Semi-supervised Learning in Generative Adversarial Networks

Ali Honarmand
57526155
Department of Computer Science
University of British Columbia
ali.honarmand@gmail.com

Ainaz Hajimoradlou
98379167
Department of Computer Science
University of British Columbia
ainaz@cs.ubc.ca

Farzad Abdolhosseini
98498165
Department of Computer Science
University of British Columbia
farzadab@cs.ubc.ca

Abstract

In this paper we explore the research at the intersection of generative adversarial networks (GANs) and semi-supervised learning. Training GANs with the additional information of class labels can enhance the quality and controllability of the generated samples. On the other hand, the structural knowledge captured by the network enables us to improve upon supervised learning. We try to categorize the research in this area by identifying the different trends. In addition, we briefly summarize the works and explain our own understanding of their strengths and weaknesses.

1 Introduction

There has been considerable interest in GANs in recent years. Most of the early research was in the realm of unsupervised learning. The focus was on generating samples that look like the true data distribution. Such models were limited in application, but recently different extensions of this framework have been used to tackle a large set of problems. One class of these methods are concerned with relaxing the data requirement of supervised learning. Another class focuses on using class label information in order to better guide the sample generation process and increasing controllability. Semi-supervised learning is at the heart of both of these problems, since gathering labeled data compared to unlabeled data is expensive in most applications.

Deep supervised learning methods have had remarkable success in the past few years in both research and industry. However, their success is highly dependent on the availability of vast amounts of labeled data. This problem motivates the use of semi-supervised learning in which easily available unlabeled data is used to guide the supervised learning process. These methods can be trained to work well with a rather small set of labeled data. In section 3.1 we will consider how the GAN framework can be integrated with almost any available neural network classifier in order to make use of unlabeled data.

In the original GAN paper, the user has no control over the samples that the model generates. In section 2.2 we briefly introduce the reader to the conditional GAN (CGAN) paper [20], which addresses this issue. There have been many works that build on CGANs to increase the controllability of the generative model. However, all of these models require labeled data to train, which is expensive to obtain. In section 3.2 we discuss the semi-supervised methods that focus on improving the quality and controllability of the produced samples.

We then explore some concrete applications of these methods in section 3.3 and present some theoretical results in section 3.4. Finally, we discuss their strengths and shortcomings in section 4.

2 Preliminaries

2.1 Generative Adversarial Networks

GAN framework introduced in [11] is a deep generative model that tries to approximately follow the data distribution. It consists of a generator G that can generate samples and a discriminator D that tries to distinguish between the fake and the true samples. This can be modeled as a minimax game. Both networks G and D can be multilayer perceptrons, and the objective function can be written as follows:

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

where $\boldsymbol{x}$ comes from the data distribution and $\boldsymbol{z}$ is an input noise vector.

2.2 Conditional Generative Adversarial Networks

GANs can be extended to a conditional model [20] if both the generator and the discriminator are conditioned on some extra information y. In the generator the input noise $p_z(\boldsymbol{z})$ is combined with the condition variable y to form a joint hidden representation which is then fed into the network. Both $\boldsymbol{x}$ and y inputs to the discriminator. This will result in the following objective:

$$\min_G \max_D L(D, G) = \mathbb{E}_{\boldsymbol{x}, y \sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x}|\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{z} \sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})|\boldsymbol{y}))].$$

2.3 Semi-Supervised Learning

Previously, semi-supervised learning has had some success by incorporating knowledge from the unlabeled data. Convolutional ladder network[24], a previous state-of-the-art method, achieves an error rate of 0.89% on MNIST using only 100 labeled samples; while the baseline supervised method that they compare to has an error rate of 6.43%. Ladder networks integrate a denoising autoencoder [34] into the existing neural network architecture and combine the original supervised learning loss with the unsupervised denoising loss. One of their advantages is that they can be incorporated into any neural network architecture, i.e. both feedforward and recurrent networks. Also, [4] argues that even though the computational cost almost triples by adding the decoder networks, the training time doesn't necessarily increase as much, since the training data is better utilized.

3 Review

3.1 Multi-class Discriminator

As in section 2.3, the methods presented here aim to improve the accuracy of the classifier by making use of the unlabeled data. A standard classifier in a supervised setting tries to classify each data point $\boldsymbol{x}$ into one of the K possible categories by outputing a K-dimensional vector and applying the softmax to get the class probabilities $p_{model}(y|\boldsymbol{x})$. This can easily be extended to the semi-supervised setting by assigning an extra label $y = K + 1$ to the fake generated samples from G. Now, unlabeled data can also be used by maximizing $p_{model}(y \neq K + 1|\boldsymbol{x})$. It is shown in [27] that this approach can also increase the quality of the generated images. A motivation for this idea is a hypothesis which argues that when humans look at an image they tend to classify the objects that they see; and they only call it "fake" if it doens't fit into any of the other known categorizes. The discriminator here is trying to do something similar.
The loss function for this method can be defined as a combination of the supervised and the unsupervised loss, such as the following used in [27]:

$$
\begin{aligned}
L &= \mathbb{E}_{\boldsymbol{x},y\sim p_{data}(\boldsymbol{x})} \log p_{model}(y|\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{x}\sim G(\boldsymbol{z})} \log p_{model}(y=K+1|\boldsymbol{x}) \\
&= L_{supervised} + L_{unsupervised}, \text{ where} \\
L_{supervised} &= \mathbb{E}_{\boldsymbol{x},y\sim p_{data}(\boldsymbol{x})} \log p_{model}(y|\boldsymbol{x}, y \neq K+1) \\
L_{unsupervised} &= \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})} \log[1 - p_{model}(y=K+1|\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{x}\sim G(\boldsymbol{z})} \log p_{model}(y=K+1|\boldsymbol{x}).
\end{aligned}
$$

Similar formulations of this idea have been leveraged in [8, 17, 21, 27, 30] to increase the accuracy of the supervised methods and achieve comparable or even better than state-of-the-art in semi-supervised learning [18, 24]. SGAN[21] shows that forcing the discriminator and the classifier to share weights improves data-efficiency but the results are evaluated on a small subset of MNIST. On the other hand, [30] gives a more thorough evaluation of their proposed categorical GAN for unsupervised and semi-supervised framework. Their proposed model also acts as a regularizer for the discriminatively trained classifier. The classification performance is competitive with the state-of-the-art models for semi-supervised image classification and they are able to generate images with high fidelity. [27] proposes some practical techniques to improve the training process, such as feature matching and minibatch discrimination. The results show that feature matching improves the accuracy of the classifier but doesn't improve the quality of the generated features. On the other hand, minibatch discrimination increases the quality of the images but doesn't help with the classification.

Even though this combined supervised and unsupervised loss has been used in many works in computer vision and NLP such as [16, 18, 35, 36], this framework naturally does not work well with textual data. This is largely due to the fact that the generator network is designed to adjust the output continuously. [9] proposes discriminative adversarial network (DAN) framework to overcome this issue. Instead of using the generator/discriminator network in GAN, they use two discriminators: a *predictor* network and a the *judge* network. Predictor, denoted by P, produces the prediction y given the data input $\boldsymbol{x}$ and judge, denoted by J, takes in a pair $(\boldsymbol{x}, y)$ as input and decides whether this is a fake or real pair. This results in the following objective:

$$
\min_P \max_J L(J,P) = \mathbb{E}_{\boldsymbol{x},y\sim p_{data}(\boldsymbol{x},y)}[\log J(\boldsymbol{x},y)] + \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log(1 - J(\boldsymbol{x}, P(\boldsymbol{x})))]
$$

The proposed approach in [9] does not consist of a generator for generating samples so, it can naturally be used in discrete domains as well as continuous. The framework can be seen as a method to learn loss functions for predictors. Moreover, the unlabeled data can be used transparently as the predictor does not need to use labeled data. More importantly, there's no need to manually define a combined loss function of supervised and unsupervised loss. The judge implicitly learns a loss function to optimize the predictor. The proposed model is evaluated on two NLP tasks, answer selection and text classification, and has shown promising results compared to the state-of-the-art methods.

Denton et al. [8] creates a conditional GAN that is conditioned on an image with a removed patch, called CC-GAN. They use the output of generator to create a full image that will be fed into the discriminator. Their goal is to use the discriminator of this conditional GAN in supervised learning and they claim that training on incomplete images will act as a regularizer for the discriminator. For a labeled image they use both the supervised and unsupervised loss to calculate the gradient and update the discriminator variables. For an unlabeled image or an image generated by the generator only the gradient from unsupervised loss can be used to update the variables of the generator and the discriminator. They use this same technique to train a normal GAN in this semi-supervised way (SSL-GAN) and they use it as a baseline. They compare the classification performance of CC-GAN's discriminator to the discriminator of SSL-GAN. In the STL-10 dataset CC-GAN has almost 4% higher accuracy. However, they fail to clearly show the advantage of CC-GAN in other datasets that that they experimented on.

3.2 Advances in Architecture For Generation of Better Samples

In this section we discuss three papers that aim to use semi-supervised learning to enhance the quality and controllability of the generated images. Sricharan et al. [32] and Bodla et al. [3] do this by making changes to the architecture of the model, while Spurr et al. [31] proposes adding two code vectors to the model which can be used as knobs to control what the model will generate.

Sricharan et al. [32] proposes an architecture for semi-supervised training of conditional GANs. The main idea is to split the discriminator into an unsupervised discriminator D_u and a supervised discriminator D_s (Figure 1). D_u outputs whether image $\boldsymbol{x}$ is real or fake while D_s outputs whether $(\boldsymbol{x}, y)$ is real or fake. D_s only receives a representation of the image from D_u and never the image itself. This intuitively has a few advantages. Firstly, it can be trained in a semi-supervised manner. Gradient from unsupervised loss can be used to update D_u and G, and the gradient from supervised loss can be used to update the whole network. Secondly, because only a representation of the image is passed to D_s, the model should be less likely to overfit to the few labeled examples that it is being trained on. This is in contrast to Denton et al. [8] feeding the image directly to the supervised discriminator. Despite the fact that this architecture makes intuitive sense, it fails to consistently do better than the semi-supervised CGAN variant that the authors are comparing to. The authors make claims about the better quality of the images produced by their model compared to the baseline, but their data does not fully support this. For example, their model produces higher quality images, but the images do not correspond well to the description they were conditioned on. They also fail to beat the baseline in the quantitative measure in various tests.

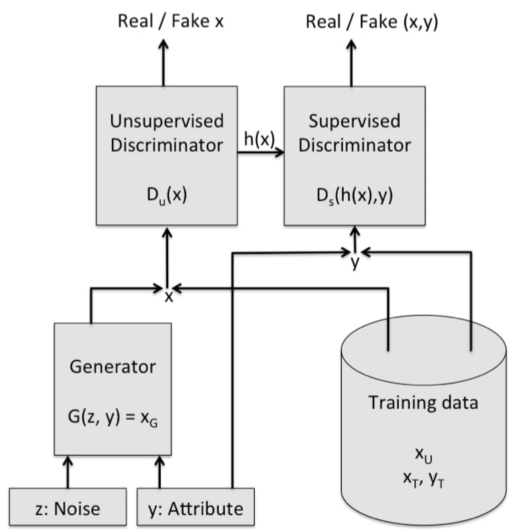

Figure 1: The Semi-supervised GAN architecture proposed by [32].

Bodla et al. [3] aims to improve the controllability of the images produced by the generator. They build on StackGAN [37] which creates a conditional GAN to produce images conditioned on a text description. Using their model called FusedGAN, one can fix the posture in the image by using a fixed noise vector and add styling to the posture using the text description. They achieve this by creating a novel GAN architecture (Figure 2). The architecture consists of an unsupervised generator G_1 and an unsupervised discriminator D_u. G_1 is further broken down into two generators G_s and G_u. The idea is that G_s outputs a representation of the posture (M_s). M_s can either be fed directly into G_u to generate an unsupervised image, or it can be concatenated with an encoding of the text description and fed into another GAN network (with its own generator G_c and discriminator D_u) to produce a an image conditioned on the text description. The conditioned image will have the posture captured in M_s, but it will have the style specified by the text description. FusedGAN is able to produce higher quality images compared to previous text to image studies. They were able to get higher inception scores and human scores compared to previous StackGAN-I [37] and GAN-INT-CLS [25] models. This is probably due to the semi-supervised nature of the approach that allows it to learn from unlabeled images.

Spurr et al. [31] is also mainly concerned with increasing the controllability of GAN. In addition to the noise vector $\boldsymbol{z}$, their model also accepts a c_{ss} vector and a c_{us} vector as input, which are dubbed semi-supervised code and unsupervised code, respectively. Their model tries to maximize the mutual information between c_{ss} and labeled examples as well as the generated examples. With some assumptions they prove that this leads to increased mutual information between real and

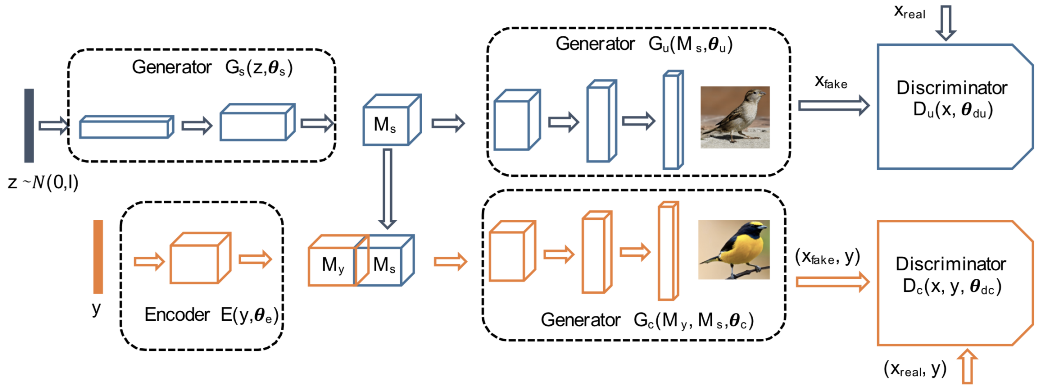

Figure 2: The FusedGAN architecture. Blue part of the graph corresponds to unconditional image generation while the orange part corresponds to conditional image generation. The posture information is captured in M_s and is used by both G_u and G_c. [3]

generated examples. They also try to maximize the mutual information between c_{us} and generated examples. In this setting c_{ss} will learn the labeled categories while the c_{us} is free to learn unsupervised semantics. They show very convincing evidence that their framework works better than the an earlier InfoGAN [6] framework that their model is based on. They were able to successfully generate images conditioned on a label by changing the input c_{ss}. In addition, c_{us} learned interpretable features such as the angle of a digit in MNIST. They also got very impressive results in terms of the proportion of labeled samples that they used. The proportion of labeled samples in their datasets varied between as little as 0.22% to a maximum of 10%. The only dataset where they did not get very good results on was the CIFAR-10. However, it should be noted that this dataset is very diverse and we have not come across any other GAN framework that generates high quality images on this dataset.

3.3 Applications

Most methods covered so far are generic approaches and mostly use standard supervised datasets such as MNIST, CIFAR-10, and SVHN for evaluation. But the most important question is whether these approaches can be useful in a real-world setting. The authors in [1, 5, 12, 14, 28, 38] apply semi-supervised learning with GANs in novel problems that we don't yet have good solutions for.

GONet [14] tackles the traversability estimation problem for a robot. The robot is equipped with a front-facing sensor, such as a camera, and it should decide whether it can move forward without colliding with an object, falling down the stairs, etc. Here, instead of using expensive sensors such as lidar [29], the authors use a cheap fish-eye camera and enhance its functionality with semi-supervised learning. The problem they face is that gathering positive (traversable) images is really easy and they even propose a simple approach to collect a huge amount of positive images. However, collecting negative (non-traversable) images can be costly and potentially dangerous to both the robot and other people. Hence, they make use of a hybrid autoencoder and DCGAN [23] network in order to learn the internal structure and similarities in the positive examples. Afterwards, the classifier makes use of this structural knowledge to better discriminate between the positive and negative samples. Their method also seems to generalize well beyond the setting in which it was trained on, i.e. it works when a person carries the camera when walking outside of a building instead of the camera being mounted on a robot inside a building.

In [28], the authors employ methods similar to what was discussed in section 3.1 for the problem of semantic segmentation of images. Getting pixel-level labeling of images for segmentation is a time-consuming and laborious task and the current state-of-the-art requires a lot more data to achieve high performance. Plus, the authors point out that even the semi-supervised methods in the past have not been very successful for this task. The proposed method seems to slightly improve on the previous approaches. Next, they introduce a method which they call *Weakly-supervised*, that is, they provide class labels for the unlabeled data but not the pixel-level segmentation. The argument is that it's a lot easier to get class labels for images than pixel-level annotation. In order to make use of these class labels, they slightly change the architecture along with the objective and use a conditional GAN instead of a normal GAN. The structure used in both networks can be found in Figure 3.

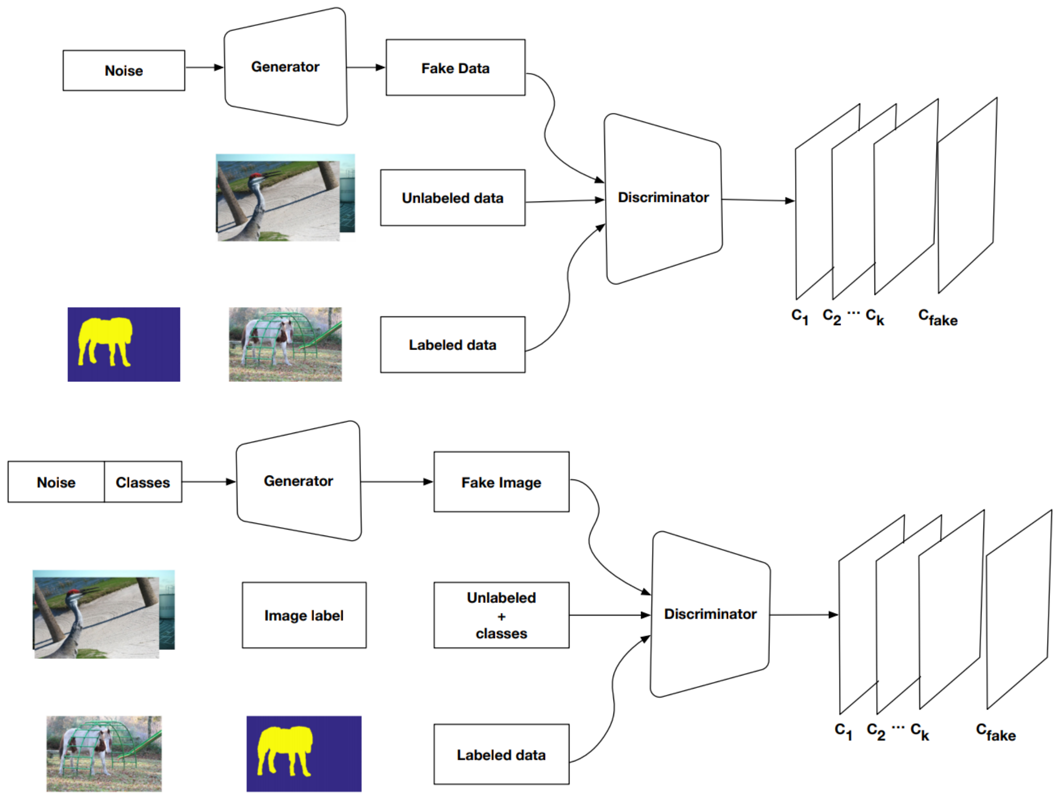

Figure 3: The network structures used in [28]. The top image corresponds to the network used for the semi-supervised task. The bottom image corresponds to the network used in the weakly-supervised task which uses a conditional GAN that gets as input the class labels.

Motivated by this result, [12] applies these techniques to the problem of detecting patches of road in an images. Similar to the previous work, they first train a semi-supervised model by making use of unlabeled data. Then, they train another model using weak-supervision by annotating the images with the shape of the road as a class label. Both of the proposed architectures seem to outperform the supervised and semi-supervised state-of-the-art, but unlike the previous work, weak-supervision does not seem to increase the accuracy much more than the simpler semi-supervised method.

Furthermore, [38] applies semi-supervised learning to the problem of cross-modal information retrieval. The idea is to develop a search algorithm that can retrieve information from one modality, e.g. image, based on a query from another modality, e.g. text. One of the main ways to solve this problem has been to use cross-modal hashing. In these methods, different modalities are projected into a common space using hash functions where similar data is assigned similar values. Then, fast Hamming distance methods can be used to retrieve similar data from other modalities. However, the presence of multiple modalities in the data makes them really hard to annotate. Here, the authors design an elaborate combination of architecture, loss function, and training scheme in order to make use of the unlabeled data. Even though their method seems to outperform the state-of-the-art, the process that they use combines the unstable adversarial learning with the erratic policy-gradients approach. Thus, without further analysis, their method might as well suffer from convergence issues.

Cai et al. [5] investigate whether the samples generated by GANs are able to do spoofing attacks on speaker recognition systems. They first obtain samples from sampleRNN [19] and waveNet [33], but these samples fail to trick a CNN-based speaker recognition system. They then propose a *mixed loss* for the discriminator which is a modification to the Wasserstein GAN (WGAN) [2] objective function. This objective encourages the discriminator to classify untargeted labels as fake. The samples generated from the trained WGAN are successful in performing both targeted and untargeted adversarial attacks on current CNN-based speaker recognition systems.

3.4 Theoretical Results

Although semi-supervised learning based on generative adversarial networks has achieved strong empirical results [3, 18, 30], it is still not clear how the discriminator benefits from joint training with the generator. Besides, it's not obvious why a good semi-supervised classifier cannot be obtained with a good generator simultaneously.

[17] tries to shed some light on the first problem. Using a case by case analysis, the author tries to demonstrate that the unsupervised loss would only be advantageous to the classifier when the generator only produces "moderate" fake samples. They argue that the unsupervised loss for "weak" samples would be approximately zero and thus wouldn't influence the supervised trainig at all. On the other hand, if the samples are too strong, we will be in one of the two settings: either the classifier has a high capacity or it doesn't. If it does, it can increase the curvature at the midpoint to discriminate between the fake and true labels, and therefore overfits to the labels. If it's regularized or it doesn't have enough capacity, the classifier can't discriminate between the fake and true samples. In this case they argue that the entropy of the distribution might increase and thus hurt the performance.

At the same time, [7] theoretically proves that in the multiclass discriminator formulation (section 3.1) good semi-supervised classification requires a bad generator. In other words, the generator distribution should not match the true data distribution. This results in a new definition, the *complement* generator, which should generate complement samples in the feature space. Intuitively, this means sampling from the complement space of the data distribution manifold (please refer to the original paper for the exact definition). The authors demonstrate that given a complement generator, a properly optimized discriminator is able to obtain correct decision boundaries in high-density areas in the feature space. [7] also proposes a new formulation of the generator and discriminator objective to improve drawbacks of feature matching which is one of the proposed techniques in [27] to improve training for GANs (section 3.1).

3.5 Beyond Traditional Architectures

Hinton et al. [13] introduced a promising framework called capsule networks which were shown to be powerful alternatives to CNNs in [26]. Moreover, the authors in [15] integrate these networks with GANs and call their model CapsuleGAN. Their network shows promising results in semi-supervised classification, but it still requires more experiments. They show that their model outperforms GANs in the semi-supervised task, but oddly enough, they don't compare their result with any of the methods that we discussed here. In contrast, they compare CapsuleGAN with their own version of convolutional GAN and show that it performs better. However, the reported error rates are much worse what was previously reported in [27] and [30].

4 Discussion

In this paper we explored the research at the intersection of GANs and semi-supervised learning. One approach improves upon supervised learning by incorporating the information from unlabeled data. These methods generally combine the discriminator and the classifier into one network and use a combined supervised and unsupervised loss for training. They have had some success over the previous state-of-the-art in semi-supervised learning. But [22] casts doubt on their real world applicability. The size of the validation sets used in many SSL methods are much larger than the realistic size. This is not representative of the real world where small validation sets lead to much noisier objective values during hyper parameter tuning. In a similar argument, we can confirm that [10] may not have much real world applicability since they optimize their hyper parametes on a validation set which is 10 times bigger than the training set. Besides, varying the number of labeled and unlabeled data heavily affects the performance. Therefore, we think that semi-supervised learning requires better established benchmarks to avoid confusion about the generalization of the methods and their real world applicability. This can also solve the latter issue of different levels of sensitivity between SSL techniques based on the number of labeled and unlabeled data.

Another approach focuses on improving the quality and controllability of the generated images by providing extra information through class labels. These methods generally make modifications to the original GAN architecture and/or the loss function to allow for semi-supervised training of the model. In some cases, providing even one extra labeled sample per class can highly improve the model to generate proper samples. This has even led to increased training speed in [31]. These semi-supervised GAN models have been able to generate impressive images in simple homogeneous datasets such as MNIST, CelebA, and SVHN. But the results are still far from perfect in more complex datasets such as CIFAR-10. Most proposed architectures [3, 31, 32] are based on DCGAN[23] which can be considered one of the most influential papers in this area. DCGAN substantially improved training, but their results were again based on homogeneous datasets such as faces and bedroom images. We believe that conditional GANs will shine when the model is trained on diverse datasets. Thus, figuring out how to do that will probably be the next step. To be able to create more general models, researchers might need to look into more flexible architectures. CapsuleGANs [15] might be one such architecture.

References

[1] Unaiza Ahsan, Chen Sun, and Irfan A. Essa. Discrimnet: Semi-supervised action recognition from videos using generative adversarial networks. *CoRR*, abs/1801.07230, 2018. URL `http://arxiv.org/abs/1801.07230`.

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *ArXiv e-prints*, January 2017.

[3] Navaneeth Bodla, Gang Hua, and Rama Chellappa. Semi-supervised fusedgan for conditional image generation. *CoRR*, abs/1801.05551, 2018. URL `http://arxiv.org/abs/1801.05551`.

[4] Rinu Boney. Introduction to Semi-Supervised Learning with Ladder Networks . `https://rinuboney.github.io/2016/01/19/ladder-network.html`, 2016. [Online; accessed 25-April-2018].

[5] Wilson Cai, Anish Doshi, and Rafael Valle. Attacking speaker recognition with deep generative models. *CoRR*, abs/1801.02384, 2018. URL `http://arxiv.org/abs/1801.02384`.

[6] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. *ArXiv e-prints*, June 2016.

[7] Zihang Dai, Zhilin Yang, Fan Yang, William W. Cohen, and Ruslan Salakhutdinov. Good semi-supervised learning that requires a bad GAN. *CoRR*, abs/1705.09783, 2017. URL `http://arxiv.org/abs/1705.09783`.

[8] Emily L. Denton, Sam Gross, and Rob Fergus. Semi-supervised learning with context-conditional generative adversarial networks. *CoRR*, abs/1611.06430, 2016. URL `http://arxiv.org/abs/1611.06430`.

[9] Cícero Nogueira dos Santos, Kahini Wadhawan, and Bowen Zhou. Learning loss functions for semi-supervised learning via discriminative adversarial networks. *CoRR*, abs/1707.02198, 2017. URL `http://arxiv.org/abs/1707.02198`.

[10] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville. Adversarially Learned Inference. *ArXiv e-prints*, June 2016.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`.

[12] Xiaofeng Han, Jianfeng Lu, Chunxia Zhao, Shaodi You, and Hongdong Li. Semisupervised and weakly supervised road detection based on generative adversarial networks. *IEEE Signal Process. Lett.*, 25(4):551–555, 2018. doi: 10.1109/LSP.2018.2809685. URL `https://doi.org/10.1109/LSP.2018.2809685`.

[13] Geoffrey E. Hinton, Alex Krizhevsky, and Sida D. Wang. Transforming auto-encoders. In *Proceedings of the 21th International Conference on Artificial Neural Networks - Volume Part I*, ICANN'11, pages 44–51, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-21734-0. URL http://dl.acm.org/citation.cfm?id=2029556.2029562.

[14] N. Hirose, A. Sadeghian, M. Vázquez, P. Goebel, and S. Savarese. GONet: A Semi-Supervised Deep Learning Approach For Traversability Estimation. *ArXiv e-prints*, March 2018.

[15] A. Jaiswal, W. AbdAlmageed, Y. Wu, and P. Natarajan. CapsuleGAN: Generative Adversarial Capsule Network. *ArXiv e-prints*, February 2018.

[16] Diederik P. Kingma, Danilo Jimenez Rezende, Shakir Mohamed, and Max Welling. Semi-supervised learning with deep generative models. *CoRR*, abs/1406.5298, 2014. URL http://arxiv.org/abs/1406.5298.

[17] Abhishek Kumar, Prasanna Sattigeri, and P. Thomas Fletcher. Improved semi-supervised learning with gans using manifold invariances. *CoRR*, abs/1705.08850, 2017. URL http://arxiv.org/abs/1705.08850.

[18] L. Maaløe, C. Kaae Sønderby, S. Kaae Sønderby, and O. Winther. Auxiliary Deep Generative Models. *ArXiv e-prints*, February 2016.

[19] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio. SampleRNN: An Unconditional End-to-End Neural Audio Generation Model. *ArXiv e-prints*, December 2016.

[20] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014. URL http://arxiv.org/abs/1411.1784.

[21] A. Odena. Semi-Supervised Learning with Generative Adversarial Networks. *ArXiv e-prints*, June 2016.

[22] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow. Realistic Evaluation of Deep Semi-Supervised Learning Algorithms. *ArXiv e-prints*, April 2018.

[23] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015. URL http://arxiv.org/abs/1511.06434.

[24] Antti Rasmus, Harri Valpola, Mikko Honkala, Mathias Berglund, and Tapani Raiko. Semi-supervised learning with ladder network. *CoRR*, abs/1507.02672, 2015. URL http://arxiv.org/abs/1507.02672.

[25] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative Adversarial Text to Image Synthesis. *ArXiv e-prints*, May 2016.

[26] Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. *CoRR*, abs/1710.09829, 2017. URL http://arxiv.org/abs/1710.09829.

[27] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *CoRR*, abs/1606.03498, 2016. URL http://arxiv.org/abs/1606.03498.

[28] Nasim Souly, Concetto Spampinato, and Mubarak Shah. Semi and weakly supervised semantic segmentation using generative adversarial network. *CoRR*, abs/1703.09695, 2017. URL http://arxiv.org/abs/1703.09695.

[29] Seydou Soumare and Akihisa Ohya. Real-time obstacle avoidance by an autonomous mobile robot using an active vision sensor and a vertically emitted laser slit. In *in Intelligent Autonomous Systems 7, IOS*. Press, 2002.

[30] J. T. Springenberg. Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks. *ArXiv e-prints*, November 2015.

[31] Adrian Spurr, Emre Aksan, and Otmar Hilliges. Guiding infogan with semi-supervision. *CoRR*, abs/1707.04487, 2017. URL `http://arxiv.org/abs/1707.04487`.

[32] K. Sricharan, R. Bala, M. Shreve, H. Ding, K. Saketh, and J. Sun. Semi-supervised Conditional GANs. *ArXiv e-prints*, August 2017.

[33] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *ArXiv e-prints*, September 2016.

[34] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390294. URL `http://doi.acm.org/10.1145/1390156.1390294`.

[35] Weidi Xu and Haoze Sun. Semi-supervised variational autoencoders for sequence classification. *CoRR*, abs/1603.02514, 2016. URL `http://arxiv.org/abs/1603.02514`.

[36] Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov, and Taylor Berg-Kirkpatrick. Improved variational autoencoders for text modeling using dilated convolutions. *CoRR*, abs/1702.08139, 2017. URL `http://arxiv.org/abs/1702.08139`.

[37] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, and Dimitris N. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. *CoRR*, abs/1612.03242, 2016. URL `http://arxiv.org/abs/1612.03242`.

[38] Jian Zhang, Yuxin Peng, and Mingkuan Yuan. SCH-GAN: semi-supervised cross-modal hashing by generative adversarial network. *CoRR*, abs/1802.02488, 2018. URL `http://arxiv.org/abs/1802.02488`.