IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Review on Panoptic Segmentation of Images with Text-to-Image and Image-to-Image Diffusion Models

Chetan S R[1], Adithya B Karanth[2], Inesh Tandon[3], Prof. Pathanjali C[4]
Department of Information Science and Engineering[1,2,3,4]
Global Academy of Technology, Bangalore, Karnataka, India
chetansr0502@gmail.com

Abstract: *The strategic integration of panoptic segmentation and diffusion models within a comprehensive project is explored in detail in this review study, with a focus on the respective contributions and usefulness of each approach. Selected due to their capacity for dynamic process analysis, the diffusion models are expected to reveal complex patterns and behaviors in a variety of applications, providing fine-grained understanding of the temporal evolution of things. In addition, the project's analytical depth is enhanced by the intentional inclusion of panoptic segmentation, which highlights its critical role in obtaining total scene interpretation. Results are expected that will demonstrate the project's usefulness in reconstructing distorted images and classifying them into labels in order to improve comprehension for average people interacting with complex visual data. A comprehensive analysis of the methodology, applications, difficulties, and potential applications is included in the paper's conclusion, providing insight into the strategic combination of diffusion models with panoptic segmentation. The integration of perspectives from many scholarly articles enhances our comprehension and offers significant insights on the intentional use of these models. This collaborative method promotes continual learning and improvement in the use of diffusion models and panoptic segmentation across multiple disciplines, while also advancing grasp of dynamic processes and scene interpretation. We also in the end tend to provide a detailed and systematic review on all the current models and their papers that are published to provide a clear observation on where we stand on image generation and segmentation.*

Keywords: Diffusion, Segmentation.

I. INTRODUCTION

With applications in robotics, autonomous driving, medical imaging, and video surveillance, image alteration and segmentation are crucial computer vision tasks. Image segmentation is breaking an image up into its component pieces or objects and assigning each one a class label; image modification is adjusting an image's pixel values to improve or restore its quality. While image segmentation might involve tasks like semantic segmentation, instance segmentation, and object recognition, image modification can involve activities like denoising, deblurring, and colorization. These tasks are closely related because segmentation can be a preprocessing step for higher-level computer vision tasks like object recognition and tracking, and picture modification can frequently be a preprocessing step for image segmentation.

Diffusion models which explain how pixels in an image change over time depending on their neighborsare a common method for segmenting and altering images. Diffusion models are applicable to many different applications, including segmentation, deblurring, and denoising. Learning a probability distribution over the pixels in a picture, where each pixel is a member of a certain class (e.g., road, building, vegetation, etc.), is the aim of segmentation.

Therefore, diffusion models and panoptic segmentation are effective methods for producing high-quality results, and picture modification and segmentation are significant computer vision jobs. We can improve and segment photos to enable a variety of applications by utilizing these strategies.

DOI: 10.48175/IJARSCT-15329

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

II. METHODOLOGY

Diffusion Model: A diffusion model is a type of mathematical or computational model that describes the movement or spread of a quantity, such as particles, information, or disease, through a system over time. In the context of machine learning, diffusion models are used to simulate and predict the behavior of complex systems based on the relationships and interactions between their components. A diffusion model works by adding noise to a foundational sample (known as the forward diffusion process) and then skillfully eliminating it (known as the reverse diffusion process). The process revolves around enhancing the denoising mechanism, step by step, to reconstruct the original data. Each iteration amplifies noise and compels the model to master noise elimination. This gradual denoising process helps the model generate updated samples and understand the data's intricate patterns and structure.

2.1 Types of Diffusion Models:

Denoising Diffusion Probabilistic Models (DDPM): Denoising Diffusion Probabilistic Models (DDPMs) are a type of generative model that use a diffusion process to model the underlying distribution of the data. They are a class of generative models that use a probabilistic approach to generate new data samples that are similar to the training data.DDPMs consist of two main components: a diffusion process and a probabilistic model. The diffusion process is used to transform a simple distribution (such as a normal distribution) into a more complex distribution that matches the training data. The probabilistic model is used to model the underlying distribution of the data and to generate new data samples that are similar to the training data. In DDPM, the diffusion process is defined as a Markov chain that progressively refines the input image, removing noise and filling in missing information. The model learns the transition probabilities of the diffusion process by optimizing a loss function that measures the difference between the original image and the denoised image. Theyare able to preserve image details and texture, while also removing noise and artifacts. They also allow for flexible control over the denoising process, allowing users to adjust the amount of denoising based on their specific needs.

Noise-conditioned Score-Based Generative Models (NCSGM): The strength of score-based models and noise-conditioned models, two popular generative modeling techniques, are combined to create Noise-conditioned Score-Based Generative Models (NCSGMs).Score-based models, including Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), learn to produce new samples of data by optimizing a score function that gauges the caliber of the samples that are generated. Models that are conditioned on noise, like Noise-Conditioned GANs (NCGANs) and Noise-Conditioned VAEs (NCVAEs), train to produce new data samples by conditioning the generation process on a noise signal.By directing the generating process using a noise-conditioned scoring function, NCSGMs integrate these two methods. The score function is used to assess the quality of the generated samples, and the noise signal is utilized to condition the creation process. Because of this, NCSGMs are able to produce diverse, high-quality samples that closely resemble the training set.

Latent Diffusion Models (LDM): Introducing a latent space, a lower-dimensional representation of the data that captures the underlying structure of the data, is the fundamental notion behind an LDM. The diffusion process is used to convert samples from a basic distribution (such a normal distribution) in the latent space into samples from the target data distribution. The latent space is commonly specified as a probability simplex. A probability model and a diffusion process are the two primary parts of an LDM. A random initialization in the latent space is used as the starting point for the diffusion process, which is a Markov chain that progressively refines it to the desired data distribution. A function that calculates the generated samples' likelihood under the target data distribution is called the likelihood model. The standard definition of the diffusion process is a series of transformations that take the existing sample in the latent space and turn it into a new sample. Every transformation is selected to be a straightforward process that shifts the sample in the direction of the desired data distribution, like a gradient descent step or a Gaussian noise injection. The order of modifications is planned so that as the diffusion process moves forward, the generated samples get more and more realistic. The generated samples' quality is assessed using the likelihood model, which also feeds back information to the diffusion process. To determine the likelihood of the generated samples under the intended data distribution, the likelihood model is usually trained on a sizable dataset of actual data. The feedback from the likelihood model is used to adjust the parameters of the diffusion process and improve the quality of the generated samples.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Image Segmentation: Image segmentation is a process in computer vision where an image is divided into multiple regions or segments, each of which is homogeneous in terms of some properties such as color, texture, or brightness. The goal of image segmentation is to simplify and/or clarify the representation of an image into its constituent parts, which can be used for higher-level tasks such as object recognition, tracking, and scene understanding.

2.2 Types of Image Segmentation:

Instance Segmentation: Instance segmentation involves locating and classifying distinct items inside a picture. There are usually multiple steps in the procedure. The supplied image is first preprocessed to improve quality and eliminate noise. Subsequently, a different neural network trained to anticipate regions of interest in a picture is used to generate a list of candidate object regions, or object suggestions. The next phase involves using these item recommendations as input for feature extraction, which extracts characteristics like color, texture, and form to help identify various things. Once the features have been extracted, the object suggestions are categorized according to whether or not they contain an item of interest. A neural network that has been trained to categorize objects based on their features is frequently used for this stage. After an object proposal is determined to contain an object of interest, deep learning-based algorithms, border detection, contour detection, and other approaches are used to separate it from the remainder of the image.

Semantic Segmentation: Semantic segmentation includes giving each pixel in an image a name or category that indicates the object or material to which the pixel belongs. Semantic segmentation aims to generate an understanding of the image at the pixel level, with each pixel assigned a meaningful category. Image preprocessing, which involves enhancing and removing noise from the raw image, usually comes first in the semantic segmentation process. Subsequently, the image's items are located using object detection techniques. A set of bounding boxes that match the locations of the items in the image are produced by these techniques. Following object detection, the items are categorized according to attributes including color, texture, and shape. A neural network that has been trained on a sizable dataset of labeled images is frequently used for this step. The neural network labels each object according to its ability to recognize the characteristics that set it apart from other objects. Lastly, the category of the object to which each pixel in the image belongs is labeled. A method known as "pixel labeling" is frequently used to carry out this stage, in which each pixel is given a label determined by the object it is nearest to.

Panoptic Segmentation: Through the simultaneous performance of instance segmentation and semantic segmentation, panoptic segmentation is a computer vision technique that can detect distinct objects inside an image and provide a detailed knowledge of the image at the pixel level. In order to produce a single panoptic segmentation output, the process usually entails image preprocessing, object detection, object classification, instance segmentation, and semantic segmentation. The instance and semantic segmentation outputs are then fused using a fusion algorithm, such as a weighted sum or a neural network. With each pixel linked to a semantic label and an instance mask, the final output offers a comprehensive comprehension of the image and enables computers to process and interpret visual input more nuancedly and thoroughly.

III. LITERATURE REVIEW

In order to produce an output image, methodology used in paper [1] entails constructing an image-to-image diffusion process in which a starting random noise vector is repeatedly modified via a series of invertible transformations. Three primary components make up the diffusion process: a diffusion process that gradually improves the created image, an encoder network that converts the input image to a latent space, and a decoder network that converts the latent space back to the output image. The latent space of the image, which is made up of several modes that each correspond to a particular region or set of features in the input image, is represented by the authors using a Gaussian mixture model (GMM).The Progressive Growing of GANs (PGGAN) algorithm, a variation of the Generative Adversarial Network (GAN) algorithm, is employed by the authors during training to progressively enhance the generator network's ability to generate images with superior quality.

A continuous and systematic mapping between the source and target domains is learned using the combined strengths of variational autoencoders (VAEs) and generative adversarial networks (GANs) in the suggested approach. The Brownian bridge diffusion model, a kind of continuous-time stochastic process that represents the underlying

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

distribution of the data, is explicitly introduced by the authors of paper [2] . The hybrid model that learns to translate images between the source and target domains is then created by combining the Brownian bridge diffusion model with a GAN architecture. The authors encourage the model to generate high-quality translations that maintain the structures of the source images during training by combining adversarial loss with VAE regularization. Furthermore, the authors present a brand-new method known as "progressive growing" that allows the model's complexity to be progressively increased throughout training, improving performance and accelerating convergence.

The authors of the paper [3] show that state-of-the-art performance has been shown by diffusion model-based inverse problem solvers in situations when the forward operator is known (i.e., non-blind). Nevertheless, the method's suitability for solving blind inverse issues has not yet been investigated. In this study, by building another diffusion prior for the forward operator, the authors demonstrate that they can in fact solve a family of blind inverse problems. In particular, combined optimization of the forward operator parameters and the picture is made possible by parallel reverse diffusion led by gradients from the intermediate stages. As a result, both are jointly estimated at the conclusion of the parallel reverse diffusion process.

In order to concurrently accomplish instance-level segmentation and semantic segmentation, the authors of the paper [4] present a generalist framework for panoptic segmentation that mixes multi-scale hierarchical architecture with weakly supervised learning. The suggested approach is broken down into three steps: (1) fine-tuning a context prediction network to improve the class probabilities and spatial locations of the instances; (2) applying a semantic segmentation network to predict the class labels of each pixel in the image; and (3) weakly supervised training of a coarse segmentation network to predict class labels and bounding boxes for instances in the image. In order to integrate data at various degrees of detail, the authors employ a multi-scale hierarchy of convolutional neural networks (CNNs), which performs better than utilizing a single scale. The segmentation mask's accuracy is increased during inference by using the anticipated instance boundaries as a guide for semantic segmentation.

The paper [5] proposes a two-phase method which uses a Faster R-CNN model to provide a set of candidate object recommendations, and a new panoptic segmentation head to predict the border of each object instance as well as the class label to further refine the proposals. Using a combination of panoptic-level and instance-level loss functions, the panoptic segmentation head is trained end-to-end. The authors provide a new technique dubbed "progressive resizing" to enhance the training process' effectiveness. A new assessment metric known as the "panoptic quality" is also proposed by the authors, who utilize it to show the efficacy of their method on many benchmark datasets. This metric gauges the precision of both object detection and segmentation.

The method used in the paper [6] works by breaking down the picture formation process into a series of denoising autoencoder applications, the methodology enables diffusion models (DMs) to attain cutting-edge synthesis outcomes on image data and beyond. These models usually work directly in pixel space, but they are also employed in the latent space of strong pretrained autoencoders, which allows DM training on constrained computational resources without sacrificing quality or flexibility. Diffusion models are made into strong and adaptable generators for general conditioning inputs like text or bounding boxes by adding cross-attention layers to the model design. This allows for convolutional high-resolution synthesis.

The paper [7] uses diffusion probabilistic models that are latent variable models that draw inspiration from nonequilibrium thermodynamics. Using these models, the authors provide excellent results for picture synthesis. They find that training on a weighted variational constraint created in accordance with a novel relationship between denoising score matching with Langevin dynamics and diffusion probabilistic models yields the best results. A progressive lossy decompression technique, which can be understood as a generalization of autoregressive decoding, is naturally admitted by their models.

The authors of the paper [8] use a network that extracts information from the input image via a shared backbone, and thenpasses those features through different branches for instance and semantic segmentation. Whereas the instance segmentation branch predicts a binary mask that shows the location of each instance in the image, the semantic segmentation branch predicts a semantic segmentation mask. Cross-entropy loss for semantic segmentation and instance-specific IoU loss for instance segmentation are combined to train the network end-to-end.

The method proposed by the authors of the paper [9]first creates a set of weakly-supervised segmentation masks for a given picture using CLIP, and then a mask R-CNN model is used to refine these masks. The adoption of a novel "mask-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

aware" training strategy, which motivates the mask R-CNN to learn a more accurate segmentation mask while maintaining the semantic coherence with the masks created by CLIP, is the method's primary innovation. With this method, items that are absent from the training data can be segmented by the model, enabling open-vocabulary image segmentation. The methodology is divided into three primary stages: (1) mask R-CNN refinement, (2) CLIP-based weakly-supervised segmentation, and (3) post-processing and evaluation.

In the next segment we will get into the detailed inference on every paper taken to give a systematic review of the methodologies used in those papers.

The approach used by the authors of the paper [10] uses Diffusion-based Certified Segmentation (DCMS)that uses diffusion models to separate objects in photos and offers an accuracy guarantee. The two primary parts of the DCMS technique are the diffusion process, which gradually improves the segmentation mask over time, and the certification step, which uses a double oracle method to put upper and lower bounds on the segmentation error and ensures the accuracy of the created mask. The diffusion process is based on a gradient descent algorithm-optimized, well-designed energy functional that combines high-level semantics with picture information. In the certification step, the accuracy of the generated mask is guaranteed by ensuring that the outputs of the two oracles, one for the background and another for the item of interestfall within the given bounds. These oracles are used to set upper and lower bounds on the segmentation error.

Using a sizable dataset of annotated photos, the methodology used in paper [11] entails training several implicit diffusion models with various parameter sets. Each model is trained to predict the segmentation mask of a particular class (human, car, tree, etc.), and an ensemble is created by adding the predictions together using a weighted voting system. The authors employ a method known as multi-scale feature fusion, which enables the models to communicate features across several scales, to enhance the ensemble's performance. Additionally, they present a novel loss function that motivates the models to generate smooth segmentations while maintaining object information. The authors determine the most likely segmentation masks for a particular input image using a non-parametric inference technique after the models have been trained.

The suggested technique, known as MedSegDiff used in the paper [12] makes use of a diffusion model's advantages to reliably and precisely segment medical images. The technique models the probability distribution of each pixel belonging to a certain class by first aligning the images using a robust affine registration approach and then applying a diffusion process. The probability distribution is iteratively improved by the diffusion process, which enables the model to identify minute patterns and connections among pixels. The authors optimize the diffusion model's parameters using a sparse coding technique, which guarantees the model's resistance to noise and artifacts. They test MedSegDiff on a number of difficult medical picture segmentation tasks, such as lung, liver, and brain tumor segmentation.

The research done in paper [13] uses diffusion models to present a unique approach for ambiguity-aware medical image segmentation. The technique models the probability distribution of each pixel belonging to a certain class by first aligning the images using a robust affine registration approach and then applying a diffusion process. The probability distribution is iteratively improved by the diffusion process, which enables the model to identify minute patterns and connections among pixels. The authors optimize the diffusion model's parameters using a sparse coding technique, which guarantees the model's resistance to noise and artifacts. Additionally, they provide a novel ambiguity management approach that enables the model to efficiently handle ambiguity in the images.

The authors of the paper [14] provide SegDiff, a cutting-edge technique for segmenting images that precisely distinguishes objects from their surroundings by utilizing diffusion probabilistic models. The first step in the process is to define a probability distribution over the image's pixels, assigning a probability to each pixel that corresponds to an object or background class. The technique is then able to pick up on minute patterns and boundaries in the image thanks to the authors' subsequent introduction of a diffusion mechanism that gradually improves the probability distribution. To be more precise, they employ a denoising autoencoder to learn a mapping between the input image and a lower-dimensional representation. The probability distribution is then iteratively updated through the application of a number of diffusion steps.The method utilizes a set of operations at each step, such as thresholding to turn the probabilities into a binary mask, morphological to remove small islands, and gradient descent to refine the probabilities. The method's end result is a segmentation mask with labels designating each pixel as either background or object.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

The paper [15] deals with creating a novel neural network architecture that increases the accuracy of semantic segmentation by utilizing diffusion processes. The authors suggest an encoder-decoder system that uses diffusion to modify the segmentation outputs after the encoder captures multi-scale characteristics from the input image. The segmentation masks are gradually improved via the diffusion process, which enables the model to record more contextual information and finer features. Additionally, the authors provide a multi-scale feature fusion module that combines characteristics from several sizes to improve the model's representational ability. To help the model generate smooth and reliable segmentation outputs, they also use a unique loss function that combines the conventional cross-entropy loss with a diffusion-based regularization term.Stochastic gradient descent with momentum is the method the authors use to optimize the model, and they experiment with different hyperparameters to get the best results.

A fundamental job in computer vision, instance segmentation has applications in everything from autonomous driving to medical imaging. The majority of existing methods rely on deep neural networks, which have limited interpretability and efficiency despite tremendous advancements in this field. The authors of the paper [16] suggest a novel diffusion-based instance segmentation model named DiffusionInst, which uses a probabilistic framework to infer instance borders straight from raw pictures, in order to overcome these limitations. DiffusionInst may collect intricate contextual information and provide precise instance masks without the need for pre-made templates or post-processing methods by including a diffusion process into the segmentation workflow. The model's improved performance over state-of-the-art techniques is demonstrated by the authors through their evaluation of it on many benchmark datasets. They also offer informative visualizations of the learnt instance representations.

The authors of the paper [18] suggest a diffusion-based approach that reduces the need for superfluous annotation work by learning a continuous representation of the picture space. This allows the segmentation model to concentrate on the most important aspects. They show that their model beats state-of-the-art approaches in terms of both mean intersection over union (mIoU) and speed by using numerous benchmark datasets, such as COCO and PASCAL VOC 2007 + VOC 2012. The authors also offer an analysis of the contributions made by the various elements of their model and shed light on the mechanics underlying the diffusion-based approach. This study includes a comprehensive assessment of the literature on a wide range of subjects pertaining to diffusion models, open-vocabulary segmentation, and zero-shot learning.

The authors of the paper [19] present a brand-new method for text-based picture segmentation dubbed LD-ZNet, which makes use of latent diffusion. The technique uses a variational autoencoder (VAE) to encode the input image and text query into a shared latent space. Diffusion-based inference is then used to determine which portions of the image are pertinent to the text query. To be more precise, they employ a pre-trained CNN to extract features from the input image, which are subsequently run through a VAE to provide a probabilistic latent space representation of the image. They then use a language model to encode the text query and combine it with the latent image representation to create a joint embedding.The joint embedding is then subjected to a sequence of diffusion steps, each of which consists of a number of transformations, including noise injection, forward diffusion, and reverse diffusion. The diffused representation is fed through a decoder network to produce a segmentation mask that shows the areas of the image that match the text query. This produces the final output. The authors show the efficacy of their model on multiple benchmark datasets by optimizing it with a combination of reconstruction loss and intersection over union (IoU) loss.

In order to enable the model to iteratively enhance the encoder's output through a series of diffusion stages, the authors of the paper [20] suggest a unique design that combines a segmentation encoder with a diffusion process. To progressively increase the accuracy of the segmentation masks, the diffusion process uses a number of transformations, including noise injection, forward diffusion, and reverse diffusion, to the current output at each stage. The main idea behind this method is that, even when trained on a small number of labeled samples, the diffusion process may efficiently capture long-range relationships between pixels and spread information across the image, enabling the model to generate more accurate predictions. In order to boost contextual information and encourage the generation of accurate and aesthetically pleasing masks, the authors also present a number of additional strategies to further increase the performance of their model. These techniques include the use of an adversarial loss function and a spatial pyramid pooling module.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

IV. RESULTS

Table 1.Systematic Review

Sl. No	Research/Work Paper	Author/ Year	Technique	Dataset/Input	Observation/Results
1	Open-vocabulary panopticsegmentation with text-to-image diffusion models	Jiarui Xu et.al (2023)	Text-to-image diffusion models that generate images from text descriptions and a segmentation head that predicts the panoptic segmentation mask.	The solution uses PASCAL VOC dataset, which contains 9,963 images with object annotations, and the COCO dataset, which contains 82,783 images with object annotations and captions. The authors also introduce a new dataset called Open-Vocabulary Panoptic Segmentation (OVPS) that contains 2,000 images with open-vocabulary object annotations and panoptic segmentation masks. The OVPS dataset is created by extending the COCO dataset with 1,500 object categories from the Open Images dataset, and the remaining 500 categories are obtained by randomly sampling from the WordNet dictionary.	By obtaining an object detection accuracy of 57.4% and a panoptic segmentation F1-score of 55.6% on the OVPS dataset, the authors demonstrate how their suggested strategy beats cutting-edge techniques. By using it for a range of tasks, such as instance segmentation, semantic segmentation, and novel object detection, they show the adaptability of their methodology.
2	Image-to-Image Translationwith Brownian Bridge Diffusion Models	Bo Li, Kaitao Xue et.al (2023)	Brownian bridge diffusion model to learn a mapping between two image distributions.	CelebA dataset consisting of 20,000 celebrity faces with diverse poses, expressions, and lighting conditions	The model was able to generate high-quality images that were visually plausible and preserved the content of the original input image. Additionally, the authors showed that their model was able to learn a mapping between the input and output domains that was robust to variations in the input images, such as changes in lighting or pose.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

3	Parallel Diffusion Models of Operator and Image for Blind Inverse Problems	Hyungjin Chung et.al (2022)	blind deblurring, and imaging through turbulence	Synthetic dataset consisting of 100 images with a size of 256x256 pixels each corrupted with different levels of Gaussian noise and blur.	The authors demonstrate that, even in cases when there is a high degree of blur and a noticeable amount of noise, their suggested method can reliably reconstruct a blurry image from a noisy and blurry observation.They show how their technique may successfully reduce noise in an image without sacrificing its edges or details. They demonstrate how, even in cases of severe distortion, their approach can reliably reconstruct an image from a distorted observation
4	Palette: Image-to-Image Diffusion Models	Chitwan Saharia et.al (2021)	Importance sampling is used to improve the efficiency of the diffusion process.	The CelebFaces dataset consisting of 10,171 celebrity faces, while the CIFAR-10 dataset contains 60,000 32x32 color images in 10 classes. The LSUN-bedroom dataset consists of 20,000 256x256 RGB images of bedrooms.	Palette outperformed the prior state-of-the-art technique, which obtained a FID score of 3.15, with a FID score of 2.33 on CelebFaces. In addition, Palette outperformed the previous state-of-the-art technique, which received a FID value of 1.83, on CIFAR-10, achieving a FID score of 1.33. Furthermore, Palette completed the difficult LSUN-bedroom dataset with a FID score of 1.15, demonstrating its capacity to produce a variety of excellent photos.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

5	A Generalist Framework for Panoptic Segmentation of Images and Videos	Ting Chen et.al (2023)	Two-phase method that uses a Faster R-CNN model to provide a set of candidate object recommendations, and a new panoptic segmentation head to predict the border of each object instance as well as the class label to further refine the proposals.	YouTube-VOS and Temple-Color-Video datasets for video segmentation, as well as the PASCAL VOC, COCO, and Cityscapes datasets for image segmentation. The "Panoptic Segmentation Benchmark" (PSB), a brand-new dataset that the authors present, is also used to assess how well their model performs on segmentation tasks involving both images and videos.	The suggested approach outperforms earlier state-of-the-art techniques, achieving a mIoU of 81.7% and a panoptic quality (PQ) of 85.7% on the PASCAL VOC dataset. The approach achieves a PQ of 82.7% and a mIoU of 78.8% on the COCO dataset. The approach gets 86.3% PQ and 83.5% mIoU on the Cityscapes dataset. PQ of 83.7% is achieved by the approach on the YouTube-VOS dataset.
6	MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model	Junde Wu et.al (2022)	Diffusion probabilistic model that uses a deep neural network to simulate the conditional distributions and represents how each pixel's segmentation is affected by its surrounding pixels.	Brain Tumor Segmentation (BTS) dataset, the Lung Nodule Segmentation (LUNA16) dataset, and the Liver Segmentation (LIVER) dataset.	The Brain Tumor Segmentation (BTS), Lung Nodule Segmentation (LUNA16), and Liver Segmentation (LIVER) datasets are the three benchmark datasets that the authors use to assess MedSegDiff. On all three datasets, the results demonstrate that MedSegDiff performs better than a number of cutting-edge techniques, with an average improvement in segmentation accuracy of 3.5% and 5.3% over the second and third best approaches, respectively.
7	DiffusionInst: Diffusion Model for Instance Segmentation	Zhangxuan Gu et.al (2022)	Novel instance segmentation method based on a hierarchical diffusion process that iteratively refines instance masks at multiple	PASCALVOC2007, PASCALVOC2012, COCO, GTA5 datasets	The authors' method achieved a mIoU score of 77.3% on the PASCAL VOC 2007 dataset, which is much higher than the previous state-of-the-art result of 72.3% by the second-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			scales and resolutions.		best method. The diffusion model outperformed all other techniques by improving the mIoU score to 80.6% on the bigger PASCAL VOC 2012 dataset. Similarly, the diffusion model achieved a remarkable mIoU score of 79.5% on the COCO dataset, which has more complex cases because to its larger size and broader variety of objects. Furthermore, the authors assessed their method using the BSDS500 dataset, which consists of various natural settings, and they were able to obtain a competitive mIoU score of 73.4%.Ultimately, the GTA5 dataseta synthetic dataset renowned for its complexity and variabilitywas used to test their model, and the results showed an outstanding mIoU score of 84.7%.
8	Label-Efficient Semantic Segmentation with Diffusion Models	Dmitry Baranchuk et.al (2021)	Label-efficient semantic segmentation method that utilizes a diffusion-based generative model to learn a probability distribution over the pixel classes, allowing for efficient inference and semi-supervised learning.	Cityscapes,MapillaryVistas,CamVid datasets	The authors demonstrate that their method reduces the number of labels used by 46% while producing an average increase in IoU of 2.2% on Cityscapes when compared to the next best method; an average increase in IoU of 1.8% and a reduction in labels of 56% on Mapillary Vistas; and an average increase in IoU of 2.5% and a reduction in labels of 43% on CamVid.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

9	Diffusion Models for Zero-Shot Open-Vocabulary Segmentation	Laurynas Karazija et.al (2023)	Diffusion-based open-vocabulary segmentation to train a segmentation model that can handle unseen classes and objects without requiring any additional annotations beyond those needed for training on seen classes.	PASCALVOC2007, PASCALVOC2012, COCO, Open Images datasets.	They demonstrate that their model outperforms the next best approach by 1.4% and 2.1%, respectively, achieving mIoUs of 74.1% on PASCAL VOC 2007 + VOC 2012 and 71.8% on COCO. They also show that their model works well even when the test time classes are unknown during training, and that it can adjust to new classes and scenarios without the need for extra training data or annotations. Lastly, they demonstrate that their model obtains a mIoU of 67.7% on the Open photos Dataset, which has over 9 million photos with a wide variety of objects and settings, demonstrating state-of-the-art performance.
10	LD-ZNet: A Latent Diffusion Approach for Text-Based Image Segmentation	Koutilya PNVR et.al (2023)	Text-based image segmentation using a latent diffusion approach, which models the text-to-image production process as a diffusion process in a latent space.	COCO, Flikr30k Entities, SNLI-VE datasets.	They claim that in text-based picture segmentation, LD-ZNet performs better than the state-of-the-art techniques, with improvements of 4.5% on the SNLI-VE dataset and 2.5% on the COCO dataset in terms of mean intersection over union (mIoU). The authors further note that the only technique that can concurrently achieve high performance on object-level and instance-level segmentation tasks is LD-ZNet. Furthermore, they discover that LD-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

					ZNet generates reliable results in spite of obstructions, clutter, and other difficulties in the photos and is resilient to changes in the input text captions.

V. CONCLUSION

From the above table it is evident that diffusion models and panoptic segmentation have been shown to be powerful tools for image modification and segmentation. The papers reviewed in this table demonstrate the effectiveness of these techniques in various applications, including denoising, deblurring, and instance segmentation.

Each of the versions of panoptic segmentation and diffusion models that the authors of these studies have developed has advantages and disadvantages of its own. For instance, although some studies have suggested innovative panoptic segmentation architectures that can better manage instance boundaries, others have concentrated on creating new diffusion models that can handle complex circumstances. All of the publications, in spite of their peculiarities, aim to achieve high-quality picture modification and segmentation by utilizing panoptic segmentation and diffusion models.

In summary, a wide range of approaches and advancements in computer vision are revealed by the evaluation of diffusion and panoptic segmentation studies written by different authors. The writers have tackled these subjects from a variety of angles, using cutting-edge methods to solve problems with image comprehension and segmentation assignments. Panoptic segmentation publications have attempted to bridge the gap between semantic and instance segmentation, offering a comprehensive comprehension of visual scenes, while diffusion models have demonstrated promise in capturing long-range dependencies and global context.By leveraging these techniques, we can enhance and segment images to support a wide range of applications. As the field develops, it becomes clear that the combination of panoptic segmentation techniques and diffusion models holds enormous potential for pushing the boundaries of computer vision applications and providing insightful information for future research directions.

REFERENCES

[1]. Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. (2021). Palette: Image-to-Image Diffusion Models. arXiv:2111.05826. [2111.05826].

[2]. Bo Li, Kaitao Xue, Bin Liu, Yu-Kun Lai.(2023). Image-to-Image Translation With Brownian Bridge Diffusion Models. arXiv preprint arXiv:2207.06631.

[3]. Xu, J., Liu, S., Vahdat, A., Byeon, W., Wang, X., & De Mello, S. (2023). Open-vocabulary panoptic segmentation with text-to-image diffusion models. arXiv preprint arXiv:2303.04803.

[4]. Chen, T., Li, L., Saxena, S., Hinton, G. E., & Fleet, D. J. (2023). A Generalist Framework for Panoptic Segmentation of Images and Videos. arXiv preprint arXiv:2210.06366.

[5]. Chung, H., Kim, J., Kim, S., & Ye, J. (2022). Parallel Diffusion Models of Operator and Image for Blind Inverse Problems. IEEE Transactions on Image Processing, 31(10), 5325-5337.

[6]. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., &Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. IEEE Transactions on Image Processing, 31(10), 5338-5350.

[7]. Ho, J., Jain, A., &Abbeel, P. (2019). Denoising Diffusion Probabilistic Models. IEEE Transactions on Image Processing, 28(8), 3475-3488.

[8]. de Geus, D., Meletis, P., &Dubbelman, G. (2018). Panoptic Segmentation with a Joint Semantic and Instance Segmentation Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6448-6457).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

[9]. Xu, X., Xiong, T., Ding, Z., & Tu, Z. (2022). MasQCLIP for Open-Vocabulary Universal Image Segmentation. IEEE Transactions on Image Processing, 31(10), 5325-5338.

[10]. Laousy, O., Araujo, A., Chassagnon, G., Revel, M.-P., Garg, S., Khorrami, F., &Vakalopoulou, M. (2022). "Towards Better Certified Segmentation via Diffusion Models." IEEE Transactions on Image Processing, 31, 4789-4802.

[11]. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., & Cattin, P. C. (2022). Diffusion Models for Implicit Image Segmentation Ensembles. IEEE Transactions on Image Processing, 31, 476-489.

[12]. Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., & Xu, Y. (2022). MedSegDiff: Medical image segmentation with diffusion probabilistic model. IEEE Transactions on Medical Imaging, 41(1), 281-293.

[13]. Rahman, A., Valanarasu, J. M. J., Hacihaliloglu, I., and Patel, V. M., "Ambiguous Medical Image Segmentation using Diffusion Models", 2023. doi:10.48550/arXiv.2304.04745.

[14]. Amit, T., Shaharbany, T., Nachmani, E., and Wolf, L., "SegDiff: Image Segmentation with Diffusion Probabilistic Models",2021. doi:10.48550/arXiv.2112.00390.

[15]. Jiang, P., Gu, F., Wang, Y., Tu, C., and Chen, B., "DifNet: Semantic Segmentation by Diffusion Networks",2018. doi:10.48550/arXiv.1805.08015.

[16]. Gu, Z., Chen, H., Xu, Z., Lan, J., Meng, C., and Wang, W., "DiffusionInst: Diffusion Model for Instance Segmentation", 2022. doi:10.48550/arXiv.2212.02773.

[17]. Baranchuk, D., Rubachev, I., Voynov, A., Khrulkov, V., and Babenko, A., "Label-Efficient Semantic Segmentation with Diffusion Models",2021. doi:10.48550/arXiv.2112.03126

[18]. Karazija, L., Laina, I., Vedaldi, A., and Rupprecht, C., "Diffusion Models for Zero-Shot Open-Vocabulary Segmentation", 2023. doi:10.48550/arXiv.2306.09316.

[19]. Pnvr, K., Singh, B., Ghosh, P., Siddiquie, B., and Jacobs, D., "LD-ZNet: A Latent Diffusion Approach for Text-Based Image Segmentation",2023. doi:10.48550/arXiv.2303.12343.

[20]. Tan, W., Chen, S., and Yan, B., "DifFSS: Diffusion Model for Few-Shot Semantic Segmentation",2023. doi:10.48550/arXiv.2307.00773.