The Aircraft Context Dataset: Understanding and Optimizing Data Variability in Aerial Domains
Supplementary Material

Daniel Steininger   Verena Widhalm   Julia Simon   Andreas Kriegler   Christoph Sulzbacher
AIT Austrian Institute of Technology
{firstname.surname}@ait.ac.at

1. Meta-annotation of data variability

To illustrate the range of available meta-annotations described in section 3.2 of the main paper, examples for the values assignable to each variability parameter are given in Figures 1 and 2 for both subsets.

2. Semantic labeling

Section 5.4 of the main paper shows the results of semantic-labeling experiments conducted on the *Seg3* variants of both subsets. This setup was derived by initially training models on dataset variants containing five target classes (denoted as MAV_{Seg5} and UAV_{Seg5}) which resulted in the performance presented in Table 1.

	$\mathbf{MAV_{Seg5}}$	$\mathbf{UAV_{Seg5}}$
Aircraft	.772	.642
Sky	.957	.931
Veg	.820	.824
Runway/Apron	.658	.490
Building	.318	.001
Overall	**.705**	**.578**

Table 1. Per-class and overall semantic-labeling results (*mIoU*) on the initial dataset variants.

3. Evaluation of classification experiments

Table 2 shows the precision and recall metrics used as a basis for computing the F1-Scores of CLS_{Ext} experiments reported in Table 2 of the main paper.

4. Impact analysis of data variability

Table 3 of the main paper shows the deviations of classification model performances for each variability parameter. This includes pure localization by the original detection module on the *Coarse1* dataset variants (*LOC*), as well as a combination of these results with the external classification model (*CLS*). The latter values are averaged across the results on the individual dataset variants for each subset, which are displayed in Table 3.

5. Qualitative results

To give a more comprehensive overview of the results discussed in Section 5 of the main paper, Figure 3 presents a composition of qualitative results under varying environmental conditions extracted from the test set. Each sample includes localization, classification by multiple model variants and preliminary semantic-labeling results.

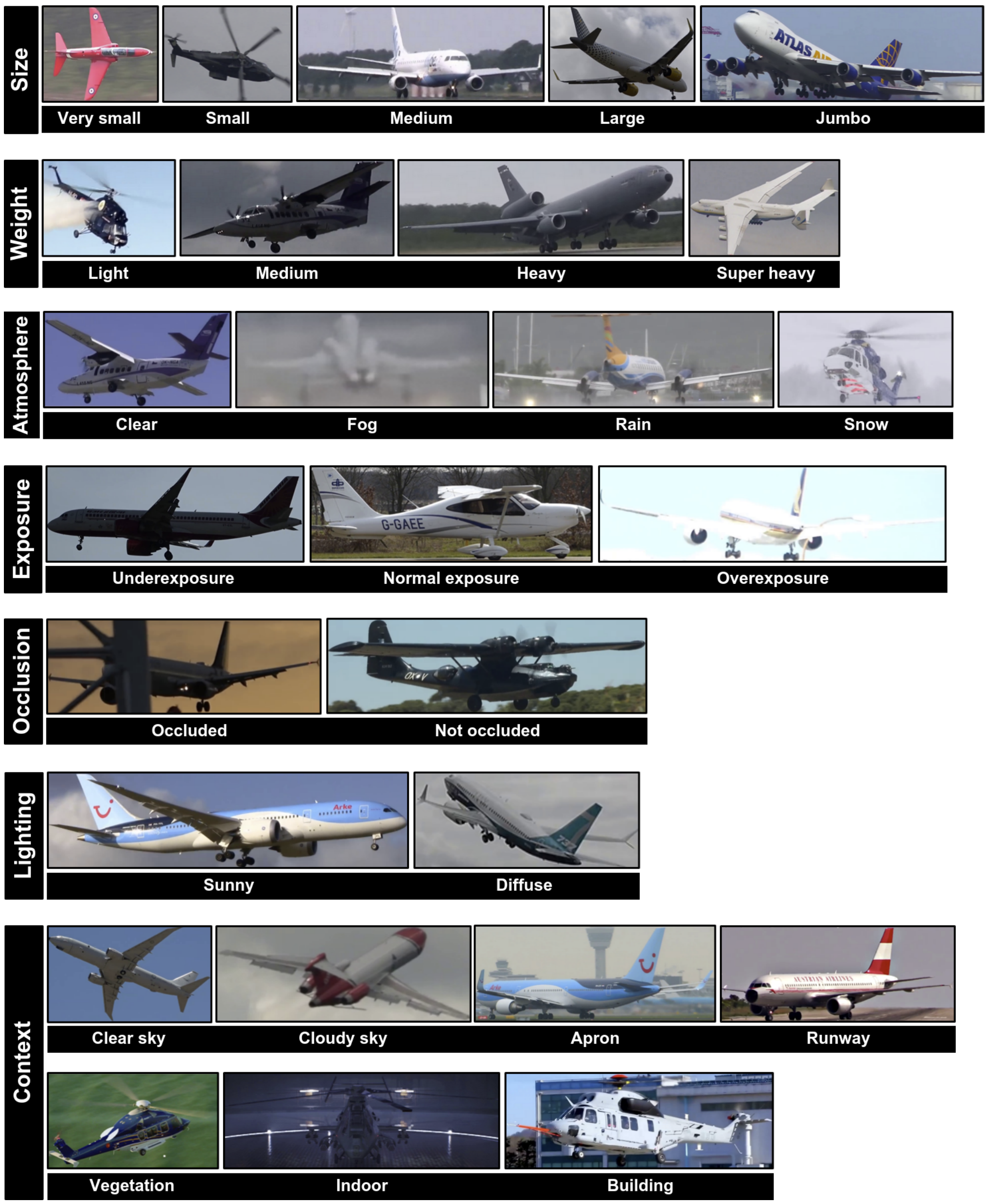

Figure 1. Representative samples of each variability parameter for the *MAV* subset.

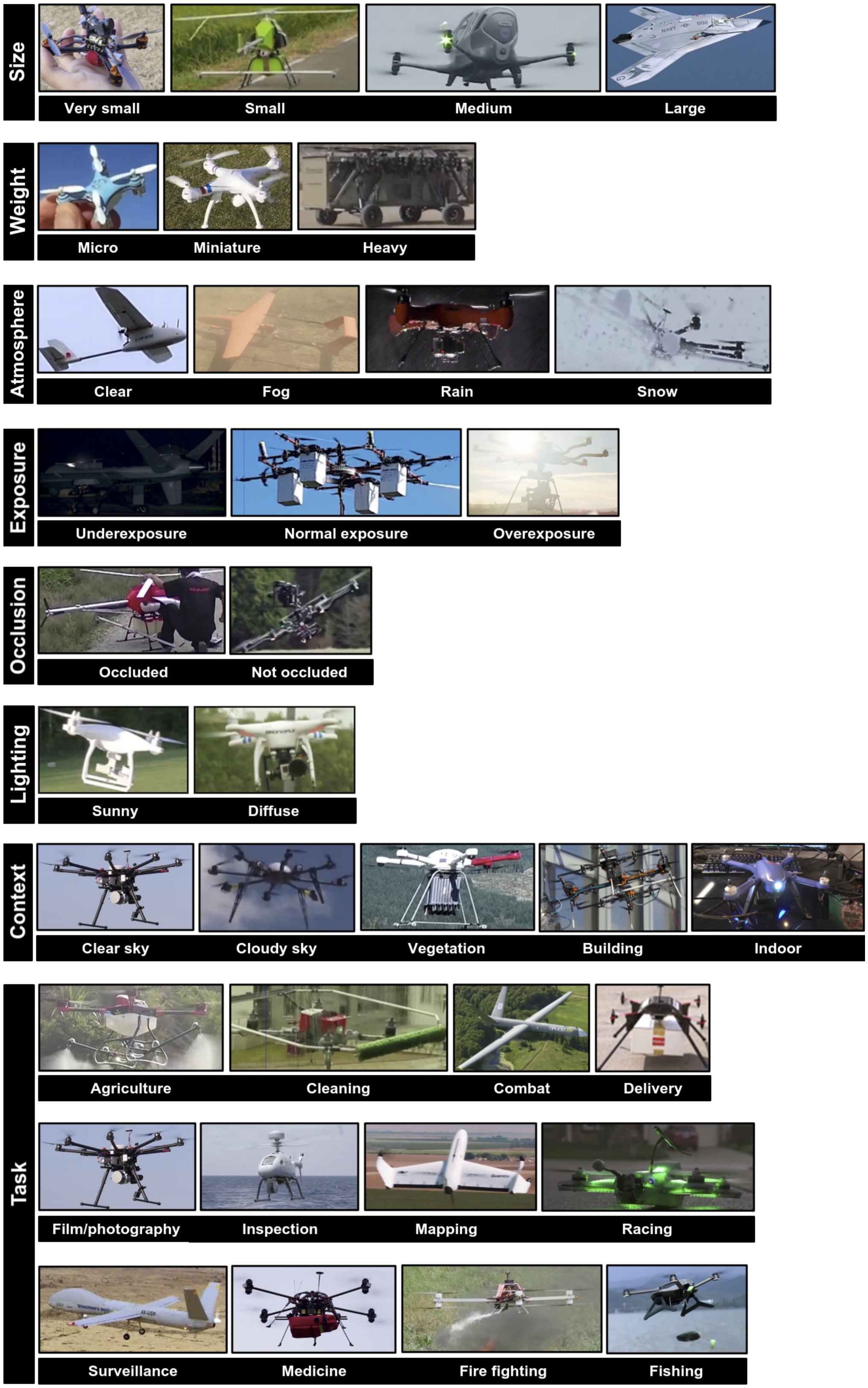

Figure 2. Representative samples of each variability parameter for the *UAV* subset.

	$\mathbf{MAV_{Fine14}}$	$\mathbf{MAV_{Domain5}}$	$\mathbf{MAV_{Prop3}}$	$\mathbf{MAV_{Air2}}$	$\mathbf{MAV_{Coarse1}}$	$\mathbf{UAV_{Fine9}}$	$\mathbf{UAV_{Domain3}}$	$\mathbf{UAV_{Prop3}}$	$\mathbf{UAV_{Air2}}$	$\mathbf{UAV_{Coarse1}}$	$\mathbf{AC_{Fine23}}$	$\mathbf{AC_{Air2}}$	$\mathbf{AC_{Coarse2}}$	$\mathbf{AC_{Coarse1}}$
mean Precision	.738	.783	.952	.947	.986	.706	.772	.924	.875	.986	.706	.941	.926	.997
mean Recall	.741	.836	.957	.941	.987	.732	.774	.927	.880	.986	.700	.936	.925	.996
mean F1-Score	.739	.806	.955	.944	.986	.715	.773	.925	.878	.986	.697	.938	.926	.996

Table 2. Classification results for experiments on all dataset variants.

	State		**Atmo**		**Object context**				**Degradation**			**Lighting**		**Occlusion**	
	ar	*nar*	*cla*	*fog*	*clr*	*cld*	*veg*	*bld*	*ndg*	*ldg*	*hdg*	*sun*	*dif*	*noc*	*oc*
MAV_{F14}	-1.1	-3.0	0.8	-7.0	-3.2	-2.7	2.9	**8.7**	0.2	2.0	**-20.6**	0.4	2.9	3.3	-10.5
MAV_{D5}	-1.6	0.6	2.3	-13.1	-4.8	-0.7	-2.0	**6.2**	4.8	0.5	-9.8	1.8	-2.4	5.3	**-14.3**
MAV_{P3}	3.3	-2.7	5.1	-3.7	5.9	5.7	2.5	-2.5	**6.7**	1.8	**-23.2**	1.1	1.5	5.0	-8.4
MAV_{A2}			6.2	-5.7	4.6	-6.0	-2.1	-1.3	**8.1**	0.3	**-16.0**	2.2	-0.9	5.9	-4.1
MAV_{C1}	5.6	-4.4	4.5	-4.7	**8.2**	6.5	3.0	-0.2	7.5	3.4	**-15.3**	1.8	0.6	6.5	-3.7
UAV_{F9}	2.0	-5.0	-1.5	6.3	**10.6**	6.7	-7.2	-16.8	0.2	8.7	-3.7	-1.1	3.3	1.4	**-22.5**
UAV_{D3}	0.6	-5.0	4.0	**-21.4**	5.4	3.7	-2.3	**16.5**	1.6	9.3	-7.3	1.6	-0.3	1.3	-14.9
UAV_{P3}	1.7	-9.1	-0.8	**11.7**	5.8	7.7	-0.2	-8.0	-5.7	7.0	-0.3	-0.4	1.0	2.3	**-30.6**
UAV_{A2}			-1.2	-18.3	-11.3	**2.0**	-3.4	-8.8	-0.1	-2.3	-4.7	0.8	-4.1	1.5	**-18.6**
UAV_{C1}	2.4	-12.4	-0.3	-4.9	7.8	**8.7**	0.4	-3.2	-6.4	6.8	-1.0	1.3	-1.0	2.4	**-22.4**
AC_{F23}	-0.4	-1.0	3.4	-9.3	-1.1	**5.6**	-0.1	-6.9	4.6	2.2	-10.7	0.6	0.5	1.9	**-15.9**
AC_{A2}			5.8	-0.7	-8.9	-9.8	-2.1	-2.5	**10.2**	2.8	**-23.0**	3.9	-1.8	6.6	-4.1
AC_{C2}	4.2	-6.8	5.4	-9.7	**10.0**	3.6	3.8	-11.1	6.6	2.6	-12.9	3.9	-1.4	4.5	**-13.3**
AC_{C1}	6.3	-6.8	4.7	0.0	**12.2**	9.1	2.2	0.2	8.9	6.2	**-9.9**	3.8	-0.5	5.6	-2.9

Table 3. Influence of variability parameters on model performance as absolute *mAP* variation for experiments combining localization on *Coarse1* with external classification for each dataset variant: airborne (*ar*) and non-airborne (*nar*) state, clear (*cla*) and foggy (*fog*) atmosphere, sky-clear (*clr*), sky-cloudy (*cld*), vegetation (*veg*) and building (*bld*), object context, no (*ndg*), low (*ldg*) and high (*hdg*) image degradation, sunny (*sun*) and diffuse (*dif*) lighting, non-occluded (*noc*) and occluded (*oc*) object.

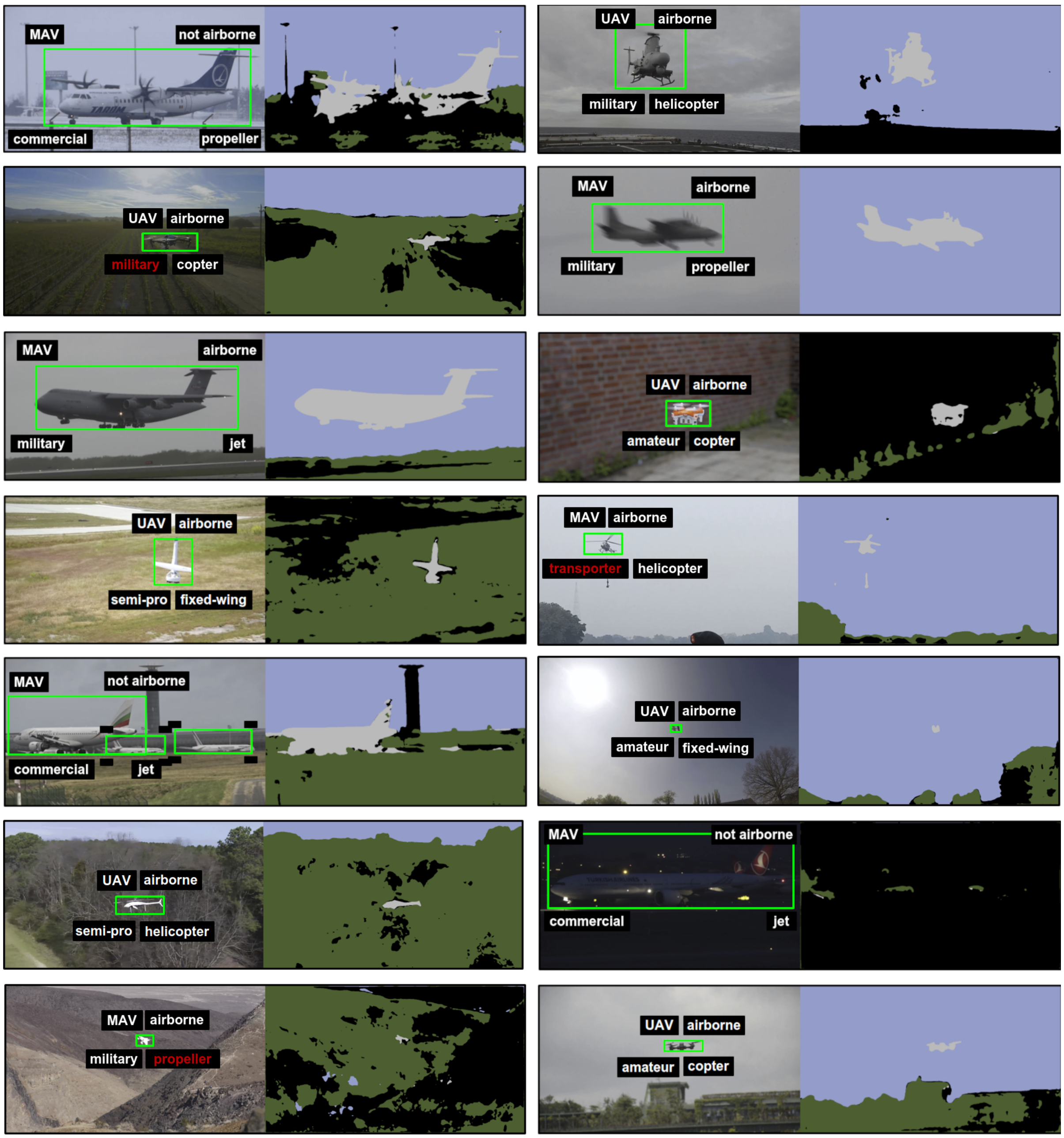

Figure 3. Representative selection of qualitative results including object localization and multiple classification variants (top left: $AC_{Coarse2}$, top right: MAV_{Air2}/UAV_{Air2}, bottom left: $MAV_{Domain3}/UAV_{Domain5}$, bottom right: MAV_{Prop3}/UAV_{Prop3}), as well as preliminary semantic labeling. Correct and incorrect classifications are indicated by white and red font, respectively.