الجمهورية الجزائرية الديمقراطية الشعبية
Peoples Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة غرداية
University of Ghardaia
كلية العلوم والتكنولوجيا
Faculty of Science and Technology
قسم الرياضيات والإعلام الآلي
Department of Mathematics and Computer Science
مخبر الرياضيات والعلوم التطبيقية
Mathematics and Applied Sciences Laboratory

Registration nŭ:
...../...../...../...../.....

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master

Domain: Mathematics and Computer Science
Field: Computer Science
Specialty: Intelligent Systems for Knowledge Extraction

Topic

Attention-Powered U-Net for Kidney Tumor Segmentation

Presented by:

DOUDOU Mohamed Elhadi & ABISMAIL Bayoub

Publicly defended on 06 30, 2025

Jury members:

Mr. AHMED Saidi	MCA	Univ. Ghardaia	President
Mr. Fekair Mohamed EL Amine	MAB	Univ. Ghardaia	Examiner
Mr. OULAD NAOUI Slimane	MCB	Univ. Ghardaia	Supervisor
Ms. LAHRACHE Ferialle	PhdS	Univ. Ghardaia	Co-Supervisor

Academic Year: 2024/2025

Acknowledgment

Above all, we would like to thank Allah, the most gracious and merciful, for granting us health, faith, and strength to complete our studies and for inspiring us to undertake and complete this work.

We extend heartfelt thanks to our supervisor, Mr. OULAD NAOUI Slimane, and our co-supervisor, Ms. LAHRACHE Ferialle, for their invaluable guidance, unwavering support, and expertise that significantly enriched our research. Their insightful direction and contributions to our project were instrumental in its completion. In addition, we express our gratitude to both supervisors for their pivotal role in selecting the important thesis title. Delving into this topic was a rewarding experience, and their guidance in this aspect was crucial.

We are overwhelmed with humility and gratitude to acknowledge the depth of our appreciation to all those who have helped us take these ideas well beyond the level of simplicity and into something concrete.

Any attempt at any level cannot be satisfactorily completed without the support and guidance of my parents and friends.

We would like to thank our parents, who provided significant support in gathering information, collecting data, and offering guidance throughout the thesis process, despite their busy schedules, they contributed valuable ideas that made this work unique.

Dedication

I dedicate this thesis to the people who have been a constant source of support, strength, and inspiration throughout my academic journey.

First and foremost, to my parents, your unconditional love, encouragement, and sacrifices have been the foundation of my life. Your belief in me gave me the strength to keep moving forward even during the most challenging times. I owe everything I have achieved to your guidance and prayers.

To my family, thank you for your understanding and patience during the long days and sleepless nights that this work demanded. Your kindness, humor, and motivation were invaluable.

To my teachers and academic advisors, your mentorship and dedication to excellence have profoundly shaped my thinking and growth. Your knowledge and feedback have guided this thesis and my development as a researcher.

I also dedicate this work to the countless researchers, educators, and innovators whose efforts have paved the way in the field of medical image analysis and deep learning. Your contributions have inspired and challenged me.

Finally, to everyone who believed in me even when I doubted myself, thank you for reminding me that persistence, humility, and hard work always lead to meaningful progress.

ABISMAIL Bayoub

Dedication

This thesis is dedicated to the people who have walked beside me with love, patience, and unwavering support.

To my dear parents, thank you for being my first teachers, my strongest pillars, and my lifelong inspiration. Your sacrifices, wisdom, and endless encouragement made this journey possible. Every page of this work is a reflection of your guidance and belief in me.

To my family, who stood by me with kindness and understanding during moments of silence, stress, and struggle, your support carried me through the toughest days.

To my teachers and academic advisors, your mentorship and dedication to excellence have profoundly shaped my thinking and growth. Your knowledge and feedback have guided this thesis and my development as a researcher.

To my friends, who listened when I needed to vent, laughed with me in moments of relief, and reminded me to breathe when things got overwhelming. thank you for your presence and loyalty.

And to the memory of loved ones who are no longer with us, but whose influence and love continue to echo in my heart you are not forgotten.

This thesis is not just an academic achievement, it is a tribute to all of you. Without your presence in my life, this would have remained a distant dream.

DOUDOU Mohamed Elhadi

مـلـخـص

يُعدّ سرطان الكلى من القضايا الصحية البارزة، حيث يشهد تزايدًا ملحوظًا في معدلات التشخيص والوفيات. ففي كل عام يُشخّص الآلاف من الأشخاص بالإصابة بهذا المرض، ويفقد الكثيرون حياتهم نتيجة التأخر في اكتشافه. وعلى الرغم من أن الطرق التقليدية للتشخيص تُعدّ مفيدة، فإنها غالبًا ما تفتقر إلى الدقة الكافية، مما يخلق تحديات في تخطيط العلاج وتحسين نتائج المرضى. تُعدّ صور الأشعة المقطعية (CT) المعيار الذهبي لتشخيص المرض، إلا أن عملية تقسيم الورم يدويًا تستغرق وقتًا طويلاً، وتُعدّ عرضة للتفاوت، وتعتمد بشكل كبير على خبرة أطباء الأشعة المتخصصين. وقد أظهرت أساليب التعلم العميق، وبشكل خاص نموذج U-Net وعودًا كبيرة في أتمتة مهام التقسيم. ومع ذلك، غالبًا ما تواجه النماذج الحالية صعوبات في التعامل مع الحدود غير الواضحة للأورام، واختلال التوازن بين الفئات، وسوء تصنيف الأكياس الحميدة. في هذه الدراسة، قمنا بتطبيق نموذجنا اعتمادًا على بنية U-Net، Attention والتي تدمج آليات الانتباه داخل إطار عمل U-Net بهدف تعزيز استخراج السمات، وتحديد موقع الورم بدقة، وتحسين دقة التقسيم لأورام الكلى من صور الأشعة المقطعية. في هذا البحث، اتبعنا نهجًا مختلفًا في عملية ما قبل المعالجة لبياناتنا، وقد أثبت هذا النهج فعاليته الكبيرة في تقسيم الكلى والورم، مما أدى إلى تحسين دقة تشخيص أمراض الكلى وتخطيط العلاج بشكل أفضل. استخدمنا في تجربتنا مجموعة بيانات KiTS19 التي تحتوي على صور أشعة مقطعية محسّنة بالتباين باستخدام أسلوب التقسيم الدلالي. وقد حقق نموذجنا متوسط درجة Dice بلغ 0.85% للكلى 0.70% لأورام الكلى، مما يبرز إمكاناته في تحسين عملية اتخاذ القرار السريري المرتبطة بالأورام الكلوي.

كلمات مفتاحية: تقسيم الصور, ورم الكلي, التعلم العميق, طريقة الانتباه, شبكة U.

Abstract

Kidney cancer is a major health concern, with rising rates of diagnosis and mortality. Each year, thousands of people are diagnosed, and many lose their lives due to late detection. Traditional diagnostic methods, while valuable, often fall short in accuracy, leading to challenges in treatment planning and patient outcomes. While computed tomography (CT) imaging is the gold standard for diagnosis, manual tumor segmentation is time-consuming, prone to variability, and highly dependent on radiologists' expertise. Deep learning-based methods, particularly U-Net, have shown a great promise in automating segmentation tasks. However, existing models often struggle with ambiguous tumor boundaries, class imbalances, and misclassification of benign cysts. In this study, we implemented a U-Net Attention model architecture, which integrates attention mechanisms into a U Net framework to enhance feature extraction, tumor localization, and segmentation accuracy of kidney tumor segmentation from CT images. In the experiment, we follow a different approach in the pre-processing pipeline of our dataset. Our approach proves a powerful way to segment kidney and tumor, leading to more accurate kidney disease diagnosis and treatment planning. We utilize the KiTS19 dataset for contrast-enhanced CT images using semantic segmentation. Our model achieves a mean Dice score of 0.85% and 0.70% for kidney and kidney tumors, respectively. It showcases the potential to improve clinical kidney method decision-making.

Keywords: Image Segmentation, Kidney Tumor, Deep Learning, Attention Mechanism, U-NET.

Résumé

Le cancer du rein représente un enjeu majeur de santé publique, avec une augmentation notable des taux de diagnostic et de mortalité. Chaque année, des milliers de personnes reçoivent un diagnostic de cette maladie, et beaucoup en meurent en raison d'une détection tardive. Bien que les méthodes de diagnostic traditionnelles soient utiles, elles manquent souvent de précision, ce qui complique la planification du traitement et nuit aux résultats cliniques. L'imagerie par tomodensitométrie (CT) est considérée comme la norme de référence pour le diagnostic, mais la segmentation manuelle des tumeurs est lente, sujette à des variations inter-opérateurs et fortement dépendante de l'expertise des radiologues. Les méthodes d'apprentissage profond, en particulier le modèle U-Net, ont montré un fort potentiel dans l'automatisation des tâches de segmentation. Toutefois, les modèles existants rencontrent fréquemment des difficultés liées à des frontières tumorales floues, des déséquilibres de classes, et des erreurs de classification des kystes bénins. Dans cette étude, nous avons mis en ovre en modèle basé sur l'architecture Attention U-Net, qui intègre des mécanismes d'attention dans l'architecture U-Net afin daméliorer l'extraction des caractéristiques, la localisation de la tumeur et la précision de la segmentation des tumeurs rénales à partir dimages CT. Nous adoptons une approche différente dans le processus de prétraitement de notre ensemble de données. Cette méthode s'est révélée efficace pour la segmentation du rein et de la tumeur, permettant un diagnostic plus précis des maladies rénales et une meilleure planification thérapeutique. Notre méthode utilise le jeu de données KiTS19, composé dimages CT rehaussées par contraste et segmentées de manière sémantique. Notre modèle a atteint un score moyen de Dice de 0,85% pour le rein et de 0,70% pour la tumeur rénale, démontrant ainsi son potentiel à améliorer la prise de décision liée aux problémes cliniques rénales.

Mots clés: segmentation dimages, tumeur rénale, apprentissage profond, Mechanism d'attention , U-NET.

Contents

List of Figures

List of Tables

List of Abbreviations

AG Attention Gate

API Application Programming Interface

CNN Convolutional Neural Network

CPU Central Processing Unit

CT Computed Tomography

CV Computer Vision

DL Deep Learning

DNN Deep Neural Network

DSC Dice Similarity Coefficient

FCM Fuzzy C-Mean

FCN Fully Convolution Network

GPU Graphical Processing Units

ILSVRC Large Scale Visual Recognition Challenge

IOU Intersection Over Union

ISBI International Symposium on Biomedical Imaging

KNN K-Nearest Neighbor

MI Medical Imaging

ML Machine Learning

MRI Magnetic Resonance Imaging

NIFTI Neuroimaging Informatics Technology Initiative

NLP Natural Language Processing

RCC Renal Cell Carcinoma

RNN Recurrent Neural Network

ROI Region Of Interest

SVM Support Vector Machine

US Ultrasound Sonography

Introduction

Kidney cancer is a disease that originates in the kidneys when healthy cells in one or both organs begin to grow uncontrollably, forming a tumor. It remains a major global health concern, with over 400,000 new cases and nearly 156,000 deaths reported annually. According to the World Cancer Research Fund International[1], kidney cancer ranks as the 14th most common cancer worldwide, placing it 10th among men and 13th among women. The most prevalent subtype, renal cell carcinoma (RCC), necessitates precise localization and segmentation for accurate diagnosis and effective treatment planning. Although imaging techniques such as computed tomography (CT) scans and magnetic resonance imaging (MRIs) are standard tools for detection, they often struggle with the accurate segmentation of tumor boundaries, leading to potential misdiagnosis and delayed treatment.

Image segmentation, a critical step in medical image analysis, involves dividing an image into meaningful regions to isolate specific anatomical structures such as tumors[16]. Manual segmentation, while commonly used, is time-intensive, error-prone, and heavily reliant on radiologists expertise. Traditional segmentation methods and even advanced deep learning models like U-Net and its extensions still face challenges, such as poor delineation of small or low contrast tumors, misclassifications of benign cysts as malignant, and limited ability to capture contextual dependencies[2]. These limitations underscore the need for more accurate and efficient solutions. Artificial intelligence, especially through deep learning, presents promising opportunities in this space. Among recent innovations, the Attention-powered U-Net architecture stands out by incorporating attention gates into the traditional U-Net framework, enabling the model to focus selectively on critical regions within CT scans and thereby enhancing segmentation performance[12].

This thesis investigates the potential of the Attention U-Net model for kidney tumor segmentation using the publicly available KiTS19 dataset[17]. It compares its performance with baseline models such as standard U-Net and ResUNet.

The study is structured into three main chapters: the first chapter covers foundational concepts including kidney anatomy, deep neural networks, Convolution Neural Network (CNN) vs Fully Convolution Network (FCN) architectures, segmentation techniques, evaluation metrics, and attention mechanisms. The second chapter discusses related work, particularly models like Attention U-Net, R2AttU-Net, and FR2PAttU-Net, analyzing their roles and limitations in medical image segmentation. The third chapter presents the implementation details, covering dataset usage, preprocessing techniques, model architecture, training procedures, and performance evaluation.

[1] `https://www.wcrf.org/preventing-cancer/cancer-statistics/kidney-cancer-statistics/`

Chapter 1

Basic Concepts

1.1 Introduction

In this chapter, we provide a comprehensive foundation for understanding kidney anatomy, image processing methods, and various kidney tumor segmentation techniques used to diagnose and treat diseases of the human kidney. The focus is on understanding the critical role of medical imaging (MI) in detecting kidney-related conditions, including kidney tumors, and exploring advances in segmentation techniques, especially those involving deep learning models. Additionally, we will discuss the importance of image-processing tools in enhancing the accuracy and efficiency of diagnostic procedures.

1.2 Kidney

The kidney is a vital organ responsible for maintaining the balance of body fluids and electrolytes by filtering and excreting waste products from the blood. They are paired, bean-shaped organs. It also secretes essential hormones and plays a key role in regulating blood pressure. The structure of the human kidneys is illustrated in Figure 1.1 [16]. In this sense, kidney injuries and diseases are serious medical concerns in urology.

Kidney tumors, most known as kidney cancer, particularly renal cell carcinoma (RCC), are among the most common malignancies that affect the urinary system. These diseases are characterized by the loss of renal function, causing kidney failure, increased risk of death, and other complications of the organ system in the human body. Notably, the exact cause of kidney cancer remains unknown, and several risk factors have been identified. These include smoking, obesity, poor diet, excessive alcohol consumption, a family history of hypertension, exposure to radiation and chlorinated chemicals, and genetic predisposition [18].

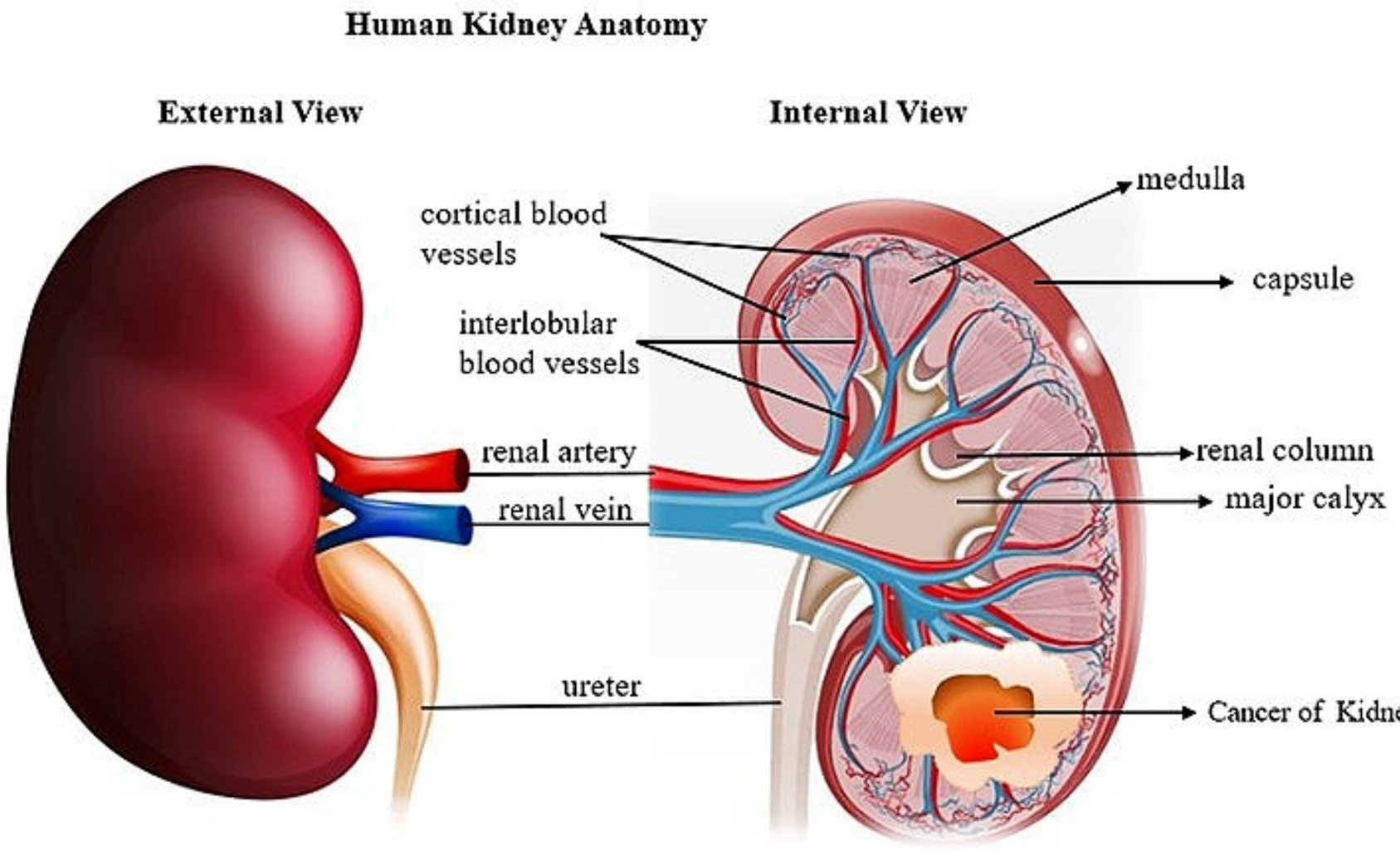

Figure 1.1: Diagram showing human kidney anatomy and renal cell carcinoma developed inside the kidney. [1]

1.3 Imaging Modalities

To evaluate and examine kidney diseases and tumors, medical imaging (MI) techniques are categorized into various types, including ultrasound sonography (US), computed tomography (CT), and magnetic resonance imaging (MRI). Medical images (MI) exhibit excellent homogeneity, which can complicate and make it challenging to identify regions of interest (ROI) and patterns, thereby blurring the boundaries between organs and other areas. Radiologists prefer CT imaging over other imaging modalities due to its ability to produce high-resolution images with good anatomical features. Additionally, it produces images with excellent contrast and exceptional spatial resolution. Therefore, CT imaging offers excellent contrast and exceptional spatial resolution, making it a crucial tool for diagnosing kidney-related diseases [16]. In addition, some CT scan results can be utilized to classify benign cancer (Figure 1.2) [2].

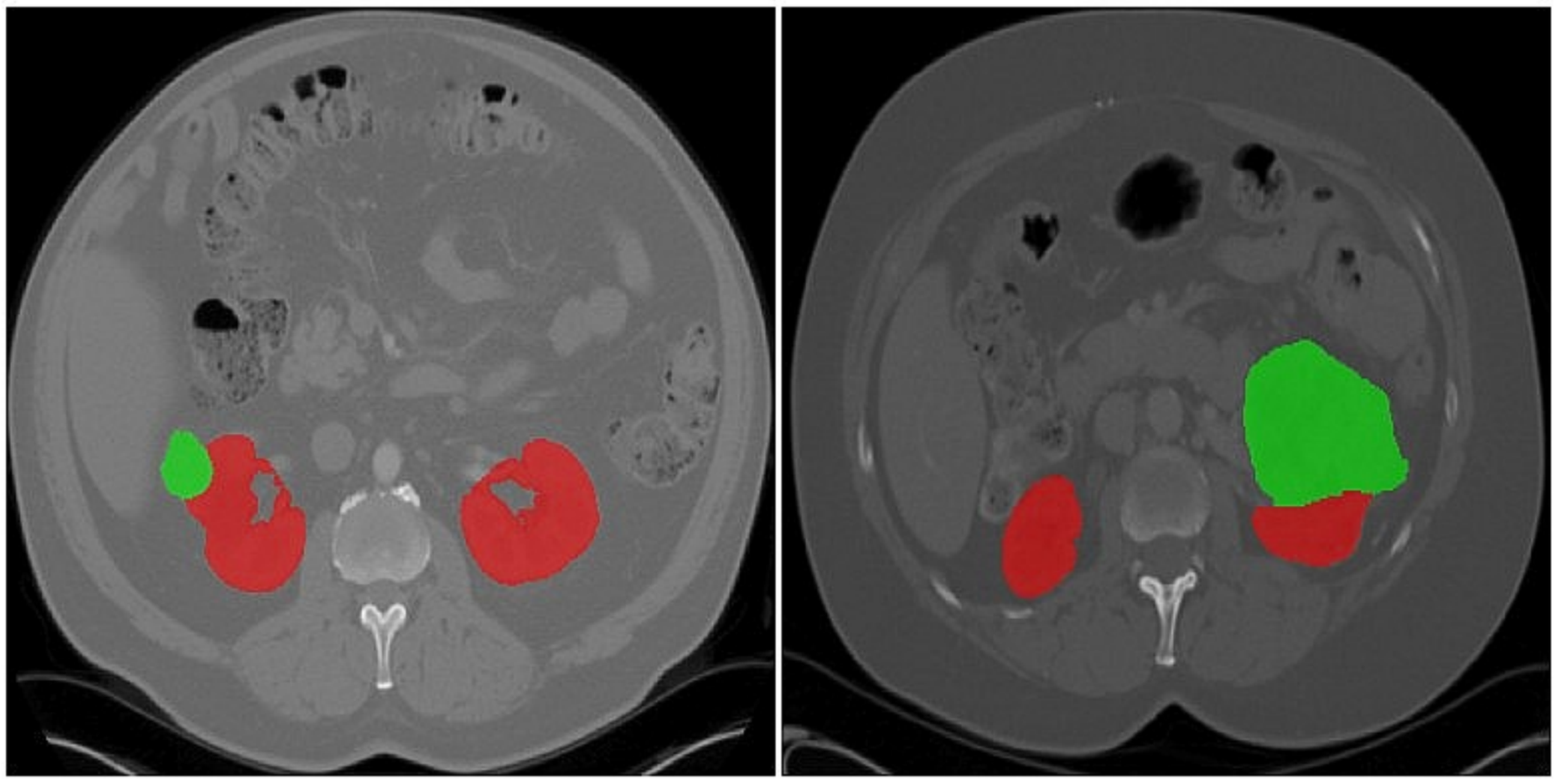

Figure 1.2: Example of an axial slice of 3D CT images of two patients in the KiTS19 dataset. Red color indicates kidneys, and green color indicates tumor region [2].

Accurate segmentation of the kidney from medical images is a fundamental step in computer aided diagnosis, therapy planning, and disease monitoring. Various imaging modalities have been applied in kidney imaging, each offering unique benefits and limitations. The most widely used modalities in the literature for kidney segmentation include computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US), as shown in Figure 1.3. This section provides a detailed explanation of the clinical use of each modality, imaging characteristics, and relevance to kidney segmentation.

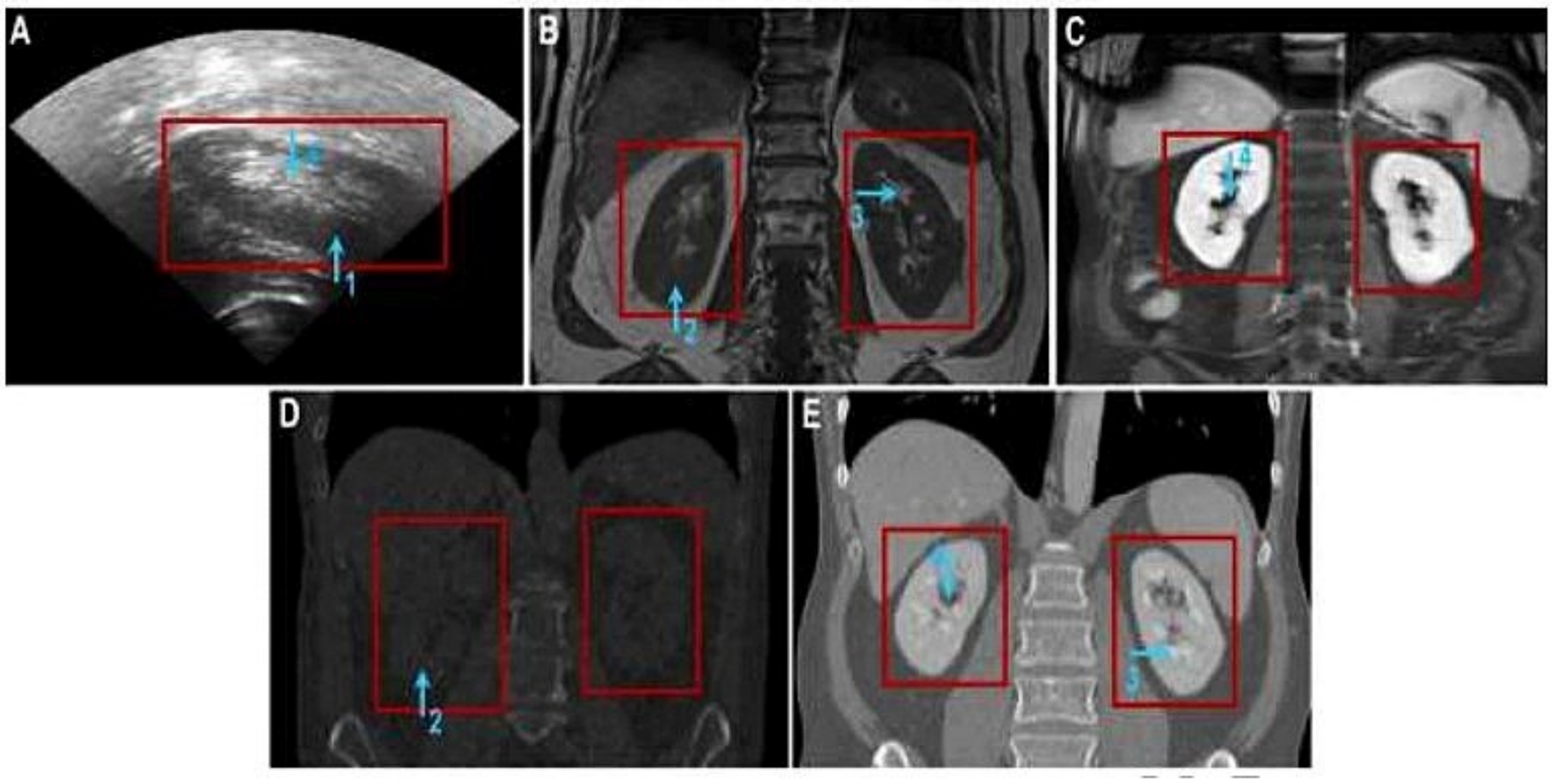

Figure 1.3: (A) US image; (B) MR image; (C) Contrast-enhanced MR image; (D) CT image; (E) Contrast-enhanced CT image; 1- Parenchyma; 2- Cortex; 3- Medulla; 4- Renal sinus [3].

1.3.1 Ultrasound (US)

Ultrasound is often the first line imaging modality in renal evaluations due to its accessibility, low cost, and lack of radiation exposure. In US images, the renal parenchyma appears hypoechoic relative to the echogenic renal sinus, with the medulla and cortex distinguishable based on slight differences in echogenicity. It is highly effective in detecting hydronephrosis, assessing renal size and morphology, and identifying cystic lesions (Figure 1.3(A)).

Despite its safety and portability, ultrasound imaging presents several challenges for segmentation algorithms. It is subject to operator dependency and artifacts such as speckle noise, acoustic shadows, and low tissue contrast. These factors introduce significant variability, making automated kidney segmentation in US a nontrivial task. Recent studies have addressed these challenges using deformable models, active shape models, and learning-based approaches to enhance segmentation accuracy [3].

1.3.2 Magnetic Resonance Imaging (MRI)

MRI provides high contrast soft tissue images without the use of ionizing radiation. In renal applications, MRI allows precise visualization of internal kidney structures such as the cortex, medulla, and renal pelvis (Figure 1.3(B)) and Figure 1.3(C)). MRI is particularly advantageous for functional evaluation using sequences like diffusion weighted imaging and dynamic contrast-enhanced MRI. in particular, it enables the assessment of renal perfusion and function by capturing the temporal evolution of gadolinium based contrast agents within the renal vasculature.

Although MRI is excellent for distinguishing soft tissue types and detecting lesions not visible on CT, it is less effective in identifying calcifications (e.g., renal stones) and is costlier and slower than CT. Additionally, MRI may be contraindicated in patients with certain implants or those at risk for nephrogenic systemic fibrosis [3].

1.3.3 Computed Tomography (CT)

CT imaging uses X-rays to generate high resolution cross-sectional images of the body. In renal imaging, CT is extensively used due to its excellent spatial resolution and contrast sensitivity. The renal parenchyma typically appears as a homogenous region, whereas the renal sinus, containing fat and urine-collecting structures, presents as lower density areas. With contrast agents, enhanced CT can clearly delineate the cortex and medulla, identify vascular structures, and detect small lesions such as stones or cysts (Figure 1.3(E)). CT is particularly valuable for detecting kidney injuries, tumors as shown in Figure 1.2, and vascular anomalies, and is often used for surgical planning.

However, exposure to ionizing radiation and potential nephrotoxicity of contrast agents remain significant drawbacks, especially in vulnerable populations such as children and patients with chronic kidney disease [3].

1.4 Image Segmentation

The segmentation stage is a critical step in the recognition of the imaging process. It involves dividing and extracting meaningful objects and regions from the entire image. Segmentation involves delineating the boundaries of the region of interest for further analysis [16]. There are several methods for segmenting images: manual segmentation, semi-automatic segmentation, automatic segmentation, and semantic segmentation. Semantic segmentation is crucial for image analysis tasks and plays a significant role in image interpretation. Image categorization, object recognition, and border localization are all required for semantic segmentation. Image segmentation can be divided into three main types.

1.4.1 Segmentation Types

- **Semantic Segmentation** Semantic segmentation assigns each pixel to a particular class, also it is a challenging task in Computer Vision (CV) systems. A lot of Deep learning techniques and methods have been developed to tackle this problem, ranging from autonomous vehicles and human-computer interaction to robotics and medical research. The left image in Figure 1.4 is an example of semantic segmentation. The pixels either belong to the person (a class) or the background (another class) [4].

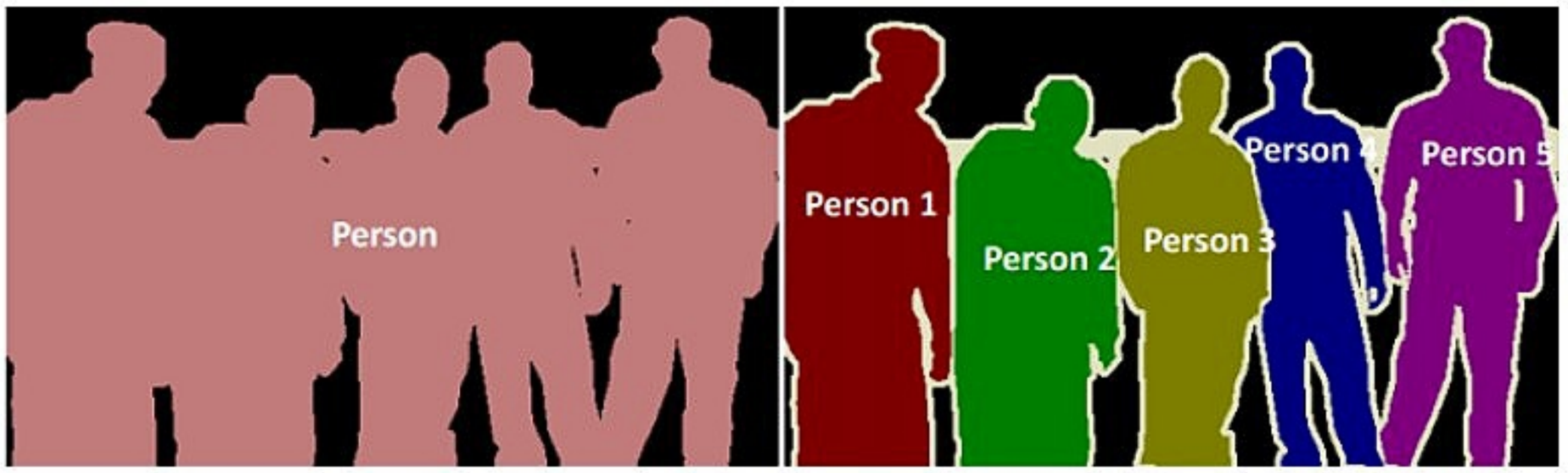

Figure 1.4: Semantic segmentation (left) and instance segmentation (right) [4].

- **Instance Segmentation** Instance segmentation is a fundamental computer vision (CV) task that assigns each pixel to a particular class. However, pixels belonging to discrete objects are labeled with a different color (mask value). Achieving accurate and robust instance segmentation in real-world scenarios such as autonomous driving and video surveillance is challenging [19]. The right image in Figure 1.4 is an example of instance, segmentation. The pixels belonging to the person's class are colored differently.

- **Panoptic Segmentation** is a unified image segmentation task which combines semantic segmentation (assigning a class label to each pixel) and instance segmentation (detecting and delineating individual object instances) Figure 1.5. In panoptic segmentation, each pixel in an image is assigned both a semantic label and an instance ID. This ensures a coherent scene understanding by distinguishing between different instances of "thing" classes (e.g., people and cars), while treating "stuff" classes (e.g sky and road) as amorphous regions without instance differentiation. The task emphasizes a

non-overlapping, complete, and interpretable segmentation of all elements in a scene [5].

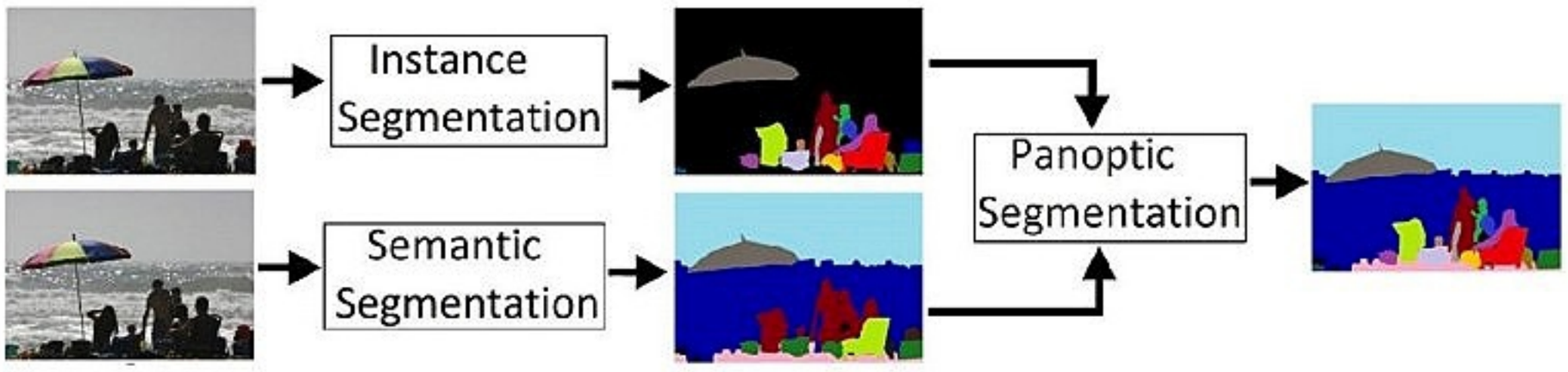

Figure 1.5: panoptic segmentation schematic [5].

In medical imaging, it consists of recognizing and extracting individual organs to help characterize tissue and improve diagnoses. [3], various imaging techniques are applied, such as magnetic resonance imaging (MRI) [20], ultrasound (US), and computed tomography (CT) [2] are commonly used for image segmentation, particularly for tasks like kidney tumor segmentation, also they are used to visualize and evaluate through renal imaging in kidney segmentation. The main motivations for kidney segmentation (renal segmentation) in clinical practice are :

- Evaluation of kidney parameters, namely its size and volume, to diagnose potential diseases.
- Assessment of renal morphology and function.
- Localization of abnormalities or pathologies present in the kidney.
- Facilitate the decision-making process, helping in treatment and intervention planning.

1.4.2 Kidney Tumor Segmentation

Accurate segmentation of kidney tumors from medical imaging modalities such as CT and MRI plays a vital role in supporting clinical decisions, including diagnosis, treatment planning, and surgical intervention. It allows clinicians to assess tumor size, location, and spread with high precision. Deep learning models, particularly convolutional neural networks (CNNs), have revolutionized medical image segmentation by offering robust performance. However, the success of such models heavily depends on various auxiliary techniques, including pre-processing, postprocessing, and data augmentation. These components work in tandem to handle the challenges posed by limited data availability, anatomical variability, and image noise.

Modern segmentation pipelines utilize these techniques to enhance model generalization and accuracy across different patient populations and imaging protocols [21]. In the context of kidney tumor segmentation, these stages are especially important due to the tumors irregular shape, variable contrast in CT images, and overlapping appearance with surrounding organs.

- **Preprocessing** Preprocessing transforms raw medical images into a standardized and cleaner format, which is critical for improving the learning performance of neural networks. Without consistent input formats and quality, segmentation models can suffer from poor generalization and inaccurate boundary detection.

  Common pre-processing steps in kidney tumor segmentation include:

  - **Intensity normalization:** Since CT and MRI intensities vary depending on acquisition devices and protocols, normalizing intensities helps reduce inter-scan variability. For CT scans, intensities are often clipped to a fixed Hounsfield Unit range (e.g., [-200, 250]) before scaling [22].
  - **Resizing or resampling:** Medical images, particularly 2D CT volumes, may vary in resolution. Resizing images to a consistent input size (e.g., 128×128 or 256×256) ensures compatibility with deep learning models while reducing computational costs [2].
  - **Cropping or patch extraction:** Focusing only on the region of interest (ROI) such as the kidney area allows the model to concentrate on relevant anatomical features. This also reduces computational cost and helps address class imbalance [23].
  - **Noise reduction:** Denoising methods, including Gaussian filtering or median filtering, are used to suppress image noise and enhance tissue contrast, improving boundary clarity for both kidneys and tumors [13].

- **PostProcessing** After a model produces segmentation predictions, postprocessing steps are employed to refine results, suppress false positives, and enforce anatomical consistency. These techniques are essential in correcting the limitations of learned models, which may occasionally predict fragmented or disconnected tumor regions. Also These methods ensure that the final segmentation maps are not only accurate but also clinically interpretable.

  Typical postprocessing strategies include:

  - **Morphological operations:** Applying operations such as dilation, erosion, opening, and closing helps to eliminate small noisy predictions and smooth the segmentation boundaries [10]. These are especially useful in removing holes within tumor masks or bridging broken regions.
  - **Connected component analysis:** To prevent over segmentation, only the largest connected region is retained as the final tumor or kidney mask, assuming that smaller components are false positives [24].
  - **Boundary refinement:** Advanced techniques like Conditional Random Fields or active contour models can be used to align segmentation boundaries more closely with image gradients and anatomical borders. This leads to more precise tumor delineation [25].

- **Data Augmentation** In medical imaging, data scarcity and class imbalance are persistent challenges, especially when tumor regions occupy only a small fraction of the entire scan. Data augmentation artificially enlarges the training dataset and introduces meaningful variability, helping to regularize the learning process and reduce overfitting.

  Augmentation techniques employed in kidney tumor segmentation include:

- **Geometric transformations:** These include rotation, flipping, scaling, translation, and shearing. Such transformations simulate real-world anatomical variability and help the model learn rotation- and scale-invariant features [26].
- **Elastic deformations:** These simulate non-rigid anatomical changes in the kidney and tumor shapes. Originally popularized in biomedical image segmentation by Ronneberger et al. in the U-Net paper[9] , elastic deformation has proven effective in training models to cope with realistic organ variations.
- **Intensity variations:** Adjusting brightness, contrast, adding Gaussian noise, or even applying histogram equalization helps models become robust to differences in scan quality and acquisition conditions.

By incorporating these augmentations during training, models are better prepared for the diversity seen in real world clinical datasets.

1.4.3 Segmentation Evaluation

In order to evaluate the performance of an image segmentation model, a set of quantitative metrics is used to assess how accurately the predicted outputs align with the ground truth annotations. These metrics are essential in determining the models effectiveness in distinguishing between different regions or classes within an image, particularly in sensitive applications such as medical imaging. Also, they help quantify not only the correctness of the classification but also the degree of overlap between the predicted and actual regions.These are the most common ones:

- **Accuracy:** The correct predictions produced by the prediction model across all suitable forecasts completed are referred to as the models accuracy [27].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1.1}$$

- **Recall:** True positive rate, the proportion of true positives, or successes, that is accurately detected is calculated as the true positive rate, also known as sensitivity [28].

$$Recall = \frac{TP}{TP + FN} = Sensitivity = TPR \tag{1.2}$$

- **Precision:** The number of correct positive scores divided by the number of positive scores anticipated by the classification algorithm is the positive predictive value, or precision [29].

$$Precision = \frac{TP}{TP + FP} \tag{1.3}$$

- **Specificity:** measures the proportion of actual negative instances (e.g., background or non-tumor pixels) that are correctly identified by the segmentation

model. It reflects the models ability to avoid false positives by correctly excluding areas that do not belong to the target class.[30].

$$Specificity = \frac{TN}{TN + FP} \tag{1.4}$$

Where:

- **TP**: True positives represent the cases where the model correctly predicted the positive class. In other words, these are instances where both the actual value and the predicted value are positive.
- **TN**: True negatives represent the cases where the model correctly predicted the negative class. These are instances where both the actual value and the predicted value are negative.
- **FP**: False positives occur when the model incorrectly predicts the positive class when the actual class is negative. In other words, these are instances where the model falsely predicts the presence of the condition.
- **FN**: False negatives happen when the model incorrectly predicts the negative class when the actual class is positive. These are instances where the model fails to detect the presence of the condition.

- **F1-score:** is defined as a harmonic mean of precision and recall The formula for F1-score is [31].

$$F1 - Score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1.5}$$

- **Intersection Over Union (IOU) :** or Jaccard Index (JI), was used to compare the statistical similarity of regions segmented (Figure 1.6) using a computational approach to hand delineations [32].

$$IoU = \frac{|S_{GT} \cap S_{DL}|}{|S_{GT} \cup S_{DL}|} \tag{1.6}$$

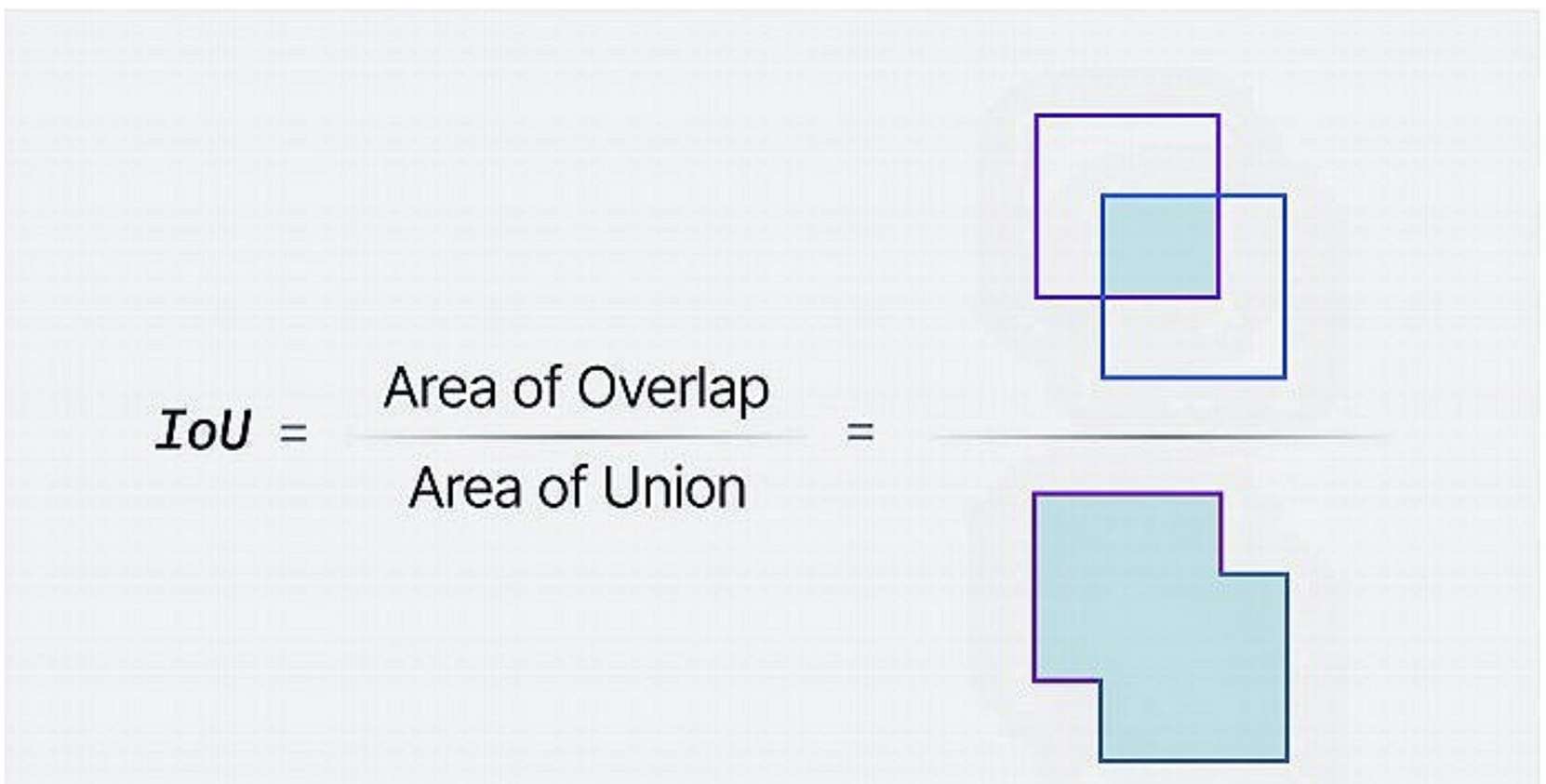

Figure 1.6: Intersection over Union [4].

- **Dice Similarity Coefficient (DSC):** The binary mask produced by the manual segmentation of the experts in the domain corresponds to the binary mask produced by the suggested approach. DSC must be close to unity to ensure that the manually drawn region corresponds to the segmented result correctly [33].

$$DSC(S_{GT}, S_{DL}) = \frac{2 \cdot |S_{GT} \cap S_{DL}|}{|S_{GT}| + |S_{DL}|} \tag{1.7}$$

  Where:

  - S_{DL} represents the predicted segmentation.
  - S_{GT} represents the ground truth.
  - $|S_{DL}|$ and $|S_{GT}|$ denote the cardinality (number of elements) of the sets S_{DL} and S_{GT}.

1.5 Deep Learning Architectures and Techniques

1.5.1 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are specialized kinds of deep learning models that have become the cornerstone of modern Computer Vision (CV), particularly for medical image analysis. A CNN is composed of several layers, including convolutional layers that extract spatial features, pooling layers that reduce dimensionality and provide translational invariance, and optionally fully connected layers for decision-making, as shown in Figure 1.7. Unlike traditional image processing techniques that rely on hand-crafted features, CNNs learn hierarchical feature representations directly from raw input data and are very effective in identifying complex structures in medical images such as tumors, lesions, and organ boundaries. According to Litjens et al[22]. CNNs have significantly improved performance in various diagnostic tasks, including classification, detection, and segmentation across a wide range of medical imaging modalities such as CT, MRI, and ultrasound.

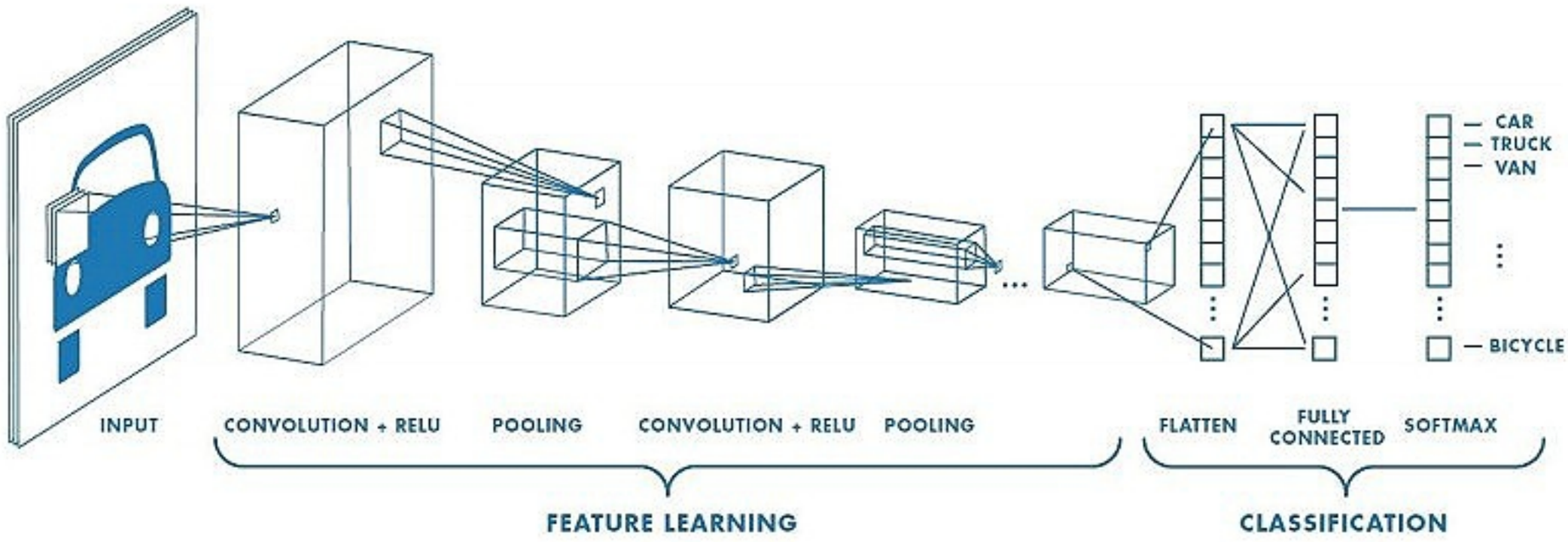

Figure 1.7: The CNN architecture [6].

A typical CNN architecture consists of the following main components:

- **Convolutional layers:** These layers perform convolution operations using learnable filters to extract spatial features from the input. Each filter detects

a specific feature, such as edges or textures, enabling the model to learn local and global representations of the image.

- **Activation functions:** Nonlinear activation functions, particularly ReLU, are used after convolution operations to feed into the model, allowing it to capture complex patterns.
- **Pooling layers:** Pooling operations reduce the spatial resolution of feature maps while preserving the underlying features while reducing computational cost and overfitting.
- **Fully connected layers:** Dense layers are typically used at the output to integrate high-level features for downstream classification or regression tasks.

In the domain of image segmentation, CNNs play a critical role in localizing regions of interest by leveraging learned spatial patterns. Hesamian et al [7]. Emphasized how CNNs have been successful in addressing challenges such as variability in organ shape, noise in imaging data, and low contrast between healthy and diseased tissues. These strengths have made CNNs a foundational component in many segmentation pipelines, particularly as an encoder within larger architectures such as Fully Convolutional Networks (FCNs) and U-Net, which further refine pixel-level predictions.

1.5.2 Fully Convolutional Network (FCN)

Fully Convolutional Networks (FCNs) are a modification of the classical CNNs architecture to enable dense, pixel-wise prediction for tasks such as semantic segmentation. Unlike classical CNNs that end with fully connected layers and output a single classification label, FCNs replace these layers with additional convolutional layers, allowing the model to produce segmentation maps that align spatially with the input image (Figure 1.8). This architectural change allows FCNs to be trained end-to-end for segmentation tasks, learning both global context and fine-grained spatial details simultaneously.

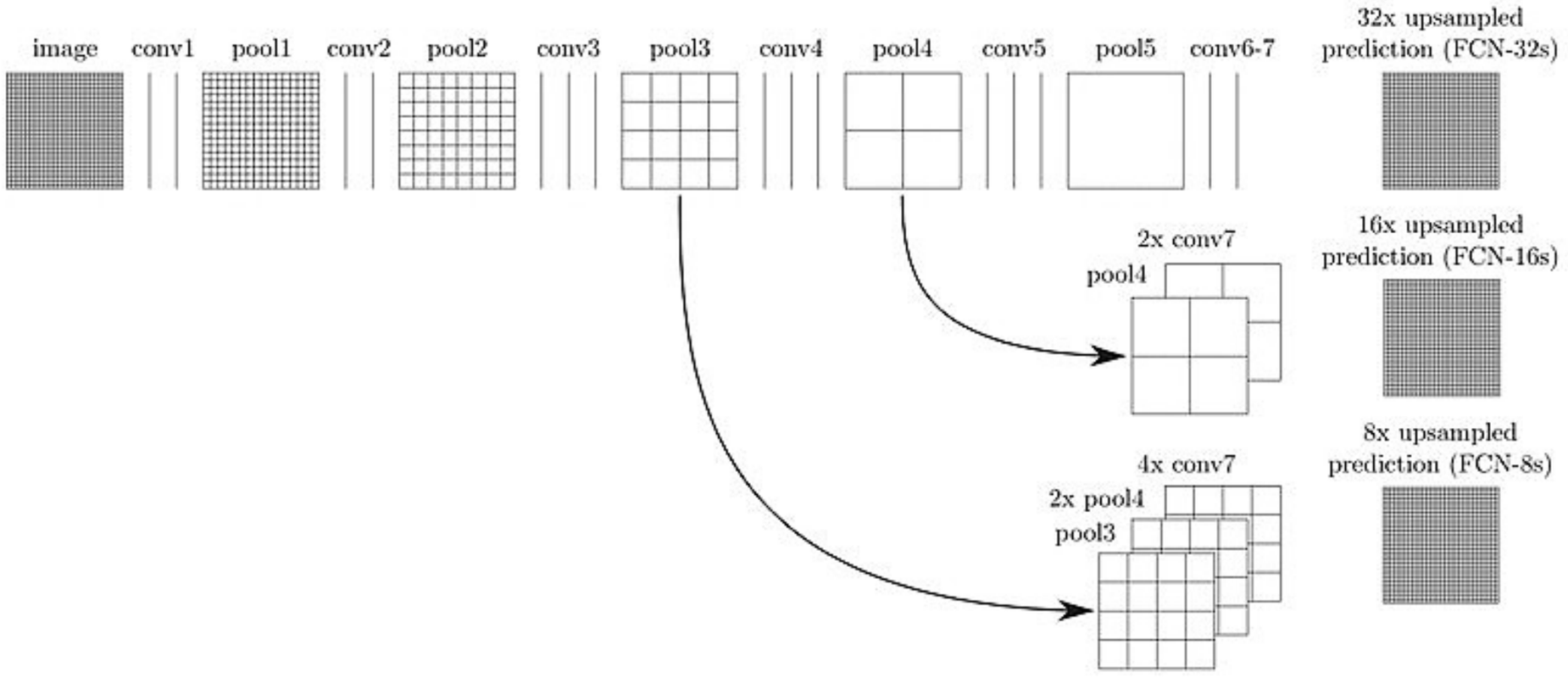

Figure 1.8: The FCN architecture [7].

Litjens et al [22], describe FCNs as a key innovation in the evolution of deep learning for medical image segmentation, particularly because of their ability to work with variable input sizes and produce high-resolution segmentation outputs.

Hesamian et al [7], pointed out that FCNs, especially in their 3D forms like V-Net, have shown significant promise in volumetric segmentation tasks such as identifying tumors within CT or MRI scans. These models not only provide high accuracy but also allow for more consistent and reproducible segmentation results, which are critical in clinical applications.

1.5.3 AlexNET

AlexNet is a leading deep convolutional neural network designed for image classification. It was presented by Geoffrey Hinton and his team, Alex Krizhevsky and Ilya Sutskever, at the 2012 International Large-Scale Visual Recognition Competition (ILSVRC). It competed with ImageNET at the ILSVRC to produce a 1,000-label classification using the ImageNET dataset, which contains over 1.2 million images. The model consists of multiple layers: five convolutional layers and three fully connected linear layers to automatically learn hierarchical representations of multiple layers of image data.

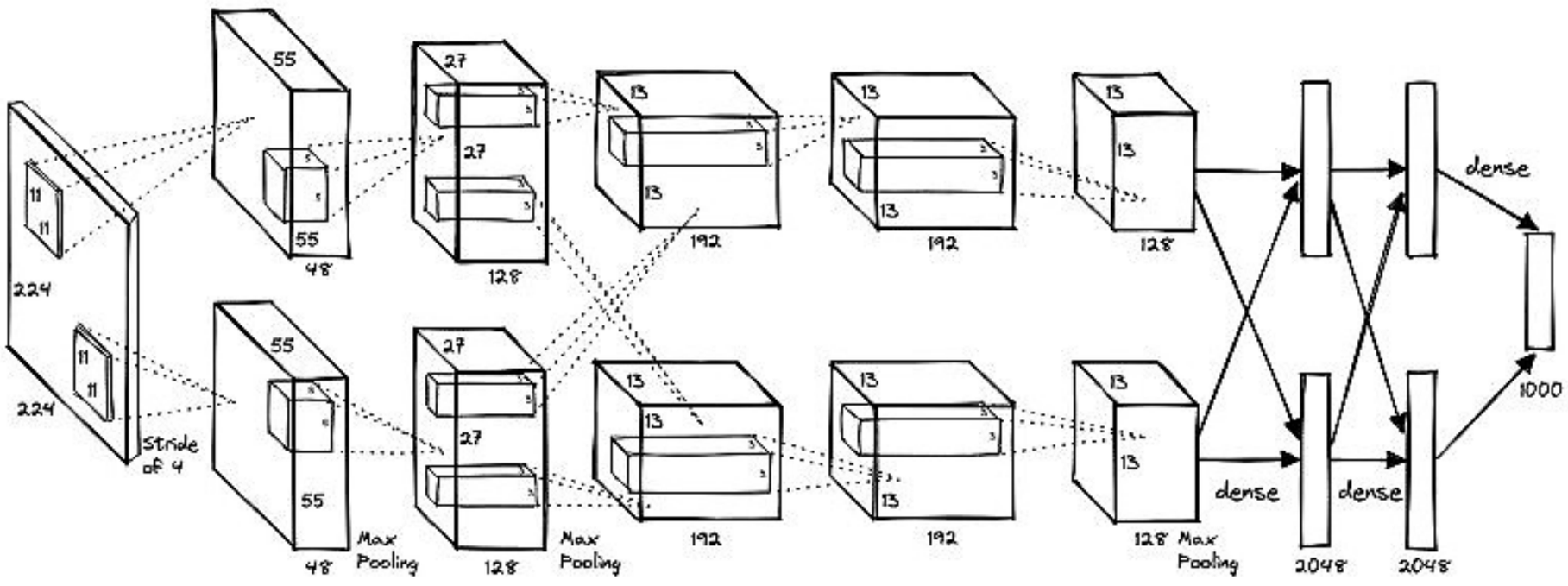

Figure 1.9: The AlexNet architecture [8].

AlexNET featured some innovative techniques that were influential in the network, most notably the ReLU activation function, dropout regularization, and GPU training, achieving an error rate of the top five of 15.3 percent, significantly outperforming ImageNET [8].

1.5.4 U-Net

U-Net is a convolutional neural network that was developed for image segmentation, designed for biomedical image segmentation. Introduced by Ronneberger et al. in 2015 for the ISBI cell tracking challenge, U-Net addressed the problem of segmenting complex structures in medical images where annotated data is often scarce. Its architecture follows a symmetric encoderdecoder structure, shaped like the letter "U" as depicted in Figure 1.10 [9].

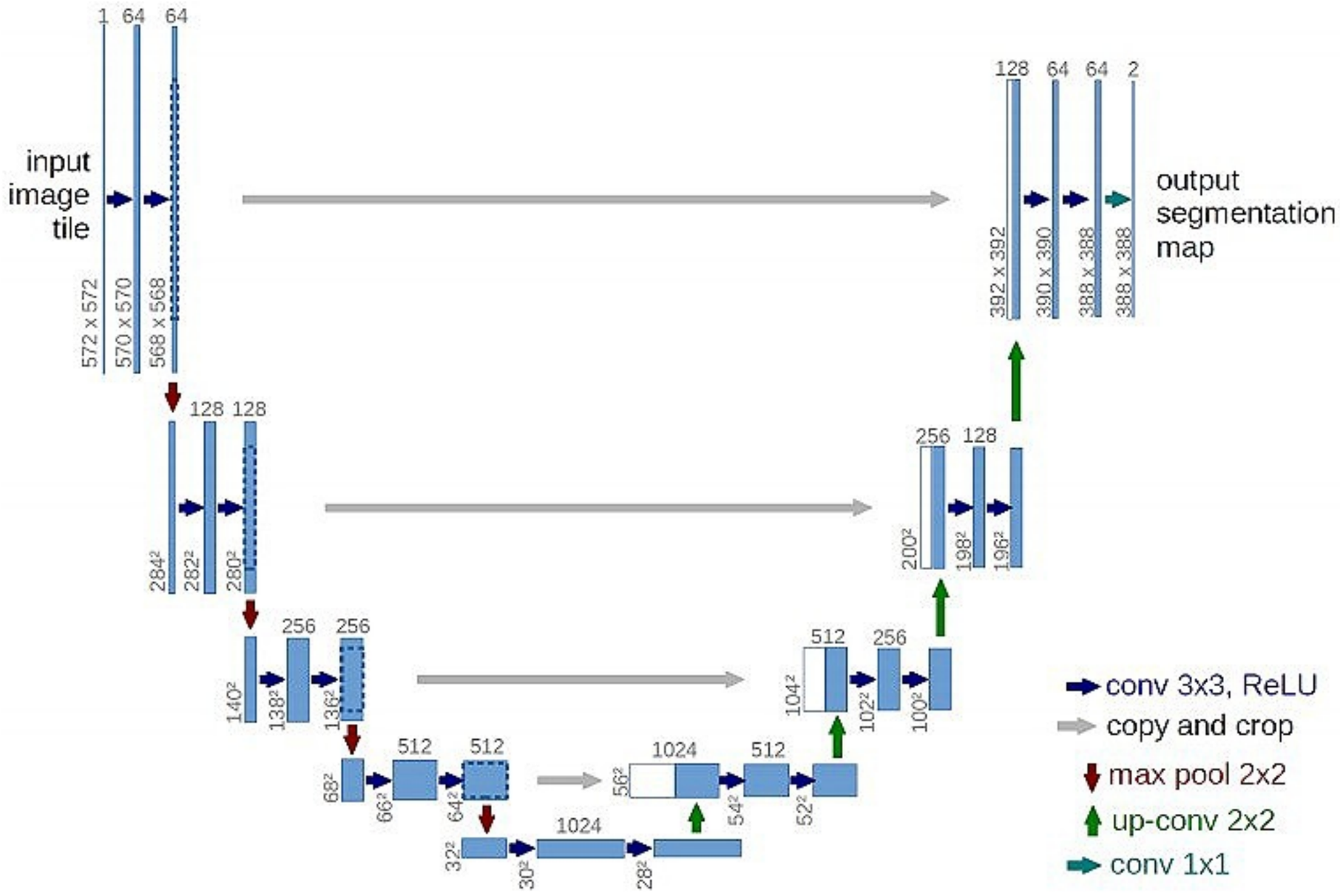

Figure 1.10: The U-Net architecture [9].

The contracting path (encoder) captures contextual information through convolutional and max-pooling layers, while the expanding path (decoder) performs precise localization using up-convolution and concatenation with high-resolution features from the encoder via skip connections. These skip connections help preserve spatial accuracy and improve segmentation quality. U-Net's ability to perform accurate, end-to-end, pixel-level segmentation with a relatively small amount of training data has made it highly effective and widely adopted in medical applications such as tumor detection, organ segmentation, and lesion delineation.

1.5.5 V-Net

V-Net is a volumetric convolutional neural network developed for 3D medical image segmentation, particularly effective for analyzing data from modalities like MRI and CT. Proposed by Milletari et al. in 2016, V-Net extends the U-Net architecture to operate directly on 3D volumetric data instead of 2D slices. It uses a fully convolutional encoderdecoder structure with residual learning, allowing deeper networks to be trained effectively. The encoder path captures hierarchical features through convolution and downsampling, while the decoder path progressively reconstructs the segmentation map using deconvolution layers and skip connections 1.11. A key innovation in V-Net is the use of a Dice loss function, which directly optimizes the overlap between predicted and ground truth volumesa crucial aspect for class-imbalanced medical segmentation tasks. V-Net has been successfully applied to tasks such as prostate segmentation from MRI volumes and other organ delineation tasks in 3D [10] .

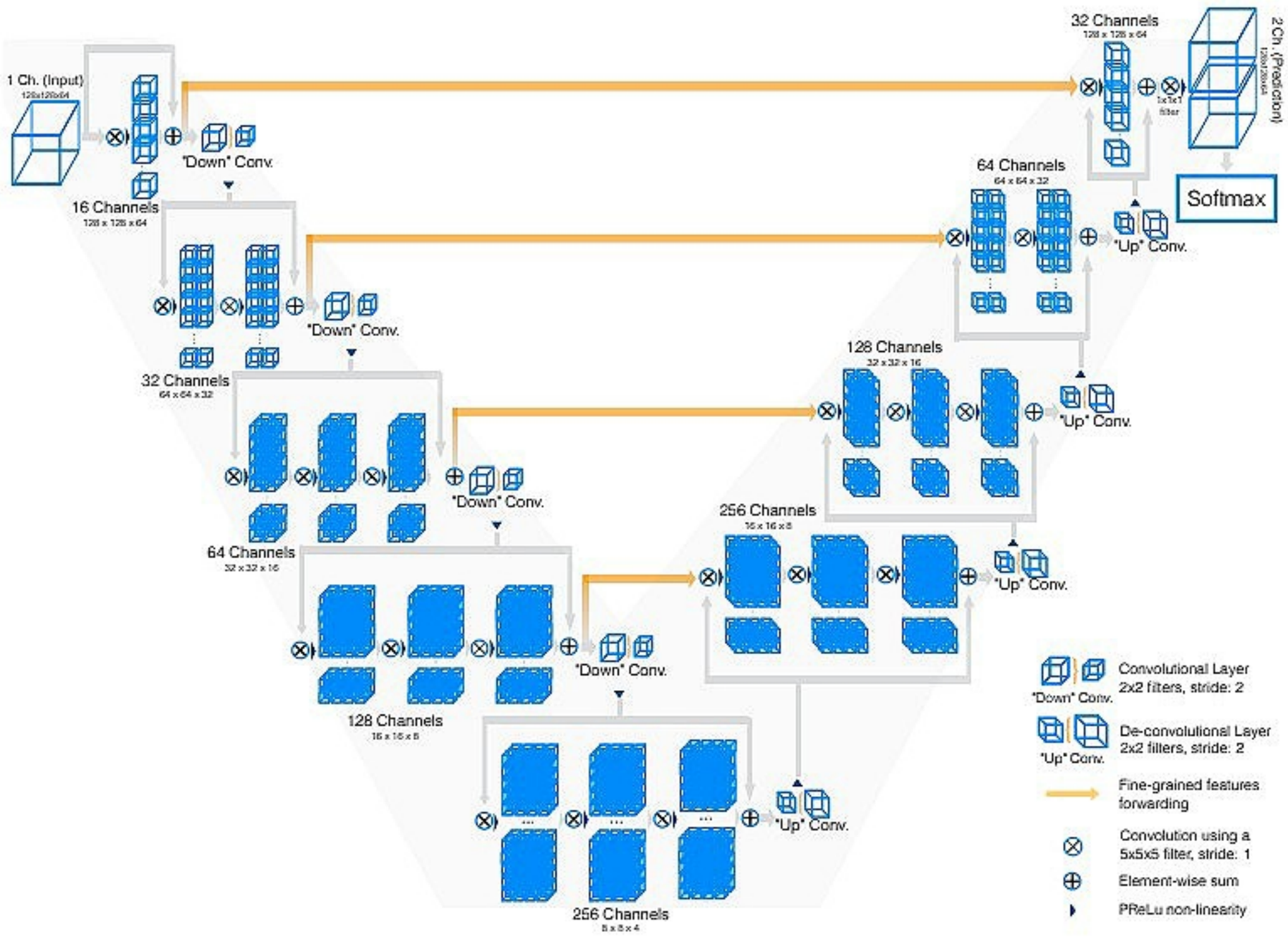

Figure 1.11: The U-Net architecture [10].

1.5.6 ResNet

ResNet (Residual Network) is a deep convolutional neural network architecture introduced by He et al. in 2015, which addressed the degradation problem in training very deep networks. As neural networks grow deeper, their performance often degrades due to vanishing gradients and difficulty in learning identity mappings. ResNet overcomes this by introducing residual learning, where shortcut or skip connections (Figure 1.12) allow gradients to flow more easily through the network by bypassing one or more layers. Instead of learning the full output, each residual block learns the difference (residual) between the input and the output. This simple yet powerful idea enables the construction of extremely deep networks, such as ResNet-50, ResNet-101, and ResNet-152, which have achieved state-of-the-art results in image classification, detection, and segmentation tasks. ResNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2015 and has become a foundational architecture in Computer Vision (CV) [11].

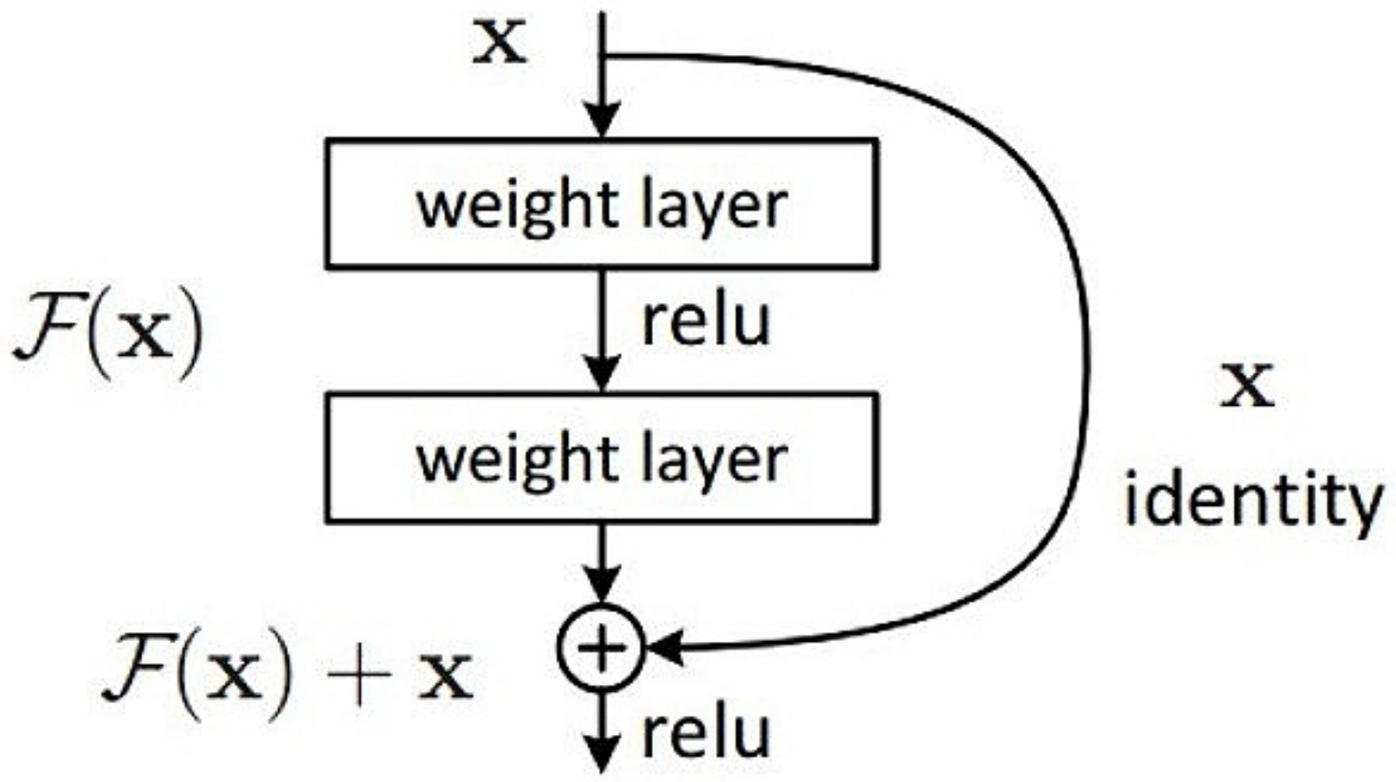

Figure 1.12: Residual learning: a building block [11].

1.5.7 EffecientNETU-Net

EfficientNet-U-Net is a deep learning architecture that integrates the powerful EfficientNet model as the encoder within the U-Net segmentation framework. EfficientNet, proposed by Mingxing Tan and Quoc Le [34] in 2019, introduced a novel compound scaling method that uniformly scales a convolutional network's depth, width, and resolution using a mathematically derived scaling rule. This approach achieves state-of-the-art performance on image classification tasks while maintaining computational efficiency. By embedding EfficientNet (typically variants such as EfficientNet-B0 to B7) as the encoder in U-Net, the model leverages EfficientNet's rich hierarchical feature extraction capabilities alongside U-Net's precise spatial localization, provided by its decoder and skip connections. This combination has been shown to significantly enhance performance in high-resolution image segmentation tasks, particularly within the medical imaging domain.

1.5.8 Attention Mechanism

The attention mechanism in deep learning is a strategy for enhancing neural models by allowing the network to refer back to the input sequence instead of forcing it to encode all information into one fixed length vector. Traditionally, sequence transduction tasks such as machine translation relied heavily on recurrent neural networks (RNNs), which face limitations in capturing long range dependencies and supporting parallelization.

Vaswani et al. (2017) proposed the Transformer, an architecture built entirely around attention, replacing recurrence and convolution with self-attention mechanisms to model global dependencies in input and output sequences [35]. The Transformer further enhances this mechanism through multi-head attention, which projects the input into multiple subspaces and performs parallel attention operations. Each head captures different aspects of the information, and their outputs are concatenated and linearly transformed. This design increases the model's ex-

pressiveness without significantly increasing computation.

Attention is used in three key ways within the Transformer:

- Self-attention in the encoder allows each token to consider all other tokens in the input.
- Masked self-attention in the decoder ensures autoregressive generation by preventing access to future tokens.
- Encoder-decoder attention lets the decoder attend to the encoders output, aligning input and output sequences effectively.

Compared to RNNs, self-attention has several advantages: it enables greater parallelization during training and provides shorter path lengths for gradient flow, which improves the models ability to learn long-range dependencies [35].

1.5.9 Transfer Learning

Building accurate machine learning models often requires large amounts of labeled data and significant computational resources. However, collecting and labeling massive datasets can be time consuming, expensive, or even impractical, especially in specialized domains like medical imaging or legal document analysis. This is where transfer learning becomes invaluable.

Transfer learning is improvement in learning a new task by transfer of knowledge from a related task that has already been learned. Also, transfer learning is essentially the use of pre-trained neural networks (e.g., Computer Vision (CV), medical imaging, natural language processing (NLP) tasks) to try to work around the perceived requirement of large datasets, and to train deep neural networks to train, two primary strategies are commonly used:

- Using a **pre-trained model** as a fixed feature extractor.
- **Fine-tuning** the pre-trained model on specific data, such as medical images.

The first approach offers a notable advantage: It eliminates the need to completely train a deep network, allowing the extracted features to seamlessly integrate into existing image analysis workflows[22].

Transfer learning has proven effective across a wide range of applications in both classical machine learning and deep learning. Two of the most prominent areas where it is widely used are Computer Vision (CV) and natural language processing (NLP).

In image recognition, training high-performing deep learning models from scratch often requires massive datasets and significant computational resources, sometimes taking days or even weeks. To address this, many research institutions and technology companies develop and release pre-trained models that others can reuse through transfer learning. These models are typically trained on large datasets such as ImageNet and serve as a foundation for downstream tasks via fine-tuning or feature extraction. Some widely used models include:

- EfficientNet developed by Google AI for optimized accuracy and efficiency across model sizes [34].
- ResNet by Microsoft [11].

In the domain of natural language processing, transfer learning is often implemented using pre-trained word embeddings, which represent words as dense vectors based on their semantic context within large corpora. These embeddings capture relationships between words such that those with similar meanings are located close to each other in the vector space. This form of representation enables efficient learning and generalization in a variety of NLP tasks. Two widely adopted word embedding techniques are:

- Word2Vec developed by Google [36].
- GloVe (Global Vectors for Word Representation), developed by Stanford, which captures global word co-occurrence statistics from a corpus [37]

These methods form the foundation for many modern NLP models and have significantly improved performance in tasks like text classification, sentiment analysis, and machine translation [37, 36].

1.6 Conclusion

In summary, this chapter has established the foundational concepts necessary for understanding and advancing kidney tumor segmentation. We reviewed kidney anatomy and the clinical imperatives for precise tumor delineation, examined the strengths and limitations of key imaging modalities (US, MRI, CT), defined the segmentation task and its variants (semantic, instance, panoptic) in the medical imaging domain, and detailed the pre-processing, postprocessing, and data augmentation techniques that underpin modern deep learning based pipelines. We also introduced essential evaluation metrics, such as accuracy, precision, recall, IoU, and Dice coefficient. We discussed important neural network architectures, from classical CNNs and FCNs to U-Net and its volumetric and attention enhanced derivatives. Finally, we discussed the transformative roles of attention mechanisms and transfer learning in overcoming challenges of limited annotated data and complex tumor boundaries. Together, these concepts form the intellectual scaffolding for our investigation of Attention UNets application to kidney tumor segmentation, setting the stage for the detailed literature review in Chapter 2.

Chapter 2

State Of The Art

2.1 Introduction

This chapter provides an overview of the state of the art in kidney tumor segmentation methods, briefly mentioning traditional techniques, machine learning-based approaches, and deep learning methods. The chapter further describes the advancements in U-Net architecture, particularly the integration of attention mechanisms, and explores how these innovations address challenges in accurately segmenting kidney tumors. Despite the progress in the area of study, several challenges remain, including variability in tumor morphology, the presence of benign cysts, and difficulties associated with poor image contrast in CT scans. As such, this chapter discusses the benefits and drawbacks of such methods, with a particular focus on their applicability to kidney tumor segmentation, and discusses ongoing challenges such as the need for large annotated datasets, computational complexity.

2.2 Classical Techniques

Traditionally, the methods used in kidney segmentation has relied on manual delineation. However, this approach is time consuming and labor-intensive, being highly prone to intra-observer variability. For these reasons, a multitude of semi-automatic and automatic methods have already been proposed. Despite, advancements kidney segmentation remains a challenging task, namely due to the presence of different renal compartments inside the kidney. Owing to their different characteristics, these structures present different intensity distributions, which leads to a higher intensity inhomogeneity inside the kidney when compared to the rest of the abdominal organs. The spatial localization of the kidney between organs with similar intensities is also a drawback in the segmentation process, given the low contrast between the kidney and its surrounding structures. Another challenge is the shape variability (in terms of length and volumes) expected between subjects. Moreover, certain congenital anomalies also modify the kidneys shape [2, 20]. These factors of the kidney may be more evident depending of the imaging modality.

In this sense, several clinical analyses require multiple imaging acquisitions (with different modalities) to improve the diagnosis and treatment process kidney

segmentation or renal segmentation such as MRI [20], US [3], and CT [2, 20]

2.3 Machine Learning-based Techniques

Earlier systems were built on traditional techniques such as edge detection filters and mathematical models[7]. These were later followed by machine learning techniques, which were based on handcrafted feature extraction, which remained dominant for a significant period. However, the process of designing and extracting these features posed major challenges, and the complexity of such methods limited their scalability as well as applicability in real-world scenarios.

Since the 2000s. Traditional machine learning methods such as support vector machines (SVM), k-nearest neighbors (KNN), and k-means clustering have been extensively utilized in medical image segmentation [7]. These methods typically operate on low- or mid-level features (e.g., intensities, textures) to segment images into tissue classes or to localize pathological regions. In medical imaging, SVMs are often used as voxel-level or superpixel classifiers, while k-means provides an unsupervised grouping of image intensities.

This section discusses how machine learning techniques have been applied in recent research on medical image segmentation, offering a deeper look into their methodologies and analyzing their respective strengths and limitations. We discusses representative studies, datasets, imaging modalities, targeted organs or diseases, and evaluation metrics.

2.3.1 Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised machine learning technique typically used for classification and regression tasks [38]. In the context of medical image segmentation, SVMs are employed to differentiate anatomical structures or pathological regions from surrounding tissues based on extracted features.

In [39], the study titled "Automatic 3D Segmentation of the Kidney in MR Images Using Wavelet Feature Extraction and Probability Shape Model", researchers developed a method for segmenting kidneys in MRI images by combining wavelet-based feature extraction with Support Vector Machines (SVMs). They employed wavelet transforms to extract texture features from different regions of the kidney and used SVMs to classify kidney and non-kidney tissues. The segmentation results were further refined using a probability shape model to adaptively identify kidney boundaries. This approach achieved a mean Dice Similarity Coefficient (DSC) of 90.6% across seven test cases, demonstrating its effectiveness in accurately delineating kidney structures in MRI images.

These studies show that SVM-based segmentation techniques are capable of high accuracy in distinguishing soft tissues and pathological areas from healthy regions in various medical imaging modalities.

2.3.2 SVM Combined with KNN

In [40], Farahani et al. proposed a hybrid machine learning approach for detecting lung nodules from chest CT images. The method combined SVM and K-Nearest Neighbors (KNN) classifiers to enhance classification accuracy. The dataset included 1,000 CT scans from the LIDC-IDRI database. Their approach achieved an overall accuracy of 93.5%.

In [41], Shah et al. proposed a hybrid classification framework for kidney tumor detection using abdominal CT scans. Their system begins with segmentation using the Fuzzy C-Means (FCM) clustering algorithm to isolate the tumor region from the kidney. Once segmented, the system extracts texture-based features using the Grey Level Co-occurrence Matrix (GLCM), which provides statistical descriptors such as contrast, energy, entropy, and correlation. These features are then passed to two classifiers, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), to determine whether the tumor is benign or malignant. The authors compared the classification results using a confusion matrix and observed improved accuracy when both classifiers were incorporated into the decision pipeline. While specific numerical metrics (e.g., sensitivity or Dice coefficient) were not reported, the hybrid approach demonstrated increased classification reliability and effectiveness for medical decision support.

These results illustrate the effectiveness of combining SVM and KNN to exploit both global decision boundaries and local neighborhood relationships, especially in cases where data is noisy or classes are highly imbalanced.

2.3.3 k-Means Clustering

k-means is an unsupervised clustering algorithm that groups image pixels into clusters based on similarity in intensity or texture. In medical image segmentation, it is often used as a pre-processing step to isolate regions of interest.

In [42], Prasad et al. developed a method for lung cancer detection using fuzzy k-means clustering followed by deep learning classifiers. The method was tested on a CT dataset of 1,200 images, achieving 99% sensitivity and 100% specificity in identifying cancerous regions.

In [43], a semantic whole-heart segmentation technique was developed using K-means clustering in combination with mathematical morphology. Applied to chest CT images, the method achieved an average Dice similarity coefficient of 90.8%. The approach was unsupervised and demonstrated robust performance in generating clinically meaningful heart segmentations.

Overall, k-means clustering serves as a fast, unsupervised segmentation tool in medical imaging. It has been applied to modalities including CT (lungs, liver), MRI (brain tumors, bone lesions), and ultrasound (kidney). Reported performance varies by task: brain-tumor Dice scores of 0.82 [44]. These studies confirm that k-means clustering is a simple yet powerful segmentation tool, especially when enhanced with postprocessing techniques like morphological filtering and edge detection.

2.3.4 Limitations of Traditional Machine Learning Methods

Despite their usefulness in structured and low-noise environments, traditional machine learning techniques face several limitations when applied to complex medical images such as CT scans [3]:

- Dependency on manual feature engineering : These methods require domain expertise to select and extract relevant features.
- Poor generalization : They often fail to generalize well across different imaging datasets or modalities.
- Lack of spatial understanding : Unlike convolutional neural networks, they struggle to capture spatial dependencies and contextual information within volumetric data.
- Computational inefficiency : Feature extraction and tuning are time-consuming and not scalable for large datasets.

2.4 Deep Learning-based Techniques

Due to the limitations of traditional machine learning approaches, such as the need for hand-crafted features and limited generalizability, there is a relative scarcity of research explicitly focused on kidney tumor segmentation using classical methods. Consequently, machine learning-based segmentation techniques have been increasingly supplanted by deep learning architectures, such as U-Net, which offer superior performance through end-to-end learning, hierarchical feature representation, and automatic feature extraction. These models have become especially valuable in complex segmentation tasks, where tumor shapes, sizes, and textures can vary significantly. However, understanding the development and impact of earlier techniques remains essential to contextualizing modern advances, particularly in architectures like Attention-Powered U-Net, which integrate mechanisms of selective focus and contextual modeling that originated in foundational segmentation research.

2.4.1 Attention U-Net

The U-Net attention network model, presented by [12], is an extension of the traditional U-Net architecture, which is widely used in medical image segmentation. It provides a solution to the difficulty traditional U-Nets face in accurately segmenting body organs with significant variations in size and shape, especially in complex medical image processing, such as kidney segmentation. To address this challenge, many segmentation techniques rely on multistage concatenated convolutional neural networks (CNN), where the first model focuses on extracting the region of interest (ROI). In contrast, the second model focuses on detailed segmentation. However, these stages result in the overextraction of features, which increases computational cost and model complexity, making them unsuitable for clinical applications.

Attention Gates (AG): are used for medical imaging that automatically learns to focus on target structures of varying shapes and sizes. Models trained with AGs implicitly learn to suppress irrelevant regions in an input image while highlighting salient features useful for a specific task. This enables them [12] to eliminate the necessity of using explicit external tissue or organ localization modules of cascaded convolutional neural networks (CNNs). AGs can be integrated into standard CNN architectures as shown in Figure 2.1 [12] . To do that, they used the U-Net model with minimal computational overhead while increasing the model sensitivity and prediction accuracy. AGs are commonly used in natural image analysis, knowledge graphs, and language processing (NLP) for image captioning [45].

Figure 2.1 shows a block diagram of the Attention U-Net segmentation model. The input image is progressively filtered and downsampled by a factor of 2 at each scale in the encoding part of the network (e.g. $H_4 = H_1/8$). N_c denotes the number of classes. Attention gates (AGs) filter the features propagated through the skip connections. Schematic of the AGs is shown in Figure 2.2.

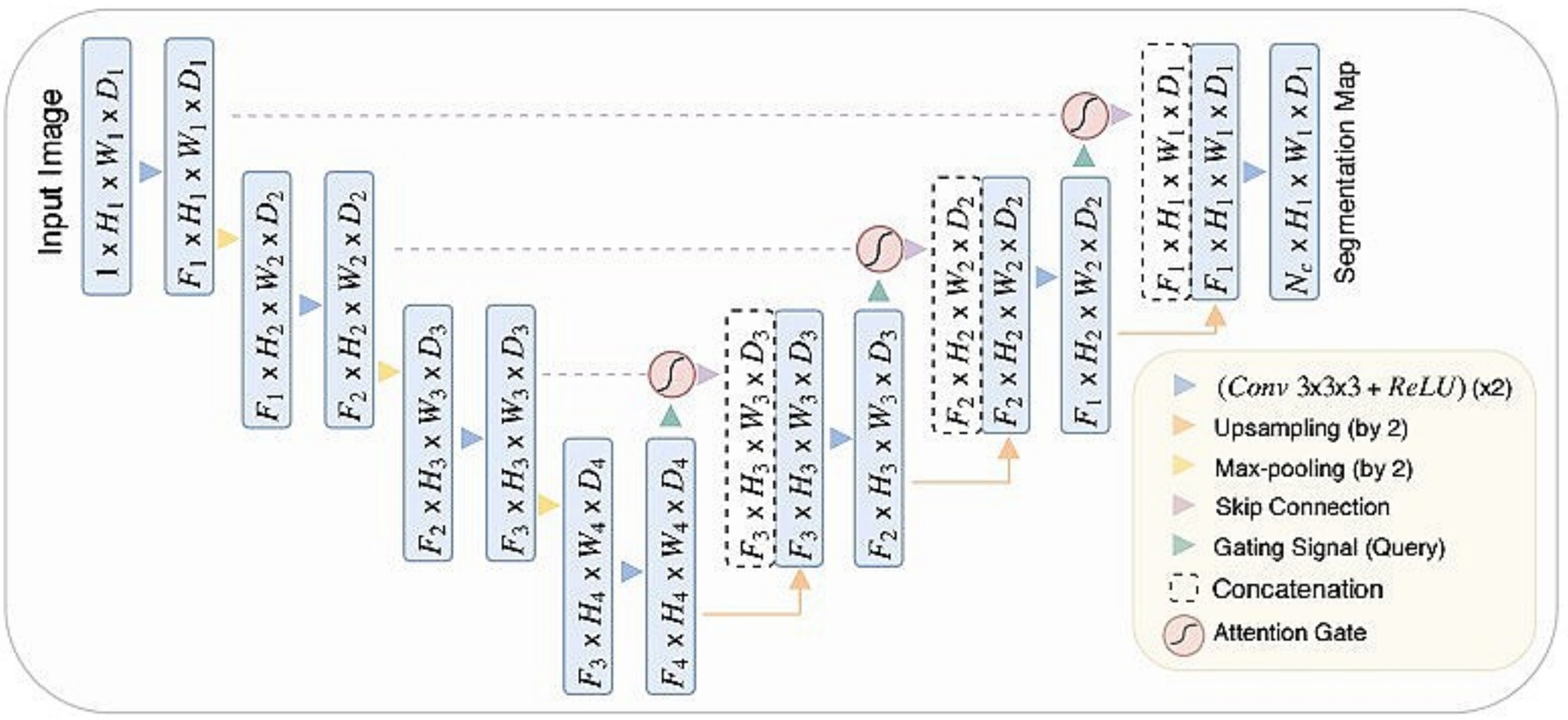

Figure 2.1: Block diagram of the Attention U-Net segmentation model [12].

As shown in Figure 2.2. Attention Gates Architecture consist of an input features (x^l) are scaled with attention coefficients (α) computed in AG. Spatial regions are selected by analyzing both the activations and contextual information provided by the gating signal (g) which is collected from a coarser scale. Grid resampling of attention coefficients is done using trilinear interpolation. Attention coefficients, $\alpha_i \in [0, 1]$, are used to identify salient image regions and prune feature responses to preserve only the activations relevant to the specific task. Therefore, each AG learns to focus on a subset of target structures, and AGs progressively suppress feature responses in irrelevant background regions without the requirement to crop an ROI between networks.and the softmax activation function is often used to normalise attention coefficients (σ_2). However, as sequential use of softmax tends to yield overly sparse activations, Oktay et al. (2018), the authors of the Attention U-Net,[12], chose a sigmoid activation function instead. This decision allows for more flexible and smoother gating of spatial features in medical image segmentation.

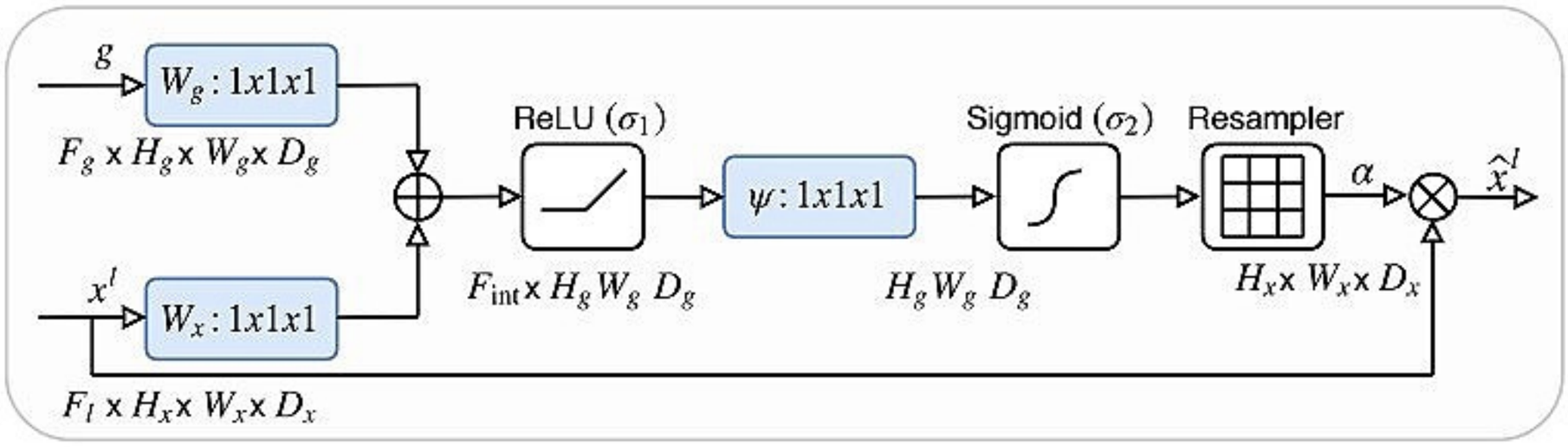

Figure 2.2: Schematic of the additive attention gate (AG) [12].

There are different types of attention that are described as follows:

- **Trainable Attention:** Trainable attention, on the other hand, is enforced by design and categorised as hard-attention and soft-attention [12].
- **Hard Attention:** is a non-differentiable attention mechanism that selects discrete regions of input (e.g., crops or proposals). Training typically relies on reinforcement learning (RL)[1] techniques such as policy gradient methods. [46].
- **Soft attention:** A differentiable, probabilistic attention mechanism that assigns continuous weights to all parts of the input. It supports end-to-end training via standard backpropagation [47].
- **Channel-wise Attention:** In [48], is employed to emphasize key feature dimensions, making it the top performer in the ILSVRC 2017 image classification challenge.
- **Self-Attention:** In [35], often referred to as intra-attention, this attention mechanism focuses on the relationships between different positions within a single sequence to generate its representation. Self-attention has proven effective in various tasks, including reading comprehension. Each element of a sequence (or feature map) attends to every other element to model internal dependencies, capturing long-range context.

2.4.2 Recurrent Residual Convolutional Neural Network U-Net

In 2018, Alom et al [14]. Proposed the recurring convolutional neural network (R2U-Net) shown in Figure 2.4 based on the U-Net architecture as an advanced version of the traditional U-Net architecture. The network incorporates recurrent and residual mechanisms, as illustrated in Figure 2.3, to enhance feature representation in image segmentation, particularly in medical imaging applications. The R2U-Net model combines three of the most potent and effective deep learning concepts: the U-Net baseline architecture, residual learning-inspired networks (ResNETs) shown

[1]Reinforcement Learning (RL) is a type of machine learning where agents learn to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. For more details, see `https://www.ibm.com/topics/reinforcement-learning`

in Figure 2.3, and recurrent convolutional neural networks (RCNNs) shown in Figure 2.3(b), to achieve superior performance in incident segmentation and feature representation.

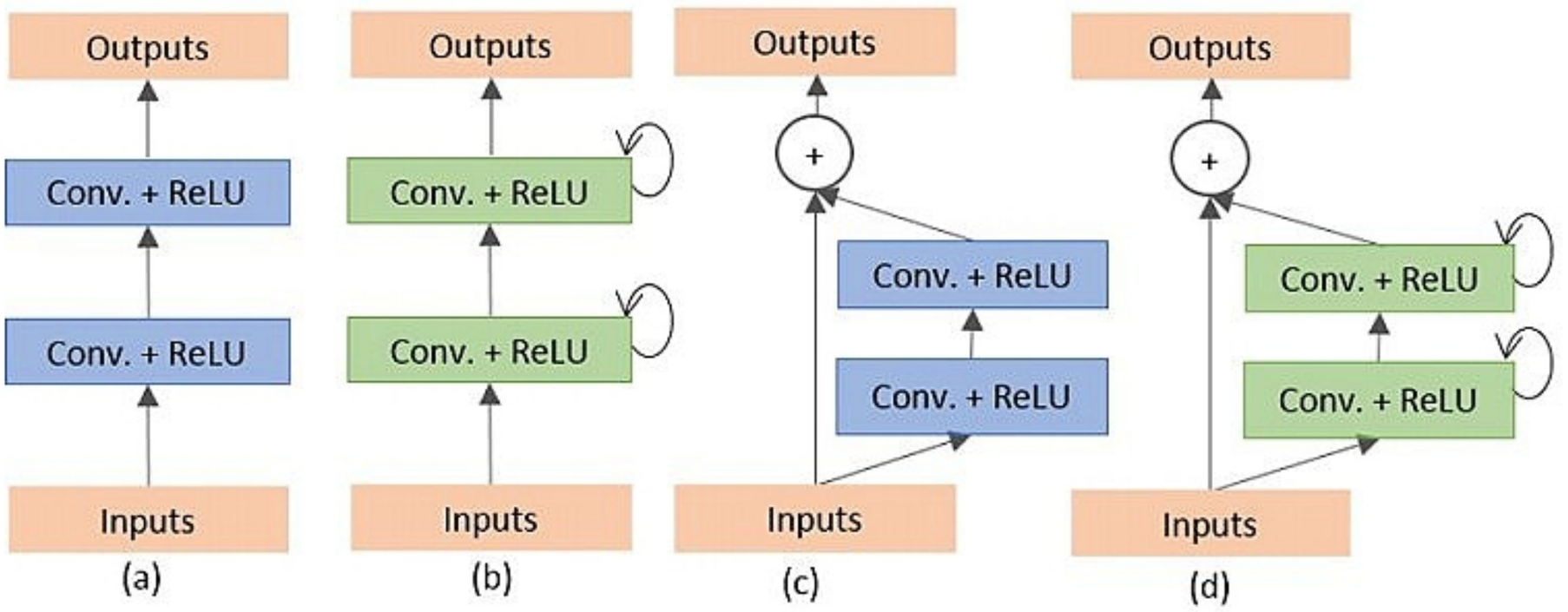

Figure 2.3: Different variants of convolutional and recurrent convolutional units (a) the forward convolutional unit, (b) the recurrent convolutional block, (c) the residual convolutional unit, and (d) the recurrent residual convolutional unit. [13].

Although the U-Net architecture is highly effective in image segmentation, it struggles with feature refinement and contextual understanding of spatial hierarchies. This is due to the accumulation of repetitive feature layers over time, rendering the model incapable of recognizing complex structures, such as blood vessels or tumor junctions.

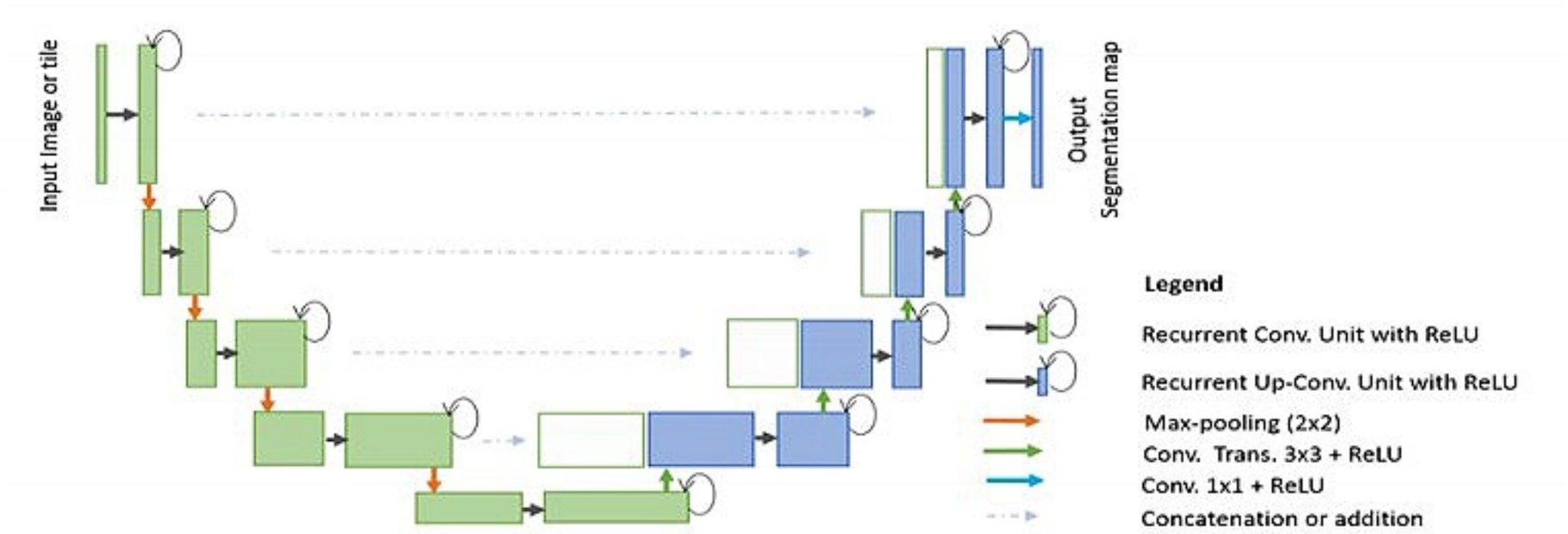

Figure 2.4: The Recurrent Residual Convolutional Neural Network U-Net architecture [14].

Recurrent Convolutional Layers (RCLs): The R2U-Net replaces standard convolutional layers with RCLs, which significantly improve features over discrete time intervals. This process is performed at each pixel using the following function statement. For a pixel at position (i, j) on the k^{th} feature map of layer l, the output at time t is computed as:

$$\mathrm{O}_{ijk}^{l} = \left(\mathrm{W}_{\mathrm{k}}^{\mathrm{f}}\right)^{T} * \mathrm{X}_{l}^{f(i,j)}\left(t\right) + \left(\mathrm{W}_{\mathrm{k}}^{\mathrm{r}}\right)^{T} * \mathrm{X}_{l}^{r(i,j)}\left(t-1\right) + \mathrm{b}_{k} \tag{2.1}$$

The function combines feedforward and recurrent convolutions at each pixel to refine feature representations over time. Here, W_k^f and W_k^r are weights for forward and recurrent convolutions, and b_k is the bias. The convolution operator $(*)$ aggregates local spatial information from the input and the previous output. This accumulation of temporal features captures the hierarchical context, critical to segmenting fine structures such as blood vessels or tumour boundaries.

Residual Connections: Inspired by the ResNet architecture, residual connections are incorporated into R2U-Net to address the vanishing gradient problem common in deep networks. These connections enable the model to learn residual functions and reuse them as the input layer, allowing for more efficient gradient flow and improving the training of deep architectures without any degradation.

The residual unit applies a nonlinear activation functiontypically a rectified linear unit (ReLU)to the output of the recurrent layer. This formulation not only facilitates deepening the network but also preserves low-level features by directly bypassing them to deeper layers, improving convergence and segmentation performance.

2.4.3 Fuzzy set Recurrent Residual Parallel and Attention U-Net

Pang et al [15], designed a deep learning architecture called FR2PAttU-Net by combining different image segmentation methods and techniques to focus and improve the segmentation of kidney tumors, even when the tumors are clear. First, they use the R2PAttU-net network, the first "R" refers to the residual network, and the second "R" refers to recurrent. Also, the letter "P" refers to parallel, which helps the model to deepen and avoid the inability to learn the gradient under the same amount of parameters, resulting in better performance. the model uses the recurrent residual block instead of the traditional Conv + ReLU layer in the encoding and decoding process, which can train a deeper network. All convolution layers are composed of successive convolution are modified to parallel convolutional networks, and combine those blocks into the attention U-Net architecture. As shown in Figure 2.5. Then, they used the fuzzy set enhancement algorithm to enhance the input image and construct the FR2PAttU-Net model to make the image obtain more prominent features to adapt to the model. To implement their model, they used the KiTS19 data set and tested the segmentation effect of the model on different convolutions and depths, and they got a score of 0.948 in kidney Dice and a 0.911 in tumor Dice, resulting in a 0.930 composite score, showing a good segmentation effect.

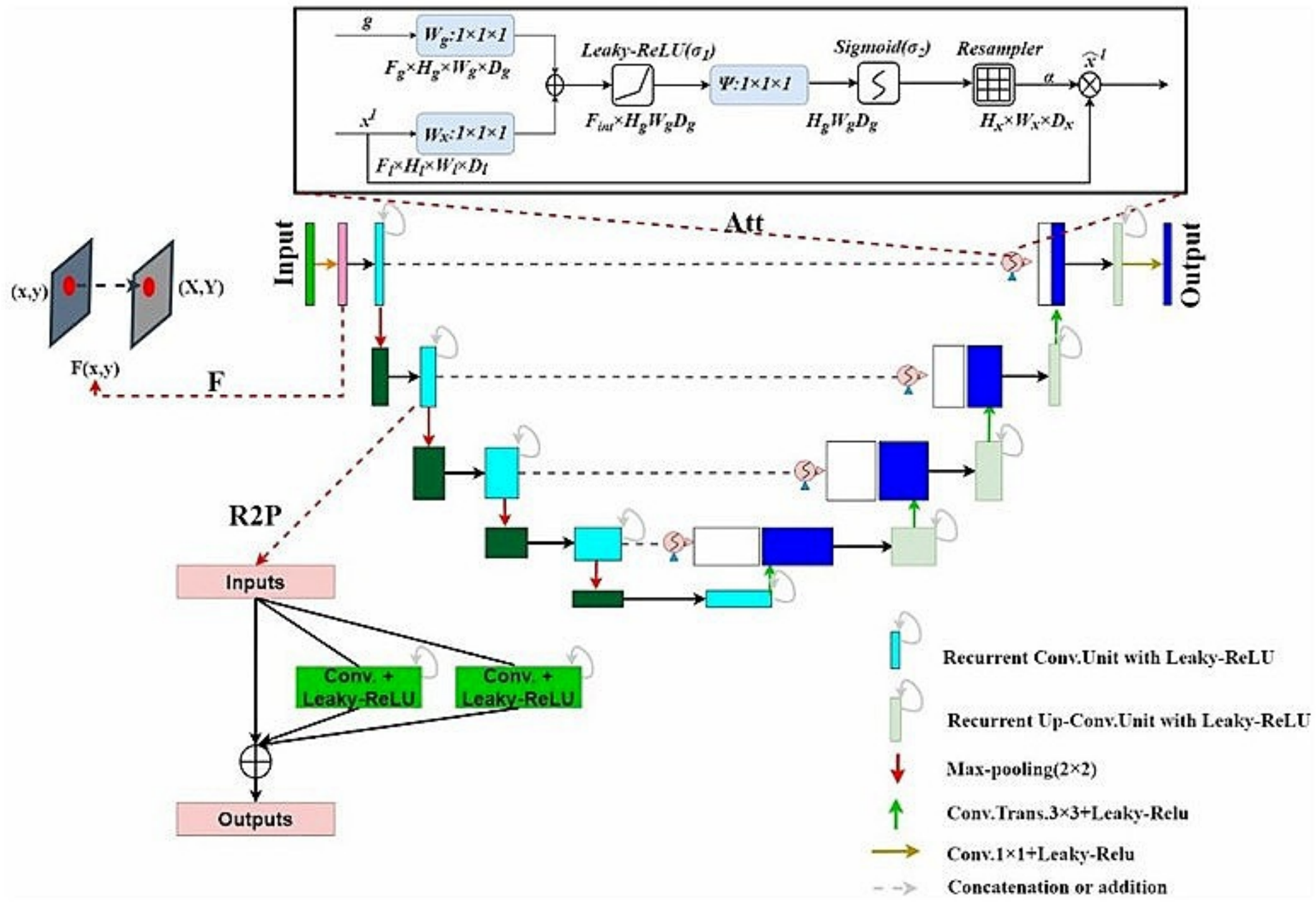

Figure 2.5: FR2PAttU-Net architecture [15].

According to Pang et al. [15], the KiTS19 dataset was used to train multiple deep learning architectures for kidney tumor segmentation, including U-Net, ResU-Net, AttU-Net, and R2U-Net. The average performance of these models, evaluated using the Dice coefficient, is summarized in Table 2.1.

Table 2.1: Comparison of algorithms on Kidney and Tumor segmentation tasks

References	architectures	Kidney Dice	Tumor Dice	Composite score
Reference [9]	U-Net	0.482	0.444	0.463
Reference [11]	ResU-Net	0.688	0.694	0.691
Reference [12]	AttU-Net	0.789	0.735	0.763
Reference [14]	R2U-Net	0.681	0.711	0.696
Reference [15]	FR2PAttU-Net	0.948	0.911	0.930

2.5 Conclusion

This chapter has traced the evolution of some representation kidney tumor segmentation techniques. There is a transition from early manual and classical image processing techniques to machine learning techniques, and in particular there is a lack of research on kidney and tumor segmentation in the latest deep learning architectures. We highlight the limitations of hand-crafted features and shallow models in handling anatomical variability, low contrast, and class imbalance. This motivates the shift toward end-to-end architectures. Key advances in UNet and its volumetric and residual variables laid the groundwork for attention-enhanced networks, which use attention gates to focus on salient regions and eliminate background noise. We reviewed the Attention UNets design, as well as subsequent

extensions like R2UNet and FR2PAtUNet, noting how each builds on skip connections, recurrent and residual blocks, and fuzzy enhancement preprocessing to improve Dice scores on the KiTS19 benchmark. Finally, we discuss complex challenges such as the small amount of labelled data, high computational cost, and the need for strong clinical validation.

Chapter 3

Implementation

3.1 Introduction

In this chapter, we detail the practical implementation of our attention-powered U-Net framework for kidney tumor segmentation. We begin by describing the computational environment and software libraries used, followed by the presentation of the KiTS19 dataset and the steps taken to preprocess the CT volumes into suitable 2D slices. Next, we define the evaluation metrics used to quantify segmentation performance. We then outline the network architecture in full, highlighting how attention gates are integrated into the classic U-Net structure, and summarize our training configuration, including hyperparameters and callbacks. Finally, we report the results obtained during both training and validation.

3.2 Implementation Setup

3.2.1 Environment

To test and validate our implementation, we chose a special working environment. We utilized Kaggle as a platform for data analytics, leveraging a high-level neural network API written in Python that offers pre-configured notebooks with open-source datasets and GPU/CPU options. Furthermore, we utilized TensorFlow as the development framework, an open-source DL framework designed for numerical computation.

Python[1]: Python is a popular high-level programming language widely used for scientific computing, web development, data analysis, and artificial intelligence. It was created in the late 1980s by Guido van Rossum and released in 1991. Python has become one of the most popular programming languages worldwide due to its simplicity, readability, and versatility.

Numpy(Numerical Python): This is a fundamental Python library used for numerical computing. It supports large, multidimensional arrays and matrices, along with a collection of high-level mathematical functions to operate on these

[1]`https://www.python.org/`

arrays. NumPy is essential for performing mathematical operations, such as linear algebra, Fourier transformations, and random number generation. It serves as the foundation for many other scientific computing libraries, including SciPy and Pandas.

Matplotlib: Is a popular Python library used for creating static, interactive, and animated visualizations. It provides an extensive collection of plotting functions and tools, enabling users to generate a wide range of charts, graphs, and plots, including line plots, bar charts, histograms, scatter plots, and more. Matplotlib is highly customizable and is often used for data visualization in scientific, statistical, and engineering applications. It integrates well with other libraries like NumPy and Pandas to visualize results.

TensorFlow[2]: It is an open-source machine learning framework and library developed by Google for building, training, and deploying machine learning (ML) and deep learning (DL) models. It provides a wide range of tools, libraries, and APIs to help developers and researchers create neural networks and other machine learning (ML) models for tasks such as image recognition, natural language processing (NLP), and time series prediction.

Kaggle: It is a global platform and community of data scientists and machine learning practitioners from diverse backgrounds and skill levels. It provides a space for collaboration, learning, and competition in the fields of data science and machine learning. Kaggle facilitates learning, experimentation, and model development by hosting competitions, providing datasets, and utilizing various notebooks and codes to help users learn, experiment, and develop models. The platform emphasizes the value of diversity and believes that embracing differences strengthens the community, fostering innovation in data science and machine learning.

3.2.2 Dataset

In this work, we used the 2019 edition of the Kidney Tumor Segmentation Challenge (KiTS19[3] database to do our experiment. The main goal of this challenge is to advance the research of kidney tumor segmentation and improve the diagnosis and treatment of patients with renal cancer. It consists of 300 contrast-enhanced CT scans and contains data from 210 patients where both image and ground truth labels were provided in an anonymized NIFTI (Neuroimaging Informatics Technology Initiative) format with shape (num_slices, height, width), along with 90 unseen test cases.

3.2.3 Data Pre-processing

To preprocess the KiTS19 dataset, we choose to work with 2D images, so we had to convert from the 3D format and save each slice (image) as an array using the NumPy package in Python. They are an efficient way to store and load NumPy arrays, which are the backbone of scientific computing and machine learning in Python. The dataset comprises 210 cases, with each case featuring three main

[2]`https://github.com/tensorflow/tensorflow`
[3]`https://kits19.grand-challenge.org/data/`

images. The first is a CT scan image, the second is a labeled mask for the kidney, and the third contains tumor masks. The next step is to resize the image. The original size was 512x512, so we downsized it to 128x128 due to memory limitations. This was necessary because our available hardware was unable to process or compress full-resolution images, which would have increased training time.

Also a very important step is to remove all irrelevant slices and keep only the ones that are related to kidney or tumor. In the final step we compressed all the images into a single file called "image.npz", all the kidney mask image into a single file called "ykid.npz", and all the kidney tumor mask images into a single file called "ytum.npz" the Figure 3.1 below shows the pre-process pipeline.

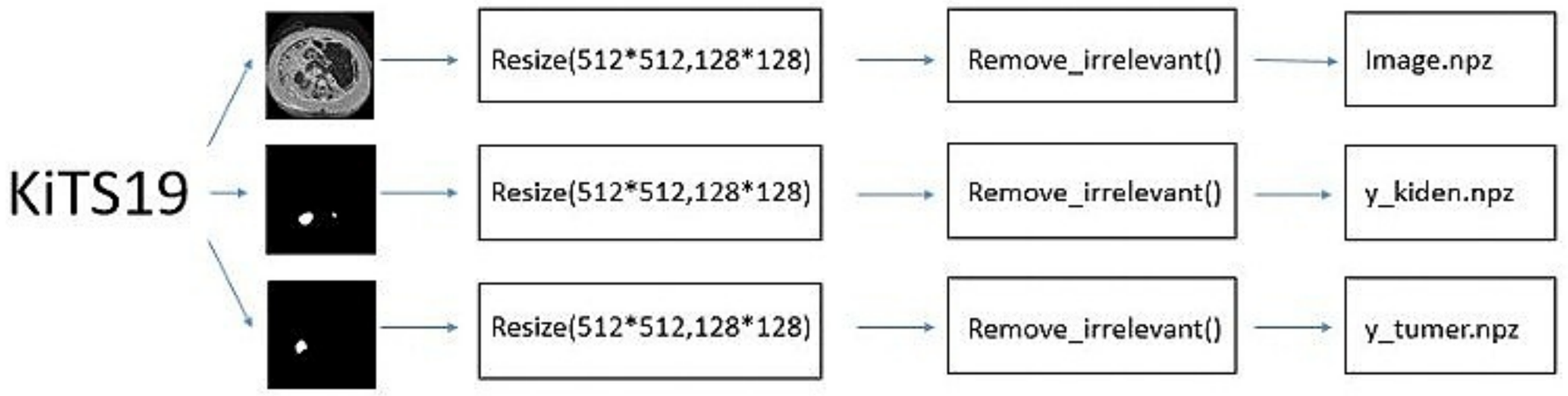

Figure 3.1: Pre-process image pipeline.

3.2.4 Evaluation Metric

As proposed by the challenge organizers, we used the most common evaluation metrics for image segmentation, The quality of the output image segmentation was evaluated with the Dice Similarity Coefficient (Eq. 3.1) computed on the tumor and kidneys, considered a single entity, and on the tumor as a standalone object. Both structures segmented with our method (S_{DL}) were compared with the ground truth segmentations (S_{GT}) provided.

$$DSC(S_{GT}, S_{DL}) = \frac{2 \cdot |S_{GT} \cap S_{DL}|}{|S_{GT}| + |S_{DL}|} \tag{3.1}$$

3.3 Architecture

To train our models, we worked on attention U-Net architecture for its high performance when it comes to medical image segmentation.The implemented model architecture follows an Attention U-Net (Figure 2.1) structure.

The encoder path progressively captures spatial information and compresses it into deeper semantic representations, while the decoder path restores the spatial dimensions using upsampling and attention mechanisms to emphasize relevant features.

Attention gates are integrated at each skip connection to highlight tumor regions and suppress irrelevant background features, enhancing segmentation accuracy.

3.3.1 Training settings and results

Properties	Kidney model Values	Kidney Tumor Values
Number of all images	16293	16293
Number of training images	13056	13034
Number of validation images	3237	3259
Image format	npy(NumPy array file)	npy(NumPy array file)
Modality	CT	CT

Table 3.1: Dataset properties used for the experimentation.

Hyperparameter	Kidney model Settings	Kidney Tumor model Settings
Activation	Sigmoid	Sigmoid
Optimizer	Adam	Adam
Learning rate	0.001	0.001
Batch size	32	1
Epochs	25	4
Metrics	Dice Similarity Coefficient(DSC)	Dice Similarity Coefficient(DSC)
Input images size	128*128	128*128

Table 3.2: Model's hyperparameter setup.

3.3.2 Evaluation Results

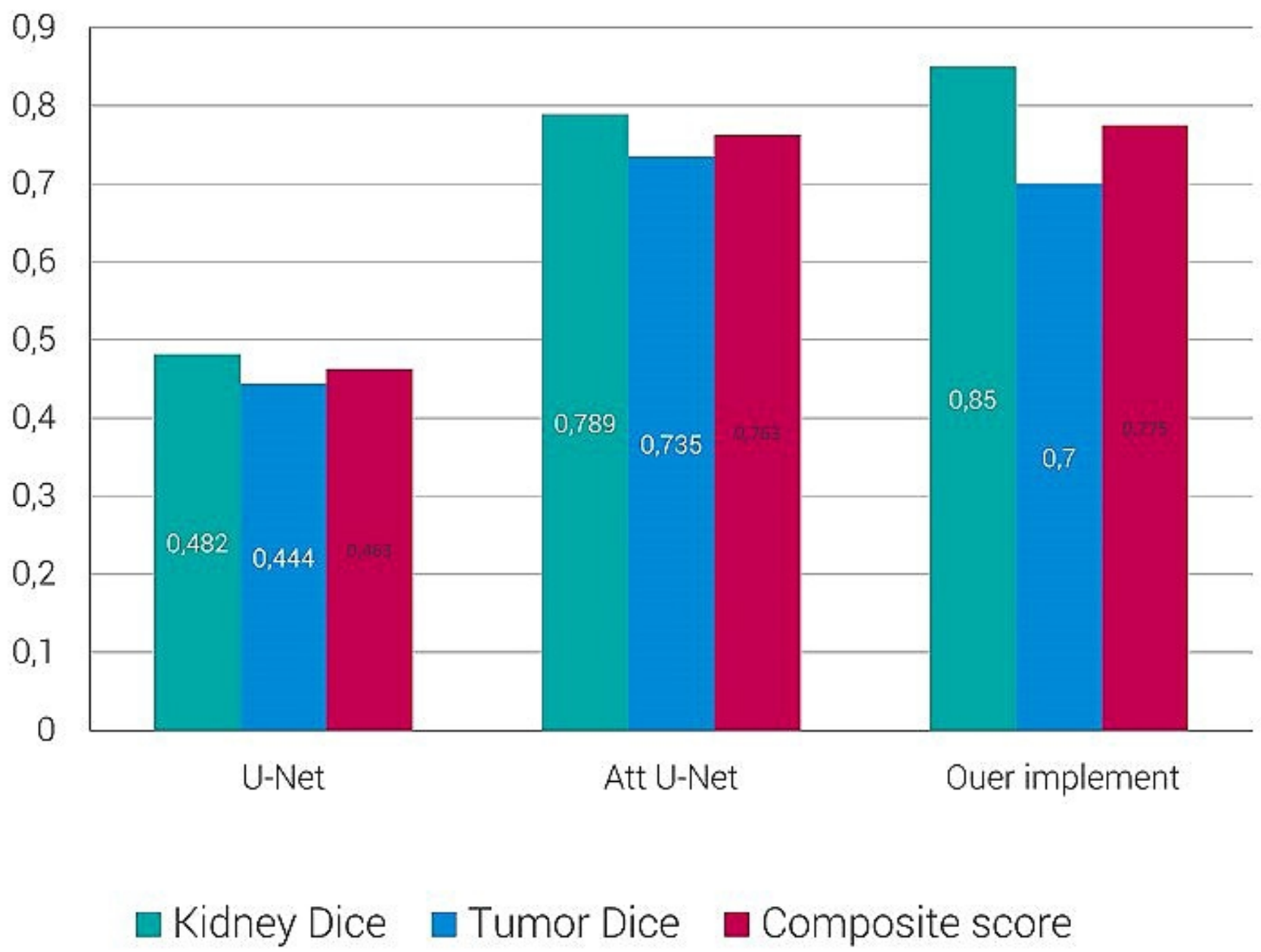

Figure 3.2: Evaluation results for kidney and kidney tumor segmentation models.

The reason for implementing into two model because each model has a different hyperparameter. With the above configuration shown in Tables 3.1 and 3.2, we obtained the results expressed in Tabel ?? on 80% for training and 20% for validation. We obtained a Dice score of 0.85% and 0.70% for kidney and tumor (Figure 3.3 and 3.4), respectively. .

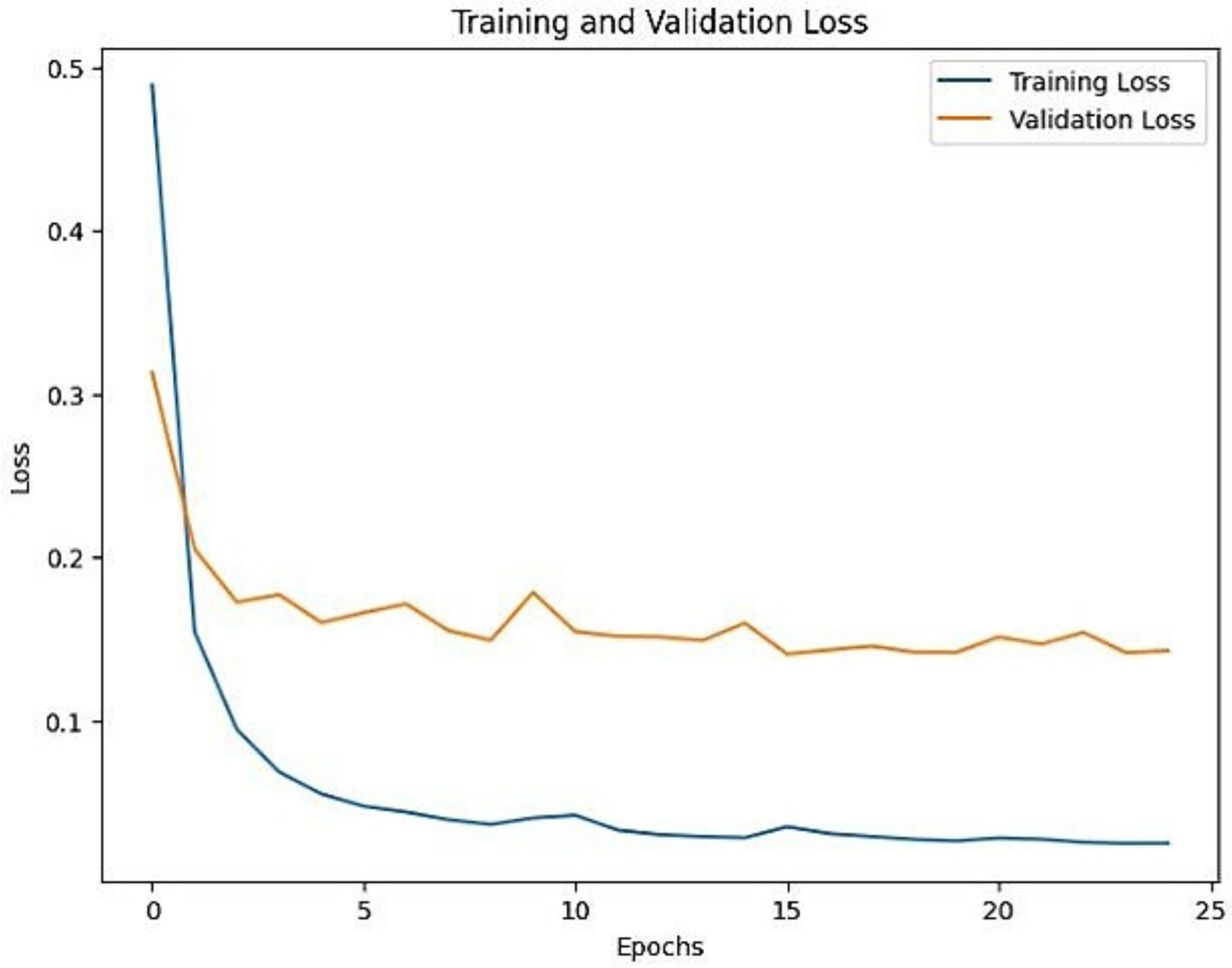

Figure 3.3: Evaluation results for kidney and kidney tumor segmentation models.

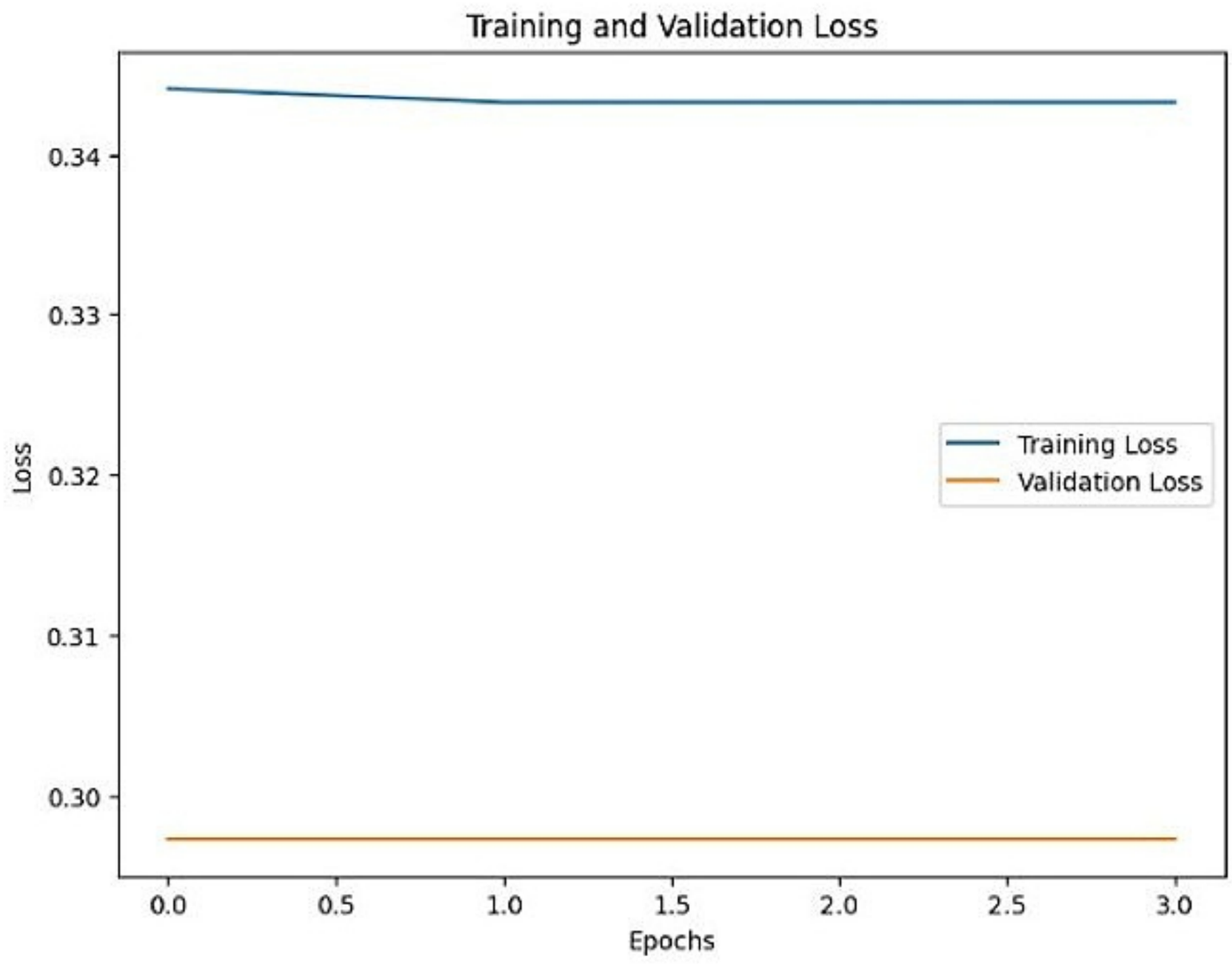

Figure 3.4: Evaluation of the tumor segmentation model.

3.4 Discussion

The baseline U-Net architecture has long been a standard in medical image segmentation due to its encoder-decoder design and skip connections that help retain spatial features. However, it has notable limitations when dealing with complex structures such as kidney tumors, especially in low-contrast or irregular shapes.

Our method is based on the same dataset (KiTS19) as U-Net, ResU-Net, AttU-Net, R2U-Net, and FR2PAtt-U-Net. We employed the Attention U-Net, which incorporates attention gates (AGs) into the standard U-Net framework. This enhancement allows the model to selectively focus on more relevant spatial regions and suppress background noise.

From the evaluation results in Table **??**, the Attention U-Net achieved a Dice score of 0.8595 for kidneys and 0.7027 for tumors. In contrast, as reported in Table 2.1, the standard U-Net achieved only 0.482 for kidney and 0.444 for tumor segmentation. In this case, our approach outperforms U-Net, ResU-Net, AttU-Net, and R2U-Net. However, our tumor Dice score is still about 0.3% lower than some of the more complex models, such as FR2PAttU-Net, suggesting room for further improvement in tumor detection performance.

3.5 Conclusion

In this chapter, we presented the implementation and evaluation of a kidney and tumor segmentation model based on the Attention U-Net architecture. The integration of attention gates enabled the network to focus on critical areas in CT images, leading to more accurate segmentation results. Compared to the standard U-Net, our model showed significant improvements: the kidney Dice score increased from 0.482 to 0.8595, and the tumor Dice improved from 0.444 to 0.7027. This confirms the effectiveness of both the attention mechanism and our tailored preprocessing pipeline.

While these results are promising, further enhancements, such as deeper architectures, 3D context modeling, or multi-model data integration, may help bridge the gap with top-performing models in tumor segmentation.

Conclusion and Perspectives

In this thesis, we presented a deep learning-based approach for automated kidney tumor segmentation using the Attention U-Net architecture. By incorporating attention gates into the conventional U-Net framework, our model was able to selectively focus on the most relevant regions within CT images, enhancing both the precision and reliability of segmentation results. Our implementation on the KiTS19 dataset demonstrated strong performance, achieving Dice scores of 0.85% for kidney and 0.70% for tumor segmentation, which indicates the effectiveness of our pre-processing strategy and model architecture. Compared to baseline models such as standard U-Net and ResU-Net, our method provided a more accurate delineation of kidney structures, particularly in challenging scenarios involving ambiguous boundaries or class imbalance. The inclusion of attention mechanisms significantly contributed to this improvement by suppressing irrelevant background noise and refining feature localization.

Despite the encouraging results, several areas remain that could be improved and enhanced in the future. Firstly, extending the model to operate on complete 3D volumes rather than 2D slices could provide richer spatial context and improve segmentation accuracy, especially for irregular or small tumors. Additionally, integrating multimodal imaging data, such as MRI, alongside CT may provide complementary information and enhance the models robustness. Exploring more advanced attention-based architectures or transformer models may also yield additional performance gains. From a practical standpoint, incorporating uncertainty estimation would enhance the interpretability and reliability of the model in clinical settings, allowing radiologists to gauge confidence in automated predictions. Moreover, domain adaptation techniques should be considered to ensure the model generalizes well across different hospitals, scanner types, and patient populations. In conclusion, the Attention U-Net architecture provides a promising foundation for automated kidney tumor segmentation, with strong potential to support enhanced diagnostic workflows and treatment planning in real-world clinical applications.

Bibliography

[1] Abubaker Abdelrahman and Serestina Viriri. Kidney tumor semantic segmentation using deep learning: A survey of state-of-the-art. *Journal of imaging*, 8 (3):55, 2022.

[2] Andriy Myronenko and Ali Hatamizadeh. 3d kidneys and kidney tumor semantic segmentation using boundary-aware networks. *arXiv preprint arXiv:1909.06684*, 2019.

[3] Helena R Torres, Sandro Queiros, Pedro Morais, Bruno Oliveira, Jaime C Fonseca, and Joao L Vilaca. Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review. *Computer methods and programs in biomedicine*, 157:49--67, 2018.

[4] Vinorth Varatharasan, Hyo-Sang Shin, Antonios Tsourdos, and Nick Colosimo. Improving learning effectiveness for object detection and classification in cluttered backgrounds. In *2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS)*, pages 78--85. IEEE, 2019.

[5] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollar. Panoptic segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2019. doi: 10.1109/cvpr.2019.00963. URL `http://dx.doi.org/10.1109/CVPR.2019.00963`.

[6] Sumit Saha. A comprehensive guide to convolutional neural networksthe eli5 way. *Towards data science*, 15:15, 2018.

[7] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32:582--596, 2019.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL `https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234--241. Springer, 2015.

[10] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565--571. Ieee, 2016.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770--778, 2016.

[12] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[13] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. A survey of mri-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 58(13):R97, 2013.

[14] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.

[15] Peng Sun, Zengnan Mo, Fangrong Hu, Fang Liu, Taiping Mo, Yewei Zhang, and Zhencheng Chen. Kidney tumor segmentation based on fr2pattu-net model. *Frontiers in Oncology*, 12:853281, 2022.

[16] Ravinder Kaur and Mamta Juneja. A survey of kidney segmentation techniques in ct images. *Current Medical Imaging Reviews*, 14(2):238--250, 2018.

[17] Nicholas Heller, Niranjan J Sathianathen, Anirban Kalapara, Erik Walczak, Kyle Moore, Henryk Kaluzniak, Jonathan Rosenberg, Patrick Blake, Zachary Rengel, Michael Oestreich, et al. The kits19 challenge data: Kidney tumor segmentation 2019. *arXiv preprint arXiv:1904.00445*, 2019. URL `https://kits19.grand-challenge.org/`.

[18] Loren Lipworth, Robert E Tarone, Lars Lund, and Joseph K McLaughlin. Epidemiologic characteristics and risk factors for renal cell cancer. *Clinical epidemiology*, pages 33--43, 2009.

[19] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, et al. Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4974--4983, 2019.

[20] Benjamin Cheong, Raja Muthupillai, Mario F Rubin, and Scott D Flamm. Normal values for renal length and volume as measured by magnetic resonance imaging. *Clinical journal of the American Society of Nephrology*, 2(1):38--45, 2007.

[21] Wentao Zhu, Yufang Huang, Liang Zeng, Xuming Chen, Yong Liu, Zhen Qian, Nan Du, Wei Fan, and Xiaohui Xie. Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*, 46(2):576--589, 2019.

[22] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60--88, 2017.

[23] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[24] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24, 2011.

[25] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5):820--838, 2021.

[26] Dan Ciresan, Alessandro Giusti, Luca Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, 25, 2012.

[27] Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, and Xiaoyi Wang. Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, 10:100--107, 2018.

[28] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

[29] Kaitlin Kirasich, Trace Smith, and Bivin Sadler. Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review*, 1(3):9, 2018.

[30] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15:1--28, 2015.

[31] Yutaka Sasaki et al. The truth of the f-measure. *Teach tutor mater*, 1(5):1--5, 2007.

[32] Vanderson Dill, Alexandre Rosa Franco, and Márcio Sarroglia Pinho. Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. *Neuroinformatics*, 13:133--150, 2015.

[33] Ravinder Kaur, Mamta Juneja, and Arup Kumar Mandal. A hybrid edge-based technique for segmentation of renal lesions in ct images. *Multimedia Tools and Applications*, 78:12917--12937, 2019.

[34] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105--6114. PMLR, 2019.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[36] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[37] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532--1543, 2014.

[38] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273--297, 1995.

[39] Hamed Akbari and Baowei Fei. Automatic 3d segmentation of the kidney in mr images using wavelet feature extraction and probability shape model. In *Proceedings of SPIE*, volume 8314, page 83143D, 2013.

[40] Farzad Vasheghani Farahani, Abbas Ahmadi, and Mohammad Hossein Fazel Zarandi. Hybrid intelligent approach for diagnosis of the lung nodule from ct images using spatial kernelized fuzzy c-means and ensemble learning. *Mathematics and Computers in Simulation*, 149:48--68, 2018.

[41] Bansari Shah, Charmi Sawla, Shraddha Bhanushali, and Poonam Bhogale. Kidney tumor segmentation and classification on abdominal ct scans. *International Journal of Computer Applications*, 164(9):1--5, 2017.

[42] J Prasad, S Chakravarty, and M Vamsi Krishna. Lung cancer detection using an integration of fuzzy k-means clustering and deep learning techniques for ct lung images. *Bulletin of the Polish Academy of Sciences Technical Sciences*, pages e139006–e139006, 2022.

[43] Beanbonyka Rim, Sungjin Lee, Ahyoung Lee, Hyo-Wook Gil, and Min Hong. Semantic cardiac segmentation in chest ct images using k-means clustering and the mathematical morphology method. *Sensors*, 21(8):2675, 2021.

[44] Ponuku Sarah, Srigiri Krishnapriya, Saritha Saladi, Yepuganti Karuna, and Durga Prasad Bavirisetti. A novel approach to brain tumor detection using k-means++, sgldm, resnet50, and synthetic data augmentation. *Frontiers in Physiology*, 15:1342572, 2024.

[45] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. Bottom-up and top-down attention for image captioning and visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6077--6086, 2018.

[46] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. Recurrent models of visual attention. *Advances in neural information processing systems*, 27, 2014.

[47] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[48] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132--7141, 2018.

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليـــم العالـي والبحـث العلمـــي

جامعة غرداية

كلية العلوم والتكنولوجيا
قيم الرياضيات والاعلام الالي

غرداية في : 09-07- 2025

شعبة: اعلام الي
تخصص: الأنظمة الذكية لاستخراج المعارف

شهادة ترخيص بالتصحيح والإيداع

أنا الأستاذ(ة) : سعيدي أحمد

الرتبة: .....MCA.....

بصفتي المشرف المسؤول عن تصحيح **مذكرة التخرج ماستر** المعنونة بـ :

Attention-Powered U-Net for Kidney Tumor Segmentation

من إنجاز الطالب (الطالبة) :
- أبي إسماعيل بايوب
- دودو محمد الهادي

التي نوقشت بتاريخ: 30-06- 2025

أشهد أن الطالب/الطالبة قد قاموا بالتعديلات والتصحيحات المطلوبة من طرف لجنة المناقشة وقد تم التحقق من ذلك من طرفنا وقد استوفت جميع الشروط المطلوبة.

إمضاء مسؤول عن التصحيح

مصادقة رئيس القسم
رئيس قسم الرياضيات والاعلام الآلي
الحاج موسى ياسين

جامعة غرداية
قسم الرياضيات والاعلام الآلي
كلية العلوم والتكنولوجيا