Table S1. Name, number of annotations per class, number of images in which each class occurs, precision, recall, mean average precision at 50% intersect-over-union (IoU) threshold (mAP50) and at 50% to 95% IoU (in 5%-steps, mAP50-95) values per class. Empty fields indicate classes which did not appear in the test set due to the low number of images available in the training set) and the metaclass.

Class	Annotations	Total images	Test set instances	Precision	Recall	mAP50	mAP50-95	Metaclass
Aluminium foil	80	52	14	0.53	0.286	0.293	0.276	Metal litter
Backpack	28	15	2	0.567	1	0.995	0.895	Clothing & Fabric
Battery	3	3						N img < 10
Broken glass	110	14	1	1	0	0.00229	0.00206	Glass
Button	1	1	1	1	0	0	0	N img < 10
Cardboard	326	247	64	0.449	0.382	0.424	0.369	Paper & Cardboard
Chair	19	12	8	1	0.226	0.246	0.137	Other
Cigarette	1	1						N img < 10
Cigarette butt	795	252	205	0.612	0.278	0.347	0.238	Other
Cigarette package	8	4	5	1	0	0.0395	0.0356	N img < 10
Cloth	89	59	18	0.316	0.111	0.113	0.0611	Clothing & Fabric
Cork stopper	1	1	1	1	0	0	0	N img < 10
Diaper	1	1						N img < 10
Dog	1	1						N img < 10
Drink carton	59	56	10	0.208	0.3	0.135	0.109	Paper & Cardboard
Egg carton	6	5	1	1	0	0.00418	0.00418	N img < 10
Facemask	9	8						N img < 10
Food waste	2	2						N img < 10
Garbage bag	37	25	12	0.507	0.259	0.291	0.189	Plastic
Glass bottle	231	171	59	0.611	0.266	0.404	0.317	Glass
Glass cup	1	1						N img < 10
Glass jar	10	8	1	1	0	0.0151	0.00754	N img < 10
Hair pin	1	1						N img < 10
Lightbulb	8	8	3	1	0	0.168	0.151	N img < 10
Lighter	24	20	2	0.759	1	0.995	0.646	Plastic

Match	1	1						N img < 10
Metal	35	24	7	0.305	0.143	0.147	0.131	Metal litter
Metal bottlecap	79	55	18	0.445	0.167	0.135	0.115	Metal litter
Metal can	597	408	108	0.695	0.633	0.627	0.516	Metal litter
Grating	91	69	27	0.712	0.296	0.34	0.151	Other
Metal lid	7	4	2	1	0	0	0	N img < 10
Metal pole	12	7	2	0.835	1	0.995	0.796	N img < 10
Metal screw cap	37	30	8	1	0	0	0	Metal litter
Other plastic	1298	660	258	0.359	0.244	0.238	0.168	Plastic
Paper	478	312	93	0.297	0.14	0.128	0.0978	Paper & Cardboard
Paper bag	6	6	3	1	0	0.0106	0.00531	N img < 10
Paper cup	77	73	14	0.316	0.286	0.216	0.194	Paper & Cardboard
Person	90	63	17	0.766	0.577	0.667	0.499	Other
Pipe	84	57	21	0.596	0.19	0.187	0.151	Other
Plant pot	3	3	1	1	0	0	0	N img < 10
Plastic bag	90	82	22	0.117	0.0909	0.116	0.0653	Plastic
Plastic bottle	1424	906	301	0.751	0.767	0.797	0.674	Plastic
Plastic bottlecap	941	640	207	0.709	0.662	0.697	0.483	Plastic
Clothes pin	10	10	2	1	0	0	0	Plastic
Plastic container	235	190	50	0.567	0.48	0.511	0.447	Plastic
Plastic cup	135	118	34	0.649	0.219	0.367	0.288	Plastic
Plastic floater	3	2						N img < 10
Plastic lid	68	64	15	0.199	0.2	0.124	0.106	Plastic
Plastic utensils	58	50	15	0.43	0.267	0.259	0.235	Plastic
Plastic wrapper & foil	1230	717	249	0.399	0.357	0.289	0.226	Plastic
Pop tab	170	128	43	0.777	0.488	0.499	0.366	Metal litter
QTip	1	1	1	1	0	0	0	N img < 10
Rope & string	223	167	45	0.737	0.436	0.477	0.337	Other
Scrap metal	46	18	8	1	0	0	0	Metal litter
Shoe	156	105	30	0.626	0.467	0.528	0.43	Clothing & Fabric
Soccer ball	1	1						N img < 10

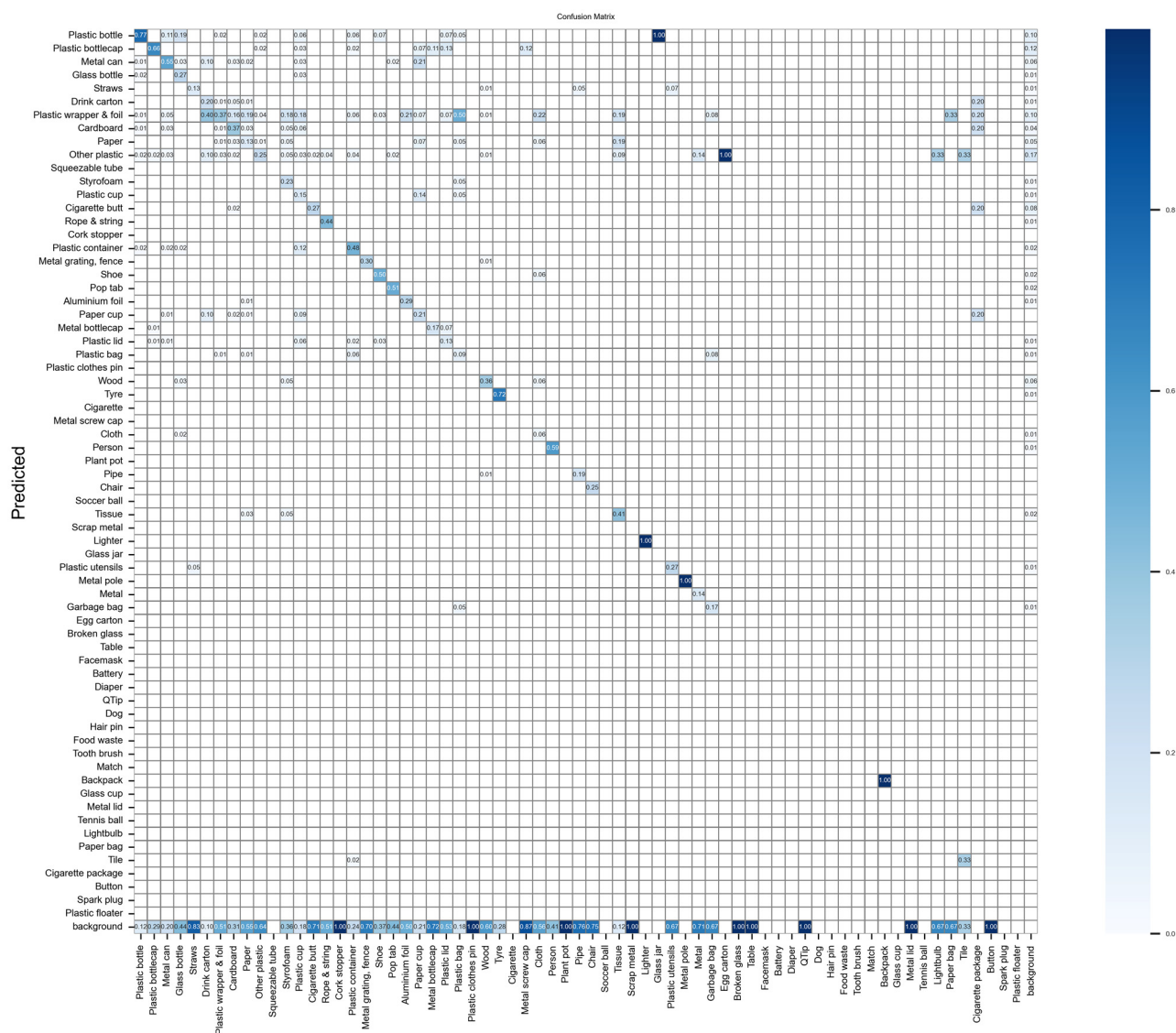|							
Spark plug	2	2						N img < 10
Squeezetube	4	4						N img < 10
Straws	213	159	63	0.463	0.127	0.141	0.101	Plastic
Styrofoam	98	79	22	0.629	0.227	0.343	0.294	Plastic
Table	3	1	3	1	0	0	0	N img < 10
Tennis ball	1	1						N img < 10
Tile	19	5	3	0.235	0.333	0.168	0.168	N img < 10
Tissue	103	65	32	0.712	0.465	0.537	0.48	Paper & Cardboard
Tooth brush	1	1						N img < 10
Tyre	115	55	25	0.785	0.72	0.762	0.586	Other
Wood	515	238	141	0.669	0.369	0.448	0.324	Other

Figure S1. Confusion matrix from the validation against the test set. The x-axis represents the actual categories of objects, while the y-axis represents the categories of the detections as predicted by the algorithm. Values in the matrix represent the proportion of predictions for detections of each category. Perfect predictions with no misclassifications will lead to a single line of 1.0 values running across the matrix from the top left to the bottom right. The fact that many misclassifications have been predicted as "background" and are therefore accumulating at the bottom of the matrix indicates false negative (FN) predictions (i.e., missed objects that should have been detected).