-loss objective function to improve the generalization of deep networks. Multi-loss functions have been proposed for a variety of medical applications: colon histology images, skin dermoscopy images or chest X-Ray images. Chen et al. [2] proposed a multi-loss learning framework for gland segmentation from histology images in which features from different layers of a deep fully convolutional network were combined

through auxiliary loss functions and added to a per-pixel classification loss. Ben-Taieb et al. [3] proposed a two-loss objective function combining gland classification (malignant vs benign) and segmentation (gland delineation) and showed that both tasks were mutually beneficial. Additionally, authors also proposed a multi-loss objective function for gland segmentation that equips a fully convolutional network with topological and geometrical constraints [4] that encourage learning topologically plausible and smooth segmentations. Kawahara et al. [5] used auxiliary losses to train a multi-scale convolutional network to classify skin lesions. More recently, adversarial loss functions were also proposed as additional forms of supervision. Dai et al. [6] leveraged an adversarial loss to guide the segmentation of organs from chest X-Ray images. While these previous works confirm the utility of training deep networks with a multi-loss objective function, they do not clearly explain how to set the contribution of each loss.

Most existing works use an empirical approach to combine different losses. Generally, all losses are simply summed with equal contribution or manually tuned hyper-parameters are used to control the trade-off among all terms. In this work, we investigate the importance of an appropriate choice of weighting between each loss and propose a way to automate it. Specifically, we utilize concepts from Bayesian deep learning [7,8] and introduce an uncertainty based multi-loss objective function. In the proposed multi-loss, the importance of each term is learned based on the model's uncertainty with respect to each loss. Uncertainty was leveraged in many medical image analysis applications (e.g. segmentation [9], registration [10]). However, to the best of our knowledge, uncertainty was only explored for the task of image registration in the context of deep learning models for medical images. Yang et al. [11] proposed a CNN model for image registration and showed how uncertainty helps highlighting misaligned regions. Previous works did not consider automating or using uncertainty for guiding the training of multi-loss objective functions designed for medical image analysis.

We illustrate our approach on the task of colon gland analysis leveraging the multi-loss objective functions proposed in previous works [3,4]. We extend these previous works by re-defining the proposed loss functions with an uncertainty driven weighting. We linearly combine classification, segmentation, topology and geometry losses weighted by the model's uncertainty for each of these terms. In the proposed uncertainty driven multi-loss, the uncertainty captures how much variance there is in the model's predictions. This variance or noise in the predictions varies for each term and thus reflects the uncertainty inherent to the classification, segmentation, topology or geometry loss.

Our contributions in this work can be summarized as follows: (i) we show how uncertainty can be used to guide the optimization of multi-loss deep networks in an end-to-end trainable framework; (ii) we combine a series of objectives that have been shown successful for gland analysis and adapt them to encode uncertainty driven weighting; (iii) we analyze the influence of different trade-offs controlling the importance of each loss in a multi-loss objective function and draw some conclusions on the adaptability of neural networks.
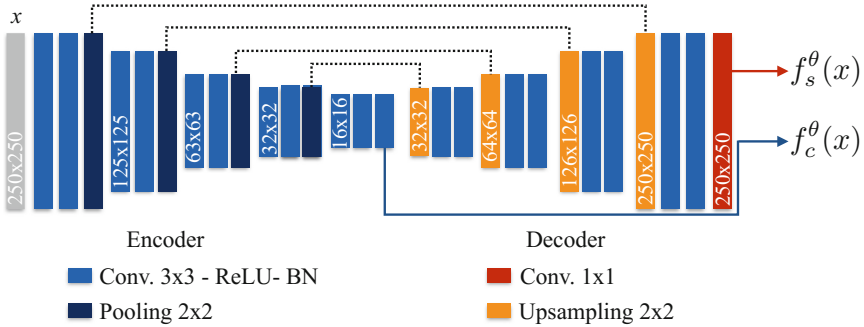
x

250x250

125x125

63x63

32x32

16x16

32x32

64x64

126x126

250x250

250x250

$f_s^\theta(x)$

$f_c^\theta(x)$

Encoder

Decoder

■ Conv. 3x3 - ReLU- BN   ■ Conv. 1x1

■ Pooling 2x2   ■ Upsampling 2x2

Fig. 1. Multi-loss network architecture. We use an encoder-decoder architecture with skip connections [12]. x is an input image. $f_c^\theta(x)$ are the activations from the last convolution layer of the encoder and are used to predict class labels (i.e. malignant vs benign tissue). $f_s^\theta(x)$ are per-pixel activations from the last convolutional layer of the decoder that are used to predict segmentations. The building blocks of the network are layers of convolution (Conv.), ReLU activation functions and batch normalization (BN). Dashed lines represent skip connections.

2   Method

Our goal is to learn how to combine multiple terms relevant to gland image analysis into a single objective function. For instance, gland classification and gland segmentation can both benefit from a joint learning framework and information about the geometry and topology of glands can facilitate learning plausible segmentations. Note that we refer to gland's geometry and topology in terms of smooth boundaries as well as containment and exclusion properties between different parts of objects (the lumen is generally contained within a thick epithelial border and surrounded by stroma cells that exclude both the lumen and the border, see Fig. 3 for an example of gland segmentation).

We train a fully convolutional network parameterized by θ, from a set of training images x and their corresponding ground truth segmentation masks S along with their tissue class label binary vector C represented by $\{(x^{(n)}, S^{(n)}, C^{(n)}); n = 1, 2, \ldots, N\}$. We drop (n) when referring to a single image x, class label C or segmentation mask S. We note K the total number of image class labels (e.g. $K = 2$ for malignant or benign tissue images of colon adenocarcinomas) and L the total number of region labels in the segmentation mask (e.g. $L = 3$ for lumen, epithelial border and stroma). The network's architecture is shown in Fig. 1. To predict class labels C, we use the network's activations $f_c^\theta(x)$ from the last layer of the encoder as they correspond to a coarser representation of x. To obtain a crisp segmentation of a color image x, we use the activations $f_s^\theta(x)$ from the last layer of the decoder and we assign a vector $S_p = (S_p^1, S_p^2, ..., S_p^L) \in \{0,1\}^L$ to the p-th pixel x_p in x, where S_p^r indicates whether pixel x_p belongs to region r, and L is the number of region labels. We assume region labels r are not always mutually exclusive such that containment

properties (e.g. glands' lumen is contained within the epithelial border) are valid label assignments.

Multi-loss networks: A multi-loss objective function is defined as follows:

$$\mathcal{L}_{total}(x;\theta) = \sum_{i=1}^{T} \lambda_i \mathcal{L}_i(x;\theta) \tag{1}$$

where θ represents the network's parameters learned by minimizing $\mathcal{L}_{total}$; T is the total number of loss functions $\mathcal{L}_i$ to minimize with respect to the network's parameters, and λ_i is a scalar coefficient controlling the importance of each loss, generally found via grid-search or set equally for all terms.

In the context of gland analysis, we define a multi-loss objective function that encodes classification, segmentation as well as gland's topology and geometry. We learn the relative weights of each term in the objective using a measure of uncertainty that reflects the amount of noise or variance in the model's predictions for each term. Using uncertainty to weight each term results in reducing the influence of uncertain terms on the total loss and hence on the model's parameters update. Formally, we write the total objective function as follows:

$$\mathcal{L}_{total}(x;\theta,\sigma_c,\sigma_s,\sigma_t,\sigma_g) = \mathcal{L}_c(x;\theta,\sigma_c) + \mathcal{L}_s(x;\theta,\sigma_s) + \mathcal{L}_t(x;\theta,\sigma_t) + \mathcal{L}_g(x;\theta,\sigma_g) \tag{2}$$

where $\mathcal{L}_c, \mathcal{L}_s, \mathcal{L}_t, \mathcal{L}_g$ are the classification, segmentation, topology and geometry loss functions and $\sigma_c, \sigma_s, \sigma_t, \sigma_g$ are learned scalar values representing the uncertainty for each loss (or amount of variance in the prediction).

Uncertainty guided classification: Similarly to Gal et al. [8], we define the classification loss $\mathcal{L}_c$ with uncertainty as:

$$\mathcal{L}_c(x;\theta,\sigma_c) = \sum_{k=1}^{K} -C_k \log P(C_k = 1|x,\theta,\sigma_c), \quad P(C_k=1|x,\theta,\sigma_c) = \frac{\exp(\frac{1}{\sigma_c^2} f_{c_k}^{\theta}(x))}{\sum\limits_{k'=1}^{K} \exp(\frac{1}{\sigma_c^2} f_{c_{k'}}^{\theta}(x))} \tag{3}$$

where K is the total number of classes, $P(C_k|x,\theta,\sigma_c)$ corresponds to the softmax function over the network's activations $f_c^{\theta}(x)$ weighted by the classification prediction's uncertainty coefficient σ_c. Note how higher values of σ_c reduce the magnitude of activations $f_c^{\theta}(x)$ over all classes (which corresponds to encouraging uniform probabilities $P(C_k|x,\theta,\sigma_c)$) and thus reflect more uncertain predictions (i.e. high activation values will be weighted lower when σ_c; the uncertainty, is high).

Assuming $\frac{1}{\sigma_c^2}\sum_k \exp\left(\frac{1}{\sigma_c^2} f_{c_k}^{\theta}(x)\right) \approx \left(\sum_k \exp(f_{c_k}^{\theta}(x))\right)^{\frac{1}{\sigma_c^2}}$ [7], we can re-write the uncertainty-guided classification loss as follows:

$$\mathcal{L}_c(x;\theta,\sigma_c) = \sum_{k=1}^{K} -C_k \log\left(\exp(\frac{1}{\sigma_c^2} f_{c_k}^{\theta}(x))\right) + \log \sum_{k'=1}^{K} \exp(\frac{1}{\sigma_c^2} f_{c_{k'}}^{\theta}(x)) \tag{4}$$

$$\approx \frac{1}{\sigma_c^2} \sum_{k=1}^{K} -C_k \log P(C_k = 1|x_p;\theta) + \log \sigma_c^2. \tag{5}$$

Note how large scale values of σ_c^2 corresponding to high uncertainty will reduce the contribution of the classification loss. The second term in Eq. (5) avoids σ_c^2 from becoming infinity and thus avoids the loss from becoming zero. We extend the above softmax with uncertainty cross-entropy classification loss to the segmentation losses.

Uncertainty guided segmentation: We learn pixel-wise predictions using a combination of a sigmoid cross entropy loss $\mathcal{L}_s$ with two higher order penalty terms (proposed in [4]): a topology loss $\mathcal{L}_t$ enforcing a hierarchy between labels and a pairwise loss $\mathcal{L}_g$ enforcing smooth segmentations.

$$\mathcal{L}_s(x;\theta,\sigma_s) = \frac{1}{\sigma_s^2} \sum_{p \in \Omega} \sum_{r=1}^{L} -S_p^r \log P(S_p^r = 1|x,\theta) + \log \sigma_s^2 \tag{6}$$

where L represents the number of regions in the segmentation mask, Ω is the set of pixels in a given image x, $P(S_p^r = 1|x,\theta,\sigma_s)$ is the output of the sigmoid function applied to the segmentation activations $f_s^\theta(x_p)$ and σ_s^2 represents the model's uncertainty for $\mathcal{L}_s$.

The topology loss defined in [4] was originally formulated as a modified softmax cross entropy loss in which the probabilities are defined to encode containment and exclusion as a hierarchy between labels. Per-pixel hierarchical probabilities are defined to penalize topologically incorrect label assignments such that their probability is set to zero. Formally, the hierarchical probabilities used to compute $\mathcal{L}_t$ are defined as:

$$P_t(S_p^r|x_p;\theta) = \frac{1}{Z} V(S_p) \prod_{r=1}^{L} \exp\left(f_{s_r}^\theta(x_p)\right) \times S_p^r, \quad Z = \sum_{r=1}^{L} \tilde{P}_t(S_p^r|x_p;\theta) \tag{7}$$

where Z is a normalizing factor, $\tilde{P}_t(S_p^r|x_p;\theta)$ is the un-normalized probability and $V(S_p)$ is a binary indicator function that identifies topologically valid label assignments ($V(S_p) = 1$) from invalid ones ($V(S_p) = 0$). Using these probabilities defined in [4] and applying the same simplification as in Eq. (5), $\mathcal{L}_t$ is formulated as the following uncertainty guided cross entropy loss where σ_t^2 is the uncertainty:

$$\mathcal{L}_t(x;\theta,\sigma_t) = \frac{1}{\sigma_t^2} \sum_{p \in \Omega} \sum_{r=1}^{L} -S_p^r \log P_t(S_p^r = 1|x,\theta) + \log \sigma_t^2. \tag{8}$$

It is worth noting that the fundamental assumption behind the sigmoid cross entropy loss $\mathcal{L}_s$ is that all segmentation labels are mutually independent whereas in the defined topology loss $\mathcal{L}_t$ inclusion and exclusion relations between the segmentation labels are set as hard constraints (i.e. enforcing containment and exclusion properties). Thus, the combination of $\mathcal{L}_s$ and $\mathcal{L}_t$ results in a soft constraint over the topology properties (as opposed to the hard constraint originally proposed in [4]).

Finally, to include uncertainty in the geometry loss, we re-define the original loss proposed in [4] such that it is weighted with an uncertainty coefficient σ_g.
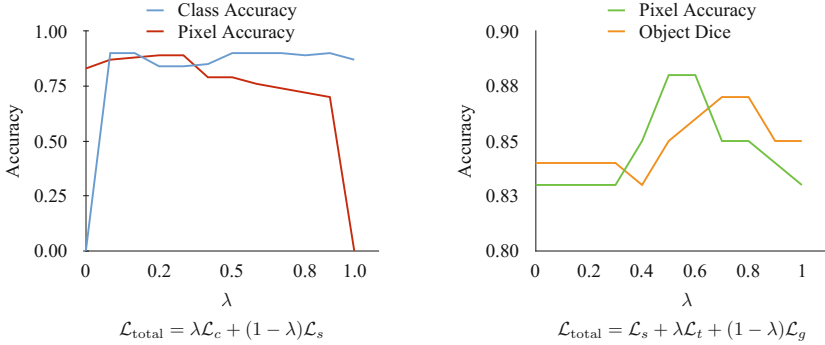
Fig. 2. Trade-off between different loss functions and influence on the network's generalization. The learning rate was kept fixed to 1e–2 in all experiments. Each graph represents the classification and segmentation accuracy on the Warwick-QU colon adenocarcinoma test set.

The geometry loss $\mathcal{L}_g$ favours smooth segmentations by minimizing the ratio of log probabilities between neighbouring pixels sharing the same labels in the ground truth segmentation.

$$\mathcal{L}_g(x; \theta, \sigma_g) = \frac{1}{\sigma_g^2} \sum_{p \in \Omega} \sum_{r=1}^{L} \sum_{q \in \mathcal{N}^p} S_p^r \left| \log \frac{P_t(S_p^r | x_p; \theta)}{P_t(S_q^r | x_q; \theta)} \right| B_{p,q} + \log \sigma_g^2 \qquad (9)$$

where $\mathcal{N}^p$ corresponds to the 4-connected neighborhood of pixel p. $\mathcal{L}_g$ trains the network to output regularized pairs of log-sigmoid label probabilities for neighbouring pixels p and q when the binary indicator variable $B_{p,q} = 1$ (i.e. when p and q share the same label in the ground truth segmentation). σ_g^2 is the uncertainty for loss $\mathcal{L}_g$. Note that in this formulation, we minimize the difference between log-probabilities so the assumption utilized in Eq. (5) still holds.

Implementation details: We implement the model using Tensorflow [13]. We train a fully convolutional architecture as describe in Fig. 1 using the proposed multi-loss function Eq. (2) optimized with stochastic gradient descent. All uncertainty parameters σ_i are learned along with the model's parameters θ. In practice, we trained the network to predict $\log \sigma_i^2$ for numerical stability [8].

3   Experiments and Discussion

We used the publicly available Warwick-QU colon adenocarcinoma dataset [14], which consists of 85 training (37 benign and 48 malignant) and 80 test images (37 benign and 43 malignant). In this dataset, each tissue image is composed of multiple glands and is labelled as benign or malignant and provided with a corresponding segmentation mask delineating each gland's lumen and epithelial border (see Fig. 3). In all experiments, we used 70 images for training, 15 for

validation and 80 for test. We extracted patches of size 250×250 pixels and used a series of elastic and affine transforms to augment the training dataset by a factor of $\sim$100. We used (image-level) classification accuracy to evaluate the model's capacity to correctly predict benign vs malignant tissue images. To evaluate the predicted segmentation masks, we used three different metrics: pixel accuracy to evaluate the accuracy in predicting a pixel as either background, lumen or epithelial border; object Dice and Hausdorff Distance to evaluate the capacity of the model in correctly identifying individual glands in an image. Object Dice and Hausdorff distance are particularly useful in evaluating the accuracy of the predicted segmentations at objects borders.

Table 1. Performance of different loss functions combined with manually tuned loss weights and uncertainty-guided weights. Results are reported on the Warwick-QU original test set.

Loss	Weights				Classification accuracy	Pixel accuracy	Object dice	Hausdorff distance
	$\mathcal{L}_c$	$\mathcal{L}_s$	$\mathcal{L}_t$	$\mathcal{L}_g$				
$\mathcal{L}_c$	1	0	0	0	0.87	–	–	–
$\mathcal{L}_s$	0	1	0	0	–	0.79	0.81	8.2
$\mathcal{L}_t$	0	0	1	0	–	0.75	0.77	8.6
$\mathcal{L}_s + \mathcal{L}_t + \mathcal{L}_g$	0	1	1	1	–	0.83	0.84	7.3
$\mathcal{L}_c + \mathcal{L}_s$	0.5	0.5	0	0	0.90	0.79	0.80	8.4
$\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_t$	0.33	0.33	0.33	0	0.94	0.78	0.80	8.4
$\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_t + \mathcal{L}_g$	0.25	0.25	0.25	0.25	0.91	0.81	0.83	7.6
$\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_t + \mathcal{L}_g$	0.1	0.6	0.22	0.08	**0.95**	**0.86**	0.85	7.1
$\mathcal{L}_c + \mathcal{L}_s$	Trained with uncertainty				**0.95**	0.78	0.80	8.4
$\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_t$					0.94	0.79	0.81	8.2
$\mathcal{L}_c + \mathcal{L}_s + \mathcal{L}_t + \mathcal{L}_g$					**0.95**	0.85	**0.87**	**7.0**

Multi-loss vs single-loss: We first tested if the combination of different loss functions without uncertainty guidance influences the classification and segmentation accuracy. We used $\mathcal{L}_{\text{total}} = \lambda\mathcal{L}_c + (1 - \lambda)\mathcal{L}_s$ and explored different values for $\lambda \in [0, 1]$. Figure 2 shows the classification as well as the per-pixel accuracy on the Warwick-QU original test set of 80 images for different values of λ. Overall, we observed that learning with multiple losses improved both segmentation and classification performance. In fact, we observed up to 3% (i.e. $\lambda = \{0.5, 0.6, 0.7\}$) increase in classification accuracy when using a combination of $\mathcal{L}_c$ and $\mathcal{L}_s$ compared to using $\mathcal{L}_c$ only (i.e. $\lambda = 1$). Similarly, for segmentation, we observed the performance improved up to 6% (i.e. $\lambda = 0.3$) in pixel accuracy when combining both losses compared to using $\mathcal{L}_s$ only (i.e. $\lambda = 0$). A similar result is shown in Table 1 when comparing $\mathcal{L}_c$ vs $\mathcal{L}_c + \mathcal{L}_s$ with equal weights.

Penalty terms trade-off: We also tested the trade-off between the topology and geometry soft constraints when combined with the segmentation loss. We

used different weighting coefficients λ and trained the network with $\mathcal{L}_{\text{total}} = \mathcal{L}_s + \lambda\mathcal{L}_t + (1 - \lambda)\mathcal{L}_g$. We only varied the importance of the soft constraints. It is interesting to note that there is a wide range of weighting coefficients for which the network produces similar (or almost identical) results. In fact, we observed a minimal change ($\leq 1e$–2) when varying the importance of each term by $\pm 20\%$ around $\lambda = 0.5$, which reflects the flexibility of deep networks to adapt to different regularization terms. We also observed that generally sigmoid cross entropy loss $\mathcal{L}_s$ was more stable than $\mathcal{L}_t$ or $\mathcal{L}_g$-only and outperformed these other losses when each of them was used alone (see Table 1, $\mathcal{L}_s$ only vs $\mathcal{L}_t$ only). However, for certain weighting configurations for each penalty term, we observed improved performance (up to 5%, see Fig. 2) in terms of pixel accuracy and object Dice (e.g. $\lambda = 0.1$ vs. $\lambda = 0.5$).

Uncertainty driven trade-off: To evaluate the utility of using uncertainty to guide the trade-off between the different loss functions, we tested different combinations of losses with uncertainty to form the total multi-loss function. Table 1 shows the performance of each tested loss configuration in terms of class accuracy, pixel accuracy, object Dice and Hausdorff distance. Overall, adding uncertainty to weigh each loss achieves competing results with other strategies (e.g. equally weighted losses) and can even outperform the best set of weights we could find using a finer grid search (in terms of classification accuracy, object Dice and Hausdorff Distance, see Table 1). Note that finding the best set of weights shown in Table 1 involved training more than 30 networks with different weights for each loss whereas using the proposed uncertainty driven weights only involved training a single network. Examples of the segmentation predictions obtained using the proposed method (Eq. 2) are shown in Fig. 3.
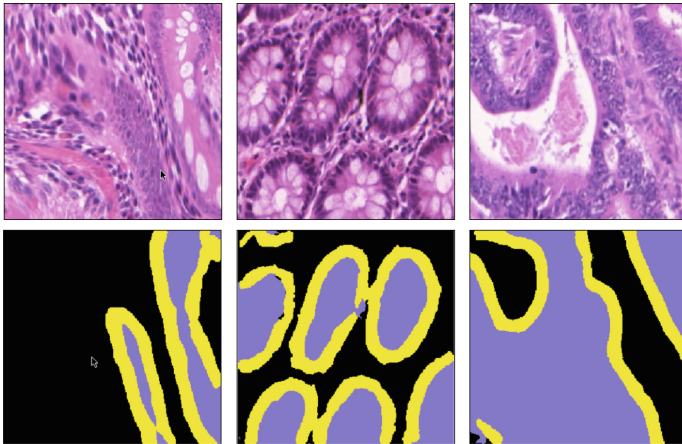

Fig. 3. Examples of predicted segmentations. Colors on the segmentation masks represent gland's central area or lumen (purple), the epithelial border surrounding the lumen (yellow) and the stroma or background (black). (Color figure online)

4   Conclusion

We showed that the combination of different loss terms with appropriate weighting can improve model generalization in the context of deep neural networks. We proposed to use uncertainty as a way to combine multiple loss functions that were shown useful for the analysis of glands in colon adenocarcinoma and we observed that this strategy helps improve classification and segmentation performance and can thus bypass the need for extensive grid-search over different weighting configurations. An interesting extension to our work could be to introduce per-instance uncertainty (as opposed to per-loss) which may be useful in situations where the data or labels are noisy.

References

1. Litjens , G., et al.: A survey on deep learning in medical image analysis. arXiv preprint arXiv:1702.05747 (2017)
2. Chen, H., Qi, X., Yu, L., Heng, P.-A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487–2496 (2016)
3. BenTaieb, A., Kawahara, J., Hamarneh, G.: Multi-loss convolutional networks for gland analysis in microscopy. In: IEEE 13th International Symposium on Biomedical Imaging, pp. 642–645 (2016)
4. BenTaieb, A., Hamarneh, G.: Topology aware fully convolutional networks for histology gland segmentation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 460–468. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_53
5. Kawahara, J., Hamarneh, G.: Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In: Wang, L., Adeli, E., Wang, Q., Shi, Y., Suk, H.-I. (eds.) MLMI 2016. LNCS, vol. 10019, pp. 164–171. Springer, Cham (2016). doi:10.1007/978-3-319-47157-0_20
6. Dai, W., et al.: Scan: structure correcting adversarial network for chest x-rays organ segmentation. arXiv preprint arXiv:1703.08770 (2017)
7. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. arXiv preprint arXiv:1705.07115 (2017)
8. Gal, Y.: Uncertainty in deep learning, Ph.D. dissertation (2016)
9. Saad, A., Möller, T., Hamarneh, G.: Probexplorer: uncertainty-guided exploration and editing of probabilistic medical image segmentation. Comput. Graph. Forum 29(3), 1113–1122 (2010). Wiley Online Library
10. Marsland, S., Shardlow, T.: Langevin equations for landmark image registration with uncertainty. SIAM J. Imaging Sci. 10(2), 782–807 (2017)
11. Yang, X., Kwitt, R., Niethammer, M.: Fast predictive image registration. In: Carneiro, G., et al. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 48–57. Springer, Cham (2016). doi:10.1007/978-3-319-46976-8_6
12. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for scene segmentation. IEEE Trans. Pattern Anal. Mach. Intell. (2017)
13. Abadi, M., et al.: Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)
14. Sirinukunwattana, K., et al.: Gland segmentation in colon histology images: the GlaS challenge contest. Med. Image Anal. 35, 489–502 (2017)