TEST Engineering & Management

March - April 2020
ISSN: 0193-4120 Page No. 4403 - 4408

Wide Class Activation Map for Generating Expanded Range of Activation Mapping

Dong In Kim[1], Sang Hun Lee[*2], Gang Seong Lee[3]

[1]Master Student, Department of Plasma Bio Display, KwangWoon University, Korea
[*2,3]Professor, Ingenium College of Liberal Arts, KwangWoon University, Korea
kimdongbee@kw.ac.kr[1], leesh58@kw.ac.kr[*2], gslee@kw.ac.kr[3]

Article Info
Volume 83
Page Number: 4403 - 4408
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 26 March 2020*

Abstract
In this paper, we proposed a wide CAM method for generating a wider range of activation maps than the CAM(Class Activation Map) method. The existing CAM method using the VGG-16 model extracts the activation map by applying GAP(Global Average Pooling) after the convolution 5_3 layer which extracts the highest dimension feature map. Because the CAM was applied to the lost feature maps rather than low-dimensional features such as edges. Not only does it determine based on objects in a much smaller area than the original image, it also only identifies objects that were relatively closed to each other in the image with many objects, or objects that are noticeable to humans. It cannot be determined. In order to improve this problem, the feature map of the convolution 4_3 layer, which was the previous step, was used to utilize the feature map of the lower level to utilize more feature of the heat map of the object. Experimental results show improved error rates in the classification top-1 and top-5 compared to the conventional methods.

Keywords: *Class Activation Map, Global Average Pooling, CNN, VGG-16, Deep learning*

1. Introduction

Recently, due to the development and popularization of deep learning, the field of computer vision has been researched and improved by using deep learning in the field of image processing such as face detection and recognition[1,2], object detection and recognition[3-7], and super resolution technique are attracting attention. The image processing method using deep learning takes much longer processing time than the existing hand-crafted method but has the advantage of showing much better results than the existing method. The CNN(Convolutional Neural Network) method of deep learning is composed of a plurality of layers for extracting features and a pooling layer for extracting important features by extracting representative values of the layers. The CAM method has attracted attention as a concern about which part of which layer is used for detecting when face detection, recognition, and object recognition using deep learning. The CAM method can be applied to various methods such as classification[8,9], object segmentation[10,11], using CAM as well as the area of interest of the basic object and is highly utilized by visually displaying the activation map of the layer. In this paper, we proposed wide CAM to obtain the activation map of wider objects in CAM and show better results than the existing CAM method.

2. Related work

2.1 Class Activation Map

Class Activation Map is a study that started with the idea of visually interpreting the CNN(Convolutional Neural Network). It is a

March - April 2020
ISSN: 0193-4120 Page No. 4403 - 4408

method to analyze which part influenced the classification of an image when the image is classified. In the conventional CNN method, low-level features can be extracted from a lower layer, and higher-level features can be obtained gradually toward a deeper layer. When this is expressed visually, it is not known which part influenced the object class, and the CAM method was proposed to confirm this. CAM used GAP(Global Average Pooling) on the first floor instead of the conventional three-layer classification method such as CNN's

FC(Fully Connected layer). When the CNN passes through many convolution layers, the weight value of the last determination layer is expressed as an activation map in object determination and the activation map is applied to the input image through filter operation. [Figure 1] shows the network of CAM and [Equation 1] shows the formula of CAM. M is CAM, x and y are coordinates, c is a discrimination class and k is each channel.

$$M_c(x,y) = \sum_k \omega_k^c f_k(x,y) \quad (1)$$

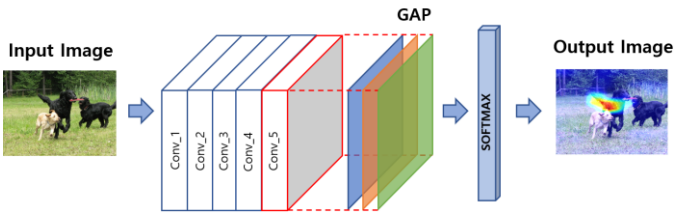


Figure 1. Class Activation Map

2.2 Global Average Pooling

The CNN method using the existing FC layer uses a flattening process to convert the result of the final convolution layer into three-dimensional convolution results into a one-dimensional matrix, and then classifies the class through three FC layers. That's the way. Since this process goes through many layers, there may be a problem of over-fitting that increases the number of parameters, consumes a lot of memory, and excessively increases the learning effect. To solve this problem, a GAP method that can preserve the location information of an object and has a relatively small amount of computation has been proposed. [Figure 2] is a visual representation of GAP calculation process, and [Equation 2] is a formula of GAP.

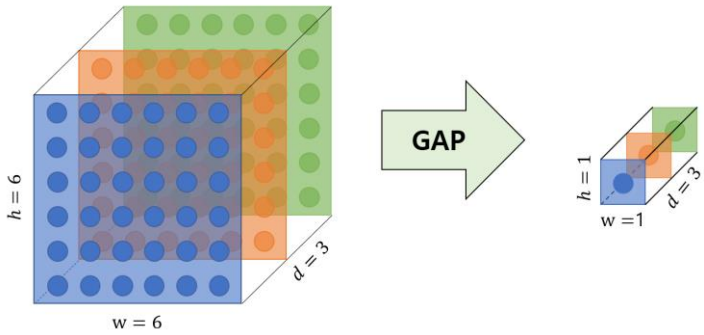


Figure 2. Global Average Pooling

3. Proposed Method

In this paper, we proposed a wide CAM method to obtain a wide activation map. In order to make up for the shortcomings of the conventional CAM method using the Conv5_3 layer, we proposed a CAM method that combined the Conv4_3 layer, which can determine objects at a relatively low level, by deconvolution. [Figure 3] shows the architecture of the wide CAM method using the proposed combination layer. In the proposed wide CAM architecture, the GAP is applied to Conv4_3 and Conv5_3 through the existing VGG-16 backbone network and then deconvolution is applied to combine the feature maps through OR operation. After that, the sorting process was performed using Softmax and the wide CAM image was finally output. The proposed method can show a wider and smoother activation map than the existing CAM method.

March - April 2020
ISSN: 0193-4120 Page No. 4403 - 4408

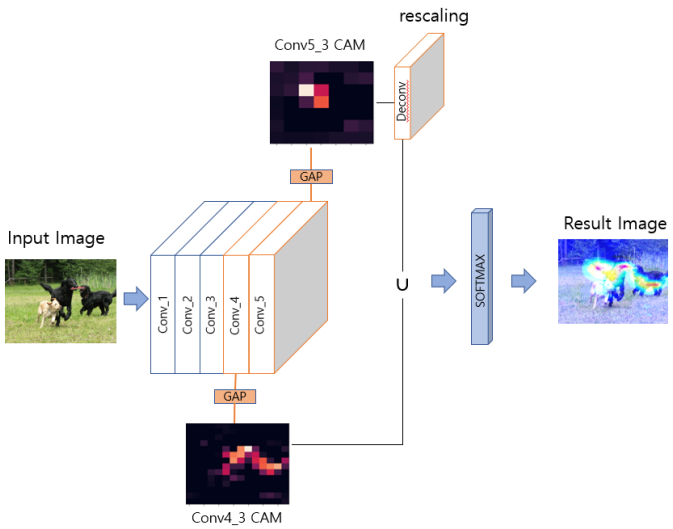


Figure 3. Proposed wide Class Activation Map Architecture

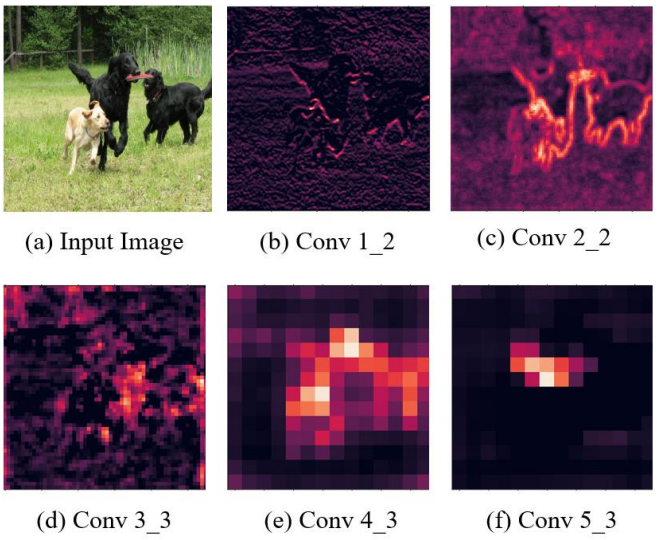


Figure 4. Feature Map Layer in Proposed Method

3.1 Modified the CAM Architecture

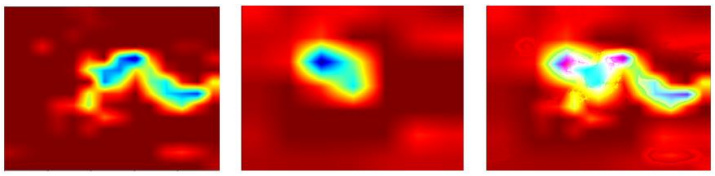
The existing Class Activation Map method is basically through the VGG-16 backbone network. This method is a method of mapping activation maps to input images by applying GAP to feature maps of Conv5_3 rare. Therefore, in the conventional CAM method, elements for determining objects of the convolution layer may appear narrow. This is a disadvantage that small objects or objects hidden behind objects are not distinguished because they use a lost feature map. Therefore, in this paper, we proposed wide CAM method to minimize the missing feature information of the existing CAM method. We used the Conv4_3 layer which has a relatively small dimension of features and contains a lot of feature information. By combining the layers of Conv4_3 and Conv5_3 and sharing the weighting parameters, we show the results of a wide range of CAM. [Figure 4] is a feature map that visually shows each convolution layer for the input image of the proposed method.

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \cup f'_k(x,y) \qquad (2)$$

3.2 Combine Feature Map

Conv1_2 and layers represent the lowest level features such as edges and show the overall image shape. As this feature map passes through the convolution layer, the overall shape is lost and an important response map for the object is obtained. Therefore, we proposed a method of combining the Conv 4_3and conv5_3 layers that can best represent the feature information of the object. First, since the feature map sizes of the Conv4_3 layer and the Conv5_3 layer are different, the Deconvolution process for the conv5_3 is performed. In the combining process, Conv 5_3 and Conv 4_3 were combined by bitwise OR operation. This results in a wider and more detailed heatmap result with the combination of feature map steps. [Figure 5] shows the heatmap combining process using heatmap to visually see the response map of the combined feature map. S_f means the size of the filter. [Equation 4] shows the process of combining feature maps of Conv4_3 and Conv5_3.

March - April 2020
ISSN: 0193-4120 Page No. 4403 - 4408



(a) Conv4_3 heatmap (b) Conv5_3 heatmap (c) Combine heatmap

Figure 5. Combine heatmap

$$S_0 = stride(S_i - 1) + S_f - 2 * pad \quad (3)$$

$$f'_k(x, y) = f_k(x, y) * r \quad (4)$$

3.3 Classification of Softmax

The Softmax process is a process for classifying and determining objects by GAP operation through the lowest convolution layer in the CAM network. The object is judged using the highest probability of the result among the classes to classify. Softmax is an activation function that has a value between 0 and 1 and the output values are always 1. Through this, the object with the highest value becomes the standard of class for discriminating the object in the image, and the proposed method showed improved Softmax classification result compared to the existing method. [Equation 5] is Softmax equation.

$$Softmax(M_c) = \frac{exp(M_c)}{\sum_i exp(M_c)} \quad (5)$$

4. Experimental and results

A variant of the VGG-16 backbone network was used for learning. For the training data, we used image data including people, animals, nature, and background from the ImageNet dataset. We trained 30 categories and 160 classes. The experimental image was composed of 5500 images from the ILSVRC 2015 dataset. Experimental results show that the proposed method had a wider response map and improved image classification than the conventional CAM method.

Figure 7. Experimental results of dog images

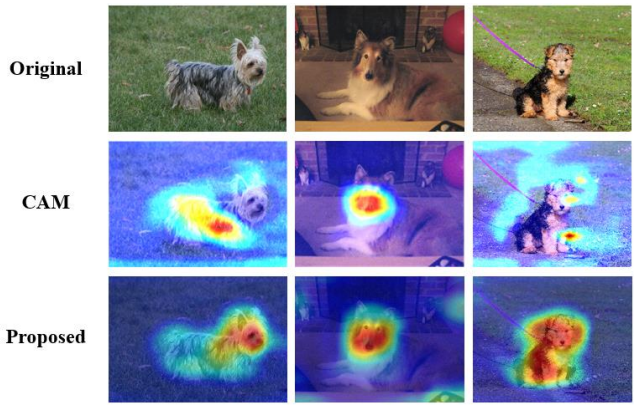


Figure 8. Experimental results of dog images

[Figure 7] and [Figure 8] are the experimental results of the dog image of ILSVRC 2015. In the existing CAM method, only a specific part of identifying a dog is shown on the activation map, but the proposed wide CAM is used to identify a wider range of objects and improved the results. When the object classification, segmentation, detection and recognition are performed using the proposed method, the result can be improved more.

March - April 2020
ISSN: 0193-4120 Page No. 4403 - 4408

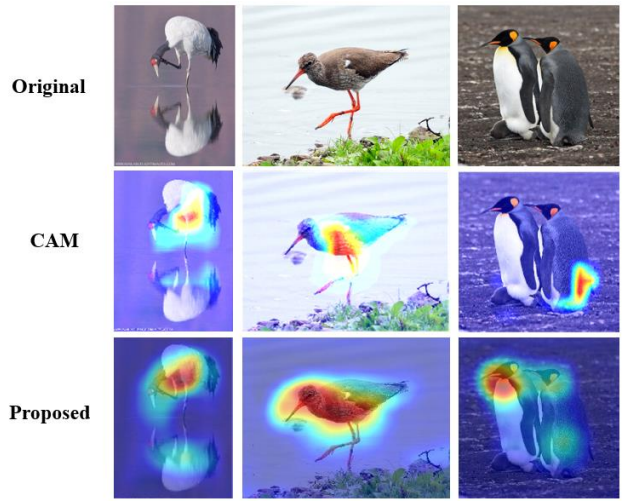


Figure 9. Experimental results of bird images

[Figure 9] shows the experimental result of bird image of ILSVRC2015. In the conventional CAM method, because the conv5_3 is used, the head of the bird becomes smaller in the feature map, which results in classifying only a part of the body of the bird. This can cause false detection when classifying, detecting, and recognizing objects, and especially affects classification. Through the proposed wide CAM, a wider range of objects can be identified and improved. When the object classification, segmentation, detection and recognition are performed using the proposed method, the result can be improved more.

Figure 10. Experimental results of vehicle images

[Figure 10] shows the experimental results of the vehicle image of ILSVRC2015. Since the conventional CAM method uses conv5_3, the wheel part of the vehicle becomes smaller in the feature map, which results in classifying only part of the vehicle. In addition, the most important wheel parts of a bicycle may not be classified, or a representative part of the classification of an object such as a headlight part of a vehicle may not be detected. This can cause false detection when classifying, detecting, and recognizing objects, and especially affects classification. Through the proposed wide CAM, a wider range of objects can be identified and improved. When the object classification, segmentation, detection and recognition are performed using the proposed method, the result can be improved more.

As an evaluation method, we compared the proposed method with the CAM and compared the average value by measuring the error rate of the top-1 and top-5 of the object rankings of the classifiers. The proposed method showed better results with lower error rates of top-1 and top-5 than the conventional CAM method. Figure 11 shows the error rates for the top-1 and top-5 classifiers.

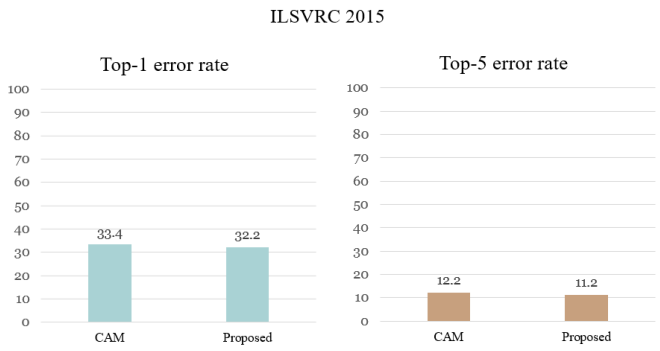


Figure 11. Classification error rate on ILSVRC 2015

5. Conclusion

In this paper, the wide CAM method was proposed to obtain a wider response map than the existing CAM method. Since the existing CAM method extracts features from the lowest layer of the convolution, it may be difficult to classify,

detect, and recognize objects by displaying only a part of the object as an activation map. This not only reduces the classification accuracy of the classifier but can also cause misclassification problems due to the narrow activation map. In order to solve this problem, we proposed a wide CAM that shows a wide range of activation maps by creating an activation map using a conv5_3 layer that combines conv5_3 layers with more detailed feature extraction. The proposed method improved the classification error rate by 1.2% for top-1 and 1% for top-5. Future research suggests CAM method suitable for various categories and CAM method that can detect and segment objects using CAM.

6. Acknowledgment

The present Research has been conducted by the Research Grant of Kwangwoon University in 2020.

7. References

1. Kim DI, Lee GS, Han GH, Lee SH. A Study on the Improvement of Skin Loss Area in Skin Color Extraction for Face Detection. Korea Convergence Society [Internet]. 2019 May 28;10(5):1–8. DOI:10.15207/JKCS.2019.10.5.001
2. Lee DW, Lee SH, Han HH, Chae GS. Improved Skin Color Extraction Based on Flood Fill for Face Detection. Korea Convergence Society [Internet]. 2019 Jun 28;10(6):7–14. DOI: 10.15207/JKCS.2019.10.6.007
3. Pyo S-K, Lee G, Park Y-S, Lee S-H. A license plate detection method based on contour extraction that adapts to environmental changes. Korea Convergence Society [Internet]. 2018 Sep 28;9(9):31–9. DOI: 10.15207/JKCS.2018.9.9.031
4. Kim H-J, Park Y-S, Kim K-B, Lee S-H. Modified HOG Feature Extraction for Pedestrian Tracking. Korea Convergence Society [Internet]. 2019 Mar 28;10(3):39–47. DOI:10.15207/JKCS.2019.10.3.039
5. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence [Internet]. 2017 Jun 1;39(6):1137–49.
6. Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.
7. Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2921-2929
8. Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra; The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626
9. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
10. Allili MS, Ziou D. Object of Interest segmentation and Tracking by Using Feature Selection and Active Contours. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition [Internet]. IEEE; 2007. Available from: http://dx.doi.org/10.1109/CVPR.2007.383449
11. Jiang X, Gao Y, Fang Z, Wang P, Huang B. An End-to-End Human Segmentation by Region Proposed Fully Convolutional Network. IEEE Access [Internet]. 2019;7:16395–405. Available from: http://dx.doi.org/10.1109/ACCESS.2019.2892973