



Title	Visual prompting and adaptive learning-based waterbody segmentation from normal-perspective imagery
Author(s)	趙, 家培
Citation	大阪大学, 2025, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.18910/103132">https://doi.org/10.18910/103132</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## Abstract of Thesis

Name (趙 家培)

Title	Visual prompting and adaptive learning-based waterbody segmentation from normal-perspective imagery (視覚プロンプティングと適応学習に基づく通常視点画像からの水域セグメンテーション)
-------	--

### Abstract of Thesis

This dissertation focuses on advancing waterbody segmentation from normal-perspective imagery, addressing key challenges such as illumination variations, reflections, and complex environmental factors. While satellite imagery has been widely used for waterbody detection, its limitations necessitate alternative approaches. To tackle these issues, this research introduces three progressive methodologies: Visual Aquatic Generalist (VAGen), Aquatic Prompt Segmentation (APS), and Adaptor-Based Fine-Tuning, forming a stepwise improvement strategy in segmentation accuracy and computational efficiency. The findings contribute to enhancing real-world applications such as urban flood detection, agricultural water management, and ecological monitoring.

Chapter 1 introduces the background and significance of waterbody segmentation using normal-perspective imagery. It discusses the limitations of satellite-based segmentation methods and the challenges posed by real-world conditions. The research questions, objectives, and key contributions of the study are highlighted, providing a foundation for the subsequent methodologies.

Chapter 2 reviews existing waterbody segmentation techniques, including traditional deep learning-based semantic segmentation methods, and their limitations in dealing with complex backgrounds and varying environmental factors. It also introduces foundational models such as Segment Anything Model (SAM) and DINOv2, providing insights into their potential for waterbody segmentation.

Chapter 3 presents the Visual Aquatic Generalist (VAGen), which utilizes Visual Prompting (VP) and In-Context Learning (ICL) to perform waterbody segmentation without retraining a Large Visual Model (LVM) from scratch. By fine-tuning only specific layers, VAGen addresses the challenges posed by limited labeled data and computational constraints, achieving a mean Intersection over Union (mIoU) of 55.71% on the ATLANTIS dataset.

Chapter 4 introduces Aquatic Prompt Segmentation (APS), which enhances segmentation performance by integrating multiple foundation models, including DINOv2, Stable Diffusion, BLIP-2, and SAM. APS employs a modular pipeline with semantic retrieval, Text Guidance, and Mask Guidance, enabling a training-free segmentation approach. It significantly improves adaptability to complex scenarios, achieving a higher mIoU of 68.18% on the ATLANTIS dataset.

Chapter 5 proposes Adaptor-Based Fine-Tuning, which advances segmentation performance by fine-tuning only task-specific adapter modules while utilizing pre-trained backbone models. This approach significantly reduces computational overhead while improving accuracy. With an mIoU of 79.38%, this method surpasses previous frameworks in handling diverse waterbody conditions, making it the most efficient and effective segmentation solution developed in this study.

Chapter 6 provides a comprehensive discussion of the experimental results, analyzing the effectiveness of each proposed method and comparing their performance against baseline models. The chapter evaluates the trade-offs between accuracy, computational efficiency, and generalizability in real-world applications.

Chapter 7 concludes the dissertation by summarizing key findings, discussing limitations, and proposing future research directions. It emphasizes the importance of integrating foundation models for scalable waterbody segmentation and explores potential applications beyond environmental monitoring, such as disaster response and autonomous navigation.

## 論文審査の結果の要旨及び担当者

氏名 ( 趙 家培 )	
	(職) 氏名
論文審査担当者	主査 教授 福田 知弘
	副査 教授 紀伊 雅敦
	副査 准教授 芳澤 信哉

## 論文審査の結果の要旨

本論文では、光線の変化、反射、複雑な環境要因などの主要な課題に対処しながら、通常の視点画像からの水域セグメンテーションを進化させることに焦点を当てている。衛星画像は水域検出に広く利用されているが、その限界から別のアプローチが必要である。これらの問題に取り組むため、本研究では 3 つの先進的な手法を導入している。すなわち、Visual Aquatic Generalist (VAGen)、Aquatic Prompt Segmentation (APS)、Adaptor-Based Fine-Tuning であり、セグメンテーションの精度と計算効率を段階的に向上させている。本研究の成果は、都市の洪水検知、農業用水管理、生態系モニタリングなどの現実のアプリケーションの改良に貢献する。

第 1 章では、通常の視点画像を用いた水域セグメンテーションの背景と意義について紹介している。また、衛星画像を用いたセグメンテーション手法の限界と、実環境がもたらす課題について述べている。さらに、本研究の課題、目的、および主な貢献を明確にし、以降の提案する方法論の基礎となる情報を提供している。

第 2 章では、従来の深層学習に基づくセマンティックセグメンテーション手法を含む、既存の水域セグメンテーション技術と、複雑な背景や様々な環境要因に対応するまでのそれらの限界についてレビューしている。また、Segment Anything Model (SAM) や DINOv2 などの基礎的なモデルを紹介し、水域セグメンテーションの可能性について考察している。

第 3 章では、Visual Prompting (VP) と In-Context Learning (ICL) を利用して、大規模視覚モデル (LVM) をゼロから再トレーニングすることなく水域のセグメンテーションを実行する Visual Aquatic Generalist (VAGen) を提案している。特定のレイヤーのみを微調整することで、VAGen は限られたラベル付きデータと計算上の制約がもたらす課題に対処し、ATLANTIS データセットにおいて 55.71% の mIoU (Intersection over Union) を達成している。

第 4 章では、DINOv2、Stable Diffusion、BLIP-2、SAM などの複数の基盤モデルを統合することで、セグメンテーション性能を向上させた APS (Aquatic Prompt Segmentation) を紹介する。APS は、セマンティック検索、テキストガイド、マスクガイドを備えたモジュラーバイオペラインを採用し、トレーニング不要のセグメンテーションアプローチを可能にしている。APS は複雑なシナリオへの適応性を大幅に改善し、ATLANTIS データセットにおいて 68.18% という高い mIoU を達成している。

第 5 章では、事前に学習されたバックボーンモデルを利用しながら、タスク固有のアダプターモジュールのみを微調整することで、セグメンテーション性能を向上させる、アダプターベースの微調整を提案している。このアプローチは、精度を向上させながら、計算オーバーヘッドを大幅に削減している。mIoU は 79.38% であり、この手法は多様な水域条件に対応する上で従来のフレームワークを凌駕し、本研究で開発された最も効率的で効果的なセグメンテーションソリューションとなっている。

第6章では、実験結果の包括的な考察を行い、各提案手法の有効性を分析し、ベースラインモデルに対する性能を比較している。本章では、実世界のアプリケーションにおける精度、計算効率、汎用性のトレードオフを評価している。

第7章では、重要な発見を要約し、限界について議論し、将来の研究の方向性を提案することで、学位論文を締めくくっている。最後に、スケーラブルな水域セグメンテーションのための基礎モデルの統合の重要性を強調し、災害対応や自律航行など、環境モニタリング以外の応用の可能性を探っている。

以上のように、本論文は環境エネルギー工学の発展に寄与すること大である。

よって本論文は博士論文として価値あるものと認める。