Image and Vision Computing 89 (2019) 67–87

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

A survey on freehand sketch recognition and retrieval☆

Xianlin Zhang[a,c,*], Xueming Li[a,b], Yang Liu[a,*], Fangxiang Feng[c]

[a]School of Digital Media and Art Design, Beijing University of Posts and Telecommunications, No. 10, Xitu Cheng Road, Haidian District, Beijing, China
[b]Beijing Key Laboratory of Network System and Network Culture, China
[c]School of Computer Science, Beijing University of Posts and Telecommunications, No. 10, Xitu Cheng Road, Haidian District, Beijing, China

ARTICLE INFO

Article history:
Received 28 October 2017
Received in revised form 16 March 2019
Accepted 27 June 2019
Available online 11 July 2019

Keywords:
Touch-screen devices
Freehand sketches
Sketch generation
SBIR
FG-SBIR

ABSTRACT

With the development of digital devices and pressure sensing equipment, research into freehand sketches from touch-screen interfaces has increased significantly in recent years. As such, we provide the first comprehensive survey of recognition tasks based on sketch generation, freehand sketch classification, sketch-based image retrieval (SBIR), fine-grained sketch-based image retrieval (FG-SBIR), and sketch-based 3D shape image retrieval. Specifically, SBIR and FG-SBIR were the main focus of the survey. Primary technologies and benchmark datasets related to all sketch-based recognition topics are also discussed, along with future trends for this promising technology.

1. Introduction

Sketches have been used for centuries to record the natural world and illustrate objects and people based on human perception [1]. With the development of touch-screen digital devices, sketch research has recently become an active area of study in the computer vision (CV) community. However, sketches are a broadly-defined topic in the CV field. For example, the exaggerated caricature sketch shown in Fig. 1 was produced by a professional artist. These cartoon sketches are exaggerated and include large shape deformations but they also contain rich target information and a variety of object details. The human face and freehand drawings produced by amateur users are all classified as sketches. Various sketch types differ significantly in the richness of object detail or the manner in which they were produced.

In this study, a sketch is defined to include a freehand drawing (independent of expertise). They are assumed to be produced by professionals, amateurs, adults, or even children using electronic devices such as mobile phones, iPads, or drawing tablets. Examples of sketches included in this study are shown in Fig. 2. This type of sketch is distinctive from a face or caricature and includes only simple strokes with little textural information. As such, Hu et al. argue that a key challenge in SBIR is overcoming the ambiguity inherent in identifying sketches [2].

All sketches are unique and, as a result, sketch-based recognition tasks encounter a variety of challenges. These can primarily be summarized as follows:

(1) Large variations exist between individual sketches. A sketch is an abstract and concise description of basic concepts or silhouettes as represented by the artist. Variations in interests, styles, and artistic capabilities produce a broad range of sketches, even for a common object. Such illustrations strongly reflect individual personality characteristics. In contrast, natural images are unbiased depictions of the world with specific and detailed features. Examples of the differences between an image and a sketch are shown in Fig. 3.
(2) A variety of common drawing modalities are available and most sketch-based tasks involve cross-modality input. Freehand sketches and real photographs (or 3D images) are two different forms of media and they belong to different domains. This represents a more challenging recognition task than recognition of two objects from the same domain (such as a photo-to-photo recognition task).
(3) The representation of sketches can be inconsistent. As opposed to physical 2D images and 3D shape models, freehand sketches are generated according to user preferences. They consist of only a few simple strokes and have little to no texture or color. As such, it is difficult to identify efficient features in freehand sketches due to a lack of useful information.

☆ This paper has been recommended for acceptance by Sinisa Todorovic.
* Corresponding authors.
E-mail addresses: zxlin@bupt.edu.cn (X. Zhang), lixm@bupt.edu.cn (X. Li), Yangliu@bupt.edu.cn (Y. Liu).

https://doi.org/10.1016/j.imavis.2019.06.010
0262-8856/

Fig. 1. Examples of different types of sketches.

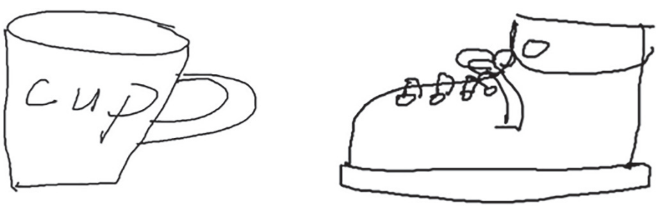

Fig. 2. Examples of the freehand sketches included in this study.

Fig. 3. An example of the differences between an image and a sketch.

(4) Specific consideration problems exist for various sketch-related tasks, which are necessary to solve specific issues. For example, memory usage and computational complexity must be considered when investigating large-scale SBIR. However, in sketch-based 3D shape retrieval, the primary issue may be the selection of multi-view 3D model projections.

We investigated specific types of hand-drawn sketches, including traditional pen-and-paper drawings and images from touch-based digital tablets. As many devices accept input from a pen-like tool, there was little distinction drawn between these modalities. Hand-drawn images also exhibited a high degree of similarity regardless of the input methodology. In addition, we primarily refer to sketches which consist of several simple black and white strokes or simply-drawn object contours with little or no artistic detail. Color-based sketches and frontal facial sketches (Ouyang et al. [3] provided a detailed description of sketch-based facial recognition), however, are beyond the scope of this study. Sketch recognition indeed including various tasks and we introduce five parts in this paper. Among the involved topics, we specifically emphasize SBIR and FG-SBIR. The primary components of each section can be described as follows:

- 1. Technological developments.
- 2. Primary datasets.
- 3. Summary and conclusions.

The primary contributions of the paper are as follows:

- 1. We conduct the first detailed review of freehand sketch recognition research in recent years.
- 2. We introduce common approaches and public benchmark datasets used in the research of sketch generation, sketch classification, SBIR, FG-SBIR, and sketch-based 3D shape retrieval.
- 3. In the summary and conclusion, we explicitly describe state-of-the-art methods and principles used to categorize existing techniques. We also provide specific recommendations and methodology evaluation criteria for readers desiring to pursue similar research. In addition, the review discusses promising directions regarding specific research tasks for sketch recognition.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of sketch generation. In Section 3, we provide a detailed review of methods and the most widely used datasets for sketch classification. In Section 4, we describe approaches for SBIR in detail. In Section 5, we focus on introducing FG-SBIR. Section 6 briefly discusses methods for matching freehand sketches to 3D shape model retrieval. Conclusions and promising future directions for freehand sketch recognition are presented in Section 7.

2. Sketch generation

2.1. Technological developments

Sketch generation, which is also named sketch synthesis, is a process that produces pseudo-sketches similar to freehand sketches of natural images. These sketches (mainly synthesized from images) are typically used for other sketch-related tasks.

The technique is intended to process a real image and produce a sketch which is similar to a freehand drawing. This process is demonstrated in Fig. 4. However, an optimal mapping function has yet to be developed between images and sketches. The specific mapping approach utilized in applications such as SBIR and FG-SBIR has a significant effect on the results. In addition to the work of sketch synthesis, there are also research efforts on how to optimize the quality of sketches. For example, Limpaecher et al. introduced a simple stroke-correction method which uses an artistic consensus model to improve strokes in real-time [4]. Lee et al. developed a shadow drawing system which can update a shadow image underlying user strokes [5]. Jiang and Liu presented a new framework for automatically transforming images into sketches using feature fusion (low-level and mid-level features) and a sketch-stroke model [6].

Despite the work of optimizing the quality of sketches, we mainly discuss the work of sketch synthesis in this part. We organize studies focusing on sketch generation methods using two classification criteria:

- Studies utilizing information from the natural images.
- Studies utilizing information from sketches (semantic strokes and stroke ordering).

2.1.1. Studies utilizing information from the natural images

Most studies on automatic sketching employ a contour detection or object segmentation technique [6-13]. These algorithms attempt to produce curves which accurately depict an image or constitute global object profiles by identifying high pixel contrast. For example, Zhu et al. exploited the inherent topological structure of salient contours [8]. Wang introduced a morphing technique for sketches using skeleton and order curves [14]. Arbelaez et al. proposed a general framework for transforming the output of any contour detector into a hierarchical region tree with the intent to generate the object contours which are most similar to human-produced image segmentation [11]. Zhao et al. presented a sketch-style face image generation algorithm utilizing the constrained least-squares estimation (LSE)

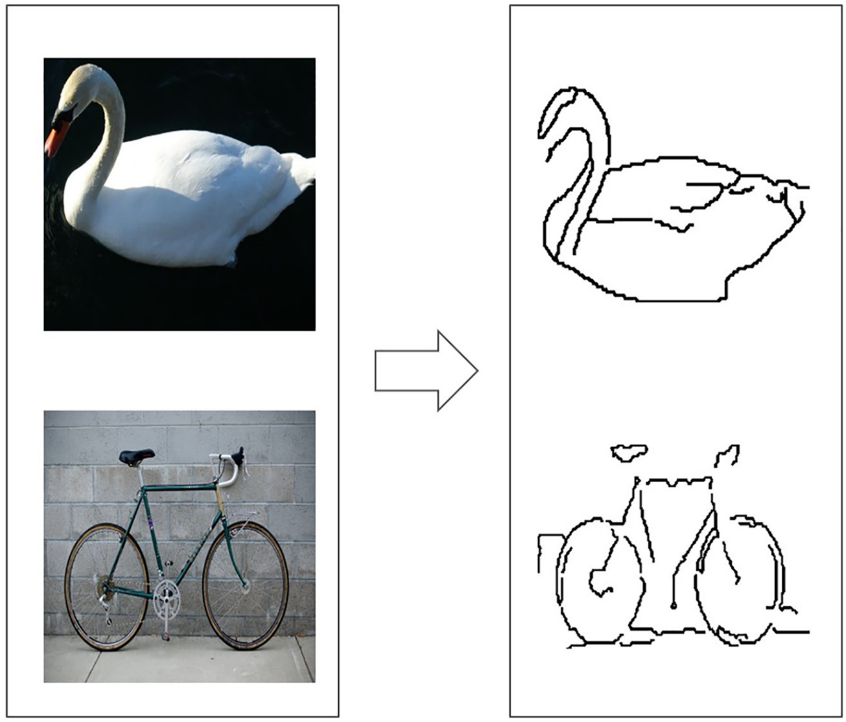

Fig. 4. An example of the sketch generation process.

and bilateral filter method [15]. Marvaniya et al. proposed sketching an object given a set of images in the same category [12]. The objective was to identify repeatable salient contours across a set of images in the same class by using Bekerley detector [16] and fast directional chamfer matching (FDCM). Lim et al. proposed a novel approach to both learning and detecting local contour-based representations for mid-level features called sketch tokens [13]. These sketch tokens were learned using supervised mid-level information in the form of hand-drawn contours in images.

In contrast, Guo et al. [17] and Qi et al. [18] proposed methods to produce a sketch from a single image, which made sketch generation more convenient and applicable. Guo et al. combined two generative models (sparse coding and Markov random field) learned from natural image statistics to present geometric structures and stochastic textures, respectively, to generate sketches [17]. Qi et al. also proposed a much simpler method exploiting perceptual grouping principles to generate improved sketches [18]. With the rapid development of deep learning (DL) in recent years, edge detection techniques [19,20] based on deep models are also experiencing increased usage in the field. As sketch generation methods overlap with image edge detection, these technologies have the potential to produce better sketches. Xie and Tu [19] proposed a method, holistically-nested edge detection (HED), which performed image-to-image prediction by means of a deep model. Later, Liu et al. proposed a richer convolutional features (RCF), which based on HED and exploited multi-scale and multilevel information of objects to perform the image-to-image prediction in a holistically manner by using convolutional neural networks (CNNs) [20]. Recently, Zhang et al. [21] proposed an improved GAN to conduct sketch synthesis by using both the information of sketches and images, the other important contribution was that they proposed a deep model to verify and determine the quality of the generated sketches.

2.1.2. Studies utilizing information from sketches

The sketch generation methods described above are all based on information contained in the original image. Additionally, sketches produced with these methods are generally applied as a preprocessing step for other CV tasks. However, sketches produced by human users typically consist of only a few simple strokes and include very few detailed features. As such, it is easier to use the characteristics of an image's shape, contours, or texture to produce sketches similar to those drawn by humans. Eitz et al. demonstrated the semantics of strokes. For example, when drawing an airplane, stroke semantics identify the aircraft's wings, body, etc. Stroke ordering is essential in producing a new sketch [1]. In addition to this, Schneider and Tuytelaars also emphasized the significance of sketch strokes [22] and studied stroke semantic segmentation.

Therefore, using sketch-stroke information is another important aspect of sketch synthesis. Despite the recent expansion in sketch research, stroke-level analysis remains sparse. Several previous studies have investigated stroke ordering and the importance of individual strokes for sketch generation and recognition [1,23,24]. While research has been conducted on the influence of sketch strokes, none of these studies have provided a means of creating stroke ordering models inside a framework. Thus, potential applications are unclear. Huang et al. were the first to take advantage of temporal stroke ordering as useful information to apply grouping constraints for segmenting freehand sketches [25]. Similarly, Li et al. proposed a deformable stroke model (DSM) to present a generative model which can automatically synthesize visually similar freehand sketches [26]. Their model first learned from a comprehensive study performed on human sketch-stroke data and then employed stroke ordering for grouping purposes. This work provided important insights into sketch strokes and presented solid evidence to support the argument that temporal ordering of strokes is meaningful, which was only considered a possibility in prior studies [25].

Further, with the breakthrough of CNN and recurrent neural network (RNN) [27] in the field of CV in recent years, there has been a combination of sketch strokes and RNN in the sketch synthesis. A sketch-rnn model, which utilized the structure of sequence-to-sequence variational auto-encoder, was used in [28] for a different purpose of conditional sketch synthesis. The work in [29] first used a reinforcement learning deep model for sketch abstraction task by sequentially removing redundant strokes, and then trained a deep model with photos only for FG-SBIR by using the proposed sketch abstraction model. Work [30] used sketch-stroke information to perform sketch synthesis as well. Concretely, they proposed a supervised-unsupervised hybrid model with dual (two-way) supervised translation sub-models and two unsupervised sub-models with shortcut cycle consistency. Different to the previous work, Li et al. [31] developed a deep network by using perceptual grouping and variational auto-encoder(VAE). Further, they proposed the largest sketch perceptual grouping (SPG) dataset to date.

2.2. Primary datasets

There have been two primary developments (sketch synthesis by using information in images and freehand sketches) in the evolving technology of sketch generation in recent years. They emerged from different points of view and have both produced effective results. The results obtained by utilizing the methods of edge detection or target saliency detection are often used as pseudo-sketches, and are then used as a pre-processing step for other sketch-related tasks to bridge the domain gap. Therefore, methods mentioned on sketch generation have some overlap with object detection, saliency detection and edge extraction algorithms. We simply list the primary datasets, the details of them are presented below.

QuickDraw dataset (https://magenta.tensorflow.org/sketch_rnn), which is the largest sketch dataset so far [28]. It is a dataset of vector drawings obtained from Quick, Draw! (http://halfdanj.dk/work/quick-draw), an online game where the players are asked to draw objects belonging to a particular object class in less than 20 seconds. QuickDraw contained 345 classes of common objects. Each class of QuickDraw is a dataset of 70 K training samples, 2.5 K validation and 2.5 K test sketches.

SPG dataset. A freehand sketch perceptual grouping dataset. The sketches in SPG are diversity, and they are chosen from the QuickDraw [28] dataset with appropriate complexity. It contains 20, 000 sketches of 25 categories with each sketch manually annotated into parts.

BSD (Berkeley Segmentation Dataset) dataset (https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/), provided an empirical basis for research on image segmentation and boundary detection, which collected 12,000 hand-labeled segmentations of Corel dataset (https://github.com/watersink/Corel5K) images from 30 human subjects. The public benchmark based on this was divided into a training set of 200 images, and a test set of 100 images, and BSDS500 was an extended version of the BSDS300 that included 200 fresh test images.

Caltech-101 was a dataset which objects of pictures were belonging to 101 categories. There were about 40 to 800 images per category and most categories had about 50 images. The size of each image was roughly 200 × 300 pixels, and Caltech-256, was an extended challenging dataset of 256 object categories containing a total of 30607 images.

ETHZ used in [32] was a dataset which had different classes of objects: Apple logo (40), Bottle (48), Giraffe (87), Mug (48) and Swan (32). The objects in the images were at various scales, orientations, illumination changes and a substantial amount of intra-class variations which made it a difficult dataset to work on.

PASCAL VOC 2007 (pattern analysis, statistical modeling and computational learning visual object classes) [33] was a dataset contained real world images with 20 labeled object categories such as people, dogs, chairs, etc.

The NYU Depth (NYUD) dataset had 1449 images. This dataset was used for edge detection. The NYUD dataset was split into 381 training, 414 validation, and 654 testing images.

2.3. Summary and conclusions

In this section, we focus on describing the technologies applicable to sketch generation. These can be divided into two categories: (1) employing information in the original images and (2) utilizing semantic sketch strokes and stroke ordering to generate sketches. Most studies utilize information in images to generate sketch, and the common methods are based on topology, contour cut, perceptual grouping and deep models. The studies on adopting sketch stroke and orders are relatively sparse. Table 1 (the "✓" in all tables of this paper presents that the methods listed are from the same author or that the results are comparable under the same experimental environment and the "×" indicates that the method is not the same experimental data setting environment and is not comparable) describes methods of sketch generation utilizing edge and contour detection to varying extents. The majority of these studies used image information to generate sketches [7,8,10-13,15,17-20,32], while others used sketch-stroke ordering information [23,25,26,28-31].

Further research into sketch generation involved hand-crafted features and classical deep features [12,13,18-20, 26]. In addition, DL algorithms have been applied to various CV tasks in recent years. These deep network models have included CNNs [35], deep belief networks (DBNs) [36], generative adversarial networks (GANs) [37], and improved network models based on GANs [38-41]. Zero-shot transfer learning [42] and one-shot learning [43] have also been utilized for CV applications. Therefore, individual sketches should be encouraged (as opposed to an image) to generate sketches as image-to-image transformation has been successful when implementing GANs [44-46].

For readers who are interested in the direction of freehand sketch synthesis, we would like you to give more attention on using sketch strokes and deep models (especially the deep structure of CNNs, RNNs and GANs), methods of [30], [28], [29], and [31] are highly recommended to consult.

3. Sketch classification

3.1. Technological developments

Sketch classification is an automated process used to identify the appropriate category for a given sketch. These algorithms were first introduced by Sutherland [47] and Herot [48] and have been developed further in recent years. Research in this area is typically divided into two types:

- Sketch classification using hand-crafted models.
- Sketch classification using deep learning models.

3.1.1. Sketch classification using hand-crafted models

Early studies were relatively simple and only recognized single object categories, including circles, straight lines, or the refining of other sketch-based tasks [49-51]. Recent studies have proposed various algorithms to successfully recognize higher level hand-written symbols and more complex sketches, such as hand-written chemical equations or mathematical formulas [52-55]. However, most of these recognition algorithms were proposed for specific tasks, which may have some limitations for identifying more general freehand sketches.

In 2012, Eitz et al. proposed a large-scale freehand benchmark dataset containing 250 categories and 20,000 freehand sketches [1]. Classification based on feature extraction was demonstrated to be a viable method for identifying both general and casual freehand sketches. Most sketch classification research focuses on designing and extracting sketch features [1, 56-58]. The modern workflow for this process is shown in Fig. 5.

Table 1
Primary methods and datasets used in sketch generation algorithms.

Used cues	Publications	Generation approach	Datasets	Deep features	Comparable
Based on images	[7]	Graph-based + Gestalt laws	||	No	×
	[8]	Topological + Graph formulation	BSD [34]	No	×
	[17]	Sparse coding + Markov filed	||	No	×
	[10]	Graph-based + Random walk	BSD	No	×
	[11]	Spectral clustering + Hierarchical tree	BSD	No	×
	[15]	Constrained LSE + Bilateral filter	Self-building	No	×
	[12]	Berkeley detector + FDCM	ETHZ [32] + Caltech	No	×
	[13]	Sketch tokens + Random forest	BSD + PASCAL	No	×
	[18]	Perceptual grouping	TU-Berlin + Caltech	No	×
	[19]	HED	BSD500 + NYUD	Yes	✓
	[20]	RCF	BSD500	Yes	✓
Based on sketch strokes	[23]	Data-Driven	Self-building	No	×
	[25]	Data-Driven	Self-building	No	×
	[26]	DSM	TU-Berlin	No	×
	[28]	RNN	QuickDraw	Yes	×
	[29]	bi-directional RNN	QuickDraw	Yes	×
	[30]	CNN + RNN	QuickDraw	Yes	×
	[31]	VAE	SPG	Yes	×

The process of sketch classification is typically divided into two steps: (1) Selecting appropriate features to represent sketch training classifiers and (2) Extracting these same characteristics from the testing set. The trained classifier is then used to determine the most appropriate category for each sketch. Eitz et al. used the histogram of oriented gradients (HOG) [59], combined with the bag-of-words (BOW) features model to conduct sketch classification on the TU-Berlin dataset [1]. Sun et al. developed a general sketch classification system to recognize semantically meaningful objects [60]. A query-adaptive shape topic model (QST) was proposed to mine object and shape topics related to the sketch in a generative process. This was done to alleviate the problem of intra-class shape variations and correct for ambiguity in unconstrained situations. Sun et al. further studied the segmentation of freehand sketches at the object level and a graph-based sketch segmentation algorithm was proposed to segment cluttered sketches into multiple parts based on a proximity factor [61]. Li et al. employed a method for representing and matching sketches through the use of a star graph, by exploiting both the local and global structure of sketches [57]. Li et al. trained a radial basis function (RBF) with a multiple kernel learning (MKL) classifier [58]. The results were then compared with multi-feature classification, including a BOW representation of HOG, self-similarity (SSIM) [62], Daisy [63], and the star graph method [57]. Schneider and Tuytelaars made use of Fisher vectors to describe features in sketches and trained a support vector machine (SVM) [64,65] to classify sketch types [56]. In addition, a new benchmark was introduced, based on TU-Berlin datasets, for better understanding and analysis of sketch classification.

3.1.2. Sketch classification using deep learning models

The methods discussed above rely heavily on hand-crafted features. Deep neural networks (DNNs) have recently outperformed traditional object recognition algorithms in multiple CV tasks. For example, improved CNN [66] and very deep CNN (i.e., residual neural network) architectures have been successfully applied to image classification [67] and person re-identification [68]. Furthermore, the latest research has demonstrated the accuracy of sketch classification has surpassed human performance using DL models [69-74]. Yu et al. designed the first sketch-DNN for classification [69]. Their deep model, which consisted of five convolution layers and two fully-connected layers, was implemented on MatConvNet [75]. Sarvadevabhatla and Babu considered sparse visual details and introduced a framework

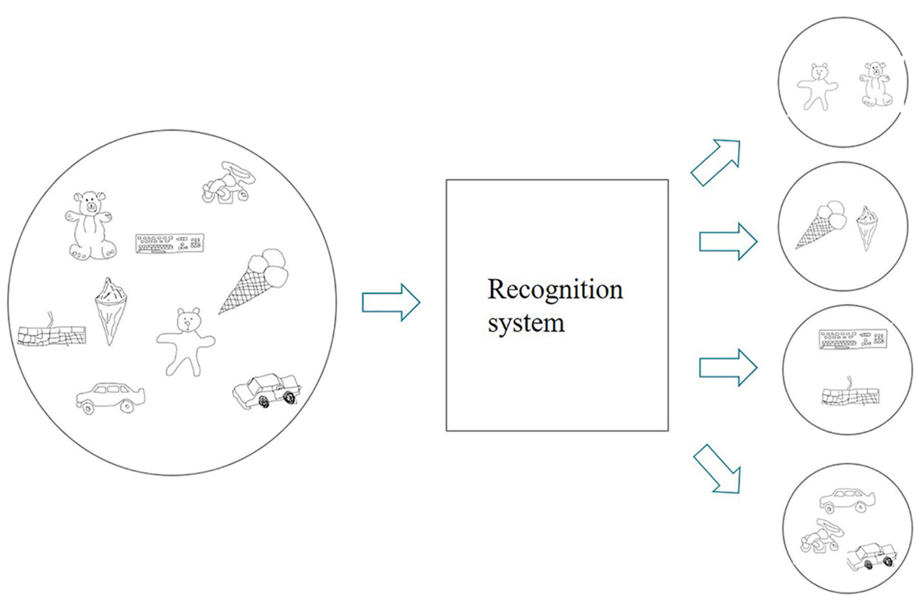

Fig. 5. An example of the framework used for sketch classification.

based on deep features extracted from CNNs [70]. They used two common CNNs for experiments: the ImageNet CNN and a modified version of the LeNet CNN.

Additionally, Yu et al. used a deep CNN model [71,72]. They designed a multi-scale multi-channel deep framework with Bayesian fusion. In addition, they encoded stroke sequential ordering into multi-channel processing to improve the results. Zhang et al. utilized SketchNet to automatically learn latent structure features from sketches [73]. They developed a triplet which was composed of the sketch, a real positive image, and a real negative image which were used as the input for the model. A ranking mechanism was then introduced to obtain robust features from positive image pairs. Finally, they formalized all constraints into a unified objective function which was used to train the model. Their experiments were conducted on Caffe, an open source deep framework [35]. Hu et al. investigated an alternative approach of synthesizing image classifiers directly from freehand sketch [76]. They found that effective photo classifiers can be synthesized using one or few freehand sketches by using the model-regression approach.

3.2. Primary datasets

The most widely used benchmarks for sketch classification are the TU-Berlin dataset [1] and the improved TU-Berlin dataset [56,73].

The TU-Berlin benchmark dataset (http://cybertron.cg.tu-berlin.de/eitz/projects/classifysketch/), developed by Eitz et al., was the first large-scale dataset for sketch recognition [1]. It was constructed using Amazon Mechanical Turk (AMT), a web-based job market, to assemble a pool of amateur artists. The participants completed a total of 351,060 sketches, each containing a median of 13 strokes.

The improved TU-Berlin dataset contains 250 categories and 20,000 sketches, with each category including exactly 80 sketches. In addition, human subjects correctly identified 73.1 percent of sketches.

Sketch-500 dataset was built based on a list of 1000 non-abstract nouns, which was a free word list collected for elementary students and contained most of commonly used nouns. For each word, a graduate student was asked to manually draw a sketch based on the top results returned from a commercial keyword-based clip-art image search engine by removing the hard words to draw a sketch and finally collected 500 sketches together with ground truth.

3.3. Summary and conclusions

Classification is an important task in sketch recognition. Early work in sketch classification was inconsistent, involving methods designed to resolve a specific problem with no uniform benchmark for evaluating the performance of different recognition algorithms. However, the field has developed rapidly since Eitz et al. established the TU-Berlin dataset [1]. Various methods using CNNs to learn deep features for sketch classification are becoming prevalent in modern studies.

Sketch classification can primarily be divided into two categories. The first involves hand-crafted features and the second uses deep model features. The first type involves extracting hand-crafted features from sketch images and inputting them into a classifier (such as SVM) [1,56-58]. Deep methods differ significantly from conventional techniques and typically produce an efficient deep network model to better learn sketch features, which are used to distinguish subtle variations in the sketches [69,71-73]. Deep learning algorithms typically outperform conventional techniques in sketch recognition. The primary methods and datasets used in sketch classification are listed in Table 2. More complex and effective deep models based on CNNs will be constructed in the future development of hand-drawn sketch classification. On the other hand, more and more large-scale hand-drawn sketch databases (including various object categories) will be created for sketch recognition research tasks.

For novice readers who are interested in sketch classification, SVM + HOG is a conventional and intuitive method which can be considered as a standard baseline model. Readers may also refer to a series of studies [69,71-73] to begin constructing deep model architectures with deep sketch features. The deep learning framework can be acquired from Caffe (http://caffe.berkeleyvision.org/), Tensorflow (https://www.tensorflow.org/), or Pytorch (http://pytorch.org/) etc.

4. Sketch-based image retrieval

4.1. Technological developments

As opposed to text-based image retrieval (TBIR), SBIR involves retrieving images from a data pool to match the sketch query. This is an important component of content-based image retrieval (CBIR). SBIR is closely correlated with touch-screen devices and has become an increasingly prominent research topic in recent years. The input for SBIR systems is typically a hand-drawn sketch (or a sketch combined with text tags) and the outputs are primarily different domain search results, such as natural images, clip-art, and comics (it is also possible to return sketches of the same domain). A typical processing framework for an SBIR system is shown in Fig. 6.

Cross-media retrieval problems typically include one of two solutions. The first involves extracting presentation features from different media domains using various processes and determining the similarity of these features with two transformation projections which make domains more comparable. For example, in the process of sketch-image retrieval, we recorded the cross-domain sketch and image as V_s^1 and V_i^2, respectively. We then extracted relevant features using an operator P^* (i.e., a histogram orientation feature or deep characteristics). After extracting features from each domain, the mathematical forms of these two media can be expressed as $P^1\left(V_s^1\right)$ and $P^2\left(V_i^2\right)$. Two transformations were then used to project these characteristics into a near comparable space W^*. After projection, the two media forms could be defined as $W^1\left[P^1\left(V_s^1\right)\right]$ and $W^2\left[P^2\left(V_i^2\right)\right]$. Finally, a similarity metric Z was used to measure the shortest distance between features:

$$Z\left\{W^1\left[P^1\left(V_s^1\right)\right], W^2\left[P^2\left(V_i^2\right)\right]\right\}$$

However, this strategy involves two difficulties. The first is selecting an appropriate feature extraction operator P^* for different media. The second is designing an appropriate projection transformation W^* to ensure the distance of a matching sketch-image pair is less than any mismatched pairs.

The second strategy for sketch-image retrieval typically involves a pre-processing step which converts images into pseudo-sketches using generation methods. Afterwards, the features $P^{1'}\left(V_s^1\right)$ and $P^{1'}\left(V_i^2\right)$ are extracted from each domain and the similarity of the presentation from the two domains is compared by a single projection $W^{2'}$, such that

$$Z\left\{W^{2'}\left[P^{1'}\left(V_s^1\right)\right], W^{2'}\left[P^{1'}\left(V_i^2\right)\right]\right\}$$

is minimized. Most approaches adopt this type of solution to perform SBIR.

As an introduction to relevant technologies, we first discuss prior research followed by more recent advances. We divide recent studies into two topics: SBIR research on mobile phones and SBIR research on computers. Computational methods are categorized as using relevance feedback (RF), not using RF, and using DL methods. We further divide non-RF discussions into common SBIR and large-scale SBIR. Table 3 presents an outline of this discussion.

Table 2
Primary methods and datasets used in sketch classification.

Methods	Publications	Generation approach	Datasets	Accuracy	Comparable
Hand-craft features	[1]	BOW + SVM	TU-Berlin	56%	✓
	[60]	QST model	Sketch-500	||	×
	[61]	Graph + semantic based	Sketch-500	||	×
	[57]	Star graph + SVM	TU-Berlin	61.5%	✓
	[56]	Fisher vectors + SVM	improved TU-Berlin	67%	×
	[58]	Star graph + MKL	TU-Berlin	65.8%	✓
Deep features	[69]	Contour + CNN	TU-Berlin	72.2%	✓
	[70]	ImageNet CNN [77] + Improved Lenet [78]	TU-Berlin	56%–77%	×
	[71, 72]	Sketch-a-Net	TU-Berlin	74.9%	✓
	[73]	SketchNet	TU-Berlin	80.4%	×
	[74]	Improved AlexNet	Self-build	||	×
	[76]	Regression model	Sketchy database	||	×

4.1.1. Prior work

SBIR has been extensively studied since the 1990s. However, due to a lack of advanced equipment and efficient solutions, it was challenging to develop a generative SBIR system. Early studies implemented techniques using global color histograms [79], spatial patterns [80], and region adjacency [81,82]. For example, color spots or predefined textures were used to match images [83] and improved robust color-based spots were utilized to complete image retrieval [79,80-82].

Sketch retrieval frameworks were essentially limited to specific categories [84-91]. For example, they were successfully applied to retrieve trademarks [86,87], documents [92], clip-art [85, 90, 91], and technical drawings [88, 89]. Shih and Chen successfully applied this sketch-based system to trademark segmentation and retrieval [86]. Fonseca and Gonçalves proposed a framework which retrieved documents using a sketch of the layout with a calligraphic interface [92]. Wang et al. introduced a system which retrieved cartoon images for a given query [91]. These studies considered the sketches to be based on contour shapes or geometric lines, thus utilizing geometric relationships to present sketches and matching images to accomplish retrieval [93-95]. However, these systems are limited, as they cannot use casual sketches as input to search for real images. Although the studies listed above are based on sketches, they represent some of the earliest SBIR research.

4.1.2. Recent work

4.1.2.1. Works on PC.

- 1. SBIR without RF

(1) **Works for common SBIR**
We begin the discussion on SBIR by highlighting research involving personal computers without RF. As shown in Fig. 7, SBIR systems typically consist of three primary steps: 1) Pre-processing. Pseudo-sketches are presented using approximate sketch synthesis method. 2) Feature extraction. An effective feature extractor is selected to present the features of query sketch and pseudo-sketches, and 3) Similarity measure. The output of the candidate results is ranked and retrieved by the similarity measure matching. Further, commonly used metrics for similarity measurements include City block distance, Cosine, Manhattan distance, and Euclidean distance. Some studies have focused on designing specific efficient matching algorithms. For example, Del Bimbo and Pala proposed a method for applying an elastic matching metric to evaluate the similarity of sketches and images [96]. Hu and Collornosse utilized the metric

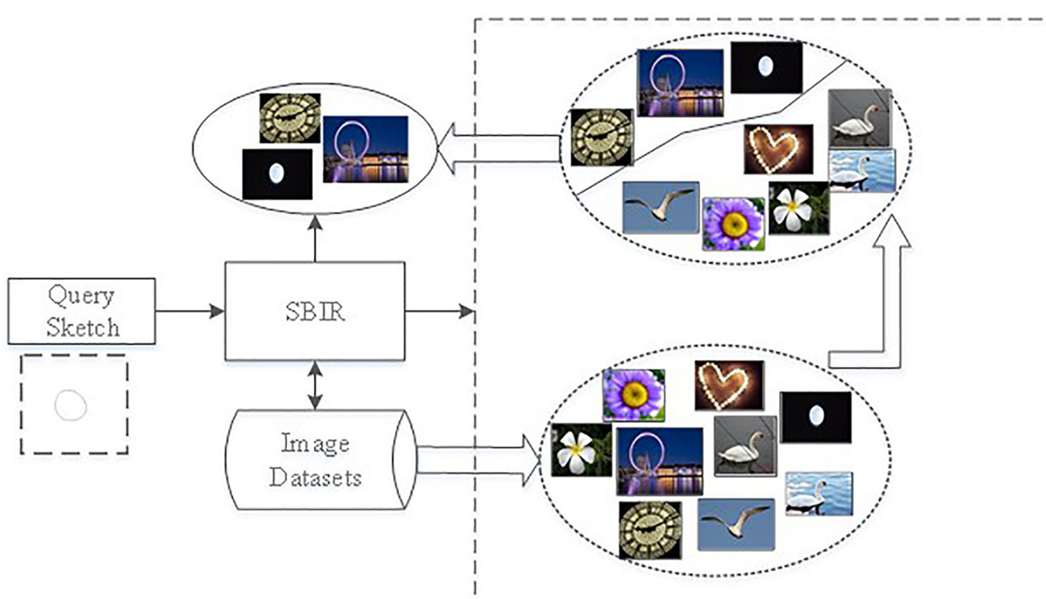

Fig. 6. An example of the SBIR framework.

Table 3
A detailed partition of SBIR discussions.

Works of SBIR	Prior work			
	Recent work	Works on PC	SBIR without RF	Works for common SBIR
				Works for large-scale SBIR
			SBIR with RF	
			SBIR with DL	
		Works on Mobile		

of histogram intersection distance for matching purposes [2]. Sciascio et al. defined a modified cosine distance to perform similarity matching [97].

a. **SBIR algorithms with global descriptors**. Most studies focus on designing feature extractors to efficiently represent sketches and several SBIR studies have started with designing global descriptors. Matusiak et al. adopted the curvature scale space to characterize object contours or edges, where contours were extracted first using contour extraction methods [98-100]. Horace et al. proposed an affine invariant contour description operator, which was based on the convex hull area of the contours used to perform SBIR [101]. Other studies have demonstrated that both the edge direction and its angle can represent contours and have successfully applied this concept to SBIR [102,84,103].
Jain and Vailaya used color and shape descriptors to conduct SBIR [102]. Tao et al. proposed combining the features of length, curvature, and spatial geometric relationships for sketch-based sky image retrieval [104]. Shih and Chen combined various features of invariant moments and the histogram of edge directions, to represent images [86]. Hernández et al. proposed the Garabato system to conduct SBIR [105]. Specifically, they first combined selective smoothing and color segmentation to represent features. The similarities between a user sketch and the image edges were then measured by considering spatial proximity and edge orientation. Pan et al. introduced a color contrast-based saliency method to solve the common problems of scale and translation [106]. In addition, an improved Hausdorff distance was proposed to increase the penalization of outliers and improve robustness. Chao et al. proposed a sketch-based retrieval representation scheme which considered sketch strokes and local features, thereby facilitating efficient retrieval with codebooks [107].
Jing et al. introduced a sketch-based method for manga image retrieval which enabled automatic retrieval of manga images for a given sketch query [108]. As the characteristics of manga images are very different from those of naturalistic images, a fine multi-scale edge orientation histogram (FMEOH) descriptor was used for retrieval. Zhang et al. proposed a method based on image contour segmentation by combining the global and saliency contour maps [109]. They further proposed a new descriptor for SBIR, namely the angular radial orientation partitioning (AROP) feature, which made full use of the gradient orientation to bridge gaps between the sketch and image. Other studies have also discussed the related technological developments of SBIR [110,111]. In contrast to our study, these two surveys presented SBIR methods from differing viewpoints and the described methods are somewhat insufficient.

b. **SBIR algorithms with local descriptors**. In recent years, sketches have generally been represented by local descriptors (or combined local and global descriptors) with various feature operators (SURF, LBP, SIFT, etc.) proposed regularly for SBIR. HOG has emerged as an effective descriptor for representing sketches [59]. HOG is a local descriptor which maintains invariance in image geometric and optical deformations. Based on this, Eitz et al. proposed a strategy using BOW to present sketches and introduced a benchmark for evaluating the performance of large-scale SBIR [112]. Hu et al. proposed a descriptor of gradient field HOG (GF-HOG) combined BOW to perform SBIR [113,114,2]. Hu and Collornosse then evaluated GF-HOG against state-of-the-art descriptors with common distance measurements and language models for SBIR [2]. They also investigated

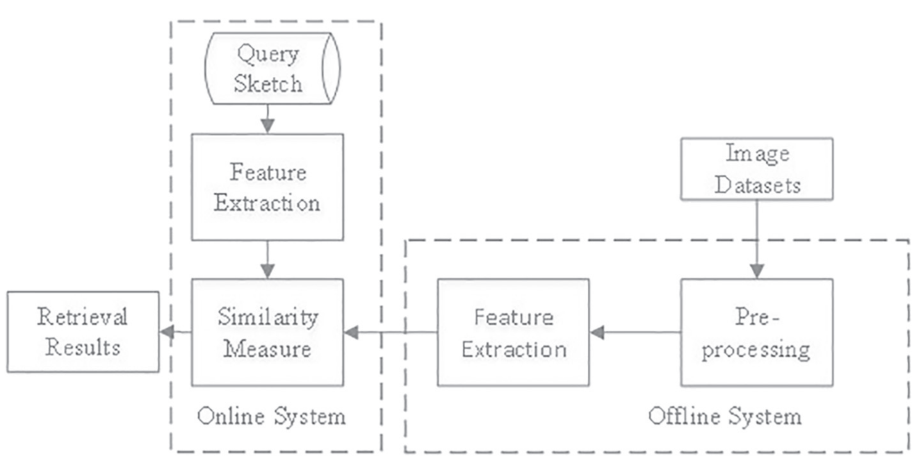

Fig. 7. A typical retrieval framework for SBIR.

affine deformations and their impact on search performance. In their work, GF-HOG was shown to consistently outperform the SIFT operator [115], multi-resolution HOG [59], Self-Similarity [62], Shape Context [116], and Structure Tensor [117] across a large Flickr dataset.

Saavedra defined a new descriptor based on efficient computation of a histogram of edge local orientations (HELO) [118]. Furthermore, they improved their descriptor by introducing soft-HELO (S-HELO), which was based on HELO descriptor. It was improved using a soft computation of local orientations, taking into account spatial information, which was superior to HELO for SBIR [119]. Aarthi and Amudha enhanced the method of indexable oriented chamfer matching (IOCM) [120] using salient feature detection algorithms [121]. Saavedra and Bustos presented a novel method for describing sketches based on detecting mid-level patterns called learned keyshapes (LKS) for poor drawing skills [122]. Bui and Collomosse extended the descriptor of GF-HOG through two technical contributions [123]. The first enabled color-shape retrieval which integrated the modalities of color and shape. The second proposed an efficient inverse-index representation for GF-HOG which delivered a scalable search over millions of images.

Kiumagai et al. improved the edge histogram descriptor (EHD) [124] and proposed an edge relation histogram (ERH) of fusing global and local features to present sketches [125]. Szanto et al. studied EHD, HOG, and SIFT for SBIR and proposed a possible application to match a forensic sketch to a gallery of mug shot images [126]. The ERH descriptor was demonstrated to be invariant to translation, scale, symmetry, and rotation. However, the method was sensitive to noise. In order to overcome the limitations of ERH, Chatbri et al. improved it by introducing the support regions descriptor (SRD) [127]. SRD was a statistical shape descriptor which combines information in near and far support regions, defined for each sketch point, to enhance robustness. They later proposed a scale-invariant shape context (SISC) descriptor [128], which used a size-adaptive window to improve robustness. Cao et al. proposed the symmetric-aware flip invariant sketch histogram (SYM-FISH) descriptor, which was also based on shape context [129]. Wang et al. proposed an SBIR framework consisting of a new line segment-based descriptor called the histogram of line relationship (HLR) and an algorithm known as object boundary selection to reduce the effects of noisy edges [130].

Some studies have focused on searching for other ways to improve the similarity of the image domain and sketch field [131-133]. Methods such as saliency detection and improved new BOW frameworks have been utilized for SBIR. The M-SBIR method generated visual words for both the sketches and images [133]. It then leveraged mapping between the two sets to identify and remove sketch visual words which did not describe the original images well. Other studies have used a method of object saliency to represent images and reduce the cross-domain gap in SBIR [131,132].

c. **SBIR algorithms with mixed queries**. Another type of retrieval involves a query composed of a sketch and keywords (such as text tags). Chen et al. proposed a framework for Sketch2Photo, which used a keyword to trigger a Google image search and returns related images [134]. The results were then filtered by matching the Shape Context of the sketches [116]. Wang et al. proposed a fusion mechanism for text-shape [135]. They later improved the sketch matching process by utilizing a raw curve-based algorithm to efficiently and precisely calculate the similarity between the salient curve representation of natural images and a user sketch query to enhance retrieval efficiency [136]. Another form of interactive sketch-based image synthesis involved generating realistic images by querying a scene with freehand sketches [134,137]. Chen et al. proposed a novel image-blending algorithm to allow seamless image composition for generating photo-realistic pictures from a casually drawn sketch with added text labels [134]. Eitz et al. introduced a system called PhotoSketch for progressively creating images through a simple sketching and compositing interface [138,137]. The compositing process was based on graph cuts and Poisson blending. Images in this database could be searched by matching user binary outline sketches.

(2) **Works for large-scale SBIR**

While most research has focused on SBIR for small-scale datasets, large-scale SBIR is highly beneficial for ensuring a system to find exact results for any sketch query. While considering the computational complexity involved, Eitz et al. proposed a real-time sketch-based large-scale SBIR system which searched a database containing over one million images [139]. Their algorithm was based on a descriptor which elegantly addressed the asymmetric shape between binary user sketches and full color images. Wang et al. demonstrated the MindFinder system, which was an interactive sketch-based image search engine and was a sketch-based multimodal search engine for million-level database. To bridge the gap between the sketch query and the full color image, they represented features by its salient curves, then described a visual word using the position of an edge pixel (edgel) and the edgel orientation at that position. They further proposed an indexable oriented chamfer matching (IOCM) to realize a fast matching between the sketch and image [120]. A novel indexed structure and a corresponding raw contour-based matching algorithm were later proposed to calculate the similarity between a sketch query and natural images [136]. This system developed a real-time sketch-based search engine by indexing more than 2 million images simultaneously, while considering storage costs, retrieval accuracy, and efficiency.

Zhou et al. used human perception mechanisms to identify two types of regions in an image: a primary region and a region of interest (ROI) [140]. Two types of candidate regions were then identified for feature extraction. To accelerate retrieval speed, they first extracted orientation features and then organized them in a hierarchal system to generate global-to-local features. Sun et al. built a sketch-based image retrieval system on a billion-level database by using raw edge pixels and oriented chamfer matching (OCM) [141].

Hashing is a useful method for increasing the speed of calculation while preserving storage space. Hashing-based frameworks have been utilized to facilitate SBIR in large-scale databases [142]. Parui and Mittal demonstrated an efficient approach for large-scale SBIR utilizing image contour chains, a fast dynamic programming-based (DP) matching algorithm, and hierarchical k-medoids indexing [143]. Liang et al. introduced a generic hashing-based approach where locality-sensitive hashing (LSH) was employed to capture content [144]. In addition, hash codes were indexed so that a query could be processed in sub-linear time to facilitate sketch-based retrieval of large of visual shape datasets. Li studied the suitability of the Kd-tree indexing mechanism to accelerate the retrieval process for a sketch-based retrieval system based on shape descriptors [145].

- 2. SBIR with RF. RF is an automated process introduced in the mid-1960s to produce improved query formulations following an initial retrieval operation [146]. It was initially developed for document retrieval (information retrieval), then transformed and introduced into content-based multimedia retrieval (mainly for content-based image retrieval - CBIR) during the early and mid-1990s [147]. Algorithms implementing RF resulted in dramatic performance improvements for CBIR systems [148-153]. Cheng et al. presented a unified RF framework for web image retrieval [151]. The method integrated texture and visual features for web retrieval in a practical manner. Meng et al. introduced an RF system using explored multiple modalities in a graph-based learning scheme, which integrated the learning of relevance scores, weights of modalities, a distance metric, and scaling for each modality into a unified frame [152]. Yang et al. proposed the use of image click-through data (which can be viewed as implicit feedback from users) to overcome the intention gap and further improve image search performance [153]. Generally, these RF methods are useful for improving the overall performance of CBIR.

  Little research has been done on the use of RF for SBIR. Most studies on SBIR using RF focus on improving retrieval efficiency [97,89, 154-156]. Others have combined query sketch and RF as techniques to improve retrieval accuracy as Di Sciascio et al. proposed gaussian Markovian random fields (GMRF) for sketch retrieval [97]. A three-layer relevance feedback architecture was proposed to progressively refine retrieval results according to user preferences. Li et al. introduced RF algorithms based on different classifiers and presented an RF method for sketch retrieval by means of a Linear Programming (LP) classifier [154,89]. This biased SVM (BSVM) classifier was then used in RF to improve retrieval effectiveness [89]. Qian et al. proposed an effective SBIR approach which included re-ranking and relevance feedback schemes [156]. The re-ranking via visual feature verification (RVFV) approach took advantage of semantics in query sketches and top-ranked images in the initial results. The contour-based RF (CBRF) was then applied to detect more relevant images for the input query sketch. The integration of these two schemes improved the overall SBIR performance.

  In contrast, applications of the RF mechanism used in SBIR were much lower than in CBIR or TBIR. This is likely due to a couple of factors. Firstly, using RF often requires manual participation in the screening process, which requires additional effort and users typically do not want to be explicitly involved in the feedback process. Secondly, using RF strategies requires greater computational complexity in the retrieval system. With further developments in SBIR research and the computing capabilities of computer hardware, there is little doubt that sketch retrieval based on relevance feedback will improve in the future.
- 3. SBIR with DL. With the rapid development of deep learning theory, there has been explosive growth in image applications using DNNs. Deep learning has been very successful for a variety of CV tasks such as image classification, object detection and localization, and saliency detection. There have been several studies on SBIR using deep learning [157-160]. Wang et al. demonstrated that deep convolutional neural networks (DCNN) performed better for SBIR [157]. The results of learned deep features with standard feasible data-argumentation largely surpassed human-level performance for benchmark datasets. Qi et al. adopted the deep model Siamese Convolutional Neural Network (SCNN) to conduct SBIR, as it was better suited for learning the deep features in sketches and images [158]. Bui et al. proposed an efficient representation for SBIR, derived from a triplet loss CNN [160]. They achieved depiction invariant embedding using SCNN. They uniquely modified the triplet loss training function and demonstrated that learned deep features were better generalized in different image categories. They also demonstrated that deep features outperformed state-of-the-art SBIR for the standard Flickr15k dataset.

  Further, GAN has produced remarkable results since its introduction. Creswell and Bharath introduced a novel GAN architecture suitable for sketch retrieval SBIR [159]. Experiments suggested that deep representations of sketch-GANs increased stability for sketch rotation, scale, and translation. It is obvious the technologies in deep architectures have provided superior performance compared to manually-designed features and are now more widely used for SBIR and FG-SBIR than low-level feature models.

4.1.2.2. Works on mobile. SBIR has been the focus of rapid development, owing to several open-access sketch benchmark datasets which have become available since 2012. The vast majority of these sketch research studies have been conducted on computers, while studies on mobile devices are sparse. This is likely because research algorithms are more easily implemented on computers. Mobile devices also have limited memory which can restrict algorithm complexity. Recent studies on mobile devices have primarily focused on reducing retrieval time and improving retrieval accuracy for large datasets [123, 160-163]. Zhang et al. proposed a retrieval system based on edge-sift for mobile devices which retrieved similar images based on an input image query [162] and Yang et al. presented an effective scalable mobile image retrieval approach by exploring contextual salient information for the input query image combined with the high-level semantic information to find relevant photos [163]. Tseng et al. proposed an SBIR framework for mobile devices which used a distance transform (DT) to precisely capture features in a sketch image [161]. This reduced the memory storage by further projecting high-dimensional DT into compact hash bits, which allowed the operation to be performed using the limited memory of a mobile device. Bui and Collomosse proposed an SBIR system accompanied by a mobile app demo (for Android) [123]. They also developed an SBIR utilizing deep models and deriving a compact image descriptor from the learned features, which was suitable for indexing data on memory-constrained mobile devices [160].

4.2. Primary datasets

In this section, SBIR research is divided into two categories: research on computers and on mobile devices. Although not a complete source, we also list primary methods and datasets. TU-Berlin and Flickr15k are widely used SBIR benchmark datasets.

The Flickr15 K dataset, developed by Hu and Collornosse, consists of approximately 15,000 images which are divided into 33 categories and iconic classes (i.e., airplane, duck, horse, etc.) [2]. It is a challenging dataset as the number of images in each category varies significantly, from 21 images (horse class) to 1600 images (Eiffel Tower class). Also, the background complexity of the images is not uniform. Additionally, the query sets consist of 10 folders, each containing 33 sketches corresponding to the 33 categories mentioned above. Freehand sketch queries were provided by 10 amateur sketchers, comprising a benchmark dataset which is widely used for SBIR tasks.

Flickr 160 database was used for measuring a retrieval system. 160 pieces of general-themed pictures have sorted from the photo sharing website called Flickr. The images can be classified into 5 classes based on their shape. A lot of images contained the same building and moments.

MSRA10 K dataset contained 10,000 images with consistent bounding box and pixel-level saliency labeling for 10 K images from MSRA dataset. MSRA dataset, which was originally provided salient object annotation in terms of bounding boxes provided by 3-9 users, was widely used in salient object detection and segmentation.

The Oxford Building dataset consisted of 5062 images collected from Flickr by searching for particular Oxford landmarks, 11 landmarks in total. 55 query images given by Oxford Buildings Dataset as test set.

GOLD was a geo-tagged large-scale web image set associated with their geographic coordinates, which is crawled from Flickr. GOLD contained more than 227 thousand images together with 80 places-of-interests which are selected from 60 world-wide cities with about 3.3 million images.

4.3. Summary and conclusions

In this section, we mainly describe PC algorithms based on several sub-categories including SBIR without RF, SBIR with RF, and SBIR with deep learning. Specifically, we present conventional methods for large-scale SBIR, methods with RF, and methods with deep learning for SBIR. Most studies which do not include RF focus on designing various operators to present sketches effectively. Operators such as Sift, Shape Context, HOG, GF-HOG, HELO, and S-HELO have been successfully applied to SBIR. Studies involving RF primarily focus on improving retrieval accuracy for a given query. Research involving the learning of deep features to represent sketches has become increasingly common in recent years. Deep architecture produces better results for SBIR and deep features have proven to be superior to low-level features in solving the majority of CV tasks. In addition, as deep features can distinguish more subtle differences at an instance level, techniques which adopt deep learning are more applicable for FG-SBIR than for SBIR. SBIR methods for mobile devices focus on compact hashing, effective index structure, and deep learning. A summary of these methods and datasets is provided in Tables 4 and 5.

SBIR has received increased attention in recent years. In contrast to CBIR, SBIR is unique as query sketches differ significantly from photos. SBIR also represents a cross-modal recognition problem but is crucial for assisting CBIR. To further improve SBIR in the future, additional deep models must be designed to improve retrieval accuracy and more extensive sketch training databases (such as ImageNet) will need to be developed. Knowledge of transfer learning and unsupervised machine learning may be beneficial for solving sketch training data problems with insufficient labels. Finally, with the high growth rate of available multimedia data, methods of compact hashing could be employed to improve retrieval effectiveness and reduce computation.

The work of SBIR has undergone the transformation of hand-crafted features (global operators and local operators) to deep features in recent years. With the use of fine-grained retrieval, many applications have gradually shifted to FG-SBIR. However, SBIR still has considerable significance. For readers interested in the work of SBIR, we suggest the following studies. There are several conventional methods for manual features [1, 2, 113, 114, 118, 119, 125, 127-129]. There are also methods used to acquire deep characteristics [157-160]. We strongly suggest the use of deep learning to establish a deep framework to solve SBIR problems.

5. Fine-grained sketch-based image retrieval

5.1. Technological developments

SBIR primarily focuses on retrieving images from the same category, based on a user sketch query, while neglecting the fine-grained characteristics of sketches. However, fine-grained sketch-based image retrieval (FG-SBIR) attempts to retrieve a particular photo based on the input query. In other words, SBIR is a category-level retrieval method while FG-SBIR is conducted at an instance level. Fig. 8 shows the framework for a cross-domain FG-SBIR system, which can be extremely challenging. The difficulty arises from several sources, including the uniqueness of abstract freehand sketches, comparing and matching across two domains, and attempting to distinguish subtle differences within a class.

Investigations into the fine-grained CBIR problem have resulted in useful methods which could be applicable to FG-SBIR. For example, latent conditional random field models have been used to discover detectable and discriminative fine-grained local attributes [164]. AirplanOID, a large dataset of airplane images annotated in detail, allowed for the study of fine-grained object descriptions and established a relationship between object parts and attribute semantics [165]. Furthermore, a coarse-to-fine technique was used to efficiently detect richer part models in the images. Wang et al. proposed a deep ranking model which learned similarity metrics directly from images [166]. An efficient triplet sampling algorithm was also proposed to train the model with distributed asynchronous stochastic gradients. Zhang et al. designed an improved CNN model aimed at differentiating subtle variations among subordinate classes [167]. Zhou et al. investigated the problem of recognizing fine-grained food images and developed an engine to identify the dish and restaurant of origin [168]. They proposed a novel approach to exploit rich relationships through bipartite-graph labels (BGL) in an overall CNN. To facilitate this study, they constructed a food benchmark dataset which contained 37,885 food images collected from 6 restaurants and 975 menus. Deep features have been widely adopted for solving fine-grained problems inherent in CBIR [166-169]. Research in FG-SBIR is divided into two aspects in this paper:

- FG-SBIR using hand-crafted models.
- FG-SBIR using deep learning models.

5.1.1. FG-SBIR using hand-crafted models

Li et al. first proposed the concept of FG-SBIR [170]. They advocated the expressiveness and efficacy of sketches under a novel FG-SBIR framework. Specifically, they introduced a learned deformable part-based model (DPM) which can both capture object pose and traverse sketch and image domains. Thus, mid-level representations established pose correspondence across two domains in order to perform FG-SBIR effectively. Later, part-level detection of visual attributes was achieved in sketches [171,172]. A corresponding dataset was introduced with 304 photos and 912 sketches, each being annotated with a semantic component and associated part-level attributes. A strongly-supervised DPM (SS-DPM) was trained using automatic detection of part-level attributes. A novel matching framework combined with canonical correlation analysis (CCA) was then constructed to integrate low- and mid-level geometric structures with high-level semantic attribute features to boost retrieval performance. Masi et al. established a fine-grained cross-domain space and proposed a joint view selection and attribute subspace learning algorithm to learn domain projection matrices for photo and sketch retrieval [173].

Table 4
Primary methods and datasets used for SBIR.

Types	Methods	Publications	Approach	Datasets	Comparable
On computers	No RF	[113, 114]	GF-HOG + BOW	Flickr160 + ETHZ	✓
..	..	[2]	GF-HOG + BOW	Flickr15 K	✓
..	..	[1]	HOG + BOW	self-build	✓
..	..	[118]	HELO	Self-build	✓
..	..	[119]	S-HELO	Flickr15 K	✓
..	..	[125]	ERH	Self-build	×
..	..	[127]	SRD	Self-build	✓
..	..	[128]	SISC	Self-build	✓
..	..	[129]	SYM-FISH	ETHZ	×
..	..	[105]	Edge-based	Self-build	×
..	..	[106]	Saliency-based	MSRA10 K	×
..	..	[108]	FMEOH	Self-build	×
..	..	[109]	AROP	Self-build	×
..	..	[121]	Modified IOCM	TU-Berlin	×
..	..	[122]	LKS	Flickr15 K	✓
..	..	[120]	HLR	dataset [1] + Flickr15 K	✓
..	Large-scale SBIR	[139]	Edge histogram + Tensor	Flickr	✓
..	..	[136]	IOCM	Flickr	✓
..	..	[140]	Hierarchal Orientation	Flickr	×
..	..	[141]	OCM	Sketch500 + TU-Berlin	×
..	..	[142]	Hashing-based	TU-Berlin	×
..	..	[143]	Contour + DP	Flickr	×
..	..	[144]	Hashing-based	Tu-Berlin + PSB	×
..	..	[145]	KD-tree	Caltech	×

5.1.2. FG-SBIR using deep learning models

Several studies have utilized deep learning to conduct FG-SBIR [174-181]. Yu et al. constructed a triplet work model to perform FG-SBIR [174]. They introduced a new database of 1432 sketch-photo pairs in two categories (shoes and chairs) and developed a deep triplet-ranking model with novel data augmentation and a pre-training strategy for FG-SBIR. Song et al. expanded this work by proposing a deep multi-task attribute-driven ranking system for FG-SBIR [175]. They constructed a deep model which could perform attribute prediction and attribute-based ranking side tasks in parallel with the primary retrieval task. As such, the main task representation was enhanced due to the ranking of attribute predictions. This triplet work model was later combined with an attention mechanism for conducting retrieval [178,180]. Sangkloy et al. proposed an FG-SBIR method using the GoogLeNet [176]. They developed the Sketchy database, the largest collection of sketch-photo pairs to date, and designed a cross-domain convolutional network which embedded sketches and photographs in a common feature space. The learned deep representations significantly outperformed hand-crafted features. Song et al. [181] constructed a multimodal quadruplet deep network which formulated to jointly model the sketch and text input modalities as well as the photo and they further introduced a multimodal fine-grained image retrieval dataset with both sketches and sentences description provided as query modalities. Pang et al. proposed a novel discriminative-generative hybrid model which used auto-encoding and decoding to narrow the domain gap [179]. Later, they proposed a novel unsupervised image style transfer model based on enforcing an embedding consistency constraint [182]. A deep FG-SBIR structure was then formulated to accommodate discriminative detail from each factorized sketch for better matching with the corresponding photo.

Taking into account the scale and computational complexity of fine-grained retrieval, Liu et al. accelerated FG-SBIR by introducing a novel binary coding method called Deep Sketch Hashing (DSH) [177]. In this process, a semi-heterogeneous deep architecture was incorporated into an end-to-end binary coding framework. Auxiliary sketch tokens were adopted as bridges to mitigate sketch-image geometric distortions during the training process. Xu et al. also developed a deep framework based on hashing for sketch retrieval [183]. They introduced a novel hashing loss based on a CNN-RNN architecture that adapted the abstract nature of sketches, and the proposed model can further yield a better generalization performance under a zero-shot setting for sketch recognition.

Table 5
Additional methods and datasets used for SBIR.

Types	Methods	Publications	Approach	Datasets	Comparable
..	DeepModels	[157]	DCNN	TU-Belin + PASCAL	×
..	..	[158]	SCNN	Flickr15 K	✓
..	..	[159]	GAN + CNN	Self-build	×
..	..	[160]	Triplet CNN	TU-Berlin + Flickr15 K	✓
..	With RF	[97]	GMRF	||	×
..	..	[154]	Linear Programming	Self-build	×
..	..	[89]	Features + BSVM	Self-build	×
..	..	[155]	Coarse-To-Fine Features	||	×
..	..	[156]	RVFV + CBRF	Sketch-500 + TU-Berlin	×
On mobiles	..	[161]	Hashing-based	Caltech + Flickr10k	×
..	..	[160]	Triplet CNN	TU-Berlin + Flickr15k	×
..	..	[162]	Edge-sift	||	×
..	..	[163]	Contextual saliency	Oxford Building + GOLD	×
..	..	[123]	extended GF-HOG	Flickr15K + self-build	×

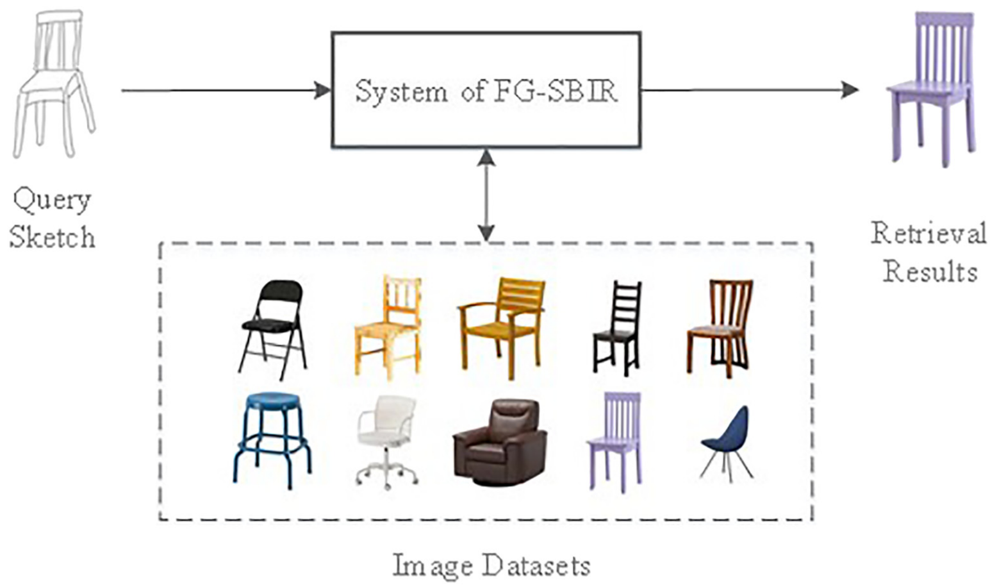

Fig. 8. A sample framework for an FG-SBIR system.

5.2. Primary datasets

FG-SBIR is a more challenging task than SBIR but has achieved significant success in recent years. Conventional benchmark datasets include the Sketchy database and the Shoe-chair dataset etc..

The QMUL-Shoe-Chair dataset (http://sketchx.eecs.qmul.ac.uk/downloads/), includes two image categories, shoes and chairs, which are collected by asking participants to hand-sketch an object after observing a photo. The dataset contains 716 sketch-photo pairs in total, with 419 pairs of shoes and 297 pairs of chairs. Additionally, a total of 32,000 ground truth triplet-ranking annotations were provided for both new model development and performance evaluation.

QMUL-Shoe-Chair-V2 (http://sketchx.eecs.qmul.ac.uk/downloads/). It is an extension of QMUL-Shoe-Chair sketch-image database. There are 6648 sketches and 2000 photos for the shoe category. For the chair category, there have 400 photos and 2000 sketches. This dataset has the largest single-category for FG-SBIR so far.

The Sketchy database, developed by Sankloy et al. (http://sketchy.eye.gatech.edu/) is the first large-scale collection dataset of sketch-photo pairs for FG-SBIR. The authors applied crowd-working to obtain sketches of particular photographic objects sampled from 125 categories, ultimately acquiring 75,471 sketches of 12,500 objects. The Sketchy database defines fine-grained associations between particular photos and sketches. It is an open-access fine-grained sketch database which can be downloaded free of charge. The Sketchy database is widely used as a benchmark for FG-SBIR and offers new opportunities for sketch and image synthesis.

5.3. Summary and conclusions

FG-SBIR involves the retrieval of images at an instance-level based on a query sketch. It represents a more difficult challenge as it must distinguish subtle differences within a class. Most research in FG-SBIR adopts a deep learning framework to accomplish this task. Multi-task CNN models, deep hashing, and triplet network models are widely used. Furthermore, several large datasets have been built to facilitate research in order to better understand fine-grained sketch-image pairs and solve the fine-grained recognition problem. The primary methods and datasets used in this section are shown in Table 6, where entries in 'Accuracy' column are assigned the top value if belonging to the same database.

FG-SBIR is more useful in practical applications than SBIR which can be implemented well by TBIR. FG-SBIR also provides several practical benefits. High-level features (such as semantic features of part-annotations) can be used to improve the retrieval accuracy of FG-SBIR [184]. Knowledge of sparse hashing or compact hashing can be adopted to reduce complexity and improve the overall retrieval efficiency. Also, by considering the uniqueness of sketches (highly abstract, large variations, etc.), the semantic information in sketch strokes can be used in deep models to improve FG-SBIR efficiency. More recently, the use of DL in FG-SBIR problems has undergone rapid development. Most studies have adopted a deep model using deep features to further reduce the domain gap [174-177, 179-181]. The Siamese network [174] (two branch CNNs are trained without sharing weights) and the heterogeneous network [176] (two branch CNNs are trained without sharing weights) are especially popular in fine-grained direction retrieval. Therefore, for those who are fond of FG-SBIR, methods which are conducted by using deep models are highly encouraged to refer.

6. Sketch-based 3D shape retrieval

6.1. Technological developments

The query modality is a central issue for 3D model retrieval as effective methods for constructing 3D models become ever more important. Query-by-3D model is a popular modality but it is often unavailable as a query option. A possible alternative involves using a set of keywords as a query. However, most 3D models lack text metadata and it is difficult to specify geometrical shapes using words alone. Another alternative is the use of 2D freehand sketches as queries for 3D shape retrieval. Sketch-based 3D model retrieval attempts to retrieve a list of 3D models based on query sketches. This process is more intuitive and convenient for novice users to learn.

Sketch-based 3D shape retrieval is growing rapidly. It is an important component of sketch recognition and has been the focus of extensive research in the last several years. Comprehensive studies have been published recently but we limit our discussion to the relevant developments in sketch-based 3D shape retrieval over the past few years [185,186]. As hand-crafted features and deep features have all been investigated [187-205] in recent years, we divide 3D model retrieval into two categories as well:

Table 6
Primary methods and datasets used for FG-SBIR.

Methods	Publications	Generation approach	Datasets	Accuracy	Comparable
Hand-crafted Features	[170]	DPM	self-build	||	×
..	[171]	SS-DPM + CCA	self-build	||	×
..	[173]	Joint Subspace	QMUL-Shoe-Chair	Shoe: 34.78%, Chair: 36.4%	✓
Deep Models	[166]	Triplet models	ImageNet + Self-build	||	×
..	[167]	Label structure + CNN	Stanford Car + Car-33	||	×
..	[168]	CNN + BGL	Self-building	||	×
..	[174]	Triplet models	QMUL-Shoe-Chair	Shoe: 39.13%, Chair: 69.07%	✓
..	[175]	Attribute + Triplet models	QMUL-Shoe-Chair	Shoe: 50.43%, Chair: 78.35%	✓
..	[176]	Modified GoogLeNet	Sketchy	37.1%	×
..	[177]	DSH	Sketchy	||	×
..	[178]	Triplet models	Sketchy	41.65%	×
..	[180]	Attention + Triplet models	QMUL-Shoe-Chair	Shoe: 61.74%, Chair: 81.44%	✓
..	[179]	Generative models	Sketchy	50.14%	×
..	[181]	Quadruplet model	self-build	||	×
..	[182]	Cyclic embedding consistency	QMUL-Shoe-V2 + Sketchy	||	×
..	[183]	CNN-RNN deep model	QuickDraw	||	×

- sketch-based 3D model retrieval using hand-crafted models.
- sketch-based 3D model retrieval using deep learning models.

6.1.1. Sketch-based 3D model retrieval using hand-crafted models

Lee and Funkhouser [193] constructed a model for 3D shape based on modeling on example (MBE) [206]. A structure-based local approach (STELA) was presented for retrieval which used the contours of 3D model [196]. Later, Eitz et al. collected a large number of benchmark datasets which related to existing 3D objects and produced a Gabor local line-based feature (GALIF) framework with BOW features for a retrieval system [189], Xie et al. also used GALIF combined with BOW for 3D shape retrieval [190]. Furuya and Ohbuchi first proposed a cross-domain manifold (CDMR) algorithm to conduct 3D shape retrieval which can form two independent manifolds of features based on sketch and 3D model [195] and then presented a novel hamming-space embedding of cross-modal manifold hashing graph model (CMMH) [191]. They trained a combination of Laplacian eigenmaps and applied iterative quantization hashing to accelerate retrieval results.

Kang et al. developed a retrieval system based on three-view sketches using a bank of Gabor filters for generating local features [188]. Tu et al. proposed a novel cascaded 3D model retrieval framework based on locality-sensitive discriminant analysis (LSDA) for automatic and accurate fitting of 3D models to 2D freehand sketches [198]. Mao et al. developed a new method for fusing global and local features based on a multi-view rendering of 3D models with 2D sketches [199]. In contrast, Li et al. proposed using global and local features which were extracted from representative 2D views of 3D models [200]. These global features represent exterior boundary shape views and the local features represent their interior details. Yasseen et al. presented a new approach for 3D object retrieval which described a 2D shape from the visual components of its silhouette [201]. In addition, they presented a list of candidate 2D projections which represent the canonical views of a 3D object.

6.1.2. Sketch-based 3D model retrieval using deep learning models

Deep model methods have been detailed in several studies [192, 202-205]. Bu et al. used deep belief networks (DBNs) to learn a model to generate the local deep feature (LDF), which was high-discriminative and effective for 3D shape applications [203]. Leng et al. [204] proposed a method of greedy layerwise strategy to train stacked local convolutional auto-encoder (SLCAE) to solve 3D shape retrieval. In addition, Su et al. [205] presented a novel CNN architecture that combined information from multiple views into a single and compact shape descriptor for 3D recognition.

Fang et al. learned features for both sketches and 3D model views using a DL framework [192]. Specifically, they drastically reduced the number of projections to two predefined directions for the entire dataset and designed a siamese convolutional neural network to learn deep features and conduct retrieval. Zhu et al. introduced a novel cross-domain neural network (CDNN) approach which further extended a pyramid cross-domain neural network (PCDNN) by cooperating with a hierarchical structure [202]. By constructing cross-domain neural networks at multiple pyramid levels, a many-to-one relationship was established between 3D shape features and sketch features.

6.2. Primary datasets

Most sketch-based 3D shape retrieval techniques focus on one of three objectives: choosing accurate multi-view projections of 3D models, constructing accurate features in 2D sketches and 3D models, or improving retrieval accuracy and efficiency for 3D shape models from a given query sketch. In order to resolve the first problem, various algorithms have been proposed to reduce the effects of view projections. Various approaches have been used to represent the features of sketches and 3D shape models, such as the combination of multilevel features using multiple pyramid levels and deep features learned by deep models. Additionally, cross-domain deep hash algorithms have been developed to improve retrieval time. To better understand the relationship between sketches and 3D shapes, and promote the progress of sketch-based 3D shape retrieval, several large and extended datasets have been established.

The 3D Shape Retrieval Contest (SHREC) (http://www.shrec.net/) is used to evaluate the effectiveness of 3D shape retrieval algorithms and has provided several resources to compare and evaluate 3D retrieval methods. The annual competition publishes a database related to the 3D works which is composed of different data parts used to meet different tasks. As such, several sketch-based 3D retrieval methods adopt this dataset as a benchmark (as shown in Table 7).

The Princeton Shape Benchmark (PSB) (http://shape.cs.princeton.edu/benchmark/) provides a repository of 3D models for evaluating shape-based retrieval and analysis algorithms. This benchmark contains a database of 3D models collected from the World Wide Web and has been split into a training database and a test database. In Version 1, the training and testing samples were comprised of 907 3D models. This is a commonly used benchmark for 3D shape retrieval.

Eitz et al. were the first to collect a significant number of sketches for the evaluation of shape retrieval performance [189]. Their dataset was based on the freely available and widely accepted set of models from the PSB. This benchmark includes all categories defined in

Table 7
Primary methods and datasets used for sketch-based 3D shape retrieval.

Methods	Publications	Generation approach	Datasets comparable	
Hand-crafted features	[193]	Extended MBE [206]	PSB	×
..	[196]	STELA	PSB	✓
..	[189]	BOW + GALIF	S-PSB	✓
..	[187]	BOW + Synthesize	dataset of [207]	×
..	[190]	GALIF + BOW	Self-build	×
..	[188]	Local features + Gabor	PSB	×
..	[195]	CDMR	S-PSB/SHREC'13	✓
..	[191]	CMMH	S-PSB + SH14 + SH14X	✓
..	[198]	LSDA	||	×
..	[199]	Multi-feature Fusion	||	×
..	[200]	Feature Fusion	||	×
Deep models	[192]	SCNN	PSB/S-PSB/SHREC'13/'14	×
..	[202]	PCNN	SHREC'14	×
..	[203]	DBN + LDF	SHREC'2007	×
..	[204]	SLCAE	PSB/SHREC'09	×
..	[205]	Multi-view CNN	Self-build	×

the PSB and gathers sketches for all models in the PSB. This set of benchmark sketches is available as a free resource and is referred to as S-PSB in this study.

Li et al. extended a large-scale sketch-based 3D shape dataset [194]. This benchmark dataset, which was included 171 distinct classes, contained 12,680 2D sketches and 8987 3D models to promote the progress of research conducted by the 3D model retrieval community. The 2D sketch query set contained 13,680 sketches from Eitz [1] (171 classes, each with 80 sketches). The 3D model dataset in this benchmark contained 8987 models (171 classes, each with around 53 models).

6.3. Summary and conclusions

We have briefly discussed the topic of sketch-based 3D shape retrieval, excluding early developments, and have provided an outline of primary techniques used in the last several years. Recent studies on sketch-based 3D shape retrieval have focused on designing effective features to bridge the gap between different domains. These design features are based on various hand-crafted methods or are based on learning deep features using deep architectures. The primary methods and benchmark datasets used in this section are listed in Table 7.

The process of 3D shape retrieval is becoming more important and common in the CV field. One promising direction related to improved learning of multimodal features involves modified deep CNNs, such as the technique proposed by Fukui et al. [208]. This process uses deep models to design a common public subspace for different modality features, which can be widely used in sketch-based 3D shape retrieval. Furthermore, transfer learning can be used to overcome the problems associated with limited training data.

For readers who are interested in sketch-based 3D shape retrieval, methods based on deep models (listed in Table 7) provide the necessary technical information for developing unique structures. The study by Eitz et al. offers a detailed discussion of sketch-based 3D shape retrieval though it is based on hand-crafted features [189].

7. Conclusions

7.1. Understandings of sketch recognition and retrieval

Sketch recognition and retrieval is one of the tasks of image recognition in computer vision. The purpose of hand-drawn sketch recognition is to enable people to express information simply and quickly in a natural way, while being directly understood by the computer, without worrying about the creative ideas or workflow interruption, and it is an efficient means of information transformation. Sketches have unique attributes and properties compared to other types of images, such as highly abstract and sparsity. Different recognition tasks have different requirements for identification technology. The rapid development of sketch recognition and retrieval technology is inseparable from the great success of deep learning in the image field and the constantly established large-scale sketch databases in recent years. The essence of sketch recognition and retrieval task is on how to propose more discriminative effective features and how to reduce the domain gaps between two or more different media data. Anyway, breakthroughs in sketch recognition and retrieval technology will inevitably promote the great development of the visual recognition in many fields.

7.2. Future sketch-based recognition research

This paper has provided a review of sketch-based recognition, highlighting techniques for solving sketch-based tasks which have achieved impressive results. It is expected that additional diverse techniques will emerge with the rapid evolution of digital intelligence-devices and future methods. We believe this promising research in sketch recognition will primarily encompass the following areas.

Further development of algorithms based on DL models will accompany the expanded use of big data. Deep models have been demonstrated to be effective tools in a variety of CV tasks. Image features learned from CNN architectures have exhibited superior performance to manually-designed features. New deep GAN models and derivative networks such as info-GAN [209], CGAN [210], and CSGAN [211] have been applied extensively in image vision processing tasks [212]. In addition, GAN and improved GAN can be used in conjunction with different deep convolution networks such as CNN or RNN for different applications. It is reasonable to expect that new deep learning models will be developed for sketch-based investigations.

The majority of sketch-based image retrieval studies currently use low features in images and sketches. However, there exists a semantic gap between human perception and low-level image understanding. As such, it is necessary to use high-level features to bridge this gap and reduce the differences among various modalities. Several studies have utilized high-level semantic features for practical image recognition [213-217]. However, they represent sparse investigation regarding high-level semantics in the field of sketch recognition. Therefore, the use of semantic attributes in sketches should be investigated for future sketch-based recognition studies.

The growth of big data studies have made large quantities of data available in a variety of fields. As mentioned, it is advantageous to use deep network models to solve various CV tasks. However, supervised learning typically requires massive training data and available sketch data (especially sketch data labeled with semantic attributes) is typically not large enough for deep network training. In order to solve this problem, recent research in one-shot and zero-shot learning have achieved good results, especially for small training samples [208,218,219,220,221]. In addition, unsupervised deep models (such as DBN and auto-encoders) can also be employed to reduce this problem [222]. Hence, transfer learning and multi-task learning should be included in sketch-based tasks in the future.

7.3. Conclusion

A complete review of current research on freehand sketches has been provided in this study. We described sketch recognition based on five aspects: sketch generation (sketch synthesis), sketch classification, sketch-based image retrieval, fine-grained sketch-based image retrieval, and sketch-based 3D shape retrieval. SBIR and FG-SBIR were the most important components involved in this investigation. For each aspect, we focused on reviewing the main techniques and primary benchmark datasets used to solve various sketch-related tasks. We also discussed promising future research directions based on three aspects discussed at the end of the paper. Sketches are as old as humanity itself and with the increasing popularity of touch-screen equipment, we expect future research in sketch recognition will develop in a more comprehensive and systematic manner.

Declaration of Competing Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the manuscript entitled “A survey on freehand sketch recognition and retrieval”.

References

[1] M. Eitz, J. Hays, M. Alexa, How do humans sketch objects? ACM Trans. Graph. 31 (4) (2012) 1–10. http://dl.acm.org/citation.cfm?doid=2185520.2335395. https://doi.org/10.1145/2185520.2335395.
[2] R. Hu, J. Collornosse, A performance evaluation of gradient field hog descriptor for sketch based image retrieval, Comput. Vis. Image Underst. 117 (7) (2013) 790–806. https://www.sciencedirect.com/science/article/pii/S1077314213000349. https://doi.org/10.1016/j.cviu.2013.02.005.
[3] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C.C. Loy, X. Wang, A survey on heterogeneous face recognition: sketch, infra-red, 3D and low-resolution, Image Vis. Comput. 56 (2016) 28–48. http://www.sciencedirect.com/science/article/pii/S0262885616301524. https://doi.org/10.1016/j.imavis.2016.09.001.
[4] A. Limpaecher, N. Feltman, A. Treuille, M. Cohen, Real-time drawing assistance through crowdsourcing, ACM Trans. Graph. 32 (4) (2013) 1–8. https://doi.org/10.1145/2461912.2462016.
[5] Y.J. Lee, C.L. Zitnick, M.F. Cohen, Shadowdraw: Real-time user guidance for freehand drawing, ACM SIGGRAPH 2011 Papers, ACM, New York, NY, USA, 2011, pp. 1–10. https://doi.org/10.1145/1964921.1964922.
[6] B. Jiang, S. Liu, Expressive image sketching with two-layer image features, Proceedings of the 2013 Seventh International Conference on Image and Graphics, IEEE Computer Society, Washington, DC, USA, 2013, pp. 927–931. https://doi.org/10.1109/ICIG.2013.187.
[7] S. Wang, T. Kubota, J.M. Siskind, J. Wang, Salient closed boundary extraction with ratio contour, IEEE Trans. Pattern Anal. Mach. Intell. 27 (4) (2005) 546–561. https://doi.org/10.1109/TPAMI.2005.84.
[8] G. Song, Q. Zhu, J. Shi, Untangling cycles for contour grouping, 2007 11th IEEE International Conference on Computer Vision(ICCV), 2007. pp. 1–8. https://doi.org/10.1109/ICCV.2007.4408929.
[9] V. Ferrari, L. Fevrier, F. Jurie, C. Schmid, Groups of adjacent contour segments for object detection, IEEE Trans. Pattern Anal. Mach. Intell. 30 (1) (2008) 36–51. http://ieeexplore.ieee.org/document/4359307/. https://doi.org/10.1109/TPAMI.2007.1144.
[10] R. Kennedy, J. Gallier, Contour cut: identifying salient contours in images by solving a hermitian eigenvalue problem, IEEE Conference on Computer Vision and Pattern Recognition, 2011. pp. 2065–2072. http://ieeexplore.ieee.org/abstract/document/5995739/. https://doi.org/10.1109/CVPR.2011.5995739.
[11] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 898–916. https://doi.org/10.1109/TPAMI.2010.161.
[12] S. Marvaniya, S. Bhattacharjee, V. Manickavasagam, A. Mittal, Drawing an automatic sketch of deformable objects using only a few images, Computer Vision-ECCV, Springer Berlin Heidelberg. 2012, pp. 63–72. https://doi.org/10.1007/978-3-642-33863-2_7.
[13] J.J. Lim, C.L. Zitnick, P. Dollar, Sketch tokens: a learned mid-level representation for contour and object detection, Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2013, pp. 3158–3165. https://doi.org/10.1109/CVPR.2013.406.
[14] H. Wang, Skeleton-based morphing between sketching images, 2010 3rd International Congress on Image and Signal Processing, 6, 2010. pp. 2665–2668. http://ieeexplore.ieee.org/abstract/document/5648112/. https://doi.org/10.1109/CISP.2010.5648112.
[15] M. Zhao, X. Yang, M. Zhao, X. Yang, Sketch-style face image synthesis via constrained least squares estimation, International Conference on Computational Problem-Solving, 2011. pp. 1–4. http://ieeexplore.ieee.org/abstract/document/6092303/. https://doi.org/10.1109/ICCPS.2011.6092303.
[16] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 530–549. https://doi.org/10.1109/TPAMI.2004.1273918.
[17] C.-e. Guo, S.-Z. Zhu, Y.N. Wu, Primal sketch: Integrating structure and texture, Comput. Vis. Image Underst. 106 (1) (2007) 5–19. https://doi.org/10.1016/j.cviu.2005.09.004.
[18] Y. Qi, J. Guo, Y. Li, H. Zhang, T. Xiang, Y.Z. Song, Sketching by perceptual grouping, IEEE International Conference on Image Processing, 2014. pp. 270–274. http://ieeexplore.ieee.org/abstract/document/6738056/. https://doi.org/10.1109/ICIP.2013.6738056.
[19] S. Xie, Z. Tu, Holistically-nested edge detection, Int. J. Comput. Vis. 125 (1) (2017) 3–18. https://doi.org/10.1007/s11263-017-1004-z.
[20] Y. Liu, M.M. Cheng, X. Hu, K. Wang, X. Bai, Richer convolutional features for edge detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. pp. 5872–5881. http://ieeexplore.ieee.org/abstract/document/8100105/. https://doi.org/10.1109/CVPR.2017.622.
[21] X. Zhang, X. Li, X. Li, M. Shen, Better freehand sketch synthesis for sketch-based image retrieval: beyond image edges, Neurocomputing 322 (1) (2018) 38–46. http://www.sciencedirect.com/science/article/pii/S0925231218311202.
[22] R.G. Schneider, T. Tuytelaars, Example-based sketch segmentation and labeling using CRFs, ACM Trans. Graph. 35 (5) (2016) 1511–1519. https://doi.org/10.1145/2898351.
[23] I. Berger, A. Shamir, M. Mahler, E. Carter, J. Hodgins, Style and abstraction in portrait sketching, ACM Trans. Graph. 32 (4) (2013) 1–12. https://doi.org/10.1145/2461912.2461964.
[24] R.G. Schneider, T. Tuytelaars, Sketch classification and classification-driven analysis using fisher vectors, ACM Trans. Graph. 33 (6) (2014) 1–9. https://doi.org/10.1145/2661229.2661231.
[25] Z. Huang, H. Fu, R.W.H. Lau, Data-driven segmentation and labeling of freehand sketches, ACM Trans. Graph. 33 (6) (2014) 1–10. https://doi.org/10.1145/2661229.2661280.
[26] Y. Li, Y.-Z. Song, T.M. Hospedales, S. Gong, Free-hand sketch synthesis with deformable stroke models, Int. J. Comput. Vis. 122 (1) (2017) 169–190. https://doi.org/10.1007/s11263-016-0963-9.
[27] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, S. Khudanpur, Recurrent neural network based language model, Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010, 2, 2010. pp. 1045–1048.
[28] D. Ha, D. Eck, A neural representation of sketch drawings, (2017) CoRR abs/1704.03477 arXiv:1704.03477. http://arxiv.org/abs/1704.03477.
[29] U.R. Muhammad, Y. Yang, Y. Song, T. Xiang, T.M. Hospedales, Learning deep sketch abstraction, (2018) CoRR abs/1804.04804. arXiv: 1804.04804. http://arxiv.org/abs/1804.04804.
[30] J. Song, K. Pang, Y. Song, T. Xiang, T.M. Hospedales, Learning to sketch with shortcut cycle consistency, (2018) CoRR abs/1805.00247. arXiv: 1805.00247. http://arxiv.org/abs/1805.00247.
[31] K. Li, K. Pang, J. Song, Y.-Z. Song, T. Xiang, T.M. Hospedales, H. Zhang, Universal sketch perceptual grouping, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision-ECCV2018, Springer International Publishing, Cham, 2018, pp. 593–609.
[32] V. Ferrari, T. Tuytelaars, L.V. Gool, Object detection by contour segment networks, European Conference on Computer Vision, 2006. pp. 14–28. https://doi.org/10.1007/11744078_2.
[33] M. Everingham, L. Gool, C.K. Williams, J. Winn, A. Zisserman, The Pascal visual object classes (VOC) challenge, Int. J. Comput. Vision 88 (2) (2010) 303–338. https://doi.org/10.1007/s11263-009-0275-4.
[34] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, IEEE International Conference on Computer Vision, 2, 2001. pp. 416–423. http://ieeexplore.ieee.org/abstract/document/937655/. https://doi.org/10.1109/ICCV.2001.937655.
[35] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, Proceedings of the 22nd ACM International Conference on Multimedia, ACM, New York, NY, USA, 2014, pp. 675–678. https://doi.org/10.1145/2647868.2654889.
[36] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527.

[37] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, International Conference on Neural Information Processing Systems, MIT Press, Cambridge, MA, USA, 2014, pp. 2672–2680. http://dl.acm.org/citation.cfm?id=2969033.2969125.

[38] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan, in: D. Cremers, I. Reid, H. Saito, M. Yang (Eds.), Nips, 9006, Springer International Publishing, Cham, 2015, pp. 1–14. https://doi.org/10.1007/978-3-319-16817-3.

[39] A. Brock, T. Lim, J.M. Ritchie, N. Weston, Neural photo editing with introspective adversarial networks, (2016) CoRR abs/1609.07093. arXiv:1609.07093. http://arxiv.org/abs/1609.07093.

[40] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, A.A. Efros, Generative visual manipulation on the natural image manifold, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision-ECCV, Springer International Publishing, Cham, 2016, pp. 597–613. https://doi.org/10.1007/978-3-319-46454-1_36.

[41] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, (2017) arXiv e-prints abs/1701.07875. arXiv:1701.07875. http://adsabs.harvard.edu/abs/2017arXiv170107875A.

[42] M. Palatucci, D. Pomerleau, G. Hinton, T.M. Mitchell, Zero-shot learning with semantic output codes, International Conference on Neural Information Processing Systems, Curran Associates Inc., USA, 2009, pp. 1410–1418. http://dl.acm.org/citation.cfm?id=2984093.2984252.

[43] L. Fei-Fei, R. Fergus, P. Perona, One-shot learning of object categories, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 594–611. http://ieeexplore.ieee.org/document/1597116/. https://doi.org/10.1109/TPAMI.2006.79.

[44] J. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, The IEEE International Conference on Computer Vision (ICCV), 2017. pp. 2242–2251. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8237506. https://doi.org/10.1109/ICCV.2017.244.

[45] S. Benaim, L. Wolf, One-sided unsupervised domain mapping, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. 2017, pp. 752–762. http://papers.nips.cc/paper/6677-one-sided-unsupervised-domain-mapping.pdf.

[46] Z. Yi, H. Zhang, P. Tan, M. Gong, Dualgan: Unsupervised dual learning for image-to-image translation, The IEEE International Conference on Computer Vision (ICCV), 2017. pp. 2849–2857. http://openaccess.thecvf.com/content_iccv_2017/html/Yi_DualGAN_Unsupervised_Dual_ICCV_2017_paper.html.

[47] I.E. Sutherland, Sketch pad a man-machine graphical communication system, Proceedings of the SHARE Design Automation Workshop, ACM, New York, NY, USA, 1964, pp. 329–346. https://doi.org/10.1145/800265.810742.

[48] C.F. Herot, Graphical input through machine recognition of sketches, Proceedings of the 3rd Annual Conference on Computer Graphics and Interactive Techniques, ACM, New York, NY, USA, 1976, pp. 97–102. https://doi.org/10.1145/563274.563294.

[49] T.M. Sezgin, T. Stahovich, R. Davis, Sketch based interfaces: early processing for sketch understanding, Proceedings of the 2001 Workshop on Perceptive User Interfaces, ACM, New York, NY, USA, 2001, pp. 1–8. https://doi.org/10.1145/971478.971487.

[50] B. Paulson, T. Hammond, Paleosketch: accurate primitive sketch recognition and beautification, Proceedings of the 13th International Conference on Intelligent User Interfaces, ACM, New York, NY, USA, 2008, pp. 1–10. https://doi.org/10.1145/1378773.1378775.

[51] H. Chatbri, K. Kameyama, Towards making thinning algorithms robust against noise in sketch images, International Conference on Pattern Recognition (ICPR), 2012. pp. 3030–3033. http://ieeexplore.ieee.org/abstract/document/6460803/.

[52] T. Hammond, R. Davis, Ladder, a sketching language for user interface developers, Comput. Graph. 29 (4) (2005) 518–532. http://www.sciencedirect.com/science/article/pii/S0097849305000865. https://doi.org/10.1016/j.cag.2005.05.005.

[53] J.J. LaViola, Jr, R.C. Zeleznik, Mathpad2: A system for the creation and exploration of mathematical sketches, ACM SIGGRAPH 2004, ACM, New York, NY, USA, 2004, pp. 432–440. https://doi.org/10.1145/1186562.1015741.

[54] T.M. Sezgin, R. Davis, HMM-based efficient sketch recognition, Proceedings of the 10th International Conference on Intelligent User Interfaces, ACM, New York, NY, USA, 2005, pp. 281–283. https://doi.org/10.1145/1040830.1040899.

[55] T.Y. Ouyang, R. Davis, Chemink: A natural real-time recognition system for chemical drawings, Proceedings of the 16th International Conference on Intelligent User Interfaces, ACM, New York, NY, USA, 2011, pp. 267–276. https://doi.org/10.1145/1943403.1943444.

[56] R.G. Schneider, T. Tuytelaars, Sketch classification and classification-driven analysis using fisher vectors, ACM Trans. Graph. 33 (6) (2014) 2691–2698. http://dl.acm.org/citation.cfm?doid=2661229.2661231. https://doi.org/10.1145/2661229.2661231.

[57] Y. Li, Y.Z. Song, S. Gong, Sketch recognition by ensemble matching of structured features, British Machine Vision Conference, 2013. pp. 1–11. https://www.eecs.qmul.ac.uk/sgg/papers/LiEtAl_BMVC2013.pdf. https://doi.org/10.5244/C.27.35.

[58] Y.-Z.S. Yi Li, Timothy M. Hospedales, S. Gong, Free-hand sketch recognition by multi-kernel feature learning, Comput. Vis. Image Underst. 137 (2015) 1–11. http://www.sciencedirect.com/science/article/pii/S1077314215000375. https://doi.org/10.1016/j.cviu.2015.02.003.

[59] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2005, pp. 886–893. https://doi.org/10.1109/CVPR.2005.177.

[60] Z. Sun, C. Wang, L. Zhang, L. Zhang, Query-adaptive shape topic mining for hand-drawn sketch recognition, ACM, New York, NY, USA, 2012, pp. 519–528. https://doi.org/10.1145/2393347.2393421.

[61] Z. Sun, C. Wang, L. Zhang, L. Zhang, Free hand-drawn sketch segmentation, Proceedings of the 12th European Conference on Computer Vision, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 626–639. https://doi.org/10.1007/978-3-642-33718-5_45.

[62] E. Shechtman, M. Irani, Matching local self-similarities across images and videos, IEEE Conference on Computer Vision and Pattern Recognition, IEEE. 2007, pp. 1–8. http://ieeexplore.ieee.org/document/4270223/. https://doi.org/10.1109/CVPR.2007.383198.

[63] E. Tola, V. Lepetit, P. Fua, Daisy: an efficient dense descriptor applied to wide-baseline stereo, IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2010) 815–830. http://ieeexplore.ieee.org/document/4815264/. https://doi.org/10.1109/TPAMI.2009.77.

[64] A. Vedaldi, B. Fulkerson, Vlfeat: an open and portable library of computer vision algorithms, Proceedings of the 18th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2010, pp. 1469–1472. https://doi.org/10.1145/1873951.1874249.

[65] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27. https://doi.org/10.1145/1961189.1961199.

[66] M. Jaderberg, K. Simonyan, A. Zisserman, K. Kavukcuoglu, Spatial transformer networks, Proceedings of the 28th International Conference on Neural Information Processing Systems, MIT Press, Cambridge, MA, USA, 2015, pp. 2017–2025. http://dl.acm.org/citation.cfm?id=2969442.2969465.

[67] S. Wu, S. Zhong, Y. Liu, Deep residual learning for image steganalysis, Multimedia Tools and Applications, 2017. pp. 1–17. https://doi.org/10.1007/s11042-017-4440-4.

[68] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, S.Z. Li, Embedding deep metric for person re-identification: a study against large variations, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision-ECCV 2016, Springer International Publishing, Cham, 2016, pp. 732–748. http://arxiv.org/abs/1611.00137.

[69] Q. Yu, Y. Yang, Y.-Z. Song, T. Xiang, Timothy Hospedales, Deep neural networks for sketch recognition, Proceedings of the British Machine Vision Conference, British Machine Vision Association. 2015, pp. 1–12. http://www.bmva.org/bmvc/2015/papers/paper007/index.html. https://doi.org/10.5244/C.29.7.

[70] R.K. Sarvadevabhatla, R.V. Babu, Freehand sketch recognition using deep features, (2015) CoRR abs/1502.00254. arXiv:1502.00254. http://arxiv.org/abs/1502.00254.

[71] Q. Yu, Y. Yang, Y.-Z. Song, T. Xiang, T. Hospedales, Sketch-a-net that beats humans, in: X. Xie, M.W. Jones, G.K.L. Tam (Eds.), Proceedings of the British Machine Vision Conference (BMVC), BMVA Press. 2015, pp. 71–712. https://www.researchgate.net/publication/280305180_Sketch-a-Net_that_Beats_Humans. https://doi.org/10.5244/C.29.7.

[72] Q. Yu, Y. Yang, F. Liu, Y.-Z. Song, T. Xiang, T.M. Hospedales, Sketch-a-net: a deep neural network that beats humans, Int. J. Comput. Vis. 122 (3) (2017) 411–425. https://doi.org/10.1007/s11263-016-0932-3.

[73] H. Zhang, S. Liu, C. Zhang, W. Ren, R. Wang, X. Cao, Sketchnet: sketch classification with web images, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 1105–1113. http://ieeexplore.ieee.org/abstract/document/7780494/. https://doi.org/10.1109/CVPR.2016.125.

[74] K. Sasaki, M. Yamakawa, K. Sekiguchi, T. Ogata, Classification of photo and sketch images using convolutional neural networks, in: A.E.P. Villa, P. Masulli, A.J. Pons Rivero (Eds.), Artificial Neural Networks and Machine Learning - ICANN2016, Springer International Publishing, Cham, 2016, pp. 283–290. https://doi.org/10.1007/978-3-319-44781-0_34.

[75] A. Vedaldi, K. Lenc, Matconvnet: c neural networks for matlab, Proceedings of the 23rd ACM International Conference on Multimedia, ACM, New York, NY, USA, 2015, pp. 689–692. https://doi.org/10.1145/2733373.2807412.

[76] C. Hu, D. Li, Y.-Z. Song, T. Xiang, T.M. Hospedales, Sketch-a-classifier: sketch-based photo classifier generation, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. pp. 9136–9144.

[77] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Proceedings of the 25th International Conference on Neural Information Processing Systems, Curran Associates Inc., USA, 2012, pp. 1097–1105. http://dl.acm.org/citation.cfm?id=2999134.2999257.

[78] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324. http://ieeexplore.ieee.org/abstract/document/726791/. https://doi.org/10.1109/5.726791.

[79] T. Kato, T. Kurita, N. Otsu, K. Hirata, A sketch retrieval method for full color image database-query by visual example, Proceedings. 11th IAPR International Conference on Pattern Recognition, 1992. pp. 530–533. http://ieeexplore.ieee.org/document/201616/. https://doi.org/10.1109/ICPR.1992.201616.

[80] G. Qiu, Indexing chromatic and achromatic patterns for content-based colour image retrieval, Pattern Recogn. 35 (8) (2002) 1675–1686. https://doi.org/10.1016/S0031-3203(01)00162-5.

[81] J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, D. Petkovic, The query by image content (QBIC) system, Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, 1995, pp. 475. https://doi.org/10.1145/223784.223888.

[82] J.R. Smith, S.-F. Chang, Visualseek: a fully automated content-based image query system, Proceedings of the Fourth ACM International Conference on Multimedia, ACM, New York, NY, USA, 1996, pp. 87–98. https://doi.org/10.1145/244130.244151.

[83] M.N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance, IEEE Trans. Image Process. 11 (2) (2002) 146–158. http://ieeexplore.ieee.org/abstract/document/982822/. https://doi.org/10.1109/83.982822.

[84] R.K. Rajendran, S.-F. Chang, Image retrieval with sketches and compositions, 2 (2000) 717–720. http://ieeexplore.ieee.org/abstract/document/871462/. https://doi.org/10.1109/ICME.2000.871462.

[85] P. Fränti, A. Mednonogov, V. Kyrki, H. Kälviäinen, Content-based matching of line-drawing images using the hough transform, Int. J. Doc. Anal. Recognit. 3 (2) (2000) 117–124. https://doi.org/10.1007/s100320070001.

[86] J.L. Shih, L.H. Chen, A new system for trademark segmentation and retrieval, Image Vis. Comput. 19 (13) (2001) 1011–1018. https://doi.org/10.1016/S0262-8856(01)00063-4.

[87] W.H. Leung, T. Chen, Trademark retrieval using contour-skeleton stroke classification, Proceedings. IEEE International Conference on Multimedia and Expo, 2, 2002. pp. 517–520. http://ieeexplore.ieee.org/document/1035662/. https://doi.org/10.1109/ICME.2002.1035662.

[88] M.J. Fonseca, A. Ferreira, J.a. Jorge, Content-based retrieval of technical drawings, Int. J. Comput. Appl. Technol. 23 (2/3/4) (2005) 86–118. http://www.inderscience.com/link.php?id=6467. https://doi.org/10.1504/IJCAT.2005.006467.

[89] S. Liang, Z. Sun, Sketch retrieval and relevance feedback with biased SVM classification, Pattern Recogn. Lett. 29 (12) (2008) 1733–1741. https://www.sciencedirect.com/science/article/pii/S0167865508001621. https://doi.org/10.1016/j.patrec.2008.05.004.

[90] P. Sousa, M.J. Fonseca, Sketch-based retrieval of drawings using spatial proximity, J. Vis. Lang. Comput. 21 (2) (2010) 69–80. https://www.sciencedirect.com/science/article/pii/S1045926X09000779. https://doi.org/10.1016/j.jvlc.2009.12.001.

[91] C. Wang, J. Zhang, B. Yang, L. Zhang, Sketch2cartoon: composing cartoon images by sketching, Proceedings of the 19th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2011, pp. 789–790. https://doi.org/10.1145/2072298.2072458.

[92] M.J. Fonseca, D. Gonçalves, Sketch-a-doc: using sketches to find documents, Proceedings of the 16th International Conference on Distributed Multimedia Systems, 2010. pp. 327–330. http://web.ist.utl.pt/mjf/publications/2014-2010/pdf/vlc10.pdf.

[93] S. Liang, Z. Sun, B. Li, Sketch retrieval based on spatial relations, International Conference on Computer Graphics, Imaging and Visualization, 2005. pp. 24–29. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1521034. https://doi.org/10.1109/CGIV.2005.67.

[94] W.H. Leung, T. Chen, Retrieval of sketches based on spatial relation between strokes, Proceedings. International Conference on Image Processing, 1, 2002. pp. 908–911. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1038173. https://doi.org/10.1109/ICIP.2002.1038173.

[95] M.J. Fonseca, A. Ferreira, J.A. Jorge, Sketch-based retrieval of complex drawings using hierarchical topology and geometry, Comput. Aided Des. 41 (12) (2009) 1067–1081. http://www.sciencedirect.com/science/article/pii/S0010448509002358. https://doi.org/10.1016/j.cad.2009.09.004.

[96] A. Del Bimbo, P. Pala, Visual image retrieval by elastic matching of user sketches, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 121–132. http://ieeexplore.ieee.org/abstract/document/574790/. https://doi.org/10.1109/34.574790.

[97] E. Di Sciascio, G. Mingolla, M. Mongiello, Content-based image retrieval over the web using query by sketch and relevance feedback, in: D.P. Huijsmans, A.W.M. Smeulders (Eds.), Visual Information and Information Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999, pp. 123–130. http://link.springer.com/10.1007/3-540-48762-X_16.

[98] F. Mokhtarian, A.K. Mackworth, A theory of multiscale, curvature-based shape representation for planar curves, IEEE Trans. Pattern Anal. Mach. Intell. 14 (8) (1992) 789–805. http://ieeexplore.ieee.org/document/149591/. https://doi.org/10.1109/34.149591.

[99] D.P. Lopresti, Feature-based approach for image retrieval by sketch, Proc. Spie 3229 (1997) 1–12. https://doi.org/10.1117/12.290343.

[100] S. Matusiak, M. Daoudi, T. Blu, O. Avaro, Sketch-based images database retrieval, Advances in Multimedia Information Systems, Springer Berlin Heidelberg. 1998, pp. 185–191. https://doi.org/10.1007/3-540-49651-3_18.

[101] H.H.S. Ip, A.K.Y. Cheng, W.Y.F. Wong, J. Feng, Affine-invariant sketch-based retrieval of images, Proceedings of the International Conference on Computer Graphics, IEEE Computer Society, Washington, DC, USA, 2001, pp. 55–61. http://dl.acm.org/citation.cfm?id=872749.873337.

[102] A.K. Jain, A. Vailaya, Image retrieval using color and shape, Pattern Recogn. 29 (8) (1996) 1233–1244. http://www.sciencedirect.com/science/article/pii/0031320395001603. https://doi.org/10.1016/0031-3203(95)00160-3.

[103] A. Chalechale, G. Naghdy, A. Mertins, Sketch-based image matching using angular partitioning, Trans. Sys. Man Cyber. Part A 35 (1) (2005) 28–41. https://doi.org/10.1109/TSMCA.2004.838464.

[104] L. Tao, L. Yuan, J. Sun, Skyfinder: attribute-based sky image search, ACM SIGGRAPH 2009, ACM, New York, NY, USA, 2009, pp. 1–5. https://doi.org/10.1145/1576246.1531374.

[105] A.M.M. Bada, G. e Jesús Hoyos Rivera, A.M. Hernández, Garabato: a proposal of a sketch-based image retrieval system for the web, 2014 International Conference on Electronics, Communications and Computers (CONIELECOMP), 2014. pp. 183–188. http://ieeexplore.ieee.org/abstract/document/6808588/. https://doi.org/10.1109/CONIELECOMP.2014.6808588.

[106] D. Pan, P. Shi, C. Li, Sketch-based image retrieval by using saliency, International Conference on Fuzzy Systems and Knowledge Discovery, 2014. pp. 825–829. http://ieeexplore.ieee.org/abstract/document/6980944/. https://doi.org/10.1109/FSKD.2014.6980944.

[107] C. Ma, X. Yang, C. Zhang, X. Ruan, M.-H. Yang, Sketch retrieval via local dense stroke features, Image Vis. Comput. 46 (2016) 64–73. http://www.sciencedirect.com/science/article/pii/S0262885615001389. https://doi.org/10.1016/j.imavis.2015.11.007.

[108] Y. Matsui, K. Aizawa, Y. Jing, Sketch2manga: sketch-based manga retrieval, 2014 IEEE International Conference on Image Processing (ICIP), 2014. pp. 3097–3101. http://ieeexplore.ieee.org/abstract/document/7025626/. https://doi.org/10.1109/ICIP.2014.7025626.

[109] Y. Zhang, X. Qian, X. Tan, Sketch-based image retrieval using contour segments, IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), 2015. pp. 1–6. http://ieeexplore.ieee.org/abstract/document/7340836/. https://doi.org/10.1109/MMSP.2015.7340836.

[110] C.V. Patil, R.K. Kumar, R. Aarthi, Sketch based image retrieval — a short survey, Int. J. Eng. 3 (2) (2014) 2475–2478.

[111] M. Indu, K.V. Kavitha, Survey on sketch based image retrieval methods, 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2016. pp. 1–4. http://ieeexplore.ieee.org/document/7530359/. https://doi.org/10.1109/ICCPCT.2016.7530359.

[112] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, Sketch-based image retrieval: benchmark and bag-of-features descriptors, IEEE Trans. Vis. Comput. Graph. 17 (11) (2011) 1624–1636. https://doi.org/10.1109/TVCG.2010.266.

[113] R. Hu, M. Barnard, J. Collomosse, Gradient field descriptor for sketch based retrieval and localization, IEEE International Conference on Image Processing, 2010. pp. 1025–1028. http://personal.ee.surrey.ac.uk/Personal/R.Hu/papers/Rui_ICIP.pdf. https://doi.org/10.1109/ICIP.2010.5649331.

[114] R. Hu, T. Wang, J. Collomosse, A bag-of-regions approach to sketch-based image retrieval, 18th IEEE International Conference on Image Processing, 2011. pp. 3661–3664. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6116513. https://doi.org/10.1109/ICIP.2011.6116513.

[115] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[116] G. Mori, S. Belongie, J. Malik, Efficient shape matching using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 27 (11) (2005) 1832–1837. https://doi.org/10.1109/TPAMI.2005.220.

[117] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, Technical section: an evaluation of descriptors for large-scale image retrieval from sketched feature lines, Comput. Graph. 34 (5) (2010) 482–498. https://doi.org/10.1016/j.cag.2010.07.002.

[118] J.M. Saavedra, B. Bustos, An improved histogram of edge local orientations for sketch-based image retrieval, in: M. Goesele, S. Roth, A. Kuijper, B. Schiele, K. Schindler (Eds.), Pattern Recognition, Springer Berlin Heidelberg. 2010, pp. 432–441. http://link.springer.com/10.1007/978-3-642-15986-2_44.

[119] J.M. Saavedra, Sketch based image retrieval using a soft computation of the histogram of edge local orientations (S-HELO), IEEE International Conference on Image Processing (ICIP), 2014. pp. 2998–3002. http://ieeexplore.ieee.org/abstract/document/7025606/. https://doi.org/10.1109/ICIP.2014.7025606.

[120] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, L. Zhang, Mindfinder: interactive sketch-based image search on millions of images, Proceedings of the 18th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2010, pp. 1605–1608. https://doi.org/10.1145/1873951.1874299.

[121] R. Aarthi, J. Amudha, Saliency based modified Chamfers matching method for sketch based image retrieval, International Conference on Innovations in Information, Embedded and Communication Systems, 2015. pp. 1–4. http://ieeexplore.ieee.org/abstract/document/7192911/. https://doi.org/10.1109/ICIIECS.2015.7192911.

[122] J.M. Saavedra, B. Bustos, Sketch-based image retrieval using keyshapes, Multimed. Tools Appl. 73 (3) (2014) 2033–2062. https://doi.org/10.1007/s11042-013-1689-0.

[123] T. Bui, J. Collomosse, Scalable sketch-based image retrieval using color gradient features, IEEE International Conference on Computer Vision Workshop (ICCVW), IEEE Computer Society. 2015, pp. 1012–1019. https://doi.org/10.1109/ICCVW.2015.133.

[124] P. Salembier, T. Sikora, Introduction to MPEG-7: multimedia content description interface, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[125] Y. Kiumagai, G. Ohashi, Query-by-sketch image retrieval using edge relation histogram, IEICE Trans. Inf. Syst. 96 (2) (2013) 340–348. http://www.mva-org.jp/Proceedings/2011CD/papers/04-12.pdf. https://doi.org/10.1587/transinf.E96.D.340.

[126] B. Szántó, P. Pozsegovics, Z. Vámossy, S. Sergyán, Sketch4match: content-based image retrieval system using sketches, International Symposium on Applied Machine Intelligence and Informatics (SAMI), 2011. pp. 183–188. http://ieeexplore.ieee.org/document/5738872/. https://doi.org/10.1109/SAMI.2011.5738872.

[127] H. Chatbri, K. Kameyama, Sketch-based image retrieval by shape points description in support regions, International Conference on Systems, Signals and Image Processing, 2013. pp. 19–22. http://adapt.cs.tsukuba.ac.jp/chatbri/web/pdf/IWSSIP2013.pdf. https://doi.org/10.1109/IWSSIP.2013.6623439.

[128] H. Chatbri, K. Kameyama, P. Kwan, Sketch-based image retrieval by size-adaptive and noise-robust feature description, International Conference on Digital Image Computing: Techniques and Applications, 2013. pp. 1–8. http://adapt.cs.tsukuba.ac.jp/chatbri/web/pdf/DICTA2013.pdf. https://doi.org/10.1109/DICTA.2013.6691528.

[129] X. Cao, H. Zhang, S. Liu, X. Guo, L. Lin, SYM-FISH: a symmetry-aware flip invariant sketch histogram shape descriptor, Proceedings of the 2013 IEEE International Conference on Computer Vision, IEEE Computer Society, Washington, DC, USA, 2013, pp. 313–320. https://doi.org/10.1109/ICCV.2013.46.
[130] S. Wang, J. Zhang, T. Xu, Z. Miao, Sketch-based image retrieval through hypothesis-driven object boundary selection with HLR descriptor, IEEE Trans. Multimedia 17 (7) (2015) 1045–1057. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7104149. https://doi.org/10.1109/TMM.2015.2431492.
[131] X. Zhang, X. Chen, Robust sketch-based image retrieval by saliency detection, in: Q. Tian, N. Sebe, G.-J. Qi, B. Huet, R. Hong, X. Liu (Eds.), Multimedia Modeling, Springer International Publishing, Cham, 2016, pp. 515–526. http://link.springer.com/10.1007/978-3-319-27671-7.
[132] T. Furuya, R. Ohbuchi, Visual saliency weighting and cross-domain manifold ranking for sketch-based image retrieval, in: C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, N. O'Connor (Eds.), Multimedia Modeling, Springer International Publishing, Cham, 2014, pp. 37–49. https://link.springer.com/chapter/10.1007/978-3-319-04114-8_4.
[133] J. Niu, J. Ma, J. Lu, X. Liu, Z. Zhu, M-SBIR: an improved sketch-based image retrieval method using visual word mapping, in: L. Amsaleg, G.Þ. Guðmundsson, C. Gurrin, B.Þ. Jónsson, S. Satoh (Eds.), Multimedia Modeling, Springer International Publishing, Cham, 2017, pp. 257–268. http://link.springer.com/10.1007/978-3-319-51814-5.
[134] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, S.-M. Hu, Sketch2Photo: internet image montage, ACM Trans. Graph. 28 (5) (2009) 1241–12410. https://doi.org/10.1145/1618452.1618470.
[135] C. Wang, Z. Li, L. Zhang, Mindfinder: image search by interactive sketching and tagging, Proceedings of the 19th International Conference on World Wide Web, 2010. pp. 1309–1312. https://doi.org/10.1145/1772690.1772909.
[136] Y. Cao, C. Wang, L. Zhang, L. Zhang, Edgel index for large-scale sketch-based image search, Computer Vision and Pattern Recognition, 2011. pp. 761–768. http://www.cvpapers.com/papers/0630. https://doi.org/10.1109/CVPR.2011.5995460.
[137] M. Eitz, R. Richter, K. Hildebrand, T. Boubekeur, M. Alexa, Photosketcher: interactive sketch-based image synthesis, IEEE Comput. Graph. Appl. 31 (6) (2011) 56–66. https://doi.org/10.1109/MCG.2011.67.
[138] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, Photosketch: a sketch based image query and compositing system, SIGGRAPH 2009:Talks, ACM, New York, NY, USA, 2009, 601–601. https://doi.org/10.1145/1597990.1598050.
[139] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, A descriptor for large scale image retrieval based on sketched feature lines, Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling, ACM, New York, NY, USA, 2009, pp. 29–36. https://doi.org/10.1145/1572741.1572747.
[140] R. Zhou, L. Chen, L. Zhang, Sketch-based image retrieval on a large scale database, Proceedings of the 20th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2012, pp. 973–976. https://doi.org/10.1145/2393347.2396360.
[141] X. Sun, C. Wang, C. Xu, L. Zhang, Indexing billions of images for sketch-based retrieval, Proceedings of the 21st ACM International Conference on Multimedia, ACM, New York, NY, USA, 2013, pp. 233–242. https://doi.org/10.1145/2502081.2502281.
[142] K. Bozas, E. Izquierdo, Large scale sketch based image retrieval using patch hashing, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M. Papka (Eds.), International Symposium on Visual Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 210–219. https://doi.org/10.1007/978-3-642-33179-4_21.
[143] S. Parui, A. Mittal, Similarity-invariant sketch-based image retrieval in large databases, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), European Conference on Computer Vision, Springer International Publishing, Cham, 2014, pp. 398–414. https://doi.org/10.1007/978-3-319-10599-4_26.
[144] S. Liang, L. Zhao, Y. Wei, J. Jia, Sketch-based Retrieval Using Content-aware Hashing, Springer International Publishing, Cham, 2014 pp. 133–142, 10.1007/978-3-319-13168-9_14.
[145] Y.H. Sharath Kumar, N. Pavithra, KD-tree approach in sketch based image retrieval, in: R. Prasath, A.K. Vuppala, T. Kathirvalavakumar (Eds.), Mining Intelligence and Knowledge Exploration, Springer International Publishing, Cham, 2015, pp. 247–258.
[146] G. Salton, C. Buckley, Improving retrieval performance by relevance feedback, J. Am. Soc. Inf. Sci. 41 (4) (1990) 288–297. https://ci.nii.ac.jp/naid/10010356145/en/. https://doi.org/10.1002/(SICI)1097-4571(199006)41:4<288::AID-ASI8>3.0.CO;2-H.
[147] X.S. Zhou, T.S. Huang, Relevance feedback in image retrieval: a comprehensive review, Multimedia Systems 8 (6) (2003) 536–544. https://doi.org/10.1007/s00530-002-0070-3.
[148] Y. Rui, T.S. Huang, S. Mehrotra, Content-based image retrieval with relevance feedback in MARS, Proceedings of International Conference on Image Processing, 2, 1997. pp. 815–818. http://ieeexplore.ieee.org/document/638621/. https://doi.org/10.1109/ICIP.1997.638621.
[149] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback: a power tool for interactive content-based image retrieval, IEEE Trans. Circuits Syst. Video Technol. 8 (5) (1998) 644–655. http://ieeexplore.ieee.org/document/718510/. https://doi.org/10.1109/76.718510.
[150] T.S. Chua, W.C. Low, C.X. Chu, Relevance feedback techniques for color-based image retrieval, Proceedings of the 1998 Conference on Multimedia Modeling, IEEE Computer Society, Washington, DC, USA, 1998, pp. 24–31. http://dl.acm.org/citation.cfm?id=823464.824330. https://doi.org/10.1109/MULMM.1998.722971.
[151] E. Cheng, F. Jing, L. Zhang, A unified relevance feedback framework for web image retrieval, Trans. Img. Proc. 18 (6) (2009) 1350–1357. https://doi.org/10.1109/TIP.2009.2017128.
[152] W. Meng, L. Hao, T. Dacheng, L. Ke, W. Xindong, Multimodal graph-based reranking for web image search., IEEE Trans. Image Process. 21 (11) (2012) 4649–4661. https://doi.org/10.1109/TIP.2012.2207397.
[153] X. Yang, T. Mei, Y. Zhang, J. Liu, S. Satoh, Web image search re-ranking with click-based similarity and typicality, Trans. Img. Proc. 25 (10) (2016) 4617–4630. https://doi.org/10.1109/TIP.2016.2593653.
[154] B. Li, Z. Sun, S. Liang, Y. Zhang, B. Yuan, Relevance Feedback for Sketch retrieval based on linear programming classification, Proceedings of the 7th Pacific Rim Conference on Advances in Multimedia Information Processing, Springer-Verlag, Berlin, Heidelberg, 2006, pp. 201–210. https://doi.org/10.1007/11922162_24.
[155] H.W.H. Leung, Representations, Feature Extraction, Matching and Relevance Feedback for Sketch Retrieval (Ph.D. thesis), Pittsburgh, PA, USA, 2003, aAI3099361.
[156] X. Qian, X. Tan, Y. Zhang, R. Hong, M. Wang, Enhancing sketch-based image retrieval by re-ranking and relevance feedback, IEEE Trans. Image Process. 25 (1) (2016) 195–208. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7317816. https://doi.org/10.1109/TIP.2015.2497145.
[157] X. Wang, X. Duan, X. Bai, Deep sketch feature for cross-domain image retrieval, Neurocomputing 207 (2016) 387–397. http://www.sciencedirect.com/science/article/pii/S0925231216303198. https://doi.org/10.1016/j.neucom.2016.04.046.
[158] Y. Qi, Y.Z. Song, H. Zhang, J. Liu, Sketch-based image retrieval via siamese convolutional neural network, IEEE International Conference on Image Processing (ICIP), 2016. pp. 2460–2464. http://www.eecs.qmul.ac.uk/yzs/qi2016sbirdeep.pdf. https://doi.org/10.1109/ICIP.2016.7532801.
[159] A. Creswell, A.A. Bharath, Adversarial training for sketch retrieval, in: G. Hua, H. Jégou (Eds.), Computer Vision-ECCV 2016 Workshops, Springer International Publishing, Cham, 2016, pp. 798–809. https://doi.org/10.1007/978-3-319-46604-0_55.
[160] T. Bui, L. Ribeiro, M. Ponti, J. Collomosse, Compact descriptors for sketch-based image retrieval using a triplet loss convolutional neural network, Comput. Vis. Image Underst. 164 (2017) 27–37. http://www.sciencedirect.com/science/article/pii/S1077314217301194. https://doi.org/10.1016/j.cviu.2017.06.007.
[161] K.-Y. Tseng, Y.-L. Lin, Y.-H. Chen, W.H. Hsu, Sketch-based image retrieval on mobile devices using compact hash bits, Proceedings of the 20th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2012, pp. 913–916. https://doi.org/10.1145/2393347.2396345.
[162] S. Zhang, Q. Tian, K. Lu, Q. Huang, W. Gao, Edge-SIFT: discriminative binary descriptor for scalable partial-duplicate mobile search, IEEE Trans. Image Process. 22 (7) (2013) 2889–2902. http://ieeexplore.ieee.org/abstract/document/6476016/. https://doi.org/10.1109/TIP.2013.2251650.
[163] X. Yang, X. Qian, Y. Xue, Scalable mobile image retrieval by exploring contextual saliency, IEEE Trans. Image Process. 24 (6) (2015) 1709–1721. http://ieeexplore.ieee.org/abstract/document/7056555/. https://doi.org/10.1109/TIP.2015.2411433.
[164] D. Parikh, Discovering localized attributes for fine-grained recognition, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Washington, DC, USA, 2012, pp. 3474–3481. http://dl.acm.org/citation.cfm?id=2354409.2355058. https://doi.org/10.1109/CVPR.2012.6248089.
[165] A. Vedaldi, S. Mahendran, S. Tsogkas, S. Maji, R. Girshick, J. Kannala, E. Rahtu, I. Kokkinos, M.B. Blaschko, D. Weiss, B. Taskar, K. Simonyan, N. Saphra, S. Mohamed, Understanding objects in detail with fine-grained attributes, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2014, pp. 3622–3629. https://doi.org/10.1109/CVPR.2014.463.
[166] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, Y. Wu, Learning fine-grained image similarity with deep ranking, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2014, pp. 1386–1393. https://doi.org/10.1109/CVPR.2014.180.
[167] X. Zhang, F. Zhou, Y. Lin, S. Zhang, Embedding label structures for fine-grained feature representation, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 1114–1123. http://ieeexplore.ieee.org/document/7780495/. https://doi.org/10.1109/CVPR.2016.126.
[168] F. Zhou, Y. Lin, Fine-grained image classification by exploring bipartite-graph labels, Computer Vision and Pattern Recognition, 2016. pp. 1124–1133. http://ieeexplore.ieee.org/abstract/document/7780496/. https://doi.org/10.1109/CVPR.2016.127.
[169] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, Z. Zhang, The application of two-level attention models in deep convolutional neural network for fine-grained image classification, IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp. 842–850. http://ieeexplore.ieee.org/document/7298685/. https://doi.org/10.1109/CVPR.2015.7298685.
[170] S. Gong, Y. Li, T.M. Hospedales, Y.-Z. Song, Fine-grained sketch-based image retrieval by matching deformable part models, British Machine Vision Conference, BMVC, 2014. pp. 1–12. http://www.eecs.qmul.ac.uk/sgg/papers/LiEtAl_BMVC14.pdf.
[171] K. Li, K. Pang, Y.Z. Song, T. Hospedales, H. Zhang, Y. Hu, Fine-grained sketch-based image retrieval: the role of part-aware attributes, IEEE Winter Conference on Applications of Computer Vision (WACV), 2016. pp.

1–9. http://ieeexplore.ieee.org/abstract/document/7477615/. https://doi.org/10.1109/WACV.2016.7477615.

[172] K. Li, K. Pang, Y. Song, T.M. Hospedales, T. Xiang, H. Zhang, Synergistic instance-level subspace alignment for fine-grained sketch-based image retrieval, IEEE Trans. Image Process. 26 (12) (2017) 5908–5921. http://ieeexplore.ieee.org/document/8016664/. https://doi.org/10.1109/TIP.2017.2745106.

[173] P. Xu, Q. Yin, Y. Qi, Y.-Z. Song, Z. Ma, L. Wang, J. Guo, Instance-level coupled subspace learning for fine-grained sketch-based image retrieval, in: G. Hua, H. Jégou (Eds.), Computer Vision-ECCV 2016 Workshops, Springer International Publishing, Cham, 2016, pp. 19–34. https://link.springer.com/chapter/10.1007/978-3-319-46604-0_2. https://doi.org/10.1007/978-3-319-46448-0.

[174] Q. Yu, F. Liu, Y.Z. Song, T. Xiang, T.M. Hospedales, C.L. Chen, Sketch me that shoe, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 799–807. http://ieeexplore.ieee.org/abstract/document/7780462/. https://doi.org/10.1109/CVPR.2016.93.

[175] J. Song, Y.Z. Song, T. Xiang, T. Hospedales, R. Xiang, Deep multi-task attribute-driven ranking for fine-grained sketch-based image retrieval, British Machine Vision Conference, 2016. pp. 1–11. http://www.bmva.org/bmvc/2016/papers/paper132/index.html. https://doi.org/10.5244/C.30.132.

[176] P. Sangkloy, N. Burnell, C. Ham, J. Hays, The sketchy database: learning to retrieve badly drawn bunnies, ACM Trans. Graph. 35 (4) (2016) 1–12. https://doi.org/10.1145/2897824.2925954.

[177] L. Liu, F. Shen, Y. Shen, X. Liu, L. Shao, Deep sketch hashing: fast free-hand sketch-based image retrieval, IEEE Conference on Computer Vision and Pattern Recognition, 2017. pp. 2298–2307. http://ieeexplore.ieee.org/abstract/document/8099730/. https://doi.org/10.1109/CVPR.2017.247.

[178] O. Seddati, S. Dupont, S. Mahmoudi, Triplet networks feature masking for sketch-based image retrieval, in: F. Karray, A. Campilho, F. Cheriet (Eds.), Image Analysis and Recognition, Springer International Publishing, Cham, 2017, pp. 296–303. https://link.springer.com/chapter/10.1007/978-3-319-59876-5_33.

[179] K. Pang, Y.-Z. Song, T. Xiang, T. Hospedales, Cross-domain generative learning for fine-grained sketch-based image retrieval, The British Machine Vision Conference (BMVC), 2017.

[180] J. Song, Q. Yu, Y.Z. Song, T. Xiang, T.M. Hospedales, Deep spatial-semantic attention for fine-grained sketch-based image retrieval, 2017 IEEE International Conference on Computer Vision (ICCV), 2017. pp. 5552–5561. http://openaccess.thecvf.com/content_ICCV_2017/papers/Song_Deep_Spatial-Semantic_Attention_ICCV_2017_paper.pdf. https://doi.org/10.1109/ICCV.2017.592.

[181] J. Song, Y.-Z. Song, T. Xiang, T. Hospedales, Fine-grained image retrieval: the text/sketch input dilemma, The British Machine Vision Conference (BMVC), 2017. https://www.research.ed.ac.uk/portal/files/41072731/song2017textSketch.pdf.

[182] K. Pang, D. Li, J. Song, Y.-Z. Song, T. Xiang, T.M. Hospedales, Deep factorised inverse-sketching, The European Conference on Computer Vision (ECCV), 2018. pp. 37–54. http://openaccess.thecvf.com/content_ECCV_2018/papers/Kaiyue_Pang_Deep_Factorised_Inverse-Sketching_ECCV_2018_paper.pdf.

[183] P. Xu, Y. Huang, T. Yuan, K. Pang, Y.-Z. Song, T. Xiang, T.M. Hospedales, Z. Ma, J. Guo, Sketchmate: d hashing for million-scale human sketch retrieval, (2018) CoRR abs/1804.01401. arXiv:1804.01401. http://arxiv.org/abs/1804.01401.

[184] A. Frome, G.S. Corrado, J. Shlens, S. Bengio, J. Dean, M. Ranzato, T. Mikolov, DeviSE: a deep visual-semantic embedding model, Proceedings of the 26th International Conference on Neural Information Processing, Curran Associates Inc., USA, 2013, pp. 2121–2129. http://dl.acm.org/citation.cfm?id=2999792.2999849.

[185] C. Ding, L. Liu, A survey of sketch based modeling systems, Front. Comp. Sci. 10 (6) (2016) 985–999. https://doi.org/10.1007/s11704-016-5422-9.

[186] B. Li, Y. Lu, A. Godil, T. Schreck, B. Bustos, A. Ferreira, T. Furuya, M.J. Fonseca, H. Johan, T. Matsuda, R. Ohbuchi, P.B. Pascoal, J.M. Saavedra, A comparison of methods for sketch-based 3D shape retrieval, Comput. Vis. Image Underst. 119 (2014) 57–80. https://doi.org/10.1016/j.cviu.2013.11.008.

[187] Y.-L. Lin, C.-Y. Huang, H.-J. Wang, W. Hsu, 3D sub-query expansion for improving sketch-based multi-view image retrieval, The IEEE International Conference on Computer Vision (ICCV), 2013. pp. 3495–3502. http://ieeexplore.ieee.org/abstract/document/6751546/. https://doi.org/10.1109/ICCV.2013.434.

[188] Y. Kang, H. Lei, S. Lin, Y. Li, Sketch-based shape retrieval using orientation histogram with Gabor filters, International Conference on Digital Home, 2014. pp. 277–281. http://ieeexplore.ieee.org/abstract/document/6996775/. https://doi.org/10.1109/ICDH.2014.60.

[189] M. Eitz, R. Richter, T. Boubekeur, K. Hildebrand, M. Alexa, Sketch-based shape retrieval, ACM Trans. Graph. 31 (4) (2012) 1–10. http://dl.acm.org/citation.cfm?doid=2185520.2335382. https://doi.org/10.1145/2185520.2335382.

[190] X. Xie, K. Xu, N.J. Mitra, D. Cohen-Or, W. Gong, Q. Su, B. Chen, Sketch-to-design: c-based part assembly, Comput. Graphics Forum 32 (8) (2013) 233–245. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.698.4844. https://doi.org/10.1111/cgf.12200.

[191] T. Furuya, R. Ohbuchi, Hashing cross-modal manifold for scalable sketch-based 3D model retrieval, Proceedings of the International Conference on 3D Vision, IEEE Computer Society, Washington, DC, USA, 2014, pp. 543–550. https://doi.org/10.1109/3DV.2014.72.

[192] F. Wang, L. Kang, Y. Li, Sketch-based 3D shape retrieval using convolutional neural networks, IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp. 1875–1883. http://ieeexplore.ieee.org/abstract/document/7298797/. https://doi.org/10.1109/CVPR.2015.7298797.

[193] J. Lee, T. Funkhouser, Sketch-based search and composition of 3D models, Proceedings of the Fifth Eurographics Conference on Sketch-Based Interfaces and Modeling, Eurographics Association, Aire-la-Ville,Switzerland, 2008, pp. 97–104. https://doi.org/10.2312/SBM/SBM08/097-104.

[194] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, M. Burtscher, H. Fu, T. Furuya, H. Johan, J. Liu, R. Ohbuchi, A. Tatsuma, C. Zou, Extended large scale sketch-based 3D shape retrieval, Proceedings of the 7th Eurographics Workshop on 3D Object Retrieval, Eurographics Association, Aire-la-Ville, Switzerland, 2014, pp. 121–130. https://doi.org/10.2312/3dor.20141058.

[195] T. Furuya, R. Ohbuchi, Ranking on cross-domain manifold for sketch-based 3D model retrieval, International Conference on Cyberworlds, IEEE. 2013, pp. 274–281. http://ieeexplore.ieee.org/document/6680126/. https://doi.org/10.1109/CW.2013.60.

[196] J.M. Saavedra, B. Bustos, M. Scherer, T. Schreck, STELA: sketch-based 3D model retrieval using a structure-based local approach, Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ACM, New York, NY, USA, 2011, pp. 1–8. https://doi.org/10.1145/1991996.1992022.

[197] S. Buoncompagni, A. Franco, D. Maio, Efficient sketch recognition based on shape features and multidimensional indexing, in: M. Kurzynski, M. Wozniak, R. Burduk (Eds.), Proceedings of the 10th International Conference on Computer Recognition Systems CORES, Springer International Publishing, Cham, 2018, pp. 159–169. https://doi.org/10.1007/978-3-319-59162-9_17.

[198] C.-T. Tu, P.-C. Lin, J.-J. Lien, Free-hand sketches for 3D model retrieval using cascaded LSDA, Multimed. Tools Appl. 76 (9) (2017) 11669–11685. https://doi.org/10.1007/s11042-016-3326-1.

[199] D. Mao, H. Yin, H. Li, Q. Cai, Sketch-based 3D model shape retrieval using multi-feature fusion, in: Y. Jia, J. Du, H. Li, W. Zhang (Eds.), Proceedings of the 2015 Chinese Intelligent Systems Conference, Springer Berlin Heidelberg. 2016, pp. 395–402. https://doi.org/10.1007/978-3-662-48365-7_40.

[200] Y. Li, H. Lei, S. Lin, G. Luo, A new sketch-based 3D model retrieval method by using composite features, Multimed. Tools Appl. (11) (2017) 1–24. https://doi.org/10.1007/s11042-017-4446-y.

[201] Z. Yasseen, A. Verroust-Blondet, A. Nasri, View selection for sketch-based 3D model retrieval using visual part shape description, Vis. Comput. 33 (5) (2017) 565–583. https://doi.org/10.1007/s00371-016-1328-7.

[202] F. Zhu, J. Xie, Y. Fang, Learning cross-domain neural networks for sketch-based 3D shape retrieval, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI Press. 2016, pp. 3683–3689. http://dl.acm.org/citation.cfm?id=3016387.3016421.

[203] S. Bu, P. Han, Z. Liu, J. Han, H. Lin, Local deep feature learning framework for 3D shape, Comput. Graph. 46 (C) (2015) 117–129. https://doi.org/10.1016/j.cag.2014.09.007.

[204] B. Leng, S. Guo, X. Zhang, Z. Xiong, 3D object retrieval with stacked local convolutional autoencoder, Signal Process. 112 (C) (2015) 119–128. https://doi.org/10.1016/j.sigpro.2014.09.005.

[205] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view convolutional neural networks for 3D shape recognition, 2015 IEEE International Conference on Computer Vision (ICCV), 2015. pp. 945–953. http://ieeexplore.ieee.org/document/7410471/. https://doi.org/10.1109/ICCV.2015.114.

[206] T. Funkhouser, M. Kazhdan, P. Shilane, P. Min, W. Kiefer, A. Tal, S. Rusinkiewicz, D. Dobkin, Modeling by example, ACM SIGGRAPH 2004, ACM, New York, NY, USA, 2004, pp. 652–663. https://doi.org/10.1145/1186562.1015775.

[207] S. Savarese, L. Fei-Fei, 3D generic object categorization, localization and pose estimation, IEEE 11th International Conference on Computer Vision, 2007. pp. 1–8. http://ieeexplore.ieee.org/abstract/document/4408987/. https://doi.org/10.1109/ICCV.2007.4408987.

[208] A. Fukui, D.H. Park, D. Yang, A. Rohrbach, T. Darrell, M. Rohrbach, Multimodal compact bilinear pooling for visual question answering and visual grounding, (2016) CoRR abs/1606.01847. arXiv:1606.01847. http://arxiv.org/abs/1606.01847.

[209] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, Proceedings of the 30th International Conference on Neural Information Processing Systems, Curran Associates Inc., USA, 2016, pp. 2180–2188. http://dl.acm.org/citation.cfm?id=3157096.3157340.

[210] M. Mirza, S. Osindero, Conditional generative adversarial nets, Comput. Sci. (2014) 2672–2680. http://cn.arxiv.org/pdf/1411.1784.

[211] Z. Yang, W. Chen, F. Wang, B. Xu, Improving neural machine translation with conditional sequence generative adversarial nets, (2014) CoRR abs/1703.04887. arXiv:1703.04887. http://arxiv.org/abs/1703.04887.

[212] P. Isola, J. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, IEEE Conference on computer vision and pattern recognition, 2017. pp. 5967–5976. https://doi.org/10.1109/CVPR.2017.632.

[213] Q. Chen, J. Huang, R. Feris, L.M. Brown, J. Dong, S. Yan, Deep domain adaptation for describing people based on fine-grained clothing attributes, IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp. 5315–5324. http://ieeexplore.ieee.org/abstract/document/7299169/. https://doi.org/10.1109/CVPR.2015.7299169.

[214] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 640–651. http://ieeexplore.ieee.org/abstract/document/7478072/. https://doi.org/10.1109/TPAMI.2016.2572683.

[215] Z. Liu, P. Luo, S. Qiu, X. Wang, X. Tang, DeepFashion: powering robust clothes recognition and retrieval with rich annotations, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 1096–1104. http://ieeexplore.ieee.org/abstract/document/7780493/. https://doi.org/10.1109/CVPR.2016.124.

[216] Y. Wang, J. Choi, V.I. Morariu, L.S. Davis, Mining discriminative triplets of patches for fine-grained classification, Computer Vision and Pattern Recognition, 2016. pp. 1163–1172. http://ieeexplore.ieee.org/abstract/document/7780493/. https://doi.org/10.1109/CVPR.2016.131.
[217] X. Zhang, H. Xiong, W. Zhou, W. Lin, Q. Tian, Picking deep filter responses for fine-grained image recognition, Computer Vision and Pattern Recognition, 2016. pp. 1134–1142. http://ieeexplore.ieee.org/abstract/document/7780497/. https://doi.org/10.1109/CVPR.2016.128.
[218] Y. Fu, L. Sigal, Semi-supervised vocabulary-informed learning, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 5337–5346. http://ieeexplore.ieee.org/document/7780945/. https://doi.org/10.1109/CVPR.2016.576.
[219] S. Changpinyo, W.L. Chao, B. Gong, F. Sha, Synthesized classifiers for zero-shot learning, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 5327–5336. http://ieeexplore.ieee.org/document/7780944/. https://doi.org/10.1109/CVPR.2016.575.
[220] Z. Akata, M. Malinowski, M. Fritz, B. Schiele, Multi-cue zero-shot learning with strong supervision, IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 59–68. http://ieeexplore.ieee.org/document/7780383/. https://doi.org/10.1109/CVPR.2016.14.
[221] Y. Qi, W.-S. Zheng, T. Xiang, Y.-Z. Song, H. Zhang, J. Guo, One-shot learning of sketch categories with co-regularized sparse coding, Advances in Visual Computing, Springer International Publishing, Cham, 2014, pp. 74–84. https://doi.org/10.1007/978-3-319-14364-4_8.
[222] M. Längkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recogn. Lett. 42 (2014) 11–24. http://www.sciencedirect.com/science/article/pii/S0167865514000221. https://doi.org/10.1016/j.patrec.2014.01.008.