Deep Vision

Project Proposal: "Title yet to come"

Elias Arnold, Moritz Althaus

June 7, 2019

1. Team

We are not able to outline, who is going to be responsible for which part. The whole point of working in a team is to work together on all aspects of the problem. We don't consider the process of splitting a task in different sub-tasks and solving them individually as teamwork.
If - for grading purposes - it must be outlined who worked on which part, we don't mind any random choice. It should be noted here that the grade of the module does not count for both of us.

2. Problem Definition

What features make a human look human alike? What are the basis vectors of the feature space of faces? We want to see if we can train a model to produce valid human faces from a small feature space. For us this is interesting because this is an area where machine learning clearly has an edge. We can't think of a good way to write a hard coded algorithm to sample images. Also, age is interesting for us. What kind of features make a person look old, what kind of features make a person look young? We want to study the connection between the age and the look of a person.
We find this interesting because in the age of automated face recognition, obtaining the age of a person can be an important factor. The prediction of the age of a person is also closely related to the prediction of other biometrics and facial attributes such as gender, ethnicity, hair color and expressions. If we are able to sample valid faces, perhaps we are able to do some kind of discrimination in the feature space to alter the age of the generated faces.

3. Dataset

3.1 Training Data

There are several public datasets of face images with age labels available. One of the largest is the *Cross-Age Celebrity Dataset (CACD)* [1] which contains 163,446 images from 2,000 celebrities with age ranging from 16 to 62. These images were collected from different search engines where the names of celebrities and a year between 2004 and 2013 were used as keywords. Due to the big ratio of the number of images and the number of celebrities we hope that this dataset provides a great variance in the age of a single person. Another possible dataset is the *IMDb dataset*. It is a huge dataset of images with gender and age labels for training. It contains 460,723 images of the most 20,284 popular actors from IMDb (Internet Movie Database). There is also a version in which the images are cropped down to faces, this versions is about the half of the original dataset in size. This is because the whole dataset contains many images in which it is hard to detect a face. We believe that both datasets together are a quite good basis for training.

3.2 Test Data

Both previous mentioned datasets can also provide a subset of images for testing purposes. Nontheless, we take another small dataset *FG-NET* [2] into consideration. This dataset contains 1002 images of 82 persons with age ranging from 0 to 69.

3.3 Problems with the Dataset

No Dataset is perfect. Because the images are obtained online and from search engines, there is no proof that the obtained information is accurate. We expect noise. Also, because many images are obtained from movies which were produced over an extended time, there is some introduced inaccuracy in the age of people.

3.4 Download

The IMDb dataset is approximately 7 GB in size, the CACD dataset about 4 GB. The metadata of both datasets are available in the MATLAB format, the images are plain png files. We don't expect any difficulty using them in python.

4. Approach

4.1 Variational Autoencoder

First we want to use a Variational Autoencoder (VAE) to represent peoples faces in latent space and then recover the faces. [3] did some interesting work on this topic. If the model converges and if we achieve a good representation with the Autoencoders in latent space, we want to have a look at the main features that are used to represent faces and how much information is needed to represent a face. This can be done by varying the size of the latent space and having a look at the quality of the generated images. Also, sampling from the latent space and interpolating in this latent space are possibilities we can explore.

4.2 Discrimination in Latent Space

Then, it will be interesting to see if there can be any kind of discrimination in latent space to make a person look older or younger. Perhaps this leads us to the point where we could alter the age of a specific person. It might also be possible to deploy a further discriminator on the image space to enhance the quality of the output image, this would be beyond the scope though.

5. Evaluation and Expected Results

5.1 Variational Autoencoder

We have described different subtasks here. It is quite hard in advance to predict the difficulties and challenges of these different subtasks. Getting the Autoencoder to converge will be a challenge but we a fairly certain that this is possible. However, we don't expect the recovered images to be of the same quality as images sampled by GANs.

5.2 Sampling and Interpolating

We are not quite sure if sampling and interpolating from the latent space will be easy to do. If the VAE achieves a good representation, this should in theory be possible. A look at the sampled and interpolated images will be an interesting evaluation.

5.3 Age progression

We are not sure if we can achieve good results here. This is because the problem is quite an extreme challenge. Discriminating in latent space doesn't sound too easy. Also for altering the age of a specific person, most of the existing work requires face images of the same person at different ages over a long span to predict future development.

6. Hardware

Hardware for setting up the models and early stage testing are two Apple Laptops. However, we don't expect the CPUs to be powerfull enough for bigger models to converge. Therefore, we would like to use HCI GPUs for later stage training. Another possibility is to use GPUs provided by AWS.

7. Uniqueness

Although the research in the field of face recognition and face generation has been going on for decades it is still a tough call to alter the age of a face in a photo-realistic manner. Of course, there are several papers tackling this problem and some of them achieve quite good results already, though the results are far from perfect. We expect by no means to yield better result but we rather see this project as a chance to approach an interesting and current field of research with slightly different methods and datasets. The most common methods used in this context are different kinds of GANs, whereas we try to deploy a VAE and some kind of discriminator. Even though we approach this topic with methodically rather blunt weapons it makes sense for two reasons:

1. It is good to learn how to implement different architectures and to learn their difficulties as well as their limitations.
2. This work can lay the basis for more advanced research in this context.

References

[1] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.

[2] Y. Fu, T. M. Hospedales, T. Xiang, Y. Yao, and S. Gong, "Interestingness prediction by robust learning to rank," in *ECCV*, 2014.

[3] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5810–5818, 2017.