UDC 620.9:.658.45

GRAD-CAM VISUALIZATION MODEL FOR LUNG DISEASE DIAGNOSIS

He Tao

gr. 267011

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Scientific supervisor: Vishniakou U.A. Dr. of Sci. (Tech.), Professor at the department of ICT

Annotation. When applying the diagnosis of lung diseases in the Internet of Things networks, the Grand-Cam model can be used to provide assistance to doctors by imaging lungs and diagnosing diseases. By creating heatmaps of attention, Cad-Cam can visualize areas in an image. Doctors can observe heat maps to make decisions about the disease and ensure that the model focuses on areas related to specific lung diseases, increasing the accuracy and reliability of diagnosis.

Keywords: Cad-Cam, solution model, heat map of attention, diagnosis of lung diseases.

Introduction. Grad-Cam is a technique for interpretable deep learning model decisions that can be applied in areas such as computer vision and natural language processing [1]. These techniques can help us understand the model's focus on the input and the basis for its decision-making, rather than just treating the deep learning model as a black box. In the application of diagnosing lung diseases in IoT networks, Grad-Cam can be used to provide interpretability of model decisions, helping doctors and researchers understand the model's focus on lung images and the basis for disease diagnosis. By generating attention heatmaps, Grad-Cam can visualize the model's areas of attention in an image, thereby revealing how much the model pays attention to different areas in the image. Specifically, when applying Grad-Cam to the diagnosis of lung diseases, you first need to train a deep learning model that can classify or locate diseases based on lung images. Grad-Cam technology is then used to generate a heat map that shows the model's areas of interest in the lung image. These areas of concern may provide doctors with important clues in diagnosing lung disease [2]. By observing the resulting heatmaps, physicians can better understand the model's decision-making process and verify that the model is focusing on areas associated with specific lung diseases. This can help doctors verify the reliability and accuracy of the model and provide additional support and evidence to make more accurate diagnoses.

Principle of Grad-Cam. Grad-Cam computes the gradient of the predicted class score relative to the activations in the last convolutional layer [3]. These gradients represent the importance of each activation map for predicting a specific class. The general process is shown in the figure. This method does not change the two-point structure of the model and does not require retraining the model. It only needs to obtain the gradient of the last stage and back-transmit it.

In order to obtain the category discriminant map Grad-Cam, it is recorded as the class positioning map $L^c_{Grad-CAM} \in R^{u \times v}$ with height h, width w, and category c. First, use y^c (logits before softmax) to calculate the gradient of class c, and define the activation value of the feature map as A^k[4]. The gradients of these backflows are globally average pooled over the width and height dimensions (indexed by i and j respectively) to obtain neuron importance weights a^c_k:

$$a^c_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \qquad (1)$$

Among them, Z represents the number of pixels of the feature map, and Aij represents the (pixel value of i, j position) of the k-th feature map.

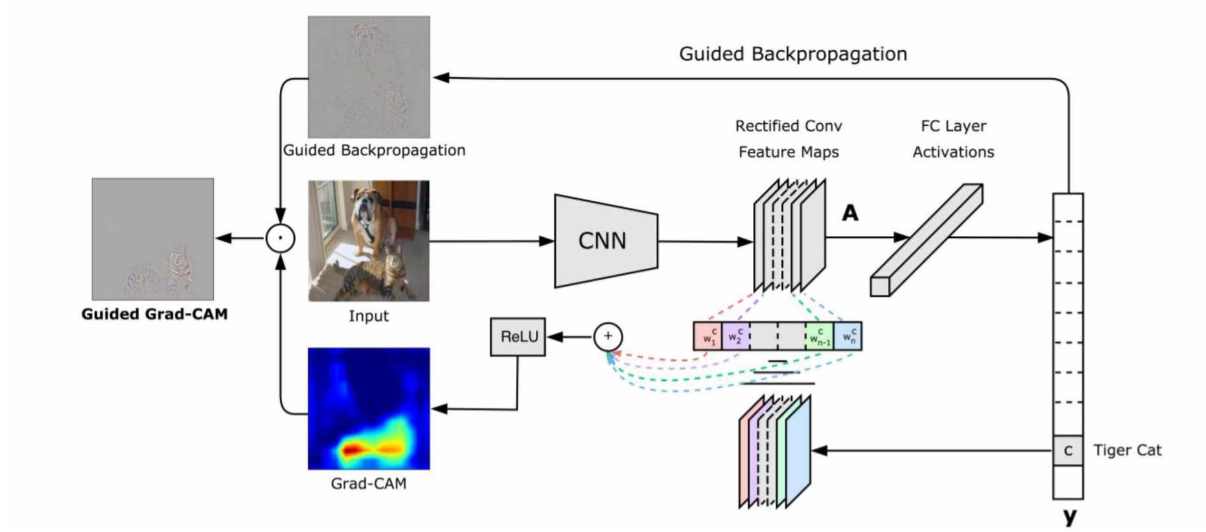

Figure 1 – Grad-Cam calculation structure diagram

While calculating a_k^c, the gradients are backpropagated with respect to the activations. The exact calculation equals the successive matrix products of the weight matrix and the gradient with respect to the activation function until the final convolutional layer to which the gradients are propagated. Therefore, this weight a_k^c represents a partial linearization of the deep network downstream of A and captures the "importance" of feature map k for target category c [5]. Calculate the product of the weight matrix and gradient of the activation function, and finally perform a weighted sum and output it after activation through ReLU:

$$L_{Grad-CAM}^{c} = ReLU(\sum_k a_k^c A^k) \quad (2)$$

It is worth noting that the rough heat map output after weighting here is consistent with the feature size of the selected convolution layer, and then the heat map is obtained through the ReLU operation. The purpose is to only consider pixels that have a positive impact on category c.

Breath sound feature optimization based on Grad-Cam. The model in this article uses the Mel spectrogram features of breath sounds as network input [6]. The extraction process includes: input audio data in wav format, resample the audio at 16kHz frequency, set the window function to Haining window, window length 25ms, step size 10ms, perform short-time Fourier transform on the audio to obtain the spectrogram, use 64-order The Mel filter group calculates the Mel spectrum on the spectrogram obtained in the previous step, calculates log(mel-spectrum+0.01) to obtain a stable Mel spectrum, and frames the frame for 0.96s per second, in which the frames are not stacked, each frame contains 64 Mel bands, and each second contains a total of 96 frames. The model in this article uses the Mel spectrogram features of breath sounds as network input. Figure 2 shows an example of a breath sound Mel spectrum.

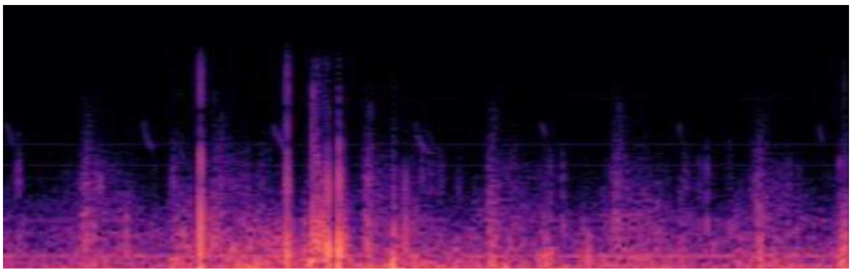

Figure 2 Breathing sound Mel spectrum sample

The features used in the basic model of this article were selected in the frequency range of 100-2000Hz. After preliminary experiments, when Grad-Cam was used to analyze the samples classified by the basic model, significant low-heat areas were found in the high-frequency areas of their spectra. As shown in Figure 3, the high-heat area in the figure represents the part that is more important to the classification model, and the low-heat area represents the part that is relatively unimportant for target prediction and classification.

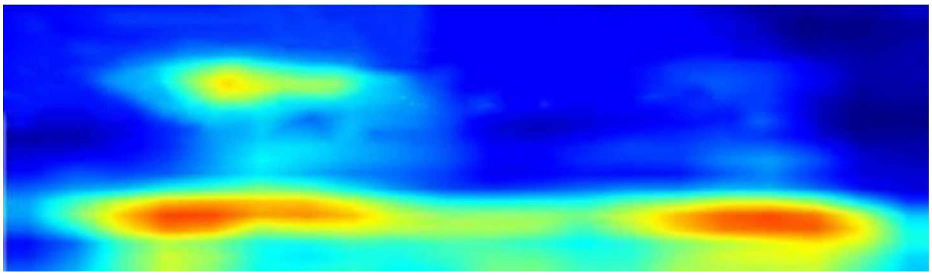

Figure 3 – Grad-Cam analyzes raw samples

The model's focus on the input image is mainly concentrated in the lower half of the input image, indicating that some areas within the audio high-frequency range cannot provide good features for the model's classification judgment [7]. Further analysis reveals that many samples have blank areas, especially in the frequency range 1500-2000Hz. This may have an adverse effect on the network performance of this article. In order to optimize the model effect, this article selectively cuts blank lines from the high-frequency areas of these spectra. The purpose of this is to ensure that the network focuses on the area of interest and reduces interference in irrelevant areas, thereby improving model performance [8]. This article chooses to cut out the area above the audio Mel frequency of 1500Hz. After pruning out the high-frequency regions, the network starts paying more attention to the lower half of the spectrogram, as shown in Figure 4.

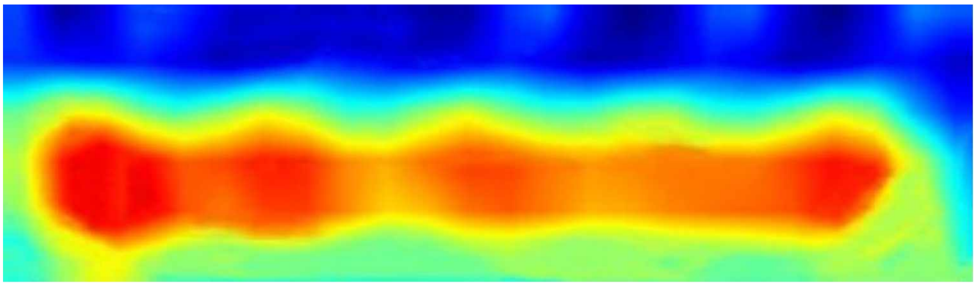

Figure 4 – Grad-Cam analyzes optimized samples

Conclusion. GRAD COM helps explain why the deep learning model focuses on breathing sounds and lung images and is the basis for disease diagnosis. The generated heat map of attention shows how much the model pays attention to different regions, thus providing detailed information about the disease. This visualization method helps to verify the reliability and accuracy of the model and increases the reliability of the diagnosis. By observing the heat map, the doctor can better make a decision and check whether the model focuses on areas related to the sound of breathing and lung diseases.

References

1. Zhou, Bolei, et al. Learning deep features for discriminative localization // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

2. Selvaraju, Ramprasaath R., et al. Grad-Cam: Visual explanations from deep networks via gradient-based localization // Proceedings of the IEEE International Conference on Computer Vision. 2017.

3. M. D. Zeiler, R. Fergu. Visualizing and understanding convolutional networks // European conference on computer vision. Springer, 2014, pp. 818–833.

4. Ba L J, Caruana R. Do Deep Nets Really Need to be Deep?. Advances in Neural Information Processing Systems, 2013:2654-2662.

5. Bastani O, Kim C, Bastani H. Interpreting Blackbox Models via Model Extraction. 2017.
6. Che Z, Purushotham S, Khemani R, et al. Distilling knowledge from deep networks with applications to healthcare domain[J]. arXiv preprint arXiv:1512.03542, 2015.
7. Xiao J, Ye H, He X, et al. Attentional Factorization Machines: Learning the Weight of Feature Interactions via Attention Networks. 2017:3119-3125.
8. Liu Q, Zeng Y, Mokhosi R, et al. STAMP: short-term attention/memory priority model for session-based recommendation // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 1831-1839.

УДК 620.9:.658.45

МОДЕЛЬ ВИЗУАЛИЗАЦИИ GRAD-CAM ДЛЯ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ЛЕГКИХ

He Tao

Гр.267011

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь

Научный руководитель: Вишняков В.А. – доктор технических наук, профессор кафедры ИКТ

Аннотация. При применении диагностики заболеваний легких в сетях Интернета вещей модель Grad-Cam может использоваться для обеспечения помощи врачам путем изображениях легких и диагностики заболеваний. Создавая тепловые карты внимания, Grad-Cam может визуализировать области на изображении. Врачи могут наблюдать за тепловыми картами, чтобы принять решений о заболевании и убедиться, что модель фокусируется на областях, связанных с конкретными заболеваниями легких, повышая точность и достоверность диагностики.

Ключевые слова: Grad-Cam, модель решения, тепловая карта внимания, диагностика заболеваний легких.