

TorqueSAM: Unsupervised Kidney CT Analysis with Localization and SAM-Integrated Torque Clustering Segmentation

by

Shahzalal Khan Rahat
22101832

Shah Miran Apurbo
22101799

Anika Tazin Pranty
22101886

Mahejabin Yesmin
22101851

Md. Asif Hasan Biplob
24241283

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
February 2026

© 2026. BRAC University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



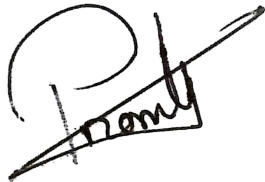
Shahzalal Khan Rahat

22101832



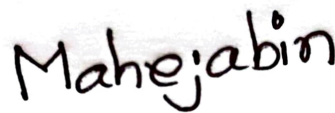
Shah Miran Apurbo

22101799



Anika Tazin Pranty

22101886



Mahejabin Yesmin

22101851



Md. Asif Hasan Biplob

24241283

Approval

The thesis titled “TorqueSAM: Unsupervised Kidney CT Analysis with Localization and SAM-Integrated Torque Clustering Segmentation” submitted by

1. Shahzalal Khan Rahat (22101832)
2. Shah Miran Apurbo (22101799)
3. Anika Tazin Pranty (22101886)
4. Mahejabin Yesmin (22101851)
5. Md. Asif Hasan Biplob (24241283)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering in January 31, 2026.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam, PhD

Professor
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Farhan Faruk

Lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam, PhD

Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Acknowledgement

In the very beginning, we would like to convey our utmost gratitude to the Almighty, the ultimate source of wisdom and power, who has helped us to find the way through this encyclopedic journey. This study would not have been possible without all the assistance, guidance, and support we have received. It is our great pleasure to thank our supervisor, Dr. Golam Rabiul Alam, PhD, who encouraged us with his everlasting support. In addition, our warmest appreciation is for Mr. Farhan Faruk, for his excellent guidance, encouragement, and unwavering assistance that helped us immensely in conducting this research.

Abstract

Chronic kidney disease is a growing public health hazard that often leads to permanent renal system failure if not diagnosed and treated by experts within a treatable time interval. In addition to that, the shortage of expert nephrologists in unprivileged areas slows down the early detection and prevention. However, AI driven solutions can play a vital role in ensuring early detection and prevention despite the shortage of experts in many cases. But most existing AI driven automated renal disease detection systems heavily rely on manually annotated data, which still needs the experts' intervention to detect any chronic renal illness. To address this issue, this study proposes and tests a fully unsupervised method to detect one critical chronic renal disease: Renal stone, often interpreted as a stone that does not rely on any annotated data. In this regard, our proposed method, namely TorqueSAM, introduces a different approach to work with modern detection and segmentation models, along with a classification module to detect any Renal Stone without relying on any ground truth labels. Our approach is tested on a dataset of 6,454 axial and coronal CT scan images. Our study focuses on one alarming illness that is highly dominant and associated with long-term kidney failure. TorqueSAM shows that the unsupervised image segmentation and classification surpass the classical supervised approaches significantly in terms of reducing the annotation overhead, time, and resources. For renal stone, TorqueSAM showed dice scores of 0.5767, IoU of 0.4183, precision of 0.9296, recall of 0.4655, accuracy of 0.9992, and specificity of 0.9998. However, the classification results of the clustering is found to be 0.8232 and 0.8472 for ARI and NMI, respectively. Despite having some limitations in the segmentation, this research shows the potential to outperform the classical supervised techniques, prioritizing the optimization of the region of interest (ROI) to enhance the predictive accuracy while reducing annotation overhead, which allows TorqueSAM to be a scalable solution to mitigate the sufferings of renal stone detection.

Keywords: Segmentation, Clustering, IoU, ARI, Renal Stone, ROI.

Table of Contents

Declaration	i
Approval	ii
Acknowledgment	iv
Abstract	v
Table of Contents	vi
List of Figures	viii
List of Tables	ix
Nomenclature	ix
1 Introduction	1
1.1 Background	1
1.2 Rationale of the Study or Motivation	2
1.3 Problem Statement	3
1.4 Research Questions	3
1.5 Objective	4
1.6 Methodology in Brief	4
1.7 Scopes and Challenges	4
1.8 Key Learnings and Insights	5
2 Literature Review	6
2.1 Preliminaries	6
2.1.1 Kidney	6
2.1.2 Renal Stone	6
2.2 Review Of Existing Research	7
2.3 Summary of Key Findings	9
3 Requirements, Impacts and Constraints	11
4 Proposed Methodology	15
4.1 Methodology Overview	15
4.2 Data Description	16
4.2.1 Data Cleaning	16
4.2.2 Data Transformation and Localization	17

4.2.3	Data Reduction	18
4.3	Proposed Unsupervised TorqeSAM Method	18
4.3.1	The Iterative Approach to Torque Clustering	22
4.3.2	Distance Metric (Euclidean):	22
4.3.3	Torque Calculation:	22
4.3.4	Feature Extraction	23
4.3.5	Torque Clustering: Iterative Approach	24
4.3.6	Key Mathematical Properties:	24
4.3.7	Parameter Tuning Guide	25
4.3.8	MinMaxScaler Normalization and Theoretical Ground	25
5	Result Analysis	26
5.1	Performance Evaluation	26
5.1.1	Evaluation Metrics and Mathematical Formulation	26
5.1.2	Segmentation and Clustering Result Analysis	29
5.2	Analysis of Design Solutions	33
5.2.1	YOLOv8 Localization Training Performance	34
5.3	Final Design Adjustments	35
5.4	Statistical Analysis	35
5.5	Comparisons and Relationships	36
5.6	Discussions	38
6	Conclusion	40
6.1	Summary of Findings	40
6.2	Contributions to the Field	40
6.3	Future Work	40
	Bibliography	43

List of Figures

4.1	Workflow of the TorqueSAM framework, illustrating kidney ROI extraction, stone masking, and torque clustering-based classification. .	15
4.2	CT Images	16
4.3	Localization	17
4.4	YOLOv8 training label visualization showing kidney bounding-box annotations used for kidney localization.	18
4.5	Top-level Overview of Proposed Methodology	19
4.6	Visualization of MedSAM based kidney segmentation and pixel level stone detection. The original CT slice is shown on the left, while the right panel overlays the segmented kidney regions (green) and detected stone candidate (red).	21
5.1	Annotated ground truth examples of kidney stone regions in CT images. Red markers indicate expert labeled kidney stones	27
5.2	Mean evaluation metrics for the proposed TorqueSAM pipeline.	30
5.3	Case 1: Axial CT slice showing stone segmentation results.	30
5.4	Case 2: Coronal CT slice demonstrating segmentation performance for a small renal stone.	31
5.5	Torque Clustering results on synthetic feature distributions, illustrating iterative convergence under uniform and density-aware mass formulations and the final stable cluster configuration after 100 iterations.	32
5.6	Torque Clustering results on real kidney CT superpixel features, showing unsupervised separation between stone related and normal renal regions in the feature space.	33
5.7	Training and validation loss curves and detection metrics for YOLOv8 kidney localization.	34

List of Tables

2.1	Comparison of Existing Unsupervised Segmentation Methods and the Proposed TorqueSAM Framework	10
4.1	Dataset Summary	16
4.2	Cleaned Dataset Summary	16
4.3	Training Hyperparameters for Kidney CT Image Classification	17
4.4	Dataset composition used for training and testing	17
4.5	Dataset composition after annotation	17
4.6	Clustering algorithm parameters	25
5.1	Ablation Study for Segmentation Pipeline Components	36
5.2	Comparison of segmentation performance metrics between the proposed pipeline and classical unsupervised baselines.	37
5.3	Clustering performance comparison using ARI and NMI	38

Chapter 1

Introduction

1.1 Background

Chronic Kidney Disease, or CKD, is a major global health concern with a significant population impact. The increasing rate of CKD, which is being driven by rising rates of diabetes, hypertension, or population aging, makes it an increasingly significant factor in global mortality. Kidney stones, kidney neoplasia (tumors), and the polycystic kidney all have an impact on important kidney functions such as blood filtration, waste disposal, fluid balance, and maintaining blood pressure. As a result, initial and accurate detection is important to avoid progression to severe renal disease.

Computed Tomography scans are the most popular method in renal imaging, offering more accurate 3D cross sectional images than ultrasonography or X-rays, particularly when identifying stones, tumors, and structural anomalies [8]. Furthermore, in some countries like south Asia, rapid diagnosis is very poor because of a significant shortage of experts, less than two per million people, which indicates enormous differences in worldwide renal treatment.

Manual renal segmentation in CT scans for evaluation, surgical preparation or after surgery, is complex, time consuming and dependent on the variation of the patient. Although supervised deep learning methods have shown great performance [4], they require huge labels which are restricted and costly in healthcare applications. As a result, unsupervised methods for segmentation are the vital alternative for targeting clustering and density estimation with no labeling.

Traditional unsupervised techniques, e.g. multiple regional clustering using dynamic active contours show potential for renal segmentation in CT images [3]. Recent improvements on fundamental models have introduced strong segmentation capabilities, where the Segment Anything Model showed exceptional zero shot segmentation performance [16]. Medical adaptations to SAM have led to improvements in cross modality generalization [18], [21], whereas unsupervised frameworks lead to improvements in reliability [40]. SAM is now able to process volumetric medical data thanks to context specific fine tuning processes and 3D extensions [37].

Torque clustering, a physics inspired but pure unsupervised approach, has a mass assignment in terms of local density and uses torque balances to automatically

determine natural clusters which perform better than previous methods on noisy, imbalanced and high dimensional data. Its capability of recognising closed contours makes it suitable for organ border detection. Unsupervised correlation learning based algorithms have shown success management as well as many complicated lesions [38].

Despite all these advancements, unsupervised segmentation of the kidney in CT scans remains a challenging task because of poor tissue contrast, malleable shape between segments, variability of the scans as well as pathological variations [10], [13]. This article overcomes these challenges by presenting TorqueSAM, an unsupervised technique that incorporates the zero shot segmentation capabilities of SAM, specialized with medical images, with the goal of producing accurate, unlabeled renal segmentation for widespread clinical use.

1.2 Rationale of the Study or Motivation

The serious and increasing global burden of CKD that is expected to become an important factor in mortality in the future decades has emphasized the critical need for better diagnostic techniques. Furthermore, illnesses such as kidney stones, kidney tumors, and polycystic kidney have a major effect on the functioning of the kidney but may go unreported due to limited knowledge of advanced medical analysis in resource-constrained settings.

The manual process of labeling CT images for segmentation is very expensive, time consuming, and erratic. Supervised methods of deep learning require large datasets that are labeled, which are limited in medical imaging [4]. On the contrary, unsupervised approaches are of great importance in democratizing the task of proper kidney segmentation strategies, decreasing their reliance on expert annotated data, and thus expanding their scale deployments in clinical situations of cases of expertise shortage. [12], [17].

The introduction of foundation models like SAM has revolutionized segmentation tasks and has given rise to zero shot capabilities without the need for domain specific training. The direct application of SAM to medical imaging has issues because of the domain gaps between real and medical images [18], [26], [30]. Recent research on pseudo label compilation using SAM [20] with hybrid frameworks [25], [32] shows the possibility of integrating fundamental models based on Vision Transformers with domain specific approaches.

The study is motivated by the capacity of physics based torque clustering to robustly distinguish closed outlines, as well as SAM’s outstanding extraction of masks. Merging these methods addresses critical problems in present unsupervised procedures, such as overly clustering and insufficient boundaries compliance [3], [13]. The ultimate goal is to improve early detection and treatment of renal diseases for better patient outcomes.

1.3 Problem Statement

Precisely and automatically segmenting the kidney in CT scans is critical for detecting and tracking renal diseases such as stones and tumors. Traditional supervised segmentation, on the opposite side, takes a long time, has significant inter patient variability, and is inappropriate for large scale clinical use. Although supervised deep learning algorithms are effective [4], but they rely on well labeled data sets, which are costly, morally complicated, and usually unavailable, particularly in impoverished places.

Unsupervised segmentation methods tend to have issues with lower tissue contrast, intensity variation, deformed organ shapes, dimension slices, scanner variation and pathological changes leading to imprecise area of interest detection and lower diagnostic reliability [10], [13]. Although fundamental models like SAM indicate promise, they have limitations in imaging for medical conditions, including shifting domains challenges [26], problems through white tissue structures [26], and necessity for suitable prompt generation [30].

Existing hybrid approaches need considerable fine tuning [28], parameter tuning for classification mechanisms [37], or fail to fully harness the collaboration among physics based methods and fundamental models [37]. As a result, there is a strong demand for a robust, label free framework capable of providing high boundary accuracy in chaotic medical images while maintaining computing efficiency and reliability.

In order to enhance diagnosis and achieve precise segmentation without manual labeling, this study focuses on developing a fully unsupervised framework, TorqueSAM, which combines SAM’s zero shot segmentation capabilities incorporating with an unsupervised classification method with autonomous density and gravity dependent classification.

1.4 Research Questions

1. **RQ1:** How can SAM’s vision transformer model features be successfully combined with physics-based torque clustering to enhance unsupervised kidney segmentation?
2. **RQ2:** Can the suggested TorqueSAM framework beat supervised approaches in segmentation without localization?
3. **RQ3:** How kidney localization can be included as an initial phase to increase overall segmentation accuracy and reduce computational overhead?
4. **RQ4:** How flexible and reliable is the TorqueSAM model to variable intensity distributions, pathological changes, and CT scanner protocols?
5. **RQ5:** What are the practical benefits of mixing physics based torque fields with SAM masks over traditional deep learning approaches?
6. **RQ6:** How does the suggested unsupervised technique stays reliable against the current medical imaging framework on the basis of precision, efficiency, and label independence?

1.5 Objective

The major goal of this study is to establish an effective strategy to outperform the traditional supervised techniques by using modern models, fundamentally based on the concepts of vision transformers, and encoder decoder models with solid incorporation with unsupervised classification methods as well.

Relevant objectives include:

1. Creating a scalable and adaptable system for better CKD diagnosis.
2. Confirming resilience across different imaging circumstances.
3. Creating a robust pipeline that eases the diagnosis process.
4. Validating the stability across various imaging circumstances.
5. Using an automatic approach to identify kidney regions from CT images.
6. Reduce the user level complexity in the diagnosis process.

1.6 Methodology in Brief

In this study, a fully unsupervised framework is designed and implemented to classify the kidney stones. The dataset used in this study has been collected from Kaggle which was developed on images from different local hospitals' CT image databases. This research is only focused on classifying the stones inside a kidney.

There are a total of 1377 images of stones and 5077 images of normal kidneys. The dataset has been scaled down to achieve class balance among the two classes, which are Normal and Stone, resulting in 1300 images per class, a total of 2600 images. For the Region of Interest, a separate YOLO model was trained on the 20% of the images to localize the kidney region with the localization of the ROI. This model forwards the predicted ROI to the SAM's prompt encoder, and SAM generates the ROI masks. Further downstream processing and classification are done using a novel algorithm called Torque Clustering at the end. The core algorithm was adapted to a specialized form and implemented by using the actual algorithm's mathematical intuition.

1.7 Scopes and Challenges

This study focuses on building and verifying an unsupervised approach designed exclusively for kidney stone structures detection in CT imaging data. There are several scopes and challenges that include:

- Segmenting stone, tumor, and cyst together is extremely challenging due to tumors having irregular shapes and intensities, while stones and cysts maintain a strong circularity in shape but behaves opposite in terms of intensity.
- Medical image noise is difficult to handle while extracting features.

- There are limitations in MedSAM while segmenting internal lesions. It is challenging to adapt MedSAM. More variants of SAM can be adapted as per the requirement.
- Kidney’s ROI prediction is difficult as the images have multiple perspectives, with that, the kidney’s position also changes.
- Adapting the core Torque Clustering to fulfill our purpose is a challenging task. Moreover, this algorithm is a novel approach, it does not have any benchmarks in this field of study.
- The integration of a gravity based clustering algorithm with a vision transformer based model like MedSAM can incredibly increase the efficiency of the entire process compared to the computational algorithms similar to Quickshift.

The framework is created to be used with 2D axial CT slices. While the technology is adaptable, this study focuses on kidney stone classification through segmentation masks to produce proof of concept and evaluate the SAM based torque clustering strategy in a clinically important domain.

1.8 Key Learnings and Insights

SAM requires domain-specific modifications to be effective for medical imaging [26], [30]. Therefore, this study implies a special variant of SAM that is trained specifically on medical images. MedSAM might not detect internal lesions due to a lack of training. In that case, a pixel level approach could be followed to detect specific stone masks explicitly. SAM with torque clustering is a hybrid approach both combined learned features and gravity-based grouping to enhance classification performance, ensuring a fully unsupervised method. Additionally, kidney localization enhances efficiency and accuracy. However, the localization is predicted through the Yolo model and might require parameter tuning while training on the annotated ROIs’ to detect the kidney ROI more accurately from different perspectives of image.

Chapter 2

Literature Review

2.1 Preliminaries

This section outlines the anatomical characteristics of the kidney and the pathological manifestations of renal stone, highlighting their clinical relevance and imaging characteristics. Understanding these aspects is essential for analyzing kidney- and stone-focused segmentation methods discussed in the subsequent literature review.

2.1.1 Kidney

The kidney plays a central role in the urinary system by filtering blood, excreting metabolic waste, and regulating fluid balance. Structurally, it is composed of the renal cortex, medulla, and collecting system, which together support filtration and urine transport. In clinical imaging, especially computed tomography (CT), the kidney appears as a well-defined organ but often exhibits intensity variability due to contrast phases, scanner settings, and patient-specific factors.

Accurate kidney delineation is critical for volumetric analysis and for isolating pathological regions such as stones or tumors. However, kidney segmentation is complicated by weak organ boundaries, similarity to surrounding tissues, and anatomical variability across patients. These challenges are well recognized in kidney-focused segmentation studies, where reliable kidney extraction is treated as a prerequisite for downstream pathological analysis [8], [13].

2.1.2 Renal Stone

Renal stones, commonly referred to as kidney stones, are solid mineral formations that develop within the renal collecting system. They represent a widespread urological condition and are associated with recurrent disease and long-term renal complications. Stones vary significantly in size, shape, location, and composition, resulting in diverse imaging appearances across patients.

Non-contrast CT is the preferred imaging modality for stone detection due to the high radiodensity of stones relative to soft tissue. Despite this contrast advantage, accurate stone segmentation remains challenging. Small stone size, irregular morphology, partial volume effects, and imaging noise often lead to

missed detections or fragmented segmentation results. While supervised approaches demonstrate strong performance when annotations are available [8], unsupervised and domain-adaptive studies highlight the difficulty of reliably identifying stones without labeled guidance [13].

2.2 Review Of Existing Research

Primary research in unsupervised medical image segmentation was based mainly on classical clustering and region based techniques. Vishnuvarthanan et al. [1] developed a strategy for unsupervised clustering for tumor identification and tissue segmentation in the brain based on MRI, showing the existence of anatomically meaningful regions that can be extracted without the aid of labeled data. While successful for relatively simple structures, the approach was extremely sensitive to noise, parameter choice and boundaries with low contrast, and it showed the low robustness of intensity only clustering approaches in complex clinical scenarios.

In order to overcome the problems of over segmentation and locality limitations of classical approaches, Aganj et al. [2] proposed an unsupervised segmentation algorithm based on the local center of mass (CM). The method incorporated global image structure into the segmentation process by computing one-dimensional local CMs in several orientations and accumulating them. Experiments on X-ray and three dimensional MR images indicated improved Dice scores and less over segmentation than watershed, GMM HMRF, and SLIC. A refined formulation was later proposed by Aganj [6], which stressed on global consistency and robustness to weak edges. However, computational complexity and dependency on manual parameter tuning restricted the scalability to high resolution volumetric data.

Beyond handcrafted grouping strategies, representation learning in combination with clustering and alignment strategies have gained grounds. Chen et al. [4] proposed SIFA, a synergistic image feature alignment framework for unsupervised cross modality medical image segmentation. Using adversarial learning and cycle consistency, both the appearance representations and semantic representations were aligned with significant improvements in Dice scores on cardiac and abdominal MRI CT datasets. Nevertheless, the slice wise two dimensional design and instability of adversarial training restricted volumetric consistency and real-world applicability.

Similarly, Yao et al. [10] introduced a three dimensional unsupervised domain adaptation framework using content style disentangled generative adversarial networks coupled with a dual attention 3D UNet. By synthesizing multi style target domain images and employing voxel wise attention, their approach significantly outperformed two dimensional domain adaptation methods, particularly for small anatomical structures. Despite these gains, the two stage design incurred high computational costs and sensitivity to style clustering quality.

Shape aware unsupervised segmentation has also been explored. Yang et al. [7] proposed learning explicit organ shape priors using a variational autoencoder to guide segmentation in unlabeled target domains. While effective for organs with stable anatomical geometry, such assumptions limit applicability to irregular or pathological structures. Recent work has emphasized contrastive objectives to enhance unsupervised representation learning. Li et al. [12] proposed a

contrastive autoencoder jointly optimizing reconstruction and clustering losses, enabling improved latent space separation across patients without annotations. However, performance depended heavily on augmentation strategies and lacked native three-dimensional support. Liu et al. [17] introduced CUTS, a fully unsupervised deep and topological framework for multigranular medical image segmentation. The method learns pixel centered patch embeddings through intra image contrastive learning and reconstruction, followed by diffusion condensation for hierarchical clustering. Strong performance was reported on retinal and brain MRI datasets, though reliance on patch-wise processing limits scalability to volumetric data.

Diffusion-based segmentation has gained attention in recent years. Wu et al. [19] proposed a diffusion-based framework incorporating dynamic conditional encoding and frequency aware filtering. Zhang et al. [29] further extended this idea using spectrum space transformers and anchor conditioning. Despite high accuracy, diffusion based approaches remain computationally intensive and predominantly supervised. Unsupervised domain adaptation using diffusion models was explored by Zhao et al. [22], where synthetic target domain image mask pairs were generated to reduce domain gaps. While near supervised performance was achieved, the approach was dependent on labelled source data. Foundation models have also had an impact on medical image segmentation research. Kirillov et al. [16] introduce the Segment Anything Model (SAM), a promptable foundation model with the ability to perform zero shot segmentation. Because of a change in domain, SAM has inconsistent performance on medical images. To overcome this issue, Ma et al. [18] proposed MedSAM, which showed better performance using medical domain adaptation.

Zhang et al. [21] further improved SAM by learnable adapter layers and interactive training using a large number of medical datasets. Although generalization improved, the approach remained dependent on prompts and extensive computational resources. Several studies incorporated SAM into weakly supervised or unsupervised pipelines. Xie et al. [20] used SAM generated masks as pseudo labels refined through clustering and knowledge distillation. However, segmentation quality remained strongly dependent on the initial SAM proposals. Slice to volume consistency was addressed by Ding et al. [9], who proposed S2VNet, a clustering propagation framework transferring segmentation knowledge across slices using two-dimensional networks. While efficient, the approach is not fully unsupervised. Kidney-focused segmentation studies provide application-specific context. Cui et al. [8] proposed an nnUNet-based framework for kidney stone segmentation from non-contrast CT under full supervision. Liu et al. [13] explored unsupervised kidney and tumor segmentation under domain shift using wavelet-based style mining and contrastive adaptation.

Classical and hybrid approaches were investigated by de Boisredon d’Assier et al. [3], combining multi-region clustering with adaptive active contours for unsupervised kidney segmentation. Feng et al. [5] applied convolutional autoencoder embeddings followed by clustering for hyperspectral biomedical segmentation. Wang et al. [14] proposed variance reduced stratified contrastive learning, while Zhou et al. [15] introduced transformer based unsupervised domain adaptation with meta attention. Efficiency oriented pipelines were proposed by Alexandru and Popescu [32], combining YOLO based kidney localization with UNet based slice wise segmentation

and three dimensional reconstruction. Although kidney segmentation accuracy was high, tumor boundary performance remained limited. Unsupervised deformable registration was explored by Balaha et al. [33], who proposed a hybrid framework combining UNet-based deformation estimation with cascaded spatial transformer networks. Efforts to improve foundation model reliability include MedSAMU, proposed by Zhou et al. [40], using uncertainty-guided multi-prompt adaptation. Generative reconstruction aware segmentation was explored by Prakash et al. [25] through SculptorGAN.

Context-aware adaptation of foundation models was studied by Zeng et al. [28] using LoRA adapters and contextual prompting. Navinchandra et al. [37] extended SAM2 to volumetric kidney and tumor segmentation, demonstrating promising zero-shot performance. In addition to studies focused on the kidney, Xu et al. [38] suggested a correlation-based unsupervised lesion clustering, and Rodrigues et al. [26] tested the generalization limits of MedSAM. Zhou and Yu [30] enhanced the segmentation of brain tumors through the integration of the prompts generated through detection into MedSAM.

2.3 Summary of Key Findings

The reviewed studies indicate a clear evolution in the development of unsupervised medical image segmentation, starting with classical clustering and region based methods, and moving towards deep representation learning, contrastive objectives, and foundation models assisted methods. Classical approaches show the feasibility of labeling free segmentation, but have issues of sensitivity to noise, parameter dependency, poor boundary handling. More recent deep learning based methods enhance feature representation and segmentation accuracy, taking advantage of adversarial learning, contrastive learning, and generative modeling, especially in cross-modality and domain adaptation cases.

Despite these advances, there are a number of persistent limitations that are evident across the literature. Although there are a lot of state-of-the-art methods showing better performances, most of them are based on complex and computationally expensive pipelines, such as diffusion models, multi stage training, or patch-wise processing, which reduces the scalability to high-resolution volumetric data. Foundation models such as SAM and MedSAM offer promising zero-shot and weakly supervised capabilities but remain sensitive to prompt selection and domain shift, often requiring additional adaptation or supervision. Consequently, achieving fully unsupervised, volumetrically consistent, and robust segmentation for complex anatomical and pathological structures remains an open challenge.

Table 2.1: Comparison of Existing Unsupervised Segmentation Methods and the Proposed TorqueSAM Framework

Aspect	Classical & Clustering Based	Adversarial / Domain Adaptation	Contrastive & Patch Based	Diffusion / Foundation Models	TorqueSAM
Supervision Level	Fully unsupervised	Unsupervised / weakly supervised	Fully unsupervised	Zero shot / weakly supervised	Fully unsupervised
Core Principle	Intensity based clustering and region heuristics	Feature alignment using adversarial learning	Latent representation learning via contrastive objectives	Prompt based or generative modeling with large pre trained models	Geometric torque modeling with Madsem structural encoding and contrastive learning
Representation Type	Hand crafted or low level features	CNN based semantic representations	Patch level or latent embeddings	Foundation model or diffusion latent space	Structure aware volumetric representations
Volumetric Consistency	Limited or absent	Partial, often slice wise	Limited due to patch wise processing	Often weak or 2D centric	Explicitly volumetric and globally consistent
Robustness to Noise	Low	Moderate	Moderate to high	High but prompt sensitive	High due to torque driven structural cues
Handling Domain Shift	Poor	Strong via explicit adaptation	Moderate	Moderate to high	Implicitly handled through geometry aware contrastive learning
Computational Complexity	Low	High due to multi stage training	Moderate to high	Very high	Moderate and scalable
Dependency on Hyperparameters / Prompts	High	High	Moderate	Very high	Low and self guided
Scalability to High Resolution 3D Data	Poor	Limited	Limited	Restricted by memory constraints	Designed for scalable 3D volumes
Generalization to Irregular Pathologies	Poor	Moderate	Moderate	Variable	Strong without rigid shape priors
Key Limitations	Noise sensitivity and over segmentation	Training instability and heavy computation	Weak global structure modeling	Domain shift and prompt dependence	Addresses robustness, scalability, and volumetric consistency
Overall Positioning	Baseline unsupervised methods	Powerful but complex pipelines	Locally effective representation learning	Strong zero shot capability	Unified, geometry aware, fully unsupervised framework

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

The proposed TorqueSAM framework is developed as a fully unsupervised kidney segmentation system which works on two dimensional axial CT slices. The core requirement is the combination of a foundation segmentation model and physics-inspired clustering mechanism to avoid the dependence on annotated data, which specifically addresses the research objectives and RQ1-RQ7 mentioned in Chapter 1. The system makes use of SAM’s pretrained image encoder to produce robust multi scale representations followed by torque based clustering for refining closed kidney contours to suppress over segmentation motivated by previous unsupervised kidney segmentation works [18], [21] and SAM based medical adaptations [9], [2], [10]. From a software point of view, it requires Python and

PyTorch as the main deep learning framework for its implementation. Publicly available medical imaging libraries for DICOM handling and preprocessing are needed to process CT scans in the same way as previous studies [29], [21]. The SAM architecture and its medical variants MedSAM as such represent a feature extractor without supervisor fine tuning leaving the unsupervised nature of the pipeline intact [9], [2]. Torque clustering and density based grouping are parameter light postprocessing modules to ensure reproducibility and interpretability as opposed to heavily tuned clustering [4], [18]. Hardware requirements are intentionally modest

so that it can be deployed in resource constrained environments. A single GPU (Nvidia) with at least 8 GB VRAM is enough for feature extraction and clustering based inference, which is consistent with computational feasibility reported in recent SAM assisted segmentation studies [10], [38]. The framework is intended to work on public CT datasets with structures and common pathologies of the kidney, without using proprietary or ethically restricted clinical data.

3.2 Societal Impact

The most important societal effect of TorqueSAM is on accessibility to automated kidney analysis in areas where there is limited access to radiological expertise, such as South Asia. Chronic kidney disease and other renal diseases need early diagnosis to prevent irreversible disease progression, but expert segmentation is

lacking in low resource healthcare systems [29], [21]. By removing the need for annotated data, the proposed system offers a way to reduce the reliance on highly trained specialists and make the system scalable. From the point of view of healthcare, unsupervised segmentation promotes clinical decision making in terms of consistent region of interest extraction for further diagnosis, treatment planning, and longitudinal monitoring. Unlike supervised models, which are prone to performance decline in scenarios with limited or biased labels, TorqueSAM makes use of domain agnostic representation and physics based reasoning, which improves robustness between scanners and patient populations [18], [9]. Legally and culturally, the framework is placed in a very clear position of being strongly a decision support tool rather than a diagnostic system in itself. This is consistent with the conservative implementation practices undertaken in the current medical AI systems, especially for foundation model based methods where limitations in generalization have been observed [32], [25]. No direct patient interaction or intervention is involved and there are reduced regulatory and safety risks.

3.3 Environmental Impact

The TorqueSAM has a relatively low environmental footprint in comparison with fully supervised pipelines of deep learners. Supervised segmentation of medical images is usually implemented in repeated training sessions on large labels at considerable computational and energy expenses [7], [22]. Conversely, the proposed system takes advantage of the pretrained SAM encoders and lightweight clustering, which lower the training cost. Computation at inference time is the most resource consuming, and conducting experiments is done on local GPUs instead of being executed on a massive cloud infrastructure, which also reduces the energy consumption. Another impact of reusing pretrained foundation models is that they facilitate sustainable AI practices by spreading the environmental cost of training large scale models to many applications [9], [2]. Generally, TorqueSAM complies with the principles of sustainable research as it targets on efficiency and reuse rather than retraining.

3.4 Ethical Issues

Ethical issues are mostly related to data utilization, favoritism, and accountable conclusion of findings. This research is based solely on publicly available datasets of anonymized CT, and patient privacy and adherence to the common ethical standards of research in medical imaging [21], [12]. No additional clinical data collection or human subject experimentation is done. One of the possible issues is bias because uneven distribution of pathologies, scanners, and demographics is present in public datasets. Nonetheless, TorqueSAM helps avoid biasing the annotation, and overfitting to labelling which is unique to an institution, a concern with supervised segmentation systems [7], [32]. Transparency is boosted by the visualization of torque field, where clinicians and researchers interpret the segmentation behavior instead of being dependent on the black box predictions. The system is not aimed at substituting the clinical judgment in an ethical way. In line with the suggestions

in SAM based medical imaging research, the outputs have to be confirmed by the medical practitioners prior to any clinical application [2], [30].

3.5 Standards (if applicable)

The project meets the set standards of medical imaging and software development. The management of CT data is based on the DICOM requirement of standardization and interoperability of scanners and datasets. Professionally, the system design is consistent with the common IEEE principles of ethical AI development, which are transparency, reproducibility, and human supervision. No regulatory certification (e.g. FDA approval) at a device level is done, since the work is in the academic research realm. The same limitations can be seen in recent studies of unsupervised and SAM based segmentation [10], [38] and thus formal clinical standards cannot be applied at this point.

3.6 Project Management Plan

The team undertaking the project is composed of five members and the tasks are split into literature review, model design, implementation, experimentation and documentation. Chapter 3 is put together as part of the system level planning and impact analysis. The entire project schedule is about half a year, the first half year covers literature review, problem formulation, model implementation, evaluation and thesis writing. The example of a schedule that takes high levels is presented that encompasses such milestones as the SAM integration, the development of the torque clustering, the experimental validation, and the final report preparation. The resource needs can consist of local access to the GPU, open source software and supervisory guidance. No additional funding is necessary other than basic calculating tools.

3.7 Risk Management

Project outcomes are susceptible to a number of risks. Shifting domain between CT scanners and patient groups can lead to low quality segmentation as it has been commonly observed in unsupervised and SAM based medical imaging literature [22], [32]. The mitigation of this risk is through the addition of torque clustering, which focuses on geometric as well as density based consistency as opposed to an entirely learnt intensity feature [18]. The other threat is over segmentation or under segmentation of the original SAM outputs [10]. This is explicitly addressed at the torque based refinement stage which imposes closed contour coherence. Computational limitations Up to now there is little risk in computing, since the system is single gpu. Frequent validation and inspection are used to identify cases of failures in time.

3.8 Economic Analysis

The economic importance of the TorqueSAM framework is that it can lessen the main structural expenses of the medical image segmentation systems. Traditional supervised methods imply a lot of expert annotation, repetitive dataset curation,

and task specific retraining which are associated with high development and maintenance costs [7], [21]. These demands restrict scalability and in many cases, restrict inter institutional or imaging protocol reuse of models. TorqueSAM has made a fundamental change to this cost structure by eliminating annotation as a repeated economic dependency since it embodies a completely unsupervised design. Additional economic efficiency gains are the integration of the pretrained foundation models that relocate the use of resources in data acquisition and instead use the model reuse. The ability to use SAM based representations enables the system to take advantage of training experiences of all large scale training in the past but without having to pay the cost of training per new dataset [9], [2]. This reuse based approach fits recent frameworks of unsupervised and pseudo label based segmentation, in which computational effort is focused on refinement, as opposed to retraining the model [10]. In the long term view, the refinement of lightweight clustering design in the framework and a small scale of computations facilitate sustainable implementation and maintenance. Continuous labelling is not done, and heterogeneous CT data are handled flexibly, which makes the operations overhead lower and makes the cost of large scale adoption more economical [18], [17]. On the whole, TorqueSAM offers a segmentation paradigm, which is economically advantageous due to the reduction of recurrent costs, the reuse of the knowledge trained once, and the ability to scale its application without relying on the expertise at expensive rates.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

This study implies an unsupervised segmentation and classification framework for identifying renal stones using CT images. Regions of interest are retrieved using localization and then utilized as box prompts for MedSAM followed by RGB conversion and resizing. To address MedSAM’s restriction in lesion-specific identification, pixel-level masking is performed to identify the stone mask smoothly. Finally, an adapted version of the novel torque clustering algorithm is used on a 2D feature space to perform unsupervised clustering for classification.

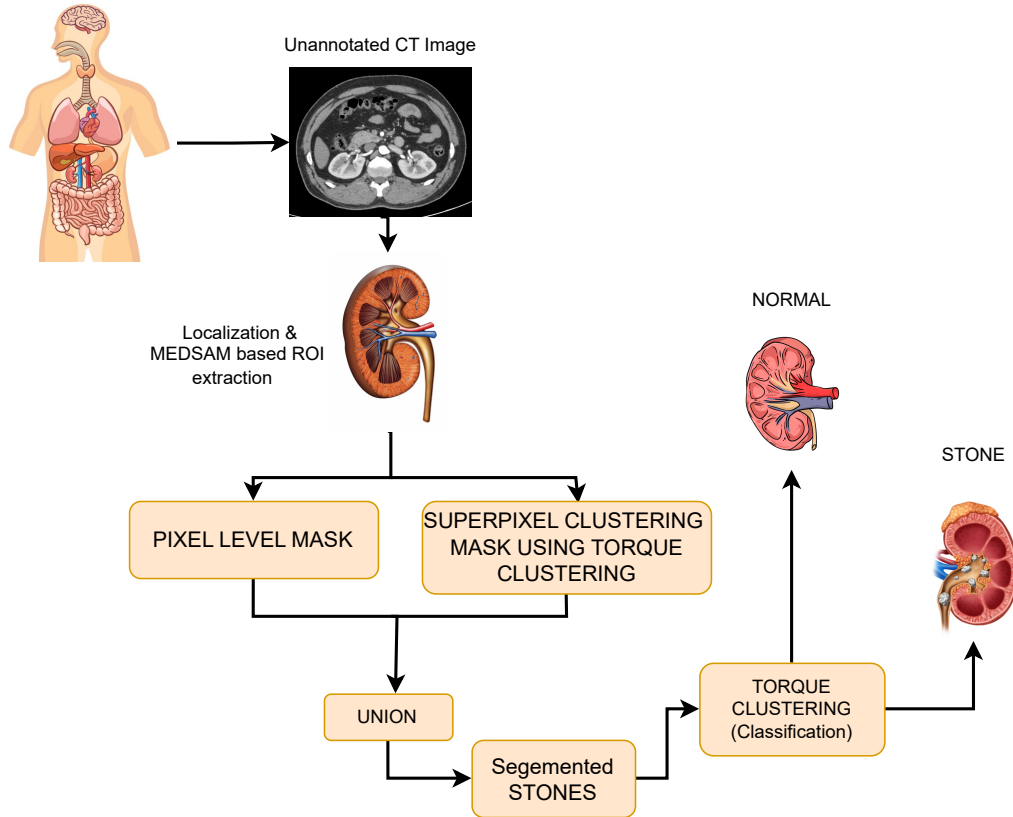


Figure 4.1: Workflow of the TorqueSAM framework, illustrating kidney ROI extraction, stone masking, and torque clustering-based classification.

4.2 Data Description

The dataset for this study comprises a comprehensive collection of medical CT images of kidneys, having four major classes: Normal (5,077 images), Stone (1,377 images), Tumor (2,283 images), and Cyst (3,709 images), developed for experiments to diagnose the disease. It consists of 12,446 CT images in coronal and axial perspectives. The dataset contains a large number of samples with a wide range of kidney conditions. It is an ideal dataset to run experiments in the field of medical imaging.[11]

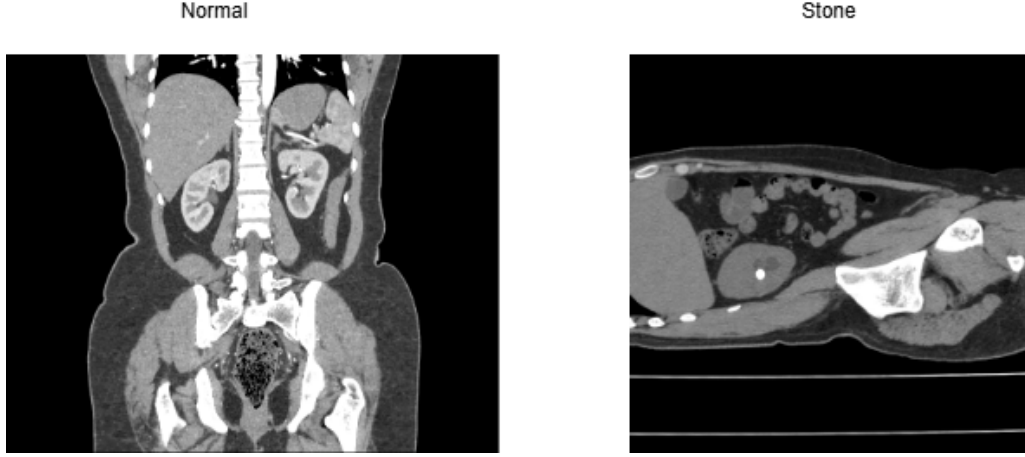


Figure 4.2: CT Images

4.2.1 Data Cleaning

The CT kidney dataset was thoroughly cleaned before any experiment due to having a massive class imbalance among the samples. In this study, only the stone and normal classes were chosen to be classified in an unsupervised approach. 1,300 images from each of the classes, a total of 2,600 images were chosen to proceed further with the detection. As a result, we achieved a perfectly balanced dataset to avoid any bias in our experiments and inferences.

Table 4.1: Dataset Summary

Class	Number of images
Normal	5077
Stone	1377
Total	6454

Table 4.2: Cleaned Dataset Summary

Class	Number of images
Normal	1300
Stone	1300
Total	2600

4.2.2 Data Transformation and Localization

In the work plan of this study, the ROI has to be predicted through a detection model. We have localized the kidney ROI with the top 20% of the sample images from the dataset to train the YOLOv8 m model. Each classes had 260 images, a total of 520 images for the YOLOv8 m training. The rest of the images were kept in the test set for further segmentation and clustering.

Table 4.3: Training Hyperparameters for Kidney CT Image Classification

Parameter	Value
Input Image Size	512×512
Batch Size	16
Epochs	80
Learning Rate	0.001667
Optimizer	AdamW

Table 4.4: Dataset composition used for training and testing

Dataset	Total images	Images per class	Classes
YOLO training	520	260	Normal, Stone
Testing	2080	1040	Normal, Stone

Table 4.5: Dataset composition after annotation

Class	Actual	Annotated
Normal	5077	No
Renal stone	1377	Yes
Total	6454	1377

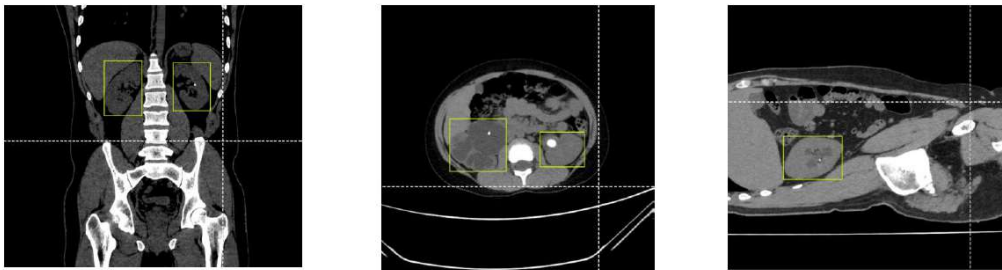


Figure 4.3: Localization

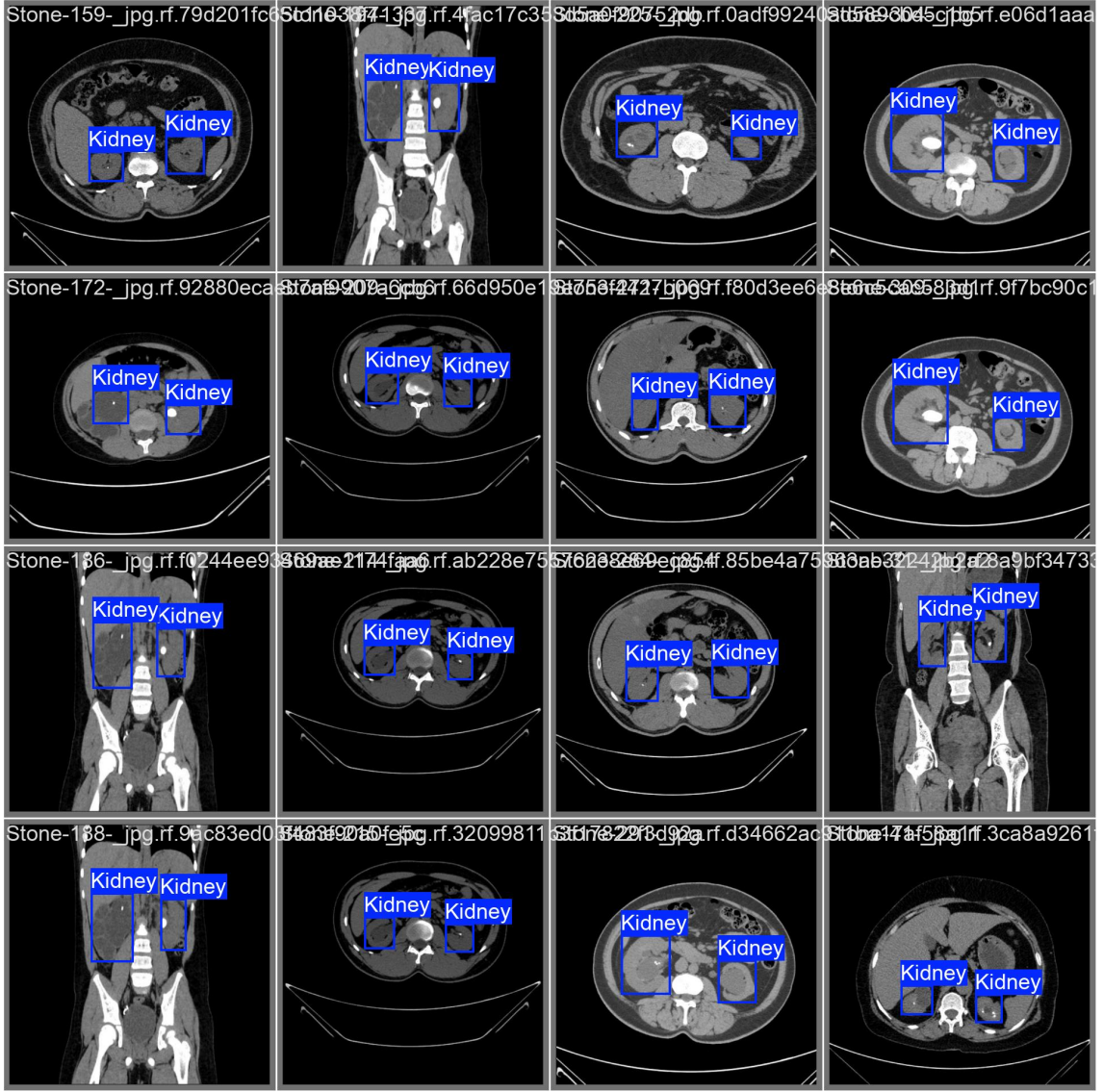


Figure 4.4: YOLOv8 training label visualization showing kidney bounding-box annotations used for kidney localization.

4.2.3 Data Reduction

After the MedSAM segmentation and detection of the kidney stone mask, the final output mask is in a binary form. We reduced the extracted features' dimension from the masks to a minimum of two to plot a 2D scatter plot while clustering. As the adapted torque clustering algorithm's time complexity is $O[n^2]$, the feature space must be kept in the lowest dimension possible to avoid computational overhead for large scale data.

4.3 Proposed Unsupervised TorqeSAM Method

This section presents the entire pipeline of a fully unsupervised method to segment and classify kidney stones as shown in 4.5. There are multiple module that works together to perform the unsupervised approach.

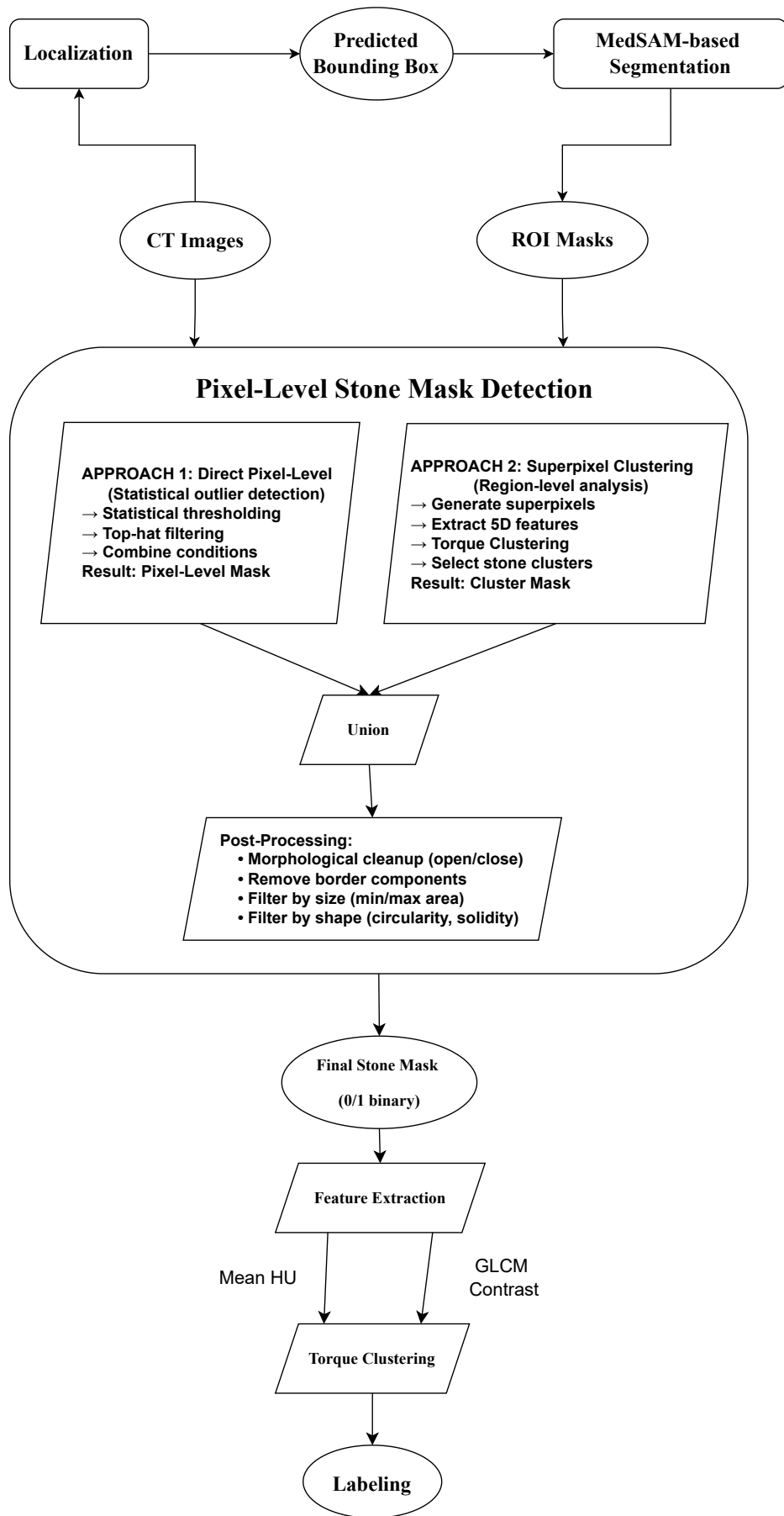


Figure 4.5: Top-level Overview of Proposed Methodology

1. **Kidney ROI Extraction:** This study uses a YOLOv8 m model for kidney localization. For the training, the processed dataset’s subset that contains the 20% of the images were localized. A bounding box was defined, only capturing the kidney area in different perspectives of the images. The training hyperparameters are presented in the table 4.3. The specific coordinates of the bounding box were used to crop the CT image to only find the specific ROI on the image.

2. **MedSAM and Pixel level Detection:**

Following kidney region localization, MedSAM is used to acquire anatomically accurate kidney segmentation masks that are used as hard spatial constraint for any follow up pathology analysis. MedSAM, which is a medical adaptation of the Segment Anything Model, is utilized in inference mode without any task specific fine tuning. Bounding boxes predicted during the kidney ROI extraction stage are fed to the MedSAM as prompts to generate high quality kidney masks by preserving the structure coherence across slices.

Let the cropped, CT image be represented as

$$I \in R^{H \times W} \quad (4.1)$$

and corresponding MedSAM output mask as

$$M \in \{0, 1\}^{H \times W}. \quad (4.2)$$

The segmentation mask M defines the anatomy of kidneys at a pixel level, and is considered to be a strong prior on anatomy. All the pixels which lie outside the kidney region are removed from further processing and the stone detection will be processed only in the anatomically meaningful regions.

In order to stabilize downstream computations, the CT image is normalized and converted to unsigned 8 bit representation while the MedSAM output is binarized. A validity check is done by calculating the number of foreground pixels,

$$N_k = \sum_{x,y} M(x, y), \quad (4.3)$$

and slices for which there is not enough kidney coverage are discarded to avoid unstable feature statistics. This step is critical in the unsupervised situation, in which unreliable region estimates can cause a significant amount of bias in clustering behavior. Once the valid kidney mask is acquired, a pixel level stone enhancement stage is applied in the kidney area. Renal stones usually present as small, high intensity structures compared to the rest of the renal structure. To use this property a morphological white top hat transformation is performed:

$$I_{\text{tophat}} = I_M - (I_M \circ B), \quad (4.4)$$

where

$$I_M = I \odot M, \quad (4.5)$$

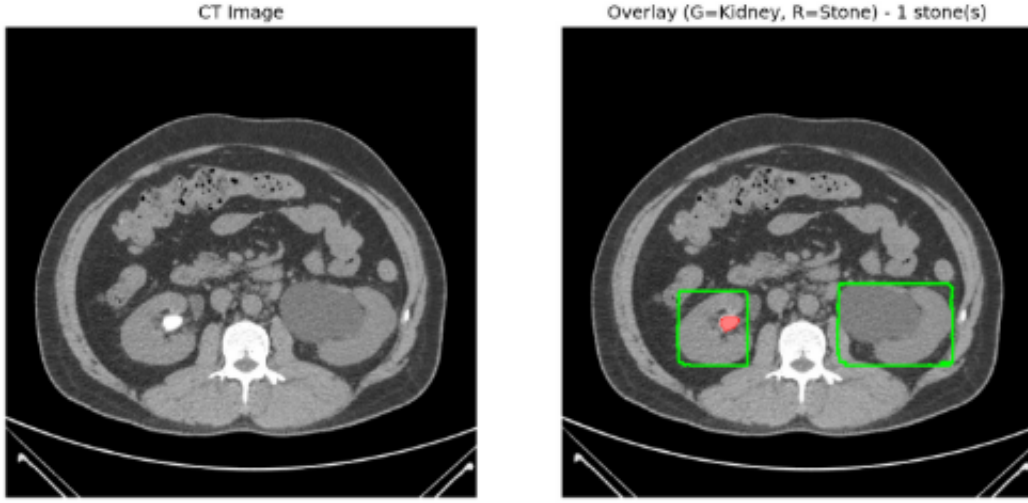


Figure 4.6: Visualization of MedSAM based kidney segmentation and pixel level stone detection. The original CT slice is shown on the left, while the right panel overlays the segmented kidney regions (green) and detected stone candidate (red).

◦ signifies morphological opening, and where B is a structuring element whose size is used to avoid large scale intensity variations, while preserving compact bright objects. This operation strengthens local intensity maxima which correspond to possible stone candidates. Candidate stone pixels are first detected on the basis of adaptive intensity and contrast criteria that are based on local statistics within the kidney region. These candidates are not considered final detections but instead are considered high confidence seeds for further analysis at the region level. Importantly, pixel level detection is made conservative, for the sake of precision in order to avoid feeding false positive statistics to the unsupervised clustering stage. The combination of MedSAM based anatomical masking and pixel level enhancement has two important purposes in the pipeline of TorqueSAM. First, it imposes strict anatomical consistency such that pathological inference is never made outside the kidney. Second, it provides a good low level signal on possible areas of stones without being agnostic on pixel wise labels. To this end, this design allows moving from pixel level cues to superpixel based feature aggregation and Torque Clustering, which provides a principled bridge between anatomical segmentation and unsupervised pathology separation.

3. **Torque Clustering:** In this study, a novel unsupervised clustering algorithm is adapted to a specialized form to achieve the purpose of this experiment. The core architecture of this module includes feature extraction to extract radiomics features such as HU and GLCM. The algorithm has two distinct variants: the Iterative Method and the Connection based Method.

4.3.1 The Iterative Approach to Torque Clustering

This algorithm displays a series of mathematical and physics based intuition to make a data point behave with gravity that pulls nearby points towards the centroid data point, depending on the torque value. Here, the Euclidean space is utilized and used for the distance calculation in 4.6.

4.3.2 Distance Metric (Euclidean):

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (4.6)$$

There are multiple mass strategies to assign mass to each data point based to density, Several methods are showcased to assign masses to data points:

1. Uniform: Every point has the same mass, specifically 1. This is an unbiased baseline setup.

$$m_i = \frac{1}{n} \quad \forall i \quad (4.7)$$

2. Density: Mass is proportional to local density (using k nearest neighbors).

$$m_i = \frac{k}{d(i, k\text{-NN})} \quad (4.8)$$

Here, k is the neighbor count, $d(i, k \text{ NN}) = \text{distance to } k\text{th neighbor}$. This setup ensures the dense region gets a stronger influence.

3. Feature Sum: Mass is proportional to the sum of a point's absolute feature values.

$$m_i = \frac{\sum_{j=1}^d |x_{i,j}|}{\sum_i \sum_j |x_{i,j}|} \quad (4.9)$$

4. Feature Variance: Mass is proportional to the variance of a point's features.

$$m_i = \frac{\text{Var across features}}{\sum_j \text{Var across features}} \quad (4.10)$$

This flexibility allows different notions of importance for each data point.

4.3.3 Torque Calculation:

A net force or “torque” is calculated for each point, summing contributions from all or k nearest neighbors. Points with higher masses pull more strongly; points far away pull more weakly (often inversely proportional to distance or its square).

Direction Vector:

$$\mathbf{d}_{ij} = \frac{\mathbf{x}_j - \mathbf{x}_i}{\|\mathbf{x}_j - \mathbf{x}_i\|_2 + \epsilon} \quad (4.11)$$

Net Torque on Point i :

$$\tau_i = \sum_{j \neq i} \frac{m_j}{d_{ij}^{p+1}} \cdot (\mathbf{x}_j - \mathbf{x}_i) \quad (4.12)$$

The intuition is that each point shall experience a pulling force from other points proportional to their mass, weakening with distance. The higher the mass, the harder the point pulls.

4.3.4 Feature Extraction

Algorithm 1 Mean HU Extraction

Require: CT image slice (float32), binary mask

Ensure: $\text{mean_hu} \in R$

```

1: function COMPUTEMEANHU(image, mask)
2:   region  $\leftarrow$  image[mask == 1] ▷ Extract masked pixels
3:   if |region| = 0 then
4:     return 0.0 ▷ Empty mask case
5:   end if
6:   mean_hu  $\leftarrow \frac{1}{|\Omega|} \sum_i \text{region}[i]$  ▷ Simple mean
7:   return float(mean_hu)
8: end function

```

Algorithm 2 GLCM Based Contrast Computation

Require: CT image, binary mask, levels = 32

Ensure: contrast $\in [0, \infty)$

```

1: (ys, xs)  $\leftarrow$  find coordinates where mask == 1
2: (x1, x2), (y1, y2)  $\leftarrow$  bounding box of mask
3: crop  $\leftarrow$  image[y1 : y2, x1 : x2]
4: region  $\leftarrow$  crop[crop_mask == 1]
5: if |region| < 10 then
6:   return 0.0 ▷ Too few pixels
7: end if
8: Normalize region to [0, levels):
9:   norm  $\leftarrow \frac{\text{region} - \min}{\max - \min} \times (\text{levels} - 1)$ 
10: Compute GLCM matrix (32 × 32):
11:   G[i, j] = count of adjacent pixel pairs with values i, j
12:   Distance: 1 pixel
13:   Direction: 0° (horizontal)
14:   Symmetric, normalized
15: Contrast  $\leftarrow \sum_n n^2 \sum_i G[i, i + n]$  ▷ Measures texture variation
16: return float(contrast)

```

4.3.5 Torque Clustering: Iterative Approach

Algorithm 3 Torque Based Clustering Algorithm

Require: data $\in R^{n \times d}$, learning rate $\alpha \in (0, 1)$, max iterations $T_{\max}$, tolerance ε , masses m_i , neighbor scope $k_{nn} \in \{\infty, k\}$, distance power $p \in \{1, 2\}$

Ensure: cluster_labels, transformed_data, iterations

```

1: data_transformed  $\leftarrow$  copy(data)
2: distances  $\leftarrow$  zeros( $n, n$ )
3: for  $t \leftarrow 1$  to  $T_{\max}$  do
4:   distances  $\leftarrow$  calculate_distance_matrix(data_transformed)
5:   for each point  $i$  do
6:      $\tau_i \leftarrow \sum_{\text{neighbors } j} \frac{m_j}{d_{ij}^p} \cdot \frac{(x_j - x_i)}{\|x_j - x_i\|}$ 
7:   end for
8:   data_transformed  $\leftarrow$  data_transformed  $+\alpha \times$  torques
9:    $\Delta^{(t)} \leftarrow \text{mean}(\|x_{\text{new}} - x_{\text{old}}\|)$ 
10:  if  $\Delta^{(t)} < \varepsilon$  then
11:    break
12:  end if
13: end for
14: return cluster_labels, transformed_data, iterations

```

Complexity: $O(n^2)$

Key Insight: Mass distribution determines where the points are attracted to converge. Denser regions with higher mass become cluster centroids.

4.3.6 Key Mathematical Properties:

The data points move according to gradient descent towards convergence following the equations 4.13 and 4.14.

Gradient descent updates and an average point displacement is set to a certain level of threshold for convergence criterion.

Points move according to gradient descent:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \alpha \mathbf{r}_i^{(t)} \quad (4.13)$$

Convergence criterion:

$$\Delta^{(t)} = \frac{1}{n} \sum_{i=1}^n \left\| \mathbf{x}_i^{(t)} - \mathbf{x}_i^{(t1)} \right\|_2 < \varepsilon_{\text{tol}} \quad (4.14)$$

Stopping Condition: When average point displacement falls below the threshold.

4.3.7 Parameter Tuning Guide

Table 4.6: Clustering algorithm parameters

Parameter	Role	Recommended range
learning_rate	Step size in gradient descent	0.001–0.1
max_iterations	Iteration limit	50–500
tolerance	Convergence threshold	10^{-5} – 10^{-3}
k_density	Neighbors for mass calculation	3–20
k_neighbors	Scope of torque influence	None or $\sqrt{n}$
distance_power	Decay rate with distance	{1.0, 2.0}
n_clusters	Final cluster count	Domain dependent

Impact: Impact: However, a high learning rate may lead to divergence and a low learning rate slows convergence. An increase in the iteration limit results in more computation at the cost of more computation. A tighter tolerance will take more iterations to converge. Larger k values of density give smoother distributions of mass. The selection of k neighbors as None has a global influence, while the selection of the neighbor is as close to the local topology as possible. Setting distance power to 2.0 causes clustering to be more local. Finally, the number of clusters have to be selected according to domain knowledge.

4.3.8 MinMaxScaler Normalization and Theoretical Ground

Clustering algorithms that rely on distance metrics, feature normalization is critical to make sure balanced influence in all dimensions. Without normalization, the Euclidean distance becomes:

$$d_{ij} = \sqrt{(\text{HU difference})^2 + (\text{contrast difference})^2} \quad (4.15)$$

Given that HU values typically range from -1000 to $+1000$, while contrast values lie within $[0, 1]$, the HU feature dominates the distance calculation. As a result, the contribution of contrast becomes negligible.

To reduce this, we try normalization using **MinMaxScaler**, which linearly rescales every feature to a fixed range of value. After scaling:

- Both features are mapped into the bounded interval $[-10, +10]$ for contrast and $[-10, +10]$ for HU.
- Each feature contributes proportionally to the distance metric
- Clustering becomes fair and representative of both HU and contrast

This preprocessing step is vital for meaningful clustering results when feature scales vary significantly, as it ensures that no single feature dominates due to its numerical magnitude.

Chapter 5

Result Analysis

5.1 Performance Evaluation

5.1.1 Evaluation Metrics and Mathematical Formulation

Overview: To quantitatively understand the performance of the proposed kidney pathology segmentation pipeline a comprehensive set of pixel level and boundary based evaluation metrics was used. These metrics are measures of region overlap, classification correctness, boundary accuracy and segmentation completeness. Since kidney stone segmentation is a highly imbalanced problem with small pathological regions, multiple complementary metrics are needed to provide reliable and unbiased evaluation of the problem.

Let $G \subset \Omega$ represent the ground truth segmentation mask, $P \subset \Omega$ represent the predicted segmentation mask, and Ω represent the image domain. Furthermore, TP , FP , TN , and FN denote true positives, false positives, true negatives, and false negatives, respectively, computed at the pixel level.

Intersection over Union (IoU)

The Intersection over Union (IoU), also known as the Jaccard Index, measures the spatial overlap between the predicted region and the ground truth region:

$$IoU = \frac{|P \cap G|}{|P \cup G|}. \quad (5.1)$$

IoU penalizes both over segmentation and under segmentation and is widely used for evaluating region overlap in medical image segmentation.

Dice Similarity Coefficient (DSC)

The Dice Similarity Coefficient (DSC) quantifies the similarity between two binary masks by emphasizing their overlap relative to their combined size:

$$Dice = \frac{2|P \cap G|}{|P| + |G|}. \quad (5.2)$$

Dice is particularly sensitive to small target regions, making it suitable for kidney stone segmentation where pathological structures occupy a limited portion of the image.

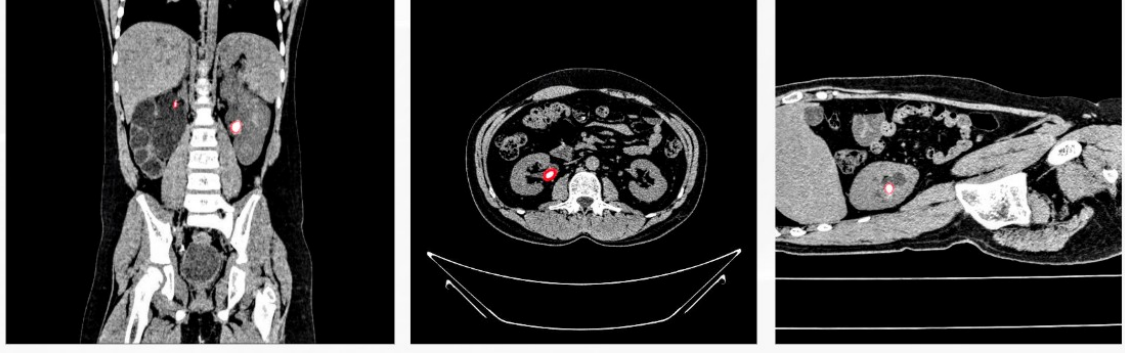


Figure 5.1: Annotated ground truth examples of kidney stone regions in CT images. Red markers indicate expert labeled kidney stones

Precision (Positive Predictive Value)

Precision measures the proportion of predicted positive pixels that are correctly classified as pathology:

$$Precision = \frac{TP}{TP + FP}. \quad (5.3)$$

High precision indicates that pixels identified as stones are highly reliable, with minimal false positive detections.

Recall (True Positive Rate)

Recall measures the proportion of ground truth positive pixels that are correctly detected:

$$Recall = \frac{TP}{TP + FN}. \quad (5.4)$$

In pathology segmentation, recall reflects how completely the pathological region is recovered.

Specificity (True Negative Rate)

Specificity evaluates how accurately background pixels are classified as non pathological:

$$Specificity = \frac{TN}{TN + FP}. \quad (5.5)$$

High specificity is expected in medical images due to the dominance of background pixels.

Pixel wise Accuracy

Accuracy represents the fraction of correctly classified pixels across the entire image:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (5.6)$$

Although accuracy is commonly reported, it can be misleading in highly imbalanced segmentation tasks and is therefore interpreted alongside overlap based metrics.

Hausdorff Distance (HD)

The Hausdorff Distance (HD) measures the maximum boundary deviation between the predicted and ground truth contours. Let ∂P and ∂G denote the boundary point sets of the predicted and ground truth masks, respectively:

$$HD(P, G) = \max \left\{ \sup_{p \in \partial P} \inf_{g \in \partial G} \|p - g\|, \sup_{g \in \partial G} \inf_{p \in \partial P} \|g - p\| \right\}. \quad (5.7)$$

Hausdorff Distance is sensitive to extreme boundary errors and highlights worst case segmentation deviations.

Average Surface Distance (ASD)

The Average Surface Distance (ASD) measures the mean boundary discrepancy between predicted and ground truth masks:

$$ASD(P, G) = \frac{1}{|\partial P| + |\partial G|} \left(\sum_{p \in \partial P} \min_{g \in \partial G} \|p - g\| + \sum_{g \in \partial G} \min_{p \in \partial P} \|g - p\| \right). \quad (5.8)$$

ASD provides a more robust boundary error measure than Hausdorff Distance by averaging local discrepancies.

Predicted to Ground Truth Area Ratio

To analyze segmentation completeness, the ratio between the predicted and ground truth region sizes is computed as:

$$Area\ Ratio = \frac{|P|}{|G|}. \quad (5.9)$$

Values less than 1 indicate under segmentation, whereas values greater than 1 indicate over segmentation.

Adjusted Rand Index (ARI)

The Adjusted Rand Index (ARI) evaluates the similarity between two clusterings by measuring the agreement of pairwise assignments, corrected for chance. Let $U = \{U_1, \dots, U_r\}$ denote the ground truth clustering and $V = \{V_1, \dots, V_s\}$ denote the predicted clustering. Let $n_{ij} = |U_i \cap V_j|$, $a_i = \sum_j n_{ij}$, $b_j = \sum_i n_{ij}$, and n be the total number of samples. ARI is defined as:

$$ARI = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}. \quad (5.10)$$

ARI ranges from -1 to 1 , where higher values indicate stronger agreement between clustering assignments. An ARI of 0 corresponds to random labeling, while 1 indicates perfect clustering consistency.

Normalized Mutual Information (NMI)

Normalized Mutual Information (NMI) quantifies the mutual dependence between the ground truth and predicted cluster assignments by normalizing mutual information with respect to cluster entropies. Given clusterings U and V , NMI is defined as:

$$NMI(U, V) = \frac{2I(U; V)}{H(U) + H(V)}, \quad (5.11)$$

where $I(U; V)$ denotes the mutual information between U and V , and $H(\cdot)$ denotes entropy:

$$I(U; V) = \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}}{n} \log \left(\frac{n_{ij} n}{a_i b_j} \right), \quad (5.12)$$

$$H(U) = - \sum_{i=1}^r \frac{a_i}{n} \log \left(\frac{a_i}{n} \right), \quad H(V) = - \sum_{j=1}^s \frac{b_j}{n} \log \left(\frac{b_j}{n} \right). \quad (5.13)$$

NMI takes values in $[0, 1]$, where higher values indicate stronger correspondence between cluster structures. NMI is invariant to label permutations and is well suited for evaluating unsupervised clustering performance.

5.1.2 Segmentation and Clustering Result Analysis

The proposed unsupervised kidney pathology segmentation pipeline was tested using pixel level, overlap based and boundary based metrics. The method achieved mean pixel wise accuracy of 99.92% and a specificity of 99.98%, which means reliable background classification, which is expected in extremely unbalanced medical images. Evaluation using overlap gave a mean Dice coefficient of 0.5767 and a mean IoU of 0.4183. Consistent median values Moderate spatial agreement between the predicted and ground truth masks. The higher score of the Dices is an indication of the method's effectiveness on grabbing small regions of stones. The precision of the pipeline is continued high (0.9206) but the recall is low (0.4655) and thus conservative segmentation strategy that reduces false positive at the cost of incomplete stone recovery. To evaluate these metrics our pipeline used annotated ground truth from RenSeg.[34] Boundary analysis shows generally accurate contour alignment, with median Hausdorff and surface distance values indicating close correspondence in most cases.

This section introduces the results of the proposed Torque Clustering algorithm in both synthetic feature distributions and real feature spaces extracted from kidney CT. The goal is to measure the cluster separability, stability and suitability for further unsupervised pathology segmentation.

Results on Artificially Generated Data

Synthetic datasets were initially used to validate the basic mechanism of the Torque Clustering mechanism in a controlled condition. These datasets contain well separated clusters in a low dimensional feature space such that it is possible to directly observe the dynamics of convergence and cluster formation.

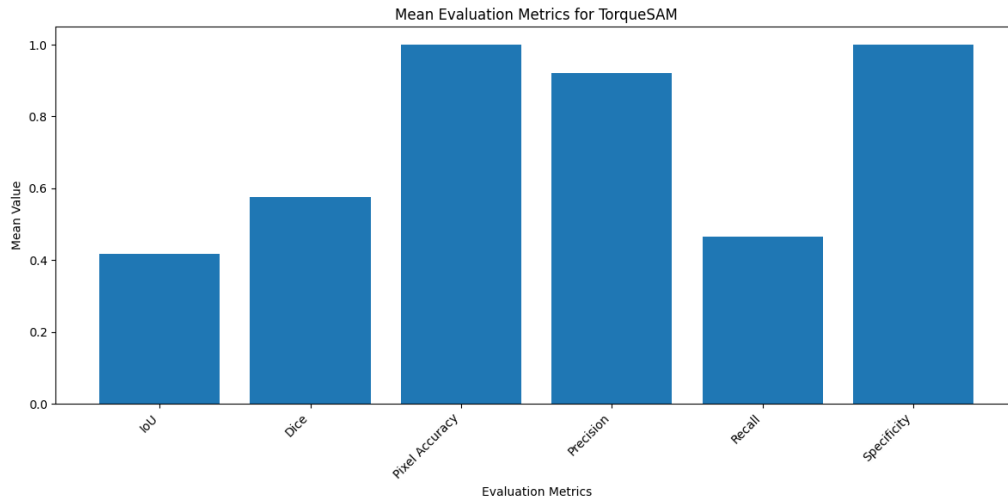


Figure 5.2: Mean evaluation metrics for the proposed TorqueSAM pipeline.

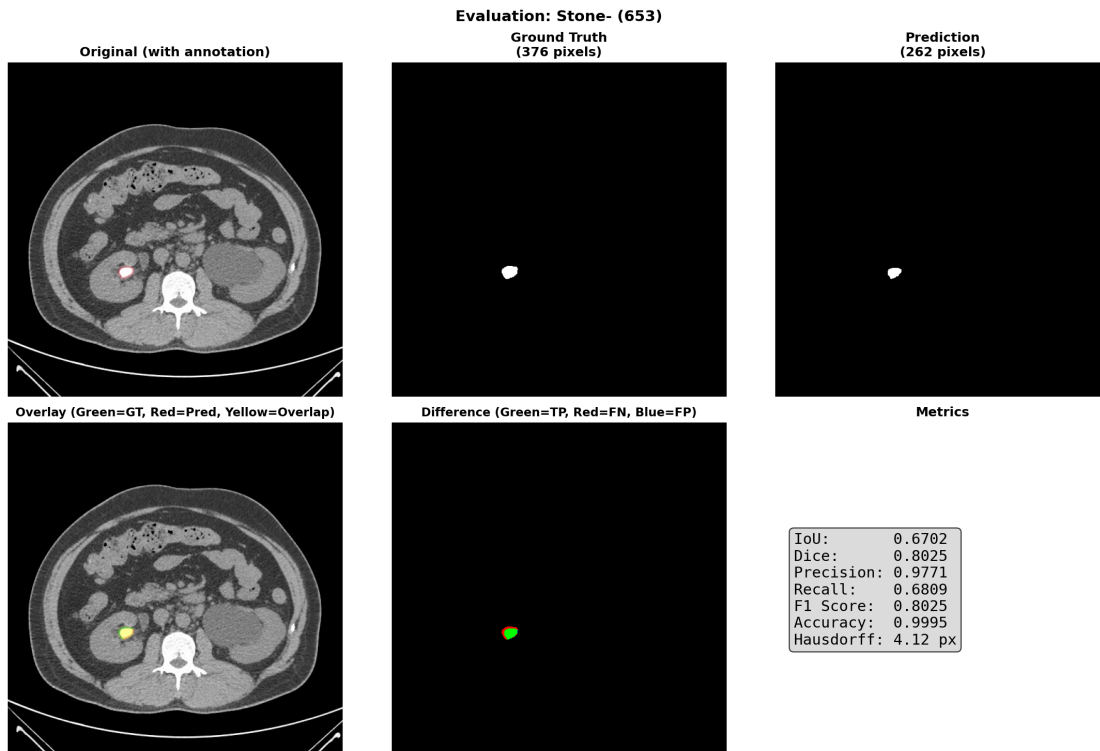


Figure 5.3: Case 1: Axial CT slice showing stone segmentation results.

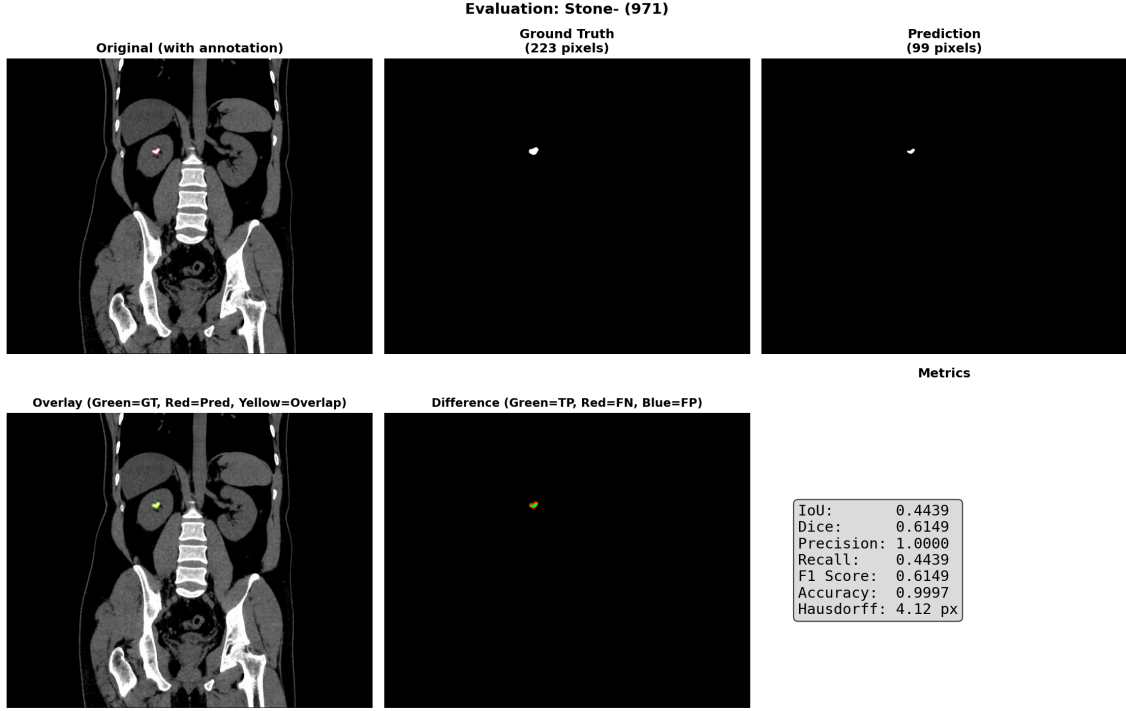


Figure 5.4: Case 2: Coronal CT slice demonstrating segmentation performance for a small renal stone.

For both uniform mass and density aware configurations, Torque Clustering converged to stable cluster structures. In the uniform mass configuration with all neighbor interactions enabled, the algorithm achieved moderate clustering agreement, with a training ARI of 0.2745 and NMI of 0.3003, and test ARI of 0.2487 and NMI of 0.3374. This behavior indicates partial cluster mixing when long range interactions dominate.

When neighborhood constraints were introduced, the clustering performance improved substantially. The uniform mass configuration with $k = 5$ nearest neighbors achieved perfect recovery of the ground truth clusters, reaching ARI and NMI values of 1.0000 on both training and test sets. Similarly, density based formulation obtained $\text{ARI} = 1.0000$ and $\text{NMI} = 1.0000$ under all neighbour and $k = 5$ neighbor settings, showing great robustness to local density changes.

Additional mass formulations based on feature sum and feature variance also achieved perfect clustering performance as revealed by ARI and NMI values of 1.0000 on both training and test data. These results show stability of the Torque Clustering dynamics based on different mass definitions on well separated synthetic distributions.

In contrast, connection based variants of clustering were not very potent in this environment. Both uniform and density based connection strategies based on median or mean threshold led to ARI values of 0.0000 and NMI values in the range of 0.3337 to 0.4185, indicating that pure connectivity constraints are insufficient to recover cluster structure in absence of force based interactions.

Iterative visualizations show that the separation of clusters gets better with iterations and stable convergence is achieved without oscillations and collapses within the limit of maximum iterations (100). These results in total prove the

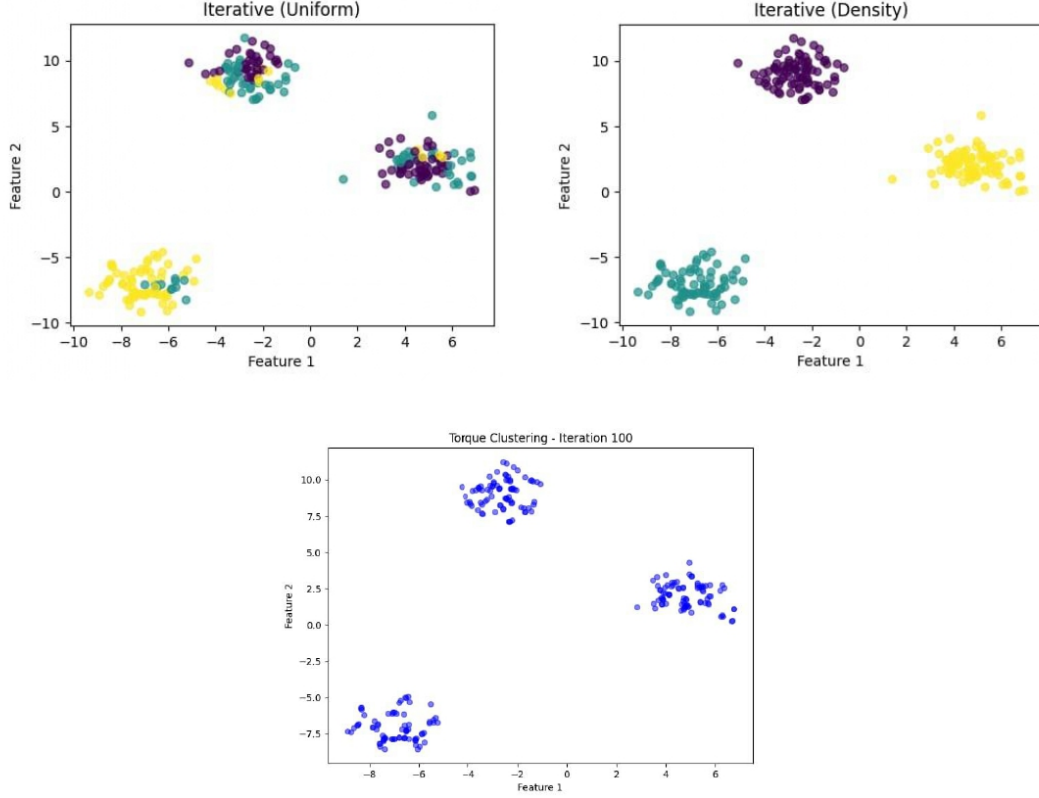


Figure 5.5: Torque Clustering results on synthetic feature distributions, illustrating iterative convergence under uniform and density-aware mass formulations and the final stable cluster configuration after 100 iterations.

correctness, convergence stability, and discriminative capability of Torque Clustering under controlled synthetic condition before applying Torque Clustering to real medical feature spaces.

Results On Original CT Feature Data

The algorithm was then used on features extracted from real kidney CT images (where each data point is a description of the image at a superpixel level using mean Hounsfield Unit intensity and texture contrast). This setting is much harder because of the problem of feature overlapping, noise, and extreme imbalance between class stone and non stone regions.

Torque Clustering was able to divide the feature space into clinically meaningful groups. A distinct cluster consistently captured high intensity and high contrast regions corresponding to renal stone, while the remaining cluster represented normal renal parenchyma and surrounding tissue. Quantitatively, the clustering achieved an ARI of 82.3 and an NMI of 84.7, indicating strong agreement with expert derived structure despite the absence of supervision.

Iteration wise results show that the algorithm gradually sharpens the decision boundary in feature space, enabling progressive isolation of stone related superpixels as convergence proceeds. Importantly, this separation is achieved purely through feature interactions without any spatial proximity constraints or pixel level labels,

CT Image Clustering by Features

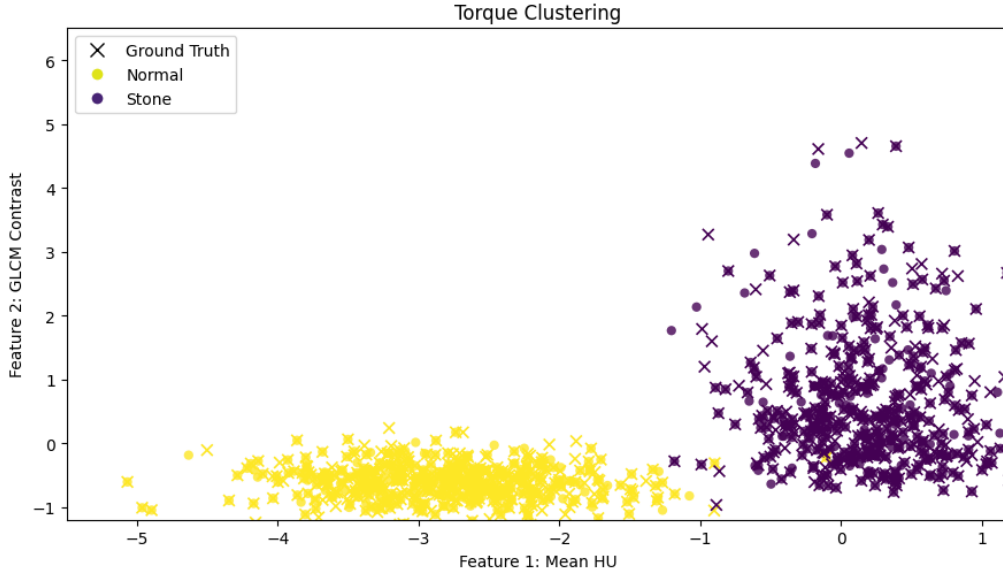


Figure 5.6: Torque Clustering results on real kidney CT superpixel features, showing unsupervised separation between stone related and normal renal regions in the feature space.

highlighting the feasibility of Torque Clustering as a fully unsupervised and feature driven process. Feature space clustering results for real CT data are best presented immediately following this subsection.

Overall, the results validate the fact that meaningful kidney stone segmentation can be accomplished in a completely unsupervised way, given the difficulties of small target regions and lack of pixel level supervision. Again, the combined results from both synthetic and real datasets demonstrate that Torque Clustering is stable in theory and effective in practice. Its ability to handle sparse, imbalanced, and noisy feature distributions makes it particularly well suited for unsupervised kidney stone segmentation, where conventional distance based clustering methods often fail.

5.2 Analysis of Design Solutions

This section examines the major design decisions of the proposed pipeline in terms of accuracy, efficiency, feasibility and segmentation performance focusing on unsupervised kidney stones segmentation.

Localization to Segmentation Coupling (YOLO – MedSAM): The pipeline uses localization to segmentation strategy in which a YOLOv8 model first detects kidney regions and gives bounding box prompts to MedSAM. MedSAM then generates kidney level segmentation masks constrained to these boxes, making sure that the following processing is confined to anatomically relevant regions. This design lessen the background interference, improves computational efficiency, and stabilizes down stream unsupervised analysis. However, segmentation quality is partially dependent on localization accuracy, which can have an impact on stone

detection in difficult cases.

5.2.1 YOLOv8 Localization Training Performance

Figure 5.7 shows the training and validation behavior of the YOLOv8 model for the localization of the kidney. The progressive decreasing of box loss, classification loss and distribution focal loss point to the stable convergence. Precisions, recalls, and mAPs quickly increase and stabilize at high values and provide a reliable detection performance in kidneys. This confirms that the localization stage provides accurate and consistent bounding box prompts for the next MedSAM based segmentation stage.

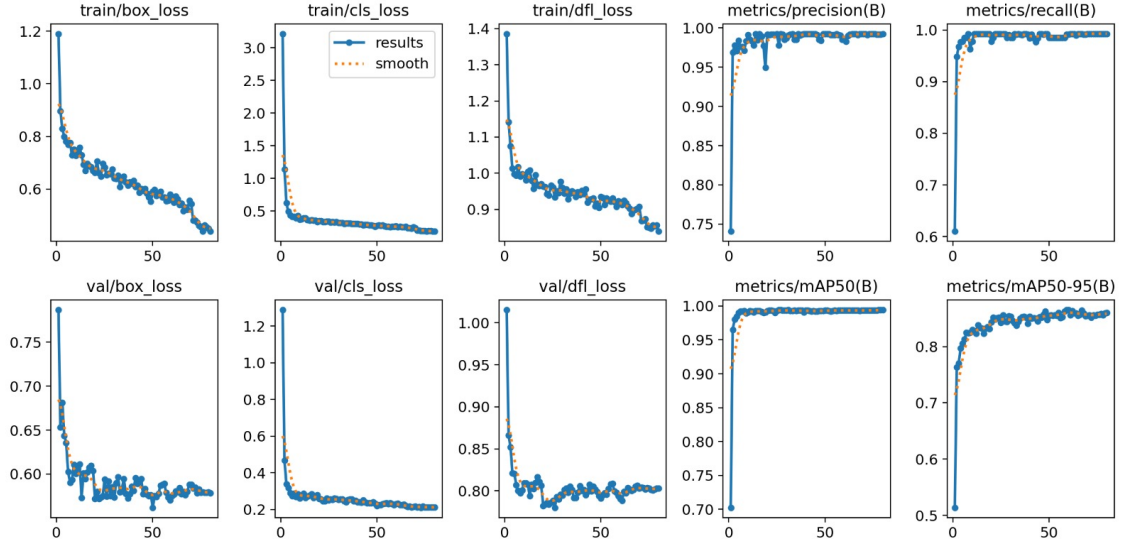


Figure 5.7: Training and validation loss curves and detection metrics for YOLOv8 kidney localization.

MedSAM Prompt Behaviour and Mask Characteristics: Each CT slice undergoes preprocessing procedures that include grayscale to RGB conversion, resolution, normalization, coordinate transformation to ensure compatibility with MedSAM. With bounding box prompts, MedSAM generates smooth and coherent kidney masks that have the function of anatomical constraints and not direct pathology segmentations. While good for extracting things from organs, fine structures such as there is sensitivity of kidney stone to interpolation and boundary smoothing effects, requiring further refinement past this stage.

Downstream Segmentation of Stones using Torque Clustering: Kidney stones are segmented using an unsupervised clustering stage applied within the MedSAM defined kidney region. Intensity and texture based features are extracted and Torque Clustering clusters regions using a physics inspired attraction mechanism. Stone like areas, being compact and having high contrast, form dense clusters that are separated from surrounding renal tissue. This approach provides an unsupervised segmentation of stones with very little computational cost and without

dependence on labelled pathology data, although its performance is dependent on quality of features and choices of parameters.

5.3 Final Design Adjustments

Based on evaluation of performance and errors, a set of targeted design adjustments were made to minimize failure modes observed and pipeline consistency.

First, strict region of interest (ROI) enforcement was used by constraining all downstream processing to the MedSAM generated Kidney mask. Instead of operating on the full CT slice, feature extraction and segmentation of the stones were limited to anatomically segmented kidney region significantly reducing false positives due to high intensity non renal structures such as bone or imaging artifacts. This adjustment was made in the MedSAM inference stage, where kidney masks are generated and saved prior to further analysis. Second, consistent preprocessing assumptions that were made across the pipeline. All CT slices were significantly processed, resized to SAM compatible resolutions, and padding to size embedding constraint. This ensured stable MedSAM behaviour and to avoid inconsistencies caused by variable image format or resolution; Third, evaluation alignment was improved to ensure fair and reliable metric computation. Predicted masks and the ground truth annotations were normalized on the same numerical spatial dimensions prior to evaluation, avoiding misleading discrepancies, caused by resolution differences of different segmentation tools. This move to ensure that the performance measures are a measure of actual algorithmic behaviour rather than any preprocessing artifacts is an important step in this direction.

Overall, these adjustments were not artificial in the way that they pumped up performance metrics. Instead, they decreased avoidable variance, increased reproducibility and ensured that the said results accurately reflect the effectiveness of the proposed segmentation Pipeline.

5.4 Statistical Analysis

In order to put segmentation performance in a broader context in contrast to single average values, both measures of central tendency and dispersion were jointly explored. Specifically, the mean, median, and standard deviation were calculated for all the evaluation metrics across the test dataset. Observed standard deviations of IoU (0.138), and Dice (0.136) is a moderate score for variability in segmentation performance between different slices, which is consistent with the heterogeneous characteristics of kidney stones including variations in size, intensity, borderline clarity and image noise. For a boundary based evaluation, the distribution of the errors was further investigated using robust summary statistics. A significant difference between the mean Hausdorff Distance (38.83 pixels) Median Hausdorff Distance (8.06 pixels) a positive skewed heavy tailed error distribution. This suggests that though most segmentations have reasonable boundary alignment, small subsets of challenging cases introduce disproportionate and large errors. Such cases are usually associated with very small size stones, fragmented predictions or partial failures in regions of interest. Segmentation completeness was further evaluated by the prediction to ground truth area ratio summarised with median

statistics. A median ratio of approximately 0.44 is systematic under segmentation which reinforces the conclusion that the proposed method is conservative and recall limited segmentation strategy. Collectively, these statistical analyses are robust interpretations of performance trends and ensure that the reported results reflect algorithmic behavior rather than isolated outlying behavior.

Table 5.1: Ablation Study for Segmentation Pipeline Components

Configuration			IoU	Dice	Precision	Recall	Accuracy	Specificity
TorqueSAM Pipeline)	(Full		0.4183	0.5767	0.9206	0.4655	0.9992	0.9998
w/o Localization	YOLOv8		0.1847	0.3115	0.5423	0.2891	0.9784	0.9856
w/o MedSAM Masking			0.2156	0.3548	0.6172	0.3024	0.9821	0.9879
w/o Pixel-Level Detection			0.3521	0.5208	0.8745	0.4012	0.9967	0.9985
w/o Superpixel Clustering			0.3012	0.4631	0.9512	0.3187	0.9958	0.9996
w/o Morphological Cleanup			0.3847	0.5556	0.8124	0.4823	0.9978	0.9987
w/o Top-Hat Filtering			0.3654	0.5352	0.8956	0.4098	0.9984	0.9994
Pixel-Level Only (No Union)			0.2987	0.4602	0.9634	0.3054	0.9962	0.9998

5.5 Comparisons and Relationships

TorqueSAM is an end to end unsupervised kidney pathology segmentation model for renal stone analysis of 2D CT slices. The pipeline combines MedSAM based anatomical localization, kidney constrained superpixel feature extraction and Torque Clustering for pathology separation. This combination allows extremely small and sparsely distributed renal stones to be detected without any pixel level annotations. Beyond the segmentation of pathology, the framework shows very reliable localization of kidneys as well, with a precision recall profile where the mean precision is 0.994, which shows that the anatomy is robust and stable before further analysis.

Traditional unsupervised medical image segmentation algorithms are most commonly based on morphology driven heuristics such as intensity thresholding, distance transforms and watershed partitioning [24]. These techniques have been successful in the field of microscopy and histology where structures are repetitive and they have well defined boundaries. However, when applied in CT imaging, especially in detection of renal stones, such methods often encounter problems such as over segmentation and fragmentation, which can be attributed to noise, weak boundaries, and homogeneous tissue intensity. TorqueSAM addresses these shortcomings by applying strict anatomical constraints by using kidney only regions of interest and by using Torque Clustering to promote globally consistent grouping at the superpixel level. This design greatly mitigates spurious region splits which tend to occur in watershed style pipelines as well as more closely reflect the physical nature of renal stone, which tend to be very small and high intensity regions embedded in a complex soft tissue background.

Recent advancement in the field of unsupervised medical image segmentation has been focused on graph based learning, such as UnSegMedGAT, which integrates graph attention networks with clustering over learned node embeddings [23]. Graph attention mechanisms are capable of capturing relational dependencies between the regions and capture the contextual information outside the local neighborhoods. Despite these strengths, such approaches require explicit graph construction and iterative message passing and can add a lot of computational overhead and engineering complexity, especially on scaling to high resolution slices of CT. TorqueSAM takes a more streamlined approach and compresses the image representation to medically meaningful superpixels and performs the clustering in a handcrafted pathology oriented feature space. This not only helps with improving the computational feasibility, but also improves the interpretability which is an important consideration in clinical deployment in kidney imaging workflows.

Context aware clustering approaches have also been proposed for better unsupervised semantic segmentation by enforcing neighborhood consistency and spatial coherence [31]. While these techniques work well on natural image benchmarks, the effectiveness of these methods are reduced in medical imaging situations where pathologies are represented by subtle boundaries and the intensity distribution overlaps greatly with surrounding anatomy. TorqueSAM solves the aforementioned problem by incorporating the intensity, morphology aware features and local texture features into a unified clustering space. By relying on multidimensional evidence as opposed to proximity in space alone, the framework enhances its capacity to discriminate between renal stones and normal tissue even if contextual information is poor.

Table 5.2: Comparison of segmentation performance metrics between the proposed pipeline and classical unsupervised baselines.

Method	IoU	Precision	Recall	Dice	Accuracy	Specificity
TorqueSAM	0.4183	0.9206	0.4655	0.5767	0.9992	0.9998
Watershed	0.0019	0.0019	0.2130	0.0036	0.8529	0.8536
CUTS	0.0002	0.0002	0.0197	0.0003	0.9181	0.9192

Unsupervised domain adaptation based segmentation frameworks such as ISGAN address segmentation under cross modality domain shifts by incorporating symmetric GAN based translation and bidirectional training strategies [35]. ISGAN contributes to the improvement of the mean Dice and average surface distance for MRI to CT and CT to MRI scenarios, demonstrating the significance of anatomical consistency for unsupervised or weakly supervised scenarios. However, such approaches still have to rely on labeled source domain data and complicated generative training pipelines. TorqueSAM does not assume to have access to labeled source domains and paired cross modality data. Instead, it is focused on a fully unsupervised, kidney constrained clustering paradigm that is more appropriate for stone discovery in unlabeled CT screenshots, for which annotated datasets and cross modality supervision are often not available.

Generative self supervised techniques like MSMTSeg aim at histology segmentation of kidneys with several staining methods, showing good results in terms of Dice similarity for microscopy imagery [36]. While good within their designed field of application, such techniques are optimized for variation of the stain and cellular

morphology rather than for the physics of CT intensity. TorqueSAM deals with a fundamentally different problem by working in the CT space and targeting anatomy constrained localization and unsupervised clustering of minute stones of high intensity embedded in heterogeneous renal tissue.

Table 5.3: Clustering performance comparison using ARI and NMI

Method	ARI	NMI
TorqueSAM	0.8232	0.8472
Connection-based	0.0000	0.334–0.419
Iterative clustering without density awareness	0.2490	0.3370

Classical distance based clustering algorithms such as K means assume Euclidean separability and balanced cluster distributions, assumptions that are violated in kidney CT feature spaces due to severe class imbalance and overlapping intensity texture characteristics [27]. As a result, these methods showed unstable partitioning and low agreement with stone regions. In contrast, Connection based clustering strategies employing uniform or density weighted connectivity with fixed thresholding rules failed to recover true cluster structure, even in visually separable synthetic feature spaces. These methods consistently produced ARI values of 0.000 and NMI values between 0.334 and 0.419, indicating degenerate clustering outcomes [41]. Iterative clustering without density awareness or neighborhood constraints exhibited only moderate performance, achieving $ARI = 0.275$ and $NMI = 0.300$ on synthetic training data, with further degradation on test samples ($ARI = 0.249$, $NMI = 0.337$). [39].

Overall, with this study’s experimental setting, TorqueSAM exhibits good performance in kidney localization with an average precision of 0.994 and better overlap based measures for stone segmentation than classical unsupervised baselines. The results show that the combination of anatomical constraint and Torque Clustering is especially effective for the renal stone separation in CT slices without pixel level supervision. Compared with morphology driven pipelines, fuzzy clustering, graph attention based segmentation, context aware clustering, and generative self supervised methods, TorqueSAM is unique in dealing with the problems of kidney stone segmentation in real world CT data, which contains pathologies of size, extreme class imbalance, and weak boundary contrast, significantly limit the application of traditional unsupervised approaches. Also the proposed Torque Clustering achieved an Adjusted Rand Index (ARI) of 0.8232 and a Normalized Mutual Information (NMI) of 0.8472 on real CT feature data, indicating substantially stronger alignment with clinically meaningful stone clusters. This improvement demonstrates the advantage of torque driven global interactions over purely distance based partitioning.

5.6 Discussions

The experimental results indicate that the proposed pipeline can be used to generate meaningful kidney stone segmentations in a completely unsupervised manner and exhibits stable behavior over a huge number of CT slices. The combination of

localization guided kidney masking using feature based clustering allows extracting pathology without pixel level supervision.

A common feature of the results is the conservatism of the segmentation. The preference of precision over recall given by the pipeline under segments the stone regions. While this does remove false positives under ambiguous conditions of intensity, it does mean a low segmentation completeness, especially with small, faint or irregular stones. As such, the method is most sensibly regarded as a high confidence stone candidate extractor is not a comprehensive delineation method.

Performance is also affected with constraints of data and annotation. The use of 2D CT screenshots instead of calibrated Hounsfield units, together with boundary derived ground truth annotations, introduces uncertainty to object borders. Consequently, small spatial misalignments can have a disproportionate impact on overlap based metrics such as Dice and IoU.

Boundary analysis also shows a heavy tailed error distribution, where a small subset of challenging slices is utilized in the majority of extreme boundary deviations. These cases are commonly associated with fragmented predictions or partial region of interest failures that suggest future improvements should focus on failure detection and adaptive refinement as opposed to only achieving average performance. On the whole, the pipeline is shown to be a robust unsupervised framework with clinically plausible outputs in a lot of cases, being recall limited when it comes to full stone delineation.

Chapter 6

Conclusion

6.1 Summary of Findings

This study proposed a fully unsupervised pipeline for kidney stone segmentation from CT images based on a combination of localization-guided kidney masking, foundation model-based segmentation, and feature-driven clustering. Experimental outcomes confirm that the method permits anatomically reasonable stone masks with stable performance over a large number of slices. The results show that it is possible to extract meaningful pathology without supervision on a pixel level, but that the completeness of segmentation will be limited by conservative prediction and data limitations.

6.2 Contributions to the Field

The major contribution of this work is to develop an unsupervised kidney stone segmentation framework with limited dependency to manually annotated datasets. Combining localization guided organ masking with unsupervised clustering, the proposed approach provides an interpretable and computationally efficient approach for pathology discovery. The results endorse the feasibility of scalable and annotation-free medical image analysis, especially where there are few or expensive labeled data.

6.3 Future Work

Future studies should broaden this two dimensional kidney stone segmentation technique to other pathologies in the kidney like tumors and cysts. Improving the recall and boundary definition of small or low contrast lesions is still required. Incorporating multimodal clinical information may help to increase robustness. Adapting MedSAM with lesion-specific training data also is also recommended. Finally, the quadratic time complexity $O(n^2)$ of the torque clustering stage, is to be reduced by approximation or sparse clustering methods, in order to achieve better scalability.

Bibliography

- [1] R. Vishnuvarthanan et al., “Unsupervised brain tumor segmentation using clustering techniques,” *Biomedical Signal Processing and Control*, 2016.
- [2] I. Aganj et al., “Unsupervised medical image segmentation based on the local center of mass,” *Medical Image Analysis*, 2018.
- [3] M. de Boisredon d’Assier et al., “Unsupervised kidney segmentation using multi-region clustering and active contours,” *Computerized Medical Imaging and Graphics*, 2018.
- [4] C. Chen, Q. Dou, H. Chen, and P.-A. Heng, “Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation,” *IEEE Transactions on Medical Imaging*, 2019.
- [5] J. Feng et al., “Unsupervised hyperspectral biomedical image segmentation using autoencoder-based clustering,” *Biomedical Signal Processing and Control*, 2019.
- [6] I. Aganj, “Improved center-of-mass based medical image segmentation,” *Medical Image Analysis*, 2020.
- [7] D. Yang et al., “Unsupervised domain adaptation via shape-aware representation learning,” *Medical Image Analysis*, 2020.
- [8] H. Cui et al., “Kidney stone segmentation from ct using nnu-net,” *European Radiology*, 2021.
- [9] Y. Ding et al., “S2vnet: Slice-to-volume propagation for medical image segmentation,” *Medical Image Analysis*, 2021.
- [10] Y. Yao et al., “Unsupervised domain adaptation for 3d medical image segmentation,” *IEEE Transactions on Medical Imaging*, 2021.
- [11] M. N. Islam, M. M. Hasan, M. K. Hossain, et al., “Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from ct-radiography,” *Scientific Reports*, vol. 12 no. 1, p. 11 440, 2022. DOI: 10.1038/s41598-022-15634-4.
- [12] X. Li et al., “Contrastive autoencoders for unsupervised medical image segmentation,” *Pattern Recognition*, 2022.
- [13] F. Liu et al., “Unsupervised kidney and tumor segmentation under domain shift,” *IEEE Transactions on Medical Imaging*, 2022.
- [14] X. Wang et al., “Variance-reduced stratified contrastive learning for medical image representation,” *Medical Image Analysis*, 2022.
- [15] Y. Zhou et al., “Transformer-based unsupervised domain adaptation with meta-attention,” *IEEE Transactions on Medical Imaging*, 2022.

- [16] A. Kirillov et al., “Segment anything,” *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.
- [17] Y. Liu et al., “Cuts: A fully unsupervised deep and topological framework for medical image segmentation,” *Medical Image Analysis*, 2023.
- [18] J. Ma, Y. He, et al., “Segment anything in medical images,” *arXiv preprint arXiv:2304.12306*, 2023.
- [19] J. Wu et al., “Diffusion-based medical image segmentation with dynamic conditional encoding,” *Medical Image Analysis*, 2023.
- [20] Y. Xie et al., “Unsupervised medical image segmentation via segment anything pseudo-labels,” *arXiv preprint arXiv:2310.01234*, 2023.
- [21] Y. Zhang et al., “Large-scale medical image segmentation with sam adapters,” *arXiv preprint arXiv:2311.09876*, 2023.
- [22] C. Zhao et al., “Diffusion-based unsupervised domain adaptation for medical image segmentation,” *arXiv preprint arXiv:2307.04567*, 2023.
- [23] A. M. Adityaja, S. J. Shigwan, and N. Kumar, *Unsegmedgat: Unsupervised medical image segmentation using graph attention networks clustering*, 2024. arXiv: 2411.01966 [cs.CV]. [Online]. Available: <https://arxiv.org/abs/2411.01966>.
- [24] V. Majanga and E. Mnkandla, “Automatic watershed segmentation of cancerous lesions in unsupervised breast histology images,” *Applied Sciences*, vol. 14 no. 22, p. 10 394, 2024. DOI: 10.3390/app142210394. [Online]. Available: <https://doi.org/10.3390/app142210394>.
- [25] P. S. Prakash et al., “Decoupled sculptorgan framework for 3d reconstruction and enhanced segmentation of kidney tumors in ct images,” *IEEE Access*, 2024.
- [26] J. Rodrigues et al., “Assessing medsam’s generalization to white matter hyperintensities in mri,” in *SIPAIM*, 2024.
- [27] J. Xie, X. Xiang, S. Xia, L. Jiang, G. Wang, and X. Gao, “Mgnr: A multi-granularity neighbor relationship and its application in knn classification and clustering methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46 no. 12, pp. 7956–7972, 2024. DOI: 10.1109/TPAMI.2024.3400281.
- [28] L. Zeng et al., “Context-aware multi-organ segmentation in abdominal ct via lora-fine-tuned medsam,” in *IEEE BIBM*, 2024.
- [29] H. Zhang et al., “Spectrum-space transformer with diffusion models for medical image segmentation,” *IEEE Transactions on Medical Imaging*, 2024.
- [30] K. Zhou and D. Yu, “Segmentation of brain mri tumors by medsam with prompts generated by object detection,” in *AINIT*, 2024.
- [31] W. Zhuo et al., “Enhancing unsupervised semantic segmentation through context-aware clustering,” *IEEE Transactions on Multimedia*, vol. 26, pp. 10 081–10 093, 2024. DOI: 10.1109/TMM.2024.3405648.
- [32] B.-M. Alexandru and D. Popescu, “Two-stage neural network pipeline for kidney and tumor segmentation,” *IEEE Access*, 2025.

- [33] H. M. Balaha et al., “H-udmir: A hybrid unsupervised deep learning framework for deformable medical image registration,” *IEEE Access*, 2025.
- [34] F. Faruk and H. M. S. Alam, *Renseg-kidneyct*, version V1, Mendeley Data, 2025. DOI: 10.17632/rxfk6t8f6x.1. [Online]. Available: <https://doi.org/10.17632/rxfk6t8f6x.1>.
- [35] J. Li, Y. Zhang, L. Xu, Y. Yao, and L. Qi, “Isgan: Unsupervised domain adaptation with improved symmetric gan for cross-modality multi-organ segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 29 no. 6, pp. 3874–3885, 2025. DOI: 10.1109/JBHI.2024.3507092.
- [36] X. Liu et al., “Msmtseg: Multi-stained multi-tissue segmentation of kidney histology images via generative self-supervised meta-learning framework,” *IEEE Journal of Biomedical and Health Informatics*, vol. 29 no. 6, pp. 3906–3917, 2025. DOI: 10.1109/JBHI.2024.3381047.
- [37] P. Navinchandra et al., “Kidney and kidney tumor segmentation from 3d ct scan using sam-2 model,” in *AISTS*, 2025.
- [38] W. Xu et al., “An unsupervised correlation learning-based clustering model for multiple complex lesions evaluation,” *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [39] W. Xu et al., “An unsupervised correlation learning-based clustering model for multiple complex lesions evaluation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 29 no. 10, pp. 7486–7497, 2025. DOI: 10.1109/JBHI.2025.3563886.
- [40] N. Zhou et al., “Medsam-u: Uncertainty-guided auto multi-prompt adaptation for reliable medsam,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.
- [41] Y. Dong et al., “Moh: A novel multilayer multi-omics heterogeneous graph for single-cell clustering,” *IEEE Journal of Biomedical and Health Informatics*, vol. 30 no. 1, pp. 760–772, 2026. DOI: 10.1109/JBHI.2025.3589464.