Image Generation for Mercedes Benz



*By*

*Harsh Pokarna (ID No. 2020B3A70754P)*

Study Oriented Project

Under the guidance of Dr. [Poonam Goyal](mailto:poonam@pilani.bits-pilani.ac.in)



[BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI,](http://www.bits-pilani.ac.in/) PILANI CAMPUS

# **TABLE OF CONTENTS**

[**TABLE OF CONTENTS 2**](#_gj2jtqohyujy)

[**INTRODUCTION 3**](#_wv4b92d1j9ot)

[**DATASETS 5**](#_skps3xmneg70)

[Income Dataset (Kaggle) (Link): 5](#_hpzpqh3ddto9)

[MNIST Dataset (Link): 5](#_hrapa4ghwukk)

[Wheels Dataset (Kaggle) (Link): 6](#_ppz6x7ylfsbt)

[Structural Dataset (Kaggle) (Link): 6](#_87vxm7m7d3na)

[**METHODOLOGY 7**](#_6nyvnlgtjcg1)

[Conditional GANs: Experimentation and Insights 7](#_6twdtjgozaoq)

[Diffusion Models: 8](#_c1a291fcny5)

[TabDDPM: 9](#_4wuu2w6xn7p4)

[Baseline diffusion model with UNET: 9](#_kzwx3388drw1)

[Denoising Diffusion Probabilistic Model (DDPM): 11](#_g3aicyegaz6l)

[Noise Condition Score Networks(NCNS): 14](#_x9varp5xsfuh)

[**FUTURE WORK 17**](#_qh3ilm8rtiwu)

[**REFERENCES 18**](#_tz9iir59bojn)

# 

# 

# **INTRODUCTION**

This project explores the application of machine learning to automate and enhance the design of beam cross-sections for Mercedes-Benz vehicles. The primary goal is to develop models that can generate innovative and optimized beam designs, considering both material properties and structural performance requirements. Our key objectives also included automating and accelerating the generation of beam cross sections and optimizing the designs for specific metal properties and structural requirements.

In the initial phase of the project, we focused on utilizing numerical data to represent beam characteristics. This data included information such as material composition, dimensions, and desired mechanical properties. We employed two prominent generative model architectures:

1. Conditional Generative Adversarial Networks (cGANs): cGANs are powerful for generating new samples from a given distribution, conditioned on specific input parameters. In this context, we used cGANs to create novel beam cross-section designs based on the provided numerical attributes.
2. Tabular Data Diffusion Probabilistic Models (TabDDPM): TabDDPM extends the concept of diffusion models, originally designed for image generation, to tabular data. This approach allows for the gradual transformation of random noise into realistic beam cross-section designs.

Later, we transitioned to working directly with images of beam cross-sections.. Our exploration of image-based generation centered around several diffusion model architectures:

1. Baseline Stable Diffusion: We began with the foundational stable diffusion model proposed in the original research paper. This architecture, based on a UNET (a type of convolutional neural network), effectively learns the underlying distribution of beam cross-section images.
2. Modified DDPM: We experimented with a variation of the denoising diffusion probabilistic model (DDPM), tailored to the specific task of beam cross-section generation.
3. Noise Conditioned Score Networks (NCSN): NCSNs represent an alternative approach to diffusion modeling, offering potential advantages in terms of sample quality and efficiency.

Throughout this project, we have progressively refined our models and experimented with various architectures. We continue to investigate advanced techniques to further improve the quality, diversity, and structural validity of the generated beam cross-sections.

### 

# **DATASETS**

The selection and preparation of datasets are fundamental steps in the development and evaluation of machine learning models. In our experiments, we utilized a diverse range of datasets, each serving a distinct purpose and contributing to the overall understanding of model performance across various tasks. This section provides an overview of the datasets employed, highlighting their key characteristics, applications, and the challenges they presented during model training and evaluation.

## **Income Dataset (Kaggle)** ([Link](https://www.kaggle.com/datasets/mastmustu/income)):

* **Purpose:** The income dataset was employed exclusively in the Conditional Generative Adversarial Network (cGAN) model. The primary objective was to generate similar data.
* **Features:** This dataset encompasses various demographic and employment-related attributes of individuals, including age, education level, occupation, and capital gains/losses. These features collectively serve as input to the cGAN model for salary prediction.

## **MNIST Dataset** ([Link](https://www.tensorflow.org/datasets/catalog/mnist)):

* **Purpose:** The MNIST (Modified National Institute of Standards and Technology) dataset served as a benchmark dataset in our experiments. It comprises a large collection of handwritten digits (0-9) in grayscale format.
* **Model Performance:** The MNIST dataset, renowned for its simplicity and widespread use, consistently yielded excellent results across a variety of machine learning models. Its primary role was to evaluate the basic image generation capabilities of our models. All the models gave very good results with MNIST dataset as the dataset itself is not very complex for the neural network to understand.

## **Wheels Dataset (Kaggle)** ([Link](https://www.kaggle.com/datasets/wajidhassanmoosa/2d-wheel-design-data)):

* **Purpose:** The Wheels dataset, containing 4,000 images of wheel cross-sections, was central to our model evaluation process.
* **Dataset Characteristics:** It is crucial to note that the images within this dataset exhibit a mix of realistic and imaginary wheel designs. This diversity posed a unique challenge for the models, as they were required to generalize their understanding of wheel structures based on both factual and fictional representations.

### 

## **Structural Dataset (Kaggle)** ([Link](https://www.kaggle.com/datasets/kimalpha/car-wheel-design-in-engineering)):

* **Purpose:** This specialized dataset, tailored for machine learning applications, consisted of 4,000 images depicting various structures. Each image was 56x56 pixels in size.
* **Feature:** Uniquely, the file names of the images included the corresponding stiffness values of the structures. This allowed us to explore the relationship between visual characteristics of structures and their mechanical properties, opening avenues for potential applications in structural engineering and design.

In case of image datasets the size of images used was 32 x 32. I was getting an out of memory error for most of the models if I increased the size more. The batch size was 32 across all the models and the learning rate varied from model to model.

# **METHODOLOGY**

In the initial phase of the project, our focus was on leveraging the available numerical data to explore the potential of generative models for beam cross-section design. This data could be material properties, dimensional specifications, or target performance metrics.

We began our investigation by experimenting with two fundamental generative model architectures: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These models offered distinct approaches to learning the underlying distribution of beam cross-section data and generating novel designs.

## **Conditional GANs: Experimentation and Insights**

To further our understanding of GANs in the context of tabular data, we delved into the realm of Conditional GANs (cGANs). Conditional GANs are a type of generative model that extends the capabilities of traditional GANs by incorporating additional information during the generation process. This information, often in the form of labels, class information, or other attributes, serves as a condition for the generation of new samples. In the context of beam cross-section design, a cGAN could take as input not only the numerical parameters defining the beam but also a label indicating the desired strength. This conditioning enables the cGAN to generate cross-sections that are tailored to the specific material, resulting in designs that are both realistic and structurally sound.

Our experiments with cGAN’s were performed on the income dataset obtained from Kaggle. The aim was to generate similar data. Below are the results obtained from cGAN on income dataset:



## **Diffusion Models:**

Next we started experimenting with diffusion models as they offer several advantages over GAN’s and VAE’s as mentioned below:

1. Training Stability**:** Unlike GANs, which often suffer from training instability due to the adversarial nature of the optimization process, diffusion models are trained using a simple likelihood-based objective, making them more stable and easier to train.
2. Sample Quality: Diffusion models have demonstrated remarkable capabilities in generating high-quality samples, often surpassing the fidelity of those produced by GANs. This is partly due to the fact that diffusion models learn a more explicit representation of the data distribution.
3. Flexibility: Diffusion models can be easily adapted to different data types and tasks. They have been successfully applied to image generation, audio synthesis, and even molecular design.

At their core, diffusion models operate by gradually adding noise to data and then learning to reverse this process. This is analogous to the physical phenomenon of diffusion, where particles spread out over time due to random motion. Mathematically, this forward diffusion process can be described as:

**q(x\_t | x\_{t-1}) = N(x\_t; √(1 - β\_t) x\_{t-1}, β\_tI)**

where *x\_t* is the noisy sample at time *t*, *x\_{t-1}* is the previous sample, *β\_t* is a time-dependent noise coefficient, and *I* is the identity matrix. By repeatedly applying this process, we can transform any data distribution into a simple isotropic Gaussian distribution.

The key innovation of diffusion models lies in their ability to learn the reverse process, gradually denoising the Gaussian noise to recover the original data distribution. This is achieved by training a neural network to predict the noise component at each timestep:

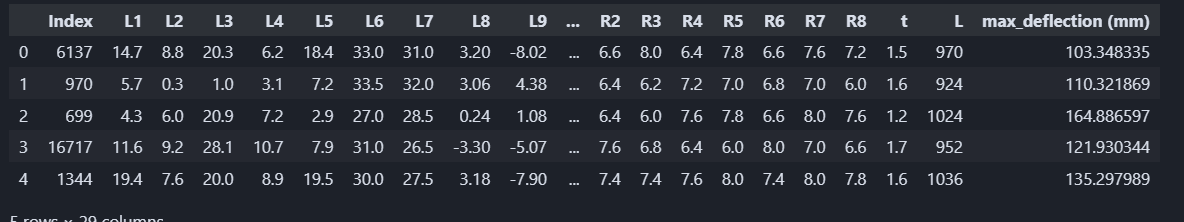
**p\_θ(x\_{t-1} | x\_t) = N(x\_{t-1}; µ\_θ(x\_t, t), Σ\_θ(x\_t, t))**

where *µ\_θ* and *Σ\_θ* are the predicted mean and covariance, parameterized by the neural network *θ*.

## **TabDDPM:**

Tabular Data Diffusion Probabilistic Model (TabDDPM) is a novel diffusion model specifically designed for generating numerical data. It was released in September 2023 and has shown promising results in various applications. The model adds noise to numerical data and aims to remove this noise in the reverse process. TabDDPM outperforms existing GAN and VAE based models for numerical data in terms of sample quality and training stability.

In our experiment, we found that TabDDPM, after making some modifications to the code, was able to generate numerical data. Below is the result obtained from TabDDPM when run on income dataset.



## **Baseline diffusion model with UNET:**

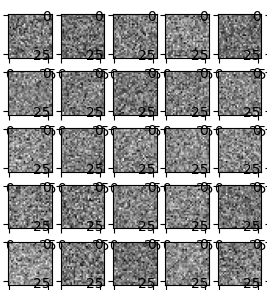
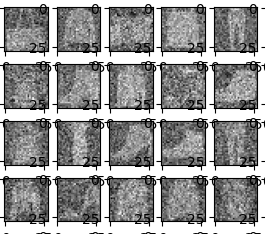
Following our investigation into numerical data generation, we shifted our focus to working directly with images of beam cross-sections. Stable diffusion, a powerful generative model framework, has demonstrated exceptional capabilities in image generation tasks. We began by experimenting with the baseline UNET architecture, which was introduced in the seminal DDPM paper.

The UNET architecture, named for its U-shaped structure, is a type of convolutional neural network specifically designed for image-to-image translation tasks. It consists of an encoder-decoder structure with skip connections that allow for the preservation of fine-grained details during the generation process.

The encoder part of the network progressively downsamples the input image, capturing hierarchical features at different scales. The decoder, on the other hand, upsamples the feature maps and reconstructs the output image. The skip connections bridge the corresponding layers in the encoder and decoder, facilitating the flow of information and enabling the network to generate high-quality images with sharp details.

We tested the baseline UNET model on the Fashion-MNIST dataset, a widely used benchmark for image classification and generation tasks. The model performed well on this dataset, generating realistic and diverse fashion item images.

Results of forwards process:



Final results:



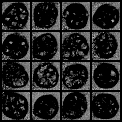
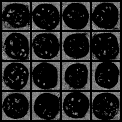
## **Denoising Diffusion Probabilistic Model (DDPM):**

Following our experimentation with the baseline UNET architecture, we turned our attention to a more refined diffusion model approach: Denoising Diffusion Probabilistic Models (DDPMs), as described in the paper "Denoising Diffusion Probabilistic Models" (Ho et al., 2020). This newer model held promise for improved image generation capabilities, particularly for complex datasets.

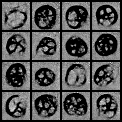
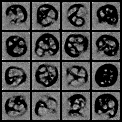
DDPMs operate by gradually adding noise to images over a series of steps, transforming them into pure Gaussian noise. The core of the DDPM lies in learning the reverse process, where a neural network is trained to progressively denoise the Gaussian noise back to the original image distribution.

While the DDPM model was theoretically promising, the initial implementation presented challenges. The available code required significant modifications and error fixes to function correctly. After addressing these issues, the model began to produce results on our beam cross-section dataset. However, the generated images, while showing some improvement over the baseline UNET, still lacked the desired level of detail and structural accuracy.

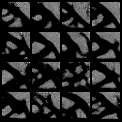
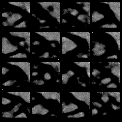
Initial outputs on wheels dataset:



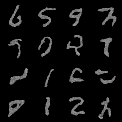
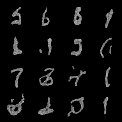
Final outputs of wheels dataset:



Output on structural dataset:



Output on MNIST dataset:



Optimization effort: We made a systematic effort to optimize the DDPM model. We experimented with various hyperparameters, including the noise schedule (β\_t), batch size, and learning rate. Additionally, we explored different neural network architectures for the denoising process. However, these modifications yielded only marginal improvements in the quality of the generated images.

Despite our best efforts, the DDPM model struggled to capture the intricate details and nuances inherent in beam cross-section images. This led us to the realization that further exploration of alternative architectures or training strategies might be necessary to achieve our goal of generating high-quality, realistic, and structurally sound beam cross-section designs.

## **Noise Condition Score Networks(NCNS):**

Following our experiments with DDPMs, we experiment with Noise Conditioned Score Networks (NCSNs), a cutting-edge approach to diffusion modeling that has shown great promise in image generation tasks.

NCSNs are a type of generative model that estimate the score function of a data distribution at various noise levels. The score function is the gradient of the log probability density function, and it provides information about the direction in which the probability density increases most rapidly. By learning the score function, NCSNs can generate new samples from the data distribution through a process known as Langevin dynamics.

The training process for NCSNs involves perturbing the training data with different levels of Gaussian noise and then training a neural network to estimate the score function for each perturbed distribution. This neural network, called the Noise Conditional Score Network, is conditioned on the noise level, allowing it to estimate the score function for a wide range of noise levels.

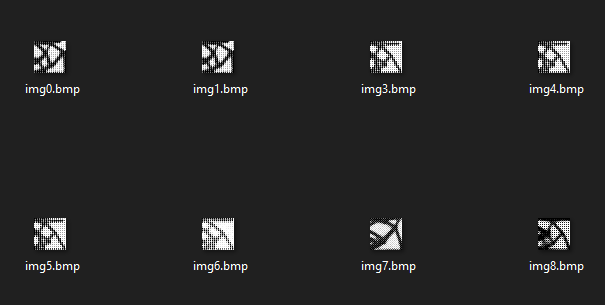
Our experiments with NCSNs on the beam cross-section dataset yielded very promising results. The generated images exhibited high quality and captured intricate details that were lacking in previous models. However, a persistent bug emerged where the model generated the same image for all samples within a single run. To obtain diverse outputs, the sampling code needed to be executed multiple times.

To further enhance the capabilities of NCSNs, we modified the model to incorporate labels corresponding to the images. This modification allows for conditional generation, where the model can generate images based on specific attributes or categories. For example, we could condition the model to generate beam cross-sections made of specific materials or designed for particular load-bearing requirements.

Outputs on wheels dataset:



Outputs on structural design dataset:



Next we also worked on a CNN model in order to estimate the labels of images generated from the above model. However as it is a CNN model it works for classification tasks rather than regression.

### 

# **FUTURE WORK**

This research represents a starting point for the integration of AI in structural engineering analysis. Several promising avenues for future work have been identified:

1. Image Generation with Labels: A natural extension would be to enable image generation conditioned on specific labels (e.g., beam strength). This would allow to visualize potential structural designs based on desired parameters. Techniques like Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs) could be explored to achieve this functionality.
2. Multimodal Input**:** Expanding the model's capabilities to accept both image and numerical data would provide a more holistic approach to image generation. This would allow for the integration of various inputs, such as material properties, load conditions, and deflection, into the prediction process. The model could then learn to correlate visual features with these quantitative parameters, potentially leading to more accurate and nuanced predictions.
3. Fixing the bug: The bug mentioned in the previous section should be removed through investigation into the model's architecture.

# **REFERENCES**

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015, November 18). Deep unsupervised learning using nonequilibrium thermodynamics [arXiv:1503.03585]. arXiv.org.<https://doi.org/10.48550/arXiv.1503.03585>

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. arXiv:2006.11239 [cs.LG].<https://doi.org/10.48550/arXiv.2006.11239>

Song, Y., & Ermon, S. (2020). Generative Modeling by Estimating Gradients of the Data Distribution [arXiv:1907.05600]. arXiv preprint arXiv:1907.05600v3.

Dhariwal, Prafulla, and Alex Nichol. "Diffusion Models Beat GANs on Image Synthesis." arXiv preprint arXiv:2105.05233 (2021).

Ho, J., Jain, A., & Abbeel, P. (2020, December 16). Denoising Diffusion Probabilistic Models [Version 2]. arXiv preprint arXiv:2006.11239.<https://doi.org/10.48550/arXiv.2006.11239>

Akim Kotelnikov, Dmitry Baranchuk, Ivan Rubachev, Artem Babenko. TabDDPM: Modelling Tabular Data with Diffusion Models [2209.15421], arXiv:2209.15421. (2022)