iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VIII **Month of publication:** August 2019

DOI: http://doi.org/10.22214/ijraset.2019.8138

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

Deep Learning based Enhanced Triplet Nework Model for Landmark Classification in Image Retrieval

K. Shanmuga Sundari, S M. Thamarai, Dr. T. Meyyappan
[1]*M.Phil Research Scholar, Department of Computer Science, Alagappa University, Karaikudi.*
[2]*Guest Lecturer, Alagappa Government Arts College, Karaikudi.*
[3]*Professor, Department of Computer Science, Alagappa University, Karaikudi.*

Abstract: ***Deep learning exist as a successful models for learning, and representing data through semantic method. These are learned as a Section of a classification task. This research work proposes the model of the triplet network, focusing on learning helpful depictions through distance comparisons. Landmark retrieval is a process of restoring a collection of images along its landmarks which is parallel to required images. In the case office reality, studies regarding landmark retrieval concentrates on utilizing the landmarks geometries regards its similarities especially the visual matches. The visual content of social images is of huge diversity in various landmarks, and a few images have similar patterns among various landmarks. At the same time, we noticed that multimodal contents will be there in social images, i.e., visual content and text tags, and landmark will be with its individual characteristic of both of visual as well as the text content. So that the approach on the basis of matching the similar images, may not be highly appealing in this environment. This research work focuses on, whether the visual and text content could be utilized over geographical correlation for landmark retrieval. Particularly, The work for landmark retrieval form an enhanced Triplet Network modal for landmark classification in order to project social image through multimodal content , by a joint model it integrates both landmark classifier and retirement through multimodal contents. The images with geo-tag will normally noted for classifier learning. On the basis of low rank matrix recovery Visual features get refined, and through automatically labeled images it can learn multimodal classification and group sparse. Finally, candidate images are ranked with the consequence of classification as well as semantic consistency between the visual and text content of the combination office. Research on real data visualize the superiority of this approach like comparing with the existing methods.***
Keywords: Landmark Recognition,Convolutional neural network,CNN,Triplet Nework, Deep Learning

I. INTRODUCTION

The number of user-contribution to social images with the content like textual tags, description, and visual contents are increasing rapidly as because of the growth of camera and mobile gadgets. Most of these are geo-tagged and close to landmarks, e.g., flickr.com and Picasa Web Album. It is tough enough to strengthen the overwhelming the context data and information regarding geometry for social images and applications [1], like landmark retrieval which returns image pattern with its landmarks highly similar to that of the query image. The concept of Image recognition has been used frequently in various applications. Till date, various attempts are made to improve the functionalities of typical image based recognition techniques. Available researches show that with the help of machine learning, these techniques can improve the knowledge bases remarkably. Any typical recognition system uses a single image data for processing purpose which provides a moderate probability of content matching which in turn reduces the overall efficiency of the system. On the other hand, if pictorial data of any angle for specific landmarks is collected, the probability of perfect matches increases outstandingly. The threshold value decreases drastically as the dataset used will contain large metadata. This ultimately decreases user efforts and error generations. Deep learning plays a very vital role for real-time applications. The efficiency of any deep learning algorithm depends on its training data.The more useable data available, more is the output availability. Also, compatibility for various devices is an important factor. When matching algorithms are used, main point to be considered is the processing time and space consumed for it. If pixel matching is discussed, it goes through the entire image and creates a sized histogram of pixel data. On the other hand, Content matching also creates a sized metadata set but it effectively reduces the processing time as it categorizes the image objects and uses these objects for comparison. While parallelizing the set of images in traditional content-based image retrieval (CBIR) systems, images with landmark are with its unique nature. In particular, though each landmark are having characteristics of its own, most of it share same visual content as in Fig. 1(a).

Fig. 1. Illustration of landmark images. (a) Two group of visually similar images taken at Trabzon - Atatürk Pavilion. (b) Two visually similar images corresponding to different landmark but tagged with different tags.

It is different from image retrieval in conventional form: at the time of retrieving a set of related images around the space of low-level feature (e.g., color and texture), things get difficult at the time of retrieving the images with same landmark owing to the landmarks with high diversity of low-level features

This work propose using unsupervised triplet networks for remote sensing image classification and Landmark recognition . A triplet network can be trained from scratch employing weakly labeled data, to some degree relieving the pressure of labeling a great number of images. And in a way, it makes a tradeoff among the availability of labeled data, the flexibility of CNN architecture and the size of a network. Most important of all, networks of triplet paired with our suggested losses gets a position of the art performance at scene classification.

II. RELATED WORK AND LITERATURE SURVEY

This work is related to vast area of literature, particularly, a shot learning through generative models, for one-shot learning an embedding space, and visual based tracking of an object.

Past research works focuses on the context of models of generative for one-shot learning, it gets differ from formulated problem as discriminative task. An old approach [10] use models of probabilistic generativeness to project the object categories and projects it invitational Bayesian framework to get needed info from training samples. Recurrent spatial attention[11],generative model is there to stipulate images with a variation Bayesian inference. While re-investigating the sample, it gets a pattern of diverse samples. Geographic referencing of images include data-driven methods and model-based methods. The landmark or geographical location of the needed image will be determined by data-driven method by getting the closest neighbors from a prebuilt database. In some case, the image databases will be constructed with tree-based structure [21] or through 3-D model [19], [22] to store efficiency office retrieval. Hays and Efros [7], with regard to the needed landmark image, they present feature matching approach to get back the K nearest neighbors , which stands as an expected images and images in data by a set of low-level features to get landmark retrieval.

Li et al. [11] gets candidates, which are similar visually by accepting its geo-visual neighbors that are geographically nearer and visually similar. In some methods, location will be given as both city or global scale [14], and the method of searching is used as part of landmark recognition or classification [15], [16], [18]–[20]. Serdyukovet al. [2] targeted Flickr image's location with a language model on the basis of the tags of users. Fang et al. [9] presents GIANT approach, to discover attributes like discriminative and representative mid-level for landmark retrieval. Wang et al. [10] proposed a method of multi query expansion to get landmark, which shows the query expansion images through discriminative patterns . The multimodal hyper graph (MMHG) is proposed to combine different types of visual features for landmark image retrieval. Through this methods landmark model retrieve through the visual content, which rejects lot of text information of social images.

Li et al. [17] and Crandall et al. [23] proposed to differentiate images of landmark through linearly combining the features of visual context and textual tags.

Cao et al. [24] proposed a method of ranking to combine numerous evidences from both textual and visual features for the estimation of image location. These approaches consider different types of features independently, which cannot exploit the inherent correlation between different types of features effectively.

III. PROPOSED WORK

In this preferred method, for the purpose of recognition of landmark we can use triplet network. In this Work, prefer a Convolutional neural network (CNN) on the basis of Triplet networks algorithm with the help of unsupervised data. This proposed work we applies deep learning a in a brief way. Then S-data will utilize, 4 conventional ML algorithms, then for training 3 algorithms and single algorithm for classification

Again it is significant to make sure that, this research work uses deep learning algorithms in a brief way.

1) K-nearest Neighbour (KNN)
2) VGG16
3) InceptionV3
4) ResNet

This work focuses on the brief review of object tracking with Triplet networks, of recent times. At the time of progressing with one-shot learning scheme in the requirement of visual tracking, in the primary frame, an exemplar is used as an object patch and search region patches which is there in consecutive frames are applied as candidate instances. It focus on parallel instances of each section in a given space, right now the object is projected as a low-dimensional vector. Learning an embedding function with both an effective and discriminative representation will be considered as a crucial move. Deep learning method is applied to understand embedding function and created a clear convolution Siamese network to reduce computation for real-time speed.

This network has various inputs with dual network processing. An exemplar branch is considered as a primary thing in order to get object object bounding box in primary frame. Instances branch which is for process the sector at the search place of next frames.

A. One-shot learning

Under the vision of computer, an object categorization problem occurs. Technique of machine learning works under the guidance of object categorization algorithms, for this work out on vast images and data will be required, one-shot learning focuses to obtain data of object which classify from one few training images.

Triplet network has triple instances of unique feed forward network (with exchanged parameters). While processing with triple samples, the network gives two intermediate values - the L_2distancefrom the embedded representation of dual inputs to the third. If all 3 inputs are mentioned as x, x+ and x-, and the embedded network will be represented as Net(x), vector will be the thing before layer: as in Fig. 2. Classification using knn as in Fig. 3.

1) Goal: Final

Learn new representation f(.) of the original image such that

$$\|f(a) - f(p)\|_2^2 \leq \|f(a) - f(n)\|_2^2$$

Add some margin α

$$\|f(a) - f(p)\|_2^2 + \alpha \leq \|f(a) - f(n)\|_2^2$$

2) Triplet Loss

$$L(a, p, n) = \max\{\|f(a) - f(p)\|_2^2 + \alpha - \|f(a) - f(n)\|_2^2, 0\}$$

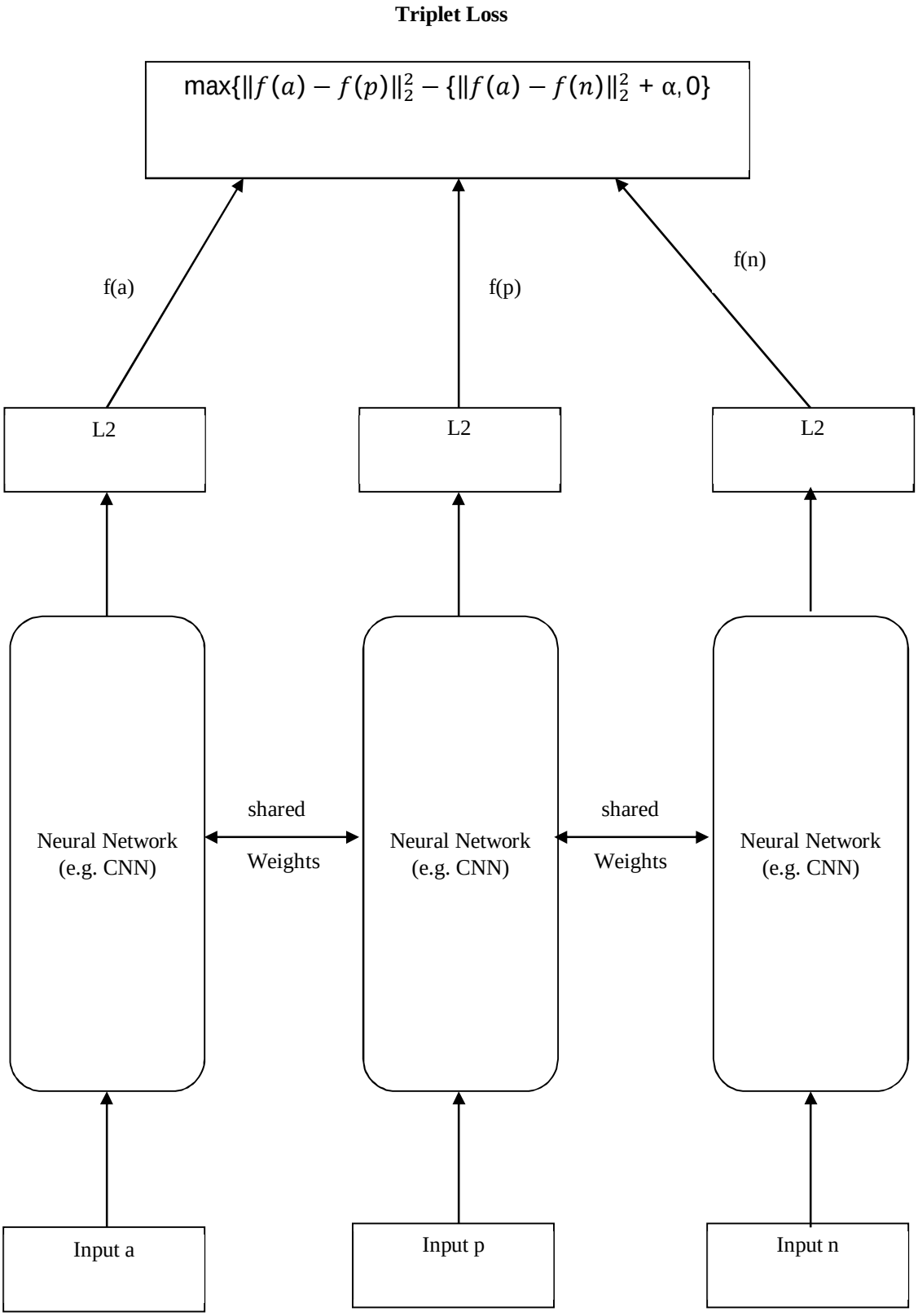


Fig. 2. Proposed Research Architecture

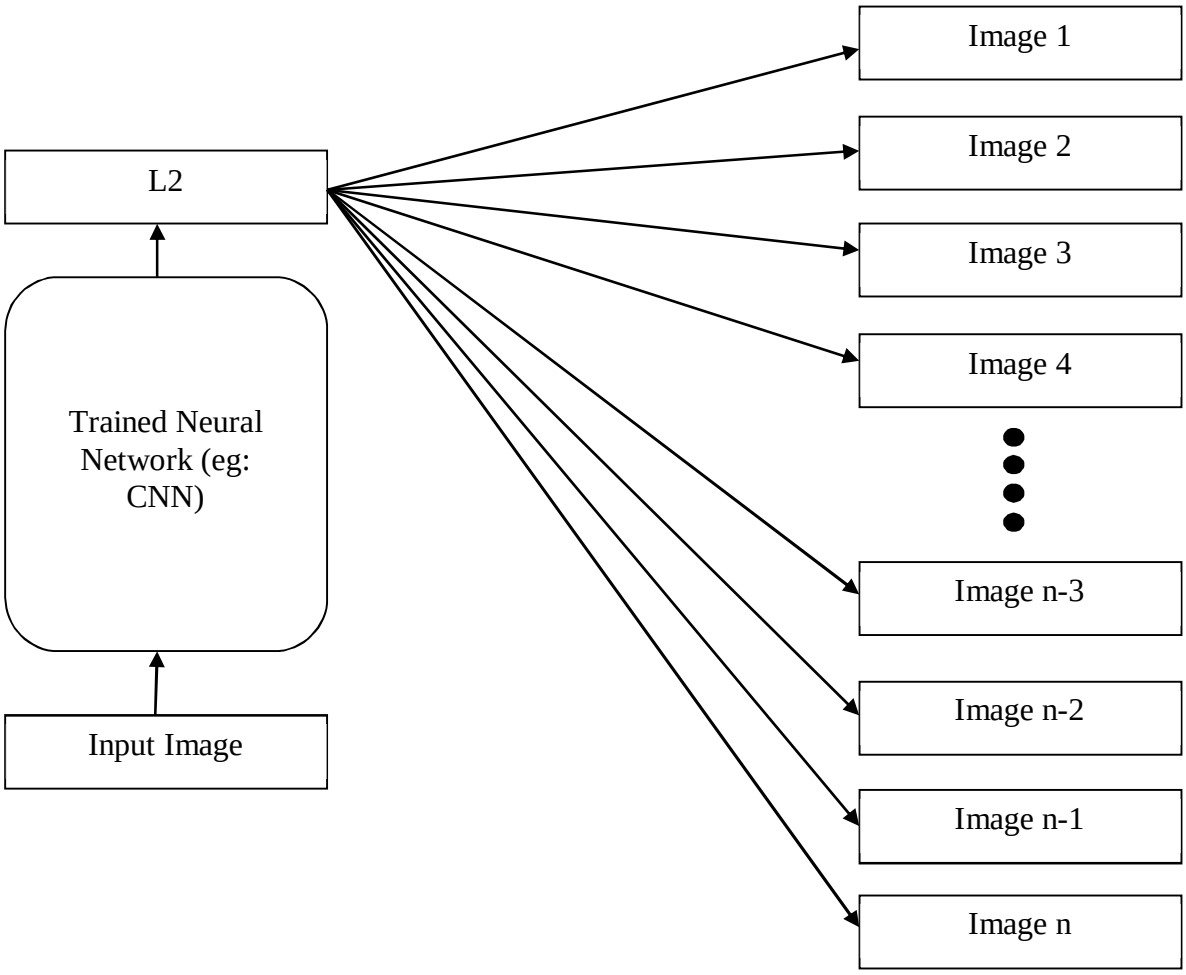


Fig. 3. Block Diagram of Classification Process

IV. EXPERIMENTAL RESULT

In order to proceed with this method, Landmark Recognition Dataset is taken from Google, which will be taken for the purpose of training and testing. Form this work 1,22,502 images drilled with 1,495 landmarks and 11,770 images were verified. On the whole, landmark classification is based on the things like, Content ,text, tags and Geo Location and color histogram as per its significance. The result is given below.Fig.4

Fig.4. Image Visualization Process

```
In [3]: train_all.head()
```

Out[3]:

	Unnamed: 0	id	url	landmark_id
0	16004	ebd7e2c9ae1d6c0f	https://lh3.googleusercontent.com/-ObREWK84G_U...	861
1	21618	ba4abad93833f781	http://lh4.ggpht.com/-m7GX-Dnk8dI/TNBhTfcyJOI/...	1107
2	10635	f9591924b056bfbc	http://lh5.ggpht.com/-NNNpffOhIOI/UDXVqK2bhXI/...	228
3	7079	5f5b50c400256f01	https://lh5.googleusercontent.com/-VGJ3tMPk7Po...	654
4	21020	808f8fdd42f779e3	http://mw2.google.com/mw-panoramio/photos/medi...	993

Fig.5 Image Training Process with landmark id

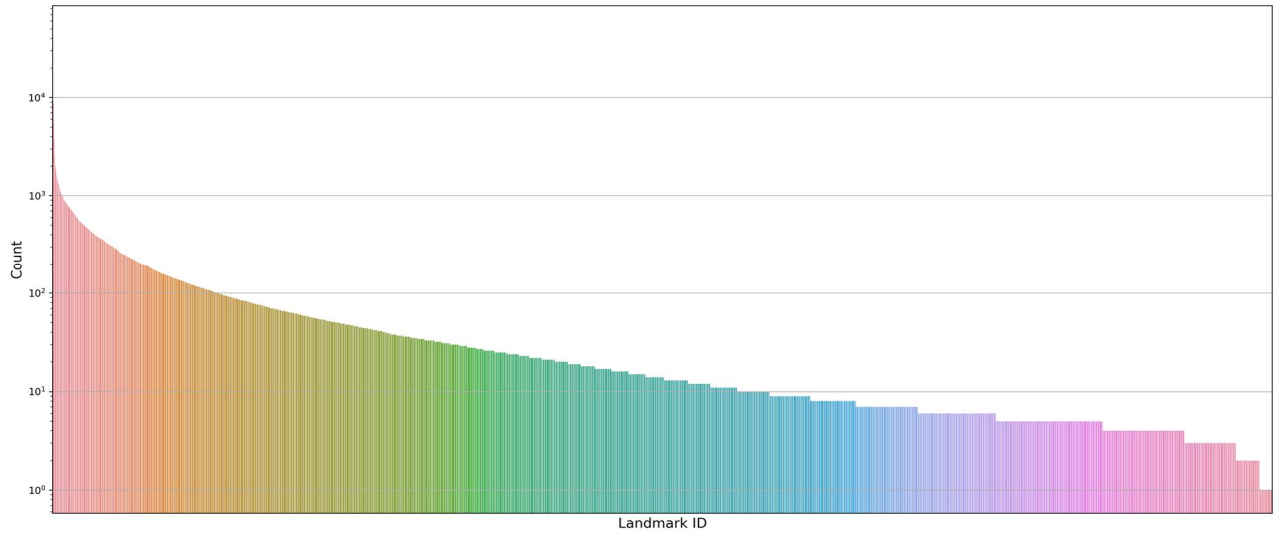


Fig.6.Image Training Result

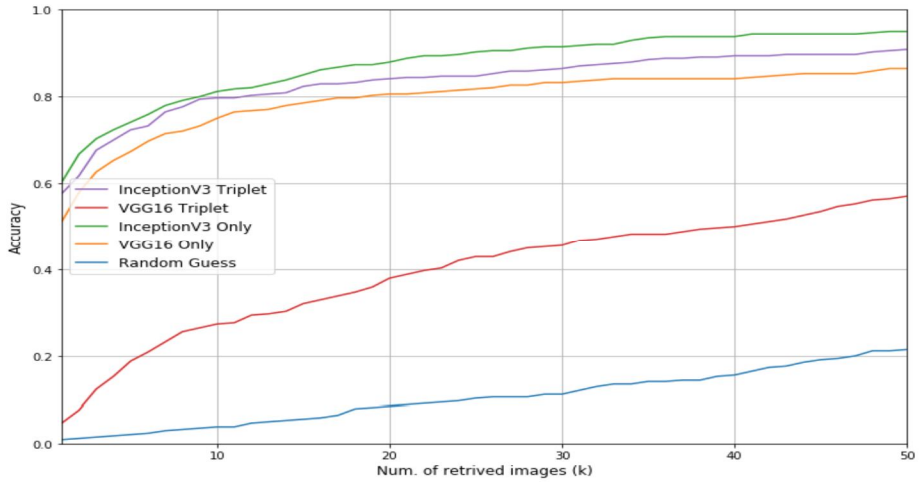


Fig.7. Proposed Model Accuracy Result

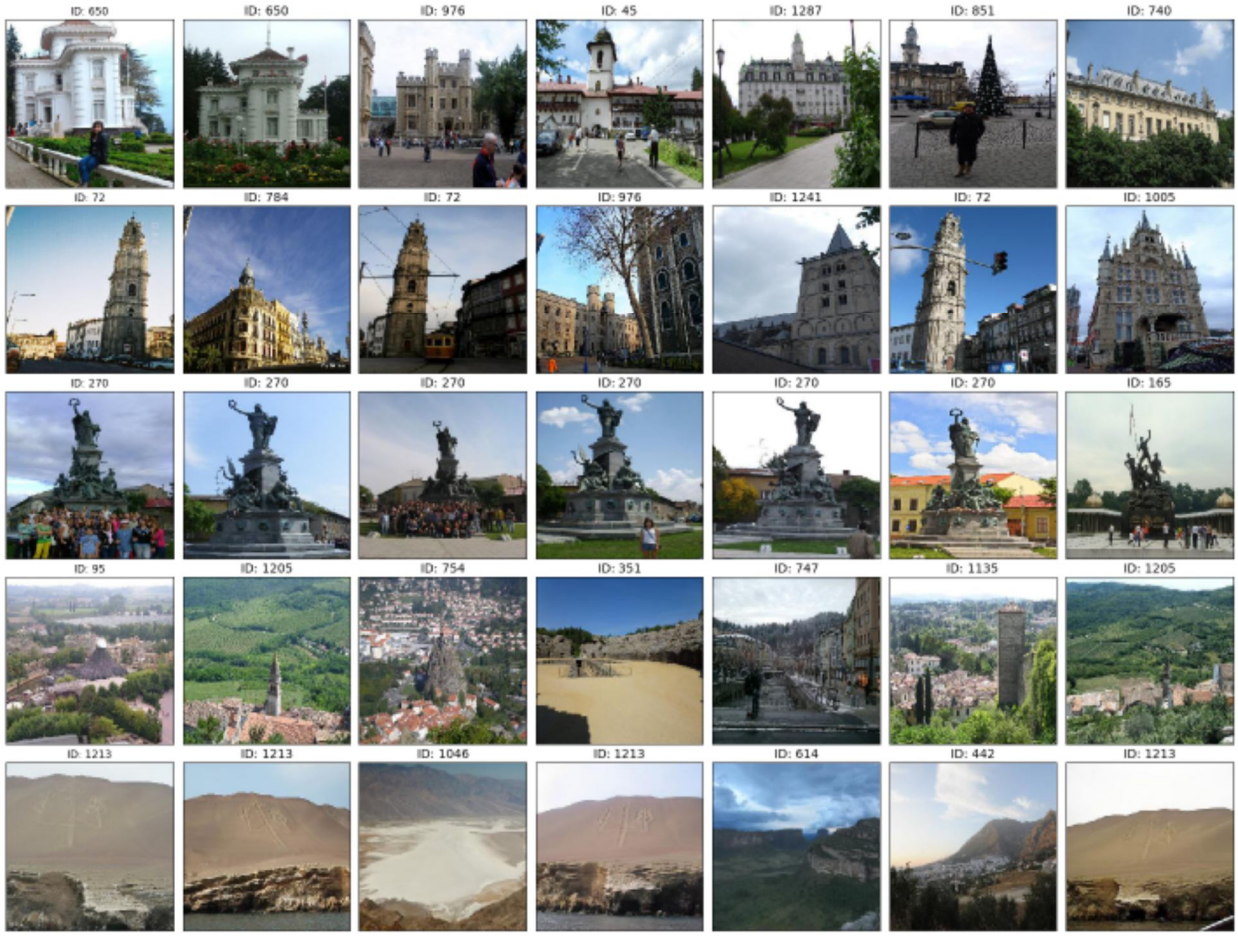


Fig.8 Model Prediction Result

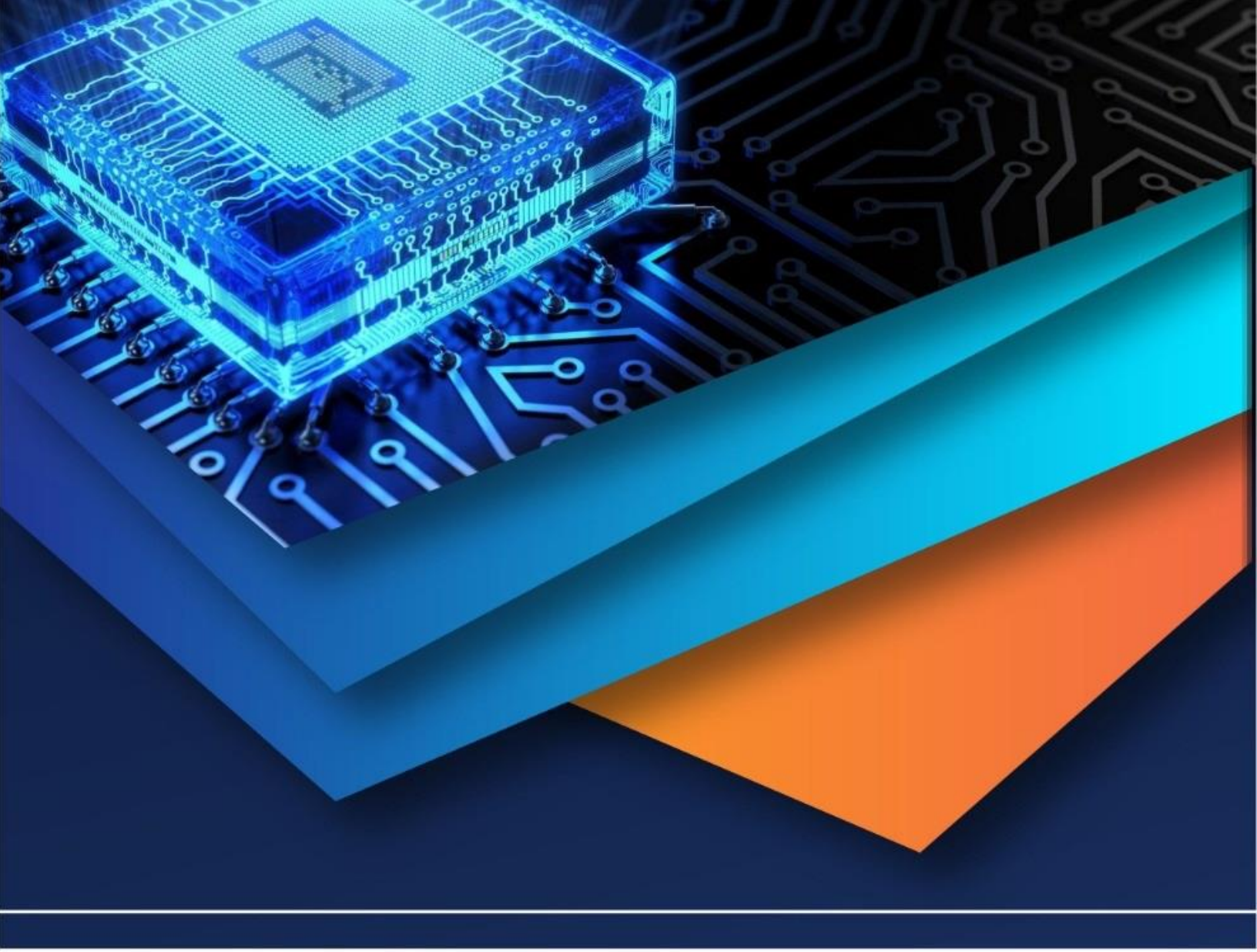
V. CONCLUSION

In this paper, activates the crisis of geo-tagged landmark retrieval images with multimodal content. So, research propose an enhanced triplet model in order to leverage the images of multimodal for the purpose landmark retrieval. Followings are the primary problems i.e., redundant and noisy visual images, and heterogeneous feature regards with both image visual and text content. Specifically, visual characteristics are clarified on the basis of low- ranking matrix recovery, and to identify landmarks with the combination of group sparse,multimodalclassificationislearnedfromautomaticallylabeledimages. Extended experiments on social picture datasets of this realmprojects the superiority of the recommended strategy to the present techniques. This work attempts to highlight landmark retrieval along automatically learned multimodal classifier from geographically structural analysis as well as the model of sparse group. It expands the present image retrieval study with the focus on the direct study of the retrieval model from raw characteristics and it too ignores the latent correlation among multimodal content.

REFERENCES

[1] H. M. Sergieh et al., “Geo-based automatic image annotation,” in Proc. Annu. ACM Int. Conf. Multimedia Retrieval, Hong Kong, 2012, Art. no. 46.
[2] P. Serdyukov, V. Murdock, and R. Van Zwol, “Placing Flickr photos on a map,’ in Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Boston, MA, USA, 2009, pp. 484–491.
[3] O. Van Laere, S. Schockaert, and B. Dhoedt, “Finding locations of Flickr resources using language models and similarity search,” in Proc. Annu. ACM Int. Conf. Multimedia Retrieval, Trento, Italy, 2011, pp. 1–8.
[4] O. Van Laere, J. Quinn, S. Steven, and B. Dhoedt, “Spatially aware term selection for geotagging,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 221–234, Jan. 2014.
[5] Z. Xia et al., “A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing,” IEEE Trans. Inf. Forensics Security, vol. 11, no. 11, pp. 2594–2608, Nov. 2016.

[6] J. Luo, D. Joshi, J. Yu, and A. Gallagher, "Geotagging in multimedia and computer vision—A survey," Multimedia Tools Appl., vol. 51, no. 1, pp. 187–211, 2011.
[7] J. Hays and A. Efros, "IM2GPS: Estimating geographic information from a single image," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Anchorage, AK, USA, 2008, pp. 1–8.
[8] C. Doersch, S. Singh, H. Mulam, J. Sivic, and A. Efros, "What makes Paris look like Paris?" ACM Trans. Graph., vol. 31, no. 4, pp. 13–15, 2012.
[9] Q. Fang, J. Sang, and C. Xu, "Discovering geo-informative attributes for location recognition and exploration," ACM Trans. Multimedia Comput. Commun. Appl., vol. 11, no. 1s, 2014, Art. no. 19.
[10] Y. Wang, X. Lin, L. Wu, and W. Zhang, "Effective multi-query expansions:Robust landmark retrieval," in Proc. ACM Multimedia, Brisbane, QLD, Australia, 2015, pp. 79–88.
[11] X. Li, M. Larson, and A. Hanjalic, "Global-scale location prediction for social images using geo-visual ranking," IEEE Trans. Multimedia, vol. 17, no. 5, pp. 674–686, May 2015.
[12] C.-Y. Chen and K. Grauman, "Clues from the beaten path: Location estimation with bursty sequences of tourist photos," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 42. Colorado Springs, CO, USA, 2011, pp. 1569–1576.
[13] E. Kalogerakis, O. Vesselova, J. Hays, A. A. Efros, and A. Hertzmann, "Image sequence geolocation with human travel priors," in Proc. IEEE 12th Int. Conf. Comput. Vis., Kyoto, Japan, 2009, pp. 253–260.
[14] G. Patterson and J. Hays, "SUN attribute database: Discovering, annotating, and recognizing scene attributes," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Providence, RI, USA, 2012, pp. 2751–2758.
[15] D. M. Chen et al., "City-scale landmark identification on mobile devices," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Colorado Springs, CO, USA, 2011, pp. 737–744.
[16] Y.-T. Zheng et al., "Tour the world: Building a Web-scale landmark recognition engine," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Miami, FL, USA, 2009
[17] Y. Li, D. J. Crandall, and D. P. Huttenlocher, "Landmark classification in large-scale image collections," in Proc. IEEE 12th Int. Conf. Comput. Vis., Kyoto, Japan, 2009, pp.
[18] R. Ji et al., "Learning compact visual descriptor for low bit rate mobile landmark search," in Proc. 22nd Int. Joint Conf. Artif. Intell., Barcelona,Spain, 2011, pp. 2456–2463.
[19] X. Xiao, C. Xu, J. Wang, and M. Xu, "Enhanced 3-D modeling for landmark image classification," IEEE Trans. Multimedia, vol. 14, no. 4, pp. 1246–1258, Aug. 2012.
[20] Z. Cheng, J. Ren, J. Shen, and H. Miao, "Building a large scale test collection for effective benchmarking of mobile landmark search," in Advances in Multimedia Modeling. Heidelberg, Germany: Springer, 2013, pp. 36–46.
[21] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Minneapolis, MN, USA, 2007, pp. 18–23.
[22] H. Liu, T. Mei, J. Luo, H. Li, and S. Li, "Finding perfect rendezvous on the go: Accurate mobile visual localization and its applications to routing," in Proc. ACM Multimedia, Nara, Japan, 2012, pp. 9–18.
[23] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. M. Kleinberg, "Mapping the world's photos," in Proc. 18th Int. Conf. World Wide Web, Madrid, Spain, 2009, pp. 761–770.
[24] J. Cao, Z. Huang, and Y. Yang, "Spatial-aware multimodal location estimation for social images," in Proc. ACM Multimedia, Brisbane, QLD, Australia, 2015, pp
[25] R. Ji et al., "When location meets social multimedia: A survey on visionbased recognition and mining for geo-social multimedia analytics," ACM Trans. Intell. Syst. Technol., vol. 6, no. 1, pp. 1–18, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY