Recognition of Facial Expressions and Gender Classification on Images

Suraj Pandey
M Tech CSE
IIIT, Delhi
New Delhi, India - 110020
Email: suraj18025@iiitd.ac.in

Tushar Agrawal
M Tech CSE
IIIT, Delhi
New Delhi, India - 110020
Email: tushar18137@iiitd.ac.in

Udit Pant
M Tech CSE
IIIT, Delhi
New Delhi, India - 110020
Email: udit18049@iiitd.ac.in

I. Problem Statement and Motivation

Facial expression and gender detection have broad applications in the field of computer vision and robotics. However, they impose quite a number of challenges. The main objective of our project is to recognize the seven key human emotions: anger, disgust, fear, happiness, sadness, surprise and neutrality along with gender detection of a person from given facial image.

II. Literature Review

Shan et al. propose a facial expression recognition system on local binary patterns [1]. The empirical study on facial representations is based on local binary pattern (LBP) features for person-independent facial expression recognition. Since LBP features are tolerant against illumination changes and are simple in implementation, these serve as capable features for recognizing facial expressions. The main contributions include the use of LBP features applied on low-resolution images and the application of Boosted-LBP by learning the most discriminative histograms using AdaBoost algorithm. The main reason for using low-resolution images is to simulate a real-world scenario where images are not obtained in the form of high-resolution frontal faces from highly controlled environments (with little or no noise). Ekman et al. identified six key emotions through facial recognition of images namely - anger, disgust, fear, happiness, sadness and surprise [2]. Savoiu et al. propose a deep-learning based approach to identify these emotions using support vector machine as the baseline model. Evaluation of performance of models has been done using the standard statistical tools such as precision and recall [3]. Li et al. propose deep learning methods for recognizing facial expressions from candid images. Baseline approaches for facial expression recognition include two popular feature extraction techniques - LBP and SIFT and a learning-based convolutional neural network model [4].

III. Dataset Details

We have used two datasets for emotion recognition and gender classification. Dataset details are as follows:

- **FER dataset:** This dataset [5] consists of 35887 images of size 48*48 pixels (8 bit grayscale). It contains images of facial expressions which form the seven key emotions in humans - Neutrality, Disgust, Sadness, Happiness, Fear, Surprise and Anger. [6]

Fig. 1: Sample of Seven key emotion in Kaggle data set

- **Adience dataset:** This dataset [7] was used for gender classification which comprises of 2,284 subjects with a total of 19370 photos(around 8K males and 11K females)[8].

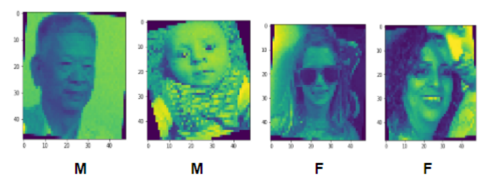

Fig 2: Sample Images with labelled genders in Adience Dataset

IV. Data Analysis and Preprocessing

Both datasets has some problems which need to be handled before going onto the task of classification.

- **FER Dataset:** This dataset is imbalanced. As seen in the figure 3, class 1 i.e. Disgust has only 400 data points.

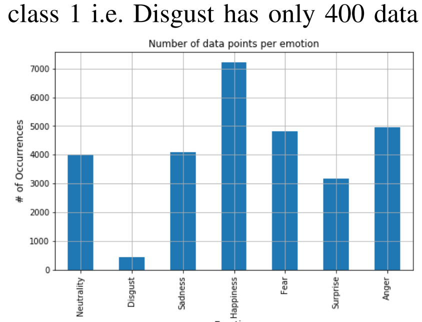

Fig 3: Classwise Datapoints frequencies in FER Dataset

So in order to balance data, we use the technique of **data augmentation**. We augment the data by using operations like zooming, horizontal translation and vertical translation. We then get augmented data with 42000 images in total(6000 images per class). From figure 4, we can say that augmentation does not change the distribution of original dataset. Since images are good in quality, so there is no need of preprocessing on images.

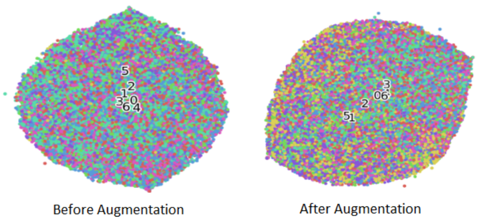

Fig 4: 2D Representation of training data before and after augmentation

- **Adience Dataset:** This dataset is also imbalanced as number of data points for male is around 8000 whereas it is 11000 for females.

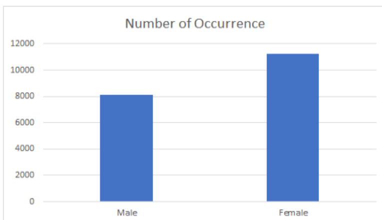

Fig 5: Classwise datapoint frequencies in Adience Dataset

So, just to balance the data, we randomly sample the data and take 8000 data points from each class. We ended up in total 16000 data points for gender detection.

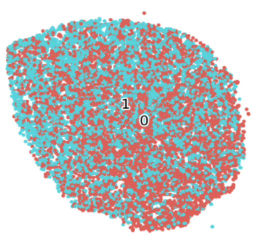

Fig 6: 2D representation of training data in Adience dataset

V. Proposed Architecture

Since we have two problems to solve i.e. facial expression recognition and gender detection, so we are splitting our proposed architecture in the two parts.

- **For Facial Expression Recognition** Landmarks detection is used to extract 68 facial features and thus plays an important role in improving the results. Apart from the space coordinates, we add two more features - distance from centroid and rotation angle to our feature set. Distance from centroid calculates the position of all points relative to each other and rotation angle is the correction angle for face alignment. The pipeline for FER using landmark detection as feature selection and logistic regression as classification technique is mentioned in the figure:

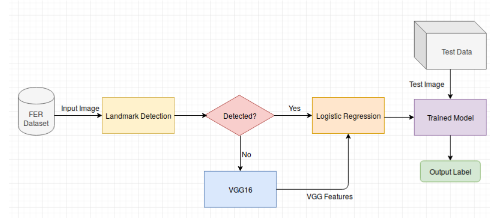

Fig: Proposed Architecture for Facial Expression Recognition

- **For Gender Detection:** We used VGG16 pretrained model, to see the classification accuracy from neural network model because VGG16 is a deep neural network with 16 hidden layers. The architecture allows even moderately powerful systems to run such a network. Due to this advantage, we preferred the use of this model for extracting features from the training images. This was achieved by taking fc7 layer output which has dimension of 4096. When we used the HOG, it gave better accuracy because HOG extract the gradient change features in images and VGG16 will extract the low level features and hardware wise feasible.

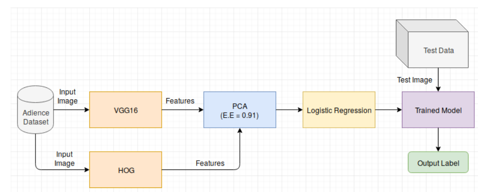

Fig: Proposed Architecture for Gender Detection

VI. Visualisations

- In FER Dataset, we are using **landmarks detection** as a feature extraction technique. Landmark detections extract 68 facial features which serve our purpose of facial expression recognition with a good accuracy.

Fig 7: Landmarks Detection in FER dataset

- Although landmarks detection is a good feature extraction technique, yet there were some images in training set whose landmarks could not be extracted as mentioned in figure 8 along with the image number. This is due to the fact that in image 4, most of the features are overlapped with the hand, in image 18, some features are overlapped with the hand and image is blurred ,image 60 is not even a face.

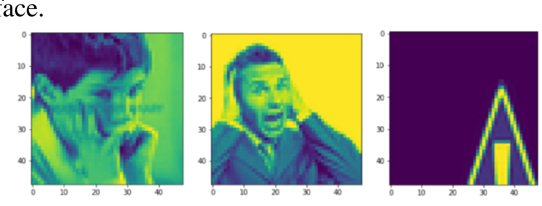

Fig 8: Training set Images whose landmarks could not be extracted in FER dataset

- On applying PCA on images in Adience dataset, we get eigen faces and it can be seen from figure 9 that they show good facial features due to which these features seem to be good enough for classifying gender.

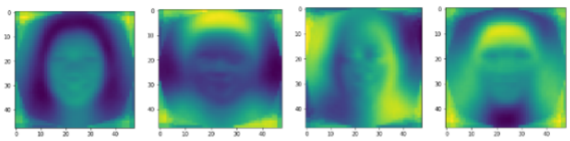

Fig 9: Eigen Faces in Adience Dataset

VII. Results

The images in both datasets were used in two ways - firstly, by using the intensity values of each pixel and secondly, by using the histogram values of each pixel. Following are the set of test accuracies obtained for different classifiers:

Classifier	**Accuracy**	
	Intensity	**Baseline**
NaiveBayes	23.11	-
Logistic Regression	35.83	-
SVM (RBF Kernel)	**30.58**	**31.8**
Adaboost	22.15	-
Bagging	36.58	-

Table 1: Cross Validation Accuracies for FER dataset

Features	**Accuracy**
LBP(PCA n=30)	25.83
Intensity (PCA n=30)	27.79
HOG (PCA n=30)	36.63
HOG + LBP (PCA n=30)	36.89
VGG16 + Alexnet (PCA n=1000)	40.27
Alexnet (PCA n=1000)	43.05
VGG16(LDA)	47.78
VGG16(PCA n=1000)	48.27
Landmarks	**55.08**

Table 2: Cross Validation Accuracies for FER dataset using Logistic Regression with different combinations of features

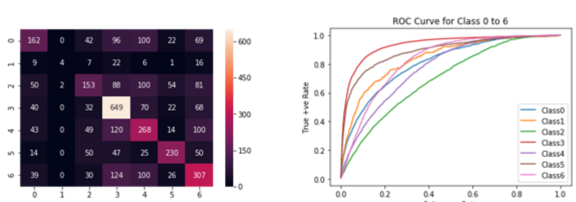

Figure: Confusion Matrix and ROC curve for FER dataset

Classifier	**Accuracy**	
	Intensity	**Baseline**
NaiveBayes	59.11	-
LogisticRegression	62.04	-
SVM (RBF Kernel)	**56.49**	**46.21**
Adaboost	42.65	-
Bagging	61.86	-

Table 3: Test Accuracy for Adience dataset

Features	**Accu-racy**	**Preci-sion**	**Re-call**	**F1-measure**
LBP (PCA n=30)	64.84	64.84	64.84	64.84
HOG (PCA n=30)	72.89	72.89	72.89	72.89
Alexnet (LDA)	73.13	73.13	73.13	73.13
HOG + LBP (PCA n=30,30 each)	74.24	74.24	74.24	74.24
VGG16	75.70	75.70	75.70	75.70
Alexnet (PCA n=1000)	75.58	75.58	75.58	75.58
Alexnet (PCA n=500) + HOG (30)+LBP(30)	75.06	75.06	75.06	75.06
VGG16(PCA n=1000) + HOG(30)	**76.1**	**77.0**	**78.1**	**79.2**

Table 4: Test Accuracy for Adience dataset using Logistic Regression with different combination of features

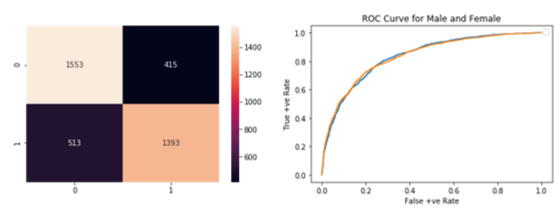

Figure: Confusion Matrix and ROC curve for gender dataset

VIII. Analysis of Results

- PCA worked better as compared to LDA - It is because LDA caused overfitting on the training data as it tries to reduce dimensions and perform classification at the same time which did not seem to work well with the test data. This means classification performed at a lower dimension worked poorly for both datasets.
- Facial Landmarks are useful features for capturing the emotional aspects of face as they boosted the classification accuracy considerably.
- Naive Bayes performed poorly on landmarks - this is due to the naive assumption of Naive Bayes which assumes features to be independent which is not the case in case of space coordinates.
- PCA in Adience dataset gives good classification accuracy - since they show good features of gender in eigen

faces.

- We used VGG16 pre-trained model, to see the classification accuracy from neural network model. The architecture (16 hidden layers) allows even moderately powerful systems to run such a network. Due to this advantage, we preferred the use of this model for extracting features from the training images. This was achieved by taking fc7 layer output which has dimension of 4096.

IX. Individual Contributions

Together we analysed the dataset, planned the work flow of the project and collectively did data visualization.

- Suraj Pandey implemented the pretrained models. Implemented some models with different combination of potential features.
- Udit Pant implemented the baseline models and used the features like HOG, LBP, etc. Implemented some models with different combination of potential features.
- Tushar Agarwal implemented baseline models with PCA and LDA. Implemented some models with different combination of potential features.

We all discussed the methodology and reasoning that why any procedure is not giving better accuracy or vice-versa. The report and poster was jointly made.

X. Conclusion

We implemented two separate models for FER dataset and Adience dataset. Since images contained in FER dataset were decent frontal face portraits, facial landmarks could be extracted which helps perform better classification. However, Adience dataset, which contains images in the wild, landmarks extraction was not possible. Instead, we used VGG16 to extract image features combined with HOG features on which PCA was applied separately. On trying various models, we found Logistic Regression to be the best working model for both the datasets.

References

[1] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[3] A. Savoiu and J. Wong, "Recognizing facial expressions using deep learning."

[4] W. Li, M. Li, Z. Su, and Z. Zhu, "A deep-learning approach to facial expression recognition with candid images," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 2015, pp. 279–282.

[5] I. Goodfellow, D. Erhan, P.-L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," 2013. [Online]. Available: http://arxiv.org/abs/1307.0414

[6] ——, "Challenges in representation learning: A report on three machine learning contests," 2013. [Online]. Available: http://arxiv.org/abs/1307.0414

[7] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4295–4304.

[8] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.