Performance elevation using Augmented Pivot Point Rotation for Kidney Stone Detection

Gorli Santoshi[1][0000−0002−4003−9538] and Ratnakar Dash[2][0000−0001−9886−2546]

[1] National Institute of Technology, Rourkela
iacr.santoshi@gmail.com
[2] National Institute of Technology, Rourkela
ratnakar@nitrkl.ac.in

Abstract. Kidney stone detection has been one of the key issues of healthcare professionals in the past. The recent development of deep learning-based models for kidney stone detection has reduced the time and workload of radiologists by assisting in the classification of kidney stone images. The contribution focuses firstly on generating the annotation of the publicly available dataset consisting of 1799 Non-Contrast Computerized Tomography(NCCT) coronal images collected from GitHub. Without manipulating the ratio of the training and testing samples, annotation of the images were carried out with bounding box instances of normal and kidney stone. A competent algorithm generates a new dataset using the augmented pivot point rotation(APPR) to the bounding box. The original and augmented datasets are trained on Single shot Detector(SSD), You Only Live Once(YOLOv7), and Faster Region-based Convolution Neural Network(RCNN) with backbones such as ResNet50, MobileNetv2, and ResNet101, and the results are compared. The result gives a trade-off between the single-stage and two-stage object detection models. The precision of YOLOv7 is 0.986 and 0.966 for normal and kidney stones, respectively, but the precision of the Faster RCNN for normal and kidney stones is balanced. Faster-RCNN training parameters are more compared to YOLOv7, resulting in an increase in training time. YOLOv7 surpasses the outcomes compared to other models with a mAP@0.5:0.95 of 0.933. An average mAP@0.5:0.95 scores for all the models trained on the augmented dataset is intensified by 18.9%. YOLOv7 and Faster-RCNN with ResNet50 provide promising results after training on augmented data. It is concluded that YOLOv7 and Faster RCNN with ResNet50 are suitable for localizing kidney stones with the proposed augmentation technique.

Keywords: Kidney stone detection· object detection · deep learning · augmentation.

1 Introduction

Medical image analysis-based interpretation is dominantly used for early detection and diagnosis of diseases. Deep learning algorithms can automatically learn complex patterns and features from the images themselves, and they are especially well-suited for medical image analysis. Based on imaging data, these algorithms can segment anatomical components, detect and classify anomalies, and forecast clinical consequences.

Kidney stone symptoms can overlap with other diseases, including appendicitis or urinary tract infections. The exact diagnosis and suitable treatment are made possible by the proper detection of kidney stones by imaging examinations[1], such as CT scans or ultrasound, which helps distinguish kidney stones from other possible causes of symptoms. Automation algorithms assist healthcare professionals in early detection, potentially saving time and improving patient outcomes.

The conventional classification technique[2] guides healthcare professionals by determining the entire image belonging to a particular class. Object detection, conversely, oblige is detecting and classifying all the objects in the image with bounding boxes. A fully automated system will be effective if the model can predict the location of the kidney having the stone. A person can have one or two kidneys according to their health conditions. Stone can be present in both kidneys or only in one or none(normal). Localization of the renal calculi reduces the burden of the radiologist.

Medical professionals are under pressure to evaluate medical images as their use increases. Due to its excellent accuracy relative to conventional methods, deep learning has recently been increasingly used to support physicians in many medical imaging tasks. Deep learning performance often depends on the volume of training data[3]. However, compared to other areas, the availability of medical picture datasets[4] is typically constrained because of the production of the image ground truth and confidentiality information. Since there are typically fewer samples of medical anomalies than real-time data, numerous strategies to produce new positive samples have been proposed and employed to produce more training samples. However, creating a compelling data augmentation understanding[5] of the target dataset and its operation is necessary.

The objectives of the proposed work are:

- Fully automating the localization and classification of normal and kidney stone regions with bounding boxes.
- Comparison of standard deep learning object detection algorithms to find an end-to-end localization and classification of kidney stones.
- Apply relevant augmentation (APPR) to the target dataset, guiding the precise classification of the stone and normal kidney regions.

2 Literature Survey

Predicting the kidney stone can be carried out in two directions. The first direction is a classification of the kidney stones and the normal images of CT images. Tremendous works have been contributed by various authors in this direction. Yildirim et al.[6] proposed classifying kidney stones for Coronal Computed tomography (CT) images by using XResNet-50 with Grad-CAM to identify the area the model is concentrated to classify. The author used the 1799 images of normal and kidney stones images. The image consists of the whole abdomen, pelvis, and parts of the thorax. The accuracy of the model is 96.82%, but it concentrates on various locations in the image. Baygin et al.[7] used the same data as above by employing ExDARK19 as a transfer learning algorithm to decrease the training time of the model and extract the features. Iterative neighborhood component analysis (INCA) was used to select the feature vector, which concentrates on each location to correctly classify the stone images and finally fed to the k-nearest neighbors algorithm to classify the stone. The work could not show the location of the stone Patro el at.[2] proposed Kronecker Convolution-deep learning techniques by modifying the CNN algorithm to reduce the redundancy in feature maps without convolution overlapping, producing an efficient and accurate result. The classification accuracy was 98.56% and did not predict the location of the stone.

Recently, Authors have also made a novel contribution in the direction of using segmentation using deep learning algorithms like Regions with Convolutional Neural Networks(RCNN), U-Net, etc., and the stone burden is also calculated. Using a dataset of 465 CT scans, Langkvist et al.[8] used CNN and a probabilistic technique to locate the ureter stone. Although the dataset is 3D and contains some slices with stones, classifying and archiving 2.68 false positives per patient is

insufficient to replace the radiologist's primary diagnostic reading. Reducing the number of preprocessing processes, like locating related components, binarizing, and choosing training data for class balance, is a challenge.

Liu et al.[9] use total variation (TV) flow for image noise reduction and Maximally Stable Extremal Regions(MSER) features for finding calculi candidates and computes a total of 7 texture and shape features to train a Support Vector Machine (SVM) for the task of identifying renal calculi in CT scans. The method was validated on a data set of 192 patients with a false positive rate of 8 per patient and a sensitivity of 69. The kidneys are segmented using a 3D U-Net model by Elton et al.[10] before being subjected to gradient-based anisotropic denoising, thresholding, and region-growing. Convolution Neural Network(CNN) is then used to separate kidney stones from spurious regions. uses 180 CT colonography scans on 6185 patients for training and validation, yielding an AUC of 95%. By discriminating between kidney stones and plagues, the accuracy can be improved. Kaviani et al.[11] collected data from 4 Vendors containing 218 CT Scans, used threshold-based segmentation, and achieved a sensitivity of 79%. Li et al.[12] proposed a two-stage segmentation algorithm; in the first stage, coarse kidneys are detected, and the cropped kidney is obtained. U-net is used to segment the kidney stone from the cropped kidney. 500 unenhanced abdominal CT scans were collected among them, 260 were selected for annotation. The dataset contains 209 scans with stone and 50 without stone. The kidney stone dice score was 80.59.

With the rise in popularity of CNN and GPU-accelerated deep learning frameworks, new approaches to developing object detection algorithms were taken. Deep-learning object detection architectures are classified into two types. YOLOv7[13] and SSD300[14] are single-stage detectors that do away with the ROI extraction step and instead classify and regress the candidate anchor boxes directly. The object identification task is divided into two steps by two-stage detectors, which first extract ROI before classifying and regressing the ROIs using R-CNN[15], Fast RCNN[16], and Faster RCNN[17]. The localization of the kidney having renal calculi is not predicted in any of the contributions. As a result, we can evaluate the one-stage detector YOLOv7 and SSD300 compared to the two-stage Faster RCNN with other backbone networks. The categorization models were developed and tested using the backbone of deep learning architectures, including ResNet50, ResNet101, Mobilenetv2, and VGG16. To enhance the performance of image recognition models, suitable augmentation strategies[18] should be automatically designed in accordance with the target data and tasks. Image transformation operations like geometric transformation, photometric transformation, and their combinations are used for medical image classification[19], providing good results. Few works are contributed to augmenting the instances for object detection.

3 Methodology

3.1 Dataset Acquisition

The dataset is collected from the publicly available GitHub repository[6]. The dataset consists of 433 patients, with 268 patients reported having kidney stones and 165 patients with normal kidneys (no stones). A total of 1799 coronal CT images of the patients were collected, 790 images with kidney stones, and 1009 images of normal scans were obtained. The training and testing dataset was divided into an 80:20 ratio.

3.2 Data Annotation

A patient can have one or two kidneys according to their health conditions. Stone can be present in both kidneys or only in one or none(normal), so the image can have one or more instances. The proposed work categorizes the annotations into two classes: normal and kidney stone. The annotation is carried out using Roboflow online platform, which annotates the instances in the dataset using bounding boxes. The total instances generated for the 1799 images is 2738. Figure 1 presents the annotation of the normal and kidney stone images.

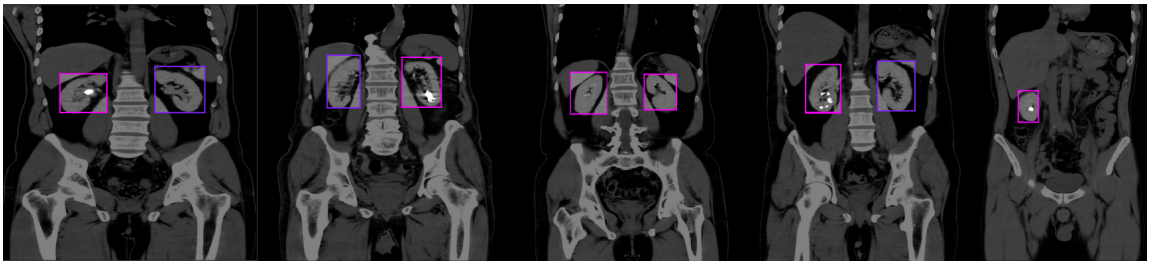

Fig. 1. The kidney stone and normal kidney with two different coloured bounding boxes, pink for kidney with stone and purple for normal kidney.

3.3 Algorithm for bounding box Augmentation

The dataset is enhanced by using color transformations to increase the contrast of the images, as the input images are non-contrast coronal Images. The augmentation of the dataset is performed to generate a new dataset and add it to the existing dataset. The proposed method applies a bounding box pivot-point rotation randomly between -15° to + 15°. The algorithm first calculates the pivot point of the bounding boxes. Then, translates the bounding box such that the pivot point coincides with the coordinate origin. Anti-clockwise rotation is applied about the coordinate origin with a specified angle. Translate back the bounding box such that the pivot point is returned to its original position. The Algorithm 1 presents the steps to generate the dataset using augmentation.

The above algorithm is used for both the bounding boxes of the image, and new bounding boxes are generated to increase the size of the dataset. Figure 2 presents the augmented images with bounding boxes. The bounding box rotation helps in training the kidney more accurately; the visual images show that the angle of the kidney in coronal CT images is tilted. The contribution supports data generation as well as the tilt in the bounding boxes provides a more accurate classification of the kidney from other organs.

3.4 Proposed Architecture

Data augmentation is used to expand the dataset, various approaches are employed to increase the amount of data depending on the type of dataset. Data augmentation serves as a regularizer[20] during the training of machine learning models and aids in preventing overfitting. Data warping augmentations change existing images while retaining their label. This includes enhancements like

Algorithm 1 Algorithm to find Augmented Pivot Point Rotation(APPR) within $[-15, 15]$ Degrees

procedure APPR($box[]$)
  Input:Bounding box with four vertices as: $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$
  $\theta \leftarrow \text{random}[-15, 15]$
  $rbox \leftarrow$ empty list ▷ rbox is the new bounding box
  $x_r \leftarrow \frac{x_1+x_2+x_3+x_4}{4}$
  $y_r \leftarrow \frac{y_1+y_2+y_3+y_4}{4}$ ▷ x_r, y_r is the pivot point
  for each (x, y) box vertices **do**
    $rotated_x \leftarrow (x - x_r) \cdot \cos((\theta)) - (y - y_r) \cdot \sin((\theta)) + x_r$
    $rotated_y \leftarrow (x - x_r) \cdot \sin((\theta)) + (y - y_r) \cdot \cos((\theta)) + y_r$
    $rbox$.append($(rotated_x, rotated_y)$)
  return $rbox$ ▷ Returns New bounding box with rotation

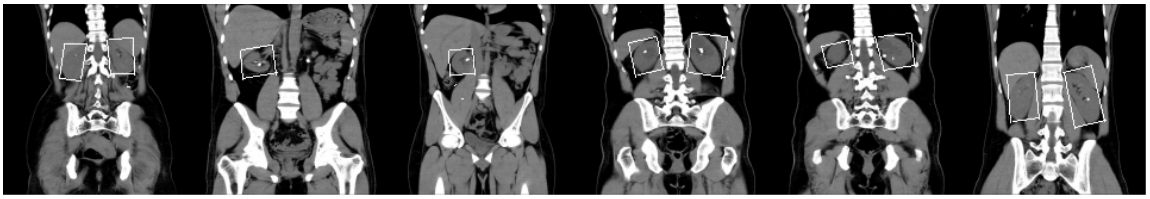

Fig. 2. Sample augmented dataset of normal and kidney stone with bounding boxes.

random erasing, adversarial training, geometric and colour changes, and neural style transfer. Many geometric transformations exist, including noise injection, flip, crop, rotation, and translation. The label of the data is no longer preserved and requires post-transformation. Creating precise labels for every unsafe Data augmentation requires much processing time. There is no image processing function[21] that cannot result in a label-changing transformation along with augmentation. A non-label preserving transformation might make it easier for the model to produce a response that is not confident about its prediction. To accomplish this, post-augmentation labels would need to be adjusted.

By directly applying the enhanced pivot point rotation to the bounding boxes rather than rotating the picture and then post-labeling to produce a reliable prediction, the proposed approach aims to overcome the post-augmentation of the labels. This is shown in figure 3, which presents the steps involved in generating the new dataset. The pivot point rotation is applied to the original dataset of 1799 images, and a new dataset is generated. Both the original and augmented datasets are combined to form a new dataset. The dataset is provided as input to the object detection algorithm for training.

Object detection algorithms provide an end-to-end localization of the objects in the image with the help of bounding boxes. More than one specific object can be detected using an object detection model. Deep learning architecture performs impressive results. The object detection task further localizes the kidney using a bounding box and related confidence score that expresses how confidently the object class—kidney stone or normal—is recognized within the bounding box.

R-CNN, Fast(R-CNN), Faster R-CNN, Region-based Fully Convolutional Network(R-FCN),single-shot detector(SSD), and You Only Live Once(YOLOv7) are some of the CNNs that have significantly raised the bar for performance on the field. Selecting an appropriate architecture depends on

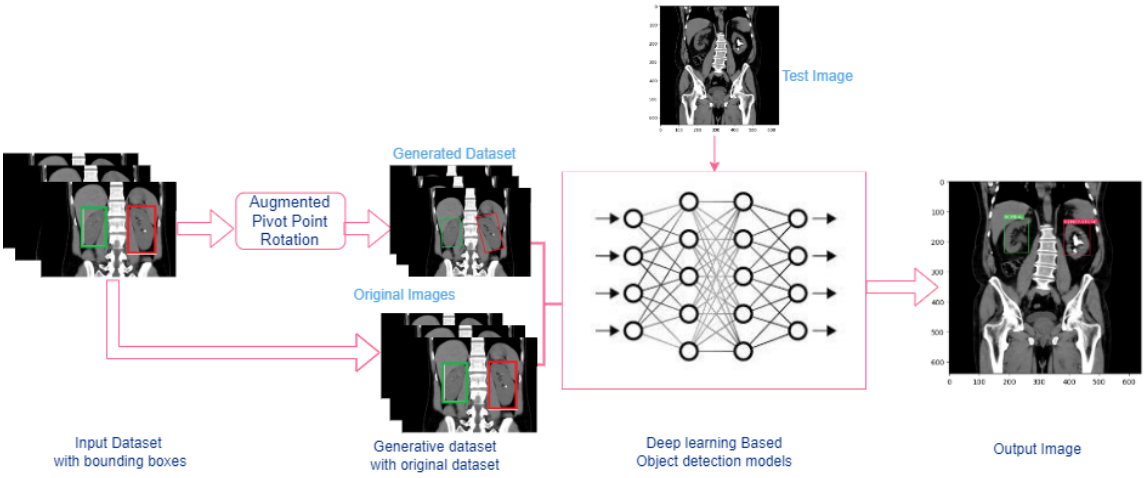

Fig. 3. Proposed Architecture to train Object detection algorithms .

the size and complexity of the dataset and the resources available. In this contribution, YOLOv7[17], SSD[14], and Faster RCNN[13] algorithms with the backbone as ResNet50, MobileNetV2, and YOLOv7 are selected as they achieved great success on large datasets, such as Pattern Analysis, Statistical Modelling, and Computational Learning Visual Object Classes(PASCAL VOC)[22], ImageNet[23], and Common Objects in Context(COCO)[24].

4 Experimental Setup

Various object detection techniques were used to validate the proposed pivot point rotation-based dataset. For experiments, we first evaluate how well YOLOv7, Faster RCNN with mobilenetv2, ResNet50 and ResNet101, and SSD300 locate the instances on original Coronal CT images. Using the augmented dataset, the above five models are retrained. Although the dataset is diverse, the same parameters are used to train the models. Pytorch is used to implement the experiments. The models are trained using an Intel® Xeon® Gold 6226R CPU running at 2.90GHz and an NVIDIA RTX A4000-GPU with 16GB of RAM. All five models are initialized using pre-trained parameters of the COCO dataset. The different hyperparameters used for training are provided in Table 1. The image size is adjusted according to the models used to train the dataset. The performance evaluation uses AP@0.5, AP@0.5:0.95, precision, and recall. These are explained in section 5.1. The comparison of the object detection algorithm with respect to the original and generated dataset is presented in section 5.

4.1 Evaluation Matrices

– **Intersection over Union (IoU)** The degree of overlapping between the instances ground truth(gt)and predicted instances(pd) is computed using equation-1

$$\text{IoU} = \frac{area(gt \cap pd)}{area(gt \cup pd)} \tag{1}$$

Table 1. Hyper parameters of the trained model

Parameters	**SSD**	**Faster RCNN**	**YOLOV7**
Batch size	8	8	8
Momentum	0.9	0.9	0.937
Weight decay	0.0005	0.0005	0.0005
Learning Rate	0.001	0.001	0.001
Optimization Algorithm	SGD	SGD	SGD
Parameters	103.76M	42M	37.2M
Image resolution	300*300	1333*800	320*320

True Positive(tp) is computed when IoU(gt,pd) is greater than the threshold, and False Positive(fp) is when IoU(gt,pd) is less than the threshold.

- **Precision and Recall**
  Precision(P) is the ratio between samples correctly classifying those true samples as true and all samples the model properly identified as positive as in equation-2. Precision is high when ground truth samples are predicted accurately, and incorrect prediction of the other classes as true is low.

$$\mathrm{P} = \frac{tp}{(tp + fp)} \tag{2}$$

  Recall(R) is the ratio between the sample correctly classifying those ground truth samples as true and the total number of instances as in equation-3. The more positive samples that are classified, the larger the recall. Recall focuses on classifying the positive samples and ignores the negative samples.

$$\mathrm{R} = \frac{tp}{(totalinstances)} \tag{3}$$

- **Mean Average Precision (mAP) @0.5** The average precision(AP) of each class is calculated using 11-point interpolation for all models. The mean Average Precision is calculated for each class, considering the IoU threshold value as 0.5. These AP values are averaged to find the mean Average Precision (mAP@0.5).

$$\mathrm{mAP@0.5} = \frac{1}{n}\sum_{r=1}^{n} AP_i, \tag{4}$$

- **Mean Average PrecisionmAP@0.5:0.95** The average of ten mAP scores was calculated with an IoU threshold of 0.5 to 0.95 with an increment of 0.05. The mAP score is the standard metric of COCO evaluation.

$$\mathrm{mAP@0.5:0.95} = \frac{1}{10}\sum_{r=1}^{10} mAP_i, \tag{5}$$

5 Experimental Results

The experiment presents the work to compare the five models trained on the original dataset as well as the performance of the same five models, trained on the proposed augmented dataset generated using the APPR algorithm. The results obtained by the ten models are categorized to address the trade-off regarding training and validation loss, comparing the Mean Average Precision(mAP@0.50)

and mAP@0.5:0.95. The models perform for the IoU threshold@0.5 in terms of precision, and recall is also considered. The training and validation loss plots are obtained after training the models on the

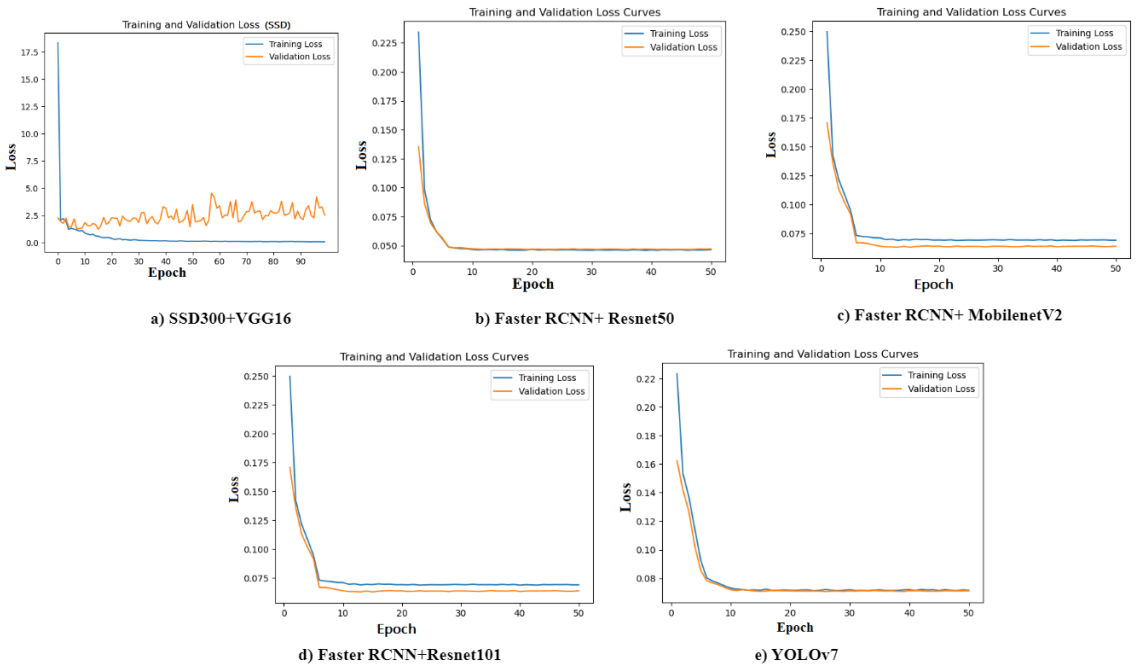

Fig. 4. Training and Validation loss plots of the five models trained on the augmented dataset.

augmented dataset. The training and validation dataset ratio is 80:20, the same as provided in the publicly available dataset. Figure4(a) shows that SSD300 generates over-fitting results on validating the new data for 100 epochs. Figure4(b) and Figure4(e) are of Faster-RCNN with ResNet50 and YOLOv7 , produces good fitting for 50 epochs only. Faster-RCNN with Mobilenetv2 and ResNet101 in Figure4(c) and(d) presents it is easier to predict the new data than the training dataset. The graphs contradict the results regarding evaluation parameters presented in Table 2.

Table 2. Mean Average Precision of five models on varied dataset

Object Detection Models	**Backbone**	**Original Dataset**		**Augmented Dataset**	
		mAP@0.5	**mAP@.5:.95**	**mAP@0.5**	**mAP@.5:.95**
SSD300	VGG16	0.965	0.602	**0.9866**	0.862
Faster-RCNN	Mobilenetv2	0.447	0.159	0.494	0.263
Faster-RCNN	ResNet50	0.977	**0.713**	0.980	0.893
Faster-RCNN	ResNet101	0.983	0.705	0.986	0.881
YOLOv7	–	**0.984**	0.706	0.984	**0.931**

Table 2 presents the mAP@0.5 and mAP@.5:0.95 results for the five object detection algorithms trained on the original dataset and augmented dataset. The results show that the Faster-RCNN with ResNet50 outperformed on the original dataset compared to Mobilenetv2 and ResNet101. Faster-RCNN also produces the best results when compared among the five models with mAP@0.5:0.95 score of 71.3%, whereas YOLOv7 performance is good for the threshold of 0.5.

After generating a new dataset by applying the APPR algorithm, YOLOv7 surpassed the result of mAP@0.5:0.95 with a spike of 22.5% and produced a magnificent result. Application of the APPR algorithm also produced a promising improvement with an average hike of 18.9% of all the five algorithms in terms of mAP@0.5:0.95, though a markable hike of 1% in mAP@0.5 results. SSD300 produces a better result of 98.66% when trained on the augmented dataset.

The precision and recall of the test dataset are computed for the IoU threshold value as 0.5. The precision is used to correctly classify the true samples as true rather than incorrectly classifying the negative samples as true. The recall concerns that every positive sample is correctly identified, but it is unconcerned if a negative sample is incorrectly labeled as positive. The performance of YOLOv7

Table 3. Precision and Recall of five models on Original dataset

Object Detection Models	**Precision**		**Recall**	
	Normal	**Kidney-Stone**	**Normal**	**Kidney-Stone**
SSD300 + VGG16	0.94	0.916	0.99	0.979
Faster-RCNN + MoblilenetV2	–	0.5869	–	0.9973
Faster-RCNN+ResNet50	0.9367	0.9470	0.9876	0.99
Faster-RCNN+ ResNet101	0.9610	0.9467	**0.9867**	**0.9920**
YOLOv7	**0.987**	**0.964**	0.981	0.979

is high in predicting the true samples more accurately than the negative sample as true with the original dataset. The precision value of the original dataset of YOLOv7 of kidney stones and normal instances is 98.7% and 96.4% respectively as in Table 3, whereas for the augmented dataset the precision score is reduced for the IoU threshold@0.5. SSD300 model identifies the true positive and false negative samples correctly from the augmented dataset for an IoU threshold@0.5 as 98.4% and 99.7% for normal and kidney stones respectively as in Table 4. The recall of the Faster RCNN

Table 4. Precision and Recall of five models on augmented dataset

Object Detection Models	**Precision**		**Recall**	
	Normal	**Kidney-Stone**	**Normal**	**Kidney-Stone**
SSD300 + VGG16	**0.9848**	**0.9818**	0.9848	0.9970
Faster-RCNN + MoblilenetV2	0.8681	0.9640	0.750	1.0
Faster-RCNN+ResNet50	0.9611	0.9746	0.9757	1.0
Faster-RCNN+ ResNet101	0.9477	0.9763	**0.9909**	**1.0**
YOLOv7	0.952	0.946	0.993	0.988

on both the original and augmented dataset is good, the true kidney stone and normal instances are correctly classified, but even some of the negative sample is also classified as true.

6 Conclusion and Future Scope

A new pivot point rotation-based augmentation technique has been proposed to enhance the performance of deep learning models for kidney stone detection. In this regards YOLOv7, SSD300, and Faster RCNN with the backbone models are trained with the original and augmented datasets. An average increase of 18.9% of mAP@0.5:0.95 is achieved by the use of an augmented dataset. We conclude that the proposed algorithm with pivot point rotation, enhances the performance of the models. The results of the work justify, YOLOV7 proves appreciating results in terms of mAP@0.5:0.95 with 93.1% when trained on the generated dataset, and the model training also results in a good fit. Faster-RCNN with ResNet50 performs skilfully at IoU threshold@0.5, confirmed by the results of mAP@0.5 as 0.980. SSD300 also produces promising results regarding the IoU threshold@0.5, but the graph presents that SSD300 convinces overfitting results. The need for a lightweight, accurate, and appropriate model with less time complexity is still a challenge in the medical field. The class imbalance problem must be addressed, as the original dataset had more normal than kidney stone instances. The contribution assists various researchers in evaluating multiple models and applying the appropriate model according to their requirements.

References

1. Raphaële Renard-Penna, Aurélie Martin, Pierre Conort, Pierre Mozer, and Philippe Grenier. Kidney stones and imaging: what can your radiologist do for you? *World journal of urology*, 33:193–202, 2015.
2. Kiran Kumar Patro, Jaya Prakash Allam, Bala Chakravarthy Neelapu, Ryszard Tadeusiewicz, U Rajendra Acharya, Mohamed Hammad, Ozal Yildirim, and Paweł Pławiak. Application of kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal ct images. *Information Sciences*, 640:119005, 2023.
3. Saman Ebrahimi and Vladimir Y. Mariano. Image quality improvement in kidney stone detection on computed tomography images. *Journal of Image and Graphics*, 3, 2015.
4. Jaskirat Kaur and Williamjeet Singh. Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, 81(27):38297–38351, 2022.
5. Fabio Garcea, Alessio Serra, Fabrizio Lamberti, and Lia Morra. Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*, page 106391, 2022.
6. Kadir Yildirim, Pinar Gundogan Bozdag, Muhammed Talo, Ozal Yildirim, Murat Karabatak, and U.Rajendra Acharya. Deep learning model for automated kidney stone detection using coronal ct images. *Computers in Biology and Medicine*, 135:104569, 2021.
7. Mehmet Baygin, Orhan Yaman, Prabal Datta Barua, Sengul Dogan, Turker Tuncer, and U. Rajendra Acharya. Exemplar darknet19 feature generation technique for automated kidney stone detection with coronal ct images. *Artificial Intelligence in Medicine*, 127:102274, 2022.
8. Martin L§ngkvist, Johan Jendeberg, Per Thunberg, Amy Loutfi, and Mats Lid©n. Computer aided detection of ureteral stones in thin slice computed tomography volumes using convolutional neural networks. *Computers in Biology and Medicine*, 97:153–160, 2018.
9. Liu J, Wang S, Turkbey EB, Linguraru MG, Yao J, and Summers RM. Computer-aided detection of renal calculi from noncontrast ct images using tv-flow and mser features. *Medical Physics*, 42:144–153, 2015.
10. Elton DC, Turkbey EB, Pickhards, and PJ Summers RM. A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast ct scans. *Medical Physics*, 49:2545–2554, 2022.
11. Parisa Kaviani, Andrew Primak, Bernardo Bizzo, Shadi Ebrahimian, Sanjay Saini, Keith J Dreyer, and Mannudeep K Kalra. Performance of threshold-based stone segmentation and radiomics for determining

the composition of kidney stones from single-energy ct. *Japanese Journal of Radiology*, 41(2):194–200, 2023.
12. Dan Li, Chuda Xiao, Yang Liu, Zhuo Chen, Haseeb Hassan, Liyilei Su, Jun Liu, Haoyu Li, Weiguo Xie, Wen Zhong, et al. Deep segmentation networks for segmenting kidneys and detecting kidney stones in unenhanced abdominal ct images. *Diagnostics*, 12(8):1788, 2022.
13. Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.
14. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.
15. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
16. Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
17. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.
18. Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.
19. Evgin Goceri. Medical image data augmentation: techniques, comparisons and interpretations. *Artificial Intelligence Review*, pages 1–45, 2023.
20. Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. *Advances in neural information processing systems*, 31, 2018.
21. Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.
22. Bird Boat Bottle Bus and Person Potted. Pascal voc-07.
23. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
24. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.