Context Based Images Processing Using Machine Learning Approaches

B.S.Murthy [1], H.V.V.D.B.Sowjanya [2]

[1]**Assistant Professor MCA, DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**

[2]**PG Student of MCA,** Dantuluri Narayana Raju College **, Bhimavaram, Andharapradesh**

Abstract With the increased use of machine learning in decision-making scenarios, there has been a growing interest in explaining and understanding the outcomes of machine learning models. Despite this growing interest, existing works on interpretability and explanations have been mostly intended for expert users. Explanations for general users have been neglected in many usable and practical applications (e.g., image tagging, caption generation). It is important for non-technical users to understand features and how they affect an instance-specific prediction to satisfy the need for justification. Here, we propose a model-agnostic method for generating context-based explanations aiming for general users. We implement partial masking on segmented components to identify the contextual importance of each segment in scene classification tasks. We then generate explanations based on feature importance. We present visual and text-based explanations: saliency map presents the pertinent components with a descriptive textual justification, visual map with a colour bar graph showing the relative importance of each feature for a prediction. Evaluating the explanations using a user study (N = 50), we observed that our proposed explanation method visually outperformed existing gradient and occlusion-based methods. Hence, our proposed explanation method could be deployed to explain models' decisions to non-expert users in real-world applications

Index Terms— visual based, text-based, machine learning, gradient and occlusion-based methods

I Introduction

Deep Neural Networks (DNN) has improved the accuracy of prediction tasks in many areas from computer vision to natural language processing. However, the inability of DNNs to show their reasoning is limiting the wide adoption of these models in real-world applications. Given that, there has been an increasing interest in explaining and understanding classifiers retrospectively and examining what they have learned during training. Most of the existing approaches explain how a model determines its final output by where the network is 'looking'. These often produce intuitive visualizations that are intended for expert users to interpret network representations and evaluate the correctness of a model. For the non-experts who do not have sufficient knowledge of computer vision systems, a justification model that can provide intelligible explanations to non-experts would be more useful. Generally, the non-technical users would be interested in knowing instance-specific explanations instead of how the model makes its decision. Thus, the utility of explanations varies with stakeholders and the necessary stakeholders' requirements should be considered.

We propose a context-based justification method detailing how visual evidence is compatible with a classifier's output. Context-based justification is an approach that examines the relationship between an input and output by modifying an input image and observing its effect on the output. Image is segmented into several interpretable regions to measure this effect. In

this work, we demonstrate our approach using both manual annotation and automated segmentation with their semantic categories. Semantic categories provides the properties of the image scene needed to assess the effect of each component in scene classification tasks. Scene classification or scene recognition in this context involves assigning a label such as a playground, kitchen, beach to an input image based on the image's overall content. The overall idea is to map all the components to a semantic space based on their contextual importance.

Context provides critical information about particular scene; such as objects in an image, their arrangement, relative physical size to other objects, and location. Contextual information provides important cues for a model to learn during training and also use during prediction. While some features have more influence on the outcome than others, the influence of each individual component is also dependent on other components. We extract the degree of importance by estimating how each feature is contributing to a prediction. This is performed through a systematic masking and observation of the prediction score. We assess the effect of context in scene prediction and use it as justification. We note that sometimes the terms explanation and justification are used interchangeably in the domain of explainable AI.

The contributions on visual explanations research are summarized as follows:

- We propose an algorithm for generating context-based explanations that can be applied to DNNs and other black-box models that output probabilities.
- We present visual and text-based explanations that aim to improve intelligibility for nonexpert users.
- Using a user study, we present evaluation results of the different aspects of visual explanation.

2 Literature survey

Researchers have been focusing on integrating explanation facilities into computer vision tasks (e.g., image captioning, visual question answering, object localization) . Generally, these explanation models can be categorized as intrinsically interpretable models and the post-hoc interpretability. Both categories produce saliency maps which simplify the image representations, making it easy to analyze. Saliency maps help to highlight the most important features for a prediction.

Interpretable models are usually specific to a certain model which focuses on the internal functioning of the model. Interpretable models analyze the interaction between neurons and what each neuron has learned. Post-hoc interpretability methods explain instance-specific predictions on the basis of how each feature (or a set of features) influences the final outcome. In general, both approaches can have some limitations and strengths. The results of interpretable models are directly explainable without requiring another model to generate explanations. However, they are limited to a specific learning model. Whereas the post-hoc interpretability methods are generally model-agnostic but they may be limited in their approximate nature .

Some of the intrinsically interpretable models apply gradients and decomposition based methods to understand the internal structure of complex black-box models. Gradient methods highlight the unit changes and emphasize the important features or regions in an image. In this way, it is possible to learn the prototypes that have the highest probability to be predicted as a certain class of a trained DNN.

Considering this, Li et al. proposed an interpretable neural network architecture whose predictions are based on the similarity of an input to a small set of prototypes learned during training. As presented by Nguyen et al, the activation maximization method synthesizes an image that highly activates a neuron and reveals the features learned by each neuron in an interpretable way. Some works proposed visual analysis by clustering important neurons based on

the features and the interactions between them . Furthermore, decomposition methods such as layer-wise relevance methods are presented to analyze which pixels are contributing to what extent to a classification result . Some post-hoc explanations also proposed using gradients to create saliency maps. One way to visualize saliency maps is by going backward through the inverted network from an output of interest. It highlights the discriminative features of the image with respect to the given class. Another method uses class activation mapping with the gradients of a target input in the last convolutional layer to produce a rough localization map highlighting the important regions . This method is further developed for explaining occurrences of multiple object instances in a single image.

Moreover, other post-hoc explanation methods suggested modifying the input image and observing the effect on the prediction. Zeiler and Fergus proposed the occlusion sensitivity method to monitor the prediction score for a specific class by masking different parts of the input image with a grey square. Results are then visualized as saliency maps to show which parts of an image are most important for the classification. Another method that uses prediction score is LIME (Local Interpretable Model-Agnostic Explanation). This method generates explanations by approximating a black-box model by perturbing an input image.

They adopt SLIC (Simple Linear Iterative Clustering) superpixel method to partition image into smaller regions and use these parts to perturb the image. Then, they present the superpixels with the highest positive weights as an explanation . Although these methods seem intuitive, they are inefficient in terms of salient features representation. In our approach, we propose observing the prediction result using semantic occlusion and measuring the effect of each semantic category on the output. The effects are then visualized to explain the main contributors. The pixel-level segmentation allows us to produce not only saliency maps but also visual and text-based explanations which are one possible means to address the nonexpert user requirements.

D. K. Sugamya, et.al, (2016), have proposed a new approach in which firstly low level features are extracted and then noisy positive examples are handle using SVM classifier . In their work a image similarity are obtained by combining different distance metrics and multiple features suing SVM classifier. The proposed approach gives efficient results in terms of shape, color and texture. The SVM classifier is trained to distinguish between irrelevant and relevant images after the selection of features. N. Tripathi, et.al, (2017), have used multi kernel Support Vector Machine (SVM) and multi-feature method for contrast enhancement in CBIR . Earlier binary SVM were used and in this new extended kernel SVM are comes in existence. In today era there is peak development in multimedia technology and images can be retrieved on the basis of texture, surface, color and features of an object using CBIR. This has found its use in national security and medical science like technology. The similarity ranking for proper retrieval and image indexing is a main challenge in this system. To solve this problem Anjali T, et.al, (2018), have used firefly optimization with decision tree classifier that reduced the computational complexity in the classification stage of feature extraction . For evaluation purpose of their proposed approach they have used recall and precision rate that shows it is more efficient in CBIR system than existing approaches. In nature association rule is one of the most important rules and each type of object in remotely sensed image is related to special association rules. So for image classification an association rules is considered as important features and for accurate classification rational and mining selection of the effective rules is the key issue.

3 Implementation Study

Researchers have been focusing on integrating explanation facilities into computer vision tasks (e.g., image captioning, visual question answering, and object localization). Generally, these explanation models can be categorized as

intrinsically interpretable models and the post-hoc interpretability. Both categories produce saliency maps which simplify the image representations, making it easy to analyze. Saliency maps help to highlight the most important features for a prediction.

Interpretable models are usually specific to a certain model which focuses on the internal functioning of the model. Interpretable models analyse the interaction between neurons and what each neuron has learned. Post-hoc interpretability methods explain instance-specific predictions on the basis of how each feature (or a set of features) influences the final outcome. In general, both approaches can have some limitations and strengths. The results of interpretable models are directly explainable without requiring another model to generate explanations. However, they are limited to a specific learning model. Whereas the post-hoc interpretability methods are generally model-agnostic but they may be limited in their approximate nature.

Some of the intrinsically interpretable models applies gradients and decomposition based methods to understand the internal structure of complex black-box models. Gradient methods highlight the unit changes and emphasize the important features or regions in an image. In this way, it is possible to learn the prototypes that have the highest probability to be predicted as a certain class of a trained DNN. Considering this, Li et al. proposed an interpretable neural network architecture whose predictions are based on the similarity of an input to a small set of prototypes learned during training. As presented by Nguyen et al. , the activation maximization method synthesizes an image that highly activates a neuron and reveals the features learned by each neuron in an interpretable way. Some works proposed visual analysis by clustering important neurons based on the features and the interactions between them. Furthermore, decomposition methods such as layer-wise relevance methods are presented to analyze which pixels are contributing to what extent to a classification result. Some post-hoc explanations also proposed using gradients to create saliency maps. One way to visualize saliency maps is by going backward through the inverted network from an output of interest. It highlights the discriminative features of the image with respect to the given class. Another method uses class activation mapping with the gradients of a target input in the last convolutional layer to produce a rough localization map highlighting the important regions. This method is further developed for explaining occurrences of multiple object instances in a single image.

Moreover, other post-hoc explanation methods suggested modifying the input image and observing the effect on the prediction. Zeiler and Fergus proposed the occlusion sensitivity method to monitor the prediction score for a specific class by masking different parts of the input image with a grey square. Results are then visualized as saliency maps to show which parts of an image are most important for the classification. Another method that uses prediction score is LIME (Local Interpretable Model-Agnostic Explanation). This method generates explanations by approximating a black-box model by perturbing an input image. They adopt SLIC (Simple Linear Iterative Clustering) super pixel method to partition image into smaller regions and use these parts to perturb the image. Then, they present the super pixels with the highest positive weights as an explanation. Although these methods seem intuitive, they are inefficient in terms of salient features representation. In our approach, we propose observing the prediction result using semantic occlusion and measuring the effect of each semantic category on the output. The effects are then visualized to explain the main contributors. The pixel-level segmentation allows us to produce not only saliency maps but also visual and text-based explanations which is one possible means to address the non-expert user requirement

3.1proposed methodology

This section presents the visual and text explanations based on image annotations and automated semantic segmentation. We first experiment with manually annotated images to generate more detailed explanations. For the best exhibition of the method, we focus on the effect of context in a scene classification problem with deep neural networks. Following that,

we present CI's potential for contrastive explanations to generate “why?—why not?” explanations. In some contexts, contrastive explanations could be more useful for model debugging. Lastly, we experiment with an automated semantic segmentation method to assess whether the explanations generated based on the semantic segmentation model are similar to explanations generated from annotations.

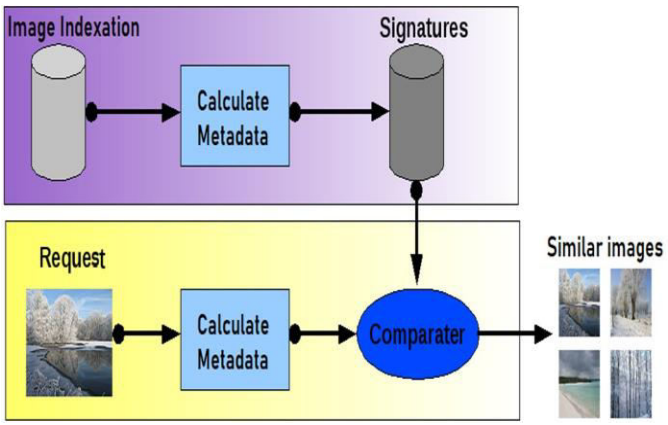

Fig 1: proposed Model

3.2 Methodology

3.2 1 Upload Covid& Diabetes dataset: button to upload dataset and to get below screen.

3,2.2 Build Context Based Image Diabetes Model: button to build machine learning model on above dataset and to get below screen.

3.2.3 Upload Test Data & Predict Disease:button to upload test data and then perform prediction on that test data.

3.3.3 Accuracy Comparison Graph: button to get below graph.

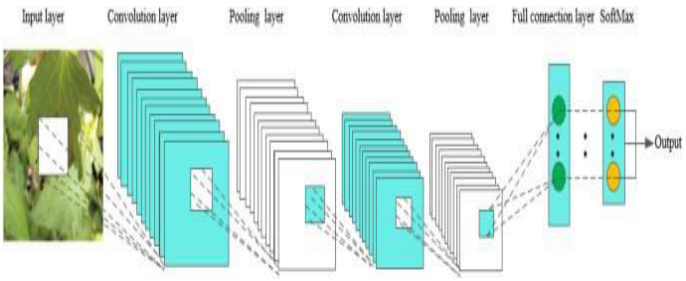

Fig 2:- CNN Model

4 Results and Evolution Metrics

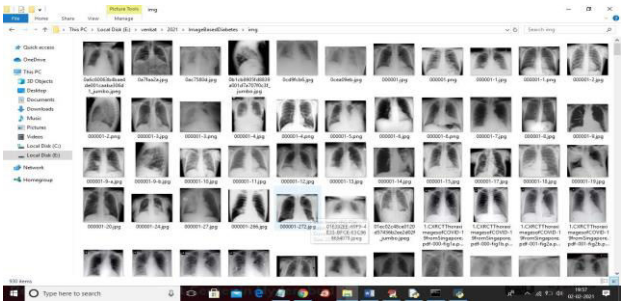

Fig 3:_ To implement diabetes we are using dataset which contains patient clinical notes given by doctor and if in this reports doctor describe about type 2 diabetes then machine learning get train with label 1 otherwise 0 label will be used and this dataset saved inside ‘dataset’ folder and below is the dataset screen shots which contains doctor reports for each patient

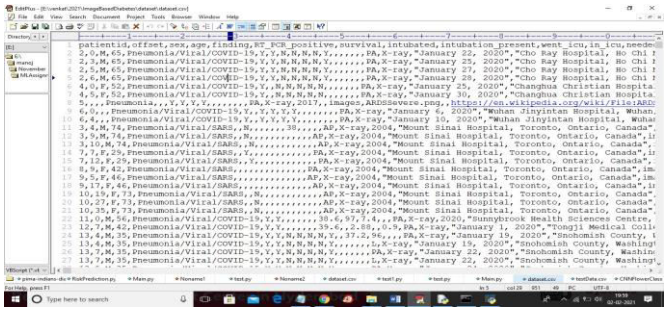

Fig 4 In above dataset we have patient id and in last column we can see doctor notes for each patient record.

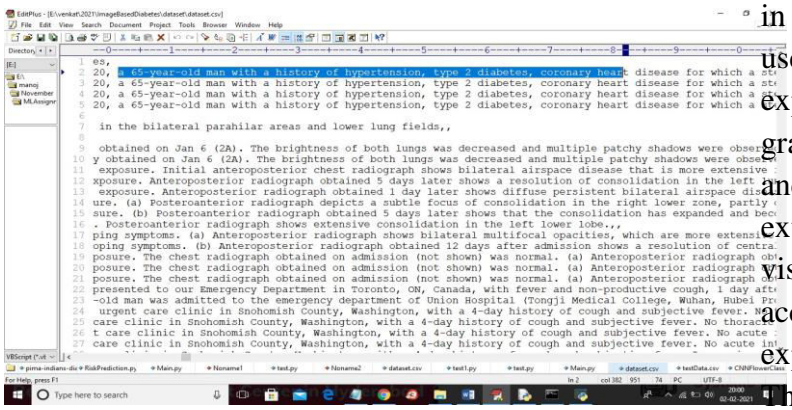

Fig 5: In above screen in dataset selected line we can see doctor given note about a patient and we used above datasets to train and predict diabetes and COVID diseases

Fig 6: In This Project We Are Using SVM Accuracy And CNN Accuracy

5 Conclusion

We presented context based visual explanation method for DNNs which can also be applied to other machine learning models for generating instance-level explanations. We have shown the implementation with both annotated and automatically segmented (semantically) images. Semantic information gives the properties of the image scene which enables us to assess the importance of each component for a prediction in a given context. We visualized the significance of each component contributing to the prediction score in a way that is easily understandable by common users. We generated visual and text based explanations using saliency maps, a colour bar graph, and descriptive phrases listing the features and their importance. These explanations could be extended to designing dynamic templates and visualizations by taking user's characteristics into account and testing the usability of these explanations in user studies.

The visual comparisons with three methods have shown varying saliency maps. Unlike these methods, our results present justifications beyond the vague representation of the important parts of an image. It uses semantic categories to distinguish the degree of importance of different parts of an image. We confirmed these findings using a user study and our claim about the elegance of the CI algorithm was validated. The results motivate current saliency map methods towards specifying not only where the network is pointing but also what it is seeing.

As the limitation of the work, the examples used are not those in which explanations are critical. In the future work, we will extend CI to be able to generalise for different input images (whether already semantically segmented or not). We will also apply CI to real-world critical examples (e.g., autonomous driving) and conduct a more immersive evaluation with different stakeholder categories in a lab setting.

6 References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art super pixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34
(11) (2012) 2274–2282.

[2] S. Anjomshoae, K. Främling, A. Najjar, Explanations of black-box model predictions by contextual importance and utility, International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, Springer 2019, pp. 95–109.

[3] S. Anjomshoae, T. Kampik, K. Främling, Py-ciu: a python library for explaining machine learning predictions using contextual importance and utility, IJCAI-PRICAI 2020 Workshop on Explainable Artificial Intelligence (XAI), 2020.

[4] S. Anjomshoae, L. Jiang, K. Främling, Visual explanations for DNNS with contextual importance, International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, Springer 2021, pp. 83–96.

[5] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PLOS ONE 10 (7) (2015) e0130140.

[6] O. Biran, K. McKeown, Justification narratives for individual classifications, Proceedings of the AutoML Workshop at ICML, vol. 2014 2014, pp. 1–7.

[7] G.J. Brostow, J. Fauqueur, R. Cipolla, Semantic object classes in video: a highdefinition ground truth database, Pattern Recognit. Lett. 30 (2) (2009) 88–97.

[8] V. Buhrmester, D. Münch, M. Arens, Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey, 2019 arXiv:1911.12116 (arXiv preprint).

[9] A. Chattopadhay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks, 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE 2018, pp. 839–847.

[10] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Proceedings of the European Conference on Computer Vision (ECCV) 2018, pp. 801–818.