

# Modified YOLOv4-DenseNet Algorithm for Detection of Ventricular Septal Defects in Ultrasound Images

Shih-Hsin Chen<sup>1</sup>, Chun-Wei Wang<sup>1</sup>, I-Hsin Tai<sup>2,\*</sup>, Ken-Pen Weng<sup>3,\*</sup>, Yi-Hui Chen<sup>4,5,\*</sup>, Kai-Sheng Hsieh<sup>6,\*</sup>

<sup>1</sup>Department of Information Management, Cheng Shiu University, Kaohsiung City 83347, Taiwan (R.O.C.)

<sup>2</sup>Department of Pediatric Cardiology, China Medical University Children's Hospital, Taichung City 40447, Taiwan (R.O.C.)

<sup>3</sup>Congenital Structural Heart Disease Center, Department of Pediatrics, Kaohsiung Veterans General Hospital, No.386, Dazhong 1st Rd., Zuoying Dist., Kaohsiung City, Taiwan (R.O.C.)

<sup>4</sup>Department of Information Management, Chang Gung University, Taoyuan 33302, Taiwan (R.O.C.)

<sup>5</sup>Kawasaki Disease Center, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung 83301, Taiwan (R.O.C.)

<sup>6</sup>Department of Pediatrics, Shuang-Ho Hospital-Taipei Medical University, New Taipei City, 23561, Taiwan (R.O.C.)

**unir**  
LA UNIVERSIDAD  
EN INTERNET

## KEYWORDS

Ventricular septal defect (VSD), Doppler echocardiographic images, Object detection, Deep Learning, YOLOv4

## ABSTRACT

Doctors conventionally analyzed echocardiographic images for diagnosing congenital heart diseases (CHDs). However, this process is laborious and depends on the experience of the doctors. This study investigated the use of deep learning algorithms for the image detection of the ventricular septal defect (VSD), the most common type. Color Doppler echocardiographic images containing three types of VSDs were tested with color doppler ultrasound medical images. To the best of our knowledge, this study is the first one to solve this object detection problem by using a modified YOLOv4-DenseNet framework. Because some techniques of YOLOv4 are not suitable for echocardiographic object detection, we revised the algorithm for this problem. The results revealed that the YOLOv4-DenseNet outperformed YOLOv4, YOLOv3, YOLOv3-SPP, and YOLOv3-DenseNet in terms of metric mAP-50. The F1-score of YOLOv4-DenseNet and YOLOv3-DenseNet were better than those of others. Hence, the contribution of this study establishes the feasibility of using deep learning for echocardiographic image detection of VSD investigation and a better YOLOv4-DenseNet framework could be employed for the VSD detection.

## I. INTRODUCTION

Wu et al. [1] studied the 7-year data of the Taiwan health insurance database and determined that, on average, 13 out of 1000 newborn infants have congenital heart disease (CHD) annually. The 5-year relative mortality rate of infants with CHDs is 5% [2]. With the advances in medical technology, CHDs can be detected using ultrasound images obtained at 18–22 weeks of gestation [3]. Echocardiographic images do not involve radioactivity, cause minimal stress on fetuses and newborns, and are cost effective [4]–[6]. Most parents and doctors begin medical planning at 18–22 weeks. Therefore, it is critical to detect CHDs by using ultrasound images in the early stage.

Two types of ultrasound images are possible, namely black-and-white and color doppler echocardiographic images. The color doppler echocardiographic images can provide critical information on velocities, accelerations, direction of the heart's blood flow (denoted by the red and blue color, respectively), flow

rate, and whether the blood pressure is diastolic or systolic [7]. The red and blue colors represent blood flowing toward and leaving the ultrasound probe, respectively. A golden-yellow color indicates a rapid blood flow. Doppler ultrasound images provide essential features of CHD. Therefore, doppler echocardiographic imaging was used in this study.

The ventricular septal defect (VSD) is the most common CHD and accounts for upto 30% of all CHDs [1]. Therefore, the VSD recognition problem was investigated in this study. Furthermore, three subtypes of VSD, namely Type 1, Type 2, and Type 4, were included in this study. Because type 3 VSD usually is associated with other congenital cardiac anomaly (so called endocardial cushion defect or atrial-ventricular canal defect) and it is very rare to have isolated type 3 VSD; thus, we exclude type 3 VSD to study. Doppler echocardiographic images of the three images are depicted in Fig. 1. Both VSD Type 1 and Type 2 typically involve the parasternal short-axis view's aortic root section. The aortic valve is at the center of the echocardiographic image. Fig. 1a in Fig. 1 shows that if the blood flow of the hole is between 11 and 1 o'clock, then the VSD may be Type 1. If the hole blood flow is between 9 and 11 o'clock, then the VSD is Type 2. The echocardiographic image of a patient with VSD Type 2 are presented in Fig. 1b of Fig. 1.

### \* Corresponding authors:

E-mail address: shchen@csu.edu.tw (Shih-Hsin Chen), m0711103@gcloud.csu.edu.tw (Chun-Wei Wang), ifaithgrace@icloud.com (I-Hsin Tai\*), kpweng@vghks.gov.tw (Ken-Pen Weng\*), kshsieh@hotmail.com (Kai-Sheng Hsieh\*), cyh@mail.cgu.edu.tw (Yi-Hui Chen\*).

For identification of Type 4 VSD, doctors evaluate the horizontal section of the mitral valve of the parasternal short-axis view and determine the doppler flow in the left ventricle from 9 to 1 o'clock. In practice, more than one hole of Type 4 VSD may occur. Therefore, the recognition of this VSD type is difficult. In Fig. 1c, a spot was observed in the 12 o'clock position. To the best of our knowledge, this study is the first to focus on the VSD object detection problem. There is much room to study this problem.

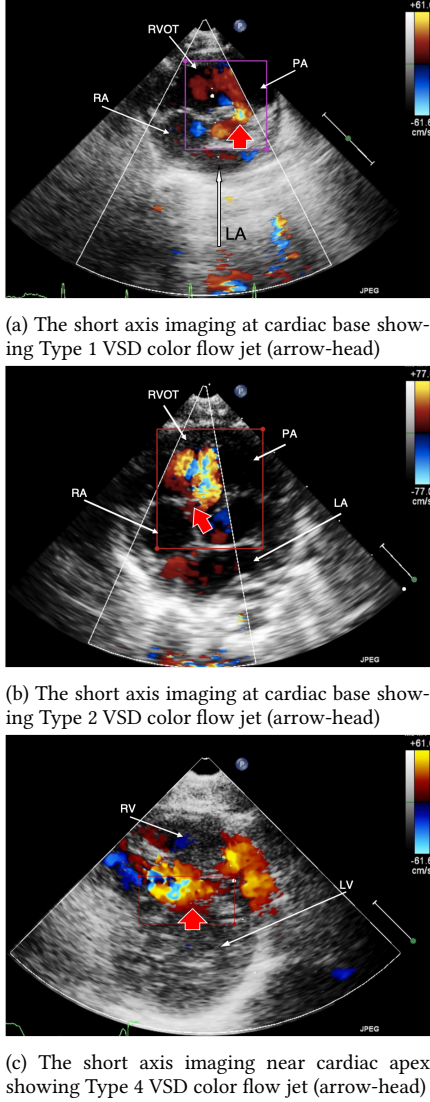


Fig. 1. Doppler echocardiographic of VSD Type 1, Type 2 and Type 4 in the parasternal short-axis view (LV: Left Ventricle, RV: Right Ventricle, RVOT: Right Ventricular Outflow Tract, LA: Left Atrium, RA: Right Atrium, AV: Aortic Valve, PA: Pulmonary Artery)

Different views and angles may lead to difficulties in identifying the characteristics of the three VSD types. Pézard et al. [8] stated that ultrasound image detection of CHDs involved challenges such as the ability and experience of physicians or radiologists [8], ambiguous images, and the nature of defects, which may affect the outcome of the judgment [9]. Furthermore, doctors may capture many ultrasound images when they check a patient; however, determining appropriate image data is time consuming. Deep learning (DL) algorithms used for automatic detection and segmentation of complex cardiac echocardiographic structures may detect the region of interest (ROI) rapidly, thus reducing time and effort required for the process [5], [10]–[15].

You only look once (YOLO)v3 [16], YOLOv4 [17], RetinaNet

[18], and Faster RCNN [19] are some popular algorithms. YOLOv4 is a state-of-the-art algorithm [17] that can improve the quality and efficiency of detection. Compared with YOLOv3, YOLOv4 integrates numerous methods. However, according to our pilot experiments, we determined that some methods may not apply to our studied problem. A modified YOLOv4 was used in the study.

YOLOv4 comprises a cross-stage partial (CSP) network [20] and DarkNet [21]. Because DenseNet [22] can extract more features than DarkNet, DarkNet was replaced with DenseNet121 in YOLOv3, which provided better results [23], [24]. Therefore, we replaced CSPDarkNet with CSPDenseNet121 in YOLOv4. The model was named YOLOv4-DenseNet121 or YOLOv4-DenseNet. To the best of our knowledge, this study is the first to attempt such modification. The proposed algorithm, YOLOv4-DenseNet, is the other major contribution of this study.

The rest of this paper is organized as follows. Section II provides detailed steps of data collection and automatic organization of the data set. The patients' ID, name, or birthday, were removed to ensure privacy and avoid information leak. In Section III, we present the YOLOv4-DenseNet algorithm. Next, we compare the revised YOLOv4-DenseNet with the unmodified YOLOv4 and some variants of YOLOv3. We present a comparison of the proposed algorithms in Section IV. Finally, we present our conclusions in Section V.

## II. MEDICAL IMAGE COLLECTION AND AUTOMATION OF DATA SET ARRANGEMENT

The ultrasound images used in this study were provided by Kaohsiung Veterans General Hospital<sup>1</sup>. Videos were transformed into frames/figures. Doctors identified each figure from the collected images to ensure the correctness of the data set. Sorting the correct classification and the required bounding boxes can be time-consuming. This study developed a standard operating procedure in Fig. 2 considering the professional ability of the doctor, protection of the patient's information, and correctness of the data. The step-by-step procedure is as follows.

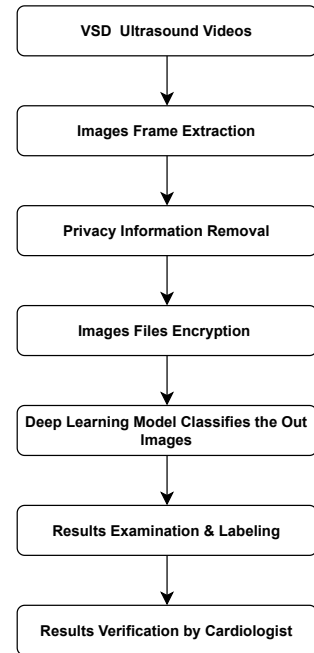


Fig. 2. Dataset arrangement procedures

<sup>1</sup><https://eng.vghks.gov.tw/>

1. Echocardiographic videos were provided by Kaohsiung Veterans General Hospital. The study protocol was approved by the Institutional Review Board (IRB) of the hospital [IRB number is 19-CT8-10(190701-2)]. Patients with VSD diagnosis were selected for the study.
2. We extracted each frame from the echocardiographic videos. The principal image resolution was  $800 \times 600$ .
3. Privacy: We removed images that contained the patient's personal information, such as name, case number, and date of birth. Thus, we removed any identifying information from each figure. The final resolution of the images was  $706 \times 532$ .
4. The images were converted to the PNG format to ensure compression without loss of information. The file name of the image was encrypted using the original video file name, and the custom private key of the project was encrypted using the advanced encryption standard algorithm combined with the video and frame number.
5. We used a few data sets to train the DL model. Then, the initial model was used to classify the images that were not arranged.
6. We then examined the classification results and labeled the images.
7. At least one cardiologist verified the classification and ROIs.

Before implementing Step 1, the coauthors obtained the required licenses from the relevant human research ethics committee. Steps 3–6 were executed after Kaohsiung Veterans General Hospital provided the required medical images. The final steps in processing these data still relied on professional doctors' judgment to provide adequate training quality.

### III. METHODS

The YOLOv4 algorithm is described in Section A. The differences between YOLOv4 and the modified YOLOv4-DenseNet are described in Section B. The metrics applied to evaluate the performance of the proposed algorithm against other algorithms are presented in Section C.

#### A. Main characteristics of the YOLOv4 framework

State-of-the-art algorithms, such as CSP, spatial pyramid pooling (SPP), feature pyramid network (FPN), path aggregation network (PANet), Mish activation function, Mosaic augmentation, dropblock, complete IoU loss (CIoU), class label smoothing, and cosine annealing scheduler, are incorporated in YOLOv4. Fig. 3 presents the modified YOLOv4 framework, which includes three parts, namely the backbone, neck, and head.

CSPDarknet is the main characteristic of the YOLOv4 backbone. The CBM, which is a combination of the convolution layer and the batch normalization (BN) and Mish activation function, was the input of CSPDarkNet. The input resolution of the first convolution layer was  $608 \times 608$ . The Mish function is a self-regular non-monotonic neural activation function that allows relevant information to penetrate the neural network. The ZCRn is composed of zero padding, CBM, and CSPRn. CSPRn denotes the CSPNet framework with  $n$  number of replications. CSPNet divided the feature maps into two parts. In the first part, the gradient changes from the beginning to the end are recorded into the feature map, which reduces the number of calculations and memory costs and ensures high accuracy. The second includes the ResNet skip connections. Finally, the first part is concatenated with the second part's feature maps. The output resolutions of ZCR1, ZCR2, and ZCR8 were  $76 \times 76$ ,  $38 \times 38$ , and  $19 \times 19$ , respectively.

In the neck area, the FPN and PANet are used in YOLOv4, whereas only the FPN is used in YOLOv3. The FPN performs the upsampling from a smaller resolution to larger resolutions

and then concatenates with the large-size ZCRn. The PANet framework employs the bottom-up path augmentation with prior local convolution layers through the upsampling operation to shorten the information path between high- and low-resolution features.

In the head area, YOLOv4 and YOLOv3 use the same head. The output resolution with the number of feature maps was  $76 \times 76/256$ ,  $38 \times 38/512$ , and  $19 \times 19/1024$ . The only change is the loss function. The CIoU is used as the loss function of YOLOv4 to measure the difference between the ground truth and predicted box.

#### B. Modified YOLOv4-DenseNet algorithm

The results of our pilot experiments revealed that not all approaches of YOLOv4 suited our problem. In particular, the performance of the mosaic data augmentation, SPP, and cosine annealing scheduler was not satisfactory. For example, although the mosaic augmentation method is the major cause of the superior performance of YOLOv4 in small object detection, the size of the CHD detection objects is large, and characteristics of CHD are located at a specific area. The cosine annealing scheduler did not yield superior performance. Therefore, the algorithm was suitably modified to address the aforementioned characteristics. Because of the GPU memory limitation, our input resolution decreased to  $416 \times 416$  instead of  $608 \times 608$  used in YOLOv4.

To enhance the performance of the YOLOv4 algorithm, this study replaced CSPDarkNet with CSPDenseNet because DenseNet extracts more information than DarkNet does. In the backbone area of Fig. 4, the ZCRn block was replaced with the CSPDn block. The values of  $n$  are 6, 16, and 24. Two subblocks belonged to the CSPDn block, namely the repeated Denseblocks (blk) and transition block. The number of times Dense blk is repeated is based on the value of  $n$ . In the proposed algorithm, the FPN and PANet were implemented in the neck area. The first difference between the proposed algorithm and YOLOv4 is that we removed the SPP because the SPP lowers performance. Second, because the SPP was removed, the number of CBL was six instead of seven in the top branch.

Finally, in the head area, we used a small input resolution; thus, the output scales were  $52 \times 52$ ,  $26 \times 26$ , and  $13 \times 13$ . Furthermore, we determined that the number of output feature maps may not be useful for our studied problem. To solve this problem, we added more feature maps in the YOLO head. The corresponding output resolution with the number of feature maps was  $52 \times 52/512$ ,  $26 \times 26/1024$ , and  $13 \times 13/1024$ . The loss function of the proposed algorithm was identical to those of the original YOLOv4 algorithm.

#### C. Evaluation metrics

The unlearned image was the target of the test. The threshold of intersection over union was set to 50. If the predicted box of the unlearned images intersected with the ground truth was less than 50, the prediction failed. The four conditions for classification were as follows: true positive, true negative, false positive, and false negative. We evaluated the performance of the model in terms of accuracy (Eq. 1), average precision (Eq. 2), average recall (Eq. 3), and F1-Score (Eq. 4). Finally, we used the average precision metric of Pascal VOC 2012. We calculated the classification mean to obtain the mAP.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

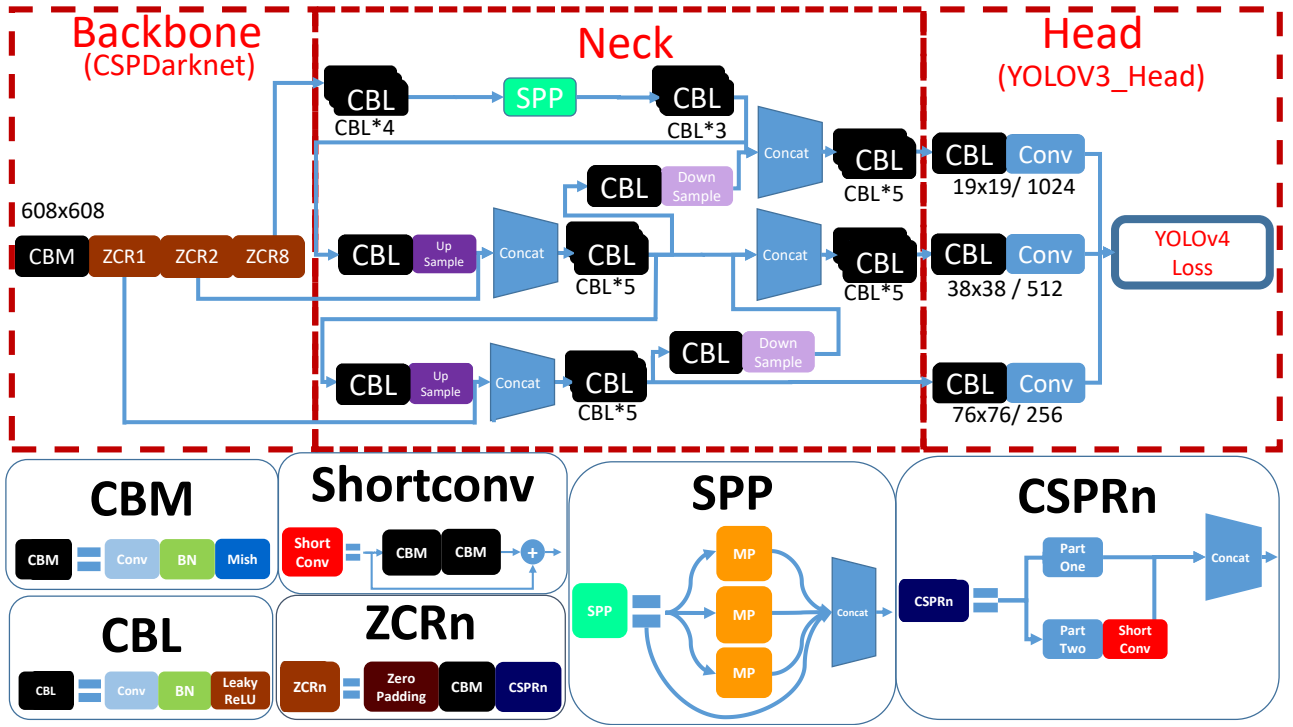


Fig. 3. YOLOv4 framework [17]

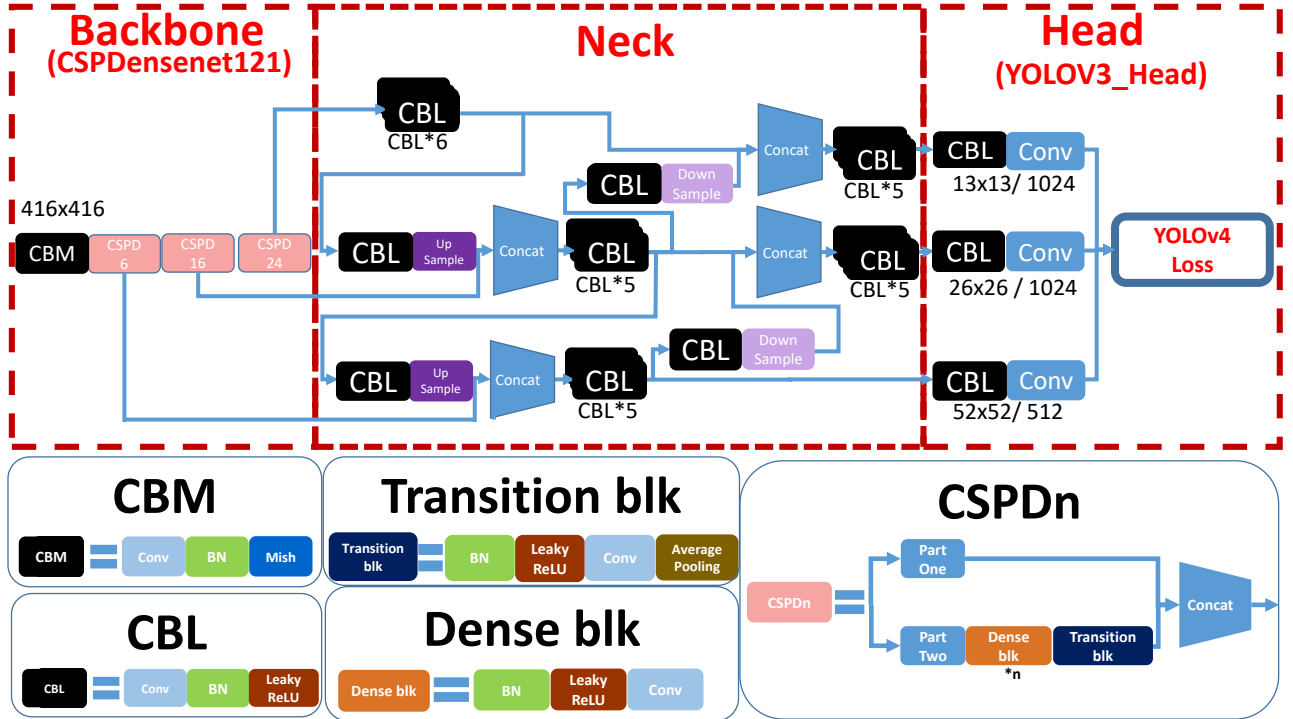


Fig. 4. Proposed algorithm: Modified YOLOv4-Densenet framework



$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

#### IV. EMPIRICAL RESULTS

This research collected 483 images of Kaohsiung Veterans General Hospital. There are 67, 129, and 287 images for type 1, type 2, and type 4 of VSD, respectively. These figures are further divided into train, validation, and test sets. The dataset distribution is shown in Fig. 5.

Based on a YOLOv3 project on Github<sup>2</sup>, we code the YOLOv3-SPP [21], YOLOv3-DenseNet [23], [24], YOLOv3-DenseNet-SPP, YOLOv4 [17], and revised YOLOv4-DenseNet framework by ourselves. Each algorithm runs 1000 epochs with three replications. We executed these algorithms on Tensorflow 1.15.3 environment and nVidia RTX 2080 GPU to experiment. The parameters of the proposed revised YOLOv4-DenseNet are shown as follows. The optimization algorithm is Adam, with the learning rate 1e-4. The number of epochs is 1,000. Due to the limitation of the GPU memory, the number of batch size is set to 4 and the input resolution to be 416\*416 for YOLOv4 instead of 608\*608. Hence, we denote YOLOv4 to be YOLOv4' to distinguish the difference. Finally, we employed the latest model trained by each algorithm to do the following comparisons.

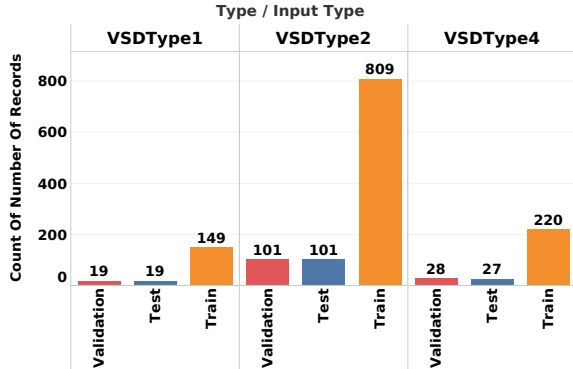


Fig. 5. Dataset arrangement PDF

We list the precision, recall, F1-score, and mAP-50 of the six algorithms in Table 1. When we compare the variants of YOLOv3, YOLOv3-DenseNet might be the best one, according to the F1-score and mAP-50. In particular, when we compare the YOLOv3 with YOLOv3-DenseNet, the F1-score is improved by 10%, and mAP is increased by 20%. The improvement is quite significant. Later on, SPP only improves the combination with YOLOv3 alone; however, SPP does not yield positive outcomes for YOLOv3-DenseNet-SPP because the variance of the mAP-50 values is high. That is the reason why our proposed algorithm does not include the SPP technique in the proposed algorithm.

When it comes to comparing the YOLOv4' and revised YOLOv4-DenseNet with the YOLOv3 variants, YOLOv4' is better than the YOLOv3. YOLOv4-DenseNet is the best one in terms of the result of F1-score and mAP-50. However, YOLOv3-DenseNet remain outperforms YOLOv4'. YOLOv4-DenseNet might be promising because this algorithm inherits the merit of YOLOv4, DenseNet [22] captures more information, and we remove some techniques that may decrease the solution quality.

Most important of all, our proposed algorithm is assisted by CSPDenseNet as the backbone. This strategy enhances the prediction quality.

Among all the algorithms, VSDType1 and VSDType2 achieve satisfactory results. However, VSDType4 does not perform well. For this issue, there is much room for the VSDType4 because most algorithms do not perform well. It is interesting to take a closer look at the correct and incorrect classifications.

In general, the locations of VSD Type 4 are distributed at variable sites within the muscular ventricular septum. Hence, it is good challenge for deep learning algorithms. To explain the other possible reason, we demonstrate the correct and incorrect detection in Fig. 7 to Fig. 9, which presented the three types done by our revised YOLOv4-DenseNet. The blue bounding box is the ground truth marked by us. The bounding box in green means the correct prediction done by the proposed algorithm. Otherwise, the bounding box color is in red. Except the Fig. 7b does not detect the VSDType1 at all, the bounding box in Fig. 8b and Fig. 9b might be too small. This problem may cause the result is not satisfactory. In addition, the symptoms of VSD Type 4 are discovered at varied places. Hence, it is quite necessary to increase the number of training dataset for VSD Type 4.

There are three ways of improving this situation. Firstly, we should increase the number of figures to train the deep learning model. Secondly, due to we set the threshold of IoU to be 50, a smaller detected bounding box yields the incorrect judgment. If we increase the IoU threshold, the mAP result should be increased. Secondly, the bounding boxes prepared by this research might be too large. As a result, we should revise the scale of the bounding boxes for the three types. In general, even though the mAP-50 result of VSDType4 is not satisfactory, it might remain useful for doctors to arrange the echocardiographic images.

#### V. CONCLUSIONS

This paper might be the first one to study the CHD in ultrasound image object detection problem. The revised YOLOv4-DenseNet algorithm is proposed in this paper. The reason for proposing the revised YOLOv4-DenseNet is that some features of YOLOv4 are not suitable for the ultrasound medical image, such as the mosaic data augmentation, SPP, and Cosine annealing scheduler. Later on, due to DenseNet could extract more features than DarkNet, we use CSPDenseNet as the backbone. The proposed algorithm was further compared with the original YOLOv3, YOLOv3-SPP, YOLOv3-DenseNet. We found the revised YOLOv4-DenseNet is the best in terms of the F1-Score and mAP-50. YOLOv4' is better than YOLOv3; however, YOLOv4' is not better than YOLOv3-DenseNet, and YOLOv3-DenseNet-SPP. These results indicated DenseNet as the backbone is effective.

For future research, we plan to improve the prediction quality of VSD Type 4 and study more CHDs, such as Atrial septal defect, Pulmonary stenosis, and Tetralogy of Fallot solved by our proposed algorithm. The hyper-parameters are not optimized; thus, we could use a genetic algorithm to do a global search. Finally, there are some ways of improving the medical image qualities [15], [25], [26]. when the Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied on COVID-19 in chest X-Ray images, the detection accuracy was greatly improved [15]. We attempt to employ the CLAHE (supported in OpenCV) in ultrasound images without modifying the proposed algorithm.

#### ACKNOWLEDGMENT

The data used in this study are restricted by the Research Ethics Review Committee of the Kaohsiung Veterans General Hospital

<sup>2</sup><https://github.com/qqwweee/keras-yolo3>

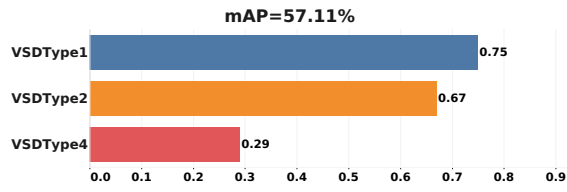
Table 1: Algorithm Models for the VSD

Algorithm	Precision(%)	Recall(%)	F1-score(%)	mAP-50
YOLOv3	99%	70%	81%	57.11%
YOLOv3-SPP	100%	77%	87%	59.54%
YOLOv3-Densenet	99%	84%	91%	71.56%
YOLOv3-Densenet-SPP	99%	83%	90%	70.56%
YOLOv4'	98%	71%	82%	58.42%
Revised YOLOV4-Densenet	97%	85%	91%	72.61%

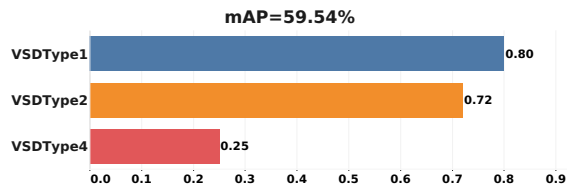
with the number 19-CT8-10(190701-2) to protect participant privacy. We thank the Ministry of Science and Technology for supporting this research with ID MOST 108-2221-E-230-004. We thank the co-author of YOLOv4, Dr. Chien-Yao Wang of the Institute of Information Science, Academia Sinica, Taiwan (R.O.C), clarified their proposed algorithm of CSPDenseNet Ref.-PRN [20]. Finally, we also thank Miss. Wen Mei of Kaohsiung Veterans General Hospital, Miss. Chin Yu worked on the patience data collection and medical images preparations, and Miss. Yu-Chi Lin further enhanced the Fig. 1 quality.

## REFERENCES

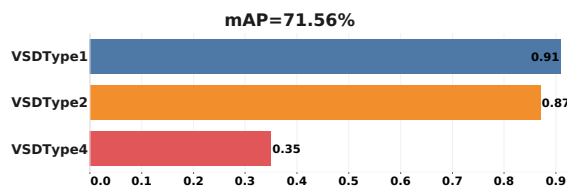
- [1] M.-H. Wu, H.-C. Chen, C.-W. Lu, J.-K. Wang, S.-C. Huang, S.-K. Huang, "Prevalence of congenital heart disease at live birth in taiwan," *The Journal of pediatrics*, vol. 156, no. 5, pp. 782–785, 2010.
- [2] S.-J. Yeh, H.-C. Chen, C.-W. Lu, J.-K. Wang, L.-M. Huang, S.-C. Huang, S.-K. Huang, M.-H. Wu, "National database study of survival of pediatric congenital heart disease patients in taiwan," *Journal of the Formosan Medical Association*, vol. 114, no. 2, pp. 159–163, 2015.
- [3] J. Carvalho, L. Allan, R. Chaoui, J. Copel, G. DeVore, K. Hecher, W. Lee, H. Munoz, D. Paladini, B. Tutschek, *et al.*, "Isuog practice guidelines (updated): sonographic screening examination of the fetal heart," *Ultrasound in Obstetrics & Gynecology*, vol. 41, no. 3, pp. 348–359, 2013.
- [4] M. Avendi, A. Kheradvar, H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Medical image analysis*, vol. 30, pp. 108–119, 2016.
- [5] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, S. K. Zhou, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 487–495, Springer.
- [6] R. P. Poudel, P. Lamata, G. Montana, "Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation," in *International Workshop on Reconstruction and Analysis of Moving Body Organs*, 2016, pp. 83–94, Springer.
- [7] G. Sutherland, M. Stewart, K. Groundstroem, C. Moran, A. Fleming, F. Guell-Peris, R. Riemersma, L. Fenn, K. Fox, W. McDicken, "Color doppler myocardial imaging: a new technique for the assessment of myocardial function," *Journal of the American Society of Echocardiography*, vol. 7, no. 5, pp. 441–458, 1994.
- [8] P. Pézard, L. Bonnemains, F. Boussion, L. Sentilhes, P. Allory, C. Lepinard, A. Guichet, S. Triau, F. Biquard, M. Leblanc, *et al.*, "Influence of ultrasonographers training on prenatal diagnosis of congenital heart diseases: a 12-year population-based study," *Prenatal diagnosis*, vol. 28, no. 11, pp. 1016–1022, 2008.
- [9] G. Hill, J. Block, J. Tanem, M. Frommelt, "Disparities in the prenatal detection of critical congenital heart disease," *Prenatal diagnosis*, vol. 35, no. 9, pp. 859–863, 2015.
- [10] C. P. Bridge, C. Ioannou, J. A. Noble, "Automated annotation and quantitative description of ultrasound videos of the fetal heart," *Medical image analysis*, vol. 36, pp. 147–161, 2017.
- [11] F. C. Ghesu, E. Krubasik, B. Georgescu, V. Singh, Y. Zheng, J. Hornegger, D. Comaniciu, "Marginal space deep learning: efficient architecture for volumetric image parsing," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1217–1228, 2016.
- [12] A. Madani, R. Arnaout, M. Mofrad, R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *npj Digital Medicine*, vol. 1, no. 1, p. 6, 2018.
- [13] M. Moradi, Y. Guo, Y. Gur, M. Negahdar, T. Syeda-Mahmood, "A cross-modality neural network transform for semi-automatic medical image annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 300–307, Springer.
- [14] J. C. Nascimento, G. Carneiro, "Multi-atlas segmentation using manifold learning with deep belief networks," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, 2016, pp. 867–871, IEEE.
- [15] F. A. Saiz, I. Barandiaran, "Covid-19 detection in chest x-ray images using a deep learning approach," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 2, pp. 11–14, 2020.
- [16] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.
- [17] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence*, 2018.
- [19] S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [20] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. Yeh, "Cspnet: A new backbone that can enhance



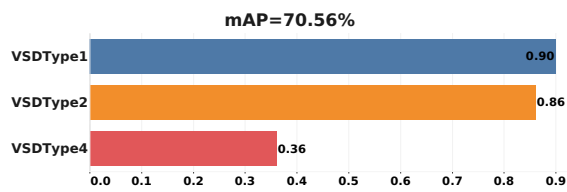
(a) YOLOv3



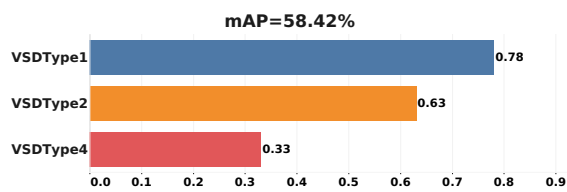
(b) YOLOv3-SPP



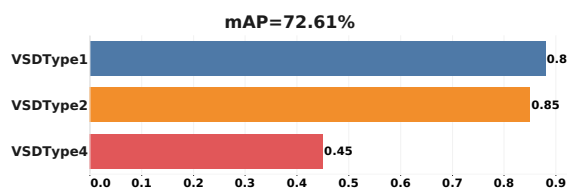
(c) YOLOv3-Densenet



(d) YOLOv3-DenseNet-SPP

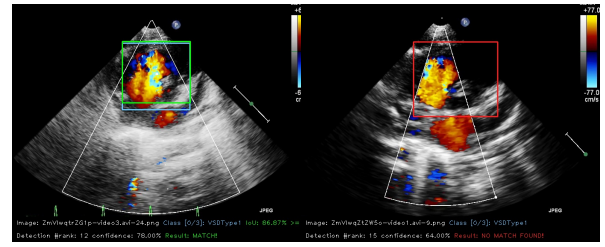


(e) YOLOv4'



(f) Revised YOLOv4-DenseNet

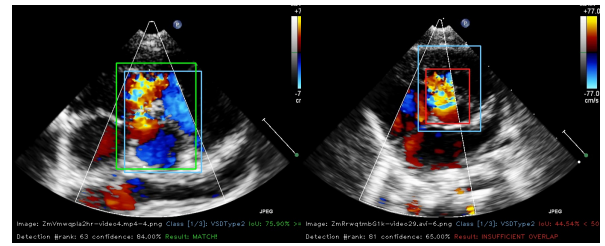
Fig. 6. The mAP performance of the compared algorithms



(a) Correct

(b) Incorrect

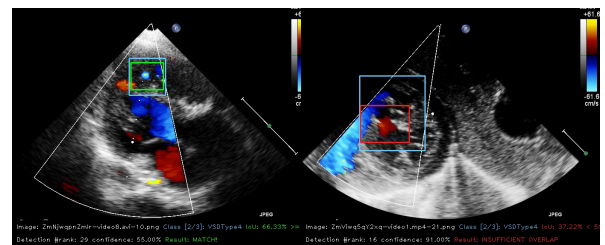
Fig. 7. The correct and incorrect detection of VSDType1



(a) Correct

(b) Incorrect

Fig. 8. The correct and incorrect detection of VSDType2



(a) Correct

(b) Incorrect

Fig. 9. The correct and incorrect detection of VSDType4

learning capability of cnn,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 390–391.

- [21] J. Redmon, “Darknet: Open source neural networks in c,” <http://pjreddie.com/darknet/>, 2013–2016.
- [22] H. Gao, Z. Liu, L. van der Maaten, K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [23] D. Xu, Y. Wu, “Improved yolo-v3 with densenet for multi-scale remote sensing target detection,” *Sensors*, vol. 20, no. 15, p. 4276, 2020.
- [24] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, Z. Liang, “Apple detection during different growth stages in orchards using the improved yolo-v3 model,” *Computers and electronics in agriculture*, vol. 157, pp. 417–426, 2019.
- [25] Z. Liao, M. H. Jafari, H. Girgis, K. Gin, R. Rohling, P. Abolmaesumi, T. Tsang, “Echocardiography view classification using quality transfer star generative adversarial networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 687–695, Springer.
- [26] Q. Nie, Y.-b. Zou, J. C.-W. Lin, “Feature extraction for medical ct images of sports tear injury,” *Mobile Networks and Applications*, pp. 1–11, 2020.



### Shih-Hsin Chen

S. H. Chen received his Ph.D. in Industrial Engineering and Management from Yuan-Ze University, Taiwan (R.O.C) in the year 2008. His research interests include medical image classification and object detection, artificial intelligence, and optimization scheduling problem. There are 30 SCI journal publications with more than 1280 citations by Google Scholar. He is currently the associate professor of the Department of Information Management and to be a group lead of Information Technology at Cheng Shiu University.

### Chun-Wei Wang



He has been a software engineer for five years and is currently a master’s degree student at the Department of Information Management, Cheng Shiu University. His master’s degree thesis studied the ultrasound images of congenital heart diseases, including Ventricular Septal Defect, Atrial septal defect, Patent ductus arteriosus, Pulmonary stenosis, and Tetralogy of Fallot. This research goal will be completed by the end of June in the year 2021.

### I-Hsin Tai



Dr. I-Hsin Tai received his medical doctor degree in 2010 from the National Defense Medical Center, Taipei. Dr. I-Hsin Tai received pediatric cardiology training in Kaohsiung Veterans General Hospital Kaohsiung Chang Gung Memorial Hospital, respectively, and continuously served in Kaohsiung Chang Gung Memorial Hospital participated in a research program of Kawasaki Disease Center. Dr. Tai had won the 2015 & 2016 original cardiology abstract price given by the Taiwan Society of Ultrasound in Medicine. He now moves to China Medical University Children’s Hospital, Taichung, where he served as a pediatric cardiologist emergency physician to take care of critically ill children.



### Ken-Pen Weng

Ken-Pen Weng, M.D. has completed his bachelor of medicine at the age of 25 years old from National Yang-Ming University, Taipei, Taiwan. He is the director of congenital/structural heart disease center, Kaohsiung Veterans General Hospital, Taiwan (R.O.C). He has published more than

50 papers in reputed journals.



### Yi-Hui Chen

She received her Ph.D. degree in computer science and information engineering at the National Chung Cheng University. Later on, she worked at Academia Sinica as a post-doctoral fellow. Later, she worked at IBM’s Taiwan Collaboratory Research Center as a Research Scientist. After that, she worked at the Department of M-Commerce and Multimedia Applications, Asia University. She is now an associate professor at the Department of Information Management, Chang Gung University. Her research interests include data mining, semantic analysis, and multimedia security.

### Kai-Sheng Hsieh



Kai-Sheng Hsieh, MD, FACC, FESC, FCCM was born in Taipei, Taiwan. He received a B.S. degree in 1975 and an M.B. degree in 1978 from National Defense Medical Center, Taipei. Prof. Hsieh was trained as a clinical fellow in pediatric cardiology at Boston Children’s Hospital and Harvard Medical School between 1982-1984. He

was chief of pediatric cardiology at the Department of Pediatrics, Taipei Veterans General Hospital in 1985. Between 1990-2014, he served as Chief of the department of pediatrics, Kaohsiung Veterans General Hospital. Between 2014-2018, he served as General Chairman of Pediatrics, Chang Medical System, Taiwan. He is currently a professor of Pediatrics, Shuang-Ho Hospital-Taipei Medical University. Prof. Hsieh is the author of 250 articles, and 15 book chapters. He has deeply involved in clinical research and clinical teaching. He was the winner of the “Major Devotion to Medical Care of Children in Taiwan” Award in year 2013. Prof. Hsieh also is exceptionally interested in biomedical engineering. He has supervised many theses from master and doctoral degree students in bioengineering fields.