



Deep Learning Based Detection Of Bone Fractures In X-Ray Images

¹Gayathri D, ²Afzal Hameed Faiz V Y, ²Mohamed Imthiyaz K,

²Abdul Kareem M, ²Arssam Basha M

¹Assistant Professor, ²Student

^{1,2} Department of Computer Science and Engineering,

^{1,2}Aalim Muhammed Salegh College of Engineering, Chennai, India

Abstract

Bone fractures are a common and critical concern in paediatric radiology, where timely and accurate diagnosis is essential for effective treatment. This study explores the application of the YOLO (You Only Look Once) object detection framework, specifically YOLOv7, for automated bone fracture detection using the publicly available GRAZPEDWRI-DX dataset, which contains paediatric wrist trauma X-ray images. The model was trained to classify and localize fractures with high precision and efficiency. Various performance metrics such as precision, recall, accuracy, mAP@0.5, and mAP@0.5:0.95 were used to evaluate the model's effectiveness. The results demonstrate that YOLOv7 can achieve strong detection performance while maintaining real-time inference capabilities, making it a viable solution for clinical decision support systems. This work highlights the potential of deep learning-based object detection models in assisting radiologists by improving diagnostic accuracy and reducing workload. Further refinement through model optimization and larger datasets could enhance its applicability in broader medical imaging tasks.

1. Introduction

Musculoskeletal injuries, especially wrist fractures, are among the most common reasons children visit the emergency department. Accurate and timely diagnosis of these injuries is essential to ensure appropriate treatment and avoid complications such as malunion or growth disturbances. While radiographic imaging remains the standard for diagnosis, interpreting paediatric X-rays can be challenging due to skeletal immaturity and the subtle appearance of certain fractures. The increasing workload on radiologists further amplifies the need for intelligent computer-aided diagnostic systems. Manual diagnosis of paediatric wrist fractures, although highly dependent on the expertise of radiologists, poses several challenges that can hinder accuracy and efficiency. One of the primary issues is the variability in radiographic interpretation. In recent years, deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful tool in the field of medical image analysis. These models are capable of learning hierarchical features directly from raw image data and have shown expert-level accuracy across various diagnostic tasks, such as detecting pneumonia, screening for breast cancer, and classifying diabetic retinopathy. In the domain of orthopaedics, deep learning algorithms have shown promise in the automated detection of bone fractures by learning complex visual patterns that might be difficult for the human eye to discern. Recent advancements in deep learning and computer vision have shown significant potential for automating medical image analysis. Object detection models, such as YOLO (You Only Look Once), have been applied to various medical imaging tasks, including fracture detection, and offer promising results in terms of identifying and localizing abnormalities in radiographic images. To enable these advancements, high-quality annotated medical datasets are crucial for training and evaluating such models. Furthermore, YOLO's ability to detect multiple objects within a single image makes it ideal for scenarios where a radiograph may contain more than one fracture site or

multiple anatomical regions. Its variants have introduced various improvements in feature extraction, attention mechanisms, and post-processing techniques, leading to significant boosts in both precision and recall in medical detection tasks. Here, we leverage the YOLO architecture for the task of automated bone fracture detection in X-ray images. The objective is to build an end-to-end deep learning pipeline that can detect and localize fractures efficiently, thereby assisting radiologists in making faster and more reliable diagnoses. The performance of the model is evaluated using metrics such as mean Average Precision (mAP), Intersection over Union (IoU), and visual interpretability through bounding box overlays. Through this work, we aim to bridge the gap between cutting-edge AI research and practical diagnostic tools in medical imaging. The GRAZPEDWRI-DX dataset containing 20,327 images that were annotated by paediatric radiologists. These annotations include bounding boxes, fracture classifications, and additional clinical descriptors such as projection type and anatomical tags. The dataset's size and detailed annotations make it an invaluable resource for training models aimed at fracture detection and classification. we apply a YOLOv7 model trained on the GRAZPEDWRI-DX dataset to detect and localize fractures in paediatric wrist X-rays. Our goal is to assess the model's effectiveness in identifying fractures and to explore its potential for supporting radiologists in clinical settings. Through both quantitative and qualitative analysis, we demonstrate the utility of the YOLOv7 model in assisting with fracture detection and highlight its relevance for clinical applications. Benefit of YOLOv7's detection performance is that it facilitates automated reporting and triage systems. By integrating such a model into clinical workflows, it becomes possible to prioritize patients with suspected fractures for faster treatment, thereby improving overall patient care and reducing wait times in emergency settings.

2. Architecture

YOLO processes the input image by dividing it into a grid of cells, with each cell tasked with predicting bounding boxes and their corresponding class probabilities. This grid-based approach enables the model to efficiently detect objects of varying sizes and positions throughout the image. By generating predictions directly from each grid cell, YOLO achieves both high speed and effective object detection performance. In YOLOv7, one of the most significant architectural advancements is the introduction of the Efficient Layer Aggregation Network (ELAN), a module that plays a critical role in improving the learning capacity of the model without increasing computational overhead. Traditional deep convolutional networks tend to suffer from problems related to inefficient information flow, vanishing gradients, and redundancy in learned features as depth increases. ELAN addresses these issues by structuring how features are learned and aggregated across multiple branches in the network. Rather than simply stacking convolutional layers in a sequential manner, ELAN strategically aggregates outputs from several intermediate layers. This design facilitates a richer and more diverse feature representation, enabling the model to learn from a broader and deeper set of activations while maintaining a consistent and stable gradient path during training. YOLOv7 further enhances this concept using Extended ELAN (E-ELAN). E-ELAN builds upon the original ELAN structure by introducing group convolution operations, a technique that divides input channels into smaller groups and applies convolution independently within each group. This significantly increases the cardinality, or the number of independent feature transformations the model can learn. Cardinality, a concept borrowed from the ResNeXt architecture, is known to be a more effective way of increasing model capacity than simply going deeper or wider. By leveraging group convolutions, E-ELAN allows the network to capture a wider variety of feature representations in parallel, making it more adept at identifying subtle and fine-grained details, such as hairline fractures in medical X-rays.

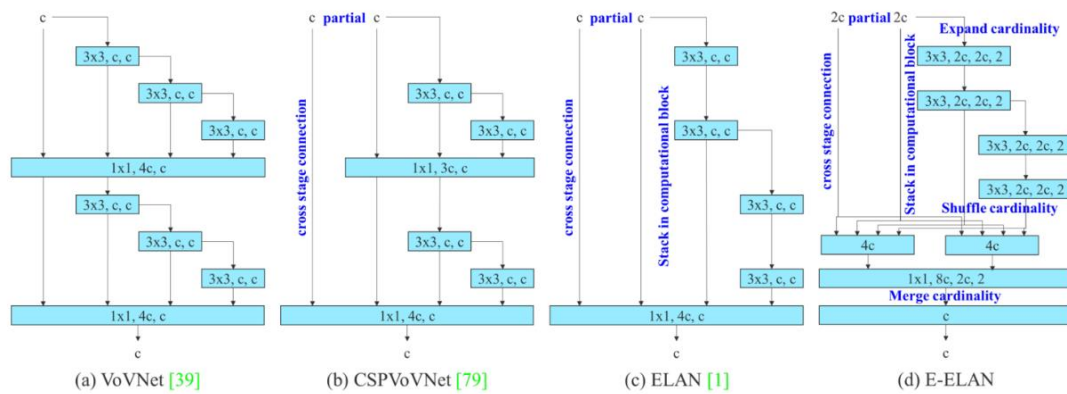


Figure 1. VoVNET, CSPVoVNET, LAN and ELAN

One of the core strengths of E-ELAN is that it retains the original gradient transmission paths of ELAN while enhancing the diversity of learned features. This is achieved through a clever shuffle and merge strategy that follows the group convolutions. After the grouped features are computed, E-ELAN mixes the channels across the groups and then merges them, encouraging better inter-group communication and reducing feature redundancy. This operation ensures that each output feature map benefits from the diversity introduced by the grouped processing while still maintaining a coherent and unified representation for subsequent layers. Another important advantage of E-ELAN is its parameter efficiency. Despite introducing more complexity in terms of learning diverse features, the module does not substantially increase the number of parameters or the computational load. This is particularly beneficial for real-time object detection systems like YOLO, where speed and accuracy must be carefully balanced. In the context of medical imaging and, specifically, bone fracture detection, this becomes even more critical. X-ray images often contain very subtle signs of fractures that require high sensitivity from the model. E-ELAN's ability to extract fine-grained features while maintaining speed makes YOLOv7 especially well-suited for such applications.

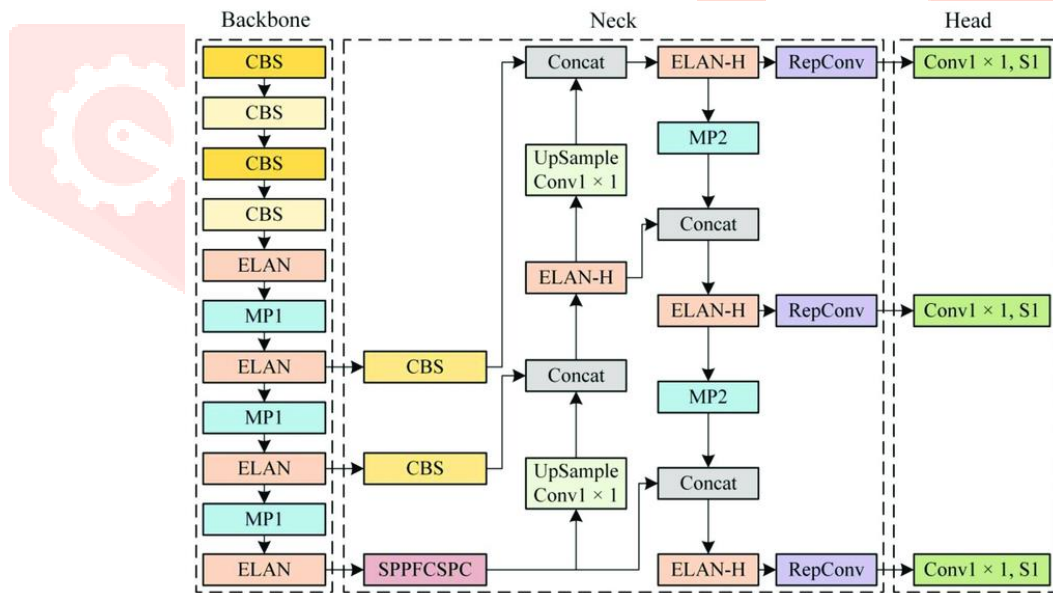


Figure 2. The YOLOv7 network structure

The YOLO structure typically has a neck, a head and backbone and the outputs are achieved in the head. YOLOv7 has made minimal modifications here too. It is not bound by a single head. It has multiple heads to do whatever it desires. Previously this technique has also been employed in Deep Supervision technique that is employed by DL models to make use of multi heads. There are two heads that. First, the main head in YOLOv7 is known as the final output head and second, the Auxiliary Head is the head that assists middle-layer training. The neck architecture in YOLOv7 resembles that of YOLOv4 and YOLOv5, utilizing a PAFPN (Path Aggregation Feature Pyramid Network) structure. Initially, the backbone output is down sampled by a factor of 32 to generate the C5 feature map. This is then processed through the SPPCSPC module, reducing the channel count from 1024 to 512. A top-down fusion is applied using features C4 and C3 to produce P3, P4, and P5. Subsequently, a bottom-up pathway is integrated with P4 and P5. Unlike YOLOv5, YOLOv7

replaces the MP2 layer with a down sampling layer and swaps the CSP module with the ELAN-H module. The P3, P4, and P5 outputs are refined using RepConv to adjust their channel dimensions. It leverages the advanced YOLOv7 object detection model, which offers improved target detection accuracy through optimized network modules and training strategies. While YOLOv7 increases the cost during training, it maintains a low inference cost and delivers faster detection speeds. As shown in Figure 9, the structure of the YOLOv7 network begins by passing the input image through the Backbone network. The Head layer then produces three feature maps of varying sizes, and finally, the RepConv layer outputs the prediction results. The Backbone of YOLOv7 consists of 50 layers, designed to effectively extract features relevant for tasks such as bone fracture detection. It begins with four convolutional layers, followed by the ELAN module, and includes three MaxPooling (MP) layers interleaved with additional ELAN modules to produce feature outputs at three different scales. The ELAN module is made up of multiple CBS blocks, each comprising a Convolution (Conv) layer, Batch Normalization (BN), and the SiLU activation function.

3. Evaluation

The evaluation of the YOLOv7 model for bone fracture detection was carried out using the GRAZPEDWRI-DX dataset, which contains annotated paediatric wrist trauma X-ray images. This dataset provided a balanced and realistic foundation to test the model's capabilities in distinguishing between healthy and fractured wrist bones. The assessment of the model's performance relied on standard object detection metrics, primarily precision, recall, F1-score, and mean Average Precision (mAP) at different IoU thresholds (e.g., mAP@0.5 and mAP@0.5:0.95). These metrics provide a comprehensive understanding of both the correctness and completeness of the detections. The dataset was split into training, validation, and test sets in a 70:20:10 ratio. Before training, we resized all images to 640x640 pixels and applied data augmentations like flipping, rotation, and contrast adjustments to improve generalization. During training, we used standard object detection loss functions to update the model weights. We closely monitored both training loss (TR-Loss) and validation loss (VL-Loss) to make sure the model was not overfitting. As training progressed, the training loss steadily decreased, showing that the model was learning well. The validation loss also reduced but at a slower rate, indicating good generalization. To evaluate the model's performance, we focused on metrics like precision, recall, and mean Average Precision (mAP). Precision tells us how accurate the model's fracture predictions are while recall tells us how well the model detected actual fractures in the images.

3.1 Precision

Precision are standard metrics used for binary classification tasks, where the ratio of true positive predictions to the total predicted positives, and recall is the ratio of true positive predictions to the total actual positives. It is calculated using the following formula:

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

True Positives (TP) refer to the cases where the model correctly identifies a positive instance.

False Positives (FP) are the cases where the model incorrectly classifies a negative instance as positive.

3.2 Recall

Recall, also known as sensitivity or true positive rate, is calculated using the following formula:

$$\text{Recall} = \frac{\text{Total correctly classified positive samples}}{\text{Total actual positive samples}}$$

True Positives (TP) are the cases in which the model correctly identifies a positive instance.

False Negatives (FN) refer to the cases where the model fails to recognize a positive instance, even though it is present.

Accurate localization and classification of fractures in X-rays can assist radiologists in confirming diagnoses more quickly and reliably. In busy hospital environments, where radiologists may need to review hundreds of images per day, a high-performing model like YOLOv7 can act as a second opinion or an initial screening

tool. This not only enhances diagnostic efficiency but also reduces the likelihood of human error, especially in subtle or borderline cases where fractures are harder to detect.

TABLE I. DETECTION EVALUATION METRICS

Bone Description	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Healthy Bone	0.931	0.912	0.833	0.723
Fractured Bone	0.903	0.883	0.841	0.694

TABLE II. EXPERIMENT RESULTS OF DATA SET

Bone Description	Accuracy Type (TR-acc) (VL-acc)		Loss Type (TR-Loss)(VL-Loss)	
Healthy Bone	93.32%	91.22%	0.078	0.1057
Fracture Bone	94.76%	88.14%	0.066	0.1301

Despite these strengths, the evaluation also revealed certain areas for improvement. The model showed some false negatives, particularly in complex fracture cases with low contrast or overlapping bone structures. This indicates that while YOLOv7 performs well, its sensitivity could still be enhanced by incorporating more diverse training data or by fine-tuning hyperparameters. In addition, experimenting with newer architectures like YOLOv8 or transformer-based models may further improve detection performance, especially under challenging imaging conditions. The evaluation results strongly support the effectiveness of YOLOv7 in detecting bone fractures in paediatric wrist X-rays. The high precision and recall, along with competitive mAP scores, suggest that the model has significant potential for clinical application. Its integration into radiology workflows can help speed up diagnosis, reduce radiologist workload, and ultimately improve patient care.

4. Conclusion

This study focused on evaluating the performance of YOLOv7 for bone fracture detection using the GRAZPEDWRI-DX dataset, which comprises paediatric wrist trauma X-ray images. The experiment provided valuable insights into the application of YOLOv7 in medical image analysis, demonstrating its potential in automating bone fracture detection. The results highlighted the model's ease of implementation and emphasized the importance of key evaluation metrics such as precision, recall, and mAP@0.5 for performance assessment. YOLOv7 achieved strong results, with a precision of 93.1% for identifying healthy bones and 90.3% for detecting fractured bones. Similarly, recall values were 91.2% and 88.3%, indicating a solid balance between identifying true positives and minimizing false negatives. The model also showed robust performance across mAP@0.5 and mAP@0.5:0.95 metrics, reinforcing its effectiveness in fracture detection. While the findings confirm YOLOv7's capability in this domain, there remains room for refinement particularly in reducing false negatives and improving sensitivity at higher IoU thresholds.

References

1. C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.
2. Nagy, E., Janisch, M., Hrzić, F. *et al.* A paediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning. *Sci Data* **9**, 222 (2022). <https://doi.org/10.1038/s41597-022-01328-z>.
3. Li, Qingge & Yang, Xiaogang & Lu, Ruitao & Fan, Jiwei & Wang, Siyu & Qin, Zhen. (2023). VisionICE: Air–Ground Integrated Intelligent Cognition Visual Enhancement System Based on a UAV. *Drones*. 7. 268. 10.3390/drones7040268.
4. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
5. Sokolova, Marina & Japkowicz, Nathalie & Szpakowicz, Stan. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science*. Vol. 4304. 1015-1021. 10.1007/11941439_114.
6. Cerqueira, Vitor & Torgo, Luís & Mozetic, Igor. (2019). Evaluating time series forecasting models: An empirical study on performance estimation methods. 10.48550/arXiv.1905.11744.
7. Nanglia, Pankaj & Kumar, Sumit & N.Mahajan, Aprana. (2020). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. *ICT Express*. 10.1016/j.icte.2020.06.007.
8. Y. Lee, J. -w. Hwang, S. Lee, Y. Bae and J. Park, "An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 752-760, doi: 10.1109/CVPRW.2019.00103.
9. C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 13024-13033, doi: 10.1109/CVPR46437.2021.01283.
10. Lee, Chen-Yu & Xie, Saining & Gallagher, Patrick & Zhang, Zhengyou & Tu, Zhuowen. (2014). Deeply-Supervised Nets. 38.
11. Huang, Lina & Miron, Alina & Hone, Kate & Li, Yongmin. (2024). Segmenting Medical Images: From UNet to Res-UNet and nnUNet. 10.48550/arXiv.2407.04353.
12. Benchoufi, M., Matzner-Lober, E., Molinari, N., Jannot, A. S., & Soyer, P. (2020). Interobserver agreement issues in radiology. *Diagnostic and interventional imaging*, 101(10), 639–641. <https://doi.org/10.1016/j.diii.2020.09.001>