# Machine Learning Project: Generating Images Using Wasserstein-GAN (W-GAN) Model

December 23, 2023

## Contents

# 1 Motivation and Summary (quad chart)

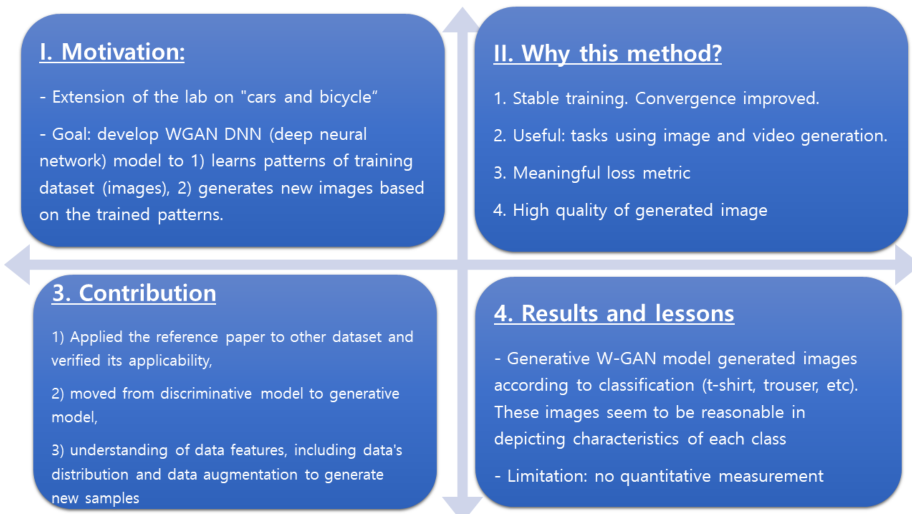


Figure 1: Quad chart

# 2 Dataset description

The data used for this project is "fashion-mnist" [1], which consists of 10 classes of clothes. The total number of training and test data is $60,000$ and $10,000$. The size of each image is $28 \times 28$. Since we do not need test images, we combine train and test data, and thus the total number of images is $70,000$.

With a generative neural network, we want to generate similar images corresponding to each class. For example, when we ask "bag" to the trained generative model, we want to observe only bag images from the output.

# 3 Generative Model

**Objective**: Serves as an executive summary. Be concise! You need to very briefly describe what is the motivation, why it should be solved by ML and the tool you picked, what is your contribution (how is it different the existing), results, summary/future directions/lessons learned.

**Answer**: In this project, we use conditional Wasserstein generative adversarial neural network with gradient penalty (WGAN with GP)[2] as an example of generative model.

## Part 1: Generative Adversarial Networks (GAN)

GAN is an algorithm used in unsupervised learning where two different networks (generator and discriminator) are trained simultaneously. The generator gets optimized to produce more realistic photos whereas the discriminator gets optimized in classifying fake and real images.

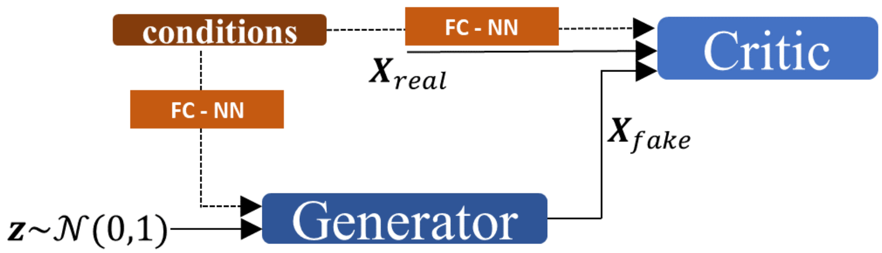


Figure 2: conditional W-GAN with gradient penalty

There are couple of potential problems with GAN; mainly the training stability and the quality of generated outputs. GAN can be difficult to train as a nice balance between the generator and discriminator is crucial. If either one becomes too powerful, the model can't be trained properly leading to unsatisfactory results. The quality of generated output is not as high as W-GAN and the biggest reason for this is because the loss function of GAN is not correlated with the quality of the output. Due to these issues based on our research, we have decided to use W-GAN instead.

## Part 2: W-GAN (Wasserstein Generative Adversarial Networks)

W-GAN is an algorithm made to improve the limitations shown in GAN related to training stability and the quality of the output as mentioned above. W-GAN

Table 1: Hyper-parameters to train WGAN-GP

| | Generator | Critic |
|---|---|---|
| Input image size | [28, 28 ,1] | |
| Learning rate | $5 \times 10^{-5}$ | |
| Optimizer | Adam ($\beta_1 = 0.5$, $\beta_2 = 0.9$) | |
| Epochs | 30 | |
| Batch size | 64 | |
| dimension of latent variable | 20 | |
| Number of parameters | 2,596,541 | 1,927,557 |

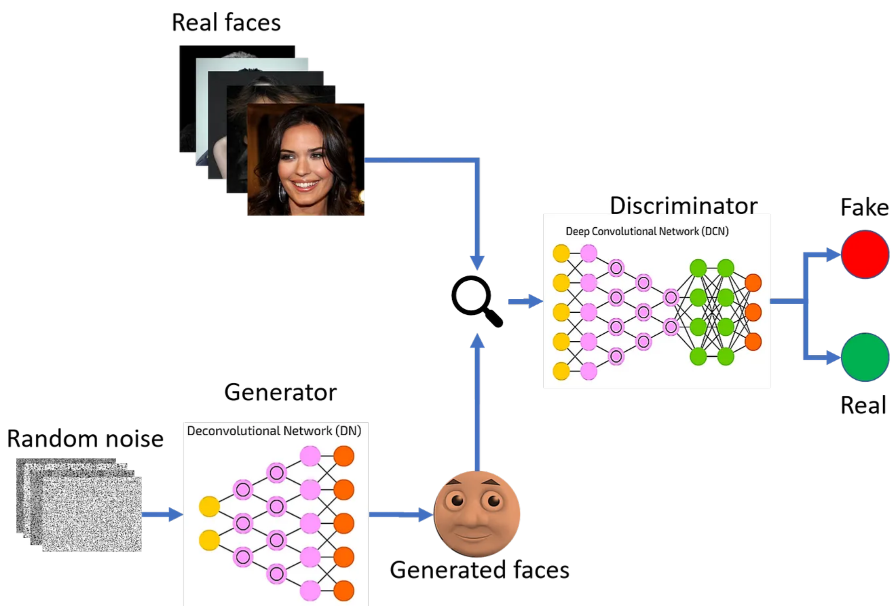


Figure 3: A pictorial explanation of a GAN model [3]

uses the Wasserstein loss function which measures the distance of the generated data and the real data resulting it more meaningful gradient for the generator.

Another difference between the W-GAN and GAN is that W-GAN uses a critic rather than discriminator. The critic estimates the Wasserstein loss which provides more useful gradient information to the generator unlike a discriminator which just classifies whether an input is fake or real. This is helpful especially in situations where the discriminator outperforms the generator too much. For these reasons, the training stability of W-GAN is better than GAN and also gives a better output. Hence, despite the fact that W-GAN is computationally more expensive, as it is able to address the issues mentioned above from GAN, we have decided to use W-GAN.

The losses of critic and generator are given as the followings:

$$
\begin{aligned}
L_{\text{critic}} &= \underset{\tilde{\boldsymbol{x}} \sim P_g}{E}[C(\tilde{\boldsymbol{x}}, \boldsymbol{u})] - \underset{\boldsymbol{x} \sim P_r}{E}[C(\boldsymbol{x}, \boldsymbol{u})] + \lambda \underset{\hat{\boldsymbol{x}} \sim P_{\hat{\boldsymbol{x}}}}{E}\left[(\|\nabla_{\hat{\boldsymbol{x}}} C(\hat{\boldsymbol{x}}, \boldsymbol{u})\|_2 - 1)^2\right] \\
L_{\text{generator}} &= -\underset{\boldsymbol{z} \sim \boldsymbol{N}(0,1)}{E}[C(G(\boldsymbol{z}, \boldsymbol{u}), \boldsymbol{u})]
\end{aligned}
$$

where $C$ is the critic model, $G$ is generator model, $\tilde{\boldsymbol{x}}$ is fake image sampled from the generator, $\boldsymbol{x}$ is real image, $\hat{x}$ is the interpolated image between the real and the fake, $\boldsymbol{u}$ is the conditional vector corresponding to each class label. The gradient penalty term is needed to avoid the *mode collapse*.

# 4 Our contributions, what we did differently

**Objective**: There can be overlap with quad chart. Explain what part of the project is different from your references, if you got any accuracy improvements, if you used new data or new method etc.

**Answer**:

### a) Applied the reference paper to other dataset and verified its applicability

- We have used new data for this project and verified the WGAN model's applicability. By applying the foundational WGAN framework to a distinct dataset, we have expanded the scope of the original research. This not only tests the model's robustness and versatility but also provides insights into its performance across different types of image data.

### b) Moved from discriminative model to generative model

- Transitioning from a discriminative model (DNN) to a generative model (WGAN) represents a paradigm shift in our approach to understanding and utilizing image data. While the discriminative model focused solely on classification tasks, identifying whether an image depicted a car or a bicycle, the generative model opens up a plethora of new possibilities.
- This move is not merely a change in technique but a fundamental shift in the model's objectives and capabilities. By generating new, realistic images, the WGAN model contributes to a broader understanding of the underlying data distribution and provides a tool for various applications ranging from data augmentation to artistic creation.
- This shift also implies a deeper engagement with the data's complexity, requiring a nuanced understanding of how to model and sample from the data distribution effectively.

### c) Understanding of data features, including data's distribution and data augmentation to generate new samples

- Generative models are inherently more intimate with the data they're trained on, as they must capture and reproduce the underlying distribution accurately.
- Through the process of training a WGAN, we've gained valuable insights into the specific characteristics that define each class (T-shirt, trouser, sandals, etc) within our datasets. This knowledge goes beyond simple classification and delves into the subtleties of shape, texture, and variation inherent in the images.

- Furthermore, our use of the WGAN model for data augmentation is a noteworthy contribution. By generating new, realistic samples of each class (T-shirt, trouser, sandals, etc), we've effectively expanded our dataset without the need for additional real-world data collection. This not only enhances the robustness and performance of any subsequent discriminative models trained on the augmented dataset but also provides a way to test and improve the resilience of these models against a wider variety of data.

### Summary of contribution

In summary, the move to a generative model and the subsequent understanding and application of data features represent significant strides in our project. These contributions are not just technical improvements but also provide deeper insights into the nature of the data we're working with and open up new avenues for application and research.

# 5 Result of generated images using W-GAN model

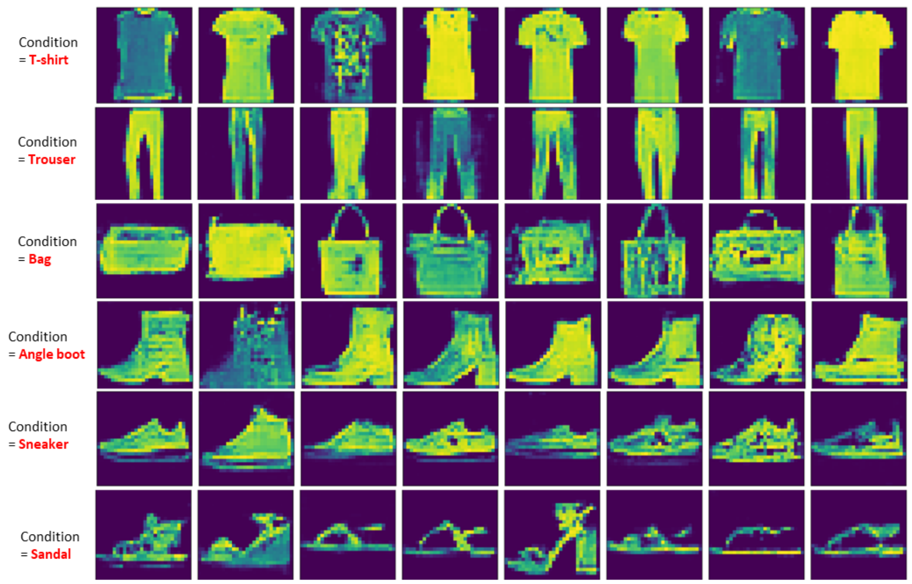


Figure 4: Result of generated images using W-GAN model

The results of generated images using our W-GAN model show promising outcomes. The model has successfully generated images representing various categories, including T-shirts, trousers, bags, ankle boots, sneakers, and sandals. Each category exhibits distinct and recognizable features, indicating the model's ability to capture and reproduce the key characteristics of different classes.

## Analysis of Generated Images

- *Qualitative Analysis*: Upon visual inspection, the generated images display a diversity in style, shape, and form, adhering to the expected variations within each category. For example, the sneakers exhibit different designs and contours, while the T-shirts vary in sleeve length and fit.

- *Comparison with Training Data*: By comparing the generated images to the original training dataset, we observed that while the W-GAN model has learned to mimic the general appearance and structure, there are subtle nuances that it has yet to capture fully. This comparison helps in identifying areas where the model might require further fine-tuning or more diverse training data.

## Limitation of the results

While the visual results are encouraging, there are inherent limitations in evaluating generative models like W-GAN:

- **Subjectivity in Quality Assessment**: The quality of generated images is often assessed subjectively. What might appear as a high-quality image to one observer might not be the case for another. This subjectivity makes it challenging to establish a consistent and objective measure of quality.
- **Quantitative Metrics**: The lack of robust quantitative metrics for evaluating the quality and diversity of generated images is a significant challenge. Traditional metrics used for discriminative models, such as accuracy or F1 score, are not applicable to generative models.
- **Mode Collapse**: Despite improvements with W-GAN, mode collapse, where the model generates a limited variety of outputs, can still occur. Detecting and mitigating mode collapse remains a challenge and an area for further research.
- **Training Stability and Convergence**: Generative models are notoriously difficult to train. Achieving a stable training process and determining convergence can be problematic, often requiring extensive experimentation with hyperparameters and model architecture.

## Future Work

To address these limitations and enhance the quality of generated images, future work might include:

- Implementing and comparing different GAN architectures and loss functions to improve image quality and model stability.
- Incorporating more diverse and larger datasets to enhance the model's ability to generate varied and high-quality images.
- Exploring advanced evaluation metrics that better capture the qualitative aspects of generated images and provide more objective quality assessments.
- Conducting user studies to understand better and quantify human perceptions of the generated image quality.

# 6 The pdf print of our codes and references

**Objective**: Make sure you put reference on top and highlight all code that you wrote yourself. If you don't highlight, I will assume all your code is copied.

**Answer**:
For implement ion, we have utilized some codes from this website [4]. We have highlighted codes that we have typed.

## References


[1] H. Xiao, K. Rasul, and R. Vollgraf. (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms.

[2] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[3] M. Bansa, "Generative adversarial networks: Unlocking the power of artificial creativity (beginner's guide)," November 2021. [Online]. Available: https://levelup.gitconnected.com/generative-adversarial-networks-unlocking-the-power-of-artificial-creativity-beginners-guide-7353a70a8709

[4] A. Persson, "Wgan implementation from scratch (with gradient penalty)," November 2020. [Online]. Available: https://www.youtube.com/watch?v=pG0QZ7OddX4ab$_channel = AladdinPersson$