Deep Collaborative Learning Approach Over Object Recognition

Yash Munot

Department of Computer Engineering, R.M.D. Sinhgad School Of Engineering,Warje, Pune, India 411058

Abstract: ***The object recognition using deep neural networks have been most used in real applications. We proposes a framework for recognizing objects in very low resolution images through the collaborative learning of two deep neural networks consists of Image enhancement network Object recognition network. The image improvement network tries to enhance extremely low resolution images into sharper and more informative images with the use of collaborative learning signals from the object recognition network. The object recognition network with trained weights for high resolution images actively participates in the learning of the image enhancement network. It also utilizes the output from the image enhancement network as augmented learning data to boost its recognition performance on very low resolution objects. We verified that the proposed method can improve the image reconstruction and classification performance.***

Keywords: Image acquisition, Image Segmentation, Feature Extraction, Artificial Neural Network.

1. Introduction

The existing work for object detection is totally based on inception models like Alex-net and Google-net through which identification and detection of object is done. The main challenge is to first work for image enhancement network for converting low resolution image to high resolutions. After getting high resolution image then applied object recognition and detection on it. So there is need of object detection system for low resolution image based on machine learning for better accuracy and high reliability. We are going to invent low resolution image object detection and prediction framework based on machine learning and image processing. We are going to work on different type of image dataset. We are using Convolutional neural network for getting high accuracy in object detection. The motive behind this system is to utilize the output from the image enhancement network as augmented learning data to boost its recognition performance on very low resolution objects. Our motivation is to propose method which can improve the image reconstruction and classification performance. The current implementation scope is to develop system for real time object recognition such as real world objects like person, vehicles, animals, fruits etc. The current implementation scope should be used in an efficient solution for to the task of very low resolution object recognition field. The proposed framework can be applied to other low resolution problems, such as faces and letters, which will be done in future studies. To implement this system to detect real world objects using computer vision and machine learning. To develop real world based object and text detection framework which will overcome existing accuracy problem. The goal of our system is to achieve high recognition accuracy for low resolution images. We are going to invent the image reconstruction and classification framework using deep neural network.

Artificial Neural Network (ANN) is going to used for future recognition in which we having the input unit of training data set of different image dataset. Next we have hidden unit which acts upon this training dataset to evaluate the output unit results train model. This entire ANN is works by considering the factors namely matrix feature of images for drafting into a train model for object recognition. We are going to face limitation while working with real time object detection will not give accurate results. In future, we are going to work with real time dataset to overcome this limitation.

2. Literature Survey

1) Very deep Convolutional networks for large-scale image recognition

In this work we investigate the effect of the Convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using architecture with very small (3×3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. These findings were the basis of our Image Net Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively.

2) Going Deeper with Convolutions

They propose a deep Convolutional neural network architecture codenamed Inception that achieves the new state of the art for classification and detection in the Image Net Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. By a carefully crafted design, we increased the depth and width of the network while keeping the computational budget constant.

3) Deep Residual Learning for Image Recognition

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. They provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth.

4) Rethinking the Inception Architecture for Computer Vision

Convolutional networks are at the core of most state of-the-art computer vision solutions for a wide variety of tasks. Since 2014 very deep Convolutional networks started to become mainstream, yielding substantial gains in various benchmarks. Although increased model size and computational cost tend to translate to immediate quality gains for most tasks (as long as enough labelled data is provided for training), computational efficiency and low parameter count are still enabling factors for various use cases such as mobile vision and big-data scenarios. Here we are exploring ways to scale up networks in ways that aim at utilizing the added computation as efficiently as possible by suitably factorized convolutions and aggressive regularization.

5) Inception-v4,Inception-ResNetand the Impact of Residual Connections on Learning

Very deep Convolutional networks have been central to the largest advances in image recognition performance in recent years. One example is the Inception architecture that has been shown to achieve very good performance at relatively low computational cost. Recently, the introduction of residual connections in conjunction with a more traditional architecture has yielded state-of-the-art performance in the 2015 ILSVRC challenge; its performance was similar to the latest generation Inception-v3 network.

6) Delving Deep into Rectifiers : Surpassing Human-Level Performance on Image-Net Classification

Rectified activation units (rectifiers) are essential for state-of-the-art neural networks. In this work, we study rectifier neural networks for image classification from two aspects. First, we propose a Parametric Rectified Linear Unit (PReLU) that generalizes the traditional rectified unit. PReLU improves model fitting with nearly zero extra computational cost and little overfitting risk. Second, we derive a robust initialization method that particularly considers the rectifier nonlinearities.

7) GenLR-Net: Deep framework for very low resolution face and object recognition with generalization to unseen categories

Matching very low resolution images of faces and objects with high resolution images in the database has important applications in surveillance scenarios, street-to-shop matching for general objects, etc. Matching across huge resolution difference along with variations in illumination, view-point, etc. makes the problem quite challenging. The problem becomes even more difficult if the testing objects have not been seen during training.

8) Studying Very Low Resolution Recognition Using Deep Networks

Visual recognition research often assumes a sufficient resolution of the region of interest (ROI). That is usually violated in practice, inspiring us to explore the Very Low Resolution Recognition (VLRR) problem. Typically, the ROI in a VLRR problem can be smaller than 16×16 pixels, and is challenging to be recognized even by human experts. We attempt to solve the VLRR problem using deep learning methods.

9) Fine-to-coarse knowledge transfer for low-res image classification

They address the difficult problem of distinguishing fine grained object categories in low resolution images. We propose a simple an effective deep learning approach that transfers fine-grained knowledge gained from high resolution training data to the coarse low-resolution test scenario. Such fine-to-coarse knowledge transfer has many real world applications, such as identifying objects in surveillance photos or satellite images where the image resolution at the test time is very low but plenty of high resolution photos of similar objects are available.

10)Accurate Image Super-Resolution Using Very Deep Convolutional Networks

They present a highly accurate single-image super resolution (SR) method. Our method uses a very deep Convolutional network inspired by VGG-net used for Image Net classification [19]. We find increasing our network depth shows a significant improvement in accuracy. Our final model uses 20 weight layers. By cascading small filters many times in a deep network structure, contextual information over large image regions is exploited in an efficient way.

3. System Design

In proposed work we are going to invent object detection framework for extremely low resolution image data set as well as high resolution datasets. We are going to train our system into two phases first is to train by using low resolution image data set. Second high resolution dataset trained separately using high resolution images same as low resolution image data count. In last case we are combining low resolution image (LRI) and high resolution image (HRI) dataset to train our model. After training our model drafted accordingly class labels which we are going to train. The whole process is done under network named image enhancement network (IEN). For train purpose we are using neural network and tensor flow framework of machine learning to gain high accuracy over object detection. In testing phase we are giving image of different objects like cars, animals , humans, and real world objects etc. After getting image inputs of low resolutions it will be converted into a high resolution images by using pre defined image enhancement network. After getting high resolution input image get recognized and detected objects from it with the help of our trained model.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

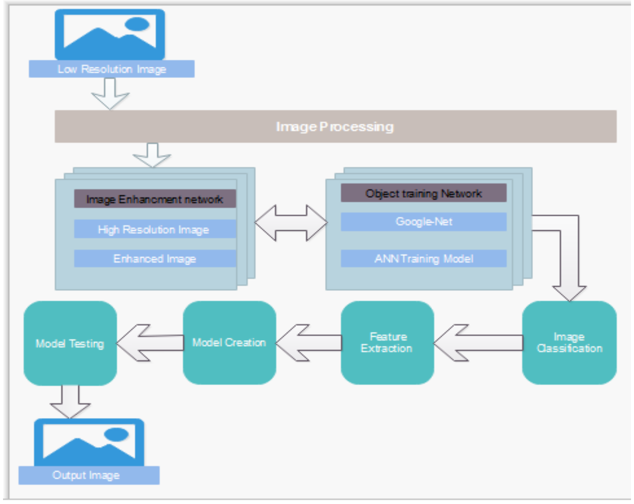

Figure: System Architecture

We are going to develop following modules:

1) Image acquisition

Open-CV (Open Source Computer Vision) is a library of programming functions used for dynamic image processing with computer vision. In our implementation we are going to use open compute vision for taking input images for further processing. After getting desire text or object from images processing applied on it for removing noise from it.

2) Image processing

After getting object image it will send for image processing module. In image processing image gets converted in gray format by removing noise in it using Gaussian filter. After gray conversion image thresholding by setting RGB colour values to zero and preserving only black and white [0 and 1] values. Gray to binary conversion is done by using OTSU's method. After getting black formatted image hand shape get extracted from image. The exact shape of hand will get by drawing edge using canny edge detection method. After processing unit extract exact objet area from input image.

3) Feature extraction

After getting exact shape of object area features get extracted from it by using pixels weight calculations. The image pixels get drafted in matrix by using weight gradient functions only on drawn area of object detected. Feature extraction done on all image dataset for training model creation and drafting. The train model creation done by using deep learning (ANN) algorithm.

4) Feature mapping & model generation

The image dataset is going through image processing and subsequent phases of feature extraction. After getting image features these statistical features get mapped on machine file which is nothing but trained model. The runtime testing image gets matched with pre trained model and respective outcomes will be generated. After outcome generation those results is nothing but our desire object and text recognition results.

4. Existing System Approach

There are different approaches for object detection is being used before. The different object images from dataset are used for further processing. Digital image processing is a field that analyses image processing methods. Mathematically, the image is a formulation of light intensity on two-dimensional field. The image to be processed by a system or computer, an image should be presented statistically with numerical values. A digital image can be stated by a two-dimensional matrix f (m, n) consisting of M columns and N rows. The colour image processing [RGB], there are different models are like hue and saturation, value (HSV) model. This model is used with an object with a certain colour can be identified and to remove the unwanted light intensity from the outside. Further Tests on images performed using six kinds of colours, ie brown, yellow, green, blue, and black and white.

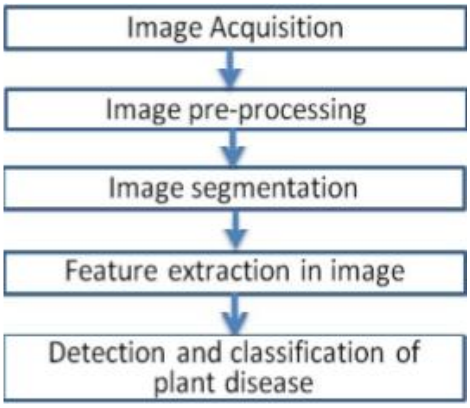

Figure: Existing Approach

a) Image Acquisition

The images of the objects or containing text are getting through the system dataset of text/object images. This image is mainly consisting of color combinations of RGB (Red, Green And Blue) form. color transformation structure for the RGB image is created, and then, a device-independent colour space transformation for the colour transformation structure is applied.

b) Image Pre-processing

To remove noise in image or other object removal, different pre-processing techniques is considered. Image grabbing is done by using grab cut like methods, i.e. cropping of the object image is to get the interested image region. The smoothing filter is generally used for image smoothing. The image contrast enhancement is for smoothing images.

c) Image Segmentation

Image segmentation means is nothing but differentiating the set of images by base of image feature similarities. The segmentation can be done using various methods like otsu' method, k-means clustering, converting RGB image into HIS model etc.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

5. Methodology Used

5.1 ALEX-NET

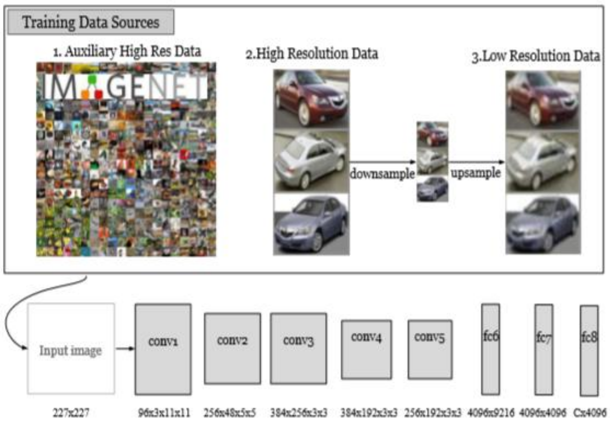

Figure: Alex-Net architecture

They proposed effective knowledge sharing method that increases fine-grained type of classification in very low resolution images. They assume that, even though the test data set has low resolution, they have to use high-resolution class training data. This is a remarkable assumption as it is much handy to label subcategories in high-res image data, and most existing datasets are high-res. They aim to transfer knowledge from such data set store al world scenarios that low resolution. The basic intuition behind their method is to use high quality distinguishing information in the training domain to guide.

5.2 Open-CV

Open-CV (Open Source Computer Vision) is a pre defined library of machine learning and image processing. The programming functions in this library mainly aim at real-time computer vision. In easy language it is library used for Image Processing. This library is primary used to do all the operations related to Image processing.

5.3 Python

Python technology is being actively used to develop machine learning applications. There are many algorithms and many functions that compose or support those algorithms. Open-CV is written natively in C++ and has a template interface that works seamlessly with STL containers.

5.4 Image Processing

Read and write images and detection of images for extracting its features. The object detection is process of drawing boxes around objects in images. The detected objects from image after get processed and get detected by detection task. Image processing works with text recognition in images. e.g. Enhancing image quality, vehicle number plate detection.

6. Conclution

We are going to invent the proposed systematic collaboration between two deep networks can serve as an efficient solution for to the task of very low resolution object recognition. To solve object recognition problem, the proposed framework will be applied to other low resolution problems, such as faces and letters.

7. Future Work

Our future work will used for developing solutions for detection of objects from video contents.

8. Acknowledgment

This work is supported in a Crop Cultivation research field in India. Authors are thankful to Faculty of Engineering and Technology (FET), SavitribaiPhule Pune University, Pune for providing the facility to carry out the research work.

References

[1] K. Simonyan and A. Zisserman, "**Very deep convolutional networks for large-scale image recognition,"** in Proc. Int. Conf. Learn. Represent., 2015, pp. 1–14.

[2] C. Szegedy et al., "**Going deeper with convolutions,**" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 1–9.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "**Deep residual learning for image recognition,**" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "**Rethinking the inception architecture for computer vision**," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2818-2826.

[5] C. Szegedy, S. Ioffe, and V. Vanhoucke, (2016) "**Inception-v4, inception-resnet and the impact of residual connections on learning**," in Proc. Int. Conf. Learn. Represent.Workshop, 2016

[6] K. He, X. Zhang, S. Ren, and J. Sun, "**Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,"** in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 1026-1034.

[7] S. P. Mudunuri, S. Sanyal and S. Biswas, "**GenLR-Net: Deep Framework for Very Low Resolution Face and Object Recognition with Generalization to Unseen Categories,**" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Workshop, 2018, pp. 602-611.

[8] Z. Wang, S. Chang, Y. Yang, D. Liu and T. S. Huang, **"Studying Very Low Resolution Recognition Using Deep Networks**," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4792-4800.

[9] X. Peng, J. Hoffman, S. X. Yu and K. Saenko, **"Fine-to-coarse knowledge transfer for low-res image classification,**" in Proc. IEEE Int. Conf. Image Process., 2016, pp. 3683-3687

[10] J. Kim, J. K. Lee, and K. M. Lee, "**Accurate image super-resolution using very deep convolutional networks,"** in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1646–1654

Author Profile

Yash Munot, B.E. Computer, R.M.D.Sinhagad School of Engineering

Chetan Jadhav, B.E. Computer , R.M.D.Sinhagad School of Engineering

Abhijeet Jadhav, B.E. Computer , R.M.D.Sinhagad School of Engineering

Namita Rode, B.E. Computer , R.M.D.Sinhagad School of Engineering