Journal of Information Hiding and Multimedia Signal Processing ©2021 ISSN 2073-4212
Ubiquitous International Volume **12**, Number **4**, December **2021**

Predication of Most Significant Features in Medical Image by Utilized CNN and Heatmap

Lubab Ahmed Tawfeeq[1], Samera Shams Hussein[2] Mahir jasem Mohammed[3] Shaimaa Sattar Abood[4]

[1]Department of Computer Science, College of Education for Pure Science
University of Baghdad, Baghdad, Iraq
lubab.a.t@ihcoedu.uobaghdad.edu.iq
[2]Department of Religious Education
Baghdad, Iraq

Received July 2021; revised September 2021

Abstract. *The growth of developments in machine learning, the image processing methods along with availability of the medical imaging data are taking a big increase in the utilization of machine learning strategies in the medical area. The utilization of neural networks, mainly, in recent days, the convolutional neural networks (CNN), have powerful descriptors for computer added diagnosis systems. Even so, there are several issues when work with medical images in which many of medical images possess a low-quality noise-to-signal (NSR) ratio compared to scenes obtained with a digital camera, that generally qualified a confusingly low spatial resolution and tends to make the contrast between different tissues of body are very low and it difficult to computed and recognized dependably. In this paper, we target to utilized CNN and heatmap to recognized most significant features that the network should focus on it. depending on class activation mapping. The goal of this study is to develop an approach that can determine the most significant features from medical images (such as x-ray, CT, MRI) through gradient the different tissue accurately by made use of heatmap. In our model, we take the gradient with regard to the final convolutional layer and after that weigh it towards the output of this layer. The model is based upon class activation mapping. However, the model is differed from traditional activation mapping based methods, that this model is the dependent on gradients via obtaining the weight of all activation map via make use of it is forward passing score over target class, then the final result is apart from linear combination of activation and weights. The results appears that the model is successfully distortion heat map of tissues in various medical image techniques and obtained better visual accuracy and fairness for interpretation the decision-making procedure.*
Keywords: Neural Networks, Deep learning, Convolutional Neural Networks, Medical Images, CNN

1. Introduction. Previously, medical imaging grew to become a common for the field of diagnosis and medical treatment which is utilized for the visual observation to recognize the features of organs and tissues and determine abnormality. With expanded usage of modern medical imaging such as x-ray, MRI, mammography, or CT scanner, there is a requirement of automatic processing for diagnosed images and detect abnormality [1]. The usage of image processing in the medical images poses an essential role in analysis for wide range of techniques because of three factors. The first factor is the fact that the almost all medical images have a weak NSR, which generally causes a bad spatial resolution. One example is the ultrasonic images, the speckle noises that represent the scattering effect of the ultrasonic beam where it is apart from microscopic tissues in homogeneities, could most certainly mask the presence of lower contrast lesions and minimizes the capability of a human observer to address the final feature. Taking into consideration these reasons, the image pre-processing strategies are used to reduce the noises and blurs the medical images are significant. The second factor is the modifications of medical image content should be done in a considerably controlled and dependable way that should not be confused in making the decision of clinical. As one example, as it is regularly appropriate to filter out the localized bright patches of the noise, there is certainly a must be aware for the case of mammography that should not clear away microcalcifications. To be able to achieve this goal, various sophisticated operations will need to be done. A considerable step in this operation is image segmentation, which can depend on the pixels in the same class which have equivalent pixel amount independent of their sites [2]. However, in MRI, the homogeneity during the magnetic field generally provides increases to intensity nonuniformity creature. This regular creature shows itself as a various, gradual, and smooth change in pixel valuations of an image and could have bad effect on the intensity efficiency based automatic segmentation approaches. However, the intensity variance that related to the field in- homogeneity, there could possibly be a loss of tissue specification of features that appear normally in MRI scan, probably for the same person attained on the identical scanner utilizing the similar protocol. The third factor is that the details obtained from two images achieved in the clinical track of events is frequently of a complementary characteristic, and a suitable integration of significant data obtained from the separated images may possibly be more desired. As a result, there is a potential benefit of medical images enhancing in such a way in which these images are merged and com- pared [3]. Image processing using neural networks (NN) typically is categorized to two classes: image restoration (Which includes denoise and image enhancement) and image reconstruction [4]. However, there are various neural networks can be utilized for this purpose, such as feed- forward NN, fuzzy NN, adaptive resonance theory NN, Hopfield neural network, self-organizing feature NN, Convolution NN, etc. [5]. Convolution NN (CNN) is a part of deep learning, which has lately taken a significant role in field of computer vision for the two learning tasks: supervised and unsupervised. The CNN includes three main parts which are, fully connected layers, pooling and convolutional layers. The key role of the convolutional layer is to determine some aspects such as lines, edges and patterns, etc. Every CNN hidden layer comprises of convolutional layers which is convolve input array together with weight-parameterized convolutional kernels. Typically, the multiple kernels produce multiple feature images and help to make have great results in several vision tasks including classification and segmentation. Class Last year's many studies have been achieved to improve CNN for side of accuracy and training/recognition time.one of these fields is focused on improving saliency map. Activation Mapping (CAM) is one strategy utilized for generating heat maps to focus on class-specific regions of images [6]. It is important to visualize exactly where a NN is looking since it can help researchers to have an understanding if the NN is looking at best suited parts of the image, or if the NN is not focused on appropriate part [7, 8]. In following some examples showing how neural networks might may be focus on wrong place in cases when making a classification decision: The CNN will classify an image when some structure is associated with a specific object. For example, the CNN is classified an image as "train" in cases that in reality it is searching for "train tracks" (which means that it will wrongly classify a train tracks photo alone as "train"). Another example that related to medical imaging, in case when utilized CNN for detect of abnormality (illness) from x-ray images of chest. The CNN may classify a chest x-ray image as "high probability of illness" depending on the physical appearance of disease, however in real x-ray images there is a metallic "R or L" token may place on that patient's right or left shoulder respectively. The key is that this token is just inserted directly on the patient's body in case they are placing down, as well as the patient is mainly going to be placed down for the x-ray if they are very weak to stand up. Hence, the CNN star learned the relationship between "metallic "R or L" token on shoulder" and "patient which he is too sick to stand". However, there is a need to make CNN focus on significant area of patent's image to discover the actual visual signs of disease and ignore not important features such as metal tokens.

2. Literature Survey.

2. Literature Survey. ANN systems are commonly used in various medical classification tasks, where the CNN is a great feature extractor, thus, utilizing it to categorize medical images can avoid expensive and complicated feature engineering. Thus, many studies had been achieved to utilize CNN in that field with improving these methods for side of accuracy, minimizing computational re- quirement and reducing time of training and recognition. Dosovitskiy and Brox (2015) [9], proposed a method to invert the image representations by using up-convolutional networks and show that this produces a lot more or a lot less accurate reconstructions of the main images, based on the invariance level of the feature representation. The NN absolutely learns natural image priors that make it possible for retrieval of information that is usually lost in the feature representation, for example brightness or color in SIFT or HOG. Their method can evaluate the visual encoding of CNNs via inverting deep features at several layers. Even while this approach can invert the Fully Connected (FC) layers, it can just show what data is being preserved in the deep features while not highlighting the relative significance of this data. The results show that this method it is fast at test time and doesn't need the gradient of the feature representation for being inverted. As a result, it can be utilized to virtually any image representation. Zhou et al. (2016) [10], proposed an ap- proach known as Class Activation Mapping (CAM) for determining discriminative regions utilized by a restricted class of image. Kermany et al. (2018) [11], make use of inception (V3) using trained weight from ImageNet and transfer learning using a medical image dataset including 108,312 images of optical coherence tomography. The result shows that this model got an average accuracy of reach up to 96.6. Zhao et al. (2019) [12], used an enhanced CNN for finger vein recognition, where they enhanced the accuracy of finger vein recognition algorithm (algorithm based on CNN that used curvature gray images) that placed under small samples, In their work, they first compute the curvature of a finger vein image making use of a 2D Gaussian template. After that they extracted two gray images from the finger vein image using several scales and put these two images to get the final curvature gray image. The curvature gray image is used as input for improved convolutional neural network and it is trained in order to recognize the identity of the input curvature gray image. The results show that this approach is efficient and better than existing schemes. Kuo et al (2021) [13], proposed a new stacked deep convolution network to enhance the overall performance of single image super-resolution (SISR). They build a simple deep-learning unit (BDU) to be a model foundation, and they train every BDU for different resolutions. Following the training process, they stack the BDUs to be a deeper model for super-resolution. Since the BDUs have been trained independently, thus the computation cost decreases significantly. The results show that the new super-resolution approach achieved sufficient performance for each side of quantitative measure indexes and visual quality. .

3. The Proposed Method In order to implement the heatmap, we use a method known as Gradient Class Activation Map (GradCAM) which is based upon work done [14]. The concept behind it is particularly simple; it is used to find the significance features and gradient it in such a way that be able to detect each tissue based on its gradient color. Based on concept of Grad-CAM, it utilizes the gradient info streaming in the latest convolutional layer of the CNN to recognize the benefit of each neuron with regard to the decision of interest. Even though the proposed model is regularly used and is usually utilized to visualize most activation in a deep network. In this method, a model based Dilated Convolutional Neural Network (DCNN) has been used to recognize feature points of medical images models. However, an information is required for object identification.

Based on Grad-CAM theory, by considering a CNN neural network *N=f(I)* that get an input ($I \in R^d$) and outputs a possibility distribution *N*, and we denote N^c as the possibility of having classified as class C_i. With a given layer Y, let it T_Y denotes the activation of layer Y. Specifically, if Y is selected as a convolutional layer, let T_Y^k signify the activation map to the k_{th} channel. In addition, signify the neuron weight k_{th} at layer Y connecting 2 layers Y and $Y+1$ as $w_{Y,Y+1}$. With take into account a model *f* consists of a global pooling layer Y which will take the output coming from the last convolutional layer Y- 1 and passes the pooled activation towards the FC layer of Y+ 1 for classification. For a class of interest C_i, the GradCAM (L_{GC})

can be defined as:

$$L_{\mathrm{GC}} = ReLU(\sum_k \alpha_k^{C_{\mathrm{i}}}) \tag{1}$$

Where,

$$\alpha_k^{C_{\mathrm{i}}} = P_G(\frac{\partial N^c}{\partial T_Y^k}) \tag{2}$$

Where, $P_G()$ denote the global pooling process

In our proposed model we improved Grad-CAM in order to be fast and most accurate and useful for medical image. We achieved that by makes the gradient info streaming to the last convolutional layer to signify the significance of every activation map. For CNN model $N = f(I)$, it takes an input I and will output the scalar N. The internal convolutional layer Y has been picking up in f and the related activation as T. Denote the k_{th}channel of T_Y by T_Y^k. For a known base-line input I_b, the contribution T_Y^k towards N is defined as

$$C_i(T_Y^k) = f(I \circ V_Y^k) - f(I_b) \tag{3}$$

Where, $V_Y^k$is a vector with the identical shape of I_b, and it given by following formula:

$$V_Y^k = S(UP(T_Y^k)) \tag{4}$$

Where, $Up()$ denotes the process that up-samples T_Y^k into the input size and $S()$ is a normalization function where it maps every part of the image (input matrix) to [0, 1] Product. The architecture of the proposed model structure is illustrated in Figure 1.

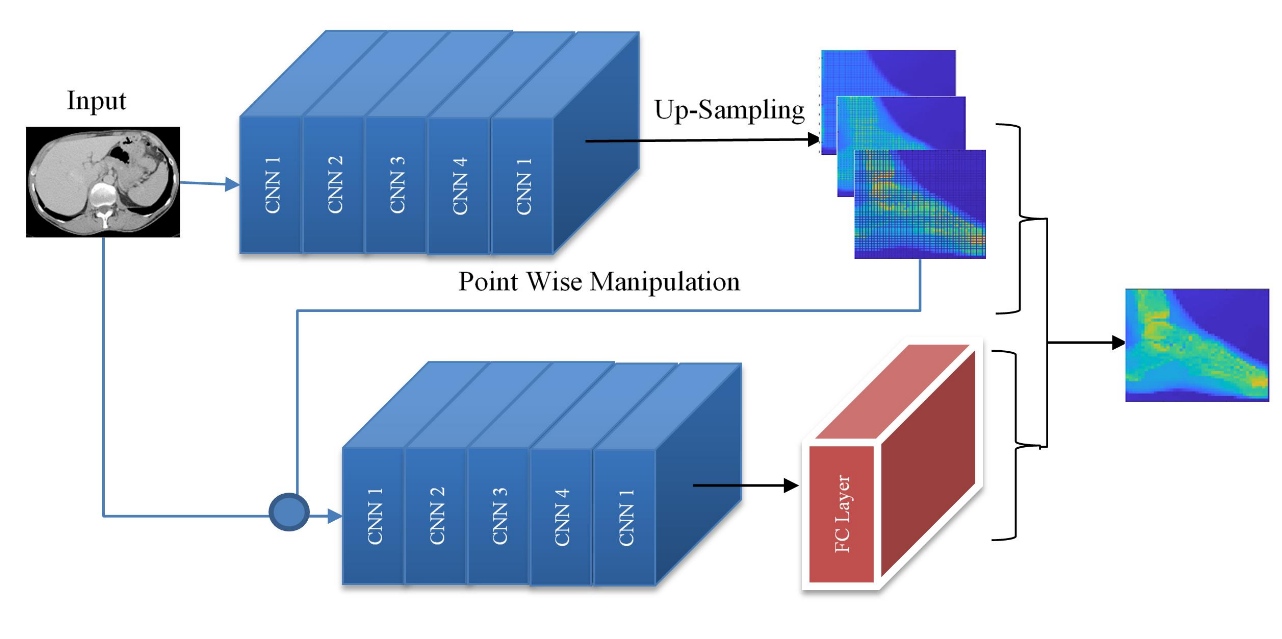

FIGURE 1. Proposed System Structure.

From figure 1, the activation maps are initially extracted in first stage (upper part). Every activation after that performs as a mask on main image and get it is forward-passing score from the target class. In stage 2, repeat steps for N times in which N represent the quantity of activation maps. At last, the result will be created by linear hybrid of activation maps and score-based weights. stage 1 and stage 2 share a same CNN model as feature extractor.

The algorithm of proposed model can be described in the following steps:

By applying a ReLU to the linear combining of maps it can just interested in the features which have a beneficial influence relating to the class of interest. Considering that the weights derive from the score related to the activation maps regarding target class, the proposed model obtains rid of the dependence on gradient. However, the last convolution layer is a more

Table 1. Proposed System Algorithm

Input:	Medical image (X-ray, CT scanner, MRI, etc.
Output:	High gradient Heat map distribution image
Step 1	Start
Step 2	Insert image I_0, base line image I_b
Step 3	Input the image row size (r), image cell size (c), size of square occlusion (s) and controls heatmap pixel size (delta), class c, layer Y
Step 4	Resize image to appropriate size (in this work we resized images to 227x227)
Step 5	Forward propagate, finds specific score
Step 6	Get activation of layer N
Step 7	Up-sample (T_Y^k), where k in $[0, \ldots, (C-1)]$
Step 8	Normalize the activation map, $S(Up(T_Y^k)$
Step 9	Hadamard product $(I \circ V_Y^k)$
Step 10	Find difference between recent score and regular score $ReLU(\sum_k \alpha_k^{C_\mathrm{i}})$
Step 11	Output image1
Step 12	End

suitable choice since it's the end point of the feature extraction, almost any kind of intermediate convolutional layer are able to be used in this model.

4. RESULTS AND DISCUSSION. In this section, we will introduce our experiment details and results. Any specific model of learned deep CNN can be preferred for visualization. For this work, we utilized ResNet with ImageNet weights, along with the last layers relearning on their data. In order to investigate the capability of this method to work with different cases.

In case 1, we processed the x-ray image of feet, the results is shown in figure 2.

From results it can see that the program start processing original image of feet (figure 2 (a)) then resize it to defined sized (in this work we used (227x227) and defined the initial parameters of row size (r), image cell size (c), size of square occlusion (s), the system start scan each pixel of image and start pre-processing image where the results from figure 2 (b) till reach to figure 2 (h). In each cycle the system re-modified and auto-tuned r, c and s value and an improvement will gain in each cycle were the final results show high level of gradient distribution for bone and tissue.

In case 2, we processed the x-ray image of the lungs, the results is shown in figure 3.

From results in figure 3, it can see that the program start processing original image of lungs (fig- ure 3 (a)) and same as case 1, the system resizes it to desired size (227x227) and defined the initial parameters of row size (r), image cell size (c), size of square occlusion (s), the system start scan each pixel of image and start preprocessing image where the results from figure 3 (b) till reach to figure 3 (h). In each cycle the system re-modified and auto-tuned r, c and s value an improvement will gain in each cycle were the final results show high level of gradient distribution for bone and clear identification of lungs which can be extracted efficiently .

In case 3, we made system processing CT image of abdomen, the results is shown in figure 4.

In case 3, we made system processing CT image of abdomen (figure 4). From result, the program start processing original image of CT (figure 4 (a)) then resize it to defined sized (in this work we used (227x227) and defined the initial parameters of row size (r), image cell size (c), size of square occlusion (s), the system start scan each pixel of image and start pre-processing image where the results from figure 4 (b) till reach to figure 4 (g). In each cycle the system

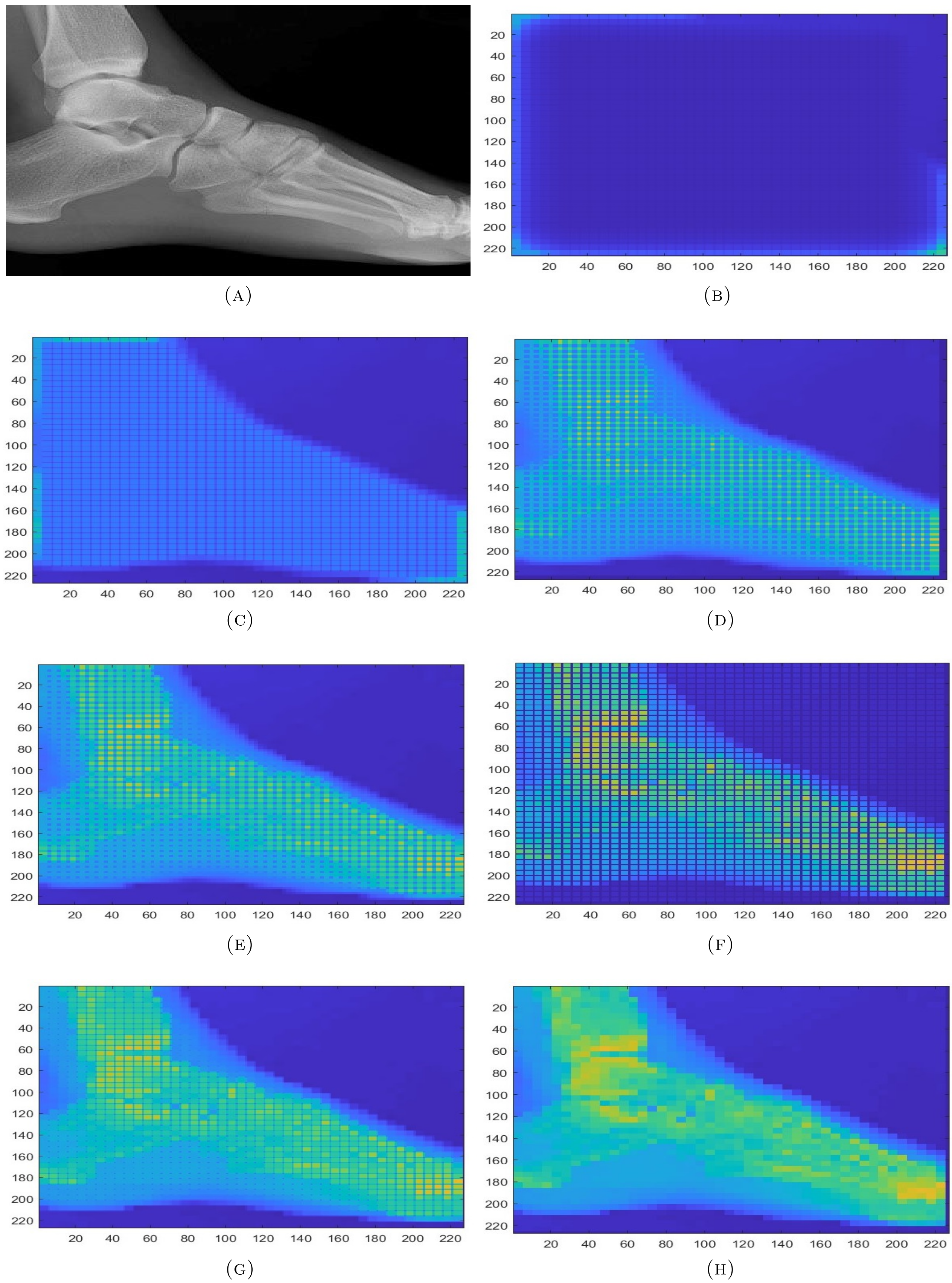

FIGURE 2. Processing Results on x-ray of foot. (a) original image, (b-h) the processing results

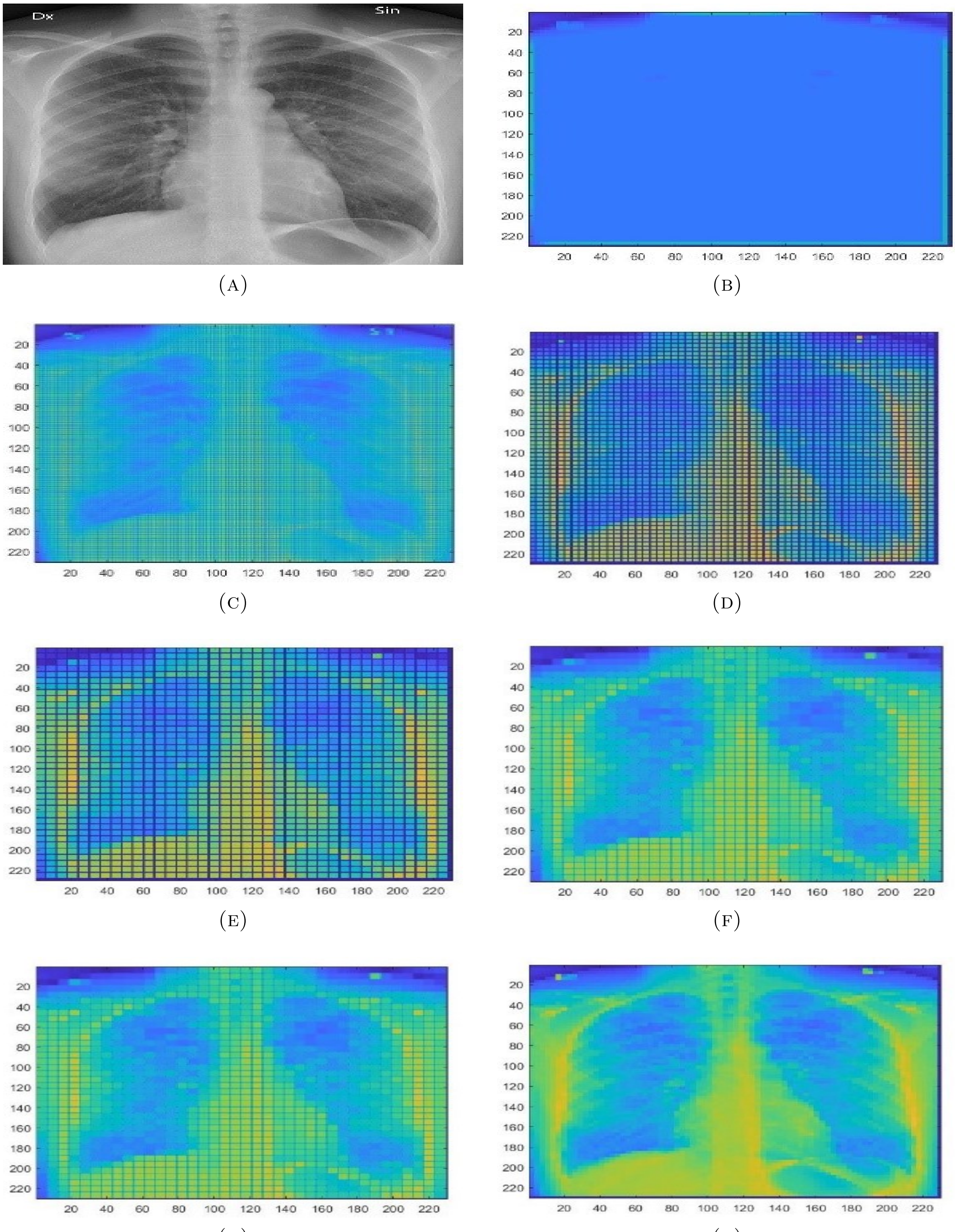

FIGURE 3. Processing Results on x-ray of lung. (a) original image, (b-h) the processing results

remodified and autotuned r, c and s value and an improvement will gain in each cycle were the final results show high level of gradient distribution for organic and bones in slice image where it can identify such as liver, vertebral disc and tissue.

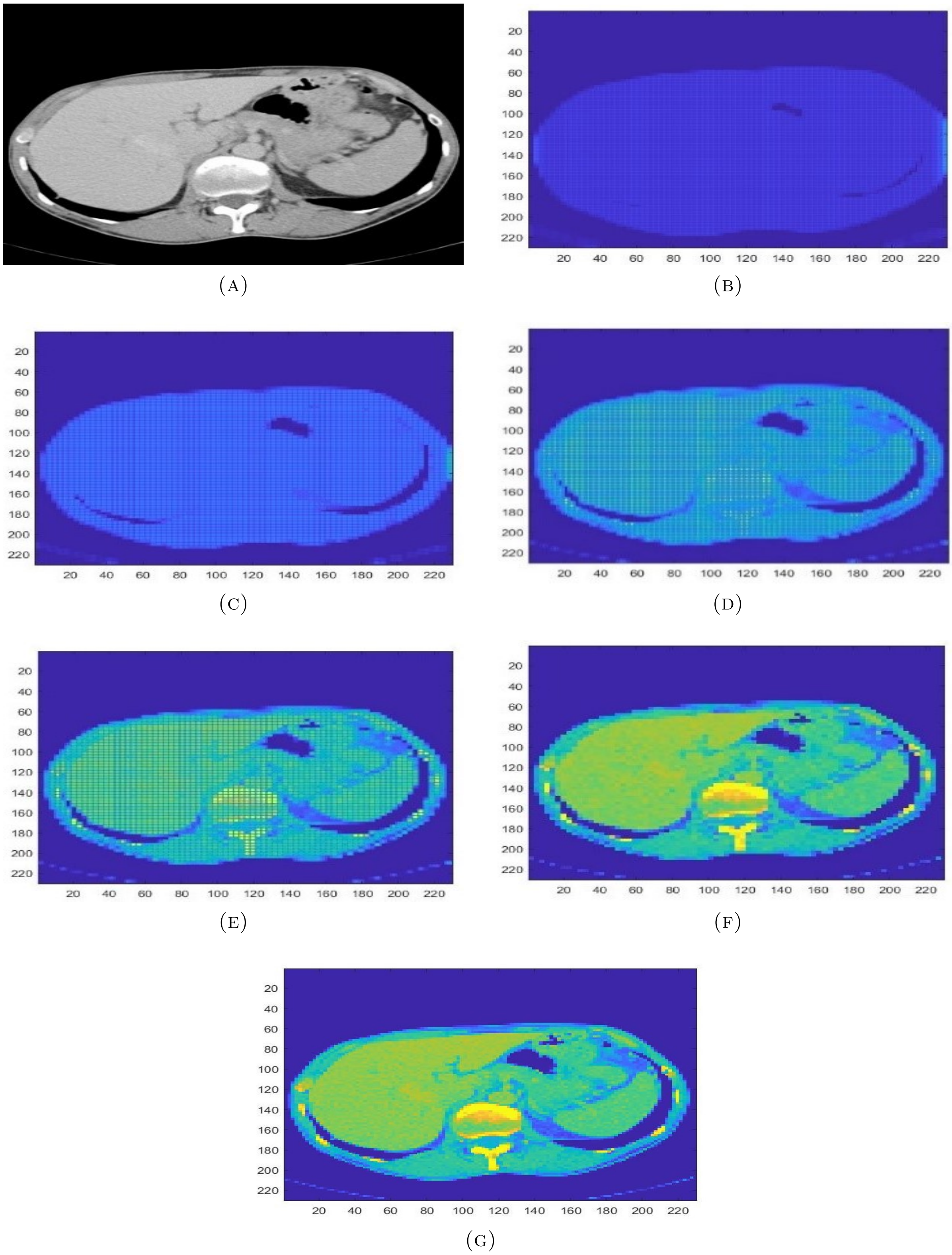

FIGURE 4. Processing Results on x-ray of lung. (a) original image, (b-h) the processing results

From overall results, it can be clear that the model successfully gradients all cases and the output image has sharp clear edges and each organ has specific colour and edges which can easily be extracted as features and identify the abnormality if it found. This can help classification program to be more accurate such as CADx system based deep learning technique .

5. Conclusion. Enhancing the medical images and recognizing saliency maps may help clear up the internal functions of CNN models. In this work, we proposed a model that improved the visual map of medical images for training of NN. The experiment results exposed an enhancement in the generated maps for various medical images. These maps had been generated by averaging gradients (in which the derivative of class score from several little perturbations of a presented image and making use of the resulting gradients to generate the GradCAM algorithm. The results show that the proposed model performs well in object localization along with distribution of heat maps for objects of the same class, which have the ability to generate maps for specific tissues, bone and other feature maps of the human body. in summary the proposed model involves an Increase in Self-confidence in the developing of weight for every activation map and eliminate the dependence on gradients as well as it has a more sensible weight representation.

References

[1] A. Hosny, C. Parmar, J. Quackenbush, L.H. Schwartz, H.J. and Aerts, Artificial intelligence in radiology, *Nature Reviews Cancer*, vol. 18, no.8, pp.500-510, 2018..

[2] Y.M.Y. Abdallah, and T. Alqahtani, Research in medical imaging using image processing techniques, *In Medical Imaging-Principles and Applications, IntechOpen*, 2019.

[3] S. Leger, S. Löck, V. Hietschold, R. Haase, H.J. Böhme and N. Abolmaali, Physical correction model for automatic correction of intensity non-uniformity in magnetic resonance imaging, *Physics and Imaging in Radiation Oncology*, vol.4, pp.32-38, 2017.

[4] L. Fan, F. Zhang, H. Fan, and C. Zhang, Brief review of image denoising techniques, *Visual Computing for Industry, Biomedicine, and Art*, vol.2, no.1, pp.1-12, 2019.

[5] S. Shanmuganathan, Artificial neural network modelling: An introduction, *In Artificial neural network modelling, Springer, Cham*, pp.1-14. 2016.

[6] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, Recent advances in convolutional neural networks, *Pattern Recognition*, vol. 77, pp.354-377, 2018.

[7] R. Yamashita, M. Nishio, R.K.G. Do and K.Togashi, Convolutional neural networks: an overview and application in radiology, *Insights into imaging*, vol.9, i4), pp.611-629, 2018.

[8] M. Xin, and Y. Wang, Research on image classification model based on deep convolution neural network, *EURASIP Journal on Image and Video Processing*, vol.2019, no.1, pp.1-11, 2019.

[9] A. Dosovitskiy and T. Brox, Inverting visual representations with convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4829-4837, 2016.

[10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, Learning deep features for discriminative localization, *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.2921-2929.

[11] D. S. Kermany, Identifying Medical Diagnoses and Treatable Diseases by Image-Based DeepLearning, Cell, vol.172, no.5, pp.1122-1131, 2018.

[12] J. Y. Zhao, J. Gong, S. T. Ma, Z. M. Lu, S. C. Chu, and J. F. Roddick, Finger Vein Recognition Scheme Based on Convolutional Neural Network Using Curvature Gray Image, J. Netw. Intell., Vol. 4, no. 3, pp.114-123, 2019.

[13] CM Kuo, H. Chaur-Heh, SCh Tseng, and JY Wu, Stacked Deep Convolution Network for Image Super-Resolution, Data Science and Pattern Recognition, Vol 5, no. 2, pp. 11–27, 2021

[14] R.R. Selvaraju, M. Cogswell, R. Das, , Vedantam, D. Parikh, and D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, *In Proceedings of the IEEE international conference on computer vision*, pp.618-626, 2017.