
DropFilter for Uncertainty Estimation in Brain Tumor Segmentation

Natalia Khanzhina¹ Maxim Kashirin¹ Andrey Filchenkov¹ Ilya Osmakov² Pavel Ulyanov²

¹Machine Learning Lab, ITMO University, 49 Kronverksky Pr., St. Petersburg 197101, Russia
²formerly Burtsev Lab, Skolkovo Innovation Center, 42 Bolshoy Blvd., Moscow 121205, Russia

Abstract

Monte Carlo (MC) Dropout technique for uncertainty estimation of deep neural networks has certainly been found useful for medical image segmentation and classification tasks. However, Dropout as a general technique is known to be harmful to convolutional neural networks and therefore has been often bypassed by other regularization techniques when it comes to its application to images. In this paper, we propose to apply an improved uncertainty estimation approach, Monte Carlo DropFilter (MCDF), for the problem of brain tumor images segmentation. Our models are able to efficiently and harmlessly incorporate the idea of Dropout for convolutional neural network architectures, which is known as Spatial Dropout, or DropFilter, into Bayesian inference based on the MC sampling. We describe the advantages of MC DropFilter technique when applied to solve the problem of brain tumor image segmentation on BraTS dataset, and, according to our evaluation results, the technique achieves superior performance over state-of-the-art baselines and MC Dropout.

1 INTRODUCTION

The majority of research concludes that the number of individuals who suffer and die from brain tumors has increased during the past few decades [Kaneko et al., 2002, Porter et al., 2010]. Despite numerous works devoted to the clinical diagnosis of brain tumors, it is difficult to differentiate this pathology of the central nervous system (CNS) from other diseases. Early detection of CNS neoplasms is still challenging, which is particularly due to the variety of options for the debut of neuro-oncological pathology and the polymorphism of its clinical and neurological manifestations. Currently, magnetic resonance imaging (MRI) is the

most common technique for brain tumors investigation.

Deep neural networks (DNNs) are successfully applied for automatic detection of tumors on MRI scans [Wang et al., 2017, 2018]. Automated systems based on DNNs can assist doctors in the disease identification at early stages, which provides more ways to reduce the risk of irreversible damage [Kan et al., 2018, Bakas et al., 2018].

At the same time, understanding what a predictive model is not confident about becomes critical in machine learning applications to medical problems. While the modern DNNs are capable to learn high dimensional representations, they often fail to generalize to real-world data. However, their outputs are blindly assumed to be accurate, which may result in error diagnoses in medical tasks [Leibig et al., 2017] and wrong decisions in other domains [16-007, 2017, Guynn, 2015].

This issue motivates researchers to develop methods that allow to distinguish and estimate the impact of different factors resulting in prediction uncertainty. Such methods and corresponding frameworks may vary [Lakshminarayanan et al., 2017, Segú et al., 2019], but they are mostly grounded in the Bayesian formalism [Bernardo and Smith, 2009].

There are two types of uncertainty measured in deep learning for computer vision (CV). The first one is *aleatoric* uncertainty is related to noise inherent in data [Wang et al., 2019a,b]. The second type, which is called *epistemic* uncertainty, represent the uncertainty in the model parameters, capturing a lack of data about the initial distribution. It can be decreased with a sufficient amount of data provided and is usually referred to as model uncertainty. Approximate variational inference (VI) is widely used to estimate epistemic uncertainty. However, it is expensive to compute for very deep NNs. Hence, approximate methods and tools have been developed by Gal and Ghahramani [2016], Li et al. [2017], Zhu and Zabaras [2018] that have been also used in brain tumor imaging [Wang et al., 2019a, Jungo et al., 2017, 2018]. One very popular method is Monte Carlo Dropout (MC Dropout, MCDO) [Gal and Ghahramani, 2016]. It has

proven to be efficient due to memory savings compared to other VI methods for very deep NNs [Gal and Ghahramani, 2016]. Rousseau et al. [2021], Mehta and Arbel [2018] also demonstrated that the technique can be successfully applied to brain tumor detection.

The problem with the usage of the classical Dropout in convolutional layers is that switching off random neurons of filters results in corrupting learned feature representations. The idea comes from structural differences between dense and convolutional layers, which are believed to be the main reason why Dropout does not work well enough for convolutional layers.

Instead of Dropout, Tompson et al. [2015] proposed to consider whole convolutional filters as drop units. The motivation for these techniques is that in convolutional layers co-adaptations tend to occur between filters rather than inside a filter. This technique referred as Spatial Dropout [Tompson et al., 2015], DropChannel [Zhang et al., 2019] or DropFilter [Tian, 2018] provides better results than dropout. In this research we will refer to that technique as DropFilter.

Based on the success of DropFilter, the epistemic uncertainty estimation technique, which we refer as *Monte Carlo DropFilter* (MC DropFilter, MCDF) has been also recently suggested by Zhang et al. [2019]. MC DropFilter has been proven its efficiency for uncertainty quantification and model calibration compared to MC Dropout, but no works on its application for the epistemic uncertainty estimation in the brain MRI problems have been found.

To close the gap, we propose to use the MC DropFilter for the epistemic uncertainty estimation in the brain tumor MRI segmentation problem based on convolutional neural networks, which is the main paper contribution.

The paper investigates the following research questions:

RQ1. Is MC DropFilter superior compared to MC Dropout for the problem of brain tumor segmentation?

RQ2. Does epistemic uncertainty quantification based on MC DropFilter help improving the accuracy of tumors segmentation?

In the article we answer these questions applying the MC DropFilter for two models, which use two different baseline backbones, namely Xception [Chollet, 2017] and DenseNet [Huang et al., 2017]. These baselines recently showed high performance compared to popular architectures for our problem [Zeineldin et al., 2020]. We analyze models calibration based on proposed MC DropFilter and classical MC Dropout as well as the impact of these techniques on the segmentation problem solving.

2 BACKGROUND

2.1 EPISTEMIC UNCERTAINTY BASED MONTE CARLO DROPOUT

Following Gal and Ghahramani [2016], Wang et al. [2019a], we first give the theoretical formulation of epistemic uncertainty for the image segmentation problem. Let X be an image, $Y = \{y_1, \dots, y_M\}$ be a segmentation mask for X , where y_i is the i -th pixel of a mask. Then $f(\cdot)$ is the solving function for segmentation problem represented by a deep neural network with parameters ω . Let $q_\theta(\omega)$ be an approximating distribution over the set of function parameters, which could be found based on minimizing the Kullback–Leibler divergence to parameters posterior. Monte Carlo sampling from $q_\theta(\omega)$ for the model $f(\omega, X)$ aims to approximate the predictive distribution $p(Y|X)$.

Gal and Ghahramani [2016] introduces the Dropout [Hinton et al., 2012] as the way of ω neural network model parameters representation. The approximating distribution $q_\theta(\omega)$ for NN $f(\omega, X)$ with Dropout applied before every layer l equals to $\prod_l q_{M_l}(W_l)$ with M_l the mean weights matrix. The approximating distribution for the layer l is $q_{M_l}(W) = M_c^K \cdot z^K$ with $z^K \sim \text{Bern}(1 - p)$ the dropping out probability vector, where p is the Dropout rate. This technique is called MC Dropout. Using it, sampling $\hat{y}_n = q_{\theta_n}(X)$ is performed with different z_N dropout masks N times giving $\mathcal{Y} = \{\hat{y}_1, \dots, \hat{y}_N\}$. The predicted segmentation mask for X image can be defined as

$$\hat{Y}^i = \arg \max_{\hat{y}^i} P(\hat{y}^i|X) \approx \text{Mode}(\mathcal{Y}^i), \quad (1)$$

where i is a pixel of segmentation mask.

Epistemic uncertainty can be measured using Mutual Information \mathcal{MI} , which is more representative for the segmentation problem, as it is the per-pixel classification [Hüllermeier and Waegeman, 2021]. This measure is used in our study and can be calculated for every pixel of the predicted segmentation mask for the image X as below:

$$\mathcal{MI}(\hat{Y}^i, \theta|D, X) = \mathcal{H}_{p(\hat{Y}^i|X, D)}(Y^i) - E_{p(\theta|D)}(\mathcal{H}_{\hat{Y}^i|X, \theta}(\hat{Y}^i)), \quad (2)$$

$$\mathcal{H}(\hat{Y}^i) \approx - \sum_{m=1}^M \hat{p}_m^i \ln \hat{p}_m^i, \quad (3)$$

where $D = \{X_1, \dots, X_K\}$, θ is the variational parameter set with the dropout mask, q_θ is the approximating distribution, $\theta \sim q_\theta$, \hat{p}_m^i is the frequency of class m for i -th pixel of segmentation mask.

Therefore, adding Dropout to a model and using it for the sampling different Dropout masks $\theta \sim q_\theta$ during inference time is a way to obtain a Bayesian NN from any DNN.

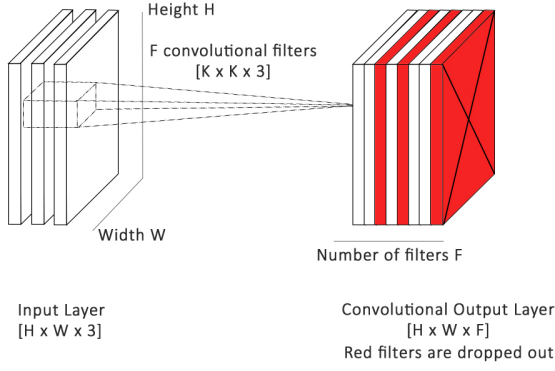


Figure 1: DropFilter technique illustration on random filters marked with red are dropped out.

2.2 DROPFILTER

DropFilter [Tompson et al., 2015], also referred to as Spatial Dropout, or DropChannel has similar idea with the Dropout scaled to convolutional filters level. DropFilter for a NN layer l can be described as

$$z^l = a((r^F \cdot F^l) * z^{l-1} + r^b \cdot b^l), \quad (4)$$

where a denotes an activation function, F denotes the convolution filter set, $*$ denotes the convolution operation, \cdot denotes the element-wise product, z^l is the l layer output, $l-1$ is the previous to l layer, b is the bias, $r^F \sim \text{Bern}(1-p)^{K_l}$, $r^b \sim \text{Bern}(1-p)^{K_l}$ are the DropFilter vectors with K_l the number of F filters and the DropFilter rate equal to p .

DropFilter can also be considered as paths suppression in depthwise separable convolution module. It can be exemplified through treating each path in this module as a convolutional filter. Thereby, DropFilter can be generalized to models that utilize depthwise separable convolutions, for example, MobileNet [Howard et al., 2017] and Xception [Chollet, 2017]. Fig. 1 represents the DropFilter illustration and Fig. 2 demonstrates its application to the depthwise separable convolution module.

3 MONTE CARLO DROPFILTER

Recently, [Zhang et al., 2019] proposed to use DropFilter for Monte Carlo sampling and showed its advantages compared to MC Dropout in terms of model calibration. In this section we propose to apply MC DropFilter for epistemic uncertainty estimation of models predictions for the brain tumor segmentation problem.

To describe MC DropFilter, let a convolutional neural network with two convolutional layers $f(\omega, X)$ be defined as:

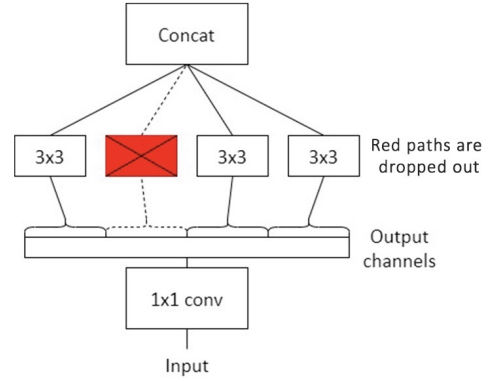


Figure 2: DropFilter technique illustration on random paths of depthwise separable convolution module marked with red are dropped out.

$$f(\omega, X) = a((r_2^{F^2} \cdot F^2) * a((r_1^{F^1} \cdot F^1) * X + r_1^{b^1} \cdot b^1) + r_2^{b^2} \cdot b^2), \quad (5)$$

$$\theta = \{F^1, F^2, r_1^{F^1}, r_1^{b^1}, r_2^{F^2}, r_2^{b^2}\},$$

where F^1, F^2 denote the convolution filters (i.e. weight matrices), $*$ denotes the convolution operation, \cdot denotes the element-wise product, X is the input image, b is the bias, $r_1^{F^1}, r_1^{b^1} \sim \text{Bern}(1-p)^{K_1}$, $r_2^{F^2}, r_2^{b^2} \sim \text{Bern}(1-p)^{K_2}$ are the DropFilter vectors with the DropFilter rate equal to p .

Given the approximating distribution $q_\theta(\omega)$ we can sample from it N times different DropFilter masks during inference or testing time in the Monte Carlo manner forming $\mathcal{Y} = \{\hat{y}_1, \dots, \hat{y}_N\}$. \mathcal{Y} can be used then for the predictive mean evaluation and epistemic uncertainty estimation the same way with Eq. 1 and 2.

The network defined in Eq. 5 can be generalized to any number of layers. The CNN with MC DropFilter applied for convolutional layers at the testing time can be treated as Bayesian NN similarly with MC Dropout.

As we show in Section 2.2, DropFilter is easily scaled to the depthwise separable convolution module. Thus, in this research we propose to use Xception [Chollet, 2017], which is based on the depthwise separable convolution module, as the first baseline architecture. Here, we demonstrate the inference with the modified Xception module using DropFilter technique (Fig. 3).

The second baseline model we chose is DenseNet [Huang et al., 2017]. Both DenseNet and Xception can be considered as mobile NNs and were successfully applied by Zeineldin et al. [2020] for the problem of brain tumor segmentation. As MC DropFilter application for DenseNet is less trivial than for Xception, and can be illustrated by Fig. 1, here we omit it.

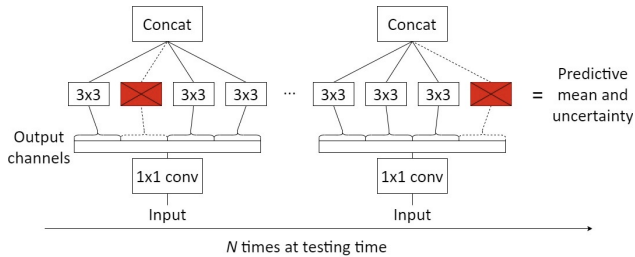


Figure 3: Monte Carlo DropFilter technique illustration for the depthwise separable convolution. Different DropFilter masks are sampled at the testing time to form the mean prediction and estimate epistemic uncertainty.

4 EXPERIMENTS

4.1 DATASET

We evaluated the models on BraTS 2018 dataset [Menze et al., 2014, Bakas et al., 2018], which is the benchmark for the brain tumor image problems (i.e. segmentation and classification). It contains 44,175 multi-modal scans of patients with brain tumors. A sample consists of four channels scan and a corresponding tumor label of size 256×256 of the axial view. Each scan consists of 4 modes, namely T1-weighted (T1), post-contrast T1-weighted (T1c), T2-weighted (T2), T2 Fluid Attenuated Inversion Recovery (FLAIR), and acquired with different clinical protocols and various scanners from multiple institutions. Labels include whole tumor (WT), active tumor (AT) and tumor core (TC) tissue. Here we only consider the WT segmentation problem as the most challenging.

4.2 MODELS

The following model architecture was used for the brain tumor image segmentation:

- Backbone subnetwork. As the backbone two neural networks were studied: Xception [Chollet, 2017], DenseNet [Huang et al., 2017]. We chose them due to their relative superiority to the other popular lightweight CNNs used by researchers for our problem [Zeineldin et al., 2020].
- Feature Pyramid Network [Lin et al., 2017] which is the U-Net-like subnetwork.
- Two “head” subnetworks: for the tumor and its border segmentation. The separating border segmentation leads to better performance, according to [Seferbekov, 2018]. Hence, we used double head loss for training, which is the combination of Dice loss and binary cross-entropy, also proposed by [Seferbekov, 2018].

Two backbone NNs lead to two baseline model used: Xception-based and DenseNet-based. For both models we

applied MC DropFilter and MC Dropout for comparison. In both Xception and DenseNet models with MCDO Dropout is applied to each convolutional layer, while in MCDF models DropFilter is applied for each convolutional layer or convolutional+BatchNorm layers. Both MCDO and MCDF techniques can be implemented trivially in any modern NN framework.

Labels preprocessing. Following the idea from [Seferbekov, 2018], we used separate mask border predictions along with the whole mask prediction (whole tumor), which was proven to significantly improve final metrics. For this purpose, we preprocessed the initial labels and made the models to predict two masks at the same time. The auxiliary model head for the border prediction aims to train the model more effectively. Thus, the border prediction is not used for metrics evaluation.

Transfer learning.

We pre-trained all the models on ImageNet dataset, which is a common practice in various computer vision problems. ImageNet is the 3-channel images dataset. At the same time, the dataset used in our experiments includes 4 MRI modes. Therefore, the transfer learning for the fourth channel was performed. The transfer learning required the modifications of the network on the framework level, to train it. To our knowledge, there is no standard interface for such modifications. There are several options in the fourth channel weights initialization. In this work, we performed the fourth channel initialization with the first channel weights.

4.3 EXPERIMENTAL SETUP

We implemented the baselines and the proposed models in the Tensorflow 2 framework. For experiments, we used single GeForce GTX 1080 Ti GPU per each model’s training and evaluation. Training of the single NN took from 2 to 3 days depending on the architecture of the model.

For a quantitative evaluation Dice coefficient was used [Zou et al., 2004]. Based on it, the training was performed using Dice loss, which is calculated as $1 - Dice$, combined with the binary cross-entropy.

We used weight decay equal to 0.0001 and Adam optimizer with learning rate equal to 0.0001. The train/test/validation BraTS split was 75/15/10, respectively. All the reported results were obtained on the validation set.

For MC techniques, the number of samples per image N was equal to 20 according to the experimental results in [Wang et al., 2019a]. For the model calibration we used three metrics, namely Negative Log Likelihood (NLL) [Loquercio et al., 2020], Brier score (BS) [Ovadia et al., 2019] and Expected Calibration Error (ECE) [Naeini et al., 2015], which are standard for the calibration evaluation [Ashukha et al., 2019, Zhang et al., 2019]. For the uncertainty estimation we

used Mutual Information described in Section 2.

5 RESULTS

5.1 QUANTITATIVE RESULTS

Table 1 presents the evaluation results of the MCDF models compared to the baseline models and the MCDO models. The experiments on BraTS were performed with Dropout/DropFilter rates equal to 0.3. The table shows that the MC DropFilter models significantly outperformed the MC Dropout models by *Dice coefficient* and also by *NLL* and *BS* calibration metrics, which confirms that Dropout rather harmful for the convolutional filters. This answers the **RQ1**. Surprisingly, the MCDO models outperformed the MCDF models by ECE. This issue requires further research.

In addition, the MC DropFilter models significantly outperform the baseline models by Dice coefficient, which answers the **RQ2**. Finally, epistemic uncertainty estimation seems to decrease False Positive (FP) rate compared to baselines, which is crucial for our problem.

Table 1: Obtained results of the proposed MC Dropfilter (MCDF) technique compared to MC Dropout on two baselines, based on Xception and DenseNet NNs, on BraTS 2018 dataset. WT DC is Whole tumor Dice coefficient. NLL is scaled $\times 10^2$, BS and ECE are scaled $\times 10^3$.

Model	WT DC, %	NLL	BS	ECE
Xception	84.71	2.92	-	-
Xception MCDO	81.70	1.75	3.1	0.8
Xception MCDF	85.39	0.98	2.4	1.3
DenseNet	85.06	2.69	-	-
DenseNet MCDO	85.05	2.01	2.2	0.2
DenseNet MCDF	86.54	1.73	2.2	0.4

5.2 QUALITATIVE RESULTS

Fig. 4 and 5 outline the comparison of the Mutual Information (MI)-based uncertainty on BraTS images for MC Dropout (Fig. 4) and MC DropFilter techniques (Fig. 5). As can be observed, the lower uncertainty values (where the model is more confident) correspond to the brighter contrast values of tumor on the input MR image, while the higher values are likely obtained near the tumor borders. Fig. (5) shows that MCDF technique provides better model calibration, than MCDO: the tumor borders, presented by MCDF epistemic uncertainty, are more precise and more clearly expressed. This confirms the answer to the **RQ1**.

Such an uncertainty information can be decisive in the disease diagnosis and treatment. For example, the uncertainty

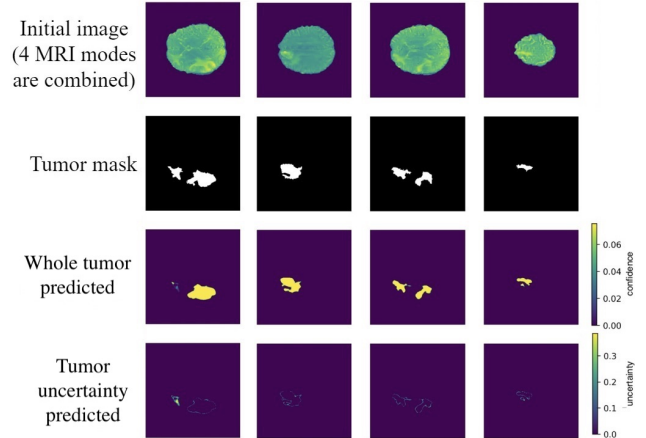


Figure 4: Examples of uncertainty estimation by *MC Dropout* technique based on the Mutual Information measure for BraTS 2018 dataset. Different MRI slices with prediction results are presented in columns. The first row demonstrates the input MR images of 4 modes combined in one, the second one presents the target masks, the third one presents the predicted mean masks, and the fourth one – the *epistemic uncertainty* masks predicted by *MC Dropout* technique, estimated based on Mutual Information.

visualization can help doctors recognize cases, when the deeper tumor investigation is required.

6 CONCLUSION

In the work presented, we have studied the application of uncertainty estimation for the brain tumor MRI segmentation problem. We have proposed to apply MC DropFilter technique for estimating epistemic uncertainty and compared this technique with classical MC Dropout. As the baselines we implemented two models, namely Xception and DenseNet, and evaluated them on BraTS 2018 dataset.

Xception MCDO and Xception MCDF models achieved 81.7% and 85.39% of Dice coefficient on BraTS dataset, respectively. DenseNet MCDO and Densenet MCDF models achieved 85.05% and 86.54% of Dice coefficient on BraTS dataset, respectively.

For the calibration ability evaluation, we performed MC Dropout/DropFilter multiple forward passes (20 times) during inference time. Xception MCDF provided the best NLL and BS metrics compared to Xception MCDO: 0.0098 and 0.0024, respectively. DenseNet MCDF also provided the best NLL and BS compared to DenseNet MCDO: 0.0173 and 0.0022, respectively. Hence, as Xception MCDF as well as DenseNet MCDF outperformed other models, they were selected as the best solutions.

Thus, the results showed that MC DropFilter is quite ef-

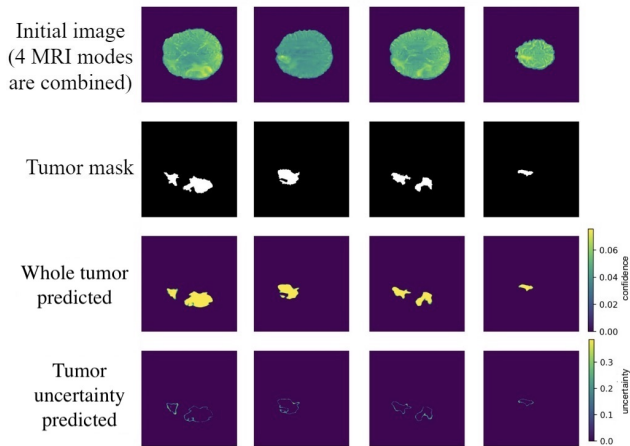


Figure 5: Examples of uncertainty estimation by the proposed *MC DropFilter* technique based on the Mutual Information measure for BraTS 2018 dataset. Different MRI slices with prediction results are presented in columns. The first row demonstrates the input MR images of 4 modes combined in one, the second one presents the target masks, the third one presents the predicted mean masks, and the fourth one – the *epistemic uncertainty* masks predicted by the proposed *MC DropFilter* technique, estimated based on Mutual Information.

fective for the problem of brain tumor image segmentation and improves model calibration. This technique can be useful for a wide range of applications, not limited to medical imaging.

In the future, further improvements and experiments are planned such as the developing of the Dropout rate impact on MC Dropout and MC DropFilter. Furthermore, we are planning to use NASNet [Zoph et al., 2018] as backbone and adopt ScheduledDropPath technique for epistemic uncertainty estimation.

Author Contributions

N. Khanzhina conceived the idea, developed the proposed methods, created the figures and wrote the paper. M. Kashirin implemented models and the proposed methods, conducted the experiments and wrote the paper. A. Filchenkov consulted on mathematical background and wrote the paper. I. Osmakov consulted on medical background. P. Ulyanov proposed the baseline models.

Acknowledgements

This work is financially supported by National Center for Cognitive Research of ITMO University, and Russian Science Foundation, Grant 19-19-00696.

The authors would like to thank Semyon Polyakov for useful ideas, Georgy Zamorin for the great help, Inna Anokhina, Alexander Grebennikov for useful comments.

References

- NHTSA PE 16-007. Tesla crash preliminary evaluation report. *Technical report, U.S. Department of Transportation, National Highway Traffic Safety Administration*, 2017.
- Arsenii Ashukha, Alexander Lyzhov, Dmitry Molchanov, and Dmitry Vetrov. Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. 2019.
- Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.
- José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.
- François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.
- Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.
- Jessica Guynn. Google photos labeled black people ‘gorillas’. *USA Today*, 2015.
- Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012. URL <http://arxiv.org/abs/1207.0580>.
- Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3): 457–506, 2021.

- Alain Jungo, Richard McKinley, Raphael Meier, Urspeter Knecht, Luis Vera, Julián Pérez-Beteta, David Molina-García, Víctor M Pérez-García, Roland Wiest, and Mauricio Reyes. Towards uncertainty-assisted brain tumor segmentation and survival prediction. In *International MICCAI Brainlesion Workshop*, pages 474–485. Springer, 2017.
- Alain Jungo, Raphael Meier, Ekin Ermis, Evelyn Herrmann, and Mauricio Reyes. Uncertainty-driven sanity check: application to postoperative brain tumor cavity segmentation. *arXiv preprint arXiv:1806.03106*, 2018.
- Zhengyan Kan, Ying Ding, Jinho Kim, Hae Hyun Jung, Woosung Chung, Samir Lal, Soonweng Cho, Julio Fernandez-Banet, Se Kyung Lee, Seok Won Kim, et al. Multi-omics profiling of younger asian breast cancers reveals distinctive molecular signatures. *Nature communications*, 9(1):1–13, 2018.
- S. Kaneko, K. Nomura, and T. et al. Yoshimura. Trend of brain tumor incidence by histological subtypes in japan: Estimation from the brain tumor registry of japan. *Neurooncol*, 60:61–69, 2002.
- Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems*, pages 6402–6413, 2017.
- Christian Leibig, Vaneeda Allken, Murat Seçkin Ayhan, Philipp Berens, and Siegfried Wahl. Leveraging uncertainty information from deep neural networks for disease detection. *Scientific reports*, 7(1):1–14, 2017.
- Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M Jorge Cardoso, and Tom Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. In *International conference on information processing in medical imaging*, pages 348–360. Springer, 2017.
- Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 936–944. IEEE Computer Society, 2017.
- Antonio Loquercio, Mattia Segù, and Davide Scaramuzza. A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, 5(2): 3153–3160, 2020.
- Raghav Mehta and Tal Arbel. Rs-net: Regression-segmentation 3d cnn for synthesis of full resolution missing brain mri in the presence of tumours. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 119–129. Springer, 2018.
- Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.
- Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.
- Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua V Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model’s uncertainty? evaluating predictive uncertainty under dataset shift. *arXiv preprint arXiv:1906.02530*, 2019.
- Kimberly R Porter, Bridget J McCarthy, Sally Freels, Yoonsang Kim, and Faith G Davis. Prevalence estimates for primary brain tumors in the united states by age, gender, behavior, and histology. *Neuro-oncology*, 12(6):520–527, 2010.
- Axel-Jan Rousseau, Thijs Becker, Jeroen Bertels, Matthew B Blaschko, and Dirk Valkenburg. Post training uncertainty calibration of deep networks for medical image segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1052–1056. IEEE, 2021.
- S. Seferbekov. Dsb2018 [ods.ai] topcoders 1st place solution. https://github.com/selimsef/dsb2018_topcoders, 2018.
- Mattia Segù, Antonio Loquercio, and Davide Scaramuzza. A general framework for uncertainty estimation in deep learning. *arXiv preprint arXiv:1907.06890*, 2019.
- Zhengsu Chen Jianwei Niu Qi Tian. Dropfilter: Dropout for convolutions. *arXiv preprint arXiv:1810.09849*, 2018.
- 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*, pages 648–656, 2015.
- Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI brainlesion workshop*, pages 178–190. Springer, 2017.
- Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In *International MICCAI Brainlesion Workshop*, pages 61–72. Springer, 2018.

- Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338:34–45, 2019a.
- Guotai Wang, Wenqi Li, Tom Vercauteren, and Sébastien Ourselin. Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. *Frontiers in computational neuroscience*, 13: 56, 2019b.
- Ramy A Zeineldin, Mohamed E Karar, Jan Coburger, Christian R Wirtz, and Oliver Burgert. Deepseg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance flair images. *International journal of computer assisted radiology and surgery*, 15 (6):909–920, 2020.
- Zhilu Zhang, Adrian V Dalca, and Mert R Sabuncu. Confidence calibration for convolutional neural networks using structured dropout. *arXiv preprint arXiv:1906.09551*, 2019.
- Yinhao Zhu and Nicholas Zabaras. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, 366:415–447, 2018.
- Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8697–8710, 2018.
- Kelly H Zou, Simon K Warfield, Aditya Bharatha, Clare MC Tempany, Michael R Kaus, Steven J Haker, William M Wells III, Ferenc A Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology*, 11(2):178–189, 2004.