Recognition of features of faces on CelebA database with Variational Autoencoders

(Az arckép jellemzőinek a felismerése CelebA adatbázison variációs autoenkóderek segítségével)

Abstract

After a short discovery of the CelebA database, a well-known discriminative model was fit to learn the labels of the CelebA database: a pretrained Inception V3 [7] was used with transfer learning with a proper validation accuracy. A massive data augmentation was used to boost the performance.

Then, a generative model, a variational autoencoder was trained to learn the features of the faces. Here, a so-called latent traversal was performed to visualize the meaning of the dimensions. Fortunately, many of the 64 dimensions seemed to be meaningful and could learn some features of the faces (i.e.: color of hair, existence of bangs, color of the skin, etc), however the disentanglement requirement could not be achieved with these computational resources.

Introduction

The features of images in a labelled dataset can be learned by using discriminative models with convolutional neural networks, while in case of unlabelled image set several unsupervised techniques are proposed. Although with GANs ([1]) have great advantages in generating new images and find implicit latents, working with Variational Autoencoders was preferred as (i) inference and the regularisation of the latent space can be directly controlled (ii) posterior inference (which might be intractable) can be made efficient by fitting an approximate inference.

Here a conditional variational autoencoder was used (hierarchical VAE was not attempted finally.)

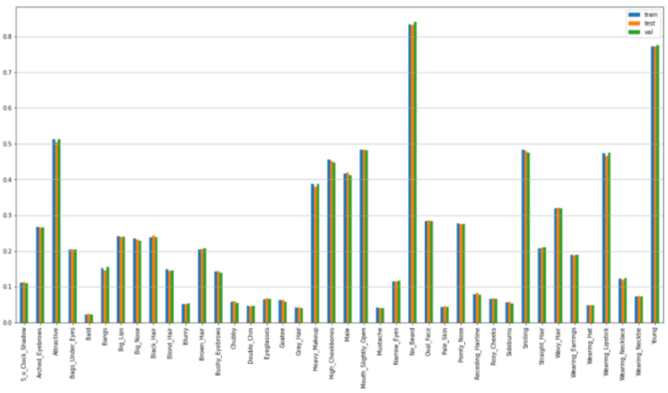
Steps

Discovery of the dataset and determine the training / test / validation dataset

The balance of the dataset is proven to be essential as shown in the following documents [8-9]

I.e. generally, the distribution of the training data has a huge impact on the performance of CNN. The balanced distribution yielded a significantly better performance than imbalanced one. The heavier the imbalance is, the worse the total classification performance. This kind of fragility when using imbalanced data in the CNN training algorithm can be eliminated by the proper selection of distribution of dataset.

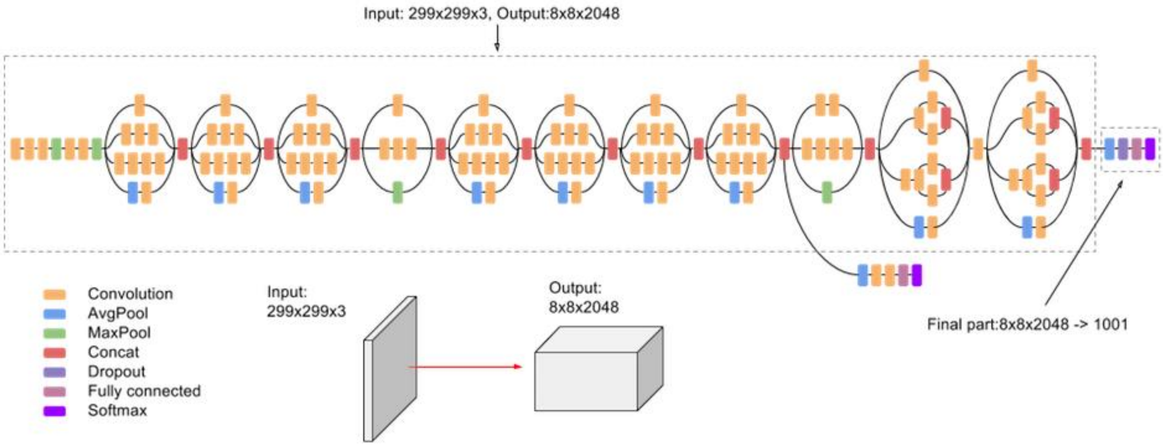
A splitting function and a check for distribution of data was performed



We investigated some solutions for dataset generation and preprocession. [A,B,C,D]

To discover the relationship between the attributes we plotted the correlation matrix between the values.

Then a pretrained network an Inception V3 was trained to learn the labels of the dataset.



Note, that those team members who would have been responsible for training hyperparameter tuning and visualization quitted the from the group.

Short summary of variational autoencoders

To observe the intrinsic structure of a dataset auto-encoders can be very efficient by finding a right pair of encoders and decoders keeping the maximum of information when encoding and the minimum reconstruction error when decoding. However, the lack of interpretable structures in the latent space (a.k.a lack of regularity of the latent space) and overfitting (because of the reconstruction component in the loss function) can result in severe inadequacy. I.e. neither : continuity (two close points in the latent space should not give two completely different contents once decoded) nor completeness (for a chosen distribution, a point sampled from the latent space should give “meaningful” content once decoded) can be guaranteed.

In variational autoencoder [1,2] we prevent the model to encode data far apart in the latent space and encourage as much as possible returned distributions to “overlap” with regularisation term, satisfying this way the expected continuity and completeness conditions.

Find the appropriate latents

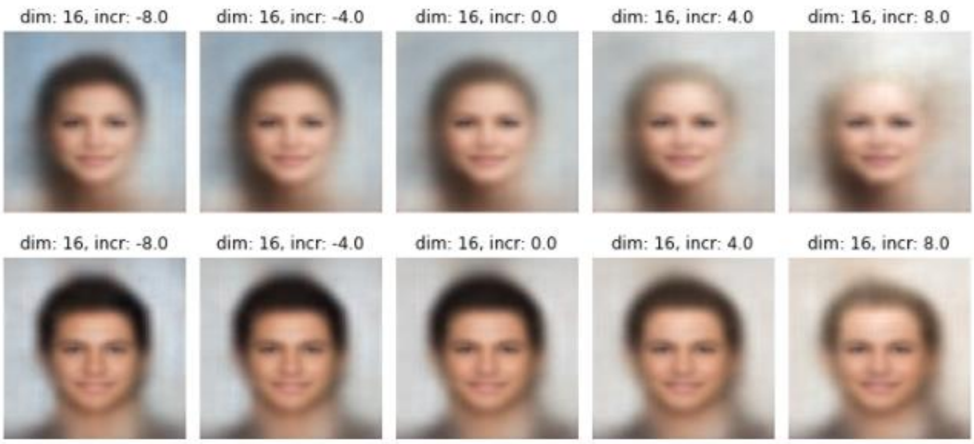
The great novelty of Variational Autoencoders was that they combined probabilistic models, i.e. variational Bayesian approach with deep learning techniques so that to perform efficient inference and continuous latent variables even in case of intractable posterior distributions, and large datasets.

With a reparameterization of the variational lower bound, a differentiable unbiased estimator of the lower bound can be achieved: the SGVB (Stochastic Gradient Variational Bayes) estimator can be used for efficient approximate posterior inference.

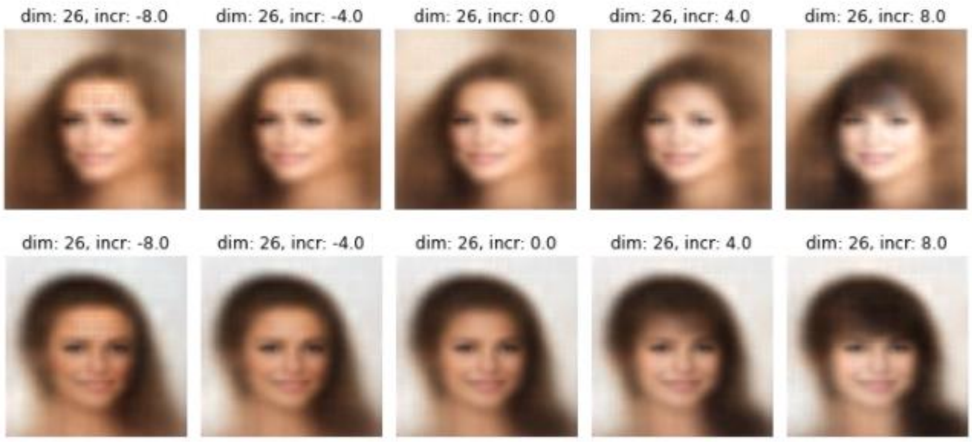
Results

Many of the latents were meaningful. Note that according to the random seed and from different network size (parameter k) and batchsize (parameter b) different results were found.

Color of the hair at dimension 16th



Existence of bangs at dimension 26th



The rotation of face at dimension 2nd dimension

The thickness of hair at dimension 9th dimension

The width of the face at 12th dimension

Future planes

Hierarchical variational autencoders should be trained to achieve a more accureate solution (for the details see Future Planes documentation)

References of literature

[1.] Diederik P Kingma and Max Welling.
Auto-encoding variational bayes.
In The International Conference on Learning Representations (ICLR), 2014.
https://arxiv.org/pdf/1312.6114.pdf

[2.] Diederik P Kingma and Max Welling.
An introduction to variational autoencoders
Foundations and Trends® in Machine Learning, 2019.
https://arxiv.org/pdf/1906.02691.pdf

[3.] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, Alexander Lerchner
ICLR 2017 conference submission, 2017.
Beta-vae: Learning basic visual concepts with a constrained variational framework
https://openreview.net/forum?id=Sy2fzU9gl

[4.] Y. Bengio, A. Courville, and P. Vincent.
Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013

[5.] Arash Vahdat, Jan Kautz
NVAE: A Deep Hierarchical Variational Autoencoder
https://arxiv.org/pdf/2007.03898.pdf

[6.] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen Progressive Growing of GANs for Improved Quality, Stability, and Variation
https://arxiv.org/pdf/1710.10196v3.pdf

[7.] SZEGEDY, Christian, et al. Going deeper with convolutions.

In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. p. 1-9.
Going Deeper With Convolutions (googleusercontent.com)

Some essays, tutorials, codebases

[8]. Paulina Hensman, David Masko

The Impact of Imbalanced Training Data for Convolutional Neural Networks

DEGREE PROJECT, IN COMPUTER SCIENCE

https://www.kth.se/social/files/588617ebf2765401cfcc478c/PHensmanDMasko_dkand15.pdf

[9] G. Weiss, Foster Provost,

Learning when training data are costly: the effect of class distribution on tree induction

Journal of Artificial Intelligence ResearchVol. 19, No. 1

https://dl.acm.org/doi/10.5555/1622434.1622445

References of used code bases, tutorials:

[A]https://www.kaggle.com/ky2019/starter-celebfaces-attributes-celeba-b5421ae1-e

[B] https://www.kaggle.com/saket0565/celebfaces-facial-attribute-recognition

[C] https://www.kaggle.com/bmarcos/image-recognition-gender-detection-inceptionv3

[D] https://www.kaggle.com/fkdplc/celeba-dcgan-for-generating-faces/notebook

[E] Conditional VAE on Faces | Kaggle

[F] VAE CelebFace | Kaggle