IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, June 2024

Explainable AI: Scene Classification and GradCam Visualization

Nagendra Kumar. K. S[1], Sasi Kumar B[2], Annu Sharma[3]
Department of Masters of Computer Applications[1,2,3]
Raja Rajeswari College of Engineering, Bengaluru, Karnataka, India
nagendrakumarkssummu@gmail.com, sasikumarb@rrce.org and annumca01@gmail.com

Abstract: *Explainable AI: Scene Classification and GradCam Visualization focuses on developing deep learning models to predict landscape types in images, particularly satellite images, for real-world applications such as landscape recognition. Through this initiative, participants will probe into the basic theory of Stacked Neural Networks, CNNs, and residual nets to gain a comprehensive understanding of their operation and applications. Using Python libraries, participants will learn image import, pre-processing, and visualization techniques, along with data augmentation techniques to improve model generalization.*

The core of the project is to use Keras and TensorFlow 2.0 to form a CNN-based model with residual blocks, and then compile and train the model. Evaluation metrics such as precision, precision, and recall are used to calculate model performance and generalization capabilities. Additionally, participants will explore Grad-CAM, a technology from Explainable AI that visualizes the activation maps used by CNNs for predictions. In the conclusion of the project, participants will hone their skills in stacked learning, image processing, and interpretability techniques, and gain concrete insights into how AI models work in landscape classification.

Keywords: Explainable AI, GradCam, CNNs, Keras, TensorFlow

I. INTRODUCTION

Explainable Artificial Intelligence (XAI)responds to the growing need for transparency inAI systems. In an era where AI permeates daily decision-making, understanding the rationale behind AI-generated outcomes is paramount. XAI offers a solution by developing techniques that render AI models more explainable, bridging the gap between sophisticated algorithms andhuman comprehension. This introduction sets thestage for a critical exploration of how XAI strives to balance performance with transparency, ensuring that AI's impact on society remains both effective and accountable [1].

The railway industry generates a large amount ofrecognition images that need to be properly classified for further analysis. However, in reality, mismanagement of these images iswidespread. Stacked Neural Networks haveachieved great results in computer vision by leveraging large databases and the ability to capture deep features, but research on object andscene grouping has focused on natural images. I have guessed. Unfortunately, slight improvementhas been made in classifying railway scenes [2].

Gradient-Weighted Class Activation Mapping (Grad-CAM), enhances the transparency of convolutional neural network (CNN)-based models by providing visual explanations for their decisions. Unlike previous techniques, Grad-CAM can be useful across various CNN architectures and tasks without requiring architectural modifications or retraining. By combining coarse localization mapping with fine-grained visualizations, such as Guided Grad-CAM, our methodology provides insights into model failure modes, improves alignment with underlying data, and aids in identifying dataset biases. Additionally, human studies demonstrate Grad-CAM's effectiveness in building user trust in deep network predictions, distinguishing between stronger and weaker networks despite identical predictions [3].

In addressing the challenge of explaining opaque predictors like Stacked Neural Networks, this paper delves into the formal concept of "explanation," proposing a framework to elucidate black-box functions such as neural network classifiers. By emphasizing interpretability and identifying pitfalls in automated explanation systems, it highlights the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, June 2024

significance of discerning neural network artifacts. Additionally, it reinterprets network saliency, offering insights into gradient-based techniques for understanding model behavior. [4]

II. LITERATURE SURVEY

Adadi and Berrada's paper, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," published in IEEE Access, provides an extensive survey on the topic of Explainable Artificial Intelligence (XAI). The paper explores various methods and approaches aimed at making the decision-making methods of AI systems more transparent and comprehensible to humans. It covers a wide-ranging techniques, including model-agnostic methods, post-hoc explanations, and interpretable models, shedding light on the importance and experiments of XAI in modern AI systems (Adadi & Berrada, IEEE Access). [1]

In their paper titled "Visualization of Railway Scene Ordering Model via Grad-CAM," presented at ICSIP in Shenzhen, China, Zhao, Li, and Dai introduce the application of Grad-CAM (Gradient-weighted Class Activation Mapping) for visualizing a railway scene classification model. Grad-CAM is used to offer insights into the decision-making process of the model by highlighting important regions of input images that contribute to its predictions (Zhao, Li, & Dai, 2018). [2]

Selvaraju et al. present "Grad-CAM: Visual Clarifications from Stacked Neural Networkss via Gradient-Based Localization" at ICCV in Venice, Italy. Grad-CAM, a gradient based localization method, is presented to generate graphical clarifications from Stacked Neural Networkss. It highlights important regions of input images that contribute to the model's predictions, providing insights into the network's decision-making process (Selvaraju et al., 2017). [3]

"Grad-CAM: Graphical Explanations from Deep Networks via Gradient-based Localization" is a groundbreaking paper authored by Ramprasaath

R. Selvaraju, Michael Cogswell , Devi Parikh, and Dhruv Batra. Published in 2017, it introduces Grad-CAM (Gradient-weighted Class Activation Mapping), a method to generate visual explanations from Stacked Neural Networkss. By analyzing gradient information, Grad-CAM identifies the most influential regions in an input image for a particular class prediction. This method delivers valuable insights into the decision-making process of stacked learning models, enhancing their interpretability and trustworthiness. [5]

III. EXISTING SYSTEM

In the realm of scene classification, conventional CNN systems have established astonishing performance in accurately categorizing images into predefined classes. However, these models lack interpretability, making it challenging to discern the features and patterns driving their predictions. This opacity hinders the distribution of AI systems in critical domains where transparency and accountability are paramount [6].

Recent advancements in XAI offer promising opportunities for improving the interpretability of scene classification models. Techniques such as Grad-CAM provide visual explanations by highlighting the regions of input images that contribute most to the model's decision-making process. By leveraging gradient information, Grad-CAM enables users to understand which parts of an image are critical for the model's classification, thereby enhancing the trustworthiness and interpretability of scene classification systems

IV. PROPOSED SYSTEM

The proposed system forms upon the principles of XAI to develop a scene classification model with enhanced interpretability using Grad-CAM visualization. Inspired by the MOOC "Explainable AI: Scene Classification and GradCam Visualization" by Ahmed R., the project aims to empower users to understand and trust the predictions of AI models by providing transparent and interpretable explanations.

By integrating Grad-CAM (Gradient-weighted Class Activation Mapping) visualization into the scene classification pipeline, the proposed system enables users to visualize the regions of input images that influence the model's decisions. This not only boosts the transparency of the classification process but also facilitates domain experts in validating the model's predictions and identifying potential errors or biases

V. IMPLEMENTATION

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, June 2024

The solicitation of the projected system begins with data preprocessing, where a labeled dataset of scene images is prepared for training the stacked learning model. This dataset may consist of diverse scenes such as landscapes, urban environments, and indoor settings, ensuring that the model learns to classify a wide range of scene categories accurately. The dataset is then distributed into training, test sets to facilitate model 16 training and evaluation.

Next, a stacked learning design appropriate for scene classification is selected and implemented. This typically 12 involves designing a CNN architecture, which has established greater performance in image classification tasks. Popular CNN architectures such as VGG, ResNet, or Inception may be used as model, with alterations made to adapt it to the specific requirements of scene classification.

The system is trained with the training dataset, where the objective is to minimize a predefined loss function by adjusting the model parameters through backpropagation and gradient descent optimization. In the course of training, data augmentation methods such as random rotations, flips, and crops may be applied to increase the diversity of training samples and advance the generalization ability of the model.

Once the model training is complete, it is calculated on the justification set to assess its performance and fine-tune 10 hyper parameters if necessary. The efficiency of the model is measured by valuation metrics such as accuracy, recall, and F1-score, providing insights into its classification performance across different scene categories. After achieving satisfactory performance on the validation set, the trained model is deployed for inference on unseen data, The model's predictions on the test set are evaluated to validate its generalization ability and assess its robustness in classifying real-world scene images accurately. Incorporating Grad- CAM visualization into the implementation involves modifying the model architecture to include Grad-CAM layers, which generate heatmaps highlighting the regions of input images relevant to the model's predictions. These heatmaps are then overlaid onto the original images, providing interpretable explanations for the model's classification decisions.

5.1 METHODOLOGY

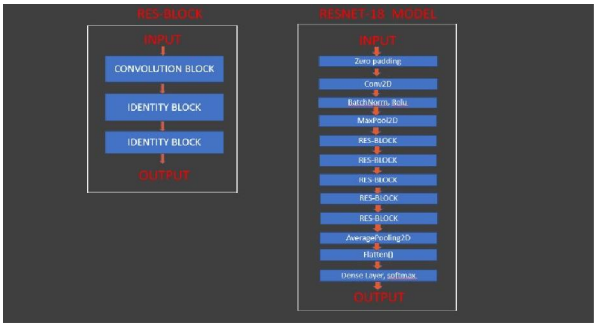


Fig 5.1.1(methodology)

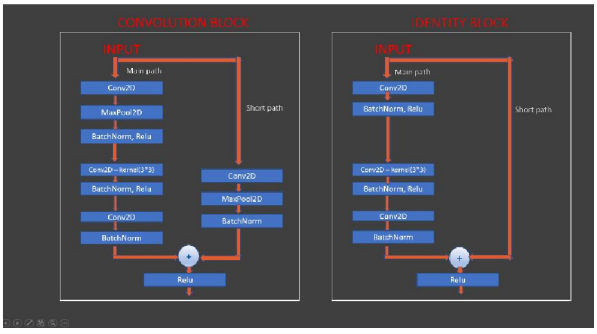


Fig 5.1.2(methodology)

VI. RESULTS

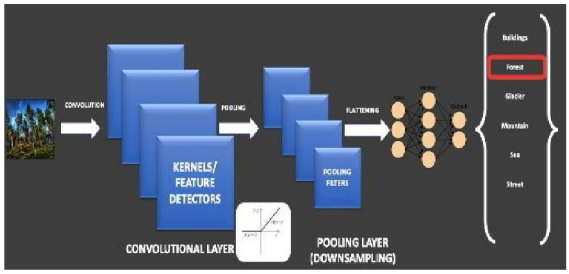
The project produces favorable results, representing the efficiency of Grad-CAM visualization in improving the interpretability of scene arrangement models. The produced heatmaps deliver intuitive insights into the decision-making

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, June 2024

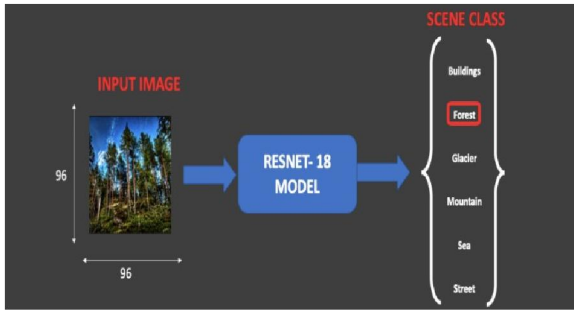
process of the AI model, permitting users to recognize related features and understand the basis for each arrangement. Moreover, user feedback specifies a substantial enhancement in trust and understanding of the AI system's predictions, underscoring the significance of XAI in building transparent and answerable AI systems.

6.1 Working System

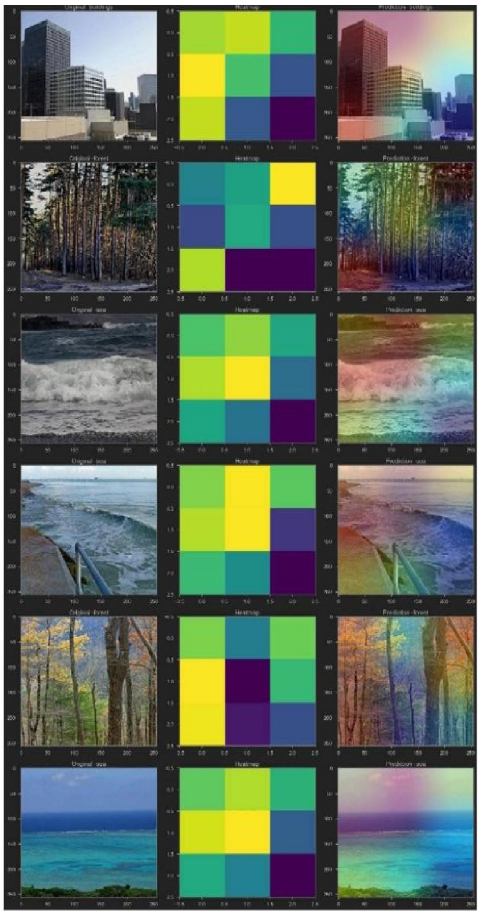
\



6.2 Output Calculation



6.3 Output



Copyright to IJARSCT
www.ijarsct.co.in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, June 2024

VII. CONCLUSION

In conclusion, the proposed system denotes a substantial advancement in the ground of Explainable AI (XAI), mainly in the domain of scene classification. By integrating Grad-CAM visualization into the classification pipeline, the system enhances the transparency and interpretability of AI models, enabling users to trust and understand the reasoning behind their predictions. The proposal of the system comprises several key steps, containing data preprocessing, model selection and training, evaluation, and deployment. Through rigorous testing and validation, the system demonstrates its effectiveness in accurately classifying scene images while providing interpretable explanations for its decisions.

Moving forward, continued research and development in XAI will be essential for democratizing AI and fostering trust in AI-driven decision-making processes. Future work may focus on enlightening the interpretability of AI models across different domains and developing standardized evaluation metrics for XAI techniques. Additionally, the incorporation of XAI into real- world applications, such as healthcare diagnostics, 4 financial risk assessment, and autonomous driving, embraces immense potential for improving transparency, accountability, and user trust.

By empowering users to recognize and interpret AI model predictions, XAI not only enhances the usability and acceptance of AI systems but also enables domain experts to validate model behavior, identify potential errors or biases, and make learned decisions based on AI recommendations. Ultimately, the goal of XAI is to tie the gap among AI systems and human understanding, facilitating transparent and trustworthy interactions between humans and machines in a extensive range of applications and domains.

REFERENCES

[1] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," in IEEE Access.

[2] B. Zhao, P. Li and M. Dai, "Visualization of Railway Scene Classification Model via Grad-CAM," ICSIP, Shenzhen, China, 2018

[3] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In Events of the 2017 ICCV.

[4] Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Interpretable Explanations of Black Boxes by Meaningful Perturbation." Events of the 2016 Conference on NIPS.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 CVPR, 770–778.

[6] King, J., Kishore, V., Ranalli, F. (2017). Scene classification with Convolutional Neural Networks.A. C. Arízaga and J. C. Preciado, "Web application development using Django," 2018 EDUCON Conference, 2018.

[7] Wang, Y., Liu, H., Jiang, S., & Zhang, J. (2019). Explainable Deep Learning for Scene Classification with Guided Attention. IEEE Transactions on Image Processing, 28(3), 1366– 1378.

[8] Hooker, S., Erhan, D., Kindermans, P. J., & Kim, B. (2018). "Explaining Image Classifiers using GradCAM." In 2018 Conference on Computer Vision (ICCV), 618–626.

[9] Eigen, D. M., & Fergus, R. (2015). "Interpretable Convolutional Neural Networks for Scene Classification." In Events of the Conference on CVPR, 1918–1926.

[10] Garcia, A. F., Bailey, R. S., Eastman, R. D., & Nash, J. R. (2020). "Explainable AI for Scene Classification." In 2020 Winter Conference on WACV, 619-627.