Journal of Real-Time Image Processing
https://doi.org/10.1007/s11554-018-0784-x

SPECIAL ISSUE PAPER

Partially shaded sketch-based image search in real mobile device environments via sketch-oriented compact neural codes

Jamil Ahmad[1] · Khan Muhammad[2] · Syed Inayat Ali Shah[3] · Arun Kumar Sangaiah[4] · Sung Wook Baik[2]

Received: 30 September 2017 / Accepted: 25 April 2018

Abstract
With the advent of touch screens in mobile devices, sketch-based image search is becoming the most intuitive method to query multimedia contents. Traditionally, sketch-based queries were formulated with hand-drawn shapes without any shades or colors. The absence of such critical information from sketches increased the ambiguity between natural images and their sketches. Although it was previously considered too cumbersome for users to add colors to hand-drawn sketches in image retrieval systems, the modern day touch input devices make it convenient to add shades or colors to query sketches. In this work, we propose deep neural codes extracted from partially colored sketches by an efficient convolutional neural network (CNN) fine-tuned on sketch-oriented augmented dataset. The training dataset is constructed with hand-drawn sketches, natural color images, de-colorized, and de-texturized images, coarse and fine edge maps, and flipped and rotated images. Fine-tuning CNN with augmented dataset enabled it to capture features effectively for representing partially colored sketches. We also studied the effects of shading and partial coloring on retrieval performance and show that the proposed method provides superior performance in sketch-based large-scale image retrieval on mobile devices as compared to other state-of-the-art methods.

Keywords Sketch-based query · Image retrieval · Hash codes · Deep learning · Convolutional neural network

1 Introduction

The rapid growth of mobile devices has contributed significantly to the volume of images being generated and consumed each day [1]. Consequently, the volume of image data on mobile devices has also increased at a rapid pace. Sketch-based image retrieval (SBIR) provides a convenient and intuitive method to query visual contents on touch screen devices. However, this mode of interaction has not seen widespread adaptation for image search despite its intuitive nature. Inherent ambiguity of hand-drawn sketches and focus only on the structural characteristics of sketches during features extraction, are the two major reasons behind its instability. Previous studies concluded that adding colors or shades to sketches are very cumbersome for users during query specification [2]. Though this conclusion holds true for users interacting with PCs using mouse and keyboard, the present touch screen devices make it very convenient to efficiently add partial colors and shades to their sketches.

With colorless sketch-based queries, it becomes difficult to understand users' intent during query processing in image retrieval systems. Researchers primarily focus on shape or structure of the sketch and attempt to match it with the edge map of the natural image during image matching [3, 4]. The absence of essential characteristics such as colors and textures restrict the matching process to a model fitting method which attempts to match sketch with natural images [5, 6]. Such methods carry a huge computational cost and restrict the use of these methods on resource constrained devices. One solution to cope with this issue is to use cloud-based services to perform the comparison; however, users typically refrain from uploading their private photos to cloud [7–9]. With the ever-growing volume of images on mobile devices and the need to effectively query these large datasets, the need

✉ Sung Wook Baik
sbaik@sejong.ac.kr

[1] Department of Computer Science, Islamia College, Peshawar, Pakistan

[2] Digital Contents Research Institute, Sejong University, Seoul, South Korea

[3] Department of Mathematics, Islamia College Peshawar, Peshawar, Pakistan

[4] School of Computing Science and Engineering, VIT University, Vellore, India

Published online: 31 May 2018

to devise efficient methods for sketch-based image search are on the rise.

Hand-drawn colorless sketches can be highly ambiguous as depicted in Fig. 1a. Typical SBIR systems attempt to reduce the semantic gap between sketches and images by enhancing features extraction and recognition methods and do not take into consideration the inherent ambiguity due to the absence of shades or colors. Even humans fail to recognize hand-drawn colorless sketches 27% of the time [10]. To overcome these problems, we present an efficient method for sketch-based image search on mobile devices using partially colored sketches and sketch-oriented compact neural codes. Instead of focusing only on the features, we also evaluate the importance of partial shading in the context of SBIR. To support partially shaded sketches recognition, an augmented dataset is developed for training deep CNNs, keeping in mind the various characteristics of sketches. The augmented dataset contain images generated by several semantics preserving transformations like coarse and fine edge maps, de-texturized, de-colorized, flipped, rotated images, hand-drawn sketches, and natural images. An efficient convolutional neural network, pretrained on the ImageNet database [11] is re-trained on the augmented dataset using the transfer learning approach to obtain compact binary codes. These binary codes are utilized as hash codes on mobile devices to allow efficient and precise access to large collections of images [12]. Some of the major contributions of this work are as follows:

1. A sketch-oriented data augmentation procedure is devised to construct an augmented dataset for training deep CNNs to recognize partially shaded sketches.
2. A compact and efficient hash code is generated for sketch representation and retrieval.
3. The effects of partial shading on retrieval performance using deep features have been thoroughly investigated.

The rest of the paper is organized as: Sect. 2 presents an overview of the recent SBIR methods using deep learning approaches. Schematics of the proposed method, implementation details, and design choices are highlighted in Sect. 3. Section 4 summarizes the experimental results and the paper concludes in Sect. 5.

2 Related work

Image retrieval systems have been significantly improved with the advancements in deep learning techniques. Convolutional neural networks [13–15], Siamese CNN [16], deep belief networks (DBN) and denoising auto-encoders (DAE) [17] have become state-of-the-art methods for image classification and retrieval [18]. Besides the traditional sample-based image retrieval, these methods have also been investigated for improving SBIR systems. A brief overview of some of deep learning based SBIR approaches are being presented here.

In the context of CBIR, Babenko et al. [18] showed that the multi-layer representations learned by CNNs during training on large datasets such as ImageNet can be used to effectively represent images in CBIR systems. They showed that lower layers learn basic features whereas higher level layers learn more complex and semantic features, allowing accurate retrieval of images. Though CNNs provide state-of-the-art performance in image retrieval systems which have been thoroughly investigated. Their effectiveness in SBIR systems is yet to be determined due to the large semantic gap between images and their hand-drawn sketches [19–21]. In an initial attempt towards SBIR, Chen et al. [22] developed a system to transform hand-drawn sketch to a photo-realistic image by seamlessly combining images discovered on the web. Extracting features from a full color image is relatively straightforward which transforms SBIR to CBIR. The first notable work in sketch recognition using CNN was carried out by Fu et al. [23], who used CNN to recognize pre-defined symbols in sketches from various engineering diagrams. Extending their work, Kiran and Babu [24] used AlexNet and LeNet models to extract features from hand-drawn sketches. These features were used to train support vector machine (SVM) classifier to recognize sketches [8, 25]. In [16], Qi et al. performed pair-wise image matching between a hand-drawn sketch and target image using Siamese CNN. A typical Siamese CNN consist of two identical

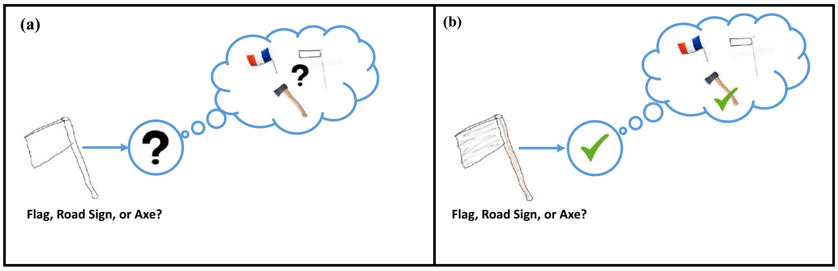


Fig. 1 **a** Difficulty in interpreting colorless sketches, **b** less ambiguity in partially shaded sketches

CNNs whose cost functions are linked together. Two images are simultaneously input to the network which generates a real-valued output, corresponding to the visual similarity between image pairs. Smaller values correspond to greater visual similarity, whereas larger values are interpreted as dissimilarity. During training, pairs of images corresponding to both similar and dissimilar classes are forward propagated. The network attempts to reduce the feature distance between same-class inputs and increases when inputs belong to different classes. Though their method provided superior performance, it carried with it a huge computational cost, making it infeasible for image search on mobile devices. Omar et al. [26] investigated features extracted from three different fully connected layers of 15-layer CNN model trained on sketches to perform sketch retrieval. They showed that features from the highest layer carry more meaningful interpretation of the sketch as compared to the lower layers. Neuronal activations from the 13th layer were used to represent and retrieve sketches. Liu et al. [27] recently introduced a semi-heterogeneous deep networks architecture where three CNNs work as hash functions to encode hand-drawn sketches, natural images, and auxiliary sketch tokens. Their method achieved state-of-the-art performance on sketch datasets. In another work [28], Wang et al. trained a CNN by combining natural images with their hand-drawn sketches or edge maps. Consequently, they used it to represent both images and sketches for SBIR systems. Different rotated versions of the edge maps were forward propagated through the network during training phase to improve the discriminative power of the network. They tested their method on a large dataset and showed that their network correctly retrieved images in response to sketch-based queries. Ahmad et al. used data augmentation and mixed sketches with edge maps and full color images to train CNN. The trained model was then used to retrieve images using partially colored sketches. These works showed that artificially augmenting training datasets allow CNNs to extract essential characteristics of both images and sketches, thereby allowing the features to be used for SBIR. Extending their work, we propose to enhance the augmented training set by including several other images generated using semantics preserving transformations. We believe that the extended dataset will further improve representation of sketches using CNNs. We also propose an efficient CNN architecture for generating short binary codes for allowing sketches to be represented efficiently on mobile devices.

3 Proposed method

Sketch-based query to access visual contents on mobile devices equipped with touch screens is an intuitive mode which has attracted attention in recent years. Consequently, several methods have been proposed consisting of both hand-engineered features as well as learned representations. The superior performance of deep learning approaches resulted in significant improvements in SBIR [26, 29]. In this work, we propose a sketch-oriented data augmentation procedure to effectively train an efficient deep CNN for representing partially shaded sketches. The network architecture being used is inspired from the recent works of Iandola et al. [30], namely SqueezeNet. The architecture has been modified to generate compact hash codes to effectively represent sketches on mobile devices. The overall framework of the proposed approach is illustrated in Fig. 2. A roughly hand-drawn, partially shaded sketch is input by the user to the SBIR system. Compact feature code is extracted from the CNN which is then matched with the stored hash codes for candidate images on the mobile device. The retrieved images are ranked on the basis of the hamming distance. Further details of the various components are provided in the subsequent sections.

3.1 Sketch-oriented data augmentation

Effective training of deep CNNs heavily rely on the availability of huge volumes of data. Semantics preserving transformation approaches have been used to effectively expand training datasets to allow CNN to effectively model abstractions from data. Data augmentation techniques have shown great promise in improving image recognition performance by enhancing the discriminative capability of the CNN model [13]. In the context of SBIR, Wang et al. [28] mixed hand-drawn sketches with natural images and their edge maps to train deep CNNs for sketch recognition. Their method effectively improved sketch recognition over existing methods. Ahmad et al. [31] used augmented dataset to allow partially shaded sketches be represented effectively by a CNN. They achieved better performance with sketches which were partially colored during query submission. Inspired by these works, we have extended the number of semantics preserving transformations to generate a more expanded dataset. A total of nine different version for each image were augmented using the following data augmentation approaches as shown in Fig. 3.

a. **Coarse and fine edge maps**

Edge maps are an effective way to transform images to sketches. Inclusion of edge maps will allow CNNs to model contours of objects. Canny and Sobel edge detectors were used to obtain fine (FEM) and coarse edge maps (CEM), respectively. The fine edge map captures contour as well as the interior fine structures of objects. On the contrary, coarse edge map only extracts their salient edge features.

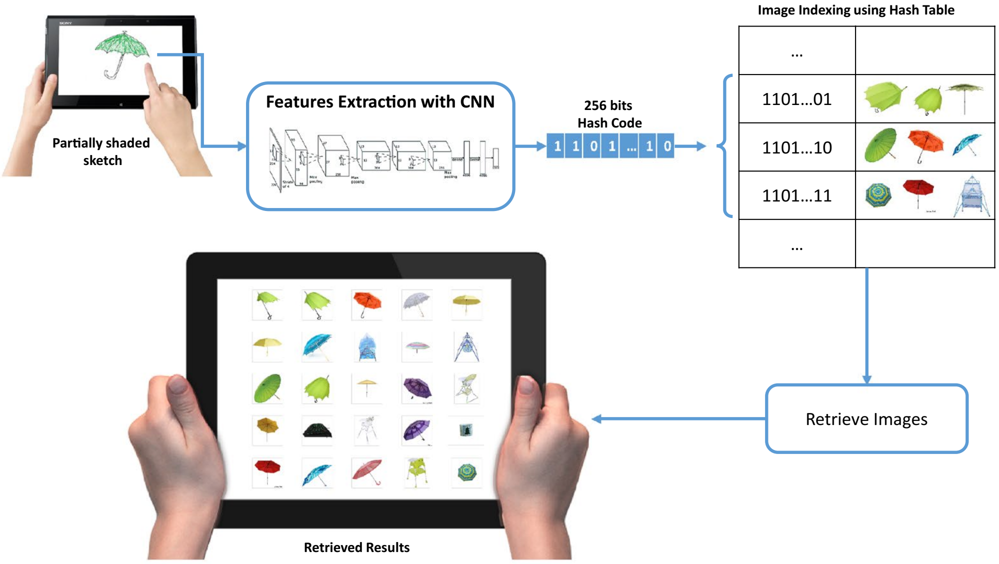
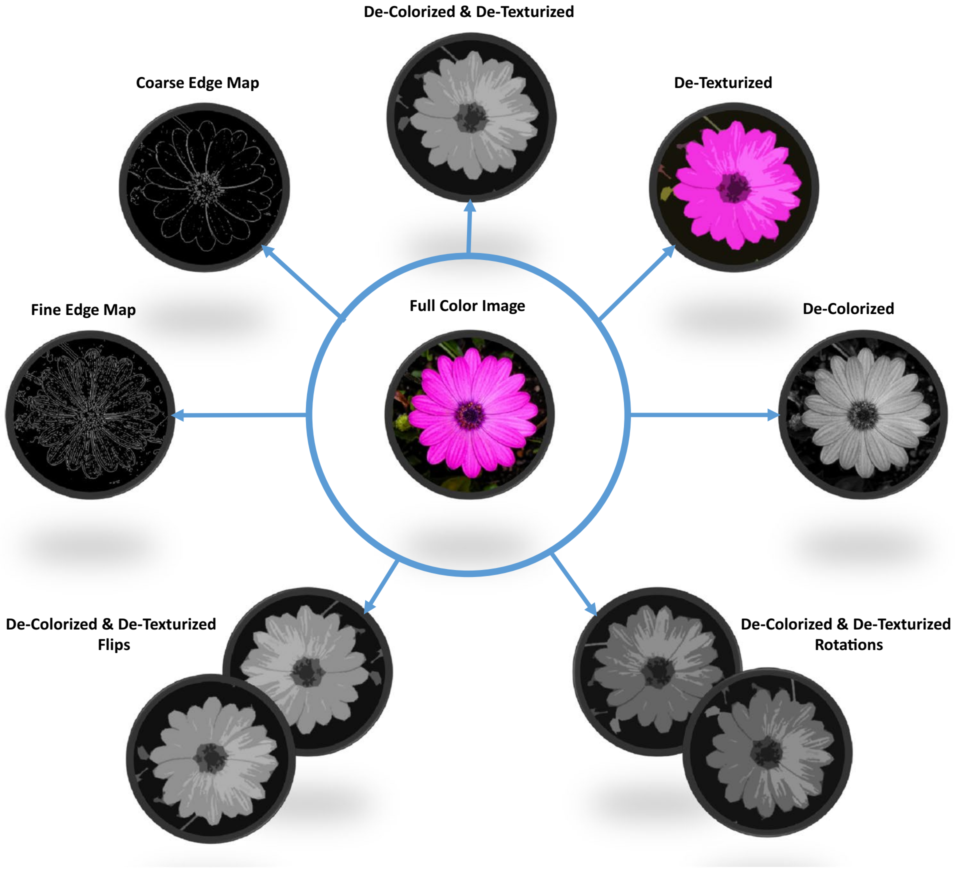


Fig. 2 Schematics of the proposed framework

Fig. 3 Semantics preserving transformations for sketch-oriented data augmentation


De-Colorized & De-Texturized
Coarse Edge Map
De-Texturized
Fine Edge Map
Full Color Image
De-Colorized
De-Colorized & De-Texturized
Flips
De-Colorized & De-Texturized
Rotations

b. **De-colorization**

Hand-drawn artistic sketches usually consist of shaded portions. Although SBIR systems usually do not require artistic sketches, our method allow users to roughly shade the interiors of their sketches for improved retrieval results. Hence, de-colorization (DC) of natural images will allow CNNs to consider shades of various objects during representation. De-colorization is achieved using the following color to gray transformation:

$$I_{\mathrm{Gray}} = 0.299 \times I_{\mathrm{R}} + 0.589 \times I_{\mathrm{G}} + 0.112 \times I_{\mathrm{B}}, \quad (1)$$

where I_R is the red component, I_G is the green component, I_B is the blue component and I_{Gray} is the grayscale version of the image.

c. **De-texturization**

SBIR users may not draw sketches as detailed as an artist and may ignore fine details. Therefore, we have included de-texturized (DT) versions of natural images to our augmented dataset. These transformations are obtained using anisotropic diffusion [32]:

$$\frac{\partial I}{\partial t} = c(x, y, t)\Delta I + \nabla c \cdot \nabla I, \quad (2)$$

$$c(||\nabla I||) = \frac{1}{1 + \left(\frac{||\nabla I||}{K}\right)^2}, \quad (3)$$

where t is the time scale parameter, c is the proposed flux function to control the rate of diffusion at any point in the image I, and K is the edge-strength parameter to consider a valid edge boundary. For our experiments, we set $t = 128$ and $K = 0.2$.

d. **De-colorized and de-texturized**

De-colorized and de-texturized (DCDT) version of images were also added to the augmented dataset, keeping in view the fact that sketches may or may not include colors.

e. **Flips/rotations**

Horizontal and vertical flipped versions (FV) as well as two rotations (ROT) (60° and 120°) of the de-texturized and de-colorized images were also included in the augmented dataset. It will give the users some degree of freedom to sketch their queries.

3.2 Efficient deep CNN

Deep learning approaches, especially deep CNNs are known for their state-of-the-art performance in image recognition and retrieval. It is due to their superior ability to automatically extract features from raw data which makes them so powerful for these tasks. CNNs have been in the works since the mid-1980s; however, extensive research started on these hierarchical architectures when AlexNet CNN won the ImageNet large scale visual recognition challenge in 2012 by a huge margin [13]. Since then CNNs have been thoroughly studied and applied to a wide variety of computer vision and non-vision tasks. Simonyan et al. [33] studied the effects of depth in CNNs and showed that smaller kernels and many layers effectively improve the internal representations of visual contents as compared to shallower models. Their work was further improved by He et al. [34]. who introduced residual units to CNNs and designed very deep models having hundreds of layers, and outperformed humans in image recognition tasks. As CNNs grow deeper, they became more computation hungry, thereby limiting their use in resource constrained devices. Hence, researchers designed efficient architectures which yielded similar performance while keeping the computation and memory requirements at minimum. CNNs such as Network-in-Network (NiN) [35], SqueezeNet [30], ZynqNet [36] and the recently introduced MobileNets [37] offer highly efficient architectures with less memory requirements. Further, compression methods have also been developed to reduce the memory and computation needs of deep CNNs [38].

Though MobileNets are highly efficient and powerful architectures, specifically developed for mobile applications, they are slightly computationally expensive than SqueezeNet. Therefore, in this work, we adapted the SqueezeNet architecture and modified it to generate compact hash codes by introducing two fully connected layers at the end before the Softmax layer. A hash layer of 256 neurons with sigmoid activation function, and an FC layer with 250 neurons. The outcome of the hash layer is converted to binary codes by simple thresholding approach. The rest of the architecture remains the same. SqueezeNet receives input of size 227 × 227 × 3. The first convolution layer applies 64 kernels of size 3 × 3 with a stride of 2 which reduces the inputs by a factor of 2. A max pooling layer with neighborhood size 3 × 3, stride 2 further down samples the input by half. SqueezeNet also consist of smaller units called fire modules which consist of a squeeze layer and an expand layer. The squeeze layer only contains 1 × 1 kernels and attempts to reduce the number of channels, whereas the expand layer consisting of both 1 × 1 and 3 × 3 kernels again increases the number of channels which introduces sparsity into the channels. A total of eight fire modules having the same architecture is shown in Fig. 4. The complete

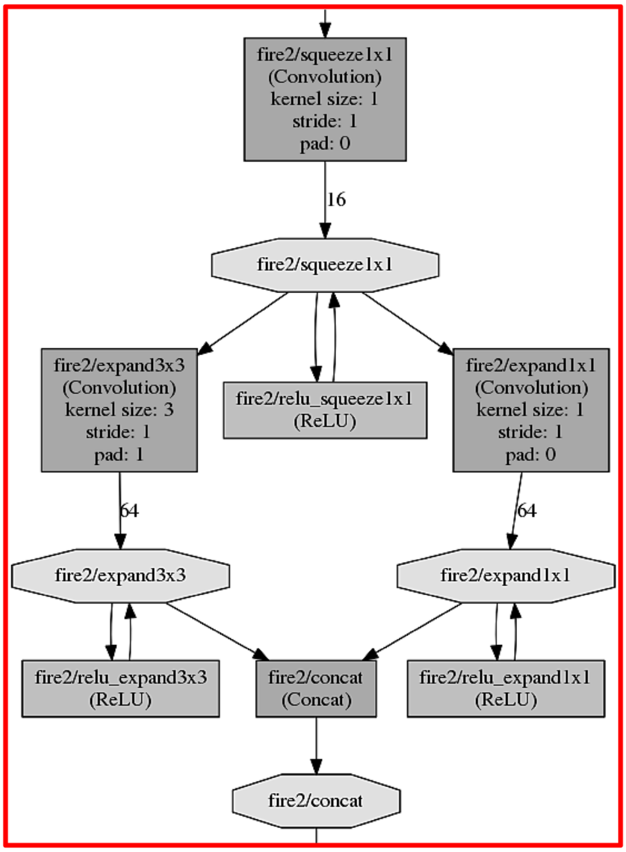


Fig. 4 SqueezeNet CNN microarchitecture (fire module)

Table 1 Architecture of efficient CNN for sketch classification

Layer name/type	Output size	Filter size/stride	Depth
Input image	227 × 227 × 3		
Conv1	113 × 113 × 64	3 × 3/2 (64 kernels)	1
MaxPool1	56 × 56 × 64	3 × 3/2	
FireModule2	56 × 56 × 128		2
FireModule3	56 × 56 × 128		2
FireModule4	56 × 56 × 256		2
MaxPool4	28 × 28 × 256	3 × 3/2	
FireModule5	28 × 28 × 256		2
FireModule6	28 × 28 × 384		2
FireModule7	28 × 28 × 384		2
FireModule8	28 × 28 × 512		2
MaxPool8	14 × 14 × 512	3 × 3/2	
FireModule9	14 × 14 × 512		2
Hash / Fully Connected	256		1
FC10/Fully Connected	250 × 256		1
Softmax	250		

architecture of the model we used is provided in Table 1. It is a 19-layer network whose increased depth gives it discriminative ability and the smaller kernels keep the memory requirements to a minimum. Although the model size increased as a result of introducing the fully connected layers, their memory needs are still less than the other models used for sketch recognition or SBIR.

3.3 Training with sketch-oriented augmented data

Training deep CNNs involves forward propagation of the training data in small batches to compute the loss cost, and backward propagation to adjust network parameters (weights and biases) using stochastic gradient descend (SGD) [13]. During the forward pass, the training cost is computed by computing the difference of predicted labels with ground truth. Parameter gradients are then computed using chain rule during the backward propagation phase to adjust their values. The process repeats until the loss cost gets sufficiently reduced. In our work, the CNN was trained on augmented dataset from scratch and then transfer learning strategy was used to fine-tune with a pre-trained ImageNet model to see if any improvements appear in the model. A significant improvement of 11% was noticed in the fine-tuned model which borrowed weights for certain layers from the original SqueezeNet model trained on ImageNet dataset [11]. During the fine-tuning process, the parameter values for the top 17 layers (conv1 to FireModule9) were acquired from the pre-trained model. The parameters for the Hash Layer and FC10 layers were initialized randomly. The learning rate for the top 17 layers was set to 0.001, so that only minor adjustments are made to their values and most of the previously learned knowledge is retained. Conversely, learning rates for the hash and FC10 layers were set to 0.01 so that they are sufficiently tuned according to the new dataset. The training batch size was set to 32 and the training process was repeated for 60 epochs. When the training finished, the freshly trained model achieved 72.6% accuracy on the validation set, whereas the fine-tuned model achieved 79.6% classification accuracy.

3.4 Sketch representation with deep sketch features

It has been observed that the features learned by any layer of the deep CNN can serve as generic descriptors for image classification and retrieval. Further, discriminative ability of the higher layers has been found to be greater than the lower layers, because high level abstractions are modeled at deeper layers. Hence, we opted to compute small hash codes for representing sketches and images by introducing two fully connected (FC) layers after the FireModule9. The first FC layer named Hash consist of 256 neurons with sigmoid activation functions. The output of this layer is fed into another FC layer (FC10), which has 250 neurons (equal to the number of classes in our training dataset). FC10 layer forwards its output to the Softmax classifier which outputs probabilities for each class. Output of the Hash layer is used to represent sketches as hash codes after applying simple threshold function on their activations.

$$H_j = \left\{ \begin{matrix} 1, & \text{Hash}_j \geqslant 0.5 \\ 0, & \mathit{Otherwise} \end{matrix} \right\} \tag{4}$$

where $j = 1, 2, 3 \ldots n$ (length of the Hash layer), Hash_j is the jth activation value, and H_j is the jth hash value. Hash code for each image is represented by $H = \{H_1, H_2, H_3, \ldots H_n\}$ where n in the present work was set to 256 after experiments.

3.5 Retrieval on mobile devices

These hash codes are stored in the hash table along with image locations to which the hash codes correspond. When a query is input, the hash code of the query image is generated and is used to access particular locations in the hash table. For instance, the entry with exactly the same code as the query is directly accessed and the images at that point are retrieved. Similarly, images at nearby locations in the hamming space are also retrieved. Which are then ranked on the basis of hamming distance with the query code. In this way, images are efficiently reduced by avoiding exhaustive searching. The advantage of hash based image search is that the search space is significantly reduced. Only a small subset of the data is searched for locating relevant images. Hash code of the query is matched with hash codes of the candidate images using hamming distance. Smaller distance correspond to greater visual similarity and vice versa. The dissimilarity score s is computed by taking the hamming distance between the query hash code H_q and the candidate image H_i as:

$$s = \left\| H_q - H_i \right\| \tag{5}$$

4 Experiments and results

The proposed SBIR system is implemented in MATLAB 2015b [39], where Caffe and MatCaffe [40] are used to extract features from the deep CNNs. Training is accomplished using NVidia DIGITS [41] with Caffe as a backend on a powerful PC equipped with 64 GB RAM, Core i7 Processor, and NVidia GeForce GTX TITAN X (Pascal) 12 GB GPU. Mobile devices used for evaluations include Samsung Galaxy Note 4, and LG G3. Details of evaluations and their results are provided in the subsequent sections.

4.1 Datasets

We used a large dataset (TU Berlin Sketches [10]) having 20,000 hand-drawn sketches organized into 250 categories with 80 sketches in each category. Sixty sketches were used for training whereas the remaining were used for testing. The training dataset is constructed by mixing natural images from Caltech256 [42] dataset and their corresponding augmented images with the training sketches. Hundred natural images and their 10 augmented versions of each image were mixed with each sketch category. The entire training dataset consisted of 95,000 images having 15,000 sketches and 80,000 natural/augmented images. For testing the retrieval performance of proposed framework, we used the Multiview objects dataset [43] which consist of 5000 images of a variety of objects without context or background. We also used images from Corel-1K dataset to evaluate retrieval performance with hand-drawn sketches from a dataset which contain full color images with varying backgrounds.

4.2 Sketch-based retrieval

In this experiment, images were retrieved from the test image set consisting of natural images by using hand-drawn sketches from the test set of sketches in the TU-Berlin dataset. No color or shading was added to the sketches during this experiment. Random sketches were selected from each category to retrieve top-N images from the dataset. Retrieval performance was measured using the standard metric mean average precision (mAP). Figure 5 shows retrieval results of top 10 images retrieved from the dataset. These results reveal that thin objects such as bicycle and chair can be retrieved from the dataset using colorless sketches. However, the rest of the results indicate the difficulty in retrieval even with powerful CNN features. Only two relevant images were retrieved for the umbrella, shoe, and cup sketches. This failure indicates the inherent ambiguity in sketches and full color images. Though some of the images were correctly retrieved, overall performance with colorless sketches was significantly poor.

4.3 Effect of partial coloring or shading on sketch based retrieval

This experiment was designed to study the effects of colors or shades on SBIR. Varying amounts of color or gray shades were added to sketches and their retrieval performance was assessed. Sketches with different amounts of shading were input as query to the SBIR system, and top-10 images were retrieved as shown in Fig. 6. In the first query, there is no shading and the retrieval results are poor. By adding slight amount of color to the sketch, we were able to retrieve 6 relevant images out of 10. Increasing the amount of shading

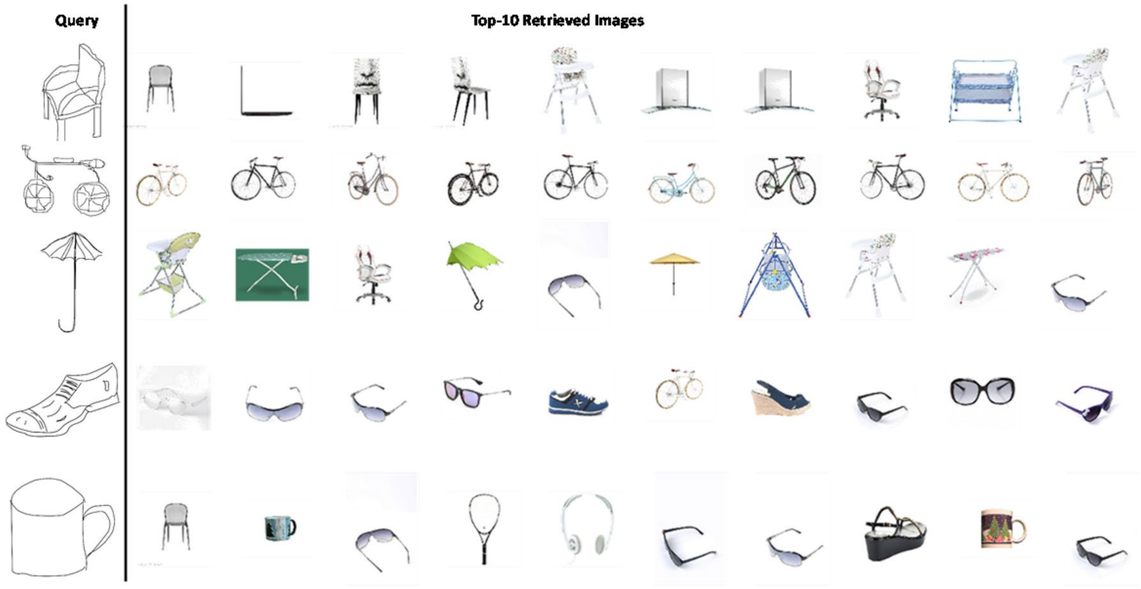


Fig. 5 Retrieval results for color-less sketch queries

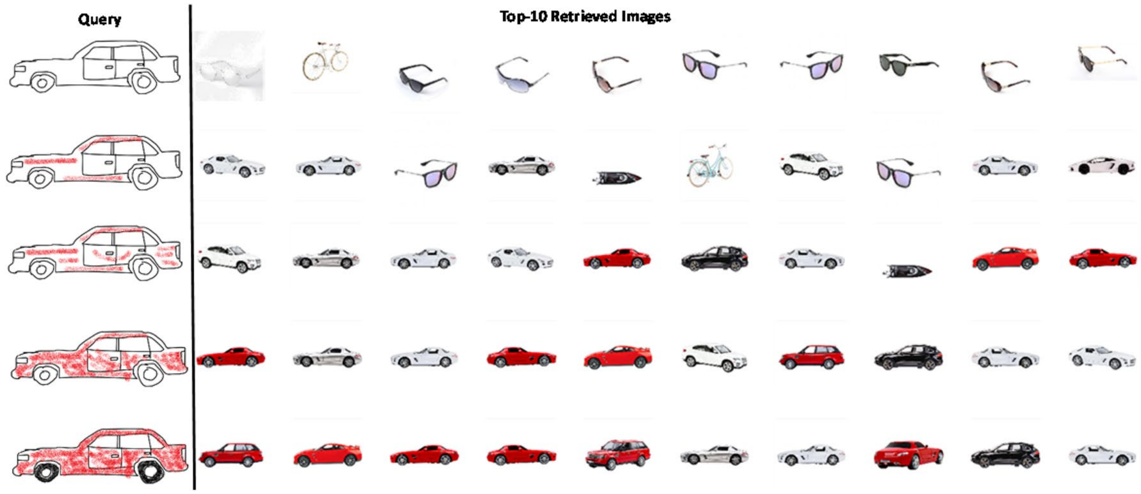


Fig. 6 Effects of colors or shades on SBIR

also improved the retrieval results. Experimental evaluations revealed that the addition to partial shading significantly improves retrieval performance. This phenomenon can be attributed to the strength of CNNs in modeling colors and textures. With colorless sketches, we were unable to utilize the full representation capability of deep CNN. This experiment shows that addition of shades can effectively decrease the semantic gap between sketches and images, which eventually improves retrieval performance.

Figure 7 shows top-10 retrieved images for partially shaded hand-drawn sketches. It is important to mention here, that the previous queries with these shapes failed to retrieve images correctly (Fig. 5). By adding shades of colors or gray, retrieval results can be significantly improved for sketches with greater ambiguity. For instance, chair and laptop sketches can be easily confused with other objects. Hence, addition of shades is essential in such sketches to overcome this confusion. Still, the shading need not be artistic and any hint of color or shade will yield improvement. In the previous queries (Fig. 5), only two correct images were retrieved for both shoe and cup. The addition of a few strokes of shading enabled the SBIR system to retrieve eight relevant images for shoe, and seven images for cup.

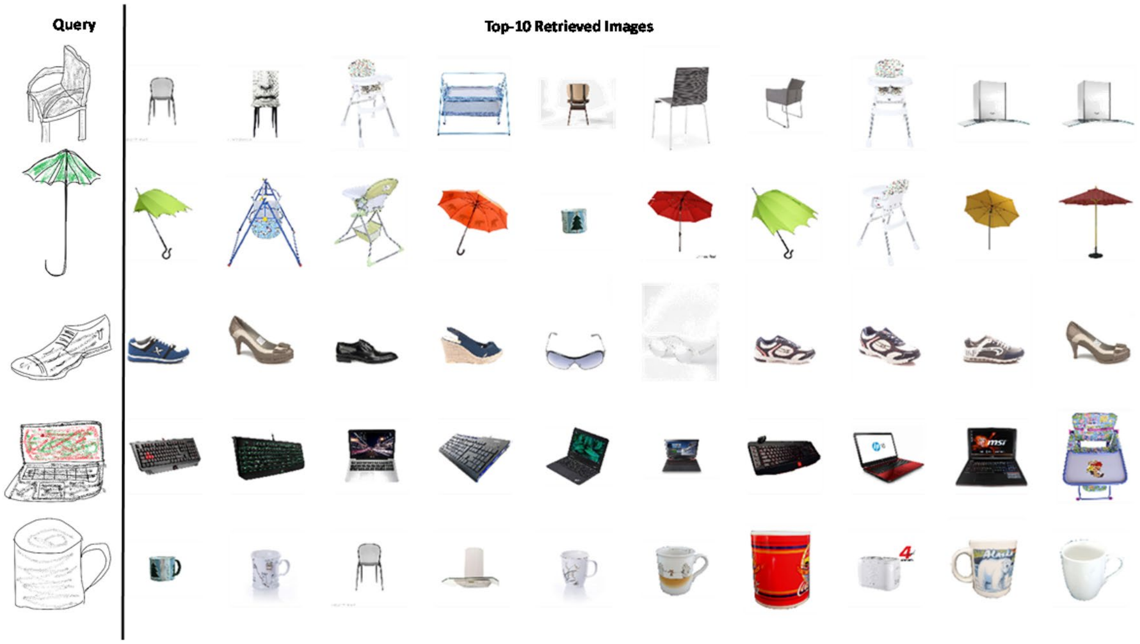


Fig. 7 SBIR results with partially colored/shaded sketches

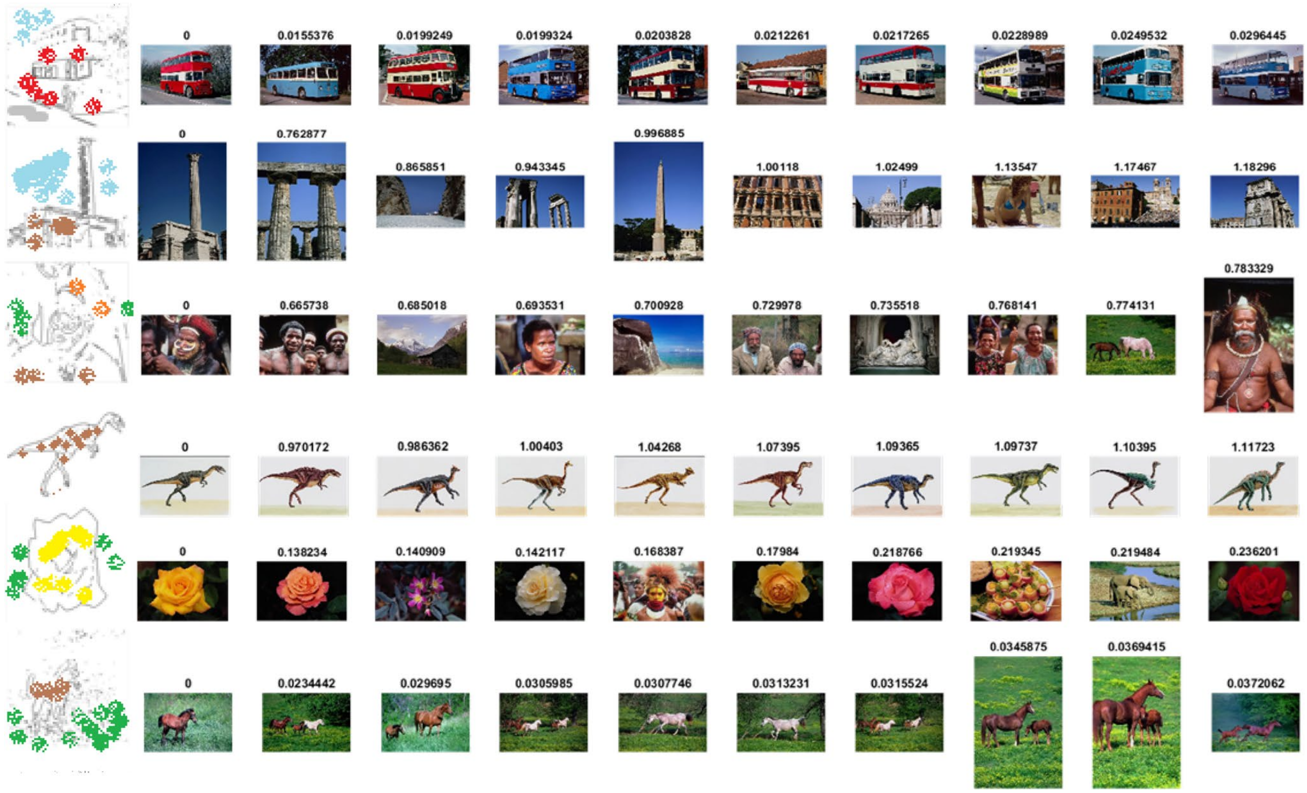


Fig. 8 SBIR results with partially colored/shaded sketches from full color images dataset

In Fig. 8, we attempted to retrieve full color images with varying backgrounds using partially shaded sketches and edge maps. The results suggest that the proposed method can successfully retrieve images when partial shading is added to the sketch. Some of the images have still been incorrectly retrieved which shows the weakness of the proposed approach. Since partial shading often introduces blobs of colors, it can sometimes confuse the CNNs. Other approaches such as Siamese CNNs could be employed to overcome these issues.

4.4 Effects of data augmentation on retrieval performance

Several models were obtained with different combinations of augmented data, to study the effects of semantic preserving transformations on SBIR. Results of evaluations are provided in Table 2. The mAP@0.5 scores correspond to the retrieval performance of images in response to partially shaded sketches. Random queries were selected to assess the performance of proposed scheme. The model trained using only sketches achieved mAP of 67.3% for partially shaded SBIR. When we added color images to the dataset, the mAP improved to 72.2 which enabled CNN to model colors and textures as well, thereby becoming more suitable for representing partially shaded sketches. Further, by adding coarse and fine edge maps of images, mAP above 75% was achieved. Extending the training set with de-colorized and de-texturized images resulted in significant improvements and the scores raised to 79%. We achieved the best scores by adding flips and rotations of the DCDT images which enhanced the discriminative ability of the CNN even further. It should also be noted that these results were obtained for sketch queries which were shaded at least 50%. These experimental results reveal that the addition of sketch-oriented data augmentation significantly improves the internal representation of CNNs which eventually improves retrieval performance of SBIR systems for partially shaded sketches.

Table 2 Performance comparison with varying data augmentation strategies

Dataset augmentations	mAP@0.5 (%)
Sketch only	67.3
Sketch + images	72.2
Sketch + images + CEM	75.6
Sketch + images + FEM	75.4
Sketch + images + CEM + FEM	76.3
Sketch + images + CEM + FEM + DC	77.6
Sketch + images + CEM + FEM + DC + DT	79.1
Sketch + images + CEM + FEM + DC + DT + DCDT	79.2
Sketch + images + CEM + FEM + DC + DT + DCDT + FV + ROT	79.6

4.5 Retrieval performance comparison with state-of-the-art

In the era of mobile devices, personalized SBIR systems are becoming popular. Traditionally SBIR systems were assessed on the basis of retrieval performance for colorless and shade-less hand-drawn sketches. In this work, we recommend users to add some degree of shading to their sketch which could significantly improve results. Here, we present a comparison of our method with other state-of-the-art methods in Table 3. Sketch-oriented data augmentations and inclusion of shade to queries have enabled our framework to outperform these methods. We strongly believe that the framework can be further enhanced with more sophisticated models and query formulation to improve SBIR performance on mobile devices.

4.6 Efficiency analysis

Image retrieval on resource constrained devices such as mobile phones require highly efficient indexing and retrieval methods. Computation and memory hungry methods could negate the very idea of mobile image search. Locality-sensitive hashing-based approaches are considered as highly efficient searching techniques. Recently, it has been shown the CNNs can be used to effectively transform high-dimensional features into low-dimensional representation without requiring explicit techniques to be used [45]. The proposed framework operate on the same principle. Our trained CNN

Table 3 Sketch-based image retrieval performance comparison

Method	Features	mAP@0.5 (%)
Eitz et al. [10]	SIFT + BoF + SVM	56.0
Deep CNN [44]	20 Layer CNN	72.2
Seddati et al. [26]	15 Layer CNN	75.4
DeepSketch [28]	8 Layer CNN trained using augmented dataset	77.3
Proposed Method	256 bits hash code from 19 Layer CNN trained using augmented dataset	**79.6**
Human		73.0

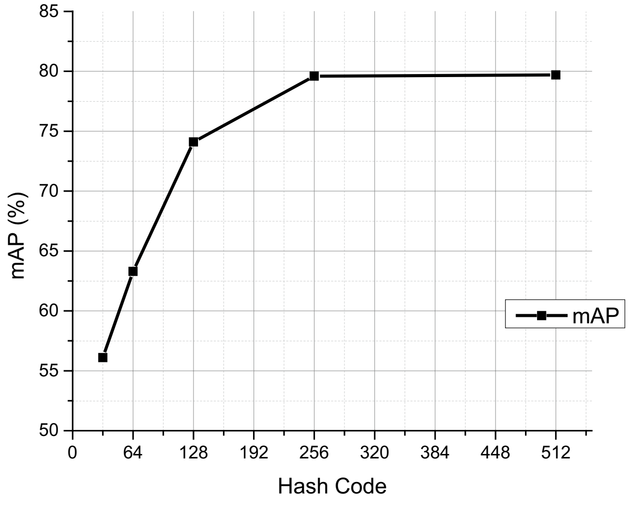


Fig. 9 Effect of hash code length on retrieval performance

generates compact hash codes of 256 bits for each sketch. The hash code length was determined through experiments as depicted in Fig. 9. The mAP@0.5 score with 32 and 64 bit codes were below 70. The score increased to 74 with 128 bits. The best score was achieved with 256 and 512 bit codes. However, the improvement in case of 512 bits was negligible, so we opted to use the 256 bits code. This hash code can be used to directly access the pool of candidate images without requiring exhaustive searching, which highly improves retrieval efficiency. It is a highly desirable property of SBIR systems for mobile devices because they lack the computational power and memory of a full scale PC.

We compared the retrieval results of proposed hash codes with codes generated by locality-sensitive hashing (LSH) [46, 47], spectral hashing (SH) [48], principal component analysis based hashing (PCAH) [18], density sensitive hashing (DSH) [49], and spherical hashing (SpH) [50] from 4096D deep features obtained from FC7 layer of a pre-trained AlexNet CNN. We chose AlexNet because the SqueezeNet model we used also offers similar performance to the AlexNet architecture. Our method generated hash codes using the CNN pipeline through a bottleneck layer with limited neurons just before the final FC layer. Comparison of the precision recall scores, shown in Fig. 10, suggests the superiority of our method against other competing approaches. The proposed method significantly outperformed other methods in 64, and 128. At 256 bits, SpH achieved higher precision at low recall, whereas our method performed the best at other recall settings. Table 4 list a comparison of the proposed approach with other state-of-the-art methods in terms of features extraction and image retrieval time. For mobile platform, it is necessary to minimize the features extraction time as well as retrieval time. We achieved optimal features extraction time through the use of a compute-friendly architecture (SqueezeNet). The hash codes generated through the network allowed us to retrieve images without needing us to compute distances between the query code and all other codes in the database. By directly accessing the neighboring region in the hamming space, we

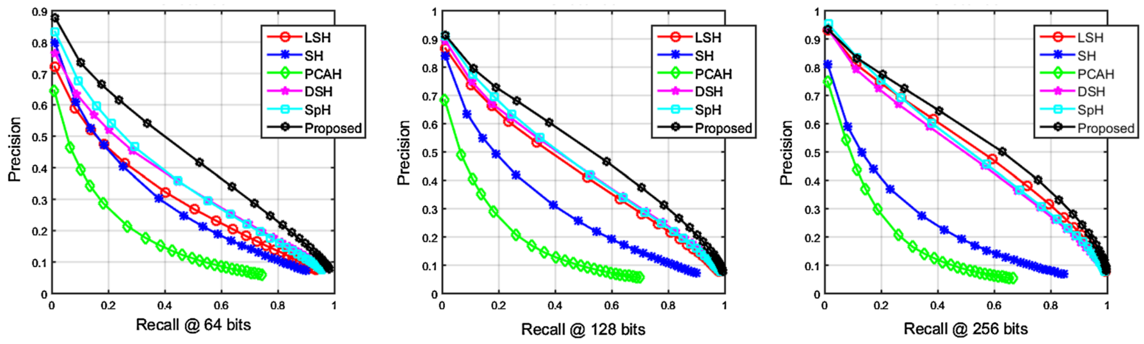


Fig. 10 Retrieval performance comparison with other hash-based image retrieval methods

Table 4 Efficiency analysis of various SBIR approaches

Methods	Feature length	Model size (MB)	Features extraction time (ms)	Image retrieval time (s) (100 k images)
Eitz et al. [10]	784	–	12.0	2.9
Deep CNN [44]	4096	180.0	47.3	13.3
Seddati et al. [26]	4096	155.6	37.6	15.6
DeepSketch [28]	4096	215.8	43.1	12.8
Proposed method	256 bits	93.2	27.0	0.2

retrieved the nearest neighbors and ranked them according to their hamming distances with the query hash code.

5 Conclusions and future work

In this paper, we presented an efficient method for sketch-based image retrieval on mobile devices. Unlike traditional SBIR systems, where the retrieval systems attempt to search images using colorless and shade-less sketches. It was believed that using mouse and keyboard as interfaces, adding colors or textures would be too cumbersome for users. However, with the popularity of touch screen devices such as mobile phones and tablets, sketching and drawing have become very intuitive and easy. Keeping in view this convenience in sketching, we proposed to perform image retrieval on mobile devices using partially shaded sketches. An efficient deep CNN has been fine-tuned on sketch oriented dataset consisting of hand-drawn sketches, full color images and its augmented versions. The augmented images consisted of fine and coarse edge maps, de-colorized and de-texturized images, and flipped/rotated images of de-colorized and de-texturized images. Through data augmentation, we attempted to allow CNN to model contents from both sketches and images to enable the features be used for SBIR using partially shaded sketches. Through extensive experiments, we have shown that adding shades or colors to hand-drawn sketches greatly improve retrieval performance. Furthermore, the proposed framework generates a short binary code for sketch which allows highly efficient retrieval performance for mobile devices. In future, we plan to enhance the data augmentation procedures, as well as the features extraction process to improve SBIR.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) Grant Funded by the Korea Government (MSIP) (No. 2016R1A2B4011712).

References

1. Ahmad, J., Muhammad, K., Lloret, J., Baik, S.W.: Efficient conversion of deep features to compact binary codes using fourier decomposition for multimedia big Data IEEE Trans. Ind. Inf. PP, 1–1 (2018)
2. Wang, S., Zhang, J., Han, T.X., Miao, Z.: Sketch-based image retrieval through hypothesis-driven object boundary selection with hlr descriptor. IEEE Trans. Multimed **17**, 1045–1057 (2015)
3. Ahmad, J., Sajjad, M., Rho, S., Baik, S.W.: Multi-scale local structure patterns histogram for describing visual contents in social image retrieval systems. Multimed. Tools Appl. **75**, 12669–12692 (2016)
4. Ahmad, J., Sajjad, M., Mehmood, I., Rho, S., Baik, S.W.: Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems J. Real-Time Image Proc. 13, 431–447 (2017)
5. Hu, R., Collomosse, J.: A performance evaluation of gradient field hog descriptor for sketch based image retrieval. Comput. Vis. Image Underst. **117**, 790–806 (2013)
6. Kim, S., Guy, S.J., Hillesland, K., Zafar, B., Gutub, A.A.-A., Manocha, D.: Velocity-based modeling of physical interactions in dense crowds. Vis. Comput. **31**, 541–555 (2015)
7. Tseng, K.-Y., Lin, Y.-L., Chen, Y.-H., Hsu, W.H.: Sketch-based image retrieval on mobile devices using compact hash bits. In Proceedings of the 20th ACM International Conference on Multimedia, pp. 913–916 (2012)
8. Al-Otaibi, N.A., Gutub, A.A.: 2-Leyer security system for hiding sensitive text data on personal computers. Lect. Notes Inf. Theory **2**(2) (2014)
9. Abdelgawad, H., Shalaby, A., Abdulhai, B., Gutub, A.A.A.: Microscopic modeling of large-scale pedestrian–vehicle conflicts in the city of Madinah, Saudi Arabia. J. Adv. Transp. **48**, 507–525 (2014)
10. Eitz, M., Hays, J., Alexa, M.: How do humans sketch objects? ACM Trans. Graph. **31**, 44:1–44:10 (2012)
11. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248–255 (2009)
12. Ahmad, J., Muhammad, K., Baik, S.W.: Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features. J. Med. Syst. **42**, 24 (2017)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **1**, 1097–1105 (2012)
14. Ahmad, J., Muhammad, K., Bakshi, S., Baik, S.W.: Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets. Fut. Gen. Comput. Syst. **81**, 314–330 (2018)
15. Ahmad, J., Sajjad, M., Mehmood, I., Baik, S.W.: SiNC: saliency-injected neural codes for representation and efficient retrieval of medical radiographs. PLoS One 12, e0181707 (2017)
16. Qi, Y., Song, Y.-Z., Zhang, H., Liu, J.: Sketch-based image retrieval via Siamese convolutional neural network. In Image Processing (ICIP), 2016 IEEE International Conference on, pp. 2460–2464 (2016)
17. Krizhevsky, A., Hinton, G.E.: Using very deep autoencoders for content-based image retrieval. In: Proceedings of the 19th European Symposium on Artificial Neural Networks, Bruges, Belgium (2011)
18. Babenko, A., Slesarev, A., Chigorin, A., Lempitsky, V.: Neural codes for image retrieval. In: Computer Vision–European Conference on Computer Vision (ECCV), Springer, pp. 584–599 (2014)
19. Gutub, A., Alharthi, N.: Improving Hajj and Umrah Services Utilizing Exploratory Data Visualization Techniques, presented at the Hajj Forum. Umm Al-Qura University–King Abdulaziz Historical Hall, Makkah (2016)
20. Gutub, A.: Exploratory data visualization for smart systems. Smart cities 2015-3rd annual digital grids and smart cities workshop, Burj Rafal Hotel Kempinski, Riyadh (2015)
21. Gutub, A.: Social media and its impact on e-Governance. ME smart cities 2015-4th middle east smart cities summit, 8–9 Dec, Dubai (2015)
22. Chen, T., Cheng, M.-M., Tan, P., Shamir, A., Hu, S.-M.: Sketch2photo: Internet image montage. ACM Trans. Graph. (TOG) **28**, 124 (2009)
23. Fu, L., Kara, L.B.: Recognizing network-like hand-drawn sketches: a convolutional neural network approach. In ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 671–681 (2009)

24. Sarvadevabhatla, R.K., Babu, R.V.: Freehand sketch recognition using deep features. arXiv:1502.00254 (2015)
25. Al-Otaibi, N.A., Gutub, A.A.: Flexible stego-system for hiding text in images of personal computers based on user security priority. In: Proceedings of: 2014 International Conference on Advanced Engineering Technologies (AET-2014), pp. 250–256 (2014)
26. Seddati, O., Dupont, S., Mahmoudi, S.: Deepsketch: deep convolutional neural networks for sketch recognition and similarity search. In: Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on, pp. 1–6 (2015)
27. Liu, L., Shen, F., Shen, Y., Liu, X., Shao, L.: Deep sketch hashing: fast free-hand sketch-based image retrieval. arXiv:1703.05605 (2017)
28. Wang, X., Duan, X., Bai, X.: Deep sketch feature for cross-domain image retrieval. Neurocomputing 207, 387–397 (2016)
29. Ahmad, J., Mehmood, I., Baik, S.W.: Efficient object-based surveillance image search using spatial pooling of convolutional features. J. Vis. Commun. Image Rep. **45**, 62–76 (2017)
30. Iandola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 1 MB model size. arXiv:1602.07360 (2016)
31. Ahmad, J., Muhammad, K., Baik, S.W.: Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search. PLoS One 12, e0183838 (2017)
32. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 629–639 (1990)
33. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014)
34. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
35. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv:1312.4400 (2013)
36. Gschwend, D.: Zynqnet: an fpga-accelerated embedded convolutional neural network. MS thesis, Swiss Federal Institute of Technology Zurich (ETH-Zurich) (2016)
37. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications arXiv:1704.04861 (2017)
38. Han, S., Mao, H., Dally, W.J.: Deep compression: Compressing deep neural network with pruning, trained quantization and huffman coding. CoRR 2, abs/1510.00149 (2015)
39. MathWorks (2015) MATLAB. Available: http://www.mathworks.com/products/parallel-computing/
40. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., et al.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia, pp. 675–678 (2014)
41. Nvidia DIGITS: Available: https://developer.nvidia.com/digits. (2016)
42. Caltech-256 Object Category Dataset: Available: http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001
43. Çalışır, F., Baştan, M., Ulusoy, Ö., Güdükbay, U.: Mobile multi-view object image search. Multimed. Tool Appl. **76**, 12433–12456 (2017)
44. Yang, Y., Hospedales, T.M.: Deep neural networks for sketch recognition. arXiv:1501.07873 (2015)
45. Lin, K., Yang, H.-F., Hsiao, J.-H., Chen, C.-S., Deep learning of binary hash codes for fast image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 27–35 (2015)
46. Gionis, A., Indyk, P., Motwani, R.: Similarity search in high dimensions via hashing. VLDB, pp. 518–529 (1999)
47. Datar, M., Immorlica, N., Indyk, P., Mirrokni, V.S.: Locality-sensitive hashing scheme based on p-stable distributions. In: Proceedings of the Twentieth Annual Symposium on Computational Geometry, pp. 253–262 (2004)
48. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. Advances in neural information processing systems, pp. 1753–1760 (2009)
49. Jin, Z., Li, C., Lin, Y., Cai, D.: Density sensitive hashing. IEEE Trans. Cybern. **44**, 1362–1371 (2014)
50. Heo, J.-P., Lee, Y., He, J., Chang, S.-F., Yoon, S.-E.: Spherical hashing. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2957–2964 (2012)

Jamil Ahmad He received his BCS degree in Computer Science from the University of Peshawar, Pakistan in 2008 with distinction. He received his Master's degree in 2014 with specialization in Image Processing from Islamia College, Peshawar, Pakistan and the PhD degree from Sejong University, Seoul, South Korea in 2018. He is also a regular faculty member in the Department of Computer Science, Islamia College Peshawar. His research interests include deep learning, medical image analysis, content-based multimedia retrieval, and computer vision. He has published several journal articles in these areas in reputed journals including Journal of Real-Time Image Processing, Multimedia Tools and Applications, Journal of Visual Communication and Image Representation, PLoS One, Computers and Electrical Engineering, SpringerPlus, Journal of Sensors, and KSII Transactions on Internet and Information Systems. He is also an active reviewer for IET Image Processing, Engineering Applications of Artificial Intelligence, KSII Transactions on Internet and Information Systems, Multimedia Tools and Applications, and IEEE Transactions on Cybernetics.

Khan Muhammad He received his B.C.S. degree in computer science from Islamia College, Peshawar, Pakistan with research in information security. Currently, he is pursuing MS leading to Ph.D. degree in digitals contents from College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea. He is working as a researcher at Intelligent Media Laboratory (IM Lab) since 2015 under the supervision of Prof. Sung Wook Baik. His research interests include image and video processing, data hiding, image and video steganography, video summarization, diagnostic hysteroscopy, wireless capsule endoscopy, and CCTV video analysis. He has published 15+ papers in peer-reviewed international journals and conferences such as Journal of Medical Systems, Biomedical Signal Processing and Control, IEEE Access, Multimedia Tools and Applications, SpringerPlus, KSII Transactions on Internet and Information Systems, Journal of Korean Institute of Next Generation Computing, NED University Journal of Research, Technical Journal, Sindh University Research Journal, Middle-East Journal of Scientific Research, MITA 2015, PlatCon 2016, and FIT 2016. He is a student member of the IEEE.

Syed Inayat Ali Shah He received his MSc and MPhil degrees in Mathematics from Quaid-e-Azam University, Islamabad in 1987 and 1990, respectively. He received the PhD degree from Saga University Japan in 2002. He is currently serving as a Professor and Dean in the Department of Mathematics, Islamia College Peshawar, Pakistan. He has published more than 30 research articles in reputed journals. His research interests include Algebraic number theory, applied mathematics, and applications of mathematical techniques in image processing and computer vision.

Arun Kumar Sangaiah He has received his Master of Engineering (M.E.) degree in Computer Science and Engineering from the Government College of Engineering, Tirunelveli, Anna University, India. He had received his Doctor of Philosophy (Ph.D.) degree in Computer Science and Engineering from the VIT University, Vellore, India. He is presently working as an associate professor in School of Computer Science and Engineering, VIT University, India. His area of interest includes software engineering, computational intelligence, wireless networks, bioinformatics, and embedded systems. He has authored more than 100 publications in different journals and conference of national and international repute. His current research work includes global software development, wireless ad hoc and sensor networks, machine learning, cognitive networks and advances in mobile computing and communications. He is an active member in Compute Society of India. Moreover, he has carried out a number of funded research projects for Indian government agencies. Also, he was registered a one Indian patent in the area of Computational Intelligence. Besides, he is responsible for Editorial Board Member/Associate Editor of various international journals like International Journal of Intelligent Information Technologies (IGI), International Journal of Cloud Applications and Computing (IGI), International Journal of High Performance System (Inderscience), International Journal of Image Mining (Inderscience), International Journal of Intelligent Engineering and Systems, International Journal of Computational Systems Engineering (Inderscience) and Institute of Integrative Omics and Applied Biotechnology (IIOAB). In addition, he has edited a number of guest editorial special issues for various journals like Future Generation Computer Systems (SCI), Neural Network World (SCI), Intelligent Automation & Soft Computing (SCI), and Scientific World Journal (SCI). Also, he has organized a number of special issues for Elsevier, Inderscience, Springer, Hindawi, and IGI publishers. Also, he has acted as a book volume editor of various publishers for Taylor andFrancis, Springer, IGI, etc. Furthermore, he made outstanding efforts and contributions on the technical programme committee member of various reputed international/national conferences.

Sung Wook Baik He received the B.S degree in computer science from Seoul National University, Seoul, Korea, in 1987, the M.S. degree in computer science from Northern Illinois University, Dekalb, in 1992, and the Ph.D. degree in information technology engineering from George Mason University, Fairfax, VA, in 1999. He worked at Datamat Systems Research Inc. as a senior scientist of the Intelligent Systems Group from 1997 to 2002. In 2002, he joined the faculty of the College of Electronics and Information Engineering, Sejong University, Seoul, Korea, where he is currently a Full Professor and Dean of Digital Contents. He is also the head of Intelligent Media Laboratory (IM Lab) at Sejong University. His research interests include computer vision, multimedia, pattern recognition, machine learning, data mining, virtual reality, and computer games. He is a member of the IEEE. He has chaired and organized several international conferences and is also an active reviewer for many reputed journals including IEEE Transactions on Cybernetics, IEEE Access, Journal of Visual Communications and Image Representation.