Effectiveness of Semi-supervised Deep-learning methods for Classification of Galaxies in the Galactic Plane

Gregory Austin
University of Cape Town
ASTGRE002@myuct.ac.za

ABSTRACT

This is a study on the effectiveness of using semi-supervised machine learning for classification of galaxies in near infrared (NIR) surveys in the area of the Galactic Plane. Not many galaxies can be found in the Galactic Plane (also called Zone of Avoidance) due to high obscuration and high star density. There has been successes in detecting galaxies in the Galactic Plane, however, the key problem remains that high resolution data is difficult to search by eye given that it is overly time consuming and laborious. Thus an automated search and classification process is necessary. Semi-supervised deep learning methods were developed for this study and used to classify galaxies. These galaxies are already classified galaxies to test the effectiveness of the semi-supervised methods. The results indicated that the best semi-supervised learning method had a high classification efficacy on the test set (>90%). As an explanation for the high efficacy classification we hypothesize that the Encoder from the autoencoder learnt some useful representations from the comparatively large unlabeled dataset.

KEYWORDS

Machine learning, semi-supervised learning, deep learning, autoencoders, extragalactic cataloging, astronomy

1 INTRODUCTION

Automated search and classification of galaxies in the ZoA is an open research problem. A solution to the automated search problem is using semi-supervised learning [33] (a sub-domain of Machine Learning classification) for classification. Semi-supervised learning is a combination of both unsupervised learning and supervised learning. The Galactic Plane is difficult to search by eye [19] so not many galaxies have been found, as a result there is not much labeled data available. By using a semi-supervised approach, less labeled data is needed because most of the training is done unsupervised (no labeled data required). By using an automated search and classification method galaxies do not have to be found manually by eye which is challenging and time consuming. In order to do automated search in the ZoA, future work can use the best performing classification method in this study. Trained classifiers can be given unlabeled data and return class predictions [22][17], for example the classifier could return '95% galaxy'.

The core components of a semi-supervised learning architecture is a unsupervised method and a supervised method. Autoencoders [1] are an existing unsupervised method and have had successes in representation learning [23] [41] [40]. Representation learning is when features of a dataset are learned, for example, in an astronomical dataset stars and galaxies are features that would be learned. An autoencoder transforms inputs into outputs with the least possible amount of distortion. Given a simple autoencoder, there are three layers: an input layer, a hidden (encoding) layer, and a decoding layer. The network is trained to reconstruct its inputs, this forces the hidden layer to learn good representations of the inputs. Autoencoders can also have a deep architecture where they are stacked and trained bottom up resulting in an autoencoder with more than one hidden layer [10]. Artificial Neural Networks [15] and Convolutional Neural Networks [25] (supervised learning approaches) can be combined with an Encoder from an autoencoder for semi-supervised learning. Convolutional Neural Networks (CNNs) are a type of Artificial Neural Network (ANN) [15] that is structured in way that is ideal for classifying images [22] [35].

The Zone of Avoidance (ZoA) obstructs astronomers' view of extra-galactic space through its high star, dust, and gas density. This makes detecting and identifying galaxies behind the ZoA very difficult [19]. This difficulty is one of the main contributing reasons to why there are currently no good classification methods for galaxy classification in the ZoA. Machine Learning approaches offer the possibility to mitigate this problem of galaxy classification in the ZoA.

1.1 Research Objectives and Hypotheses

The aim of this study is to evaluate the effectiveness of using Semi-supervised Deep-learning methods for Classification of Galaxies in the Galactic Plane. The effectiveness is defined as classification task performance and was measured as the percentage correctly classified in the test sets, refer to section 3.1.4 for more information on datasets. The hypothesis of the study is:

Hypothesis 1. *A semi-supervised deep-learning method for classification of galaxies in the Galactic Plane will yield high classification efficacy with respect to a range of test data from the Galactic Plane. Where high classification efficacy is classification accuracy >90%.*

In order to adequately measure the effectiveness of the semi-supervised methods, they were compared against supervised methods. This study used an autoencoder with a deep architecture for the unsupervised learning component of the semi-supervised learning methods. Previous work that used autoencoders for semi-supervised learning methods show these methods can have a high classification efficacy (> 90%) across a range of problem domains [43][41][40]. It was hypothesized that the same will hold for classification with respect to a test set from the Galactic Plane.

The autoencoder was trained on the near infrared astronomical survey: VVV (VISTA Variables in the Via Lactea) [28] in the Ks filter. The Ks filter captures light with a specific wave length, specifically from 1.990 µm to 2.310 µm. The Ks filter was chosen because the abundance of observations in this filter and extinction is lowest in comparison to other NIR filters. Extinction is the absorption and

scattering of electromagnetic radiation by dust and gas between a point source, for example a galaxy, and the observer [39].

For the supervised learning component of semi-supervised learning methods the study used an ANN and a CNN. When an ANN or a CNN is combined with the Encoder from an autoencoder the result is a semi-supervised learning method. The ANN and CNN were trained using the pre-trained Encoder from the autoencoder as input and the Encoder used the images of galaxies in the Ks filter as input. The ANN and CNN were also trained on their own using the labeled dataset for a comparison of supervised and semi-supervised learning methods.

2 BACKGROUND

The core architecture used in this research was an autoencoder with a deep architecture which was trained on unlabeled astronomical data and an ANN and CNN which were trained on labeled encoder data (the encoder takes astronomical data as input). The ANN and CNN were also trained on labeled galaxy astronomical data for comparison of supervised and semi-supervised learning methods. This section provides some background information on galaxy characteristics, astronomical datasets, and the semi-supervised learning methods which includes the unsupervised learning method and the supervised learning methods.

2.1 GALAXY CHARACTERISTICS

It was important to use galaxies visual characteristics to choose the way that data was extracted and formatted so that the right characteristics were represented clearly and precisely.

Hubble [12] realized there are a small number of morphological types of galaxies that apply to almost all galaxies. This section discusses the fundamental morphological types of regular galaxies: elliptical and spiral galaxies, as well as: irregular galaxies and how to identify each.

2.1.1 Elliptical Galaxies. Elliptical galaxies were named because they look like elliptical blobs of light. Ellipticals have a triaxial smooth shape, the luminosity fades smoothly from the core to indefinite edges [12]. This can be described by equation 1 (de Vaucouleurs law [34]).

$$I(R) = \ln I_e + 7.669[1 - (\frac{R}{R_e})^{0.25}] \quad (1)$$

In equation 1, R is the distance from the center, R_e is the radius of the isophote containing half of the total luminosity. An isophote is a curve on an illuminated surface that connects points of equal brightness. In addition, I_e is the surface brightness at R_e. Using equation 1 it is possible to confirm elliptical galaxies through their light distribution [5]. The light distribution of a galaxy describes how the brightness decreases from center to edge.

Further classification can be made with the degree of elongation of elliptical galaxies. Elliptical galaxies can be round or elongated. Elliptical galaxies have been designated with the symbol "E" followed by a single figure, numerically equal to the ellipticity $(a-b)/a$ where a is the measure of the long axis and b is the measure of the short axis (the decimal point is omitted). The series of ellipticals is E0...E7 with E7 being more flat. There are no elongated ellipticals greater than E7. A problem with ellipticals is identifying their real shape: a flat elliptical may look round if seen from above [12].

2.1.2 Spiral Galaxies. The visual difference between ellipticals and spirals is mainly based on the light distribution, see 1 and 2. Spiral galaxies have a flat rotating disc with a thicker, smooth-shaped bulge in the centre. The central bright nucleus of spiral galaxies appear reddish because it is composed of older stars. The outer disks appear bluish because of the presence of young stars. Spirals can be round (when seen from above) and thus have no ellipticity, or have any ellipticity up to very elongated (when seen from the edge), but in all cases they have a bulge and an exponential disc. Surface brightness in the disk decreases radially outward described by equation 2.

$$I(r) = I_c exp(-r/a) \quad (2)$$

In equation 2, length a is scale factor (a distance over which the brightness drops by a given amount), r is the radius and I_c is the central brightness [20]. Using equation 2 it is possible to confirm spiral galaxies using the light distribution.

There is a fundamental difference between spirals that show an axisymmetrical light distribution from center to edge (SA galaxies) and galaxies whose centers have a luminous bar across the center (barred spiral galaxies, type SB). The difference between SA and SB is only due to the absence (SA) or presence (SB) of the bulge. SA galaxies have a central nuclear region surrounded by spiral features curving out symmetrically. SB galaxies are typically two armed spirals with the arms originating from the luminous bar across the center of the galaxy (the milky way is SB) [20].

Ellipticity is less contributive to the classification of spirals because the ellipticity is only dependent on the orientation of the galaxy to the observer. Instead of ellipticity, conspicuous structural features are used for classification such as: (1) size of central bulge; (2) extent to which arms are unwound; (3) the arm-interarm brightness contrast [12]. Using these structural features subtypes of spiral galaxies can be defined.

There are five subtypes of spiral galaxies to the main types of spiral galaxies (SA and SB): subtype *a*,*b*,*c*,*d*, and *m*. Bulge-to-disk ratio decreases with each subtype, the arms are less tightly wound and open up, and the arm-interarm brightness contrast increases through the sequence. Subtype *a* spiral galaxies have the largest bulge-to-disk ratio, tightest wound spiral arms, and the lowest brightness contrast. Subtype *m* has the smallest bulge-to-disk ratio, the most unwound spiral arms, and the highest brightness contrast [20][12].

2.1.3 Irregular Galaxies. Irregular galaxies show little, if any, symmetry in their structure. The appearance is irregular and asymmetric this is why Hubble [12] classified them as irregular. This irregular shape does not have any consistent features, making such galaxies harder to discern than regular galaxies [36].

2.2 ASTRONOMICAL DATASETS

The FITS (Flexible Image Transport System) file format was designed in order to facilitate the interchange of astronomical image data between observatories. FITS provides a means of transporting arrays and tables of data and keyword and value pairs of metadata.

FITS is defined by standards documents that are approved by the International Astronomical Union and published in refereed journals [32].

2.3 UNSUPERVISED LEARNING

Semi-supervised learning combines both unsupervised and supervised learning. Unsupervised learning learns feature representations of a dataset, for example, galaxies and solar systems from an astronomical dataset. One can utilize these feature representations by using supervised learning [33]. This section describes the advantages of using autoencoders with a deep architecture (unsupervised method).

Unsupervised learning techniques use unlabeled data. Before the work of Le et al. [23] with autoencoders with a deep architecture, only ”edge” or ”blob” detectors were possible using unsupervised methods. Specifically, RBMs (Restricted Boltzmann Machines) [11], autoencoders [10], sparse coding [26] and K-means [4]. The main problem with these methods is that they are time consuming to train and typically require a lot of manual parameter tuning by the researcher in order to perform optimally on a given task and sample dataset. For instance researchers typically reduce dataset size and model size in order to train networks in a practical amount of time [23].

In the work of Le et al. [23] a high-level feature detector was created by scaling up core components involved in training: the dataset, model, and computational resources. Simply put, the model can be viewed as a deep autoencoder with three important ingredients: local receptive fields, pooling and local contrast normalisation. Local receptive fields [30] are used to scale the autoencoder to large images, to do this, each feature in the autoencoder can only connect to a small region of the lower layer. Next, to achieve invariance to local deformations, L2 pooling [7] and local contrast normalisation [16] is used. This could be used for galaxy detection because there are local deformations such as: translated, scaled and out-of-plane rotation of galaxies. L2 pooling and local contrast normalisation allow learning of variant features, L2 pooling in particular [13]. Refer to figure 1 for one stage of the network. The overall structure of the network is nine layers. In the figure, RF is Local Receptive Fields, and LCN is Local Contrast Normalisation. The sizes of each layer in the figure were chosen by Le et al. [23] to maximise representation learning.

In the work of Le et al. [23] the dataset used was 10 million unlabeled 200x200 pixel images sampled from YouTube trained on a large computer cluster. The autoencoder by Le et al. [23] was trained using 1000 machines (16 000 cores) for three days. Experimental results confirm that it is possible to learn useful high level features from unlabeled data. Despite lacking supervisory signals given during training, the best neuron in the network achieves 81.7% accuracy in detecting faces [23]. To test each neuron, its maximum and minimum activation values were found, then 20 equally spaced thresholds in between the minimum and maximum were picked. The reported accuracy is the best classification accuracy among 20 thresholds.

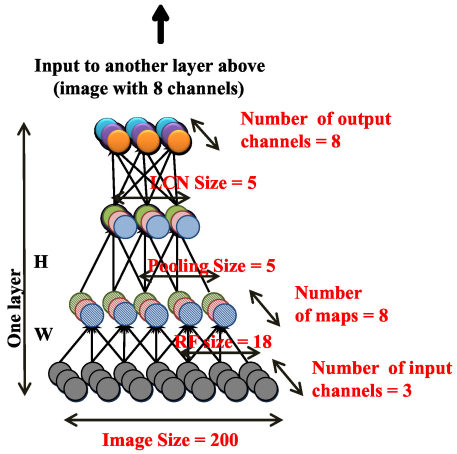

Figure 1: The architecture and parameters in one stage of Le et al.’s network from their study [23]. The overall network replicates this structure three times. For simplicity, the images are in 1D.

2.4 SEMI-SUPERVISED LEARNING

Supervised learning can be used with labeled data and the activations from any unsupervised technique. This way it can use the unsupervised feature representations. This section briefly discusses ANNs and CNNs which are structured in a way which is suited to image classification [22] [35].

Artificial Neural Networks (ANNs) are modelled after the way biological brains work and consist of a collection of nodes organized in layers; input, output, and hidden layers in-between. Nodes are linked to each other by weighted connections and fire signals based on an activation function. ANNs learn to approximate a function to a solution by adjusting their connection weights with regards to training data [15].

CNNs are feed-forward ANNs which means the input and output layers are not directly connected [25]. The input and output layers are connected by a series of interconnected hidden-layers through which input signals are forward propagated. The topology of a CNN can be imaged as a three-dimensional space with multiple interconnected layers. The first is the input layer which feeds to the convolutional layer which breaks the image into regions (filters) and computes the output for neurons in local regions. The rectifier linear unit (RELU) applies activation functions on the regions’ outputs before the pooling layer performs down sampling to produce a feature map [22]. This process may be repeated until a down sampled image can be classified by the last layer called the fully connected layer. The classification depends on the dataset the CNN is trained on and the labels of the dataset.

3 METHODS

This section contains information on the data preprocessing and treatment, and the neural network architecture. This includes details about training data size, test data size, and data transformations, as well as neural network layers and training details.

3.1 Data

3.1.1 Unlabeled Galaxy Dataset. The dataset is from the VVV survey [28]. The VVV survey consists of tiles where each tile is a section of the sky, refer to figure 2. The VVV survey uses NIR filters J, H and Ks, with Ks filter observations being the most common. The tile names start with "b" for bulge and "d" for disk tiles, but only bulge tiles were used in the Ks filter as FITS images. The bulge tiles chosen were tile 328 to tile 338, these tiles were chosen so that there is more data from the Galactic Plane. There are several observations per tile, however in the chosen dataset one observation was chosen per tile with the criteria of best seen observation. The criteria for best seen observation was decided by an expert at VISTA Science Archive (VSA). This data was used to train the autoencoder in section 3.2.1.

3.1.2 Unlabeled Galaxy Dataset preprocessing and treatment. The autoencoder in section 3.2.1 took 96x96 pixel squares as input. Each square was taken from the dataset in slightly bigger slices, for example 110x110, because there was a random crop transformation that crops it to 96x96. The dimensions of each slice was decided so that the galaxies are a better fit in the frame. Dimensions 512x512, 256x256, 128x128, and 96x96 were tried and 96x96 frame sized galaxies yielded the highest autoencoder task performance. Each slice was interleaved horizontally and vertically so that image features were complete. The transformations used were random crop, random horizontal flip, random vertical flip, random rotate, random noise, and the data was normalized. The data was randomly shuffled. The dataset was randomly split into a mutually exclusive training set and test set, 90% and 10% respectively.

3.1.3 Labeled Galaxy Dataset. The dataset consists of images of galaxies labeled from 1 to 9 indicating whether the image is a galaxy (1 being most likely a galaxy, and 9 being least likely a galaxy). In the dataset images range form 200x200 to 3500x3500 pixels and in each image the point source is centered. The dataset is in the Ks filter, this is used because the autoencoder is trained on data in the Ks filter.

3.1.4 Labeled Galaxy Dataset Preprocessing and Treatment. In the dataset label 1 (galaxy) and label 9 (not galaxy) are the most common class, because of this, targets 1-4 are changed to 0 (for galaxy) and targets 5-9 are changed to 1 (for non-galaxy). This way the data is more evenly spread and the classification problem becomes binary classification. All images are cropped to 110x110 pixels and then random cropped to 96x96 pixels. The dataset is small in comparison to standard Deep Learning datasets [23][21][42][22][4]. The dataset is small because of the difficulty and time consuming nature of correctly labeling many galaxies in the ZoA [19]. Data transformations were used to artificially increase the dataset size for better training [9], these transformations were random crop, random horizontal flip, random vertical flip, random rotate, random noise, and the data was normalized. The data was also randomly shuffled. For the training set there were 1725 galaxies and 1006 non galaxies, the training set thus had 63.16% labeled galaxies. The test set had 200 galaxies and 200 non-galaxies.

3.1.5 CIFAR-10. This is a widely used dataset and is used in this study for benchmark task performance comparisons. The CIFAR-10 dataset [21] consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The 10 classes are: airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships and trucks. The classes are completely mutually exclusive.

3.2 Unsupervised Network Architecture

3.2.1 Autoencoder. The details and parameters for each layer of the autoencoder are described in table 1, 2 and 3. A more visual representation of the encoding and decoding layers is given by figure 3 and 4. The input to the network can be 1 channel 96x96 pixel images (galaxy data) or 3 channel (RGB) 32x32 pixel images (CIFAR-10). Pooling is used because it helps the autoencoder learn more robust features, pooling is invariant to local deformations [23]. The convolutional layers are inspired by LeCun et al.'s [24] LeNet which used convolutional layers.

3.3 Supervised Network Architectures

There were two supervised ANNs used in this study. First, a very simple ANN is used to test the learned representation of the Autoencoder. Second, an image detection CNN is used to test the learned representation and as a comparative method. Both ANNs use the same training parameters (Table 4).

3.3.1 Simple Artificial Neural Network. The Simple ANN was set up to take the encoders output as input. Alternatively the ANN can be set up to take raw input. There was only one fully connected layer in the Simple ANN. A fully connected layer is where each neuron receives input from every node of the previous layer. Before the encoder output was given to the fully connected layer, it went through batch normalisation to improve the stability of the network [14]. Batch normalisation normalises the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. Following batch normalisation the output went through the fully connected layer which maps to two outputs (galaxy or non-galaxy). Alternatively the fully connected layer can map to ten outputs for CIFAR-10. Finally the outputs are given to a softmax function. The softmax function is an activation function which is used for multiclass classification [17].

3.3.2 Convolutional Neural Network. The Convolutional Neural Network has a total of 12 layers, 5 convolutional layers, 5 max pooling layers, and 2 fully connected layers. Refer to figure 5 for a visual overview of the net. The CNN uses max pooling where the objective is to down-sample input, reducing its dimensionality and allowing for assumptions to be made about the down-sampled input [7]. Refer to figure 6 for a visual representation of max pooling. Max pooling also helps decrease over-fitting by providing an abstracted form of the representation [8]. After the convolutional and max pooling layers, there are two fully connected layers. The first fully connected layer uses ReLU [29] as an activation function. The final fully connected layer uses softmax. Softmax is an activation function which is used for multiclass classification [17].

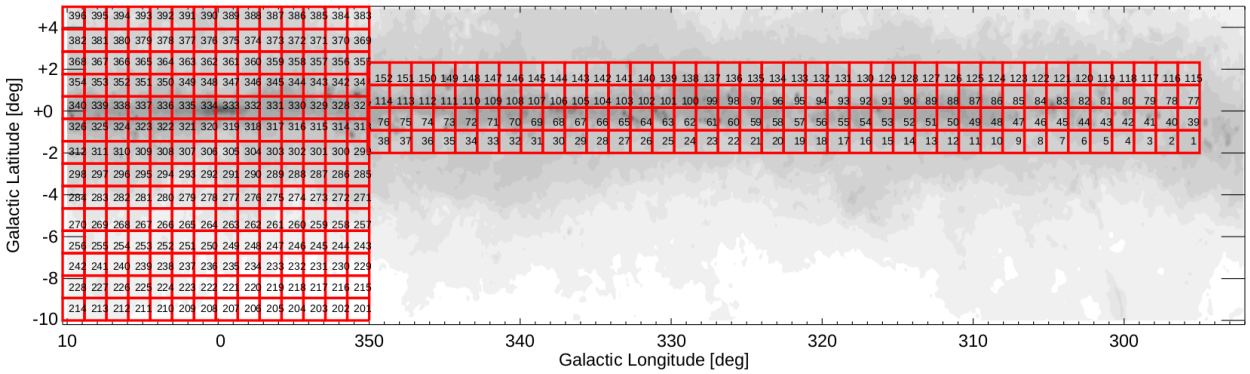

Figure 2: VVV Survey area and tile numbering from the VVV data release [28]. The tile names start with "b" for bulge and "d" for disk tiles, followed by the numbering as shown in the figure.

Parameter	Value
Activation function per layer	ReLU [29]
Network optimizer	Adam [18]
Learning rate	1e-3 [27]
Loss function	MSE [38]

Table 1: Autoencoder training parameters.

Layer type	Input channels	Output channels	Kernel size	Stride	Padding
Conv	1 or 3	8	3x3	1	1
Conv	8	16	3x3	1	1
Conv	16	8	3	2x2	1
Maxpool	8	8	2x2	2	0
Conv	8	4	3x3	1	1

Table 2: The encoding layers of the autoencoder (the encoder) described in section 3.2.1.

Layer type	Input channels	Output channels	Kernel size	Stride	Padding
Deconv	4	8	3x3	1	1
Unpool	8	8	2x2	2	0
Deconv	8	16	3x3	2	1
Deconv	16	8	3x3	1	1
Deconv	8	1 or 3	3x3	1	1

Table 3: The decoding layers of the autoencoder (the decoder) described in section 3.2.1.

4 EXPERIMENTS AND EVALUATION

The experiments in this study were designed to measure the effectiveness of semi-supervised deep-learning methods for classification of galaxies in the ZoA. The classification efficacy was measured with respect to test data from the Galactic Plane. The classification efficacy was also measured with respect to test data from CIFAR-10. CIFAR-10 was used to show the effectiveness of these methods for

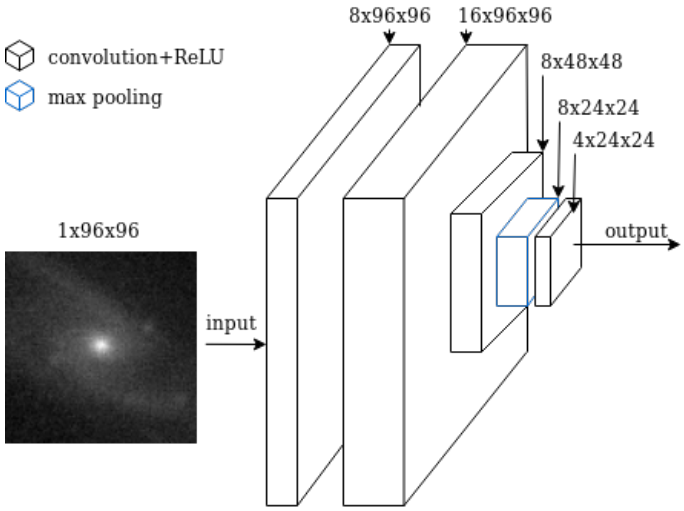

Figure 3: The encoder (encoding network) for the autoencoder with actual input image.

an image recognition task on an established dataset. Each classification model was run 10 times for 50 epochs, the best classification efficacy per run was used for comparison.

There were two different experiments trained and tested on two different datasets. The datasets were the galaxy dataset and CIFAR-10. The two experiments compare a semi-supervised method against a supervised method. The first experiment compared the task performance of the Simple ANN (supervised) versus the task performance of the Simple ANN with the Encoder (semi-supervised). The second experiment compared the task performance of the CNN (supervised) versus the task performance of the CNN with the Encoder (semi-supervised). After training, the task performance is calculated as the percentage correctly classified in the test sets. The task performance is measured for all experiments. These experiments show the effectiveness of the semi-supervised method in comparison to the supervised methods.

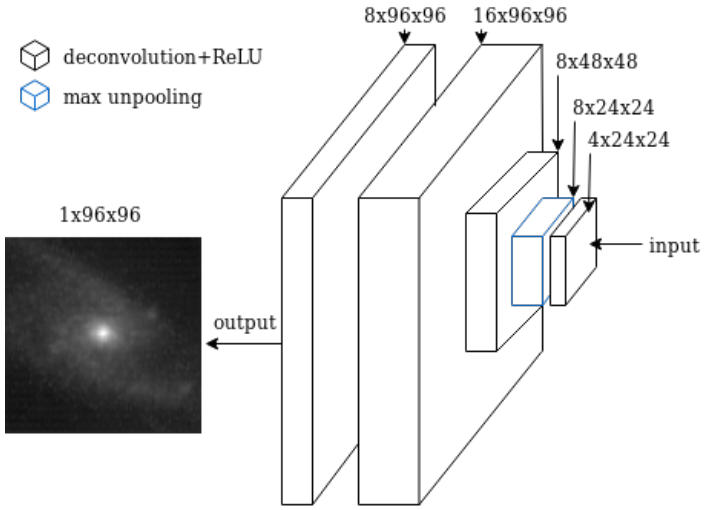

Figure 4: The decoder (decoding network) for the autoencoder with actual output image.

Parameter	Value
Network optimizer	Adam [18]
Learning rate	1e-4 [27]
Loss function	Cross Entropy Loss [31]

Table 4: Supervised networks training parameters.

For the experiments the parameters in table 1, 4 and 5 were used. Tables 1, 4 and 5 specify the autoencoder, supervised and semi-supervised parameters. These parameter values were derived via parameter tuning experiments and were selected since they resulted in effective learning for all experiments. The autoencoder is only run on the unlabeled galaxy dataset for one epoch (refer to Table 5), this is because the unlabeled galaxy dataset is 2.4GB and the method converges in the first epoch.

5 RESULTS AND DISCUSSION

The results of the study indicated that a semi-supervised deep-learning method for classification of galaxies in the Galactic Plane yields high classification efficacy with respect to a range of test data from the Galactic Plane. Where high classification efficacy is classification accuracy >90%. The best semi-supervised method achieved a mean best of 96% on the test data set. This result supports the hypothesis in section 1.1. The results also show that the Encoder + Simple ANN (semi-supervised) is more effective than the Simple ANN (supervised) with respect to both test sets (galaxy and CIFAR-10). In addition, they show that the Encoder + CNN (semi-supervised) is less effective than the CNN (supervised) with respect to the CIFAR-10 test set. However with respect to the galaxy test set the results show that the Encoder + CNN (semi-supervised) is more effective than the CNN (supervised). Refer to table 6 and 7 for results. These results are supported by statistical significance tests (table 8 and 9).

A t-test for two independent means [37] was used to compare the supervised and semi-supervised methods. The t-test used a two-tailed hypothesis with a significance level of 0.05. The t-test was used to determine if there was a statistically significant difference between supervised and semi-supervised learning methods on the same dataset, refer to table 6, 7, 8 and 9.

The results show that the Encoder + ANN is more effective than the ANN on both test sets. The Simple ANN has only one fully connected layer, refer to section 3.3.1, and was expected to be an ineffective classifier. This was expected because a contributing factor to successes in supervised learning is attributed to deeper models [22][6][3]. The Simple ANN was intended to illustrate that the encoder from the autoencoder (unsupervised) learnt some useful representation by performing better than only the Simple ANN. The Simple ANN + Encoder performed better than the Simple ANN on both test sets and therefore learnt some useful representations of the datasets (refer to section 1, paragraph two for how autoencoders learn feature representations). Using these representations the semi-supervised learning method could utilize the data better than a supervised method. This result is supported by previous work having success with autoencoders and representation learning [23][41][40][43]. The topic of future research is the analysis and visualisation of the representations learnt by the autoencoder. The future research will use some of the analysis and visualisation methods used by Le et al. [23].

In support of this study's results, related work that also used autoencoders for semi-supervised learning also yielded high efficacy classification. A dataset called *rect* was used by Vincent et al. in two studies [41][40] that used autoencoders for semi-supervised learning. The dataset consists of 28x28 pixel images and there are a total of two classes. This is an example of a binary classification problem, which is the same as the galaxy classification problem. The goal is to discriminate between tall and wide rectangles. The rectangles are white on a black background. The autoencoders in both studies have high efficacy classification (>90%). It is theorised that the performance of using these autoencoders [41][40] for a binary classification problem support the result of high efficacy performance on the binary classification of galaxies and non-galaxies.

Another dataset used by Vincent et al. in the same two studies [41][40] is *bg-rand*, which is MNIST with added random noise. MNIST [24] is a dataset with 10 classes which are handwritten numbers from zero to nine. Each image in the *bg-rand* dataset is 28x28 pixels with added Guassian noise. Images of galaxies in the ZoA have added noise as a result of the high star, dust, and gas density in addition to added Guassian noise to the training set. The autoencoders in both studies almost achieved high efficacy classification (>90%) with 89.7% [41] and 89.62% [40] with respect to the *bg-rand* test set. The galaxy dataset and *bg-rand* have noise in common. It is theorised that the similar performance on the galaxy test set and the *bg-rand* test set is because of the shared characteristic of noise.

The results of the CNN (supervised) with respect to the galaxy test set show that it is less effective than the Encoder + CNN (semi-supervised), refer to table 6. It is theorised that this is due to the CNN overfitting to the small dataset, refer to section 3.1.4. Overfitting in machine learning is when a trained model corresponds too closely to the training set, and may therefore fail to fit additional

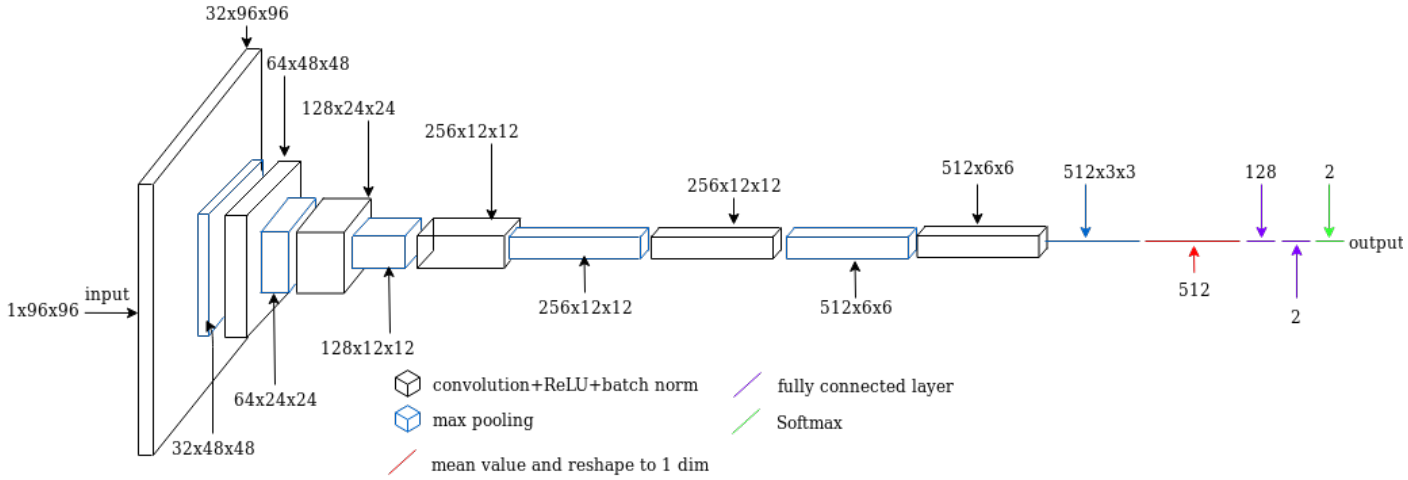

Figure 5: The structure of the CNN described in section 3.3.2. This configuration is for galaxy images.

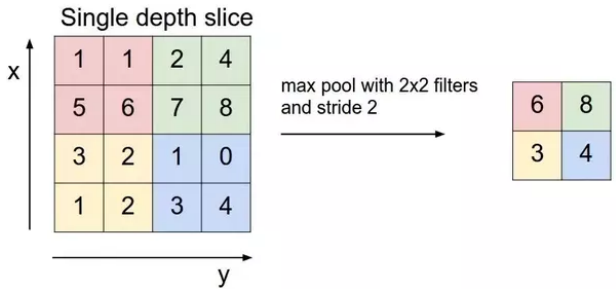

Figure 6: Visual illustration of max pooling.

Supervised and semi-supervised method parameters	
Network runs / epochs	10 / 50
Autoencoder network and experiment parameters	
Runs / Epochs on galaxy dataset	10 / 1
Runs / Epochs on CIFAR-10	10 / 20
Training data transformation parameters	
Gaussian noise	0.01
Random crop probability	1
Random horizontal flip probability	0.5
Random vertical flip probability	0.5
Random rotate 0/90/180/270	0.25
Batch size galaxy dataset	1
Batch size CIFAR-10	128

Table 5: Runs and epochs for different methods and data transformation parameters.

data or predict future observations reliably. In the last epoch of one run the CNN got 58% test set classification efficacy and 96% on the training set. The difference between training and test set classification efficacy in the last epoch is indicative of overfitting.

Typically supervised deep learning methods with high efficacy classification use large datasets [23][21][42][22][4]. Larger datasets help remedy overfitting with supervised learning methods because with more data, supervised models become better at generalising [8]. The comparatively better performance of the CNN on CIFAR-10 (table 7), which is a larger dataset, supports the idea that larger datasets remedy overfitting.

In the last epoch of one run of the Encoder + CNN with respect to the galaxy dataset the model got 94.75% test set classification efficacy and 95.203% on the training set. This difference shows that the Encoder + CNN is a better fit to the galaxy dataset than the CNN. A contributing factor to this result is because the Encoder has learnt some useful representations from the larger unlabeled dataset, refer to section 3.1.1. The Encoder + CNN is thus better at generalising with the smaller labeled dataset than the CNN.

The result of the CNN (supervised) being more effective than the Encoder + CNN (semi-supervised) on the CIFAR-10 test set is supported by the difference between state of the art supervised learning method and state of the art semi-supervised learning method [43][6] with respect to the CIFAR-10 test set. These methods are state of the art at the time of writing according to Benenson [2]. The state of the art supervised learning method achieves 96.53% on CIFAR-10 [6] and the state of the art semi-supervised method achieves 92.75% on CIFAR-10 [43]. This shows that in the field of objects classification the best performing supervised learning method [6] performs better than the best performing semi-supervised learning method [43] with respect to the CIFAR-10 test set (according to Benenson [2]). This supports the result of the CNN (supervised) being more effective than the Encoder + CNN (semi-supervised) with respect to the CIFAR-10 test set.

The CNN was run once with an unnormalized galaxy training and test set and achieved 90%+ on the training set, however only achieved 50% on the galaxy test set. Although there was a high percentage accuracy on the training set the difference between the training and test set showed the model was overfitting. With normalized data, the CNN achieved a mean best of 82.5% with respect to the test set (refer to table 6). These results are supported by previous work on normalizing data and how it can improve neural networks and reduce overfitting [14][8].

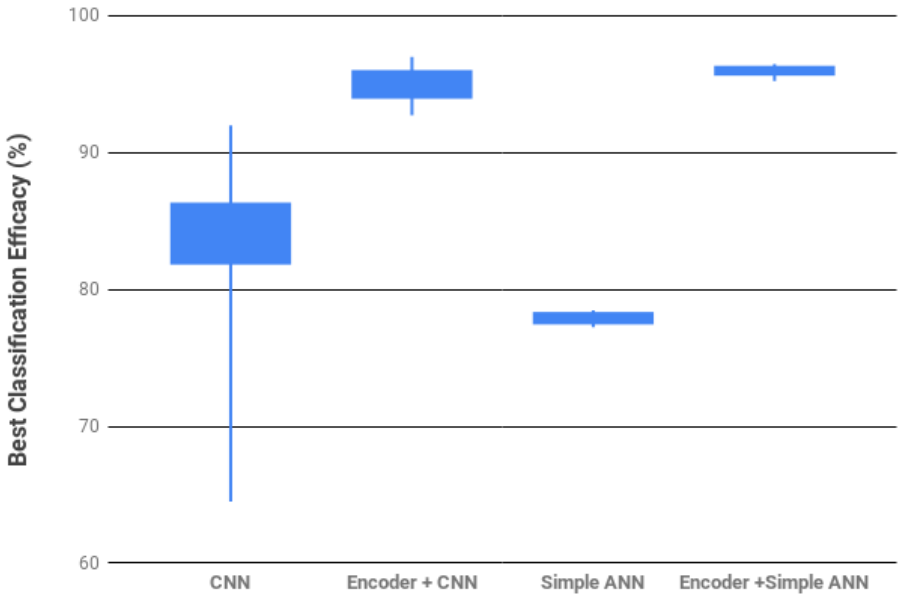

Figure 7: Box and whiskers plot for best classification efficacy with respect to the galaxy test set. Refer to section 3.2 and 3.3 for more details on the the Encoder, CNN and ANN.

Method	min	mean	max
Simple ANN	77.25%	77.9%	78.25%
Simple ANN + Encoder	95.25%	96%	96.25%
CNN	64.5%	82.5%	92%
CNN + Encoder	92.75%	95.03%	97%

Table 6: Highest percentage efficacy with respect to the galaxy test set from 10 runs of each model.

Method	min	mean	max
Simple ANN	40.05%	40.18%	40.32%
Simple ANN + Encoder	50.01%	51.71%	53.05%
CNN	82%	82.63%	82.84%
CNN + Encoder	64.99%	66.635%	67.7%

Table 7: Highest percentage efficacy with respect to the CIFAR-10 test set from 10 runs of each model.

Experiment	T-value	P-value	Significance
CNN vs CNN + Encoder	-5.29	0.00005	significant
Simple ANN vs Simple ANN + Encoder	-93.12	<0.00001	significant

Table 8: Results of t-tests from galaxy experiments.

Experiment	T-value	P-value	Significance
CNN vs CNN + Encoder	50.29	<0.00001	significant
Simple ANN vs Simple ANN + Encoder	-42	<0.00001	significant

Table 9: Results of t-tests from CIFAR-10 experiments.

6 CONCLUSIONS

Automated search and classification of galaxies in the ZoA is an open research problem. A solution explored in this study was automated search using semi-supervised learning [33] for classification in the ZoA. The aim of this study was to evaluate the effectiveness of using Semi-supervised Deep-learning methods for Classification of Galaxies in the Galactic Plane. In order to adequately test the methods, the semi-supervised learning methods were compared against supervised learning methods. Results indicated that the best semi-supervised learning method had a high classification efficacy on the galaxy test set (>90%) and was superior to using a supervised learning method. The semi-supervised learning method (Encoder + Simple ANN) achieved a mean best classification efficacy of 96%. As an explanation for the high efficacy classification we hypothesize that the Encoder from the autoencoder learnt some useful representations from the comparatively large unlabeled dataset. Using these representations the semi-supervised learning method could utilize the data better than a supervised method. This result is supported by previous work having success with autoencoders and representation learning [23] [41] [40] [43].

ACKNOWLEDGMENTS

The author would like to thank Dr Anja Schroeder from SAAO, for the labeled galaxy dataset; Nicholas Cross from VSA, for the unlabeled galaxy dataset; Professor Geoff Nitschke for his help in writing this study, supervision and support; Gerard van Wyk for his advice on Machine Learning architecture.

REFERENCES

[1] Pierre Baldi. 2012. Autoencoders, Unsupervised Learning, and Deep Architectures. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*

Score	Method
96.53%	Fractional Max-Pooling
89%	ImageNet Classification with Deep Convolutional Neural Networks
82.63%	**CNN (section 3.3.2)**
78.67%	PCANet: A Simple Deep Learning Baseline for Image Classification?

Table 10: CNN (section 3.3.2) vs other studies on CIFAR-10.

(Proceedings of Machine Learning Research), Vol. 27. 37–49.

[2] Rodrigo Benenson. 2016. Classification datasets results. https://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html

[3] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. 2015. PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing* 24, 12 (2015), 5017–5032.

[4] Adam Coates, Andrew Ng, and Honglak Lee. 2011. An Analysis of Single-Layer Networks in Unsupervised Feature Learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Vol. 15. 215–223.

[5] Gabriella De Lucia, Volker Springel, Simon D. M. White, Darren Croton, and Guinevere Kauffmann. 2006. The formation history of elliptical galaxies. *Mon. Not. Roy. Astron. Soc.* 366 (2006), 499–509.

[6] Benjamin Graham. 2014. Fractional Max-Pooling. *CoRR* abs/1412.6071 (2014). arXiv:1412.6071

[7] Karol Gregor and Yann LeCun. 2010. Emergence of Complex-Like Cells in a Temporal Product Network with Local Receptive Fields. *CoRR* abs/1006.0448 (2010). arXiv:1006.0448

[8] Douglas M. Hawkins. 2004. The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences* 44, 1 (2004), 1–12.

[9] Alex Hernández-García and Peter König. 2018. Do deep nets really need weight decay and dropout? *CoRR* abs/1802.07042 (2018). arXiv:1802.07042

[10] Geoffrey Hinton and Ruslan Salakhutdinov. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science* 313, 5786 (2006), 504 – 507.

[11] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. 2006. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* 18, 7 (July 2006), 1527–1554.

[12] Edwin Hubble. 1926. Extragalactic nebulae. *Astrophysical Journal* 64 (1926), 321–369.

[13] Aapo Hyvrinen, Jarmo Hurri, and Patrick O. Hoyer. 2009. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.* (1st ed.). Springer Publishing Company, Incorporated.

[14] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15)*. 448–456.

[15] Anil K. Jain, Jianchang Mao, and K. Mohiuddin. 1996. Artificial Neural Networks: A Tutorial. *IEEE Computer* 29 (1996), 31–44.

[16] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. 2009. What is the best multi-stage architecture for object recognition?. In *2009 IEEE 12th International Conference on Computer Vision*. 2146–2153.

[17] Bridle J.S. 1990. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. *Neurocomputing: Algorithms, Architectures and Applications* 68, 1 (1990), 227–236.

[18] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980

[19] R. Kraan-Korteweg and O Lahav. 2004. The Universe behind the Milky Way. *The Astron Astrophys Rev* 10 (2004), 211–261.

[20] Katarina Kraljic. 2012. The two-phase formation history of spiral galaxies traced by the cosmic evolution of the bar fraction. *The Astrophysical Journal* 757 (2012), 14.

[21] Alex Krizhevsky. 2009. Learning Multiple Layers of Features from Tiny Images. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)*. 1097–1105.

[23] Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, and Andrew Y. Ng. 2012. Building High-level Features Using Large Scale Unsupervised Learning. In *Proceedings of the 29th International Coference on International Conference on Machine Learning (ICML'12)*. 507–514.

[24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[25] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. 1999. Object Recognition with Gradient-Based Learning. In *Shape, Contour and Grouping in Computer Vision*. 319–.

[26] Honglak Lee, Rajat Raina, Alex Teichman, and Andrew Y. Ng. 2009. Exponential Family Sparse Coding with Applications to Self-taught Learning. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence (IJCAI'09)*. 1113–1119.

[27] Hao Li, Zheng Xu, Gavin Taylor, and Tom Goldstein. 2017. Visualizing the Loss Landscape of Neural Nets. *CoRR* abs/1712.09913 (2017). arXiv:1712.09913

[28] D. Minniti, P. W. Lucas, J. P. Emerson, R. K. Saito, M. Hempel, P. Pietrukowicz, A. V. Ahumada, M. V. Alonso, J. Alonso-Garcia, J. I. Arias, R. M. Bandyopadhyay, R. H. Barbá, B. Barbuy, L. R. Bedin, E. Bica, J. Borissova, L. Bronfman, G. Carraro, M. Catelan, J. J. Clariá, N. Cross, R. de Grijs, I. Dékány, J. E. Drew, C. Fariña, C. Feinstein, E. Fernández Lajús, R. C. Gamen, D. Geisler, W. Gieren, B. Goldman, O. A. Gonzalez, G. Gunthardt, S. Gurovich, N. C. Hambly, M. J. Irwin, V. D. Ivanov, A. Jordán, E. Kerins, K. Kinemuchi, R. Kurtev, M. López-Corredoira, T. Maccarone, N. Masetti, D. Merlo, M. Messineo, I. F. Mirabel, L. Monaco, L. Morelli, N. Padilla, T. Palma, M. C. Parisi, G. Pignata, M. Rejkuba, A. Roman-Lopes, S. E. Sale, M. R. Schreiber, A. C. Schröder, M. Smith, L. S. , Jr., M. Soto, M. Tamura, C. Tappert, M. A. Thompson, I. Toledo, M. Zoccali, and G. Pietrzynski. 2010. VISTA Variables in the Via Lactea (VVV): The public ESO near-IR variability survey of the Milky Way. *New Astronomy* 15 (July 2010), 433–443.

[29] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10)*. 807–814.

[30] Rajat Raina, Anand Madhavan, and Andrew Y. Ng. 2009. Large-scale Deep Unsupervised Learning Using Graphics Processors. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09)*. 873–880.

[31] Reuven Rubinstein. 1999. The Cross-Entropy Method for Combinatorial and Continuous Optimization. *Methodology And Computing In Applied Probability* 1, 2 (01 Sep 1999), 127–190.

[32] Allen. S. 2018. MIME Sub-type Registrations for Flexible Image Transport System (FITS). https://fits.gsfc.nasa.gov/

[33] Razieh Sheikhpour, Mehdi Agha Sarram, Sajjad Gharaghani, and Mohammad Ali Zare Chahooki. 2017. A Survey on Semi-supervised Feature Selection Methods. *Pattern Recogn.* 64 (April 2017), 141–158.

[34] F. Simien and G. de Vaucouleurs. 1986. Systematics of bulge-to-disk ratios. *Astrophysical Journal* 302 (March 1986), 564–578.

[35] K. Simonyan and A. Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv e-prints* (Sept. 2014). arXiv:1409.1556

[36] N. A. Tikhonov. 2006. Stellar structure of irregular galaxies: Edge-on galaxies. *Astronomy Reports* 50 (July 2006), 517–525.

[37] Kim TK. 2015. T test as a parametric statistic. *Korean Journal of Anesthesiology* 68, 6 (2015), 540–546.

[38] Carlos Toro-Vizcarrondo and T. D. Wallace. 1968. A Test of the Mean Square Error Criterion for Restrictions in Linear Regression. *J. Amer. Statist. Assoc.* 63, 322 (1968), 558–572.

[39] R. J. Trumpler. 1930. Preliminary results on the distances, dimensions and space distribution of open star clusters. *Lick Observatory Bulletin* 14 (1930), 154–188.

[40] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th International Conference on Machine Learning (ICML '08)*. 1096–1103.

[41] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* 11 (Dec. 2010), 3371–3408.

[42] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao. 2015. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop. (June 2015). arXiv:1506.03365

[43] J. Zhao, M. Mathieu, R. Goroshin, and Y. LeCun. 2015. Stacked What-Where Auto-encoders. *ArXiv e-prints* (2015). arXiv:1506.02351