Comparison of Deep Learning Algorithms with Single- and Multi-Channel Input for Brain Metastases Detection

Jang-Hoon Oh, Hyug-Gi Kim, Kyung Mi Lee

Department of Radiology, Kyung Hee University Hospital, Kyung Hee University College of Medicine, Seoul, Republic of Korea

roineri5@gmail.com, khyukgi@gmail.com, bandilee@khu.ac.kr

Abstract

To investigate the effect of multi-channel input data in deep learning networks, deep learning models using single- and multi-channel input data were trained for brain metastases detection, and their performance was evaluated and compared. The YOLO v2 networks based on the pre-trained ResNet50 were trained using a single (SPACE GD)- and multi (MPRAGE, SPACE GD, MPRAGE GD)-channel input data., the overall performances of the deep learning model with multi-channel input data showed lower false positive average but lower sensitivity compares to the single-channel model. In this study, a deep learning model using multi-channel input data does not guarantee improved performance, however, it may improve performance in certain cases.

I. Introduction

Data collection in the medical image domain is difficult compared with the general image domain because of Patient Privacy Policy[1], such as the Health Insurance Portability and Accountability Act (HIPAA)[2], hence, deep learning studies in medical image domain are often performed using a limited amount of data. Special medical imaging techniques can obtain the images in same dimension with various pattern of intensities, such as magnetic resonance images (MRIs) with different sequences [3], and by using images with multi sequence images as multi-channel input data, the deep learning model can be trained with more data and patterns.

Brain metastases (BMs), which commonly originate from lung cancer, breast cancer, or malignant melanoma, are the most common intracranial tumors, and contrast-enhanced T1-weighted imaging (CE T1WI) magnetic resonance (MR) sequence and black-blood (BB) imaging are key in the diagnosis of BM[4]. Recently, deep learning-based algorithms that automatically detect or segment BM lesions have been proposed [5].

In this study, to compare the performance with different amounts of input data, we developed two deep learning models with single- and multi-channel input data for BMs and investigated their performance.

II. Materials and Methods

MR data were acquired using a 3T MR system (3T MAGNETOM VIDA; Siemens, Erlangen, Germany). The MRI protocol for BM included pre-contrast-enhanced 3D-MPRAGE T1 weighted image (T1WI), post-contrast-enhanced 3D BB image (sampling perfection with application-optimized contrasts using different flip angle evolutions, SPACE GD), and post-contrast-enhanced 3D MPRAGE T1WI (MPRAGE GD). After acquire the 3D MPRAGE, MPRAGE GD, and BB images in sagittal plane, image reconstruction in axial plane was performed with the slice thickness of 3 mm.

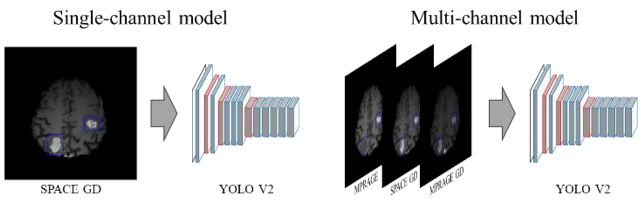
This retrospective study collected anonymized MR data from Kyung Hee University Hospital, and a total of 113 individuals were acquired from May 2019 to February 2021, and all individuals was randomly separated into five subgroups for five-fold cross-validation. The MR dataset consisted of 6,196 slices with 585 metastases in 1,055 images for each sequence.

The labeling process was performed by a board-certified radiologist with drawing a rectangular boundary box with the single label "brain metastases" to include all BM lesions in SPACE GD image on all slices where the BM was located. These boundary boxes were replicated into MPRAGE and MPRAGE GD to enable multi-channel data. The labeling process was conducted based on 2D slice, and the object detection network was trained BM labels corresponding to each slice. However, radiologists generally determine a diagnosis based on overall BM mass, rather than individual slices. Accordingly, the BM was evaluated based on total mass rather than each slice. To do this, a process of re-labeling BM labels when overlapping 0.3 intersection over union (IoU) or more in adjacent slices as one mass was performed, and an independent order was assigned to each BM mass. According to response assessment criteria for brain metastases[6], any BMs of less than 5 mm were excluded.

Two YOLO v2 [7] networks with pre-trained ResNet-50 [8] as a backbone were trained using single(SPACE GD)- and multi(MPRAGE, SPACE GD, and MPRAGE GD)-channel input data (Figure 1.) with the following parameters: number of anchor boxes of 7, optimizer using 'Adam', factor for L2 regularization of 0.0001, the initial learning rate of 0.001, mini-batch size of 64, and maximum epochs of 1,000. For data augmentation, random image rotation (0°, 90°, 180°, and 270°) and random horizontal flipping were applied while training both networks. Skull stripping using a brain extraction tool (BET, v1.3) and image intensity normalization were performed for all training data as pre-processing.

All training and pre/post-processing were conducted using MATLAB (MathWorks, R2020b, Natick, MA, USA) on a single-server workstation with a Windows operating system (Window Server 2016) with double NVIDIA V100 graphic cards and 32-GB memory. To compare the performance of deep learning networks for single- and multi-channel inputs, performance was evaluated in terms of sensitivity, precision, F1-Score, and false positive average (FP_{avg}). Furthermore, each lesion was evaluated as a true positive if even only part of the lesion was predicted as brain metastases by the deep learning model.

Figure 1. Simple diagram of single- and multi-channel input model



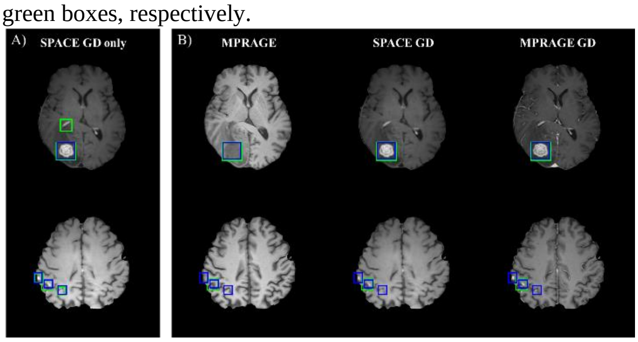
III. Results

Table 1 show the performance of the two deep learning models for single- and multi-channel input data for five-fold cross validation. The overall performances of the deep learning models with single-channel input data was 88.03%, 23.24%, 36.77%, 15.05 for sensitivity, precision, F1-Score, FP_{avg}, respectively. The overall performances of the deep learning models with multi-channel input data was 82.27%, 34.87%, 48.98%, 7.97 for sensitivity, precision, F1-Score, FP_{avg}, respectively. Figure 2 shows examples of two deep learning models. The multi-channel input model did not detect the false positive (upper) which is detected by the single-channel input model, but some BMs are not detected (lower).

Table 1. Summary of the performances of deep learning models with single- and multi-channel input data.

	Sensitivity	Precision	F1-Score	FP_{avg}
Single-channel				
Overall	88.03%	23.24%	36.77%	15.05
Cross validations				
Dataset1	87.12%	23.39%	36.88%	21.14
Dataset2	86.55%	23.02%	36.36%	21.52
Dataset3	80.88%	19.57%	31.52%	10.27
Dataset4	89.52%	29.84%	44.76%	9.61
Dataset5	97.44%	20.54%	33.93%	12.78
Multi-channel				
Overall	82.27%	34.87%	48.98%	7.97
Cross validations				
Dataset1	79.25%	38.18%	51.53%	9.27
Dataset2	85.12%	36.02%	50.62%	11.55
Dataset3	73.91%	20.00%	31.48%	8.87
Dataset4	82.86%	46.77%	59.79%	4.50
Dataset5	88.75%	34.98%	50.18%	5.74

Figure 2. Examples of two deep learning models for BM detection. the ground truth and predicted box were presented with blue and green boxes, respectively.



IV. Discussion and Conclusion

In this study, deep learning-based object detection networks with single- and multi-channel input data were applied to detect BM, and their performance was compared. Deep learning models were trained from data, and researcher or developer usually expected that a more amount of data would improve the performance of deep learning network, which is a common idea in deep learning [9]. From the result, the performance of the multi-channel input data model improved by 12.21% of the F1-score, which integrates the sensitivity and precision. However, the sensitivity of the multi-channel model decreased by 5.76%, and the decrease in sensitivity is not desirable if it is applied as computer-aided diagnostic algorithm for radiologists. From a statistical perspective, the independent multi-channel input data that has been tripled increases the complexity of the deep learning algorithm, and consequently, it does not improve the sensitivity of the deep learning model.

From the previous study, Al-masni et al. [10] compared three-channel input imaging types: SWI, phase, and magnitude images. The two-channel model using pre-processed SWI and phase showed the best performance, whereas the three-channel model using SWI, phase, and magnitude showed the lowest sensitivity. Fei et al. [11] proposed a multimodal computing model for MRI synthesis based on deep learning with a feature-disentanglement strategy. They compared the performance of FLAIR synthesis between the T1, T1+T2, and T1+T2+T1c models and reported that although the triple-input model (T1+T2+T1c) produced the highest PSNR value, there were no significant differences between the three models.

From the results of this study and the above experiments, we did not observe a consistent improvement in performance as the number of input channels increased. Although multichannel input does not guarantee improved performance, however, it may improve performance in specific cases.

ACKNOWLEDGMENT

This study is a part of a doctoral thesis of an author (JHO) [12]. This study was performed with received grants from the National Research Foundation of Korea (NRF) grant funded by the Government of South Korea (Ministry of Science and ICT, MSIT) (NRF-2020R1C1C1006623, NRF-2021R1F1A1050515) and Bracco Company. This research was the result of a study on the "HPC Support" Project, supported by the 'Ministry of Science and ICT' and NIPA.

References

[1] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, "How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?," 2015.
[2] R. Nosowsky and T. J. Giordano, "The Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule: Implications for clinical research," *Annu. Rev. Med.*, vol. 57, pp. 575–590, 2006.
[3] Y. Kurata et al., "Automatic segmentation of uterine endometrial cancer on multi-sequence MRI using a convolutional neural network," Sci. Rep., vol. 11, no. 1, pp. 1–10, 2021.
[4] S. Yang, Y. Nam, M.-O. Kim, E. Y. Kim, J. Park, and D.-H. Kim, "Computer-Aided Detection of Metastatic Brain Tumors Using Magnetic Resonance Black-Blood Imaging," *Invest. Radiol.*, vol. 48, no. 2, pp. 113–119, Feb. 2013.
[5] O. Charron, A. Lallement, D. Jarnet, V. Noblet, J. B. Clavier, and P. Meyer, "Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network," *Comput. Biol. Med.*, vol. 95, no. November 2017, pp. 43–54, 2018.
[6] N. U. Lin *et al.*, "Response assessment criteria for brain metastases: Proposal from the RANO group," *Lancet Oncol.*, vol. 16, no. 6, pp. e270–e278, 2015.
[7] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *Cvpr2017*, no. April, pp. 187–213, Dec. 2016.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.
[9] A. A. Tavor *et al.*, "Do not have enough data? Deep learning to the rescue!," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 7383–7390, 2020.
[10] M. A. Al-masni, W. R. Kim, E. Y. Kim, Y. Noh, and D. H. Kim, "Automated detection of cerebral microbleeds in MR images: A two-stage deep learning approach," *NeuroImage Clin.*, vol. 28, p. 102464, 2020.
[11] Y. Fei, B. Zhan, M. Hong, X. Wu, J. Zhou, and Y. Wang, "Deep learning-based multi-modal computing with feature disentanglement for MRI image synthesis," *Med. Phys.*, vol. 48, no. 7, pp. 3778–3789, Jul. 2021.
[12] Jang-Hoon Oh, "Development and Feasibility Evaluation of Deep Learning-Based Medical Imaging Assistance Technology," Dissertation, *Kyung Hee Univ.*, 2022.