Zewail City of Science, Technology, and Innovation
University of Science and Technology
Communications and Information Engineering

A Summarization and Criticization about the paper

UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS

Prepared By
Rahma Hassan

2021/2022

Date
24, February 2022

I. Problem Definition

CNN is used in supervised learning for computer vision tasks. GAN is used in unsupervised learning. In this paper, a kind of CNN called deep convolutional generative adversarial network (DCGAN) is introduced which has specific architecture and constraints and shows strong results in learning images representations. Unsupervised learning is highly needed in tasks related to computer vision field which data are unlabeled. Due to few research tries to clarify GAN and what it can do, authors of this paper tried to understand GANs architecture and demonstrate that GANs can learn representation for images' features to be used later in supervised learning tasks like classification. They tried to clarify the intermediate representation of multi-layer GANs. GAN is unstable to train; therefore, authors of this paper propose a more stable set of architectures to train GAN by making some modifications on it (they called these architectures DCGAN) and they prove that it enhances the ability of GAN to learn good images' representations for supervised learning tasks.

II. Importance of solving this problem

There is a need for learning representations of images which have no label for doing machine learning tasks on it. Unsupervised learning is already used in tasks related to computer vision field and some techniques achieved good performance. Computer vision field has great importance these days due its implication in different fields which deal with images or frames of videos. The importance of this paper is to make training GAN more stable and to show that the generator and discriminator of deep convolutional GAN can learn good hierarchical representation of images. Hence, they can be used in several supervised ML tasks. Therefore, solving this problem will be useful in problems related to computer vision field and in problems use images in AI field.

III. Proposed Approach
(Architecture guidelines for stable Deep Convolutional GANs)

There were different trails for using CNN to scale up GAN and enhance its training, but these trails were unsuccessful. After intensive research, the authors of this paper became able to reach set of architectures can be used to stably train GAN across some datasets. Therefore, they applied three changes to CNN architectures to get stable GAN. In addition, they use different activation functions in generators and discriminators from those used in original paper of GAN.
So, all these changes are clarified here:

-All pooling layers are replaced by convolutional layers with stride in generator and disseminators.

-Fully connection between layers is removed. The output of first input layer (fully connected) is reshaped to 4-dimensional tensor as a start for the following convolutional layers.

-Batch normalization is used because it makes the learning more stable. This applied to all layers except for generator output layer and discriminator input layer to avoid instability of the model and oscillation of samples.

-ReLU activation function is used in generator except for output layer which uses Tanh function because it is found that makes model saturate faster.

-LeakyReLU activation function is used in all layers of discriminator.

IV. Experiment

The DCGAN is trained on three datasets, Large-scale Scene Understanding (LSUN) Imagenet-1k and a newly assembled Faces dataset. No pre-processing was applied to training data. Mini-batch stochastic gradient descent (SGD) was applied in all models. The slope of leak was equal to 0.2 in LeakyReLU which used in all layers of discriminator. Authors used Adam optimizer with tuned hyperparameters although that momentum is used in previous versions of GAN to accelerate the model with keeping it optimized. They used learning rate equal to .0002 instead of suggested value which equals to .001. Moreover, they reduced momentum value from 0.9 to 0.5 to help in stabilization of training.

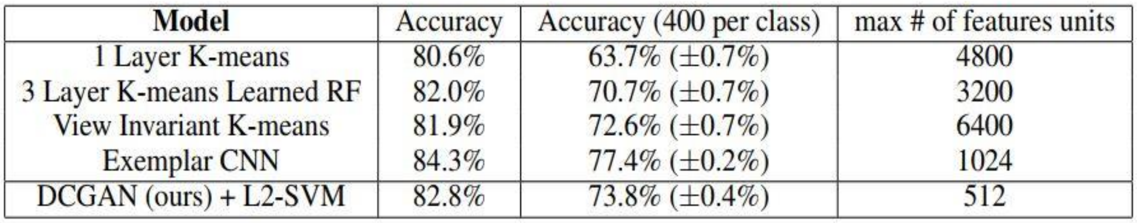
First, a model is trained 3 million images of LSUN bedroom dataset. Authors think that no overfitting was found because they used small learning rate and minibatch SGD. No data augmentation was applied to images. Deduplication was applied to reduce the probability of the generator of GAN memorizing the data and hence overfit. 275,000 duplicates were removed by this technique. Second, a model was applied on 3 million of images of people’s faces for 10K persons collected from searching on web by names. OpenCV library on python was used to detect faces on images so 350,000 face boxes were obtained and used in training. No data augmentation was applied in this case too. Third, a model is trained by natural images dataset, Imagenet-1k and, no data augmentation is applied.

V. analysis

The model is trained by experimenting it on supervised dataset, Imagenet-1k but the model is used to evaluate its performance in learning representation of CIFAR-10. Authors used their proposed model DCGAN on classifying CIFAR-10 and following it by regularized SVM (L2-SVM). They compared their results by results obtained by using different k-means based approaches. Their model gives better results which gives accuracy equals to 82.8% with 512 feature maps compared to k-means based approaches which gives lower accuracy and use much feature maps than proposed model. They did the same experiment on another dataset, on the Street View House Numbers dataset (SVHN).

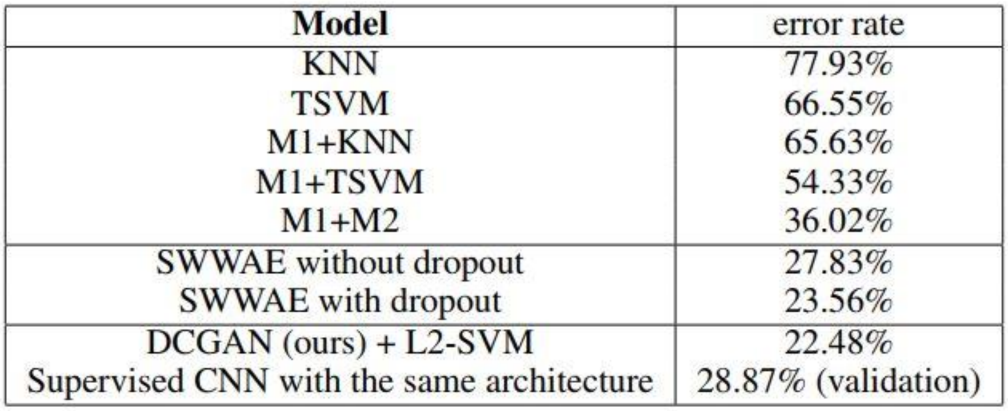
Results on both datasets, CIFAR-10 and SVHN were illustrated in the papers in the following tables:

Table 1: CIFAR-10 classification results using our pre-trained model. Our DCGAN is not pre-trained on CIFAR-10, but on Imagenet-1k, and the features are used to classify CIFAR-10 images.



Model	Accuracy	Accuracy (400 per class)	max # of features units
1 Layer K-means	80.6%	63.7% (±0.7%)	4800
3 Layer K-means Learned RF	82.0%	70.7% (±0.7%)	3200
View Invariant K-means	81.9%	72.6% (±0.7%)	6400
Exemplar CNN	84.3%	77.4% (±0.2%)	1024
DCGAN (ours) + L2-SVM	82.8%	73.8% (±0.4%)	512

Table 2: SVHN classification with 1000 labels



Model	error rate
KNN	77.93%
TSVM	66.55%
M1+KNN	65.63%
M1+TSVM	54.33%
M1+M2	36.02%
SWWAE without dropout	27.83%
SWWAE with dropout	23.56%
DCGAN (ours) + L2-SVM	22.48%
Supervised CNN with the same architecture	28.87% (validation)

Authors wanted to asse the model; so some things are done for this purpose like walking in the latent space. The model learn interesting representation if walking in latent space shows significant semantic changes among images. In addition, they demonstrated that a discriminator of unsupervised DCGAN trained on a large image dataset can also learn a hierarchy of important features. They visualized the features learnt by discriminator to check its ability to learn important parts in images of datasets. Moreover, they conducted some experiments to test what is the generator learns. They wanted to make the generator forget to draw windows in bedrooms and interestingly, the generated images showed that the model replaced windows by other objects.

VI. Criticize the Paper (Strengths and Weaknesses)

In this section, criticization of the paper will be illustrated from my point of view according to my understanding. Strengths will be shown first and then weaknesses.

Strengths:

Strength behind this paper is its contribution in proposing a version of GAN different than the original one to increase stability in training GAN. They tried in this paper to remove the gap between CNN used for supervised learning and unsupervised learning by making some modifications on architecture of GAN and introducing DCGAN. Also, strength is clear in the effect of this architecture (DCGAN) in learning hierarchical representation of images through layers which is important in images-based applications. In addition, this representation can be used in supervised learning tasks which clarify the importance of this paper in this field. Moreover, they conducted different changes in structure of original GAN like changing in used activation functions, removing any pooling, removing fully connected layers to provide a stable DCGAN. One of the strengths is experimenting the model on three different datasets which provide better evaluation of models. Also, they proved that no overfitting occurred in their testing of the models because of usage of minibatch SGD and small learning rate. When we look at results, we found that proposed model gives accuracy better than k-means based model on CIFAR-10 dataset with much fewer feature maps. It gives accuracy equals to 82.8% although the DCGAN model was never trained on CIFAR-10 which means that the model is strong in learning features. In addition, the model gives error rate equals to 22.48% on SVHN classification task with 1000 labels compared to higher error rate values using other ML models. In this paper, authors used different methods to asse their models and test performance and they avoided evaluating them by poor metrics. Last words on strengths of this paper is that it provided more stable set of architectures for training GAN and opened the door for further enhancement and exploration of GAN architecture and applications in different domains too.

Weaknesses:

Although the good results of the model and that it gives better accuracy on CIFAR-10 dataset than k-means based models, it couldn’t exceed the accuracy of exemplar CNNs. The model gives accuracy equals to 82.8% but exemplar CNN gives accuracy equals to 84.3%. In addition, from my point of view, although the organization of the paper is good, the language of the paper is difficult to be understood in some parts specially (INVESTIGATING AND VISUALIZING THE INTERNALS OF THE NETWORKS) part. Someone who has some knowledge in GAN but doesn’t aware about each detail of it might find some difficulty in understanding details of this paper. Last weaknesses is that there is still some instability in training models, and it was found when models were trained for long time, they sometimes collapse. Therefore, I see there are some weaknesses and further work is needed as suggested from the authors themselves too.

VII. Influence of the Paper

By the date of writing this report, this paper is cited more than 11482 times which shows its importance in enhancing training of GAN. This paper proposed a more stable way for training GAN on some datasets by making some modifications on CNN and original version of GAN. Although, some instability still exist when the model trained for long time, the results showed obtaining good representation of images. This representation is used in computer vision field tasks which are importantly needed these days.

VIII. Conclusion

Understanding GAN and how it works in intermediate layers was not clarified much in literature. In addition, it is known that obtaining stable GAN in training is a challenge and there were some unsuccessful trails for obtaining this before. Therefore, this paper tries to understand GAN and moreover, it makes some modification on CNN architectures and uses different activation functions in generator and discriminator of GAN to obtain stable trainable version of GAN. This version is called deep convolutional generative adversarial network (DCGAN) which gives stable training on some datasets.

The authors prove that DCGAN can learn good representation of images which can be used later for supervised ML learning tasks like classification. On the other hand, they finds that some instability is remaining when the model is trained for longer time. Therefore, further work can be done to solve this problem. In addition, this work can be extended to other domains like videos for frame prediction.