Convolutional Neural Networks to detect Gravitational Lensing phenomena

The Euclid spacecraft, which is scheduled to launch in 2022, will measure the redshift of galaxies and unearth the expansion of the universe, distances to them, and the contribution of dark energy in the accelerating expansion of the universe. It will do so by leveraging the phenomenon of gravitational lensing. For this purpose, a Convolutional Neural Network (CNN) may prove promising due to its excellent feature extraction capacity from the training set and its humongous applications in the image classification arena. Even though human eyes are adept at recognizing visual particulars and provide precise conclusions, an equally proficient, or at times yielding stupendous results, deep learning systems are a must to keep up with the current data rates, especially in fuel-intensive and cost demanding instances.

Lensing phenomena are informative cosmological models since they allow observation of fainter galaxies and stars and provide paramount insight about the dark matter and dark energy substructures. The non-trivial quandary of lensing phenomena is the inadequacy of training data as they are infrequent. Apart from this, lensing types are variegated, including single and binary sources and lenses. Hence the model must be robust and experienced enough to classify different orientations of lensing systems. The questions of dire importance whose tactful handling would help improve binary lensing classifications are:

1. Will data augmentation help mitigate the issue of insufficient training data, or will it depreciate the performance?
2. Will the transfer learning approach prove to be a boon, or will relatively shallow networks (~5 -10 layers) outperform deeper networks?
3. Can generative models like the General Adversarial Networks (GANs) help to provide significant improvement in learning by manifesting heterogeneous data from the training set?
4. Can traditional data augmentation techniques work well if hybridized with the above promising generative models?
5. Apart from this, does the number of false-positive cases reduce at the end? This is a burning problem since the lensing phenomena are rare and need meticulous inspection before classifying an event as a real lensing phenomenon.
6. Finally, which set of hyperparameters results in the foremost classification accuracy? – A common question in all classification tasks.

Previous approaches included automated algorithms for classifying lenses, although accurate and robust, they fail to handle enormous data. Similar proposals in machine learning, particularly Support Vector Machines (SVMs), logistic regression, and others, have been used. On the other hand, CNN's can work better than almost all algorithms implemented before, by splitting features (color indices) in various filter bands. Previous work has indicated observations using four passbands, giving better overall classification accuracy than observing in a single filter range [1].

The training examples include both verified systems with a lensed source and simulated images of similar lensed systems. Since some simulated systems consisted of larger Einstein radius and fainter reproduction, such images were removed from the original work. This may lead to lower prediction accuracy, but such colossal einstein radius lensing phenomena are easily identifiable [1].

Original CNN network architecture consisted of four convolutional 2D layers with 2×2 max-pooling introduced after the first two convolutional layers, finally feeding into two fully connected dense layers. The final dense layer will have a sigmoid activation because it is a two-class classification. It is worth noting that using softmax or other activations crafted for multi-classification problems may lead to worse results and hence must be avoided. Using Batch-Normalization in between the convolution layers (may prove useful since the batch size is high-250) and ReLU activation, maxpooling after the activations between the convolutional layers, dropout with a drop probability of 0.2-0.25 between the two dense layers may give better results than not using any normalization. Moreover, using an adam optimizer with binary cross-entropy as the loss function (as there are two classes – a binary classification problem) seems viable since an adam optimizer adapts to a reasonable learning rate automatically during training and is a preferred choice by many, although experimenting with other optimizers maybe left as a choice. Using dropout between convolutional layers may not improve the result much. Using a blend of dropout and batch normalization regularization techniques can ultimately prove consequential for validation loss and accuracy.

An exciting attempt will to implement a transfer learning approach and gauge the models' adaptability to different lensing images. All the images would have a range of bright and more or less smeared pixels. Moreover, the model needs to be able to better generalize on new images, adopting a deeper network of around 100-200 layers, familiar in utilizing a pre-trained network, may return satisfactory results.

Utilizing a generative model – GANs- to increase the size of the training set may significantly improve the model's robustness. Running the generator and discriminator network on GPU will save time since both must be trained for a sufficient number of epochs with a lower limit of around 30 epochs. More the number of epochs, the adroiter the generator will be. It is worth attempting to compare results obtained from the simulated images (from software) and using generative models and distinguish which set of images lead to a better classification accuracy or whether not to use any augmentation in the first place, although a deep instinct favors augmentation.

The number of epochs to train will be decided after assessing the training and validation loss and accuracy. A helpful procedure to authenticate this is to plot the losses and accuracies and stick to a number where both are unwavering and converge to an optimal value. To avoid overfitting, a callback function may be used to terminate training after achieving a baseline accuracy of about 90% and loss less than about 0.5.

Model constructed by diligently augmenting images after deciding an optimal image augmentation category, designing layers with regularization techniques, terminating at a desired accuracy and loss, and committing on the most exquisite set of hyperparameters (after constructive trial and error) will develop in a more generalized network which when deployed on any lensing phenomena, will result in a competent accuracy on predictions.

REFERENCES 1. Davies, A., Serjeant, S. and Bromley, J., 2019. Using convolutional neural networks to identify gravitational lenses in astronomical images. Monthly Notices of the Royal Astronomical Society, 487(4), pp.5263-5271.