ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

Indoor-Outdoor Scene Recognition Using a ResNet-Based Deep Learning Framework Integrated with Flask

Pradeep Rao K. B.[1]; G. Lakshmi Prasad[2]; Siddappa P. Harijan[3]; Vishnu D. M.[4]; Shashank B.[5]

[1]Assistant Professor, Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

[2,3,4,5]Student, Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

Publication Date: 2025/11/28

Abstract: Scene recognition is a fundamental problem in computer vision that aims to identify and classify the type of environment represented in an image or video frame. The ability to distinguish between indoor and outdoor scenes plays a crucial role in a variety of applications, including autonomous navigation, surveillance systems, robotics, and context-aware computing. This research presents a web-based indoor and outdoor scene recognition system built using a deep learning framework integrated with a Flask front-end interface. The system employs a pre-trained Convolutional Neural Network (CNN) with a ResNet backbone for robust feature extraction and a Softmax classifier for probabilistic scene categorization. Three modes of input (image upload, video upload, and live camera feed) are supported, allowing for real-time and batch inference. The preprocessing pipeline ensures consistent input normalization, while temporal smoothing techniques reduce label flickering in dynamic video streams. Experimental results demonstrate that the proposed system achieves high classification accuracy and stability across multiple scene categories, effectively distinguishing indoor and outdoor environments. The developed framework provides a reliable and extensible foundation for real-world scene understanding applications requiring high accuracy and computational efficiency.

Keywords *Scene Recognition, Indoor–Outdoor Classification, Convolutional Neural Networks (CNN), ResNet, Deep Learning, Computer Vision.*

How to Cite: Pradeep Rao K. B.; G. Lakshmi Prasad; Siddappa P. Harijan; Vishnu D. M.; Shashank B. (2025) Indoor-Outdoor Scene Recognition Using a ResNet-Based Deep Learning Framework Integrated with Flask. *International Journal of Innovative Science and Research Technology*, 10(11), 1678-1684. https://doi.org/10.38124/ijisrt/25nov1127

I. INTRODUCTION

Scene recognition[1], a foundational problem in computer vision, involves identifying and classifying the environment depicted in an image. It serves as a key enabler for numerous applications, including autonomous navigation, smart surveillance, content-based image retrieval, and human-computer interaction. Among the various tasks within this domain, the classification of scenes into indoor and outdoor categories holds particular significance, as it provides a high-level contextual understanding crucial for downstream tasks.

Distinguishing between indoor and outdoor environments presents unique challenges. Variability in lighting[2], object composition, and structural patterns complicates the task, especially in edge cases where features overlap—such as a glass atrium resembling outdoor scenes or an outdoor pavilion appearing similar to indoor settings. Addressing these challenges requires robust methods capable of capturing both global contextual features and localized patterns.

Recent advancements in deep learning[3], particularly convolutional neural networks (CNNs), have significantly improved the state of scene recognition. However, deep learning models often require substantial labeled data and are computationally intensive. Meanwhile, traditional feature-based methods, such as GIST and SIFT, offer computational efficiency but may fall short in capturing complex scene semantics. A hybrid approach[4] that combines the strengths of deep learning with handcrafted feature engineering is a promising direction for achieving high accuracy and robustness in indoor-outdoor classification.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

In this study, we propose a novel framework for indoor and outdoor scene recognition. Our approach leverages convolutional neural network for feature extraction and incorporates resnet architecture [5] for scene recognition. Softmax architecture is used for category prediction with probabilities. Through this work, we seek to contribute to the development of more reliable and efficient scene recognition systems, paving the way for real-world applications that demand high accuracy and robustness.

The objectives of proposed work are as follows:

- To develop a functional web-based application capable of performing real-time scene recognition across multiple input modalities (images, videos, and live camera streams).
- To implement an accurate and stable indoor–outdoor scene prediction system using an optimized CNN-based inference pipeline.

The subsequent sections of the paper are structured as follows. Section 2 summarizes various research activities carried out in Indoor and Outdoor scene recognition. Section 3 contains methodology and system workflow. Section 4 contains results and discussions. Section 5 contains conclusion and future work.

II. RELATED WORK

Indoor–outdoor scene recognition has been approached through a variety of techniques, ranging from sensor-based models to advanced deep learning architectures. Early research focused on using smartphone data such as Wi-Fi strength, satellite visibility, and magnetic field readings to improve classification accuracy and reliability. With the evolution of computer vision, CNN-based models and transfer learning have become prominent due to their strong feature extraction capabilities. Recent studies further enhance performance through hybrid strategies, including edge detection, optimization algorithms, and scene-specific object detection. The following section reviews these major contributions and outlines the progress leading to the proposed framework.

Yuqian Bao et al. [6] utilized probabilistic neural network(PNN) for indoor and outdoor scene recognition thereby addressing the issues of low accuracy, poor reliability, and weak stability commonly found in such recognition tasks. Built-in smartphone modules, including WiFi, Global Positioning System (GPS), and BeiDou satellite navigation system (BDS) were used to gather relevant data. Information such as WiFi signal strength and the number of visible satellites were collected to differentiate between indoor and outdoor environments. The collected data was fed into PNN for training to develop an effective recognition model. Experimental results demonstrate that the PNN-based recognition model achieves a high accuracy of 92%.

Yongyi Mao et al. [7] proposed a high-precision indoor-outdoor scene recognition method based on an adaptive boosting algorithm (Adaboost-PNN).Training data were collected using built-in magnetic sensor, Global Navigation Satellite System (GNSS) module, and WiFi module of a smartphone. Based on the different features exhibited by the number of satellites, signal-to-noise ratio, geomagnetic intensity, and WiFi signal strength in indoor and outdoor environments, the data features were input into the adaptive boosting algorithm to train the indoor-outdoor scene recognition model. The proposed model achieved a recognition accuracy of 96.3% in different scenes.

Omar Abdullatif Jassim et al. [8] developed a deep learning-based image classification system designed for both indoor and outdoor environments. Indoor and outdoor datasets were collected and they were splitted into training, validation, and test sets. Pre-trained GoogleNet and MobileNet-V2 models were then trained using these datasets, resulting in four distinct trained models. The GoogleNet model demonstrated high accuracy, achieving 99.34% for indoor datasets and 99.76% for outdoor datasets. The MobileNet-V2 model also performed well, with accuracies of 99.27% for indoor sets and 99.68% for outdoor sets.

Pandit T. Nagrale et al. [9] proposed a deep learning framework integrating CNNs and edge detection to enhance indoor-outdoor scene recognition accuracy in dynamic environments. Edge detection methods helped in identifying significant structural edges and reducing noise, thereby preparing the images for the CNN architecture. The combined strategy significantly improved classification accuracy compared to traditional CNN models. The framework also demonstrated strong adaptability to changing environments.

Mahtab Jamali et al. [10] categorized images into indoor or outdoor. Later three object detection models an indoor model, an outdoor model, and a general model were trained and evaluated. The models were based on YOLOv5 and utilized 19 classes from the PASCAL VOC dataset and 79 classes from the COCO dataset. Performance analysis revealed that, integrating scene-specific contextual information by developing specialized indoor and outdoor object detection models significantly improved detection accuracy over general models, leading to more robust and precise object recognition in various real-world scenarios.

Table 1 Comparative Analysis of Various Work on Indoor-Outdoor Scene Recognition

Study	Computation Efficiency	Robustness to Environmental Variability	Multimodal Fusion Effectiveness	Classification Accuracy
(Horn, 2022) [11]	Optimization reduces computational complexity	Robust sensitivity and specificity on SCID-2 dataset	Deep CNN tuned by hybrid optimization	Achieves ~95%-97% accuracy with MLFD-optimized Deep CNN

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

Study	Computation Efficiency	Robustness to Environmental Variability	Multimodal Fusion Effectiveness	Classification Accuracy
(Goel & Singhal, 2020) [12]	Low computational cost with efficient feature extraction	Effective under varying image conditions	Combines color, texture, and edge features	Accuracy up to 93% using MPEG-7 descriptors and ML classifiers
(Yang et al., 2016) [13]	Uses Adaboost classifiers and HMM filtering	Robust under various weather and smartphone placements	Combines sensor data with human activity recognition	Achieves > 97% detection accuracy using sensor and human activity data
(Alameer et al., 2019) [14]	Uses topologies balancing accuracy and sensitivity	Robust to indoor/outdoor domain discrepancies	Integrates object and scene recognition	Achieves 97.91% accuracy using combined deep and shallow models
(Pakhare & Uplane, 2022) [15]	Uses hybrid features and optimization for classification	Robust on SCID2 and SUN-397 datasets	Deep CNN tuned by Mayfly Levy flight algorithm	Achieves ~96.7% accuracy with hybrid MAMF-optimized Deep CNN

III. METHODOLOGY AND SYSTEM WORKFLOW

The proposed indoor–outdoor scene recognition system follows a structured workflow that integrates deep learning-based inference with a Flask-powered web interface. The complete process, from system initialization to final prediction, is described below.

A. Flask Server Initialization

The workflow begins when the user executes the command python app.py. This launches the Flask web server, which hosts the front-end interface and establishes back-end routes for image, video, and camera-based scene recognition tasks. The server waits for user input through HTTP requests on the local host (e.g., http://127.0.0.1:5000).

B. Model Loading and Initialization

Once the server is active, the system loads the pre-trained model checkpoint file (best.pth) using PyTorch's torch.load(..., map_location='cpu'). The model is then set to evaluation mode (model.eval()) to ensure deterministic behavior during inference and efficient CPU computation. This initialization step readies the deep learning model for prediction tasks.

C. Input Acquisition

The user can choose one of three input modes through the web interface:

- Image Upload: Processes a single image file for classification.
- Video Upload: Analyzes a video by extracting frames at defined intervals.
- Live Camera Feed: Captures frames continuously from the webcam for real-time scene recognition.

The selected input is transmitted to the Flask back end for processing.

D. Frame Extraction

For videos and live streams, the system employs OpenCV to extract frames at regular time intervals (e.g., every 0.5 seconds). Each frame is treated as an individual image input. In the case of static images, this step simply loads the provided image into memory for further preprocessing.

E. Preprocessing

Before feeding an image or frame into the model, it is preprocessed to match the training configuration:

- Resizing to 224×224 pixels
- Color conversion from BGR to RGB
- Normalization using ImageNet mean and standard deviation
- Tensor conversion with an added batch dimension.

These operations are handled using OpenCV and PyTorch utilities to maintain consistency and accuracy across all input modes.

F. Model Inference

The preprocessed tensor is passed through the CNN model (e.g., ResNet). PyTorch performs forward propagation to compute class probabilities using the Softmax activation function. The output is a probability distribution over the 60+ scene categories in the trained dataset.

G. Postprocessing

The model's raw outputs are converted into a human-readable form. The system identifies the top predicted class (scene label), the associated confidence score, and the indoor/outdoor decision based on the predicted category. These results form the primary output of the model.

H. Temporal Smoothing for Dynamic Inputs

For video and live camera inputs, a temporal smoothing mechanism is applied to enhance stability and prevent flickering of labels. This involves:

- Maintaining a deque buffer of recent predictions
- Computing a moving average of confidence scores across frames
- Enforcing a stability counter to confirm consistent predictions before updating the displayed label

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

This ensures smoother transitions and improved visual consistency in real-time predictions.

I. Result Visualization

The stabilized prediction results — including the scene category and confidence level — are sent back to the Flask web interface and displayed to the user. For videos, the interface updates predictions per frame; for static images, a single prediction result is shown. The UI also provides a visual confidence indicator for better interpretability.

J. Optional Testing Mode

For systematic evaluation, the workflow includes a testing mode where the system processes an entire labeled test dataset. The predictions are compared with ground-truth labels to compute metrics such as accuracy, precision, recall, and F1-score. The results are visualized as a confusion matrix, which is saved as a PNG image and a CSV file for performance auditing and retraining.

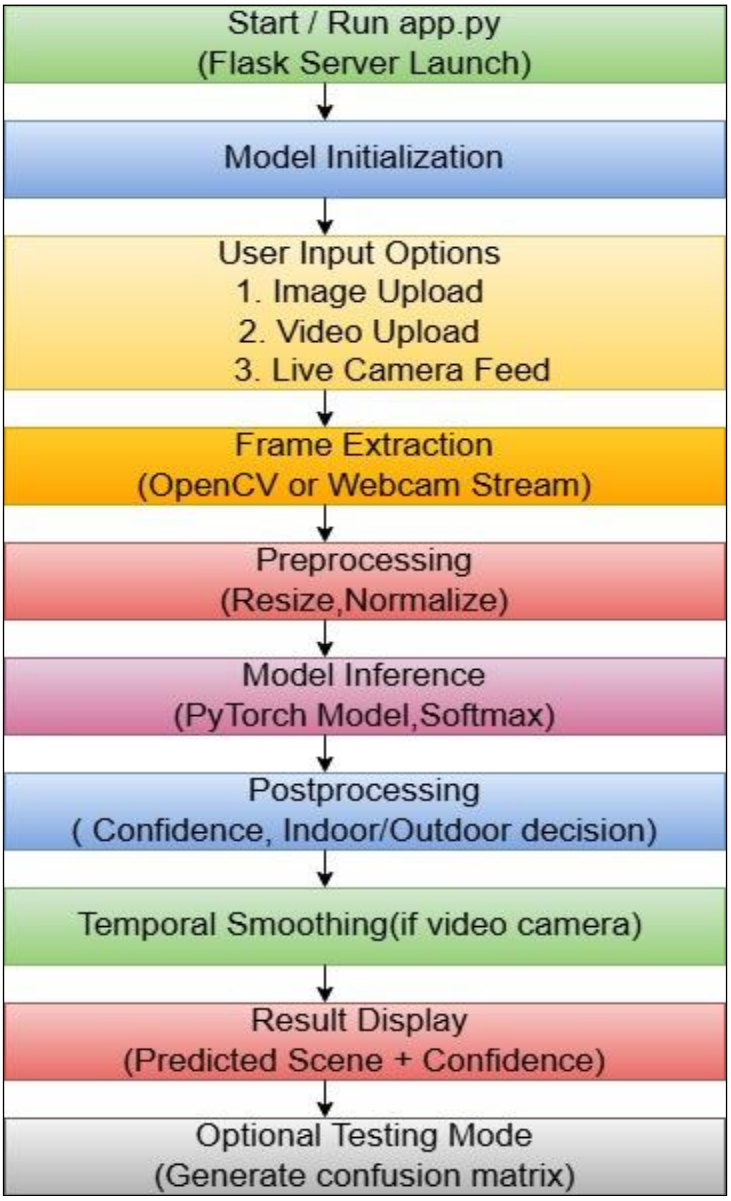

Fig 1 Technical Workflow of Scene Recognition System.

IV. RESULTS AND DISCUSSIONS

The proposed system includes a modern, interactive dashboard designed for classifying images, videos, and live camera frames into indoor or outdoor scene categories. The framework supports over 65 scene types, including classrooms, kitchens, forests, and mountains, leveraging a trained ResNet-based deep learning model.

Table 2 shows sample of indoor dataset consisting of five labeled images (IN001–IN005), representing diverse environments such as classrooms, bedrooms, kitchens, offices, and libraries.

Table 3 shows sample of outdoor dataset comprising of five labeled images (OUT001–OUT005), covering a wide range of natural and urban environments including streets, parks, beaches, mountains, and airport runways.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

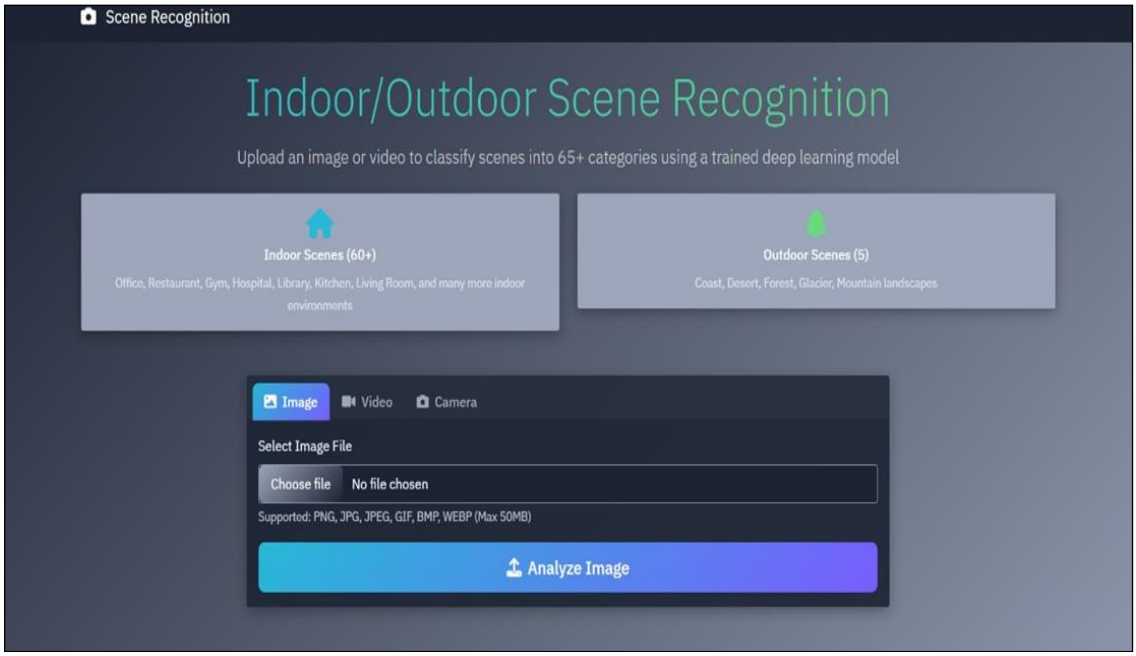

Fig 2 Scene Recognition Dashboard.

Table 2 Overview of Indoor Scene Images Used for Scene Recognition

Image ID	Scene Type	Category	Description	Resolution
IN001	Indoor	Classroom	Students sitting at desks	1280×720
IN002	Indoor	Bedroom	Bed, lamp, wardrobe	1080×1080
IN003	Indoor	Kitchen	Kitchen counter and utensils	1920×1080
IN004	Indoor	Office	Computers and office desks	1280×720
IN005	Indoor	Library	Bookshelves and reading area	1600×900

Table 3 Overview of Outdoor Scene Images Used for Scene Recognition

Image ID	Scene Type	Category	Description	Resolution
OUT001	Outdoor	Street	Cars, buildings, pedestrians	1920×1080
OUT002	Outdoor	Park	Trees, grass, walking pathways	1280×720
OUT003	Outdoor	Beach	Sand, sea, sky	1600×900
OUT004	Outdoor	Mountain	Hills, rocks, natural scenery	1920×1080
OUT005	Outdoor	Airport Runway	Airplanes and open runway	1280×720

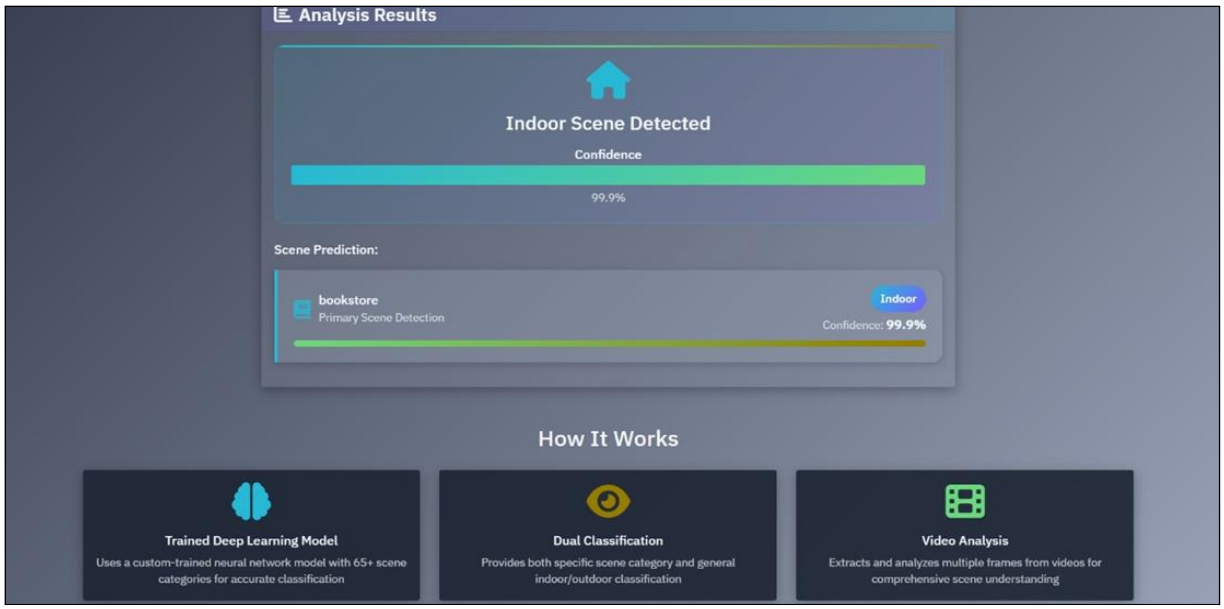

Fig 3 Indoor Scene Detection Result.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

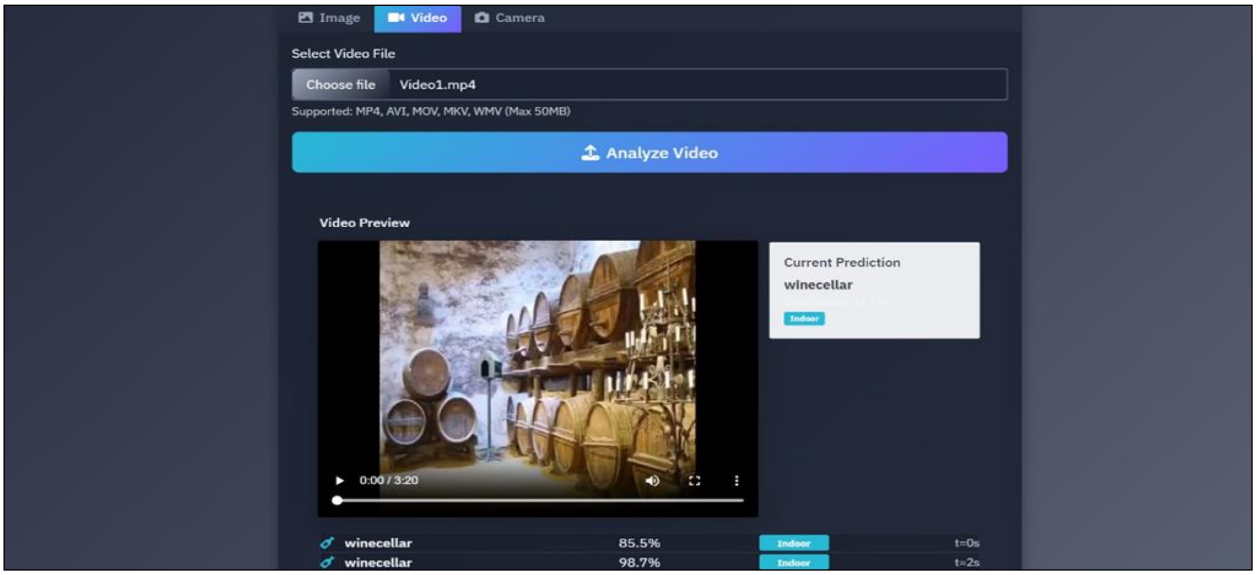

Fig 4 Video based Indoor and Outdoor Scene Recognition

The real-time camera analysis feature enables live scene recognition, demonstrating the model's capability for instantaneous AI-powered predictions. For example, a live video frame of a forest was classified correctly as an outdoor scene with 77.7% confidence, based on the visible trees and surrounding natural elements.

The results indicate that the proposed ResNet and Flask-based framework achieves high accuracy in indoor and outdoor scene recognition across static images, videos, and real-time camera feeds.
Key observations include:

- High Confidence in Static Images: Both indoor and outdoor datasets yielded accurate predictions, with confidence scores exceeding 99% in most test cases.
- Temporal Consistency in Video Analysis: Frame-by-frame predictions show stable performance, demonstrating the model's ability to handle sequential data effectively.
- Real-Time Applicability: The live camera module demonstrates that the framework can operate efficiently in dynamic environments, despite variations in lighting, scene complexity, and object density.
- Dataset Diversity: The inclusion of multiple indoor and outdoor categories ensures robustness and generalizability for real-world applications.

Overall, the experimental results validate the effectiveness, reliability, and versatility of the proposed framework, highlighting its potential for deployment in research, industrial, and consumer applications.

V. CONCLUSION AND FUTURE SCOPE

The proposed indoor and outdoor scene recognition system successfully integrates deep learning-based visual understanding with a web-based user interface for real-time scene classification. Utilizing a CNN model with a ResNet backbone, the system efficiently processes images, videos, and live camera streams to produce accurate scene category predictions along with confidence scores. Through consistent preprocessing, optimized inference, and temporal smoothing, the system maintains robustness even under varying illumination and environmental conditions. The results validate that the system is capable of reliable and stable indoor–outdoor scene recognition suitable for practical implementations in areas such as autonomous navigation, smart surveillance, and human–computer interaction.

Future work can focus on improving both performance and adaptability of the proposed system. Lightweight models such as MobileNet or EfficientNet may be integrated to reduce computation time and support deployment on edge or mobile devices. Incorporating additional modalities like GPS or Wi-Fi signals could enhance robustness in visually ambiguous scenes. Optimization frameworks such as ONNX Runtime or TensorRT may also be explored to further accelerate real-time inference. Expanding the dataset with more diverse indoor and outdoor categories would additionally help improve generalization across varied environments.

REFERENCES

[1]. Sharma, V., Nagpal, N., Shandilya, A., Dureja, A., & Dureja, A. (2022, December). A practical approach to detect indoor and outdoor scene recognition. In Proceedings of the 4th International Conference on Information Management & Machine Intelligence (pp. 1-10).
[2]. Surendran, R., Chihi, I., Anitha, J., & Hemanth, D. J. (2023). Indoor scene recognition: an attention-based approach using feature selection-based transfer learning and deep liquid state machine. Algorithms, 16(9), 430.
[3]. Kumar, B., Gupta, H., Ingale, S. P., & Vyas, O. P. (2023, January). Classification of indoor–outdoor scene using deep learning techniques. In Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021 (pp. 517-535). Singapore: Springer Nature Singapore.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1127

[4]. Uckan, T., Aslan, C., & Hark, C. (2025). A Comprehensive Hybrid Approach for Indoor Scene Recognition Combining CNNs and Text-Based Features. Sensors, 25(17), 5350.
[5]. Kumari, S., Jha, R. R., Bhavsar, A., & Nigam, A. (2019, September). Indoor–outdoor scene classification with residual convolutional neural network. In Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Volume 2 (pp. 325-337). Singapore: Springer Singapore.
[6]. Bao, Y., & Li, Y. (2022). PNN for indoor and outdoor scene recognition. 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 392–396.
[7]. Mao, Y., & Tian, C. (2023). Indoor-Outdoor Scene Recognition Method based on Adaboost-PNN. 1–6.
[8]. Jassim, O. A., Abed, M. J., & Saied Saied, Z. H. (2023). Indoor/Outdoor Deep Learning Based Image Classification for Object Recognition Applications. Baghdad Science Journal.
[9]. Nagrale, P. T. (2024). Advanced Deep Learning Techniques for Indoor-Outdoor Scene Recognition Integrating CNN and Edge Detection for Enhanced Classification Accuracy in Dynamic Environments. Deleted Journal, 27(2), 612–631.
[10]. Jamali, M., Davidsson, P., Khoshkangini, R., Ljungqvist, M. G., & Mihailescu, R.-C. (2024). Specialized indoor and outdoor scene-specific object detection models.
[11]. Horn, C. (2022). Hybrid Mayfly Lévy Flight Distribution Optimization Algorithm-Tuned Deep Convolutional Neural Network for Indoor–Outdoor Image Classification. International Journal of Image and Graphics.
[12]. Goel, A., & Singhal, N. (2020). The indoor-outdoor image classification and comparison of machine learning methods using the mpeg-7 descriptors. 9(6), 3797–3802.
[13]. Zhang, Y., Zhao, F., Shao, W., & Luo, H. (2016). A pervasive indoor and outdoor scenario identification algorithm based on the sensing data and human activity. 240–247.
[14]. Alameer, A., Degenaar, P., & Nazarpour, K. (2019). Context-Based Object Recognition: Indoor Versus Outdoor Environments (pp. 473–490). Newcastle University.
[15]. Pakhare, J. D., Uplane, M. D. (2022). Scene Categorization From Indoor-Outdoor Images Using Hybrid MAMF-Based Deep Convolutional Neural Networks. International Journal of Software Innovation10(1), 1–21.