YACHAY TECH UNIVERSITY

"Comparative Analysis of Two Approaches for Image Regeneration: U-Net and VAEGAN"

Ricardo Farinango
Wilter Menoscal
Anthony Salazar

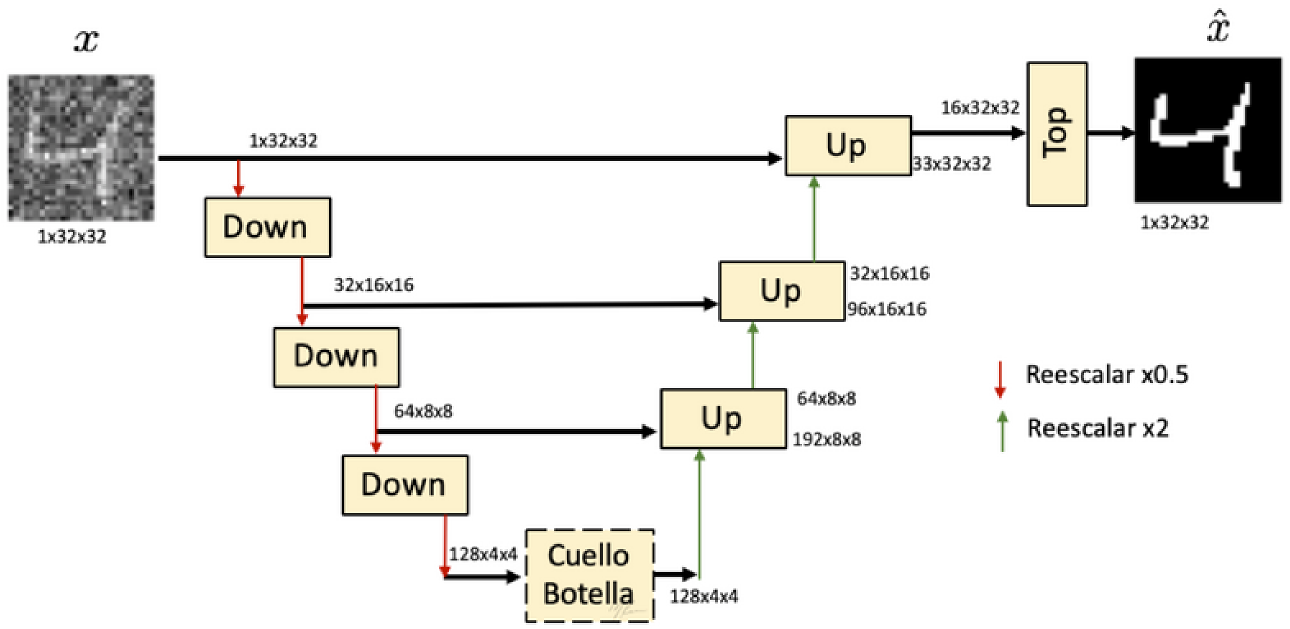
Urcuquí, August 03 2023

School of Mathematical and Computational Sciences

TABLE OF CONTENTS

INTRODUCTION

U_NET ARCHITECTURE



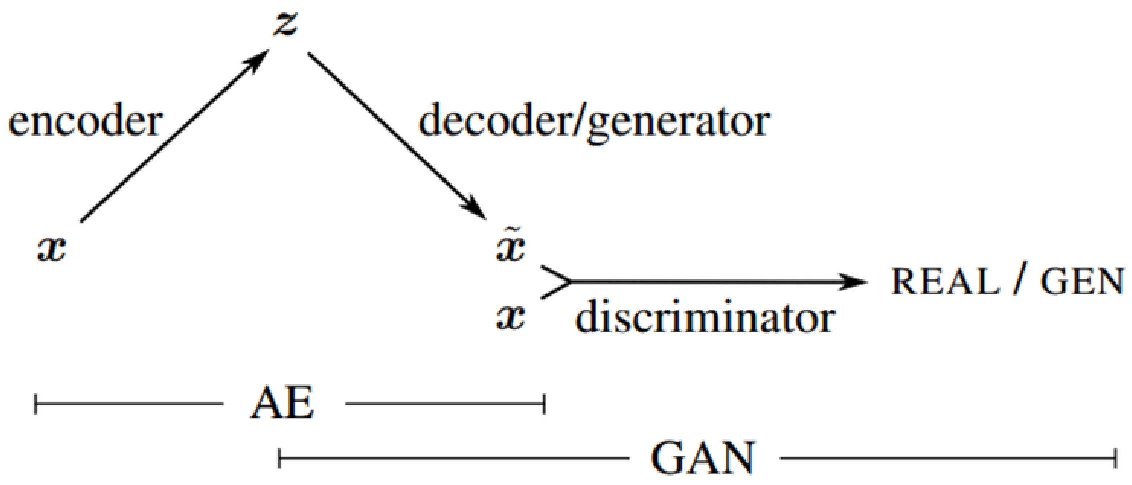
Encoder-Decoder with Skip Connections
U-Net is a convolutional network that uses an encoder-decoder structure with skip connections to reconstruct images or perform semantic segmentation.

Segmentation and Inpainting
It is commonly used to segment images, assigning labels to each pixel, and also for *inpainting*, which consists of filling in missing or damaged regions in an image.

U-shape
It is named after its "U" shape in the network diagram, where features are reduced in one path and then reconstructed in another.

Overfitting Reduction
Hop connections allow detailed reconstruction and reduce overfitting by combining information from different levels of features.

VAEGAN



VAE (Variational Autoencoder)
It is a type of neural network that learns to represent the input data in a continuous latent space. It uses a combination of encoder and decoder to map the original data into the latent space and then reconstruct it.

GAN (Generative Adversarial Network)
This is another type of neural network consisting of a generator and a discriminator. The generator creates data samples (e.g. images) from random noise, while the discriminator tries to distinguish between real and generated samples.

Combination of VAE and GAN
In VAEGAN, the VAE architecture is combined with the GAN architecture. VAE is used to learn the distribution of data in the latent space, while GAN is used to generate more realistic samples from the latent space.

Benefits
The combination of VAE and GAN improves the quality and diversity of the generated images.

OBJETIVES

General Objetive

Develop a system for regeneration of damaged or missing images using a combination of U-Net and VAEGAN architecture techniques.

01

Implement a generative network using the U-Net architecture to perform the inpainting process on missing images.

02

Integrate the VAEGAN architecture, which combines a Variational Autoencoder (VAE) with a Generative Adversarial Network (GAN), to improve the quality and diversity of the generated images.

03

Train the model using a dataset suitable for image regeneration, optimizing the hyperparameters to obtain optimal results.

METHODS

Model 1

Dataset

- The CelebA dataset is a large-scale face attributes dataset.
- It contains over 200,000 celebrity images, each with 40 attribute annotations.
- In our case, the images are loaded, resized to 64x64 for compatibility with our model, and normalized so that pixel values are in the range [0, 1].
- This Dataset will be used for both our models
- For this model we use 101300 images from the dataset

Building the U-Net Model

- The U-Net model is a type of Convolutional Neural Network (CNN) that is optimized for tasks like image inpainting.
- The model has an encoder path to extract context (low level features) and a decoder path to enable precise localization (high level features).

Constructing the U-Net Model

- Encoder Path: Uses Conv2D and MaxPooling2D layers to extract image features and reduce spatial dimensions. It employs a 'same' padding strategy and ReLU activation.
- Additional Layers: An extra Conv2D layer enhances the model's performance by learning more complex representations.
- Decoder Path: Utilizes UpSampling2D and Conv2D layers to expand spatial dimensions and reconstruct the original image. Feature maps from the encoder path are concatenated to provide localization information.
- Final Layers: The last Conv2D layer outputs the inpainted image, using 3 filters for RGB channels and a sigmoid activation to keep pixel values in the range [0, 1].
- This design allows the model to understand image context and accurately fill in missing parts.

Compiling the Model

- After defining the model, we compile it using the Adam optimizer and the Mean Squared Error (MSE) as the loss function.
- The optimizer improves the model based on the loss function, which measures the difference between the model's predictions and the actual values.

Training the Model

- The model is trained for 4000 epochs with a batch size of 128.
- During each epoch, a random batch of images is selected, and a blank region is created in each image.
- The model then learns to fill in the blank region in a plausible way, effectively learning to "inpaint" images.

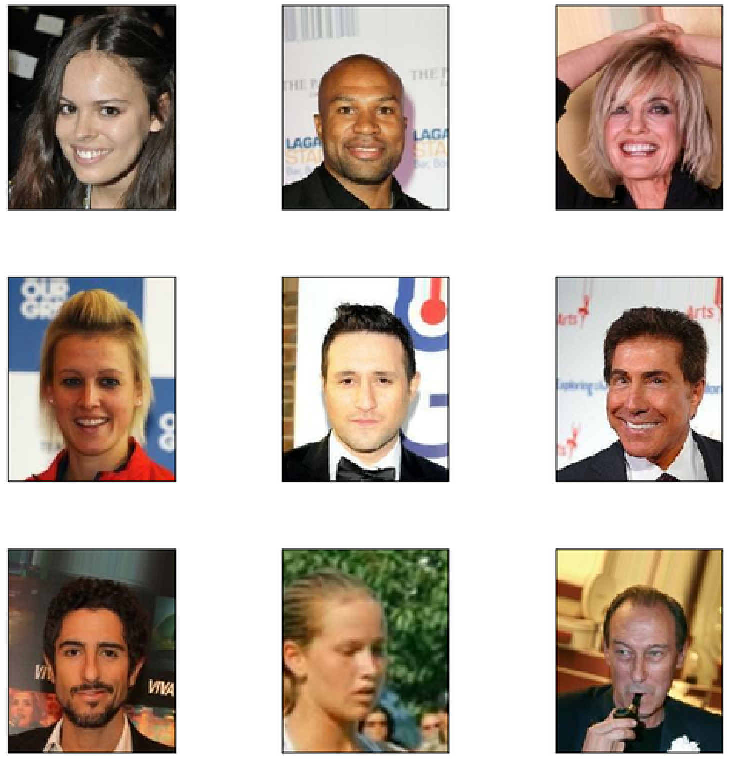
Testing the Model

- After training, the model is tested on new images from the CelebA dataset.
- A blank region is created in each test image, and the model is used to inpaint the image.
- The original image, the image with the blank region, and the inpainted image are displayed side by side for comparison.

Model 2

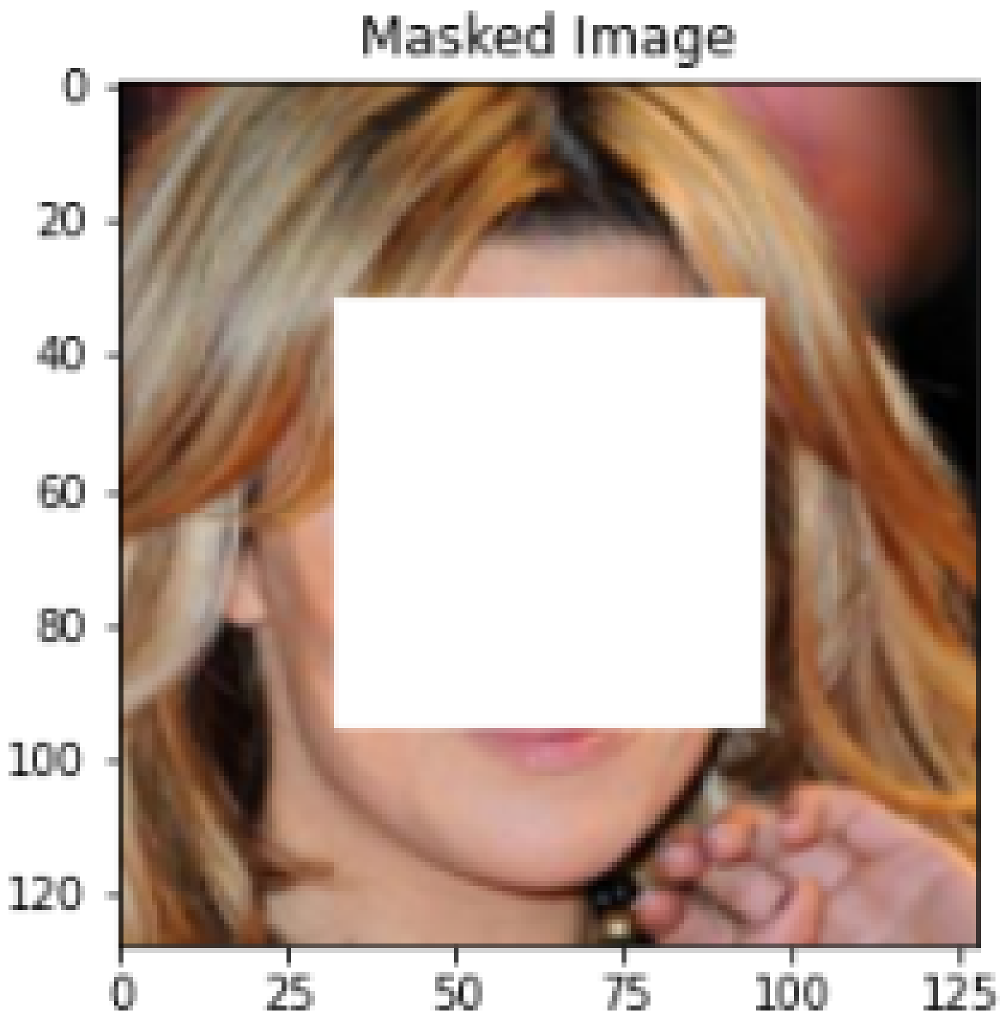
Dataset

The dataset used is the same as the first model, which is Celeb A, which consists of images of faces of well-known people. This model was trained with just 30,000 data. Which was divided between 28,000 training and the rest for validation and testing.

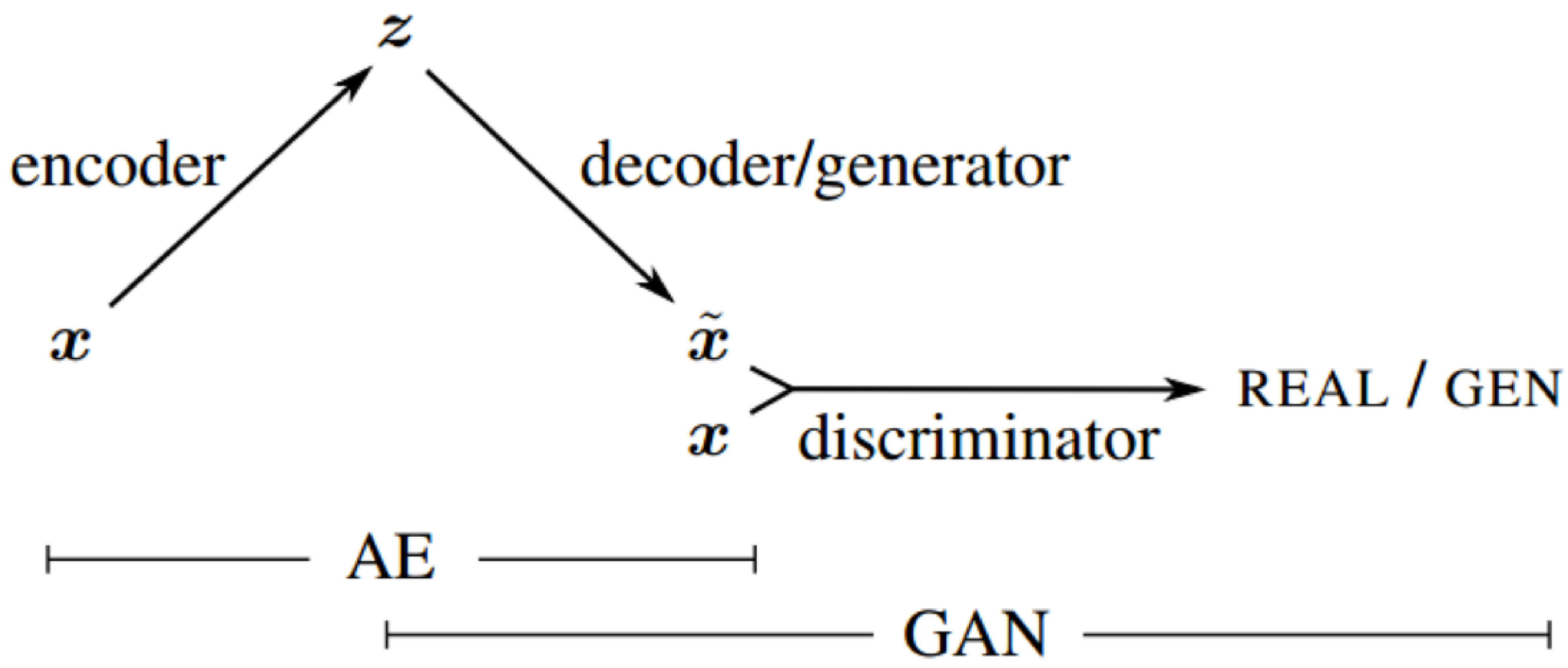


Mask

The mask used in this case is basically based on a central square, which will be applied to the images at the time of training, so that the network learns to generate the missing information.



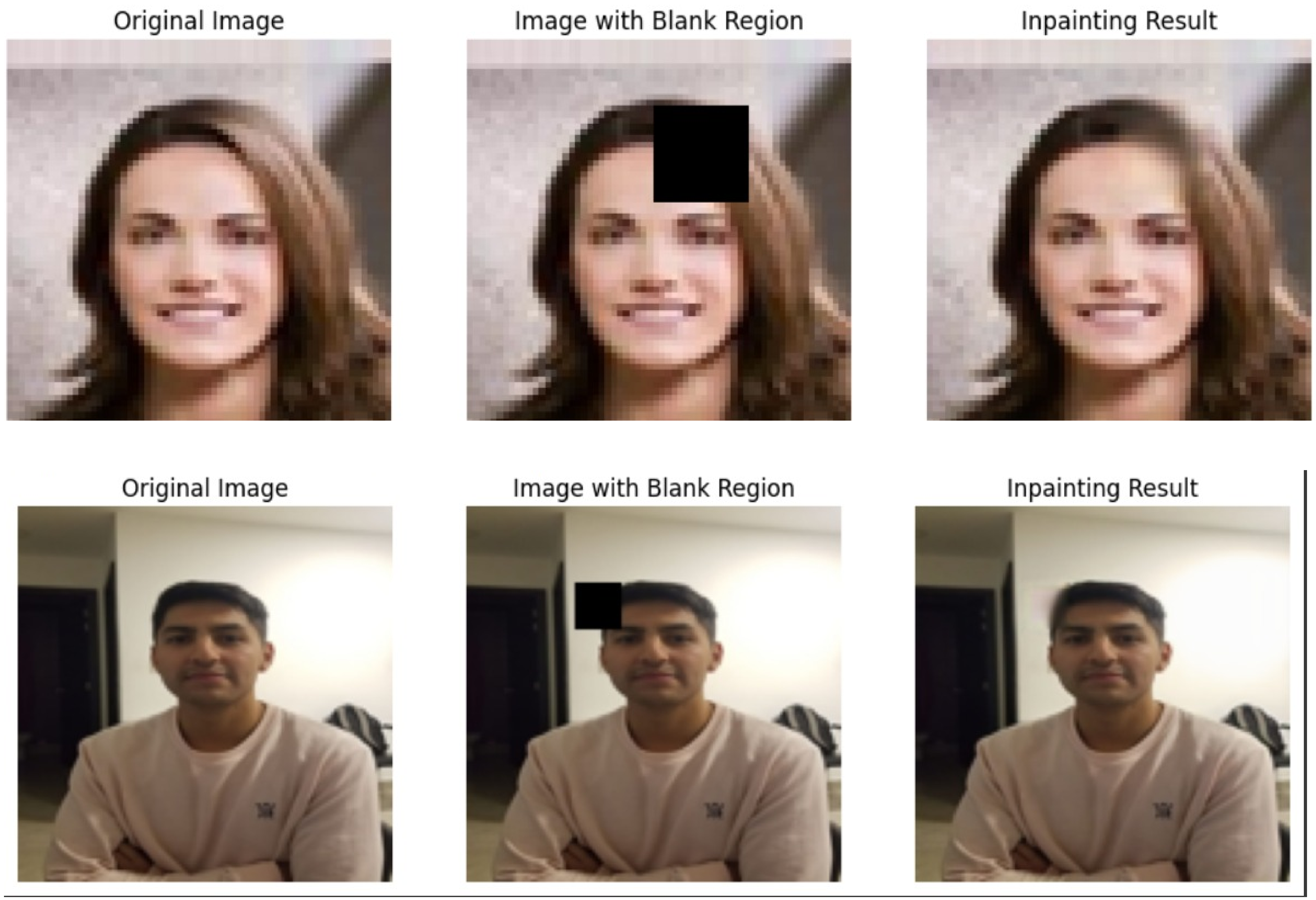
Architecture



Combining a VAE with the GAN is likely to enhance the quality and controllability of the generated images. The VAE component is responsible for learning a meaningful latent representation of the input data, capturing the underlying structure of the images

Results

Model 1


Original Image
Image with Blank Region
Inpainting Result
Original Image
Image with Blank Region
Inpainting Result

Model 2


Original Image
Masked Image
Inpainted Image
Original Image
Masked Image
Inpainted Image


Original Image
Masked Image
Inpainted Image
Original Image
Masked Image
Inpainted Image

Conclusions

- **The U-net architecture, consisting of a contracting path (encoder) and an expanding path (decoder), allows it to effectively capture context (low-level features) and localize (high-level features), respectively.**
- **Image inpainting is the process of filling missing or corrupted parts of images in a plausible way. This is an essential task in various fields, including art restoration, photo editing, and even in medical imaging.**
- **Our use of the U-Net model for the image inpainting task on the CelebA dataset demonstrated the model's effectiveness and potential.**

Conclusions

Use a VAEGAN in inpainting tasks can significantly improve the quality and realism of the generated completions. The combination of VAE and GAN provides a powerful framework for learning expressive representations of the data and generating high-quality inpaintings that seamlessly blend with the rest of the image.

Thanks