Project Report - ECE 285
Diffusion-Based Image Restoration: A Zero-Shot Approach

Sharvari Satish Deshmukh
ECE (Signal & Image Processing)
A59022896

Abstract

There is a need for high-quality images in various applications, such as medical imaging, remote sensing, and surveillance. The problem of image restoration is to recover a high-quality image from a degraded observation. The degradation can be caused by various factors, such as noise, blur, or missing pixels. Traditional image restoration methods often require task-specific training and cannot generalize to different types of degradation. The core part of the problem is to design a generic restoration model that can effectively remove the degradation and recover the original image for variety of degradation. This project is based on one of the solutions provided by the DDNM paper - "Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model" [6] The performance of the proposed method is compared after applying it on pre-trained as well as self- trained diffusion models for an image inpainting task using evaluation metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The dataset used for self-trainig and evaluation is the DREAMING - Diminished Reality for Emerging Applications in Medicine through Inpainting Dataset provided as part of DREAMING 2024 Grand challenge by IEEE ISBI 2024 [2].

1 Introduction

Image inpainting, the process of restoring missing or corrupted parts of an image, remains a significant challenge in the field of image restoration (IR). The ability to accurately reconstruct the missing portions of an image is crucial for various applications, including photo editing, film restoration, and enhancing images degraded by occlusions or artifacts. Traditional inpainting model-based approaches often rely on predefined priors and tend to produce less realistic results. On the other hand, end-to-end learning methods, though capable of generating more realistic images, require large datasets for training and are not easily generalizable across different tasks. The advent of diffusion models has significantly simplified the task of image generation, allowing for easy adaptation to perform image inpainting tasks.

In this project, I integrate Range-Null space Decomposition (RND), a well-known concept in linear algebra, to modify the denoising process of diffusion models for image inpainting tasks. This approach allows us to leverage existing pre-trained foundational diffusion models, avoiding the costly and compute-intensive process of training new diffusion models from scratch.

The dataset on which the method performance is tested is provided as part of DREAMING 2024 Grand challenge by IEEE ISBI 2024. The DREAMING challenge aims to introduce Diminished Reality (DR) into oral and maxillofacial surgery. The dataset consists of synthetic, photorealistic surgery scenes in an operating room, featuring humans and/or objects obstructed by medical tools and/or hands of the surgeon. The target is to seamlessly remove these obstructions, ensuring an unobstructed view of the operative site, which could significantly enhance surgeons' visibility.

The inpainting task challenge posed by this dataset appropriately fits as an use-case for the method proposed by the DDNM paper. This project implements one of the sampling methods proposed by the paper and applies it on both pre-trained and self-trained models for performance comparison. As we'll see in the next sections, the results are comparable with the benchmark set by the paper.

2 Related Work

Generative Adversarial Networks (GANs) were a groundbreaking development in the field of generative models, introduced in 2014. However, GANs had some limitations, such as instability during training and the inability to provide explicit guidance for generating specific types of images. To address these limitations, alternative approaches to generative models were developed, leading to the emergence of Diffusion Models, which include Denoising Diffusion Probabilistic Models (DDPMs) [3] and Denoising Diffusion Implicit Models (DDIMs) [5].

The key idea behind DDPMs is to start with a data sample (e.g., an image) and gradually add noise to it through a diffusion process, ultimately converting it into pure noise. During training, the model learns to reverse this diffusion process, starting from pure noise and iteratively removing the noise to reconstruct the original data sample. DDPMs are easier to train compared to GANs and can generate diverse and high-quality samples, as they learn the entire data distribution rather than just the high-density regions like GANs. It was shown that diffusion models can do better than GANs with class coverage, image quality, and stability.[1]

DDIMs, introduced in 2021, are a variant of DDPMs that aim to improve the sampling efficiency during the generation process. While DDPMs require iteratively sampling from the learned reverse diffusion process, which can be computationally expensive, DDIMs use a more efficient sampling strategy. DDIMs leverage the fact that the denoising process can be approximated as a deterministic function of the noisy input and the learned model parameters. This allows for faster sampling by directly computing the denoised output in a single step, rather than iteratively sampling from the reverse diffusion process. Despite the faster sampling, DDIMs can still generate high-quality samples comparable to or better than DDPMs.

The solution provided by the DDNM paper, on which this project is based on, builds on top of the DDIM sampling. It was chosen as the tentative method for the inpainting task because it claims advantages over existing image restoration methods by being Zero-Shot i.e. it does not require any task-specific training data or optimization, making it applicable to a wide range of image restoration tasks, including super-resolution, denoising, deblurring, inpainting, and colorization. By refining only the null-space contents during the reverse diffusion sampling, it only requires an off-the-shelf diffusion model to yield realistic and data-consistent results. It also claims to be robust to noise levels, degradation types, and image content, making it suitable for a variety of practical applications.

3 Method

The project leverages the DDNM+ (Denoising Diffusion Null-Space Model +) technique for image restoration provided by the paper. It works by refining the null-space components of a degraded image during the reverse diffusion process. This process effectively reconstructs missing or corrupted image data, making it particularly suitable for inpainting tasks. DDNM+ builds upon the foundation of diffusion models, which operate through a two-step process, forward process that gradually introduces noise into an image and reverse process which recovers the original image from the noisy version. DDNM+ also incorporates a time-travel trick to improve the harmony between restored image parts, enhancing the overall quality. To avoid inferior realness caused when inpainting with a large mask, this method produces a better past, which in turn produces a better future.

Range-Null Space Decomposition: Central to DDNM+ is the concept of Range-Null space decomposition. This technique deconstructs an image into two components:

- Range-Space: Contains information directly related to the observed, degraded image.
- Null-Space: Holds the missing information lost during degradation.
- DDNM+ excels by precisely restoring the range-space component based on the degraded input. It then utilizes the diffusion model to generate plausible null-space content, effectively filling in the missing information.

DDNM+ Implementation Details:

- There are two versions of DDNM suggested in the paper. One is SVD-based version, which is precise in solving noisy tasks and a simplified version, which is flexible for users to define their own degradations. I implemented the simplified version as is enough for any inpainting task.
- As mentioned in [7] This includes finding the variant of $x_{0|t}$ in the sampling algorithm, and modifying the sampling of x_{t-1} using following equations:

$$\gamma_t = \sigma_t^2 - (a_t \lambda_t \sigma_y)^2 \tag{1}$$
$$\text{where } \lambda_t = 1, \text{if } \sigma_t \geq a_t \sigma_y \tag{2}$$
$$\text{and } \lambda_t = \sigma_t/(a_t \sigma_y), \text{if } \sigma_t < a_t \sigma_y \tag{3}$$
$$\hat{x_{0|t}} = x_{0|t} - \Sigma_t A^{\dagger}(Ax_{0|t} - y) \tag{4}$$

- Detailed sampling algorithm and the time travel trick as adopted from the paper is represented in Figure 1

Algorithm 1 Sampling of DDNM

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, ..., 1$ **do**
  $\mathbf{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t - \mathcal{Z}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\sqrt{1-\bar{\alpha}_t}\right)$
  $\hat{\mathbf{x}}_{0|t} = \mathbf{A}^{\dagger}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{x}_{0|t}$
  $\mathbf{x}_{t-1} \sim p(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{x}}_{0|t})$
return $\mathbf{x}_0$

Algorithm 2 Sampling of DDNM^{+}

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, ..., 1$ **do**
  $L = \min\{T - t, l\}$
  $\mathbf{x}_{t+L} \sim q(\mathbf{x}_{t+L}|\mathbf{x}_t)$
  for $j = L, ..., 0$ **do**
    $\mathbf{x}_{0|t+j} = \frac{1}{\sqrt{\bar{\alpha}_{t+j}}}\left(\mathbf{x}_{t+j} - \mathcal{Z}_{\boldsymbol{\theta}}(\mathbf{x}_{t+j}, t+j)\sqrt{1-\bar{\alpha}_{t+j}}\right)$
    $\hat{\mathbf{x}}_{0|t+j} = \mathbf{x}_{0|t+j} - \boldsymbol{\Sigma}_{t+j}\mathbf{A}^{\dagger}(\mathbf{A}\mathbf{x}_{0|t+j} - \mathbf{y})$
    $\mathbf{x}_{t+j-1} \sim \hat{p}(\mathbf{x}_{t+j-1}|\mathbf{x}_{t+j}, \hat{\mathbf{x}}_{0|t+j})$
return $\mathbf{x}_0$

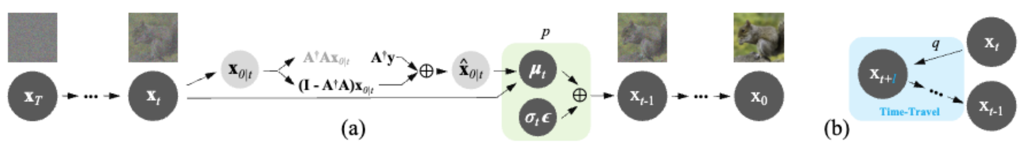


Figure 1: (a)DDNM & DDNM+ Sampling (b) Time travel trick [6]

Training and Testing Algorithms:

- Pre-trained Model: In this project, I employed a basic CelebA-HQ model provided by SDEdit [4] as the foundation for DDNM+ sampling.
- Self-Trained Model: Since the CelebA-HQ model is trained specifically on celebrity images, I trained the model with the DREAMING dataset before employing DDNM+ and compared the performance with pre-trained Model. The training involved adding Gaussain noise with variance schedule in the forward process and predicting the noise added while using MSE Loss as the denoising loss.
- Testing: During sampling, the masked image and the corresponding mask are provided as input. The algorithm iteratively recovers the image following the DDNM+ sampling method.

4 Experiments

Dataset: Experiments were conducted using the IEEE Dreaming dataset, which consists of realistic surgical scenes, chosen for its relevance to the inpainting task. The dataset includes:

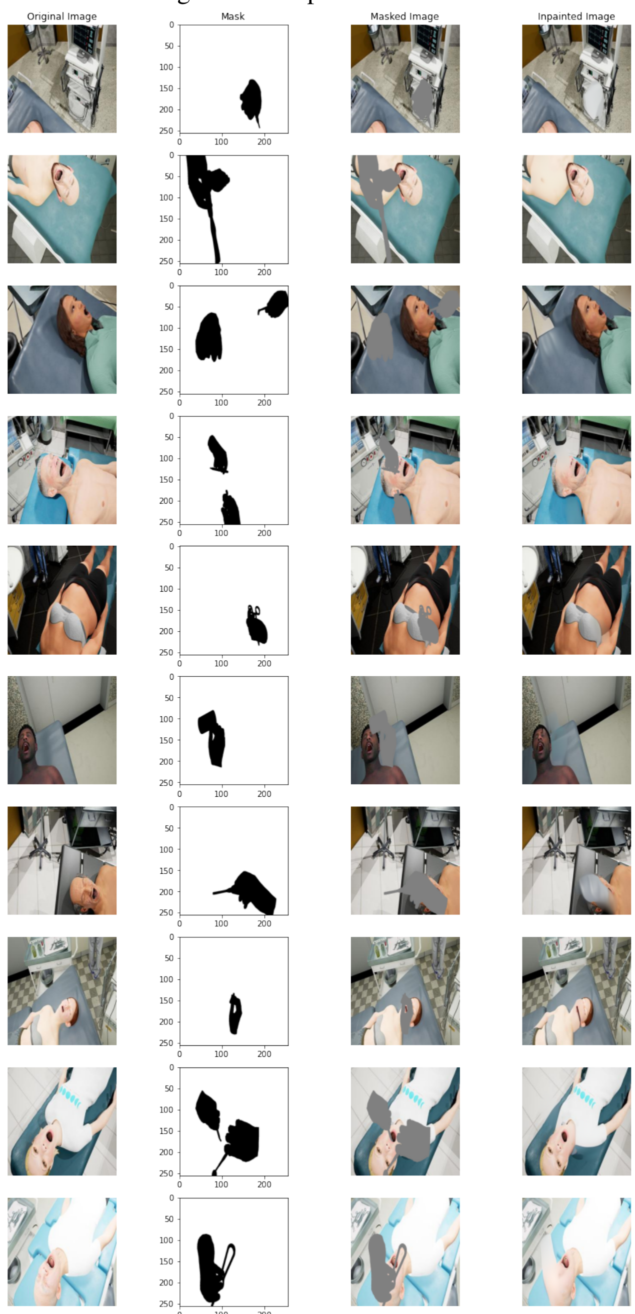
- Occluded RGB images
- Ground truth RGB images
- Binary masks defining inpainting areas

- Camera pose and intrinsic data

The dataset is available in 10 sets, each with size $\sim$ 2GB. Only the images from scenes 0001-0006 of part-1 of the dataset were used for training the model (total $\sim$ 5000 images) for max of 15 epochs, constrained by the compute resource and time available. Randomly sampled images from scene 0007-0009 were used for testing the performance of the method in terms of evaluation metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

Results: The proposed diffusion-based image restoration approach achieved a comparable performance on the test data for the inpainting task. Figure 2 displays the set of sample images used for testing the inpainting performance of different models.

Figure 2: Sampled Test Dataset



The figures Figure 3 Figure 4 Figure 5 display the qualitative outputs obtained by applying DDNM+ on pretrained CelebA-HQ model and self-trained models (with epochs 4, 8, 12 and 15) (row-wise) tuned using best training and sampling hyperparameters including learning rate, batch size, betas, sigmas etc. (Corresponding average PSNR and SSIM values can be found in the code zip submitted.) We can see the qualitative improvements after using self-trained models and with increasing epochs but it is not consistent across different kinds of occlusions. So, there's a future scope for improvements in the sampling algorithm and model architecture along with training method.

Figure 3: Qualitative results for mask covering face, body and objects

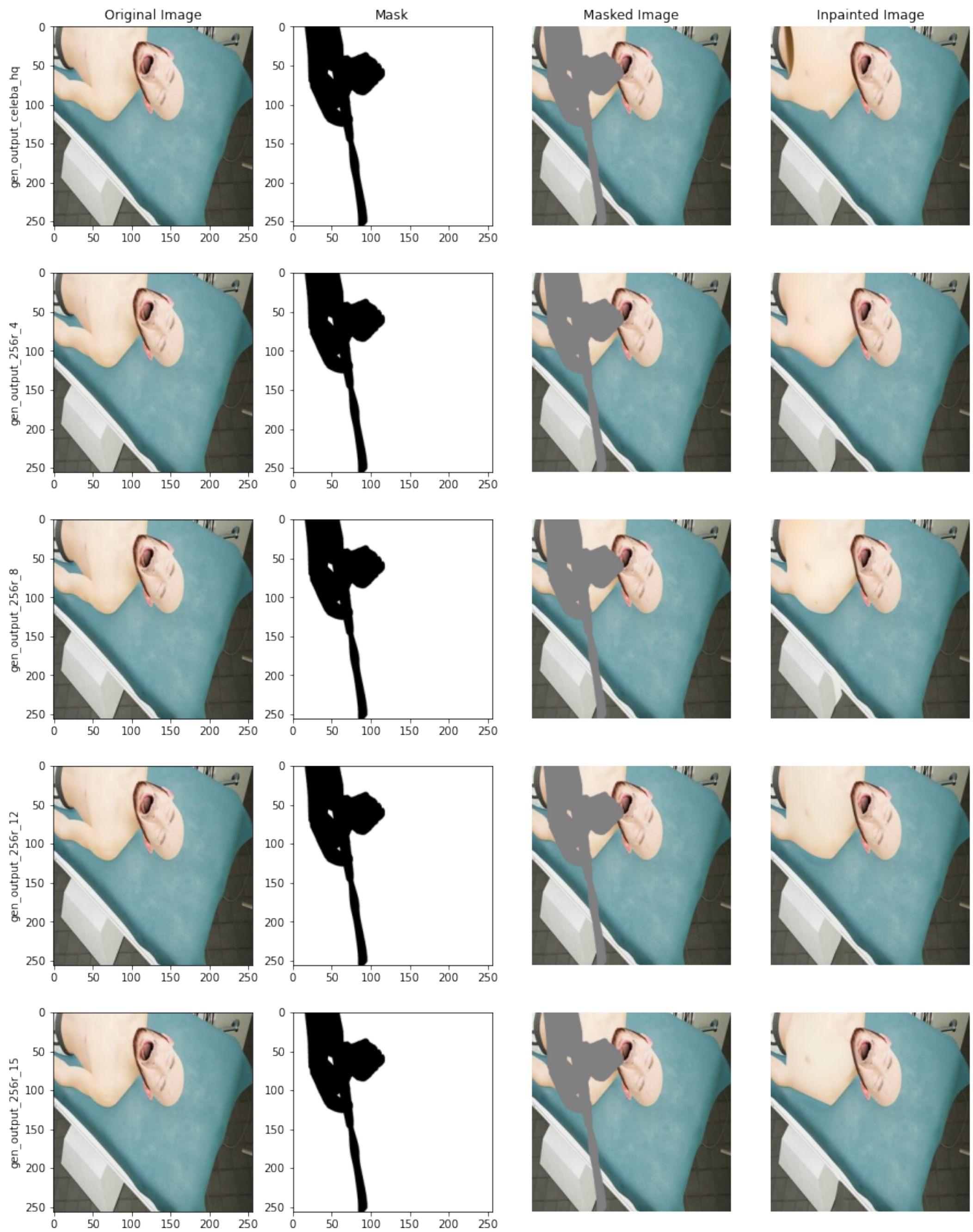


Figure 4: Qualitative results for mask covering face and objects

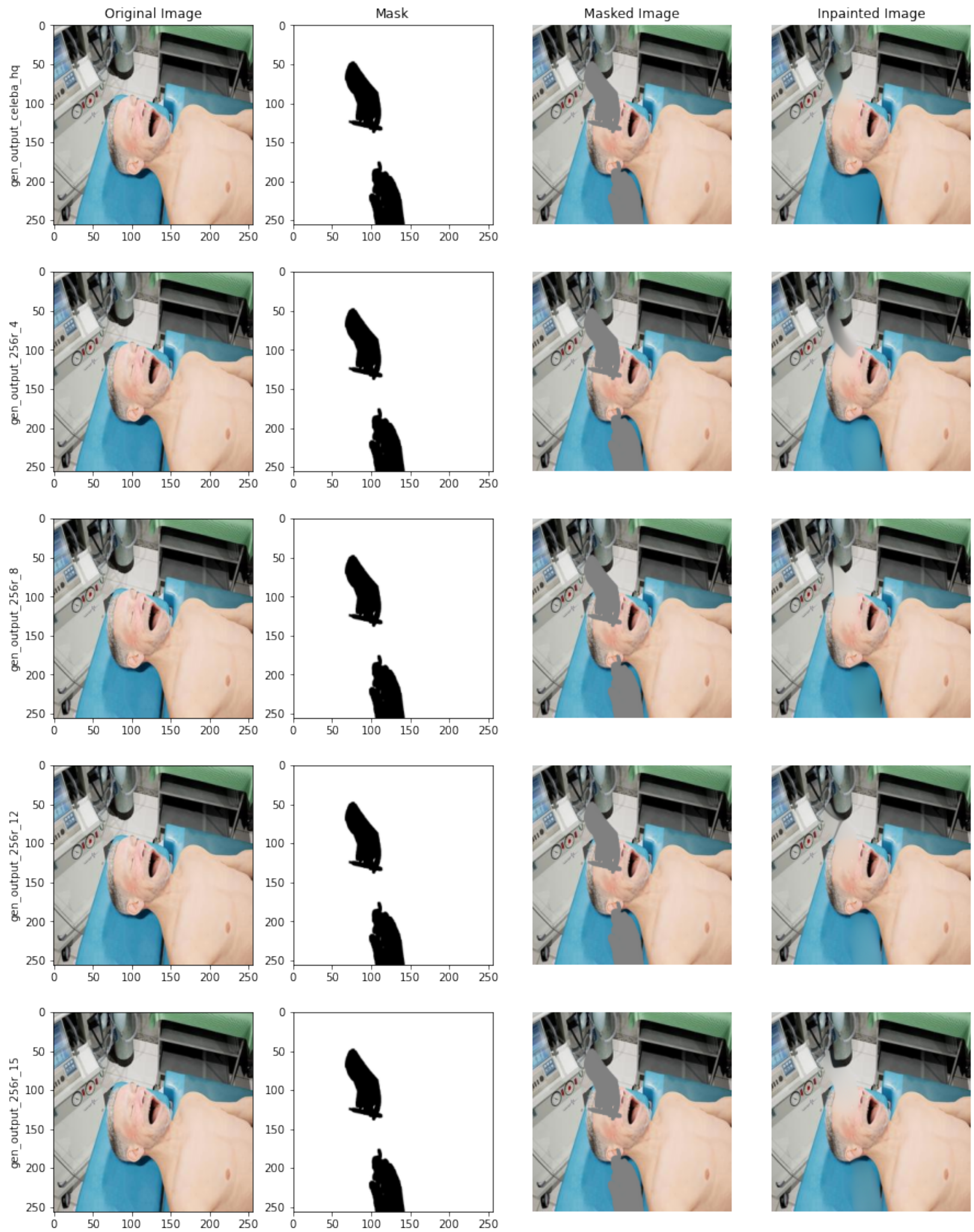
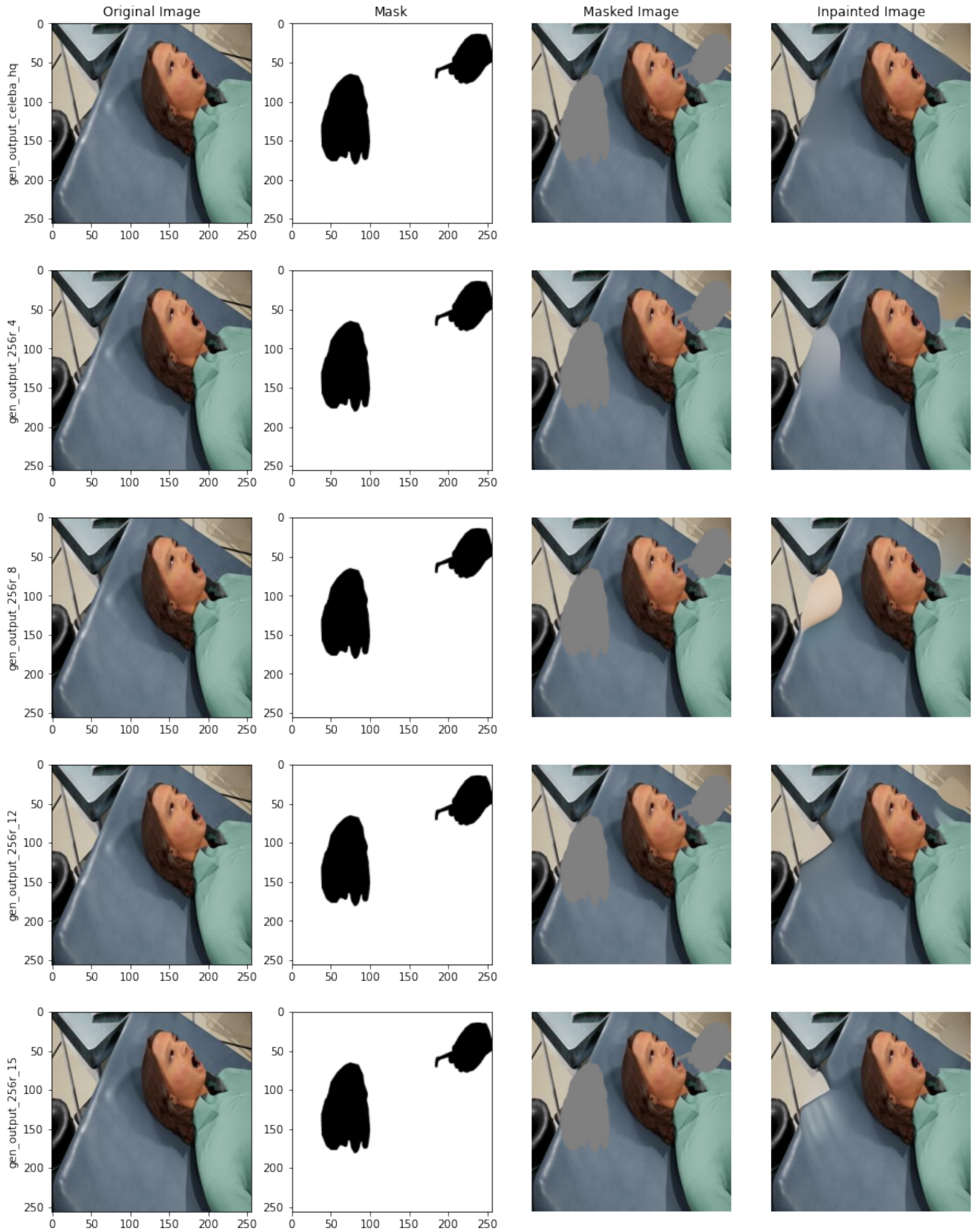


Figure 5: Qualitative results for mask covering mostly objects



The table 1 displays the quantitative performance on the inpainting task obtained by applying DDNM+ on pretrained CelebA-HQ model and trained models (with epochs 4, 8, 12 and 15). (TM - e(i) stands for self-trained model with i epochs)

$Inpainting$	$Celaba-HQ$	$TM-e4$	$TM-e8$	$TM-e12$	$TM-e15$	$DDNM-Benchmark$
PSNR	28.53	28.64	28.31	29.28	28.51	35.64
SSIM	0.96±0.02	0.96±0.01	0.96±0.02	0.96±0.02	0.96±0.02	0.982

Table 1: Quantitative results of DDNM+ on multiple models

5 Implementation

Code Implementation is divided into two files - diffusion_training.ipynb and diffusion_sampling.ipynb. The diffusion model architecture implemented is based on/ inspired by the model mentioned in the DDNM paper and the implementation by SDEdit [4]. I implemented a simple training agorithm for training the model on DREAMING dataset. The diffusion sampling algorithm is based on the DDNM+ sampling algorithm mentioned in the DDNM paper. I added evaluation metrics as well as visualisations for the dataset. The original implementation by the paper is quite complex and utilises multiple models, optimization schemes and hyperparameters for optimum resuls. I simplified the same and implemented a minimised version for the inpainting task which this project is focused on. Also, as the sampling method mentioned in the DDNM paper relies on not needing to train existing foundational models, it also utilises complex and large pre-trained models to achieve this. But as per the results received on the relatively simple model I used and the feedback on the project proposal from TA, I also trained the model on the DREAMING dataset to check for any performance improvements.

The ouput for the sampled 10 images are present in the inpainted_ouputs folder while the comparative improvements for each image with different models and epochs can be found in the compare_output folder.

References

[1] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[2] Christina Gsaxner, Timo van Meegdenburg, Gijs Luijten, Viet Duc Vu, Behrus Puladi, and Jan Egger. DREAMING - Diminished Reality for Emerging Applications in Medicine through Inpainting Dataset, Jan. 2024.

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[4] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022.

[5] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[6] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *The Eleventh International Conference on Learning Representations*, 2023.

[7] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *The Eleventh International Conference on Learning Representations*, 2023.