A Survey of Image Segmentation Techniques and Models

www.surveyx.cn

Abstract

Image segmentation is a fundamental task in computer vision, crucial for partitioning images into meaningful segments to enhance analysis and interpretation. This survey paper explores the landscape of image segmentation, categorizing it into instance, semantic, and panoptic segmentation, each serving distinct analytical purposes. The survey delves into key models such as SAM2, SegFormer, BiRefNet, DeepLab, and Mask2FormerForUniversalSegmentation, highlighting their architectural innovations and contributions to segmentation accuracy and efficiency. Comparative analyses reveal the strengths and limitations of these models, emphasizing performance metrics like mean Intersection over Union (mIoU) and Dice Similarity Coefficient (DSC) for evaluating efficacy. The paper also addresses challenges in scalability, efficiency, and adaptability, underscoring the importance of ongoing research to refine and enhance segmentation methodologies. Applications in healthcare, autonomous driving, and remote sensing are examined, showcasing the transformative impact of segmentation technologies in these fields. Future directions emphasize the integration of transformer-based architectures, advancements in unsupervised learning, and the exploration of new segmentation paradigms to address current limitations. This comprehensive survey underscores the critical role of image segmentation in advancing computer vision, highlighting its significance across diverse applications and its potential for future innovation.

1 Introduction

1.1 Importance of Image Segmentation

Image segmentation is fundamental in digital image processing and computer vision, serving as a critical component for effective object recognition in applications such as face detection [1]. In medical imaging, precise segmentation of anatomical structures is essential for accurately delineating brain tumors, which enhances model generalization on Out-of-Distribution (OOD) tasks [2] and aids in the detection of small, irregularly shaped tumors, thereby contributing significantly to computer-aided diagnosis.

Beyond healthcare, image segmentation drives advancements in computer vision by enabling the clustering of unlabeled data samples into distinct categories, crucial for meaningful image analysis. In multi-human parsing, segmentation requires both instance-level and fine-grained category-level information [3]. Referring image segmentation (RIS) further illustrates this importance by localizing and segmenting specific image regions based on natural language descriptions, which is vital for understanding complex visual scenes.

The segmentation of farmlands with contour levees from high-resolution aerial imagery exemplifies the challenges posed by small objects and features that distort object boundaries, highlighting the complexities in remote sensing tasks [4]. Despite the limitations of existing methods, including convolutional neural networks (CNNs) and statistical models that often falter in raw medical image processing [5], ongoing advancements in segmentation methodologies are critical. These develop-

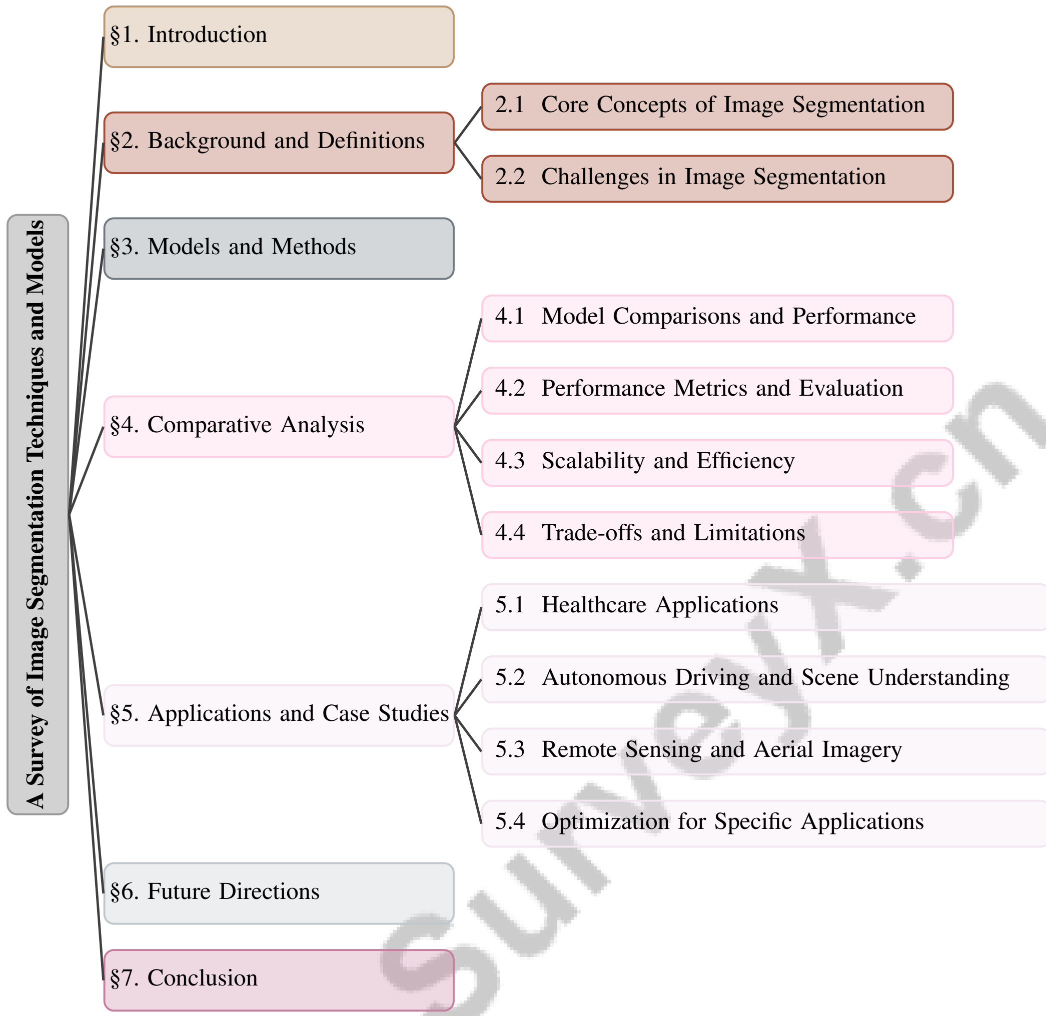


Figure 1: chapter structure

ments are essential for addressing the ill-posed nature of segmentation problems and enhancing the efficacy of computer vision technologies across diverse domains, facilitating the creation of adaptable systems capable of incremental learning in dynamic environments.

1.2 Types of Image Segmentation

Image segmentation encompasses three primary categories: instance segmentation, semantic segmentation, and panoptic segmentation, each targeting different aspects of visual scene interpretation [6]. Instance segmentation focuses on identifying and delineating individual object instances within an image, crucial for applications requiring precise scene analysis [7]. This task is particularly important in biomedical imaging, where accurate segmentation of anatomical structures is necessary for clinical diagnostics and effective medical image analysis across various datasets [8]. The complexity of instance segmentation is further highlighted by the challenge of segmenting referred regions based on natural language expressions, necessitating the identification of specific instances described by attributes like action, category, color, and position [9].

In contrast, semantic segmentation classifies each pixel in an image into categories without distinguishing between individual instances. This approach is vital in biomedical image segmentation, where traditional methods are often enhanced by deep learning architectures [10]. Underwater imaging presents unique challenges for semantic segmentation, requiring accurate delineation of regions of interest [11]. The task is complicated by domain shifts and varying feature spaces, emphasizing the need for sophisticated models [12]. Weakly supervised learning approaches also significantly contribute by leveraging image-text pairs to achieve pixel-level semantics, addressing the segmentation of fine-grained masks [13].

Panoptic segmentation unifies the goals of instance and semantic segmentation by assigning both a semantic class and an instance identity to each pixel [6]. This approach is essential for detailed scene parsing in applications such as autonomous driving and advanced robotics, where comprehensive scene context understanding is critical. The integration of segmentation tasks for open-vocabulary settings further underscores the necessity for models capable of handling diverse and complex visual inputs [10].

These segmentation tasks collectively tackle specific challenges and applications within computer vision, each contributing uniquely to enhancing visual scene understanding. The development of robust and efficient segmentation models remains vital for advancing the capabilities of computer vision systems across various domains [14].

1.3 Structure of the Survey

This survey offers an extensive examination of deep learning-based image segmentation methods, highlighting significant advancements while addressing key challenges and future directions in the field [15]. The paper is organized into several sections, each focusing on distinct aspects of image segmentation. The introduction emphasizes the importance of image segmentation and delineates the types of segmentation tasks, including instance, semantic, and panoptic segmentation. Following this, the background section provides definitions and explanations of core concepts and challenges inherent in image segmentation.

The survey then delves into models and methods, analyzing key models such as SAM2, SegFormer, BiRefNet, DeepLab, and Mask2FormerForUniversalSegmentation, alongside a discussion of their architectural features and innovations. A comparative analysis section evaluates the performance of these models, considering metrics, scalability, efficiency, trade-offs, and limitations.

The applications and case studies section explores the real-world impact of segmentation techniques across domains including healthcare, autonomous driving, and remote sensing. The survey concludes with a discussion on future directions, emphasizing emerging trends, advancements in learning techniques, and new segmentation paradigms aimed at addressing current limitations. This structured approach ensures a comprehensive understanding of the current state and future potential of image segmentation in computer vision.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts of Image Segmentation

Image segmentation, a crucial task in computer vision, partitions images into meaningful segments to enhance analysis and interpretation [10]. It is vital in applications like medical image analysis and autonomous driving, where precise object delineation is essential [16]. Methods in image segmentation include instance, semantic, and panoptic segmentation, each serving distinct purposes [6].

Instance segmentation identifies individual object instances, posing challenges in scenarios like multi-human parsing, where distinguishing between individuals and body parts is essential for scene understanding [3]. Semantic segmentation assigns categories to each pixel without distinguishing instances, crucial for applications like underwater imaging, where weak object boundaries challenge existing models [11]. Panoptic segmentation combines instance and semantic objectives, assigning each pixel a semantic class and an instance identity, crucial for detailed scene parsing in complex environments [6].

Salient object detection focuses on detecting salient objects by generating saliency maps, addressing limitations of heuristic methods that struggle with generalization across datasets [17]. Referring image segmentation (RIS) segments objects based on natural language descriptions, necessitating a fine-grained correlation between textual and visual data [18]. Weakly supervised learning approaches significantly enhance RIS by enabling accurate segmentation without pixel-level semantic supervision [13].

AI-generated, for reference only.

The distinction between bottom-up and top-down processing methods is fundamental, affecting model effectiveness [7]. The ill-posed nature of segmentation, where multiple valid segmentations can exist for a single image, underscores the need for robust computational models [14]. These core concepts are foundational for advancing image segmentation, addressing limitations, and expanding applicability across various domains.

2.2 Challenges in Image Segmentation

Image segmentation faces numerous challenges that hinder the development of robust models. A significant issue is the computational complexity inherent in segmentation tasks, often exacerbated by bottom-up processing methods, imposing substantial burdens [7]. This complexity is evident in medical image segmentation, where feature loss in tiny tumors or organs after deep convolutions poses challenges, necessitating high precision [19]. Current pruning techniques, primarily developed for image classification, struggle to transfer effectively to semantic segmentation due to the need for preserving detailed pixel-level structures [20].

Another challenge is the inability to manage local and global information simultaneously, leading to suboptimal segmentation outcomes [16]. Moreover, gradient-based techniques fail to address segmentation challenges, resulting in less precise saliency maps [17]. The inflexibility of current models, which cannot dynamically adjust parameters and architectures with new data, further impedes solutions [21].

In remote sensing, segmenting and classifying farmlands with contour levees using high-resolution imagery is challenging due to small objects and features [4]. The lack of hierarchical structures complicates parameter tuning and leads to unstable results across scales [22]. The ill-posed nature of segmentation poses challenges in achieving uniqueness and stability in solutions [14].

A key obstacle is the separate operation on instance and category information, preventing effective end-to-end optimization and hindering unified frameworks [3]. In medical image segmentation, existing methods' inefficiency, particularly the inability of CNNs to capture global features and the high computational cost of transformers, remains problematic [5]. Noise in pixel-level semantic responses from image-text pairs complicates segmentation, especially in weakly supervised learning scenarios [13]. Additionally, the lack of tolerance to diverse background changes often leads to incorrect classifications, highlighting a critical limitation [1].

Continuous research and innovation are essential for improving the robustness, accuracy, and efficiency of segmentation technologies. Enhancements will facilitate applications across diverse domains, including language-based image segmentation systems, diverse datasets utilization, and advanced frameworks that adapt to industrial needs while reducing computational costs. Addressing current algorithms and datasets' limitations will be crucial in ensuring segmentation technologies meet real-world application demands [23, 24, 25, 18, 26].

In recent years, advancements in image segmentation have been pivotal in enhancing the accuracy and efficiency of various applications in computer vision. A comprehensive understanding of these developments necessitates an examination of the underlying models and architectural innovations that have emerged in this field. As illustrated in Figure 2, the hierarchical organization of key models, including notable contributions such as SAM2 and SegFormer, is highlighted alongside the challenges that persist, particularly in terms of contextual integration and data efficiency. This visual representation not only encapsulates the complexity of the landscape but also serves as a foundation for discussing the implications of these innovations on future research trajectories.

3 Models and Methods

3.1 Key Models and Methods

Recent progress in image segmentation is marked by the development of various models, including bottom-up approaches, superpixel techniques, interactive methods, object proposals, and semantic image parsing. Table 1 provides a comprehensive overview of recent advancements in image segmentation methods, illustrating their diverse applications and technical innovations. These innovations cater to diverse fields such as medical imaging, document analysis, and remote sensing, as evidenced by a comprehensive review of 180 publications which highlights the evolution, challenges,

Method Name	Model Types	Application Domains	Technical Innovations
AquaSAM[11]	Bounding Box Prompt	Underwater Image Segmentation	Fine-tuning Approach
LF-SAM[27]	Ladder-Side Tuning	Medical Imaging	Flexible Additional Cnn
UM[16]	Transformer-based Architectures	Medical Imaging	Cross-domain Adaptability
TUnet[5]	Transformer Models	Medical Imaging	Transformer Layers
ACOSP[20]	Segnet, Pspnet	Medical Imaging	Auto-compression
PPT[13]	Transformer-based Point	Medical Imaging	Curriculum Learning Strategy
VN[4]	End-to-end	Remote Sensing	Voting Mechanism
UP[3]	Correlation Representation Learning	Multi-human Parsing	Joint Optimization

Table 1: Summary of key image segmentation methods, detailing their model types, application domains, and technical innovations. This table highlights the diverse approaches and advancements in the field, showcasing methods tailored for specific applications such as underwater image segmentation, medical imaging, and remote sensing.

datasets, and evaluation metrics shaping the current landscape [14, 26]. Prominent models like SAM2, SegFormer, BiRefNet, DeepLab, and Mask2FormerForUniversalSegmentation each contribute unique advancements.

SAM2, an extension of the Segment Anything Model, exemplifies cross-domain adaptability, with AquaSAM tailored for underwater segmentation through specialized prompts [11]. The Ladder Fine-tuning method enhances SAM's effectiveness in medical imaging by integrating a complementary CNN [27].

SegFormer leverages a transformer-based architecture to capture global context, crucial for interpreting complex scenes [6]. BiRefNet refines object boundaries through bidirectional processing [6].

DeepLab advances dense prediction tasks with atrous convolution, capturing multi-scale contextual information efficiently [6]. Mask2FormerForUniversalSegmentation unifies segmentation tasks, addressing both instance and semantic challenges [6].

U-Netmer combines U-Net's local feature extraction with Transformers' global context learning, enhancing performance in complex scenarios [16]. TUnet showcases advanced learning by integrating transformers within U-Net [5].

The AUTO-COMPRESSING SUBSET PRUNING (ACOSP) method enhances model efficiency by pruning convolutional filters based on importance during training [20]. The Point PrompTing (PPT) framework connects CLIP and SAM for weakly supervised referring image segmentation, advancing mask generation [13].

VoteNet applies a voting mechanism to segment and classify farmlands from high-resolution aerial imagery, demonstrating deep learning's role in remote sensing [4]. UniParser merges instance and category-level representations in a unified learning framework, benefiting multi-human parsing [3].

These advancements address challenges like contextual integration, data efficiency, and multi-scale data incorporation, significantly enhancing segmentation accuracy across domains such as medical imaging and complex scene analysis [28, 2, 26]. These innovations pave the way for robust and adaptable segmentation technologies.

3.2 Architectural Features and Innovations

Architectural innovations in image segmentation enhance model adaptability, efficiency, and precision across applications. The Visual Hint Boundary Segment Algorithm exemplifies dynamic adaptability, allowing real-time model adjustments, contrasting with static approaches that require complete retraining [21]. This adaptability is crucial for applications needing rapid updates.

The Segment Anything Model (SAM) demonstrates versatility through extensive benchmarking across configurations, optimizing performance with selective fine-tuning to minimize resource use while maintaining accuracy [8]. TUnet integrates transformers with CNNs, balancing detailed local feature extraction with broader contextual understanding [5].

The Point PrompTing (PPT) framework uses a trainable point generator to guide mask decoding, improving segmentation accuracy and efficiency [13]. The context window approach in convolution-based segmentation optimizes weights via a non-parametric function, enhancing adaptability to contextual information [1].

Hierarchical graph-based segmentation adheres to locality and causality principles, ensuring stable results across scales [22]. UniParser unifies instance and category-level features, producing pixel-level outputs that integrate multiple objectives [3].

The voting mechanism in contour levee segmentation refines predictions through majority voting, improving accuracy [4]. This collaborative decision-making enhances reliability.

These architectural advancements, such as dense representation and prompt fusion in generalist models like SegNext, improve segmentation quality while maintaining low latency. The Augmentation-based Model Re-adaptation Framework (AMRF) enhances generalization through data augmentation, expanding applicability across domains [23, 29, 24, 30, 26].

4 Comparative Analysis

4.1 Model Comparisons and Performance

Comparative evaluations of image segmentation models are crucial for assessing their effectiveness across various tasks and datasets. The UniParser model excels in instance-aware metrics, proving its capability in complex segmentation scenarios [3]. VoteNet achieves an average accuracy of 94.34

The AUTO-COMPRESSING SUBSET PRUNING (ACOSP) method maintains high segmentation accuracy across various compression ratios, highlighting efficient model compression techniques [20]. U-Netmer, integrating U-Net and Transformer architectures, shows effectiveness across multiple datasets, including ACDC and BraTS [16]. Fine-tuning the Segment Anything Model (SAM) consistently outperforms traditional methods like UNet, indicating its adaptability for optimizing segmentation performance [8]. TUnet, combining transformers with U-Net architectures, achieves superior segmentation outcomes by capturing both local and global features [5]. The Point PrompTing (PPT) framework demonstrates superiority in weakly supervised referring image segmentation tasks, emphasizing guided learning techniques [13].

These analyses reveal the strengths and weaknesses of segmentation models, addressing challenges such as region-level textual annotation scarcity and balancing latency with quality. Insights into architectural differences and parameter choices inform current practices and guide future advancements in image segmentation methodologies, including diverse prompts and cost minimization strategies [18, 30, 24].

4.2 Performance Metrics and Evaluation

Benchmark	Size	Domain	Task Format	Metric
GOSS[31]	118,287	Image Segmentation	Pixel Classification And Clustering	GOSS Quality
SAM2[32]	2,000	Image Segmentation	Instance-level Segmentation	AP, F max
SAM[33]	7,451	Medical Imaging	Segmentation	Dice overlap
FMS[34]	1,056	Medical Imaging	Segmentation	Dice Similarity Coefficient, Std-GM
AL3D-MIS[35]	3,000	Medical Imaging	Segmentation	Dice Score, Hausdorff Distance
CIRA[36]	5,000	Semantic Segmentation	Pixel-wise Classification	mIoU
FoodSeg[37]	9,490	Food Image Segmentation	Image Segmentation	mIoU, mAcc
FP-k[38]	2,100	Image Segmentation	Few-Shot Learning	Mean Intersection over Union

Table 2: Table ef presents a comprehensive overview of representative benchmarks utilized for evaluating image segmentation models. The table details various datasets across different domains, highlighting their size, task format, and the specific metrics employed for performance assessment. This information is crucial for understanding the diverse evaluation criteria applied in image segmentation research.

Evaluating image segmentation models requires a comprehensive suite of metrics, each offering unique insights into segmentation quality. The mean Intersection over Union (mIoU) is widely used for assessing the overlap between predicted and ground truth segmentations, crucial for evaluating model generalization across diverse datasets [39]. The Dice Similarity Coefficient (DSC) is significant in medical image segmentation, quantifying overlap between predictions and actual segmentations, thus providing a robust precision measure [5].

Additional metrics like pixel accuracy, precision, recall, and the Jaccard index offer nuanced perspectives on performance. The Jaccard index measures similarity and diversity of sample sets, providing insights into accuracy [16]. Precision metrics are crucial in contexts requiring high accuracy, particularly in detecting small or overlapping objects [18]. The choice of metrics significantly influences perceived effectiveness, as different metrics may yield varying results [40].

Table 2 provides an insightful summary of key benchmarks used in the evaluation of image segmentation models, illustrating the diversity in dataset characteristics and evaluation metrics. Employing a combination of metrics, including mIoU, DSC, and precision, alongside parameters like sensitivity and Hausdorff distance (HD), creates a well-rounded framework for evaluating segmentation models [41]. Sensitivity, measuring the true positive rate, is crucial for applications demanding high detection accuracy. This multifaceted framework enables systematic assessment of different algorithms across tasks and datasets, guiding informed decision-making and facilitating advancements in image segmentation technologies [42, 43].

4.3 Scalability and Efficiency

Evaluating scalability and efficiency of image segmentation models is crucial for real-world applications, especially with large datasets. Scalability refers to a model's ability to maintain performance as input data size increases, while efficiency pertains to computational resources required for processing [4]. The AUTO-COMPRESSING SUBSET PRUNING (ACOSP) method exemplifies efficiency improvements through a channel selection mechanism that prunes convolutional filters based on significance, reducing computational load without sacrificing accuracy [20].

Models like SAM2 and SegFormer demonstrate scalability through their architecture, effectively balancing global context capture with computational efficiency [6]. SegFormer leverages a transformer-based design to efficiently process large images by capturing long-range dependencies, crucial for maintaining performance at scale. TUnet enhances scalability by combining transformers and CNNs, allowing efficient processing of large datasets while retaining high segmentation accuracy [5]. VoteNet's voting mechanism refines predictions, showcasing an efficient approach to segmentation in large datasets, particularly in remote sensing applications [4].

The U-Netmer model effectively manages local and global feature extraction, demonstrating scalability for extensive medical imaging data analysis [16]. Recent advancements emphasize the critical role of architectural innovations and resource optimization techniques in enhancing scalability and efficiency. The Segment Anything Model (SAM) shows promise in medical image analysis, utilizing parameter-efficient learning strategies for superior performance across diverse datasets. This underscores the importance of developing robust, adaptable algorithms to meet the demands of large-scale image processing applications [8, 14, 24].

4.4 Trade-offs and Limitations

Despite significant advancements, image segmentation models face various trade-offs and limitations impacting their performance across tasks. A notable limitation of many methods, including non-parametric approaches, is their dependency on careful training context selection, where the absence of critical context adversely affects performance [1]. The UniParser model, while innovative, demands increased memory and time resources, constraining its use in environments with limited computational capacity [3]. Similarly, hierarchical graph-based segmentation methods may struggle with dynamic datasets due to their lack of parameter flexibility [22].

VoteNet's reliance on the parameter K poses risks of over-segmentation or under-segmentation, affecting quality based on its selection [4]. The effectiveness of segmentation methods is often contingent on the specific characteristics of images and desired outcomes, making it challenging to develop universally effective models across diverse datasets and applications [14].

These trade-offs and limitations highlight the need for ongoing research to enhance adaptability, efficiency, and generalization capabilities of segmentation models. Addressing challenges, particularly the scarcity of datasets with joint vision and text annotations, is essential for advancing the field.

Leveraging existing large-scale vision-only and text-only datasets can improve models for image segmentation from referring expressions, facilitating better performance across applications. A comprehensive review of recent advancements, including bottom-up approaches, superpixel techniques, and interactive methods, underscores the necessity for continued innovation to overcome limitations and expand the utility of segmentation technologies [18, 26].

5 Applications and Case Studies

Image segmentation significantly influences various domains, including healthcare, autonomous driving, and remote sensing, each presenting unique challenges that necessitate specialized techniques. The following subsections delve into these applications, highlighting their impact on diagnostics, navigation, and land use analysis.

5.1 Healthcare Applications

In healthcare, image segmentation is crucial for medical imaging, allowing precise delineation of anatomical structures necessary for accurate diagnosis and treatment planning [6]. Advanced models like the Segment Anything Model (SAM) enhance segmentation accuracy and reduce annotation workload in 3D imaging [8]. SAM's performance in brain tumor segmentation, as benchmarked by [44], facilitates comparisons of prompting strategies.

Innovative approaches such as ADSNet improve polyp segmentation in colonoscopy images, enhancing diagnostic accuracy [45]. Techniques improving calibration and Out-of-Distribution (OOD) detection bolster diagnostic tool reliability [46]. Attention-based methods like the CVFC model demonstrate enhanced performance across pathology datasets [47], while the MS-Twins model shows notable accuracy improvements on datasets like Synapse and ACDC [48].

Lightweight networks proposed by [49] enhance performance without compromising computational efficiency, suitable for real-time emergency diagnostics. Innovations such as Auxiliary Online Learning (AuxOL) and the Part-aware Personalized Segment Anything Model (P^2SAM) significantly boost segmentation accuracy in diverse clinical scenarios, addressing variability in patient data and enhancing patient-specific treatments [50, 51, 29]. These advancements promise improved diagnostics and treatment planning, offering enhanced precision and reliability.

As depicted in Figure 3, advanced computational techniques have significantly improved medical diagnostics and treatment planning. The first subfigure illustrates medical image segmentation, crucial for identifying anatomical structures or lesions. The second subfigure compares image augmentation techniques, SAMAug and SAMAug-C, which enhance model training by introducing varied noise and artifacts. The third subfigure showcases a comparative analysis of brain segmentation methods, highlighting the effectiveness of techniques like "Ours-3

5.2 Autonomous Driving and Scene Understanding

In autonomous driving, image segmentation is essential for scene understanding and navigation, enabling vehicles to accurately interpret complex visual scenes. A primary challenge is managing uncontrolled environmental conditions, which can significantly impact segmentation accuracy [54].

Advanced models using divisive normalization enhance robustness against environmental variability, improving scene interpretation reliability [54]. These models help vehicles identify road boundaries, track lane markings, and predict surrounding agents' behavior. Recent advancements, such as the Fast Segment Anything Model (SAM) and its real-time adaptations, achieve impressive performance metrics, including 90 FPS inference speed and a mean IoU of 73

5.3 Remote Sensing and Aerial Imagery

In remote sensing and aerial imagery, image segmentation is crucial for land use analysis. Segmenting aerial images enables classification of various land cover types, essential for urban planning, environmental monitoring, and resource management. The LandCoverAI dataset, categorized into four classes—Buildings, Woodland, Water, and Road—serves as a benchmark for evaluating segmentation methods [55].

Advanced techniques improve accuracy with limited labeled data, particularly in large-scale remote sensing scenes where obtaining labeled data is challenging. Multi-view approaches enhance inter-view consistency, providing a coherent representation of segmented areas [56]. Integrating noisy labels into the training process has improved model performance, leveraging variability in remote sensing data to enhance robustness [57]. These advancements underscore the importance of developing robust models capable of addressing the unique challenges of remote sensing and aerial imagery. Innovations like end-to-end trainable networks and enhanced voting mechanisms significantly improve segmentation precision, facilitating accurate land use analysis and informed decision-making [4, 26].

5.4 Optimization for Specific Applications

Optimizing segmentation techniques for specific applications is critical for tailoring models to meet various domain demands. Enhancements in generalized open-set semantic (GOSS) segmentation improve the processing of unknown objects, increasing applicability in dynamic environments compared to traditional open-set semantic (OSS) models [31].

In industrial applications, refining segmentation methods to accommodate different kernels is a promising research direction. Optimizing the cost function associated with these methods can lead to more efficient and accurate outcomes [24]. In medical imaging, targeted optimization strategies are crucial for bridging the gap between 2D and 3D segmentation techniques, particularly through the exploration of 2.5D methods [58].

Comprehensive optimization strategies encompass enhancing adaptability, refining methodological components through innovative algorithms, and investigating architectures that leverage online learning and diverse prompt integration. Advances in models like the Segment Anything Model (SAM) and evolutionary computation methods illustrate the potential for improving segmentation accuracy and efficiency across various domains, including medical imaging and interactive segmentation tasks [50, 59, 30, 24]. These efforts are essential for advancing the effectiveness and applicability of segmentation technologies across diverse industrial and medical domains.

As illustrated in Figure 4, advanced techniques in image processing and segmentation reflect the multifaceted nature of modern computational strategies. The first example, "Shapes and Objects," presents various geometric forms and everyday objects, emphasizing clarity in distinguishing these elements. The "Watershed Algorithm for Image Segmentation" highlights the procedural intricacies of the algorithm, refining input images into distinct segments. The "Residual Skip Decoder Network" exemplifies the integration of neural network architectures in image processing, employing convolutional layers enhanced by residual skip connections for efficient image transformation. Collectively, these examples underscore the tailored application of optimization techniques in enhancing image segmentation capabilities, reflecting their pivotal role in advancing computational imaging technologies [60, 59, 38].

6 Future Directions

Navigating the evolving landscape of image segmentation requires identifying emerging trends and innovations that are shaping the field's future. As technological and methodological advancements unfold, researchers focus on integrating novel approaches to enhance segmentation models' capabilities and performance. This section explores the latest developments driving progress in image segmentation, emphasizing transformer-based architectures and unsupervised learning advancements. By examining these trends, we gain insight into the trajectory of image segmentation research and its implications for various applications.

6.1 Emerging Trends and Innovations

Image segmentation is experiencing a transformative phase, with trends aimed at enhancing model capabilities and broadening applicability across domains. Transformer-based architectures are increasingly adopted to capture long-range dependencies and improve accuracy, particularly in complex tasks like panoptic segmentation [6]. Unsupervised learning techniques are gaining traction, reducing reliance on large labeled datasets, which are costly and time-consuming to produce [10]. This shift towards autonomous learning is exemplified by self-supervised techniques that enhance performance without extensive human intervention.

Future research focuses on refining context modules and exploring additional contextual features to improve segmentation performance. Integrating pseudo-labels for high-quality predictions and utilizing weighted priority scores are approaches that aim to leverage contextual information effectively, addressing challenges in diverse applications, including medical imaging and natural image segmentation [61, 24, 28, 18, 2].

AUTO-COMPRESSING SUBSET PRUNING (ACOSP) applied to state-of-the-art architectures enhances performance through tailored loss functions and efficient model compression techniques [20]. Optimizing parameters like queried boxes or voting mechanisms is pursued to improve performance across varying conditions [4]. Hybrid models, combining strengths from various approaches, improve accuracy and efficiency by leveraging complementary methodologies [14]. Real-time models and weakly-supervised learning aim to create more adaptable segmentation algorithms [10].

Emerging trends and innovations in image segmentation suggest a promising future with sophisticated and adaptable models tackling diverse challenges. Transformer-based architectures, such as Transformer-Unet, enhance segmentation performance in complex tasks like medical image analysis. Unsupervised learning advancements, exemplified by models like CLUSTERFORMER, improve representation learning and versatility across vision tasks [62, 5, 26].

6.2 Advancements in Learning Techniques

Recent advancements in learning techniques have significantly enhanced segmentation performance by introducing innovative approaches addressing traditional methods' limitations. Deep learning frameworks, particularly those leveraging convolutional neural networks (CNNs) and transformers, have been pivotal in improving segmentation accuracy and efficiency across applications. Hybrid models, combining CNN strengths for local feature extraction with transformers for global context capture, exemplify this trend [5].

Self-supervised and unsupervised learning techniques are reducing reliance on large labeled datasets, which are resource-intensive to produce [10]. Attention mechanisms, integrated into segmentation models, improve focus on relevant features and enhance interpretability [47]. Domain adaptation techniques address challenges posed by domain shifts and variations, enabling models to maintain high performance across datasets [12].

Multi-scale and multi-modal learning techniques enrich segmentation models' capabilities, allowing them to process and integrate information from multiple scales and modalities, enhancing their ability to handle complex tasks [6]. Recent advancements signify a substantial leap in the field, offering solutions that are robust, efficient, and adaptable to real-world challenges. These include innovative algorithms minimizing costs by leveraging diverse segmentation hypotheses, enhancing accuracy in domains like medical imaging. Online learning methods, such as Auxiliary Online Learning (AuxOL), allow real-time improvements in segmentation quality by utilizing rectified annotations during test time [50, 24, 26].

6.3 Exploration of New Segmentation Paradigms

Exploring new segmentation paradigms is crucial for addressing current limitations and enhancing model adaptability. Future research could extend frameworks like UniverSeg to support 3D segmentation tasks and multi-label maps, increasing versatility and applicability [63]. PoissonSeg's potential extensions focus on segmenting volumetric medical images, improving semi-supervised learning capabilities [64].

Enhancing frameworks like Pixel Objectness for instance segmentation involves refining models' ability to distinguish between individual object instances [65]. Exploring new paradigms is essential for overcoming challenges and expanding segmentation technologies' applicability. Innovative frameworks, methodologies, and techniques, such as cost minimization, auxiliary online learning, and augmentation-based model re-adaptation, enhance segmentation models' capabilities, enabling robust and adaptable solutions for evolving demands across applications [50, 23, 24].

6.4 Addressing Current Limitations and Challenges

Addressing current challenges in image segmentation involves refining model architecture, enhancing training methodologies, and adapting to specific domain requirements. A significant limitation is dependency on large labeled datasets, which are resource-intensive and time-consuming to produce [10]. Future research should develop efficient algorithms reducing the need for extensive labeled data and computational resources, improving model generalizability across diverse tasks.

Computational efficiency and real-time application optimization remain critical areas for improvement. Hierarchical graph-based methods, effective in producing stable results, face challenges in computational efficiency and require optimization for real-time applications [22]. Enhancing scalability and speed is crucial for deployment in time-sensitive applications, such as autonomous driving and emergency medical diagnostics.

Exploring correlation representation learning offers promising improvement avenues, particularly with limited image samples. This technique could achieve higher accuracy and robustness without relying on vast training data [3]. Addressing these challenges through innovative research will evolve image segmentation, offering robust, accurate, and adaptable solutions across diverse applications. Integrating advanced learning techniques, focusing on computational efficiency and generalizability, is essential for addressing challenges, particularly in medical imaging. This approach enhances segmentation models' accuracy, aligning with recent findings in human visual perception, enabling effective image decomposition into meaningful components. Leveraging diverse segmentation hypotheses through cost minimization frameworks achieves improved and stable results, facilitating better image understanding and interpretation across applications [50, 7, 24, 26].

7 Conclusion

Image segmentation remains a cornerstone of computer vision, crucial for extracting detailed information across diverse fields such as healthcare, autonomous driving, and remote sensing. The latest innovations in segmentation techniques, exemplified by models like SAM2, SegFormer, BiRefNet, DeepLab, and Mask2FormerForUniversalSegmentation, have significantly enhanced the precision, efficiency, and adaptability of segmentation processes, addressing the intricate demands of contemporary tasks.

Emerging methodologies such as MiSuRe have improved model interpretability, providing valuable insights through effective saliency maps that deliver both broad and detailed explanations. The GASE framework exemplifies progress in adapting to distribution shifts, achieving superior segmentation accuracy and offering a visualizable manifold that aids in understanding model dynamics. These developments contribute to a more profound comprehension of model operations, thereby reinforcing the dependability of segmentation results.

Moreover, the MM tuning approach has propelled advancements in self-supervised learning for semantic segmentation, achieving leading performance on benchmark datasets like PASCAL VOC2012 and CityScapes. This highlights the promise of innovative learning paradigms in minimizing the need for large annotated datasets while enhancing model generalization capabilities.

As the discipline evolves, ongoing research and innovation are essential to overcoming existing limitations and seizing new opportunities in image segmentation. The fusion of advanced learning strategies, along with a focus on computational efficiency and generalizability, will be pivotal in tackling current challenges and expanding the scope of segmentation technologies across various sectors.

References

[1] Upendra Kumar, Tapobrata Lahiri, and Manoj Kumar Pal. Non-parametric convolution based image-segmentation of ill-posed objects applying context window approach, 2012.

[2] Lingdong Shen, Fangxin Shang, Xiaoshuang Huang, Yehui Yang, Haifeng Huang, and Shiming Xiang. Segicl: A multimodal in-context learning framework for enhanced segmentation in medical imaging, 2024.

[3] Jiaming Chu, Lei Jin, Junliang Xing, and Jian Zhao. Uniparser: Multi-human parsing with unified correlation representation learning, 2024.

[4] Abolfazl Meyarian and Xiaohui Yuan. Voting network for contour levee farmland segmentation and classification, 2023.

[5] Youyang Sha, Yonghong Zhang, Xuquan Ji, and Lei Hu. Transformer-unet: Raw image processing with unet, 2021.

[6] Rui Li, Chenxi Duan, Shunyi Zheng, Ce Zhang, and Peter M. Atkinson. Macu-net for semantic segmentation of fine-resolution remotely sensed images, 2022.

[7] Emanuel Diamant. Paving the way for image understanding: A new kind of image decomposition is desired, 2005.

[8] Hanxue Gu, Haoyu Dong, Jichen Yang, and Maciej A. Mazurowski. How to build the best medical image segmentation algorithm using foundation models: a comprehensive empirical study with segment anything model, 2024.

[9] Jiaxing Yang, Lihe Zhang, Jiayu Sun, and Huchuan Lu. Spatial semantic recurrent mining for referring image segmentation, 2024.

[10] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey, 2020.

[11] Muduo Xu, Jianhao Su, and Yutao Liu. Aquasam: Underwater image foreground segmentation, 2023.

[12] Gaetano Scebba, Jia Zhang, Sabrina Catanzaro, Carina Mihai, Oliver Distler, Martin Berli, and Walter Karlen. Detect-and-segment: a deep learning approach to automate wound image segmentation, 2021.

[13] Qiyuan Dai and Sibei Yang. Curriculum point prompting for weakly-supervised referring image segmentation, 2024.

[14] Laura Antonelli, Valentina De Simone, and Daniela di Serafino. A view of computational models for image segmentation, 2022.

[15] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542, 2021.

[16] Sheng He, Rina Bao, P. Ellen Grant, and Yangming Ou. U-netmer: U-net meets transformer for medical image segmentation, 2023.

[17] Syed Nouman Hasany, Fabrice Mériaudeau, and Caroline Petitjean. Misure is all you need to explain your image segmentation, 2024.

[18] Ronghang Hu, Marcus Rohrbach, Subhashini Venugopalan, and Trevor Darrell. Utilizing large scale vision and text datasets for image segmentation from referring expressions, 2016.

[19] Fuchen Zheng, Xuhang Chen, Weihuang Liu, Haolun Li, Yingtie Lei, Jiahui He, Chi-Man Pun, and Shounjun Zhou. Smaformer: Synergistic multi-attention transformer for medical image segmentation, 2024.

[20] Konstantin Ditschuneit and Johannes S. Otterbach. Auto-compressing subset pruning for semantic image segmentation, 2022.

[21] Yu Su and Margaret H. Dunham. Visual-hint boundary to segment algorithm for image segmentation, 2010.

[22] Silvio Jamil F. Guimarães, Jean Cousty, Yukiko Kenmochi, and Laurent Najman. An efficient hierarchical graph based image segmentation, 2012.

[23] Zheming Zuo, Joseph Smith, Jonathan Stonehouse, and Boguslaw Obara. An augmentation-based model re-adaptation framework for robust image segmentation, 2024.

[24] Marc Bosch, Christopher M. Gifford, Austin G. Dress, Clare W. Lau, Jeffrey G. Skibo, and Gordon A. Christie. Improved image segmentation via cost minimization of multiple hypotheses, 2018.

[25] Xu Zhao, Wenchao Ding, Yongqi An, Yinglong Du, Tao Yu, Min Li, Ming Tang, and Jinqiao Wang. Fast segment anything, 2023.

[26] Hongyuan Zhu, Fanman Meng, Jianfei Cai, and Shijian Lu. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation, 2015.

[27] Shurong Chai, Rahul Kumar Jain, Shiyu Teng, Jiaqing Liu, Yinhao Li, Tomoko Tateyama, and Yen wei Chen. Ladder fine-tuning approach for sam integrating complementary network, 2023.

[28] Maciej A. Mazurowski, Haoyu Dong, Hanxue Gu, Jichen Yang, Nicholas Konz, and Yixin Zhang. Segment anything model for medical image analysis: an experimental study, 2023.

[29] Seyed M. R. Modaresi, Aomar Osmani, Mohammadreza Razzazi, and Abdelghani Chibani. Boosting medical image segmentation performance with adaptive convolution layer, 2024.

[30] Qin Liu, Jaemin Cho, Mohit Bansal, and Marc Niethammer. Rethinking interactive image segmentation with low latency, high quality, and diverse prompts, 2024.

[31] Jie Hong, Weihao Li, Junlin Han, Jiyang Zheng, Pengfei Fang, Mehrtash Harandi, and Lars Petersson. Goss: Towards generalized open-set semantic segmentation, 2022.

[32] Jialun Pei, Zhangjun Zhou, and Tiantian Zhang. Evaluation study on sam 2 for class-agnostic instance-level segmentation, 2024.

[33] Sheng He, Rina Bao, Jingpeng Li, Jeffrey Stout, Atle Bjornerud, P. Ellen Grant, and Yangming Ou. Computer-vision benchmark segment-anything model (sam) in medical images: Accuracy in 12 datasets, 2023.

[34] Qin Li, Yizhe Zhang, Yan Li, Jun Lyu, Meng Liu, Longyu Sun, Mengting Sun, Qirong Li, Wenyue Mao, Xinran Wu, Yajing Zhang, Yinghua Chu, Shuo Wang, and Chengyan Wang. An empirical study on the fairness of foundation models for multi-organ image segmentation, 2024.

[35] Josafat-Mattias Burmeister, Marcel Fernandez Rosas, Johannes Hagemann, Jonas Kordt, Jasper Blum, Simon Shabo, Benjamin Bergner, and Christoph Lippert. Less is more: A comparison of active learning strategies for 3d medical image segmentation, 2022.

[36] Levente Halmosi and Mark Jelasity. On evaluating the adversarial robustness of semantic segmentation models, 2023.

[37] Xiongwei Wu, Xin Fu, Ying Liu, Ee-Peng Lim, Steven C. H. Hoi, and Qianru Sun. A large-scale benchmark for food image segmentation, 2021.

[38] Sean M. Hendryx, Andrew B. Leach, Paul D. Hein, and Clayton T. Morrison. Meta-learning initializations for image segmentation, 2020.

[39] Jiawei Zhang, Yanchun Zhang, and Xiaowei Xu. Objectaug: Object-level data augmentation for semantic image segmentation, 2021.

[40] Zizhao Zhang, Fuyong Xing, Hanzi Wang, Yan Yan, Ying Huang, Xiaoshuang Shi, and Lin Yang. Revisiting graph construction for fast image segmentation, 2017.

AI-generated, for reference only.

[41] Shiqiang Ma, Xuejian Li, Jijun Tang, and Fei Guo. Eaa-net: Rethinking the autoencoder architecture with intra-class features for medical image segmentation, 2022.

[42] Majid Harouni and Hadi Yazdani Baghmaleki. Color image segmentation metrics, 2020.

[43] Fernando A. Fardo, Victor H. Conforto, Francisco C. de Oliveira, and Paulo S. Rodrigues. A formal evaluation of psnr as quality measurement parameter for image segmentation algorithms, 2016.

[44] Peng Zhang and Yaping Wang. Segment anything model for brain tumor segmentation, 2024.

[45] Quang Vinh Nguyen, Van Thong Huynh, and Soo-Hyung Kim. Adaptation of distinct semantics for uncertain areas in polyp segmentation, 2024.

[46] Improving calibration and out-of-distribution detection in medical image segmentation with convolutional neural networks.

[47] Liangrui Pan, Lian Wang, Zhichao Feng, Liwen Xu, and Shaoliang Peng. Cvfc: Attention-based cross-view feature consistency for weakly supervised semantic segmentation of pathology images, 2023.

[48] Jing Xu. Ms-twins: Multi-scale deep self-attention networks for medical image segmentation, 2024.

[49] Dian Qin, Jiajun Bu, Zhe Liu, Xin Shen, Sheng Zhou, Jingjun Gu, Zhijua Wang, Lei Wu, and Huifen Dai. Efficient medical image segmentation based on knowledge distillation, 2021.

[50] Tianyu Huang, Tao Zhou, Weidi Xie, Shuo Wang, Qi Dou, and Yizhe Zhang. Improving segment anything on the fly: Auxiliary online learning and adaptive fusion for medical image segmentation, 2024.

[51] Chenhui Zhao and Liyue Shen. Part-aware personalized segment anything model for patient-specific segmentation, 2024.

[52] Pengfei Gu, Zihan Zhao, Hongxiao Wang, Yaopeng Peng, Yizhe Zhang, Nishchal Sapkota, Chaoli Wang, and Danny Z. Chen. Boosting medical image classification with segmentation foundation model, 2024.

[53] Weijia Feng, Lingting Zhu, and Lequan Yu. Cheap lunch for medical image segmentation by fine-tuning sam on few exemplars, 2023.

[54] Pablo Hernández-Cámara, Jorge Vila-Tomás, Paula Dauden-Oliver, Nuria Alabau-Bosque, Valero Laparra, and Jesús Malo. Image segmentation via divisive normalization: dealing with environmental diversity, 2024.

[55] Santiago Rivier, Carlos Hinojosa, Silvio Giancola, and Bernard Ghanem. Learning semantic segmentation with query points supervision on aerial images, 2024.

[56] Zipeng Qi, Chenyang Liu, Zili Liu, Hao Chen, Yongchang Wu, Zhengxia Zou, and Zhenwei Sh. Multi-view remote sensing image segmentation with sam priors, 2024.

[57] Chenying Liu, Conrad M Albrecht, Yi Wang, and Xiao Xiang Zhu. Task specific pretraining with noisy labels for remote sensing image segmentation, 2024.

[58] Yichi Zhang, Qingcheng Liao, Le Ding, and Jicong Zhang. Bridging 2d and 3d segmentation networks for computation efficient volumetric medical image segmentation: An empirical study of 2.5d solutions, 2022.

[59] Roohollah Aslanzadeh, Kazem Qazanfari, and Mohammad Rahmati. An efficient evolutionary based method for image segmentation, 2017.

[60] Alexander Aue and Thomas C. M. Lee. On image segmentation using information theoretic criteria, 2012.

[61] Zhenchao Jin, Bin Liu, Qi Chu, and Nenghai Yu. Isnet: Integrate image-level and semantic-level context for semantic segmentation, 2021.

[62] James C. Liang, Yiming Cui, Qifan Wang, Tong Geng, Wenguan Wang, and Dongfang Liu. Clusterformer: Clustering as a universal visual learner, 2023.

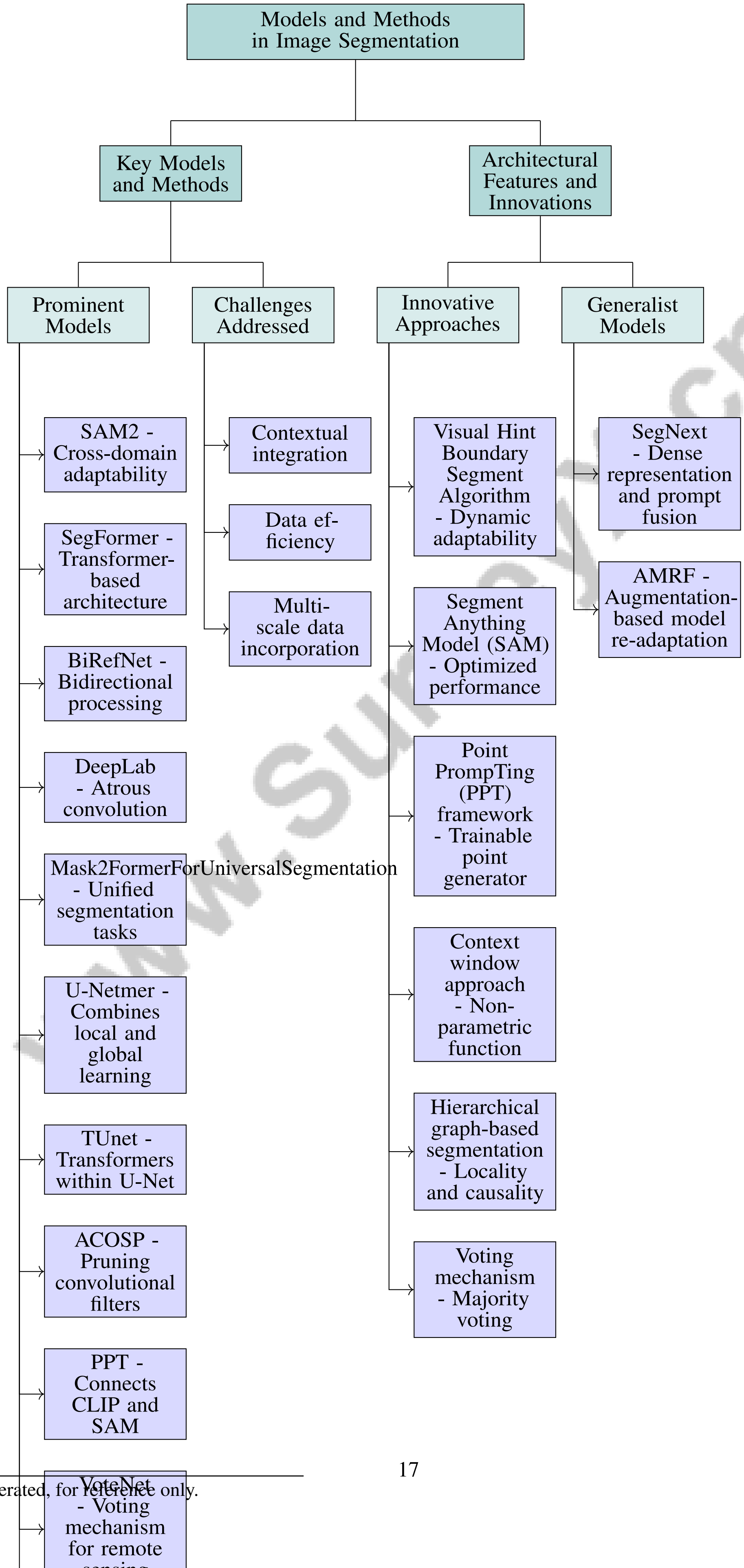
[63] Victor Ion Butoi, Jose Javier Gonzalez Ortiz, Tianyu Ma, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Universeg: Universal medical image segmentation, 2023.

[64] Xiaoang Shen, Guokai Zhang, Huilin Lai, Jihao Luo, Jianwei Lu, and Ye Luo. Poissonseg: Semi-supervised few-shot medical image segmentation via poisson learning, 2021.

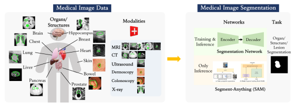
[65] Bo Xiong, Suyog Dutt Jain, and Kristen Grauman. Pixel objectness: Learning to segment generic objects automatically in images and videos, 2018.

Disclaimer:

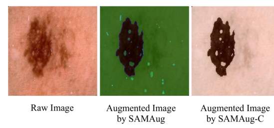
SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.



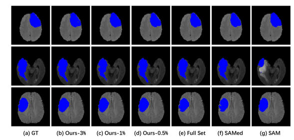
Models and Methods in Image Segmentation
Key Models and Methods
Architectural Features and Innovations
Prominent Models
Challenges Addressed
Innovative Approaches
Generalist Models
SAM2 - Cross-domain adaptability
SegFormer - Transformer-based architecture
BiRefNet - Bidirectional processing
DeepLab - Atrous convolution
Mask2FormerForUniversalSegmentation - Unified segmentation tasks
U-Netmer - Combines local and global learning
TUnet - Transformers within U-Net
ACOSP - Pruning convolutional filters
PPT - Connects CLIP and SAM
VoteNet - Voting mechanism for remote
Contextual integration
Data efficiency
Multi-scale data incorporation
Visual Hint Boundary Segment Algorithm - Dynamic adaptability
Segment Anything Model (SAM) - Optimized performance
Point PrompTing (PPT) framework - Trainable point generator
Context window approach - Non-parametric function
Hierarchical graph-based segmentation - Locality and causality
Voting mechanism - Majority voting
SegNext - Dense representation and prompt fusion
AMRF - Augmentation-based model re-adaptation



(a) Medical Image Segmentation[33]

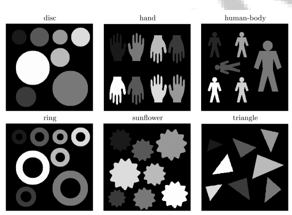


(b) Comparison of SAMAug and SAMAug-C Image Augmentation Techniques[52]

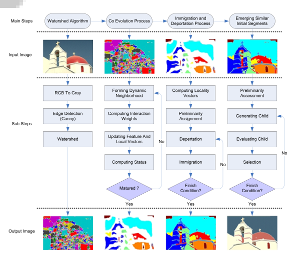


(c) Comparison of Brain Segmentation Results Using Different Methods[53]

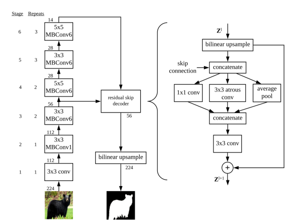
Figure 3: Examples of Healthcare Applications



(a) Shapes and Objects[60]



(b) Watershed Algorithm for Image Segmentation[59]



(c) Residual Skip Decoder Network[38]

Figure 4: Examples of Optimization for Specific Applications