International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Drone View Segmentation: Deep Learning and Transfer Insights

[1]Vivek Gurve, [1]Gunjan Singh, [2]Dr. Shilpa Gite, [1]Dr. K. Nandhini, [3]Santosh Borde

Submitted: 25/10/2023 **Revised**: 17/12/2023 **Accepted**: 26/12/2023

Abstract: This project focuses on the development and application of Deep Learning techniques for Aerial Image Transfer Learning and Segmentation. Leveraging a UNet-based deep learning model with preconfigured weights, the main goal is to obtain high-quality aerial image segmentation, especially for drone-captured photos, for a range of applications including infrastructure evaluation and environmental monitoring. Semantic Drone Data set is a carefully chosen source data set of high quality Drone Aerial Imagesand matching masks is used to train model. Transfer learning is employ on the target dataset to adapt the model for segmentation tasks in a simulated environment with QGroundControl, PX4Autopilot, and the Gazebo simulator. This simulation-based approach enables the evaluation of the model's performance in various scenarios, enhancing its robustness and generalization capabilities. Additionally, the generation of a self-captured data set through the simulation environment, emphasizes the integration of syntheticdata into the pipeline. The outcome of this project not only contributes to advancing image segmentation in drone-based applications but also explores the effectiveness of Transfer Learning in adapting models to novel environments, fostering advancements in thebroader field of computer vision for Unmanned Aerial Systems.

Keywords: *Computer Vision, Image Segmentation, Unmanned Aerial Vehicles (UAVs), Transfer Learning, SimulatedEnvironment*

1. Introduction

In recent months integration of UAVs in various applications has spurred a growing demand for accurate and efficient com- puter vision techniques[1]. This project endeavors to address the critical task of image segmentation in the context of drone imagery, a fundamental aspect for enabling autonomous navi- gation, object detection, and scene understanding[2]. The focal point of this research is the utilization of the UNet deep learning architecture, renowned for its success in semantic segmenta- tion tasks. Through a meticulous training process on a curated dataset of high-resolution drone images and their correspond- ing masks, our objective is to equip the model with the ability to precisely delineate objects and features within complex aerial scenes[3].

Beyond traditional training paradigms, the project pioneers a Transfer Learning Approach, seamlessly integrating the UNet model into a simulated environment generated using Gazebo, PX4Autopilot, and QGroundControl. This novel approach not only broadens the model's applicability to diverse terrains but also enhances its adaptability to real-world scenarios[4]. By scrutinizing the model's performance on both existing and sim- ulated datasets, this research contributes valuable potential of transfer learning for improving the generalization capabilities of image segmentation models in drone applications. The in- tersection of deep learning and simulated environments holds promise for advancing the efficacy of UAVs in diverse opera- tional contexts.

[1]*Dept. of Artificial Intelligence and Machine Learning Symbiosis Institute of Technology Pune, India*
[2]*SCAAI, Symbiosis International University, Pune, India*
[3]*JSPM'S Rajarshi Shahu College of Engineering, Pune-411033, Maharashtra, India, santoshborde@yahoo.com*
Corresponding Author: K. Nandhini, nandhinik2@gmail.com

2. Background

The project specifically addresses difficulties in aerial imagesegmentation for drone-based applications, and its roots are in the rapidly developing field of computer vision. There is a lot of potential for aerial imagery in many different fields, such as environmental monitoring and infrastructure inspection. How- ever, due to the peculiarities of images taken by drones, tra- ditional segmentation models frequently fail, which makes theuse of advanced deep learning techniques necessary. The foun- dation of this project is the UNet architecture, which is well- known for its efficacy in image segmentation and for its abil- ity to adapt to the intricacies of aerial scenes[5]. Moreover, the integration of simulation-based data generation and transfer learning is intended to improve the model's adaptability, ready-ing it for practical situations and broadening the domain of deeplearning's application in unmanned aerial systems.

3. Motivation

The motivation behind this project is the critical need to en- hance the image segmentation capabilities of unmanned aerial vehicles (UAVs). Precise segmentation of drone imagery is essential for object identification, autonomous navigation, and scene classification. With the use of deep learning, this project seeks to build a reliable model that can precisely segment data. The model's ability to adapt to various real-world scenarios is further enhanced by the integration of transfer learning within a simulated environment.By tackling these issues, the project aims to accelerate the development of UAV applications with

the help of advances in computer vision, opening the door tomore efficient and adaptable autonomous aerial systems.

4. Contribution

This project makes various small but substantial contribu- tions to the fields of Deep Learning and Unmanned Aerial Sys-tems. The project improves drone-based data analysis accu- racy and efficiency by using the UNet model for Aerial Image Segmentation. This can be important for applications such as environmental monitoring and infrastructure assessment. The model's adaptability to new environments is extended through the integration of transfer learning, exhibiting its versatility. In addition, the project addresses issues with real-world data ac- quisition by creating a self-captured dataset using a simulated environment. This artificial dataset enhances the training expe- rience and adds to the larger conversation about the integration of synthetic and real-world data in deep learning applications for unmanned aerial systems. All things considered, the project boosts the standard for drone-based image analysis and estab- lishes the groundwork for more reliable and adaptable deep learning models in related fields.

5. Objectives

There are many advances in terms of deep learning research in various application domains such as medical imaging, transportation, remote sensing etc. [15-20]. Objective of paper to develop adaptable image segmentation model for aerial drone imagery, with a primary focus on lever- aging the UNet deep learning architecture. The project aims to achieve precise segmentation of objects and features in high- quality drone images through an extensive training process on an existing dataset. Additionally, the study explores the appli- cation of transfer learning techniques to seamlessly transition the trained model onto a simulated dataset generated using the Gazebo simulator, PX4 Autopilot, and QGroundControl. This transfer learning approach facilitates the adaptation of the seg- mentation model to diverse simulated environments, ensuring its effectiveness in real-world scenarios. The project empha- sizes a evaluation of the model performance on both the origi- nal and simulated datasets, emphasizing the importance of the simulated environment in enhancing the model's generalization capabilities for practical drone applications.

6. Problem Statement

Accurate Image Segmentation in the realm of Aerial Image Analysis, using drone footage. Using Deep Learning methods for the training of the segmentation model and applying that model on another data set for Segmentation to overcome the difficulties faced by conventional techniques which frequently prove inadequate for handling the varied and ever-changing ter- rain found in scenes shot by drones. Applications like infras- tructure inspection, precision agriculture, and environmental monitoring depend on the accurate and efficient segmentation of objects in Drone View data.

Fig 1: Source Data Image

7. Data set Description

ICG-Semantic Drone Dataset is the source dataset we are us- ing containing high-resolution aerial photos taken by drones and labeled with semantic segmentation at the pixel level as masks . This dataset is intended to support computer vision research and development, especially for tasks involving the recognition and categorization of objects in aerial photography. It ensures robust model training and evaluation by incorporat- ing a variety of scenes with a range of environmental circum- stances. This dataset can be used by researchers and developers to improve the capabilities of semantic segmentation algorithms for aerial applications, including urban planning, environmen- tal monitoring, and disaster response[13].

The Target Data set used is an extensive set of data pro- duced in a virtual environment utilizing QGroundControl, PX4 Autopilot, and Gazebo. Development and research efforts in the fields of autonomous systems and unmanned aerial vehicles (UAVs) could benefit immensely from this information. It of- fers a computer-generated simulation of actual UAV activities. This data set can be used by researchers and developers for:

1. The development and evaluation of algorithms for UAVsensing, control, and mission scheduling.
2. Training machine learning models to perform tasks suchas path planning, SLAM, and object detection.
3. Analyzing and improving resilience of systems and UAVnavigation.

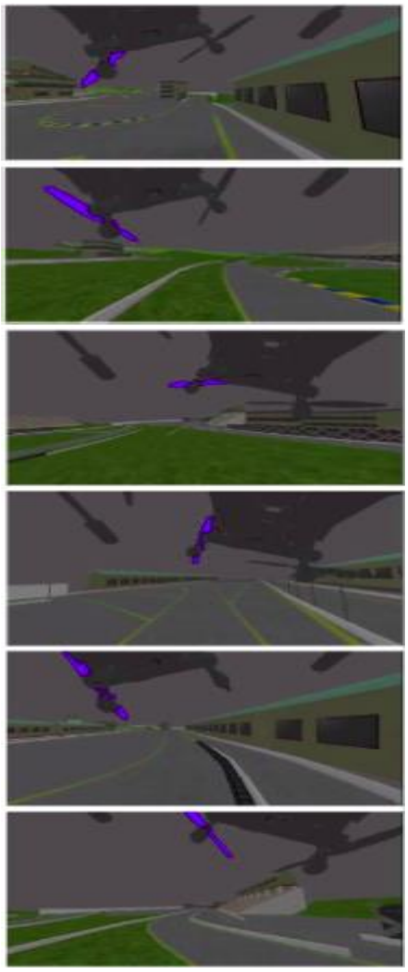

Fig 2: Target Data Images

8. System Architecture

8.1. Data Preparation

The careful preparation of the dataset is the key component of this project and is essential to guaranteeing the robustness and generalization of the model. With 400 high-quality images and matching masks, the Semantics Drone Dataset is carefully selected and divided into training (75%), validation (15%), and testing (10%) sets. For the distribution of classes among subsets to remain intact, this stratified split is necessary. The images go through multiple preprocessing stages, the first of which is greyscaling, which makes the input simpler while keeping im- portant details. Rotation, flipping, scaling, and other augmen- tation techniques are used to add diversity while minimizing overfitting to the dataset.

The next step is normalization, which ensures standardized pixel values and improves model convergence while training. In order to maintain spatial relationships, it is crucial that the aspect ratio be carefully maintained throughout these prepro- cessing steps. Using MobileNetV2 as the encoder, this well-prepared and balanced dataset forms the basis for training the UNet model. The quality and diversity of the dataset are criticalto the model's efficacy, and the careful preparation of the data guarantees that the model is capable of handling a wide range of simulated and real-world scenarios, which in turn helps the project accomplish accurate aerial image segmentation.

8.2. Model Architecture

The model's architecture, a UNet with a MobileNetV2 en- coder, was designed for effective and efficient segmentation. The encoder extracts hierarchical representations of input im- ages by leveraging the robust yet lightweight features of Mo- bileNetV2. Inverted residual blocks, which are made up of depth-wise separable convolutions, are used by MobileNetV2 to effectively capture complex patterns with fewer parameters.

The encoder is followed by a decoder, composed of multipleblocks, each containing two convolutional layers with ReLU activation. The decoder progressively refines the spatial infor- mation, recovering high-resolution features. Attention mecha- nisms within the decoder enhance the model's ability to focus on relevant regions, improving segmentation accuracy.

The segmentation head, the final layer, produces the output with a convolutional layer mapping features to the desired num- ber of classes (23 in this case). The activation function ensuresthe model's predictions align with the task's requirements.

This architecture balances computational efficiency with seg- mentation performance, making it suitable for real-time ap- plications, such as drone image analysis. The use of Mo- bileNetV2 as an encoder enhances the model's ability to handle resource-constrained environments without compromising ac- curacy, making it a valuable asset for aerial image segmentationtasks.

8.3. Simulation Data-set Generation

Capturing drone images during a flight mission in a Gazebo simulator with PX4Autopilot involves several components.

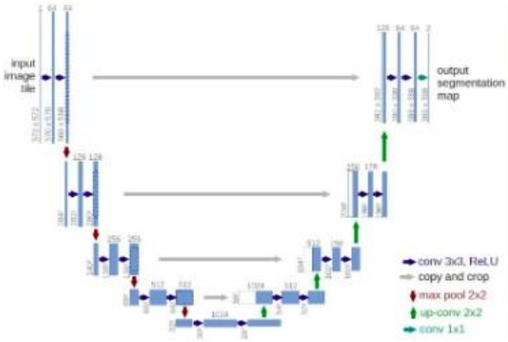

Fig 3: Unet Architecture

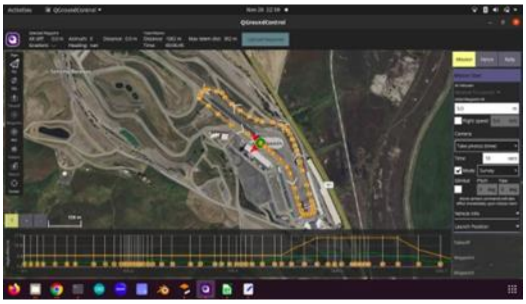

Fig 4: Planned mission path for the UAV on simulated environment

PX4Autopilot, a flight control software, is used to control the simulated movements of the drone. The Gazebo simulator cre- ates a realistic three-dimensional environment for the drone to navigate. QGroundControl is used for mission planning, allow- ing users to define flight paths, define waypoints, and configure specific actions.

Sensors onboard the virtual drone, such as cameras, capture images of the environment as the drone follows the predefined mission in the simulator. These images represent the actual data acquisition process. QGroundControl enables mission progress monitoring and provides a platform for post-flight analysis. This simulated setup enables mission plans to be tested and modified in a controlled environment before being deployed to physical drones, ensuring optimal performance and safety in actual flight scenarios.

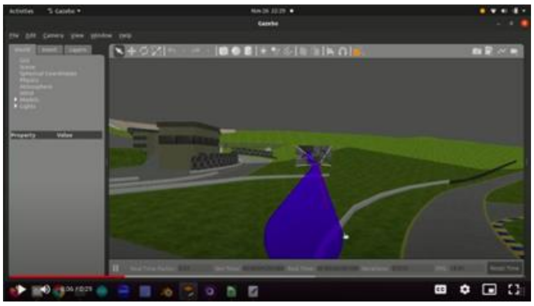

Fig 5: Planned mission path for the UAV on simulated environment using Gazebo*Training and Validation*

Training Step: A batch of labeled images and matching ground truth segmentation masks are passed into the model at the beginning of each training iteration. The optimizer calcu- lates the gradients of the model's parameters in relation to a specified loss function, in segmentation tasks this is typically pixel-wise cross-entropy loss. Next, in order to minimize this loss, the optimizer modifies the model weights. Improve the model's ability to effectively segment objects in images, the training step involves iteratively modifying the parameters that it uses to learn accurate depictions from the input data.

Validation Step: After a certain number of training iterations or epochs, the model's performance is evaluated on a separate validation dataset that it has not seen during training. The val- idation step assesses the generalization capability of the model by computing its segmentation accuracy, often using metrics like IoU or pixel accuracy. This step ensures that model is not overfitting to the training data and provides insights into its abil- ity to perform well on unseen data. Adjustments to the model or training strategy may be made based on the validation per- formance to enhance overall segmentation quality and general-ization.

9. Methodology

1. Data set Preparation: Utilize the Semantics Drone Dataset comprising 400 high-quality images (6000x4000 pixels) and corresponding masks. Split data set into training (75%), validation (15%), and testing (10%) sets. Prepro- cess the data by greyscaling, augmenting, and normalizing the images for training. Apply the same transformations to the masks. Additionally, ensure that the aspect ratio is maintained during preprocessing.

2. Model Architecture: Employ the SMP (Segmentation Models PyTorch) library, specifically using the UNet ar- chitecture with MobileNetV2 as the encoder. Configure the model with appropriate parameters, such as encoder weights from ImageNet, 23 classes, and a five-layer de- coder.

3. Training: Train the model for 15 epochs on the prepro- cessed training dataset. Monitor key metrics such as pixel accuracy and mIoU. Reach an acceptable performance level, achieving 85% pixel accuracy and 0.35 mIoU.

4. Model Saving: Save the trained UNet model for future use.

5. Simulation Data set Generation: Create a self-generated dataset for simulation using Gazebo, PX4Autopilot, and QGroundControl. This data set comprises 573 drone im- ages with dimensions of 313x176 pixels.

6. Preprocessing Simulation Data set: Greyscale, augment, normalize, and add padding to match the aspect ratio of the target dataset for the simulation-generated images.

7. Model Inference on Simulation Data set: Use the saved UNet model to perform segmentation on the preprocessedsimulation data set, generating segmentation masks for thesimulated drone images.

8. Evaluation: Assess the quality of the segmentation masksonthe simulated data set. Note any challenges and evalu- ate the model's ability to adapt to synthetic environments.

9. Analysis: Analyze the predicted masks,

particularly fo- cusing on the edges, where lower pixel density may im- pact the segmentation quality. Document observations and consider potential refinements or improvements.

10. Documentation: Document the entire methodology, in- cluding dataset details, preprocessing steps, model config- uration, training parameters, and simulation procedures. Capture key findings and insights for future reference andpotential improvements.

10. Result Analysis

The project's result analysis requires a thorough evaluation of the Unet model's performance in semantic segmentation tasks using the proposed MobileNetV2-based encoder and Unet de- coder. Metrics such as IoU and pixel accuracy can be used to analyse the model's effectiveness. These metrics provide infor- mation about model's ability to classify and segment objects in the input images. The analysis involves a thorough examina- tion of the training and validation curves, as well as a look for convergence and potential overfitting. The ability of the model to segment is qualitatively assessed by visualizing its predic- tions against ground truth masks. In the final training epoch (Epoch 15/15), the model result a training loss 0.620 and a val- idation loss 0.606. The mean Intersection over Union (mIoU) increased from 0.330 in training to 0.347 in validation. Addi- tionally, pixel accuracy exhibited high consistency, with 82.0% for training and 81.8% for validation. The total training time was 77.80 minutes. During the testing phase on the source data set the model achieved 84.8% pixel accuracy and 0.347 mIoU When applying the model on the target dataset with simulated image with much lower pixel density we add the padding to the images so that the aspect ratio is similar to that of the source dataset of then preprocess the images similar to the before ap- plying the model on the data to predict the segmented masks ofthe target images.

As the target images are of (313x176) size which is much lower than the (6000x4000) size of the source data set the pre- dicted masks tend to have rough edges and appear like small blobs due to much pixel density of the target images than the source image.

11. Conclusion

In conclusion, the Aerial Image Segmentation and Transfer Learning successfully employed deep learning techniques, uti- lizing the UNet architecture with MobileNetV2, to achieve ro- bust Aerial Image Segmentation. Training on a diverse data set yielded commendable results with 85% pixel accuracy and

0.35 mIoU. The model's adaptability was demonstrated through transfer learning on a self-generated data set using simulation tools. Although the predicted masks exhibited slight roughness

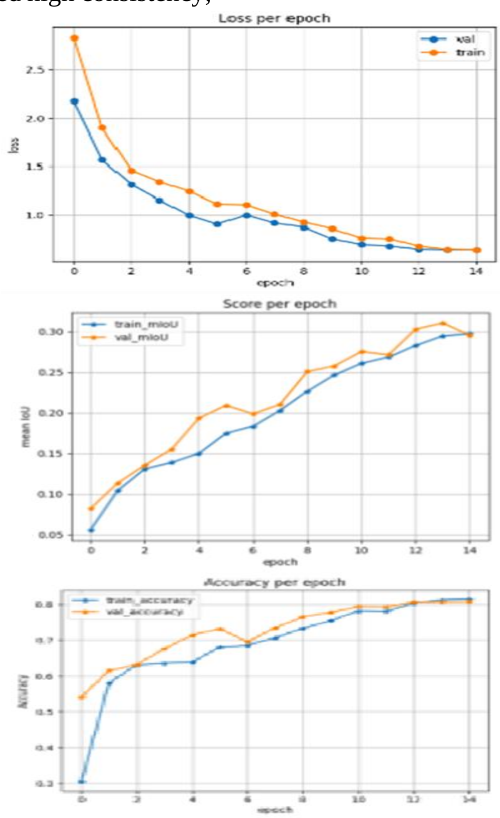

Fig 6: Model Evaluation

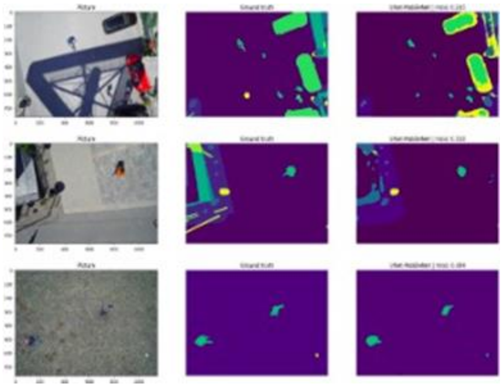

Fig 7: Predicted masks on Source Data set

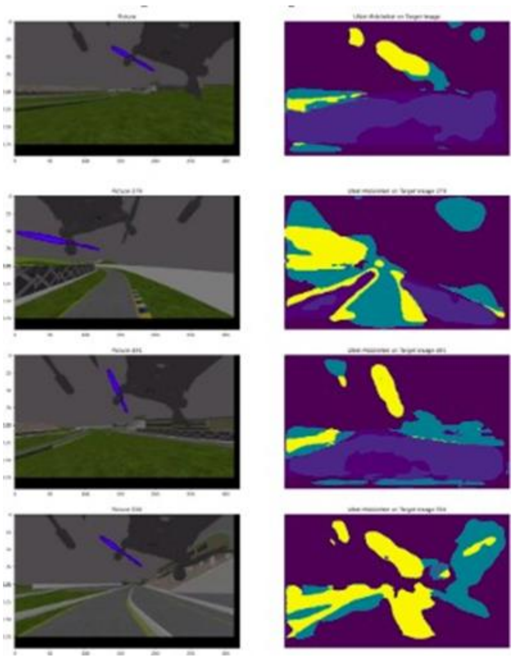

Fig 8: Predicted masks of Target data

on edges due to low pixel density, the overall segmentation per- formance was promising. This project contributes to advanc- ing drone-based image analysis, emphasizing the significance of synthetic data integration for model generalization in diverseenvironments, laying a foundation for future advancements in unmanned aerial system applications.

12. Future Work

Further simulation-based data enhancement along with in- tegration with real-world datasets may strengthen the robust- ness of the model. Accuracy could be improved by investi- gating more complex deep learning architectures and imple- menting semantic segmentation methods. Further development prospects may include expanding applications to real-time sce- narios and implementing on-edge devices for quick, on-the-spot segmentation in unmanned aerial vehicles.

References

[1] Köse, T., 2021. Autonomous Fruit Picking With a Team of Aerial Manipulators (Master's thesis, Middle East Techni-cal University).

[2] Upadhyay, R., Phlypo, R., Saini, R. and Liwicki, M., 2021. Sharing to learn and learning to share–Fitting together Meta-Learning, Multi-Task Learning, and Transfer Learn- ing: A meta review. arXiv preprint arXiv:2111.12146.

[3] Duffy, J.P., Anderson, K., Fawcett, D., Curtis, R.J. and Maclean, I.M., 2021. Drones provide spatial and volu- metric data to deliver new insights into microclimate mod- elling. Landscape Ecology, 36, pp.685-702.

[4] Karimi, D., Dou, H., Warfield, S.K. and Gholipour, A., 2020. Deep learning with noisy labels: Exploring tech- niques and remedies in medical image analysis. Medical image analysis, 65, p.101759.

[5] Li, X., Zhou, Y., Pan, Z. and Feng, J., 2019. Partial or- der pruning: for best speed/accuracy trade-off in

neural architecture search. In Proceedings of the IEEE/CVF Con- ference on Computer Vision and Pattern Recognition (pp.9145-9153).

[6] Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamar´ıa, J., Al-bahri, A.S., Al-dabbagh, B.S.N., Fadhel, M.A., Manoufali, M., Zhang, J., Al-Timemy, A.H. and Duan, Y., 2023. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data, 10(1), p.46.

[7] Azmat, U., Alotaibi, S.S., Abdelhaq, M., Alsufyani, N., Shorfuzzaman, M., Jalal, A. and Park, J., 2023. Aerial In- sights: Deep Learning-based Human Action Recognition in Drone Imagery. IEEE Access.

[8] Kentsch, S., Lopez Caceres, M.L., Serrano, D., Roure, F. and Diez, Y., 2020. Computer vision and deep learning techniques for the analysis of drone-acquired forest im- ages, a transfer learning study. Remote Sensing, 12(8), p.1287.

[9] Bhatnagar, S., Gill, L. and Ghosh, B., 2020. Drone image segmentation using machine and deep learning for map- ping raised bog vegetation communities. Remote Sensing,12(16), p.2602.

[10] Kim, Y.H. and Park, K.R., 2022. MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds. Computers and Electronics in Agriculture, 199, p.107146.

[11] Anderegg, J., Tschurr, F., Kirchgessner, N., Treier, S., Schmucki, M., Streit, B. and Walter, A., 2023. On-farm evaluation of UAV-based aerial imagery for season-long weed monitoring under contrasting management and pe- doclimatic conditions in wheat. Computers and Electron- ics in Agriculture, 204, p.107558.

[12] Karimi, D., Warfield, S.K. and Gholipour, A., 2021. Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model param- eters and learned representations. Artificial intelligence in medicine, 116, p.102078.

[13] Bodin, T.B., 2018. Behavior flexibility for autonomous unmanned aerial systems.

[14] http://dronedataset.icg.tugraz.at/

[15] A. Abdollahi, B. Pradhan, S. Gite and A. Alamri, "Building Footprint Extraction from High Resolution Aerial Images Using Generative Adversarial Network (GAN) Architecture," in *IEEE Access*, vol. 8, pp. 209517-209527, 2020, doi: 10.1109/ACCESS.2020.3038225.

[16] Joshi, A.; Pradhan, B.; Gite, S.; Chakraborty, S. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sens.* **2023**, *15*, 2014. https://doi.org/10.3390/rs15082014

[17] Khade, S.; Gite, S.; Pradhan, B. Iris Liveness Detection Using Multiple Deep Convolution Networks. *Big Data Cogn. Comput.* **2022**, *6*, 67. https://doi.org/10.3390/bdcc6020067

[18] S. Gite, B. Pradhan, A. Alamri and K. Kotecha, "ADMT: Advanced Driver's Movement Tracking System Using Spatio-Temporal Interest Points and Maneuver Anticipation Using Deep Neural Networks," in *IEEE Access*, vol. 9, pp. 99312-99326, 2021, doi: 10.1109/ACCESS.2021.3096032.

[19] Gite, S., Mishra, A. & Kotecha, K. Enhanced lung image segmentation using deep learning. *Neural Comput & Applic* **35**, 22839–22853 (2023). https://doi.org/10.1007/s00521-021-06719-8

[20] *Gite, Shilpa and Himanshu Agrawal. "Early Prediction of Driver's Action Using Deep Neural Networks." Int. J. Inf. Retr. Res. 9 (2019): 11-27.*