Steganography Applications of StyleGAN: A Short Analytical Investigation from Hiding Message in Face Images

FARHAD SHADMAND - NUNO GONÇALVES - LUIZ SCHIRMER

INSTITUTE OF SYSTEMS AND ROBOTICS
UNIVERSITY OF COIMBRA

vis team
isr-uc

29th Portuguese Conference on Pattern Recognition

Abstract

In this investigation, we delve into the latent codes denoted as w, pertaining to both original and encoded images in steganography models, which are projected through StyleGAN—a generative adversarial network renowned for generating aesthetic synthesis.
We present evidence of disentanglement and latent code alterations between the original and encoded images. This investigator possesses the potential to assist in the concealment of messages within images through the manipulation of latent codes within the original images, resulting in the generation of encoded images.
The message into encoded renderings is facilitated by the employment of CodeFace, serving as a steganography model.
CodeFace comprises an encoder and decoder architecture wherein the encoder conceals a message within an image, while the decoder retrieves the message from the encoded image.
By gauging the average disparities amid the latent codes belonging to the original and encoded images, a discerning revelation of optimal channels for concealing information comes to light. Precisely orchestrated manipulation of these channels furnishes us with the means to engender novel encoded visual compositions.

Methods

The experimental protocol unfolds as follows:

1. Face Detection and Extraction: The initial phase entails the application of a specialized face detection algorithm.
2. Hiding message into face images and produce encoded images ,
3. Latent Space Projection: A critical dimension of this investigation involves the projection of the latent space across the entire spectrum of encoded images, encompassing $w+_{CF}$, in tandem with the unaltered originals denoted as $w+_{or}$.
4. Comparative Analysis: The ultimate phase encompasses an intricate comparative analysis, shedding light upon the variances that distinguish the latent spaces across the aforementioned image categories.

Results

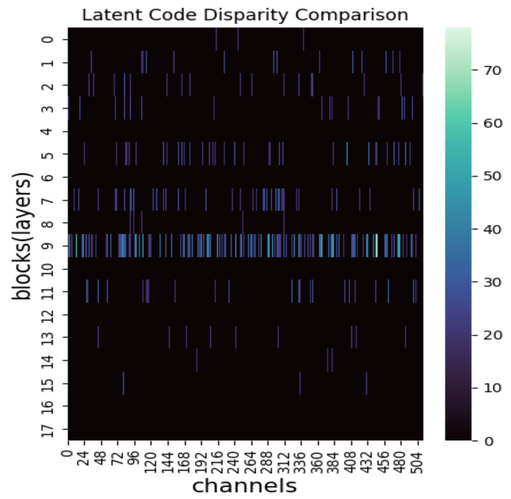
The initial results in the tables below,

blocks	mean
block_9	11.1665
block_5	3.0773
block_7	2.9486
block_11	1.8379
block_3	1.4243
block_1	1.2588
block_2	1.2308
block_13	0.571
block_8	0.2821
block_15	0.2516
block_0	0.1835
block_14	0.1488
block_4	0.0368
block_10	0.0
block_12	0.0
block_6	0.0
block_16	0.0
block_17	0.0

blocks	channels	distance
9.0	445.0	78.0
9.0	400.0	57.89
9.0	366.0	55.98
9.0	271.0	55.9
9.0	477.0	55.16
9.0	13.0	54.47
9.0	418.0	52.85
9.0	383.0	52.28
9.0	290.0	51.38
9.0	109.0	48.17
9.0	463.0	48.0
9.0	99.0	47.34
9.0	93.0	46.42
9.0	216.0	46.14
9.0	494.0	45.71
9.0	296.0	45.33
9.0	25.0	45.32
9.0	349.0	44.92

The first left-side table lists the 18 preeminent channels identified for the covert embedding of messages. These channels, characterized by substantial alterations within latent codes between original and encoded images, emerge as focal points of significance.
The other left table highlights the main channels for message hiding, ranked by importance. On the left, another table lists the top eighteen channels crucial for concealing messages in images.

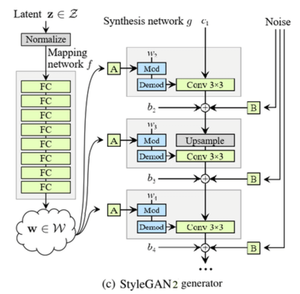
The Figure below manifests as a heatmap delineating the average discrepancy between original and encoded images. Adjacent to this, a tabular exhibition enumerates the mean disparity values corresponding to each individual block.



Why StyleGAN Latent space?

A generative adversarial network (GAN) has two parts: generator and discriminator. The key innovation of StyleGAN lies in its ability to control both high-level features, like overall structure and objects, and low-level features, such as textures and fine details, separately. This separation of control is achieved through a unique two-part generator network: a mapping network and a synthesis network.

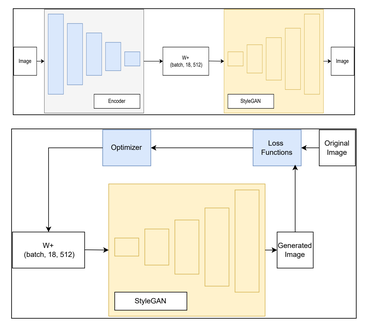
The mapping network takes a 512-dimensional random vector $z \in N(0,1)$ is mapped to an intermediate latent space $w+ \in R^{l\times 512}$, where l is the number of blocks (layers). This intermediate space is designed in such a way that different dimensions control different features of the image. For instance, one dimension might control the age of a generated face, while another might control the hairstyle. This separation of features enables fine-grained control over the generated content.



(c) StyleGAN2 generator

Steganography

Image steganography is a technique to hide a secret message in a cover image or video while minimizing the distinctiveness of the encoded and original images.



References

1. Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan++: how to edit the embedded images? in 2020 ieee. In CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8293–8302.
2. Cian Eastwood and Christopher KI Williams. A framework for the quantitative evaluation of disentangled representations. In International conference on learning representations, 2018Lastname, F. (2000). Title of article. *Journal., Volume*, page-page.
3. Farhad Shadmand, Iurii Medvedev, and Nuno Gonçalves. Codeface: A deep learning printer-proof steganography for face portraits. IEEE Access, 9:167282–167291, 2021.
4. Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 4401–4410, 2019.

farhad.shadmand@isr.uc.pt