“ScatSpotter” — A Dog Poop Detection Dataset

Jonathan Crall
Kitware
jon.crall@kitware.com

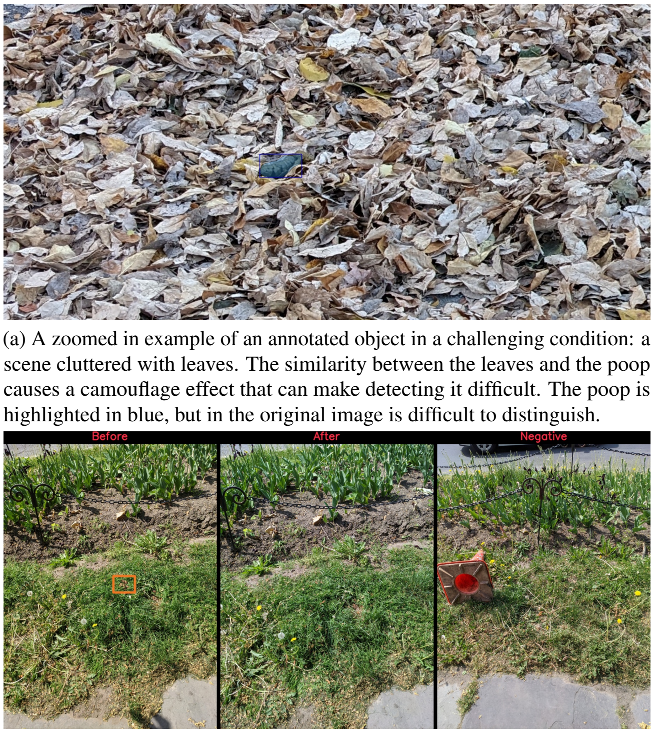
Abstract

Small, amorphous waste objects such as biological droppings and microtrash can be difficult to see, especially in cluttered scenes, yet they matter for environmental cleanliness, public health, and autonomous cleanup. We introduce “ScatSpotter”: a new dataset of phone images annotated with polygons around dog feces, collected to train and study object detection and segmentation systems for small potentially camouflaged outdoor waste. We gathered data in mostly urban environments, using a “before/after/negative” (BAN) protocol: for a given location, we capture an image with the object present, an image from the same viewpoint after removal, and a nearby negative scene that often contains visually similar confusers.

Image collection began in late 2020. This paper focuses on two dataset checkpoints from 2025 and 2024. The dataset contains over 9000 full-resolution images and 6000 polygon annotations. Of the author-captured images we held out 691 for validation and used the rest to train. Via community participation we obtained a 121-image test set that, while small, is independent from author-collected images and provides some generalization confidence across photographers, devices, and locations. Due to its limited size, we report both validation and test results.

We explore the difficulty of the dataset using off-the-shelf VIT, MaskRCNN, YOLO-v9, and DINO-v2 models. Zero-shot DINO performs poorly, indicating limited foundational-model coverage of this category. Tuned DINO is the best model with a box-level average precision of 0.69 on a 691-image validation set and 0.70 on the test set. These results establish strong baselines and quantify the remaining difficulty of detecting small, camouflaged waste objects.

To support open access to models and data (CC-BY 4.0 license), we compare centralized and decentralized distribution mechanisms and discuss trade-offs for sharing scientific data. Code for experiments and project details are hosted on GitHub.



(a) A zoomed in example of an annotated object in a challenging condition: a scene cluttered with leaves. The similarity between the leaves and the poop causes a camouflage effect that can make detecting it difficult. The poop is highlighted in blue, but in the original image is difficult to distinguish.

(b) The “before/after/negative” protocol. The orange box highlights the location of the poop in the “before” image. In the “after” image, it is the same scene and viewpoint but the poop has been removed. The “negative” image is a nearby similar scene, potentially with a distractor. Note that the object is small relative to the image size.

Figure 1. (a) A challenging annotation case due to clutter and camouflage. (b) An image triplet from the BAN protocol.

1. Introduction

Autonomous and AI-assisted waste monitoring is increasingly achievable with modern object detection and segmentation methods [35, 49, 51, 57] combined with large annotated datasets. Substantial progress has been made in detecting large or conspicuous objects, especially those well represented in foundational training corpora. However, small and irregular waste objects — such as biological droppings or microtrash — are underrepresented in existing datasets and

Name	#Cats	#Images	#Annots	Image W × H	Annot Area$^{0.5}$	Disk Size	Annot Type
ImageNet[48]	1,000	594,546	695,776	500 × 374	239	166GB	box
MSCOCO[33]	80	123,287	896,782	428 × 640	57	50GB	polygon
CityScapes[12]	40	5,000	287,465	2,048 × 1,024	50	78GB	polygon
ZeroWaste [3]	4	4,503	26,766	1,920 × 1,080	200	10GB	polygon
TrashCanV1[25]	22	7,212	12,128	480 × 270	54	0.61GB	polygon
UAVVaste[29]	1	772	3,718	3,840 × 2,160	55	2.9GB	polygon
SpotGarbage[41]	1	2,512	337	754 × 754	355	1.5GB	category
TACO[46]	60	1,500	4,784	2,448 × 3,264	119	17GB	polygon
MSHIT[39]	2	769	2,348	960 × 540	99	4GB	box
Ours	1	9,296	6,594	4,032 × 3,024	87	60GB	polygon

Table 1. Related datasets. Columns list dataset name, number of categories, images, and annotations. Image W × H gives median image dimensions; Ann Area$^{0.5}$ is the median square root of annotation area (pixels); Size is disk requirements in GB; Annot Type is the labeling method. Figure 2 shows the distribution of annotation shapes, sizes, and locations.

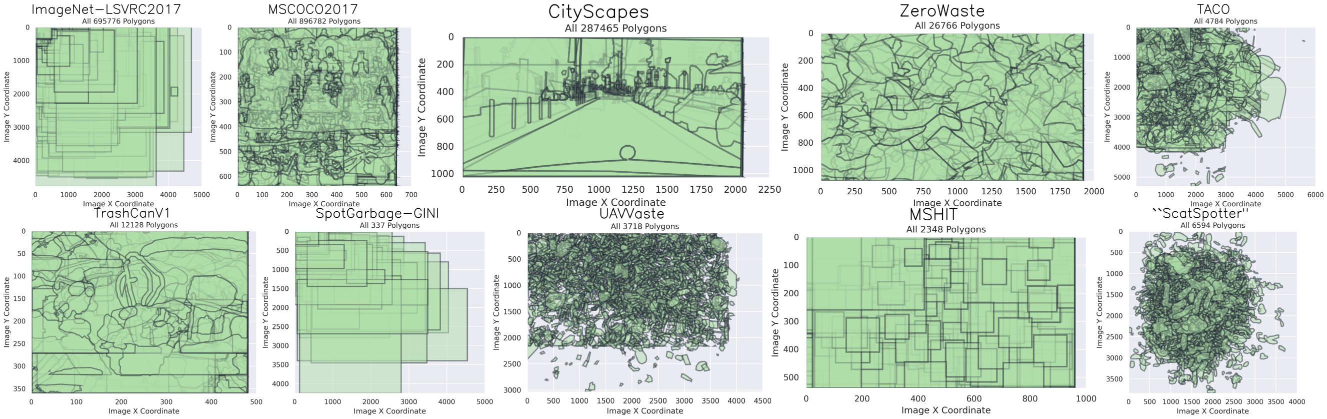


Figure 2. A comparison of all of the annotations for different datasets including ours. All polygon annotations drawn in a single plot with 0.8 opacity to demonstrate the distribution in annotation location, shape, and size with respect to image coordinates.

remain difficult to detect. These objects — such as the example illustrated in Figure 1a — are often low contrast, variable in appearance, and confusable with natural clutter, making them challenging for both humans and vision systems.

To address this gap we introduce a new dataset which, in formal settings, we call "ScatSpotter". Our dataset contains high-resolution images of dog poop in most of which are from urban, outdoor environments in a single city. The dataset exhibits variation in appearance, season, lighting, and background clutter despite biases toward the author's dogs and geographic region. Poops are annotated with polygons, making the dataset suitable for both detection and segmentation models. To assist with annotation and provide counterfactual examples we collect images using a "before/after/negative" (BAN) protocol as shown in Figure 1b.

One motivating use case, which originally inspired this work, is a phone application that assists dog owners in locating their dog's poop in a leafy park for easier cleanup. Other applications include automated waste disposal to keep sidewalks, parks, and backyards clean, tools for monitoring wildlife populations via droppings, and warning systems in smart glasses to prevent people from stepping in poop. Although we focus on a single class, dog poop provides an accessible prototypical example for the broader problem of detecting small, amorphous, and often camouflaged waste in outdoor environments — a challenge in common with tasks such as litter detection, microtrash identification, and wildlife monitoring. The visual difficulty of the domain, rather than the specific species, is the focus of this work.

Beyond the dataset itself, we are also interested in how large datasets can be shared efficiently and robustly. Centralized methods such as Girder [42] and HuggingFace Datasets [32] are a typical choice, offering high speeds, but they can be costly for individuals, often requiring institutional support or paid hosting services. They are also prone to outages and lack built-in data validation. In contrast, decentralized methods allow volunteers to host data and offers built-in validation of data integrity. This motivates us to compare

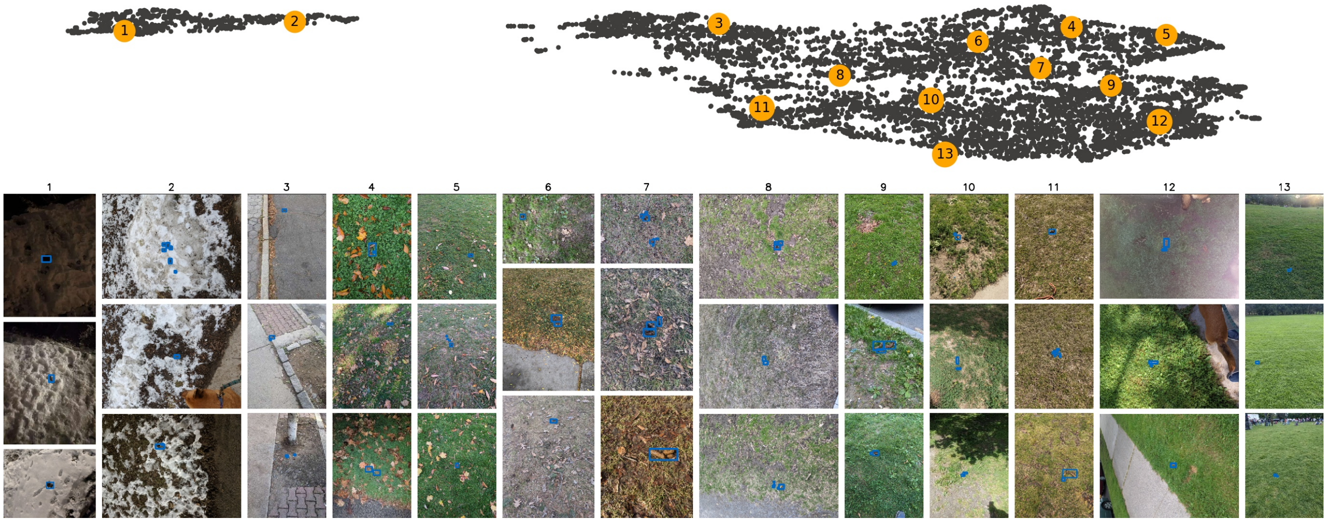


Figure 3. Example images from 2D UMAP clusters [38]. Each point in the top image represents a 2D-projected embedding, with numbered orange dots indicating nearby images in the bottom columns. Blue annotation boxes are shown. A clear separation emerges between snowy (columns 1-2) and non-snowy images (columns 3-13).

and contrast BitTorrent [8], and IPFS [4] as mechanisms for distributing datasets.

Our contributions are: 1) A challenging new **open dataset** of images with polygon annotations for small, camouflaged waste objects (using dog poop as a case study). 2) A set of trained **baseline models**. 3) A **comparison of dataset distribution** methods. Together, these contributions are intended to support future work on small-object waste detection, smart waste monitoring, and environmentally focused computer vision applications. For F.A.Q., see Appendix A.

2. Related Work

To the best of our knowledge, our dataset is currently the largest publicly available collection of annotated dog poop images, but it is not the first. A dataset of 100 dog poop images was collected and used to train a FasterRCNN model [43] but this dataset and model are not publicly available. The company iRobot has a dataset of annotated indoor poop images used to train Roomba j7+ to avoid collisions [21], but as far as we are aware, this is not available. In terms of available poop detection datasets we are only aware of MSHIT [39] which is much smaller, only contains box annotations, and the objects of interest are plastic toy poops.

Compared to benchmark object localization and segmentation datasets [12, 33, 48] ours is much smaller and focused only on a single category. However, when compared to litter and trash datasets [3, 25, 29, 41, 46] ours is among the largest in terms of number of images / annotations, image size, and total dataset size. ZeroWaste [3] uses a “before/after” protocol similar to our BAN protocol. We provide an overview of these related datasets in Table 1. Among all of these, ours stands out for having the highest resolution images and the smallest objects relative to that resolution. For a review of additional waste related datasets, refer to [40].

Section 5 discusses the logistics and tradeoffs between dataset distribution mechanisms with a focus on comparing centralized and decentralized methods. IPFS [4] and BitTorrent [8] are the decentralized mechanisms we evaluate, but others exist such as Secure Scuttlebut [53] and Hypercore [17], which we did not test.

3. Dataset

Our first contribution is the creation of a new open dataset which consists of images of dog poop in mostly urban, mostly outdoor environments, from mostly a single city. The data is annotated to support object detection and segmentation tasks. The majority of the images feature fresh poop from three specific medium sized dogs, but there are a significant number of images with poops of unknown age and from unknown dogs.

Despite these biases, the dataset has significant image variations. To provide a gist, we computed UMAP [38] embeddings using ResNet50 [22] descriptors, and display images corresponding with clusters in Figure 3.

More details about the dataset are available in a standardized datasheet [18] that covers the motivation, composition, collection, preprocessing, uses, distribution, and maintenance. This is distributed with the data itself.

3.1. Dataset Collection

A single researcher on dog walks photographed fresh dog poop, mostly their own dogs, but often others. Distance was varied for diversity. Most images were taken following the “before/after/negative” (BAN) protocol. A BAN triple comprises a “before” shot of the poop, an “after” shot post removal, and a “negative” shot of a nearby lookalike (e.g., pine cones, leaves). We only use them for negative sampling, but they could enable contrastive triplet losses [50].

The majority of images follow the BAN protocol, but there are exceptions. The first six months of data collection only involved the “before/after” part of the protocol. We began collecting the third negative image after a colleague suggested it. In some cases, the researcher failed or was unable to take the second or third image. These exceptions are often programmatically identifiable.

We also received 121 contributor images, mostly outside the BAN protocol, which we use as a test set. Due to the small size, our main results also include validation scores.

3.2. Dataset Annotation

Images were annotated using labelme [27]. Most annotations were initialized using SAM and a point prompt. All AI polygons were manually reviewed. In most cases only small manual adjustments were needed, but there were a significant number of cases where SAM did not work well and fully manual annotations were needed. Regions with shadows seemed to cause SAM the most trouble, but there were other failure cases. Unfortunately, there is no metadata to indicate which polygons were manually created or done using AI. However, the number of vertices may be a reasonable proxy to estimate this, as polygons generated by SAM tend to have higher fidelity boundaries. The boundaries of the annotated polygons are illustrated in Figure 2.

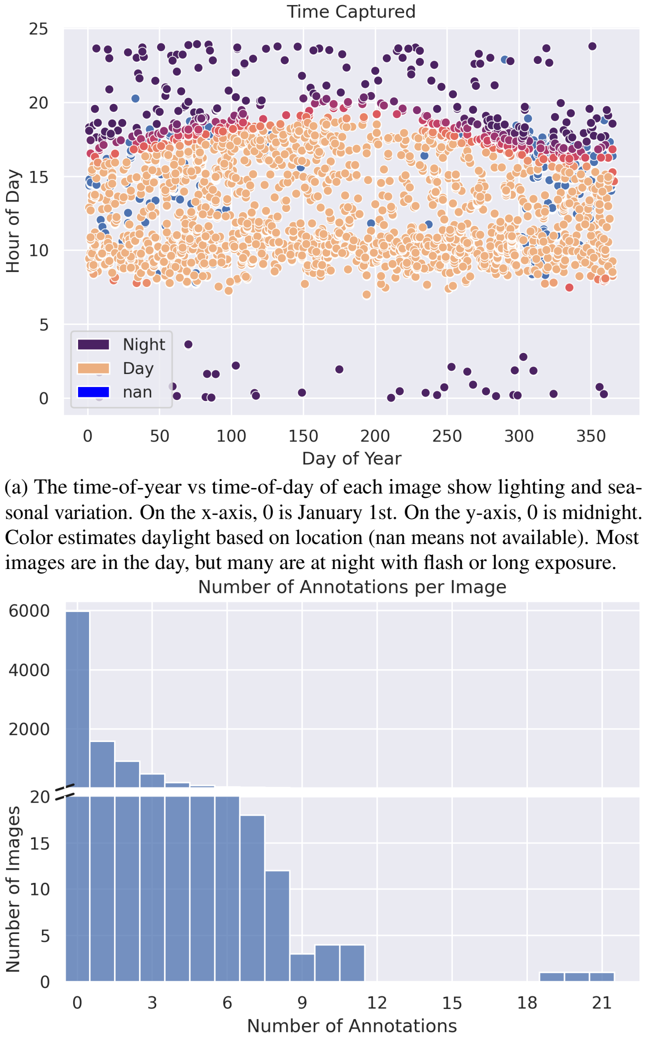
Data collected after 2024-07-03 was annotated with the help of models trained on prior data. Again, all predictions were manually verified or corrected. In these later cases, false positive annotations were labeled (e.g. stick, leaf), but because these categories are not labeled exhaustively, we exclude them from all analysis in this paper.

3.3. Dataset Properties and Statistics

The data was captured at a regular rate over 4.3 years, primarily in parks and sidewalks within a small city. Weather conditions varied across snowy, sunny, rainy, and foggy. A visual representation of the distribution of seasons, time-of-day, daylight, and capture rate is provided in Figure 4a.

The dataset is available in full resolution. Almost all images were taken using the same phone-camera, with a consistent size of 4,032 × 3,024 (up to EXIF rotation). The images are stored as 8-bit JPEGs with RGB channels, and most include overviews (i.e., image pyramids), allowing for fast loading of downscaled versions.

Due to the BAN protocol, about one-third of the images contain annotations, the rest were taken after the object(s) were removed. Consequently, most images have no annotations. When present, annotations are typically singular, but multiple are common due to: 1) fragmented droppings, 2) dogs pooping together, 3) repeated poops in the same area over time (sometimes hard to distinguish from dirt). The number of annotations per image is given in Figure 4b.



(a) The time-of-year vs time-of-day of each image show lighting and seasonal variation. On the x-axis, 0 is January 1st. On the y-axis, 0 is midnight. Color estimates daylight based on location (nan means not available). Most images are in the day, but many are at night with flash or long exposure.

(b) The histogram show variation in annotations per image. Only 35% (3,314) of images contain annotations; 65% (5,982) are known negatives. About half of the negatives were taken immediately after pickup; the rest are from nearby locations with potential lookalikes. Note the split y-axis.

Figure 4. Dataset distributions. (a) Time and daylight scatterplot. (b) Annotation count histogram.

3.4. Dataset Splits

Our dataset is split into training, validation, and test sets based on the year and day of image capture and photographer. Only data captured by the authors is used for training and validation. Of these, images from 2021-2023, 2025 and beyond are assigned to the training set. Images from 2020 are used for validation. For data from 2024, we consider the ordinal date n of each image and include it in the validation set if $n \equiv 0 \pmod 3$; otherwise, it is assigned to training.

For testing data, we use contributor images to not bias our results based on the way the authors took images. These splits are provided in the COCO JSON format [33] as well as a WebDataset [54] on HuggingFace.

4. Baseline Models

As our second contribution, we trained and evaluated models to establish a baseline for future comparisons. Specifically we train 7 model variants. We trained a semantic segmentation vision transformer variant (VIT-sseg-s) [13, 20], which was only trained from scratch. We trained two MaskRCNN [23] models (specifically the `R_50_FPN_3x` configuration), one starting from pretrained ImageNet weights (MaskRCNN-p), and one starting from scratch (MaskRCNN-s). Similarly we trained YOLO-v9 [56] both from scratch and using pretrained ImageNet weights. Lastly, we evaluated the foundational Grounding DINO [35] model. In the zero-shot setting we used `IDEA-Research/grounding-dino-tiny` using the prompt: "animalfeces". Finally, we fine-tune evaluate the same DINO model using [58].

The number of parameters for MaskRCNN, VIT, GroundingDINO, and YOLO are 44M, 26M, 172M, and 51M. Hyperparameters are given in Appendix D.

For these baseline models, the training data was limited to an older subset taken before 2024-07-03. Our training dataset consists of 5,747 images and is identified by a suffix of `1e73d54f`, which is the prefix of its content hash. The validation set contains 691 images and has a suffix of `99b22ad0`. The test set, consists of the 121 images, has a suffix of `6cb3b6ff`, and includes contributor images up to 2025-04-20. The evaluated models were selected based on their Box-AP validation scores.

The primary detection “Box” evaluation computes standard COCO object detection metrics [33]. MaskRCNN, GroundingDINO, and YOLO-v9 natively output scored bounding boxes, but for the VIT-sseg model, we convert heatmaps into boxes by thresholding the probability maps and taking the extend of the resulting polygons as bounding boxes. The score is taken as the average heatmap response under the polygon. Bounding box evaluation has the advantage that small and large annotations contribute equally to the score, but it can also be misleading for datasets where the notion of an object instance can be ambiguous.

To complement the box evaluation, we performed a pixelwise evaluation, which is more sensitive to the details of the segmented masks, but also can be biased towards larger annotations with more pixels. The corresponding truth and predicted pixels were accumulated into a confusion matrix, allowing us to compute standard metrics [45] such as precision, recall, false positive rate, etc. For the VIT-sseg model, computing this score is straightforward, but for MaskRCNN we accumulate per-box heatmaps into a larger full image heatmap, which can then be scored. Because YOLO-v9 and GroundingDINO do not produce masks, they were excluded from pixelwise evaluation.

Quantitative results for each of these models on box and pixel metrics are shown in Table 2. Because the independent test set is only 121 images, we also present results on the larger validation dataset. Corresponding validation results are illustrated in Figure 5 and test results in Figure 6.

All models were trained on a single machine with an Intel Core i9-11900K CPU and an NVIDIA GeForce RTX 3090 GPU. Our environmental impact [1] was manageable.

5. Dataset Transfer Experiment

Our third contribution is an experiment that studies transfer rates of decentralized and centralized data distribution methods. For centralized distribution, we use a self-hosted instance of Girder [42] and the HuggingFace datasets [32] platform. For decentralized clients, we use Transmission [31] (BitTorrent) and Kubo [26] (IPFS). As a baseline, we also measure direct transfers using Rsync [55].

For data transfer experiments, we use the 2024-07-03 version of the dataset. This is content-addressed with the IPFS CID (content identifier): `bafybeiedwp2zvmdyb2c2axrcl455xfbv2mgdbhgkc3dile4dftiimwth2y`. The torrent magnet URL is: `magnet:?xt=urn:btih:ee8d2c87a39ea9bfe48bef7eb4ca12eb68852c49`, and is tracked on Academic Torrents [10]. More details in Appendix C.1.

The HuggingFace results stand out, as they are faster than rsync. We believe this is due to an optimized client and content delivery networks, utilizing CAKE [24] to minimize buffer bloat [19]. However, this speed relies on costly centralized infrastructure. The expected speed from a more modest centralized service is $\sim$20× slower.

There is an additional $\sim 4\times$ slowdown between compressed and uncompressed rsync baselines, which needs to be considered when comparing decentralized results. The minimum time column shows that decentralized methods can be competitive with rsync, but on average decentralized mechanisms are significantly slower and can be stifled by long peer-discovery times.

6. Conclusion

We have introduced the largest open dataset of high resolution images with polygon segmentations of dog poop. While only focused on a single class, it is prototypical of challenges that arise in small-waste detection relevant to waste

[1] Over all of our experimentation, prediction and evaluation took 14 days, consuming 108 kWh and emitting 23 CO_2 kg (CodeCarbon [30]). Training was estimated at 164 days and 1359 kWh, yielding 285 CO_2 kg, assuming a 345W GPU draw and a 0.21 $\frac{\text{kgCO}_2}{\text{kWh}}$ emission factor. At \$0.16/kWh and \$25/tonne CO_2, total cost was \$242.37. More details in Appendix E.

(a) Validation (n=691)

Model	AP Box	AUC Box	F1 Box	TPR Box	AP Pixel	AUC Pixel	F1 Pixel	TPR Pixel
MaskRCNN-p	0.61	**0.72**	0.62	0.57	0.74	0.91	**0.74**	0.68
MaskRCNN-s	0.26	0.58	0.35	0.31	0.43	0.89	0.48	0.50
VIT-sseg-s	0.48	0.53	0.60	0.51	**0.76**	**0.97**	0.74	**0.69**
GroundingDINO-t	**0.69**	0.63	**0.74**	**0.68**	–	–	–	–
GroundingDINO-z	0.08	0.21	0.20	0.25	–	–	–	–
YOLO-v9-p	0.41	0.59	0.50	0.42	–	–	–	–
YOLO-v9-s	0.33	0.41	0.44	0.37	–	–	–	–

(b) Test (n=121)

Model	AP Box	AUC Box	F1 Box	TPR Box	AP Pixel	AUC Pixel	F1 Pixel	TPR Pixel
MaskRCNN-p	0.61	**0.70**	0.65	0.60	**0.81**	**0.85**	**0.78**	**0.73**
MaskRCNN-s	0.25	0.47	0.34	0.30	0.39	0.80	0.41	0.44
VIT-sseg-s	0.39	0.40	0.52	0.41	0.41	0.82	0.48	0.37
GroundingDINO-t	**0.70**	0.67	**0.76**	**0.68**	–	–	–	–
GroundingDINO-z	0.23	0.30	0.39	0.38	–	–	–	–
YOLO-v9-p	0.44	0.55	0.51	0.50	–	–	–	–
YOLO-v9-s	0.36	0.36	0.48	0.37	–	–	–	–

Table 2. Baseline model performance on validation and test sets. Suffixes indicate training conditions: -p (pretrained), -s (scratch), -t (tuned), -z (zero-shot). Metrics include box- and pixel-level AP (area under precision-recall), AUC (area under ROC), F1, and TPR (recall), computed using scikit-learn [44]. Pretrained models outperform with the tuned foundational Grounding DINO model performing best. Note: VIT-sseg was tuned more extensively and operated at full resolution, while MaskRCNN, DINO, and YOLO used off-the-shelf settings (that resized images) and may improve with additional tuning.

monitoring, pollution tracking, and environmental surveillance. The dataset includes amorphous objects, occlusion, multi-season variation, difficult distractors, daytime / nighttime variation. We have described the dataset collection and annotation process and reported statistics on the dataset.

We provided a recommended train/validation/test split of the dataset, and trained baseline segmentation models that perform well, but could likely be improved. In addition to providing quantitative and qualitative results of the models, we also estimate the resources required to perform these training, prediction, and evaluation experiments.

Method	Zipped	μ	σ	Min	Max
BitTorrent	No	8.36h	5.16h	2.21h	14.39h
IPFS	No	10.68h	9.54h	1.80h	24.62h
Rsync	No	4.84h	1.39h	3.10h	6.10h
Girder	Yes	2.85h	2.31h	1.05h	6.24h
HuggingFace	Yes	**0.14h**	0.03h	0.11h	0.18h
Rsync	Yes	1.10h	0.03h	1.07h	1.13h

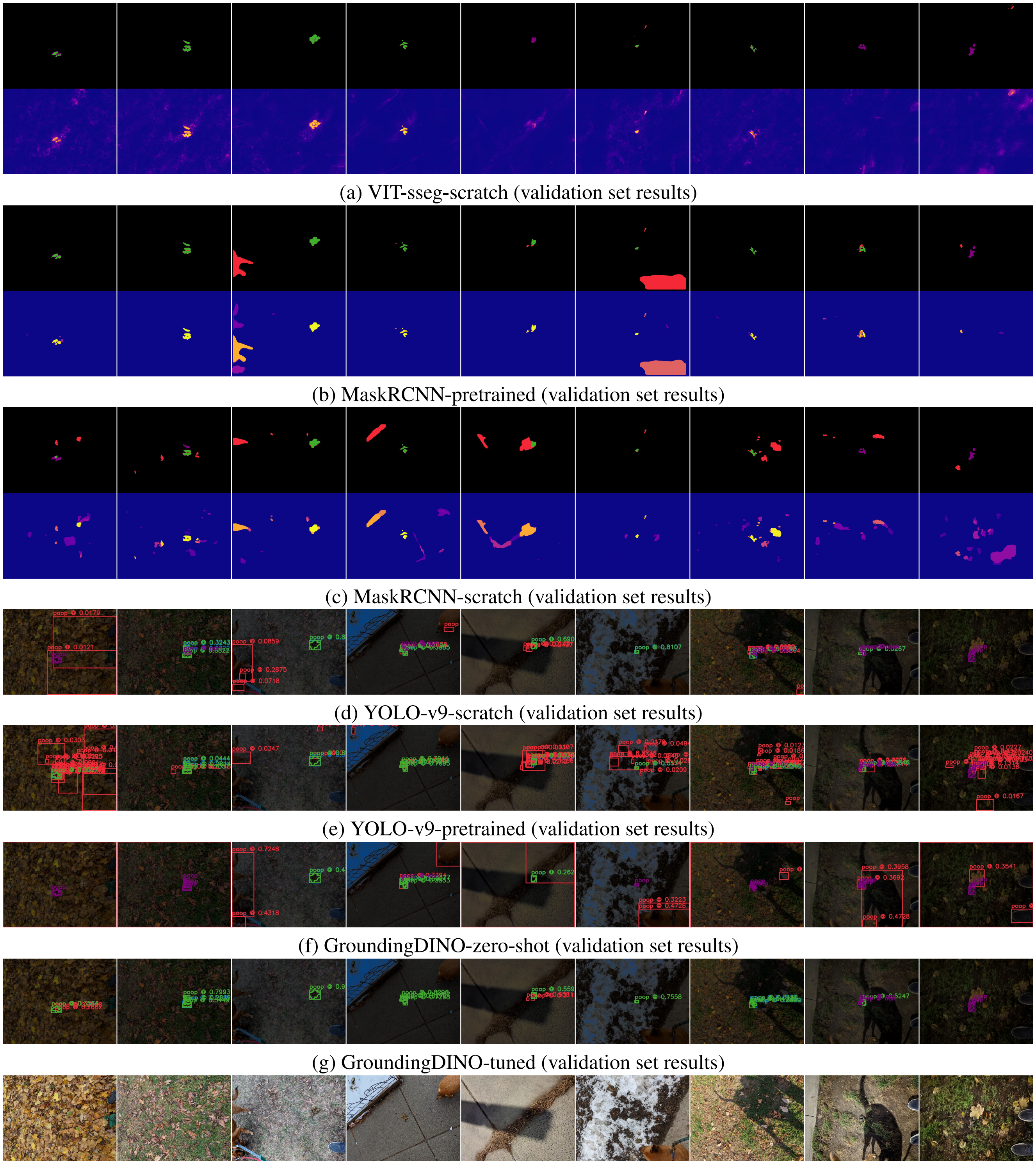
Table 3. Transfer times (in hours) for our 42GB dataset: trials (n), mean (μ), std (σ). Each experiment was run 5 times. Uncompressed transfers provide granular access to individual files, while compressed (zipped) transfers are faster.

We have published our data and models under a permissive license, and made them available through both centralized (Girder and HuggingFace) and decentralized (BitTorrent and IPFS) mechanisms. Decentralized methods are robust, but suffer from significant network transfer overhead. HuggingFace has exceptionally fast transfer speeds, has some decentralized properties, but lacks content identifiers.

Our dataset enables applications such as mobile feces detection, urban cleanliness monitoring, and augmented-reality collision warnings. Because it trains models to recognize small, irregular, low-contrast objects in cluttered scenes, we predict that including "ScatSpotter" in foundational training corpora will improve robustness to camouflage and small-object ambiguity in a broad range of ecological and waste-monitoring downstream tasks.

7. Acknowledgements

We would like to thank all of the dogs that produced subject matter, all of the contributors for helping to construct the test set, and Anthony Hoogs for several suggestions including taking the third negative picture. This work is dedicated to Bezoar, a weird and good girl; Honey, a sweet red-fox lookalike; and Roadie, a vicious soft-faced cuddlebug.



(a) VIT-sseg-scratch (validation set results)

(b) MaskRCNN-pretrained (validation set results)

(c) MaskRCNN-scratch (validation set results)

(d) YOLO-v9-scratch (validation set results)

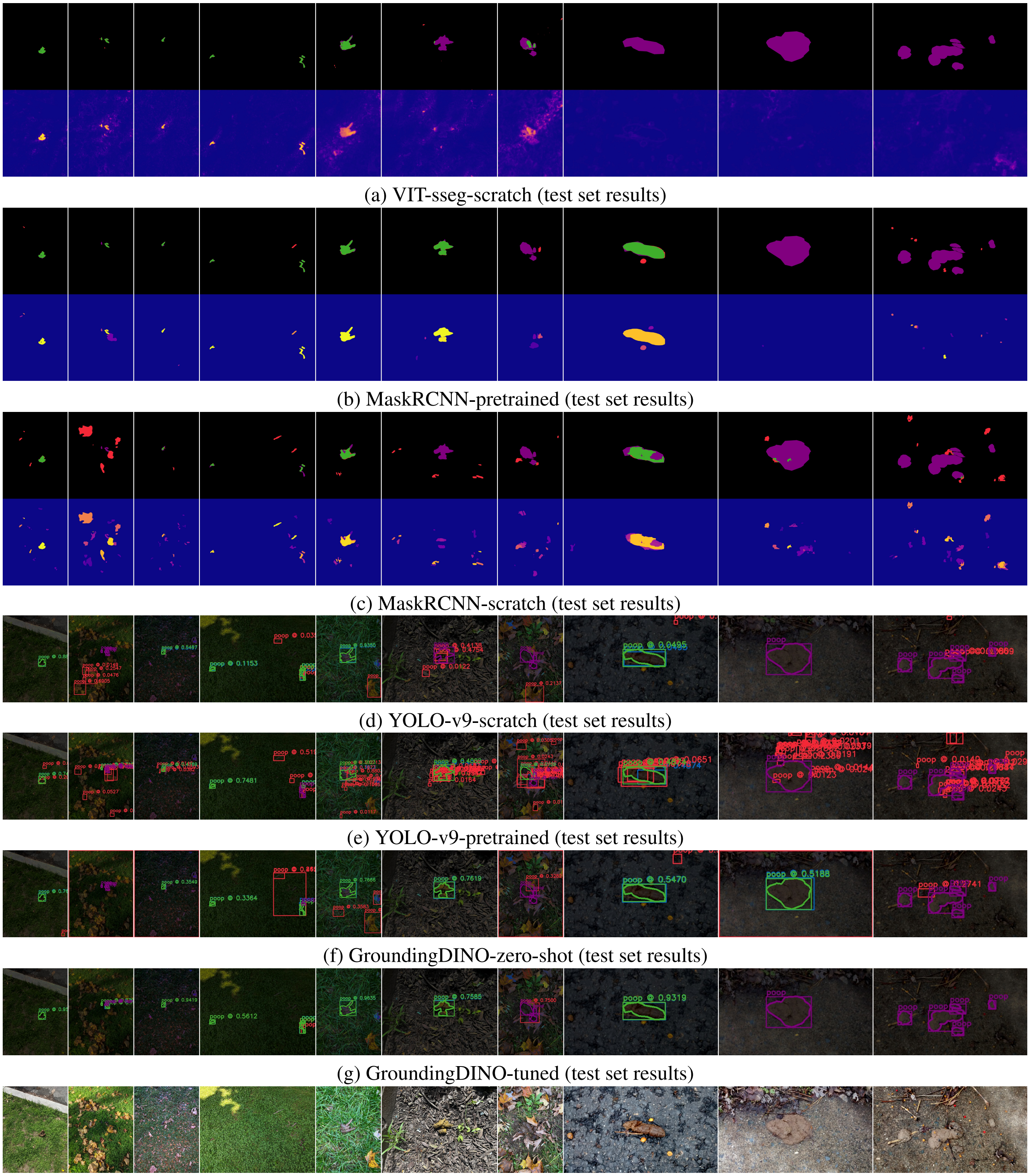
(e) YOLO-v9-pretrained (validation set results)

(f) GroundingDINO-zero-shot (validation set results)

(g) GroundingDINO-tuned (validation set results)

(h) Inputs from the validation set

Figure 5. Qualitative results from validation-selected models applied to the same validation images. Subfigures (a-c) show results for VIT and MaskRCNN, including both the binarized classification map (true positives in green, false positives in red, false negatives in purple, true negatives in black) and the predicted heatmap before binarization. Subfigures (d-g) show bounding-box detections from YOLO-v9 and Grounding DINO, using the same color scheme (blue = true-positive predicted boxes; green = matched ground truth). Subfigure (h) shows the input image.



(a) VIT-sseg-scratch (test set results)

(b) MaskRCNN-pretrained (test set results)

(c) MaskRCNN-scratch (test set results)

(d) YOLO-v9-scratch (test set results)

(e) YOLO-v9-pretrained (test set results)

(f) GroundingDINO-zero-shot (test set results)

(g) GroundingDINO-tuned (test set results)

(h) Inputs from the test set

Figure 6. Qualitative results from validation-selected models applied to test images. Subfigures (a-c) show results for VIT and MaskRCNN, including both the binarized classification map (true positives in green, false positives in red, false negatives in purple, true negatives in black) and the predicted heatmap before binarization. Subfigures (d-g) show bounding-box detections from YOLO-v9 and Grounding DINO, using the same color scheme (blue = true-positive predicted boxes; green = matched ground truth). Subfigure (h) shows the input image.

References

[1] Jordan T. Ash and Ryan P. Adams. On Warm-Starting Neural Network Training. In *NeurIPS*. Curran Associates, Inc, 2020. 19, 20

[2] Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, 2016. 13

[3] Dina Bashkirova, Mohamed Abdelfattah, Ziliang Zhu, James Akl, Fadi Alladkani, Ping Hu, Vitaly Ablavsky, Berk Calli, Sarah Adel Bargal, and Kate Saenko. ZeroWaste Dataset: Towards Deformable Object Segmentation in Cluttered Scenes. In *CVPR*, pages 21147–21157, 2022. 2, 3

[4] Juan Benet. IPFS - Content Addressed, Versioned, P2P File System. *ArXiV*, abs/1407.3561, 2014. 3, 13

[5] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *ICML*, page 4, 2021. 19

[6] Christian Bieri. An overview into the InterPlanetary File System (IPFS): use cases, advantages, and drawbacks. *Communication Systems XIV*, 28, 2021. 13

[7] Scott Chacon and Ben Straub. Pro git, 2014. 13

[8] Bram Cohen. Incentives Build Robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer systems*, pages 68–72, 2003. 3, 13

[9] Bram Cohen. The BitTorrent Protocol Specification v2. https://www.bittorrent.org/beps/bep_0052.html, 2017. Accessed: 2024-08-23. 13

[10] Joseph Paul Cohen and Henry Z. Lo. Academic torrents: A community-maintained distributed repository. In *Annual Conference of the Extreme Science and Engineering Discovery Environment*, 2014. 5

[11] Christian Collberg and Todd A Proebsting. Repeatability in computer systems research. *Communications of the ACM*, 59 (3):62–69, 2016. 13

[12] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset. In *CVPRW*, page 1, 2015. 2, 3

[13] Jon Crall, Connor Greenwell, David Joy, Matthew Leotta, Aashish Chaudhary, and Anthony Hoogs. GeoWATCH for Detecting Heavy Construction in Heterogeneous Time Series of Satellite Images. In *IGARSS*, 2024. 5, 19

[14] Abhyuday Desai, Mohamed Abdelhamid, and Nakul R. Padalkar. What is Reproducibility in Artificial Intelligence and Machine Learning Research? *ArXiV*, abs/2407.10239, 2024. 13

[15] Shibhansh Dohare, J. Fernando Hernandez-Garcia, Parash Rahman, Richard S. Sutton, and A. Rupam Mahmood. Loss of Plasticity in Deep Continual Learning. *ArXiV*, abs/2306.13812, 2023. arXiv:2306.13812 [cs]. 19, 20

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR*, 2021. 19

[17] Paul Frazee and Mathias Buss. DEP-0002: Hypercore - Dat Protocol. https://www.datprotocol.com/deps/0002-hypercore/, 2018. Accessed: 2024-08-23. 3

[18] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. 3

[19] Jim Gettys and Kathleen Nichols. Bufferbloat: dark buffers in the internet. *Communications of the ACM*, 55(1):57–65, 2012. 5

[20] Connor Greenwell, Jon Crall, Matthew Purri, Kristin Dana, Nathan Jacobs, Armin Hadzic, Scott Workman, and Matt Leotta. Watch: Wide-area terrestrial change hypercube. In *WACV*, pages 8277–8286, 2024. 5, 19

[21] Devindra Hardawar. iRobot's latest Roomba can detect pet poop (and if it fails, you'll get a new one). https://www.engadget.com/irobot-roomba-j-7-object-poop-detection-040152887.html, 2021. Accessed: 2024-08-23. 3

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[23] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, pages 2961–2969, 2017. 5

[24] Toke Høiland-Jørgensen, Dave Täht, and Jonathan Morton. Piece of cake: a comprehensive queue management solution for home gateways. In *2018 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, pages 37–42. IEEE, 2018. 5

[25] Jungseok Hong, Michael Fulton, and Junaed Sattar. Trashcan: A semantically-segmented dataset towards visual detection of marine debris. *ArXiV*, abs/2007.08097, 2020. 2, 3

[26] Jeromy Johnson, Juan Benet, Steven Allen, et al. ipfs/kubo. https://github.com/ipfs/kubo, 2024. Accessed: 2024-08-23. 5

[27] Wada Kentaro. Labelme: Image polygonal annotation with python. https://github.com/labelmeai/labelme, 2016. Accessed: 2024-08-23. 4

[28] Keith Kirkpatrick. The Carbon Footprint of Artificial Intelligence. *Communications of the ACM*, 66(8):17–19, 2023. 21

[29] Marek Kraft, Mateusz Piechocki, Bartosz Ptak, and Krzysztof Walas. Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle. *Remote Sensing*, 13(5), 2021. 2, 3

[30] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning. *ArXiV*, abs/1910.09700, 2019. 5, 21

[31] Jordan Lee, Josh Elsasser, Eric Petit, and Mitchell Livingston. Transmission. https://github.com/transmission/transmission, 2024. Accessed: 2024-08-23. 5

[32] Quentin Lhoest, Albert Villanova Del Moral, Yacine Jernite, Abhishek Thakur, Patrick Von Platen, Suraj Patil, Julien Chaumond, Mariama Drame, Julien Plu, Lewis Tunstall, et al.

Datasets: A community library for natural language processing. *arXiv preprint arXiv:2109.02846*, 2021. Accessed: 2025-04-26. 2, 5, 13

[33] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *Computer Vision - ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing. 2, 3, 5

[34] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *CVPR*, pages 2980–2988, 2017. 19

[35] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection, 2024. arXiv:2303.05499 [cs]. 1, 5

[36] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *ICLR*, 2019. 19

[37] Petar Maymounkov and David Mazières. Kademlia: A Peer-to-Peer Information System Based on the XOR Metric. In *Peer-to-Peer Systems*, pages 53–65. Springer, Berlin, Heidelberg, 2002. 13

[38] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiV*, 2020. 3

[39] Mikian. Dog Poop (MSHIT). https://www.kaggle.com/datasets/mikian/dog-poop, 2020. Accessed: 2024-08-23. 2, 3

[40] Agnieszka Mikołajczyk. Waste datasets review. https://github.com/AgaMiko/waste-datasets-review, 2024. Accessed: 2024-09-07. 3

[41] Gaurav Mittal, Kaushal B Yagnik, Mohit Garg, and Narayanan C Krishnan. Spotgarbage: smartphone app to detect garbage using deep learning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 940–945, 2016. 2, 3

[42] Zack Mullen, Brian Helba, David Manthy, et al. Girder: a data management platform. https://girder.readthedocs.io/en/latest, 2024. Accessed: 2024-08-23. 2, 5

[43] Neeraj Madan. Dog Poop Detection: Deep Learning (Details). https://www.youtube.com/watch?v=qGNbHwp0jM8, 2019. Accessed: 2024-08-23. 3

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 6, 20

[45] David Martin Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011. 5, 20

[46] Pedro F Proença and Pedro Simões. Taco: Trash annotations in context for litter detection. *ArXiV*, abs/2003.06975, 2020. 2, 3

[47] Edward Raff. A step toward quantifying independently reproducible machine learning research. In *NeurIPS*. Curran Associates, Inc., 2019. 13

[48] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 115(3):211–252, 2015. 2, 3

[49] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *CVPR*, pages 4510–4520, 2018. 1

[50] Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *CVPR*, pages 815–823, 2015. 4

[51] Mennatullah Siam, Mostafa Gamal, Moemen Abdel-Razek, Senthil Yogamani, and Martin Jagersand. RTSeg: Real-Time Semantic Segmentation Comparative Study. In *ICIP*, pages 1603–1607, 2018. 1

[52] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, pages 369–386. SPIE, 2019. 19

[53] Dominic Tarr, Erick Lavoie, Aljoscha Meyer, and Christian Tschudin. Secure Scuttlebutt: An Identity-Centric Protocol for Subjective and Decentralized Applications. In *ACM Conference on Information-Centric Networking*, pages 1–11, New York, NY, USA, 2019. Association for Computing Machinery. 3

[54] The WebDataset Contributors. Webdataset. GitHub repository and documentation, 2023. Accessed: 2025-04-26. 5

[55] Andrew Tridgell, Paul Mackerras, et al. rsync. https://github.com/RsyncProject/rsync, 2024. Accessed: 2024-08-23. 5

[56] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. In *ECCV*, pages 1–21. Springer, 2024. 5

[57] Kang Yu, Guoxin Tang, Wen Chen, Shanshan Hu, Yanzhou Li, and Haibo Gong. MobileNet-YOLO v5s: An Improved Lightweight Method for Real-Time Detection of Sugarcane Stem Nodes in Complex Natural Environments. *IEEE Access*, 11:104070–104083, 2023. 1

[58] Wei Li Zuwei Long. Open grounding dino:the third party implementation of the paper grounding dino. https://github.com/longzw1997/Open-GroundingDino, 2023. 5