Universidade Federal da Bahia
Instituto de Matemática e Estatística

Programa de Graduação em Ciência da Computação

REMOVING EYEGLASSES THROUGH ADVERSARIAL TRAINING

Gabriel Lecomte Pinho e Souza

TRABALHO DE GRADUAÇÃO

Salvador
14 de Dezembro de 2018

GABRIEL LECOMTE PINHO E SOUZA

REMOVING EYEGLASSES THROUGH ADVERSARIAL TRAINING

Este Trabalho de Graduação foi apresentado ao Programa de Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Maurício Pamplona Segundo

Salvador
14 de Dezembro de 2018

Sistema de Bibliotecas - UFBA

LECOMTE PINHO E SOUZA, Gabriel.
Removing eyeglasses through adversarial training / Gabriel Lecomte Pinho e Souza – Salvador, 2018.
18p.: il.

Orientador: Prof. Dr. Prof. Dr. Maurício Pamplona Segundo.
Monografia (Graduação) – Universidade Federal da Bahia, Instituto de Matemática e Estatística, 2018.

1. GAN. 2. Deep learning 3. Aprendizado não-supervisionado.. I. PAMPLONA SEGUNDO, Maurício. II. Universidade Federal da Bahia. Instituto de Matemática e Estatística. III Título.

CDD – XXX.XX
CDU – XXX.XX.XXX

TERMO DE APROVAÇÃO

GABRIEL LECOMTE PINHO E SOUZA

REMOVING EYEGLASSES THROUGH ADVERSARIAL TRAINING

Este Trabalho de Graduação foi julgado adequado à obtenção do título de Bacharel em Ciência da Computação e aprovado em sua forma final pelo Programa de Graduação em Ciência da Computação da Universidade Federal da Bahia.

Salvador, 14 de Dezembro de 2018

Maurício Pamplona Segundo, PhD
Universidade Federal da Bahia

André Brasil Vieira Wyzykowski, MsC
Universidade Federal da Bahia

Leone da Silva de Jesus, MsC
Universidade Federal da Bahia

RESUMO

Óculos tem um impacto conhecido em reconhecimento facial automatizado. Até as abordagens mais recentes baseadas em *deep learning* mostram uma queda na acurácia quando óculos são introduzidos no cenário. O uso de óculos também pode ser um problema mesmo quando o reconhecimento facial é realizado por humanos. Em situações onde a pessoa não conhece o indivíduo a ser reconhecido, como no caso de um criminoso procurado pela polícia, a presença de óculos torna essa tarefa ainda mais difícil.

Por isso, nós propomos uma abordagem não-supervisionada baseada em SimGAN para remover óculos de fotos. A solução consiste em duas redes distintas: o refinador e o discriminador. O refinador consiste em uma rede residual que recebe uma imagem e retorna a mesma imagem sem óculos. O discriminator é uma rede completamente convolucional que deve diferenciar as imagens reais sem óculos das imagens refinadas. Nós modificamos a *loss* de regularização, introduzindo uma matriz de pesos e o erro quadrático médio como um substituto da distância L1.

Nosso método é avaliado quantitativamente e qualitativamente. Apesar de ocorrerem artefatos, apresenta resultados promissores, especialmente para fotos de pessoas usando óculos de sol.

Palavras-chave: GAN, *deep learning*, aprendizado não-supervisionado.

ABSTRACT

Eyeglasses have a known impact on automated facial recognition. Even the most recent deep learning approaches show a decrease in accuracy when glasses are accounted for. Furthermore, glasses can also be a problem if the recognition is performed by humans. In situations where the person is not familiar with the subject to be recognized, such as a wanted criminal, the addition of eyeglasses can make the task even harder.

Therefore, we propose an unsupervised adversarial approach based on SimGAN to remove eyeglasses from pictures. The solution consists in two different networks: the refiner and the discriminator. The refiner is a Residual Network that receives an image and outputs the same image without glasses. The discriminator is a fully convolutional network that has to discern between real faces without glasses and refined images. We modified SimGAN's regularization loss, introducing a weight matrix and mean squared error as a replacement for the L1 distance.

Our method is evaluated quantitatively and qualitatively. Even though artifacts occur, shows promising results, specially for pictures of people wearing sunglasses.

Keywords: GAN, deep learning, unsupervised training.

CONTENTS

Chapter 1—Introduction 2

1.1 Problem . . . 3
1.2 Solution . . . 4

Chapter 2—Methodology 6

2.1 SimGAN . . . 6
2.1.1 Network Architecture . . . 7
2.1.1.1 Refiner . . . 8
2.1.1.2 Discriminator . . . 8
2.1.2 Regularization Loss . . . 8

Chapter 3—Results 10

3.1 Dataset . . . 10
3.2 Validation . . . 10
3.2.1 Qualitative . . . 11
3.2.2 Quantitative . . . 12
3.3 Related Work . . . 13

Chapter 4—Conclusion 16

4.1 Future work . . . 16

LIST OF ACRONYMS

CNN	Convolutional Neural Network	8
GAN	Generative Adversarial Network	4
ResNet	Residual Network	8
SimGAN	Simulated + Unsupervised learning GAN	4
SRCNN	Super-Resolution Convolutional Neural Network	13

Chapter 1

INTRODUCTION

Figure 1.1 The classic example of how human facial recognition can fail

Throughout the years, Superman fooled millions by posing as Clark Kent by only wearing eyeglasses, as it can be seen in Figure 1.1. Eyeglasses and other elements that disguise a person's face pose a challenge not only for automated face recognition, but also for human face recognition.

Deep learning models have been performing really well for human face recognition in multiple scenarios (SCHROFF; KALENICHENKO; PHILBIN, 2015). However, their practical applications face a few challenges, as glasses are commonly worn nowadays, either as a fashion statement or to correct shortsighted vision. These kinds of accessories that occlude the region around the eyes (or over them if you consider sunglasses), have shown to have a negative impact on the performance of such recognition models (LIANG et al., 2015) (GUO et al., 2018).

When the same task is done by humans, glasses are also a challenge, specially when the individual to be recognized is not known by the subject (RIGHI; PEISSIG; TARR, 2012). When a person is wanted by the police, for example, artifacts that occlude the face in the released pictures, as glasses, may make it difficult for the average citizen to

Figure 1.2 The pictures released by the FBI of the subjects responsible for the Boston marathon bombing.

recognize the subject. In 2013, after the Boston Marathon bombings, pictures of the suspects (Figure 1.2) were released to the public, in the hope of locating them. However, one of the suspects was wearing sunglasses and that was true challenge for his recognition.

In the following sections we discuss the problem tackled in this monograph (Section 1.1) and our proposed solution (Section 1.2).

1.1 PROBLEM

Removing glasses from pictures is not a trivial task. Several factors, such as different poses, eyeglasses frames and shades of lenses, need to be taken into account, as the removal process will be different for each of these cases.

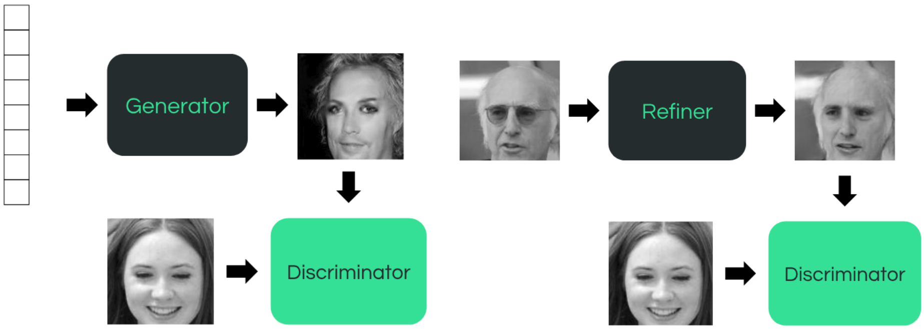
The different poses of a face may affect the removal in a couple of ways. The first one being that the steeper the angle between the camera and the glasses is, the more refraction is seen in the picture (Figure 1.3(a)), and an ideal removal solution should consider that in order to generate a realistic output respecting the subject's eye size. The second one is that the different poses may result in different reflections in the lenses and glasses' frames occluding different parts of the face (Figure 1.3(b)).

(a) (b) (c)

Figure 1.3 Examples from the CelebA face database showcasing: (a) refracion, (b) different poses and (c) dark lens shades.

As for the shades of the lenses, the darker a lens is, the more distorted the information we have on the subject's eyes is (Figure 1.3(c)). And if sunglasses start to be taken into account, the task differs and becomes more difficult, as the eyes need to be completely reconstructed because we no longer have any information on them.

1.2 SOLUTION



(a) (b)

Figure 1.4 (a) The GAN model. (b) The SimGAN model.

Our problem has several factors that need to be accounted for, such as ethnicity, gender, different poses, lighting conditions and different types of glasses. Due to all of the possible variations that emerge from pictures of faces, removing glasses using traditional algorithms can be a difficult task. Considering that we decided to approach the task of removing eyeglasses from pictures with deep learning, one problem is the absence of large databases that have pictures with glasses and their ground truth available (the same face in the pose and same environment conditions, but without glasses). Such databases would be necessary to train deep neural network models in a supervised manner. However, there are large datasets with images annotated according to the presence or absence of glasses, enabling us to work in an unsupervised way with an adversarial deep learning approach.

The Generative Adversarial Network (GAN) (GOODFELLOW et al., 2014) consists in a framework that learns how to map a random input to a given distribution through adversarial training. It proposes a minimax game between two networks: the generator and the discriminator. The discriminator is responsible for identifying if a given input is real (an image from the dataset) or fake (an image supplied by the generator), while the generator tries to fool its adversary, as it can be seen in Figure 1.4(a). The base GAN implementation, however, is not ideal for our task, as it expects to expand its given input and doesn't keep any characteristics from it, which in our case would be important to preserve the subject's facial traits. GANs also struggle with the introduction of artifacts, as the generator tends to emphasize certain features learned by the discriminator to try to fool it. That's the reason why we chose to replicate the Simulated + Unsupervised learning GAN (SimGAN) (SHRIVASTAVA et al., 2017) architecture (as seen in Figure 1.4(b)), replacing the generator with a refiner to work from existing images, and introducing a couple of mechanisms to avoid overly modifying the input and adding artifacts to the resulting images. Therefore, we propose a solution that uses adversarial training to remove eyeglasses from pictures without their ground truth.

In the following chapters we detail our solution (Chapter 2), expose our results (Chapter 3) and then explain our conclusions (Chaper 4).

Chapter 2

METHODOLOGY

In order to achieve the goal of removing eyeglasses from pictures, we considered an approach that does not require knowledge of their ground truth (the face in the same pose and scenario of the input picture but without glasses), as none of the existing datasets provide such information. Therefore, we went towards a GAN-based approach, as its primary task is to match a given distribution (in our case faces without glasses) through unsupervised learning and there is no need to know which output each input should produce.

However, the original GAN model was designed to generate images from a random signal, and not to modify an existing image. Thus, we looked for an architecture that could take an image as an input and remove eyeglasses from it while mantaining the original face in the result. Given these requirements, the SimGAN model appeared as the most adequate solution. In the following section we give a brief explanation of SimGAN and how we changed it to fit our scenario.

2.1 SIMGAN

The SimGAN architecture is an adaptation of the GAN framework. It is intended to work from existing input images, instead of a random input, to avoid some of the existing issues with GANs, such as catastrophic forgetting and the introduction of artifacts from the training dataset. These artifacts are derived from image features common to the dataset that are over-emphasized with the intent of fooling the discriminator (e.g. makeup in a database of celebrity faces).

Those issues are eased through the introduction of three new factors: a regularization loss, a history of refined images and a local adversarial loss.

The regularization loss consists of a new loss term introduced in the refiner, that is calculated as the L1 distance between the refiner's input and output images. It intends to reduce the difference between the two, thus keeping important features from the input. In this work, these features are the image background and the subject's identity.

The history of refined images consists of a small set of refined images. A small part of the history is randomly replaced after each refiner training iteration, and half of the fake samples in a discriminator's training batch comes from it. Its goal is to prevent the reintroduction of artifacts in the refined images, an issue that is often seen in GANs due to their lack of memorization capacity.

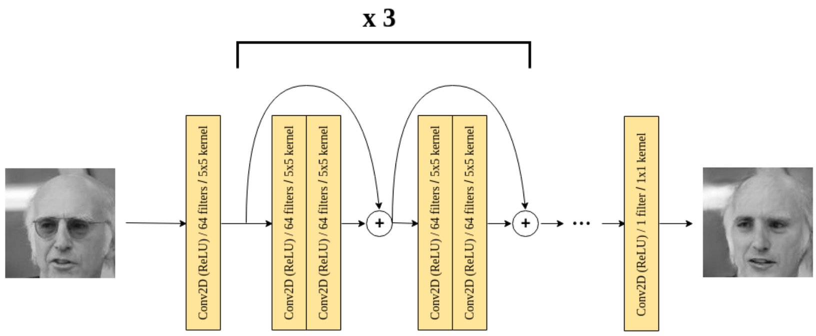
Finally, the local adversarial loss consists of replacing the discriminator with a single output by a fully convolutional one that produces a probability for different patches of the image. This also helps avoiding the introduction of artifacts in the refiner's output.

The model was originally designed to add realism to synthetic images with the intent of augmenting datasets. For that to be possible, the images need not only to be realistic, but also keep features from the simulated images to guarantee variability. In the performed experiments, the SimGAN provided great results both for gaze and hand pose estimation.

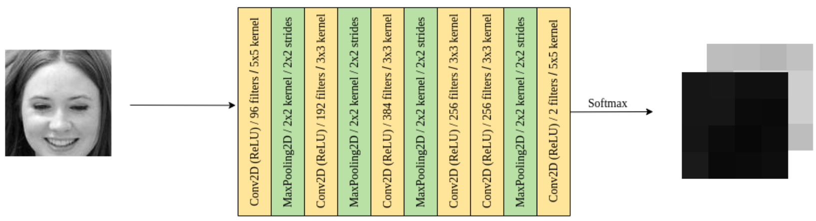
In the following sections we explain our adaptation of the SimGAN architecture (Section 2.1.1) and the changes we made to the regularization loss (Section 2.1.2).

2.1.1 Network Architecture

The SimGAN architecture consists of two intertwined networks: the refiner and the discriminator. The first one receives an image as input, and returns a refined image of the same size that preserves important characteristics (in our case, the unique traits of the face), as seen in Figure 2.1(a). The second one is responsible for discerning between real images from the aimed distribution (faces without glasses) and images that were refined to reproduce such distribution, as seen in Figure 2.1(b).



(a)



(b)

Figure 2.1 (a) The refiner network. (b) The discriminator network.

Sections 2.1.1.1 and 2.1.1.2 describe in details the architecture of each network.

2.1.1.1 Refiner For our refiner, we used an architecture similar to the architectures proposed for eye gaze estimation and hand pose estimation in the original SimGAN, which consists of a Residual Network (ResNet) (HE et al., 2016). The input first passes through a convolutional layer with 64 filters and a 5x5 kernel, then goes through six ResNet blocks, and finally through a last convolutional layer with 1 filter and 1x1 kernel. Tests were performed using the eye gaze estimation architecture, which consists of 4 ResNet blocks with a 3x3 kernel, but the results were not satisfactory. That lead to the decision of incresing the kernel and the number of ResNet blocks to 5x5 and 6, respectively.

2.1.1.2 Discriminator Our discriminator differs from the one used in the original SimGAN implementation. The architecture we used derives from a Convolutional Neural Network (CNN) (LECUN et al., 1998) designed to detect glasses (SHAO; ZHU; ZHAO, 2016) with changes in the last layers to make it fully convolutional. The network receives an image as input and returns a probability map of each image patch having glasses. Our CNN is one layer deeper than the one in the original SimGAN implementation and has more filters per layer. The complete architecture can be seen in Table 2.1.

Conv2D (ReLU)	96 filters	5x5 kernel	same padding
MaxPooling2D	2x2 pool size	2x2 strides	
Conv2D (ReLU)	192 filters	3x3 kernel	same padding
MaxPooling2D	2x2 pool size	2x2 strides	
Conv2D (ReLU)	384 filters	3x3 kernel	same padding
MaxPooling2D	2x2 pool size	2x2 strides	
Conv2D (ReLU)	256 filters	3x3 kernel	same padding
Conv2D (ReLU)	256 filters	3x3 kernel	same padding
MaxPooling2D	2x2 pool size	2x2 strides	
Conv2D	2 filters	5x5 kernel	valid padding
Softmax			

Table 2.1 The architecture of the discriminator network.

2.1.2 Regularization Loss

In the original implementation of the SimGAN, the regularization loss is the L1 distance between the original image and the refined image. However, in order to condition the refiner to prioritize modifications in the region of the glasses over the rest of the face, we introduced a weight matrix that reduces the regularization loss in the eyes region. Considering that the refiner receives an aligned face image, the introduction of the weight matrix provides better control over what is modified.

We performed experiments using either the L1 distance associated with a weight matrix or a masked mean squared error based in the content-similarity loss (BOUSMALIS et al., 2017). The masked MSE (Figure 2.2), using the weight matrix as the mask,

Figure 2.2 A visual representation of the mean squared error regularization loss. The black strip of the mask represent the matrix positions with 1×10^{-4} and the remaining white parts represent the matrix positions with 1.

provided better results with fewer modifications in the background of the refined images. Therefore, it replaces the L1 distance as the regularization loss.

Chapter 3

RESULTS

Evaluating a task so visual as the removal of eyeglasses statistically can be quite difficult, therefore we propose two different validation methods for our solution: a qualitative one and a quantitative one. In the following sections we discuss the dataset used for our experiments (Section 3.1), our validation methods (Section 3.2) and a comparison with other approaches to this problem (Section 3.3).

3.1 DATASET

In order for our method to work with a large variation of faces and poses, we need a dataset with a considerable amount of faces both with and without glasses. We also need these faces to be aligned to each other. This way, the area we have to modify will be roughly the same for all pictures. For those reasons we chose the aligned version of the CelebA database (LIU et al., 2015). It contains 13193 pictures of faces with glasses and 189407 pictures of faces without glasses, which make a total of 202600 pictures of 10177 distinct people.

To both train and validate our model we had to split the dataset in two parts: training and validation. The training set contains all faces without glasses and 11138 faces with glasses. The validation set contains the 2055 remaining faces with glasses. The faces of the validation set were picked randomly from the whole set of faces with glasses and the split is subject-independent.

3.2 VALIDATION

The first step towards validation was to pretrain our networks. The refiner was trained for 60 steps with the goal of minimizing the regularization loss. Then the discriminator was trained for 300 steps, using outputs from the refiner pretrain as fake samples, and real images from the base as real samples. Our entire network was then trained for 10000 steps. Each step consists in two training steps for the refiner and one training step for the discriminator. For both pre-training and training our method, the learning rate of the refiner was 2×10^{-4} and the one of the discriminator was 5×10^{-6}. Both the discriminator and the refiner were optimized using the stochastic Adam method (KINGMA; BA, 2014) during the whole process.

With the trained model, we ran all the images from our validation set through the refiner, in order to get the samples needed for both qualitative and quantitative evaluation. For a qualitative analysis of the proposed solution, we present some visual results in Section 3.2.1. Then, quantitative results are discussed in Section 3.2.2.

3.2.1 Qualitative

Our qualitative validation consists in an analysis of the samples generated by our trained model. In our validation set there are images with different properties, such as gender, ethnicity and types of glasses.

In the early stages of training, it was clear that the refiner was trying to imitate certain characteristics of our dataset. It is noticeable that the outputs of our refiner had some enhanced contrast in the eyes region and smoothed faces, as it can be seen in Figure 3.1. As we were working with pictures of celebrities, these traits can be easily correlated with the use of makeup by most of the subjects present in the training set.

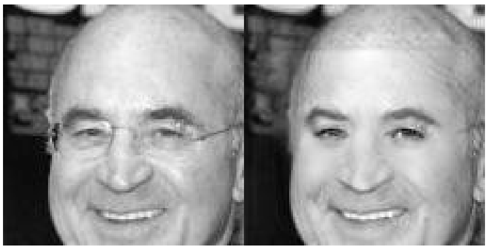


Figure 3.1 On the left the original picture of one of the subjects from the validation set and on the right the result of that picture going through the refiner.

Another interesting behavior of our model is seen when faces wearing sunglasses are used as input. In most of those pictures, the eyes of the subject are hidden due to the shade of the lenses. Then, in order to fool the discriminator, the refiner tries to recreate eyes on the subject's face, as it can be seen in Figure 3.2. Although the results are not very realistic, it is fascinating to see that our refiner had the capacity of trying to recreate human eyes. So is the fact that the discriminator was considering the presence of eyes to discern real images from fake ones.

The model also had some issues with a few different poses, resulting in remains of the glasses' frames, specially on the sides of a subject's face, as it can be seen in Figure 3.3(a). That may occur for multiple reasons, one of those being the lack of half-frontal faces without glasses in the dataset, or an insignificant amount when compared with frontal faces. A similar issue happens when glasses with white frames are in the picture. Although the lenses are mostly gone, the frames are still present in the output, as in Figure 3.3(b).

In frontal face pictures, however, the model performs well, removing the glasses from the subject's face, but some issues are still present. Most of the samples do not present

Figure 3.2 The result of the refiner when given a picture with sunglasses as an input.

(a) (b)

Figure 3.3 Examples of how the model struggles with: (a) sideways faces and (b) white glasses frames.

a high degree of realism due to distortions in the resulting faces. Those can be either a result of glasses going beyond the subjects' faces (Figure 3.4(a)) or simply an artifact of the network (Figure 3.4(b)).

In Figure 3.5 we can see the evolution of the refiner during training for 10 images from the validation set. It is noticeable that the progress for people wearing sunglasses is slower than the progress for subjects wearing reading glasses. We can also see that the network shows more realistic outputs for thinner frames of lighter colors, as it struggles a bit more with thick black frames.

Therefore, through our qualitative validation we could infer that, although the model produced promising results, artifacts and generalization are still issues. The model doesn't respond quite well to images that escape the characteristics of the majority of the base.

3.2.2 Quantitative

In order to perform a quantitative validation, we analyzed how well a CNN could distinguish pictures with and without glasses using the images generated by our refiner. For that we used a network with the same architecture of the discriminator (Table 2.1) with the addition of an average pooling layer in the end, to generate a single output. We used two different datasets for training and one for validation. The two training datasets were composed of 2000 real images of people with glasses, randomly sampled from our SimGAN training dataset, and 2055 fake images of people without glasses, the same images analyzed in Section 3.2.1. Our validation set was composed of 2000 real images of people without glasses, also randomly sampled from our SimGAN training dataset. The network was trained for 600 steps, each step being training a mini batch of size 16

(a) (b)

Figure 3.4 (a) Glasses that go over the subjects' faces. (b) Network artifacts.

composed uniformly by samples from both training sets with a learning rate of 2×10^{-4} and using the Adam optimizer.

Our accuracy in the validation set was of 95.35%, meaning that of the 2000 real images of people without glasses, the network correctly classified 1908 images as people without glasses. Therefore, we can infer that although the outputs provided by our network are not very realistic, a CNN was not able, in most of the cases, to differentiate between real and fake image with glasses. However, nothing can be said regarding improvements in facial recognition, as this experiment did not take into account the identity of the subjects.

3.3 RELATED WORK

Non-adversarial solutions have approached the problem of removing of eyeglasses from pictures. Although we could not find any other unsupervised learning methods, there were a couple of supervised learning models.

A version of Super-Resolution Convolutional Neural Network (SRCNN) was trained for this purpose using images and their ground truth (LIANG et al., 2017). To avoid the lack of big datasets with ground truths, the researchers used a set of aligned face images and introduced glasses artificially using facial landmarks. Although the results shown have a lot less face modification as ours, the use of artificially placed glasses seems to have made the model leave frame remains in the pictures. Our approach leaves close to no remains of the frame in most cases when applied to the same scenario chosen to showcase their results (frontal faces). It also seems to be more resistant to different kinds of frames and different glasses positions (in relation to the eyes), due to the use of a large database with several types of glasses in real scenarios. No comparison can be made in the case of half-frontal faces, as only front facing subjects were showcased in their work.

Solutions unrelated to deep learning also tackled this problem. An earlier work using

Figure 3.5 The first column is composed by input images and the following columns represent the results every 1000 training steps in increasing order.

PCA reconstruction (DU; SU, 2005) was proposed. Although the work provided interesting results, the extraction of the eyeglasses' region can be quite difficult in in-the-wild scenarios, specially considering different poses. Due to the use of PCA, the variation in lighting conditions can also be a problem for reconstruction. Their method provides more realistic results, with way less modifications in the face. However, our model is not as restrictive, as it is resistant to different lighting conditions and more well suited for the in-the-wild scenario.

Chapter 4

CONCLUSION

We proposed a method derived from SimGAN for removing eyeglasses from facial images through unsupervised learning. Our method provided good results for front-facing images, with the introduction of a few artifacts. It showed to be resistant to different kinds of frames, but struggled with lighter coloured ones. On faces with different poses, the model struggled with the frame temples, with remaining parts of them being seen in most samples. The solution performed surprisingly well with sunglasses, imitating eyes when the lenses' shades hid the subject's eyes, but those results were not realistic enough.

4.1 FUTURE WORK

In our current model, a few improvements could be done in training to achieve more realistic results. One of these improvements is to balance the training database considering the different characteristics of the subjects (such as gender and ethnicity) and of the pictures themselves (such as poses and lighting). The use of incremental resolutions during training has also shown to provide good results for GANs (KARRAS et al., 2017) and could be done with this model. Lastly, the introduction of a new loss related with facial recognition could be another way to preserve the facial traits of the pictures in order to provide more realism to the results.

Also, considering that this architecture was though for gray-scale aligned images, other works should explore the use of similar techniques for coloured pictures and unaligned faces. Improvements should also be done in the current model to reduce the artifacts generated by the network and allow better results with different poses. The task of removing eyeglasses is very different from removing sunglasses, due to the lack of eye information in the latter. Therefore, separate models for each task should be studied, to determine if they would have a better performance than a combined solution.

BIBLIOGRAPHY

BOUSMALIS, K. et al. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. v. 1, n. 2, p. 7.

DU, C.; SU, G. Eyeglasses removal from facial images. *Pattern Recognition Letters*, Elsevier, v. 26, n. 14, p. 2215–2220, 2005.

GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680.

GUO, J. et al. Face synthesis for eyeglass-robust face recognition. *arXiv preprint arXiv:1806.01196*, 2018.

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

KARRAS, T. et al. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LIANG, A. et al. Face recognition despite wearing glasses. In: IEEE. *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*. [S.l.], 2015. p. 1–8.

LIANG, M. et al. Deep convolution neural networks for automatic eyeglasses removal. *DEStech Transactions on Computer Science and Engineering*, n. aiea, 2017.

LIU, Z. et al. Deep learning face attributes in the wild. In: *Proceedings of International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015.

RIGHI, G.; PEISSIG, J. J.; TARR, M. J. Recognizing disguised faces. *Visual Cognition*, Routledge, v. 20, n. 2, p. 143–169, 2012. Disponível em: ⟨https://doi.org/10.1080/13506285.2012.654624⟩.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 815–823.

SHAO, L.; ZHU, R.; ZHAO, Q. Glasses detection using convolutional neural networks. In: SPRINGER. *Chinese Conference on Biometric Recognition*. [S.l.], 2016. p. 711–719.

SHRIVASTAVA, A. et al. Learning from simulated and unsupervised images through adversarial training. In: *CVPR*. [S.l.: s.n.], 2017. v. 2, n. 4, p. 5.