Facial Emotion Recognition

By Manoj Kumar
Liverpool John Moores University
Under Supervision of Karthik O S

Introduction

- Humans utilize facial expressions to communicate a variety of meanings

- Rapid advancement of artificial intelligence has caused automatic facial expression detection systems to gain a lot of attention

- Facial emotion recognition (FER) acts as a tool to automatically detect human expressions.

- FER is a key branch of computer vision research

- It has made significant progress in distinguishing basic human emotions such as happy, angry, fear, sad, surprise, disgust, and neutral in controlled situations.

Introduction

- Applications of FER exist in many industries such as:
  - Finding mental illnesses
  - Human social interaction systems.
  - Fatigued driving monitoring
  - Market Research
  - Engagement recognition systems for students
  - Interactive games utilizing human expressions.
  - Customized movie summaries depending on the audience’s emotions
  - Medical devices
  - Social robots many other human–computer interaction systems.

Problem Statement

- FER has made significant progress in distinguishing basic human emotions in controlled situations.
- Still there are challenges when we use these systems to classify emotions on natural and uncontrolled condition
- Though FER has been the subject of countless studies, it is still one of the most challenging problems for image classification systems and key reasons are:
  - Substantial similarity between the basic emotion types
  - Cultural variations in emotion expression
  - Identity of an individual
  - Variables which are unrelated to expressions
  - Subject dependence
  - Shortage of wide variety of training data – DL models require a lot of training data to prevent overfitting
  - Head attitude, pose invariant, direction, illumination
  - Bias due to class imbalance
  - Image resolution challenges
  - No cross-database analyses - due to which the accuracy is deceptively high.

Aims and Objectives

- The main aim is to propose an effective automated Facial Emotion Recognition (FER) to recognize basic emotions like anger, happiness, fear, sadness, surprise, disgust and neutral on images with difficult naturalistic conditions and challenges
- Objectives of the study are as follows:
  - To check the robustness of algorithm with various data augmentation techniques
  - To implement various model pruning techniques for improved performance and deliver a compact model
  - To explore existing well known CNN architecture as well and train on dataset to get the best performance
  - To compare the results and performance of models achieved with these different approaches

Literature Review

Key findings

- Since 2013, contests like FER2013 and Emotion Recognition in the Wild (EmotiW) have indirectly encouraged the shift of FER from lab-controlled to in-the-wild settings the by gathering reasonably sufficient training data from difficult real-world circumstances. (Goodfellow et al., 2013) (Dhall et al., 2015, 2017)
- Dynamic-based methods consider the temporal relationship among consecutive frames in the input facial expression sequence. (Zhao and Pietikäinen, 2007; Jung et al., 2015; Zhao et al., 2016)
- Static-based methods solely encode the feature representation with spatial information from the current single image. (Shan et al., 2009; Liu et al., 2014; Mollahosseini et al., 2015)
- Based on these two vision-based techniques, multimodal systems have also employed other modalities to aid in the identification of expression, including audio and physiological channels. (Corneanu et al., 2016)

Literature Review

Key findings

- Challenges have been observed in recognizing human emotions under real circumstances
- Recent research work has been around solving these challenges to recognize human emotions under real circumstances.
- The quality of feature vectors may have an impact on how well the classifier performs, making feature extraction an essential step in FER.
- Deep learning approaches are being used more frequently to address the difficult aspects of emotion recognition in the wild due to well designed architecture and dramatically increased chip processing abilities/ GPUs
- In the recent era, CNN has been the most often utilized DL method for extracting important characteristics.
- Researches have tried handle the challenge of variations in real circumstances by augmentatng the databy scaling, horizontally/ vertically shifting, rotation, cropping etc. (Khaireddin and Chen, n.d.)
- Researchers presented a way to optimize the hyperparameters of CNN to improve accuracy specifically for FER. (IEEE Romania Section et al., n.d.)

Literature Review

Key findings

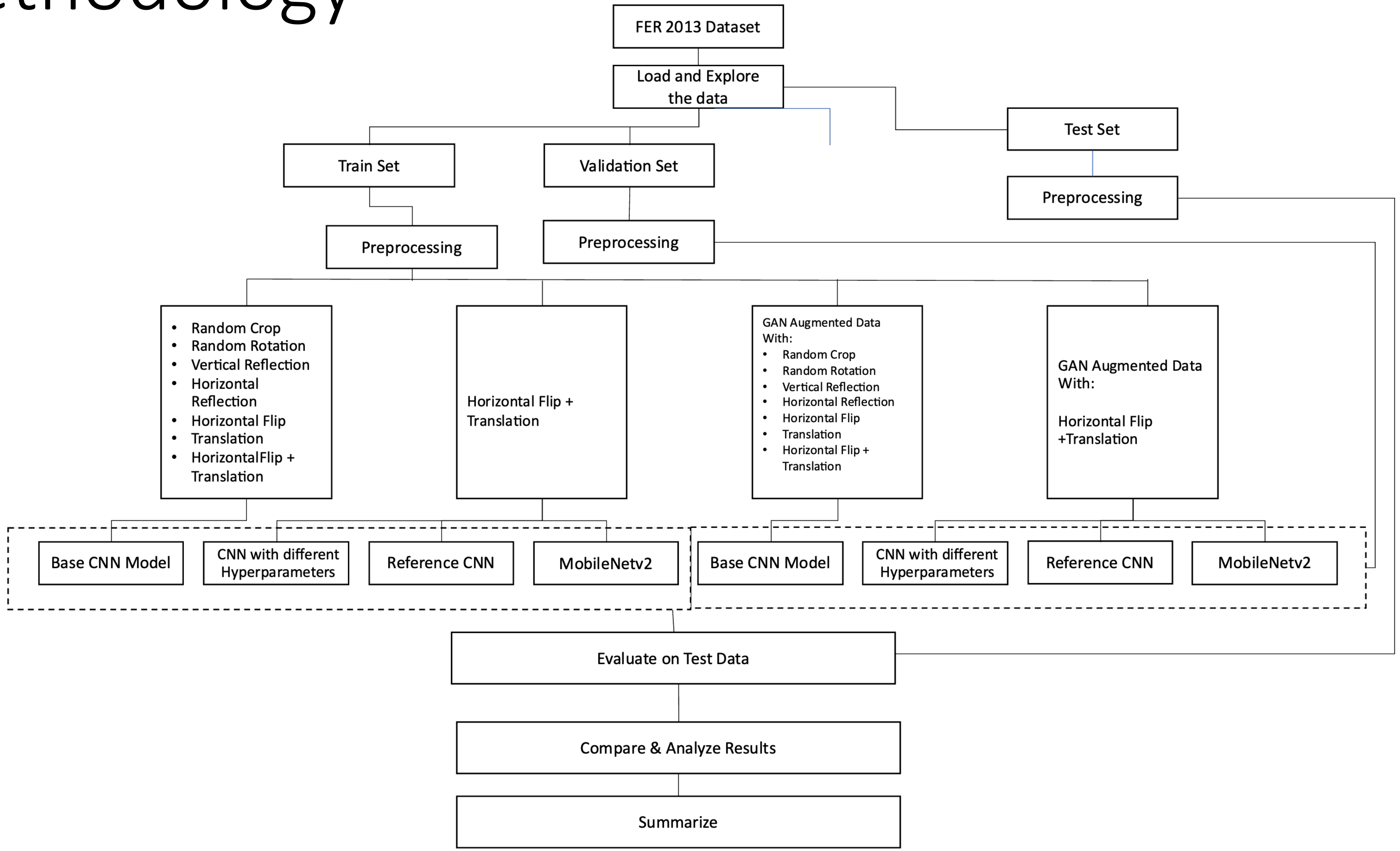
- Researchers hypothesized that by combining information from diverse domains gender, emotion, race, and age etc. a CNN subjected to multi-task learning on the same data can perform better on individual tasks (Saroop et al., 2021)
- Researchers have tried ton address the challenges occlusion-resistant and pose-invariant (Wang et al., 2019).
- Researchers have used Super Resolution method to up-scale the low-resolution images and address the challenge of different size images (Vo et al., 2020)
- GAN framework pits one deep network (the generative) against the other (the discriminative). The neural network can generate fresh data using this method. (Goodfellow et al., n.d.).
- GAN has been used in FER to address various problems including subject dependence, Inter-class imbalance, Noisy labels, limited training datasets, individual differences etc.

Literature Review

Key findings

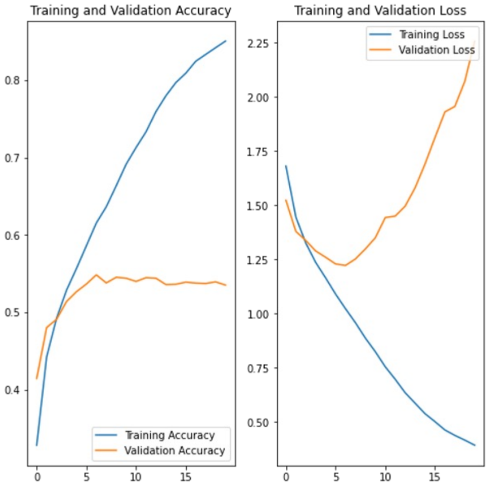
- Researchers have handled the problem of subject dependent by proposed Auxiliary Classifier Generative Adversarial Network (AC-GAN) which redevelops from input face ten expressions (Dharanya et al., 2021)
- Researchers have used GAN to augment the data (Porcu et al., 2020)
- (Kim and Won, 2020) has used GAN to work on the issue of Inter-class imbalance and Noisy labelling
- Feature separation exchange-GAN has been proposed by (Yang et al., 2020) to handle the interference of individual differences in FER accuracy
- GAN has also been used for image super-resolution, restoration of image and re-identification of individuals to improve and extract richer features. (Ledig et al., 2016)

Methodology

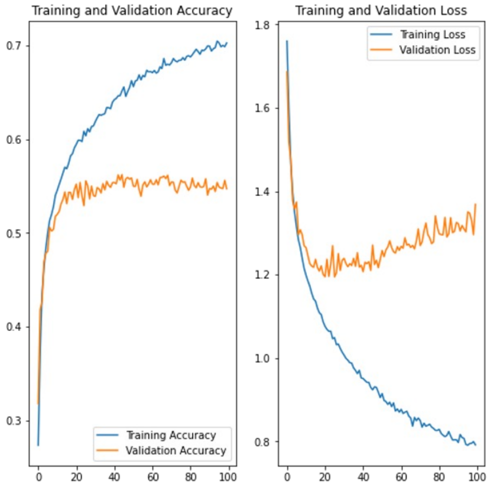


Result and Discussion

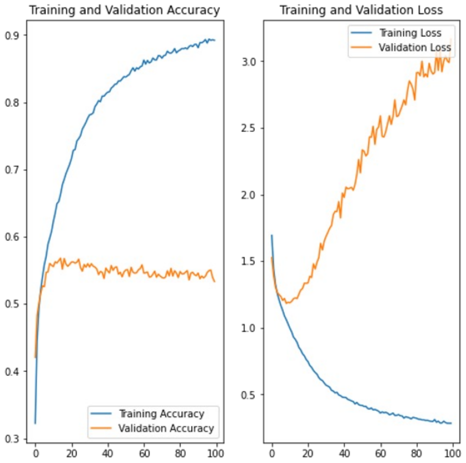
Impact of Experimented Geometric based augmentation on base CNN



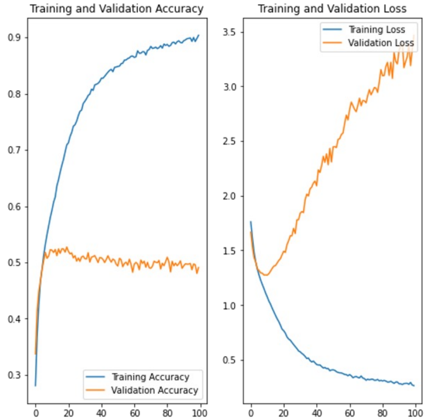
No Augmentation



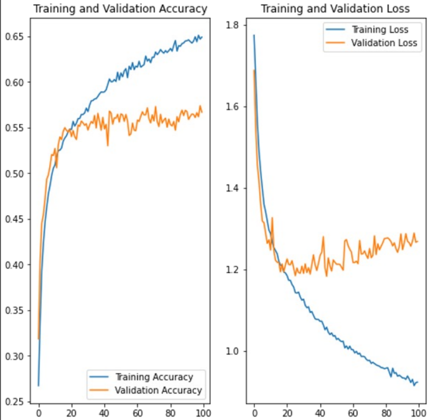
Random Crop



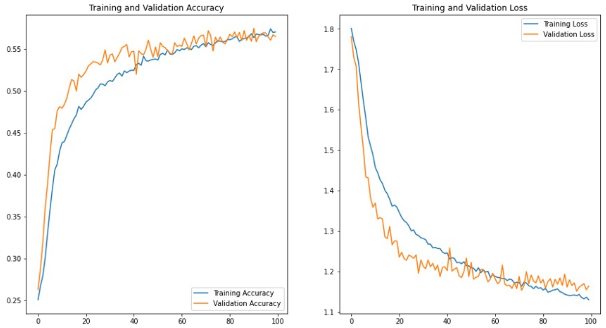
Random Rotation



Vertical Reflection



Horizontal Flip



Translation

GAN Augmentation

Bico GAN (Jaiswal et al., 2017) with Generator having 2,405,209 parameters and discriminator having total 1,659,621 parameters was trained for 18000 + epochs. Generator was saved to provide synthetic images (3000 images for each emotion) to augment with custom code,.

Epoch 17800

Epoch 18000

Epoch 18200

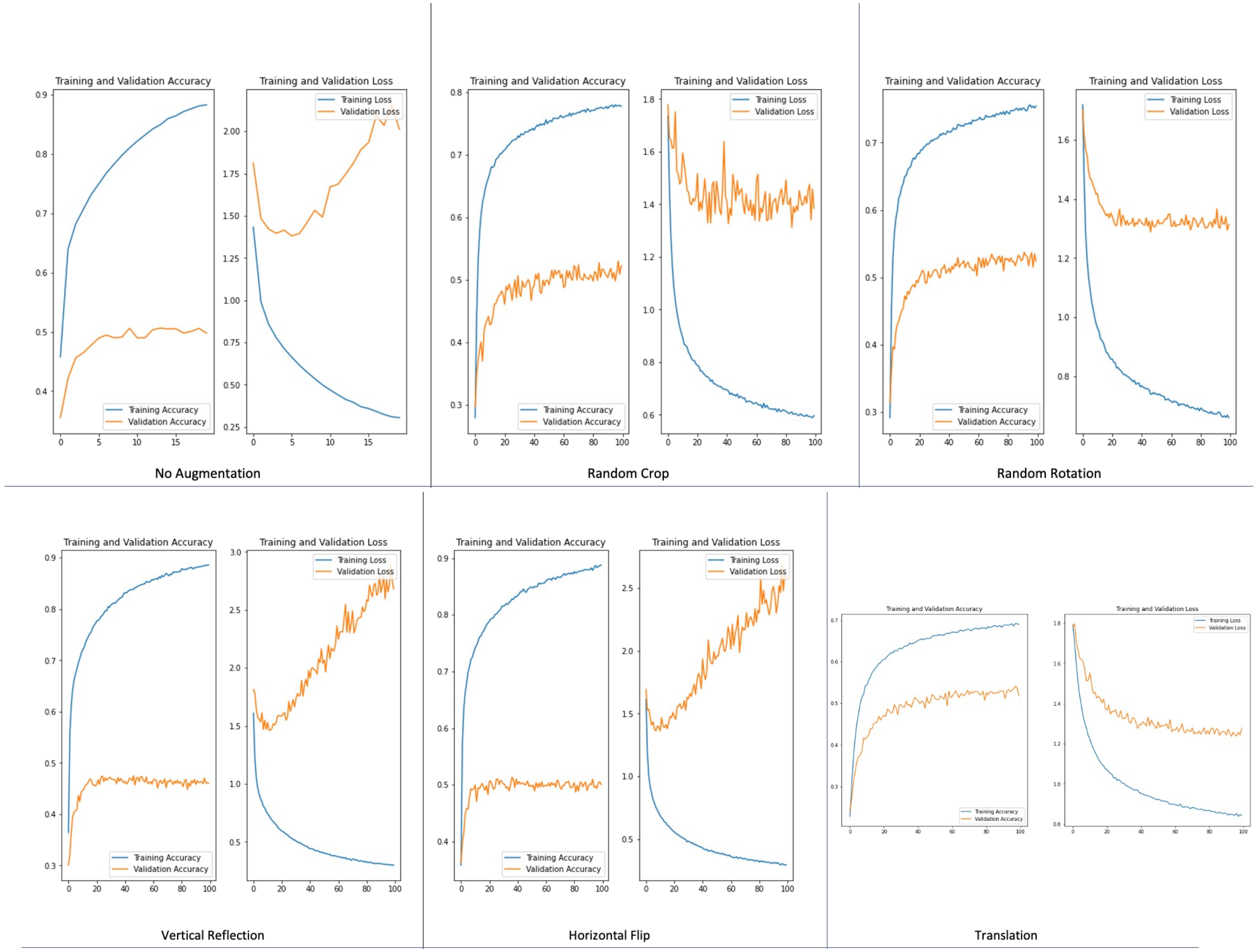
Epoch 18400

Epoch 18600

Label	FER2013 Train Images	GAN Generated Images	GAN Augmented Data
surprise	3171	3000	6171
sad	4830	3000	7830
angry	3995	3000	6995
neutral	4965	3000	7965
disgust	436	0	436
fear	4097	3000	7097
happy	7215	3000	10215
Total	28709	18000	46709

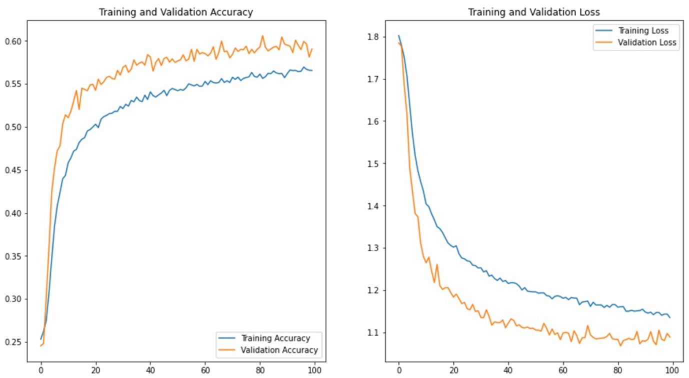
Result and Discussion

Impact of Experimented Geometric based augmentation on base CNN (GAN augmented data)

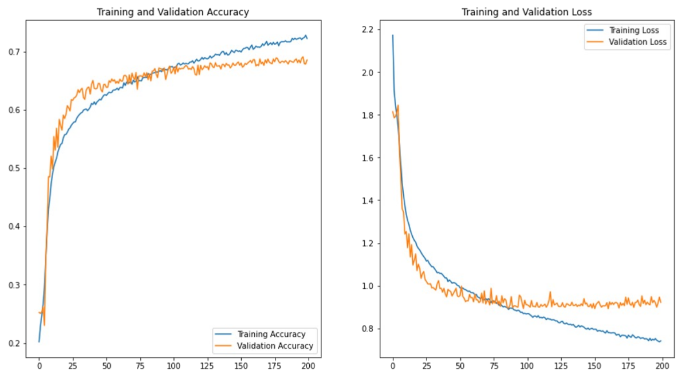


Result and Discussion

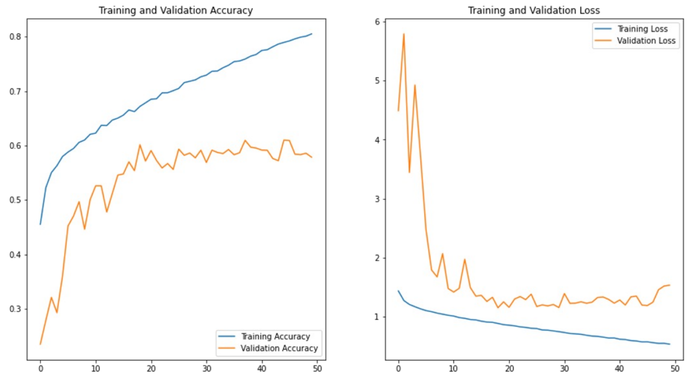
Performance of models with chosen Augmentation



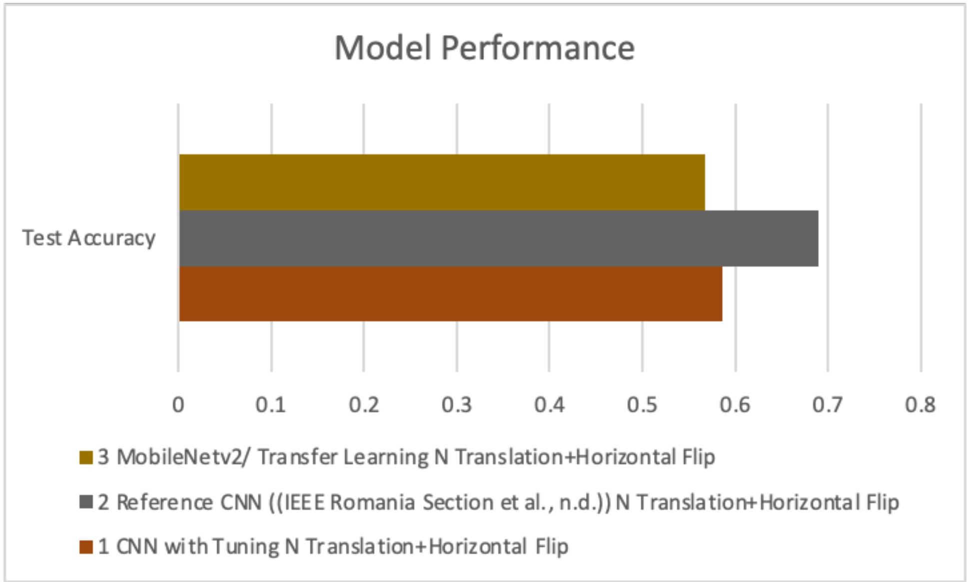
Base CNN with Tuning



Reference CNN



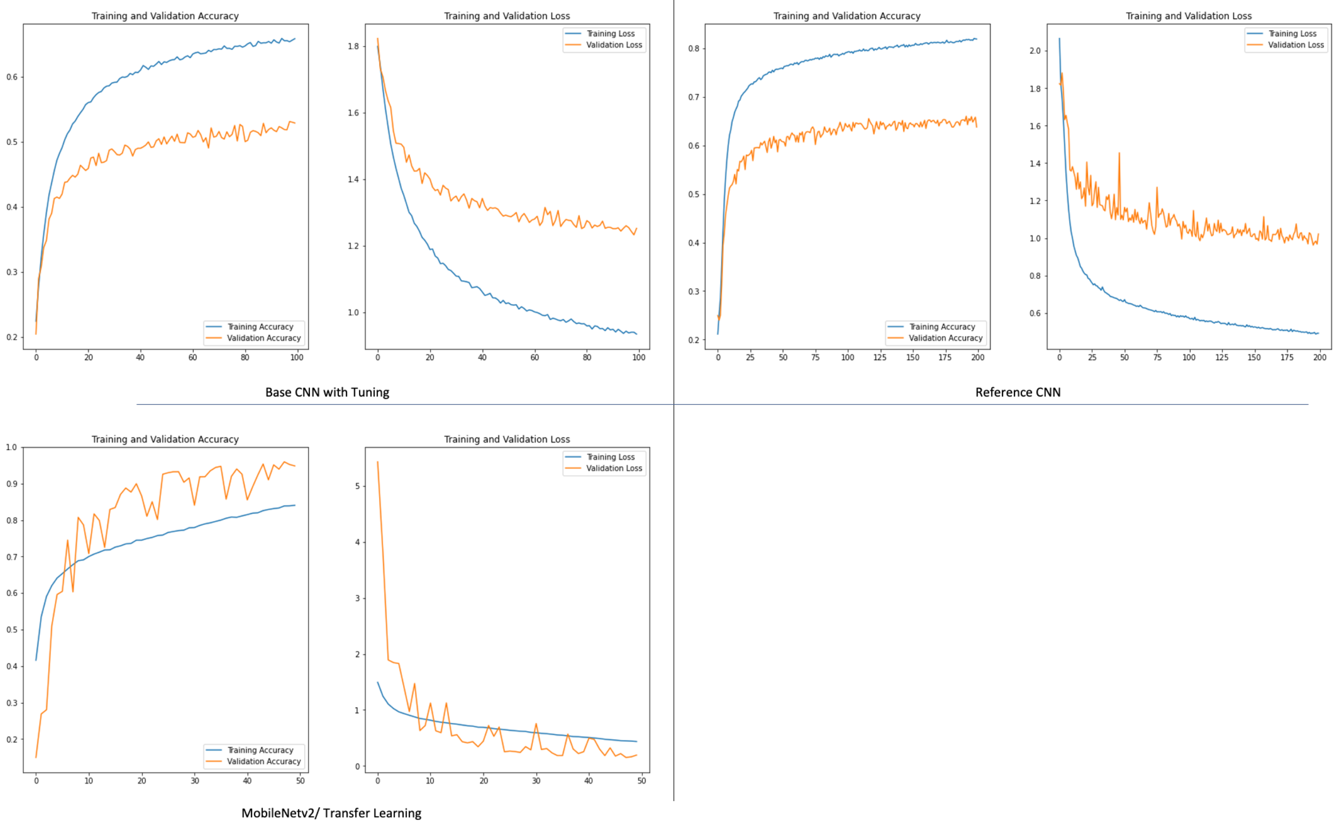
MobileNetv2/ Transfer Learning

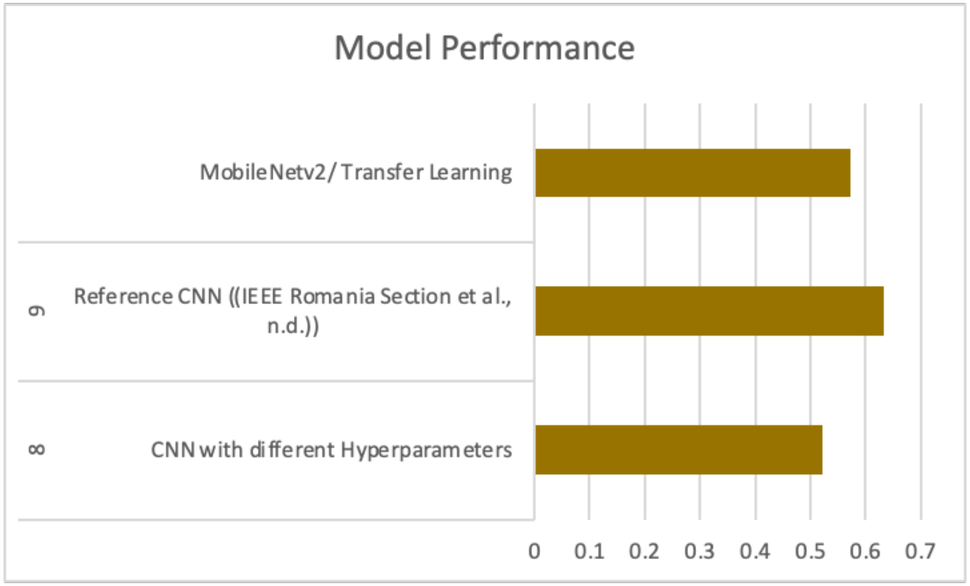


SN	Model	Data Augmentation	Trainable Parameters	Train	Validation	Test	Train	Validation	Test
1	CNN with Tuning	Translation+Horizontal Flip	2,45,927	1.1607	1.0682	1.0867	0.5565	0.6062	0.5863
2	Reference CNN ((IEEE Romania Section et al., n.d.))	Translation+Horizontal Flip	51,75,111	0.7473	0.8989	0.9358	0.7229	0.6908	0.6892
3	MobileNetv2/ Transfer Learning	Translation+Horizontal Flip	23,96,551	0.5749	1.1984	2.1595	0.7894	0.6098	0.5682

Result and Discussion

Performance of models with chosen Augmentation (GAN Augmented Data)





					Loss			Accuracy		
SN	Model	GAN Augmentation	Data Augmentation	Trainable Parameters	Train	Validation	Test	Train	Validation	Test Accurac
8	CNN with different Hyperparameters	Y	Translation+Horizontal Flip	2,45,927	0.9401	1.2431	1.24	0.654	0.5313	0.5231
9	Reference CNN ((IEEE Romania Section et al., n.d.))	Y	Translation+Horizontal Flip	51,75,111	0.4909	0.9682	1.0115	0.8186	0.6603	0.633
10	MobileNetv2/ Transfer Learning	Y	Translation+Horizontal Flip	23,96,551	0.4476	0.1507	1.7632	0.8389	0.9599	0.5723

Conclusion and future works

- Observed the impact of geometric transformation based augmentation techniques with and without the presence of GAN generated images in source data
- Horizontal Flip and Translation were best performing techniques as observed in this work.
- Generated synthetic images using BicoGAN model post a heavy training process, those augmentation though helped to achieve high train accuracy but validation and test accuracy did not improve with GAN augmentation.
- GAN can be useful if generate quality images with very close match to real images and then can help to improve model performance, gain high accuracy and generalize well on test data.

Conclusion and future works

Review Goals

- To check the robustness of algorithm with various data augmentation techniques
  - Thoroughly analyzed the impact of augmentation techniques individually as well as the combination chosen based on performance seen with individual techniques.
  - Analyzed how GAN augmentation impacts and compared results without GAN augmentation.
- To implement various model pruning techniques for improved performance and deliver a compact model
  - Study experimented with a CNN with 319527 parameters, reference CNN with 5175111 parameters and MobileNetV2/ transfer learning with 2396551 parameters. Compared to top performing model on FER 2013 datasets given parameters are very low and thus are very compact.
- To explore existing well known CNN architecture as well and train on dataset to get the best performance
  - Explored MobileNetv2 with transfer learning to train on FER 2013 dataset
  - Trained MobileNetV2 on GAN augmented data as well to see the impact
- To compare the results and performance of models achieved with these different approaches
  - This study compares all the results achieved with the experiments made and summarize to give a conclusion.

Conclusion and future works

Future Work

- Further research can be done by having a high performance GAN to generate quality images on FER 2013 data set with a close match to real images
- “disgust” emotion images were not augmented with GAN in this study. Further research can address this and augment these emotion images as well to handle imbalance
- This can be doable by using other GAN variant which can generate this emotion images by converting another emotion images.

Thank You!