Generation Based on Diffusion Model

Team ID	Corgi's Writing		
Student No.	Name	Identity	Main Contributions
21373272	高嘉轩	Leader	Abstract, Introduction, and Conclusion Writing. Discussion and classification.
21371195	魏润田	Member	Review and critical analysis Part 1 Writing.Discussion and classification.
21373219	胡斌	Member	Review and critical analysis Part 2 Writing. Analysis and design.
21371188	李昱哲	Member	Review and critical analysis Part 3 Writing. Analysis and design.
21371255	刘兆丰	Member	Slide production. Inspection and evaluation
21373278	赵乾宇	Member	Slide production. Inspection and evaluation

Abstract

Introduction

The advent of machine learning and artificial intelligence has propelled numerous advancements in generating high-quality data, such as images, text, and audio. Among the various generative models, diffusion models have emerged as a powerful and promising approach for data generation. The general problem addressed by diffusion models is the creation of realistic and high-fidelity synthetic data from noise, leveraging a probabilistic framework that iteratively refines the generated outputs.

The ability of diffusion models to generate realistic data has profound implications across numerous fields. High-quality synthetic data can enhance training datasets, improve the performance of machine learning models, and enable simulations in environments where real data is scarce, expensive, or ethically challenging to obtain. For instance, in the medical field, generating synthetic medical images can augment datasets for training diagnostic models, reducing the reliance on limited and sensitive patient data. In entertainment, diffusion models can be used to create lifelike animations and special effects, pushing the boundaries of visual storytelling. Additionally, in natural language processing, these models can generate coherent and contextually relevant text, aiding in tasks such as automated content creation and conversational agents.

The potential applications of diffusion models are vast and varied. They can be employed in the development of advanced image editing tools, where users can manipulate and enhance images with unprecedented precision. In robotics, synthetic data generated by diffusion models can be used to train autonomous systems in virtual environments, ensuring better performance and safety in the real world. Moreover, these models can contribute to scientific research by simulating complex phenomena, enabling scientists to explore scenarios that are otherwise infeasible in laboratory settings. Thus, diffusion models represent a significant leap forward in generative modeling, offering innovative solutions to complex problems across diverse domains.

Review and critical analysis

1. Text-to-Image

The text-to-image diffusion model is a model used for generating images, and its core idea is to convert text descriptions into images. This model takes natural language descriptions as input and generates photo-realistic images corresponding to the descriptions. In recent years, the diffusion models have made a big splash in the field of text-to-image generation, and their unique way of combining text and image information allows them to generate highly relevant images based on the text. From another perspective, these generative models suggest precise correlations between words and pixels.

In the text-to-image diffusion model, a well-trained text encoder is ffrst needed to encode natural language descriptions into a vector representing the textual content. Then, the model uses this textual representation along with some random noise as input to progressively generate the image through diffusion process. During the diffusion process, the model continuously updates the states of pixels to make the generated image gradually approach a real image.

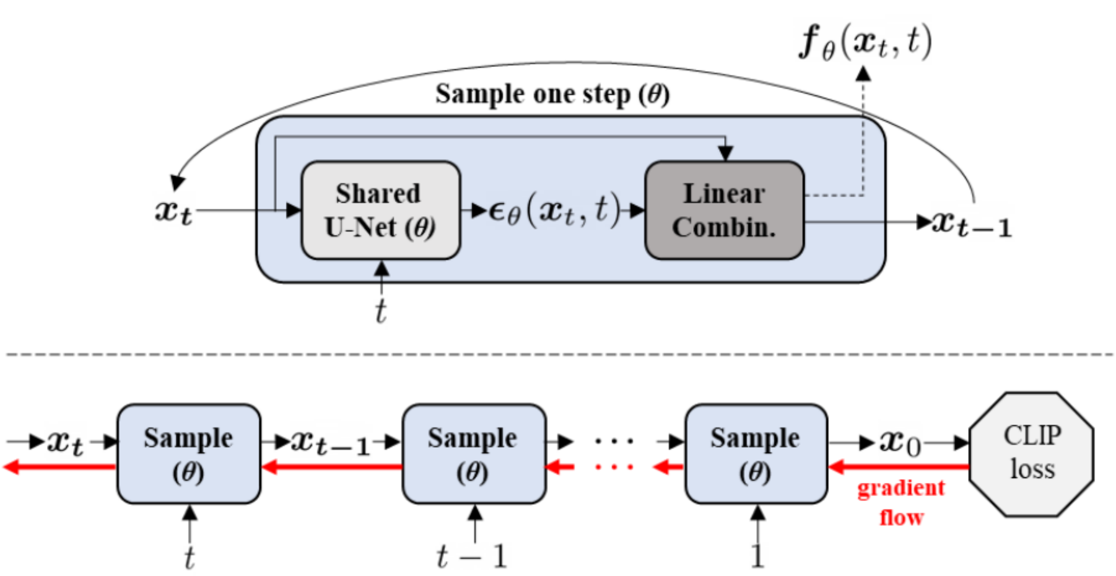
During the diffusion process, the model generates new pixel values at each step based on the textual representationand noise, and combines attention mechanisms to ensure the consistency between pixel generation and text description. Attention mechanisms help the model focus on image features that are relevant to the description, enabling better generation of images that match the description.

- DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation

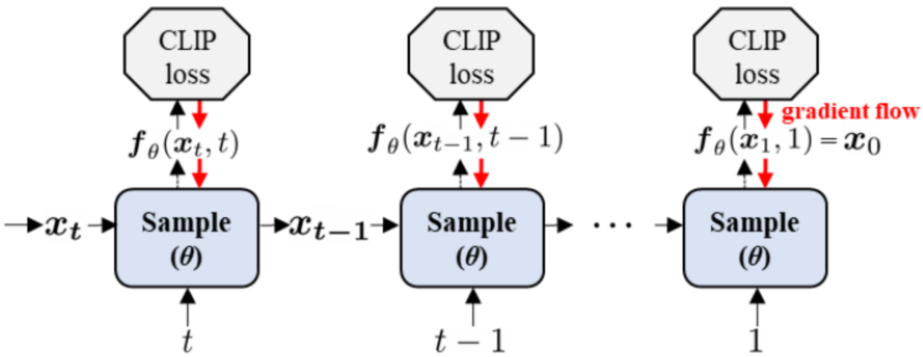
The paper introduces DiffusionCLIP, a method for text-guided image manipulation using diffusion models. It leverages the full inversion capability of diffusion models to perform zero-shot image manipulation guided by text prompts, enabling tasks such as image translation between unseen domains and multi-attribute transfer. The methodology fine-tunes the reverse diffusion model using a CLIP loss to control image attributes based on text prompts, showcasing robust manipulation performance compared to existing baselines. The study highlights the importance of deterministic processes in both forward and reverse diffusion for successful image manipulation, while also discussing the benefits of stochastic sampling for image translation between unseen domains. The article's contribution lies in its fast sampling strategy, which accelerates the forward diffusion process and reduces fine-tuning time, making it a valuable tool for image generation tasks.

The methodology employed in DiffusionCLIP demonstrates a novel approach to text-guided image manipulation by combining diffusion models with CLIP loss, showcasing superior performance in diverse image manipulation tasks. The relationship between the aims, methodology, and results is well-aligned, with the methodology effectively addressing the research goals of enabling robust image manipulation guided by text prompts. This article stands out in comparison to others on the same topic by offering a comprehensive solution that leverages diffusion models for zero-shot image manipulation, showcasing the potential for advancements in the field of image generation.

This article would be highly beneficial to your research on Image Generation Based on Diffusion Model, as it introduces a novel method, DiffusionCLIP, that demonstrates the effectiveness of combining diffusion models with text guidance for image manipulation tasks. The insights and techniques presented in this paper could provide valuable guidance and inspiration for your research endeavors in the field of image generation using diffusion models.



(a) Original fine-tuning

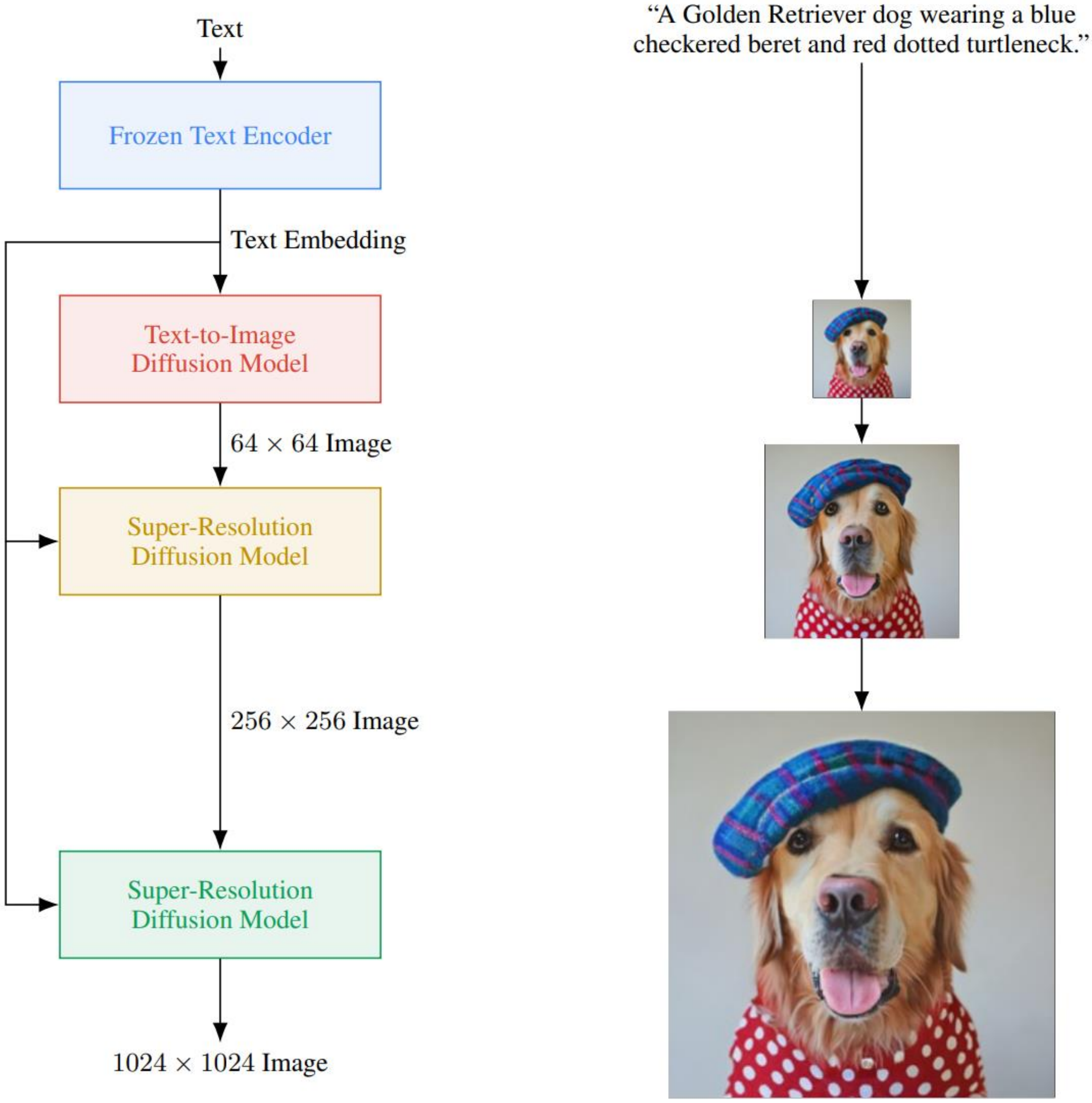


- Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding

The paper "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding" introduces Imagen, a text-to-image diffusion model that leverages large transformer language models and diffusion models to generate photorealistic images with a deep understanding of language. The key finding is that large language models pretrained on text-only corpora are highly effective at encoding text for image synthesis, leading to improved sample fidelity and image-text alignment in Imagen. Imagen achieves a new state-of-the-art FID score on the COCO dataset without direct training on COCO, and human raters prefer Imagen over other models in terms of sample quality and image-text alignment.

The methodology used in the paper, particularly the combination of large pretrained language models with diffusion models, showcases a novel approach to text-to-image synthesis, emphasizing the importance of text encoding for image generation. The relationship between the aims, methodology, and results is well-established, with the paper demonstrating how scaling language models has a significant impact on overall performance in text-to-image generation.

For research on Image Generation Based on Diffusion Models, this article provides valuable insights into the effectiveness of frozen large pretrained language models as text encoders and the impact of scaling language models on performance. The comparison with other models in terms of sample quality and image-text alignment, as well as the introduction of DrawBench as a benchmark for text-to-image models, adds to the understanding of current advancements in the field.



- GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models

The paper "GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models" explores the application of guided diffusion models for text-conditional image synthesis, achieving photorealistic image generation by conditioning on natural language descriptions. The study compares two guidance techniques, CLIP guidance and classifier-free guidance, finding that classifier-free guidance produces higher-quality images according to human and automated evaluations. The research also demonstrates the effectiveness of fine-tuning diffusion models for image inpainting, enabling text-driven image editing. The methodology involves training a 3.5 billion parameter text-conditional diffusion model and a 1.5 billion parameter upsampling diffusion model, as well as utilizing noised CLIP models for guidance.

The methodology used in the paper, particularly the adoption of guided diffusion models for text-conditional image synthesis, showcases a novel approach to achieving photorealistic image generation. The comparison between CLIP guidance and classifier-free guidance provides valuable insights into the effectiveness of different guidance techniques in image synthesis.

Compared to other studies in the field of image generation based on diffusion models, this paper stands out for its focus on text-guided image synthesis and editing. The incorporation of text prompts for image generation sets this research apart from traditional image synthesis approaches.

This article would be highly useful for research focused on image generation based on diffusion models, especially for those interested in text-conditional image synthesis and guided diffusion models. The findings and methodologies presented in this paper could provide valuable insights and inspiration for further research in this area

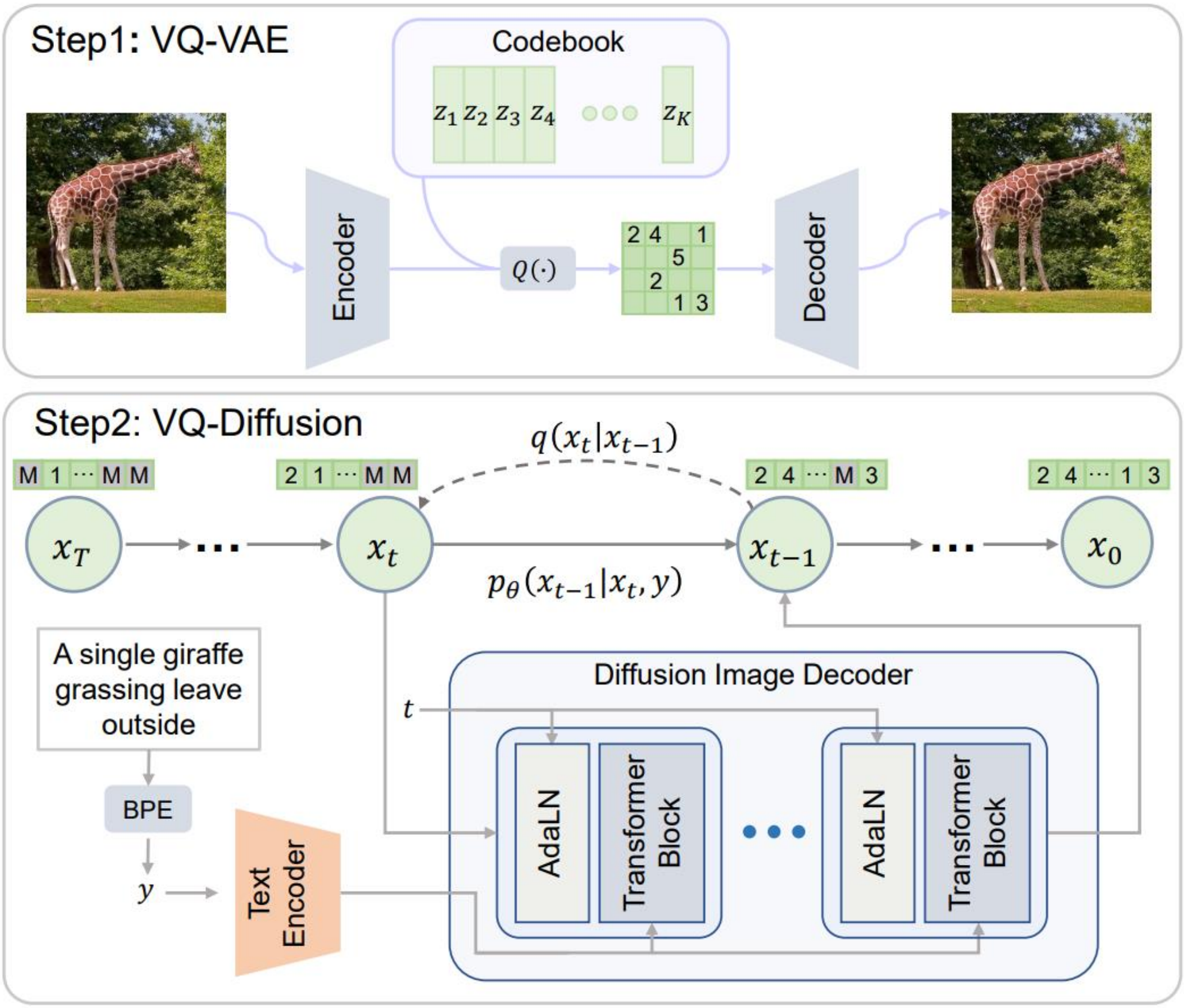
- Vector Quantized Diffusion Model for Text-to-Image Synthesis

The paper introduces the Vector Quantized Diffusion (VQ-Diffusion) model for text-to-image generation, which addresses issues such as unidirectional bias and accumulated prediction errors in existing methods 3. The VQ-Diffusion model leverages a combination of a vector quantized variational autoencoder (VQ-VAE) and a conditional variant of the Denoising Diffusion Probabilistic Model (DDPM) to improve image synthesis and quality .

In comparison to autoregressive models and previous GAN-based methods, the VQ-Diffusion model demonstrates significantly better performance in handling complex scenes, enhancing image quality, and achieving a more favorable balance between quality and speed . Notably, the model offers faster inference speed by providing global context for each token prediction and making it independent of image resolution through reparameterization, resulting in a substantial speed improvement while maintaining image quality

The methodology employed in the VQ-Diffusion model, incorporating a mask-and-replace diffusion strategy, effectively addresses biases and errors present in previous text-to-image generation methods The alignment between the research aims of improving text-to-image generation quality and speed, the methodology of combining VQ-VAE and DDPM, and the results showcasing superior performance highlights the success of the study.Compared to other approaches in the field, such as GAN-based methods and autoregressive models, the VQ-Diffusion model stands out for its ability to handle complex scenes, improve image quality, and enhance inference speed.

The article is highly relevant to your research on Image Generation Based on Diffusion Model, as it introduces a novel approach that significantly improves text-to-image generation quality and speed, addressing key limitations of existing methods.



2. Application

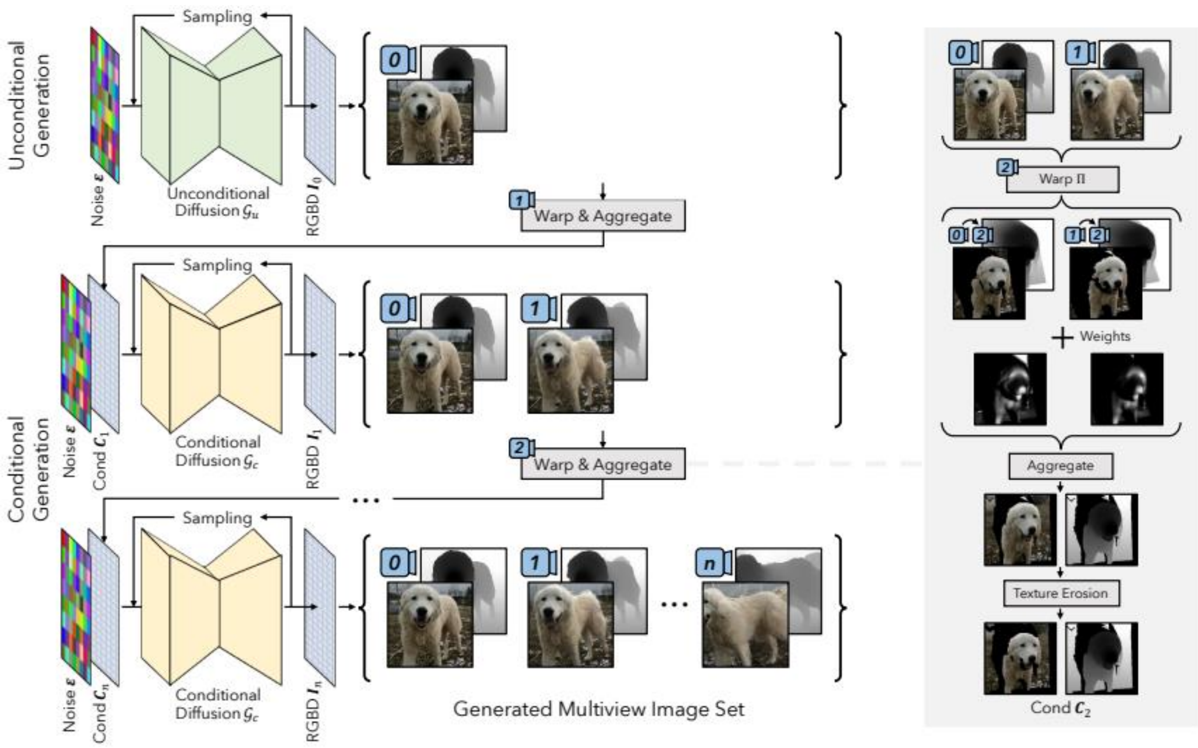
Image generation based on diffusion models has shown strong application potential in multiple fields. In the research article, diffusion models demonstrate how to generate images with 3D perception ability through 2D diffusion models at the application level, achieving breakthroughs in virtual reality and 3D modeling. The application of potential diffusion models in brain imaging generation is explored, and guided diffusion methods are used to make the image generation process more controllable. Fine adjustments can be made according to user input, improving the quality and flexibility of generated images. The application of composable diffusion models in visual generation is emphasized, which is widely used in advertising design, art creation, and content generation by combining different elements to generate complex images. Overall, the diverse applications of diffusion models in image generation technology are continuously driving the development of related fields.

- 3D-Aware Image Generation Using 2D Diffusion Models

This article introduces a new method for 3D perceptual image generation using 2D diffusion models. The researchers formalized the 3D perception image generation task into multi-view 2D image set generation, and further transformed it into a sequential unconditional-conditional multi-view image generation process. They used the 2D diffusion model to improve the generative modeling ability of the method, and constructed the training data of the conditional diffusion model through the depth information of the monocular depth estimator, using only static images. They trained their method on ImageNet, a large-scale unstructured 2D image dataset, and the resulting high-quality images were significantly superior to the previous method. In addition, their method demonstrates its ability to generate instances with large viewing angles, even when training images are from a "wild" real-world environment, diverse, and misaligned.

In this paper, the researchers adopt an innovative approach to solve the problem of 3D perceptual image generation, introducing 2D diffusion models into this field. The methodological nature of this method is positive, as it successfully leverages the generative modeling capabilities of the 2D diffusion model and realizes the training of the conditional diffusion model by introducing depth information. The method in this paper has a strong relationship with the results, and by introducing a 2D diffusion model, the researchers have achieved high-quality image generation results that are superior to those of previous methods. This article takes a different approach than other articles on the same topic, making significant progress by utilizing 2D diffusion models instead of traditional GANs.

This article is very useful for your research on diffusion model-based image generation, as it presents a novel approach that demonstrates the potential of utilizing 2D diffusion models in 3D perception image generation tasks with impressive results.



- Brain Imaging Generation with Latent Diffusion Models

The paper "Brain Imaging Generation with Latent Diffusion Models" explores the use of Latent Diffusion Models (LDMs) to generate synthetic brain images from high-resolution 3D MRI data, addressing the challenge of limited dataset sizes in medical imaging projects. The study demonstrates that LDMs can produce high-quality synthetic images with sharp details and realistic textures, outperforming traditional GAN-based methods in stability and convergence, especially at high resolutions. The methodology involved training an autoencoder for compression and using diffusion models for generative modeling, conditioning the models on variables such as age, gender, and brain structure volumes.

The methodology used in the study, combining autoencoders for compression and diffusion models for generation, appears appropriate for generating realistic synthetic brain images. The conditioning on relevant variables enhances the model's ability to replicate properties from the training images.

Compared to other studies in the field of medical image generation, this paper showcases the effectiveness of Latent Diffusion Models in producing high-quality synthetic brain images and highlights the advantages over traditional GAN-based methods, especially in stability and convergence at high resolutions.

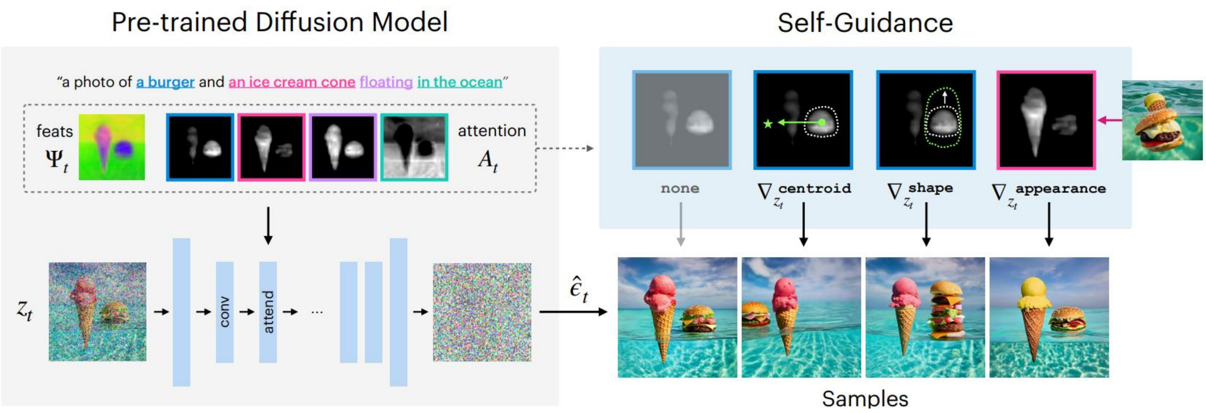
This article would be highly useful for your research on Image Generation Based on Diffusion Models as it provides insights into the application of LDMs for generating synthetic brain images, addressing dataset limitations in medical imaging projects, and demonstrating the potential of these models for future research in medical image generation

- Diffusion Self-Guidance for Controllable Image Generation

This article introduces the concept of self-guidance, which provides a more precise control over the generated images through an internal representation of the guided diffusion model. The authors show that the properties of objects can be extracted from these representations and used to guide the sampling process, enabling fine-grained manipulation of image attributes without the need for additional models or training. Self-guidance makes it possible to perform complex image manipulations, such as moving or adjusting objects, merging the appearance of different images, and combining objects. This approach offers a unique way of manipulating images that opens up new possibilities in the field of image generation.

In this paper, the self-guided approach proposed by the authors provides a new avenue for zero-shot control for image generation without additional modeling or supervision. The method allows fine-grained manipulation of attributes such as the shape, position, and appearance of the generated image by utilizing the internal representation of the pre-trained text-to-image diffusion model. The innovation of this approach lies in the fact that complex image manipulations such as moving, adjusting objects, merging the appearance of different images, etc., can be achieved without the need for expensive optimization processes. However, despite the more control provided by the method, there are still some simple editing operations that are still difficult to implement, which may require further research and refinement.

For your research on diffusion model-based image generation, this article provides a new approach, self-guidance, that can enhance control over the generated images without the need for additional modeling or supervision. The article shows how to extract attributes from the internal representation of a pretrained model for fine-grained manipulation of the generated images. This may be of great significance to your research, as it can provide new ideas and methods for your image generation work, and help you better control the properties and appearance of the generated images.

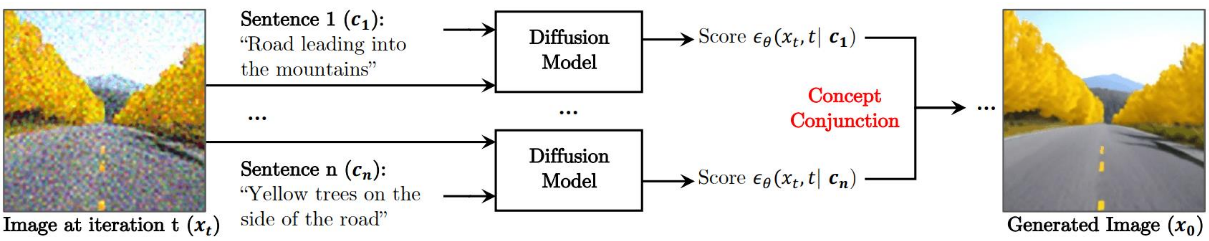


- Compositional Visual Generation with Composable Diffusion Models

The paper "Compositional Visual Generation with Composable Diffusion Models" introduces a novel approach for compositional visual generation using diffusion models. The main idea is to compose a set of diffusion models, each modeling a specific component of an image, to generate complex scenes at test time. The proposed method interprets diffusion models as energy-based models, allowing for structured generalization in visual generation tasks. By explicitly combining data distributions, the approach enables the generation of scenes with complex compositions, including sentence descriptions, object relations, human facial attributes, and new combinations rarely seen in the real world. The key findings highlight the effectiveness of the method in promoting structured generalization for visual generation.

The methodology used in this article, which interprets diffusion models as energy-based models and composes them for visual generation, is innovative and shows promise in addressing the limitations of existing text-guided diffusion models. The relationship between the aims, methodology, and results is well-established, as the proposed approach successfully generates high-quality images with complex compositions, showcasing the effectiveness of the compositional operators introduced. This article stands out in its focus on structured generalization in visual generation tasks, offering a unique perspective compared to other works in the field .

This article can be highly useful for research on Image Generation Based on Diffusion Models, as it presents a novel approach that enables the generation of complex scenes and compositions not commonly seen in training data. The findings and methodology outlined in the paper provide valuable insights for researchers interested in advancing the field of visual generation using diffusion models.



3. Multiple types of diffusion models

In recent years, multiple types of diffusion models have been developed to tackle various challenges in image generation, each introducing unique methodologies to enhance performance and efficiency. These papers illustrate the diversity and innovation in diffusion models, each contributing valuable insights and techniques to the field of image generation.

3.1 Multilevel diffusion: Infinite dimensional score-based diffusion models for image generation

The paper "Multilevel Diffusion: Infinite Dimensional Score-Based Diffusion Models for Image Generation" introduces a novel approach to image generation using score-based diffusion models (SBDM) in an infinite-dimensional setting. The authors aim to create a well-posed infinite-dimensional learning problem that can be discretized consistently on multiple resolution levels, leading to diffusion models that generalize across different resolutions and improve training efficiency. The key contributions include modifying the forward process to ensure a well-defined latent distribution in the infinite-dimensional setting, deriving reverse processes for finite approximations, and demonstrating the benefits of approximating the score function with an operator network for multilevel training.

In terms of critical comments, the methodology used in this article, particularly the development of SBDMs in an infinite-dimensional setting, showcases a novel and promising approach to image generation. The relationship between the aims, methodology, and results is well-established, with the authors successfully addressing the challenges of infinite-dimensional SBDMs and demonstrating the effectiveness of their approach through experimental results on MNIST and Fashion-MNIST datasets. This article stands out in its exploration of infinite-dimensional diffusion models for image generation, offering a unique perspective compared to existing finite-dimensional approaches.

Regarding the usefulness of the article to research on Image Generation Based on Diffusion Model, this paper provides valuable insights into the development of multilevel diffusion models for image generation in an infinite-dimensional setting. Researchers interested in exploring advanced techniques for image generation and generative modeling may find the theoretical foundations, methodologies, and experimental results presented in this article beneficial for their own investigations.

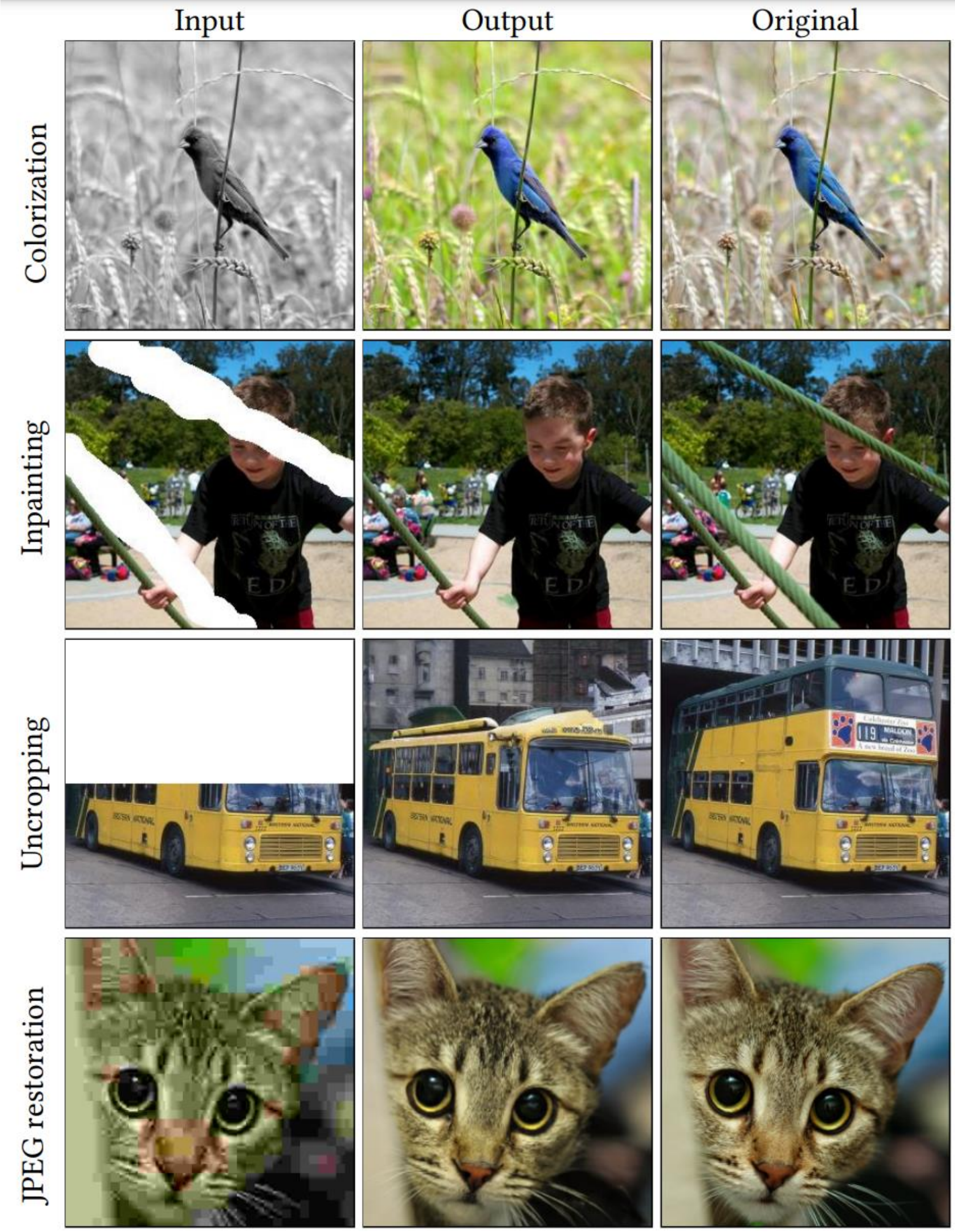
3.2 Palette: Image-to-image diffusion models

The paper "Palette: Image-to-Image Diffusion Models" introduces a unified framework for image-to-image translation based on conditional diffusion models and evaluates its performance on challenging tasks such as colorization, inpainting, uncropping, and JPEG restoration. The research questions addressed include the impact of loss functions on sample diversity, the importance of self-attention in neural architecture, and the effectiveness of a generalist diffusion model compared to task-specific specialist models.

In my critical comment on the summary, I would focus on the appropriateness of the methodology used in the paper. The authors demonstrate the effectiveness of their approach across various image-to-image translation tasks without the need for task-specific customization or auxiliary loss, which highlights the robustness and versatility of their framework. The relationship between the aims, methodology, and results is well-established, as the authors successfully address the research questions posed and provide empirical evidence to support their claims.

Compared to other works in the field, the paper stands out for advocating a unified evaluation protocol based on ImageNet and showcasing the performance of a generalist diffusion model against task-specific specialist counterparts. This approach contributes to advancing image-to-image translation research by providing a standardized evaluation framework and demonstrating the efficacy of a multi-task diffusion model.

Regarding the usefulness of the article to your research about Image Generation Based on Diffusion Model, this paper can serve as a valuable reference due to its comprehensive exploration of conditional diffusion models for image-to-image translation tasks. The insights provided on loss functions, neural architecture, and multi-task training can inform and inspire further research in the field of image generation based on diffusion models.

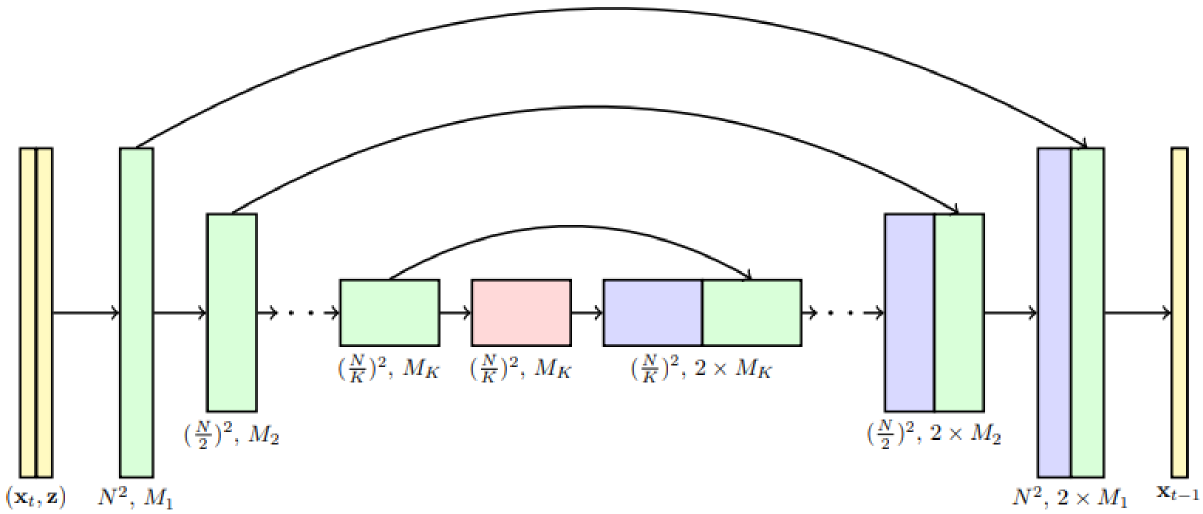


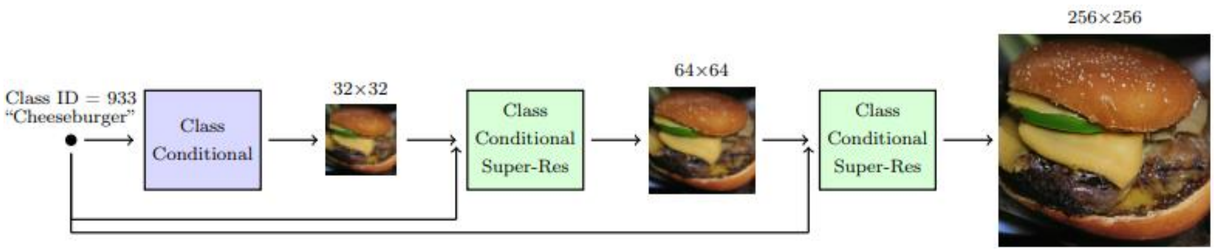
3.3 Cascaded diffusion models for high fidelity image generation

The article discusses Cascaded Diffusion Models (CDM) for high-fidelity image generation, emphasizing the importance of conditioning augmentation in improving sample quality in cascading pipelines. The key contributions include showcasing the superiority of CDM over BigGAN-deep and VQ-VAE-2 in terms of FID scores and classification accuracy scores, without the need for additional classifiers. The methodology involves using conditioning augmentation for super-resolution models, exploring augmentation policies, and training models amortized over varying levels of conditioning augmentation. The results demonstrate the effectiveness of cascading diffusion models in generating high-quality images on the class-conditional ImageNet benchmark.

The article's methodology of incorporating conditioning augmentation in cascading pipelines is crucial for achieving high sample fidelity and outperforming existing generative models. The relationship between the aims, methodology, and results is well-established, as the study successfully demonstrates the effectiveness of CDM with conditioning augmentation in generating high-fidelity images. In comparison with other works on diffusion models, this article stands out for its in-depth exploration of conditioning augmentation and its focus on improving sample quality through cascading alone, without the use of additional classifiers.

This article would be highly useful for research on Image Generation Based on Diffusion Models, providing insights into the effectiveness of cascaded diffusion models and the importance of conditioning augmentation in achieving high sample fidelity.





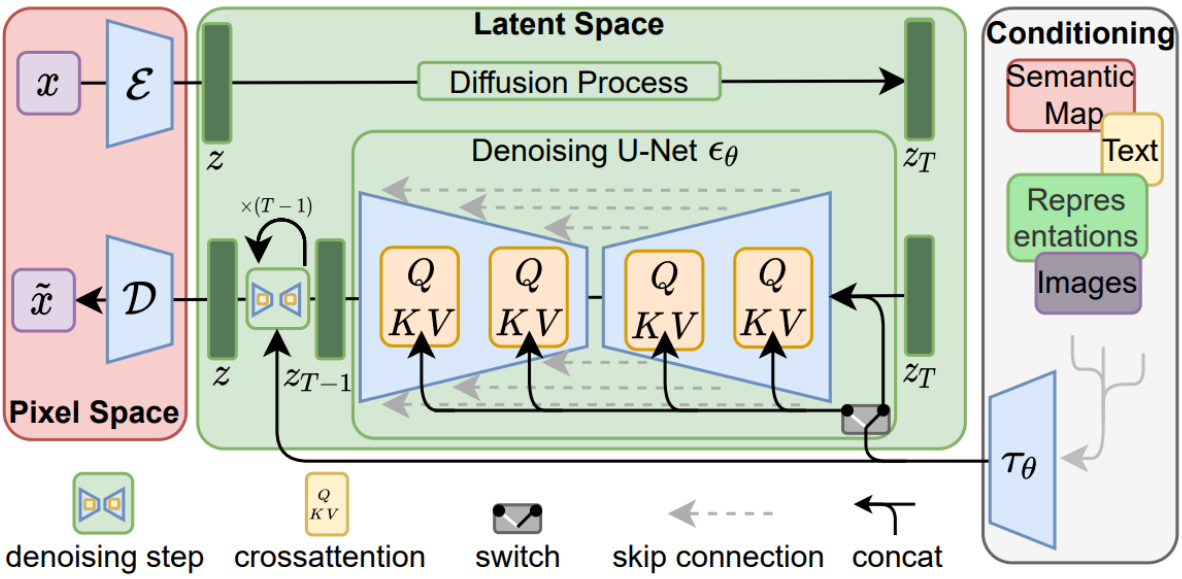
3.4 High-Resolution Image Synthesis with Latent Diffusion Models

The paper "High-Resolution Image Synthesis with Latent Diffusion Models" introduces a novel approach to training diffusion models (DMs) for high-resolution image synthesis by utilizing a latent space obtained from a pretrained autoencoder, leading to reduced computational complexity while maintaining visual fidelity. The authors demonstrate the effectiveness of their Latent

Diffusion Models (LDMs) in achieving state-of-the-art results for various image synthesis tasks, including image inpainting, class-conditional image synthesis, text-to-image synthesis, and super-resolution. The key contributions of the paper include the scalability of LDMs to higher-dimensional data, efficient image generation from the latent space, and the design of a general-purpose conditioning mechanism based on cross-attention.

The methodology employed in paper, particularly the use of latent diffusion models trained in a perceptually equivalent latent space, is innovative and addresses the computational challenges associated with high-resolution image synthesis. By leveraging pretrained autoencoders and cross-attention mechanisms, the authors have demonstrated significant advancements in image generation tasks. The relationship between the aims, methodology, and results is well-established, showcasing how the proposed approach effectively bridges the gap between computational efficiency and visual fidelity in image synthesis tasks.

This article is highly relevant to research on image generation based on diffusion models, offering valuable insights into improving computational efficiency without compromising on image quality. Researchers in the field of image synthesis can benefit from the novel methodology presented in this paper, especially in tasks requiring high-resolution image generation and conditional image synthesis. The release of pretrained latent diffusion and autoencoding models further enhances the utility of this work for various image-to-image and text-to-image tasks.

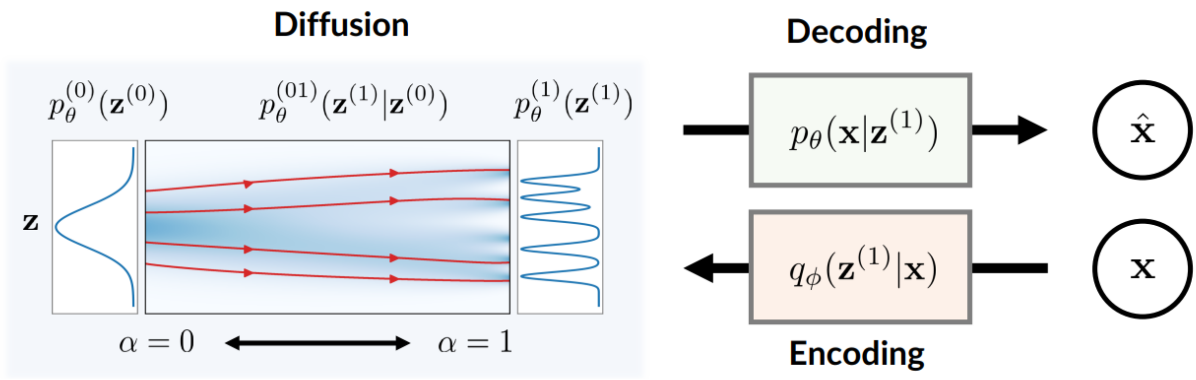


3.5 D2C: Diffusion-Decoding Models for Few-Shot Conditional Generation

The paper "D2C: Diffusion-Decoding Models for Few-Shot Conditional Generation" introduces a novel paradigm called D2C for training unconditional VAEs for few-shot conditional image generation. D2C utilizes contrastive self-supervised learning and a diffusion-based prior over latent representations to enhance generation and representation quality. It can adapt to new generation tasks with minimal labeled examples and outperforms state-of-the-art VAEs and diffusion models in conditional generation tasks. The methodology of D2C involves a diffusion component operating over the latent space and a decoding component mapping latent representations to images. The generative process of D2C involves drawing latent variables from a diffusion process and decoding images from these latent variables. The paper provides detailed algorithms for conditional generation with D2C and discusses how manipulation constraints can be incorporated into the model. The article is significant as it addresses the challenge of learning conditional generative models with minimal paired data, showcasing superior performance and efficiency compared to existing methods.

The methodology used in the paper, combining contrastive self-supervised learning with a diffusion-based prior, is innovative and effective in addressing the challenge of few-shot conditional image generation. The relationship between the aims, methodology, and results is well-established, with the paper demonstrating superior performance over existing models in conditional generation tasks. The article's comparison with other works in the field highlights its contributions to advancing the state-of-the-art in generative models.

Regarding the usefulness of the article to your research about Image Generation Based on Diffusion Model, this paper provides valuable insights and methodologies that can enhance your understanding and approach to conditional image generation tasks. The innovative techniques introduced in D2C could potentially improve the efficiency and performance of your research in this area.



Conclutions and recommendation

The exploration of diffusion models for image generation reveals significant advancements and diverse applications across multiple domains. Our review categorizes the state-of-the-art research into three main areas: Text-to-Image, Application, and Multiple Types of Diffusion Models.

Research in t**ext-to-Image** category demonstrates the powerful capabilities of diffusion models to generate high-quality images guided by textual descriptions. Notable works like DiffusionCLIP, Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, GLIDE, and Vector Quantized Diffusion Model for Text-to-Image Synthesis highlight the robustness and versatility of these models in understanding and translating text into detailed and photorealistic images. These advancements indicate significant potential for applications in fields such as content creation, advertising, and interactive media.

Studies focused on specific applications illustrate the broad utility of diffusion models beyond general image synthesis. For instance, 3D-Aware Image Generation Using 2D Diffusion Models and Brain Imaging Generation with Latent Diffusion Models reveal the potential in 3D modeling and medical imaging, respectively. Diffusion Self-Guidance for Controllable Image Generation introduces methods to enhance control over the generation process, while Compositional Visual Generation with Composable Diffusion Models demonstrates the ability to create complex, composite visuals. These applications underscore the adaptability of diffusion models to meet domain-specific needs.

Research in m**ultiple types of diffusion models** category explores various diffusion model architectures and methodologies to enhance image generation capabilities. Multilevel Diffusion, Palette, Cascaded Diffusion Models, High-Resolution Image Synthesis with Latent Diffusion Models, and D2C highlight innovations that improve fidelity, resolution, and flexibility of generated images.

These studies contribute to a deeper understanding of the theoretical underpinnings and practical implementations of diffusion models.

Recommendations

Based on our comprehensive review, we propose the following recommendations for future research and development in the field of diffusion-based image generation.

Future research should focus on optimizing diffusion models to reduce computational requirements and improve generation speed. This could involve developing more efficient algorithms or leveraging hardware advancements.

Despite significant progress, there is room to enhance the accuracy and detail of images generated from textual descriptions. Integrating more advanced natural language processing techniques and expanding training datasets could yield better results.

Researchers should explore additional domains where diffusion models can be applied. Potential areas include autonomous vehicle navigation, augmented reality, and fashion design. Tailoring diffusion models to these specific applications could drive innovative solutions.

Collaboration between experts in machine learning, domain-specific fields, and industry practitioners can foster the development of more practical and impactful applications of diffusion models. Interdisciplinary research can also help in addressing real-world challenges and enhancing model robustness.

As diffusion models become more powerful, it is crucial to consider the ethical implications of their use. Researchers should prioritize transparency, fairness, and accountability in model development to mitigate potential misuse and ensure that these technologies benefit society as a whole.

In conclusion, diffusion models for image generation represent a rapidly evolving and highly promising area of research. By addressing current limitations and exploring new applications, these models have the potential to revolutionize various industries and contribute significantly to technological advancements.

Referece

[1] Ho, Jonathan, et al. "Cascaded Diffusion Models for High Fidelity Image Generation." Cornell University - arXiv,Cornell University - arXiv, May 2021.
[2] Epstein, Dave, et al. Diffusion Self-Guidance for Controllable Image Generation
[3] Gu, Shuyang, et al. "Vector Quantized Diffusion Model for Text-to-Image Synthesis." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, https://doi.org/10.1109/cvpr52688.2022.01043.

[4] Croitoru, Florinel-Alin, et al. Diffusion Models in Vision: A Survey. Sept. 2022.
[5] Nichol, Alex, et al. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models.
[6] Saharia, Chitwan, et al. "Palette: Image-to-Image Diffusion Models." Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings, 2022, https://doi.org/10.1145/3528233.3530757.
[7] Liu, Nan, et al. Compositional Visual Generation with Composable Diffusion Models. June 2022.
[8] Pinaya, Walter HL, et al. "Brain imaging generation with latent diffusion models." MICCAI Workshop on Deep Generative Models. Cham: Springer Nature Switzerland, 2022.
[9] Hagemann, Paul, et al. Multilevel Diffusion: Infinite Dimensional Score-

Based Diffusion Models for Image Generation. Mar. 2023.
[10] Zhang, Chenshuang, et al. Text-to-Image Diffusion Model in Generative AI: A Survey. Mar. 2023.
[11] Kim, Gwanghyun, et al. “DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, https://doi.org/10.1109/cvpr52688.2022.00246.
[12] Saharia, Chitwan, et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding.
[13] Sinha, Abhishek, et al. “D2C: Diffusion-Decoding Models for Few-Shot Conditional Generation.” Neural Information Processing Systems,Neural Information Processing Systems, Dec. 2021.
[14] Xiang, Jianfeng, et al. 3D-Aware Image Generation Using 2D Diffusion Models. Mar. 2023.