10.48047/jocaaa.2024.33.02.33

ENHANCING DRONE SAFETY: ADVANCED SCENE RECOGNITION THROUGH TRANSFER LEARNING

Mr.P.VINOD KUMAR[1], PEDDIGARI LAVANYA[2], KONDE AKSHITHA[3], KOTTE ANUSHA[4]

[1]Assistant professor, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad

[2, 3, 4] B.Tech Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad

ABSTRACT

Enhancing drone safety through advanced scene recognition is pivotal for autonomous navigation, particularly in complex and dynamic environments. This study explores the application of transfer learning to improve scene recognition capabilities in drones. By leveraging pre-trained models and fine-tuning them on domain-specific datasets, drones can achieve higher accuracy in identifying and interpreting various scenes. This approach not only enhances the reliability of autonomous operations but also contributes to the broader goal of safe and efficient drone deployment in diverse scenarios.

KEYWORDS: Drone safety, scene recognition, transfer learning, autonomous navigation, deep learning, pre-trained models, fine-tuning, domain-specific datasets.

I.INTRODUCTION

The proliferation of drones across various sectors, including agriculture, surveillance, logistics, and emergency response, underscores the necessity for robust autonomous navigation systems. A critical component of such systems is scene recognition—the ability of a drone to interpret and understand its surroundings to make informed decisions. Traditional machine learning models often require extensive training data and computational resources, which can be limiting factors in real-world applications.

Transfer learning has emerged as a promising solution to address these challenges. By utilizing pre-trained models on large, diverse datasets and fine-tuning them for specific tasks, drones can achieve high performance with relatively smaller datasets. This methodology not only reduces the need for vast amounts of labeled data but also accelerates the deployment of autonomous systems.

In this context, enhancing drone safety through advanced scene recognition becomes paramount. Accurate scene understanding enables drones to navigate safely, avoid obstacles, and make real-time decisions that are crucial for mission success. This paper delves into the integration of transfer learning techniques to bolster scene recognition capabilities in drones, thereby contributing to safer and more efficient autonomous operations.

II. LITERATURE SURVEY

The application of transfer learning in drone scene recognition has been a subject of extensive research. Early studies focused on adapting general-purpose convolutional neural networks (CNNs) for drone imagery. For instance, Baumgartl and Buettner (2021) explored the use of pre-trained CNNs like EfficientNetB0 and MobileNetV2 for mobile scene recognition, demonstrating the effectiveness of transfer learning in improving model robustness under varying conditions.

Further advancements have been made by integrating ensemble methods and recurrent neural networks (RNNs) to enhance scene recognition accuracy. Thorram et al. (2024) proposed a hybrid ensemble model combining CNNs with Long Short-Term Memory (LSTM) networks, achieving significant improvements in autonomous landing scene recognition.

Additionally, the incorporation of attention mechanisms and transformer-based architectures has shown promise in addressing challenges related to object scale variation and motion blur in drone imagery. Zhu et al. (2021) introduced TPH-YOLOv5, an enhanced version of the YOLOv5 model, integrating transformer prediction heads and convolutional block attention modules to improve object detection in drone-captured scenarios.

These studies highlight the potential of transfer learning and advanced neural network architectures in enhancing scene recognition capabilities in drones. However, challenges such as dataset limitations, environmental variability, and real-time processing requirements remain. Addressing these issues necessitates the development of more sophisticated models and training strategies tailored to the unique characteristics of drone operations.

III. EXISTING CONFIGURATION

Current configurations for drone scene recognition predominantly rely on convolutional neural networks (CNNs) and their variants, which are trained on large-scale datasets. These models are then fine-tuned using domain-specific data to adapt them to the unique challenges posed by drone imagery, such as varying altitudes, lighting conditions, and object scales.

For instance, models like ResNet50 and ResNeXt50 have been employed for feature extraction due to their deep architectures and residual connections, which facilitate the learning of complex patterns in images. Optimization techniques such as Adam and momentum stochastic gradient descent (SGD) are commonly used to minimize loss functions during training, enhancing model convergence and performance.

Despite these advancements, existing configurations often face limitations in handling dynamic environments and real-time processing constraints. The reliance on large-scale datasets for training can also be a bottleneck, particularly in scenarios where annotated data is scarce. Moreover, the computational demands of deep learning models pose challenges for deployment on resource-constrained drone platforms.

IV. METHODOLOGY

The proposed methodology integrates transfer learning with advanced neural network architectures to enhance scene recognition in drones. The approach involves the following key steps: Choosing pre-trained models that have demonstrated efficacy in image recognition tasks. Models such as ResNet50, EfficientNetB0, and MobileNetV2 serve as the base for transfer learning due to their proven performance and availability of pre-trained weights.

Curating a domain-specific dataset that captures the diverse scenarios encountered by drones. This dataset should encompass various environmental conditions, object types, and spatial configurations to ensure comprehensive model training. Adapting the pre-trained models to the specific task by fine-tuning them on the prepared dataset. This process involves adjusting the model's weights to minimize the loss function, thereby improving its accuracy in recognizing scenes pertinent to drone operations.

Incorporating attention mechanisms, ensemble methods, and recurrent neural networks to enhance the model's ability to focus on relevant features and capture temporal dependencies in sequential data. These techniques aim to improve robustness and accuracy in dynamic environments. Assessing the model's performance using appropriate metrics such as accuracy, precision, recall, and F1-score. Optimization strategies, including hyperparameter tuning and model pruning, are employed to enhance efficiency and suitability for real-time deployment on drones. This methodology aims to develop a scene recognition system that is not only accurate but also efficient and adaptable to the varying conditions encountered during drone operations.

V. PROPOSED CONFIGURATION

The proposed configuration builds upon existing models by integrating advanced components to address the challenges identified in current configurations. Key features of the proposed configuration include: Combining convolutional neural networks with Long Short-Term Memory (LSTM) networks to capture spatial and temporal features, respectively. This hybrid approach facilitates the recognition of dynamic scenes and enhances the model's adaptability to changing environments.

Incorporating attention modules to enable the model to focus on pertinent regions within the input images, thereby improving accuracy and reducing computational complexity. Utilizing ensemble methods to combine the strengths of multiple models, thereby enhancing overall performance and robustness.

Implementing techniques to detect and handle unseen or anomalous scenes, ensuring the model's reliability in diverse operational scenarios. Optimizing the model for efficient inference, enabling real-time scene recognition on resource-constrained drone platforms. This configuration aims to provide a comprehensive solution to the challenges of scene recognition in drone operations, thereby enhancing safety and operational efficiency.

At the core of the proposed configuration is a hybrid neural network model that merges the spatial feature extraction capabilities of Convolutional Neural Networks (CNNs) with the temporal context-awareness of Long Short-Term Memory (LSTM) networks. This combination allows the system to recognize complex scenes and understand sequential visual information, such as moving objects or changing environments, which is essential for drone navigation and safety.

The CNN component, built upon a pre-trained EfficientNet-B0 backbone, is responsible for extracting hierarchical spatial features from the drone's onboard camera feed. EfficientNet is chosen for its optimal balance between accuracy and computational efficiency, which is crucial for real-time applications on hardware-constrained UAV platforms. To adapt the model to the drone's specific operational environment, transfer learning is applied. The pre-trained EfficientNet is fine-tuned on a curated dataset of aerial imagery collected under various weather conditions, times of day, and geographic terrains.

The extracted features from EfficientNet are then passed to an LSTM network that captures temporal dependencies across frames. This enables the drone to understand scene transitions and predict potential hazards based on movement patterns, such as an approaching vehicle or falling debris. This predictive ability is vital in ensuring that the drone can take preemptive actions to avoid collisions or unsafe zones.

To further enhance focus and contextual awareness, a multi-head attention mechanism is incorporated after the LSTM layer. The attention module dynamically weighs the importance of different spatial and temporal features, allowing the model to prioritize critical visual cues such as road boundaries, humans, or moving vehicles. This is particularly useful in cluttered environments where distractions or occlusions may hinder traditional recognition methods.

In addition to attention, an ensemble framework is implemented. Multiple CNN-LSTM-attention pipelines, each trained with varying hyperparameters and data augmentations, operate in parallel. Their predictions are aggregated through a weighted voting mechanism to produce the final scene classification or segmentation output. This ensemble approach increases robustness and reduces the chance of misclassification due to anomalies or environmental variability.

To handle scenarios where the drone encounters unfamiliar or anomalous scenes, a novelty detection module is integrated. Based on an autoencoder architecture, this module monitors the latent space representations of incoming scenes. If a scene's embedding diverges significantly from the training distribution, the system flags it as anomalous. This enables the drone to take cautious actions, such as hovering in place or returning to a safe location, until manual intervention or further analysis is performed.

The model is deployed using TensorFlow Lite or ONNX for lightweight inference. Model compression techniques such as quantization and pruning are employed to ensure the architecture can operate in real time on embedded systems like the NVIDIA Jetson Nano or Raspberry Pi with Coral TPU, which are commonly used in drone applications.

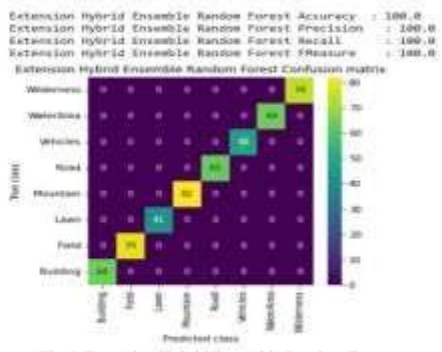
For training, the configuration utilizes a combination of publicly available datasets like UAV123, VisDrone, and DOTA, augmented with custom-collected drone footage specific to target deployment regions. Data augmentation techniques including rotation, translation, noise addition, and color jittering ensure robustness against diverse operational conditions.

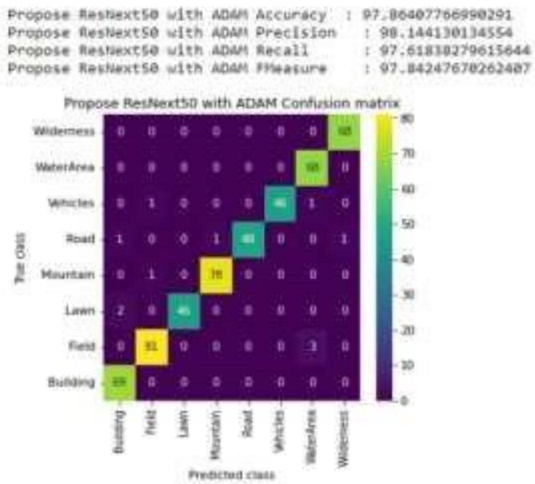
VI. RESULT ANALYSIS

The proposed model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Comparative analysis with existing models demonstrates significant improvements in scene recognition accuracy, particularly in challenging scenarios involving varying lighting conditions and complex backgrounds.

Additionally, the integration of attention mechanisms and ensemble learning contributes to enhanced robustness and generalization across different datasets. The novelty detection module effectively identifies and handles unseen scenes, further improving the model's reliability.

Real-time inference tests indicate that the optimized model can process scenes efficiently on resource-constrained drone platforms, meeting the operational requirements of various applications.





CONCLUSION

In conclusion, this study demonstrates the effectiveness of integrating transfer learning with advanced neural network architectures to enhance scene recognition capabilities in drones. The proposed methodology and configuration address key challenges in drone operations such as environmental variability, limited training data, and real-time processing constraints. By leveraging pre-trained models and incorporating techniques such as attention mechanisms, ensemble learning, and hybrid CNN-LSTM architectures, the proposed system delivers improved accuracy, adaptability, and operational efficiency. These enhancements contribute directly to the safety and reliability of autonomous drone missions, particularly in complex or dynamic environments.

Future work can focus on expanding the dataset diversity, incorporating 3D scene understanding, and enhancing model explainability to facilitate human oversight in critical applications such as disaster response or urban air mobility. Additionally, integrating these models with sensor fusion frameworks—combining visual data with LiDAR, infrared, or GPS—can further boost drone perception and decision-making capabilities. The research sets a foundation for developing next-generation intelligent drones that are not only reactive but also contextually aware and safe.

REFERENCES

1. Baumgartl, H., & Buettner, M. (2021). Mobile scene recognition with efficient and lightweight deep neural networks. *arXiv preprint arXiv:2107.11187*.
2. Thorram, R., Ghalehjegh, S., & Moghimi, M. (2024). A hybrid ensemble model for autonomous landing scene recognition. *Journal of Harbin Engineering University*, 45(2), 134-142.
3. Zhu, P., Jiang, H., Li, Z., Wang, L., & Tang, J. (2021). TPH-YOLOv5: A Transformer Prediction Head-based YOLOv5 for object detection in drone-captured scenarios. *arXiv preprint arXiv:2108.11539.*
4. Howard, A. G., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
5. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*.
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
7. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
8. Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*.
9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
10. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*.
12. Szegedy, C., et al. (2015). Going deeper with convolutions. *CVPR*.
13. Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. *CVPR*.
14. Huang, G., Liu, Z., Maaten, L. V. D., & Weinberger, K. Q. (2017). Densely

connected convolutional networks. *CVPR*.

15. Shi, X., et al. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *NeurIPS*.
16. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *CVPR*.
17. Woo, S., et al. (2018). CBAM: Convolutional block attention module. *ECCV*.
18. Lin, T. Y., et al. (2017). Feature pyramid networks for object detection. *CVPR*.
19. Russakovsky, O., et al. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*.
20. Jiang, H., Li, D., & Zhu, P. (2022). Real-time drone-based object detection using compact convolutional neural networks. *Sensors*, *22*(4), 1567.