Annotated Image Data Version 1

Deliverable D2.2

Version FINAL

Odeuropa

NEGOTIATING OLFACTORY AND SENSORY EXPERIENCES IN CULTURAL HERITAGE PRACTICE AND RESEARCH

European Commission

The Odeuropa project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004469. This document has been produced by the Odeuropa project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.

Grant Agreement No.	101004469
Project Acronym	ODEUROPA
Project full title	Negotiating Olfactory and Sensory Experiences in Cultural Heritage Practice and Research
Funding Scheme	H2020-SC6-TRANSFORMATIONS-2020
Project website	http://odeuropa.eu/
Project Coordinator	Prof. Dr. Inger Leemans KNAW Humanities Cluster Email: inger.leemans@huc.knaw.nl
Document Number	Deliverable D2.2
Status & version	FINAL
Contractual date of delivery	March 2022
Date of delivery	March 2022
Type	Other
Security (distribution level)	Public
Number of pages	20
WP contributing to the deliverable	WP2
WP responsible	WP2
EC Project Officer	Hinano Spreafico
Authors: Mathias Zinnen,[1] Vincent Christlein[1]	
Internal reviewers: Pasquale Lisena[2], Marieke van Erp[3]	
Affiliations: (1) FAU, (2) EURECOM, (3) KNAW	
Keywords: image annotation, benchmark, computer vision	
Abstract: This document explains the creation of the annotated image data version 1 (aidv1) dataset. The dataset is published on Zenodo under the doi *10.5281/zenodo.6367776* (https://doi.org/10.5281/zenodo.6367776). This document describes how annotations and images have been retrieved, reports on the labeling scheme for the raw annotations and on the distribution of collected annotations. Furthermore, mappings to two different datasets, most notably to aidv1, are presented. Class distributions of these derived datasets are documented, and mapping decisions are justified.	

Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.1	Feb 2022	First draft	Mathias Zinnen, Vincent Christlein	all
0.2	Feb 2022	Check and draft approval by project manager	Marieke van Erp	-
0.3	Mar 2022	Revision	Mathias Zinnen	-
0.4	Mar 2022	Revision	Vincent Christlein	-
0.5	Mar 2022	Revision	Pasquale Lisena, Mathias Zinnen	-
1.0	March 2022	Final check and approval by project manager	Marieke van Erp	-

Executive Summary

This document introduces the annotated image data version 1 (aidv1) dataset, a dataset of artworks annotated with olfactory references forming the basis for training systems to automatically detect olfactory references. The dataset is published on Zenodo under the doi 10.5281/zenodo.6367776. This document reports on how images are collected from various digital museum collections and how annotations are created for them. Furthermore, mappings to multiple derived datasets are explained, most notably to the aidv1 dataset, which is also being used as training dataset for the ODeuropa Competition on Olfactory Object Recognition (ODOR) challenge (Deliverable D2.5, due in M24).

Contents

1 Introduction

How can past smells and their historical conceptualizations be accessed when the smells itself have long evaporated? Apart from studying the texts that have been written about smells, we can analyze historical artworks to look for traces of past smells. Visual olfactory references can come in various ways: Smells might be alluded by the use of specific metaphors, smell sources can be depicted as well as smell related narratives or spaces. Building upon the taxonomy of smell references that was developed in D2.1 (Taxonomy of Olfactory Phenomena in Images), we assemble a dataset of visual smell references by annotating a selection of historical artworks with categories of olfactory phenomena and their positions in the images.

The resulting annotations do not only enrich the European Olfactory Knowledge Graph (EOKG)[1] but also serve as a vital ground truth for machine learning models that enable the automatic extraction of smell references (D2.3, Object Detection/Image Analysis). Subsequently, making use of our models, we will extend the annotations with support by the machine, to create a second, enlarged version of the dataset (D2.4, Annotated Image Data Version 2).

This document describes the process of how the images and annotations are obtained, from which self-contained datasets can be derived. We outline how this "raw" annotation data is mapped to two different self-contained datasets, one initial experiment dataset that has been used internally, and most notably the aidv1 dataset (D2.2, Annotated Image Data Version 1). The aidv1 dataset will furthermore be used as a challenge dataset in the ICPR 2022[2] ODOR challenge[3] that we organize starting in March 2022 (D2.5, Computer Vision Challenge).

By combining the release of the aidv1 dataset with the organization of the computer vision challenge, we hope to increase the visibility of the dataset publication and maximize the impact in both computer vision and digital humanities research.

2 Image Annotation

The following section describes the collection of image data and annotation generation that have been undertaken to create a dataset of visual olfactory phenomena in historical artworks.

2.1 Image Material

As a prerequisite to the annotation of olfactory phenomena in historical artworks, the underlying image data has to be collected and downloaded. We therefore queried multiple digitized museum collections using a list of search terms that allegedly lead to images with olfactory relevance. Table 1a lists the collections that have been used for data collection, next to the number of images downloaded from each collection.

Our strategy in defining the search terms is two-fold: In a first step, we defined an initial list of search terms that reflect our expectations at the start of the project: in which contexts did we expect to find meaningful smell-references? This first, relatively open image query lead to a collection of 30,134 historical artworks. To incorporate our developing knowledge about actual smell representations, we design the further data processing as an iterative process, that is still in progress. There are continuous discussions and exchange with the other Work Packages (WPs) that include regular meetings and collaborations on drafts. Olfactory and art historians were also involved in defining the annotation format (D2.1). This collaboration between WPs on the one hand, and the ongoing analysis and annotation of the existing data on the other hand, leads to a steadily evolving understanding of visual smell representations that exceeds our state of knowledge at the start of the project.

[1] cf. D4.1 (European Olfactory Knowledge Graph)
[2] https://www.icpr2022.com/
[3] https://odor-challenge.odeuropa.eu

collection	# images
RKD	27 104
Fondazione Federico Zeri	11,223
Bildindex der Kunst und Architektur	4,776
Catalogo Beniculturali	2,380
Artmuseum Princeton	1,925
University of Notre Dame	1,202
Web Gallery of Art	952
SLUB Dresden	683
Museum Boijmans	597
Ashmolean Museum Oxford	553
National Gallery of Art	539
Plateforme ouverte du patrimoine (POP)	321
arkuBiD Bonn	262
Fondazione Giorgi Cini	224
Städelmuseum Frankfurt	16
Réunion des musées nationaux	3
Total	52,760

(a)

search term	# images
Smell[a]	618
Senses[b]	2217
Lazarus[c]	4215
Still Life[d]	21074
Gloves	901
Donkey[e]	2,483
Goat	5,177
Cheese	365
Pomander	146
Tobacco	1,922
Whale[f]	229
Censer[g]	195
Total	41,552[h]

(b)

Table 1: (a) List of digitized collections, sorted by the number of images that have been collected from the respective collection. (b) Overview of initial and later search terms with number of images collected for each. Initial search terms are marked in green, those who were added later in pink. Search term variations: [a]Geruch, odore, geur, odeur; [b]sens, sensi, Sinne, zintuig; [c]Lazare, Lazarro; [d]natura morta, natura morte, stillleben, stilleben, stilleven; [e]Ezel; [f]Walvis.
[g]A censer is an incense burner that was used to burn incense or perfume in solid form. [h]Additionally, 11,319 images have been collected in the initial data gathering phase without saving the query term, hence the difference to the total number of images in (a).

We therefore repeatedly queried for specific keywords that turned out to be relevant throughout the project work and extended our initial collection of artworks with 11,418 images (up to now). Table 1b gives an overview of initial and later search terms and the number of images that have been collected for each.

2.2 Annotation Categories

Another prerequisite for the creation of annotated image data of olfactory phenomena is the design of an annotation scheme that can be used to label the collected images. In collaboration with the other WPs, we created multilingual lists of olfactory vocabularies in which we collected all kinds of olfactory phenomena that might be found in texts or images. These olfactory vocabulary lists are subdivided into the four broader categories of olfactory objects, olfactory iconography, fragrant spaces, and olfactory gestures [Lisena et al., 2022]. Based on the vocabularies, we derived a high level taxonomy of visual olfactory phenomena that was presented in Deliverable D2.1 *Taxonomy of Olfactory Phenomena in Images*.

Iconographies and spaces are related to the image as a whole, while objects and gestures can be localized on the canvas of an artwork, as described in more detail in the detection techniques chapter in Deliverable D2.1. The aim of our annotation campaign is to mark occurrences of olfactory objects and gestures with manually drawn bounding boxes, whereas the other two categories are initially disregarded. We assume that they can later either be recognized using image metadata, or indirectly, by analyzing spatial relations and co-occurrences of olfactory objects present in an artwork.

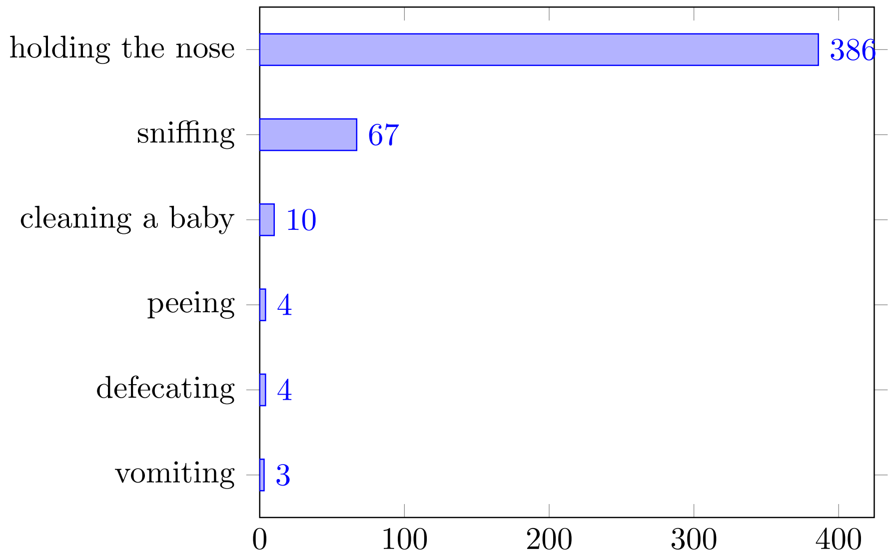

Figure 1: Annotated smell gestures and their respective number of occurrences.

Annotations on the artworks were created manually by the two student assistants, and partly by the other team members of WP2, using an image annotation tool that is described in Subsection 2.3. For gestures, we annotated 474 samples in six categories of smell gestures, with the vast majority of the annotations coming from the two categories *sniffing* and *holding the nose*. Figure 1 gives an overview of annotated gestures.

With 22,381 samples in 213 categories, olfactory objects constitute the vast majority of annotations collected. This imbalance is grounded in the data, as there are simply more objects than depictions of smell gestures in the artworks we collected. Explicitly querying for specific gestures might lead to an increased number of samples for smell gestures. However, since initial experiments with a direct recognition of smell gestures did not show promising results, we decided to focus on objects first and recognize gestures in a second step, presumably indirectly by analyzing the spatial relationship between objects, faces, and body parts.

The high number of object categories, including objects that are very rare and particular, suggests the usage of a hierarchical structure of classes, which has multiple advantages: (1) It makes it easier to find specific object categories, simplifying the annotation process. (2) Detection systems can resolve to a fallback solution in cases where the exact object category cannot be determined but a broader classification can be made (e. g., detecting a flower instead of flower species). D2.1 describes the rationale between this hierarchization in detail and presents a WordNet-based [Miller, 1995] strategy. In contrast to the initial proposition, we decided to simplify the WordNet concept hierarchy and incorporate only two levels of abstraction, since a more complex hierarchy remains mostly unused and complicates annotation and detection architectures without adding much extra value. From the leave nodes, the full WordNet hierarchy can however still be created in case this will become necessary at a later stage. Since WordNet is also used by WP3 and WP4 the different data sources can easily be linked and integrated.

The selection of the supercategories from the WordNet concepts is based on pragmatic considerations such as visual similarity, assumed familiarity of concepts, and simplicity.

Table 2 lists the supercategories that have been used in the annotation scheme and how many subcategories have been defined for each as well as the number of samples in each supercategory.

Figures 2a and 2b show the exemplary subcategory distributions of the *mammal* and *seafood* categories, respectively. The full distribution of samples on both hierarchy levels is listed in appendix B.

supercategory	# subcategories	# samples
flower	20	8,484
fruit	28	5,196
mammal	38	2,126
bird	13	1,185
vegetable	26	1,088
smoking equipment	16	958
insect	17	708
beverage	5	553
jewellery	11	433
seafood	10	321
reptile/amphibia	3	105
nut	3	78
other	14	1,094

Table 2: Supercategories of the annotation scheme. The middle column gives the number of subcategories that have been defined for each of the supercategories. The right column reports the number of samples that have been annotated for the supercategory including its subtypes. *Other* subsumes all top-level categories that do not have further subcategories.

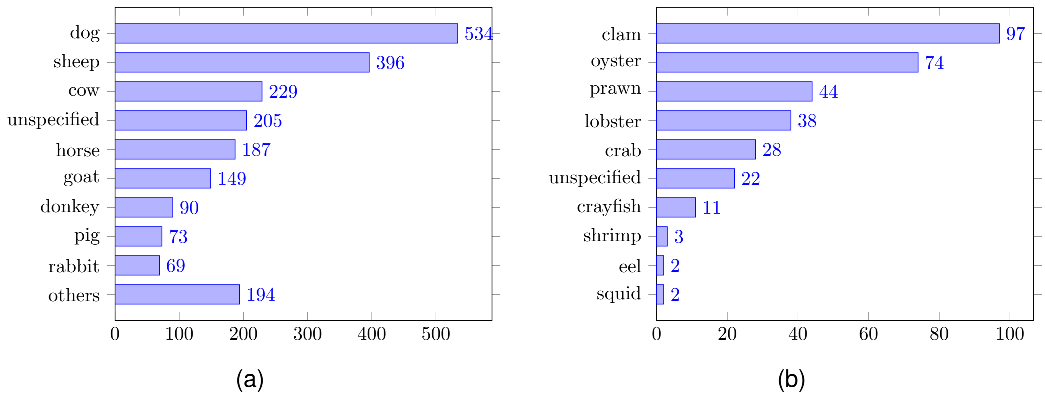

Figure 2: Distribution of subcategory annotations of (a) mammals and (b) seafood supercategories.

2.3 Annotation Tool

There are many different tools available for the annotation of images with bounding boxes, e. g., labelme,[4] image-annotator,[5] or labelImg[6]. Prior to our annotation campaign, we conducted an investigation about which tool suits our needs the best and decided to use the Computer Vision Annotation Tool (CVAT)[7]. Our reasons for selecting CVAT over the alternatives include its capability to easily set-up a self-hosted instance, its support of many different input and output data formats, and the possible integration into *fiftyone*,[8] an open source tool for the management of computer vision datasets. The tool is well documented by its developers.[9] Additionally, we created an annotation guidelines document[10] to help annotators with issues that are specific to our use-case. While the tool is used mostly by our two student assistants to create the annotations, some

[4] https://github.com/wkentaro/labelme
[5] https://github.com/Abbe98/image-annotator
[6] https://github.com/tzutalin/labelImg
[7] https://github.com/openvinotoolkit/cvat
[8] https://voxel51.com/docs/fiftyone/
[9] https://openvinotoolkit.github.io/cvat/docs/
[10] https://github.com/Odeuropa/wp2-annotations/blob/master/annotation-guidelines/annotation-guidelines.md

Metadata field	Description
File Name	Unique file name of the image file
Artist	Artist of the artwork
Title	Artwork title
Query	Query term used for retrieval
Earliest Date	Earliest assumed date of artwork creation
Latest Date	Latest assumed date of artwork creation
Genre	Artwork genre
Current Location	Location of the artwork by the time of retrieval
Repository Number	Repository number of the source collection
Photo Archive	URL of the source collection
Image Credits	URL of Image Credits
Details URL	Download URL for the image

Table 3: Explanation of fields in the metadata csv file.

annotations are also created by other team members (e. g., in cases where many annotations needed to be created quickly, or where art historians' expertise is needed.).

2.4 Data Format

The annotations are provided in *JavaScript Object Notation (JSON)*, and more specifically in *COCO JSON*, the format defined for the Common Objects In Context (COCO) challenge [Lin et al., 2014], which is the de-facto standard for object detection annotations.[11]

Moreover, we release a comma separated values (csv) file containing image metadata such as *artist*, *title*, and *genre*, as well as links to download the images from their respective original collections. Table 3 gives a listing of metadata attributes present in the csv file. The mapping between the annotations and the image metadata can be established via unique filenames that are present in the *images* array of the annotation *JSON* file, as well as in the metadata csv file. Unfortunately, we cannot provide the images as such since many of the artworks in our datasets are licensed in a way that does not allow their redistribution. Instead we resort to including the download links in the metadata csv file, and provide a script for dataset users to download the images on their own.

3 Datasets

The annotations collected throughout the project form a solid corpus of visual knowledge that can directly be fed into the EOKG. In principle, the full set of annotations, together with the associated images, could also be used as a dataset to train an object detection model. With 24,560 annotations on 3,012 images in 223 categories, the raw annotations provide a large basis for training detection systems. In practice however, directly applying the full annotation set has multiple drawbacks:

(1) The accuracy of classification decreases drastically as the number of classes increases [Liu et al., 2020]. The large number of different categories that are present in our annotation set thus makes the implementation of a detection system with decent results very difficult. For initial experiments particularly, it is therefore sensible to decrease the number of classes by using only a subset of the annotation set;

[11] A definition of the data format can be accessed at https://cocodataset.org/#format-data

(2) To reliably detect a class, classic object detection algorithms need to be presented with a minimum number of samples.[12] The full set of annotations has a long-tailed distribution of classes, with many of the categories having very few samples. Since object detection accuracy is usually measured by averaging the detection accuracy over all categories, these rare classes can drastically impair the overall detection performance;

(3) Additionally, an imbalance in the class distribution of object detection datasets can harm the detection performance [Buda et al., 2018]. The long-tailed distribution of the full annotation set leads to a considerable imbalance in the distribution of annotations that can be mitigated by considering only a subset of classes.

We generated two distinct datasets from the full set of annotations that will be described in the following subsections.

3.1 Initial Experiments Dataset

The first dataset we created was used to conduct initial experiments and set up an object detection pipeline. It consists of 1,126 images with 10,818 annotations in total, from which 100 images were split for a test set, resulting in a training set of 1,026 images with 9,519 annotations, and a test set of 100 images with 1,299 annotations.

We selected 29 categories from the 213 categories of the full annotation set and included all available samples, as well as the associated images. The main rationale behind reducing the number of categories was to simplify the detection task since first results with the full category set were not promising. The categories and the respective number of samples are listed in Table 4.

Like in the case of the full set of annotations, the initial experiments dataset exhibits a large class imbalance which has to be taken into account when using it to train object detection models, e. g., by oversampling underrepresented classes [Buda et al., 2018].

Since the dataset was not meant to already extract information for the EOKG, but to be used for initial computer vision experiments only, the category selection was mainly guided by technical considerations rather than their olfactory relevance. One challenge in the detection of olfactory objects in historical artworks is that object detection systems are usually designed and trained to be working on photographic data. Huge labeled datasets of photographs such as ImageNet [Russakovsky et al., 2015], COCO [Lin et al., 2014], and OpenImages (OI) [Kuznetsova et al., 2020] contain millions of annotated images and enable to train detection models that have an impressive performance on photographic data using the categories that are used in these datasets. Since there are no datasets of comparable size available for historical artworks, we designed the initial experiments dataset in a way that enables us to leverage the high quality photographic datasets for our purpose. The categories of our initial experiments dataset where thus restricted to classes that (1) are present in the OI dataset, and (2) have samples in our full set of annotations.

We conducted experiments with varying pretraining schemes, using combinations of the OI, IconArt [Westlake et al., 2016], and PeopleArt [Gonthier et al., 2018] datasets. A scientific publication of the results has been accepted to the *Digital Humanities 2022* conference,[13] held by the Alliance of Digital Humanities Organizations [Zinnen et al., 2022].[14]

3.2 Annotated image data version 1 (aidv1)

The second dataset we generated from the full annotation set constitutes the *aidv1* dataset as defined in the grant agreement. It is publicly and persistently available on Zenodo.[15] In addition, to increase the dissemination of the project and this dataset in particular, the dataset serves as our

[12] Although strategies for the detection of objects with only few samples exists, which are subsumed under the name of few-shot detection, cf. [Huang et al., 2021].

[13] https://dh2022.adho.org/

[14] https://adho.org/

[15] https://doi.org/10.5281/zenodo.6367776

Category	# Samples
Flower	6,479
Bird	883
Grape	601
Insect	584
Peach	431
Dog	358
Fish	331
Pear	191
Apple	137
Oyster	126
Horse	117
Lemon	89
Bread	65
Pomegranate	62
Wine	52
Strawberry	49
Pumpkin	49
Rabbit	47
Sheep	39
Lobster	33
Deer	27
Goat	25
Candle	19
Cheese	13
Pig	9
Whale	2

Table 4: Categories present in the initial experiments dataset, ordered by number of samples.

challenge dataset for the competition (D2.5) we are organizing as part of International Conference on Pattern Recognition (ICPR) 2022,[16] one of the premier conferences in computer vision.

Unifying aidv1 and the challenge datasets has the advantage that we can leverage the accomplishments of the challenge participants in terms of detection approaches to enrich the EOKG and to extend the annotations for version 2 (D2.4) of the annotated image data with machine assistance. In contrast to the initial experiments dataset, we consider olfactory relevance a key requirement for selecting the categories for the challenge dataset. However, since we still need to provide a minimum number of samples for each category to enable deep learning methods, we only include classes where we can provide at least 10 samples (distributed over training, test, and validation set).

In the ICPR2022-ODOR challenge,[17] we ask the participants to implement object detection models capable of recognizing a large range of objects with olfactory relevance on historical artworks. The dataset consists of 2,989 artworks, annotated with 24,391 bounding boxes, from which 797 images will serve as test and validation sets that remain unpublished until after the competition.

The dataset contains 87 classes, of which most correspond to subcategories of the full annotation set. In general, the challenge dataset can thus be considered as a flattened subset of the full annotation set. However, since we do not want to lose the hierarchical information and encourage innovative detection approaches leveraging the object hierarchy, we still provide the original hierarchical structure of the categories. In some supercategories, e. g., for the *bird* category,

[16] https://www.icpr2022.com/
[17] https://odor-challenge.odeuropa.eu

Category	# samples	# subcategories
Flower	5 030	21
Fruit	4 026	13
Vertebrate	3 303	13
Invertebrate	556	8
Vegetable	479	8
Drinking Vessel	644	10
Jewellery	224	4
Other	1 561	11

Table 5: Categories and number of samples as well as subcategories.

we assume that specifying the subcategory does not add any additional olfactory information. In these cases, we directly use the supercategories in the challenge dataset. To increase logical consistency, we subsume the mammal, bird, and fish categories under the newly introduced *vertebrate* supercategory, whereas insects and seafood are being classified as subtypes of the *invertebrate* supercategory. Another change we make for the challenge dataset is to assign all instances of dead mammals and birds to the newly introduced *animal carcass* category because we expect the odor difference between dead and living animals is more significant than that between different species of dead animals. Note that we do not establish this mapping for fish and seafood since we assume that those animals share a specific smell that both living and dead individuals have in common.

Table 5 gives an overview of most frequent (super)categories occurring in the dataset. The hierarchy information is included in the *supercategory* key of the entries of the *categories* array of the COCO JSON document in which the annotations are stored (see listing 1 for an example).

Listing 1: Example entries of the categories-array in the COCO JSON annotations document of aidv1.

```
[
    [...]
    {"id": 20, "name": "violet", "supercategory":"flower"},
    {"id: 21, "name": "other flower", "supercategory": "flower"},
    {"id": 101, "name": "apple", "supercategory": "fruit"}
    [...]
]
```

The full distribution of categories annotated in aidv1, including the hierarchy information is listed in appendix A.

4 Conclusion

Based on the taxonomy of visual olfactory references that we developed earlier and described D2.1, we have collected a corpus of nearly 25,000 annotations on more than 3,000 artworks. To account for class imbalance and rare categories, we do not use these annotations directly to train object detection algorithms. Instead, they serve as a basis for the creation of multiple self-contained datasets.

Section 3.1 describes the creation of the initial experiments dataset, which has a largely decreased number of categories and was used to conduct initial experiments with pre-training schemes for our object detection pipeline.

The creation of a larger, cleaned version of the annotations and associated image links is described in Section 3.2. As the annotated image data version 1 (aidv1) dataset, it serves both as the basis of the present deliverable, and as the training set for the ODeuropa Competition on Olfactory Object Recognition (ODOR) competition (D2.5). The dataset is published on Zenodo.[18] It represents the first dataset of olfactory objects on artworks, and to the best of our knowledge also the largest object detection datasets in artworks.

The corpus of annotation that was used for the creation of the dataset may be used to create further datasets for specific purposes, e. g., a dataset of olfactory gestures that might serve as training data or ground truth for gesture detection models.

Apart from their use as a basis for the creation of datasets, the corpus of manual annotations can also directly be fed into the EOKG, opening up the olfactory heritage inscribed into four centuries of visual culture.

References

[Buda et al., 2018] Buda, M., Maki, A., and Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259.

[Gonthier et al., 2018] Gonthier, N., Gousseau, Y., Ladjal, S., and Bonfait, O. (2018). Weakly supervised object detection in artworks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0.

[Huang et al., 2021] Huang, G., Laradji, I., Vazquez, D., Lacoste-Julien, S., and Rodriguez, P. (2021). A survey of self-supervised and few-shot object detection. *arXiv preprint arXiv:2110.14711*.

[Kuznetsova et al., 2020] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., et al. (2020). The open images dataset v4. *International Journal of Computer Vision*, 128(7):1956–1981.

[Lin et al., 2014] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

[Lisena et al., 2022] Lisena, P., Schwabe, D., van Erp, M., Troncy, R., Tullet, W., Leemans, I., Marx, L., and Ehrich, S. (2022). Capturing the Semantics of Smell: The Odeuropa Data Model for Olfactory Heritage Information. In *19th Extended Semantic Web Conference (ESWC), Resource Track*, Hersonissos, Greece.

[Liu et al., 2020] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., and Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision*, 128(2):261–318.

[Miller, 1995] Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

[Russakovsky et al., 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.

[Westlake et al., 2016] Westlake, N., Cai, H., and Hall, P. (2016). Detecting people in artwork with cnns. In *European Conference on Computer Vision*, pages 825–841. Springer.

[Zinnen et al., 2022] Zinnen, M., Madhu, P., Bell, P., Maier, A., and Christlein, V. (2022). Transfer learning for olfactory object detection.

[18] https://doi.org/10.5281/zenodo.6367776

A Label Overview AIDV1

Category	Supercategory	Occurrences
grapes	fruit	944
carnation	flower	790
peach	fruit	694
rose	flower	595
pipe	pipe	530
animal carcass	vertebrate	514
bird	vertebrate	497
anemone	flower	462
sheep	vertebrate	428
cherry	fruit	411
dog	vertebrate	401
other fruit	fruit	376
fish	vertebrate	353
jasmine	flower	347
plum	fruit	322
tulip	flower	317
pear	fruit	313
other vessel	drinking vessel	306
gloves	gloves	302
iris	flower	300
other flower	flower	283
lily	flower	257
cow	vertebrate	251
heliotrope	flower	232
butterfly	invertebrate	230
columbine	flower	219
apple	fruit	213
daffodil	flower	200
currant	fruit	199
cornflower	flower	197
lemon	fruit	190
fig	fruit	181
horse	vertebrate	176
goat	vertebrate	174
lily of the valley	flower	148
mushroom	vegetable	146
hyacinth	flower	140
pig	vertebrate	140
other vertebrate	vertebrate	140
melon	fruit	139
ring	jewellery	125
nut	nut	122
violet	flower	119
glass with stem	drinking vessel	119
bread	bread	116
bivalve	invertebrate	116
geranium	flower	113
jug	drinking vessel	110
onion	vegetable	110

whale	vertebrate	105
cheese	cheese	101
donkey	vertebrate	99
smoke	smoke	97
fire	fire	92
poppy	flower	84
olive	vegetable	76
petunia	flower	74
meat	meat	73
pomander	jewellery	71
ashtray	ashtray	65
lilac	flower	65
neroli	flower	62
glass without stem	drinking vessel	60
carrot	vegetable	48
other invertebrate	invertebrate	44
strawberry	fruit	44
caterpillar	invertebrate	44
fly	invertebrate	41
candle	candle	37
pumpkin	vegetable	35
artichoke	vegetable	32
garlic	vegetable	32
lobster	invertebrate	29
bracelet	jewellery	28
prawn	invertebrate	28
lavender	flower	26
censer	censer	26
cat	vertebrate	25
bug	invertebrate	24
carafe	drinking vessel	15
wine bottle	drinking vessel	15
chalice	drinking vessel	8
coffeepot	drinking vessel	7
cup	drinking vessel	2
teapot	drinking vessel	2

Table 6: Full label distribution of all categories present in aidv1, middle column gives the hierarchy information.

B Label Overview Raw Annotations

Category	Supercategory	Occurrences
flower		7,596
fruit		2,192
grapes	fruit	1,014
beverage		652
animal corpse	mammal	651
bird		630
peach	fruit	535
pipe	smoking equipment	524
dog	mammal	502
sheep	mammal	476
cherry	fruit	473
fish		440
vegetable		411
apple	fruit	390
gloves		388
holding the nose		364
pear	fruit	350
butterfly	insect	294
cow	mammal	272
plum	fruit	222
pig	mammal	213
bread		202
mushroom	vegetable	201
insect		197
ring	jewellery	197
horse	mammal	195
nut		195
mammal		194
lemon	fruit	187
goat	mammal	186
cheese		182
currant	fruit	181
onion	vegetable	160
smoking equipment		139
whale		138
meat		125
rose	flower	124
fire		122
pomander	jewellery	116
smoke		114
fig	fruit	104
olive	vegetable	102
donkey	mammal	102
pomegranate	fruit	89
melon	fruit	85
tulip	flower	84
carrot	vegetable	80
oyster	seafood	80
wine	beverage	69

censer		61
strawberry	fruit	60
pumpkin	vegetable	59
ashtray	smoking equipment	58
clam	seafood	57
tobacco-packaging	smoking equipment	57
cat	mammal	56
parrot	bird	55
tobacco	smoking equipment	54
bracelet	jewellery	53
candle		53
artichoke	vegetable	53
caterpillar	insect	49
prawn	seafood	46
lobster	seafood	41
garlic	vegetable	40
match	smoking equipment	38
blackberry	fruit	37
monkey	mammal	32
radish	vegetable	32
rabbit	mammal	31
necklace	jewellery	31
tobacco-box	smoking equipment	30
reptile/amphibian		30
lizard	reptile/amphibian	30
snail	insect	29
snake	reptile/amphibian	28
aspargus	vegetable	26
fly	insect	26
deer	mammal	25
crab	seafood	25
orange	fruit	23
dragonfly	insect	21
seafood		21
bug	insect	20
walnut	nut	19
physalis	fruit	18
cigarette	smoking equipment	17
celery	vegetable	17
lion	mammal	17
ant	insect	16
duck	bird	16
spring onion	vegetable	15
frog	reptile/amphibian	15
cucumber	vegetable	15
handkerchief	other	14
spring onion	vegetable	13
crayfish	seafood	13
lily	flower	13
cauliflower	vegetable	13
daffodil	flower	10
gooseberry	fruit	10
spider	insect	10
cabbage	vegetable	9

tree		9
cocoon	insect	9
raspberry	fruit	8
salad	vegetable	8
chili	vegetable	7
goose	bird	7
apricot	fruit	7
sniffing		7
camel	mammal	7
earring	jewellery	7
rosehip	flower	7
flacon		6
rooster	bird	6
owl	mammal	6
guinea pig	mammal	6
eggplant	vegetable	6
iris	flower	6
quince	fruit	6
moth	insect	6
rat	mammal	6
other		5
swan	bird	5
rose hip	flower	5
dianthus	flower	5
hazelnut	nut	5
bear	mammal	5
grasshopper	insect	5
peacock	bird	5
mouse	mammal	5
hand fan		5
beer	beverage	5
punica	fruit	5
corn	vegetable	5
squirrel	mammal	4
bee	insect	4
watermelon	fruit	4
cigar-holder	smoking equipment	4
amulet	jewellery	4
herbs		4
leopard	mammal	4
jewellery		4
fennel	vegetable	4
poppy	flower	4
rope tobacco	smoking equipment	4
chicken	bird	3
crown	jewellery	3
oil lamp	lamp	3
elephant	mammal	3
boar	mammal	3
torch	candle	3
otter	mammal	3
coffee	beverage	3
eagle	bird	3
cheetah	mammal	3

cale	vegetable	3
shrimp	seafood	3
tiger	mammal	3
cigar	smoking equipment	3
myosotis	flower	2
ox	mammal	2
cleaning a baby		2
squid	seafood	2
eel	seafood	2
runner bean	vegetable	2
wig		2
cookie	other	2
bat	mammal	2
turkey	bird	2
red cabbage	vegetable	2
fox	mammal	2
cone	jewellery	2
polar bear	mammal	2
peeing		2
hen	bird	1
aries	mammal	1
vomiting		1
ferret	mammal	1
scallion	vegetable	1
coconut drink	beverage	1
cocoa	fruit	1
pineapple	fruit	1
beetle	insect	1
panther	mammal	1
tangerine	fruit	1
sunflower	flower	1
aspargurs	vegetable	1
parsley	other	1
comment	vegetable	1
paprika	vegetable	1
chilli	vegetable	1
savoy	vegetable	1
gas lamp	lamp	1
defecating		1
guineau pig	mammal	1
gull	bird	1
stag	mammal	1
belt	jewellery	1
banana	fruit	1
armadillo	mammal	1
tea	beverage	1
watepipe	smoking equipment	1
pipe tamper	smoking equipment	1
worm	insect	1
hedgehog	mammal	1
cockroach	insect	1
washing	other	1
tabacco	smoking equipment	1
pipie-tamper	smoking equipment	1

grashopper	insect	1
dove	bird	1
cigar-box	smoking equipment	1
mandarin	fruit	1
stranded	whale	1
melonnnn	fruit	1
bellflower	flower	1
glove	other	1
wulf	mammal	1
hair jewelry	jewellery	1
faeces		1

Table 7: Full label distribution for raw annotations, middle column gives the hierarchy information.