International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 6, December 2023, pp. 6293~6301
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i6.pp6293-6301

Novel hybrid generative adversarial network for synthesizing image from sketch

Pavithra Narasimha Murthy[1], Sharath Kumar Yeliyur Hanumanthaiah[2]
[1]Maharaja Institute of Technology Mysuru, Visvesvaraya Technological University, Belagavi, India
[2]Department of Information Science and Engineering, Maharaja Institute of Technology,
Mysuru Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received Feb 6, 2023
Revised May 31, 2023
Accepted Jun 4, 2023

Keywords:

Deep learning
Generative adversarial network
Hybrid model
Image synthesis
Sketch-based image retrieval

ABSTRACT

In the area of sketch-based image retrieval process, there is a potential difference between retrieving the match images from defined dataset and constructing the synthesized image. The former process is quite easier while the latter process requires more faster, accurate, and intellectual decision making by the processor. After reviewing open-end research problems from existing approaches, the proposed scheme introduces a computational framework of hybrid generative adversarial network (GAN) as a solution to address the identified research problem. The model takes the input of query image which is processed by generator module running 3 different deep learning modes of ResNet, MobileNet, and U-Net. The discriminator module processes the input of real images as well as output from generator. With a novel interactive communication between generator and discriminator, the proposed model offers optimal retrieval performance along with an inclusion of optimizer. The study outcome shows significant performance improvement.

Corresponding Author:

Pavithra Narasimha Murthy
Maharaja Institute of Technology, Visvesvaraya Technological University
Belagavi, India
Email: pavithra.apr02@gmail.com

1. INTRODUCTION

Search engines and tools has become one of the efficient ways to look for various forms of information both in standardized as well as in customized manner [1]. With constraints towards text-based search, image-based search option is one of the simplest and best option to look for information [2]. However, image-based search can be further more customized when the query image is given in the form of free hand drawing also known as sketch [3]. In this form of search technique, a sketch related to certain real object is drawn by user and it acts as an input to the system which looks for all matching images respective to the sketch. In this perspective, the relevance factor between the query images and real images in other sources are matched by the user using various shape-based information [4]. In such a process, generally the content-based information is not much prioritized, and this is one of the prime reasons for multiple outliers in the outcome of related images. The application of sketch-based image retrieval system is not merely about matching and extracting the related images, but the complexity is to generate the synthesized or reconstructed image by the machine [5]. At present, there are multiple approaches and techniques applied for this purpose with claimed beneficial characteristics. From the conventional scheme of information retrieval system, it is known that there are basically three types of standard approaches viz. i) text-based, ii) content-based, and iii) sketch based [6]. The existing research work towards this domain of problem is mainly emphasized towards improving various internal operations which has a direct impact towards retrieval process, and it is outcome. From this perspective, enough significance is given towards feature-based processing using different schemes viz. edge detection, patch hashing, histogram of oriented gradient, gradient field, and bag of visual words. [7]. Further, usage of descriptors e.g., edge histogram descriptor and tensor descriptor are also used in this process. It was also noticed that the core complexity in this retrieval process is to offer a high-end accuracy, so that retrieved image bears higher degree of equivalency with the input of hand-drawn sketch. This is the core issue which can be effectively addressed by using machine learning or deep learning techniques [8]. However, there are various rising concern in this area of investigation viz. i) there is a need of an efficient technique which can offer cost effective computational model as involvement of deep learning offers accuracy but at the cost of computational burden; ii) identifying the efficient and accurate features from the dynamic form of strokes involved in free hand drawing is quite a computationally burden task, which still many existing system are trying to investigate; iii) an efficient model which can reconstruct the highly relevant method is requite to adhere with various conditional logic as well as multi-objective function, which are less seen in existing model. Out of all the techniques, generative adversarial network (GAN) is found to offer many beneficial aspects towards the process of retrieval. With its capability to generate image which is nearly similar to source query image, this approach can perform deeper investigation of an image which becomes as a complementary advantage for further machine learning task [9]. With various upcoming studies on generative adversarial models on image processing being reviewed, the prime motivation of proposed scheme is to harness its advantageous features along with improvising the learning model for better retrieval performance. The proposed solution uses a hybrid generative adversarial model to address the problem.

Various existing research methodologies have been reviewed to understand existing schemes. One of the significant problems in image synthesis from sketches is associated with inadequate samples to take up efficient decision for image reconstruction. Study in such direction has been carried out by Chaudhuri *et al.* [10] who has presented a retrieval of intermodal data using zero-shot concept. The study model uses loss function projection as well as cross triplet in order to enhance the overlapping among domains. Similar concept of zero-shot has also been used by Deng *et al.* [11] where a semantic network of cross-modal is deployed. According to this concept, the retrieval is carried out considering aligned features of sketch and images with respect to a common space. The implementation of zero-shot concept was also reported in work of Dutta *et al.* [12] where a unique generation of a counterfeited images are carried out. The performance of retrieval is enhanced by selecting only the images belonging to unknown classes with the query image. Further problems associated with degradation of quality of retrieved image due to challenges involved in understanding features from free hand drawing is reported to be addressed in work of Choi *et al.* [13]. According to this model, a prediction system is designed for forecasting the consecutive strokes of sketches from massive dataset using deep learning. Segmentation also plays a significant role in the current domain of investigation. Such segmentation-based scheme is implemented by Hu *et al.* [14] where the neural network is used toward this purpose by synthesizing the weights. All the categories are generalized by the model and are applicable for both instance-based and category-based processing. Work carried out by Jiao *et al.* [15] have used manifold ranking system for establishing a correspondence system of cross domains. According to this work, features are extracted from sketches and convolutional neural network (CNN) model is built in two parts followed by generation of undirect graphs. For this purpose, semantic labels as well as learning features have been used between three-dimensional model and sketches. Adoption of CNN was also reported in work of Qi *et al.* [16] where a transfer learning is utilized in order to improve feature extraction performance. The system also constructs a training dataset of personalized model using neural network. Lei *et al.* [17] have implemented an integrated embedding network in order to learn semi-heterogeneous features belonging to cross domains. The study further implements a co-attention model using common features of edge map and natural image. It has also been observed that there are higher dependencies on an appropriate edge map, which is not applicable for real world scenario. This problem is addressed in work of Li *et al.* [18] where a GAN is designed for representation of extracted classes. The implemented architecture assists in harnessing the potential of all unpaired images followed by performing more accurate form of learning.

Setumin *et al.* [19] have addressed the problem of shape exaggeration of generalized sketch where a Gradient histogram on the basis of Gaussian value is designed for managing dynamic local features. A block-based operation is carried out for feature management using histogram of oriented gradient while the canonical correlation analysis is adopted to increase the correlation. The performance issues towards applying deep learning when applied for corrupted sketch image is addressed in work of Wan *et al.* [20]. The solution to this problem is via constructing a restoration network using feedback in order to identify the corrupted area of sketches. The technique is claimed to offer balance between cost of memory and higher quality of images. The methodology of clustering is also found to be utilized for image retrieval schemes as noticed in work of Wang *et al.* [21]. According to this work, a multi-clustering approach with re-ranking method is applied using semantic information based on natural image, object images and edge map. The final retrieval values depend upon the clustering being carried out on the above three sets of information. Similar author has also improvised the similar re-ranking mechanism using CNN [22]. An unique method is presented by Weng and Zhu [23] in order to optimize the adoption of supervised information for better accomplishment of similarity scores. According to this implementation, a hashing scheme of supervised nature is designed where the supervised information of data is constructed semantically. The study outcome claims of lower storage and time complexity while adopting this hashing scheme. The work of Xu *et al.* [24] addressed the problem of processing large scale target images by presenting a deep learning technique using adversarial training for embedding the spaces. An interesting study is presented by Xu *et al.* [25] where sketches are retrieved from video files by solving retrieval issue of cross-modal origin. A deep learning network is constructed with annotated dataset and it can perform efficiently in presence of both weakly supervised network as well as strongly supervised network. Xu *et al.* [26] have used combination of CNN and recurrent neural network (RNN) for encoding both temporal and spatial information of strokes of sketches. Further, deep learning hashing is used for retrieving the sketches as well as zero-shot concept is used for deep embedding operation. Retrieval of three-dimensional shapes is reported in work of Xu *et al.* [27] where an algorithm for view selected is constructed. Further triplet loss is developed for network metric training. The dependency problem of edge information towards retrieval system is addressed by model of Yang *et al.* [28] which is capable of processing dark segments and edge regions using GAN. The problems associated with domain adaptation is addressed in work of Zheng *et al.* [29] where edge map has been used for approximating sketches while a transformation is applied to reduce the semantic gaps.

After reviewing the existing methodologies adopted for image synthesis from sketches, it is noticed that there are considerable number of beneficial outcomes. It is also noticed that the existing system has offered varied ranges of solutions towards addressing wide ranges of research problems in depth. Developed over a controlled research environment, the existing schemes have witnessed some of the significant contributions mainly in form of accuracies obtained as well as novelty in the methodologies being used. However, a closer insight into existing scheme is found to possess two perspectives: i) usage of similar methodologies with common properties without many changes and ii) higher emphasis towards accuracy in its generated synthesized images. Hence, the outlines of the research problems identified are as follows:

a. Frequent adoption of zero-shot method: It is noticed that zero-shot based concept is dominantly adopted in existing system. One of the advantages of this technique is its capability to investigate the classes which are usually not seen during training operation. Although, this concept offers reduction of dependencies towards labelling system; however, they also witness various computational issues. These issues are mainly based on quality of the data/features being obtained from the dynamic and heterogeneous query image. At present, there is no reported study to formalize this problem using any learning approach in zero-shot concept [30].
b. Usage of deep learning: There is no denying the fact that deep learning and its numerous variants have proven effective towards retrieval system. A deeper insight to existing methodologies shows direct usage of deep learning without further amendments in it. Such models do have problems associated with class imbalance, surplus gradient, overfitting, sophisticated training operation, consumption of longer duration of training. Although, these studies have achieved better accuracy at the end as well as they have also offered solution towards solving one specific problem out of all the above-mentioned set of problems, but still such problems remain a higher concern for hybridizing deep learning method.
c. Inadequate focus on operation cost: All the existing studies discussed in prior section perform three set of work operation i.e., taking direct input of images, applying methods, and fine-tuning to achieve higher accuracy. However, there is not much refinement of the last stage of operation associated with fine-tuning. This results in inclusion of higher number of processing which elevates the cost of operation at the end. As the core focus of all the existing studies are mainly towards accuracy, therefore, there was not enough emphasis towards reducing the cost involved in operation especially using all forms of deep learning approaches.
d. Lack of simplified optimization: Majority of the existing studies claims of higher retrieval performance; however, there are few studies which has offered optimization of the training model involved in it. More efforts towards incorporation of optimization will have better probability towards reduction of cost of operation balanced with higher accuracy.
e. Less emphasis toward deeper investigation: There are certain methods which have used GAN, which offers independence from labelled data and capability towards generating high quality synthesized image. Irrespective of a better perspective of GAN, there is less scope of improvement being carried out towards modelling it. There is a need to investigate such networks with more learning option for deep learning.

The statement of the problem can be defined as “developing a robust and comprehensive learning approach towards facilitating a simplified sketch-based image retrieval is computationally complex task.” Therefore, the proposed scheme contributes towards addressing the above-mentioned outlines of research problems as follows: i) a simplified and robust sketch-based image retrieval process is presented using hybridized GAN; ii) the study model uses flexible and multiple pretraining method incorporated within the generator module for optimized outcome of retrieval; iii) the study model emphasizes on both higher accuracy as well as reduced computational burden during the process of retrieval; and iv) an extensive assessment environment is constructed in order to benchmark the presented scheme. The next section discusses about the adopted method.

2. METHOD

The proposed study is an extension of our prior work towards sketch-based image retrieval [31], [32]. In the present work, the agenda is to offer more improvisation towards the retrieval process using a novel analytical research methodology. A deep learning-based approach is presented in the proposed scheme in order to address the identified problems briefed in prior section. The adopted methodology for achieving the study aim is highlighted in Figure 1. According to the adopted architecture exhibited in Figure 1, the proposed system implements a novel and yet simplified design using hybrid form of GAN. It consists of generator and discriminator which are meant to process the input of query image and real-images from the dataset. Upon receiving the input of query images, the generator model applies three explicit set of deep learning individually i.e., ResNet, MobileNet, and U-Net. The processed information from the generator module is forwarded to discriminator module which further yields real and not-real images. This set of intermediate classified information from discriminator is further forwarded to an optimizer that interacts with generator module. This process results in more improved generation of the synthesized image.

This architecture consists of a generator module and discriminator module. Pretrained modules like ResNet, MobileNet, U-Net are incorporated in generator module. Discriminator module classify whether generated image is real or not real image. The unique advantage of this methodology is that it is less iterative and more progressive which can mitigate the computational burden observed in existing deep learning techniques. Apart from this, owing to adoption of interactivity between novel generator module and discriminator, the operation becomes more simplified with presence of optimizer to incorporate enhanced outcome. The next section elaborates about algorithm design.

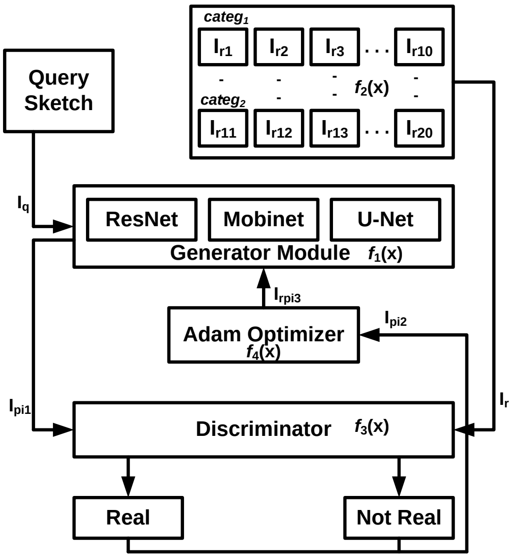


Figure 1. Adopted architecture

2.1. Algorithm implementation design

This section presents the discussion about the algorithm design involved in implementing the methodology briefed in prior section. Upon giving a query sketch to the algorithm, the study model offers an outcome of reconstructed image using unsupervised learning scheme. However, there are various essential operations being carried out within the system model which can be well illustrated using algorithm. The proposed scheme implements a GAN which comprises of two core modules i.e., generator and discriminator. These two modules are interactive enough to finally yield a reconstructed image. The essential steps involved in the proposed algorithm is as follows:

Algorithm for synthesizing image from sketch

```
Input: I_q (query image), I_r (real image)
Output: I_f (final image)
Start
1. For i_q=1:n
2.     i_pi1→f_1(i_q)
3.     I_r=f_2('categ_m')
4.     I_dis=[i_pi1, i_r]
5.     I_pi2→f_3(I_dis)
6.     I_pi3→f_4(I_pi2)
7.     I_pi4→f_1(I_p3)
8.     Go to Line-5
9.     Extract I_f
10. End
End
```

The discussion of the algorithmic steps are as follows: the algorithm takes an input of I_q (query image) and I_r (real image), which yields an outcome of I_f (final image). The formulated algorithm is in its generalized form towards synthesizing image from the input of sketch image. Considering that there are n number of the query hand-drawn digitized sketch images i_q (Line-1), the first part of the algorithmic operation resides towards forwarding i_q to the generator module using an explicit function $f_1(x)$ (Line-2). This function is mainly meant for generating new samples which are slightly different from that of the real-images in dataset. By utilizing a random order, function $f_1(x)$ generates the new samples in adherence with the preciseness of original information. The proposed system implements three more deep learning operation within $f_1(x)$ i.e., ResNet, MobileNet, and U-Net. In parallel, there is another set of operation being carried out associated with the dataset building. For this purpose, the proposed algorithm implements a function $f_2(x)$ for performing web scrapping towards extracting images from networks under m number of categories *categ* (Line-3). All the real images can be stored in a defined directory under explicit categories. For example, there could be 10 images of chairs under the category name '*chair*' or there could be 10 images of tables under category name '*table*'. Hence, a dataset is constructed for all these real-images that are structured in one defined and easily accessible location with the implemented scripts. The next part of the algorithm implementation is to prepare an input for the discriminator module. An array I_{dis} represents an input image to the discriminator module that consists of two elements i.e., outcome obtained from generator I_{pi1} and real images I_r (Line-4). This input array I_{dis} is further subjected to a function $f_3(x)$ which generates an outcome of I_{pi2} (Line-5). This function $f_3(x)$ is responsible for investigating the network that could have been generated from the initial real image dataset or if it has been yielded by the generator module. The criticism generated by the discriminator is used for learning the generator as the latter does not have any form of direct accessibility with the real image dataset. Further, the discriminator checks for both the types of images i.e., counterfeited images from generator and real images from the test.

The error for the discriminator is provided by looking upon the frequency of the order of counterfeited samples as an outcome from the generator. It should be noted that the proposed algorithm is one of the smart mechanisms to train a generator module where both generator and discriminator modules are used for framing up as supervised learning problem. This training operation leads the generator to generation of new samples while classification of these samples is carried out by discriminator in the form of generated sample (counterfeited) and real sample (from domain). By adopting a zero-shot game adversarial approach, the proposed system trains both generator and discriminator. This process is continued until and unless the discriminator model is misled about majority of the time. This will eventually mean that a greater number of plausible samples are now generated by generator module. The output from the discriminator module I_{pi2} is now subjected to function $f_4(x)$ that carries out an optimization that considers multiple gradient descent methods (Line-6). It is one of the suitable alternatives of the conventional stochastic algorithm in order to perform updating of the iteratives of network weights depending upon the training data. The outcome of optimizer i.e., I_{pi3} is reforwarded to generator using similar function $f_1(x)$ to generator outcome of I_{pi4} that is again forwarded to discriminator (Line-7/8) which finally generates a reconstructed final image I_f (Line-9). Apart from this, the rate of learning is exchanged between both the network models; however, rates of learning are required to be differentiated from discriminator to generator with respect to certain goal which cannot be exceeded by neither the discriminator nor generator. Another significant contribution of the proposed algorithm is the usage of an enhanced U-Net architecture that adopts BConvGRU framework for the purpose of segmenting the images. It should be noted that proposed mechanism performs segmentation of an image for obtaining the masked image as this form of mask is capable of retaining the significant information of an image while blocking the residual segment of an image. The prime reason behind this operation is the presence of various redundant information that is eventually not required to be processed during the image query operation. Hence, segmentation is carried out to retain the core part of an image which is essential for reducing computational burden as well as to achieve better retrieval image. Following is an empirical expression which carry out the above-mentioned operation:

$$\mathrm{argmin}_{\mathrm{u}}\gamma||\nabla \mathrm{u}||_{\mathrm{o}} + \int (\mathrm{u}-\mathrm{f})^2 \mathrm{dx} \tag{1}$$

In (1), the proposed system further adds analysis of transform domain for improving the performance of above-mentioned empirical expression x as (2).

$$\mathrm{argmin}_{\mathrm{u,k}}\gamma|\mathrm{K}| + \mu \int_K^C |\nabla \mathrm{u}|^2 \mathrm{dx} + \int (\mathrm{u}-\mathrm{f})^2 \mathrm{dx} \tag{2}$$

In the expression (2), the variable f, u, ∇u, K represents function of image pixel, average strength of pixel, pixel variance, and constant that is optimized by gated recurrent unit (GRU).

3. RESULTS AND DISCUSSION

This section discusses about the accomplished outcome from the implementation of proposed algorithm discussed in prior section. The complete assessment of the outcome of proposed system is carried out using sketch image dataset [33] which consists of 70,000 total images with 56,300 training images and 14,100 testing ratios. From the viewpoint of training aspect involved in proposed study, the assessment is carried out considering 80% of training images and 20% of testing images. Further, the results have been recorded for 1,000 epochs for 32 batch sizes with binary cross entropy as the loss function. The outcome of the discriminator module has been improved using Adam optimizer. Figure 2 exhibits random 9 sketch query images given as input to proposed system out of 250 objects.

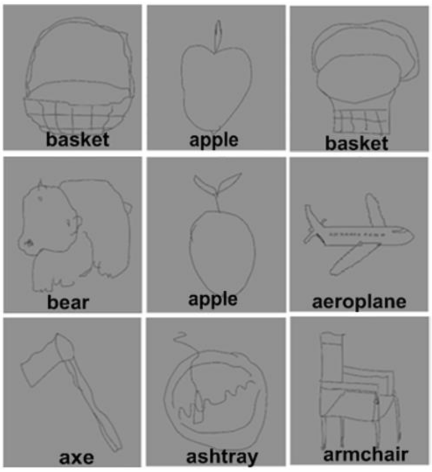


Figure 2. Sample query sketch images

The proposed system carries out training operation as well as fine-tuning over the presented model using Keras framework while it runs TensorFlow on the background. The training operation of the model is carried out over GeForce RTX 2060 GPU and Intel(R) i7 CPU.2060Ti GPU. The assembling of the training dataset is carried out in the form of triplets obtained from the similarity matrix. Figure 2 highlights the visuals of the query images being used for validating the proposed scheme. A deep variational autoencoder is constructed from the normal VGG16 architecture by eliminating the all the connected layers of VGG16. A single layer of variational autoencoder is used for augmenting the left model in order to construct an encoder that is further integrated with an analogous decoder.

Different forms of parameters were used for assessing the variational autoencoder model viz. dimension of variance encoding, statistical mean, number of convolution layers. Considering different numbers of pseudo sketches of dataset, the analysis is carried out over 5 layers of convolution blocks and 128 dimensions of encoding generated by the optimal accuracy of reconstruction. A batch size of variational autoencoder is used for training while 0.001 is considered as preliminary rate of learning. Figure 3 shows reconstruction of arm chair sketch by using 3 pretrained networks like ResNet, MobileNet, and U-Net.

From Figure 4 for comparative analysis, the outcome shows that proposed scheme offers better performance for MobileNet in comparison to ResNet and U-Net. A closer interpretation of Figure 4(a) depicts a comparison outcome for accuracy vs. epoch, where MobileNet offers significantly higher accuracy with consistency in contrast with ResNet and U-Net for the increased epoch values. On the other hand, as depicted in Figure 4(b), MobileNet outperforms both ResNet and U-Net for loss analysis. The prime rationale behind this outcome is that MobileNet offers faster performance with reduced network size for which reason both accuracy and loss values are optimized. On the other hand, ResNet and U-Net have more dependency for higher consumption time. Yet, they are quite good enough for existing applications.

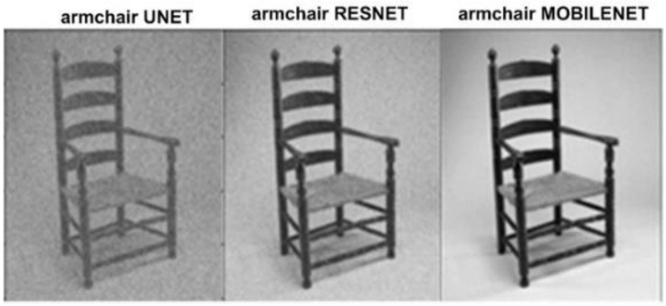
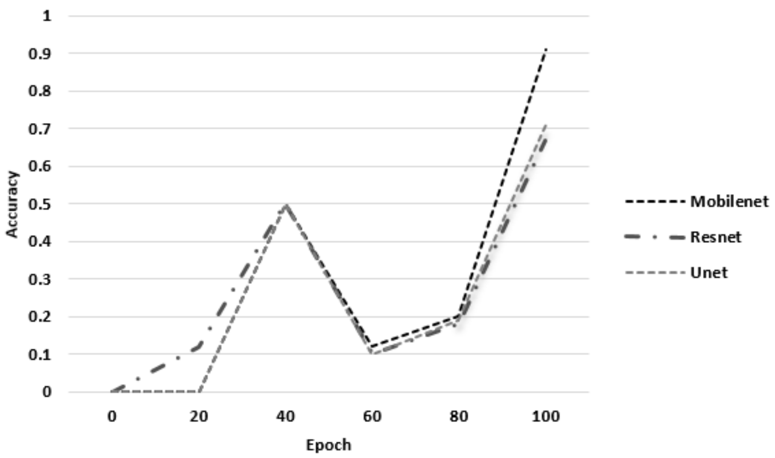
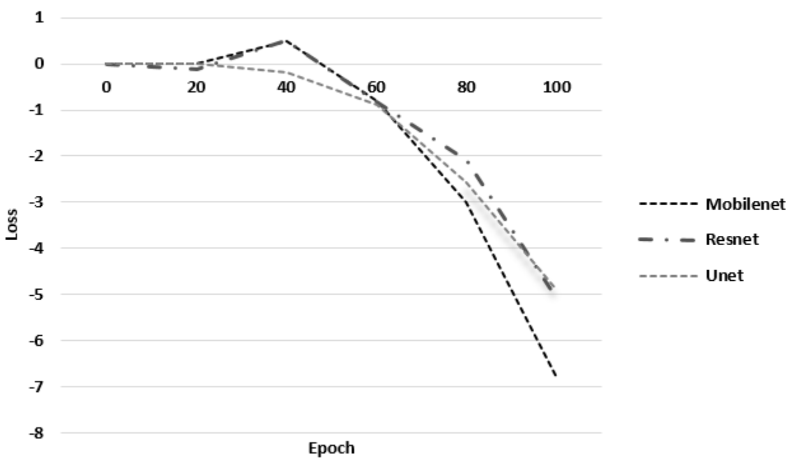


Figure 3. Reconstruction of the armchair sketch by three different networks



(a)



(b)

Figure 4. Comparative analysis of (a) accuracy vs epoch and (b) loss vs epoch between pretrained networks like ResNet, MobileNet, U-Net. It shows that MobileNet performs better compared to other networks

4. CONCLUSION

From the discussion carried out in this paper, it can be said that performing an identification as well as synthesizing of images from queried sketches is never an easy task. Existing learning-based approaches and other techniques do offer beneficial features, but they are equally associated with challenges. Hence, the proposed scheme offers a novel synthesizing scheme considering a unique hybrid GAN. The presented generator module is capable of training the network with massive number of layers with no dependency towards maximizing score of error proportion. Owing to unsupervised scheme of network, there is no dependency towards training labelled data while it also offers finest quality of reconstructed images. The proposed model incorporates segmentation benefits as it permits the simultaneous adoption of global location as well as context leading to reduction of computational effort. With an extensive test environment, it was noticed that MobileNet offers better performance in comparison to ResNet and U-Net architecture for presented hybrid generator module. The novelty accomplishes by the proposed scheme are as follows viz. i) existing problems associated with zero-shot has been addressed in proposed system by improvising the feature management carried out in considered generator module. Proposed model offers a hybridized deep learning methods which offers varied characteristic to mitigate trade-off between computational effort and accuracy and thereby it overcomes the identified problems of existing deep learning techniques; ii) further, the closer look into methodology and formulated algorithm exhibits that complete operation is controlled by interactivity between generator and discriminator module. This phenomenon reduces higher number of sophisticated operational steps and thereby it addresses the identified problems of higher operational cost of existing scheme; iii) further, incorporation of deep learning methods e.g., ResNet, MobileNet, and U-Net within generator module not only offers simplified optimization, which is also a serious concern with existing techniques, but also offers more emphasis towards deeper investigation. Therefore, the proposed system has successfully offered solution towards identified research problem.

REFERENCES

[1] H. Laaroussi, F. Guerouate, and M. Sbihi, "A novel hybrid deep learning approach for tourism demand forecasting," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 1989–1996, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1989-1996.

[2] B. Ionescu, W. A. Bainbridge, and N. Murray, Eds., *Human perception of visual information*. Cham: Springer International Publishing, 2022.

[3] N. Lakshmi and M. P. Arakeri, "A novel sketch based face recognition in unconstrained video for criminal investigation," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 1499–1509, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1499-1509.

[4] Y. Song, J. Lei, B. Peng, K. Zheng, B. Yang, and Y. Jia, "Edge-guided cross-domain learning with shape regression for sketch-based image retrieval," *IEEE Access*, vol. 7, pp. 32393–32399, 2019, doi: 10.1109/ACCESS.2019.2903534.

[5] D. Birari, D. Hiran, and V. Narawade, "Survey on sketch based image and data retrieval," in *Lecture Notes in Electrical Engineering*, Springer Singapore, 2020, pp. 285–290.

[6] J. Sheng, F. Wang, B. Zhao, J. Jiang, Y. Yang, and T. Cai, "Sketch-based image retrieval using novel edge detector and feature descriptor," *Wireless Communications and Mobile Computing*, pp. 1–12, Feb. 2022, doi: 10.1155/2022/4554911.

[7] R. Hu and J. Collomosse, "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval," *Computer Vision and Image Understanding*, vol. 117, no. 7, pp. 790–806, Jul. 2013, doi: 10.1016/j.cviu.2013.02.005.

[8] S. R. Dubey, "A decade survey of content based image retrieval using deep learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2687–2704, May 2022, doi: 10.1109/TCSVT.2021.3080920.

[9] X. Wu, K. Xu, and P. Hall, "A survey of image synthesis and editing with generative adversarial networks," *Tsinghua Science and Technology*, vol. 22, no. 6, pp. 660–674, Dec. 2017, doi: 10.23919/TST.2017.8195348.

[10] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, "A zero-shot sketch-based intermodal object retrieval scheme for remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3056392.

[11] C. Deng, X. Xu, H. Wang, M. Yang, and D. Tao, "Progressive cross-modal semantic network for zero-shot sketch-based image retrieval," *IEEE Transactions on Image Processing*, vol. 29, pp. 8892–8902, 2020, doi: 10.1109/TIP.2020.3020383.

[12] T. Dutta, A. Singh, and S. Biswas, "StyleGuide: zero-shot sketch-based image retrieval using style-guided image generation," *IEEE Transactions on Multimedia*, vol. 23, pp. 2833–2842, 2021, doi: 10.1109/TMM.2020.3017918.

[13] J. Choi, H. Cho, J. Song, and S. M. Yoon, "SketchHelper: real-time stroke guidance for freehand sketch retrieval," *IEEE Transactions on Multimedia*, vol. 21, no. 8, pp. 2083–2092, Aug. 2019, doi: 10.1109/TMM.2019.2892301.

[14] C. Hu, D. Li, Y. Yang, T. M. Hospedales, and Y.-Z. Song, "Sketch-a-segmenter: sketch-based photo segmenter generation," *IEEE Transactions on Image Processing*, vol. 29, pp. 9470–9481, 2020, doi: 10.1109/TIP.2020.3028292.

[15] S. Jiao, X. Han, F. Xiong, F. Sun, R. Zhao, and L. Kuang, "Cross-domain correspondence for sketch-based 3d model retrieval using convolutional neural network and manifold ranking," *IEEE Access*, vol. 8, pp. 121584–121595, 2020, doi: 10.1109/ACCESS.2020.3006585.

[16] Q. Qi, Q. Huo, J. Wang, H. Sun, Y. Cao, and J. Liao, "Personalized sketch-based image retrieval by convolutional neural network and deep transfer learning," *IEEE Access*, vol. 7, pp. 16537–16549, 2019, doi: 10.1109/ACCESS.2019.2894351.

[17] J. Lei, Y. Song, B. Peng, Z. Ma, L. Shao, and Y.-Z. Song, "Semi-heterogeneous three-way joint embedding network for sketch-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 9, pp. 3226–3237, Sep. 2020, doi: 10.1109/TCSVT.2019.2936710.

[18] Z. Li, C. Deng, E. Yang, and D. Tao, "Staged sketch-to-image synthesis via semi-supervised generative adversarial networks," *IEEE Transactions on Multimedia*, vol. 23, pp. 2694–2705, 2021, doi: 10.1109/TMM.2020.3015015.

[19] S. Setumin, M. F. C. Aminudin, and S. A. Suandi, “Canonical correlation analysis feature fusion with patch of interest: a dynamic local feature matching for face sketch image retrieval,” *IEEE Access*, vol. 8, pp. 137342–137355, 2020, doi: 10.1109/ACCESS.2020.3009744.

[20] J. Wan, K. Zhang, H. Li, and A. B. Chan, “Angular-driven feedback restoration networks for imperfect sketch recognition,” *IEEE Transactions on Image Processing*, vol. 30, pp. 5085–5095, 2021, doi: 10.1109/TIP.2021.3071711.

[21] L. Wang, X. Qian, X. Zhang, and X. Hou, “Sketch-based image retrieval with multi-clustering re-ranking,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4929–4943, Dec. 2020, doi: 10.1109/TCSVT.2019.2959875.

[22] L. Wang, X. Qian, Y. Zhang, J. Shen, and X. Cao, “Enhancing sketch-based image retrieval by CNN semantic re-ranking,” *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3330–3342, Jul. 2020, doi: 10.1109/TCYB.2019.2894498.

[23] Z. Weng and Y. Zhu, “Online supervised sketching hashing for large-scale image retrieval,” *IEEE Access*, vol. 7, pp. 88369–88379, 2019, doi: 10.1109/ACCESS.2019.2926303.

[24] F. Xu, W. Yang, T. Jiang, S. Lin, H. Luo, and G.-S. Xia, “Mental retrieval of remote sensing images via adversarial sketch-image feature learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7801–7814, Nov. 2020, doi: 10.1109/TGRS.2020.2984316.

[25] P. Xu *et al.*, “Fine-grained instance-level sketch-based video retrieval,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1995–2007, May 2021, doi: 10.1109/TCSVT.2020.3014491.

[26] P. Xu *et al.*, “On learning semantic representations for large-scale abstract sketches,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 9, pp. 3366–3379, Sep. 2021, doi: 10.1109/TCSVT.2020.3041586.

[27] Y. Xu, J. Hu, K. Wattanachote, K. Zeng, and Y. Gong, “Sketch-based shape retrieval via best view selection and a cross-domain similarity measure,” *IEEE Transactions on Multimedia*, 2020, doi: 10.1109/TMM.2020.2966882.

[28] Z. Yang, X. Zhu, J. Qian, and P. Liu, “Dark-aware network for fine-grained sketch-based image retrieval,” *IEEE Signal Processing Letters*, vol. 28, pp. 264–268, 2021, doi: 10.1109/LSP.2020.3043972.

[29] Y. Zheng, H. Yao, and X. Sun, “Deep semantic parsing of freehand sketches with homogeneous transformation, soft-weighted loss, and staged learning,” *IEEE Transactions on Multimedia*, vol. 23, pp. 3590–3602, 2021, doi: 10.1109/TMM.2020.3028466.

[30] B. Zhao, B. Wu, T. Wu, and Y. Wang, “Zero-shot learning posed as a missing data problem,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2017, pp. 2616–2622, doi: 10.1109/ICCVW.2017.310.

[31] N. Pavithra and K. Y. H. Sharath, “Sketch-based image retrieval: effectivity notion of recent approaches,” in *2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, pp. 1–6, doi: 10.1109/INCET49848.2020.9153999.

[32] N. Pavithra and Y. Sharath Kumar, “FSBRS: framework for sketch-based retrieval system of the color images,” in *Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020*, 2021.

[33] H. Wang, S. Ge, E. P. Xing, and Z. C. Lipton, “Learning robust global representations by penalizing local predictive power,” *Advances in Neural Information Processing Systems*, vol. 32, May 2019.

BIOGRAPHIES OF AUTHORS

Pavithra Narasimha Murthy is a research scholar at Maharaja Institute of Technology, Mysuru, Visvesvaraya Technological University (VTU), India. Her areas of interest are digital image processing, big data analytics, data science, deep learning, data mining. She has around 15 years of experience. She can be contacted at email: pavithra.apr02@gmail.com.

Sharath Kumar Yeliyur Hanumanthaiah is a Professor and Head, Department of Information Science and Engineering, Maharaja Institute of Technology, Mysuru, Visvesvaraya Technological University (VTU), India. His areas of interest are image processing, pattern recognition, and information retrieval. He has around 15 years of experience in teaching. He has one year of industry experience. He can be contacted at email: sharathyhk@gmail.com.