Computer Vision Challenge

Deliverable D2.5

Version DRAFT

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 KNAW Humanities Cluster Email: inger.leemans@huc.knaw.nl
Document Number	Deliverable D2.5
Status & version	FINAL
Contractual date of delivery	December 2022
Date of delivery	20 December 2022
Type	Other
Security (distribution level)	Public
Number of pages	16
WP contributing to the deliberable	WP2
WP responsible	WP2
EC Project Officer	Hinano Spreafico
Authors: Mathias Zinnen,[1] Vincent Christlein[1]	
Internal reviewers: Sofia Ehrich,[2] Ali Hürriyetoğlu[2]	
Affiliations (1) FAU, (2) KNAW	
Keywords: Challenge, object detection	

Abstract: We held the Odeuropa ICPR2022-ODOR competition where we challenged participants to recognise smell-active objects in historical artworks using methods of object detection. All in all, we had 36 registered teams, of which 4 submitted to the final phase. The resulting detection systems pushed the performance of object detection on the ODOR dataset (cf. D2.2) to 11.5 % mAP, thus improving over the recognition capabilities of our own systems at that time. The winning team *Thousandwords* applied a state-of-the art one-stage object detection model and a variety of tricks which we plan to adopt to improve our own recognition system. Specifically small objects and fine grained classification turned out to be as difficult as expected which was reflected in the evaluation results. For larger objects, however, and for some specific, smell-significant categories such as censers, gloves, or pomanders, the participants could achieve decent detection capabilities. The challenge results encourage us to keep working on the improvement of our algorithms to automatically extract smell references from historical artworks.

Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.1	18 November	Full Draft	Mathias Zinnen	all
0.2	29 November	Internal review	Vincent Christlein	all
0.3	7 December	Comments internal review processed	Sofia Ehrich, Ali Hürriyetoğlu	all
0.4	12 December	Comments internal review processed	Mathias Zinnen	all
1.0	20 December 2022	Final check and approval by project manager	Marieke van Erp	-

Executive Summary

In this deliverable, we describe the *OD*europa Challenge on *O*lfactory Object *Recognition* that was held in the context of the International Conference on Pattern Recognition 2022 (ICPR2022-ODOR).[1] We describe the dataset that was used in the challenge, give a short introduction on how the data was collected, and present dataset statistics. Furthermore, we report on the challenge rules, its different phases, and the respective participation. Before presenting the challenge results, we describe the evaluation metric that we applied to rate the submissions, and describe the methods of the four finalists. Finally, we evaluate the results quantitatively and qualitatively and interpret them in the context of the Odeuropa project and the general task of recognising visual smell references.

Summary table	
Challenges	Most importantly, there was a significant mismatch between the number of participating teams and the number of final submissions. We hypothesize that many participants wanted to gain access to the dataset but were not interested in investing work on solving the challenge. Another reason might be that participants were discouraged by the difficulty of the dataset, which leads to very low values in our evaluation metric when compared to object detection on photographic image datasets.
Barriers	A barrier was posed by amount of the work needed to organise the challenge and by the number of work hours we had available in our working package. We learned that organising a challenge is much more laborious than we expected and needs help by many co-organizers. Another barrier was that typically first-time iterations of challenges struggle with low participation. We tried mitigating this by actively promoting the challenge via mailing lists, on the Odeuropa and pattern recognition lab homepages, and via twitter.
Practices	There are several learning points that we can use as practices for the organisation of future, and follow-up competitions: • To motivate active participation, also in the final phase, offering a reward, e. g., small amount of price money, is likely to be helpful. • To ensure all participants write a short paragraph about how their systems work, it turned out to be helpful to personally contact and encourage them. • To prevent participants from being discouraged by their results, publish baseline results and a simple, reproducible baseline method from the start. • Think carefully about the naming of the challenge. We believe that the name of the challenge (ODOR) might have been misleading and made people think that they would have to predict smells from molecular structures instead of performing object detection. The actual challenge objective should be part of the challenge name. • The relatively low number of final participants might have also been caused by choosing a lesser-known competition platform. Selecting a more well-known platform such as kaggle[2] might have helped to attract more participants (unfortunately kaggle is a commercial platform in contrast to codalab).
Guidelines	As guidelines for the technical implementation of the competition, we used the codalab user manual available at https://github.com/codalab/codalab-competitions/wiki. For the implementation of the final evaluation, we relied on the pycocotools library that is available under the cocoapi tools on GitHub (https://github.com/cocodataset/cocoapi).

[1] https://odor-challenge.odeuropa.eu/
[2] https://www.kaggle.com

Layman's Summary

We organized the ICPR2022-ODOR challenge, a competition where the participants were asked to localize and classify a set of 87 smell-active objects in historical artworks. Therefore, we provided a set of annotated artworks for the participants to train and evaluate their object detection algorithms on. This public part of the challenge datasets is identical to the annotated image dataset v1 which was presented in D2.2. For the evaluation of the participants' submissions, we created another set of annotated images, which we kept secret. We asked all participants to perform object detection on this set of images without releasing the correct annotations and compared the resulting predictions with our correct solution. Depending on how close the predictions were to this gold-standard, we ranked the participants. All in all, we had 36 participating teams, of which 4 submitted to the final phase of the challenge. The winning team submitted a recognition system that exhibited a better performance for the detection of olfactory objects than our own algorithms at that time. Since all teams were asked to provide method descriptions of their submissions, we were able to adopt many of the tricks and methods that the winning team used to improve their performance.

Contents

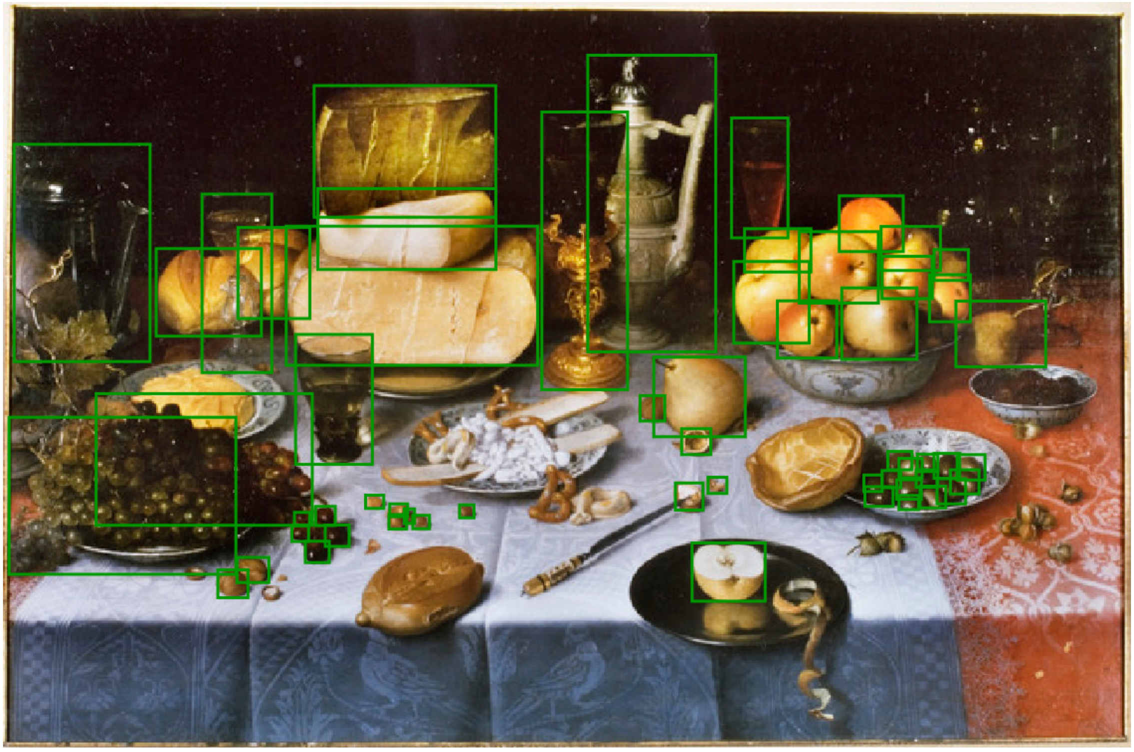


Figure 1: Example image from the challenge dataset exhibiting a large number of small, partially occluded objects. Image credit: *Laid Table with Cheese and Fruit*. 1610. Floris van Dyck. Public Domain, via Wikimedia Commons.

1 Introduction

Finding smell references in historical artworks is a very challenging task. These references can be implicit in a painting's narrative, the actions of depicted characters, or the depicted spaces. We try to approximate the recognition of complex and implicit smell references by first detecting objects with olfactory relevance, based on which more complex smell references might be recognized. The detection of olfactory objects in historical artworks is challenging in multiple aspects:

(1) Object detection in the artistic domain requires algorithms to cope with varying degrees of abstraction and artistic styles, which leads to a considerably higher intra-class variance than photographic depictions.

(2) In contrast to the famous COCO [Lin et al., 2014] and ImageNet [Russakovsky et al., 2015] datasets, where the images usually contain repetitive objects with huge per sample instances, historical artworks usually contain many object instances of diverse sizes, which are often partially occluded (cf. Fig. 1).

(3) Smell-relevant objects can be particular, leading to a fine-grained classification of target objects. Different types of flowers, for example, might have a different smell although looking very similar. Even if the challenge participants did not have to assign smells to the flowers, we still want to use the results to enable an analysis of depicted smells later.

(4) Since the dataset covers a period over multiple centuries, the appearance of some target objects is subject to historical change. Particularly, man-made objects like cigars or beverages might have changed their look over the years, whereas others like flowers or animals remained mostly invariant.

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

Table 2: Overview of search terms with the number of images collected for each.
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 used to burn incense or perfume in solid form.

The category and domain gap between photographic datasets and our target domain poses a challenge that encourages new approaches to increase object detection models' robustness and transfer capability. In posing the double challenge of overcoming a domain and category gap, we want to foster the development of domain adaptation techniques in object detection and promote a multisensory cultural heritage perspective on computer vision that acknowledges the importance of olfaction.

We allow and encourage the use of different kinds of pre-training on photographic data to enable various domain adaptation methods, e. g., transfer learning or style transfer. Along with our annotated dataset, we provide a hierarchy of object categories, which facilitates the implementation of hierarchical approaches to object detection.

2 Dataset

We provide the first dataset of olfactory objects within artworks for the challenge. This section shortly describes the collection, annotation, and a brief description of class distribution. The training set matches the Annotated Image Dataset version 1 as reported in D2.2. In this report, we also provide a more detailed analysis of the dataset and its creation process.[3] For a more detailed analysis of the dataset and its creation process, please refer to the according report.

2.1 Image Collection & Annotation

As a prerequisite for the assembly of the dataset, we queried multiple digitized museum collections using a list of search terms (cf. Table 2) that allegedly led to images with olfactory relevance. Our image collection strategy is two-fold: In the first step, we defined an initial list of search terms, which led to a collection of 30 134 artworks. As our knowledge about contexts in which smell active objects might appear evolves in the annotation process, we extended the image base with new search terms that have become relevant in multiple iterations.

[3]Please refer to https://odeuropa.eu/wp-content/uploads/2022/05/D2_2_Annotated_Image_Data_version_1.pdf for the report and to the zenodo record for the dataset.

The objects were annotated manually using cvat[4] and Amazon mechanical turk (only flower subcategories).

We predefined a set of categories that were then iteratively extended resulting in a list of 222 classes to date. The high number of object categories, including objects that are 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). In contrast to a WordNet-based concept hierarchy as it is applied by Redmon *et al.* [Redmon and Farhadi, 2017], we 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. Based on our fine-grained object categories the complete WordNet hierarchy can, however, still be created. The selection of the supercategories is based on pragmatic considerations such as visual similarity, assumed familiarity with concepts, and simplicity.

Finally, we filtered out supercategories that had less than ten samples for creating the challenge dataset, resulting in a list of 87 categories.

2.2 Label Distribution

Table 3 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.

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

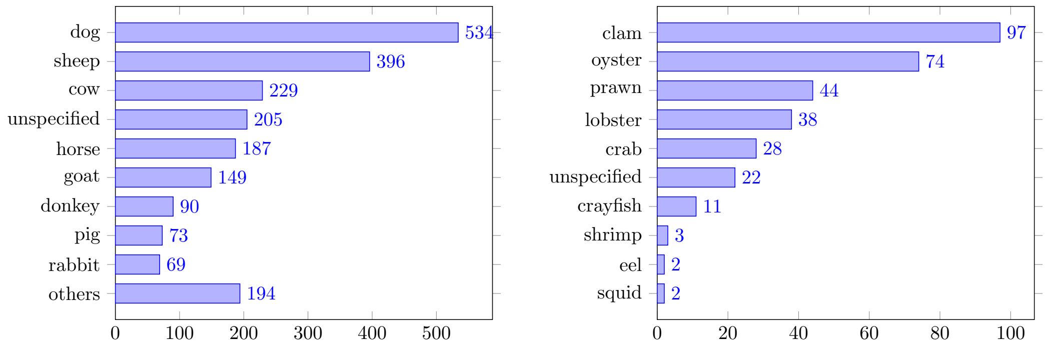
Table 3: Supercategories of the annotation scheme. The middle column gives the number of subcategories that have been defined for each of the supercategories. The right column reports the number of samples that have been annotated for the supercategory including its subtypes. *Other* subsumes all top-level categories that do not have further subcategories.

Figures 2a and 2b show the exemplary subcategory distributions of the *mammal* and *seafood* categories, respectively.

2.3 Distribution Format

Due to license compliance, we cannot publish the images directly. Instead, we provide a CSV file with URLs pointing to the image sources and a script to conveniently download them. The annotations are provided in COCO JSON format[5] which defines a bounding box as $[x, y, w, h]$,

[4] https://openvinotoolkit.github.io/cvat/
[5] https://cocodataset.org/#format-data



(a) (b)

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

with x and y denoting the coordinates of the upper left corner of a box, and w, h the box width and height, respectively. Additionally, each bounding box is assigned to one of the predefined categories via the *category_id* attribute. Apart from the publication via codalab,[6] the challenge training set is also published on zenodo [Zinnen et al., 2022] including additional metadata.

3 Challenge Overview

The aim of the ODOR challenge was to locate and classify a diverse range of odor-active objects on historical artworks. The participants are provided with a training set of artwork images along with the bounding box annotations for the target objects. Additionally, they are provided with a validation set of images without annotations.

The competition started with a preliminary warm-up phase, where the participants were given training data and a starter kit enabling them to perform exploratory data analysis and build initial prototypes and setup their code. Subsequently, the main challenge was conducted in two phases: (1) a development phase (2) and a final phase. For both, *development* and *final* phases, submissions were expected as a zip file containing the predictions as a COCO-*JSON* format.

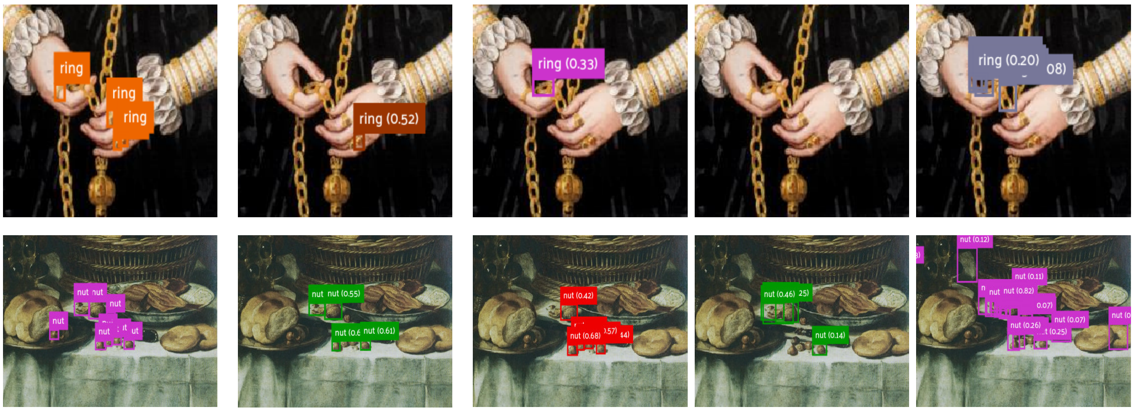
Development phase.

For the development phase, the bounding box annotations of the validation set were not provided to the participants. During this phase, participants were allowed to upload their predictions on the validation set and used only the annotated training set to train their algorithms. The validation set bounding boxes were used to evaluate each participant's submission and provide feedback as per the COCO evaluation metric. Each participant was allowed to upload one submission per day.

Final phase.

During the final phase, the validation annotations and the test set (without annotations) were provided to the teams to further fine-tune their models and present robust and generic algorithms on the test set. As the validation set annotations were released, the participants could use these as additional training data. Similar to the development phase, they were required to submit their results on the test set. For this phase, each participant could submit a total of six submissions.

[6]https://codalab.lisn.upsaclay.fr/competitions/1939



(a) Ground Truth (b) Thousandwords (c) None (d) DeadlyDL (e) angelvillar96

Figure 3: Qualitative comparison of small-object prediction results of the four finalists. The first row shows predictions for rings in a portrait, whereas the second row shows predictions of partially occluded nuts.
Image credits: *(top)* *Portrait of an 18-year old woman*. Attributed to Pieter Pourbus. 1574. Oil on panel. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/280945. *(bottom)* Detail from *Stilleven met een mand met kazen*. Pieter Claesz. 1645 – 1661. Oil on panel. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/108716.

3.1 Evaluation Metrics

We use *COCO metric* as the evaluation metric which determines the participants ranking in the final leaderboard. To understand any object detection metric, we need to understand Intersection over Union (IoU). IoU measures the overlap of the areas of a correct object localisation and a models' prediction. It thus measures if a predicted bounding box is correct with respect to the ground truth object bounding box or not. It is defined as the ratio of intersection and union between the predicted and actual bounding box. A prediction is considered to be correct (True Positive) if IoU is greater than a predefined threshold value. If this is not the case, then it is considered a False Positive. For COCO evaluation, the predefined IoU thresholds range from 0.5 to 0.95 with a step size of 0.05. We evaluate *COCO metric* by calculating the mean average precision (mAP) based on the average over all threshold values (IoU 0.5:0.05:0.95). Since our dataset contains many small objects that are particularly difficult to detect, we also report the mAP for small, medium, and large objects separately.

3.2 Participation

A total of 36 teams registered for the challenge, out of which 6 teams submitted during the development phase, and 4 teams submitted their predictions for the final phase. Although we are happy with the contribution of the existing participants, we initially expected more submissions. One reason might be the challenging nature of the dataset which might discourage some scholars. By skimming through the available codalab challenges, some scholars might have also misinterpreted the challenge name which, in its abbreviated form, does not explicitly link to object detection. We plan to create a follow-up where we consider these findings and attract more participants.

To simplify participation, we provided a simple *baseline method* that was published on GitHub during the training stage.[7] Retrospectively, it might have been helpful to provide the baseline method already at the start of the warm-up phase to attract more participants. For the baseline, we used an ImageNet pre-trained Faster-RCNN with a Resnet-50 FPN backbone. First, we fine-tuned only the head for 10 epochs using a learning rate of 1e-3, followed by 50 epochs of training the

[7] https://github.com/Odeuropa/ICPR-ODOR-starting-kits/

whole network with the same learning rate of 1e-3 before using a lower learning rate of 1e-4 for another 50 epochs. We used mild data augmentation as provided by the albumentation library and normalized the input using ImageNet-based mean and standard deviation.

Team *Thousandwords* consists of Ten Long (University of Amsterdam), Sadaf Gulshad (University of Amstedam), Stuart James (Istituto Italiano di Tecnologia), Noa Garcia (Osaka University), and Nanne von Noord (University of Amsterdam). They proposed the use of a strong object detector network called PPYOLO-E [Long et al., 2020] with a CSP-Resnet [Wang et al., 2019] backbone. The final results were obtained by training the network for 150 epochs using a batch size of 10, base learning rate (LR) of 2.5e-3. They used stochastic gradient descent with momentum as optimizer for the final model. The final model training used a LR scheduler for 5 epochs of *LinearWarmup* and maximum 360 epochs of *CosineDecay*. For augmentations, they used *BatchRandomResize* with target random sizes of [320, 352, 384, 416, 448, 480, 512, 544, 576, 608, 640, 672, 704, 736, 768] and random interpolation. They normalized the images with a mean of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225]. They experimented with various training schemes such as using grayscale images as augmentation, excluding small bounding boxes for robust learning and style transfer as augmentation for domain adaptation. Interestingly, they reported that none of these techniques work better than using a strong object detection model.

Team *None*[8] proposed the use of a YoloV5 [Jocher et al., 2022] model pre-trained on COCO. They fine-tuned the model using the Ultralytics platform[9] for 50 epochs with a learning rate of 1e-3 and a batch size of 16. For training, the team applied mild data augmentation as given by the *aug_tfms* function of the albumentations [Buslaev et al., 2020] library.

Team *DeadlyDL* with the single member Badhan Kumar Das (Siemens Healthineers) used a Faster RCNN [Ren et al., 2015] model for this task. The model was trained for 80 epochs with a learning rate of c. 5e-4 (0.000478), determined by the learning rate finder [Smith, 2017], a batch size of 2 and ADAM optimizer. For preprocessing, the team used padding and data normalization before passing the images to the neural network.

Team *angelvillar96* (Angel Villar-Corrales, University of Bonn) used a single-shot object detection network called RetinaNet [Lin et al., 2017b] with a Resnet50-FPN [Lin et al., 2017a] backbone pretrained on COCO-2017 dataset. The team used the Adam optimizer with an initial learning rate of 3e-4 with a decay factor of 10 (3e-5, 3e-6). The batch size was set to 32 due to hardware limitations and the network was trained for 50 epochs, with the best performance at 45th epoch. The final model was trained on a machine with an NVIDIA RTX A6000 with 48GB. Training for 50 epochs took about 1.5 hours.

4 Challenge Results

The submissions were ranked according to the *COCO metric* and the final score is listed in Table 4. The winner was team *Thousandwords* with members from the University of Amsterdam, Istituto Italiano di Tecnologia, and the Osaka University, the second place went to team *None*, *DeadlyDL* from Siemens Healthineers achieved the 3rd place, and *Angelvillar96* from the University of Bonn scored the 4th place.

To comprehensively evaluate the submissions, we also report the mean average precision (mAP) for small, medium and large bounding boxes. Table 5 shows that team *Thousandwords* achieved the highest mAP for all three types of bounding boxes. As expected, all submissions were struggling with small boxes. Compared with middle-sized boxes, we observe a performance decrease of more than 100 % for the first and second ranked team, and an even higher drop of about 350 % for the other participants.

[8]The participants preferred not to be mentioned in the paper.
[9]https://github.com/ultralytics/

	COCO mAP(%)	mAP@.5(%)	mAP@.75(%)
baseline	3.99	8.92	2.95
Thousandwords	11.49	18.93	12.00
None	7.52	12.16	8.29
DeadlyDL	4.58	10.00	3.77
angelvillar96	3.82	8.41	2.65

Table 4: Results on the final test set in terms of COCO mAP, Pascal VOC mAP (mAP@.5), and strict evaluation (mAP@.75).

	mAP-small(%)	mAP-medium(%)	mAP-large(%)
baseline	1.07	3.50	10.25
Thousandwords	4.19	11.71	25.24
None	3.03	7.36	15.74
DeadlyDL	1.00	4.50	10.43
angelvillar96	0.84	3.76	9.19

Table 5: Evaluation of COCO mAP for different object sizes

5 Discussion

Objects with few training samples, small objects, periodically changing objects with varying styles, and overlapping objects posed the major challenges of this competition. Figure 3 gives examples for some of the most challenging categories. The first two rows visualize detections of *small objects*, i. e., a portrait with three rings in the first row, and a still-life containing a large number of (partially occluded) nuts in the second row. Considering the object size, both nuts and rings are reasonably well detected by team Thousandwords and None. While the models of the teams DeadlyDL and angelvillar96 seem to largely overestimate the number of instances, the confidence score is below 0.5 for all instances, meaning that the false predictions do not decrease the COCO metric. However, the large number of overlapping predictions suggest that the usage or modification of non-maximum-suppression might improve the results. What surprised us was the detection performance for the allegedly challenging categories of smoke and fire. We expected both categories to be very challenging to detect since, especially in the case of smoke, they lack clear boundaries and their localisation is ambiguous. As Table 7 shows, our expectation was met for the teams None and angelvillar96 who both achieved a 0.0 precision for these categories. Surprisingly however, the teams Thousandwords and DeadlyDL achieved precision values considerably higher than their average over all categories. Figure 4, where the Thousandwords and DeadlyDL models both detect instances of smoke with blurry boundaries, emphasizes this finding.

Another positive outcome was the robustness of the participants towards deviations in stylistic representation of the target objects. Figure 5 shows detection of the Thousandwords method for three different representations of pipes. Although the right image exhibits a completely different artistic style, the pipe detection is still detected successfully. Furthermore, the different variations of the pipe object exhibited by the leftmost and the middle image do not prevent their successful detection.

Challenging as expected was the detection of large numbers of objects partially occluding each other. Figure 7 shows detections of a heap of apples for three participants. None of the participant models managed to find the majority of the apples in the heap. This motivates an evaluation approach similar to the OpenImages [Kuznetsova et al., 2020] evaluation protocol where groups of objects with at least five overlapping instances are counted as successful detections if at least one instance in the bonding box around the group is being detected. We might adapt this evaluation protocol in a possible future challenge. Interestingly, we do not observe a confusion between the visually relatively similar categories of apples, peaches, and pears, which is reflected in the confusion matrix between those categories (cf. Table 8).

(a) Ground Truth (b) Thousandwords (c) None (d) DeadlyDL (e) angelvillar96

Figure 4: Qualitative comparison of prediction results for the challenging smoke and fire categories. Image credit: Detail from *Solomon's idolatry (1 Kings 11:7–8)*. Circle of Claude Vignon. 1650–1674. Oil on canvas. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/114441.

Figure 5: Exemplary pipe detections of the winning model over different stylistic representations. Image credits: *(l)* Detail from *Self portrait in the studio*. Jan Toorop. 1883. Oil on panel. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/128870. *(m)* Detail from *Portrait of a man smoking*. Anonymous. 1800–1850. Oil on panel. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/294941. *(r)* Detail from *Peasant seated with pipe*. Adriaen van Ostade. 1625–1685. Graphite on paper. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/198724.

	smoke AP	fire AP
Thousandwords	0.44	0.33
None	0.00	0.00
DeadlyDL	0.12	0.20
angelvillar96	0.00	0.00

Table 7: Average precision of smoke and fire categories for all finalists. All precision values are reported according to COCO evaluation.

	apple	pear	peach	none	other
apple	6	0	0	133	0
pear	0	0	0	34	0
peach	0	0	0	11	66

Table 8: Confusion matrix for detections of apples, pears, and peaches for team Thousandwords

(a) Ground Truth (b) Thousandwords (c) DeadlyDL (d) angelvillar96

Figure 6: Apple detections on a heap of occluded and overlapping apple instances. Team None did not have any detections. Image credits: Detail from *Still life with a lobster, glasswork, bread, cheese and parrots*. Artus Claessens. 1615–1644. Oil on canvas. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/16311.

(a) Ground Truth (b) Thousandwords (c) DeadlyDL (d) angelvillar96

Figure 7: Apple detections on a heap of occluded and overlapping apple instances. Team None did not have any detections. Image credits: Detail from *Still life with a lobster, glasswork, bread, cheese and parrots*. Artus Claessens. 1615–1644. Oil on canvas. RKD – Netherlands Institute for Art History, https://rkd.nl/explore/images/16311.

6 Conclusion

We held the Odeuropa Challenge on Olfactory Object Recognition to promote object detection in the challenging domain of digital heritage. A total of 36 teams participated in the challenge, among which 6 submitted to the development phase, and 4 teams submitted to their final predictions. By raising the attention of digital humanities and computer vision alike, the challenge increased the respective visibility and cooperation. We hope to promote an interdisciplinary approach that considers computational methods, particularly in olfactory heritage studies. We briefly introduced the four final submissions and analyzed their results qualitatively and quantitatively. The winning team shows promising results in terms of small object detection and robustness towards different styles. To further monitor the progress and enable easy benchmarking of newly developed algorithms, we will reopen the challenge for new submissions.

References

[Buslaev et al., 2020] Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2).

[Jocher et al., 2022] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., TaoXie, Fang, J., imyhxy, Michael, K., Lorna, V, A., Montes, D., Nadar, J., Laughing, tkianai, yxNONG, Skalski, P., Wang, Z., Hogan, A., Fati, C., Mammana, L., AlexWang1900, Patel, D., Yiwei, D., You, F., Hajek, J., Diaconu, L., and Minh, M. T. (2022). ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference.

[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., 2017a] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017a). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125.

[Lin et al., 2017b] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017b). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

[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.

[Long et al., 2020] Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., Ding, E., et al. (2020). Pp-yolo: An effective and efficient implementation of object detector. *arXiv preprint arXiv:2007.12099*.

[Redmon and Farhadi, 2017] Redmon, J. and Farhadi, A. (2017). Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271.

[Ren et al., 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

[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.

[Smith, 2017] Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 464–472. IEEE.

[Wang et al., 2019] Wang, C.-Y., Liao, H.-Y. M., Yeh, I.-H., Wu, Y.-H., Chen, P.-Y., and Hsieh, J.-W. (2019). Cspnet: A new backbone that can enhance learning capability of cnn.

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