J.N. Andrews
Honors Program
Andrews University

Proposal for Senior Honors Thesis

HONS 497 Senior Honors Thesis **Credits** 2 (2 minimum required)

Directions: Please return signed proposal to the Honors Office **at least one week prior to your scheduled meeting with the Honors Council.** This proposal must be accepted by Honors Council the semester before presentation.

Student's Name: Solomon Kim

Primary Advisor: Dr. Rodney Summerscales
Secondary Advisor:
Thesis Title: Generative Adversarial Neural Networks for Labeled Facial Creation
Local phone: 630-636-8518 Email: ksolomon@andrews.edu

Expected date of Graduation:

1. Provide goals and brief description of your project or research.

My proposed research focuses on generating realistic faces from labeled descriptions of emotion and facial features. These labels will generally include verbal descriptions of the emotions portrayed by the face, for example, happy or sad, and descriptions of the features of their face, for example, their skin tone, the size of their nose or shapes of their lips. If we were to give the program the input of, "A young, white man who is slightly overweight with glasses, blue eyes, and a small nose" the program would expectedly output an image of a face corresponding to those descriptors that we had given as input.This type of facial generation has applications that can contribute to both the real world and the field of computer science. Using labeled descriptions to direct the generation of realistic faces has not been researched extensively, leaving an opportunity for new findings and techniques. Neural network construction relies heavily on personal preference and experimentation, meaning that my research, even if previously conducted in a similar fashion, would be a contribution to the overall field given that my methods and techniques would most likely differ from another researcher. In terms of my project's contribution to the real world the main application would be to automize police sketches. Police sketches are currently fairly inaccurate and do not have any obvious path to improvement. With my research a victim could simply insert labels and be given several different outputs, or images of faces, to choose from. There are similar services that have been very successful, but these services are based off of mixing different preset features and creating a composite image, not artificial intelligence based. While there is already research out there about text to image creation I believe that there is an opportunity for improvement upon current results' accuracy and efficiency.

I plan to use Generative Adversarial Networks (GAN) to improve upon past research in this field. In order to use a GAN we must define a couple of variables from the outset of the research. The first variable, or input that we will be giving the GAN will be the labels, or descriptions of the faces that we want to create. These features are the roadmap for the GAN to begin creating an image out of random noise. These labels act as a guide in a similar way that an actual police sketch artist would use these labels as guides to their depiction. The images that we are going to generate will be smaller images, typically around 250 x 250 pixels. We are aiming to create realistic looking faces that are distinguishable from other faces based on the initial labels given.

II. Outline your methodology. **Please be specific**. How does this achieve your goals and how reliable is it?

To understand GANs we must first understand neural networks. In a traditional neural network you have many different variables, or inputs, that will be fed into various layers of calculations. From these calculations a final answer may be given to a specific problem. For example, in figure 1 we feed our network a picture of a cat.This would be considered our input layer. After the network has an input it begins to make guesses in each of the hidden layers. There are many types of hidden layers that may perform various different types of calculations, but in general the concept remains constant. These hidden layers are meant to hold individual weights, or probabilities. Once the network as computed each node, the circles in the graphs which represent weights, it will compile those probabilities into an output layer. The network will then return a probability of whether or not the input is a cat or a dog.

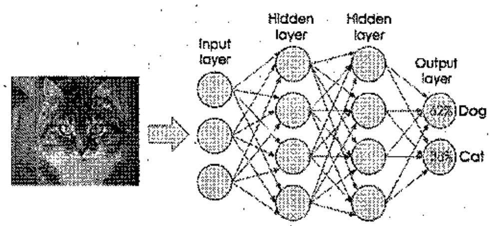

Figure 1 - A visual description of a neural network[1]

Now that we can see the steps through which a neural network can identify a photo we must understand the second stage of a neural network, the learning stage. For each input picture that we give to the network we also give an answer. In the example shown in figure 1, we would have given our network the answer that this was a picture of a cat. Unfortunately our neural network incorrectly gave a higher probability that this picture was of a dog. Because we have labeled the picture, our network can recognize that this was incorrect and go back and correct itself. For each hidden layer in the middle, the calculations of smaller probabilities, the network can go back and make small adjustments to those probabilities. These adjustments change the weights given to each node in each layer, making the output layer's probability change as well. This process of going back to each layer and adjusting the probabilities is called backpropagation. After back propagation we can feed our network another image and hopefully we will get closer to the

[1] https://medium.com/@RosieCampbell/demystifying-deep-neural-nets-efb726eae941

correct answer. We keep doing this until our network can successfully recognize that the input picture given was of a cat. This process is referred to as training and is necessary for any neural network to operate. Once the network is trained we would then give it a random set of new images and measure how accurate the predictions are. This is called testing and will give us an idea of how well we have constructed our network.

Now that we have an understanding of a neural network we can understand how a GAN works. A GAN is unique because it pits two neural networks against each other as adversaries in order to improve each other. The first neural network is the generator as seen in figure 2. This generator will take in noise, or a random input, and will output a sample image. In the beginning this image will probably look nothing like a face, but will still be used as one of two samples given to a discriminator. The discriminator is the second neural network. Its job is to take a real world image of a face, the second input, and the sample generated from the generator network and decide which of the two images is a real picture of a face and which is a computer generated picture of a face. Using similar methods of back propagation and training, as previously described, these two neural networks will continually battle against each other, improving each other's performance. This means that when the discriminator identifies the generator's image as fake, this will feed a negative response to the generator and cause it to readjust its weights, producing a more realistic image. At the same time this will make the discriminator's job more difficult, given the increasing performance of the generator. Eventually, the generator will become so good at creating computer generated faces that the discriminator will not be able to differentiate between the real world sample images and the generator's images. Once the generator is able to fool the discriminator a majority of the time the process is finished and both neural networks have been trained and tested.

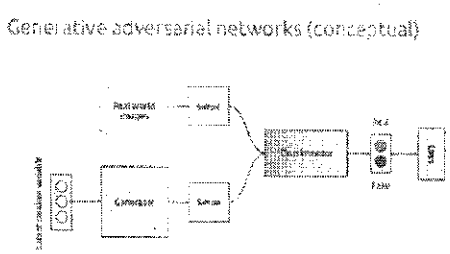

Figure 2 - A conceptual visualization of a generative adversarial neural network[2]

For the purposes of my research I am only interested in the output of the generator part of the GAN. This will be the neural network that I will use to generate facial images. There has been research already conducted on the efficacy and techniques used when creating GAN for facial generation, however, my research differs in a couple of key aspects. Firstly, I am going to be measuring my GAN with an extra feature. Typically we would primarily be concerned with whether or not the network created a

[2] https://medium.com/coinmonks/celebrity-face-generation-using-gans-tensorflow-implementation-eaa2001eef86

realistic face, however, I am concerned with the type of face produced. This means that the dataset that I am going to be using will need to not only label whether or not the image is realistic, but also what type of emotion the face is portraying and other distinguishing features of the face. Theoretically when my GAN is complete I will be able to give it an input of several defining emotions and features of the face and my generator network will be able to produce an image accordingly.

During the course of my research I will be taking bits and pieces from several created datasets, but the majority of my data will come from images of faces that are not yet labeled. Unfortunately there are not large enough datasets of labeled faces, so I will have to create my own. The ideal size of my dataset would be around 10,000 images. I plan to repeat this experiment many times until I reach a satisfactory level of loss or error. A unique part of research in neural networks and machine learning in general is the human variability. There are no set ways to create a GAN or any network, so much of the structure of the networks will be reliant on trial and error as well as personal preference. However, there are standard, empirical practices and quantifiable methods to determine the correct construction of networks that can adapt to my specific problem. I do not foresee any costs for the university in conducting my research as I already have all of the equipment and resources necessary to complete my analysis. I will possibly be using the University of Notre Dame's Center for Research Computing for my higher level computations. This will essentially allow me to use greater computing power, expediting my results. This would be at no cost considering I have an agreement with the university through my employment, and would only entail a reference to the center in my final paper.

III. Explain in what sense your project is original, unique, or beyond normal senior expectations. How does it relate to current knowledge in the discipline?

The problem is unique due to the end goal of producing computer generated faces from a description of text. This task has been done, but not with the end result being an image of a face. My approach to this problem is original because of its unique combination of convolutional neural networks, StackedGANs, and natural language processing. While these ideas have been combined in pairs, I have yet to see all three used, especially in the application of facial generation, in the field of machine learning and artificial intelligence. Also, when working with GANs there is a certain amount of artistic liberty that the programmer may take with the creation of the network. While there is research done on which GANs may perform best, the performance depends on the type of problem and also on what sequence certain calculations are done. Ultimately, the performance of a network may come down to personal preference, therefore leaving room for addition of my personal experimentation and preference.

Machine Learning, and more specifically realistic facial generation, has been a hot topic within the computer science field for the past few years. In its resurgence there has been a steady increase of funding and interest in the topic within both government and private corporations. Having the experience of working with this rising field would allow me to gain valuable leverage and knowledge that would help secure interest from employers and greater personal edification.

IV. Include a substantive annotated bibliography of similar or related work.

S. Liu, J. Yang, C. Huang and M. Yang, Multi-objective convolutional learning for face labeling, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

This paper aimed to use convolutional neural networks to label different facial features in a dataset of faces. The most accurate results were found on their HELEN dataset and ultimately could identify eyes, brows, noses, mouths, lips, and color of skin with an overall accuracy of 0.854, improving upon a previous papers' results of 0.804. These results were fulfilled through the method of network regularization in a convolutional neural network. They also converted all images to an RGB image to serve as the input to their network. These methods allowed them to regularize the network and maintain competitive results between both large and small models. This paper sheds light on the importance of regularization and possibly a different approach of convolutional neural nets to the discriminator function that I will be constructing.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

The concept of Deep Convolutional GANs (DCGANS) allows for a new approach to GANs used in computer vision research. Through their research they noticed that there is model instability when the models are trained longer due to a collapse of filters. A useful strategy proposed for facial construction based on description labels is to use a vector of descriptions in order to direct the model toward image creation. This would require using an X, Y, and Z vector for each space, resulting in more accurate and compact calculations. By dealing in vector space the network may have more control over scale, angle, and will reduce the amount of noisy overlap due to incorrect alignment of images in pixel space.

Jon Gauthier. Conditional generative adversarial nets for convolutional face generation. Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, 2014.

In this paper the author discovers that using a conditional set of data combined with a GAN will produce a more controlled output. By having a conditional dataset you may be able to deterministically control the output of the generator and therefore have closer results to what you may initially describe. This is important to my research because it provides a small example of a potential way to incorporate descriptive labels into a GAN system. The research found that with this initial conditional information we can generate faces with specific attributes, perfect for our application of GANs.

S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396, 2016.

This research also used DCGANs in order to translate text descriptions of images into actual images. Both of the networks performed feed-forward inference conditioned on the text feature. The research used natural language processing, similar to the goal of our project in order to turn the descriptions into actual images. The research concluded that a GAN-CLS generator network combined with a temporal structure would produce accurate results of different text variations. Because our labels would be fairly consistent I think the most valuable finding of this research is the conclusion that a GAN-CLS network is most accurate at providing descriptive images of text. This model will provide us with a general basis for generating images from text.

Ledig C; Wang Z; Shi W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Computer Vision and Pattern Recognition. IEEE, 2017.

This research aims to decrease the loss incurred by single image super resolution. They accomplish this decrease using a combination of convolutional neural nets and GANs, further confirming the popular use of CNNs with GANs to produce images. Using their SRGAN they could upscale images up to 4x the original resolution with very little loss. These deep residual networks would be applicable to my research due to the potential high definition and accuracy of our final images. The techniques used, specifically their choice of both type of layer and layer sequence, will guide the creation of my final network.

Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Huang, X.; Wang, X.; and Metaxas, D. N. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. arXiv preprint arXiv:1612.03242. 2016.

This paper provides the most intriguing and closest method to my research. The researchers decided to utilize a StackGAN, in other words a series of GANs feeding into one another to break up the process of image generation into sections to create a more accurate final image. The initial GAN, or Stage-I GAN breaks down the text description into a general sketch of color and shape in a lower resolution image. The second GAN will improve upon the the Stage-I results. This stacked GAN approach is an approach that I plan to use in my research considering the results of this paper, and the results of other methods mentioned before. By using a double GAN approach we can focus each stage on a specialty calculation, allowing for greater control over the conditional text descriptions and final output.

J. Choe, S. Park, K. Kim, J. H. Park, D. Kim, and H. Shim. Face generation for low-shot learning using generative adversarial networks. IEEE International Conference on Computer Vision Workshops, pages 1940–1948, 2017.

This paper is useful to my research given that the training data available may be limited. The low-shot learning method suggested by this paper essentially takes attributes that we are looking to recreate and creates other images, using a convolutional neural net, that possess those same attributes to continue challenging our network during training. By using a GAN to also create some training data, we are essentially giving our network a stacked approach, similar to the work described before. Each of the metrics used to measure the success of this method, proved to show that the addition of low-shot, augmented data allowed for higher accuracy and coverage of the network. With this stacked approach we can bolster our training set and allow us to further train our network.

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. Stackgan++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. arXiv: 1710.10916, 2017.

Similar to the other paper on StackGANs this paper did a further study into stacking GANs. These researchers also found that by having multiple layers of GANs they were able to break the problem of image generation down into more easily controllable sub problems. The results for both conditional and unconditional image generation were quite impressive. This method of GANs beat out CLSGANs and most other popular variations of GANs in qualitative and quantitative comparisons. The Stage-II GAN that they used essentially generated 4x higher resolution images than the Stage-I GAN create. Ultimately I hope to use this method of StackGANs in my research as it will improve my unconditional image generation task.

Diederik P. Kingma and Jimmy Lei Ba. Adam : A method for stochastic optimization. arXiv:1412.6980v9. 2014

This algorithm is going to be the basis for most of our optimization when using GANs. The stochastic, or random, nature of the algorithm allows for a constant change in the information presented to the network. An Adam optimization is best used when the dataset is relatively large, especially when compared to stochastic gradient descent or just regular gradient descent. In this paper it is demonstrated through experimental results that Adam, either with or without dropout, provides the least amount of loss when used with smaller batches and convolutional neural networks. In my research I may apply an AdaMax extension to the Adam optimizer if my initial training dataset is relatively small. Another advantage of this optimization function is that it requires little memory and accelerates the rate of convergence for image recognition problems. I would consider using this optimizer on at least the first level of my StackGAN.

V. Provide a statement of progress to date and list the research methods coursework completed.

I have taken the Machine Learning and Data Science classes offered by Andrews University, in which I have studied neural networks and many forms of data manipulation and management. In conjunction with these courses I have also been working at Notre Dame in their Center for Research Computing, assisting on research projects that use GANs and other forms of neural networks and machine learning. I am familiar with both the literature and current popular methods through my exposure to current research and Notre Dame and my coursework in the aforementioned courses.

Department Chair Approval

- **This student's performance in his/her major field is acceptable.**
- **He/she has completed the requisite research methods coursework for the research to be pursued.**
- **I understand that he/she plans to graduate with Honors.**

Department Chair (signature)

Research Advisor Approval

I have read and support this proposal:

Primary Advisor (signature)

I have read and support this proposal:

Secondary Advisor (signature)

If human subjects or if live vertebrate animals are involved, evidence of approval from the Institutional Review Board or an Animal Use Committee is needed through the campus scholarly research offices (Ext. 6361).