Universidad ORT Uruguay
Facultad de Ingeniería

Assessment
of
Data Augmentation Techniques
with
Synthetic Images
in
Uncommon Datasets Cases

Entregado como requisito para la obtención del título de
Máster en Big Data

A. Mauricio Repetto - 143045
Federico de León - 252047

Tutor: Franz Mayr

2023

Declaración de Autoría

Nosotros, A. Mauricio Repetto y Federico de León, declaramos que el trabajo que se presenta en esa obra es de nuestra propia mano. Podemos asegurar que:

- La obra fue producida en su totalidad mientras realizábamos la Tesis de Maestria en Big Data;

- Cuando hemos consultado el trabajo publicado por otros, lo hemos atribuido con claridad;

- Cuando hemos citado obras de otros, hemos indicado las fuentes. Con excepción de estas citas, la obra es enteramente nuestra;

- En la obra, hemos acusado recibo de las ayudas recibidas;

- Cuando la obra se basa en trabajo realizado conjuntamente con otros, hemos explicado claramente qué fue contribuido por otros, y qué fue contribuido por nosotros;

- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.

A. Mauricio Repetto
11-09-2023

Federico de León
11-09-2023

Abstract en Español

En una era donde la escasez de datos de entrenamiento, especialmente en ciertos dominios, plantea desafíos para los modelos de *Machine Learning* (*ML*), los datos sintéticos ofrecen una solución convincente. Aunque este problema es particularmente visible en enfermedades raras, de ninguna manera es exclusivo de ellas. Los datos reales, como reflejo de la realidad, no abarcan todas las posibles condiciones o eventualidades. Al complementarlo con datos sintéticos, podemos tener en cuenta casos extremos y condiciones no vistas, permitiendo aplicaciones de *ML* donde la escasez de datos podría haber dado lugar a modelos inutilizables debido a sesgos o a escenarios raros o sin precedentes.

Para este estudio, nuestro objetivo es centrarnos en el área de la *Computer Vision* (*CV*), y por lo tanto, nos enfocamos en el uso de imágenes y en problemas de clasificación relacionados con ellas. Proponemos la generación de datos sintéticos a través de técnicas modernas generativas de *Inteligencia Artificial* (*IA*) para imágenes, como los modelos de difusión (cuya notoriedad ha crecido significativamente recientemente), como una alternativa a las técnicas convencionales de *Data Augmentation*.

Corroboramos que los modelos entrenados con una combinación de datos reales y sintéticos pueden superar a los entrenados sólo con datos reales. Ésta mejora, sin embargo, mostró variaciones significativas dependiendo del conjunto de datos y de la técnica generativa empleada. En un dataset particular, caracterizado por su simplicidad y uniformidad, el rendimiento demostró ser bueno. Por otro lado, en conjuntos de datos más variados, raros o especializados, los desafíos en la adaptación de los modelos generativos resaltaron la necesidad de un análisis cuidadoso y, posiblemente, de ajustes finos. Además, el conocimiento previo de los modelos generativos en los conceptos utilizados en el *fine-tuning* resultó ser crucial para obtener imágenes sintéticas de calidad. Esto subraya la importancia de entrenar los modelos generativos en una amplia variedad de datos para que puedan reproducir los conceptos necesarios de manera efectiva. En resumen, los datos sintéticos sirven en algunos casos como una herramienta valiosa para mejorar la eficiencia de los modelos de *ML* en tareas de *CV*, particularmente con conjuntos de datos complejos debido a las características específicas dentro de sus clases o el desbalanceo de las mismas.

Abstract in English

In an era where training data scarcity, especially in certain domains, poses challenges for *Machine Learning* (*ML*) models, synthetic data offers a compelling solution. While this problem is particularly visible in rare diseases, it's by no means exclusive to them. Real data, as a reflection of reality, does not encompass every possible condition or eventuality. By supplementing it with synthetic data, we can account for edge cases and unseen conditions, allowing *ML* applications where data scarcity might have otherwise led to unusable models due to bias or rare or unprecedented scenarios.

For this study, we aim to focus on the area of *Computer Vision* (CV), and therefore, we focus on the use of images and classification problems related to them. We propose the generation of synthetic data through modern generative *Artificial Intelligence* (AI) techniques for images, such as diffusion models (whose notoriety has grown significantly recently), as an alternative to conventional *Data Augmentation* techniques.

We corroborated that models trained with a combination of real and synthetic data can outperform those trained solely with real data. However, this improvement showed significant variations depending on the dataset and the generative technique employed. In one particular dataset, characterized by its simplicity and uniformity, the performance proved to be good. On the other hand, with more varied, rare, or specialized datasets, the challenges in adapting generative models highlighted the need for careful analysis and possibly fine-tuning. Furthermore, generative model's prior knowledge of the concepts used in fine-tuning proved to be crucial for obtaining quality synthetic images. This emphasizes the importance of training generative models on a wide variety of data to effectively reproduce the necessary concepts. In summary, synthetic data can often serve as a valuable tool for enhancing the efficiency of ML models in CV tasks, particularly with complex datasets due to the distinct characteristics within their classes or the presence of class imbalance.

Palabras clave/Key words

Artificial Inteligence; Machine Learning; Deep Learning; Supervised Learning; Generative Model; Diffusion Model; Generative Adversarial Network; Stable Diffusion; Data Augmentation; DreamBooth; Synthetic Data; LoRa

Table of Contents

List of Tables

List of Figures

1 Introduction

Artificial Intelligence (AI) and deep learning model development rely heavily on the availability of large data sets. However, fields like medicine often face data scarcity due to the rarity of certain conditions or diseases, challenges in data collection, or exorbitant acquisition costs, resulting in the limited performance of AI models [1]. This research attempts to overcome these limitations by generating synthetic data via contemporary generative AI techniques such as diffusion models. These techniques offer an alternative to more traditional data augmentation procedures, which have been extensively discussed in various studies, one example of this is Mingle Xu et al. [2], typically involving data transformations such as rotation, zooming, cropping, among others.

Synthetic data, as opposed to real-world data, is artificially generated, mimicking real-world data's statistical properties and correlations. While the best insights come from real data, it's often expensive, imbalanced, unavailable, or unusable due to privacy regulations or other factors. Hence, synthetic data can be an effective complement or alternative and when merged with real data, it can create enhanced datasets that address the gaps and weaknesses often found in real data and improve AI models by increasing robustness and generalizability.

In addition to our work, there has been a growing interest in the use of synthetic data for ML, Gartner in 2021 predicted that "*by 2030, most of the data used in AI will be artificially generated by rules, statistical models, simulations or other techniques*" [3] and later the same year it made another prediction saying "*by 2024, 60% of the data used for the development of AI and analytics projects will be synthetically generated*" [4]. These predictions suggest that synthetic data is a rapidly growing field with the potential to revolutionize the way we develop and train ML models, showing the importance to make advances in the subject. Proof of this trend is evident in the growing emergence of platforms dedicated solely to the process of generating synthetic data, such as *Gretel* [5] and more recently *MONAI* [6].

In collaboration with Cogniflow [7], a No-Code AI platform with AutoML capabilities, this project pursues enhancing the performance of various image classification models by applying generative techniques. Through research and analysis of the state of the art (SOTA) in synthetic data generation, we developed and tested different models to address problems where the lack of training data or the imbalance between classes negatively affects the performance of deep learning models.

This project is carried out using public datasets, and although the available datasets focus on the health area, the results and applications of this study will not be limited

to this industry. The research's findings are available to the scientific community, and eventually, they can be integrated into the Cogniflow platform to enhance its ML service offerings.

The project methodology included the following steps:

- Studying the SOTA in synthetic data generation.
- Analyzing the available data and information.
- Defining evaluation metrics to compare results.
- Generating different models based on the studied techniques.
- Developed small proof of concepts (PoC) that demonstrates the complete training process for a given dataset.
- Presenting the conclusions and lessons learned throughout the project.

This document provides a detailed description of the project, its objectives, expected outcomes, and technological features, including the use of neural network models, GANs, diffusion models, cloud computing, and various related tools and technologies.

The project was expected to take 6 months to complete where the major milestones along the way included:

- **Month 1-2**: Complete the literature review, check available models and techniques, among others and develop a research plan.
- **Month 3-6**:
  - Collect and preprocess the data.
  - Implement and train the models.
- **Month 6**: Evaluate the models and present the findings.

The project faced a number of challenges, including:

- The complexity of the generative models.
- The diversity of data between datasets.
- The availability of existing generative models.
- The computing resources needed to train or finetune such models.
- The potential for bias in the generated data.

We managed to mitigate some of these challenges and eventually, we hope that the insights presented in this work contribute to the broader research community and future endeavors in synthetic data generation.

2 Related Work

2.1 Deep Learning and Supervised Learning

Deep Learning (DL) [8] is a subclass of ML focused on algorithms inspired by the structure and function of the brain, known as artificial neural networks (NNs). Supervised Learning is an approach where the model learns from examples in the training set to generalize to new, unseen data and it forms the backbone of many ML applications, including DL.

2.1.1 The Importance of Data Augmentation

Data is a critical resource in ML and particularly in DL. However, collecting a large and diverse dataset is often impractical due to time, scale, financial, and other constraints. This is where data augmentation comes into play. Data augmentation is the process of artificially increasing the size and diversity of a dataset, traditionally by applying various transformations like rotation, scaling, and flipping to the existing data or more complex techniques like the ones this work focuses on. Research from Luis Perez et al. [9] gives a good comparison among older and newer techniques, proposing a method based on NNs to choose the best one for a given classifier.

2.2 Generative Models

Generative models [10] are a class of statistical models that aim to learn the underlying data distribution of the training set so that new data points can be generated. They have been successfully employed in various applications, including but not limited to, image synthesis and text generation. In the current work, they will be utilized for data augmentation.

2.2.1 Generative Adversarial Networks

Generative Adversarial Networks (GANs) [11] are a specialized form of DL models designed for generating synthetic yet realistic data. A GAN comprises two neural networks:

the Generator (G) and the Discriminator (D). These networks are trained in tandem, engaging in a zero-sum game where the Generator aims to produce data indistinguishable from real data, and the Discriminator strives to distinguish between real and generated data.

Architecture: The Generator network takes a random noise vector z as input and produces synthetic data $G(z)$. The Discriminator network takes both real data x and generated data $G(z)$ as input and outputs a probability score indicating the likelihood that the data is real. A schematic representation of this architecture can be seen in Figure 2.1.

Loss Function: The training process involves optimizing specific loss functions for both networks. The Discriminator aims to maximize the following loss function:

$$\text{Discriminator Loss} = -\log(D(x)) - \log(1 - D(G(z)))$$

Conversely, the Generator tries to minimize:

$$\text{Generator Loss} = -\log(D(G(z)))$$

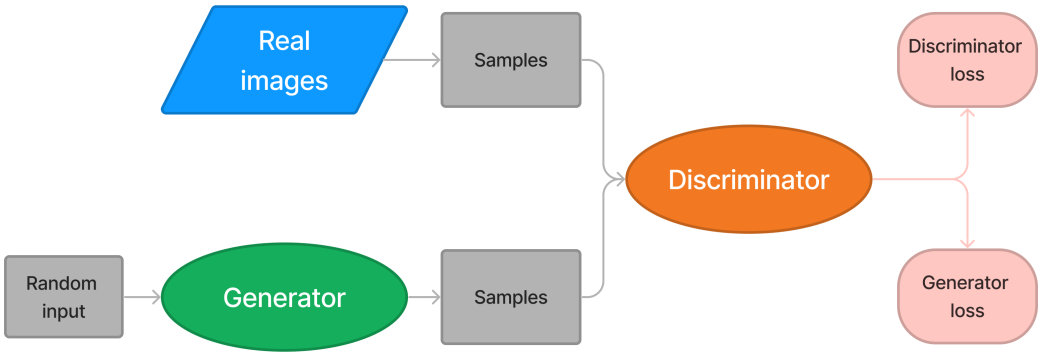


Figure 2.1: Schematic representation of the GANs training process

Applications in Data Augmentation: GANs have proven effective in enhancing dataset diversity and improving model performance across various tasks [11, 12]. For instance, Sundaram and Hulkund (2021) utilized GANs for data augmentation on the CheXpert chest X-ray dataset. Their study revealed that GAN-based augmentation significantly improved performance in underrepresented classes, particularly when data is scarce, thereby highlighting the promise of GANs in scenarios where data collection is expensive or impractical [13].

2.2.2 Diffusion Models

Diffusion models (DMs) [14] are deep generative models that work by adding noise (Gaussian noise) to the available training data (the forward diffusion process) and then reversing the process (the denoising or the reverse diffusion process) to recover the data.

Training Process: During the forward diffusion process, Gaussian noise is added to the image, corrupting the data gradually. The reverse diffusion process aims to recover the original data from the noise. This is achieved by minimizing a specific loss function that measures the difference between the original and the recovered data. A visual diagram of this process is illustrated in Figure 2.2

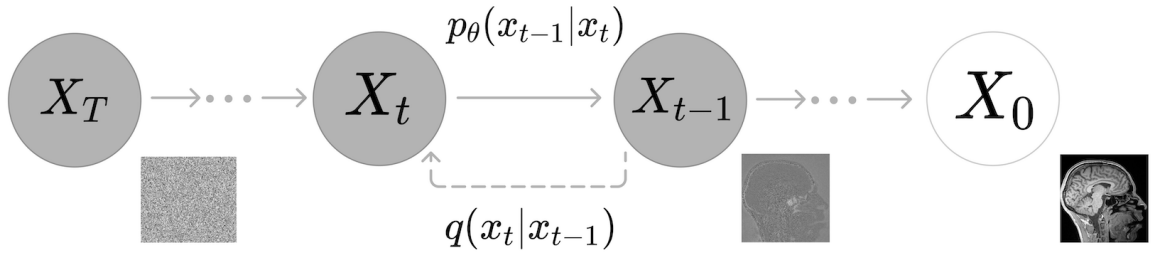


Figure 2.2: Schematic representation of the DMs training process.

DMs have been successfully applied to a variety of image generation tasks, including processing, colorization, super-resolution, inpainting, and semantic editing (Ruifei et al. [15], Trabucco et al. [16], Azizi et al. [17]).

In particular, diffusion models have been used to generate large-scale images from text descriptions. A few notable examples, some of them with different versions, include:

- *Stable Diffusion* [18].
- *DALL-E 2* [19].
- *Imagen* [20].
- *eDiff* [21].
- *GLIDE* [22].
- *Midjourney* [23].
- CM3leon [24]

These models have produced incredible and high-resolution synthetic images.

Recently, DMs have been used to augment training data, for instance, synthetic data generated with GLIDE has been shown to improve the performance of zero-shot and few-shot image classification [15]. Other studies have explored strategies for augmenting

individual images using pretrained DMs, with promising results in few-shot settings [16]. During the course of our work, Azizi et al. [17] demonstrated that the *Imagen* model can be fine-tuned for ImageNet [25] to produce class-conditional models, achieving SOTA results.

However, images and domain-specific language in the medical field differ from natural images and texts, which can make it challenging for text-to-image models to generalize well into this domain. In the same direction, Chambon et al. [26] adapted a pretrained DM to a corpus of chest X-rays and their respective radiology reports. The resulting RoentGen model is proficient at generating high-fidelity, diverse synthetic medical images that are controllable via text prompts. In a similar vein to our work, RoentGen model has also been employed for data augmentation, leading to enhanced classifier performance. Nevertheless, the computational overhead for these advancements is substantial, RoentGen required the use of 64 A100 GPUs. This makes the approach less accessible and commercially viable for broader applications. To quantify this, we consider the cost of running such experiments on Google Cloud [27] , known for its cost-effective A100 GPU rates at $1.25 per hour:

- Fine-tuning a model for 1,000 training steps incurs an approximate cost of $26.67.
- Extending the training to 12,500 steps raises the cost to around $400.
- A full-scale training run of 60,000 steps would demand about $1,920.

These estimates solely cover the GPU usage and do not account for other associated costs like storage, data transfer, or additional cloud services. Therefore, the actual expenditure could be significantly higher, further accentuating the computational and financial barriers for the adoption of such sophisticated models.

3 Goals

3.1 Main goal

The main objective of the project is to evaluate impact and improve models performance when trained using augmented datasets with synthetic image generation techniques.

To achieve this goal, the project aimed to develop a pipeline shown in Figure 3.1 capable of generating synthetic data using generative models, which could be generalized and enable further enhancements to existing classification models. Through the collaboration with Cogniflow, our research gained valuable resources and guidance. We were granted unlimited access to Cogniflow's platform, simplifying the development and management of our classification models and facilitating the seamless execution, comparison, and tracking of experiments. Additionally, we held two consultative meetings with the Cogniflow team, which provided specific insights for transitioning our development to a production environment should the outcomes be promising. The initial project idea was also conceived jointly with Cogniflow, aligning our research goals from the onset.

As part of our mutual objective with Cogniflow, we are dedicated to formulating a solution that demonstrates efficiency, scalability, and minimal computational resource demands. Scalability is of particular significance, especially for small businesses, as it empowers them to implement better classification models in a commercial context without the need for substantial infrastructure investments or extensive resource allocation. This scalability should be adaptable to varying data input sizes and computational requirements, not only ensuring versatility but also commercial viability and facilitating the seamless integration of these models, even for enterprises operating within resource constraints.

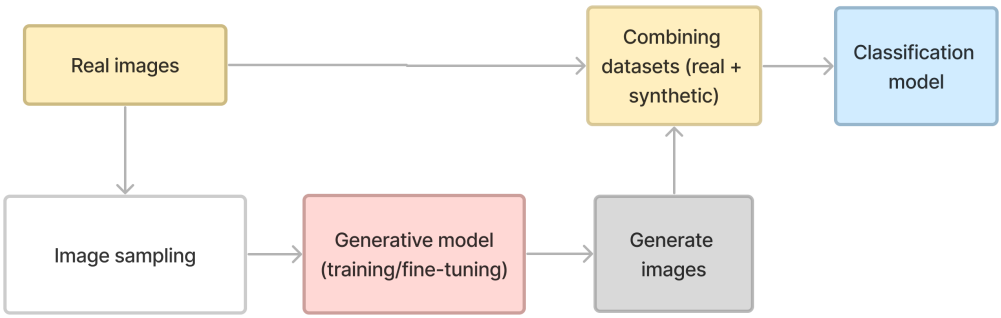


Figure 3.1: Depicts the major steps of the designed pipeline illustrating the project goal.

The introduced pipeline incorporates a parallel path alongside the traditional training steps for a classification model. This includes image sampling, generative model training, image generation, and combination steps, each posing unique challenges and objectives.

3.2 Secondary goal: Contributions

In addition to the mentioned principal objective, we established a distinct goal to enrich public repositories and contribute valuable resources to the broader community. This endeavor not only serves to assist future research in this domain but also extends its potential benefits to other related subjects.

4 Datasets

For our research into enhancing image classification through synthetic data augmentation, we utilized four datasets, primarily comprising chest X-ray and other medical images with diverse composition as summarized in Table 4.1.

Property	**Chongqing**	**NIH**	**Human Brain MRI**	**Diabetic Retinopathy**
Train Size	566	108,948	5,712	3,652
Test Size	140	NA	1,311	719
# Classes	2	8	4	5
Channels	1	1	3	3
Resized to	512x512	512x512	300x300	300x300

Table 4.1: Table summarizing overall datasets characteristics.

4.1 Chongqing Pneumoconiosis Detection

The *Chongqing Pneumoconiosis Detection Dataset* [28] is designed to facilitate the detection of pneumoconiosis, an occupational lung disease caused by inhaling mineral dust. The dataset includes 706 images, 142 of which are positive cases of pneumoconiosis. The goal of this dataset is to enhance detection and early diagnosis of pneumoconiosis.

Examples of these images are presented in Figure 4.1.

(a) Image 1 (b) Image 2 (c) Image 3

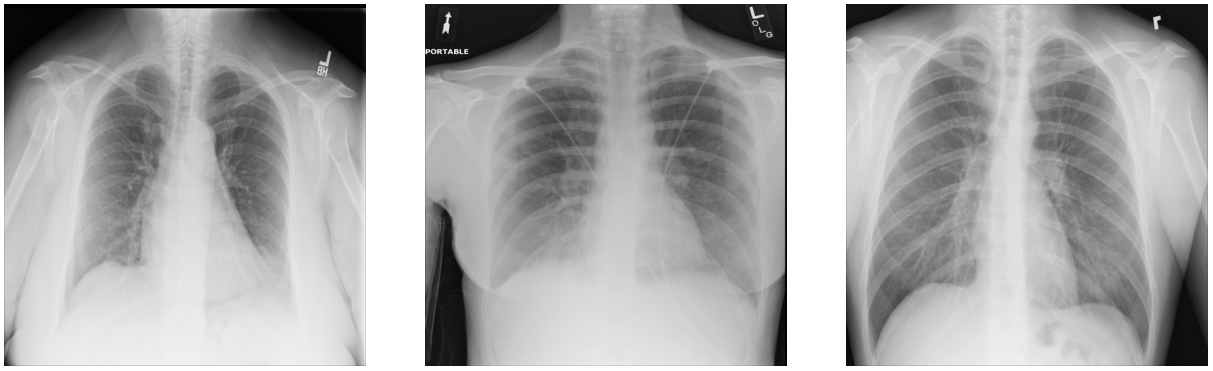
Figure 4.1: Sample images from the Chongqing Pneumoconiosis Detection Dataset.

4.1.1 NIH Chest X-rays

The *NIH Chest X-rays Dataset* [29] comprises 108,948 chest X-ray images with corresponding disease labels for 32,717 patients. The aim of this dataset is to aid in detection and computer-aided diagnosis of diseases seen in chest X-rays. It provides a vast resource for training and validation of deep learning models in the medical field.

Each dataset, NIH and Chongqing, served a unique purpose in our research. Both were used for pretraining models and generating new images. Specifically, we utilized the Chongqing dataset to train and validate our baseline model for the detection and classification of pneumoconiosis. We aim to use NIH dataset in pneumoconiosis detection to compensate the limitations of the Chongqing dataset. This additional dataset contributed to a more comprehensive range of images for training, allowing the model to learn from various pneumoconiosis manifestations and improving its detection capabilities.

You can see some of these images at Figure 4.2



(a) Image 1 (b) Image 2 (c) Image 3

Figure 4.2: Sample images from the NIH Chest X-rays.

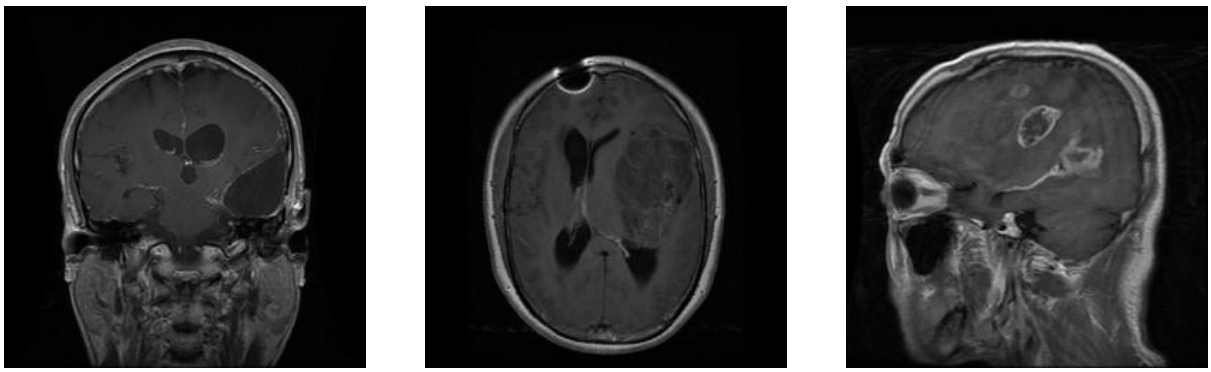
4.2 Human Brain MRI

The *Human Brain MRI Dataset* [30] is crucial for the detection and classification of brain tumors. It merges three separate datasets—figshare, SARTAJ, and Br35H—resulting in 7,023 human brain MRI images. These images are classified into four categories:

- Glioma
- Meningioma
- No Tumor
- Pituitary

Because the original images varied in size, we performed a pre-processing phase to resize all images to a uniform size of 300x300 pixels. Additionally, to manage the dataset size, we randomly reduced the image count in each class by 80 percent. This smaller dataset was utilized for effective data augmentation during model training.

Examples of these images are presented in Figure 4.3.



(a) Image 1 (b) Image 2 (c) Image 3

Figure 4.3: Sample images from the Human Brain MRI Dataset.

4.3 Diabetic Retinopathy Gaussian Filtered

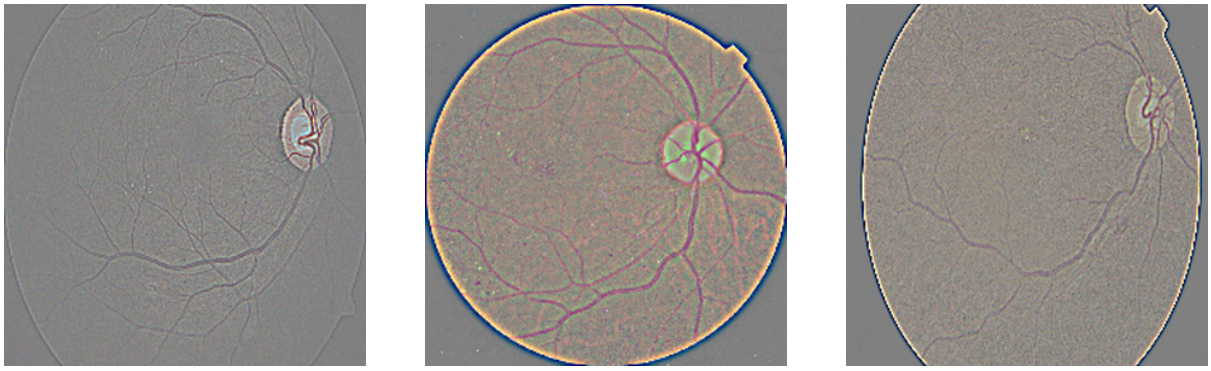
Diabetic Retinopathy is a retinal disease caused by diabetes and can lead to blindness. We utilized a dataset originated from the APTOS 2019 Blindness Detection dataset [31], which contains Gaussian filtered retina scan images optimized for diabetic retinopathy detection. These images have been resized to 224x224 pixels to ensure compatibility with various pre-trained deep learning models.

The images in this dataset are organized according to the severity or stage of diabetic retinopathy:

- No DR
- Mild
- Moderate
- Severe
- Proliferate DR

To match our model's input requirements and facilitate a balanced evaluation, we further processed this dataset by resizing the images to 300x300 pixels and splitting it into training and test sets, using an 80-20 split.

Examples of these images are presented in Figure 4.4.



(a) Image 1 (b) Image 2 (c) Image 3

Figure 4.4: Sample images from the Diabetic Retinopathy Gaussian Filtered Dataset.

5 Pipeline implementation

In this chapter are described some of the implementation decisions taken for the different pipeline phases depicted in Figure 3.1.

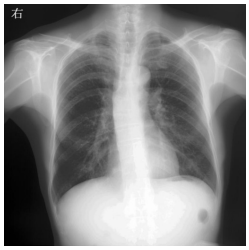
As a brief summary of key aspects of each phase:

- **Image sampling**: refers to the process of selecting specific images from the larger original dataset. It directly impacts subsequent steps by capturing representative samples that approximate the data distribution, requiring considerations of factors like image size, diversity, and adaptability for future datasets.
- **Generative model**: refers to the process of first, evaluating generative models by collecting and analyzing information to determine their suitability and second, implementing the chosen models using diverse approaches to train/fine-tune them.
- **Generate images**: involves ensuring that generated images meet quality and fidelity standards. Also that they are compatible with real-world data, minimizing the domain gap between synthetic and real data for optimal model performance in practical applications.
- **Combining datasets**: this is about how to balance real and synthetic data to train a classification model, as an excessive reliance on synthetic data can introduce biases or inconsistencies that affect the model's ability to generalize effectively, necessitating careful consideration to strike the optimal balance.
- **Classification model**: implies the training of the classification model based on the steps and decisions previously taken.

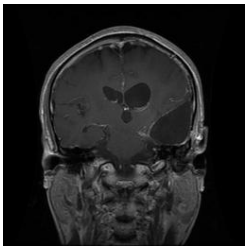
5.1 Image sampling

Image sampling is a critical component for securing high-quality training data, particularly in pipelines where the result of each step heavily influences subsequent ones. The effective execution of this process is essential for capturing representative samples that embody the critical features necessary to closely approximate the underlying data distribution. During this phase, considerations must include a range of factors such as image size, diversity, distribution, complexity, and resolution. The goal is to devise a methodology that is not only effective for the datasets at hand but also adaptable for future collections of data.

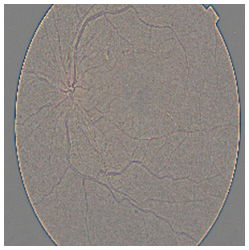
One of the primary challenges we encountered was the inherent diversity among images within each dataset. Our work with X-ray images presented limited variations, resulting in a high degree of similarity among samples. However, datasets like MRI (Figure 4.3) and Retinopathy (Figure 4.4) presented a broader spectrum of image differences, as detailed in Table 5.1.



(a) X-ray - Image 1

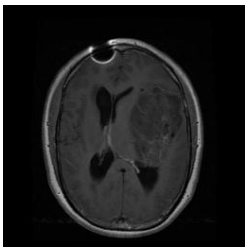


(b) MRI - Image 1

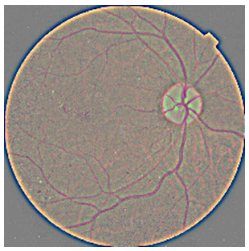


(c) Scan - Image 1

(d) X-ray - Image 2



(e) MRI - Image 2



(f) Scan - Image 2

Table 5.1: Image diversity among and within datasets.

To address the challenge posed by the diversity within each dataset, we devised a solution comprising the following steps:

- Initially, we propose to extract features from each image, with the purpose of transforming them into representative feature vectors or embeddings. To achieve this different approaches were explored:
  - Using a pretrained EfficientNet model [32].
  - Using the recently released CLIP model [33].

  While GeneCIS [34] was also under consideration, it was ultimately discarded due to challenges related to its implementation. These features served as a basis for subsequent clustering.

- Upon extraction of features from all images, we used the Faiss k-means clustering algorithm [35] to group images into clusters based on feature similarity. This process allowed us to group together images with similar content.

- For each cluster, we calculated the distances between the cluster centroid and the individual images within it. We then selected the top N images closest to the centroid, thereby ensuring their representativeness of the overall characteristics within the cluster.

- To evaluate the consistency and representativeness of the selected images, we assessed the average similarity and standard deviation of similarities within each cluster. Using a predetermined threshold, we prioritized clusters with higher average similarity and lower standard deviation. This method ensured that the chosen images were both indicative and consistent with their respective clusters.
- Finally, we employed Stable Diffusion techniques to generate synthetic images for each cluster separately, allowing each model to learn from better representations and generate synthetic images based on more uniform images.

Our proposed solution aimed to manage the diversity of images within each cluster. As a result, it facilitated the generation of higher-quality synthetic images through the selected models. We propose that the models by learning from a more homogeneous set of consistent images could potentially enhance both the quality and coherence of the generated images.

5.2 Generate model & Generate images

In the rapidly evolving landscape of generative models, two primary approaches have gained prominence: GANs and DMs. While GANs are known for their ability to produce high-quality images, they come with the drawbacks of computational intensity and lengthy training processes, making them less suitable for our commercialization objectives [13]. Conversely, DMs have shown remarkable advancements in both generation quality and efficiency, aligning more closely with our goals [17] [26].

Given this context, our research is directed toward meeting two broad sets of challenges. The first is the generative model assessment, which involves diligently collating, reading, analyzing, and assessing a wealth of information to identify which generative models are most appropriate for our purposes. This step also includes the time-consuming task of implementing these chosen models using various approaches.

The second challenge revolves around the quality and fidelity of generated images. It is crucial to ensure that the synthetic images produced meet established quality and fidelity benchmarks and are compatible with real-world data. Moreover, there is a significant challenge in bridging the domain gap between real and synthetic data to minimize any potential negative impacts on the model's performance in real-world applications.

While we have made considerable progress in both exploring the use of DMs and evaluating synthetic image quality, complete success remains an ongoing objective. In addition to using automated metrics for assessment, another possible approach is to involve human experts in the evaluation process to gain valuable insights into the quality of generated samples.

5.2.1 Progressive Growing of Generative Adversarial Networks

One of the generative techniques analyzed in this study is the Progressive Growing of Generative Adversarial Networks (PGGAN) by Karras et al. [36]. PGGANs are known for their ability to produce high-quality, high-resolution images. This type of GAN involves using a generator and discriminator model with the same general structure and starting with very small images, such as 4×4 pixels. During training, new blocks of convolutional layers are systematically added to both the generator model and the discriminator models as shown in Figure 5.1.

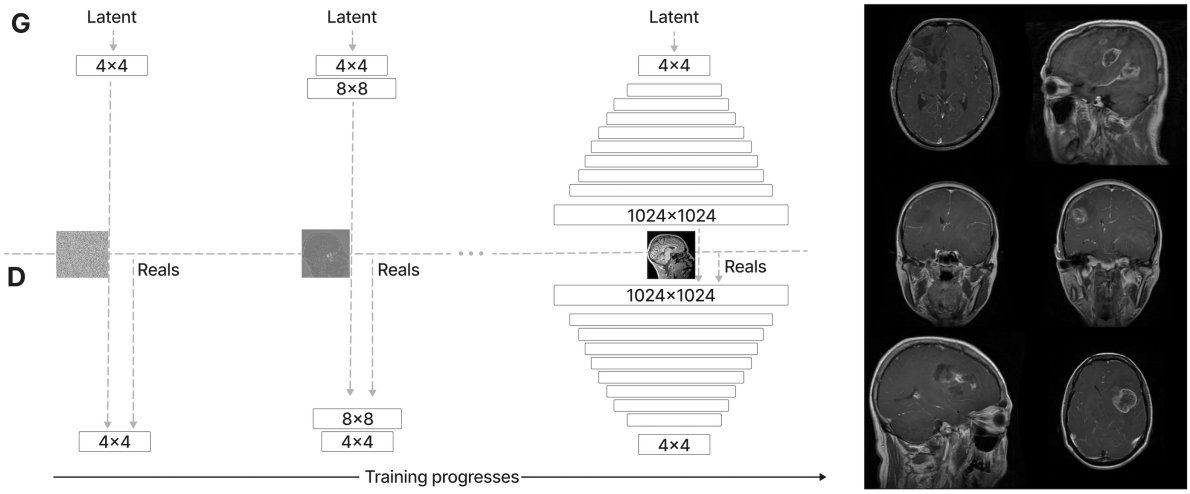


Figure 5.1: Diagram that outlines the PGGAN training process.

The incremental addition of the layers allows the models to effectively learn coarse-level detail and later learn ever finer detail, both on the generator and discriminator side. This incremental nature allows the training to first discover large-scale structure of the image distribution and then shift attention to increasingly finer scale detail, instead of having to learn all scales simultaneously.

In our specific case, our first trial was to generate synthetic chest X-ray images that exhibit pneumoconiosis features, which can be added to the original dataset to improve the classification model's performance.

The implementation of the PGGAN technique was based on an unofficial PyTorch code repository [37], developed by Karras and collaborators available to the public.

As part of this work we improved the repository to enable the use of pretrained models, this enhancement allowed us to fine-tune the PGGAN model using our chest x-ray dataset more effectively.

Table 5.2 presents a comparison of the original images from the Chongqing dataset (a, b, c) with images generated using PGGAN (d, e, f) and images generated using PGGAN fine-tuned with chest X-ray images (g, h, i). The table illustrates how the

generated images from both PGGAN and PGGAN fine-tuned with chest X-ray images can potentially enhance the dataset by providing additional examples for the classification model to learn from. By incorporating these generated images, we try to improve the model's ability to detect pneumoconiosis with higher accuracy and efficacy.

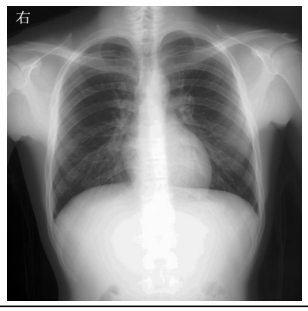
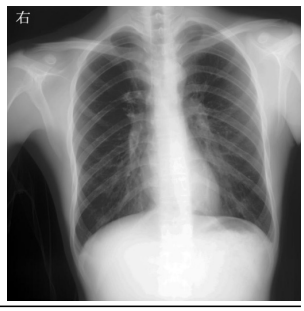
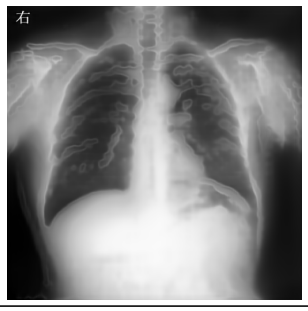
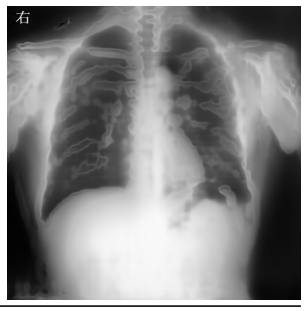
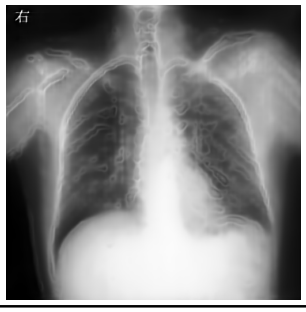
	Image 1	Image 2	Image 3
Original Dataset			
PGGAN			
PGGAN with NIH Chest X-ray Fine-Tuning			

Table 5.2: Comparison of original images (a, b, c) with those generated using PGGAN (d, e, f) and PGGAN fine-tuned with chest X-ray images (g, h, i) based on the different types of noise added.

As observed in Table 5.2, the quality of the generated images may not be optimal, possibly due to insufficient training. For example, the rib structures in the generated images appear irregular and unnatural, particularly in images (e) and (i), where the ribs seem distorted and not well-defined. This observation suggests that the PGGAN model may require further fine-tuning and additional training data to generate more accurate representations of chest X-ray images exhibiting pneumoconiosis features.

Given these results, the training time insumed and the relatively disappointing performance on the Chongqing dataset, we decided not to pursue PGGAN for the rest of the datasets. To provide some context regarding the time investment, we committed 18 days to training and fine-tuning these models using a dedicated Google Cloud instance equipped with a T4 GPU:

- 4 days were allocated to training a model exclusively with the Chongqing dataset.
- 7 days were dedicated to training a model exclusively with the Chest X-ray dataset.
- The remaining 7 days were spent on fine-tuning the model initially trained on the Chest X-ray dataset with additional data from the Chongqing dataset.

Future work might involve further refinement of the PGGAN model and exploration of its potential benefits on a larger and more diverse dataset.

5.2.2 Stable Diffusion

Despite several diffusion models being named for image generation, at the time of this work, only Stable Diffusion was both publicly released and provided code pieces for finetuning. Table 5.3 gives a better idea of the evaluated models and our decisions behind SD. Hence our work mainly focuses on using the different versions of SD models for synthetic data generation.

Model	**Company**	**Accessible thru API/APP**	**Publicly available**	**Weights available**	**Fine-tuning scripts**
Stable Diffusion	Stability.ai	✓	✓	✓	✓
DALL-E 2	OpenAI	✓	✗	✗	✗
Imagen	Google	✓	✗	✗	✗
eDiff	NVIDIA	✗	✗	✗	✗
GLIDE	OpenAI	✗	✗	✗	✗
Midjourney	Midjourney	✓	✗	✗	✗
CM3leon	Meta	✓	✗	✗	✗

Table 5.3: Table summarizing aspects of the models contemplated for this project.

Within the scope of this study, we employed the Stable Diffusion technique [18] as a generative method to augment our dataset. This approach not only enhances the diversity and quality of the original training dataset but is also recognized for its proficiency in generating images closely resembling real-world instances. An overview of its arquitecture can be seen at Figure 5.2

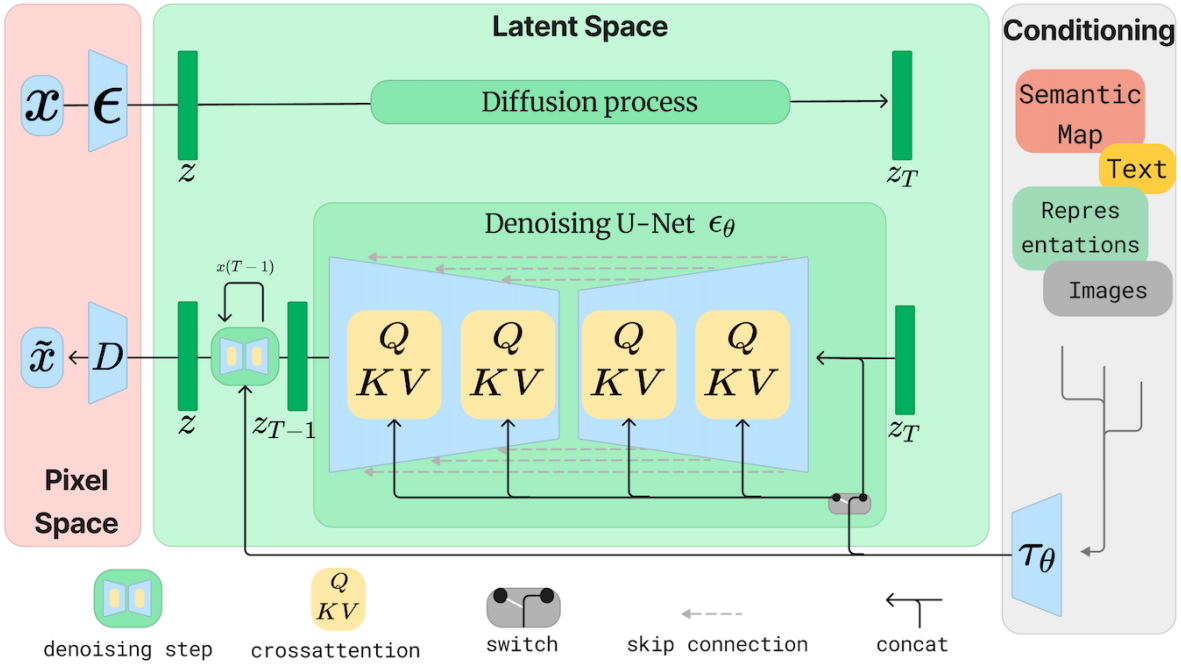


Figure 5.2: Stable Diffusion architecture.

The initial stage of our investigation involved the implementation of the Stable Diffusion method through the Hugging Face's StableDiffusionImg2ImgPipeline [38]. The process involved introducing noise and later prompting the model to reconstruct the altered image. This approach enabled us to swiftly conduct testing and generating synthetic data from an initial image.

Subsequently, we explored various models (versions) by adopting the widely employed approach in diffusion models, where images are generated from text using Hugging Face's StableDiffusionPipeline [39]. This class stores all components from the SD architecture (models, schedulers, and processors presented in Figure 5.2) for diffusion pipelines and provides methods for loading, downloading and saving SD models. At inference time the steps followed by the pipeline are described in Figure 5.3

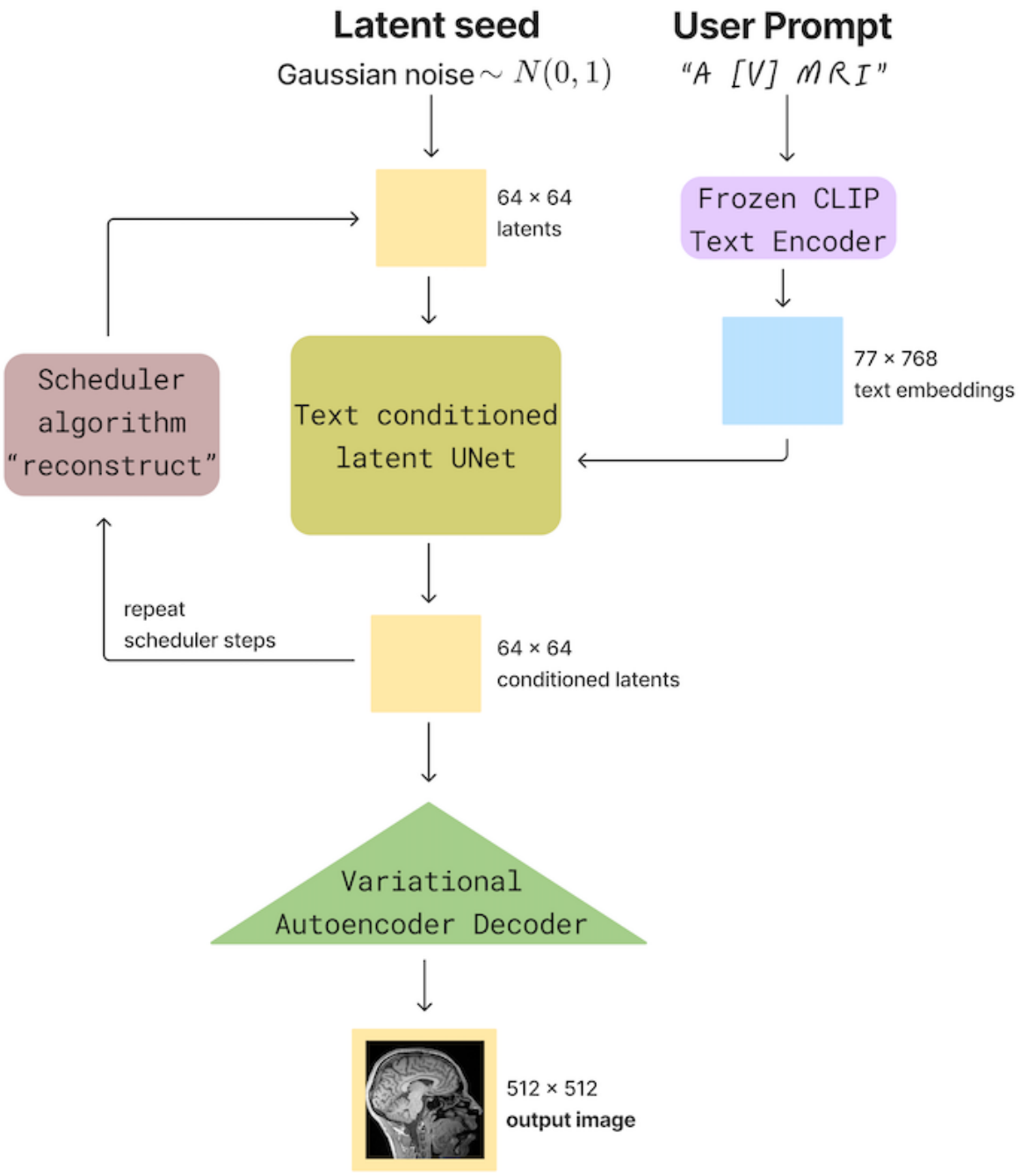


Figure 5.3: Stable Diffusion during inference.

The techniques researched and utilized for fine-tuning these models are summarized in the Table 5.4.

Version/Technique	Text. Inv.	DreamBooth	LoRa	HyperDreamBooth
v1.4	✗	✓	✗	✗
v1.5	✓	✓	✓	✓
v2.0	✗	✓	✗	✗
v2.1	✗	✓	✓	✗
XL 0.9	✗	✗	✓	✗
XL 1.0	✗	✗	✓	✗

Table 5.4: Table showing SD versions and fine-tuning techniques used during the course of our project.

- **Textual Inversion** [40]:

  In DMs, a text encoder translates input prompts into embeddings that steer the model's output. This involves tokenizing the input into indexed tokens via a predefined dictionary, and then passing these tokens through the text encoder. The result is a set of unique embedding vectors, accessible through indexing. These vectors serve a dual purpose: they guide a downstream UNet model and work in conjunction with latent image input. This enables the modification of token embeddings to either generate a variety of images or to facilitate the learning of new concepts.

  Textual inversion capitalizes on this embedding space to introduce new vocabulary to the text model by using a limited set of example images to fine-tune embeddings that are closely aligned with visual representations, as illustrated in Figure 5.4. This process involves adding new tokens to the existing vocabulary and training the model with representative images. The outcome is a set of novel embedding vectors that encapsulate specific concepts. A designated placeholder string S marks these new concepts, replacing the original tokenized string vector with a newly acquired embedding v_*. Importantly, this addition does not alter the foundational generative model, preserving its utility for new applications.

  In our experiments, Textual Inversion proved to be an efficient and accessible method for fine-tuning SD models. However, as shown in Table 11.2 in the Appendix, the results were less than optimal, leading us to discontinue further work in this direction.

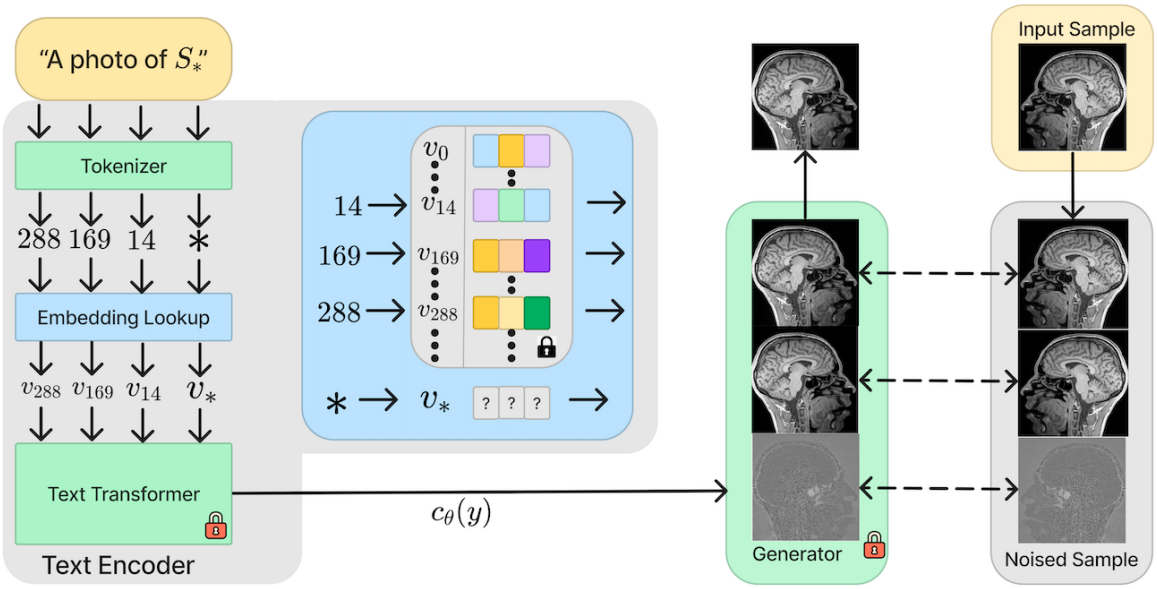


Figure 5.4: Diagram that outlines the text-embedding and inversion process.

- **Dreambooth** [41]:

  As shown in Figure 5.5 this technique involves refining a text-to-image diffusion model with a small collection of images depicting a subject. This refinement entails pairing the input images with a text prompt encompassing a distinctive identifier and the designated class to which the subject pertains (for instance, "A [V] MRI"). Concurrently, they employ a class-specific prior preservation loss capitalizing the model's inherent semantic knowledge pertaining to the class, fostering the generation of a varied array of instances associated with the subject's class. This is achieved by incorporating the class name into a text prompt (for example, "An MRI"), thus encouraging diverse and representative outputs.

  The initial description in the Dreambooth paper explains a method to improve the UNet part of the model while keeping the text encoder unchanged. However, based on an article published from Hugging Face [42] the text encoder was also adjusted since it has been demonstrated that tweaking the text encoder leads to the most favorable results. This adjustment translates into more realistic images and it reduces the risk of fitting the model too closely to the training data while also improving the model's ability to understand and work with more complex prompts.

  In our experiments, this technique has proven itself to be an effective and accessible method for refining SD models. Unlike Textual Inversion, Dreambooth exhibited notably positive outcomes. Our research was primarily dedicated to investigating and enhancing this specific technique, proving to be instrumental in facilitating the learning of new concepts by the SD models with minimal computational burden.

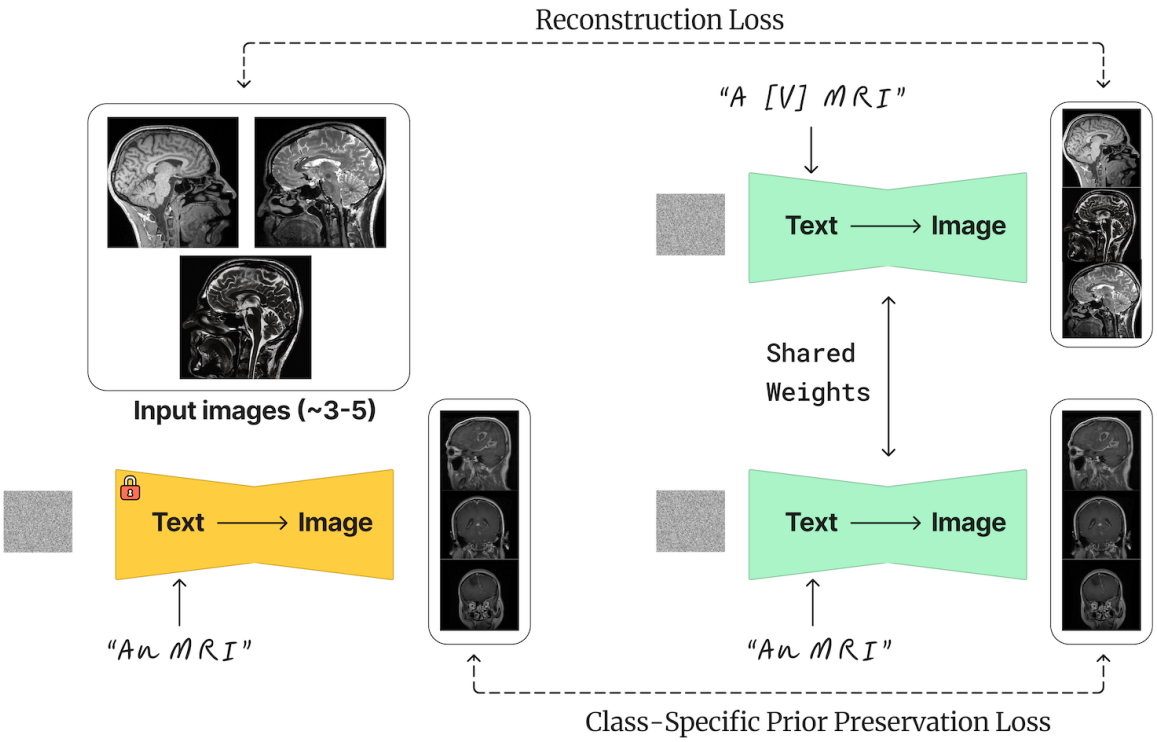


Figure 5.5: Diagram that outlines the DreamBooth fine-tuning process.

- **LoRA** [43] (in combination with Dreambooth):

  LoRA, which stands for Low-Rank Adaptation of Large Language Models, is a fresh approach devised by Microsoft researchers to address the challenge of refining extensive language models. Coping with the expense of fine-tuning large models like GPT-3, which possess billions of parameters, can be daunting. LoRA introduces a strategy where the original model's weights are kept fixed, and instead, trainable layers (in the form of rank-decomposition matrices) are integrated into each transformer block. This innovation significantly diminishes the count of trainable parameters and the GPU memory prerequisites, as most model weights do not necessitate gradient computations. The researcher's findings indicate that by concentrating on the Transformer attention blocks within expansive language models, employing LoRA for fine-tuning yields comparable quality to full model fine-tuning, all while being notably faster and demanding less computational resources.

  While originally introduced for large-language models and showcased through transformer blocks, LoRA's applicability extends beyond these bounds. In the context of fine-tuning Stable Diffusion, LoRA can be effectively implemented within the cross-attention layers, which establish the connection between image representations and corresponding descriptive prompts. The Figure 5.6 gives a better idea on where are this cross-attention layers present in the SD model architecture.

  When combined with Dreambooth, as documented by Hugging Face [44], LoRA demonstrated efficiency in fine-tuning SD models. It is worth noting, however, that the results were not as impressive as when using Dreambooth alone. For our research, we exclusively used the SD XL model variant in combination with LoRA, as it offered to us visually superior results.

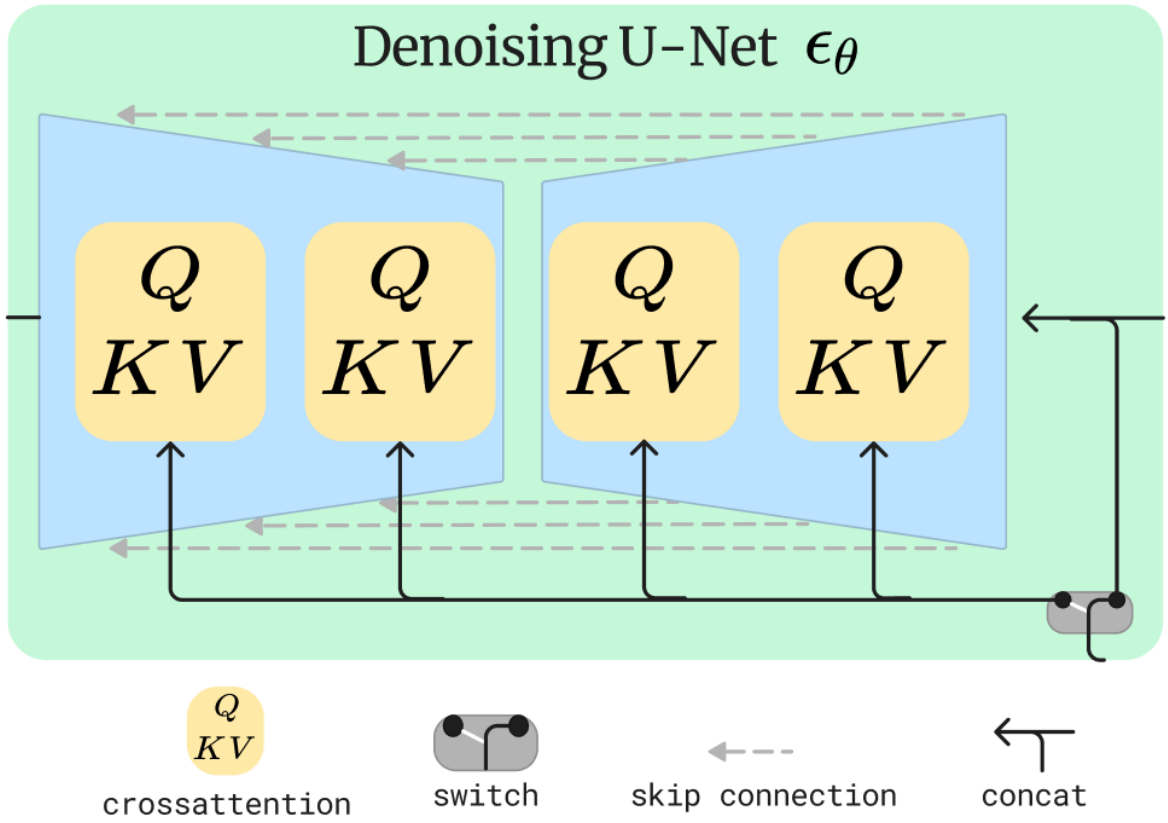


Figure 5.6: Zoom-in of the Denoising U-Net. Yellow blocks are the cross-attention layers, the ones in charge of building the relationship between image and text representations where the trainable matrices are injected.

- **Perfusion** [45]

  Perfusion, a promising technique introduced recently, aims to address two conflicting goals in personalized text-to-image (T2I) models: (1) Avoiding overfitting to ensure that the model doesn't replicate example images too precisely and (2) Preserving the identity or theme across generated images, even when those images vary in presentation. However, these two goals go against each other since when the model copies the examples too closely, it's good at keeping the idea the same, but it struggles to make new and unique pictures when asked to be creative and vice versa.

  The authors propose what they term a "naïve solution" to address both goals. They distinguish between two pathways within the model: the "*K pathway*", which governs the features of the generated content (the *What*), and the "*V pathway*", which dictates the spatial arrangement of these features (the *Where*). To simplify, one part decides where objects should be in the picture, while the other part determines what these objects should look like. By isolating these functions, the model can better balance creativity and consistency. To achieve these objectives effectively, they suggest that whenever the encoded content contains the desired concept, ensure that its cross-attention keys align with those of its broader category, a technique referred to as "*Key Locking*". Furthermore, they aim for the cross-attention values to accurately represent the concept within the multi-resolution latent space. Addi-

tionally, they have developed another gated rank-1 approach that enables control over the influence of a learned concept during inference and allows for combining multiple concepts. Figure 5.7 illustrates these pathways and the gated rank-1 concept.

As of now, the authors have yet to release the source code for Perfusion, limiting its current applicability.

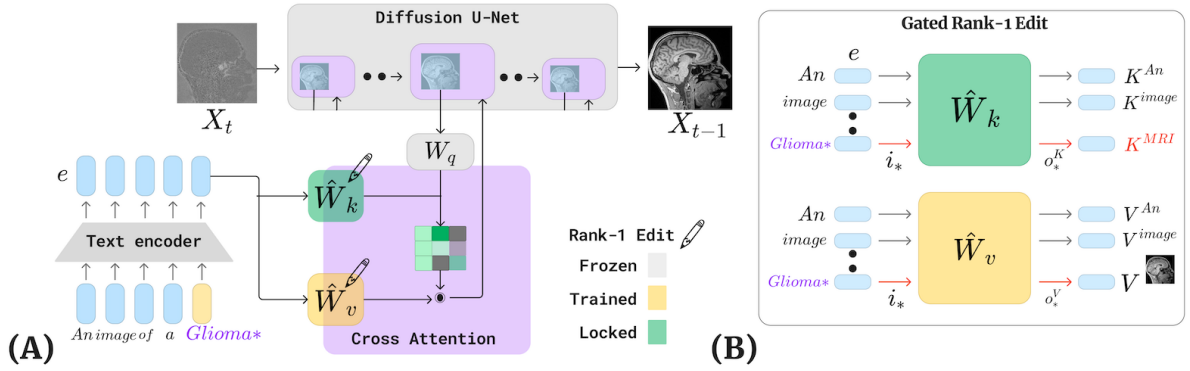


Figure 5.7: Diagram showing *Architecture outline (A)* and the *Gated Rank 1 Edit (B)*.

- **HyperDreamBooth** [46]

  This is a very recent technique introduced during the advanced stages of our work and developed by the same creators of DreamBooth. Their research aims to address the challenges posed by the size and speed of their previous technique while maintaining model quality, editability, and subject faithfulness. As illustrated in Figure 5.8 this is achieved through a two-step process: (1) utilizing a HyperNetwork to create an initial estimation of a portion of network weights, which are subsequently (2) enhanced through rapid fine-tuning to capture subject-specific details with high fidelity. This method ensures the preservation of model consistency and style variety while closely approximating the subject's essence and details.

  The motivation behind exploring the HyperNetwork approach lies in the understanding that to create highly accurate representations of specific subjects using a predefined generative model, it is imperative to adjust the way the model works. This involves incorporating information about the desired subjects by modifying the model's weights.

  Although this technique was originally designed with faces in mind, we decided to investigate its potential. It is important to note that there is no official implementation available, and the current code available is still a work in progress [47]. However, as shown in Table 11.3 in the Appendix, the results obtained from this method were not very promising, leading us to discontinue further exploration in this direction.

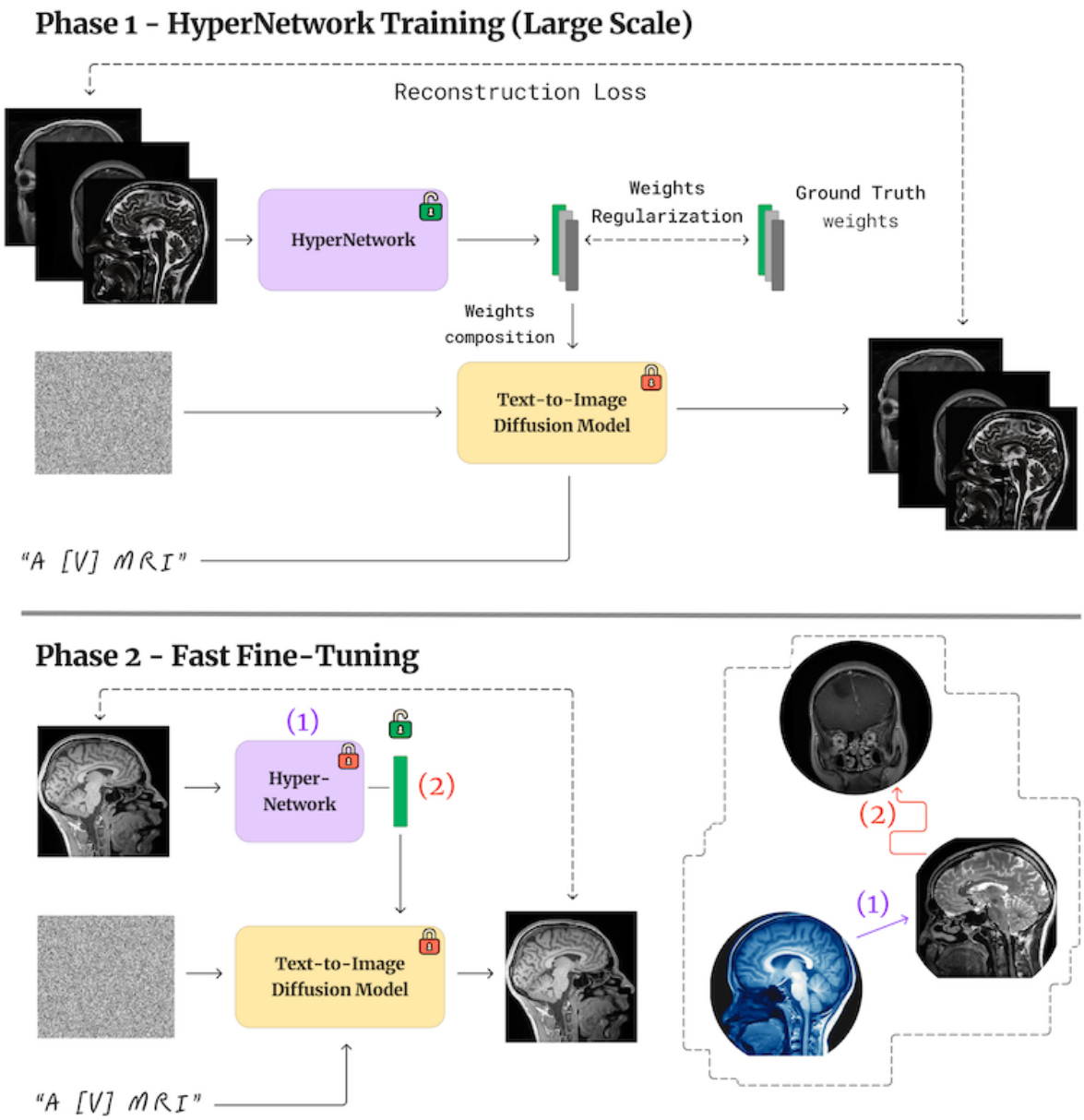


Figure 5.8: Diagram illustrating the *HyperDreamBooth Training* and *Fast Fine-Tuning* phases.

For more detailed visualizations of the images generated by the discarded techniques, please refer to Section 11.3 in the Appendix.

Stable Diffusion *v1*

During the initial phase, we utilized the stable-diffusion-v1-5 model [48] from Runway [49] to introduce varying degrees of noise to our images and we asked the model to reconstruct them. Specifically, we added noise levels ranging from 5 to 20 percent to each image in our dataset over 24 inference steps and, in one instance, over 50 steps. Table 5.5 provides a comparison between the original images from the Chongquing dataset and their synthetic counterparts.

Although the differences between the original and noise-added images may appear minimal at first glance, closer inspection reveals subtle variations. Each row in the table represents a unique image from our dataset, with noise levels incrementally increasing from left to right across the columns. Also the addition of further noise than 0.15 caused a clear difference in images.

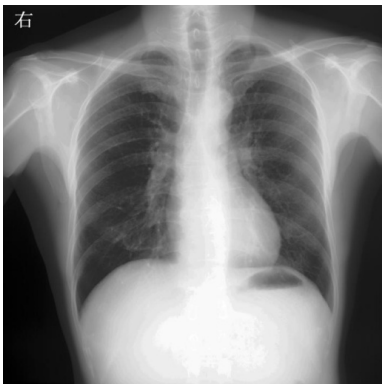
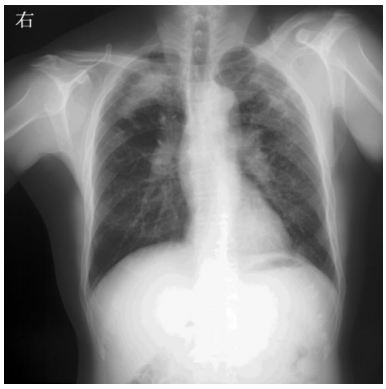
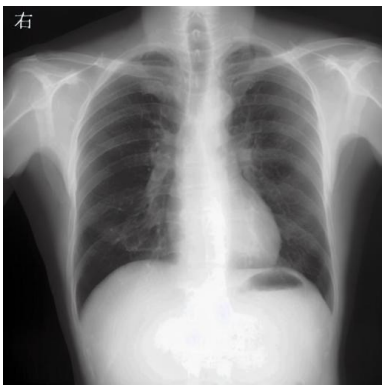
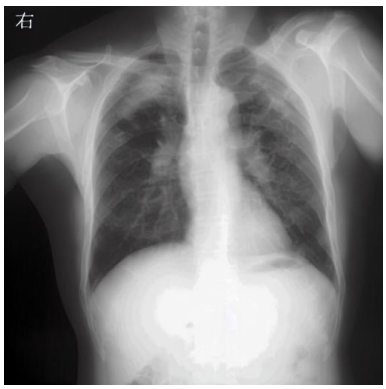
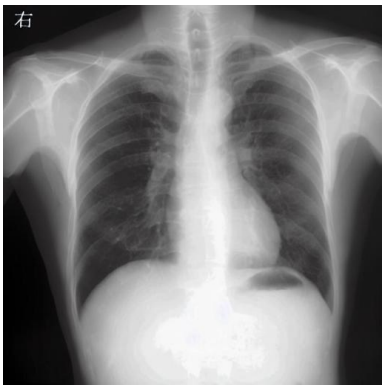
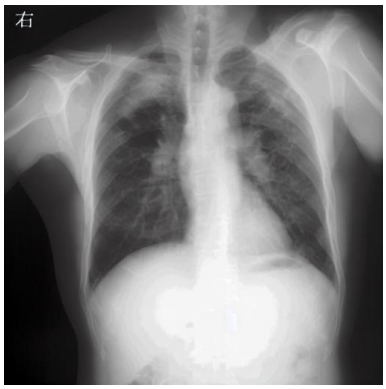
	Image 1	Image 2
Original		
Noise 0.1		
Noise 0.15		
Noise 0.2		

Table 5.5: Comparison of original images with those generated by Stable Diffusion adding different levels of noise.

The performance of the model trained on the Chongqing dataset, which was augmented with noise using runwayml/stable-diffusion-v1-5, is outlined in Table 6.1 in the final chapter on results.

To build upon this initial exploration, we applied the principles of DreamBooth, a technique for personalizing text-to-image diffusion models, as outlined in a guide elaborated by Tryolabs [50]. Our objective was to expand the model's language-vision dictionary, enabling it to associate new words with specific subjects we wanted to generate.

We started by fine-tuning the Stable Diffusion model using the Chongqing dataset, focusing on the positive cases. We introduced a prompt for these images: "image of a **pneumoconiosis** xray" using model's existing class knowledge through the "*xray*" class. This approach allowed the model to leverage its prior knowledge of the subject class while associating the class-specific instance with the unique identifier.

For each real image in the Chongqing true cases, we generated one synthetic image. The results of this fine-tuning process were promising, with the model demonstrating an enhanced ability to generate novel renditions of a subject in different contexts while maintaining its distinctive features. Furthermore, as evidenced in Table 6.1 in the final chapter on results, under the title "*Stable Diffusion with Single-Class Fine-Tuning*", the performance of the fine-tuned model improved the base model as outperform the original dataset across all metrics. This advancement affirms the potential of SD as a data augmentation strategy, and underscores the potential of such synthetic data generation techniques in improving ML model performance.

Considering the previous results, we deemed it interesting for this project to involve a medical radiologist with specialized expertise. Her role was to evaluate and validate the authenticity and realism of the synthetically generated images. The radiologist meticulously examined a subset of 30 images, assessing factors including their fidelity to actual X-rays, anatomical precision, and the capability to accurately detect pneumoconiosis. A thorough report detailing the results of this evaluation can be found in the Appendix, specifically within Section 11.1.

In addition, we attempted multi-class fine-tuning instead of single-class using the same dataset, with the objective of enhancing the model's capacity to differentiate between instances of pneumoconiosis and non-cases. To clarify single-class refers to fine-tune using only the positive cases whereas multi-class refers to fine-tune the model using both classes, positive and negative. We generated one synthetic image for each real image in the Chongqing true cases. However, the results did not meet our expectations and are outlined in Table 6.1 in the results chapter, under the title "*Stable Diffusion with Multi-Class Fine-Tuning*".

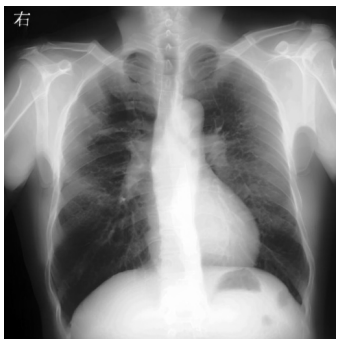
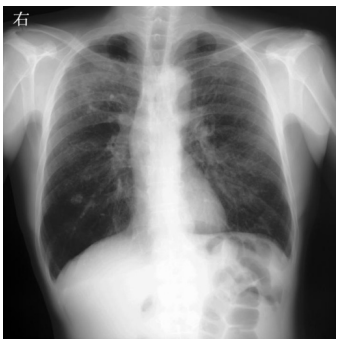
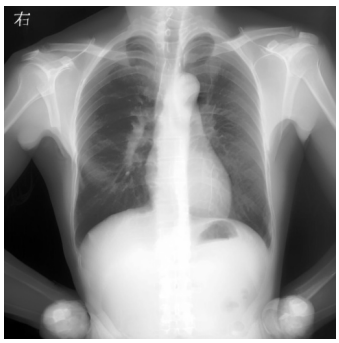
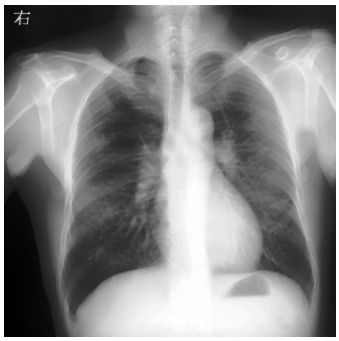
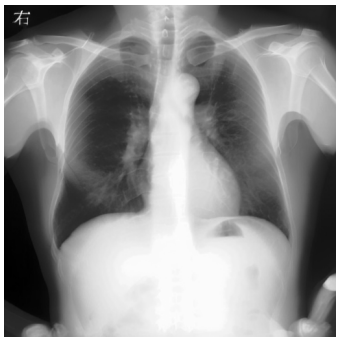
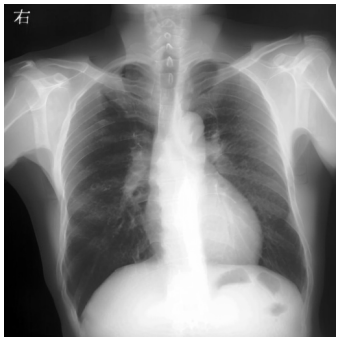
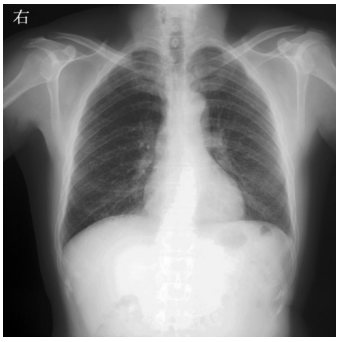
	Original	Stable Diffusion: Single-Class Fine-Tuning	Stable Diffusion: Multi-Class Fine-Tuning
Image 1			
Image 2			
Image 3			

Table 5.6: Comparison of original images with those generated by Stable Diffusion with different fine-tuning approaches.

Table 5.6 presents a comprehensive comparison between the authentic images from the Chongqing real cases and their corresponding synthetic versions generated by the two distinct model fine-tuning approaches: single-class and multi-class fine-tuning. The results show a noticeable disparity in the quality of generated images. Specifically, the images produced through single-class fine-tuning exhibit a higher level of quality compared to those generated via multi-class fine-tuning. Notably, the single-class fine-tuning outperforms in terms of visually assessing ribs, whereas the multi-class fine-tuning portrays anomalies in arm appearance and demonstrates significant variation in lung size.

Building on our experience with the Chongqing dataset, we moved into the application of SD to the other medical imaging contexts described before, continuing our work by fine-tuning the Stable Diffusion model using the Glioma class of the Human Brain MRI Dataset. The prompt used for these images was "an image of a **glioma-brain-tumor**

MRI", using once more model's prior concepts about MRI and allowing the model to utilize its previous knowledge of the class while associating the class-specific instance with the unique identifier.

For the MRI dataset we doubled the size of each Glioma class by generating new images, however, due to the diverse MRI scan views within the dataset (top-down, rear, left profile, right profile, etc.) the quality of the generated images was not satisfactory. We believe this diversity introduced substantial noise and confusion for the SD model, impeding its ability to generate coherent and realistic synthetic images. As depicted in Table 5.7, the Stable Diffusion v1 with Dreambooth approach generates synthetic Glioma images that exhibit an invariant perspective. This invariant perspective is inconsistent with the diverse viewing angles present in the original Glioma images.

As a response to the mentioned problem an intermediate approach was executed based on image clustering, as detailed in Section 5.1. Our solution of clustering proved to reduce the diversity of MRI views, potentially enabling SD to generate higher-quality images by learning from more consistent and uniform MRI views. This is can be observed with the synthetic images generated from multiple perspectives and more accurately, reflecting the original dataset's diversity.

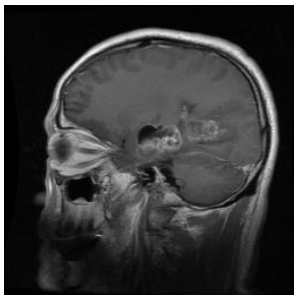
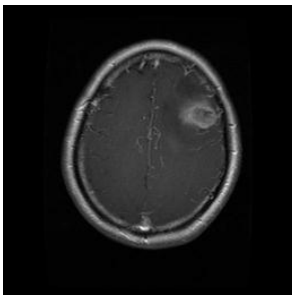
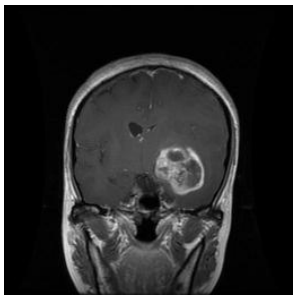
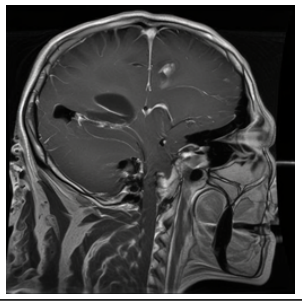
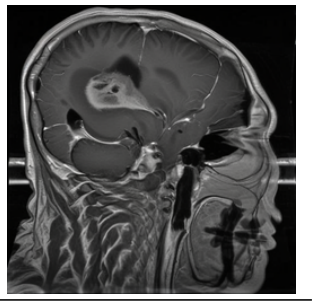
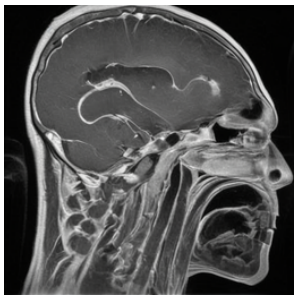
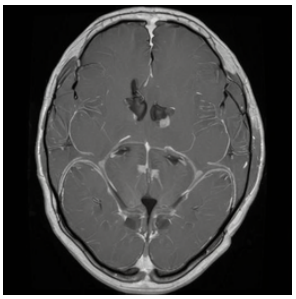
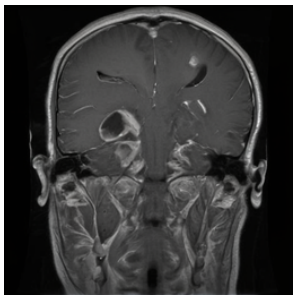
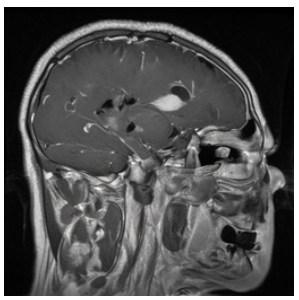
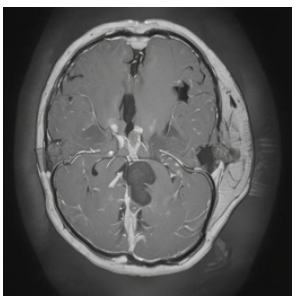
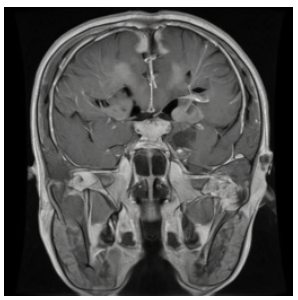
	Image 1	Image 2	Image 3
Glioma Original			
Stable Diffusion: Glioma Fine-Tuning			
Stable Diffusion: Glioma Clustering with EfficientNet			
Stable Diffusion: Glioma Clustering with Clip			

Table 5.7: Comparison of original images with those generated by Stable Diffusion v1 with different fine-tuning approaches.

In a similar vein, our efforts with the Diabetic Retinopathy dataset to generate synthetic images of mild retinopathy were met with challenges. At this time the prompt used was "an image of a **mild** retinopathy", making use again of model's prior concepts. The results were not satisfactory, and in some instances, the images generated even include two retinas, as can be seen in Table 5.8. These challenges appear to stem from the complexity of the retinal structures and the specific characteristics of mild retinopathy, mirroring the obstacles faced with the MRI dataset.

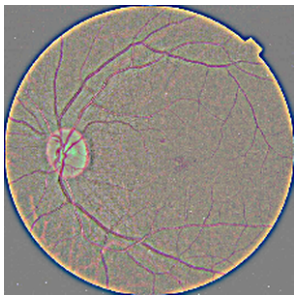
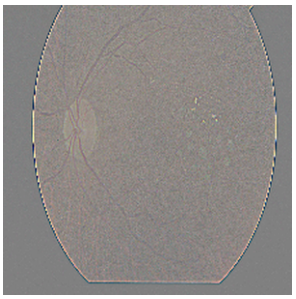
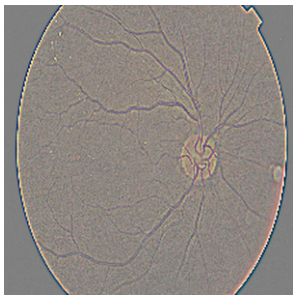
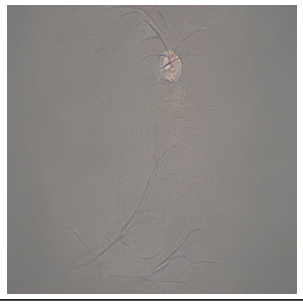
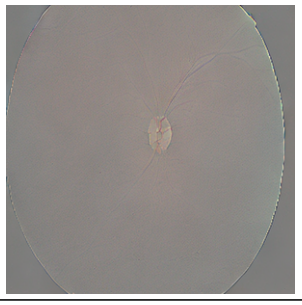
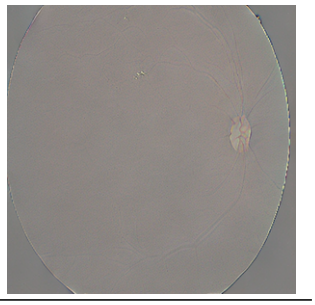
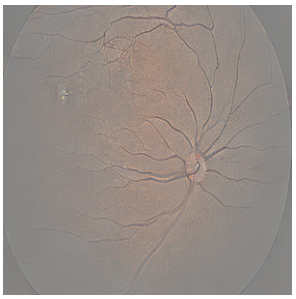
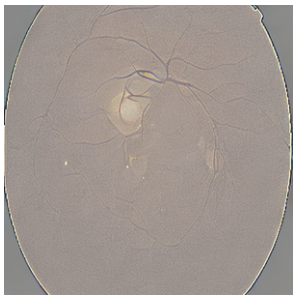
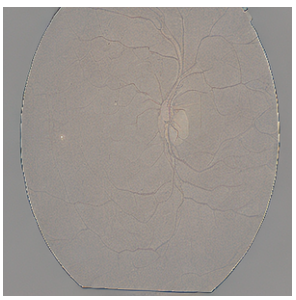
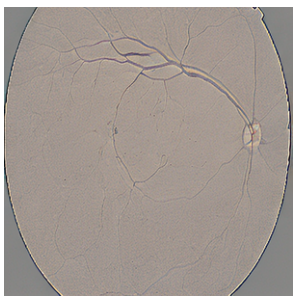
	Image 1	Image 2	Image 3
Retinopathy Original			
Stable Diffusion: Mild Retinopathy Fine-Tuning			
Stable Diffusion: Mild Retinopathy Clustering with EfficientNet			
Stable Diffusion: Mild Retinopathy Clustering with Clip			

Table 5.8: Comparison of original images with those generated by Stable Diffusion v1 with different fine-tuning approaches.

Nonetheless, despite the noticeable improvement in image generation quality, the SD $v1$ with Dreambooth and using Clustering strategy failed to outperform the original SD $v1$ with Dreambooth in terms of classification metrics. This outcome emphasizes that improved visual fidelity of synthetic images does not necessarily translate to better classification performance, indicating the complexity and challenges in working with diverse medical datasets.

The results of the clustering strategy will be discussed in the results chapter (Table 6.2 under "Stable Diffusion v1 Dreambooth Clustering"). A parallel analysis for the Diabetic Retinopathy dataset can be found in the results chapter as well, under Table 6.3, where the challenges and outcomes closely resemble those encountered with the MRI dataset.

Stable Diffusion *v2*

In our research, we also explored stable-diffusion-2-1 [51] from Stability AI [52] for the generation of synthetic medical images. Although the *v*2 of this model represents a departure from its predecessor, with modifications to the CLIP model in the text encoder for increased accuracy and the inclusion of a depth model, it did not fulfill the specific requirements of our study.

The orientation of SD *v*2.1 towards photorealism, while a significant advancement in certain contexts, did not align with our needs, as the dataset used for training did not seem to fit the specific styles required in medical images. Furthermore, the conceptual understanding of medical themes within SD *v*2.1 appeared to be less refined compared to earlier versions, leading to a decrease in the quality of generated images as depicted in Table 5.9.

In conclusion despite the underlying improvements and the potential for future enhancements, the immediate application of SD *v*2.1 within our project yielded results that were not satisfactory. Consequently, our research necessitated the continued utilization of earlier versions and alternative models, which demonstrated a more suitable alignment with the complexities of medical image synthesis.

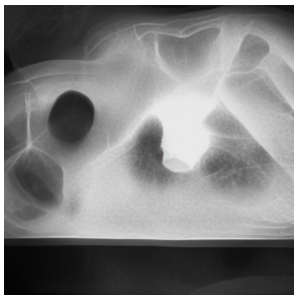
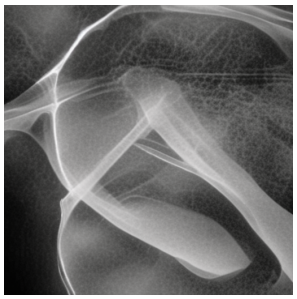
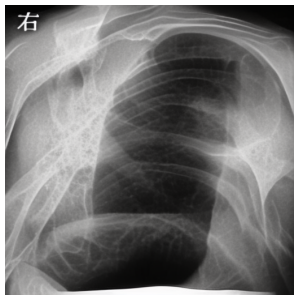
	Image 1	Image 2	Image 3
Pneumoconiosis			

Table 5.9: Images generated with Dreambooth using Stable Diffusion *v2.1*.

Stable Diffusion XL

The stable-diffusion-xl-base-1.0 [53] model, introduced also by Stability AI along with the research conducted by D. Podell et al. [54], builds upon foundational diffusion models. This model incorporates several noteworthy enhancements. For instance, it employs a heterogeneous distribution of transformer blocks within the architecture of the UNet [55]. Additionally, it leverages powerful pre-trained text encoders to bolster its capabilities. Notably, the SDXL model boasts a substantial parameter count of 2.6 billion, signifying its complexity and potential. One of its novel contributions is the incorporation of micro-conditioning techniques, including size-conditioning to modify the appearance of an output corresponding to a given prompt. An overview of its architecture can be seen in Figure 5.9.

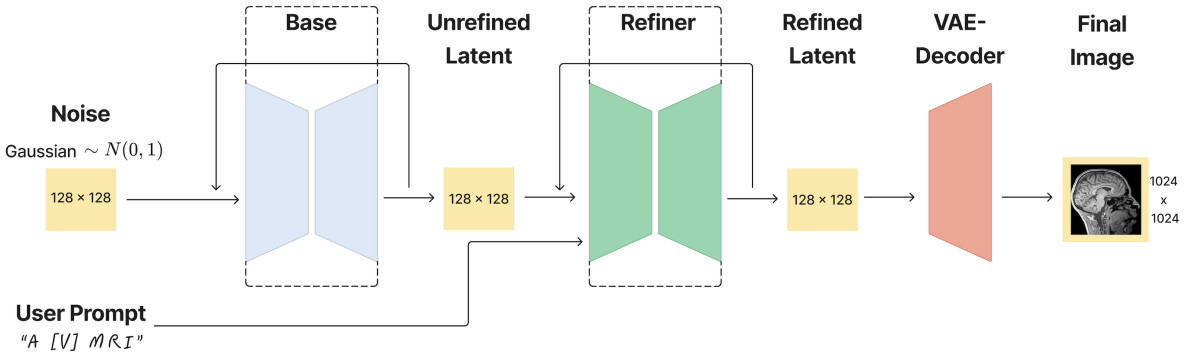


Figure 5.9: Stable Diffusion XL two-stage overview architecture.

However, due to the substantial scale of the model, fine-tuning is only achievable through the utilization of the LoRA technique and in combination once more with DreamBooth. This technique has demonstrated commendable performance in terms of training times and computing costs, although accompanied by the limitation of producing results that may not exhibit the same level of quality. Consequently, even when employing a more advanced model like SDXL with the LoRA technique on medical datasets, the outcomes were not optimal.

Although this combination outperformed earlier iterations of SD models and introduced innovative functionalities, the images generated within our specific medical context were of unsatisfactory quality. The unique characteristics of medical datasets, coupled with the constant evolution of the underlying architecture, did not align well with the intricacies involved in synthesizing medical images. As a result, we persisted in relying on previous model versions and alternative approaches that demonstrated a better fit for addressing our specific requirements. Table 5.10 gives a clear idea of what was mentioned before.

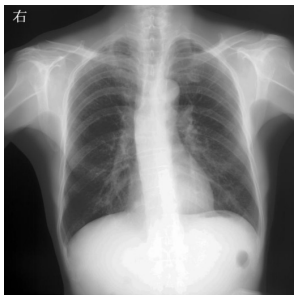
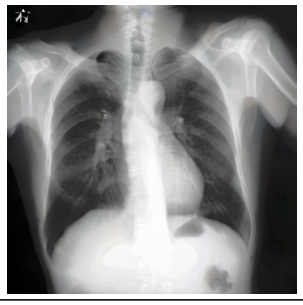
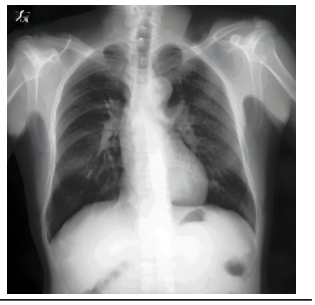
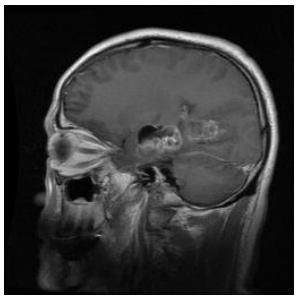
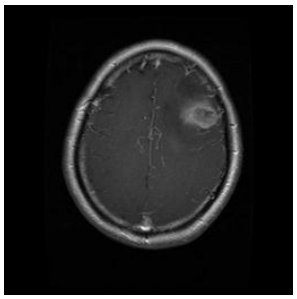
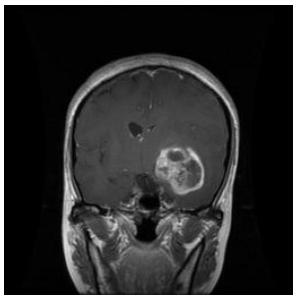
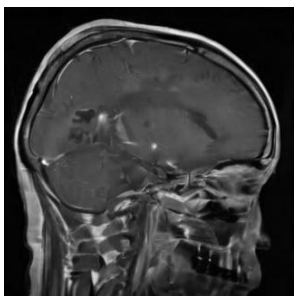
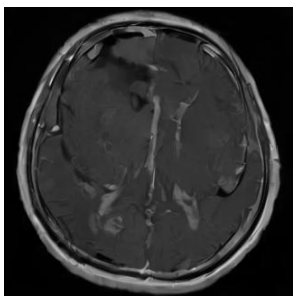
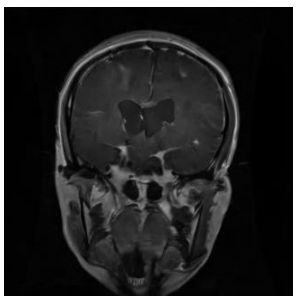
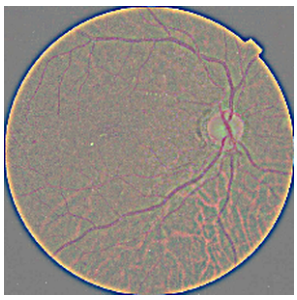
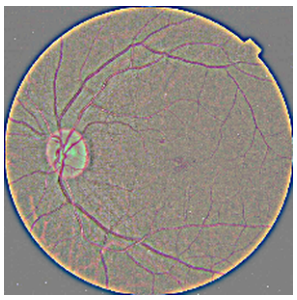
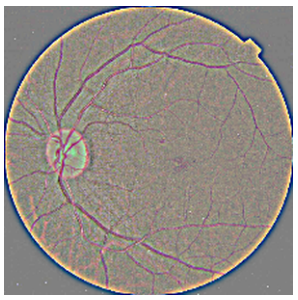
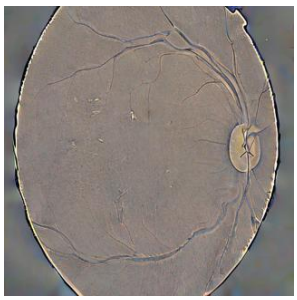
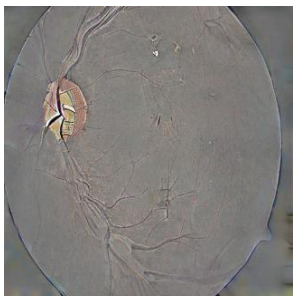
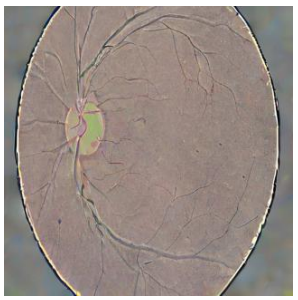
	Image 1	Image 2	Image 3
Pneumoconiosis Original			
Stable Diffusion: Retinopathy Clustering with Clip			
Glioma Original			
Stable Diffusion: Glioma Clustering with Clip			
Retinopathy Original			
Stable Diffusion: Retinopathy Clustering with Clip			

Table 5.10: Comparison of original images with those generated by Stable Diffusion XL *v1.0*.

The XL architecture, as presented in Figure 5.9, comes with a second model called "*Refiner*" [56] precisely for a more refined end in images using a two-stage pipeline when generating them. First, the base XL model is used to generate latents of the desired output size. In the second step a specialized high-resolution model is used, applying an "*img2img*" technique to the latents generated in the first step, using the same prompt. This technique is slightly slower than the regular one, as it requires more function evaluations. However this only gave to the images cartoon-like aspect and hence this step was removed. Some example images generated by this two-stage approach can be seen in Table 11.4 within the Appendix.

Problem: Complexity of Medical Concepts and Lack of Pre-existing Visual Priors

The application of the different techniques methods in image generation for specific medical conditions like mild retinopathy, pneumoconiosis chest X-ray, or glioma brain tumor brings up two significant challenges:

1. **Complexity of Medical Concepts:** Each medical condition carries a unique set of complexities and variations that the model needs to understand. For instance, recognizing a glioma brain tumor involves understanding its appearance in diverse patient groups, various disease progression stages, and different MRI settings. Without such knowledge, the model may fail to generate accurate and clinically meaningful images.

2. **Lack of Pre-existing Visual Priors:** The model might not have a pre-existing "visual prior" or conceptual understanding of specific medical conditions. This is due to the likelihood that these particular conditions were underrepresented or not present at all in the model's initial training data. Without a robust prior, the model may not accurately generate images of these medical conditions.

To visualize the different models' existing priors and what we have previously commented, the Table 5.11, Table 5.12 and Table5.13 show their current concept understanding for the worked datasets across different SD versions.

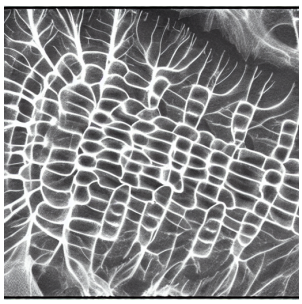
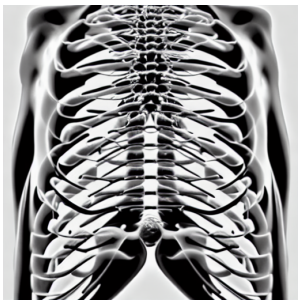
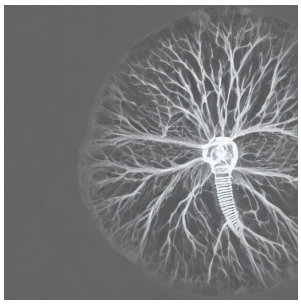
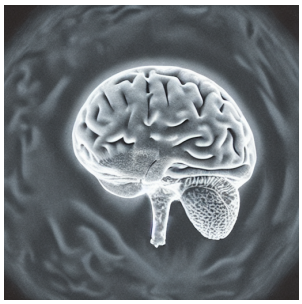
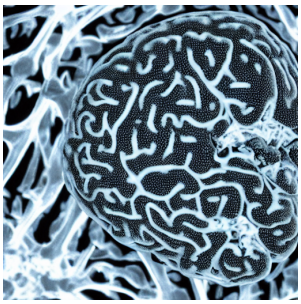
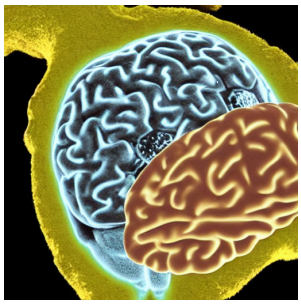
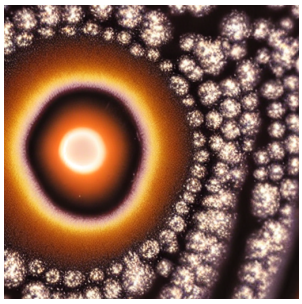
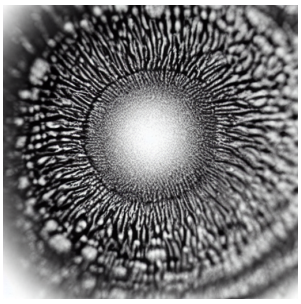
	Image 1	Image 2	Image 3
X-ray			
MRI			
Retinopathy			

Table 5.11: Image samples of several concepts from the model's visual priors using Stable Diffusion v1.5.

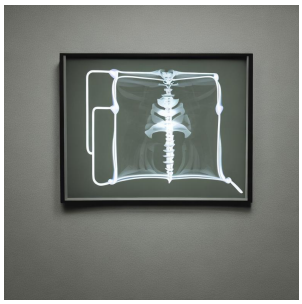
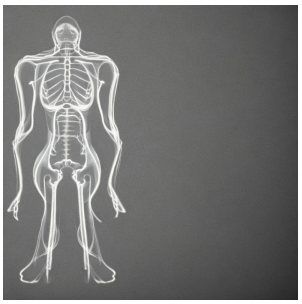
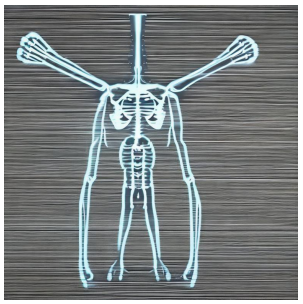
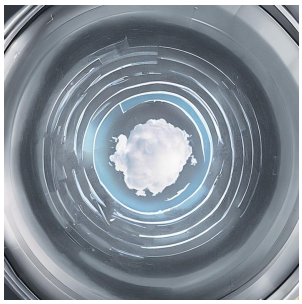
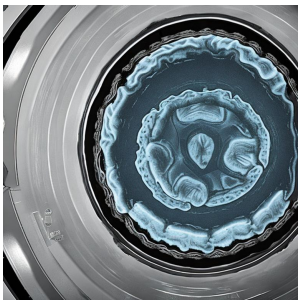
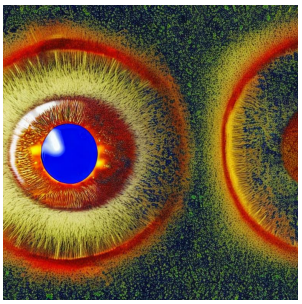
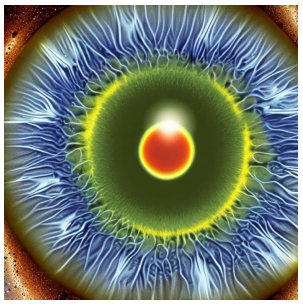
	Image 1	Image 2	Image 3
X-ray			
MRI			
Retinopathy			

Table 5.12: Image samples of several concepts from the model's visual priors using Stable Diffusion v2.1.

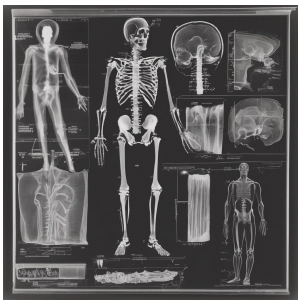
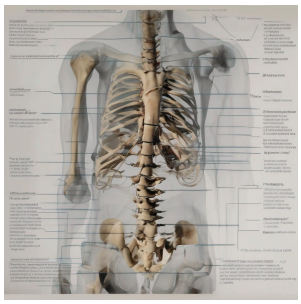
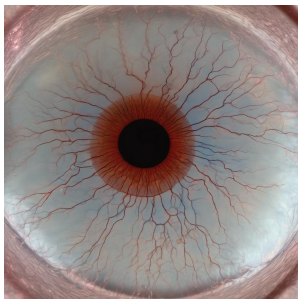
	Image 1	Image 2	Image 3
X-ray			
MRI			
Retinopathy			

Table 5.13: Image samples of several concepts from the model's visual priors using Stable Diffusion XL.

5.3 Combining datasets

Combining real and synthetic images for training a classification model introduces several challenges that require careful consideration. One of the main issues is determining the optimal balance between real and synthetic data. While synthetic data augmentation can enhance the diversity of the training set, an excessive reliance on synthetic data may introduce biases or inconsistencies that hinder the model's generalization capability. Striking the right balance between real and synthetic data is crucial to avoid overfitting or under-representing certain real-world scenarios.

Our own study did not primarily focus on a comprehensive evaluation of these challenges. Azizi et al.'s work [17] offers critical insights that underscore the crucial balance between image quality and diversity. Specifically, they found that models exclusively trained on synthetic data fell short in performance when benchmarked against those trained on real-world data, especially in the context of ImageNet classification tasks.

We recognize the importance of this phase, which holds weight comparable to other

aspects of the research process. However, a thorough examination of this stage would entail significant effort, involving rigorous testing and evaluation to ascertain the optimal mix of synthetic and real images. Such an endeavor could feasibly be a separate research project. Given these considerations, we opted for a straightforward strategy: we doubled the size of the selected classes while maintaining a consistent methodology throughout.

5.4 Classification Model

In this study, we used Cogniflow to develop classification models aimed at solving three distinct medical challenges: pneumoconiosis detection, brain tumor identification, and diabetic retinopathy classification. The datasets employed for these tasks are the Chongqing dataset for pneumoconiosis, the Human Brain MRI dataset for brain tumors, and the Diabetic Retinopathy dataset for retinopathy classification.

	Chongqing	Human Brain MRI	Diabetic Retinopathy
Model	FFNN	SVC (rbf)	Logistic Regression
Learning Rate	0.001	NA	NA
Batch Size	64	NA	NA
Max Epochs	300	NA	NA
Optimizer	Adam	NA	NA
Kernel Regularizer	L2	NA	NA
Kernel Regularizer Value	0.02	NA	NA
Hidden Layers	[256, 256, 256, 256]	NA	NA
Activations	['relu', 'relu', 'relu', 'relu']	NA	NA
Batch Normalization	False	NA	NA
Gamma	NA	scale	NA
Kernel	NA	rbf	NA
C	NA	10	10
Maximum Iterations	NA	1000	300
Tolerance	NA	0.01	NA
Random State	NA	123	123
L1 ratio	NA	NA	0.5
Multi Class	NA	NA	ovr
Penalty	NA	NA	elasticnet
Solver	NA	NA	saga
Evaluation	Improved with synthetic data	No improvement with synthetic data	No improvement with synthetic data

Table 5.14: Summary of Classification Models for Different Datasets

It is crucial to clarify that the primary aim of our project is not the optimization or fine-tuning of these classification models for their respective medical tasks. Rather, our central focus lies in the assessment of data augmentation techniques. The models,

as summarized in Table 5.14, were automatically selected by Cogniflow for each dataset. We deliberately kept the model parameters consistent across the different medical tasks to isolate and evaluate the impact of data augmentation on model's performance. This methodology provides a controlled environment for comparison and assessment.

For those interested in more granular details of the classification models, please refer to Appendix 11.2.

6 Results

In this chapter we present the outcomes derived from the application of distinct generative models across a variety of datasets, along with an analysis of the impact of synthetic data on the performance of our classification models.

A short performance improvement was achieved exclusively in the Chongqing dataset. This success was realized by utilizing Stable Diffusion *v1* with DreamBooth and single-class fine-tuning, as detailed in Table 6.1.

Contrarily, the experiments conducted on the Human Brain MRI Dataset did not yield improvements. As shown in Table 6.2, the overall results reported either equal or worse metrics compared to the original data. This observation holds true across the entirety of our experimentation.

A similar outcome was observed with the Diabetic Retinopathy dataset, where none of the experiments led to improved results, as exhibited in Table 6.3.

Some possible reasons behind these varied outcomes, and the specific success with the Chongqing dataset, will be interpreted and discussed further in the conclusion section.

6.1 Chongqing Dataset

	Model Performance			
	Accuracy	Precision	Recall	F1-score
Original Dataset (OD)	0.90	0.87	0.80	0.83
OD + Conventional DA	0.89	0.84	0.81	0.83
GANS				
PGGAN	0.90	0.87	0.80	0.83
PGGAN with NIH Fine-Tuning	0.90	0.87	0.80	0.83
SD v1 with noise:				
0.05 24 steps	0.90	0.90	0.78	0.82
0.10 24 steps	0.89	0.87	0.77	0.81
0.15 24 steps	0.89	0.86	0.79	0.81
0.20 24 steps	0.89	0.83	0.79	0.81
0.05 50 steps	0.89	0.86	0.79	0.81
SD v2 with noise:				
0.05 24 steps	0.90	0.88	0.79	0.82
0.10 24 steps	0.86	0.79	0.79	0.79
0.15 24 steps	0.84	0.74	0.76	0.75
0.20 24 steps	0.84	0.75	0.79	0.76
0.05 50 steps	0.87	0.80	0.81	0.80
SD v1 Dreambooth				
SD with Single-Class Fine-Tuning	**0.91**	**0.89**	**0.81**	**0.84**
SD with Multi-Class Fine-Tuning	0.88	0.84	0.75	0.78
SD v2 Dreambooth finetuned				
SD LoRa	0.86	0.78	0.82	0.79

Table 6.1: Chongqing Dataset Results

6.2 Human Brain MRI Dataset

	Model Performance			
	Accuracy	Precision	Recall	F1-score
Original Dataset	0.9	0.89	0.88	0.88
SD v1 Dreambooth				
SD with Glioma Fine-Tuning	0.89	0.88	0.88	0.88
SD v1 Dreambooth Clustering				
SD + CLIP with Glioma Fine-Tuning	0.89	0.88	0.88	0.88
SD + EfficientNet with Glioma Fine-Tuning	0.89	0.88	0.88	0.88

Table 6.2: Human Brain MRI Dataset Results

6.3 Diabetic Retinopathy Dataset

	Model Performance			
	Accuracy	Precision	Recall	F1-score
Original Dataset	0.96	0.95	0.89	0.92
SD v1 Dreambooth				
SD with Mild Fine-Tuning	0.95	0.95	0.89	0.92
SD v1 Dreambooth Clustering				
SD + EfficientNet with Mild Fine-Tuning	0.95	0.95	0.89	0.91
SD + CLIP with Mild Fine-Tuning	0.95	0.95	0.89	0.91

Table 6.3: Diabetic Retinopathy Dataset Results

7 Contributions to public Repositories

This project embarked on a secondary mission to support existing public repositories by contributing valuable resources and tools that benefit the broader community. This goal was successfully met, as evidenced by the enhancements we made to some key repositories.

7.1 pro_gan_pytorch repository

We added new functionalities to one of the most used non-official PyTorch implementations [37] of the PGGAN paper [36], making them available for the community. Our approved and merged pull request contributes to repository enhancement and can be found on GitHub: `github.com/akanimax/pro_gan_pytorch/pull/69`.

7.2 diffusers repository

We made a few enhancements to this leading Python library for SOTA pretrained diffusion models [57], improving its functionalities and code quality. Our approved and merged pull request can be found on GitHub: `github.com/huggingface/diffusers/pull/3807`.

7.3 data_augmentation_using_synthetic images repository

In an endeavor to contribute to the community, we developed and maintained a repository, housing the work done throughout this project. The repository can be found at GitHub: `github.com/fededemo/data_augmentation_using_synthetic_images`

8 Conclusions

Our research set out to investigate the use of advanced AI methods for generating synthetic data, with a particular focus on fields such as medicine, where a shortage of data can hinder the effectiveness of AI models. In partnership with Cogniflow, our goal was to enhance the performance of various image classification models by employing generative techniques to tackle issues related to insufficient training data and imbalanced classes.

Our findings revealed the complexity and diversity of challenges in applying generative models to different datasets. Synthetic data has immense promise for AI and analytics, but our study found that it is challenging to achieve satisfactory data augmentation without fully fine-tuning or retraining the models. Not doing so is essential for our objective of using them in production, as it allows automation for more generalized application and scalability. However, this approach presented difficulties in most analyzed datasets, where concepts are barely known by the techniques studied. The models' original weights, possibly trained on common datasets like ImageNet or others with similar concepts, may not readily adapt to the unique characteristics of specialized medical datasets.

One notable exception was the Chongqing dataset, where a small improvement was achieved. This dataset is characterized by its simplicity, with all the images bearing a strong similarity, in particular including the same chinese character in all the X-rays. As detailed earlier in our study, this uniformity is not common in the other datasets we worked with, and it likely contributed to the distinct results obtained with it.

The project's journey was not without challenges, reflecting the complexity of the field. These challenges included the intricate nature of generative models, the diversity of data across different datasets, the substantial computing resources required, the potential for bias in both the data and the models, and not insignificantly, the constant emergence of new findings, models, and techniques — almost on a weekly basis — throughout the course of our work. Despite these obstacle, the insights we have gathered help to highlight the complex relationship between synthetic data and its use in the real world applications, emphasizing the need for careful consideration, potential fine-tuning, and continued research in this area.

In summary, this research underscores the significance of considering dataset characteristics and the limitations of generative techniques without proper adjustment. The potential to transform the development and training of ML models with synthetic data in specialized datasets and domains is yet to be proved, however this work gives additional basis for ongoing exploration and progress in this rapidly evolving field.

9 Acknowledgments

We would like to express our appreciation to our advisor, Franz Mayr, for his guidance, wisdom, and patience throughout this project. His support and insights have been instrumental in bringing this work to fruition.

Furthermore, we would like to extend our gratitude to Waldemar Lopez, CTO of Cogniflow, for providing us with the necessary resources and for his technical assistance that enhanced our research.

Special thanks to Dr. Micaela Mandacen for her valuable time and expertise in evaluating the pneumoconiosis datasets. Her evaluation significantly enriched our research.

We would also like to take a moment to appreciate the love and understanding from our families, friends and loved ones, who have been a constant source of strength and support throughout this demanding process.

Lastly, we would like to acknowledge the assistance of both, the GPT-4 from OpenAI [58] and Bard from Google [59] for the help with wording, formatting, and styling throughout this work.

10 Bibliography

[1] S. Decherchi, E. Pedrini, M. Mordenti, A. Cavalli, and L. Sangiorgi, "Opportunities and challenges for machine learning in rare diseases," *Frontiers in Medicine*, vol. 8, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmed.2021.747612

[2] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, vol. 137, p. 109347, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320323000481

[3] Gartner, "Maverick research: Forget about your real data — synthetic data is the future of ai," Jun. 2021. [Online]. Available: https://www.gartner.com/en/documents/4002912 Accessed on: March, 01, 2023.

[4] WSJ, "Fake it to make it: Companies beef up ai models with synthetic data," Jul. 2021. [Online]. Available: https://www.wsj.com/articles/fake-it-to-make-it-companies-beef-up-ai-models-with-synthetic-data-11627032601 Accessed on: Mar., 01, 2023.

[5] GretelLabs, "Gretel: The synthetic data platform for developers," Jan. 2019. [Online]. Available: https://www.gretel.ai Accesed on: Mar., 01, 2023.

[6] W. H. L. Pinaya, M. S. Graham, E. Kerfoot, P.-D. Tudosiu, J. Dafflon, V. Fernandez, P. Sanchez, J. Wolleb, P. F. da Costa, A. Patel, H. Chung, C. Zhao, W. Peng, Z. Liu, X. Mei, O. Lucena, J. C. Ye, S. A. Tsaftaris, P. Dogra, A. Feng, M. Modat, P. Nachev, S. Ourselin, and M. J. Cardoso, "Generative ai for medical imaging: extending the monai framework," Jul. 2023. [Online]. Available: https://arxiv.org/abs/2307.15208

[7] Cogniflow, "Cogniflow: No-code ai platform," Mar. 2022. [Online]. Available: https://www.cogniflow.ai Accesed on: Mar., 01, 2023.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[9] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," Dec. 2017. [Online]. Available: https://arxiv.org/abs/1712.04621

[10] A. Karpathy, P. Abbeel, G. Brockman, P. Chen, V. Cheung, Y. Duan, I. Goodfellow, D. Kingma, J. Ho, R. Houthooft, T. Salimans, J. Schulman, I. Sutskever, and W. Zaremba, "Generative models," Jun. 2016. [Online]. Available: https://openai.com/research/generative-models Accesed on: Mar., 01, 2023.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," *CoRR*, vol. abs/1406.2661, 2014. [Online]. Available: https://arxiv.org/abs/1406.2661

[12] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, pp. 1–48, 2019.

[13] S. Sundaram and N. Hulkund, "Gan-based data augmentation for chest x-ray classification," *CoRR*, vol. abs/2107.02970, Jul. 2021. [Online]. Available: https://arxiv.org/abs/2107.02970

[14] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

[15] R. He, S. Sun, X. Yu, C. Xue, W. Zhang, P. Torr, S. Bai, and X. QI, "IS SYNTHETIC DATA FROM GENERATIVE MODELS READY FOR IMAGE RECOGNITION?" in *The Eleventh International Conference on Learning Representations*, Kigali, Rwanda, 2023. [Online]. Available: https://openreview.net/forum?id=nUmCcZ5RKF

[16] B. Trabucco, K. Doherty, M. Gurinas, and R. Salakhutdinov, "Effective data augmentation with diffusion models," Feb. 2023. [Online]. Available: https://arxiv.org/abs/2302.07944

[17] S. Azizi, S. Kornblith, C. Saharia, M. Norouzi, and D. J. Fleet, "Synthetic data from diffusion models improves imagenet classification," Apr. 2023. [Online]. Available: https://arxiv.org/abs/2304.08466

[18] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on CVPR*, Vancouver, Canada, June 2022, pp. 10 684–10 695.

[19] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," Apr. 2022. [Online]. Available: https://arxiv.org/abs/2204.06125

[20] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, "Photorealistic text-to-image diffusion models with deep language understanding," May. 2022. [Online]. Available: https://arxiv.org/abs/2205.11487

[21] Y. Balaji, S. Nah, X. Huang, A. Vahdat, J. Song, Q. Zhang, K. Kreis, M. Aittala, T. Aila, S. Laine, B. Catanzaro, T. Karras, and M.-Y. Liu, "ediff-i: Text-to-image diffusion models with an ensemble of expert denoisers," 2023. [Online]. Available: https://arxiv.org/abs/2211.01324

[22] A. Q. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. Mcgrew, I. Sutskever, and M. Chen, "GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models," in *Proceedings of the 39th International*

Conference on Machine Learning, ser. Proceedings of Machine Learning Research, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds., vol. 162. PMLR, 17–23 Jul 2022, pp. 16 784–16 804. [Online]. Available: https://proceedings.mlr.press/v162/nichol22a.html

[23] J. Oppenländer, "The creativity of text-to-image generation," Jun. 2022. [Online]. Available: https://arxiv.org/abs/2206.02904

[24] L. Yu, B. Shi, and A. Aghajanyan, "Scaling autoregressive multi-modal models: Pretraining and instruction tuning," https://ai.meta.com/research/publications/scaling-autoregressive-multi-modal-models-pretraining-and-instruction-tuning/, 2023.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248–255.

[26] P. Chambon, C. Bluethgen, J.-B. Delbrouck, R. Van der Sluijs, M. Połacin, J. M. Z. Chaves, T. M. Abraham, S. Purohit, C. P. Langlotz, and A. Chaudhari, "Roentgen: Vision-language foundation model for chest x-ray generation," Nov. 2022. [Online]. Available: https://arxiv.org/abs/2211.12737

[27] CloudGPUs, "Cloud gpus," Jan. 2023. [Online]. Available: https://cloud-gpus.com Accessed on: Mar., 01, 2023.

[28] C. Hao, N. Jin, C. Qiu, K. Ba, X. Wang, H. Zhang, Q. Zhao, and B. Huang, "Balanced convolutional neural networks for pneumoconiosis detection," *International Journal of Environmental Research and Public Health*, vol. 18, no. 17, p. 9091, 2021.

[29] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jul. 2017. [Online]. Available: https://doi.org/10.1109%2Fcvpr.2017.369

[30] Kaggle, "Brain tumor mri dataset," Jan. 2021. [Online]. Available: https://www.kaggle.com/dsv/2645886 Accesed Mar., 01, 2023.

[31] Kaggle, "Diabetic retinopathy 224x224 gaussian filtered," Jan. 2020. [Online]. Available: https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered Accesed on: Mar., 01, 2023.

[32] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[33] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable

visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 8748–8763. [Online]. Available: https://proceedings.mlr.press/v139/radford21a.html

[34] S. Vaze, N. Carion, and I. Misra, "Genecis: A benchmark for general conditional image similarity," in *CVPR*, 2023.

[35] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with gpus," Feb. 2017. [Online]. Available: https://arxiv.org/abs/1702.08734

[36] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=Hk99zCeAb

[37] GitHub, "pro_gan_pytorch," Jan. 2020. [Online]. Available: https://github.com/akanimax/pro_gan_pytorch Accesed Mar., 01, 2023.

[38] HuggingFace, "Image-to-image generation," Jun. 2023. [Online]. Available: https://huggingface.co/docs/diffusers/api/pipelines/stable_diffusion/img2img Accesed on: Jun., 01, 2023.

[39] HuggingFace, "Text-to-image generation," May. 2023. [Online]. Available: https://huggingface.co/docs/diffusers/api/pipelines/stable_diffusion/text2img Accesed on: Jun., 01, 2023.

[40] R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. H. Bermano, G. Chechik, and D. Cohen-or, "An image is worth one word: Personalizing text-to-image generation using textual inversion," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=NAQvF08TcyG

[41] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 22 500–22 510.

[42] HuggingFace, "Training stable diffusion with dreambooth using diffusers," Nov. 2022. [Online]. Available: https://huggingface.co/blog/dreambooth Accessed on: Mar., 01, 2023.

[43] E. J. Hu, yelong shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=nZeVKeeFYf9

[44] HuggingFace, "Using lora for efficient stable diffusion fine-tuning," Jan. 2023. [Online]. Available: https://huggingface.co/blog/lora Accessed on: Mar., 01, 2023.

[45] Y. Tewel, R. Gal, G. Chechik, and Y. Atzmon, "Key-locked rank one editing for text-to-image personalization," in *ACM SIGGRAPH 2023 Conference Proceedings*, ser. SIGGRAPH '23, 2023.

[46] N. Ruiz, Y. Li, V. Jampani, W. Wei, T. Hou, Y. Pritch, N. Wadhwa, M. Rubinstein, and K. Aberman, "Hyperdreambooth: Hypernetworks for fast personalization of text-to-image models," Jul. 2023. [Online]. Available: https://arxiv.org/abs/2307.06949

[47] GitHub, "Hyperkohaku," Jul. 2023. [Online]. Available: https://github.com/KohakuBlueleaf/HyperKohaku Accesed Jul., 01, 2023.

[48] HuggingFace, "runwayml/stable-diffusion-v1-5," Jun. 2022. [Online]. Available: https://huggingface.co/runwayml/stable-diffusion-v1-5 Accesed on: Jul., 01, 2023.

[49] RunwayAI, "Runway: Advancing creativity with artificial intelligence." Jan. 2023. [Online]. Available: https://runwayml.com Accesed on: Mar., 01, 2023.

[50] Tryolabs, "The guide to fine-tuning stable diffusion with your own images." Oct. 2022. [Online]. Available: https://tryolabs.com/blog/2022/10/25/the-guide-to-fine-tuning-stable-diffusion-with-your-own-images Accesed on: Jul., 01, 2023.

[51] HuggingFace, "stabilityai/stable-diffusion-2-1," Jun. 2022. [Online]. Available: https://huggingface.co/stabilityai/stable-diffusion-2-1 Accesed on: Jul., 01, 2023.

[52] Stability, "Stability ai: Ai by the people for the people," Jan. 2023. [Online]. Available: https://stability.com Accesed Mar., 01, 2023.

[53] HuggingFace, "stabilityai/stable-diffusion-xl-base-1.0," Jul. 2023. [Online]. Available: https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0 Accesed on: Jul., 15, 2023.

[54] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "SDXL: Improving latent diffusion models for high-resolution image synthesis," Jul. 2023. [Online]. Available: https://arxiv.org/abs/2307.01952

[55] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[56] HuggingFace, "stabilityai/stable-diffusion-xl-refiner-1.0," Jul. 2023. [Online]. Available: https://huggingface.co/stabilityai/stable-diffusion-xl-refiner-1.0 Accessed Jul., 15, 2023.

[57] P. von Platen, S. Patil, A. Lozhkov, P. Cuenca, N. Lambert, K. Rasul, M. Davaadorj, and T. Wolf, "Diffusers: State-of-the-art diffusion models," https://github.com/huggingface/diffusers, 2022.

[58] OpenAI, "Gpt-4 technical report," Mar. 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

[59] J. Manyika, "An overview of bard: an early experiment with generative ai," Mar. 2023. [Online]. Available: https://ai.google/static/documents/google-about-bard.pdf Accessed on: Mar., 21, 2023.

11 Appendix

11.1 Expert Evaluation of Pneumoconiosis Generated Images

In order to gain insights into the realism and quality of our generated images, we solicited the expertise of a medical radiologist. 31 images, were subjected to a thorough examination. The evaluation was structured around four questions:

Question	**Yes**	**No**	**Uncertain/Doubtful**
Do they seem like real X-rays?	31	0	0
Do the images seem to be of a person effectively?	31	0	0
Are there any malformations, strange bones, or anomalies in them?	0	31	0
Is pneumoconiosis identifiable?	4	24	3

Table 11.1: Summary of the radiologist's responses to the posed questions.

11.1.1 Observations and Interpretations

In examining the images, the radiologist noted some recurring patterns that they termed as "errors." The middle thirds of both lungs in the images appear more white ("subexposed") than other areas. This inconsistency could complicate diagnosis due to the uneven distribution of exposure.

The majority of the images show a rightward rotation. This rotation could cause the left hilum to be more prominently visible. Interestingly, in these images, the right hilum is consistently emphasized, a characteristic usually associated with pathological conditions or leftward rotation.

Another noteworthy observation is the overexposure of the lower third of the right side. This could be an optical phenomenon, where the darker appearance of the lower area is due to the lighter ("white") appearance of the areas above.

The radiologist hypothesized that these recurring patterns could potentially be artifacts from the original dataset used in the image generation process.

11.1.2 Conclusion

The expert evaluation reveals that, overall, the generated images exhibit a high level of realism and accurately represent the human anatomy. However, the presence of recurrent patterns, interpreted as "errors," could introduce diagnostic challenges. Among the generated images, the presence of pneumoconiosis was questionable in some instances.

It is critical to acknowledge that these "errors" might be an inherent limitation of the original dataset used to train the model. Despite their overall realism, certain features in the generated images—such as the overemphasized right hilum and the consistent rightward rotation—could potentially mislead non-expert observers.

11.2 Classification Models Technical Details

In this appendix, we present the technical details involved in the creation of the different baseline classification models for the different datasets.

11.2.1 Chongqing Dataset

For pneumoconiosis detection using the Chongqing dataset the following steps were undertaken:

- **Data Preprocessing:** The Chongqing dataset was preprocessed, which included resizing images to 512x512. The dataset, already split into training and test sets, was further subjected to normalization and consistent image orientation. To balance the class distribution, the dataset was augmented as necessary.
- **Model Selection and Training:** With the help of Cogniflow's No-Code AI, we selected the Feed-Forward Neural Network (FFNN) as the classification model. This model, vectorized by InceptionResNetV2, was found to be appropriate for the task. The model took approximately 0 hours, 3 minutes, and 41 seconds to run.
- **Configuration Parameters:** The model was configured as follows:
  - Learning Rate: 0.001
  - Batch Size: 64
  - Max Epochs: 300
  - Optimizer: Adam
  - Kernel Regularizer: L2
  - Kernel Regularizer Value: 0.02

  - Hidden Layers: [256, 256, 256, 256]
  - Activations: ['relu', 'relu', 'relu', 'relu']
  - Batch Normalization: False

- **Data Augmentation:** Due to the limited data in the Chongqing dataset, we applied generative AI techniques to generate synthetic data. The synthetic data served to augment the original dataset, resulting in a more diverse and balanced training dataset for the classification model.

- **Evaluation:** The performance of the classification model was evaluated using the test set from the Chongqing dataset. Macro averages of precision, recall, F1-score, and accuracy were utilized to assess the effectiveness

- **Comparison:** The model trained with synthetic data showed a slight improvement in all metrics.

11.2.2 Human Brain MRI Dataset

For brain tumor detection using the Human Brain MRI dataset the following steps were undertaken:

- **Data Preprocessing:** The Human Brain MRI dataset was preprocessed, which included resizing images to 300x300. The dataset was already split into training and test sets and was further normalized for consistent image quality.

- **Model Selection and Training:** We used Cogniflow's No-Code AI to select the Support Vector Classifier (SVC) with a radial basis function (rbf) kernel. The images were vectorized using the Xception model. This selection was found to be well suited for the task at hand.

- **Configuration Parameters:** The model was configured with the following parameters:
  - Gamma: scale
  - Kernel: rbf
  - C: 10
  - Maximum Iterations: 1000
  - Tolerance: 0.01
  - Random State: 123

- **Data Augmentation:** We applied generative AI techniques to generate synthetic data. The synthetic data served to augment the original dataset, resulting in a more diverse and balanced training dataset for the classification model. This augmentation was performed both for the class with the least instances and across all classes.

- **Evaluation:** The performance of the model was evaluated using the test set from the Human Brain MRI dataset. Macro averages of precision, recall, F1-score, and accuracy were utilized to assess the effectiveness of tumor detection.
- **Comparison:** The model trained with synthetic data did not showed an improvement in the metrics.

11.2.3 Diabetic Retinopathy Gaussian Filtered Dataset

For eye's retinopathy detection using the Diabetic Retinopathy Gaussian Filtered dataset, the following steps were undertaken:

- **Data Preprocessing:** The Diabetic Retinopathy Gaussian Filtered dataset was preprocessed, which involved resizing images to 300x300. The dataset, already split into training and test sets, was further normalized to maintain consistent image quality.
- **Model Selection and Training:** Logistic Regression was chosen as the classification model using Cogniflow's No-Code AI. The images were vectorized using the InceptionV3 model. This model was found to be appropriate for the task, and it took approximately 0 hours, 8 minutes, and 18 seconds to run.
- **Configuration Parameters:** The model was configured with the following parameters:
  - C: 10
  - L1 ratio: 0.5
  - Maximum Iterations: 300
  - Multi Class: ovr
  - Penalty: elasticnet
  - Random State: 123
  - Solver: saga
- **Data Augmentation:** We applied generative AI techniques to generate synthetic data. The synthetic data served to augment the original dataset, resulting in a more diverse and balanced training dataset for the classification model. This augmentation was performed both for the class with the least instances and across all classes.
- **Evaluation:** The performance of the model was evaluated using the test set from the Diabetic Retinopathy Gaussian Filtered dataset. Macro averages of precision, recall, F1-score, and accuracy were utilized to assess the effectiveness of retinopathy detection.
- **Comparison:** The model trained with synthetic data did not showed an improvement in the metrics.

11.3 Discarded techniques results

We present an overview of the outcomes achieved through our experimentation with various fine-tuning approaches. However, these outcomes were not integrated into our work due to their under-performance in generating images when compared to the quality of the original images. These were:

- **Textual inversion** using Stable Diffusion v1.5 (Table 11.2). While the generated images displayed a certain degree of resemblance to those in the dataset, the finetuned model exhibited an inability to faithfully replicate identical outcomes. Instead, it exhibited tendencies to introduce hallucinatory details, potentially stemming from its creative interpretation of the provided prompt.

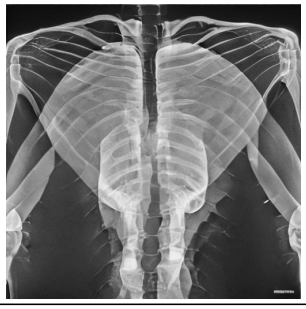
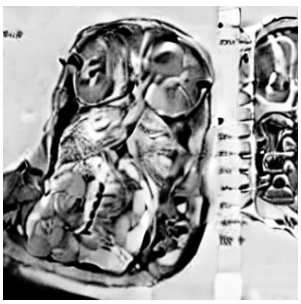
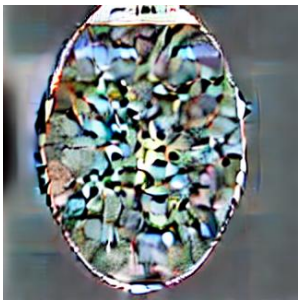
	Image 1	Image 2	Image 3
Pneumoconiosis x-ray			
Human Brain MRI			
Retinopathy			

Table 11.2: Images generated using *Textual Inversion* technique with SD v1.5

- **HyperDreamBooth** using Stable Diffusion v1.5 (Table 11.3). It is evident that this technique would benefit from further refinement, both in terms of time investment for enhanced implementation and empirical assessment given its recent discovery. The images presented in the corresponding table fell notably short of anticipated outcomes, underscoring the necessity for a more effective parameter configuration during the finetuning process.

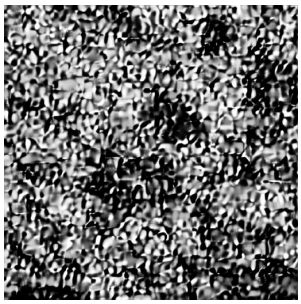
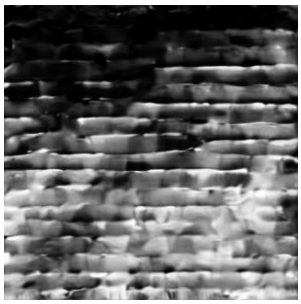
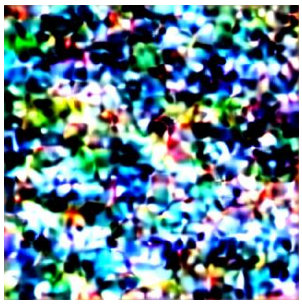
	Image 1	Image 2	Image 3
Pneumoconiosis x-ray			
Human Brain MRI			
Retinopathy			

Table 11.3: Images generated using *HyperDreamBooth* technique with SD v1.5

The results for these techniques exhibited inferior performance when compared to the DreamBooth technique. This was observed both when utilizing out-of-the-box scripts and when employing similar training epochs.

- **Stable Diffusion XL Refiner Model**: As mentioned before we can use a two-stage pipeline when generating images by applying two models however this technique proved to be not appropriate for our purposes as depicted in Table 11.4

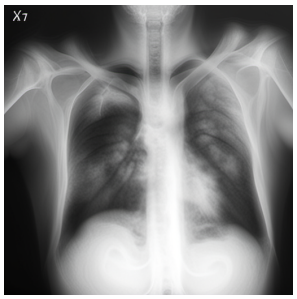
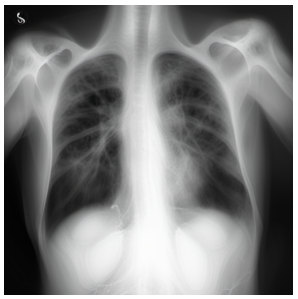
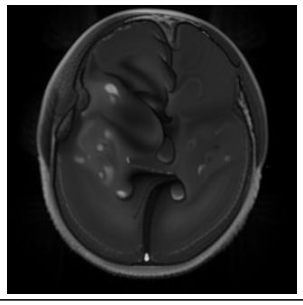
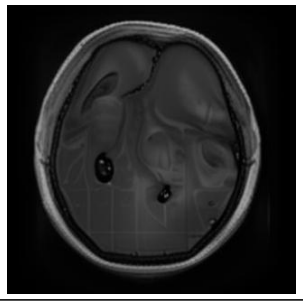
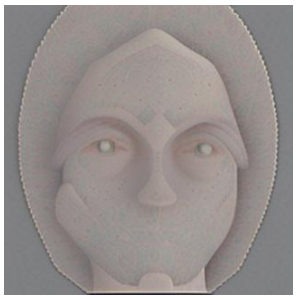
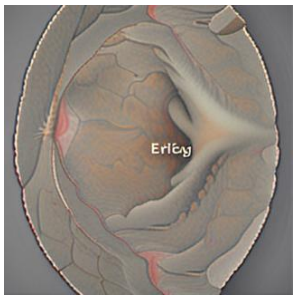
	Image 1	Image 2	Image 3
Pneumoconiosis with Refiner			
Glioma with Refiner			
Retinopathy with Refiner			

Table 11.4: Images with use of Refiner model for Stable Diffusion XL *v1.0*.