CONTROLLING THE MEMORABILITY OF REAL AND UNREAL FACE IMAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Everyday, we are bombarded with many face photographs, whether on social media, on television, or smartphones. From evolutionary perspective, faces are intended to be remembered, mainly due to survival and personal relevance. However, all these faces do not have the equal opportunity to stick in our minds. It has been shown that memorability is an intrinsic feature of an image but yet, it's largely unknown what attributes make the images more memorable. In this work, we aim to address this question by proposing a fast approach to modify and control the memorability of face images. In our proposed method, we first find a hyperplane in the latent space of StyleGAN to separate high and low memorable images. We then modify the image memorability (while keeping the identity and other facial features such as age, emotion, etc.) by moving in the positive or negative direction of this hyperplane normal vector. We further analyzed how different layers of the StyleGAN augmented latent space contribute to face memorability. These analysis showed how each individual face attribute makes images more or less memorable. Most importantly, we evaluated our proposed method for both real and unreal (generated) face images. The proposed method successfully modifies and controls the memorability of real human faces as well as unreal (generated) faces. Our proposed method can be employed in photograph editing applications for social media, learning aids, or advertisement purposes.

1 INTRODUCTION

In our everyday life, we are exposed to many pictures of scenes, objects or faces. But recent research have shown that they do not have the same likelihood to be recalled later (Isola et al., 2011; Bainbridge et al., 2013; Isola et al., 2014). Although different people have different abilities in memorizing visual contents (image or video), it has been shown that memorability is an intrinsic feature of an image and it is consistent across different observers (Isola et al., 2011; Bainbridge et al., 2013; Isola et al., 2014). In other words, memorability of an image is an attribute of that image which can be measured, predicted or manipulated (Isola et al., 2011). So far there have been several studies on understanding and predicting image memorability (Khosla et al., 2013a; 2015; Goetschalckx & Wagemans, 2019; Fajtl et al.; Needell & Bainbridge, 2021; Squalli-Houssaini et al., 2018; Almog et al.), and a few work attempted to modify image memorability (Khosla et al., 2013b; Siarohin et al., 2017; Sidorov, 2019; Goetschalckx et al., 2019). However, for practical applications (e.g. education, advertisement), it is most important to have methods for memorability modifications. Moreover, such approaches will help understanding what constitute as image memorability, i.e. "What makes an image memorable?".

This work proposes a new framework based on Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) for modifying face memorability as a facial attribute. Older approaches on modifying face memorability manually used face features (Khosla et al., 2013b), such as SIFT (Lowe, 2004), HOG2x2 (Dalal & Triggs, 2005), and Local Binary Pattern (LBP) (Ojala et al., 2002). Recently, just a few work used generative adversarial networks for memorability modification (Goetschalckx et al., 2019; Sidorov, 2019). Goetschalckx et al. (2019) leverage latent vector modification to change the memorability of the fake food, scenes, and animal images generated by BigGAN (Brock et al., 2018). This memorability modification affects several attributes of the image, such as size, color, and shape. In our work, we aim to modify memorability of face images of real people while keeping their identity, consequently, their method cannot be used here.

The largest dataset of the human faces with their memorability annotations is US 10k Face (Bainbridge et al., 2013), which includes 2222 face images with their memorability scores acquired from human observers in an experiment. Training deep neural network based demands for require a large dataset. For this, we employed pre-trained StyleGAN1 (Karras et al., 2019) and StyleGAN2 (Karras et al., 2020) on the FFHQ dataset (Karras et al., 2019) to generate 100k fake faces. Next, we adopted computational memorability models which we trained on the US 10k Face Database, to predict the memorability of the generated face images and organized them into high-memorable and low-memorable faces. Inspired by Shen et al. (2020), we found a hyperplane in latent subspace to separate the high-memorable and low-memorable faces. We showed that both latent space and extended latent space can be used for finding the separating hyperplane. After finding the hyperplane, we can move the latent vector of each image, in the positive or negative direction of the normal vector of that hyperplane and change the distance of the latent vector from the separating hyperplane to manipulate the memorability of that image. We call this hyperplane's normal vector, memorability modification vector. With this proposed approach, we can control the amount of memorability change by using different weights for the memorability modification vector. In contrast to the method proposed by Sidorov (2019), our method do not require training another auxiliary network for modifying face memorability. Furthermore, the amount of change in memorability is controlled by a hyperparameter in our method.

Since different hyperplanes for different facial attributes in styleGAN latent space (Shen et al., 2020; Härkönen et al., 2020) can be found, with our method we can modify the memorability of the images conditionally. For example, we are able to change the memorability of the face while fixing the length of the hair and the existence of eyeglasses. For this, we first find the corresponding hyperplanes for these attributes, and then leverage projection to those subspaces to have the desired attributes fixed while changing memorability. StyleGAN produces high quality real-looking images which are near impossible to differentiate from real images. To make sure our memorability-modified faces still look real, we considered the Frechet Inception Distance (FID) (Heusel et al., 2017) and Kernel Inception Distance (KID) (Bińkowski et al., 2018) scores of the generated faces from the StyleGAN as the baseline. Then by calculating the FID score and KID score of our memorability modified images, we showed that our modified faces still look real.

For real faces, we first embedded the face images into GAN latent space using Image2Style (Abdal et al., 2019) and the method provided by Karras et al. (2020). After finding the image latent vector, we modified the real face memorability in a similar way explained for synthetic faces. Figure 1 illustrates the general idea of our approach. Finally, we examined how different layers of extended latent space of StyleGAN affect the face image and the face memorability.

1.1 Related works

Image Memorability. People have different capabilities in memorizing different visual events (Hunt et al., 1981). In spite of these differences, through a series of experiments, Isola et al. (2011) showed that people consistently remember some images and forget others. They designed an online memory game experiment and recruited a large number of participants through Amazon Mechanical Turk. In this experiment, participants observed a series of images presented in a sequence and their task was to detect repetitive images in the sequence. Then Isola et al. (2011) measured a memorability score of each image, which is the rate of people remembering an image after single exposure to that image in the sequence. Khosla et al. (2015) created the largest annotated image memorability dataset (LaMem), which consists of 60,000 images, mostly objects, scenes, and animals. Using this dataset, Khosla et al. (2015) introduced the first deep model for predicting image memorability. This model uses AlexNet (Krizhevsky et al., 2012) as its backbone architecture. Moreover, Needell & Bainbridge (2021) introduced new architectures based on residual networks (He et al., 2016) to improve the performance of memorability prediction. In addition to these models, Fajtl et al. leveraged Attention Maps to introduce AMNet for predicting memorability.

Generative Adversarial Networks (GANs). With the development of Generative Adversarial Networks (Goodfellow et al., 2014), now we are capable of generating real-looking synthetic images that are indistinguishable from real images. Generally, these networks are composed of two parts; a generative and a discriminative network. The goal of the generative network is to generate real-looking images to fool the discriminator and the goal of the discriminator is to learn to distinguish generated images from real images. These two networks are optimized through a minmax game

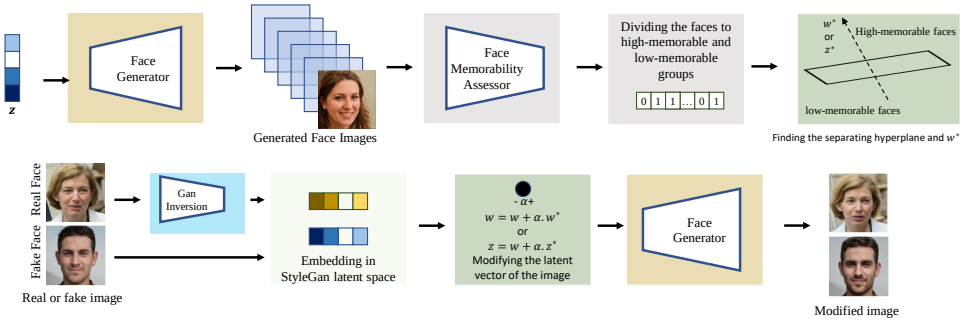


Figure 1: **The proposed method.** We first generate synthesized face images from random latent vectors and predict their memorability scores by a face memorability assessor network. Then, divide them into high-memorable and low-memorable faces. Using either latent vector or extended latent vector subspace we find a memorability separating hyperplane in that subspace. The second row shows the proposed framework for modifying the memorability score of face images. Using a GAN inversion technique we first map faces to StyleGAN's latent space. Then modify the face memorability by moving the latent vector (or extended latent vector) towards negative or positive direction of memorability discriminating hyperplane obtained in previous step. The modified latent vector (or extended latent vector) is fed to the GAN to generate the face image with modified memorability. It is optional to ovalize the face images before feeding to the assessor.

when competing to reach their specified goals. In recent years, there have been huge improvements in this area and many different GANs have been introduced to produce natural-looking images such as Progressive GAN (Karras et al., 2017), DCGAN (Radford et al., 2015), and CycleGAN (Zhu et al., 2017). In this work, we used pre-trained StyleGAN1 (Karras et al., 2019) and StyleGAN2 (Karras et al., 2020) on the FFHQ dataset (Karras et al., 2019) to generate real-looking faces.

Modifying Image Memorability. While image memorability modification has many potential applications (e.g. in education or advertisement), just a few works have focused on it. Khosla et al. (2013b) proposed the pioneer method for changing face memorability. It leveraged Active Appearance Models (AAMs (Cootes et al., 2001)) to represent faces by their shape and appearance. Then the loss function was defined based on the cost of modifying identity, modifying facial attributes, and memorability. As the result, in their method, the identity was fixed. Another approach (Sidorov, 2019) used VAE/GAN (Larsen et al., 2016), StarGan (Choi et al., 2018), and AttGan (He et al., 2019) and trained them with three memorability levels (poorly memorable, moderately memorable, and highly memorable) of faces and modified the memorability of different faces to these three levels only. Siarohin et al. (2017) with another approach utilized style transfer to increase the memorability of an input image. However, the added style adversely effects the realness of the modified images, such that it barely can be used in real-world applications. Most recently, Goetschalckx et al. (2019) trained a transformer network to change the memorability of each generated image by BigGAN (Brock et al., 2018) through modifying their latent vectors. Their method works on generated images of objects and scenes.

2 Method

The overview of our proposed method is depicted schematically in Figure 1. Below we explain the approach step by step.

2.1 Creating the Dataset

For the purpose of analyzing the latent vectors of the GANs and their relation to memorability, we need a large dataset of face images with their memorability scores. The largest dataset available for face images is the 10k US Adult Faces Database (Bainbridge et al., 2013). This database contains 10,168 natural human face images and 2,222 of these images are annotated with memorability

scores. To create a larger dataset for face images with their corresponding memorability scores, we leveraged StyleGAN1 and StyleGAN2, which are the state-of-the-art models for creating realistic-looking face images. These models were pre-trained on the FFHQ dataset (Karras et al., 2019) which consists of 70,000 high-quality face images with 1024×1024 resolution with variations in age, gender, and glasses. We created two different datasets with StyleGAN1 and StyleGAN2. We randomly sampled 100k $\boldsymbol{z} \in \mathbb{R}^{1\times512}$ from a standard normal distribution with truncation to produce high-quality synthetic face images and saved their mappings in the extended latent space ($\mathbb{R}^{18\times512}$) of both GANs.

2.2 Preprocessing step

The generated images from the StyleGAN have 1024×1024 resolution in three channels. For the purpose of acquiring the memorability of the generated images, we had to preprocess them to predict their memorability with the computational memorability predicting models (assessor). We leveraged VGG16 (Simonyan & Zisserman, 2014), ResNet50 (He et al., 2016), and SENet50 (Hu et al., 2018) that are pre-trained on VGGFace2 (Cao et al., 2018) and fine-tuned them on the US 10k Face Database to correctly estimate face memorability scores. This dataset consists of oval-shaped human faces with white backgrounds. 10k US Adult Face images have the same height of 256 with different widths. Hence, we had to first ovalize the generated faces, then compute their memorability scores. We used MTCNN (Zhang et al., 2016) for detecting the face in the generated images and masked an oval on it to make all the images similar to the US 10k Face Database (See Figure 2).

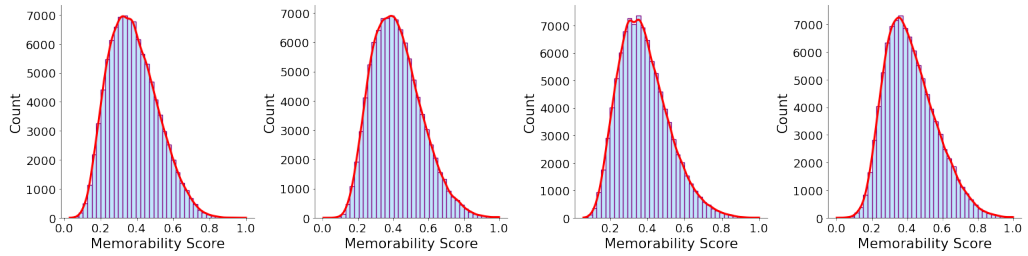
Figure 2: Preprocessing step to make the shape of the synthesized images similar to the dataset that the assessor is trained on.

2.3 Predicting the memorability scores

As explained in previous section, memorability assessors are trained on 10K US face database which includes ovalized faces. We first, tested whether the ovalization process is necessary to obtain memorability scores for generated faces. For this, we fine-tuned pre-trained SENet50 (Hu et al., 2018), VGG16 (Simonyan & Zisserman, 2014), and ResNet50 (He et al., 2016) for predicting face memorability scores on 10K US face database. We calculated the memorability of both the oval-shaped and square-shaped faces generated from StyleGAN1 and StyleGAN2. The distributions of the memorability scores using SENet with oval and squared faces are shown in Figure 3. As can be seen, the memorability distributions are very similar for oval and square faces. We further calculated the Kendall rank correlation (Schaeffer & Levitt, 1956) and Spearman's rank correlation for the paired oval-shaped and square-shaped images. We observed that SeNet50 and VGG16 have a high-rank score (see Table 1). It's an interesting result which shows these models can be used for the square-shaped face images with some backgrounds too. The benefit of using square-shaped faces is that we can have more control on the person's hair. Please note that the performance of memorability predicting models are measured with rank correlation (Khosla et al., 2015).

2.4 The proposed method for modifying face memorability

For the next step, we aimed to find a hyperplane to separate the high memorable and low-memorable images. First, we needed to label the faces into high-memorable and low-memorable faces. For this, we used the mean of memorability scores as the threshold. We labeled an image as high or low memorable if its memorability score was higher or lower than the mean. In our experiments,



(a) StyleGan1 (oval) (b) StyleGan1 (square) (c) StyleGan2 (oval) (d) StyleGan2 (square)

Figure 3: Distributions of the memorability scores of the generated images of StyleGan1 and StyleGan2 based on using an ovalization step or not.

Table 1: Correlations of memorability scores of square-shaped and oval-shaped faces. High correlation score suggests the assessor performs well on square faces too.

Assessor	**Kendall Tau correlation**	**Spearman's correlation**
ResNet50	0.1606	0.2391
SENet50	0.4419	0.6217
VGG16	0.4720	0.6562

we also tried using median of the memorability scores as the threshold for labeling high and low memorable groups. Next, using logistic regression, we attempted to find a hyperplane to separate low-memorable and high-memorable images. For this purpose, we used either $\boldsymbol{z} \in \mathbb{R}^{1\times512}$ latent vectors or $\boldsymbol{w} \in \mathbb{R}^{18\times512}$ extended latent space of stylegan1 and stylegan2. After finding the separating hyperplane, we can move the latent vector or the augmented latent vector in the positive or negative direction of the normal vector of that hyperplane to control the image memorability. This hyperplane denotes the moderately memorable images, as we chose it to separate the faces into high-memorable and low-memorable faces based on the mean (or median) of the memorability scores. Memorability of each image is related to the distance of its latent vector (or extended latent vector if extended latent vector is used to find the hyperplane) from this hyperplane. Consider image i with corresponding latent vector $\boldsymbol{z}_i \in \mathbb{R}^{1\times512}$ or $\boldsymbol{w}_i \in \mathbb{R}^{18\times512}$, and its memorability score mem_i. We note the normal vector of this hyperplane by $\boldsymbol{z}^* \in \mathbb{R}^{1\times512}$ or $\boldsymbol{w}^* \in \mathbb{R}^{18\times512}$ and we show the distance from the hyperplane by function d. We will have:

$$mem_i \propto d(\boldsymbol{z}^*, \boldsymbol{z}_i) = \boldsymbol{z}^{*^T}.\boldsymbol{z}_i \tag{1}$$

$$mem_i \propto d(\boldsymbol{w}^*, \boldsymbol{w}_i) = \boldsymbol{w}^{*^T}.\boldsymbol{w}_i \tag{2}$$

Hence, we can change the memorability of each image, by changing the distance of its latent vector (or extended latent vector) from the separating hyperplane; $\boldsymbol{z}_{edit} = \boldsymbol{z} + \alpha\boldsymbol{z}^*$ or similarly $\boldsymbol{w}_{edit} = \boldsymbol{w} + \alpha\boldsymbol{w}^*$ and we will have:

$$d(\boldsymbol{z}^*, \boldsymbol{z}_{edit}) = \boldsymbol{z}^{*^T}.\boldsymbol{z}_{edit} = \boldsymbol{z}^{*^T}.(\boldsymbol{z} + \alpha\boldsymbol{z}^*) = \boldsymbol{z}^{*^T}.\boldsymbol{z} + \alpha = d(\boldsymbol{z}^*, \boldsymbol{z}) + \alpha \tag{3}$$

$$d(\boldsymbol{w}^*, \boldsymbol{w}_{edit}) = \boldsymbol{w}^{*^T}.\boldsymbol{w}_{edit} = \boldsymbol{w}^{*^T}.(\boldsymbol{w} + \alpha\boldsymbol{w}^*) = \boldsymbol{w}^{*^T}.\boldsymbol{w} + \alpha = d(\boldsymbol{w}^*, \boldsymbol{w}) + \alpha \tag{4}$$

The results of this classification task for finding the separating hyperplane are presented in Table 2 when different assessors are used for memorability score predictions. In each case, the accuracy was about 10 percent higher when we used extended latent space (w), hence we showed the results for extended latent space in Table 2. According to the results, we decided to use SENet50 as our assessor and mean memorability as the threshold for our further analysis.

The performance of logistic regression is higher when the extended latent space is used to find the separating hyperplane of low- and high-memorable images. Further, as we show in the next section working with the extended latent space yields better results in modifying face memorability.

Table 2: Accuracy of the separating hyperplane, based on the method for dividing images into high-memorable and low memorable images, the shape of the images, and the assessor.

Assessor	Median		Mean	
	Oval	Square	Oval	Square
ResNet50	0.8131	0.7933	0.8149	0.7928
SENet50	0.8157	0.8291	**0.8207**	**0.8317**
VGG16	0.7938	0.8037	0.7952	0.8071

One of the benefits of our approach is that we can control fixing specific attributes while changing memorability. It has been shown that latent vector play an important role in determining different attributes of a face and we can find hyperplanes to separate faces based on those attributes, such as glasses, age, and smile. Consider the norm vectors of these hyperplanes as $\mathbb{A} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_k\}$. We can fix these attributes by subspace projection as follows: $\boldsymbol{w}^*_{new} = \boldsymbol{w}^* - \sum_{i=1}^{i=k} (\boldsymbol{w}^{*T}.\boldsymbol{a}_i)\boldsymbol{a}_i$

2.5 Latent Vector Recovery for Real faces

The efficiency of our method to modify real human faces depends on how well the latent vector of the real face can be obtained to reconstruct the original image. After acquiring the latent vector of the real face image, we can repeat the process the same as what we did for the synthesized images and modify its memorability. In this work, we used image2style (Abdal et al., 2019) to embed the real faces to latent space of stylegan1, which its loss function consists of a VGG-16 perceptual loss (Johnson et al., 2016) and a pixel-wise MSE loss term. Furthermore, for projecting real faces to stylegan2 latent space, we employed the same algorithm described by Karras et al. (2020) after using the dlib library (King, 2009) to align 68 face landmarks in the preprocessing step. In this work, we assume that we have the latent vector of the generated faces, however, if the latent vector of a generated face is not available, a similar approach can be employed to retrieve the latent vector.

3 Experiments

3.1 Changing memorability of synthesized faces

As described in Section 2.1, we created two different datasets with StyleGAN1 and StyleGAN2. We randomly sampled 100k $\boldsymbol{z} \in \mathbb{R}^{1\times512}$ from a standard normal distribution with truncation to produce high-quality synthetic face images and saved their mappings in $\mathbb{R}^{18\times512}$ augmented latent space of both GANs. We calculated their memorability scores using SENET50, then labeled them high- or low-memorable images when compared to the mean memorability score in the dataset, finally we used the logistic regression to find the separating hyperplane as described in Section 2.4. After finding the separating hyperplane, we evaluated our method in modifying the memorability of generated faces with StyleGAN1 and StyleGAN2 using our proposed method. As we discussed earlier in 2.4, we found separating hyperplanes in either latent space or extended latent space of StyleGAN1 and StyleGAN2. Figure 4 and Figure 5, demonstrates some samples of memorability modification with StyleGAN1 and StyleGAN2 respectively (See A.1 for more examples).

Besides evaluation of our method on plenty of face images qualitatively, we also tested the performance of our method in modifying memorability, quantitatively. For this, we used 10k synthesised faces and tried different weights of memorability modification on them. Figure 6 depicts the distributions of memorability scores when shifted by our method as well as the mean of the distributions.

3.2 Realness of the modified faces

In this experiment, we evaluated the realness of the faces when our memorability modification was applied to them. StyleGAN is a state-of-the-art model in generating real-looking faces. Hence, we considered the realness of the generated images from StyleGAN (before memorability modification) as our baseline and compared the realness of modified faces with it. We utilized two well-known measures for this purpose; FID and KID. As demonstrated in Figure 7, the FID and KID scores of

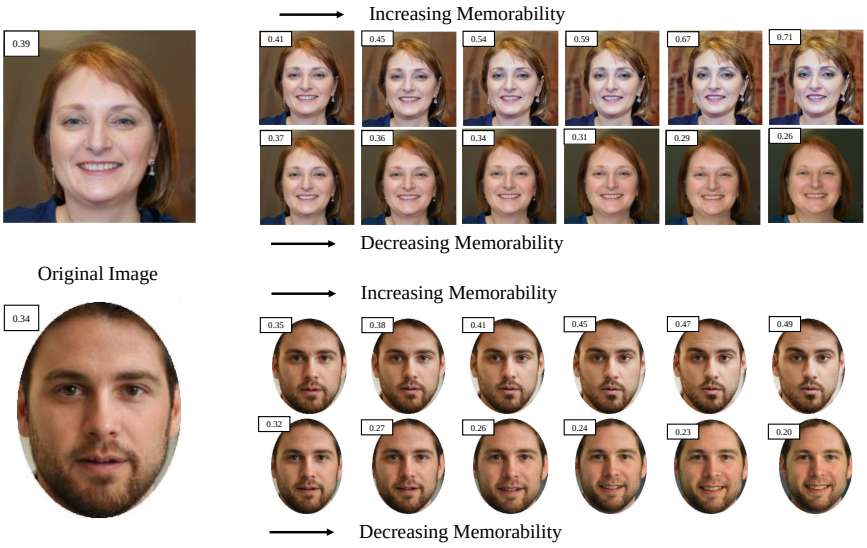


Figure 4: **Modifying memorability of faces generated by StyleGAN1.** Exemplar faces and their counterparts when our memorability modifying approach is applied to them. The second row depicts images when an extra step of ovalization is applied before feeding to the assessor.

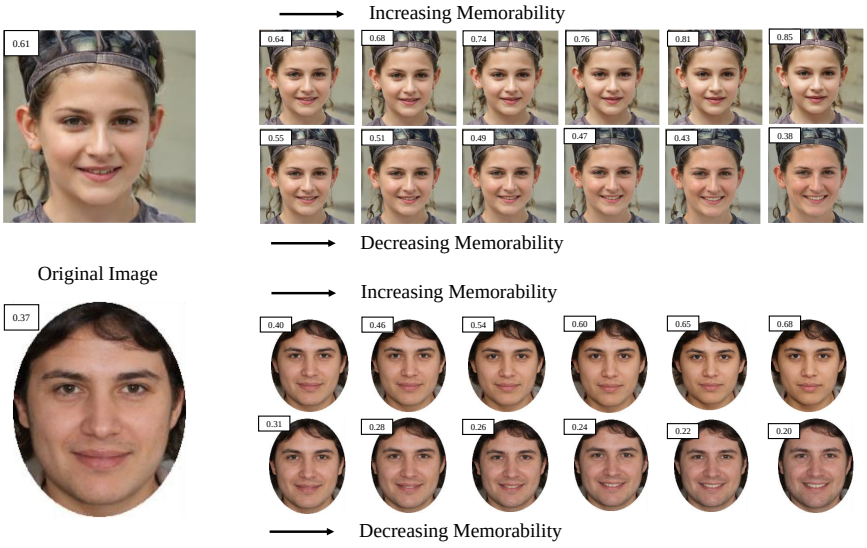
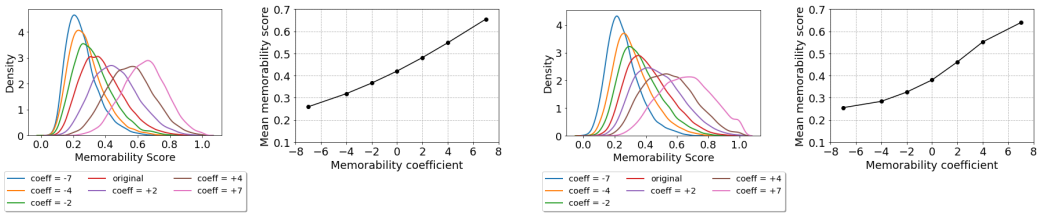


Figure 5: **Modifying memorability of faces generated by StyleGAN2.** Exemplar faces and their counterparts when our memorability modifying approach is applied to them. The second row depicts images when an extra step of ovalization is applied before feeding to the assessor.

the modified images are close to unmodified faces and therefore, the realness of faces is not affected by our method.



Oval-shape faces

Square-shape faces

Figure 6: The effectiveness of our method for modifying memorability scores. As depicted the distribution and mean memorability score of images changes with the coefficient used for memorability modification.

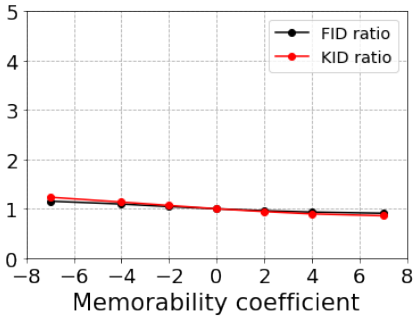


Figure 7: The realness of the generated images measured by the FID and KID ratios for different memorability modification coefficients. FID and KID score of the unmodified images are the respective baselines. A close to one ratio indicates similar level of realness with the baseline.

3.3 Changing memorability of real faces

For the next experiment, we evaluated our method on real human faces. First, we computed their extended latent vector using the GAN inversion method described in the Methods section and then applied our proposed method to them to change their memorability. For real human faces, we chose to work with the extended latent vectors, because the regenerated images from the extended latent space are more similar to the original images (See Figure 8).

3.4 Layerwise memorability modification

For the last experiment, we specifically investigated what are the most contributing layers in extended latent vectors in face memorability, i.e., which layers are most responsible for modifying a face memorability score. In each experiment, we only changed one layer and kept other layers fixed. We plotted the faces after these layerwise changes to examine what kind of changes these layerwise modifications will cause in faces. We observed modifications in the first 11 layers were mostly responsible for the face changes, and the other 7 layers mostly just affected the color and background of the image. Hence, we only focused on the first 11 layers. Then repeated the same layerwise changes on 3k synthesized faces and calculated the mean memorability of these images before layerwise and after layerwise modifications (See Figure 9, for the other layers see A.2).

4 Conclusion

In this work, we proposed a new method to modify the memorability of face images. Our approach does not suffer from the previous methods limitations and it is able to modify the memorability of faces (being synthetic or real) with an arbitrary continuous range. Moreover, we demonstrated

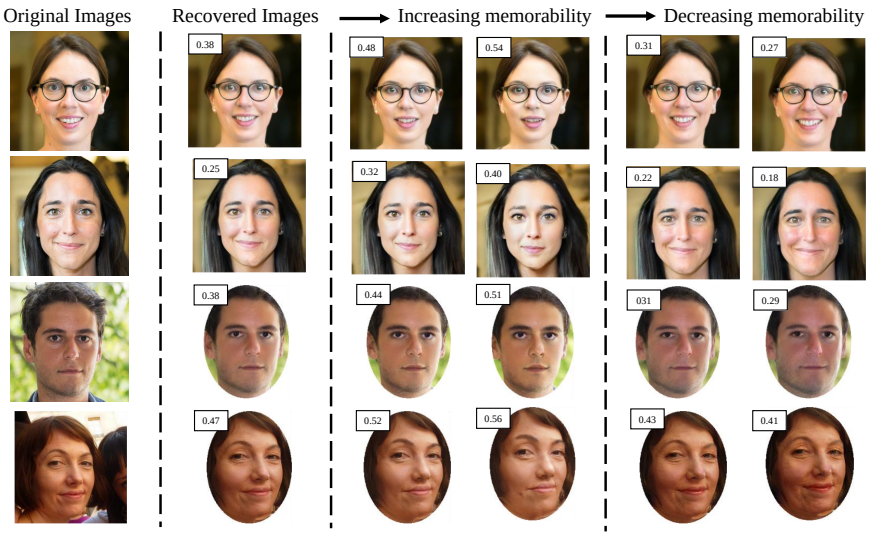


Figure 8: **Memorability modification of real faces.** The first three rows were encoded to StyleGAN2 latent space and the last one was encoded to StyleGAN1 latent space.

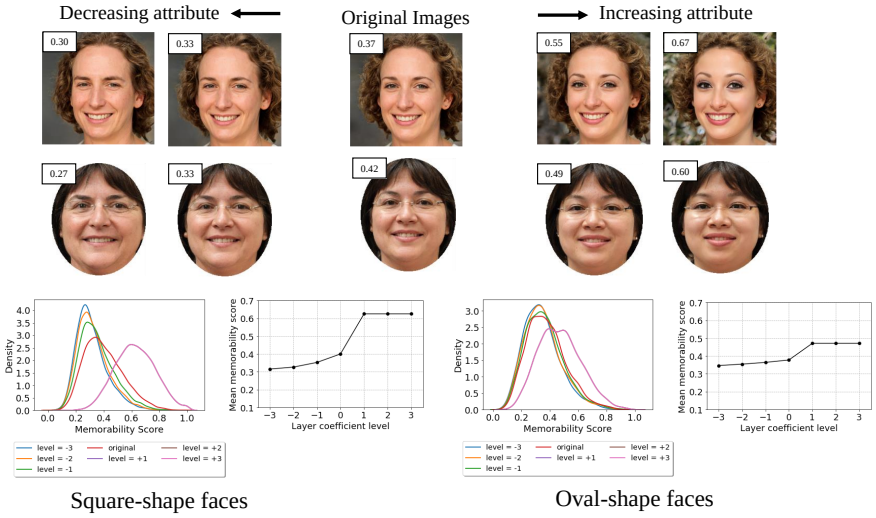


Figure 9: **The effect of changes in the 7th layer of the extended latent space on memorability score.** This layer mostly affects make-ups and facial hair. We observe that when we move this layer in the positive direction of the corresponding layer of the memorability vector (w^*), the memorability score will increase hugely and the person's make-up is highlighted. The lips thickness increases, and the eyebrows shape and the person's skin changes.

these changes will not affect the realness of the images. However, if a very large weight for the memorability modification vector is chosen, it will affect the face identity and realness. We showed that our method is effective by applying it to 10k synthesized faces. Further, we employed our method on real human faces and showed it is effective in changing the memorability of real faces as well. Finally, we studied how layerwise modifications will affect the face and its memorability score and discussed one of the benefits of the proposed method is modifying the memorability score conditionally by leveraging subspace projection method.

REFERENCES

Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4432–4441, 2019.

Gal Almog, Saeid Alavi Naeini, Yu Hu, Emma G Duerden, and Yalda Mohsenzadeh. Memoir dataset: Quantifying image memorability in adolescents.

Wilma A. Bainbridge, Phillip Isola, and A. Oliva. The intrinsic memorability of face photographs. *Journal of experimental psychology. General*, 142 4:1323–34, 2013.

Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018.

Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pp. 67–74. IEEE, 2018.

Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8789–8797, 2018.

Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pp. 886–893. Ieee, 2005.

J Fajtl, V Argyriou, D Monekosso, and P Remagnino. Amnet: Memorability estimation with attention. arxiv 2018. *arXiv preprint arXiv:1804.03115*.

Lore Goetschalckx and Johan Wagemans. Memcat: A new category-based image set to study image memorability. 05 2019. doi: 10.13140/RG.2.2.19105.15202.

Lore Goetschalckx, Alex Andonian, Aude Oliva, and Phillip Isola. Ganalyze: Toward visual definitions of cognitive image properties. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5744–5753, 2019.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. *arXiv preprint arXiv:2004.02546*, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Attgan: Facial attribute editing by only changing what you want. *IEEE transactions on image processing*, 28(11):5464–5478, 2019.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

Earl B Hunt, Janet Davidson, and Marcy Lansman. Individual differences in long-term memory access. *Memory & Cognition*, 9(6):599–608, 1981.

Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. What makes an image memorable? In *CVPR 2011*, pp. 145–152. IEEE, 2011.

Phillip Isola, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva. What makes a photograph memorable? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(7): 1469–1482, 2014.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pp. 694–711. Springer, 2016.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

Tero Karras, Samuli Laine, and Timo Aila. 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, 2019.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8110–8119, 2020.

Aditya Khosla, Wilma Bainbridge, Antonio Torralba, and Aude Oliva. Modifying the memorability of face photographs. pp. 3200–3207, 12 2013a. doi: 10.1109/ICCV.2013.397.

Aditya Khosla, Wilma A Bainbridge, Antonio Torralba, and Aude Oliva. Modifying the memorability of face photographs. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3200–3207, 2013b.

Aditya Khosla, Akhil S. Raju, Antonio Torralba, and Aude Oliva. Understanding and predicting image memorability at a large scale. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 2390–2398, 2015. doi: 10.1109/ICCV.2015.275.

Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10: 1755–1758, 2009.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. In *International conference on machine learning*, pp. 1558–1566. PMLR, 2016.

David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

Coen D Needell and Wilma A Bainbridge. Embracing new techniques in deep learning for estimating image memorability. *arXiv preprint arXiv:2105.10598*, 2021.

Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Maurice S Schaeffer and Eugene E Levitt. Concerning kendall's tau, a nonparametric correlation coefficient. *Psychological Bulletin*, 53(4):338, 1956.

Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE transactions on pattern analysis and machine intelligence*, 2020.

Aliaksandr Siarohin, Gloria Zen, Cveta Majtanovic, Xavier Alameda-Pineda, Elisa Ricci, and Nicu Sebe. How to make an image more memorable? a deep style transfer approach. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pp. 322–329, 2017.

Oleksii Sidorov. Changing the image memorability: From basic photo editing to gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0, 2019.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Hammad Squalli-Houssaini, Ngoc QK Duong, Marquant Gwenaëlle, and Claire-Hélène Demarty. Deep learning for predicting image memorability. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2371–2375. IEEE, 2018.

Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

A APPENDIX

A.1 ADDITIONAL EXAMPLES

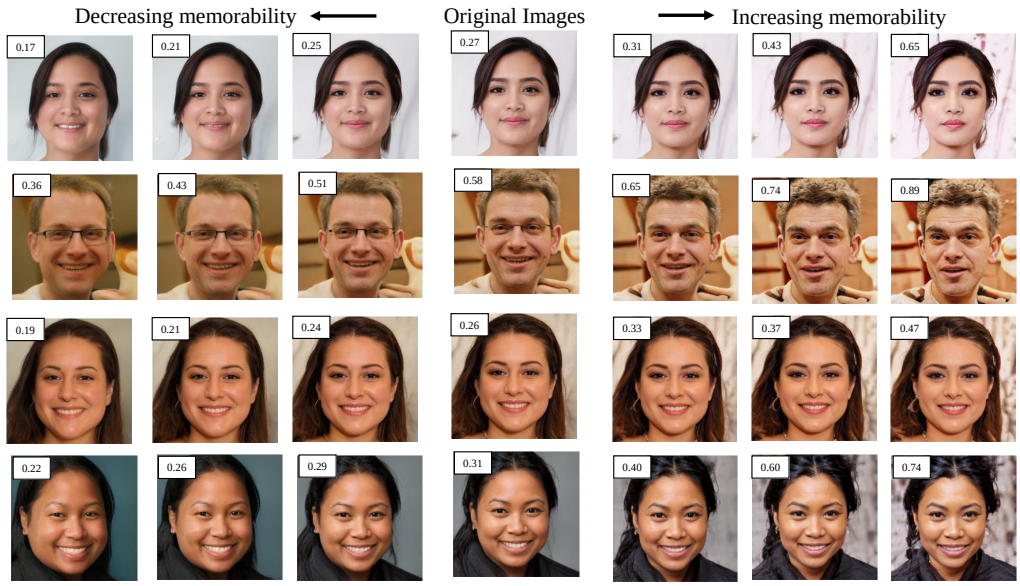


Figure 10: **Generated images by StyleGAN1 with their corresponding memorability scores.** Modified images when extended latent space of the StyleGAN1 was used to determine the separating hyperplane. Square-shaped faces were fed to the assessor.

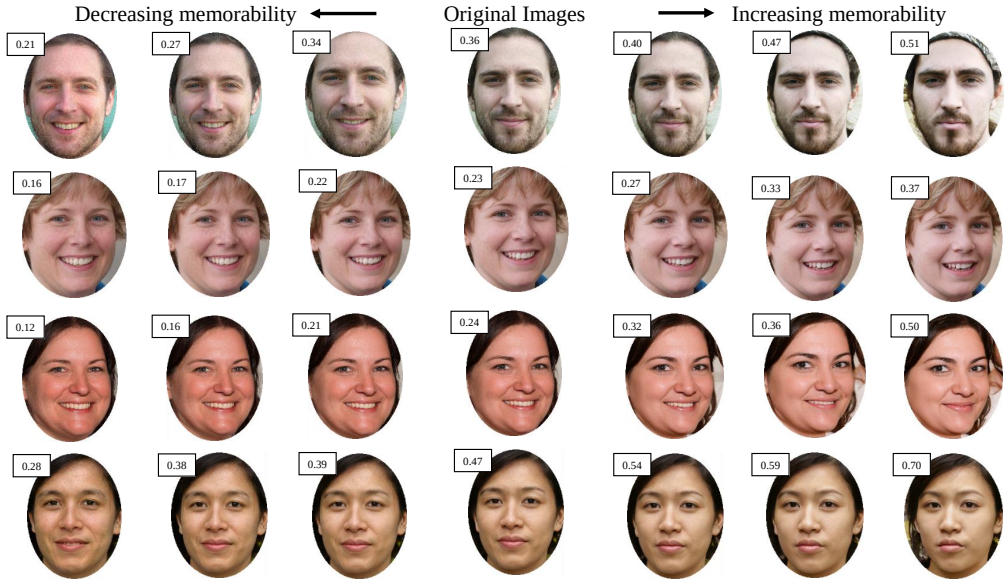


Figure 11: **Generated images by StyleGAN1 with their corresponding memorability scores.** Modified images when extended latent space of the StyleGAN1 was used to determine the separating hyperplane. Oval-shaped faces were fed to the assessor.

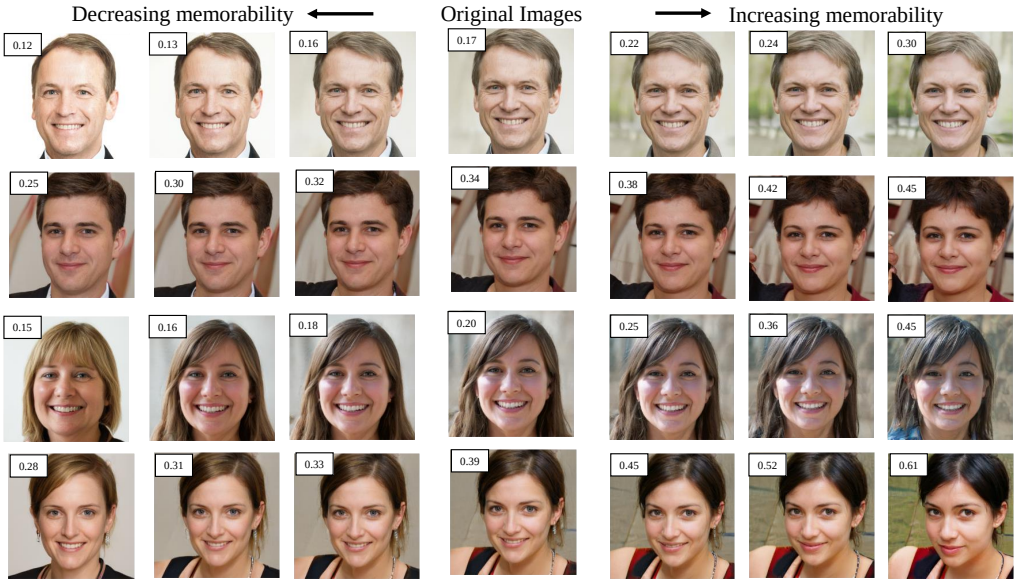


Figure 12: **Generated images by StyleGAN1 with their corresponding memorability scores.** Modified images when latent space of the StyleGAN1 was used to determine the separating hyperplane. Square-shaped faces were fed to the assessor.

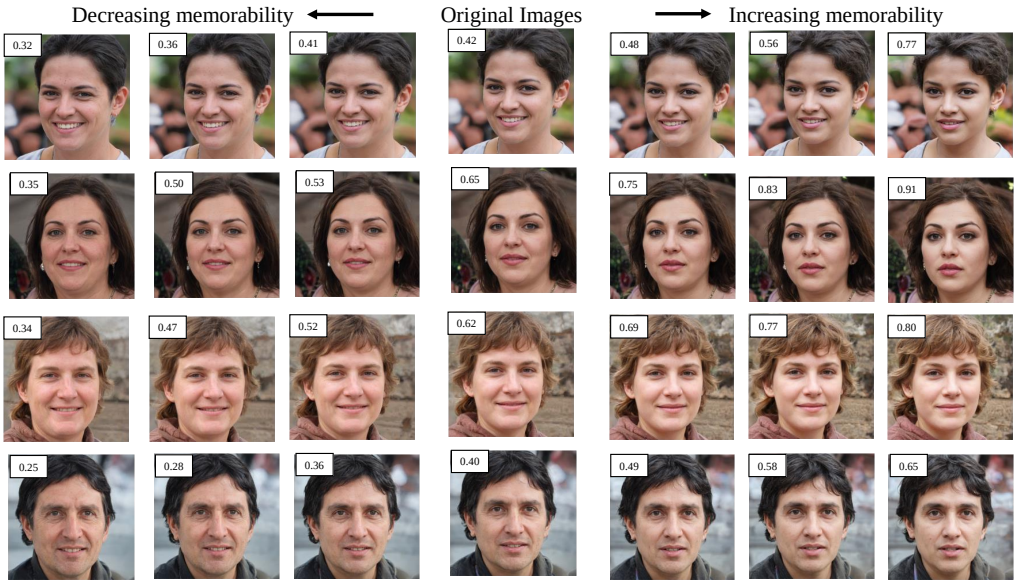


Figure 13: **Generated images by StyleGAN2 with their corresponding memorability scores.** Modified images when extended latent space of the StyleGAN2 was used to determine the separating hyperplane. Square-shaped faces were fed to the assessor.

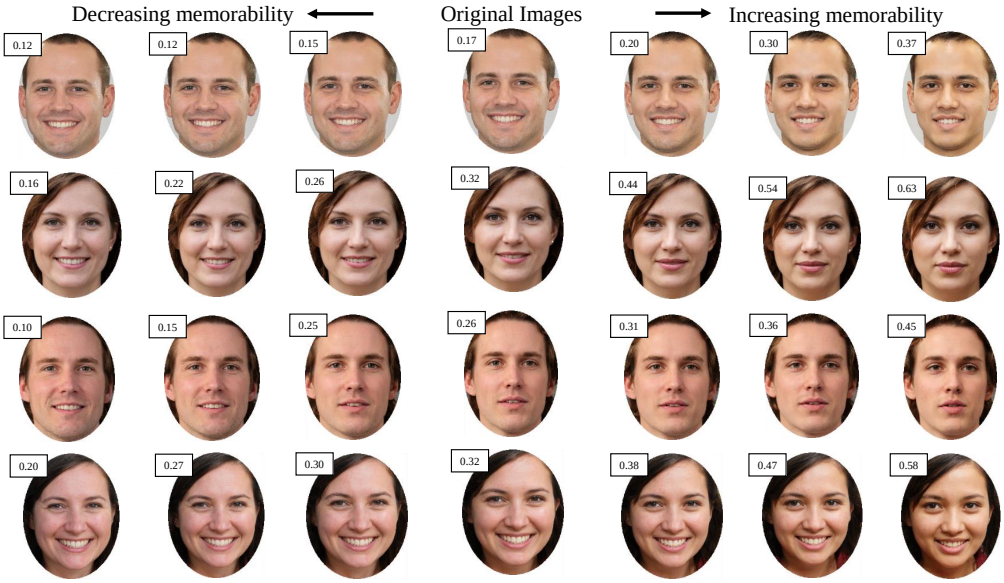


Figure 14: **Generated images by StyleGAN2 with their corresponding memorability scores.** Modified images when extended latent space of the StyleGAN2 was used to determine the separating hyperplane. Oval-shaped faces were fed to the assessor.

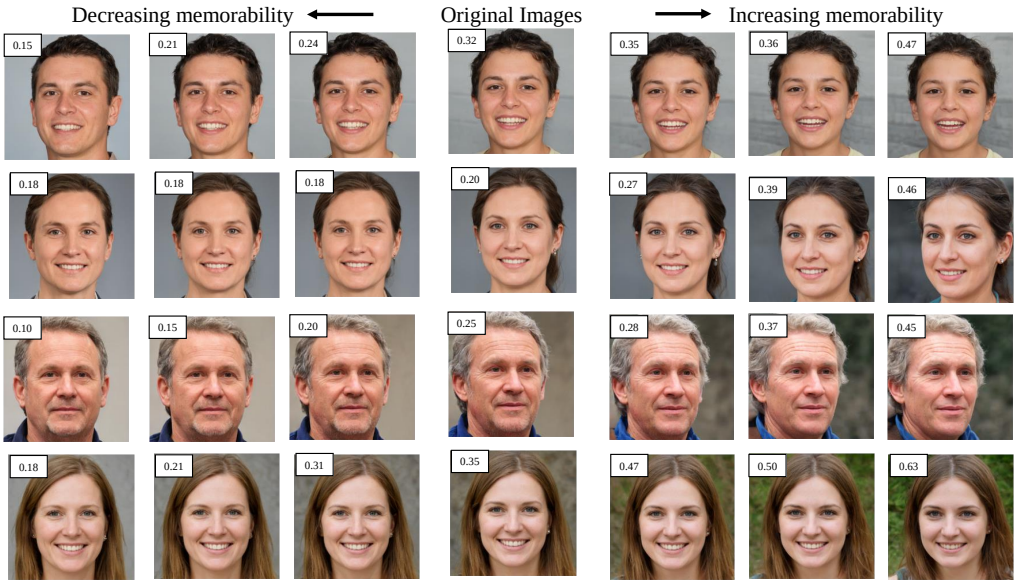


Figure 15: **Generated images by StyleGAN2 with their corresponding memorability scores.** Modified images when latent space of the StyleGAN1 was used to determine the separating hyperplane. Square-shaped faces were fed to the assessor.

A.2 LAYERWISE MODIFICATIONS

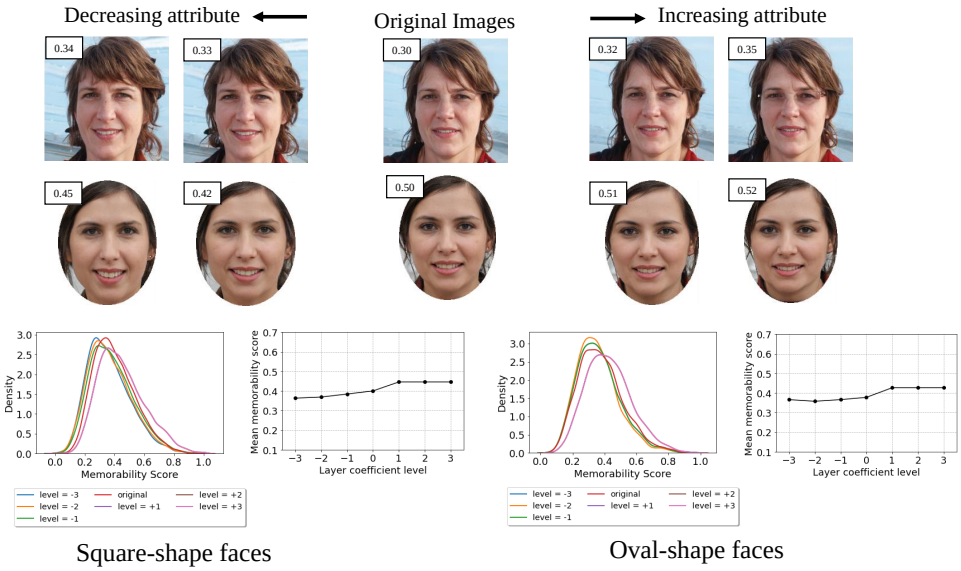


Figure 16: **The effect of the changes in the 1st layer of the extended latent space on memorability score.** This layer mostly affects shape of the face. Increasing this attribute make the face smaller, whereas decreasing this attribute makes the face larger, which usually may cause a decrease in the memorability score.

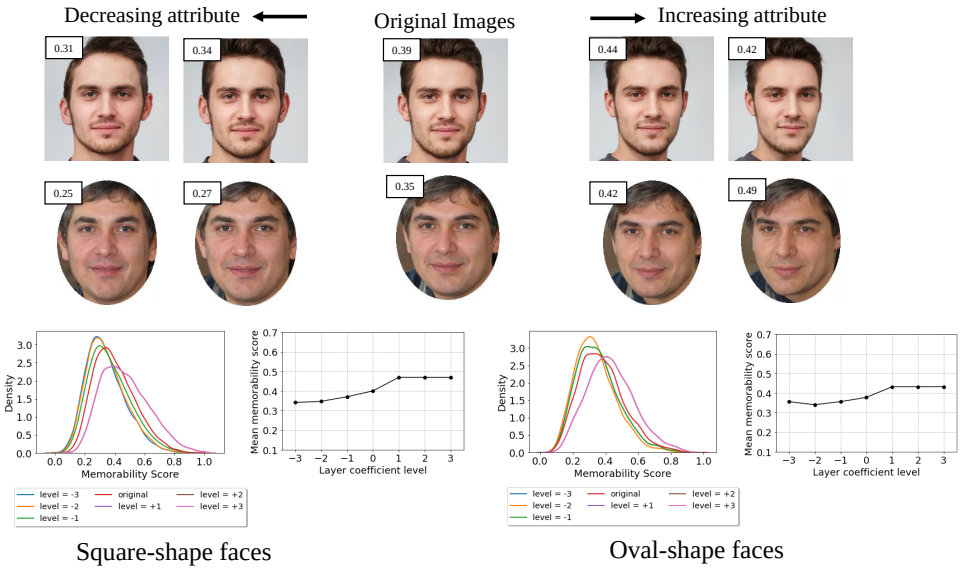


Figure 17: **The effect of the changes in the 2nd layer of the extended latent space on memorability score.** This layer mostly affects hair, pose of the face, and the direction of the eyes and shows how these attributes make contribution to the memorability score.

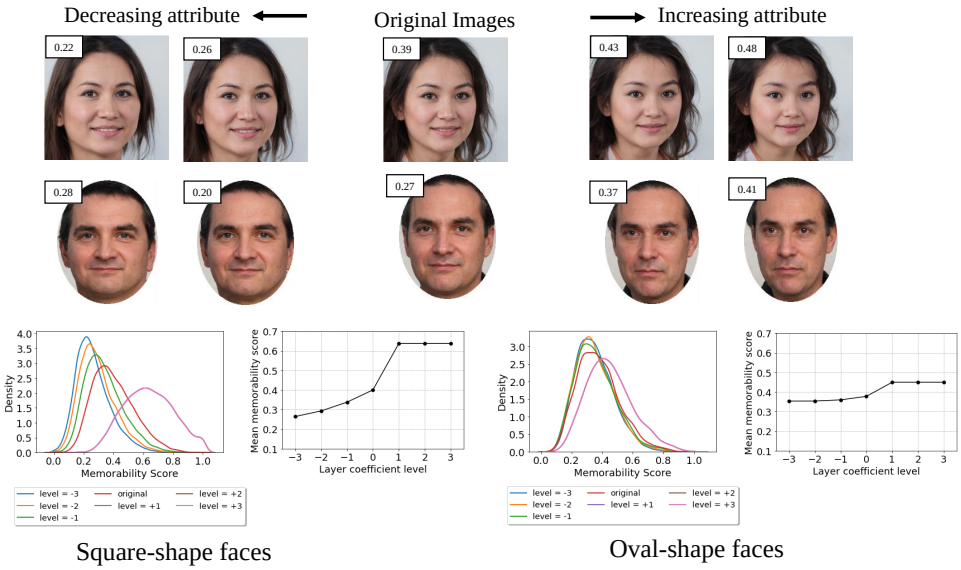


Figure 18: **The effect of the changes in the 3rd layer of the extended latent space on memorability score.** This layer mostly affects shape and seriousness of the faces. We observed that moving this layer, in the positive direction of the corresponding layer in the memorability vector (w^*), will highly increase the memorability score.

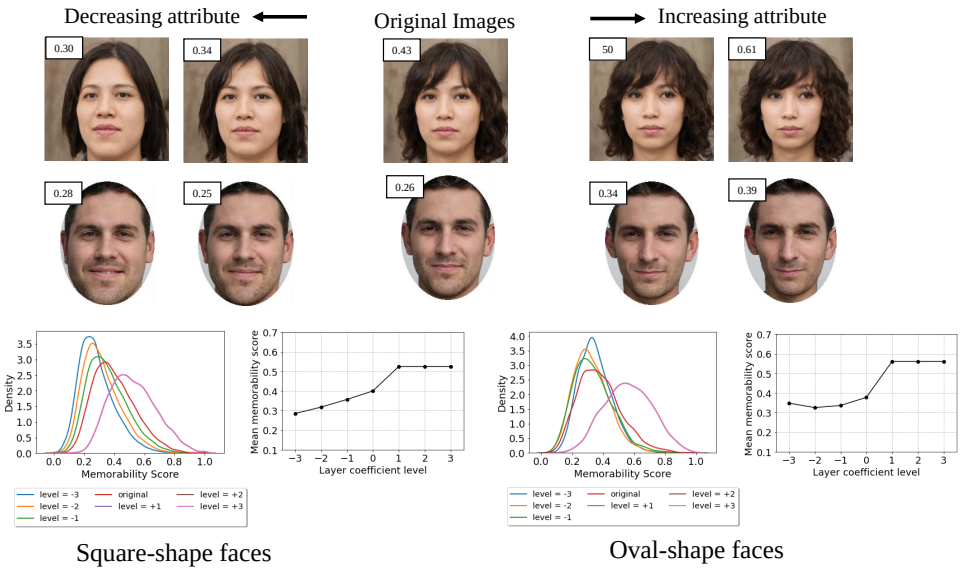


Figure 19: **The effect of the changes in the 4th layer of the extended latent space on memorability score.** This layer mostly affects shape of the face (especially the chin). We can observe that changes in this attribute, largely contribute to the memorability of the face and moving this layer in the positive direction of the corresponding layer in the memorability vector (w^*), will increase the memorability score.

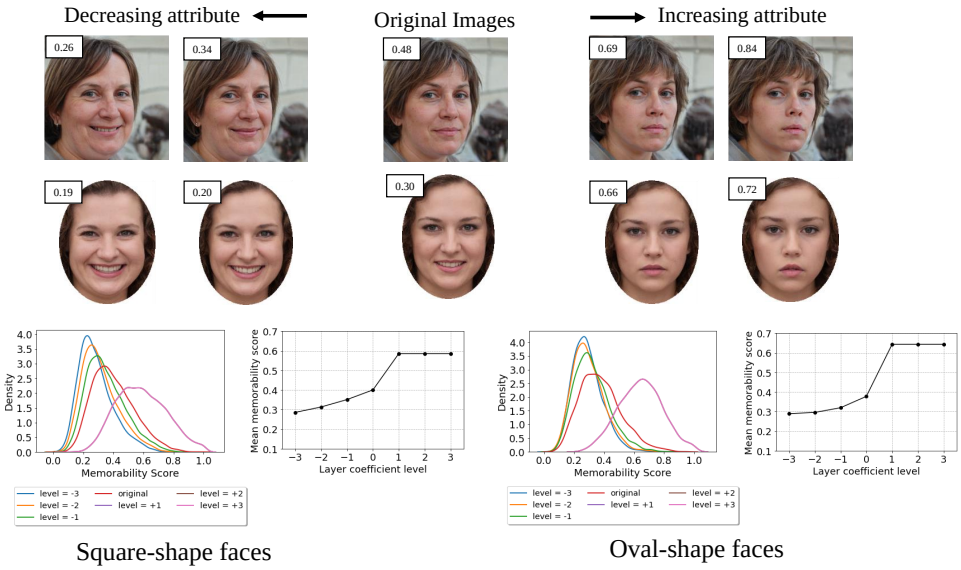


Figure 20: **The effect of the changes in the 5th layer of the extended latent space on memorability score.** This layer mostly affects nose, lips, facial weight, and smile. This is one of the most important layer playing role in determining the memorability score.

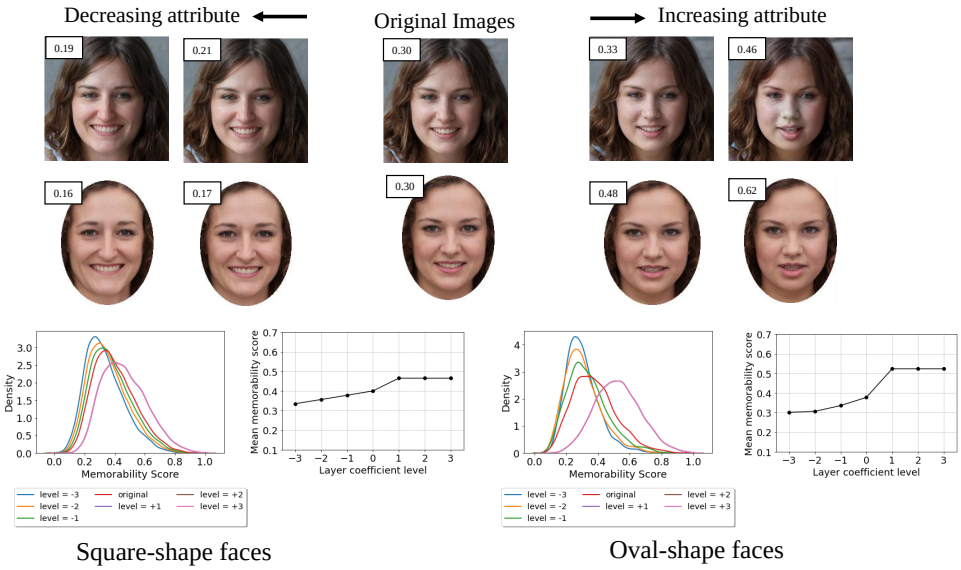


Figure 21: **The effect of the changes in the 6th layer of the extended latent space on memorability score.** This layer mostly affects the smile and form of the lips. We observe that when we move this layer, in the positive direction of the corresponding layer of the memorability vector (w^*), the memorability score will increase hugely and the person's lips will become thicker. However, when you move it in the opposite direction, the person's lips will become thinner and the memorability score will decrease.

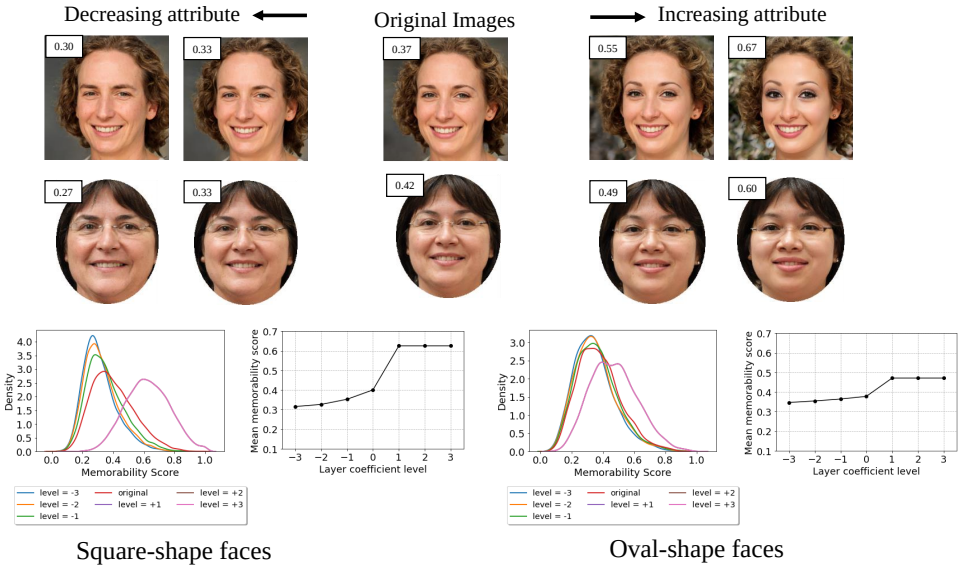


Figure 22: **The effect of the changes in the 7th layer of the extended latent space on memorability score.** This layer mostly affects make-ups and facial hair. We observe that when we move this layer, in the positive direction of the corresponding layer of the memorability vector (w^*), the memorability score will increase hugely and the person will get some makeups. The lips will become thicker and the eyebrows shape and the person's skin changes.

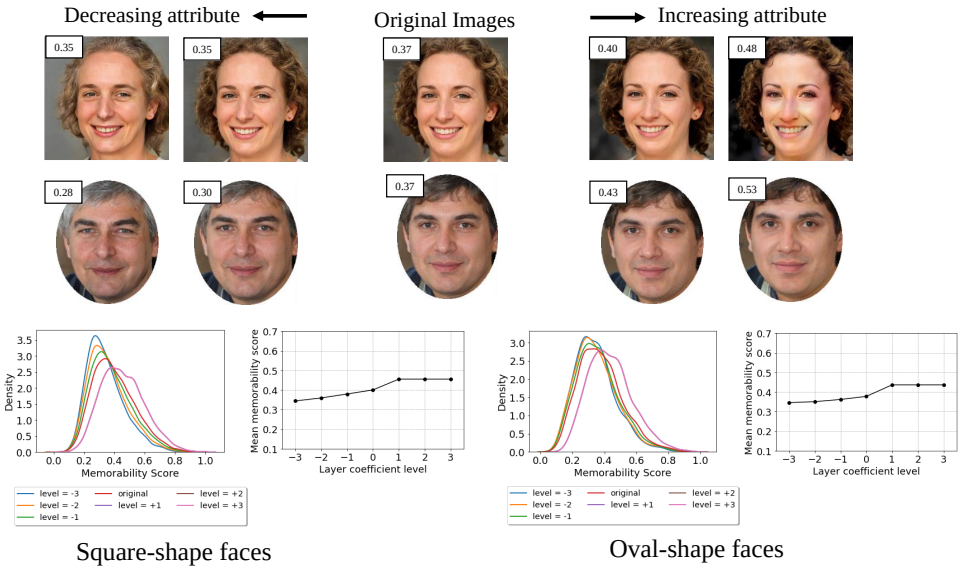


Figure 23: **The effect of the changes in the 8th layer of the extended latent space on memorability score.** This layer mostly affects the eyes and the hair color.

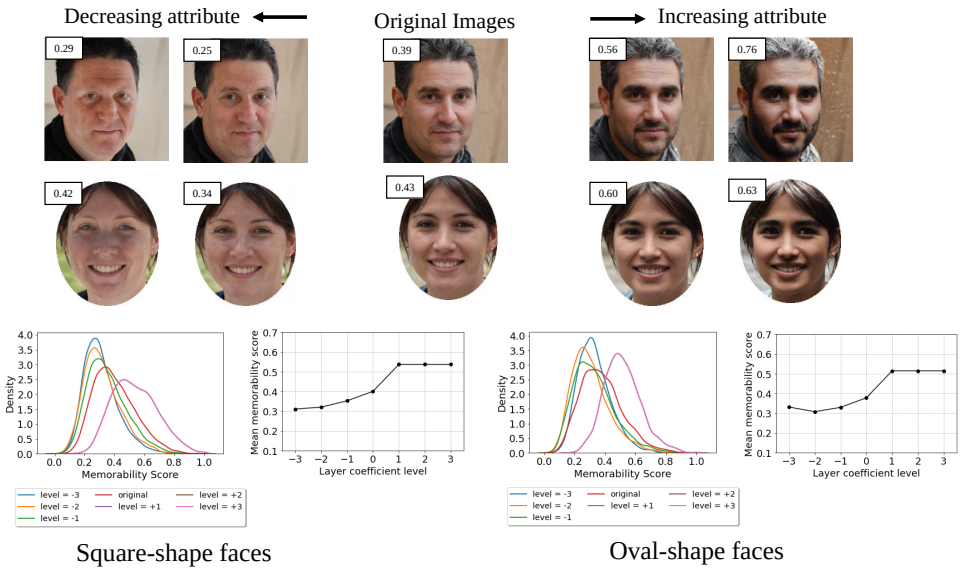


Figure 24: **The effect of the changes in the 9th layer of the extended latent space on memorability score.**This layer mostly affects the facial hair, hair color type and skin color. moving this layer, in the positive direction of the corresponding layer of the memorability vector (w^*), will make the hair color gold and some shadows and facial hair (if the person is a man), will appear on the face.

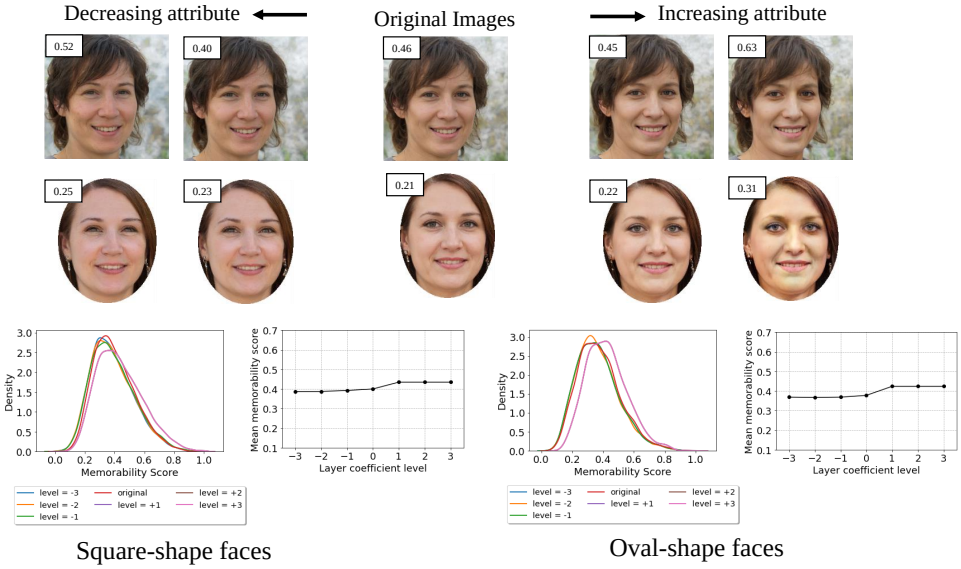


Figure 25: **The effect of the changes in the 10th layer of the extended latent space on memorability score.** The changes in this layer or mostly responsible for modifications on skin color and eyes. We observe, the changes in this layer are not as effective as other layers to modify the memorability score.

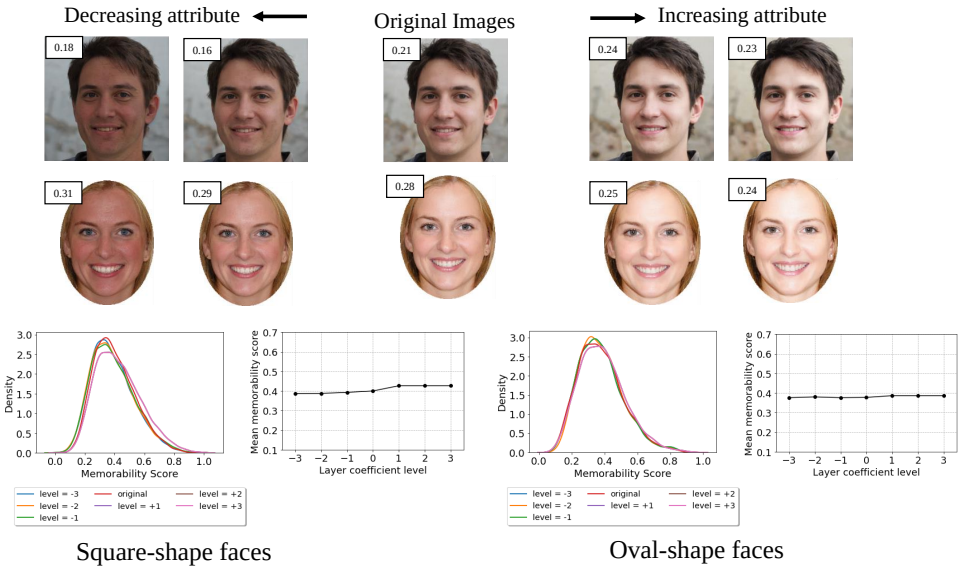


Figure 26: **The effect of the changes in the 11th layer of the extended latent space on memorability score.** Changes in this layer control the skin color and make the face brighter or darker by manipulating this layer. As it is shown, this layer does not play an important role in modifying memorability score of the face.