IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 June 2022, accepted 11 July 2022, date of publication 19 July 2022, date of current version 2 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3192389

RESEARCH ARTICLE

Computer Aided Facial Bone Fracture Diagnosis (CA-FBFD) System Based on Object Detection Model

GWISEONG MOON[1,2], SEOLA KIM[2], WOOJIN KIM[3], YOON KIM[1,2], YEONJIN JEONG[4], AND HYUN-SOO CHOI[1,2], (Member, IEEE)

[1]Department of Computer Science and Engineering, Kangwon National University, Chuncheon 24341, Republic of Korea
[2]Ziovision, Chuncheon 24341, Republic of Korea
[3]Department of Biomedical Informatics, Kangwon National University, Chuncheon 24341, Republic of Korea
[4]Department of Plastic and Reconstructive Surgery, Kangwon National University Hospital, Chuncheon 24289, Republic of Korea

Corresponding authors: Yeonjin Jeong (no15blade@naver.com) and Hyun-Soo Choi (choi.hyunsoo@ziovision.co.kr)

This work was supported by the Promotion of Innovative Businesses for Regulation–Free Special Zones funded by the Ministry of Small and medium-sized enterprises (SMEs) and Startups (MSS, South Korea) under Grant P0020626.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the International Review Board of Kangwon National University Hospital under Application Nos. KNUH-2021-01-004 and KNUH-2021-11-007.

ABSTRACT Facial bone fractures must be diagnosed and treated as early as possible to avoid complications and sequelae. CT images need to be analyzed to detect fractures, but the analysis is time-consuming, and enough specialists are not available to analyze them. Many classification and object detection studies are being conducted to address these issues. The ability of classification-based studies to pinpoint the exact location of fractures is limited. Object detection-based research, by contrast, is problematic because the shape of a fracture is ambiguous. We propose a computer-aided facial bone fracture diagnosis (CA-FBFD) system to address the aforementioned challenges. This system adopts the object detection model YoloX-S, which is trained using only IoU Loss for box prediction, along with CT image Mixup data augmentation. For training, we used only nasal bone fracture data, whereas for testing, we used several other facial fracture data. During evaluation, the CA-FBFD system achieved an average precision of 69.8% for facial fractures, which is better than the baseline YoloX-S model by a large margin of 10.2%. In addition, the CA-FBFD system achieved a sensitivity/person of 100% for facial fractures, which is considerably better than that exhibited by the baseline YoloX-S model by a margin of 66.7%. Therefore, the CA-FBFD system can effectively minimize the labor of doctors who need to determine facial bone fractures in facial CT.

INDEX TERMS CT images, fracture detection, facial bone fracture, computer-aided diagnoses, nasal bone fracture, YoloX, deep learning, image processing.

I. INTRODUCTION

Many patients seen in emergency departments have facial trauma. The most common causes of facial injuries are assault (44–61%), traffic accidents (15.8%), and falls (15%) [1]–[3]. Patients with facial trauma should receive appropriate treatment depending on whether or not they have fractured facial bones. If fractured facial bones are not treated properly or are left untreated for more than two weeks, complications and sequelae such as nasal canal rupture and eyeball retraction may occur. For these reasons, it is critical to detect facial bone fractures early. In the past decade, Computed tomography (CT) has become widespread in the United States, and the advancement of radiographic imaging technology has greatly improved the capability of finding fractures to reach an accuracy of 95%. As a result, various methods [4]–[8] have been proposed to treat facial fractures using CT images. Despite technological advances, interpreting a CT

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandra Bertoldo.

scan and diagnosing a patient takes time because the CT scan consists of multiple images. Although facial bone fracture is a common injury that frequently occurs in many places, there is a marked shortage of specialists in radiology, plastic surgery, and oral surgery that can diagnose facial bone fractures.

Many studies have been conducted to detect fractures using engineering techniques to alleviate the aforementioned difficulty. In some of earlier works of fracture detection, researchers mainly focused on detecting fractures in specific bone regions using computer graphics and machine learning [9], [16]–[20]. In [21], a stacked random forest based on feature fusion was used to detect fractures in X-ray images. After extracting edge and shape features from bones, a combined classifier was designed to detect fractures by fusing several classifiers such as Back Propagation Neural Network, K-Nearest Neighbor, Support Vector Machine, Max/Min Rule, and Product Rule [22], [23]. These methods judge the presence of fractured bone in the entire image but do not locate the fractured bone region. As a result, for the classifier to be useful in practice, specialized doctors must detect and locate fractures in different types of bones.

Fracture detection research is being driven to find and locate fractures to replace the specialized doctors. A Class Activation Map (CAM) can provide the approximate location in the classification model. However, its localization capability is limited, and it is difficult to find all fractures in a single image. In addition to using CAMs, other studies use an object detection model [24]–[26]. The above studies were conducted in varying ways because determining the location of the fracture is difficult. The difficulty comes from the ambiguous shape of the fracture, unlike general object detection. When using a classification model, fracture locations are often missed, and when using an object detection model, ambiguous bounding box predictions adversely affect the classification of fractures.

To tackle the remaining challenges, we propose a Computer-Aided Facial Bone Fracture Diagnosis (CA-FBFD) system based on an object detection model called YoloX. The first key point is to use IoU loss for bounding box regression. A general object detection model uses L_1 loss or L_2 loss that directly compares the box's upper-left coordinate, width, and height to regress the bounding box. The upper-left coordinate, width, and height of the box are unimportant for regression of an ambiguous bounding box, such as that of fracture, when compared to general bounding box regression. To properly regress the bounding box, we use IoU loss rather than L_1 (or L_2) loss. Using the IoU loss aids in properly locating the bounding box and has the effect of reducing the negative impact on the classification task.

The second key point is a data augmentation technique adopted for the fracture detection task. Objects from nature images are of various sizes and shapes. As a result, the object detection model for natural images employs data augmentation techniques to generate synthetic images in a variety of ways by utilizing the given training images. Objects from CT images, by contrast, have less deformation than objects from natural images. Therefore, we appropriately reduce the deformation range such as rotation, size, and movement to suitably reproduce the deformation of the CT images. Finally, we also use Mixup [34] to maximize the effect of data augmentation. We build a CA-FBFD system using these methods that is trained on only nasal fracture data but can detect other facial fractures as well.

The main contributions of this paper are summarized as follows:

- By using IoU loss for the object detection model, we reduce the adverse effect on the classification performance.
- We apply an appropriate data augmentation technique to improve the performance of facial bone fracture detection.
- Our system achieves 69.8% AP on the test set containing diverse bone types by training the model using only the nasal bone fracture, which outperforms the baseline YoloX model by a large margin of 10.2%.

II. RELATED WORKS

A. FRACTURE CLASSIFICATION

A deep learning fracture diagnosis system is being researched to determine whether or not an image contains a fracture. Yu Jin Seol *et al.* [28] proposed automatic diagnosis of nasal bones based on 3D deep learning. This algorithm has the advantage of reducing false positives by analyzing multiple images at once rather than analyzing each image one by one. However, 3D deep learning algorithm [28] requires much labor from doctors to find the nasal bone in facial CT. Furthermore, even if these algorithms classify the input as a fracture, the doctor must still perform the laborious task of pinpointing the location of the fracture. Because one of the primary benefits of automatic fracture diagnosis is that it saves doctors' time, it is critical to developing a qualified automatic fracture diagnosis system for 2D images containing fractures.

For fracture classification in 2D images, Jun Luo *et al.* [27] proposed a multi-view deep learning algorithm to classify elbow fractures. A multi-view deep learning algorithm makes multi-angle analysis possible. To accurately identify an elbow fracture, this multi-view approach necessitates viewing the fracture from various angles. Leonardo Tanzi *et al.* [26] proposed a femur fracture classification algorithm using the Vision Transformer model. Tanzi *et al.* [26] not only improved the performance of classification models, but also demonstrated the effectiveness of collaboration between physicians and CAD systems. These algorithms have great significance in improving the accuracy of fracture diagnosis.

Hojjat Salehinejad *et al.* [25] raised the problem of disproportionate fracture data. Compared to the amount of data for normal persons without fractures, that for patients with fractures is extremely small. Furthermore, even in the CT images of patients with fractures, most CT images of patients with fractures do not contain fracture parts because they occupy only a tiny area on a face. Hojjat Salehinejad *et al.* [25] proposed a bidirectional long and short-term memory model

to address these issues. Amelia Jim′enez-S′anchez *et al.* [24] investigated the classification of femur fractures into seven types based on X-ray images, with six types classified based on fracture location and, type, and the number of fragments in the fracture, and the remaining type classified as a normal class without fracture. Amelia Jim′enez-S′anchez *et al.* [24] used the curriculum learning method to solve the data imbalance problem. Curriculum learning is a method of scheduling key training samples to learn more by scoring training samples. Hojjat Salehinejad *et al.* [25] can determine an approximate location of the fracture using the Grad-CAM [33], and Amelia Jim′enez-S′anchez *et al.* [24] can determine whether the fracture is anterior or posterior through location classification. However, both algorithms have limitations in their ability to express the location of a fracture.

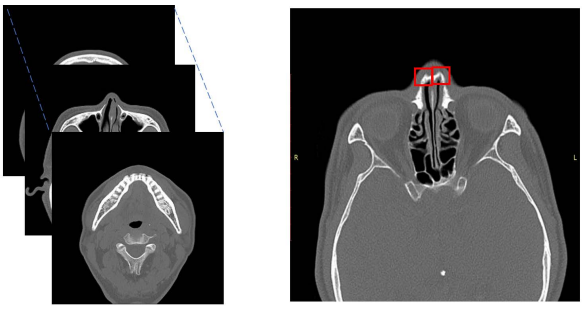
B. FRACTURE REGION DETECTION

The Fracture Classification study has a limitation in that pinpointing the exact location of the fracture is difficult. To overcome this limitation, fracture region detection algorithms that find the location immediately are being researched. Yangling Ma *et al.* [31] proposed the two-stage system for bone fracture detection. The first step is to find the bone structure as a bounding box. Then, in the second step, each bone is checked for fractures with a classification model. Compared to the 3D deep learning model for classification [28], this detection algorithm saves the labor of doctors who need to find bone structures and put them as inputs to a classification model. However, doctors' labor is still required to determine the exact location of a fracture.

Wang *et al.* [29] suggested a fracture detection method that uses a weakly supervised learning model to find fracture candidate regions by creating a fracture probability map. This method locates a fracture in a manner similar to Grad-CAM [33], which is known to be limited in its ability to pinpoint the exact location of a fracture. Fırat Hardala *et al.* [30] used five models to locate fractures as bounding boxes and developed an ensemble method to increase the performance by synthesizing the results of the five models. This ensemble method outperforms the previous methods. However, this method also cannot clearly express fractures as bounding boxes. For further enhancement of fracture location accuracy, we propose a method using IoU loss based on an anchor-free strategy and a decoupled head structure. Also, we try a multi-positive strategy to overcome the problem of imbalanced fracture data.

III. METHODS

In this section, we first describe our data configuration on facial bone CT data, including data collection, labeling by boundary box, and pre-processing to make the bones more visible. Then we introduce YoloX as the baseline for our work, along with five reasons to adopt the YoloX model. In addition, we present the modified data augmentation techniques to fit the facial bone CT images from those for natural images. Finally, we discuss evaluation metrics and the IoU threshold applied to the small fracture box.



(a) Example of a person's facial bone CT (b) Example of fracture on facial bone CT

FIGURE 1. Example of facial bone CT data.

A. DATA

1) DATASET CONFIGURATION

In this study, we used facial bone CT data of patients prescribed for nasal bone fractures between 2014 and 2020 at a university hospital. Only axial CT images were used in this study, and nasal bone fractures were newly labeled with bounding boxes to indicate which part of the bone was fractured. The nasal bone fracture bounding box was pre-labeled by the deep learning developer, who had been trained by a plastic surgeon to detect nasal bone fractures. Then, referring to the pre-labeling, the plastic surgeon corrected, deleted, and added the facial bone fracture boundary box to complete the labeling.

For training purposes, 65,205 facial bone CT images from 690 patients with approximately 5,000 nasal bone fracture bounding boxes were created. For validation data, 4,681 facial bone CT images from 50 patients with approximately 500 nasal bone fracture bounding boxes were used. For the test data, we used facial bone CT images of 20 patients without fractures and facial bone CT images of 20 other patients with nasal bone fractures. The test dataset includes about 400 fracture bounding boxes. Although the training data only contains nasal bone fracture data, the test data includes a small number of facial bone fractures other than the nasal bone fracture to see if the detector can detect other facial bone fractures besides the nasal bone fracture. Figure 1 depicts a sample of facial CT data.

This study was approved by the International Review Board of Kangwon national university hospital (KNUH-2021-01-004, KNUH-2021-11-007)

2) DATASET PREPROCESS

Generally medical professional views and analyzes CT images pre-processed by image filters that are suitable for diagnostic purposes. When diagnosing lung diseases, for example, a filter is used to make the lungs visible. As a result, we use facial CT images pre-processed by filters to make

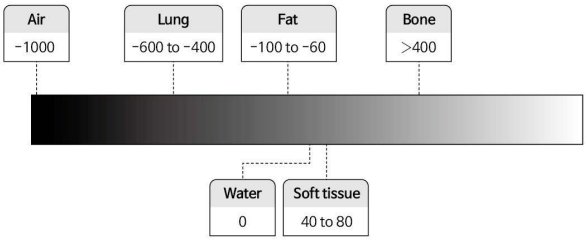


FIGURE 2. The range of Hounsfield units.

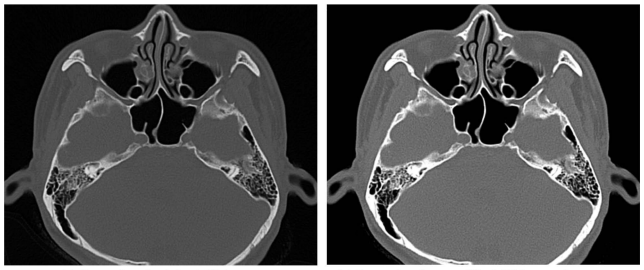


FIGURE 3. Comparison of pre-processing of CT image and original CT image.

the bones more distinguishable as a deep learning input for analyzing nasal bone fractures.

As shown in Figure 2, the range of Hounsfield Units (HU) in the CT data value is approximately −1024 to 3000. From experience, we know that HU of bone is over 400, HU of soft tissue is from 40 to 80, HU of water is 0, HU of fat is from −100 to −60, HU of the lung is from −600 to −400 and HU of air is −1000. Considering this experience, we pre-process the CT image to highlight the bones as

$$I(p) = \begin{cases} 0 & \text{if HU(p)} < \min \\ \frac{\text{HU(p)}-\min}{\max-\min} * 255 & \text{if } \min \le \text{HU(p)} < \max \\ 255 & \text{if HU(p)} \ge \max \end{cases}, \quad (1)$$

where $I(p)$, and HU(p) are the pre-processed CT image value and HU value at the pixel p, respectively, whereas min and max values are set to -800 and 1200, respectively. Figure 3 compares the pre-processed CT image ($I(p)$) to the original CT image.

B. YOLOX-BASED OBJECT DETECTION

1) DATA AUGMENTATION

YoloX was designed to detect objects in natural images. As a result, the majority of YoloX data augmentation methods reflect various object deformations that can occur in natural images. The data augmentation methods for YoloX include Mosaic [35], MixUp [34], scale, rotate, translation, and flip. For convenience, this augmentation method is denoted by *BaseAug*, which stands for baseline augmentation. *BaseAug* must be modified to fit the CT environment to perform well in CT image analysis. *BaseAug* fills the augmented void with meaningless 114, but a value of 114 on CT means there is some substance. Therefore, the augmentation method that transforms *BaseAug* to fit the CT environment is defined as *CTbase*, and *CTbase* fills the augmented empty space with 0, which denotes an air layer with nothing in the CT. Because the natural image is made up of RGB channels, *BaseAug* employs hue, saturation, and value (HSV) augmentation to change the saturation, brightness, and so on. However, because CT images are black and white, it is not appropriate to use HSV augmentation; therefore, *CTbase* does not use it. *CTbase* should be improved further because CT images of the facial bone have far less variation than natural images. Therefore, we revise the data augmentation method so that the augmented facial bone CT images have small variations. To this end, we modify *CTbase* by moderately reducing the deformation range and eliminating Mosaic [35] and Mixup [34], referred to as *CTaug*, which means a CT image augmentation. Specifically, the range of size-changing ratio is reduced from [0.2, 2] for *BaseAug* to [0.8, 1.2] for *CTaug*. As a result, the object size in CT image is varied from 0.8 to 1.2 times the reference size depending on the face size of a patient. In *BaseAug*, the images are flipped with a probability of 0.5, but the flip is not applied in *CTaug*. The degree of angular deformation and position change in natural images is considered to be a value that can properly reflect small changes in posture and position of patients taking CT, so the same values are applied in *CTaug*.

In addition, to augment images including diverse fractures, we add Mixup [34] data augmentation to *CTaug*, which is referred to as *CTmixup*. The Mixup [34] data augmentation method helps to find a little more variety of fractures by superimposing two images in CT images. The difference from the existing Mixup [34] is that the size change ratio during Mixup is reduced from [0.5, 1.5] for *BaseAug* to [0.8, 1.2] for *CTmixup*.

2) YoloX MODEL

In our study, we used YoloX as the baseline model for the five reasons listed below. First, YoloX's performance on the coco benchmark dataset was validated.

Second, YoloX uses an anchor-free strategy, unlike other Yolo series models. In an anchor-based strategy, the ground truth bounding boxes need to be statistically analyzed because the anchor should be set based on statistical analysis. Because the nasal bone fracture bounding box is inconclusive and ambiguous, defining a box criterion in the problem of nasal bone fracture detection is inappropriate. Deep learning models using anchors usually predict multiple object bounding boxes per grid, whereas those without anchors predict one object per grid in many cases. Because nasal bone fracture bounding boxes rarely overlap, the anchor-free strategy works well for our task.

Third, the IoU loss in YoloX for regression is appropriate for the ambiguous bounding box for nasal bone fracture. Generally, the object detection loss consists of a combination of the classification loss and regression loss between ground

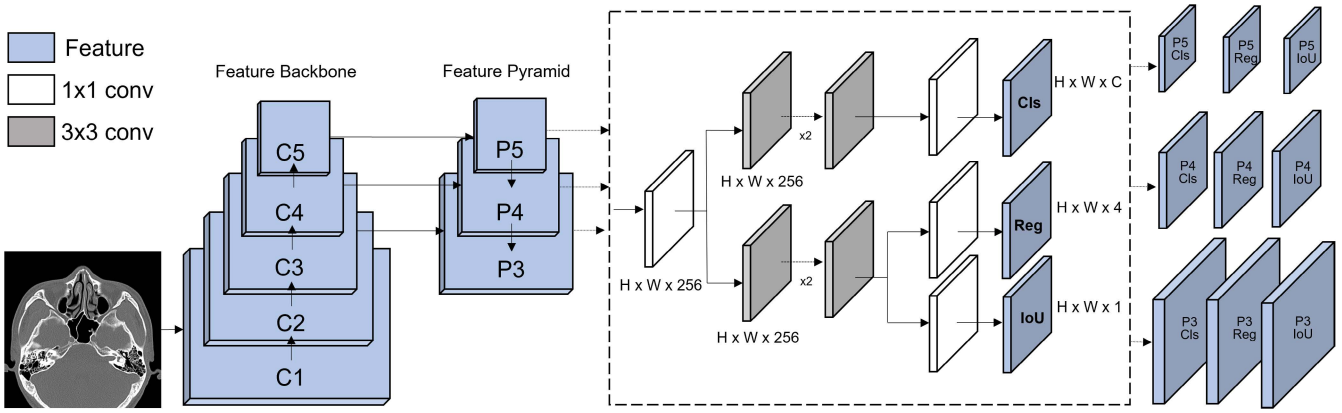


FIGURE 4. YoloX-S model structure used in our study.

truth and prediction as in Equation (2).

$$C_{ij} = L_{ij}^{cls} + \lambda L_{ij}^{reg}. \tag{2}$$

The regression loss directly compares the upper-left coordinates and the height and width of the ground truth with the prediction boxes. YoloX uses IoU loss in addition to the coordinate-based regression loss. Unlike natural images, we use only IoU loss ($\mathcal{L}_{\text{IoU}}$) in (3) for regression because the corner points of the nasal bone fracture bounding box are ambiguous.

$$\mathcal{L}_{\text{IoU}} = 1 - \text{IoU}^2. \tag{3}$$

By using only IoU loss without using ambiguous upper-left coordinates, the deep learning model can learn the fracture bounding box flexibly.

Fourth, YoloX uses a decoupled head structure, unlike other Yolo series models, to mitigate the interference of bounding box regression and class classification with each other. A single model that simultaneously handles classification and bounding box regression, leads to an imperfect trade-off [13], [14]. Because a classification model learns salient areas, whereas, a bounding box regression model learns a region around the boundary, an imperfect trade-off occurs. In addition, vague fracture bounding boxes affecting regression will exacerbate the degree of imperfect trade-off. Therefore, a coupled head structure in which the structure outputs the result in a bundle causes an imperfect trade-off in all layers of the model. On the other hand, decoupled heads, as shown inside the dotted line in Fig. 4, yield classification results and bounding box regression results independently to mitigate the degree of imperfect trade-off. As a result, many object detection models [14]–[16] employ a decoupled head structure. And other Yolo series models use a coupled head, but YoloX has adopted the decoupled head structure.

Fifth, YoloX employs a multi-positive strategy to address the problem of data imbalance, which is common in medical data such as fracture data. Imbalanced datasets cause serious problems in classification tasks in machine learning (ML) [32]. As Hojjat Salehinejad *et al.* [25] points out, there is a data imbalance problem because there are significantly fewer fracture data than non-fracture data. The multi-positive strategy is a sampling strategy that trains a deep learning model by considering the predicted bounding box in a 3×3 region centered on the fracture location as positive samples, which is called "center sampling" in FCOS [15], which alleviates the problem of data imbalance in fracture data by compensating for the insufficient number of positive samples.

YoloX is divided into YoloX-Nano, YoloX-Tiny, YoloX-S, YoloX-M, YoloX-L, and YoloX-X according to the backbone structure. Among them, we use the YoloX-S model. The reason is that YoloX-S model can achieve good performance with one graphics card. Figure 4 depicts the overall configuration of the YoloX-S model used in our study.

C. PERFORMANCE ASSESSMENT

For performance evaluation, each predicted bounding box is classified as true positive (TP), false positive (FP), or false negative (FN). A positive class defines a fracture class, whereas a negative class defines a non-fracture box. For evaluation, we measure average precision (AP) in (6) calculated by using precision, recall, and AP in (4),(5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

$$\text{Average Precision} = \int_0^1 p(r)dr, \tag{6}$$

where $p(r)$ is the precision curve depending on the recall value (r).

In addition, we measure the fracture diagnosis time per person. Because the size of a natural image's bounding box varies greatly, the criterion for judging TP is set to IoU 0.5 or greater. However, the size of the detection box for a nasal bone fracture is quite small, measuring approximately 19 x 20 pixels on average. In this case, IoU is usually small. Thus, we assign TP to a box with IoU ≥ 0.1.

TABLE 1. Ablation study for fracture detection by YoloX-S Model depending on loss functions and data augmentation techniques.

Model	L_1	L_2	$\mathcal{L}_{IoU}$	BaseAug	CTbase	CTaug	CTmixup	AP	sensitivity/person	specificity/person	F1 score/person
M_1	√							0.607	42.9%	100%	0.600
M_2		√						0.591	23.8%	100%	0.385
M_3	√		√					0.602	38.1%	100%	0.552
M_4		√	√					0.588	28.6%	100%	0.444
M_5			√					0.577	33.3%	100%	0.500
M_6	√			√				0.591	38.1%	100%	0.552
M_7		√		√				0.588	33.3%	100%	0.500
M_8	√		√	√				0.596	33.3%	100%	0.500
M_9		√	√	√				0.592	81.0%	73.7%	0.791
M_{10}			√	√				0.612	95.2%	78.9%	0.889
M_{11}	√		√		√			0.655	81.0%	94.7%	0.872
M_{12}			√		√			0.688	95.2%	63.2%	0.833
M_{13}			√			√		0.692	95.2%	73.7%	0.870
M_{14}			√				√	0.698	100%	84.2%	0.933

BaseAug: Baseline augmentation used in YoloX-S for natural images.
CTbase: Modified BaseAug for CT images.
M_8: Baseline of YoloX-S.
$M_{11}, M_{12}, M_{13}, M_{14}$: New Trials for CT images.
M_{14}: Proposed Method; Best one for CT images

IV. RESULTS

This section explains how the test dataset was built for evaluation. Subsequently, we present an ablation study using quantitative and qualitative analyses to validate the proposed components using the loss and data augmentation method. Finally, a discussion is provided.

A. TEST DATASET

The test dataset consists of CT scans of 40 patients. In the test dataset, 19 patients are normal patients without fractures, and 2 patients have both nasal bone fractures and other facial fractures.

One patient has only other facial fractures and no nasal bone fractures, while the other 18 patients have only nasal bone fractures. We only use data from patients with nasal bone fractures as training data.

B. EXPERIMENT SETTING

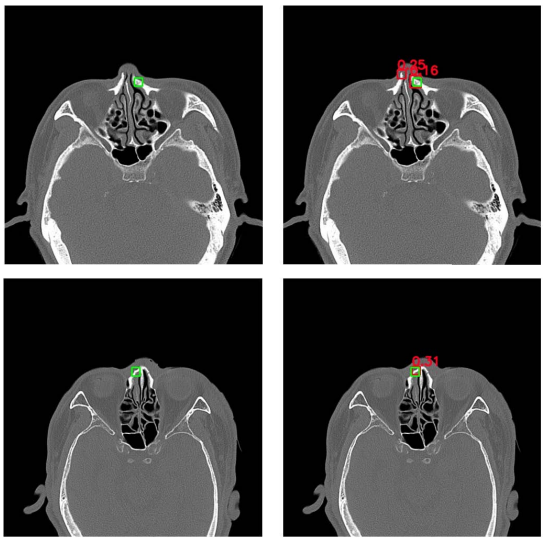
A single GeForce 3080 graphics card was used in an experiment. For prediction boxes with a confidence level of 0.1 or higher, the AP was calculated in the range IoU 0.05 to 0.6. Sensitivity/person and specificity/person were measured as a result of classifying the presence or absence of fractures for each person. If each person has at least one fracture box, the correct answer is considered the person with a fracture. The deep learning model then analyzes the person's face CT data sequentially. If a fracture boundary box was detected in two consecutive CT images, the person was classified as having bone fractures.

C. QUANTITATIVE ANALYSIS

Table 1 shows the results of the ablation study for fracture detection by the YoloX-S Model depending on loss functions and data augmentation techniques. M_1 - M_5 are models trained by changing only the regression loss without applying data augmentation. The AP performances of the M_1 and M_2 models learning the coordinates of the bounding box are better than those of the M_3 and M_4 models with IoU loss added. The M_5 model using only IoU loss has the lowest AP performance. The specificity/person performance of M_1 - M_5 is 100%, while the sensitivity/person performance does not exceed 50%. These findings suggest that fractures are uncommon in all five models. Because data augmentation is not used, it is most likely due to a lack of training data.

M_6 - M_{10} are variant models trained by applying *BaseAug*. Variant models M_6 - M_8 are obtained by applying *BaseAug* data augmentation to M_1 - M_3 models. The AP performances of these models are worse than those without *BaseAug*. By contrast, M_9 and M_{10} improve AP performance by adding *BaseAug* data augmentation to the M_4 and M_5 models. The specificity/person performance decreases slightly, but the sensitivity/person performance improves by more than 80%. This means that unlike M_1 - M_8, M_9 and M_{10} are models that find facial bone fractures well. When comparing before and after adding *BaseAug*, the model with IoU Loss shows less degradation or improves AP performance compared with the model without IoU Loss. This implies the IoU loss is beneficial to find fractures in the case that the training data have plenty of features via data augmentation. In particular, M_{10}, which uses only IoU Loss for regression, shows the best AP performance and the best sensitivity/person performance among M_1 - M_{10} models.

M_{11} - M_{14} models are data augmentation models that have been modified to fit the CT image. The difference between M_{11} and M_{12} is whether L_1 loss is used. The AP performance of M_{11} with L_1 loss is 0.655, and the AP performance of M_{12} without L_1 loss is improved to 0.688. When using L_1 Loss, comparing the sensitivity/person performance and specificity/person performance of M_{11} and M_{12} reveals that the model is trained to find only obvious fractures. When



(a) Results with L_1 Loss (b) Results without L_1 Loss

FIGURE 5. Results of analysis of nasal bone fracture patients depending on the use of L_1 Loss. (a) analysis result of the model using L_1 Loss, M_{11}. (b) analysis result of the model without L_1 loss, M_{12}. Green box is ground true box. Red box is predicted box.

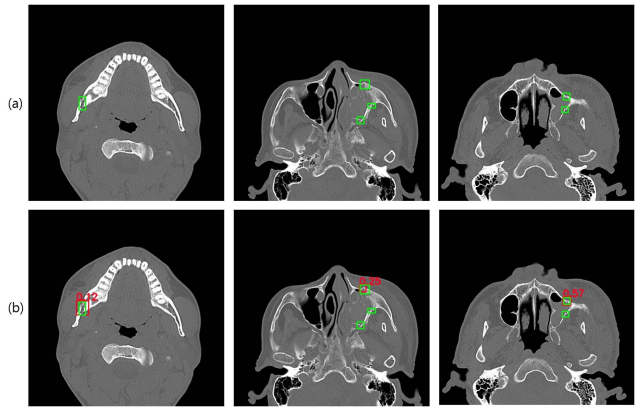


FIGURE 6. Results of analysis of patients with facial bone fractures different from nasal bone fractures depending on the use of L_1 Loss. (a) analysis result of the model using L_1 Loss, M_{11}. (b) analysis result of the model without L_1 loss, M_{12}. Green box is ground true box. Red box is predicted box.

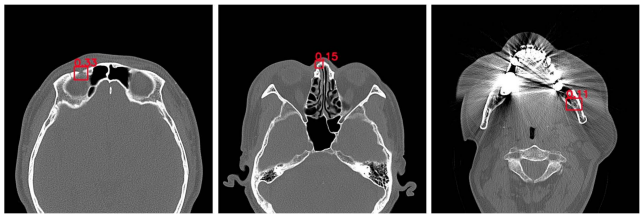


FIGURE 7. False positives in model analysis results without L_1 loss, M_{12}. Red box is predicted box.

collaborating with doctors and the CA-FBFD system, it is more important that the CA-FBFD system detects all fractures even if there are a few false positives. So, we decided to use only IoU Loss. Equation 3 shows the IoU loss equation.

M_{12} uses only IoU loss as regression loss and applies *CTbase* data augmentation technique, named *YoloX-S-CTbase*. M_{13} is a model in which *CTaug* data augmentation is applied, named *YoloX-S-CTaug*. M_{14} is a model that adds Mixup [34], with size parameters adjusted to increase the recall rate of fractures, to *YoloX-S-CTaug*, named *YoloX-S-CTmixup*. As shown in Table 1, the AP performance is improved to 0.688 for *YoloX-S-CTbase* and 0.692 for *YoloX-S-CTaug*, and the highest value reached 0.698 for *YoloX-S-CTmixup*. In addition, the sensitivity/person performance of *YoloX-S-CTmixup* means that all patients with fractures are found.

D. QUALITATIVE ANALYSIS

Figure 5 shows the results of the M_{11} and M_{12} models predicting patients with nasal bone fractures, respectively. (a) is the predicted result of the model with L_1 Loss, and (b) is the predicted result of the model without L_1 Loss. The red bounding box is the fracture bounding box predicted by the model, and the green bounding box is the ground truth bounding box. The number above the red box is the confidence score predicted by the model with a value between 0 and 1. The model with L_1 Loss does not detect the nasal bone fracture, whereas the model without L_1 Loss does. However, the model without L_1 Loss detects a false positive.

The prediction results for patients with facial bone fractures other than nasal bone fractures are shown in Figure 6. (a) is the result predicted by the model using L_1 Loss (M_{11}), and (b) is the predicted result by the model not using L_1 Loss (M_{12}). The model using L_1 Loss does not find any other facial bone fractures. This is because other facial bone fractures are not precisely learnt from the training data. However, the model without L_1 loss detects other facial bone fractures.

The model without L_1 Loss, by contrast, has the disadvantage of detecting more false positives. Figure 7 shows the results predicted by the model without L_1 loss (M_{12}) for patients without fractures. The detected figure does look like a fracture but is a false positive. By contrast, the model with L_1 Loss (M_{11}) predicts no fracture in the above cases. We performed data augmentation to compensate for the case of too many false positives in the model without L_1 loss(M_{12}).

Figure 8 shows the results of analyzing the same patients in Figure 5 with three models. (a) is the result of *YoloX-S-CTbase*, M_{12}, (b) is the result of *YoloX-S-CTmixup*, M_{14}, and (c) is the result of *YoloX-S-CTaug*, M_{13}. *YoloX-S-CTbase* predicts one false positive. But, the proposed *YoloX-S-CTaug* and *YoloX-S-CTmixup* do not yield any false positives.

Figure 9 depicts the detection results from three models for the same patients tested in Figure 6. As seen in (c), *YoloX-S-CTaug* finds only one facial bone fracture different from a nasal bone fracture. As shown in (b), *YoloX-S-CTmixup* does not find as many facial bone fractures different from nasal bone fractures as *YoloX-S-CTbase* but finds more of them than *YoloX-S-CTaug*. However, as shown in Figure 7,

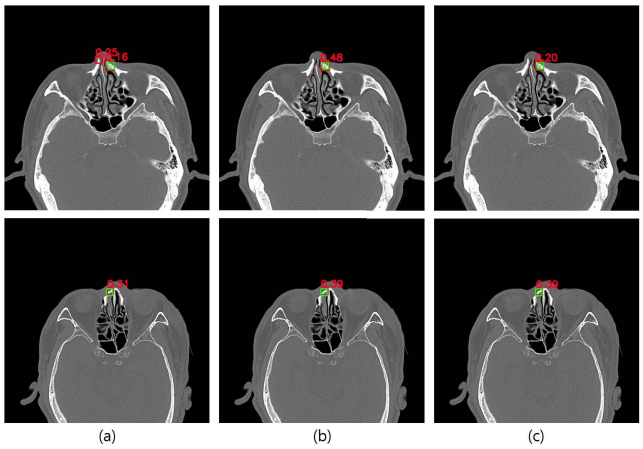


FIGURE 8. Results of analysis of patients with nasal bone fractures of three models depending on data augmentation methods. (a) analysis result of *YoloX-S-CTbase*, M_{12}. (b) analysis result of *YoloX-S-CTmixup*, M_{14}. (c) analysis result of *YoloX-S-CTaug*, M_{13}. Green box is ground true box. Red box is predicted box.

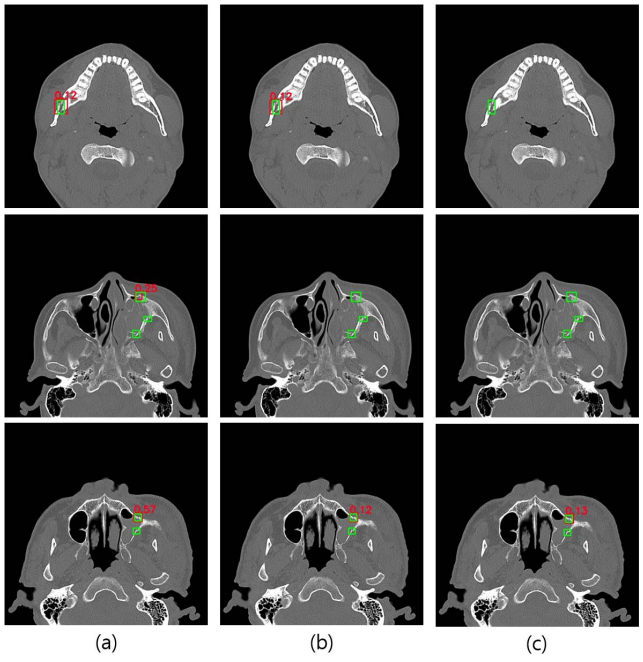


FIGURE 9. Results of analysis of patients with facial bone fractures different from nasal bone fractures of three models depending on data augmentation methods. (a) detection result by *YoloX-S-CTbase*, M_{12}. (b) detection result by *YoloX-S-CTmixup*, M_{14}. (c) detection result by *YoloX-S-CTaug*, M_{13}. Green box is ground true box. Red box is predicted box.

YoloX-S-CTbase detects false positives, but the *YoloX-S-CTaug* and *YoloX-S-CTmixup* do not detect false positives.

In the terms of overall performance, *YoloX-S-CTmixup*, M_{14} is the best, which does not explicitly exploit ambiguous bounding boxes such as its coordinates, and applies the data augmentation method reproducing the deformation of the medical image. All of the models tested above take an average of 0.1 second to analyze a single CT image. Therefore, the analysis speed is sufficient to be used actually for our nasal bone fracture analysis in hospitals.

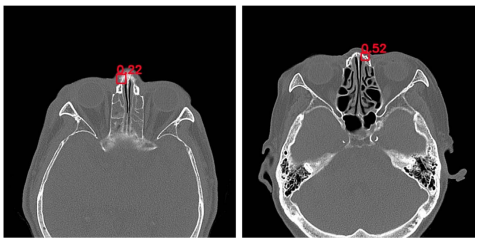


FIGURE 10. Difficult cases to accurately identify a fracture. Left image is a CT image with severe noise. Right image is a CT image of past fracture deformation.

V. DISCUSSION

Automatic fracture detection in facial bone CT images is a challenging task. This is due to the fact that the facial bone has a more complex shape than other bones, and a fracture is difficult to define clearly by a bounding box. In this paper, we use only IoU Loss for the bounding box regression by YoloX-S to solve this problem, and propose *CTmixup* data augmentation method to improve the performance.

Our CA-FBFD system works with the picture archiving and communication system (PACS) server already installed in the hospital so that doctors can use it conveniently. The patient's facial bone CT data is imported through the PACS server, fracture detection is analyzed, and the analyzed results are stored in the hospital database through the PACS server again. Therefore, doctors can easily see the results of analyzing facial bone fractures with the previously used PACS viewer program. Our CA-FBFD system will be a very useful tool for doctors, allowing them to reduce examination time and focus on other more important tasks, allowing them to make faster and more accurate diagnoses. Also, our CA-FBFD system can be of service to patients who are difficult to obtain a fracture diagnosis.

Our CA-FBFD system is good at classifying whether a person has a fracture or not, but the AP performance is only 69.8%. Although significant advances have been made in object detection, it is still difficult to detect small objects. In particular, the deeper the layer in the Feature Pyramid Network, the more the features of small objects may disappear. Yuqi Gong *et al.* [36] uses Fusion Factor in the Feature Pyramid Network to better detect small objects. As we also use the FPN-based YoloX model, we will continue to study to better detect facial bone fractures with many small objects.

We will conduct research to detect nasal bone fractures as well as other types of facial bone fractures. The current study can detect some facial bone fractures that are distinct from nasal bone fractures, but in order to detect all facial bone fractures, the coronal view must also be examined. So, we will also do 2.5D, 3D, and multi-view research. In addition, as illustrated in Figure 10, doctors may find it difficult to distinguish a fracture when only looking at CT images of physical distortion bones after a previous fracture or CT

images with severe noise. Since the site that requires immediate treatment is the site of an acute fracture, it is necessary to differentiate an acute fracture from noise and past fractures. We will investigate the uncertainty score, which indicates how certain the results of deep learning analysis are, to identify cases in which acute fractures cannot be distinguished solely by CT image analysis.

REFERENCES

[1] E. K. Ludi, S. Rohatgi, M. E. Zygmont, F. Khosa, and T. N. Hanna, "Do radiologists and surgeons speak the same language? A retrospective review of facial trauma," *Amer. J. Roentgenol.*, vol. 207, no. 5, pp. 1070–1076, Nov. 2016.

[2] K. H. Lee, "Interpersonal violence and facial fractures," *J. Oral Maxillofacial Surg.*, vol. 67, no. 9, pp. 1878–1883, Sep. 2009.

[3] A. Bakardjiev and P. Pechalova, "Maxillofacial fractures in southern bulgaria—A retrospective study of 1706 cases," *J. Cranio-Maxillofacial Surg.*, vol. 35, no. 3, pp. 147–150, Apr. 2007.

[4] L. Gentry, W. Manor, P. Turski, and C. Strother, "High-resolution CT analysis of facial struts in trauma: 2. Osseous and soft-tissue complications," *Amer. J. Roentgenol.*, vol. 140, no. 3, pp. 533–541, Mar. 1983.

[5] J. L. Marsh, M. W. Vannier, M. Gado, and W. G. Stevens, "*In vivo* delineation of facial fractures: The application of advanced medical imaging technology," *Ann. Plastic Surg.*, vol. 17, no. 5, pp. 364–375, Nov. 1986.

[6] P. N. Manson, B. Markowitz, S. Mirvis, M. Dunham, and M. Yaremchuk, "Toward CT-based facial fracture treatment," *Plastic Reconstructive Surg.*, vol. 85, no. 2, pp. 202–212, Feb. 1990.

[7] F. J. Laine, W. F. Conway, and D. M. Laskin, "Radiology of maxillofacial trauma," *Current Problems Diagnostic Radiol.*, vol. 22, no. 4, pp. 145–188, Jul-Aug. 1993.

[8] E. A. Luce, "Developing concepts and treatment of complex maxillary fractures," *Clinics Plastic Surg.*, vol. 19, no. 1, pp. 125–131, Jan. 1992.

[9] O. Bandyopadhyay, A. Biswas, and B. B. Bhattacharya, "Long-bone fracture detection in digital X-ray images based on digital-geometric techniques," *Comput. Methods Programs Biomed.*, vol. 123, pp. 2–14, Jan. 2016.

[10] A. A. A. Setio, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Stateline, NV, USA, Dec. 2012, pp. 1097–1105.

[12] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[13] G. Song, Y. Liu, and X. Wang, "Revisiting the sibling head in object detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11560–11569.

[14] Y. Wu, Y. Chen, L. Yuan, Z. Liu, L. Wang, H. Li, and Y. Fu, "Rethinking classification and localization for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10183–10192.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[16] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9626–9635.

[17] M. A. Badgeley, J. R. Zech, L. Oakden-Rayner, B. S. Glicksberg, M. Liu, W. Gale, M. V. McConnell, B. Percha, T. M. Snyder, and J. T. Dudley, "Deep learning predicts hip fracture using confounding patient and healthcare variables," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–10, Apr. 2019.

[18] H. Chen, S. Miao, D. Xu, G. D. Hager, and A. P. Harrison, "Deep hierarchical multilabel classification of chest X-ray images," in *Proc. Int. Conf. Med. Imag. With Deep Learn.*, vol. 102, Feb. 2019, pp. 109–120.

[19] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, and M. P. Lungren, "ChexNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," in *Proc. PMED* Nov. 2018, pp. 1–7.

[20] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[21] Y. Cao, H. Wang, M. Moradi, P. Prasanna, and T. F. Syeda-Mahmood, "Fracture detection in X-ray images through stacked random forests feature fusion," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 801–805.

[22] N. Umadevi and S. Geethalakshmi, "Multiple classification system for fracture detection in human bone X-ray images," in *Proc. ICCCNT*, Jul. 2012, pp. 1–8.

[23] V. L. F. Lum, W. K. Leow, Y. Chen, T. S. Howe, and M. A. Png, "Combining classifiers for bone fracture detection in X-ray images," in *Proc. ICIP*, vol. 1, Nov. 2005, p. 1149.

[24] A. Jiménez-Sánchez, D. Mateus, S. Kirchhoff, C. Kirchhoff, P. Biberthaler, N. Navab, M. A. G. Ballester, and G. Piella, "Curriculum learning for improved femur fracture classification: Scheduling data with prior knowledge and uncertainty," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102273.

[25] H. Salehinejad, E. Ho, H.-M. Lin, P. Crivellaro, O. Samorodova, M. T. Arciniegas, Z. Merali, S. Suthiphosuwan, A. Bharatha, K. Yeom, M. Mamdani, J. Wilson, and E. Colak, "Deep sequential learning for cervical spine fracture detection on computed tomography imaging," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 1911–1914.

[26] L. Tanzi, A. Audisio, G. Cirrincione, A. Aprato, and E. Vezzetti, "Vision transformer for femur fracture classification," 2021, *arXiv:2108.03414* .

[27] J. Luo, G. Kitamura, D. Arefan, E. Doganay, A. Panigrahy, and S. Wu, "Knowledge-guided multiview deep curriculum learning for elbow fracture classification," in *Proc. MICCAI*, vol. 12966, Sep. 2021, pp. 555–564.

[28] Y. J. Seol, Y. J. Kim, Y. S. Kim, Y. W. Cheon, and K. G. Kim, "A study on 3D deep learning-based automatic diagnosis of nasal fractures," *Sensors*, vol. 22, no. 2, p. 506, Jan. 2022.

[29] Y. Wang, L. Lu, C.-T. Cheng, D. Jin, P. Adam Harrison, J. Xiao, C.-H. Liao, and S. Miao, "Weakly supervised universal fracture detection in pelvic X-rays," in *Proc. MICCAI*, Sep. 2019, pp. 459–467.

[30] F. Hardalaç, F. Uysal, O. Peker, M. Çiçeklidag, T. Tolunay, N. Tokgöz, U. Kutbay, B. Demirciler, and F. Mert, "Fracture detection in wrist X-ray images using deep learning-based object detection models," *Sensors*, vol. 22, no. 3, p. 1285, Feb. 2022.

[31] Y. Ma and Y. Luo, "Bone fracture detection through the two-stage system of crack-sensitive convolutional neural network," *Inform. Med. Unlocked*, vol. 22, no. 4, Nov. 2020, Art. no. 100452.

[32] J. Luengo, A. Fernández, S. García, and F. Herrera, "Addressing data complexity for imbalanced data sets: Analysis of SMOTE-based oversampling and evolutionary undersampling," *Soft Comput.*, vol. 15, no. 10, pp. 1909–1936, 2011.

[33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[34] H. Zhang, M. Cisse, N. Yann Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," in *Proc. ICLR Conf.*, Feb. 2018, pp. 1–13.

[35] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[36] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao, and Z. Han, "Effective fusion factor in FPN for tiny object detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1159–1167.

GWISEONG MOON received the B.S. degree in mechanical engineering from Sungkyunkwan University, South Korea, in 2015, and the M.S. degree in computer science and engineering from Kangwon National University, South Korea, in 2018, where he is currently pursuing the Ph.D. degree in computer science and engineering. From 2016 to 2017, he was with Geomex-Soft, Chuncheon, Kangwon, South Korea. Since 2017, he has been with Ziovision, Chuncheon. His research interests include machine learning, deep learning, and computer vision.

SEOLA KIM received the B.S. degree in physics from Korea University, South Korea, in 2016, and the Engineering degree in material science with a University's Degree (D.U.) in biomedical engineering from Polytech Paris-Saclay, France, in 2022. Since 2022, she has been with Ziovision, Chuncheon, Kangwon, South Korea, conducting and assisting researches on machine learning models for analysis of medical data, including medical images and sensor data.

WOOJIN KIM received the B.S. and M.S. degrees from the Medical College, Seoul National University, in 1994 and 2004, respectively, and the Ph.D. degree in medicine from Hallym University, in 2006. In 2004, he joined the Department of Internal Medicine, Kangwon National University, where he is currently a Professor. His research interest includes biomedical informatics.

YOON KIM received the B.S., M.S., and Ph.D. degrees in electronic engineering from Korea University, in 1993, 1995, and 2003, respectively. In 2004, he joined the Department of Computer Engineering, Kangwon National University, where he is currently a Professor. Since 2016, he has been with Ziovision, Chuncheon, Kangwon, South Korea. His research interests include deep learning and computer vision.

YEONJIN JEONG received the M.D. degree from the College of Medicine, The Catholic University of Korea, in 2009, and the integrated master's/Ph.D. degrees in plastic and reconstructive surgery from the Catholic University Graduate School, South Korea, in 2018. She received the Boardship of Plastic and Reconstructive Surgery in 2014. After obtaining the Boardship, she had worked at St. Mary's Hospital, Seoul. She is currently working as a Clinical Assistant Professor at Kangwon National University Hospital. Her research interests include facial trauma, and general plastic and reconstructive surgery.

HYUN-SOO CHOI (Member, IEEE) received the B.S. degree in computer and communication engineering for the first major and in brain and cognitive science for the second major from Korea University, in 2013, and the integrated M.S./Ph.D. degrees in electrical and computer engineering from Seoul National University, South Korea, in 2020. From 2020 to 2021, he was a Senior Researcher with Vision AI Labs, SK Telecom. Since March 2021, he has been working as an Assistant Professor at the Education Research Team for Medical Big-Data Convergence, Department of Computer Science and Engineering, Kangwon National University, South Korea. Since October 2021, he has been a Chief Technical Officer with Ziovision.