IntersectGAN: Learning Domain Intersection for Generating Images with Multiple Attributes

Zehui Yao¹, Boyan Zhang¹, Zhiyong Wang¹, Wanli Ouyang²,³, Dong Xu², Dagan Feng¹
¹School of Computer Science, The University of Sydney, Australia
²School of Electrical and Information Engineering, The University of Sydney, Australia
³The University of Sydney, SenseTime Computer Vision Research Group, Australia
{zyao3112,bzha8220}@uni.sydney.edu.au, {zhiyong.wang,wanli.ouyang,dong.xu,dagan.feng}@sydney.edu.au

ABSTRACT

Generative adversarial networks (GANs) have demonstrated great success in generating various visual content. However, images generated by existing GANs are often of attributes (e.g., smiling expression) learned from one image domain. As a result, generating images of multiple attributes requires many real samples possessing multiple attributes which are very resource expensive to be collected. In this paper, we propose a novel GAN, namely IntersectGAN, to learn multiple attributes from different image domains through an intersecting architecture. For example, given two image domains \( X_1 \) and \( X_2 \) with certain attributes, the intersection \( X_1 \cap X_2 \) denotes a new domain where images possess the attributes from both \( X_1 \) and \( X_2 \) domains. The proposed IntersectGAN consists of two discriminators \( D_1 \) and \( D_2 \) to distinguish between generated and real samples of different domains, and three generators where the intersection generator is trained against both discriminators. And an overall adversarial loss function is defined over three generators. As a result, our proposed IntersectGAN can be trained on multiple domains of which each presents one specific attribute, and eventually eliminates the need of real sample images simultaneously possessing multiple attributes. By using the CelebFaces Attributes dataset, our proposed IntersectGAN is able to produce high quality face images possessing multiple attributes (e.g., a face with black hair and a smiling expression). Both qualitative and quantitative evaluations are conducted to compare our proposed IntersectGAN with other baseline methods. Besides, several different applications of IntersectGAN have been explored with promising results.

1 INTRODUCTION

In recent years, generative adversarial networks (GANs) have achieved promising progress on image generation due to its adversarial training strategy. A GAN model generally consists of two key components, a generator and a discriminator. The generator outputs generated samples (i.e., fake samples) that are as indistinguishable as possible from real samples, while the discriminator differentiates real samples from fake samples (i.e., generated samples) as much as possible. As a result, the generator is able to match the distribution of real samples.

Motivated by the success of generating synthetic data (e.g., digits [4]) for training a better classifier, various GANs-based frameworks have been proposed for different applications such as image style transfer [8, 29, 32], image super resolution [13], high resolution image generation [9, 27], text-based image generation [28, 31], faces image synthesis [2, 19, 26] and fashion designs [20, 34].

When generating images with multiple attributes using conventional GANs, we often need to collect real samples possessing all...
the required attributes and train the networks with the collected images or apply an extension of GAN to a conditional setting. However, as shown in Figure 1, when the number of specified attributes increases, in general, the number of samples decrease dramatically. It is challenging to collect a large number of real samples possessing multiple attributes for training deep networks, and also resource expensive to produce supervision labels. It is anticipated that it will be increasingly challenging when the number of attributes increases.

To overcome such limitation of the existing GANs, we propose a novel GAN model, namely IntersectGAN, to generate images with multiple attributes by learning the distribution of a new domain intersected from existing domains. For simplicity of explanation, we use two domains in the following discussions. Given two image domains \( X_1 \) and \( X_2 \) with corresponding attributes \( A_1 \) and \( A_2 \), respectively. The intersection \( Y = X_1 \cap X_2 \) denotes a new domain where images possess the attributes from both \( X_1 \) and \( X_2 \) domains, and the new intersected attribute is denoted as \( A_Y \). That is, the intersected domain \( X_1 \cap X_2 \) possesses a new attribute \( A_Y \) denoted as \( A_1 \cup A_2 \), which is the combination of both attributes \( A_1 \) and \( A_2 \). As illustrated in Figure 2, \( A_1 \) is \( \text{smiling} \) and \( A_2 \) is \( \text{male} \). The images in the intersected domain \( X_1 \cap X_2 \) possess these two attributes simultaneously. That is, \( A_{X_1 \cap X_2} = A_1 \cup A_2 = \{ \text{smiling, male} \} \).

As illustrated in Figure 3, when generating images with \( n \) specified attributes, our proposed IntersectGAN model consists of \( n+1 \) generators (i.e., one generator for each attribute and one generator for combined attributes) and \( n \) discriminators (i.e., one for each attribute). In comparison with conventional GANs, the discriminators are further challenged to be more discriminant by also taking the input from the intersection generator \( G_Y \). During the training phase, the intersection generator aims to produce content which is as indistinguishable as possible to all \( n \) input domains. We define the overall adversarial loss as the summation of the losses from \( n+1 \) generators. That is, the intersection generator \( G_Y \) is trained to fit the distribution of samples possessing the attributes from all the given domains. As a result, our proposed IntersectGAN is able to produce synthesized samples possessing multiple attributes without relying on real samples simultaneously possessing those attributes.

Furthermore, we explore the capacity of our proposed IntersectGAN in several other image generation tasks, such as generating images of a blended attribute by intersecting two opposite attributes (e.g., male and female) and content-aware domain intersection. In addition, by exploiting the interaction among three generators with inter-layer weight sharing, we are able to adapt the IntersectGAN to generate trio image samples, which better illustrates the idea of learning domain intersection.

In summary, the key contributions of our work are as follows:

(1) We propose a novel IntersectGAN model which is able to generate image samples possessing multiple attributes without using real samples simultaneously possessing those attributes. To the best of our knowledge, this is the first GAN model specially developed for generating images with multiple attributes from noise input by learning a domain intersection without introducing supervision labels.

(2) We formulate the intersection learning problem under the GAN framework to learn the distribution of an intersection of multiple domains. New adversarial training strategies are proposed to take into account the new architecture: the discriminators are trained against the parallel generators of individual domains plus the intersection generator while the intersection generation is trained against all the discriminators.

(3) We demonstrate the capacity of the proposed IntersectGAN in other image generation applications, such as generating images with a blended attribute and generating content-aware domain intersected images. In particular, a weight sharing architecture (as illustrated in dashed lines in Figure 3) is proposed to explore the relationship among generators and generate trio images. More details of weight sharing are explained in Section 5.3.

2 RELATED WORK

In this section, we organize GAN based image generation methods into two categories: image generation and image-to-image translation, in terms of whether a real input image is needed for the generator. The first category aims to generate image samples from a noise vector after learning the distribution of an image domain, while the second category aims to generate a new version of each given image by translating the input sample to another image domain.

2.1 Image Generation

GAN [4] was first proposed to learn a target distribution with an adversarial learning strategy, and has dramatically accelerated the progress on image generation. It has been widely utilized to generate image samples of the same style as of the given training image set. Various other improvements for GANs have been developed, such as network architectures, loss functions, and optimization strategies. For example, DCGAN [25] was proposed to boost the
Generative performance of GAN with a new architecture consisting of convolution layers instead of using max pooling or fully connected layers. Many GANs have also been proposed by using new forms of cost functions (e.g., L2 loss and L1 loss etc.) to boost the training process, such as LSGAN [21], WGAN [1], WGAN-GP [5]. However, these GANs take only one image set (i.e., image domain) and follow the one-to-one adversarial learning strategy, namely, one generator is trained against one discriminator. When generating images of multiple attributes through these GANs, real samples simultaneously possessing those attributes are required. However, building qualified dataset is expensive and sometimes even impractical.

In order to influence the content of a generated image in addition to the style, conditional GAN (cGAN) [22] was proposed by adding content labels as an additional input to both generator and discriminator. As a result, the content of generated images is supposed to be of the same as of the specified label, and the style of generated images will be of the same as of the given training images. However, as a supervised learning approach, even for its state-of-the-art cGAN with Projection Discriminator [23], cGAN requires a large number of samples with supervision labels, which are expensive to obtain. Our proposed IntersectGAN is able to address this limitation through its intersection architecture. That is, only the samples possessing individual attributes are required to train IntersectGAN for generating images with multiple attributes and no supervision labels are systematically needed.

2.2 Image-to-Image Translation

Image-to-image translation is a vision and graphics problem aiming to learn a mapping between an input image domain and a target image domain with two training datasets (one dataset for each domain). Pix2pix [8] was first proposed to perform one-way image-to-image translation through a supervised learning approach. A pair of samples (e.g., a sketch image and a color image) is required for both the generator and discriminator. In order to loose the requirement on paired training samples, various GANs have been proposed, such as UNIT [15] and CycleGAN [33], DiscoGAN [10], DualGAN [30]. However, these GANs are limited to translate a given image of one domain to an image of the other domain.

Three recently proposed GANs, IsGAN [24], conditional CycleGAN [18] and StarGAN [2], are able to translate a given image into an image with multiple attributes through conditional input labels. However, label information is required during the training as the conditional configuration. Differently, our IntersectGAN is proposed for image generation from noise input, instead of image-to-image translation, and supervision label information is not required during training.

3 INTERSECTGAN

In this section, we first present the formulation of IntersectGAN and then explain its implementation details.

3.1 Formulation

Given image domains $X_1, X_2, ... X_n$ specified with $n$ attributes $A_1, A_2, ... A_n$, respectively. IntersectGAN aims to learn a mapping from random noise to the target domain $Y = \bigcap_{1 \leq i \leq n} X_i$ where generated
samples are supposed to possess all the specified attributes. We denote the distribution of real image samples as $x_i \sim p_X$, for $1 \leq i \leq n$. In addition, the distribution of random noise is denoted as $z \sim p_Z$.

During the training phase, our model learns $n$ parallel mappings $G_i : Z \rightarrow X_i$ and an intersection mapping $G_{\text{Intersect}} : Z \rightarrow Y$. Meanwhile, there are $n$ discriminators $D_1, D_2, \ldots, D_n$ distinguishing between real and generated samples with regard to the corresponding domains $X_1, X_2, \ldots, X_n$, respectively. In the training stage, the expected output of a discriminator for a real image is set to 1, and that for a generated image is set to 0.

Motivated by GAN, our proposed IntersectGAN is formulated through the definition of adversarial loss. According to the objective of GAN, for an image domain $X$, a generator $G$ and a corresponding discriminator $D$, the adversarial loss function is defined as:

$$
\mathcal{L}_\text{GAN}(X, G, D) = \mathbb{E}_{x \sim p_X} [\log D(x)]
+ \mathbb{E}_{z \sim p_Z} [\log (1 - D(G(z)))] ,
$$

(1)

where $G$ is trained to generate synthetic images $G(z)$ possessing the attribute of domain $X$ while $D$ tries to differentiate real image samples $x$ from generated samples $G(z)$.

As illustrated in Figure 3, IntersectGAN has $n$ parallel generators $G_1, G_2, \ldots, G_n$ as well as $n$ corresponding discriminators $D_1, D_2, \ldots, D_n$. Their adversarial losses can be expressed as:

$$
\mathcal{L}_i = a_i \mathcal{L}_\text{GAN}(X_i, G_i, D_i),
$$

(2)

where $a_i$ indicates the objective importance of the $i$-th parallel generator.

The intersection generator $G_Y$ produces image samples containing all the specified attributes $A_1, A_2, \ldots, A_n$. Ideally, each discriminator $D_i$ cannot distinguish between the generated samples $G_Y(z)$ and real images from its corresponding domain $X_i$. Therefore, we can formulate the objective of $G_Y$ as a weighted sum of $n$ adversarial losses:

$$
\mathcal{L}_\text{Intersect} = \sum_{1 \leq i \leq n} a_i \mathcal{L}_\text{GAN}(X_i, G_Y, D_i),
$$

(3)

where $a_i$ is the importance constant mentioned above and $G_Y$ aims to simultaneously minimize two adversarial losses against both discriminators.

For the whole model, the overall adversarial loss is:

$$
\mathcal{L}_\text{IntersectGAN} = \sum_{1 \leq i \leq n} \mathcal{L}_i + \lambda_\text{Intersect} \mathcal{L}_\text{Intersect},
$$

(4)

where $\lambda_\text{Intersect}$ is a constant controlling the relative significance among different objectives. During the training phase, the optimization problem can be formulated as a minimax game where the generative model tries to minimize the objective while discriminators aim to maximize it, and the goal is to find appropriate parameters for the intersection generator. This is given by:

$$
G_Y^* = \arg \min_{G_Y} \max_{D_1, D_2, \ldots, D_n} \mathcal{L}_\text{IntersectGAN},\text{ where } 1 \leq i \leq n.
$$

(5)

The optimization aims to find an appropriate set parameters where the intersection generator can produce most realistic samples.

The core idea of the proposed IntersectGAN model is reflected in two parts:

1. The existence of parallel generators $G_i$ is for boosting their adversaries $D_i$ to better distinguish real samples and fake samples;
2. All the discriminators $D_i$ are the adversaries of $G_Y$, which allows the intersection generator to learn all the attributes presented in those given domains.

### 3.2 Implementation

**Revised Adversarial Loss Function** In order to stabilize the training outcomes and help models to converge better, we follow the strategy used in Wasserstein GAN [1] to define a new objective function by introducing the gradient penalty [5]. Then the loss function defined in Equation (1) can be re-formulated as:

$$
\mathcal{L}_\text{GAN}(X, G, D) = \mathbb{E}_{x \sim p_X} [D(x)] - \mathbb{E}_{z \sim p_Z} [D(G(z))]
- \lambda_{\text{pg}} \mathbb{E}_{x \sim p_X} \left[ (\|\nabla_x D(\tilde{x})\|_2 - 1)^2 \right],
$$

(6)

where $\tilde{x}$ is uniformly sampled from the linear space between a pair of real and generated images.

**Architecture** Deep convolutional neural networks [25] are used to implement IntersectGAN, which is constructed with $n + 1$ structurally identical generator networks and $n$ discriminator networks. When generating sample images with the size of $128 \times 128$, we fix the dimension of input noise to 32. All the generators consist of one fully connected layer and seven deconvolutional layers. All the discriminators contain six convolutional layers and two fully connected layers to output a single logit indicating the probability whether the input sample is real. Batch normalization [7] and layer normalization [14] are utilized for generators and discriminators, respectively. Note that the relationship among generators (discriminators) can be explored through weight sharing as explained in Section 5.3.

**Training** During the training phase, we apply random cropping and flipping to real input images in order to enhance sample variety. All the input vectors of random noise are sampled from the standard normal distribution. The batch size of input data is 32. We use Adam optimizer [11] to train the model with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. In addition, the initial learning rate was set to 0.0002. We train the model for 20,000 iterations and apply linear decay on the learning rate.
rate in the second 10,000 iterations to stabilize the training phase. We assume all the attributes are equally important (i.e., $\alpha_i = 1.0$ for all $1 \leq i \leq n$) and set $\lambda_{\text{Intersect}}$ and $\lambda_{\text{up}}$ to 1.0 and 10.0 respectively in all the experiments.

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we first present two experiments for qualitative evaluation: a preliminary experiment on generating colored squares, and a comprehensive experiment on generating face images with multiple attributes. Then a set of experiments are introduced for quantitative evaluation against several baseline algorithms.

4.1 Experiments

Generating Multi-Attribute Squares In the experiment, as illustrated in Figure 4, 5 specific visual attributes are utilized to describe the colored square images: vivid (i.e. the color of a square is closer to pure red), vivid background (i.e. the background color is closer to pure green), small, left and up. An image dataset of Colored Squares was purpose-built for validating the objective of IntersectGAN. It consists of 5 groups of images and each of them contain 2000 image samples of a colored square. Each group corresponds to one visual attribute and all the images in that group contain the a specific attribute and other attributes are randomly generated. All the samples are generated randomly by our computer program and all the random factors (e.g. color) are sampled from uniform distributions. It is likely that there is no exact the same image that appears in two image groups. In the experiment, the goal is to generate images possessing all the specific attributes after training our proposed IntersectGAN with the given training datasets. Sample results of each experiment can be found in Figure 6, which indicates that IntersectGAN is able to generate image samples with multiple specific attributes.

Generating Multi-Attribute Faces We use CelebFaces Attributes (CelebA) Dataset [17] to train baseline models for generating faces images with multiple attributes. There are 202,599 annotated image samples of celebrity faces in the dataset. According to the annotations, before training the model, we partition the dataset into...
different domains in terms of specified visual attributes. As shown in Figure 5, our proposed IntersectGAN is able to produce high quality two-attribute and three-attribute face images. In each row, the samples generated are gradually changed from one face to the other by linearly changing the input noise vector. For example, each of the two-attribute samples black hair + smiling clearly present such two attributes, although the gender attribute does not affect the results.

4.2 Metrics
Fréchet Inception Distance (FID) [6] As a quantitative measurement of GAN proposed recently, FID score is used to estimate the generating quality of baseline models by measuring the Fréchet distance between generated samples and real samples. For each face generation experiment with certain combination of specified attributes, we collect ground truths from CelebA datasets according to the given annotations and face image samples generated by each baseline model that is supposed to simultaneously contain all the specified attributes.

Perceptual Study As evaluating image realism and identifying facial attributes can be subjective, we conduct two perceptual studies with Amazon Mechanical Turk (AMT) system. To estimate the realism of the samples generated by each model, we obtain a Realism Score for each model by using AMT to perform human based perceptual survey. At first, 1000 groups of image samples were generated from different models (i.e., our proposed IntersectGAN and the baseline models) and each group contains one image sample generated by each model with the same noise input. In the survey, a Tucker was shown a number of image groups and asked to select the most realistic one from each group. To ensure the quality of data collection from Turkers, we set additional qualifications to only allow workers whose Approval Rate is greater than 95% to participate in the survey and also assign a few questions with apparent answers to validate their effort. After all the image groups are examined, we obtain the Realism Score for each model by calculating the percentage of sample images generated by the model that Turkers selected as the most realistic ones.

To further examine the presence or absence of specified attributes in the generated samples, we design another human based perceptual survey using AMT to obtain Attribute Score for each model. In the survey, a Tucker was shown a number of generated image samples and for each given sample the Tucker was asked to answer three multiple choice questions in regard to specific attributes including hair color, expression, and gender. For example, for gender attribute, the given choices are Male, Female and Cannot identify. In order to reduce the bias caused by random selection, in the experiment, 300 image samples were generated by each model for examination. After all the tasks are finished, we derive the Attribute Score for each model by calculating the percentage of image samples generated by the model that simultaneously possesses all the specified attributes.

4.3 Baselines
Conditional GAN with Projection Discriminator (cGAN PD) As one of the state-of-art GAN-based architectures with additional supervision inputs, conditional GAN with projection discriminator [23] is used as one of the baseline model. n labels (i.e., 0/1) are given to indicate whether an output image should respectively contain each attribute. As a supervised learning method, in its implementation, the label of each image needs to be available, which means that it has stronger or stricter requirements than ours.

GAN with Dual Discriminators (D2GAN*) This model has the identical structure with the D2GAN model proposed by Nguyen et al. [3]. Two discriminators in the original D2GAN learn from the same datasets. However, in the comparison experiments, they were trained from different image domains. When the number of specified attributes is more than 2, we increase the number of discriminators correspondingly. The comparison with this model can help tell how much the quality of generated image samples would increase when a model contains multiple parallel generators. For the sake of convenience, we denote the modified model as D2GAN* in this paper.

4.4 Comparisons
In each comparison experiment, we specify n attributes and collect real samples of these attributes to build the training datasets \(X_i\) for \(1 \leq i \leq n\). The CelebA dataset is used in all the following comparison experiments. For fair comparison, we use an identical architecture to implement all the models and train them under the same configuration.

For qualitative comparison, as shown in Figure 7, the image samples generated by our proposed IntersectGAN are generally of higher visual quality, while those by other baselines are of lower visual quality (e.g., artifacts and unnatural skin color). In addition, the image samples generated by our IntersectGAN present clear attributes than those generated by other GANs, as IntersectGAN aims to fool all the discriminators corresponding to each attribute.

For quantitative comparison, we conduct experiments with 5 combined attributes, including black hair + male, male + smiling, female + smiling, black hair + male + smiling, and black hair + female + smiling. As illustrated in Table 1, our proposed IntersectGAN achieves lowest FID scores in all the comparative experiments. D2GAN* is inferior in all the comparisons as well, which indicates the boosting effect of parallel generators. Without any supervision labels, our proposed IntersectGAN performs even better than cGAN PD for the lower FID scores. The comparison results indicate that our proposed IntersectGAN performs better for generating samples of higher quality and diversity than other baseline models.

For perceptual studies, as shown in Table 2, IntersectGAN achieves significantly higher Realism Score than other baselines models.
Figure 7: Sample multi-attribute face images generated by different GAN models trained on the CelebA images. The example on the left represents the combined attributes blackhair and smiling and the example on the right represents the combined attributes blackhair and female.

| Methods      | B+M | M+S | F+S | B+M+S | B+F+S |
|--------------|-----|-----|-----|-------|-------|
| cGAN PD      | 31.7% | 40.0% | 22.6% | 27.8% | 21.8% |
| D2GAN*       | 16.2% | 16.5% | 23.9% | 25.5% | 32.5% |
| IntersectGAN | 52.1% | 43.5% | 53.5% | 46.7% | 45.7% |

Table 2: Comparison results in terms of Realism Score for different attribute combinations.

| Methods      | B+M | M+S | F+S | B+M+S | B+F+S |
|--------------|-----|-----|-----|-------|-------|
| cGAN PD      | 80.6% | 69.0% | 83.0% | 41.6% | 46.6% |
| D2GAN*       | 52.6% | 41.0% | 68.6% | 23.6% | 20.6% |
| IntersectGAN | 87.0% | 81.3% | 91.3% | 47.6% | 53.6% |

Table 3: Comparison results in terms of Attribute Score for different attribute combinations.

which means that our IntersectGAN is able to generate more realistic image samples than other baseline models. It is expected that the advantage will be more obvious when the number of attributes increases.

As shown in Table 3, our IntersectGAN also achieves considerably high Attribute Scores in the experiments. Percentages of image samples possess all the specified attributes are given as a reference. Our proposed IntersectGAN achieves highest Attribute Scores in all the comparisons which are much higher than the ground truth percentage. It demonstrates the advantage of our model that it is able to synthesize high quality images of multiple attributes without relying on real image samples possessing those attributes. As for cGAN PD and D2GAN*, in addition to the potentially poorer capability to enforce output images to contain both specified attributes, the low realism of their generating samples can also negatively influence the Attribute Score.

5 OTHER APPLICATIONS

In this section, we explore the generation capacity of IntersectGAN for three interesting applications: generating images of a blended attribute from two opposite attributes, generating content-aware domain intersected images, and generating trio images.

5.1 Generating Images of Blended Attributes

Semantic attribute indicates the presence or absence of certain characteristics in an image, which can be regarded as a binary value. When these attributes are independent of each other, intersecting such attributes leads to the combination of these attributes. However, when such attributes contradict to each other, such as gender, intersection would lead to a blending effect on such attributes which is a unique outcome of our IntersectGAN, in comparison with cGAN and its variants.

In the experiment, we divide the CelebA dataset into two sets: male image set and female image set, according to the annotations.

As shown in Table 3, our IntersectGAN also achieves considerably high Attribute Scores in the experiments. Percentages of image samples possess all the specified attributes are given as a reference. Our proposed IntersectGAN achieves highest Attribute Scores in all the comparisons which are much higher than the ground truth percentage. It demonstrates the advantage of our model that it is able to synthesize high quality images of multiple attributes without relying on real image samples possessing those attributes. As for cGAN PD and D2GAN*, in addition to the potentially poorer capability to enforce output images to contain both specified attributes, the low realism of their generating samples can also negatively influence the Attribute Score.
This task is to demonstrate that the proposed IntersectGAN is able to produce high quality images possessing multiple attributes without using real samples simultaneously possessing those attributes. To the best of our knowledge, this is the first GAN model generating multi-attribute images from noise input without introducing extra supervision labels. Both qualitative and quantitative evaluations have demonstrated that our proposed IntersectGAN is able to produce high quality images possessing multiple attributes from separate image sets possessing each individual attribute, rather than from real image samples possessing multiple attributes which are generally expensive to collect. The generation capacity of the proposed IntersectGAN is further explored for three applications: generating images of a blended nature of IntersectGAN, it is expected that the generated image samples will possess both male and female attributes.

In other words, generated images of such two attributes would be gender-neutral. As shown in Figure 8, it is not easy to tell the gender of the generated image samples in the bottom row.

### 5.2 Generating Content-aware Domain Intersected Images

This task is to demonstrate that the proposed IntersectGAN is able to generate images with both content and attribute intersected from two domains. Assume we have two image domains $X_1$ and $X_2$ where $X_1$ contains images of content $c_1, c_2, \ldots$, and images in $X_2$ are of content $c_3, c_4, \ldots$. For both image domains, we denote the set of content items as $C_1$ and $C_2$, and the set of intersected content items as $C_1 \cap C_2$.

We train our model on the MNIST dataset [12] for this task. As shown in Figure 9, domain $X_1$ contains images of digits 0, 1, 2, 3, 4, 5, domain $X_2$ contains images of digits 6, 7, 8, 9, and the intersected domain $Y$ contains generated samples of 2 digits 0 and 1 only which is the intersection of two sets $\{0, 1, 2, 3, 4, 5\}$ and $\{0, 1, 6, 7, 8, 9\}$.

### 5.3 Generating Trio Images

To better visualize the objective of multi-attribute image generation by learning domain intersection, we change the network architecture of IntersectGAN to generate image samples in trio which are visually similar. We follow the idea from CoGAN [16] and share the weights in the first several layers of three generators in our model. In addition, discriminators $D_1$ and $D_2$ share the weights except the first two layers. This architecture allows IntersectGAN to generate samples in pair. Then for each input noise $z$, we can generate samples in trio denoted as $G_1(z), G_2(z)$ and $G_Y(z)$, respectively, where $G_1(z)$ and $G_2(z)$ are the samples produced by two parallel generators and the sample $G_Y(z)$ would contain the typical attributes of both domain $X_1$ and $X_2$. Ideally, all three generated samples are supposed to look visually similar in some ways.

In addition to the existing objective of IntersectGAN, to improve the correspondence of each trio of output samples, we introduce an additional loss for image generation, namely similarity loss denoted as a L1 loss:

$$L_{\text{Similarity}} = \mathbb{E}_{z \sim p_z} [||G_1(z) - G_Y(z)||_1] + \mathbb{E}_{z \sim p_z} [||G_2(z) - G_Y(z)||_1].$$

Its objective is to enforce the similarity between $G_Y(z)$ (the output of the intersection generator) and $G_1(z)$ and $G_2(z)$ (the outputs of the two domain generators). A coefficient $\lambda_{\text{Similarity}}$ is introduced to balance the importance of such similarity loss and we set it to 10 in our experiments. Without $L_{\text{Similarity}}$, although the generated images $G_Y(z)$ contain both specified attributes (e.g., black hair and smiling expression), they may look far from the corresponding samples generated by the parallel generators on other aspects (e.g., skin color and image background), as shown in Figure 10.

As shown in Figure 11, we generate several groups of trio-image samples with IntersectGAN. We observed that each trio set appears visually similar while the image samples in domain $X_1 \cap X_2$ still contains the attributes intersected from domains $X_1$ and $X_2$.

### 6 CONCLUSION

In this paper we present a novel IntersectGAN framework to learn intersection of multiple image domains for generating image samples possessing multiple attributes without using real samples simultaneously possessing those attributes. To the best of our knowledge, this is the first GAN model generating multi-attribute images from noise input without introducing extra supervision labels. Both qualitative and quantitative evaluations have demonstrated that our proposed IntersectGAN is able to produce high quality images possessing multiple attributes from separate image sets possessing each individual attribute, rather than from real image samples possessing multiple attributes which are generally expensive to collect. The generation capacity of the proposed IntersectGAN is further explored for three applications: generating images of a blended attribute, content-aware domain intersection based image generation, and trio-image generation. In our future work, we will further extend this model to deal with more attributes with more scalable network architectures.

### ACKNOWLEDGEMENTS

This research was partially supported by ARC (Australian Research Council) grant DP160103675. Dong Xu is supported by ARC Future Fellowship FT180100116. We would like to thank the NVIDIA Corporation for GPU support through NVIDIA Hardware Grant.
