InVar-100: Industrial Objects in Varied Contexts Dataset

Jan Lehr Vivek Chavan* Paul Koch Marian Schlüter Clemens Briese

Fraunhofer IPK, Berlin, Germany

1 Introduction

This datasheet introduces the InVar-100 dataset. Please read it carefully before using the dataset for your application or research work. The dataset is being shared under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This means that you are free to use, adapt, and distribute the dataset for any purpose, even commercially, as long as you provide appropriate credit to us as the creators of the dataset. This datasheet is a supplement to our ICCV 2023 paper: **Towards Realistic Evaluation of Industrial Continual Learning Scenarios with an Emphasis on Energy Consumption and Computational Footprint** [2]. Please always cite the original work from the conference proceedings. A preprint will be available on https://arxiv.org and will be linked to the GitHub repository.

The following is a summary of the files aggregated together in our dataset:

- **Image Data:** The image data folder contains 100 classes. Each class contains four subclasses/subcategories.
- **Metadata:** The metadata file includes various properties of the objects digitised in the InVar dataset. This includes the weight, length, breadth, height, superclass, material, shape, colour and additional properties. These tags and descriptors allow for further general research work, including modality fusion.
- **Datasheet:** This document describes the particulars of the dataset and provides more context on the objects, subcategories and related statistics.

2 Industrial Data Issues

Figure 1 shows some images from a digitalisation station at the sorting facility during research with the EIBA project[1]. The images were taken by operators under strict time constraints while working on large batches of components. This can result in poor image quality and has further knock-on effects for Machine Learning applications. The InVar-100 dataset simulates such environments and highlights the issues in industrial setups.

3 Dataset

The Industrial Objects in Varied Contexts (InVar) Dataset was internally produced by our team and contains 100 objects in 20800 total images (208 images per class). The objects consist of common automotive, machine and robotics lab parts. Each class contains 4 sub-categories (52 images each) with different attributes and visual complexities. **White background (D_{wh}):** The object is against a clean white background and the object is clear, centred and in focus. **Stationary Setup (D_{st}):** These images are also taken against a clean background using a stationary camera setup, with uncentered objects at a constant distance. The images have lower DPI resolution with occasional cropping. **Handheld (D_{ha}):** These images are taken with the user holding the objects, with occasional occluding. **Cluttered background (D_{cl}):** These images are taken with the object placed along with other objects from the lab in the background and no occlusion. The dataset was produced by our staff at different workstations and labs in Berlin. Human subjects, when present in the images, (e.g. holding the object) remain anonymised. More details regarding the objects used for digitisation are available in the metadata file.

There are other larger datasets on industrial objects, such as the ABC dataset [6], MECCANO [7] and the MCB project [5]. While other datasets contain a higher number of classes and images, the four subcategories in our dataset

*Correspondence: vivek.chavan@ipk.fraunhofer.de
InVar-100 Dataset: Fraunhofer Fordatis
DOI: 10.24406/fordatis/266.2
GitHub Repo: https://github.com/Vivek9Chavan/RECIL

[1]The work (EIBA project 033R226) is funded by the German Federal Ministry of Education and Research (BMBF) in the ReziProK program over the FONA platform for sustainable research.

Figure 1: Sample industrial images taken in by operators at the EIBA project location depicting miscellaneous issues such as clutter, dirt, cropping, occlusion, and blur. Our dataset and the visual contexts address such issues with industrial data collection.

simulate the different visual contexts in which industrial objects are generally digitised during inference time. The context of the images changes, but the underlying features of the target object remain constant, making it ideal for our investigation. Datasets such as NICO [4] and NICO++ [8] also present object classes in different visual contexts. However, the industrial objects in our dataset are unlikely to be present in general large pretraining datasets such as ImageNet [3]. The dataset can, thus, serve as a useful downstream dataset for research investigations. Figure 2 shows sample images for the four subcategories.

Attribute	D_{wh}	D_{st}	D_{ha}	D_{cl}
Object is centered	✓	✓*	✗	✗
Object in focus	✓	✓	✗	✗
High Resolution	✓	✗	✓	✓
Cropping	✓*	✓*	✗	✗
Occlusion	✗	✗	✓*	✗
Clutter	✗	✗	✓*	✓
Blur	✗	✗	✓*	✓*

Table 2: Details on the InVar-100 dataset(* means only a fraction of images have the attribute). Taken from [2]

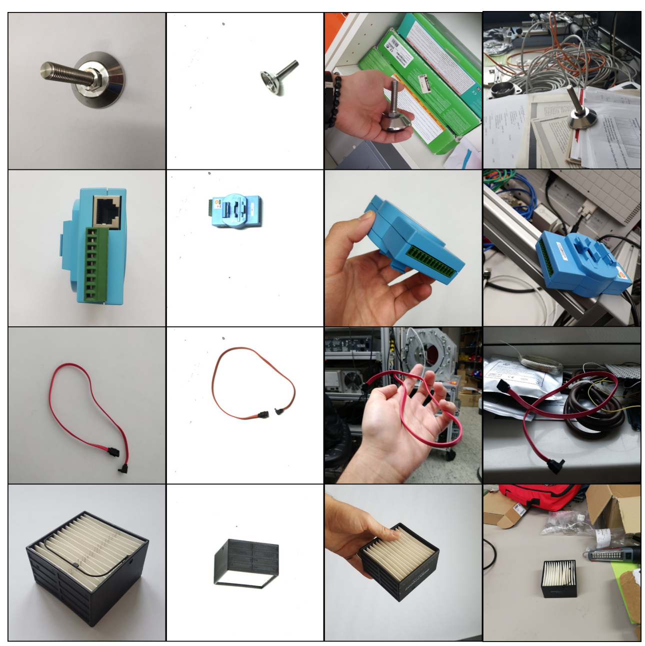
.

4 Details

Table 1 shows the performance of each subcategory on the other subcategories as the validation data. While the objects being recognised remain the same, the differing contexts make it extremely challenging for the model to identify the object accurately. Figure 4 shows the 100 objects from the InVar dataset clustered in 2D. Figure 3 gives histogram plots for the objects in the dataset based on the superclass, the weight, and the length. The attached file contains more properties for the dataset that can be used for further research.

Training Data	Validation Data			
	D_{wh}	D_{st}	D_{ha}	D_{cl}
D_{wh}	98.6%	3.1%	3.4%	3.6%
D_{st}	4.1%	93.1%	1.5%	1.4%
D_{ha}	31.7%	2.0%	89.4%	13.5%
D_{cl}	35.2%	1.2%	14.6%	88.1%

Table 1: Joint learning accuracy matrix for the subsets of the InVar-100 dataset. The results show that it is necessary to introduce the context while training for accurate recognition in a given context.



(a) White (b) Stationary (c) Handheld (d) Cluttered

Figure 2: Example of images from the InVar-100 dataset with the subcategories. Taken from [2].

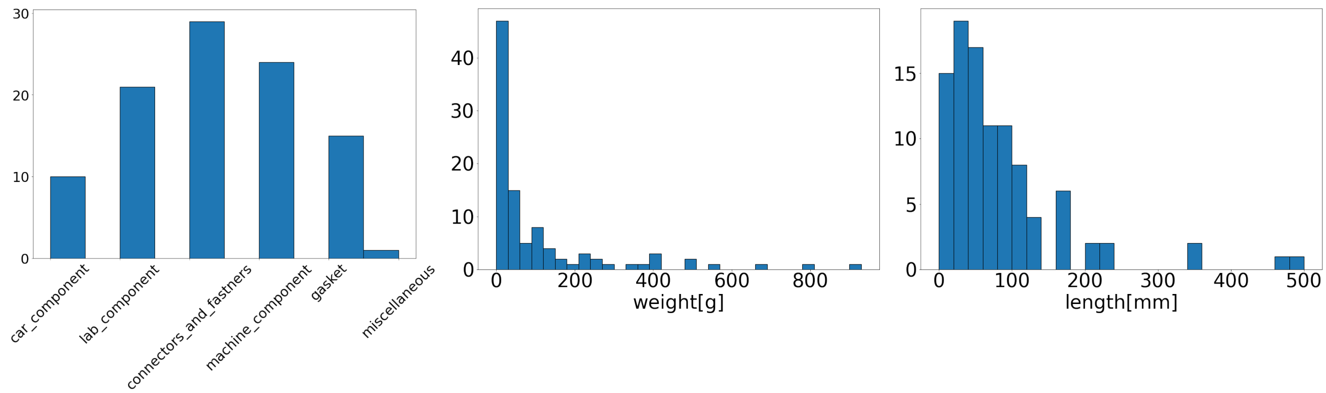


Figure 3: **Left:** A histogram of the Superclasses for the InVar-100 Dataset. **Middle:** Weight distribution between the objects. **Right:** Lengths of the objects.

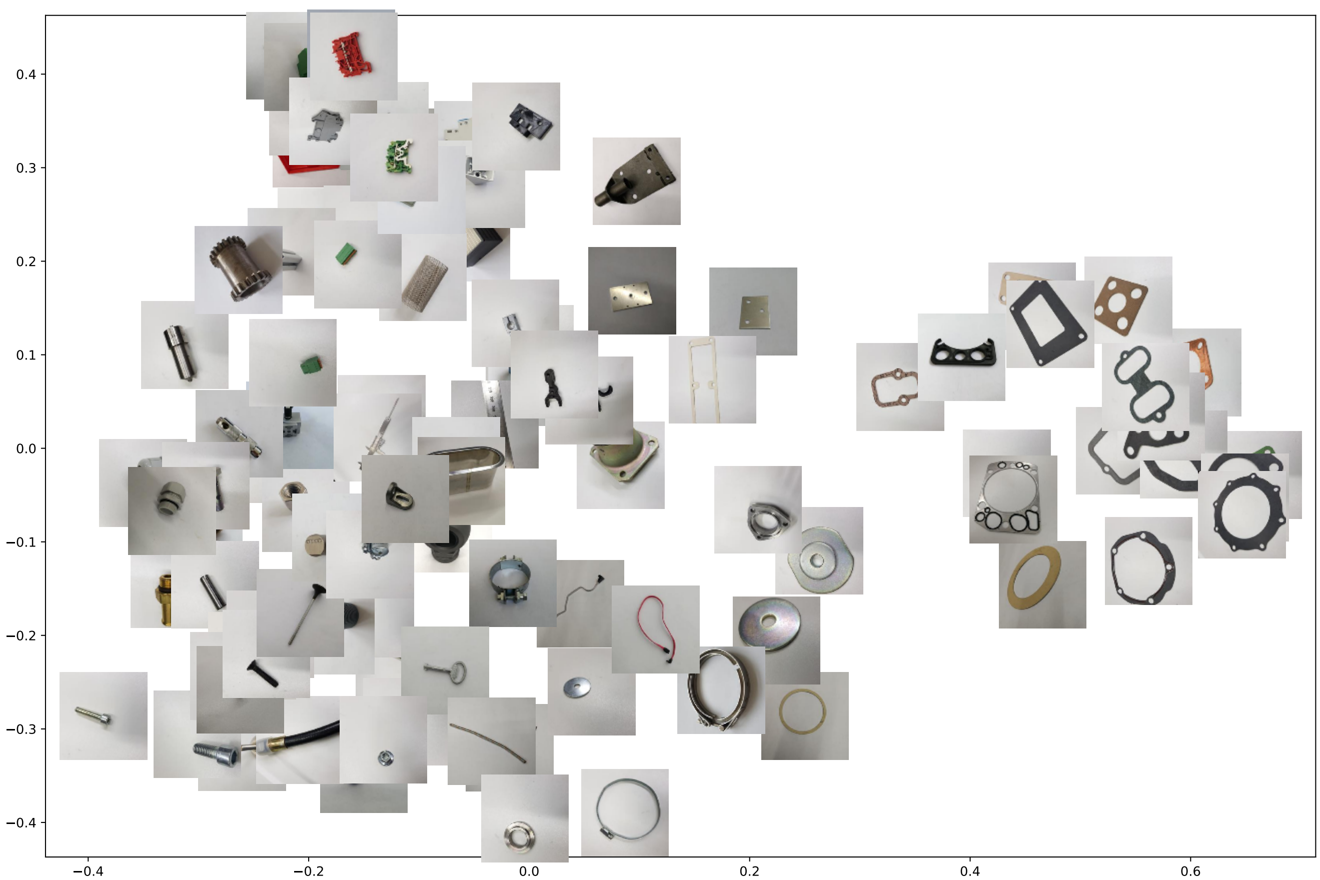


Figure 4: The 100 objects in the InVar dataset, clustered using the embeddings from a Vision Transformer (ViT-S) pretrained using Dino [1]. Please note that many similar components are placed closely and cannot be seen clearly in the figure.

References

[1] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. 3

[2] Vivek Chavan, Paul Koch, Marian Schlüter, and Clemens Briese. Towards realistic evaluation of industrial continual learning scenarios with an emphasis on energy consumption and computational footprint. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2023. 1, 2

[3] 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, 2009. 2

[4] Yue He, Zheyan Shen, and Peng Cui. Towards non-iid image classification: A dataset and baselines. *Pattern Recognition*, 110:107383, 2021. 2

[5] Sangpil Kim, Hyung-gun Chi, Xiao Hu, Qixing Huang, and Karthik Ramani. A large-scale annotated mechanical components benchmark for classification and retrieval tasks with deep neural networks. In *Proceedings of 16th European Conference on Computer Vision (ECCV)*, 2020. 1

[6] Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. Abc: A big cad model dataset for geometric deep learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1

[7] Francesco Ragusa, Antonino Furnari, Salvatore Livatino, and Giovanni Maria Farinella. The meccano dataset: Understanding human-object interactions from egocentric videos in an industrial-like domain. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1569–1578, January 2021. 1

[8] Xingxuan Zhang, Yue He, Renzhe Xu, Han Yu, Zheyan Shen, and Peng Cui. Nico++: Towards better benchmarking for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16036–16047, June 2023. 2