ISSN 2277-2685
IJESR/June. 2023/ Vol-13/Issue-2/1-11
Mrs.E.Ramola *et. al.,* / International Journal of Engineering & Science Research

Medical image improvement for lesion detection based on Class-aware attention

Mrs.E.Ramola ramolae@stellamaryscoe.edu.in
Mrs.V.Jino Shiny jinoshiny@stellamaryscoe.edu.in
Mr.G.Biju George bijugeorge@stellamaryscoe.edu.in
Mrs . R.Ayshwarya ayshwarya@stellamaryscoe.edu.in
Department Of Electronics and Communication Engineering
Stella Mary's College Of Engineering, Tamilnadu, India

Abstract -The lesion detection on medical images has been a challenging problem in computer vision due to the lack of high-quality annotations at the bounding box level and the difficulty to identify the lesion on gray-scale medical images which are commonly used in radiology tests like X-ray, CT and MRI. In this paper, we propose a novel framework of medical image enhancement based on the class-aware attention weight extraction and the deep colorization of medical images with generative adversarial networks motivated by human visual characteristics and cell staining. The evaluation conducted on the real public medical image datasets proves that, the performance of lesion detection based on the existing detectors is improved after enhancing the medical images with the proposed enhancement framework.

Keywords— Lesion detection, Medical image enhancement, Deep colorization, Class-aware attention weight extraction

I. INTRODUCTION

There has been significant progress in medical image analysis since the prevalence of deep learning technologies, such as lesion detection [1], disease classification [2] and lung segmentation [3]. Lesion detection plays an important role in Computer Aided Diagnosis (CADx). However, lesion detection is still confronted with two important challenges: (1) There are very few annotations of lesions on the bounding box level compared to the image-level annotations. (2) Most medical images such as ultrasonic, X-ray, CT (Computed Tomography) images are gray-scale which are not compatible with the advanced models like ResNet [4] pretrained on the color ImageNet dataset [5]. It is not straightforward to fine-tune the lesion detection model on top of these advanced pretrained models due to the difference in the color space.

*Corresponding Author

For the first challenge, some research [2] attempts to selfproduce bounding boxes based on image-level annotations through weakly supervised learning. However, these weakly supervised methods usually have high error rates. Thus, this is not the best solution for the first challenge.

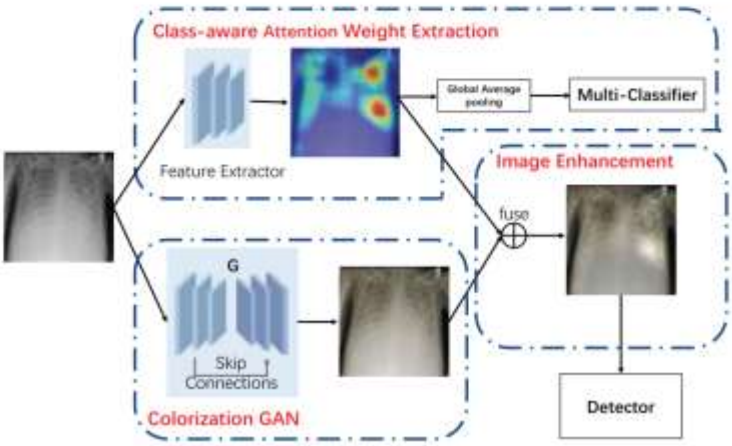

Fig. 1: The proposed framework for medical image enhancement, which consists of class-aware attention weight extraction, colorization GAN and image enhancement.

For the second challenge, some research [6] attempts to pretrain a model on ImageNet after converting color images to gray-scale images, which discards some useful information that the gray-scale cannot represent. On the opposite side, it is a straightforward thinking to colorize the gray-scale medical images. With the advances of Generative Adversarial Network (GAN) [7], some studies have sought to use GAN to colorize the gray-scale images [8], which inspires us to consider medical image colorization with deep neural networks.

To tackle the above issues, we propose a novel medical image enhancement framework to extract the lesion location information via class-aware attention and colorize medical images towards medical image enhancement in this paper. The enhanced images are further used to train the lesion detection model. Fig. 1 illustrates the proposed framework.

We summarize the major contributions of this paper as follows: (1) We propose a novel framework of medical image enhancement to benefit the lesion detection problem. The framework consists of class-aware attention weight extraction, colorization and image fusion. (2) We investigate a class-aware attention weight extraction method with weakly supervised learning that reveals the location of lesions to some degrees. (3) We propose to use the

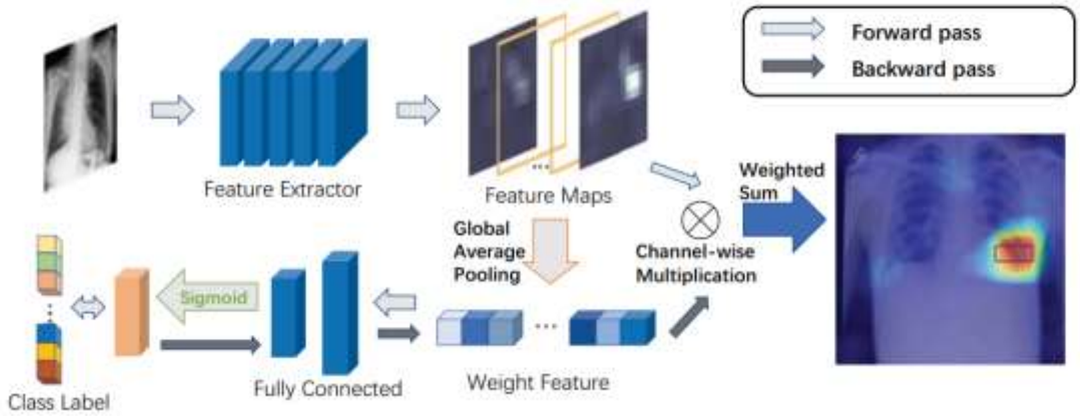

Fig. 2: The illustration of CAWE: for a given image, forward pass the image through the network to get the activated classes. The gradients of the activated classes are backward propagated to get the weight features, which are combined with the feature maps in the forward pass to generate the weight map of lesion locations. The blue rectangle is the ground-truth lesion location.

Generative Adversarial Network to colorize the medical images, whose results are supposed to be compatible with the advanced models pretrained on the color ImageNet. (4) The experimental evaluation conducted on publicly available dataset shows the lesion detection models trained on the enhanced images using the proposed framework outperform those trained with the existing enhancement methods or without any enhancement. This proves the effectiveness of the proposed framework for the medical image enhancement problem.

II. LITERATURE SURVEY:

The generative adversarial network (GAN) is an emerging deep learning technique for modeling high-dimensional data distributions and has been widely used in computer vision tasks.Image translation is one important application of GAN models . To overcome the need for perfectly aligned input-output pairs, CycleGAN uses a cycle consistency loss; an unsupervised approach is also taken by DualGAN and UNIT for image-to-image translation. Medical image synthesis is becoming an active research topic in medical imaging. However, most existing work focuses on synthesizing across imaging modalities rather than restoring the image in some way. For example, map from MRI to CT, while map from MRI to PET and map from multiple MRI modalities to other modalities.infer the manifold of normal tissue using a GAN architecture and develop an anomaly scoring scheme to predict abnormal tissue. In the GAN generates synthetic retinal images using segmentation labels, but pathological patterns are not considered.

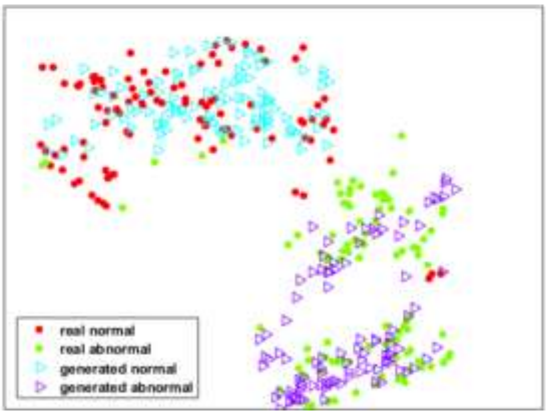

Fig. 3: 100 normal (red) and 100 abnormal (green) medical images from the BratS dataset embedded in R2 using t-SNE. We also show their embeddings after mapping by their respective generators. Red maps to purple (N2A) and green maps to blue (A2N).

III. PROPOSED METHOD:

In this section, we introduce our medical image enhancement framework, which consists of class-aware attention weight extraction, colorization GAN and image fusion for lesion detection.

3.1.Class-aware Attention Weight Extraction (CAWE)

Our intuition is that the semantic and spatial features about the lesion in the image can be extracted to enhance the original image. Thus, we propose the Class-aware Attention Weight Extraction (CAWE) method which extracts the high-level semantic and spatial information via a deep convolutional neural network. CAWE is inspired by CAM [9] that replaces the last fully connected layer with a global average pooling layer to preserve the class-specific semantic information.

Thus, we propose CAWE that takes both the semantic and spatial information into account as shown in Fig. 3. The global average pooling is used after the last convolutional layer to obtain the semantic (lesion class) information [9]. Meanwhile, the finally generated weight map by CAWE also relies on the feature maps from the convolutional layers, which preserves the lesion location information [10].

Formally, let F ∈ R^H×H×K denote the K feature maps generated by the last convolutional layer in Fig. 3, and let F^k denote the k-th feature map of F. The class-aware attention weights are generated based on the weighted sum of F^k. We propose to use the backward gradient flow from the activated classes back to the last convolutional layer as the weights. Specifically, we compute the weight of the k-th feature map:

$$w_k = \overbrace{\frac{1}{H \times H}\sum_i \sum_j}^{mean\,of\,the\,features} \underbrace{ReLU\Big(\frac{\partial \sum_{\sigma_l \geq \alpha_l} \sigma_l}{\partial F^k_{ij}}\Big)}_{multi-class\,gradients}, \quad (1)$$

where H is the size (width and height) of the feature maps, Fk ij is the feature variable of cell (i, j) on feature map Fk , and σl≥αl σl is the sum of probability of all activated classes. We use the weight wk to indicate the importance of Fk. Then, the class-aware attention weights can be computed as the weighted sum of all feature maps:

$$L_{CAWE} = ReLU\Big(\sum_k w_k \times F^k\Big). \quad (2)$$

We propose two ReLU operations in Eq. (1) and Eq. (2), respectively. The reason is that ReLU can effectively filter out the noisy signals (the noisy cell values in the feature map) outside the region-of-interest and preserve the signals from the class-of-interest.

The output of CAWE, LCAW E, has the same size as the last convolutional feature map which needs to rescale to the original image size through bilinear interpolation to perform image enhancement later.

The goal of CAWE is to extract the semantic and spatial information of lesions through a pretrained CNN classification network, which is independent from the lesion detection network. Thus, we can either train the CNN network on the dataset of the same source with that used for lesion detetion (denoted as intra-dataset) or on the dataset of other sources (denoted by inter-dataset) and the evaluation on the inter-dataset is conducted later. It means that CAWE is not limited on the same dataset as that used for lesion detection. CAWE uses weakly supervised learning to reveal the location information of lesion information to solve the first challenge mentioned in Sec. 1.

3.2. Colorization GAN for Medical Images

Despite the traditional colorization methods, GAN has been very popular in the colorization of the naturally taken images (not artificially transformed images) [11]. Inspired by that, we propose to use GAN for the colorization of medical images. We bring forward the Colorization GAN model (include the generator and the discriminator) based on the selfattention GAN [12] to colorize medical images. We train our colorization GAN model based on the ImageNet dataset where we convert the color images into grayscale images to obtain the training pairs. In the colorization GAN, the generator fools the discriminator by generating color image from gray-scale image. Let LG and LD denote the losses of the generator and the discriminator, respectively. LG and LD are computed separately in the training. Generator loss LG consists of two parts: the content loss LC and the adversarial loss LA:

$$L_G = L_C + \frac{1}{2} L_A, \quad (3)$$

Let x, y and G(x) denote the input gray-scale image, the ground-truth color image and the output colorized image from the generator. Then, for a given image, LC is composed of the reconstruction loss and the additional loss of feature map difference inspired by [13]:

$$L_C = |y - G(x)| + \frac{1}{w_i \times h_i} \sum_i |\phi_i(y) - \phi_i(G(x))|, \quad (4)$$

where, |·| is the L1 loss. x, y and G(x) are the input gray-scale image, the real ground-truth color image and the pseudo color image output by generator, respectively. Similar to [13], we use the additional loss, 1 wi×hi i |φi(y)−φi(G(x))|, to force the model to generate similar intermediate features when inputting either the real color image or the pseudo color image. φi is the i-th feature map. wi and hi are the width and the height of φi. Meanwhile, the adversarial loss LA of a given image is:

$$L_A = -log(D(G(x))), \quad (5)$$

where D(·) is the output (binary classification, i.e. real or pseudo color image) probability of the discriminator. Besides the generator, the discriminator is trained to distinguish the generated pseudo color image from the real color image. Thus, the loss of discriminator for a given image is defined as:

$$L_D = -log(D(y)) - log(1 - D(G(x))). \qquad (6)$$

3.3. Image Fusion for Lesion Detection

Based on CAWE and colorization GAN introduced in the previous sections, we propose to fuse the weight image and the colorized image to get the enhanced result:

$$I_{en} = \alpha I_{in} + (1 - \alpha) L_{CAWE} \odot I_{in}, \qquad (7)$$

where Iin and Ien denote the input image and the enhanced image, respectively. Iin can either be the original gray-scale image or the colorized image. is the element-wise product operator. LCAW E are the extracted class-aware attention weights. α (0 ≤ α ≤ 1) is a hyper-parameter to balance the importance of the enhanced image in the result and we empirically set it to 0.5 by default.7

he performance of the proposed medical image enhancement framework on public datasets.

3.4. Datasets and Metrics

For experimentation, we use two public datasets. (1) RSNA Pneumonia Detection. This dataset is introduced by the RSNA Pneumonia Detection Challenge [18] which consists of over 20,000 chest X-ray radiographs. The goal of this challenge is to detect the pneumonia in medical images. It is the main dataset in the evaluation and we split it at the patient level into the training and the testing sets at the ratio of 8:2. (2) CheXpert [19]. This dataset consists of 224,316 chest radiographs of 65,240 patients and is used as the inter-dataset to train the CAWE model alone. Each radiograph is annotated with image-level annotations in radiology reports. In our experiment, we evaluate the lesion detection performance by Free-Response ROC (FROC) [20] and Sensitivity at N False Positives (Sens@N). FROC is originally proposed to evaluate the diagnostic performance. Similar to ROC metric, FROC is computed by replacing the false positive rate with the average number of false positive predictions per image on X-axis and preserving the true positive rate on the Y-axis. The average number of false positive predictions is: the number of all false positive predictions divided by the number images. Sensitivity is commonly used in biomedicine, and it is computed as:

$$Sensitivity = \frac{TP}{TP + FN}, \qquad (8)$$

where T P and F N are the numbers of true positives and false negatives, respectively. We use Sens@N to denote the sensitivity at the N false positive predictions.

3.5. Experimental Settings

For CAWE, we use ResNet-50 as the backbone to extract image features. The network uses adam optimizer with initial parameters and the early stopping is used to avoid over-fitting. For colorization GAN, we use ResNet-34 pretrained on ImageNet as the feature extractor in the generator. To train the colorization GAN, the learning rate of the discriminator is five times larger than that of the generator based on Two Time-Scale Update Rules [21].

Methods	Sens@1	Sens@2	Sens@3	Sens@4
Faster R-CNN [14]	60.4%	69.5%	73.6%	76.6%
Faster R-CNN+AFDA [15]	65.5%	72.2%	74.9%	76.5%
Faster R-CNN+MedGA [16]	62.9%	71.4%	74.9%	77.5%
Faster R-CNN+Color (Ours)	63.6%	71.6%	76.3%	78.4%
Faster R-CNN+CAWE (Ours)	65.0%	71.8%	75.6%	78.1%
Faster R-CNN+Color+CAWE (Ours)	**67.7%**	**75.9%**	**78.7%**	**80.2%**
RetinaNet [17]	58.1%	69.9%	74.2%	77.3%
RetinaNet+AFDA [15]	62.5%	70.9%	75.9%	78.3%
RetinaNet+MedGA [16]	62.6%	73.2%	77.7%	80.8%
RetinaNet+Color (Ours)	62.0%	71.6%	76.4%	79.7%
RetinaNet+CAWE (Ours)	62.1%	71.6%	76.5%	80.2%
RetinaNet+Color+CAWE (Ours)	**64.9%**	**74.8%**	**79.7%**	**83.0%**

Table 1: The comparison of lesion detection with or without image enhancement, w.r.t. inter dataset. The detection models are trained on the RSNA Pneumonia Detection dataset. CAWE is trained on the CheXpert dataset.

Since the focus of this paper is on medical image enhancement instead of lesion detection, we experiment with the existing detection models, RetinaNet [17]1 and Faster RCNN [14], with ResNet-50 pretrained on ImageNet as the backbone. We train these detection models using the stochastic gradient descent (SGD) algorithm with learning rate of 0.02 on an NVIDIA Tesla P40 GPU, each of which contains 32 images per iteration. The input image size is 512×512 and the number of region-of-interest proposals is 128. In testing, the threshold of Intersection over Union (IoU) is set to 0.5 for the Non-Maximum Suppression (NMS).

3.6.Experimental Results

The performance of the proposed framework is evaluated on inter-dataset way (train CAWE and detector on different datasets), which could be seen as a semi-supervised learning. In our experiments, we train CAWE model on the CheXpert dataset and evaluate the performance of lesion detection on the RSNA Pneumonia Detection datasets. Besides, the proposed method was compared with several medical enhancement methods, including the Adaptive Fractional Differential Approach (AFDA) [15] and Medical Image Enhancement based on Genetic

Algorithms (MedGA) [16], both of them are set as default parameters. Table 1 shows the results on the RSNA Pneumonia Detection dataset.The maximum absolute improvements of the proposed methods compared to the baseline are 7.3%, 6.4%, 5.1% and 3.6% on Faster R-CNN, respectively, and those on RetinaNet are 6.8%, 4.9%, 5.5% and 5.7%, respectively. Meanwhile, the proposed method outperforms baseline and the competing medical enhancement approaches.

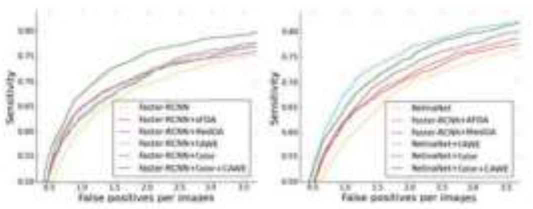

Fig. 4: The FROC curve of lesion detection on RSNA Pneumonia Detection dataset.
The X-axis represents the average number of false positive predictions per image.

Besides, the FROC curves on the RSNA dataset are illustrated in Fig. 3. We can also see that the proposed methods outperform the baseline and the other enhancement approaches. Specifically, the proposed methods with both deep colorization and class-aware attention weight extraction obtains the best performance among all the comparative methods. According to Table 1 and Fig. 3, the experimental results prove that the proposed framework is capable of transferring knowledge from other data sources to enhance the images in the target dataset, and the performance of lesion detection is also improved.

IV.CONCLUSION

In this paper, we propose a novel framework of medical image enhancement for lesion detection. The framework consists of the class-aware attention weight extraction and the colorization GAN, which deal with the two main challenges in the lesion detection on medical images, respectively. The evaluations conducted on the real public datasets prove that the performance of lesion detection can be improved by enhancing the medical images with the proposed framework.

References:

[1] Ke Yan, Xiaosong Wang, Le Lu, and Ronald M Summers, “Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning,” Journal of Medical Imaging, vol. 5, no. 3, 2018.

[2] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in CVPR, 2017.

[3] Dakai Jin, Ziyue Xu, Youbao Tang, Adam P Harrison, and Daniel J Mollura, "Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation," in MICCAI, 2018.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in CVPR, 2016.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in CVPR, 2009.

[6] Yiting Xie and David Richmond, "Pre-training on grayscale imagenet improves medical image classification," in ECCV, 2018.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in NIPS, 2014.

[8] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich, "Learning representations for automatic colorization," in ECCV, 2016.

[9] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba, "Learning deep features for discriminative localization," in CVPR, 2016.

[10] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in ICCV, 2017.

[11] Yun Cao, Zhiming Zhou, Weinan Zhang, and Yong Yu, "Unsupervised diverse colorization via generative adversarial networks," in ECML PKDD, 2017, pp. 151– 166.

[12] Xiaolin Zhang, Yunchao Wei, Guoliang Kang, Yi Yang, and Thomas Huang, "Self-produced guidance for weakly-supervised object localization," in ECCV, 2018.

[13] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Ca- ´ ballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image superresolution using a generative adversarial network," in CVPR, 2017.

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in NIPS, 2015.

[15] Bo Li and Wei Xie, “Adaptive fractional differential approach and its application to medical image enhancement,” Computers & Electrical Engineering, vol. 45, pp. 324–335, 2015.

[16] Leonardo Rundo, Andrea Tangherloni, Marco S Nobile, Carmelo Militello, Daniela Besozzi, Giancarlo Mauri, and Paolo Cazzaniga, “Medga: a novel evolutionary method for image enhancement in medical imaging systems,” Expert Systems with Applications, vol. 119, pp. 387–399, 2019.

[17] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar, “Focal loss for dense object detection,” ´ in ICCV, 2017.

[18] G. Jagga Rao, Y. Chalapathi Rao, "Human Body Parts Extraction in Images Using JAG-Human Body Detection (JAG-HBD) Algorithm Through MATLAB" Alochana Chakra Journal, Volume IX, Issue V, May/2020.

[19] G. Jagga Rao, Y. Chalapathi Rao, Dr. Anupama Desh Pande "A Study of Future Wireless Communication: 6G Technology Era " volume 14, issue 11,2020.

[20] Ai Bandos, T Rockette, Hesong, and D Gur, “Area under the free-response roc curve (froc) and a related summary index,” Biometrics, vol. 65, no. 1, pp. 247–256, 2009.

[21] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” in NIPS, 2017.