ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

SDE U-Net: Disentangling Aleatoric and Epistemic Uncertainties in Medical Image Segmentation

Chuxin Zhang[1], Ana Maria Barragan-Montero[1] and John A. Lee[1,2] *

1- UCLouvain - IREC/MIRO
Avenue Hippocrate 55 B1.54.07, 1200 Brussels - Belgium

2- UCLouvain - ICTEAM
Place du Levant 3, 1348 Louvain-la-Neuve - Belgium

Abstract. Quantifying uncertainty is crucial in artifical intelligence (AI) applications, particularly in high-stakes healthcare settings. This paper introduces SDE U-Net, a novel architecture that integrates stochastic differential equations (SDEs) with the U-Net framework, effectively distinguishing between *aleatoric* and *epistemic* uncertainties. By incorporating a randomness component, SDE U-Net directly captures and quantifies *aleatoric* uncertainty, while *epistemic* uncertainty is assessed through multiple forward passes. Comparative results show that SDE U-Net not only matches but also exceeds benchmark performance, achieving similar results in just 500 epochs, half the epochs required by the benchmark. This approach enhances the reliability of AI in medical decision-making by providing a clear, comprehensive representation of uncertainty, marking a significant advancement in the field of medical image segmentation.

1 Introduction

Innovative AI applications are reshaping medical practices, offering unprecedented accuracy and efficiency in the whole pipeline. Unlike applications in other areas, tasks in healthcare are critically important; even minor errors can lead to significant risks and even human live losses. Therefore, developing deep learning models with uncertainty estimation is essential. Uncertainty is a metric that measures how trustworthy the predictions are. There are two kinds of uncertainty: *aleatoric* uncertainty, which arises from noise in the data, and *epistemic* uncertainty, which refers to the model's lack of knowledge.

Monte Carlo dropout (MCDO) and deep ensemble (DE) are two prevalent methods in deep learning. For instance, Vanginderdeuren et al. [13] explored the estimation of uncertainty in radiation oncology dose prediction, employing dropout and bootstrap techniques within U-Net models to enhance the safety and reliability of treatment planning . In contrast, Senousy et al. [5] engineered an uncertainty-aware network using DE, noted for its high accuracy and efficiency. However, these common approaches for uncertainty estimation have significant drawbacks: they require extensive computational resources and cannot effectively differentiate between the two types of uncertainties.

*C.Z. is funded by the F.R.S.-FNRS Télévie grant #.7.4558.22. J.A.L. is a Research Director with the Belgian F.R.S.-FNRS.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Importantly, distinguishing between these two types of uncertainty is vital for enhancing the reliability of medical decisions. To address this challenge, this work introduces a novel method capable of distinguishing *epistemic* and *aleatoric* uncertainties. Specifically, our approach combines SDE-Net [7] and U-Net [6]. SDE-Net uses stochastic differential equations (SDE) to model neural networks, incorporating Brownian motion into the forward propagation.

The rest of this paper is organized as follows: Section 2 presents the theoretical principles and model design. Section 3 describes the datasets and experimental setup. Section 4 discusses the results obtained from these experiments. Finally, Section 5 draws conclusions from our study and sketches perspectives for future research.

2 Method

2.1 Neural Neworks as Dynamical Systems

Our method is inspired by SDE-Net [7], which conceptualizes neural networks as dynamical systems. A neural network typically comprises a series of hidden layers, analogous to the states in a dynamic system. The transformation between two consecutive layers can be described with the differential equation:

$$dx_t \quad = f(x_t, t)dt \ , \tag{1}$$

where t represents the levels in the network, f is a function of x and t, and x_t is the hidden state at level t. This equation follows the Euler method for solving ordinary differential equations (ODEs) [8]. Equation (1) relates to the concept of a Neural Ordinary Differential Equation (Neural ODE) [8], a deterministic model that does not inherently estimate uncertainty. To account for uncertainty, we extend this model by introducing a stochastic term, leading to a Neural Stochastic Differential Equation (Neural SDE), given by:

$$dx_t = f(x_t, t)dt + g(x_t, t)dW_t \ , \tag{2}$$

where g is a function dependent on x and t, and dW_t represents the Brownian motion. The solution to this SDE can be approximated using the Euler-Maruyama method [11], as follows:

$$X_{t+\Delta t} = X_t + f(X_t, t)\Delta t + g(X_t, t)\Delta W_t \ . \tag{3}$$

The stochastic term introduces variance from the noise into the model, enabling the direct capture and quantification of *aleatoric* uncertainty. Conversely, *epistemic* uncertainty is addressed through the variability in model predictions under different realizations of the stochastic process dW_t. By conducting multiple forward passes through the model, each with a distinct realization of dW_t, we generate a distribution of outputs. This distribution reflects the *epistemic* uncertainty.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

2.2 The architecture of SDE U-Net

In our proposed network, depicted in Fig. 1, we integrate the SDE into the traditional U-Net architecture [6]. Fig. 1 one pair of an encoder and a decoder from the SDE U-Net architecture, where these components are connected via skip connections. Our model retains the classical U-Net structure, which consists of four encoder layers, a bottleneck layer, and four decoder layers.

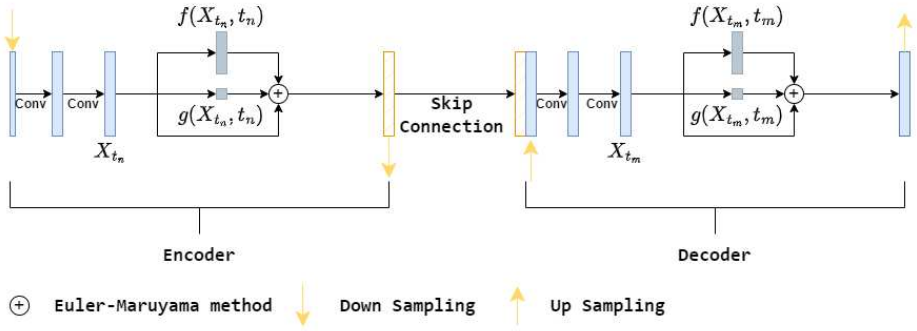

Fig. 1: One pair of encoder and decoder of the SDE U-Net architecture.

Euler-Maruyama Method: Compared to the standard U-Net, we utilize drift f and diffusion g modules to implement the Euler-Maruyama method within each level of the network, as shown in Fig. 2. The drift component f is applied to generate deterministic changes, while the diffusion component g accounts for stochastic variations. These are combined at each level to form the final output as defined by Eq. (3).

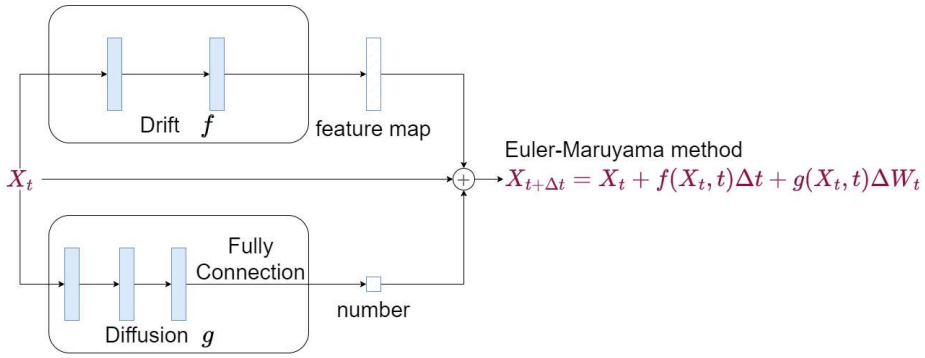

Fig. 2: Drift and diffusion modules used to implement the Euler-Maruyama method in the SDE framework.

Time Convolution: Additionally, we employ time convolution in place of standard convolution. This involves adding a time channel to the feature maps

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

before applying the convolution operation, without altering the number of output feature map channels, as illustrated in Fig. 3. Here, the variable t is represented as a continuous value proportionate to the level number, calculated as $t = \text{constant} \times \frac{\text{current level}}{\text{total levels}}$, reflecting the relative depth within the network since our medical imaging task does not involve temporal data.

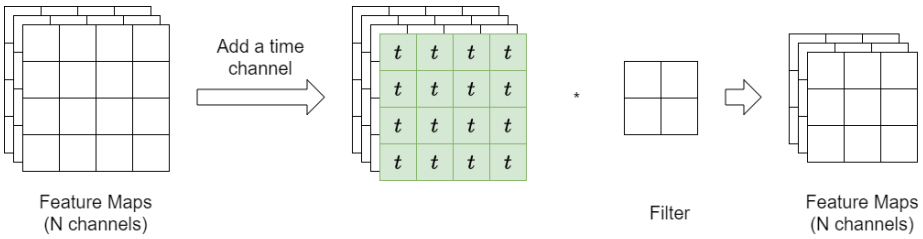

Fig. 3: Illustration of how time convolution is implemented by adding a time channel to the feature maps before the convolution operation.

3 Experiments and Results

3.1 Dataset and Settings

For our experiments, we employed two distinct datasets. The first dataset, from a previous study on left breast cancer radiotherapy [9], includes CT scans, segmentation of organs at risk (OARs), and dosimetric data from 60 patients, divided into training, validation, and testing subsets of 42, 9, and 9 patients, respectively. The second dataset, sourced from the NSCLC-Radiomics database [14], consists of CT scans for GTV segmentation from 422 non-small cell lung cancer patients. These 3D images were processed as 2D slices, resulting in 5930 images (from 379 patients) for training and 652 images (from 43 patients) for validation, providing a broader evaluation in a different clinical context.

Both models were trained on Nvidia A100 GPUs. The CT images underwent Z-score normalization for consistency, and only slices containing object masks were used for training. Post-processing involved sigmoid activation followed by thresholding at 0.85 to generate the final segmentation masks.

3.2 Results

In the first dataset, our SDE U-Net achieved a Dice score of 0.9598, while nnUNet scored 0.9788. On the second dataset, our SDE U-Net achieved a Dice score just above 0.4, compared to nnUNet's slightly higher score above 0.5. Despite the slightly higher score of nnUNet, our model showed comparable performance with significantly fewer parameters- 62.5 million versus nnUNetâs 126.6 million- highlighting its robustness and potential, especially in resource-constrained settings.

We also explored the relationship between Dice scores and uncertainty, focusing on both *aleatoric* and *epistemic* uncertainties. Epistemic uncertainty was

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

quantified by multiple forward passes during inference, calculating the variance across predictions, while *aleatoric* uncertainty came from the SDE U-Net's final diffusion layer. As shown in Table 1, we found a moderate negative correlation between Dice scores and *epistemic* uncertainty, consistent with the idea that higher uncertainty reflects lower confidence and accuracy. *Aleatoric* uncertainty, however, showed varying correlations with Dice scores across patients (Pearson: 0.38 to 0.72), indicating the model's capture of intrinsic variability.

Table 1: Correlation coefficients between Dice scores and uncertainty for different patients.

Patient	**Epistemic Uncertainty**						**Aleatoric Uncertainty**	
	Pearson			**Spearman**			**Pearson**	**Spearman**
	10	**20**	**50**	**10**	**20**	**50**		
Patient 1	-0.8691	-0.8695	-0.8705	-0.6429	-0.6429	-0.6429	0.56	0.80
Patient 2	-0.8144	-0.8147	-0.8144	-0.7716	-0.7822	-0.7733	0.70	0.51
Patient 3	-0.8336	-0.8329	-0.8327	-0.7771	-0.7873	-0.7802	0.55	0.61
Patient 4	-0.6522	-0.6608	-0.6570	-0.6944	-0.7048	-0.6975	0.49	0.55
Patient 5	-0.5074	-0.5058	-0.5036	-0.7277	-0.7246	-0.7300	0.65	0.55
Patient 6	-0.7881	-0.7837	-0.7888	-0.4708	-0.4598	-0.4653	0.72	0.32
Patient 7	-0.2188	-0.2166	-0.2138	-0.8950	-0.8831	-0.8861	0.38	0.59
Patient 8	-0.9022	-0.9054	-0.9020	-0.8338	-0.8269	-0.8292	0.64	0.28
Patient 9	-0.3695	-0.3667	-0.3729	-0.8468	-0.8425	-0.8504	-0.43	-0.08

4 Conclusion

This study introduces a novel integration of SDEs into the U-Net architecture to incorporate *aleatoric* uncertainty directly into the model. Unlike typical approaches that use a density map for uncertainty quantification, our model uniquely outputs *aleatoric* uncertainty as a numerical value, enhancing the interpretability of uncertainty in medical image segmentation.

Noticeably, the diffusion output, interpreted as *aleatoric* uncertainty, shows a decreasing trend during training and validation, inversely related to Dice scores, yet positively correlates with Dice scores during testing. This observation raises intriguing questions about the dynamics between learned uncertainty and model generalization, indicating a complex interplay that requires further exploration.

The study is limited by its lack of comparison with other uncertainty estimation methods, which could provide deeper insights into the efficacy of our approach. Looking ahead, there is potential to extend this methodology to other medical domains such as dose prediction and 3D segmentation, suggesting broad implications for predictive models in medical practice.

In conclusion, the integration of *aleatoric* uncertainty within our framework presents a promising direction for improving the robustness and reliability of medical imaging analysis. Future work will focus on further elucidating the dynamics of uncertainty and expanding the model applications.

References

[1] Z. Xiao, B. Liu, L. Geng, F. Zhang, and Y. Liu. Segmentation of lung nodules using improved 3D-UNet neural network, *Symmetry*, 12(11), 1787, 2020.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

[2] S. Iqbal, G. F. Siddiqui, A. Rehman, L. Hussain, T. Saba, U. Tariq, and A. A. Abbasi. Prostate cancer detection using deep learning and traditional techniques, *IEEE Access*, 9, 27085-27100, 2021.

[3] M. T. Lu, A. Ivanov, T. Mayrhofer, A. Hosny, H. J. Aerts, and U. Hoffmann. Deep learning to assess long-term mortality from chest radiographs, *JAMA network open*, 2(7), e197416-e197416, 2019.

[4] Z. Klanecek, T. Wagner, Y. K. Wang, L. Cockmartin, N. Marshall, B. Schott, and R. Jeraj. Uncertainty estimation for deep learning-based pectoral muscle segmentation via Monte Carlo dropout, *Physics in Medicine & Biology*, 68(11), 115007, 2023.

[5] Z. Senousy, M. M. Abdelsamea, M. M. Gaber, M. Abdar, U. R. Acharya, A. Khosravi, and S. Nahavandi. MCUa: Multi-level context and uncertainty aware dynamic deep ensemble for breast cancer histology image classification, *IEEE Transactions on Biomedical Engineering*, 69(2), 818-829, 2021.

[6] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer assisted intervention MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III*, 18, 234-241, Springer International Publishing, 2015.

[7] L. Kong, J. Sun, and C. Zhang. SDE-Net: Equipping deep neural networks with uncertainty estimates. *arXiv preprint arXiv:2008.10546*, 2020.

[8] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud,. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[9] L. Vander Veken, G. Van Ooteghem, A. Razavi, S. D. R. Quaresma, E. Longton, C. Kirkove, and X. Geets. Voluntary versus mechanically-induced deep inspiration breath-hold for left breast cancer: A randomized controlled trial. *Radiotherapy and Oncology*, 183, 109598, 2023.

[10] F. Isensee, P. F. Jaeger, S. A. A. Kohl, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*, 18, 203-211, 2021. doi:10.1038/s41592-020-01008-z.

[11] G. Maruyama. Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4, 48-90, 1955.

[12] M. Huet-Dastarac, S. Michiels, S. T. Rivas, H. Ozan, E. Sterpin, J. A. Lee, and A. Barragan-Montero. Patient selection for proton therapy using Normal Tissue Complication Probability with deep learning dose prediction for oropharyngeal cancer, *Medical Physics*, 50(10), 6201-6214, 2023.

[13] A. Vanginderdeuren, M. Huet-Dastarac, and A. M. Barragan Montero, J.A. Lee. Estimating uncertainty in radiation oncology dose prediction with dropout and bootstrap in U-Net models. In: ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. doi:10.14428/esann/2021.es2021-117.

[14] H. J. W. L. Aerts, L. Wee, E. Rios Velazquez, et al. Data From NSCLC-Radiomics (version 4) [Data set]. *The Cancer Imaging Archive*, 2014. doi:10.7937/K9/TCIA.2015.PF0M9REI.