Title

Automatic detection and segmentation of multiple brain metastases on magnetic resonance image using asymmetric UNet architecture.

Journal

Physics in Medicine & Biology, 66(1)

Authors

Cao, Yufeng
Vassantachart, April
Ye, Jason
et al.

Publication Date

2021-01-13

DOI

10.1088/1361-6560/abca53

Peer reviewed

HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2024 August 05.

Published in final edited form as:

Phys Med Biol. ; 66(1): 015003. doi:10.1088/1361-6560/abca53.

Automatic detection and segmentation of multiple brain metastases on magnetic resonance image using asymmetric UNet architecture

Yufeng Cao[1], **April Vassantachart**[1], **Jason C Ye**[1], **Cheng Yu**[2], **Dan Ruan**[3], **Ke Sheng**[3], **Yi Lao**[3], **Zhilei Liu Shen**[1], **Salim Balik**[1], **Shelly Bian**[1], **Gabriel Zada**[2], **Almon Shiu**[1], **Eric L Chang**[1], **Wensha Yang**[1]

[1]Department of Radiation Oncology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States of America

[2]Department of Neurological Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States of America

[3]Department of Radiation Oncology, University of California Los Angeles, Los Angeles, CA, United States of America

Abstract

Detection of brain metastases is a paramount task in cancer management due both to the number of high-risk patients and the difficulty of achieving consistent detection. In this study, we aim to improve the accuracy of automated brain metastasis (BM) detection methods using a novel asymmetric UNet (asym-UNet) architecture. An end-to-end asymmetric 3D-UNet architecture, with two down-sampling arms and one up-sampling arm, was constructed to capture the imaging features. The two down-sampling arms were trained using two different kernels ($3 \times 3 \times 3$ and $1 \times 1 \times 3$, respectively) with the kernel ($1 \times 1 \times 3$) dominating the learning. As a comparison, vanilla single 3D UNets were trained with different kernels and evaluated using the same datasets. Voxel-based Dice similarity coefficient (DSC_v), sensitivity (S_v), precision (P_v), BM-based sensitivity (S_{BM}), and false detection rate (F_{BM}) were used to evaluate model performance. Contrast-enhanced T1 MR images from 195 patients with a total of 1034 BMs were solicited from our institutional stereotactic radiosurgery database. The patient cohort was split into training (160 patients, 809 lesions), validation (20 patients, 136 lesions), and testing (15 patients, 89 lesions) datasets. The lesions in the testing dataset were further divided into two subgroups based on the diameters (small S = 1–10 mm, large L = 11–26 mm). In the testing dataset, there were 72 and 17 BMs in the S and L sub-groups, respectively. Among all trained networks, asym-UNet achieved the highest DSC_v of 0.84 and lowest F_{BM} of 0.24. Although vanilla 3D-UNet with a single $1 \times 1 \times 3$ kernel achieved the highest sensitivities for the S group, it resulted in the lowest precision and highest false detection rate. Asym-UNet was shown to balance sensitivity and false detection rate as well as keep the segmentation accuracy high. The novel asym-UNet segmentation network showed overall competitive segmentation performance

Wensha.Yang@med.usc.edu .

and more pronounced improvement in hard-to-detect small BMs comparing to the vanilla single 3D UNet.

Keywords

brain metastasis; medical imaging; asymmetric Unet

1. Introduction

Brain metastasis (BM) develops in approximately one-third of cancer patients throughout their disease courses (Walker *et al* 1985). Treatment options for BMs include whole-brain radiation therapy (WBRT) or stereotactic radiosurgery (SRS), with or without surgical resection (Suh *et al* 2020). In patients with a limited number of BMs, SRS has been shown to reduce cognitive decline and improve quality of life compared to WBRT (Brown *et al* 2016, Sahgal *et al* 2015, Yamamoto *et al* 2017). However, close surveillance with magnetic resonance imaging (MRI) is necessary to assess for local and distant recurrences (Chang *et al* 2009). Earlier detection of BMs with advanced neuroimaging and improvement of systemic therapies has increased survival in patients with BMs, further demanding accurate assessment of follow-up MRIs (Chang *et al* 2018). These factors necessitate the review of multiple, longitudinal MRIs, which can be challenging due to the volume of images needing evaluation and astute detection of sub-centimeter BMs. Fast and accurate detection and segmentation of new and recurrent BMs with the assistance of a computer algorithm is thus of paramount clinical importance.

Computer-aided diagnosis (CAD) has become one of the major research topics in medical imaging and diagnostic radiology. In CAD, clinicians use the computer as a ‘second opinion’ to assist their final decision (Abd-Ellah *et al* 2019). CAD is useful for lesion detection and segmentation of BMs for SRS treatment planning and can serve as a second check of tumor contours. However, vanilla computer vision methods have not been widely adopted because they still heavily rely on human supervision, which did not significantly alleviate the workload (Doi 2007). The development of deep learning has pushed the limits of what is possible in the domain of digital image processing. Significantly improved performance has been reported on solving difficult problems, including image colorization, classification, segmentation, and detection using deep learning. Among them, convolutional neural networks (CNNs) have been mostly studied to improve prediction performance using a large amount of pre-labeled data (Lecun *et al* 1989, Chen *et al* 2016). Specifically, CNNs have achieved remarkable success in 2D and 3D medical image segmentation (Ronneberger *et al* 2015, Long *et al* 2015, Tong *et al* 2018, 2019, Tang *et al* 2019, Gou *et al* 2020). Nonetheless, relatively few studies have reported the detection and segmentation of the BMs (Liu *et al* 2017). Among them, Losch *et al* were the first to investigate the use of deep learning to detect and segment BMs (Losch *et al* 2015). Charron *et al* adapted an existing 3D CNN (DeepMedic) by training 164 patients to detect and segment BMs on multimodal MR images. The authors used Dice similarity coefficient (DSC) which quantifies the normalized overlap between the manual and automated segmentation, and achieved an average DSC of 0.77 with the best DSC score of 0.79 (Charron *et al* 2018). Grovik *et al* built a 2.5D

GoogleNet CNN to detect and segment BMs on multisequence MRI (Grøvik *et al* 2020). One hundred five training data points were involved, and the average of voxel DSC is 0.79 calculated based on patients. Although the average performance could be clinically useful, these studies used a combined DSC measurement for a patient with multiple BMs of varying sizes, which is insensitive to the detection and segmentation errors of small BMs. Realizing this limitation, Hu *et al* divided the brain metastases into small size (<1500 mm^3) and large size (>1500 mm^3) when combining 3D UNet and DeepMedic to detect and segment BMs (Hu *et al* 2019). Their average detection sensitivities are 0.61 and 0.98 for small- and large-sized BMs, respectively. Their average voxel-based DSC is 0.47 and 0.82 for small- and large-sized BMs. Besides the relaxed threshold set for small BM, which has an equivalent diameter of 14 mm, the steep decline in detection sensitivity and accuracy with BM size suggests the challenge of small BM detection. Further improvement in the segmentation performance, particularly for lesions smaller than 10 mm, is essential for the clinical utility of automated methods. In the study, we propose a novel asymmetric UNet (asym-UNet) architecture to detect and segment brain metastases using a single T1 contrast-enhanced MR sequence.

2. Methods and materials

2.1. Clinical data

Following IRB approval, contrast-enhanced T1 MR images from 195 consecutive patients with multiple BMs were solicited from our institutional SRS database. We focus on more challenging multiple BMs detection and segmentation, so patients with single BM were not included. The BM segmentations were verified by radiation oncologists with >5 years of experience of treating BMs with SRS. A total of 1034 BMs in the 195 patients were manually contoured for SRS planning and treatment. Figure 1 shows the distribution of BMs in axial, sagittal, and coronal views. BMs from all patients were overlaid in a common brain mask to qualitatively demonstrate the locations of metastases in the main from all cancer types. The patient cohort was split into training (160 patients, 809 lesions), validation (20 patients, 136 lesions), and testing (15 patients, 89 lesions) datasets. The means and range of ages for training, validation, and testing data are 60.4 (19–92), 59.5 (28–81), 56.8 (25–68), respectively. The 89 lesions in the testing dataset were further divided into two groups based on their largest diameter: small, S (1–10 mm, N = 72) and large, L (11–26 mm, N = 17). Figures 2(a) and (b) show the distribution of BM binned by the diameters in the training and validation datasets.

2.2. Image pre-processing

Treatment planning and subsequent surveillance MR scans were performed with T1 sequence after intravenous gadolinium contrast injection on 1.5T Symphony™ (Siemens, München, Germany), with original volumetric dimensions of 256 × 256 × 80, a spatial resolution of 1.016 × 1.016 × 2 mm, TR (repetition) of 738 ms, and TE (echo time) of 16 ms. FOV (percent phase field of view) of 100 mm^2 and flip angle of 180 °. BMs contours were created by neurosurgeons, physicists, and radiation oncologists at the time of the gamma knife treatment planning. DICOM files containing the images and tumor contours were exported to VelocityAI™ (Varian, Palo Alto, CA). Whole-brain contours were then

created, excluding the skull and skin. The organized dataset, including original MR images, the tumor, and whole-brain contours, were then exported to a LINUX computational server, equipped with 4× RTX 2080 Ti 11GB GPU and Devbox (X299, I9–7900x, 16GB × 4, 1× 240GB OS SSD, 1× 10TB DATA HDD, Ubuntu 16.04) for model training and testing.

2.3. Asymmetric 3D UNet architecture

As shown in figure 3, a 3D asym-UNet architecture was built with two down-sampling arms and one up-sampling arm to capture the spatial information. For the down-sampling arms, 3 × 3 × 3 (k1) and 1 × 1 × 3 (k2) were used to capture image features with different encoding paths, which avoids fusing from the early stage. We selected k1 and k2 based on the image resolution (1.016 × 1.016 × 2 mm) and typical BM diameters (>50% of BM are smaller than 7 mm) in our datasets. BM and neighboring brain tissue were noted to have distinct features. When a BM was small and surrounded by a large amount of normal brain tissue, a smaller kernel may be more attentive to local BM texture and not be overly distracted by the neighborhood. For this reason, k2 dominated the learning process. k1 was necessary as a supplement to learn features from neighboring tissue because more image features were incorporated in the 3 × 3 × 3 kernel. Inside each encoding path, the corresponding kernel convolution was applied twice with a rectified linear unit, a dropout layer between the convolutions with a dropout rate of 0.3 and a 2 × 2 × 2 max-pooling operation in each layer (Srivastava *et al* 2014). The number of feature channels was doubled after the max-pooling operation. In the up-sampling path, each layer consists of an up-convolution of 2 × 2 × 2, followed by two 3 × 3 × 3 convolution operations to reduce the feature channels by half. We then concatenated the channels from the corresponding layers. In the final step, a 1 × 1 × 1 convolution and softmax were used to map the feature vector to binary classes. Glorot (Xavier) normal initializer was used for this asym-UNet, which drew samples from a truncated normal distribution centered on zero with stdev = sqrt $(2/(\text{fan}_{\text{in}} + \text{fan}_{\text{out}}))$ (Glorot and Bengio 2010). fan_{in} and fan_{out} are the numbers of input and output units, respectively, in the weight tensor. The energy function was computed using the cross-entropy loss function with a pixel-wise softmax at the final step, which was defined as:

$$p_k(x) = \frac{exp(f_k(x))}{\sum_{k'=1}^{C} exp(f_{k'}(x))} \tag{1}$$

where $f_k(x)$ was the activation in feature channel k at the pixel position x. C was the number of classes and $p_k(x)$ was the approximated maximum function. $p_k(x) \approx 1$ for channel k had the maximum activation $f_k(x)$, otherwise $p_k(x) \approx 0$. Here, the cross-entropy at each position was defined as:

$$E = -\sum_{x} y_k(x) \log(p_k(x)) + (1 - y_k(x)) \log(1 - p_k(x)) \tag{2}$$

where y_k was the true label at the position, which was 0 or 1 for this binary case. Adam optimizer was applied to train this model (Kingma and Ba 2014). The learning rates ranging

from 1×10^{-4} to 1×10^{-6} were investigated in this project. The model did not converge with the learning rate of 1×10^{-6} and up to 1200 epochs. For the learning rate 1×10^{-4}, the network converged quickly at epochs around 200. However, it converged to an undesirable local minimum, and all BM less than 4 mm were missed, and DSC_V was lower. With the learning rate 6×10^{-5}, the results, including the detection sensitivity and voxel parameters, were stable when the number of epochs was set between 500 and 800. For our final model, the learning rate 6×10^{-5} with 500 epochs was chosen to balance learning speed and network stability. Encoder depth size from 3 to 6 was studied. We observed that the deeper of the network, the harder it was to capture the features from small kernels. Therefore, we chose an encoder depth size of 4 and a batch size of 8 to balance among the size of the image data, GPU memory, and the training speed.

For comparison, vanilla UNet with different kernels were trained with the same batch size, learning rate, and epochs as those used in the asym-UNet (Çiçek *et al* 2016). In addition to the k1 and k2 used in the asym-UNet training, k3 ($2 \times 2 \times 2$) and k4 ($3 \times 3 \times 1$) were also tested.

2.4. Model performance assessment

To evaluate the performance of the asym-UNet model, we defined the following quantitative measures. The overall performance of the segmentation was given by the voxel-based DSC_V, defined in equation (3),

$$DSC_V = \frac{2 \times TP_V}{2 \times TP_V + FP_V + FN_V} \tag{3}$$

where TP_V was the number of overlapping voxels, FN_V was the number of undetected voxels manual contoured, and FP_V was the number of automatically segmented voxels outside the manual contours. DSC_V of 1 indicated FP_V and FN_V are both zero.

Voxel-based sensitivity (S_V) and precision (P_V) were the similarity measure often used in the medical image processing to evaluate the performance of the segmentation algorithm that had predefined ground truth (Fenster and Chiu 2006). S_V and P_V were defined in equations (4) and (5), respectively.

$$S_V = \frac{TP_V}{TP_V + FN_V} \tag{4}$$

$$P_V = \frac{TP_V}{TP_V + FP_V} \tag{5}$$

Individual BM-based sensitivity (S_{BM}) and false detection rate (F_{BM}) were used to evaluate the detection capability, and they were defined as:

$$S_{BM} = \frac{TT_{BM}}{TT_{BM} + TN_{BM}} \tag{6}$$

$$F_{BM} = \frac{FD_{BM}}{TT_{BM} + TN_{BM}} \tag{7}$$

where TT_{BM} was the number of metastases correctly detected (CD), TN_{BM} was the number of undetected (UD) metastases, and FD_{BM} was the number of metastases falsely detected. In the study, undetected metastasis is defined as the metastasis that was missed by the algorithm but was deemed by the experts as true metastasis; falsely detected metastasis was the automated solitary contour that was not at the location confirmed by experts and had zero overlaps with the manual expert contour.

3. Results

Figure 4(a) shows the distribution of BM binned by the diameters in the test dataset. Figure 4(b) shows the receiver operating characteristic (ROC) curves for the six trained models. The asym-UNet classifier clearly shows the best diagnostic ability among all six models. The area under the curve values are 0.86, 0.76, 0.75, 0.67, 0.65, 0.62 for asym-UNet, sym-UNet, UNet with k1, UNet with k2, UNet with k3, and UNet with k4, respectively.

Table 1 shows the detailed model validation results. For the vanilla UNet model trained using k1 alone, the numbers of CD and UD BMs are 46 (64%), 26 (36%), respectively, for the group S that contained smaller (1–10 mm) lesions. The number of falsely detected tumors are 17. As a comparison, the UNet model trained using k2 or k3 alone improved the CD to 64 (89%) and 50 (69%) and reduced the number of UD to 8 (11%), 22 (31%) for the S group. However, the UNet model trained using k2 or k3 alone has a large number of false detected tumors of 50, 40 for the S group, respectively. For group L consisting of larger (11–26 mm) BMs, the UNet models trained with either k1 alone, k2 alone, or k3 alone detected all 17 (100%) BMs correctly, and yet the UNet with k2 alone falsely detected five lesions, with k3 alone detected two. A similar observation can be made for DSC that under a single kernel UNet setting, a smaller kernel performed better in detection sensitivity but tend to report higher false positivity while the larger kernels are more specific but less sensitive. Table 1 also shows that the performance could not be substantially improved using a single differently sized kernel. In comparison, our asym-UNet model trained with combined k1 and k2 kernels improved the balance of CD/UD/FD BMs, achieving the highest DSC_V of 0.65 for the S group among all models tested. For the L group, asym-UNet was able to maintain the $DSC_V = 0.84$ and high sensitivity (≥0.9). For the S group, the asym-UNet balanced between high sensitivity while keeping the false detection ratio reasonably low ($F_{BM} = 0.24$).

Figure 5 shows examples from two typical patients comparing ground truth, the prediction from vanilla UNet, and our new asym-UNet. Patient 1 is a 45 year old female with metastatic colorectal adenocarcinoma. There are multiple heterogeneously enhancing lesions noted on the axial T1 + C MRI scan image, including a right paramedian occipital lobe metastasis measuring 2.0 × 1.7 × 1.8 cm and lateral right occipital lobe metastasis measuring 1.6 × 1.3 × 1.3 cm. A 3 mm rim-enhancing lesion is also seen along the inferior mid-right temporal lobe. Patient 2 is a 52 year old female with metastatic breast cancer. Multiple BM were detected, including an 8 mm focus in the left frontal lobe, 1 mm focus in the left periventricular frontal white matter, and 3 mm focus in the right parietal lobe. The first column of the image shows the axial slice of the T1 contrast-enhanced MR image, and the second column shows the ground truth contours from treating neurosurgeon and radiation oncologist. The third column shows the prediction result from the vanilla single-kernel UNet (k1 = 3 × 3 × 3). The fourth one is the result of our asym-UNet architecture. It is shown that the 3D asym-UNet can detect the metastases closer to the scalp boundary and with smaller sizes that have been missed in vanilla single UNet architecture. Neither vanilla UNet nor asym-UNet was able to detect lesions of 1 mm or less on a single slice contour.

4. Discussion

In this study, we adapted the 3D UNet deep learning classification method for the detection of multiple BMs on standard T1-weighted post-contrast axial MRI used for SRS planning purposes. Compared with the vanilla UNet using a single kernel, our asym-UNet used two different sized kernels to capture the image features of small tumors as well as the boundary information of larger metastases. Compared with vanilla UNet with only a single 3 × 3 × 3 kernel, the proposed asym-UNet achieved improved performance for both the voxel DSC and the detection sensitivity. The improvement is particularly remarkable for small BMs less than 4 mm in diameter, where the detection sensitivity was improved by 25%. For BMs greater than 6 mm in diameter, where the detection sensitivity was 100% only with a 1.1 false positive ratio (FP) per patient.

Due to the demonstrated advantages of using deep learning for segmentation for BMs and other tasks, the discussion focuses on the comparison with deep learning-based BM segmentation studies. Table 2 lists a few published BM detection papers using deep learning and comparison to our results. Without size stratification, our BM cohort is similar to Charron's (Charron *et al* 2018). Charron *et al* tested 18 patients with 38 lesions. They reported a sensitivity of 0.93 and 7.8 FP per patient without excluding tumors smaller than 4 mm. Our test dataset has 15 patients with 89 lesions, which have more metastases per patient than Charron's. With a comparable detection sensitivity of 0.90, our FP rate is only 1.1 per patient, more than sevenfold lower than Charron *et al*. The DSC from Charron and our group are 0.79 and 0.80, respectively. With size stratification, our S (1–10 mm) group is equivalent to 0.5–500 mm^3 and is smaller than Hu's small metastases group, however our detection sensitivity is still 20% higher than Hu's small group (Hu *et al* 2019). Our small group result comparison to Hu's small group result showed improved voxel-based DSC as well. Our L (11–26 mm) group is equivalent to 696–9203 mm^3. Our L group improved sensitivity to 1 as compared to 0.98 reported in Hu's large group result. Grovik reported a 50% sensitivity to detect BM less than 7 mm (Grøvik *et al* 2020). Although the sensitivity for small group

was not directly reported, we estimated it based on the figure from the published manuscript as 73% for S (1–10 mm), which is lower than 81% achieved using our asym-UNet method. Note that for the DSC calculation, our results and Hu's results are the averages of all BMs, while other studies averaged all patients, which are insensitive to the detection error of small BMs for patients with multiple metastases.

Pairs of kernels were studied in this project. First, we observed that smaller kernels performed better in detection sensitivity but worse in false positivity. The first thought was that vanilla UNet with a larger kernel ($3 \times 3 \times 3$) dominated the learning processing and then added more features which can be learned by smaller kernels, such as $1 \times 1 \times 3$ or $1 \times 1 \times 1$. Therefore, combinations of a larger kernel and a smaller kernel were investigated. Unfortunately, the performances were very similar with the vanilla UNet when the learning processing was dominated by a larger kernel. However, the performances can keep higher voxel DSC and the detection sensitivity, but lower false detection when the proportion of kernel pair was 15k1:1k2, where k1 was $1 \times 1 \times 3$ and k2 was $3 \times 3 \times 3$, The optimal combination ratio is problem-dependent, and we speculate that when the overall targets of detection are relatively large (e.g. the scale of 6 mm of diameter) a combination in the range of 1 to 1 would work reasonably well, and when the major targets of detection are small, then the range of 10 to 1 is suggested.

5. Conclusion

A novel 3D asym-UNet model was successfully constructed for automated detection and segmentation of brain metastases. The model compared favorably in detection sensitivity, false positivity, and dice similarity index with previous studies using deep learning. The improvement is particularly significant for small BM with a diameter <10 mm.

References

Abd-Ellah MK, Awad AI, Khalaf AAM and Hamed HFA 2019 A review on brain tumor diagnosis from MRI images: practical implications, key achievements, and lessons learned Magn. Reson. Imaging 61 300–18 [PubMed: 31173851]

Brown PD et al. 2016 Effect of radiosurgery alone vs radiosurgery with whole brain radiation therapy on cognitive function in patients with 1 to 3 brain metastases: a randomized clinical trial JAMA 316 401–9 [PubMed: 27458945]

Chang EL, Brown P, Lo SS, Sahgal A and Suh J 2018 Adult CNS Radiation Oncology: Principles and Practice (Berlin: Springer)

Chang EL et al. 2009 Neurocognition in patients with brain metastases treated with radiosurgery or radiosurgery plus whole-brain irradiation: a randomised controlled trial Lancet Oncol. 10 1037–44 [PubMed: 19801201]

Charron O, Lallement A, Jarnet D, Noblet V, Clavier J-B and Meyer P 2018 Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network Comput. Biol. Med 95 43–54 [PubMed: 29455079]

Chen H, Dou Q, Wang X, Qin J, Cheng JC and Heng P-A 2016 3D fully convolutional networks for intervertebral disc localization and segmentation Int. Conf. on Medical Imaging and Augmented Reality (Berlin: Springer) pp 375–82

Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T and Ronneberger O 2016 3D U-Net: learning dense volumetric segmentation from sparse annotation Int. Conference on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 424–32

Doi K 2007 Computer-aided diagnosis in medical imaging: historical review, current status and future potential Comput. Med. Imaging Graph 31 198–211 [PubMed: 17349778]

Fenster A and Chiu B 2006 Evaluation of segmentation algorithms for medical imaging 2005 IEEE Engineering in Medicine and Biology 27th Annual Conf. (Piscataway, NJ: IEEE) pp 7186–9

Glorot X and Bengio Y 2010 Understanding the difficulty of training deep feedforward neural networks Proc. of the thirteenth international conference on artificial intelligence and statistics pp 249–56

Gou S, Tong N, Qi SX, Yang S, Chin RK and Sheng K 2020 Self-channel-and-spatial-attention neural network for automated multi-organ segmentation on head and neck CT images Phys. Med. Biol 65 245034 [PubMed: 32097892]

Grøvik E, Yi D, Iv M, Tong E, Rubin D and Zaharchuk G 2020 Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI J. Magn. Reson. Imaging JMRI 51 175–82 [PubMed: 31050074]

Hu S-Y, Weng W-H, Lu S-L, Cheng Y-H, Xiao F, Hsu F-M and Lu J-T 2019 Multimodal Volume-Aware Detection and Segmentation for Brain Metastases Radiosurgery Workshop on Artificial Intelligence in Radiation Therapy (Berlin: Springer) pp 61–69

Kingma DP and Ba J 2014 Adam: a method for stochastic optimization Int. Conf. on Learning Representation (arXiv: 1412.6980)

Lecun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W and Jackel LD 1989 Backpropagation applied to handwritten zip code recognition Neural Comput. 1 541–51

Liu Y et al. 2017 A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery PloS One 12 e0185844 [PubMed: 28985229]

Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE) pp 3431–40

Losch M, Sullivan J and Lansner A 2015 A detection and segmentation of brain metastases with deep convolutional networks M.S. thesis, KTH, Comput. Vision and Active Perception, CVAP

Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Berlin: Springer) pp 234–41

Sahgal A, Aoyama H, Kocher M, Neupane B, Collette S, Tago M, Shaw P, Beyene J and Chang EL 2015 Phase 3 trials of stereotactic radiosurgery with or without whole-brain radiation therapy for 1–4 brain metastases: individual patient data meta-analysis Int. J. Radiat. Oncol 91 710–7

Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting J. Mach. Learn. Res 15 1929–58

Suh JH, Kotecha R, Chao ST, Ahluwalia MS, Sahgal A and Chang EL 2020 Current approaches to the management of brain metastases Nat. Rev. Clin. Oncol 17 279–99 [PubMed: 32080373]

Tang H, Chen X, Liu Y, Lu Z, You J, Yang M, Yao S, Zhao G, Xu Y and Chen T 2019 Clinically applicable deep learning framework for organs at risk delineation in CT images Nat. Mach. Intell 1 480–91

Tong N, Gou S, Yang S, Cao M and Sheng K 2019 Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images Med. Phys 46 2669–82 [PubMed: 31002188]

Tong N, Gou S, Yang S, Ruan D and Sheng K 2018 Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks Med. Phys 45 4558–67 [PubMed: 30136285]

Walker AE, Robins M and Weinfeld FD 1985 Epidemiology of brain tumors: the national survey of intracranial neoplasms Neurology 35 219–26 [PubMed: 3969210]

Yamamoto M et al. 2017 A multi-institutional prospective observational study of stereotactic radiosurgery for patients with multiple brain metastases (JLGK0901 Study Update): irradiation-related complications and long-term maintenance of mini-mental state examination scores Int. J. Radiat. Oncol 99 31–40

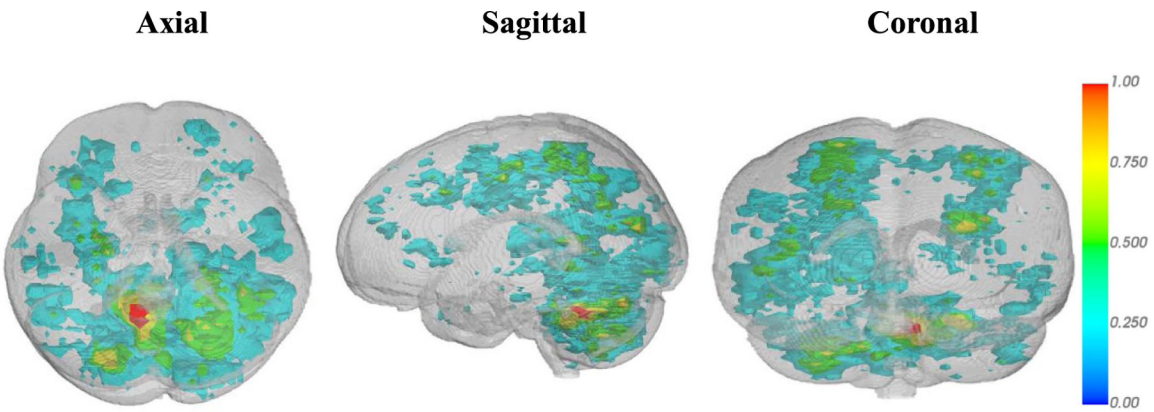

Figure 1.

Distribution of brain metastases in a common brain mask. Red indicates a greater and blue indicates a smaller number of BM distribution in the area.

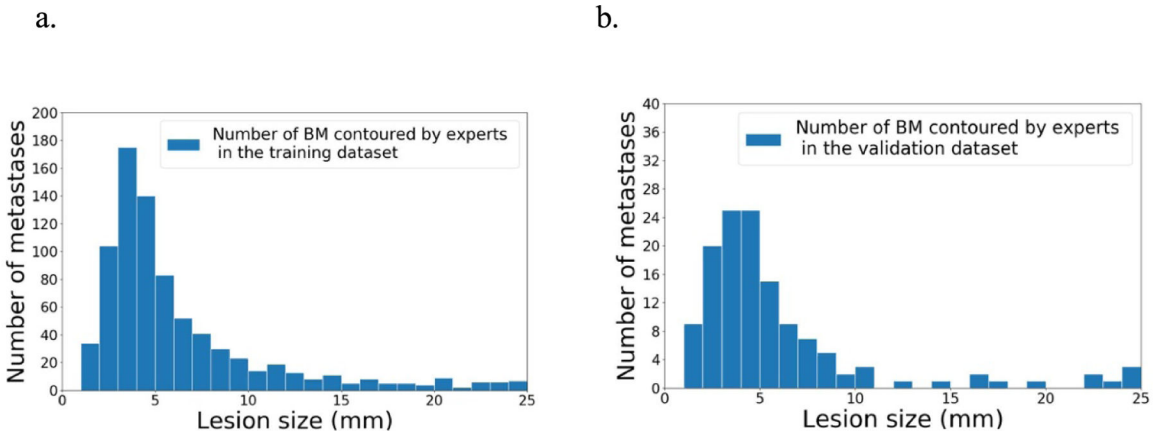

Figure 2.
Histogram of BM diameters in the training dataset (a) validation dataset (b).

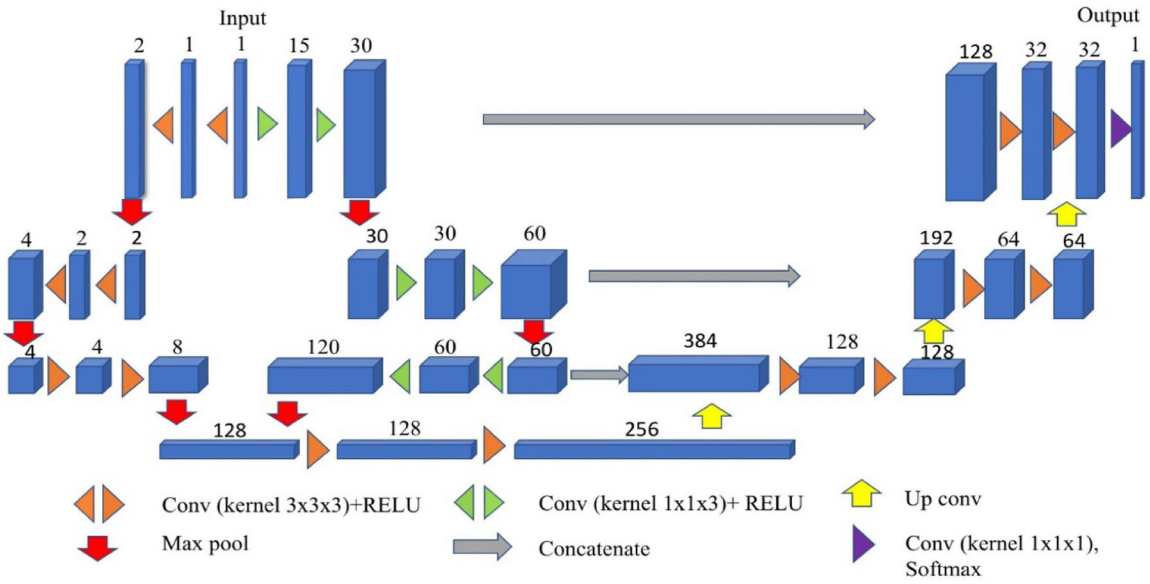

Figure 3.

Asym-UNet architecture. The diagram is composed of two down-sampling arms and one up-sampling arm. In the left and right of down sampling paths, the convolutional kernels are 3 × 3 × 3, 1 × 1 × 3, respectively. Each blue cuboid corresponds to feature maps. The number of channels is denoted on the top of cuboid.

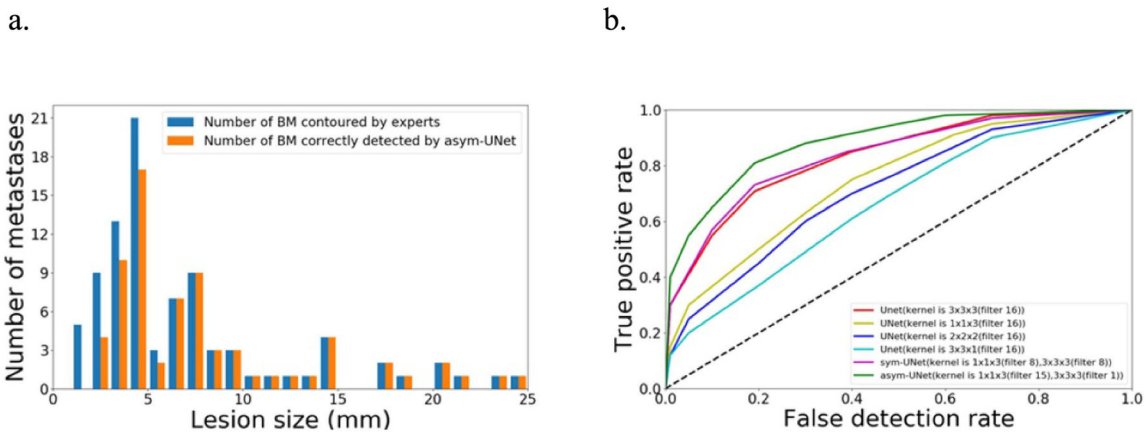

Figure 4.

(a) Histogram of BM diameters from expert manual contours (blue) and asym-UNet automatic contours (orange) in the test dataset. (b) ROC curves from the six kernel setups.

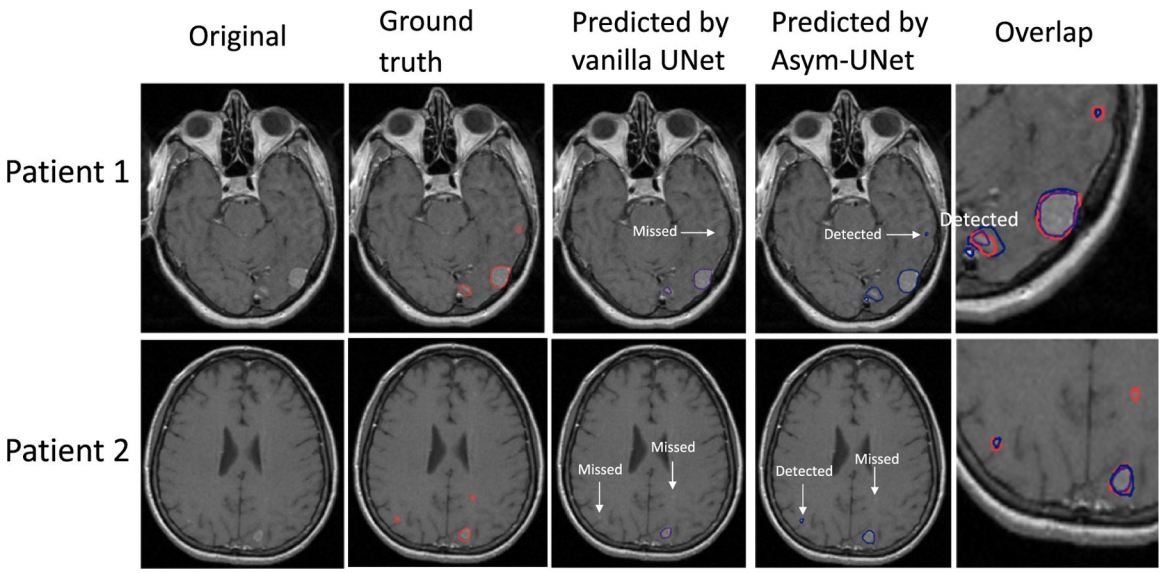

Figure 5.
Examples from two typical patients comparing ground truth, the prediction from vanilla UNet, and our asym-UNet.

Table 1.

Summary of detection and segmentation performance comparing vanilla UNet and asym-UNet.

	Diameter	# of BM	CD/UD/FD[b]	DSC_V	S_V	P_V	S_{BM}	F_{BM}
UNet	S (1–10 mm)	72	46/26/17	0.58	0.61	0.81	0.64	0.24
16k1[a]	L (11–26 mm)	17	17/0/0	0.84	0.79	0.91	1	0
UNet	S (1–10 mm)	72	64/8/50	0.43	0.80	0.20	0.89	0.69
16k2[a]	L (11–26 mm)	17	17/0/5	0.52	0.95	0.43	1	0.29
UNet	S (1–10 mm)	72	50/22/40	0.20	0.60	0.30	0.69	0.55
16k3[a]	L (11–26 mm)	17	17/0/2	0.41	0.75	0.45	1	0.12
UNet	S (1–10 mm)	72	40/32/36	0.55	0.61	0.78	0.55	0.50
16k4[a]	L (11–26 mm)	17	17/0/7	0.81	080	0.89	1	0.41
sym-UNet	S (1–10 mm)	72	48/24/18	0.60	0.65	0.80	0.67	0.25
8k1:8k2	L (11–26 mm)	17	17/0/1	0.82	0.80	0.89	1	0.06
asym-UNet	S (1–10 mm)	72	58/14/17	0.65	0.76	0.72	0.81	0.24
1k1:15k2	L (11–26 mm)	17	17/0/0	0.84	0.94	0.82	1	0

[a] k1, k2, k3 and k4 stand for kernel 3 × 3 × 3, kernel 1 × 1 × 3, kernel 2 × 2 × 2 and kernel 3 × 3 × 1, respectively.

[b] CD = number of BM correctly detected; UD = number of undetected BM; FD = number of BM falsely detected.

Table 2.

Comparing our asym-UNet model to those published in the literature for BM.

	# Training/Testing	BM size	Sensitivity	DSC
Liu *et al* (2017)	440/50	NA	NA	0.67[c]
Losch *et al* (2015)	490/265(glioma)	NA	0.82[a]	0.66[c]
Charron *et al* (2018)	164/18	Diameter (1.6–27 mm)	0.93[a]	0.77[c]
Grøvik *et al* (2020)	105/51	NA	0.83[a]	0.79[c]
Hu *et al* (2019)	305/36	Volume (<1500 mm^3)	0.61[b]	0.47[d]
Hu *et al* (2019)	305/36	Volume (>1500 mm^3)	0.98[b]	0.82[d]
Our study (asym-UNet)	180/15	Diameter (1–10 mm)	0.81[b]	0.65[d]
Our study (asym-UNet)	180/15	Diameter (11–26 mm)	1.00[b]	0.84[d]

[a]The BM detection sensitivity is calculated per patient. The average sensitivity from all patients is then reported.

[b]The detection sensitivity is calculated from all BM in the testing dataset regardless of which patient it came from.

[c]The DSC is calculated per patient. The average DSC from all patients is then reported.

[d]The DSC is calculated per BM. The average DSC from all BM in the testing dataset is then reported.