Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 38(1), ss. 35-39, Mart 2023
Cukurova University Journal of the Faculty of Engineering, 38(1), pp. 35-39, March 2023

Performance Analysis of CNN Channel Attention Modules for Image Classification Task

Mehmet SARIGÜL[*1,2] **ORCID** *0000-0001-7323-6864*

[1]*Iskenderun Technical University, Computer Engineering Department, Iskenderun, Hatay*
[2]*Cukurova University, Computer Engineering Department, Adana, Türkiye*

Geliş tarihi: 05.01.2023 Kabul tarihi: 28.03.2023

Atıf şekli/ How to cite: SARIGÜL., M., (2023). Performance Analysis of CNN Channel Attention Modules for Image Classification Task. Cukurova University, Journal of the Faculty of Engineering, 38(1), 35-39.

Abstract

Increasing the representation power of convolutional neural networks, a popular deep learning model, is one of the hot study topics recently. Channel attention is a common strategy followed in this regard. In this strategy, the inter-channel relationship is exploited by a module placed after the convolution operation. Recently, successful channel attention modules are proposed in this context. In this article, a performance analysis of three popular channel attention structures which are Squeeze-and-Excitation Networks (SeNet), Efficient Channel Attention Networks (Eca-Net), and Convolutional Block Attention Module (CBAM), is performed using five different image datasets for the classification task. According to the obtained results, SeNet is the most successful channel attention module surpassing the other's performance in the majority of the experiments. In experiments with the ResNet18 and ResNet34 base models, the SeNet module showed the highest performance in three of the five datasets. For the ResNet50 baseline, SeNet was the channel attention module with the highest accuracy values for all datasets.

Keywords: Convolutional neural networks, Channel Attention, Image classification

Görüntü Sınıflandırma Görevi için CNN Kanal Dikkat Modüllerinin Performans Analizi

Özet

Popüler bir derin öğrenme modeli olan evrişimli sinir ağlarının temsil gücünün arttırılması, son zamanlarda sıcak çalışma konularından biridir. Kanal dikkati bu konuda izlenen yaygın bir stratejidir. Bu stratejide, konvolüsyon işleminden sonra yerleştirilen bir modül ile kanallar arası ilişkiden yararlanılır. Son zamanlarda, bu bağlamda başarılı kanal dikkat modülleri önerilmiştir. Bu makalede, üç popüler kanal dikkat yapısı olan Sıkıştır-ve-uyarım ağları (SeNet), Etkin Kanal Dikkat Ağları (Eca-Net) ve Evrişimsel Blok Dikkat Modülü (CBAM) üzerine beş farklı görüntü veriseti kullanılarak sınıflandırma görevi için performans analizi yapılmıştır. Elde edilen sonuçlara göre SeNet, deneylerin çoğunda diğerlerinin performansını geride bırakan en başarılı kanal dikkat modülü olmuştur. ResNet18 ve ResNet34 temel

[*]Corresponding author (Sorumlu yazar): Mehmet SARIGÜL, *msarigul@cu.edu.tr*

modelleriyle yapılan deneylerde, SeNet modülü beş veri kümesinden üçünde en yüksek performansı göstermiştir. ResNet50 temel modeli içinse SeNet, tüm veri kümeleri için en yüksek doğruluk değerlerine sahip kanal dikkat modülü olmuştur.

Anahtar kelimeler: Evrişimli sinir ağları, Kanal dikkati, Görüntü sınıflandırması

1. INTRODUCTION

Channel attention is a recent idea to improve the performance of a deep convolutional neural network. The convolution operation focuses on the spatial relationship of the input, while the channel attention focuses on and takes advantage of the inter-channel relationships. According to this idea, each channel entering the convolution layer has different importance and can be weighted accordingly. In order to achieve this, the features coming out of the CNN layer are first compressed along the channel, then the relationship between the channels is calculated and each channel is weighted.

Recently, there are many successful plug-and-play modules that have emerged with channel attention idea. Some of them are Squeeze-and-Excitation Networks (SeNet) [1], Efficient Channel Attention Networks (ECA-Net) [2], and Convolutional Block Attention Module (CBAM) [3]. These additional modules add a negligible amount of computational cost to the deep convolutional neural networks and provide a good performance increase. In this article, the performances of these three modules are compared on five different datasets using different versions of the ResNet [4] structures as base models. According to the test results, the most successful module is determined as SeNet. SeNet surpasses other modules' performance along the baseline model and achieves the best accuracy results in the majority of the experiments.

2. METHODS AND MATERIALS

In this section, brief information about the baseline model ResNet is given, three popular channel attention modules which are SeNet, ECA-Net and CBAM are explained, and, information about the datasets used in the experiments is demonstrated.

2.1. Residual Network (ResNet)

Residual Network (ResNet) was proposed by He et al [4] to enable the training of very deep network structures. It includes a large number of ResBlocks (a ResBlock is demonstrated in Figure 1) and achieved the best accuracy values in ILSVRC 2015 for the classification task of ImageNet dataset [5]. There are different ResNet structures named accordingly with their depths such as ResNet18, ResNet34, ResNet50, etc.

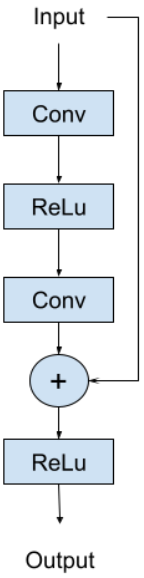


Figure 1. A standard residual block containing two convolutional layers

2.2. Squeeze-and-Excitation Networks (SeNet)

Squeeze-and-Excitation Networks (SeNet) proposed by Hu et al [1] is a recent channel attention technique enhancing convolutional operator's efficiency by considering channel-wise relationship. This technique proposes a module that adaptively recalibrates feature responses on a channel-by-channel basis. This module contains three parts which are squeeze block, excitation block and scale block. While squeeze block is used to reduce the spatial sizes of the feature maps, excitation block is utilized to learn the adaptive scaling weights for all input channels. The scaling module, on the other hand, allows each channel to be scaled with the weight calculated for it.

2.3. Efficient Channel Attention Networks (ECA-Net)

Efficient Channel Attention (ECA) block is designed to focus on the most important channels in a given image, and is similar in structure to a Squeeze-and-Excitation block. It has two parts: a compression module which collects global spatial information, and an excitation module which examines interactions between channels. Unlike a Squeeze-and-Excitation block, an ECA block only looks at the direct relationship between each channel and its closest k neighbours [6], which helps to keep the model simple.

2.4. Convolutional Block Attention Module (CBAM)

Convolutional Block Attention Module (CBAM) was proposed by Woo et al. CBAM considers inter-channel relationships of the input features. It squeezes the spatial dimension and, then collects spatial information of a feature map using both average pooling and maximum pooling. The squeezed features are then passed to a multilayer perceptron (MLP[7]) which contains one hidden layer. Finally, the output feature vectors are combined using element-by-element summation.

2.5. Datasets

In this paper, five different datasets are utilized in the experiments which are KTH Animals [8], Leeds Butterflies [9], Oxford Flowers [10], Ponce Birds [11] and Ponce Butterflies [12]. KTH Animal contains 1740 images of 19 different animals. Leeds butterflies dataset includes 832 images of 10 different butterfly species. Oxford Flowers consist of 1360 images belonging to 17 different flower types. Ponce Birds contains 600 images of 6 bird species and Ponce Butterflies includes 619 images of 7 different butterfly species. Some images from these datasets are given in Figure 2. All images in the datasets are resized to 224x224 which is the standard size for the ResNet model used as a baseline. Eighty percent of each dataset is split as training data and twenty percent is reserved as validation set.

Figure 2. Some samples from the datasets [8-12]

3. EXPERIMENTS, RESULTS AND DISCUSSION

In the experiments, different ResNet structures (ResNet18, ResNet34 and ResNet50) containing different numbers of ResNet blocks were utilized as the base models. Each channel attention module was embedded in these base network structures and the performance of each channel attention module was evaluated, separately. Each model was trained for 100 epochs for each dataset by using Adam Optimizer with a learning rate of 0.0001. Beta1 and beta2 parameters of Adam optimizer were set to 0.9 and 0.99. Twenty percent of the images in each dataset were reserved as test data and training was carried out with the remaining images. The performance of each model on the test data was evaluated and reported.

For the ResNet18 and ResNet34 base models, the SeNet structure performed the best accuracy values in three of the five datasets. For the ResNet50 base model, SeNet reached the highest performance value for all datasets. The results for the ResNet18 base model and channel attention modules are given in Table 1, the results for the ResNet34 base model and channel attention modules are shown in Table 2, and the results for the Resnet50 base model and channel attention modules are given in Table 3. The test loss and test accuracy values of the ResNet50 model and the Oxford Flowers dataset are given in Figure 3. Figure 3 shows that the Squeeze-and-Excitation module reaches the highest accuracy values and the lowest loss values during training.

In general, at least one of the channel attention modules increases the performance of the base models in the experiments. However, the highest and most consistent performance increase was achieved by the Squeeze-and-Excitation channel attention module.

Table 1. Performance values of different channel attention modules on ResNet18 baseline

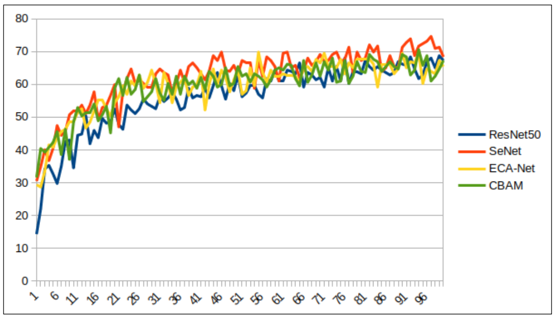
Dataset	Resnet	SeNet	EcaNet	CBAM
KTH Animals	73.851	**77.874**	75.575	69.253
Leeds butterflies	**90.419**	**90.419**	89.222	**90.419**
Oxford flowers	68.382	70.221	**70.588**	67.279
Ponce birds	83.333	**89.167**	**89.167**	87.5
Ponce butterflies	**95.161**	92.742	94.355	93.548

Table 2. Performance values of different channel attention modules on ResNet34 baseline

Dataset	Resnet	SeNet	EcaNet	CBAM
KTH Animals	72.126	**76.149**	**76.149**	74.425
Leeds butterflies	90.419	89.82	**91.018**	89.222
Oxford flowers	72.059	**74.632**	69.118	70.221
Ponce birds	**91.667**	90.833	90.833	**91.667**
Ponce butterflies	**95.968**	**95.968**	**95.968**	91.935

Table 3. Performance values of different channel attention modules on ResNet50 baseline

Dataset	Resnet	SeNet	EcaNet	CBAM
KTH Animals	72.701	**76.724**	76.437	74.138
Leeds butterflies	90.419	**92.216**	**92.216**	89.222
Oxford flowers	68.75	**74.632**	69.853	70.588
Ponce birds	87.5	**91.667**	90.833	**91.667**
Ponce butterflies	91.935	**93.548**	92.742	**93.548**



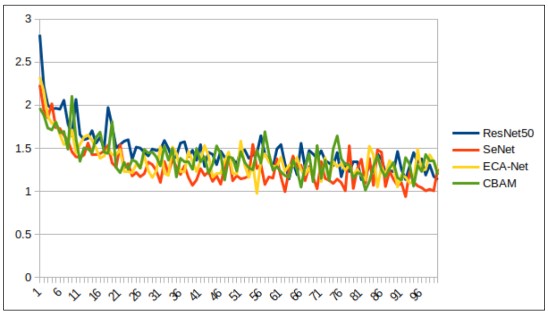


Figure 3. Accuracy values(top) and Cross-Entropy loss values(bottom) for different Channel attention modules on ResNet50 baseline for Oxford Flowers validation set

4. CONCLUSION

Channel attention is one of the strategies developed recently to improve the performance of convolutional neural networks. This strategy focuses on the inter-channel relationship within the feature maps that emerge from the convolutional layer. It calculates scale values for each channel and weights the channels according to their importance.

In this paper, classification accuracy values on five different datasets are reported and compared by adding different channel attention modules which are SeNet, ECA-Net and CBAM, on ResNet baseline models. It has been observed that all channel attention modules provide performance increases at different rates. As a result of the experiments, the most consistent performance increase was achieved by the SeNet module. It gave the best accuracy values in 73.3 percent of the experiments.

5. ACKNOWLEDGEMENT

This work has been supported by project number 120E447 from the TUBITAK (Scientific and Technological Research Council of Türkiye). The authors contributed equally to this work.

6. REFERENCES

1. Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7132-7141.
2. Wang, Q., Wu, B, Zhu, P., Li, P., Zuo, W., Hu Q., 2020. Supplementary Material for ‘ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA, 13-19.
3. Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, 3-19.
4. He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.
5. Krizhevsky, A., Sutskever, I., Hinton, G. E., 2017. Imagenet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 60(6), 84-90.
6. Peterson, L.E., 2009. K-Nearest Neighbor. Scholarpedia, 4(2), 1883.
7. Gardner, M. W., Dorling, S. R., 1998. Artificial Neural Networks (The Multilayer Perceptron)-A Review of Applications in the Atmospheric Sciences. Atmospheric Environment, 32(14-15), 2627-2636.
8. Afkham, H.M., Targhi, A.T., Eklundh, J.O., Pronobis, A., 2008. Joint Visual Vocabulary for Animal Classification. In 2008 19th International Conference on Pattern Recognition, 1-4.
9. Wang, J., Markert, K., Everingham M., 2009. Learning Models for Object Recognition from Natural Language Descriptions. In BMVC, 1, 2.
10. Nilsback M.E., Zisserman, A., 2006. A Visual Vocabulary for Flower Classification. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 1447-1454.
11. Lazebnik, S., Schmid, C., Ponce, J., 2005. A Maximum Entropy Framework for Part-based Texture and Object Recognition. In Tenth IEEE International Conference on Computer Vision. 1, 832-838.
12. Lazebnik, S., Schmid, C. Ponce, J., 2004. Semi-local Affine Parts for Object Recognition. In British Machine Vision Conference, 779-788.