Multimedia Tools and Applications (2024) 83:14637–14670
https://doi.org/10.1007/s11042-023-15968-9

A comprehensive survey on object detection in Visual Art: taxonomy and challenge

Siwar Bengamra[1,2] · **Olfa Mzoughi**[3] · **André Bigand**[2] · **Ezzeddine Zagrouba**[1]

Received: 29 December 2022 / Revised: 19 March 2023 / Accepted: 26 May 2023 /
Published online: 3 July 2023

Abstract
Cultural heritage data plays a key role in the understanding of past human history and culture, enriches the present and prepares the future. A wealth of information is buried in artwork images that can be extracted via digitization and analysis. While a huge number of methods exists, a deep review of the literature concerning object detection in visual art is still lacking. In this study, after reviewing several related papers, a comprehensive review is presented, including (i) an overview of major computer vision applications for visual art, (ii) a presentation of previous related surveys, (iii) a comprehensive overview of relevant object detection methods for artistic images. Considering the studied object detection methods, we propose a new taxonomy based on the supervision learning degree, the adopted framework, the adopted methodology (classical or deep-learning based method), the type of object to detect and the depictive style of the painting images. Then the several challenges for object detection in artistic images are described and the proposed ways of solving some encountered problems are discussed. In addition, available artwork datasets and metrics used for object detection performance evaluation are presented. Finally, we provide potential future directions to improve object detection performances in paintings.

Keywords Computer vision · Painting · Object detection · Deep learning · Explainability

✉ Siwar Bengamra
bengamra.siwar@gmail.com

Olfa Mzoughi
olfa.mzoughi@gmail.com

André Bigand
bigand@univ-littoral.fr

Ezzeddine Zagrouba
e.zagrouba@gmail.com

[1] LIMTIC laboratory, Higher Institute of Computer Science, University of Tunis El-Manar, 2, Rue Abou Raihan El Bayrouni, Ariana 2080, Tunisia

[2] LISIC laboratory, University of the Littoral Opal Coast (ULCO), 50, Rue Ferdinand Buisson, Calais 62228, France

[3] Department of Computer Sciences, Prince Sattam Bin Abdulaziz University, Al-Aflaj 16733, Kingdom of Saudi Arabia

1 Introduction

Cultural heritage images present an important source of information, as it reflects the bond to the past, enriches the present, and informs the future. Transmitting these heritages to the next generations is an important issue on the name of humanity to protect the history and the identity of the nations [166]. With the rapid development of computer vision and multimedia technologies, a wide range of interesting applications have grown to analyze, understand and create visual artworks (i.e artistic images). A brief summary of computer vision applications is given in Table 1.

Among recent developments, the Generative Adversarial Networks (GAN) have succeed to generate novel paintings simulating a given artwork's style (e.g. Vincent van Gogh, Gustave Dore, Monet or Cezanne) [36, 56, 58, 65, 94, 151, 172, 176, 179]. This process is called image style transfer. As example, the work "Portrait of Edmond de Belamy" presented in [65], is the first Artificial Intelligence painting created using the GAN algorithm [68] which learned from a dataset of 15,000 portraits painted between the 14th to the 20th one. it is worth noting that most of the researches are conducted on automatic recognizing and classifying artworks

Table 1 Major computer vision applications for visual art analysis and understanding

Art history problems	Computer vision approaches	References
Style transfer	Generative Adversarial Networks (GAN) algorithms, Creative Adversarial Networks (CAN), AI-Creative Adversarial Network (AICAN).	[36, 56, 58, 65, 94, 140, 151, 176, 179]
Image classification	Hand-crafted image features extraction and classification, Deep Convolutional Neural Network (CNN).	[4, 24, 33, 35, 43, 46, 47, 74, 79, 86, 91, 100, 104, 131, 148–150, 155, 167, 169, 175, 181]
Forgery detection	Visual pattern extraction based Deep CNN for classification, Hand-crafted features with machine learning algorithms for classification.	[6, 20, 52, 53, 122]
Image retrieval	Feature extraction based Deep CNN for similarity matching, Template generation and matching.	[17, 44, 57, 90, 102, 106, 134, 135]
Object detection	Deep CNN, Hand-crafted feature with classifiers and regressors.	[2, 10, 41, 62, 66, 67, 77, 80, 83, 84, 117, 139, 140, 144, 159, 170]
Emotion classification/Analysis	Deep CNN, Specific colors extraction for Luscher test.	[1, 93, 107, 123, 152, 171]
Clustering	Feature extraction for clustering algorithm.	[7, 26, 69, 142]
Illumination analysis	Shape from Shading, Face recognition, Occluding contour estimation.	[82, 146, 147]
Aesthetics quality assessment	Computation of statistical properties for rating analysis, Deep feature extraction for aesthetic prediction.	[34, 71, 87, 132, 178]
3D reconstruction	Tactile Fine Art Printing, Linear statistical model([12]).	[108, 127, 154, 157]
Image captioning	Deep CNN, Neural encoder-decoder, Retrieval-based methods, object detectors	[30, 31, 97, 137]

based on artists [35, 86, 100, 129, 131, 155], styles [4, 46, 47, 91, 104, 175], genre [74, 167, 181], materials [104, 168, 169], years of creation [104], scenes [50], artistic media [129, 169] or even their associations [33, 148, 150]. Moreover, there are other applications that are interested in indexing and searching artwork databases. The Detection of forgeries in paintings has also gained increasing attention to help art historians in the authentication of copies or disputed paintings [6, 20, 52, 122]. Significant attention is also devoted to image retrieval from digitized artwork collections in order to query similar images [27, 28, 44, 57, 102, 106, 134–136]. Furthermore, automatic emotion recognition across auditory and visual modalities has received increasing attention in computer vision [76, 112]. So, several tools have been developed for emotion recognition from artworks [1, 93, 107, 123, 152, 171]. In fact, these tools facilitate the detection of evoked emotions related to the visual stimuli received from artworks, and possibly provide associated explanations. Another important direction is the clustering of artistic images that can be useful for visual link and knowledge discovery in artwork datasets [26, 69, 142, 165]. The idea of clustering is to arrange artworks in clusters (or groups). So the images in the same cluster are similar according to predefined criteria. Additionally, some computer vision methods [145–147] were proposed to analyse lighting and illumination in order to determine the position or the direction of lighting (i.e. illuminant) in the image. This can be used for determining the studio conditions once the painting was executed. Another application of developing computational methods in visual art is to assess aesthetic quality of paintings [34, 71, 132, 178]. It could be used to determine the age of some paintings. In addition to that, a growing interest has also been observed in the reconstruction of three-dimensional shapes of artworks [108, 127, 154, 157]. The segmentation problem is also addressed in visual art for several tasks, in particular artwork restoration [3] and color change tracking over centuries [133]. It is worth noting that, the list is not exhaustive and many other applications of computer vision methods for visual art could be found in literature, representing the broad scope of proposed solutions.

Object detection in artistic images, as one of the most interesting topics, plays important role as it can help art historians in performing several tasks such as portrait analysis [170], illuminant position estimation [10], facial expression dataset annotation [105], image captioning [97] or improving augmented reality experiences [140]. Although object detection is a classic topic in computer vision, it remains challenging in visual art due to the wide diversity of painting images compared to natural images that makes them difficult to understand or analyse, even by an expert. So, considerable research efforts have been devoted towards developing various methods for the detection of different types of objects in artistic images (e.g. bodies, faces, animals, musical notation and vehicles).

In recent years, advances in deep learning have led to significant improvements in different applications such as image processing [141, 163, 180], speech processing [109–111], and video processing [89, 115]. The deep learning is a sub-field of machine learning, which in turn is a sub-field of artificial intelligence (AI). It involves the use of artificial neural networks with multiple hidden layers to model and solve complex problems. In particular, Convolutional Neural Network (CNN) as a type of deep learning neural network, has created new opportunities in visual art research with automatic tools able to extract, analyze and understand information from artworks. However, training deep neural networks can be challenging, and requires large amounts of data.

While the growth of demand in computer vision and deep learning for visual art, few surveys were published in the field. Table 2 highlights related survey papers.

The first survey was the Stork's work [146] which presents the power of several computer techniques to process and understand paintings and drawings. Then, in the survey [146], the author proposes a categorization of computer vision methods used for automatic analysis of

Table 2 Recent related surveys

Main focus/ Topic	Reference
Computer methods for painting analysis	[146]
Computer Vision Algorithms for recognising objects in artwork and in photographs	[22]
Recent developments in computational aesthetics	[15]
A Survey on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia	[39]
Machine Learning for Cultural Heritage: A Survey	[49]
Computer Vision Applications for Art History	[51]
Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview	[29]
The use of AI technologies for understanding and creation of art	[32]

artworks. In [22], the authors synthesize current popular algorithms for object recognition in artwork and in photograph. The survey of Brachmann and Redies [15] provides an overview of recent developments and results in the field of computational aesthetics. The recent survey [39] focuses mainly on the evaluation and prediction of multimedia art quality by using deep learning. In [49], machine learning algorithms within cultural heritage has been a subject of extensive investigation. In fact, a detailed taxonomy of surveyed works according to supervision degree is presented in order to compare highlighted machine learning methods, their major uses and limitations. The recent survey by Cetinic et al. [32] reviewed artificial intelligence systems for analysis and production of artworks. They presented a detailed overview of artwork datasets and recent works that address a variety of tasks such as classification, object detection, similarity retrieval, multimodal representations, computational aesthetics, etc. The work in [51] proposes a short review of computer vision applications studying articles in art history journals from a computer science perspective. In fact, technical details are missing in their study. Finally, Castellano and Vessio [29] propose a general overview of deep learning methods for pattern extraction and recognition in visual art. Authors in [29] presented a detailed taxonomy of methods according to the task being solved namely, artwork attribute prediction, information retrieval, content generation, object recognition and detection.

Although several surveys regarding visual arts exist, a deep review of object detection methods in artistic images is still lacking. Object detection in visual arts is currently an area of great research interest. Motivated by the growing literature that has recently emerged on this topic, a review study is required to help interested researchers to quickly and effectively grasp the important concepts and key issues of object detection in artistic images. So this present review intends to provide the reader with a state-of-the-art and future research trends, which could help enter this research field and explore it in details. The contributions of this paper can be summarized as follows:

- We propose in this research review a taxonomy of existing methods, that is extensively and deeply carried out about object detectors for visual art data.
- We provide an overview of issues and challenges faced by object detection in artistic images.
- We provide a critical review of research findings and give insights for future research directions.

The rest of the paper is organized as follows. Section 2 presents the background of object detectors in artistic images. In this section, we firstly reviewed available artwork datasets and metrics used in this context. Then, we survey the basic principles of the recent works and we describe our proposed new taxonomy of object detection in artistic images. Finally, the various challenges are explained. Section 3 is devoted to provide critical summary of the recent literature and directions for future research. Finally, Section 4 concludes the survey.

2 Background of object detectors in artistic images

In this section, we firstly review the datasets and the evaluation metrics adopted in the context of object detection methods in artworks. Then, we introduce a new taxonomy of the studied methods according to different aspects namely, methodology of work, framework, supervision learning degree, depictive style and type of objects. Finally, we discuss the challenges of current studies and present possible solutions to overcome the encountered problems.

2.1 Main object detection artwork datasets

In object detection, a large number of artwork datasets have been collected and annotated to evaluate and to compare the performances of detectors in artistic images. These specific datasets now open new research perspectives in this particular research field. Table 3 summarizes characteristics of the well-known datasets used in the studied works on object detection in artistic images, provided in Table 5 with references. For each dataset, the style of the images, the subject classes, the number of images, the type of annotation which could be image level or object level, and the the number of object instances if the dataset contains object level annotations are specified. In fact, the image level annotation is the task of assigning labels (e.g the style or the artist of the painting) to the images. On the other side, the object instance annotation is the process of labeling objects in an image with rectangular bounding boxes. There are multiple formats of bounding boxes, namely COCO and PASCAL VOC. In fact, the COCO format represents the bounding box by the coordinates of its top-left corner along with its width and height. In the PASCAL VOC format, the bounding box is identified by the coordinates of its top-left corner and the coordinates of its bottom-right corner.

For automatic face detection, five datasets of digitized paintings are provided for several tasks. In order to make progress on the visual Turing test [61], researchers created a new-dataset MAFD-150 (Modern Art Face Detection) [99] for face detection containing digitized artworks from modern art that cover much diversity in style and artist ranging from 15th to 20th century. The Artistic-Faces dataset [5] is also created to evaluate facial landmark detections in artistic portraits. It could be used for creating a style signature for portrait artworks or geometry-aware portrait style transfer. The Artistic-Faces dataset includes artistic portrait of various art genres, styles (ranging from High Renaissance through Cubism to Comics) and 16 different artists, with a large variation in both geometry and texture. In fact, labels related to the artist name, artwork title, style, date and source, in addition to facial landmarks are provided. KaoKore and Kouhon Datasets are also created for automatic face detection from Japanese artworks to help art historians create a new facial expression dataset. The KaoKore dataset [153] is an open dataset of pre-modern Japanese artworks with annotated metadata for each face. It describes the social status in the pre-modern Japanese society, the genre and the orientation. Nevertheless, the Kouhon dataset is closed within

Table 3 Detailed list of artwork datasets used for object detection

Dataset	Artisitc Style	Subject matter	Total images	Total instances	Annotation
Artistic-Faces	High Renaissance through Cubism to Comics	Face	160	68 facial landmarks per image	Image level and Facial landmarks
Al-Qazvin	Islamic painting	10 categories including devils, angels and amphibian animals.	67	29	Object level (COCO Format)
Brueghel	Flemish(Bruegel family)	5 categories (carts, cows, windmills, rowboats and sailbaots)	1587	273	Object level (COCO Format)
CASPA paintings	Cartoons, Sketches, Paintings	8 categories: bear, bird, cat, cow, dog, elephant, giraffe, horse, sheep, and zebra	1391	2834	Object level (COCO Format)
Clipart1k	Clipart	20 categories including person, horse and plant	1K	3165	Object level (PASCAL VOC Format)
Comic2k	Comic	6 categories including bike, bird, cat, car, dog and person	2K	6389	Image and Object level (PASCAL VOC Format)
Dataset of cultural sites	Not available	16 objects including sculptures, paintings and books	75.3K	-	Image level
Getty	Drawings, Painting	>22 including child Jesus, Mary, fruit, horse, flower and face	7872	-	Image level
IconArt	Paintings	7 classes: Jesus Child, Saint Sebastian, Angel, Crucifixion, Mary, Nudity and Ruins	6K	3009	Image and object level (PASCAL VOC Format)
KaoKore	Pre-modern Japanese	Face	1470	8573	Image and Object level (COCO Format)

Table 3 continued

Dataset	Artisitc Style	Subject matter	Total images	Total instances	Annotation
Kotenseki	classical and ancient Japanese style	6 classes: person, floor, shoji, tree, roof and animal	3126	609631	Image and Object level (CCOCO Format)
Kouhon	Japanese(owned by Shojo-Kouji)	Face	346	-	Image level
MAFD-150	29 styles belonging to Modern art	Face	150	398	Object level (PASCAL VOC Format)
MMSD	Medieval music	Phylactery, Folio, Book, Altar et Lectern	Not available	693	Object level
Paintings	Different styles (e.g. photo-realistic, abstract and impressionism)	10 classes: Aeroplane, bird, boat, chair, cow, table, dog, horse, sheep and train	8629	-	Image level
PeopleArt	Photo, cartoon and 41 styles from [160]	People	4.6K	$> 4.6K$	Object level (PASCAL VOC Format)
PhotoArt-50	Photo, oil painting, drawing, cartoon, stick-figures, etc.	50 classes including person, horse, car, bike, giraffe and face.	≈ 100 for each class	$> 5K$	Object level (COCO Format)
Picasso	Cubism	People	218	-	Image level
Tenebrism	Tenebism	Face	409	1159	Object level (PASCAL VOC Format)
Watercolor2k	Watercolor	6 classes: bike, bird, cat, car, dog and person	2K	3315	Image and Object level (PASCAL VOC Format)
WikiArt	27 styles including Abstract, Baroque and Byzantine	Landscape, Interior, face, person, etc.	>80K	-	Image level

the research group [105], owned by Shojo-Kouji and digitized by Yugyoji Museum. The artworks of Kouhon were collected from picture scrolls in the 14th century with 346 sheets of paper in 10 volumes. They are characterized by many faces appearing across the scroll. Similarly, researchers created the Tenebrism dataset [10] to estimate the illuminant position within a painting, and thereby to answer technical questions. This dataset includes artworks in the Tenebrism style of the 17th century. It is worth noting that Tenebrism paintings are challenging for face detection, since they are characterized by violent contrasts of light and dark, and they exhibit large variation in viewpoint, pose and occlusion.

In addition to that, some visual art datasets are created to assess the domain shift problem (i.e. applying natural images-trained detectors to paintings) for different classes in different depictive styles (photographs, drawings, paintings etc.) [40, 83]. As an example, the Paintings dataset [116], a subset of the 'Art UK' dataset is mentioned. It contains oil paintings of different styles (e.g photo-realistic,abstract and impressionist) and eras (18th, 19th and mid-20th century) annotated for popular PASCAL VOC categories. Similarly, the Getty dataset, provided by the Getty Research Institute, used in [83], contains 7872 images among which only 156 images are object-level annotated. IconArt dataset in [78], is also an artwork dataset used for object detection in several works [66, 67, 83], with challenging domain shifts . It includes painting images from Wikicommons [161] that hold instances from iconographic categories (e.g. Jesus Child, Saint Sebastian), ranging from the 11th to the 20th century.

For studying the cross-depiction object detection problem (i.e. detect objects regardless the depictive style), many cross-domain datasets (i.e. a collection of images across different domains, photographic images and artworks, containing the same target object classes) are produced such as PhotoArt-50 [121] and PeopleArt [120] datasets. The PhotoArt-50 consists of 50 object categories. Each category contains around 100 images with different instances. Approximately half of the images in each class are artworks covering a wide gamut of style. The other imaging are photographic images. The PeopleArt dataset is made of photos, cartoons and artistic images from 41 different movements. It has the single class people. This dataset is particularly challenging due to the high variability in styles and depiction techniques: from Picasso's cubism to Disney's Sleeping Beauty. Other work [62] has sought to design algorithms that mimic the human visual system for object detection outside the realm of natural images. So the Cubist paintings of Picasso dataset are examples used to depict objects that are not normally seen in nature.

In addition to that, some artwork datasets such as Kotenseki [88] and Al-Qazvin [2] are created to develop content-based image retrieval systems based on automatic object detection [139]. The Al-Qazvin dataset includes Islamic paintings of the 12th century where most of the objects in these paintings are imaginary objects. The Kotenseki is a collection of classical and ancient Japanese paintings. They are comprised of both colored and black-and-white drawings of different objects. In [139], only 1104 images were selected to be used. The paintings in Kotenseki dataset are annotated in COCO format for object detection. In addition to object instance annotation, additional informations are provided for Kotenseki images, such as image path, tagname, page of Kotenseki book, title, and tagid. In literature, the WikiArt dataset [160] is also investigated. It represents one of largest online available dataset of digitized paintings from different artists, styles and genres. However, it is not annotated for object detection.

The Brueghel dataset [19] is used for the authentication of artworks based object detection. This dataset includes artworks made in different media (e.g. oil, watercolor, chalk) and on different materials (e.g. paper, panel copper), representing a wide variety of scenes (e.g. landscape, religious, still life).

Watercolor2k, Clipart1k and Comic2k datasets are also used in few studied works [66, 67, 80]. Watercolor2k is composed of two thousand watercolor paintings made by artists in the 20th century. It contains object instances from 6 categories (i.e. bike, bird, cat, car, dog and person) in common with the PASCAL VOC dataset. It is splitted halfly into training and testing sets. The Clipart1k includes thousand clipart images containing object instances belonging to the same 20 categories of PASCAL VOC. It is splitted into training and test sets. Each includes 500 images. The Comic2k contains two thousand comic images sharing 6 categories with Pascal VOC and splitted halfly into training and testing sets. The CASPApaintings dataset used in [66] is the painting subset of the CASPA (Cartoons, Sketches, Paintings) dataset [25] which covers animal COCO categories.

Recently, the Dataset of cultural sites is created to recognize artworks in cultural sites using images acquired from the visitor [118]. This can be used to improve augmented reality experiences. This new dataset consists of synthetic and real images of 16 artworks from different points of view.

More Recently, Medieval Musicological Studies Dataset (MMSD) is created and annotated with medieval artworks (paintings or drawings). It holds musical scenes of persons in solo or in group-singing situations, whether accompanied or not by musical instruments. MMSD is used in [77] to detect signing performances in order to better understand the physical postures of singers, their relationship, and their location inside the building.

Even though museums and cultural institutions constantly make artwork collections available, the choice of the dataset remains application specific, and the most important factors to consider could be the target artistic styles or/and classes, the artwork period and the available annotations (i.e. instance level or image level). For example, it is possible to find a dataset covering the target classes or target style [67, 83], but only offers image level annotations which are not helpful for object detection purpose. In this case, considerable effort should be made to annotate bounding boxes around objects for its images, which sequentially requires a lot of times.

2.2 Evaluation metrics

The results of object detections are bounding boxes around objects with confidence score representing the class probability. Several metrics are provided to evaluate object detection performances in artworks. The four commonly used criteria are Precision, Recall, F-measure (or F1-score) and Average Precision (AP). To compute these metrics, the concept of Intersection Over Union (IoU) is used to determine if a detection result is True Positive (TP), False Positive (FP) or False Negative (FN). In fact, the IoU of two bounding boxes, the bounding box for the ground truth and the predicted bounding box, is the intersection area divided by the union area. A detection result is considered TP if IoU is greater than a predefined threshold, otherwise it is considered as FP. When a ground truth bounding box is present in the image and the object detection model failed to detect it, then it will be considered as FN. True Negative (TN) is every part of the image (i.e. background) where no object is detected. TN is not useful for object detection. In Table 4, we describe the metrics aforementioned above whose detection results in artworks are available in mentioned papers.

The precision recall curve (PR-curves) of classes has been drawn in several works [10, 40, 62, 84, 164] to show the tradeoff between precision and recall for different thresholds. Indeed, when the precision stays high as its recall increases, the object detector is considered good. So, high area under the curve tends to indicate both high precision and high recall. Hence, the Average Precision AP metric, that consists to estimate the area under the precision

Table 4 Metrics used for object detection in artworks

Metric	Description	Formula	Papers
Precision (P)	The fraction of all positive predictions that are true positives.	$P = \frac{TruePositive(TP)}{TruePositive(TP)+FalsePositive(FP)}$	[2, 40, 62]
Recall (R)	The fraction of all actual positives that are predicted positive.	$R = \frac{TruePositive(TP)}{TruePositive(TP)+FalseNegative(FN)}$	[2, 62, 66, 105]
F-measure (F)	The harmonic mean of precision and recall.	$F = 2 \times \frac{Precision \times Recall}{Precision+Recall}$	[2, 62, 158]
Average Precision (AP)	The average precision at a set of eleven equally spaced recall levels [0,0.1,...,0.9,1]. The precision at each recall level r is interpolated by taking the maximum precision $P_{interp}(r)$ whose corresponding recall value is greater than r.	$AP = \frac{1}{11} \sum_{r \in \{0,0.1,\ldots,1\}} P_{interp}(r)$ where $P_{interp}(r) = \max_{\widetilde{r}:\widetilde{r} \geq r} P(\widetilde{r})$ and $P(\widetilde{r})$ is the measured precision at recall $\widetilde{r}$	[10, 40, 62, 66, 67, 77, 80, 81, 83, 84, 105, 114, 117, 118, 139, 140, 159, 164]
Normalised Mean Error (NME)	The mean Euclidean distance between estimated landmarks and ground truth landmarks divided by the number of points and the normalized distance.	$NME = \frac{1}{n} \sum_{i=1}^{n} \frac{\left\|x_i - x_i^*\right\|_2}{d}$ d = reference distance for normalization (e.g. interocular distance or minimal bounding box size), n= number of facial landmarks, x_i and x_i^* are ground-truth points and model prediction points, respectively.	[170]

recall curve, was considered the popular metric for evaluating object detectors. To do that, the 11-point interpolation approach is commonly used, which averages the maximum precision values at a set of 11 equally spaced recall levels [0, 0.1, 0.2,...,1]. The mAP is computed by taking the average precision over all object categories. In addition to that, the Normalised Mean Error (NME) metric was also adopted in [170] to evaluate facial landmark detection in artistic portraits. Finally, training and testing times are computed for some object detectors in artistic images. Thus, we termed ATrT for the average training times and ATeT for average testing times.

It is also worth noting that some computer vision researchers give attention to the learning process evaluation to identify potential issues such as underfitting, overfitting and convergence problems. For example, in [10], authors give the accuracy and loss curves of the proposed models to control overfitting.

While the training parameters of deep models are adapted during the training phase, the values of the hyper-parameters have to be fixed before the learning phase and nevertheless make an important impact on model's accuracy. The hyper-parameters tuned in the studied papers are mainly the batch size, epoch, learning rate, weight decay and momentum. Indeed, epoch is the number of repetitions of the learning process. The batch size is the number of samples processed by the deep model in one epoch before updating the model parameters. In addition to that, the learning rate is a hyper-parameter that controls the gradient descent algorithm to find the minimum value of the error function for adjusting the weights of the network. The weight decay (commonly called L2 regularization) is a regularization parameter that controls the trade-off between having a powerful model and over-fitting the model. Finally, the momentum controls the rapidity of the learning process.

Setting optimal hyper-parameters is one of the main challenge during the training of deep models. In several works [10, 62, 77, 80], the authors chose to tune the hyper-parameters of their models using the default values of the original architectures or the same setting used in the original papers. Others researchers proposed the grid-search cross validation method for hyper-parameters tuning [77] by choosing the values that give the most optimal results. For instance, the values of hyper-parameters that minimise the loss function are selected in [67]. A 3-fold cross validation is also performed in [66] for determining the main hyper-parameters of the SVM. Besides, in [140], authors proposed a Bayesian optimization strategy [18] to efficiently find a good combination of hyper-parameters. In [66, 84, 105], an analysis of the hyper-parameters of the proposed model on the detection performances was performed. For instance, it was demonstrated in [105] how the choice of some values for the intersection over union (IoU) that represents an important metric of the proposed method, enables a larger improvement on precision than on recall.

2.3 New taxonomy of object detectors in visual art

As previously discussed (Section 1), a large number of methods have been developed for object detection in artistic images in the last decade. Figure 1 shows the trend of this topic in terms of number and publication year of the reviewed papers. It can be seen an increasing number of publications demonstrating the growing interest of the research community on the topic.

In order to understand the basic principles of the works investigating object detection in atristic images, a schematic overview is provided in Table 5. The best mean average precision scores (AP_{50}) are given with at least 50% intersection over union (IoU) overlap. For each studied work, we present used artwork dataset, used object detection frameworks, obtained

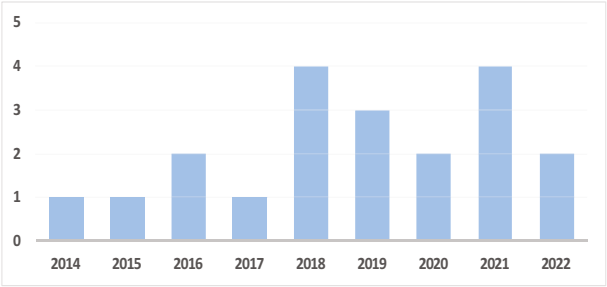

Fig. 1 Count of papers published in object detection from artworks (above) and year of publication (below)

performances and highlights of the work. This schematic overview serves as a comprehensive guide, to choose framework, dataset, metric and approach, ultimately leading to successful solutions for object detection. Figure 2 illustrates examples of object detection results on different used datasets.

Based on the previous meticulous review of papers published for object detection in artistic modalities, five major categories are identified. In fact, the reviewed methods can be classified according to their supervision learning degree into supervised, weakly supervised and unsupervised methods. They could also be categorized according to their methodology of work into two classes: deep-learning based methods and traditional computer vision methods. An especially important separation of the object detection methods in artistic modalities could also be done based on the category of object detector framework adopted in the work: one-stage or two-stages. Additionally, another aspect of dividing the methods is based on the target object to detect: face, person, animal, etc. At last, one crucial aspect that should be taken into consideration is the style of artistic images in which target objects are depicted. Figure 3 presents a mind-map highlighting the different aspects that can be used to classify an object detection method in artistic modality. These aspects should be taken into consideration by practitioners with respect to the problems they want to solve. It is worth noting that this classification of existing methods is not absolute, it depends on the characteristics of the methods. So the existing methods can be classified into many overlapping or non-overlapping categories simultaneously.

2.3.1 Methodology: deep learning vs traditional computer vision methods

Existing methods could be classified into two categories according to the adopted methodology by using classical computer vision techniques or deep learning methods (CNNs). In fact, the classical object detection methods consist of traditional hand-crafted features extraction and classification. The hand-crafted features such as edge detection and texture analysis, are designed manually by researchers [70]. On the other hand, the CNN is designed to automatically learn high-level deep feature representations for prediction [54, 101, 113]. One of the main advantages of CNNs is their ability to learn and recognize complex patterns in images without the need for explicit feature engineering. Recently, several techniques are proposed to select optimal features in a CNN to enhance its performance. For example, in [112], authors proposed a two-stream CNN to extract spatial-spectral features which are then fused and fed to the iterative neighborhood component analysis to select the most discriminative optimal features for the final prediction.

Table 5 Brief overview of various research works on object detection in painting images

Year/Ref	Artwork Dataset	Used framework	Highlights	Performances
2022/[77]	MMSD	YOLOv4 (m/s), Faster RCNN, Mask RCNN, SWIN-T	A dataset collection and annotation, Bi-training object detection model based finetuning	AP_{50}=83.62%
2022/[66]	PeopleArt, IconArt, Watercolor2k, Clipart1k, Comic2k and CASPA paintings	Faster RCNN	An extension of [67], Transfer learning of pretrained CNN, A polyhedral separation for classification	AP_{50}=60%, R=94%
2021/[84]	PeopleArt	Faster-RCNN	AdaIn Style transfer for artistically-styled images generation, Finetuning the model	AP_{50}=68%
2021/[10]	Tenebrism	Faster RCNN	Collection and annotation of new dataset, Finetuning pretrained Faster RCNN, Application of several data augmentation techniques (rotation, flipping).	AP_{50}=86.51%
2021/[83]	Getty, IconArt	Faster-RCNN, NudeNet, YOLOFace, YOLO9000, ResNe(X)t and WBF	Style transfer for artwork-like images generation, Fine-tuning Faster RCNN.	AP_{50}=82%
2021/[118]	Dataset of cultural site [117]	DA-RetinaNet, RetinaNet, Faster RCNN, DA-Faster-RCNN, Strong-Weak	New Dataset creation, Proposition of DA-RetinaNet	AP_{50}=55.54%
2020/[81]	Brueghel, PeopleArt	YOLO	Neural style transfer, Data augmentation, Training object detector.	AP_{50}=40.6%
2020/[105]	KaoKore, Kouhon	SSD300, Faster RCNN, Cascade RCNN	Image patching, object detection in each patch, Non Maximum-Suppression	AP_{50}=82.9%,R=91.1%
2019/[158]	MAFD-150	Four face detectors including Viola & Jones	New artwork dataset creation, performance evaluation of face detectors, challenges of face detection in artistic images	F<35%
2019/[170]	Artistic-Faces	Facial landmark detection framework [177]	Artistic data augmentation, Heat-Maps Network for landmark detection, Landmarks correction using a pre-trained point distribution model (PDM)	NME=2.01

Table 5 continued

Year/Ref	Artwork Dataset	Used framework	Highlights	Performances
2019/[67]	IconArt, PeopleArt, Watercolor2k	Faster RCNN , MI-max framework	New database creation, Multiple instance learning classification coupled with Faster RCNN	AP_{50}=85.2%
2018/[2]	Al-Qazvin	SURF detector, Hash Function	Improved SURF algorithm for features extraction, Hash function to index features in database	F=81%, P=87%, R= 77%, ATeT = 1.4 sec to 4.3 sec
2018/[80]	Clipart1k, Watercolor2k, Comic2k	Faster RCNN, SSD300, YOLO	New dataset construction, CycleGAN for generation of artistic images, Finetuning pretrained fully supervised detector	AP_{50}=86.3%
2018/[140]	Paintings, WikiArt	VGG-19 architecture	Artistic Style Transfer, Transfer learning, Training CNN models	AP_{50}=82%
2018/[114]	Tenebrism	AlexNet	Finetuning pre-trained model, Data augmentation.	AP_{50}=25.89%
2017/[139]	Kotenseki	Faster RCNN	Fine-tuning Faster RCNN, L2 Normalization, Data augmentation.	AP_{50}=98.57%
2016/[159]	PeopleArt, Picasso	Fast RCNN, DPM, YOLO	Collection and annotation of a new dataset, Fine-tuning pre-trained model.	AP_{50}=59%
2016/[40]	Paintings	Faster RCNN	Data augmentation, Application of a pre-trained Faster RCNN.	P=100%, AP_{50}=59%
2015/[62]	Picasso	Dalal and Triggs (D&T), DPM, Poselets, R-CNN	Comparison of object detectors	P=44.4%, R=48.6%, F=45.8%, AP_{50}=37.8%
2014/[164]	PhotoArt-50	DPM	Model visual objects with graphs	AP_{50}=89.1%, ATrT = 4 to 5 min, ATeT= 4.5 to 5 min

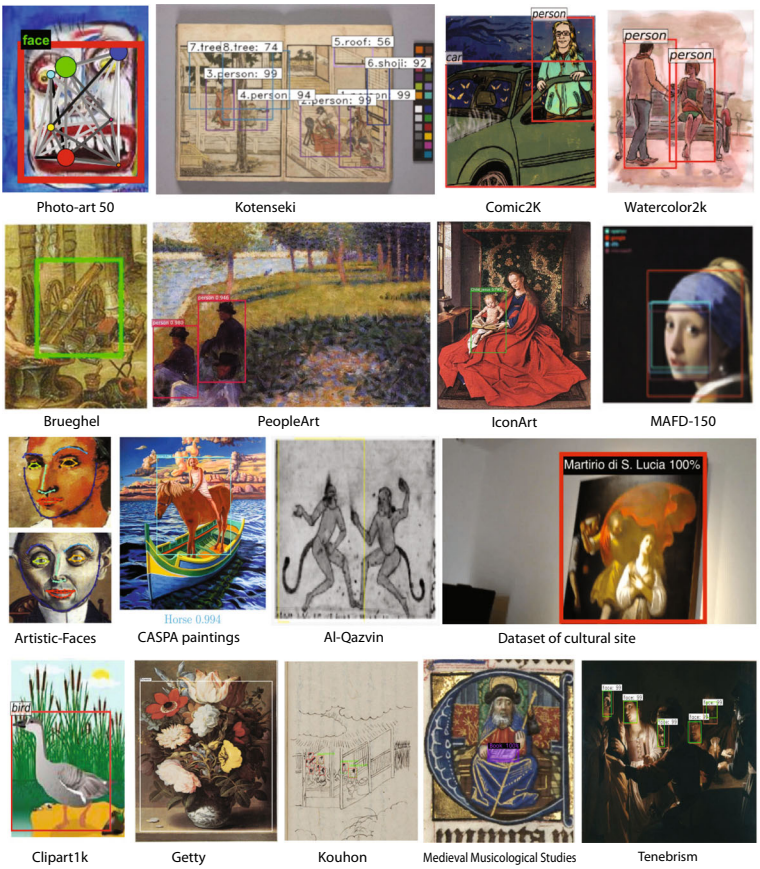

Fig. 2 Examples of available object detection results on different used artwork datasets. The results are taken from papers cited in Table 5

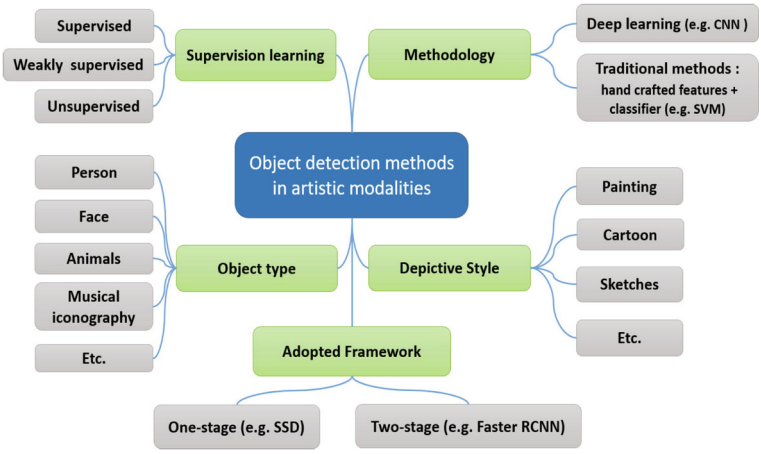

Fig. 3 Taxonomy mind-map of object detection methods in artistic images

Before the success of deep convolutional neural networks in artistic images, a variety of traditional methods for object detection based on hand-crafted features have been investigated. Wu et al. [164] presented a graph modelling method for object description with multi-labeled nodes and discriminative weights on arcs and nodes to encode relative importance. The visual class models are learned from different depiction styles. For the object detection, a matching algorithm is used to find the graph in the given artistic image that best matches the given visual class model. In [62], the Deformable Part-based Models (DPM) detection method [48] has demonstrated the best performances on person detection in paintings compared to other object detectors. The DPM method extracts oriented gradient (HOG) features [42] at different resolutions since they are invariant to changes in lighting and small deformations, then localizes objects by sweeping the features with support vector machine (SVM) [72] classifiers. Similarly, in the same work [62], authors explored the Poselet detector which leads to good performances on person detection in cubist paintings. The poselet detector [14] is a HOG part-based linear support vector machine classifier. Indeed, Poselets are parts of the human pose trained under specific viewpoint and each poselet detection casts a vote towards the full bounding box of the person. Although the poselet detector is robust over a wide range of human poses but it has a high computational demand. The template-based Dalal and Triggs (D&T) method [42] is also investigated for person detection in [62], but provides inferior performances compared to DPM [48] and Poselets [14]. The (D&T) method describes objects by histograms of orientations of local edge gradients binned over a dense image grid (HOG) and uses the combined feature vector in a conventional SVM for detection. The algorithm of Viola-Jones Haar face detector is used in [158] to detect faces on varying art categories (e.g. Impressionism, Fauvism, Pop art, etc.) but provided very low face detection performances (F1-score = 22%). The Viola-Jones algorithm computes at first the integral image to quickly extract Haar-like features. The AdaBoost algorithm is then used to both select important features and train the classifier. In the literature, the Viola-Jones is considered as one of the best in detection effectiveness/speed of work ratio [119, 162]. Recently, in [2], the authors proposed to adopt a modified version of the SURF (Speeded-Up Robust Features) algorithm to construct a set of description features for every paintings images, mainly scale and orientations features. Then, a matching score between the desired object and the painting images is used to complete the detection process. More recently, in [77], authors have tested training a model based on histogram of oriented gradients, followed by a support vector machine classifier, to detect signing representations in medieval artworks. The model gives very low precision since it confused many of the other parts of the image as an object, dropping the F1-score to 23.9%.

Recently, deep learning have been intensively employed for detecting objects in artistic images with highly promising results. In the reviewed papers, there are some techniques used to enhance the performance of CNN models such as transfer learning, fine-tuning, data augmentation and style transfer.

Fine-tuning has been widely adopted in transferring learned information from one domain (e.g natural image) to another domain (e.g. painting). This technique involves training a pre-trained CNN on a new smaller dataset. By freezing the lower layers of the CNN and only training the higher layers, the model can learn new feature representations specific to the new dataset. In a very recent work [77], authors proposed a bi-training technique for performing object detection in medieval artworks. In this work, authors classified object classes in base and novel classes. Base classes are the samples highly represented in the dataset and novel classes are the less represented ones. The first step is to fine-tune the object proposal model (not the classifier) on the whole dataset for potential region proposals. In the second step, authors fine-tune the classifier and bounding box regressor with a particular configuration (i.e.

specific weights and learning rate) forcing the model to perform better with the novel classes. In [10], authors fine-tune the Faster RCNN [126] pre-trained on a dataset of photograph faces (AFLW) in different ways. Firstly, face detection improvements are observed by retraining only the higher layers of the model on painting images. Then, authors retrain all layers on the target artwork dataset to have a model that achieves high face detection performances. Another work in [159] uses a model pre-trained on ImageNet and fine-tunes it on People-Art dataset with different settings in order to choose the one that maximizes the performances. Similarly, authors in [139] fine-tune pre-trained model with different configurations (e.g freezing all or some convolutional layers layers, no fine-tuning, training all convolutional layers, etc.) in order to obtain the optimal results. So we remark that transfer learning via fine-tuning has been widely employed by object detectors in artworks. It is worth noting that fine-tuning pre-trained models is better than training models from scratch. To do this, different settings for fine-tuning are investigated in the reviewed papers to determine the optimal configuration of the model.

Furthermore, it is worth noting that object detector models based deep learning need large scale datasets annotated with objects. So, in order to avoid the time consuming annotation task, many researchers proposed image style transfer to generate new images that look like artistic ones and will be used for fine-tuning pre-trained models. Image style transfer consists in combining the content of an input image I_c with the style of a reference image I_s to generate new image I that preserves the content of I_c but has the style of I_s. The images I_c and I_s are passed through a pre-trained convolutional neural network (CNN) to extract their content and style features. In fact, the content features capture the high-level features such as structure and positions, while the style features capture the textures, colors, and patterns. The output stylized image I is then optimized to minimize a total loss using gradient descent or other optimization algorithms. Neural style transfer methods are classified into two categories. The first category represents optimization-based methods that transfer the style by iteratively optimizing an image. The second category, called model-based neural methods, optimizes a generative model offline, and produces the stylized image with a single forward pass. Although image optimisation-based methods are able to yield impressive stylised images, they are computationally expensive. So model-based neural methods speed up the style transfer process.

The optimization-based style transfer method proposed by Gatys et al. [59] is employed in [81, 140, 170] to generate stylized images. In fact, in [81], the COCO dataset is used as a content image, and some images from Brueghel [19] and People-Art [120] datasets are used as a style images. In [140], authors propose the transfer of artistic style to the natural images in PASCAL VOC dataset. The authors in [170] investigate geometric and texture style transfer. The total loss function of the method proposed by Gatys et al. [59] is defined by Equation 1.

$$L_{total}(I_s, I_c, I) = \alpha L_{content}(I_c, I) + \beta L_{style}(I_s, I) \tag{1}$$

where α and β balance the style and content losses. The content loss is used to measure the distance (e.g squared euclidean distance) between the content features of the I_c and I. The style loss is used to measure the difference between the style of I_s and the style of I.

In [84], a model-based style transfer method, called AdaIn (Adaptive Instance Normalization), is used to generate stylized images. The AdaIn loss function involves normalizing the mean and standard deviation of the feature maps F of the content image I_c to match those of the style image I_s, as shown in the Equation 2.

$$AdaIn(F(I_s), F(I_s)) = \sigma(F(I_s))\left(\frac{F(I_c) - \mu F(I_c)}{\sigma F(I_c)}\right) + \mu F(I_s) \tag{2}$$

Another work [80] explored the use of CycleGAN [179] algorithm to generate clipart, watercolor and comic images from natural images (PASCAL VOC). CycleGAN is capable of learning the mapping between two different domains without the need for paired training data. So, two generators, G_{CS} and G_{SC}, are required to map images from the content domain C (e.g. natural images) to the style Domain S (e.g. clipart images). Two discriminators, D_C and D_S, are also required to distinguish between real and fake images. CycleGAN minimizes a loss function during training for style transfer, as shown in Equation 3. This loss function includes adversarial loss, cycle-consistency loss and identity loss. The adversarial loss is used to ensure that the generated images are similar to the real images. The cycle-consistency loss ensures that the generated images are consistent with the input images. It is computed by comparing the difference between the input image and the image generated after passing through both the generator networks. The identity loss helps to preserve the content of the original image while transferring the style.

$$L_{cycleGAN} = L_{identity} + L_{cycle-consistency} + L_{adv} \tag{3}$$

Finally, in [83], authors propose the use of four different functions offered by openCV library [16] for style transfer namely, edgePreservingFilter, pencilSketch, stylization and xphoto.oilPainting. To conclude, it has demonstrated in the several previous works that artistic style transfer enhance object detection performance in digitized fine-art. However, style transfer requires experimentation for adjusting parameters to achieves good results.

Data augmentation is considered as a traditional strategy to artificially increase the size and variability of the original dataset by applying certain transformations to the training artistic images deriving new examples. This can help to improve the generalization performance of the CNN. In fact, the transformations are generally geometric (e.g. flipping, cropping and rotation) or photometric (e.g. contrast changes and color perturbations), and they can be performed at training time (called online or on-the-fly augmentation) or before training (called offline augmentation). Numerous works performed data augmentation for learning the object detector models on artistic images in order to be invariant to these transformations. Consequently, data augmentation helps these models to obtain better accuracy by reducing overfitting and improving generalization. Nevertheless, the training phase will require significant computing time. In [139], researchers proposed several data augmentation techniques, namely shearing, zooming, shifting width and height, and flipping. The experimental results in [139] demonstrated that data augmentation during the training process could significantly improve the accuracy of the model. Authors in [40] proposed the use of four data augmentation techniques available in MatConvNet toolbox [156], mainly crop with different configurations and stretch. They showed that augmentation increases the mAP of the model and the stretch technique produces the highest performance. Recently, authors in [10] investigate the effects of online data augmentation on face detection performances by evaluating separately each technique used. It has demonstrated that data augmentation improve detection performances and the best results are obtained with rotation technique. In general, despite the object detection performance improvement using data augmentation, the choice of the best or appropriate technique remains critical. Thus, it is recommended to perform an in-depth investigation of the influence of data augmentation techniques on object detection performances for optimal choice.

2.3.2 Adopted framework

According to literature, the deep learning frameworks are the most used to detect objects in visual arts. The existing deep learning frameworks are roughly classified into two categories

based on their architectures: single-stage and two-stage detectors. The two-stage frameworks have complex architecture and are slow, but still provide more accurate results than one-stage frameworks. Nevertheless, although the one-stage object detectors have simple architecture and save computational time, they suffer accuracy loss, particularly for tiny/small objects which was a challenge for several object detectors in artistic images [67, 77, 105]. So, single stage frameworks are recommended for real time applications.

Figure 4 illustrates a schematic description of the two type of object detectors.

On one hand, the two-stage object detectors firstly proposes a set of regions of interests (ROIs) by selective search or using Regional Proposal Network (RPN). Then, a classification of each ROI, as well as its regression to the ground-truth locations are performed. The anchor boxes are predefined bounding boxes created at different sizes, aspect ratios, and locations to be used for the ROIs generation in RPN. On other hand, the single stage object detectors use a single feed-forward neural network to learn class probabilities and bounding box coordinates according to the calculation of the loss function. Both types of object detectors have a backbone composed of successive convolutional and pooling (or down-sampling) layers for deep features extraction. In the convolutional layer, the input image is convoluted with a set of learnable filters to extract features such as edges, corners, and textures. The pooling layer reduces the dimensionality of the output by downsampling the features. The ROI pooling layer is used to solve the problem of fixed size of feature maps required for the detection generator. It utilizes max-pooling to convert features inside valid ROIs (i.e. region proposal) into feature maps with fixed size which will be fed then to the fully connected layers for prediction. The purpose of the classification layer is to judge which class a region belongs to. For binary classification, the sigmoid function is used in the classification layer

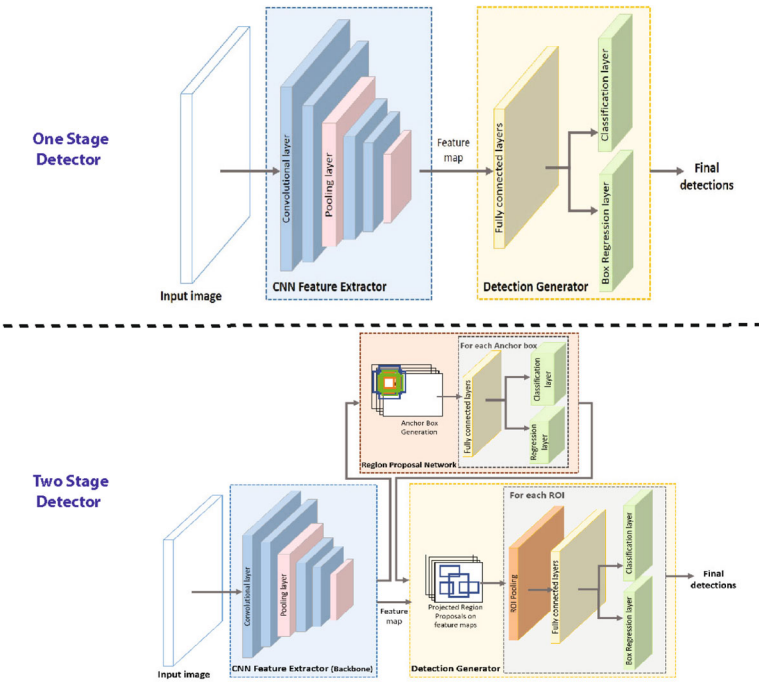

Fig. 4 Basic architectures of single stage and two-stage object detection frameworks

to produce a value between 0 and 1 that represents the class probability. For multi-class classification, the activation function is typically the softmax function, which produces a probability distribution over the possible output classes. The bounding box regression layer adjusts the positions of regions of interest (i.e. four bounding box coordinates) to obtain the final object detection result.

Many works in the studied literature have employed two-stage object detectors. The used frameworks are R-CNN (Region-based CNN) family networks which mainly include original RCNN, Fast RCNN, Faster RCNN, Cascade RCNN and Mask RCNN. For instance, the RCNN network proposed in [64] performs object detection by using a deep ConvNet to classify object proposals. RCNN was used in [62] for person detection in abstract paintings but it does not perform well compared to traditional methods. The obtained performance is justified by the fact that RCNN is not a part-based model and over-fits to the natural images. The Fast RCNN [63] which takes as inputs an entire image and a set of object proposals, learns to classify object proposals and refine their spatial locations. In [159], a finetuned Fast RCNN based VGG16 gives the best object detection performances on the People-Art Dataset (59% mAP) compared to other object detector models, namely YOLO [124] and DPM [48]. The Faster RCNN [126] improves Fast R-CNN by introducing the Region Proposal Networks (RPN). Faster RCNN has been widely used for object detection in artworks for its strong performances [10, 40, 66, 67, 77, 80, 83, 84, 105, 139]. Indeed, performance analysis were performed in several works for different architectures such as VGG16 and RES-152-COCO, as backbone network of Faster RCNN used for feature extraction. We observed that the residual network (ResNet) appears to be the best architecture for transfer learning by feature extractions [10, 67]. The two-stage model Cascade RCNN was investigated in [105] for face detection on the Kouhon dataset. Indeed, the Cascade RCNN is composed of a sequence of detectors trained with increasing IoU thresholds that address the limitations of Faster RCNN as mentioned in [21], to enable high quality object detection. Experimental results in [105] show that Cascade RCNN yields similar face detection results as Faster RCNN. Recently, the Mask RCNN framework is used in [77] and compared with YOLO and Faster RCNN for detection of objects (e.g. Lectern, Phylactery) in medieval singing images. The Mask RCNN, proposed in [73], adopts the same two-stage approach by adding a branch (in the second stage) in parallel with the existing branch for classification and bounding box regression, to predict a binary mask for each RoI. In [77], Mask RCNN achieves better performances in average precision than Faster RCNN, but it is less accurate than YOLOv4-s.

Other methods such as [77, 80, 83, 105, 117] explored one-stage object detectors frameworks. YOLO (You Only Look Once), proposed in [124] as a single stage object detector model, was investigated in several related works [77, 81, 83, 159] for object detection in artworks. This framework divides an image into S $\times$ S grids, and if the center coordinate of the ground truth of an object falls into a grid, the grid is responsible for detecting the object. YOLO is fast by design since it throws out the region proposal generation stage entirely. Several versions of YOLO are proposed to ensure a tradeoff between speed and accuracy. The more highly used in the studied works are YOLOv5 [174], YOLOv4-s [13], YOLOv4-m [13], YOLOv2 or YOLO9000 [125] and YOLOFace [173]. In [77], YOLOv4-s provides the best object detection performances on medieval singing dataset (MMSD) compared to Mask RCNN and YOLOv4-m. Experimental results in [83] show a varying performance of different YOLO models where each model is trained for a specific type of object (animals, structures, produce and person). The Animal category scored the best performance with 82% mAP. Similarly, YOLOFace, proposed in [37], is a pre-trained model used in [83] for the detection of the Face category. In fact, YOLOFace is based on the YOLOv3 architecture trained to predict faces. Furthermore, authors in [81] applied YOLOv5 pretrained on COCO

dataset for object detection in artworks to achieve 40.6% mAP on PeopleArt dataset. In [159], YOLO was also the best performing object detector model on the Picasso dataset (53% mAP) compared to DPM [48] and Fast RCNN [63] models which make it possible to believe that YOLO's design is more robust to abstract forms of art. Moreover, the SSD300 model was also used as one stage object detector in some research works. The SSD300 is a pretrained model of SSD (Single Shot MultiBox Detector) [95] using images with 300 x 300 size resolution. SSD predicts a fixed number of scores and bounding boxes which is more than that of YOLO, followed by a non-maximum suppression step to remove duplicate predictions pointing to the same object. In [80], a fine-tuned SSD300 outperforms the best-performing baselines in terms of mAP across all datasets (i.e. Clipart1k, Watercolor2k and Comic2k). In [105], SSD300 is trained on the KaoKore dataset for face detection on images from Kouhon dataset. The best obtained result of mAP was 72.7%. Finally, the single-stage object detector called RetinaNet [92] was used in the recent work [118] to detect artworks in cultural sites. Authors showed that RetinaNet is more robust than Faster RCNN to domain shift, in which synthetic images are used for training and real images are used for test.

More Recently, the new model SWIN-T [96] is used in [8] to detect objects in medieval singing images. This model belongs to the Visual Transformers (ViT) architecture family that is different from the CNN architecture. In comparison to CNN, ViT exhibits an extraordinary performance with few computational resources. The authors in [8] have concluded that SWIN-T model is very powerful, but they require more training data.

2.3.3 Supervision learning degree

The studied works on object detection for cultural heritage visual data can be classified into three categories of methods according to the supervision learning degree, namely supervised, weakly-supervised, and unsupervised learning.

In supervised learning, the model is trained on a labeled dataset, and during training, it learns to make predictions by minimizing a loss function using an optimization algorithm such as stochastic gradient descent (SGD). In object detection, the goal is to locate and classify objects in image. So the loss function for many object detection models consists of two components namely, localization loss and classification loss, weighted by hyperparameter that determine the relative importance of each loss (Equation 4). The localization loss measures the error between the predicted bounding boxes and the ground truth bounding boxes of the objects in the image. The classification loss measures the error between the predicted class probabilities and the ground truth class labels of the objects in the image.

$$Loss = Loss_{localization} + \lambda Loss_{classification} \tag{4}$$

Several works [8, 10, 84] propose supervised approaches for object detection in artistic images. In [8, 10], the cross-entropy loss function is used as a classification loss and the smooth L1 loss function is used as a regression loss in the used CNN models. [92] The supervised learning algorithms, Support Vector Machine (SVM) and structured support vector machine (SSVM), are also used in [42, 164].

In weakly supervised learning, the model is trained to make predictions despite the lack of complete and accurate labels for the training data. In [66, 67], the Multiple Instance Learning (MIL) approach is adopted to detect specific classes, such as Nudity, Mary, Jesus as a child

or the crucifixion of Jesus, that are not available in photographic images. In this approach, the artistic images are organized into bags, where each bag contains multiple instances. The labels are assigned to bags instead of instances. In [80], authors use Weakly Supervised Deep Detection Network (WSDDN) [11] and ContextLocNet [85] for training on artwork datasets, where only image level annotations are provided.

In unsupervised learning, the model is trained using unlabeled data. The goal of unsupervised learning is to find patterns and relationships in the data without any prior knowledge. Recently, in [118] authors investigate the use of unsupervised techniques for artwork detection in cultural sites images. This consists of combining feature alignment techniques based on adversarial learning [55] with the RetinaNet architecture [92]. So three discriminators with Gradient Reversal Layer are added to the architecture of RetinaNet to distinguich between real and fake data. The resulting network, called DA-RetinaNet, is then trained to minimize the supervised loss and the discriminator loss. The proposed framework DA-RetinaNet outperforms RetinaNet [92], Strong Weak [130] and DA-Faster RCNN [38].

2.3.4 Type of target object

We identify several types (or classes) of target objects that could be detected in artworks. A recent work has focused on detection of musical representation, namely book hold in the hands or placed on a lectern, unfolded phylactery and musical notation, from medieval artworks [8, 77]. Some other works in the literature, such as [105, 114, 144], address face detection in artwork. Other applications were focused on recognising persons [84, 103] or even specific ones, like Leonardo [154] and other artists [143]. In addition, other works were interested in detection of iconographic characters such as Mary, Saint-Sebastian, Jesus as a child or the crucifixion of Jesus [66, 83]. Animals were also successfully detected in paintings, such as horse, cock and cow [40, 83, 103]. Finally, efforts have been made in [40] to develop method for detection of objects in different classes (e.g. table, aeroplane and rose) from painting images.

To conclude, it is worth noting that the class of the object influence the object detection performance. For example, it is hard to predict some categories of objects, such as Landforms (e.g. Mountains and Hills) and Produce (e.g. Meat, Fish and Flower), than others. Moreover, we note that low object detection performances are obtained with IconArt dataset due to the semantic complication of the classes. On the other hand, the bike class has scored the best performances in many works, as cited in [67, 80].

2.3.5 Depictive style

Object detection for visual cultural heritage was explored across different depictive styles of paintings explained by the use of several different datasets. The explored styles used with the studied objects detectors include High Renaissance [158, 170], Post-impressionism [158], Cubism [62, 170], Tenebrism [10, 114], Comics and Caricatures [170], Expressionism [170] and numerous other styles [159]. Object detection were also applied to pre-modern Japanese artworks [105], Al-Qazvin Cosmography book [2], Japanese old books [139], paintings of medieval and early modern annunciations [67, 98] and Western artworks [158]. Using a style transfer model, such as CycleGAN [75], it was possible to transfer an artistic style of paintings to the content of real world images. However, although many style-transferred datasets are generated in literature [98, 133, 140, 170], irregular artifacts are observed in the style-transferred images and the expert opinion remains essential to validate the quality of generated images (i.e. painting-like images).

It should be noted that some styles of artworks require substantially more effort than others for object detection, like the Abstract art or the Tenbrism style. This is surely due to the specific use of shadows in these painting styles. In fact, it has been demonstrated in [164] how performances of human and methods degrade with increased painting abstractness.

2.4 Challenges

Painting images are generally considered more challenging to analyze and interpret compared to photo-realistic images. They are customarily much harder when it comes to detect objects. In fact, in addition to classical challenges of object detection in photographic images (e.g. different viewpoints, illuminations, resolutions and occlusions), paintings hold many particularities making their study a more complex task. These are mainly artistic creativity, complexity of scene and lack of large annotated dataset for certain styles.

Painting images are often seen as a form of creative expression. Unlike photo-realistic images, paintings are man-made representations of real-life objects, animals, peoples, scenes or also an inspiration of them as a result of creative imagination. They can incorporate abstract elements and stylized object representations, allowing the artist to express their emotions and ideas in a unique and subjective way that cannot be captured by a camera. Creativity can be expressed in terms of a particular or exaggerated composition of elements such as texture, form, shape, color, tone and line. In addition to these elements, several other artistic concepts such as movement, unity, variety, harmony, pattern, thickness of brush strokes, media (e.g. oil, watercolor) and painting materials (e.g. paper, copper) [45] are added. This leads to significant differences in geometry, texture and color of the artistically rendered objects reflecting time-specific dimensions (e.g. hairstyle, shape of faces, physical postures, clothes worn, etc.) and different artist intentions. Therefore, an interesting related issue posed by the wide variation in the visual appearance exhibited by the same object across different styles, is the high intra-class variations. For instance, in figure 5-(a), we show the representation of persons across different styles to exhibit the significant visual differences between instances of the same class, revealing the high intra-class variations. The figure 5-(b) illustrate also the wide variation in visual appearance exhibited by the horse across different depictive styles.

The high visual complexity of scenes (figure 5-(c)) appears to also pose a challenge for object detection in paintings. It was proven that it is difficult to detect objects in busy scenes characterized by small regions of interest, especially when the style of the images is very different from recent photos [77, 84]. The inherent complexity of the style can also be seen in the image surrounded by the green dashed bounding box in figure 5-(c). In fact, the extreme contrast between light and dark areas introduces an element of mystery and ambiguity bringing an element of drama [10, 114]. Moreover, the detection of abstract objects (like people in the pink dashed bounding box in fig. 5-(c)) represents also a complex task [84].

Another major challenge is the lack of large annotated data with high quality digitized paintings for training. Even though a large number of artwork datasets were continuously published, the available images remain content, technique, artist or style dependent. Various works have been proposed to overcome the lack of large-scale annotated paintings and hence make use of the promising deep neural networks as explained in Section 2.3.

The occlusion of depicted objects was also one of the critical issue yet to be solved for object detection in painting [2, 10, 40, 159]. Indeed, painting images often present overlapping, occluded or truncated objects (blue and green bounding boxes in figure 5-(e)). Generally, occlusion in object detection task occurs under three situations: self-occlusion,

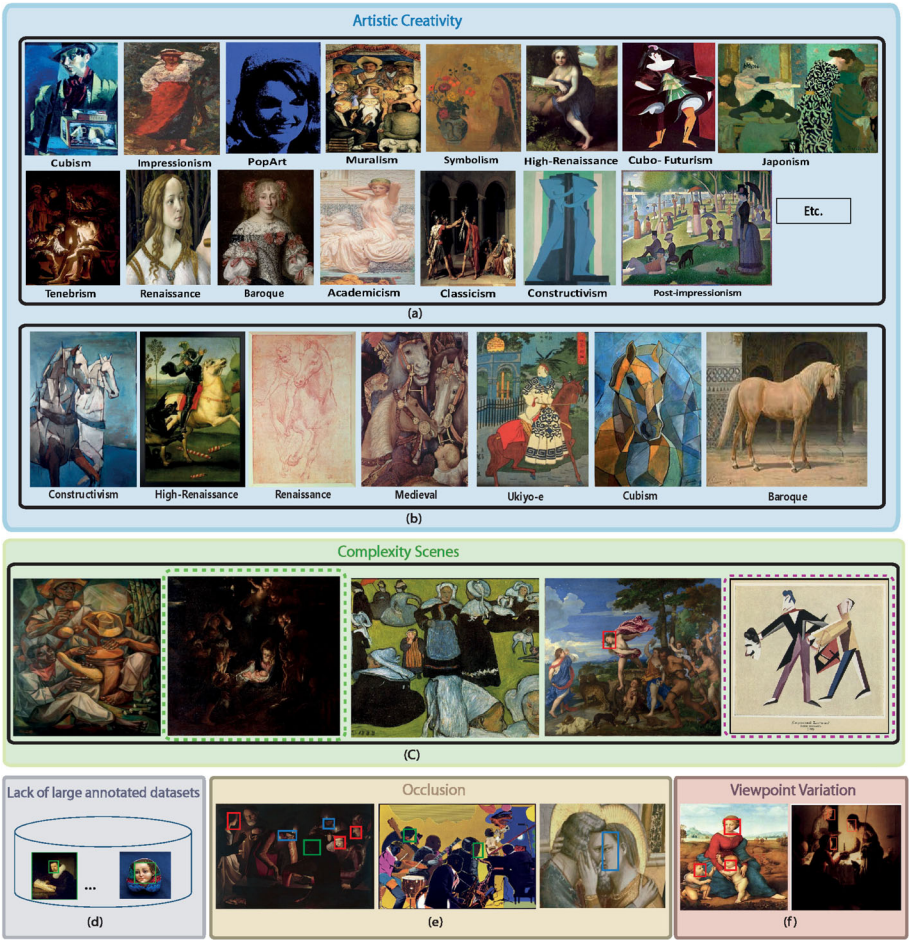

Fig. 5 Five different object detection Challenges in painting. (**a**) et (**b**) Difficulties to detect objects depicted by different artists: significant differences with photographic images and wide intra-class variation. (**c**) Detection of objects in complex scenes. (**d**) The scarcity of paintings annotated in the form of bounding boxes around objects. (**e**) The problem of occlusion in object detection. (f) Detection of objects viewed from different viewpoints

inter-object occlusions or occlusion by background. Self-occlusion arises when a part of the object occlude another whereas inter-object occlusion occurs when two objects occlude each other. The occlusion by the background occurs when a structure in the background occludes the object. A new occlusion situation is considered in artistic images when the painter depict an occlusion that is obviously intended. For example, in figure 5-e, several faces (surrounded by green-bounding boxes) appear occluded by the painter.

Finally, we notice that objects with a variety of viewpoints and sizes (red bounding boxes in figure 5-(c), (e) and (f)) make the object detection in artistic images a challenge to existing methods [2, 10, 40, 114].

3 Discussion and future directions

Considering the reviewed literature, it is clear that object detection plays an increasingly important role in visual art analysis. Despite the improvements that have been realised using deep learning neural networks for the other art history problems such as style, genre and artist recognition, these deep neural network models still face challenges when it comes to identifying objects in paintings. It was worth pointing out that painting images are the result of creativity. In fact, they can incorporate abstract elements and stylized representations of objects besides to a variety of brush stroke, shapes, textures and lighting conditions. In addition, they may depict complex and intricate scenes with multiple elements, characters, and details. These artistic characteristics, coupled with the small number of available images per style, have made the task of object detection more difficult than in photo-realistic images. This problem was stated in literature as the cross-depiction problem [22]. A primary question that arises in this context is why the human mind is capable of recognizing objects depicted in various artistic styles, whereas computers are unable to do so with equal ease. [23] have shown that using object representations (expressed within the BoW-SIFT feature space), there is a larger variance across photo and art domains than there is in each one alone. [60] have demonstrated that CNNs are surprisingly biased toward texture, using ImageNet dataset, whereas human rely heavily on shape similarity among objects for shape identification and recognition. As a result, researchers should explore new techniques in order to improve object detection accuracy in paintings. We will next discuss some of the promising research directions for object detection in paintings.

In our context, domain generalization aims at finding objects regardless they are photographed, painted, drawn, etc. Considering neural networks, constructing a generalized object detection models require a huge number of data in different depictive styles. Such goal remains largely untapped due to the limited availability of paintings in digital form or even the scarcity of certain styles besides to the labor-intensive process of manually annotating them. As stated throughout this review, existing object detectors for paintings are generally initiated from a model pre-trained on photographs since they are largely more abundant in comparison with paintings. They are then fine-tuned using transfer learning. Different transfer learning strategies and configurations were investigated for paintings towards studying the generalization abilities of models. For example, in [66] authors have presented a comparison between different feature extraction methods, loss functions and model' hyperparameters. [10] have also evaluated different backbone feature extraction architectures towards enhancing object detection results of Faster RCNN in the Tenerism Style. In addition, transfer learning approaches are generally associated with data augmentation techniques to encounter data scarcity in paintings. These latters include mainly standard augmentation techniques [10], generative adversarial networks [138] and neural style transfer [84]. It is worth noting that despite the improvement of detection results that has been achieved, crucial difficulties lie ahead. Particularly, certain artwork styles are more difficult than others for object detection. [67, 80] have shown that performance varies depending on the used artwork style. In fact, the more different is the target style from photograph, the more difficult become the object detection task. Particularly, It has been stated that difficulties increase also along with high level of painting abstractness [84, 164]. Generally, abstract paintings are highly deviated from visual reality. Recognizable elements that could be found within abstract painting are generally distorted or defined using composition of simplified shapes. Existing style transfer techniques [84, 164] can make it possible to shift the color and texture but not shapes. This drawback explains the performance degradation of object detection in abstract

paintings. Another complex style that were stated in literature is the Tenebrism style. The mysterious and dramatic illumination with violent contrasts of light and dark and the dominance of darkness make the task of face detection particularly difficult [10]. Besides to style complexity, another source of bias that should also been investigated is related to the training sources themselves. Both the photograph-based datasets (such as ImageNet or COCO dataset which were used for pre-training) and the painting datasets (which were used for transfer learning) hold biases related to local context and cultural disparities. For example, [84] have claimed that his model trained on StyleCOCO (obtained by style transfer from COCO dataset) performs better on Renaissance paintings than on a print from the Japanese shin-hanga movement. This was explained by the predominance of training images from North America and Europe. To this end, it becomes important to establish a detailed quantitative evaluation of object detection among different image styles, different objects and different datasets separately to pinpoint difficult ones. Another way of research that could be investigated is to design specific approaches for each difficult style. For instance, studying opportunities to deploy low-light enhancement techniques to improve object detection in the Tenebrist Style. On the other hand, another important line of research is to construct models that generalize well to novel test-domain. This could benefit from multi-task style and object identification architectures or also from recent stable diffusion image generators [128]. Finally, it is still important to continue developing new digital large-scale datasets for paintings.

The use of eXplainable AI (XAI) methods could also be efficient to produce explanations and reasons for decisions made by object detectors in paintings. A first attempt in this direction was established in [9]. In this work, the D-RISE method, a perturbation-based method that were designed for generating visual explanations of object detection, were used to examine saliency maps of Faster-RCNN face detection model in the Tenebrism style. A first investigation has proven that face detection in Tenebrism heavily rely on context unlike photograph-based face detection models. This is encouraging to persue a comprehensive study to understand principal modes of failures and to interpret the features learned by painting-based object detection models.

Another research axis that could be explored to improve object detection models for paintings is related to combining neural network models with semantic metadata information associated with them. Recently, a first try was realised by [103] were time-specific dimensions were associated with images to detect the most probable objects that fit the time period of the painting. In this direction, other contextual constraints such as cultural and geographical dimensions could be investigated for better object detection in paintings.

4 Conclusion

The continued and vigorous progress of computer vision has made a great advent in the development of methods for visual art analysis, particularly due to deep learning techniques. This review provides a comprehensive overview of computer vision applications in visual art. Among these applications, object detection in paintings represents a task that remains acute nowadays. In this paper, a review of the recent progress in this field was presented. We broadly categorized the methods into into five categories based on various aspects namely, methodology of work, framework, supervision degree, depictive style and type of objects. We list the highlights and challenges of each one, and we show that deep learning based methods can bring new tools for art history studies. This survey will enables new ways to

investigate visual art object detection and new strategies for doing quantitative research on art history and visual cultural heritage. It has been concluded that significant effort will be required in the future to tackle the challenges and open issues with object detection in visual art.

Data Availability All data analysed during this study are available at locations cited in the reference section.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Achlioptas P, Ovsjanikov M, Haydarov K, et al (2021) ArtEmis: Affective language for visual art. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 11569–11579. https://doi.org/10.1109/cvpr46437.2021.01140
2. Al-Yasiri D, Obaid AJ (2018) A new approach for object detection, recognition and retrieving in painting images. Journal of Advance Research in Dynamic and Control System 10(2):2345–2359
3. Amura A, Tonazzini A, Salerno E et al (2020) Color segmentation and neural networks for automatic graphic relief of the state of conservation of artworks. Cultura e Scienza del Colore-Color Culture and Science 12(02):07–15
4. Arora RS, Elgammal A (2012) Towards automated classification of fine-art painting style: A comparative study. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE, pp 3541–3544
5. Artistic-faces dataset (2019). https://faculty.runi.ac.il/arik/site/foa/artistic-faces-dataset.asp, Accessed: 2023-03-06
6. Bai Y, Guo Y, Wei J, et al (2020) Fake generated painting detection via frequency analysis. 2020 IEEE International Conference on Image Processing (ICIP) pp 1256–1260
7. Barnard K, Duygulu P, Forsyth D (2001) Clustering art. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, IEEE, pp II–II
8. Bekkouch IEI, Constantin ND, Eyharabide V, et al (2021) Adversarial domain adaptation for medieval instrument recognition. In: Lecture Notes in Networks and Systems. Springer International Publishing, pp 674–687. https://doi.org/10.1007/978-3-030-82196-8_50
9. Bengamra. S, Mzoughi. O, Bigand. A, et al (2023) Towards explainability in using deep learning for face detection in paintings. In: Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods - ICPRAM,, INSTICC. SciTePress, pp 832–841. https://doi.org/10.5220/0011670300003411
10. Bengamra S, Mzoughi. O, Bigand. A, et al (2021) New challenges of face detection in paintings based on deep learning. In: Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP,, INSTICC. SciTePress, pp 311–320. https://doi.org/10.5220/0010243703110320
11. Bilen H, Vedaldi A (2016) Weakly supervised deep detection networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2846–2854
12. Blanz V, Vetter T (1999) A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques SIGGRAPH '99, pp 187–194
13. Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934
14. Bourdev L, Malik J (2009) Poselets: Body part detectors trained using 3d human pose annotations. In: 2009 IEEE 12th International Conference on Computer Vision. IEEE, pp 1365–1372. https://doi.org/10.1109/iccv.2009.5459303
15. Brachmann A, Redies C (2017) Computational and experimental approaches to visual aesthetics. Front Comput Neurosci 11. https://doi.org/10.3389/fncom.2017.00102
16. Bradski G (2000) The opencv library. Dr Dobb's Journal: Software Tools for the Professional Programmer 25(11):120–123

17. Bredow T, Alder N, Büßemeyer M (2021) Image retrieval. In: Deep learning for computer vision in the art domain: proceedings of the master seminar on practical introduction to deep learning for computer vision, HPI WS 20/21, Universitätsverlag Potsdam, p 59
18. Brochu E, Cora VM, De Freitas N (2010) A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599
19. Brueghel dataset (2019). https://imagine.enpc.fr/~shenx/ArtMiner/, Accessed: 2023-03-06
20. Buchana P, Cazan I, Diaz-Granados M, et al (2016) Simultaneous forgery identification and localization in paintings using advanced correlation filters. 2016 IEEE International Conference on Image Processing (ICIP) pp 146–150
21. Cai Z, Vasconcelos N (2018) Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6154–6162
22. Cai H, Wu Q, Corradi T, et al (2015a) The cross-depiction problem: Computer vision algorithms for recognising objects in artwork and in photographs. arXiv preprint arXiv:1505.00110
23. Cai H, Wu Q, Hall P (2015b) Beyond photo-domain object recognition: Benchmarks for the cross-depiction problem. In: Proceedings of the IEEE international conference on computer vision workshops, pp 1–6. https://doi.org/10.1109/iccvw.2015.19
24. Carneiro G, da Silva NP, Bue AD, et al (2012) Artistic image classification: An analysis on the PRINTART database. In: Computer Vision – ECCV 2012. Springer Berlin Heidelberg, pp 143–157. https://doi.org/10.1007/978-3-642-33765-9_11
25. Caspa dataset (2018). https://people.cs.pitt.edu/~chris/artistic_objects/, Accessed: 2023-03-08
26. Castellano G, Vessio G (2022) A deep learning approach to clustering visual arts. Int J Comput Vision 130(11):2590–2605
27. Castellano G, Lella E, Vessio G (2021) Visual link retrieval and knowledge discovery in painting datasets. Multimedia Tools and Applications 80(5):6599–6616
28. Castellano G, Vessio G (2020) Towards a tool for visual link retrieval and knowledge discovery in painting datasets. In: Italian research conference on digital libraries, Springer, pp 105–110
29. Castellano G, Vessio G (2021) A brief overview of deep learning approaches to pattern extraction and recognition in paintings and drawings. In: International Conference on Pattern Recognition, Springer, pp 487–501
30. Cetinic E (2021a) Iconographic image captioning for artworks. In: International Conference on Pattern Recognition, Springer, pp 502–516
31. Cetinic E (2021b) Towards generating and evaluating iconographic image captions of artworks. Journal of Imaging 7(8):123
32. Cetinic E, She J (2022) Understanding and creating art with AI: Review and outlook. ACM Trans Multimed Comput Commun Appl 18(2):1–22. https://doi.org/10.1145/3475799
33. Cetinic E, Lipic T, Grgic S (2018) Fine-tuning convolutional neural networks for fine art classification. Expert Syst Appl 114:107–118. https://doi.org/10.1016/j.eswa.2018.07.026
34. Cetinic E, Lipic T, Grgic S (2019) A deep learning perspective on beauty, sentiment, and remembrance of art. IEEE Access 7:73694–73710. https://doi.org/10.1109/access.2019.2921101
35. Cetinic E, Grgic S (2013) Automated painter recognition based on image feature extraction. In: Proceedings ELMAR-2013, IEEE, pp 19–22
36. Chen X, Xu C, Yang X et al (2019) Gated-gan: Adversarial gated networks for multi-collection style transfer. IEEE Trans Image Process 28:546–560
37. Chen W, Huang H, Peng S et al (2021) Yolo-face: a real-time face detector. Vis Comput 37:805–813
38. Chen Y, Li W, Sakaridis C, et al (2018) Domain adaptive faster r-cnn for object detection in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3339–3348
39. Chu WT, Motomura H, Tsumura N et al (2019) [invited papers] a survey on multimedia artworks analysis and attractiveness computing in multimedia. ITE Transactions on Media Technology and Applications 7(2):60–67
40. Crowley EJ, Zisserman A (2016) The art of detection. In: European conference on computer vision, Springer, pp 721–737
41. Crowley E, Zisserman A (2014) The state of the art: Object retrieval in paintings using discriminative regions. In: Proceedings of the British Machine Vision Conference 2014. British Machine Vision Association. https://doi.org/10.5244/c.28.38
42. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), IEEE Computer Society, pp 886–893. https://doi.org/10.1109/cvpr.2005.177
43. Del Chiaro R, Bagdanov AD, Del Bimbo A (2019) Webly-supervised zero-shot learning for artwork instance recognition. Pattern Recogn Lett 128:420–426

44. Dominguez V, Messina P, Parra D, et al (2017) Comparing neural and attractiveness-based visual features for artwork recommendation. In: Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems. ACM, pp 55–59. https://doi.org/10.1145/3125486.3125495
45. Elgammal AM, Saleh B (2015) Quantifying creativity in art networks. CoRR abs/1506.00711
46. Elgammal A, Liu B, Kim D, et al (2018) The shape of art history in the eyes of the machine. Proceedings of the AAAI Conference on Artificial Intelligence 32(1). https://doi.org/10.1609/aaai.v32i1.11894
47. Falomir Z, Museros L, Sanz I et al (2018) Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (QArt-learn). Expert Syst Appl 97:83–94. https://doi.org/10.1016/j.eswa.2017.11.056
48. Felzenszwalb PF, Girshick RB, McAllester D et al (2009) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
49. Fiorucci M, Khoroshiltseva M, Pontil M et al (2020) Machine learning for cultural heritage: A survey. Pattern Recogn Lett 133:102–108. https://doi.org/10.1016/j.patrec.2020.02.017
50. Florea C, Badea M, Florea L, et al (2017) Domain transfer for delving into deep networks capacity to de-abstract art. In: Scandinavian Conference on Image Analysis, Springer, pp 337–349
51. Foka A (2021) Computer vision applications for art history: Reflections and paradigms for future research. In: Proceedings of EVA London 2021. BCS Learning & Development, pp 73–80. https://doi.org/10.14236/ewic/eva2021.12
52. Folego G, Gomes O, Rocha A (2016) From impressionism to expressionism: Automatically identifying van gogh's paintings. 2016 IEEE International Conference on Image Processing (ICIP) pp 141–145
53. Frank SJ (2021) State of the art: This convolutional neural network can tell you whether a painting is a fake. IEEE Spectr 58(9):26–31. https://doi.org/10.1109/MSPEC.2021.9531029
54. Fujiyoshi H, Hirakawa T, Yamashita T (2019) Deep learning-based image recognition for autonomous driving. IATSS research 43(4):244–252
55. Ganin Y, Lempitsky V (2015) Unsupervised domain adaptation by backpropagation. In: International conference on machine learning, PMLR, pp 1180–1189
56. Gao X, Tian Y, Qi Z (2020) Rpd-gan: Learning to draw realistic paintings with generative adversarial network. IEEE Trans Image Process 29:8706–8720
57. Garcia N, Vogiatzis G (2019) How to read paintings: Semantic art understanding with multi-modal retrieval. In: Lecture Notes in Computer Science. Springer International Publishing, pp 676–691. https://doi.org/10.1007/978-3-030-11012-3_52
58. Gatys LA, Ecker AS, Bethge M (2016a) Image style transfer using convolutional neural networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 2414–2423
59. Gatys LA, Ecker AS, Bethge M (2016b) Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2414–2423
60. Geirhos R, Rubisch P, Michaelis C, et al (2019) Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In: International Conference on Learning Representations,
61. Geman D, Geman S, Hallonquist N et al (2015) Visual turing test for computer vision systems. Proc Natl Acad Sci 112(12):3618–3623
62. Ginosar S, Haas D, Brown T et al (2015) Detecting people in cubist art. AI Matters 1(3):16–18. https://doi.org/10.1145/2735392.2735398
63. Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448
64. Girshick R, Donahue J, Darrell T, et al (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
65. Goenaga MA (2020) A critique of contemporary artificial intelligence art: Who is edmond de belamy? AusArt 8(1):51–66. https://doi.org/10.1387/ausart.21490
66. Gonthier N, Ladjal S, Gousseau Y (2022) Multiple instance learning on deep features for weakly supervised object detection with extreme domain shifts. Comput Vis Image Underst 214(103):299
67. Gonthier N, Gousseau Y, Ladjal S, et al (2019) Weakly supervised object detection in artworks. In: Lecture Notes in Computer Science. Springer International Publishing, pp 692–709. https://doi.org/10.1007/978-3-030-11012-3_53
68. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al (2014) Generative adversarial networks. arXiv preprint arXiv:1406.2661
69. Gultepe E, Conturo TE, Makrehchi M (2018) Predicting and grouping digitized paintings by style using unsupervised feature learning. J Cult Herit 31:13–23
70. Gupta S, Kumar M, Garg A (2019) Improved object recognition results using sift and orb feature detector. Multimedia Tools and Applications 78:34157–34171

71. Hayn-Leichsenring GU, Lehmann T, Redies C (2017) Subjective ratings of beauty and aesthetics: Correlations with statistical image properties in western oil paintings. i-Perception 8(3):204166951771,547. https://doi.org/10.1177/2041669517715474
72. Hearst MA, Dumais ST, Osuna E et al (1998) Support vector machines. IEEE Intelligent Systems and their applications 13(4):18–28
73. He K, Gkioxari G, Dollar P, et al (2017) Mask r-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp 2961–2969. https://doi.org/10.1109/iccv.2017.322
74. Hosain MK, Harun-Ur-Rashid, Taher TB, et al (2020) Genre recognition of artworks using convolutional neural network. In: 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE, pp 1–5. https://doi.org/10.1109/iccit51783.2020.9392688
75. Hu X (2018) Tensorflow implementation of cyclegan. https://github.com/xhujoy/CycleGAN-tensorflow
76. Hu M, Wang H, Wang X et al (2019) Video facial emotion recognition based on local enhanced motion history image and cnn-ctslstm networks. J Vis Commun Image Represent 59:176–185
77. Ibrahim BIE, Eyharabide V, Page VL et al (2022) Few-shot object detection: Application to medieval musicological studies. Journal of Imaging 8(2):18. https://doi.org/10.3390/jimaging8020018
78. Iconart dataset (2018). https://wsoda.telecom-paristech.fr/downloads/dataset/, Accessed: 2023-03-08
79. Iliadis LA, Nikolaidis S, Sarigiannidis P et al (2021) Artwork style recognition using vision transformers and mlp mixer. Technologies 10(1):2
80. Inoue N, Furuta R, Yamasaki T, et al (2018) Cross-domain weakly-supervised object detection through progressive domain adaptation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp 5001–5009. https://doi.org/10.1109/cvpr.2018.00525
81. Jeon HJ, Jung S, Choi YS, et al (2020) Object detection in artworks using data augmentation. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, pp 1312–1314. https://doi.org/10.1109/ictc49870.2020.9289321
82. Johnson MK, Stork DG, Biswas S, et al (2008) Inferring illumination direction estimated from disparate sources in paintings: an investigation into jan vermeer's girl with a pearl earring. In: Computer image analysis in the study of art, International Society for Optics and Photonics, p 68100I
83. Junger A, Metzenthin E, Wullenweber P (2021) Object detection. In: Deep learning for computer vision in the art domain: proceedings of the master seminar on practical introduction to deep learning for computer vision, HPI WS 20/21, Universitätsverlag Potsdam, p 33
84. Kadish D, Risi S, Lovlie AS (2021) Improving object detection in art images using only style transfer. In: 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, pp 1–8. https://doi.org/10.1109/ijcnn52387.2021.9534264
85. Kantorov V, Oquab M, Cho M, et al (2016) Contextlocnet: Context-aware deep network models for weakly supervised localization. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14, Springer, pp 350–365
86. Keren D (2002) Painter identification using local features and naive bayes. In: Object recognition supported by user interaction for service robots. IEEE Comput. Soc, pp 474–477. https://doi.org/10.1109/icpr.2002.1048341
87. Khalili A, Bouchachia H (2021) An information theory approach to aesthetic assessment of visual patterns. Entropy 23(2):153. https://doi.org/10.3390/e23020153
88. Kotenseki dataset (2019). http://codh.rois.ac.jp/pmjt/, Accessed: 2023-03-14
89. Kumar KK, Venkateswara Reddy H (2022) Crime activities prediction system in video surveillance by an optimized deep learning framework. Concurrency and Computation: Practice and Experience 34(11):e6852
90. Lang S, Ommer B (2018) Attesting similarity: Supporting the organization and study of art image collections with computer vision. Digital Scholarship in the Humanities 33(4):845–856. https://doi.org/10.1093/llc/fqy006
91. Lecoutre A, Negrevergne B, Yger F (2017) Recognizing art style automatically in painting with deep learning. In: Zhang ML, Noh YK (eds) Proceedings of the Ninth Asian Conference on Machine Learning, Proceedings of Machine Learning Research, vol 77. PMLR, Yonsei University, Seoul, Republic of Korea, pp 327–342
92. Lin TY, Goyal P, Girshick R, et al (2017) Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp 2980–2988. https://doi.org/10.1109/iccv.2017.324
93. Lin Y (2020) Sentiment analysis of painting based on deep learning. In: International Conference on Application of Intelligent Systems in Multi-modal Information Analytics, Springer, pp 651–655. https://doi.org/10.1007/978-3-030-51556-0_96
94. Liu Y (2021) Improved generative adversarial network and its application in image oil painting style transfer. Image Vis Comput 105(104):087

95. Liu W, Anguelov D, Erhan D, et al (2016) Ssd: Single shot multibox detector. In: European conference on computer vision, Springer, pp 21–37
96. Liu Z, Lin Y, Cao Y, et al (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 10012–10022
97. Lu Y, Guo C, Dai X et al (2022) Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training. Neurocomputing 490:163–180
98. Madhu P, Kosti R, Mührenberg L, et al (2019) Recognizing characters in art history using deep learning. In: Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents, pp 15–22
99. MAFD-150 dataset (2018). https://github.com/andeeptoor/mafd-150, Accessed: 2023-03-06
100. Ma D, Gao F, Bai Y, et al (2017) From part to whole: Who is behind the painting? In: Proceedings of the 25th ACM international conference on Multimedia. ACM, pp 1174–1182. https://doi.org/10.1145/3123266.3123325
101. Maji B, Swain M, Mustaqeem (2022) Advanced fusion-based speech emotion recognition system using a dual-attention mechanism with conv-caps and bi-gru features. Electronics 11(9). https://doi.org/10.3390/electronics11091328
102. Mao H, Cheung M, She J (2017) Deepart: Learning joint representations of visual arts. In: Proceedings of the 25th ACM international conference on Multimedia. ACM, pp 1183–1191. https://doi.org/10.1145/3123266.3123405
103. Marinescu MC, Reshetnikov A, López JM (2020) Improving object detection in paintings based on time contexts. In: 2020 International Conference on Data Mining Workshops (ICDMW), IEEE, pp 926–932
104. Mensink T, Van Gemert J (2014) The rijksmuseum challenge: Museum-centered visual recognition. In: Proceedings of International Conference on Multimedia Retrieval, pp 451–454
105. Mermet A, Kitamoto A, Suzuki C, et al (2020) Face detection on pre-modern japanese artworks using r-CNN and image patching for semi-automatic annotation. In: Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritAge Contents. ACM, pp 23–31. https://doi.org/10.1145/3423323.3423412
106. Messina P, Dominquez V, Parra D, et al (2017) Exploring content-based artwork recommendation with metadata and visual features. ArXiv abs/1706.05786
107. Mohammad SM, Kiritchenko S (2018) Wikiart emotions: An annotated dataset of emotions evoked by art. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)
108. Moutafidou A, Fudos I, Adamopoulos G, et al (2018) Reconstruction and visualization of cultural heritage artwork objects. In: International Conference on Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage, Springer, pp 141–149
109. Mustaqeem, Kwon S (2020) Clstm: Deep feature-based speech emotion recognition using the hierarchical convlstm network. Mathematics 8(12). https://doi.org/10.3390/math8122133
110. Mustaqeem, Kwon S (2021a) 1d-cnn: Speech emotion recognition system using a stacked network with dilated cnn features. Cmc-computers Materials & Continua 67:4039–4059
111. Mustaqeem, Kwon S (2021b) Att-net: Enhanced emotion recognition system using lightweight self-attention module. Applied Soft Computing 102:107101. https://doi.org/10.1016/j.asoc.2021.107101
112. Mustaqeem Kwon S (2021) Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network. Int J Intell Syst 36:5116–5135
113. Mustaqeem Ishaq M, Kwon S (2022) A cnn-assisted deep echo state network using multiple time-scale dynamic learning reservoirs for generating short-term solar energy forecasting. Sustainable Energy Technol Assess 52:102275. https://doi.org/10.1016/j.seta.2022.102275
114. Mzoughi O, Bigand A, Renaud C (2018) Face detection in painting using deep convolutional neural networks. In: Advanced Concepts for Intelligent Vision Systems. Springer International Publishing, pp 333–341. https://doi.org/10.1007/978-3-030-01449-0_28
115. Nasir IM, Raza M, Shah JH, Wang SH, Tariq U, Khan MA (2022) Harednet: A deep learning based architecture for autonomous video surveillance by recognizing human actions. Comput Electr Eng 99:107805. https://doi.org/10.1016/j.compeleceng.2022.107805
116. Paintings dataset (2014). https://www.robots.ox.ac.uk/~vgg/data/paintings/, Accessed: 2023-03-06
117. Pasqualino G, Furnari A, Farinella GM (2021a) Unsupervised domain adaptation for object detection in cultural sites. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE. https://doi.org/10.1109/icpr48806.2021.9412661
118. Pasqualino G, Furnari A, Signorello G, et al (2021b) An unsupervised domain adaptation scheme for single-stage artwork recognition in cultural sites. Image and Vision Computing 107:104098

119. Peleshko D, Soroka K (2013) Research of usage of haar-like features and adaboost algorithm in viola-jones method of object detection. In: 2013 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), IEEE, pp 284–286
120. PeopleArt dataset (2014). https://github.com/BathVisArtData/PeopleArt, Accessed: 2023-03-06
121. PhotoArt50 dataset (2016). https://github.com/BathVisArtData/PhotoArt50, Accessed: 2023-03-14
122. Polatkan G, Jafarpour S, Brasoveanu A, et al (2009) Detection of forgery in paintings using supervised learning. 2009 16th IEEE International Conference on Image Processing (ICIP) pp 2921–2924
123. Ranjgar B, Azar MK, Sadeghi-Niaraki A et al (2019) A novel method for emotion extraction from paintings based on luscher's psychological color test: Case study iranian-islamic paintings. IEEE Access 7:120857–120871. https://doi.org/10.1109/access.2019.2936896
124. Redmon J, Divvala S, Girshick R, et al (2016) You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 779–788. https://doi.org/10.1109/cvpr.2016.91
125. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
126. Ren S, He K, Girshick R et al (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 28:91–99
127. Rodrigues JB, Ferreira AVM, Maia IMO, et al (2018) Image processing of artworks for construction of 3d models accessible to the visually impaired. In: International Conference on Applied Human Factors and Ergonomics, Springer, pp 243–253
128. Rombach R, Blattmann A, Lorenz D, et al (2022) High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://github.com/CompVis/latent-diffusion, https://arxiv.org/abs/2112.10752
129. Sabatelli M, Kestemont M, Daelemans W, et al (2019) Deep transfer learning for art classification problems. In: Lecture Notes in Computer Science. Springer International Publishing, pp 631–646. https://doi.org/10.1007/978-3-030-11012-3_48
130. Saito K, Ushiku Y, Harada T, et al (2019) Strong-weak distribution alignment for adaptive object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 6956–6965
131. Saleh B, Elgammal A (2015) Large-scale classification of fine-art paintings: Learning the right metric on the right feature. ArXiv abs/1505.00855
132. Sargentis GF, Dimitriadis P, Koutsoyiannis D (2020) Aesthetical issues of leonardo da vinci's and pablo picasso's paintings with stochastic evaluation. Heritage 3(2):283–305. https://doi.org/10.3390/heritage3020017
133. Sarı C, Salah AA, Akdag Salah AA (2019) Automatic detection and visualization of garment color in western portrait paintings. Digital Scholarship in the Humanities 34(Supplement_1):i156–i171
134. Schlecht J, Carqué B, Ommer B (2011) Detecting gestures in medieval images. In: 2011 18th IEEE International Conference on Image Processing, IEEE, pp 1285–1288
135. Seguin B, Striolo C, Kaplan F, et al (2016) Visual link retrieval in a database of paintings. In: European conference on computer vision, Springer, pp 753–767
136. Shen X, Efros AA, Aubry M (2019) Discovering visual patterns in art collections with spatially-consistent feature learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9278–9287
137. Sheng S, Moens MF (2019) Generating captions for images of ancient artworks. In: Proceedings of the 27th ACM International Conference on Multimedia, pp 2478–2486
138. Sindel A, Maier A, Christlein V (2023) Artfacepoints: High-resolution facial landmark detection in paintings and prints. In: Karlinsky L, Michaeli T, Nishino K (eds) Computer Vision - ECCV 2022 Workshops. Springer Nature Switzerland, Cham, pp 298–313
139. Sirirattanapol C, Matsui Y, Satoh S, et al (2017) Deep image retrieval applied on kotenseki ancient japanese literature. In: 2017 IEEE International Symposium on Multimedia (ISM). IEEE, pp 495–499. https://doi.org/10.1109/ism.2017.98
140. Smirnov S, Eguizabal A (2018) Deep learning for object detection in fine-art paintings. In: 2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo), IEEE, pp 45–49. https://doi.org/10.1109/MetroArchaeo43810.2018.9089828
141. Song Y, Ren S, Lu Y, et al (2022) Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. Computer Methods and Programs in Biomedicine p 106821
142. Spehr M, Wallraven C, Fleming RW (2009) Image statistics for clustering paintings according to their visual appearance. Computational Aesthetics 2009: Eurographics Workshop on Computational Aesthetics in Graphics. Visualization and Imaging, Eurographics, pp 57–64

143. Srinivasan R, Rudolph C, Roy-Chowdhury AK (2015) Computerized face recognition in renaissance portrait art: A quantitative measure for identifying uncertain subjects in ancient portraits. IEEE Signal Process Mag 32(4):85–94. https://doi.org/10.1109/msp.2015.2410783
144. Srinivasan R, Roy-Chowdhury A, Rudolph C, et al (2013) Recognizing the royals: Leveraging computerized face recognition for identifying subjects in ancient artworks. In: Proceedings of the 21st ACM International Conference on Multimedia. Association for Computing Machinery, New York, NY, USA, MM '13, p 581–584. https://doi.org/10.1145/2502081.2502153
145. Stork DG (2011) Computer analysis of lighting style in fine art: steps towards inter-artist studies. In: Computer Vision and Image Analysis of Art II, vol 7869. SPIE, p 786903. https://doi.org/10.1117/12.873190
146. Stork D (2009) Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. International Conference on Computer Analysis of Images and Patterns. Springer, CAIP, pp 9–24
147. Stork D, Johnson MK (2006) Computer vision, image analysis, and master art: Part 2. IEEE Multimedia 13:12–17
148. Strezoski G, Worring M (2017) Omniart: Multi-task deep learning for artistic data analysis. ArXiv abs/1708.00684
149. Surapaneni S, Syed S, Lee LY (2020) Exploring themes and bias in art using machine learning image analysis. In: 2020 Systems and Information Engineering Design Symposium (SIEDS), IEEE, pp 1–6
150. Tan WR, Chan CS, Aguirre HE, et al (2016) Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In: 2016 IEEE international conference on image processing (ICIP), IEEE, pp 3703–3707. https://doi.org/10.1109/ICIP.2016.7533051
151. Tan WR, Chan CS, Aguirre HE, et al (2017) Artgan: Artwork synthesis with conditional categorical gans. 2017 IEEE International Conference on Image Processing (ICIP) pp 3760–3764
152. Tan W, Wang J, Wang Y et al (2018) Cnn models for classifying emotions evoked by paintings. Technical Report, SVL Lab, Stanford University, USA, Tech. rep
153. Tian Y, Suzuki C, Clanuwat T, et al (2020) Kaokore: A pre-modern japanese art facial expression dataset. arXiv preprint arXiv:2002.08595
154. Tyler CW, Smith WAP, Stork DG (2012) In search of Leonardo: computer-based facial image analysis of Renaissance artworks for identifying Leonardo as subject. In: Rogowitz BE, Pappas TN, de Ridder H (eds) Human Vision and Electronic Imaging XVII, International Society for Optics and Photonics, vol 8291. SPIE, pp 407 – 413
155. Van Noord N, Hendriks E, Postma E (2015) Toward discovery of the artist's style: Learning to recognize artists by their artworks. IEEE Signal Process Mag 32(4):46–54
156. Vedaldi A, Lenc K (2015) Matconvnet: Convolutional neural networks for matlab. In: Proceedings of the 23rd ACM international conference on Multimedia, pp 689–692
157. Volpe Y, Furferi R, Governi L et al (2014) Computer-based methodologies for semi-automatic 3d model generation from paintings. International Journal of Computer Aided Engineering and Technology 6(1):88–112
158. Wechsler H, Toor AS (2019) Modern art challenges face detection. Pattern Recogn Lett 126:3–10. https://doi.org/10.1016/j.patrec.2018.02.014
159. Westlake N, Cai H, Hall P (2016) Detecting people in artwork with CNNs. In: Lecture Notes in Computer Science. Springer International Publishing, pp 825–841. https://doi.org/10.1007/978-3-319-46604-0_57
160. Wikiart: visual art encyclopedia (2010). https://www.wikiart.org/, Accessed: 2023-03-06
161. Wikicommons (2004). https://commons.wikimedia.org/wiki/MainPage, Accessed: 2023-03-08
162. Winarno E, Hadikurniawati W, Nirwanto AA, et al (2018) Multi-view faces detection using viola-jones method. In: Journal of Physics: Conference Series, IOP Publishing, p 012068
163. Winston JJ, Hemanth DJ, Angelopoulou A, et al (2022) Hybrid deep convolutional neural models for iris image recognition. Multimedia Tools and Applications pp 1–23
164. Wu Q, Cai H, Hall P (2014) Learning graphs to model visual objects across different depictive styles. In: European Conference on Computer Vision, Springer, pp 313–328. https://doi.org/10.1007/978-3-319-10584-0_21
165. Xie J, Girshick R, Farhadi A (2016) Unsupervised deep embedding for clustering analysis. In: International conference on machine learning, PMLR, pp 478–487
166. Yakar M, Doğan Y (2018) Gis and three-dimensional modeling for cultural heritages. International Journal of Engineering and Geosciences 3(2):50–55
167. Yang Z (2021) Classification of picture art style based on VGGNET. J Phys: Conf Ser 1774(1):012043. https://doi.org/10.1088/1742-6596/1774/1/012043

168. Yang H, Min K (2019) Classification of basic artistic media based on a deep convolutional approach. The Visual Computer 36(3):559–578. https://doi.org/10.1007/s00371-019-01641-6
169. Yang H, Min K (2019b) A deep approach for classifying artistic media from artworks. KSII Trans Internet Inf Syst 13:2558–2573
170. Yaniv J, Newman Y, Shamir A (2019) The face of art: landmark detection and geometric style in portraits. ACM Transactions on graphics (TOG) 38(4):1–15
171. Yanulevskaya V, Uijlings J, Bruni E, et al (2012) In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In: Proceedings of the 20th ACM international conference on multimedia, pp 349–358
172. Yi R, Liu YJ, Lai YK, et al (2019) Apdrawinggan: Generating artistic portrait drawings from face photos with hierarchical gans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 10743–10752
173. Yoloface (2019). https://github.com/sthanhng/yoloface, Accessed: 2023-03-08
174. Yolo-v5 (2023). https://github.com/ultralytics/yolov5, Accessed: 2023-03-08
175. Young-Min K (2019) Feature visualization in comic artist classification using deep neural networks. Journal of Big Data 6(1):1–18. https://doi.org/10.1186/s40537-019-0222-3
176. Zhang C, Lei K, Jia J, et al (2018a) Ai painting: an aesthetic painting generation system. In: Proceedings of the 26th ACM international conference on Multimedia, pp 1231–1233
177. Zhang H, Li Q, Sun Z, et al (2018b) Combining data-driven and model-driven methods for robust facial landmark detection. IEEE Transactions on Information Forensics and Security 13(10):2409–2422
178. Zhao L, Shang M, Gao F et al (2020) Representation learning of image composition for aesthetic prediction. Comput Vis Image Underst 199:103024. https://doi.org/10.1016/j.cviu.2020.103024
179. Zhu JY, Park T, Isola P, et al (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp 2242–2251. https://doi.org/10.1109/iccv.2017.244
180. Zhu Y, Yan WQ (2022) Traffic sign recognition based on deep learning. Multimedia Tools and Applications 81(13):17779–17791
181. Zujovic J, Gandy L, Friedman S, et al (2009) Classifying paintings by artistic genre: An analysis of features & classifiers. In: 2009 IEEE International Workshop on Multimedia Signal Processing. IEEE, pp 1–5. https://doi.org/10.1109/mmsp.2009.5293271