



# **Semantic Segmentation of Aircraft and Ships from Satellite Images**

## **Members:**

Harshit Ranjan (BT22CSA029)

Harsh Suhan (BT22CSA034)

Sharit Vaishnav (BT22CSA042)

Prajwal Prasad (BT22CSA043)

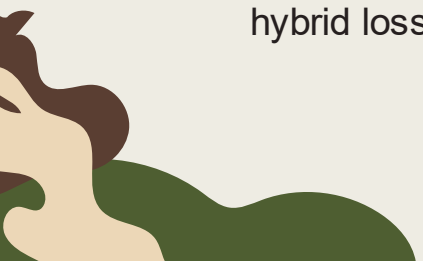
Assigned by :  
Dr. Khushboo Thakkar Jain

Under the mentorship of: **Dr. Prerna Mishra**

# Problem Statement



## **Semantic Segmentation of Aircraft and Ships in Satellite Imagery**

- Highlight the need for accurate semantic segmentation in defense, surveillance, and maritime monitoring.
  - Emphasize shortcomings of traditional models in handling complex backgrounds, small objects, and limited data.
  - Traditional CNNs with fixed receptive fields struggle to adapt to variable object shapes, scales, and orientations in aerial imagery, especially for small or irregular targets like ships and occluded aircraft.
  - Semantic segmentation in real-world satellite imagery often suffers from class imbalance and limited labeled data, requiring advanced techniques such as deformable convolutions and hybrid loss functions to ensure robust generalization.
- 

# Table of Contents



**01**

**Introduction  
and Aim**

**02**

**Literature Review**

**03**

**Methodology**

**04**

**Results and  
Conclusion**

# INTRODUCTION

Semantic segmentation of aerial and satellite imagery plays a crucial role in defense surveillance, coastal monitoring, and disaster response. However, challenges such as **class imbalance**, **small object detection**, **complex backgrounds**, and **limited labeled data** significantly hinder model performance in real-world scenarios.

- ❖ **Robust Semantic Segmentation:** Aimed at accurately segmenting aircraft and ships in satellite imagery, addressing challenges like small object detection, occlusion, and complex backgrounds.
- ❖ **Advanced Model Architecture:** Utilizes **Attention U-Net enhanced with Deformable Convolutions** for adaptive feature extraction and improved spatial sensitivity to object shapes.
- ❖ **Improved Data Handling:** Incorporates **RoboFlow-based annotation**, **data augmentation**, and a **hybrid loss function (Crossentropy + Dice Loss)** to boost accuracy and generalization in real-world scenarios.



Design a robust Attention U-Net architecture integrated with **Deformable Convolution Layers** to accurately segment aircraft and ships in complex aerial imagery.

## AIMS



Utilize **hybrid loss functions** (Sparse Categorical Crossentropy + Dice Loss) and **data augmentation** to enhance model performance on small and irregular targets.



Leverage **RoboFlow** for streamlined data labeling and build a scalable, real-time-ready system suitable for deployment in **military and surveillance applications**.



# **Literature Review**

# Semantic Segmentation with Contextual Features and using U-Net CNN Model

- Zhang, Y., Yin, J., Gu, Y., & Chen, Y. (2024). Multi-level Feature Attention Network for medical image segmentation. Expert Systems with Applications, 125785.
- Yin, X. X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). [Retracted] U-Net-Based Medical Image Segmentation. Journal of healthcare engineering, 2022(1), 4189781.

The study introduces the **Multi-level Feature Attention Network (MFAN)**, which enhances medical image segmentation by addressing the limitations of U-Net-based architectures. MFAN leverages a hierarchical Swin Transformer for global context modeling, improving its ability to handle complex images.

Key components include the Cross-connection Multi-level Attention (CMA) module, which refines feature representation for accurate localization, and the Pyramid Collaborative Attention (PCA) module, which captures multi-scale semantic features to enhance segmentation performance. These advancements achieve state-of-the-art results on datasets like ACDC and ISIC2017, demonstrating improved precision, robustness, and generalization with fewer parameters.

The study explores the evolution and impact of U-Net-based architectures in medical image segmentation. Introduced in 2015, the original U-Net gained prominence for its encoder-decoder structure with skip connections, enabling accurate segmentation even with limited data. Its adaptability made it a cornerstone for tasks like multi-organ segmentation and lesion detection. Advancements in U-Net variants, such as Attention U-Net, refine encoder-decoder interactions using attention gates, while UNet++ introduces nested skip connections and deep supervision to enhance segmentation accuracy for complex, multi-scale images. **TransUNet leverages transformers** to model global context, addressing U-Net's challenges in handling long-range dependencies while retaining precision in fine details.

# CNN Using Deformable Convolution

- Xin Zhang ,Yingze Song ,Tingting Song ,Degang Yang , Yichen Ye (2024) LDConv: Linear deformable convolution for improving convolutional neural networks
- Feng Chen, Fei Wu, Jing Xu, Guangwei Gao, Qi Ge, Xiao-Yuan Jing (2020) Adaptive deformable convolutional network

The paper introduces **Linear Deformable Convolution (LDConv)**, a novel convolutional operation designed to overcome the limitations of standard and deformable convolutions in CNNs. Unlike traditional fixed-size kernels, LDConv enables **arbitrary kernel shapes and parameter sizes**, allowing for flexible and efficient feature extraction tailored to object geometry. It addresses the issue of quadratic parameter growth in deformable convolutions by enabling **linear parameter scaling**, making it more hardware-friendly. LDConv also explores the effect of different initial sampling shapes on network performance. The paper demonstrates its effectiveness through object detection tasks on datasets like COCO2017, VOC, and VisDrone-DET2021.

This paper proposes **Adaptive Deformable ConvNet (A-DCN)**, an enhanced deformable convolutional network that improves geometric transformation modeling for semantic segmentation and object detection tasks. By introducing a refined **adaptive deformable convolution module**, the paper integrates spatial attention, channel attention, and their interdependency, enabling better focus on relevant image regions. Additionally, it explores optimal arrangements and combinations of deformable convolutions to maximize performance. Extensive experiments on benchmarks like **PASCAL VOC 2012**, **Cityscapes**, and **COCO** demonstrate that A-DCN achieves state-of-the-art results. The work provides both theoretical insights and practical enhancements for integrating deformable operations into deep CNN architectures.



# Evaluation Metric Used

- Hoeser, T., & Kuenzer, C. (2020). Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. Remote Sensing, 12(10), 1667.

Image segmentation in remote sensing involves pixel-level classification to create accurate masks. While CNNs provide high-level semantic information, their feature maps often lose spatial resolution as they deepen, posing challenges for precise pixel-wise predictions. Effective segmentation also depends on contextual relationships influenced by segment size, continuity, and density. This has led to the development of segmentation architectures focused on multi-scale context modeling and resolution preservation.

The PASCAL-VOC 2012 dataset is a standard benchmark for segmentation, using Mean Intersection over Union (**mIoU**) as the main evaluation metric. This metric assesses the overlap between ground truth and predicted segmentation masks for each class, providing an overall accuracy score. Encoder-decoder models like U-Net address the limitations of simple decoders with skip connections that preserve spatial resolution during upsampling. However, challenges remain, such as the slow and manual nature of data annotation, difficulty detecting small or occluded objects, and balancing speed with accuracy. Additionally, U-Net models struggle with long-range dependencies, and object detection models like **YOLO-v3** are not fully suited for aerial imagery. Many models also face issues with data imbalance and domain variability. Our work addresses these through tools like SAM for efficient annotation, tailored model fine-tuning, and advanced architectures that incorporate attention mechanisms, along with data augmentation and transfer learning, making the system more adaptable and robust for practical applications.



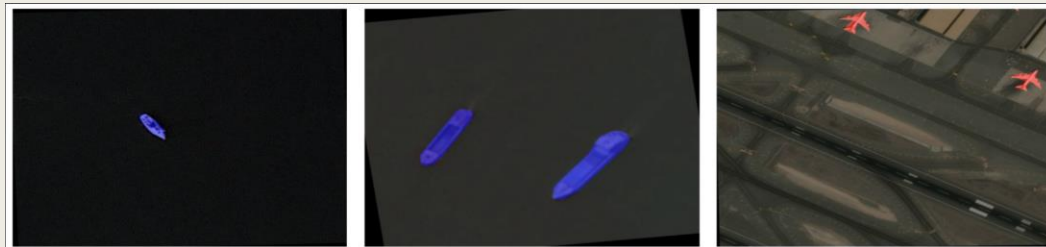
# **Work Done (Methodology)**

# DATA DESCRIPTION

- A custom dataset was created using **RoboFlow**, enabling accurate labeling of aircraft and ships in satellite images with polygon and bounding box annotations.
- Applied techniques like **rotation, flipping, scaling, and brightness adjustments** to increase dataset size and variability, improving model robustness and generalization.



(a) Using SAM Model



(b) Self Labeled Using Robo-Flow

Annotated Training Images

# Model Development



## **Dataset Creation and Annotation :**

- Custom dataset built using Roboflow
- Annotated using polygon masks for aircraft and ships
- Exported in COCO segmentation format
- Applied data augmentation (flip, rotate, brightness, scale)



## **Model Architecture:**

### Base Model Selection :

1. Started with U-Net architecture for initial segmentation.
2. Used for generating baseline masks to evaluate performance.

### Architecture Enhancement :

1. Upgraded to Attention U-Net to focus on important spatial regions
2. Integrated Deformable Convolutional Layers
  - Allows the model to adapt to irregular object shapes
  - Enhances performance in cluttered or noisy backgrounds

# Model Development



## Training and Evaluation

### Multi-Class Output:

Added support for multiple classes:→ Aircraft, Ships, Background  
Updated final softmax layer to handle multi-class pixel-wise prediction.

### Training and Fine-Tuning :

Unfroze last 10 layers for fine-tuning.

Trained on augmented dataset using Google Colab GPU. Validated performance using Dice coefficient and visual outputs



## Loss Function and Optimization

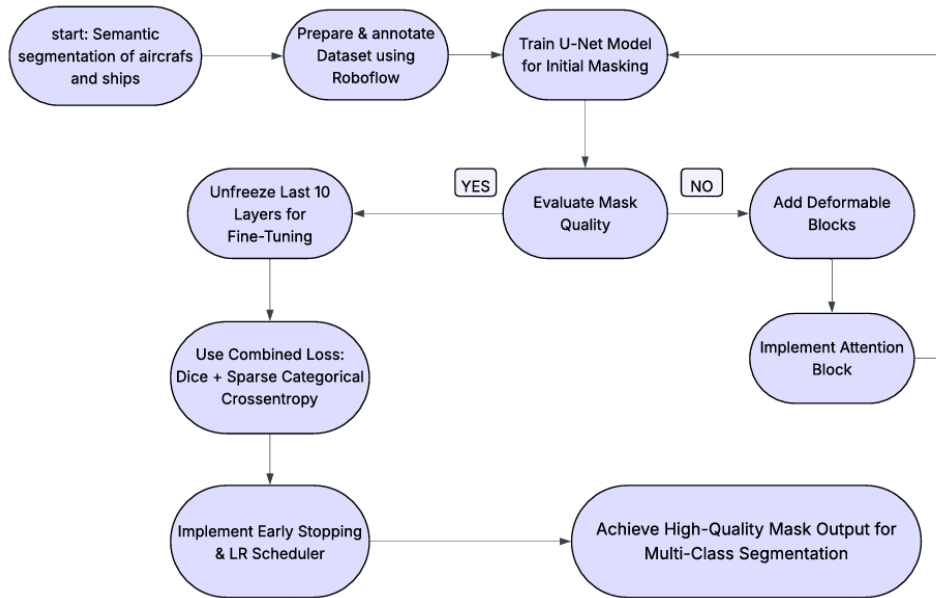
Used a combined loss:

- Sparse Categorical Crossentropy (for pixel-wise classification)
- Dice Loss (for region overlap accuracy)

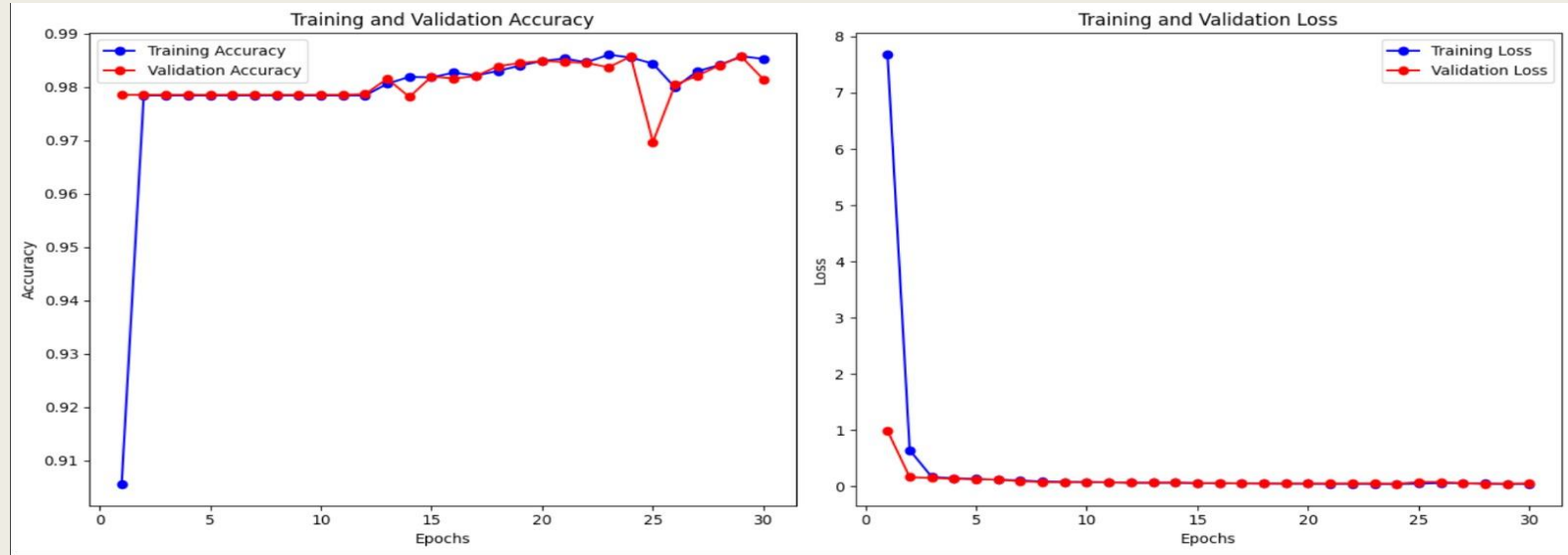
Optimizer: Adam

Added learning rate scheduler and early stopping

# Workflow Diagram

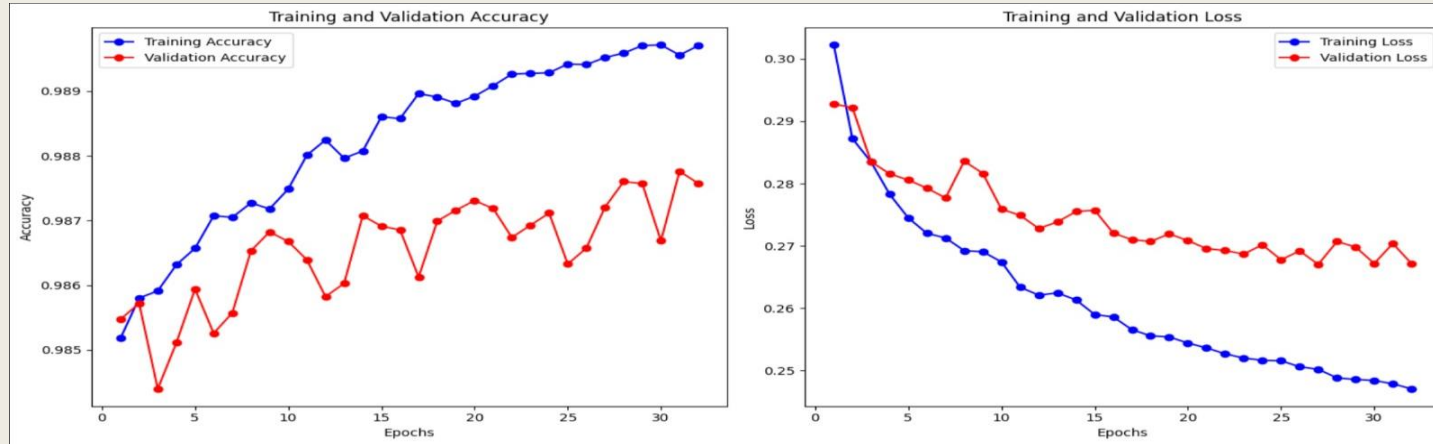


# 1. Results for Attention U-Net Model without Early Stopping



- ❑ Training Accuracy : 98.64%
- ❑ Validation Accuracy : 98.14%
- ❑ Training Loss : 0.168 to 0.0407
- ❑ No. of epochs : 30
- ❑ Validation Loss decreased by : 0.0513

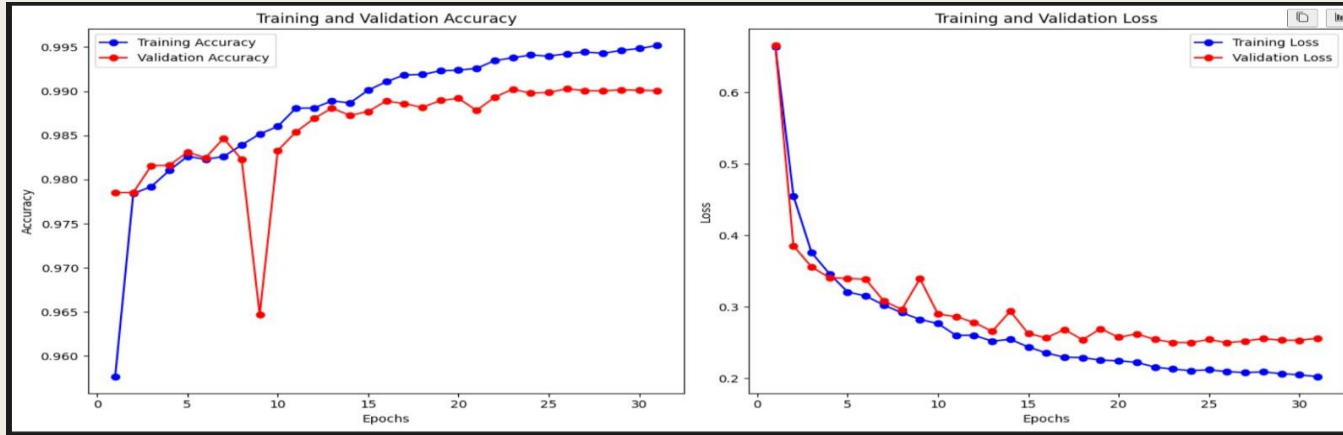
## 2. Results for Attention U-Net Model with Early Stopping



- ❑ Training Accuracy : 99.07%
- ❑ Validation Accuracy : 98.88%
- ❑ Training Loss : 0.31 to 0.237
- ❑ Validation Loss: 0.291 to 0.27
- ❑ No. of epochs : 31/50

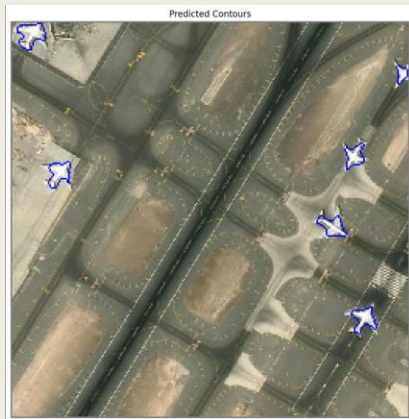


### 3. Results for Attention U-Net Model with Deformable Layer

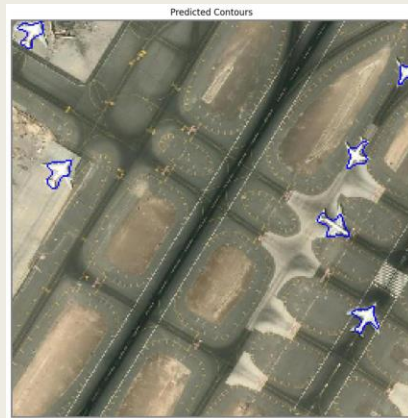


- ❑ Training Accuracy : 99.44%
- ❑ Validation Accuracy : 99.00%
- ❑ Training Loss : 0.77 to 0.2087
- ❑ Validation Loss:0.2529
- ❑ No. of epochs : 31

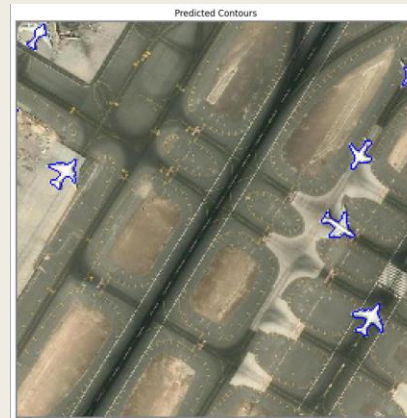
## Final Result and Visualization (Aircrafts)



Attention U-Net

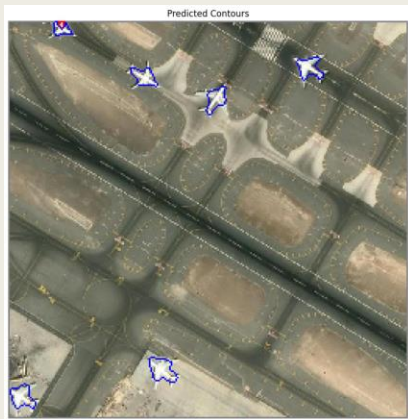


Attention U-Net with Early  
Stopping

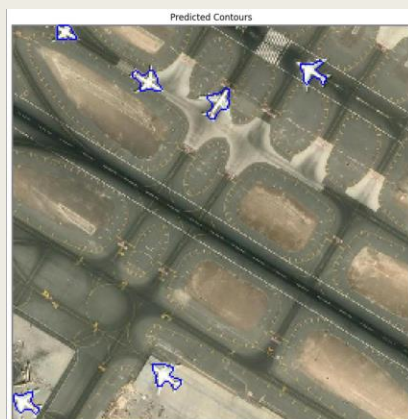


Attention U-Net with  
Deformable Layer

## Final Result and Visualization (Aircrafts)



Attention U-Net



Attention U-Net with Early  
Stopping



Attention U-Net with  
Deformable Layer

## Final Result and Visualization (Aircrafts)



Attention U-Net

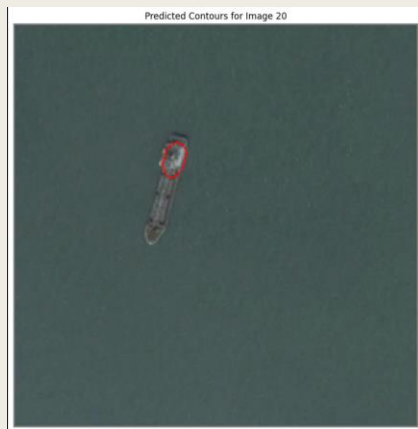


Attention U-Net with Early  
Stopping

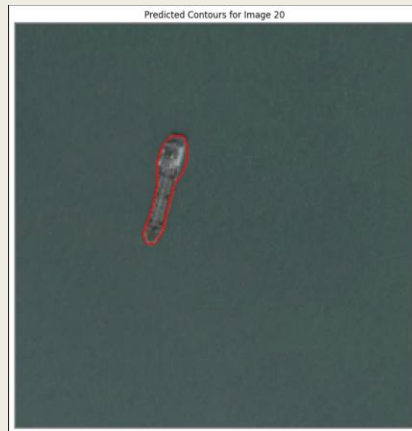


Attention U-Net with  
Deformable Layer

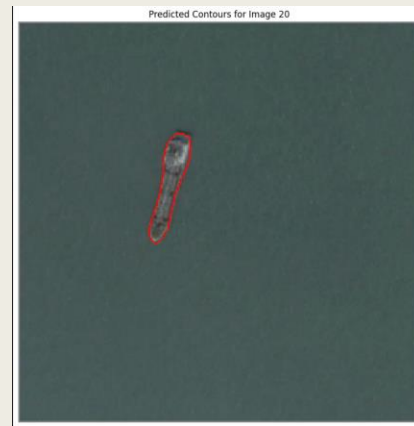
## Final Result and Visualization (Ships)



Attention U-Net



Attention U-Net with Early  
Stopping



Attention U-Net with  
Deformable Layer



# Future work

## **Integration with Real-Time Systems**

- Optimize the model for real-time deployment using techniques like model pruning, quantization, and hardware acceleration (GPUs/TPUs).

## **Expansion to Multimodal Data**

- Incorporate thermal or radar images to improve robustness and accuracy under diverse conditions (e.g., low light, heavy cloud cover).

## **Transfer Learning for Specific Domains**

- Fine-tune the model for related tasks, such as detecting drones, vehicles, or ships in military scenarios or monitoring environmental changes during disasters.



# CONCLUSION

- Custom Dataset Creation using Roboflow enabled accurate, multi-class pixel-wise annotation, effectively addressing the lack of publicly available satellite imagery datasets.
- The enhanced Attention U-Net architecture with Deformable Convolutional Layers significantly improved segmentation quality, particularly in complex and cluttered backgrounds. Attention gates in the model prioritized relevant features, improving segmentation accuracy in complex backgrounds.
- Using a combined loss function (Dice + Sparse Categorical Cross entropy) led to better handling of class imbalance and improved region-wise segmentation accuracy.
- The final model achieved ~99.00% validation accuracy and outstanding Dice scores, demonstrating strong performance for real-world multi-class segmentation of aerial objects.

# Thank You

