SKETCH BASED IMAGE RETRIEVAL (SBIR)

RUSHIKESH KALE, SOURAB MARPHALE, YASH OZA

Department of Electronics & Telecommunication Engineering,
G H Raisoni College of Engineering & Management, Pune.

Abstract : ***In this project, we look into the dilemma of zeroshot sketch-based image retrieval (ZS-SBIR), which involves using human sketches as queries to retrieve images from previously unseen categories. We make a significant contribution to prior art by proposing a ZS-SBIR scenario that represents a major advancement in its practical implementation. To exploit the domain void, highly abstract amateur human sketches are intentionally sourced rather than those found in existing datasets, which are mostly semi-photorealistic. The ZS-SBIR system is then used to model sketches and images into a single embedding space. To alleviate the domain gap, a novel strategy for mutual knowledge between domains has been created. To aid semantic transition, external semantic awareness is further embedded. Surprisingly, retrieval efficiency outperforms on existing datasets with a simplified version of our model. On the newly proposed dataset, we compare our complete model to a variety of alternatives and show that it outperforms them.***

Keywords: Convolutional Neural Network, Deep Learning, Image Retrieval, Sketch Based Image Retrieval, zero shot.

I. INTRODUCTION

The surge in internet users, along with expanded computing space, improved internet access, and higher bandwidths, has resulted in an exponential increase in multimedia content on the Web. Image advertising, in particular, has become ubiquitous and is increasingly used to engage followers on social media and consumers on numerous e-commerce platforms. Users' knowledge needs and browsing habits have changed as a result of the increase in image content. Users are increasingly searching for photographs (rather than documents) by offering either a textual summary of the image or another image that is identical to the requested image. The former is referred to as text-based image retrieval, while the latter is referred to as content-based image retrieval.

The paper suggests a SBIR scheme that requires users to draw a sketch, after which the system finds a matching image from the data collection. The key advantage of sketch-based image retrieval over text-based retrieval is that it is easier to articulate the necessary image's orientation and pose in the query sketch rather than defining these characteristics in the query text. The biggest difficulty with a SBIR system is that it needs knowledge of both the sketch and image domains before comparing them. Traditional methods have relied on hand-drawn features that use gradients or edges as features that are inherently invariant across image and sketch contexts, but these techniques can be significantly improved. Image detection activities in both the image and sketch domains have improved since the introduction of deep convolutional networks. As a result, the purpose of this approach is to train a deep convolutional network to learn the cross-domain representation for sketches and photographs, as well as to retrieve images with the same pose and orientation as the query sketch.

II. PROBLEM DEFINITION

Given a sketch query from a previously unseen category, the goal of this approach is to retrieve semantically meaningful images from a given database. The knowledge-gap between the seen and unseen categories along with sketch-image domain shift makes this an extremely challenging problem. Hence deep learning based approach will be effective to retrieve the images based on the provided sketch accurately.

III. LITERATURE SURVEY

SBIR is a useful technique for querying massive image databases. The aim of all of the studies is to figure out how to bridge the difference between sketch and image matching. There are several approaches to this distance that have been explored or found. We recently evaluated an approach that addresses the three key points of sketch-based image retrieval. This section discusses the relevant literature and explains how an image retrieval system works. The scientific articles used in the literature review were selected to have an adequate context for solving the research sub-problems. Despite the fact that a large number of scholars have worked on content-based image retrieval using an example image as a query, some related works on sketch-based image retrieval have only been published in the last three years.

In this regard, the work of Eitz et al. [2] is one of the most significant. They suggest two methods for extracting relevant knowledge from sketch representations based on the well-known SIFT and Shape Context approaches. The derived data is clustered into feature vectors to form a codebook, which is then included in the Bag of Features (BoF) approach. The authors propose the first systematically constructed benchmark for the SBIR problem, which makes this work much more important. This benchmark has the unique feature of taking into account the user's perception of the similarities between sketches and test photographs. This is important since the overarching purpose of a retrieval system is to meet the needs of the customer. Eitz, et aldataset's contains 31 drawings, each with 40 images ranked by similarity. In addition, they create new descriptors based on the bag-of-features approach.

Hu et al. [3] suggested a different solution. This thesis, like Eitz's, is built on the Bag of Feature approach, but it approaches the extraction information stage in a somewhat different way. The transformation of sketch representations into gradient field (GF) images is a novel concept in this work. Edge maps are used to transform the test images into sketch representations using the Canny operator. The HOG descriptors are then computed in three different scales with respect to each edge pixel using the GF images. A BoF technique is then used to construct a frequency histogram. This approach necessitates the solution of a sparse linear equation scheme with the number of variables equal to the order of the underlying image's dimension. The most of SBIR methods use HoG to compute either a global or local representation.

HOG appears to be the most common descriptor in the group, according to Rui Hua and John Collomossea [4]. HOG descriptors, on the other hand, may be sparse due to the sparseness of drawings, which may have a negative influence on the final efficacy. In [3,] an exhaustive assessment of many image descriptors is performed, showing that HOG-like characteristics are superior in SBIR. Two SBIR databases, Flickr160 and Flickr15k, were also made public by the publishers.

Chalechale et al. [5] used angular partition to remove compact and efficient features in the spatial domain of photos. The first step in the process is to obtain the edge map of all natural images so that they can be converted into a format that can compete with binary sketches. To best fit the edge maps derived from the photographs, queries were preprocessed with a morphological thinning filter. An picture is partitioned into K angular regions using an angular partition. To obtain hierarchical coarse to fine representations, K can be modified. Each slice function is represented by the number of edge points for each area Ri, I = 1, 2,..., K. Size and rotation invariance are ensured by computing a 1-D Discrete Fourier Transform (DFT) for each area of the image and retaining only the DFT magnitude. The authors also point out that this system is resistant to translation. The l1 distance between the two function vectors is used to determine how similar photographs and drawings are. A database of 3,600 photographs was used to validate this method. In order to obtain the right file, at least 13% of the images had to be retrieved from the database, highlighting the noise sensitivity of the DFT.

Jun Guo et al. present their work on developing optimal use of edges through perceptual grouping in CVPR 2015. The data from that paper is used to compare the success of our scheme.

Dipika R. Birari and J.V. Shinde [7] suggested an SBIR scheme based on constraints with a descriptor. Edge extraction, descriptor architecture, and edge selection are all part of their scheme. This paper's method is closely similar to the paper's. The edge matching algorithm is the difference. Our system's performance is compared to theirs. For edge extraction of both query and database images, the proposed SBIR method in this paper uses the Canny edge detection algorithm. This work develops a block histogram matching by sliding window system for effective sketch and image matching. Image Retrieval System is a software programme that allows you to search for images. Information retrieval systems have been a hot topic in computer science over the past few years.. The major subject of information retrieval research is multimedia information retrieval, and in particular, image retrieval systems, which has drawn a large number of researchers from various communities such as computer vision, multimedia retrieval, and data mining, among others, resulting in a large number of related publications. The fact that photographs are abundant and convenient to collect by users is the main reason why image retrieval is the subject of so much study.

IV. BLOCK DIAGRAM

The block diagram of the proposed Sketch Based Image Retrieval System is as shown in Fig.1.

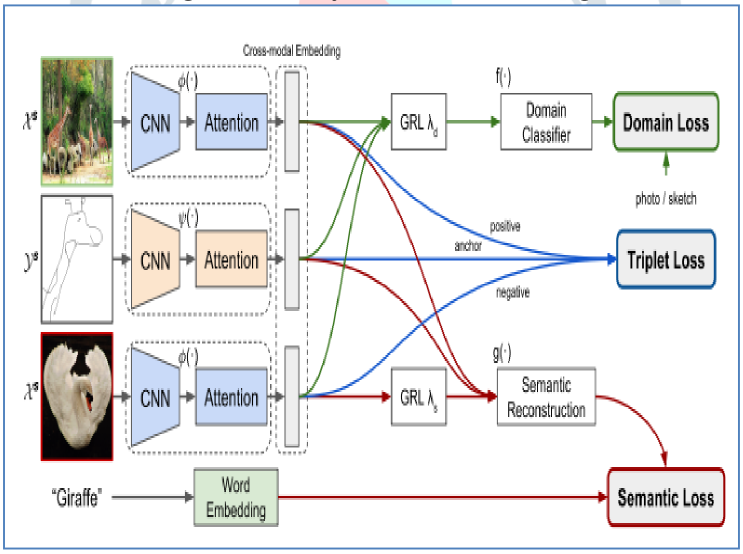

Fig. 1. Block diagram of the proposed system

A. Sketchy-Extended Dataset

Originally intended as a fine-grained retrieval link between sketches and individual images. This dataset was created specifically for the ZS-SBIR mission. In the one hand, Shen suggested using 25 random classes as a research range, with the remaining 100 classes being used for instruction. As a result, participants tended to draw the models realistically, resulting in drawings that closely resembled true edge-map drawings. This effectively bridges the distance between sketch and image domains.

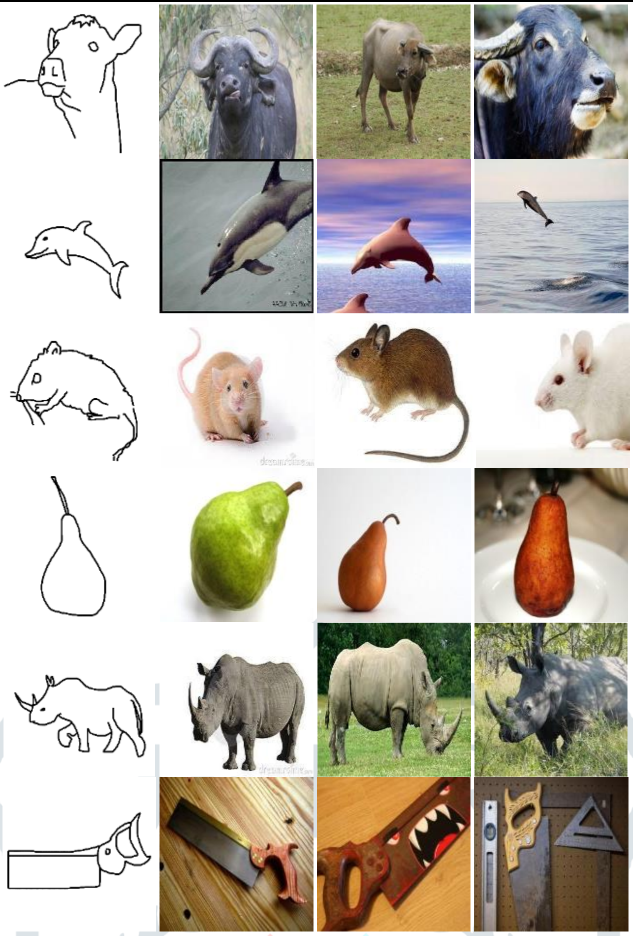

Fig. 2 Database image samples

B. Problem Formulation

Let C represent the list of all conceivable classes in a dataset, and $X = \{x_i\}_{i=1}^{N}$ and $Y = \{y_i\}_{i=1}^{M}$, respectively, the set of photographs and sketches. $I_x: X \rightarrow C$ and $I_y: Y \rightarrow C$ be are two picture and sketch labelling keys, respectively. An optimum ranking of gallery images can be achieved by including a feedback sketch. Training and trial sets are split into seen $C^s \subset C$ and unseen $C^\mu \subset C$ types in a zero-shot system, where $C^s \cap C^\mu = \phi$. The model must learn an aligned space between drawings and photographs in order to do so on test data for classes that have never been used in training.

There are two major components of the proposed structure. The encoder transforms the input image to the appropriate embedding space. The second part is the cost function, which guides the learning process to give the embedding the desired properties. Figure 1 depicts the proposed technique.

C. Encoder Networks

The embedding functions φ(·) and ψ(·) are two CNNs constructed with caution, with the last fully-connected layer replaced to fit the desired embedding size D. In all modalities, the focus process aids our method in locating the relevant features. Soft emphasis is the most general, and since it is differentiable, it can be trained end-to-end like the rest of the network. Our soft-attention model learns an attention mask from a function map that assigns various weights to different regions of an image. These weights are used to emphasize essential features, so $f + f \cdot att.$ is the output of the attention module when an attention mask att and a function map f are given. To compute the attention mask, 1×1 convolution layers are applied to the corresponding feature diagram.

D. Learning objectives

The suggested framework's learning goal combines: (i) Triplet Loss; (ii) Domain Loss; and (iii) Semantic Loss. The encoder network receives visual and semantic input from these objective functions.

V. ALGORITHM

Algorithm: Sketch based image retrieval training algorithm

Input: Sketch data $\{X, Y\}$; Class semantics S; $\lambda s = 0.5$ and maximum iterations T
Output: Encoder networks parameters $\{\Theta\varphi, \Theta\psi\}$.

1: repeat
2: Get a random minibatch $\{y_i, x\,p_i, xn_i, s_i\}$ NB i=1; where ,y_i, $x\,p_i$ belong to the same class and x n_i does not.
3: $\lambda_d \leftarrow \text{clip}(z_i(\cdot), \min = 0, \max = 1)$
4: $L \leftarrow$ Eq. 4
5: $\Theta \leftarrow \Theta - \Gamma(\nabla\Theta L)$
6: **until** max training iterations T

VI. RESULT

The results of the system are presented in this section. The sketch is taken as an input to the system. The sketch images is tested with the trained model. Sketchy dataset is used for training as well as testing. The input sketchy image will be tested with the trained model and it produces the output as most relevant images from the dataset. The qualitative analysis of the proposed system is as shown in Fig. 3.

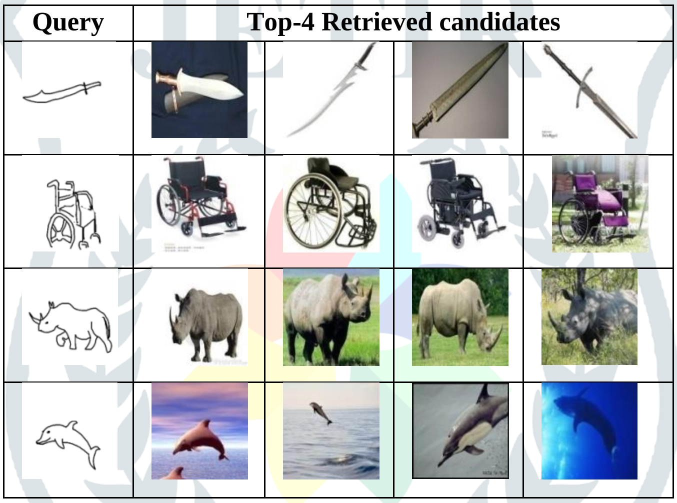

Fig. 3. Qualitative analysis of the system

The input query image is shown in first column of the Fig. 3 and its top four retrieved images are shown in the second column of the Fig. 3. The qualitative analysis of the proposes system shows the satisfactory results.

VII. CONCLUSION

This paper presents the deep learning approach for sketch based image retrieval system. The performance of the trained network is evaluated on "Sketchy" dataset. The performance of the system is presented in qualitative analysis and it is observed that the system is able to retrieved the relevant image.

REFERENCES

[1]. John Eakins, M.G.: Content-based image retrieval, University of Northumbria at Newcastle.
[2]. Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors". IEEE Transactions on Visualization and Computer Graphics, 2011.
[3]. Rui Hu, M. Barnard, and J. Collomosse. "Gradient field descriptor for sketch based retrieval and localization". In 17th IEEE International Conference on Image Processing (ICIP), 2010.
[4]. Rui Hua and John Collomossea, "A Performance Evaluation of Gradient Field HOG Descriptor for Sketch Based Image Retrieval", Centre for Vision, Speech and Signal Processing, University of Surrey, UK.
[5]. A. Chalechale, G. Naghdy, and A. Mertins. "Sketchbased image matching using angular partitioning". IEEE Transactions on Systems, Man and Cybernetics, 2005.
[6]. Yonggang Qi, Yi-Zhe Song, Tao Xiang, Honggang Zhang, Timothy M. Hospedales, Yi Li, and Jun Guo, "Making better use of edges via perceptual grouping," in CVPR, 2015.
[7]. Dipika R. Birari, J. V. Shinde, "A Sketch based Image Retrieval with Descriptor based on Constraints," International Journal of Computer Applications, Volume 146 – No.12, July 2016.