CS 475/675 Project Proposal

Chang Yan, Jingguo Liang
cyan13, jliang35

Abstract

Face detection and recognition is a popular and useful technic in many areas today, like automated tagging of photo, real-time security cameras, etc. In this project, we will develop deep convolutional networks and ensemble models, combined with other optimizations like feature engineering and distortion reduction, aiming to make predictions on the age, gender and ethnicity of a person based on an image of his face.

1 Project choice

Choose either a **methods** or **applications** project, and a subarea from the below table.

☒ **Applications**				
☐ Genomics data	☐ Healthcare data	☐ Text data	☒ Image data	☐ Finance data
☐ **Methods**				
☐ Fairness in ML	☐ Interpretable ML	☐ Graphical Models	☐ Robust ML	☐ Privacy in ML

2 Introduction

Face detection and recognition has been used nowadays in many areas, ranging from entertainment to security. In this project, we will develop a neural network that can make certain predictions given an image of human face. The input for the network will be face of a human. The single human face should be the dominant element in the image. The output will be predictions made to the person based on the image. Possible predictions include age, gender, and race.

3 Dataset and Features

We currently plan to use UTKFace dataset on the training of the network. A link to the dataset is included in reference part.[1] Each entry of data consists of the following components: the image; four labels: age, gender, race, and date of collection of the image; landmarks on the face. We will not use the date as it is irrelavent to our task. Image are in RGB form with varying sizes. We will annotate the data by giving the position of the square where the face in the image resides in. We will then crop and transform the image into a fixed size, with the face being the dominant element on the image. We will normalize the brightness of the image such that the impact from lighting in the surrounding environment will be eliminated.The raw features from the data would be the RGB values of each pixel, and we also plan to add feature engineering like edge detection as new features.

4 Methods

The hypothesis class would be the deep convolutional networks that we designed or modified from other models, and we also plan to use an ensemble with multiple different network structures to boost the performance. A 5-fold cross validation will be used to evaluate the performance, and we will benchmark using a logistics regression model. We would use cross entropy loss as our first try, and we plan to try others like a MSE loss or combination of cross entropy and MSE. Our hypothesis is that it would be better to use MSE to predict age (as age would be a real-number), and cross entropy loss would be better for predicting sex and ethnicity as classes. The simple optimization would be hyperparameter tuning and changing of the nextwork structurs, and we also plan to use ensemble of multiple networks and/or attention models which focus on specific local structures[2], nested structures which samples each layer of the network[3], semi-supervised learning, etc. to better optimize it. Moreover, we would try feature engineering: The raw input has 3 channels, RGB,

and we plan to add more channels like intensity, edges, etc. to improve the performance. Also, as the dataset provides key points, we could potentially first register all the image to the same space using the key points, creating transformations of each image to a standard space, and then input the transformed image (which are in the same sapce now) to the networks. This would account for the projective deformation in the images and produce better performance[4].

5 Deliverables

These are ordered by how important they are to the project and how thoroughly you have thought them through. You should be confident that your "must accomplish" deliverables are achievable; one or two should be completed by the time you turn in your Nov 19 progress report.

5.1 Must accomplish

1. A program with a trained deep convolutional network with optimized structure and best tuned parameters, that can predict the age of the face in the input image
2. The prediction of the program should be distinctly better than simple benchmark run, like a logistic regression.
3. The program should use appropriate data preprocessing mothods to improve the performance.

5.2 Expect to accomplish

1. The program should be able to predict not only age but also sex and ethnicity of the face.
2. The program should use multiple different network structures, ensembles or other methods to further improve the performance.
3. The prediction of the program should reach an accuracy that is usable in real-life, like over 90%,.
4. Adding feature extraction methods before the network (like edge detection) to create more features from the image.

5.3 Would like to accomplish

1. The program should be able to first identify the location of the face in the image, and then predict using the cropped face area. The position of the face should also be an output.
2. The program should be able to take in different sizes of image files.
3. The program should be able to identify an image with no face, instead of giving random output.
4. The program should work with both RGB and grayscale images, and uses the RGB channels to achieve better performance than grayscale.
5. The speed of the prediction should be fast enough to be done in real-time.

References

This section should include citations for: (1) Any papers on related work mentioned in the introduction. (2) Papers describing methods that you used which were not covered in class. (3) Code or libraries you downloaded and used.

[1] UTKFace dataset https://susanqq.github.io/UTKFace/

[2] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham.

[3] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[4] We learned this from CS 601.461 Computer Vision class