JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IMAGE BASED SEARCH ENGINE USING DEEP LEARNING - Review

GUNDU MOUNIKA
Department of Master of Computer Science
Miracle Educational Society Group of Institutions
Vizianagram– 535216 (AP) India

SARAGADAM SRIDHAR
Department of Master of Computer Science
Miracle Educational Society Group of Institutions
Vizianagram– 535216 (AP) India

ABSTRACT:

During previous couple of years, the World Wide Web (WWW) has become an extremely well-liked information source. To successfully utilize the vast quantity of information that the web provides, we want an effective way to explore it. Image data is much more voluminous than textual data, and visual information cannot be indexed by traditional strategies developed for indexing textual information. Therefore, Content-Based Image Retrieval (CBIR) has received an excellent deal of interest within the research community. A CBIR system operates on the visible features at low-level of a user's input image that makes it troublesome for the users to devise the input and additionally doesn't offer adequate retrieval results. In CBIR system, the study of the useful representation of features and appropriate similarity metrics is extremely necessary for improving the performance of retrieval task. Semantic gap has been the main issue which occurs between image pixels at low level and semantics at high-level interpreted by humans. Among varied methods, machine learning (ML) has been explored as a feasible way to reduce the semantic gap. Inspired by the current success of deep learning methods for computer vision applications, in this project, we aim to confront an advance deep learning method, known as Convolutional Neural Network (CNN), for studying feature representations and similarity measures. Extracting the last-but-one fully connected layer from the retraining of CNN model served as the feature vectors for each image, computing Euclidean distances between these feature vectors and that of our query image to return the closest matches in the dataset.

Keywords: World Wide Web(WWW), Content Based Image Reterival(CBIT),Machine Learning(ML), Convolutional Neural Networks(CNN)

INTRODUCTION

In the days of Internet boom where social networks and reasonable smartphones are capable of taking high-quality photos and videos, users have automatic access to several images across the Web. Given these circumstances, the necessity to search, filter and organize the images is a lot more crucial. In the case of small collections, it is possible to search for the specified pictures or duplicates manually. This becomes impractical if the quantity of items increases. To deal with this fast growth there is a need to develop the image retrieval systems that will operate extensively. The main intention is to model a reliable retrieval system that will manage and enquire database of images in a precise manner. CBIR [1] is the strategy of automatically indexing pictures by the extraction of visible features at low-level, like shape, color and texture and these indexed features are entirely responsible for the retrieval of images. In typical CBIR systems (Figure 1) [1], the visible information of the pictures in the database of images is separated and illustrated by multidimensional vectors of features. The vectors of features derived from the pictures present in the database then form a database of features. To fetch similar pictures, the query image is provided by users to retrieval system. Image retrieval system then modifies these query images into a representative model of feature vectors. The resemblance between the query picture's feature vector and the vectors of pictures in the database is then studied, and retrieval is executed with the help of an indexing strategy. The indexing procedure specifies an economical manner to find out similar pictures in the image database. Representation of features and similarity measurements are critical for the retrieval performance of a CBIR system. Various approaches have been suggested, but even then it remains as a challenging task due to the semantic gap present between the image pixels and high-level semantics perceived by humans. One favorable approach is ML that aims to solve this problem in the long-term. Deep learning represents a category of ML approaches where several layers of data processing steps in hierarchical layouts are utilized for classification task and study of features [2]. Deep learning frameworks have attained great achievements in image classification. However, the ranking of similar images is inconsistent with the classification of images. For classification of images, "black boots," "white boots" and "dark-gray boots" are all boots, but for ranking of similar images, if a query image is a "black boot," we conventionally want to rank the "dark gray boot" higher than the "white boot." CNNs [2] are a specific type of ANN for handling data that features a grid-like

opology like, image data, which is a 2D grid of pixels. CNNs are merely ANNs that involvesthe use of convolution instead of conventional matrix multiplication operation in a minimum of one in all their layers. Convolution supports three essential concepts that can facilitate in improving a ML system: parameter sharing, equivariant representations, and sparse interactions. CNNs are eminent for their potential to learn shapes, textures, and colors, making this problem suitable for the application of neural networks. In this paper, we investigated an architecture of deep learning for CBIR systems by applying an advanced deep learning system, that is, CNNs for studying feature representations from picture data. Overall, our approach is to retrain the pre-trained CNN model, that is, Inception-v3 model of GoogleNet deep architecture on our dataset. Then, the trained network is used to perform twotasks: classify objects into its appropriate classes, and perform a nearest-neighbors analysis to return the most similar and most relevant images to the input image [3][4]

LITERATURE SURVEY

There has been a remarkable research in image retrieval systems over the last years. Krizhevsky et al. [5] trained a deep CNN to classify ImageNet dataset consisting of 1.2 million images into 1000 different classes. The authors worked on a network containing eight layers, where first five were convolutional layers, and last three were fully connected layers. Since a single GTX 580 GPU with 3GB memory bounds the maximum network size for training, therefore, this network has been trained on two GTX 580 3GB GPUs. The authors used the features extracted from 7th layer to fetch similar pictures and achieved the top-1 error rate of 37.5% and top-5 error rates of 17.0%. However, because of the highdimensionality of CNN features and inefficiency of similarity computation between two 4096-dimensional vectors, Babenko et al. [6] suggested to compress the features using dimensionality reduction method, and attained a good performance. Deep models have been used for hash learning. Xia et al. [7] proposed a supervised hashing method to study binary hash codes to retrieve images using deep learning and revealed the revolutionary performance of retrieval on datasets that are publicly available. In a pre-processing step, they have used a matrix decomposition algorithm for studying the codes to represent the data. But, this stage iscritical in case of large data as it consumes storage and requires more computational time. Lin et al. [8] proposed a straightforward and efficient supervised learning model for fast image retrieval system using hashing-based methods that project the highdimensional features to low-dimensional feature space and produce the binary hash codes. This approach used binary pattern matching methods or Hamming distance calculation that greatly reduces the computational time and also optimizes the search efficiency. The authors have claimed that Euclidean distance computation between two 4096-dimensional feature vectors requires 109.767ms while Hamming distance computation between two 128 bits binary codes require 0.113ms, thus reducing the time complexity. The most easy way of enhancing the performance of Deep Neural Networks (DNNs) is by increasing the number of layers in the network as well as the number of neurons in each layer. Szegedy et al. [3] presented a deep CNN architecture, Inception, that achieved the state-of-the-art performance for image classification and image detection tasks in the ImageNet dataset. The primary indicator of thimodel is the effective use of computing resources in the network. The authors have increased the width and depth of the network. The architectural decisions are based on the Hebbian principle to optimize quality. This structure helps to increase the number of neurons at each step remarkably without increasing computational complexity in later steps. The improved usage of computational resources permits the increment of the width of each step and the number of steps without getting into computational problems. Chen et al. [9] explored Deep Learning with CNNs with an aim of solving clothing style classification and similar clothing retrieval. To lower the complexity of training, transfer learning is used by fine tuning pre- trained structures on large datasets. Since the parameters are enormous for any deep network, the model is designed to use multiple deep networks trained with a sub-dataset. Compared with the existing approaches that use ML algorithms with shallow structure, this method provided more likely outcomes on three clothing datasets, particularly on the large dataset with 80,000 images where an improvement of 18% in accuracy was recognized. The approach of Khosla and Venkataraman [10] research is to train other CNNs on the shoe dataset and then use these trained networks to classify input shoe image into appropriate shoe class and perform the nearest neighbors evaluation to return K most similar shoes to the given input shoe image. The authors used Caffe as neural network architecture and Euclidean distance metric to return the closest matches to the input image. This approach of computing Euclidean distance between the features vectors of the images has achieved 75.6% precision on retrieval process and an average score of

Keywords: CNN(Convolutional Neural Network,DNN(Deep Neural Network)

PROPSED SYSTEM:

In this project, we investigated an architecture of deep learning for CBIR systems by applying an advanced deep learning system, that is, CNNs for studying feature representations from picture data. Overall, our approach is to retrain the pre-trained CNN model on our dataset. Then, the trained network is used to perform two tasks: classify objects into its appropriate classes, and perform a nearest-neighbours analysis to return the most similar and most relevant images to the input image.

ADVANTAGES:

- Accuracy is Very high.
- The trained Inception-v3 model has been used as a feature extractor rather than as a classifier to retrieve images similar to query image based on Euclidean distance as similarity measure.

CNN

A convolutional neural network (CNN or convnet) is a subset of machine learning. It is one of the various types of artificial neural networks which are used for different applications and data types. A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data.

There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice. This makes them highly suitable for computer vision (CV) tasks and for applications where object recognition is vital, such as self-driving cars and facial recognition.

The CNN is another type of neural network that can uncover key information in both time series and image data. For this reason, it is highly valuable for image-related tasks, such as image recognition, object classification and pattern recognition. To identify patterns within an image, a CNN leverages principles from linear algebra, such as matrix multiplication. CNNs can also classify audio and signal data.

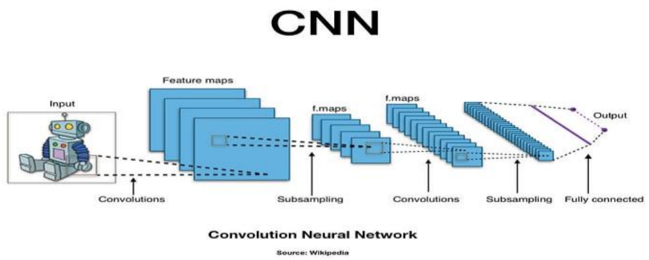

Fig-1

DNN

Deep learning uses neural networks to learn useful representations of features directly from data. For example, you can use a pretrained neural network to identify and remove artifacts like noise from images.

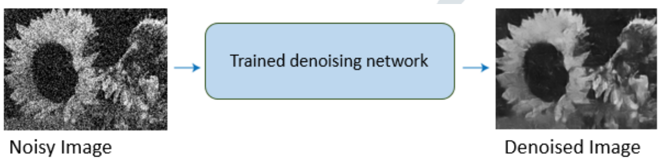

Fig-2

Deep Neural Networks (DNNs) have become a promising solution to inject AI in our daily lives from self-driving cars, smartphones, games, drones, etc. In most cases, DNNs were accelerated by server equipped with numerous computing engines, e.g., GPU, but recent technology advance requires energy-efficient acceleration of DNNs as the modern applications moved down to mobile computing nodes. Therefore, Neural Processing Unit (NPU) architectures dedicated to energy-efficient DNN acceleration became essential. Despite the fact that training phase of DNN requires precise number representations, many researchers proved that utilizing smaller bit-precision is enough for inference with low-power consumption. This led hardware architects to investigate energy-efficient NPU architectures with diverse HW-SW co-optimization schemes for inference. This chapter provides a review of several design examples of latest NPU architecture for DNN, mainly about inference engines. It also provides a discussion on the new architectural researches of neuromorphic computers and processing-in-memory architecture, and provides perspectives on the future research directions.

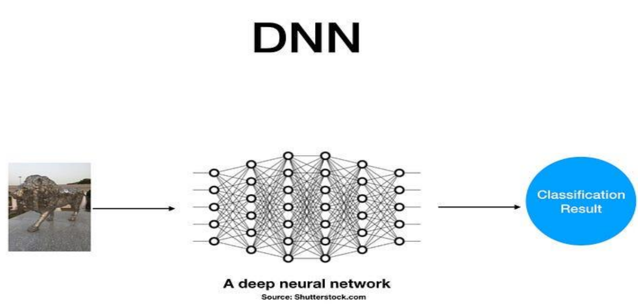

Fig-3

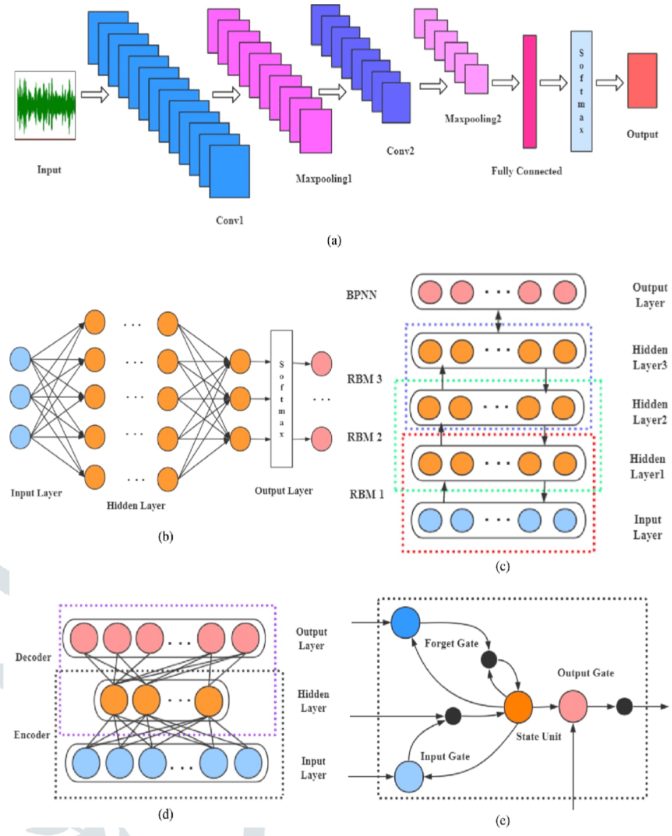

Fig-4

Results

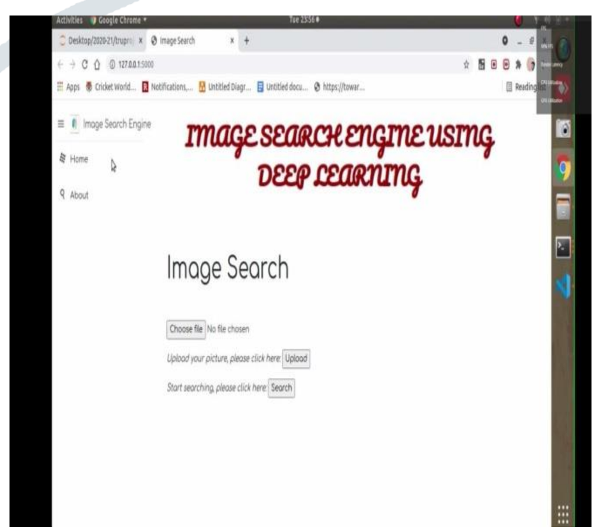

Fig-5

Uploading the picture and starting the picture

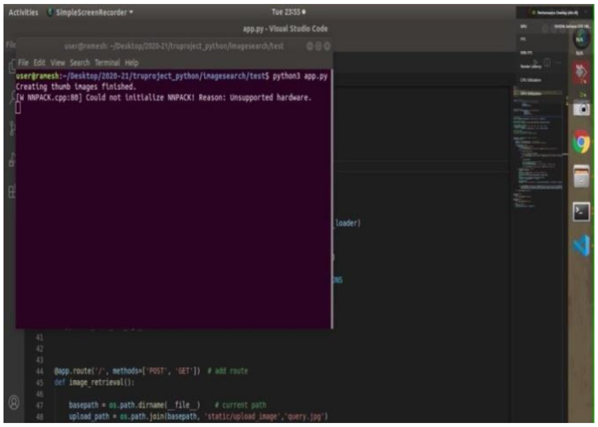

Fig-6

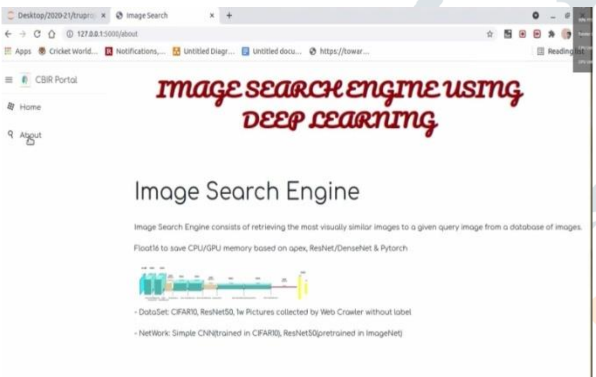

Fig-7

The image search engine will reriev the most visually similar images To the given query as output

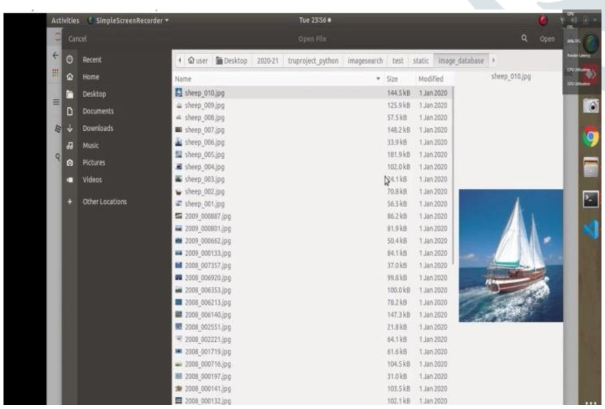

Fig-8

Showing the list of the images which are matching with the query

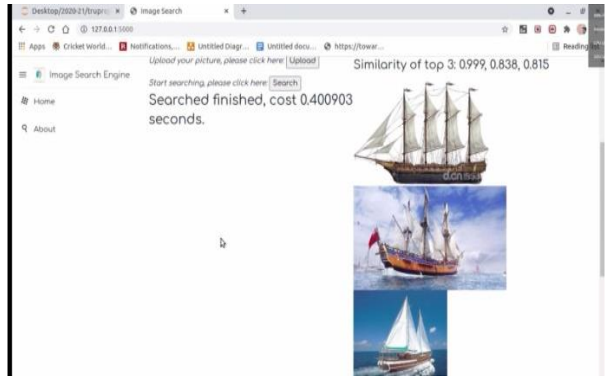

Fig-9

Output showing the list of images,when we query for selecting ship images.

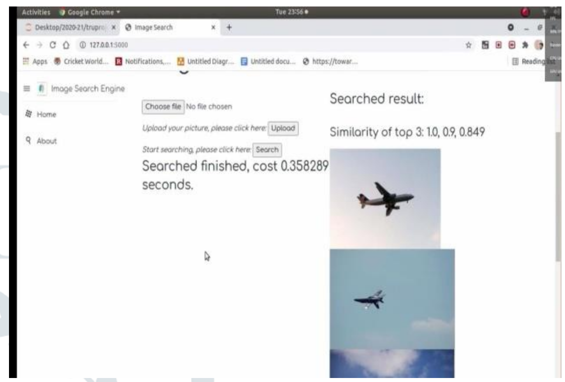

Fig-10

Output showing list of images of aeroplane.

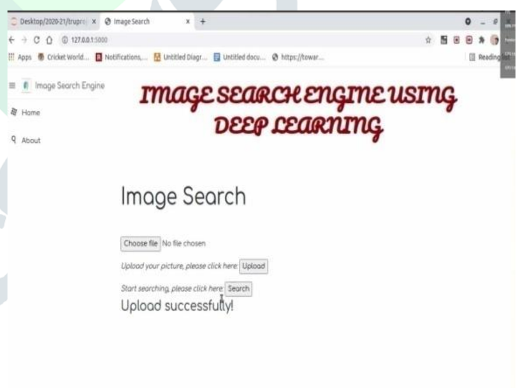

Fig-11

Screen showing the image loaded as query for image search and the upload is successful

Fig -12

Search show the images list as ouput once the image is Uploaded.

CONCLUSION

This paper presents a technique to retrieve similar images using transfer learning. We demonstrated an application of transfer learning on Yahoo! Shopping Shoes dataset and Kitchen Appliances dataset to introduce the concept of retraining the Inception-v3 model released by Google for classification. We achieved top-1 accuracy as 82.12% and 98.78% accuracy for top-10 predictions in correctly classifying shoes images in test data. The accuracy obtained for classification task on kitchen appliances dataset for top- 1 and top-3 predictions are 93.75% and 98.44% respectively. To retrieve images similar to query image, Euclidean distance was used as similarity metric on the last-but-one fully connected layer of the find-tuned inception-v3 model. The Inception-v3 model consists of 22 layers that make the training of this network difficult on the machines having only CPUs and not GPUs. Also, the proposed method of finding an image similar to query image does not always give a semantically accurate result. For the query image of slipons shown in Figure 14, the algorithm also suggests the image of hiking shoes to be similar to slipons. Also, evaluation of retrieval results corresponding to query image is based only on visual perception and have not been evaluated on the basis on any evaluation metric. In future work, the focus will be on the investigation of a various range of publicly available pre-trained deep CNN models for the fine-tuning purpose. Also, the major challenge is to test the approach against photos of shoes taken in real life to see how well they adapt to real-world images. One of the most exciting approaches that can be done is Ensemble Learning, where results of various deep CNN models can be combined to obtain the result. We further hope to refine this work and expand this work into other consumer goods and products

REFERENCES

[1]Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, “A survey of content-based image retrieval with high-level semantics,” Pattern recognition, vol. 40, no. 1, pp. 262–282, 2007.

[2]Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[3]C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[4]C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818– 2826.

[5]A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[6]A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, “Neural codes for image retrieval,” in European conference on computer vision. Springer, 2014, pp. 584–599.

[7]R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, “Supervised hashing for image retrieval via image representation learning.” in AAAI, vol. 1, 2014,p. 2.

[8]K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, “Deep learning of binary hash codes for fast image retrieval,” in Proceedings of the IEEE Coference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 27–35.

[9]J.-C. Chen and C.-F. Liu, “Visual-based deep learning for clothing from large database,” in Proceedings of the ASE BigData & SocialInformatics 2015. ACM, 2015, p. 42.

[10]N. Khosla and V. Venkataraman, “Building image-based shoe search using convolutional neural networks,” CS231n Course Project Reports, 2015.

[11]A. Iliukovich-Strakovskaia, A. Dral, and E. Dral, “Using pre-trained models for fine-grained image classification in fashion field,” 2016.

[12]J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, “Decaf: A deep convolutional activation feature for generic visual recognition.” in Icml, vol. 32, 2014, pp. 647–655.

[13]G. Shrivakshan, C. Chandrasekar et al., “A comparison of various edge detection techniques used in image processing,” IJCSI International
Journal of Computer Science Issues, vol. 9, no. 5, pp. 272–276, 2012.

[14]A. Maurya and R. Tiwari, "A novel method of image restoration by using different types of filtering techniques," International Journal of Engineering Science and Innovative Technology (IJESIT) Volume, vol. 3, 2014.

[15]R. Kandwal, A. Kumar, and S. Bhargava, "Review: existing image segmentation techniques," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 4, 2014.

[16]K. Roy and J. Mukherjee, "Image similarity measure using color histogram, color coherence vector, and sobel method," International Journal of Science and Research (IJSR), vol. 2, no. 1, pp. 538–543, 2013. Textbooks:

1.Programming Python, Mark Lutz

2.Head First Python, Paul Barry

3.Core Python Programming, R. Nageswara Rao

4.Learning with Python, Allen B. Downey

Websites:

•https://www.w3schools.com/python/

•https://www.tutorialspoint.com/python/index.htm

•https://www.javatpoint.com/python-tutorial

•https://www.learnpython.org/

•https://www.pythontutorial.net/