Colorectical histology texture synthesis through conditional diffusion

October 24, 2023

Ruben Seror [1]

Abstract

In this work, we introduce an innovative approach to texture synthesis based on probabilistic diffusion models. The model is trained to synthesize texture about colorectal hystologies.

1. Introduction

Texture synthesis is a fundamental problem in the fields of computer graphics and image processing. In this work, we introduce an innovative approach to texture synthesis based on Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020). Diffusion models are designed to capture complex data distributions, making them useful in various applications, such as image generation, and inpainting (Lugmayr et al., 2022). Diffusion models are capable of, such as beating GANs on image synthesis (Dhariwal & Nichol, 2021).The project aims to generate texture images about colorectal cancer hystologies. The adopted dataset, collected by (Kather et al., 2016), contains more than 5000 images of texture about 8 different classes (i.e. Tumor, Stroma, Complex, Lympho, Debris, Mucosa, Adipose, Empty). In Figure 1 it is represented an example for each class. The approach proposed in this work consists to train a DDPM, conditioning the learned distribution with label of the samples in order to better represents texture of specific a class. Diffusion Models work by destroying training data through the successive addition of Gaussian noise, and then learning to recover the data by reversing this noising process. Following the training process, we can utilize the Diffusion Model for data generation by feeding randomly sampled noise through the acquired denoising procedure. The goal of training a diffusion model is to learn the reverse process. The report is organized as follows. Section 2 describe state-of-the-art about texture synthesis. Section 3 describe architecture and method proposed in this work. Section 4 reports experimental results. Section 5 concludes the paper.

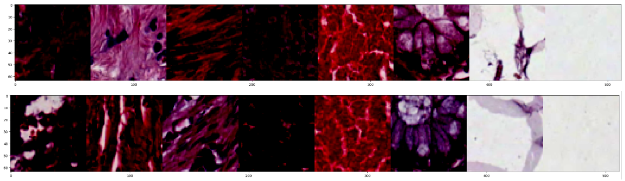

Figure 1. A sample of images contained in the dataset. From left to right a sample of: Tumor, Stroma, Complex, Lympho, Debris, Mucosa, Adipose, Empty.

Code. `https://github.com/rubser98/Diffusion-based-Texture-Synthesis`.

2. Related works

(Gatys et al., 2015) introduced a new parametric texture model based on the feature spaces of convolutional neural networks optimised for object recognition. (Xian et al., 2018) proposed a method based on Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) to synthesize objects consistent with these texture suggestions developing a local texture loss in addition to adversarial and content loss to train the generative network. GANs allow for efficient sampling of high resolution images with good perceptual quality, but are difficult to optimize. Recently, Diffusion Models have achieved state-of-the-art results in sample quality (Dhariwal & Nichol, 2021), especially in text-to-image task, with DALL-E (Ramesh et al., 2022), Stable diffusion (Rombach et al., 2021) and Imagen (Saharia et al., 2022).

3. Method

Diffusion model works in two phases: the forward process that consists of progressively adding noise to the image, and the reverse process, in which noise is transformed back into a sample from target distribution. Equation 1 represents the forward step.

$$q(x_{1:T}|x_0) := \prod_{t=1}^{T} q(x_t|x_{t-1}) := \prod_{t=1}^{T} \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}) \tag{1}$$

[1] **AUTHORERR: Missing \dlaiaffiliation.** Email: Ruben Seror <seror.1815399@studenti.uniroma1.it>.

Deep Learning and Applied AI 2023, Sapienza University of Rome, 2nd semester a.y. 2022/2023.

In this work T is set to 1000. β_1,..., β_T is a variance schedule which each value is evenly spaced in the interval $[0.0001, 0.02]$. During training, model learns to reverse the diffusion process. Starting from pure Gaussian noise $p(x_T) := \mathcal{N}(x_T, 0, I)$, the model learns the joint distribution $p_\theta(X_{0:T})$ defined as follows:

$$p_\theta(X_{0:T}) := p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{2}$$

$$p_\theta(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2\mathbf{I}) \tag{3}$$

$$\sigma_t^2 = \beta_t \tag{4}$$

The model predicts the distribution mean given the noisy image and time step.

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\hat{\alpha}}}\epsilon_\theta(x_T, t))) \tag{5}$$

$$\alpha_t := 1 - \beta_t \tag{6}$$

$$\hat{\alpha_t} := \prod_{s=1}^{t} \alpha_s \tag{7}$$

The loss is a scaled version of Mean-Square Error (MSE) between the real added noise and the one predicted by the model.

$$\mathcal{L}(\theta) = \mathbb{E}_{t,x_0,\epsilon}[||\epsilon - \epsilon_\theta(\sqrt{\hat{\alpha}_t}x_0 + \sqrt{1-\hat{\alpha}_t}\epsilon, t)||^2] \tag{8}$$

The backward process is implemented with the U-Net (Ronneberger et al., 2015) with self attention. The model inputs are conditioned with sinusoidal time step embedding. Classifier free guidance (CFG) (Ho & Salimans, 2022) is also applied to condition the model and improve generation adding information about classes. CFG allows the model to generate textures of an arbitrary class instead of generating just random textures avoiding posterior collapse. The information of the class sample is added to time step embedding. 10% of the training is done unconditionally and 90% is done conditioned by the labels. Sampling phase consists firstly to sample an image from Gaussian distribution. Then, iteratively for each time step is used re-parametrization trick to sample x_{t-1}.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\hat{\alpha}_t}}\epsilon_\theta(x_t, t)) + \sigma_t z \tag{9}$$

z is sampled from $\mathcal{N}(0, \mathbf{I})$ and for x_0, the final generated image, z is valued 0. Sampling is done by linearly interpolating the conditional and unconditional predicted noise.

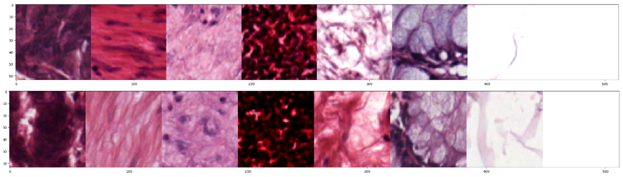

Figure 2. Generated 64x64 textures. From left to right a sample of: Tumor, Stroma, Complex, Lympho, Debris, Mucosa, Adipose, Empty.

4. Experimental results

The model is trained on Google Colab with NVIDIA Tesla T4 for 300 epochs. Due to limited availability of resources the model is trained on 64x64 images. AdamW is used as optimizer. The average sampling time of an image is around 30 seconds. In Figure 2 some generated images for each class are shown. The generated textures result more enlightened than the original images. The model has more difficult to generate texture of Adipose and Empty class because there is a dominance of white in their pattern. To generate textures with greater size is applied a bilinear interpolation to the sampled image.

5. Conclusion and future works

A diffusion based texture synthesis model is presented. The model is able to generate textures similar to images of the dataset. Currently, the model is able to generate 64x64 images. In order to allow the model to synthesize images of greater shapes can be used a state-of-the-art model for augmenting resolution, such as ESRGAN (Wang et al., 2018), or can be trained another diffusion model to do that.

References

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis, 2021.

Gatys, L. A., Ecker, A. S., and Bethge, M. Texture synthesis using convolutional neural networks, 2015.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks, 2014.

Ho, J. and Salimans, T. Classifier-free diffusion guidance, 2022.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models, 2020.

Kather, J. N., Zöllner, F. G., Bianconi, F., Melchers, S. M., Schad, L. R., Gaiser, T., Marx, A., and Weis, C.-A. Collection of textures in colorectal cancer histology, May 2016. URL https://doi.org/10.5281/zenodo.53169.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Gool, L. V. Repaint: Inpainting using denoising diffusion probabilistic models, 2022.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. Hierarchical text-conditional image generation with clip latents, 2022.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models, 2021.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation, 2015.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., and Norouzi, M. Photorealistic text-to-image diffusion models with deep language understanding, 2022.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., and Tang, X. Esrgan: Enhanced super-resolution generative adversarial networks, 2018.

Xian, W., Sangkloy, P., Agrawal, V., Raj, A., Lu, J., Fang, C., Yu, F., and Hays, J. Texturegan: Controlling deep image synthesis with texture patches, 2018.