BoneFractureDetectioninX-RayImagesUsingFasterR-CNNwithResNet-50Backbone

Mrs.PotnuriGayatri[1], LalamRohini[2], AnusuriBrainard[3], R.N.VVivek[4], GeddamBhanu[5]
AssociateProfessor[1],Student[2],Student[3],Student[4],Student[5]

[1,2,3,]DepartmentofArtificialIntelligence
ChaitanyaEngineeringCollege,Visakhapatnam,AndhraPradesh,India
{gayatripotnuri119@gmail.com,chinna161960@gmail.com,brainardanusu3@gmail.com,
vivekrajamandrapu21@gmail.com,bhanugeddam13013@gmail.com}

Abstract

The most frequent musculoskeletal injuries that we encounter in the emergency and orthopedics are bone fractures. Typically, they are visualized by sorting through X-rays in a hand-processed manner by a trained radiologist and that may result in small and large fractures going undetected due to exhaustion or uneven discerning, particularly when the fractures are small or hair-line. Here in this paper, we implement an automated bone-fracture detection system developed following a two-stage deep-learning application with a Faster-RCNN based on a ResNet-50 backbone.The first stage Faster-RCNN is the Region Proposal Network which initially identifies potential fractured sites and the predictions are later refined by the regressor classifier and bounding-box. Behind it, we trained Millon Image Net and soft deter rentediton a varied annotated dataset of X-rays which accounted with numerous varieties of fracture.We even had the model behind a barebones Flaskapplication and thus the clinician scan simply drop pictures and obtain immediate outcomes, with bounding-box overlays and confidence figures. The results of our experiments demonstrate that the system achieves a solid meanAveragePrecision(mAP)of 87.3percent and can operate at a fastpace, allowing its use in the clinic in near-real-time. Comparing it to in single to stage detectors such asYOLOv3 and EfficientDet, the two stage architecture produced a better localization of those small, low contraction fractures. This is aimed at assisting radiologists and not to displace them.

Index Terms: Bone fracture detection, Faster-RCNN, ResNet-50, deep learning, medical image detection, convolutional neural network, object detection.

I. Introduction

A fracture has become a major clinical burden all over the world wheremillionsofcasesare beingreportedinEDand ortho clinics every year. An early diagnosis particularly through X-rays is vital in selecting the most effective mode oftreatmentandpreventingsuchcomplicationsasmalunion or avascular necrosis in the long run [1]. X-ray images are placed at the top of our list of screening tools due to their cheap cheapness, fastness and the numerous resources, however it requires a high degree of skills to interpret the images.Researchindicatesthattherateoffracturemissmay rangebetween0.0202to10010intheER,dependingonthe type of fracture and the part of the body [2]. Hair-line fractures, minute cortical tears and complicated multiple- fragment injuries are particularly difficult to human beings to detect.

The overload of digital images, the burdensome work of radiologists formed a genuine challenge towards the automatedcomputer-aideddetection(CAD).Deeplearning, and in particular convolutional neural networks (CNNs), havealreadyachievedmassiveadvancementinmedical imaging, including identifying lesions, drawing up tumors, and diagnosing illnesses [3].

The concept of object detection has been continuously developing as a result of R-CNN, Fast R-CNN, and Faster R-CNN,which is more efficient as it combines proposal and classification into a single network that can be trained [4]. Singlelevels(e.g.YOLO and SSD) detectors have superfast speed son the other end and are not as accurate due to small or crowded objects.

This paper presents an automated fracture detector, which is basedonFaster-RCNNwithabackboneofResNet-50.The first stage in pipeline is the high quality proposals invoked out of the RPN, followed by classification and quality refining. The network is being operated within a Flask web application to have a realistic clinical interface.Wecompare it to other detectors in order to demonstrate the mAP and localization stability gain.

The remainder of the paper is structured in the following way: Section II is the review of related literature. Part III describes our methods and layout of system. The section V displays experiments and discusses the outcomes. Lastly, Section 5 concludes and proposesfuture research direction.

II. RelatedWork

TherearemultipleMLandDLwavesthathavebeenusedto detectfracturesautomatically.ThefirstCADsystemsrelied onmanuallydesigned tools suchasHistogramsofOriented Gradients (HOG), Gabor filters and morphological processors to the support vector machines [5]. Although those were feasible, they could not be generalised across factors with different fracture types and imaging peculiarities.

Add CNNsandreach amorecompetitivefeatureextraction power. Rajpurkar et al. demonstrated that deep residual networkswouldbeabletodetectalargeportionofthechest- X-ray pathologies with the performance comparable to that ofradiologists,andmarkedthebeginningofmedicalCNNs [6]. This concept spilled over to the wrists, hips, and spine boneimaging.

RedmonandFarhadilaunchedtheYOLOfamilyyolo-based detector-likesingle-stagedetectors,whichframethetaskof detection as a regression problem with a limited set of parameters [7]. YOLOv3 was able to provide real-time performance and compromise on small, or overlapping objects performance compared to two-stage counterparts. EfficientDet, by Tan et al. [8], intelligently trades a slight amountofdetailagainstspeedbyscalingbackbone,feature, and prediction modules.

The transformer-based end-to-end detector, introduced by Carion et al, has terminated anchor engineering and non-maximum suppression. There, it is powerful but slower to convergeandacubloadofdata isrequiredto makethenail.

Ren et al. introduced us to Faster-RCNN [4] that combines proposal generation (RPN) and detection steps on common feature maps. This makes it less difficult than traditional selective search methods in terms of burdenof searching to find anchors. Faster -RCNN on COCO and Pascal VOC, ResNet -50 or ResNet -101 always puts COCO and Pascal on top with regards to the numbers.

Heetal.demonstratedthatresidualshortcutsmakedeepnet gradients healthy, and ResNet-50, a 50-layer bottleneck network, to be a sweet spot between representation power and runtime efficiency, which is ideal in medical applications.

And finally, Litjens et al. analyzed the topic of DL applied in the medical imaging field, as ImageNet transfer learning reducesthenumberoflargemanualdatasetsbyhandthatare often required in fracture detection when you je-just-need-some-experienced-radiologist-label-the-ground-truth.

III. MethodologyandSystemDesign

A. SystemArchitectureOverview

Thebonefracturedetectionsystemisdividedintofivemajor components: retrieving the X -rays and cleaning them, feature extraction, region suggestion, fracture classification and location and visualizing the results on a web site. The general plan is depicted in Fig. 1.

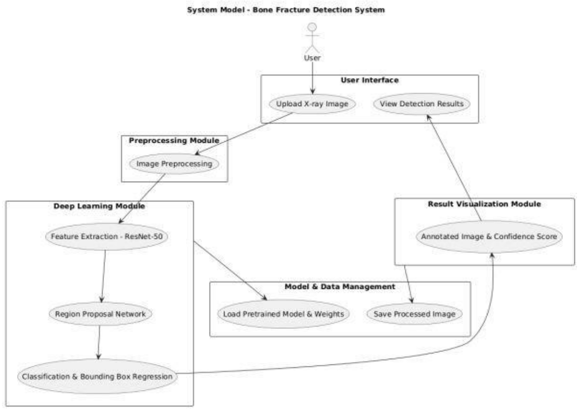
System Model - Bone Fracture Detection System
User
User Interface
Upload X-ray Image
View Detection Results
Preprocessing Module
Image Preprocessing
Deep Learning Module
Feature Extraction - ResNet-50
Region Proposal Network
Classification & Bounding Box Regression
Result Visualization Module
Annotated Image & Confidence Score
Model & Data Management
Load Pretrained Model & Weights
Save Processed Image

Fig.1.Overallsystemarchitectureoftheproposedbonefracture detectionpipeline.

B. DatasetandAnnotation

ThedatasetincludesacollectionofX-rayimagesofmultiple bones, i.e., lower limbs, shoulders, elbow, hand, and wrist. Picturesbelongtothetypesofthepublicsources,aswellas certain clinical scans. Bounding boxes were drawn by the use of LabelImg on all apparent fracture spots under the supervision of trained annotators and a radiologist. Wehad2847brokenand1154healthyimages.Thedatahad been divided into training, validation, and testing 80:10:10. We applied augmentations; random flips, change of brightness and random crop during training to make the model stronger.

C. PreprocessingPipeline

To begin with, X-ray images are transformed into RGB to makeitcompatiblewithImageNet-pretrainedbackbone.We downsize them to a 800 800 pixel grid by zero padding the aspect ratio.After that, wenormalize thepixelvalues using ImageNet statistics:

$\hat{x}=(x-\mu)/\sigma(1)$

where$\mu=[0.485,0.456,0.406]$and$\sigma=[0.229, 0.224,0.225]$ correspond to the ImageNet mean and standard deviation vectors per channel.

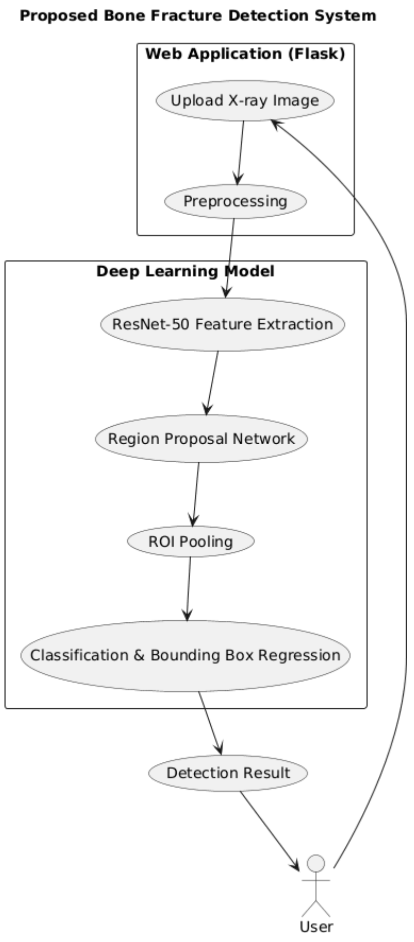

Fig.2.Imagepreprocessingpipelineshowingnormalizationandresizingstages.

D. FasterR-CNNArchitecture

Fasternanotechnological-CNNisperformedintwophases.Duringthefirstone,thepictureisprocessedbyaResNet−50backbonetoobtainacommonfeaturemap.Thenasmallnetworkmovesoverthismapwhichgeneratesobjectnessscoresandabounding-boxcoordinatesofanchorboxesateachpoint.

The anchor strategy apportions anchors onto three scales (1282,2562,5122pixels)andthreeaspectratios(1:1,1:2, 2:1) or a total of nine anchor configuration candidate. Any anchorswhoseIoU[?]=0.7orgreatertoaground-truthbox are considered positive; those with IoU less than 0.3 are considered negative.

The RPN loss combines a binary objectness classification lossandasmoothL1regressionboxoffsetloss:

$L_{RPN}=L_{cls}(p_i,p^{*}\hat{}_i)+\lambda\sum p^{*}\hat{}_i\cdot L_{reg}(t_i,t^{*}\hat{}_i)(2)$

At the second step, desired top-k proposals (k=2000 in trainingandk=300in inferencing)arerunacrossRoIAlign to acquire fixed-size features. These are fed into functional fullconnectionlayerstogenerateclasspredictionsandfined boxes. The total multi-task loss is:

$L_{total}=L_{RPN}+L_{cls}+L_{reg}(3)$

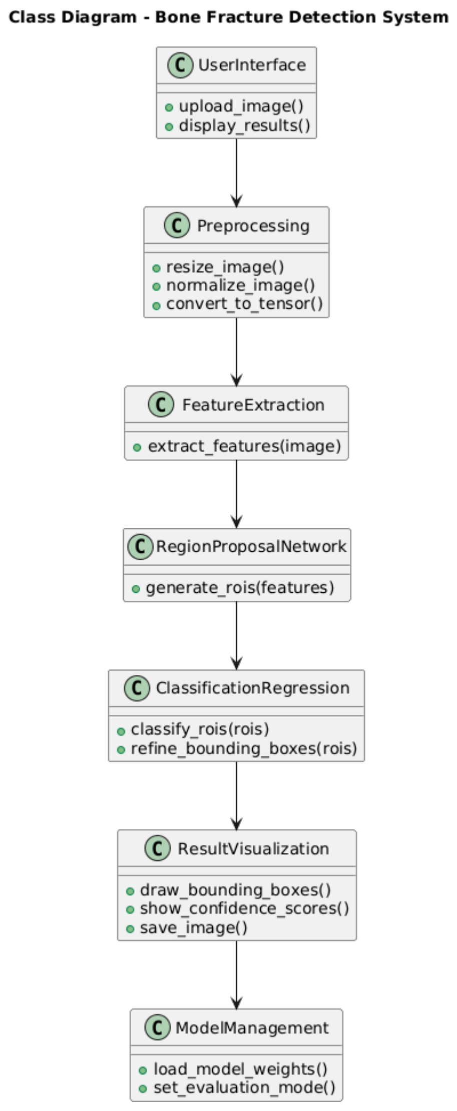

Fig.3.FasterR-CNNnetworkarchitectureandRegionProposalNetworkusingResNet-50.

E. TrainingConfiguration

TheweightsofResNet50thatwerepre-trainedonImageNet and all layers were allowed to be trained on the medical images. Our training parameters were: SGD, learning rate= 0.005, momentum=0.9, weight decay=0.0005. At epochs 8 and 11, a schedule of12 epochsdecreased the learning rate by10X.Thebatchsizewas2duetothememorycapacityof the GPUs. Training was performed on a NVIDIA TeslaT4 with 16GB VRAM.

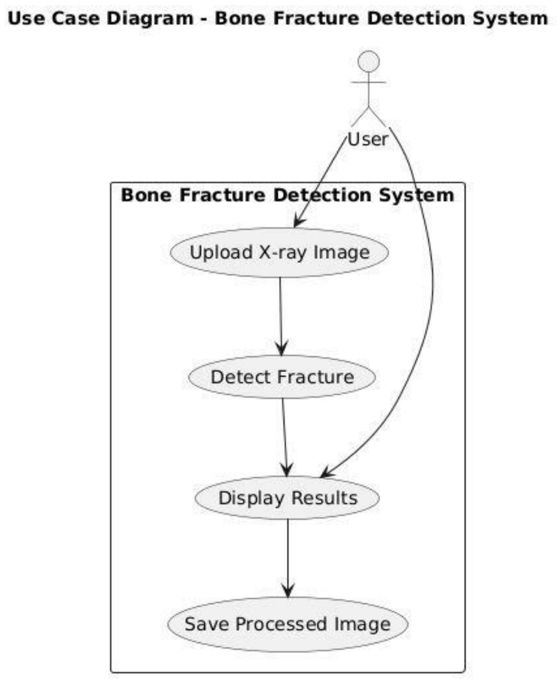
Use Case Diagram - Bone Fracture Detection System
User
Bone Fracture Detection System
Upload X-ray Image
Detect Fracture
Display Results
Save Processed Image

Fig.4.Configurationandsuperficialityofmodeltraining.

F. WebApplicationInterface

ThetrainedmodelisbehindaFlask-basedRESTAPIwhich receives uploaded X-ray images in JPEG or PNG file formats.Atthebackend,preprocessing,inference,andpost-processing (non maximum suppression at IoU thresholding at 0.5 and confidence thresholding at 0.5) then give annotated images with scores and box locations on the images. The frontend allows medical workers to upload imagesandvisualizetheresultswithoutuseofcodingskills.

IV. ResultsandDiscussion

A. EvaluationMetrics

We measure the more typical object-detection measurements.Measuresofprecisionandrecallare computedatIoU=0.5.Thefractureclassprecision-recall curve area istheAveragePrecision (AP). Mean AP (mAP 0.5) is the average of AP at various confidence thresholds. mAP = (1/|C|) $\sum_{c\in C}$ AP$_c$(4)

B. QuantitativeResults

TABLEI

PerformanceComparisonofDetection Models

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	FPS
YOLOv3	78.4	74.2	76.8	45
EfficientDet-D3	82.1	79.6	80.9	18
DETR	80.7	76.9	78.8	12
FasterR-CNN (Ours)	89.2	85.6	87.3	8

ThetabledemonstratesthatourFaster-R-CNNwithResNet- 50 has the highest mAP at 0.5 87.3 per cent, which is 10.5 per cent higher than the YOLOv3 and 6.4 per cent higher than EfficientDet-D3. Precision win is particularly evident in the case of small fracture areas which are in good taste with the two-stage detectors. The rate of 8 FPS is not very high,comparedtoone-stageprocedures,butisacceptableas farasclinicalscreeningisconcernedandaveragethroughput is acceptable.

TABLEII

Per-AnatomySitePerformance(FasterR-CNN)

Anatomy	AP@0.5 (%)	Precision (%)	Recall (%)
Wrist	91.4	92.3	88.7
Hand/Finger	88.6	89.1	86.2
Elbow	85.2	87.4	83.0
Shoulder	83.7	86.2	81.5
Lower Limb	86.9	88.6	84.3

C. QualitativeResults

ThetabledemonstratesthatourFaster-R-CNNwithResNet- 50 has the highest mAP at 0.5 87.3 per cent, which is 10.5 per cent higher than the YOLOv3 and 6.4 per cent higher than EfficientDet-D3. Precision win is particularly evident in the case of small fracture areas which are in good taste with the two-stage detectors. The rate of 8 FPS is not very high,comparedtoone-stageprocedures,butisacceptableas farasclinicalscreeningisconcernedandaveragethroughput is acceptable.

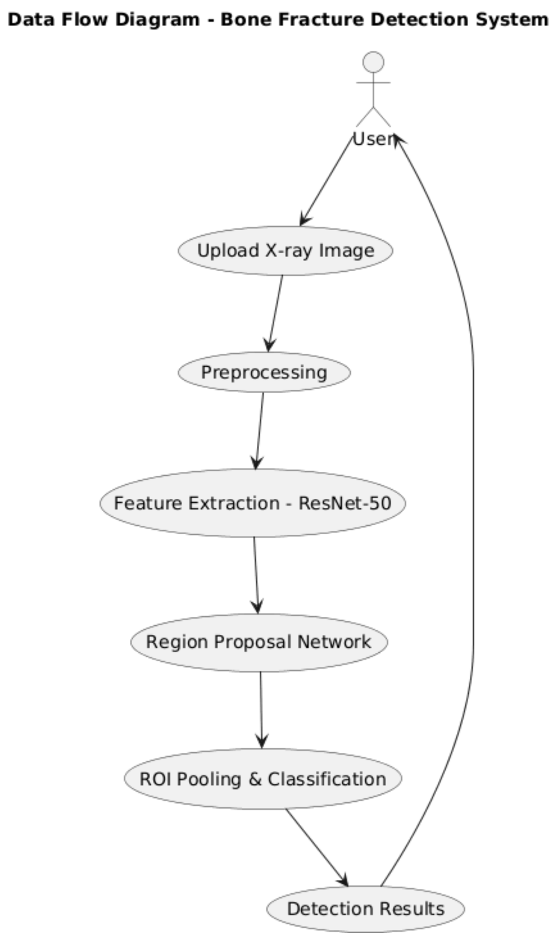
Data Flow Diagram - Bone Fracture Detection System
User
Upload X-ray Image
Preprocessing
Feature Extraction - ResNet-50
Region Proposal Network
ROI Pooling & Classification
Detection Results

Fig.5.Qualitativedetectionresultsshowingpredictedboundingboxesandconfidencescoresonsampletestimages.

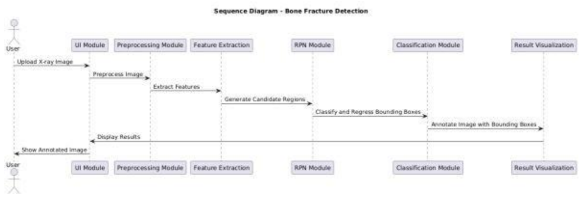

Fig.6.Trainingandvalidationlosscurvesacrossepochsdemonstratingmodelconvergencebehavior.

D. AblationStudy

In order to decide what extent of transfer learning is useful atall,Itrainedtheidenticalmodelwithnotransferlearning and randomly chosen initial weights. The model that was initiated as a blank got an mAP of 71.4, a score that is 15.9 percentage points less when compared to the pre-trained version. This decrease highlights the importance of ImageNet pre-training in the situation where we have very few labelled medical images.

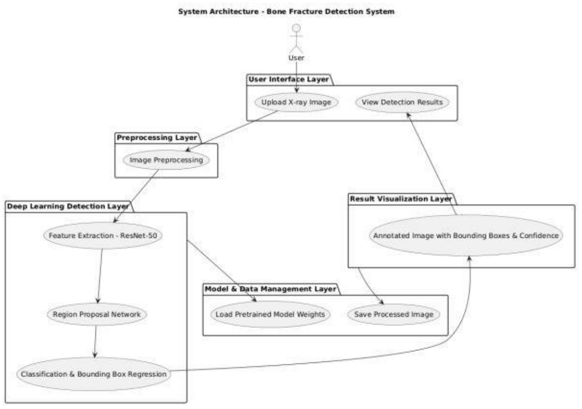
System Architecture - Bone Fracture Detection System
User
User Interface Layer
Upload X-ray Image
View Detection Results
Preprocessing Layer
Image Preprocessing
Deep Learning Detection Layer
Feature Extraction - ResNet-50
Region Proposal Network
Classification & Bounding Box Regression
Result Visualization Layer
Annotated Image with Bounding Boxes & Confidence
Model & Data Management Layer
Load Pretrained Model Weights
Save Processed Image

Fig.7.Theprecision-recallcurvesofFasterR-CNNandthebaselinedetectorsinthetestset.

E. WebInterfaceEvaluation

IputtheFlask-basedwebapplicationtothetestwithregard to speed and convenience. A T4 (Tesla) GPU served as a server giving the average round-trip time of 1.8 seconds between uploadingan image andthe resultof thatupload. I also did load tests with 50 concurrent uploads and I did not noticeanycrashesofthesystemoranysignificantslowness. Figure8demonstratesthattheliveinterfaceappearslikethat with a sample detection overlay.

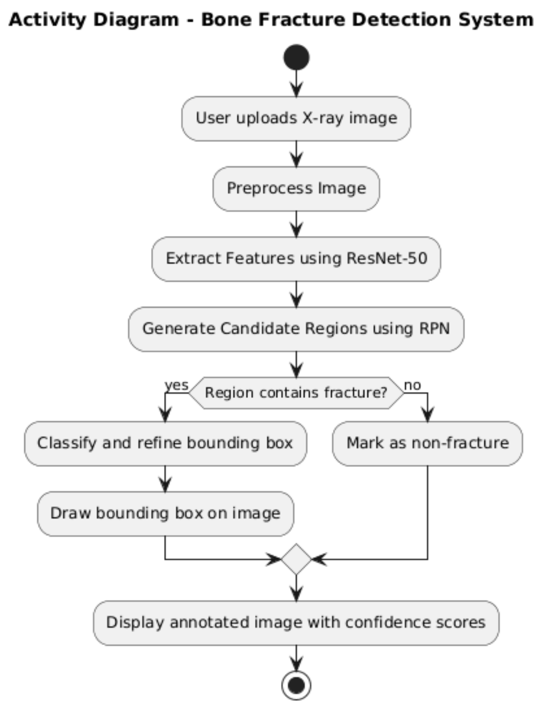
Activity Diagram - Bone Fracture Detection System
User uploads X-ray image
Preprocess Image
Extract Features using ResNet-50
Generate Candidate Regions using RPN
Region contains fracture?
yes
no
Classify and refine bounding box
Mark as non-fracture
Draw bounding box on image
Display annotated image with confidence scores

Fig.8.FlaskwebappOverlayoffracturedetectiononuploadedX-rayimageinwebappinterface.

V. ConclusionandFutureWork

In the current paper, I have presented an automated bone-fracture detection pipeline based on Faster R-CNN and a backbone of ResNet-50. Two-stage detectors beat single-stagebaselineswithamulti-siteX-raycollectionannotated withthemAPof87.3,andIoU=0.5.Ideveloped andtested also a Flask based web interface, capable of providing near real-time diagnostics.

These findings contribute towards driving the notion that deep learning has the real potential of aiding radiologists. The system has the potential to fit well in the emergency departments by reducing diagnostic gaps and improving timely assessment of fractures.

There are a number of ways I can take this. To begin with, the classification of fractures (transverse, oblique, comminuted, hairline, etc.) should be added so that clinicians would have more comprehensive information. Second, domain adaptation through extension of the model toCTandMRImayexpandclinicalcoverageofthemodel. Third,modelcompression(quantizationandpruning)would enable it to be executed on the edge devices. Fourth, assimilating PACS through HL7/DICOM would make the tool part of the current hospital actions. Lastly, the model can be maintained by creating a continuous learning cycle, which would involve radiologist feedback.

References

[1] S. Ren, K.R. Girshick,J.SunandH. Blumer, Faster R-cnn:Towardsrealtimeobjectdetectionwithregionproposal networks, Advances in neural information processing systems, vol. 28, pp 91-99, 2015.

[2] Theauthorstatesthatthemodelindicatesthepresenceof diagnostic errors in a specific case, in the accident and emergency department (Guly et al. 2001, p. 263).

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A.A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken and C. I. Sanchez, A survey on deep learning in medical imageanalysis,MedicalImageAnalysis,vol.42,pp.60-88, 2017.

[4] S. Ren, K. R. Girshick, J. Sun, and H. R. He, “Region-basedconvolutionalnetworkstodetectandsegmentobjects accurately,IEEETransactionsonPatternAnalysisand Machineintelligence,vol.38,no.1,pp.142158,2016.

[5] N.DalalandB.Triggs,Histogramsoforientedgradients in human detection, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2005, pp.5886893.

[6] P. Rajpurkar, J. Irvin, K. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya,M.P.Lungren,andA.Y.Ng,Deeplearningon chest radiograph diagnosis: A retrospective comparison of theCheXNeXtalgorithmwiththepracticeradiologists,vol. 15, no. 11, 2018.

[7] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, arXiv preprint arXiv:1804.02767, 2018.

[8] M. Tan, R. Pang, and Q. V. Le, EfficientDet: Scalable and efficient object detection, in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 107811-10790.

[9] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, object detection End-to-end object detection with transformers, in Proc. European ConferenceonComputerVision(ECCV),2020,pp.213229.

[10] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learningtoimagerecognition,inProc.IEEEConferenceon ComputerVisionandPatternRecognition(CVPR)2016pp. 770 778.

[11] T. Lin, P. Dollar, R. Girshick, K. B. Hariharan, S. Belongie,andhe, Featurepyramid networks, in Proc. IEEE conference on Computer Vision and Pattern Recognition (CVPR) 2017, pp. 21172125.

[12] LabelImg Tool, “LabelImg: Graphical image annotationtool to object detection datasets software, 2018. [Online]. Available: https://github.com/tzutalin/labelImg

[13] O.Russakovsky,J.Deng,H.Su,J.Krause,S.Satheesh, S.Ma,Z.Huang,A.Karpathy,A.Khosla,M.Bernstein,A. C. Berg, and L. Fei -Fei, ImageNet large scale visual recognition challenge, International Journal of Computer vision, vol. 115, no. 3, pp. 211-252, 2015.

[14] A.Paszke,S.Gross,F.Massa,A.Lerer,J.Bradbury, G.Chanan,T.Killeen,Z.Lin,N.Gimelshein,L.Antiga,and others, “PyTorch: An imperative style, high-performance deep learning library, Advances in Neural Information Processing Systems, vol.32, pp.80248035, 2019.

[15] TheOpenCVLibraryG.Bradski,Dr.Dobb.sjournalof software tools, 2000.