Learning a Deep Reinforcement Learning Policy Over the Latent Space of a Pre-trained GAN for Semantic Age Manipulation

Kumar Shubham*, Gopalakrishnan Venkatesh*, Reijul Sachdev[†§], Akshi*
Dinesh Babu Jayagopi[‡] , G. Srinivasaraghavan[‡]
International Institute of Information Technology, Bangalore, India* [‡]
Indian Institute of Technology, Madras, India[†]
{kumar.shubham, gopalakrishnan.v, akshi.025}@iiitb.org*,
{cs20d004}@smail.iitm.ac.in[†] {jdinesh, gsr}@iiitb.ac.in[‡]

Abstract—Learning a disentangled representation of the latent space has become one of the most fundamental problems studied in computer vision. Recently, many Generative Adversarial Networks (GANs) have shown promising results in generating high fidelity images. However, studies to understand the semantic layout of the latent space of pre-trained models are still limited. Several works train conditional GANs to generate faces with required semantic attributes. Unfortunately, in these attempts, the generated output is often not as photo-realistic as the unconditional state-of-the-art models. Besides, they also require large computational resources and specific datasets to generate high fidelity images. In our work, we have formulated a Markov Decision Process (MDP) over the latent space of a pre-trained GAN model to learn a conditional policy for semantic manipulation along specific attributes under defined identity bounds. Further, we have defined a semantic age manipulation scheme using a locally linear approximation over the latent space. Results show that our learned policy samples high fidelity images with required age alterations, while preserving the identity of the person.
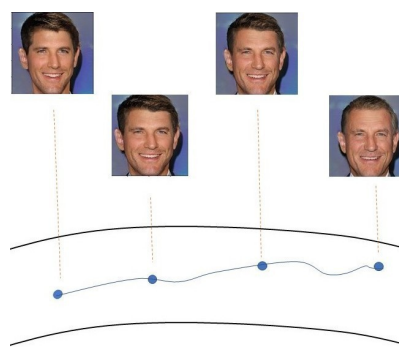
Fig. 1: A non-linear trajectory learned by the RL policy, over the latent space of a pre-trained ProgressiveGAN, performs the required age manipulation.

I. INTRODUCTION

The task of performing attribute manipulation in human face images has multiple applications. For example, face aging has often been used for cross face verification [25] and even in forensic art [9]. There have been various supervised and unsupervised approaches proposed for age-specific semantic manipulation [36], [37]. Unfortunately though, the outputs are low in resolution or not comparable to the images generated by the state-of-the-art GANs like ProgressiveGAN [15] and StyleGAN [16]. This often limits the application of such models in downstream tasks, which require high-resolution images with particular facial attributes. A custom generative model can generate high-resolution images with required facial attributes, but training it from scratch is an arduous task requiring enormous computational resources and specific datasets.

Recent studies have tried to understand and utilize the latent structure of generative models. The authors of [29] have shown

that the vector arithmetic over the latent space has a direct association with the semantic changes over the generated output. The authors of InterfaceGAN [35] have shown promising results in generating semantically rich high resolution images by traversing the latent space of ProgressiveGAN [15] and StyleGAN [16]. InterfaceGAN [35] proposes a linear traversal scheme over the latent space defined by the d-dimensional standard normal distribution, the base distribution used to train the ProgressiveGAN. This method has a lot of potential as it allows us to utilize pre-trained state-of-the-art models for downstream tasks.

Even though InterfaceGAN generates a rich set of high-quality images, it is often noticed that sometime generated output fail to preserve identity. The authors of InterfaceGAN have also mentioned that a linear traversal for age manipulation has a high correlation with gender which leads to undesired changes in the gender of the generated output. Such variations impact the identity of the person in the generated images which makes it unsuitable for many downstream tasks. To mitigate this issue, the authors of InterfaceGAN have proposed a complement scheme, wherein the projection of age hyperplane along the gender hyperplane is subtracted

arXiv:2011.00954v2 [cs.CV] 28 Apr 2021

from the original direction, to negate such unwanted changes. However, considering the large number of unwanted variations generated in such a trajectory it becomes impractical to train a hyperplane for all the attributes. Moreover, a linear traversal may also sample intermediate vectors which lie in the low-density regions of a high dimensional normal distribution [18].

In this work, we propose a non-linear traversal scheme (Figure 1) over the latent space of ProgressiveGAN [15] using Reinforcement Learning (RL). Our approach generates a path through the latent space using locally linear approximations of the desired manifold [2], [17], of the typical set [24] of a standard normal distribution, as shown in Figure 2. Unlike InterfaceGAN, our formulation places explicit emphasis on identity preservation over the sampled trajectory. We have formulated a Markov Decision Process (MDP) for such a traversal, which incentivizes the sampling of images with the required age variation while preserving identity.

Our contribution can be summarized in the following:

- We propose a non-linear traversal scheme over the latent space of the pre-trained ProgressiveGAN using RL. Our policy is conditioned on the base state and the required age alteration, giving it additional freedom to learn different manipulation schemes for different latent vectors.
- We have formulated an MDP for semantic manipulation without making any inherent assumption about the network architecture of the pre-trained GAN and the attribute over which the manipulation is performed. This approach provides the additional flexibility of easily extending our formulation to other state-of-the-art GAN models and other semantic attributes like head pose transfer.
- Our approach utilizes the local linear property of a manifold to generate a trajectory that would sample images with the required age variation while preserving the base image's identity. The results show the efficacy of this approach compared to InterfaceGAN.
- We propose a low-rank traversal scheme over the typical set [24] of a standard normal distribution to prevent out-of-distribution sampling, thereby preventing mode collapse in the sampled output.

II. RELATED WORK

Latent space for attribute manipulation: Latent space of GANs has been modeled as Riemannian manifold in many works [5], [34]. [14] studied the possibilities and limitations of GAN models to generalize beyond the training distribution and produce image transformations by exploring walks in the latent space.

Face aging with identity preservation: There have been several works specific to facial attribute manipulation in recent years [36], [37]. A necessary requirement in this task is to preserve the person's identity, which can be extremely challenging. Dual conditional GANs proposed in [36] attempts to learn identity-preserving face aging from unlabelled datasets of various age groups. Siamese GANs [13] can transform a low resolution input image to high resolution output. IPCGAN [37]

forces the high level features of the input image and synthesized image to be similar to preserve identity and uses an age classifier module to ensure that the generated image lies in the target age group. While these variants of conditional GANs do not rely on sequential training data and preserve identity, they still require training large networks and even fail to generate high resolution images compared to state-of-the-art models like ProgressiveGAN or StyleGAN.

Reinforcement learning for face ageing: The authors of [8] have used Reinforcement Learning (RL) to generate facial aging in videos. However, unlike our approach, the authors have used RL and associated reward schemes for the temporal consistency of facial aging in the frames of genereated video. They have relied on the linear traversal scheme over the activation space of VGG-19 model, based on clusters of semantically opposite images, and used a policy network to sample appropriate images from the reference dataset. This approach restricts the generation to the choice of the reference dataset which would thereby impact the quality of the overall output. Our work is fundamentally different from this, as we have used RL to learn a non-linear trajectory over the latent space of the generative model which generates a single image and not for temporal consistency. We have also relied on the learned distribution of ProgressiveGAN to generate high fidelity images and have not used any reference dataset.

III. METHOD

Recently, there have been multiple successful applications of Reinforcement Learning (RL) algorithms in real world problems [21]. Unfortunately, such applications are limited to solving low dimensionality formulations. Learning policies directly over visual input is still far from being plausible and often requires an intermediate representation learning procedure to learn a condensed embedding of the necessary visual information. Once such a representation is learned, a complete MDP is defined over the low dimensional representation to learn the required policy over the latent space. For the task of representation learning, deep generative models have shown a lot of promise. Researchers have used Variational Autoencoders (VAEs) [41] to learn such representations. The authors of [20] have even shown that the latent space of InfoGAN [6] is capable of learning goal-directed visual plans.

Contrary, to learning a condensed representation of visual input. Our work proposes an MDP formulation to exploit the semantics of representations in pre-trained GANs to generate high fidelity images for facial aging. For this task, we have binned the continuous range of plausible age variations into consecutive buckets. Inspired by goal-reaching RL problems, we learn a conditional policy which samples face images belonging to different age bins; conditioned on the base state (latent vector defines the base identity to be generated) and the required age alteration i.e., making the face older (ascending) or making the face younger (descending). In our MDP formulation, the states and the transition function are defined over the latent space of the ProgressiveGAN [15].

Rather than learning a policy defined over the full R^d dimension space, we have restricted the search space to the typical set [24] of a d-dimensional standard normal distribution (the base distribution used to train the ProgressiveGAN [15]). We have used the locally linear approximation of manifolds to learn a linear affine subspace to generate a new state under the Markovian assumption. The agent gets a positive reward if it samples from a new age bin (to improve image generation with age diversity) while respecting the defined identity bound (to preserve identity) and the typical set of the distribution (to avoid mode collapse). In later sub-sections, we will discuss each of these components individually.

A. Latent Space and Goal

The latent space of ProgressiveGAN, defined over the d-dimensional standard normal distribution, is capable of generating high fidelity images with required semantic alteration [35]. In our work, the RL policy learns a sampling scheme that can sample the required latent vector and thereby generate the corresponding image with the necessary facial properties. For this, our state space comprises of the d-dimensional latent space of ProgressiveGAN ($\mathcal{G}_{model}$), where each latent vector s is associated with its corresponding image based on the following equation.

$$I \leftarrow \mathcal{G}_{model}(s) \tag{1}$$

For the given sampling policy, we have also defined the corresponding goal g to be the task of preserving the identity (I^{base}) while inducing age variations over the set of generated images. Hence we represent our goal (g) as the concatenation of latent vector associated with the given base image (s^{base}) and the vector of 0's or 1's, R^d, to define the required age alteration on the base image: younger ($\mathcal{C}^{dsc}$) or older ($\mathcal{C}^{asc}$). Hence, the goal (g) for a given base image and required age alteration is represented as $[s^{base}, \mathcal{C}]$ where $C \in \{\mathcal{C}^{asc}, \mathcal{C}^{dsc}\}$. Similar to the goal-based RL formulations, we have used the concatenation of state and goal vectors in the policy neural networks [3], [32].

B. Locally linear approximation

InterfaceGAN [35] has shown that a linear traversal across a hyperplane defined over the latent space of ProgressiveGAN, can generate the required semantic variation for a given base image. Often, such a linear traversal isn't generic enough to sample images across different age groups while preserving the person's identity. Also, the linear traversal scheme often fails to consider the latent space's manifold structure and can even lead to traversing low-density region where, by definition, the model has not been trained well [18].

In our work, we circumvent this issue by generating a non-linear trajectory using a locally linear approximation of the manifold, as shown in Figure 2. This approximation strategy gives the model an additional freedom to learn a shared local manifold structure across different latent vectors and utilize it for a given task. Manifold structures for face aging around a given conditioned base state can be approximated to be
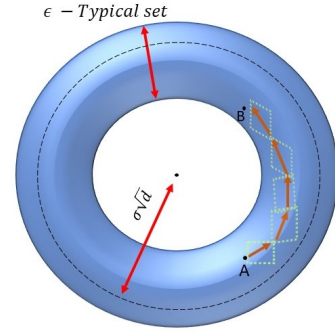

Fig. 2: Non linear traversal, using the locally linear approximation, over the typical set of a normal distribution from point A to point B.

locally Euclidean and the geometry can be estimated using linear approximations (planar subspace) [2], [17]. Any planar approximation of a d-dimensional latent space requires at least two R^d dimensional basis vector, which can span the full space.

For the given problem, we have taken inspiration from InterfaceGAN and have used one of the basis as the planar hyperplane direction learned for facial aging (K^{hyp}). This choice is motivated by the high degree of disentanglement along that direction and also the promising level of achieved generalization [35]. Based on the defined age alteration, we have used K_g^{hyp} as K^{hyp} if $\mathcal{C} = \mathcal{C}^{asc}$ or $-K^{hyp}$ if $\mathcal{C} = \mathcal{C}^{dsc}$. The other basis for the local approximation, K_t^{gen}, is generated by our policy network $\pi_g(a_t|s_t)$, as part of the action required at a given state. Our action space a_t is R^{d+2} dimensional. Here, the d-dimensions provide the predicted basis vector and the other two dimensions provide the scalars, w_t^1 and w_t^2, to weigh K_g^{hyp} and the generated basis vector, K_t^{gen}, respectively. Therefore, the action a_t is the concatenation of $[k_t^{gen}, w_t^1, w_t^2]$. Hence the transition function for given action a_t at state s_t is defined as:

$$s_{t+1} = \mathcal{T}s_t + (1 - \mathcal{T})(s_t + w_t^1 * k_g^{hyp} + w_t^2 * k_t^{gen}) \tag{2}$$

For our experiments, we have introduced an additional hyper parameter $\mathcal{T}$ which controls the smoothness of the generated trajectory [38]. An episode starts with s_{base} i.e., $s_0 = s^{base}$ and the full trajectory is unrolled from there.

C. Typical set

The typical set for a probability distribution is defined as a set of elements whose information content is sufficiently close to the expected information content of the distribution [24]. As mentioned in [24], an (ϵ, N) - typical set for a distribution P(x) with support $x \in \mathcal{X}$ is defined as a set of all N-length sequence that satisfy

$$\mathcal{H}[P(x)] - \epsilon \leq \frac{1}{N} - \log(P(x_1, x_2, x_3...x_N)) \leq \mathcal{H}[P(x)] + \epsilon \tag{3}$$

where $\mathcal{H}[P(x)]$ represents the entropy of the distribution, $P(x)$, and is defined as $\int_{\mathcal{X}} P(x)[-log(P(x))]dx$. A simpler factorized formulation for the same equation can be written as

$$\mathcal{H}[P(x)] - \epsilon \leq \frac{1}{N} \sum_{n=1}^{N} -\log(P(x_n)) \leq \mathcal{H}[P(x)] + \epsilon \quad (4)$$

In [24], the authors have shown that a high likelihood for an out-of-distribution dataset in a generative model results from a mismatch between the typical set of a distribution and the region with a high likelihood. During the training of the generative model, most of the latent vectors are sampled from a subset of the model's full support and especially from the typical set of the distribution. For a d-dimensional standard normal distribution $N(0, I)$ an $(\epsilon, 1)$ - typical set is defined as $\frac{1}{2}|d - ||x||_2^2| \leq \epsilon$. Interestingly, this equation is similar to the Gaussian Annulus Theorem [39] defined over the same distribution. For a normal distribution this region lies at a $\sigma\sqrt{d}$ distance from the mode as shown in Figure 2. Even the authors of InterfaceGAN [35] have mentioned that out-of-distribution sampling performed during their proposed linear manipulation leads to mode collapse.

We have tackled this issue by restricting the learned trajectory to fall over the typical set of the normal distribution. For each newly generated state, a check is done to determine if the generated state is within the defined typical set of the distribution. If the new state is not in this region, then the episode is terminated with high negative reward. We have used the following function to formulate an $(\epsilon, 1)$ - typical set defined over the d-dimensional latent space of ProgressiveGAN.

$$\mathcal{Z}_g(t) = \begin{cases} 1 & \text{if } \frac{1}{2}|d - ||s_t||_2^2| \leq \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $\epsilon \in R^+$ and is a hyper parameter.

D. Age Manipulation

In our RL framework, we have associated the *age* property with each generated image (obtained from the sampled latent vector - equation 1). For the task of predicting the age of a face image, we have trained an age regression (A_{reg}) model using a ResNet-50 [11] network over the IMDB Wiki Faces dataset [30]. The following equation enlists the prediction mechanism.

$$\begin{aligned} I_t &\leftarrow \mathcal{G}_{model}(s_t) \\ age_t &\leftarrow \mathcal{A}_{reg}(I_t) \end{aligned} \quad (6)$$

Our reward scheme then incentivizes the policy to sample face images belonging to different age groups (bins). For example, say the base image is of a 30-year-old person and the policy is conditioned to generate older faces (ascending order). In such a scenario, the agent will receive its reward only if it samples images belonging to the higher age group, i.e., age > 30. The exact reward structure is described in Section III-F.

As the ProgressiveGAN [15] is trained over CelebA-HQ dataset [15], [22], there is an age-associated bias over the learned latent space. Most of the generated images from the model have an associated age value that lies between 20-60 year group. Hence, for our experiments, we have sub-divided the given age range into multiple buckets ($\mathcal{S}$) of equal size $\mathcal{B}$, to group similarly aged face images into the same bucket.

To achieve the desired diverse age variation within a given episode, we incentivize the policy to cover as many buckets as possible, in a given order, while discouraging it from sampling images from the same bucket. For this, our policy is encouraged to generate one image from each of the buckets, under the given ascending or descending conditioning, while being within the set identity bound (described Section III-E). To ensure this, we reward the agent only when it transitions to an unvisited age bucket. This approach ensures that, for an episode, the policy looks for a diverse set of age variations and thereby does not sample images from the same age bucket. We define an indicator function $\delta(age)$, which specifies, if the newly generated face image has an age value associated with an unvisited bucket. Using this formulation, we define a function $\mathcal{M}_g(t)$, which indicates if the newly sampled face image qualifies for a reward.

$$\mathcal{M}_g(t) = \begin{cases} 1 & \text{if } age_t > age_{base} \text{ and } \delta(age_t) = 1 \text{ and } \mathcal{C} = \mathcal{C}^{asc} \\ 1 & \text{if } age_t < age_{base} \text{ and } \delta(age_t) = 1 \text{ and } \mathcal{C} = \mathcal{C}^{dsc} \quad (7) \\ 0 & \text{otherwise} \end{cases}$$

E. Identity preservation

One of the key components of our proposed formulation is the task of preserving identity during face sampling. To preserve the identity of the generated image during state transitions, we have defined a reward scheme where an incentive is given for sampling images for a new age bin under a given threshold over identity.

Generally, an age manipulation over an image leads to multiple changes in the face; this makes the task of identity comparison even more challenging. The authors of [40] have shown that a Mean Square Error (MSE) in a pixel space between the base image and generated image does not capture the face aging attributes like hair color, beard, and wrinkles. Instead, a perceptual loss based on the image's content was proposed as a surrogate for identity preservation during age change. The authors have used the lower feature layers of an AlexNet [19] model trained on ImageNet and have empirically shown that the conv5 feature vector is the best suited surrogate. Inspired by [40], we have also used an MSE based comparison of the conv5 layer of an AlexNet model. The squared distance is computed between the feature vectors of the base image and all the associated images generated in the given trajectory. For each new state generated by the transition function, we generate the corresponding image using our ProgressiveGAN. This image is then passed through the AlexNet model to get the corresponding conv5 feature vector (F).

$$\mathcal{I}_g(t) = ||\mathcal{F}(I_t) - \mathcal{F}(I_{base})||^2 \quad (8)$$

Preserving identity while transitioning over large age differences is challenging. This problem arises, as a lot of the facial attributes have to be modified. For example, an image in the (20-25) age bucket will not have wrinkles, but such attributes would be present in the older age group of (55-50). This makes the sampling of images across extreme age differences difficult. It becomes extremely tough to distinguish it from a facial attribute modification to an identity change, as these have a very high correlation with the identity of a person. One way of solving this is by having a more lenient threshold on the identity (8). However, this tends to introduce noisy intermediary samples.

To solve this issue, we define two thresholds: a soft threshold P_1, under which identity will be better preserved, and a hard threshold P_2, where an agent will be able to sample images from the extreme age differences without much deterioration in identity. The thresholds respect the $P_1 < P_2$ relationship. Our reward scheme thereby incentivizes the agent to make decisions by following the locally linear approximation to sample images under soft threshold while the hard threshold gives the flexibility to sample across extreme age variations.

F. Reward Scheme

For each transition, the corresponding s_t is passed through ProgressiveGAN (equation 1) to get the associated image I_t. This image, I_t, is further passed through age regressor ($\mathcal{A}_{reg}$) and through Alexnet [19] model to get the corresponding age and identity feature associated with it. If the generated state is associated with an unvisited bucket under given sampling order (ascending or descending) and within a given identity bound (based on equations 7 and 8), then the policy is credited with a positive reward.

The maximum number of steps in an episode is set to be $\mathcal{E}_{len}$. As mentioned in Section III-E, we have considered a dual thresholding scheme where P_1 is the soft threshold and P_2 is the hard threshold such that $P_1 < P_2$. If the MSE score based on equation 8 over the identity is less than P_1 and $\mathcal{M}_g(t) = 1$, then the agent is credited with mr reward. If it is between P_1 and P_2, we provide a reward r, and if it crosses over P_2 then we give a large negative reward and terminate the episode. The idea of having a reward scheme based on soft and hard margin is to promote sampling of images that have a large age difference. Such large variations cannot be captured under tighter identity bound. The P_2 margin over the trajectory also serves as a filter to remove spurious face images with plausibly different identity. Our reward scheme incentivizes the agent to reach an unvisited age bucket under P_2 identity bound while providing the flexibility to search for better identity preserving images under P_1 bound. If the model fails to generate a new state that satisfies the aforementioned conditions, it gets -1 reward. The concept of giving a negative reward is inspired by simulation-based goal-reaching RL tasks like Fetch-Push [27] where the agent tries to finish the goal as early as possible. Once the agent samples from all the age buckets, under given conditioning, the episode is terminated. A sizeable negative

Parameter	r	n	$\mathcal{E}_{len}$	P_1	P_2	$\mathcal{T}$	ϵ	d	$\mathcal{B}$	m
Value	2	25	60	750	900	0.3	3	512	5	2

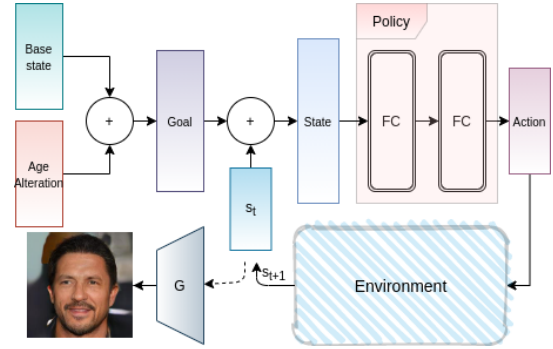TABLE I: Hyper-parameters used in our experiments

Fig. 3: Architecture of our proposed RL formulation. The reward from the environment has not been shown for brevity. The policy network comprises of two Fully Connected (FC) layers with *tanh* activations. As per equation 1, the face image is decoded by the generator *G*.

reward is also given if the new state does not fall in the typical set of the isotropic normal distribution.

$$\mathcal{R}_g(s_t, a_t, s_{t-1}) = \begin{cases} -n & \text{if } \mathcal{I}_g(t) > P_2 \text{ or } \mathcal{Z}_g(t) = 0 \\ mr & \text{if } \mathcal{I}_g(t) \leq P_1 \text{ , } \mathcal{M}_g(t) = 1 \text{ \& } \mathcal{Z}_g(t) = 1 \\ r & \text{if } P_1 < \mathcal{I}_g(t) \leq P_2 \text{ , } \mathcal{M}_g(t) = 1 \text{ \& } \mathcal{Z}_g(t) = 1 \\ -1 & \text{otherwise} \end{cases}$$

$$(9)$$

where $r, n \in R^+$

IV. EXPERIMENTS

A. Implementation Details

For our experiments[§], we use a ProgressiveGAN pre-trained on $1024 \times 1024 \times 3$ resolution face images from the CelebA-HQ dataset [15]. For the defined environment formulation, both Proximal Policy Optimization (PPO) [33] and Advantage Actor Critic (A2C) [23], policy-gradient based model-free RL algorithms were trained to learn the corresponding conditioned policy. For training of the RL policy, 60,000 latent vectors and their corresponding images were sampled. The starting state of an episode is chosen randomly from this set of 60,000 training samples. After every 100 episodes of training, the age alteration order i.e., $\mathcal{C}^{asc}$, $\mathcal{C}^{dsc}$ is switched. The hyperparameters associated with training are listed in Table I. The results presented from here onwards are from policies which have been trained for 1 million steps. The trained RL policy was tested on another unseen set of sampled vectors.

Figure 3 illustrates the complete architecture of our proposed RL framework. The associated environment was defined and implemented in Pytorch [26]. We used the TensorFlow 2

[§]The source code associated with the environment definition and the RL algorithms is available on https://github.com/kyrs/rl_semantic_manipulation.
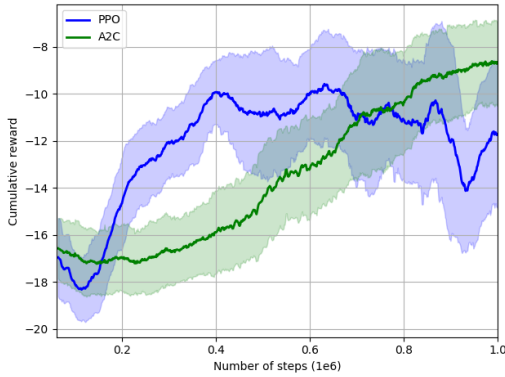
Fig. 4: Cumulative reward plots of the policies learnt by the PPO and A2C algorithms. The curve represents the average reward and the shaded region depicts the variance.

Models	Altertion	VGG Face	SSIM	PSNR
ProgressiveGAN	Young	0.85 ± 0.10	0.52 ± 0.77	57.18 ± 2.25
InterfaceGAN	Young	0.91 ± 0.11	0.64 ± 0.18	58.95 ± 4.22
PPO (Ours)	Young	$\mathbf{0.94} \pm 0.05$	$\mathbf{0.68} \pm 0.10$	$\mathbf{59.78} \pm 2.54$
A2C (Ours)	Young	$\mathbf{0.94} \pm 0.04$	0.64 ± 0.10	59.04 ± 2.67
ProgressiveGAN	Old	0.81 ± 0.11	0.51 ± 0.08	57.33 ± 1.99
InterfaceGAN	Old	0.90 ± 0.08	0.61 ± 0.12	59.07 ± 2.89
PPO (Ours)	Old	$\mathbf{0.95} \pm 0.04$	$\mathbf{0.67} \pm 0.09$	$\mathbf{59.99} \pm 2.50$
A2C (Ours)	Old	0.93 ± 0.05	0.61 ± 0.10	58.81 ± 2.23

TABLE II: Comparison of the cosine similarity scores of VGG Face, SSIM and PSNR between linear trajectories from ProgressiveGAN, InterfaceGAN and the non-linear trajectories from the policies of the RL algorithms. The best scores have been marked in **bold**.

[1] branch of the baselines [7] package to implement and test the model-free RL algorithms. Inspired by [31], our policy network consists of two Fully Connected (FC) layers each of 64 dimensions with *tanh* activations. The experiments were performed on a machine with Intel (R) Xeon (R) Silver 4114 Processor @ 2.20GHz, 40 cores, 128GB RAM, and 2 Nvidia GeForce RTX 2080Ti each with 11016 MB.

B. Identity preservation and catastrophe avoidance

The unfavourable condition of leaving the typical set or failure to preserve identity has been termed as a catastrophe. As mentioned in equation 9, the agent is given a large negative reward when it generates states associated with a large identity difference from the base image. A negative reward is also given whenever the generated state does not fall in the defined $\epsilon-$typical set. Such negative rewards, particularly for identity preservation and for falling in the typical set, introduces a notion of *fear* [10] in the model to generate images in a limited sub-region where there is a better possibility to preserve identity and associated semantics while maintaining high generation quality. The negative reward hyperparameter, *n*, can be changed based on user-specific needs. In our experiments, we have explored the model's feasibility to generate diverse age groups while respecting the above-mentioned conditions. Figure 4 shows the cumulative reward curve for both the policy-gradient based algorithms.

C. Comparison with other algorithms

We evaluate our proposed RL methods against two standard algorithms. The first baseline is computed on the pre-trained ProgressiveGAN by estimating the latent direction for age manipulation. This direction is computed by subtracting the centroids of two clusters [8] of young and old face images. For our evaluation, each of these clusters had 32 points. Secondly, we compare our RL algorithms against the state-of-the-art InterfaceGAN algorithm.

1) Quantitative Evaluation: To compare the identity preservation aspect of different models, we have computed the cosine similarity between the feature vectors from VGG Face [28]
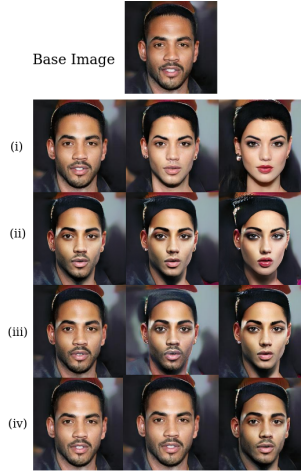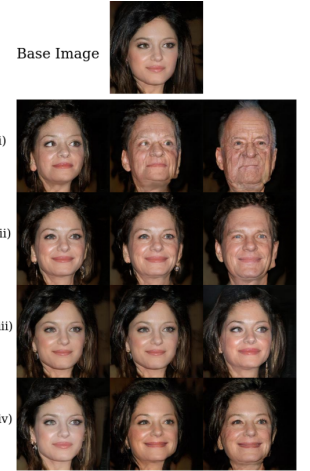

Fig. 5: Age alteration: young   Fig. 6: Age alteration: old

Qualitative samples from the trajectories generated by (i) ProgressiveGAN (ii) InterfaceGAN (iii) PPO (iv) A2C for the specified base image and the mentioned age alteration.

of the base image and the images sampled in the trajectory. We have further computed SSIM and PSNR [12] score to evaluate quality of generated image. For this evaluation, we have generated images from each algorithm for 300 randomly sampled unseen base images.

From Table II, it is clear that our RL models perform better than the baseline ProgressiveGAN and state-of-the-art InterfaceGAN in terms of identity preservation, as measured in the VGG Face measure. From the superior SSIM and PSNR scores (Table II), it is also evident that the RL algorithms maintain a very high generation quality while sampling semantically modified versions of the base image. This high image generation quality could be attributed to the RL policy's ability to sample latent points from the Typical set (Section III-C) of the distribution on which the ProgressiveGAN was trained.

2) Qualitative Evaluation: Figures 5 and 6 show age alterations introduced by our RL algorithms when compared against the baseline linear traversal in ProgressiveGAN and the current state-of-the-art InterfaceGAN. For a fair comparison between the methods, we have used *age hyperplane without compliment* as the primary basis in our models and the

Fig. 7: Age alteration: young

Fig. 8: Age alteration: old

Trajectories sampled by PPO and A2C algorithms for the specified age alteration and base image.
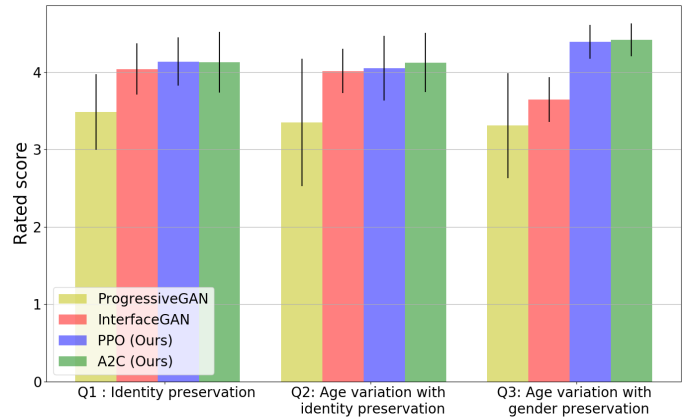

Fig. 9: User ratings for the trajectories generated by the algorithms on the three queries. The height of the bar denotes the average score and the error bar specifies the standard deviation.

same has been used as a hyperplane for linear traversal in InterfaceGAN. From the presented examples, it can be seen that the images generated by our RL algorithms does a better job of preserving the gender while inducing necessary age alterations in the base image.

According to [35], there is a high degree of entanglement between the linear latent directions for age and gender. Therefore a linear traversal across the semantic boundary could cause unnecessary changes in other attributes like gender. Even our findings, like in Figures 5 and 6, show that by following a linear trajectory, a gender change is often noted when the age alteration is introduced. On the other hand, the non-linear trajectories from the RL policies mitigates this issue by relying on an identity-based comparison of the sampled image and the base image (equation 8). If the RL policy encounters such a discrepancy, the episode is terminated (Section IV-B) with a large negative reward. This negative signal would prevent the policy from making significant identity changes, including gender alterations, while introducing the required semantic modification. Representative examples from our proposed RL framework involving PPO and A2C is presented in Figures 7 and 8. From these samples, it is clear that the proposed RL algorithms induce the required age alteration without compromising the identity of the base image.

3) User Study: Considering the subjective nature of perceiving ageing of faces, we further evaluated the algorithms by involving human raters. Thereby, 300 trajectories from the unseen test set were evaluated by four humans who rated the algorithms on a 5-point scale, with 5 being the best score.

The results presented in Sections IV-C1 and IV-C2 show that

our RL methods achieve commendable performance in identity preservation while not compromising on the generation quality. As identity is deeply coupled with other facial aspects, like gender, we wanted to further evaluate if our non-linear traversal schemes are capable of retaining the gender of the base image. Hence, the evaluators rated the algorithms on the following dimensions: **Q1**: Evaluate the identity preservation aspect in the generated trajectories. **Q2**: Rate the algorithms based on its ability at inducing the required age alteration while preserving the identity of the base image. **Q3**: Rate the algorithms based on its ability at introducing the age alteration while preserving the gender of the base image.

The Intraclass Correlation Coefficient (ICC) [4], between the ratings from the evaluators, is 0.59 indicating a moderate to good agreement. As shown in Figure 9, our non-linear trajectories from the RL algorithms perform better than the linear traversal algorithms on all the queries. From the results of **Q3** in Figure 9 and the qualitative samples presented in Figures 5 and 6, it is evident that the non-linear trajectories from the RL method achieves a higher degree of disentanglement between the age and gender attributes than the linear traversal algorithms. We would attribute this emergence of disentanglement to the non-linear trajectory's ability in preserving the identity of the base image. The RL policies have been trained with a *fear* factor, of not compromising on identity, thereby forcing it to sample images where the required age alteration is introduced while keeping other attributes, including gender, intact.

V. CONCLUSION

In this work, we have formulated a locally linear traversal scheme over the latent space of a ProgressiveGAN for the task of semantic age manipulation of a given image. Our approach is independent of the generative model's architecture and can be easily extended to other state-of-the-art GAN models and facial attributes. We obtain superior results against the current state-of-the-art method, InterfaceGAN, for the task of face

ageing and have shown that the learned conditional policy does a better job at preserving the base state's identity while introducing the required semantic variations. We have also shown the effectiveness of our approach in learning a non-linear trajectory while generating samples that belong to the distribution's typical set. In the future, we will extend this formulation to other attributes, like pose variation and smile change. We also intend to experiment with different GAN architectures and generative models. Furthermore, theoretical investigations into the convergence of these RL based methods could be interesting future work.

REFERENCES

[1] Martín Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016.

[2] Talal Ahmed. *Geometric manifold approximation using locally linear approximations*. PhD thesis, Rutgers University-Graduate School-New Brunswick, 2016.

[3] Dzmitry Bahdanau, Felix Hill, Jan Leike, Edward Hughes, Arian Hosseini, Pushmeet Kohli, and Edward Grefenstette. Learning to understand goal specifications by modelling reward. *arXiv preprint arXiv:1806.01946*, 2018.

[4] John J Bartko. The intraclass correlation coefficient as a measure of reliability. *Psychological reports*, 19(1):3–11, 1966.

[5] Nutan Chen, Alexej Klushyn, Richard Kurle, Xueyan Jiang, Justin Bayer, and Patrick Smagt. Metrics for deep generative models. volume 84 of *Proceedings of Machine Learning Research*, pages 1540–1550. PMLR, 2018.

[6] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2172–2180, 2016.

[7] Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, Yuhuai Wu, and Peter Zhokhov. Openai baselines, 2017.

[8] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, Nghia Nguyen, Eric Patterson, Tien D Bui, and Ngan Le. Automatic face aging in videos via deep reinforcement learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10013–10022, 2019.

[9] Yun Fu, Guodong Guo, and Thomas S Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010.

[10] Javier Garcıa and Fernando Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480, 2015.

[11] 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*, pages 770–778, 2016.

[12] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[13] Chih-Chung Hsu, Chia-Wen Lin, Weng-Tai Su, and Gene Cheung. Sigan: Siamese generative adversarial network for identity-preserving face hallucination. *IEEE Transactions on Image Processing*, 28(12):6225–6236, 2019.

[14] Ali Jahanian, Lucy Chai, and Phillip Isola. On the "steerability" of generative adversarial networks, 2019.

[15] 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.

[16] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[17] Sofia Karygianni and Pascal Frossard. Tangent-based manifold approximation with locally linear models. *Signal Processing*, 104:232–247, 2014.

[18] Yannic Kilcher, Aurélien Lucchi, and Thomas Hofmann. Semantic interpolation in implicit models. *arXiv preprint arXiv:1710.11381*, 2017.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[20] Thanard Kurutach et al. Learning plannable representations with causal infogan. In *Advances in Neural Information Processing Systems*, pages 8733–8744, 2018.

[21] Sascha Lange, Martin Riedmiller, and Arne Voigtländer. Autonomous reinforcement learning on raw visual input data in a real world application. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2012.

[22] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[23] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937, 2016.

[24] Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, and Balaji Lakshminarayanan. Detecting out-of-distribution inputs to deep generative models using typicality. *arXiv preprint arXiv:1906.02994*, 2019.

[25] Unsang Park, Yiying Tong, and Anil K Jain. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):947–954, 2010.

[26] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library, 2019.

[27] Matthias Plappert et al. Multi-goal reinforcement learning: Challenging robotics environments and request for research. *arXiv preprint arXiv:1802.09464*, 2018.

[28] Zakariya Qawaqneh, Arafat Abu Mallouh, and Buket D Barkana. Deep convolutional neural network for age estimation based on vgg-face model. *arXiv preprint arXiv:1709.01664*, 2017.

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

[30] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018.

[31] Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017.

[32] Tom Schaul, Daniel Horgan, Karol Gregor, and David Silver. Universal value function approximators. In *International conference on machine learning*, pages 1312–1320, 2015.

[33] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[34] Hang Shao, Abhishek Kumar, and P. Thomas Fletcher. The riemannian geometry of deep generative models, 2017.

[35] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9243–9252, 2020.

[36] Jingkuan Song et al. Dual conditional gans for face aging and rejuvenation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 899–905. International Joint Conferences on Artificial Intelligence Organization, 2018.

[37] X. Tang, Z. Wang, W. Luo, and S. Gao. Face aging with identity-preserved conditional generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, 2018.

[38] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, pages 1195–1204, 2017.

[39] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[40] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7939–7947, 2018.

[41] Manuel Watter, Jost Springenberg, Joschka Boedecker, and Martin Riedmiller. Embed to control: A locally linear latent dynamics model for control from raw images. In *Advances in Neural Information Processing Systems*, pages 2746–2754, 2015.