Supervised vs Semi-Supervised CNN Using Handwritten Recognition

Abstract—**Semi-supervised techniques have experienced an increase in interest due to recent developments in deep learning. One contemporary method for semi-supervised learning is generative adversarial networks (GANs) (SSL). GAN has been used for SSL utilizing GANs is presented. Pseudo-labeling/classification, encoder-based, TripleGAN-based, two GAN, manifold regularization, and stacking discriminator techniques are some of the previous GAN applications to SSL that have been studied. It presents a quantitative and qualitative study of the various strategies. We present a comparative study between supervised and semi-supervised techniques in hand written digits recognition. Future research prospects including the adaption of SSL components into GAN-based solutions are also identified, given the recent success of non-GAN-based techniques for SSL.**
Index Terms—**Generative Adversarial Networks, Semi-Supervised Learning, Deep Learning**

I. Introduction

Recent advancements in deep learning and its applications have resulted in an expansion and diversification of research opportunities in the field. Semi-supervised learning is among these approaches (SSL). SSL, in contrast to supervised learning, is a method of learning that may learn from incomplete data in which only a portion of the data is labeled [1]. In supervised learning, the training data consist of a collection of data points and a label for each point. In contrast, in unsupervised learning, the training data consist simply of data points without any output, necessitating a procedure that identifies novel data structures and groups [2]. In instances where a small number of labeled training samples and a large number of unlabeled data points are available, semi-supervised learning is utilized. While supervised learning has been the technique of choice for the majority of classification tasks, labeled data can be difficult to obtain and the labeling procedure can be costly and time-consuming [3]. SSL eliminates the requirement for big, labeled datasets by employing some labeled data but primarily unlabeled data.

Semi-supervised learning is based on the premise that the distribution of data over the input space encodes significant information about the distribution of labels in the output space [1]. Most SSL methods will fail if this assumption is not met, as the input space would include no information about the actual labels, and it would therefore be unable to improve accuracy using unlabeled data. According to [3], if the sample distribution of the data does not contain substantial information, the ensuing learning may not be superior to unsupervised learning and may result in an increase in incorrect predictions. The fundamental assumption can be subdivided into the smoothness assumption, the low-density assumption, and the manifold assumption.

II. Related Works

A. Taxonomy

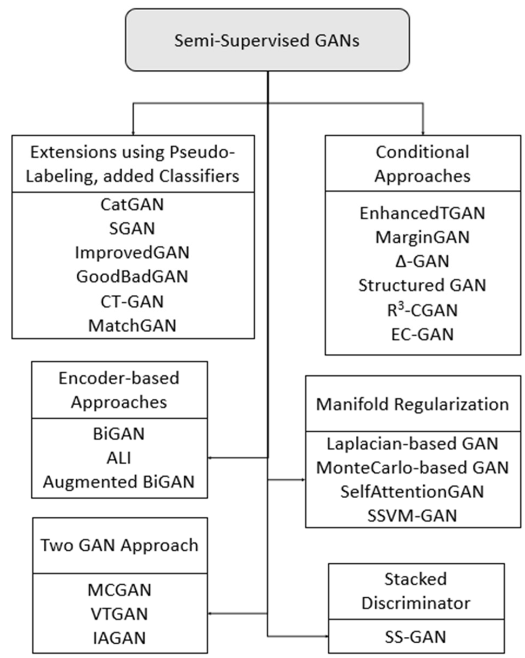

Fig. 1. Surveyed taxonomy papers

Figure 1 depicts a classification of the surveyed papers. As depicted in Figure 1, early efforts to SSL GANs typically comprised pseudo-labeling or the addition of a classifier component to existing GAN models. Numerous models, including CatGAN [4], SGAN, Improved GAN, GoodBadGAN, CT-GAN [5], and MatchGAN, utilized this strategy. Many others employed a conditional method in which both the image and label were supplied into the GAN. EnhancedTGAN, MarginGAN, Triangle-GAN, Structured GAN, R3-CGAN, and EC-GAN all demonstrated this. A third technique comprised of models employing encoder-based approaches, in which an encoder was added to the GAN architecture to map images into a latent space, which aided the training process. BiGAN, ALI, and Augmented BiGAN models utilized this strategy. Recent methods have utilized manifold regularization techniques to make the model more robust to input perturbations. This category includes Laplacian-based GAN, Monte Carlo-based

GAN, SelfAttentionGAN, and SSVM-GAN. Other novel approaches utilized two GANs, such as MCGAN, VTGAN, and IAGAN, and SS-GAN used conditional GANs in a stacked discriminator strategy to distinguish between predicted qualities.

III. Methodologies

A. Generative Adversarial Networks

GANs, or Generative Adversarial Networks, is an approach to generative modeling that employs deep learning techniques such as convolutional neural networks.

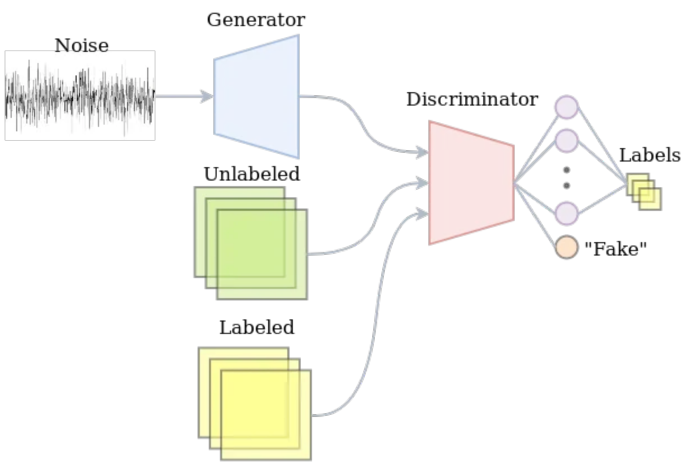

Fig. 2. GAN Architecture

Generative modeling is an unsupervised learning problem in machine learning that entails detecting and learning the regularities or patterns in incoming data so that the model can be used to create or output new instances that plausibly may have been chosen from the original dataset

The generator model is trained to generate new instances, while the discriminator model attempts to categorize samples as either real (from the domain) or fraudulent (generated). The two models are trained together in an adversarial zero-sum game until the discriminator model is deceived approximately fifty percent of the time, indicating that the generator model generates believable examples.

GANs are an exciting and rapidly evolving field, delivering on the promise of generative models by generating realistic examples across a variety of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake.

B. Artificial neural network

Artificial neural networks, also known as ANNs, are computational networks that are inspired by biological neural networks. In this chapter, we concentrate on multilayer perceptrons (MLPs) with backpropagation learning methods rather than any of the other varieties of ANNs that are available. MLPs, which are the kind of ANNs that are utilized the most frequently for a wide range of issues, are built on the premise of a supervised method and consist of three layers: input, hidden, and output. Several features of MLPs, such as their structure, algorithms, data pretreatment, overfitting, and sensitivity analyses, are topics that we cover in this article. In addition, we discuss the benefits and drawbacks of MLPs and advocate for their application in ecological modeling. At long last, an illustration showing how MLP can actually be used in ecological modeling is provided here.

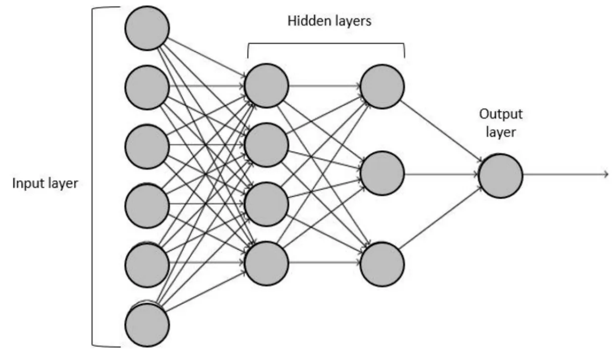

Fig. 3. ANN Architecture

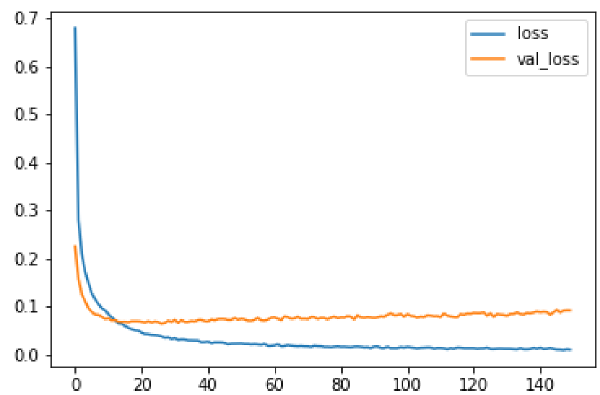

Fig. 4. Loss Curve

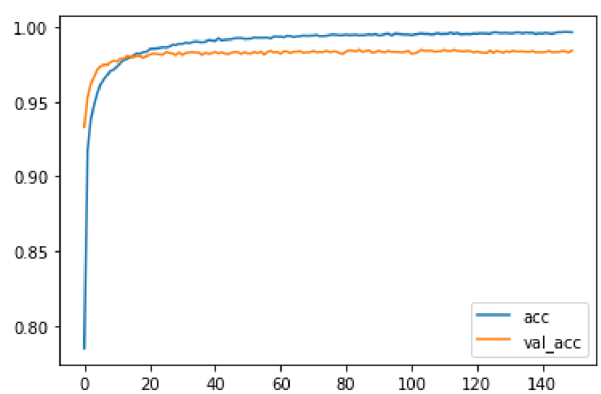

Fig. 5. Accuracy Curve

TABLE I
ANN PARAMETER SIZE

ANN		
conv2d_15 (Conv2D)	(None, 28, 28, 32)	320
conv2d_16 (Conv2D)	(None, 28, 28, 32)	9248
max_pooling2d_6 (MaxPooling 2D)	(None, 14, 14, 32)	0
dropout_16 (Dropout)	(None, 14, 14, 32)	0
conv2d_17 (Conv2D)	(None, 14, 14, 64)	18496
conv2d_18 (Conv2D)	(None, 14, 14, 64)	36928
max_pooling2d_7 (MaxPooling 2D)	(None, 7, 7, 64)	0
dropout_17 (Dropout)	(None, 7, 7, 64)	0
flatten_9 (Flatten)	(None, 3136)	0
dense_32 (Dense)	(None, 256)	803072
dropout_18 (Dropout)	(None, 256)	0
dense_33 (Dense)	(None, 10)	2570
dense_33 (Dense)	(None, 10)	2570
Total params: 870,634		
Trainable params: 870,634		
Non-trainable params: 0		

C. Convolutional Neural Network

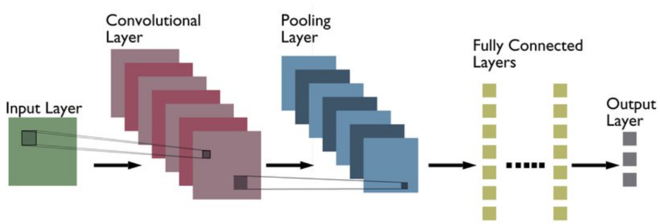

Fig. 6. CNN Architecture

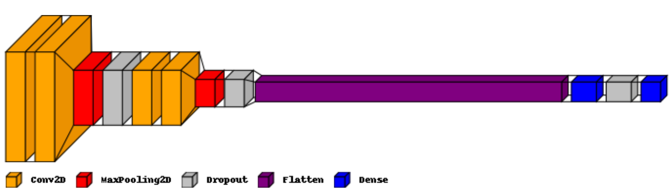

Fig. 7. Visual representation of proposed CNN architecture

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an image as input, assign importance (learnable weights and biases) to various aspects/objects in the image, and distinguish between them. ConvNet requires significantly less pre-processing than other classification techniques. ConvNets are capable of learning these filters/characteristics with sufficient training, whereas filters in basic approaches are hand-crafted. ConvNet architecture is comparable to the connectivity pattern of neurons in the human brain and was inspired by the structure of the visual cortex. The Receptive Field is a confined portion of the visual field where individual neurons respond to inputs. These fields overlap to cover the whole visual field. We began with photos with a resolution of 28 by 28 pixels. Four (3,3) conv2D layers were utilized. We use a MaxPooling layer of (2,2) to reduce computing power.

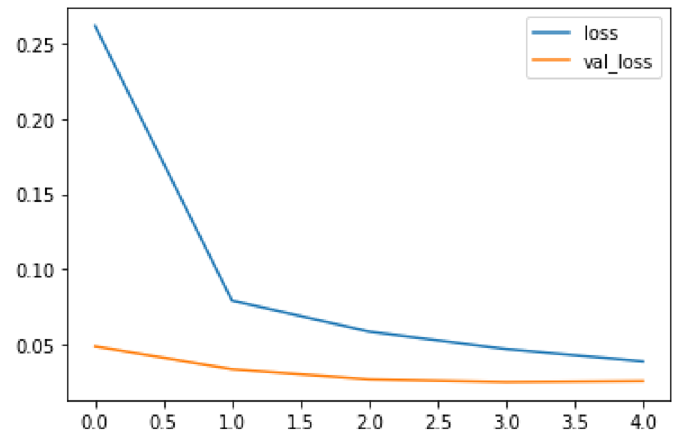

Fig. 8. Loss curve

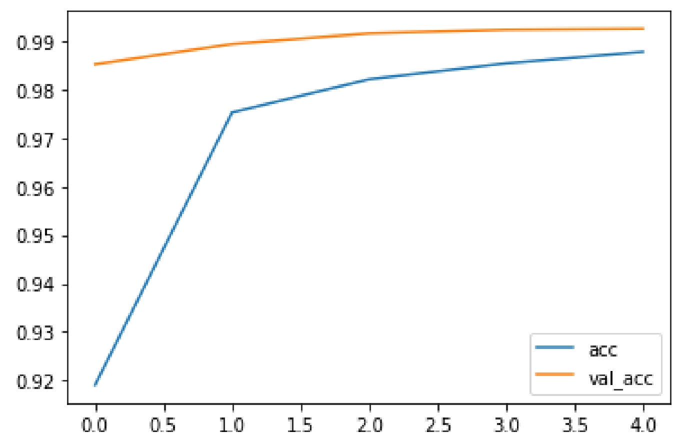

Fig. 9. Accuracy curve

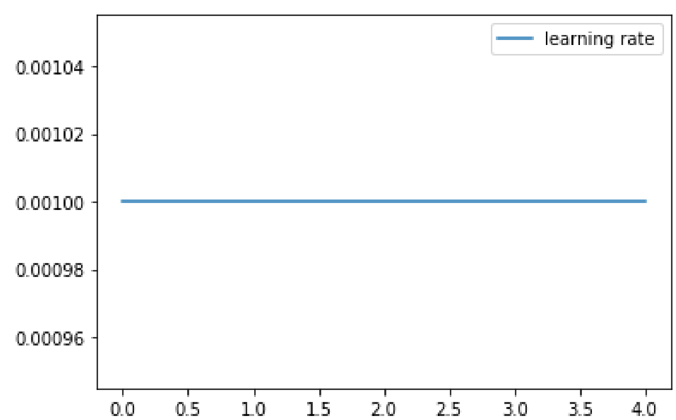

Fig. 10. Learning rate

TABLE II
CNN PARAMETER SIZE

CNN		
Layer (type)	Output Shape	Param
conv2d_15 (Conv2D)	(None, 28, 28, 32)	320
conv2d_16 (Conv2D)	(None, 28, 28, 32)	9248
max_pooling2d_6 (MaxPooling 2D)	(None, 14, 14, 32)	0
dropout_16 (Dropout)	(None, 14, 14, 32)	0
conv2d_17 (Conv2D)	(None, 14, 14, 64)	18496
conv2d_18 (Conv2D)	(None, 14, 14, 64)	36928
max_pooling2d_7 (MaxPooling 2D)	(None, 7, 7, 64)	0
dropout_17 (Dropout)	(None, 7, 7, 64)	0
flatten_9 (Flatten)	(None, 3136)	0
dense_32 (Dense)	(None, 256)	803072
dropout_18 (Dropout)	(None, 256)	0
dense_33 (Dense)	(None, 10)	2570
Total params: 870,634		
Trainable params: 870,634		
Non-trainable params: 0		

IV. Evaluation Metrics

After any image classification task, the quality of the model is evaluated based on performance assessment metrics. For quantitative evaluation, well-known performance evaluation metrics such as Accuracy1, Recall2, Precision3, and F1-Score4 are used to determine the performance of the proposed method.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1, Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{4}$$

Accuracy is defined as the proportion of correct predictions made by a model relative to the total number of predictions made. Precision is used to calculate the fraction of accurate identifications. Precision is determined by dividing the number of true positive outcomes (TP) by the number of true positive outcomes plus anticipated positive outcomes (TP + FP). Recalls are used to determine what proportion of true positives are accurately discovered. Recall is computed by dividing the number of true positives (TP) by the entire amount of data (TP + FN). The crucial aspect is the ratio of TP to total data. In a similar manner, the F1 Score is a well-known metric for evaluating the efficacy of machine learning algorithms. It is computed using the arithmetic mean of precision and recall. It is between 0 and 1 in value. F1 scores show the number of instances correctly detected by the learning models.

V. Result and Analysis

We trained SGAN with less amount of data compared to ANN and CNN. Among all three models accuracy of CNN is the highest. From accuracy and loss curves we can conclude that CNN is a better fit also and it requires less epochs to reach the highest accuracy and other metrics.

TABLE III
S GAN EPOCH

Epoch	D loss supervised	D loss unsupervised	Gloss
100	0.0218	0.0405	0.212609
200	0.0159	0.0036	0.144474
300	0.0176	0.0170	2.709530
400	0.0042	0.0060	0.173067
500	0.0035	0.0015	0.178969

TABLE IV
SGAN ACCURAY RATE

Model	Testing Accuracy
SGAN	73.20%

VI. Dataset

There are 60,000 training examples and 10,000 test cases in the MNIST collection of handwritten digits. It is a piece of the overall NIST collection in its whole. Now all of the digits are the same size, and they are centered within the image that has been provided. Since it was made publically accessible for the first time in 1999, this standardized collection of handwritten images has been utilized for the purpose of evaluating categorization techniques. Even when more advanced machine learning techniques become more readily available, MNIST maintains its position as a dependable resource for both academic researchers and students.

VII. Conclusion and Future Work

In light of the growing interest in semi-supervised learning and the quick advancements in generative learning, a survey was done to examine recent research on the use of GANs for semi-supervised learning. Prior work was categorized according to the proposed advancement, the model design, and the training processes. In addition, each paper's methodology was reviewed before a quantitative analysis was made based on the experimental results produced by each of the works. Finally, a qualitative analysis of the various categories was conducted to better comprehend the advantages and disadvantages of the various approaches, after which a number of possible future research directions were identified to encourage advancements in the field of semi-supervised learning using generative adversarial networks.

References

[1] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, pp. 373–440, 2019.

TABLE V
ANN CNN ACCURAY RATE

Model	**Testing Accuracy**	**Recall**	**Precision**	**AUC**	**F1 Score**
ANN	98.30%	98.25	98.35	98.35	98.29
CNN	99.35%	99.29	99.98	99..98	99.63

[2] A. I. Károly, R. Fullér, and P. Galambos, “Unsupervised clustering for deep learning: A tutorial survey,” 2019.
[3] X. Yang, Z. Song, I. King, and Z. Xu, “A survey on deep semi-supervised learning,” 2021. [Online]. Available: https://arxiv.org/abs/2103.00550
[4] J. T. Springenberg, “Unsupervised and semi-supervised learning with categorical generative adversarial networks,” 2015. [Online]. Available: https://arxiv.org/abs/1511.06390
[5] X. Wei, B. Gong, Z. Liu, W. Lu, and L. Wang, “Improving the improved training of wasserstein gans: A consistency term and its dual effect,” 2018. [Online]. Available: https://arxiv.org/abs/1803.01541