Taxonomy of Olfactory Phenomena in Images

Deliverable D2.1

Version Final

Odeuropa

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 D2.1
Status & version	Final
Contractual date of delivery	June 2021
Date of delivery	31 June 2021
Type	Report
Security (distribution level)	Public
Number of pages	13
WP contributing to the deliverable	WP2
WP responsible	WP2
EC Project Officer	Hinano Spreafico

Authors: Mathias Zinnen,[1] Peter Bell[1]

Internal reviewers: Vincent Christlein,[1] Lizzie Marx,[2] Raphael Troncy,[3] William Tullett,[4] Inger Leemans,[2] Marieke van Erp[2]

Affiliations (1) FAU, (2) KNAW, (3) EUR, (4) ARU

Keywords: Taxonomy, Visual, Images, Mappings, WordNet

Abstract: This report documents the creation of a taxonomy of odour active phenomena. It explains decisions that have been taken regarding:

- The identification of relevant phenomena
- Their hierarchisation and ordering
- The connection to different image datasets and metadata systems
- The connection to other taxonomies that are developed in the Odeuropa project.

Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.1	May 2021	First Draft	Peter Bell, Vincent Christlein, Mathias Zinnen	all
0.2	May 2021	Internal Review	Raphael Troncy, Inger Leemans, William Tullett, Mathias Zinnen	all
0.3	June 2021	Revised Version	Peter Bell, Mathias Zinnen	4,A–E
0.4	June 2021	Revised Version	Peter Bell, Lizzie Marx, Mathias Zinnen	1–6
0.5	28 June 2021	Internal review	Marieke van Erp	1–6
0.6	29 June 2021	Revised version	Mathias Zinnen	1–6
1.0	soon	approval by project manager	Marieke van Erp	-

Executive Summary

This report documents the creation of a taxonomy of odour active phenomena, their identification, and their mapping to other ordering systems. It is used to document and communicate the decisions that have been taken in the process of developing the taxonomy.

Contents

1 Introduction

Being directly linked to our memories and emotions, smell is an essential part of how we experience the world. However, in the context of tangible and intangible cultural heritage, smell remains severely undervalued. The Odeuropa project aims at tracing, preserving, recreating and promoting the olfactory heritage of Europe. To find out how people in the European past thought about smells, and to reconstruct historical smells, state-of-the-art methods of artificial intelligence are applied to a large corpus of visual and textual data from 17th – 20th century Europe.

For the research on visual sources, we plan to automatically extract pictorial representations of smell and references to olfaction using a combination of computer vision algorithms. Computer vision techniques have successfully been applied to cultural heritage [Belhi et al., 2019] and historical artworks ([Madhu et al., 2019], [Bell and Impett, 2019]) in digital heritage research and the digital humanities. However, their application to sensory, and in particular olfactory, phenomena has not yet been attempted.

A prerequisite for the collection of olfactory phenomena in images is the identification of visual representations of smell, which can be very challenging. Aside from exceptions such as smoke, the odour is usually invisible, which requires us to look for references to odours instead of the odours themselves [Marx, 2021]. Many visual references to smell must thus rely on indirect representations such as reactions to a smell or depictions of odorous objects, which is why no classification system has yet been developed. This report documents the process of developing a taxonomy of visual olfactory phenomena for their detection in historical artworks and contains the following:

- It describes methods that are applied for the creation of the taxonomy.
- It presents mappings to existing non-olfactory taxonomies of visual objects such as ImageNet [Deng et al., 2009].
- It explains the relation to other taxonomies that are used in the Odeuropa project.

The resulting taxonomy will be used to derive a system of labels with which images can be annotated to create training data for computer vision algorithms. Learning from the label information, these algorithms will be trained to automatically detect and localize olfactory phenomena in historical artworks. Figure 1 shows a screenshot of CVAT [1], the interface by which the training images will be annotated with the visual olfactory phenomena that make up the taxonomy.

2 Gathering Olfactory Phenomena

We start by identifying different kinds of references to smell that can be expressed visually and collecting them in a list. These references might be distinct or subtle, they might be direct depictions of smell or its sources, or they might indirectly insinuate smells via metaphors or iconographies. Contributions to this list are made by project partners, taking multiple perspectives into account:

- Members of the Odeuropa project have backgrounds in Art History, Cultural Studies, Olfactory Museology, History, Linguistics, Cultural Heritage, Digital Humanities, Computer Science, and Chemistry. We make use of this wealth of interdisciplinary knowledge by encouraging every project member to contribute to shared lists of olfactory objects, spaces, and iconographies. This list is then used for image processing, text mining as it is done in WP3, and for the creation of an ontology of olfactory phenomena.
- Specifically in WP2, but also among other Odeuropa project members, there is art-historical expert knowledge available. Incorporating the art-historical perspective, we identify objects,

[1] https://github.com/openvinotoolkit/cvat

Figure 1: Screenshot of the the interface for the annotation of historical artworks. Each of the coloured boxes correspond to one label and contains information about the object's location and category. In this example, green boxes mark birds, violet boxes oysters, and yellow boxes grapes.

iconographies, and metaphors which are used to visually evoke associations to smell in specific periods or artistic genres. As an additional source of art-historical expert knowledge, we draw on existing iconographic taxonomies, such as Iconclass[2] [Couprie, 1978].

- Advised by experts on the chemistry of olfaction, we assemble a list of objects that carry or emit a strong smell. As the perception of smell can be highly subjective and vary from person to person, let alone between time periods, this chemistry-based list represents a more "objective" or naturalistic approach that focuses on the physical odorous attributes of objects, as opposed to the aforementioned human-centered approach that focuses rather on the perception of smell as a human experience.

The olfactory references collected in these steps are integrated into a single set of lists which are accessible for all Odeuropa team members and serve as a shared resource for further project development. Furthermore, a unique id is assigned to each of the list elements, providing a clear reference for their usage in multiple contexts. Although the specific reference identifiers will remain unchanged, these lists can and will be extended as we find more references, especially in the text side of the project. References to current snapshots of these lists can be found in table 3.

3 Detection Techniques

To automatically extract as many types of olfactory references as possible, we will apply and combine the following detection techniques from the field of computer vision:

- **Image classification (IC)**: In image classification, images are classified as a whole, depending on the object they display. An image classifier is trained with image-level labels, where each image is assigned only one label. The classifier predicts the class of the image from a predefined set of possible class labels;

[2] accessible online at www.iconclass.org

- **Object detection (OD)**: Object detection is the combined localisation and classification of objects in images or videos. Whereas the class assignment of each object is often similar to image classification techniques, it differs from image classification in that there is usually more than one object per image present. Training an object detector requires object-level labels, which contain not only a class assignment but furthermore information about the position of the object in the image (bounding boxes). Big image datasets usually contain much fewer object-level than image-level labels since the former are more expensive to achieve;

- **Face Detection (FD)**: Face detection is a specific kind of object detection problem, where the objects to be found are restricted to faces;

- **Pose estimation (PE)**: The aim of pose estimation is to recognize a set of key points (e. g. hands, shoulders, feet) that allow the estimation of orientation, poses, and gestures of the persons in an image;

- **Scene understanding (SU)**: Scene understanding techniques are applied to understand the semantic content of an image. Instead of only focusing on a local area (e.g. the shape and texture of a single object), scene understanding methods take a broader context into account.

4 Deriving A Taxonomy

We structure the elements of the taxonomy in a hierarchy and follow pragmatic considerations in defining the categories:

- The top-level categories (objects, gestures, iconographies, spaces) are given by the fact that there might be different methods necessary to detect these entities;

- Categories below the top level are defined in accordance with the requirements of the respective detection method.

Figure 2 shows the top-level categories of the taxonomy. The categories are meant to be mutually disjoint, i.e. if a phenomenon can be interpreted as an object and as an iconographical allusion, it will occur in each of the categories as a distinct entry.

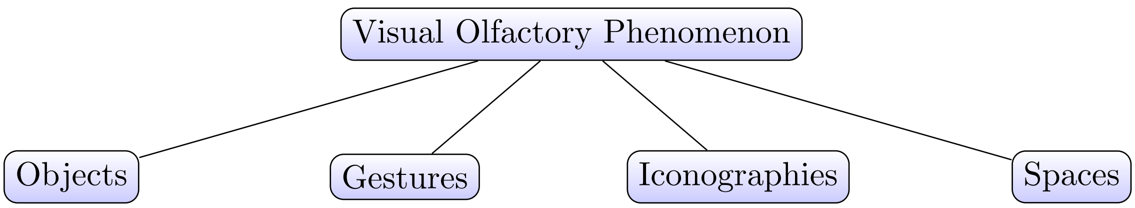

Figure 2: Top level categories for a taxonomy of visual olfactory phenomena

Olfactory objects are objects that carry or emit smells, such as tobacco pipes and flowers. They represent the most direct references we are dealing with. We assume that we can detect these kinds of olfactory references by applying an object detection method and annotating the images accordingly. **Olfactory Gestures** can either be reactions to a smell experience (e.g. holding the nose, bringing to the nose), or actions that produce a smell (e. g. urinating). We will experiment with different detection methods to recognise olfactory-relevant gestures. A person bringing a flower to the nose might for example be detected by a combination of object detection (flower, nose), pose estimation, and face recognition. **Olfactory iconographies** relates to visual narratives with olfactory insinuations on the one hand, and to metaphors and allusions on the other. By visual narratives, we mean depictions of iconographic scenes like *The Raising of Lazarus*. Apart from their visual content, they often carry additional meaning, since in the past iconographic scenes might have been linked to textual sources of the narrative. The example *Raising of Lazarus* illustrates this: Since the biblical verse talks about the stench of the dead Lazarus (John 11.39, [Bible, 1995]), a depiction of the scene might insinuate the stench even if there is no direct visual reference to olfaction in the painting. The detection of these kinds of references requires knowledge about these stories as well as about particular visual codes that hint at a specific iconographic scene or narrative. We will search for ways to incorporate art-historical background knowledge that enables us to recognize olfactory iconographies. Since in most cases, there is exactly one visual narrative present per image, we can rely on image-level classification methods. Metaphors and allusions on the other hand are often located on an object level. Like olfactory narratives, they represent allusions to smell that require the incorporation of knowledge about specific iconographic codes and symbols. Examples of olfactory metaphors and allusions are depictions of dogs, which signify the *sense* of smell, or depictions of flies and other insects, which allude to the presence of a stench, without emitting it themselves. Metaphors can be presented in the form of singular objects that have a specified position in the image, which requires us to apply object localization methods. Still, they cannot be detected like olfactory objects, since their symbolic reference is dependant on context information like the artwork's creation period and genre. We will presumably combine image-level metadata like the time and location of creation with object detection and scene understanding to consider the allusions and metaphorical objects in a broader context. **Olfactory spaces** are locations that are strongly related to smell. Examples of such locations are gardens, kitchens, or slaughterhouses. The detection of olfactory spaces differs from objects in that they can be detected by methods of image-level classification, allowing us to use image-level labels. Furthermore, we might make use of detected objects to infer whether an image depicts an olfactory space. Conversely, using the image-level information about olfactory spaces might help to predict correct object categories.

Table 1 lists the detection methods we plan to apply and combine for each of the top-level categories mentioned above.

Top-level Category	Detection Methods
Objects	Object Detection
Gestures	Object Detection, Pose Estimation
Iconographies	Image Classification, Object Detection, Scene Understanding, Metadata
Spaces	Image Classification, Object Detection, Scene Understanding, Metadata

Table 1: Planned Method Combinations for the detection of each of the top-level categories. Note that similar lists of detection methods might still differ in the way they are combined

.

Categories below the top-level categories are determined according to the technical requirements of the applied recognition methods. Since for gestures, iconographies, and spaces, we did not decide on a concrete implementation yet, the subcategories remain unfixed as well. Presumably, we will integrate the respective entities as plain lists. This is different for the olfactory object category.

It is in the nature of our task that the number of object types we would like to recognise is very large and that the relevant objects can be very specific. Literally every physical object can emit a smell, and the character of said smell can vary greatly between objects that visually seem very similar, for example two flower species that look similar might smell completely different. By implementing a recognition system as fine-grained as possible, we want to capture these nose-wise differences which are hardly visible to the untrained eye.

On a technical level, the difficulty of a correct object classification grows with the number of possible object classes. We aim to find a balance between prediction accuracy and model capacity in terms of detectable categories by starting with a small number of categories which we then gradually expand to recognise as many object types as possible while maintaining reasonable accuracy. In order to cope with objects that cannot be identified given the aforementioned technical limitations, we plan to apply a fallback strategy for object categories that are too specific for our models. Guided by the structure of WordNet synsets [Miller, 1995], we will exploit semantic information that is given by the WordNet hierarchy of concepts.

Figure 3: Woman wearing a pomander hanging from her belt. Virgo Tiguriensis / Ein Züricher Iungfraw, Wenceslaus Hollar, 1649. Retrieved from https://www.rijksmuseum.nl/en/collection/RP-P-OB-11.550.

A pomander, for example, is a fragranced pendant that was worn on the body, such as around the neck, the wrist, or from a belt for medical reasons or as a piece of scented jewellery (cf. fig. 3). In the past, depictions of pomanders carried great olfactory meaning. At the same time, no existing dataset has bounding boxes for the detection of this object class. Using the information from a hierarchy of hypernyms, we can fall back to objects of a more general category where an object class is too specific to be detectable. If we consider the semantic parents, we thus might at least be able to classify the object as an item of jewelry when recognizing it as a "pomander" is not possible. Figure 4 illustrates the semantic tree for the example of an English daisy. Whereas recognising the precise species (*bellis perennis*) might be too difficult, detecting a daisy, or a flower is likely to be feasible.

The ‘WordNet parent‘ of table 2 illustrates the notation of the hierarchical structure of objects. Where the WordNet semantic tree requires supertypes that are not yet part of the object list, they are appended as table rows (cf. IDs 101, 103, and 104).

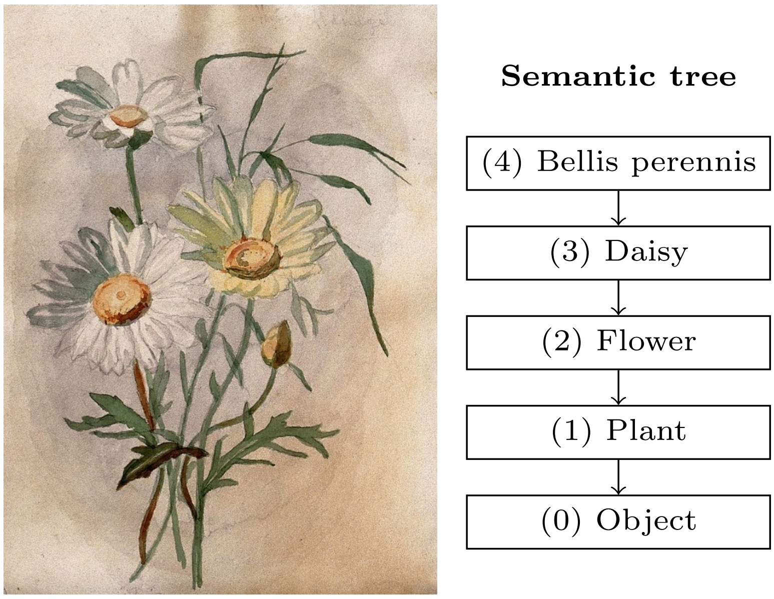

Figure 4: Semantic tree for a daisy, adapted from [Ridnik et al., 2021]. Painting: Ox-eye or marguerite daisy (Leucanthemum vulgare): flowers and leaves. Watercolour.. [Watercolors]. Retrieved from https://library.artstor.org/asset/25348720

The approach of incorporating hierarchical information from WordNet has been applied in computer vision before, incorporating a multi-class loss [Redmon and Farhadi, 2017] or multiple loss functions [Ridnik et al., 2021]. Using WordNet has the additional advantage that it forms the backbone for the labels of ImageNet [Russakovsky et al., 2015], a large dataset of photographic data.

ID	Name	WordNet ID	OpenImages ID	Iconclass code	WordNet parent	Nose-First parent
100	Bellis perennis	11960168	-	-	101	floral
101	Daisy	11959960	-	25G41(DAISY)	102	floral
102	Flower	11690372	17	25G41	103	floral
103	Plant	00017402	-	25G	104	-
104	Object	00002684	-	-	-	-

Table 2: Example snippet from the table of olfactory objects. The ‘WordNet parent’ column specified the conceptual hierarchy described above (cf. fig. 4. The ‘WordNet ID’, ‘OpenImages ID’, and ‘Iconclass code’ columns define the mappings to existing classification systems and datasets (cf section 5). The ‘Nose-First parent’ column defines a hierarchy that originates from a nose-first perspective (cf section 6). There is no complete match between these example entries and those of the living document that is used in the Odeuropa project. The former have been chosen for illustration purposes only.

5 Mappings

We collect images from various large archives of art and cultural heritage (Ambrosiana [3], Arkubid Uni Bonn [4], ArtPrice [5], Artstor [6], Ashmolean [7], Beni Culturali [8], Boijmans [9], British Library [10], Cini [11], Joconde [12], Foto Marburg [13], NGA Washington [14], Princeton Medieval Art [15], Princeton Art Museum [16], Prometheus [17], RKD [18], RMN [19], SLUB Dresden [20], SMB Berlin [21], Städel Museum, Warburg [22], Web Gallery of Art [23], Zeri [24]). to create a basis for training our models. The resulting dataset will be expanded with OmniArt[25] [Strezoski and Worring, 2018], which is particularly suited for our needs since it is the largest dataset of artistic data publicly available. By expanding our selection of digital archives with a larger, more generic dataset of artworks, we want to minimize possible biases and presumptions from the selection of data sources that might restrict the qualities of the detected olfactory phenomena to what we were already expecting.

Although these sources provide some metadata with the images, they are lacking annotations that are required for the recognition of olfactory references. Labels for classification can be located on the *image level*, where the whole image is annotated with a class based on what is depicted, as well as on the *object level*, where possibly multiple objects per image are annotated with a class and their position in the image (bounding box).

There are several large datasets which contain images that are annotated with image-level or object-level labels which greatly facilitates the classification and localization of relevant objects such as flowers or dogs. To be able to use existing labels, we need to map the olfactory references we have identified to other classification systems that are used by these datasets. We create mappings to the labels used by ImageNet, OpenImages, and Iconclass labels.

- **ImageNet** is arguably the most important dataset for computer vision. The full version [Deng et al., 2009] contains image-level labels for over 20.000 classes, and object-level labels for over 10.000 classes. Unfortunately for our application, it consists of photographic data and hardly contains artworks or prints. This requires us to apply domain adaptation techniques to exploit the dataset for training for the Odeuropa domain. The authors aim to provide on average 1.000 images to illustrate each concept of WordNet which theoretically would enable the detection of about 20.000 possible objects. In practice, however, about half of the 21.841 possible classes have fewer than 500 samples, which does not suffice to effectively train their recognition [Ridnik et al., 2021]. Still, ImageNet by far exceeds any other image dataset regarding the number of label classes and their respective samples. Being able to use this wealth for the training of our recognition systems is very promising. We

[3] https://www.ambrosiana.it
[4] http://www.arkubid.uni-bonn.de
[5] https://de.artprice.com
[6] https://www.artstor.org
[7] https://collections.ashmolean.org/
[8] http://www.catalogo.beniculturali.it
[9] https://www.boijmans.nl/en/collection/
[10] https://www.bl.uk/catalogues-and-collections/digital-collections
[11] http://arte.cini.it/
[12] http://www2.culture.gouv.fr/documentation/joconde/fr/pres.htm
[13] https://www.uni-marburg.de/de/fotomarburg
[14] https://www.nga.gov/collection.html
[15] https://ima.princeton.edu/digital-image-collections/
[16] https://artmuseum.princeton.edu/collections
[17] https://www.prometheus-bildarchiv.de/
[18] https://rkd.nl/en/collections/archives
[19] https://photo.rmn.fr/collections
[20] https://digital.slub-dresden.de/en/digital-collections
[21] http://www.smb-digital.de
[22] https://sammlung.staedelmuseum.de/en
[23] https://www.wga.hu/
[24] https://fondazionezeri.unibo.it/en/photo-archive/
[25] accessible online at http://www.vistory-omniart.com/

establish the taxonomy between ImageNet and our taxonomy by appending the elements of our taxonomy with the respective WordNet IDs. Until now, this connection has been established manually, but we plan to explore methods for the automated mapping such as the string2vocabulary tool[26] developed by the DOREMUS project [Achichi et al., 2015].

- **OpenImages** [Kuznetsova et al., 2020] in its current version (V6) contains object-level labels on 600 categories, and about 600.000 image-level labels on 19.957 categories. Similarly to ImageNet, it consists of photographic images. Although OpenImages is not comparable to ImageNet in terms of size and number of categories, we will still include OpenImages in the training of our models to see whether it improves the results. Mappings to OpenImages will be established manually.

- **Iconclass** is a classification system for iconographical description of visual arts [Couprie, 1978] that is applied by many digital collections to enrich their images with metadata. In contrast to ImageNet and OpenImages labels, which are applied to photographic data, annotations in Iconclass usually refer to artworks. This has the advantage that we can use Iconclass labels to directly train our models on digital collections of artworks, without having to rely on techniques of *transfer learning* [Pan and Yang, 2009] to transfer the knowledge from the photographic to the artistic domain. Mappings to Iconclass will be established manually.

The columns 'WordNet ID', 'OpenImages ID', and 'Iconclass code' of table 2 illustrate how mappings between the identified olfactory references and identifiers of the above-mentioned classification systems are persisted in the Odeuropa databases.

6 Other Odeuropa Ordering Schemes

Ordering the visual olfactory references according to the requirements of computer vision reproduces an ocular-centric perspective. In addition to a *visual* processing method, it will also be possible to order the olfactory references according to the principles of the Odeuropa project, by promoting a perspective that is sensitized for the non-visual, olfactory dimensions of cultural heritage.

We aspire to conceptualise our taxonomy as a purely pragmatic one, meaning that the categories and hierarchisation we have chosen are not meant to reflect the intrinsic properties of their elements by any means. Since all elements of the taxonomy can be addressed with a fixed identifier, which is defined in the common Odeuropa entity lists described in 2, the taxonomy enables and encourages the reordering of its elements according to other, non-visual aspects.

Other taxonomies and hierarchisations of the olfactory references have indeed been developed in the course of the Odeuropa project. One example is a nose-first approach, which groups olfactory references according to the scents they are associated with. By the definition of smell families and the assignment of associated smells to olfactory references, an olfactory similarity can be described where two objects are distinct from a visual point of view. The connection to this nose-first taxonomy is illustrated by the 'Nose-First parent' column of table 2. Entities of each of the categories can be assigned to a smell category which defines their position in a nose-first hierarchisation.

The object lists described in section 2 are explicitly open for extension such that they can arbitrarily be brought into different orderings that reflect a diversity of perspectives.

[26] https://github.com/DOREMUS-ANR/string2vocabulary

A Links to olfactory entity lists

Current snapshots of the lists of olfactory objects, gestures, iconographies, and fragrant spaces, annotated with the parent relations as described above can be found at the following URLs:

Olfactory objects	https://github.com/Odeuropa/d2.1-visualTaxonomy/blob/master/entity-list-snapshots/20210629_olfactory-objects.pdf
Olfactory gestures	https://github.com/Odeuropa/d2.1-visualTaxonomy/blob/master/entity-list-snapshots/20210629_olfactory-gestures.pdf
Olfactory Iconographies	https://github.com/Odeuropa/d2.1-visualTaxonomy/blob/master/entity-list-snapshots/20210629_olfactory-iconographies.pdf
Fragrant spaces	https://github.com/Odeuropa/d2.1-visualTaxonomy/blob/master/entity-list-snapshots/20210629_fragrant_spaces.pdf

Table 3: Table of taxonomy URLs

References

[Achichi et al., 2015] Achichi, M., Bailly, R., Cecconi, C., Destandau, M., Todorov, K., and Troncy, R. (2015). Doremus: Doing reusable musical data. In *ISWC: International Semantic Web Conference*.

[Belhi et al., 2019] Belhi, A., Gasmi, H., Al-Ali, A. K., Bouras, A., Foufou, S., Yu, X., and Zhang, H. (2019). Deep learning and cultural heritage: The ceproqha project case study. In *2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, pages 1–5. IEEE.

[Bell and Impett, 2019] Bell, P. and Impett, L. (2019). Ikonographie und interaktion. computergestützte analyse von posen in bildern der heilsgeschichte. *Das Mittelalter*, 24(1):31–53.

[Bible, 1995] Bible, H. (1995). New american standard bible. *Grand Rapids: World*.

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

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

[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*, pages 1–26.

[Madhu et al., 2019] Madhu, P., Kosti, R., Mührenberg, L., Bell, P., Maier, A., and Christlein, V. (2019). Recognizing characters in art history using deep learning. In *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents*, pages 15–22.

[Marx, 2021] Marx, L. (2021). Odours in art – depicting the invisible. *Fleeting - Scents in Colour, exh. cat. Mauritshuis, The Hague*.

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

[Pan and Yang, 2009] Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

[Redmon and Farhadi, 2017] Redmon, J. and Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, Honolulu, HI. IEEE.

[Ridnik et al., 2021] Ridnik, T., Ben-Baruch, E., Noy, A., and Zelnik-Manor, L. (2021). ImageNet-21K Pretraining for the Masses. *arXiv:2104.10972 [cs]*. arXiv: 2104.10972.

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

[Strezoski and Worring, 2018] Strezoski, G. and Worring, M. (2018). Omniart: a large-scale artistic benchmark. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):1–21.