# Annotated Image Data version 2

Deliverable D2.4

Version FINAL


NEGOTIATING OLFACTORY AND SENSORY EXPERIENCES IN CULTURAL HERITAGE PRACTICE AND RESEARCH

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<br>KNAW Humanities Cluster<br>Email: inger.leemans@huc.knaw.nl |
| **Document Number** | Deliverable D2.4 |
| **Status & version** | FINAL |
| **Contractual date of delivery** | August 2022 - Extension granted to September 2022 |
| **Date of delivery** | September 2022 |
| **Type** | Data |
| **Security (distribution level)** | Public |
| **Number of pages** | 15 |
| **WP contributing to the deliberable** | WP2 |
| **WP responsible** | WP2 |
| **EC Project Officer** | Hinano Spreafico |
| **Authors:** Mathias Zinnen[1] | |
| **Internal reviewers:** Vincent Christlein[1], George Alexopoulos[2], Pasquale Lisena[3] | |
| **Affiliations** (1) FAU , (2) UCL, (3) EURECOM | |
| **Keywords:** Computer vision, dataset | |
| **Abstract:** This deliverable introduces the second version of the Odeuropa Dataset of Smell-Related Objects. Compared to the first version, we extend the image metadata, most notably the semantic field describing the image contents, we introduce 83 new categories, revise their hierarchy, and increase the number of annotated images from 2116 to 4696. Version two of the dataset is published on Zenodo under the doi 10.5281/zenodo.7125961 (https://zenodo.org/record/7125961). | |

# Table of Revisions

| Version | Date | Description and reason | By | Affected sections |
|---|---|---|---|---|
| 0.1 | Sept 26, 2022 | Draft for internal review | Mathias Zinnen | - |
| 0.2 | Sept 29, 2022 | Incorporate reviewer comments | Mathias Zinnen | all |
| 1.0 | Sept. 30, 2022 | Final check and approval by project manager | Marieke van Erp | - |

# Executive Summary

In this deliverable, we introduce the second version of the Odeuropa Dataset of Smell-Related Objects. We describe changes from the first version of the dataset in three aspects: 1. Metadata, where we include semantic fields, describing the image contents as well as license and language information. 2. Categories, where we include 83 new categories to include more smell-relevant objects and gestures and rework the class hierarchy 3. Dataset size, where we increase the number of annotations from 15 484 to 36 663, and the number of annotated images from 2 116 to 4 969. Furthermore, we report on the updated dataset statistics, justify the decisions that have been taken, and give a short outlook on the next steps that we plan to further improve the dataset.

| **Summary table** | |
|---|---|
| **Challenges** | Annotating objects in artworks require lots of manual work. Additionally, ensuring a sufficient annotation quality becomes increasingly difficult as the number, and granularity of categories increase. This is best exemplified by the annotation of the 27 flower species we found, where the correct annotation requires expert botanical knowledge and can be extremely time-consuming. While we could account for this problem with our expert annotators who gained more and more experience in analyzing historical artworks, this is even more challenging for crowdsourced annotations with Amazon Mechanical Turk where anonymous workers with varying expertise are asked to create annotations. |
| **Barriers** | Many objects with high-olfactory relevance, and particularly smell gestures, are very difficult to find in existing digital collections. One reason for this might be that the depictions of the respective objects or gestures simply do not exist in the digital collections, or, more likely, that they do exist but are not annotated according to the smell-relevant categories, that are of interest for us. Another barrier is the challenging nature of our dataset which permits the application of a semi-automated annotation approach, where machine-generated annotations only need to be corrected manually. The detection systems could not achieve a high enough accuracy on the dataset to realize this approach. |
| **Practices** | When annotating images with multiple annotators, and particularly when using crowdsourcing, we found that it is crucial to implement measures to ensure internal consistency of the annotations. To warrant consistency across our annotators, we keep a guidelines document in which annotators are asked to report the decisions they take while annotating (e.g. annotation of a flower entails only the blossom, not the stem). For the crowdsourced annotations, we created example documents that have been accessible while performing the annotation task on Mechanical Turk to give the annotators guidance on what each of the categories looks like[1]. |
| **Guidelines** | We follow the annotation guidelines that have been developed in the creation of dataset version one and are accessible on GitHub[2]. |

Table 1: Summary table about challenges, barriers, practices, and guidelines learned/developed in the deliverable.

[1] cf. https://htmlpreview.github.io/?https://raw.githubusercontent.com/Odeuropa/wp2-annotations/master/annotation-guidelines/examples.html for a document of flower species examples

[2] https://github.com/Odeuropa/wp2-annotations/blob/78cd55376b3e838476e37c9747d152565d396300/annotation-guidelines/annotation-guidelines.md

**Layman's summary**

We improved the dataset of visual smell references by adding more context-related metadata, increasing the overall number of images and annotations, and by extending the set of annotated objects, most notably we also included smell gestures in the dataset.

# Contents

# 1 Introduction

Collecting information about visual smell references and compiling them in form of a machine-readable dataset sets the prerequisite to both, the development of automatic extraction algorithms, and quantitative analysis of smells in past visual cultures.

For this reason, we have published the Odeuropa Dataset of Smell-Related Objects [Zinnen et al., 2022a] which contains 2 937 artworks, annotated with 24 391 bounding boxes from 87 categories of smell-related objects. The generation process of the dataset is described in deliverable D2.2. Apart from its project-internal usage for the training of machine learning models and quantitative research, the dataset has served as the training set of the *OD*europa Challenge on *O*lfactory Object *R*ecognition (ODOR) which has been held in the context of the ICPR2022 [Zinnen et al., 2022b]. In an adapted version, it has also been used to provide visual data for the participants of the Odeuropa PastScent workshop[3] that took place on August 22–23 in Amsterdam, and as the visual part of the multi-modal MUSTI [Hürriyetoğlu et al., 2023] challenge dataset, which will be held at the MediaEval22 workshop.[4]

Throughout its various applications, we noticed that there is room for improvement in multiple aspects:

1. The image *metadata* records miss some very useful fields that are present in at least some of the source collections we query for;

2. The number of object *categories* is insufficient to find relevant data for some research questions;

3. Increasing the number of *training samples* would likely lead to better performance in the automatic extraction of smell references (cf. D2.3 Object detection/ image analysis).

In the following, we report which changes have been made in the second, revised version of the dataset to address these issues. Apart from these changes the metadata, and image annotation format remain the same as in version 1 of the dataset with the specifications being reported in section 2.4 of D2.2 Annotated image data version 1.

# 2 Metadata Changes

Table 2 lists the metadata fields that we provided with version one of the dataset. These fields include all information necessary to unambiguously identify the artwork, view it in the context of its original collection, and download a local copy from its source collection. However, apart from the title field, which is often very instructive, it barely contains information about the content of the image. Since, in many cases, the metadata records of the source collection do provide information about the image contents, we extend the metadata by various “semantic” columns that will be discussed below. Additionally, we include license information and the language of the metadata records where we could obtain the relevant information. Table 3 gives an overview of all newly included metadata columns.

One problem we encountered when trying to use the data from our image dataset is that many images could not be redistributed or published because of their restrictive licensing. Manually looking up the licenses for each image that should be used in a publication or presentation turned out to be a laborious task. Thus, we adapted our metadata extraction script to automatically recognize image licenses and added this information to our metadata records under the ‘License’ key. Currently, we only distinguish between Public Domain, CC-BY, and missing license information, but we plan to extend this in the future.

[3] https://odeuropa.eu/pastscent
[4] https://multimediaeval.github.io/

| 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 2: Metadata fields of dataset version one.

| Metadata field | Description |
|---|---|
| Iconclass code | standardized codes describing the image contents |
| Description | free-form description of the image contents |
| Keywords | keywords describing the image contents, standardized within the collection |
| Language | Language of the metadata records |
| License | License of the image |

Table 3: Metadata fields added in dataset version two.

The language column specifies the language of the metadata associated with the image, which is useful for multi-modal approaches that leverage both textual information from the metadata and visual information from the images as it is done in the MUSTI challenge.

The Iconclass code, description, and keywords columns can be considered semantic insofar as they describe the content of the associated image. Iconclass[5] is a classification scheme for artworks that initially was published in printed form between 1973 and 1985 [Couprie, 1978] and is currently maintained by Hans Brandhorst and Etienne Posthumus [Brandhorst and Posthumus, 2016]. Iconclass is employed by various digital museum collections to tag artworks according to their content. Using Iconclass codes for the classification of artwork contents has the advantage that it is invariant to changes in metadata language or different words for the same concepts. The description column, on the other hand, captures textual descriptions of images that are included in some of the target collections. These descriptions can have varying lengths, ranging from single sentences to whole paragraphs. Somewhere in between formalized iconclass codes and free-form textual descriptions are the keywords: these are used by some collections in a way that is standardized within the collections but lacks interoperability between collections.

# 3 Category Changes

The most apparent change in the categories is their number. Compared to the 87 categories of dataset version one, the 170 categories of dataset version two exhibit an increase of 95%. This increase has two main reasons. First, while continuously annotating new artworks, our expert annotators identified new classes that appear frequently in the artworks. This led, for example, to the introduction of five new flower subclasses. Second, we specifically searched for categories that have been identified as smell-significant in interdisciplinary exchange with the

[5] https://iconclass.org

| Gesture | # Occurrences |
|---|---|
| Sniffing | 68 |
| Holding the nose | 368 |
| Vomiting | 5 |
| Cleaning a baby | 10 |
| Defecating | 5 |

Table 4: Number of smelling gesture examples in the dataset

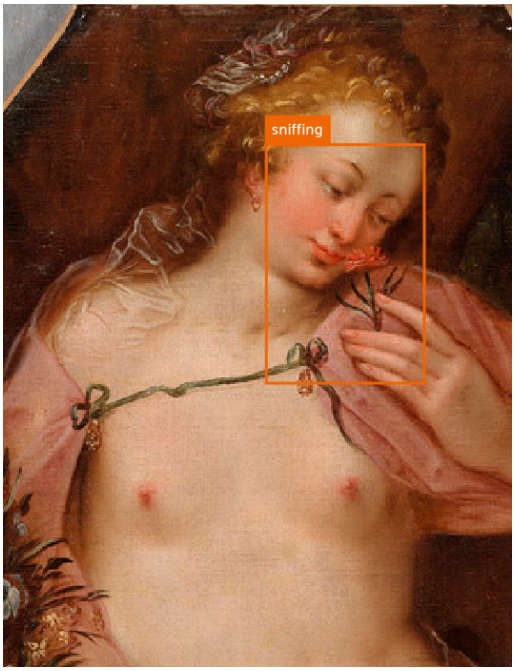
other work-packages, most specifically during a recurring sync meeting between representatives of WP2, WP5, WP6, and WP7. The most prominent examples of categories that have been added due to their high olfactory relevance are smell gestures. Their immense significance for past conceptualizations of smells should be immediately understandable, a detailed motivation for their automatic recognition can be found in section 2.2 of deliverable 2.2 or in section 6 of deliverable 2.3. Table 4 lists the five different smell gestures that we have in our dataset and their number of occurrences. Section 3 shows examples from the dataset. Considering the high olfactory relevance of smell gestures, it is noticeable that the number of annotated smell gestures is very low compared to the number of samples of some of the objects in the dataset. This is because it is very hard to find samples of smell gestures in digital collections due to a lack of smell-related meta-data in the museum collections [Ehrich et al., 2021] – a shortcoming that we are continuously trying to amend by searching for examples of smell gesture in a large variety of digital collections.

Another category of objects that turned out to be interesting in terms of potential smell research questions are chimneys, which have a strong connection to a history of air pollution and the development of urban smellscapes. We annotated a total of 350 chimneys with bounding boxes, of which some examples are provided in Fig. 2.

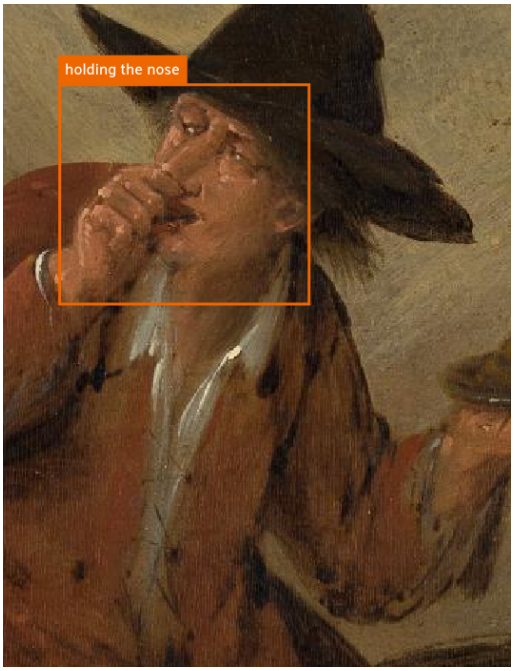
Apart from the added categories, we revised the category hierarchy that was used in version 1 of the dataset. Table 5 compares both class hierarchies in conjunction with the number of subcategories and samples for each of the supercategories. The most apparent change in the class hierarchy is that we refined the invertebrate, and vertebrate categories into mammal, bird, insect, fish, and reptile/amphibia categories. The rationale behind this was that the more fine-grained classification leads to objects of the same superclass having more similar visual features which enable the application of hierarchical approaches in their automatic recognition. The downside is a decrease in taxonomic consistency, but since the usage of the dataset as a training set for machine learning models is a major objective, we decided to prioritize visual consistency over taxonomic and biological consistency. This consideration also guided the subsumption of specific categories under seemingly inappropriate supercategories. One example is the whale category, which we subsumed under fish instead of mammal because its taxonomically correct classification as a mammal would make hierarchy-based classification more difficult since it shares most visual features with fish. Another example is the flacon category. Flacons are small bottles often used to store perfumes (hence their smell significance). We stored them under the category of drinking vessels although people usually did not drink out of flacons, but their visual appearance is quite close to that of a wine bottle.

# 4 Quantitative Changes

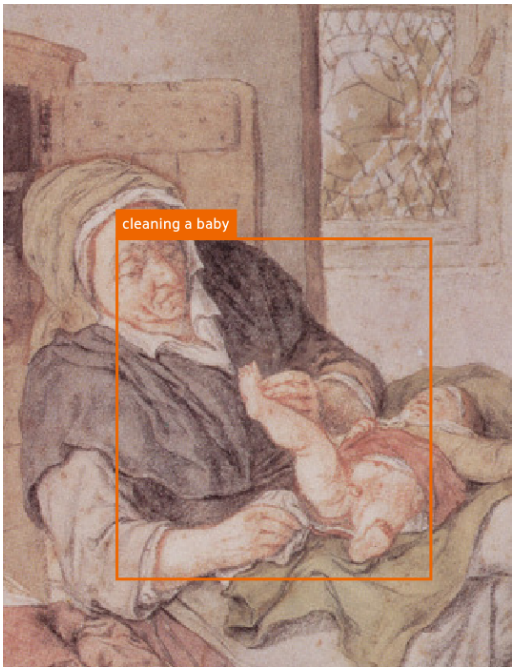
In version two of the dataset we were able to increase the number of annotated images from 2116 to 4969, the number of ground truth bounding boxes from 15 484 to 36 663, and the number of categories from 87 to 170. While this naturally implied a change in the class distribution of the dataset, the general tendency of having a long tail distribution remained unchanged as well as many of the most frequent classes. See Fig. 3 for a comparison of the class distributions of both




(a) Sniffing




(b) Holding the nose




(c) Cleaning a baby

(d) Vomiting

Figure 1: Four examples of smell-related gestures from the dataset. Note that the vomiting gesture box is not displayed because it would cover the gesture. Image credits (Details from):
(a) *Female personification of one of the five senses: Smell*. Anonymous. 1600 – 1610. RKDImages (108716). (b) *The five senses: smell*. Jan Molenaer. 1670 – 1700. RKDImages (278370). (c) *A woman changing a baby's nappy (allegory of smell)*. Cornelius Dusart. 1687. RKDImages (198783). (d) *Woman in an inn*. Egbert van Heemskerck. 1649–1704. RKDImages (271618).


(a) (b)

(c) (d)

Figure 2: Examples of annotated chimneys. Image credits (Details from):
(a) *Lange Bleekerssloot: view toward the Kostverlorenvaart*. Piet Mondriaan. c. 1898. RKDImages (263365). (b) *View of a street, a firehouse beyond*. Anonymous. 18th century. RKDImages (26303). (c) *House in the dunes, Scheveningen*. Hendrik Willem Mesdag. 1865–1915. RKDImages (287599). (d) *The royal wax candle factory*. Piet Mondriaan. 1900–1901. RKDImages (263410).

(a) Version 1

| supercat | subcats | samples |
|---|---|---|
| flower | 21 | 4855 |
| fruit | 12 | 4005 |
| vertebrate | 13 | 3192 |
| drinking vessel | 10 | 640 |
| invertebrate | 8 | 549 |
| pipe | 1 | 530 |
| vegetable | 8 | 479 |
| gloves | 1 | 302 |
| jewellery | 4 | 224 |
| nut | 1 | 122 |
| bread | 1 | 112 |
| cheese | 1 | 101 |
| smoke | 1 | 97 |
| fire | 1 | 75 |
| meat | 1 | 73 |
| ashtray | 1 | 65 |
| candle | 1 | 37 |
| censer | 1 | 26 |

(b) Version 2

| supercat | subcats | samples |
|---|---|---|
| flower | 27 | 9869 |
| fruit | 20 | 8317 |
| mammal | 21 | 5097 |
| smoking equipment | 11 | 2686 |
| drinking vessel | 13 | 2683 |
| vegetable | 22 | 1663 |
| smoke-related | 5 | 1064 |
| bird | 1 | 1043 |
| misc | 3 | 838 |
| fish | 2 | 811 |
| food | 4 | 785 |
| insect | 13 | 771 |
| jewellery | 7 | 532 |
| seafood | 4 | 504 |
| gesture | 5 | 474 |
| nut | 4 | 430 |
| clothing | 1 | 400 |
| reptile/amphibia | 4 | 115 |
| lamp | 2 | 17 |

Table 5: Overview of supercategories, their respective subcategories, and number of samples for dataset versions 1 and 2.

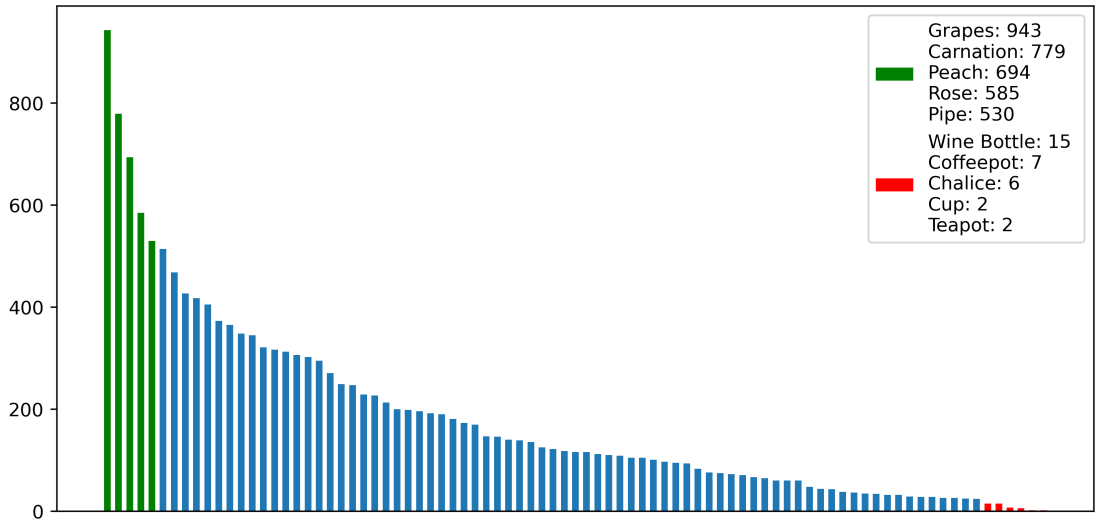
| Object | v1 occurrences | v2 occurrences |
|---|---|---|
| Wine Bottle | 15 | 209 |
| Coffeepot | 7 | 81 |
| Chalice | 6 | 56 |
| Cup | 2 | 251 |
| Teapot | 2 | 236 |

Table 6: Comparison of occurrences for classes that were rare in dataset version one.

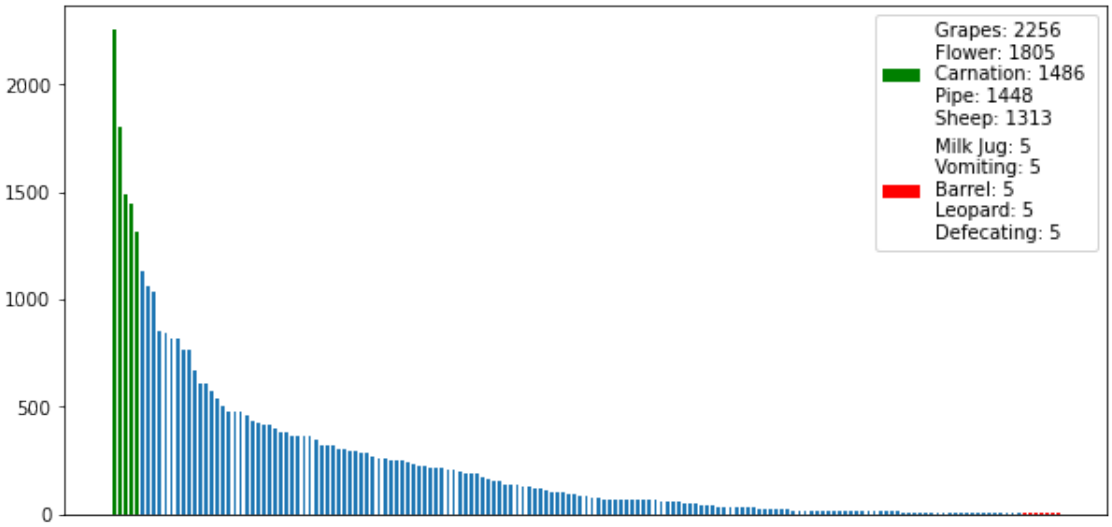
versions of the dataset. We specifically tried to increase the number of occurrences of classes that had very few samples in version one of the dataset. This is exemplified by Table 6 which lists the changes of occurrences of the five least frequent classes of dataset version one. However, as we considerably increased the number of classes in the dataset, the rare classes have been replaced by new categories for which we could not yet annotate a larger number of samples.

Initially, we had planned to create even larger amounts of annotations by leveraging computer vision algorithms to generate suggestions for bounding boxes in unseen images and then manually correcting these. Unfortunately, the automatic detection turned out to be even more challenging than we expected, which can be explained by the challenging properties of the dataset, such as the prevalence of small objects, the amount of occlusion, and the varying degrees of artistic abstraction (cf. D2.3 Object detection/ image analysis).

This, in turn, prevented us from implementing detection systems that have a high enough accuracy to successfully use them for such a semi-automatic annotation approach. Additionally, the semi-automated approach does not allow the continuous extension of the set of detection categories that we perform to meet the changing requirements that come up in the interdisciplinary exchange with the other work packages. We therefore adhered to the manual annotation process that we developed in dataset version one. However, we still plan to apply the machine-guided




(a) Class distribution of dataset version one.




(b) Class distribution of dataset version two.

Figure 3: Comparison of class distributions for dataset version one and two.

approach as soon as we succeed in implementing detection systems with a sufficiently high accuracy.

# 5 Summary and outlook

The annotated image dataset that has been introduced in the context of Deliverable D2.2 Annotated image data version 1 Meanwhile, derived versions of it has been put to multiple uses: It served as the training set of the ODOR challenge, as part of the multi-modal MUSTI challenge dataset, as basis for the visual analysis of past smells in the context of the PastScent workshop, and as the training data for algorithms to automatically detect olfactory references in historical artworks.

Throughout these uses, we have identified three main areas of possible improvements, i.e. *metadata*, *object categories*, and *number of samples*. In version two, we have extended and improved the dataset in each of these aspects. We have extracted five additional metadata keys that provide basic semantic information about the image contents as well as license and language information. In terms of categories, we have nearly doubled the number of classes. The categories now include highly smell-relevant categories such as smell gestures. Furthermore, we reworked the class hierarchy in a way that facilitates its exploitation for hierarchical approaches in object recognition.

Quantitatively, we were able to increase the number of annotations and images by more than 100%, raising the number of images from 2000 to 5000, and the number of annotations from 15 000 to more than 36 000. We specifically put an emphasis on those classes that have been less represented in version one of the dataset and could increase the number of samples significantly. Unfortunately, we were not able to annotate a sufficiently large number of smell gestures because it turned out to be extremely challenging to collect depictions of these gestures from digital collections. However, since we are aware of the high significance of smell gestures, we are still evaluating a large list of digital collections and intend to find more examples in the future.

Another direction of future activities is the semi-automatic annotation where we want to generate annotation suggestions with object detection algorithms and only manually correct them afterwards. This approach would enable a large quantitative extension of the dataset but requires an improvement of the object detection algorithms on which we are currently working.

The recently published version two of the dataset represents one of the largest datasets for object detection in historical artworks and the only dataset for the detection of olfactory references. We hope that it will be used widely and turns out to foster research in both computer vision for the visual arts, and the application of computer vision in the context of olfactory heritage.

# 6 Acknowledgements

First and foremost, we want to thank our tireless expert annotators Hang Tran and Azhar Hussian whose diligent work was the most important factor for the creation of the dataset. Apart from that, we thank all the Odeuropa team for all the fruitful discussions, valuable feedback and their patient explanations of how visual data might be used for a reconstruction of olfactory heritage.

# References

[Brandhorst and Posthumus, 2016] Brandhorst, H. and Posthumus, E. (2016). Iconclass: a key to collaboration in the digital humanities. In *The Routledge Companion to Medieval Iconography*, pages 201–218. Routledge.

[Couprie, 1978] Couprie, L. D. (1978). Iconclass, a device for the iconographical analysis of art objects. *Museum International*, 30(3-4):194–198.

[Ehrich et al., 2021] Ehrich, S. C., Verbeek, C., Zinnen, M., Marx, L., Bembibre, C., and Leemans, I. (2021). Nose first: Towards an olfactory gaze for digital art history. *First International Workshop on Multisensory Data and Knowledge*. Online accessed, December 09, 2021.

[Hürriyetoğlu et al., 2023] Hürriyetoğlu, A., van Erp, M., Akdemir, K., Troncy, R., Menini, S., Lisena, P., Tonelli, S., Paccosi, T., and Zinnen, M. (2023). Musti - multimodal understanding of smells in texts and images. unpublished.

[Zinnen et al., 2022a] Zinnen, M., Madhu, P., Kosti, R., Bell, P., Maier, A., and Christlein, V. (2022a). Odeuropa dataset of smell-related objects.

[Zinnen et al., 2022b] Zinnen, M., Madhu, P., Kosti, R., Bell, P., Maier, A., and Christlein, V. (2022b). Odor: The icpr2022 odeuropa challenge on olfactory object recognition. In *2022 26th International Conference on Pattern Recognition (ICPR)*. forthcoming.