Neural Computing and Applications (2021) 33:121–157
https://doi.org/10.1007/s00521-020-05565-4

Artificial Neural Networks and Deep Learning in the Visual Arts: a review

Iria Santos[1] · Luz Castro[1] · Nereida Rodriguez-Fernandez[1] · Álvaro Torrente-Patiño[1] · Adrián Carballal[1]

Received: 22 May 2020 / Accepted: 1 December 2020 / Published online: 12 January 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd. part of Springer Nature 2021

Abstract
In this article, we perform an exhaustive analysis of the use of Artificial Neural Networks and Deep Learning in the Visual Arts. We begin by introducing changes in Artificial Intelligence over the years and examine in depth the latest work carried out in prediction, classification, evaluation, generation, and identification through Artificial Neural Networks for the different Visual Arts. While we highlight the contributions of photography and pictorial art, there are also other uses for 3D modeling, including video games, architecture, and comics. The results of the investigations discussed show that the use of Artificial Neural Networks in the Visual Arts continues to evolve and have recently experienced significant growth. To complement the text, we include a glossary and table with information about the most commonly employed image datasets.

Keywords Artificial Neural Networks · Generative Adversarial Networks · Convolutional Neural Networks · Deep Learning · Visual Arts · Machine Learning · Prediction · Classification · Evaluation · Generation · Identification · Transfer Learning · Datasets

1 Introduction

Since the 1943 paper of McCulloch and Pitts [1], Artificial Neural Networks (ANNs) have been used for applications in many fields, such as health [2–4], optimization of structural design problems from civil engineering [5], traffic accident prediction [6], renewable energies [7], electrochemistry [8], video games generation [9], text translation [10], voice recognition [11–13], as well as in the applications of several types of commercialized hardware [14], etc.

Recent years have seen considerable growth in the use of Deep Learning, both in research and in industrial applications. Despite this recent growth, it has been used for decades, with examples including work from Giebel [15], Fukushima [16], and LeCun et al. [17]. But it was work like that of Krizhevsky et al. [18] that improved the state of the art of image classification. They trained a Deep Convolutional Neural Network (DCNN) to classify 1.3 million high-resolution images in ImageNet LSVRC-2010 dataset [19] (1000 different classes). The tests achieved error rates of 39.7% (top 1) and 18.9% (top 5). To perform fast training, they used non-saturating neurons and a GPU to implement convolutional nets.

Techniques such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have been widely used. The basic architecture of a CNN consists of a convolution layer, a layer that applies the activation function in the matrix elements, a dimension reduction layer and a final layer with the number of neurons to be classified. This type of network can be made more complex, enhanced, and modified with other layers [20, 21]. GANs are another type of ANN in which two unsupervised networks of neurons compete. One of the networks (the generative network) generates candidates that the other ANN (the discriminative network) evaluates, following a scheme similar to that of co-evolutionary systems [22].

Few tasks are as characteristic of human beings as artistic ones. It is therefore natural that many scientists working in Artificial Intelligence (AI) are interested in modeling aspects of art using computer systems [23, 24].

✉ Iria Santos
iria.santos@udc.es

1 Department of Computer Science, Faculty of Computer Science, CITIC, University of A Coruña, 15071 A Coruña, Spain

However, artistic tasks are challenging. The value of artistic work is dependent on the cultural environment and the aesthetic tastes of the human observer. This aesthetic value may even depend on time, as a person's taste or emotional state may change over the years. Some artistic tasks, such as the creation of new images, are related to the ability to create new content or explore large search spaces.

These difficulties slowed the application of AI to artistic endeavors for many years. But since the 1990s, the application of soft-computing techniques, mainly evolutionary computing and neural networks, has revitalized the area. In recent years, a growing number of articles have proposed the use of AI techniques for the creation or analysis of artistic work in different fields, such as painting, architecture, sculpture, music, and even dance or poetry. There are now annual international conferences dedicated to AI in these areas, such as the International Conference on Computational Creativity [25], the Bridges Conference [26], and the International Conference on Artificial Intelligence on Music, Sound, Art and Design [27]. Special issues devoted to the application of AI in art have also been published in journals such as *ALIFE* [28], *Mathematics and the Arts* [29], and *Complexity* [30].

There are constant advances in the state of the art for specific uses of AI. For example, Galanter conducted a 2012 workshop on computational aesthetic evaluation [31], and Spratt and Elgammal [32] analyzed the use of computer vision systems in the analysis of paintings in 2014, including analysis of the stylistic influences between different paintings, as well as the reaction of art historians to the use of these technologies. Toivonen and Gross [33] provided an overview of the possible uses of Data Mining and Machine Learning in creative systems in 2015; Upadhyaya et al. [34] reviewed low- and high-level features employed on content-based image retrieval (CBIR) in 2016; and Colin et al. [35] analyzed different studies on the psychology of aesthetics, including the relationship between complexity and aesthetics, measures of complexity, and complexity predictors from the perspective of AI, relating this research with evolutionary computation fitness functions measuring aesthetics.

Several papers have analyzed the contributions of other techniques, such as evolutionary computing, to the artistic domain. For example, Todd [36], curiously, did so in a neural networks reference book, and in 2007, Lewis published a review that analyzed more than 150 works in Evolutionary Computation applied to the arts [37]. Some works analyze the use of Deep Learning techniques applied to the generation and analysis of music, such as the 2017 paper and 2019 book on the topic by Briot et al. [38, 39].

In our paper, we present the use of different ANN-based techniques for the creation and analysis of visual art. The methodology used has the following phases: (1) An exhaustive search of work by different authors on AI applied to the Visual Arts conducted through the scientific portals Google Scholar and ResearchGate using keywords including AI, ML, ANN, image, style transfer, art generation, artistic prediction, visual art, painting, painter, and aesthetics. Conference proceedings and special issues commented before in this paper have been consulted. (2) The references of these papers have been examined—both papers that cite them and those cited by them up to the third degree. (3) All papers using ANN have been filtered. Some papers related to other techniques (such as Support Vector Machines, SVM) were included because they present datasets that were explored in later ANN papers.

We have grouped investigations by similar topics into sections that we present in order of increasing subjectivity. We start with work that deals with tasks that are simple and objective for a human being, such as the detection of objects. Then we proceed in the difficulty of the tasks addressed and in their subjectivity, such as the detection of styles or assessment of the aesthetic value of an image. We end with tasks that involve creating new images, such as style transfer. In some sections, we have grouped similar papers; within each section or grouping, the order is chronological, to show the evolution of technology.

Section 2 discusses the specific use of ANNs in object detection in artworks. In Sect. 3, we discuss articles dealing with the classification of visual works according to their style or author, while Sect. 4 deals with the classification of works according to their visual characteristics. In Sect. 5, we deal with those that evaluate the aesthetic value or quality of visual works. Section 6 covers papers that use ANNs for style transfer, while Sect. 7 describes various systems that use ANNs for automatic generation or reconstruction of images. Finally, a series of general conclusions are presented in the final section. We present in Table 1 the number of papers per year presented in each section. At the end of the document, we list in Table 9 the datasets used in the cited texts. In some sections, we have grouped similar papers; within each section or grouping, chronological order is used to show the evolution of technology. Finally, we provide a short glossary describing and defining some concepts that are discussed in the paper.

We have compiled an index of all the articles covered in this state of the art [40].

2 Detection of elements in works of art

To delve into the field of aesthetics and Visual Arts, we will begin with visual perception, and in this section, we will address the research that uses ANNs to detect elements in painting works and comics.

Table 1 Summary of the number of articles on each topic analyzed in this paper by year: (2) detection of elements in works of art; (3) classification according to style and/or authorship; (4) classification based on quality, complexity and visual characteristics; (5) evaluation based on photo quality or aesthetics; (6) style transfer; and (7) pictorial generation or reconstruction

Type of papers	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
All	1	1	7	12	15	19	20	16	5	92
2	–	–	1	1	2	–	4	1	1	10
3	–	1	3	2	3	5	4	3	1	22
4	–	–	–	2	2	2	1	2	1	10
5	–	–	2	4	4	7	6	4	1	28
6	–	–	–	2	2	2	–	1	–	7
7	–	–	–	1	1	2	6	6	1	17

ANNs have been used extensively for the detection of elements (objects, people, animals) in photographs [41–43]. They have also been used to detect visual relationships between these elements (e.g., if an object is behind or inside another object) within image content [44–48] and even to detect cracks in painted surfaces [49]. In this section, we highlight several works in which ANNs are used to detect elements in paintings and comics. We begin by analyzing some methods that detect objects in paintings. Some of these methods are trained with a dataset of real-world photographs and tested with paintings. Next, we describe some work aimed at detecting humans in paintings and finally describe some that detect objects in comics.

2.1 Object detection in paintings

Hall et al. [50] tested an approach in which several classifiers are trained using different sets of characteristics (including those extracted from a CNN). In all cases, the classifier is trained with photographs and tested with paintings. They sought to use Deep Learning to solve the problem of cross representation, but could exceed a detection rate of 40% using CNN methods.

Seguin et al. [51] tested different machine vision techniques to detect visual elements (people, animals, and things) linked to each other (e.g., eyeglasses on people, a table near a chair). The study then tried to use these objects and relationships to find similarities between different pictorial works. They found that a "pre-trained Convolutional Neural Network can perform better for this task than other machine vision methods aimed at photograph analysis" and that the neural network's success is improved by fine-tuning.

Inoue et al. [52] presented a framework to cross-domain weakly supervised object detection, the detection of common object in images from different domains (natural image, painting, design) that require little or no human supervision. They trained a GAN on photographs and applied the framework to object detection in clipart, watercolor and comic images, achieving some improvements of between 5 and 20 points over current baselines.

In the same year, Gonthier et al. [53] presented a Multiple Instance Learning (MIL) technique for object detection in photographs, drawings, and paintings. They used the detection network Faster R-CNN [54], maintaining the Region Proposal Network (RPN) which takes an image with no specific size as input and outputs a set of rectangular object proposals, each with an objectivity score, and the features corresponding to each proposed region. They compared their approach to the artwork object detection methods of Inoue et al. (previously mentioned) [52] and Westlake et al. (explained in the next subsection) [55], among others. For the specific case of person, Gonthier et al. obtained an Average Precision (AP) of 55.4% with their method, and 59% with the method of Westlake et al. They also obtained a (mean) AP of 50.1% with their method, and 54.3% with the method of Inoue et al. Therefore, in these cases their results were worse, although they surpassed those of other methods. In Fig. 1, we can see an example of the use of MI-max-C detection. The contents used to train their dataset (IconArt [56]), whose detection is shown in the figure, are angels, the baby Jesus, the crucifixion, the Virgin Mary, Saint Sebastian, and ruins.

Later, Gonthier et al. [57] used Faster R-CNN [54] as a feature extractor to train a system of for weakly supervised object detection with extreme domain shifts. They trained the model with photographs in two phases (ImageNet [18] and MS COCO [58]). They applied a Multiple Instance Learning (MIL) to this network—a multiple instance extension of the perceptron [59]. They used six different non-photographic database for testing: PeopleArt [55], Watercolor2k, Clipart1k, Comic2k [52], IconArt [53, 56] and CASPApaintings [60]. The authors detected five problems: in some images, specific areas were detected instead of the entire element (for example, a body part

Fig. 1 Examples of the use of the MI-max-C detection scheme [53] with a rate greater than 0.75

instead of a complete character); some sets were grouped rather than separated into individual instances; the bounding box was sometimes cropped incorrectly; and there were images of outstanding semantic complexity in which the labels fail. The best results they obtained were with the PeopleArt [55] dataset, achieving 94% recall.

2.2 Detection of humans in paintings

Some work has focused on the detection of humans in paintings. Crowley and Zisserman [61] used four techniques to detect people in cubist paintings. One is R-CNN, which did "not perform well on this task" and "overfits to the natural visual world and fails at adapting to the domain of paintings." In this context, it is relevant to highlight the great complexity of detecting objects in unrealistic artistic styles such as cubism. Along the same lines, Westlake et al. [55] tested a dataset composed of photographs, cartoons, and images of people. They obtained an average accuracy of 45% for detecting people using an untrained CNN, increasing to 58% when the CNN was adjusted.

2.3 Object detection in comics

Several studies have focused on object detection in comics. Nguyen et al. [62] detected several elements (panel, balloon, text, comic character, and face) in comics. They employed traditional approaches, as well as innovative models of Deep Learning and text recognition using the LSTM (Long Short-Term Memory) model. Ogawa et al. [63] proposed a CNN-based method for object detection in comics using Manga109 Annotations [64] (dataset information in Table 9). The categories in their method are frame, text, face, and body. The method used is a SSD300-fork implemented in YOLOv2 [65]. They compared their method with the original version of YOLOv2, and those proposed using ChainerCV [66] for Faster R-CNN [54] and SSD300 [67]. The structure uses the VGG-16 [68] network trained with ImageNet [18]. Tests of SSD300-fork were carried out in 10 volumes of different subjects with a total of 880 pages, including comics such as UltraEleven (108 pages of sports) and UnbalanceTokyo (82 pages of science fiction). The results of each method were measured in mean Average Precision (mAP) (i.e., by the average of the average precision scores for each query). The highest value obtained was with the SSD300-fork, 84.2, compared to 49.9 (Faster R-CNN [54]), 81.3 (SSD300), and 59.7 (YOLOv2). Finally, in the next year, Dubray and Laubrock [69] automatically segmented text balloons. They used a CNN approach inspired by the U-Net architecture, combined with a VGG-16 [68]-based encoder trained on annotated pages of the Graphic Narrative Corpus [70], resulting in an F1-score of over 0.94.

3 Classification by style and/or authorship

When creating an ANN for recognition of pictorial works, one of the first questions we can ask ourselves is how humans recognize the authorship of a work. What characteristics, what we call "style," does work possess that allow us to be sure that an image is the work of Velázquez or that it is a cubist painting? This task, although it can be objectively evaluated (e.g., in the detection of the author of a painting), presents more complexity than those discussed in the previous section. The work and artistic styles of one painter are often related to those of other painters and styles. Furthermore, it is common for an artist to go through different periods with clearly differentiated styles. In this section, we focus on work that extracts characteristics from painting works, photographs, illustrations, or comics to group them according to style or authorship. We will begin by showing some classification systems for paintings and photographs, then focus on drawings, and finally analyze comics and architectural works.

3.1 Classification of painting works and photographs

Some works extract characteristics from painting works and photographs to group them according to their style or authorship, as we describe below.

Murray et al. [71] presented a large-scale database for visual aesthetic analysis (AVA) (dataset information provided in Table 9). The database contains more than 250,000 images with a variety of metadata (aesthetic scores for each image, semantic labels for more than 60 categories, and photographic style labels) to support research on computational models of aesthetic preference. They trained three different classifiers—a large-scale aesthetic quality categorizer, content-based aesthetic categorizer, and style categorizer—to show their application in computational aesthetics. They did not use a CNN, but their reference database has been used by many later studies to compare the effectiveness of methods (e.g., [72]).

Karayev et al. [73] presented two datasets consisting of 80,000 Flickr [74] photographs annotated with 20 style tags and 85,000 paintings with 25 style/gender tags. They used a multilayer network [75] for art style prediction. They divided the Flickr categories into six types: optical techniques (macro, bokeh, depth-of-field, long exposure, HDR), atmosphere (hazy, sunny), mood (serene, melancholy, ethereal), composition styles (minimal, geometric, detailed, texture), color (pastel, bright) and genre (noir, vintage, romantic, horror). They used the winning Caffe convolutional arquitecture [76] with is open-source, which uses ImageNet [18] annotated images. The AP varied for each class, ranging from 0.17 (depth of field) to 0.62 (macro). The accuracy also varied, from 68% (romantic, depth of field) to 85% (sunny, noir, macro); the average accuracy per class was 78%. There were confusions that the authors considered normal, such as confusing depth of field with macro, romantic with pastel and vintage with melancholy. They did consider some of the mistakes observed, such as confusion between macro and bright/energetic, surprising.

To see whether the results were line with the assessments of humans, they solicited ratings through Amazon's Mechanical Turk, with three assessments per image. The average accuracy of human raters was 75% (ranging from 61% for romantic to 92% for macro). The best results of the algorithm were worse than those of the users for macro and horror, and better for vintage, romantic, pastel, detailed, HDR, and long exposure. In the experiments with WikiArt (see dataset information in Table 9), the mAP results were 0.441. Accuracy by class ranged from 72% (symbolism, expressionism, art nouveau) to 94% (ukiyo-e, minimalism, color field painting). In AVA experiments [71], the mAP was 0.579. Finally, they applied the style classifiers learned from Flickr to a new dataset of 80,000 images collected from Pinterest for the organization of paintings and photographs.

Bar et al. [77] proposed the use of Binarized Features derived from a Deep Neural Network for the classification of artistic styles and applied the method to the Wikiart dataset [78]. Their baseline descriptors were extracted from a Decaf implementation [41] of a CNN trained in ImageNet [18]. They combined Decaf encoding with PiCoDes [79], an optimized method of joining low-level features [34]. They tried different classifiers: SVM, AdaBoost, Bayes Naive and k-Nearest Neighbors (KNN), and preferred the latter. After several configurations "the best features fusion descriptor incorporates PiCoDes (1024-dimensionality), PiCoDes (2048-dimensionality), encoded $Decaf_5$ (405-dimensionality) and encoded $Decaf_6$ (405-dimensionality) and matches the best (non-encoded) features fusion result using a binary descriptor with 63% compression." They obtained an accuracy of 0.43 and an AP of 0.47. Khan et al. [80] employed visual characteristics to identify artistic styles (abstract, expressionism, baroque, constructivism, cubism, impressionism, neoclassical, pop art, post-impressionism, realism, renaissance, romanticism, surrealism, and symbolism) in paintings using a CNN. They also classified the paintings of 91 artist by artist. They used binary representations combined with PiCoDes descriptors [79]. Examples of the best and worst results of the dataset categorization are shown in Fig. 2.

Mensink and Van Gemert [81] proposed four classifiers to predict the artist, type, material, and year of creation. They used 1-vs-Rest linear SVM with 112,039 photographic reproductions of the artworks of 6629 artists exhibited in the Rijksmuseum in Amsterdam (see dataset information in Table 9). The result "improves the tools of a museum curator while improving content-based exploration by online visitors of the museum collection." This study also does not use ANN, but its dataset has been subsequently used by several authors, for example Van Noord et al. [82] or Jboor et al. [83].

Castro et al. [84] presented the results of two experiments that compare the operation of a computer system with that of a group of humans in the performance of two tasks: painter identification and aesthetic appreciation. The

Fig. 2 Examples of best and worst results of the automatic classification by author in the dataset of 2014 trained by Khan et al. [80] with a score above 0.75

first experiment consisted of identifying the author of a painting within a dataset of 666 paintings (212 by Picasso, 339 by Monet and 115 by Kandinsky). The second used Maitland Graves' test of aesthetic appreciation [85, 86], which consists of evaluating some aspects of the viewer's aptitude for appreciating an art form. This test of appreciation of drawings involved 90 sets of images in which one broke some aesthetic principle defined by the author (unity, predominance, balance between elements, variety, continuity, symmetry, proportion, and rhythm). There are examples of similar studies from other authors with different samples of humans [85, 86], as well as using mathematical and computational models [87]. For instance, Machado et al. [87], in 2008, obtained a success rate of 64.9% using a heuristic approach and 71.67% with the use of ANN. Castro et al. [84] selected 30 pairs of elements at random, which they presented to individuals. The results of the computer system were superior to those obtained by humans in both tasks.

Later, Van Noord et al. [82] trained a CNN corresponding to the architecture (PigeoNET) of AlexNet [18] on a large collection of digitized artworks. They used artworks from 2260 individuals from the Rijksmuseum Challenge dataset [81]. Their goal was for PigeoNET to be able to classify works by author, and for this purpose, trained it to identify characteristics of each one of them. They reported an accuracy of greater than 70%.

Saleh et al. [88] developed a machine capable of predicting style, genre, and artist. They used the WikiArt dataset [78] and extracted the low-level features [34] with GIST features [89] and high-level semantic features [34] with Classeme [90], PiCoDes [79], and a pre-trained CNN. They obtained 45.97% accuracy for style classification, which they compared with the 43% accuracy achieved by the previous similar experiment by Bar et al. [77].

In the next year, Tan et al. [91, 92] presented a large-scale classification of fine art paintings using a DCNN. They used images from the ImageNet dataset [18] for pre-training, with the goal of training an end-to-end Deep Convolution model through a CNN with five convolutional layers—three of maximum clustering, and three connected (an AlexNet-inspired design [18]). In the fine-tuning process, they pre-trained the network using ImageNet [18] (see dataset information in Table 9). They used Principal Component Analysis (PCA) and Support Vector Machine (SVM) for the extracted features. They also performed a similar set of experiments with a new softmax layer on the pre-trained CNN without removing the final layer. They used a set of more than 80,000 WikiArt paintings [78] and selected a subset of about 20,000 images from 23 artists [93]. The authors carried out three types of classification experiments based on styles (26 styles, such as abstract expressionism, baroque, minimalism or post-impressionism), genre (abstract painting, cityscape, genre painting, illustration, landscape, nude painting, portrait, religious painting, sketch and study and still life) and authorship (e.g., Claude Monet, Edgar Degas, or Rembrandt). In terms of styles, the results highlight the differentiation of Ukiyo-e (86%), a type of Japanese art. In other styles the performance of the CNN was poorer: synthetic cubism (46%), analytical cubism (50%), rococo (56%), and baroque (64%). In the classification based on genre, there were better results—for example, for portraits (81%) and landscapes (86%). When identifying artist, the CNN had a preference for some techniques or objects in paintings, such as Gustave Dore (engraving and lithography) or Eugene Boudin (outdoor scenes, mostly marine or seaside). However, it gave bad results for works by Salvador Dalí (33%) and confused them with the work of artists like Picasso. The authors indicated their belief that this was due to issues of influence among the painters and left the use of CNN to find influences among authors for future research.

Banerji and Shinha [94] explored the use of a pre-trained CNN as a feature extraction tool combined with different classifiers for the classification of paintings from the dataset Painting-91 [80]. The dataset used has 91 different artists and 4,266 fine art images divided into 13 styles (see Table 9 for more information). The number of paintings by each artist varies between 31 (Frida Kahlo) and 56 (Sandro Botticelli). For the classification of styles, 2388 images were used, discriminating between those that are ambiguous (where the general theme of the paintings is not evident) or those of artists that cover several styles. They used KNN, EFM-KNN (or PCA), and SVM Linear. The authors used 25 images from each class to train the classifiers and the rest for testing. They used OverFeat [95], a CNN similar to the one used in ImageNet [18]. With OverFeat, there are eight stages: the first six involve convolution and grouping and the last involves connected layers. The highest accuracy the authors obtained for artist classification was 45%; for style classification, the highest accuracy was obtained with the SVM classifier (64.5%). Raw CNN worked best for styles such as baroque, neoclassical, realism, and symbolism. Tag-based representation won in categories such as constructivism, cubism, and surrealism.

Baumer and Chen [96] used CNN to classify images from a collection of 40,000 digitized artworks by artist, genre, and location. They pre-processed and reduced the samples and then used a modified VGGNet architecture [68] for training. The accuracy was 62.2% for artist classification and 68.5% for genre classification.

Bianco et al. [97] proposed a novel deep multibranch neural network that uses different scales of the image to automatically predict a painting's artist and style. They applied the network to the Painting-91 dataset [80]. For artist recognition, they used approximately 50% of the images in the dataset, while for style recognition they used only 2,338. They obtained an accuracy of 78.8% and 85% for the classification of artists and styles, respectively.

Lecoutre et al. [98] used Deep Residual Neural Network to detect the artistic style of paintings. They applied it to the Wikipaintings dataset, a set of images from WikiArt [78] using the techniques of Karayev et al. [73] and Tan et al. [91]. They used AlexNet [18] and ResNet50 [99] for training. To increase the dataset, they distorted the input images in different ways: by flipping horizontally, rotating, moving axially, and zooming. They achieved better than 60% accuracy in the selection of data from Wikiart [78]. To evaluate the generality of the identified styles, they used an additional dataset from an independent source (ErgSap [100]) that contains almost 6000 paintings. To make the datasets compatible, they removed the classes that are not represented in both. The accuracy results for each art style are provided in Table 2.

Mao et al. [101] presented DeepArt [102], a predictive model of artistic styles that uses CNN to simultaneously assess the content and style of works of art. They used Art500K [103], a set of data collected from the Rijksmuseum [81], Google Arts and Culture [104], WikiArt [78] and Web Gallery of Art [105]. The dataset includes more than 500,000 digitized artworks with labels classifying features such as artist, genre, art movement, event, historical figure and description. It recognizes five sets of categories: origin (West/East), art movement (55 classes), artist (1000 classes), genre (42 classes), and medium (112 classes). To evaluate the results of the experiment, the authors calculated the precision (90%) and the normalized discounted cumulative gain (0.970).

Strezoski and Worring [106] proposed a method of recognizing works of art based on artist, period of creation, type of work of art and style, with a focus on the recognition of artists. They used the Rijksmuseum collection [81], the Met collection [107] and the Web Gallery of Art collection [105] to create the Omniart dataset [108], which features 432,217 photographic reproductions of works of art. They removed ambiguous labels such as anonymous or unknown and added new ones: IconClass [109], ColorCodes, current location, actual size and geography, origins, and techniques. They experimented with the best-performing deep architectures on ImageNet, including ResNet-50 [99], VGG-16 [68], VGG-19 [68] and Inception-v2 [110]. The best results were obtained with ResNet-50. Their results show that as the sample threshold decreases, the accuracy increases: for 100 artists the accuracy was 78.5%; for 200 artists, 74%; for 300 artists, 70.8%; and for the total dataset, 52.2%. The other results can be seen in Table 3.

Hicsonmez et al. [111] exploited the CNNs to categorize illustrations according to the style of their illustrator. The dataset used contains 223 books and includes a total of 6,468 images from 24 illustrators. CNN uses the multiple training models: AlexNet [18], VGG-19 [68] and GoogLeNet [110]. The best result was obtained with GoogLeNet—94.07% accuracy—versus 93.47% for VGG-19 and 68.75% for AlexNet. The authors carried out experiments with VGG-19 and GoogLeNet to categorize illustrators per page and per book and found that GoogLeNet offered better results for page cataloging (79.27% vs.

Table 2 Accuracy of the prediction of artistic styles in the experiments of Lecoutre et al. [98] for each of the 25 styles they examined

Style	Accuracy	Style	Accuracy
Ukiyo-e	99.695	Minimalism	99.146
Color field painting	99.110	Early renaissance	99.061
Magic realism	98.927	High renaissance	98.915
Art informel	98.793	Pop art	98.781
Mannerism (late renaissance)	98.671	Abstract art	98.427
Naive art (primitivism)	98.220	Rococo	98.208
Northern renaissance	98.098	Cubism	97.976
Neoclassicism	97.964	Abstract expressionism	97.891
Baroque	96.878	Art nouveau (modern)	96.196
Symbolism	96.000	Surrealism	93.989
Post-impressionism	93.806	Romanticism	93.513
Expressionism	93.281	Impressionism	91.623
Realism	89.745		

Table 3 Results of the experiments of Strezoski and Worring [106] with Omniart

Type of estimation	Number of artists	Result
Artist attribution accuracy (%)	390	64.5%
	87	80.8%
	23	87.5%
	8	94.1%
Type prediction mAP (%)	112	99.4%
	75	99.7%
	39	98.8%
	21	97.9%
Material prediction mAP (%)	1424	99%
	803	98.8%
	94	85.5%
	63	76.8%
Period estimation: means abs. error (years)	544	77.9
	510	67.8
	358	52.2
	237	28.5

78.96%) and VGG-19 offered better results for book cataloging (90% vs. 88.33%). Experiments were also carried out with the use of SVM (combined with Dense SIFT [112, 113] and Color Dense SIFT [114]), with inferior results to those of the neural networks (except AlexNet), obtaining accuracies of 82.71% and 84.35%, respectively. They also carried out style transfer experiments with GoogLeNet, finding good results for all but one illustrator. The authors indicated they believe this is because that individual uses a wide variety of styles in his work.

In the next year, Rodríguez et al. [115] described a method of style classification based on transfer learning and classification of sub-regions or patches of the painting. The experimental validation was based on two art classification datasets (the first has 30,870 images from Wikiart [78] and the second 19,320 images from Pandora 18K [116, 117]) and six pre-trained CNNs (AlexNet [18], VGG-16 [68], VGG-19 [68], GoogLeNet [110], ResNet-50 [99] and Inception-v3 [118]). In both cases, 80% of the data were used for CNN training and 20% for testing. The most accurate results were obtained with Inception-v3 training and the worst with AlexNet.

Finally, Hua et al. [119] presented an image classifier for sorting paintings by artist. They used a CNN to determine the class label and Markov random fields (MRF) to help model the relationship between patches in an image, without the requirement to determine the order of the patches. The proposed CNN-MRF method was applied to the PaintingDB dataset (it do not currently publish), which comprises 1,300 images of paintings from 13 European artists (100 images per artist: 80 for training and 20 for testing) including Goya, Monet and Klee. The evaluation used 20,405 images classified by artist from 23 painters (WikiArt [78]) and the previous dataset. In 2016, this same dataset was used by Jangtjik et al. [120] for an artist classification using a CNN and a weighted fusion scheme to adaptively combine the decisions. They obtained an 88.08% score for recall. The results obtained by Hua et al. when testing the approach on the WikiArt dataset were 72.89% accuracy, 67.40% recall, and an F-score of 70.04%. The results on the trained dataset were 76.92% accuracy, 77.31% recall, and 77.12% F-score.

3.2 Classification of drawings

Elgammal et al. [121] proposed a computational approach for analyzing strokes in line drawings to attribute them to individual artists. Their aim was to facilitate the attribution of artists' drawings to make forgery more difficult. They carried out experiments on a dataset of 300 digitized drawings with more than 80,000 different strokes by artists including Pablo Picasso, Henry Matisse, and Egon Schiele. Their approach succeeded in classifying individual strokes with an accuracy of 70–90%, and recognizing them in drawings with an accuracy of more than 80%. Chen and Deng [122] compared the effectiveness of using SVM and CNN for art classification. They used a set of 7462 paintings by 15 artists, with 5982 images used for training, 737 for validation, and 743 for testing. They tested six different SVM feature classifiers (GIST descriptors, Hu moments, color histograms, SIFT keypoints, histogram of oriented gradient, and Haralick textures). The best results were obtained with GIST descriptors, Hu moments, and color histograms. The training dataset for the CNN was increased with the use of rotation, zoom, flip, and slicing of the images, resulting in 23,928 training images. The best result was obtained from the CNN, with 74.7% accuracy, with 68.1% of the best result from SVM. Sandoval et al. [123] presented a new approach to image classification. They used three datasets: The first includes 30,870 images from 6 artistic styles (Australian aboriginal art, expressionism, impressionism, post impressionism, realism and romanticism). The last five styles were selected from WikiArt and the first was collected manually by volunteers. The second covers a larger number of artistic styles (23, in 26,400 images), adding all WikiArt classes and merging the ones related to cubism into one class. The third contains the 19,329 images of Pandora 18K [116, 117]. The authors used two stages to improve the accuracy of style classification. First, they divided the input image into five parts and applied a DCNN to train and classify each part individually. Then, they merged the five parts into the decision-making module, which applies a shallow neural network (with only one hidden layer) trained by the probability vectors of the first-stage classifier. The method was tested using six pre-trained CNNs (AlexNet [18], VGG-16 [68], VGG-19 [68], GoogLeNet [110], ResNet-50 [99], and Inception-v3 [118]) as the first-stage classifiers, and a shallow neural network as the second-stage classifier. The average accuracy results can see in Table 4. The best results were obtained with ResNet-50 and Inception-v3.

3.3 Classification of comic pages

Young-Min [124] classified comic pages from five different styles of Japanese manga using a CNN similar to AlexNet [18]. The dataset used, Manga 109 [64], comprises 109 volumes of professional Japanese manga, each by a different artist. Color pages and introductions were manually removed, resulting in a total of 823 pages. The average accuracy was 0.86, with accuracies by style ranging from a low of 0.79 (A2) to a high of 0.93 (A3). The author also carried out experiments in style transfer, but those results were unexpected and left for future work. Young-Min [125] later proposed a method for classifying pages of Japanese manga comics using a CNN to classify artistic style. Two different approaches were tried: examining the comic pages and the internal panels of the comic pages. Each image in the Manga 109 [64] was labeled with its author. Young-Min selected eight volumes of comics with different styles with a total of 1330 pages. Free extraction software [126] was used for the classification based on cartoon panels. The panels were also divided into the eight styles for a total of 1417 training images for a modified version of AlexNet. The trained model obtained a mean F1-score of 84% for the classification of full pages and 50% for the classification of panels.

3.4 Classification of architectural works

Yoshimura et al. [127] trained a DCNN model for classification of the designer of architectural works. The model was is trained using photographs of work of 34 architects (recent winners of the Pritzker Prize). The authors used photographs they took themselves and public photographs

Table 4 Average accuracy results of the experiments of Elgammal et al. [121] in each dataset

Dataset	1 (%)	2 (%)	3 (%)
AlexNet [18]	62.46	60.27	73.11
VGG-16 [68]	62.69	62.11	73.61
VGG-19 [68]	62.81	62.49	73.36
GoogleNet [110]	64.41	64.27	74.37
ResNet-50 [99]	66.64	66.02	76.14
Inception-v3 [118]	67.16	66.71	77.53

collected from the Internet, for a total of 19,568. They achieved 73% classification accuracy.

4 Classification based on quality, complexity and visual characteristics

In addition to object detection, ANNs have also been used for the detection, classification, and comparison of visual characteristics [128–131], such as complexity, quality or the presence of common objects. The most recent examples related to the Visual Arts are of the detection, classification, and comparison of stylistic similarities.

4.1 Quality classification

Tian et al. [132] proposed a model for automatic extraction of abstract characteristics through mass training with a DCNN. The dataset was composed of images from the CUHK-PQ [133] and AVA [71] datasets. The network contains five layers: two convolutional and three fully connected; the connected layers contain 16 neurons each. This network was used to classify photos into two classes: high- or low-quality. This quality is based on different aspects like deep/shadow, colorful/monotone, simplicity/complexity or sharpness/blur. The categories in which the images are divided are animal, architecture, human, static, night, landscape, and plant. The authors considered the performance optimal, though 20% of the images were poorly classified.

Later, Wagner et al. [134] applied CNN-based Deep Learning methods for image classification using Objective Image Quality Assessment (IQA). They used Inception-v2 [110] with the ImageNet dataset [18]. The results did not exceed the best single method, KonCept512 (Spearman Order Correlation coefficient [SROCC] of 0.871 vs. 0.908).

4.2 Complexity classification

Machado et al. [135] presented a model of Machine Learning with an ANN for predicting image complexity. They used a dataset of 800 images divided into 5 categories: 252 abstract artistic (AA), 141 abstract non-artistic (AN), 149 representational artistic (RA), 48 representational non-artistic (RN), and 200 photographs of natural and human-made scenes (NHS). They performed feature extraction based on the responses of 240 humans to 800 stimuli, and then trained an ANN using a backpropagation algorithm [136] and a 10-fold cross-validation strategy. The best configuration obtained a mean prediction error of 0.095 and a correlation of 0.833 (normalized intervals from 0 to 1). The metrics with the best performance were edge density and compression error.

4.3 Classification of visual characteristics

Denzler et al. [137] applied Deep Learning to the analysis of common statistical properties used in the Visual Arts for understanding aesthetic perceptions. They used as representative sample of the category "art," the entire JenAesthetic dataset [138] (1625 paintings by 410 artists from 11 different periods/styles), in which 1047 paintings have two labels and 425 have three. The categories are abstract, portrait of a person, portrait of many people, nudes, port or coast, sky, seascape, still life, animals, flowers or vegetation, urban scene, building, interior scene, and other themes (see information in Table 9). For the non-art category, they used 175 photographs of building facades, 528 of entire buildings, 225 of urban scenes, and included a dataset from Redies [138–140] (289 photographs of the natural landscape, 289 of vegetation and 316 of plant details). They used AlexNet [18] as the architecture, and trained several models. The first model is called imagenet_CNN, since it is used for image recognition. This model was trained with 1.5 million images and 1000 common categories of objects [138]. The second is called places_CNN, and was trained for scene-based cataloging with 7 million images from 205 categories. The third, natural_CNN, classifies nature scenes and was trained with 125,000 images from 128 categories. It stands out that natural_CNN obtained an accuracy of 70%. Denzler et al. carried out a study with the classified images to check the capacity of differentiation between photographs and pictorial works (art vs. non-art). Of the three sets they trained, those that show the greatest difference between art and non-art are those of imagenet_CNN and places_CNN, increasing up to the fourth layer of convolution. It is not suitable for nature images. Finally, they analyzed the change in specific properties of images when transfer learning methods were applied to transform them into artistic works. They concluded that the transfer of images to art is directly related to the transfer of intrinsic properties of the images, such as self-similarity.

Carballal et al. [141] used an ANN to distinguish paintings from photographs with edge detection, comprehension, and entropy estimation methods closely related to the complexity of the illustrations. They used two different types of image: 2625 National Geographic photographs (nature, animals, landscapes, documentation, and abstract photographs) and 2610 paintings (including artists such as Caravaggio, Kandinsky, Picasso, Van Gogh and Dalí). The results indicate that these estimates achieve better values than previous results based on perceptual borders, texture, and color. The ANN that provided the best results uses filters and the full set of metrics. The success rate was balanced between sets: 94.67% for paintings and 94.97% for photographs.

Later, Prasad et al. [142] proposed a CNN for the classification of flower images. They tested different architectures to obtain greater accuracy using a database of 9500 flower photographs for experimentation and categorized them into four types (single flower with good lighting; single flower with poor lighting; flower along with leaves; and images with several of the same flowers). CNN training was carried out in five batches and tested on all sets, with a maximum accuracy of 97.78%.

Collomosse et al. [143] proposed an automated search engine for graphics, paintings, and drawings based on the measurement of similarities. The network they implemented consists of three branches that augment GoogLeNet [110], each adding an inner-product layer. They used 65 million contemporary artworks from the Behance website [43]. The artworks were annotated with seven semantic categories (bicycle, tree, cat, bird, car, dog, flower, people), different artistic media (3D, comics, pencil or pencil sketches, oil paintings, vector images and watercolors) and four emotional categories for the viewer (happiness, sadness, peace and fear). They obtained a 90% accuracy in their labels with the TU-Berlin dataset [144], which contains 20,000 sketches divided into 250 categories (key, chair, pineapple, bear...) classified by 1350 people through Amazon Mechanical Turk, a crowdsourcing marketplace to realize online surveys. See information on the dataset in Table 9. They demonstrated "that learning a projection of structure and style features through this network further enhances retrieval accuracy, evaluating performance against baselines in a large-scale (Amazon Mechanical Turk) experiment."

In the next year, Lu [145] proposed a technique called Deformable Convolutional Networks for classification of sketches (DeepSketch) using the dataset TU-Berlin [144]. The number of images used for training is 16,000 for validation and 2000 for testing. The model proposed is a CNN with 8 layers (5 convolutional, 1 deformable, and 2 connected). In the test it obtained an accuracy of 62.5%. Later it reached results of 75.4% (DeepSketch) and 77.7% (DeepSketch2, which considers the order of strokes).

Shen et al. [146] developed a method for discovering similar patterns in art collections and reproducing them as accurately as possible. They used the pre-trained features of ImageNet [18] to obtain matches, and the Brueghel dataset [147], which contains 1587 works of art made by different media (ink, chalk, watercolor, oil...), with different materials (paper, panel, copper...) and with a wide variety of scenes (landscape, religion, still life, etc.) for training. See information in Table 9. They selected the 10 details most repeated in the dataset in collaboration with art historians (Fig. 3). They annotated the images using the VGG Image Annotator tool [148], and used the DocExplore dataset [149], which detects repeated patterns in manuscripts, to validate the detection approach. They also tested the performance of the algorithm for object recognition on photographs. For the photo test, they used the LTLL dataset [150], which contains 225 historical and 275 modern photos from 25 locations. DocExplore has 1500 images and 35 tags, but only considers 18 of them. They also evaluated its performance on the Oxford5K dataset [151], which contains 5,062 images and 11 different tags. To show the generality of their approach, they used paintings by other artists from the WikiArt dataset [78]: 378 paintings by Peter Paul Rubens, 195 by Dante Gabriel Rossetti and 166 by Canaletto (see information about the dataset in Table 9). The best results for the detection of the DocExplore dataset were 75.3% cosine similarity and 76.4% discovery score with the Brueghel dataset training [147]. The LTLL and Oxford5K datasets were used for visual pattern recognition, With LTLL training the classification accuracy results were 88.5% (LTLL) and 83.6% (Oxford5K), while with Oxford5K training, the classification accuracy results were 85.6% (LTLL) and 85.7% (Oxford5K).

Finally, Castellano and Vessio [152] presented a framework based on a DCNN for the extraction of visual characteristics from digitized paintings to search for paintings of a similar style given a starting painting. The tool they proposed learns visual attributes through a CNN with the VGG-16 [68] network trained through ImageNet [18]; the network is capable of building a hierarchy of visual features. The results obtained were of too-high dimensionality (25,088 dimensions), so they used PCA to reduce it. They tested the proposed method with a dataset of 8446 paintings from 50 very different popular painters (such as Giottodi Bondone, Leonardo da Vinci, Michelangelo, Pablo Picasso, or Salvador Dalí) provided by the Kaggle platform [153]. Three examples of the results are shown in Fig. 4. They found similarities showing stylistic influences between works by different painters.

5 Evaluation based on photo quality or aesthetics

We include two closely related approaches in this section: photographic quality and aesthetic value. Though they are different measures, they are often related, and both are subjective values. For example, in the Photo.net dataset [154], the aesthetic value of a series of images is measured by online voting. The article in which Photo.net dataset is presented states that two measures are obtained on the web, of "originality" and "aesthetics," and that these are highly correlated. Thus, it is difficult to assume that visitors are not considering photographic quality in addition to

Fig. 3 10 categories noted in the Brueghel dataset [147] for the experiments of Shen et al. [146]

Fig. 4 Examples of search results for similar works from the experiments of Castellano and Vessio [152]

aesthetic value when voting on just one. The DPChallenge [155] dataset is used as an aesthetic dataset, while the website collect votes from its users.

We will first present the most relevant datasets, and then dedicate two subsections to work focused on photographic quality and work focused on aesthetic evaluation.

Of the datasets used for evaluation, the best known are Photo.net [154], DPChallenge [155] and AVA [71]. More information about each can be found in Table 9.

Photo.net has been described by Datta et al. [154]. In their original article, Datta et al. used SVM with a series of ad hoc metrics to obtain aesthetic classifications of 0.6877 (accuracy), 0.8089 (AUROC), and 0.6890 (Pearson). Several later articles outlined in Table 5 used different combinations of metrics to successively improve upon those results.

Ke et al. [155] obtained a 27.8% error rate when applying the Bayes Naive technique to a DPChallenge.com dataset. The study shows blurring as the most discriminating quality metric, and it achieved essentially the same results Tong et al. [160] obtained using a smaller dataset (27.8% vs. 27.7% error rate). Ke et al. [155] reduced the error rate to 24% by training all features in combination with Real-AdaBoost [161], with a classification accuracy of 72%. Later, in 2008, Luo and Tang [162] used AdaBoost for this task; the methodology they used extracted the subject region from a photo and then formulated semantic characteristics based on that subject and the background division. This resulted in a classification rate is 93%, and for web image search reclassification and the accuracy of the photo and video is over 95%.

Murray et al. [71] introduced the AVA dataset. AVA contains 3,581 images from Photo.net, 12,000 images from CUHK (DPChallenge.com [155]), 17,613 images from CUHK-PQ [133] and the MIRFLICKR dataset (containing 1 million images). The AVA dataset was created to combat the problems described by Wu et al. [163]. Murray et al. [71] trained eight independent SVMs for each semantic category and obtained a mAP of 53.85%.

5.1 Quality evaluation

The most relevant work on photographic quality assessment has been by Tan et al. [164], Gao et al. [165], Meng et al. [166], Talebi and Milanfar [167] and Zhang et al. [168].

Tan et al. [164] presented a method of photographic quality assessment based on an ANN combined with an automatic encoder for the prediction of aesthetic assessment of high- and low-quality photography. The work was divided into three phases: image collection, feature extraction, and training. They used the datasets DPChallenge [155] and Photo.net [154]. The extracted features were frequency-tuned saliency region: local features (rule of thirds, top5-patches and visual attention center), and global features (aspect ratio, brightness, saturation, dark channel, Kolmogorov complexity, NSCT texture, wavelet-based texture, depth of field and contrast). The ANN used backpropagation (BP), a type of direct multilayer network (with input, output, and implication layers). The improved BP-ANN of this experiment contains one input layer, one output layer, and three hidden layers, to improve classification accuracy. It was then combined with an autoencoder—a feedforward, non-recurrent neural network. The average accuracy was 82.1%, with 84.6% accuracy for high-quality images and 79.7% for low-quality images.

Verkoelen et al. [169] trained Restricted Boltzmann Machines (RBM) [170] with Deep Neural Networks for image classification. They used Exactitude's dataset [171], which contains 154 series of portraits of people. They applied Auto-encoding Neural Network models to reduce the dimensionality of the data. The resulting series was subjected to different experiments: straight paths between feature vectors pairs, random points in feature spaces, activation distribution per feature, highest and lowest activating portraits per feature, single feature variations in portrait context, single activation features vectors, portrait distribution over features space, feature pair visualizations, series classification and best and worst classifiable portrait. The most outstanding results were those of the last two experiments focusing on portrait classification; the average classification correctness was 0.622.

Gao et al. [165] proposed a system, DeepSim, that performs an aesthetic evaluation using a ConvNet model [68] trained for the classification of objects. The test and reference images are fed to the VGGnet system separately, creating a feature. The system calculates the local similarities between the feature maps and groups them to obtain an overall quality score and scale. Their experiments used

Table 5 Summary of articles that have sought to improve upon Datta's [154] results with Photo.net

Author/s	Method	AUROC	Accuracy	Pearson
Wong and Low [156]	SVM	0.8590	0.7367	–
Marchesotti et al. [157]	SVM with SGD	–	0.7585	–
Wang et al. [158]	SVM	0.8956	0.8240	0.7900
Xia et al. [159]	GSP-GMM	–	0.8614	–

the four largest IQA databases: Categorical Subjective Image Quality (CSIQ) [172], Laboratory of Image and Video Evaluation (LIVE) [173], LIVE Multiply Distorted (LIVEMD) [174] and Tampere Image Database 2013 (TID2013) [175]. Each database contains several images and an average subjective quality score assigned by several subjects (MOS) or the MOS difference and perfect quality score (DMOS). See information about datasets in Table 9. Spearman's rank correlation coefficient (SRCC) results show an average value of 0.904 and a weight average of 0.884.

Meng et al. [166] propose the used of various levels of CNN to learn models for the aesthetic evaluation of images. Their system extracts features from several layers and adds them for score prediction. They created 3 Multilayer Aggregation Networks (MLANs) based on several reference networks (MobileNet [176], VGG-16 [68] and Inception-v3 [118]) and applied the result to the AVA dataset [71]. The best result they obtained were an accuracy of 79.38% (Inception-v2).

Talebi and Milanfar [167] used a CNN called Neural Image Assessment (NIMA) to predict the average aesthetics score for images. They explored different classification architectures (VGG-16 [68], Inception-v2 [110, 177] and MobileNet [176]) to assess image quality in this task, selecting Inception-v2 as the most appropriate architecture. They trained two separate models in AVA [71], TID2013 [175] and LIVE [173], using 20% of the set for testing. The correlation results of NIMA (Inception-v2) with the training and testing models for the different datasets are shown in Tables 6 and 7.

Zhang et al. [168] proposed a Deep Bilinear model for Blind Image Quality Assessment (BIQA). The model has two CNN, each specialized for a distortion scenario. The first is pre-trained using large-scale training data to classify the type of distortion. The second is pre-trained for image classification. The two CNNs were grouped for final quality prediction, and experiments were performed on three synthetic and distorted IQA datasets (LIVE [178], CSIQ [172] and TID2013 [179]). The CNN used was VGG-16 [68] pre-trained with ImageNet [18]. The resulting Deep Bilinear CNN (DB-CNN) obtained the best results when training with TID-2013: 0.891 (LIVE), 0.809 (CSIQ), and 0.457 (LIVE Challenge). Tests were conducted using the Waterloo Exploration Database [180], a collection of 4744 pristine natural images classified into seven categories (human, animal, plant, landscape, cityscape, still-life, and transportation). The best results were obtained with the DB-CNN trained using SCRATCH for classification of distortion types without taking into account the level of distortion. These results were an average of 0.968 (LIVE), 0.946 (CSIQ), 0.816 (TID2013) and 0.851 (LIVE Challenge).

Table 7 Spearman's rank correlation coefficient between NIMA [167] and the testing and training models for the LIVE [173], TID2013 [175] and AVA [71] datasets

Train dataset	LIVE [173]	TID2013 [175]	AVA [71]	Average
LIVE	0.698	0.547	0.238	0.491
TID2013	0.178	0.827	0.101	0.369
AVA	0.552	0.514	0.636	0.567

5.2 Aesthetic evaluation

ANN have been widely used for the evaluation of aesthetics. For example, Carballal et al. [181] described a series of characteristics that allow us to estimate the complexity of an image as a whole; of the elements that make it up; and of its focus. They used these characteristics to evaluate the aesthetic composition of landscapes and videos. To do so, they used a neural network as a classifier based on ad hoc characteristics, achieving an accuracy of over 85% in an aesthetic composition binary classification task for image and video.

Lu et al. [182] presented RAPID (RAting PIctorial aesthetics using Deep Learning), a method for assessing image aesthetics with a Single-Column Deep Convolutional Neural Network (SCNN). They used the AVA dataset [71] to train their SCNN, and proposed a Double-Column DCNN architecture to recognize the global aesthetic characteristics of images. They also employed two different attributes to perform the categorization: semantic attributes and style. They present a categorization approach through the Regularized Double-Column Deep

Table 6 Linear correlation coefficient between NIMA [167] and the testing and training models for the LIVE [173], TID2013 [175] and AVA [71] datasets

Training dataset	LIVE [173]	ID2013 [175]	AVA [71]	Average
LIVE	0.637	0.327	0.200	0.388
TID2013	0.155	0.750	0.087	0.331
AVA	0.543	0.432	0.612	0.529

Convolutional Neural Network (RDCNN). For the categorization of aesthetic quality, they are based on the aspect with medium accuracy (mAP), where they reach 56.81% compared to AVA [71], which reached 53.85%. The results are presented in Table 8.

Based on their study mentioned above [182], Lu et al. [72] presented a new dataset with 1.5 million images, IAD, in 2015. It comprises 300K images from DPChallenge [155] and 1.2 million from Photo.net [154]. The training results for this dataset were divided into two categories (high and low esthetics), with 747K and 696K, respectively. They evaluated the results with the AVA dataset [71]. The best results for the SCNN were 73.85%, and the DCNN achieved 74.6% accuracy—both improvements over the AVA training set (73.25%). They also tested an alternative strategy, using the top 20% rated images as positive samples and the bottom 20% images as negative samples; with this approach, the best accuracy of the SCNN was 72.65% and the DCNN, 72.9%, without optimal results compared to large-scale IAD data.

In the same year, Zhou et al. [183] presented a method based on Deep Learning for the aesthetic evaluation of photographs. They combined the use of an ANN with an autoencoder. The dataset they used contains 3,581 photographs from Photo.net [154] and 28,896 photographs from DPChallenge.com [155]. The classification accuracy was 82.14%.

Dong and Tian [184] presented a method of aesthetic evaluation in which they used ANN; they used SVM as a photographic quality classifier for each feature type (color, subject-background contrast, sharpness, depth of field, and image size) combined with the RBF kernel function. For high dimensional descriptors, they applied a 4096-d DCNN descriptor and 1024-d Dense SIFT descriptor [112, 113]. The tests were performed on two datasets: CUHK-PQ [133] and AVA [71]. The performance was better for the CUHK-PQ dataset. With DCNN descriptors and rule-based features, it work well in the AVA dataset. They also performed tests on a multi-level dataset by dividing the images into "good" and "bad." In this case, the DCNN achieved the best performance with an overall accuracy of 73.59% versus the 66.72% obtained by the rule-based approach. The direct concatenation of all functions achieved 75.9% accuracy. Subsequently, three different MKL-based feature fusion schemes were applied and obtained better results for accuracy, of 76.05–77.21. The MKL-based feature fusion schemes also offered good recognition accuracy: 73.89% for bad images, 64.17% for common images, and 88.52% for good images.

Campbell et al. [185] proposed a classifier to divide images into two groups—with high and low aesthetic value—based on images created by IMAGENE (an art generation machine based on genetic programming) [186]. Their models were trained using a separate Boltzmann Restricted Machine [170] for each dataset (high value and low value); they then joined the datasets, and trained them again with the Restricted Boltzmann Machine. They analyzed the results and trained a Deep Belief model with 10 layers. The highest classification accuracy of the learned features was achieved in the second hidden layer: 84%.

Wang et al. [187] presented a Multi-Scene Deep Learning Model (MSDLM) for aesthetic evaluation. They established Alex_CNN [18] on the first 4 layers of the network (previously trained with an ImageNet dataset [18]). They designed a scene convolution layer for the descriptors to distinguish between seven categories (animal, architecture, human, landscape, night, plant and static) of CUHK-PQ [133]. Images were randomly divided into six groups of similar size: four for training, one for validation, and one for testing. The authors added images and rotated the highest quality images by 90° and 270° so that the whole dataset was balanced into high- and low-quality images for training. The total accuracy for their experiments was 0.9259. They also trained the model and enabled it in the AVA dataset [71] through two experiments: one with 51,106 randomly divided images and another with 74,673 low-quality and 180,856 high-quality images. The accuracies were 84.88% and 76.94%, respectively.

Jin et al. [188] proposed a new DCNN structure, ILG-Net, for the aesthetic classification of images. The structure introduced an Inception module and connected the local intermediate layers to the global one. The authors used GoogLeNet, [110]. For verification, Jin et al. used the same subsets of the AVA dataset as Wang et al. [187]. Jin et al. obtained accuracies of 85.62% and 79.25%.

Kao et al. [189] proposed a framework for evaluating the aesthetics of images. They divided the images into three categories (scene, object, and texture) and trained a CNN associated with each category. They also developed an A&C CNN to simultaneously evaluate aesthetic quality

Table 8 Accuracy results of the experiments of Lu et al. [72, 182] with the AVA dataset [71]

δ	AVA (%)	SCNN (%)	AVG_SCNN (%)	DCNN (%)	RDCNN_style (%)	RDCNN_semantic (%)
0	66.7	71.20	69.91	73.25	74.46	75.42
1	67	68.63	71.26	73.05	73.70	74.2

and categorize. The classification and regression models were developed separately for aesthetic prediction (high or low) and scoring. They used 5000 images from the AVA dataset [71] and obtained overall 91.3% accuracy. The accuracy for each category is 76.04% (scene), 73.30% (object) and 71.6% (texture), with the accuracy for the regression method 74.51%. They also experimented with using category training to predict categorization.

Kao et al. [190] proposed a method of aesthetic evaluation using Multi-Task Deep Learning (MTCNN). They used the AVA dataset [71], with δ of 0 and 1 with different values of λ in their MTCNN training algorithm to check the accuracy they achieve. As best value they get $\lambda = 1/29$ with a $\delta = 0$ of 76.15% accuracy and a $\delta = 1$ of 75.90%. They reach an accuracy of 76,58% ($\delta = 0$) and 76,04% ($\delta = 1$).

Malu et al. [191] used Deep CNN for automatic evaluation of aesthetics; they used the Deep Residual Network (ResNet60) for training. The network evaluates eight aesthetic attributes (balancing elements, content, color harmony, depth of field, light, object, rule of thirds, and vivid color) along with the overall aesthetic score. They used the dataset AADB [192] for training and testing. The total correlation result (ρ) obtained was 0.689.

Tan et al. [193] presented an aesthetic photo classifier with a DCNN based on GoogLeNet [110] for aesthetic quality classification. They used images from DPChallenge.com [155]: 22,104 photographs from the category of landscapes and 28,913 photographs from the nature gallery, all voted on by 100 users. They manually deleted images until they had 20,114 photographs landscapes and 27,190 nature photographs. They built two groups of images, of high and low-quality, respectively, and considered a value of $\delta = 1.0$. Images with a rating of $\geq 5.5 + \delta/2$ are considered very good and those with a rating of $\leq 5.5 - \delta/2$ are considered very bad, since there are few above 8 and below 3. The classification accuracy of the method was 87.10%.

Li et al. [194] proposed an Embedded Learning Convolutional Neural Network (ELCNN) that uses an image's content to evaluate its aesthetic quality. They compared its performance with that of AlexNet [18] and VGG_S, two other Deep Learning methods. They carried out experiments on the CUHK-PQ [133] dataset, which has 17,613 images classified into animal, architecture, human, landscape, night, plant, and static. All of these images have been manually labeled as high- or low-quality. The authors used ImageNet [18] to train the images. They used 300 low-quality and 100 high-quality images for testing, leaving the rest for training. The classification accuracy for each of the categories was 0.9712 (animal), 0.9325 (architecture), 0.9660 (human), 0.9520 (landscape), 0.9300 (night), 0.9360 (plant) and 0.9350 (static).

Lemachand [195] applied methods for computational aesthetic quality classification of photographs to video content. Their method extracts features of orientation distribution, curvature distribution, HSB color distribution (hue, saturation, brightness), and reflectional symmetry on the cardinal and diagonal axes. It then employs a Deep Neural Network composed of 3 hidden layers to learn visual preferences. The results may be distorted, because they do not consider the audio and movement content, since the learning is based on images from the AVA dataset [71]. The author used a series of Youtube videos from the dataset of Tzelepis et al. [196], which contains 700 short videos evaluated by five people, as a test set. This test set was used to analyze the evolution of aesthetics in feature films by highlighting interesting patterns related to filmmakers' decisions. Wes Anderson's symmetry approach obtained good scores on this metric with results like 56.16% (The Great Budapest Hotel), 22% (Moonrise Kingdom), 20.60% (Fantastic Mr. Fox) and 58.83% (The Royal Tenenbaums). In contrast, Stanley Kubrick's films scored low on this metric; 12.10% (Full Metal Jacket), 17.75% (A Clockwork Orange), 14.12% (The Shining), and 16.21% (Space Odyssey).

Murray and Gordo [197] focused on training aesthetic prediction models and presented a model called APM. They used a CNN trained end-to-end to predict aesthetic scores, and a network based on ResNet101 [198] that preserves the 1000 semantic categories used for ImageNet classification [18], for training. They used the AVA dataset [71], and their best results were 80.3% accuracy.

Bianco et al. [199] proposed an aesthetic evaluation model using a CNN. They used the AVA dataset [71], and the Caffe network architecture [76] (inspired by AlexNet [18]), but modified and adjusted it for their purposes. They replaced the last connected layer with a single neuron layer to produce an aesthetic score (as a value between 1 and 10). The network used was Hybrid-CNN [200], with the original training combining the scene categories from the Places datasets [200] and the object categories from ImageNet [18], for a total of 1183 classes. They measured the results with MRSSE, reporting their best result was 0.3727. Some 99% of their predictions had an error less than or equal to the standard deviation.

Lemarchand [201] proposed a system for the aesthetic classification of photographs into datasets. It uses two CNNs that are trained separately and then converge into a common final layer. Finally, it uses one CNN as a control and feeds it raw RGB images. The CNN uses hyperparameters similar to other CNNs, with a learning rate of 0.001 and a dropout probability of 0.5. Its AI system extracts percentage distributions of curvature, metric, orientation and color, knowing only the aesthetic judgments of people of the images (preference for the color blue, presence of horizontal lines...). The author used the Photo.net [154] and AVA [71] (with 4,000 items and complete) datasets to test the network. The accuracy results for the classification of low-quality images achieved 69.42%. For high-quality images, the accuracy ratings achieved 70.23%. The average rating accuracy results were 61.43%, 58.01%, and 69.83%, which do not exceed those of Lu et al. [182].

Zhang et al. [202] proposed a CNN model for classifying images according to their aesthetics. They use 250,000 images from the AVA dataset [71] (230,000 for training and 20,000 for testing). They used Global Average Precision (GAP) to create the Aesthetic Activation Map (AesAM) and the Attribute Activation Map (AttAM). They created a single branch aesthetic prediction module named AesCNN, and added attributes to it to create AesAttCNN. This second module did not obtain optimal results (according to the authors), so they create others with weights based on their score: AesCNN-W and AesAttCNN-W. The models they used were AlexNet [18] and VGG-16 [68]. The accuracy for the weighted modules was 77.39% (AesCNN-W with AlexNet), 78.87% (AesCNN-W with VGG-16), 77.18% (AesAttCNN-W with AlexNet) and 78.62% (AesAttCNN-W with VGG-16). The code and trained model are available online [203].

Jin et al. [204, 205] presented a new dataset that distributes images uniformly by their aesthetics (IDEA). They proposed a spatial aggregation perception neural network architecture. For the creation of a balanced set of aesthetic images (with scores from 0 to 9), they collected 1000 images with each score (except for 9, for which they collected 191) from DPChallenge [155] and Flickr [74]. They used 8191 for the training set and selected 1,000 at random for the test set. The results had a MRSSE of 0.2856 for the AVA dataset [71].

Later, Apostolidis and Mezaris [206] presented a method for evaluating the aesthetic quality of images and used the version of the AVA dataset [71] divided into 2 subsets previously used by other authors [187, 188] for testing. They addressed existing shortcomings by introducing a CNN as a classifier to feed images with the highest possible resolution, maintain the aspect ratio of the input images to avoid distortion, and combine local and global features. The architecture they used is VGG-16 [68], implemented with the Keral Neural Network API [207] to turn it into FCN [208]. The best results in the preliminary tests were obtained with an input size of 336 $\times$ 336 (batch size: 8) without freezing, with an average accuracy for the second AVA dataset of 88.44%. For images of 672 $\times$ 672, the average input inference time was 790 ms, so it was not recommended. Not freezing any layer provides higher accuracy. After these tests, they continued the search for better performance while maintaining the original aspect ratio of images; the highest accuracy they obtained in this case was 89.94% using three clippings from the original image to include the entire image surface in the network. Finally, they added a skip connection, which aimed to combine the output of the final layers with the image in its initial layers, obtaining a top accuracy of 91.01%.

Sheng et al. [209] performed aesthetic image evaluation from the perspective of feature learning without manual annotation. They manipulated images into easily controlled parameters (total loss function, entropy-based weighting, degradation identification loss, and triplet loss), resulting in predictable perceptual quality. The training consists of two parts: a pre-training stage to learn the visual characteristics with unlabeled images and a task adaptation stage to evaluate the aesthetic quality of the images. In the pre-training the first layers follow the structure of AlexNet [18]. The datasets used were AVA [71], AADB [192] and CUHK-PQ [133]. The average performance of the model was 80.02% (AVA), 66–32% (AADB), and 83.36% (CUHK-PQ).

Carballal et al. [210] proposed a new approach that completely automates the process of creating metrics, without the need for human subjectivity. The authors used the Photo.net [154], DPChallenge [155] and AVA [71] databases. Metrics were obtained by transfer learning from the ResNet-50 [99] and GoogLeNet [110] networks. This system was based on the integration of CNN and Correlation by Genetic Search (CGS). The best results for GoogLeNet were 0.9378 (Pearson's correlations), 0.9293 (AUROC) and 0.9315 (accuracy). For ResNet50, the results were 0.9177 (Pearson's correlations), 0.9128 (AUROC) and 0.9162 (accuracy). Therefore, the CNN-CGS obtained the best results for transfer learning from GoogLeNet.

In the next year, Cetinic et al. [211] employed CNN for the prediction of scores related to three subjective aspects of human perception (aesthetic evaluation, evoked feeling and memorability). They used different datasets for training and testing in each category (listed in Table 9). They used the models AlexNet [18], GoogLeNet [110] and ResNet50 [99] for training. They analyzed the WikiArt dataset with their methods for style, genre, artist, and period. The results suggest that image content and lighting have a significant influence on aesthetics, while the emphasis on objects has an impact on recall.

Finally, Dai [212] presented a model to evaluate the aesthetic features of photographs. It focuses on the use of Repetitive Self-Revised Learning (RSRL) to train a CNN-based aesthetics classification network with an unbalanced dataset. The dataset used (xiheAA [213]) has 3100 photographs evaluated by a professional photographer (with scores between 2 and 9). The photographs in the xiheAA dataset were taken by students, and their teachers evaluated them. Also included were 310 images from 500px [214].

Table 9 Table of compilation of the data sets used in the different experiments with ANN addressed throughout the article, which are available online

Data sets
WikiArt [78]
They have used it in both classification and art generation experiments
Type of content: works of art
WikiArt presents 250,000 works of art by 3000 artists. These works of art can be found in museums, universities, city halls, and civic or cultural buildings in more than 100 countries. The works are classified according to their style, genre, and type (watercolor, color pencils, silkscreen, glass, etc.)
The works are available through the website https://www.wikiart.org/
Web Gallery of Art [105]
It has been used by Strezoski and Worring [106] in their method of recognizing works of art according to the artist, period of creation, type of work of art and style, and Mao et al. [101] for their DeepArt model [102] of predicting artistic styles
Type of content: works of art
The contents are divided according to their artist (in alphabetical order) and periods (early Christian art, pre-Romanesque art, and medieval art). The contents according to their author are labeled by their birth-die-campus, the period to which they belong (mannerism, realism, early renaissance, high renaissance, baroque, romanticism, neoclassicism, etc.), the nationality and the profession. It also has two specific databases on decorative and architecture
The access is made directly through the website https://www.wga.hu/index.html
Rijksmusem [81]
Used for classification experiments
Type of content: painting works
Usable content from users of the Rijkmuseum website in https://www.rijksmuseum.nl/ through the API, which allows the use and access to the collection metadata
Berkeley/Brueghel [147]
Used by Shen et al. [146] to discover similar patterns in art collections and reproduce them as accurately as possible
Type of content: painting works
Gathers all the known works attributed to Pieter Bruegel (811)
The access to the works can be carried out through the web http://pieterbruegel.net/
Painting-91 [80]
Type of content: painting works
It is used for classification
Data set of 4266 paintings from 91 different authors labeled according to their painter and style. The styles included in this data set are: expressionism, abstract, baroque, constructivism, cubism, impressionism, neoclassical, pop art, postimpressionism, realism, renaissance, romanticism, surrealism, and symbolism
To access the data set, please send an email to fahad@cvc.uab.es
OmniArt [108]
Type of content: painting works
It was created by Strezoski and Worring [106] with photographs by Rijksmuseum collection [81], the Met collection [107] and the Web Gallery of Art collection [105]
The total of photographic reproductions of works of art that this dataset is 432,217
Google Arts and Culture [104]
It was used by Mao et al. [101] for their DeepArt model [102] of predicting artistic styles
Type of content: photographs and paintings
It has numerous collections from different partners, including museums such as the Momma, the Musée d'Orssay (Paris), or the Van Gogh Museum. The works are classified by collections, themes, artists, date, as well as techniques and artistic trends, and historical events and characters
Access is through the website itself https://artsandculture.google.com/
AVA [71]
Used in classification, evaluation and generation
Type of content: photographs
The entire data set has been split into 64 7z files: About 32 GB and 250,000 images with a variety of metadata, including aesthetic scores for each image, semantic tags with more than 60 categories, and photo style-related tags

Table 9 (continued)

Data sets
How to get them: contact isp@uv.es, via Torrent at http://academictorrents.com/details/71631f83b11d3d79d8f84efe0a7e12f0ac001460 or via Mega at https://mega.nz/folder/hIEhQTLY#RkOnZv8Fz7EbYreHsiEzvA/file/IAUE1YzC
Pandora 18k [116]
The data set was used for classification experiments by Rodríguez et al. [115] and Sandoval et al. [123]
Type of content: photographs
The original data set was created specifically for the location of the center of the head, the position of the head, and the estimation of the position according to the shoulder, so it has images of different subjects with various positions. Pandora contains more than 250 k (1920 × 1080 pixels) and depth images (512 × 424) with their corresponding annotations: 110 tagged sequences using 10 men and 12 women. They have subsets of clips from the images
You can download the files by requesting access to them hosted on Google Drive through an email and institution form on the web https://aimagelab.ing.unimore.it/pandora/
ImageNet [18]
Used for classification, evaluation and generation training by numerous authors
This is a data set that currently has 14197122 images indexed in 21841 syntactic categories (synsets/tags)
Type of content: photographs
The download is available directly from the website http://image-net.org/index
Oxford5K [151]
It was used by the same authors as the Bruegel data set [147], Shen et al. [146], for testing their method
Type of content: photographs. There are 5062 images collected from Flickr, through a search of 17 Oxford landmarks (All Souls Oxford, Balliol Oxford, Christ Church Oxford, Hertford Oxford, Jesus Oxford, Keble Oxford, Magdalen Oxford, New oxford, Oriel Oxford, Trinidad Oxford, Radcliffe Camera Oxford, Cornmarket Oxford, Bodleian Oxford, Pitt Rivers Oxford, Ashmolean Oxford, Worcester Oxford, and Oxford. The images were manually labeled based on 4 different labels: good (nice, clear image of the object/building), OK (more than 25% of the object is visible), bad (the object is not present) and rubbish (less than 25% of the object is visible or there is distortion)
The data set can be downloaded via the link https://www.robots.ox.ac.uk/~vgg/data/oxbuildings/
The same authors created another database: Paris 6K [303] which has 66412 images also collected from Flickr [74] by searching 12 identifiers (La Defense Paris, Eiffel Tower Paris, Hotel des Invalides Paris, Louvre Paris, Moulin Rouge Paris, Musee Rouge Paris, Musee d'Orsay Paris, Notre Dame Paris, Pantheon Paris, Pompidou Paris, Sacre Coeur Paris, Arc de Triomphe Paris and Paris)
The data set can be downloaded via the link https://www.robots.ox.ac.uk/~vgg/data/parisbuildings/
Photo.net [154]
Used in classification and aesthetic evaluation experiments
Type of content: photographs
Used in classification and aesthetic evaluation experiments
This data set has an extensive catalog of photographs of different people divided into 31 categories and with evaluations from other users of the web
The data set can be accessed directly via the website https://www.photo.net
DPChallenge [155]
Type of content: photography
Used in classification and aesthetic evaluation experiments
It is also a data set hosted on the web and has images of various authors divided into 67 categories and with ratings from different users
The data set can be accessed directly via the website https://www.dpchallenge.com
AADB [192]
Type of content: photographs
The data set contains 10,000 images with aesthetic quality ratings and attribute labels provided by five different individual evaluators
The data set is in Google Drive and to download it you only have to enter the following link https://drive.google.com/drive/folders/0BxeylfSgpk1MOVduWGxyVlJFUHM
Places Database [200]
Used by Bianco et al. [199] in their experiment of aesthetic evaluation
Type of content: photographs
It has 2.5k images with a category label per image (205 categories total)
To obtain it you must register on http://places.csail.mit.edu/index.html
Large-scale Scene Understanding (LSUN) [246]

Table 9 (continued)

Data sets
Type of content: photographs
Used by Radford and Metz [245] in their imaging experiment
Set of one million images tagged in 10 scene categories and 20 object categories
Downloadable content at https://www.yf.io/p/lsun
MIT-Adobe FiveK [254]
Type of content: photographs
Used by Tabeli and Milanfar [253] in the training and evaluation of his network for the improvement of photography
It includes 5000 photographs, versioned by 5 experts, and with semantic tags for each one of them
The content is downloadable through the link https://data.csail.mit.edu/graphics/fivek/
Exactitude website [171]
It was used by Verkoelen et al. [169] in their image classification training.
Type of content: portrait photographs
It contains 154 series of 12 portraits of different people
Access is free and direct through the website https://exactitudes.com/collectie/?v=s
LIVE [178] It is used in evaluation
Type of content: distorted images. This is a data set of distorted images with different types of distortion evaluated by human subjects
The public data set has 2 versions available in https://live.ece.utexas.edu/research/quality/subjective.htm
CSIQ [172] It is used in quality evaluation
Type of content: distorted images
The set has 30 original distorted images with 6 different types of distortion (of 4 or 5 different levels of distortion). It contains 5000 subjective classifications of 35 humans
The data set can be downloaded via the link http://vision.eng.shizuoka.ac.jp/mod/page/view.php?id=23
TID2013 [179] It is used for aesthetic classification
Type of content: distorted images. This is an extension of TID2008 and contains 25 reference images and 3000 distorted images, i.e., 34 types of distortion and 5 levels of distortion
It can be downloaded directly by clicking on the link http://www.ponomarenko.info/tid2013/tid2013.rar or from the page http://www.ponomarenko.info/tid2013.htm
TU-Berlin [144]
It was used in the experiments of Lu [145] and Collomosse et al. [143], both for classification
Type of contents: sketches
data set containing 20,000 sketches distributed in 250 object categories
The data set can be obtained by direct download through the url http://cybertron.cg.tu-berlin.de/eitz/projects/classifysketch/
ModelNet [280]
Type of content: 3D CAD models for objects
Used to generate the experiments of the generation of 3D models
This is a selection of several images from 10 popular object categories labeled by category
Both the 10 class version and the full data set can be downloaded in https://modelnet.cs.princeton.edu/#
Manga109 [64]
Used for object detection and classification
Type of content: comics
The data set includes 109 volumes of manga drawn by professional Japanese artists between 1970 and 2020. The comics are cataloged by title, author, year of publication, publisher, target, genre, number of pages, volume, and whether the content is downloadable for commercial use or not
To access the data set you must fill out the form by accessing it from the website http://www.manga109.org/en/download.html
Kaggle [153]
It was used by Castellano and Vessio [152] in a classification experiment.
The website https://www.kaggle.com/datasets has different data sets with different characteristics.
ILSVRC 2012 dataset [262]
It is used in a pre-trained object detection DCNN model that Tanjil and Ross [261] have used to employed to investigate evolutionary computing art generation

Table 9 (continued)

Data sets
Type of contents: photographs
It contains 150.000 photographs of Flickr [74] with 1.000 object categories
The dataset is available here: http://image-net.org/challenges/LSVRC/2012/
The following data sets were used in experiments by Cetinic et al. [211]
FLICKR-AES [304]
Type of content: photographs
It contains 40,000 Flickr images [74] tagged using Amazon Mechanical Turk
Available on Google Drive together with the data set REAL-CUR (contains 14 real user photo albums with aesthetic ratings provided by the owners) through the link https://drive.google.com/drive/folders/1XLlPu_lgHqRstF7DBmXQ2QPSK9KPx1Yu
Twitter DeepSent [305]
Type of content: photographs
It has 1269 Twitter images
The download is available at https://www.cs.rochester.edu/u/qyou/DeepSent/deepsentiment.html
Flickr Sentiment [306]
Type of content: photographs
It has 90139 images with feeling tags downloaded from Flickr [74]
The data set is available at https://mm.doshisha.ac.jp/en/senti/CrossSentiment.html along with a similar data set containing 65,439 images downloaded from Instagram
LaMem [307]
Type of content: photographs
It is a data set ordered according to its memorability value. It has four categories: Emotions, COCO, Abnormal and SUN
It is available on the page http://memorability.csail.mit.edu/download.html
SUN Memorability [138]
Type of content: photographs
It contains 131,067 Images with 908 Scene categories and 313,884 Segmented objects with 4479 Object categories
It is available from https://groups.csail.mit.edu/vision/SUN/
JenAesthetic [211]
Type of content: painting works.
It features 1628 aesthetic paintings from 11 different western art periods. It has 16 different classes in which the images are labeled
The data set is available through a form at http://www.inf-cv.uni-jena.de/en/jenaesthetics
WikiEmotions [308]
Type of contents: artistic pieces
It has a data set of 4105 pieces, mainly pictorial works labeled by the emotion they evoke in the observer. They were selected from the WikiArt data set [78]
It can be downloaded at the following link http://saifmohammad.com/WebPages/wikiartemotions.html
MART [309]
Type of content: abstract art
It contains 100 photographs of abstract paintings
Content can be downloaded at the following link http://disi.unitn.it/~sartori/datasets/ after filling in a form
We can find other useful data sets on websites such as
http://personal.ie.cuhk.edu.hk/~ccloy/datasets.html
https://www.istockphoto.com/es/collaboration/boards/RW2QOGD7IUul6JkfKCQddw
https://www.pexels.com/es-es/
https://unsplash.com/
https://gratisography.com/
https://pixabay.com/es/
https://foter.com/
https://stocksnap.io/
https://freestocks.org/

Table 9 (continued)

Data sets
https://picography.co/
https://www.lifeofpix.com/
https://www.foodiesfeed.com/
https://www.metmuseum.org/art/online-features
https://www.artic.edu/collection
https://cv.iri.upc-csic.es/
https://sketchfab.com

The accuracy of the test dataset for the 10th retrained network was 0.42, but the accuracy of the training network was 0.99.

6 Style transfer

In the last 5 years, many researchers have used ANN and Deep Learning for a task called style transfer. This task consists of transferring the style of an image, or a set of images (for example a pictorial work or the work of a painter) to another image (usually a photograph or a video). In many cases, the style of painting is transferred to photographs [215], although it is also done in other areas, such as comics.

Gatys et al. [216] proposed a model of texture generation from a previous image using a CNN, with VGG-19 [68] used for training. They compared the results with those of Portilla and Simonelli's [217] experiments with steerable pyramid representation and complex analytic filters.

In the same year, Gatys et al. [218] presented one of the first examples of style transfer. They discovered that the representations of content and style of CNN are separable, and therefore, each representation can be manipulated independently to create new images. In their approach, the machine synthesizes a new image showing the content of the photograph and the style representation of the artwork and uses the CNN to create artistic images from a combination of paintings and photographs. To obtain a representation of the style of the painting, a space is created for the capture of texture information. The authors used the VGG-Network [68] in the experiments. An example they showed is the result of combining a photograph by Andreas Praefcke from Neckarfront in Tubinger, Germany with different paintings from several painters (J. M. W Turner, Vincent van Gogh, Edvar Munch, Pablo Picasso and Wassily Kandinsky).

In the next year, Gatys et al. [219] also performed style transfer using a DCNN. The method used separates and recombines the image content and its natural style using the VGG-Network [68]. The algorithm allowed production of new, high-quality images by combining the content of a random photo with known artwork. Some results with good transfer were obtained by combining photographs with the styles of "The Shipwreck of the Minotaur" by J.M.W. Turner (1805), "The Starry Night" by Vincent van Gogh (1889), "The Scream" by Edvard Munch (1893), "Seated Nude" by Pablo Picasso (1910) and "Composition VII" by Wassily Kandinsky. Gatys et al. [220] also presented a model to preserve the original color and luminance in neural artistic style transfer; examples of the results are shown in Fig. 5.

Chen and Hsu [221] proposed a model of Deep Style Transfer inspired by the 2016 work of Champandard [222], which used a CNN to include semantic information in style transfer. They also considered the model of Li and Wang [223], which combines MRFs and DCNN for image synthesis, and that of Gatys et al. [219]. These works that inspired Chen and Hsu used the VGG-19 network [68]; the major difference that stands out in Chen and Hsu's experiments is the use of two restrictions on the network: where and what to transfer. To restrict where to transfer, they proposed a "masking out" process to specify the spatial correspondence. In the case of what to transfer, they proposed new high-order feature statistics to better capture and combine the representation of the style.

Bhautik et al. [224] proposed the use of Neural Style Transfer to transfer the impressionist style from the image "Come Swim" to images of a short film written and directed by Kristen Stewart, which was inspired by the image. The result is shown in Fig. 6.

Chen et al. [225] used a CNN to transform photographs into comics and also trained a CNN classifier that separates photographs from comic book drawings. They were inspired in this work by the previous work of Gatys [219].

(a) Content (b) Style (c) Output using [2]

(d) Output with color preservation (color transfer) (e) Output with colour preservation (luminance matching)

Fig. 5 Results of the experiments for the conservation of color and luminance in neural artistic style transfer by Gatys et al. [220]. **a** Original photograph, **b** Painting by Pablo Picasso ("Seated Nude"). **c** Transformed content image, using the original neural style transfer algorithm. **d** Transformed content image, using color transfer to preserve colors. **e** Transformed content image, using style transfer in the luminance domain to preserve colors

Fig. 6 Result of the use of Neural Style Transfer of the impressionist style to "Come Swim" [224]. Left: content image, middle: style image, right: upsampled result

The CNN model they used is VGG-16 [68] with the Caffe framework [76].

Two years later, Krishnan et al. [226] proposed a model of style transference using CNN for the fusion of multiple paintings by one artist. They also proposed an algorithm for the evaluation of such transference. They used five layers of VGG-19 [68] for training, and a test dataset consisting of eight different artist styles. They compared their results with those of an earlier experiment by Jing et al. [227]; the average confidence score of the authors' method was 0.553, better than the 0.539 of Jing et al.

Other authors, for example Surma [228], have carried out public experiments on image generation and transfer using CNNs. These works have largely not been published in scientific fora, but are accessible through GitHub.

7 Pictorial generation or reconstruction

The last and most difficult task for which AI is used in the Visual Arts is the creation of images. In this section, we address, among others, ANNs used in reconstruction of images, including damaged artwork and 3D model recovery. We also address work to generate new artistic images using different aesthetic adaptation metrics; some of the results are currently exhibited in art museums. Finally, we will analyze some examples of the creation of 3D models and graphic content for games.

7.1 Generation of photographs and paintings

Several works employ Deep Convolutional Generative Adversarial Networks (DCGANs). In a GAN, there are two ANNs trained in parallel: one generates images that meet a criterion (e.g., look real) while the other tries to identify images that don't meet that criterion (i.e., detects images that aren't real).

Previous artistic works have followed a similar automatic generator–evaluator approach, although not all of their components were ANNs. For example, Machado et al. [87] presented an art generation system in which an ANN is used for classification and a Evolutionary Computation for a generation. Their work aimed to generate images with aesthetic value without user input. For this purpose, a reference of 3322 paintings from 14 famous artist is used as a positive reference. A series of random images are generated as a negative reference and a generator based on genetic programming is used to generate new images. A backpropagation neural network is then trained, whose role is to distinguish between all the images generated (including the previous iterations) and the painting dataset. With this newly trained ANN, the genetic programming system is again used to generate a new iteration of images that seek to be classified as paintings and not as generated works. The generation and classification steps are repeated; in each iteration, the system tries to explore a different "style." The ANN obtained significantly better classification results in external datasets after the iterative process than before. This approach has also been used for the classification and generation of faces [229–232], as well as for the generation of ambiguous images [233].

Concerning the image quality assessment we mentioned in the previous section, [234] proposed an attention-driven noise removal CNN (ADNet). The ADNet code can be accessed on GITHUB [235]. Noise reduction of the image will provide improvement in the image quality.

Colton et al. [236] built The Painting Fool, an automated painter. It uses several AI techniques: natural language processing [237], constraint solving [238], evolutionary search [239], design grammars [240] and Machine Learning [241]. The Painting Fool was applied in different artistic contexts, and the resulting work has been exhibited in several art galleries and exhibitions, notably the Amelies Progress gallery, Pencils, Pastels and Paint gallery, and "No Photos Harmed exhibition" [242]. Finally, the authors held an exhibition, "You Can't Know My Mind" [243], which focused on the creative intent of the software. To achieve a system with a more sophisticated sense of appreciation and levels of internationality, they created the ability for the machine to associate visual stimuli with linguistic concepts. For this task, they used part of the DARCI system (a CNN-based visual–linguistic association approach) that contains a public dataset with images tagged by volunteers [244]. From the DARCI system, The Painting Fool collected a set of 236 association networks (ANs) that are equivalent to particular adjectives, and a method of turning an image into the numerical inputs to the ANs. The authors eliminated all ANs that could create controversy for the machine or which had a low trigger value. For example, *red* generated better results for images with green than for clearly red ones. This process resulted in 65 usable ANs, which were used to title images and to compare images based on a given adjective.

Radford and Metz [245] proposed the use of Deep Learning in a GAN configuration for image generation. They carried out training on three datasets: Large-Scale Scene Understanding (LSUN) [246] (an LSUN room dataset, containing over 3 million images),

ImageNet-1k [18] and their new Faces dataset. The ImageNet dataset was used as a source of natural image data for unsupervised training. For their human face dataset, they selected 3 million images from random searches of names in DBpedia [247]. They ran the images through the OpenCV face detector to obtain the highest resolution images, retaining 350,000 face images. To evaluate the performance of their machine based on classification, they trained using the ImageNet database [18] with the image classifiers of CIFAR-10 [248] and achieved an accuracy of 82.8%. They then tested DCGAN for supervised purposes when labeling was poor, using the same characteristics as the previous experiment, and obtained a best result (for the classification with 1000 labels) of a test error of 22.48%, after which they carried out generation experiments.

In the next year, Tan et al. [249] proposed an extension to the GANs called ArtGAN. They used the WikiArt dataset [78], reserving 30% for testing and using the rest for training. They performed three trainings, based on characteristics of genre, artist, and style. For the tests of natural image generation, they used their training and that of DCGAN [245] with the dataset CIFAR-10 [248]. The results of the models using log-likelihood measured by the Parzen-window estimate were 2348 ± 67 (DCGAN) and 2564 ± 67 (ArtGAN).

Elgammal et al. [250] proposed a new art generation system, CAN, based on GANs. They demonstrated that it is possible to generate novel images through computer-based learning. Their method analyzed 81,449 paintings by 1119 artists (from the fifteenth to the twentieth century) from three WikiArt datasets [78]. To obtain the results, they surveyed 18 users of Amazon Mechanical Turk, obtaining 10 different responses per image about the quality of the results. They concluded that 85% of the expressionist sets were of better value to the human artists; that 53% of the machine-generated images were seen as images by contemporary artists; and that CAN's images are considered to have 60% more creativity than those created with generic GANs.

Neumann et al. [251, 252] trained GANs to create new images from previous images scored with high and low aesthetic values. They used two sets of data (faces and butterflies). When using a single-dimensional feature, more realistic images were obtained for the faces dataset than for the butterflies dataset, as the butterflies dataset contains more varied images. When experimenting with two-dimensional features the more realistic images were produced by minimizing smoothness and maximizing saturation for both datasets, with the results skewed towards more colorful and rugged images.

Tabeli and Milanfar [253] used a CNN trained with a large-scale dataset tagged with human aesthetic preferences to create an image enhancement machine they call Neural Image Assessment (NIMA). They used the AVA dataset [71] for training and testing. They tested several classifiers (VGG-16 [68], Inception-v2 [110] and MobileNet [176]), and found the best results were obtained using Inception-v2: an accuracy of 81.88%, a linear correlation coefficient (LCC) of 0.660, a Spearman's rank correlation coefficient (SRCC) of 0.636 and an EMD of 0.048. The network was trained and evaluated for the improvement of the photographs in the MIT-Adobe FiveK dataset [254].

Bontrager et al. [255] described the generation high-quality image through the combination of GANs with interactive evolutionary computation (IEC). They produced 2D images from an initial configuration of three datasets CelebA face dataset [256], UT Zappos50K shoes [257], and 3D Chairs images [258] and a vanilla version of Deep Interactive Evolution (DeepIE). Figure 7 illustrates a test by the authors who were trying to arrive at an image similar to Nicolas Cage. The results of selecting three different photographs (1) and the steps toward the best result (2, 3, 4, 5, and 10) are shown.

Van Noord and Postma [259] proposed a model of image painting capable of generating missing content in paintings, Pixel Content Encoders (PCE), using dilated convolutions and PatchGAN. They compared the results PCE with Context Encode (CE) [260], a CNN developed by Pathak et al. [260]. CE was trained to generate content from a region of the image in a manner dependent on the surrounding pixels.

In the next year, Jboor et al. [83] presented a method of completing paintings intended for recovery of damaged works of art. They used the WikiArt [78], Rijksmuseum [81] and MET [107] datasets. A DCGAN with a VGG-16 [68] based architecture gave results that were not effective for a dataset with different characteristics and contexts. They then designed a framework that improves visual quality with a GAN-based semantic inpainting using a split and response strategy. Instead of training a single GAN, they used K-means grouping for category classification.

Later, Tanjil and Ross [261] investigated the use of a pre-trained object detection DCNN model (using the ILSVRC 2012 dataset [262]) for the generation of art with evolutionary computing. See Table 9 for information of the dataset. They developed a heuristic technique, Mean Minimum Matrix Strategy (MMMS), and their experiments showed that Genetic Programming (GP) can create "procedural texture images that...have the same high-level feature" [34]. The user provides a label and the machine develops an image of the content. GP is a type of evolutionary computing, which allows automatic problem solving without the need for the user to specify the shape or structure of the problem. The results sometimes result in the machine-generated art looking like a key point in the image, rather than exhibiting the full expected theme. In other words, the content of the image would not be confused with the class to which it refers, since the machine has a different form of observation than the human eye, but this can be useful when creating artistic works.

Elgammal [263] created a creative adversarial network algorithm called AICAN. It tries to learn from existing works of art to generate images, but penalizes the creation of works that emulate a style too similar to existing art. AICAN training uses 80,000 images representing the last five centuries of Western art. It also creates titles for its works, based on the known titles of existing works. To evaluate the images created by AICAN, the author used a previously developed algorithm. To determine if humans could tell that the works of art were created by a machine, the author asked those attending an exhibition held at Art Basel; 75% thought that the works of art were created by human artists. The first work of art created by AICAN, shown in Fig. 8, was entitled "St George Killing the Dragon" and was sold in New York in November 2017 for $16,000.

Blair [265] created an artist and artificial critic, Hercule LeNet, using Adversarial Coevolution between a Genetic Program (HERCL) and a DCNN (LeNet). This artificial artist produces images of low algorithmic complexity, which resemble a set of real photographs and can deceive the human visual system. An example of Hercule LeNet's results from a photograph of the Sydney Opera House is shown in Fig. 9.

Finally, Shen et al. [266, 267] proposed a two-stage process to align images: a feature-based parametric coarse alignment using one or more homographies (through RANSAC [268–275]) and a non-parametric fine pixel-wise alignment with an unsupervised way by a deep network which optimizes a standard structural similarity metric

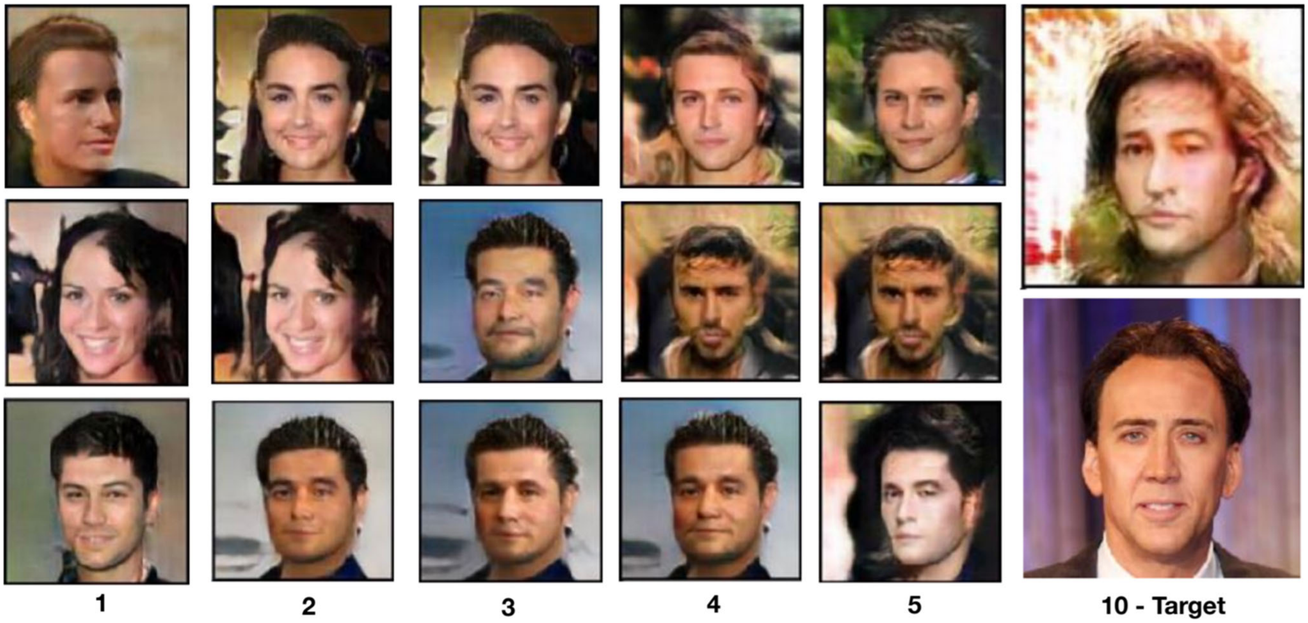

Fig. 7 Results from Bontrager et al.'s [255] attempt to approach a result like the one in image 10—Target (top), an image of Nicolas Cage. The results of selecting three different photographs (1) in each of the steps (2, 3, 4, and 5) are shown. Picture 10, Target (top) is the best version achieved in step 10

Fig. 8 "St George Killing the Dragon," made by AICAN [264], sold in New York in November 2017 for $16,000

(SSIM) [276–278] between the two images. This was the first time that the model has been used in the alignment of works of art, using the Brueghel dataset [147]. The results and code can be viewed online [267].

7.2 Generation of 3D models

Temizel [279] proposed the use of GANs for the generation of 3D objects. The author's implementation used the ModelNet dataset [280] and selects 989 chair class samples, 615-bed class samples, and 780 sofa class samples. He performed two tests with different object classes for 2 and 4 conditions (orientations/rotations) in a batch of 128 pairs. Results were measured in Average Absolute Difference between generated matrices (AAD) and Average Voxel Agreement Ratio (AVAR). For both experiment the best results were for sofa.

Li et al. [281] proposed a model for the generation and reconstruction of 3D models using a GAN that adds the class information to the generator and the discriminator—a 3D conditional GAN. They used the ModelNet10 subset of the ModelNet dataset [280] to train the generation network; this dataset contains 4899 3D models of 10 classes (bathtub, bed, chair, desk, dresser, monitor, nightstand, sofa, table, and toilet). Examples of the generated results are shown in Fig. 10. For the search of 3D objects the authors used 10 classes of the ShapeNet dataset [282] (bathtub, bed, boat, bookcase, car, chair, monitor, plane, sofa, and table). They collected random images from the Internet to use as a background. They tested the IKEA dataset [283], which contains 759 images related to six categories (bed, bookcase, chair, monitor, plane, sofa, and table). Finally, they trained the network with both datasets. The results obtained were of an average accuracy of 70.9% in the IKEA dataset.

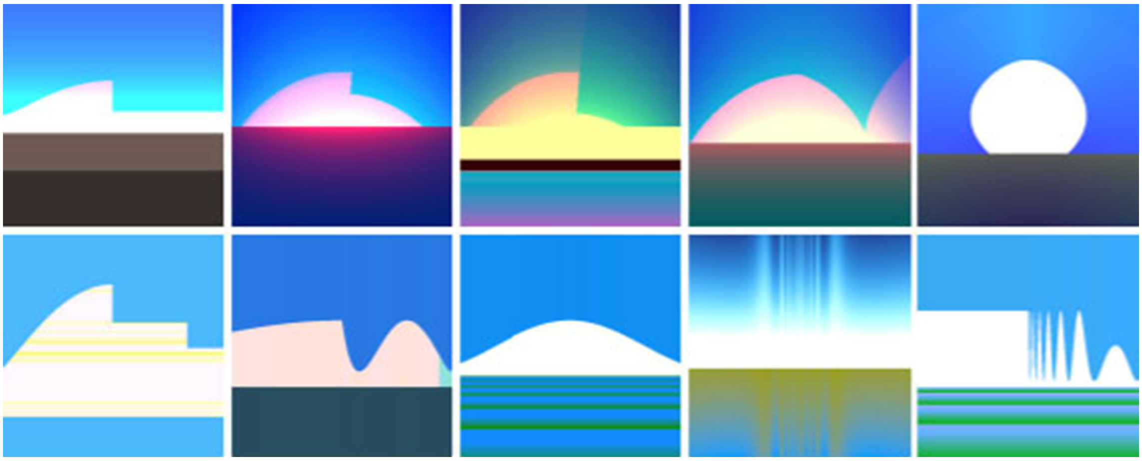

Fig. 9 Example of the results of image generation with Hercule LeNet [265]

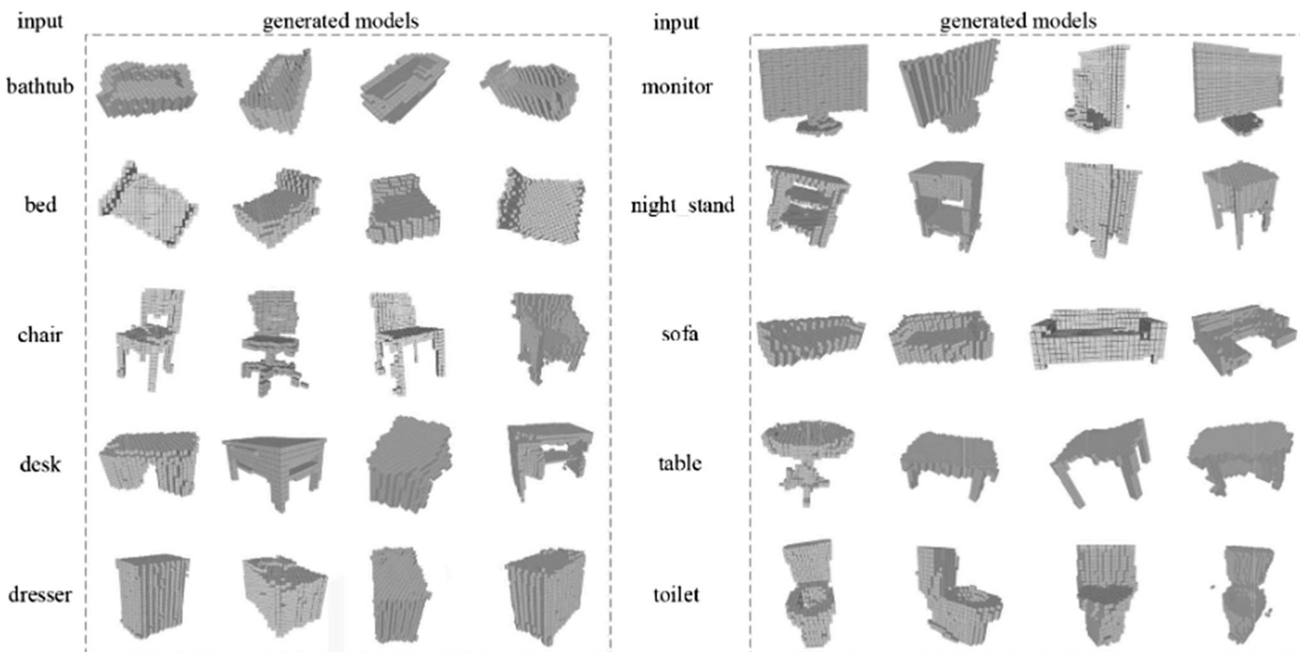

Fig. 10 Examples of 3D object generation by the 3D conditional GAN of Li et al. [281]

7.3 Generation of graphics for games

Some work has used ANNs to generate of graphic content for games (such as levels or textures). For example, Volz et al. [284] trained a GAN to create levels in Super Mario Bros using the Video Game Level Corpus [285] and enhanced it with the application of a covariance matrix adaptation evolution strategy (CMA-ES). They used the A champion agent from the 2009 Mario AI competition [286] to evaluate whether a level is playable and how many jumping actions are required to clear it.

Hollingsworth and Schrum [287] presented the Infinite Art Gallery, a game that uses Procedural Content Generation (PCG) with Compositional Pattern Producing Networks to allow its users to explore an art world adapted to their visual preferences. They conducted a study with 30 users to evaluate responses to the game, and measured an average enjoyment rate of 4.2 (on a 5-point scale).

8 Discussion

This article has analyzed a large number of articles dedicated to ANN in the Visual Arts from the last 8 years. The number of articles per year grew constantly from 2012–2015 and has been the same since 2016 (excluding 2020, for obvious reasons). In 2019, the more complex tasks (quality and aesthetic evaluation, and image generation) show greater growth compared to more objective tasks such as object detection, for which only one article was found.

The categories with the most examples are evaluations based on quality and aesthetics (28) and classification by style and author. It is important to note for aesthetics-based evaluation that some of these tasks are being used in real-world applications, for example, evaluation of real estate images [288].

We must highlight differences in the difficulty and degree of subjectivity of the different tasks. While the detection of objects presents total objectivity, the evaluation of the aesthetic components of artistic (and non-artistic) images appears to be totally subjective. In fact, it makes no sense to speak of aesthetic evaluation without defining who is evaluating the image (either an individual or the average of evaluations of a group of people, and their cultural relationships).

9 Conclusions

The use of ANN in the Visual Arts is currently an area of great research interest, with many advances made and excellent prospects for the future. In this article, we have introduced the studies that have been carried out in recent years, and summarized their content in Table 1. We must point out that the current year has not been taken into account for our conclusions, and we assume that there may be research from 2019 that we have missed.

We have carried out an exhaustive survey of recent advances in the most common uses of ANN and Deep Learning in Visual Arts—recognition, classification, evaluation, prediction, transfer, and creation. We have found a constant increase in the number of studies in these areas. The oldest focused mainly on the detection of objects in works of art, while the most recent, especially since 2018, have investigated the possibilities of generating images with artistic value.

As a complement to the work described here, we provide a table with information on the image datasets used by the different authors in recent years that are available online (Table 9). We intend this will serve as a reference and manual for other researchers and thus facilitate the search both for articles of interest and for datasets for future research or applications.

We have been struck by the scarcity of commercial applications that make use of the results of this research. We believe that this is a field ripe for the creation of various types of solutions, from systems that can generate filters and images to the user's taste, to applications for searching images on the Internet personalized according to our personal preferences. These and other utilities could also have applications in the areas of advertising and marketing.

The popularization of the use of ANNs in art has made available many libraries to test GAN on any system, just as there are numerous resources available in the cloud that can be tested for free. We have also found numerous works that are published directly on the web (on GitHub and other websites) as personal projects, that do not have associated research articles. This popularization of the use of ANNs in the artistic field can also have a negative aspect: the capacity of current systems is so great that most users are limited to using existing resources and do not seek new solutions.

We highlight the existence of works of art with commercial results generated by these techniques. We trust that this dynamic will be strengthened in the future, and that Ada Lovelace's dream of seeing computer systems capable of generating new and interesting images autonomously, or even of generating new artistic styles, will come true. In this sense, it may be relevant to increase the input of these systems, through the incorporation of larger datasets, or the possibility of obtaining huge sets of images from the Internet.

Acknowledgements CITIC, as a Research Centre of the Galician University System, is financed by the Regional Ministry of Education, University and Vocational Training of the Xunta de Galicia through the European Regional Development Fund (ERDF) with 80% of funding provided by Operational Programme ERDF Galicia 2014–2020 and the remaining 20% by the General Secretariat of Universities (Ref. ED431G 2019/01).

Funding Statement This work has also been supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (Ref. ED431G01, ED431D 201716), and Competitive Reference Groups (Ref. ED431C 201849).

Availability of data and material No data were used to support this study.

Compliance with ethical standards

Conflict of interest The authors declares that there are no conflicts of interest regarding the publication of this paper.

Code availability No code was used to support this study.

Glossary

This section explains a number of concepts that are mentioned throughout the document that readers may not be familiar with. The article is extensive and has numerous annotations, so those that are considered most difficult to understand or most commonly used have been selected.

- 1-vs-rest or One-vs-Rest strategy [289]: splits a multi-class classification into one binary classification problem per class.
- 10-Fold Cross-Validation strategy [290]: the original sample is divided into 10 samples of the same size and one of these subsamples is kept as test data, with the rest used as training data. The same process is carried out with all samples and the results can be subsequently averaged.
- Accuracy: How close the measured result is to the actual value. It is common to use this value as a percentage.
- AUROC (area under the receiver operating characteristic): a measure of discrimination, which discriminates between positive and negative examples. For example, a randomly selected x image will have a value set to look like y. If x is close to y the value will be high, and if it is not similar, the value will be low.
- Autoencoder/automatic encoder or Auto-encoding Neural Network [291]: type of Artificial Neuron Network, used for unsupervised learning of efficient data encoding.
- CNN or ConvNet (Convolutional Neural Networks): class of Deep Neural Network that uses a mathematical operation called convolution.
- Convolutional layers [292]: the convolutional layer is the main nucleus of a CNN.
- DB-CNN (Deep Bilinear-CNN) or Deep Bilinear model [168, 293]: grouping in a single representation of two bilinial models with pre-trained characteristics.
- Deep Convolutional Network or DCNN (Deep Convolutional Neural Network): consists of many neural network layers and uses convolution.
- Deep Neural Network [294, 295]: ANN with multiple layers between input and output.
- Degradation identification loss: probability of loss due to degradation.
- Dense SIFT: descriptor that divides the image into overlapping cells before using Histogram of Oriented Gradients (HOG) to describe the interest points. Important not to confuse with SIFT that detects interest points using Difference of Gaussian Filtering (DoG) and before using HOG to describe these interest points. Color Dense SIFT [114] is similar to Dense SIFT except that it also contains color information.
- Entropy estimation: estimation of differential entropy with an observing system. Commonly with histograms.
- F1-score: measure of a test's accuracy with the precision and the recall. Precision is the number of correctly identified positive results divided by the number of all positive results. Recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive, the relevance.
- GAN (Generative Adversarial Network) [296]: artificial intelligence algorithms used in unsupervised learning. GAN has two neural networks, a generator and an evaluator. DCGANs (Deep Convolutional Generative Adversarial Networks) [245] are a direct extension of the GAN that use convolutional layers in the discriminator and convolutional-transpose layers in the generator.
- GP (Genetic Programming): extension of the Genetic Algorithm (GA), in which the structures that are adapted are hierarchical computer programs, which vary in size and structure.
- GAP (Global Average Precision): average precision based on the top 20 predictions.
- Histogram of oriented gradient: feature descriptor when the distribution (histograms) of directions of gradients (oriented gradients) are used as features. This technique is used to detect objects.
- Hu moments: weighted average of pixel intensities within an image.
- kNN (k-Nearest): supervised instance algorithm. Not to be confused with k-means, that is unsupervised.
- KRCC (Kendall's rank correlation coefficient): statistic used to measure the ordinal association between two measured quantities.
- LSTM (Long Short-Term Memory) [297]: artificial recurrent neural network (RNN) architecture composed of a cell, an input gate, an output gate and a forget gate (commonly).
- mAP (mean Average Precision): mean of the average precision scores for each query.
- MMMS (Mean Minimum Matrix Strategy) [261]: reduces dimensions and identifies the most relevant high-level activation maps using reduced activation matrices for a skill.
- MRSSE (Mean Residual Sum of Squares Error): mean of the residual sum of squares (RSS), a statistical technique used to measure the amount of variance in a dataset that is not possible to explain by a regression model.
- PCG (Procedural Content Generation) [298, 299]: automation of media production, for example, PCG for games is the use of algorithms to produce game content that would otherwise be created by a designer.
- Real-AdaBoost [161]: is the use of decision trees for Adaptive Boosting (AdaBoost), a machine learning meta-algorithm. Each node on the sheet is modified to produce half of the transformations.
- Recall [300]: measure of quantity calculated as the number of true positives divided by the total number of true positives and false negatives.

- RBM (Restricted Boltzmann Machines) [301]: generative stochastic Artificial Neural Network that can learn a probability distribution over its set of inputs.
- SIFT keypoints [113]: keypoints uses to detect and describe local features in images with scale-invariant feature transform (SIFT), a feature detection algorithm in computer vision.
- SVM (Support Vector Machine) [302]: supervised learning models that are formed by hyperplane or set of hyperplanes in high or infinite dimensional space, which can be used for tasks such as classification or regression.
- Total loss function: expected loss (in average) of a group of items.

References

1. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133
2. Wani IM, Arora S (2020) Deep neural networks for diagnosis of osteoporosis: a review. In: Proceedings of ICRIC 2019. Springer, pp 65–78
3. Negassi M, Suarez-Ibarrola R, Hein S, Miernik A, Reiterer A (2020) Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects. World J Urol 38:2349–2358. https://doi.org/10.1007/s00345-019-03059-0
4. Moon S, Ahmadnezhad P, Song H-J, Thompson J, Kipp K, Akinwuntan AE, Devos H (2020) Artificial neural networks in neurorehabilitation: a scoping review. NeuroRehabilitation 1–11
5. Yucel M, Nigdeli SM, Bekdaş G (2020) Artificial neural networks (anns) and solution of civil engineering problems: anns and prediction applications. In: Artificial intelligence and machine learning applications in civil, mechanical, and industrial engineering. IGI Global, pp 13–37
6. Pradhan B, Sameen MI (2020) Review of traffic accident predictions with neural networks. In: Laser scanning systems in highway and safety assessment. Springer, Cham, pp 97–109
7. Kalogirou SA (2001) Artificial neural networks in renewable energy systems applications: a review. Renew Sustain Energy Rev 5(4):373–401
8. Sundararaj A, Ravi R, Thirumalai P, Radhakrishnan G (1999) Artificial neural network applications in electrochemistry—a review. Bull Electrochem 15(12):552–555
9. Risi S, Togelius J (2015) Neuroevolution in games: state of the art and open challenges. IEEE Trans Comput Intell AI Games 9(1):25–41
10. Lipton ZC, Berkowitz J, Elkan C A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019
11. Mammone RJ (1994) Artificial neural networks for speech and vision, vol 4. Chapman & Hall, London
12. Lippmann RP (1989) Review of neural networks for speech recognition. Neural Comput 1(1):1–38
13. Kamble BC (2016) Speech recognition using artificial neural network—a review. Int J Comput Commun Instrum Eng 3(1):61–64
14. Dias FM, Antunes A, Mota AM (2004) Artificial neural networks: a review of commercial hardware. Eng Appl Artif Intell 17(8):945–952
15. Giebel H (1971) Feature extraction and recognition of handwritten characters by homogeneous layers. In: Zeichenerkennung durch biologische und technische Systeme/Pattern Recognition in Biological and Technical Systems. Springer, pp 162–169
16. Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36(4):193–202
17. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1(4):541–551
18. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
19. Berg A, Deng J, Fei-Fei L (2010) Large scale visual recognition challenge (ilsvrc). http://www.image-net.org/challenges/LSVRC, p 3
20. Le QV et al (2015) A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks. Google Brain 1–20
21. Wu J (2017) Introduction to convolutional neural networks, National Key Lab for Novel Software Technology. Nanjing Univ China 5:23
22. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA (2018) Generative adversarial networks: an overview. IEEE Signal Process Mag 35(1):53–65
23. Romero JJ (2008) The art of artificial evolution: a handbook on evolutionary art and music. Springer, Berlin
24. Romero J (2020) Artificial intelligence in music, sound, art and design: 9th international conference, EvoMUSART 2020, Held as Part of EvoStar 2020, Seville, Spain, April 15–17, 2020, Proceedings. Springer
25. Association for computational creativity. https://computationalcreativity.net/home/
26. The bridges conference. https://www.bridgesmathart.org/
27. Evomusart. https://evomusart-index.dei.uc.pt/
28. Dorin A (2015) Artificial life art, creativity, and techno-hybridization (editor's introduction). Artif Life 21(3):261–270
29. Greenfield G, Machado P (2012) Guest editor's introduction, special issue on mathematical models used in aesthetic evaluation. J Math Arts 6(2–3):59–64
30. Romero J, Johnson C, McCormack J (2019) Complex systems in aesthetics and arts. Complexity 2019:2. https://doi.org/10.1155/2019/9836102
31. Galanter P (2012) Computational aesthetic evaluation: steps towards machine creativity. In: ACM SIGGRAPH 2012 courses. pp 1–162
32. Spratt EL, Elgammal A (2014) Computational beauty: aesthetic judgment at the intersection of art and science. In: European conference on computer vision. Springer, pp 35–53
33. Toivonen H, Gross O (2015) Data mining and machine learning in computational creativity. Data Min Knowl Disc 5(6):265–275
34. Upadhyaya N, Dixit M, Pradesh DM (2016) A review: relating low level features to high level semantics in cbir. Int J Signal Process Image Process Pattern Recognit 9(3):433–444
35. Johnson CG, McCormack J, Santos I, Romero J (2019) Understanding aesthetics and fitness measures in evolutionary art systems. Complexity 2019:14. https://doi.org/10.1155/2019/3495962
36. Todd PM, Werner GM (1999) Frankensteinian methods for evolutionary music. Musical networks: parallel distributed perception and performance, pp 313–340
37. Lewis M (2008) Evolutionary visual art and design. In: The art of artificial evolution. Springer, Berlin, pp 3–37
38. Briot J-P, Hadjeres G, Pachet F-D, Deep learning techniques for music generation—a survey. arXiv preprint arXiv:1709.01620

39. Briot J-P, Hadjeres G, Pachet F (2019) Deep learning techniques for music generation, vol 10. Springer, Berlin
40. Papers index of this soa. https://cutt.ly/4fCCWGs
41. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) Decaf: a deep convolutional activation feature for generic visual recognition. In: International conference on machine learning, pp 647–655
42. Elhoseiny M, Cohen S, Chang W, Price B, Elgammal A, Automatic annotation of structured facts in images. arXiv:1604.00466
43. Wilber MJ, Fang C, Jin H, Hertzmann A, Collomosse J, Belongie S (2017) Bam! the behance artistic media dataset for recognition beyond photography. In: Proceedings of the IEEE international conference on computer vision, pp 1202–1211
44. Masui K, Ochiai A, Yoshizawa S, Nakayama H (2017) Recurrent visual relationship recognition with triplet unit. In: 2017 IEEE international symposium on multimedia (ISM). IEEE, pp 69–76
45. Nguyen K. Relational networks for visual relationship detection in images
46. Zhang J, Shih K, Tao A, Catanzaro B, Elgammal A. Introduction to the 1st place winning model of openimages relationship detection challenge. arXiv:1811.00662
47. Peyre J, Laptev I, Schmid C, Sivic J, Detecting rare visual relations using analogies. arXiv:1812.05736
48. Detección del logotipo del vehículo utilizando una red neuronal convolucional y una pirámide de histograma de gradientes orientados, in: 11ª Conferencia Internacional Conjunta de 2014 sobre Ciencias de la Computación e Ingeniería de Software (JCSSE)
49. Dulecha TG, Giachetti A, Pintus R, Ciortan I, Jaspe A, Gobbetti E (2019) Crack detection in single- and multi-light images of painted surfaces using convolutional neural networks. In: Eurographics Workshop on Graphics and Cultural Heritage. https://doi.org/10.2312/gch.20191347
50. Hall P, Cai H, Wu Q, Corradi T (2015) Cross-depiction problem: recognition and synthesis of photographs and artwork. Comput Visual Media 1(2):91–103
51. Seguin B, Striolo C, Kaplan F et al (2016) Visual link retrieval in a database of paintings. In: European conference on computer vision. Springer, Berlin, pp 753–767
52. Inoue N, Furuta R, Yamasaki T, Aizawa K (2018) Cross-domain weakly-supervised object detection through progressive domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5001–5009
53. Gonthier N, Gousseau Y, Ladjal S, Bonfait O (2018) Weakly supervised object detection in artworks. In: Proceedings of the European conference on computer vision (ECCV)
54. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp 91–99
55. Westlake N, Cai H, Hall P (2016) Detecting people in artwork with cnns. In: European conference on computer vision. Springer, pp 825–841
56. Seguin B, Costiner L, di Lenardo I, Kaplan F (2018) New techniques for the digitization of art historical photographic archives-the case of the cini foundation in venice. In: Archiving conference, vol 2018. Society for Imaging Science and Technology, pp 1–5
57. Gonthier N, Ladjal S, Gousseau Y. Multiple instance learning on deep features for weakly supervised object detection with extreme domain shifts. arXiv:2008.01178
58. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft coco: common objects in context. In: European conference on computer vision. Springer, Berlin, pp 740–755
59. Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386
60. Thomas C, Kovashka A (2018) Artistic object recognition by unsupervised style adaptation. In: Asian conference on computer vision. Springer, pp 460–476
61. Crowley EJ, Zisserman A (2014) In search of art. In: European conference on computer vision. Springer, pp 54–70
62. Nguyen N-V, Rigaud C, Burie J-C (2018) Digital comics image indexing based on deep learning. J Imaging 4(7):89
63. Ogawa T, Otsubo A, Narita R, Matsui Y, Yamasaki T, Aizawa K. Object detection for comics using manga109 annotations. arXiv:1803.08670
64. Matsui Y, Ito K, Aramaki Y, Fujimoto A, Ogawa T, Yamasaki T, Aizawa K (2017) Sketch-based manga retrieval using manga109 dataset. Multimed Tools Appl 76(20):21811–21838
65. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 7263–7271
66. Niitani Y, Ogawa T, Saito S, Saito M (2017) Chainercv: a library for deep learning in computer vision. In: Proceedings of the 25th ACM international conference on multimedia. pp 1217–1220
67. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) Ssd: single shot multibox detector. In: European conference on computer vision. Springer, pp 21–37
68. Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
69. Dubray D, Laubrock J, Deep cnn-based speech balloon detection and segmentation for comic books. arXiv:1902.08137
70. Dunst A, Hartel R, Laubrock J (2017) The graphic narrative corpus (gnc): design, annotation, and analysis for the digital humanities. In: 2017 14th IAPR international conference on document analysis and recognition (ICDAR), vol 3. IEEE, pp 15–20
71. Murray N, Marchesotti L, Perronnin F (2012) Ava: a large-scale database for aesthetic visual analysis. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 2408–2415
72. Lu X, Lin Z, Jin H, Yang J, Wang JZ (2015) Rating image aesthetics using deep learning. IEEE Trans Multimed 17(11):2021–2034
73. Karayev S, Trentacoste M, Han H, Agarwala A, Darrell T, Hertzmann A, Winnemoeller H, Recognizing image style. arXiv:1311.3715
74. Website: Flickr. https://www.flickr.com/
75. All data, trained predictors, and code of sergey karayev. http://sergeykarayev.com/recognizing-image-style
76. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. pp 675–678
77. Bar Y, Levy N, Wolf L (2014) Classification of artistic styles using binarized features derived from a deep neural network. In: European conference on computer vision. Springer, pp 71–84
78. Wikiart. visual art encyclopedia. https://www.wikiart.org/en/artists-by-painting-school
79. Bergamo A, Torresani L, Fitzgibbon AW (2011) Picodes: learning a compact code for novel-category recognition. In: Advances in neural information processing systems. pp 2088–2096
80. Khan FS, Beigpour S, Van de Weijer J, Felsberg M (2014) Painting-91: a large scale database for computational painting categorization. Mach Vis Appl 25(6):1385–1397

81. Mensink T, Van Gemert J (2014) The rijksmuseum challenge: museum-centered visual recognition. In: Proceedings of international conference on multimedia retrieval. ACM, p 451
82. 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
83. Jboor NH, Belhi A, Al-Ali AK, Bouras A, Jaoua A (2019) Towards an inpainting framework for visual cultural heritage. In: 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT). IEEE, pp 602–607
84. Castro L, Perez R, Santos A, Carballal A (2014) Authorship and aesthetics experiments: comparison of results between human and computational systems. In: International conference on evolutionary and biologically inspired music and art. Springer, pp 74–84
85. Götz KO, Götz K (1974) The maitland graves design judgment test judged by 22 experts. Percept Mot Skills 39(1):261–262
86. Eysenck H, Castle M (1971) Comparative study of artists and nonartists on the maitland graves design judgment test. J Appl Psychol 55(4):389
87. Machado P, Romero J, Manaris B (2008) Experiments in computational aesthetics. In: The art of artificial evolution. Springer, pp 381–415
88. Saleh B, Elgammal A (2015) A unified framework for painting classification. In: 2015 IEEE international conference on data mining workshop (ICDMW). IEEE, pp 1254–1261
89. Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42(3):145–175
90. Torresani L, Szummer M, Fitzgibbon A (2010) Efficient object category recognition using classemes. In: European conference on computer vision. Springer, pp 776–789
91. Tan WR, Chan CS, Aguirre HE, Tanaka K (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
92. Tan WR, Chan CS, Aguirre HE, Tanaka K, Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification
93. Saleh B, Elgammal A, Large-scale classification of fine-art paintings: learning the right metric on the right feature. arXiv:1505.00855
94. Banerji S, Sinha A (2016) Painting classification using a pre-trained convolutional neural network. In: International conference on computer vision, graphics, and image processing. Springer, pp 168–179
95. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y, Overfeat: integrated recognition, localization and detection using convolutional networks. arXiv:1312.6229
96. Baumer M, Chen D, Understanding visual art with cnns
97. Bianco S, Mazzini D, Schettini R (2017) Deep multibranch neural network for painting categorization. In: International conference on image analysis and processing. Springer, pp 414–423
98. Lecoutre A, Negrevergne B, Yger F, Recognizing art style automatically with deep learning
99. Cazenave T (2017) Residual networks for computer go. IEEE Trans Games 10(1):107–110
100. ergart. https://www.ergsart.com
101. 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
102. Deepart. http://deepart2.ece.ust.hk/
103. Art500k. http://deepart2.ece.ust.hk/ART500K/art500k.html
104. Google arts and culture. https://artsandculture.google.com/
105. Web gallery of art. https://www.wga.hu/
106. Strezoski G, Worring M, Omniart: multi-task deep learning for artistic data analysis. arXiv:1708.00684
107. Met. https://www.metmuseum.org/art/collection
108. Omniart dataset. http://www.vistory-omniart.com/
109. Couprie LD (1983) Iconclass: an iconographic classification system. Art Librar J 8(2):32–49
110. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1–9
111. Hicsonmez S, Samet N, Sener F, Duygulu P (2017) Draw: deep networks for recognizing styles of artists who illustrate children's books. In: Proceedings of the 2017 ACM on international conference on multimedia retrieval. pp 338–346
112. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
113. Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision, vol 2. IEEE, pp 1150–1157
114. Sener F, Samet N, Sahin PD (2012) Identification of illustrators. In: European conference on computer vision. Springer, pp 589–597
115. Rodriguez CS, Lech M, Pirogova E (2018) Classification of style in fine-art paintings using transfer learning and weighted image patches. In: 2018 12th international conference on signal processing and communication systems (ICSPCS). IEEE, pp 1–7
116. Florea C, Condorovici R, Vertan C (2018) Pandora. http://imag.pub.ro/pandora/pandora_download.html
117. Florea C, Toca C, Gieseke F (2017) Artistic movement recognition by boosted fusion of color structure and topographic description. In: 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 569–577
118. Mooers CN (1977) Preventing software piracy. Computer 10(3):29–30
119. Hua K-L, Ho T-T, Jangtjik K-A, Chen Y-J, Yeh M-C (2020) Artist-based painting classification using Markov random fields with convolution neural network. Multimed Tools Appl 79:12635–12658. https://doi.org/10.1007/s11042-019-08547-4
120. Jangtjik KA, Yeh M-C, Hua K-L (2016) Artist-based classification via deep learning with multi-scale weighted pooling. In: Proceedings of the 24th ACM international conference on Multimedia. pp 635–639
121. Elgammal A, Kang Y, Den Leeuw M (2018) Picasso, matisse, or a fake? Automated analysis of drawings at the stroke level for attribution and authentication. In: Thirty-second AAAI conference on artificial intelligence
122. Chen J, Deng A (2018) Comparison of machine learning techniques for artist identification
123. Sandoval C, Pirogova E, Lech M (2019) Two-stage deep learning approach to the classification of fine-art paintings. IEEE Access 7:41770–41781
124. Kim Y-M (2018) What makes the difference in visual styles of comics: from classification to style transfer. In: 2018 3rd international conference on computational intelligence and applications (ICCIA). IEEE, pp 181–185
125. Young-Min K (2019) Feature visualization in comic artist classification using deep neural networks. J Big Data 6(1):56
126. Furusawa C, Hiroshiba K, Ogaki K, Odagiri Y (2017) Comicolorization: semi-automatic manga colorization. In: SIGGRAPH Asia 2017 Technical Briefs. pp 1–4
127. Yoshimura Y, Cai B, Wang Z, Ratti C (2019) Deep learning architect: classification for architectural design through the eye of artificial intelligence. In: International conference on

computers in urban planning and urban management. Springer, pp 249–265
128. Liao W, Lan C, Zeng W, Yang MY, Rosenhahn B, Exploring the semantics for visual relationship detection. arXiv:1904.02104
129. Zhang J, Shih KJ, Elgammal A, Tao A, Catanzaro B (2019) Graphical contrastive losses for scene graph parsing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 11535–11543
130. Gu J, Zhao H, Lin Z, Li S, Cai J, Ling M (2019) Scene graph generation with external knowledge and image reconstruction. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1969–1978
131. Zhang J, Kalantidis Y, Rohrbach M, Paluri M, Elgammal A, Elhoseiny M (2019) Large-scale visual relationship understanding. In: Proceedings of the AAAI conference on artificial intelligence, vol 33. pp 9185–9194
132. Tian X, Dong Z, Yang K, Mei T (2015) Query-dependent aesthetic model with deep learning for photo quality assessment. IEEE Trans Multimed 17(11):2035–2048
133. Luo W, Wang X, Tang X (2011) Content-based photo quality assessment. In: 2011 International conference on computer vision. IEEE, pp 2206–2213
134. Wagner M, Lin H, Li S, Saupe D, Algorithm selection for image quality assessment. arXiv:1908.06911
135. Machado P, Romero J, Nadal M, Santos A, Correia J, Carballal A (2015) Computerized measures of visual complexity. Acta Psychol 160:43–57
136. Haykin S (1994) Neural networks: a comprehensive foundation. Prentice Hall PTR, London
137. Denzler J, Rodner E, Simon M (2016) Convolutional neural networks as a computational model for the underlying processes of aesthetics perception. In: European conference on computer vision. Springer, pp 871–887
138. Amirshahi SA, Denzler J, Redies C, Jenaesthetics—a public dataset of paintings for aesthetic research. Computer Vision Group [Google Scholar], Jena
139. Redies C, Amirshahi SA, Koch M, Denzler J (2012) Phog-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. In: European conference on computer vision. Springer, pp 522–531
140. Amirshahi SA, Redies C, Denzler J (2013) How self-similar are artworks at different levels of spatial resolution?. In: Proceedings of the symposium on computational aesthetics, pp 93–100
141. Carballal A, Santos A, Romero J, Machado P, Correia J, Castro L (2018) Distinguishing paintings from photographs by complexity estimates. Neural Comput Appl 30(6):1957–1969
142. Prasad M, Jwala Lakshmamma B, Chandana AH, Komali K, Manoja M, Rajesh Kumar P, Sasi Kiran P (2018) An efficient classification of flower images with convolutional neural networks. Int J Eng Technol 7(11):384–391
143. Collomosse J, Bui T, Wilber MJ, Fang C, Jin H (2017) Sketching with style: visual search with sketches and aesthetic context. In: Proceedings of the IEEE international conference on computer vision, pp 2660–2668
144. Eitz M, Hays J, Alexa M (2012) How do humans sketch objects? ACM Trans Graphics TOG 31(4):1–10
145. Lu NGM, Deformsketchnet: Deformable convolutional networks for sketch classification
146. Shen X, Efros AA, Mathieu A, Discovering visual patterns in art collections with spatially-consistent feature learning. arXiv:1903.02678
147. Brueghel family: Jan brueghel the elder. University of california, Berkeley. https://www.janbrueghel.net/
148. Dutta A, Zisserman A, The vgg image annotator (via). arXiv:1904.10699
149. En S, Nicolas S, Petitjean C, Jurie F, Heutte L (2016) New public dataset for spotting patterns in medieval document images. J Electron Imaging 26(1):011010
150. Fernando B, Tommasi T, Tuytelaars T (2015) Location recognition over large time lags. Comput Vis Image Underst 139:21–28
151. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
152. 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
153. Kaggle. https://www.kaggle.com/ikarus777/best-artworks-of-all-time
154. Datta R, Joshi D, Li J, Wang JZ (2006) Studying aesthetics in photographic images using a computational approach. In: European conference on computer vision. Springer, pp 288–301
155. Ke Y, Tang X, Jing F (2006) The design of high-level features for photo quality assessment. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), vol 1. IEEE, pp 419–426
156. Wong L-K, Low K-L (2009) Saliency-enhanced image aesthetics class prediction. In: 2009 16th IEEE international conference on image processing (ICIP). pp 997–1000
157. Marchesotti L, Perronnin F, Larlus D, Csurka G (2011) Assessing the aesthetic quality of photographs using generic image descriptors. In: 2011 international conference on computer vision. IEEE, pp 1784–1791
158. Wang W, Cai D, Wang L, Huang Q, Xu X, Li X (2016) Synthesized computational aesthetic evaluation of photos. Neurocomputing 172:244–252
159. Xia Y, Liu Z, Yan Y, Chen Y, Zhang L, Zimmermann R (2017) Media quality assessment by perceptual gaze-shift patterns discovery. IEEE Trans Multimedia 19(8):1811–1820
160. Tong H, Li M, Zhang H-J, He J, Zhang C (2004) Classification of digital photos taken by photographers or home users. In: Pacific-Rim conference on multimedia. Springer, pp 198–205
161. Friedman J, Hastie T, Tibshirani R et al (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Statist 28(2):337–407
162. Luo Y, Tang X (2008) Photo and video quality evaluation: focusing on the subject. In: European conference on computer vision. Springer, pp 386–399
163. Wu O, Hu W, Gao J (2011) Learning to predict the perceived visual quality of photos. In: 2011 international conference on computer vision. IEEE, pp 225–232
164. Tan Y, Zhou Y, Li G, Huang A (2016) Computational aesthetics of photos quality assessment based on improved artificial neural network combined with an autoencoder technique. Neurocomputing 188:50–62
165. Gao F, Wang Y, Li P, Tan M, Yu J, Zhu Y (2017) Deepsim: deep similarity for image quality assessment. Neurocomputing 257:104–114
166. Meng X, Gao F, Shi S, Zhu S, Zhu J (2018) Mlans: image aesthetic assessment via multi-layer aggregation networks. In: 2018 Eighth international conference on image processing theory, tools and applications (IPTA). IEEE, pp 1–6
167. Talebi H, Milanfar P (2018) Nima: neural image assessment. IEEE Trans Image Process 27(8):3998–4011
168. Zhang W, Ma K, Yan J, Deng D, Wang Z (2018) Blind image quality assessment using a deep bilinear convolutional neural network. IEEE Trans Circuits Syst Video Technol 30(1):36–47. https://doi.org/10.1109/TCSVT.2018.2886771

169. Verkoelen SD, Lamers MH, van der Putten P (2017) Exploring the exactitudes portrait series with restricted boltzmann machines. In: International conference on evolutionary and biologically inspired music and art. Springer, pp 321–337
170. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
171. Exactitude website. https://exactitudes.com/collectie/?v=s
172. Larson EC, Chandler DM (2010) Most apparent distortion: full-reference image quality assessment and the role of strategy. J Electron Imaging 19(1):011006
173. Ghadiyaram D, Bovik AC (2015) Massive online crowd sourced study of subjective and objective picture quality. IEEE Trans Image Process 25(1):372–387
174. Jayaraman D, Mittal A, Moorthy AK, Bovik AC (2012) Objective quality assessment of multiply distorted images. In: 2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR). IEEE, pp 1693–1697
175. Ponomarenko N, Ieremeiev O, Lukin V, Egiazarian K, Jin L, Astola J, Vozel B, Chehdi K, Carli M, Battisti F et al (2013) Color image database tid2013: peculiarities and preliminary results. In: European workshop on visual information processing (EUVIP). IEEE, pp 106–111
176. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H, Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861
177. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2818–2826
178. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans Image Process 15(11):3440–3451
179. Ponomarenko N, Jin L, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F et al (2015) Image database tid2013: peculiarities, results and perspectives. Sig Process Image Commun 30:57–77
180. Ma K, Duanmu Z, Wu Q, Wang Z, Yong H, Li H, Zhang L (2016) Waterloo exploration database: new challenges for image quality assessment models. IEEE Trans Image Process 26(2):1004–1016
181. Carballal A, Perez R, Santos A, Castro L (2014) A complexity approach for identifying aesthetic composite landscapes. In: International conference on evolutionary and biologically inspired music and art. Springer, pp 50–61
182. Lu X, Lin Z, Jin H, Yang J, Wang JZ (2014) Rapid: rating pictorial aesthetics using deep learning. In: Proceedings of the 22nd ACM international conference on Multimedia. pp 457–466
183. Zhou Y, Li G, Tan Y (2015) Computational aesthetics of photos quality assessment and classification based on artificial neural network with deep learning methods. Int J Signal Process Image Process Pattern Recognit 8(7):273–282
184. Dong Z, Tian X (2015) Multi-level photo quality assessment with multi-view features. Neurocomputing 168:308–319
185. Campbell A, Ciesielksi V, Qin AK (2015) Feature discovery by deep learning for aesthetic analysis of evolved abstract images. In: International conference on evolutionary and biologically inspired music and art. Springer, pp 27–38
186. Xu Q, D'Souza D, Ciesielski V (2007) Evolving images for entertainment. In: Proceedings of the 4th Australasian conference on Interactive entertainment. RMIT University, p 26
187. Wang W, Zhao M, Wang L, Huang J, Cai C, Xu X (2016) A multi-scene deep learning model for image aesthetic evaluation. Signal Process Image Commun 47:511–518
188. Jin X, Chi J, Peng S, Tian Y, Ye C, Li X (2016) Deep image aesthetics classification using inception modules and fine-tuning connected layer. In: 2016 8th international conference on wireless communications and signal processing (WCSP). IEEE, pp 1–6
189. Kao Y, Huang K, Maybank S (2016) Hierarchical aesthetic quality assessment using deep convolutional neural networks. Signal Process Image Commun 47:500–510
190. Kao Y, He R, Huang K, Visual aesthetic quality assessment with multi-task deep learning. arXiv:1604.04970 5
191. Malu G, Bapi RS, Indurkhya B, Learning photography aesthetics with deep cnns. arXiv:1707.03981
192. Kong S, Shen X, Lin Z, Mech R, Fowlkes C (2016) Photo aesthetics ranking network with attributes and content adaptation. In: European conference on computer vision. Springer, pp 662–679
193. Tan Y, Tang P, Zhou Y, Luo W, Kang Y, Li G (2017) Photograph aesthetical evaluation and classification with deep convolutional neural networks. Neurocomputing 228:165–175
194. Li Y-X, Pu Y-Y, Xu D, Qian W-H, Wang L-P (2017) Image aesthetic quality evaluation using convolution neural network embedded learning. Optoelectron Lett 13(6):471–475
195. Lemarchand F et al (2017) From computational aesthetic prediction for images to films and online videos. AVANT. Pismo Awangardy Filozoficzno-Naukowej (S):69–78
196. Tzelepis C, Mavridaki E, Mezaris V, Patras I (2016) Video aesthetic quality assessment using kernel support vector machine with isotropic gaussian sample uncertainty (ksvm-igsu). In: 2016 IEEE international conference on image processing (ICIP). IEEE, pp 2410–2414
197. Murray N, Gordo A, A deep architecture for unified aesthetic prediction. arXiv:1708.04890
198. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
199. Bianco S, Celona L, Napoletano P, Schettini R (2016) Predicting image aesthetics with deep learning. In: International conference on advanced concepts for intelligent vision systems. Springer, pp 117–125
200. Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A (2014) Learning deep features for scene recognition using places database. In: Advances in neural information processing systems. pp 487–495
201. Lemarchand F (2018) Fundamental visual features for aesthetic classification of photographs across datasets. Pattern Recogn Lett 112:9–17
202. Zhang C, Zhu C, Xu X, Liu Y, Xiao J, Tillo T (2018) Visual aesthetic understanding: sample-specific aesthetic classification and deep activation map visualization. Signal Process Image Commun 67:12–21
203. Zhang C, Zhu C, Xu X, Liu Y, Xiao J, Tillo T, Modelos de cnn de zhan c et al. https://github.com/galoiszhang/AWCU
204. Jin X, Wu L, Zhao G, Zhou X, Zhang X, Li X (2020) IDEA: a new dataset for image aesthetic scoring. Multimed Tools Appl 79(21):14341–14355
205. Jin X, Wu L, Zhao G, Zhou X, Zhang X, Li X (2018) Photo aesthetic scoring through spatial aggregation perception dcnn on a new idea dataset. In: International symposium on artificial intelligence and robotics. Springer, pp 41–50
206. Apostolidis K, Mezaris V (2019) Image aesthetics assessment using fully convolutional neural networks. In: International conference on multimedia modeling. Springer, pp 361–373
207. Keras neural network api. https://keras.io/
208. Implementation en keras neural network api. https://github.com/bmezaris/fullyconvolutionalnetworks

209. Sheng K, Dong W, Chai M, Wang G, Zhou P, Huang F, Hu B-G, Ji R, Ma C, Revisiting image aesthetic assessment via self-supervised feature learning. arXiv:1911.11419
210. Carballal A, Fernandez-Lozano C, Heras J, Romero J (2019) Transfer learning features for predicting aesthetics through a novel hybrid machine learning method. Neural Comput Appl 32:5889–5900. https://doi.org/10.1007/s00521-019-04065-4
211. Cetinic E, Lipic T, Grgic S (2019) A deep learning perspective on beauty, sentiment, and remembrance of art. IEEE Access 7:73694–73710
212. Dai Y, Cnn-based repetitive self-revised learning for photos' aesthetics imbalanced classification. arXiv:2003.03081
213. Dai Y, Sample-specific repetitive learning for photo aesthetic assessment and highlight region extraction. arXiv:1909.08213
214. 500px. https://web.500px.com/
215. Semmo A, Isenberg T, Döllner J (2017) Neural style transfer: a paradigm shift for image-based artistic rendering?. In: Proceedings of the symposium on non-photorealistic animation and rendering. pp 1–13
216. Gatys L, Ecker AS, Bethge M (2015) Texture synthesis using convolutional neural networks. In: Advances in neural information processing systems. pp 262–270
217. Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. Int J Comput Vis 40(1):49–70
218. Gatys LA, Ecker AS, Bethge M, A neural algorithm of artistic style. arXiv:1508.06576
219. Gatys LA, Ecker AS, Bethge M (2016) Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2414–2423
220. Gatys LA, Bethge M, Hertzmann A, Shechtman E, Preserving color in neural artistic style transfer. arXiv:1606.05897
221. Chen Y-L, Hsu C-T (2016) Towards deep style transfer: a content-aware perspective. In: BMVC
222. Champandard AJ, Semantic style transfer and turning two-bit doodles into fine artworks. arXiv:1603.01768
223. Li C, Wand M (2016) Combining markov random fields and convolutional neural networks for image synthesis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2479–2486
224. Joshi B, Stewart K, Shapiro D (2017) Bringing impressionism to life with neural style transfer in come swim. In: Proceedings of the ACM SIGGRAPH digital production symposium. pp 1–5
225. Chen Y, Lai Y-K, Liu Y-J (2017) Transforming photos to comics using convolutional neural networks. In: 2017 IEEE international conference on image processing (ICIP). IEEE, pp 2010–2014
226. Krishnan U, Sharma A, Chattopadhyay P, Feature fusion from multiple paintings for generalized artistic style transfer. Available at SSRN 3387817
227. Jing Y, Yang Y, Feng Z, Ye J, Yu Y, Song M (2020) Neural style transfer: a review. IEEE Trans Visual Comput Graph 26(11):3365–3385. https://doi.org/10.1109/TVCG.2019.2921336
228. Portfolio. greg surma. https://gsurma.github.io/
229. Correia J, Martins T, Martins P, Machado P (2016) X-faces: the exploit is out there. In: Proceedings of the seventh international conference on computational creativity
230. Machado P, Correia J, Romero J (2012) Improving face detection. In: European conference on genetic programming. Springer, pp 73–84
231. Machado P, Correia J, Romero J (2012) Expression-based evolution of faces. In: International conference on evolutionary and biologically inspired music and art. Springer, pp 187–198
232. Correia J, Martins T, Machado P (2019) Evolutionary data augmentation in deep face detection. In: Proceedings of the genetic and evolutionary computation conference companion. pp 163–164
233. Machado P, Vinhas A, Correia J, Ekárt A (2015) Evolving ambiguous images. In: Twenty-fourth international joint conference on artificial intelligence
234. Tian C, Xu Y, Li Z, Zuo W, Fei L, Liu H (2020) Attention-guided CNN for image denoising. Neural Netw 124:117–129. https://doi.org/10.1016/j.neunet.2019.12.024
235. Adnet, URL: http://github.com/hellloxiaotian/ADNet
236. Colton S, Halskov J, Ventura D, Gouldstone I, Cook M, Ferrer BP (2015) The painting fool sees! new projects with the automated painter. In: ICCC. pp 189–196
237. Krzeczkowska A, El-Hage J, Colton S, Clark S (2010) Automated collage generation-with intent. In: ICCC. pp 36–40
238. Colton S (2008) Automatic invention of fitness functions with application to scene generation. In: Workshops on applications of evolutionary computation. Springer, pp 381–391
239. Colton S (2008) Experiments in constraint-based automated scene generation. In: Proceedings of the 5th international joint workshop on computational creativity. pp 127–136
240. Colton S, Ferrer BP (2012) No photos harmed/growing paths from seed: an exhibition. In: Proceedings of the symposium on non-photorealistic animation and rendering. pp 1–10
241. Colton S (2012) Evolving a library of artistic scene descriptors. In: International conference on evolutionary and biologically inspired music and art. Springer, pp 35–47
242. The painting fool. about me. http://www.thepaintingfool.com/about/index.html
243. Colton S, Ventura D (2014) You can't know my mind: a festival of computational creativity. In: ICCC. pp 351–354
244. Dataset darci. http://darci.cs.byu.edu
245. Radford A, Metz L, Chintala S, Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434
246. Yu F, Seff A, Zhang Y, Song S, Funkhouser T, Xiao J, Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv:1506.03365
247. Dbpedia. https://wiki.dbpedia.org/
248. Krizhevsky A, Hinton G et al (2009) Learning multiple layers of features from tiny images
249. Tan WR, Chan CS, Aguirre HE, Tanaka K (2017) Artgan: artwork synthesis with conditional categorical gans. In: 2017 IEEE International conference on image processing (ICIP). IEEE, pp 3760–3764
250. Elgammal A, Liu B, Elhoseiny M, Mazzone M, Can: creative adversarial networks, generating "art" by learning about styles and deviating from style norms. arXiv:1706.07068
251. Neumann A, Pyromallis C, Alexander B (2018) Evolution of images with diversity and constraints using a generative adversarial network. In: International conference on neural information processing. Springer, pp 452–465
252. Neumann A, Pyromallis C, Alexander B, Evolution of images with diversity and constraints using a generator network. arXiv:1802.05480
253. Talebi H, Milanfar P (2018) Learned perceptual image enhancement. In: 2018 IEEE international conference on computational photography (ICCP). IEEE, pp 1–13
254. Bychkovsky V, Paris S, Chan E, Durand F (2011) Learning photographic global tonal adjustment with a database of input/output image pairs. In: CVPR 2011. IEEE, pp 97–104
255. Bontrager P, Lin W, Togelius J, Risi S (2018) Deep interactive evolution. In: International conference on computational intelligence in music, sound, art and design. Springer, pp 267–282

256. Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision. pp 3730–3738
257. Yu A, Grauman K (2014) Fine-grained visual comparisons with local learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 192–199
258. Aubry M, Maturana D, Efros AA, Russell BC, Sivic J (2014) Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 3762–3769
259. Van Noord N, Postma E, Light-weight pixel context encoders for image inpainting. arXiv:1801.05585
260. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA (2016) Context encoders: feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2536–2544
261. Tanjil F, Ross BJ (2019) Deep learning concepts for evolutionary art. In: International conference on computational intelligence in music, sound, art and design (Part of EvoStar). Springer, pp 1–17
262. Deng J, Berg A, Satheesh S, Su H, Khosla A, Li F (2012) Large scale visual recognition challenge 2012. In: ILSVRC 2012 workshop
263. Elgammal A (2019) Ai is blurring the definition of artist: advanced algorithms are using machine learning to create art autonomously. Am Sci 107(1):18–22
264. Aican.io. https://www.aican.io
265. Blair A (2019) Adversarial evolution and deep learning—How does an artist play with our visual system?. In: International conference on computational intelligence in music, sound, art and design (part of EvoStar). Springer, pp 18–34
266. Shen X, Darmon F, Efros AA, Aubry M, Ransac-flow: generic two-stage image alignment. arXiv:2004.01526
267. Shen X, Darmon F, Efros AA, Aubry M (2020) Ransac-flow: generic two-stage image alignment. In: 16th European conference on computer vision
268. Barath D, Matas J (2018) Graph-cut ransac. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 6733–6741
269. Barath D, Matas J, Noskova J (2019) Magsac: marginalizing sample consensus. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 10197–10205
270. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
271. Plötz T, Roth S (2018) Neural nearest neighbors networks. In: Advances in neural information processing systems. pp 1087–1098
272. Qi CR, Su H, Mo K, Guibas LJ (2017) Pointnet: deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 652–660
273. Raguram R, Chum O, Pollefeys M, Matas J, Frahm J-M (2012) Usac: a universal framework for random sample consensus. IEEE Trans Pattern Anal Mach Intell 35(8):2022–2038
274. Ranftl R, Koltun V (2018) Deep fundamental matrix estimation. In: Proceedings of the European conference on computer vision (ECCV). pp 284–299
275. Zhang J, Sun D, Luo Z, Yao A, Zhou L, Shen T, Chen Y, Quan L, Liao H (2019) Learning two-view correspondences and geometry using order-aware network. In: Proceedings of the IEEE international conference on computer vision. pp 5845–5854
276. Jason JY, Harley AW, Derpanis KG (2016) Back to basics: unsupervised learning of optical flow via brightness constancy and motion smoothness. In: European conference on computer vision. Springer, pp 3–10
277. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
278. Yin Z, Shi J (2018) Geonet: unsupervised learning of dense depth, optical flow and camera pose. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1983–1992
279. Temizel A et al (2018) Paired 3d model generation with conditional generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV)
280. Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J (2015) 3d shapenets: a deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1912–1920
281. Li H, Zheng Y, Wu X, Cai Q (2019) 3d Model generation and reconstruction using conditional generative adversarial network. Int J Comput Intell Syst 12(2):697–705
282. Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, Savarese S, Savva M, Song S, Su H et al Shapenet: an information-rich 3d model repository. arXiv:1512.03012
283. Lim JJ, Pirsiavash H, Torralba A (2013) Parsing ikea objects: fine pose estimation. In: Proceedings of the IEEE international conference on computer vision. pp 2992–2999
284. Volz V, Schrum J, Liu J, Lucas SM, Smith A, Risi S (2018) Evolving mario levels in the latent space of a deep convolutional generative adversarial network. In: Proceedings of the genetic and evolutionary computation conference. pp 221–228
285. Summerville AJ, Snodgrass S, Mateas M, Ontanón S, The vglc: the video game level corpus. arXiv:1606.07487
286. Togelius J, Karakovskiy S, Baumgarten R (2010) The 2009 mario ai competition. In: IEEE congress on evolutionary computation. IEEE, pp 1–8
287. Hollingsworth B, Schrum J (2019) Infinite art gallery: a game world of interactively evolved artwork. In: 2019 IEEE congress on evolutionary computation (CEC). IEEE, pp 474–481
288. Romero J, Automatic real estate image evaluation by artificial intelligence. Present. https://cutt.ly/DfVT6VI
289. Bishop CM (2006) Pattern recognition and machine learning. springer, Berlin
290. McLachlan GJ, Do K-A, Ambroise C (2005) Analyzing microarray gene expression data, vol 422. Wiley, New York
291. Kramer MA (1991) Nonlinear principal component analysis using autoassociative neural networks. AIChE J 37(2):233–243
292. Géron A (2019) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Newton
293. Lin T-Y, RoyChowdhury A, Maji S (2015) Bilinear cnn models for fine-grained visual recognition. In: Proceedings of the IEEE international conference on computer vision. pp 1449–1457
294. Kingma DP, Ba J, Adam: a method for stochastic optimization. arXiv:1412.6980
295. Bengio Y (2009) Learning deep architectures for AI. Now Publishers Inc, New York
296. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems. pp 2672–2680
297. Hochreiter S, Ja1 4 rgen schmidhuber (1997) "long short-term memory". Neural Comput 9(8)
298. Barriga NA (2019) A short introduction to procedural content generation algorithms for videogames. Int J Artif Intell Tools 28(02):1930001
299. Togelius J, Kastbjerg E, Schedl D, Yannakakis GN (2011) What is procedural content generation? Mario on the borderline. In:

Proceedings of the 2nd international workshop on procedural content generation in games. pp 1–6
300. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manag 45(4):427–437
301. Smolensky P (1986) Information processing in dynamical systems: foundations of harmony theory, Technical report. Colorado University at Boulder Department of Computer Science
302. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
303. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2008) Lost in quantization: improving particular object retrieval in large scale image databases. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
304. Ren J, Shen X, Lin Z, Mech R, Foran DJ (2017) Personalized image aesthetics. In: Proceedings of the IEEE international conference on computer vision. pp 638–647
305. You Q, Luo J, Jin H, Yang J (2015) Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Twenty-ninth AAAI conference on artificial intelligence
306. Katsurai M, Satoh S (2016) Image sentiment analysis using latent correlations among visual, textual, and sentiment views. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 2837–2841
307. Khosla A, Raju AS, Torralba A, Oliva A (2015) Understanding and predicting image memorability at a large scale. In: Proceedings of the IEEE international conference on computer vision, pp 2390–2398
308. Mohammad S, 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)
309. Yanulevskaya V, Uijlings J, Bruni E, Sartori A, Zamboni E, Bacci F, Melcher D, Sebe N (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

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.