Solving Privacy Issues in Imposter Detection Training with AI Generated Artificial Face Images

Costas D. Koufogazos, Jesse D. Flint
Research Associate I, Senior Research Associate III
Orlando, Florida
costas.koufogazos@designinteractive.net
jesse.flint@designinteractive.net

Nick Brawand
AI Researcher
Orlando, Florida
nbrawand@anl.gov

ABSTRACT

Organizations across DoD and DHS use training to develop the visual search skills necessary to compare facial features of individuals to identification document (ID) photos and determine if the individual is an imposter. This task presents a challenge, as individuals may differ in age and appearance compared to the provided ID photo. Training to effectively gain these visual search skills is typically accomplished through instructor-led presentations in a classroom setting, with instructors describing the techniques and highlighting the critical visual cues (e.g., facial features) needed to perform the task. However, few training platforms currently have a means to visualize performance during the task and obtain objective feedback of how well they visually interrogated critical cues. Creating training platforms to address these issues is challenging due to the lack of images that are (1) photorealistic, (2) appropriately different across ID and individual photos, and (3) publicly available for use in a training platform. Advances in Generative Adversarial Network (GAN) Artificial Intelligence (AI) algorithms allow for the creation of photorealistic artificial face images that can (1) avoid privacy issues and (2) allow for artificial aging and other methods to make the images look more appropriately different. Advantages to using GAN artificial face images also include the ability to update an image database on a continual basis and the ability to more closely control the demographics of the images used. However, the act of creating and changing face images results in challenges associated with validating that pairs of face images are the "same person" or an "imposter". The current paper examines the issues associated with creating a database of facial images appropriate for imposter detection training, the methodology utilized to create a set of training images, and the process used to validate the images before being implemented.

ABOUT THE AUTHORS

Costas D. Koufogazos is a Research Associate I at Design Interactive in Orlando, Florida. He holds an M.S. Degree in Human Factors from Embry-Riddle Aeronautical University and a B.S. Degree in Health Sciences from the University of Central Florida. He has been in his current role on the Performance Augmentation team at Design Interactive for 1.5 years.

Jesse D. Flint is a Senior Research Associate III at Design Interactive in Orlando, Florida with over a decade of experience in cognitive psychology research. Mr. Flint's work focuses on human factors research and technology capability design for Department of Homeland Security components such as the Federal Law Enforcement Training Center (FLETC).

Nick Brawand is a postdoc in theoretical physics at Argonne. His research focuses on understanding the fundamental forces of nature that bind nuclei. He has worked as a data scientist at Disney and Design Interactive.

Solving Privacy Issues in Imposter Detection Training with AI Generated Artificial Face Images

Costas D. Koufogazos, Jesse D. Flint
Research Associate I, Senior Research Associate III
Orlando, Florida
costas.koufogazos@designinteractive.net
jesse.flint@designinteractive.net

Nick Brawand
AI Researcher
Orlando, Florida
nbrawand@anl.gov

INTRODUCTION TO IMPOSTER DETECTION TRAINING

Experts in imposter detection training use cues within different facial features to compare individuals to an identification document (ID). This can be a challenging task to learn, as individuals may differ significantly from their appearance in an ID photo. Individuals may be older, have a different hair color or hairstyle, have gained or lost weight, etc. The cues that experts look for in these scenarios are traditionally taught in a classroom setting with instructor-led presentations. There are training platforms that can visualize the process of imposter detection. An example of this visualization is through eye-tracking, which obtains objective feedback on how individuals are investigating facial identification cues. Customs and Border Protection Officers have recently utilized one of these technologies for training imposter detection. This led to increased performance at the task, with an increase in their ability to correctly identify imposters and a decrease in the amount of time it took to identify imposters (Department of Homeland Security Science & Technology Directorate, 2020).

With these training platforms, comes a unique set of challenges when selecting face images for training. Using a real person's face requires the photo to not have any privacy restrictions. Coupled with the need for a diverse, expanding database creates a roadblock for this type of training long-term. There are challenges involved when selecting photos for imposter detection training. Public availability and diversity of the images must be taken into consideration during the creation of imposter detection training materials, as these issues coincide. Due to privacy concerns, training content creators cannot choose any photos they find to use in the training content. They must ensure the images selected are available for public use and will not cause any legal issues. Additionally, they must be able to gather an image pool diverse enough to serve the purpose of the training. A sample of images too small will not provide enough reinforcement or training content for continued training. When creating imposter detection training content, the content creators must find two images of the same person, in different environments, with slightly altered appearances to pass off as a cleared pair. They must also find images of two different individuals but must keep it similar enough to fool someone if they are not carefully looking at the proper cues. These challenges make this process an inefficient and difficult way to train imposter detection.

For the study in this paper, fake face images were created to be used in place of real face images to solve the challenges of using real face images for imposter detection training. A general adversarial network (GAN) was leveraged to create images for the purpose of imposter detection training. Specifically, a StyleGAN was used, which is an expansion on the traditional GAN. Before the images are used in an imposter detection training platform, they must be validated by experts. Attempted validation of the images was done by measuring accuracy on an imposter detection assessment, using the created StyleGAN images. This assessment was given to both experts with experience in ID validation, and novices with no experience in ID validation.

METHODOLOGY FOR CREATION AND VALIDATION OF ARTIFICAL FACE IMAGES

The fake faces were generated using NVIDIA's StyleGAN ([Links to versions of StyleGAN], n.d.). StyleGAN is a type of GAN that is trained to generate realistic-looking images using large datasets of faces. A GAN is composed of two neural networks, one is called a generator and the other is a discriminator. These networks are tasked with determining if the image is real or not. These two networks train against each other to the point where the discriminator cannot determine if the image is fake (Khan, 2019).

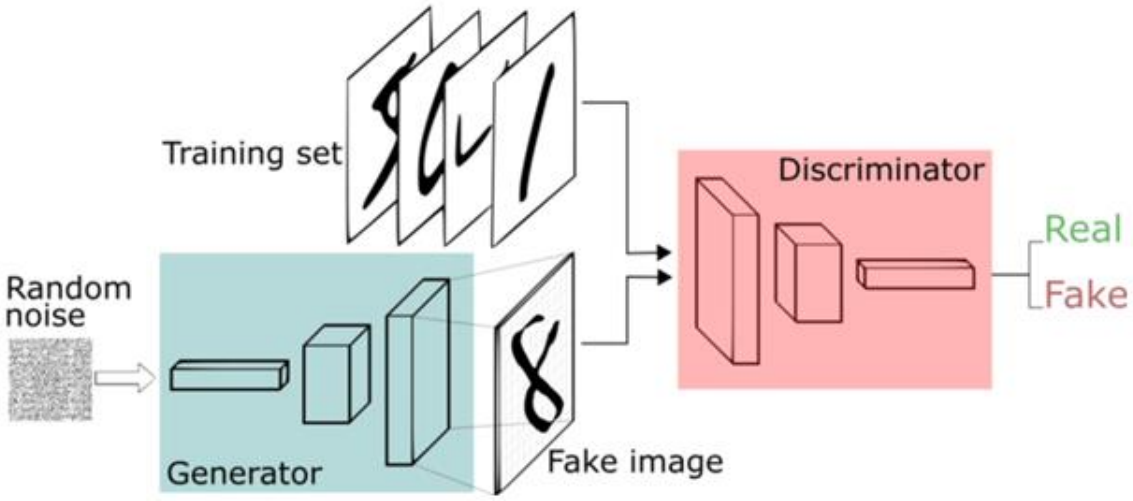


Figure 1. Basic GAN Framework (Silva, 2018)

The StyleGAN takes the features of the tradition GAN and expands on them, allowing for finer creation of images. A traditional GAN generates one image, while the StyleGAN generates a large set of images with different characteristics (Karras, 2019). Facial features of generated images are controlled by moving along preidentified directions in the latent space of the neural network which corresponds to different high-level features such as sex and age. NVIDIA has multiple versions of StyleGAN available (https://nvlabs.github.io/stylegan2/versions.html). For this study, StyleGAN 2018 was utilized to generate the fake face images. The tool used to edit the faces was built by a user on github, spiorf. Spiorf identified the directions to manipulate the facial features, added the sliders and code using NVIDIA's official code release. This resulted in sets of artificially generated faces (Spiorf, 2019). Please see the references for more details on the technical aspects of GANs and StyleGAN.

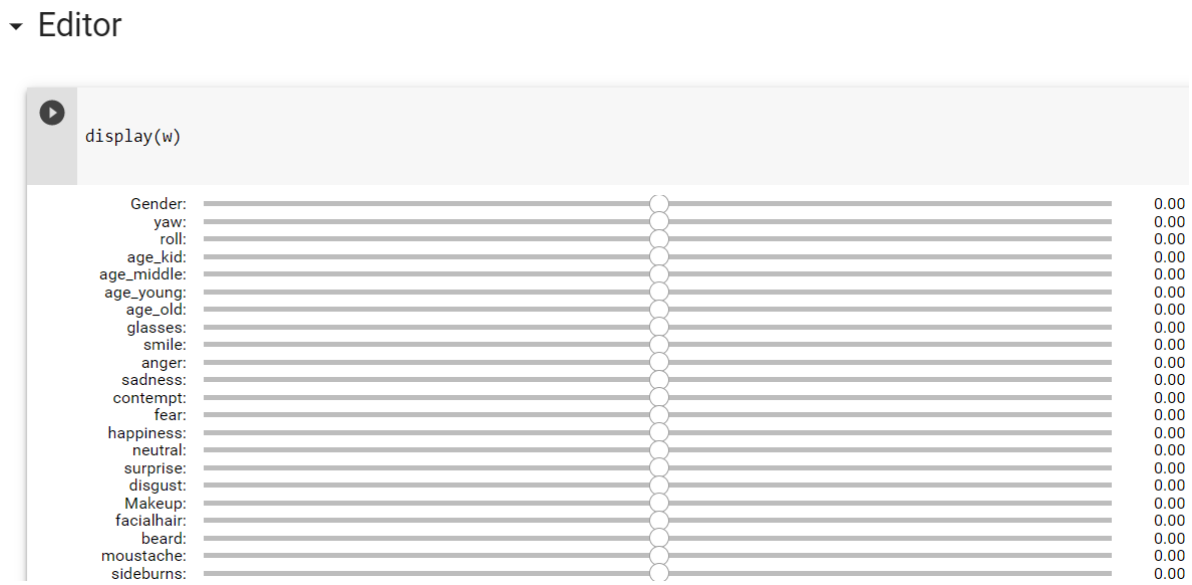


Figure 2. Face Editing Tool (Spiorf, 2019)

Image Creation Process

The above explained a brief technical aspect of how a GAN works, the following will explain the process the user followed to generate images for training. The process used to generate the fake face images was created inside

Google Colaboratory by an artificial intelligence (AI) expert, using the code from NVIDIA StyleGAN along with the sliders generated from spiorf.

The first step for the user is to find any free-to-use image of someone's face. Once the user has their raw image, they insert it into the program. The program will return a generated image based on the raw image that was uploaded. Once the user advances the program further, they reach the step to edit the generated image via the sliders.

Once the image is available beside the sliders, it can be manipulated. If the user wished to create a cleared pair, they saved the original generated image, then moved a slider such as age to create an older version of the generated image. Other sliders were manipulated to create imposter pairs, such as gender or various emotions. See the example below of how a generated facial image is aged (Figure 3).

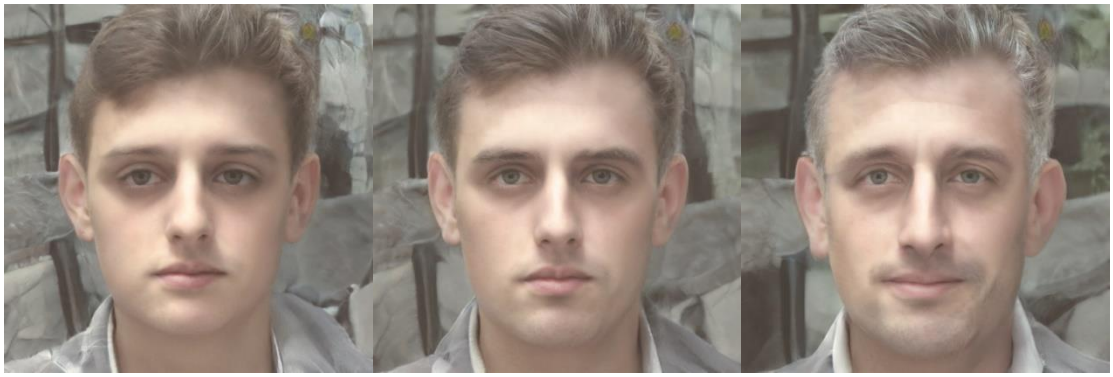


Figure 3. Aging a Face

Once generated and saved, the images underwent background edits to eliminate noise behind the faces. (See Figure 3 for an example of background noise, and Figures 4 and 5 for an example of the edited backgrounds). The photos were then presented side by side as a pair. On the left was the "ID" photo, or a photo with a plain background as you would find on a person's ID. The photo on the right included a "checkpoint" background, or a background you would expect to see in an area where ID verification is performed.

Figure 4. Male Pair with Edited Backgrounds

Figure 5. Female Pair with Edited Backgrounds

Using Artificially Generated Images to Address Challenges

This process can be used to solve the previously mentioned challenges of using real face images for imposter detection training. By using NVIDIA StyleGAN, the challenges of diversity and privacy of images is eliminated. Using this process, a user can manipulate any base generated image to match their needs. This allows for an indefinite amount of faces that can be used for imposter detection training. This alleviates the eventual problem of repetitive training content and lack of diversity across face pairs. A user can generate new faces any time they wish to update their imposter detection training database. Since the images are generated inside a neural network, they are not real, and are not subject to any privacy restrictions. An early concern was the lack of photorealism that would come out of artificially generating face images. As seen above, this concern was mitigated, and the photos generated displayed photorealism.

Validation of Face Images

To use the generated images on a training platform, they must first be validated. This is to ensure the created "imposter" and "clear" pairs are able to meet the training objectives, based on imposter detection performance from experts and novices. There are several criteria that were used for creating the images for the validation process. Several sliders were identified by the user as being most relevant to the creation of imposter and clear pairs. The gender slider, which made the person appear either more masculine or feminine, appeared to alter the overall facial structure of the person. The makeup slider, and various emotion sliders had a similar effect. These were the primary sliders used in creating imposters since it appeared to create a different looking person, but similar enough to try and pass off as an imposter. Sliders such as age, facial hair, smile, altered the look of a person but did not seem to alter their facial structure. This allowed for the creation of pairs of the same person, but slightly aged, a different smile, or the inclusion/absence of facial hair (for males). This was done at the discretion of the user, the images created did not follow any guidelines from experts, but rather on the judgement of the user who generated the images.

The experts chosen to evaluate the face image pairs were individuals who have experience in law enforcement and have validated IDs in their career. The novices selected were non-law enforcement and have never performed ID validation before. Both experts and novices received a PowerPoint slide deck with the created imposter and clear pairs of images in a randomized order. Before starting the task, experts were asked about their experience validating IDs. They were law enforcement with varying years of experience (13, 20, and 24 years). Participants were instructed to look through the entire slide deck and mark down whether each pair was cleared or if the person pictured on the right was an imposter. The responses were marked on a corresponding excel spreadsheet. In addition, participants marked how confident they felt in marking each image pair.

For the images to be considered in a training context, the experts would need to correctly identify pairs as imposters or cleared. The novices would be expected to perform poorly on this task, indicating they would need further

training to correctly identify the imposter and cleared pairs. This would designate that the faces generated have the cues, or lack thereof, that experts use to identify an imposter.

RESULTS

The results to the study yielded a return from 3 experts and 10 novices. Three measurements were taken to evaluate performance of experts and novices on identifying faces to be either imposter or cleared pairs: percentage correct, confidence, and inter-rater reliability. Percentage correct identified how many of the pairs were marked correctly by the participant. The average percentage correct across experts (n=3) was 65%. The average percentage correct across novices (n=10) was 66%. See the charts below for distribution of the individual scores

Table 1: Percentage Correct Across Experts

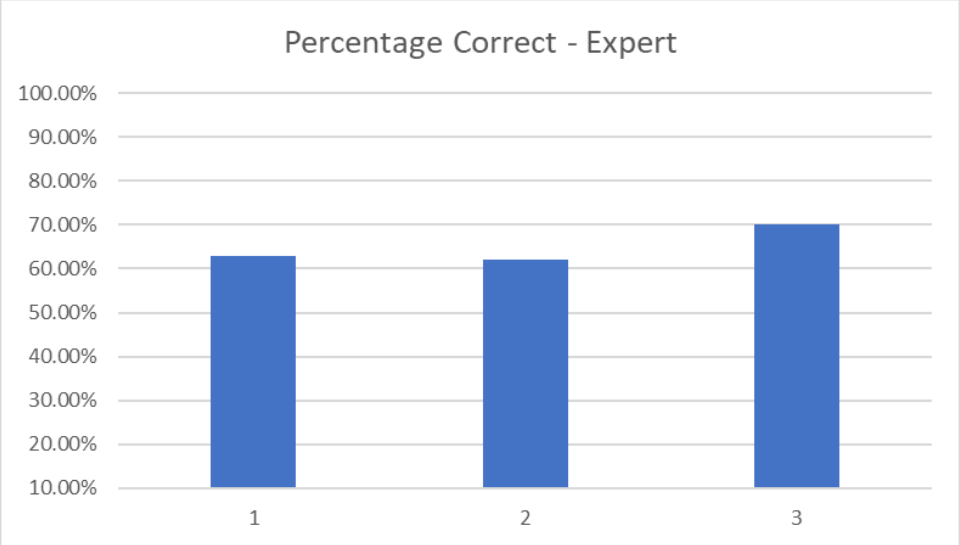
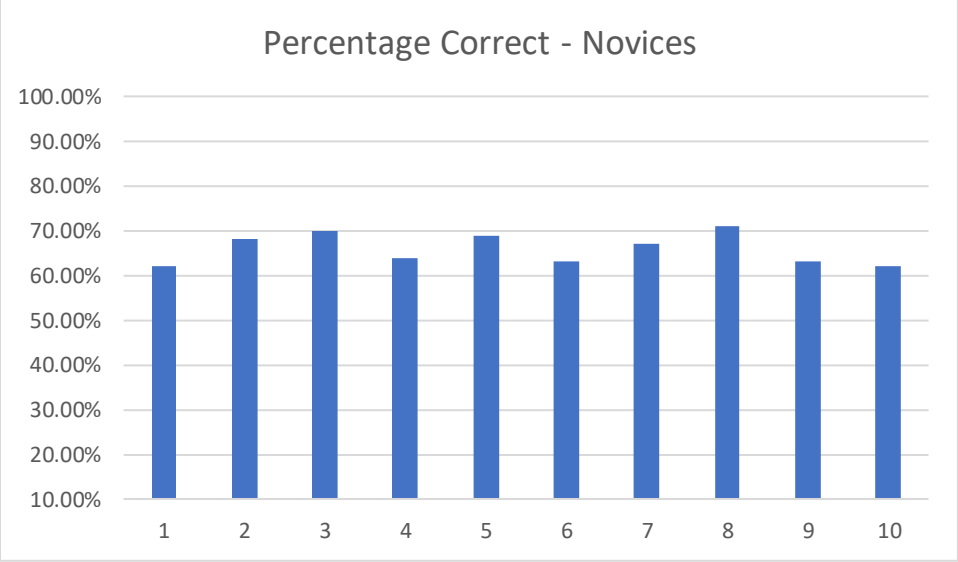


Table 2: Percentage Correct Across Novices



After marking each pair as imposter or cleared, each participant rated the confidence they felt in that choice on a scale of 1-5. Experts reported an average of 2.95 out of 5 confidence rating and novices reported an average of 3.94 out of 5 confidence rating. One unique comment received from an expert stated this was very difficult and in their line of work, they would have done fingerprint scans on many of these people to verify the identity.

Inter-rater reliability was calculated to determine the agreement amongst each group of participants. Fleiss' kappa was chosen for its ability to calculate agreement between two or more raters (Laerd Statistics, 2019). Fleiss' kappa showed that there was fair agreement between the experts, $\kappa = .282$. It also showed fair agreement between the novices, $\kappa = .229$.

DISCUSSION

The results indicate that the imposter pairs were difficult to identify for both experts and novices. Due to this, these images could not be used in an imposter detection training platform without further knowledge and studies on this process. On average, experts correctly identified 65% of the pairs, and novices correctly identified 66% of the pairs. Without further research, it is difficult to determine why the experts did not score at a much higher rate than the novices. However, several hypotheses can be made as to why this happened. Experts in imposter detection look for specific facial cues. If the generated image did not contain a specific cue that they were looking for, they may have marked it as an imposter, even though it was generated to be a cleared pair. This potentially answers why the experts reported lower confidence ratings as well. Novices with no training do not know what specific facial cues to look for when performing imposter detection, and most likely felt more confident with wrong choices because of this. Experts, on the other hand look for very specific facial cues. If they did not see those in the generated images, it was difficult for them to determine if an image was an imposter.

A low kappa between the experts, coupled with similar percentage correct across all 3 suggests they could be using different perceptual cues to make their decisions. They did not have much agreement on which photos they determined to be imposters, but in the end still reported similar results in overall percentage correct. However, without further research into what perceptual cues experts look for it is difficult to tell why this phenomenon is happening. Another possible reason for low agreement between the experts may have been due to the training they received and the years of experience they have. The three experts reported 13, 20, and 24 years of experience. It is also likely they came from different training experiences, where one may have received more training on specific facial cues over others. If the images generated did not change those specific cues, the expert easily may have perceived a generated imposter pair as cleared. The novices, who showed a slightly lower kappa score may have been guessing on many of the image pairs or fooled by the similarity of most of the pairs.

While the objective was to create a similar looking face for an imposter, it must also contain the cues that experts look for in imposter detection. The potential hinderance of privacy issues and lack of diversity and size of a training database are eliminated by following a process for generating "fake faces." Since a process has been determined to efficiently create face images, it can be leveraged in future studies in the field of imposter detection training. There are possible explanations as to why the results from this small study came out the way they did, and future research options should be suggested. The first and crucial step in this would be to identify the specific features that subject matter experts (SMEs) look for when performing ID validation. This could be done through an ideal observer analysis. An ideal observer analysis is a way to determine the optimal way to assess a task, using specific physical properties of the environment and stimuli. Using an ideal observer model, ideal performance at a task can be determined using the precise measure of physical properties. This creates a benchmark to compare and train performance (Geisler, 2003). In the context of imposter detection training, this could be done for the specific features identified by SMEs. An example of a facial feature that could be used in this ideal observer model would be the distance between the eyes. The distance between the eyes on each image of the image pair would be measured, and the difference found. If the difference between the eye gaps in the pair is very small, then it would be identified as a cleared pair. If the difference between the eye gaps in the pair is large, it would be identified as an imposter. This could be done for each important feature identified by SMEs. The results could be compared with actual user responses to see if they match with any of the ideal observer models. From there, an AI expert could further manipulate the StyleGAN code to allow for specific measurements of various facial features that the content creators could integrate in the image generation process. This would make the process of generating images more precise, as the user could generate very specific features that an expert is looking for when validating an ID image.

There were some limitations in this study. Lack of access to an abundance of ID validation experts resulted in a small sample size. In this study, ten experts were contacted, but only three produced results in the timeframe required for the submission of this paper. Since there were few experts who participated in the study, there was a limited amount of information on participant training background and how that could have affected scores. Due to the small sample size, the results are underpowered for generalizable findings. Future studies would involve expert input during the image creation process, and a larger sample size of experts to participate in the validation process.

CONCLUSION

The method of generating face images with GANs solves several key challenges that traditional imposter detection training is hindered by. The results of the study performed at the end of the face generation process did not show that the experts were able to correctly identify the imposter and clear pairs at a higher rate than the novices. Therefore, the images used for this study should not be used in imposter detection training. This creates a foundation for further research into methods to improve imposter detection training via artificial face generation. Leveraging what was learned from this study and its limitations, future steps include further research into cues that experts look for when performing imposter detection tasks. More experts should be involved in both the creation of artificial face images and the validation task. Information from experts about specific facial cues used in imposter detection could be implemented and translated by an AI expert into a GAN/StyleGAN to create a custom program to manipulate specific cues that experts look for. As GANs and StyleGANs evolve and become more complex, they continue to show potential to be used in imposter detection training. However, in the current state as used in the study, they do not meet the cut for use in imposter detection training.

REFERENCES

DHS Science & Technology. (2020). *Snapshot: CBP Officers Leverage S&T-Developed Imposter Detection Training Tech*. Department of Homeland Security. https://www.dhs.gov/science-and-technology/news/2020/02/25/snapshot-cbp-officers-leverage-st-developed-imposter-detection-training-tech

Geisler, W. S. (2003). Ideal observer analysis. *The visual neurosciences*, *10*(7), 12-12.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4401-4410). https://openaccess.thecvf.com/content_CVPR_2019/papers/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.pdf

Khan, J. (2019). *StyleGANs: Use machine learning to generate and customize realistic images*. Heartbeat. https://heartbeat.fritz.ai/stylegans-use-machine-learning-to-generate-and-customize-realistic-images-c943388dc672

Laerd Statistics (2019). Fleiss' kappa using SPSS Statistics. *Statistical tutorials and software guides*. https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php

[Links to versions of StyleGAN]. (n.d.). https://nvlabs.github.io/stylegan2/versions.html

Spiorf. (2019). *Spiorf/stylegan-encoder*. GitHub. https://github.com/spiorf/stylegan-encoder