Roboflow 100: A Rich, Multi-Domain Object Detection Benchmark

Floriana Ciaglia*, Francesco Saverio Zuppichini*, Paul Guerrie*, Mark McQuade*, and Jacob Solawetz*

*{floriana, paul, francesco, mark, jacob}@roboflow.com

Abstract

The evaluation of object detection models is usually performed by optimizing a single metric, e.g., mAP, on a fixed set of datasets, e.g., Microsoft COCO and Pascal VOC. Due to image retrieval and annotation costs, these datasets consist largely of images found on the web and do not represent many real-life domains that are being modelled in practice, e.g. satellite, microscopic and gaming, making it difficult to assert the degree of generalization learned by the model.

We introduce the Roboflow-100 (RF100) consisting of 100 datasets, 7 imagery domains, 224,714 images, and 805 class labels with over 11,170 labelling hours. We derived RF100 from over 90,000 public datasets, 60 million public images that are actively being assembled and labelled by computer vision practitioners in the open on the web application Roboflow Universe. By releasing RF100, we aim to provide a semantically diverse, multi-domain benchmark of datasets to help researchers test their model’s generalizability with real-life data. RF100 download and benchmark replication are available on GitHub.

1 Introduction

In object detection research, Microsoft COCO[Lin et al., 2014] and Pascal VOC[Everingham et al., 2010] have become the de facto benchmark standards to train and evaluate the performance of models. These models are then released to the public who fine-tuned them on a smaller dataset consisting of specific imagery and targets of interest. While the general object detection benchmarks provide a proxy for how the model will perform in a similar setting, there is no substitute for domain-specific training.

There is a strong research interest in evaluating models on a more diverse task set. For example, 13 Roboflow community open source datasets have organically been used by researchers to create The Object Detection in the Wild (ODINW) [Li et al., 2021] benchmark. This benchmark was utilized to assert model performance for object detection and zero-shot capabilities in Florence [Yuan et al., 2021] and GLIP [Li et al., 2021]. Through curating a larger set of narrow task datasets, we build on this interest for deeper domain-specific assessment and introduce a stronger benchmark we called Roboflow 100 (RF100).

RF100 consists of a collection of 100 crowdsourced object detection (OD) datasets, specifically constructed by Roboflow users to represent a chosen subject of study. In Figure 1 we show a small sample of annotated image domains that the RF100 benchmark encompasses, which range from satellite and aerial
By introducing a benchmark of narrow subject matter datasets, we accomplish two goals. Firstly, our benchmark provides a strong collection of closed-domain tasks used in the wild that are of demonstrated interest to practitioners for researchers that are designing models with the intent of them being fine-tuned. Secondly, researchers building general models can test transfer between object detection tasks in zero-shot or few-shot fashion.

2 Related Work

Historically, object detection datasets are created by gathering a large corpus of images and sourcing annotators to label objects in a fixed set of classes.

The Pascal VOC project [Everingham et al., 2010] is a collection of datasets made to enhance object detection tasks and encourage researchers to create models that recognize objects in realistic scenes in the form of a challenge. The Pascal VOC challenges started in 2005 and laid the groundwork for a new generation of state-of-the-art benchmarks.

ImageNet [Deng et al., 2009] is an image dataset comprised of over 14 million images each described by word phrases called "synset". It was specifically created to answer the need in the industry for a high-quality object categorization benchmark with clearly established evaluation metrics. Similarly to RF100, ImageNet was created to encourage the creation of more generalizable machine learning models.

Open Images [Kuznetsova et al., 2020] is an image collection with over 9 million annotated images and 600 object classes. It was created to enable the study of tasks such as image classification, object detection, visual relationship detection, instance segmentation, and multimodal image descriptions all from one joined resource to stimulate progress towards image scene comprehension.

The Common Objects in Context [Lin et al., 2014] (COCO) benchmark is a large-scale object detection and segmentation dataset with a total of 2.5M labelled instances in 328K images. Ever since its release, the COCO benchmark has established itself as the state-of-the-
The Objects365 [Shao et al., 2019] dataset is a large collection of separate object detection datasets that is comprised of images from the website Flicker. Images are gathered, categories are assigned, and then a labelling team is employed to create annotations. Pretraining models on Objects365 are shown to benefit performance on downstream tasks, such as COCO.

The Object Detection in the Wild (ODINW) [Li et al., 2021] dataset was released in the same spirit as RF100. The original ODINW version uses 13 Roboflow object detection datasets to assess the generalizability of their zero-shot model, GLIP [Li et al., 2021]. Despite advances in open vocabulary object detectors like GLIP, accurate and fast object detection still requires custom training to be performed on quality, annotated data with a closed vocabulary.

Unlike prior related work, we assemble a large-scale object detection dataset that is sourced via image upload and annotation by practitioners on a web application who are using computer vision to accomplish real tasks.

3 Methods

In this section, we describe the RF100 dataset creation process and our initial modelling experiments on the new benchmark.

3.1 Dataset Collection and Distribution

Roboflow Universe is a public repository of computer vision dataset that over 100,000 Roboflow users have assembled and labeled.
for their own custom use cases. We selected 100 datasets from Roboflow Universe for our benchmark using the following criteria:

- Effort - the user spent substantial labelling hours working on the task;
- Diversity - the user was working on a novel task;
- Quality - the user annotated with high fidelity to the task;
- Substance - the user assembled a substantial dataset with nuance;
- Feasibility - the user was attempting a learnable task;

After selection for inclusion, all datasets were processed in the following way:

- All images resized to 640x640 pixels following best practices [Wang et al., 2022]
- Eliminate class ambiguity, i.e. if a class was labeled by the original author as 0 to represent a flower, that class label would be changed to a word descriptive of the actual subject such as *daisy*
- The train, validation, and test split were manipulated only in the instances where either one or more split sets were missing completely, or where one, or more, of the split sets were extremely under-represented. In all other cases, we respected the split ratio chosen by the original author of the dataset.
- Underrepresented classes were filtered when they represented less than 0.5 percent of all objects in a dataset. These classes tended to be labeling errors.

The datasets are available for download from GitHub or from the Roboflow Universe website by clicking on the *Export* button.

| Category      | Datasets | Images  | Classes |
|---------------|----------|---------|---------|
| Aerial        | 7        | 9683    | 24      |
| Videogames    | 7        | 11579   | 88      |
| Microscopic   | 11       | 13378   | 28      |
| Underwater    | 5        | 18003   | 39      |
| Documents     | 8        | 24813   | 90      |
| Electromagnetic| 12      | 36381   | 41      |
| Real World    | 50       | 110615  | 495     |
| **Total**     | **100**  | **224,714** | **805** |

Table 2: Overview of per-category metadata, including number of datasets, number of images, and number of classes across categories.

### 3.2 Semantics

We selected seven different semantic categories to achieve comprehensive coverage of real-life possible domains: Aerial, Videogames, Microscopic, Underwater, Documents, Electromagnetic and Real World.

The Real World category is the biggest in the RF100 benchmark since the majority of use cases for computer vision involve everyday scenes. We included indoor, outdoor, Vehicles, animals, plants, damage control, safety, electronics, geology, board games and various human activity images.

The Video Games category includes virtual reality scenes, robot fighting, first-person shooters and MOBA; The Underwater category includes fishery sights from both seas and aquariums, as well as inanimate objects found underwater (i.e. pipes).

The Microscopic category is comprised of items that can only be seen with the aid of a microscopic lens like bacteria and human cells; the Aerial category includes images taken from an overhead view, including images from space and drones; the Electromagnetic category includes scenarios where electromagnetic waves were used to capture the
pictures, X-rays, MRIs, IR, thermal and night vision cameras; and finally, in the Documents category, we include all images that relate to articles, papers, tables, diagrams and social media.

Figure 4 shows samples for each category.

3.3 Data Statistics and Analysis

Table 2 summarizes the benchmark’s metadata including the number of datasets, images and classes present in each category. Per dataset statistics and results can be seen in Table 4.

Figure 2 reports different statistics grouped by category, such as number of classes and bounding boxes area. Most notably, Aerial, Microscopic and Electromagnetic have smaller bounding boxes compared to the rest. Moreover, the average number of classes across different categories is only ten, meaning in practice people need to identify a small set of objects.

Figure 3 shows a scatter plot produced to analyze the vector clustering degree of the RF100 categories using CLIP embeddings [Radford et al., 2021] generates for each of its datasets. This illustration shows that the datasets in each category do tend to cluster together. You can also view an image-level clustering of these semantics on the RF100 web exploration web demo.

3.4 Experiments and Evaluation

We trained popular object detection model architectures on RF100 and report the results. Only one model instance was trained per dataset.

Finetuning: We finetuned two comparable versions of YOLOv5 [Jocher et al., 2020] and YOLOv7[Wang et al., 2022]: YOLOv5s and YOLOv7, with 7.2M parameters and 36.9M parameters respectively and similar FPS when evaluated on a Tesla V100.

We trained both models with default hyperparameters for 100 epochs at 640x640 resolution.

Zero Shot: We also evaluated GLIP [Li et al., 2022], a zero-shot detector that can solve open vocabulary detection by rephrasing it as
Table 3: Experiments results on RF100. We recorded the average mAP@.50 value for the YOLOv5 and YOLOv7 models and the mAP@.50:.95 for the GLIP model for each category.

| Category      | YOLOv5 | YOLOv7 | GLIP |
|---------------|--------|--------|------|
| Aerial        | 0.636  | 0.504  | 0.230|
| Videogames    | 0.859  | 0.796  | 0.188|
| Microscopic   | 0.650  | 0.591  | 0.159|
| Underwater    | 0.560  | 0.662  | 0.019|
| Documents     | 0.716  | 0.722  | 0.024|
| Electromagnetic| 0.689  | 0.607  | 0.058|
| Real World    | 0.752  | 0.699  | 0.108|
| **Total**     | **0.694** | **0.654** | **0.112** |

4 Discussion

Our initial benchmarks show that there is variation in model performance between models across datasets and datasets domains that may run contrary to the model’s ranking on incumbent benchmarks. A given model may perform better on one dataset and worse on another. While we do not investigate the underlying reasons for performance differentials, our results suggest that there are likely significant improvements to be made to object detection models to expect a wider array of custom datasets that they may be applied to.

Lastly, our evaluation shows that zero-shot object detection models lose considerable performance when extended to new domains.

5 Conclusion

We introduce the RF100 object detection benchmark of 100 datasets to encourage the evaluation of object detection model performance to test model generalizability across a wider array of imagery domains. Our initial evaluation shows that the new RF100 benchmark will provide valuable insights into how new object detection models will perform in the wild. RF100 is available for download on GitHub.

Acknowledgments

We thank all of the advisors we have had on our research both internally at Roboflow, and externally in the machine learning community. We would also like to thank Intel for sponsoring the work involved in constructing the RF100 benchmark.

Finally, we thank everyone working on public computer vision datasets on Roboflow.
Universe and in particular, the creators of the RF100 datasets: Abhishek Dada, Adam Crenshaw, Adrian Rodriguez, Ahmad Rabbiee, Alex Hyams, Aman Ahuja and Alan Devera, Ammar Abdalmutalib, Amro, Anshul Rankawat, Kapil Verma, Shubhankar Rawat, Manisa Mondal, Pranav Arora, Arifianti Nur Sayidah, Brad Dwyer, Brad Dwyer, CC Moon, Chang Yuan, Dane Sprsiter, David Lee, Djamel Mechklouf, Abrisse Cerine, Anfal Lanna, Yasmin Emekhlouf, Evan Kim, MJ Kim, Graham Doerksen, Ilyes Talbi, Jan Douwe, Jason Zhang, Cadin Li, Jhonathan, Joao Paulo Martins, Jordan Bird, Leah Bird, Carrie Ijichi, Aurelie Jolivald, Salisu Wada, Kay Owa, Chloe Barnes, Joseph Nelson, Brad Dwyer, and Cheng Hsun Teng, Justin Henke, Reginald Viray, Kais Al Hajjih, Karen Weiss, Kat Laura, Lao and Shiguang, Lukas D. Ringle, Matteo Pacini, Melanie S. Capalungan, B-Jay Daguio, Isaac Balbuena, and Reanne Joy Rafael, Mevlir Crasta, Miguel Fernández Cruchaga, Mike Drickramer, Minoj Selvaraj, Mohamed Attia, Mohamed Refai, Abarna, Amjad Hafiz, Sutheshan Maiu, and Thanshua Sritharan, Mohamed Sabek, Monika Patel, Kartik Attri, Aniket Dhanotia, Divyam Jha, Pankaj, kanchan, Ujjwal Sharma, Garvita Vijay, Aniket Choudhary, Pearl Rathour, Roshni Ghai, Kavya Shukla, Preeti Sharma, Ananya Kharatay, Krishna Gambhir, Ayush Sahu, Ujjwal Sharma, Divyam Jha, Kanchan, Kartik Attri, Lav Naruka, Kas, Preeta Sharma, Terada Shoma, Thuan Phat Nguyen, Vanitchaporn, Victor Perez, Stephen Groff, Mason Hintermeister, Wang Tianyi, Wilfred Shu and Adrian Stuart, Wojciech Blachowski, Wojciech Przydzial, Dorota Przydzial, Magdalena Przydzial, Mazur, and Bartlomiej Mazur, Xingwei He, Yilong Zheng, Yimin Chen, Yousef Ghanem, Yuanyu Anpei, Yudha Bhakti Nugraha and Kris, Yuntaewon, Hwanghyeyun, Gimminseo, Gimnohyeon, Sindahong, Gimseongsu, Yuyang Li, Zhang Kaimin, Zhe Fan.
References

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. Int. J. Comput. Vis., 88(2):303–338, 2010. URL http://dblp.uni-trier.de/db/journals/ijcv/ijcv88.html#EveringhamGWWZ10.

Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammmana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, Guilhen, Hatovix, Jake Poznanski, Jiacong Fang, Lijun Yu, changyu98, Mingyu Wang, Naman Gupta, Osama Akhtar, PetrDvo-racek, and Prashant Rai. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements, October 2020. URL https://doi.org/10.5281/zenodo.4154370.

Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, Tom Duerrig, and Vittorio Ferrari. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. IJCV, 2020.

Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, Kai-Wei Chang, and Jianfeng Gao. Grounded language-image pre-training, 2021. URL https://arxiv.org/abs/2112.03857.

Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, et al. Grounded language-image pre-training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10965–10975, 2022.

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2014. URL https://arxiv.org/abs/1405.0312.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. URL https://arxiv.org/abs/2103.00020.

Shuai Shao, Zeming Li, Tianyuan Zhang, Chao Peng, Gang Yu, Xiangyu Zhang, Jing Li, and Jian Sun. Objects365: A large-scale, high-quality dataset for object detection. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 8429–8438, 2019. doi: 10.1109/ICCV.2019.00852.

Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.

Lu Yuan, Dongdong Chen, Yi-Ling Chen, Noel Codella, Xiyang Dai, Jianfeng Gao, Houdong Hu, Xuedong Huang, Boxin Li, Chunyuan Li, Ce Liu, Mengchen Liu, Zicheng Liu, Yumao Lu, Yu Shi, Lijuan Wang, Jianfeng Wang, Bin Xiao, Zhen Xiao, Jianwei Yang, Michael Zeng, Luowei Zhou, and Pengchuan Zhang. Florence: A new foundation model for computer vision, 2021. URL https://arxiv.org/abs/2111.11432.
A Appendix

(a) **Areal** images gather from drone, space and static cameras.

(b) **Video Games** screen recordings from different games, such as Far Cry, Apex Legends, CS Go an League of Legends.

(c) **Microscopic** images, mostly human diseases recorded with medical equipment showing cells, parasites, bacteria.

(d) **Underwater** images of various sea plants and animal collected in the sea or aquariums.

(e) **Documents** such as tweets, tables and activity diagrams.

(f) **Electromagnetic** images from X-ray/thermal cameras and MRI.

(g) **Real World** images from a wide array of domains, animals, vehicles, human activities, paintings and electronics.

Figure 4: Examples of images samples from different categories. *Real World* was samples more due to its bigger size.
| Dataset                  | Category     | Images | Labeling | YOLOv5 mAP@.50 | YOLOv7 mAP@.50 | GLIP  mAP@.50:95 |
|-------------------------|--------------|--------|----------|----------------|----------------|-------------------|
| aerial pool             | aerial       | 673    | 96       | 177            | 421            | 0.513             |
| secondary chains        | aerial       | 103    | 16       | 43             | 201            | 0.341             |
| aerial spheres          | aerial       | 318    | 51       | 104            | 177            | 0.993             |
| soccer players          | aerial       | 114    | 16       | 33             | 0              | 0.660             |
| weed crop               | aerial       | 823    | 118      | 235            | 0              | 0.820             |
| aerial cows             | aerial       | 1084   | 299      | 340            | 179            | 0.854             |
| cloud types             | aerial       | 3528   | 504      | 1008           | 0              | 0.271             |
| apex videogame         | videogames   | 2583   | 415      | 691            | -1             | 0.839             |
| farcry6 videogame       | videogames   | 82     | 14       | 24             | 0              | 0.619             |
| csgo videogame         | videogames   | 1774   | 207      | 446            | 0              | 0.974             |
| avatar recognition      | videogames   | 225    | 30       | 59             | 3              | 0.889             |
| halo infinite           | videogames   | 462    | 71       | 136            | 4              | 0.921             |
| team fight              | videogames   | 1162   | 112      | 307            | 88             | 0.961             |
| robomasters 285km       | videogames   | 1945   | 278      | 556            | 27             | 0.816             |
| stomata cells           | microscopic  | 1482   | 209      | 414            | 0              | 0.840             |
| bcccd ouzjz             | microscopic  | 255    | 36       | 73             | 0              | 0.912             |
| parasites 1s07h         | microscopic  | 1484   | 215      | 411            | 0              | 0.848             |
| cells uyemf             | microscopic  | 16     | 2        | 4              | 210            | 0.249             |
| 4 fold                  | microscopic  | 503    | 33       | 134            | 279            | 0.970             |
| bacteria ptywi          | microscopic  | 30     | 10       | 10             | 472            | 0.162             |
| cotton plant            | microscopic  | 724    | 102      | 198            | 259            | 0.204             |
| mitosis gjs3g           | microscopic  | 213    | 30       | 61             | 0              | 0.931             |
| phages                  | microscopic  | 1155   | 103      | 164            | 74             | 0.854             |
| liver disease           | microscopic  | 2782   | 400      | 794            | 31             | 0.592             |
| asbestos                | microscopic  | 932    | 133      | 266            | 126            | 0.596             |
| underwater pipes        | underwater   | 5617   | 779      | 1575           | 316            | 0.995             |
| aquarium qlnqy          | underwater   | 448    | 63       | 127            | 0              | 0.790             |
| peixos fish             | underwater   | 821    | 118      | 261            | 0              | 0.148             |
| underwater objects      | underwater   | 5320   | 760      | 1520           | 0              | 0.693             |
| coral lwptl             | underwater   | 427    | 74       | 93             | 165            | 0.174             |
| tweeter posts           | documents    | 87     | 9        | 21             | 0              | 0.708             |
| tweeter profile         | documents    | 425    | 61       | 121            | 0              | 0.988             |
| document parts          | documents    | 906    | 150      | 318            | 192            | 0.677             |
| activity diagrams       | documents    | 259    | 45       | 74             | 192            | 0.427             |
| signatures xc8up        | documents    | 257    | 37       | 74             | 0              | 0.961             |
| paper parts             | documents    | 8472   | 1209     | 2359           | 211            | 0.590             |
| tabular data            | documents    | 3251   | 206      | 409            | 271            | 0.752             |
| paragraphs co84b        | documents    | 4209   | 633      | 1221           | 228            | 0.626             |
| thermal dogs            | electromagnetic | 142  | 20  | 41  | -1   | 2   | 0.967 |
| solar panels            | electromagnetic | 112  | 19  | 30  | 175  | 5   | 0.413 |
| radio signal            | electromagnetic | 1954 | 278 | 566  | 0   | 2 | 0.673 |
| thermal cheetah | electromagnetic | 90 14 25 | 0 2 | 0.931 0.513 0.028 |
|-----------------|----------------|----------|-----|-------------------|
| x ray           | electromagnetic | 135 16 34 | 16 12 | 0.722 0.506 0.000 |
| acl x           | electromagnetic | 2141 306 612 | 0 1 | 0.995 0.998 0.000 |
| abdomen mri     | electromagnetic | 1887 238 479 | 0 1 | 0.965 0.958 0.021 |
| axial mri       | electromagnetic | 253 39 79 | 0 2 | 0.638 0.549 0.039 |
| gynecology mri  | electromagnetic | 2122 253 526 | 7 3 | 0.323 0.171 0.000 |
| brain tumor     | electromagnetic | 6930 1990 | 0 3 | 0.768 0.809 0.003 |
| bone fracture   | electromagnetic | 326 44 88 | 0 4 | 0.085 0.090 0.000 |
| flir camera     | electromagnetic | 9306 1452 2854 | 17 4 | 0.796 0.824 0.073 |
| hand gestures   | real world | 642 94 178 | -1 14 | 0.995 0.995 n/a |
| smoke uvylj     | real world | 522 76 148 | 7 1 | 0.959 0.962 0.431 |
| wall damage     | real world | 325 40 96 | -1 3 | 0.500 0.434 n/a |
| corrosion bi3q3 | real world | 820 105 304 | 186 3 | 0.768 0.764 0.003 |
| excavators czvg9| real world | 2244 144 267 | 0 3 | 0.946 0.895 0.274 |
| chess pieces    | real world | 202 29 58 | 0 13 | 0.977 0.830 0.017 |
| road signs      | real world | 1376 229 488 | 0 21 | 0.963 0.944 0.036 |
| street work     | real world | 611 87 175 | 2 11 | 0.478 0.708 0.148 |
| construction safety | real world | 997 90 119 | 505 5 | 0.915 0.915 0.259 |
| road traffic    | real world | 494 133 187 | -1 12 | 0.597 0.847 n/a |
| washroom rf1fa  | real world | 1885 318 775 | 449 10 | 0.619 0.634 0.146 |
| circuit elements | real world | 672 36 64 | 311 46 | 0.063 n/a 0.001 |
| mask wearing    | real world | 105 15 29 | 0 2 | 0.788 0.513 0.008 |
| cables m42k     | real world | 4816 794 1220 | 0 11 | 0.688 0.722 0.010 |
| soda bottles    | real world | 1547 216 486 | 243 6 | 0.964 n/a 0.098 |
| truck movement  | real world | 740 107 215 | 282 7 | 0.786 0.846 0.007 |
| wine labels     | real world | 3172 630 841 | 249 12 | 0.569 0.632 0.045 |
| digits t2eg6    | real world | 2912 367 824 | 144 10 | 0.999 0.989 0.003 |
| vehicles q0x2v  | real world | 2634 458 966 | 1121 -1 | 0.454 0.464 0.029 |
| peanuts sd4kf   | real world | 268 42 77 | 212 2 | 0.995 0.997 0.358 |
| printed circuit | real world | 548 44 80 | 311 34 | 0.091 n/a 0.000 |
| pests 2x1v      | real world | 509 55 153 | 188 28 | 0.136 0.029 0.004 |
| cavity rs0uf    | real world | 287 38 93 | 165 2 | 0.782 0.799 0.029 |
| leaf disease    | real world | 1589 296 616 | 143 3 | 0.531 0.560 0.027 |
| marbles         | real world | 54 32 19 | 133 2 | 0.992 0.473 0.030 |
| pills sx8ht     | real world | 316 45 90 | 0 8 | 0.869 0.867 0.194 |
| poker cards     | real world | 964 128 193 | 0 53 | 0.886 0.251 -0.000 |
| number ops      | real world | 4869 623 1636 | 28 15 | 0.990 0.992 0.055 |
| insects mytwu   | real world | 696 100 199 | 0 10 | 0.890 0.858 0.024 |
| cotton 20xz5    | real world | 367 20 19 | 17 4 | 0.569 0.591 0.157 |
| furniture ngpea | real world | 454 74 161 | 0 3 | 0.983 0.968 0.836 |
| cable damage    | real world | 919 134 265 | 2 2 | 0.910 0.574 0.006 |
| animals ij5d2   | real world | 700 100 200 | 12 10 | 0.761 0.342 0.249 |
| coins 1apki     | real world | 6121 699 1599 | 0 4 | 0.932 0.977 0.175 |
| apples fvp5     | real world | 489 30 178 | -1 2 | 0.779 0.791 n/a |
|                      |       |       |       |       |       |       |       |       |       |
|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| people in            | real world | 634 | 81 | 194 | 5 | 1 | 0.575 | 0.678 | 0.168 |
| circuit voltages     | real world | 92  | 15 | 25 | 11 | 6 | 0.797 | 0.257 | 0.009 |
| uno deck             | real world | 6295 | 899 | 1798 | 0 | 15 | 0.993 | 0.994 | 0.013 |
| grass weeds          | real world | 1661 | 245 | 580 | 105 | 1 | 0.781 | 0.781 | 0.106 |
| gauge u21wv          | real world | 158  | 25 | 52 | 141 | 2 | 0.642 | 0.668 | 0.217 |
| sign language        | real world | 504  | 72 | 144 | 0 | 26 | 0.870 | 0.255 | 0.006 |
| Valentines chocolate | real world | 68  | 6 | 13 | 4 | 22 | 0.110 | 0.059 | 0.013 |
| fish market          | real world | 14180 | 1202 | 3116 | 252 | 21 | 0.920 | 0.988 | 0.013 |
| lettuce pallets      | real world | 1060 | 151 | 299 | 168 | 5 | 0.945 | 0.966 | 0.031 |
| shark teeth          | real world | 191  | 36 | 53 | 154 | 4 | 0.948 | 0.863 | 0.025 |
| bees jt5in           | real world | 5640 | 836 | 1604 | 163 | 1 | 0.891 | 0.680 | 0.009 |
| sedimentary features | real world | 156  | 21 | 45 | 31 | 5 | 0.327 | 0.244 | 0.000 |
| currency v4f8j       | real world | 576  | 82 | 155 | 1 | 10 | 0.583 | 0.514 | 0.099 |
| trail camera         | real world | 941  | 131 | 239 | 4 | 2 | 0.966 | 0.969 | 0.512 |
| cell towers          | real world | 705  | 101 | 202 | 25 | 2 | 0.939 | 0.942 | 0.053 |

Table 4: The above table reports metadata about each dataset in RF100. It includes each data-set’s name, number of classes, labeling hours spend by the original author, the train/validation/test split used, model’s mAP@.50 value on the three benchmarked models and the source link. In the labeling hours column, a zero value denotes that the dataset was annotated outside of the Roboflow app, and a n/a value denotes that the dataset was created before Roboflow started keeping track of the labeling hours data.