UDC 620.9:.658.45

GRAD-CAM VISUALIZATION MODEL FOR LUNG DISEASE DIAGNOSIS

HE TAO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 28, 2024

Abstract. 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 uses the gradient of any target concept (such as the logits of a certain class in the classification category, or even the output in the caption task), flows into the final convolutional layer, and generates a rough positioning map to highlight the features in the image. important areas for prediction [3]. By preceding the last global average pooling layer, the category activation map image is generated as a cumulative weighted activation, which is upscaled to the original image size.

The basic principle of the Grad-Cam method is to calculate the weight of each feature map in the last convolutional layer to the image category, then find the weighted sum of each feature map, and finally map the weighted sum feature map to the original image.

As shown in Figure 1, the input image first passes through multiple CNN convolutional layers, calculates global average pooling on the feature map of the last convolutional layer, and then flattens the pooling result into one dimension to make it a fully connected layer [4]. Then predict the classification result through the softmax activation function, and at the same time calculate the weight of all feature maps in the last convolution layer to the image category, then perform a weighted sum of these feature maps, and finally map the feature maps to the original in the form of a heat map get the results in the picture.

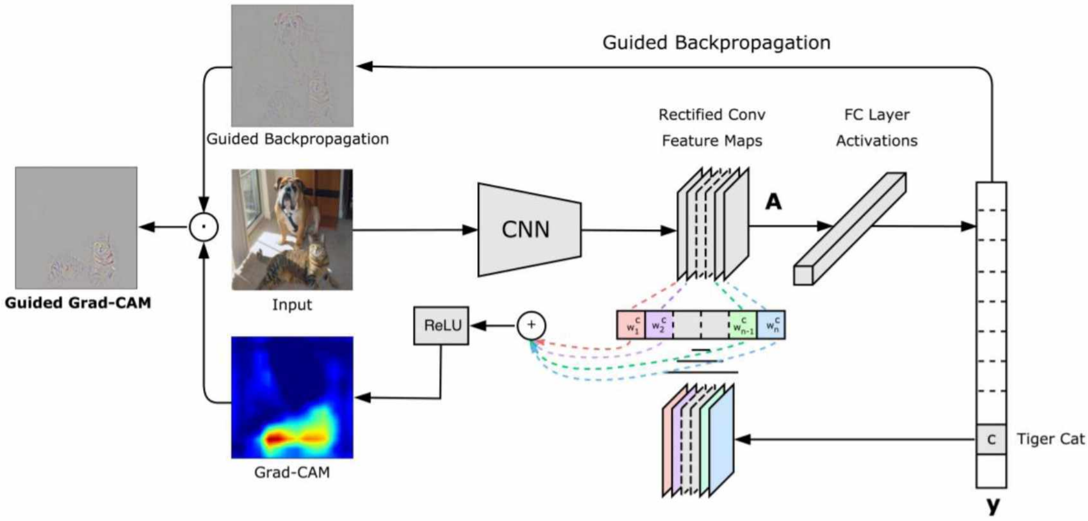

Figure 1. Grad-Cam calculation structure diagram

Given an image and a class of interest ("Tiger Cat" or any other kind of differentiable output) as input, the image is forward-propagated through the CNN part of the model, which then performs task-specific calculations to obtain a raw score for that category [5]. The gradients are set to zero for all classes except the desired class (Tiger Cat), which is set to 1. This signal is backpropagated to the rectified convolutional feature map of interest, which is combined to compute a coarse-grained CAM localization (blue heat map), which represents where the model must focus to make a specific decision. The heatmap is finally dot-multiplied with guided backpropagation to obtain a high-resolution and concept-specific guided gravity cam visualization.

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.

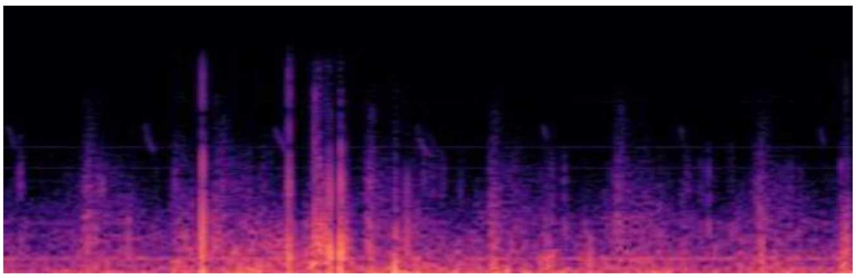

Figure 2. Breathing sound Mel spectrum sample

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.

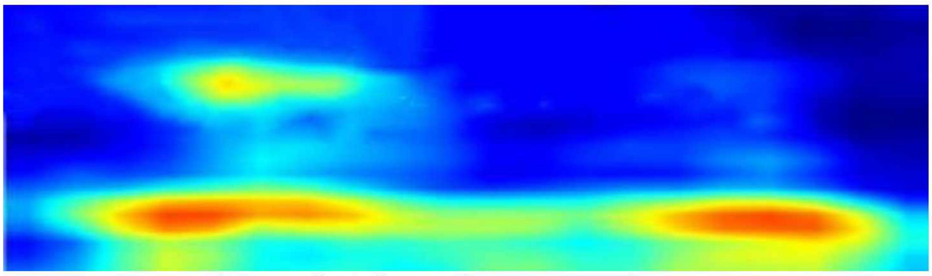

Figure 3. Grad-Cam analyzes raw samples

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.

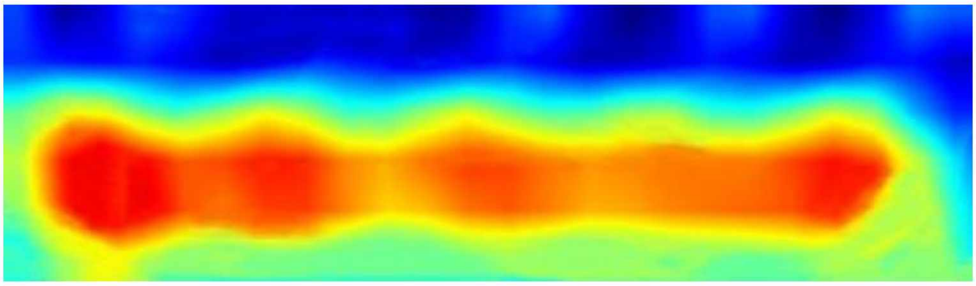

Figure 4. Grad-Cam analyzes optimized samples

Conclusion

Grad-Cam 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. Selvaraju R R., Cogswell M., Das A. // Grad-cam: Visual explanations from deep networks via gradient-based localization. 2017. Vol. 1. P. 21.
2. Zhang, Yunya. // Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. 2021. Vol. 2. P. 11.
3. Ramaswamy H G. // Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization. 2020. Vol. 1. P. 12.
4. Morbidelli P., Carrera D., Rossi B. // Augmented Grad-CAM: Heat-maps super resolution through augmentation. 2020. Vol. 1. P. 13.
5. Kim J Y., Kim J M. // Bearing fault diagnosis using grad-CAM and acoustic emission signals. 2020. Vol. 12
6. Xiao M., Zhang L., Shi W. // A visualization method based on the Grad-CAM for medical image segmentation model. 2021. Vol. 1. P. 8.
7. Kim J K., Jung S., Park J. // Arrhythmia detection model using modified DenseNet for comprehensible Grad-CAM visualization. 2022. Vol. 2. P. 4.
8. Jiang H., Xu J., Shi R. // A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification. 2020. Vol. 7