Deliberate Attack

Analyze attack trends

Table 1: We run a logistic model regressing success against perturb-target distance (relative to image width/height) and perturb box length (relative to image width/height) in the deliberate attack experiment. Longer perturb box length (relative to image width/height) or shorter perturb-target distance (relative to image width/height) cause success rates to significantly increase for all model and attack combinations, except for perturb box length (relative to image width/height) in untargeted attack on Cascade R-CNN. The interaction terms, even when significant, are negligibly close to 0. Table headers are explained in Appendix ??.

	Group	Regression							
	Attack	term	sig	estimate	std.error	statistic	p.value	conf.low	conf.high
YO	LOv3								
	Vanishing	distance	*	-6.969	1.182	-5.898	0.000	-9.308	-4.674
		length	*	7.744	0.453	17.110	0.000	6.873	8.648
		distance * length	*	-15.800	3.262	-4.844	0.000	-22.195	-9.400
	Mislabeling	distance	*	-7.516	1.198	-6.276	0.000	-9.891	-5.194
		length	*	5.962	0.354	16.853	0.000	5.277	6.664
		distance * length		-3.757	2.981	-1.260	0.208	-9.565	2.127
	Untargeted	distance	*	-9.214	1.496	-6.157	0.000	-12.205	-6.335
		length	*	2.276	0.257	8.860	0.000	1.775	2.782
		distance * length		3.324	2.896	1.148	0.251	-2.311	9.048
SSI)								
	Vanishing	distance	*	-11.391	1.488	-7.656	0.000	-14.362	-8.527
		length	*	4.449	0.281	15.841	0.000	3.903	5.004
		distance * length		-2.946	2.946	-1.000	0.317	-8.682	2.871
	Mislabeling	distance	*	-9.002	1.596	-5.641	0.000	-12.194	-5.934
		length	*	5.846	0.303	19.286	0.000	5.257	6.446
		distance * length	*	-13.088	3.183	-4.111	0.000	-19.300	-6.813
	Untargeted	distance	*	-10.046	1.760	-5.710	0.000	-13.580	-6.679
		length	*	3.873	0.280	13.823	0.000	3.328	4.426
		distance * length	*	-9.039	3.479	-2.598	0.009	-15.823	-2.178
Ret	tinaNet								
	Vanishing	distance	*	-22.822	2.997	-7.615	0.000	-28.866	-17.129
		length	*	2.657	0.333	7.978	0.000	2.008	3.313
		distance * length		-10.710	6.207	-1.725	0.084	-22.836	1.492
	Mislabeling	distance	*	-23.172	4.211	-5.503	0.000	-31.759	-15.284

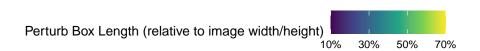
	length	*	1.352	0.411	3.293	0.001	0.552	2.162
	distance * length		3.035	8.178	0.371	0.711	-12.833	19.200
Untargeted	distance	*	-23.262	2.438	-9.540	0.000	-28.144	-18.587
	length	*	2.805	0.310	9.056	0.000	2.203	3.417
	distance * length	*	36.482	4.181	8.726	0.000	28.422	44.811
Faster R-CNN								
Vanishing	distance	*	-31.003	3.728	-8.317	0.000	-38.489	-23.891
	length	*	2.099	0.353	5.954	0.000	1.413	2.795
	distance * length		-0.776	7.406	-0.105	0.917	-15.251	13.769
Mislabeling	distance	*	-27.655	4.100	-6.745	0.000	-35.955	-19.909
	length	*	1.000	0.384	2.606	0.009	0.253	1.757
	distance * length		9.942	7.822	1.271	0.204	-5.291	25.353
Untargeted	distance	*	-29.452	2.752	-10.702	0.000	-34.974	-24.189
	length	*	1.949	0.294	6.625	0.000	1.375	2.529
	distance * length	*	34.982	4.521	7.738	0.000	26.255	43.974
Cascade R-CNI	N							
Vanishing	distance	*	-30.415	3.888	-7.823	0.000	-38.197	-22.974
	length	*	3.974	0.397	10.014	0.000	3.205	4.761
	distance * length	*	-25.436	8.343	-3.049	0.002	-41.832	-9.131
Mislabeling	distance	*	-34.967	4.698	-7.442	0.000	-44.420	-26.025
	length	*	1.927	0.388	4.971	0.000	1.173	2.693
	distance * length		2.023	9.140	0.221	0.825	-15.793	20.012
Untargeted	distance	*	-43.934	3.758	-11.692	0.000	-51.456	-36.729
	length	*	0.830	0.313	2.654	0.008	0.220	1.447
	distance * length	*	44.859	6.141	7.305	0.000	32.936	57.013

Table 2: We combined the data in the randomized and deliberate attack experiments to run a logistic model regressing success against object (versus non-object), with perturb-target distance (relative to image width/height) and perturb box size (relative to image width/height) as covariates. The "object" term codes object as 1 and non-object as 0. Perturbing an object (in the randomized attack) rather than a non-object (in the deliberate attack) significantly decreases success rates for all model and attack combinations, after controlling for perturb sizes and perturb-target distances. Table headers are explained in Appendix ??.

Group	Regression							
Attack	term	sig	estimate	std.error	statistic	p.value	conf.low	conf.high
YOLOv3								
Vanishing	object	*	-0.702	0.065	-10.865	0.000	-0.828	-0.575
	distance	*	-8.962	0.501	-17.896	0.000	-9.958	-7.995
	size	*	14.550	0.738	19.702	0.000	13.134	16.030
	distance * size	*	-38.136	4.331	-8.806	0.000	-46.722	-29.740

Mislabeling distance * -0.652 0.062 -10.521 0.000 -0.774 17.661 0.000 -0.377 17.661 0.000 -0.377 17.661 0.000 -0.377 17.661 0.000 -0.377 17.661 0.000 -0.471 17.661 0.000 -0.471 17.661 0.000 -0.471 17.661 0.000 -0.471 17.661 0.000 -0.471 17.661 0.000 -0.471 17.661 0.000 -0.471 17.661 0.000 0.041 17.671 17.671 0.065 0.078 0.065 0.									
Size	-0.531	-0.774	0.000	-10.521	0.062	-0.652	*	object	Mislabeling
Mislabelling Misl	-7.507	-9.377	0.000	-17.661	0.477	-8.427	*	distance	
Part	9.654	7.841	0.000	18.883	0.462	8.731	*	size	
Mislabeling Mislance * -1.15.8 Mislabeling Mislance * -1.15.8 Mislabeling Mislance * -1.15.8 Mislabeling Mislance * -1.5.78 Mislabeling Mislabeling Mislance * -1.5.78 Mislabeling M	-4.945	-17.177	0.000	-3.540	3.119	-11.042	*	distance * size	
Size * 2.107 0.265 7.957 0.000 1.589 Vanishing Object 0.079 0.063 1.242 0.214 0.045 Size * 1.5053 0.0713 -21.121 0.000 -16.475 Size * 5.106 0.307 16.639 0.000 4.510 Size * 6.1080 0.066 4.241 0.000 -0.409 Size * 6.5517 0.310 17.805 0.000 4.915 Size * 5.517 0.310 17.805 0.000 4.915 Size * 5.517 0.310 17.805 0.000 4.915 Size * 6.1290 0.070 3.163 0.002 0.357 Size * 3.430 0.275 12.493 0.000 2.895 Size * 8.2013 1.755 1.5965 0.000 -3.1533 Size * 8.2013 1.755 1.5965 0.000 -3.1533 Size * 2.736 0.339 8.081 0.000 2.078 Size * 2.736 0.339 8.081 0.000 2.078 Size * 2.736 0.339 8.081 0.000 2.078 Size * 2.736 0.339 8.081 0.000 0.718 Size * 2.736 0.339 8.081 0.000 0.078 Size * 2.736 0.329 0.581 0.561 8.297 Size * 2.736 0.329 0.581 0.561 8.297 Size * 2.0039 8.456 2.370 0.018 3.311 Untargeted Size * 1.027 0.423 2.426 0.015 0.200 Size * 3.363 0.280 1.2025 0.000 0.520 Size * 3.343 0.025 0.000 0.520 Size * 3.343 0.027 0.025 0.000 0.520 Size * 3.343 0.025 0.025 0.000 0.520 Size * 3.363 0.280 1.2025 0.000 0.520 Size * 3.363 0.280 1.2025 0.000 0.520 Size * 3.363 0.280 1.316 0.000 0.520 Size * 3.363 0.280 1.316	-0.633	-0.940	0.000	-10.026	0.078	-0.786	*	object	Untargeted
Note	-10.023	-13.118	0.000	-14.616	0.789	-11.538	*	distance	
Vanishing Object Size	2.627	1.589	0.000	7.957	0.265	2.107	*	size	
Vanishing object 0.079 0.063 1.242 0.214 -0.045	18.235	7.649	0.000	4.786	2.700	12.922	*	distance * size	
Vanishing object 0.079 0.063 1.242 0.214 -0.045									SD
Mislabeling Size * 5.106 0.307 16.639 0.000 4.510	0.203	-0.045	0.214	1.242	0.063	0.079		object	
Mislabeling object * -0.280 0.066 -4.241 0.000 -0.409 Mislabeling object * -0.280 0.066 -4.241 0.000 -0.409 distance * -14.939 0.800 -18.674 0.000 -16.538 size * 5.517 0.310 17.805 0.000 4.915 distance * size -2.191 2.933 -0.747 0.455 -7.946 Untargeted object * -0.220 0.070 -3.163 0.002 -0.357 distance * 15.788 0.918 -17.194 0.000 -17.628 size * 3.430 0.275 12.493 0.000 -2.895 distance * size 3.110 3.178 0.979 0.328 -3.137 RetinaNet Vanishing object * -0.552 0.087 -6.327 0.000 -0.724 distance * size * 27.36 0.339 8.081 0.000 -31.533 size * -0.469	-13.681	-16.475	0.000	-21.121	0.713	-15.053	*	distance	
Mislabeling distance object * -0.280 0.066 -4.241 0.000 -0.409 distance * -14.939 0.800 -18.674 0.000 -16.538 size * 5.517 0.310 17.805 0.000 4.915 distance * size -2.191 2.933 -0.747 0.455 -7.946 Untargeted object * -0.220 0.070 -3.163 0.002 -0.357 distance * 15.788 0.918 -17.194 0.000 -17.628 size * 3.430 0.275 12.493 0.000 2.895 distance * size 3.110 3.178 0.979 0.328 -3.137 RetinaNet Object * -0.552 0.087 -6.327 0.000 -0.724 distance * 28.013 1.755 -15.965 0.000 -31.533 size * 2.736 0.339 8.081 0.000 2.078 distance * size * 3.445 5.929 0.581 0.561 <td>5.714</td> <td>4.510</td> <td>0.000</td> <td>16.639</td> <td>0.307</td> <td>5.106</td> <td>*</td> <td>size</td> <td></td>	5.714	4.510	0.000	16.639	0.307	5.106	*	size	
distance	8.638	-2.073	0.230	1.201	2.731	3.280		distance * size	
Size	-0.150	-0.409	0.000	-4.241	0.066	-0.280	*	object	Mislabeling
Distance Size Sist Sis	-13.402	-16.538	0.000	-18.674	0.800	-14.939	*	distance	
Untargeted Object	6.130	4.915	0.000	17.805	0.310	5.517	*	size	
distance * -15.788 0.918 -17.194 0.000 -17.628 size * 3.430 0.275 12.493 0.000 2.895 distance * size 3.110 3.178 0.979 0.328 -3.137 RetinaNet Vanishing object * -0.552 0.087 -6.327 0.000 -0.724 distance * -28.013 1.755 -15.965 0.000 -31.533 size * 2.736 0.339 8.081 0.000 2.078 distance * size 3.445 5.929 0.581 0.561 -8.297 Mislabeling object * -0.469 0.125 -3.748 0.000 -0.718 distance * -30.658 2.722 -11.265 0.000 -36.164 size * 1.027 0.423 2.426 0.015 0.200 distance * size * 20.039 8.456 2.370 0.018 3.311 Untargeted object * -0.359 0.082 -4.382 0.000 -0.520 distance * -13.947 1.029 -13.554 0.000 -16.015 size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142 Size * 2.888 0.384 7.528 0.000 2.	3.555	-7.946	0.455	-0.747	2.933	-2.191		distance * size	
Size	-0.084	-0.357	0.002	-3.163	0.070	-0.220	*	object	Untargeted
Size 3.490 0.275 12.435 0.000 2.535	-14.029	-17.628	0.000	-17.194	0.918	-15.788	*	distance	
Vanishing	3.972	2.895	0.000	12.493	0.275	3.430	*	size	
Vanishing object * -0.552 0.087 -6.327 0.000 -0.724 distance * -28.013 1.755 -15.965 0.000 -31.533 size * 2.736 0.339 8.081 0.000 2.078 distance * size 3.445 5.929 0.581 0.561 -8.297 distance * size 3.445 5.929 0.581 0.561 -8.297 distance * -30.658 2.722 -11.265 0.000 -36.164 size * 1.027 0.423 2.426 0.015 0.200 distance * size * 20.039 8.456 2.370 0.018 3.311 0.000 distance * size * 20.039 8.456 2.370 0.018 3.311 0.000 distance * -13.947 1.029 -13.554 0.000 -0.520 distance * -13.947 1.029 -13.554 0.000 -16.015 size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	9.327	-3.137	0.328	0.979	3.178	3.110		distance * size	
Vanishing object * -0.552 0.087 -6.327 0.000 -0.724 distance * -28.013 1.755 -15.965 0.000 -31.533 size * 2.736 0.339 8.081 0.000 2.078 distance * size 3.445 5.929 0.581 0.561 -8.297 distance * size 3.445 5.929 0.581 0.561 -8.297 distance * -30.658 2.722 -11.265 0.000 -36.164 size * 1.027 0.423 2.426 0.015 0.200 distance * size * 20.039 8.456 2.370 0.018 3.311 0.000 distance * size * 20.039 8.456 2.370 0.018 3.311 0.000 distance * -13.947 1.029 -13.554 0.000 -0.520 distance * -13.947 1.029 -13.554 0.000 -16.015 size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142									RetinaNet
Size * 2.736 0.339 8.081 0.000 2.078	-0.382	-0.724	0.000	-6.327	0.087	-0.552	*	object	
Mislabeling Object * -0.469 0.125 -3.748 0.000 -0.718	-24.656	-31.533	0.000	-15.965	1.755	-28.013	*	distance	
Mislabeling object * -0.469 0.125 -3.748 0.000 -0.718 distance * -30.658 2.722 -11.265 0.000 -36.164 size * 1.027 0.423 2.426 0.015 0.200 distance * size * 20.039 8.456 2.370 0.018 3.311 Untargeted object * -0.359 0.082 -4.382 0.000 -0.520 distance * -13.947 1.029 -13.554 0.000 -16.015 size * 33.63 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888	3.406	2.078	0.000	8.081	0.339	2.736	*	size	
distance * -30.658 2.722 -11.265 0.000 -36.164 size * 1.027 0.423 2.426 0.015 0.200 distance * size * 20.039 8.456 2.370 0.018 3.311 Untargeted object * -0.359 0.082 -4.382 0.000 -0.520 distance * -13.947 1.029 -13.554 0.000 -16.015 size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	14.956	-8.297	0.561	0.581	5.929	3.445		distance * size	
Size	-0.227	-0.718	0.000	-3.748	0.125	-0.469	*	object	Mislabeling
Untargeted distance * size * 20.039 8.456 2.370 0.018 3.311	-25.501	-36.164	0.000	-11.265	2.722	-30.658	*	distance	
Untargeted object * -0.359 0.082 -4.382 0.000 -0.520 distance * -13.947 1.029 -13.554 0.000 -16.015 size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	1.859	0.200	0.015	2.426	0.423	1.027	*	size	
distance * -13.947 1.029 -13.554 0.000 -16.015 size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	36.476	3.311	0.018	2.370	8.456	20.039	*	distance * size	
Size * 3.363 0.280 12.025 0.000 2.816 distance * size * 31.443 3.041 10.341 0.000 25.570 Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	-0.199	-0.520	0.000	-4.382	0.082	-0.359	*	object	Untargeted
Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	-11.982	-16.015	0.000	-13.554	1.029	-13.947	*	distance	
Faster R-CNN Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	3.913	2.816	0.000	12.025	0.280	3.363	*	size	
Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142	37.493	25.570	0.000	10.341	3.041	31.443	*	distance * size	
Vanishing object * -1.040 0.113 -9.179 0.000 -1.266 distance * -27.692 2.080 -13.316 0.000 -31.877 size * 2.888 0.384 7.528 0.000 2.142									aster R-CNN
size * 2.888 0.384 7.528 0.000 2.142	-0.821	-1.266	0.000	-9.179	0.113	-1.040	*	object	
SIZC 2.000 0.404 1.020 0.000 2.142	-23.729	-31.877	0.000	-13.316	2.080	-27.692	*	distance	
	3.646	2.142	0.000	7.528	0.384	2.888	*	size	
distance * size -13.106 7.590 -1.727 0.084 -28.136	1.621	-28.136	0.084	-1.727	7.590	-13.106		distance * size	
Mislabeling object * -1.009 0.141 -7.175 0.000 -1.292	-0.740	-1.292	0.000	-7.175	0.141	-1.009	*	object	Mislabeling
distance * -23.256 2.228 -10.440 0.000 -27.775	-19.048	-27.775	0.000	-10.440	2.228	-23.256	*	distance	

	size	*	1.152	0.411	2.805	0.005	0.348	1.958
	distance * size		3.841	7.680	0.500	0.617	-11.396	18.717
Untarge	ted object	*	-0.530	0.090	-5.900	0.000	-0.708	-0.356
	distance	*	-18.563	1.238	-14.997	0.000	-21.049	-16.197
	size	*	2.349	0.285	8.254	0.000	1.793	2.909
	distance * size	*	32.699	3.330	9.821	0.000	26.239	39.295
Cascade R-	CNN							
Vanishir	g object	*	-1.019	0.100	-10.149	0.000	-1.218	-0.824
	distance	*	-33.425	2.225	-15.020	0.000	-37.885	-29.166
	size	*	5.056	0.422	11.968	0.000	4.241	5.898
	distance * size	*	-40.938	8.862	-4.619	0.000	-58.577	-23.819
Mislabel	ing object	*	-0.898	0.123	-7.302	0.000	-1.143	-0.660
	distance	*	-31.929	2.726	-11.713	0.000	-37.433	-26.754
	size	*	2.875	0.392	7.341	0.000	2.111	3.646
	distance * size		-10.216	9.136	-1.118	0.263	-28.219	7.593
Untarge	ted object	*	-0.642	0.097	-6.605	0.000	-0.835	-0.454
	distance	*	-29.381	1.881	-15.623	0.000	-33.163	-25.791
	size	*	1.493	0.306	4.884	0.000	0.895	2.094
	distance * size	*	32.873	5.009	6.563	0.000	23.086	42.736



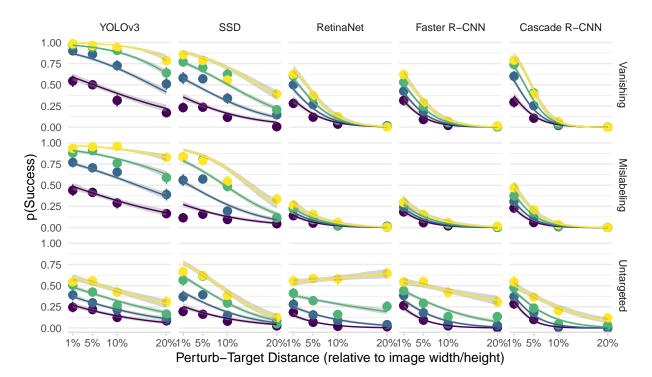


Figure 1: A deliberate attack obfuscates intent with increased success for all models and attacks: We implement intent obfuscating attack by perturbing an arbitrary non-overlapping square region to disrupt a randomly selected target object at various lengths and distances. The binned summaries and regression trendlines graph success proportion against perturb-target distance (relative to image width/height) and perturb box length (relative to image width/height) in the deliberate attack experiment. Errors are 95% confidence intervals. and every point aggregates success over 200 images. The deliberate attack multiplies success as compared to the randomized attack (Figure ??), especially at close perturb-target distance (relative to image width/height) and large perturb box length (relative to image width/height). Full details are given in Section ??.