Review Article | Technology, Experiment, and Physics

eISSN 2005-8330
https://doi.org/10.3348/kjr.2022.0765
Korean J Radiol 2023;24(7):698-714

Korean Journal of Radiology
KJR

Developing and Evaluating Deep Learning Algorithms for Object Detection: Key Points for Achieving Superior Model Performance

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

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

In recent years, artificial intelligence, especially object detection-based deep learning in computer vision, has made significant advancements, driven by the development of computing power and the widespread use of graphic processor units. Object detection-based deep learning techniques have been applied in various fields, including the medical imaging domain, where remarkable achievements have been reported in disease detection. However, the application of deep learning does not always guarantee satisfactory performance, and researchers have been employing trial-and-error to identify the factors contributing to performance degradation and enhance their models. Moreover, due to the black-box problem, the intermediate processes of a deep learning network cannot be comprehended by humans; as a result, identifying problems in a deep learning model that exhibits poor performance can be challenging. This article highlights potential issues that may cause performance degradation at each deep learning step in the medical imaging domain and discusses factors that must be considered to improve the performance of deep learning models. Researchers who wish to begin deep learning research can reduce the required amount of trial-and-error by understanding the issues discussed in this study.

Keywords: Deep learning; Object detection; Diseases with small sizes; Disease subclass; Image modality; Deep learning workflow; Data augmentation; Hyperparameter optimization

INTRODUCTION

Recently, deep learning technologies in computer vision have rapidly developed owing to the advances in and widespread use of graphic processor units optimized for parallel operation [1]. Object detection [2] is a deep-learning task that simultaneously identifies the location and label of a target object. Interesting results for object detection have been reported in various studies, such as face detection [3], recognition [4], pedestrian detection [5], and car detection [6]. Furthermore, object detection has been applied in the medical imaging domain, which has shown remarkable results in developing models to predict lesions, such as brain cancer [7], liver disease [8], and wrist, rib, and pediatric skull fractures [9–12] using various imaging modalities, such as radiography, computed tomography (CT), and magnetic resonance imaging (MRI).

While majority of deep learning studies in the medical image domain have demonstrated remarkable results, certain approaches have exhibited poor performance [13]. In a previous study [14], deep-learning-based false-positive reduction demonstrated lower performance than rule-based false-positive reduction. Performance degradation can be caused by various factors, such as insufficient data, unoptimized hyper-parameters, or the application of an incorrect evaluation strategy. However, it is often difficult to understand the cause of poor performance because humans cannot understand the intermediate process of the deep learning model owing to the black-box problem

Received: October 7, 2022 **Revised:** April 29, 2023
Accepted: May 16, 2023
*These authors contributed equally to this work.
Corresponding author: Kyung Mi Lee, MD, PhD, Department of Radiology, Kyung Hee University Hospital, Kyung Hee University College of Medicine, 23 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Korea.
• E-mail: bandilee@khu.ac.kr

[15]. Moreover, conducting comparative experiments on all variables that can affect deep learning performance has practical limitations.

The purpose of this article is to explain and demonstrate the key considerations for applying object detection to the medical imaging domain across each step of deep learning research. These considerations are typically performed in the following order: target disease selection, data collection, data labeling, deep learning network training, and performance evaluation (Fig. 1). We hope that this article can help junior researchers who aim to apply object detection in the medical image domain to understand the potential issues that may occur at each step and efficiently conduct their research by reducing trial-and-error.

Target Disease Selection

Suitability of Object Detection

Deep learning techniques in medical imaging analysis can be categorized into classification, object detection, and segmentation. Examples of previous studies that have applied these three deep learning methods in medical image analysis are summarized in Supplementary Table 1, and an example explaining the differences between the three methods are shown in Figure 2.

Object detection (Fig. 2A) detects trained objects in

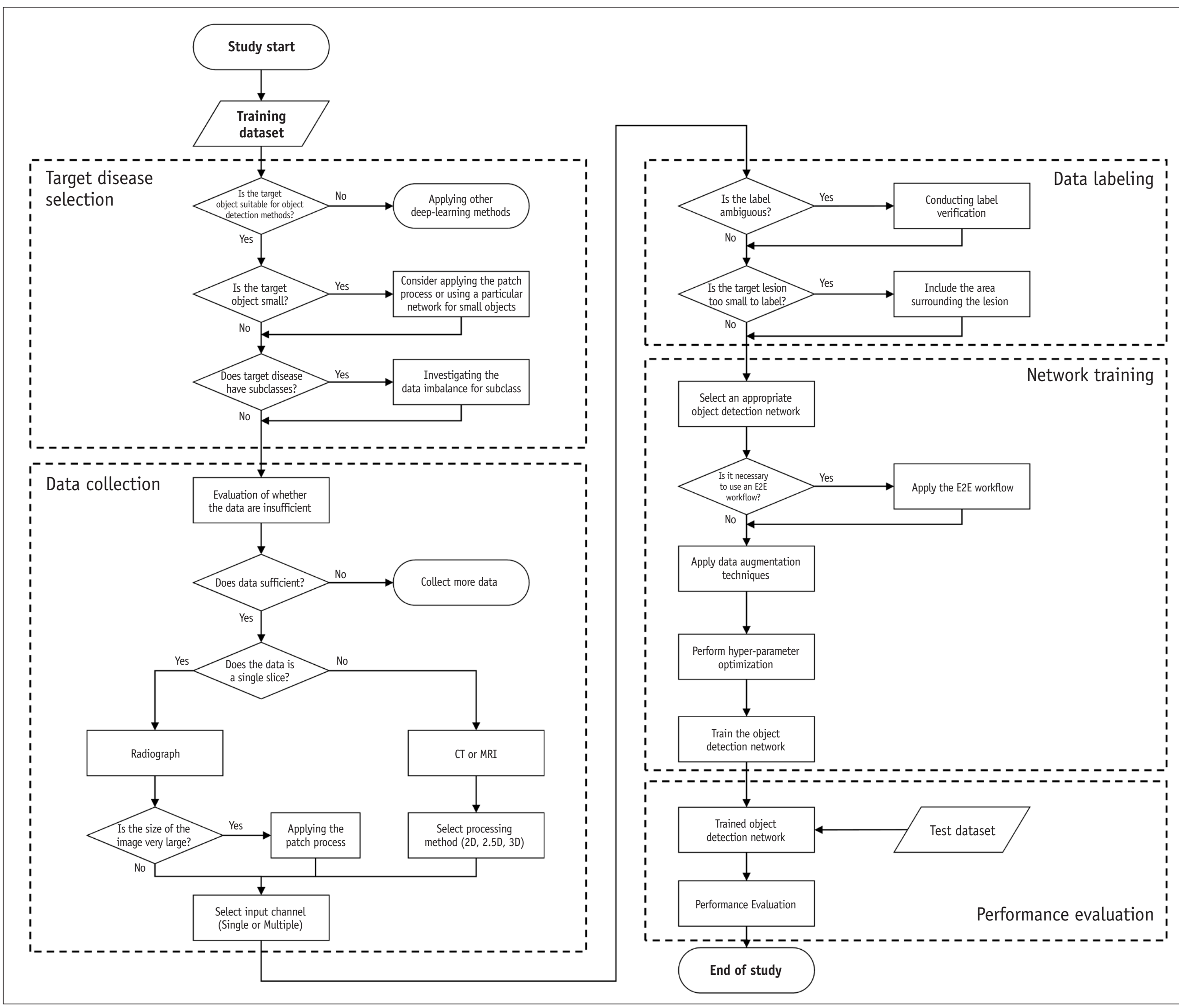

Fig. 1. Flowchart of the process for deep learning research by applying the issues introduced in this study. The flowchart organizes the issues that should be considered when conducting deep learning research in a sequential manner, and categorized into target disease selection, data collection, data labeling, network training, and performance evaluation. 2D = two dimensional, 2.5D = two and a half dimensional, 3D = three dimensional, CT = computed tomography, MRI = magnetic resonance imaging, E2E = end-to-end

 https://doi.org/10.3348/kjr.2022.0765

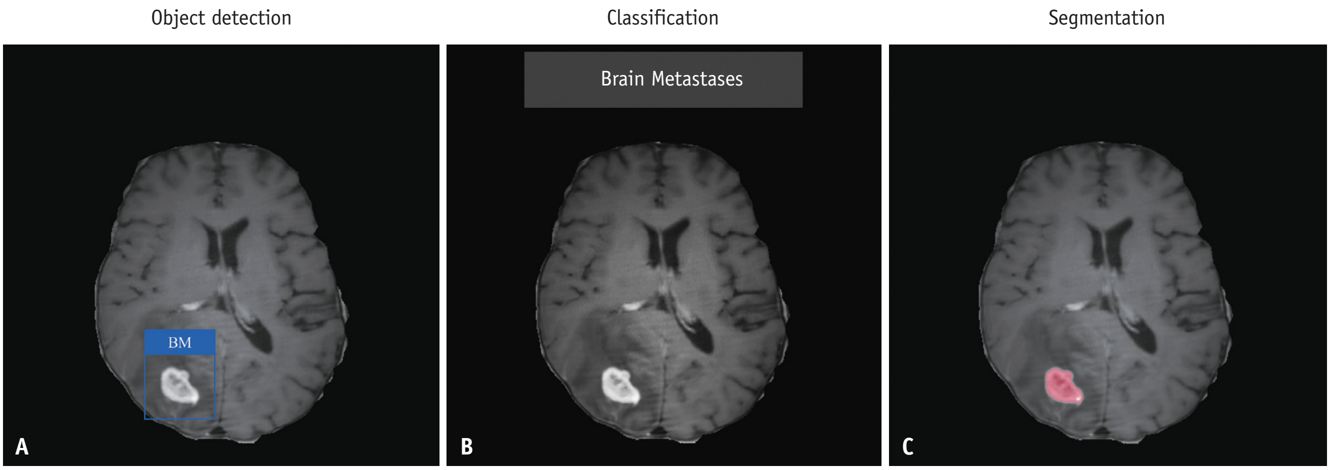

Fig. 2. An example showing the differences between object detection, classification, and segmentation. The object detection method **(A)** shows the result with a blue bounding box with the label "BM", the classification method **(B)** presents the result with the label "Brain Metastases", and the segmentation method **(C)** displays the result on the image using a red mask. BM = brain metastases.

an image using a bounding box or circle with its class. In contrast, the classification method (Fig. 2B) determines the class of the entire input image. Object detection offers the advantage of being able to identify multiple lesions in an image and does not suffer from the multilabel classification problem [16]. For example, previous studies have utilized object detection methods to detect lesions such as brain metastases [17], liver lesions [18], maxillary sinusitis [19], and cerebral microbleeds [14,20]. Other studies have applied it to identify the location of an object in an image, such as facial region [19], wrist region [9], and various organs, to extract a patch image. As shown in Figure 2C, the segmentation method yields results at the pixel level and is suitable for detecting lesions that require pixel-level evaluation, such as measuring the volume of the lesion [21,22] or supporting a radiotherapy plan. However, the cost of labeling is higher than that of the object detection method. Therefore, unless an evaluation in units of pixels is required, the object detection method is appropriate for detecting visible diseases or lesions in images, where the evaluation of the location information contained in images can help recognize each lesion separately.

Target Disease with a Small Size

Small-object detection is a fundamental challenge in computer vision [23]. A small object is defined as an object with a size less than or equal to 32 x 32 pixels [24], which results in issues such as indistinguishable features, low resolution, and limited context information, making it difficult to detect the target object [25]. Therefore, using an object detection algorithm for detecting diseases with small sizes, such as small calcifications and early-stage cancers, may result in poor performance. Examples of previous studies that demonstrated lower performance on small lesions than on large lesions are listed in Table 1 [17,22,26-28]. Zhou et al. [17] reported that four deep-learning networks exhibited lower sensitivity (10%–40%) in detecting brain metastases smaller than 3 mm. In another study on the detection of breast calcifications, Akselrod-Ballin et al. [26] reported that removing calcifications with radii smaller than 10 pixels can significantly improve performance. To address the challenges of detecting small objects, applying the patch process to increase the proportion of lesions in an image may improve performance [29]. Moreover, using specialized models for small objects, such as M2Det [30], multi-scale deconvolutional single-shot detector (SSD) [31], and improved faster region-based convolutional neural network (R-CNN) for small object detection [32], which have been recently published, may enhance the detection of small lesions. However, to the best of our knowledge, these models have not yet been applied in the medical imaging domain.

Identifying Data Distribution of Subset Groups

Certain target diseases in the medical imaging domain can be sub-classified. For example, maxillary sinusitis can be sub-classified into full opacification, air/fluid level, cysts, and mucosal thickening [19]; cancer labels can be grouped based on lesion size [17]; and metastases can be grouped according to their origin. If the object detection model is trained with integrated labels, the performance for each subclass might vary owing to the differences in the lesion features and amount of training data. In addition, the

Table 1. Examples of Previous Studies that Applied a Deep Learning Approach to Lesion Detection, Including Small-sized Lesions

Studies	Modality	Target Disease	Deep Learning Task	Algorithm	Lesion Size	Contents
Zhou et al. [17]	MRI	Brain metastases	Object detection	SSD SSD + ResNet50 SSD + focal loss RetinaNet	< 3 mm	Four deep learning networks showed lower sensitivity (10%–40%) for detecting brain metastases smaller than 3 mm while higher sensitivity (92%–98%) for larger than 6 mm.
Akselrod-Ballin et al. [26]	Breast mammography	Small calcifications	Object detection	Faster R-CNN	< 10 pixels	The performance of the proposed models for the INBreast dataset yielded a significant improvement by the removal of small calcification (radius < 10 pixels).
Kang et al. [22]	MRI	Meningioma	Segmentation	Attention and 2D U-net with GMM	< 1.5 cm^3	Two deep learning models that performed the best did not recognize any parts of the tumors in five cases with a size smaller than 1.5 cm^3.
Takao et al. [27]	CT	Brain metastases	Object detection	SSD	< 3 mm	The sensitivity of the CE + NECT model was 41.5% for < 3 mm lesions; 81.6% for 3–6 mm lesions; and 96.5% for ≥ 6 mm lesions.
Nam et al. [28]	Chest radiography	Lung cancer	-	Commercially available deep learning-based algorithm (Lunit INSIGHT CXR; Lunit)	≤ 2.0 cm	45% of nodules smaller than 2.0 cm were detected by the algorithm, while 63% of nodules larger than 2.0 cm were detected.

MRI = magnetic resonance imaging, SSD = single-shot detector, R-CNN = region-based convolutional neural network, GMM = Gaussian mixture model, CT = computed tomography, CE + NECT = contrast-enhanced + non-enhanced computed tomography

integrated performance of the test dataset may be altered if the composition of the subclass is different from that of the training dataset. If the composition of the subclass with lower accuracy is higher in the external validation set, the performance may be lower than that of the internal validation set, which can be mistakenly considered as the result of overfitting (Fig. 3). Therefore, the ratio of the subclasses should be verified during the composition of the test dataset, and the performance of each subclass must be investigated.

Data Collection

Evaluation of Diminished Performance Owing to Insufficient Data

Previous studies that applied the deep learning approach used various amounts of data (Supplementary Table 1). A deep learning network is trained with features from the data; therefore, training with more data can improve performance. However, data collection in the medical image domain is difficult compared with that in the general image domain, and studies that apply deep learning are often performed using a limited amount of data. Hence, the question "How much data is sufficient?" is commonly asked by researchers interested in artificial intelligence, and the amount of data required may vary depending on the target disease and imaging modality. For example, if the target disease has numerous variables, such as various sizes, locations in different regions, and varying lesion textures, specifying the amount of data required is difficult.

To address this issue, it is possible to evaluate whether the data are insufficient. If the target disease is determined and a certain amount of data is collected at the initial stage of the study, the relationship between the amount of data and performance of the deep learning model can be investigated by training and evaluating each model while increasing the amount of training data (Fig. 4). By estimating the amount of data required, a researcher can determine whether the amount of data for the study must be increased and how much data must be collected. For example, Cho et al. [33] investigated the relationship between accuracy and the amount of training data. Their estimation results predicted 98% accuracy for a training data size of 1000 per body class.

Single-Slice Images Such as Radiographs and Patch Process

A radiography image is relatively large compared with

 https://doi.org/10.3348/kjr.2022.0765

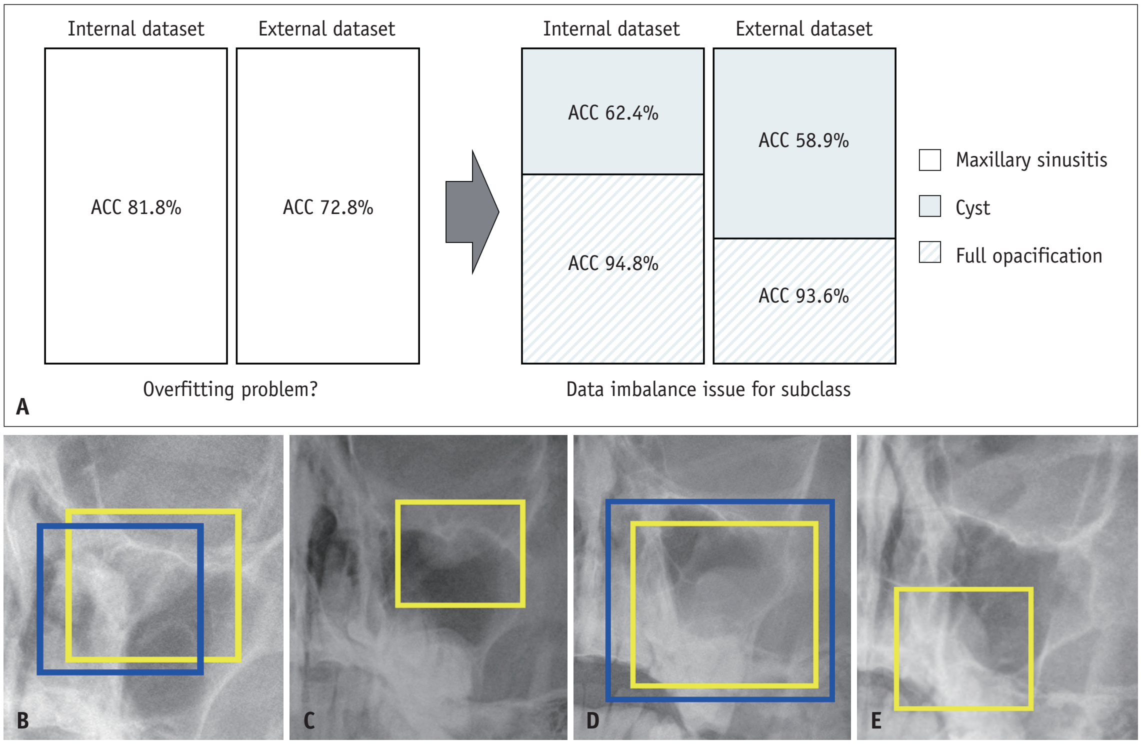

Fig. 3. Simple example of the data imbalance issue for a subclass. **A:** The scenario where subclass data imbalance can be misinterpreted as overfitting. **(B-E)** The predicted results of the object detection model for cyst cases, which are a subclass of sinusitis. The sample images for the internal **(B, C)** and external **(D, E)** datasets indicate ground truth and predicted results as blue and yellow bounding boxes, respectively. ACC = accuracy

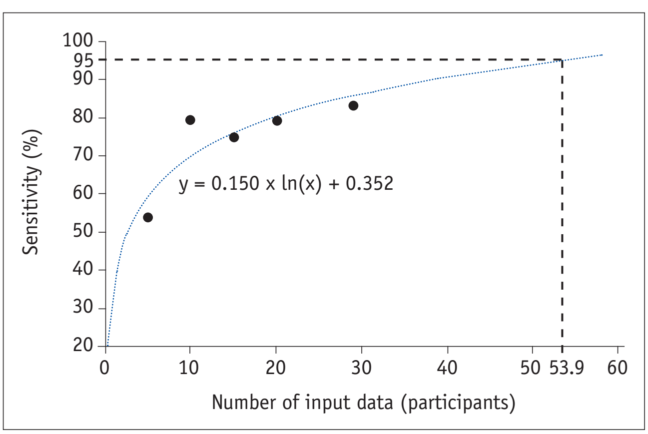

Fig. 4. Relationship between the amount of data and performance of the object detection deep learning network for brain metastases. The black dots represent the number of data used for model training and their corresponding sensitivity. The blue dashed line represents the trend line of the black dots, and the black dashed line represents the number of data required to achieve a 95% sensitivity. The results, shown as a logarithmic function, indicate that the sensitivity of the deep learning model increases as the amount of input data increases. In this example, to reach a sensitivity of 95%, the researcher can predict that they need the data of approximately 54 participants.

the average size of an image in ImageNet (approximately 1600–2000 pixels in the horizontal and vertical axes vs. approximately 400 x 350 pixels, respectively) [34]. A large image size significantly increases the computational power required and may also result in a dimensionality problem [35]. Additionally, radiographs usually contain a substantial portion of background that is unrelated to the diagnosis of the disease. Applying the patch process, which crops only the essential part of the image, can produce a cropped image without the loss of lesion information owing to shrinking, and unnecessary parts, such as the background, can be removed. However, the patch process has traditionally been performed by humans. In the medical imaging domain, certified radiologists or other medical doctors performed the manual patch process, which is time-consuming. As a result, the associated costs are considerably high.

Recently, an automated patch process was applied in the medical imaging domain by employing the object detection approach to address the disadvantages of large images, such as radiographs. Table 2 lists previous studies that used a manual patch process or applied deep learning approaches [9,19,36-42]. Previous studies [9,19,39-42] have applied an

Table 2. Examples of Deep Learning Approach-based Studies that Applied the Patch Process

Studies	Modality	Target Disease or Object	Patch Region	Patch Method	Contents
Kim et al. [36]	Water's view	Maxillary sinusitis	Facial region	Hand-craft	A handcrafted patch process including both maxillary sinuses was performed for 5020 Water's view images.
Kim et al. [37]	Water's view	Maxillary sinusitis	Maxillary sinus	Hand-craft	9540 Water's view images were cropped to 7 x 7 cm^2 based on the central coordinates of the bilateral maxillary sinuses.
Han et al. [38]	CT	Renal cancer	Renal cell carcinoma region	Hand-craft	Images were cropped based on the region of interest for the renal cell carcinoma lesion drawn by the radiologist.
Lee et al. [39]	Hand AP, PA	Hand	Hand mask	Object detection-based preprocessing engine	Images were cropped based on the hand mask patch generated by multiple-processing steps, including normalization, object detection on hand radiographs, and reconstruction.
Ebsim et al. [9]	Wrist radiographs (PA and LAT)	Wrist fracture	Wrist region	Machine learning	The patch process consisted of global search (RFRV) and local search (RFCLM).
Al-antari et al. [40]	Mammogram	Breast cancer	Mass region	Deep learning (object detection)	The predicted regions from the trained YOLO model for mass detection were cropped for the mass segmentation stage.
Jeon et al. [41]	Caldwell and Water's view	Sinusitis	Frontal, maxillary, and ethmoid sinusitis	Deep learning (object detection)	Input images had cropped based on bounding boxes which were predicted from a detector for localizing each sinus area, and cropped images were used for the second network that classifies each sinus patch with diagnostic labels.
Oh et al. [19]	Water's view	Maxillary sinusitis	Facial region	Deep learning (object detection)	The YOLO v2 detection network was applied to detect the facial region that includes both maxillary sinuses.
Oh et al. [42]	Mammogram	Mammography phantom image evaluation	Phantom region	Deep learning (object detection)	The YOLO v2 detection network was applied to detect the phantom region to extract a patch for the next step.

CT = computed tomography, AP = anteroposterior, PA = posteroanterior, LAT = lateral, RFRV = random forest regression voting, RFCLM = random forest regression voting constrained local model, YOLO = you only look once

 https://doi.org/10.3348/kjr.2022.0765

automated patch process based on deep learning or machine learning approaches, such as facial region detection [19], phantom region detection [42], radius location identification [9], and mass detection [40], as pre-processing to limit the area of the disease location, reducing the loss of lesion information caused by shrinking and removing the background that is unrelated to the disease.

Multi-Slice Images Such as CT or MRI

CT and MRI scans involve acquiring a multi-slice image instead of a single image and using 2D, 2.5D, or 3D methods depending on the type of image input required for the deep learning model. Compared with the 3D method, a 2D-based deep learning network can be used with a larger image as an input, for example, 256 x 256 or 512 x 512 images [14,17]. However, besides extremely small lesions, most target lesions are located over multiple image slices. The 2D deep learning model predicts the lesion in each image individually; therefore, to evaluate the lesion on a mass unit, post-processing that evaluates the lesion in an adjacent slice as a mass must be performed [18]. The 3D method has the advantage that it can train a deep learning model using more information from adjacent slices, and its performance can be improved. The disadvantage of the 3D method is that it requires a significant amount of computational power; therefore, small-sized input data are generally used [43]. The 2.5D method uses a 2D-based deep learning model in three orthogonal directions—coronal, sagittal, and axial—and can improve the performance of a deep learning algorithm through a majority decision step with three deep learning networks. However, the labeling process must be performed for all three directions, and post-processing for labeling must be performed to apply the label to all three directions. Moreover, a majority decision step must be performed during additional post-processing [44].

Single- and Multi-Channel Input Data

Special imaging techniques, such as dual-energy CT [45] and MRIs with multiple different sequences [46], have made it possible to acquire the same images with different intensities. Examples of previous studies that have investigated multi-channel input data for the deep-learning approach are listed in Table 3 [47-51]. Using images with different intensities as multi-channel input data, the deep learning model can be trained with more data and patterns, and its performance can be improved [49]. For example, if images before and after using a contrast agent are used as a multichannel input, the performance of the deep learning model can be improved by recognizing the intensity differences before and after using the contrast agent [52].

However, using multi-channel data for performance improvement does not guarantee a statistically significant difference [48]. In a scenario where a model is being developed for detecting lesions, if images that are not related to the diagnosis of the lesion are included in the multi-channel input data, the required computational power increases, which increases the dimension of the data. This, in turn, may lead to a decrease in the performance of the deep learning network [47]. Therefore, researchers who aim to utilize multi-channel data in developing deep learning models should exclude unnecessary data during the data collection process and evaluate the performance using data that can affect the performance.

Data Labeling

Labeling Verification for Ambiguous Objects

Object detection requires training with the correct labels, and its performance may decrease when the training data include noisy or incorrect labels. According to Rolnick et al. [53], the performance of a classification model using the ImageNet dataset with 5% incorrect labels decreases by approximately 20%.

In the medical imaging domain, the labeling process is typically performed by a radiologist, and most lesions are clearly labeled. However, several lesions may be ambiguous to diagnose or label, and using labels for ambiguous lesions in training and evaluation can lead to performance degradation. This degradation can be overcome by verifying ambiguous labels. Verification can be performed by comparing with other imaging modalities or biopsy results, and labels can be determined by aggregating the opinions of several raters, which reduces ambiguity. In a previous study, Kim et al. [37] used paranasal sinus CT scans as the reference standard for sinusitis to compare the overall diagnostic performance of a deep-learning algorithm with that of radiologists. In another study [54], subtype labels were confirmed by the pathological examination of surgically removed tumors to diagnose kidney cancer.

Labeling Small Objects

In object detection, the labeling process is typically performed by drawing a bounding box or circle around a target object. For small objects, such as cerebral microbleeds

Table 3. Examples of Previous Studies that Compared Multi-channel Input Data for the Deep Learning Approach

Study	Imaging	Purpose	Comparison Models	Performance	Comments
Al-masni et al. [47]	Brain MRI	Cerebral microbleeds detection	One-channel: SWI Two-channel: SWI + Phase Three-channel: SWI + Phase + Magnitude Two-channel: SWI and Complement phase Two-channel: SWI and Complement phase with averaging of adjacent slices*	91.67 97.22 88.89 94.44 100.00* (Sensitivity)	
Fei et al. [48]	Brain MRI	Synthesize the FLAIR modality	T1 T1 + T2* T1 + T2 + T1c models*	23.70 ± 2.16, 0.86 ± 0.02 24.80 ± 1.85, 0.88 ± 0.02* 24.93 ± 1.96*, 0.87 ± 0.02 (PSNR, SSIM)	The triple-input model achieved the highest PSNR values, and the T1 + T2 group achieved the highest SSIM values.
Feng et al. [49]	Breast MRI (DWI, DCE)	Classifying the breast cancer	OA patch PT patch AP-DCE patch DCE ensemble*	70.0 78.2 80.0 84.6* (Accuracy)	The classification performance of multi-channel input was always better than using only a single-channel input.
Chen et al. [50]	Brain MRI	Brain segmentation for GM, WM, CSF	T1 T1-IR T2-FLAIR All All + auto-context*	86.96, 89.70, 79.58 80.61, 85.89, 76.44 81.13, 83.21, 75.34 88.08, 90.93, 82.51 88.50, 91.06, 82.70* (Dice coefficient for GM, WM, CSF, respectively)	
Park et al. [51]	Brain MRI	Segmentation for brain metastases	3D BB + 3D GRE* 3D BB 3D GRE	93.1* 92.6 76.8 (Sensitivity)	

*The detection networks with the best performance for each test dataset. MRI = magnetic resonance imaging, SWI = susceptibility-weighted imaging, FLAIR = fluid-attenuated inversion recovery, PSNR = peak signal-to-noise ratio, SSIM = structural similarity index, DWI = diffusion-weighted image, DCE = dynamic contrast-enhanced, OA = over-appearance, PT = peripheral tissue, AP-DCE = all phases of DCE-MRI, GM = gray matter, WM = white matter, CSF = cerebrospinal fluid, T1-IR = T1 inversion recovery, 3D = three dimensional, BB = black blood, GRE = gradient echo

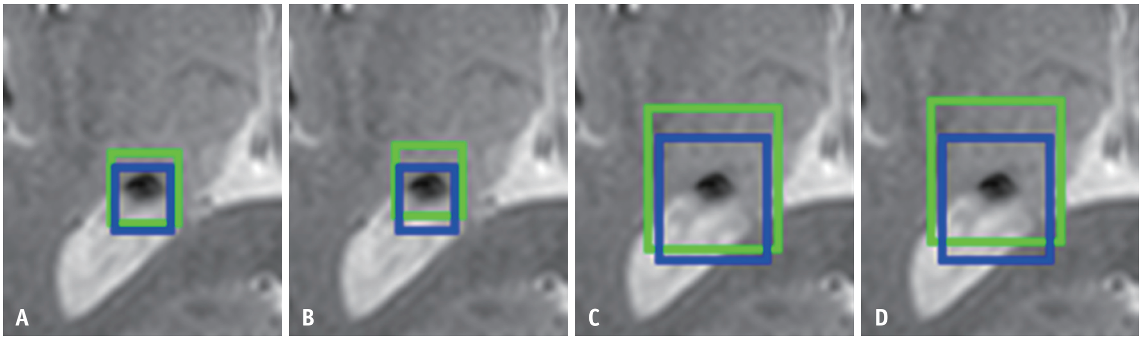

Fig. 5. Example of intersection over union (IoU) degradation with different boundary box sizes for small-object detection. The ground truth and predicted results are represented by blue and green boundary boxes, respectively. The IoU for the small bounding boxes **(A)** and bounding boxes that include a sufficient area around the lesion **(C)** was estimated to be 0.605, and 0.614, respectively. For the small bounding boxes **(B)** and bounding boxes that include a sufficient area around the lesion **(D)**, the predicted boundary box was moved upward by two pixels and the IoU was estimated to be 0.485, and 0.511, respectively.

or cancers less than 3 mm in size, the size of the label box is significantly small. The intersection over union (IoU) [55] in the small label box can be easily decreased, even with a difference of 1–2 pixels between the ground truth and prediction results (Fig. 5). In Figure 5B, the IoU decreased to 0.485 owing to the upward movement of the prediction box by two pixels. If an IoU of 0.5 is used as the threshold for a true positive, it will not be evaluated as a true positive even if the prediction box includes the lesion. Therefore, when the labeling process is performed for small objects, the stability of the IoU can be improved by labeling a large area that includes a sufficient area around the lesion. In the study of [47], to detect cerebral microbleeds based on a deep learning approach, a bounding box with a size of 20 x 20 was applied for the labeling process, which included a sufficient area around the cerebral microbleeds.

Network Training

Object Detection Networks

Object detection architectures have been continuously developed, and several review papers have described their progress from early models to state-of-the-art technologies [56-58]. Therefore, this article does not provide a detailed description of each model. Instead, this article briefly describes the classification of object detection architectures into 2-stage and 1-stage detectors, and a brief description of several object detection models is presented. Table 4 summarizes the performance of some well-known object detection architectures [2,59-69].

The 2-stage detector performs localization and classification separately, whereas the 1-stage detector performs them simultaneously. Generally, a 2-stage detector is recognized for achieving higher accuracy but lower speed. The R-CNN family is a representative 2-stage detector. The R-CNN [59] was the first model to apply a CNN to object detection and consists of region proposal (selective search), feature vector acquisition using a CNN, class classification using a support vector machine, and bounding box regression. However, R-CNN has the disadvantage of long training time owing to the multiple stages of learning. To improve this, a fast R-CNN [2] with a region of interest pooling and faster R-CNN [62] with a region proposal network were developed. Although not in the R-CNN family, the region-based fully convolutional network (R-FCN) model [65], which performs position-sensitive pooling using position-sensitive score maps, showed similar performance to the faster R-CNN but was 2.5 to 20 times faster. Moreover, the feature pyramid network (FPN) [64], which employs a method for recognizing target objects of various sizes, and mask R-CNN [63], which adds a mask branch to enable instance segmentation in the bounding box, have been introduced.

For real-time screening, a 1-stage detector, which has the advantage of high speed, is appropriate. It is known that the 1-stage detector shows lower performance than the 2-stage detector. However, owing to recent developments in the 1-stage detector, its accuracy has become similar to that of the 2-stage detector. The you only look once (YOLO) family is a representative 1-stage detector. YOLO [66], which is the first introduced model, redefines localization and classification, which are separately performed in a 2-stage detector, as a single-regression problem. Consequently, a single neural network predicts the bounding box and class probability using a single process. However, the YOLO model exhibits a lower mean average precision (mAP) value with

Table 4. Well-known Object Detection Networks and Their Performances in the Literature

Proposed Model	Region Proposal	Trained Dataset	Training Time, h	Backbone	Test Dataset	mAP, %	Run-time, s
2-stage detector							
R-CNN [59]	Selective search	ILSVRC2012 + ILSVRC2013	13	-	ILSVRC2013	31.4	60 (CPU)
		ILSVRC2012 + VOC 2012		-	VOC 2010	53.7	
				-	VOC 2007	58.5	
				OxfordNet	VOC 2007	66.0	
Fast R-CNN [2]	Selective search	VOC 2007 + 2012	9.5	VGG16	VOC 2012	70.0	0.3
				VGG16	VOC 2010	68.8	
				VGG16	VOC 2007	68.4	
Faster R-CNN [62]	RPN	VOC 2007 + 2012 + COCO	-	VGG16	VOC 2012	75.9	0.2
				VGG16	VOC 2007	78.8	
				VGG16	COCO	42.7	
R-FCN [65]*	RPN	VOC 2007 + 2012 + COCO	-	ResNet101	VOC 2012	82.0	0.42
				ResNet101	VOC 2007*	83.6*	
				ResNet101	COCO	53.2	1
FPN [64]	RPN	COCO	8 (8 GPUs)	ResNet101	COCO	57.1	0.148
Mask R-CNN [63]	RPN	COCO	44 (8 GPUs)	ResNeXt101FPN	COCO	60.0	0.2
1-stage detector							
YOLO [66]	-	VOC 2007 + 2012	-	-	VOC 2012	57.9	0.02
				-	VOC 2007	63.4	
YOLO v2 [67]	-	ImageNet	-	Darknet19	VOC 2012	73.4	0.025
				Darknet19	VOC 2007	78.6	
				Darknet19	COCO	44.0	
YOLO v3 [68]	-	-		Darknet53	COCO	57.9	0.05
YOLO v4 [69]	-	COCO	-	CSPDarknet53	COCO	65.7	0.03
SSD [61]*	-	VOC 2007 + 2012 + COCO	-	VGG16	VOC 2012*	82.2*	
				VGG16	VOC 2007	83.2	0.045
				VGG16	COCO	48.5	
RetinaNet [60]	-	COCO	10–35	ResNeXt101FPN	COCO*	61.1*	0.198

*The detection networks with the best performance for each test dataset. mAP = mean average precision, R-CNN = region-based convolutional neural network, CPU = central processing unit, RPN = region proposal network, R-FCN = region-based fully convolutional network, FPN = feature pyramid network, GPU = graphic processor unit, YOLO = you only look once, SSD = single-shot detector

missing small objects. To overcome the disadvantages of the YOLO model, an SSD consisting of a multiscale feature layer and default box was proposed [61]. Subsequently, several improvements were introduced: YOLO v2 [67] improved performance by applying batch normalization and using an anchor box; RetinaNet [60] used focal loss to solve the class imbalance problem caused by the difference in the number of positive/negative samples used during model training; YOLO v3 [68] improved performance by using DarkNet53 as the backbone architecture and three feature maps; and YOLO v4 [69] combined various methods that affect performance.

If there is no time constraint, it is appropriate to select 2–3 recent and well-known models supported in the relevant development environment, compare them, and select the model with satisfactory performance. In addition, the proposed object detection network can be modified by applying models such as VGG-19 [70], ResNet-50 [71], or Inception v3 [72] as the backbone.

Deep Learning Workflow Based on the Diagnostic Process Performed by Radiologists

Radiologists are empirically trained in anatomy, pathology, imaging techniques, and disease patterns, and they make decisions based on their own experiences and criteria [73]. However, a deep-learning network is not trained with any criteria with anatomical and pathological bases, and the weights of each node are adjusted by training on a specific dataset. Consequently, a deep-learning network cannot provide any criteria for its decision, and even if a deep-learning network presents a correct decision, it may not be accepted if the deep learning model does not provide any decision criteria. Therefore, rather than simply developing a deep learning network to detect disease, one can take a step towards developing a more reliable algorithm by understanding each process of disease diagnosis performed by a radiologist and applying it in similar manner to the deep learning workflow. In this way, the deep learning model can present the results of each step to the radiologists, helping them to better understand and trust the model's results.

Previous studies have proposed deep-learning workflows that mimic the diagnostic processes of radiologists for maxillary sinusitis assessment and mammography phantom image evaluation. For the assessment of maxillary sinusitis [19], the diagnostic processes of a radiologist include finding the facial region, adjusting the window level, increasing the contrast difference, diagnosing the lesion, and generating a clinical report. These processes were imitated and applied to the deep learning workflow in the form of preprocessing, facial patch detection, facial region extraction, image intensity normalization, maxillary sinusitis detection, and detection result generation, which highlights the image with a bounding box and a report regarding the original image space. Figure 6 shows a diagram that compares the process of maxillary sinusitis detection by radiologists and that of the deep learning model. For a mammography phantom image [42], the evaluation processes of a radiologist include finding the phantom region, adjusting the window level and width, evaluating each phantom object, summing phantom scores for each group according to the guidelines of the American College of Radiology digital mammography quality control, and generating reports. These processes were applied to the deep learning workflow in the form of phantom region detection, image intensity normalization, phantom object detection that yields location information as a bounding box with its group and score, summation of each phantom score for each group, and generating reports.

Imaging Data Argumentation

Data collection in the medical field is often limited, and data augmentation is performed to compensate for the insufficient amount of data in the training dataset. Data augmentation may be applied selectively or randomly and includes image processing, such as flipping, rotating, translating, and scaling the image size (magnification or reduction). The labels in the augmented image are identical to those in the original image but with slightly different features, thereby allowing the deep learning model to learn a wider variety of patterns, which can improve its performance. Yadav et al. [74] investigated the effect of data augmentation to distinguish pneumonia images from normal images in a chest X-ray dataset. They set two different augmentation models using different augmentation parameters, and the model that included the augmentation parameters of rotation range, shear range, zoom range, horizontal flip, and vertical

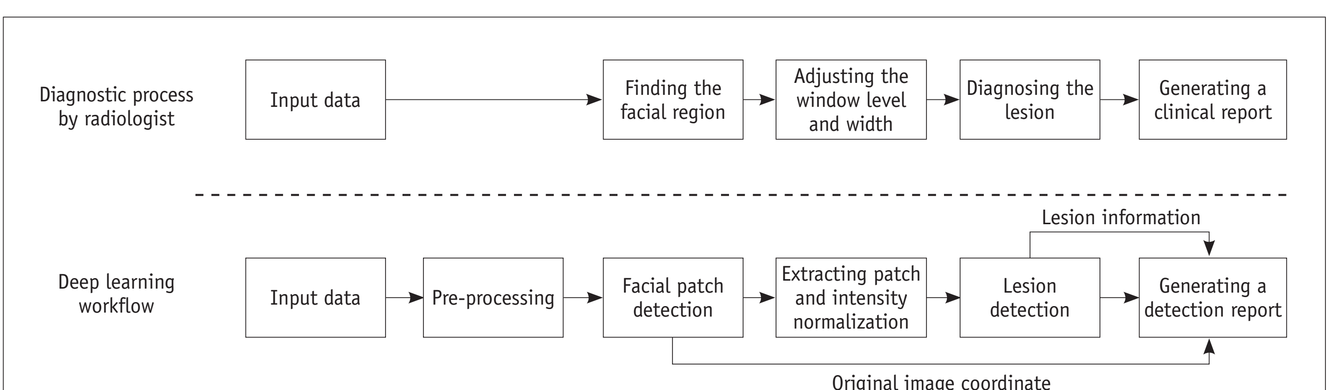

Fig. 6. Diagram of the diagnostic process for maxillary sinusitis performed by a radiologist and deep learning workflow that mimics the radiologist's diagnostic process. The upper row represents the diagram of the diagnostic process conducted by a radiologist, while the bottom row represents the diagram of the deep learning workflow designed to mimic each step of the radiologist's diagnostic process.

https://doi.org/10.3348/kjr.2022.0765

flip showed better results. Liu et al. [75] trained a deep learning network to detect cerebral microbleeds, and their data augmentation comprised 3D rotation, translation, and random left-to-right flipping to avoid overfitting. In a study evaluating the diagnostic performance of deep learning networks on panoramic radiographs, Yang et al. [76] augmented their training set by applying horizontal and vertical flipping, translation, and scaling.

Hyperparameter Optimization

A hyperparameter [77] that affects the performance of a deep learning algorithm is not the main variable optimized through the training process but is instead a variable that humans set as a priori knowledge before training the network. Hyperparameters include the activation function, batch size, dropout rate, number of dense nodes, input image size, epochs, initial learning rate, and a factor for L2 regularization [19,78]. Manual search [79], grid search [80], random search [81], and Bayesian optimization are known hyperparameter optimization methods. The Bayesian optimization method [82,83] uses prior knowledge to generate a statistical model based on experimental results, and it effectively determines the next search direction for the optimal hyperparameters by evaluating the objective function [84]. It has the advantage of efficiently finding the optimal hyperparameters in a shorter time than random or grid searches [85]. In our previous study [19], to enhance the performance of the maxillary sinusitis detector, Bayesian hyperparameter optimization was applied using the following parameters: input image size, number of anchor boxes, maximum epochs, initial learning rate, and a factor for L2 regularization. Bayesian hyperparameter optimization usually attempts to find values of hyperparameters that minimize an objective function and to find hyperparameters that increase the accuracy during the Bayesian optimization process. Ait Amou et al. [78] applied Bayesian optimization to obtain optimal hyperparameters for the complete training of their model to distinguish brain tumors. The activation function, batch size, dropout rate, number of dense nodes, and gradient descent optimization function were selected for Bayesian hyperparameter optimization, and their accuracy was evaluated as an objective function.

Performance Evaluation

Quantitative Performance Metric

For the object detection task, IoU, precision, recall, average precision (AP), and mAP are mainly used to evaluate model performance. IoU is a metric that evaluates how much the predicted boxes overlap with the ground-truth bounding boxes and can be represented by Eq. (1) The IoU is used as a criterion to determine true and false positives and is the most popular evaluation metric used for object detection [55]. In general image domains, such as the PASCAL VOC [86] and MS COCO benchmark challenges [87], the performance of object detection models is commonly evaluated using a fixed IoU threshold of 0.5 [88] or multiple thresholds [62,69]. However, in the medical imaging domain, a fixed IoU threshold of 0.5 [89] or lower, such as 0.2 [17], may be used, depending on the specific study. However, it is also important to ensure that the lesions are included in the prediction box of the deep learning model. Precision (Eq. (2)) indicates the proportion of true positives among the total number of objects predicted by deep learning, and recall (Eq. (3)) indicates the proportion of true positives among all ground truths. The area under the curve in the precision-recall graph is calculated and expressed as an AP to quantitatively evaluate the model's performance [90]. The mAP is the mean of the AP values of each target class [91], and it is the same as the AP when only one target class exists. The false positive rate is calculated by dividing the total number of false positives by the number of slices or participants.

$$\text{IoU} = \frac{|\text{Ground truth} \cap \text{Predicted boxes}|}{|\text{Ground truth} \cup \text{Predicted boxes}|} \qquad \text{Eq. (1)}$$

$$\text{Precision} = \frac{\text{True positives}}{\text{Whole predictions by model}} \qquad \text{Eq. (2)}$$

$$\text{Recall (Sensitivity)} = \frac{\text{True positives}}{\text{Ground truth}} \qquad \text{Eq. (3)}$$

In a clinical setting, the developed computer-aided diagnostic algorithm is often evaluated using sensitivity, specificity, and area under the receiver operating characteristic (AUROC), which are evaluation metrics for distinguishing between normal and abnormal individuals. However, from Eq. (2) and (3), the precision and recall for the performance evaluation of the object detection model are the metrics evaluated in the target object unit.

To evaluate the performance for distinguishing normal and abnormal using the object detection model, secondary processing steps, such as considering predictions as normal or abnormal according to the presence or absence of the prediction result, should be performed. The researcher should then evaluate the accuracy, specificity, AUROC, etc.

https://doi.org/10.3348/kjr.2022.0765

Evaluation of Deep Learning Models

In a mammography phantom image evaluation study [42], the inter-rater correlation coefficient for the total group score of the deep learning model and radiologist was 0.54–0.62, which is in the poor-to-acceptable range. However, as a result of evaluating each of the 16 phantom objects, the agreement between the deep learning model and ground truth was low only at the ambiguous point, and a similar pattern was observed in the results from humans. In previous studies on brain metastasis detection [17], although the overall detection sensitivity was 81%, the sensitivity for the small metastasis group (< 3 mm) was only 15%. By not only evaluating the integrated result but also investigating the deep learning performance for subclasses, researchers can identify deep learning that performs well.

CONCLUSION

In this study, we have addressed the potential challenges and important considerations that arise at each step of deep learning research when employing object detection methods. Although recent studies that have applied deep learning have shown remarkable performance, they have not always guaranteed the best results. Researchers can more efficiently perform deep learning research by identifying issues that may pose problems in each step of the research, thereby reducing trial-and-error.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2022.0765.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Kyung Mi Lee. Formal analysis: Jang-Hoon Oh. Funding acquisition: Kyung Mi Lee, Hyug-Gi Kim. Methodology: Hyug-Gi Kim, Jang-Hoon Oh. Project administration: Kyung Mi Lee. Supervision: Hyug-Gi Kim, Kyung Mi Lee. Validation: Hyug-Gi Kim. Visualization: Jang-Hoon Oh. Writing—original draft: Jang-Hoon Oh. Writing—review & editing: Kyung Mi Lee.

ORCID iDs

Jang-Hoon Oh
https://orcid.org/0000-0002-4251-5470
Hyug-Gi Kim
https://orcid.org/0000-0002-6786-9531
Kyung Mi Lee
https://orcid.org/0000-0003-3424-0208

Funding Statement

Kyung Mi Lee 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). Hyug-Gi Kim 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-2021R1F1A1050515).

Acknowledgments

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. Wang YE, Wei GY, Brooks D. Benchmarking TPU, GPU, and CPU platforms for deep learning. arXiv:1907.10701v4 [Preprint]. [posted July 24, 2019; revised October 22, 2019; cited October 4, 2022]. https://arxiv.org/abs/1907.10701
2. Girshick R. *Fast R-CNN*. In: *2015 IEEE International Conference on Computer Vision (ICCV)*; 2015 December 7-13; Santiago, Chile. Danvers: The Institute of Electrical and Electronics Engineers, Inc.; 2015. p. 1440-1448
3. Sun X, Wu P, Hoi SCH. Face detection using deep learning: an improved faster RCNN approach. *Neurocomputing* 2018;299:42-50
4. Balaban S. *Deep learning and face recognition: the state of the art*. In: Kakadiaris IA, Kumar A, Scheirer WJ, editors. *Biometric Surveill Technol Hum Act Identif XII*; 2015 April 20-24; Baltimore, United States. Bellingham: Society of Photo-Optical Instrumentation Engineers; 2015. p. 94570B
5. Brunetti A, Buongiorno D, Trotta GF, Bevilacqua V. Computer vision and deep learning techniques for pedestrian detection and tracking: a survey. *Neurocomputing* 2018;300:17-33
6. Feyzabadi S. Joint deep learning for car detection. arXiv:1412.7854v2 [Preprint]. [posted December 25, 2014; revised July 14, 2016; cited October 4, 2022]. https://arxiv.org/abs/1412.7854
7. Zhang M, Young GS, Chen H, Li J, Qin L, McFaline-Figueroa JR, et al. Deep-learning detection of cancer metastases to the brain on MRI. *J Magn Reson Imaging* 2020;52:1227-1236

8. Wang CJ, Hamm CA, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *Eur Radiol* 2019;29:3348-3357
9. Ebsim R, Naqvi J, Cootes TF. *Automatic detection of wrist fractures from posteroanterior and lateral radiographs: A deep learning-based approach*. In: Vrtovec T, Yao J, Zheng G, Pozo JM, editors. *Computational methods and clinical applications in musculoskeletal imaging*. 6th International Workshop, MSKI 2018; 2018 September 16; Granada, Spain. Cham: Springer; 2019. p. 114-125
10. Weikert T, Noordtzij LA, Bremerich J, Stieltjes B, Parmar V, Cyriac J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol* 2020;21:891-899
11. Zhou QQ, Wang J, Tang W, Hu ZC, Xia ZY, Li XS, et al. Automatic detection and classification of rib fractures on thoracic ct using convolutional neural network: accuracy and feasibility. *Korean J Radiol* 2020;21:869-879
12. Choi JW, Cho YJ, Ha JY, Lee YY, Koh SY, Seo JY, et al. Deep learning-assisted diagnosis of pediatric skull fractures on plain radiographs. *Korean J Radiol* 2022;23:343-354
13. Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJPC. Identifying pneumonia in chest X-rays: a deep learning approach. *Measurement* 2019;145:511-518
14. Myung MJ, Lee KM, Kim HG, Oh J, Lee JY, Shin I, et al. Novel approaches to detection of cerebral microbleeds: single deep learning model to achieve a balanced performance. *J Stroke Cerebrovasc Dis* 2021;30:105886
15. von Eschenbach WJ. Transparency and the black box problem: why we do not trust AI. *Philos Technol* 2021;34:1607-1622
16. Zhao Z, Guo Y, Shen H, Ye J. *Adaptive object detection with dual multi-label prediction*. In: Vedaldi A, Bischof H, Brox T, Frahm JM, editors. *Computer Vision - ECCV 2020*. 16th European Conference; 2020 August 23-28; Glasgow, United Kingdom. Cham: Springer; 2020. p. 54-69
17. Zhou Z, Sanders JW, Johnson JM, Gule-Monroe MK, Chen MM, Briere TM, et al. Computer-aided detection of brain metastases in T1-weighted MRI for stereotactic radiosurgery using deep learning single-shot detectors. *Radiology* 2020;295:407-415
18. Kim K, Kim S, Han K, Bae H, Shin J, Lim JS. Diagnostic performance of deep learning-based lesion detection algorithm in CT for detecting hepatic metastasis from colorectal cancer. *Korean J Radiol* 2021;22:912-921
19. Oh JH, Kim HG, Lee KM, Ryu CW, Park S, Jang JH, et al. Effective end-to-end deep learning process in medical imaging using independent task learning: application for diagnosis of maxillary sinusitis. *Yonsei Med J* 2021;62:1125-1135
20. Chen H, Yu L, Dou Q, Shi L, Mok VCT, Heng PA. *Automatic detection of cerebral microbleeds via deep learning based 3D feature representation*. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*; 2015 April 16-19; Brooklyn, United States. New York: The Institute of Electrical and Electronics Engineers, Inc; 2015. p. 764-767
21. Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage Clin* 2015;8:367-375
22. Kang H, Witanto JN, Pratama K, Lee D, Choi KS, Choi SH, et al. Fully automated MRI segmentation and volumetric measurement of intracranial meningioma using deep learning. *J Magn Reson Imaging* 2023;57:871-881
23. Cao G, Xie X, Yang W, Liao Q, Shi G, Wu J. *Feature-fused SSD: fast detection for small objects*. In: Yu H, Dong J, editors. *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)*; 2017 October 14-16; Qingdao, China. Qingdao: International Society for Optics and Photonics (SPIE); 2018. p. 106151E
24. Tong K, Wu Y, Zhou F. Recent advances in small object detection based on deep learning: a review. *Image Vis Comput* 2020;97:103910
25. Liu Y, Sun P, Wergeles N, Shang Y. A survey and performance evaluation of deep learning methods for small object detection. *Expert Syst Appl* 2021;172:114602
26. Akselrod-Ballin A, Karlinsky L, Hazan A, Bakalo R, Horesh AB, Shoshan Y, et al. *Deep learning for automatic detection of abnormal findings in breast mammography*. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, et al editors. *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Third International Workshop, DLMIA 2017, and 7th International Workshop; 2017 September 14; Québec, Canada. Cham: Springer; 2017. p. 321-329
27. Takao H, Amemiya S, Kato S, Yamashita H, Sakamoto N, Abe O. Deep-learning single-shot detector for automatic detection of brain metastases with the combined use of contrast-enhanced and non-enhanced computed tomography images. *Eur J Radiol* 2021;144:110015
28. Nam JG, Hwang EJ, Kim DS, Yoo SJ, Choi H, Goo JM, et al. Undetected lung cancer at posteroanterior chest radiography: potential role of a deep learning-based detection algorithm. *Radiol Cardiothorac Imaging* 2020;2:e190222
29. Meng Z, Fan X, Chen X, Chen M, Tong Y. *Detecting small signs from large images*. In: Zhang C, Palanisamy B, Khan L, Sarvestani SS, editors. *2017 IEEE International Conference on Information Reuse and Integration (IRI)*; 2017 August 4-6; San Diego, United States. Danvers: The Institute of Electrical and Electronics Engineers, Inc.; 2017. p. 217-224
30. Zhao Q, Sheng T, Wang Y, Tang Z, Chen Y, Cai L, et al. *M2det: a single-shot object detector based on multi-level feature pyramid network*. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, The Thirty-First Conference on Innovative Applications of Artificial Intelligence, The Ninth Symposium on Educational Advances in Artificial Intelligence*; 2019 January 27-Feb 1; Honolulu, United States. Honolulu: Association for the Advancement of Artificial Intelligence (AAAI); 2019. p. 9259-9266
31. Cui L, Ma R, Lv P, Jiang X, Gao Z, Zhou B, et al. MDSSD: multi-

scale deconvolutional single shot detector for small objects. *Sci China Inf Sci* 2020;63:120113
32. Cao C, Wang B, Zhang W, Zeng X, Yan X, Feng Z, et al. An improved faster R-CNN for small object detection. *IEEE Access* 2019;7:106838-106846
33. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv:1511.06348v2 [Preprint]. [posted November 19, 2015; revised January 7, 2016; cited October 4, 2022]. https://arxiv.org/abs/1511.06348
34. Fei-Fei L, Deng J, Li K. ImageNet: constructing a large-scale image database. *J Vision* 2009;9:1037
35. Sun Q, Yang Y, Sun J, Yang Z, Zhang J. *Using deep learning for content-based medical image retrieval*. In: Cook TS, Zhang J, editors. *Medical Imaging 2017: Imaging Informatics Healthcare, Research, and Applications*; 2017 February 11-16; Orlando, United States. Orlando: International Society for Optics and Photonics (SPIE); 2017. p. 1013812
36. Kim HG, Lee KM, Kim EJ, Lee JS. Improvement diagnostic accuracy of sinusitis recognition in paranasal sinus X-ray using multiple deep learning models. *Quant Imaging Med Surg* 2019;9:942-951
37. Kim Y, Lee KJ, Sunwoo L, Choi D, Nam CM, Cho J, et al. Deep Learning in diagnosis of maxillary sinusitis using conventional radiography. *Invest Radiol* 2019;54:7-15
38. Han S, Hwang SI, Lee HJ. The classification of renal cancer in 3-phase CT images using a deep learning method. *J Digit Imaging* 2019;32:638-643
39. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging* 2017;30:427-441
40. Al-antari MA, Al-Masni MA, Choi MT, Han SM, Kim TS. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Inform* 2018;117:44-54
41. Jeon Y, Lee K, Sunwoo L, Choi D, Oh DY, Lee KJ, et al. Deep learning for diagnosis of paranasal sinusitis using multi-view radiographs. *Diagnostics (Basel)* 2021;11:250
42. Oh JH, Kim HG, Lee KM, Ryu CW. Reliable quality assurance of X-ray mammography scanner by evaluation the standard mammography phantom image using an interpretable deep learning model. *Eur J Radiol* 2022;154:110369
43. Janssens R, Zeng G, Zheng G. *Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks*. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*; 2018 April 4-7; Washington, DC, United States. Danvers: The Institute of Electrical and Electronics Engineers, Inc.; 2018. p. 893-897
44. Kern D, Mastmeyer A. 3D Bounding box detection in volumetric medical image data: a systematic literature review. *J Image Graph* 2022;10:17-27
45. Zhu X, Zhu L, D Song, Wang D, Wu F, Wu J. Comparison of single- and dual-energy CT combined with artificial intelligence for the diagnosis of pulmonary nodules. *Clin Radiol* 2023;78:e99-e105
46. Kurata Y, Nishio M, Moribata Y, Kido A, Himoto Y, Otani S, et al. Automatic segmentation of uterine endometrial cancer on multi-sequence MRI using a convolutional neural network. *Sci Rep* 2021;11:14440
47. Al-masni MA, Kim WR, Kim EY, Noh Y, Kim DH. Automated detection of cerebral microbleeds in MR images: a two-stage deep learning approach. *NeuroImage Clin* 2020;28:102464
48. Fei Y, Zhan B, Hong M, Wu X, Zhou J, Wang Y. Deep learning-based multi-modal computing with feature disentanglement for MRI image synthesis. *Med Phys* 2021;48:3778-3789
49. Feng H, Cao J, Wang H, Xie Y, Yang D, Feng J, et al. A knowledge-driven feature learning and integration method for breast cancer diagnosis on multi-sequence MRI. *Magn Reson Imaging* 2020;69:40-48
50. Chen H, Dou Q, Yu L, Qin J, Heng PA. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage* 2018;170:446-455
51. Park YW, Jun Y, Lee Y, Han K, An C, Ahn SS, et al. Robust performance of deep learning for automatic detection and segmentation of brain metastases using three-dimensional black-blood and three-dimensional gradient echo imaging. *Eur Radiol* 2021;31:6686-6695
52. Zhu Y, Man C, Gong L, Dong D, Yu X, Wang S, et al. A deep learning radiomics model for preoperative grading in meningioma. *Eur J Radiol* 2019;116:128-134
53. Rolnick D, Veit A, Belongie S, Shavit N. Deep learning is robust to massive label noise. arXiv:1705.10694v3 [Preprint]. [posted May 30, 2017; revised February 26, 2018; cited October 4, 2022]. https://arxiv.org/abs/1705.10694
54. Uhm KH, Jung SW, Choi MH, Shin HK, Yoo JI, Oh SW, et al. Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography. *NPJ Precis Oncol* 2021;5:54
55. Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. *Generalized intersection over union: a metric and a loss for bounding box regression*. In: *The IEEE/CVF Conference on Computer Vision and Vision and Pattern ion (CVPR), 2019*; 2019 June 16-20; Long Beach, United States. California: Institute of Electrical and Electronics Engineers (IEEE); 2019. p. 658-666
56. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. *IEEE Trans Neural Netw Learn Syst* 2019;30:3212-3232
57. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, et al. Deep learning for generic object detection: a survey. *Int J Comput Vis* 2020;128:261-318
58. Hoeser T, Kuenzer C. Object detection and image segmentation with deep learning on Earth observation data: a review-part I: evolution and recent trends. *Remote Sens* 2020;12:1667
59. Girshick R, Donahue J, Darrell T, Malik J. *Rich feature hierarchies for accurate object detection and semantic segmentation*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014; 2014 June 23-28; Ohio, United States. Colombus: Institute of Electrical and

Electronics Engineers (IEEE); 2014. p. 580-587
60. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* 2020;42:318-327
61. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. *SSD: single shot multibox detector*. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer vision – ECCV 2016*. 14th European Conference; 2016 October 11-14; Amsterdam, Netherlands. Cham: Springer; 2016. p. 21-37
62. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017;39:1137-1149
63. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* 2020;42:386-397
64. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. *Feature pyramid networks for object detection*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*; 2017 July 21-26; Honolulu, United States. Honolulu: Institute of Electrical and Electronics Engineers (IEEE); 2017. p. 2117-2125
65. Dai J, Li Y, He K, Sun J. *R-FCN: object detection via region-based fully convolutional networks*. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, editors. *30th Conference on Neural Information Processing Systems (NIPS 2016)*; 2016 December 5-10; Barcelona, Spain. Barcelona: Neural Information Processing Systems Foundation, Inc. (NeurIPS); 2016. p. 379-387
66. Redmon J, Divvala S, Girshick R, Farhadi A. *You only look once: unified, real-time object detection*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*; 2016 June 26-July 1; Las Vegas, United States. Las Vegas: Institute of Electrical and Electronics Engineers (IEEE); 2016. p. 779-788
67. Redmon J, Farhadi A. *YOLO9000: Better, faster, stronger*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*; 2017 July 21-26; Honolulu, United States. Honolulu: Institute of Electrical and Electronics Engineers (IEEE); 2017. p. 7263-7271.
68. Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv:1804.02767 [Preprint]. [posted April 8, 2018; cited October 4, 2022]. https://arxiv.org/abs/1804.02767
69. Bochkovskiy A, Wang CY, Mark Liao HY. YOLOv4: optimal speed and accuracy of object detection. arXiv:2004.10934 [Preprint]. [posted April 23, 2020; cited October 4, 2022]. https://arxiv.org/abs/2004.10934
70. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v6 [Preprint]. [posted September 4, 2014; revised April 10, 2015; cited October 4, 2022]. https://arxiv.org/abs/1409.1556
71. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*; 2016 June 26-July 1; Las Vegas, United States. Las Vegas: Institute of Electrical and Electronics Engineers (IEEE); 2016. p. 770-778
72. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. *Rethinking the inception architecture for computer vision*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*; 2016 June 26-July 1; Las Vegas, United States. Las Vegas: Institute of Electrical and Electronics Engineers (IEEE); 2016. p. 2818-2826
73. Stassa G, Evans JA. Radiology, anatomy, and the medical student. *Radiology* 1969;92:1562-1563
74. Yadav SS, Jadhav SM. Deep convolutional neural network based medical image classification for disease diagnosis. *J Big Data* 2019;6:113
75. Liu S, Utriainen D, Chai C, Chen Y, Wang L, Sethi SK, et al. Cerebral microbleed detection using susceptibility weighted imaging and deep learning. *Neuroimage* 2019;198:271-282
76. Yang H, Jo E, Kim HJ, Cha IH, Jung YS, Nam W, et al. Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. *J Clin Med* 2020;9:1839
77. Aszemi NM, Dominic PDD. Hyperparameter optimization in convolutional neural network using genetic algorithms. *Int J Adv Comput Sci Appl* 2019;10:269-278
78. Ait Amou MA, Xia K, Kamhi S, Mouhafid M. A novel MRI diagnosis method for brain tumor classification based on CNN and Bayesian Optimization. *Healthcare (Basel)* 2022;10:494
79. Bergstra J, Bardenet R, Bengio Y, Kégl B. *Algorithms for hyper-parameter optimization*. In: Shawe-Taylor J, Zemel R, Bartlett P, Pereira F, Weinberger KQ, editors. *25th Annual Conference on Neural Information Processing Systems (NIPS 2011)*; 2011 December 12-14; Granada, Spain. Granada: Neural Information Processing Systems Foundation, Inc. (NeurIPS); 2011. p. 2546-2554
80. Huang Q, Mao J, Liu Y. *An improved grid search algorithm of SVR parameters optimization*. In: Yang Y, editor. *2012 IEEE 14th International Conference on Communication Technology*; 2012 November 9-11; Chengdu, China. Chengdu: The Institute of Electrical and Electronics Engineers, Inc.; 2012. p. 1022-1026
81. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res* 2012;13:281-305
82. Zhang Y, Sohn K, Villegas R, Pan G, Lee H. *Improving object detection with deep convolutional networks via Bayesian optimization and structured prediction*. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015; 2015 June 7-12; Boston, United States. Boston: Institute of Electrical and Electronics Engineers (IEEE); 2015. p. 249-258
83. Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE* 2016;104:148-175
84. Brochu E, Cora VM, de Freitas N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599 [Preprint]. [posted December 12, 2010; cited October 4, 2022]. https://arxiv.org/abs/1012.2599
85. Snoek J, Larochelle H, Adams RP. *Practical Bayesian optimization of machine learning algorithms*. In: Pereira F,

Burges CJ, Bottou L, Weinberger KQ, editors. *26th Annual Conference on Neural Information Processing Systems 2012*; 2012 December 3-6; Lake Tahoe, United States. Lake Tahoe: Neural Information Processing Systems Foundation, Inc. (NeurIPS); 2012. p. 2951-2959
86. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. *Int J Comput Vis* 2010;88:303-338
87. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. *Microsoft COCO: Common objects in context*. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer Vision - ECCV 2014*. 13th European Conference; 2014 September 6-12; Zurich, Switzerland. Cham: Springer; 2014. p. 740-755
88. Girshick R, Donahue J, Darrell T, Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans Pattern Anal Mach Intell* 2016;38:142-158
89. Hardalaç F, Uysal F, Peker O, Çiçeklidağ M, Tolunay T, Tokgöz N, et al. Fracture detection in wrist X-ray images using deep learning-based object detection models. *Sensors (Basel)* 2022;22:1285
90. Zhang E, Zhang Y. *Average precision BT*. In: Liu L, Özsu T, eds. *Encyclopedia of database systems.* Boston: Springer, 2009:192-193
91. Beitzel SM, Jensen EC, Frieder O. *MAP BT*. In: Liu L, Özsu T, eds. *Encyclopedia of database systems.* Boston: Springer, 2009:1691-1692

https://doi.org/10.3348/kjr.2022.0765