Robust MSFM Learning Network for Classification and Weakly Supervised Localization

Komal Kumar, Balakrishna Pailla, Kalyan Tadepalli, Sudipta Roy
Artificial Intelligence & Data Science, Jio Institute, Navi Mumbai-410206, India

INTRODUCTION

- Accurate classification and precise localization of bone fractures in X-ray images are crucial for medical diagnosis and treatment.
- Current state-of-the-art classification methods prioritize accuracy but lack reliability.
- Existing approaches for fracture localization often require costly annotated data.
- We proposed Multistage Feature Map (MSFM) learning network.

- This approach aims to enhance both classification accuracy and localization in weakly-supervised manner without relying on costly annotated bounding boxes.
- A feature augmentation technique is introduced to focus the model on discriminative regions, refining localization.

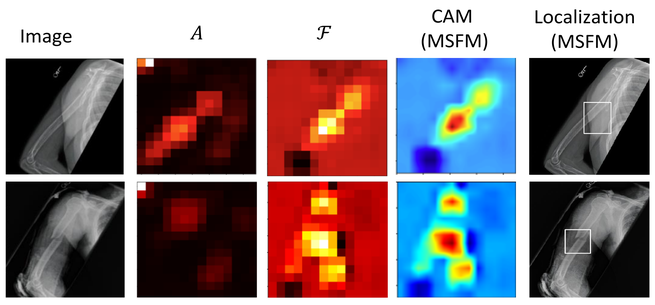

Figure 1. This image shows components of the Class activation map from a classifier, activation map (A), and weighted feature map (F).

- Experiments conducted on the MURA dataset, covering diverse X-ray images, showcase the effectiveness of the MSFM model.
- The MSFM model holds potential for advancing the medical image classification and localization.

METHODOLOGY

Class activation map decomposition [1]

$$CAM(X) = W_{cl}^{T} F(X) = \| W_{cl} \| \| F(X) \| \cos\theta \leq \| W_{cl} \| \cdot \| F(X) \|$$

For activation map, $A = \|F(X)\|$. $\widehat{M_{(i,j)}}$ is the possible resign to localize.

$$\widehat{M_{(i,j)}} = \begin{cases} 1, & \text{if } A(i,j) \geq \bar{A} \\ 0, & \text{if } A(i,j) < \bar{A} \end{cases}, \text{ and } \bar{A} = \frac{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} A(i,j)}{h \times w}$$

Multi-stage feature map learning (MSFM) network architecture

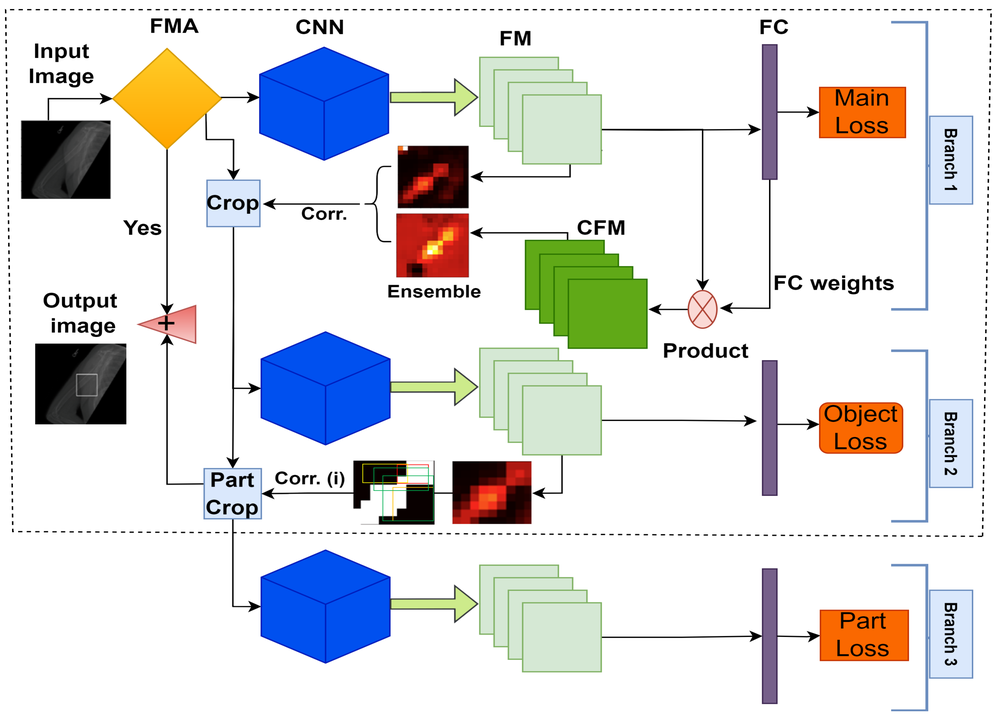

Figure 2. The full MSFM architecture of the multi-stage network that consists of three stages starting with CNN in the training phase and dotted black box is the structure in the test phase. The class feature map (CFM) represents the product of Fully Connected (FC) weights and FM. "Yes" represent the condition to the output for a specific class (fractured image in our case), and '+' in the output image is to paste the combined part in the original image. The CNN and FC layers of the same color represent parameter sharing.

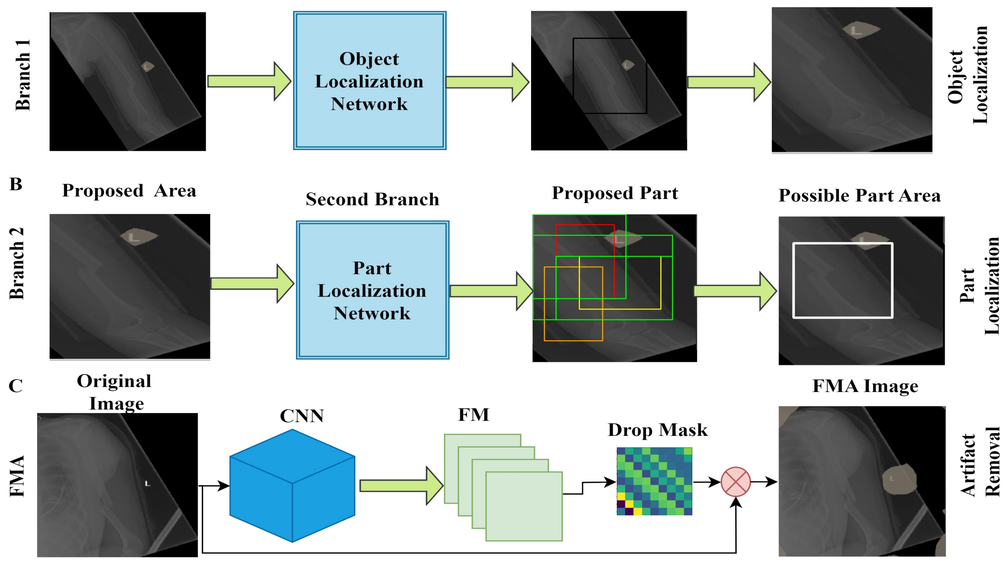

Figure 3: Visual representation of the main model's core operational objectives.

A. This network trains to localize the key area in a weakly supervised manner. The informative area is represented by the bounded box.
B. This pipeline is based on the second stage of MSFM. Here red, orange, yellow, and green colors indicate the order of the proposed window, and the white window indicates the final window by combining all the windows using discounted factor method in the same order.
C. This figure represents the feature map-based augmentation. CNN of different colors represents the different pre-train CNN for removing the high-intensity values based on the drop mask that produces FMA with the product of the original image.

EXPERIMENTAL RESULTS

Table 1. Comparison based on classification results to baseline models. The highest accuracy of each part is highlighted. MSFM-R-n is our proposed method with the backbone of R=Resnet for n= [18, 34, 50].

Part Name	Elbow	Finger	Forearm	Hand	Humerus	Shoulder	Wrist	Av.	Tumor
Res Net	0.79	0.65	0.73	0.72	0.61	0.75	0.78	0.72	0.94
Dense Net	0.85	0.73	0.58	0.75	0.76	0.75	0.76	0.74	0.76
Inception	0.69	0.61	0.70	0.64	0.71	0.75	0.76	0.70	0.90
APGA [2]	0.83	0.77	0.83	**0.79**	0.86	0.76	0.84	0.81	-
MSFMR50	**0.87**	**0.81**	**0.84**	**0.79**	**0.89**	**0.80**	**0.87**	**0.85**	**0.97**
MSFMR18	0.84	0.73	0.68	0.72	0.85	0.71	0.81	0.77	0.89
MSFMR34	0.80	0.78	0.77	0.79	0.82	0.75	0.82	0.79	0.94

Table 2. Results based on Cohen's kappa statistic score and comparison to the baseline technique and three unbiased radiologists. RD1, RD2, and RD3 represent three radiologists. The highest Cohen's kappa score in all RDs/models and the models marked as bold. Av. is the average kappa score of all parts.

RDs/models ↓ Parts Name →	Elbow	Finger	Forearm	Humerus	Shoulder	Wrist	Hand	Av.
RD1 [3]	**0.85**	0.30	0.79	0.86	**0.86**	0.79	0.66	0.74
RD2 [3]	0.71	0.40	**0.80**	0.73	0.79	**0.93**	**0.92**	0.72
RD3 [3]	0.71	0.41	0.79	**0.93**	**0.86**	**0.93**	0.78	**0.77**
Dense	0.71	0.38	0.50	0.60	0.72	0.62	0.55	0.58
Res Net	0.59	0.34	0.46	0.22	0.50	0.54	0.39	0.44
Inception	0.39	0.26	0.41	0.43	0.50	0.53	0.17	0.42
MSFMR34	0.61	0.57	0.55	0.65	0.50	0.64	0.54	0.58
MSFMR18	0.68	0.47	0.37	0.71	0.43	0.61	0.38	0.54
MSFMR50	**0.72**	**0.62**	**0.64**	**0.77**	**0.73**	**0.67**	**0.56**	**0.70**

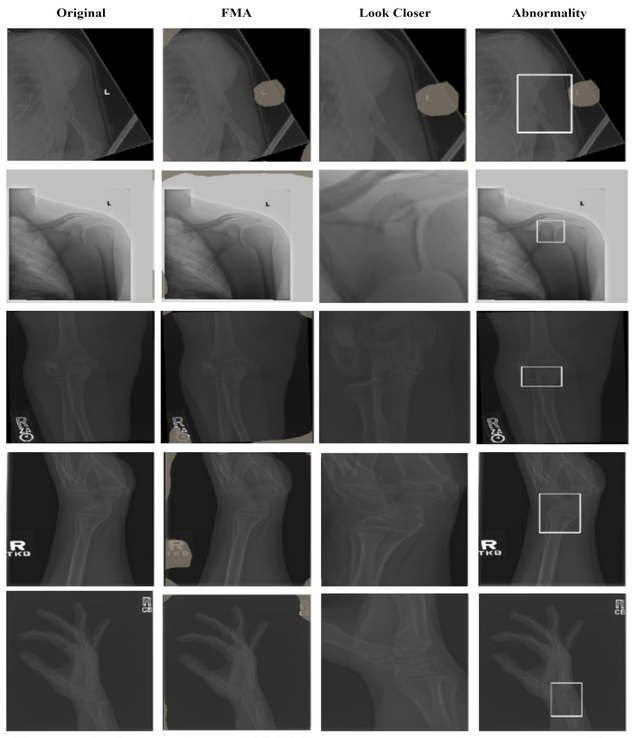

Figure 4. Output from MSFM at each step of the model. The first column consists of an original image from each part followed by the columns of FMA, Look Closer in the original image, and abnormality detection in the original image.

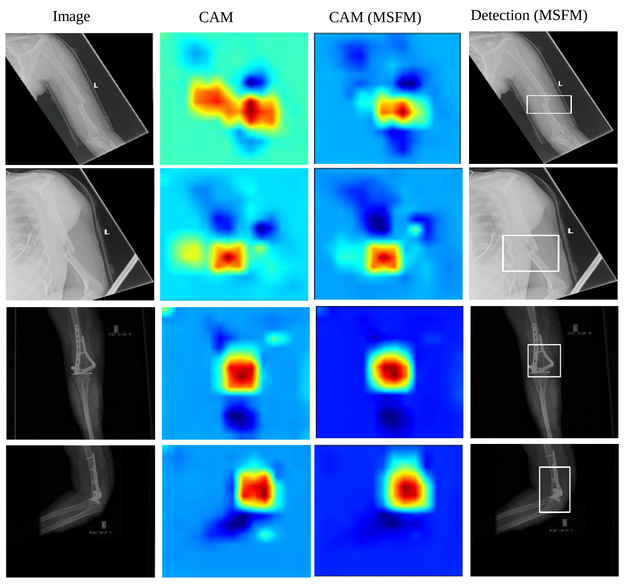

Figure 5. Visual comparison with weakly supervised CAM.

Table 3. Localization performance of the first stage of MSFM on CUB.

Methods	Training from scratch	PCL (%)
ACOL [4]	No	46.0
ADL [5]	No	62.3
SCDA [6]	No	76.8
MMAL [7]	No	71.1
MSFM (ours)	Yes	77.4

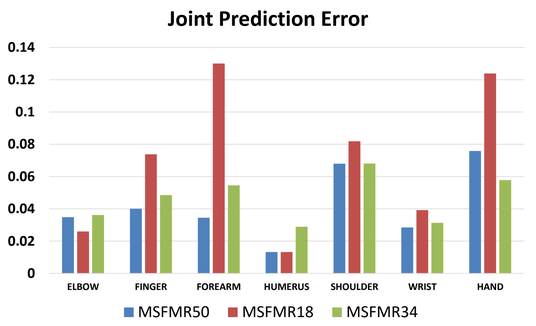

Figure 6. Comparison of MSFM on backbone models using JPE.

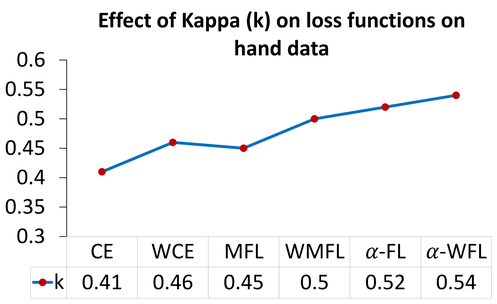

Figure 7. Effect of loss function on Cohen's kappa statistic.

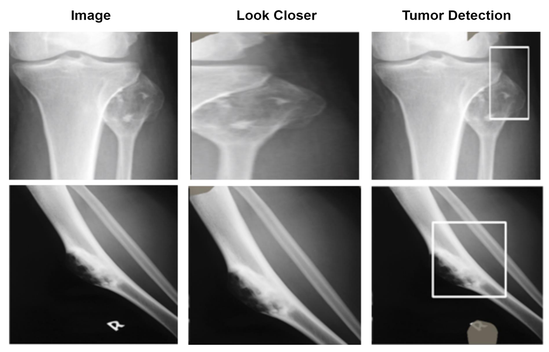

Figure 8. Tumor detection output visualization.

SUMMARY AND CONCLUSIONS

- The proposed Multistage Feature Map (MSFM) learning network offers a robust solution for bone fracture classification and localization in X-ray images, surpassing the limitations of existing methods that rely on annotated bounding boxes.
- MSFM builds upon CAM, utilizing feature map-based activation (A) and class feature map (F) to capture comprehensive object information. By incorporating multiple stages, MSFM enhances robustness to image variations and extends the applicability of CAM to the entire object.
- The paper's contributions encompass the development of the MSFM-net architecture, a novel weakly supervised fracture localization technique, a feature augmentation method, and an extensive analysis of loss functions.
- Through comprehensive experiments, the proposed MSFM model achieves state-of-the-art results on the standard MURA dataset, which includes the elbow, finger, forearm, humerus, shoulder, wrist, hand, and bone tumor dataset., enhancing both classification accuracy and detection visualization.
- The advancements presented in this paper hold promise for improving the efficiency and accuracy of disease localization in X-ray images, ultimately benefiting patient care and outcomes in the field of medical image analysis.
- Our future endeavors involve extending the application of the MSFM model to a wide range of medical and natural image datasets. This expansion will encompass moving beyond pretrained models and exploring more advanced backbone models.

References

[1] Eunji Kim et al., "Bridging the gap between classification and localization for weakly supervised object localization," CVPR 2022.
[2] Kaiyang Cheng et al., "Adversarial policy gradient for deep learning image augmentation," MICCAI 2019.
[3] Pranav Rajpurkar et al., "Mura: Large dataset for abnormality detection in musculoskeletal radiographs," arXiv 2017.
[4] Xiaolin Zhang et al., "Adversarial complementary learning for weakly supervised object localization," CVPR 2018.
[5] Junsuk Choe and Hyunjung Shim, "Attention-based dropout layer for weakly supervised object localization," CVPR 2019.
[6] Xiu-Shen Wei et al., "Selective convolutional descriptor aggregation for fine-grained image retrieval," IEEE TIP 2017.
[7] Fan Zhang et al., "Multibranch and multi-scale attention learning for fine-grained visual categorization," MMM 2021.