# Synthetic Data-driven Approaches to Evaluate Convolutional Neural Network Robustness

Ronald Nap, Cristian Espinosa, Baixi Guo, Kyle Wright, Cory McCullough
School of Natural Science, University of California, Merced

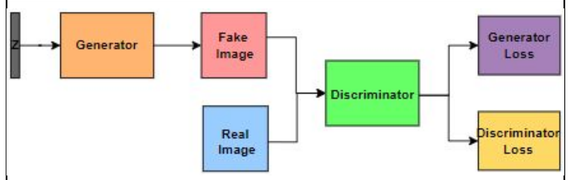
## Introduction

**Motivation:** Accurate and robust deep learning models require large amounts of data to solve complex tasks such as medical image analysis and emotion recognition. However, it poses a challenge due to the scarcity of authentic data.

**Goals:** Generate synthetic images based on original human emotion images and investigates its impact on robustness of the deep learning model to increase amount of data available utilized for deep learning models. This study involves 3 stages:

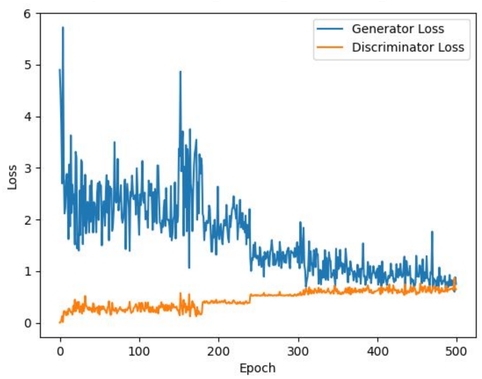
- Stage 1: Synthesize data via a Generative Adversarial Network (GAN) to supplement a limited amount of data.
- Stage 2: Compare the accuracy and robustness of Convolutional Neural Network (CNN) models when trained on authentic data vs synthetic data.
- Stage 3: Investigate the optimal split between synthetic and authentic training datasets to assess the accuracy and robustness of our model utilizing synthetic data.

## Stage 1: GAN




Generative Adversarial Network Architecture

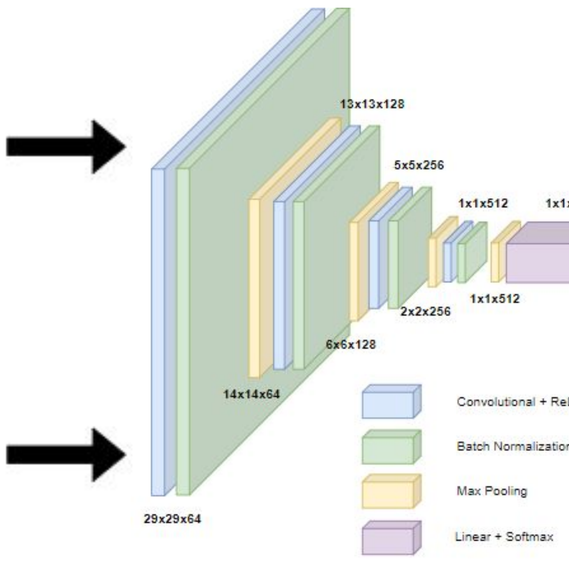
- Generator takes random noise (z) as input and generates synthetic data resembling the real data.
- Discriminator distinguishes between real and synthetic data, and classifies them as real or fake.
- The training process is adversarial as the generator and discriminator compete against each other.
- As the generator improves in generating realistic data, it challenges the discriminator's task, leading to the discriminator providing more accurate feedback as it improves in distinguishing between real and fake data.
- Once trained, the generator can independently produce new data samples resembling the original training data.




## Stage 2: CNN Classifier

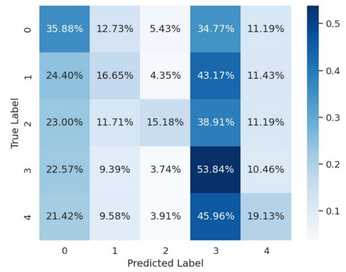
Synthetic Image Inputs

Authentic Image Inputs



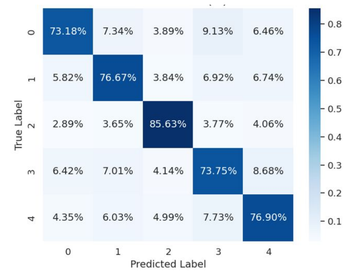

Convolutional Neural Network Architecture:

| | |
|---|---|
| 0 | Fear |
| 1 | Angry |
| 2 | Happy |
| 3 | Sad |
| 4 | Neutral |




Performance Metric:
Confusion Matrix

| | Precision | Recall |
|---|---|---|
| Fear | 0.38 | 0.30 |
| Angry | 0.14 | 0.35 |
| Happy | 0.15 | 0.44 |
| Sad | 0.24 | 0.36 |
| Neutral | 0.31 | 0.27 |

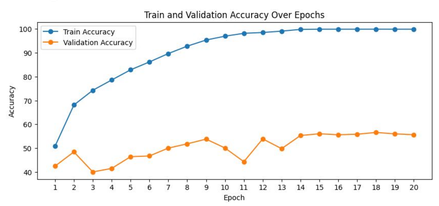


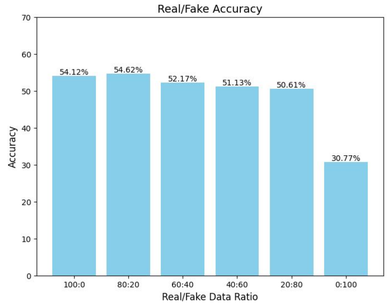

Performance Metric:
Precision and Recall

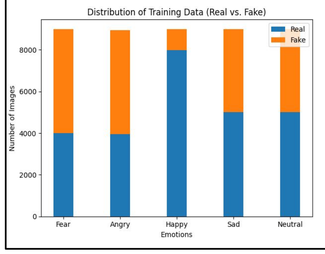
| | Precision | Recall |
|---|---|---|
| Fear | 0.44 | 0.40 |
| Angry | 0.46 | 0.49 |
| Happy | 0.64 | 0.70 |
| Sad | 0.40 | 0.40 |
| Neutral | 0.55 | 0.48 |

## Stage 3: Authentic & Synthetic Data Split

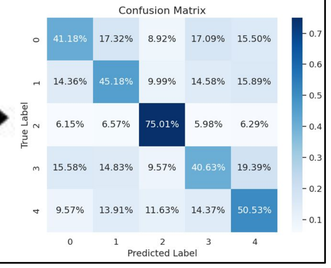
- Synthetic images generated by the GAN are used to address class imbalance issues in the original dataset.
- Only real images are used for inference, while generated images are exclusively utilized for training the model.
- 45,000 evenly distributed real images per class in the training set and 5,000 real images in the validation set.




**CNN Classifier**

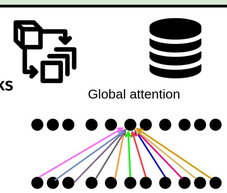



## Results and Conclusion

- The split of synthetic and authentic data reveals no significant difference in accuracy when utilizing synthetic data.
- Incorporating the generated images alongside the original dataset results in a 2% increase in accuracy compared to using the original data alone.

## Next Steps

- Train on extracted feature vectors opposed to original images
- Explore different types of GAN Architecture, Loss Functions
- Add Self-Attention layer to generator and discriminator networks after convolutional layers.
- Utilize the LSUN dataset, a larger and more complex dataset containing a collection of diverse indoor and outdoor images.




## Acknowledgements & References

- The authors acknowledge the support of the NSF funding to the RESUME Applied Math program at UC Merced

Reference:
- Goodfellow, Pouget-Abadie, et al, “Generative Adversarial Nets.” (2014).