International Journal of Combinatorial Optimization Problems and Informatics, 14(1), Jan-April 2023, 76-85. ISSN: 2007-1558.

www.editada.org

Using a novel XAI algorithm for Data Augmentation in image classification problems

Tonantzin Marcayda Guerrero Velázquez [1] and Juan Humberto Sossa Azuela[1]

[1] Instituto Politécnico Nacional – CIC. Laboratorio de Robótica y Mecatrónica
tmguerrerov@gmail.com, humbertosossa@gmail.com

Abstract. Nowadays, Machine learning solutions are increasing their presence in the industry, and the benefits associated with the use of this technology are reflected in a reduction of time, costs, and a clear economic benefit, also most of these solutions use supervised learning, where a dataset with labeled real examples is needed.

However, the many challenges associated with the implementation of a supervised machine learning solution are sometimes difficult to overpass, so one of the most important challenges is the creation of a big labeled dataset to feed the algorithms, since most of the time there is no any dataset available for the proposed solution, and most companies cannot afford to have a hand-created dataset with tens of thousands of records, also because they even cannot be sure that the model will work.

And the problem is that current machine learning techniques fail to improve the problem understanding on training with very small datasets, so here it is shown how by using a novel XAI method used to explain the decisions of a machine learning model in a computer vision problem, we can augment the labeled dataset focused on the important regions for the image classification, and increase the model performance, not just for training but for validation and testing, also this method turns out to be superior to the most used data augmentation methods if we further reduce the amount of information.

Keywords: Data Augmentation, XAI, CNN, Explainability, Machine Learning.

Article Info
Received Oct 01, 2022
Accepted Jan 20, 2023

1 Introduction

With Data Augmentation, basically, new samples are created through some modifications to the original dataset and in this way allow a better abstraction of the system behavior and reduce overfitting, it is commonly needed and used on computer vision problems to improve the performance of this pretty complicated task where often we have not enough data and also helps face some common problems inherent to some machine learning algorithms, like rotational invariance for the Convolutional Neural Networks.

Sometimes Machine learning models are able to make inferences on many problems with very high precision, but often this precision is due to the big amount of information available for the problem in turn. However, in some other cases, there are just a few samples available, or they are difficult to obtain; such is the case of the analysis of medical images [1] where furthermore we have a very complex problem, also is too complex to obtain the data.

Last years some new data augmentation techniques have been created trying to solve the problem of having a small dataset in a reliable way. Such is the case of the work described in [2] where they present a new technique based on randomly cutting out regions of four different images and putting them together to form a new image. This new image will be part of the training dataset, and its label will be proportionally composed of the classes to which the cropped images belong.

Also, in [3] a technique called Smart Augmentation is described, which is based on creating a new neural network that learns how to generate new data during the training process of the base model. Another interesting technique is the one described in [4], which uses a set of functions to combine different input images to create a new image.

More recent work is that described in [5], where they use a technique based on randomly selecting a rectangular region of the image, then its pixels with random values are deleted during the training. Also, in [6], it is proposed a new Explanation-driven Data Augmentation method to Improve Model and Explanation Alignment.

Although there is currently a constant search for new data augmentation methods, for problems with images, only some of these methods are commonly used and with them, it is more than enough to achieve the desired results, however, sometimes the problem turns out to be too complex, or the data is insufficient, therefore, in the present work, a new alternative to these methods is presented, which uses XAI to generate the data augmentation and thus improve the performance of the model even when the common methods fail.

2 Data Augmentation

Data Augmentation is useful to increase the number of training samples by small valid transformations to the original samples. These transformations can be split into two categories: geometric and photometric [7].

Geometric transformations are those that alter the image geometry of the image moving each of its pixels in some direction. On the other hand, the photometric or color transformations refer to those that map the value of the RGB channels by changing each pixel value (r, g, b) to a new value (r', g', b') according to some predefined heuristics [7].

One of the issues present with a small dataset is that the model training cannot generalize adequately the problem, which is known as overfitting. There are several methods to solve this problem, such as Dropout, Batch normalization, Transfer Learning, and Pretraining, and yes, overfitting can also be resolved using data augmentation [8]. However, unlike these techniques, data augmentation deals directly with the root of the problem: the lack of data in the training dataset, as is shown in [1].

Some of the most common techniques used when working with images are horizontal and vertical shift augmentation (HVA), horizontal and vertical flip augmentation (HVF), random rotation augmentation (RRA), random brightness augmentation (RBA), and random zoom augmentation (RZA).

3 Explainable Artificial Intelligence XAI

Today, machine learning models are widely used to solve complex problems and automate processes, both in industry and everyday life. These models reach surprising levels of precision and good performance, which have often exceeded the results obtained by a human being.

Although the use of machine learning generates such good results, there is an inherent concern about the risk associated with the use of algorithms whose results cannot be justified, and worse, still it is not possible to have absolute certainty that the result of a machine learning model will be correct.

Because the machine learning models are used in critical areas such as medicine, medical care, autonomous vehicle management, credit approval, legal and justice issues, etc., it is critical to have techniques that generate confidence in these models and their inferences. These tools are included in the so-called Explainable Artificial Intelligence (XAI).

Explainability techniques are currently used not only to justify the decisions made by a model but also to understand the importance and relationship that input data have with each other when making predictions, they are also used when deploying the model in a production environment as a monitoring tool for evaluation and retraining.

Although there are different types of explainability techniques, the present work focuses on working with images. In general, the mechanism of this images focused tools is by marking the attention points on the image, i.e, coloring the regions of the image that are most important to the model.

4 Proposed Method

This paper is based on a novel explainability method described in [9], that uses the model itself as an explainability tool. The authors of this paper show the efficiency of this method with an image classifier, whose predictions are explained by marking with colors the regions that the model considers to have the greatest influence to reach that result. This algorithm generally has three stages, in the first stage they use selective search to locate candidate regions as regions that may have some specific meaning in the image, which is shown in Figure 1(b), the second stage consists of reducing the number of regions using some heuristics and thus having the minimum number of possible candidate regions Figure 1(c), and finally using the model itself to find out the so-called useful regions through a classification process, then statistically encapsulate them into three categories, significant (green), relevant (yellow) and futile (red), which according to the previous order of appearance are more or less important for the prediction of the model Figure 1(d). The entire process can be seen in Figure 1 where it has been explained why the classifier asserts that the image is a cat.

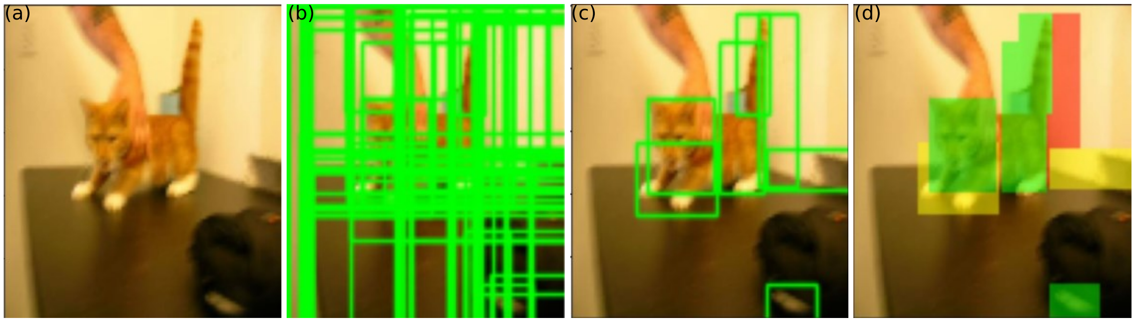

Figure 1. Generation of the visual explanation of the prediction of an image classifier model. (a) Original image. (b) Set of Candidate regions. (c) Set of useful regions (d) Visual explanation.

The intuitive reasoning behind the Data augmentation method proposed here is that if the colored green regions on the explanation denote the most important parts of the image for the classification purpose, then these regions can be used as an extension to the original dataset. So as an example, for the image in Figure 1(a) there are three green significant regions, the most important regions for the classifier to make its prediction, these three images after being resized can be used as new training samples for the model with the same label as the original image. These resized images are shown in Figure 2

Something important to mention here is that just the correctly classified images must be used in this data augmentation process in order to have very focused feedback on what is the right path when training. Then after doing this process of classifying the entire dataset, then choosing the correctly classified images, explaining their results, and expanding the original dataset with these explanation results the hope is that new images can be enough to improve even a little the performance of the model.

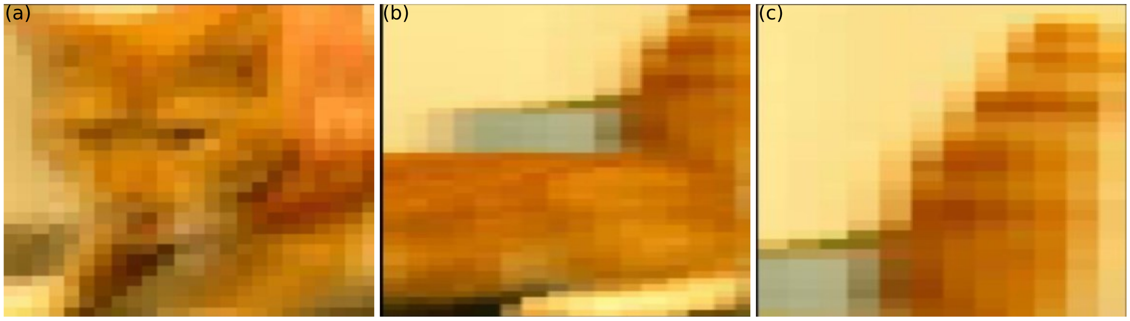

Figure 2. Example of augmented data generated by applying the explainability method for Figure 1(a).

On the images shown in Figure 2, it is observable that these images will have a lower resolution than the original because of the resizing and probably a distortion but it could even help to increase the variety of shapes for training the model. So, the data augmentation algorithm is:

1. Train the classifier.
2. Classify the entire training dataset.
3. Take the correctly classified images and discard the others.
4. Explain this new subset of images.
5. For each explanation:
   a. Cut individually the significant regions.
   b. Resize each significant region to the original image size.
   c. Add these new images to the training dataset.

5 Results and Discussion

The main intention here is to prove the efficacy of this new method when there is a very small dataset to work so, it is used some subsets of the classical Dogs vs. Cats dataset that was taken from Kaggle [10], which contains a total of 25,000 images for the training dataset and 12,500 images for the test dataset. The experiments also are divided into three sections, section 1 called Data Augmentation XAI Based Method, where it is proved the efficiency of this proposal, in the second part we compare this method with the common ones and finally in the third part called a very small dataset, it is shown the superiority of this method ahead of other common data augmentation transformations used in computer vision.

The common data transformations used as data augmentation methods to compare with are horizontal and vertical shift augmentation HVA, horizontal and vertical flip augmentation HVF, random rotation augmentation RRA, random brightness augmentation RBA and random zoom augmentation RZA. The metric used to compare these data augmentation methods will be the accuracy obtained by applying each of these methods to the training, validation, and test data sets. The number of augmented images that are added to the original dataset will depend on the size of the batch and the size of the dataset. For the case of the dataset with a size of 1000 images, an approximate of 4200 images with a batch size of 32 was used and using the data augmentation technique with XAI an approximate of 2168 images. For the case of the data set of 500 images an approximate of 1800 images with a batch size of 32 was used and using the data augmentation technique with XAI an approximate of 2761 images.

During this work it was important to avoid overfitting, to achieve this a common practice is to hold out part of the available data as a test set, however it could have some type of leak, to solve this we work on subpopulations and use cross validation where the training dataset is split into k smaller parts or folds and different models are trained with different folds for training and validation, so in this way, we have a major advantage in problems with few data.

Throughout this work, its used a CNN (Convolutional Neural Network) model with *Adam Optimization* and *categorical_crossentropy* as the loss function, also it was used always early stopping with 25 epochs. The architecture of the implemented model is shown in Figure 3.

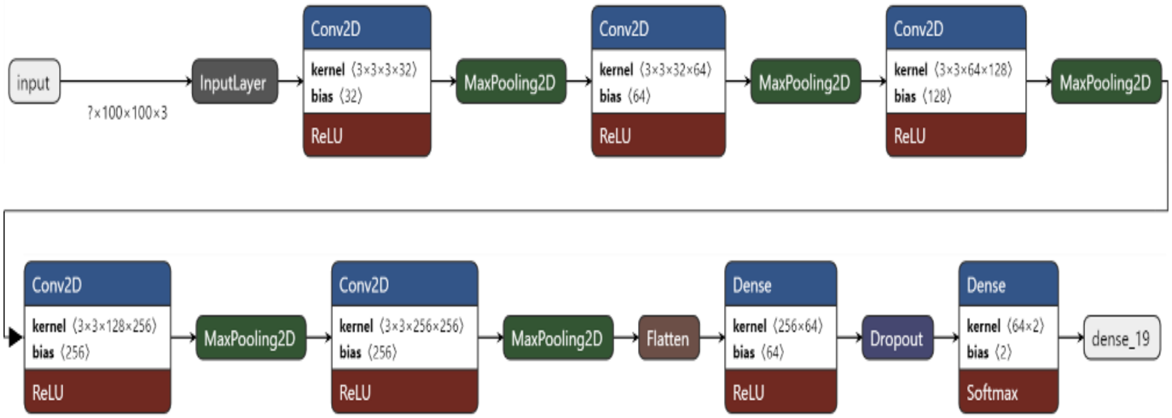

Figure 3. CNN used to prove the data augmentation proposed method.

5.1 Data Augmentation for a small dataset

Here a subset of 1000 (100x100 pixels) images of the Dogs vs. Cats dataset was used, in order to simulate a small dataset. Also, from the 1000 images of the new small dataset, 90% are used for training and 10% for validation. Another different set of 200 images will be used for testing.

The best model version could be obtained with the CNN architecture proposed before, it was obtained an accuracy of 0.9278 for training and 0.77 for validation, then using the data augmentation method with XAI as described before we get an accuracy of 0.98 for training and 0.78 for validation. A comparison between the loss and accuracy evolution in training with and without data augmentation using XAI is shown in Figure 4.

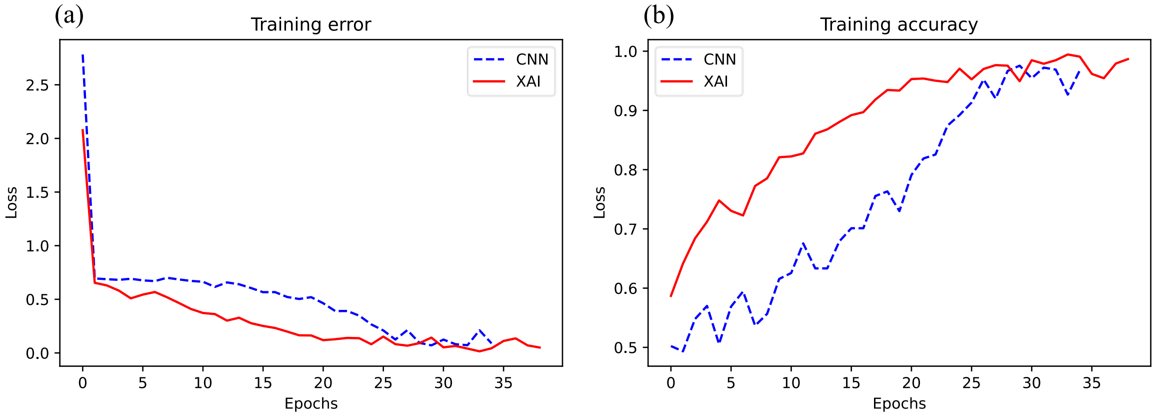

Figure 4. Comparison curves. (a) Training error with and without data augmentation XAI. (b) Training accuracy with and without data augmentation XAI.

When we compare Figures 4 (a) and (b), and according to the accuracy of the best model obtained, the data augmentation method using XAI proposed here slightly improves the model performance and its training. Furthermore, if we use the AUC (Area Under the Curve) of the ROC (Receiver Operating characteristic Curve) as shown in Figure 5, the model trained with data augmentation shows a better performance with 0.83 over 0.78

without data augmentation, and it definitively will be chosen in a real production environment where even 0.01 of improvement could make a real difference.

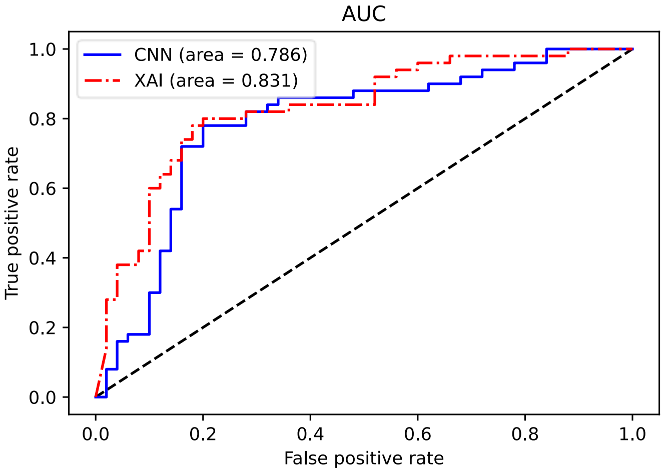

Figure 5. Curves ROC comparison

A better improvement is shown in Table 1 when it is compared the results with the test dataset which is the really important one, here the XAI method for data augmentation shows a clear increase or accuracy with 0.65 vs 0.60 without data augmentation.

Table 1. Results on the test dataset with and without data augmentation XAI.

Model	*Training ACC*	*Validation ACC*	*Testing ACC*	*AUC*
CNN	0.9278	0.7700	0.6000	0.786
XAI	0.9789	0.7800	0.6500	0.8371

5.2 XAI Method versus Known Methods

It is fair and necessary to compare the method against others commonly used. To do this, we use five different commonly used methods, horizontal and vertical shift augmentation HVA, horizontal and vertical flip augmentation HVF, random rotation augmentation RRA, random brightness augmentation RBA and random zoom augmentation RZA. In Figures 6 (a), a comparison of the training error is shown where we can observe that the XAI method error is always lower than the others, and in Figure 6 (b), we compare the training accuracy curve where the XAI method is also better. When we compare the AUC for the different methods, we found that the model created with the XAI method is one of the best methods, as shown in Figure 7.

Table 2 shows a comparison between the accuracies obtained from the different methods of data augmentation on the different datasets.

Table 2. Comparison between the accuracies of the different methods.

Model	*Training ACC*	*Validation ACC*	*Testing ACC*	*AUC*
XAI	0.9789	0.7800	0.6500	0.831
HVA	0.6467	0.6900	0.6600	0.707
HVF	0.6289	0.6600	0.6200	0.685
RRA	0.6589	0.7000	0.6250	0.714
BRA	0.9644	08000	06450	0.846
RZA	0.7256	0.7800	0.6700	0.798

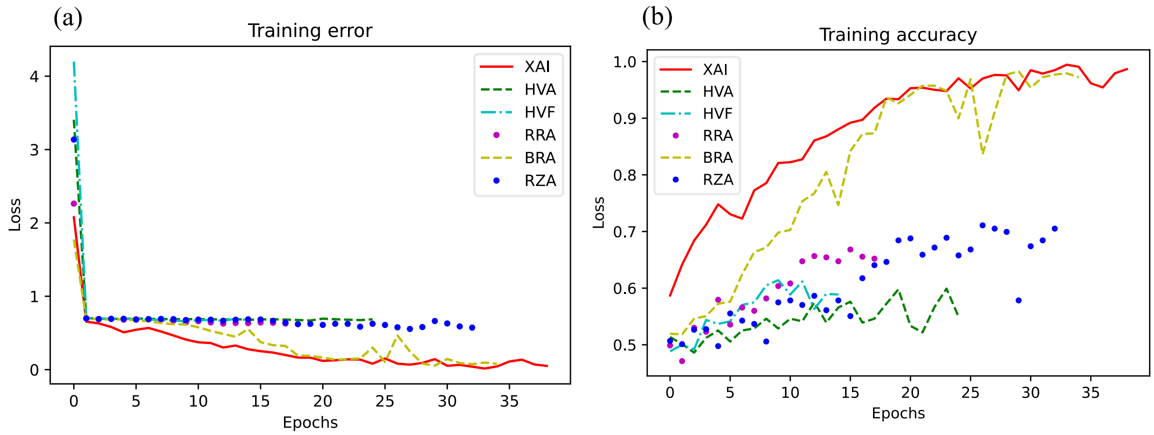

Figure 6. Comparison curves. (a) Training error data augmentation methods comparison. (b) Training Accuracy data augmentation methods comparison.

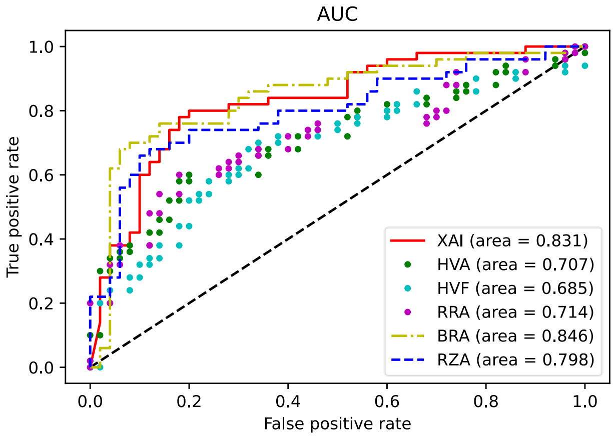

Figure 7. AUC for different Data augmentation methods for a small dataset comparison

5.3 Using a very small dataset

Here a subset of 500 images of the Dogs vs. Cats dataset was used, in order to simulate a very small dataset. Also, from the 500 images of the new small dataset, 90% are used for training and 10% for validation. Another different set of 200 images will be used for testing as in the first experiment.

Table 3. Results on the test dataset with and without data augmentation XAI.

Model	*Training ACC*	*Validation ACC*	*Testing ACC*	*AUC*
CNN	0.9925	0.6600	0.6500	0.6996
XAI	0.9950	0.7100	0.7400	0.7720

In Table 3 It can observe how the performance of the model obtained using XAI data augmentation is far superior to that without data augmentation, it is even more clear and also more important on testing with 0.74 over 0.65 without data augmentation which is 10% better.

It can also be compared the AUC that is shown in Figure 8 for the different approaches with and without data augmentation, and this time the AUC using data augmentation is almost 10% better and a clear winner.

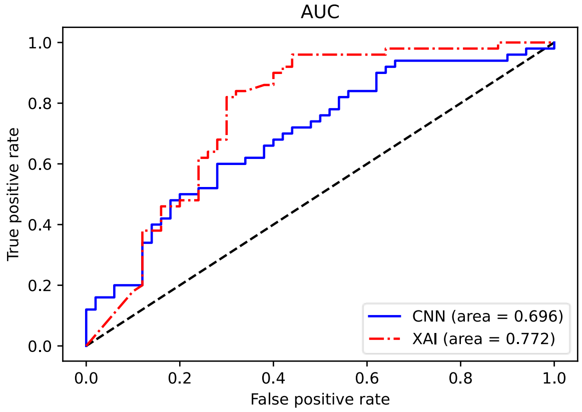

Figure 8. Curves AUC comparison for a very small dataset with and without XAI data augmentation

Now it would be interesting to explore what happens if we compare the XAI data augmentation method with the other well-known data augmentation methods as we can see in Table 4. Then it is observed that the XAI data augmentation method proposed here is clearly better than the others, both in training and validation, and testing.

Table 4. Comparison of the different data augmentation methods for a very small dataset.

Model	*Training ACC*	*Validation ACC*	*Testing ACC*	*AUC*
XAI	0.9950	0.7100	0.7400	0.772
HVA	0.6675	0.6600	0.6400	0.663
HVF	0.6775	0.6900	0.6800	0.716
RRA	0.7350	0.7000	0.7300	0.742
BRA	0.9800	0.7000	0.7200	0.761
RZA	0.5725	0.6000	0.6800	0.702

It is not surprising that the XAI data augmentation obtained better results for the AUC as shown in Figure 9

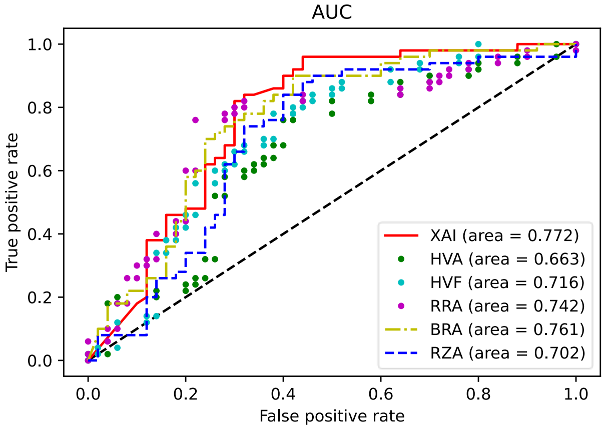

Figure 9. AUC comparison for different data augmentation methods using a very small data set

6 Conclusions and Directions for Further Research

In the present work, a new XAI-based data augmentation method for image classifiers is proposed, with this method a dataset can be extended, but also guide the model training in the right direction by creating the data augmentation using important regions on the images for the correct classification. It was also shown in the experiments that this method was effective in increasing the test accuracy of the original model from 0.65 to 0.74 and obtained the best results compared to the most commonly used data augmentation methods in image classification. In future work, it would be interesting to be able to apply this method to a practical problem and obtain results that have an impact in the real world. In addition, we can go further on this research, and since unlike the other methods mentioned here this one is not based on the images but on the explanation algorithm, then it can be easily extended to work with text just like the explainability method here used.

Acknowledgements

Tonantzin Guerrero thanks CONACYT for the scholarship to undertake her doctoral studies.The authors thank the Instituto Politécnico Nacional for the economic support under projects SIP 20200630 and 20210788, and CONACYT under projects 65 (Fronteras de la Ciencia) and 6005 (FORDECYT-PRONACES).

References

1. Shorten, C. and Khoshgoftaar T. M. (2019) “A survey on image data augmentation for deep learning”. Journal of Big Data. Vol.6, No. 1, pp. 1-48.
2. Takahashi R., Matsubara T. and Uehara K. (2019) “Data augmentation using random image cropping and patching for deep CNNs”. IEEE Transactions on Circuits and Systems for Video Technology. Vol. 30, No. 9, pp. 2917-2931.
3. Lemley J., Bazrafkan, S. and Corcoran, P. (2017) “Smart augmentation learning an optimal data augmentation strategy”. IEEE Access. Vol.5, pp. 5858-5869.
4. Summers, C., and Dinneen, M. J. (2019) “Improved mixed-example data augmentation”. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1262-1270.
5. Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020) “Random erasing data augmentation”. In Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34, No. 7, pp. 13001-13008.
6. Li, R., Zhang, Z., Li, J., Sanner, S., Jang, J., Jeong, Y., and Shim, D. (2021) “EDDA: Explanation-driven Data Augmentation to Improve Model and Explanation Alignment”. arXiv preprint arXiv:2105.14162.
7. Taylor, L., and Nitschke, G. (2018) “Improving deep learning with generic data augmentation”. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI). pp. 1542-1547.
8. Perez, L., and Wang, J. (2017) “The effectiveness of data augmentation in image classification using deep learning”. arXiv preprint arXiv:1712.04621.
9. Velázquez T.M.G. and Azuela J.H.S. (2021) “New explainability method based on the classification of useful regions in an image”. Computación y Sistemas. Vol. 25, No. 4.
10. Microsoft. “PetFinder.com: (Dogs vs. Cats dataset)”. https://www.kaggle.com/c/dogs-vs-cats/data.
11. Feng, S. Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T., and Hovy, E. (2021) “A survey of data augmentation approaches for NLP”. arXiv preprint arXiv:2105.03075.
12. Montserrat, D. M., Lin, Q., Allebach, J., and Delp, E. J. (2017) “Training object detection and recognition CNN models using data augmentation”. Electronic Imaging. Vol. 10, pp. 27-36.
13. Hernández-García, A., and König, P. (2018) “Further advantages of data augmentation on convolutional neural networks”. In International Conference on Artificial Neural Networks. pp. 95-103, Springer, Cham.
14. Kukačka, J., Golkov, V., and Cremers, D. (2017) “Regularization for deep learning: A taxonomy”. arXiv preprint arXiv:1710.10686.
15. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014) “Dropout: a simple way to prevent neural networks from overfitting”. The journal of machine learning research. Vol. 15, No. 1, pp. 1929-1958.
16. Ioffe, S., and Szegedy, C. (2015) “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In International Conference on Machine Learning. pp. 448-456, PMLR.
17. Molnar, C. (2020) “Interpretable Machine Learning. A guide for making black box models explainable”. https://christophm.github.io/interpretable-ml-book/.
18. Gunning, D. (2017) “Explainable Artificial Intelligence (XAI)”. DARPA.

19. Yu, J., Choi, J., & Lee, Y. (2022). “Mixing Approach for Text Data Augmentation Based on an Ensemble of Explainable Artificial Intelligence Methods”. Neural Processing Letters, 1-17.
20. Kim, D., & Lee, J. (2022). “Predictive evaluation of spectrogram-based vehicle sound quality via data augmentation and explainable artificial Intelligence: Image color adjustment with brightness and contrast”. Mechanical Systems and Signal Processing, Vol. 179, p. 109363.
21. Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). “Explainable artificial intelligence: objectives, stakeholders, and future research opportunities”. Information Systems Management, Vol. 39, No. 1, pp. 53-63.
22. Anaya-Isaza, A., & Mera-Jiménez, L. (2022). “Data Augmentation and Transfer Learning for Brain Tumor Detection in Magnetic Resonance Imaging”. IEEE Access, Vol. 10, pp. 23217-23233.
23. Maharana, K., Mondal, S., & Nemade, B. (2022). “A Review: Data Pre-Processing and Data Augmentation Techniques”. Global Transitions Proceedings.