

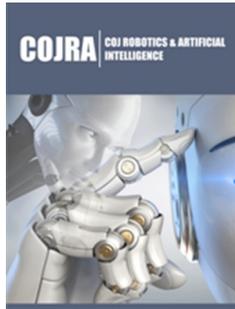
Transition Yourself - Emotion, Gender and Region

Gayathri R^{1*}, Neelanarayanan V² and Malav M²

¹School of Computer Science and Engineering, Vellore Institute of Technology, India

²School of Computer Science and Engineering, Vellore Institute of Technology, India

ISSN: 2832-4463



***Corresponding author:** Gayathri R, School of Computer Science and Engineering, Vellore Institute of Technology, India

Submission: 📅 May 24, 2022

Published: 📅 August 10, 2022

Volume 2 - Issue 3

How to cite this article: Gayathri R Neelanarayanan V, Malav M. A Short History of Cataract Surgical Training. COJ Rob Artificial Intel. 2(3). COJRA. 000536. 2022.

DOI: [10.31031/COJRA.2022.02.000536](https://doi.org/10.31031/COJRA.2022.02.000536)

Copyright@ Gayathri R, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract

The Image-to-image translation process belongs to a class of vision and graphics problem area, where the goal is to learn the mapping between an input and output image using a training set consisting of aligned image pairs. This research will focus on utilisation of image transitioning to make a Java Spring based Web-App which can transition a portrait into different human-oriented emotion (i.e., happy, sad, angry and frown), and simultaneously identify and transition its demographic region (i.e., America, Europe, Asia, South Asia) and same can be applied to the biological gender (i.e., Male, Female). The research study will use a proven deep learning methodology, which is Generative Adversarial Networks (GAN).

Keywords: General Adversarial Network (GAN); Deep learning; Pix-to-pix GAN; Cycle GAN; Webapp; Image-to-image transition

Introduction

Many issues in image processing, computer graphics, and computer vision may be simplified to “translating” an input image into an output image that looks comparable. A scene can be represented as an RGB image, a gradient field, an area map, a semantic label map, and so on, in the same way that a concept can be communicated in both English and French. In a similar fashion to automatic language translation, we describe automated image-to-image translation as the task of converting one feasible representation of a scene into another given enough training data. Traditionally, each of those functions has been handled by distinct, special-motive gear, despite the fact that the environment remains consistent.: Pixels should be expected from pixels. The goal of this study is to provide a framework for changing the transition of a dataset of individuals from one area, gender, and emotion to another [1]. Deep learning technologies and generative adversarial networks, often known as cycle GANs, are responsible for this.

It would be wonderful if we could provide a high-level objective, such as “make the output indistinguishable from reality,” and then have the image-to-image transition to the goal’s associated result. Fortunately, utilising the recently suggested Generative Adversarial Networks, this is precisely what is implemented (GANs). GANs learn a loss that tries to classify whether the output picture is genuine or false while also training a generative version to lower the loss. Blurry photos will no longer be accepted since they look to be clearly false. GANs are discovered in the conditional situation in this work. Conditional GANs (cGANs) evaluate a conditional generative version of data in the same way as GANs do [2]. This qualifies cGANs for image translation tasks, which need us to locate an input image and create a corresponding output image. In the last two years, GANs have been extensively investigated, and the numerous solutions we discuss in this study have been explicitly suggested. However, early research focused on specific applications, and it’s unclear how effective image GANs

may be as a general-reason solution for image translation. Our first contribution is to show that conditional GANs can solve a wide range of problems at a reasonable cost.

Related Work

Generative Adversarial Networks (GANs) have performed impressive results in image generation, photo editing, and representation learning. Recent strategies undertake the identical concept for conditional image generation applications, which include image to region, image to gender, and image to emotion, in addition to different domain names like videos and 3D data [3]. The key to GANs' fulfilment is the concept of an adverse loss that forces the generated images to be, in principle, indistinguishable from actual photos. This loss is in particular effective for image generation tasks, as that is precisely the goal that a whole lot of computer graphics targets to optimize. We undertake an adversarial loss to study the mapping such that the translated images cannot be prominent from images in the target domain.

Image-to-Image Translation The concept of image-to-image translation is going back at least to Hertzmann et al. 's Image Analogies, who rent a non-parametric texture model on a single input-output training image pair. More recent techniques use a dataset of enter-output examples to examine a parametric translation feature the usage of CNNs. Our method builds on the "Cycle GAN" framework, which uses a conditional generative adversarial network to examine a mapping from input to output images [4]. Similar thoughts had been implemented to various tasks which included generating images from people images or from characteristic and semantic layouts.

Method

GANs are generative models that examine a mapping from random noise vector z to output image y , $G: z \rightarrow y$ [24]. In contrast, conditional GANs examine a mapping from observed image x and random noise vector z , to y , $G: x \rightarrow y$. The generator G is trained to produce outputs that can't be outstanding from "real" images via the means of an adversarially trained discriminator, D , that is trained to do in addition to detecting the generator's "fakes" [5].

Understanding and implementing of cycle gan in tensor flow

The paper we are going to implement is titled "Image-to-Image Translation to region, emotion and gender using Cycle-Consistent Adversarial Networks". The title is quite a mouthful, and it helps to look at each phrase individually before trying to understand the model all at once. However, the above adversarial training strategy has a flaw. Citing the original paper's authors: Adversarial training can, in theory, research mappings G and F that produce outputs identically allotted as target domains Y and X respectively. However, with massive sufficient capacity, a network can map the equal set of input images to any random permutation of images in the goal domain, in which any of the learned mappings can induce an output distribution that fits the target distribution [6]. Thus, an adversarial loss alone cannot assure that the learned feature can map an

individual input x_i to a preferred output y_i . To regularize the model, the authors introduce the constraint of cycle-consistency - if we transform from supply distribution to target and then back again to supply distribution, we have to get samples from our supply distribution.

Process to Implementation

In the project many technologies we are going to use, so to start the implementation we need to install some packages like tensor flow because we are going with the cycleGAN so to install the tensor flow package we need to give the following command - (Figure 1) After installing the above packages we need to download the dataset. In my project I am collecting the man and female dataset. To learn and start the training i am going to train the different dataset which is horse and zebra dataset. In the dataset there are four folders TrainA, TrainB, ValA, ValB. in the TrainA folder horse and zebra picture exists. We collected more than 100 pictures and differentiated the images into the different folders. 90% images are in the TrainA folder and the rest of the 10% images in the ValA folder. Because when we are going to train, we need to validate the images that are real or fake. So separated the images into different folders [7]. Further, after setting up with the dataset we have to set which folder we are going to train, but in our case, we are going with the TrainA folder. So, we need to add the images and download them to the terminal.

```
[ ] import tensorflow as tf
[ ] import tensorflow_datasets as tfds
    from tensorflow_examples.models.pix2pix import pix2pix
import os
import time
import matplotlib.pyplot as plt
from IPython.display import clear_output
AUTOTUNE = tf.data.AUTOTUNE
```

Figure 1:

```
bash ./scripts/download_cyclegan_model.sh horse2zebra
```

Figure 2:

```
bash ./scripts/download_cyclegan_model.sh horse2zebra
```

Figure 3:

To download the images, we need some commands so we will apply the following command - (Figure 2) When we apply the above command terminal will connect to the server which dataset is available and download the dataset. After the process we need to set which folder images we are going to train. And the folder is TrainA. When we run the command terminal will download all the images on the terminal and will saved to the './checkpoints/horse2zebra_pretrained/latest_net_G.pth' folder. So doing all this we set a model which we call horsezebra model. To test the model we need to run the following command-(Figure 3) After running the command, it will start the testing. It will take some time to test because the model will get all the images and test them. After successfully testing we

B. For pix2pix and your own models, you need to explicitly specify `--netG`, `--norm`, `--no_dropout` to match the generator architecture of the trained model.

After successfully running the command, we will train the model. To train the model we need to run the following command -

```
python train.py --dataroot ./datasets/horsetozebra --name
facades_pix2pix --model pix2pix
--direction BtoA
```

Dataset

Affectiva's emotion database has now grown to nearly 6 million faces analysed in 75 countries. To be precise, we have now gathered 5,313,751 face videos, for a total of 38,944 hours of data, representing nearly 2 billion facial frames analysed. This global data set is the largest of its kind-representing spontaneous emotional responses of consumers while they go about a variety of activities. To date, the majority of our database consists of viewers watching media content (i.e, ads, movie trailers, television shows and online viral campaigns) [8]. In the past year, we have expanded our data repository to include other contexts such as videos of people driving their cars, people in conversational interactions and animated gifs.

Literature Review

Paper - 1: generative adversarial network

In the paper they recommend a new framework for estimating generative models thru an adversarial process, wherein we concurrently train models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G . The training system for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax -player game. In the space of arbitrary functions G and D , a completely unique solution exists, with G recovering the training data distribution and D identical to 1/2 everywhere. In the case in which G and D are defined through multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during both training and generation of samples [9]. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

Paper-2 Image to image translation with conditional adversarial network

They investigate conditional adversarial networks as a general-purpose solution to image-to-image translation troubles. These networks now no longer only learn the mapping from input image to output image, however, also learn a loss function to train this mapping. This makes it possible to apply the identical generic approach to problems that traditionally might require very different loss formulations. We show that this approach is effective at synthesizing photos from label maps, reconstructing objects

from edge maps, and colorizing images, amongst other tasks. Indeed, considering the fact that the release of the pix2pix software program associated with this paper, a big number of internet users (lots of them artists) have posted their personal experiments with our system, further demonstrating its extensive applicability and simplicity of adoption without the need for parameter tweaking. As a community, we no longer hand-engineer our mapping functions, and this work suggests we are able to acquire reasonable results without hand-engineering our loss functions either.

Paper-3 Unpaired image-to-image translation using cycle consistent adversarial network

Image-to-image translation is a category of vision and graphics issues in which the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for plenty of tasks, paired training data will not be available. We present a method for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y the use of an adversarial loss. Because this mapping is highly under-constrained, we couple it with an inverse mapping $F : Y \rightarrow X$ and introduce a cycle consistency loss to enforce $F(G(X)) \approx X$ (and vice versa). Qualitative results are presented on several tasks in which paired training data does now no longer exist, which includes collection style transfer, object transfiguration, season transfer, photo enhancement, etc. Quantitative comparisons against several prior techniques demonstrate the superiority of our approach.

Paper - 4 Cycle-gan based emotion style transfer as data augmentation for speech emotion recognition

Cycle consistent adversarial networks (CycleGAN) have shown great fulfilment in image style transfer with unpaired data sets. In-spired with the aid of using this, we check out emotion style transfer to generate synthetic data, which aims at addressing the data scarcity problem in speech emotion recognition. Specifically, we propose a CycleGAN-primarily based totally method to transfer feature vectors extracted from a huge unlabeled speech corpus into synthetic features representing the given target emotions. We extend the CycleGAN framework with a classification loss which improves the discriminability of the generated data. To show the effectiveness of the proposed method, we present results for speech emotion recognition the use of the generated feature vectors as (i) augmentation of the training data, and (ii) as standalone training set. Our experimental results reveal that when utilising synthetic function vectors, the classification overall performance improves within-corpus and cross-corpus evaluation [10].

Paper - 5 Segmentation enhanced cycle-gan

Algorithmic reconstruction of neurons from electron microscopy data traditionally requires training system learning models on dataset-particular ground truth annotations that are expensive and tedious to acquire. We enhanced the training

procedure of an unsupervised image-to-image translation approach with extra components derived from an automated neuron segmentation approach. We show that this approach, Segmentation-Enhanced CycleGAN (SECGAN), enables close to perfect reconstruction accuracy on a benchmark connectomics segmentation dataset despite operating in a “zero-shot” setting wherein the segmentation model was trained using only volumetric labels from a different dataset and imaging approach. By decreasing or eliminating the need for novel ground fact annotations, SECGANs alleviate one of the main practical burdens involved in pursuing automated reconstruction of extent electron microscopy data.

Dataset

Affectiva’s emotion database has now grown to nearly 6 million faces analysed in 75 countries. To be precise, we have now gathered 5,313,751 face videos, for a total of 38,944 hours of data, representing nearly 2 billion facial frames analysed. This global data set is the largest of its kind-representing spontaneous emotional responses of consumers while they go about a variety of activities [11]. To date, the majority of our database consists of viewers watching media content (i.e, ads, movie trailers, television shows and online viral campaigns). In the past year, we have expanded our data repository to include other contexts such as videos of people driving their cars, people in conversational interactions and animated gifs.

Link :

<https://blog.affectiva.com/the-worlds-largest-emotion-database-5.3-million-faces-and-counting>

Conclusion

The consequences on this paper recommend that conditional adversarial networks are a promising method for lots of image to image translation tasks, particularly the ones concerning highly structured graphical outputs. These networks learn a loss adapted to the task and information at hand, which makes them relevant in an extensive sort of setting. in the paper given input could be changed into any region, emotion and gender with their respective fields [12,13]. to provide the security and recognition for any

corporates this method could be useful. The technology used is deep learning through general adversarial networks is the main factor of the project.

References

1. Isola P, Yan Zhu J, Zhou T, Efros A (2018) Image-to-image translation with conditional adversarial networks. Berkeley AI Research (BAIR) Laboratory, University of California, Berkeley, California, USA.
2. Goodfellow IJ, Abadie JP, Mirza M, Xu B, Farley DB, et al. (2014) Generative adversarial nets. Department of Computer Science and Operations Research, Universite de Montreal, Montreal, Canada.
3. Bao F, Neumann M, Thang Vu N (2019) CycleGAN-based emotion style transfer as data augmentation for speech emotion recognition. Institute for Natural Language Processing (IMS), University of Stuttgart, Stuttgart, Germany.
4. Zhu JY, Park T, Isola P, Efros AA (2020) Unpaired image-to-image translation using cycle-consistent adversarial networks. Berkeley AI Research (BAIR) laboratory, University of California, Berkeley, California, USA.
5. Januszewski M, Jain V (2019) Segmentation-enhanced cycleGAN, bioRxiv.
6. Brownlee J (2019) How to develop a cyclegan for image-to-image translation with keras.
7. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, et al. (2016) The cityscapes dataset for semantic urban scene understanding. 2016 IEEE Conference on Computer Vision and Pattern Recognition CVPR, Las Vegas, Nevada, USA.
8. Bousmalis K, Silberman N, Dohan, Erhan D, Krishnan D (2017) Unsupervised pixel-level domain adaptation with generative adversarial networks. CVPR.
9. Goodfellow I (2017) Generative adversarial networks. OpenAI.
10. Ledig C, Theis L, Huszar F, Caballero J, Aitken A, et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. CVPR.
11. Xiong F, Wang Q, Gao Q (2019) Consistent embedded gan for image-to-image translation. State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an, China.
12. Wang H, Yu CN (2019) A direct approach to robust deep learning using adversarial network. Published as a conference paper at ICLR 2019.
13. Bansal H, Rathore A. Understanding and implementing CycleGAN in tensorflow.

For possible submissions Click below:

[Submit Article](#)