Authors' accepted manuscript. Paper presented at the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (Berlin, Germany, 23-27 July 2019) under the title “Epistemic Uncertainty Modeling for Vessel Segmentation”. The final published version is available at https://doi.org/10.1109/EMBC.2019.8857785

Epistemic uncertainty modeling using Bayesian convolutional networks for vascular structure segmentation

Rémi Martin[1], Joaquim Miró[2] and Luc Duong[3]

***Abstract*— X-ray angiograms are currently the gold-standard in percutaneous guidance during cardiovascular interventions. However, due to lack of contrast, to overlapping artifacts and to the rapid dilution of the contrast agent, they remain difficult to analyze either by cardiologists, or automatically by computers. Providing, a general yet accurate multi-arteries segmentation method along with the uncertainty linked to those segmentations would not only ease the analysis of medical imaging by cardiologists, but also provide a required pre-processing of the data for tasks ranging from 3D reconstruction to motion tracking of arteries. The proposed method has been validated on clinical data providing an average accuracy of 94.9%. Additionally, results show good transposition of learning from one type of artery to another. Epistemic uncertainty maps provide areas where the segmentation should be validated by an expert before being used, and could provide identification of regions of interest for data augmentation purposes.**

I. INTRODUCTION

Congenital heart diseases are, to date, one of the major causes of death in the world. In Canada alone, one birth over a hundred is affected by a congenital heart disease and that number is growing every year [1]. For instance, congenital stenosis - a narrowing of the vascular structure developed by the fetus - affects the blood flow of the patient and, when left untreated, can result in severe consequences. When possible, interventional cardiologists perform angioplasty. It consists in dilating the narrowed part of the artery using an inflatable balloon, and when required, a stent is placed. The current gold-standard imaging technique used during percutaneous interventions is X-ray angiography. It consists in acquiring X-ray images of the patient's artery while, at the same time, injecting a contrast agent, such as iodine, using a catheter to reveal the topology of vascular structures. However, the analysis of vascular structures from X-ray angiography is challenging due to the rapid dilution of the contrast agent, the superposition of complex vessel topologies and the radiographic noise. Recently, augmented and virtual reality technologies have been introduced to the medical community [2], [3], [4]. Those techniques rely on the use of an accurate 3D modeling of the region of interest. Most 3D reconstruction techniques require the use of prior segmentation [5], hence rendering the reconstruction heavily dependent on the accuracy of the segmentation.

[1]Rémi Martin is with the Department of Software and IT Engineering, École de technologie Supérieure, Montréal, Canada remi.martin.1@ens.etsmtl.ca

[2]Joaquim Miró is with the Department of Pediatrics, CHU Sainte-Justine, Montréal, Canada joaquim_miro@ssss.gouv.qc.ca

[3]Luc Duong is with the Department of Software and IT Engineering, École de technologie Supérieure, Montréal, Canada luc.duong@etsmtl.ca

Vessel segmentation from X-ray angiography has been a prolific area of studies. The most popular approach is based upon the analysis of the Hessian matrix of the X-ray images - known as vesselness filters, to enhance elongated vessel-like structures [6]. Recent works on blood vessel segmentation from angiograms focus on the use of graph-based methods [7] or convolutional neural networks (CNN) [8]. However, most of those studies focus on the segmentation of only one type of artery and none, to our knowledge, tries to generalize the learning to multiple different types. Medical imaging analysis is often impacted by a shortage of available data. We believe that mixing different types of arteries, and not only specializing the learning to one, is an efficient way to perform data augmentation that can be particularly efficient on learning the textures and patterns contained in the background.

Even if CNN provide state-of-the-art results in numerous domains [9], including medical imaging, the output often consists in just a prediction and no measure of confidence regarding the said-prediction is given. Recent approaches towards Bayesian convolutional neural networks [10] allow for the modeling of uncertainties. There are two uncertainties that one can model: aleatoric uncertainty and epistemic uncertainty. The former gives information about the uncertainty contained in the input while the later provides uncertainty values regarding what has and, particularly, what has not been learned by a network. Providing information about how confident the model is can, specifically in a medical task, provide a way to raise awareness on predictions that are not confidently enough predicted by the network.

For those reasons, we wish to provide a method to automatically segment different types of arteries while also providing a measure of uncertainty that reflects the correctness of the yielded predictions. Hence, the contributions of this paper are threefold:

- Automatic segmentation of various blood vessels: aorta, coronary arteries, and pulmonary arteries,
- Pixel-wise epistemic uncertainty measurements in the context of blood vessel segmentation,
- Calculation of pixel-weights designed to optimize the distribution of uncertainty by pushing towards its correlation with the correctness of the predictions.

II. METHODOLOGY

A. Dataset and pre-processing

A dataset of 368 images from X-ray angiography was acquired in our institution during various cardiac interventions

and annotated by an experienced technician. This study was approved by our IRB. The images contain three different types of arteries: aorta, coronary arteries and pulmonary arteries. All the images have been assigned to 3 subsets: training set, validation set and test set. The configuration of the three subsets is the following:

- Training set: 25 different patients representing 254 images in total. Half of those images contain only background information. The other half is composed of: 50 aortas, 27 pulmonary arteries and 50 coronary arteries.
- Validation set: 26 images in total. This set is composed of 9 aortas, 8 pulmonary arteries and 9 coronary arteries.
- Test set: 88 images in total extracted from 10 patients. Amongst those 10 patients 5 are different from those of the training set. 34 images contain aortas, 24 images contain pulmonary arteries and 30 contain coronary arteries.

Several steps of data augmentation have been performed including horizontal flipping and random cropping.

B. Uncertainty modeling

As mentioned in the introduction, epistemic uncertainty is directly yielded by what the model was and was not capable of learning. On the other hand, aleatoric uncertainty represents the amount of noise already contained in the original images. The objective is to obtain a map showing where the classification might be failing according to the model's beliefs. The uncertainty modeling is done solely using epistemic uncertainty. By activating dropout at test time, and performing Monte-Carlo sampling over the predictions yielded by the network, the uncertainty is calculated using the entropy on the average prediction probabilities (eq. 1) [10].

$$U(\hat{y}) = \frac{1}{T}\sum_{t}^{T} H(\hat{y}_t) \tag{1}$$

With T representing the number of Monte Carlo sampling and H Shannon's entropy defined as follows:

$$H(\hat{y}) = \sum_{c}^{C} \hat{y}_c \log(\hat{y}_c) \tag{2}$$

In which C represents the number of classes.

C. Pixel-weighting

We propose a pixel-weighting technique designed to adapt the distribution of uncertainty and make it correlate more with the correctness of the classification.

For that purpose we propose the following equation to dynamically define a weight for each pixel of the dataset during the training.

$$\omega_i = 1 + \bar{U}(\hat{y}_i) \times \begin{cases} \alpha, & \text{if the pixel belongs in FP} \\ \beta, & \text{if the pixel belongs in FN} \\ 0, & \text{else} \end{cases} \tag{3}$$

With i representing one pixel in the dataset and $\bar{U}(\hat{y}_i) = 1 - U(\hat{y}_i)$. FN and FP are respectively defined as the sets of False Positive and False Negative pixels. The constants α and β are used to define a compromise between FN and FP pixels and hence adapt the solution yielded by the model. The weighting increases the loss of the wrongly classified pixels by applying a penalty based on the complement of the uncertainty.

Because the calculation of the epistemic uncertainty is a complex time-consuming process, we propose to approximate that uncertainty, during the training only, by removing the costly Monte Carlo sampling. Hence our formulation of the uncertainty, during the training process, becomes:

$$U(\hat{y}) = H(\hat{y}) = \sum_{c}^{C} \hat{y}_c \log(\hat{y}_c) \tag{4}$$

The final loss used during the training can be expressed in the following way:

$$\begin{aligned} \mathscr{L}_{\omega}(y_i, \hat{y}_i) &= \omega_i \mathscr{L}_{XH}(y_i, \hat{y}_i) \\ &= \omega_i \sum_{c}^{C} y_{i,c} \log(\hat{y}_{i,c}) \end{aligned} \tag{5}$$

Where $\mathscr{L}_{XH}$ represents the cross-entropy loss function and C is the number of classes.

To reduce that weighted loss, the model now has two choices for the wrongly classified pixels. It can try to reduce the unweighted loss and hence improve the accuracy of the predictions or decrease $\bar{U}(\hat{y}_i)$ (increase the uncertainty).

D. Segmentation model

The segmentation is performed using a FC-DensetNet103 convolutional neural network, a densely connected CNN designed for semantic segmentation [11]. It is composed of 103 layers. The layers are defined inside dense blocks, transitions down and transitions up. A Dense block is formed with first a Batch Normalization layer, then a ReLU activation, followed by a 3x3 convolution and finally a dropout layer with probability $p = 0.2$. The transition down block is composed of a batch normalization layer, a ReLU activation, a 1x1 convolution, a dropout layer with probability $p = 0.2$ and finally a 2x2 non-overlapping max-pooling. Finally, the transition up block is composed of a 3x3 inverse convolution with stride 2.

The model is first trained on the entire training set using the categorical cross-entropy. At the end of the training, the top layers are removed and all the other layers are freezed. We then add three dense layers of the following respective sizes: 500, 100, 2. Each of the dense layers is also preceded by a Batch Normalization and a Dropout layers. The last dense layer also contains a softmax activation function. The training is later resumed using our weighted version of the loss function

III. RESULTS AND DISCUSSION

The initial training has been conducted on 1150 epochs using Keras with Adam for the optimizer and a categorical cross-entropy loss function. The second training was conducted on 150 epochs for our weighted version of the categorical cross-entropy.

A. Accuracy of the segmentation

Table I gives the overall and artery-specific accuracy, specificity and sensitivity scores. As can be seen, the best overall performance of the network is obtained on the aorta. This is due to the relatively simple shape of this type of artery. Indeed, the main trunk of the aorta is large compared to coronary arteries and most parts of the pulmonary arteries. Also, an aorta generally presents a low number of bifurcations and overlapping, which renders the segmentation task easier. According to the results yielded by our methodology, pulmonary arteries seem to be the most difficult type of vascular structure to segment. Indeed, it contains a very large number of bifurcations that overlap and occlude each other. Furthermore, the contrast agent is not necessarily evenly diluted in the entire arteries. Additionally, all the bifurcations and the problems of contrast rendered the manual segmentation difficult, which could explain part of the lower results obtained on pulmonary artery segmentation.

Fig. 1 presents visual results of the obtained segmentation for one sample of each type of artery. As can be seen, the segmentation yielded by our method (row (c)) appears really close to the ground-truth manual segmentation, even on patients for which no image has been used during the training process. Furthermore, our model correctly classified some thin bifurcations deemed too small during the manual segmentation (see first column of Fig. 1).

B. Visualization of the uncertainty from the segmentation

Rows (d) and (e) of Fig. 1 give visual uncertainty maps for the results obtained by the segmentation model. Row (d) represents the uncertainty for pixels correctly classified by the network while row (e) focuses on misclassified pixels. Those results are displayed as uncertainty maps: for each pixel an intensity is given according to its uncertainty. Hence, the brighter the color, the more important the uncertainty. This uncertainty shows multiple things.

First, we can see that the background generally contains low to no uncertainty. It means that ribs and other surrounding structures have properly been learned by the model and are not considered as arteries. However, the last column shows ribs segmented as part of the artery due to some lack of knowledge from the network, noticeably regarding the specific camera configuration (the image is either zoomed out or the patient is smaller than the other patients). Same conclusions can be drawn on the inside of the arteries where low uncertainty is usually yielded. The portions of maximum uncertainty in the images seem to be located on pixels at the border of the arteries. This is mostly due to the manual segmentation that can be challenging in the case of pulmonary and coronary arteries and also because of the lack of contrast in some regions at the borders of the arteries. The problem at the borders of the arteries mostly arises when the contrast agent is very diluted in the blood and when structures with really dark intensities are located around the arteries (see columns 3 and 4 of Fig. 1).

A second source of high uncertainty is generated by the overlapping of nearby structures. Indeed, because angiograms are 2D projections of a body, bones or other organs can overlap and sometimes occlude parts of an artery. Those particular regions are particularly difficult to classify. One example of that is displayed in the 4^{th} column of Fig. 1 where ribs with similar intensity than the aorta, and some cables overlap with the vessel, rendering the segmentation difficult to perform. However, even though those regions are flagged as false negatives, they yield high uncertainty that can give an idea to the model or the operator to use the segmentation with caution.

The last two columns contain examples for which the model provided some errors of segmentation. In accordance to previously drawn conclusions, a part of the high uncertainty regions belong on the boundaries of the arteries. Those columns also display that large uncertain patches also lie in regions containing false positive and false negative classifications. This configuration of uncertainty is what seem, to us, the most important to consider. It means that even though the model made mistakes in the segmentation, by either over-segmenting (ribs segmented as arteries in th 5^{th} column) or under-segmenting (missing part of the aorta in the 4^{th} column), it still has an idea of where those mistakes have been made.

Those images also show that wrongly classified pixels contain higher uncertainty values than well classified points.

C. Comparison weighted vs unweighted loss

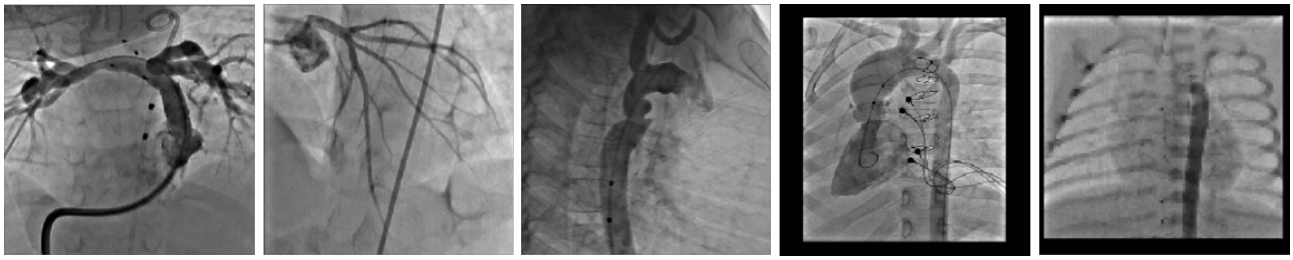
Our proposed weighting of the categorical cross-entropy loss function does not impact the accuracy of the predictions. Indeed, we obtained respective general accuracies of 94.9% for the weighted version and 94.7% for the regular cross-entropy after an average of 10 predictions for each.

Fig. 2 shows a comparison of the distribution of uncertainties for the weighted loss compaired to the original, unweighted, categorical cross-entropy for one image of our dataset. As can be seen, the unweighted version of the categorical cross-entropy tends to yield a lot of overconfident predictions, even for the missclassified points. This is, in part, due to an inflation of the most probable class created by the softmax loss function [12]. On the other hand, our weighting of the cross-entropy seems to counter that inflating effect. What can be observed is that, even though some good classifications are yielded with less confidence (more uncertainty), the number of confident errors is decreased. This is particularly useful in areas, such as medical imaging,

TABLE I

ACCURACY, SENSITIVITY AND SPECIFICITY SCORES FOR THE DIFFERENT TYPES OF ARTERIES CONTAINED IN THE TEST SET.

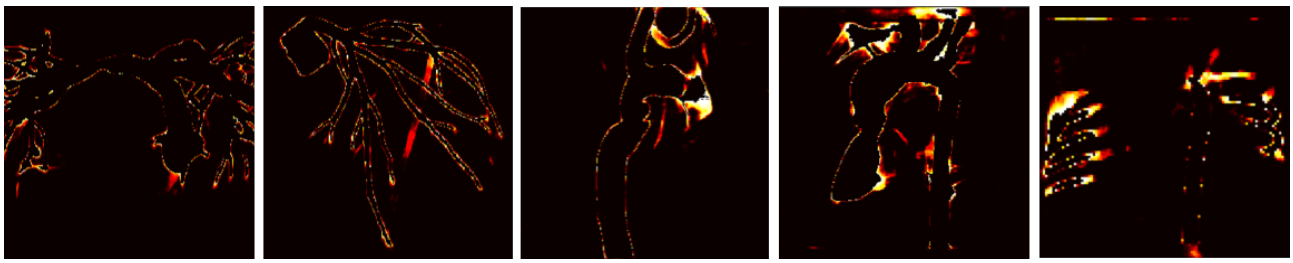
	Accuracy	Sensitivity	Specificity
Aorta	95.2%	97.2%	92.3%
Coronary	96.8%	97.1%	92.2%
Pulmonary	93.3%	95.0%	82.0%
Overall	94.9%	96.3%	84.6%



(a) Original angiograms

(b) Ground-truth manual segmentations

(c) Corresponding segmentations

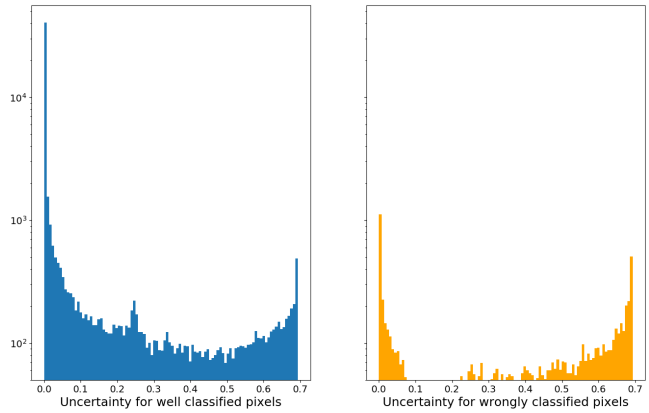


(d) Corresponding uncertainty map for well classified pixels

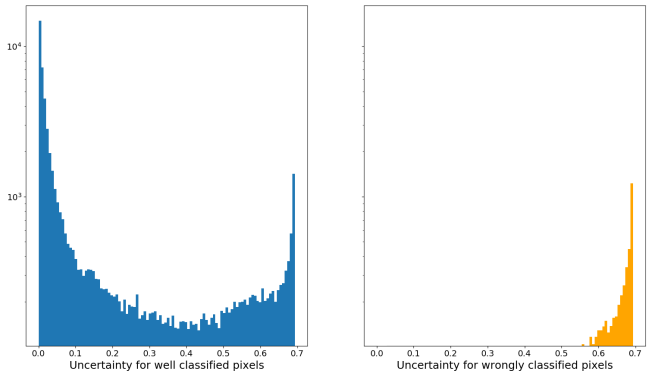
(e) Corresponding uncertainty map for misclassified pixels

Fig. 1. Segmentation for the three types of vascular structures considered in this paper (columns 1-3). Columns 4 and 5 present difficult cases with low contrast in the vascular structure. Rows (d) and (e) are presenting the uncertainty maps as follows: dark colors for low uncertainty and bright colors for high uncertainty.

where an error can have severe consequences. This observation is further confirmed in the table II that presents average uncertainties of well and incorrectly classified pixels for both versions of the loss.



(a) Distribution of uncertainty for the unweighted loss.



(b) Distribution of uncertainty the weighted loss.

Fig. 2. Histograms for the distributions of uncertainty for the unweighted and the weighted losses. Histogram displayed in log scale.

TABLE II

COMPARISON OF AVERAGE UNCERTAINTY VALUES FOR WELL AND WRONGLY CLASSIFIED PIXELS IN BOTH THE REGULAR CATEGORICAL CROSS-ENTROPY AND OUR PROPOSED WEIGHTED CATEGORICAL CROSS-ENTROPY.

Mean uncertainty	**Unweighted loss**	**Weighted loss**
Well classified pixels	2.5 * 10^{-2}	7.8 x 10^{-2}
Wrong classifications	2.5 * 10^{-1}	4.1 * 10^{-1}

IV. CONCLUSION

We proposed a new method to perform blood vessel segmentation on X-ray angiography using CNN. The approach presented in this paper does not focus on one single type of blood vessel, but instead, tries to model blood vessels using patterns and features generated by the contrast agent. The uncertainty modeling proposed along the segmentation identifies regions where the model is not confident enough about the predictions. Generally, high uncertainty regions seem to correlate with wrong classifications, in accordance with the proposed formulation of pixel weighting. Further investigations are required to draw a strong link between the outcome of the classification and the yielded uncertainty. This research paves the way for automatic handling of epistemic uncertainty in the context of X-ray imaging. Further works could consider incorporating the uncertainty in an automatic process designed to produce data augmentation using high uncertain patches and hence, refine the segmentation results.

ACKNOWLEDGMENT

This project was funded by NSERC Discovery grant. The Titan X used for this research was donated by the NVIDIA Corporation.

REFERENCES

[1] B. Irvine, W. Luo, and J. Leon, "Congenital anomalies in canada 2013: a perinatal health surveillance report by the public health agency of canadas canadian perinatal surveillance system," *Health promotion and chronic disease prevention in Canada: research, policy and practice*, vol. 35, no. 1, p. 21, 2015.

[2] S. Fumagalli, G. Torricelli, M. Massi, S. Calvani, S. Boni, A. T. Roberts, E. Accarigi, S. Manetti, and N. Marchionni, "Effects of a new device to guide venous puncture in elderly critically ill patients: results of a pilot randomized study," *Aging clinical and experimental research*, vol. 29, no. 2, pp. 335–339, 2017.

[3] Y.-S. Hsu, Y.-H. Lin, and B. Yang, "Impact of augmented reality lessons on students stem interest," *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 1, pp. 1–14, 2017.

[4] W. S. Khor, B. Baker, K. Amin, A. Chan, K. Patel, and J. Wong, "Augmented and virtual reality in surgerythe digital surgical environment: applications, limitations and legal pitfalls," *Annals of translational medicine*, vol. 4, no. 23, 2016.

[5] R. Martin, E. Vachon, J. Miró, and L. Duong, "3d reconstruction of vascular structures using graph-based voxel coloring," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 1032–1035.

[6] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998, pp. 130–137.

[7] F. M'hiri, L. Duong, C. Desrosiers, M. Leye, J. Miró, and M. Cheriet, "A graph-based approach for spatio-temporal segmentation of coronary arteries in x-ray angiographic sequences," *Computers in biology and medicine*, vol. 79, pp. 45–58, 2016.

[8] J. Fan, J. Yang, Y. Wang, S. Yang, D. Ai, Y. Huang, H. Song, A. Hao, and Y. Wang, "Multichannel fully convolutional network for coronary artery segmentation in x-ray angiograms," *IEEE Access*, 2018.

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[10] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in neural information processing systems*, 2017, pp. 5574–5584.

[11] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1175–1183.

[12] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Advances in Neural Information Processing Systems*, 2018, pp. 3183–3193.