HILOD21 - A High Level-Of-Detail Image-Dataset for Object Detection

Stefan Wagenpfeil[1], Paul Mc Kevitt[2], and Matthias Hemmje[1]

[1] University of Hagen, Faculty of Mathematis and Computer Science, Germany
{firstname.lastname}@fernuni-hagen.de
[2] FTK e.V. Research Institute for Telecommunications and Cooperation, Germany
p.mckevitt@ftk.de

Abstract. This paper introduces a novel large dataset for experiments on high resolution image processing called *HILOD21*. This dataset contains 800 images, automated and manual annotations, and manual descriptions for every image and is available for public academic use. Various object detection algorithms have been implemented and performed on this dataset during the *Object Detection Challenge 2021* at the University of Hagen to validate and optimize the annotations of the dataset. Some results of this evaluation are also outlined in this paper and show, that the *HILOD21* dataset can be applied as data source for object detection algorithms in different programming languages running on various devices.

Keywords: high-resolution image dataset · object detection · level of detail · multimedia feature detection · pattern recognition

1 Introduction and State of the art

Every year, more than 1.2 trillion photos are taken on Smartphones and digital cameras [11]. The resolution of these images increased during the last years up to 100 megapixel - even on Smartphones [17]. These high resolution images also provide a higher Level-Of-Detail (LOD), which has to be utilized in object and/or feature detection algorithms. For the evaluation of object and/or feature detection algorithms, a validated and reliable reference dataset is required. With such a dataset, machine learning models can be trained [5], and an automated verification of algorithms can be performed [13]. For image processing, the Flickr30k set [18], the DIV2K dataset [3], the IAPTRC12 dataset [9], or the PASCAL VOC dataset [7] are some of the most relevant collections. However, these datasets either contain only low-resolution images with manual annotations, or high-resolution images without annotations. As shown in our previous work [15][16], for the automated processing and evaluation of high-LOD images, these datasets need to be optimized. Therefore, we decided to construct a high-resolution dataset with automated and manual annotations, which will be described in the next section.

2 Modeling and Design

As basis for our dataset, we selected 800 random images from the DIV2K dataset [3], which is available for public academic use. The images have a resolution of 2040x1356 pixels (in average) and typically contain various objects of various topic areas with a high LOD. During the *Object Detection Challenge 2021 (ODC21)* at the University of Hagen, these images have been manually annotated with keywords and textually described. The resulting dataset is named *High-Level-Of-Detail Dataset 2021 (HILOD21)* and also available for public academic use. The *HILOD21* has been designed for a broad variety of feature detection algorithms. Examples of the dataset are shown in Figure 1. Each of the images in the dataset has been manually described, validated, reviewed and approved. The descriptions are available as XML and CSV-file. In addition, during the *ODC21*, these manual annotations have been refined by automated algorithms.

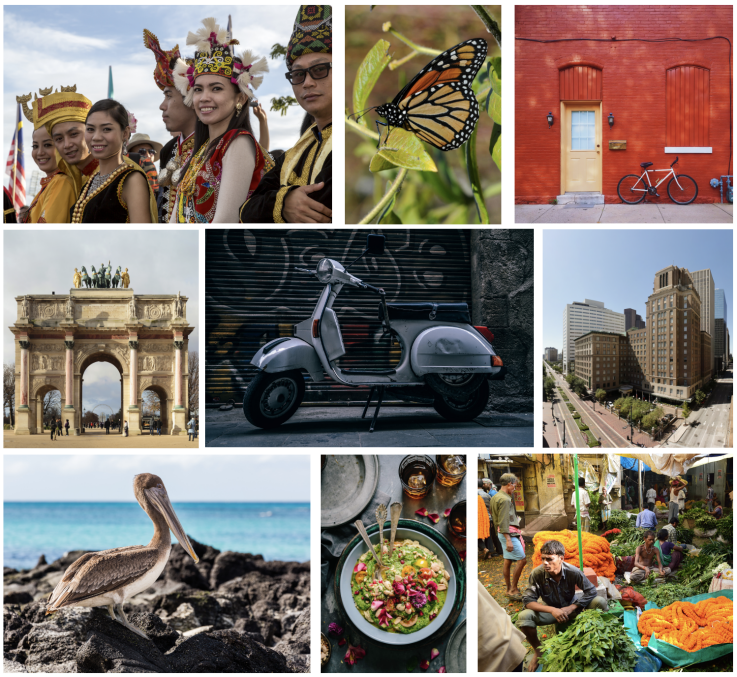


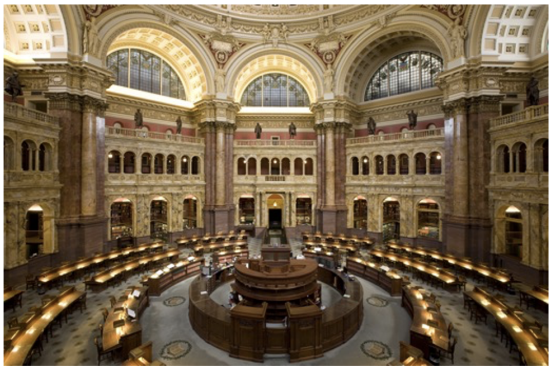
Fig. 1. Example images of the HILOD21 dataset

In the *HILOD21*, each image will be described and annotated by the following attributes:

– image-name: identifies the corresponding .png image of the dataset. Image names are UUIDs.

– keywords: manually annotated keywords containing general aspects of the image. These keywords are also used to describe overall settings, like e.g. "football", "landscape", "black & white". Keywords are also used for the description of landmarks, like e.g. "fontana di trevi".
– primary objects: a list of the most prominent objects or features of the image.
– other objects: a list of all other objects, that are shown on the image.
– description: a textual description of the image.
– automated keywords: a list of keywords, that have been detected automatically during the *ODC21*

An example of a complete record is given in Figure 2.



Attribute	Value
Name	ab307b8f-662b-4429-9123-804a73e1038a.png
Keywords	library, building, congress of the united states
Primary Objects	library
Other Objects	building, window
Description	library of congress of the united states of america. The image is taken inside the building and shows seats for reading in a large hall.
Automated detection	fixture, interior design, wall, symmetry, facade, real estate, glass, landmark, city

Fig. 2. HILOD21 annotation attributes and values.

Table 1 lists the most relevant object classes of the *HILOD21*. It contains only objects, that are detected or described in at least 40 images of the dataset. A list of all available keywords is part of the dataset download. All detected object term have been validated against a english dictionary.

During the *ODC21* at the University of Hagen, a set of 25 algorithms (i.e. plugins) has been implemented and tested with the dataset. The various plugins

#	object class	#	object class	#	object class
218	animal	86	window	55	whiskers
211	person	77	hat	52	gesture
208	plant	77	facade	51	cloud
171	cloud	70	art	48	human body
150	face	67	eye	46	vehicle
122	organism	66	tree	44	food
122	building	65	grass	44	clothing
111	wood	62	wheel	42	luggage
92	sleeve	59	snout	41	city
92	water	58	fawn	40	bird

Table 1. Selection of the most prominent object classes and the corresponding number of images in the dataset.

address different object classes and thus provide a good coverage of the various aspects of the *HILOD21*. These implementations were both used to optimize the quality of the dataset and to evaluate the algorithms themselves. Some of these results are shown in section 3 of this paper.

3 Implementation and Evaluation

For the evaluation of the dataset and the selected algorithms, a Precision & Recall experiment has been performed for each plugin, where Precision P is the ratio of correctly positive observations to the total predicted positive observations, Recall R is the ratio of correctly predicted positive observations to all possible observations. As we employed various plugins with special focuses (e.g. the Carnet.AI plugin [6], which is trained to detect cars and car models), the results of these experiments have been manually compared with the annotations in the dataset to refine the descriptions of the images. It should be mentioned, that we also made a reference run with Google Vision [8], which showed very good results in our evaluation and can be regarded as a well established reference implementation for object and feature detection. Table 2 shows an excerpt of the plugins, we have employed to prove and enhance the quality of the dataset annotation. It also illustrates some of the most relevant object types (i.e. classes), that have been processed, verified, and enhanced with this plugin and their number of occurrences in the dataset.

To illustrate, that the images of *HILOD21* can be also employed on devices like the Nvidia Jetson Nano, Raspberry PI, or Smartphones, we also performed experiments on such hardware or limited memory of CPU configurations. This experiment showed, that object / feature detection plugins vary according to the relevant object classes, especially on devices with limited memory. This is due to the training set of machine learning components, and also due to the algorithms for pattern recognition, that are implemented by these APIs. More detailed information about the specific implementations of the plugins and their

refinement plugin	**example class**	**# images**
Tensorflow.js[2]	fruit	64
	plant	208
	apple	23
	orange	55
YOLO [12]	dog	23
	animal	218
	cat	57
	fish	38
Image AI [1]	person	211
	tree	198
	building	161
	car	180
Amazon AWS [4]	mountain	69
	cloud	83
	water	171
	snow	35
Microsoft Visioning [10]	boat	42
	train	14
	building	122
	music	9
Google Vision [8]	woman	34
	man	79
	smile	20
	face	150
Carnet.AI[6]	wheel	62
	train	14
	truck	4
	bicycle	16

Table 2. Plugins applied to additionally validate and refine the dataset.

underlying algorithms can be found in the reference section of this paper. However, this experiment shows, that state-of-the-art APIs, algorithms, and tools can be evaluated on basis of the *HILOD21* independent from programming language (e.g. Java, C#, or Nvidia CUDA) or hardware (e.g. Computers, Smartphones, or Devices).

To prove, that the *HLOD21* dataset provides a significant improvement of LOD, we performed an additional experiment, in which the LOD has been recursively applied to an image. This means, that for each detected object, the bounding box of the object has been calculated and an additional image only with the bounding box has been generated and re-processed. Figure 3 shows the original input image, Table 3 the resulting objects according to the LOD.

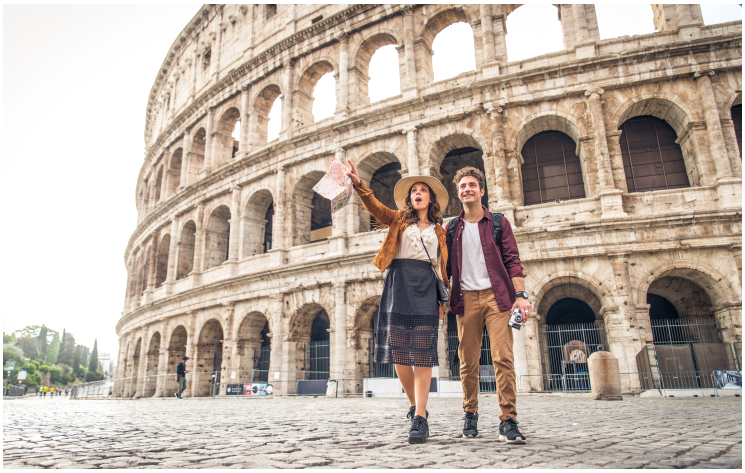


Fig. 3. Level-Of-Detail and feature detection.

4 Discussion and Summary

In this paper, we presented the details of the *HILOD21*, which is made available for further public research. The dataset contains 800 high-resolution images and is annotated both manually and automatically. The dataset and the contained objects / features have been evaluated during the *ODC21* and results of this have been shown in the previous section. The results of this evaluation show, that the dataset provides data for high-LOD object and features detection, which can be used to train, optimize, evaluate object detection algorithms. Thus, *HILOD21* will be an important component in multimedia information retrieval research.

The *HILOD21* dataset is available for academic research purposes only and can be downloaded at [14] (3.54 GB zipped file).

References

1. Image ai - state-of-the-art recognition and detection ai with few lines of code. (08 2021), http://www.imageai.org
2. Tensorflow (08 2021), https://en.wikipedia.org/wiki/TensorFlow
3. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017), http://www.vision.ee.ethz.ch/ timofter/publications/Agustsson-CVPRW-2017.pdf
4. Amazon.com: Amazon webservices. Tech. rep., Amazon Inc., https://aws.amazon.com/de/api-gateway/ (09 2020)
5. Beyerer, J., Richter, M., Nagel, M.: Pattern Recognition - Introduction, Features, Classifiers and Principles. Walter de Gruyter GmbH & Co KG, Berlin (2017)

LOD	# n	# e	**selected exemplary feature terms**
0	53	204	Person, Travel, Piazza Venezia, Joy
1	71	274	Pants, Top, Pocket, Street fashion
2	119	475	Camera, Bermuda Shorts, Shoe
3	192	789	Sun Hat, Fashion accessory, Watch, Bergen County (Logo)
4	228	956	Mouth, Lip, Eyebrow, Pocket, Maroon
5	274	1189	Leather, Grey, Single-lens reflex camera

Table 3. Level-Of-Detail in Figure 3 with selected feature terms. n is the number of nodes (i.e. detected objects), e is the number of edges (i.e. relationships between objects).

6. Carnet-AI: Carnet.ai (08 2021), https://carnet.ai
7. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision **88(2)**, 303–338 (2010)
8. Google.com: Google vision ai – derive insights from images. Tech. rep., Google.com, http://cloud.google.com/vision (07 2020)
9. Grubinger, M., Clough, P., Müller, H., Deselaers, T.: The iapr tc12 benchmark: A new evaluation resource for visual information systems. Workshop Ontoimage (10 2006)
10. Microsoft.com: Machine visioning. Tech. rep., Microsoft.com, http://azure.microsoft.com/services/cognitive-services/computer-vision (07 2020)
11. Nudelman, M.: Smartphones cause a photography boom. Tech. rep., Statista / Business Insider, http://www.businessinsider.com/12-trillion-photos-to-be-taken-in-2017-thanks-to-smartphones-chart-2017-8 (09 2020)
12. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. pp. 779–788 (06 2016). https://doi.org/10.1109/CVPR.2016.91
13. Semertzidis, T., Rafailidis, D., Tiakas, E., Strintzis, M., Daras, P.: Multimedia Indexing, Search and Retrieval in Large Databases of Social Networks, pp. 43–63 (10 2013). https://doi.org/10.1007/978-1-4471-4555-4-3
14. Wagenpfeil, S.: Generic multimedia analysis framework (gmaf) (09 2021), http://www.stefan-wagenpfeil.de/public/GMAF/
15. Wagenpfeil, S., Engel, F., Kevitt, P.M., Hemmje, M.: Ai-based semantic multimedia indexing and retrieval for social media on smartphones. Information **12**(1) (2021). https://doi.org/10.3390/info12010043, https://www.mdpi.com/2078-2489/12/1/43
16. Wagenpfeil, S., McKevitt, P., , Hemmje, M.: Graph codes - 2d projections of multimedia feature graphs for fast and effective retrieval (03 2021)
17. Xiaomi: Redmi note 10 pro - the 108mp voyager (03 2021), https://www.mi.com/global/product/redmi-note-10-pro/overview

18. Young, P.: From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions (2016), http://shannon.cs.illinois.edu/DenotationGraph/