Classification of galaxies

Problem Statement

The objective of the project is to determine class of a galaxy from its image or video. Galaxy shapes forms the basis of classification of Galaxies. Three major shapes of galaxies are noticed by scientists: elliptical, spiral and irregular. As we know orbit of planets are elliptical due to gravitational force. Similarly galaxies which are group of stars also form shape similar to symmetric curves due to mutual gravitation between stars. Stars are in constant motion with respect to each other due to gravitational pull which cause galaxies to rotate. Irregular galaxies are exception from above circular shapes and do not have any symmetric shape. This is due to lot of elemental Hydrogen or dust in these galaxies. As we also see on earth dust particles do not form any regular shape thus the reason behind shape of irregular galaxies.

Background

The Galaxy classification was given initially by Edwin P Hubble in 1926 and later extended by scientist: Gerard de Vaucouleurs.

Elliptical Galaxies: The spherical form in 3D appears to us as circular shape in 2D. These are classified into E0 which are perfectly circular to E7 which are most flattened. These are brightest at center and brightness diminishes moving away from center.

Spiral Galaxies: These have three major components: bulge, disk and halo. Bulge is central portion which consists of old stars. Arm is the linear portion of stars which are in circular motion around bulge and is made of dust and younger stars. Halo is the spherical part around bulge and covers some part of disk. Arms emerge directly from bulge (ordinary spiral) or from a bar of material around bulge (barred spiral). They are also further classified into lower case letters: 'a', 'b', 'c'.. on how tightly the arms are bound to bulge. The category 'a' have most tightly bound arms.

Irregular Galaxies: These are classified into Irregular I or Irregular II. First category has lot of elemental Hydrogen and young stars. Irregular II has lot of dust that makes the distinct stars not clearly visible.

Reference url: https://www.spacetelescope.org/images/heic9902o/
http://pages.astronomy.ua.edu/keel/galaxies/classify.html

[1] John Kormendy and Ralf Bender explains about S0 galaxies which are intermediate between E7 (elliptical) and Sa (true spiral). They have a bulge and a disk, but no spiral so differs from both elliptical and spiral galaxies and are called lenticular galaxies. Authors make a parallel classification of S0 galaxies to Sa, Sb, Sc and gives them names S0a, S0b, S0c. This classification is done based on B/T ratio i.e. bulge divided by total light. This value decreases from a to c for both spiral and lenticular galaxies.

[2] Ronald J. Buta et al. proposes a new classification of galaxies by the name CVHRS (Comprehensive de Vaucouleurs revised Hubble-Sandage). In this authors classify galaxies in nota-

tions of the form: S<u>a</u>b - An Sab galaxy that is closer to Sa than to Sb, Sa<u>b</u> - An Sab galaxy that is closer to Sb than to Sa and so on for other galaxies.

[3] Lior Shamir describes automatic classification of galaxy images into elliptical, spiral or edge on galaxies. Manually classified images are used to extract image features and given Fisher score. Test images are classified using Nearest Weighted neighbour using the Fisher score as weights. Author finds automatic classification into elliptial, spiral or edge on was done with 90 % accuracy.

[4] Edward J. Kim and Robert J. Brunner describes that most star-galaxy classifiers use reduced information from catalogs, this requires careful feature extraction and selection. With latest advances in machine learning which use deep CNN allows machine to automatically learn the features directly from images and minimizes need for human input. Authors present a star-galaxy classification framework using (ConvNets) directly on galaxy images pixel values.

Methodology

Step 1: Collection of photos of several galaxies.

Step 2: Manually classifying photos in one of three major classes.

Step 3: Simulating the prediction by a CNN image classification. Training data consists of images of galaxies in 3 types into 3 folders which will be read by CNN as 3 types of images.

Step 4: On giving a test image the CNN will automatically tell the type of galaxy as which training class it is matching to.

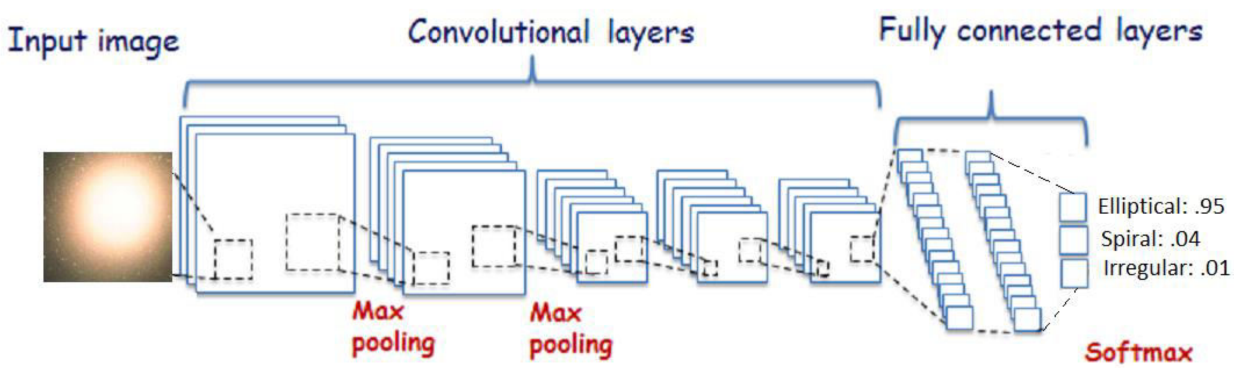

Fig 1: Architecture Diagram of Work Flow of Image classification using CNN

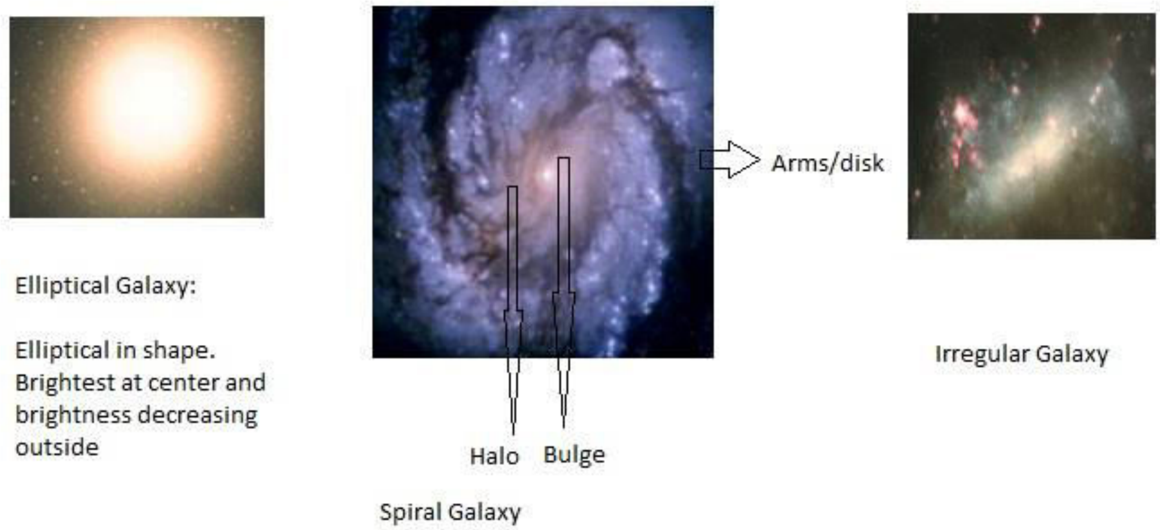

Fig 2: Photos of three types of galaxies

Experimental Design

Dataset:

Step 1: The galaxy photos are collected from url: *http://hubblesite.org/images/gallery* and *https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge/data*

Step 2: CNN is trained using python keras and tensor flow to create a weights files from training dataset.

Step 3: Testing image is given to trained CNN and checked which class it is predicting the image to be.

Evaluation Measures: The test validation set is prepared before testing on a new image and when CNN predicts the class images belong to we count the number of images correctly classified and divide by total number of images to get the accuracy value.

Software and Hardware Requirements: Anaconda with spyder is used for CNN which uses python libraries of keras and tensorflow. The hardware needed will be of multi core fast processor or a GPU machine to train on large dataset with epochs more than 40. This will take training time nearly equal to 1 hour. After saving these weights we get a trained model and this is used to predict new image class.

The CNN can be multi layer with 3-4 hidden layers and 3 classes or categories with Relu (Rectified Linear Unit) activation function. The loss function used will be adams optimizer and categorial cross entropy.

References:

[1] A Revised Parallel-Sequence Morphological Classification of Galaxies: Structure and Formation of S0 and Spheroidal Galaxies, John Kormendy and Ralf Bender, The Astrophysical Journal, 19 Oct 2011

[2] A Classical Morphological Analysis of Galaxies in the Spitzer Survey of Stellar Structure in Galaxies (S^4G), Ronald J. Buta et al., the Astrophysical Journal Supplement Series 2 Jan 2015

[3] Automatic morphological classification of galaxy images, Lior Shamir, Monthly Notices of the Royal Astronomical Society, Volume 399, Issue 3, 1 November 2009

[4] Star-galaxy Classification Using Deep Convolutional Neural Networks, Edward J. Kim and Robert J. Brunner, arXiv:1608.04369v2 [astro-ph.IM] 13 Oct 2016