**Image Super-Resolution Using Variational Autoencoders (VAE)**

Amit Tzahar, Noam Tsfaty, Alon Gotlib, Ravit Shagan Damti

**Introduction**

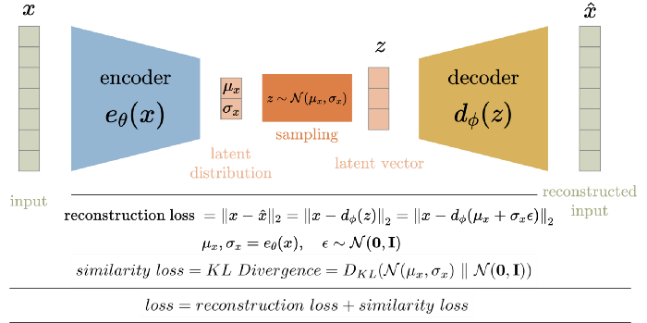
 Variational AutoEncoders (VAEs) represent a class of generative models that employ a probabilistic framework to encode input data into a latent space and subsequently decode it to synthesize new data. In the context of image super-resolution, VAEs are designed to map low-resolution images into a latent space that encapsulates crucial details, thereby facilitating the generation of high-resolution images. This approach enables VAEs to infer and reconstruct missing information, making them particularly effective in enhancing both the resolution and quality of images.  
  
  
  
  
  
  
  
  
  
Image super-resolution (SR) constitutes a fundamental task in computer vision, aimed at enhancing the resolution of low-quality images by generating high-resolution counterparts. This study investigates two distinct VAE-based approaches to SR: the first utilizes perceptual loss to optimize visual quality, while the second employs an unsupervised generative model that operates without the need for paired data. The implementation of these models on the DIV2K dataset facilitated a comparative analysis of their effectiveness, emphasizing both quantitative and qualitative performance metrics. The primary objective was to evaluate the potential of these models to enhance image detail and quality across diverse real-world scenarios.

fig 1- VAE architecture

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטי

fig 2- DEV2K dataset distribution

Figure 2 displays the distribution of image resolutions across different classes in the dataset. The high-resolution (HR) images are concentrated around 3 million total pixels, while the lower resolution (LR) images for x2, x3, x4, and x8 upscaling factors are distributed at significantly smaller pixel counts. The higher frequency of LR images at lower resolutions indicates a greater number of lower-resolution samples compared to the high-resolution ones.

**Methodology**

Network Architecture

This project evaluated two distinct approaches to image super-resolution utilizing Variational Autoencoders (VAEs). The first approach, known as the Perceptual Loss Approach, was proposed by Nujitha Wickramasurendra. It incorporates a perceptual loss layer into the VAE framework, leveraging pre-trained deep neural networks to calculate loss in feature space rather than in pixel space. This strategy is designed to produce images that are not only visually appealing but also closely aligned with human perception.

The second approach, titled Unsupervised Real Image Super-Resolution via Generative VAE, was developed by Zhi-Song Liu et al. This method introduces a novel generative architecture within the VAE framework, distinguishing itself by not relying on paired high-resolution and low-resolution images. Instead, it adopts an unsupervised approach, utilizing a generative model trained on unpaired data. This makes the model particularly resilient to various forms of degradation encountered in real-world scenarios.

While both models share the conventional encoder-decoder architecture characteristic of VAEs, they diverge significantly in their loss functions and training methodologies. These differences are pivotal to their performance in the task of image super-resolution.

**Training the Network**

The training process for these networks involved the optimization of specific loss functions tailored to the unique objectives of each approach:

1. Perceptual Loss Approach:

- The primary goal during training was to minimize a weighted combination of pixel-wise reconstruction loss and perceptual loss. The perceptual loss was computed as the discrepancy in activations between the generated images and the original high-resolution images, using a pre-trained VGG network. This method encourages the model to produce images that are not only accurate in terms of pixel-level detail but also visually consistent with the original high-resolution images, as perceived by human observers.

2. Unsupervised Generative Approach:

This model was trained by incorporating an adversarial loss in conjunction with the standard VAE reconstruction and KL-divergence losses. The adversarial loss is derived from a discriminator network, which compels the generator to create realistic high-resolution images that are indistinguishable from actual images by the discriminator.

A diagram of a bird

Description automatically generated

Fig 3 - second article architecture

In figure 3 we can see the complexed architecture of the second article.  
The encoder transforms the input image into a latent space using a VGG network to enhance feature extraction.

The decoder reconstructs the image from the latent space, removing noise but possibly missing some fine details.

The VAE takes the data and distribution it in the latent space to generate a denoised image using the VAE equation:

This equation can be too complex to calculate so both side will be logged (Lower bound assumption) and rearranged:

To summarize this denoising- we get the noisy image X that given as input to the encoder, using the encoder learns to compress the noisy image to latent variable z and capturing the underlying structure of the image while disregarding the noise. Then the decoder takes this latent variable z and reconstruct the image without the noise.

The denoised image is refined by a GAN, with a Discriminator enhancing realism.

Cycle Training Iteratively improves the image to capture detailed features, with MAE loss ensuring structural consistency:

The loss function in Equation (4) focuses more on the low-frequency components of the image. However, high-frequency components are also important, especially for making the SR image look realistic. To solve this problem the authers use GANs discriminator from VGG19 to extract features map from both the SR and the LR images, and using the loss function in (7) the images will be more realistic:

**Experiments**

Experimental Setup

To assess the performance of the two models, a set of controlled experiments were conducted using a subset of 100 images from the DIV2K dataset. These images were deliberately downscaled to simulate low-resolution inputs, after which the models were tasked with restoring them to their original resolution.

**Perceptual Loss:**

Original Image:



Restore Image:

fig 4- Perceptual Loss Approach article result

The image shows an unsuccessful super-resolution attempt of the first article, where the model failed to produce clear and detailed output, suggesting the need for further model refinement Although this attempt was unsuccessful, we plan to use a larger dataset and a more complex backbone (like ResNet-50) network in future experiments.

**Unsupervised Generative**תמונה שמכילה בקבוק, אוסף, שתיה קלה, מדף

התיאור נוצר באופן אוטומטי

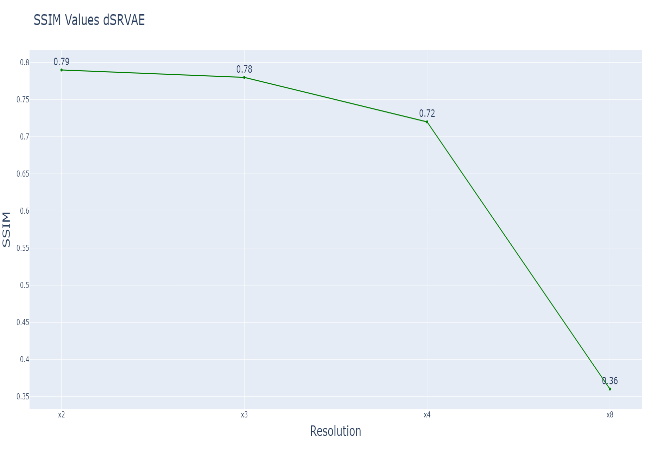
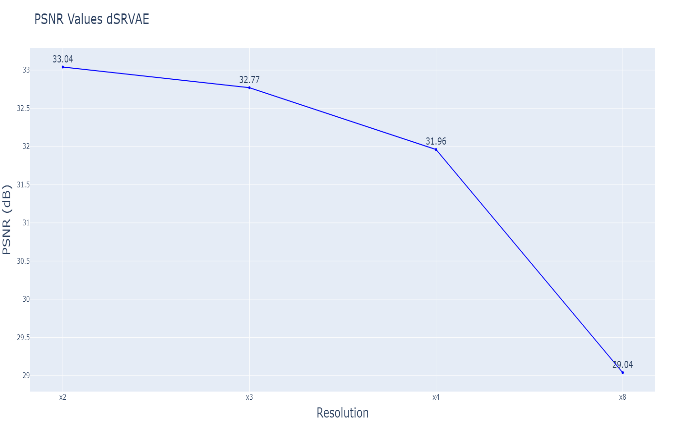
fig 5- VAE and GAN article result

figure 4 show a successful super-resolution experiment, where the model effectively enhances the resolution of an image at various upscaling factors (x2, x3, x4, x8). The results demonstrate that, despite increasing the upscaling factor, the model manages to preserve significant details, closely approximating the high-resolution (HR) reference image.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Data** | **PSNR** | **SSIM** | **MSE** |
| dSRVAE | x2 | 33.04 dB | 0.79 | 35.58 |
| dSRVAE | x3 | 32.77 dB | 0.78 | 37.69 |
| dSRVAE | x4 | 31.96 dB | 0.72 | 44.95 |
| dSRVAE | x8 | 29.04 dB | 0.36 | 86.65 |

Tabel 1: second article outputs:  
  
The table shows that as the upscaling factor increases from x2 to x8, the dSRVAE model's performance degrades. PSNR decreases from 33.04 dB to 29.04 dB, SSIM drops from 0.79 to 0.36, and MSE increases from 35.58 to 86.65. This indicates that higher upscaling factors result in lower image quality and greater error.

fig 6- PSNR and SSIM results



The graphs illustrate the performance of the dSRVAE model across different resolutions. As the resolution increases from x2 to x8, both PSNR (top graph) and SSIM (bottom graph) show a noticeable decline, indicating a decrease in image quality and structural similarity. This suggests that the model's ability to maintain high-quality images diminishes with higher upscaling factors.

**conclusion**

Choosing the Best Model

The best models were selected based on their performance on validation datasets. Key evaluation metrics included Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and perceptual quality scores. The model with the highest average performance across these metrics was chosen for further evaluation and testing.

The table below summarizes the average performance of both models across the test set.

Tabel 2: Comparison between the two models:

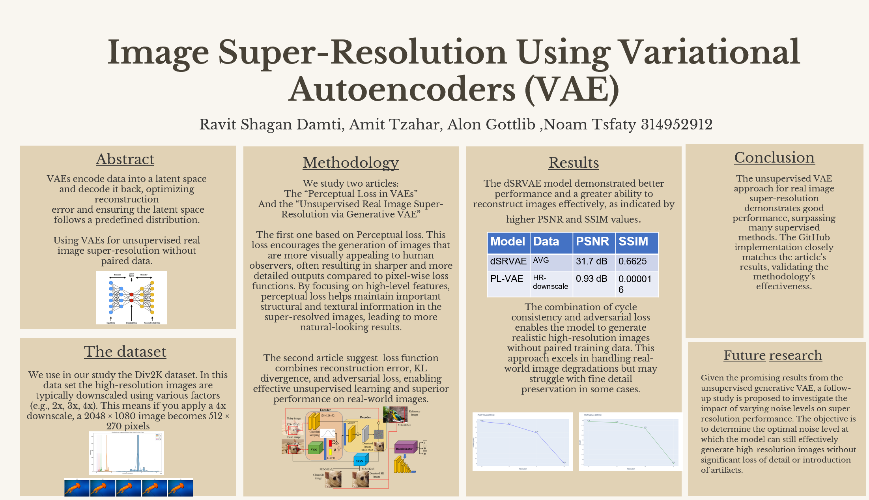
|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Data** | **PSNR** | **SSIM** |
| dSRVAE | AVG | 31.7 dB | 0.6625 |
| PL-VAE | HR- downscale | 0.93 dB | 0.000016 |

As we can see, the **Unsupervised Generative VAE** achieves higher PSNR and SSIM values.

**Future research**

**Investigating the Impact of Varying Noise Levels on Super-Resolution Performance**

Given the promising results from the unsupervised generative VAE, a follow-up study is proposed to investigate the impact of varying noise levels on super-resolution performance. The objective is to determine the optimal noise level at which the model can still effectively generate high-resolution images without significant loss of detail or introduction of artifacts.

**Our poster**

**Bibliography**

1. “dSRVAE (Generative Variational AutoEncoder for Real Image Super-Resolution)” GitHub website

<https://github.com/Holmes-Alan/dSRVAE>

1. “Unsupervised Real Image Super-Resolution via Generative Variational

AutoEncoder” by Zhi-Song Liu, Wan-Chi Siu, Li-Wen Wang1, Chu-Tak Li1, Marie-Paule Cani2, and Yui-Lam Chan1  
<https://arxiv.org/pdf/2004.12811>

1. “Image Super-Resolution using Perceptual Loss with Variational Autoencoder (VAE)” GitHub website

<https://github.com/nujitha99/PerceptualLossVAE/tree/main>

1. “Pereceptual Loss in (Variational) Autoencoder” by Nujith Wickramasurenda

https://github.com/nujitha99/PerceptualLossVAE/blob/main/5P77Project.pdf