Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
Paper No. MVML 139
DOI: 10.11159/mvml25.139

A Comparative Study of Segmentation Models for the Identification of the Trapezium Bone in X-rays

Youssef FRIKEL[1,2], Victor MAIGNÉ[3], Thomas GRÉGORY[2,3], Mélanie COURTINE[1,2]
[1]Limics, Université Sorbonne Paris Nord, Sorbonne Université, INSERM UMR S-1142,
Bobigny, France
youssef.frikel1@sorbonne-paris-nord.fr ; melanie.courtine@sorbonne-paris-nord.fr
[2]LaMSN, Université Sorbonne Paris Nord
Saint-Denis, France
[3]Orthopaedic Surgery Department, AP-HP Avicenne,
Bobigny, France
victor.maigne@aphp.fr ; thomas.gregory@aphp.fr

Abstract **-** Recent advancements in deep learning have rendered the identification of bones in X-ray images imperative for tasks such as anomaly detection and surgical procedures. However, the presence of overlapping bones, such as the trapezium, has the potential to compromise the efficacy of this identification. Segmenting the trapezium in X-ray images poses a significant challenge due to its overlap with surrounding bones, including the scaphoid and trapezoid. This study explores the use of deep learning techniques to assist surgeons in accurately localizing the trapezium bone in X-ray images of the hand. This can be helpful in surgical procedures such as trapeziometacarpal joint replacement surgery. The efficacy of a set of models (namely SAM, Mobile-SAM and U-Net) was tested by utilizing radiographic images. Furthermore, a hybrid approach integrating object detection and segmentation was developed. Initially, the object detection model YOLOv8 was trained to localize the region of the image containing the trapezium. This model demonstrated a high level of performance in identifying the trapezium. The utilization of the segmentation model, U-Net, resulted in the identification of pixels belonging to the trapezium bone, thereby achieving a Dice score of 94%. This two-step approach underscores the benefits of this algorithm by reducing the computational load while maintaining high performance.

Keywords**:** Deep Learning, Object Detection, Object Segmentation, X-ray images, Trapezium.

1. Introduction

The trapezium bone poses a significant challenge to identify in the analysis of X-ray images due to its superposition with other bones, particularly the scaphoid and trapezoid (see Fig.1). This complicates the precise extraction of its edges. The trapezium is a crucial component in numerous surgical procedures, including trapeziometacarpal (TMC) joint replacement surgery [1]. In this surgery, the precise identification of the trapezium's location is essential. TMC surgery is a procedure performed on patients afflicted with Rhizarthrosis, a disease characterized by progressive deterioration of the cartilage at the base of the thumb, leading to impaired moving thumb mobility. This surgical intervention involves the replacement of the damaged TMC joint with a prosthesis, with the aim of restoring patients' functionality and quality of life. To ensure optimal outcomes, surgeons must meticulously examine the trapezium through X-ray imaging of the hand.

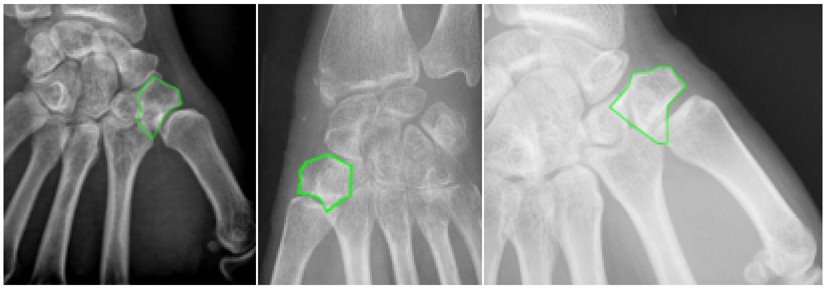

Fig. 1: X-ray images

Recent advancements in deep learning have transformed the field of medical imaging by providing robust algorithms capable of efficiently identifying anatomical structures. These algorithms, which are based on convolutional neural networks (CNNs) [2], have exhibited remarkable capabilities in radiographic images with high precision [3]. CNNs serve as the foundation for numerous algorithms, including YOLO and U-Net. The efficacy of these algorithms in the domain of medical imaging has been well-documented [3].

The objective of this study is to develop a deep learning model capable of efficiently identifying trapeziums, considering the overlap between bones in the region where the trapezium exists. The proposed solution utilizes a two-step approach: (1) object detection is employed to identify the region of interest (ROI) of the trapezium; and (2) object segmentation is implemented to refine the ROI identified by the first step. The efficacy of this two-step approach is evaluated by comparison with segmentation models applied directly. The primary objective of this comparison is to ascertain whether initiating the segmentation process with a detection model enhances its overall effectiveness.

2. Related works

Deep learning [2], a subfield of machine learning, is predicated on artificial neural networks, which enable the development of sophisticated architectures capable of performing complex tasks such as the analysis of medical images [3]. However, these architectures require a large amount of data to learn a concept, which is critical for ensuring optimal performance in learning tasks.

Object recognition is defined as the identification of the location of objects within an image or video. The predominant approach in the field of computer vision entails the detection of objects and their segmentation (see Fig.2). The process of object detection involves the identification and location of the region of interest in an image, that is, the region in which the object is located. Architectures such as YOLO [4] have demonstrated efficacy in this task, exhibiting real-time capabilities with high precision in identifying objects. Segmentation, also referred to as pixel-level classification, is the process of extracting pixels that belong to the object of interest within a given region. This process is critical in fields such as image analysis, robotics, and medical imaging. Architectures such as U-Net [5] have shown notable proficiency in this domain, as substantiated by numerous studies in the field of medical imaging [6].

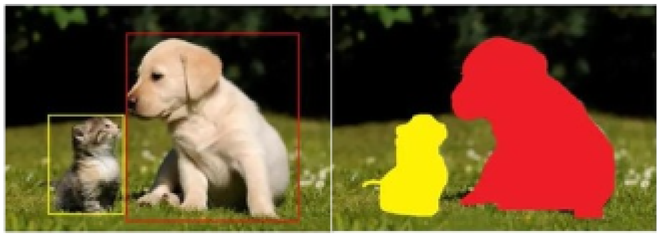

Fig. 2: The difference between object detection (on the left) and object segmentation (on the right).

Recent advancements in the domain of medical imaging have been largely driven by the integration of deep learning algorithms, which have been instrumental in various applications, including cancer detection, tumor detection, fracture detection, and, most notably, the identification of bone structure, a critical component of diagnostic and surgical procedures. The integration of Convolutional Neural Networks (CNN) and advanced deep learning technologies has led to significant advancements in the field of bone detection and segmentation in medical images. YOLO (You Only Look Once) [4] is a renowned object detection algorithm that has been instrumental in facilitating the identification of bones in X-ray images [7]. Its real-time capabilities ensure rapid and accurate detection. In addition to YOLO architecture, models such as U-Net [5] are regarded as leading the field in bone segmentation. This performance can be attributed to the encoder-decoder type architecture of the U-Net. Another critical task in which deep learning models have been applied is fracture detection, a process that involves identifying fractures in X-ray images [8] [9]. Recent techniques, including OSA-YOLOv5 [10] and GRU-UNet [11], have demonstrated enhanced accuracy in hand bone segmentation, attaining up to 14.7% higher accuracy compared to traditional methods. Additionally, attention mechanisms have been integrated into the U-Net model to facilitate the detection of subtle alterations in bone morphology. Aibinder et al. [12] combined self-attention layers with U-Net to enhance the model's ability to focus on the most informative image regions at any given moment in time, leading to more precise detection of slight or irregular patterns in the bone structure.

The integration of detection and segmentation models into a unified architecture has been demonstrated to facilitate the efficient segmentation of the desired object. This strategy has already been applied to localize objects such as brain tumors and cell detection. Nur Iriawan et al. [13] developed a methodology that integrates the YOLO and U-Net algorithms to facilitate the detection and segmentation of brain tumors in MRI images. The proposed YOLO-U-Net architecture exhibited a high degree of performance, achieving a Correct Classification Ratio (CCR) of up to 97% on testing data, including noisy MRI images. In a seminal paper, Bizhe Bai et al. [14] proposed a two-stage architecture, named YUSEG, which integrated the YOLO and U-Net models for cell instance segmentation. In their work, the researchers combine YOLOv5 for object detection and U-Net with an EfficientNet backbone for a high-performance segmentation.

3. Materials and methods

3.1. The proposed model

The proposed approach integrates two distinct architectures into an unified model to enhance the recognition of the trapezium.

The initial architectural framework employed is YOLOv8 [4]. YOLO is a real-time detection model that identifies and locates objects in images or videos with a single forward pass through a neural network. The system is designed to swiftly and precisely detect objects, such as the trapezium object of interest.

The second architecture employs this region to train a U-Net model [5] to segment the trapezium. This model is trained exclusively on the crop images, extracted in the preceding step, encompassing a smaller area of the image containing the trapezium. The U-Net model is composed of two components: an encoder and a decoder. The encoder is responsible for encoding the image into an intermediate representation by applying a set of convolutional layers in a progressive manner, followed by a pooling layer. The decoder, in turn, is tasked with reconstructing the mask segmentation from the intermediate representation.

3.2. Dataset preparation

The dataset utilized in this study encompasses 519 X-ray images of the hand, with a particular emphasis on the trapezium region (see Fig.1). These images were obtained from Avicenne Hospital AP-HP in Bobigny, France. These images play a crucial role in the development of a model based on deep learning, which aims to facilitate more precise of the trapezium position. The objective of this study is to train a model that can accurately detect and segment the trapezium in X-ray images. A medical expert meticulously annotated the contour of the trapezium, and this annotation is subsequently employed to assess the efficacy of the learned model.

It is imperative to acknowledge the challenges associated with X-ray images. The initial challenge pertains to the quality and variability of the images. It is imperative that the models exhibit excellent quality and incorporate images from diverse angles to facilitate the model's generalization capabilities and enhance its ability to accurately identify the trapezium. The second challenge pertains to the quantity of images. The architecture of deep learning technologies is intricate, necessitating substantial data sets for proper operation. To this end, data augmentation techniques were employed without compromising the integrity of the transformations, including rotation and flipping, to ensure a sufficient number of images for training.

3.3. Evaluation metrics

Several metrics [15] [16] are utilized to assess the efficacy of the trained models.

In the context of object detection models, mean average precision (mAP) is employed to assess the efficacy of the detection process. The value of mAP is defined as the average value of the average precision of values. The calculation of average precision entails the determination of the mean precision value for a given recall value. This value signifies the area under the precision-recall curve.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

In the context of object segmentation models, two metrics are used: intersection over union (IOU) and Dice score. IOU is a key evaluation metric for object detection, image segmentation, and computer vision tasks. It quantifies the similarity between a predicted bounding box or mask and the ground truth, or the actual object location. Like IOU, the Dice score also measures this similarity, but with a formula that doubles the importance of the intersection. This makes it more sensitive to small structures.

$$IOU = \frac{Area\ of\ intersection}{Area\ of\ union} = \frac{A \cap B}{A + B}$$

$$Dice\ Score = \frac{2\ x\ Area\ of\ intersection}{Area\ of\ union} = \frac{2\ x\ A \cap B}{A + B}$$

4. Experiments

4.1. Training details

In the following sections, we will present the results of our trained models. We will start with the detection results and present the results of YOLOv8, which was the only model used to detect the trapezoid object. For the segmentation part, we used various models, such as U-Net, SAM (Segment Anything Model), and Mobile-SAM. U-Net will be trained in two ways using a ResNet encoder with ImageNet weights (see Fig. 3): first, it will be applied directly to the image, and then to the region of interest. We will compare the two results to check if starting with a detection improves the segmentation results. We will also use other segmentation models, such as SAM and Mobile-SAM [17]. SAM is a real-time segmentation model that is trained on a billion masks and can only be used for inference; no training is needed. For this model to segment our desired object, it needs information about the object, such as a textual prompt, a points prompt, or a bounding-box prompt. SAM has demonstrated its capability of segmenting medical images using inference alone [18] [19]. In our case, we will use a bounding box extracted by YOLOv8 and pass it to SAM to segment the trapezium. We will use a lightweight variant of SAM, called Mobile-SAM, to segment our desired object. We will apply a comparative study to evaluate all segmentation models. All models are trained with 32 GB of RAM and an RTX 3500 Ada with 12 GB of VRAM.

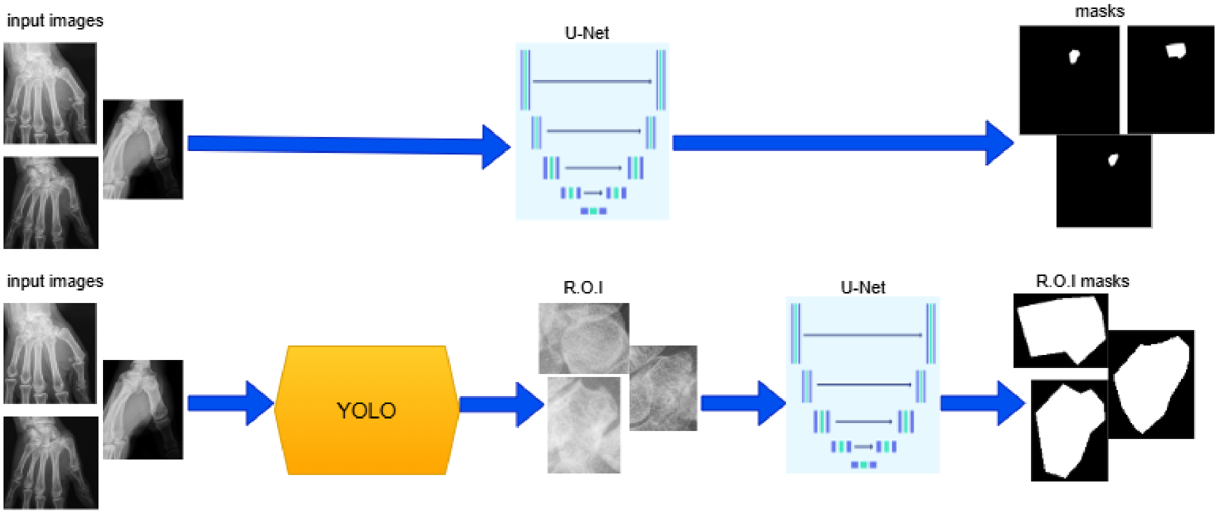

Fig. 3: Classical approaches vs. the two-step proposed approach.

4.2. Results and discussion

To conduct our experiment, we divided our dataset of 519 images into two parts: an 80% training set and a 20% testing set. Then, we augmented only the training dataset using the data augmentation techniques described in Section 3.B. We divided the dataset beforehand to avoid including the same patient in both sets.

We conducted a performance evaluation of the models. We made a comparison (see Tab.1) between a YOLOv8-based model, a U-Net model trained on images without a detection part, and a U-Net model trained on ROI extracted by the YOLO part. We also made a comparison with SAM models.

Table 1: The results of the different models.

Model	**mAP**	**IOU**	**Dice Score**
YOLOv8 (detection and segmentation)	99.50%	75.31%	84.86%
YOLOv11 (detection and segmentation)	99.10%	81.08%	89.02%
SAM	-	80.30%	88.80%
Mobile-SAM	-	80.50%	88.90%
U-Net	-	81.17%	89.52%
YOLOv8+U-Net	-	89.13%	94.21%

YOLOv8 and YOLOv11 demonstrate a high degree of efficacy in detecting the trapezium, as evidenced by their respective mAP scores of 99.5% and 99.1%. This indicates the models' precision and performance in locating the trapezium under various anatomical conditions and radiological image complexities. These results demonstrate YOLO's powerful performance in distinguishing between bones and the trapezium.

YOLOv8 and YOLOv11 can also perform segmentation tasks. YOLOv8 achieved a Dice score of 84.86%, considered an acceptable result, while YOLOv11 achieved a Dice score of 89.02%, demonstrating better performance in segmentation tasks. SAM and Mobile-SAM achieved an overall result of 89%, demonstrating good performance in identifying the trapezium without training and using only inference on the image with bounding box prompting. These results demonstrate the power of foundation models trained on large amounts of data. SAM is a foundation model that was trained on over 1 billion segmentation masks to learn promptable segmentation. Mobile-SAM demonstrates performance similar to SAM's, despite having an architecture that is 5x lighter and 7x faster, and being trained on only 1% of the original dataset used to train SAM. This difference in complexity and speed makes Mobile-SAM more suitable for mobile apps, where approximate detection is sufficient. The U-Net model's efficacy is demonstrated by its ability to attain a Dice score of nearly 90%, underscoring its remarkable capacity to segment objects. Additionally, reducing the image size provided to U-Net during training with YOLOv8 improved performance metrics, achieving a Dice score of 94.21% and an IOU of 89.13%. These results suggest that integrating YOLO and U-Net improves trapezium segmentation performance. Beginning with the detection stage is advantageous for helping the U-Net model accurately segment the trapezium.

A visual analysis of the found trapezium compared to the annotation reveals the limitations of this approach. Despite the obtained results, it is evident that the YOLO+U-Net model has difficulty perfectly segmenting the edges (see Fig.4), due to its difficulty inability to overlap with other bones. However, integrating the YOLO+U-Net model significantly improves the U-Net model's performance. Using the region of interest as input for U-Net optimizes its search.

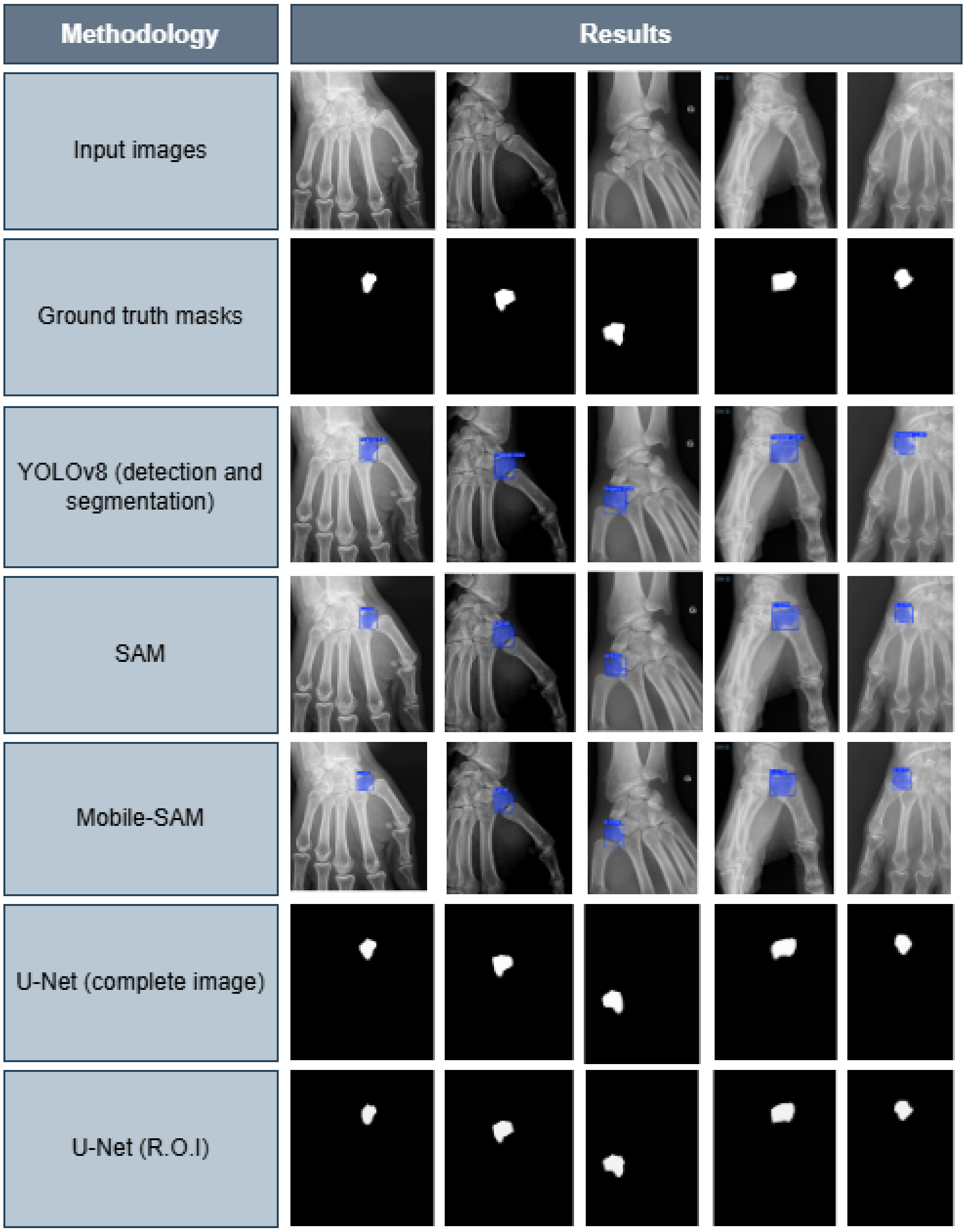

Fig. 4: Examples of trapezium identification using different approaches.

5. Conclusion

This paper demonstrates the challenge of segmenting a bone, such as the trapezium. This is especially the case when the bone to detect overlaps with other bones, making it difficult to identify its true edges. This is because the bone does not show up clearly in our 2D X-ray images (Fig.1). Additionally, observing the ground truth masks in Fig.4 reveals that the trapezium polygon takes a different form with different edges from a 2D perspective. In this paper, we propose a detection-segmentation strategy to improve trapezium bone segmentation. The first step is detection, which identifies the ROI containing the desired object. We used Yolov8 for this task, and it successfully detected the trapezium, achieving a mAP of 99.5%. Next, the ROI is sent to the U-Net model for segmentation and refinement, achieving a Dice score of 94.21%. This demonstrates the benefits of a two-step approach and the effectiveness of U-Net in identifying bones in X-ray images.

This study demonstrates how two deep learning models can improve object segmentation when working together. Specifically, YOLOv8 is used for object detection, while U-Net is used for segmentation. YOLO identifies the location of the trapezium on all images, and the detected region serves as input for the segmentation process. Despite variations in anatomical complexity and image quality, the U-Net model produces highly accurate segmentation results by effectively segmenting the trapezium. These results demonstrate the efficacy of deep learning architectures in identifying bones despite their complexity. Applying these technologies to medical subdomains, such as surgery, could help develop applications that identify bones or anomalies, enabling surgeons to plan procedures more effectively.

The focus of future work will be on leveraging artificial intelligence and deep learning technologies to enhance the planning process for TMC surgical procedures.

Statement on data and research ethics

This research project adheres to all research ethics requirements stipulated by the AP-HP. The study was approved by Avicenne Hospital's ethics committee with ID: CLEA-2025-434. Informed consent was obtained from all participants, and their privacy rights were strictly observed.

References

[1] L. Lajoinie and S. Barbary, « Total trapeziometacarpal prosthesis: Radio-clinical advice on cup implantation », Hand Surgery and Rehabilitation, vol. 42, May 2023.

[2] Y. LeCun, Y. Bengio and G. Hinton, « Deep Learning », Nature, vol. 521, pp. 436-44, May 2015.

[3] J. Ker, L. Wang, J. Rao and T. Lim, « Deep Learning Applications in Medical Image Analysis », IEEE Access, vol. 6, pp. 9375-9389, 2018.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, « You only look once: Unified, real-time object detection », Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[5] O. Ronneberger, P. Fischer and T. Brox, « U-net: Convolutional networks for biomedical image segmentation », Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, 2015.

[6] L. Ding, K. Zhao, X. Zhang, X. Wang and J. Zhang, « A Lightweight U-Net Architecture Multi-Scale Convolutional Network for Pediatric Hand Bone Segmentation in X-Ray Image », IEEE Access, vol. 7, pp. 68436-68445, 2019.

[7] P. Samothai, P. Sanguansat, A. Kheaksong, K. Srisomboon and W. Lee, « The evaluation of bone fracture detection of yolo series », 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 2022.

[8] A. Alam, A. S. Al-Shamayleh, N. Thalji, A. Raza, E. A. Morales Barajas, E. B. Thompson, I. de la Torre Diez and I. Ashraf, « Novel transfer learning based bone fracture detection using radiographic images », BMC Medical Imaging, vol. 25, p. 5, 2025.

[9] T. Aldhyani, Z. A. T. Ahmed, B. M. Alsharbi, S. Ahmad, M. H. Al-Adhaileh, A. H. Kamaland, M. Almaiah and J. Nazeer, « Diagnosis and Detection of Bone Fracture in Radiographic Images Using Deep Learning Approaches », Frontiers in Medicine, vol. 11, p. 1506686, 2025.

[10] H. Shen, Z. Dong, Y. Yan, R. Fan, Y. Jiang, Z. Chen and D. Chen, « Building roof extraction from ASTIL echo images applying OSA-YOLOv5s », Applied optics, vol. 61, p. 2923–2928, 2022.

[11] H. Du, H. Wang, C. Yang, L. Kabalata, H. Li and C. Qiang, « Hand bone extraction and segmentation based on a convolutional neural network », Biomedical Signal Processing and Control, vol. 89, p. 105788, 2024.

[12] D. Aibinder, M. Weisberg, A. Ghidotti and M. Weiss Cohen, « Enhanced Attention Res-Unet for Segmentation of Knee Bones », Mathematics, vol. 12, 2024.

[13] N. Iriawan, A. Pravitasari, U. S. Nuraini, N. Nirmalasari, T. Azmi, M. Nasrudin, A. Fandisyah, K. Fithriasari, S. Purnami, Irhamah and W. Ferriastuti, « YOLO-UNet Architecture for Detecting and Segmenting the Localized MRI Brain Tumor Image », Applied Computational Intelligence and Soft Computing, vol. 2024, pp. 1-14, February 2024.

[14] B. Bai, J. Tian, S. Luo, T. Wang and S. Lyu, « YUSEG: Yolo and Unet is all you need for cell instance segmentation », Proceedings of The Cell Segmentation Challenge in Multi-modality High-Resolution Microscopy Images, 2023.

[15] R. Padilla, S. L. Netto and E. A. B. da Silva, « A Survey on Performance Metrics for Object-Detection Algorithms », 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), 2020.

[16] D. Müller, I. Soto-Rey and F. Kramer, « Towards a guideline for evaluation metrics in medical image segmentation », BMC Research Notes, vol. 15, p. 210, 2022.

[17] Z. Ma, Y. Sun, S. Zhai, G. Lei and K. Zhang, « A Review of Segment Anything Model and Its Applications », 2024 IEEE International Conference on Unmanned Systems (ICUS), 2024.

[18] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, S. Liu, H. Chi, X. Hu, K.Yue, L. Li, V. Grau, D.P. Fan, F. Dong and D. Ni, « Segment anything model for medical images? », Medical Image Analysis, vol. 92, p. 103061, 2024.

[19] Y. Zhang, Z. Shen and R. Jiao, « Segment anything model for medical image segmentation: Current applications and future directions », Computers in Biology and Medicine, p. 108238, 2024.