Abnormality Detection and Segmentation in Musculoskeletal Radiographs

ECE 9202 -Advanced Image Processing Report

1st Amina Tabassum
Electrical and Computer Engineering
Western University
London ON, Canada
atabass4@uwo.ca

1st Xiang Li
Electrical and Computer Engineering
Western University
London ON, Canada
xli2824@uwo.ca

1st Thisali Senarath Rathnayake
Electrical and Computer Engineering
Western University
London ON, Canada
tsenarat@uwo.ca

***Abstract*—Abnormality detection in Musculoskeletal (MSK) radiographs has important clinical considerations and hence needs the expertise of radiologists. In order to help combat radiologist fatigue, we proposed a unique automated abnormality detection and localization employing weakly supervised segmentation. We used EfficientNet as the classification model, and Class Activation Maps (CAM) was used to visualize the decision-making process of the classifier. These activation maps were then used to create bounding boxes, which were used as segmentation masks for Segment Anything Model (SAM). SAM then segmented the region of abnormality with a dice score similarity coefficient value of 0.89. The work proposed in this study is preliminary and needs further improvements in terms of bounding box sizes and classification accuracy.**

***Index Terms*—Musculoskeletal , radiographs , automatica medical image segmentation , weekly supervised segmenation**

I. Introduction

The growing prevalence of musculoskeletal disorders has necessitated the development of advanced diagnostic tools to assist healthcare professionals in identifying abnormalities in medical imaging. Over 1.7 billion people worldwide suffer from chronic pain and disability due to musculoskeletal disorders. [1]. X-ray imaging is one of the oldest and most widely used diagnostic tools in detecting bone abnormalities, fractures, and joint-related issues [2]. However, the manual interpretation of X-ray images requires expert intervention but also is a time-consuming and error-prone task, which can lead to misdiagnoses and delayed treatment [3].

In recent years, deep learning models have revolutionized the analysis of medical images and have shown great promise in automating the process of diagnosis, by providing rapid and accurate results. Among these, Convolutional Neural Networks (CNNs), have gained significant traction for automating the classification of medical images [4]. CNNs, through their ability to automatically extract features from images, have revolutionized the landscape of computer-aided diagnostics (CAD), providing potential solutions for high-throughput, accurate, and efficient image analysis [5]. Among the various CNN architectures, EfficentNet has demonstrated, success in detecting and classifying abnormalities in medical imaging [6]. The architecture of EfficientNet allows for more efficient use of parameters compared to traditional CNNs, where typically only one of these factors is adjusted at a time [6].

While EfficientNet can successfully predict whether an image is normal or abnormal, it often operates as a black-box model [7]. Since understanding why a model made a specific decision is crucial for the medical professional's decision-making and diagnosis, the lack of interpretability hinders their full integration into clinical practice [8]. In this paper, while classifying whether an X-ray is normal or abnormal, we address the interpretability challenge by using Class Activation Mapping (CAM) to identify the regions of the image that contribute most to the model's decision. These techniques will help us highlight the areas of the X-ray that are indicative of abnormalities, thereby improving the trustworthiness and transparency of the AI system.

II. Related Work

Many studies have been done in computer-aided diagnostics (CAD) to detect abnormalities in X-rays. Traditional methods, such as image processing and feature extraction, are replaced by end-to-end deep learning models capable of learning relevant features directly from raw image data, making the process more efficient and accurate [9]. Among the deep learning techniques, CNNs are researched extensively for image classification and localization tasks because they can automatically learn spatial hierarchies of features such as edges, textures, and shapes which are crucial for identifying abnormalities [10]. Architectures such as AlexNet, ResNet, VGGNet, and InceptionNet have been studied to capture unique patterns in X-ray images, leading to accurate detection and classification of bone abnormalities [11] [12] [13]. Among these, Efficinet-based models have shown high accuracy and precision in detecting bone fractures compared to other models [14] [6].

One of the challenges with deep learning in healthcare is the lack of interpretability [15]. Medical professionals need to understand why a model makes a particular decision, especially when dealing with critical diagnoses [16]. To address this, methods like Class Activation Mapping (CAM) and

Score-weighted Class Activation Mapping (SAM) have been introduced to provide interpretability to CNNs [17]. CAM generates a heatmap that highlights the regions of the image that most influenced the model's decision [18] [19]. SAM can enhance the segment and localization of abnormalities, making it particularly useful in identifying specific regions of the X-ray that are most indicative of disorders [20].

III. Dataset

Musculoskeletal Radiographs (MURA) is one of the largest publicly available MSK radiographic image datasets [21], comprising x-rays of seven upper body MSK extremities, including elbow, finger, forearm, hand, humerus, shoulder, and wrist. It comprises 40,561 images from 14,863 studies, where each study is manually labeled as either normal or abnormal by board-certified radiologists at Stanford University Hospital in the diagnostic radiology environment between 2001 and 2012.

IV. Methodology

A. Data Preprocessing

The clinical images have a different resolution and aspect ratio. The data preprocessing pipeline is implemented such that the images are resampled to 224x224x3 to ensure consistency across all images. It is then followed by standardized normalization to limit all the pixel values to the 0 and 1 range and reduce computation load.

B. Deep Learning Models

1) Clasification Models: Residual Networks (ResNet) is a deep convolutional network that uses skip connections to mitigate the vanishing gradient problem. We used pre-trained ResNet50 for abnormality detection. We also used EfficientNet B3 for abnormality detection. It uses compound scaling to balance network depth and width, and hence is more efficient and accurate in image classification tasks.

2) Class Activation Maps: Since the goal of the project is segmentation and segmentation masks are not provided with this dataset, we are proposing a unique weakly supervised segmentation approach by using CAM to generate segmentation masks. CAM is one of the tools to visualize and understand the decision-making process of the classifier, i.e., EfficientNet B3. In order to visualize the regions of an image that contribute most to the model's prediction, CAM is used and it comprises three main steps:

- Global Average Pooling (GAP): As shown in Figure 1, after the last convolutional layer, GAP is applied to the feature map and hence computes the average value of all pixels in the feature map.
- Weighted Sum: The average values from GAP are then multiplied by the weights of the output layer, and the weighted sum highlights the most important regions for the model's prediction.
- Upsampling: The resulting activation map is then upsampled to the size of the input image, i.e., 224x224, creating a heatmap that shows the regions within an image that contributed to the classification decision.

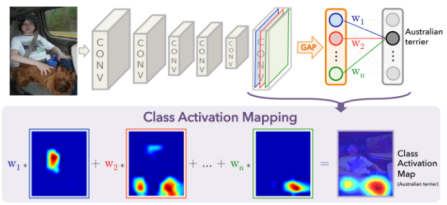


Fig. 1: CAM architecture

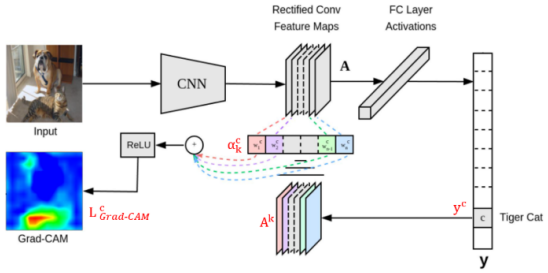


Fig. 2: Grad-CAM architecture

3) Gradient Class Activation Maps (Grad-CAM): In order to improve the performance of CAM and use a pre-trained model, Grad-CAM is used, and its architecture is shown in Figure 2. Its working principle is similar to CAM. It first uses the original image and inputs it into the CNN model, which consists of a feature extraction part and a classification part. We select the last convolutional layer of the feature extraction part. Unlike CAM, GAP is not added at the end; instead, the fully connected layer is directly connected. During backward propagation, we calculate the gradients of the last convolutional layer's feature maps, where each pixel corresponds to the partial derivative of a specific category.

It is then followed by finding brightest point and generating 50x50 bounding box around the region of highest importance as shown in figure 3.

Fig. 3: Bounding boxes to show the abnormality using Grad-CAM

4) SAM: SAM is developed by Meta AI [22] and consists of the following key components:

- Promptable Segmentation: It can segment objects using different prompts such as points, text descriptions, or bounding boxes. We use a bounding box to prompt the segmentation mask.
- Zero-Shot Generalization: It is trained on 11 million images and can accurately segment images without prior training.

Figure 4 shows its overall architecture. Its working principle is:

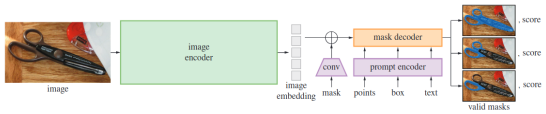


Fig. 4: SAM overview

- Image Embedding: It is a network that can input any picture and generate a C×H×W image embedding. It is obtained by pretraining a model called MAE (ViT), using 14×14 windowed attention and four equally spaced global attention blocks. The prompt is computed only once per image, not per prompt, to improve real-time processing capability. Following other standard practices, it will then generate 256-channel embeddings.
- Prompt Processing: It processes the input prompts to identify the regions of interest and generate 256-dimensional vector embeddings using different encoding.
- Segmentation Mask Generation: Based on the prompts and image embeddings, SAM generates segmentation masks for objects which can be further refined.

V. Results

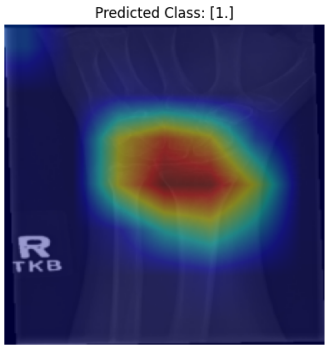


Fig. 5: CAM generated from EfficientNet B3 classification results

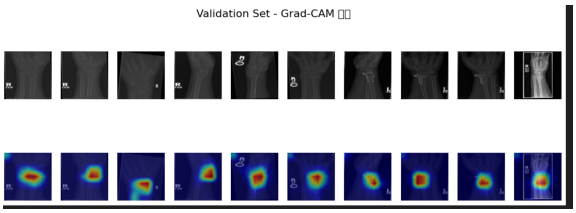


Fig. 6: Results of Grad-CAM

VI. Discussion

The goal of this project was to identify and segment abnormality in MSK radiographs. For classification, we trained both ResNet and EfficientNet B3. ResNet deals with the vanishing gradient problem in CNNs while EfficientNet has better efficiency because of compound scaling. ResNet gave 71.93% accuracy while classification accuracy for EfficientNet-B3 was 95.92%. We used EfficientNet B3 as our final classification model because of its better efficiency and performance. The dataset is labeled as normal and abnormal images, which is good for abnormality detection. However, for the chosen problem, in order to correctly identify the abnormal region, semantic segmentation is required and for that, segmentation masks are needed. This dataset does not have segmentation masks, so we used CAM to create a bounding box around the abnormal region and use those bounding boxes as segmentation masks for SAM. CAM helps us visualize the regions within an image that the model is using for classification. Feature maps with the highest value serve as the center point for creating the bounding box. This bounding box does highlight the region of abnormality. However, it does not segment the abnormality. SAM is trained on 11 million images and 1.1 billion mask annotations and is chosen because it has the additional feature of zero-shot generalization. It allows promptable segmentation. Figure 9 demonstrates the result when there is no prompt provided. In this case, it segments all the regions within the image. However, our goal was just to segment the abnormal region and for that, the model needs a prompt or bounding box. We used the bounding box generated by CAM and used those as segmentation masks

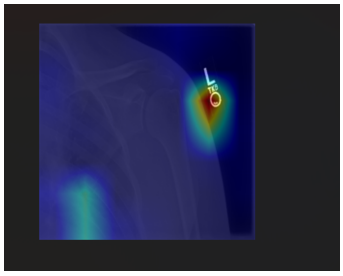


Fig. 7: Example for error in generated heat map

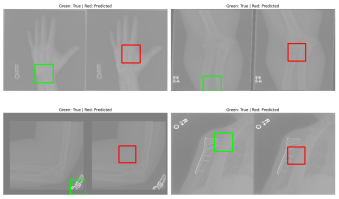


Fig. 8: Bounding boxes to be used as segmentation masks for SAM

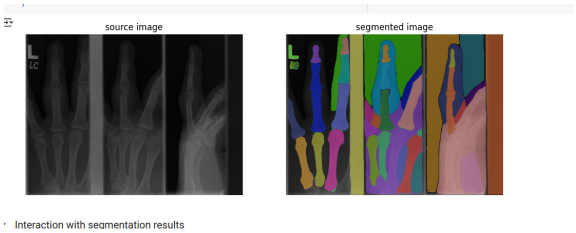
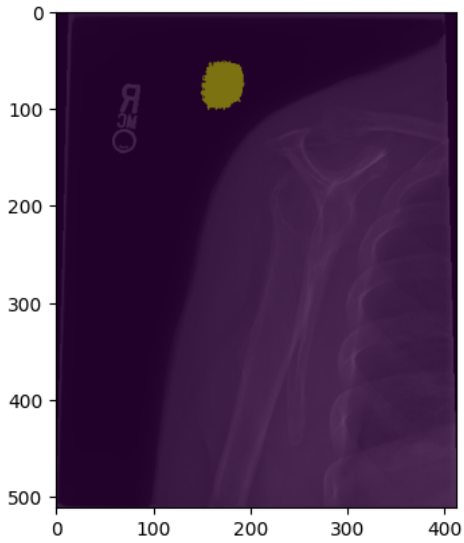
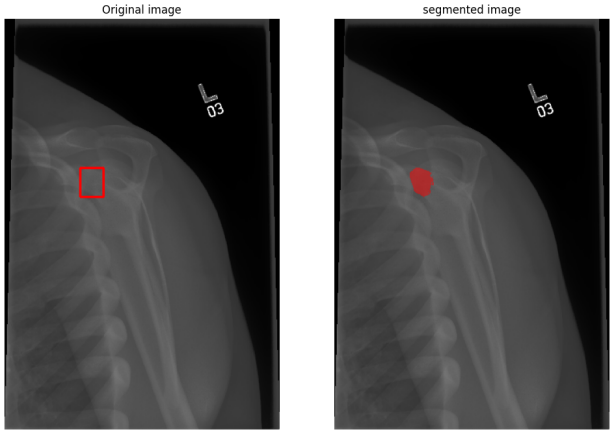


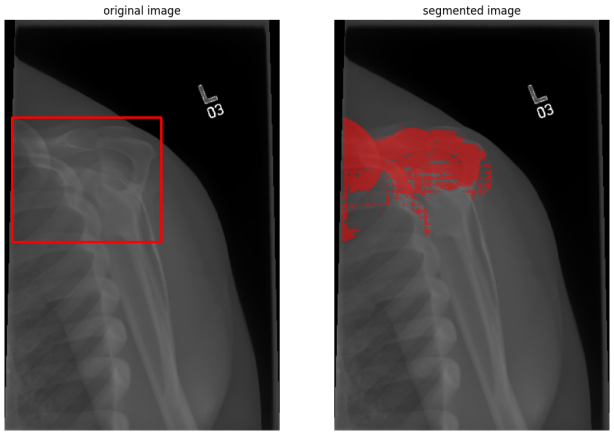
Fig. 9: Results after applying SAM without bounding box



(a) SAM segmentation from bounding box that does not precisely localize abnormality



(b) SAM segmentation from small bounding box



(c) SAM segmentation from precise bounding box

Fig. 10: Results after using bounding boxes generated from CAM as masks to the SAM model

for SAM. In this way, this unique weakly supervised segmentation approach performs segmentation. The performance metric for abnormality classification is accuracy. Resnet gives 94.32% and EfficientNet-B3 gives 95.92% accuracy. We used EfficientNet-B3 to predict bounding boxes and used these as masks for SAM. The performance metric for SAM is the dice similarity coefficient and our model gives a dice score value of 0.89. Currently, one of the limitations of our proposed approach is that we are using CAM to generate segmentation masks for SAM. The generated segmentation masks are not very precise as it can be seen from Figure 7,10 and 8. The smaller bounding box results in segmenting a smaller region of abnormality, not the entire region, and hence it affects the dice similarity score coefficient value. In order to improve the results, future work is to improve prediction accuracy and hence the bounding box size generated by CAM which will eventually improve the segmentation results.

VII. Conclusion

The goal of the study was to identify and precisely localize the abnormality in radiographs. We proposed a unique architecture that employs weakly supervised segmentation for identifying abnormalities in medical images.

References

[1] L. M. Waddell *et al.*, "Responsiveness of subjective and objective measures of pain and function following operative interventions for musculoskeletal conditions: A narrative review," *Arthritis care & research*, vol. 76, no. 6, pp. 882–888, 2024.

[2] J. D. Howell, "EARLY CLINICAL USE OF THE X-RAY," *Transactions of the American Clinical and Climatological Association*, vol. 127, pp. 341–349, 2016.

[3] L. C. Lawrence, *Machine Learning Integration in Healthcare: Human Error Reduction in Breast Cancer Screening*. PhD thesis, 2024. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2024-07-11.

[4] N. M. Pillai, S. Manimala, A. S. Rongali, A. Gugnani, A. S. Kumar, and G. V. Sriramakrishnan, "Automated classification of medical images using convolutional neural networks," *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, 2024.

[5] H. Yu, L. Yang, Q. Zhang, D. Armstrong, and M. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021.

[6] H. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet," *IEEE Access*, vol. PP, pp. 1–1, 2022.

[7] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2017.

[8] M. Khatri, Y. Yin, and J. Deogun, "Enhancing interpretability in medical image classification by integrating formal concept analysis with convolutional neural networks," *Biomimetics*, vol. 9, 2024.

[9] J. Walsh, N. O' Mahony, S. Campbell, A. Carvalho, L. Krpalkova, G. Velasco-Hernandez, S. Harapanahalli, and D. Riordan, "Deep learning vs. traditional computer vision," 04 2019.

[10] L. Nanni, S. Ghidoni, and S. Brahnam, "Ensemble of convolutional neural networks for bioimage classification," *Applied Computing and Informatics*, vol. ahead-of-print, 06 2018.

[11] M. Kabir, T. J. Tahiti, and T. A. Prome, "A comparative study of certain convolutional neural network architectures for x-ray image analysis in bone fracture detection and identification," *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, pp. 1–8, 2024.

[12] A. Saad, U. U. Sheikh, and M. S. Moslim, "Developing convolutional neural network for recognition of bone fractures in x-ray images," *Advances in Science and Technology Research Journal*, 2024.

[13] V. V, S. K. Natarajan, A. M, N. P, M. C. A, and N. M. Hosahalli, "Efficient cnn-based bone fracture detection in x-ray radiographs with mobilenetv2," *2024 2nd International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)*, pp. 72–77, 2024.

[14] I. Bouslihim, W. Cherif, and M. Kissi, "Application of a hybrid efficientnet-svm model to medical image classification," in *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pp. 1–6, 2023.

[15] D. Jin, E. Sergeeva, W. Weng, G. Chauhan, and P. Szolovits, "Explainable deep learning in healthcare: A methodological survey from an attribution view," *WIREs mechanisms of disease*, p. e1548, 2021.

[16] J. P. Amorim, P. Abreu, A. Fernández, M. Reyes, J. A. M. Santos, and M. Abreu, "Interpreting deep machine learning models: An easy guide for oncologists," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 192–207, 2021.

[17] H. Hu, R. Wang, H. Lin, and H. Yu, "Unioncam: enhancing cnn interpretability through denoising, weighted fusion, and selective high-quality class activation mapping," *Frontiers in Neurorobotics*, vol. 18, 2024.

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, pp. 336–359, Feb. 2020.

[19] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," 2016.

[20] F. Chen, L. Chen, H. Han, S. Zhang, D. Zhang, and H. Liao, "The ability of segmenting anything model (sam) to segment ultrasound images.," *Bioscience trends*, 2023.

[21] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, *et al.*, "Mura: Large dataset for abnormality detection in musculoskeletal radiographs," *arXiv preprint arXiv:1712.06957*, 2017.

[22] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023.