Looks Like Magic: Transfer Learning in GANs to Generate New Card Illustrations

Matheus K. Venturelli, Pedro H. Gomes, Jônatas Wehrmann
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
Rio de Janeiro, Brazil
{mventurelli,pedro.gomes}@inf.puc-rio.br, wehrmann@puc-rio.br

Abstract—In this paper, we propose MAGICSTYLE-GAN and MAGICSTYLE-GAN-ADA – both incarnations of the state-of-the-art models StyleGan2 and StyleGan2 ADA – to experiment with their capacity of transfer learning into a rather different domain: creating new illustrations for the vast universe of the game “Magic: The Gathering” cards. This is a challenging task especially due to the variety of elements present in these illustrations, such as humans, creatures, artifacts, and landscapes – not to mention the plethora of art styles of the images made by various artists throughout the years. To solve the task at hand, we introduced a novel dataset, named MTG, with thousands of illustration from diverse card types and rich in metadata. The resulting set is a dataset composed by a myriad of both realistic and fantasy-like illustrations. Although, to investigate effects of diversity we also introduced subsets that contain specific types of concepts, such as forests, islands, faces, and humans. We show that simpler models, such as DCGANs, are not able to learn to generate proper illustrations in any setting. On the other side, we train instances of MAGICSTYLE-GAN using all proposed subsets, being able to generate high quality illustrations. We perform experiments to understand how well pre-trained features from StyleGan2 can be transferred towards the target domain. We show that in well trained models we can find particular instances of noise vector that realistically represent real images from the dataset. Moreover, we provide both quantitative and qualitative studies to support our claims, and that demonstrate that MAGICSTYLE-GAN is the state-of-the-art approach for generating Magic illustrations. Finally, this paper highlights some emerging properties regarding transfer learning in GANs, which is still a somehow under-explored field in generative learning research.

Index Terms—gan, transfer learning, stylegan, mtg

I. INTRODUCTION

MAGIC: THE GATHERING (MTG)\(^1\) is one of the most popular collectible card games in the world. The game is known for its iconic cards, always accompanied by fascinating illustrations. Throughout its almost 30 years of existence, MTG assembled a vast universe of creatures, lands, and artifacts. Each new expansion set comes with different art styles, with 32 possible combinations of colors that alter the design of the cards. This leads to an extensive range of possible gameplays, and also brings many challenges for generating new illustrations based on a dataset of existing cards.

In this paper we explore several approaches for training and fine-tuning generative neural networks for generating cards for the game Magic. We begin exploring straightforward strategies which are indeed limited by several aspects of the architecture, and then present options based on transfer learning of large pre-trained GANs. This study shows that it is possible to leverage pre-trained information even when the network was trained on rather distinct domains. For instance, we clearly demonstrate that StyleGAN2 previously trained to generate human faces can be used to generate a plethora of Magic-related concepts, such as forests, landscapes, human-like characters and the like. Moreover, we present ablative studies carried out in separate subsets of the proposed dataset, allowing us to understand the visual aspects that such networks can easily learn, and those that are harder for them.

In summary, we propose a deep learning method, hereby

\(^1\)https://magic.wizards.com/en
referred as MAGICSTYLEGAN (see Figure 1), to create new card illustrations for the game MAGIC: THE GATHERING with the aid of Generative adversarial networks (GAN) [10]. We introduced a novel dataset composed of more than 60,000 high quality $1024 \times 1024$ card illustrations. Our dataset is labeled with rich metadata, that can be used to identify the original author of the illustration, as well as information regarding category, type of illustration etc. Note that it is complex and computationally costly to generate images in such high dimensionality. To circumvent that we first proposed a DCGAN-based [16] baseline that generates lower dimensional images.

The main contributions of the paper are three-fold: (i) empirical analysis of generative networks for generating artistic-looking cards; (ii) extensive evaluation of transfer learning capabilities of state-of-the-art generative models; and (iii) a novel, large dataset containing cards for the game Magic, along to several subsets that can be used to validate transfer learning with few instances in more controlled settings.

II. RESEARCH PROBLEM

In this paper we propose to handle the problem of card illustration generation for the game MAGIC: THE GATHERING. The generation of images for card games is an interesting research problem in many distinct facets. Such illustrations are made out of the time-demanding creative process of multiple artists worldwide, and notably, there is room to improve the artistic capabilities of AI systems. Moreover, it is not a simple task to train a generative model to create such illustrations, given the diversity among the images and the current datasets that are limited in number of samples. Not to mention that such images are in high resolutions, e.g., $1024 \times 1024$. GANs tend to generate better results when trained in more specific and controlled domains, such as generation of human faces, horses, cars, beaches, rooms, etc. Therefore, it is unclear whether the current state-of-the-art approaches would be able to represent such large diversity of concepts as those presented in the illustrations. On top of that, one has to recall that the process of training GANs from scratch is often prone to convergence problems and mode collapse. Those can be handled using some best practices, but often a large training dataset is also paramount for the model to be able to generalize and generate high-quality high-resolution data.

In addition to experimenting with GANs training from scratch, we aim to uncover some properties of a somewhat unexplored area: transfer learning in GANs. Finetuning options are widely explored when applied to visual recognition and language modeling, although there are only few studies on the reuse of pre-trained GANs for different tasks. This paper proposes to answer the following question: “Does transfer-learning in GANs benefit the generation of artistic illustrations of cards for the game Magic?”

To the best of our knowledge, this is the first study that handles the task of MAGIC card generation, as well as studying Generative Models transfer-learning capabilities across models trained in completely different domains.

III. BUILDING THE MAGICSTYLEGAN

There are plenty of different GAN architectures nowadays, but their transfer learning capabilities are still unclear. Recently, StyleGAN-based architectures did show evidence that they could establish themselves as proper base architectures in order to allow that. This paper helps to shed more light on those questions and try to show their strengths and weaknesses. One should recall that deep neural networks, such as GANs, are data-driven models and require datasets to be trained on.

Formally, let $G$ be the generator function and $D$ be the discriminator function. Both are incarnated as deep convolutional neural networks. $G: z \mapsto v$ is the network that produces an artificial image from a random noise $z \sim N(\mu, \sigma^2)$, where $N(\mu, \sigma^2)$ is a normal distribution defined with mean $\mu$ and standard deviation $\sigma$. $D: v \mapsto k$ is a neural network that tries to identify whether the given image $v$ is generated or drawn from the real world training data. $k$ is the label that indicates whether the $v$ is real ($v_R$) or fake ($v_F$), so $P(k = Real|k = Fake(G(z))) = D(G(z))$. Typically, the loss value of the generator would be the opposite of the discriminator loss, i.e., $L_G = -L_D$, which per se defines an instance of Adversarial Training [10].

In this setting, we optimize the discriminator weights $\theta_d$ by minimizing the binary cross entropy of the predictions for both real and generated images

$$\Delta \theta_d \frac{1}{m} \sum_{i=1}^{m} \left[ \log D(v_i) + \log(1 - D(G(z_i))) \right]$$ (1)

where $m$ is the number of instances in the mini-batch, and $v_i$ is the $i^{th}$ image drawn from the real data distribution $Z$, and $z_i$ is the noise sampled from $Z$ for that iteration. For optimizing the weights $\theta_g$ of the synthesis network, we use the opposite of the loss function for the $D$ as shown in Equation 2.

$$\Delta \theta_g \frac{1}{m} \sum_{i=1}^{m} \left[ \log(1 - D(G(z_i))) \right]$$ (2)

The overall optimization problem objective is then formulated as the following $\min \max$ game,

$$\min_D \max_G \mathbb{E}_{v \sim Z}[\log(D(v))] + \mathbb{E}_{z \sim Z}[\log(1 - D(G(z)))]$$ (3)

where the synthesis net is forced to generate images more alike to the real data distribution to fool the discriminator net – and at the same time, the discriminator has to improve to be able to distinguish between fake $v_F$ and real $v_R$ images.

Such training is much more unstable than supervised ones, and require a fine balance between the learning capability of both $D$ and $G$. There has been extensive work on improving GAN training stability, for instance Spectral Normalization, SAGAN, DCGAN, ProGAN and StyleGAN.
A. MAGICSTYLEGAN

In recent architectures, such as the StyleGAN, the noise vector \( z \in \mathbb{Z} \) is first projected via a mapping network \( f \) to an intermediate latent code \( w \in \mathbb{W} \), i.e., \( f : \mathbb{Z} \rightarrow \mathbb{W} \). Then, \( w \) is used in multiple layers to control the synthesis network \( G \). Such an approach has been widely used to allow style transfer networks [12] which gives the StyleGAN its name.

The subsequent version of StyleGAN, namely StyleGAN2 [13], brings some improvements to its predecessor and introduces larger networks, which with the aid of Progressive Growing, can stably be optimized to generate high-resolution images with sizes up to \( 1024 \times 1024 \). StyleGAN2 model has been shown by many researchers as a powerful tool for unconditional image generation [1], [2], [23], [24], [26], and it still matched its successor StyleGan3 in FID comparison [9]. We chose to adopt such architecture as the base for this study considering the generated images quality. Therefore, we leverage available pre-trained networks for evaluating the StyleGAN 2 transfer learning capability.

Considering that StyleGAN2 promotes disentangled and hierarchical noise introduction to the network, as well as improvements to function inversion, we believe that it is possible to finetune such a model so as to smoothly adjust the features learned by the model into the Magic illustrations domain. Models tuned in this strategy are called MAGICSTYLEGAN.

B. MAGICSTYLEGAN-ADA

Notably, there is a rather limited amount of Magic card illustrations. Thus, in the context of limited data, the Style- Gan2 model may suffer from discriminator overfitting, causing the feedback to the generator to become meaningless and the training to diverge. One of the most common alternatives to prevent overfitting is to adopt stronger data augmentation schemes. Note that in GANs, augmentation practices may affect the generated images, causing transformation leaking or making the discriminator gradient useless.

To circumvent those issues we employ the strategy presented by StyleGAN2-ADA [8] to train MAGICSTYLEGAN-ADA models. They use a wide range of invertible augmentations applied to both nets in an adaptive way. Invertible transformations have been shown to be non-leaking and by using them in both fake and real images the discriminator will not be able to simply overfit the augmentation style to discern between them. Such improvements may as well benefit transfer-learning capabilities in small dataset regimes, which is the case for generating Magic illustrations.

C. Transfer learning

One of the main goals of this paper is to understand whether pretrained StyleGAN-based ease the process of generating illustrations for the Magic cards. We expect them to be able to do a smooth domain shift towards the target application. To observe this effect, we have to register the progression of the generated images during each epoch of the finetuning phase. A smooth domain shift would be characterized by a sequential morphing from the original generated images towards the target domain. Note that if the training diverges, or the pretrained features are not useful, the network would most likely forget its original patterns, to learn new ones from the target data.

We suspect that by using the previous knowledge acquired from the pre-training phase as base to shift towards a new target domain will reduce the need for large datasets to achieve high quality image synthesis. In addition, it is likely the case that we would achieve significantly better results than training from scratch.

In our experiments, we proposed a variety of transfer learning scenarios to understand the impact of the pre-training data distribution. For instance, we uncover the impact of finetuning models that had source domains similar or very distinct to the target dataset. We also test cases where source and target datasets have similar or very distinct levels of diversity. Results of those experiments can be seen in Section VI.

IV. MTG DATASET

For training models that can generate Magic card illustrations it is necessary to have a dataset with real illustrations at hand. To the best of our knowledge there is a public dataset, referred as Magic1024 \( \times \) 1024\(^2\), that contains 49 thousand images from different card illustrations. We did preliminary experiments using a sample from that dataset, and all models failed to properly approximate the real data distribution. That dataset contains a diverse set of cards, but no information regarding card types and card content. Therefore, making it an overly complex function for the models to learn.

In order to fix the previously mentioned issues, we gathered a novel dataset that we hereby call MTG. It has more card illustrations samples (62,000 images versus 49,000), and also richer in metadata that allows us to precisely create subsets for specific concepts.

We built the MTG dataset by downloading card images from the Scryfall\(^3\) website. Then, we upscaled lower resolution images and extracted 1024 \( \times \) 1024-sized center crops from each illustration. Hence, we standarize all images to have the same resolution. Our dataset contains a multitude of card types (creatures, lands, artifacts) and colors (white, blue, black, red, green, and its combinations). The collected metadata dataframe contains 82 columns, such as: keywords, name of the card, oracle text, produced mana, color indicator, and many more. Generation of card illustration is one but many other possible tasks that can be handled using this dataset, such as: (i) text-to-image generation; (ii) card name prediction; (iii) modeling of oracle text; (iv) card authorship identification; and (v) keyword generation.

Focusing on the goal of this study, we created five main subsets out of MTG, namely Cards, Islands, Forest, Humans, and Faces. MTG-Cards is the complete dataset, otherwise referred simply as MTG; MTG-Islands has illustrations of “basic land” type cards named “Island”; MTG-Forest has

\(^3\)https://www.kaggle.com/joaopedromattos/magic-the-gathering-cards-
1024x1024

\(^3\)https://scryfall.com/
images of “basic land” type cards called “Forest”; MTG-Humans contains a reduced set of human-like creatures of the compound types: “human warrior” and “human soldier”; and MTG-Faces comprises manually extracted crops from the MTG-Humans illustrations focusing only on the character faces. See Table I for more details and statistics on the introduced subsets to allow scale comparison.

| Dataset              | Image Size | Samples |
|----------------------|------------|---------|
| Flickr-Faces-HQ (FFHQ) | 1024 × 1024 | 70,000  |
| AFHQ Dogs            | 1024 × 1024 | 5,000   |
| MetFaces             | 1024 × 1024 | 1,336   |
| Magic 1024×1024      | 1024 × 1024 | 49,000  |
| MTG-Cards            | 1024 × 1024 | 62,000  |
| MTG-Islands          | 1024 × 1024 | 614     |
| MTG-Forests          | 1024 × 1024 | 625     |
| MTG-Islands+Forests  | 1024 × 1024 | 1,239   |
| MTG-Humans           | 1024 × 1024 | 1,376   |
| MTG-Faces            | 1024 × 1024 | 1,161   |

It’s worth noticing that these datasets, although more constrained, still have great deal of variety in design and style from the various artists that created their illustrations throughout the years.

V. EXPERIMENTAL SETUP

In this section we explain the methodology for running all experiments, including baseline models, evaluation procedures and metrics, as well as the hyper-parameters used.

A. Baseline models

For the experiments listed in this paper, we trained GANs both from scratch and applying transfer learning on pre-trained models. In order to validate our hypotheses regarding the behavior and advantages of MAGICSTYLEGAN and MAGICSTYLEGAN-ADA we use DCGAN and StyleGAN trained from scratch as main baselines. These baselines can help us understand if it is advantageous to perform finetuning, and if ADA-related models perform better in limited data regimes.

For running the transfer learning experiments we used the StyleGan2 and StyleGan2 ADA pre-trained models made available by NVIDIA in their public Github repositories. Code was written in the TensorFlow framework. For training the MAGICSTYLEGAN model, we used the StyleGAN pre-trained with the FFHQ dataset as the backbone of our experiments. Training procedures with MAGICSTYLEGAN-ADA used a series of other trained models: FFHQ, AFHQ Dogs and MetFaces.

B. Hyper-parameters

All of our models were trained in single GPUs, with a batch size of 4, though with gradient accumulation for 8 steps to get gradients from 32 instances. We made use of Nvidia Tesla P100 and V100 GPUs. Complete finetuning procedure takes 24 hours in a single P100 GPU, while the V100 is almost three-fold faster. We estimate that all experiments took more than 480 GPU hours to run. All models were optimized with Adam with $\beta_1 = 0.0, \beta_2 = 0.99$ and learning rate of $2 \times 10^{-3}$ for both discriminator and generator networks. For data augmentation sampled images were randomly mirrored left to right, but not top to bottom. Augmentation hyperparameters used for MAGICSTYLEGAN-ADA were pixelblitting and geometric transformations. Generated images had truncation at 0.7.

C. Model evaluation

This study employs two widely used metrics for quantitative evaluation of GANs, specifically, FID and KID. FID is a well accepted metric for the context of GAN performance [6]. It uses InceptionV3 model to measure the distance between the generated and real images, as shown in Equation (4). Low metric values represent high similarity, and therefore
lower FID↓ values are better, as indicated by the down arrow. In this work, FID values were calculated for 10,000 generated images.

\[
FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}) \tag{4}
\]

Kernel Inception Distance (KID) is somewhat similar to FID, and describes the squared MMD (Maximum Mean Discrepancy) between Inception representations [7]. KID uses the polynomial kernel, defined in Equation (5), to avoid correlations with the objective of GANs and to avoid tuning any kernel parameter.

\[
k(x, y) = \left(\frac{1}{d}x^T y + 1\right)^3 \tag{5}
\]

KID estimates are unbiased, and standard deviations shrink quickly even for small datasets. Similar to FID, the lower the KID↓ value, the better.

**VI. EXPERIMENTAL ANALYSIS**

The experiments tested the generation of illustrations in a number of datasets with large aesthetics and pattern variation between them. In this section we discuss results achieved by MAGICSTYLEGAN and MAGICSTYLEGAN-ADA, that leveraged transfer learning from pre-trained StyleGan2 and StyleGan2 ADA models.

**A. Transfer learning yields far better results**

For the main results of the paper we used the FFHQ pre-trained StyleGan2 model as the source domain dataset. Interestingly, for most of the experiments FFHQ images are quite different: they differ in texture, proportions and semantics. When trained in MTG-Cards, our approach was able to produce good results despite the large diversity in the training data. It performed significantly better than DCGAN in terms of KID (Table II). Nonetheless, most of the generated images did not have the potential to be mistaken by real MTG illustrations. This shows that for better results, one should consider training separate models for each card type.

**B. Subsets allow for better generation**

After training with MTG-Cards, the next experiments tested more restricted datasets by using the created subsets: Islands, Forests, Humans, and Faces. Since these datasets have a small number of samples, we opted for the reporting StyleGan2 ADA architecture first. The curated output from the generated islands and forests can be observed in Figure 2. The training is successful in transferring the knowledge from the FFHQ source dataset to the target domains, even when the domains are very different. The generated images are far better looking and more consistent than the results from section VI-A, which suggests that creating per-concept subsets positively affect the result of the model training.

Furthermore, we ran an experiment that included both images from a combined version of MTG-Islands and MTG-Forests, that is the MTG-Islands+Forests. Figure 3 presents curated generated images from MAGICSTYLEGAN-ADA model trained in that data set. The model was still able to learn differences between the two concepts, and was able to generate both types of classes. That being said, the model was less consistent in generating high quality samples.
TABLE II
QUANTITATIVE RESULTS.

| Dataset           | MagicStyleGAN | MagicStyleGAN-ADA | DCGAN | MagicStyleGAN | MagicStyleGAN-ADA | DCGAN |
|-------------------|---------------|-------------------|-------|---------------|-------------------|-------|
|                   | FID↓          |                   |       | KID↓ ×10²     |                   |       |
| MTG-Cards         | 0.678         | 0.803             | 0.792 | 2.693         | 4.043             | 22.820|
| MTG-Islands       | 0.986         | 1.201             | 29.743| 1.457         | 3.581             | 700.400|
| MTG-Forests       | 1.250         | 1.547             | 6.237 | 2.243         | 4.581             | 142.140|
| MTG-Islands+Forests| 0.845       | 1.026             | 20.341| 3.115         | 3.505             | 230.830|
| MTG-Humans        | 1.141         | 1.038             | 8.331 | 5.442         | 3.080             | 264.280|
| MTG-Faces         | 0.747         | 0.840             | 1.649 | 2.327         | 4.872             | 31.880|

Fig. 4. Per epoch evolution of the generated image for the model MagicStyleGAN trained with MTG-Forest dataset. First to penultimate columns are from subsequent training epochs. The last one showcases the image generated by the end of the training procedure. Note: image was compressed due to maximum paper size policy for the submission.

C. Variety on target domain impacts generation quality

Perhaps the more challenging subset from MTG is the MTG-Humans. That is possibly the case given that humans do present themselves as complex forms composed by a variety of body parts, poses, movements, and clothing. Experiments ran in this subset demonstrated that MagicStyleGAN was able to learn some high level human features, though failed to draw them in detail and in a consistent fashion. The network is able to understand the basic structure of humans to a certain point, such as the position of the head and the presence of eyes. MagicStyleGAN-ADA was also able to synthesize the texture of armors (widely present in “MTG-Humans”). Though, all models failed to generate realistic samples, since this dataset is incredibly complex with substantial diversity among the dataset samples – specially when compared to the number of instances.

In contrast, MagicStyleGAN and MagicStyleGAN-ADA were able to quickly learn to generate proper samples for the “MTG-Faces” dataset, which is focused on generating new illustrations of character faces. Figure 3 shows faces generated by MagicStyleGAN-ADA. We observed that the manual crop of each face significantly helped the learning process of the network. That is expected once the face pattern becomes easier for the network to understand and to related to the origin features.

D. Effects of the base model domain

In this experiment we explore the relation between the source domain dataset and the results of the generated images. It has been pointed out that transfer learning success seems to depend primarily on the diversity of the source dataset, instead of similarity between subjects [8].

The results displayed in Table III show that the similarity does have an important effect on transfer learning. For MTG-Faces, the best results were achieved with the source domains of both faces in paintings and high quality faces pictures. Notably, the best results for MTG-Islands+Forests results are from models finetuned from FFHQ pretrained weights, which indicates that well trained models can provide useful features to a plethora of domains. Nonetheless, AFHQ Dogs also provided competitive results. Such results help demonstrating that transfer learning could be much more further explored in generative models.
TABLE III
EFFECTS OF THE SOURCE DOMAIN ON TRAINING MAGICSTYLEGAN-ADA.

| Dataset                | FFHQ | AFHQ Dogs | MetFaces |
|------------------------|------|-----------|----------|
| MTG-Faces              | 0.840| 1.038     | 0.804    |
| MTG-Islands+Forests    | 1.026| 1.111     | 1.463    |

VII. QUANTITATIVE RESULTS

We first report the quantitative metrics for our models, and a DCGAN baseline, considering all datasets. Table II depicts that MAGICSTYLEGAN and MAGICSTYLEGAN-ADA presented very similar performance, and both greatly outperformed the baseline trained from scratch. It is important to mention that some results can be counter-intuitive. For instance, StyleGan2-ADA trained from scratch using MTG-Forest got a FID of 1.498 and KID of 0.039, which is similar to the transfer learning results with the same setup. That being said, we noticed that the transfer learning model generates reasonable results much faster. In addition, experiments shown in Figure 6 clearly demonstrate that the projection of target data from MTG-Forest on the latent space of StyleGAN2-ADA and MAGICSTYLEGAN-ADA shows a big difference in quality, with the latter getting the upper hand.

Relative low scores for MTG-Cards and MTG-Humans were achieved on all GANs while a manual evaluation of the generated images points to a different conclusion, as we discuss in Section VIII-B. FID and KID rely heavily on InceptionV3 model [20], which loads weights pre-trained on ImageNet [5]. Therefore, it is unclear how well these metrics work for a domain such as MAGIC: THE GATHERING, that is rather distinct from the ImageNet-related ones. We aim to address this considerations in future works, maybe proposing a more reliable metric considering newer models, such as CLIP [15].

VIII. VISUALIZING THE LEARNED REPRESENTATIONS

In this section we analyze the trained models to evaluate their representation learning capacity. First, we validate our assumption that when finetuning StyleGAN-like architectures we can see a clear domain shift as the network leverages human-face like features to generate the target domain shapes. Then we evaluate the quality of the prior distribution when we try to find the noise vector $z$ that when fed to the synthesis network ($G$) generates the real world image.

A. Visualizing the representation shift

Figure 4 clearly shows that when using a model pre-trained to generate high-quality faces as base, the network learns to adapt such features as to generate shapes, colors and textures from the target domain. It is amusing to see that during the initial stages of the optimization it makes a mashup of both origin and target domains. As the training continues, the network forgets about face features and concerns more in generating detailed card illustrations.

It is relevant to notice that even for very distinct target domains, training MAGICSTYLEGAN and MAGICSTYLEGAN-ADA could in fact leverage from transfer learning based on StyleGan2 models. Our models were able to quickly understand the visual aspects of the proposed datasets and to generate possible illustrations for the game MAGIC: THE GATHERING.

B. Finding the noise vector for real images

Figure 5 shows the projection of target images onto latent space of MAGICSTYLEGAN-ADA. i.e., we optimize the noise vector $z$ that generates the closest image of the given real image. By analyzing the projection, it becomes clear that the network is able to generate better approximations when dealing with more uniform datasets like MTG-Forests, MTG-Islands and MTG-Faces. On the other hand, for models trained with more diverse data, e.g., MTG-Humans and MTG-Cards, representations fell short of their counterparts.

IX. LIMITATIONS

As discussed throughout this paper one of the main challenges for unsupervised generative models to face is the diversity within the datasets. Some models can leverage those when conditioned to classes [3], [14] or text [4], [19], [25]. Although, in this paper we made it evident that for finetuning purposes it is often paramount to present clearer patterns for the networks to achieve better results. Hence, image and pattern diversity is still a challenge in the field, specially in small data regimes. In addition, we note that even for very small datasets, such as MTG-Forests (600 instances), it was necessary at least a couple of hours for the model to generate reasonable illustrations – perhaps the area still lacks in efficient ways to do zero and few-shot transfer.

X. RELATED WORK

To the best of our knowledge, this is the first paper to handle the specific problem of card illustration generation for the
Magic card game. Most of the GAN work has been dedicated to generating images in more constrained datasets [8], [19]. Although, there is some similar related work to some extent. Generating images in a high diversity dataset has been explored by [23] in the context of illustrations for Yu-Gi-Oh card game. Training GANs with limited data is a topic that has receiving attention in recent works, such as [8], [21], [22].

[2] proposed finetuning StyleGAN2 for cartoon face generation with some transfer learning approaches. StyleGAN2 is also used in [11] for image-to-image translation. AgileGAN [18] uses the pre-trained weights from StyleGAN2 to create a framework that can generate high quality stylistic portraits via inversion-consistent transfer learning. Style transfer with StyleGAN has also been explored in [1] and [17].

XI. CONCLUSION

In this paper, we proposed MAGICSTYLEGAN and MAGICSTYLEGAN-ADA, which are finetuned incarnations of StyleGAN2 and StyleGAN2 ADA for generating MAGIC: THE GATHERING illustrations. To train those models we introduced MTG, a novel dataset that contains 62 thousand illustrations and it is rich in metadata. We show that our models achieve significant results even for very small datasets. Results depicted throughout this paper show that transfer learning is indeed an effective method for generating images even when the source and target domains are rather different from each other – though models do benefit from pre-trained weights in domains that share structural and aesthetic patterns. We clearly demonstrated that pre-trained features are reused and adapted during tuning. Our experiments also demonstrated that it is paramount to control data diversity for training in such small data regimes. An example of that is that we achieved far better image generative results when training separate models for more specific subsets. It is worth noticing that these models still take a considerable time to generate images even for small datasets. For future work, we aim to improve generative performance of MAGICSTYLEGAN, specially to generate more complex forms, such as Humans, Animals, Goblins and the like. To achieve that, we intent to explore Deep Probabilistic Models, as well as allow generation of illustrations conditioned to classes or text.

REFERENCES

[1] R. Abdal, Y. Qin, and P. Wonka, “Image2stylegan: How to embed images into the stylegan latent space?” 2019.
[2] J. Back, “Fine-tuning stylegan2 for cartoon face generation,” 2021.
[3] A. Brock, J. Donahue, and K. Simonyan, “Large scale gan training for high fidelity natural image synthesis,” arXiv:1809.11096, 2018.
[4] A. Dash, J. C. B. Gamboa, S. Ahmed, M. Liwicki, and M. Z. Afzal, “Tac-gan-text conditioned auxiliary classifier generative adversarial network,” arXiv:1703.06412, 2017.
[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
[6] M. H. et. al., “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” 2018.
[7] M. B. et. al., “Demystifying mmd gans,” 2021.
[8] T. K. et. al., “Training generative adversarial networks with limited data,” 2020.
[9] ——, “Alias-free generative adversarial networks,” 2021.
[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” NIPS, vol. 27, 2014.
[11] J. Huang, J. Liao, and S. Kwong, “Unsupervised image-to-image translation via pre-trained stylegan2 network.” 2020.
[12] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in CVPR, 2019, pp. 4401–4410.
[13] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of stylegan,” 2020.
[14] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier gans,” in ICML. PMLR, 2017, pp. 2642–2651.
[15] A. Radford and et. al., “Learning transferable visual models from natural language supervision,” in ICML. PMLR, 2021, pp. 8748–8763.
[16] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv:1511.06434, 2015.
[17] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, “Encoding in style: a stylegan encoder for image-to-image translation,” in CVPR, 2021, pp. 2287–2296.
[18] G. Song, L. Luo, J. Liu, W.-C. Ma, C. Lai, C. Zheng, and T.-J. Cham, “Agilegan: stylizing portraits by inversion-consistent transfer learning,” arXiv:2006.05338, vol. 2, p. 3, 2020.
[19] H.-Y. Tseng, L. Jiang, C. Liu, M.-H. Yang, and W. Yang, “Regularizing generative adversarial networks under limited data,” in CVPR, 2021, pp. 7921–7931.
[20] V. Vavilala and D. Forsyth, “Controlled gan-based creature synthesis via a challenging game art dataset – addressing the noise-latent trade-off,” 2021.
[21] X. Wu, Y. Wu, X. Li, X. Zhang, and X. Zhou, “Urine red blood cells generation using stylegan2 network,” 10 2020, pp. 244–248.
[22] T. e. a. Xu, “Attgan: Fine-grained text to image generation with attentional generative adversarial networks,” in CVPR, 2018.
[23] N. Yang, M. Zhou, B. Xia, X. Guo, and L. Qi, “Inversion based on a detached dual-channel domain method for stylegan2 embedding,” IEEE Signal Processing Letters, vol. 28, pp. 553–557, 2021.