Table 2. Comparison of various Datasets for Small Object Video Detection

DataSet	Area Size						Classes	MedianArea	Avg Obj/Frame	Instances	Images/videos
	$0-12^2$ (es)		12^2-20^2 (rs)		20^2-32^2 (gs)						
ImageNet VID	968	0.05%	13k	0.65%	47k	2.36%	30	187	1	2M	1.2M/4417
NPS-Drone	20k	34.65%	35k	60.74%	2k	4.43%	1	14	1	59K	45K/50
UAVDT	29k	3.30%	204k	23.03%	265k	29.81%	6	30	22	889K	40k/50
VisDrone2019 VID	7k	0.53%	115k	7.51%	316k	20.60%	10	45	49	1.5M	30k/58
XS-VIDours	**49k**	**19.29%**	**94k**	**36.61%**	36k	13.97%	**8**	**18**	**22**	260K	12k/38