A2I2-Haze: A multi-purpose real haze benchmark with quantifiable haze levels and ground-truths

I. Introduction to A2I2-Haze Dataset

In A2I2-Haze, we leverage US Army's unique capability to produce and measure nonlethal smoke/obscurant munitions to generate hazy condition in a controlled fashion. We collect imagery and metadata from unmanned air vehicles (UAVs) and stationary sensors of target objects such as civilian vehicles, mannequins, and potential man-made obstacles encountered during UGV maneuver. We then develop an image dataset with metadata for realistic, accurate and fine-grained algorithm evaluation in hazy DVEs.

The A2I2-Haze Dataset consists of paired haze, haze-free aerial imagery acquired by a small UAS. The aerial data set is synchronized with insitu smoke measurements and altitude measurement acquired from the flight controller.

Data collection was conducted at the DEVCOM Chemical and Biological Center's (CBC) M-field test range utilizing the world's most comprehensive obscuration assessment methodologies and facilities to accurately measure the smoke concentration. These measurements were time synchronized with the UAV and UGV sensors to relate the obscuration properties to the sensor's response.

The facility has a platoon of six M56E1 Smoke Generating systems that provide large area obscuration by disseminating obscurants either simultaneously or separately while stationary or mobile. The current on-board obscuration capabilities include visual obscuration using fog oil, infrared obscuration using graphite, and radar obscuration using carbon fiber. These obscurants thus have the ability to degrade the perception capabilities of threat weapons and Reconnaissance, Intelligence, Surveillance, Targeting and Acquisition (RISTA) operating in the visual, IR, and MMW portions of the electromagnetic spectrum. Specifically the obscurant materials are designed to effectively absorb and scatter energy making target acquisition difficult.

As part of the Phase 1 DVE data collection, ARL collected a wide range of imagery of non-military targets in the DEVCOM CBC M-Field test area using visual obscurants (fog oil) generated with the M56E1 Smoke Generating System. The target objects included civilian vehicles, mannequins, and potential man-made obstacles such as traffic cones, barriers, barricades, etc. Imagery and metadata were collected from aerial, ground platforms and stationary sensors.

In summary, the contribution of A2I2-Haze are as follows:

a) We present A2I2-Haze, the first real haze dataset with in situ smoke measurement aligned to aerial imagery. This multi-purpose dataset has paired haze, haze-free imagery that will allow fine-grained evaluation of computer vision algorithms.

b) We perform comprehensive study and evaluation of state-of-the-art single image dehazing and object detection algorithms using A2I2 Haze.

II. Quantitative Evaluation on Dehazing

A. Baseline Approaches

We test three state-of-the-art homogeneous dehazing approaches: (1) **GCANet**; (2) **FFA-Net**; and (3) **MSBDN**. We also test three state-of-the-art non-homogeneous approaches: (4) **SRKT**; (5) **DWDehaze**; and **Trident-Dehazing-Network**. All dehazing models adopt official implementation and pretrained models.

We employ detection results as semantic metrics to evaluate performance of dehazing algorithms. We use four state of the art detectors: (a) **YOLOv5**; (b) **YOLOX**; (c) **faster R-CNN**; and (d) **CenterNet**; For comparisons, we re-train these detectors on **VisDrone2019-DET**, **UAVDT-M** and a mix-up dataset of **VisDrone2019-DET** and **UAVDT-M**. All detectors adopt official COCO based pretrained models.

B. Evaluation and Analysis

Fig. 1 shows detection examples of original hazy images and dehazed images. Table I is mAP score comparison on different detectors, different dehazing algorithms and different datasets Fig. 1 argues different dehazing algorithms show different effects on detectors. Non homogeneous dehazing algorithms have better mAP scores than homogeneous algorithm. Mostly, FFA-Net performs worst on detection results. Trident-Dehazing-Netwwork has a best mAP score on YOLOv5, faster R-CNN and CenterNet.

Detectors performs differently on different datasets. As Table. I shows, detectors degrade significant detection performance on UAVDT dataset. We analyze that UAVDT dataset consists of lots of small size objects, while detection objects in test dataset are in medium and large size. A large domain gap exists between train and test datasets. Since YOLOX is more sensitive on data, detector will tend to detect small object as target and thus perform worse when UAVDT dataset is included.

We test four detectors on proposed dataset. For fair comparison, all detectors are trained on VisDrone2019-DET dataset. The overall detection performance has a mAP of 39.4% on hazy images using YOLOv5, 34.2% using faster R-CNN, 40.9% using CenterNet and 38.9% using YOLOX. Faster R-CNN and CenterNet are more stable on different datasets compared with YOLOv5 and YOLOX.

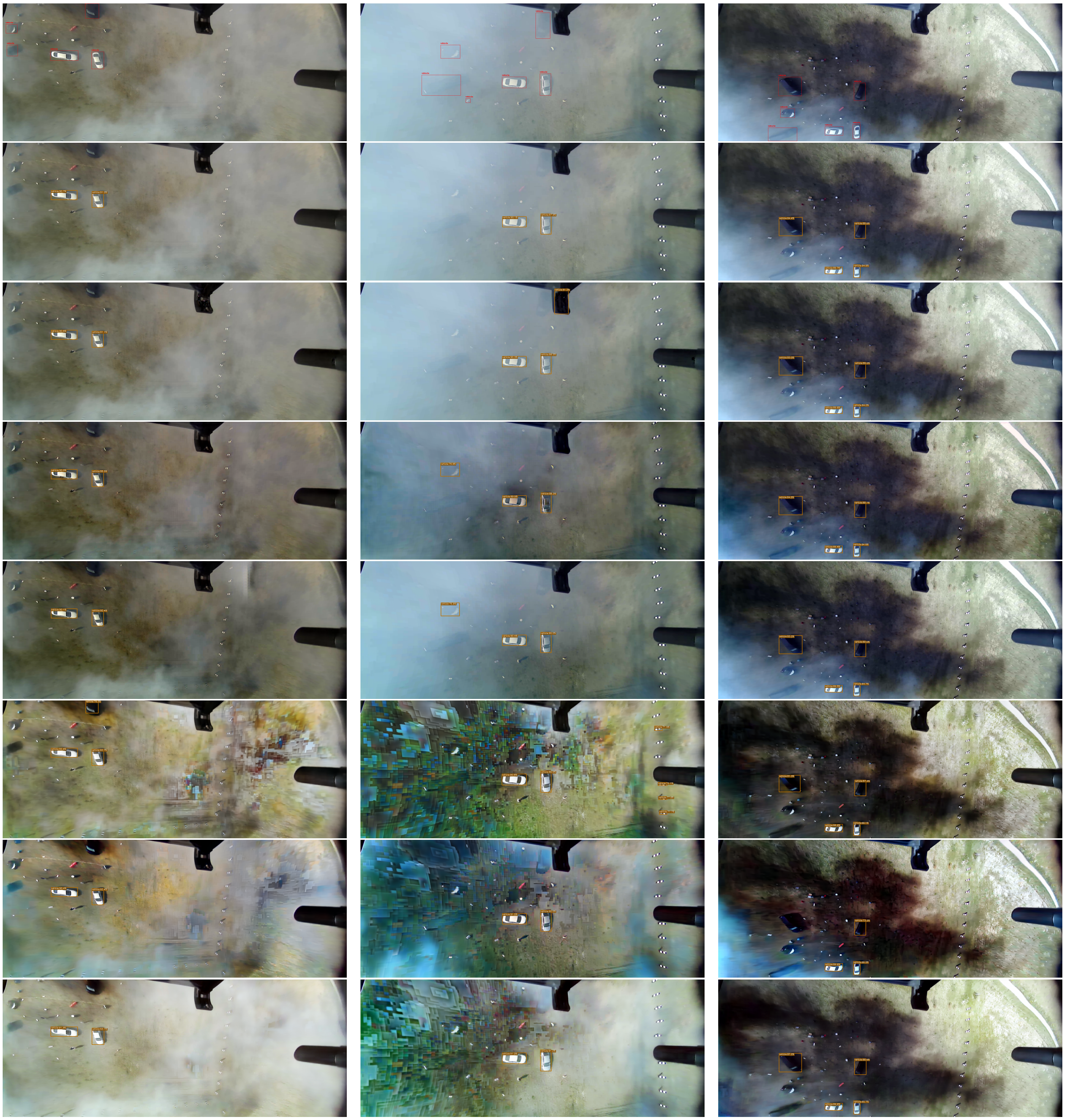

Fig. 1: Examples of Object detection of hazy images and dehazed images. All detection results are from YOLOX-m re-trained on mix up dataset. From top to bottom: ground-truth, no dehazing, FFA-Net, GCANet, MSBDN, SRKT, Trident and DWDehazing.

TABLE I: Detection Results With Different Dehazing Approaches

Pretraining	Dehazing	YOLOv5			FasterRCNN+FPN			CenterNet			YOLOX		
		AR	$AP_{0.5}$	$AP_{0.5:0.95}$	AR	$AP_{0.5}$	$AP_{0.5:0.95}$	AR	$AP_{0.5}$	$AP_{0.5:0.95}$	AR	$AP_{0.5}$	$AP_{0.5:0.95}$
VisDrone	hazy	0.410	0.469	0.394	0.380	0.463	0.342	0.546	0.591	0.409	0.415	0.500	0.389
	GCANet	0.424	0.480	0.411	0.438	0.515	0.390	0.568	0.614	0.440	0.412	0.485	0.383
	FFA-Net	0.402	0.455	0.383	0.395	0.461	0.346	0.557	0.560	0.397	0.416	0.497	0.387
	MSBDN	0.430	0.490	0.417	0.430	0.512	0.386	0.568	0.625	0.443	0.421	0.495	0.387
	Trident	**0.503**	**0.574**	**0.482**	**0.489**	0.556	**0.414**	**0.596**	**0.678**	**0.486**	0.383	0.432	0.333
	SRKT	0.491	0.567	0.470	0.487	0.550	0.406	0.573	0.635	0.451	0.405	0.452	0.356
	DWDehaze	0.487	0.547	0.466	0.445	**0.568**	0.395	0.592	0.650	0.466	**0.450**	**0.508**	**0.411**
UAVDT	hazy	0.099	0.117	0.096	0.216	0.264	0.196	0.493	0.457	0.323	0.296	0.325	0.256
	GCANet	0.142	0.176	0.139	0.245	0.310	0.222	0.520	0.503	0.361	0.279	0.295	0.224
	FFA-Net	0.115	0.138	0.113	0.218	0.149	0.181	0.506	0.457	0.326	0.299	**0.328**	**0.256**
	MSBDN	0.145	0.176	0.145	0.240	0.301	0.219	0.520	0.510	0.363	0.288	0.307	0.240
	Trident	**0.198**	**0.242**	**0.188**	**0.358**	**0.455**	**0.311**	**0.557**	0.554	0.395	0.307	0.296	0.217
	SRKT	0.166	0.207	0.163	0.299	0.384	0.268	0.541	**0.570**	**0.402**	**0.309**	0.280	0.214
	DWDehaze	0.145	0.188	0.143	0.306	0.281	0.259	0.550	0.561	0.390	0.287	0.303	0.223
UAVDT+VisDrone	hazy	0.427	0.491	0.411	0.341	0.434	0.305	0.543	0.586	0.423	0.384	0.460	0.370
	GCANet	0.463	0.530	0.447	0.360	0.467	0.328	0.580	0.604	0.446	0.387	0.458	0.370
	FFA-Net	0.439	0.498	0.418	0.359	0.424	0.302	0.551	0.579	0.418	0.382	0.456	0.366
	MSBDN	0.466	0.541	0.452	0.392	0.504	0.354	0.571	0.627	0.456	0.379	0.450	0.364
	Trident	**0.533**	**0.626**	**0.510**	0.491	0.634	0.426	0.594	**0.683**	**0.489**	0.386	0.456	0.367
	SRKT	0.519	0.608	0.501	**0.500**	**0.641**	**0.437**	0.586	0.668	0.483	0.374	0.442	0.355
	DWDehaze	0.504	0.580	0.487	0.489	0.621	0.420	**0.596**	0.671	0.484	**0.431**	**0.506**	**0.410**