Determining Human Emotions from Facial Images.

Kevin L. Markley, Kennesaw State University, Marietta, GA 30060 USA

e-mail: kmarkley@students.kennesaw.edu

***Abstract*—This paper is meant to explore facial expression recognition and determine the efficiency of several different previously devised models on the same hardware infrastructure and compare the results of each model's speed and precision. Simultaneously, providing an in depth look at the techniques and methods used be each of the various models and allowing multiple types of people, including students and instructors to gain a large amount of knowledge of each model through the use of an interactive web based framework. This is an important topic as research has steadily progressed regarding similar machine vision areas such as object and face detection, and recognizing human faces, but we have not had as much universally acclaimed research into the next step of information extraction: gaining more than a surface level recognition or comparison of a person.**

Ideally, this will provide some context and possible answers to the question of can a computer understand human nonverbal communication? This question is ever evolving and will take place in many different contexts ranging from unsportsmanlike conduct to individual therapy sessions. All of which can fall into one of two categories, human to robotic interactions and human to software interactions. Throughout the paper, these topics will be discussed with unique interactions kept in mind. In future works, this question will transform towards how well and quickly can a computer understand human facial expressions.

***Index Terms*—Deep Learning, Facial Expression Recognition, Machine Learning, Machine Vision.**

I. Introduction

A. Introduction to Facial Expression Recognition

Facial expression recognition is the task of determining the emotional response that is present in human nonverbal communication, specifically through the muscle movements of the face. Research into this area has been limited but has started to grow with the advent of fast and accurate object detection using deep machine learning algorithms inspired by the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [1]. As can be seen by the results of the latest competition, held in 2017, most of the competing teams have found a way to consistently exceed a 95% accuracy rate [2]. Due to the success rates of the teams, the challenge that was posed by the competition has been diminished, and the competition itself is being restructured to find a more engaging challenge [3, 4]. This has allowed researchers the opportunity to focus on creating new algorithms to perform other computer vision tasks while being able to rely on previous efforts into the tasks of localization, detection, and video detection for extremely high accuracy and performance [2]. Outside of the competition, computer vision research has evolved over the years until computers could more accurately detect many objects, including human faces, with higher precision than humans could [5]. Thus, we can now conclude that the field of computer vision has had extraordinary success in the measure of detecting objects across multiple situations and conditions.

The next step for research in this topic was to go above simply detecting an object in the image but now detecting several unique types of any object. This would allow images to be classified on the number of different models of cars or separate instances of a car that were in the image rather than just classification based on the image containing at least one car. This leads to recognizing the objects and classifying them as at a board object category and again in subcategories such as unique people, which is known as instance segmentation [6, 7, 8]. Knowing now that there is not just a car in the image but a car with a certain make, model and even year and how many cars of that category there are can be extremely useful information. That classification system allows computers to distinguish any one car from any other cars in the image, depending on the amount and type of categories and subcategories that are used. Taking that information, humans can attribute other information such as engine type, miles per gallon, maintenance requirements and possibly even the owner for any car in a given image. Altogether the information gleaned from instance segmentation provides a simple tool that humans can use with powerful leverage across a multitude of professions and tasks, as can be seen through the creation of autonomous vehicles [9, 10]. The tools and methods for extracting that information have gotten quicker and and easier over time as well [10].

Thus the next unique and yet similar area of research is to extract even more information from an two dimensional image. For objects that don't visibly change with much frequency, such as cars, there is limited additional information that can be found after recognizing and classifying it. However, humans have nonverbal yet visible communication

which can lead to a lot of nuanced additional information about the image that necessitates a new unique method of classifying and understanding. For instance, we can detect and recognize the person in any image, and yet we still wouldn't know how they felt or what they might have been thinking without additional processing of the image. This is where Facial Expression Recognition technology has a massive role to play and not just in assisting human tasks, but also in terms of computer to human interaction, yet this field is still in its infancy [18].

B. Introduction to Nonverbal Communication

There has been extensive research into nonverbal human communication since before computers were even suggested and from such renowned researchers as Charles Darwin [12]. Darwin was one of the first to associate emotional expression through bodily movement to a learned habit which was proven to be understandable across many humans and animals. Some of the common emotions that were displayed included fear, disgust, anger, sadness, shock, happiness, and more complex emotions like shyness. Understanding that nonverbal communication can take place between all people and can include many emotional states leads into the belief that it is essential to communication and this is shown through a multitude of studies [14, 15, 16, 17]. A conservative estimate from many researchers suggest that approximately two thirds of all communication is done nonverbally [15, 16]. Using the previous research and new testing methodologies, Ekman et al. developed a codified system that can be used to describe facial expressions [13]. This system describes the unique movements of each muscle in the face to create changes in its appearance, thus the system is called the Facial Action Coding System (FACS). There are 52 Action Units (AU), eight of which describe the position of the head. Thus there is a system in place that can be used to attribute emotional meaning to each of the different facial movements, which should aid computers and humans alike, in understanding each other.

C. Research Problem

Starting with Charles Darwin's "The Expressions of the Emotions in Men and Animals" the study of nonverbal communication begun [12]. Since his initial curiosity lead him to study the topic for both humans and other animals, major strides have been taken to better understand how this communication occurs. One major point to note is that it has been discovered that nonverbal communication is roughly two thirds of all communications [14, 15, 16]. This means that in order to truly grasp two thirds of human communication, a machine would need to learn how to decipher these signals, otherwise computers will never be able to communicate effectively and appropriately. Thus the research into this field will greatly increase the believability, relatability, and understandability of human computer interactions. However, there are some prerequisites to this research such as face detection and recognition, since it is required to understand what the general shape of a human face is so that computers do not attempt to find communication where there is none and to allow them to determine the idiosyncrasies of each human's nonverbal communication [6, 7, 18]. Because of the needed preliminary learning steps, this topic has only recently become more effective in performance times and precision [2, 3, 5]. However, as those problems have received more and more attention, there has been a growth in research and the beginning of the creation of models and techniques that outperform humans in object and face detection for accuracy and speed [5]. This has made the field of facial expression recognition more accessible, but since it is still a budding area, it requires more explanation of current implementations of solving this problem as well as new ways of solving it. Any novel methods will be described in the future works section.

This research requires the use of a lot of data in order to train the machine learning algorithms. Thus the data that will be used is the open source Facial Expression Recognition (FER) dataset that was created by Pierre-Luc Carrier and Aaron Courville, working with Microsoft [30]. Results will be based on the accuracy and speed of the models that was created after training on the FER dataset.

D. Purpose of the Study

The purpose of this study is evaluate different implementations of human facial expressions and to determine the significance of each in a few important contexts. Simultaneously, that research will not only produce results regarding runtime performance and viability, but also a set of actual implementations of each of the different techniques. As an insightful courtesy, those implementations may be shifted to an online and freely available set of open source code that should increase general learning and understanding of the processes, requirements and general methodology of performing facial expression recognition. It is expected that the tangential output of this paper will be more greatly utilized and referenced than the paper itself, since there appear to be a lack of full implementation details in this research topic as well as it being mostly confined to the arena of research papers.

E. Audience

The intended audience of this paper is anyone who desires to enter the field of computer based facial expression recognition. This can be anyone from computer science students and teachers to developers to people who simply have a passing interest and no prior computer science background. This is due to the computations and determinations of facial expressions will happen in real time and be hosted on the internet it provides a simple yet powerful interface for the learning and determination of the appropriate use of implementations in this field. However, the traditional research paper will still be produced so that it may not only be referenced but allow learners and experienced researchers the opportunity to utilize the information contained herein with their keen eye.

F. Contribution

The contribution of this paper is to directly overview several models of determining human facial expression

recognition which should allow practitioners new to the field to understand the intricacies of each and when to use them. This inherently requires the implementations to be recreated and tested on the same hardware specifications and with the same data sets. Since that requirement exists, an additional step with limited overhead would be to translate the code into a form that would be easily accessible, such as a web based situation. That will allow readers to view the underlying structure of the code without much requiring any additional and overly descriptive sections regarding the layout and operation of the code. Therefore the paper will mostly contribute graphs and statistical evaluations of the results of each implementation while the implementations themselves will be directly accessible and manipulatable.

G. Motivation

The driving force behind this research is the desire to allow computers more complete understanding of human interactions and for them to to engage with us in a way that is more meaningful than simply performing calculations. It is the belief that this research will lead to amazing and simple innovations that will revolutionize human and computer interactions and should be beneficial to both not only in the short term but for the rest of time.

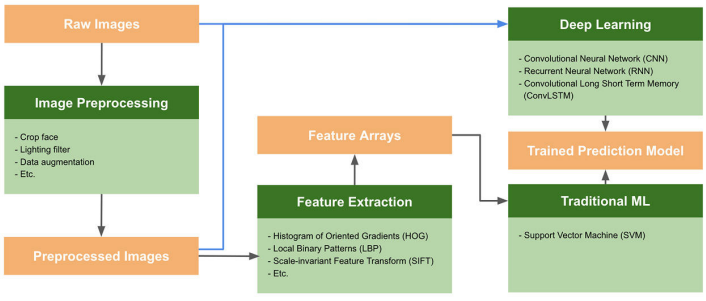
H. Paper Goals and Organization

The project has a few objectives. Namely, creating a robust graphical interface for facial expression recognition and doing so in a real time web based setting. On top of that, it needs to provide a comparison of more than two methods and techniques. The reason that the research needs to create a robust graphical interface is because the interface will need to work for all possible implementations and simultaneously provide an easy to understand explanation of a complicated and heavily mathematically based operation. This will be the one of the more difficult parts of the project and will require much fine tuning and effort to implement on top of the other objectives. As of right now, it will be accomplished by providing a graphic of a matrix that will represent the image and each step will be highlighted and show how the matrix is manipulated as it is used in each step until it produces a final product which labels the image as a face with one of the possible expressions/classifiers given. Below the graphic, will be an explanation in words of the step showing the underlying code and describing the reasoning behind it. Another objective is to actually implement facial expression recognition, which is important to this research since that is its foundation and it is a valuable field to research as it will allow machines to better automatically interact with humans. This should be accomplished through the use of a programming language and following as many steps as are described in several seminal research papers about the subject. The third objective which is built directly on top of that facial expression recognition is that of the web based accessibility of the code's output. The reasoning behind adding that is to ensure that the code can be run anywhere at any time and by anyone while also allowing for a more detailed viewing of its process than would be allowed simply by a research paper detailing and showing some output. This will be accomplished as of right now through the use of either GitHub pages, which will provide a simple domain name for the project while allowing complete customization and control on what is visible and it will also allow for the execution of the programming language source code that is used to model facial expression recognition. The last objective of making it real time is simply to allow for a responsive viewing of the process and does not enforce hardware requirements on the user or set up time before hand. They may simply come to the site and then view the project without any wait time or preexisting knowledge about the subject.

This paper has the basic layout six numbered sections with a reference section at the end. The six sections move in order of introductory to techniques used in the study, proposed method, experiments and results, conclusions and future works and acknowledgements. The introduction section, which is almost finished at this point, introduces the concepts that will be used for the rest of the paper and describes related works and their implementations and findings. It also covers the motivations of the research and why the research is critical in the motivation and research problem subsections.

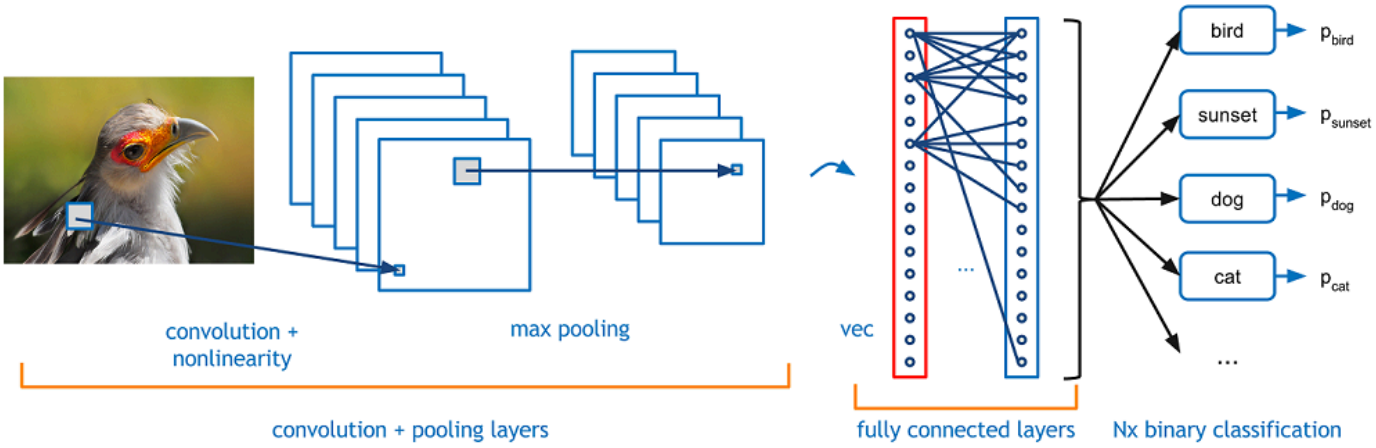
The techniques used in this study section will primarily deal with the different variations between each implementation that was recreated as part of this research problem. It will detail these differences to a rather high level degree at the moment, as provided by the referenced papers. However, this limitation will be overcome with the lower level details that will be generated through the rebuilding of their implementations on our own systems, which may be more closely viewed when the code becomes accessible through the web.

The following section will cover the proposed method of creation and deployment of the code base in order to test it and compare the results of each implementation. Experiments and results will be described and reviewed in the section foregoing that. Finally, conclusions will be drawn and debated and acknowledgements will be given.



I. Related Works

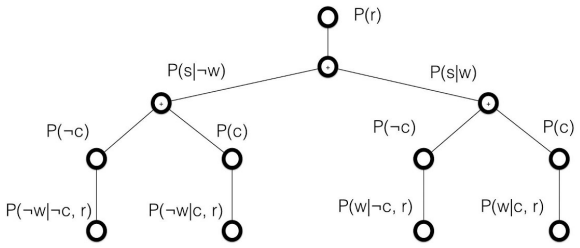
Computer vision is a deep and complex field requiring practitioners to have knowledge of subjects ranging from algebra to performance optimization of any chosen programming language [1, 11, 18]. Thus for simplicity's sake the research will only explain the programming and computational portions required to perform the determination of human nonverbal facial communication. This will primarily be the ideas used in convolutional neural networks and bayesian networks since those are the two distinct ways that



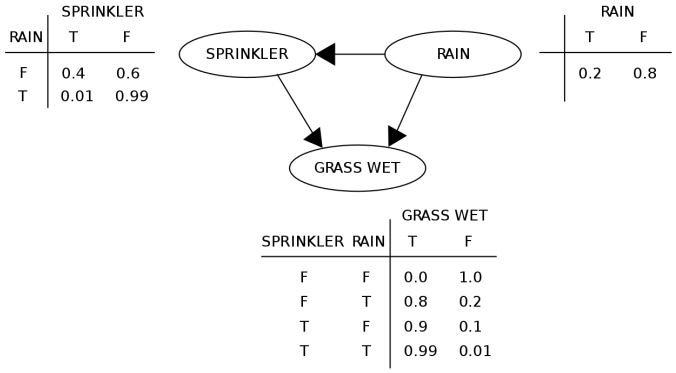
facial expression recognition is done [11, 18, 24, 25, 26].

Convolutional Neural Networks (CNNs) are best described in the words of [26]. The latest training architectures being explored are all deep neural networks which perform better under spontaneous uncontrolled circumstances. Convolutional Neural Networks are currently considered the first choice of neural networks for image classification, because they pick up on patterns in small parts of an image, such as the curve of an eyebrow [7]. CNNs apply kernels, which are matrices smaller than the image, to chunks of the input image. By applying kernels to inputs, new activation matrices, sometimes referred to as feature maps, are generated and passed as inputs to the next layer of the network. In this way, CNNs process more granular elements within an image, making them better at distinguishing between two similar emotion classifications [10, 21].

$$P(r|s) \propto P(r) \sum_{w} P(s|w) \sum_{c} P(c)P(w|c,r)$$



Bayesian networks are instead meant to represent logical relationships, which is done through the use of inference [24, 27]. The relationship between each node in a Bayesian graph is shown in figure 1. This allows any node to be affected by any other and thus models conditional dependence and causation. In other words, when a specific situation is occurring, this graph can show the probability of another situation occurring after or because of the first. This becomes helpful for facial expression recognition because we can relate, through training, each individual action unit, or facial movement, to each other and determine which emotion is currently playing out on the person's face [27]. Some specific actions bear a greater weight on determination than others, and this should be learned by the model through training on the images.



II. Techniques Used in the Study

In the amount of time that was available, this research opted to implement two different systems of facial expression recognition. The first is a deep learning technique known as a convolution neural network. Convolutional neural networks have been described before in the related works section, but for this implementation, some of the specifics have changed. An important distinction is that CNN will be using raw images and labels only. The figure below shows how several different neural network types can be used to fulfill similar purposes, but is in this case describing our implementation of just the CNN. The other method that is being used is a Bayesian network. Bayesian networks work on an internal directed graph where each node represents an action and that action leads to several different states through the use of directed edges [24]. This leads to certain learning algorithms that can use inference to predict any action from any state. Below is an example graph of a Bayesian network. Thus we can use either network type to extract features and understand correlations between them.

Several issues need to be understood and adapted for before continuing into the actual implementation. One of the largest issues to consider is how many images are in the dataset that is

being used [1, 18, 20, 21, 22, 24]. The larger the dataset the more likely the computer is to learn certain facial expressions and what areas of the face contribute to making those expressions and thus how important any one feature is in determining a certain facial expression [21, 22]. However, since people and cultures each have their own way of making and using an expression, this can cause the computer to overfit if many of the images are too closely related in any area or taken from a sample population that is too similar. Some example situations of overfitting include but are not limited to, orientation of the camera to the face, different locations and sizes of facial features for each person, skin color and lighting, and the way each person contracts their facial muscles to perform the expressions and the locations of those muscles [21]. There are ways of sidestepping these problems, namely transfer learning and data augmentation. Transfer learning is the process of using a pre-trained neural network and training it again with images from a separate dataset [28, 29]. Data augmentation is the use of a separate neural network to slightly modify each image in the existing dataset so that we artificially increase the amount of images and images per class (i.e. anger, fear, or sadness) we have [21, 22].

Another problem is the processing speed, even with high accuracy if the program takes too long to execute it could cause issues. For example, in the case of a police officer looking for a certain type of car that a criminal was last seen driving it would not be effective if the confirmation came after the suspected car has already driven away.

The final gap and most import for this research is that this is a field of research that has been on the cutting edge but has not seen many teaching directives yet [24]. Thus the main goal of this research is to implement and provide an easy to view and operate way of comparing current implementations of techniques and models used in this field.

The way that this will be accomplished will be done by going step by step through most of the algorithms and techniques described in several key research papers. Each method will have its own labeled implementation which should be accessible through not just the paper, but a website. Each implementation will have show each step of the process, but not every action taken in that step for performance reasons, and will show a labeled description of the step below a graphical output. This should then easily provide a simple method of learning the algorithms while also establishing each one's strengths and weaknesses. The hope of this research is that it will inspire more students, teachers, developers to learn and teach how to perform facial expression recognition or adopt its usage and that it will provide a useful learning tool without any additional cost to them, which may even inspire a general interest in not just computer vision but also computer science.

III. Proposed Method

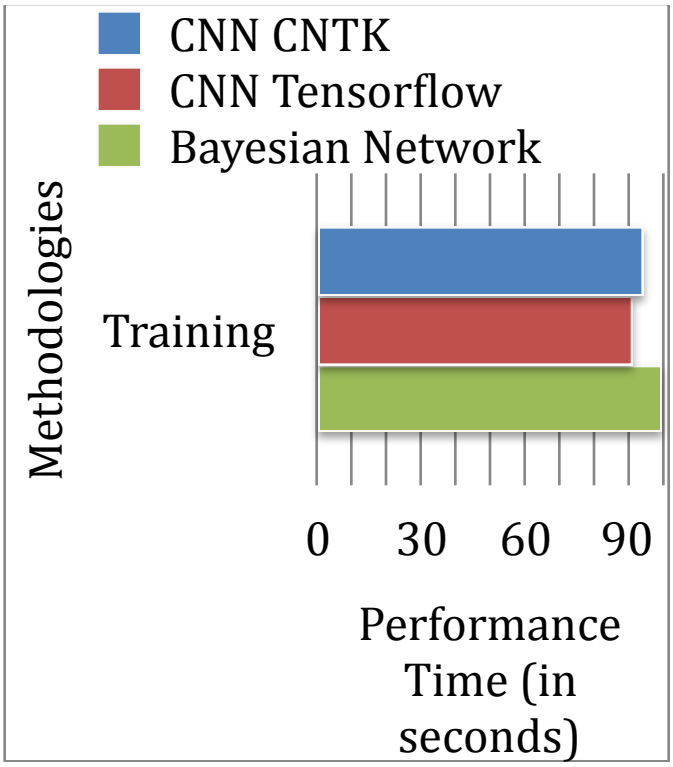
Both the Convolution Neural Network (CNN) and Bayesian network implementations were trained on the raw images in Microsoft's open source Emotion Facial Expression Recognition+ dataset, which carries slightly over 28,700 labeled images. The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered and adjusted so that the face is more or less centered and occupies about the same amount of space in each image. Each image is human annotated so that the expression shown is placed into one of the seven categories: 0 is Angry, 1 is Disgust, 2 is Fear, 3 is Happy, 4 is Sad, 5 is Surprise, 6 is Neutral. The training set is represented with a comma separated file that contains two columns, "emotion" and "pixels". The emotion column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The pixels column contains a string surrounded in quotes for each image. The contents of this string are the space-separated pixel values for each image in row major order. Thus each row will have in the first column a singular number inclusively between zero (0) and six (6), while the second column will contain forty eight by forty eight, or 2,304, pixel values separated by spaces which represent the image. The testing set has a similar file, but it only describes the pixels column, and the testing set used consists of a separate 3,589 images which are not included in the training set. The goal of our machine learning algorithms is to categorize each face based on the emotion shown in the facial expression in to one of seven categories: angry, disgust, fear, happy, sad, surprise, and neutral, and to have the predicted category correctly match the real emotion column's values. That dataset was augmented through the use of transfer learning and data augmentation [21, 22, 24]. This allowed the use of an artificially increased dataset and the creation and use of several similar datasets that would produce better results for the classification of the expressions into angry, disgust, fear, happy, sad, surprise, and neutral. In addition to the first set of testing images, in a novel approach, there will be an additional testing set that is entirely composed of computer generated images representing human faces. These images will be as closely normalized and centered as the images in FER+ dataset but will come from animated movies such as Pixar's "Coco". Since the animated images tend to exaggerate expressions, the results expected should be similar when the conditions of the image match closely to the original dataset that the algorithms were trained on.

There was a preprocessing requirement for both implementations as well. Any preprocessing that was shared between the two was done before initialization and execution of either program. Thus, performance times were not measured using the data augmentation portion of the code. For any specific preprocessing requirements for either implementation that was necessarily included as part of the initialization portion of the code for the required implementation. However, since the FER+ dataset already came as a grayscale image and normalized to center faces with the same height and width, we could avoid much of the performance increase of preprocessing directives. For other datasets and for use in real world and real time applications, these preprocessing initiatives would significantly increase the run times, if training was needed. It has been shown that testing speeds of each deep learning algorithm is much faster than the training period, so it would go to reason that if such an application was needed in a real time context, that the machine would only be testing the classifiers created by the

deep learning algorithms [1, 30].

In order to decrease the required memory space for both implementations, the preprocessed dataset was available as a read-only, global variable for them. This worked as preprocessing that was needed for either implementation was again done as part of its own training run.

Finally, both the CNN and Bayesian network were run on the same machine independently of each other and concurrently. This allowed them to be tested using the same underlying hardware and so both would have the same amount of memory and processing speed to work with. The reason they were also run concurrently as separate threads, was to see if they could run with the same speed and memory requirements when being run simultaneously. They were both created in a serial fashion with no concurrency built into their natural function.



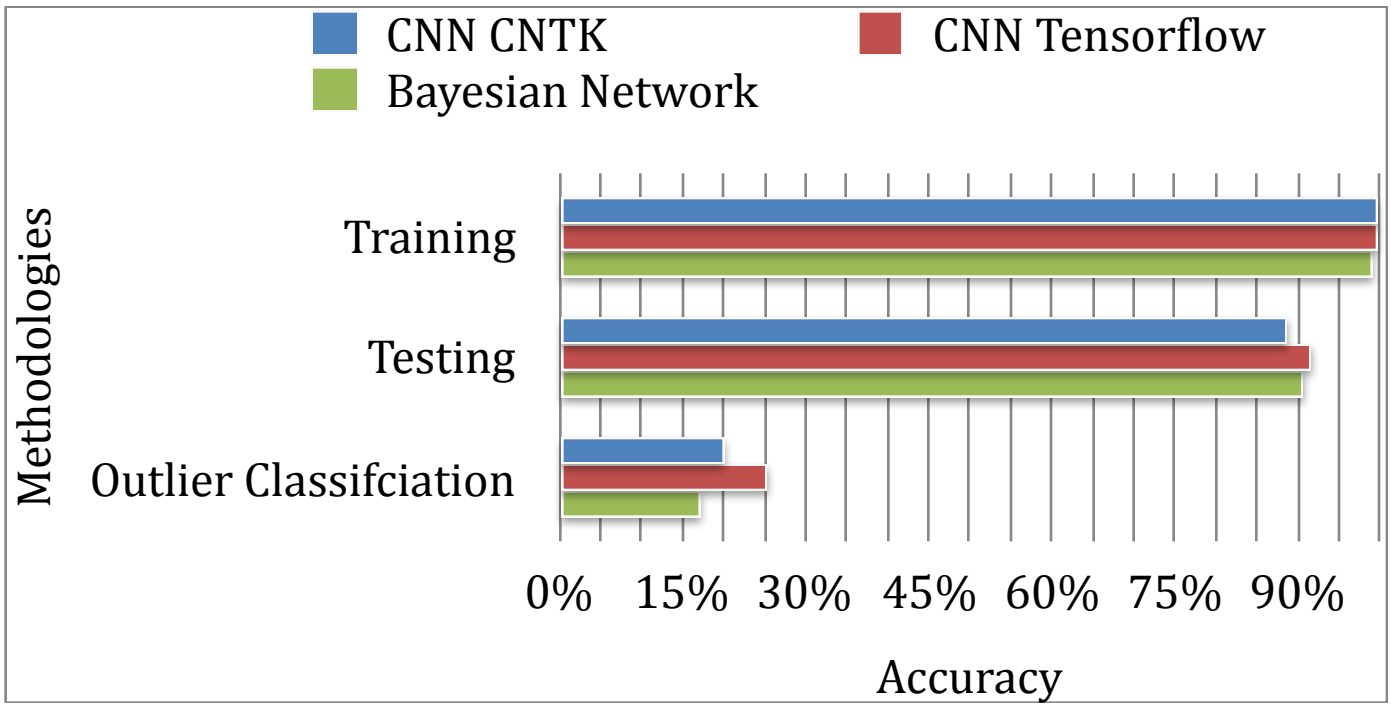
IV. Experiments and Results

The experiments produced results that were similar in a few ways. The first is that both achieved a relatively high accuracy and were competitively close to each other's accuracy. This is not entirely unexpected since they were trained on the same or a very similar dataset, since the data augmentation was run multiple times on the FER+ dataset to generate additional images but could have generated some that differed between runs, and the output was bounded between only 7 possible values for emotions. When compared to the number of possible classifications in other computer vision tasks, this produces a much smaller possible maximal difference. Roughly allowing for 28,000 ^ 7 choices, while object detection with 1,000 different classes allows on the same dataset 28,000 ^ 100 choices. In terms of training, the Bayesian network was a little quicker to train since its connections between features on the human face that it could use was limited to the number of features on the face, which were bound to the eye pupil dilation, corners of the mouth and eyes, nose position, and eyebrows position. Even though they all interacted with each other, there was only seven connections between each node or feature. That still produced a possible seven by seven traversal of the features to determine what any one expression was, but this was limited again by the usage of a naive search. Whereas, the convolutional neural network suffered in performance because of large forty eight by forty eight set of neurons being connected and sending input and output to each other repeatedly in order to learn features. In terms of the testing time, both ran at roughly the same speed with only microseconds of difference between them in classifying the much smaller testing set.

There were also tested a second time against images not in the dataset and were created through human usage of three dimensional computer modeling. These images were taken from computer animated films by Pixar, namely "Coco" and "Incredibles 2". Both machine learning techniques were able to achieve a similar accuracy on this additional testing set when the images were similarly centered and focused on the faces as the images in the FER+ dataset. It was noticeable though, that the CNN was able to achieve a higher accuracy when compared to Bayesian network when the faces were not as similarly centered and focused. Therefore both were limited in the ability to classify images across a wide spectrum of diverse conditions, but could more easily label images that were similar to well defined and clear images used in the data set. This is the expected outcome from the that test since it stands to reason that if the machine was not taught to understand the emotions of the characters in diverse conditions then it should not. However, it was expected and interesting that they were able to maintain a similar accuracy for images where the model used was not truly a person but a completely computer generated image.

V. Conclusions and Future Works

While both techniques have different runtimes in terms of training, this is not a terrible issue as computers still have some room to increase performance and training is expected to be the most time consuming task, The more important performance conclusion is that, when considering run times for testing, that both are able to achieve a speed that is similar over a small sized dataset, assuming that nearly 4,000 images is a small dataset. That means that if they are used in real time applications to determine only a single person's expression, that they will be able to do so without any large performance issues and may be able to contribute to those real time applications. However, the CNN is able to handle a much wider band of circumstances and still be able to relate those different images back to the original dataset it was trained on. While, the Bayesian network fails to make comparable judgements in those circumstances. Thus, Bayesian networks may be more adapted to situations that require a fast



implementation for specific circumstances, and a CNN is better suited to handle any situations that may require a range of input that is outside of its training knowledge, yet not require that it be implemented quickly for production use.

Future works are expected to come from this in that we would like to assemble a dataset of images that come from Computer Generated Images (CGI), as well as testing on those images and then determining if these FER algorithms will be able to apply their knowledge from them to real people and vice versa. Our preliminary results are something that shows a high level of confidence for continuing research into that direction without having to reinvent the wheel, so to speak.

VI. Acknowledgement

The author would like to thank ThoughtWorks and Angelica Perez for their contribution to the open source community through the creation of EmoPy for the task of facial expression recognition, as well as Pierre-Luc Carrier and Aaron Courville, for preparing the FER+ dataset as part of an ongoing research project under Microsoft.

References

1. Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.
2. Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution, Jul 17, 2017) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015. http://image-net.org/challenges/LSVRC/2017/results. image-net.
3. "New computer vision challenge wants to teach robots to see in 3D". *New Scientist*. 7 April 2017.
4. Y. Xiang, W. Kim, W. Chen, J. Ji, C. Choy, H. Su, R. Mottaghi, L. Guibas, S. Savarese, "ObjectNet3D: A large scale database for 3D object recognition", *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 160-176, 2016.
5. Chaochao Lu and Xiaoou Tang. Surpassing Human-Level Face Verification Performance on LFW with GaussianFace. Appearing in Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15), 2014. arXiv 1404.3840.
6. Hariharan, B., Arbelaez, P., Girshick, R., Malik, J.: Simultaneous detection and segmenta- ´ tion. In: European Conference on Computer Vision (ECCV). Springer (2014) 297–312.
7. Chen, Y.T., Liu, X., Yang, M.H.: Multi-instance object segmentation with occlusion handling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 3470–3478
8. Romera-Paredes B., Torr P.H.S. (2016) Recurrent Instance Segmentation. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9910. Springer,
9. Cham. Li Z., Gavves E., Mensink T., Snoek C.G.M. (2014) Attributes Make Sense on Segmented Objects. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham.
10. Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. Mask R-CNN The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969.
11. I.Michael Revina, W.R. Sam Emmanuel. A Survey on Human Face Expression Recognition Techniques. Journal of King Saud University - Computer and Information Sciences, 2018. ISSN 1319-1578. https://doi.org/10.1016/j.jksuci.2018.09.002.
12. Darwin, Charles (1972). The Expression of the Emotions in Man and Animals. AMS Pres.
13. Ekman, Paul & Rosenberg, Erika & Editors,. (1997). What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). 10.1093/acprof:oso/9780195179644.001.0001.
14. Hogan, K.; Stubbs, R. (2003). *Can't Get Through: 8 Barriers to Communication*. Grenta, LA: Pelican Publishing Company. ISBN 978-1589800755. Retrieved 14 May 2016.
15. Burgoon, Judee K; Guerrero, Laura k; Floyd, Kory (2016). "Introduction to Nonverbal Communication". *Nonverbal communication*. New York: Routledge. pp. 1–26. ISBN 978-0205525003.
16. Calder, A., Young, A. Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci* **6,** 641–651 (2005) doi:10.1038/nrn1724M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
17. Antonio Damasio. The Feeling of what happens. Harcourt, Inc - ISBN 978-0-15-601075-7, 2000.

18. Revina, I.Michael & Emmanuel, W.R. Sam. (2018). A Survey on Human Face Expression Recognition Techniques. Journal of King Saud University - Computer and Information Sciences. 10.1016/j.jksuci.2018.09.002.
19. Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, Thomas S. Huang, Facial expression recognition from video sequences: temporal and static modeling, Computer Vision and Image Understanding, Volume 91, Issues 1–2, 2003, Pages 160-187, ISSN 1077-3142, https://doi.org/10.1016/S1077-3142(03)00081-X.
20. Identifying and detecting facial expressions of emotion in peripheral vision Smith FW, Rossit S (2018) Identifying and detecting facial expressions of emotion in peripheral vision. PLOS ONE 13(5): e0197160.https://doi.org/10.1371/journal.pone.0197160.
21. Recognizing human facial expressions with machine learning. Angelica Perez.
22. Classifying Facial Emotions via Machine Learning. Sumit Kumar Singh.
23. Unsupervised Pre-Training of Image Features on Non-Curated Data. Mathilde Caron, **Piotr Bojanowski**, Julien Mairal, **Armand Joulin**.
24. I. Cohen, N. Sebe, F. G. Gozman, M. C. Cirelo, and T. S. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages I–595 – I–601 vol.1, 2003.
25. Ben Gal I (2007). "Bayesian Networks" (PDF). In Ruggeri F, Kennett RS, Faltin FW (eds.). *Encyclopedia of Statistics in Quality and Reliability*. John Wiley & Sons. doi: 10.1002/9780470061572.eqr089. ISBN 978-0-470-01861-3.
26. Caifeng Shan, Shaogang Gong, Peter W. McOwan, Facial expression recognition based on Local Binary Patterns: A comprehensive study, Image and Vision Computing, Volume 27, Issue 6, 2009, Pages 803-816, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2008.08.005.
27. Simplicio, Carlos & Prado, José & Dias, Jorge. (2010). Comparing Bayesian Networks to Classify Facial Expressions. 10.2316/P.2010.706-065.
28. George Karimpanal, Thommen, and Roland Bouffanais. "Self-Organizing Maps for Storage and Transfer of Knowledge in Reinforcement Learning." Adaptive Behavior 27.2 (2018): 111–126. Crossref. Web.
29. Niculescu-Mizil and R. Caruana. Inductive transfer for Bayesian network structure learning. In Conference on AI and Statistics, 2007.
30. Emad Barsoum and Cha Zhang and Cristian Canton Ferrer and Zhengyou Zhang. Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution. 2016. 1608.01041 arXiv.
31. By AnAj - Own work (Original text: self-made), Public Domain, https://commons.wikimedia.org/w/index.php?curid=19734596