Investigating Data Augmentation Strategies for Computer Vision Facial Expression Recognition

Jack Liu
Irvington High School
jacliuwave@gmail.com

Abstract

Autism is a neurodevelopmental disorder. A major symptom is a difficulty communicating and understanding social cues such as emotions. I aim to help people with autism better recognize emotions by developing improved artificial intelligence (AI) models to recognize facial expressions. Such models can be and have been integrated into digital therapeutics for children with autism. A crucial step to achieving performant models is to apply data augmentation to increase the dataset size and the generalization capacity. I compare and contrast data augmentation strategies on the Facial Expression Recognition (FER) 2013 dataset to determine which method leads to a maximal increase in performance. I then examine the benefit of data augmentation at various training set sizes. Among the strategies I evaluate, I find that shifting the width of the image provides the greatest increase in performance when compared to not applying data augmentation. Furthermore, I find that at several training dataset sizes ranging from 100 to 20,000 images, applying all data augmentation strategies consistently outperforms no data augmentation. These strategies can inform the development of digital therapies for autism which focus on the evocation and subsequent automatic detection of facial expressions.

Introduction

Autism spectrum disorder (ASD) is a developmental disorder caused by genetic and environmental factors that can cause difficulties in communicating, repetitive behaviors, particular interests, sensory processing difficulties, speech or cognitive issues, and other characteristic behaviors [1,2]. These effects cause autistic people to act, interact, and learn differently than others [3]. Autism affects around 1 in 44 people in the United States [4]. Signs of autism can appear when a child is around 2 or 3 years old but can be detected within 18 months of birth [5]. Autism affects everyone differently with some still being high-functioning and being able to live independently, while others may need a lot of assistance or are severely challenged [5].

Symptoms and effects of autism vary across individuals, but some people with autism may find trouble recognizing emotions effectively and communicating with others [3-4]. For example, they may be unable to maintain eye contact, unable to effectively use hand gestures, have scripted speech patterns, or struggle to make friends [3-4]. As a result of struggling to communicate, they may feel ostracized or lonely and are more likely to develop conditions such as anxiety and depression [3-4]. When people with autism are unable to identify emotional cues, it is a condition known as alexithymia [5]. Around half of all people with autism also have alexithymia [5]. Studies have shown that many autistic people with alexithymia know that they are experiencing some sort of emotion, but are unable to identify and react to it [6]. Therefore, facial emotion AI could potentially be used to help these children better understand and recognize the emotions of others, which could improve their social interactions. In addition, it can be used to

provide feedback to children with autism about their own emotional expressions. This could help them learn to better understand and regulate their own emotions as well as recognize the emotions of others.

A multitude of researchers has made innovations in the development and application of emotion recognition AI models towards providing remote, scalable, and accessible therapy to children with autism. For example, the SuperpowerGlass project is a digital device that relays real-time social cues to the wearer [7-12]. The device uses Google Glass to scan faces and sends the detected facial expressions and emotions to a smartphone app [13].

Data augmentation is a technique used to artificially increase the size of a dataset by generating new data samples from the existing ones, for example rotating, cropping, or flipping them. This can improve the performance of machine learning models. It is especially important when the given dataset is small and the model needs a large dataset to train with.

In this work, I trained a machine learning model using the FER 2013 dataset and applied several data augmentation strategies to understand which strategy would provide the highest performance gains to my AI model. The data augmentation strategies I experimented with were featurewise center, featurewise standard normalization, zero-component analysis whitening, zero-component analysis epsilon, rotation, width shift, and height shift. I then measured the accuracy, precision, and recall of this model on a held-out test dataset from FER 2013 to determine how useful or efficient the data augmentation strategies are. Among the strategies I evaluated, I found that shifting the width of the image provided the greatest increase in performance when compared to no data augmentation. Furthermore, I found that at several training dataset sizes ranging from 100 to 20,000 images, applying all data augmentation strategies consistently outperformed not applying data augmentation.

Related Work

I reviewed the most recent literature describing facial emotion recognition using computer vision techniques. I selected works that explicitly document the data augmentation strategies used. The selected works discussed employ the following data augmentation strategies: Signal-based Audio Augmentation (SA), SA with replacement (SAR), SA with replacement of the majority class only (SARM), SAR adding only Background Noise (SARB), SAR using only TS and PS (SARS), face detection crop, grayscale conversion, image normalization, image augmentation, rotation, reflection, flip, scale, shift, noise, color jitter, random erasing, mirroring, rescaling, geometric transformations, linear transformations, flipping, and mixing images together. The range of emotions tested for is anger, disgust, fear, happiness, sadness, surprise, and neutrality. Accuracy percentages ranged from 53% to 97%. Chatziagapi et al. proposed a General Adversarial Network (GAN) to deal with the issue of data imbalance in Speech Emotion Recognition (SER) on the emotions of anger, happiness, and sadness. The data augmentation strategies that were analyzed were sample copying (CP), signal-based audio augmentation (SA), SA with replacement (SAR), SA with replacement of the majority class only (SARM), SAR adding only background noise (SARB), and SAR using time stretch (TS) and pitch shift (PS) (SARS) with a dataset consisting of many spectrograms. These researchers reached an accuracy of 53.6% with data augmentation and 49.0% accuracy without data augmentation [14].

Ahmed et al. developed a facial expression classifier on 5 popular emotion recognition datasets and studied the effect of horizontal flipping, rotating, rescaling, shearing, and zooming as data augmentation strategies. The authors reached 96.24% accuracy with data augmentation and 92.95% accuracy without data augmentation [15]. Tong et al. created improved convolution neural networks by using second-order pooling. Because they had a small dataset, they applied data augmentation strategies to enhance the training process. Their group obtained an 88.625% accuracy with data augmentation [16]. Xu et al. investigated bias and fairness in facial recognition AI while applying data augmentation strategies to the RAF-DB and CalebA datasets. They achieved an accuracy of up to 74.8% with data augmentation and 62.2% without data augmentation [17]. Sajjad et al. implemented different algorithms for face and emotion detection. They also used data augmentation to improve training within their dataset. They reached an 80.5% validation accuracy without data augmentation and 94% with data augmentation [18]. Yang et al. researched and developed a model to scan and identify if someone is angry, disgusted, fearful, happy, sad, or surprised. They implemented pre-processing approaches like face detection, rotation rectification, and data augmentation during training. They achieved a 97.02% accuracy with the CK+ dataset, 92.21% with the JAFFE dataset, and 92.89% with the Oulu-CASIA dataset [19]. Khaireddin and Zhuofa used the VGGNet Architecture in this work along with rescaling, shifting, and rotating images in the dataset to experiment with optimization. The emotions analyzed were anger, disgust, fear, happiness, sadness, surprise, and neutrality, and they achieved 73.28% accuracy [20]. Zhu et al. analyzed emotion classification using GAN-based data augmentation techniques. The new images generated after data augmentation techniques were applied were a combination of high-level facial features from the original images. The emotions that were analyzed were neutral, fear, anger, disgust, sadness, happiness, and surprise. The experiments saw a 5% to 10% increase in accuracy after data augmentation was applied [21]. Tan et al. researched group facial recognition using different convolutional neural networks. They achieved an accuracy of 80.9% on the test set [22]. Psaroudakis and Kollias experimented with data augmentation strategies to propose a newer, more effective data augmentation strategy in relation to emotion recognition. The precision, recall, and F1 scores increased by 10%, 5%, and 6% when their data augmentation strategy was applied [23].

Methods

To perform my experiments, I implemented a MiniVGGNet convolutional neural network (CNN) implemented in TensorFlow. I did not make any modifications to this base architecture except to replace the final layer with a fully connected layer with 7 nodes, one per emotion, with softmax activation. I optimized the model with categorical cross-entropy loss.

A crucial step to achieving performant models is to apply data augmentation to increase the dataset size and the generalization capacity. I compared and contrasted data augmentation strategies on the Facial Expression Recognition (FER) 2013 dataset, which consists of 48x48 grayscale images of faces, to determine which method leads to a maximal increase in performance. I compared the following data augmentation strategies (Figure 1): featurewise center, featurewise standard normalization, zero-component analysis whitening, zero-component analysis epsilon, rotation, width shift, and height shift. I then examined the

benefit of data augmentation at various training set sizes, ranging from 100 images to 20,000 images.

To increase my confidence about my results, I applied 5 iterations of bootstrapped sampling to each condition and measured the mean and standard deviation of the bootstrapped samples.

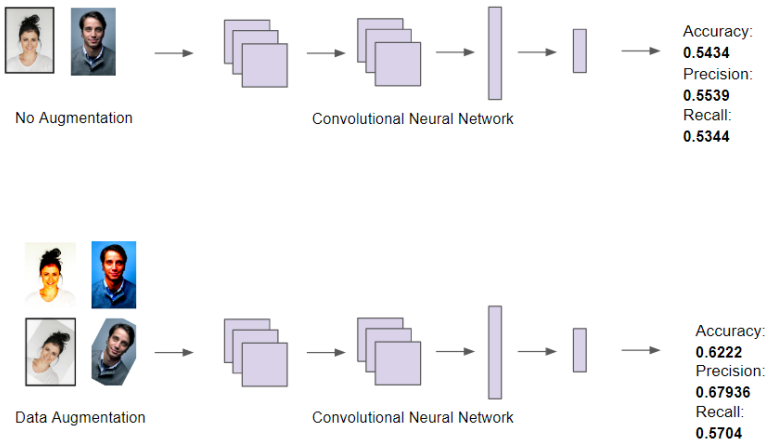

Figure 1. Summary of the experimental procedures. A convolutional neural network (CNN) is trained with and without data augmentation. The accuracy, precision, and recall on a held-out test set are compared for both models.

Results

When comparing data augmentation strategies, I found that the best-performing strategy was the width shift with a mean accuracy of 59.192% (Table 1). The worst-performing strategies were samplewise centering and samplewise standard normalization with a mean accuracy of 26.018% and 29.166% respectively (Table 1). However, there was no significant difference across augmentation strategies besides the significantly lower percentages for the samplewise centering and samplewise standard normalization.

Data augmentations that were used	**Test accuracy (note: this is the val_accuracy at epoch 50 / the last epoch)**
Featurewise Centering	0.54654 +/- 0.00862
Samplewise Centering	0.26018 +/- 0.02680
Featurewise Standard Normalization	0.54850 +/- 0.01064
Samplewise Standard Normalization	0.29166 +/- 0.03539
Zero-phase Component Analysis Whitening	0.54906 +/- 0.00366
Zero-phase Component Analysis Epsilon	0.55304 +/- 0.00587

Rotation Range	0.58722 +/- 0.00733
Width Shift Range	0.59192 +/- 0.00596

Table 1. Test accuracy (mean +/- standard deviation) on a held-out set from FER 2013 for several data augmentation strategies.

I also compared the effect of training set size on the effect of data augmentation (Table 2). I found that at all training set sizes, the performance metrics are consistently higher with data augmentation applied compared to when data augmentation is not applied.

Training set size	**Accuracy (mean +/- stdev)**		**Precision (mean +/- stdev)**		**Recall (mean +/- stdev)**	
	No augmentation	All augmentation strategies	No augmentation	All augmentation strategies	No augmentation	All augmentation strategies
100	**0.1766 +/- 0.00951**	**0.24048 +/- 0.00667**	**0.1839 +/- 0.01328**	**0.24846 +/1 0.00818**	**0.11774 +/- 0.01793**	**0.20772 +/- 0.00786**
200	**0.26 +/- 0.01377**	**0.27664 +/- 0.01235**	**0.2854 +/- 0.01225**	**0.28654 +/- 0.01484**	**0.2221 +/- 0.00937**	**0.24446 +/- 0.01491**
500	**0.2945 +/- 0.00489**	**0.3337 +/- 0.00830**	**0.32022 +/- 0.00550**	**0.3461 +/- 0.00875**	**0.2403 +/- 0.00312**	**0.31176 +/- 0.00739**
1000	**0.34788 +/- 0.00557**	**0.3917 +/- 0.01526**	**0.37432 +/- 0.00400**	**0.40582 +/- 0.01479**	**0.30412 +/- 0.00517**	**0.37418 +/- 0.01699**
2000	**0.383 +/- 0.00725**	**0.4474 +/- 0.00299**	**0.40694 +/- 0.00990**	**0.46238 +/- 0.00461**	**0.36326 +/- 0.02382**	**0.4286 +/- 0.00400**
5000	**0.434 +/- 0.00711**	**0.53128 +/- 0.00296**	**0.45064 +/- 0.00738**	**0.54636 +/- 0.01098**	**0.41826 +/- 0.00653**	**0.51674 +/- 0.00568**
10,000	**0.4853 +/- 0.00735**	**0.588 +/- 0.00130**	**0.50388 +/- 0.00939**	**0.60518 +/- 0.00279**	**0.47204 +/- 0.00455**	**0.57 +/- 0.00255**
15,000	**0.5066 +/- 0.00890**	**0.6192 +/- 0.00287**	**0.52106 +/- 0.00989**	**0.63532 +/- 0.00256**	**0.4882 +/- 0.00866**	**0.6024 +/- 0.00354**

20,000	**0.5434 +/- 0.00552**	**0.6222 +/- 0.00331**	**0.5539 +/- 0.00533**	**0.67936 +/- 0.00171**	**0.5344 +/- 0.00468**	**0.5704 +/- 0.00795**

Table 2. Comparison of data augmentation vs. no data augmentation for various training set sizes.

Discussion and Conclusion

The primary finding of this work is that data augmentation consistently improves the performance of an emotion recognition computer vision model regardless of the amount of training data used. From the quantification of the marginal utility of some of the most popular image-based data augmentation approaches for computer vision, I found that including out-of-the-box augmentation strategies can yield detrimental performance, likely due to maladaptive distribution shifts.

There are increasing efforts to build state-of-the-art emotion recognition models for digital therapies and behavioral phenotyping for increased automation of screening and detection tools [24-32]. Ultimately, the strategies I explored in this work can inform the development of digital therapies for autism which focus on the evocation and subsequent automatic detection of facial expressions.

The primary limitation of this study was the reduced selection of data augmentation strategies and the use of only a single emotion recognition dataset for both training and evaluation. In future iterations of this work, I will evaluate additional data augmentation strategies for an expanded array of emotions. It would also be fruitful to explore generative models for data augmentation using variational autoencoders or generative adversarial networks [33-35].

References

1. “Signs and Symptoms of Autism Spectrum Disorders.” *Centers for Disease Control and Prevention*, 9 Dec. 2022, www.cdc.gov/ncbddd/autism/signs.html.
2. “What Is Autism?” *Autism Speaks*, www.autismspeaks.org/what-autism.
3. “What Is Autism Spectrum Disorder?” *Centers for Disease Control and Prevention*, 9 Dec. 2022, www.cdc.gov/ncbddd/autism/facts.html.
4. “What Is Autism Spectrum Disorder?” *American Psychiatric Association*, https://www.psychiatry.org/patients-families/autism/what-is-autism-spectrum-disorder.
5. Dattaro, Laura. “Difficulty Identifying Emotions Linked to Poor Mental Health in Autistic People: Spectrum: Autism Research News.” *Spectrum*, 12 Nov 2020, www.spectrumnews.org/news/difficulty-identifying-emotions-linked-to-poor-mental-health-in-autistic-people/.
6. Brewer, Rebecca, and Jennifer Murphy. “People with Autism Can Read Emotions and Feel Empathy.” *Spectrum*, 12 July 2016, www.spectrumnews.org/opinion/viewpoint/people-with-autism-can-read-emotions-feel-empathy/.

7. Kline, Aaron, et al. "Superpower glass." *GetMobile: Mobile Computing and Communications* 23.2 (2019): 35-38.
8. Voss, Catalin, et al. "Superpower glass: delivering unobtrusive real-time social cues in wearable systems." *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 2016.
9. Voss, Catalin, et al. "Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial." *JAMA pediatrics* 173.5 (2019): 446-454.
10. Daniels, Jena, et al. "Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism." *NPJ digital medicine* 1.1 (2018): 32.
11. Washington, Peter, et al. "SuperpowerGlass: a wearable aid for the at-home therapy of children with autism." *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1.3 (2017): 1-22.
12. Washington, Peter, et al. "SuperpowerGlass: a wearable aid for the at-home therapy of children with autism." *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1.3 (2017): 1-22.
13. Kulkarni, Nitish. “Stanford Researchers Treat Autism with Google Glass.” *TechCrunch*, 19 Oct. 2015, techcrunch.com/2015/10/19/stanford-researchers-treat-autism-with-google-glass/.
14. Chatziagapi, Aggelina, et al. "Data Augmentation Using GANs for Speech Emotion Recognition." *Interspeech*. 2019.
15. Ahmed, Tawsin Uddin, et al. "Facial expression recognition using convolutional neural network with data augmentation." *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE, 2019.
16. Tong, Xiaoyun, Songlin Sun, and Meixia Fu. "Data augmentation and second-order pooling for facial expression recognition." *IEEE Access* 7 (2019): 86821-86828.
17. Xu, Tian, et al. "Investigating bias and fairness in facial expression recognition." *European Conference on Computer Vision*. Springer, Cham, 2020.
18. Sajjad, Muhammad, et al. "Human behavior understanding in big multimedia data using CNN based facial expression recognition." *Mobile networks and applications* 25.4 (2020): 1611-1621.
19. Yang, Biao, et al. "Facial expression recognition using weighted mixture deep neural network based on double-channel facial images." *IEEE access* 6 (2017): 4630-4640.
20. Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." *arXiv preprint arXiv:2105.03588* (2021).
21. Zhu, Xinyue, et al. "Emotion classification with data augmentation using generative adversarial networks." *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Cham, 2018.
22. Tan, Lianzhi, et al. "Group emotion recognition with individual facial emotion CNNs and global image based CNNs." *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 2017.
23. Psaroudakis, Andreas, and Dimitrios Kollias. "MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

24. Carpenter, Kimberly LH, et al. "Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism." *Autism Research* 14.3 (2021): 488-499.
25. Chi, Nathan A., et al. "Classifying Autism from Crowdsourced Semi-Structured Speech Recordings: A Machine Learning Approach." *arXiv preprint arXiv:2201.00927* (2022).
26. Egger, Helen L., et al. "Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study." *NPJ di*
27. Kalantarian, Haik, et al. "A mobile game for automatic emotion-labeling of images." *IEEE transactions on games* 12.2 (2018): 213-218.
28. Kalantarian, Haik, et al. "Guess What? Towards Understanding Autism from Structured Video Using Facial Affect." *Journal of healthcare informatics research* 3 (2019): 43-66.
29. Penev, Yordan, et al. "A mobile game platform for improving social communication in children with autism: a feasibility study." *Applied clinical informatics* 12.05 (2021): 1030-1040.
30. Sapiro, Guillermo, Jordan Hashemi, and Geraldine Dawson. "Computer vision and behavioral phenotyping: an autism case study." *Current Opinion in Biomedical Engineering* 9 (2019): 14-20.
31. Washington, Peter, et al. "Improved Digital Therapy for Developmental Pediatrics Using Domain-Specific Artificial Intelligence: Machine Learning Study." *JMIR Pediatrics and Parenting* 5.2 (2022): e26760.
32. Washington, Peter, et al. "Training an emotion detection classifier using frames from a mobile therapeutic game for children with developmental disorders." *arXiv preprint arXiv:2012.08678* (2020).
33. Antoniou, Antreas, Amos Storkey, and Harrison Edwards. "Data augmentation generative adversarial networks." *arXiv preprint arXiv:1711.04340* (2017).
34. Calimeri, Francesco, et al. "Biomedical data augmentation using generative adversarial neural networks." *Artificial Neural Networks and Machine Learning–ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part II 26*. Springer International Publishing, 2017.
35. Zhu, Xinyue, et al. "Emotion classification with data augmentation using generative adversarial networks." *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part III 22*. Springer International Publishing, 2018.