

YOLOv9 for fracture detection in pediatric wrist trauma X-ray images

Chun-Tse Chien,¹ Rui-Yang Ju,²  Kuang-Yi Chou,³ and Jen-Shiun Chiang¹, 

¹Department of Electrical and Computer Engineering, Tamkang University, New Taipei City, Taiwan

²Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei City, Taiwan

³School of Nursing, National Taipei University of Nursing and Health Sciences, Taipei City, Taiwan

✉ E-mail: jsken.chiang@gmail.com

The introduction of YOLOv9, the latest version of the you only look once (YOLO) series, has led to its widespread adoption across various scenarios. This paper is the first to apply the YOLOv9 algorithm model to the fracture detection task as computer-assisted diagnosis to help radiologists and surgeons to interpret X-ray images. Specifically, this paper trained the model on the GRAZPEDWRI-DX dataset and extended the training set using data augmentation techniques to improve the model performance. Experimental results demonstrate that compared to the mAP 50–95 of the current state-of-the-art model, the YOLOv9 model increased the value from 42.16% to 43.73%, with an improvement of 3.7%. The implementation code is publicly available at <https://github.com/RuiyangJu/YOLOv9-Fracture-Detection>.

Introduction: The accurate interpretation of X-ray images is critical to the success of surgery. However, there is a shortage of experienced radiologists in some areas, which poses a challenge [1, 2]. Computer-assisted diagnosis (CAD) helps specialists such as radiologists and surgeons to interpret medical images, including magnetic resonance imaging (MRI), computed tomography (CT), and X-ray images. However, it cannot be relied on as the sole means of detection due to uncertainties in its prediction accuracy. With the rapid development of deep learning in computer vision tasks [3, 4], its application to medical image processing has obtained satisfactory results [5].

You only look once (YOLO) series [6–9] are the main neural networks for real-time object detection task, widely employed in fracture detection [10–12]. Wrist fractures in children are more common cases and the GRAZPEDWRI-DX dataset [13] provides 20,327 X-ray images of pediatric wrist trauma that can be used in fracture detection tasks. Ju et al. [11] first used the YOLOv8 [8] model for fracture detection on this dataset. Since attention mechanisms [14–18] have excellent results in enhancing the performance of neural network models, Chien et al. [12] achieved the state-of-the-art (SOTA) performance by incorporating different attention mechanisms into the YOLOv8 model.

With the presentation of YOLOv9 [9], which achieved remarkable model performance on the MS COCO 2017 [19] benchmark dataset, achieving an impressive increase in mAP from 53.9% in YOLOv8 to 55.6%. This paper first trained the YOLOv9 model on the GRAZPEDWRI-DX dataset and obtained the SOTA performance, as shown in Figure 1.

The main contributions of this paper are as follows:

1. This paper is the first to apply YOLOv9 to the fracture detection task, demonstrating that the model not only has the excellent performance in real-time object detection across real-life scenarios, but also has good results in medical image recognition.
2. This paper addresses the issue of information loss in fracture detection on X-ray images by employing YOLOv9 algorithm, and aims to preserve more information during model training on low-features X-ray images, enhancing the performance of the model.
3. The mAP 50–95 of YOLOv9 model trained on the GRAZPEDWRI-DX dataset is significantly improved, achieving the SOTA level.

Related works: In the field of object detection task, detectors typically employ either one-stage or two-stage algorithms. Compared to two-stage object detectors, YOLO series models offer a balanced combination of

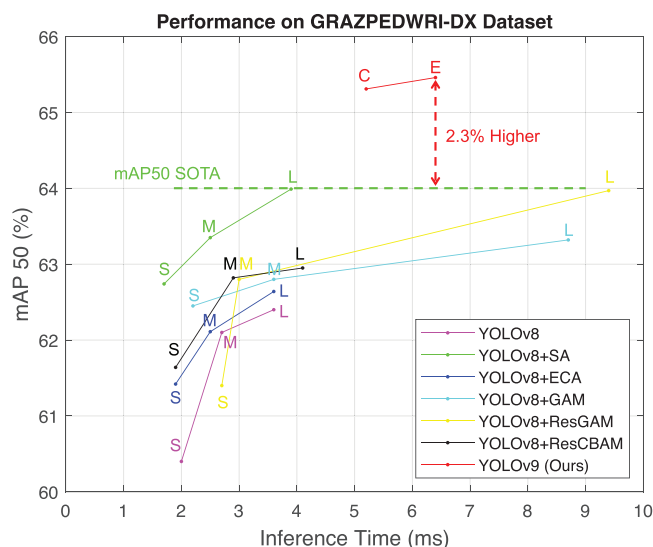


Fig. 1 Comparisons of fracture detection models on the GRAZPEDWRI-DX dataset when the input image size is 640. In terms of accuracy, our models outperform all previous models at the state-of-the-art level

accuracy and inference speed, making them suitable for deployment on mobile computing platforms for medical image recognition. Hrzić et al. [10] employed the YOLOv4 [7] model for fracture detection on the GRAZPEDWRI-DX dataset [13], which was the first to demonstrate that the models of YOLO series can assist radiologists in more accurately predicting wrist injuries in children on X-ray images. Ju et al. [11] developed the “fracture detection using YOLOv8 app” to aid surgeons in interpreting X-ray images for fractures, aiming to reduce misclassification and enhance the information available for fracture surgery. Chien et al. [12] proposed the YOLOv8-AM model, incorporating four attention modules [15–18], and trained them on the GRAZPEDWRI-DX dataset [13]. Specifically, the ResCBAM-based YOLOv8-AM model achieved a significant improvement in mAP 50 value, increasing from 63.6% to 65.8% compared to the YOLOv8 model, which reached the performance of SOTA. While the application of the models of YOLO series to medical image recognition is a hot research topic, to date, no one has utilized YOLOv9 [9] for fracture detection.

Method: Neural networks often have the challenge of information loss since the input data undergoes multiple layers of feature extraction and spatial transformation, resulting in the loss of the original information. This issue is particularly pronounced in X-ray images, where the low-features present significant difficult in fracture detection tasks. Specifically, models trained on such low-featured images tend to perform poorly, and addressing the problem of information loss could substantially enhance the accuracy of model predictions. To address this, we utilizing the YOLOv9 algorithm, which leverages the programmable gradient information (PGI) and the generalized efficient layer aggregation network (GELAN) to more effectively extract key features.

Programmable gradient information: PGI is an auxiliary supervision framework designed to manage the propagation of gradient information across various semantic levels, to improve the detection capability of the model. PGI comprises three main components: main branch, auxiliary reversible branch, and multi-level auxiliary information. During the inference process, it exclusively employs the main branch, which handles both forward and back propagation. As the network becomes deeper, an information bottleneck may occur, leading to loss functions that fail to produce useful gradients. In such cases, auxiliary reversible branch employs reversible functions to preserve information integrity and mitigate information loss in the main branch. Additionally, multi-level auxiliary information addresses the issue of error accumulation from the deep supervision mechanism, improving the learning capacity of the model through the introduction of supplementary information at different levels. Notably, research [9] highlighted the efficacy of PGI in preserving information during training, particularly in scenarios with limited

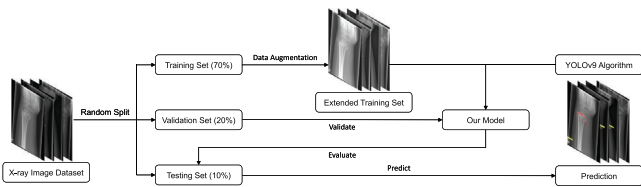


Fig. 2 An overall flowchart of the YOLOv9 algorithm model applied to the fracture detection task

features. This provides the theoretical basis for the YOLOv9 model to have excellent performance in fracture detection tasks.

Generalized efficient layer aggregation network: To enhance information integration and propagation efficiency in model training, YOLOv9 introduced a novel lightweight network architecture named generalized efficient layer aggregation network (GELAN). GELAN integrates CSP-Net [20] and ELAN [21] to efficiently aggregate network information, reducing information loss in propagation and enhancing inter-layer information interaction. This architecture is particularly suitable for fracture detection in environments with limited computing resources due to its lower parameters and computational complexity.

Data processing and augmentation: Figure 2 illustrates the flowchart of the experiments conducted in this study. Since the publisher of the GRAZPEDWRI-DX [13] dataset did not provide predefined training, validation, and test sets, we randomly assigned 70% to training set, 20% to validation set, and 10% to test set during the data processing. Moreover, due to the limited brightness diversity of low-featured X-ray images, models trained only on these images may not generalize well to X-ray images in other environments. To enhance the robustness [22–24] of the model, we followed the data augmentation technique used by Ju et al. [11] to extend the training set. Specifically, we fine-tuned the contrast and luminance of the X-ray images using the addWeighted function from the OpenCV library.

Dataset: GRAZPEDWRI-DX [13] is a public dataset provided by the Medical University of Graz, which contains 20,327 X-ray images of pediatric wrist trauma. These X-ray images were collected by a team of pediatric radiologists at the University Hospital Graz from 2008 to 2018. The dataset comprises 6,091 patients and 10,643 studies, with a total of 74,459 labelled images, representing 67,771 labelled objects. There is a significant class imbalance in this dataset. Specifically, the top two labelled objects are “text” (23,722 instances) and “fracture” (18,090 instances), whereas the number of instances for “bone anomaly” and “soft tissue” are only 276 and 464, respectively. All models use the same training data.

Experiment setup: The experiments in this paper utilized one single NVIDIA GeForce RTX 3090 GPU, employing Python with the PyTorch framework. Before training our model, we employed the YOLOv9 model weights pretrained on the MS COCO 2017 [19] dataset. In the training process, we trained the model using the SGD optimizer, with a weight decay rate set to $5e-4$ and a momentum of 0.937. We followed the research [11] to set the initial learning rate to $1e-2$, the number of epochs to 100. Due to resource limitations (24 GB memory) imposed by a single GPU, a batch size of 16 was employed for training the model.

Experimental results: It is well known that the YOLO series models are designed for real-time object detection and are suitable to be deployed in web or mobile applications to assist surgeons in pediatric wrist fracture diagnosis in all regions of the world. The research [13] has shown that the GRAZPEDWRI-DX dataset is the first publicly available and pertinent pediatric dataset, so we evaluate the performance of different models on it.

To evaluate the performance of YOLOv9 and other SOTA models in real diagnostic scenarios, this study compares model size (parameters and floating point operations), accuracy (F1 score, mean average precision at 50% (mAP 50), and mean average precision from 50% to 95% (mAP 50–95)), and inference time. It is widely recognized that us-

Table 1. Quantitative comparison with other state-of-the-art models for fracture detection on the GRAZPEDWRI-DX datasets when the input image size is 640

Model	Params (M)	FLOPs (G)	F1 Score	mAP 50 (%)	mAP 50–95 (%)	Speed ^a (ms)
YOLOv8 [8]	43.61	164.9	0.59	62.44	40.32	3.6
YOLOv8+SA [17]	43.64	165.4	0.62	63.99	41.49	3.9
YOLOv8+ECA [16]	43.64	165.5	0.61	62.64	40.21	3.6
YOLOv8+GAM [18]	49.29	183.5	0.60	63.32	40.74	8.7
YOLOv8+ResGAM [12]	49.29	183.5	0.62	63.97	41.18	9.4
YOLOv8+ResCBAM [15]	53.87	196.2	0.62	62.95	40.10	4.1
YOLOv9-C (ours)	51.02	239.0	0.64	65.31	42.66	5.2
YOLOv9-E (ours)	69.42	244.9	0.64	65.46	43.32	6.4

Note: The model size of all YOLOv8 and its variants listed in the table is large.

^aSpeed is the total time for preprocessing, inference, and post-processing.

Table 2. Quantitative comparison with other state-of-the-art models for fracture detection on the GRAZPEDWRI-DX datasets when the input image size is 1024

Model	Params (M)	FLOPs (G)	F1 score	mAP 50 (%)	mAP 50–95 (%)	Speed ^a auto (ms)
YOLOv8 [8]	43.61	164.9	0.62	63.63	40.41	7.7
YOLOv8+SA [17]	43.64	165.4	0.63	64.25	41.64	8.0
YOLOv8+ECA [16]	43.64	165.5	0.65	64.26	41.94	7.7
YOLOv8+GAM [18]	49.29	183.5	0.65	64.26	41.00	12.7
YOLOv8+ResGAM [12]	49.29	183.5	0.64	64.98	41.75	18.1
YOLOv8+ResCBAM [15]	53.87	196.2	0.64	65.78	42.16	8.7
YOLOv9-C (ours)	51.02	239.0	0.66	65.57	43.70	12.7
YOLOv9-E (ours)	69.42	244.9	0.66	65.62	43.73	16.1

Note: The model size of all YOLOv8 and its variants listed in the table is large.

^aSpeed is the total time for preprocessing, inference, and post-processing.

ing larger input image sizes improves prediction accuracy but also requires more computational resources. Therefore, we conducted two experiments with input image sizes of 640 and 1024 for various scenarios, and the results are presented in Tables 1 and 2. With the input size of 640, both YOLOv9-C (Compact) and YOLOv9-E (Extended) demonstrate significantly improved mAP, while maintaining a reasonable inference speed. Specifically, YOLOv9-E achieves mAP 50–95 of 43.32%, which is 4.4% higher than 41.49% achieved by the current SOTA model YOLOv8+SA. When the input image size is 1024, the mAP 50–95 of YOLOv9-E reaches 43.73%, which also obtains the SOTA performance. However, due to the increased inference time, it is more suitable for deployment on devices with high computing resources. In summary, the YOLOv9 model effectively balances inference time with fracture detection accuracy, making it suitable for deployment in web applications as a CAD system to assist surgeons in analysing X-ray images.

To facilitate further performance enhancement, we visualized the accuracy of various models in predicting each class, as illustrated in Figure 3. Across all models, the ability to accurately predict “fracture”, “metal”, and “text” classes exceeded 90%, with YOLOv9 notably achieving a 94.8% accuracy in “metal” class prediction. However, the performance in predicting bone anomalies and soft tissue classes was poorer, with YOLOv9 achieving only 11.3% and 28.2% accuracy, respectively. This significantly impacted the average accuracy of YOLOv9 in fracture detection.

Discussion: The performance of YOLO series models is significantly influenced by the quality and diversity of the training datasets. As mentioned above, the limited ability of these methods to accurately

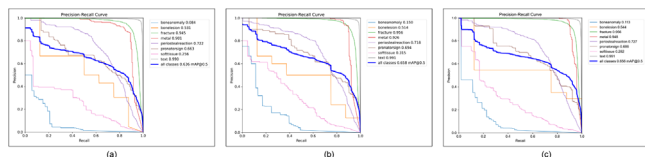


Fig. 3 Visualization of the accuracy of predicting each class using different models on the GRAZPEDWRI-DX dataset with the input image size of 1024. (a) YOLOv8, (b) YOLOv8+ResCBAM (SOTA), (c) YOLOv9 (ours)

predict “bone anomaly” and “soft tissue” classes is due to the insufficient number of relevant data in the GRAZPEDWRI-DX dataset. Therefore, a future research can enhance performance by improving the models with additional information on “bone anomaly” and “soft tissue” classes.

This study focuses exclusively on the detection of pediatric wrist fractures, limiting its applicability to fractures in other body regions. To enhance the model’s utility across new cases, we have made the trained weight files available on our GitHub. Users can leverage these files to train the model using X-ray images of different fracture types, enabling the model to achieve reliable predictions with minimal training data.

Conclusion: The models of YOLO series can serve as CAD to assist radiologists and surgeons in interpreting X-ray images. However, the predictions from the previous models are often unsatisfactory due to the low features of X-ray images. This paper first introduces the application of YOLOv9 to fracture detection, addressing the issue of information loss during model training by employing the newly proposed PGI and GELAN. Experimental results indicate that the YOLOv9 model achieves SOTA performance on the GRAZPEDWRI-DX dataset, proving the effectiveness of this method.

Author contributions: **Chun-Tse Chien:** Conceptualization; methodology; software; visualization; writing—review and editing. **Rui-Yang Ju:** Conceptualization; data curation; methodology; validation; writing—original draft. **Kuang-Yi Chou:** Formal analysis; investigation; visualization. **Jen-Shiun Chiang:** Funding acquisition; project administration; supervision; writing—review and editing.

Acknowledgements: This research is supported by National Science and Technology Council of Taiwan, under Grant Number: NSTC 112-2221-E-032-037-MY2.

Conflict of interest statement: The authors declare no conflicts of interest.

Data availability statement: The dataset that support the findings of this study are openly available in figshare at <https://doi.org/10.6084/m9.figshare.14825193.v2>, reference number Nagy, Eszter; Janisch, Michael; Hrzic, Franko; Sorantin, Erich; Tschauer, Sebastian (2022). GRAZPEDWRI-DX. figshare. Dataset.

The implementation code and the trained model that support the findings of this study are openly available our Github at <https://github.com/RuiyangJu/YOLOv9-Fracture-Detection>.

© 2024 The Author(s). *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

This is an open access article under the terms of the Creative Commons Attribution-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.

Received: 21 March 2024 Accepted: 26 May 2024
doi: 10.1049/ell2.13248

References

- Rimmer, A.: Radiologist shortage leaves patient care at risk, warns royal college. *BMJ* **2017**, 359 (2017)
- Burki, T.K.: Shortfall of consultant clinical radiologists in the UK. *Lancet Oncol.* **19**(10), e518 (2018)
- Sezer, A., Altan, A.: Detection of solder paste defects with an optimization-based deep learning model using image processing techniques. *Soldering Surf. Mount Technol.* **33**(5), 291–298 (2021)
- Yağ, İ., Altan, A.: Artificial intelligence-based robust hybrid algorithm design and implementation for real-time detection of plant diseases in agricultural environments. *Biology* **11**(12), 1732 (2022)
- Adams, S.J., et al.: Artificial intelligence solutions for analysis of X-ray images. *Can. Assoc. Radiol. J.* **72**(1), 60–72 (2021)
- Redmon, J., et al.: You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788. IEEE, Piscataway, NJ (2016)
- Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: optimal speed and accuracy of object detection. arXiv:2004.10934 (2020)
- Glenn, J.: Ultralytics YOLOv8. GitHub. <https://github.com/ultralytics/ultralytics>. Accessed 11 Jan 2023
- Wang, C.Y., Yeh, I.H., Liao, H.Y.M.: YOLOv9: learning what you want to learn using programmable gradient information. arXiv:2402.13616 (2024)
- Hrzić, F., et al.: Fracture recognition in paediatric wrist radiographs: an object detection approach. *Mathematics* **10**(16), 2939 (2022)
- Ju, R.Y., Cai, W.: Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. *Sci. Rep.* **13**(1), 20077 (2023)
- Chien, C.T., et al.: YOLOv8-AM: YOLOv8 with attention mechanisms for pediatric wrist fracture detection. arXiv:2402.09329 (2024)
- Nagy, E., et al.: A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning. *Sci. Rep.* **9**(1), 222 (2022)
- Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141. IEEE, Piscataway, NJ (2018)
- Woo, S., et al.: CBAM: convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19. Springer, Cham (2018)
- Wang, Q., et al.: ECA-Net: efficient channel attention for deep convolutional neural networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11534–11542. IEEE, Piscataway, NJ (2020)
- Zhang, Q.L., Yang, Y.B.: Sa-net: Shuffle attention for deep convolutional neural networks. In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2235–2239. IEEE, Piscataway, NJ (2021)
- Liu, Y., Shao, Z., Hoffmann, N.: Global attention mechanism: retain information to enhance channel-spatial interactions. arXiv:2112.05561 (2021)
- Lin, T.Y., et al.: Microsoft COCO: common objects in context. In: *Proceedings of the 13th European Conference on Computer Vision—ECCV 2014*, pp. 740–755. Springer, Cham (2014)
- Wang, C.Y., et al.: CSPNet: a new backbone that can enhance learning capability of CNN. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 390–391. IEEE, Piscataway, NJ (2020)
- Wang, C.Y., Liao, H.Y.M., Yeh, I.H.: Designing network design strategies through gradient path analysis. arXiv:2211.04800 (2022)
- Ansari, M.Y., et al.: A lightweight neural network with multiscale feature enhancement for liver ct segmentation. *Sci. Rep.* **12**(1), 14153 (2022)
- Mohanty, S., Dakua, S.P.: Toward computing cross-modality symmetric non-rigid medical image registration. *IEEE Access* **10**, 24528–24539 (2022)
- Ansari, M.Y., et al.: Dense-PSP-UNet: a neural network for fast inference liver ultrasound segmentation. *Comput. Biol. Med.* **153**, 106478 (2023)