BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 69(3), 2021, Article number: e136750
DOI: 10.24425/bpasts.2021.136750

Deep learning-based framework for tumour detection and semantic segmentation

Estera KOT[1], Zuzanna KRAWCZYK[1]*, Krzysztof SIWEK[1], Leszek KRÓLICKI[2], and Piotr CZWARNOWSKI[2]

[1] Warsaw University of Technology, Faculty of Electrical Engineering, Pl. Politechniki 1, 00-661 Warsaw, Poland
[2] Medical University of Warsaw, Nuclear Medicine Department, ul. Banacha 1A, 02-097 Warsaw, Poland

Abstract. For brain tumour treatment plans, the diagnoses and predictions made by medical doctors and radiologists are dependent on medical imaging. Obtaining clinically meaningful information from various imaging modalities such as computerized tomography (CT), positron emission tomography (PET) and magnetic resonance (MR) scans is the core method in software and advanced screening utilized by radiologists. In this paper, a universal and complex framework for two parts of the dose control process: tumours detection and tumours area segmentation from medical images is introduced. The framework formed the implementation of methods to detect glioma tumour from CT and PET scans. Two deep learning pre-trained models: VGG19 and VGG19-BN were investigated and utilized to fuse CT and PET examination results. Mask R-CNN (region-based convolutional neural network) was used for tumour detection – the output of the model is bounding box coordinates for each object in the image (tumour). U-Net was used to perform semantic segmentation: segment malignant cells and tumour area. The transfer learning technique was used to increase the accuracy of models while having a limited collection of the dataset. Data augmentation methods were applied to generate and increase the number of training samples. The implemented framework can be utilized for other use cases that combine object detection and area segmentation from grayscale and RGB images, especially to shape computer-aided diagnosis (CADx) and computer-aided detection (CADe) systems in the healthcare industry to facilitate and assist doctors and medical care providers.

Key words: deep learning; medical imaging; tumour detection; semantic segmentation; image fusion.

1. Introduction

Based on Cancer Research UK reports, brain, other central nervous systems (CNS) and intracranial tumours are the 9th most common cancer grouping in the UK. For over a decade in the UK (between 2005–2007 and 2015–2017), the cancer incidence rate for this specific group of cancers in women and men combined increased by 15%. In females, age-standardized incidence rates grew by 22%, and in males, rates increased by 8% [1].

Diagnoses and prognoses made by medical doctors and radiologists are dependent on medical imaging, i.e. obtaining clinically meaningful information from various imaging modalities such as CT and/or MR imaging. In certain advanced clinics and hospitals, PET scan is executed during CT scan, but since PET is invasive, its use is not a widespread practice.

In the Nuclear Medicine Department of the Medical University of Warsaw, at the Central Clinical Hospital, clinical trials are conducted using new treatment strategies for cases diagnosed with glioblastoma [2]. Based on the medical images, radiologists determine the tumour location, size, and the level of advancement, but also the amount of radiological substance injected into the interior of the tumour during the invasive examination or postoperatively after resection. The substance doses inflict damage on the tumour cell along with normal, noncancerous brain cells which render the process of measuring the tumour area/volume a critical factor in modern intensity-modulated radiotherapy (IMRT) planning and further treatment course selection.

Deep learning algorithms have been applied in numerous industries with great success. Researchers, together with developers, have produced evidence for many concept solutions that fully utilize algorithms based on advanced statistics. Nevertheless, there have been relatively few revolutionary deployments of deep learning systems over the last decade in common radiology practices. In modern research, there is a trend of attempting to work using medical data in healthcare environment application and transformation, which requires appropriate standards and frameworks in order to achieve applicable outcomes. Initial efforts have the potential result of shaping medical image processing, which in turn has influence on computing software deployment.

This paper proposes a universal and complex framework for two parts of the dose control process: tumours detection and tumours area segmentation, based on medical images. The framework is comprehensive and addresses the main challenges when working with medical data and applying modern deep learning techniques for automated tumour detection and segmentation. The framework is applied to shape CADx and CADe systems in the healthcare industry to assist doctors and facilitate their work. CADx/CADe objectives are achieved with the use of classification, detection, and prediction, while image processing tasks are performed with image segmentation, registration, and generation [3].

*e-mail: zuzanna.krawczyk@ee.pw.edu.pl

Manuscript submitted 2020-10-20, revised 2020-12-15, initially accepted for publication 2020-12-21, published in June 2021

The framework formed the implementation of the methods to detect glioma tumour from CT and PET scans developed and presented in [4]. Glioma, called glioblastoma, is the most malicious primary brain tumour [5], with a 100% mortality rate and the fourth histologic classification of the World Health Organization.

The framework consists of six key components presented in Fig. 1. The method commences with the pre-processing of image operations. The implementation was performed in the Python environment; however, not all Python libraries support image transformation in DICOM format, which is the optimal and desired choice for medical data. Consequently, conversion of DICOM to PNG/JPG format is required. DICOM contains essential information which is used in the subsequent analysis; for example, for volume computation, there is a pixel spacing attribute required. Saving original files or exporting DICOM parameters to a database can facilitate further development. To execute fusion, images must be rescaled to identical sizing. Fusion of the images can be executed in numerous manners; considerations are addressed in the exposition below. For tumour detection, the so-called "YOLO" network was considered, particularly in the section of related work, where numerous references to that deep learning model can be found. An alternative network to YOLO is Mask R-CNN, proposed as it can also perform object instance segmentation, which can be beneficial in further research. For tumour segmentation, the U-Net model was selected as this model is dedicated to biomedical images and, in previous papers, it produced promising results. Volume computation is based on the tumour square on each slice and pixel spacing, i.e. the physical distance in the patient between the centre of each pixel. The final component, i.e. visualization of the tumour, is essential for the radiologist.

All deep learning architectures used in this research are based on a convolutional neural network (CNN). A CNN refers to a network architecture composed of several stacked convolutional layers [6]. The convolution procedure recognizes locally connected information, i.e. the neighbouring voxels or pixels. Expanding the area of view of the network is possible due to the presence of pooling layers between the convolution layers. It takes a part of the locally connected nodes of the input layer and results in an output having a smaller spatial footprint [6]. A series of fully connected layers combines all activations of the previous layers. Finally, the model outputs the final set of feature values relevant to the given task.

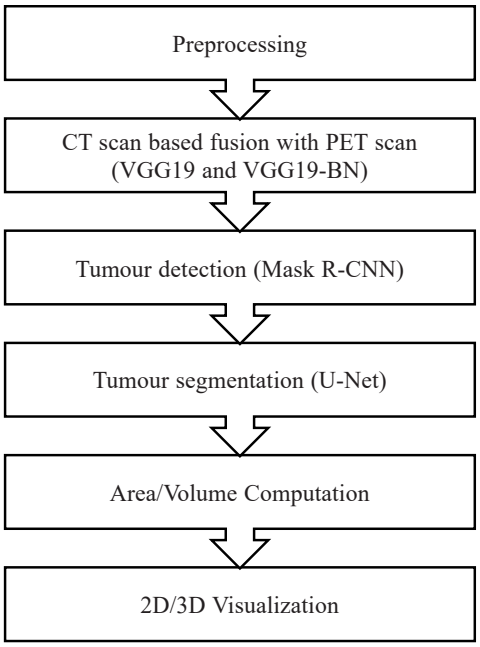


Fig. 1. Key components of the proposed framework

2. Related works

A common pre-requisite for therapeutic planning is detection and identification of the region, i.e. the voxel positions associated with the diseased condition. The authors [7] have successfully used the deep learning method, ROI-based convolutional neural network (CNN) or "YOLO," for detecting changes in mammography images, obtaining as much as 96.33% accuracy. The model in question conducted automatic detection and classification. A model capable of processing a single image in 43 ms was built [8]. Based upon the possibilities presented in the paper while maintaining accuracy within statistical reason, a complete analysis of the study for a single patient may take up to 10 seconds. The YOLO network is also used to detect and locate tumours in the lungs employing CT images [9]. Detection of early stage brain tumours based on MR imaging using the AlexNet model and Faster R-CNN making use of transfer learning has been described in [10, 11]. The new XmasNet architecture completion of the classification of the senile neoplasm based on MRI examinations was presented in [12]. Finally, in [13] a brain tumour detection and classification system were implemented using CWT, DWT and SVMs.

The segmentation of medical images is also a topic covered by contemporary scientists. Fully automated segmentation of brain tumours from MR images was presented in paper [14] using the U-Net [15, 16]. MRI processing was also examined in [17], where the Faster R-CNN network was deployed to classify and segment tumours. Custom CNN network was also used in [18] to perform localization of multiple sclerosis lesions in MRI images.

3. Image fusion

Tumour structures are nearly undetectable on CT scans. This type of image presents the soft tissues as well as the structure of the brain. To base a diagnostic decision on CT, a PET scan is executed during a medical examination via the same CT-PET scanner. As a result, two series of slices per each type of examination are returned. Slices are correlated but not merged. For further analysis processes, both for the radiologist and these research purposes, the fusion of CT and PET scans must be executed.

CT scan is 512×512 which results in 262 144 pixels. The pixels with the lowest value (zeros) are displayed as black,

and the highest values as white. The structure of the brain is reflected in CT scans as a level of grey, which indicates the degree of absorption of X-rays determined by Hounsfield units (HU).

The proper sets of training and validating data determine whether the developed model distinguishes between abnormal and normal brain matter regions and classifies each correctly as a benign tumour, a malignant tumour or healthy tissue. The dataset of twenty-two patients used for this research purposes was delivered and acquired by the Nuclear Medicine Department. Images were taken using a Siemens Biograph 64 PET-CT scanner during the years 2016–2018, both 148 slices for PET and CT scan. Each slice weighs around 520 KB netting approximately 110 MB for a single examination of a CT and correlated PET scan.

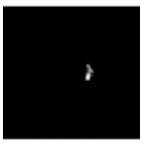
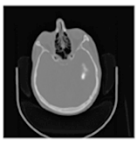
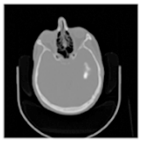
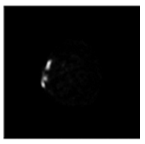
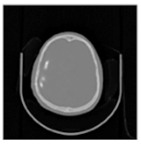
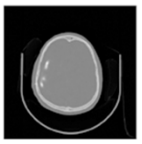
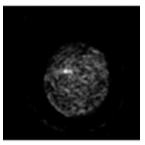
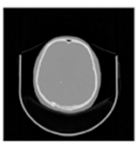
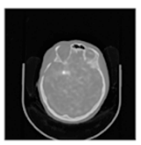
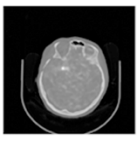
With the increase in the volume of data, especially when working with MR scans that can weigh 1.5 GB for a single examination, traditional machine learning algorithms may be incapable. Modern deep learning models are most promising. The basic computational unit is the neuron. This concept grew from the study of the human brain, which takes many signals as inputs, combines them linearly using weights, and then transfers the combined signals by nonlinear operations to generate output signals [3]. To produce accurate results, deep learning models require massive amounts of training datasets, where annotations are made by medical experts, which makes a medical dataset harder to obtain. To mitigate this challenge, two approaches can be taken: transfer learning technique and data augmentation. Transfer learning facilitates the transfer of features from non-medical data. Data augmentation facilitates generating and increasing the number of training samples via methods such as cropping, random rotation, transposing, elastic transform, etc.

3.1. VGG19 and VGG19-BN. To fuse CT and PET scans, the method described in [19] was applied, initially developed to fuse infrared and visible images. Experiments confirmed the methods could be successfully applied for medical image fusion. The methods firstly take the source images (CT and PET scans) and decompose them into base parts and detailed content. Then a weighted-averaging fusion strategy is applied to fuse the base parts. To extract the detailed content, a pretrained deep learning network (VGG19 or VGG19-BN) is applied to compute multi-layer features.

VGG19 is a variant of the VGGNet – a deep convolutional neural network proposed in [20]. The name of this model was inspired by the name of the research group 'Visual Geometry Group (VGG)'.The VGG19 has 19 layers. VGG-19-BN is an extended VGG19 model with batch normalization, i.e. a layer between each convolutional and activation unit layer as well as between each inner product and activation unit layer [21]. Implementation is available in [22]. Both models are trained on ImageNet to extract deep features. ImageNet dataset consists of various types of objects which can be used for training purposes. The transfer learning from a pre-trained ImageNet network significantly improves the results on every dataset, which is an attempt at compensating for the lack of adequate training data [23].

The implementation of the fusion experiment coded in Python utilizing PyTorch library can be found in [24]. Table 1 presents a selection of obtained graphical results.

Table 1
Results for fusion experiment

Slice no.	PET scan	CT scan	Fusion with VGG19	Fusion with VGG19-BN
1				
2				
3				

Each slice processing takes approximately 6 to 7 s of computation for VGG19 and VGG19-BN, approximately 8 s on 2 nodes of Virtual Machine with 224 GB of memory and 4 GPUs (Tesla K80).

3.2. Jaccard distance. A fused image is expected to present more shaped, contrasting details and less artificial noise. For quantitative comparison of the obtained images, as well as for measuring the performance of segmentation, methods such as the dice coefficient (DICE) or average symmetric surface distance coefficient (ASSD) [25], accuracy, precision, recall, F-measure, G-measure, and Jaccard distance [26] can be used. In this paper, the Jaccard coefficient was applied to measure the similarity between two fusion results. Jaccard distance reaches its best value at 1 and worst value at 0, indicating a lower similarity [26]. The results comparing five fusion outcomes are presented in Table 2.

Table 2
Jaccard coefficient for fused images with VGG19 and VGG19-BN models

Method	Slice no.				
	1	2	3	4	5
Jaccard coefficient	0.9823	0.9837	0.9851	0.9931	0.9479

Results confirm the impact of applying batch normalization [27] in a fusion model, as long as the same deep learning model architecture was chosen. Fig. 2 presents the differences in pixel intensities in the obtained fused images. Moreover, the histogram is useful to confirm if there is any significant noise affecting the image. The original examination results are saved

in DICOM format, and in those cases, the histograms should reflect intensities in HU. Any transformation of an image, especially conversion to PNG/JPG formats, can change pixel intensity values. The maximum pixel intensity value obtained from the PNG files was 255 when on the original CT was 2,692, and on the original PET DICOM file, 23,706.

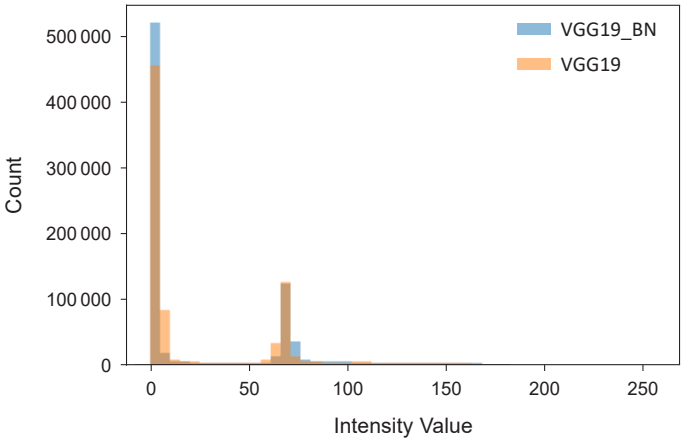


Fig. 2. Histogram for fused images

Choosing a different deep learning model to fuse images can lead to producing different output images and modifying pixel intensity values. Detecting tumours, especially segmenting tumour area, is based on pixel intensities, independent of an image format. For volume computation to determine a further treatment plan; pixel intensity is crucial, which is why choosing the correct method for fusion can influence the results.

Fusion with VGG19-BN model produces smaller contrast results, and for further experiments, results obtained with the usage of VGG19 are used.

4. Tumour detection

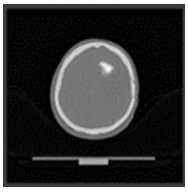
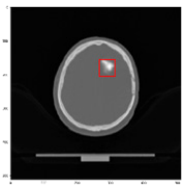
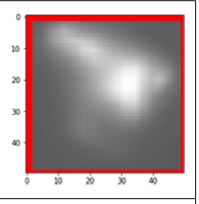
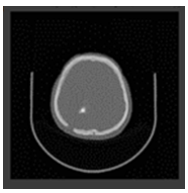
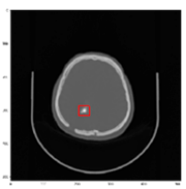
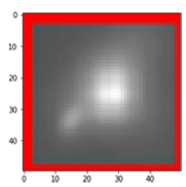
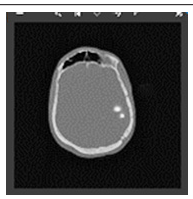
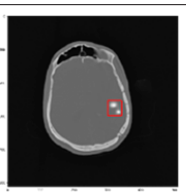
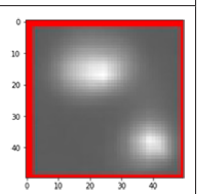
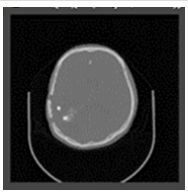
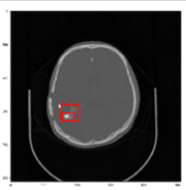
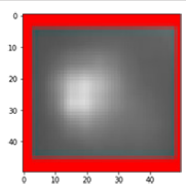
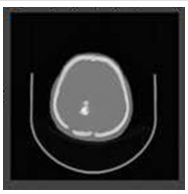
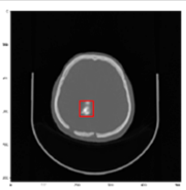
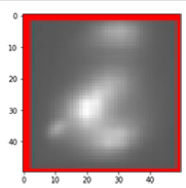
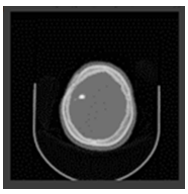
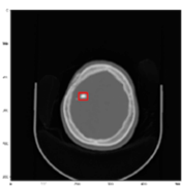
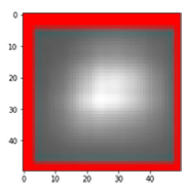
Mask R-CNN is an extension of Faster R-CNN [28]. The model is flexible and can be utilized for object classification, object detection and instance segmentation. Mask R-CNN returns the class label, bounding box coordinates and the object mask for each object in the image.

4.1. Implementation details. The pretrained model was used on the COCO dataset [29]. Pretrained models contain the weights and biases that represent the features the dataset was trained on. Learned features can be transferable to different data. Init models can be downloaded from [30] – Mask R-CNN 2.0. Next, the first training for all layers (except the initial layer which was frozen) was implemented. The chosen parameters were five epochs and a learning rate of 0.001. After the initial run, a second training of all layers during ten epochs and with a learning rate of 0.0001 was completed.

Recreation of the model in inference mode and load trained weights was the primary stage. Random testing followed, and finally, 10 images not used during the training were taken. The best-achieved model is available in [31].

4.2. Results. Tumour detection takes approximately 10 s in the previously discussed infrastructure. Cropping is accomplished in less than 1 s. Results are presented in Table 3.

Table 3
Mask R-CNN model results

Slice no.	Source image	Detection	Cropping
1			
2			
3			
4			
5			
6			

Mask R-CNN can be successfully applied to detected tumours. For all slides where tumour was clearly present, the model gives the correct results and detects the tumour. However, the sensitivity of the trained model is too high, especially for the slices where the analysed tumour (glioma) cannot exist. The results of the working model presented in Fig. 3 are not acceptable. The results for slice 6 from Table 3 present the algo-

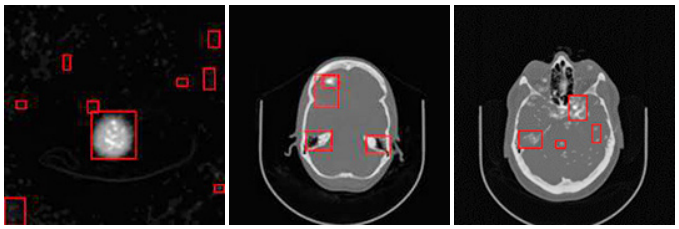


Fig. 3. Overdetection in slices where glioma cannot be present due to its nature

rithm behaviour when two boxes are drawn, one should not be inside the other.

The solution for this challenge is to analyse only the slices in which the tumour is visible. Based on statistical experiments, the tumour is not visible in slices from 0 to 60 and from 110 to 148. Processing exclusively the scans where the tumour can be located not only reduces over-detection but also drastically improves the performance via reducing the time of computation by almost 2/3.

The final algorithm cropped a 64×64 size image as the input for the next step of the proposed method, i.e. segmentation. An alternative approach could be to calculate the size of a final output based on the equation:

$$\rho = max\,(\alpha, \beta) + 30\ pixels, \tag{1}$$

where α and β are the tumour width and height and 30 pixels were added as small surrounding area [32].

5. Tumour segmentation

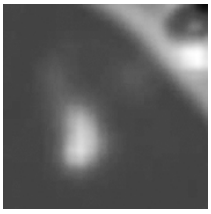
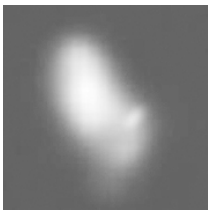
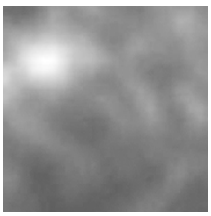
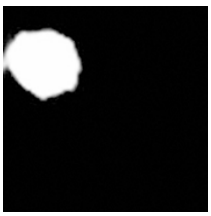
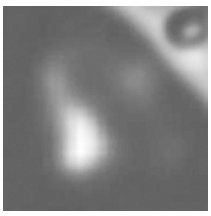
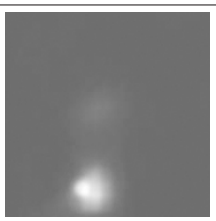
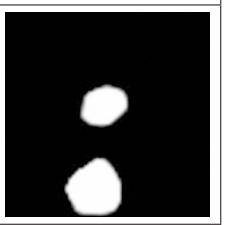
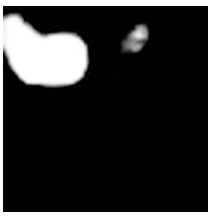
The objective of the semantic segmentation is to assign an object category label for each pixel in the image. The U-Net is the extension of a fully convolutional network (FCN) [33], i.e. the first end-to-end architecture proposed for semantic segmentation. The U-Net produces good results even with a large dataset absent when data augmentation techniques are used [15].

5.1. Implementation details. The U-Net model implemented in Keras was adjusted to run on smaller images than originally designed (64×64 resolution images). The adjusted model architecture implementation is presented in [34]. The model was trained with a prepared dataset where tumour areas were manually masked. The steps per epoch number were calculated as the number of trained images divided by the batch size set to 16. Data augmentation was used to increase the dataset variance to be able to feed the deep learning neural network. There were several techniques utilized in augmenting data samples such as rotations, shear intensity, zoom, channel shifts, and horizontal flips.

Output from the network is 64×64, which represents a mask. In the sigmoid activation function, mask pixels are in [0, 1] range. The loss function for the training is simply a binary cross-entropy. The model was trained for 10 epochs achieving the accuracy of almost 97% on the training dataset and similar on the validation dataset.

5.2. Results. Tumour area segmentation for a single patient (about 50 slices) takes around 15 s in the same infrastructure as fusion was completed. The visual results are presented in Table 4. The infrastructure choice increased in its importance, especially during training. The main inconveniences of CNNs are both the enormous amount of computational power and the amount of time necessary to train the networks. The strict

Table 4
Results for U-Net semantic segmentation

Slice no.	Source image	Segmented tumour area
1		
2		
3		
4		
5		
6		

requirement for training is GPU. For interference mode, it was concluded that the CPU working well and validating FPGA (field-programmable gate arrays) is necessary in order to possibly accelerate the calculations.

The obtained results are entirely satisfactory. Loading of pretrained weights significantly increases the accuracy and reduces the time for model training.

6. Conclusions and future work

In this paper, the framework for tumour detection and segmentation in medical images is presented. It is understood that the frames for operations involving object detection and segmentation tasks using grayscale images can be applied in other cases when dealing with solid structure tumours, such as the detection of tumours from MRI images.

The primary challenging aspect of applying deep learning methods in medicine is the size of any given dataset which will feed and validate the CNNs. Our research for this paper combines CT and PET scans, with fused images. The obtained results are similar to MRI images, which gives hope that the methods and models trained on MRI images, due to the transfer learning technique, can be successfully run on the fusion of CT and PET. Reciprocally, models trained on fused images can work on MRI. However, experimental confirmation is necessary.

The object detection on greyscale images is of greater difficulty compared to the identical task using RGB images. The presented framework can be applied to object detection in general and the Mask R-CNN model is recommended. When there is no need for volume or area computation, then the segmentation step can be omitted. For tumour segmentation, one should consider using instance segmentation for a separate analysis of small tumours, e.g. primary research focus was on the main body of a tumour. However, glioma can consist of the main tumour with additional lesser in the vicinity.

The objective of tumour detection and segmentation could be achieved in one step, i.e. using only U-Net. The main challenge is that setting up training as a heterogeneous dataset of different shapes of glioma should be prepared. Moreover, for that task, standard methods such as local thresholding and processing slices only when the global threshold for tumour is met should be applied.

The subsequent development for these methods is to follow the created framework and evaluate the performance of the proposed image segmentation algorithm using the Dice coefficient. Evaluation of the proposed algorithm in MR scans with brain tumours and prostate should continue. It would be beneficial to work on 3D tumour visualisation inside a 3D skull with colour-coded tumours. A proposal to release the algorithm as a web app for initial use by the radiologists in cooperation with the research team has been considered.

Over the previous year, due to an increase of detected cancers and SARS-COV-2 global pandemic, there has been increased human awareness of the reliance on the healthcare sector along with overall demand for higher, faster, and more accurate standards of care. However, the interpretations of medical data can be made only by medical experts. Their number is not limitless, and they are in high demand in every hospital or radiological centre. It can be confidently attested that all automatic systems which utilize modern deep learning techniques will be in high demand. Radiologists can use the results of the model for tumour detection during the screening tests, thereby decreasing the time necessary while increasing the accuracy. Oversensitivity of the model detects all areas of pixel intensity including abnormalities which cannot be observed by humans.

Acknowledgements. This research is part of the "Design and development of algorithms for automatic identification, volume measurement and visualization of brain tumours and prostate cancer using deep learning methods" project and has been supported by the Warsaw University of Technology as part of the grant 504/04560/1042/43.040001 since 8th of July 2020.

References

[1] Cancer Research UK Statistics from the 5th of March 2020. [Online]. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours/incidence#ref-

[2] E. Kot, Z. Krawczyk, K. Siwek, and P.S. Czwarnowski, "U-Net and Active Contour Methods for Brain Tumour Segmentation and Visualization," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9207572.

[3] J. Kim, J. Hong, H. Park, "Prospects of deep learning for medical imaging," *Precis. Future. Med*. 2(2), 37–52 (2018), doi: 10.23838/pfm.2018.00030.

[4] E. Kot, Z. Krawczyk, and K. Siwek, "Brain Tumour Detection and Segmentation Using Deep Learning Methods," in *Computational Problems of Electrical Engineering*, 2020.

[5] A.F. Tamimi and M. Juweid, "Epidemiology and Outcome of Glioblastoma," in: *Glioblastoma* [Online]. Brisbane (AU): Codon Publications, 2017, doi: 10.15586/codon.glioblastoma.2017.ch8.

[6] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," in: *Advances in Neural Information Processing Systems*, 2012, p. 1097–1105.

[7] M.A. Al-masni, et al., "Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network," *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, 2017, pp. 1230–1233, doi: 10.1109/EMBC.2017.8037053.

[8] P. Yin, R. Yuan, Y. Cheng, and Q. Wu, "Deep Guidance Network for Biomedical Image Segmentation," *IEEE Access* 8, 116106–116116 (2020), doi: 10.1109/ACCESS.2020.3002835.

[9] R. Sindhu, G. Jose, S. Shibon, and V. Varun, "Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans", Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis, 105751I, 2018, doi: 10.1117/12.2293699.

[10] R. Ezhilarasi and P. Varalakshmi, "Tumor Detection in the Brain using Faster R-CNN," *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Palladam, India, 2018, pp. 388–392, doi: 10.1109/I-SMAC.2018.8653705.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-timeobject detection with region proposal networks," in *Advances in neuralinformation processing systems*, 2015, pp. 91–99.

[12] S. Liu, H. Zheng, Y. Feng, and W. Li, "Prostate cancer diagnosis using deeplearning with 3D multiparametric MRI," in *Proceedings of Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, Bellingham: International Society for Optics and Photonics (SPIE), 2017. p. 1013428.

[13] M. Gurbină, M. Lascu, and D. Lascu, "Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and Support Vector Machines," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, 2019, pp. 505–508, doi: 10.1109/TSP.2019.8769040.

[14] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in: *Medical image understanding and analysis*, pp. 506–517, eds. Valdes Hernandez M, Gonzalez-Castro V, Cham: Springer, 2017.

[15] 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, Lecture Notes in Computer Science*, vol 9351, doi: 10.1007/978-3-319-24574-4_28.

[16] K. Hu, C. Liu, X. Yu, J. Zhang, Y. He, and H. Zhu, "A 2.5D Cancer Segmentation for MRI Images Based on U-Net," in *2018 5th International Conference on Information Science and Control Engineering (ICISCE)*, Zhengzhou, 2018, pp. 6–10, doi: 10.1109/ICISCE.2018.00011.

[17] H.N.T.K. Kaldera, S.R. Gunasekara, and M.B. Dissanayake, "Brain tumor Classification and Segmentation using Faster R-CNN," *Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2019, pp. 1–6, doi: 10.1109/ICASET.2019.8714263.

[18] B. Stasiak, P. Tarasiuk, I. Michalska, and A. Tomczyk, "Application of convolutional neural networks with anatomical knowledge for brain MRI analysis in MS patients", *Bull. Pol. Acad. Sci. Tech. Sci.* 66(6), 857–868 (2018), doi: 10.24425/bpas.2018.125933.

[19] L. Hui, X. Wu, and J. Kittler, "Infrared and Visible Image Fusion Using a Deep Learning Framework," *24th International Conference on Pattern Recognition (ICPR)*, Beijing, 2018, pp. 2705–2710, doi: 10.1109/ICPR.2018.8546006.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[21] M. Simon, E. Rodner, and J. Denzler, "ImageNet pre-trained models with batch normalization," arXiv preprint arXiv:1612.01452, 2016.

[22] VGG19-BN model implementation. [Online]. https://pytorch.org/vision/stable/_modules/torchvision/models/vgg.html

[23] D. Jha, M.A. Riegler, D. Johansen, P. Halvorsen, and H.D. Johansen, "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation," *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, MN, USA, 2020, pp. 558–564, doi: 10.1109/CBMS49503.2020.00111.

[24] Jupyter notebook with fusion code. [Online]. https://github.com/ekote/computer-vision-for-biomedical-images-processing/blob/master/papers/polish_acad_of_scienc_2020_2021/fusion_PET_CT_2020.ipynb

[25] E. Geremia et al., "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images", *NeuroImage* 57(2), 378–390 (2011).

[26] D. Anithadevi and K. Perumal, "A hybrid approach based segmentation technique for brain tumor in MRI Images," *Signal Image Process.: Int. J.* 7(1), 21–30 (2016), doi: 10.5121/sipij.2016.7103.

[27] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," arXiv preprint arXiv:1502.03167.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.* 39(6), 1137–1149, (2017), doi: 10.1109/TPAMI.2016.2577031.

[29] T-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Lawrence Zitnick, "Microsoft COCO: common objects incontext" in *Computer Vision – ECCV 2014*, 2014, p. 740–755.

[30] Original Mask R-CNN model. [Online]. https://github.com/matterport/Mask_RCNN/releases/tag/v2.0

[31] Mask R-CNN model. [Online]. https://github.com/ekote/computer-vision-for-biomedical-images-processing/releases/tag/1.0, doi: 10.5281/zenodo.3986798.

[32] T. Les, T. Markiewicz, S. Osowski, and M. Jesiotr, "Automatic reconstruction of overlapped cells in breast cancer FISH images," *Expert Syst. Appl.* 137, 335–342 (2019), doi: 10.1016/j.eswa.2019.05.031.

[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2015, pp. 3431–3440.

[34] The U-Net architecture adjusted to 64×64 input image size. [Online]. http://bit.ly/unet64x64