Neural Nets to Score Colliding Galaxy Models

Shawn Mace
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
sdm2w@mtmail.mtsu.edu

Aric Moilanen
Department of Physics and Astronomy
Middle Tennessee State University
Murfreesboro, United States
atm4w@mtmail.mtsu.edu

Matthew Ogden
Computational Science, PhD Program
Middle Tennessee State University
Murfreesboro, United States
mbo2d@mtmail.mtsu.edu

Derrick Reckers
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
dmr3p@mtmail.mtsu.edu

Jarrett Shaver
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
Jarrett.shaver@gmail.com

Devon Wilson
Department of Computer Science
Middle Tennessee State University
Murfreesboro, United States
djw4s@mtmail.mtsu.edu

Abstract—**Information that best describes galaxy mergers is sparse. Complications arise when forming postulations due to missing information such as orbital mass-ratios, orbital velocities, and orientations. To remedy this, we devised two neural networks to analyze images and human scores from Galaxy Zoo: Mergers in the hopes that a well-fitting model - one that can find true dynamic parameters for galaxy collisions. In the single layer neural network, the information is parsed to inform the convolution net about hyperparameters such as batch or epoch sizes. The convolution model is designed to assign a fitness score to an input galaxy image. That score represents an approximation of how well the image fits into predicted galaxy collision simulations. Through the limited scope of this project there are points of interest. Both networks achieved successful validation while eventually overfitting. The convolution net managed with good accuracy to predict scores of mergers when compared to human grading and, perhaps, more accurate than the human scores. Though the data set we used is small and unevenly distributed, there are hopeful interpretations of the results. Better and more data sets along with more network layers may yield better results in the future. With more tuning, our neural networks may eliminate the need for further human rankings**

Index Terms—**convolutional neural networks, single layer neural networks, CNN, regression model, galaxy collisions, image scoring**

I. Introduction

When two galaxies merge, it causes a drastic change in the evolution of the resulting galaxy. Understanding ongoing collisions in the night sky will aid scientists in their exploration to uncover more about merging galaxies. Unfortunately, information such as orbital velocities, orientations, and mass-ratios between the two galaxies cannot be obtained from observation alone.

To obtain this information, past research has looked into modeling the dynamics of interacting galaxies and developing galactic models to simulate galaxy interactions. By finding models and analyzing how well they fit known observations of galaxies, models that best fit observations could reflect the true parameters for the galaxy collisions.

This project takes data from galactic models and associated human fitness scores, with the goal to develop both a single layer neural net and a convolutional neural net that will help scientists better understand this data. These two nets will be used to produce a machine fitness score for each model image that we hope will closely match its respective human fitness score. We will compare these two networks based on how well they were able to match human scores with their predictions. By having theses tools like these neural networks, scientists can quicken the search to find those well-fitting models, explore the galactic model parameter space, and hopefully find the true dynamic parameters for observable galaxy collisions.

II. Background

Our project is based on several pieces of research. First is the work of Toomre and Toomre who demonstrated merging galaxies would cause distinct tidal distortions [1]. These tidal distortions often result in the galaxies forming bridges and tails, star regions that appear to bridge the gap between the two galaxies, and arc away from each other. Toomre and Toomre performed a variety of simulations showing that the gravitation interactions caused these distinct shapes [1].

Second is the work of SPAM. SPAM is O(n) public release software that builds upon the restricted three-body model to simulate galactic models merging. SPAM takes galactic models and simulates the two galaxies merger. In addition, point particles are added, undergoing realistic tidal distortions and produce the iconic bridge and tail shapes [2].

Lastly, most of our data has origins from Galaxy Zoo: Mergers. Galaxy Zoo: Mergers is a citizen scientist effort to explore the unknown parameter space for the galactic models. Using the quick efficiency of SPAM to visual the final shape of galactic models, millions of galactic models were simulated and views by thousands of volunteers. Citizen scientists would select models with a similar shape to the target image the models are attempting to represent. From there, those selected models went through a tournament like competition with other

models to receive the human fitness score. The citizen scientist would choose the galactic model whose shape better matches the target images. After going through the tournament, all the models would have an associated fitness scores [3].

The galactic models and their associated human fitness score are what were used in this project. The human fitness scores were directly applied, while the galactic models had to be visualized again. The galactic models were sent through SPAM again with a larger number of particles to obtain a higher resolution image of the colliding galaxies. These model images and the human fitness scores are used [2] [3].

III. METHODS

A. Data
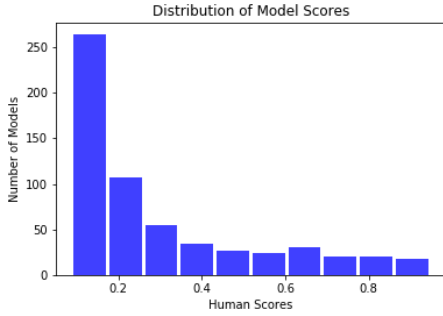
Fig. 1. Distribution of sample models across a range of human scores from 0-1.

We chose one of the 62 pairs of galaxies and pulled the associated data. This set consisted of 1294 grayscale model images in .png format and their associated human scores in a text file. The set was already sorted in descending order by score. However, we realized that this data was heavily skewed towards worse models, with more than half of the models having scores below 0.10. In an effort to combat overfitting on bad model images, we used only the first 600 images, which, while still skewed towards worse models (see Fig. 1), presented a much more evenly distributed data set.
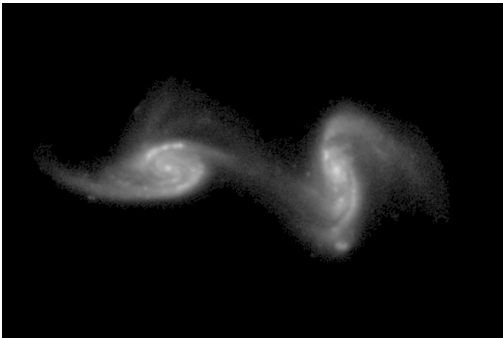

Fig. 2. The chosen colliding galaxy pair for this project. A higher model score should represent a model that closely resembles this image.

The images and scores were read in using a technique similar to the MNIST problem presented in OLA 6 [4].

The main difference being that the grab_image function was modified to read in our images as greyscale rather than full RGB and at a reduced resoultion of 100x100. We then shuffled the images, taking care to manipulate them in such a way that the scores would retain the same index as their respective images. This shuffling ensured that models of similar scores would be randomly distributed, preventing the model from learning soley on one quality of model at a time.

The 600 shuffled images were then split into a training and validation set consisting of 510 images and a testing set with the remaining 90 images. When training our nets, a validation split of 0.3 was used, resulting in the nets training on 357 images and validating on the remaining 153.

B. Single Layer Neural Net

For the single layer net, the images were flattened to 1D arrays then passed into a single dense layer. This layer was given an input size of 1 and a Softmax activation function in order to return a single score between 0 and 1. This net, and the nets discussed below, were all compiled with mean squared error, which is the best fit for regression problems, as the loss parameter.

In training our nets, we found that a small batch size was an absolute necessity for decent learning, most likely due to the small data set and skew towards worse models. We used a batch size of 4 for all of the nets discussed in this project.

C. Convolutional Neural Net

In our second net, we attempted to utilize the superior image recognition characteristics of a convolutional neural net. As a base, we used the convolutional neural net used on the cat problem in OLA 6 [4]. In this net, there are two 2D convolution layers that feed into a pooling layer. The main changes made were to kernel size and the final dense layer. There are two 2D convolution layers that use a standard ReLu activation function, which are then followed by pooling and dropout layer. It is then passed into a flattening layer and a single dense layer, still with a ReLu activation function, before passing though a final dropout layer and into the final dense layer. The main changes we made to this architecture from OLA 6 were in terms of kernel size, unit size of the dense and convolution layers, and the setup of the final dense layer.

We adjusted the kernel sizes to be slightly larger. This was done in an attempt to better capture what we believed were the most important characteristics of the image, namely the tidal distortions around the edges of each galaxy and the bridge, or lack thereof, between the galaxies. The unit sizes in the convolution layers were made smaller in order to fit our relatively small data set. Without doing this, we experienced drastic overfitting or no training at all. The final dense layer was also changed to match our plan for a regression model. The input for the final layer was reduced to 1 and the activation changed to Softmax in order to return a single score between 0 and 1 for the model image.
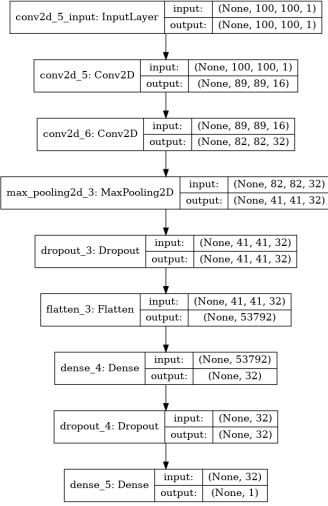
Fig. 3. Layout of the convolutional neural net

IV. RESULTS
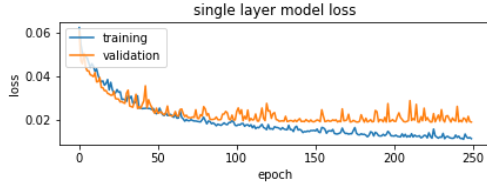
A. Single Layer Net Results

Fig. 4. Loss (Mean Squared Error) for the single layer net over 120 epochs

Shown in Fig. 4 is the training and validation loss history over 120 epochs. At the end of 120 epochs, the model had decreased to a MSE loss of 0.012 for the training data and 0.019 for the validation data. We can also plainly see the overfitting that plagued both of our neural nets, regardless of the hyperparameter tuning we performed to try to remove it.
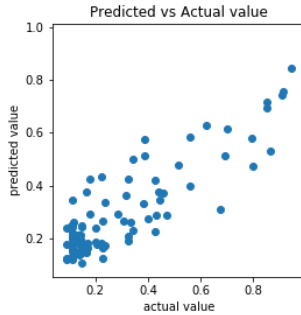
Fig. 5. Predicted values generated by single layer net vs actual value

Fig. 5 shows the values predicted by the single layer net on the 90 testing images verus their generated scores. We see a linear correlation between the two, which is what we hoped

to observe. A perfect linear correlation would have meant that the model was guessing exactly the same as the human score.
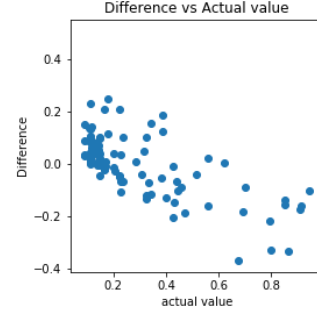

Fig. 6. Difference between predicted values generated by single layer net and actual values vs actual value

Fig. 6 shows the difference between the values predicted by the single layer net for the 90 test images and their human scores versus the human scores. From this plot, it is easy to see that the net had a tendency to overscore models with low human scores and underscore models with high human scores. We can also see that the most accurate predictions occured for models with lower human scores, which makes sense given the data's skew towards worse models. Over the full 90 image set, the net attained a mean difference of 0.093 between its predicted values and the human scores.
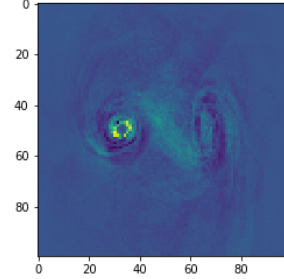

Fig. 7. Visualization of weights for the single layer net

The final weights for the single layer net are shown in Fig. 7. We see that the net is placing emphasis on the bridge forming between the two galaxies as well as on the tidal distortions around the edges. Interestingly, the net also thinks that the core of the left galaxy is extremely important. This may be due to the fact that all of our model images are positioned roughly the same way, and in every image there is a bright spot marking the center of a galaxy at roughly the hotspot location shown in Fig. 7.

Shown in Fig. 8 is the model that the single layer net scored the highest (0.846). Its real human score was 0.946, so this shows good agreement between the net's perception and human perception of what the best model is. We can also visually verify that this chosen model closely resembles the target image shown in Fig. 2.
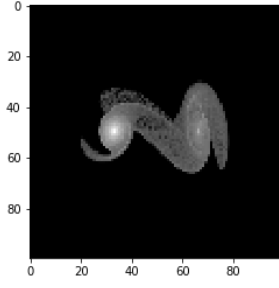
Fig. 8. Model that was scored the highest by the single layer neural net
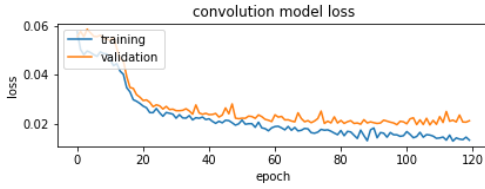
B. Convolutional Neural Net Results

Fig. 9. Layout of the convolutional neural net

In Fig. 9 we see very similar learning behavior to the single layer net. Both display overfitting issues and, after 120 epochs, reach approximately the same MSE. For the convolutional net, the final MSE for the training data was 0.013 and 0.021 which is just slightly worse than the single layer net. Qualitatively, the only difference in learning behavior was the prescence of a slight plateau for the fist 10 or so epochs before the net actually began to learn quickly.
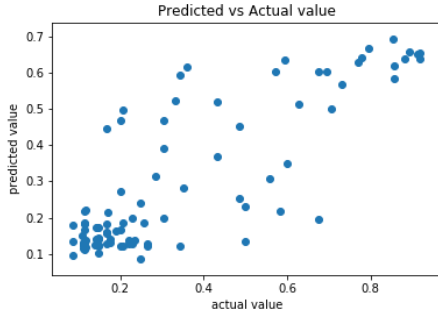

Fig. 10. Predicted values generated by convolutional net vs actual value

Like with the single layer net, Fig. 10 shows a linear correlation between the predicted values and human scores. However, as the human scores increase, we see a wider spread in predicted values when compared to the single layer net.

Once again, Fig. 11 shows that the convolutional net has a tendency to overscore worse models and underscore better models. However, when compared to the single layer net, we see a much tighter grouping of worse models (human scores
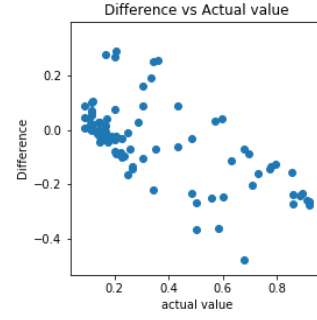

Fig. 11. Difference between predicted values generated by convolutional net and actual values vs actual value

≤ 0.3) that the model was able to accurately guess with a difference of $-0.1 \leq \text{Diff} \leq 0.1$. For the 90 image testing set, the convoltional net attained a mean difference of 0.112, which is slightly worse than the single layer net.
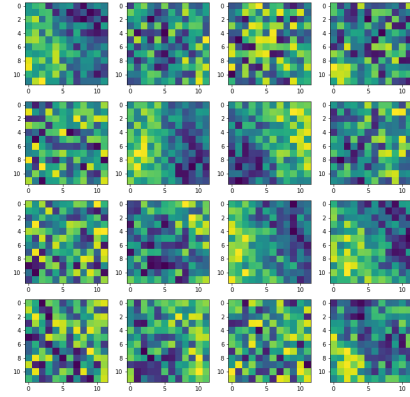

Fig. 12. Visualization of weights for the first convolution layer

Fig. 12 and Fig. 13 show the weights for the first and second convolution layers respectively. However, it is hard to glean and useful information from just these weights on their own. That is why we have shown the activations of units in the first layer in Fig. 14. We can see that for this specific activation, the net was processing a model that was very similar to the target image. We can see units that are picking up the major features: galactic cores, tidal distortions, and the bridge. We can also plainly see the units that are detecting the edges of the galaxies, as well as some units that aren't really being utilized.

Fig. 15 shows the image that was given the highest score by the convolutional net (0.695). This predicted value is significantly lower than the human score of 0.853, but it shows that the convolutional net was still atleast somewhat successful in deciding on what the best model should look like. Once again, we can visually verify that the convolutional
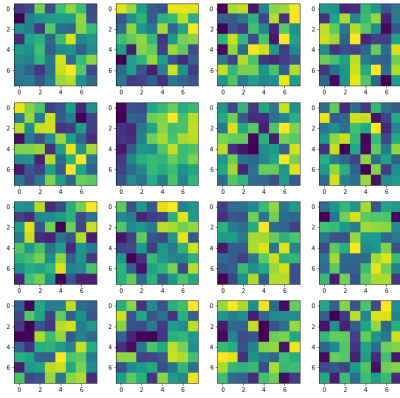
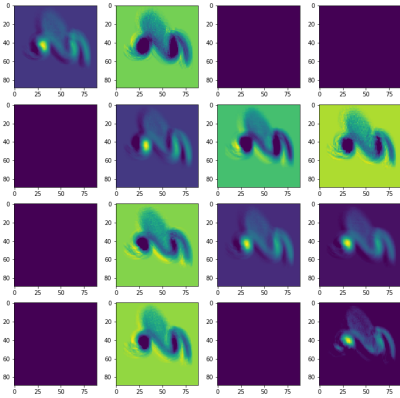Fig. 13. Visualization of weights for the second convolution layer

Fig. 14. Visualization of network activations for the first convolution layer
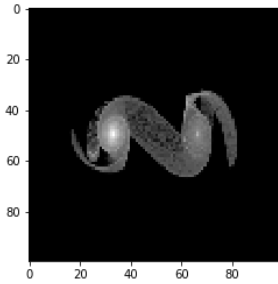

Fig. 15. Model that was scored the highest by the convolutional neural net

net's predicted best model closely resembles the target image shown in Fig. 2.

V. DISCUSSION

Both the single layer neural network and convolutional neural network show promise in correctly predicting how good of a fit a model is for a target galaxy collision. Both nets achieved a MSE of ≈ 0.02 on the validation data and a mean difference of ≈ 0.1 when tasked with predicting the scores of 90 test models. Both nets successfully chose an image out of a set of 90 that they decided was the best model for the target image. Upon visual comparison, these two chosen models very closely resemble the target galaxy collision. However, both models struggled with overfitting which can most likely be attributed to both our data set's small size and its uneven distribution of model quality.
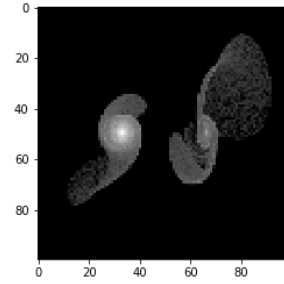

Fig. 16. Example of a model that the convolutional net scored "poorly" when just comparing against human score.

Another factor to consider is the noise associated with the human scores. As a citizen scientist project, the models in the Galaxy Zoo: Mergers data set were ranked by people with minimal training. There are also no checks in place to prevent against people clicking though models without consideration, artificially inflating or deflating their scores. We must also acknowledge the randomness of the tournament system. It is entirely possible that some models were "unlucky" and went against rounds of exclusively good models, causing them to be scored lower than they actually should have or vice versa.

Take for example the model shown in Fig. 16. This was one of the convolutional net's worst predictions, with a difference of nearly 0.5 between the net's predicted score of 0.195 and the human score of 0.676. However, if we visually compare this model with the target image, one could argue that the model did not deserve such a high human score. Instead, the model might better represent a score between the net's prediction and the human score.

Future work could pertain to the expansion of the convolutional net to include more layers. Training with these extra layers and units would necessitate a larger data set, hopefully with a more even distribution of model quality. We believe that such a net would obtain better results than either of the nets described in this paper. Also, the nets we created for this paper could be applied to the other 61 galaxy pairs, and, with

enough training, possibly eliminate the need for further human rankings. This would allow for new models to be created and scored much faster than they have been previously.

REFERENCES

[1] A. Toomre and J. Toomre, "Galactic bridges and tails," *The Astrophysical Journal*, vol. 178, pp. 623–666, Dec. 1972.

[2] J. Wallin, A. Holincheck, and A. Harvey, "Jspam: A restricted three-body code for simulating interacting galaxies," *Astronomy and Computing*, vol. 16, 11 2015.

[3] A. Holincheck, J. Wallin, K. Borne, L. Fortson, C. Lintott, A. Smith, S. Bamford, W. Keel, and M. Parrish, "Galaxy zoo: Mergers - dynamical models of interacting galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 459, no. 1, pp. 720–745, Jun. 2016.

[4] J. Phillips, "Open lab assignment 6." [Online]. Available: https://www.cs.mtsu.edu/ jphillips/courses/CSCI4850-5850/