EXPLOITING UNCERTAINTY OF DEEP NEURAL NETWORKS FOR IMPROVING SEGMENTATION ACCURACY IN MRI IMAGES

Alireza Norouzi[1], *Ali Emami*[1], *Kayvan Najarian*[2,3], *Nader Karimi*[1]
Shadrokh samavi[1,3], *S.M.Reza Soroushmehr*[2,3]

[1] Department of Electrical and Computer Engineering,
Isfahan University of Technology, Isfahan 84156-83111, Iran.
[2] Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, USA
[3] Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, USA

ABSTRACT

Deep neural networks have shown great achievements in solving complex problems. However, there are fundamental challenges which limit their real world applications. Lack of a measurable criterion for estimating uncertainty of the network predictions is one of these challenges. However, we can compute the variance of the network output by applying spatial transformations, distortions or noise injection to network inputs and interpret these variances as uncertainty of the network predictions. In other words, as long as the deformations do not conceptually alter target of interest, we expect the network to produce the same result. Hence, any outputs changes can be a sign of uncertainty in the network predictions. In order to estimate the prediction uncertainty of deep convolutional neural networks we use simple random transformations. By exploiting the network uncertainty, we improve the overall performance of the system. For a real use case, we apply the proposed method to segment left ventricle in MRI cardiac images. Experimental results demonstrate state–of–the–art performance and highlight the potential capabilities of simple ideas in conjunction with deep neural networks.

Index Terms— adaptive thresholding, conditional random fields, deep convolutional networks, segmentation

1. INTRODUCTION

The explosion of interest in using deep neural networks, fueled by their success in solving complex real world problems, has led to birth of impressive automated medical image analysis systems. Despite their capacity to learn rich hierarchical features, their output is usually a single number, resembling a kind of belief the network has about the input. However, there is usually no clue about how confident the model is about its prediction. This could be an alarm, especially in critical applications such as medical diagnosis where an estimate about reliability or confidence of the network output is a need. In this work, we try to model uncertainty of the output probability map using Monte Carlo sampling from a linear manifold on which the input image lies. This process is done by applying random affine transformation on input images. In contrast to other approaches which sample output by direct injection of noise to model internal representation [1], we perturb the input and check whether the system can respond properly or not. Since we know the perturbation process and its effect on the input, our approach offers more interpretable results about the model uncertainty compared to [1] which injects noise into the network neurons. Having computed the heat maps and uncertainty of the model using this technique, we apply an adaptive thresholding method based on a modified Conditional Random Fields (CRFs) to get the final segmentation mask.

For a use case of our proposed method, we use left ventricle segmentation in MRI cardiac images because of their importance. Cardiac diseases are one of the top causes of death especially in developed nations around the world. Despite massive investments on developing equipment, medicines and pre-caution strategies, there is still a huge gap between annual death reports caused by cardiac failures and an ideal world in which these diseases are fully under control [2].

To summarize, the contribution of this paper is three–fold: First, we compute uncertainty of the model output using Monte Carlo sampling in the input space using a well-defined affine transformation. Second, we utilize this information to improve the final segmentation accuracy by extending CRF formulation. Third, we apply our method to left ventricle MRI image segmentation and achieve state-of-the-art results.

The rest of this paper is organized as follows: in section 2, a literature review is presented, section 3 gives a detailed explanation of the proposed methods. In section 4, experimental results are discussed and finally we conclude in section 5.

2. LITERATURE REVIEW

Invention of fully convolutional neural networks(FCNs) [3] paved the way for popularizing the use of deep neural net-

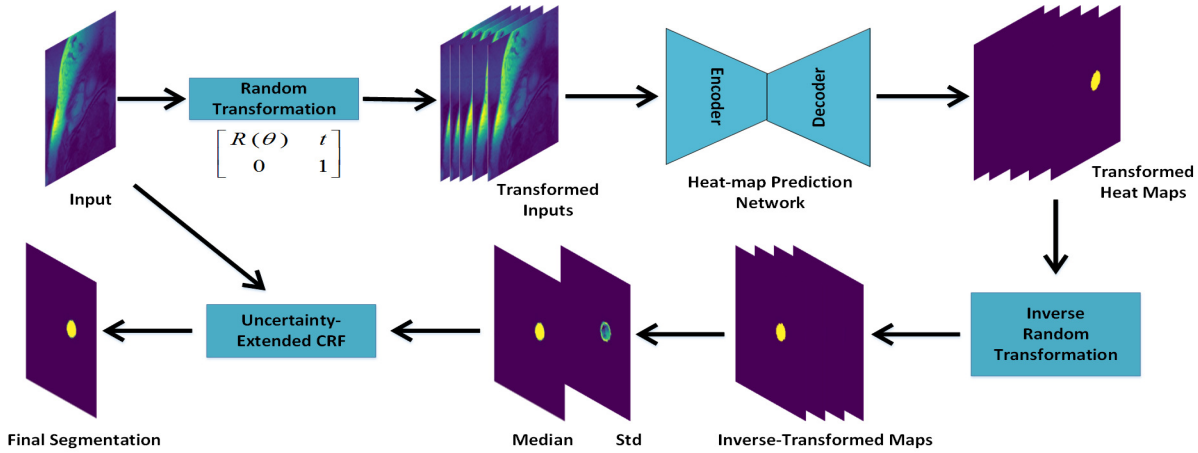

Fig. 1. Block diagram of the proposed method

works in semantic segmentation tasks. Since then, many enhancements and architectural changes have been proposed for improving accuracy [4] and operational speed [5].

In the field of computational medical image analysis, a specific architecture called U-Net [6] is successfully applied for solving a range of complex tasks, including segmentation. Recently, a number of research works have demonstrated the capability of deep learning methods in solving medical tasks comparable to that of human performance [7], [8]. Particularly, for the problem of left ventricle segmentation a wide range of methods have been proposed. Algorithms based on active contours and shape models are arguably one of the earliest and most popular tools used in this context [2]. More recently, a method based on dynamic programming has been proposed by [9]. Deep neural networks are of recent proposals in solving left ventricle segmentation problem [10], [11].

For incorporating model uncertainty in deep neural networks, one suggestion is to inject noise into network internal representations by various means such as test time dropout [1]. Another line of work is to make the network model a distribution family over the input rather than direct prediction of the desired output [12]. Compared to previous methods, our approach is certainly more controllable and interpretable, due to our knowledge of input space and perturbation process.

In this work, we employ an adaptive thresholding algorithm based on conditional random field models. Adaptive thresholding by using context or combination of local and global information to improve the segmentation is a mature topic in classical image processing. Otsu and Sauvola [13] is one of the most popular methods. Conditional random fields is another extensively investigated schemes among the deep learning community, especially after the introduction of fully connected CRFs introduced in [14]. In this work we extend the formulation presented in [14] to incorporate our proposed belief about the uncertainty of segmentation results.

3. PROPOSED METHOD

The proposed pipeline consists of three major components, as demonstrated in Figure 1. First of all, there is a random transformation generator which takes an image as input and generates several random transformations. The second component at the core of the pipeline, is a deep FCN while the last module in the pipeline uses these outputs to compute appropriate statistics for the final segmentation result, which utilize an uncertainty Extended CRF (UE-CRF). In the following subsections we elaborate on details of each module.

3.1. Heat-map prediction

The core of the pipeline is built upon the idea of fully convolutional neural networks and is shown in Figure 1. Among vast variety of architectures, U-Net is one of the most well-known ones for medical image analysis. Due to great success of deep learning in general and specifically U-Net architecture in solving complex medical tasks [6], we utilize a similar structure by proposing an improved version of U-Net [15]. In a nutshell, U-Net is an auto-encoder based architecture in which the encoder extracts the most salient features with regards to the input-output relationship. Given the encoded features of input image, decoder tries to predict the final answer, which in our case is a segmentation map. Furthermore, there are data flow connections between each encoder and its corresponding decoder. These shortcut links are crucial, especially in problems such as semantic segmentation to preserve spatial information which might be corrupted as a result of down-sampling procedures, such as max-pooling or strided convolution [1]. To encourage the network towards further generalization and prevent overfitting, a spatial dropout layer is used after the last decoder block. This is shown to be more effective than vanilla dropout in working with highly correlated spatial data like image and video data [16].

3.2. Variance Estimation

We feed various distortions or transformations of an input image to the network and calculate statistical variance of all segmentation results, as an indicator of the network uncertainty. There are various ways to transform images. For instance, it may be based on geometric operation such as affine transformation. There are also methods based on the intensity values, e.g. contrast jitter or intensity histogram mappings. In this work we limit our approach to geometric operations. For this purpose, a random affine transformation is applied on each input image prior to feeding into the deep CNN. The random affine matrix A, is built based on the following distributions.

$$A(t_x, t_y, \theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

$$t_x, y_y \sim Uniform(-\frac{T}{2}, +\frac{T}{2}) \tag{2}$$

$$\theta \sim Uniform(-\frac{\pi R}{360}, +\frac{\pi R}{360}) \tag{3}$$

where scalars T and R in (2) and (3), determine range of translation (in pixels) and rotation (in degrees) respectively. As is evident from (1), we only use translation and rotation transformations to build the affine matrix. Each generated matrix is stored in memory to compute the inverse transformation in later stages in which we apply it on the output of the CNN.

3.3. Conditional Random Field

Having computed the FCN uncertainty or variance of the network outputs, we apply an extension of CRF which takes the network uncertainty into account for improving the segmentation accuracy. We use fully conditional random fields as proposed in [14] which efficiently captures the interactions between every pairs of random variables. At inference, we assign a state to every random variable in order to minimize a total energy as defined by Equation (4).

$$E(\mathbf{x}) = \sum_i \mu(x_i) + \sum_i \sum_{j>i} \rho(x_i, x_j) \tag{4}$$

$$\mu(x_i) = f_\mu(m_i, \sigma_i) \tag{5}$$

$$\rho(x_i, x_j) = \delta(x_i, x_j) \sum_k w^{(k)} f^k_\varphi(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j) \tag{6}$$

In (4), (5) and (6), $\mu(x_i)$ and $\rho(x_i, x_j)$ are unary and pairwise terms as defined by general Equations (5) and (6). Functions $f_\mu(m_i, \sigma_i)$ and $f^k_\varphi(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j)$ are user defined to control the behavior of unary and pairwise terms. m_i and σ_i are median and standard deviation of each pixel as computed in section 3.2. $\boldsymbol{\varphi}_i$ is a feature vector extracted from location i, such as intensity, location or any handcrafted feature. The set of

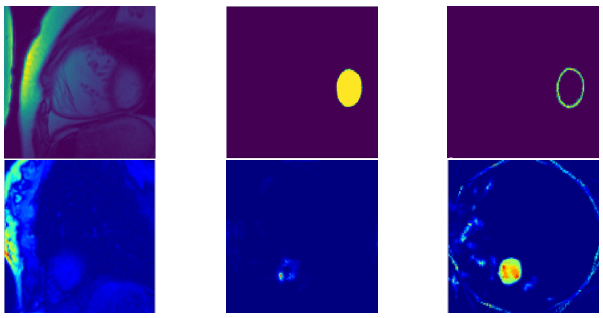

Image Median Standard deviation

Fig. 2. Exploratory analysis of dataset and FCN behavior. Top row shows the case of a successful segmentation. Bottom row shows the second case in which the region of interest is captured by uncertainty mask rather than the median.

kernel functions, f^k_φ, computes the similarity between their input arguments. Moreover, learnable weights shown as $w^{(k)}$ are for each kernel and $\delta(x_i, x_j)$ is called label compatibility function, which captures the dependency between random variables x_i and x_j.

The intuition behind incorporating uncertainty into the energy function of Equation (4) is inspired by our exploratory analysis of training data and their uncertainty masks. We observe two major cases in our experiments, as shown in Figure 2. The first one is when the heat map prediction network correctly identifies the region of interest which is the left ventricle in our experiments. In this scenario, the uncertainty map is indicative of the segmented region boarder. The second case is when the network is not able to consistently locate and segment the left ventricle in most of the transformed MRI images. Interestingly, in this situation the uncertainty map is highly correlated with the segmentation ground truth. In the light of above case analysis, the unary term is defined as:

$$\mu(x_i) = f_\mu(m_i, \sigma_i) = -\log(m_i + \lambda\sigma_i) \tag{7}$$

where λ is a hyper-parameter. We call this model UE-CRF(U). Pairwise terms are also defined as follows:

$$\rho(x_i, x_j) = w^{(1)} exp(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\sigma_s^2}) + w^{(2)} exp(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\sigma_s^2} - \frac{\|\mathbf{l}_i - \mathbf{l}_j\|^2}{2\sigma_l^2}) \tag{8}$$

in which $\mathbf{p}_i$ and $\mathbf{l}_i$ are pixel intensity and location respectively. σ_s and σ_l are hyper–parameters that control the sensitivity of the pairwise term to variations of intensity and location.

For the sake of comparison, we also compare our results with vanilla CRF. The unary term in this model is the negative log of the predicted probabilities of the original image.

4. EXPERIMENTAL RESULTS

We use the York dataset of cardiac MRI sequence [2] for training and evaluation. Dataset consists of 33 patients, for each of which, there are 20 MRI series with 8 to 15 slices. Each slice is a single channel image with spatial size of 256x256. In total, dataset contains 7980 MRI images, while only 5011 of them have segmentation ground-truth. In training and testing phases, we solely use the frames with segmentation labels. Some samples from the dataset are shown in Figure 3 with their corresponding ground truth.

4.1. Training and parameters tuning

Because there are no official training and validation splits available for this dataset, fair comparison with other methods is hard. In order to alleviate this problem, we use 11-fold cross-validation and report the average results. For each iteration, the test set consists of frames from 3 patients and other 30 patients are used in the training phase. To save computational power, we resize every frame to 128×128 pixels. The deep FCN (c.f. section 3.1) is trained for 3000 epochs using Adam optimization [17] with initial learning rate of 10^{-3}. In order to improve the performance in terms of Dice coefficient, we define a new loss function to directly incorporate Dice coefficient as follows.

$$Loss(\mathbf{y}_t, \mathbf{y}_p) = BCE(\mathbf{y}_t, \mathbf{y}_p) - exp(1 + Dice(\mathbf{y}_t, \mathbf{y}_p)) \quad (9)$$

where $\mathbf{y}_t$ is ground truth and $\mathbf{y}_p$ is the prediction of the model. BCE is the binary cross-entropy, widely used for binary classification problems. Dice coefficient in its original formulation is not differentiable, so we use the widely adapted soft version introduced in [6]. Soft Dice score is also shifted by one unit in the exponential function, so we are in a regime where gradient magnitude is bigger than one. This definition helps to alleviate vanishing gradient problem. However, it can introduce exploding gradients and oscillation at the end of training. Gradient clipping by norm with threshold 5 is used to prevent exploding gradient. For regularization, spatial dropout rate is set to 0.5 and data augmentation such as vertical/horizontal flips, rotations, translations and zooming are also applied during training. For hyper-parameters of random affine transformation and CRF in Equations (1) to (3), as well as (7) and (8), we set $T = 20$ *px* and R=180° and optimize λ using cross validation. λ values in range $[10^{-2}, 5 * 10^{-1}]$ provide good performance. We set $\lambda = 0.1$ for our experiments.

4.2. Results and comparisons

To evaluate and compare the proposed method, we use 3 different metrics. These metrics are Dice score, mean surface distance (MSD) also known as average perpendicular distance and Hausdroff distance (HD). Dice score computes the intersection over union ratio of ground truth segmentation and proposed segmentation areas. Maximum value of 1 is obtained when the proposed and ground truth segmentation perfectly overlaps. On the other hand, mean surface distance and Hausdroff distance focus on the boundary or contour of the segmentation. Mean surface distance measures the average distance from the manually drawn contour points to the proposed segmentation contour points, while the Hausdroff use the maximum distance. The overall results are shown in Table 1.

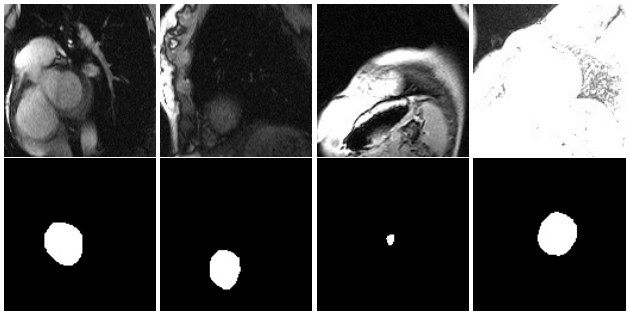

Fig. 3. Samples from the York MRI dataset [2] to show variations in the data. MRI frames are shown on the top row with their segmentation ground truth on the bottom row.

Table 1. Models evaluation and comparison with other methods.

Method	Dice(%)	MSD(mm)	HD(mm)
Andreopoulos [2]	N/A	1.4 ± 1.3	N/A
Fast-Segment [9]	0.859 ± 0.083	2.1 ± 0.7	N/A
Multiphase B-Spline [18]	0.9052 ± 0.0260	N/A	$\mathbf{3.4407 \pm 0.0187}$
FCN Only	0.8859 ± 0.0247	2.3916 ± 1.3702	5.1672 ± 0.0386
FCN + Vanilla CRF	0.8973 ± 0.0173	1.9377 ± 0.9443	4.3742 ± 0.0251
FCN + UE-CRF(U)	$\mathbf{0.9104 \pm 0.0218}$	$\mathbf{1.4184 \pm 0.9318}$	3.6437 ± 0.0210

5. CONCLUSION & FUTURE WORKS

In this paper we introduced a method to utilize a measure of uncertainty in predictions of deep convolutional neural networks to achieve state of the art results on left ventricle segmentation in MRI images. The uncertainty is computed by sampling from a linear manifold the input image lies on. Sampling precedes with generating random affine transformation matrix and applying it on the input image. Having computed multiple outputs and the approximate uncertainty (standard deviation) for each pixel, we used an adaptive thresholding algorithm based on extended CRF formulation to obtain the final segmentation map. Further investigation for modeling uncertainty of the deep CNN in their predictions is crucial to expand our understanding about internal mechanisms of these networks.

6. REFERENCES

[1] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla, "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding," nov 2015.

[2] Alexander Andreopoulos and John K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038, 2014.

[4] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han, "Learning Deconvolution Network for Semantic Segmentation," in *IEEE International Conference on Computer Vision (ICCV)*. dec 2015, pp. 1520–1528, IEEE.

[5] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello, "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation," *arXiv preprint arXiv:1606.02147*, 2016.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing & Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[7] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115, 2017.

[8] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, and Others, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint arXiv:1711.05225*, 2017.

[9] C. Santiago, J.C. Nascimento, and J.S. Marques, "Fast segmentation of the left ventricle in cardiac MRI using dynamic programming," *Computer Methods and Programs in Biomedicine*, vol. 154, pp. 9–23, 2018.

[10] Jacinto C. Nascimento and Gustavo Carneiro, "Deep Learning on Sparse Manifolds for Faster Object Segmentation," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4978–4990, 2017.

[11] S Molaei, M E Shiri, K Horan, D Kahrobaei, B Nallamothu, and K Najarian, "Deep Convolutional Neural Networks for Left Ventricle Segmentation," in *Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 668–671.

[12] Yann N Dauphin and David Grangier, "Predicting distributions with linearizing belief networks," *arXiv preprint arXiv:1511.05622*, 2015.

[13] Mehmet Sezgin and Bülent Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging*, vol. 13, no. 1, pp. 146–166, 2004.

[14] Philipp Krähenbühl and Vladlen Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.

[15] Abhishek Chaurasia and Eugenio Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," *CoRR*, vol. abs/1707.03718, 2017.

[16] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648–656.

[17] Diederik P. Kingma and Jimmy Lei Ba, "Adam: a Method for Stochastic Optimization," *International Conference on Learning Representations 2015*, pp. 1–15, 2015.

[18] Van-Truong Pham, Thi-Thao Tran, Kuo-Kai Shyu, Lian-Yu Lin, Yung-Hung Wang, and Men-Tzung Lo, "Multiphase b-spline level set and incremental shape priors with applications to segmentation and tracking of left ventricle in cardiac mr images," *Mach. Vision Appl.*, vol. 25, no. 8, pp. 1967–1987, Nov. 2014.