

Marc Minnee

AUGMENTED FACE RECOGNITION

Capstone Proposal
Udacity nanodegree Machine Learning Engineer



Domain

BACKGROUND

- Face recognition: the good, the bad and the ugly
- The good: authentication
- The bad: state surveillance
- The ugly: false positives
- Enhance face recognition models to prevent things going south

Image and face recognition is getting a lot of attention in AI and machine learning. Applications can be used for both good and bad purposes.

The good: easy and swift image recognition solves the problem of authentication in ever expanding granular digital authentication workflows. No more password hassle.

The bad: state surveillance threatens human rights when people are constantly being watched and checked upon.

The ugly: predictive policing with image recognition shows too many false positives based on biased (racial) inputs and poor performing models, not recognising subtle differences in facial attributes from human races.

The last one is especially interesting to solve. We should be able to enhance models by reducing biased inputs and account for differences in human race. In order to prevent the ugly becoming the bad...



The Bad

<https://bitterwinter.org/chinas-high-tech-surveillance-state-a-digital-despotism/>



The Ugly

<https://www.theguardian.com/technology/2016/feb/04/us-police-data-analytics-smart-cities-crime-likelihood-fresno-chicago-heat-list>

Problem

STATEMENT

- Transparency in model use is uncommon
- Biased inputs by historical data

If we'd like to combat the ugly before getting worse, we need to understand why modelling predictive policing is controversial.

First of all, used models are opaque in the sense that authorities don't want to open source these in order to outsmart criminals. This way we cannot objectively determine the performance and evaluate the inputs used to train the model.

And if we are to investigate the inputs, we likely find datasets, based on historical data of imprisoned persons which are biased towards Afro-Americans and Latinos.

So, the problem has its origin in biased datasets, training models with more images of a certain kind, also described by Cathy Neil.

If we can fix this, we can at least account for unbiased input when training our models.

Datasets &

INPUTS

- Dog breed classification for human race recognition?
- CelebA dataset, Gender Shades dataset
- Using a GAN to augment dataset with "race" label
- Train a model for augmented face recognition

We start out with the capstone project of dog breed classification. It would be interesting to see how well this model performs on a dataset of human faces, what kind of 'breed' the outcome would be. But that's just for fun. Of course we need to train a comparable model with inputs of human faces, preferably of different human race.

As a base input we use a processed CelebA dataset from Kaggle, with 202,599 number of face images of various celebrities. However we can extract many facial features, the label we are interested in is race or skin type. We need to infer this label from the other features in our dataset, by labeling a sample of the images by hand and then use a generative modelling technique, a GAN, to get more training data with race or skin type labels. It might be better to use the Gender Shades dataset, which already accounts for different skin types.

Mixing these datasets, we can train a model which is better capable of recognising differences in gender and race, as a means to:

1. Show more transparency in model use
2. Account for (un)biased input
3. Perform better at recognising faces of different races


Data Sources

list_attr_celeba.csv	41 columns
list_bbox_celeba.csv	5 columns
list_eval_partition.csv	2 columns
list_landmarks_align_celeba.csv	11 columns
img_align_celeba	
img_align_celeba	
000001.jpg	
000002.jpg	
000003.jpg	

About this file

No description yet

000001.jpg (11.17 KB)



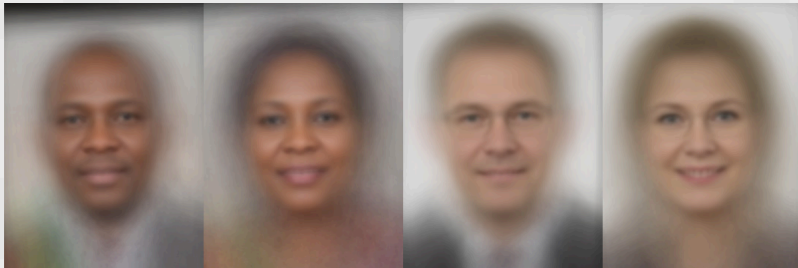
Kaggle sample

<https://www.kaggle.com/jessicali9530/celeba-dataset#000001.jpg>


es

Gender Shades project
audits the accuracy of AI
gender
classification products.

valuation focuses on gender
 classification as a motivating
 example to show the need for
 increased transparency in the
 performance of any AI products
 or services that focused on
 human subjects. Bias in this
 context is defined as having
 systematic differences in gender
 classification error rates between
 groups.



Gender Shades



Gender Shades project

<http://gendershades.org/overview.html>

Solution

STATEMENT

- Bring more transparency in model use
- Transparency in Inputs: balanced dataset
- Transparency in Outputs: enhanced model performance
- Better performance of face recognition as a whole

So it would be interesting to know if we could model face recognition with face features corresponding to gender and race or skin type labeling, as a means to recognise different gender and races in images in order to augment:

1. Model inputs: is the model fed with a balanced set of images of different gender and races?
2. Model outputs: are we able to enhance model performance given the (unbiased) features of the input dataset?
3. Performance of face recognition

This model should at least perform better at recognising different race in images, as an intermediate feature to be used in unbiased predictive policing and to help audits performed on these models. A tool to counterbalance state surveillance.

Benchmark

MODEL

- Best benchmark model: Gender Shades project
- Alternative: NIST report

As a benchmark we use the recent Gender Shades Pilot Parliament project, which is best related to our project and has metrics already evaluated and explained.

Second best is the recent NIST report, but datasets are quite dated and biased in the samples being used.

Evaluation

METRICS

- Copy benchmark metrics
- Evaluate skin type recognition

We copy the evaluation matrix of the benchmark model, which means we are trying to minimize false positives and negatives for subgroups. Gender classification performance is measured by the positive predictive value (PPV), error rate (1-PPV), true positive rate (TPR), and false positive rate (FPR).

On top of what is evaluated in the benchmark model, we might also do the same for race attributes, albeit we must then label the dataset accordingly with skin types.

Project

DESIGN

- Investigate datasets
- Augment datasets with GAN
- Train a PyTorch network
- Tune and resample
- Develop an API
- Publish on Medium

The workflow needed for this project:

1. Investigate the Gender Shades and Kaggle dataset for gender and skin type labeling
2. Augment both datasets with a GAN to get more samples
3. For each dataset:
 - shuffle the data and split in test and training
 - feed it into a PyTorch network, captured from the dog breed classification project
 - Evaluate against the benchmark model
4. Tune the parameters of the network to increase the performance
5. Go back to step 2 and 3 to test resampling
6. When performance is OK, develop an API for external evaluation purposes
7. Publish an article on Medium