Araştırma Makalesi   Research Article

DETECTION OF STANDARD PLANE FROM ULTRASOUND SCANS BY DEEP LEARNING METHODS FOR THE DIAGNOSIS OF DEVELOPMENTAL HIP DYSPLASIA

Kerim Kürşat ÇEVİK[1*] , Şeyda ANDAÇ[2]

[1] Akdeniz University, Faculty of Applied Sciences, Management Information Systems, Antalya, Türkiye
[2] Bagcilar Training and Research Hospital, Radiology Department, Istanbul, Türkiye

Keywords	Abstract
DDH, *Standard Plane,* *Deep Learning,* *Transfer Learning,* *YOLO.*	The term developmental dysplasia of the hip (DDH) describes a range of hip abnormalities affecting newborns where the femoral head and acetabulum are in improper alignment or grow abnormally, or both. The ultrasonographic evaluation technique rely on the capability of the ultrasonographer to pick up the accurate frame used for exact calculations. In our study we developed a new computer aided system that determines the exact frame from real time 2D ultrasound images and calculates the accuracy rate for each result. The deep learning architectures recently used in literature were utilized for these processes. In addition, transfer learning was carried out to increase the performance of the system using pretrained networks (SqueezeNet, VGG16, VGG19, ResNet50 and ResNet101). One of the best methods of object detection, You Only Look Once (YOLO) model, was used with pre-trained networks to determine DDH location. As a result of the study, the performance of the deep neural network model proposed with the help of these pre-trained networks was evaluated. When the obtained results were compared with expert opinions, frames (standard planes) in 605 of 676 (89.05%) test images were correctly detected. The accuracy rates for the used pre-trained networks were obtained as SqueezeNet 0.79, VGG16 0.95, VGG19 0.96, ResNet50 0.88 and ResNet101 0.93.

GELİŞİMSEL KALÇA DİSPLAZİSİ TANISINDA DERİN ÖĞRENME YÖNTEMLERİYLE ULTRASON TARAMALARINDAN STANDART DÜZLEM TESPİTİ

Anahtar Kelimeler	Öz
GKD, *Standart Düzlem,* *Derin Öğrenme,* *Transfer Öğrenme,* *YOLO.*	Gelişimsel kalça displazisi (GKD) terimi, femur başı ve asetabulumun yanlış hizada olduğu, anormal şekilde büyüdüğü veya her ikisinin birden olduğu yeni doğanları etkileyen bir dizi kalça anormalliği olarak tanımlanır. Ultrasonografik değerlendirme tekniği, ultrasonografi uzmanının kesin hesaplamalar için kullanılan doğru çerçeveyi(standart düzlem) seçme yeteneğine dayanır. Çalışmamızda, gerçek zamanlı 2B ultrason görüntülerinden standart düzlemi belirleyen ve her bir sonuç için doğruluk oranını hesaplayan yeni bir bilgisayar destekli sistemi geliştirilmiştir. Bu işlemler için literatürde son zamanlarda kullanılan derin öğrenme mimarilerinden yararlanılmıştır. Ayrıca önceden eğitilmiş ağlar (SqueezeNet, VGG16, VGG19, ResNet50 ve ResNet101) kullanılarak, sistemin performansını artırmak için transfer öğrenmesi gerçekleştirilmiştir. Nesne algılamanın en iyi yöntemlerinden biri olan You Only Look Once (YOLO) modeli, DDH konumunu belirlemek için önceden eğitilmiş ağlarla birlikte kullanılmıştır. Çalışma sonucunda önceden eğitilmiş bu ağlar yardımıyla önerilen derin sinir ağı modelinin performansı değerlendirilmiştir. Elde edilen sonuçlar uzman görüşleri ile karşılaştırıldığında 676 test görüntüsünün 605(%89,05) 'inde doğru kareler (standart düzlemler) doğru olarak tespit edilmiştir. Kullanılan önceden eğitilmiş ağlar için doğruluk oranları SqueezeNet 0.79, VGG16 0.95, VGG19 0.96, ResNet50 0.88 ve ResNet101 0.93 olarak elde edilmiştir.

* İlgili yazar / Corresponding author: kcevik@akdeniz.edu.tr, +90-242-310-6369

1. Introduction

Developmental dysplasia of the hip (DDH) describes a wide spectrum of abnormal hip development that takes place during infancy and early development, and has an important share in childhood disabilities. In the definition of DDH, a wide range of severity conditions that range from mild acetabular dysplasia without hip dislocation to detect hip dislocation are encompassed (Yang, Zusman, Lieberman, & Goldstein, 2019), 29% of primary hip replacements in people aged under 60 years are reported to be caused by DDH (Dezateux & Rosendahl, 2007). In addition, in women under 40, DDH is reported to be the most common cause of hip arthritis, and it has been also reported that 5% to 10% of all total hip replacements in the United States occurred because of DDH (Shaw & Segal, 2016). Clinical hip imbalance occurs in 1% to 2% in infants who were born on time, and can be detected from imaging studies of hip imbalance or hip maturation at rates up to 15%. Universal clinical screening should be conducted as a part of the physical examination of the newborn.

Although clinical examination continues to be the mainstay in the process of diagnosing the DDH in the newborns, it is not possible to detect DDH in all cases by physical examination. Imaging by ultrasonography or radiography is one of the methods widely used for screening or confirmation of the diagnosis as well as for the classification of the severity of the dysplasia.

The ultrasound hip screening method invented by Graf in 1980 is used for assessing the appropriate maturation of the hip joints of infants (Schams, Labruyère, Zuse, & Walensi, 2017). With this method, ultrasound assessment of the hip is performed by quantifying the maturity of the cartilaginous and bony acetabular roof and the position of the femoral head based on sonographic structures (Graf, 2006). The hip types, as shown in Figure 1, are assessed based on the Graf method range between type I and IV. Type I is related to a normal hip that has a good bony modeling (large arrow in Figure 1), a sharp bony rim (arrowhead) and a narrow, covering cartilage roof triangle (small arrow). Type II embraces physiologically immature hips with rounded rim and cartilage roof triangle (Type 2a) and encompasses hips with delayed pathological ossification cations (Type 2a-c). In addition, it encompasses hips with deficient bony modeling, rounded/flattened bony rim and displaced cartilage roof (Type D). Types 3 and 4 include non-concentric hips with a weak bone model, a flattened bony rim, and a displaced cartilage triangular roof (Dezateux & Rosendahl, 2007).
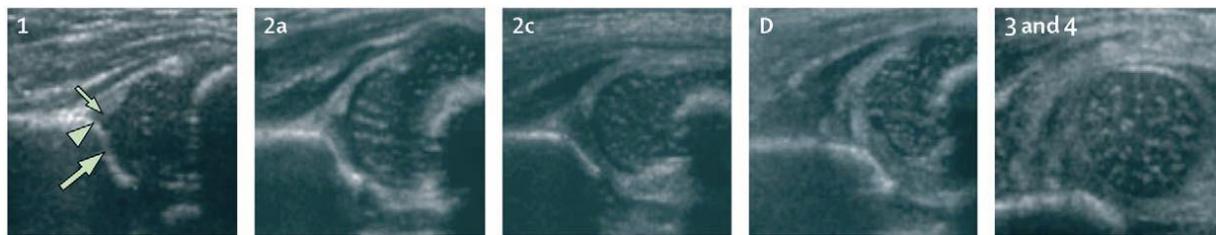
Figure 1. Hip types (1–4) in Graf method (Dezateux & Rosendahl, 2007)

When the same sonographic section passing through the hip joint is used in the same plane, the Graff method can be reproducible. For hip ultrasound, these sections are as follows: The lower limb of the bony ilium in the depth of the acetabular fossa, mid-section of the bony acetabular roof and the acetabular labrum. A sonogram will not be helpful and thus it cannot be used for diagnosis of the condition when one of these points is either missing or unclear. For ultrasound purposes, the lower limb of the os ilium is the center of the acetabulum. In addition, it becomes impossible for the sectional plane to pass through the center of the acetabulum if this landmark is not seen on ultrasound images. Even if the acetabular labrum and the sectional plane are correctly shown, diagnosis may not be made in a centered hip joint (Graf, 2006).

When the Graf method is examined, it is understood that the acquisition of the standard plane angle is required for the diagnosis, which is primarily proportional to placing the patient at the right angle and using the correct ultrasound probe angle. In addition, the skill and knowledge of the expert is very important for this process. The

development of a computer-aided and automated standard plane image detection system may reduce the influence of variables on DDH diagnosis.

2. Literature Survey

In the current study, a software based on deep learning methods (DLM) was developed to capture standard plane frames in ultrasound images. The system aims to help the expert by capturing the standard plane frame from real-time ultrasound images. This study aims to help the diagnosis of DDH by determining the correct diagnostic image using DLMs over real-time 2D ultrasound images. It is seen that DLMs have been used frequently by researchers in the diagnosis, treatment planning or visualization stages of DDH.

Golan et al. conducted a study in 2016 that aimed to provide the fully automating of Graf's Method using deep convolutional neural network (CNN). The angle values of the Graf method were determined by segmenting the ilium and acetabular roof areas in the two-dimensional images. It was reported that there was less than 5% discrepancy between 77% of the test results obtained with the proposed deep architecture and the estimates made by the expert (Golan, Donner, Mansi, Jaremko, & Ramachandran, 2016).

In the study conducted by Paserin et al. in 2017, the researchers proposed near real-time classification process for 3D scans using CNN. It was aimed to determine whether the scans obtained based on the classification process were adequate or inadequate for DDH diagnosis. In their study, they designed a 12-layer CNN architecture using a ReLu activation function with 3 convolutions and 3 fully connected layers. They stated that the fact that the application was not full real-time was a limitation in their study and clinical studies were to be conducted in their subsequent studies. They also indicated that the study had limitations since it was not a real time application, and further stated that they would conduct tests in their follow-up clinical study (Paserin, Mulpuri, Cooper, Hodgson, & Abugharbieh, 2017).

Hareendranathan et al. proposed a method to automatically segment the acetabular bone based on training a CNN network using multi-scale super-pixels. They tested this method on 2D ultrasound images of 50 infant hips and the root mean square error was obtained as 1.8 ± 0.7 mm when compared to manual segmentation. They argued that the proposed method could be used for accurately classifying normal vs. dysplastic hips and automatic diagnosis of hip dysplasia in infants (Hareendranathan et al., 2017).

In another study conducted by Paserin et al. in 2018, the researchers performed a near real-time classification of 3D ultrasound scans using transfer learning approach in CNN networks. They implemented Squeezenet in transfer learning and demonstrated that their approach achieved 93% classification rate on 40 datasets taken from 15 pediatric patients (Paserin, Mulpuri, Cooper, Abugharbieh, & Hodgson, 2018).

In a study conducted in 2018, Tang et al. aimed to classify images at a sufficient level for diagnosis of DDH using the Region Proposal Network (RPN) method. They used the VGG-16 model and 3D U-Net based architecture to learn convolutional features of the backbone. The average Intersection over union (IoU) was obtained as 0.709 (Tang, Zhang, Cobzas, Jagersand, & Jaremko, 2018).

In another study, Pasarin et al. aimed to rapidly and automatically detect the accurate scan for DDH diagnosis using Long Short-Term Memory (LSTM) based CNN and Recurrent Neural Network (RNN) architectures. They reported to achieve 82% accuracy using 200 3D US volumes taken from 25 pediatric patients where each runtime was performed under 2 seconds (Paserin, Mulpuri, Cooper, Hodgson, & Garbi, 2018).

In 2020, Chen et al. compared two different methods proposed for femur and acetabular roof segmentation in 2D ultrasound images. In the first method, mean filtering, morphological processing and least squares operation were used while in the second method, a CNN named FNet was utilized. In conclusion, the proposed deep neural network architecture method was found to provide better segmentation than other methods (Chen et al., 2020).

When the literature is examined; Researchers have carried out many studies on ultrasound images for the diagnosis of DDH, using various segmentation methods. They also focused on Deep Neural Networks (DNN) methods for the classification of these images. However, no study has been found in the literature to capture the image that can make an accurate diagnosis from flowing ultrasound images. For this reason, this study will fill a large gap in the literature.

In this study, it was aimed to detect the standard plane to be used in DDH diagnosis from real-time ultrasound images. Pre-trained networks were also utilized in order to increase the classification success of the system in the application realized through the deep learning method and You Only Look Once (YOLO) object recognition

infrastructure. In addition, models created by combining SqueezeNet, VGG16, VGG19, ResNet50 and ResNet101 pre-trained CNN networks with YOLO were tested and their performance was discussed. The most important and first step to make the correct diagnosis of DDH is to capture the right frame. It is thought that the success of the diagnosis of DDH will be increased thanks to the study carried out.

3. Material and Method

Deep Neural Networks is an artificial neural network model with one or more layers and is applied extensively in image classification, segmentation and object detection studies (Qassim, Verma, & Feinzimer, 2018). It was introduced by (Fukushima, 1980) and improved by (LeCun, Bottou, Bengio, & Haffner, 1998) to be used in various fields and studies as summarized in (Ciresan, Meier, Gambardella, & Schmidhuber, 2011; Cireşan, Meier, Gambardella, & Schmidhuber, 2010) and is known to achieve higher success rates depending on the increased number of layers as structured into two modules. Feature extraction is performed first by detecting the distinguishable features across training images which is then followed by analysis of the extracted features for classification of images into an image category (Rhu, Gimelshein, Clemons, Zulfiqar, & Keckler, 2016).

The higher classification success rate achieved by Hinton's team in ImageNet competition in 2012 increased the interest in DNNs. 26.1% classification success of ImageNet, which is currently used with the name AlexNet, has been reduced to 15.3% by Hinton's team. The error rates have been further reduced (Krizhevsky, Sutskever, & Hinton, 2012) by the architectures developed in the following years (AlexNet-2012 (Krizhevsky et al., 2012), GoogleNet-2014 (Szegedy et al., 2015), VGGNet-2014 (Simonyan & Zisserman, 2014), ResNet-2015 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2016), SqueezeNet-2016 (Iandola et al., 2016), NasNet-2017 (Krizhevsky, Sutskever, & Hinton, 2017) etc.).

In traditional machine learning methodology, training data and testing data are taken from the same domain, and therefore input feature space and data distribution characteristic are the same, which directly affects the system performance. On the contrary, in some real-world machine learning scenarios where training data is expensive or difficult to obtain, this assumption does not hold. In addition, training processing times of these data are not at acceptable levels for normal users. Therefore, high-performance models (pre-trained networks) that are trained using more easily obtained data from different domains are required to be created. This methodology is called transfer learning (Weiss, Khoshgoftaar, & Wang, 2016). In order to use DNN networks with more effective performance, transfer learning approach is frequently employed using pre-trained networks, some of which are mentioned.

YOLO was presented by Redmon et al. as an approach different from other methods of object detection (Redmon, Divvala, Girshick, & Farhadi, 2016). The methods are used to detect a class for the objects and evaluate it at various locations and scales in the test image, and therefore to determine the class to which the object belongs. This method is complex and difficult to optimize. YOLO aims to detect the object from image pixels, to conduct all processes including bounding box coordinates and class probabilities using a single regression formula (Redmon et al., 2016). As single network architecture is used in YOLO to process the input image and generate the output results, RPNs are used for all object recognition predictions. This feature of the YOLO model significantly increases its operating speed (Wong et al., 2019). When using this system, by employing the YOLO process once, object detection is provided without continuously recalling the classifier to predict the objects that are present and their location. In YOLO, training is conducted on full images and detection performance is directly optimized. It has a simple and extremely fast running structure compared to other methods and thus performs faster object detection in real-time webcams (Redmon et al., 2016).

Additionally, the input image is divided into an S*S grid by YOLO. If an object's center falls into a grid cell, then the detection of that object is conducted by that grid cell. Bounding boxes and confidence scores of those boxes will be predicted by each grid cell. The higher confidence scores indicate higher classification accuracy for the system. The confidence score becomes zero when no object in that cell exists. There are five predictions in each box as x, y, w, h, and confidence predictions (cp). The x and y predictions represent the coordinates, w and h predictions represent the width and height relative to the whole image. The confidence prediction represents whether there is a relationship between the used box and the images to be classified. Each grid cell also predicts a probability, and these probabilities are conditioned for the object (Van Rijthoven, Swiderska-Chadaj, Seeliger, van der Laak, & Ciompi, 2018). YOLO Regression Formula is then S*S*(B*5+C), where S*S is the grid size, B is the bounding boxes, 5 is the prediction number and C is the class probability of the boxes, as depicted in Figure 2.
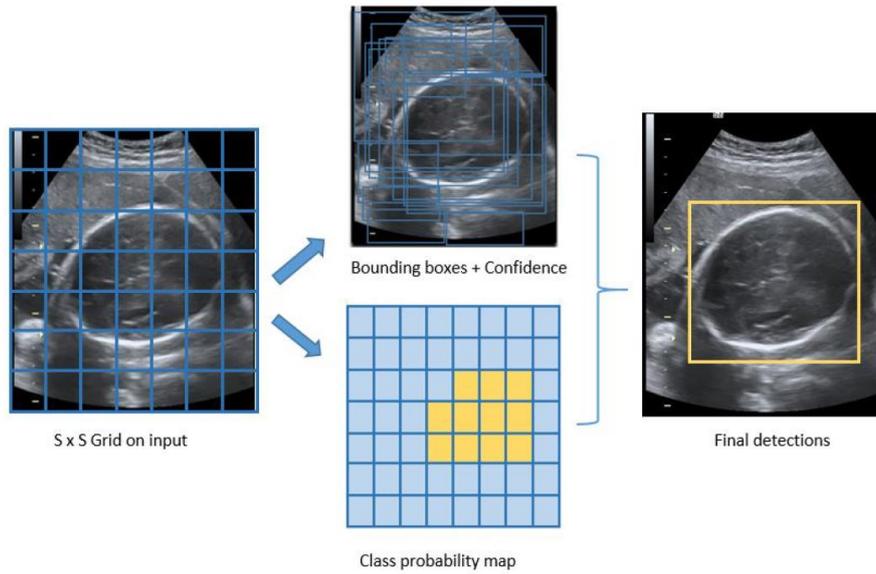
Figure 2. Sample grid and result with YOLO (Irene, Haidi, Faza, & Chandra, 2019)

1000-class ImageNet 2012 dataset is used to train YOLO network. YOLO training results are compared with other real-time detection systems on PASCAL VOC 2007 and VOC 2012. YOLO makes more localization errors when compared to Fast R-CNN, which is the most successful network of VOC 2007 (Redmon et al., 2016). Although Fast R-CNN makes less localization errors, it makes more background errors (Redmon & Farhadi, 2017). When compared to the models in VOC 2012, YOLO scores 8-10% lower than R-CNN and Feature Edit on categories like bottle, sheep, and tv/monitor. However, on other categories like cat and train, YOLO achieves higher performance. When YOLO was combined with R-CNN and, a higher performance was obtained. 2.3% improvement was provided with Fast R-CNN and YOLO combination, boosting it 5 spots up on the leaderboard (Redmon et al., 2016). YOLO architecture is presented in Figure 3.
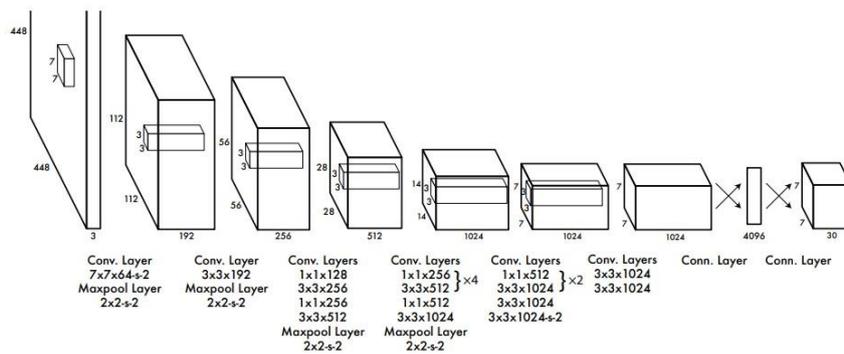
Figure 3. YOLO architecture (Redmon & Farhadi, 2017)

In addition, recall is relatively low in when compared to other RPN-based methods. YOLOV2 was created to improve these shortcomings by maintaining YOLO's classification accuracy. It was obtained by adding a high resolution detector to the YOLO model. It was observed that the shortcomings of YOLO decreased and the success rate increased by 1.8% with the addition of a high resolution detector (Redmon & Farhadi, 2017).

3.1. Proposed Method

The flow chart of the method proposed in this study is shown in Figure 4. The main and most basic stage of the study is the acquisition of images used for diagnosis of DDH. At this stage, the patient records (retrospectively) obtained from Bagcilar Training and Research Hospital (Istanbul, Turkey) were used. For training phase of the system, 140 single-frame images in which the specialist physician performed the diagnosis were selected. For testing the system, 8 of the 5-10 second videos containing the recording made during the diagnosis process were obtained. An ultrasound system (Toshiba Aplio 400) was used to obtain single frame images and videos. All images had 800*600 resolution and were in RGB color mode. The single frame images were saved in JPEG format and the videos were stored in MP4 format. In Figure 5, sample images in the data set are given.

1018
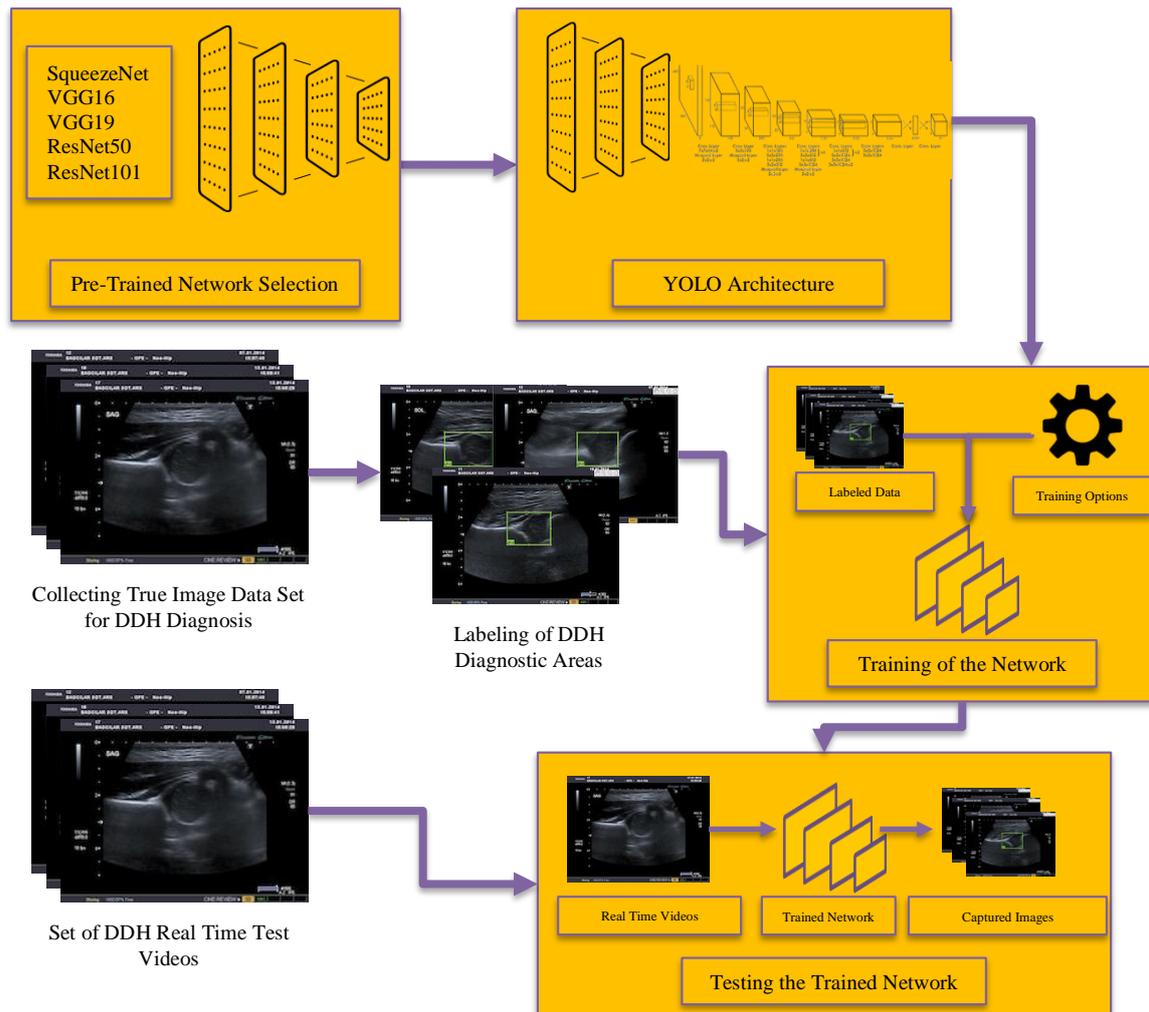
Figure 4. Flow chart of the designed system

Figure 5. Examples of the images used in study

In the second stage, the obtained data set was labeled. The sections used for diagnosis of the training images used for DDH were labeled and verified under the supervision of a specialist physician. During the test process, the similarity of these sections in the streaming image was used. An example labeling process that was performed with Matlab Image Labeler (Matlab, 2020) is shown in Figure 6.

Figure 6. Image labeling process

After the data were labeled and prepared, the data were trained with DNN using the software developed within the scope of the study. Some pre-trained networks that have been previously tested and validated were used for DNN design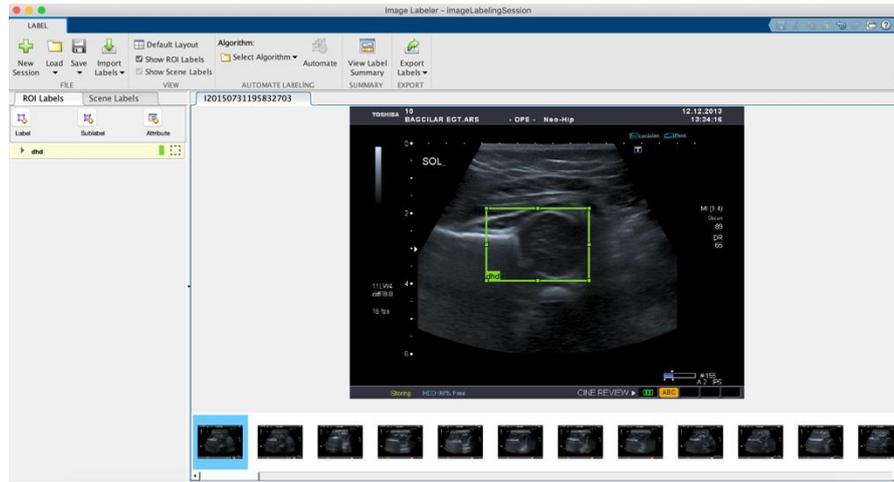. These are YOLO architecture based DNN networks such as SqueezeNet, VGG16, VGG19, ResNet50 and ResNet101 pre-trained networks.

The input data of all these networks were set to 224*224*3. For each image, the DDH value with a single class was given as the output data. The prepared deep networks were subjected to the training process with the same parameters. These settings are given in Table 1.

Table 1. Training parameters for the proposed methods

OPTION	VALUE
NUMBER OF EPOCHS	100
BATCH SIZE	16
LEARNING RATE	0.001
OPTIMIZER	SGDM
INPUT SIZE	224*224*3
OUTPUT SIZE	1 (DDH)
VERBOSE	True
VERBOSE FREQUENCY	1
SHUFFLE	Never

LOSS values of the model created during training are plotted in Figure 7, for different types backbone architectures. The change of LOSS values indicate how the model's estimate is different from the real value. In figure, the horizontal axis indicates the number of iterations and the vertical axis indicates the values.
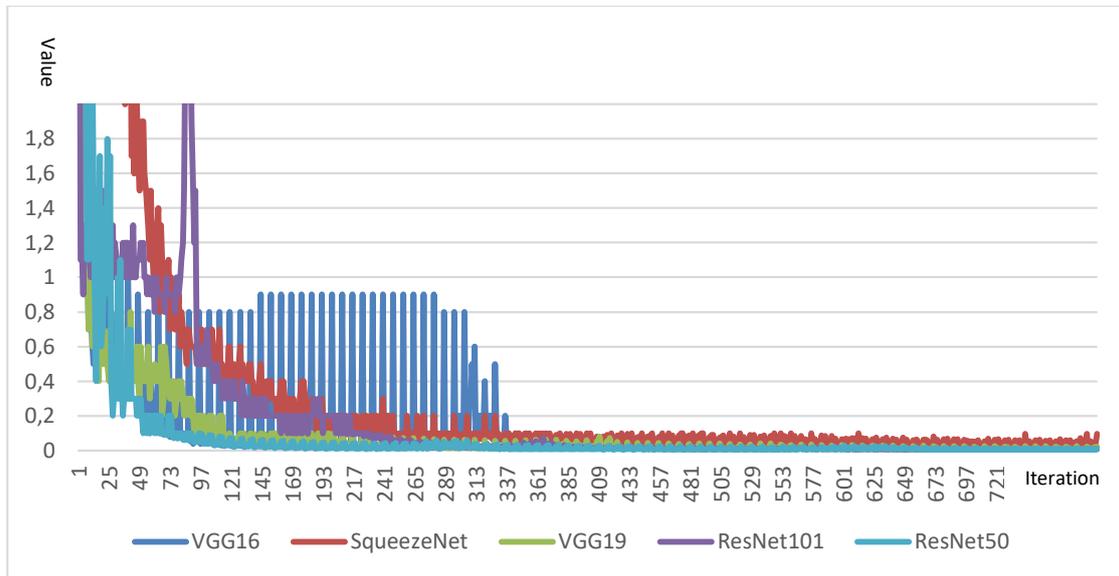
Figure 7. LOSS values of ResNet50, ResNet101, VGG16, VGG19 and SqueezeNet architectures during training

In addition, training processes were completed at different times depending on the structure of the network. Training times depending on the network structure and the minimum Root Mean Square Error (RMSE) and LOSS values obtained by each network are shown in Table 2.

Table 2. Training time for the proposed methods

BACKBONES	TRAINING TIME (HH:MM:SS)	MIN RMSE	MIN LOSS
RESNET50	02:15:03	0.06	0.0037
RESNET101	04:21:11	0.07	0.0056
VGG16	03:37:04	0.06	0.0036
VGG19	04:20:07	0.07	0.0044
SQUEEZENET	01:05:08	0.14	0.0190

In the last stage, the trained DNNs were tested with real-time videos that were in the data set but were not previously presented to the system. The single-frame images captured according to a certain DDH similarity ratio (for example, 0.9 and above treshold) of the trained network in the streaming image were re-evaluated by the expert and it was decided whether these images could be used for DDH diagnosis. In Figure 8, the images with a similarity value of over 0.9 obtained during the test process are given.
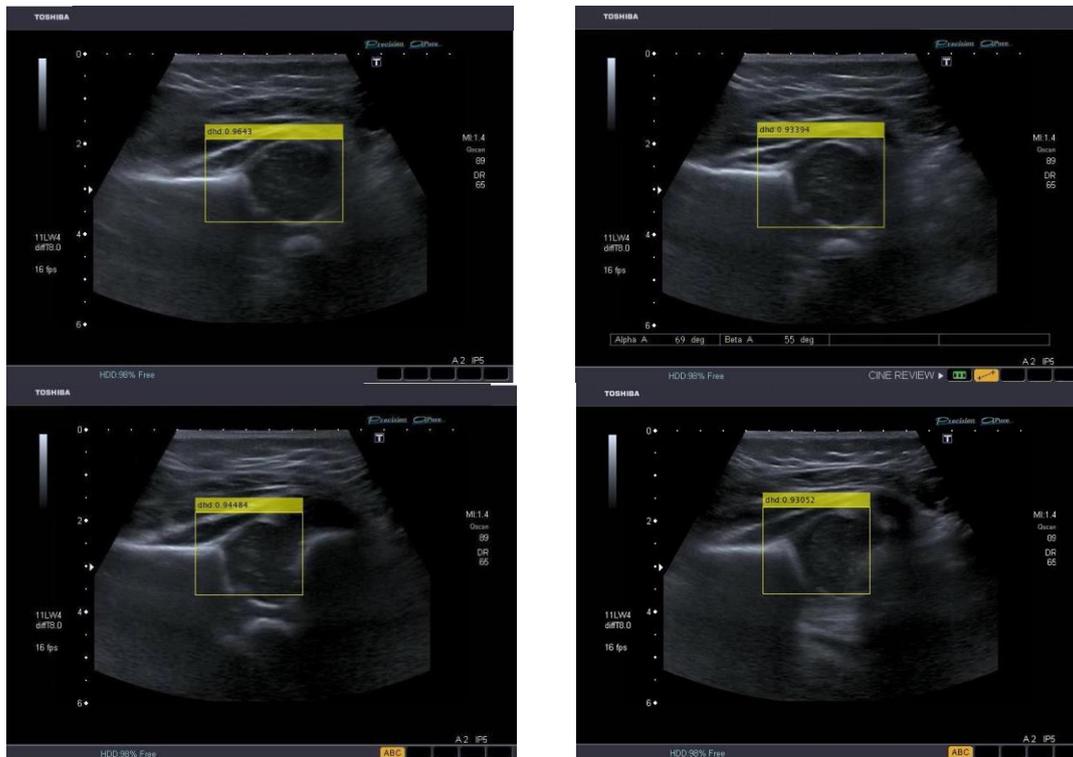
Figure 8. The frames determined as a result of the test process

4. Experimental Results

Video tests were conducted on YOLO architectures powered by pre-trained networks designed within the scope of the study. For testing with 8 videos containing the image sets previously used by the specialist physician for DDH diagnosis were selected. In these videos, there were frames that were appropriate for more than one diagnosis. The specialist doctor had performed the diagnosis from the frame that could be captured manually from the video. With the designed DNN model, the frames that could be used for DDH diagnosis in the video were determined. These obtained frames were then examined by the specialist physician and it was determined whether they could be used in the diagnosis of DDH.

Intersection over Union (IoU) evaluation metric is used to evaluate the results of models working on object detection from images. In this study, IoU metric was used to determine the correct image. A metric that helps measure whether the region proposition suggested by the model are acceptable for the location of the object in the image. The IoU evaluates the overlap between the predicted region and the actual reference value. It is formulated as shown in Equation 1 and Figure 9 (Gamage, Wijesinghe, & Perera, 2019)

$$IoU = \frac{area\ (reference\ value\ \cap\ predicted)}{area\ (reference\ value\ \cup\ predicted)} \qquad (1)$$
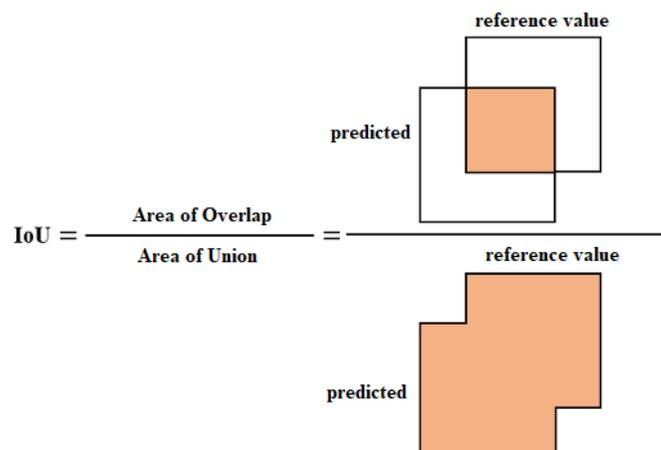

Figure 9. IoU formula

In determining the frames, the threshold rate for IoU was chosen as 0.9 and above. For each network, a set of frames with 0.9 or more threshold for IoU in DDH detection was extracted from each video. The obtained values and expert opinions are shown in Table 3.

Table 3. Test results in real-time videos for detection standard plane

VIDEO	TIME (MINUTE)	FRAME	BACKBONES	0.9 OR MORE IOU FRAMES	0.93 OR MORE IOU FRAMES	TOTAL TEST TIME (SECOND)
VIDEO (1).MP4	9	175	ResNet50	42	32	33.25
			Resnet101	18	4	44.28
			VGG16	35	21	42.82
			VGG19	40	4	51.39
			SqueezeNet	23	5	26.72
VIDEO (2).MP4	10	188	ResNet50	21	4	29.41
			Resnet101	58	17	44.7
			VGG16	21	6	41.18
			VGG19	38	10	48.88
			SqueezeNet	19	1	22.44
VIDEO (3).MP4	10	188	ResNet50	11	1	32.76
			Resnet101	61	18	44.88
			VGG16	22	5	42.71
			VGG19	56	4	47.93
			SqueezeNet	41	26	24.45
VIDEO (4).MP4	10	188	ResNet50	45	21	30.57
			Resnet101	2	1	45.83
			VGG16	55	30	42.42
			VGG19	35	15	48.4
			SqueezeNet	64	55	24.76
VIDEO (5).MP4	10	188	ResNet50	85	54	30.37
			Resnet101	31	4	45.83
			VGG16	64	34	42.65
			VGG19	90	45	48.94
			SqueezeNet	42	15	24.55
VIDEO (6).MP4	10	188	ResNet50	75	48	32.47
			Resnet101	57	4	44.92
			VGG16	92	41	45.05
			VGG19	96	38	51.55
			SqueezeNet	61	14	25.31
VIDEO (7).MP4	7	128	ResNet50	3	1	21.22
			Resnet101	12	1	30.4
			VGG16	27	11	29.71
			VGG19	17	1	33.65
			SqueezeNet	64	57	17.65
VIDEO (8).MP4	4	79	ResNet50	18	5	13.13
			Resnet101	23	8	19.04
			VGG16	12	1	18.11
			VGG19	28	13	20.84
			SqueezeNet	2	1	10.56

When Table 4 is examined, it is seen that the total number of frames of 8 videos is 1322, and the number of frames that can be used in DDH diagnosis from each treshold rate (0.9 and 0.93) is respectively 1606 and 676. When the test process durations were examined, it was found that the processing of a frame was 1.68 seconds for ResNet50, 2.41 seconds for Resnet101, 2.30 seconds for VGG16, 2.65 seconds for VGG19, and 1.33 seconds for SqueezeNet. Tests were performed on data sets with fewer number of frames having a value of 0.93. The results obtained as a result of examining the obtained DDH diagnosis frames (0.93 or More IoU Frames) are given in Table 4 and Table 5.

Table 4. Video based comparison of results for experts and proposed method

VIDEO	BACKBONES	0.93 OR MORE IOU FRAMES	EXPERT'S CORRECTED FRAME	ACCURACY	AVERAGE ACCURACY
VIDEO (1).MP4	ResNet50	32	31	0.97	0.9747
	Resnet101	4	4	1.00	
	VGG16	21	19	0.90	
	VGG19	4	4	1.00	
	SqueezeNet	5	5	1.00	
VIDEO (2).MP4	ResNet50	4	4	1.00	0.9482
	Resnet101	17	16	0.94	
	VGG16	6	6	1.00	
	VGG19	10	8	0.80	
	SqueezeNet	1	1	1.00	
VIDEO (3).MP4	ResNet50	1	1	1.00	0.8959
	Resnet101	18	15	0.83	
	VGG16	5	4	0.80	
	VGG19	4	4	1.00	
	SqueezeNet	26	22	0.85	
VIDEO (4).MP4	ResNet50	21	20	0.95	0.8911
	Resnet101	1	1	1.00	
	VGG16	30	28	0.93	
	VGG19	15	14	0.93	
	SqueezeNet	55	35	0.64	
VIDEO (5).MP4	ResNet50	54	40	0.74	0.8415
	Resnet101	4	4	1.00	
	VGG16	34	34	1.00	
	VGG19	45	45	1.00	
	SqueezeNet	15	7	0.47	
VIDEO (6).MP4	ResNet50	48	44	0.92	0.9492
	Resnet101	4	4	1.00	
	VGG16	41	38	0.93	
	VGG19	38	37	0.97	
	SqueezeNet	14	13	0.93	
VIDEO (7).MP4	ResNet50	1	1	1.00	0.986
	Resnet101	1	1	1.00	
	VGG16	11	11	1.00	
	VGG19	1	1	1.00	
	SqueezeNet	57	53	0.93	
VIDEO (8).MP4	ResNet50	5	5	1.00	0.9846
	Resnet101	8	8	1.00	
	VGG16	1	1	1.00	
	VGG19	13	12	0.92	
	SqueezeNet	1	1	1.00	

Table 5. Network based comparison of results for experts and proposed method (0.93 Threshold)

BACKBONES	DETECTED FRAME	EXPERT RESULT	ACCURACY
RESNET50	166	146	0.8795
RESNET101	57	53	0.9298
VGG16	149	141	0.9463
VGG19	130	125	0.9615
SQUEEZENET	174	137	0.7874
TOTAL	**676**	**602**	**0.8905**

5. Result and Discussion

In the study, a deep learning-based model is presented to capture the standard plane image from real-time ultrasound images to be used for applying Graf method in diagnosis of DDH. YOLO object recognition infrastructure was used to determine the standard plane within the image. In addition, pre-trained networks were utilized to increase the classification success of the system. The models created by combining SqueezeNet, VGG16, VGG19, ResNet50 and ResNet101 pre-trained CNN networks with YOLO were tested and their performance was discussed.

When the training and test durations of the designed models were examined, it was seen that the fastest system was the SqueezeNet model. When SqueezeNet was used in training, it was seen that the training phase was completed in at least half time shorter and in at most a quarter time shorter than the other networks. In addition, in the testing process, SqueezeNet was found to be much more successful than the other networks. When the training and test error rates were examined, it was determined that the architecture designed with VGG19 produced very little difference success rates compared to other models. However, there was not much difference in the system success. When the number of frames captured in the test process was examined, it was seen that the

VGG19 model captured the accurate frames more than the others.

When the results of the study were evaluated in general, it was observed that the proposed deep learning architectures used in obtaining the adequate and required standard plane for the experts to make the correct diagnosis were quite successful. In future studies, the design of these architectures for user application and the development of systems that assist experts can be handled.

Conflict of Interest

No conflict of interest was declared by the authors.

References

Chen, L., Cui, Y., Song, H., Huang, B., Yang, J., Zhao, D., & Xia, B. (2020). Femoral head segmentation based on improved fully convolutional neural network for ultrasound images. Signal, Image and Video Processing, 1-9.

Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2011). Convolutional neural network committees for handwritten character classification. Paper presented at the 2011 International Conference on Document Analysis and Recognition.

Cireşan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. Neural computation, 22(12), 3207-3220.

Dezateux, C., & Rosendahl, K. (2007). Developmental dysplasia of the hip. The Lancet, 369(9572), 1541-1552.

Fukushima, K. (1980). A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern., 36, 193-202.

Gamage, H., Wijesinghe, W., & Perera, I. (2019). Instance-based segmentation for boundary detection of neuropathic ulcers through Mask-RCNN. Paper presented at the International Conference on Artificial Neural Networks.

Golan, D., Donner, Y., Mansi, C., Jaremko, J., & Ramachandran, M. (2016). Fully automating Graf's method for DDH diagnosis using deep convolutional neural networks. In Deep Learning and Data Labeling for Medical Applications (pp. 130-141): Springer.

Graf, R. (2006). Hip sonography: diagnosis and management of infant hip dysplasia: Springer Science & Business Media.

Hareendranathan, A. R., Zonoobi, D., Mabee, M., Cobzas, D., Punithakumar, K., Noga, M., & Jaremko, J. L. (2017). Toward automatic diagnosis of hip dysplasia from 2D ultrasound. Paper presented at the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. J. a. p. a. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size.

Irene, K., Haidi, H., Faza, N., & Chandra, W. (2019). Fetal Head and Abdomen Measurement Using Convolutional Neural Network, Hough Transform, and Difference of Gaussian Revolved along Elliptical Path (Dogell) Algorithm. arXiv preprint arXiv:1911.06298.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Paper presented at the Advances in neural information processing systems.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Matlab. (2020). Image Labeler. Retrieved from https://www.mathworks.com/help/vision/ug/get-started-with-the-image-labeler.html

Paserin, O., Mulpuri, K., Cooper, A., Abugharbieh, R., & Hodgson, A. J. (2018). Improving 3D ultrasound scan adequacy classification using a three-slice convolutional neural network architecture. CAOS, 2, 152-156.

Paserin, O., Mulpuri, K., Cooper, A., Hodgson, A. J., & Abugharbieh, R. (2017). Automatic near real-time evaluation of 3D ultrasound scan adequacy for developmental dysplasia of the hip. In Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures (pp. 124-132): Springer.

Paserin, O., Mulpuri, K., Cooper, A., Hodgson, A. J., & Garbi, R. (2018). Real time RNN based 3D ultrasound scan adequacy for developmental dysplasia of the hip. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.

Qassim, H., Verma, A., & Feinzimer, D. (2018). Compressed residual-VGG16 CNN model for big data places image recognition. Paper presented at the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Rhu, M., Gimelshein, N., Clemons, J., Zulfiqar, A., & Keckler, S. W. (2016). vDNN: Virtualized deep neural networks for scalable, memory-efficient neural network design. Paper presented at the 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO).

Schams, M., Labruyère, R., Zuse, A., & Walensi, M. (2017). Diagnosing developmental dysplasia of the hip using the Graf ultrasound method: risk and protective factor analysis in 11,820 universally screened newborns. European Journal of Pediatrics, 176(9), 1193-1200.

Shaw, B. A., & Segal, L. S. (2016). Evaluation and referral for developmental dysplasia of the hip in infants. Pediatrics, 138(6).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). Going deeper with convolutions. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Tang, M., Zhang, Z., Cobzas, D., Jagersand, M., & Jaremko, J. L. (2018). Segmentation-by-detection: A cascade network for volumetric medical image segmentation. Paper presented at the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).

Van Rijthoven, M., Swiderska-Chadaj, Z., Seeliger, K., van der Laak, J., & Ciompi, F. (2018). You only look on lymphocytes once.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), 9.

Wong, A., Famuori, M., Shafiee, M. J., Li, F., Chwyl, B., & Chung, J. (2019). YOLO nano: A highly compact you only look once convolutional neural network for object detection. arXiv preprint arXiv:1910.01271.

Yang, S., Zusman, N., Lieberman, E., & Goldstein, R. Y. (2019). Developmental dysplasia of the hip. Pediatrics, 143(1).