www.igi-global.com

Data Augmentation Using GANs for 3D Applications

Ioannis Maniadis, Vassilis Solachidis, Nicholas Vretos, Petros Daras
Centre for Research and Technology Hellas Information Technologies Institute, Thessaloniki, Greece

ABSTRACT

Modern Deep Learning techniques have proven that they have the capacity to be successful in a wide area of domains and tasks, including applications related to 3D and 2D images. However, their quality is directly dependent on the quality and quantity of the data with which models are trained. This fact becomes increasingly relevant as the capacity of deep learning models increases, and data availability becomes the most significant obstacle with regard to their application. To counter this issue various techniques are utilized, including data augmentation. Data augmentation refers to the practice of expanding the original dataset with artificially created samples, in order to train a model with data interpretations that will hopefully equip it with better generalization properties. With regard to data augmentation, one approach that has been found to show great promise, are Generative Adversarial Networks (GANs). Unlike other methods that apply domain-agnostic transformations on the original data to produce new samples (e.g. noise, rotation, flip etc.), the GAN's objective is specifically to produce diverse samples that belong to a given data distribution. Taking advantage of this property, a multitude of GAN architectures has been leveraged for data augmentation applications. The subject of this paper is to review and organize implementations of this approach on 3D and 2D imagery, examine the methods that were used, and survey the areas in which they were applied.

Keywords: Generative Adversarial Networks, GAN, Data Augmentation, Adversarial Learning, Domain Transfer GAN,

INTRODUCTION

The advances that have been made in the field of deep learning have provided us with ever more potent tools, which can be applied in an increasing number of tasks, computer vision being foremost among them. Concurrently, the potential value of data has become apparent, and so data gathering and mining are now employed in several domains in order to make it possible for deep learning techniques to be applied in those domains. Deep learning models require data for their training which constitute a representative sampling of a given task. When the available data only relate to a subset of cases, a model will only learn to address those cases only and fail in the task overall. For this reason, deep learning models generally require significant amounts of data for their training. Nonetheless, in many cases the available data is not sufficient for training models that generalise adequately. That may occur either because data gathering might be difficult in a given setting (due to scarcity of subject cases or due to difficulties in collecting them) or because the available data are not annotated. In all the above cases, other methods must be employed to make deep learning feasible.

Several techniques have been developed in order to tackle the above-mentioned problems, particularly when dealing with 2D and 3D image data. One point of focus is to develop architectural modifications that make models generalize better in a given task, such as dropout and weight regularization (Sutskever, Hinton, Krizhevsky, & Salakhutdinov, 2014). Another approach suggests expanding the initial dataset by manipulating the existing data and creating new synthetic samples. This approach is usually refered to as data augmentation. The most frequent implementations of data augmentation are the addition of random noise to the data and the application of geometric and/or other transformations (Taylor & Nitschke, 2019). The latter is particularly effective in image data, whose features have spatial properties. Those data augmentation methods however, while suitable for image data, are domain-agnostic, since they apply transformations without taking into account the nature, characteristics and features of the original data, and produce synthetic samples that could deviate from the original distribution.

In order to achieve these two objectives, that is to augment a dataset with meaningfully and significantly diverse samples, a method would be required that augments a dataset in ways specific to its properties, so that the generated samples would cover the largest area of the sample space possible, without deviating from it. GANs (Goodfellow et al., 2014), (Zoumpourlis, Doumanoglou, Vretos, & Daras, 2017), (Shijie, Ping, Peiyi, & Siping, 2017) as is demonstrated in (Shijie, Ping, Peiyi, & Siping, 2017), can be used to augment data in this exact way, and so they are an attractive alternative to the above domain agnostic methods. Ideally, GANs produce samples that belong to the original data distribution, while at the same time they differ from any given sample of that distribution. In this way they fulfill the essential objective of data augmentation, which is to provide to the model a diverse sample pool with which to train, which is representative of a given task. Additionally, the fundamental formulation of GANs has proven to be remarkably flexible, in that GANs can be modified to generate samples in many different ways, and can be combined with a variety of architectures to tackle different data augmentation tasks. However, GANs are also remarkably hard to train (Goodfellow I. , 2016), and so have been the subject of intense study in an effort to develop mode efficient architectures (Arjovsky & Bottou, 2017), (Salimans et al., 2016). It is important to note at this point that unlike other GAN review papers, such as (Z. Wang, She, & Ward, 2019) and (Pan et al., 2019), this work does not provide a broad review of GAN models. Rather, it limits its purview to cases where GANs have been used in the context of 2D and 3D image data augmentation for the purpose of improved performance in classification, segmentation, object detection/identification and motion tracking tasks. This work studies these cases with regard to the GAN architecture that was used, the domain in which it was used, and the specific way it was leveraged to augment the available data.

Considering the volume of our findings and their diversity, the authors decided that the first step of our survey should be to provide the reader with a detailed description of the significant terms and symbols that will be used in this work (Section *Background*). Following that, the authors will outline the tasks for which GAN-based data augmentation was utilized, the types of GANs encountered, and the ways in which data augmentation could be done depending on the dataset and the problem in question (Section *Tasks & Augmentation Techniques*). Then, the studied specific GAN models are described with regard to their function and architecture (Section *GAN models*), Finally, the domains and applications in which GAN-based data augmentation was used are presented, organised according to their domain (Section Applications). Table 1 summarises our findings by providing an extensive listing of the works encountered, including their domain of application, the task in question, as well as the GAN and dataset(s) used. With regard to the distinction between 2D and 3D data, while applications that deal with

3D data are noted as such, it must be stressed that, as will be made apparent in section *GAN models* and subsection *3D Generative Adversarial Network (3D-GAN)* in particular, the transition from 2D to 3D data generation usually requires no significant architectural modifications and the working principles of each model are maintained.

BACKGROUND

Notation and Definitions

Labeled sample

In the context of this work, it is considered that a sample is labeled when it is paired with the information we are interested in and that we expect a trained model to be able to infer, depending on the task in question. That might be the sample's class(es), a segmentation mask, the location of objects etc.

Conditioning input

In the earliest GAN formulations, the input, based on which samples are generated, is a random vector. However, later models introduced conditionality. In those models, the input includes user defined information which aims to restrict generated samples to specific subspaces. This is referred to as conditioning input. For example, if a sample is generated with conditioning information that corresponds to a specific class c, we expect that sample to belong to that class. The term *conditioning input* will then refer to use-defined information fed to a GAN, which restricts a synthetic sample to a specific subset of the datasets' broader distribution.

True (real) & fake (synthetic) samples

True (or real) samples are those that belong to the original dataset in each given task. Fake (or synthetic) samples are those that were generated by a GAN or other augmentation approach.

Noise vector z

Most GAN variants make use of a noise vector z as a latent input variable from which to draw to generate diverse samples. This vector is frequently drawn from a uniform distribution $N(0,1)$ and its length depends on the architecture in question. It is also referred to as noise prior, symbolised with $p(z)$.

Domain

A domain can be thought of as a specific subset of existing data. Each domain's properties may be defined by annotations (e.g. classes) or other less discernible properties determined by the samples themselves. Broadly speaking, a domain is defined by the common attributes of the group of samples it consists of. This definition becomes particularly relevant in domain transfer applications (see section *Domain Transfer (DT)*), where models are used to transform samples from one domain to another, without user defined knowledge of the properties of each domain.

Data augmentation

Data augmentation been established as an effective method to improve a model's performance in various tasks with regard to generalization. The term data augmentation relates to a number of techniques used to expand a dataset with artificially created samples, so that a model trained with that dataset will generalize better. Data augmentation has been found to be particularly effective in image-related tasks, where label-preserving transformations such as rotation, flipping, noise, cropping etc. have been studied and applied to great success (Simard, Steinkraus, & Platt, 2003). These methods are used to artificially expand datasets to assist the model in learning more robust interpretations of the available data. Building on these augmentation approaches, more complex techniques have been proposed over time, such as Colour

Jittering, Edge Enhancement and Fancy PCA (Taylor & Nitschke, 2019), which, except for the latter, are again mostly applicable to image-related tasks. Finally, there are also algorithms such as SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) and its variants (Bunkhumpornpat, Sinapiromsaran, & Lursinsap, 2012), which artificially expand a dataset by synthesizing new samples based on the original dataset's feature space. This technique has the advantage of being domain-agnostic, in that it can be used on non-image datasets. Finally, generative models such as GANs and Variational Autoencoders (VAEs) (Kingma & Welling, 2013) can also be used to create synthetic samples for data augmentation purposes. It is important to note that the above approaches are not necessarily mutually exclusive, and can be used in various combinations at a given task.

Generative Adversarial Networks

Generative Adversarial Networks (GANs) were first proposed in 2014 by (Goodfellow et al., 2014). They are a class of generative networks whose fundamental operating principle is that of a competition between two distinct models, the generator and the discriminator, that are trained jointly. The original GAN's architecture is demonstrated in Figure 1, and their training process is as follows. In each iteration, the generator produces a batch of synthetic (fake) samples. The discriminator is then fed these samples, along with true samples from the original dataset. The objective of the discriminator is to accurately distinguish true from fake samples, while the generator's objective is to generate fake samples that fool the discriminator. That is, samples that the discriminator falsely assigns *True* labels to. This function, which amounts to a minimax game between the two components, ideally results in a discriminator having learned to perfectly discern true from fake samples, and a generator that still manages to fool it, which would mean that it has learned to create perfectly realistic samples. At this point we consider the GAN to have converged. This process is illustrated in Algorithm 1 below. It should be noted that the above description corresponds to the GAN as described in (Goodfellow et al., 2014). Subsequent works have modified this approach is significant ways, however the principle of the adversarial game between the generator and the discriminator remains the central idea in which all GAN variants are rooted.

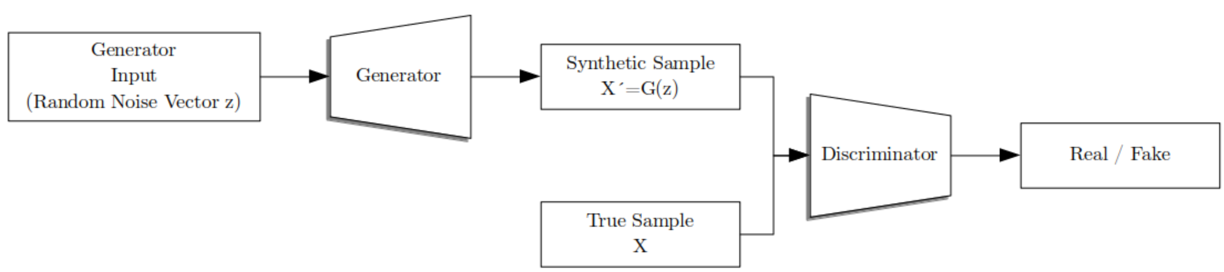

Figure 1 ***GAN Architecture****: The discriminator receives true samples X and fake samples X', which, through its training, it learns to distinguish and assign appropriate True and Fake labels to. The generator is provided with a noise vector z, and is trained to generate sample X', such that the discriminator would label as True. Through their iterative training, as described in Algorithm 1, the generator learns to create samples X' that are realistic enough to be indistinguishable from true samples X by the discriminator.*

Algorithm 1
Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, *k*, is a hyperparameter whose value is set to *k=1* based on experimentation.

for number of training iterations **do**
    for k steps **do**
        - Sample minibatch of *m* noise samples $\{z^{(1)}, \ldots, z^{(m)}\}$ from noise prior $p_g(z)$.

- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{data}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$$

end for

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(z^{(i)}))$$

end for

This algorithm amounts to a minimax game of:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \sim p_{data(x)}}[\log D(x)] + \mathbb{E}_{z \sim p_{data(z)}}[\log(1 - D(G(z)))] \qquad (1)$$

GANs present a simple yet powerful concept which, particularly after it was expanded on in ways that are examined in section *GAN models*, can be leveraged to create high quality, realistic and diverse data. Specifically from a data augmentation perspective, GANs can be used to generate synthetic samples that share the specific properties of the original dataset, while at the same time being more diverse. Also, while GANs have been used to generate various types of data, including text and sound, most GAN models focus on image data generation, which means that there is a wide array of approaches that can be used for augmenting 2D and 3D image data. Those that we found to have been utilized in this context are presented in section *GAN models*. It should also be noted that GANs do not prohibit the use of other augmentation methods (see subsection *Data Augmentation*) and can be employed in conjunction with them. GANs, however, also have significant drawbacks that have been the focus of extensive study, in an effort to identify and mitigate them. The most significant of those drawbacks are presented below.

Mode collapse

A frequent problem in GAN training, mode collapse refers to the cases where the generator only produces a very small number of distinct samples. Those samples might fulfill the objectives defined by the GANs architecture, that is to be realistic enough to fool the discriminator, which means that the GAN shows signs of convergence. However, the fact that the synthetic samples are not diverse means that the model overall has failed. Mode collapse is then an issue that relates with the quantity of diverse samples that the GAN synthesizes.

Overfitting

Overfitting occurs when the discriminator learns to only assign true labels to samples that are almost identical to those in the original dataset. This, in turn, forces the generator to produce samples that are, in fact, near exact copies of the original dataset. Although overfitting is a problem that seems similar to mode collapse, it is not, in fact these drawbacks are mutually exclusive. Their difference is that given N real samples, in the case of mode collapse the GAN produces $M \ll N$ distinct samples, each of which may or may not be identical to some sample in N. When a GAN overfits, it produces as many as $M \approx N$ distinct samples, where for each synthetic sample $m \in M$, there exists an almost identical sample $n \in N$. In that sense, overfitting relates to the diversity of synthetic samples compared to the original dataset.

Non-convergence

The adversarial nature of a GAN's training process means that both the generator and discriminator have no fixed objective. Rather, they receive feedback from each other. It is then possible they reach a point where neither provides the other with useful information and both diverge from the point of convergence. This development is referred to as non-convergence, a point at which both discriminator and generator no longer provide each other with beneficial feedback and the synthetic samples deteriorate. It is common for GANs to deteriorate in this way if they keep being trained after having converged.

Diminished & exploding gradients

The nature of GANs means that the generator and discriminator depend on each other for guidance. However, it is possible possible for either one to significantly outpace the other (Arjovsky & Bottou, 2017). In this case, the component that has fallen behind may not be able to perceive a viable path through which to progress. A generator might produce samples more realistic than the discriminator can identify, which usually leads to its gradients increasing rapidly and it being unable to make meaningful updates. Conversely, the discriminator might discern true from synthetic samples so well that the generator is unable to identify ways to fool it, which leads to its gradients converging to zero and it not being trained at all. In both cases, training fails.

Sensitivity to hyper-parameters & computational load

In addition to the above issues, GANs are particularly difficult to be designed with regard to their hyperparameters. Due to their unsupervised training, significant experimentation is required to identify the optimal design for a given task. Also, given the minimax nature of GANs' function, even when a successful tuning has been identified, a small change in hyperparameters might disturb the balance at some stage of the training process, and lead to any of the problems outlined above. This is an ongoing challenge, though significant progress is being made, such as in the case of (Gong, Chang, Jiang, & Wang, 2019), which applied Neural Architecture Search (NAS) (Elsken, Metzen, & Hutter, 2019) in the context of GANs.

Lack of evaluation metrics

Evaluating the output of a GAN is a major problem in evaluating their quality. The discriminator only provides a relative evaluation of how realistic a sample is. It is then a statement related to the state of the generator at each time, rather than relevant to the actual quality of the synthetic samples. The diversity of synthetic samples is also hard to bequantified. Considerable work has been done in this area (Lucic, Kurach, Michalski, Gelly, & Bousquet, 2017), (Shmelkov, Schmid, & Alahari, 2018) and several metrics have been proposed, most prominently the Inception Score (IS) (Heusel, Ramsauer, Unterthiner, Nessler, & Hochreiter, 2017), the Frechet Inception Distance (FID) (Salimans et al., 2016) and the Classification Accuracy Score (CAS) (Ravuri & Vinyals, 2019). Regardless, evaluating GANs remains a significant open problem. In the specific context of data augmentation however, the criterion that is used is simply how much the performance is improved in a given task when a given data augmentation method is applied, compared with using only the original data or other augmentation techniques.

Various proposals have been made proposed to address the above-mentioned issues and some of them are presented on Section *GAN models*, along with the respective models. It should be noted that, while some approaches have been proven to be more effective than others, there has yet to emerge a clear consensus regarding optimal GAN training parameters, and finding such a consensus remains the subject of a number of works (Arjovsky & Bottou, 2017), (Lucic et al., 2017), (Salimans et al., 2016), (Kurach, Lucic, Zhai, Michalski, & Gelly, 2018), (Odena et al., 2018).

TASKS & AUGMENTATION TECHNIQUES

In this section definitions will be provided with regard to the tasks in which data augmentation was found to have been applied and the approaches that were used. Given the quantity and diversity of the works examined in this survey, this section is important in understanding *GAN models*, as well as Table 1, where these works are enumerated for the reader to examine collectively.

Tasks

Classification (CL)

The model is trained to assign the correct class label(s) to the samples it examines. While there are many ways to augment a dataset in the context of a classification task (see *Samples Generation Method* subsection), the most frequent way is to use GANs that generate labeled synthetic samples through some conditional input.

Object Detection (OD)

The model is expected to locate user-specified objects in an image. To do this, the model is most often expected to define a space with a bounding box, in which space the object has been found to exist. This task may or may not include an object recognition aspect, wherein objects are not only located, but also classified.

Segmentation (SG)

The model is expected to detect the exact shape of an object in a given image. This can be seen as an expansion of object detection, where not only the position but also the shape of an object are requested. In most cases, this amounts to creating a segmentation mask, by assigning a label to each pixel in an image, where pixels that belong to the same object have the same label. Data augmentation in this case is done by generating samples paired with their corresponding desired segmentation masks.

Object Tracking (OT)

An extension of object detection, the task of object tracking requires for the model in question to not only be able to locate an object in a given image, but also to track that object's movements in the various frames that make up a video sequence (Doumanoglou, Vretos, & Daras, 2019).

Person Re-Identification (PID)

Given an image of a specific person and a set of other images, the model is expected to identify that person in those images (if indeed they appear in them). It differs from classification in that models that do classification learn to assign a finite number of identities to the samples they examine. Rather than assign identities, PID models learn to use people's images to detect if those people appear in other samples.

Samples Generation Method

A significant distinguishing factor among GAN variants relates to their inputs and, more specifically, if and how those inputs apply restrictions to the generated samples. Three broad categories are presented: a) GANs generating samples unconditionally, b) GANs generating samples conditionally via some conditioning input and c) and Domain Transfer GANs, which do not generate entirely new synthetic samples, but rather a transition of an input sample to another domain. Finally, considering the increasing importance of 3D applications, as well as the fact that the handling of 3D imagery differs in some cases to that of 2D, cases where 3D imaging is either the input or output of the examined GAN will be noted with (*3D*).

Unconditional Samples Generation (USG)

Unconditional Samples Generation refers to the approach by which samples are generated independent to any label or other condition, only under the constraint that they belong to the sample space which is defined by the dataset used to train the GAN. GANs adhering to this approach generally use a noise vector as input. It should be noted that "Unconditional" does not require synthetic samples to be unlabeled. It only means that the user cannot influence the kind of sample (e.g. with regard to its class) the GAN will generate. While this approach seems impractical in the context of data augmentation, several methods were found to have been used to apply it. 1) A separate unconditional model is trained to generate samples for each label (USGa). While this approach could be effective in producing labeled samples, it becomes computationally impractical as the number of distinct labels increase. 2) Samples are generated unconditionally, drawing from the entire dataset in question, and are labeled by a model trained on that original dataset. The model is then re-trained or fine-tuned with some combination of original and synthetic data (USGb). 3) The labels are themselves included in the samples (USGc). This method is, for example, applicable when each sample is an N dimensional vector, in which case m class labels can be included via concatenation. It can also be applied in segmentation tasks, where the segmentation mask may be generated along with the original image via depth-wise concatenation. In adding the labels to the samples, the assumption can be made that a realistic synthetic sample is one whose content and label are a) realistic and b) match each other. 4) For a dataset with *n* classes, all generated samples are assigned class memberships of *1*/*n* for each class and are then used along the original dataset (USGd). This methods is limited to classification tasks.

Conditional Samples Generation (CSG)

Per this approach, in order to produce synthetic samples the GAN receives (alongside other possible inputs, most often a noise vector) some conditioning input, which restricts the sample space to which the synthetic sample is expected to belong. Depending on the task in question, the conditioning input might be related to a particular class, the location and/or ID of an object, information regarding segmentation etc. This method is also the simplest one conceptually in the context of data augmentation, in that it produces labeled synthetic samples. It should be noted that, while the conditioning input is usually expressed as a vector, it can also be represented in other ways. An example is DAGAN, where the generator receives a sample image as input, and is expected to generate a sample of the same class as that image. In this case, we consider the conditioning input to be the image itself, as a representation of its class.

Domain Transfer (DT)

Also referred to as image translation when applied to images, this approach to generate samples is distinct in that the generator's inputs include the same kind of data as those that it is expected to produce (e.g. *Image* → *Image*). Its function is to receive input sample *X* that belongs to a domain *A*, and generate sample *Y* that belongs to a different domain *B*. An example would be to have generic images of *angry faces* (domain *A*), and expect them to be changed to *happy faces* (domain *B*). Some GANs in this category also enforce that transferred samples *Y* differ from their origin samples *X only* in the properties required to make the transition *A* → *B*. In the prior example, that would mean that for each *angry face*, we expect its transition to *happy face* not to alter that face's distinguishing features and for the person to remain identifiable. We will refer to this property as *identity-preservation*. Regarding cases where the GAN in question allows for Domain Transfer to any one of multiple domains, which are chosen in accordance with some conditioning input, we will refer to their function as Conditional Domain Transfer (DTc).

A significant subcategory of domain transfer applications relates to transitioning visual data from 2D to 3D or the inverse. These applications (Alexiadis et al., 2016) are rapidly becoming more relevant in terms of their academic and industrial impact, due to the increasing number of 3D applications being developed and deployed, which increases the availability of 3D datasets, as well as the requirement for additional annotated data. We will refer to this function as Domain Transfer 2D to 3D (DT2t3) when the GAN in question transitions samples from 2D to 3D, Domain Transfer 3D to 2D (DT3t2) when the transition is from 3D to 2D, and Domain Transfer 2D and 3D (DT2a3) when the transition can happen in either direction.

Augmentation Approach

A final distinction that should be made in this section relates to the specific ways in which a dataset may be augmented. Augmentation is not necessarily done in a uniform way. Depending on the dataset, a model may benefit from augmenting only specific sample subsets, or by augmenting various subsets to uneven degrees. As such, two distinct approaches can be discerned.

Uniform Data Augmentation (UA)

In this approach, samples are generated uniformly, without distinctions being made among existing subsets (i.e. labels). An open question within the context of this approach regards to the ratio of real to synthetic samples. A fixed figure does not exist and is determined in a case-to-case basis via experimentation, with values ranging from *0.2* to *10* synthetic samples for each real one. This is the most frequent approach and, as such, in *Applications* section it will be assumed to be the one implemented, unless stated otherwise.

Dataset Balancing (DB)

In cases where, in a dataset, some labels have considerably fewer samples than others, this imbalance may significantly degrade a model's performance in a given task. In those cases, data augmentation can be used to generate samples that belong to those specific labels' distributions, in order to mitigate this imbalance, a practice with is referred to as dataset balancing. GANs have been used in this way, and can perform this function via all three sample generation methods described later in *Samples Generation Method* subsection. In terms of the quantity of synthetic data, the most frequent approach is to generate enough so that the minority class(es) have approximately the same number of samples as the other(s).

GAN MODELS

In this section various GAN architectures are examined. Those architectures have been selected either because they have been applied in Data Augmentation tasks with 3D or 2D imagery, and so are of direct interest to this paper, or because their presentation is considered to be fundamental in understanding subsequent GAN models. It should be understood that each model will not be discussed in depth and small variations of each that might have been applied will be presented, but will not be separately analyzed. The objective of this section is to provide sufficient technical information for each GAN model for the reader to be able to understand their functions and how they were used in each application. Subsequently, not all GANs that are presented in this work will be analyzed in equal length. We will expand only on the models that we consider to be the most significant in terms of their impact on the field, their novelty, or the frequency with which they were found to have been used. Also, in order to maintain cohesion, GANs will be organised in accordance with the way they generate samples, a distinction made in *Samples Generation Method* subsection. While this distinction is not absolute and in the case of some GANs it is unclear whether their function is closer to CSG or DT (e.g. the TAGAN and GANs used for content infilling), we believe it is a strong taxonomy criterion for the purposes of this survey.

Unconditional GANs

For the original GAN, proposed in (Goodfellow et al., 2014), see *Generative Adversarial Networks* in the *Background* section.

Deep Convolutional GAN (DCGAN)

At the time they were first formulated, GANs had significant difficulties in incorporating deep convolutional architectures. This issue was tackled by the Deep Convolutional GAN (DCGAN), proposed in (Radford, Metz, & Chintala, 2015). The authors studied the use of CNNs in GANs, proposed guidelines for stable training, and suggested an architecture template. The DCGAN was notable for using no fully connected layers in its generator (seen in Figure 2), instead relying entirely on convolutions. The guidelines that were suggested in this work are as follows:

- Rather than pooling layers, use strided convolutions in the discriminator and fractional-strided convolutions in the generator.
- Use Batch Normalization in both the generator and the discriminator.
- Use no fully connected hidden layers for deeper architectures.
- Use ReLU activation in the generator for all layers except for the output, which uses Tanh.
- Use LeakyReLU activation in the discriminator for all layers.

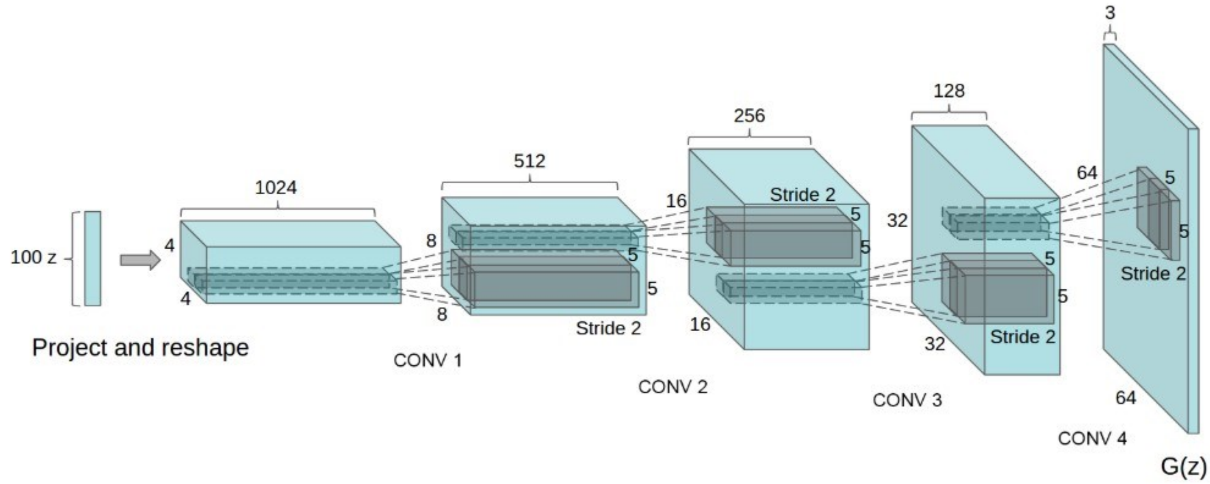

Figure 2 DCGAN generator used for LSUN scene modeling. The discriminator's architecture follows the same architectural principles as a standard deep convolutional classifier, subject to the guidelines outlined for DCGANs

3D Generative Adversarial Network (3D-GAN)

Based on the DCGAN, (J. Wu, Zhang, Xue, Freeman, & Tenenbaum, 2016) proposed 3D-GAN, the first instance of a GAN being used to create synthetic 3D images. Its architecture is similar to the DCGAN, the most significant alteration being that 3D-GAN used volumetric convolutional layers to produce 3D samples (Figure 3). Its impact is significant in that it proved that GANs can also be applied in the field of 3D imagery. A conditional variant called 3D-CGAN was later suggested by (Jin, Xu, Tang, Harrison, & Mollura, 2018), which, rather than creating samples by drawing from a random vector, uses as its input a distorted real sample. In the context of the original work, the 3D-CGAN is used for content infilling, where the distortion takes the form of cropping specific areas of the original samples.

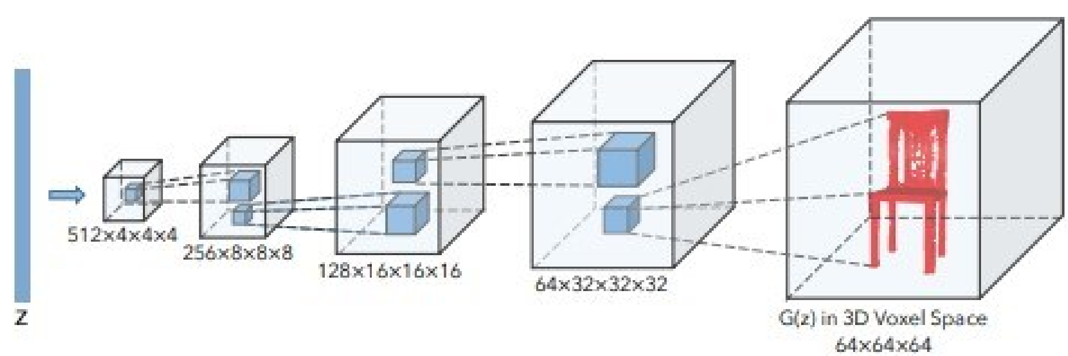

Figure 3 3D-GAN generator. The discriminator mostly mirrors that architecture

Wasserstein GAN (WGAN) & WGAN-Gradient Penalty (WGAN-GP)

In (Arjovsky, Chintala, & Bottou, 2017), the authors observed that, up until that point, the objective functions used by GANs were limited to variations of the Jensen Shannon and Kullback-Leibler divergences. However, the paper proves that the Earth Mover (EM) distance provides superior convergence properties, and is thus more sensible to use for GANs. To enforce the K-Lipschitz discriminator requirement of the EM distance, the paper proposes that 1-Lipschitz continuity be enforced via weight clipping. Thus, the paper concludes to a novel GAN formulation, the Wasserstein-GAN (WGAN), which is trained per Algorithm 2, and displayed far superior results with regard to its stability and the quality and diversity of its generated samples compared to other contemporary models.

Algorithm 2

WGAN algorithm. Suggested default values by (Arjovsky & Bottou, 2017) are $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{critic} = 5$

Require: α, the learning rate. c, the clipping parameter. m, the batch size. n_{critic}, the number of iterations of the critic per generator iteration. w_0, initial critic parameters. θ_0, initial generator's parameters.

while θ has not converged **do**

for $t = 0,\ldots,n_{critic}$ **do**

- Sample $\{x^{(1)},\ldots,x^{(m)}\} \sim \mathbb{P}_r$, a batch from the real data.
- Sample $\{z^{(1)},\ldots,z^{(m)}\} \sim p(z)$, a batch of prior samples.
- $g_w \leftarrow \nabla_w \left[\frac{1}{m}\sum_{i=1}^{m} f_w\left(x^{(i)} - \frac{1}{m}\sum_{i=1}^{m} f_w(g_\vartheta(z^{(i)}))\right)\right]$
- $w \leftarrow w + a \cdot RMSProp(w, g_w)$
- $w \leftarrow clip(w, -c, c)$

end for

- Sample $\{z^{(1)},\ldots,z^{(m)}\} \sim p(z)$, a batch of prior samples.
- $g_w \longleftarrow \nabla_\theta \frac{1}{m}\sum_{i=1}^{m} f_w(g_\vartheta(z^{(i)}))$
- $\theta \longleftarrow a \cdot RMSProp(w, g_\theta)$

end while

Algorithm 3

WGAN with gradient penalty (WGAN-GP). Suggested default values by (Gulrajani, Ahmed, Arjovsky, Dumoulin, & Courville, 2017) are $\lambda = 10$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$, $n_{critic} = 5$

Require: The gradient penalty coefficient λ. The number of critic iterations per generator iteration n_{critic}. The batch size m. Adam hyperparameters α, β_1, $\beta_2.\alpha$. Initial critic parameters w_0. Initial generator's parameters θ_0.

while θ has not converged **do**

for $t = 1,...,n_{critic}$ **do**

for $i = 1,...,m$ **do**

- Sample real data $x \sim \mathbb{P}_r$, latent variable $z \sim p(z)$, a random number $\varepsilon \sim U[0,1]$
- $\tilde{x} \leftarrow G_\theta(z)$
- $\hat{x} \leftarrow \varepsilon x + (1-\varepsilon)\tilde{x}$
- $L^{(i)} \leftarrow D_w(\tilde{x}) - D_w(x) + \lambda(\| \nabla_{\hat{x}} D_w(\hat{x}) \|_2 - 1)^2$

end for

$w \leftarrow Adam(\nabla_w \frac{1}{m}\sum_{i=1}^{m} L^{(i)}, w, a, \beta_1, \beta_2)$

end for

- Sample a batch of latent variables $\{z^{(1)},...,z^{(m)}\} \sim p(z)$.
- $w \leftarrow Adam(\nabla_w \frac{1}{m}\sum_{i=1}^{m} -D_w(G_\vartheta(z)), \theta, a, \beta_1, \beta_2)$

end while

The weight clipping approach is problematic in the blunt way in which it manipulates the model's weights, which (Arjovsky & Bottou, 2017) acknowledges. (Gulrajani et al., 2017) sought to solve this issue and proposed a modification of the WGAN which used Gradient Penalty (WGAN-GP). In their work, they proposed imposing the Lipschitz constraint through an additional objective in the GAN's discriminator loss function, as can be seen in the WGAN-GP's training Algorithm 3 above. This modification was proven to fulfil the Lipschitz constraint requirement, as well as to provide significantly improved performance compared to the original WGAN.

It should be noted that the WGAN and WGAN-GP essentially propose methodologies that enforce constraints which improve GAN training, more so than they constitute distinct GAN models in themselves. As such, these approaches can be combined with other conditional and unconditional GAN architectures, which will in fact be the case for many of the applications that will be examined in Section *Applications*. Until spectral normalization was proposed by (Miyato, Kataoka, Koyama, & Yoshida, 2018), the WGAN-GP's formulation served as the basis for most subsequent GAN variants.

Other Unconditional GANs

Laplacian GAN (LAPGAN). As mentioned earlier, training GANs with deep CNNs was particularly difficult before the DCGAN proposed a concrete methodology for doing so. Prior to that, LAPGAN (Denton, Chintala, Szlam, & Fergus, 2015) proposed overcoming that difficulty by leveraging the methodology of the Laplacian Pyramid (Burt & Adelson, 1983) to combine multiple shallow, and thus easier to train, CNNs in a single GAN architecture. Per this approach, LAPGAN consists of a cascade of small GANs using CNNs. A smaller sized sample is generated by the first GAN and is progressively enlarged as it passes through the pyramid until it reaches the desired level. Each GAN contributes via the application of a mask to the sample, so that images of higher resolution and quality are produced progressively.

Progressive Growing GGAN (PGGAN). The Progressively Growing GAN (Karras, Aila, Laine, & Lehtinen, 2017) expands on previous works that attempted to use GANs in a cascade structure. However, rather than training multiple GANs, PGGAN consists of a single generator-discriminator pair. Both models are initially shallow. As training progresses, however, more layers gradually are added to both, so that they *are both mirror images of each other and grow in synchrony*. This continues until the generated samples has the desirable properties in terms of dimensions, diversity and realism. A conditional variant of the PGGAN is proposed in (Han, Murao, et al., 2019).

Consistency Term GAN (CT-GAN) Proposed in (Wei, Gong, Liu, Lu, & Wang, 2018), it builds on the WGAN-GP by modifying its loss function. Specifically, it introduces a component named consistency regularization which, introduced to the GP loss, is claimed by the authors to enforce Lipschitz continuity in a way that improves the model's performance, particularly with regard to avoiding overfitting.

WaveGAN Introduced in (Donahue, McAuley, & Puckette, 2018), WaveGAN is the result of an effort to make advances with regard to using GANs for the generation of sound. The authors did so by building on DCGAN and WGAN-GP to propose an architecture more appropriate to this task, by modifying the components proposed by the DCGAN, using WGAN-GP's training objectives, and suggesting that the discriminator use the proposed *Phase Shuffle* method for evaluating samples.

Conditional GANs

Conditional GAN (cGAN)

Following the publication of the first GAN, (Mirza & Osindero, 2014) proposed the Conditional GAN, which could produce samples dependent on user defined information, as seen in Figure 4. The cGAN approach modifies the original minimax game from equation 1 to:

$$min_G max_D V(D,G) = \mathrm{E}_{x\sim p_{data}(x)}[\log D(x|y)] + \mathrm{E}_{z\sim p_{data}(z)}[\log(1 - D((G(z|y)))] \quad (2)$$

Per the cGAN formulation, synthetic samples are not generated randomly, rather they are generated based on conditioning input vector y. Realistic samples are then required be the discriminator to be such that they belong to the specific distribution defined by the conditioning input y.

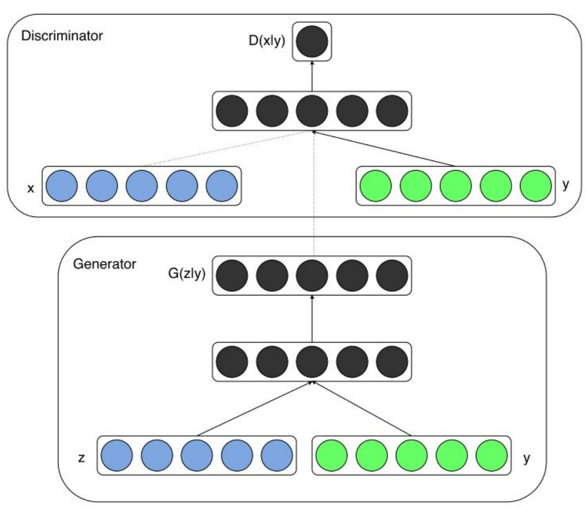

Figure 4: cGAN Architecture: The generator is provided with a noise vector z, drawn from a uniform distribution, and a vector y, which includes the information to which samples must be conditioned, which it uses to generate sample x. x is then passed on to the discriminator, along with information y, which determines whether x is real or synthetic.

Auxiliary Classifier GAN (AC-GAN)

Seeking to improve on the cGAN in the task of generating synthetic samples using a conditional input, (Odena, Olah, & Shlens, 2016) introduced the Auxiliary Classifier GAN. Similar to the cGAN, this model suggests that the generator use a random noise vector z and a conditioning input vector y in order to produce synthetic samples $X_{fake}=G(y,z)$. Where the AC-GAN deviates from the cGAN is that, given a sample X, the AC-GAN's discriminator produces both an assessment S={True,False} regarding how realistic X is, and an estimate C, representing the sample's class. The two outputs $D(X) = P(S|X), P(C|X)$ of the discriminator are used to estimate the log-likelihood of the correct source L_S and that of the correct class L_C, as displayed below in the equations below:

$$L_S = E[\log P(S = real|X_{real}] + E[\log P(S = fake|X_{fake}] \quad (3)$$

$$L_C = E[\log P(C = y|X_{real}] + E[\log P(C = y|X_{fake}] \quad (4)$$

Working with these equations, the discriminator D is trained to maximise L_C+L_S, and the generator G is trained to maximize L_C-L_S. In that way, the discriminator's objective is changed to assign a *True* or *Fake* label to each sample, but also to assign a correct class label. Concurrently, the generator's objective is now to not only produce realistic synthetic samples, but also to provide samples appropriate to its conditioning input y. This modification, combined with deeper convolutional architectures such as those suggested in (Radford et al., 2015), yielded significantly improved results compared to previous approaches.

Conditional Variational Auto-encoder GAN (CVAE-GAN)

CVAE-GAN (Bao, Chen, Wen, Li, & Hua, 2017) is an architecture that draws from both Conditional GANs and autoencoders. Specifically, it combines a GAN and a Variational Autoencoder (VAE) into a single architecture, as seen in Figure 5. The model consists of an encoder (E), a generator (G), a discriminator (D), and a classifier (C), which are trained jointly (contrary to most GAN architectures where the discriminator and generator are trained separately). Following the notation of Figure 5, the encoder is given data sample x and it's corresponding class c. It then maps x to an encoding z, through a learned distribution $P(z|x,c)$, which is then used by G, along with c, to generate a synthetic sample x'. Subsequently, G and D functions as they would in a GAN. D learns to identify synthetic samples and G tries to create samples that can fool D. Finally, the Classifier C tries to accurately assess the class c that samples x and x' correspond to.

Other than its architecture, the model is notable in that it is trained with a combination of six loss functions. They each apply to specific components of the model and combined aim to ensure that the model is stable in its training and produces diverse, realistic and class-appropriate samples.

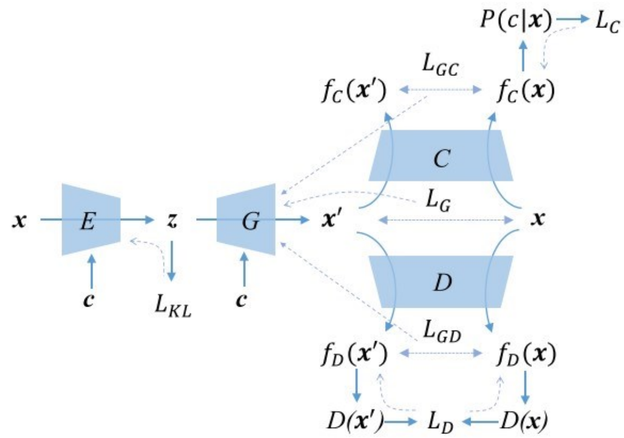

Figure 5: Overview of CVAE-GAN, including its losses.

Data Augmentation GAN (DAGAN)

Data Augmentation GAN (DAGAN) was proposed by (Antoniou, Storkey, & Edwards, 2017) as a GAN aimed specifically towards Data Augmentation, and was tested on Few-Shot and One-Shot Learning classification tasks. It's architecture and function can be seen in Figure 6. It is an example of a GAN where the generator's conditional input is not a vector, but rather a sample. DAGAN receives samples as inputs and generates synthetic samples that match each input sample's class, through the process described in Figure 6. The fact that the DAGAN uses a real sample as input to produce class preserving synthetic samples, means that it can also augment classes it has not been trained on, and thus can tackle one-shot learning problems as well.

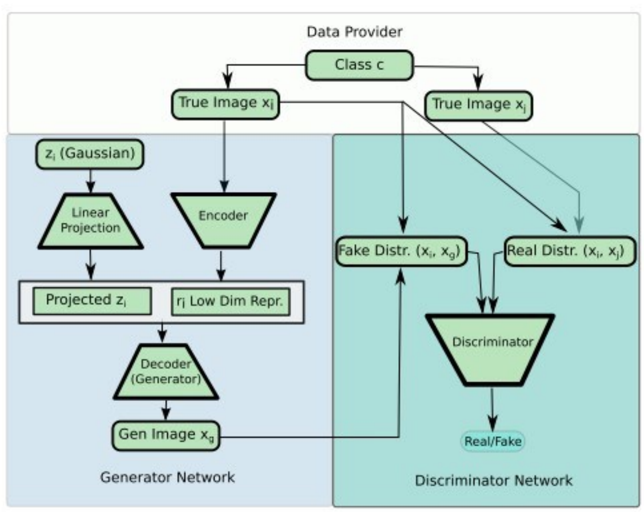

Figure 6: Overview of DAGAN. For each class c for which a synthetic sample is to be generated, a pair of random samples $(x_i,x_j) \in c$ are chosen. The generator then produces a synthetic sample $x_g = G(x_i,z)$, where z is a random noise vector. The samples are evaluated as real or fake by the discriminator, which receives either Fake pair (x_i,x_g) or Real pair (x_i,x_j). In this way the DAGAN's architecture promotes the generation of realistic, conditional samples that avoid mode collapse, all while using only the standard adversarial loss (the paper uses the WGAN-GP loss proposed by (Gulrajani et al., 2017)

Other Conditional GANs

Deep Adversarial Data Augmentation (DADA). Expanding approaches proposed in (Salimans et al., 2016) and (Odena et al., 2016), (X. Zhang, Wang, Liu, & Ling, 2018) proposed a novel GAN based data augmentation framework. DADA recommends that the GAN's discriminator not assign $k + 1$ probabilities to each sample (where k probabilities correspond to each of k classes and the $(k + 1)$th probability corresponds to a *True*/*Fake* label). Rather, it was proposed that the model and its loss function be modified so that the discriminator would assign $2k$ probabilities, where to each class c_k corresponds 2 possible outcomes $c_{k_{True}}$ and $c_{k_{Fake}}$.

Multiple Distribution GAN (MD-GAN). Proposed by (Yirui Wu, Yue, Tan, Wang, & Lu, 2018), the MD-GAN enforces conditionality by drawing its input from multiple distinct distributions per label via a Gaussian Mixture Model, rather than a noise vector z drawn drom a single distribution. This modification, the paper claims, also results to more diverse samples.

3D Multi-Conditional GAN (3D MCGAN). The 3D MCGAN (Han, Kitamura, et al., 2019) is proposed for 3D conditional infilling. It's generator receives 3D images, areas of which are replaced with noise, and which are concatenated with conditioning information along the input samples' 4th dimension. The model is expected to fill the specified area with content that is realistic and appropriate to the conditioning input. It uses two discriminators, the first of which evaluates whether the sample as a whole is realistic, and the other determines whether the infilling matches the conditions provided.

MetaGAN. In (R. Zhang, Che, Grahahramani, Bengio, & Song, 2018), MetaGAN is proposed, which leverages Meta-Learning approaches, such as MAML (Finn, Abbeel, & Levine, 2017) and RN (Sung et al., 2017), to tackle supervised and semi-supervised few-shot learning tasks, both in the sample level (when the dataset in question includes unlabeled samples) and in the task level (when the dataset includes multiple tasks, of which some are unlabeled). MetaGAN is also noteworthy in that it trains the classifier as part of the adversarial process, theorizing that even imperfect synthetic samples will provide beneficial information to the classifier.

Balancing GAN (BAGAN). Proposed as a tool to balance datasets with minority classes by (Mariani, Scheidegger, Istrate, Bekas, & Malossi, 2018), the BAGAN utilizes the architectural similarities between autoencoders and GANs to leverage the advantages of both. Specifically, autoencoders are easier to train, but GANs produce more diverse samples. Per BAGAN's proposed methodology, an autoencoder is trained first. Then, the weights are used to initialize a GAN, which is subsequently trained adversarially. Notably, for n classes the discriminator has $n + 1$ outputs, n of which correspond each of the classes, and the $(n + 1)th$ indicates a *fake* sample. Additionally, the BAGAN uses the trained autoencoder's encoder to develop a class-conditioned input vector generator, which provides the GAN with input samples drawn from class-dependent distributions. The resulting architecture can generate synthetic samples conditionally and, per the paper, is easy to train due to the strong initialization point provided by the autoencoder.

Three-Player GAN. This model, suggested in (Vandenhende, De Brabandere, Neven, & Van Gool, 2019), is used to augment data in the context of classification tasks. The three components of the model are a generator, a discriminator and a classifier which, unlike the AC-GAN, are distinct architectures. Per the proposed method, the generator and discriminator are pre-trained as a conditional GAN. After that, the classifier is included in the training. The classifier is trained with both original and synthetic samples, while the generator is concurrently trained to produce samples that are harder for the classifier to classify. It is important to note that in each step both models are trained.

Conditional Infilling GAN (ciGAN). CiGAN (Yirui Wu et al., 2018) tackles the task of conditional image infilling. That is, to fill specific areas of a given image with content that is a) realistic and b) appropriate to conditions that ciGAN received as input. To achieve this, ciGAN's inputs consist of a sample from the original dataset, with the content of the area to be filled replaced by random noise, a class-related conditioning input, and a mask which identifies the area to be filled. This input information is fed to various stages of the generator as the sample is gradually upscaled to match the desired dimensions. CiGAN also uses a slightly modified loss function compared to regular GANs, more suitable to the task in question.

Adversarial in-painting based framework (AIPBF). This model, suggested in (Jie Yang et al., 2018), was unnamed by the writers of the paper. Used for conditional infilling of 3D images, it's inputs are a class label and a 3D image with a mask which determines the area to be filled. The model has two generators that function sequentially in a coarse-to-fine scheme and two discriminators that both promote realistic and class adhering sample generation, though one of the two enforces those qualities locally and the other globally.

Domain Transfer GANs

pix2pix

The pix2pix architecture was first presented in (Isola, Zhu, Zhou, & Efros, 2017) and was one of the first GAN models to tackle the image-to-image translation task. Its objective is to establish a way for images to transition from one domain to another (e.g. grayscale to colored, Google Maps to aerial photo etc.).

For the pix2pix framework to function, pairs of the same sample in both examined domains are needed. The generator G receives a sample x from source domain X and generates sample $\hat{y} = G(x)$, which we expect to be a realistic interpretation of x in target domain Y. The discriminator D then attempts to distinguish *true* pairs (x,y) from *fake* pairs (x, $\hat{y}$ of samples using the loss function in equation 5 below. To improve the quality of generated images, another component is added to the model's objective, which tasks the generator with producing samples $\hat{y}$ that not only succeed in fooling the discriminator, but that are close to the ground truth image y in terms of their L_1 distance (equation 6). These two components combine to form the model's final objective, as seen in equation 7, where in the paper $\lambda = 100$. It is also notable that pix2pix uses a Markovian discriminator (C. Li & Wand, 2016) and U-NET generator (Ronneberger, Fischer, & Brox, 2015).

$$L_{cGAN}(G,D) = E_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + E_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1 - D(x,G(x,z)))] \tag{5}$$

$$L_1(G) = E_{x,y\sim p_{data}(x,y),z\sim p_z(z)}[\| y - G(x,z) \|_1] \tag{6}$$

$$G^* = arg\ min_G\ max_D\ L_{cGAN}(G,D) + \lambda L_{L_1}(G) \tag{7}$$

Cycle-GAN

Pix2pix is limited by the fact that it requires pairs of the same content in both the source and target domains. Expanding on pix2pix, Cycle-GAN (J. Y. Zhu, Park, Isola, & Efros, 2017) suggested an approach that allows for domain transition without using paired samples. Furthermore, it allows for transition from each domain to the other, rather than from the source to the target domain only.

Given two domains X and Y, with samples x and y respectively, the model consists of two mapping functions (generators) $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and their corresponding discriminators D_Y and D_X, which encourage G and F to generate samples indistinguishable from those in domains Y and X respectively. Each pair (G,D_y) and (F,D_X) have their own adversarial loss objective, per equation below.

$$L_{GAN}(G,D_Y,X,Y) = E_{y\sim p_{data}(y)}[\log D_Y(y)] + E_{x\sim p_{data}(x)}[\log(1 - D_Y(G(x))] \tag{8}$$

The paper also introduces the Cycle Consistency Loss, an objective that tasks the generators with not only producing samples that belong in their respective target domain, but that also correspond to their input sample from the source domain. This is achieved by forcing them to satisfy *backward cycle consistency*: $y \to F(y) \to G(F(y)) \approx y$ and $x \to G(x) \to F(G(x)) \approx x$. The loss function for this objective can be seen in equation 9.

$$L_{cyc}(G,F) = E_{x\sim p_{data}(x)}[\| F(G(x)) - x \|_1] + E_{y\sim p_{data}(y)}[\| G(F(y)) - y \|_1] \quad (9)$$

The above combine for the Cycle-GAN's full objective function (in experiments $\lambda = 10$) and its corresponding minimax formulation in equations below. The model's function can also be seen in Figure 7 below.

$$L(G,F,D_X,D_Y) = L_{GAN}(G,D_Y,X,Y) + LGAN(F,D_X,Y,X) + \lambda Lcyc(G,F) \quad (10)$$

$$G_*,F_* = arg\ min_{G,F}\ max_{D_X,D_Y} L(G,F,D_X,D_Y) \quad (11)$$

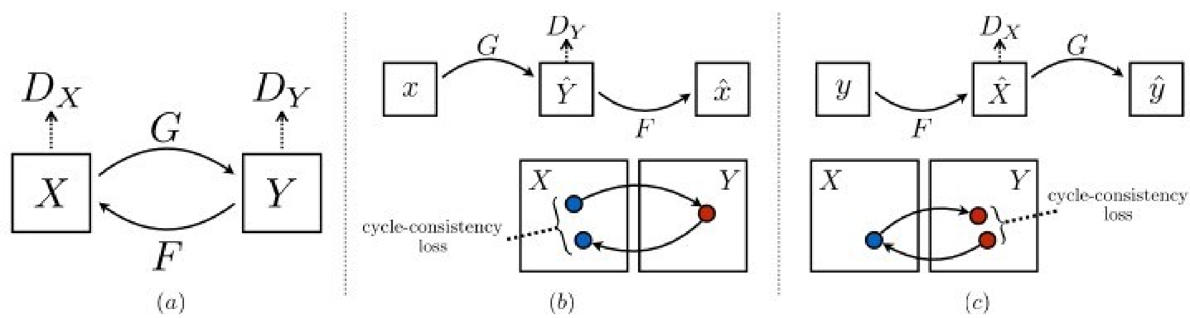

Figure 7: The Cycle-GAN's function. (a) An overview of the model with regard to its two mapping functions $G : X \to Y$ and $F : Y \to X$, and their corresponding discriminators DX and DY . (b) The forward cycle-consistency loss $x \to G(x) \to F(G(x)) \approx x$. (c) The backward cycle-consistency loss $y \to F(y) \to G(F(y)) \approx y$

The Cycle-GAN served as the basis for a number of variants such as the covariance-preserving conditional cycle-GAN (cCov-GAN) (Gao, Shou, Zareian, Zhang, & Chang, 2018), which expanded on the Cycle-GAN's objective to allow for conditional domain transfer and preserving intra-class covariance information, as well as other variants that will be examined in the next subsection.

Other Domain Transfer GANs

StarGAN Proposed by (Choi et al., 2018), StarGAN expands on the Cycle-GAN. Each Cycle-GAN is limited to domain transfer between two specific domains X and Y. However, StarGAN allows for conditional transition to multiple domains. Trained with data from N domains, it can transfer any sample x from domain $X \in N$ to any of domain $Y \in N$. To achieve that, during training it receives a sample as input, but it also receives a label indicative of the domain the sample is expected to transition to. A classification error is also incorporated in the training process to facilitate this function, that is to restrict each generated sample to a domain determined by a conditioning input. StarGANs allow for N to N domain transition with only one model, whereas the use of Cycle-GANs would have required the training of a distinct model for each domain pair.

A StarGAN variant named CTGAN was also encountered in (Zhou, Ke, & Luo, 2019), where an *ID consistent* loss is added to the objective in the form of an L_1 distance between the synthetic image and its original.

DavinciGAN. In (K. Lee, Choi, & Jung, 2019) the DavinciGAN is proposed as an alternative to Cycle-GAN for unpaired image-to-image translation. Its contribution lies in the use of loss functions that promote the generation of samples that are realistic with regard, belong to the desired domain, and whose background (i.e. the parts of the image that are irrelevant to the source and target domains) has not been altered. The later objective is achieved via a combination of unsupervised segmentation and attention. The DavinciGAN is then particularly suitable for applications which require that only very specific sections and/or objects of each image should be altered as the image transitions between domains, with the rest of it remaining relatively intact.

Tonality-Alignment GAN (TAGAN). Proposed in (L. Chen et al., 2018), TAGAN is a domain transfer GAN designed to augment 2D and 3D hand gesture datasets for the purpose of hand pose estimation. It consists of a generator and discriminator. The generator receives as input the shape drawing of a palm in a given gesture and a color map, to generate a 2D or 3D rendering a realistic palm with the corresponding gesture and color properties. The TAGAN uses multiple objectives to encourage the generation of samples that are realistic and maintain the required shape and color consistency. It should be noted that, while we group TAGAN as a domain transfer GAN, in the sense that it transitions samples in the domain of shape drawings to that of images of hand gestures, it could also be thought of as a conditional GAN, generating samples based on the conditioning input of the shape drawing and the color map.

AugGAN. AugGAN is a variant of the Cycle-GAN suggested in (S. W. Huang et al., 2018) meant for use in the context of object detection and segmentation tasks. In AugGAN's architecture, the generator produces a segmentation mask along with the synthetic image, which is guided by a loss function to correspond to a specific object. Via this modification, AugGAN learns to transition samples between domains without distorting the shape and placement of the object that is under examination.

Differential GAN (D-GAN). Proposed in (Gu, Kim, Kim, Baddar, & Ro, 2017) and focused on facial expression alteration, D-GAN transfers an image x from domain X to image y at domain Y, as defined by a conditional input, using pairs of the same image in both domains as ground truth for training. The mask is created by projecting a class conditioning vector to the shape of the image in a learned way, so that it can be depth-wise concatenated with the image to be modified. The resulting tensor includes both the original image and the conditioning mask. It is then fed to the generator, with produces a synthetic image. The model then uses two discriminators, the *standard* one, tasked with determining how realistic the synthetic image is, and the *differential* one, which determines if the differential image $x - y$ is acceptable. In that way, the generator is forced to create realistic synthetic samples that alter the original image only in areas and ways that are required by the domain transition task.

Expression GAN (ExprGAN). Combining elements of autoencoders and conditional GANs, the ExprGAN proposed by (Ding, Sricharan, & Chellappa, 2017) performs identity preserving domain transfer. ExprGAN is novel in that its conditioning input relates to information not only about the synthetic sample's the desired class, but its intensity as well. In the context it was used in the original paper, this formulation related to the intensity with which certain emotions were expected to appear in synthetic images of faces. Another notable element of the ExprGAN is that, due to the fact it uses multiple loss functions, each relating to specific components of its architecture, the authors suggest training those components independently in distinct stages. The ExprGAN was designed for the purposes of face expression alteration, but is not by design restricted to this application.

Identity Preserved Conditional GAN (IPCGAN). (X. Tang, Wang, Luo, & Gao, 2018) is designed to perform conditional domain transfer in the context of age on images of faces. In addition to the discriminator, the model makes use of a classifier and a pre-trained AlexNet model. The classifier determines the age group a given image belongs to and is used to force the generator to produce synthetic samples that correspond to the age group determined by the conditioning input. The AlexNet is used to extract features of the original and the synthetic image, whose distance the generator is motivated to minimize, so that the identity of the subject is preserved. Those three components ensure that the synthetic images are realistic, belong to the expected age group, and are of the same person as that of their input. Similar to the ExprGAN, it is designed and tested in the context of a specific task, that of face aging, but is not restricted to it with regard to its applicability.

Deep Attention GAN (DA-GAN). DA-GAN (Not to be confused with DAGAN (Antoniou et al., 2017)) was introduced by (Ma, Fu, Chen, & Mei, 2018). It uses a deep attention encoder to produce latent space representations based on the localized properties of each sample. The generator uses those representation to produce synthetic samples, which both locally and collectively have features that correspond to those of the target domain (as defined by samples of that domain). The paper proposes a combination of four loss function components to facilitate the use of DAEs and to enforce desirable properties on the model. Additionally, due to the fact that the synthetic samples are generated based on a localized interpretation of their properties projected to a latent space (which in the case of image-to-image translation is extracted via the DAE), DA-GAN is flexible regarding its applications which include, for instance, text-to-image generation.

Pedestrian Synthesis GAN (PS-GAN). This model, proposed in (Ouyang, Cheng, Jiang, Li, & Zhou, 2018), focuses on content infilling. Its input is two versions of the same image, the second of which has had a specific part of it corrupted by noise. The PS-GAN is expected to fill that area with realistic content. This application is used specifically in the context of a pedestrian detection task (though it can be used in other domains), wherein the PS-GAN is tasked with inserting pedestrians in sections of images defined by the corrupting noise. To achieve this, PS-GAN uses two discriminators rather than one. The first determines if the area in question is filled with realistic content in itself. The second discriminator determines if the synthetic image as a whole is realistic, when considered along with its noisy original.

Stacked GAN (SGAN) The SGAN, proposed in (Y. Tang et al., 2018), uses two GANs, whose architectures draw heavily from the SRGAN (Ledig et al., 2016), to perform two consecutive domain transfers on their input images in order to achieve desirable properties in the end result. In the case of (Y. Tang et al., 2018), the SGAN is used to improve an image's quality and resolution, and the two functions of the SGAN are to first denoise, and then to increase the resolution of its input image samples.

Domain Invariance & Feature Augmentation (DIFA) A framework designed for data augmentation for unsupervised domain adaptation tasks proposed in (Volpi, Morerio, Savarese, & Murino, 2018). It expands on the cGAN so that it performs data augmentation in the feature space, rather than by synthesizing new samples. The architecture is trained in 3 steps. Initially, a classifier is trained on the source domain. Then, the trained model is used as a feature extractor and a GAN is trained to produce realistic class-conditioned synthetic feature vectors. Finally, the encoder is trained again in the context of a GAN on both the source and target domains. Its objective is for the encoder to create common feature vector representations for samples in both domains. Ultimately, the framework results in an encoder trained for domain-invariant feature extraction, which can be used for augmentation, as well as for class inference in the cases of unlabeled datasets.

3DMM Cycle-GAN. (Gecer, Bhattarai, Kittler, & Kim, 2018) propose an adversarial approach for generating 2D images conditioned by a 3D Morphable Model. Being a method with similar objectives to the Cycle-GAN, its objective is to transfer samples generated by a 3DMM to a photorealistic domain. To do that, the model requires unpaired samples of both domains, and a proportionally small number of paired samples. Similar to the Cycle-GAN, the model uses a pair of generators $G : X \rightarrow Y$ and $F : Y \rightarrow X$ and their respective generators, but on the other hand it uses a Classifier along with the discriminator, as well as different loss functions to guide its training, in order to enforce identity preservation and pair matching, with limited paired samples from each domain.

Background Augmentation Generative Adversarial Network (BAGAN). BAGAN (M. Yan et al., 2018) is designed to function as part of a framework that generates 2D samples by drawing from renderings of 3D models of objects and augmenting them with diverse synthetic backgrounds while maintaining the objects' identities. It should be noted that the full pipeline of the framework involves multiple techniques and components that inform and utilize the actual GAN component, but the framework's purpose overall is to generate diverse and realistic annotated 2D samples from a limited 3D dataset. The Background Augmentation Generative Adversarial Network is distinct from the Balancing GAN (Mariani et al., 2018), though their acronyms are identical.

APPLICATIONS

*Table 1: Data augmentation with GANs - Applications. Datasets marked with * were gathered for the purposes of each work and have not necessarily been made public.*

Dataset	GAN Model	Reference	Sample Gen.
Domain: Medicine			
DDSM (Heath, Bowyer, Kopans, Moore, & Kegelmeyer, 2001)	ciGAN	(Yirui Wu et al., 2018)	CSG
CBIS-DDSM (R. S. Lee et al., 2017)	DADA	(X. Zhang et al., 2018)	CSG
LIDC (McNitt-Gray et al., 2008)	3D-CGAN	(Jin et al., 2018)	CSG(3D)
LIDC	3D MCGAN	(Han, Kitamura, et al., 2019)	CSG(3D)
LIDC	AIPBF	(Jie Yang et al., 2018)	CSG(3D)
Cardiac CT & MRI scans *	3D Cycle-GAN Var.	(Z. Zhang, Yang, & Zheng, 2018)	DT(3D)
Combination of fMRI datasets	Cond. WGAN-GP (3D) Var.	(Zhuang, Schwing, & Koyejo, 2019)	CSG
Liver Lesions *	DCGAN, AC-GAN	(Frid-Adar et al., 2018)	CSG, USGa
Chest X-Rays dataset *	DCGAN	(Salehinejad, Valaee, Dowdell, Colak, & Barfett, 2017)	USGa
Chromosome Karyotyping Cell Dataset *	MD-GAN	(Yirui Wu et al., 2018)	USGb
Pulmonary nodule CT dataset * Breast Cancer Wisc. (Diagnostic) Data Set (UCI, 2011)	WGAN	(Onishi et al., 2019)	USGa
BRATS (Menze et al., 2015)	PGGAN	(Han, Murao, et al., 2019)	USGa
Surgery Images *	DavinciGAN	(K. Lee et al., 2019)	DT
BCDR (Guevara Lopez et al., 2012), INbreast (Moreira et al., 2011), CBIS-DDSM (R. S. Lee et al., 2017)	Cycle-GAN	(Jendele, Skopek, Becker, & Konukoglu, 2019)	DT
MITOS-ATYPIA-14 Challenge[1]	Cycle-GAN	(R. S. Lee et al., 2017; Shaban, Baur, Navab, & Albarqouni, 2018)	DT
TCGA (Kandoth et al., 2013), CINJ Histopathology Dataset *	GAN var.	(Ren, Hacihaliloglu, Singer, Foran, & Qi, 2018)	DT
MITOS-ATYPIA-14 Challenge, MICCAI’16(Sirinukunwattana et al., 2016), Ovarian Carcinoma whole slide images *	AC-GAN var.	(Bentaieb & Hamarneh, 2017)	DTc

[1] https://mitos-atypia-14.grand-challenge.org

Dataset	GAN Model	Reference	Sample Gen.
DeepLesion (K. Yan et al., 2018)	SGAN	(Y. Tang et al., 2018)	DT
Blood smears images *	pix2pix	(Bailo, Ham, & Shin, 2019)	DT
BRATS, ADNI[2]	pix2pix	(Shin et al., 2018)	DT(3D)
CT & FLAIR datasets *	PGGAN	(Bowles, Gunn, Hammers, & Rueckert, 2018)	USGc
Brain MRI dataset *	Cond. PGGAN	(Han, Murao, et al., 2019)	CSG
Domain: People & Faces			
VGG Face (Parkhi, Vedaldi, & Zisserman, 2015)	DAGAN	(Antoniou et al., 2017)	CSG
CASIAWebFace (D. Yi, Lei, Liao, & Li, 2014)	CVAE-GAN	(Bao et al., 2017)	CSG
VGG Face, LFW (G. B. Huang, Ramesh, Berg, & Learned-Miller, 2007), IJB-A (Klare et al., 2015)	3DMM Cycle-GAN	(Gecer et al., 2018)	DT3t2(3D)
CASIA Palmprint Dataset [3] IIT Delhi Palmprint database[4]	DCGAN	(G. Wang, Kang, Wu, Wang, & Gao, 2019)	USGa
CACD (B.-C. Chen, Chen, & Hsu, 2014)	IPCGAN	(X. Tang et al., 2018)	DTc
Oulu-CASIA (Zhao, Huang, Taini, Li, & Pietikäinen, 2011)	ExprGAN	(Ding et al., 2017)	DTc
FER2013 (Dhall, Goecke, Lucey, & Gedeon, 2011), SFEW (Goodfellow et al., 2015), JAFFE (Kamachi, Lyons, & Gyoba, 1997)	Cycle-GAN	(Y. Zhu, Aoun, Science, Krijn, & Vanschoren, 2018)	DT
MMI (Valstar & Pantic, 2010), LFW	D-GAN	(Gu et al., 2017)	DT
Tsinghua-Daimler Cyclist Benchmark (X. Li et al., 2016),	PS-GAN	(Ouyang et al., 2018)	DT
CASIA gait dataset (Shiqi Yu, Daoliang Tan, & Tieniu Tan, 2006), VOT2013, VOT2014, VOT2015 (Kristan et al., 2015), OTB100 (Y Wu, Lim, & Yang, 2013)	StarGAN	(K. Chen, Zhou, Zhou, & Xu, 2019)	DTc
Market-1501 (L. Zheng et al., 2015), DukeMTMC-ReID (Ristani, Solera, Zou, Cucchiara, & Tomasi, 2016)	CTGAN	(Zhou et al., 2019)	DTc
Market-1501, DukeMTMC-ReID	Cycle-GAN	(Zhong, Zheng, Zheng, Li, & Yang, 2017)	DT
Market-1501, DukeMTMC-ReID, CUHK03 (W. Li, Zhao, Xiao, & Wang, 2014)	DCGAN	(Z. Zheng, Zheng, & Yang, 2017)	USGd
Domain: Other			
ImageNet	cCov-GAN	(Gao et al., 2018)	DTc
X-Rays of various items*	DCGAN, CT-GAN,	(J Yang, Zhao, Zhang, & Shi, 2019)	USGb
ObjectNet3D (Xiang et al., 2016), ShapeNet (Chang et al., 2015)	BAGAN	(M. Yan et al., 2018)	DT3t2(3D)
Omniglot (M Lake, Salakhutdinov, & B Tenenbaum, 2015), EMNIST (Cohen, Afshar, Tapson, & van Schaik, 2017)	DAGAN	(Antoniou et al., 2017)	CSG
RHP (Zimmermann & Brox, 2017), STB (J. Zhang et al., 2016), CMU-PS (Simon, Joo, & Sheikh, 2017)	TAGAN	(L. Chen et al., 2018)	DT, DT2t3(3D)
HMDB51, UCF101	WGAN	(Y. Zhang, Jia, Chen, Zhang, & Yong, 2019)	USGa
Omniglot, Mini-Imagenet (Vinyals, Blundell, Lillicrap, Kavukcuoglu, & Wierstra, 2016)	MetaGAN	(R. Zhang et al., 2018)	CSG
CURE-TSR (Temel, Kwon, Prabhushankar, & AlRegib, 2017)	Three-Player GAN	(Vandenhende et al., 2019)	CSG

[2] http://adni.loni.usc.edu/

[3] CASIA Palmprint Database, http://biometrics.idealtest.org/

[4] IIT Delhi Palmprint Image Database version 1.0, http://www4.comp.polyu.edu.hk/csajaykr/ITD/

Dataset	GAN Model	Reference	Sample Gen.
MNIST (Lecun, Bottou, Bengio, & Haffner, 1998), CIFAR-10 (Krizhevsky, 2012), Flowers[5]	BAGAN	(Mariani et al., 2018)	CSG
CUB-200-2011 (Wah, Branson, Welinder, Perona, & Belongie, 2011)	DA-GAN	(Ma et al., 2018)	DT
MNIST, USPS (Denker et al., 1989), SVHN (Netzer et al., 2011), SYN DIGITS (Ganin & Lempitsky, 2015), NYUD (Silberman, Hoiem, Kohli, & Fergus, 2012)	DIFA	(Volpi et al., 2018)	DT
SYNTHIA (Ros, Sellart, Materzynska, Vazquez, & Lopez, 2016), GTA (Richter, Vineet, Roth, & Koltun, 2016), KITTI (Geiger, Lenz, & Urtasun, 2012), ITRI*[6]	AugGAN	(S. W. Huang et al., 2018)	DT
Cityscapes (Cordts et al., 2016)	PS-GAN	(Ouyang et al., 2018)	DT
CVPPP 2017 LSC Plant Dataset ("Jonathan Bell and Hannah M Dee. Aberystwyth leaf evaluation dataset. 2016.," n.d.)	cGAN	(J. Y. Zhu et al., 2017)	CSG

Having surmised in section *GAN models* the technical features of each GAN that will be mentioned, and having defined the tasks they were used for in section *Tasks*, this section presents the specific Data Augmentation applications that GANs were identified to have been used in by our research on the topic. We organize this section in three distinct groups based on the fields in which each application belongs to. The first groups relates to medicine, the second to people and faces, and the third includes all remaining applications that could not be grouped into a broader category. It is this section's objective to list the ways in which GANs have been used to augment datasets in each domain and application. We will refer to sample generation techniques using the notation described in subsection *Samples Generation Method*. The specific datasets used in each work described in this section is listed in Table 1.

Medicine

GAN-based data augmentation was found to have been used extensively in applications related to medicine, where scarcity of data and difficulty in annotating them coincides with enormous potential benefits if machine learning could be applied to facilitate various tasks, especially diagnostic ones. The motivating promise of significant returns for healthcare systems can explain the fact that most applications we found were related to this domain. 2D and 3D imagery analysis is particularly important in extracting information from CT, MRI and other diagnostic tools.

In this fields, GAN-based data augmentation has been used most frequently in augmenting datasets consisting of MRIs, X-Rays and other medical images to improve on classification or segmentation related diagnostic tasks. With regard to classification, (Han, Rundo, et al., 2019) used USGa with the PGGAN on brain MRIs for cancer diagnoses, and subsequently (Han, Murao, et al., 2019) used a conditional PGGAN in a similar application with a tumor detection objective (OD). PGGAN was also used by (Bowles et al., 2018) in a transfer learning framework to augment brain CTs and MRIs for improved segmentation via USGc. Another application of USGa was found in (Frid-Adar et al., 2018) and (Salehinejad et al., 2017), where the DCGAN was used to augment datasets consisting of liver lesions CT scans and chest X-rays respectively. With regard to Conditional Samples Generation (CSG), (Frid-Adar et al., 2018) also used the AC-GAN to generate samples conditionally to augment a dataset of CT scans. An interesting application of data augmentation via USGa is found in (Onishi et al., 2019), where a

[5] https:// www.kaggle.com/alxmamaev/flowers-recognition
[6] Dataset real driving images

WGAN is used on a pulmonary CT dataset, but the classifier is trained only on synthetic samples, and then fine-tuned with the originals. (X. Zhang et al., 2018) used the DADA framework to augment a dataset of scanned film mammography studies after experimenting with CIFAR-10, while (E. Wu, Wu, Cox, & Lotter, 2018) used the ciGAN in a similar dataset, to augment mammograms via content infilling (filling specific sections of each sample conditionally to facilitate classification) for dataset balancing. Moving on to GANs performing domain transfer, (Bentaieb & Hamarneh, 2017) proposed a combination of an AC-GAN and an autoencoder in order to perform conditional domain transfer (DTc) on various histopathology image datasets, in order to solve the problem of stain color inconsistencies for both classification and segmentation purposes. On the same subject, (Shaban et al., 2018) used the Cycle-GAN on a histopathology dataset to improve on tumor classification, and (Ren et al., 2018) used an architecture combining a GAN with a Siamese network to augment two prostate histopathology image datasets in an unsupervised way, also to improve classification results. (Y. Tang et al., 2018) used the SGAN to augment CT images for lesion segmentation purposes and (Jendele et al., 2019) used the Cycle-GAN for breast cancer diagnosis (classification). Regarding 3D imagery, (Jin et al., 2018) used for content infilling to create synthetic lung nodules which are then used to fine tune a nodule segmentation model. Using the same dataset of lung nodules, (Han, Kitamura, et al., 2019) proposes the 3D MCGAN for 3D conditional infilling in the context of object detection, and (Jie Yang et al., 2018) proposed the AIPBF GAN based framework for augmented classification. (Shin et al., 2018) utilized the pix2pix methodology, adjusted for 3D data, to augment brain MRIs and improve on a tumor segmentation task. (Z. Zhang et al., 2018) proposed a modification on the Cycle-GAN, such that allowed for domain transition between two 3D domains, which was utilized to improve segmentation performance using CT and MRI cardiac scans by transitioning samples to either domain. In (Zhuang et al., 2019), a conditional WGAN-GP variant was used to augment 3D fMRI datasets for a classification task.

Diagnostics aside, GANs have also been used in other applications. (Yirui Wu et al., 2018) used MD-GAN to improve on chromosome classification (Karyotyping). (Bailo et al., 2019) augmented a dataset of blood smear slides with pix2pix and improved on segmentation performance. In (K. Lee et al., 2019), the DavinciGAN is used to augment a dataset consisting of frames of surgeries which include the tools used be the surgeon, to improve on instrument classification.

For a broader overview of GANs used in medical applications, we suggest reading (Kazeminia et al., 2018) and (X. Yi, Walia, & Babyn, 2018).

People & Faces

This section lists works that tackled the datasets whose subjects were people, whether in terms of faces, body shapes or other characteristics. Regarding face data augmentation techniques more broadly, we also suggest (X. Wang, Wang, & Lian, 2019), which reviews approaches that are not limited to GANs.

In this domain, the only cases were samples were generated unconditionally were (G. Wang et al., 2019) and (Z. Zheng et al., 2017), in both of which the DCGAN was used. The former applied USGa on a palmprint dataset, applying classification in terms of identifying the owner of each print. The later used USGd in a person identification (PID) task using datasets consisting of images of pedestrians captured by cameras. (Gecer et al., 2018) used the 3DMM Cycle-GAN to augment 2D face datasets with samples originating from using 3D models. (Antoniou et al., 2017) used the DAGAN in a CSG application to augment a dataset of faces to improve person classification, which was also done by (Bao et al., 2017) with the CVAE-GAN. The rest of the works in this domain used Domain Transfer. In the area of classification, (X. Tang et al., 2018) used the IPCGAN to augment dace datasets age-wise in an identity preserving way (DTc) and (Y. Zhu et al., 2018) used the Cycle-GAN to augment face datasets emotionwise, which was done both uniformly (UA) and for specific classes (DB) to improve on emotion classification. Finally, (Gu et al., 2017) used the ExprGAN for DTc and (Ding et al., 2017) used the D-GAN both for UA and DB, to augment face datasets with regard to their expresions. It should be noted that works performing domain transfer on faces provide for identity-preservation, as defined in section *Domain Transfer (DT)*. Regarding person identification (PID), (Zhou et al., 2019) uses the CTGAN on images of pedestrians and (Zhong et al., 2017) uses the Cycle-GAN to improve on a multi-camera person re-identification task. Finally, object detection (OD) is tackled in (Ouyang et al., 2018) on a dataset of cyclist images using the PS-GAN, and (K. Chen et al., 2019) does object tracking (OT) with the StarGAN on various videos of people.

Other Data

This subsection relates to works whose area of focus does not fall into a concrete application domain, and includes publications that tackled popular machine learning datasets, as well as data specific to particular applications.
Regarding classification tasks, USGb was used in (J Yang et al., 2019), which compared the DCGAN, WGAN-GP and CT-GAN's performances in augmenting a dataset consisting of X-rays of various items which were to be classified. Conditional approaches included (Antoniou et al., 2017), which used the DAGAN on two datasets with written characters and (R. Zhang et al., 2018) which used the MetaGAN on datasets of characters and various images. CSG was also used in (Vandenhende et al., 2019), which used the Three-Player GAN to augment a dataset of traffic signs to improve classification accuracy. Finally, (Mariani et al., 2018) use the BAGAN on various machine learning datasets performing CSG and DB. Domain Transfer for classification purposes is used by (Ma et al., 2018) with the DA-GAN on a dataset of birds that are augmented pose-wise, by (Volpi et al., 2018), where DIFA is used to augment data using other similar datasets or domains, and by (Gao et al., 2018), where a variation of the Cycle-GAN, the cCov-GAN, was used to augment a downsampled ImageNet dataset in a low-shot learning classification task. Lastly, (J. Y. Zhu et al., 2017) utilized a cGAN conditioned on plant masks to conditionally generate diverse plant images to facilitate accurate leaf segmentation and counting, (S. W. Huang et al., 2018) used the AugGAN to perform day/night DT on various datasets of images of cars in streets in order to improve object detection efficiency, and (Ouyang et al., 2018) performed DT on a dataset of city street images with the PS-GAN, also to improve OD. Data augmentation in video was applied in (Y. Zhang et al., 2019), where WGAN was used as part of a pipeline to augment video classification datasets. In the area of 3D to 2D transitions, BAGAN (M. Yan et al., 2018) was used with the ObjectNet3D and ShapeNet databases to augment a 2D classification task with samples generated from 3D models, arguing that 3D models are easier to generate than it is to gather annotated 2D data, and so using them to synthesize diverse 2D samples is more efficient. TAGAN, which is proposed in (L. Chen et al., 2018) to augment datasets used for hand pose estimation is also used to generate both 2D and 3D samples.

CONCLUSIONS

In summarizing our findings, we first have to observe that the number and diversity of the GAN models that were found to have been used for Data Augmentation on 2D and 3D imagery, as well as the various techniques that were employed, make it difficult to suggest a strict methodology for choosing the optimal approach for any given problem. We would, however, make note of the following conclusions that can be drawn from the works that were studied. The first conclusion is that using Unconditional Samples Generation is rarely the best option for Data Augmentation. While there are ways that it might be applied, those approaches have significant drawbacks and were found to be largely outdated. Considering the need for annotated samples in the context of data augmentation, Conditional Samples Generation is the most straightforward tool, particularly when augmenting data for a classification problem. Among conditional GANs, the AC-GAN presents itself as a simple, efficient and versatile model. Finally, with regard to Domain Transfer, the Cycle-GAN proved to be a very potent and adaptive architecture, which can also be attested to by the number of variants it has inspired (e.g. the StarGAN and the cCov-GAN). The above are in no way guaranteed to augment all datasets in an satisfactory way. They constitute, however, potent and versatile architectures that can easily be modified to match the specific requirements of various problems, which was the case in most of the cases that were examined in this survey.

It is also important to point out that many of the papers we studied used approaches that were largely outdated by the time those papers were written. Expanding on this observation, it must be made clear that this work does not include more recent and advanced GANs that have not, to the best of our knowledge, been used in the context of image data augmentation. For the study of state of the art models and the techniques they employ we recommend (H. Zhang, Goodfellow, Metaxas, & Odena, 2018), (Brock, Donahue, & Simonyan, 2018) and (Lucic et al., 2019) with regard to CSG and (Park, Liu, Wang, & Zhu, 2019) with regard to Domain Transfer. We attribute the relative delay in applying state of the art GAN models to data augmentation tasks to the fact that GANs are considerably complex architectures, and their level of complexity increases steeply as more advanced models are examined. Consequently, it is not easy to adapt such models to new applications. It should also be pointed out that the increased capacity of state of the art GANs may not be required in cases where data are limited, as well as that more powerful GANs are more difficult to train successfully, particularly with limited data and computational resources. Consequently, GANs that perform better in large benchmark datasets are not necessarily appropriate for data augmentation purposes.

Finally, it is important to examine the applications listed in this survey with regard to the 2D/3D distinction. 3D data augmentation applications were considerably more scarce than 2D. This mirrors the general trend in machine learning research overall, where 3D data and tasks have only recently attracted interest. This can be attributed to the fact that 3D applications are only now becoming prominent enough to motivate extensive study and the creation of annotated datasets. GAN-based data augmentation however is proven in the works we surveyed to be a potent tool in overcoming the relative scarcity of annotated 3D data. Most 2D architectures can easily be adapted to generate 3D data (e.g. the 3D-GAN in section *3D Generative Adversarial Network (3D-GAN)*), while domain transfer GANs can be used to transfer images from 2D to 3D and the inverse. This is an interesting application related to the generation of diverse samples from 3D models, which can also serve as a source of annotated data.

Overall, research in data augmentation can be shown by the surveyed works to be largely driven by real-life data shortages in specific applications rather than academic interest. However, there is consensus in the works we surveyed that GANs can be an invaluable tool in overcoming data scarcity, which is itself a major impediment in the use of machine learning in many domains. It is then our expectation that, considering the potential for applying machine learning solutions in many of the subject areas we examined, considerable effort will be invested in the future toward making advanced GAN models more accessible and easier to use

ACKNOWLEDGMENT
This work was supported by the EU funded project NADINE H2020 under the grant agreement 822601

REFERENCES

Alexiadis, D. S., Chatzitofis, A., Zioulis, N., Zoidi, O., Louizis, G., Zarpalas, D., & Daras, P. (2016). An integrated platform for live 3D human reconstruction and motion capturing. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(4), 798–813.

Antoniou, A., Storkey, A., & Edwards, H. (2017). Data Augmentation Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/1711.04340

Arjovsky, M., & Bottou, L. (2017). Towards Principled Methods for Training Generative Adversarial

Networks. Retrieved from http://arxiv.org/abs/1701.04862
Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. Retrieved from http://arxiv.org/abs/1701.07875
Bailo, O., Ham, D., & Shin, Y. M. (2019). Red blood cell image generation for data augmentation using Conditional Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/1901.06219
Bao, J., Chen, D., Wen, F., Li, H., & Hua, G. (2017). CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training. Retrieved from http://arxiv.org/abs/1703.10155
Bentaieb, A., & Hamarneh, G. (2017). Adversarial Stain Transfer for Histopathology Image Analysis. *IEEE Transactions on Medical Imaging*, *PP*, 1. https://doi.org/10.1109/TMI.2017.2781228
Bowles, C., Gunn, R., Hammers, A., & Rueckert, D. (2018). GANsfer Learning: Combining labelled and unlabelled data for GAN based data augmentation, 1–10. Retrieved from http://arxiv.org/abs/1811.10669
Brock, A., Donahue, J., & Simonyan, K. (2018). Large Scale GAN Training for High Fidelity Natural Image Synthesis, 1–35. Retrieved from http://arxiv.org/abs/1809.11096
Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. (2012). DBSMOTE: Density-based synthetic minority over-sampling technique. *Applied Intelligence*, *36*(3), 664–684. https://doi.org/10.1007/s10489-011-0287-y
Burt, P. J., & Adelson, E. H. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*. https://doi.org/10.1109/TCOM.1983.1095851
Chang, A. X., Funkhouser, T. A., Guibas, L. J., Hanrahan, P., Huang, Q.-X., Li, Z., … Yu, F. (2015). ShapeNet: An Information-Rich 3D Model Repository. *CoRR*, *abs/1512.03012*. Retrieved from http://arxiv.org/abs/1512.03012
Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*.
Chen, B.-C., Chen, C.-S., & Hsu, W. H. (2014). Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. In *Proceedings of the European Conference on Computer Vision ({ECCV})*.
Chen, K., Zhou, X., Zhou, Q., & Xu, H. (2019). Adversarial Learning-based Data Augmentation for Rotation-robust Human Tracking. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1942–1946). IEEE. https://doi.org/10.1109/icassp.2019.8683451
Chen, L., Lin, S.-Y., Xie, Y., Tang, H., Xue, Y., Xie, X., … Fan, W. (2018). Generating Realistic Training Images Based on Tonality-Alignment Generative Adversarial Networks for Hand Pose Estimation. *CoRR*, *abs/1811.09916*. Retrieved from http://arxiv.org/abs/1811.09916
Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8789–8797. https://doi.org/10.1109/CVPR.2018.00916
Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). {EMNIST:} an extension of {MNIST} to handwritten letters. *CoRR*, *abs/1702.05373*. Retrieved from http://arxiv.org/abs/1702.05373
Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., … Schiele, B. (2016). The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 3213–3223. https://doi.org/10.1109/CVPR.2016.350
Denker, J. S., Gardner, W. R., Graf, H. P., Henderson, D., Howard, R. E., Hubbard, W., … Guyon, I. (1989). Neural Network Recognizer for Hand-Written Zip Code Digits. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 1* (pp. 323–331). Morgan-Kaufmann. Retrieved from http://papers.nips.cc/paper/107-neural-network-recognizer-for-hand-written-zip-code-digits.pdf
Denton, E., Chintala, S., Szlam, A., & Fergus, R. (2015). Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. Retrieved from http://arxiv.org/abs/1506.05751

Dhall, A., Goecke, R., Lucey, S., & Gedeon, T. (2011). Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 2106–2112). https://doi.org/10.1109/ICCVW.2011.6130508

Ding, H., Sricharan, K., & Chellappa, R. (2017). ExprGAN: Facial Expression Editing with Controllable Expression Intensity. Retrieved from https://arxiv.org/abs/1709.03842

Donahue, C., McAuley, J., & Puckette, M. (2018). Adversarial audio synthesis. *ArXiv Preprint ArXiv:1802.04208*.

Doumanoglou, A., Vretos, N., & Daras, P. (2019). Frequency--based slow feature analysis. *Neurocomputing*, *368*, 34–50.

Elsken, T., Metzen, J. H., & Hutter, F. (2019). Neural Architecture Search: A Survey. *Journal of Machine Learning Research*, *20*(55), 1–21. Retrieved from http://jmlr.org/papers/v20/18-598.html

Finn, C., Abbeel, P., & Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *CoRR*, *abs/1703.03400*. Retrieved from http://arxiv.org/abs/1703.03400

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, *321*, 321–331. https://doi.org/10.1016/j.neucom.2018.09.013

Ganin, Y., & Lempitsky, V. (2015). Unsupervised Domain Adaptation by Backpropagation. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37* (pp. 1180–1189). JMLR.org. Retrieved from http://dl.acm.org/citation.cfm?id=3045118.3045244

Gao, H., Shou, Z., Zareian, A., Zhang, H., & Chang, S.-F. (2018). Low-shot Learning via Covariance-Preserving Adversarial Augmentation Networks, (NeurIPS), 1–13. Retrieved from http://arxiv.org/abs/1810.11730

Gecer, B., Bhattarai, B., Kittler, J., & Kim, T.-K. (2018). Semi-supervised Adversarial Learning to Generate Photorealistic Face Images of New Identities from 3D Morphable Model. https://doi.org/10.1007/978-3-030-01252-6_14

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3354–3361). https://doi.org/10.1109/CVPR.2012.6248074

Gong, X., Chang, S., Jiang, Y., & Wang, Z. (2019). AutoGAN: Neural Architecture Search for Generative Adversarial Networks. Retrieved from https://arxiv.org/abs/1908.03835

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., … Bengio, Y. (2015). Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, *64*, 59–63. https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.005

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/1406.2661

Gu, G., Kim, S. T., Kim, K., Baddar, W. J., & Ro, Y. M. (2017). Differential Generative Adversarial Networks: Synthesizing Non-linear Facial Variations with Limited Number of Training Data. Retrieved from http://arxiv.org/abs/1711.10267

Guevara Lopez, M. A., González Posada, N., Moura, D., Pollán, R., Franco-Valiente, J., Ortega, C., … Ferreira M Araújo, B. (2012). BCDR: A BREAST CANCER DIGITAL REPOSITORY (pp. 1065–1066).

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved Training of Wasserstein GANs. Retrieved from http://arxiv.org/abs/1704.00028

Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., … Li, Y. (2019). Synthesizing Diverse Lung Nodules Wherever Massively: 3D Multi-Conditional GAN-based {CT} Image Augmentation for Object Detection. *CoRR*, *abs/1906.04962*. Retrieved from http://arxiv.org/abs/1906.04962

Han, C., Murao, K., Noguchi, T., Kawata, Y., Uchiyama, F., Rundo, L., … Satoh, S. (2019). Learning More with Less: Conditional PGGAN-based Data Augmentation for Brain Metastases Detection

Using Highly-Rough Annotation on MR Images. Retrieved from http://arxiv.org/abs/1902.09856
Han, C., Rundo, L., Araki, R., Furukawa, Y., Mauri, G., Nakayama, H., & Hayashi, H. (2019). Infinite Brain MR Images: PGGAN-based Data Augmentation for Tumor Detection. Retrieved from http://arxiv.org/abs/1903.12564
Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, P. (2001). The digital database for screening mammography. *Proceedings of the Fifth International Workshop on Digital Mammography*. https://doi.org/10.1007/978-94-011-5318-8_75
Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Retrieved from https://arxiv.org/abs/1706.08500
Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*.
Huang, S. W., Lin, C. T., Chen, S. P., Wu, Y. Y., Hsu, P. H., & Lai, S. H. (2018). AugGAN: Cross domain adaptation with GAN-based data augmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11213 LNCS*, 731–744. https://doi.org/10.1007/978-3-030-01240-3_44
Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Vol. 2017-Janua, pp. 5967–5976). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CVPR.2017.632
Jendele, L., Skopek, O., Becker, A. S., & Konukoglu, E. (2019). Adversarial Augmentation for Enhancing Classification of Mammography Images. Retrieved from http://arxiv.org/abs/1902.07762
Jin, D., Xu, Z., Tang, Y., Harrison, A. P., & Mollura, D. J. (2018). CT-Realistic Lung Nodule Simulation from 3D Conditional Generative Adversarial Networks for Robust Lung Segmentation. *CoRR*, *abs/1806.04051*. Retrieved from http://arxiv.org/abs/1806.04051
Jonathan Bell and Hannah M Dee. Aberystwyth leaf evaluation dataset. 2016. (n.d.).
Kamachi, M., Lyons, M., & Gyoba, J. (1997). The japanese female facial expression (jaffe) database. *Availble: Http://Www. Kasrl. Org/Jaffe. Html*.
Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., … Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*. https://doi.org/10.1038/nature12634
Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive Growing of GANs for Improved Quality, Stability, and Variation, 1–26. Retrieved from http://arxiv.org/abs/1710.10196
Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., & Mukhopadhyay, A. (2018). GANs for Medical Image Analysis. Retrieved from http://arxiv.org/abs/1809.06222
Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. Retrieved from https://arxiv.org/abs/1312.6114
Klare, B. F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., … Jain, A. K. (2015). Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1931–1939). https://doi.org/10.1109/CVPR.2015.7298803
Kristan, M., Matas, J., Leonardis, A., Felsberg, M., Cehovin, L., Fernandez, G., … Solis Montero, A. (2015). The Visual Object Tracking VOT2015 Challenge Results. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)* (pp. 564–586). https://doi.org/10.1109/ICCVW.2015.79
Krizhevsky, A. (2012). Learning Multiple Layers of Features from Tiny Images. *University of Toronto*.
Kurach, K., Lucic, M., Zhai, X., Michalski, M., & Gelly, S. (2018). A Large-Scale Study on Regularization and Normalization in GANs. Retrieved from https://arxiv.org/abs/1807.04720
Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791
Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A. P., Tejani, A., … Shi, W. (2016). Photo-

Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *CoRR*, *abs/1609.04802*. Retrieved from http://arxiv.org/abs/1609.04802

Lee, K., Choi, M.-K., & Jung, H. (2019). DavinciGAN: Unpaired Surgical Instrument Translation for Data Augmentation. In M. J. Cardoso, A. Feragen, B. Glocker, E. Konukoglu, I. Oguz, G. Unal, & T. Vercauteren (Eds.), *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning* (Vol. 102, pp. 326–336). London, United Kingdom: PMLR. Retrieved from http://proceedings.mlr.press/v102/lee19a.html

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). Data Descriptor: A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*. https://doi.org/10.1038/sdata.2017.177

Li, C., & Wand, M. (2016). Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. Retrieved from https://arxiv.org/abs/1604.04382

Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 152–159). https://doi.org/10.1109/CVPR.2014.27

Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., … M. Gavrila, D. (2016). A new benchmark for vision-based cyclist detection (pp. 1028–1033). https://doi.org/10.1109/IVS.2016.7535515

Lucic, M., Kurach, K., Michalski, M., Gelly, S., & Bousquet, O. (2017). Are GANs Created Equal? A Large-Scale Study. Retrieved from http://arxiv.org/abs/1711.10337

Lucic, M., Tschannen, M., Ritter, M., Zhai, X., Bachem, O., & Gelly, S. (2019). High-Fidelity Image Generation With Fewer Labels. Retrieved from http://arxiv.org/abs/1903.02271

M Lake, B., Salakhutdinov, R., & B Tenenbaum, J. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*, 1332–1338. https://doi.org/10.1126/science.aab3050

Ma, S., Fu, J., Chen, C. W., & Mei, T. (2018). DA-GAN: Instance-Level Image Translation by Deep Attention Generative Adversarial Networks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 5657–5666). https://doi.org/10.1109/CVPR.2018.00593

Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., & Malossi, C. (2018). BAGAN: Data Augmentation with Balancing GAN. Retrieved from http://arxiv.org/abs/1803.09655

McNitt-Gray, M., Armato III, S., Meyer, C., Reeves, A., Mclennan, G., Pais, R., … Clarke, L. (2008). The Lung Image Database Consortium (LIDC) Data Collection Process for Nodule Detection and Annotation. *Academic Radiology*, *14*, 1464–1474. https://doi.org/10.1016/j.acra.2007.07.021

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., … Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694

Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. *CoRR*, *abs/1411.1784*. Retrieved from http://arxiv.org/abs/1411.1784

Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral Normalization for Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/1802.05957

Moreira, I., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M., & Cardoso, J. (2011). INbreast: Toward a Full-field Digital Mammographic Database. *Academic Radiology*, *19*, 236–248. https://doi.org/10.1016/j.acra.2011.09.014

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading Digits in Natural Images with Unsupervised Feature Learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*. Retrieved from http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf

Odena, A., Buckman, J., Olsson, C., Brown, T. B., Olah, C., Raffel, C., & Goodfellow, I. (2018). Is Generator Conditioning Causally Related to GAN Performance? Retrieved from http://arxiv.org/abs/1802.08768

Odena, A., Olah, C., & Shlens, J. (2016). Conditional Image Synthesis With Auxiliary Classifier GANs. Retrieved from http://arxiv.org/abs/1610.09585

Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., … Fujita, H. (2019). Automated Pulmonary Nodule Classification in Computed Tomography Images Using a Deep Convolutional Neural Network Trained by Generative Adversarial Networks. *BioMed Research International*, *2019*, 1–9. https://doi.org/10.1155/2019/6051939

Ouyang, X., Cheng, Y., Jiang, Y., Li, C.-L., & Zhou, P. (2018). Pedestrian-Synthesis-GAN: Generating Pedestrian Data in Real Scene and Beyond. Retrieved from http://arxiv.org/abs/1804.02047

Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., & Zheng, Y. (2019). Recent Progress on Generative Adversarial Networks (GANs): A Survey. *IEEE Access*, *7*, 36322–36333. https://doi.org/10.1109/ACCESS.2019.2905015

Park, T., Liu, M.-Y., Wang, T.-C., & Zhu, J.-Y. (2019). Semantic Image Synthesis with Spatially-Adaptive Normalization. *CoRR*, *abs/1903.07291*. Retrieved from http://arxiv.org/abs/1903.07291

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. In *British Machine Vision Conference*.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/1511.06434

Ravuri, S., & Vinyals, O. (2019). Classification Accuracy Score for Conditional Generative Models. Retrieved from http://arxiv.org/abs/1905.10887

Ren, J., Hacihaliloglu, I., Singer, E. A., Foran, D. J., & Qi, X. (2018). Adversarial Domain Adaptation for Classification of Prostate Histopathology Whole-Slide Images. Retrieved from http://arxiv.org/abs/1806.01357

Richter, S. R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for Data: Ground Truth from Computer Games. *CoRR*, *abs/1608.02192*. Retrieved from http://arxiv.org/abs/1608.02192

Ristani, E., Solera, F., Zou, R. S., Cucchiara, R., & Tomasi, C. (2016). Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. *CoRR*, *abs/1609.01775*. Retrieved from http://arxiv.org/abs/1609.01775

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Retrieved from https://arxiv.org/abs/1505.04597

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., & Lopez, A. M. (2016). The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3234–3243). https://doi.org/10.1109/CVPR.2016.352

Salehinejad, H., Valaee, S., Dowdell, T., Colak, E., & Barfett, J. (2017). Generalization of Deep Neural Networks for Chest Pathology Classification in X-Rays Using Generative Adversarial Networks. Retrieved from https://arxiv.org/abs/1712.01636

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved Techniques for Training GANs. Retrieved from http://arxiv.org/abs/1606.03498

Shaban, M. T., Baur, C., Navab, N., & Albarqouni, S. (2018). StainGAN: Stain Style Transfer for Digital Histological Images. Retrieved from http://arxiv.org/abs/1804.01601

Shijie, J., Ping, W., Peiyi, J., & Siping, H. (2017). Research on data augmentation for image classification based on convolution neural networks. *Proceedings - 2017 Chinese Automation Congress, CAC 2017*, *2017-Janua*(201602118), 4165–4170. https://doi.org/10.1109/CAC.2017.8243510

Shin, H. C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., … Michalski, M. (2018). Medical image synthesis for data augmentation and anonymization using generative adversarial networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11037 LNCS*, 1–11. https://doi.org/10.1007/978-3-030-00536-8_1

Shiqi Yu, Daoliang Tan, & Tieniu Tan. (2006). A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 4, pp. 441–444). https://doi.org/10.1109/ICPR.2006.67

Shmelkov, K., Schmid, C., & Alahari, K. (2018). How good is my GAN. In *Proceedings of the European Conference on Computer Vision (ECCV 2018)* (pp. 1–20). Retrieved from

http://arxiv.org/abs/1807.09499
Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor Segmentation and Support Inference from RGBD Images. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part V* (pp. 746–760). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-33715-4_54
Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR* (Vol. 2003-Janua, pp. 958–963). https://doi.org/10.1109/ICDAR.2003.1227801
Simon, T., Joo, H., & Sheikh, Y. (2017). Hand Keypoint Detection in Single Images using Multiview Bootstrapping. *CVPR*.
Sirinukunwattana, K., Pluim, J. P. W., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., … Rajpoot, N. M. (2016). Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest. *CoRR*, *abs/1603.00275*. Retrieved from http://arxiv.org/abs/1603.00275
Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H. S., & Hospedales, T. M. (2017). Learning to Compare: Relation Network for Few-Shot Learning. *CoRR*, *abs/1711.06025*. Retrieved from http://arxiv.org/abs/1711.06025
Sutskever, I., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. R. (2014). Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*.
Tang, X., Wang, Z., Luo, W., & Gao, S. (2018). Face Aging with Identity-Preserved Conditional Generative Adversarial Networks (pp. 7939–7947). https://doi.org/10.1109/CVPR.2018.00828
Tang, Y., Cai, J., Lu, L., Harrison, A. P., Yan, K., Xiao, J., … Summers, R. M. (2018). CT Image Enhancement Using Stacked Generative Adversarial Networks and Transfer Learning for Lesion Segmentation Improvement. Retrieved from https://arxiv.org/abs/1807.07144
Taylor, L., & Nitschke, G. (2019). Improving Deep Learning with Generic Data Augmentation. In *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018* (pp. 1542–1547). https://doi.org/10.1109/SSCI.2018.8628742
Temel, D., Kwon, G., Prabhushankar, M., & AlRegib, G. (2017). {CURE-TSR:} Challenging Unreal and Real Environments for Traffic Sign Recognition. *CoRR*, *abs/1712.02463*. Retrieved from http://arxiv.org/abs/1712.02463
UCI. (2011). UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set. *Http://Archive.Ics.Uci.Edu/Ml/Datasets/Breast+Cancer+Wisconsin+%2528Diagnostic%2529*.
Valstar, M., & Pantic, M. (2010). Induced disgust, happiness and surprise: An addition to the mmi facial expression database. *Proc. Int'l Conf. Language Resources and Evaluation, Workshop EMOTION*, 65–70.
Vandenhende, S., De Brabandere, B., Neven, D., & Van Gool, L. (2019). A Three-Player GAN: Generating Hard Samples To Improve Classification Networks. Retrieved from http://arxiv.org/abs/1903.03496
Vinyals, O., Blundell, C., Lillicrap, T. P., Kavukcuoglu, K., & Wierstra, D. (2016). Matching Networks for One Shot Learning. *CoRR*, *abs/1606.04080*. Retrieved from http://arxiv.org/abs/1606.04080
Volpi, R., Morerio, P., Savarese, S., & Murino, V. (2018). Adversarial Feature Augmentation for Unsupervised Domain Adaptation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 5495–5504. https://doi.org/10.1109/CVPR.2018.00576
Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2011). *The Caltech-UCSD Birds-200-2011 Dataset*.
Wang, G., Kang, W., Wu, Q., Wang, Z., & Gao, J. (2019). Generative Adversarial Network (GAN) Based Data Augmentation for Palmprint Recognition. *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 1–7. https://doi.org/10.1109/DICTA.2018.8615782
Wang, X., Wang, K., & Lian, S. (2019). A Survey on Face Data Augmentation. Retrieved from https://arxiv.org/abs/1904.11685

Wang, Z., She, Q., & Ward, T. E. (2019). Generative Adversarial Networks: A Survey and Taxonomy. Retrieved from https://arxiv.org/abs/1906.01529

Wei, X., Gong, B., Liu, Z., Lu, W., & Wang, L. (2018). Improving the Improved Training of Wasserstein GANs: A Consistency Term and Its Dual Effect. Retrieved from https://arxiv.org/abs/1803.01541

Wu, E., Wu, K., Cox, D., & Lotter, W. (2018). Conditional infilling GANs for data augmentation in mammogram classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11040 LNCS, pp. 98–106). Springer Verlag. https://doi.org/10.1007/978-3-030-00946-5_11

Wu, J., Zhang, C., Xue, T., Freeman, W. T., & Tenenbaum, J. B. (2016). Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *CoRR*, *abs/1610.07584*. Retrieved from http://arxiv.org/abs/1610.07584

Wu, Y, Lim, J., & Yang, M. (2013). Online Object Tracking: A Benchmark. In *2013 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2411–2418). https://doi.org/10.1109/CVPR.2013.312

Wu, Yirui, Yue, Y., Tan, X., Wang, W., & Lu, T. (2018). End-To-End Chromosome Karyotyping with Data Augmentation Using GAN. *Proceedings - International Conference on Image Processing, ICIP*, 2456–2460. https://doi.org/10.1109/ICIP.2018.8451041

Xiang, Y., Kim, W., Chen, W., Ji, J., Choy, C., Su, H., … Savarese, S. (2016). ObjectNet3D: A Large Scale Database for 3D Object Recognition (Vol. 9912, pp. 160–176). https://doi.org/10.1007/978-3-319-46484-8_10

Yan, K., Wang, X., Lu, L., Zhang, L., Harrison, A. P., Bagheri, M., & Summers, R. M. (2018). Deep Lesion Graphs in the Wild: Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-Scale Lesion Database. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00965

Yan, M., Liu, K., Guan, Z., Xu, X., Qian, X., & Bao, H. (2018). Background Augmentation Generative Adversarial Networks (BAGANs): Effective Data Generation Based on GAN-Augmented 3D Synthesizing. *Symmetry*, *10*, 734. https://doi.org/10.3390/sym10120734

Yang, J, Zhao, Z., Zhang, H., & Shi, Y. (2019). Data Augmentation for X-Ray Prohibited Item Images Using Generative Adversarial Networks. *IEEE Access*, *7*, 28894–28902. https://doi.org/10.1109/ACCESS.2019.2902121

Yang, Jie, Liu, S., Grbic, S., Setio, A. A. A., Xu, Z., Gibson, E., … Comaniciu, D. (2018). Class-Aware Adversarial Lung Nodule Synthesis in {CT} Images. *CoRR*, *abs/1812.11204*. Retrieved from http://arxiv.org/abs/1812.11204

Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning Face Representation from Scratch. *CoRR*, *abs/1411.7923*. Retrieved from http://arxiv.org/abs/1411.7923

Yi, X., Walia, E., & Babyn, P. (2018). Generative Adversarial Network in Medical Imaging: A Review. Retrieved from http://arxiv.org/abs/1809.07294

Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2018). Self-Attention Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/1805.08318

Zhang, J., Jiao, J., Chen, M., Qu, L., Xu, X., & Yang, Q. (2016). 3D Hand Pose Tracking and Estimation Using Stereo Matching. *CoRR*, *abs/1610.07214*. Retrieved from http://arxiv.org/abs/1610.07214

Zhang, R., Che, T., Grahahramani, Z., Bengio, Y., & Song, Y. (2018). MetaGAN: An Adversarial Approach to Few-Shot Learning. *Advances in Neural Information Processing Systems 31*. Retrieved from http://papers.nips.cc/paper/7504-metagan-an-adversarial-approach-to-few-shot-learning.pdf

Zhang, X., Wang, Z., Liu, D., & Ling, Q. (2018). DADA: Deep Adversarial Data Augmentation for Extremely Low Data Regime Classification, 2807–2811. Retrieved from http://arxiv.org/abs/1809.00981

Zhang, Y., Jia, G., Chen, L., Zhang, M., & Yong, J. (2019). Self-Paced Video Data Augmentation with Dynamic Images Generated by Generative Adversarial Networks.

Zhang, Z., Yang, L., & Zheng, Y. (2018). Translating and Segmenting Multimodal Medical Volumes with Cycle- and Shape-Consistency Generative Adversarial Network. *CoRR*, *abs/1802.09655*.

Retrieved from http://arxiv.org/abs/1802.09655
Zhao, G., Huang, X., Taini, M., Li, S. Z., & Pietikäinen, M. (2011). Facial expression recognition from near-infrared videos. *Image and Vision Computing*, *29*(9), 607–619. https://doi.org/https://doi.org/10.1016/j.imavis.2011.07.002
Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable Person Re-identification: A Benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 1116–1124). https://doi.org/10.1109/ICCV.2015.133
Zheng, Z., Zheng, L., & Yang, Y. (2017). Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 3774–3782. https://doi.org/10.1109/ICCV.2017.405
Zhong, Z., Zheng, L., Zheng, Z., Li, S., & Yang, Y. (2017). Camera Style Adaptation for Person Re-identification. Retrieved from http://arxiv.org/abs/1711.10295
Zhou, S., Ke, M., & Luo, P. (2019). Multi-camera transfer GAN for person re-identification. *Journal of Visual Communication and Image Representation*. https://doi.org/10.1016/j.jvcir.2019.01.029
Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 2242–2251. https://doi.org/10.1109/ICCV.2017.244
Zhu, Y., Aoun, M., Science, C., Krijn, M., & Vanschoren, J. (2018). Data Augmentation using Conditional Generative Adversarial Networks for Leaf Counting in Arabidopsis Plants. *Computer Vision Problems in Plant Phenotyping (CVPPP2018)*, 1–11. Retrieved from https://www.semanticscholar.org/paper/Data-Augmentation-using-Conditional-Generative-for-Zhu-Aoun/0636eb841bf3480309a346587010f43f2a87633e
Zhuang, P., Schwing, A. G., & Koyejo, S. (2019). {FMRI} data augmentation via synthesis. *CoRR*, *abs/1907.06134*. Retrieved from http://arxiv.org/abs/1907.06134
Zimmermann, C., & Brox, T. (2017). *Learning to Estimate 3D Hand Pose from Single RGB Images*. Retrieved from https://lmb.informatik.uni-freiburg.de/projects/hand3d/
Zoumpourlis, G., Doumanoglou, A., Vretos, N., & Daras, P. (2017). Non-linear convolution filters for cnn-based learning. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4761–4769).