Casual2Professional: Image-to-Image Translation for Professional Headshots using Generative Adversarial Nets (GANs)

Bi Yaqi (A0218845W), **Hu Zhongda** (A0218879H), **Wang Guoqing** (A0218979E), **Wang Huijuan** (A0218881W), **Wong Fu Jie** (A0218916X)

Abstract

We explore the use of Generative Adversarial Nets (GANs) to translate headshot photos of individuals in casual attire into professional attire that are suitable for use as LinkedIn profile photos. We used a cycle-GAN algorithm for cross-domain image translation and trained the model on real-life photos scraped from LinkedIn profiles from employees of companies in various sectors. Our final model is able to learn the professional dressing style from the professional data domain and apply the attire to another casual domain. It successfully transfers casual dressing like t-shirts into suit-and-tie after 200 epochs training, which achieves our initial objective. More training data with diversity in images would be useful to further improve the output image such as reducing distortion of other parts of the images such as face. We also discuss possible solutions that could address these issues.

1. Introduction

1.1 Problem Background

LinkedIn is the world's largest professional online networking platform, and currently has more than 700 million members in over 200 countries and territories. With the advent of working from home and virtual networking due to the global COVID-19 situation in 2020, more recruiters and employers are expected to rely on LinkedIn to identify and connect with potential talent internationally. Statistics from LinkedIn showed that members with a photo in their profile receive 21 times more profile views and 9 times more connection requests (Jalan, 2017). Having a good and suitable profile photo that allows the viewer to form a positive first impression of the candidate will further enhance his success rates in professional networking using the LinkedIn platform.

Several online resources provide useful tips on good profile photos (Abbot, 2019; Meero.com, 2020) which commonly emphasise on coming across as approachable and friendly as well as natural and authentic. At the same time, the individual should be attired professionally and should reflect his/her current look as far as possible. Therefore, it is commonly recommended to engage a professional photographer to take a headshot photo for use as profile. However, as such professional photos tend to be in staged poses, individuals may find that such professional photos come across as stiff and does not convey the approachable look.

Therefore, we explore Machine Learning methods to develop an algorithm using Generative Adversarial Nets (GANs) to generate professional-looking headshots from the same individual's own photos in a casual setting and attire. In casual photos, he individual would be more relaxed and would be more likely to put across a friendly and approachable look. Hence, the success of our model would be for the algorithm to convert the individual's attire to business attire while keeping other features e.g. face, background as unchanged as possible.

1.2 Machine Learning Methods for Image Generation

The image generation topic falls under the generative modelling field in unsupervised machine learning, whereby the algorithm identifies and learns the patterns in the input data and generates new data that could plausibly been in the original dataset. There has been much development and applications in the last few years following the invention of the Generative Adversarial Networks (GANs) by Ian Goodfellow et al. (2014). The GAN model architecture involves two sub-models that combines both supervised and unsupervised learning concepts: (a) a generator model that is trained on the input data and generates or creates new observations and (b) a discriminator model that aims to distinguish "fake" observations generated by the generator model from real observations drawn from the training set. In the machine learning process, the generator model and the discriminator model are updated in a zero-sum or adversarial manner with the generator model updated based on how well the output it generated successfully fooled the discriminator (i.e. generated photo was misclassified as real by the discriminator) whereas the discriminator would be updated

Masters of Science in Business Analytics (MSBA), NUS.

based on how well the discriminator correctly identified the generated images from the real images.

Since then, researchers have made much progress and development in the use of GANs for various image generation applications (Brownlee, 2019). Some of the applications of GANs in photo generation include the research by NVIDIA researchers Karras et al. (2018) to generate highly realistic photographs of human faces based on celebrity photos, by Ma et al. (2017) to generate new photos of human models with new poses, and by Perarnau et al. (2016) using a GAN to reconstruct photographs of faces with specific specified features, such as changes in hair colour, style, facial expression, and even gender.

2. Dataset

Image recognition is a common task for deep neural networks. In order to find a legal public source for large amount of professional personal images, we used LinkedIn (http://www.linkedin.com/) as our main data source to train and test the neural networks. On LinkedIn, every user can their image photo as their profile photo and their profile photo is visible to other LinkedIn members.

2.1 Data Collection

We randomly selected users and collected their public profile photos. In the process of collecting, we have established filtering rules, so that the selected users are diverse in terms of gender, age, and skin colour. A data collection workflow is established to obtain these image data. First, we analysed the URLs and HTML structure of the LinkedIn website to locate the user profile photo files we need. In this process, Hou Yi Collector, a free data collection software, was used to help us analyse web page elements. Then, we generated URLs and XPath codes with filtering rules. These URLs and XPath codes accurately locate the user profile picture files we need on the web page. Finally, we input the generated URLs and XPath codes into the Hou Yi collector. After setting the related parameters, a task of automatically collecting and downloading pictures is created and executed.

The process of image data collection is time-consuming and unstable due to the traffic restriction from the LinkedIn website. Through multiple collections, we collected a total of 5953 colour photos in 100*100 resolution.

2.2 Data Cleaning

We cleaned the data to reduce noise and make the dataset more suitable for model training. First, using algorithms, we removed duplicate photos. Then, we manually cleaned the photos according to some rules. We removed those photos that were not real people, with blurred facial information, have complicated photo backgrounds, and shoot from abnormal camera angles. The important features of the photos we selected are: (a) only one person appears in the photo; (b) showing clear attire and facial features; (c) uncluttered photo background; (d) the camera angle is frontal.

2.3 Manually Labelling

The user profile photos collected from LinkedIn do not contain information other than the image itself. Manual classification is necessary to get the training labels of the images.

2.3.1 CASUAL TO INDUSTRY

Intuitively, we divided our dataset into casual pictures and professional pictures. Then we selected different industries furtherly in that different industries have various dressing styles although all of them are professional. Take high-tech and banking as examples, staff's attire in big-tech companies are obviously different from staff in banking. Finally, we set all of our professional dataset into 8 industries, banking, big-tech, government, law, oil and gas, retail, transport.

Unfortunately, after the implementation of industry division, our model's performance was found to be unsatisfactory. Therefore, we regrouped our dataset to focus on performing casual to professional translation.

2.3.2 CASUAL TO PROFESSIONAL

We manually labeled the cleaned data into professional and casual photos. The standard for photo classification is whether the person in the picture is dressed in a professional style e.g. individuals in jacket and tie would be classified as professional whereas photos of persons wearing T-shirts were classified as casual photos.

The raw data we collected are imbalanced because most of the user profile photos in LinkedIn were professional photos. We adopted the under-sampling method to deal with this problem and got 835 casual photos and 918 professional photos, which is a balanced data set. Finally, we split the labeled data set into a training set and a test set.

There are 715 casual photos and 798 professional photos in the training set. In the test set, there are 120 casual photos and 120 professional photos. We also used the group members' own casual photos for the model evaluation.

2.4 Image augmentation

Data collection on LinkedIn websites is time-consuming and limited. In order to expand the training sample size effectively, and improve the performance of the model, image augmentation is conducted. We used random brightening, random cropping, random sharpness enhancement, random contrast enhancement, and other methods to augment the raw data, and expanded the sample size of the training set by five times.

3. Modelling

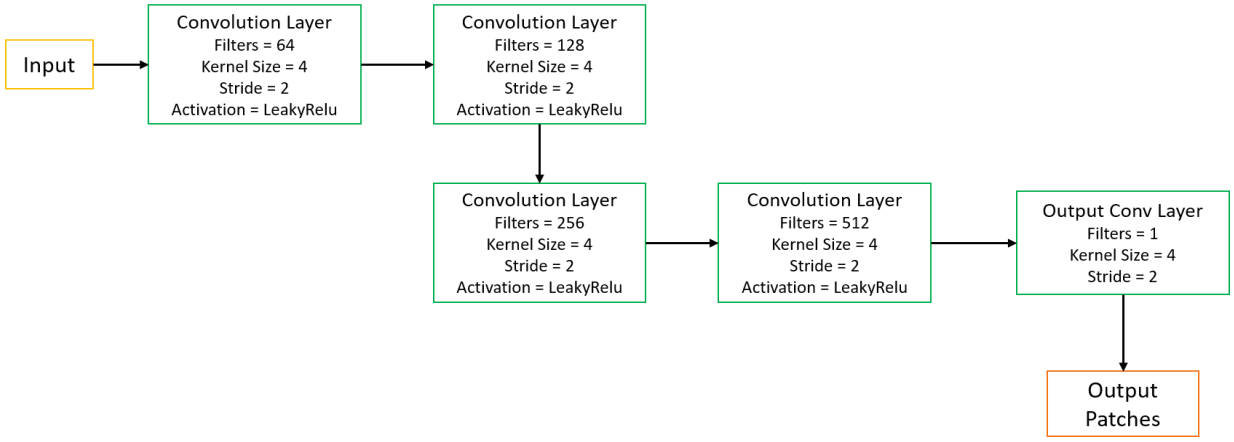

Figure 1. CycleGAN Discriminator Model

Due to the disparity between the 2 sets of profile images, we needed to employ a model that can be trained with unpaired datasets for image translation. Thus, we adopted the CycleGAN model (Zhu et. al., 2017) for our machine learning approach.

3.1 CycleGAN Model

The CycleGAN model is a GAN variant, training both a generator and discriminator to iteratively improve image generation and differentiation outputs. However, the unique proposition of CycleGAN is the usage of 2 generators, G and F, and 2 discriminators, X and Y, to perform the translation. Generator G translates from casual (domain A) to professional (domain B) and is also the main generator of interest while generator F performs the reverse translation. Discriminator X will differentiate between real and fake photos from domain B while Discriminator Y will do the same for photos from domain A.

Another characteristic of the model is the usage of instance normalisation as opposed to batch normalisation.

3.2 Discriminators

Figure 1 shows the CycleGAN's discriminator model. Both domains use the same 70 x 70 PatchGAN architecture which breaks up the input image into multiple patches and determines whether each patch is real or fake. Therefore, the discriminator outputs a vector of patch predictions rather than just one binary value.

The PatchGAN model enables the discriminator to take in arbitrarily sized photos and also has fewer parameters than full image discriminators, allowing it to train more quickly.

3.3 Generators

The CycleGAN generator is an autoencoder which uses several downsampling blocks to encode the input image before transforming it using several ResNet blocks. The transformed output is then fed into a decoder and upsampled to produce the final translated image. Figure 2 shows the overall generator model while Figure 3 drills into an individual ResNet block.

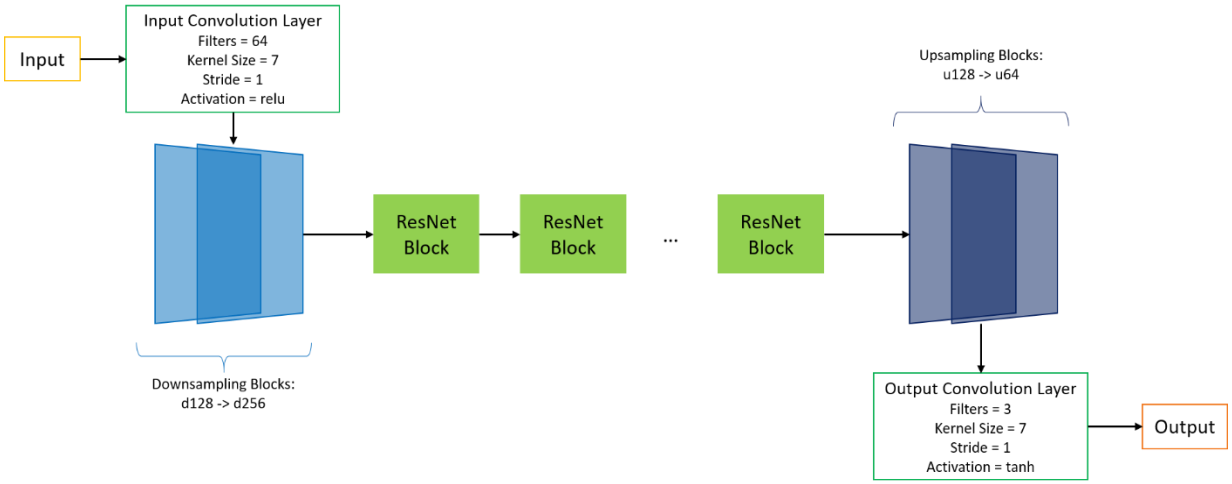

Figure 2. CycleGAN Generator Model

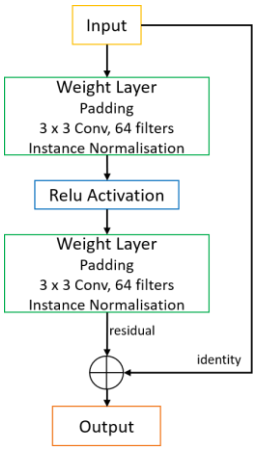

Figure 3. ResNet Blocks

ResNet blocks are designed to improve training efficiency for deep learning networks (He et. al., 2016). Rather than using stacked layers to train a mapping between the input and output values, $H(x)$, it instead attempts to determine the residual output through its convolution layers, $F(x)$. This is then concatenated with the original input, x, so that $H(x) = F(x) + x$.

The reference model recommends 2 different settings based on input image sizes; 6 ResNet blocks for 128 x 128 images and 9 ResNet blocks for 256 x 256 images. Our model is trained using both settings by first resizing the training photos before feeding them into the model.

3.4 Model Objective & Loss Functions

3.4.1 Adversarial Loss

As with other GANs, adversarial loss measures the difference between the generated image and how well the discriminator can distinguish the images. In the CycleGAN model, 2 values are produced; one between generator G and discriminator X, the other between generator F and discriminator Y.

3.4.2 Cycle Consistency Loss

To prevent mode collapse situations, where a generator maps all input to the same image, cycle consistency is introduced, wherein a translated image should be able to be retranslated back to its original image using the other generator. That is, $a \rightarrow G(a) \rightarrow F\big(G(a)\big) \approx a$ (forward consistency) and $b \rightarrow F(b) \rightarrow G\big(F(b)\big) \approx b$ (backward consistency). Both forward and backward consistency losses are computed.

3.4.3 Identity Loss

Identity mapping means that any image from domain A should experience no translation when fed into generator F and vice versa. The identity loss is employed to help preserve image colour.

3.4.4 Overall Objective

The overall generator loss is the summation of the 3 losses while the discriminator loss is the difference between the real/fake input image and the predicted patch output. A weight of 0.5 is given to the discriminator loss to slow down learning relative to the generator. The CycleGAN model aims to minimise generator loss while maximising discriminator loss. Therefore, CycleGAN tries to solve

$$G^*, F^* = \arg min_{G,F} max_{X,Y}\, L(G, F, X, Y)$$

where L represents the loss values of each generator and discriminator.

3.5 Training & Model Deviations

Each type of image input size was trained for a total of 200 epochs. Due to the amount of time required to train the CycleGAN model, multiple sessions were required which caused a deviation to the reference model. While learning rates of all optimisers were initialised to 0.0002 as per reference, the subsequent decay rates was not consistent due to the training breaks. This resulted in 'jumps' in the loss values when the training was restarted.

In total, the 128 x 128 input images with 6 ResNet blocks took around 24 hours to train while the 256 x 256 input images with 9 ResNet blocks took close to 48 hours.

4. Model Evaluation Methods

We considered various quantitative and qualitative evaluation approaches to evaluate the success of our trained model.

4.1 Quantitative Evaluation Approach

Assessing the quality of generated image is an open question. There are several different quality metrics are conducted by experts to evaluate the quality of the generated images from GAN models. Specifically, there are three commonly used full reference image quality assessment methods: (a) Structural Similarity Index (SSIM), (b) Feature Similarity Index (FSIM), and (c) Gradient Magnitude Similarity Deviation (GMSD).

4.1.1 STRUCTURAL SIMILARITY INDEX (SSIM)

The single-scale SSIM measure (Zhou et al., 2004) is a well-characterized perceptual similarity measure that aims to discount aspects of an image that are not important for human perception. If we have the ground truth professional version of a casual photo, we can use SSIM to compare corresponding pixels and their neighbourhoods in two images, denoted by x and y, using three quantities—luminance (I), contrast (C), and structure (S). The three quantities are combined to form the SSIM score: $\mathrm{SSIM(x,y)} = \mathrm{I(x,y)}^{\alpha}\mathrm{C(x,y)}^{\beta}\mathrm{S(x,y)}^{\gamma}$ where X and Y are the original and reconstructed images. Further, contrast and structure components can be weighted at each scale. The final measure is:

$$\mathrm{MS}-\mathrm{SSIM}(\mathrm{x},\mathrm{y})=\mathrm{I}(\mathrm{x},\mathrm{y})^{\alpha \mathrm{M}}\prod_{j=1}^{M} C_j\,(\mathrm{x},\mathrm{y})^{\beta_j}\mathrm{S}_j(\mathrm{x},\mathrm{y})^{\gamma_j}$$

where j denotes the scale which the original images are down sampled with a factor of 2^{j-1}.

This score ranges between 0 (low similarity) and 1 (high similarity) (Snell et al.,2017) defined a loss function for training GANs, which we can sum up this score for all images:

$$L(x,y)=-\sum_i MS-SSIM(x_i,y_i)$$

where i denotes an index over image pixels.

The SSIM index brings image quality assessment (IQA) from pixel-based stage to structure-based stage.

4.1.2 FEATURE SIMILARITY INDEX (FSIM)

Another method, the feature similarity index (FSIM) (Zhang et al., 2011) is proposed based on the fact that human visual system (HVS) understands an image mainly according to its low-level features. It measures the dissimilarity between two images(f_1, f_2) based on local phase congruency(PC) and gradient magnitude(GM). the similarity measure for $PC_1(x), PC_2(x)$ is defined as

$$S_{pc}(x)=\frac{2PC_1(x)*PC_2(x)+T_1}{PC_1^2(x)+PC_2^2(x)+T_1}$$

where T_1 is a positive constant to increase the stability of S_{PC}.

Similarly,

$$S_G(x)=\frac{2G_1(x)*G_2(x)+T_2}{G_1^2(x)+G_2^2(x)+T_2}$$

where T_2 is a positive constant depending on the dynamic range of GM values.

Combining these two components, we get:

$$S_L(\mathbf{x})=\mathrm{S}_{PC}(\mathbf{x})\cdot\mathrm{S}_G(\mathbf{x})$$

Having obtained the similarity $S_L(\mathbf{x})$ at each location $\mathbf{x}$, the overall similarity between f_1 and f_2 can be calculated.

$$FSIM=\frac{\sum_{x\in\Omega}S_L(x)*PC_M(x)}{\sum_{x\in\Omega}PC_M(x)}$$

where $PC_m(\mathbf{x})$ = max($PC_1(\mathbf{x})$, $PC_2(\mathbf{x})$) weights the importance of $S_L(\mathbf{x})$ in the overall similarity between f_1 and f_2. However, computing metrics based on HVS is very time consuming especially for PC measures.

4.1.3 GRADIENT MAGNITUDE SIMILARITY DEVIATION (GMSD)

GMSD (Xue et al.,2014) is another effective method for image quality assessment, where the pixel-wise gradient magnitude similarity (GMS) is used to capture image local quality, and the standard deviation of the overall GMS map is computed as the final image quality index. Such a standard-deviation-based strategy measures the difference between local quality of two photos.

4.2 Qualitative Evaluation Approach

These methods reflect the score from human visualization experience, denote picture quality from various aspects. Larger score of previous two methods mean better quality pictures and smaller score for the last method means better quality. Since all three methods rely on paired dataset, and we do not have the ground truth of casual photos in professional version, it is difficult for us to conduct these evaluation measurements.

One possible solution is performing a task of retouching the photos by experts, to generate a target domain for train dataset. Our test dataset consists 119 casual photos, then we can retouch each photo into professional version manually using photoshop. After that, we can conduct a quantitative comparison between the retouched photo and the professional images that produced by CycleGAN. We can measure the difference between our result images and expert-retouched images using common metrics.

4.3 Final Evaluation Approach for Model Selection

Though we have proposed the method of conducting quantitative measures for image quality assessment in previous part, human visual experience is the golden standard for assessing the quality of generated images. In this project, we conduct a 'opinion rank' testing by human. To realize this, we design a ranking system that asks each scorer to give a ranking indication of the perceived quality of each image generated by CycleGAN models with different epoch. Specifically in this task our members formed a scorer team to rank the images according to naturality and preference among different models for the output from the same test casual photo. Then we can evaluate the different models' performance based on average ranking for 119 testing photos, to make a decision on the final model.

5. Results and Evaluation

5.1 Initial results

After we put the industry classification of our dataset into CycleGAN model, the performance is not satisfactory. The initial results are as follows.

From Figure 4, we could see that all of faces in pictures were distorted and it was difficult to identify the background. The photo also appeared to take on a canvas-like appearance. A possible reason for this is that some of the pictures in our industry dataset have complex background, which is hard for the model to identify and clarify. If the complicated background could be removed,

there is a high probability that the performance will be improved.

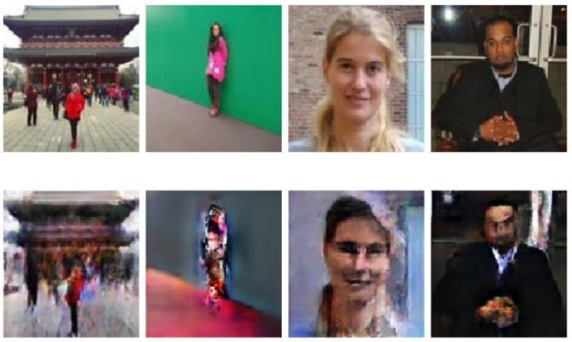

Figure 4. Initial results by industry classification.

5.2 Final results

Considering the bad results of initial results, we conducted a refinement into our industry dataset. We kept all simple background pictures and cut the raw data into casual and professional photos. After we employed our model into the refinement of professional pictures, the results are much better.

It is obvious that our CycleGan model translate casual dressing like T-shirt into professional dressing after 200 epochs training, and some males were put on ties. Most of our test results are changed in dressing styles. Although there are some distorted faces, it has a significant improvement, as observed in Figure 5. Generally speaking, the final results achieved our initial purpose.

Figure 5. Final results by professional refinement

5.3 Training loss functions

As shown in Figure 6, the generator loss decreased quickly in the first 10 epochs, and kept stable in the remaining 190 epochs.

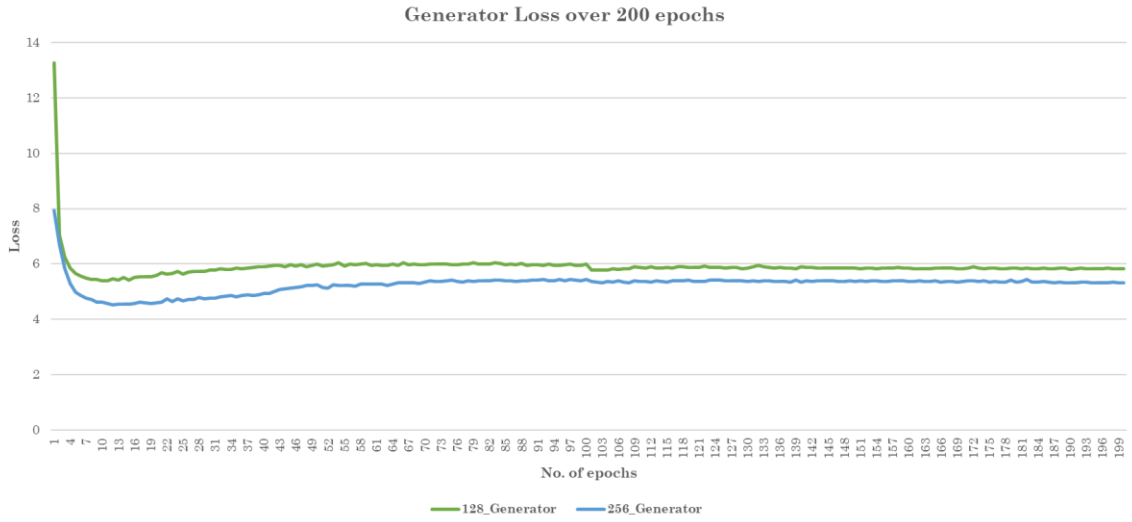

Figure 6. Generator Loss over 200 epochs

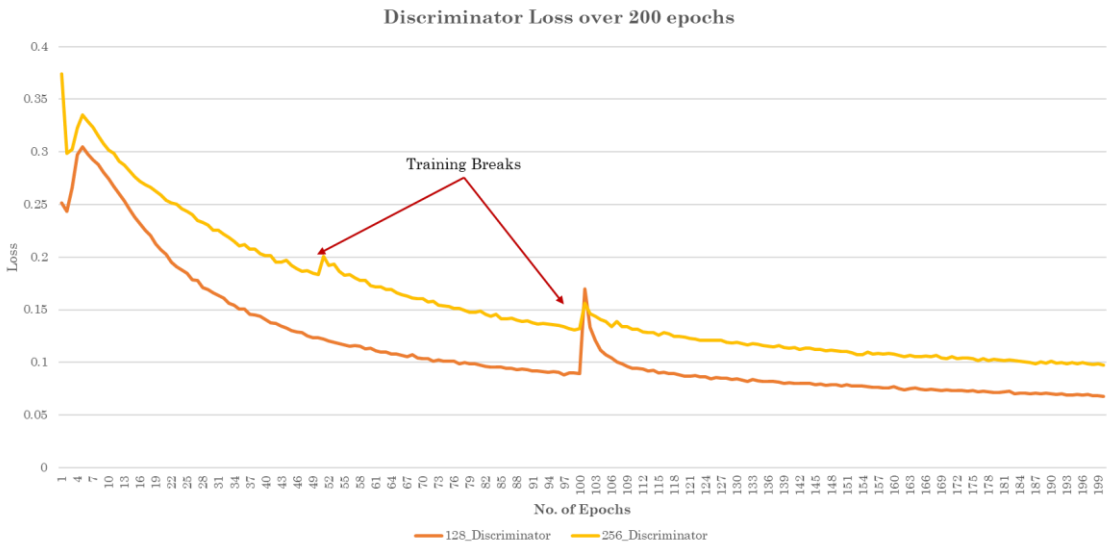

Figure 7. Discriminator Loss over 200 epochs

Discriminator loss chart, in Figure 7, displays a different trend that it decreased slowly in the first half period and there are some occasionally rise. The obvious rise in discriminator loss chart is in that the training period is much time consuming, and there are some manual breaks in training.

5.4 Evaluation

As stated in section 4.3 our team formed a scorer team to rank the images according to naturality and preference among different models for same training casual picture.

Taking the resolutions and property of neural network into consideration, we selected 6 epochs as our best epoch pool based on 2 different image dimensions (128*128 pixels, 256*256 pixels) with 3 completed epochs count (190, 195 and 200 epochs) for each. Then our team members ranked each picture in 6 epochs from 1 to 6 with 1 being the best transformed image and 6 being the worst transformed image for the same test image.

In doing the ranking, we used a combination of criteria focusing on: (a) how much the model managed to, or attempted to, transform the attire of the person to a professional attire. (b) how much the model distorted the facial and other features during the transformation. The ranking for each set of testing and output images were done independently and comparison was not made between output images from different test images in the same model/ epoch pool. This was because different test images had different extent of distortion and hence the worst output image in a specific test image set could actually have better resolution than the best output image from a different test image set.

The ranking results are as follows.

Table 1. Ranking Table

SIZE / EPOCH	128	128	128	256	256	256	sum
	190	195	200	190	195	200	
zhongda	5.31	4.69	3.90	2.81	2.39	1.90	21.0
yaqi	4.60	3.71	2.89	2.85	3.29	3.66	21.0
fujie	4.21	3.68	2.58	3.92	3.66	2.96	21.0
huijuan	3.59	4.04	3.18	3.34	3.79	3.05	21.0
guoqing	3.43	2.44	1.61	4.07	4.48	4.98	21.0
mean	4.23	3.71	2.83	3.40	3.52	3.31	21.0

In the table, each cell represents the average of the member ranking score in the corresponding epoch, with a lower score reflecting a better performance. The last row is the average of all ranking scores for that particular dimension-epoch pool. From the table, we found that 200-epoch with 128 pixels is the best epoch from our team members. Three out of five members had also on average ranked the images from this epoch pool as the best compared to the remaining epoch pools. Hence, we chose the 200-epoch for 128*128 pixels as our best epoch pool.

This final model was then used to transform the group members' own casual photos. We found that the performance of the final model on the team members' casual photos are comparable to photos in the testing set, as described in section 6 below.

6. Findings and Insights

6.1 Overall performance of model

In general, the performance of the model improved after the refinement to the training set such as only using training images with uncluttered background. While distortion was still observed in the faces of the photos, this is not surprising as the training set contained photos of different individuals or faces. This means that the model would also learn the patterns in the different faces and would apply these learnt patterns to the new testing photos, leading to transformation in the photo.

In particular, we found that as there was more variation in females' attire for business wear compared to males (standard jacket and tie), the model tended to have more difficulties doing a good transformation for females. This can be overcome with more training images for females.

6.2 Likely factors giving better image translation

There were still some success factors that gave better image translation. First, we found that our model is relatively successful when the casual photo to be converted has an uncluttered background which is likely because this makes it more similar to the training images and hence the algorithm focuses on other aspects of the image for translation.

We also found that images that were a headshot (i.e. the face taking up most of the photo area) were more successfully converted with less distortion compared to photos which captured the full-body pose. This is likely because this led to more differences between the casual test photos and the training professional photos to be transformed. To improve the performance of the model, more epochs to train the algorithm would be needed.

Lastly, we found that there tended to be greater success and less distortion when there was greater colour intensity differentiation between different segments of the photo e.g. between the person's hair or clothes and background. This is likely because the algorithm was more able to distinguish that there was a change in the image component and apply different rules on whether to do a transformation or to leave it unchanged.

6.3 Problem and possible implementation methods

One of the problems we faced was that the professional version photos generated from CycleGAN have lower resolution and quality compare to the original casual pics. There are some algorithms for image-to-image translation can solve this kind of problem, for example, the pix2pix HD. pix2pixHD takes a pyramid approach (Zhu, 2018): First, it returns as output low-resolution pictures. Second, it uses the previously output low-resolution picture as the input to another network, and then generate a higher-resolution picture. It also can be applied on face enhancing and face dehazing. We can perform this task of mapping function which transforms input images into enhanced images, if we have a large number of original and enhanced image pairs.

But in many cases, it is difficult to obtain paired data. For example, in our project if we want to develop an application to help people generate professional headshot from his casual headshot, there is no corresponding real photo in real life. Therefore, CycleGAN is our best possible algorithm here. The reason for CycleGAN's success is that it separates style and content. It is difficult to manually design such a separation algorithm, but with a neural network, it is easy for its learners to automatically maintain the content, which means the background and people's face and change the style. From our training example, it can be seen that CycleGAN can find the position of the person's attire more accurately and turn it into a dark tone version in suit and tie.

However, we have some failure cases. For example, as shown in Figure 8, when the input photo is a woman wearing headscarf, the output headshot only changes the colour tone of the headscarf and in some cases convert it to resemble hair but fail to translate the attire to a professional attire. This is because there does not exist a woman wearing scarf in our professional picture domain. The difficulty is that a comprehensive picture dataset is difficult to obtain.

Figure 8. Left: original casual photo of the woman wearing headscarf; Right: professional version generated by our model.

One possible method to separate the influence on the woman's clothing and her headscarf is to use the Mask R-CNN algorithm (He et al., 2017). This algorithm can simultaneously perform object detection and instance segmentation in a network, and is widely used in detecting cars for self-driving cars. So, for a given image, Mask R-CNN, in addition to the class label and bounding box coordinates for each object, will also return the object mask. Its framework is developed upon Faster R-CNN, which first use a ConvNet to extract feature maps from imagines (Sharma, 2020), these features mapped would be passed through a Region Proposal Network (RPN) to generate bounding box. The segmentation mask feature enables the model to segment all the objects in the image. This is the final step in Mask R-CNN where we predict the masks for all the objects in the image.

By implying this technique, we can separate input photo into different boxes, therefore only transfer the dressing part of the photo. There is one remaining concern for this implement algorithm, which is there are overlapping part between the woman's headscarf and her clothes. It may be difficult to make CycleGAN's generator only transfer her clothes.

The large pose discrepancy between training images is one of the key challenges in photo transformation. Most of LinkedIn headshots are frontal facing, however some photos we collected have invariant poses, for example, some photos have extreme profile views, both for their face and body Our current model has not successfully transferred invariant-posing dressing, as seen in Figure 9.

Figure 9. Left: original casual photo of the woman has a profile pose; Right: professional version generated by our model.

One possible future implement is to employ frontalization to synthesize a frontal pose for those photos, then apply our algorithm to change the dressing style. Disentangled Representation Learning-Generative Adversarial Network (DR-GAN) is one novel method which also evolves a two-player game between a generator G and a discriminator D, similar to the CycleGAN model. The input to the encoder is a face image of any pose, the output of the decoder is a synthetic face at a target pose, and the learnt representation bridges encoder and decoder. This technique is widely used by law enforcement practitioners to identify suspects from CCTV. However, currently DR-GAN only attempts to recognize and transfer faces in variant angle, whether it can be applied on body angle transferring is still unknown.

7. GitHub Reference

Our codes and a representative set of our data is stored on GitHub and can be accessed via the following link: https://github.com/wong-fu-jie/casual2professional. Translations of group member's profiles using the best evaluated model is also available.

The generated photos from the scraped LinkedIn profile photos are however not uploaded onto GitHub in view of privacy considerations.

8. Conclusion and Future Opportunities

This report presents an image-to-image translation method: CycleGAN to demonstrate a style transfer for profile photos. It involves generating a synthetic version of a given casual image with a specific modification. Without pairing data, CycleGAN is trained in an unsupervised manner using a collection of images and translate them into a professional version which is suitable to be used as a LinkedIn profile photo.

Our model is built upon GAN architecture using unpaired collections of images from two different domains, casual and professional. Aiming for improving model performance, we set strict rules on selecting the input photo and use several techniques to further perform argumentations to expand dataset effectively. Our final model learns the professional dressing style from the professional data domain and apply the dressing code to another casual domain. It successfully transfers casual dressing like t-shirts into suit-and-tie after 200 epochs training, which achieves our initial objective.

In general, we have found that a CycleGAN-based algorithm is suited for such a domain-to-domain photo translation or transformation between domains (casual and professional in this case). The success of the algorithm depends heavily on the training set that needs to have good quality with diverse data with little noise (cluttered background in our case). It is however anticipated that with further advances in machine learning methods, such a photo translation application can be improved and made available for use.

In the meantime, our algorithm can potentially be applied to users who have simpler photo translation needs. For the same context of attire transformation, fashion retail companies could adopt such an algorithm to help customers see how a new design would look on them without physically trying on the outfits. The algorithm could be applied in other context or domains that involves changing one part of the image e.g. product repackaging.

It should however be acknowledged that similar to other photo generation applications using machine learning, there would be ethical concerns and other risks if such machine learning methods are misused. For example, as the algorithm improves in generating more authentic-looking photos, there is the risk of the photos being used for improper purposes e.g. masquerading as persons in uniformed-based services such as police or military.

References

Abbot, L. (2019, August 5). *10 Tips for Picking the Right LinkedIn Profile Picture*. LinkedIn. https://business.linkedin.com/talent-solutions/blog/2014/12/5-tips-for-picking-the-right-linkedin-profile-picture

About LinkedIn. (n.d.). https://about.linkedin.com/.

Brownlee, J. (2019, July 12). *18 Impressive Applications of Generative Adversarial Networks (GANs)*. https://machinelearningmastery.com/. https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. (eds.). *Proceedings of the International Conference on Neural Information Processing Systems* (NIPS 2014). pp. 2672–2680. Red Hook, NY: Curran Associates, Inc, 2014. Ghahramani, Z, Welling, M, Cortes, C, et al.

He, K., Gkioxari, G., Dollár P., and Girshick, R. Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV 2017)*, pp. 2961–2969, Venice, Italy, October 2017. IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Houyi Collector. (2019). [Data crawling application]. KuaiYi Tech. http://www.houyicaiji.com/

Jalan, A. (2017, March 14). *LinkedIn Profile Photo Tips: Introducing Photo Filters and Editing*. LinkedIn. https://blog.linkedin.com/2017/march/14/linkedin-profile-photo-tips-introducing-photo-filters-and-editing

Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *Proceedings of International Conference on Learning Representations (ICLR)*, 2018.

Ma, L., Jia, X., Sun, Q., Schiele, B., Tuytelaars, T., and. Van Gool, L. Pose guided person image generation. *Advances in Neural Information Processing Systems (NIPS)*, pp. 406-416. Research Collection School of Computing and Information Systems, 2017.

Meero.com. (2020, July 16). *11 Dos & Dont's for your LinkedIn profile picture in 2020*. Meero.Com. https://www.meero.com/en/news/corporate/411/11-Tips-To-Follow-For-The-Perfect-Linkedin-Profile-Picture-In-2019

Perarnau, G., Weijer, J.V., Raducanu, B., and Álvarez, J.M. Invertible Conditional GANs for image editing. *NIPS Workshop on Adversarial Training*, 2016

Sharma, P. (2020, November 28). *Computer Vision Tutorial: Implementing Mask R-CNN for Image Segmentation (with Python Code)*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/

Snell, J., Ridgeway, K., Liao, R., Roads, B. D., Mozer, M. C., and Zemel, R. S. Learning to generate images with perceptual similarity metrics. *2017 IEEE International Conference on Image Processing, (ICIP 2017)*. pp. 4277-4281, IEEE.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4):600–612, 2004.

Xue, W., Zhang, L., Mou, X., and Bovik, A. C. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Transactions on Image Processing* 23(2): 684–695, 2014.

Zhang, L., Zhang, L., Mou, X., and Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378-2386, 2011.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

Read GAN, pix2pix, CycleGAN and pix2pixHD in one article. (2018). ProgrammerSought. https://www.programmersought.com/article/61574161988/