DOT/FAA/TC-24/29

Federal Aviation Administration
William J. Hughes Technical Center
Aviation Research Division
Atlantic City International Airport
New Jersey 08405

Small Unmanned Aircraft-Based Foreign Object Debris Detection on Airport Surfaces

September 2024

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-24/29		

4. Title and Subtitle	5. Report Date
SMALL UNMANNED AIRCRAFT-BASED FOREIGN OBJECT DEBRIS DETECTION ON AIRPORT SURFACES	September 2024
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Logan Shaffer*, David Hall*, Sheldon Menezes*, Paul Rybski**, and Rorry Brenner**	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Department of Transportation Federal Aviation Administration William J. Hughes Technical Center Airport Technology R&D Branch Atlantic City International Airport, NJ 08405 Woolpert, Inc.* 600 Aviation Research Blvd Egg Harbor Township, NJ 08234 Neya Systems** Applied Research Associates 555 Keystone Drive Warrendale, PA 15086	
	11. Contract or Grant No. 692M15-20-T-00034

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Department of Transportation Federal Aviation Administration UAS R&D Portfolio Branch (ANG-C21) 800 Independence Ave. Washington, DC 20591	Final Report
	14. Sponsoring Agency Code ANG-C21 and AAS-100

15. Supplementary Notes

The Federal Aviation Administration (FAA) Airport Technology Research and Development (ATR) Branch Technical Monitor was Garrison Canter. The Federal Aviation Administration Aviation Research Division COR was Michael DiPilato.

16. Abstract

Foreign Object Debris (FOD) can create significant safety implications for aircraft and personnel and is a continuous concern in the airport environment. The FAA Airport Technology Research and Development Branch conducted a research effort to explore the feasibility and maturity of using commercially available small unmanned aircraft systems (sUAS) and artificial intelligence/machine learning (AI/ML) algorithms to detect FOD on airport surfaces. The objectives of this research effort were to develop a novel, proof-of-concept sUAS-based FOD detection workflow using AI/ML algorithms and to assess the workflow to determine whether it is capable of meeting all, some, or none of the requirements in FAA Advisory Circular (AC) 150/5220-24, *Airport Foreign Object Debris Detection Equipment*.

The research team conducted initial testing of this workflow at Cape May County Airport and validation testing at Atlantic City International Airport. The sUAS-based FOD detection workflow, which used the FastFlow ML deep learning algorithm, was capable of meeting some of the AC 150/5220-24 requirements, including achieving a 96% detection rate for FOD items specified in the AC. However, further research and development will be needed for this technology to meet the full set of AC 150/5220-24 requirements, including reducing the false positive rate, reducing the data processing time, implementing a software interface for displaying and recording FOD detection alerts, and detecting FOD in low-light and inclement weather conditions.

17. Key Words	18. Distribution Statement
Small Unmanned Aircraft System (sUAS), FOD detection, Machine learning, Artificial intelligence	This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	95	

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AC	Advisory Circular
ACY	Atlantic City International Airport
AGL	Above ground level
AI	Artificial intelligence
AOA	Airport Operations Area
ATR	Airport Technology Research and Development Branch
AUC	Area under the curve
CAD	Computer-aided design
CFR	Code of Federal Regulations
CPU	Central processing unit
CRF	Conditional random field
DRBA	Delaware River & Bay Authority
EO	Electro-optical
FAA	Federal Aviation Administration
FBO	Fixed-based operator
FOD	Foreign object debris
GNSS	Global navigation satellite system
GPS	Global Positioning System
GPU	Graphics processing unit
GSD	Ground sampling distance
ML	Machine learning
NCAT	NGS Coordination Conversion and Transformation Tool
NGS	National Geodetic Survey
NOTAM	Notice to Air Mission
OCR-GAN	Omni-frequency channel-selection reconstruction- Generative adversarial network
PVC	Polyvinyl chloride
RAM	Random-access memory
RGB	Red, green, blue
ROC	Receiver operating characteristic
RTK	Real-time kinematic
SD	Standard deviation
SME	Subject matter expert
sUAS	Small unmanned aircraft system
WHJTC	William J. Hughes Technical Center
WWD	Cape May County Airport

TERMINOLOGY

Airframe. The airlifting component of a small Unmanned Aircraft System (sUAS); does not include the sensor.

Area Under the Curve. Area under the curve (AUC) is a standard metric to quantitatively determine the performance of a binary classifier system.

Artificial Intelligence. A branch of computer science seeking to simulate intelligent behavior.

Binary Classifier. A supervised learning algorithm that categorizes observations into one of two classes.

Data Labeling. The process by which elements in a dataset are manually labeled to indicate which category they belong to. This information is used by supervised learning algorithms as part of their training process and used by all machine learning algorithms in order to analyze their performance.

Deep Learning. A Machine Learning (ML) system where an architecture of artificial neurons (simulated in computer software or hardware) are constructed in layers. The depth of the network is based on the number of neuron layers.

False Negative. In terms of the output of the foreign object debris (FOD) detection algorithm, a “Negative” result is an indication that FOD has NOT been detected. A "False Negative” is a negative result that is incorrect. In this case, there is FOD present, but the algorithm did not detect it.

False Positive. In terms of the output of the FOD detection algorithm, a “Positive" result is an indication that FOD has been detected. A "False Positive” is a positive result that is incorrect. In this case, the algorithm detected FOD where there was none.

Foreign Object Debris. Any object located in an inappropriate location in the airport environment that has the capacity to injure airport or airline personnel and damage aircraft.

Ground Sampling Distance. The distance, as measured along the ground, between adjacent pixels in an image. Ground Sampling Distance governs how much information can be inferred about features from image measurements. The Ground Sampling Distance and image resolution determine the size of the footprint of the image captured by the sensor.

Machine Learning. A field in computer science devoted to creating methods where software can learn from data to improve performance on a specified task.

Nadir. Nadir is the angle between the image acquisition platform and the Earth's surface. This angle is typically measured in degrees, with nadir being 0 degrees (looking straight down, perpendicular to the Earth's surface) and the horizon being 90 degrees.

Oblique. Oblique orientation captures imagery at any non-nadir angle, typically between the horizon and -45 degrees to the ground.

Overlap. Forward overlap is the amount of coverage between consecutive images along the same flight line, as measured in percentages. Side overlap is the amount of coverage between images in adjacent flight lines, as measured in percentage.

Receiver Operating Characteristic curves. Graphical plots that summarize the performance of a binary classifier system that show the true positive rate (sensitivity) on the y-axis and the false positive rate (1 – specificity) on the x-axis across multiple confidence thresholds.

Supervised Learning. A machine learning paradigm whereby the algorithm learns to distinguish between two or more categories of data. The training data presented to the algorithm is pre-labeled to indicate which category each individual datum belongs to.

Sensor. The sensor in a digital camera is a light-sensitive chip that records the image as a pattern of tiny squares, called pixels. The more photosites a sensor has, the higher the resolution of the image will be.

Small Unmanned Aircraft. An unmanned aircraft weighing less than 55 pounds on takeoff, including everything that is on board or otherwise attached to the aircraft.

Small Unmanned Aircraft System. A small, unmanned aircraft and its associated elements (including communication links and the components that control the small, unmanned aircraft) that are required for the safe and efficient operation of the small, unmanned aircraft in the national airspace system.

Test Imagery Limits. A smaller and more refined area of the training imagery limits to facilitate smooth test imagery collection times and field-testing logistics. A reduced subset of imagery compared to the training imagery limits was collected from these limits.

Training Imagery Limits. A determined area or areas of pavement that contain a vast diversity of surface conditions, types, and airport features or structures found in the airport environment. The limits are the imagery collection areas for training the deep learning network's library of information.

True Negative. In terms of the output of the FOD detection algorithm, a "Negative" result is an indication that FOD has NOT been detected. A "True Negative" is a negative result that is correct and matches the state of the world. In this case, there is no FOD present, and the algorithm did not report any.

True Positive. In terms of the output of the FOD detection algorithm, a “positive" result is an indication that FOD has been detected. A “true positive” is a positive result that is correct and matches the state of the world. In this case, the algorithm correctly identified a real piece of FOD.

Unsupervised Learning. An ML paradigm where an algorithm learns to identify patterns in a dataset where no knowledge or category information is provided.

EXECUTIVE SUMMARY

Foreign Object Debris (FOD) can create significant safety implications for aircraft, personnel and is a continuous concern in the airport environment. The Federal Aviation Administration (FAA) Airport Technology Research and Development (ATR) Branch conducted a research effort to explore the feasibility and maturity of using commercially-available small unmanned aircraft systems (sUAS) and artificial intelligence/machine learning (AI/ML) algorithms to detect FOD on airport surfaces. The objectives of this research effort were to develop a novel, proof-of-concept sUAS-based FOD detection workflow using AI/ML algorithms and to assess the workflow to determine whether it is capable of meeting all, some, or none of the requirements in FAA Advisory Circular (AC) 150/5220-24, *Airport Foreign Object Debris Detection Equipment*.

The research team developed and trained an AI/ML deep learning network as part of a sUAS-based FOD detection workflow. The research team conducted initial testing of this workflow at Cape May County Airport (WWD). The initial test effort at WWD comprised three stages: small-scale testing, calibration, and full-scale testing, each serving specific purposes. In the small-scale test stage, the primary focus was to identify initial standardized test items using black, white, and gray spheres, as specified in AC 150/5220-24 (i.e., the size of a standard golf ball), while simultaneously collecting essential data for training AI/ML algorithms. Subsequently, the calibration test stage introduced additional complexities to challenge the trained algorithms, incorporating factors such as variations in ground sampling distance (GSD) and the inclusion of additional FOD items. This stage aimed to enhance the algorithms' adaptability and performance. Finally, the full-scale test stage was conducted on an entire runway, encompassing the full range of AC 150/5220-24 FOD items. Leveraging the knowledge acquired from the previous test stages, this final stage served to assess the detection efficacy of the AI/ML algorithms while incorporating lessons learned and refining the detection process.

The research team then performed validation testing at Atlantic City International Airport (ACY) to validate the initial testing at WWD. This included collecting data on a runway at ACY, training an anomaly detection AI/ML algorithm, testing against a variety of FOD targets and addressing requirements such as implementing geolocation, and adhering to accuracy requirements set forth in FAA AC 150/5220-24.

The sUAS-based FOD detection workflow used the FastFlow ML deep learning algorithm and was capable of meeting some of the AC 150/5220-24 requirements, including achieving a 96% overall detection rate for FOD items specified in the AC. However, to meet the full set of AC 150/5220-24 requirements, further research and development on this technology is needed , including reducing the false positive rate, reducing the data processing time, implementing a software interface for displaying and recording FOD detection alerts, and detecting FOD in low-light and inclement weather conditions.

Finally, this testing led to the identification of best methods for data collection. Images captured from a nadir vantage point had a significantly higher true positive detection rate than images captured from an oblique vantage point. Positive detection rate was also improved by flying the sUAS lower to the runway for a 0.2-cm GSD, which increased the image resolution.

1. INTRODUCTION

The Federal Aviation Administration (FAA) Airport Technology Research and Development (ATR) Branch conducted a research effort to evaluate the feasibility of using commercially available small unmanned aircraft systems (sUASs)[1] and artificial intelligence/machine learning (AI/ML) algorithms to detect foreign object debris (FOD) on airport surfaces. This report summarizes the research conducted and results obtained from testing conducted in two different airport environments.

1.1 BACKGROUND

The presence of FOD on airport surfaces is a continuous concern for airport operators and poses a significant hazard to aviation safety and the traveling public. The standard, and most practiced, FOD identification method at airports is visual inspections. The challenges of frequent inspections are unique to each airport but are mostly centered on aircraft activity (runway usage) and the size of area inspected relative to the human resource constraints and visual inspection capabilities.

FOD can include anything from a small rock to a large bird, and it can cause extensive damage to aircraft engines, landing gear, and other critical components of safe aircraft operations. Airlines have implemented various strategies, such as developing checklists, that detail the specific areas of the aircraft that need to be inspected to mitigate FOD damage, but the cost of damage and delays continue to rise. The FAA report, *Foreign Object Debris Characterization at a Large International Airport* (2015), noted:

> Costs [for FOD-caused damage] to one major airline average $15,000 per aircraft, which represents an industry cost of over $60 million per year. This is the equivalent of one new medium-sized transport category jet.

Therefore, airports are looking for new ways to improve their FOD control programs. One potential solution is the use of automated FOD detection technologies that include mobile, radar, electro-optical (EO), or hybrid technologies; however, these are typically costly systems that may require specific installation. Interest in automated detection systems to identify FOD at airports, including leveraging sophisticated AI and ML algorithms and sUAS-based solutions, is growing due to the cost-effective nature of commercially available sUAS.

sUASs deliver proven capabilities in collecting high-resolution imagery and video for a variety of practical use cases. Their data collection ability, tied with low upfront cost and the efficiency of the technology, allows for implementation in applications such as mapping, surveying, obstruction analysis, condition assessments, and surveillance. With the increased capability and integration of sUAS across industries, research is being conducted to determine the feasibility, considerations, technical specifications, and policies of using sUAS for FOD detection in an airport environment.

[1] Small UASs are defined in Title 14 Code of Federal Regulations (C.F.R.) Part 107.3, Definitions, as unmanned aircraft weighing less than 55 pounds on takeoff, including everything that is on board or otherwise attached to the aircraft (Definitions, 2016).

This report provides a comprehensive overview of the research conducted. The report encompasses an in-depth account of the research methodology employed and outlines the outcomes of multiple iterative tests, including the results obtained from testing conducted in two different airport environments.

1.2 PURPOSE

The purpose of this research effort is to assess the feasibility and maturity of sUAS-based FOD detection on airport surfaces.

1.3 OBJECTIVES

This research effort had the following primary objectives:

1. Develop a novel, proof-of-concept sUAS-based FOD detection workflow using AI/ML algorithms.
2. Assess the performance of the sUAS FOD detection workflow to determine whether it is capable of meeting all, some, or none of the requirements in FAA AC 150/5220-24, *Airport Foreign Object Debris Detection Equipment* (2009).

1.4 RELATED DOCUMENTS

1. FAA AC 150/5220-24, *Airport Foreign Object Debris (FOD) Detection Equipment*
2. FAA AC 150/5210-24, *Airport Foreign Object Debris (FOD) Management*
3. 14 Code of Federal Regulations (CFR) Part 107, *Small Unmanned Aircraft Systems*

2. PERFORMANCE REQUIREMENTS FOR FOD DETECTION SYSTEMS

FAA performance specifications for FOD detection systems are contained in AC 150/5220-24 (FAA, 2009). This AC is directly relevant to this research effort as it set a baseline set of requirements against which the sUAS-based FOD detection technology could be assessed. Chapters 1 and 2 of the AC provide a general introduction to FOD such as FOD hazard, types, and sources. Four types of detection systems are discussed in this AC, including: stationary radar; stationary EO; stationary hybrid radar and EO; and mobile radar. AC 150/5220-24 defines requirements for basic functions, performance, system output, data presentation and data management standards.

Tables 1 through 4 below provide a summary of these requirements.

Table 1. Basic Performance Requirements

Category	Basic Performance Requirement
Basic Functions	Provide surveillance in the Airport Operations Area (AOA) as specified by the airport
	Detect and locate single and multiple FOD items on the AOA
	Provide an alert to the user when FOD has been detected
	Operate in conjunction with, and not interfere with, airport and aircraft communication, navigation, and surveillance systems
	Operate in conjunction with, and without interference from, normal airport and aircraft operations
	Provide a data record of detected FOD, allowing for equipment calibration and maintenance, and for analysis of the FOD event

Table 2. Detection Performance Requirements

Category	Detection Performance	Object Dimensions
Object Detection	An unpainted metal cylinder	1.2 in. (3.1 cm) high and 1.5 in. (3.8 cm) in diameter
	A white, gray, or black sphere	1.7 in. (4.3 cm) in diameter (i.e., a standard-sized golf ball)
	90% of the following group of objects when placed within a 100-ft by 100-ft (30-m by 30-m) square in the desired coverage area. One item from each category must be included in the group.	
	A “chunk” of asphalt or concrete	No larger than 4 in. (10 cm) in any dimension
	Any portion of a runway light fixture (in-pavement or edge light)	No larger than 4 in. (10 cm) in any dimension
	An adjustable crescent wrench	Up to 8 in. (20 cm) in length
	A deep socket	At least 2 in. (5 cm) in length
	A piece of rubber from an aircraft tire	No larger than 4 in. (10 cm) in any dimension
	A distorted metal strip	Up to 8 in. (20 cm) in length
	A fuel cap	No larger than 4 in. (10 cm) in any dimension
	A lug nut	No larger than 4 in. (10 cm) in any dimension
	A hydraulic line	Up to 8 in. (20 cm) in length
	A white polyvinyl chloride (PVC) pipe	2 in. (5 cm) in diameter
	Any two of the objects above, located no more than 10 ft (3 m) apart from each other, identified as separate objects	

Table 3. Detection Performance Requirements

Category	Detection Performance	Performance Requirement
Location Accuracy	Must provide location information for a detected object	Within 16 ft (5.0 m) of the actual FOD object location
Inspection Frequency	Mobile detection systems	Airport dependent
Surveillance Area	Airport operator will specify the desired surveillance (detection) area in the AOA requiring FOD detection.	Manufacturer of a FOD detection system must notify the airport operator of any locations within the specified surveillance area where detection would not be possible.
Weather	Must demonstrate the detection performance under both clear and inclement weather conditions	Detect FOD under rain snow, clear and inclement weather, lighting conditions, and time required for the system to recover after inclement weather
Alerts and Alarms	Must be able to alert the system operator to the presence of FOD in scanned areas with enough information to assess the severity of the hazard in order to determine if immediate object removal is necessary	False alarms should be minimal—1 per day with visual capabilities, 3 per day without visual capabilities

Table 4. System Output Requirements

Category	System Output	Performance Requirement
Detection Data	Data record on detected FOD	Alert time, date, location (at minimum)
Data Presentation	Coordinate scheme on maps of the airport, in an operator's console, or broadcast to mobile units	As specified by the airport
Data Management	Digital record	Capability to retain the data for at least 2 years after the detection event

3. RESEARCH APPROACH

Sections 3.1–3.6 provide an overview of the research approach and methodology.

3.1 TEST STAGES

As shown in Figure 1, the testing effort took place in three stages: small-scale, calibration, and full-scale/validation testing, each serving specific purposes. In the small-scale test stage, the primary focus was on identifying black, white, and gray spheres specified in AC 150/5220-24 as initial standardized items, while simultaneously collecting essential data for training AI/ML

algorithms. Subsequently, the calibration test stage introduced additional complexities to challenge the trained algorithms, incorporating factors such as variations in ground sampling distance (GSD) and the inclusion of additional FOD items. This stage aimed to enhance the algorithms' adaptability and performance. Finally, the full-scale test stage was conducted on an entire runway, encompassing the full range of AC 150/5220-24 FOD items. Leveraging the knowledge acquired from the previous test stages, this final stage served to assess the detection efficacy of the AI/ML algorithms while incorporating lessons learned and refining the detection process.

The research team conducted most of the testing at Cape May County Airport (WWD) and performed final testing at Atlantic City International Airport (ACY) to validate the WWD testing. This included collecting data on a runway at ACY, training an anomaly detection AI/ML algorithm, testing against a variety of FOD targets, and addressing requirements conveyed in FAA AC 150/5220-24, *Airport Foreign Object Debris Detection Equipment* (2009), such as implementing geolocation and adhering to accuracy requirements.

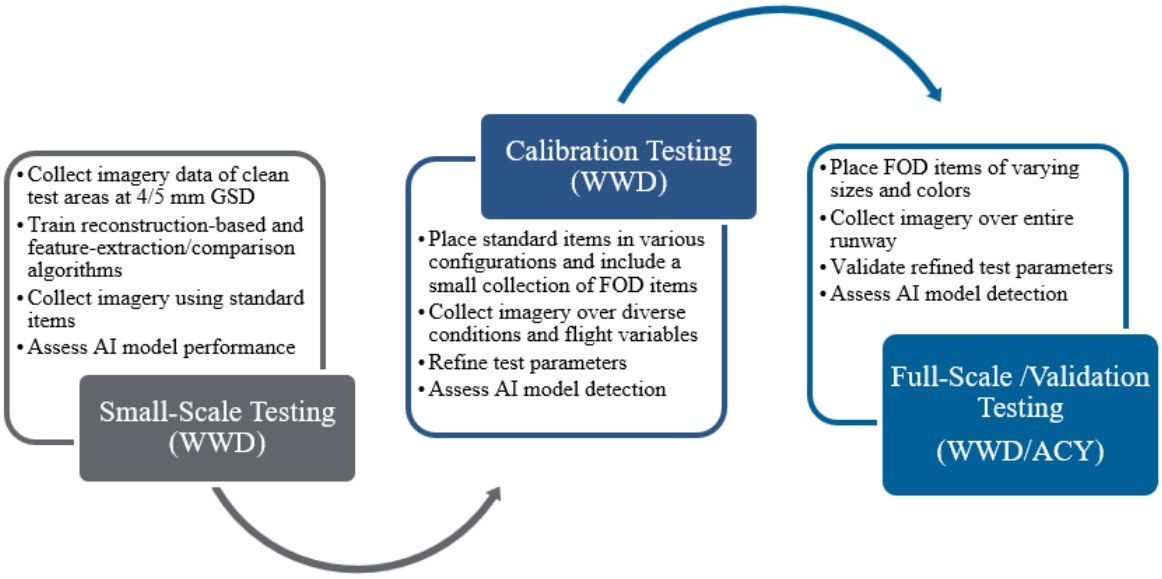

Figure 1. Study Conception and Design

3.2 SMALL UNMANNED AIRCRAFT SYSTEM PLATFORMS AND PAYLOADS

The research effort incorporated two commercially available sUAS models: the DJI M210 real-time kinematic (RTK) v2 and the DJI M300 RTK. Figure 2 depicts the DJI M210 RTK v2, which was equipped with a Zenmuse X7 EO sensor with a focal length of 50 mm. However, the M210 RTK v2, X7 airframe was solely used during the small-scale test stage of the research. Subsequently, the DJI M300 RTK, coupled with a Zenmuse P1 sensor, was selected for the calibration, full-scale, and validation testing. The decision to switch to the DJI M300 RTK airframe, as depicted in Figure 3, was based on its sensors' enhanced image resolution, which would result in each image crop having significantly more detail, thereby improving the detection algorithm's performance. Moreover, the adoption of the DJI M300 RTK brought

advantages such as extended flight time, enhanced position accuracy, and increased hardware reliability. The transition from the DJI M210 and X7 used during the initial testing to the DJI M300 and P1 represented an overall upgrade without any adverse effects on the quality of the training data parameters or data processing. For detailed specifications of the airframe and sensors, refer to Appendix A.

Figure 2. DJI M210 RTK v2 with a Zenmuse X7

Figure 3. DJI M300 RTK with a Zenmuse P1

During testing, two sUAS mission planning software applications were employed. Initially, DJI GS Pro software was used exclusively in the initial test stage, specifically designed to support the DJI M210 RTK v2 airframe. Subsequently, for all other test stages, DJI Pilot 2 software was adopted, which is compatible with the DJI M300 RTK airframe. Both software solutions exhibited ample flexibility in facilitating the collection of nadir and oblique imagery. The research team gathered imagery in either the “RAW” or “DNG” data format, depending on the specific airframe used.

3.3 AI/ML ALGORITHMS

The research team conducted a literature review to identify the most suitable AI/ML algorithms for testing. Based on this literature review, the research team selected the following two deep learning algorithms: omni-frequency channel-selection reconstruction—generative adversarial network (OCR-GAN) and FastFlow. A significant benefit of both algorithms is that they do not require training data that is manually annotated with ground truth labels. These are described in further detail in Sections 3.3.1 and 3.3.2.

3.3.1 Omni-Frequency Channel-Selection Reconstruction—Generative Adversarial Network

OCR-GAN is a reconstruction-based deep learning algorithm that has been shown in various studies to be highly effective for anomaly detection (Zhou & Paffenroth, 2017; Deecke et al., 2019; Di Mattia et al., 2021; Liang et al., 2022). Reconstruction-based algorithms such as OCR-GAN perform anomaly detection by deconstructing an input image into a set of "features" that are learned during the training process; then, the algorithm reconstructs the image using those features.

The OCR-GAN algorithm was selected based on its performance with MVTec, which is one of the most popular datasets for testing anomaly detection algorithms (Bergmann et al., 2021). This algorithm uses a unique approach that fixes the problems other GAN-based methods sometimes encounter where they become so good at reconstruction that even the anomalies may still be reproduced. While many alternative deep learning systems must supplement MVTec training data to achieve their competitive results on the test set, OCR-GAN does not require this and achieves its performance solely on the provided training data. This smaller data requirement greatly simplifies the process by which training data must be obtained and processed.

3.3.2 FastFlow

FastFlow is a feature-extraction-based algorithm. Feature-extraction algorithms use various methods to extract features and convert images into quickly comparable combinations of features (Rao et al., 2017; Tayeh et al., 2020; Cohen & Avidan, 2021; Bergmann et al., 2020; Reiss et al., 2021; Yu et al., 2021). However, these algorithms often require additional training data to match or surpass the performance of reconstruction-based methods. In feature-extraction-based algorithms, images are dissected into collections of distinctive features, and a database is maintained to track encountered features during training. Anomalies are detected when the input images contain features that differ from those in the training dataset.

FastFlow was selected for this study due to its superior performance on the MVTec dataset. Another reason for investigating feature-extraction-based methods is their relatively lower computational resource requirements, including random-access memory (RAM) and central processing unit (CPU) time, compared to reconstruction-based methods. This advantage enables the processing of a greater number of images within the same timeframe and enhances the functionality of real-time systems that necessitate continuous image processing at a high pace. Figure 4 shows an overview of the FastFlow architecture (Yu et al., 2021).

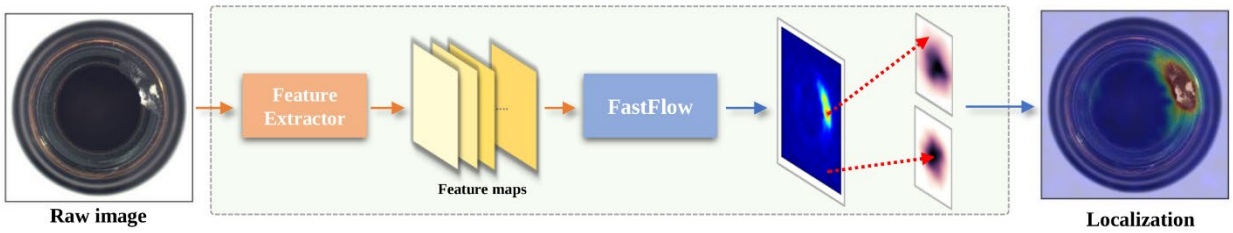

Figure 4. Overview of the FastFlow Architecture

3.4 FOREIGN OBJECT DEBRIS TEST ITEMS

Two primary types of test items were used: standardized targets with uniform dimensions (golf balls) and the FOD items listed in AC 150/5220-24.

1. Standardized Targets—Golf balls in white, gray, and black.

   Golf balls, as shown in Figure 5, were used as standard targets throughout each test stage. The utilization of a single shape, the sphere, in three different colors, intended to control the test item variable to assess the efficacy of the AI/ML algorithms of detecting items of varying contrast relative to the color of pavement. Contrast is one of the most important variables affecting the success of ML algorithms using red, green, blue (RGB) imagery.

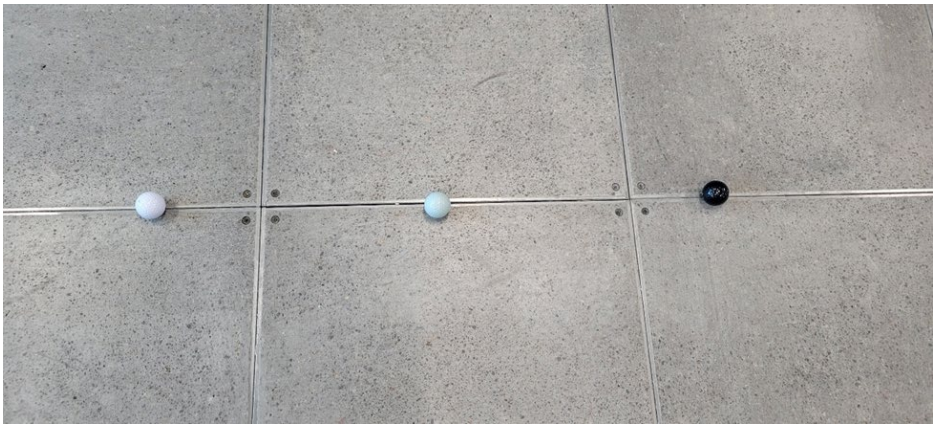

Figure 5. Golf Balls in White, Gray, and Black

2. AC 150/5220-24 Test Items

   During full-scale and validation testing, a variety of test items were introduced, consisting of fuel caps, hydraulic hoses, taxi lights, wrenches, PVC pipes, rubber scraps, sockets, nuts and bolts, asphalt chunks, metal pipes, and metal scraps. These test items are shown in Figure 6.

Figure 6. Array of FOD Test Items

3.5 OPERATIONS AND SAFETY FOR SMALL UNMANNED AIRCRAFT SYSTEMS

To conduct the research testing efforts as intended, the research team diligently carried out airport coordination measures. All sUAS operations were conducted during scheduled runway maintenance closures. Prior to commencing operations, an sUAS Operations License Agreement was established with the Delaware River & Bay Authority (DRBA), acting as the sponsor to WWD. This agreement outlined the general operational characteristics and limitations related to sUAS activities. Additionally, the research team obtained an airspace authorization for ACY and coordinated with airport operations personnel with the South Jersey Transportation Authority (SJTA).

As a further safety step, the research team effectively communicated and submitted a Notice of Proposed sUAS Operation form to each respective airport when scheduling sUAS flight operations. This form served as an operational communication to the airport operator to simplify the Notice to Air Mission (NOTAM) process by providing comprehensive mission details, such as date, time, duration, altitude, and the specific location(s) of operations on the airfield. Upon approval, the airport issued a NOTAM to ensure traditional aviation pilots were aware of the sUAS operations, including the designated time, location, and maximum altitude.

To reduce the inherent risk involved in placing FOD on an airport surface, the research team implemented procedures and followed safety precautions when placing and removing FOD. As a safety measure for FOD being left on the airfield, the research team employed an inventory and

storage solution as illustrated in Figure 7. Each FOD item was accounted for when being placed, and again when being picked up. The solution allowed the team to manage inventory and accountability prior to the departure from the AOA. Additionally, the research team inspected the test area prior to testing and removed any FOD that was present. Following each testing effort, a FOD walk was conducted to ensure that the test areas were left FOD-free and safe for aircraft activity.

Figure 7. Inventory and Storage Solution for FOD Items

3.6 DATA PROCESSING AND ANALYSIS WORKFLOW

The following section outlines the processes in the data workflow methodology developed for the testing. The data analysis workflow consists of multiple steps, summarized in Figure 8.

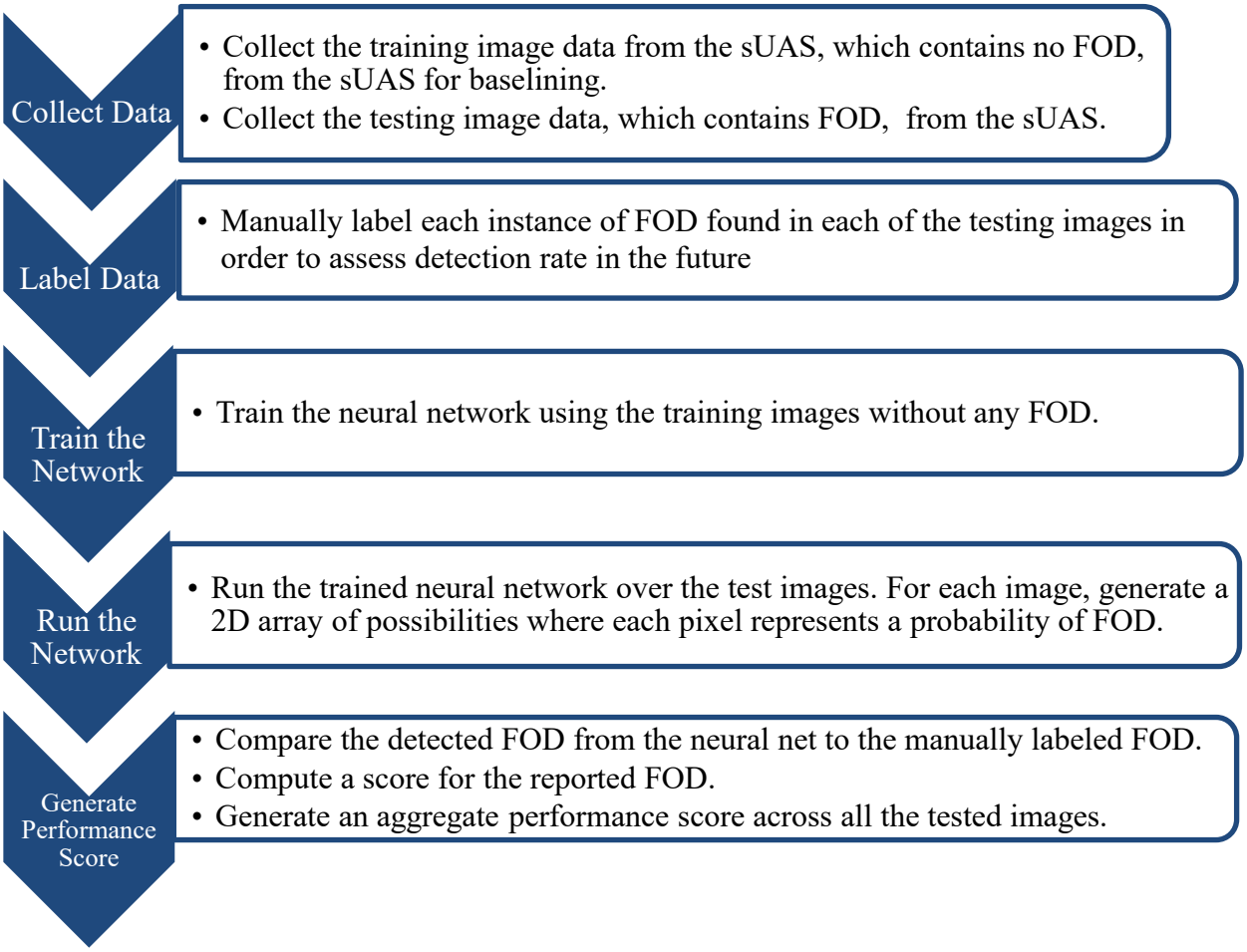

Figure 8. Data Analysis Workflow

3.6.1 Collect Data

The Collect Data step involved deploying the sUAS over the test areas to collect imagery. The first dataset was collected over the test areas with no FOD present. The second dataset was collected over the test areas with FOD present. The data without FOD was intended to train and provide the AI/ML algorithms reference images, whereas the data with FOD was intended to assess the algorithm's performance at detecting FOD as anomalies. After each data collection exercise, the images were copied from the sUAS onto a computer that ran the FOD detection algorithms. Images were separated into 256- by 256-pixel crops for batch processing. The research team also used crops from the "test" images to supplement additional training. All FOD crops were used for testing, and the non-FOD crops used for testing were randomly sampled from both sets of images.

3.6.2 Label Data

The Label Data step was only required for analyzing the performance of the algorithms as part of this research and would not be performed in a real-world deployment of the system. This step required that the ground truth location of each piece of FOD on the test areas be known ahead of time. The Label Data step was used exclusively to determine the accuracy of the algorithms'

performance in the post-processing stage. The process of image labeling consisted of manually using an image manipulation program to paint a white pixel overlay on a copy of each image for all pixels corresponding to FOD items and a black overlay for all non-FOD pixels.

3.6.3 Train the Network

Before the neural network could be used to detect FOD, it had to be trained on an initial dataset. The training dataset contained images of a clean test area without any FOD. The intent was to teach the neural network what a FOD-free area looks like so that any FOD encountered in future imagery would appear as anomalies to the network and be flagged. The neural network operated on image sizes that are smaller than the raw images taken from sUAS cameras. Therefore, each of these large individual images were cropped into many small overlapping sub-image blocks. To ensure that there was enough overlap between the crops to avoid missing any test area details, the input image width and height was divided by the crop width and height and rounded up. Crop locations were selected by spreading this number of crops equally across the width of the image. This method of cropping ensured that all pixels collected were used for training. It also minimized the amount of overlap that would cause the algorithms to process the same pixels multiple times.

After the training dataset was generated, the neural network went through a traditional deep learning training procedure, in which involved the dataset was repeatedly passed through the network architecture. Each image processed by the neural network would calculate the output error for what the architecture produced versus what it should have produced. This error value was used to adjust the parameters of the neural network via an algorithm called backpropagation. As the process repeated itself, the error values seen at the output of the networks were reduced, and the parameters were optimized. The training process stopped when the error was reduced to a level where further training would not generate any improvement.

In the next steps of the research process, the set of cropped images were augmented by duplicating and transforming them in various ways to provide additional perspective in training the neural network. These additional cropped images were created by rotating and/or mirroring the original image crops. Rotating the original images provided the neural network with multiple valuable perspectives, as the sUAS could potentially view the test area from any heading or direction. However, this is only true when the sUAS is taking images from directly over the runway (e.g., in a nadir orientation). If the sUAS were to view the runway from the side (e.g., in an oblique orientation), rotating the image could create “upside down” images of the horizon that would never actually be encountered. In addition, where possible, horizontal mirroring was performed to augment the final dataset.

Finally, to refine the datasets and lower the false positive rate, a standard deviation (SD) filter was applied to the input images, which required all images to surpass a threshold of pixel intensity SD values before it could be passed to the network for training. The purpose of the SD filter was to curate the dataset to have a higher concentration of images with paint, cracks, and generally interesting features, while reducing the concentration of just plain flat pavement. The SD filter reduced the number of false positives found by increasing the percentage of ”interesting” non-FOD pavement images used for training, while reducing “uninteresting”

uniform pavement that had no distinguishing marks like paint, tar, or cracks. The SD filtering process only removed the plainest images.

3.6.4 Run the Network

The Run the Network step involved running the trained network on each of the testing dataset images to identify the presence of FOD. As with training, each of the raw images were broken into a set of overlapping sub-images. The neural network processed each of these input sub-images and for each sub-image, generated, an output image of the same size where each pixel represented a probability that it contains a piece of FOD.

3.6.5 Generate Performance Score

The Generate Performance Score step provided the research team with quantitative results to analyze algorithm performance. The performance score was generated by taking the individual pixel probabilities of the output images and comparing them to the pixels in the ground truth imagery to score each report. The individual scores were then aggregated to generate a total performance score across all images. The scores were used to determine whether reported FOD constituted a true positive, false positive, false negative, or true negative when compared to the FOD ground truth of the Label Data step.

A set of example outputs from the algorithm are shown in Figure 9. Column 1 is the RGB crop. Column 2 shows the manually created ground truth labels. Column 3 shows a probability heatmap for all pixels in the image overlayed on the original image (brighter corresponds to higher likelihood). Column 4 shows the predicted pixels whose confidence in the heatmap were above the determined threshold for a detection. Column 5 shows small clusters on the original image from the predicted detection pixels.

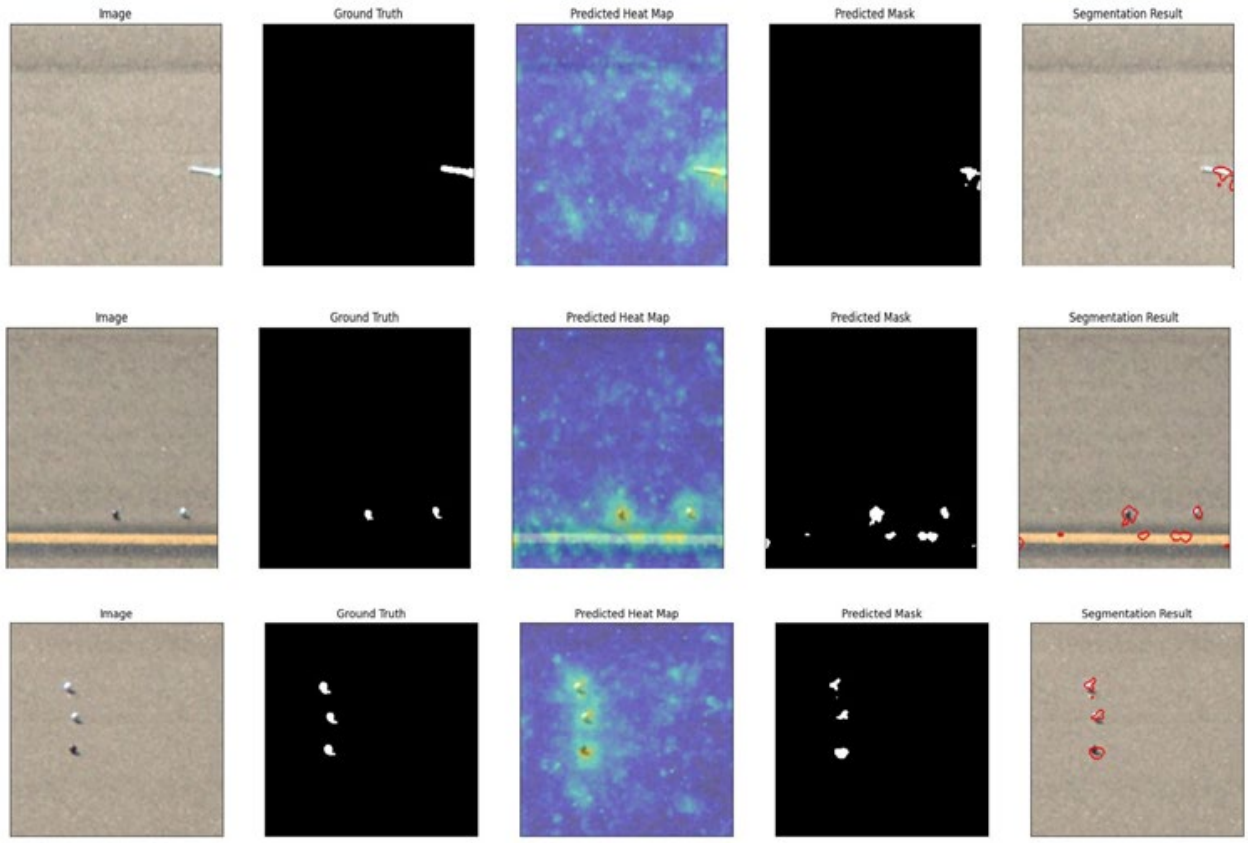

Figure 9. Example Outputs from the FOD Detection Algorithm

Quantitative results for the datasets were reported in Receiver Operating Characteristic (ROC) curves, which are graphical plots that summarize the performance of a binary classifier system. They show the true positive rate (sensitivity) on the y-axis and the false positive rate (1—specificity) on the x-axis across 90% (initially 60%) confidence thresholds for counting a pixel or image as an anomaly (e.g., containing an instance of visible FOD). The ROC curve is a useful tool for evaluating the performance of a binary classifier, as it provides a visual representation of the trade-off between sensitivity and specificity. A classifier with a high true-positive rate and a low false-positive rate will have a high Area Under the Curve (AUC) value, indicating good performance. On the other hand, a classifier with a low true-positive rate and a high false-positive rate will have a low AUC value, indicating poor performance. ROC curves can be used to compare the performance of different classifiers, and to choose the optimal discrimination threshold for a given classifier. Figure 10 shows a representative ROC curve for one dataset processed as part of this research.

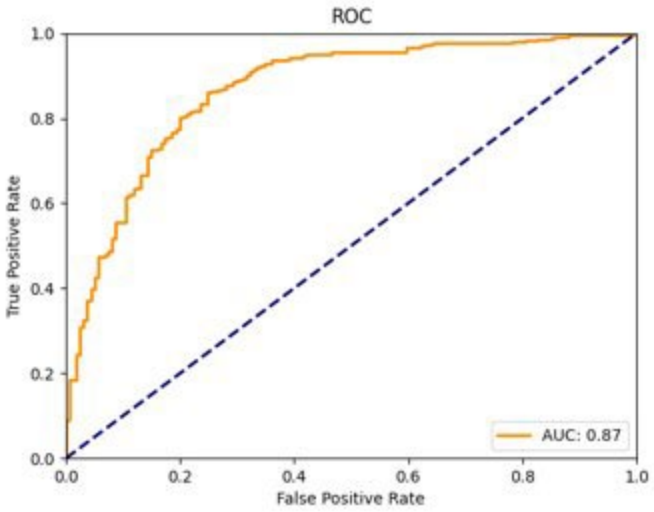

Figure 10. Representative ROC Curve

ROC curves were calculated for both pixel-wise and image-wise accuracy. Pixel-wise accuracy is a measure of how accurately a model can predict the label or class of each pixel in an image. For example, if the model is being used for image segmentation, pixel-wise accuracy measures the percentage of pixels that are correctly classified by the model. Image-wise accuracy is a measure of how accurately a model can classify an entire image. For example, if the model is being used for image classification, image-wise accuracy measures the percentage of images that are correctly classified by the model.

It is important to note that pixel-wise accuracy and image-wise accuracy are not necessarily the same, and a model with high pixel-wise accuracy may not necessarily have high image-wise accuracy, and vice versa. For example, a model with high pixel-wise accuracy can still err when classifying an entire image, if it has misclassified many pixels in the image. Similarly, a model with high image-wise accuracy may still have low pixel-wise accuracy if it has erred in classifying individual pixels in the image.

For this application, the research team focused on image-wise accuracy instead of pixel-wise accuracy. As the objective was to identify instances of FOD in an image, the presence of any identified FOD pixels in an image would be sufficient to notify an operator to look at that image and/or location on the runway.

4. INITIAL TESTING: CAPE MAY AIRPORT

Because of its proximity to the FAA William J. Hughes Technical Center, WWD was chosen as the initial test site for sUAS-based FOD detection. WWD presents a diverse array of pavement materials, paint markings, and pavement distress, making it an ideal location for gathering a comprehensive dataset to train deep learning algorithms. Moreover, the airport's low traffic density and uncontrolled airspace, in compliance with Title 14 Code of Federal Regulations, Part 107, Small Unmanned Aircraft Systems (14 CFR Part 107), make it an optimal location for conducting sUAS operations. It is important to highlight that ATR holds a Memorandum of

Agreement with the DRBA, which enables airport safety and pavement research to be performed at WWD and further emphasizes WWD's suitability for conducting sUAS-based FOD detection testing.

4.1 SMALL-SCALE TESTING

For the small-scale testing, imagery of "clean" (FOD-free) training data was collected from test areas to train reconstruction-based and feature-extraction/comparison deep learning algorithms. Once trained, these algorithms were tested against standard FOD objects to assess detection performance under various flight parameters and pavement types.

4.1.1 Algorithm Training Requirements

The research team's approach for deep learning-based anomaly detection made use of a feature extraction-based algorithm called FastFlow, described in Section 3.3.2. The FastFlow algorithm first decomposes an input image into a set of "features" that are learned during the training process. The training datasets containing only "clean" images without FOD or other anomalies are encoded into a database of coordinates in a feature space. During inference, the features of new images are compared to the database of "clean" image features, and anomalies are determined based on their distance from the "clean" features.

Images consist of cropped image segments, known as "blocks," that are extracted from a single source image collected via the sUAS. Depending on the input needs of the neural network being trained, these blocks can be produced in different sizes, such as 256x256 or 512x512 pixels. Using a 24-megapixel camera, such as the Zenmuse X7, the sUAS can capture approximately 500 to 750 source images from a single data collection that can be used to generate 15,000 to 20,000 image blocks, which can then be used to train AI/ML algorithms.

For the algorithm to perform successfully, the training dataset must offer:

- An extensive variety of images with diverse appearances
- Different surface materials, such as asphalt and concrete
- A range of different painted markings
- Images captured at different times of day

4.1.2 Test Areas

The research team selected three test areas at WWD, with an emphasis on operational safety and parameters, such as material type and paint color. The sUAS performed flights parallel to the length of the operating area, like a runway, in the scoped operating areas. These test areas also met the algorithm training requirements discussed above.

The data collection was designed to comply with the 14 CFR Part 107 regulations. The training imagery limits were intended to collect a vast diversity of airport surface imagery to build a library of training data for the deep learning networks. The test imagery limits were smaller than the training imagery limits to allow for more efficient field data collection times and to better account for FOD items placed on the airfield.

4.1.2.1 Test Area 1

Test Area 1, identified as "Taxiway," is an asphalt taxiway located on the FAA ATR research Taxiway C at WWD. This operating area is approximately 1,100 ft long by 100 ft wide, containing an asphalt runway with yellow markings. This test area is illustrated in Figure 11.

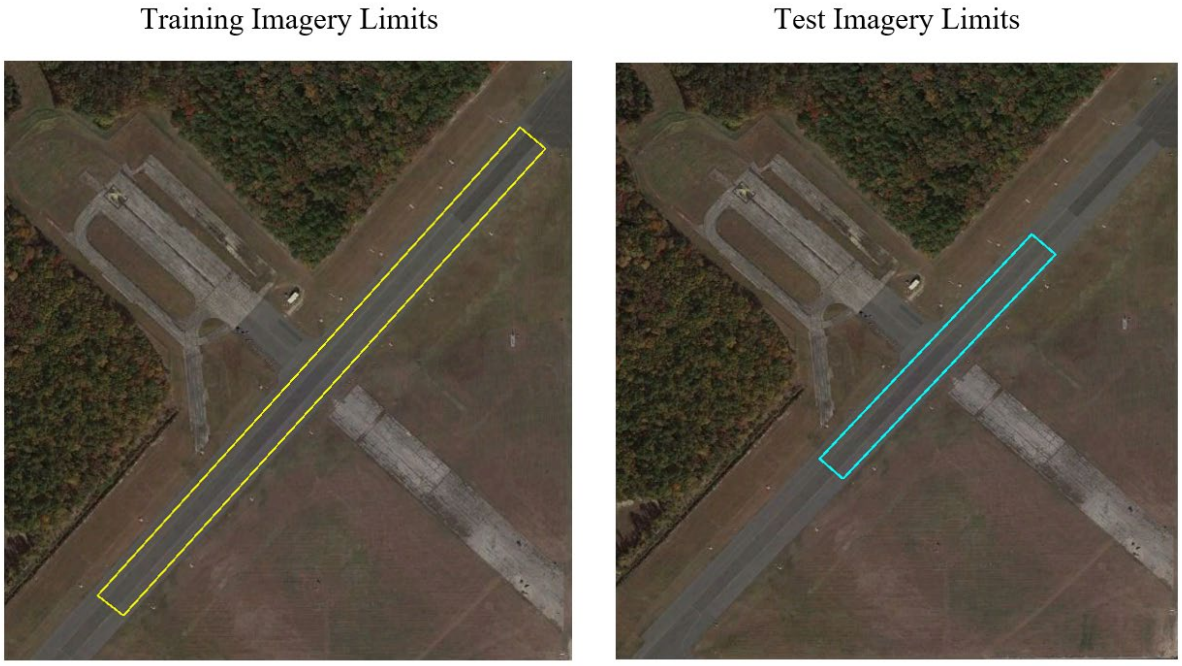

Figure 11. Initial Testing—Test Area 1

4.1.2.2 Test Area 2

Test Area 2, identified as "Runway," encompasses the western (approach) end of Runway 10. This section of the runway is approximately 500 ft long by 150 ft wide and consists of asphalt pavement with various white runway paint markings and rubber buildup. This test area is illustrated in Figure 12.

Figure 12. Initial Testing—Test Area 2

4.1.2.3 Test Area 3

Test Area 3, identified as “Fixed-Based Operator (FBO) RAMP”, is a ramp area on the southwestern portion of WWD consisting of asphalt and concrete pavement. The test area is approximately 720 ft long by 180 ft wide, with aircraft tie downs and paint markings of yellow and black color. This test area is illustrated in Figure 13.

Figure 13. Initial Testing—Test Area 3

4.1.3 FOD Selection

This section discusses the selection of FOD items for initial testing and their placement within the selected test areas.

4.1.3.1 FOD Items

The research team chose uniform standardized items and an additional test item from FAA AC 150/5220-24 to act as a primary control FOD target for the initial testing, as outlined in Section 3.4 of this report. The set of three standardized items, connected by a string, provided a consistent shape but with a color difference to evaluate the detection of object color contrasts in different situations; the control target, in this case a standard wrench, provided a larger object that served to evaluate the detection of objects with a different shape.

Table 5 summarizes the type and number of FOD items that were placed on the test areas.

Table 5. Type and Number of FOD Items

Type of FOD	Number of FOD Items
Golf Ball (White)	34
Golf Ball (Gray)	34
Golf Ball (Black)	34
Wrench	6
Total	**108**
Item Totals by Test Area	
Runway	41
Taxiway	41
FBO Ramp	26

4.1.3.2 FOD Placement

Prior to data collection, the research team surveyed primary characteristics of the test areas such as pavement type, condition, and markings, also making note of features such as elevated/flush lights and rubber buildup. The research team placed test items in such a way that a FOD target would be within proximity to one or more of the various characteristics or features within each test areas. Ample spacing between FOD targets was done to prevent overlap between different targets. Each item placed consisted of either one wrench or a string of three golf balls. Additionally, to mitigate the risk of FOD being left on the airfield, the research team utilized the inventory and storage solution as previously discussed in Section 3.5 of this report.

Test Area 1 contained 15 locations where FOD was placed, Test Area 2 contained 15 FOD items, and Test Area 3 contained 10 FOD items, as shown in Figures 14 15., and 16, respectively.

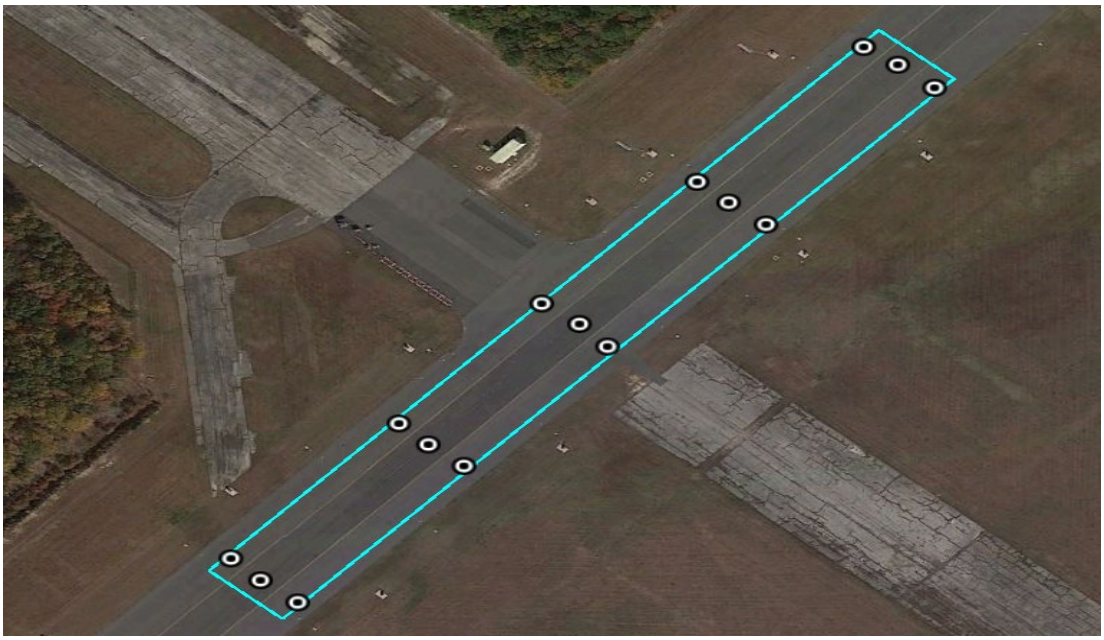

Figure 14. Initial Testing FOD Placement—Test Area 1

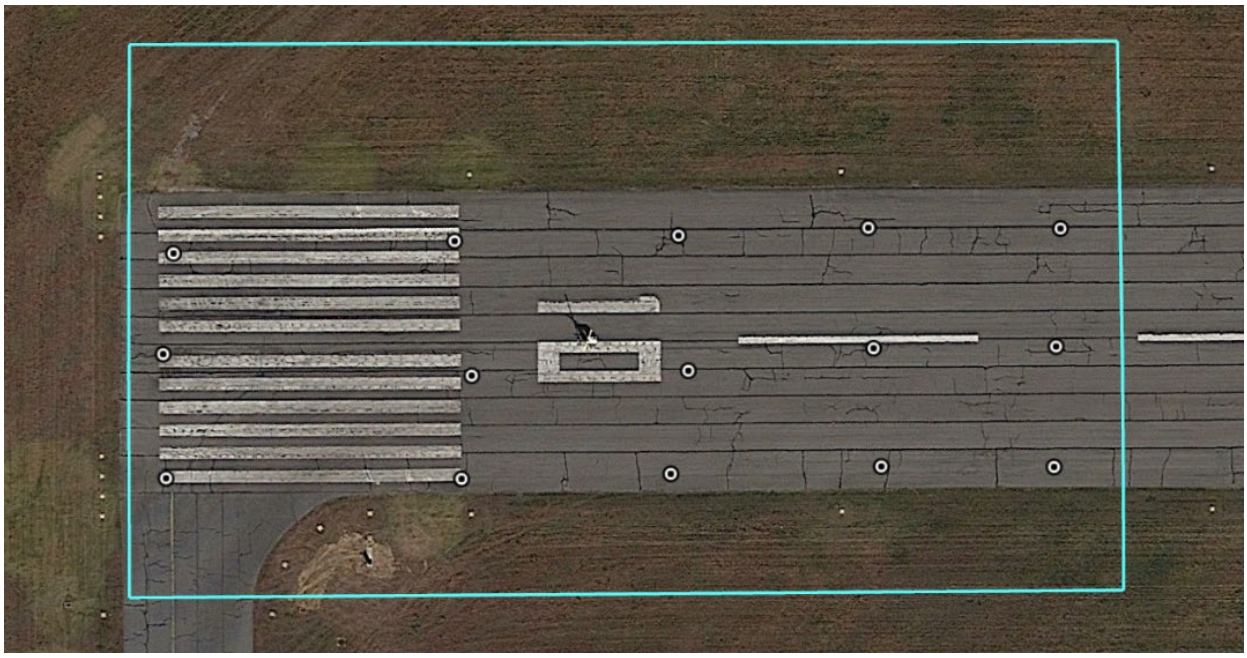

Figure 15. Initial Testing FOD Placement—Test Area 2

Figure 16. Initial Testing FOD Placement—Test Area 3

4.1.4 sUAS Data Collection

The research team collected RGB data with and without selected FOD items for 1 week. At each of the three test areas, multiple datasets were collected with controlled and uncontrolled variables. Controlled variables included the use of airframe, sensor, FOD items, and their placement location. uncontrolled variables included sun angle and the orientation of FOD items . During initial testing, the team chose a single airframe and sensor to eliminate any variables that

could occur due to system variances. Therefore, the DJI M210 RTK v2 with a Zenmuse X7 (50 mm focal length) was selected.

4.1.4.1 Flight Planning Characteristics and Parameters

The research team collected imagery data using an industry standard sUAS at a GSD of approximately 0.4 cm for initial testing. The goal of this testing was to collect data with and without FOD. The data without FOD was intended to train the AI/ML algorithms, while the data with FOD was intended to assess the algorithms' performance at detecting FOD as anomalies. During initial testing, variables such as the airframe, sensor, and detection items were controlled to allow for parameters such as surface type, sun angle, and image capture angle to be systematically tested.

Imagery was collected in two different orientations: nadir and oblique. Nadir orientation refers to imagery that is captured while looking perpendicular to the ground. Oblique orientation covers a wide range of possible capture parameters including all camera angles not perpendicular to the ground. Both the camera's angle relative to the ground as well as the camera's angle relative to the aircraft's flight path are variable. These variables require the test to strike a balance between operational efficiency, GSD consistency, and safety.

Oblique imagery was collected with the sensor aimed at 22.5 degrees below parallel (a 67.5-degree angle above nadir) and rotated 90 degrees perpendicular to the flight path, because it offered the best balance between consistent GSD across the imagery and the theorized benefits of oblique imagery. The aircraft was programmed to fly with a lateral offset from the runway centerline as required by calculation of an image center GSD of approximately 0.4 cm. Figure 17 illustrates example images of various oblique angles. Table 6 summarizes the flight parameters implemented during small-scale initial testing.

Figure 17. Illustration of 20-, 30-, and 45-Degree Oblique Angles, Respectively

Table 6. Flight Parameters – Initial Testing

Flight Parameters	
Mission Planning Software	DJI GS Pro
Data Format	DNG
sUAS Orientation	Nadir and Oblique
Overlap	15/15

4.1.4.2 sUAS Data Collection Workflow

The research team developed and implemented the testing workflow shown in Figure 18.

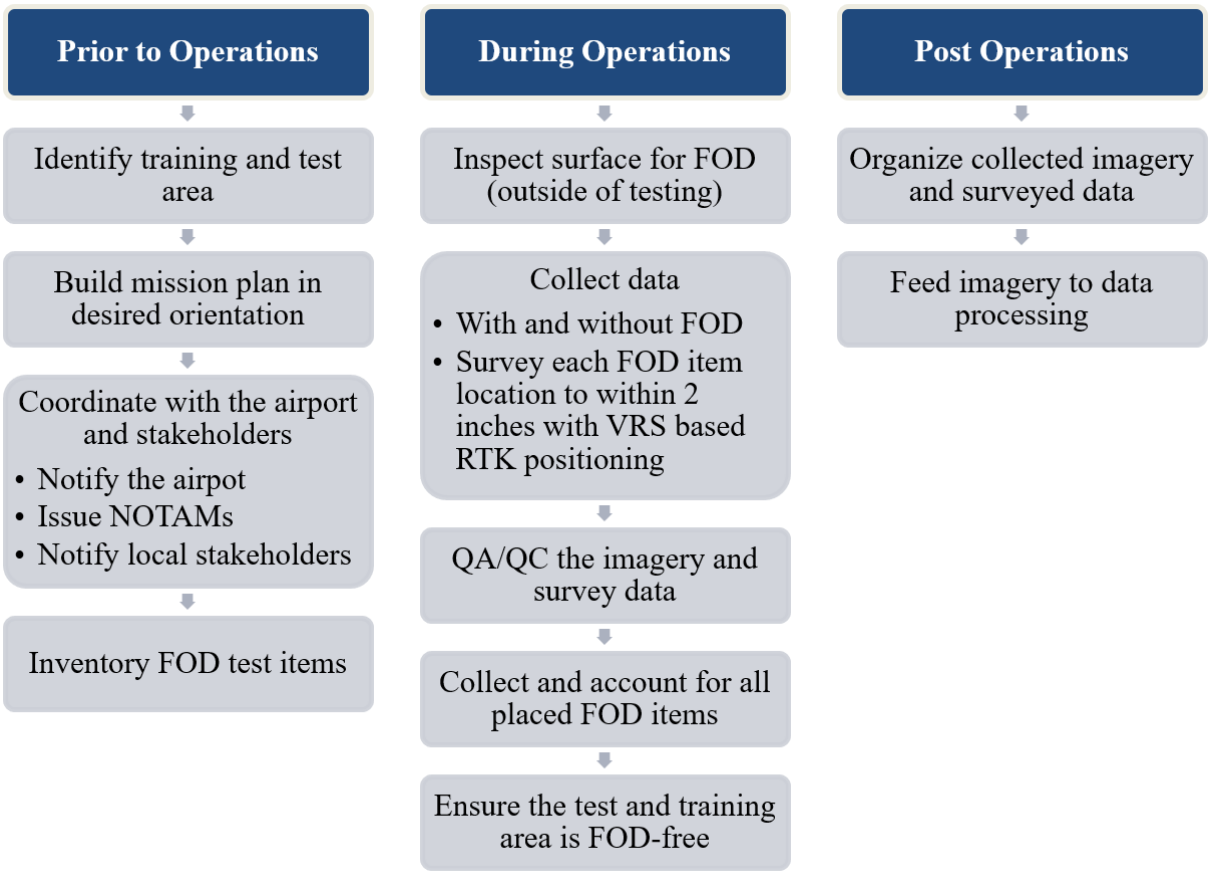

Figure 18. sUAS Data Collection Workflow

4.1.5 Data Processing

This section discusses the data processing techniques using the two deep learning methodologies discussed earlier in Section 3.3 of this report.

4.1.5.1 Algorithm Training

The methodology for the algorithm training process is outlined in Section 3.6, wherein the data captured by the sUAS was first labeled and then used for training the networks. The network training involved splitting the large raw imagery into many smaller image "blocks" that would pass through the algorithm repeatedly until no further training would yield improvement in the algorithm's parameters.

During the initial evaluation stage, the team was unable to get the open-source OCR-GAN downloaded code repository working properly. When tested on the runway data, the OCR-GAN algorithm appeared to reproduce anomalous portions of the images as well as the non-FOD sections. This behavior was unexpected given the results detailed in the published literature, and the research team identified that the acquired dataset was significantly different from the one it was designed for. Due to time and resource limitations, the decision was made to abandon the use of OCR-GAN. Therefore, all training was performed using the FastFlow algorithm only.

4.1.6 Data Analysis

A set of example outputs from the FastFlow algorithm are shown in Figure 19 below. Column 1 is the RGB crop. Column 2 shows the manually created ground truth labels. Column 3 shows a probability heatmap for all pixels in the image overlayed on the original image (brighter corresponds to higher likelihood). Column 4 shows the predicted pixels whose confidence in the heatmap were above the determined threshold. Column 5 shows small clusters on the original image from the predicted pixels.

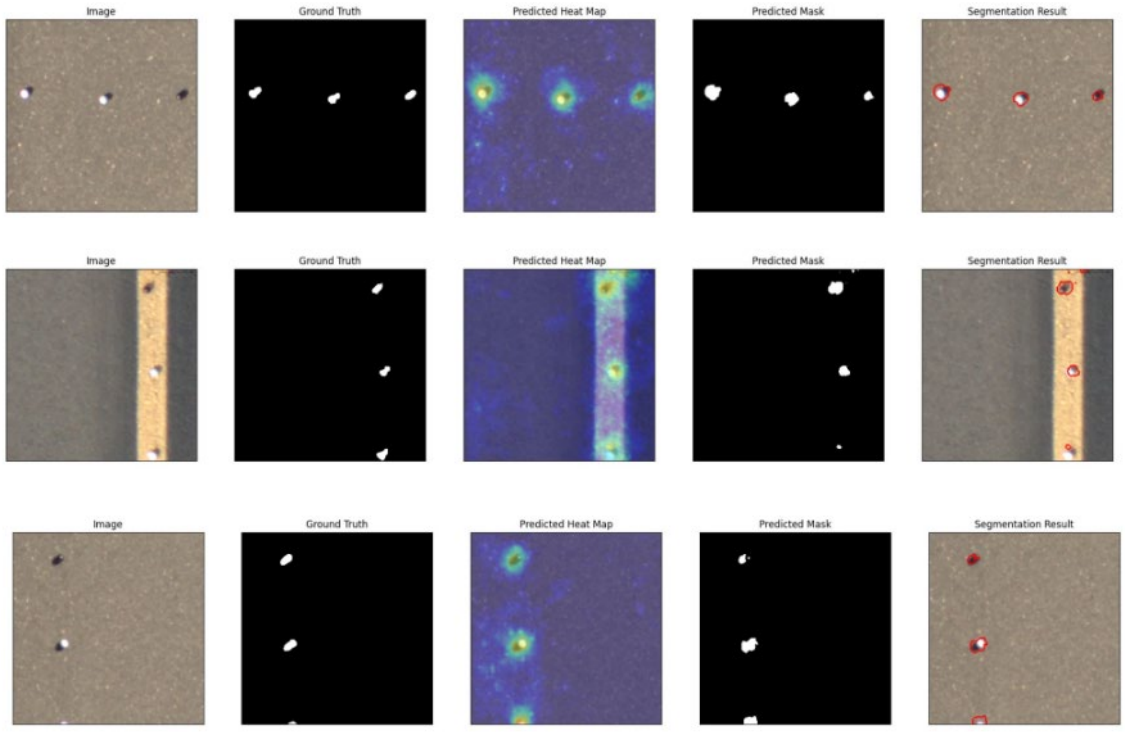

Figure 19. Example Outputs from the FastFlow Algorithm

FAA AC 150/5220-24 requires a positive detection rate of 90% for FOD detection equipment. Throughout the small-scale testing, the research team used a lower confidence threshold of 60%, which is an industry standard, to calculate the initial testing datasets. This does not reflect a true positive rate of 60%; rather, it is an indication that the algorithm is 60% confident that an anomaly is present, so those pixels and images are labeled as anomalous. The small-scale testing results provided baseline parameters that could then be refined to the point where the algorithm would meet the FAA AC 150/5220-24 requirement.

Quantitative results for the oblique and nadir datasets were reported in the form of ROC curves. ROC curves summarize the performance of the algorithm in terms of true positive rate to false positive rate. The AUC value indicates the performance results, where a high AUC value represents good performance, and a low AUC value is poor performance. ROC curves for both pixel-wise and image-wise accuracy were calculated.

Figure 20 shows the image-wise ROC curve on the left image and pixel-wise ROC curve on the right image for the nadir dataset.

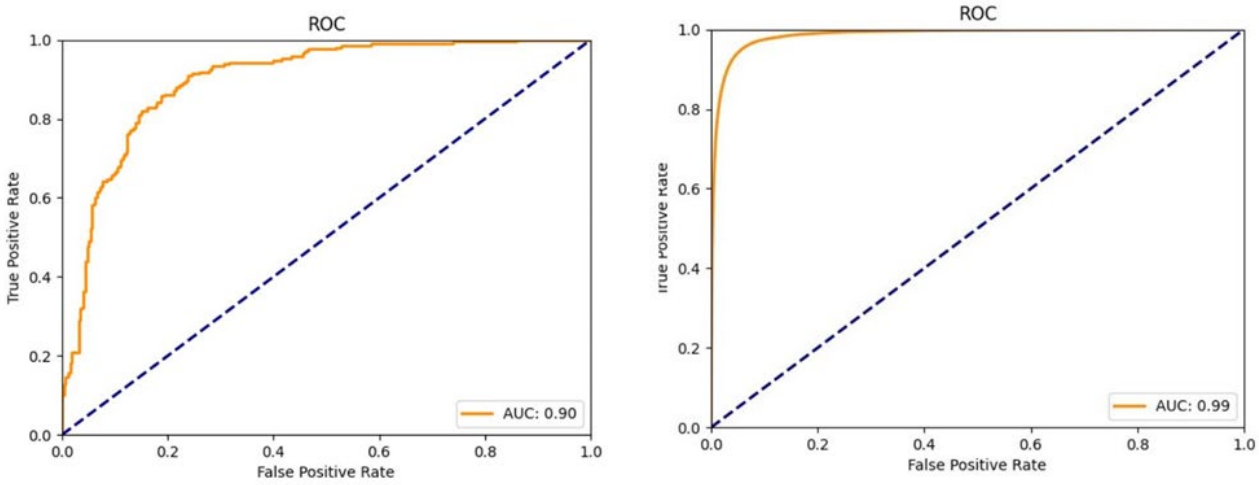

Figure 20. Image-Wise (Left) vs Pixel-Wise (Right) ROC Curve, Nadir

Figure 21 shows the image-wise ROC curve on the left image and pixel-wise ROC curve on the right image for the oblique dataset.

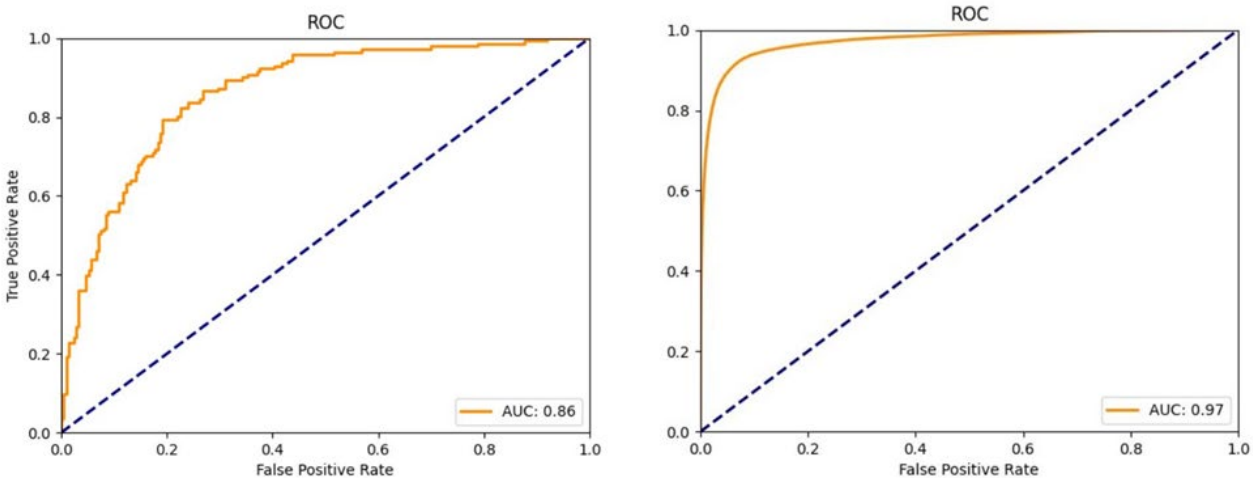

Figure 21. Image-Wise (Left) vs Pixel-Wise (Right) ROC Curve, Oblique

Additionally, the AUCs were calculated for each ROC and are shown in Table 7.

Table 7. Values of AUC

ROC	AUC Value
Nadir Image-Wise ROC	0.90
Nadir Pixel-Wise ROC	0.99
Oblique Image-Wise ROC	0.86
Oblique Pixel-Wise ROC	0.97

A detailed evaluation was completed to see how FastFlow performed with each FOD type on various backgrounds. These metrics were manually recorded by tabulating the details of each FOD image.

- For wrenches, an image was counted if greater than 50% of the wrench was in the image.
- FOD was counted as found if even a single pixel in the FastFlow output overlapped with the ground truth.
- False positives were counted if no pixel in the FastFlow output overlapped the ground truth FOD.

The results of FOD detection are shown in Table 8.

Table 8. Statistics for FOD Detection

Oblique	
White Balls	242 Find, 24 Miss
Gray Balls	227 Find, 36 Miss
Black Balls	187 Find, 77 Miss
Wrench	25 Find, 21 Miss
Nadir	
White Balls	155 Find, 12 Miss
Gray Balls	151 Find, 19 Miss
Black Balls	117 Find, 37 Miss
Wrench	26 Find, 1 Miss

The oblique images were analyzed based on where the crop appeared in the image. FOD items closer to the camera appeared larger than items that were further from the camera. A test was performed to determine if the range from the camera significantly affected the results. The crops were split into two categories based on whether they were from the front or back half of the image. Figure 22 shows the ratio of Find vs Miss, where a value of 0 means no misses, and a value above 1 means more FOD was missed than found.

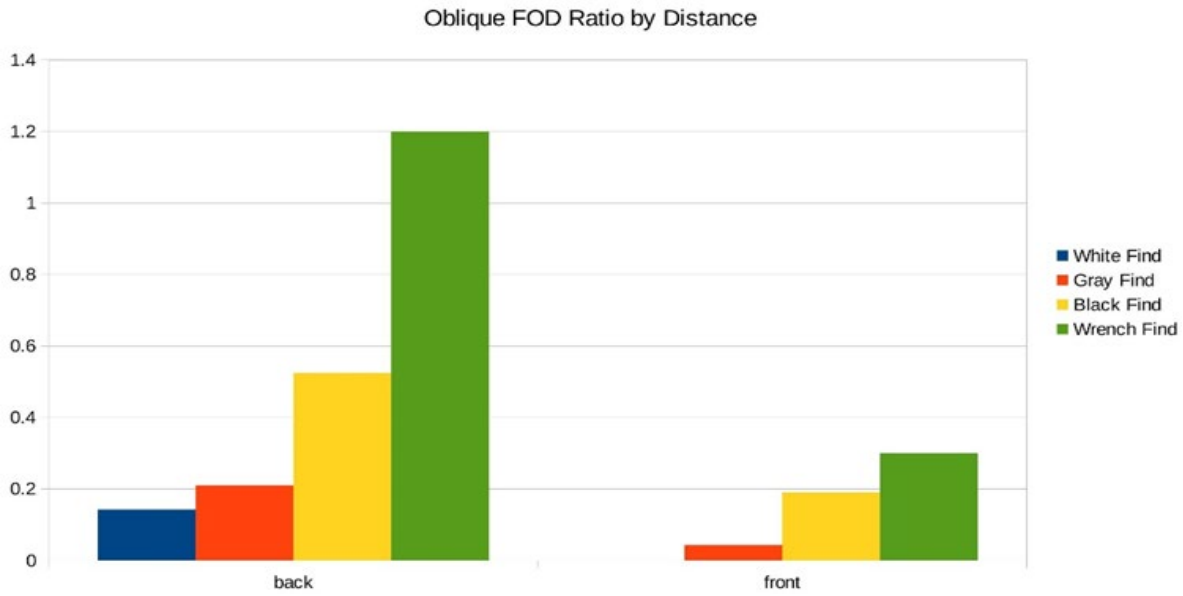

Figure 22. Ratio of Found FOD vs Missed FOD by Location in Oblique Image

As seen in Figure 22, the distance of the FOD from the camera played a significant role in the detection rate from oblique imagery. FOD located toward the front of the image are detected; however, FOD located toward the back of an oblique image are likely to be missed.

The effect of the background on which a piece of FOD appears was also analyzed. The results were organized based on the backdrop of each piece of FOD and were graphed as a ratio of hits

vs misses. The results from the oblique and nadir data can be seen in Figures 23 and 24, respectively.

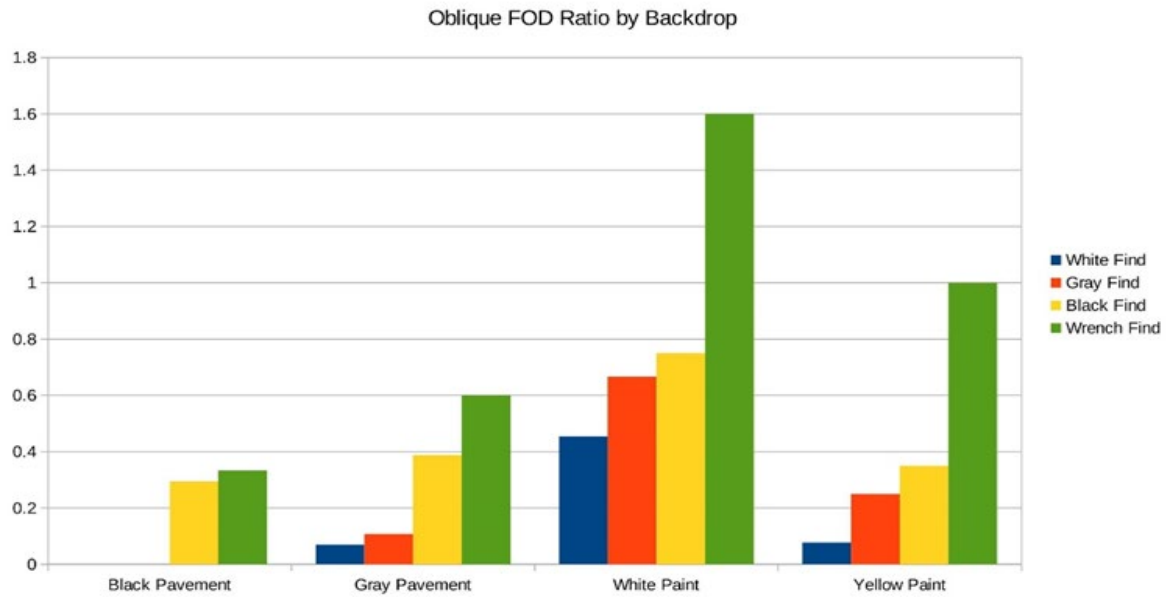

Figure 23. Results from Oblique Dataset

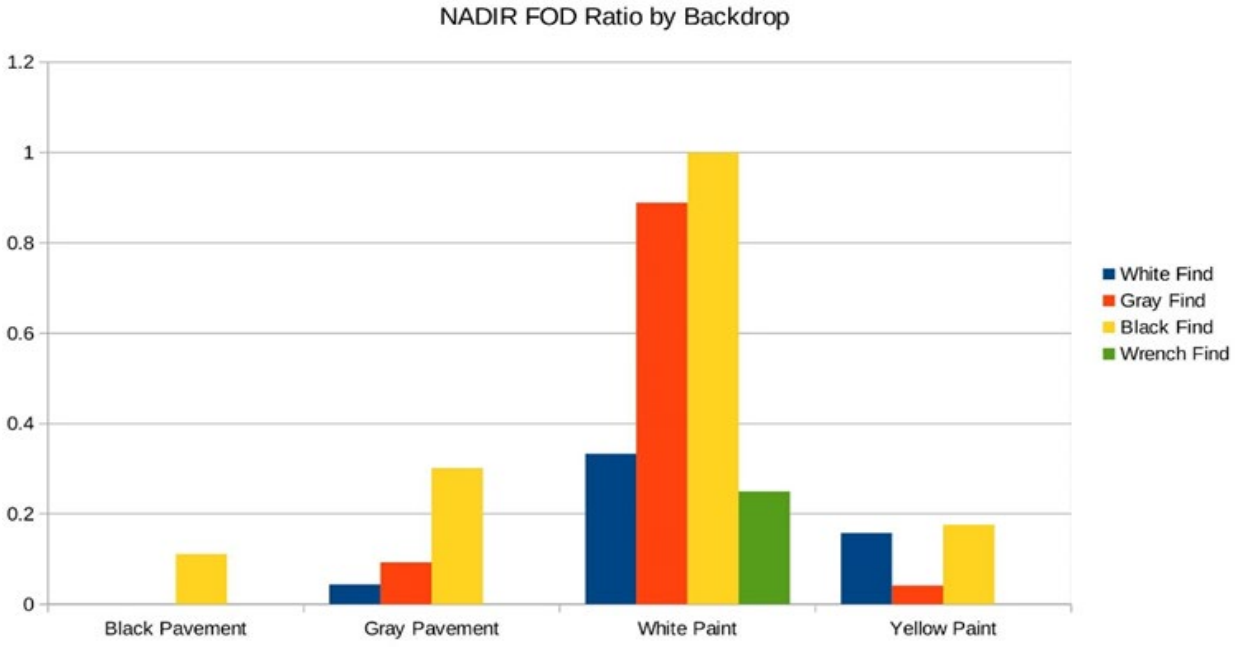

Figure 24. Results from Nadir Dataset

Figures 23 and 24 indicate that black pavement was the best backdrop against which FOD can be detected, while white paint was the worst.

False positives from each backdrop type were also included in this analysis, as shown in Table 9.

Table 9. False Positives by Picture Count

Oblique	False Positives by Picture Count
Black Pavement	10 of 36
Gray Pavement	117 of 293
White Paint	41 of 58
Yellow Paint	26 of 51
Nadir	
Black Pavement	7 of 23
Gray Pavement	56 of 237
White Paint	32 of 35
Yellow Paint	28 of 47

Table 9 indicates that the white paint (in Test Area 2), which was mostly chipped, produced a significant number of false positives.

4.1.7 Small-Scale Test Findings

Key takeaways from the analysis during small-scale testing were as follows:

- Multiple FOD test items inside a single cropped image block could lead to uncertainty in the performance of the ML algorithm.
- Results were significantly worse on paint than on pavement.
- White paint in worn condition had a high false positive rate.
- Results were worse on gray pavement than on black pavement.
- Oblique results were worse for the crops that were further from the camera, indicating that the distance to the camera is significant.
- The research team found that grass growing through cracks in the pavement impacted algorithm performance, in addition to different pavement types and markings producing false positives. Grass growth in pavement cracks was often the source of false positives and could not be easily filtered out by Global Positioning System (GPS) geofencing in the way that grass growing along the edges of the runway could be. Grass was minimized in the training dataset as the goal was to assess FOD detection on the runway surface instead of the grass surface.

4.2 CALIBRATION TESTING

The purpose of the Calibration Testing stage was to implement the lessons learned from the small-scale testing with revised test parameters for greater FOD detection accuracy. Coupling the use of both oblique and nadir orientation on a sUAS, test environments were used to capture data with and without FOD. The data without FOD was used to continue to build a library that helped train the AI/ML algorithms, while the data with FOD was used to assess the algorithms' performance.

The small-scale testing highlighted several areas in need of improvement with regard to data collection and analysis. Data collection improvements to be addressed included increasing sUAS data collection GSD. Assessment of FastFlow performance in the small-scale testing noted areas needing improvement, such as changing the test parameters. The changes would create new levels of accuracy in the test environment, ensuring FOD items can be surveyed and captured in a manner that provided the algorithms with absolute positive detection results.

4.2.1 Algorithm Training Requirements

Based on the performance and lessons learned from the initial dataset, the FastFlow algorithm was used again for the tests using the calibration dataset. As with the initial dataset collection, FastFlow was trained using a set of "clean" images that did not contain any FOD. During testing, images that contained FOD as well as images that were clear of FOD were shown to the trained network. The calibration dataset included additional categories of FOD in addition to the types used for the initial data collection.

4.2.2 Test Location

The research team continued the testing efforts at WWD for the calibration stage. The calibration testing included three test areas, with an emphasis on the following factors:

- Paint coloring amount
- Minimization of grass in imagery
- Higher resolution (decreased GSD) samples
- An increase in FOD variety

The flight areas were designed to accommodate high-level requirements for the training data established by the AI/ML subject matter expert (SME). The imagery was collected with an sUAS in compliance with the 14 CFR Part 107 regulations. The FOD item placements were focused on or around pavement markings, structures, and cracks to meet calibration testing goals.

4.2.2.1 Test Area 1

Test Area 1, identified as "Taxiway," is an asphalt taxiway located on the FAA ATR research Taxiway C at WWD. This operating area is approximately 1,100 ft long by 100 ft wide and includes asphalt pavement with yellow markings. Figure 25 illustrates this test area.

Figure 25. Calibration Testing—Test Area 1

4.2.2.2 Test Area 2

Test Area 2, identified as "Runway," includes the western (approach) end of Runway 10. This section of the runway is approximately 500 ft long by 150 ft wide and consists of asphalt with various white runway paint markings and rubber buildup. Figure 26 illustrates this test area.

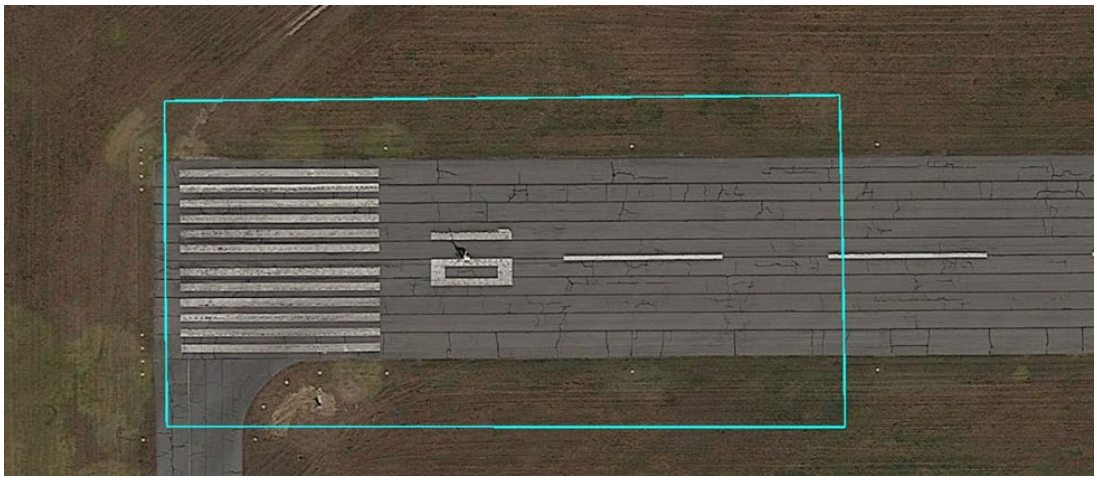

Figure 26. Calibration Testing—Test Area 2

4.2.2.3 Test Area 3

Test Area 3, identified as "FBO Ramp," is a ramp area on the southwestern portion of WWD consisting of asphalt and concrete. The flight area is approximately 720 ft long by 180 ft wide with aircraft tie-downs and yellow and black paint markings. Figure 27 illustrates this test area.

Figure 27. Calibration Testing—Test Area 3

4.2.3 FOD Selection

This section discusses the selection of FOD items for calibration testing and their placement within the selected test areas.

4.2.3.1 FOD Items

During the calibration test stage, the research team used test items similar to those implemented in the small-scale test stage and introduced new items for detection. However, during this stage, the golf ball string arrays were replaced with individual golf balls.

As learned during the small-scale test stage, multiple FOD items inside a single cropped image block could lead to uncertainty in the performance of the ML algorithm. To eliminate this uncertainty, the golf ball strings were disassembled, and a layout was designed that prevented any two test items from being present in the same image block.

Additional common FOD items, as defined in FAA AC 150/5220-24, were included as part of this testing. These additional items included nuts, bolts, fuel caps, and painted crescent wrenches. Adding these items gave the FastFlow algorithm increased opportunities to identify items of various shapes, sizes, and colors.

Table 10 summarizes the type and number of FOD items that were placed during testing.

Table 10. Type and Number of FOD Items

Type of FOD	Number of FOD Items
Golf Ball (White)	4
Golf Ball (Gray)	3
Golf Ball (Black)	12
Wrench	12
Nuts & Bolts	6
Sockets	6
Scrap Metal	6
Metal Pipe	6
Total	**55**
Item Totals by Test Area	
Runway	16
Taxiway	19
FBO Ramp	20

4.2.3.2 FOD Placement

Prior to the field effort, the research team designed a layout prioritizing the placement of test items on or near pavement markings, features such as lights and pavement cracks, with sufficient spacing between each item to ensure no test item appeared in more than a single image. Each item consisted of one of the types of FOD listed in Table 10 above. Additionally, to mitigate the risk of FOD being left on the airfield, the research team used the inventory and storage solution discussed in Section 3.5 of this report.

Test Area 1contained 19 FOD items, Test Area 2, contained 16 FOD items, and Test Area 3contained 20 FOD items, as shown in Figures 28, 29, and 30, respectively.

Figure 28. Calibration Testing FOD Placement—Test Area 1

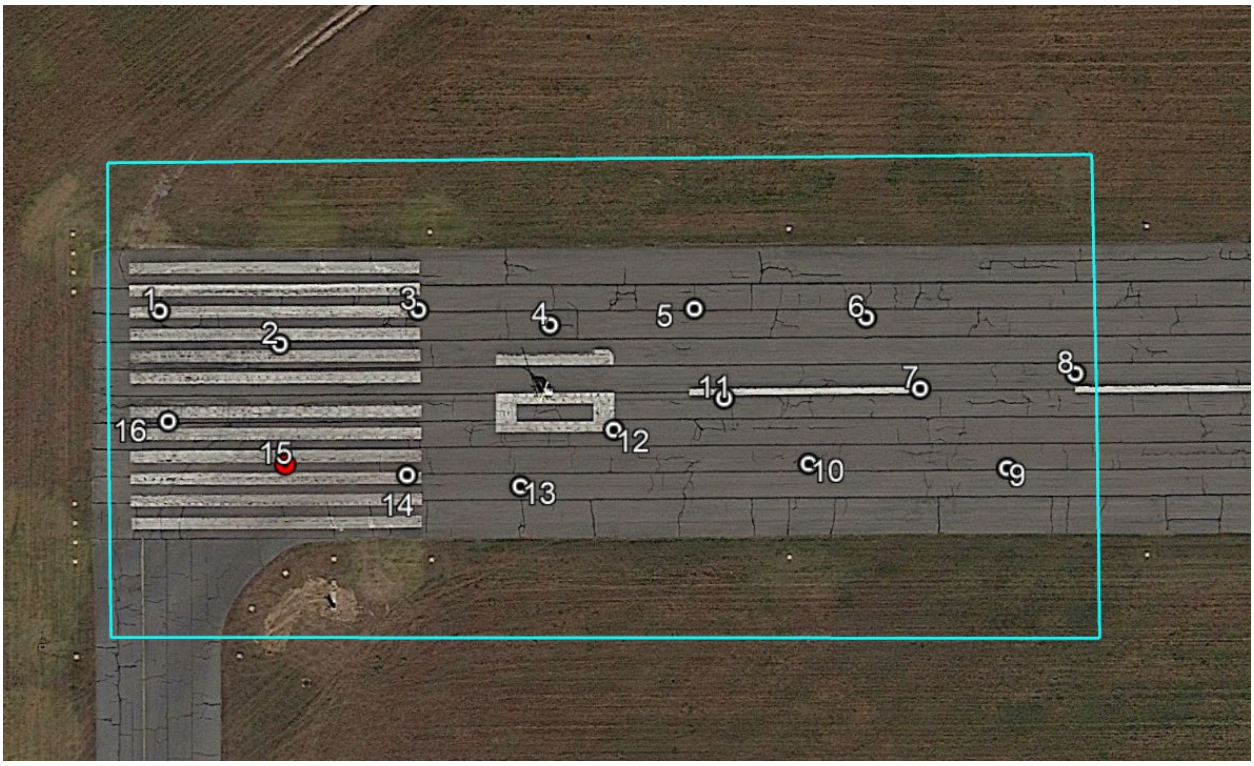

Figure 29. Calibration Testing FOD Placement—Test Area 2

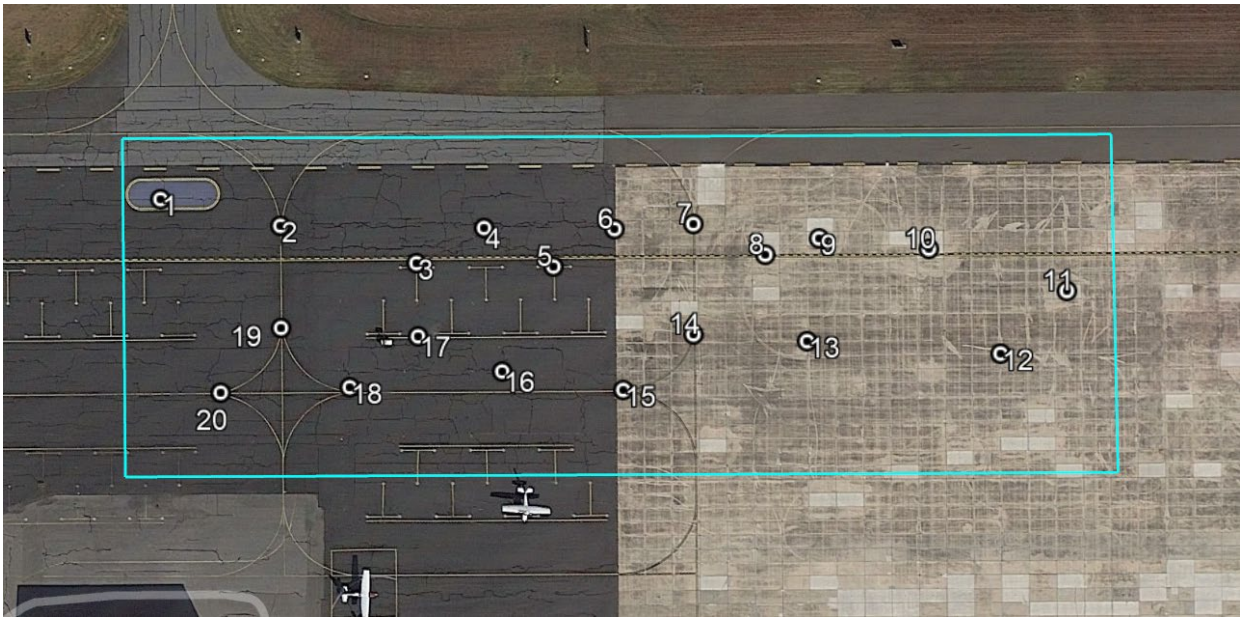

Figure 30. Calibration Testing FOD Placement—Test Area 3

4.2.4 sUAS Data Collection

The research team chose at this stage to switch solely to the DJI M300 RTK and the Zenmuse P1 sensor for continued calibration testing. Imagery was collected with and without selected FOD items for 1 week. At each of the three test areas, multiple datasets were collected with controlled and uncontrolled variables. Controlled variables included the airframe, sensor, pavement materials, and FOD items. Other variables, such as GSD and orientation of FOD items, varied.

With the lessons learned from the small-scale test stage, calibration testing focused on specific pavement aspects. To improve labeling efficiency, a flight plan was developed in which each FOD test item would be centered in its image. The flight plans contained the same number of images as there are FOD present per flight area. The flights were flown in a point-to-point waypoint format to capture the images and align the FOD items to be as centered as possible.

4.2.4.1 Flight Planning Characteristics and Parameters

The research team collected imagery in two different orientations: nadir and oblique. Oblique imagery was collected with the sensor at 22.5 degrees below parallel (a 67.5-degree angle above nadir) and rotated 90 degrees perpendicular to the flight path. The aircraft was programmed to fly with a lateral offset from the runway centerline as required by calculation of an image center GSD of approximately 0.4 cm. Figure 31 shows computer-aided design (CAD) drawings of data collection angles.

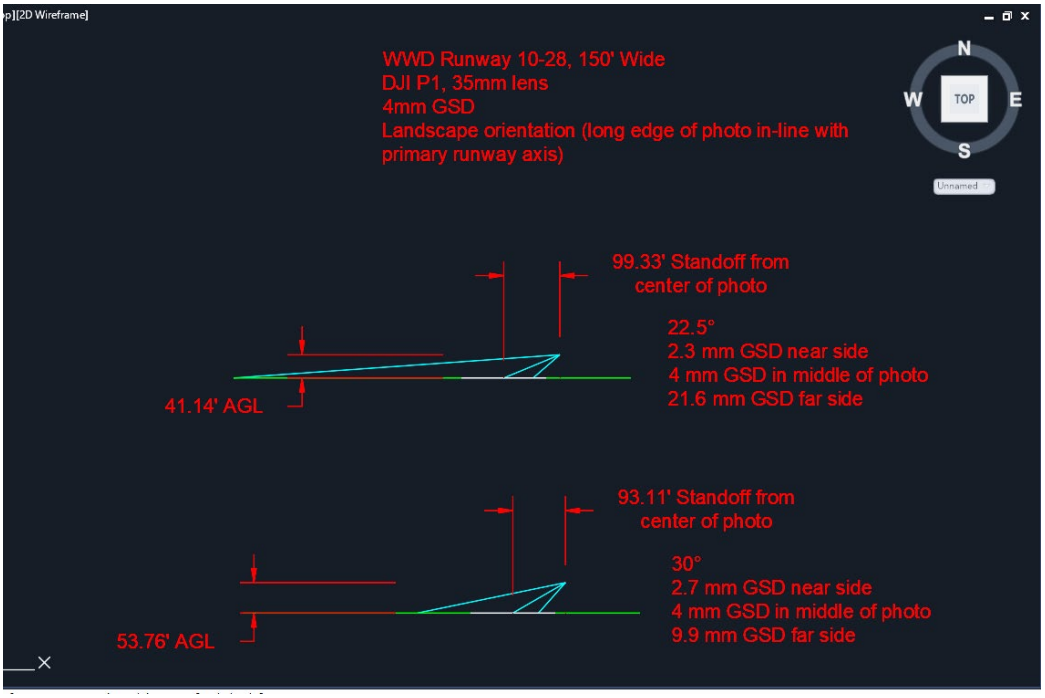

Figure 31. Data Collection Angles—CAD Drawing

In small-scale testing, only one type of nadir flight was conducted as part of determining a start point of the research effort. For calibration testing, two types of nadir flight collections were

conducted to assist in refining the path taken for future testing. The only difference between the flights of nadir high and nadir low were the altitude at which the data collection was conducted, with nadir high denoting the flight at 0.4 cm GSD and nadir low denoting a flight GSD of 0.2 cm. Nadir high had the same flight planning parameters as small-scale testing, while nadir low had the new lower flight altitude and improved GSD.

Table 11 summarizes the flight parameters for calibration testing.

Table 11. Flight Parameters—Calibration Testing

Flight Parameters	
Mission Planning Software	DJI Pilot 2
Data Format	RAW
sUAS Orientation	Nadir and Oblique
Overlap	N/A

4.2.4.2 sUAS Data Collection Workflow

The research team utilized the sUAS data collection workflow summarized in Section 3.6.

4.2.5 Data Processing

The data from the calibration collection effort was processed in the same manner as the initial collection effort but with two additional modifications. The first modification was to augment the dataset with additional sub-images that were created by rotating and/or mirroring the pixels. Image rotation cannot always be performed in datasets as, for example, an upside-down landscape would never be encountered by the algorithm during testing. The team experimented to determine how much augmentation with rotated images would benefit the algorithm's performance. In addition, where possible, both vertical and horizontal mirroring was performed to augment the final dataset.

The second modification was the choice to use two independent datasets, one of which went through a process of additional SD filtering. The SD filter curated the dataset to have a higher concentration of images with paint, cracks, and generally interesting features, while reducing the amount of "uninteresting" uniform pavement. The SD filter datasets were used to see what impact such a filter would have on the number of false positives.

4.2.6 Data Analysis

Quantitative results for each dataset were measured in the same manner, generating ROC curves and calculating AUCs. Throughout the small-scale testing, the research team used a confidence threshold of 60% to determine the baseline parameters of future testing. As part of the move to calibration testing, the confidence threshold was increased to 90%, and all the statistics in this section reflect that more stringent confidence level.

Figure 32 shows an example output of the SD filter applied to the imagery. The top row is the SD-filtered imagery, and the bottom row is imagery processed without the SD filter being applied to imagery.

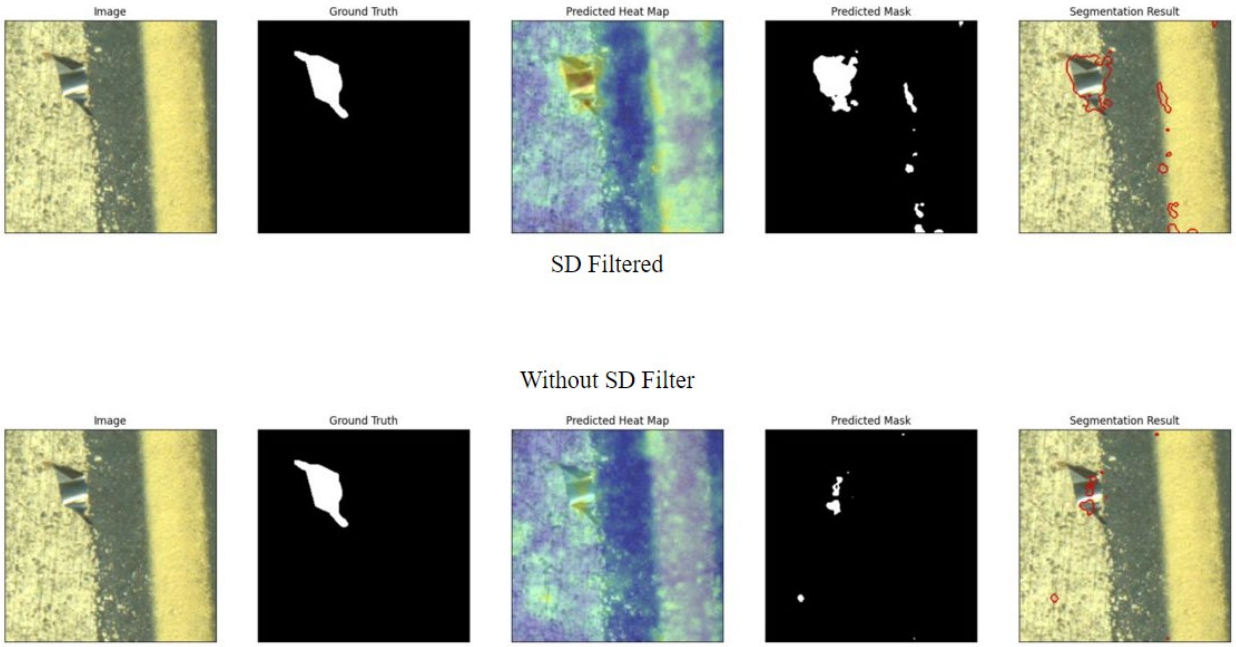

Figure 32. Example Output with SD Filter

Figure 33 shows the pixel-wise ROC curves for the nadir high dataset. The left image is for all crops and the right image is for the SD-filtered crops.

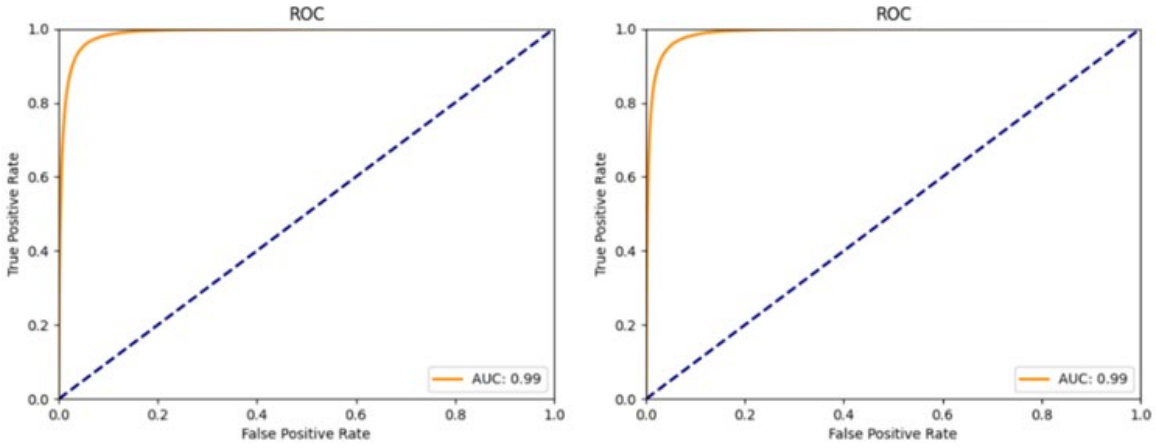

Figure 33. Pixel-Wise ROC, Nadir High

Figure 34 shows the pixel-wise ROC curves for the nadir low dataset. The left image is for all crops and the right image is for the SD-filtered crops.

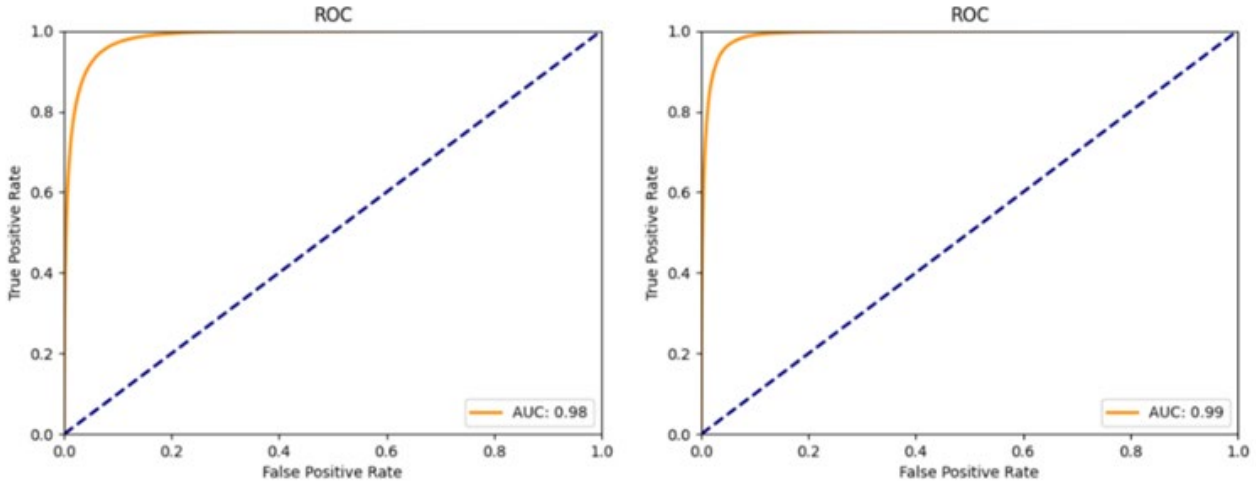

Figure 34. Pixel-Wise ROC, Nadir Low

Figure 35 shows the pixel-wise ROC curves for the oblique dataset. The left image is for all crops and the right image is for the SD-filtered crops.

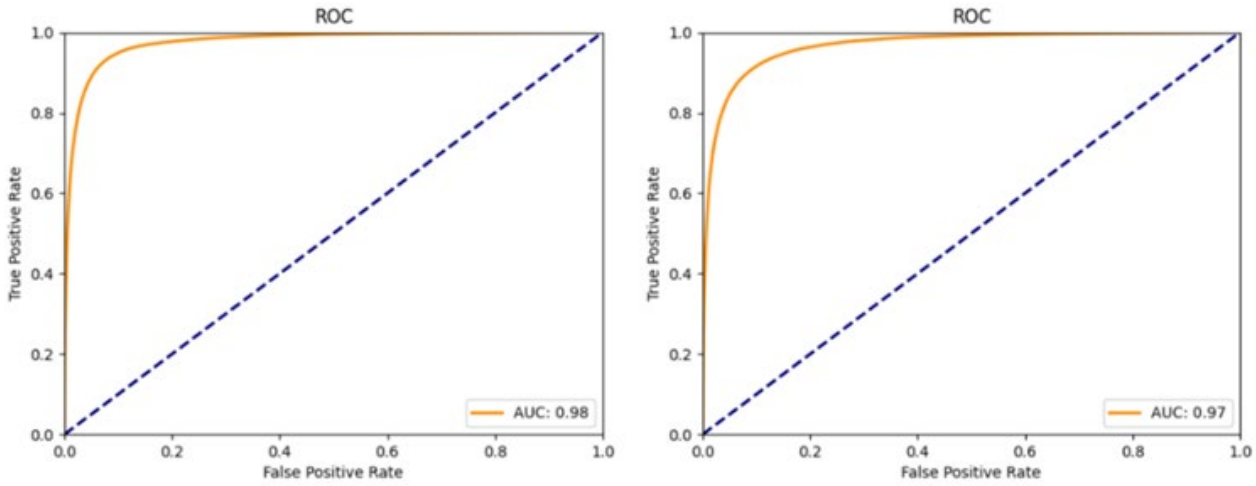

Figure 35. Pixel-Wise ROC, Oblique

Figure 36 shows the image-wise ROC curves for the nadir high dataset. The left image is for all crops and the right image is for the SD-filtered crops.

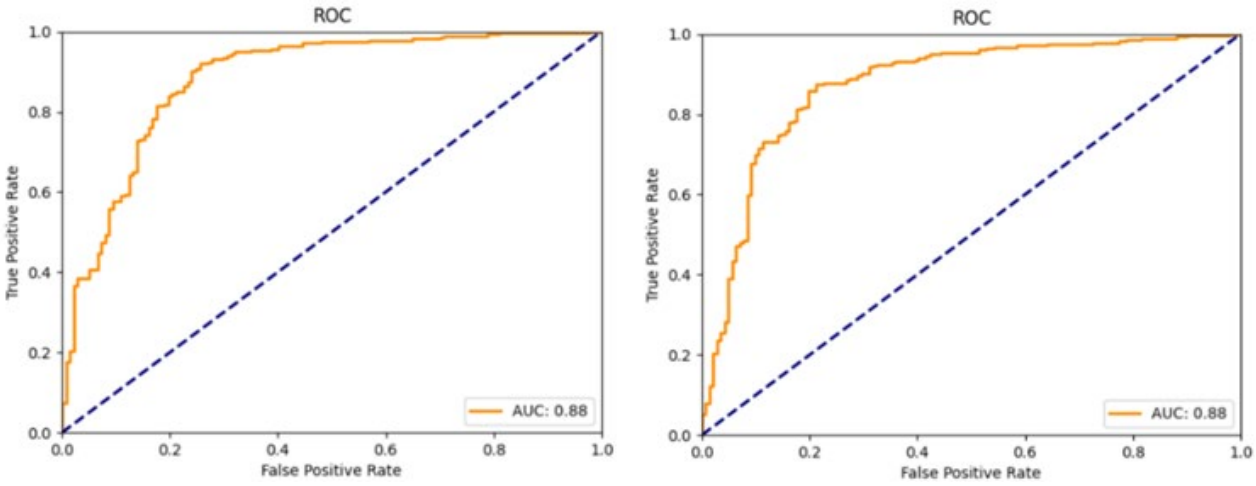

Figure 36. Image-Wise ROC, Nadir High

Figure 37 shows the image-wise ROC curves for the nadir low dataset. The left image is for all crops and the right image is for the SD-filtered crops.

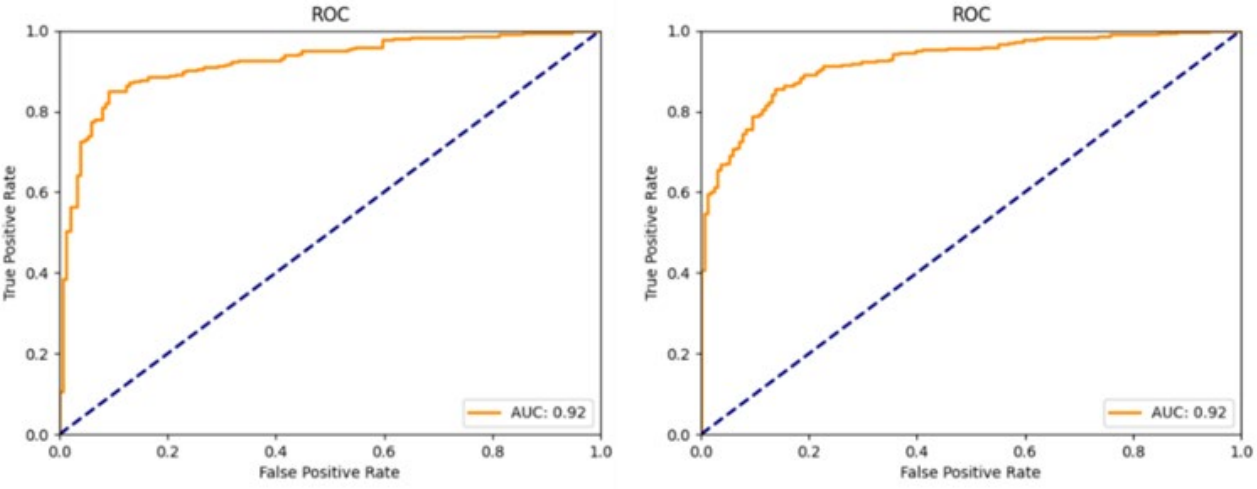

Figure 37. Image-Wise ROC, Nadir Low

Figure 38 shows the image-wise ROC curves for the oblique dataset. The left image is for all crops and the right image is for the SD-filtered crops.

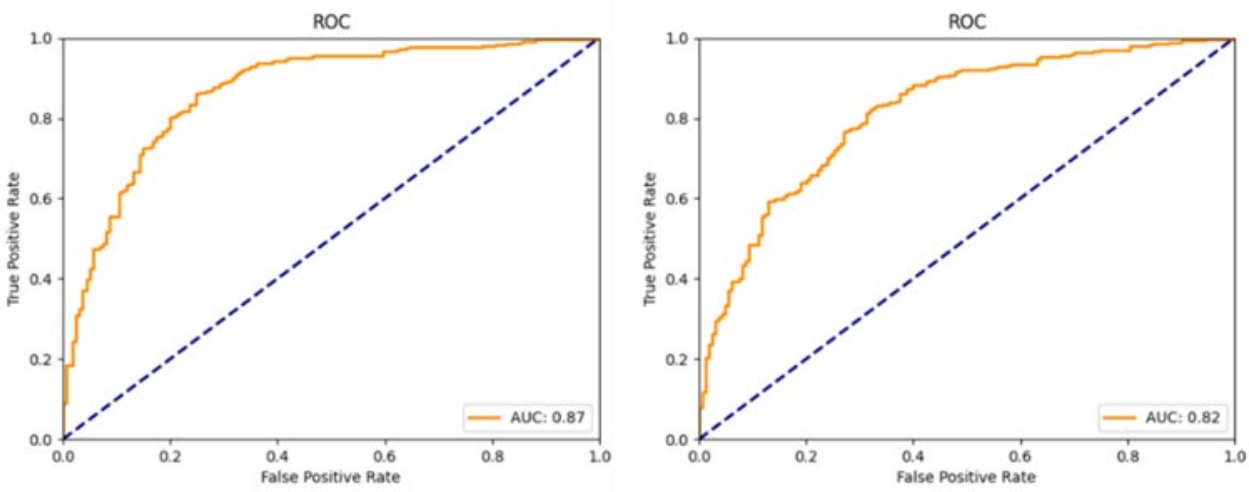

Figure 38. Image-Wise ROC, Oblique

Table 12 summarizes all AUCs for the datasets with the best result noted in green.

Table 12. Values of AUC

	Image AUC	Pixel AUC
Nadir High—All	0.89	0.99
Nadir Low—All	0.92	0.98
Oblique—All	0.87	0.98
Nadir High—SD filtered	0.88	0.99
Nadir Low—SD filtered	0.92	0.99
Oblique—SD filtered	0.82	0.97

Nadir Low—SD filtered had the best image and pixel AUCs with 706 false positives per image.

Table 13 presents the Nadir Low—SD filtered analysis results in relation to target identification.

Table 13. Calibration Testing Results

Dataset	90% Confidence Threshold	False Positive Crops Per Image	True Positives	False Negative	True Negatives	Total FOD	FOD Found	FOD Missed
Nadir Low—SD filtered	0.709	706	536	56	247	55	51	4

4.2.7 Calibration Testing Findings

The lessons learned include difficulties in identifying FOD with oblique imagery, and a higher margin of error in FOD detection at the 0.4-cm GSD. Recognizing these challenges, the research team decided not to collect oblique imagery for the full-scale test stage. Additional routes to

research oblique orientation imagery are present; however, budget constraints and insufficient sensor quality options warrant a recommendation for future testing.

Finally, it was noted that FastFlow did not display additional challenges with the new FOD types added in this data collection effort. The recommendation from these results was to pursue capturing nadir images at low altitude and to continue the SD filtering preprocessing step implementation to train and test the FastFlow network.

4.3 FULL-SCALE TESTING

The full-scale test stage implemented lessons learned during initial and calibration testing and provided the AI/ML algorithms with a full-scale dataset. The research team collected imagery over the entire test area runway, implementing standard items and an array of common FOD items listed in AC 150/5220-24. This testing refined test parameters and accounted for real-world scenarios, such as larger test areas and a suite of actual FOD items. This dataset provided a comprehensive collection of FOD items and images with paint, cracks, and other pavement markings, all captured at a GSD of 0.2 cm in nadir orientation.

4.3.1 Algorithm Requirements

In addition to implementing lessons learned, this effort evaluated FastFlow's FOD detection capability in a full-scale data collection effort.

4.3.2 Test Area

Full-scale testing consisted of one test area at WWD as requested by the FAA, with an emphasis on the following factors:

- Paint coloring amount
- Minimization of grass in imagery
- Higher resolution (decreased GSD) samples
- An increase in FOD variety

The imagery was collected with an sUAS in compliance with the 14 CFR Part 107 regulations carrying a RGB sensor. The FOD item positions were focused on or around pavement markings, structures, and cracks to meet full-scale test goals. The sUAS followed a preprogrammed path to ensure that the collected imagery captured all FOD items placed with approximately 25 ft between items.

4.3.2.1 Test Area

The test area identified as "Runway" was the full runway of 10/28 at WWD, illustrated in Figure 39. This runway is approximately 5,000 ft long by 150 ft wide and consists of a mix of freshly rehabilitated and older asphalt with various white runway paint markings and rubber buildup.

Figure 39. Full-Scale Testing—Test Area

4.3.3 FOD Selection

This section discusses the selection of FOD items for full-scale testing and their placement within the selected test areas.

4.3.3.1 FOD Items

The research team chose the standardized items used in the prior initial and calibration testing in addition to the test items laid out in AC 150/5220-24. Several additional FOD items from AC 15/5220-24 were added into the full-scale test stage, including PVC pipe, hydraulic hose, gas cap, tire rubber, taxi light, and asphalt chunk. Additionally, to mitigate the risk of FOD being left on the airfield, the research team utilized the inventory and storage solution as discussed in Section 3.5 of this report.

Table 14 summarizes the type and number of FOD items that were placed during testing.

Table 14. Type and Number of FOD Items

Type of FOD	Number of FOD Items
Golf Ball (White)	5
Golf Ball (Gray)	4
Golf Ball (Black)	4
Wrench	7
Nuts & Bolts	3
Sockets	3
Scrap Metal	3
Metal Pipe	3
PVC Pipe	3
Hydraulic Hoses	3
Gas Caps	3
Tire Rubber	3
Taxi Lights	3
Asphalt Chunks	3
Total	**50**
Item Totals by Test Area	
Runway	50

4.3.3.2 FOD Placement

Prior to the field effort, the research team designed a layout prioritizing the placement of test items on or near pavement markings, features such as lights, and pavement cracks, with sufficient space between each item to ensure no test item overlaps. The flight area, as shown in Figure 40, had two FOD zones: Runway 10 End (FOD items 1–25) and Runway 28 End (FOD items 26–50), with a size of 100 ft x 100 ft and 25 items laid out on each in a grid format.

Figure 40. Full-Scale Testing—Test Zones

Each area consisted of at least one of the following: a golf ball colored white, gray, or black; fuel caps; hydraulic hoses; taxi lights; wrenches; PVC pipes; rubber scraps; sockets; nuts and bolts; asphalt chunks; metal pipes; and metal scraps. The FOD placement in the testing zone is shown in Figures 41 and 42.

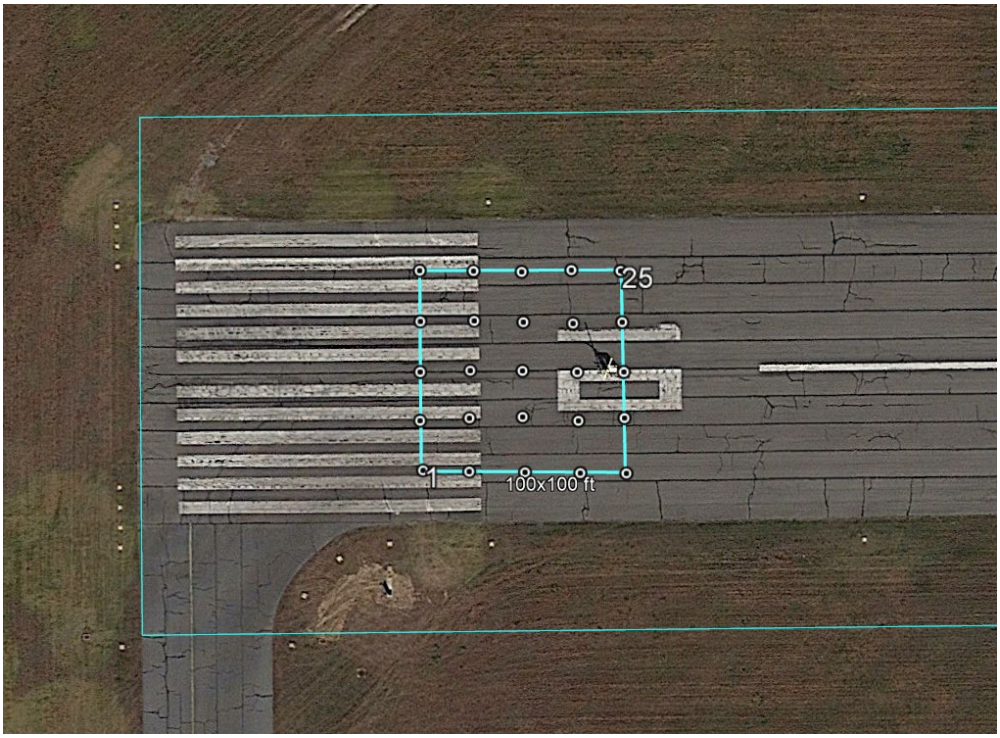

Figure 41. Full-Scale Testing FOD Placement—Runway 10 End

Figure 42. Full-Scale Testing FOD Placement—Runway 28 End

4.3.4 sUAS Data Collection

The research team collected RGB data with FOD test items for 1 week. Two datasets were collected with variables such as pavement material, item type, and the presence or lack of test items. Additionally, the DJI M300 RTK and P1 sensor were chosen for the full-scale testing. The continued implementation of the DJI M300 and P1 during full-scale testing was due to its performance during calibration testing.

The use of ground control points during data collection was not necessary; however, to analyze detection accuracy, the team surveyed the locations of test items placed within the test area. To meet the full-scale testing objectives, images of the FOD were collected in a traditional "lawnmower" pattern. With no requirement to produce an orthophoto, overlap requirements remained low. The research team used a forward overlap of 10% and side overlap of 15% to ensure the complete collection of the flight area. The set overlap fulfilled collection requirements while minimizing flight time and excess data.

4.3.4.1 sUAS Data Collection Workflow

The research team used the sUAS data collection workflow identified during small-scale testing in Section 4.2.4.

4.3.4.2 Flight Planning Characteristics and Parameters

Nadir imagery was collected in the most efficient format, with the aircraft flying parallel to the center line of the runway and the widest portion of the sensor positioned perpendicular to the center line. The data output from this sUAS data collection effort was standalone imagery rather than an orthomosaic output.

Table 15 summarizes the flight parameters for full-scale testing.

Table 15. Flight Parameters—Full-Scale Testing

Flight Parameters	
Mission Planning Software	DJI Pilot 2
Data Format	RAW
sUAS Orientation	Nadir
Overlap	10/15

4.3.5 Data Processing

Data processing for the full-scale dataset followed nearly the same protocol as the processing for the calibration data. However, since this collection was only for the test data, the data used for training was the Nadir Low – SD Filtered set from the calibration dataset. This was chosen since it had the best AUC numbers for both FOD images and FOD pixels when tested with the calibration imagery. The method for generating image crops from the raw imagery was different than with the previous tests. Because the data collected in the full-scale set was mainly focused

on testing and not training, a cropping method was selected that ensured each piece of FOD would appear towards the center of at least one crop.

In the previous datasets, FOD was sometimes missed if it was only found on the edge of an image as shown in Figure 43. In the left image, a small portion of a piece of FOD is highlighted by a red circle. The middle image shows the ground truth label for that FOD in the image. The right image shows the output of Fastflow where that piece of FOD was missed and included a false positive at the lower right.

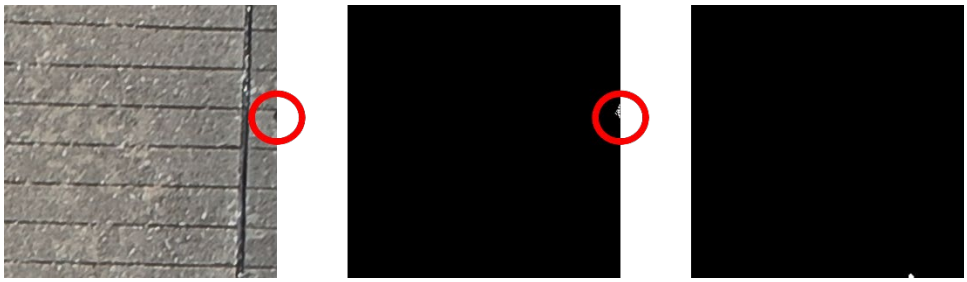

Figure 43. Example Image Crop Where Only a Small Portion of FOD is Visible

For this purpose, the 256 x 256-pixel crops were simply offset by 128 pixels in an evenly distributed grid as shown in Figure 44. After all the crops were generated, a final row and column of crops was taken to capture the final pixels at the edges of the image. This ended up producing about 300 image crops that contained FOD and 3.29 million crops which did not. Due to memory constraints, the non-FOD imagery was randomly subsampled to reduce the number of crops, bringing the total count closer to the number of FOD images. This resulted in a total of 329 non-FOD crops being processed through the FastFlow setup.

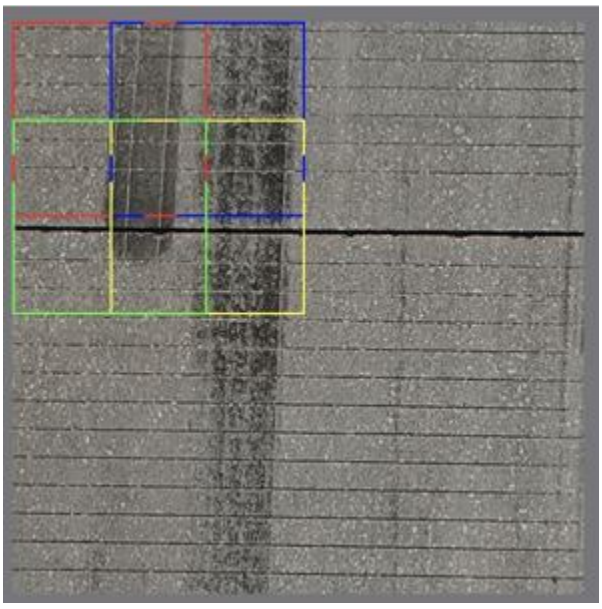

Figure 44. Visualization of Test Cropping Method

4.3.6 Results and Analysis

Quantitative results for full-scale testing showed similar values to other tests when training on the calibration testing Nadir Low – SD Filtered dataset and testing on the images from the new data collection.

ROC Curves are shown in Figure 45. The left image is pixel-wise ROC with an AUC of 0.98, and the right image is image-wise ROC with an AUC of 0.90.

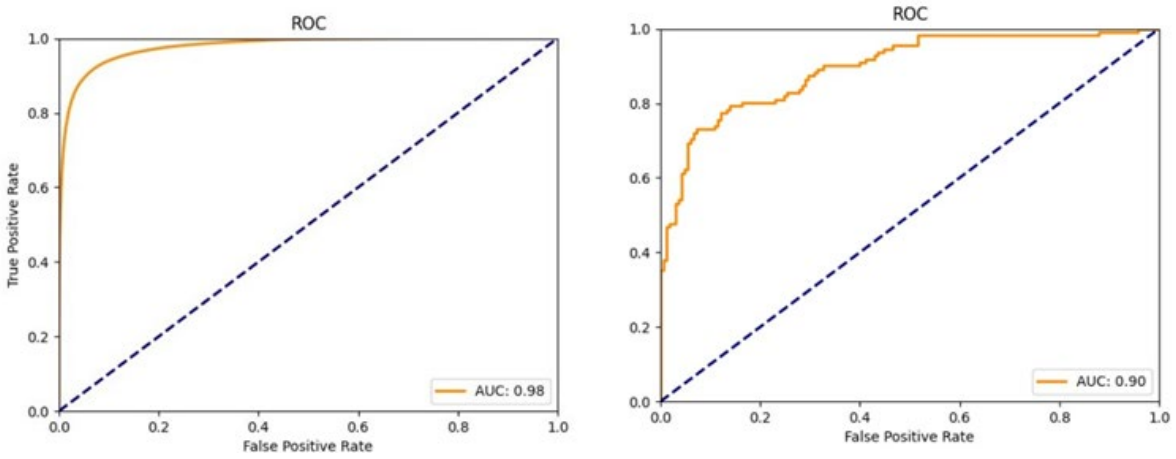

Figure 45. Pixel-Wise vs Image-Wise ROC

To generate scores for this dataset, the crop where the FOD was most centered was chosen to be the only one that would be scored. As with the other data collection efforts, a true positive was indicated if even a single pixel overlapped with the FOD. Additionally, the team analyzed this dataset at the 60% confidence threshold, and at the 90% confidence threshold required by AC 150/5220-24 in order to compare the differences.

In the case of the 60% confidence threshold, the true positive results for full-scale testing showed five missed pieces of FOD over the 53 pieces that were captured in the imagery. The three extra pieces of FOD were from FOD that were in more than one picture. All five of these missed FOD were in difficult situations:

- Two were asphalt chunks that were the same color as the asphalt they were sitting on.
- Two were a black gas cap and a black golf ball on black pavement.
- One was a rubber strip that looked like the tar used to repair cracks, which FastFlow learned not to be counted as anomalous.

There were also a significant number of false positives throughout the imagery. On plain pavement images there were often scatterings of small clumps of pixels that surpassed the confidence threshold to be counted as anomalous. Additionally, as seen in previous tasks, false positives occurred primarily in images that contained cracks, paint, and grass. Examples can be seen in Figures 46–50.

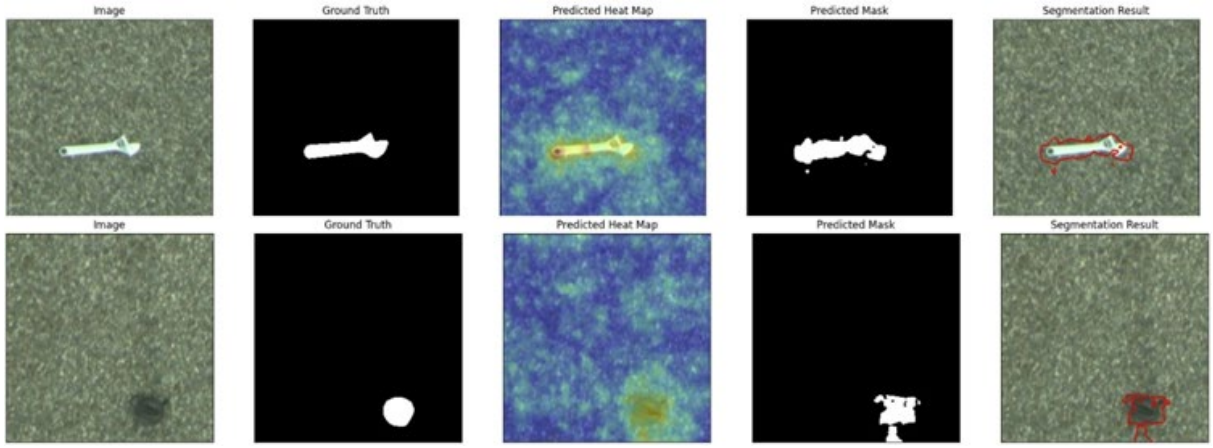

Figure 46. True Positives

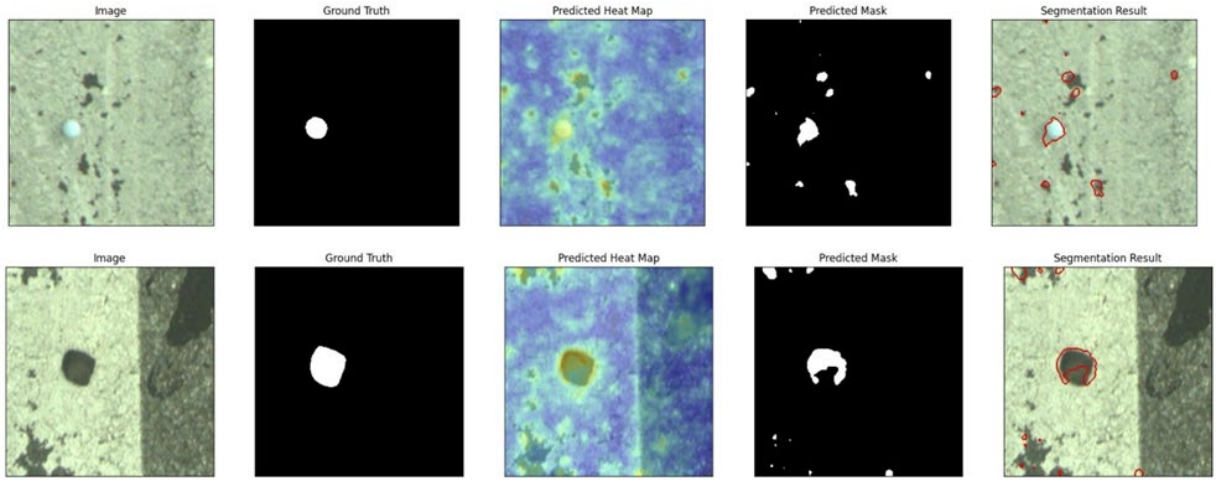

Figure 47. True Positives with False Positives

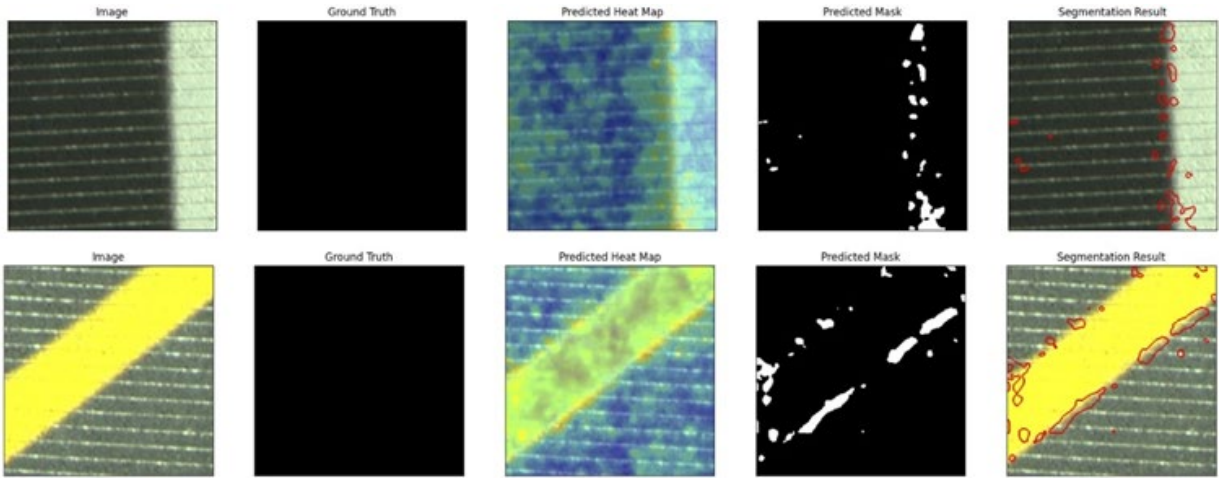

Figure 48. False Positives

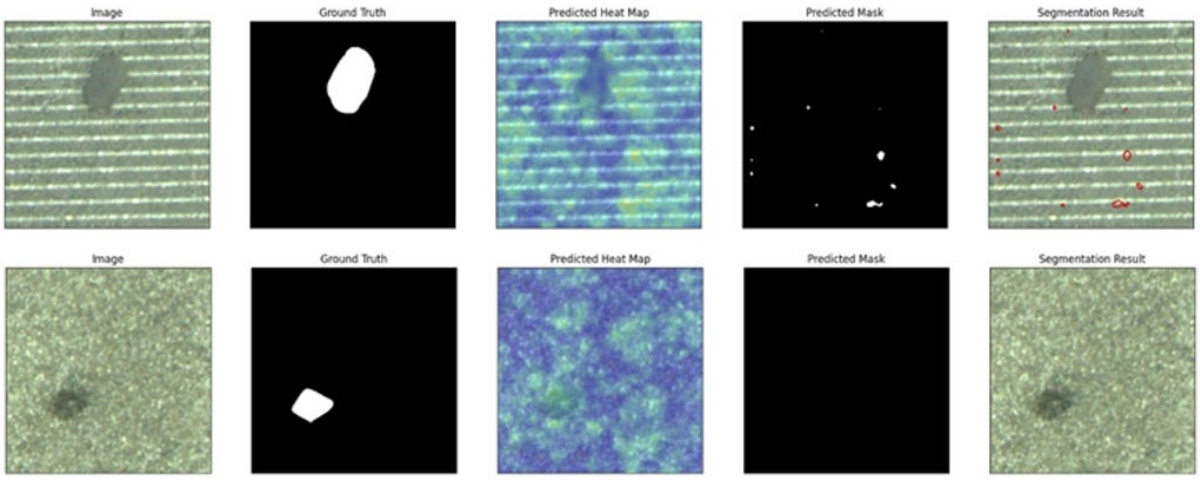

Figure 49. False Negatives

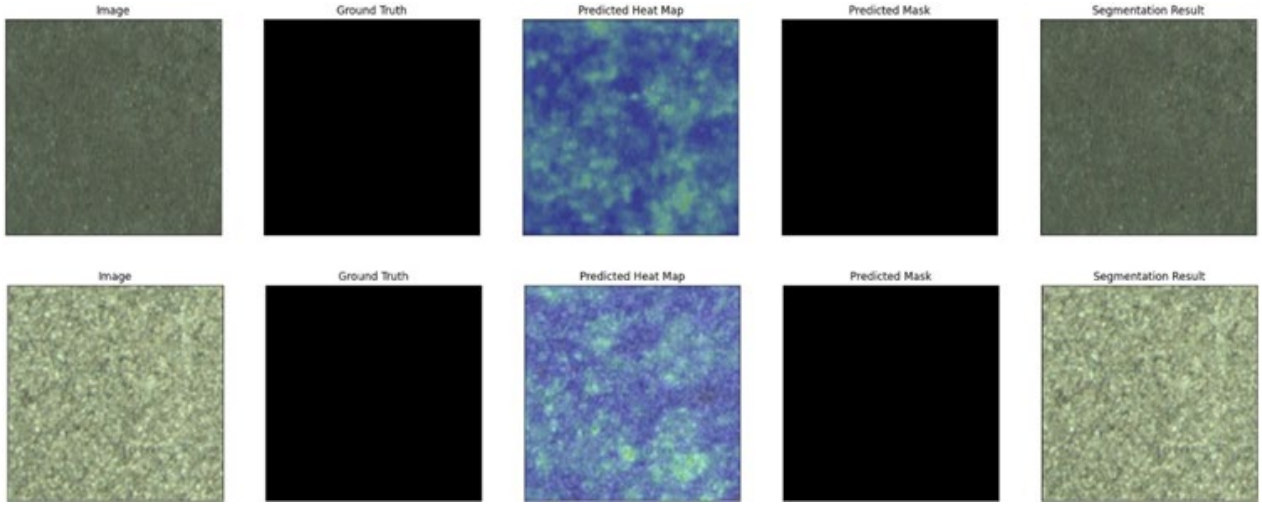

Figure 50. True Negatives

In the case of the 90% confidence threshold, the true positive results for full-scale testing showed a miss of one FOD target out of the 53. The missed FOD happened to be the small V-shaped scrap metal target. This marks a notable difference between the 60% and 90% thresholds in what is more difficult to detect target wise, with 60% being color contrast and 90% being object size. This dataset also had a total of 1,737 false positive crops per image, where false positives appeared around paint, markings, and cracks.

4.3.7 Full-Scale Test Findings

The lessons learned include the viability and successful FOD detection with AI/ML of all types of FOD outlined in AC 150/5220-24. It also presented challenge areas for the AI/ML algorithms in the positive identification of FOD items that were of small sizes or closely matched the visual appearance of the pavement surface they were placed on. The items of specific difficulty included nuts and bolts, sockets, asphalt chunks, and rubber strips that matched the pavement

surface color. Paths to increase positive identification of challenging FOD items are outlined in Section 7.

Finally, it was noted that FastFlow had a high number of false positives focused on certain visual features, such as chipped paint, cracks, or tar patches. The high false positive rate is believed to be directly linked to an inadequate number of training images containing such features.

4.4 INITIAL TESTING SUMMARY

The FastFlow anomaly detection algorithm, originally designed for detecting defects in manufactured parts, was successfully applied to the challenge of detecting FOD in runway images taken from an sUAS. The research team found that FOD items that differ in appearance from the background pavement on which they are placed were nearly always found by the FastFlow algorithm. In contrast, FOD items with an appearance that very closely matched the pavement on which they were placed were more difficult to detect and were mostly missed by the algorithm when using a 60% confidence threshold. The algorithm is reliant upon the capabilities of the sUAS's camera to generate imagery, which can mean that over- or under-saturated sections of images and areas of low contrast could cause the algorithm to potentially miss pieces of FOD.

Another challenge revealed by the testing is the high likelihood of false positives (e.g., false FOD reports) being reported by the algorithm. While the FastFlow models were trained using images taken from the runway, certain visual features, such as chipped paint, cracks, or tar patches, did not appear in large quantities. As a result, the algorithm appeared not to learn that these features were not FOD and would still flag some of them as FOD. Thus, the default FastFlow algorithm still requires very careful curation of images for training, and additional development and testing would be required to reduce false positives while still maintaining a high true positive detection rate.

Finally, the testing led to the identification of best methods for data collection. Images captured from a nadir vantage point had a significantly higher true positive detection rate than images captured from an oblique vantage point. Positive detection rate was also improved by flying the sUAS lower to increase the image resolution. The full-scale testing was conducted using the low-altitude nadir flights, so the final detection results detailed in Tables 16 and 17 include both those results and the results from the calibration test, which was flown in the same fashion. No results from the initial test were included because they were all captured at a higher altitude and would not be comparable to the subsequent tests.

Table 16 summarizes the initial testing Fastflow results at WWD.

Table 16. Initial Testing Results

Test Stage	90% Confidence Threshold	False Positive Crops Per Image	True Positives	False Negative	True Negatives	Total FOD	FOD Found	FOD Missed
Calibration	0.709	706	536	56	247	55	51	4
Full-Scale	0.623	1737	272	28	113	50	49	1

Table 17 summarizes the overall AC 150/5220-24 FOD target detection results of initial testing at WWD. The table breaks down the difference in target detections between the 60% and 90% confidence thresholds.

Table 17. FOD Detection Confidence Threshold Comparison

FOD Target	sUAS/AI Detection Rate at 90% Confidence Threshold	sUAS/AI Detection Rate at 60% Confidence Threshold
An unpainted metal cylinder	100% (9 of 9)	100% (9 of 9)
A white, gray, or black sphere	100% (32 of 32)	100% (32 of 32)
A “chunk” of asphalt or concrete	100% (3 of 3)	33% (1 of 3)
Any portion of a runway light fixture (in-pavement or edge light)	100% (3 of 3)	100% (3 of 3)
An adjustable crescent wrench	100% (19 of 19)	100% (19 of 19)
A deep socket	89% (8 of 9)	100% (9 of 9)
A piece of rubber from an aircraft tire	100% (3 of 3)	67% (2 of 3)
A distorted metal strip	78% (7 of 9)	100% (9 of 9)
A fuel cap	100% (3 of 3)	67% (2 of 3)
A lug nut	78% (7 of 9)	100% (9 of 9)
A hydraulic line	100% (3 of 3)	100% (3 of 3)
A white PVC pipe	100% (3 of 3)	67% (2 of 3)
Total	95 % (100 of 105)	95 % (100 of 105)

Based on the results of the initial testing at WWD, the research team implemented a Conditional Random Field (CRF) algorithm to reduce the number of false positives. These solutions tackled the false positive rate problem from the standpoint of post-processing the output and carefully controlling what input data are used to train the model. The CRF algorithms make use of reported region size, region probability, and the RGB values in the input image of the region to remove instances of low-probability FOD while promoting higher-probability FOD.

To address the FAA AC 150/5220-24 location accuracy requirements, the research team implemented a proof-of-concept FOD geolocation workflow and benchmarked its performance against surveyed positions of the placed FOD test items.

5. VALIDATION TESTING: ATLANTIC CITY INTERNATIONAL AIRPORT

Validation testing at ACY was aimed at implementing lessons learned from initial testing at WWD. During the full-scale testing at Cape May, tests were conducted over a full runway with refined testing parameters to evaluate the training and learning the algorithms had undergone in a real-world case. Validation testing provided datasets at a new airport (ACY) to help validate the performance and demonstrate the ability to generalize the processes and workflows previously developed, which included evaluating the performance of CRF filtering methods to reduce false positive rates while maintaining true positive performance. Additionally, the validation testing sought to demonstrate the minimum viable product requirements for using sUAS and AI as a means of FOD detection per AC 150/5220-24.

5.1 TEST AREAS

ACY was chosen as the location for this sUAS-based FOD detection validation testing due to its close proximity to the William J. Hughes Technical Center (WJHTC). ACY offers an appropriate amount of diversity for testing conditions. The specific testing site at ACY was chosen to encompass as much runway area as possible while not interfering with normal airport operations taking place during the time of testing.

The research team selected two test sections of Runway 04/22 at ACY, with an emphasis on a novel location and runway material while staying outside the intersecting area of Runway 13/31 and Taxiway B, thereby minimizing impact on the airport during testing. The sUAS performed flights parallel to the length of the operating area. These test areas also met the algorithm training requirements discussed in previous sections of this report.

The validation testing effort had two separate flight operations in these areas on Runway 04/22. These flights correspond to the two stages where the first stage was the collection of training data over the “clean” novel runway, and the second was a collection of testing data over the same areas with FOD present.

The flight area identified as “RWY 04/22” was made up of two parts of Runway 04/22 and is shown in Figure 51. This runway is 6,144 ft long by 150 ft wide and consists of a mix of freshly rehabilitated and older asphalt with various white runway paint markings and rubber buildup. There is an additional 75-ft buffer space to the pavement as defined by the required airspace authorization.

Figure 51. Test Areas at ACY

The flight area identified as “Test Area 1” is a portion of Runway 04/22 on the 04-approach side of the runway and is shown in Figure 52. This runway section is approximately 3,300 ft long by 150 ft wide, with an additional buffer area of 75 ft to either side of the runway.

Figure 52. Test Area 1 at ACY

The flight area identified as “Test Area 2” is a portion of Runway 04/22 on the approach side of Runway 22 and is shown in Figure 53. This area is approximately 2,200 ft long by 150 ft wide with a buffer area of 75 ft to either side of the runway.

Figure 53. Test Area 2 at ACY

5.2 DATA COLLECTION

The research team collected nadir RGB data from a DJI M300 RTK and P1 sensor with and without selected FOD items during the appropriate test stage. At each of the test areas, multiple datasets were collected with controlled and uncontrolled variables. Controlled variables included the use of airframe, sensor, and FOD items and their placement location, while other variables, such as sun angle and orientation of FOD items, are uncontrolled.

The use of ground control points during data collection was not necessary; however, to analyze detection accuracy, the team surveyed the locations of test items placed within the test area. To meet the validation testing objectives, images of the FOD were collected in a traditional "lawnmower" pattern. With no requirement to produce an orthophoto, overlap requirements remained low and comparable with full-scale testing. The research team determined the use of a forward overlap of 10% and side overlap of 15% for the training stage to ensure ample collection of flight areas. The team determined a forward overlap of 10% and side overlap of 10% for the

test stage would meet collection requirements while minimizing flight time and excess or duplicate data.
Nadir imagery was collected in the most efficient format, with the aircraft flying parallel to the center line of the runway and the widest portion of the sensor positioned perpendicular to the center line. The validation testing datasets focused on maintaining FOD objects of increased variety and imagery containing paint, cracks, and other pavement markings at the 0.2-cm GSD. Table 18 summarizes the flight plan for full-scale testing.

Table 18. sUAS Flight Parameters – ACY Training & Testing Stages

sUAS Flight Parameters	
Mission Planning Software	DJI Pilot 2
Data Format	RAW
sUAS Orientation	Nadir
Overlap (Training Stage)	10/15
Overlap (Testing Stage)	10/10
Ground Sample Distance	0.2 cm

5.3 TRAINING DATA

The research team collected imagery over the two test areas without FOD present in the imagery to provide the FastFlow algorithm a "clean" dataset to compare FOD imagery against. This testing effort was designed to build out the Fastflow training library of the test site by implementing baseline FOD-free training data at ACY. The training dataset follows a nearly identical structure to what was done in initial testing. The primary differences from the initial testing were that the dataset was collected at a 0.2-cm GSD from the start and included grass along the edge of the runway surface.

The training dataset processing followed the same processes as the WWD full-scale testing until the point of running the FastFlow network. There is no significant analysis for this stage as this data collection was focused on presenting the FastFlow algorithm training data that could be compared to the testing data in the next stage.

5.4 TESTING DATA

The validation testing dataset implemented lessons learned during the WWD full-scale testing and provided the AI/ML algorithms with a new FOD dataset. This dataset provided a comprehensive collection of FOD items and images with paint, cracks, and other pavement markings, all captured at a GSD of 0.2 cm in nadir orientation. During validation testing, the research team also focused on improving the detection performance of FastFlow by implementing a new image post-processing workflow that incorporates CRF filtering.

5.4.1 FOD Selection

This section discusses the selection of FOD items for validation testing and their placement within the selected test areas.

5.4.1.1 FOD Items

The research team decided to incorporate standardized items from previous performance assessments into validation testing. In addition to the standardized items, the research team picked an array of test items out of AC 150/5220-24.To mitigate the risk of FOD being left on the airfield, the research team used the inventory and storage solution discussed in Section 3.2 of this report.

Table 19 summarizes the type and number of FOD items that were placed during testing.

Table 19. Type and Number of FOD Items

Type of FOD	Number of FOD Items
Golf Ball (White)	4
Golf Ball (Gray)	4
Golf Ball (Black)	4
Wrench	8
Nuts and Bolts	6
Sockets	6
Scrap Metal	6
Metal Pipe	6
PVC Pipe	6
Hydraulic Hoses	6
Gas Caps	6
Tire Rubber	6
Taxi Lights	6
Asphalt Chunks	6
Total	**80**
Item Totals by Test Area	
Test Area 1	40
Test Area 2	40

5.4.1.2 FOD Placement

Prior to data collection, the team surveyed primary characteristics of the test areas such as the pavement type, condition, markings, presence of features such as elevated/flush lights, and rubber buildup. The team identified notable attributes and placed test items across each attribute with enough space between them to prevent any overlaps. Each FOD item consisted of either a golf ball or an item identified in AC 150/5220-24.

The following figures display the defined areas where FOD was located during the validation testing effort. The FOD zones did not have FOD in a predefined placement orientation. Each zone within the two test areas contained 20 pieces of FOD, for a total of 80 FOD in the testing.

Figure 54 displays the collection area limits for the training dataset and highlights that there is no FOD zone layout in the collection areas. The collection area limits are defined by green borders and measure 3,200 ft by 150 ft for Test Area 1 and 1,900 ft by 150 ft for Test Area 2.

Figure 54. Training Stage—No FOD Placement

Figure 55 displays the overview of the test areas and the two sets of two FOD zones in each. Each of the four FOD zones are 100 ft by 100 ft and contained 20 FOD items each.

Figure 55. Testing Stage—FOD Zones Identified

Figure 56 provides greater perspective on FOD zone placement during the testing for Test Area 1.

Figure 56. Test Area 1 Testing Stage—FOD Zone

Figure 57 provides greater perspective on FOD zone placement during the testing for Test Area 2.

Figure 57. Test Area 2 Testing Stage—FOD Zone

5.4.2 Data Processing

Data processing for the ACY dataset followed nearly the same protocol as the processing for the full-scale dataset at WWD. The primary difference from the WWD testing was the application of an additional postprocessing CRF filter to the output. A CRF is a tool which has been used effectively to help clean up the output imagery in other applications, such as segmentation networks. The research team determined that it would work well to help remove small instances of false positive detections that appear in many of the output images. A CRF uses a combination of filters to improve probability maps by running the filters over both the probability map and the original RGB image together. The process ends with a final probability map that is skewed towards 0 and 1 without as many values of medium probability. The combination of the probability map with the RGB image results in probability maps where the edges of the probability are more consistent with the edges in the real image. Examples of this algorithm are shown in Figure 58 and Figure 59.

Figure 58 displays an output of no FOD in the image. On the left is the pre-CRF raw RGB crop. The second image is the manually labeled ground truth. The third image is the pre-CRF output crop containing many false positives. On the right is the post-CRF output crop that resulted in zero false positives being reported.

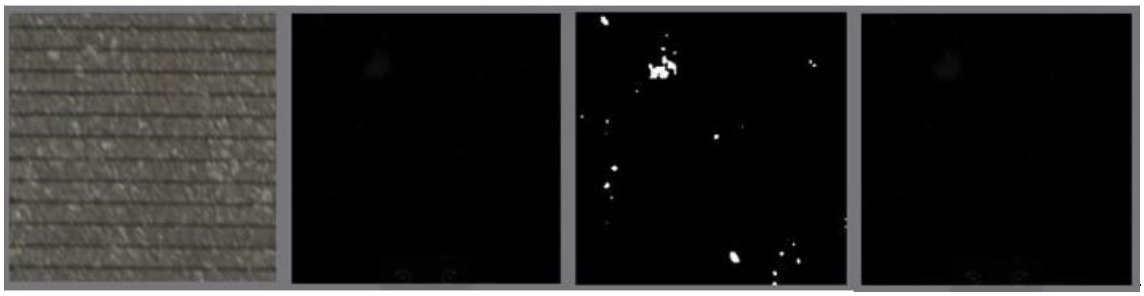

Figure 58. Example of CRF on an Image Containing No FOD

Figure 59 displays an output of a black golf ball FOD in the image. On the left is the pre-CRF raw RGB crop. The second image is the manually labeled ground truth crop. The third image is the pre-CRF output crop containing several false positives in addition to the FOD. On the right is the post-CRF output that resulted in only the true positive of the FOD being reported.

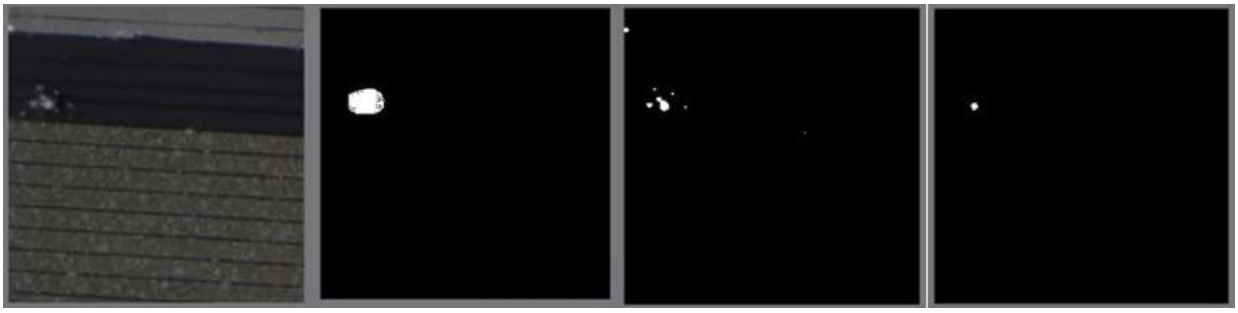

Figure 59. Example of CRF on an Image Containing FOD

In addition to CRF implementation, the research team investigated how two different cropping algorithms for generating crops from the raw imagery would impact the number of false positives a user would see. The first method, referred to as “training cropping,” segmented a raw image into a series of crops by minimizing the amount of overlap between the cropped images. Some overlap was to be expected because the crops were all fixed to a size of 256 pixels by 256 pixels, and the raw imagery size could not be evenly divided by 256. The minimal amount of overlap was to ensure that FastFlow was presented with mostly new pixels in each crop. The training cropping generated about 704 crops per raw image. The second method, referred to as “test cropping,” segmented the raw image such that 50% of the pixels in the X and/or Y dimensions could be duplicated in each subsequent crop. The large amount of overlap was intended to minimize the instances where FOD might only be partially seen on the boundary of more than one crop. The test cropping generated about 2,646 crops per raw image. Figure 60 illustrates both cropping methods on the same segment of a raw image.

Figure 60. Examples of the Cropping Methods—Testing (Left) and Training (Right)

5.4.3 Results and Analysis

Quantitative results for ACY testing showed similar values to full-scale tests when comparing results to the pre-CRF values. However, there is a significant reduction in the number of false positives reported due to CRF implementation. It should be noted that the CRF filter also slightly decreased the true positive rate in instances where the FOD target was poorly or barely detected pre-filtering.

ROC curves are shown in Figure 61. The left image is the image-wise ROC of no-grass training, and the right image is the image-wise ROC of with-grass training. Their AUCs are 0.87 and 0.90 respectively.

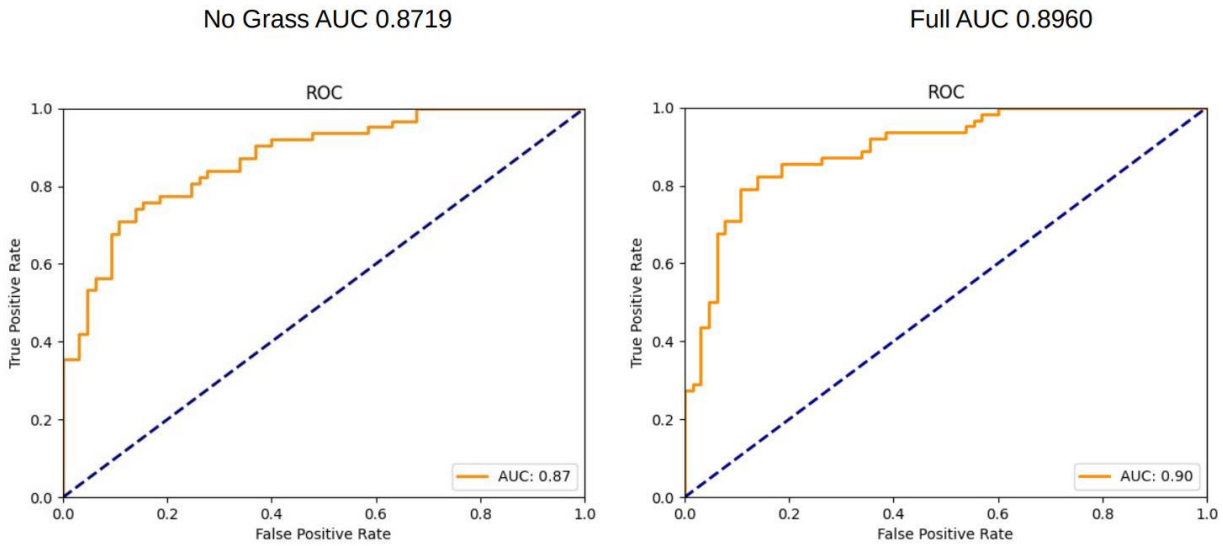

Figure 61. No-Grass Image-Wise ROC vs With-Grass Image-Wise ROC

To generate scores for this dataset, the research team decided that the crop where the FOD was most centered would be the only one scored. As with the other data collection efforts, a true positive was indicated if even a single pixel overlapped with the FOD. The true positive results for ACY testing datasets showed two missed pieces of FOD alongside the 75 pieces that were captured in the sUAS imagery. The FOD that were missed were on the smallest side of both size and visible profiles. The missed FOD were a lug nut and V-shaped metal scrap.

In previous tests, there was a focus on excluding grass from the data. However, the ACY test analysis determined that including grass in the FastFlow training proved to be beneficial for the algorithm. The grass inclusion had trained FastFlow to recognize the variable textures of grass and correlate that to some of the variable features like paint and cracks. The result was a visible difference in recorded false positives between the two datasets represented in Table 20.

Table 20. Validation Testing Results from ACY

Test Stage	90% Threshold	False Positive Crops per Image	True Positives	False Negatives	True Negatives	Total FOD	FOD Found	FOD Missed
ACY – No Grass	0.693	1,139	414	52	74	75	74	1
ACY – No Grass (CRF)	0.693	814	392	74	90	75	74	1
ACY – Grass	0.721	875	422	44	87	75	75	0
ACY – Grass (CRF)	0.721	447	394	72	108	75	73	2

To further reduce the number of false positives, the team reviewed the data with different cropping methods. The test cropping method, which had been used predominately in initial testing analysis, was expected to outperform the training cropping method because the FOD would be completely present in at least one of the overlapping images. However, in practice, the training cropping algorithm had nearly half as many false positives, and the number still fell within the FAA's required 90% threshold, as shown in Table 20.

Table 21 displays the reduction of false positives and FOD detections based on cropping method.

Table 21. Cropping Method Comparison

Test Cropped	Detection Rate	FOD Found	FOD Missed	False Positives
ACY – No Grass (CRF)	99%	74	1	814
ACY – Grass	97%	73	2	447
Training Cropped				
ACY – No Grass (CRF)	92%	69	6	216
ACY – Grass (CRF)	93%	70	5	119

The advantage of the training cropping method is the reduced number of image crops that were required to be processed. Searching for FOD in an image segmented by the training cropping method resulted in nearly a 75% reduction in the number of image crops compared to the test cropping method. The high level of redundancy in the test cropping method would result in a corresponding number of false positives being repeated for a given image. These false positives could be reduced by projecting them back into a single raw image and then recomputing the number of false positives across the unified image, but that would introduce additional processing time. Because the training cropping method can still perform above the 90% required threshold, it is a viable method in FOD detection, with fewer false positives.

5.4.4 ACY Test Findings

During the validation testing, the sUAS hardware was found to have a technological limitation at low altitude and low overlaps that had gone undetected during initial testing. It was discovered that the sUAS was collecting imagery approximately +/- 5 ft horizontally of that ideal collection position. In most use cases this would not be an issue; however, in this case, the combined GSD and overlap values means there was approximately 4 ft of imagery overlap planned. When two image spacing variance errors coincided, blank spaces were left that did not appear in the imagery, thus resulting in entirely missed FOD. It is unclear if there was a hardware issue resulting in inaccurate image collection timing or if it is a latent limitation of current technological capabilities. In total, five FOD were missed in data collection due to this occurrence.

From this testing, it was determined that 10% overlaps are insufficient to guarantee complete coverage. From planned overlaps, it is estimated that an overlap value of 20% forward overlap (i.e., 7 ft of overlap) and 15% side overlap (i.e., 8 ft of overlap) will present complete coverage while taking hardware limitations at high GSD into consideration for the hardware setup used in this research effort.

Both test and training cropping meet the 90% detection requirement of AC 150/5220-24; however, each has different areas of impact. The test cropping focuses on equal cropping overlap, which results in capturing a majority of FOD, but also results in more false positives per image. The training cropping focuses on reducing cropping overlap, which significantly reduces false positives at the cost of missing additional FOD. As both meet the 90% detection requirement, it becomes a choice between the cropping that is most effective in detecting FOD and the one that is more efficient for the user in sorting out false positives.

Table 22 displays the FOD target detection breakdown for CRF with grass dataset compared to initial testing results.

Table 22. Validation Testing FOD Detection Breakdown

FOD Target	sUAS/AI Detection Rate Initial Testing*	sUAS/AI Detection Rate Validation Testing*	sUAS/AI Detection Rate Validation Testing**
An unpainted metal cylinder	100% (9 of 9)	100% (6 of 6)	100% (6 of 6)
A white, gray, or black sphere	100% (32 of 32)	100% (10 of 10)	100% (10 of 10)
A “chunk” of asphalt or concrete	100% (3 of 3)	100% (6 of 6)	100% (6 of 6)
Any portion of a runway light fixture (in-pavement or edge light)	100% (3 of 3)	100% (5 of 5)	100% (5 of 5)
An adjustable crescent wrench	100% (19 of 19)	100% (8 of 8)	100% (8 of 8)
A deep socket	89% (8 of 9)	100% (6 of 6)	100% (6 of 6)
A piece of rubber from an aircraft tire	100% (3 of 3)	83% (5 of 6)	100% (6 of 6)
A distorted metal strip	78% (7 of 9)	67% (4 of 6)	83% (5 of 6)
A fuel cap	100% (3 of 3)	100% (4 of 4)	100% (4 of 4)
A lug nut	78% (7 of 9)	83% (5 of 6)	83% (5 of 6)
A hydraulic line	100% (3 of 3)	83% (5 of 6)	100% (6 of 6)
A white PVC pipe	100% (3 of 3)	100% (6 of 6)	100% (6 of 6)
Total	95 % (100 of 105)	93% (70 of 75)	97% (73 of 75)

* Displays results using training cropping method.
** Displays results using test cropping method.

5.4.5 ACY Testing Geolocation

As part of this research, a proof-of-concept FOD geolocation workflow was developed and implemented using Microsoft Excel. The goal of this geolocation workflow was to demonstrate

the ability to meet location accuracy performance specifications from FAA AC 150/5220-24 without requiring the use of additional expensive survey equipment or software.

5.4.5.1 Workflow

To perform geolocation, the three following data sources were used:

- Image Centers from the sUAS imagery
    - These contain the image name and the coordinates (latitude/longitude) of the center of the image. This information is contained in the imagery Exchangeable Image File Format (EXIF) metadata from DJI camera payloads and the imagery metadata of most commercially available global navigation satellite system (GNSS)-enabled sUAS.
    - The image center latitude and longitude coordinates should be converted to the appropriate State Plane coordinates (a grid coordinate system where the X value represents the easting, and the Y value represents the northing).
- While Blue Marble Geographics' Global Mapper software was used to produce these grid coordinates from the sUAS's native latitude and longitude data, free conversion utilities, such as the National Geodetic Survey's (NGS) Coordination Conversion and Transformation Tool (NCAT) (NCAT, n.d.), can be used to perform this step.
- sUAS Imagery Payload Metadata
    - The image height and width in pixels, and the dimensions of each pixel (the GSD) are needed to compute distances.
- FastFlow anomaly output
    - A .csv file output by the FastFlow routine containing a list of all image names in which an anomaly was detected and the position within each image (in X, Y pixel coordinates) of the detected anomaly.

These data sources were entered into a Microsoft Excel workbook that performs the following computations:

- The FastFlow anomalies are delivered in pixel distances relative to the upper-left corner of the image. These distances are converted to a spatial distance (using the GSD) and a bearing relative to the center of the image.
- An orientation is calculated for each sUAS-captured image by comparing the heading between it and the two images taken prior and the two images captured subsequently.
    - The consistency of these headings is evaluated to determine if the sUAS was flying in a straight line or if it was turning.

- If the sUAS was determined to be flying in a straight line, then the heading computed between the prior image and the current image is assumed to be representative of the image's orientation.

- The image orientation bearing is summed with the FastFlow anomaly's bearing from center.

- This combined bearing and the previously calculated FastFlow anomaly position's distance from the center are then applied to the image center coordinates to calculate and output each anomaly's position in the State Plane coordinate system.

- In a commercial implementation of this system, a program can be developed to automate this workflow, resulting in coordinates that would then be transferred to a hand-held GPS receiver for manual retrieval of the detected FOD. In this proof-of concept implementation, the calculated FOD positions were compared to survey measurements collected when the FOD was placed on the runway. The accuracy of these comparisons is shown in Table 23. The survey measurements were performed via RTK measurements with an accuracy greater than 4 cm.

5.4.5.2 Results

Table 23 displays the implemented geolocation workflow for some FOD detections within the ACY datasets.

Table 23. Geolocation Workflow Accuracy Assessment

	Computed Location		Surveyed Position		
Type of FOD	**Northing**	**Easting**	**Northing**	**Easting**	**Error (m)**
Rubber strip	224725.16	468177.27	224725.86	468178.23	0.363
Fuel Cap	224721.19	468147.38	224720.96	468147.38	0.069
Wrench	225430.90	468564.78	225433.75	468564.93	0.869
PVC	225443.52	468539.97	225443.85	468541.66	0.525
PVC	229440.80	470701.79	229441.65	470702.98	0.444
Metal scrap	229455.32	470679.55	229455.51	470679.81	0.098
Asphalt chunk	228895.18	470476.90	228896.88	470476.72	0.522
Taxi light	228873.16	470464.05	228873.47	470463.97	0.098

As shown in the table, all FOD objects tested via this methodology had their position determined with an accuracy exceeding the AC 150/5220-24 requirements.

5.4.5.3 Limitations

- The accuracy of the sUAS image center position has a direct bearing on the computed FOD location. For this reason, it is recommended to use an sUAS capable of RTK GNSS navigation.
- The location computations all assume nadir photography and a stable imaging payload. If imagery is captured during high winds or other conditions that might impact camera

stability, then additional computational steps might be needed to correct for camera tilt and rotation.

- All the images captured during this research had a small (approximately 16 m by 11 m) footprint. If imagery is collected with a significantly larger footprint, accumulated errors will increase in magnitude.
- The present workflow requires the entire dataset to be processed before computing FOD locations.

6. ASSESSMENT BASED ON AC 150/5220-24 SPECIFICATIONS

The research team assessed performance against the requirements listed in AC 150/5220-24; however, not all requirements were applicable to sUAS and AI. Several requirements were that of a mature detection system, as opposed to the nascent nature of this technology, while other requirements were outside the scope of this research effort and are listed with recommendations to meet the AC requirements.

Table 24 presents the basic performance requirements a FOD detection equipment must perform, along with ATR findings. The sUAS-based FOD detection workflow was able to meet four of the six required basic functions required by AC 150/5220-24. The two functions not demonstrated (providing detection alerts to users and providing a record of detected FOD) could be implemented in the future via development of a robust software interface.

Table 24. Basic Performance Requirements

Category	Basic Performance Requirement	ATR Finding
Basic Functions	Provide surveillance in the AOA as specified by the airport.	Demonstrates ability to meet AC specifications for detecting FOD in the AOA.
	Detect and locate single and multiple FOD items on the AOA.	Detected single and multiple FOD items. Geolocation meeting the AC standards demonstrated.
	Provide an alert to the user when FOD has been detected.	Alerts possible, requires software implementation.
	Operate in conjunction with, and not interfere with, airport and aircraft communication, navigation, and surveillance systems.	Demonstrated successful operation at towered and non-towered airports with no interference reported.
	Operate in conjunction with, and without interference from, normal airport and aircraft operations.	Demonstrated successful operation at towered and non-towered airports with no interference reported.
	Provide a data record of detected FOD, allowing for equipment calibration and maintenance, and for analysis of the FOD event.	Robust reporting requires software implementation of false positive filtering and combining true positives with the developed geolocation workflow.

Table 25 outlines the combined FOD object detection results for each item listed within AC 150/5220-24 for the initial testing and validation testing. Of 180 FOD items, 173 (96%) were detected overall. The FOD items not detected included the lug nut (12 of 15 detected), the deep socket (14 of 15 detected), and the distorted metal strip (12 of 15 detected).

Table 25. Detection Performance Requirements

Category	Detection Performance	Object Dimensions	ATR Finding
Object Detection	An unpainted metal cylinder	1.2 in. (3.1 cm) high and 1.5 in. (3.8 cm) in diameter	100% (15 of 15)
	A white, gray, or black sphere	1.7 in. (4.3 cm) in diameter (i.e., a standard size golf ball)	100% (42 of 42)
	90% of the following group of objects when placed within a 100 ft by 100 ft (30 m by 30 m) square in the desired coverage area. One item from each category must be included in the group.		
	A "chunk" of asphalt or concrete	no larger than 4 in. (10 cm) in any dimension	100% (9 of 9)
	Any portion of a runway light fixture (in-pavement or edge light)	no larger than 4 in. (10 cm) in any dimension	100% (8 of 8)
	An adjustable crescent wrench	up to 8 in. (20 cm) in length	100% (27 of 27)
	A deep socket	at least 2 in. (5 cm) in length	93% (14 of 15)
	A piece of rubber from an aircraft tire	no larger than 4 in. (10 cm) in any dimension	100% (9 of 9)
	A distorted metal strip	up to 8 in. (20 cm) in length	80% (12 of 15)
	A fuel cap	no larger than 4 in. (10 cm) in any dimension	100% (7 of 7)
	A lug nut	no larger than 4 in. (10 cm) in any dimension	80% (12 of 15)
	A hydraulic line	up to 8 in. (20 cm) in length	100% (9 of 9)
	A white PVC pipe	2 in. (5 cm) in diameter	100% (9 of 9)
	Any two of the objects above, located no more than 10 ft (3 m) apart from each other, identified as separate objects		Successfully distinguished

Table 26 presents different feature aspects of detection performance. The implemented workflow was able to achieve the geolocation, inspection frequency, and surveillance area requirements. However, further research and development would be needed to develop an sUAS system that can operate in low light and inclement weather conditions. Additionally, the sUAS-based workflow greatly exceeds the maximum false alarm rate for FOD detection systems with visual sensing capability of one per day.

Table 26. Detection Performance Requirements

Category	Detection Performance	Performance Requirement	ATR Finding
Location Accuracy	Must provide location information for a detected object	Within 16 ft (5.0 m) of the actual FOD object location	Met AC specification
Inspection Frequency	Mobile detection systems: must provide a mobile operations capability to enhance mandated airport safety self-inspections	Airport dependent	Met AC specification
Surveillance Area	Airport operator will specify the desired surveillance (detection) area in the AOA requiring FOD detection	The manufacturer of a FOD detection system must notify the airport operator of any locations within the specified surveillance area where detection would not be possible	Met AC specification.
Weather	Must demonstrate the detection performance under both clear and inclement weather conditions	Detect FOD under rain, snow, clear and inclement weather, lighting conditions, and time required for the system to recover after inclement weather	Further research required under low light and inclement weather conditions
Alerts and Alarms	Must be able to alert the system operator to the presence of FOD in scanned areas with enough information to assess the severity of the hazard to determine if immediate object removal is necessary	False alarms should be minimal—1/day with visual capabilities, 3/day without visual capabilities	Alerting and alarms require additional software programming and implementation Further research required to reduce false positive rate

Table 27 outlines data and user software performance requirements. Due to the proof-of-concept nature of the testing, these were not assessed. However, these features could be implemented through a robust software interface.

Table 27. System Output Requirements

Category	System Output	Performance Requirement	ATR Finding
Detection Data	Data record on detected FOD	Alert time, date, location at minimum	Requires software implementation
Data Presentation	Coordinate scheme, on maps of the airport, in an operator's console, or broadcast to mobile units	As specified by the airport	Requires software implementation
Data Management	Digital record	Capability to retain the data for at least 2 years after the detection event	Requires software implementation

7. NEXT STEPS AND RECOMMENDATIONS

The sUAS-based FOD detection workflow based on the FastFlow ML deep learning algorithm was capable of meeting some of the AC 150/5220-24 requirements, including achieving a 96% detection rate for FOD items specified in the AC. However, further research and development will be needed for this technology to meet the full set of AC 150/5220-24 requirements, including reducing the false positive rate, reducing the data processing time, and implementing a software interface for displaying and recording FOD detection alerts. Additionally, further research and development would be needed on sUAS and sensing technologies capable of detecting FOD in low light and inclement weather conditions. These are discussed in Sections 7.1–7.5.

7.1 DETECTION OF FOD ITEMS WITH LOW CONTRAST

Although the FastFlow algorithm had a high true positive rate for objects of larger size and/or with distinctive visual appearance compared to the pavement around them, the algorithm struggled with objects of small sizes, such as lug nuts and metal strips, that had visual reflective similarities to the runway markings. Additionally, FastFlow reported a 16.9% rate of false positives per image regardless of cropping method, which must be reduced before the algorithm can be put into service as a product. Pieces of FOD that differed in appearance from the background pavement on which they were placed were nearly always detected by the FastFlow algorithm. In contrast, FOD that very closely matched the pavement on which they were placed were much more difficult to detect and were almost always missed by the algorithm. The algorithm is reliant upon the capabilities of the sUAS's camera to generate imagery, which can mean that over- or under-saturated sections of image and areas of low contrast could cause the algorithm to miss pieces of FOD.

To address the limitation of FOD with a visual appearance that is nearly indistinguishable from the surrounding runway, the team proposes exploring two different solutions. The first approach

is to test image preprocessing algorithms, such as image normalization and contrast enhancement, to determine whether the FOD can be made more visually distinctive. The potential downside to this approach is that small non-FOD visual variations in color and pattern on the runway might also be enhanced and be more difficult to distinguish from FOD. A second approach is to explicitly reason about height differences in the runway by computing a 3D model using photogrammetry methods applied to the sUAS imagery. Generating the height value for each of the image pixels could allow for visually indistinct FOD or smaller objects, such as lug nuts or bolts, to be detectable outside the use of FastFlow.

7.2 FALSE POSITIVES

Another challenge revealed by this testing is the high likelihood of false positives being reported by the algorithm. While the FastFlow models were trained using images taken from the runway, certain visual features (such as chipped paint, cracks, or tar patches) did not appear in large quantities. As a result, the algorithm appeared not to learn that these features were not FOD and still flagged some of these non-FOD items as FOD. Thus, the default FastFlow algorithm still requires very careful curation of images for training, and more research will be needed to reduce false positives while still maintaining a high true positive detection rate.

To address the problem of false positives, the research team proposes exploring several different solutions. One approach is to carefully curate the image data used to train the FastFlow models. For example, because the paint patterns, asphalt coloring, and weathering for each runway are unique, the team proposes generating runway-specific models rather than trying to make a generic model that will work for all runways. The expectation is that runway-specific models would allow the algorithms to learn that specific patterns of chipped paint and tar patches are not anomalous but are expected and should not be labeled as FOD.

Another avenue the team recommends exploring is training a group of multiple different runway-specific FastFlow models that are each focused on separate categories of normal runway, including painted lines, cracks, and tar patches, and uniform segments of pavement. This approach would make use of a pre-sorting algorithm that will determine whether a crop contains any painted lines, cracks, or tar patches, or is a “clean” segment of pavement. Each collection of sorted images will be fed into a different instance of FastFlow that has been trained exclusively on images of that type. The expectation is that specialized FastFlow models will be able to reduce false positives caused by paint, tar, or rubber tracks.

Airfield infrastructure, such as lights that are embedded in the runway, typically show up as false positives in this testing. The main reason is that there are no sufficient examples of this sort of structure in the training set to allow the algorithm to fully generalize this as a normal and expected feature. If runway-specific models and feature-specific models are insufficient to address this issue, then a specialized detector that is trained via supervised learning to recognize these individual lights can be explored. This light detector would run on the images and flag any instance of lights that it finds so that any false positive instances reported by FastFlow on those lights can be discarded and not passed to the user.

In addition to finding FOD on the surface of the runway, FastFlow also flagged several other non-FOD anomalies that could be indicative of pavement stress on the runway that could require

inspection and maintenance. Such stresses include spalling and loose chunks of asphalt which might soon become FOD. Because FastFlow is fundamentally only looking for visually apparent deviations from a "normal" runway appearance, it will likely flag FOD and surface wear equally. Thus, FastFlow could be used in both FOD detection and runway pavement condition inspection activities.

7.3 PROCESSING TIME

Another limitation of the implemented workflow is the extensive processing time due to computing constraints. FastFlow's runtime performance on an Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz with an NVIDIA 2080Ti graphics processing unit (GPU) processes images at a rate of one full image of data every 4 minutes (approximately three crops per second). The processing time will scale linearly with the number of images that are captured for a given runway. The process of searching for FOD can be greatly accelerated by increasing the number of computers/GPUs that are in use at the same time, as each crop can then be processed independently of each other. One potential solution to increasing the processing speed is to explore how to modify FastFlow to process multiple crops simultaneously on a single GPU and measure the performance increase. Another potential solution is to upload all the raw images to a cloud-based service such as Amazon Web Service or Microsoft Azure cloud with multiple GPU nodes and have each node process one of the images. At such a scale, the entire dataset could be processed in the time it currently takes to process a single image. Additional data management infrastructure would be required to activate each of the cloud nodes, upload each image to the appropriate nodes, collect the resulting FOD detections, and aggregate all the results into a single output that can be presented to the user.

7.4 SOFTWARE INTERFACE DEVELOPMENT

The default FastFlow algorithm must be built into a robust FOD-detection software interface before it can be deployed for testing in situ at an airport. An application must be built around the algorithm that will input a set of geotagged images captured from the sUAS and generate a set of runway locations where a technician would go to retrieve the reported FOD. This application would address all the sections labeled "requires software implementation" in the AC 150/5220-24 requirements tables. The research team anticipates that a FOD application would directly involve human interaction to visually inspect the images that contain instances of reported FOD to determine whether they are true or false positives before dispatching someone to the runway to find the FOD to remove it. This application could also include a feature that allows a user to tag false positive anomalies that repeatedly show up on a given runway (e.g., light fixtures) as being "safe" so that they are not reported as FOD. In addition, the application should allow its users to periodically update the runway model should its appearance begin to change by routine wear and tear, resurfacing, or repainting, and the visual differences caused by seasonal variations. This update process could also be performed incrementally whereby only some portion of the runway would need to be updated rather than the entire stretch. Additional work would be required to build such a data management system which could store, catalog, and process historical collections of images.

7.5 OPERATIONS IN LOW-LIGHT AND INCLEMENT WEATHER CONDITIONS

Due to the proof-of-concept nature of the implemented workflow, testing was only conducted in fair weather and daytime lighting conditions. Further research and development would be needed on sUAS and sensing technologies capable of detecting FOD in low light and inclement weather conditions. Visual camera sensors are limited in their ability to capture imagery in twilight and nighttime conditions. sUAS platforms also have specific operation tolerances for wind, temperature, and precipitation, which limits the ability to use sUAS in inclement weather conditions. Additionally, detecting FOD on wet pavement would likely require additional ML model training.

8. SUMMARY

The FAA ATR Branch conducted a research effort to explore the feasibility and maturity of using commercially available sUAS and AI/ML algorithms to detect FOD on airport surfaces. The objectives of this research effort were to develop a novel, proof-of-concept sUAS-based FOD detection workflow using AI/ML algorithms and to assess the workflow to determine whether it is capable of meeting all, some, or none of the requirements in FAA AC 150/5220-24.

The research team developed and trained an AI/ML deep learning network as part of a sUAS-based FOD detection workflow. The research team conducted initial testing of this workflow at WWD. The initial testing effort at WWD comprised three stages: small-scale, calibration, and full-scale testing. The research team then performed validation testing at ACY to validate the initial testing at WWD. This included collecting data on a runway at ACY, training an anomaly detection AI/ML algorithm, testing against a variety of FOD targets, and addressing requirements such as implementing geolocation and adhering to accuracy requirements set forth in FAA AC 150/5220-24.

The sUAS-based FOD detection workflow based on the FastFlow ML deep learning algorithm was capable of meeting some of the AC 150/5220-24 requirements, including achieving a 96% detection rate for FOD items specified in the AC. However, further research and development will be needed for this technology to meet the full set of requirements laid out in AC 150/5220-24, including reducing the false positive rate, reducing the data processing time, implementing a software interface for displaying and recording FOD detection alerts, and detecting FOD in low-light and inclement weather conditions.

Finally, the testing led to identification of best methods for data collection. Images captured from a nadir vantage point had a significantly higher true positive detection rate than images captured from an oblique vantage point. Positive detection rate was also improved by flying the sUAS lower to the runway for a 0.2-cm GSD, thereby increasing the image resolution.

9. REFERENCES

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2021). The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, *129*, 1038–1059. https://doi.org/10.1007/s11263-020-01400-4

Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4182–4191. https://doi.org/10.1109/CVPR42600.2020.00424

Cohen, M. J., & Avidan, S. (2021). Transformaly—Two (Feature Spaces) are better than one. *ArXiv:2112.04185 [Cs]*. https://arxiv.org/abs/2112.04185

Deecke, L., Vandermeulen, R., Ruff, L., Mandt, S., & Kloft, M. (2019). Image anomaly detection with generative adversarial networks. In: Berlingerio, M., Bonchi, F., Gärtner, T., Hurley, N., Ifrim, G. (eds.), *Machine Learning and Knowledge Discovery in Databases*, 3–17. https://doi.org/10.1007/978-3-030-10925-7_1

Definitions, 14 Code of Federal Regulations § 107.3 (2016). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-107/subpart-A/section-107.3

Di Mattia, F., Galeone, P., De Simoni, M., & Ghelfi, E. (2021). A Survey on GANs for Anomaly Detection. *ArXiv:1906.11632 [Cs, Stat]*. https://arxiv.org/abs/1906.11632

FAA. (2009). *Airport foreign object debris (FOD) detection equipment* (Advisory Circular [AC] 150/5220-24). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_150_5220-24.pdf

FAA. (2015). Foreign object debris characterization at a large international airport. https://www.airporttech.tc.faa.gov/DesktopModules/EasyDNNNews/DocumentDownload.ashx?portalid=0&moduleid=3682&articleid=36&documentid=60

Liang, Y., Zhang, J., Zhao, S., Wu, R., Liu, Y., & Pan, S. (2022). Omni-frequency channel-selection representations for unsupervised anomaly detection. *ArXiv:2203.00259 [Cs]*. https://arxiv.org/abs/2203.00259

National Geodetic Survey (NGS) Coordinate Conversion and Transformation Tool (NCAT). (n.d.) https://www.ngs.noaa.gov/NCAT/

Rao, D. J., Mittal, S., & Ritika, S. (2017). Siamese neural networks for one-shot detection of railway track switches. *ArXiv:1712.08036 [Cs]*. https://arxiv.org/abs/1712.08036

Reiss, T., Cohen, N., Bergman, L., & Hoshen, Y. (2021). PANDA: Adapting pretrained features for anomaly detection and segmentation. *ArXiv:2010.05903 [Cs]*. https://arxiv.org/abs/2010.05903

Tayeh, T., Aburakhia, S., Myers, R., & Shami, A. (2020). Distance-based anomaly detection for industrial surfaces using triplet networks. *ArXiv:2011.04121 [Cs]*. https://arxiv.org/abs/2011.04121

Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., & Wu, L. (2021). FastFlow: Unsupervised anomaly detection and localization via 2D normalizing flows. *ArXiv:2111.07677 [Cs]*. https://arxiv.org/abs/2111.07677

Yuan, Z.D., Li, J.Q., Qiu, Z.-N., & Zhang, Y. (2020). Research on FOD Detection System of Airport Runway Based on Artificial Intelligence. *Journal of Physics: Conference Series*, *1635*(1), 012065. https://doi.org/10.1088/1742-6596/1635/1/012065

Zhou, C., & Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3097983.3098052

APPENDIX A—AIRFRAME AND SENSOR SPECIFICATIONS

DJI Matrice 210 RTK v2	
Type	Rotary Aircraft (4)
Wingspan	25.3” motor-to-motor cross measurement
Weight	10.83 lb with batteries only
Max Flight Time	± 25 min
Average speed of flight during image capture	± 15 mph
Operating Temperature Range	-4 °F – 122 °F
Transmitter Range	5 mi (unobstructed)
Communication with Transmitter	Radio (2.4000 – 2.4835 GHz; 5.725 – 5.850 GHz)
Maximum sustained wind speed limit for safe flight	Up to 27 mph
Lost Link Procedure (if > 3 s)	Autonomous return-to-home at predetermined above ground level (AGL) with manual override available once link has been reestablished.
Low Battery Procedure	Pilot override from autonomous to manual control and return sUAS to launch location and land when battery percentage reaches 20%. If battery decreases to level where flight computer can no longer maintain current altitude, sUAS will initiate autonomous land and current position.
Operational Area Procedure	On board, pre-programed flight area prohibits flying outside of pre-determined GeoFence.
Obstacle Avoidance	Forward, Down, Above, DJI Airsense (ADS-B Receiver)
Ingress Protection Rating	IP43

Zenmuse X7	
Airframe Compatibility	DJI Inspire 2, M210 RTK V2
Gimbal Control (3D Stabilized) (Detachable Mount*)	Pitch: -125° to +40° Pan: ±300° Roll: +90° to -50°
Still Image Size	24.0 MP (6016x4008) – 3:2
Ground Sample Distance (GSD) @ 400 ft AGL 16 mm	1.17 in.
Ground Sample Distance (GSD) @ 400 ft AGL 24 mm	0.78 in.
Ground Sample Distance (GSD) @ 400 ft AGL 3 5mm	0.54 in.
Ground Sample Distance (GSD) @ 400 ft AGL 50 mm	0.38 in.
Sensor Type	CMOS – Super 35 APS-C
Still Image Format	JPEG, RAW, RAW + JPEG
Shutter Mode	Electronic Linear for 16 mm & 50 mm Mechanical Global for 24 mm & 35 mm
Max Video Resolution*	4K @ 59.94 FPS
Max Video Bitrate	100 Mbps
Video Format	MP4/MOV

DJI Matrice 300 RTK	
Type	Rotary Aircraft (4)
Wingspan	38” motor-to-motor cross measurement
Weight	13.9 lb with batteries only
Max Flight Time	± 55 min
Average speed of flight during image capture	± 15 mph
Operating Temperature Range	-4 °F – 122 °F
Transmitter Range	5 mi (unobstructed)
Communication with Transmitter	Radio (2.4000 – 2.4835 GHz; 5.725 – 5.850 GHz)
Maximum sustained wind speed limit for safe flight	Up to 33 mph
Lost Link Procedure (if > 3 s)	Autonomous return-to-home at predetermined AGL with manual override available once link has been reestablished.
Low Battery Procedure	Pilot override from autonomous to manual control and return sUAS to launch location and land when battery percentage reaches 20%. If battery decreases to level where flight computer can no longer maintain current altitude, sUAS will initiate autonomous land and current position.
Operational Area Procedure	On board, pre-programed flight area prohibits flying outside of pre-determined GeoFence.
Obstacle Avoidance	Forward, Backward, Left, Right, Down, Above,
Ingress Protection Rating	IP45

Zenmuse P1	
Airframe Compatibility	M300 RTK
Gimbal Control (3D Stabilized) (Detachable Mount*)	Pitch: -130° to +40° Pan: ±320° Roll: +55° to -55°
Still Image Size	45.0 MP (8192x5460) – 3:2
Ground Sample Distance (GSD) @ 400 ft AGL 24 mm	0.85 in.
Ground Sample Distance (GSD) @ 400 ft AGL 35 mm (currently only available option)	0.59 in.
Ground Sample Distance (GSD) @ 400 ft AGL 50 mm	0.41 in.
Sensor Type	35 mm (Full Frame)
Still Image Format	JPEG, RAW, RAW + JPEG
Shutter Mode	Mechanical Global
Max Video Resolution*	4K @ 60 FPS
Max Video Bitrate	100 Mbps
Video Format	MP4/MOV