Challenge #3: Generative Adversarial Networks+DeepFakes

Course: ImSecu Spring 2023

Supervisors: Sahar Husseini; Prof. Jean-Luc Dugelay

Deadline: 21 April 2023

Introduction:

The rapid evolution of artificial intelligence (AI) has enabled the development of sophisticated technologies, such as Generative Adversarial Networks (GANs) and deepfake algorithms, which can produce highly realistic yet misleading media content. The proliferation of these technologies highlights the critical importance of deepfake detection in accurately identifying manipulated media and mitigating the dissemination of false information. The goal of this challenge is to enhance our understanding of the underlying mechanisms involved in the generation of fake visual content and the effectiveness of detection methods.

Objective:

This challenge includes two distinct tasks. The primary objective of the first task is to generate deepfake images using an existing algorithm and subsequently evaluate the efficacy of a detector in identifying these manipulated images. In the second task, we will create synthesized images using a basic GAN model. Through this exercise, we will gain practical experience in creating fake datasets.

TASK 1: How easy is it to create a deepfake?

The goal of this exercise is to familiarize you with deepfake generation and detection. In this exercise, you use a reenactment method to animate a source face image by a driving video.
To complete this task:

- **Deepfake Generator**
  1. Familiarize yourself with the reenactment concept. Explain what is reenactment in the report.
  2. Use demo.ipynb in Demo for paper "First Order Motion Model for Image Animation" [2] to create deepfake videos.
     - Try to take a few pictures of yourself under different lighting or with different accessories (different conditions in general) and try to generate deepfake videos that are as realistic as possible and as fake as possible. You can also use the SoF dataset in the following link:
       https://drive.google.com/file/d/1ufydwhMYtOhxgQuHs9SjERnkX0fXxorO/edit
     - In the notebook, you need to add `!pip install -U scikit-image==0.18.0` to the first section.
     - You can also upload your own driving videos in the Colab. **1)** Navigate through Files to “/content/demo/videos”. **2)** Use the three dots to upload a video. The video has to be named correctly, depending on the sequence - e.g. if the last video is 4.mp4, your has to be 5.mp4. **3)** Make sure to also change the range value in the notebook so the script can process it: video_tab.children = [ipywidgets.HBox([create_video(i) for i in range(5)])] # here change 5 to 6 if you added a video
- **Deepfake Generator**
  3. Extract a number of frames from your fake videos
  4. Feed a few of your generated images to the following deepfake detector Deepfake Detection - a Hugging Face Space by aaronespasa

5. Report your deepfakes and the results gained by the deepfake detector with explainability map (The detection algorithm outputs the face image with the explainability map). Show the results at least for 4 videos and explain your result.
6. What are the potential benefits and risks associated with deepfake technology?

TASK 2: From Data to Dollars: Using GANs to Generate Bags for Your Online Shop

You are a designer working for an online fashion store that wants to stand out from the competition by offering unique, stylish bags that customers can't find anywhere else. To determine which designs are worth producing, the online shop has decided to use GAN-generated fake images on the website before producing the actual bags. This is a good approach to test the market's new designs and ensure that they have the potential to sell well.

To achieve this, you have been tasked with using Generative Adversarial Networks (GANs) to create realistic-looking bag designs. You will be working with the MNIST fashion dataset, which consists of images of clothing and accessories, including bags. Your goal is to create bags that are so realistic that customers won't be able to guess that they're fake

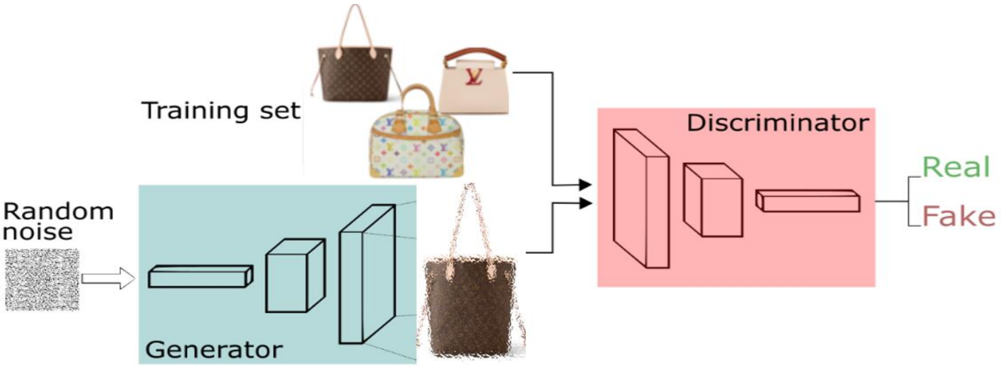

Figure 1. Generative Adversarial Network

The goal of this exercise is to familiarize you with the generation of fake data using a simple GAN architecture. You can use your own computer or google Colab. To generate bags, you will need to use bags (label==8) from MNIST fashion dataset.

To complete this task:

Familiarize yourself with the code inside the notebook (imageSecurityChallenge.ipynb) and make sure that it runs as expected. Please note that the code in the .ipynb script is completed, however, the GAN includes a very simple generator and discriminator consisting of a few linear layers.

1. Train the model from scratch, and change the generator and discriminator architectures to improve the network's result. Additionally, change network hyperparameters to improve the network's performance.
2. During training check the generated images and the loss curve in each step to find the best step to stop the training (Refer to [1] for more details). The generated images and the loss curve plot are saved in the "results_baseline" folder.
3. Report your experiments (new parameters and architectures) with the generated images and loss curves and return the .ipynb file.
4. Choose the best-generated image and name it best_generated_by_yourFirstName_lastName.

5. After the deadline, you will be provided with a Google form to perform a subjective test on other students' results.

References:

[1] https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/

[2] Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E. and Sebe, N., 2019. First order motion model for image animation. *Advances in Neural Information Processing Systems*, *32*.