Ongoing Work: A Prototype Dataset for Low-flying Autonomous Medical UAS Operations

J. Tanner Slagel*, Sarah M. Lehman, Kyle M. Smalling, Joshua M. Fody, and David Wagner
NASA Langley Research Center, Hampton, Virginia, 23666, USA

Jack T. Fortner-Monegan
NASA Glenn Research Center, Cleveland, Ohio, 44135, USA

Ricardo A. Arteaga
NASA Armstrong Flight Research Center, Edwards, California, 93523, USA

This paper presents ongoing work to create a dataset for low-flying autonomous medical Uncrewed Aerial Systems (UAS) operations, focused on human stance recognition. This is an exploration of the viability of airborne classification for the Drone as a First Responder (DFR) concept in which a UAS arrives at the scene of an incident before emergency response personnel can get there and provides some level of situational awareness for the personnel arriving to the scene. Future incarnations could also see the UAS administer some level of care to injured parties at the scene. The data set, focused on detecting human stance, being developed here is the result of 30 test flights at NASA Langley Research Center in early 2024. In addition to flights where the participant (an anthropomorphic testing device or human) is alone in the viewing area holding a particular stance, two emergency scenes have been fabricated and collected through video - "bike crash" and "difficult camping". These test flights include four human participants. The contribution of this work upon completion will be a publicly available data set for the development of classification engines focused on human stance, and in the future, even triage.

I. Introduction

NASA's Convergent Aeronautics Solutions (CAS) project is focused on investing in ideas that lead to solutions in aeronautics and related industries*. The CAS project has identified an opportunity for NASA to help increase the quality of human health by enabling the burgeoning medical Uncrewed Aerial Systems (UAS) industry, e.g., [1]. One technology in this space is using a Drone as a First Responder (DFR). The idea of DFR is that, in an emergency response scenario, a drone is deployed to the target area at the same time that a dispatcher deploys traditional resources such as an ambulance, firetruck, or police unit. In many environments, a drone is capable of getting to the scene earlier than other resources; this is true in both rural and crowded urban areas. Currently, implementations of the DFR concept have a human pilot flying the drone, manually mapping the scene, and sending textual information back through the dispatcher to emergency medical personnel. This approach suffers from latency issues as well as the possibility of human error due to the stressful nature of the scenario.

Rapid advances in processing capacity and resource efficiency for onboard UAS sensors are helping to address these issues. The availability of data streams such as coarse- and fine-grained positioning information, inertial and movement data, Wi-Fi and cellular network signals, etc. means that *automatic scene mapping* can be conducted without requiring manual input from the operator. Additionally, the concept of *automatic scene perception* is of great importance to medical UAS applications, where objects of interest within the environment are automatically detected and that information is populated into a common operating picture for first responders. Examples of perception in a medical emergency environment might include door locations, structural integrity of infrastructure, animals in the area, detection and triage of humans, presence of weapons, obstructions to entrances or exits, prescription medicine, drugs, biological material, and more.

*Corresponding author: j.tanner.slagel@nasa.gov
*https://www.nasa.gov/directorates/armd/tacp/cas/

Engagement with the medical UAS industry revealed a need for low-altitude aerial datasets that capture objects of interest related to medical UAS use cases; effective use of such a dataset would also require the development and assurance of perception engines that work for this unique altitude range and optical orientation. Typically, image classification and object detection datasets do not have the top-down perspective of a drone flying hundreds of feet above the visual target; thus, they are ill-suited for the development of perception engines for UAS operations. One effort towards obtaining related data is the NASA Airvue project, which is focused on developing a crowdsourced video dataset of helicopter flight in urban environments [2]. While the focus of the Airvue project does share some characteristics with medical UAS applications, it cannot wholly capture all aspects of the medical UAS use cases; this gap has motivated the development of a similar dataset directly focused on the needs of medical UAS, with the opportunity for future collaboration between these two efforts in the future.

The contributions of this paper are summarized as follows:

- **Prototype Dataset for Low-Altitude Medical UAS:** The primary contribution of this work is the creation of a prototype low-altitude aerial dataset for development of perception engines related to the burgeoning medical UAS domain, thereby addressing a gap that has been frustrating industrial and academic practitioners alike. The dataset comprises a collection of images showing different human stances captured using a drone flying at altitudes ranging from 50 to 200 feet. This altitude band was chosen so that human subjects do not occupy the majority of the image real estate; this is in keeping with popular "in-context" scene understanding datasets such as Microsoft's "Common Objects in Context" (COCO) dataset [3], and is important for maintaining a realistic representation of normal UAS operating conditions.

- **Proof-of-concept Perception Engine:** In addition to the creation of the dataset, we developed a proof-of-concept model for object detection which demonstrates the feasibility of our dataset for medical UAS operations. The model is based on the You Only Look Once (YOLO) neural network architecture [4], which has demonstrated high degrees of performance for aerial applications.

The rest of the paper is structured as follows. Section II describes the related work relevant to this effort. Section III describes the procedures followed to collect the raw UAS data, while Section IV describes the procedures followed to clean the data and assemble the final dataset. Section V discusses the development of the proof-of-concept classifiers, and Section VI concludes with a discussion of future work.

II. Related Work

UAS technology is no longer a dream for the future; the future has arrived. The UAS market has experienced explosive growth in the last few decades, garnering billions of dollars in venture capitalist investment across manufacturing, infrastructure, and operational support [5]. With this increased availability comes the opportunity for new applications, such as medical operations. The possibility of using drones in emergency medical care [6, 7] has garnered great interest in recent years within the research community, with work examining DFR services for pre-hospital care [8, 9], location and delivery of Automated External Defibrillator (AED) devices [10], disaster relief [11, 12], and even battlefield medicine [13]. There has also been an enthusiastic response from industry with solutions offered by organizations such as Draganfly Inc.[†], BRINC Drones[‡], Airobotics[§], Paladin Drones[¶], and Skyfire Consulting[‖]; and training resources provided by organizations such as Argus Rising[**] and DRONERESPONDERS[††].

As interest in UAS applications increases, it brings with it a demand for high-quality datasets. NASA's Airvue project facilitated the collection of a crowd-sourced video dataset of helicopter flight data in urban environments[2]. The VisDrone-2018 dataset [14] also focused on urban environments with data collected across 14 different cities; this resulted

[†] https://draganfly.com
[‡] https://brincdrones.com
[§] https://www.airoboticsdrones.com
[¶] https://paladindrones.io
[‖] https://skyfireconsulting.com
[**] https://www.argusrising.com
[††] https://www.droneresponders.org

(a) "Bike Crash" Scenario

(b) "Difficult Camping" Scenario

Fig. 1 Sample Emergency Response Scenarios

in a dataset comprising 8,599 images spanning ten different classes (pedestrian, person, car, van, bus, truck, motor, bicycle, awning-tricycle, tricycle). DroneCrowd [15] and CrowdFlow [16] focused on human trajectory tracking in urban environments with datasets comprising 2420 and 3200 images respectively. The UAVDT dataset [17] consists of roughly 80,000 images representing particularly challenging conditions for UAS including high-density areas, small objects in the frame, camera motion, and latency-sensitive classification. AIDER [18, 19] focused on disaster scenario classification with their dataset, collecting roughly 500 images each for fire/smoke, flood, collapsed building, and traffic accidents; as well as roughly 4000 "null" images where no disaster is present. RescueNet [20] and UAVid [21] also focus on natural disasters and landscape perception with an emphasis on semantic segmentation. NOMAD, the Natural Occluded Multi-scale Aerial Dataset, is a dataset consisting of 42,825 images focused on detecting humans who are occluded in some way from an aerial perspective for emergency response purposes [22]; individuals in this dataset are walking, laying, or hiding and metadata is included to indicate the percentage of the human body which is visible in the frame.

The dataset being developed in this work represents a contribution of data for aerial perception related to the DFR released to the general public. Although less comprehensive and smaller in size than other related datasets, two configurations have been modeled as emergency response operations, with flight paths similar to how a UAS might initially approach an emergency medical response scenario. Replicating an emergency scenarios for aerial video capture is not something the authors at this time have seen available in public data sets. These efforts represent the first steps to enabling automatic triage at an emergency medical scene with UAS.

III. Data Collection Through Test Flights

Raw data for the human stance dataset was collected through 30 test flights representing a total of 480 minutes of airtime. For 16 of the flights, an Anthropomorphic Testing Device (ATD) was used as a stand-in for a human. The use of the ATD allowed time for NASA's Institutional Review Board (IRB) to approve the data collection test flights with human participants. Use of the ATD - which was a male model, 5 foot 8 inches tall and weighing 150 pounds - introduced a few challenges. The configuration of the ATD made it impossible for him to stand unsupported, so test flights including the ATD were in either a sitting, prone, or supine position. To introduce variety in the data collection, the ATD's attire was changed periodically between scenario configurations.

Upon approval by NASA's IRB, four human participants were used for the final 14 data collection flights. Hazard mitigation plans were developed and enacted by the team in accordance with expectations from NASA's UAS Operations Office (UASOO); for a brief discussion of the hazard mitigation plan used during this data collection effort, see Section III.C. For the test flights with human subjects, two flights were conducted similar to those executed with the ATDs, where the human subject was alone in an area in a variety of stances, including standing, sitting, prone, and supine. The other 12 flights were dedicated to certain emergency response scenarios with the goal of using these videos to test developed classifiers. The two scenarios, entitled "bike crash" and "difficult camping", are reflected in Figure 1.

Fig. 2 Approximate locations of ATDs and human participants during data collection flights. Between different configurations, the location was subject to change within 50 feet. Location A identifies an area where the participant was surrounded by a grassy field, and Location B identifies an area was the participant was on a strip of sidewalk with grass on both sides.

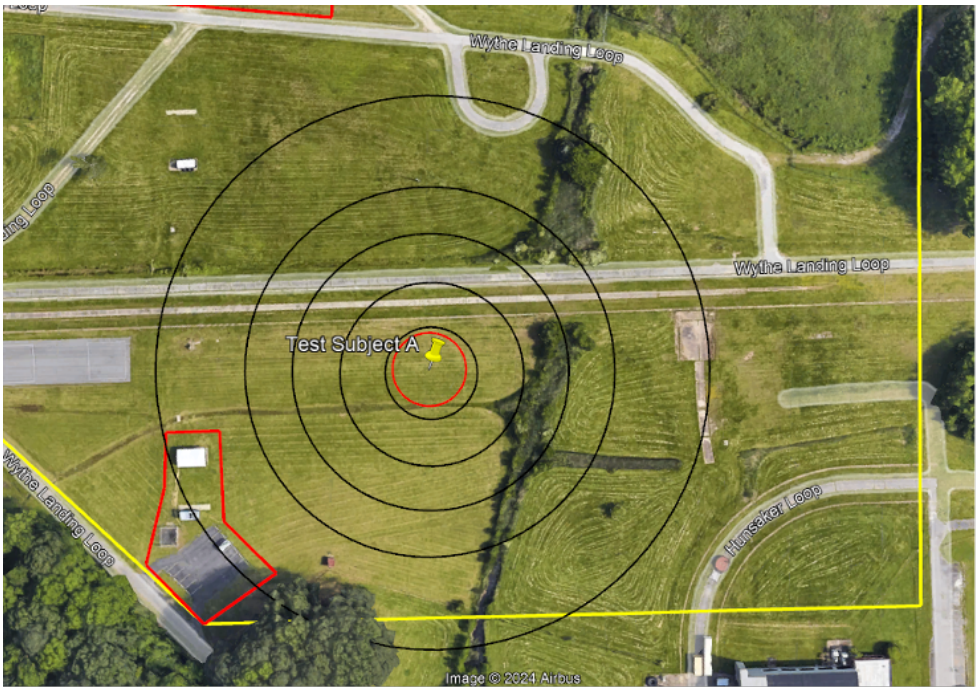
In "bike crash", there are two bicycles, a human, and an ATD, the poses of which are modified to reflect different specific conditions. In one variant, the ATD is laying down with the human seemingly checking the ATD's vitals (shown in Figure 1a), while another variant consists of the human standing with the bicycle between their legs while the ATD is sitting. Another configuration variant has both the ATD and human sitting with their bicycles askew. The scenario "difficult camping" consists of a human and the ATD lying amid wrecked camping equipment, as shown in Figure 1b.

Aerial data was collected using a Skydio2 supplied by the UASOO office at NASA Langley Research Center. Data was collected in the form of 100 20-minute videos covering a range of paths from various altitudes. Only a single human / ATD test subject stance configuration - standing, sitting, or laying down - was captured per video. Half of the test subject configurations were in a grassy field, and half were on a concrete walking path within NASA Langley's CERTAIN 1 range, as reflected in Figure 2.

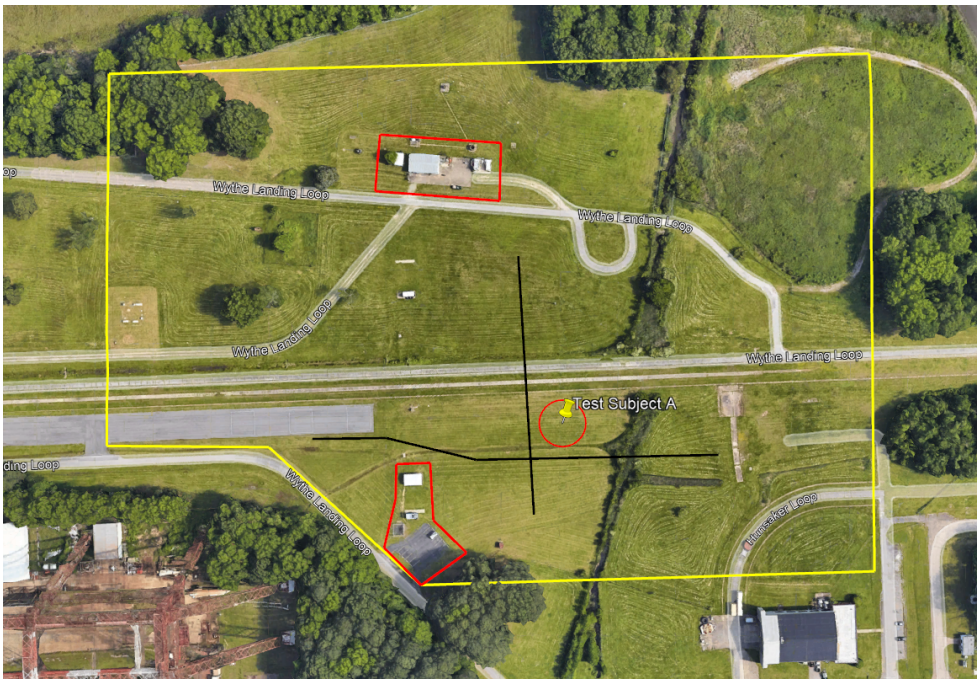
A. Data Collection Procedure

Figure 2 describes the operational area of the data collection test flights, all of which were conducted at NASA Langley's CERTAIN 1 range. The operational area contains two sites at which the human and ATD test subjects could be positioned. Site A placed the participant in a grassy field, while Site B had the participant on a concrete walking path down the center of the range. To preserve safety of the test flight operators and human participants, sections of the operational area were marked out as no-fly zones in order to avoid flying over or possibly colliding with a human.

Two types of paths were flown during data collection. The first was an *orbital path*, as shown in Figure 3a. For this type of flight path, the UAS follows a series of concentric circles at prescribed distances and altitudes around the subject, ranging from an altitude of 50 to 200 ft and horizontal distance of 50 to 300 feet. The other type of path was a *fly-by path*, which was a roughly straight line approximately 50 feet to the left or right of the participant, as shown in Figure 3b. This path occurred at altitudes ranging from 50 to 200 feet and was flown back and forth on an east-west and north-south path. For this path, the angle of the camera was toggled instead of continually focusing on the subject, which is more toward what an initial dispatch medical UAS footage might look like. The different angles used were 0,



(a) Orbital Flight Path



(b) Fly-by Flight Path

Fig. 3 Sample flight paths for data collection, encompassing a series of concentric circles at prescribed distances and altitudes around the test subject.

(a) Configuration 1 - Front

(b) Configuration 1 - Back

(c) Configuration 2

Fig. 4 Sample ATD Stances

45, and 90 degrees below horizontal.

B. Flight Test Configurations

Data was collected for 21 different configurations of human and ATD test subjects. This section describes those configurations in detail. A selection of sample ATD stances are shown in Figure 4. Summaries of the flight distances and camera angles for ATD scenarios are reflected in Figures 5 and 6 respectively. Configuration details for the human participant scenarios are described in Figure 7.

- **ATD Configuration 1:** The first configuration was the ATD, sitting upright with its arms behind for support, and its legs extended (Figures 4a and 4b). The ATD was wearing a black toboggan and plaid blue button-up with blue jeans. For this configuration, the ATD was sitting at Site A. Three test flights were conducted collecting data from this configuration, with altitudes ranging from 50-200 feet and horizontal distances 50-300 feet.

- **ATD Configuration 2:** The second configuration was the ATD, laying on its back with legs bent and arms slightly splayed by the sides (Figure 4c). The ATD was wearing a black toboggan and plaid blue button-up with blue jeans and sunglasses. For this configuration, the ATD was sitting in Site A. Three test flights were conducted in this configuration, with altitudes ranging from 50-200 feet and horizontal distances 50-300 feet. Additionally, the fly-by flight path was undertaken at two different camera angles and altitudes.

- **ATD Configuration 3:** The third configuration was the ATD sitting up with legs extended, and arms in its lap. The outfit for this configuration was a yellow baggy sweatshirt with hood up, flat bill hat, blue jeans. For this configuration, the ATD was sitting in Site A. Two test flights were conducted collecting data from this configuration, with altitudes ranging from 50-200 feet and horizontal distances 50-200 feet.

- **ATD Configuration 4:** The forth configuration was the ATD sitting up with legs extended, and arms in its lap. The outfit for this configuration was a yellow baggy sweatshirt with hood up, flat bill hat, and black pants. This configuration is the same as configuration 3, but with different pants. One test flight was conducted collecting data from this configuration, with orbital paths at 300 feet horizontal distance as 50-200 feet altitude, as well as a fly-by path path at 200 feet.

- **ATD Configuration 5:** The fifth configuration was the ATD in the prone position, laying flat on its stomach with legs fully extended. It is noted in this configuration the angel of the legs and the torso could not be made to be flat, so the flattest position possible was made. The outfit for this configuration was a yellow baggy sweatshirt with hood up, flat bill hat, and black pants, same as Configuration 4. Two test flights were conducted collecting data from this configuration, with orbital paths at 300 feet horizontal distance as 50-200 feet altitude, as well as a fly-by

		Horizontal Distance (ft)				
		50	100	150	200	300
Altitude (ft)	50	x	x	x	x	x
	100	x	x	x	x	x
	150	x	x	x	x	
	200	x	x		x	

(a) Configuration 1

		Horizontal Distance (ft)				
		50	100	150	200	300
Altitude (ft)	50	x	x	x		
	100	x	x	x	x	
	150	x	x	x	x	
	200	x	x	x	x	

(b) Configuration 2

		Horizontal Distance (ft)			
		50	100	150	200
Altitude (ft)	50	x	x	x	x
	100	x	x	x	x
	150	x	x	x	x
	200	x	x	x	x

(c) Configuration 3

		Horizontal Distance (ft)
		300
Altitude (ft)	50	x
	100	x
	150	x
	200	x

(d) Configuration 4

		Horizontal Distance (ft)				
		50	100	150	200	300
Altitude (ft)	50	x	x	x	x	x
	100	x	x	x	x	x
	150	x	x	x	x	x
	200	x	x	x	x	x

(e) Configuration 5

		Horizontal Distance (ft)				
		50	100	150	200	300
Altitude (ft)	50	x	x	x	x	x
	100	x	x	x	x	x
	150	x	x	x	x	x
	200	x	x	x	x	x

(f) Configuration 6

		Horizontal Distance (ft)	
		50	100
Altitude (ft)	50	x	x
	100	x	x
	150	x	x
	200	x	x

(g) Configuration 7

		Horizontal Distance (ft)	
		75	175
Altitude (ft)	50	x	x
	100	x	x
	150	x	x
	200	x	x

(h) Configuration 8

Fig. 5 Orbital Path Altitude and Horizontal Distance Settings for ATD Flight Tests

		Camera Angle (degrees)		
		-90	-45	0
Altitude (ft)	50E	x	x	x
	50W	x	x	x
	100E	x	x	x
	100W	x	x	x

(a) Configuration 2

		Camera Angle (degrees)
		-45
Altitude (ft)	200W	x
	200E	x

(b) Configuration 4

		Camera Angle (degrees)	
		-90	-45
Altitude (ft)	50EW	x	x
	100EW	x	x
	150EW	x	x
	200EW	x	x

(c) Configuration 5

		Camera Angle (degrees)	
		-90	-45
Altitude (ft)	50EW		x
	100S	x	
	200NS	x	x

(d) Configuration 6

		Camera Angle (degrees)	
		-90	-45
Altitude (ft)	50NS		x
	100NS	x	

(e) Configuration 7

		Camera Angle (degrees)
		-45
Altitude (ft)	100EW	x
	150EW	x

(f) Configuration 8

Fig. 6 Fly-by Camera Angle Settings for ATD Flight Tests

		Horizontal Distance (ft)	
		75	175
Altitude (ft)	50	x	x
	100	x	x
	150	x	x
	200	x	x

(a) Altitude and Horizontal Distance for Orbital Paths

		Camera Angle (degrees)
		-45
Altitude (ft)	75EW	x
	175S	x

(b) Camera Angle for Fly-by Paths

Fig. 7 Flight Test Settings for Human Participant Configurations

path at 50-200 feet and multiple camera angles.

- **ATD Configuration 6:** The sixth configuration was the ATD in the sitting position with its legs extended, and one arm up in the arm as if it is waving. The outfit for this configuration was a brown sweatshirt, with a black curly-hair wig and black pants. Three test flights were conducted collecting data from this configuration, with orbital paths at 300 feet horizontal distance as 50-200 feet altitude, as well as a fly-by path at 50-200 feet and multiple camera angles.

- **ATD Configuration 7:** The seventh configuration was the ATD laying in the prone condition with its arms and legs splayed. The ATD wore the same outfit as Configuration 6 (black curly wig, brown sweatshirt, black pants). Two test flights were conducted collecting data from this configuration, with orbital paths at 50 feet and 100 feet horizontal distance as 50-200 feet altitude, as well as a fly-by path at 50-100 feet and multiple camera angles.

- **ATD Configuration 8:** The eighth configuration had that ATD laying on a cot, in the "difficult camping" scene. This was the first flight of the emergency scenarios, and served as a warm-up for the human participant flights, that would be in the "difficult camping" or "bike crash scenes". The ATD wore the same outfit as Configuration 6 (black curly wig, brown sweatshirt, black pants). One test flight was conducted to collect data from this configuration, with orbital horizontal distances between 75-175 feet and altitudes between 50-200 feet.

- **Human Participant Configurations:** Configurations 9-21 included a set of four human participants. Six flights were of the "difficult camping" scene in the grassy field location (Site A). Each participant had one flight where they were sitting, laying, and standing while signaling towards the UAS. All flights followed the same flight paths, covering horizontal distances of 75-175 and altitudes of 50-200 feet, with fly-by paths at 100 and 200 feet altitude, as shown in Figure 7. Five flights with humans were the "bike crash" scene and had the same flight path parameters as the "difficult camping" scenario, although the scene was in the walking path location (Site B). Additionally, to augment the corpus of stance-only data, two flight tests were taken with just human participants and no scene in the field of view. Each of these flights went through three different poses: standing, laying, and sitting. When the participant was standing, they would signal occasionally towards the UAS.

C. Hazard Mitigations

The following hazards were identified as applicable, normal risk of flight for Visual Line-of-Sight (VLOS) operations in this data collection effort, per the hazard analysis definition in LPR 1710.16, Section 5.5.6[‡‡]: Personal Injury, Collateral Damage, Vehicle Fly-Away, and Hazardous Chemicals. The following plan was developed and observed to mitigate these hazards during test flights according to NASA procedures:

- **Flight Conditions:** The Skydio2 was kept at altitudes between 55 and 300 feet. No-fly zones were established in 40 foot radii around human participants, buildings, and obstacles such as trees and power lines. Whenever an obstacle fly-over was unavoidable, an additional 150 feet of elevation was observed.

- **Lost Link Procedure:** In case of link loss exceeding five seconds, disarm motors would activate automatically. The UAS would slowly descend and attempt a return to launch, while the test flight operators make contact with test subject participants either in-person or over radio.

- **Range Containment:** To prevent the UAS leaving the approved operational area, an altitude geofence was established at 350 feet. A lateral geofence was not implemented, as a human operator would be in control of this movement. For emergencies, a physical controlling button can also be pressed to force the Skydio to return to launch, though the path back might include a flight path over human participants.

- **Non-Standard Equipment:** The mitigation plan also included some non-standard equipment both to improve the quality of data collection, such as the ATD itself, and to facilitate ease of communication and comfort of human participants. Human participant support equipment included VHF radios, chairs, and a wheelchair.

[‡‡] https://lmse.larc.nasa.gov/admin/public_docs/LPR_1710-16J_Final(final).pdf

D. Human Subject Protections

As this experiment involved human subjects, there were a number of important considerations to protect their safety and privacy. Before any data collection was attempted, approval was received from NASA's IRB to ensure that the treatment of human subjects is in compliance with NASA Procedural Requirement (NPR) 7100.1C "Protection of Human Research Subjects" [§§] and Code of Federal Regulations Title 14, Part 1230, "Protection of Human Subjects" [¶¶]. Before taking part, participants were asked to sign an informed consent form confirming their agreement to the data collection and the ultimate use of the data by the general public.

For physical comfort, the use of a chair or cot in the poses described above were in place to protect the health and safety of human participants in any cold and potentially damp experimental conditions. Additionally, to maintain privacy, participants were given the option of facial coverings to hide their identity. The subjects were also visually scanned to make sure they had no personally identifiable information on them during data collection, such as employee IDs or clothing with easily recognizable designs.

Some non-standard procedures were also instituted for the purposes of this data collection. For physical safety, four humans were required to reposition the ATD due to its weight. Additionally, the human participants were provided with VHF radios and instructed to radio the operations crew if they became uncomfortable. The UAS would then either land or transition to a holding pattern while the participant cleared the test area.

IV. Creation of Prototype Dataset

In developing this dataset, our focus was to provide an initial starting point from which practitioners can begin to prototype, train, and test medical UAS models; the availability of such a domain-specific dataset would also provide a common baseline from which the medical UAS industry and related academic contributors could compare preliminary results. Therefore, the quality of the samples as well as their associated labels and metadata was of the utmost importance. Perfect execution would have resulted in the collection of approximately 72,000 still images and 60 one-minute videos after all test flights were completed. However, due to factors such as set-up and tear-down time, excluding samples due to poor quality, and other factors relevant to outdoor testing, we anticipate the final sample count being in the range of 50,000 - 70,000 images and approximately 50 videos. Although the initial size of this data set is small, it is considered appropriate for a first-pass data set according to subject matter experts in the data science field.

Post-processing raw data for use in a machine learning dataset is an involved, labor-intensive task. For a first-pass labeling effort, we used the You Only Look Once (YOLO) classifier [4] to detect when humans were in a given video frame. YOLO-based models have demonstrated success in vision tasks such as detecting humans with disabilities [23], evaluating posture and detecting falls [24], recognizing license plates [25], and identifying and tracking marine vessels [26]. Within the UAS research space, YOLO has shown promise in obstacle detection [27–29], human location tracking [30, 31], and other coarse-grained, first-pass perception tasks [32]. It therefore made sense to use this model as a coarse-grained way to separate useful frames from non-useful in the raw video feed.

We do acknowledge the possibility for error in relying on a non-human method for labeling data samples, but as this is only the first step in developing the dataset, we anticipate mitigating this risk in future iterations of data curation. The benefit in applying YOLO as a first-pass filter was in quickly identifying video frames of interest to a human, thus decreasing the labeling effort. For example, if a flight test configuration had a participant sitting, for every frame with a YOLO-detected human, we could relabel that data point "sitting" and add it to the training data. Figure 8 shows an example of a post-processed batch of training samples, where 0 indicates sitting, 1 indicates standing, and 2 indicates laying.

One interesting challenge in this process was that, in some videos, the pilot and other operations personnel related to the experiment were in the UAS's field of view. For these images, automatic labeling would not be successful since these personnel ran the risk of detection and thus being added to the data set with the wrong label. This required human intervention to check what areas of the video had personnel and to avoid collecting data from these sections of the videos.

[§§] https://nodis3.gsfc.nasa.gov/npg_img/N_PR_7100_001C_/N_PR_7100_001C_.pdf

[¶¶] https://www.ecfr.gov/current/title-14/chapter-V/part-1230

Fig. 8 Example of a post-processed training batch where 0 indicates sitting, 1 indicates standing, and 2 indicates laying.

Fig. 9 Example of a validation batch with model predictions and confidence values.

Table 1 Description of YOLOv8 hyper-parameters used in training the medical UAS model. All other parameters were kept at default values as designated by API documentation[†††].

PARAM	DESCRIPTION	DEFAULT VALUE	ASSIGNED VALUE	PHASES USED
epochs	Total number of training epochs	100	10	Train
batch	Number of samples to include in each training batch	16	8	Train, Val
imgsz	Size of input image (assumed square)	640	3840	Train, Val
conf	Minimum detection confidence threshold	0.001	Null	Val
iou	Intersection Over Union (IOU) threshold (reduces duplicate detections)	0.6	0.7	Val
half	Enabling flag for half-precision computation (resource consumption saving mode)	True	False	Val
kobj	Keypoint "objectness" weight (trade-off between confidence and accuracy)	2.0	1.0	Train

As this paper describes ongoing work, we are still in the process of cleaning and curating the collected data. The proof-of-concept model described in Section V was trained on a dataset of almost 1600 samples with appreciable validation accuracy (shown in Figure 9). As time progresses, more data will be cleaned and added to the dataset so that more mature models may be developed and deployed.

V. Perception Model Proof-of-Concept

As the focus of this work was on the creation and curation of a specialized dataset, we elected to use an existing, high-performing neural network model rather than creating our own from scratch. We selected You Only Look Once (YOLO) as our base model for its versatility and performance. The specific model we selected was YOLOv8[***], the architecture for which is reflected in Figure 10. The architecture is split into three portions - the *backbone*, which performs the general feature extraction tasks common to most YOLO versions; the *neck*, which aggregates features from different layers in preparation for final output; and the *head*, which generates the final output. For YOLOv8, the head consists of a set of detection blocks to process inputs of different sizes; each block then contains a pair of pipelines which calculate the bounding box (BBox) and classification (Class) results in parallel, thereby boosting processing efficiency [33]. A description of the unique hyper-parameter values used in training are described in Table 1; default values were used for all other parameters.

The first round of data cleaning resulted in a preliminary dataset of 1598 images. Of these, 1021 images were in the "standing" pose, 347 were "sitting", and 230 were "laying". For training, 1442 images were used and the remainder reserved for validation. The best-case and macro prediction metrics observed during the training process are displayed in Figures 11 and 12 respectively. Performance results after training was complete are shown in Figure 13. The model achieved high scores in precision and recall both during and after training, with a best-case F1-confidence score of 0.85 and a recall of 0.904 when mean average precision (mAP) was 0.5. These are encouraging initial results, which we expect to improve upon as the dataset grows.

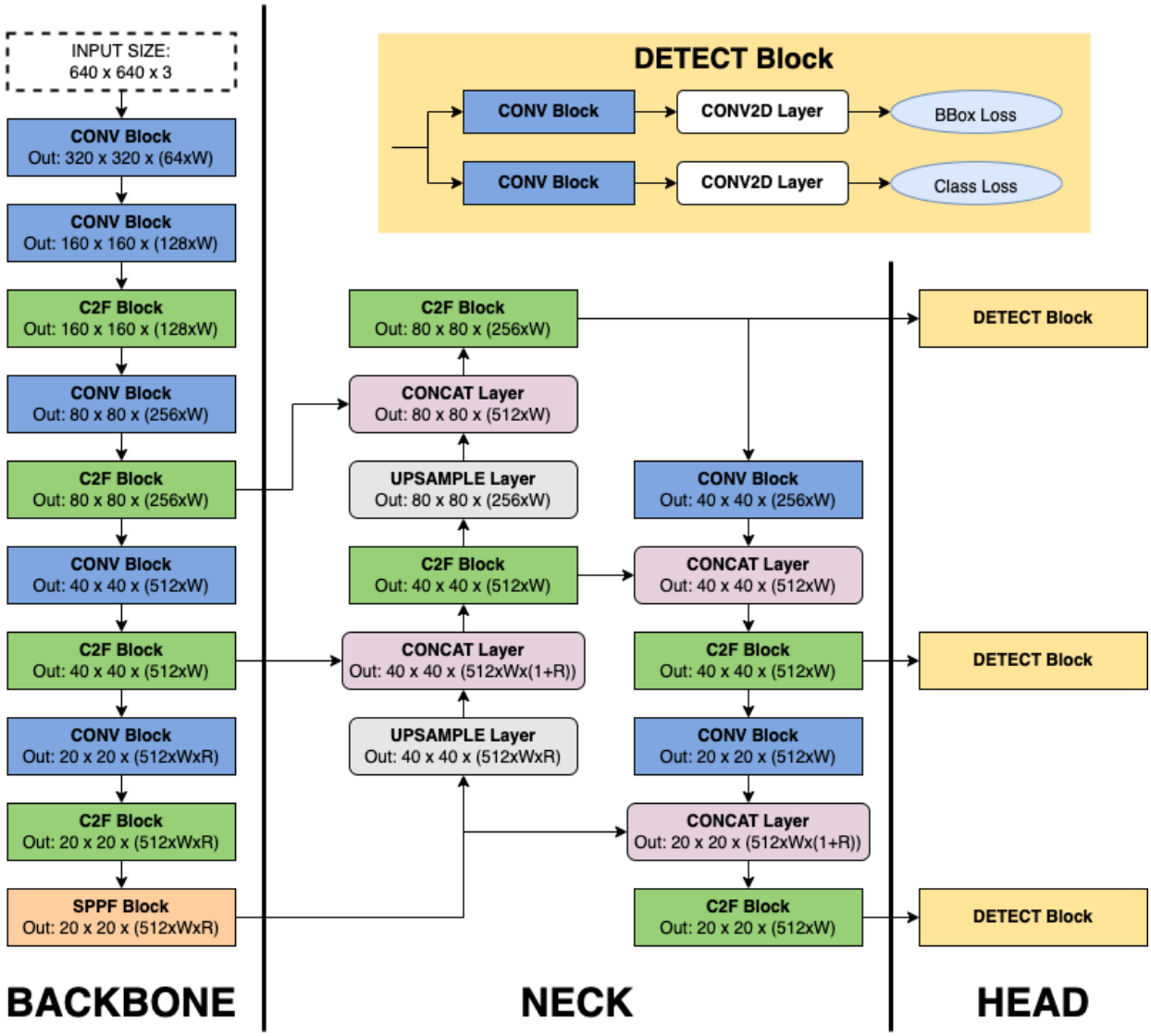
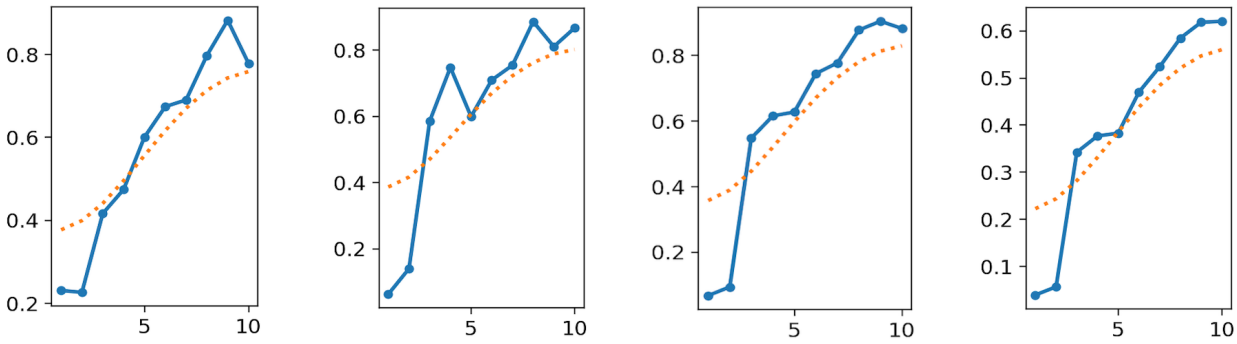
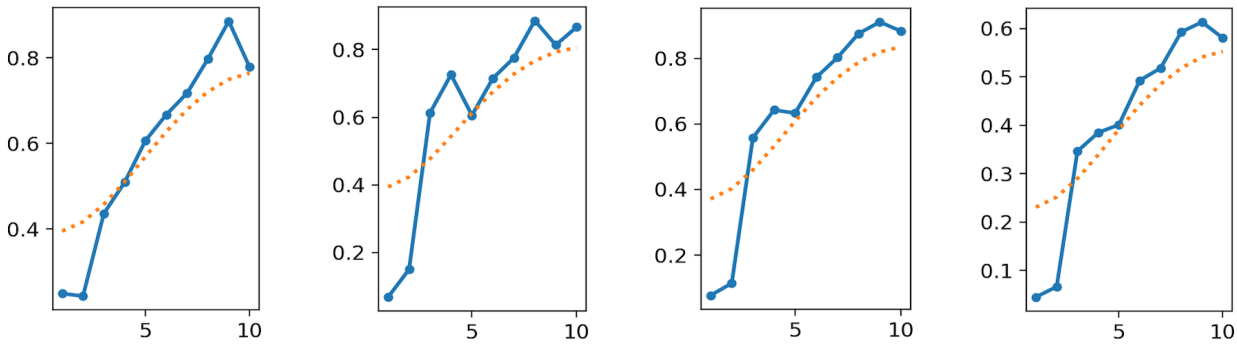


Fig. 10 YOLO v8 architectural diagram. Note that all data sizes are described in (Height, Width, Channel) tuples where the "Channel" portion is parameterized by the width multiplier (W) and ratio (R), allowing the network to scale. For the purposes of this work, we selected the "nano" sized model for which $W = 0.25$ and $R = 2.00$.



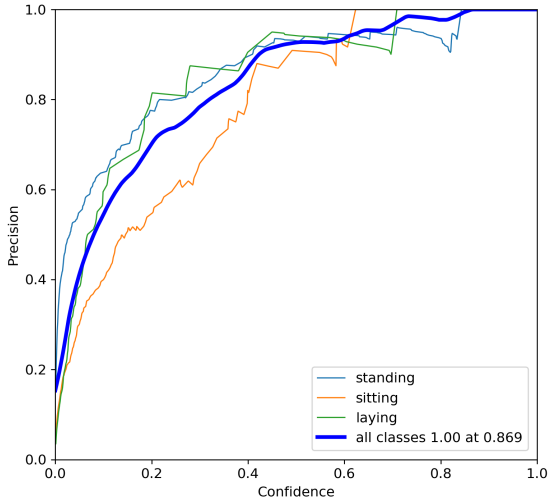
(a) Precision **(b) Recall** **(c) mAP @ 50** **(d) mAP @ 50-95**

Fig. 11 Best prediction metric performance across ten training epochs, where the solid line represents the raw curve and the dotted line represents the smoothed curve.

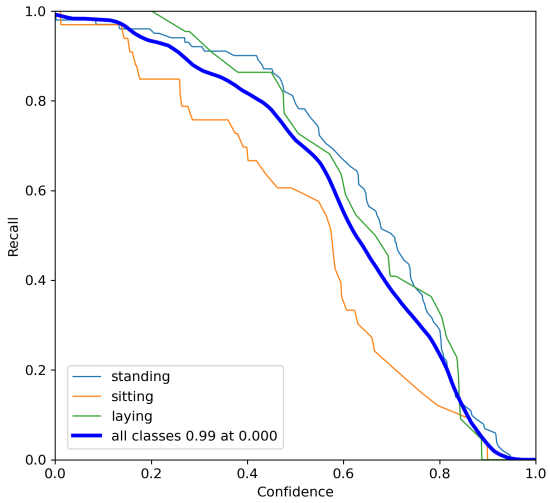


(a) Precision **(b) Recall** **(c) mAP @ 50** **(d) mAP @ 50-95**

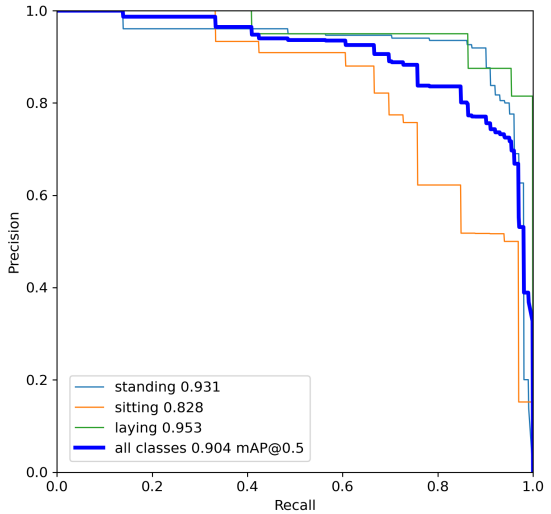
Fig. 12 Macro prediction metric performance across ten training epochs, where the solid line represents the raw curve and the dotted line represents the smoothed curve.



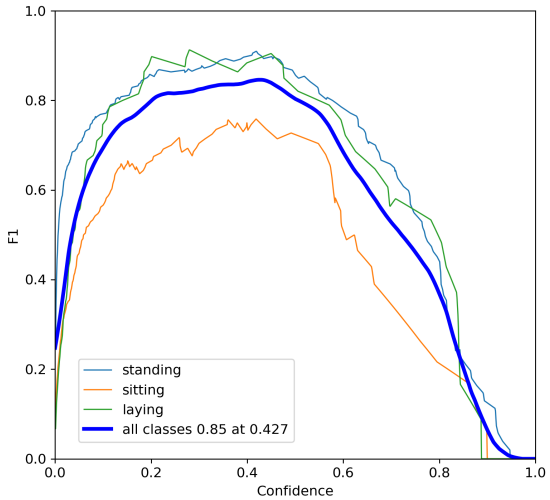
(a) Precision-Confidence Curve



(b) Recall-Confidence Curve



(c) Precision-Recall Curve



(d) F1-Confidence Curve

Fig. 13 Expanded post-training performance results for YOLOv8 model trained on our collected dataset.

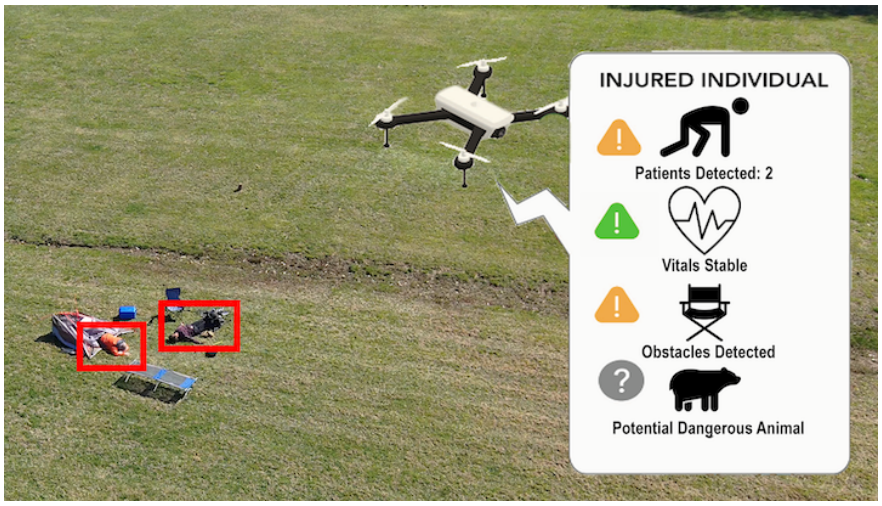


Fig. 14 An example of possible future medical UAS operations, where the vehicle autonomously scans the scene and reports relevant information back to a first responder coordinator.

VI. Future Work and Conclusions

This paper detailed ongoing work for collection and processing of a prototype dataset for low-flying, medical UAS use case. This included the experimentation process, data collection, and data processing. The paper also included a nominal classifier trained on a subset of this data and transfer learned from the YOLO computer vision system. This work is ongoing, and although approximately 1600 images have been labeled so far, there is still work to complete the dataset that will be made publicly available. This includes processing all 30 of the raw videos into classified images complete with all available metadata, such as altitude, horizontal distance, camera angle, subject information, etc. The next major goal in this effort is to finish labeling this dataset and make it publicly available.

Development of this dataset is merely the first step toward creating a larger, more robust machine learning apparatus for medical UAS use cases, such as the example shown in Figure 14. The scope for this particular paper has been kept narrow to allow realistic completion of the effort while also setting up future expansion efforts for success. In the future, we will explore the possibility of crowd-sourcing new additions to the dataset, similar to the Airvue project at NASA. Future work will also seek to identify stakeholders in industry and academia with emergency medical UAS use cases that would be enhanced or improved with aerial perception. Additionally, this prototype dataset is intended for future work in creating assured classifiers for safety-critical applications. The data set and nominal classifier developed in this work will serve as an artifact for neural network assurance metrics being developed by the Aerial **AI**d team.

Acknowledgments

The authors thank the Convergent Aeronautics Solutions (CAS) Project at NASA for sponsoring the exploratory phase of the Aerial **AI**d activity.

References

[1] Johnson, A. M., Cunningham, C. J., Arnold, E., Rosamond, W. D., and Zègre-Hemsey, J. K., "Impact of using drones in emergency medicine: What does the future hold?" *Open Access Emergency Medicine*, 2021, pp. 487–498.

[2] Brown, N., "AIRVUE (Crowdsourcing) Airborne Instrumentation for Real-world Video of Urban Environments," *Transformational Tools and Technologies (T3) Project*, 2023.

[3] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L., "Microsoft coco: Common objects in context," *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, Springer, 2014, pp. 740–755.

***`https://yolov8.com`

[4] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[5] Cohn, P., Green, A., Langstaff, M., and Roller, M., "Commercial drones are here: The future of unmanned aerial systems," *McKinsey & Company*, 2017.

[6] Roberts, N. B., Ager, E., Leith, T., Lott, I., Mason-Maready, M., Nix, T., Gottula, A., Hunt, N., and Brent, C., "Current summary of the evidence in drone-based emergency medical services care," *Resuscitation Plus*, Vol. 13, 2023, p. 100347.

[7] Pulsiri, N., and Vatananan-Thesenvitz, R., "Drones in emergency medical services: A systematic literature review with bibliometric analysis," *International Journal of Innovation and Technology Management*, Vol. 18, No. 04, 2021, p. 2097001.

[8] Robakowska, M., Ślęzak, D., Żuratyński, P., Tyrańska-Fobke, A., Robakowski, P., Prędkiewicz, P., and Zorena, K., "Possibilities of using UAVs in pre-hospital security for medical emergencies," *International Journal of Environmental Research and Public Health*, Vol. 19, No. 17, 2022, p. 10754.

[9] Pulsiri, N., Vatananan-Thesenvitz, R., Sirisamutr, T., and Wachiradilok, P., "Save lives: a review of ambulance technologies in pre-hospital emergency medical services," *2019 Portland International Conference on Management of Engineering and Technology (PICMET)*, IEEE, 2019, pp. 1–10.

[10] Pulver, A., Wei, R., and Mann, C., "Locating AED enabled medical drones to enhance cardiac arrest response times," *Prehospital Emergency Care*, Vol. 20, No. 3, 2016, pp. 378–389.

[11] Al-Wathinani, A. M., Alhallaf, M. A., Borowska-Stefańska, M., Wiśniewski, S., Sultan, M. A. S., Samman, O. Y., Alobaid, A. M., Althunayyan, S. M., and Goniewicz, K., "Elevating healthcare: Rapid literature review on drone applications for streamlining disaster management and prehospital care in Saudi Arabia," *Healthcare*, Vol. 11, MDPI, 2023, p. 1575.

[12] Abbas, N., Abbas, Z., Liu, X., Khan, S. S., Foster, E. D., and Larkin, S., "A survey: future smart cities based on advance control of Unmanned Aerial Vehicles (UAVs)," *Applied Sciences*, Vol. 13, No. 17, 2023, p. 9881.

[13] Braun, J., Gertz, S. D., Furer, A., Bader, T., Frenkel, H., Chen, J., Glassberg, E., and Nachman, D., "The promising future of drones in prehospital medical care and its application to battlefield medicine," *Journal of trauma and acute care surgery*, Vol. 87, No. 1S, 2019, pp. S28–S34.

[14] Zhu, P., Wen, L., Bian, X., Haibin, L., and Hu, Q., "Vision Meets Drones: A Challenge," *arXiv preprint:1804.07437*, 2018.

[15] Wen, L., Du, D., Zhu, P., Hu, Q., Wang, Q., Bo, L., and Lyu, S., "Detection, Tracking, and Counting Meets Drones in Crowds: A Benchmark," *CVPR*, 2021.

[16] Schröder, G., Senst, T., Bochinski, E., and Sikora, T., "Optical flow dataset and benchmark for visual crowd analysis," *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2018, pp. 1–6.

[17] Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., and Tian, Q., "The unmanned aerial vehicle benchmark: Object detection and tracking," *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 370–386.

[18] Kyrkou, C., and Theocharides, T., "Deep-Learning-Based Aerial Image Classification for Emergency Response Applications Using Unmanned Aerial Vehicles." *CVPR workshops*, 2019, pp. 517–525.

[19] Kyrkou, C., and Theocharides, T., "EmergencyNet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 13, 2020, pp. 1687–1699.

[20] Rahnemoonfar, M., Chowdhury, T., and Murphy, R., "RescueNet: A high resolution UAV semantic segmentation benchmark dataset for natural disaster damage assessment," *arXiv preprint arXiv:2202.12361*, 2022.

[21] Lyu, Y., Vosselman, G., Xia, G.-S., Yilmaz, A., and Yang, M. Y., "UAVid: A semantic segmentation dataset for UAV imagery," *ISPRS journal of photogrammetry and remote sensing*, Vol. 165, 2020, pp. 108–119.

[22] Russell Bernal, A. M., Scheirer, W., and Cleland-Huang, J., "NOMAD: A Natural, Occluded, Multi-scale Aerial Dataset, for Emergency Response Scenarios," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 8584–8595.

[23] Alruwaili, M., Atta, M. N., Siddiqi, M. H., Khan, A., Khan, A., Alhwaiti, Y., and Alanazi, S., "Deep Learning-Based YOLO Models for the Detection of People With Disabilities," *IEEE Access*, 2023.

[24] Tran, T.-H., Nguyen, D. T., and Nguyen, T. P., "Human posture classification from multiple viewpoints and application for fall detection," *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, IEEE, 2021, pp. 262–267.

[25] Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., and Menotti, D., "A robust real-time automatic license plate recognition based on the YOLO detector," *2018 international joint conference on neural networks (ijcnn)*, IEEE, 2018, pp. 1–10.

[26] Han, X., Zhao, L., Ning, Y., and Hu, J., "ShipYolo: an enhanced model for ship detection," *Journal of Advanced Transportation*, Vol. 2021, No. 1, 2021, p. 1060182.

[27] Liu, M., Wang, X., Zhou, A., Fu, X., Ma, Y., and Piao, C., "Uav-yolo: Small object detection on unmanned aerial vehicle perspective," *Sensors*, Vol. 20, No. 8, 2020, p. 2238.

[28] Lalak, M., and Wierzbicki, D., "Automated detection of atypical aviation obstacles from UAV images using a YOLO algorithm," *Sensors*, Vol. 22, No. 17, 2022, p. 6611.

[29] Hu, Y., Wu, X., Zheng, G., and Liu, X., "Object detection of UAV for anti-UAV based on improved YOLO v3," *2019 Chinese Control Conference (CCC)*, IEEE, 2019, pp. 8386–8390.

[30] Boudjit, K., and Ramzan, N., "Human detection based on deep learning YOLO-v2 for real-time UAV applications," *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 34, No. 3, 2022, pp. 527–544.

[31] Zhang Sr, D., Shao, Y., Mei, Y., Chu, H., Zhang, X., Zhan, H., and Rao, Y., "Using YOLO-based pedestrian detection for monitoring UAV," *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*, Vol. 11069, SPIE, 2019, pp. 1141–1145.

[32] Chen, C., Zheng, Z., Xu, T., Guo, S., Feng, S., Yao, W., and Lan, Y., "Yolo-based uav technology: A review of the research and its applications," *Drones*, Vol. 7, No. 3, 2023, p. 190.

[33] Dsilva, L. R., Tantri, S. H., Sampathila, N., Mayrose, H., Muralidhar Bairy, G., Belurkar, S., Saravu, K., Chadaga, K., and Hafeez-Baig, A., "Wavelet scattering-and object detection-based computer vision for identifying dengue from peripheral blood microscopy," *International Journal of Imaging Systems and Technology*, Vol. 34, No. 1, 2024, p. e23020.