On Learning Deep Models with Imbalanced Data Distribution

Puspita Majumdar,[1] Richa Singh,[2], Mayank Vatsa[2]
[1]IIIT-Delhi, India,
[2]IIT Jodhpur, India
pushpitam@iiitd.ac.in, {richa, mvatsa}@iitj.ac.in

Abstract

The availability of large training data has led to the development of sophisticated deep learning algorithms to achieve state-of-the-art performance on various tasks and several applications have been benefited immensely. Despite the unparalleled success, the performance of deep learning algorithms depends significantly on the training data distribution. An imbalance in training data distribution affects the performance of deep models. Our research focuses on designing and developing solutions for different real-world problems, specifically related to facial analytic tasks, with imbalanced data distribution. These problems include injured face recognition, fake image detection, and estimation and mitigation of bias in model prediction.

Introduction

Machine learning and deep learning algorithms have achieved tremendous success, and significant advancements are made in various applications such as face analytic tasks, including face recognition and attribute prediction. The widespread use of deep neural networks and open-source deep learning libraries have enabled to build deep models that achieve state-of-the-art performance in different applications. The performance of these algorithms depends on the training data distribution and an imbalance leads to degradation in model performance. Imbalanced data distribution, i.e., under-representation of some classes or subclasses within a class and over-representation of others, is a common problem in machine learning and deep learning. This results in deep models that have low performance, specifically for the under-represented class or subclass. Imbalance in training data could be in a (i) subclass within a class or (ii) in a particular class (or set of classes). Real-world problems such as recognition of injured victims is an example, where the training data has an imbalance in the subclass within a class. For recognition of an injured victim, both injured and non-injured samples of the victim are required. Generally, the number of injured samples is smaller than non-injured samples, which in turn leads to an imbalance in the subclasses within a class. On the other hand, real world classification problems such as fake/altered image detection (Rossler et al. 2019; Seibold et al. 2020) have an imbalanced class distribution, where altered images constitute a small portion of the training data compared to the original images. Recently reported bias in model predictions is also an example of imbalanced class distribution problem, where the performance of the models is biased towards the over-represented class (Gong, Liu, and Jain 2020; Drozdowski et al. 2020).

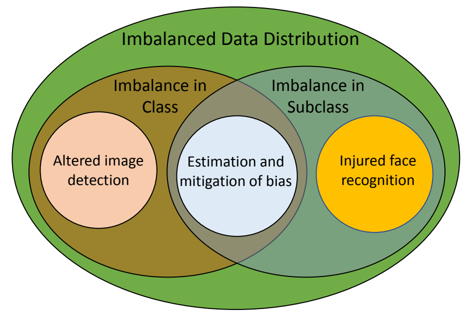

Figure 1: This research focuses on real-world problems with imbalanced data distribution on real world face analysis problems.

In our research, as shown in Figure 1, we focus on developing algorithms for different real-world problems with imbalanced data distribution. We have introduced the problem of injured face recognition, created novel databases, and provided initial solutions for the same. Next, we have developed a sophisticated algorithm to enhance the performance of models in the presence of facial injuries. Along with this, we have proposed a novel framework for handling the challenges of digital alterations such as morphing and detecting fake images. Currently, we are developing algorithms for handling the problem of bias in model prediction. In this direction, we are working on developing a solution to reduce bias and improve the model performance for the facial attribute prediction task. We are also analyzing the effect of image distortions on biased predictions of deep models.

Detection and Recognition of Injured Facial Images

Facial injuries change the facial structure and appearance due to swelling, bruises, blood clots, and accidental cuts. For the first time, we have proposed a novel database, termed as the **Domestic Violence Face (DVF)** database of 450 subjects, with 150 subjects belonging to domestic violence class and 300 belonging to non-domestic violence class. The performance of existing face recognition systems is analyzed towards detecting domestic abuse via faces. Further, we have proposed a novel framework using activation maps of deep learning features for detecting the victims of domestic violence. The proposed framework achieves a detection accuracy of 80% at Equal Error Rate (EER).

For the recognition of injured faces, we have proposed a novel loss function, termed as **Subclass Contrastive Loss (SCL)**. The proposed loss function minimizes the distance between subclasses of the same subject and maximizes the distance from subclasses of others, where the two subclasses refer to the set of non-injured images and set of injured images. To the best of our knowledge, this is the first work towards injured face recognition. Also, a novel **Injured Face (IF)** database of 100 subjects with injured and non-injured face images is created for evaluation. The performance of existing face recognition systems is analyzed, and the proposed loss function is compared with state-of-the-art algorithms. The proposed approach gives 36.70% Rank-1 identification accuracy, which is 11.27% better than the best performing baseline algorithm.

To further enhance the performance of deep models for injured face recognition, a novel loss function, **Subclass Injured Face Identification (SCIFI)** loss is proposed. The proposed loss function optimizes subclass space (introduced in this research), which in turn optimizes the feature space. Subclass space is a 2-dimensional space of scores among the subclasses (injured and non-injured subclasses). The aim is to achieve an optimized feature space where the interclass separability is maximized while maintaining approximately equal distance between the feature representations of the samples of different subjects as well as the same distance between the samples of the same subject. Extensive sets of experiments are performed to showcase the efficiency of the proposed SCIFI loss. For this purpose, comparison is performed with 12 state-of-the-art loss functions, including SCL, with UMAP and tSNE visualizations to interpret the results, ablation study, and choice of hyper-parameters.

Altered Image Detection

Images are digitally altered using different techniques such as morphing, retouching, and using Generative Adversarial Networks (GANs). In morphing, a new face image is generated digitally using the information available from multiple source face images of different subjects to elude their own identity or gain the identity of others. We have exploited the vulnerabilities of deep face recognition systems towards partial replacement and morphing of facial regions. For this purpose, 8136 digitally altered samples are generated, and the verification performance of deep face recognition systems are evaluated on the altered and original samples. Further, a **Partial Face Tampering Detection (PFTD)** network is proposed for the detection of partially altered images. The network captures the inconsistencies among the original and partially altered images by combining the raw and high-frequency information of the input images. The proposed network surpasses the performance of the existing baseline deep neural networks for altered image detection.

With the advent of easy-to-use image editing tools and GANs, retouching and generation of digital images have become an easy task. The adverse effect of retouched images in the biometric identification process and misuse of GANs generated images in *Deep-nude* and *fake news* demands an automatic system for the detection and classification of these images. Therefore, we proposed **Digital Alteration Detection using Hierarchical Convolutional Neural Network (DAD-HCNN)** for detecting retouched and GANs generated images at different levels of the framework. The proposed framework achieves more than 99% accuracy for altered image detection.

Estimation and Mitigation of Bias in Model Prediction

One of the major concerns of imbalanced training data distribution is the biased prediction of deep models, where the models favor the over-represented class(es). Estimation and mitigation of bias are important for fair and unbiased model predictions. It is important to understand the bias in model prediction for designing algorithms for its mitigation. Therefore, we are exploring different factors that could lead to biased predictions. Currently, we are working on proposing a metric to jointly measure bias in model prediction and the overall model performance. Additionally, we aim to use this metric in the bias mitigation algorithm for performance enhancement (Majumdar et al. 2020).

Acknowledgments

Puspita Majumdar is partly supported by DST INSPIRE PhD Fellowship.

References

Drozdowski, P.; Rathgeb, C.; Dantcheva, A.; Damer, N.; and Busch, C. 2020. Demographic bias in biometrics: A survey on an emerging challenge. *IEEE Transactions on Technology and Society* .

Gong, S.; Liu, X.; and Jain, A. K. 2020. Jointly debiasing face recognition and demographic attribute estimation. In *European Conference on Computer Vision*, 330–347. Springer.

Majumdar, P.; Chhabra, S.; Singh, R.; and Vatsa, R. 2020. Subgroup Invariant Perturbation for Unbiased Pre-trained Model Prediction. *Frontiers in Big Data* doi:10.3389/fdata.2020.590296.

Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; and Nießner, M. 2019. Faceforensics++: Learning to detect manipulated facial images. In *IEEE International Conference on Computer Vision*, 1–11.

Seibold, C.; Samek, W.; Hilsmann, A.; and Eisert, P. 2020. Accurate and robust neural networks for face morphing attack detection. *Journal of Information Security and Applications* 53: 102526.