Semantic segmentation of Breast Cancer Pathology Images using the U-Net model

Mohammad Hossein Mohammadi

Supervisors: Dr. Mehdi Jabalameli

Department of Computer Engineering
Islamic Azad University Najafabad Branch

This dissertation is submitted for the degree of
Bachelor's Degree

September 2021

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Mohammad Hossein Mohammadi
September 2021

Acknowledgements

I would like to express my sincere gratitude to my esteemed supervisor, Dr. Mehdi Jabal Ameli, for teaching me the importance of ethical standards alongside scientific research, and for generously dedicating his valuable time to guide me through this work.

My deepest thanks also go to my dear father, mother, and sister, whose unyielding love and support have always sustained me through the most challenging times.

Abstract

Breast cancer is the most common malignancy among women and causes more fatalities than any other type of cancer. Nevertheless, early detection of breast cancer can play a crucial role in reducing morbidity and mortality rates. Mammography-based diagnosis faces several limitations, such as imaging noise, low resolution during image capture, and the presence of minute cancerous patterns throughout the image, making accurate detection challenging. Furthermore, previous diagnostic methods often suffer from uncertain boundaries between classes, where most classification algorithm errors occur. Restricting these ambiguous boundaries can significantly improve the detection rate of breast cancer. Therefore, developing methods with acceptable accuracy, precision, and performance remains a central topic in many studies, leading to various proposed techniques whose accuracies are frequently compared. The aim of this study, alongside reviewing breast cancer detection, is to evaluate proposed methods for semantic segmentation of breast cancer using deep learning approaches.

Keywords: Semantic Segmentation, Breast Cancer, Deep Learning, U-Net Model

Table of contents

List of figures

Chapter 1

Introduction

1.1 Introduction and Basic Definitions

In this chapter, concepts are presented that facilitate a better understanding of this thesis. First, we define the problem and explain the significance of the topic. Next, we review the concept of segmentation and its various types. We then define concepts related to pathology images and their characteristics. Finally, basic concepts regarding deep learning will be explained [7, 6].

1.2 Definition of Segmentation

Dividing an image into meaningful structures, known as image segmentation, is often a fundamental and critical step in image analysis, object recognition, visualization, and many other subfields of image processing [22]. Numerous segmentation methods have been introduced over the past decades, and understanding them helps gain a better grasp of the problem. Although segmentation methods can be distinguished from one another, they also share common characteristics. Therefore, differentiating these segmentation techniques is primarily based on the underlying approaches used rather than a completely isolated classification.

1.2.1 Types of Approaches for Image Segmentation

The various approaches used for image segmentation are categorized as follows:

Threshold-Based Segmentation

This method is the simplest approach to image segmentation. By using a predefined threshold, a grayscale image can be converted into a binary image. Figures 1.1 and 1.2 illustrate how thresholds are applied to images. Histogram thresholding and entropy-based thresholding are commonly used for image segmentation. These methods can be applied directly or after executing preprocessing or postprocessing steps.

Fig. 1.1 Original image

Fig. 1.2 Image after applying pixel-level thresholding

Edge-Based Segmentation

This method involves identifying edges, which corresponds to detecting the boundaries of objects within an image to enable object recognition. Figure 1.3 illustrates how segmentation is performed using edge detection techniques.

Region-Based Segmentation

In edge-based segmentation, edges are identified first, and the internal regions bounded by those edges are extracted as the target object. In contrast, region-based segmentation operates

in the reverse direction: it begins with an internal seed point inside an object and expands that point outward until it reaches the boundary edge, thereby recovering the object's extent.

Fig. 1.3 Segmentation using edge detection in the image

Segmentation Using Clustering Methods

Although clustering is sometimes used synonymously with segmentation methods, here it refers to a technique for analyzing patterns in high-dimensional measurement data [5]. In this context, clustering methods aim to group similar patterns together. This objective closely aligns with what is performed in segmentation; thus, certain clustering methods can be effectively applied to image segmentation [17]. Figure 1.4 shows an example of using the k-means method with 6 clusters for image segmentation.

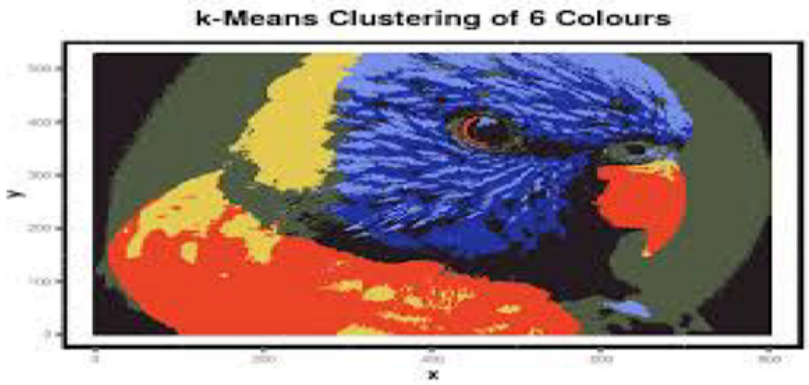

Fig. 1.4 Segmentation based on *k*-means clustering

Template Matching-Based Segmentation

When we possess prior knowledge about the characteristics and visual properties of the object we are searching for within an image, we can utilize this information to locate the target object. This method is known as template matching. In Figure 1.5, the goal is to locate a

human figure within the image by exploiting the specific image characteristics of a human body.

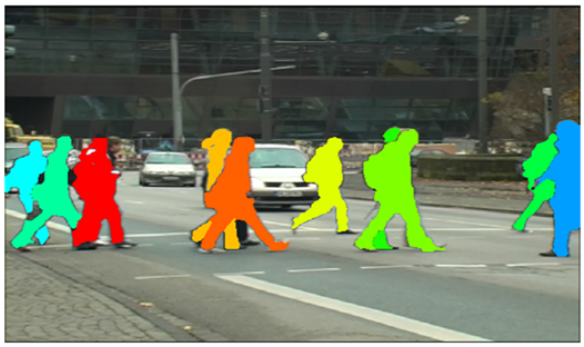

Fig. 1.5 Template matching using human image features

1.3 Types of Medical Imaging

1.3.1 Mammography Imaging

Mammography involves using low-dose X-rays to examine breast tissue [11, 25]. This examination is conducted to detect and diagnose breast diseases in women presenting with clinical abnormalities such as breast pain or nipple retraction. Furthermore, it is routinely performed as a periodic screening tool for the early prevention of breast cancer in healthy women aged 40–45 years and older [14, 20]. This test helps identify potential malignant or benign tumors as well as breast tissue cysts before they become palpable or detectable through physical examination by a physician or the individual.

Mammography does not definitively diagnose cancer; however, if the results indicate a high probability of breast cancer, the suspicious tissue is referred to a pathology laboratory for biopsy. In a biopsy procedure, a sample of the suspicious tissue is extracted using a specialized needle for microscopic examination.

Mammography has been practiced for over thirty years. However, over the past fifteen years, advancements in X-ray imaging technology have enabled significantly more accurate detection of potential abnormalities within breast tissue.

Classification of Different Types of Mammography in Medical Imaging

Mammography is classified into three advanced branches: digital mammography, computer-aided detection systems, and breast tomosynthesis.

1. **Digital Mammography**
   Digital mammography, or Full-Field Digital Mammography (FFDM), is an imaging system in which radiographic films are replaced by X-ray detectors that convert radiation directly into digital breast images. Digital mammography closely resembles digital camera systems, producing high-quality images at low radiation doses. These mammograms are reviewed by radiologists and transferred to computer systems for long-term storage and retrieval.

2. **Computer-Aided Detection (CAD) Systems**
   Computer-Aided Detection (CAD) systems identify regions of abnormal density, masses, or microcalcifications in digital mammography images [22]. CAD systems highlight these suspicious areas on the images, enabling radiologists to evaluate them with greater precision and scrutiny [13].

3. **Breast Tomosynthesis**
   Breast tomosynthesis, also known as Digital Breast Tomosynthesis (DBT), is essentially a three-dimensional mammography technique. In this method, images of the breast are captured from multiple angles, synthesizing a three-dimensional representation of the tissue. Similar to Computed Tomography (CT) scans, this 3D volume is generated by acquiring and combining thin cross-sectional slice images.

   The radiation dose in some breast tomosynthesis systems is slightly higher than that of standard 2D mammography. However, the exposure dose for breast tomosynthesis remains fully approved by the FDA.

4. **Applications of 3D Mammography in Breast Cancer Screening and Diagnosis**
   Studies have demonstrated that screening with breast tomosynthesis leads to improved breast cancer detection rates [9]. In 3D mammography, small cancerous tumors are clearly discernible. Breast tomosynthesis is particularly beneficial for:

   - Increasing the probability of detecting multiple breast tumors
   - Reducing unnecessary biopsies or secondary follow-up examinations
   - Generating clearer visual details of abnormalities in dense breast tissue
   - Achieving higher accuracy in determining the size, shape, and spatial location of breast abnormalities

The Mammography System and Breast Tissue Imaging Process

A mammography unit consists of a rectangular cabinet housing the X-ray tube. This apparatus is designed exclusively for breast X-ray examinations. It includes a dedicated compression mechanism that holds the breast securely in place and compresses it, allowing imaging from various projections. On average, a complete mammography examination takes approximately 30 minutes.

Mammography is performed on an outpatient basis. Before imaging commences, the radiologic technologist positions the patient's breast on a specialized platform within the unit. The breast is compressed using a transparent plastic paddle, with the technician gradually increasing the pressure. Compressing the breast tissue offers several key advantages:

- Thinning out the tissue thickness to achieve clearer images
- Decreasing X-ray scatter to enhance spatial contrast and clarity
- Utilizing a lower X-ray dose across a thinner tissue volume
- Eliminating motion artifacts by immobilizing the tissue to prevent image blur
- Spreading out tissue structures to reveal small hidden abnormalities between overlapping layers

During the procedure, the patient may be asked to adjust their position. The most standard mammographic views taken are the craniocaudal (top-down) and mediolateral oblique (side) views. This process is subsequently repeated for the opposite breast. Breast compression is equally essential in tomosynthesis imaging to minimize motion artifacts and improve overall image quality. During digital breast tomosynthesis screening, synthetic 2D images are often reconstructed from the acquired 3D volumetric dataset.

The patient must remain completely motion-free during the exposure and may be instructed to hold their breath for a few seconds to maintain sharpness and prevent motion blurring.

Duration of the Mammography Procedure

The actual exposure time required for mammography imaging takes only about 10 minutes. However, the entire appointment—including registration, waiting, preparation, and exiting the clinic—takes approximately 30 minutes. The total duration depends on several factors:

- Time spent in the waiting room

- Time required to complete pre-examination medical questionnaires
- Time needed to disrobe and prepare for the exam
- Time taken by the technologist to properly position the patient's breasts
- Time required to inspect the images and ensure repeat exposures are unnecessary due to improper tissue coverage

Recommended Age for Mammography

The American Cancer Society recommends annual mammography screening starting at age 40 [3]. The optimal baseline age to begin screening is 35, followed by annual screenings after age 40, and biannual screenings after age 50.

For women under 25, alternative diagnostic modalities are preferred over mammography unless a highly suspicious clinical finding is present. Individuals in this younger age bracket who undergo early screening typically have a strong family history of breast cancer.

Physicians advise individuals with a family history of breast cancer in close relatives to start regular clinical breast exams and monthly breast self-examinations starting at age 25, ideally performed the day following the end of their menstrual cycle.

Optimal Timing for Scheduling a Mammogram

When scheduling a mammogram, it is essential to monitor your natural menstrual cycle. The week immediately following the end of menstruation is the optimal window for the exam. Conversely, the week prior to menstruation is unsuitable, as breast tissue reaches maximum tenderness and sensitivity during this phase.

Importance and Clinical Benefits of Mammography

Clinical research indicates that women who undergo annual mammography screenings after age 40 experience a significant reduction in breast cancer mortality [11, 20]. Although mammography does not prevent the onset of breast cancer, early detection of tumors and tissue lesions allows for prompt intervention, substantially decreasing the likelihood of requiring aggressive chemotherapy to eliminate tumors.

Furthermore, routine mammography empowers women with a clearer understanding of their breast health, providing up to a 98% sense of reassurance regarding the absence of breast cancer when results are negative.

Discomfort and Pain During Mammography

The perception of pain during mammography varies considerably among individuals. However, most women experience mild discomfort due to the pressure exerted on the breast tissue, which is entirely normal. Pain intensity depends on factors such as breast size and timing relative to the menstrual cycle.

To minimize discomfort, certain precautionary guidelines should be followed. Women with fibrocystic breast tissue are more prone to experiencing pain. Scheduling the appointment post-menstruation and abstaining from dietary triggers such as chocolate, coffee, and heavy spices for at least 48 hours prior to the exam can significantly reduce discomfort.

Mammography During Pregnancy and Lactation

Mammography is not strictly contraindicated during pregnancy because it utilizes low radiation doses. However, routine screening mammograms are generally postponed during pregnancy. If a physician detects a suspicious mass or physical abnormality, a diagnostic mammogram may be prescribed under specific protective protocols.

During lactation, milk accumulation within the mammary glands increases overall tissue density, making the detection of potential lesions more challenging. It is generally advised to wait three months post-weaning before undergoing a mammogram. Nevertheless, women over 40 should consult their physician regarding mammography, as the baseline risk for breast cancer increases in this age group.

Side Effects and Risks of Mammography

The belief that mammography itself causes cancer due to X-ray exposure is a misconception; the radiation dose used in mammography is exceptionally low and clinically negligible.

Approximately 5% to 15% of screening mammograms require supplementary evaluation, such as diagnostic views or breast ultrasound, the vast majority of which yield normal findings. If an abnormality is confirmed, further diagnostic workups or biopsies are pursued. It is worth noting that ultrasound is purely a complementary modality rather than a replacement for mammography and is primarily recommended for primary evaluation in women under 30.

Overall, the diagnostic and lifesaving benefits of mammography far outweigh its minimal associated risks.

Limitations of Mammography

Mammography currently remains the most effective and accurate screening method for detecting breast cancer. However, the procedure has distinct inherent limitations; for instance, it can never provide an absolute, definitive histological diagnosis of cancer on its own.

False-Positive Results A false-positive result occurs when a mammogram incorrectly indicates the presence of a malignant mass. This error occurs more frequently in younger women—particularly during pregnancy and lactation—due to higher dense fibroglandular tissue. Approximately half of all women who undergo regular annual mammograms over a 10-year period will experience at least one false-positive result.

False-Negative Results A false-negative result occurs when a mammogram fails to detect an existing cancerous lesion. Standard mammography screening misses approximately one in every five confirmed breast cancer cases.

Overdiagnosis and Overtreatment In some instances, mammography may detect indolent or non-progressive tumors that are incorrectly categorized as dangerous. This overdiagnosis can lead to unnecessary pharmacological or surgical interventions, exposing patients to treatment side effects for lesions that would not have caused clinical harm.

Mammography in Breast Implants

In individuals who have undergone breast augmentation surgery (mammoplasty), mammography must be conducted under specialized clinical protocols led by experienced radiologic technologists.

Due to the delicate nature of breast implants and the potential risk of rupture during compression, the integrity of the implant must be verified prior to imaging. Excessive compression could potentially cause silicone or saline leakage into the surrounding tissue. Furthermore, displacing the implant during positioning may cause temporary discomfort.

Following mammography, slight changes in implant shape or tissue softness may occur. However, uniform shape retention across both breasts post-procedure cannot be guaranteed.

Interpretation of Mammography Results

Mammographic imaging produces two-dimensional or three-dimensional grayscale images of breast tissue, which are evaluated on high-resolution medical diagnostic displays. A specialist in radiology reviews these images and generates a comprehensive report for the

referring physician. Based on the imaging findings, the radiologist assesses the presence of malignant or benign lesions and may recommend short-interval follow-up exams. Standard mammographic interpretations evaluate parameters such as:

- Microcalcifications or macrocalcifications in ductal structures and parenchymal tissue
- Abnormal dense masses or focal tissue asymmetries
- Structural architectural distortions across breast quadrants
- Focal dense regions isolated to a single breast
- Developing tissue densities compared against prior mammograms

When abnormal features are identified, the radiologist may recommend spot-magnification views, targeted ultrasound, or Magnetic Resonance Imaging (MRI) to definitively characterize subtle lesions that remain inconclusive on standard mammograms.

1.3.2 Breast Ultrasound Imaging

Breast ultrasound is not recommended as a standalone diagnostic tool for breast cancer screening; rather, it serves as an essential complementary modality alongside other diagnostic procedures [9].

When an abnormality or mass is identified on a mammogram or detected via clinical palpation, ultrasound is the premier method to differentiate whether the mass is solid (such as a benign fibroadenoma or a malignant tumor) or fluid-filled (such as a benign simple cyst). However, ultrasound alone cannot definitively confirm whether a solid lesion is malignant, nor can it effectively visualize microcalcifications.

In women under 30 years of age, clinicians generally prefer ultrasound over mammography to evaluate palpable breast masses. Due to the dense glandular tissue structure in younger women (BIRADS categories 3 and 4), mammography offers reduced diagnostic sensitivity, whereas older women exhibit higher adipose (fatty) tissue replacement. Most palpable masses in young women represent benign cysts or focal dense glandular tissue. Furthermore, real-time ultrasound is widely utilized to guide needle core biopsies of suspicious breast lesions.

Introduction to Ultrasound Imaging

Ultrasound imaging is a non-invasive, radiation-free medical imaging technique that utilizes high-frequency acoustic waves to visualize internal anatomical structures. Breast ultrasound

is routinely employed for annual check-ups or as a diagnostic adjunct in suspicious cases. When physical exams, mammograms, or MRI scans reveal ambiguous findings, ultrasound is utilized to gather additional diagnostic information regarding the lesion.

Breast ultrasound requires no special patient preparation; patients simply report to the imaging facility at their scheduled appointment time.

Why Breast Ultrasound is Needed Alongside Mammography

While mammography uses X-rays to detect structural masses, ultrasound is frequently used as a complementary or alternative approach. By relying on sound waves rather than ionizing radiation, ultrasound eliminates radiation exposure risks to the fetus during pregnancy and avoids cumulative X-ray exposure risks, offering several distinct advantages:

- Effective diagnostic utility in younger women whose high tissue density limits mammographic sensitivity
- Highly accurate measurement of exact lesion dimensions with minimal error
- Precise characterization of lesion composition (solid vs. cystic)
- Exceptionally low diagnostic risk profile with high safety margins

Principles of Breast Ultrasound

As noted previously, ultrasound is a non-invasive procedure that generates real-time images using high-frequency acoustic waves emitted by a transducer (probe). A specialized acoustic coupling gel is applied to the skin to ensure smooth probe movement and efficient wave transmission. The sound waves propagate through tissue, reflect off anatomical boundaries, and return to the transducer, where they are converted into digital signals and processed into real-time images of the internal breast architecture.

Ultrasound is a flexible clinical tool. Additionally, specialized modalities such as Doppler ultrasound evaluate vascular blood flow within targeted tissue regions. When requested by a physician, Doppler imaging assesses vascularity within breast masses, providing critical diagnostic insight into tumor vascularization.

Clinical Indications for Breast Ultrasound

Breast ultrasound is primarily indicated for three major diagnostic purposes:

- **Differentiating Mass Composition**: Differentiating between solid masses and fluid-filled cysts to help assess benign vs. malignant characteristics. Doppler ultrasound further evaluates internal blood flow.
- **Supplemental Cancer Screening**: Complementing screening data to confirm or rule out suspicious lesions. Ultrasound is preferred over invasive procedures or high-cost modalities like MRI, especially in pregnant patients or high-risk individuals.
- **Interventional Guidance**: Providing precise real-time visual guidance during core-needle biopsies and tissue aspiration procedures.

How Breast Ultrasound Works

Ultrasound operates on acoustic principles analogous to echolocation used by bats to navigate in the dark. High-frequency sound waves emitted by the probe penetrate breast tissue and bounce off structures, masses, or cysts. The velocity and attenuation of the returning echoes determine the depth and spatial distance of the target lesion.

Furthermore, acoustic reflection patterns indicate whether a mass is solid tissue or a fluid-filled cavity. Doppler ultrasound measures frequency shifts in returning echoes to map blood flow dynamics inside and surrounding the lesion.

Advantages and Disadvantages of Breast Ultrasound

Breast ultrasound presents specific clinical advantages and limitations:

Advantages

- Non-invasive and painless procedure
- Superior soft-tissue contrast resolution compared to conventional X-ray imaging
- Real-time image generation, enabling dynamic biopsy guidance
- Radiation-free imaging, rendering it ideal for dense tissue evaluation

Disadvantages Although safe, ultrasound has technical limitations. It can produce false-positive diagnoses, though error rates are minimized when high-resolution scanners are operated by experienced sonographers.

1.3.3 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a diagnostic radiology modality that combines strong magnetic fields, radiofrequency (RF) pulses, and computer processing to capture detailed images of internal anatomical structures [9]. The scanner consists of a cylindrical bore surrounded by a powerful superconducting magnet. MRI is widely utilized in medical and veterinary diagnostics and operates on the principles of Nuclear Magnetic Resonance (NMR).

Image formation in MRI is a complex physical process. It relies on the nuclear spin properties of hydrogen protons placed inside a primary static magnetic field (B_0). Upon applying radiofrequency pulses (B_1), the proton spins undergo excitation and subsequently return to their thermal equilibrium state during relaxation. Relaxation rates vary across different tissue types; for instance, at a field strength of 1.5 Tesla, the longitudinal relaxation time (T_1) is approximately 260 ms for adipose tissue and 900 ms for cerebral gray matter.

By adjusting pulse sequence parameters, image contrast can be customized to emphasize T_1 or T_2 relaxation properties—a core strength of MRI technology. For example, fat appears bright on T_1-weighted sequences and dark on T_2-weighted sequences. Each 2D image slice is spatially encoded along the x-axis and y-axis using frequency and phase encoding gradients generated by gradient coil magnets. The raw acquired signals are stored in a complex frequency domain known as k-space and reconstructed into spatial images via Inverse Fast Fourier Transform (IFFT).

Image quality is evaluated using spatial resolution metrics. Each image voxel is assigned a intensity value representing the average magnetic resonance properties of the underlying tissue. Voxel dimensions determine spatial resolution and structural detail. While scan times are often extended to improve Signal-to-Noise Ratio (SNR), patient comfort must be maintained. Adult MRI protocols typically last around 20 minutes, whereas pediatric brain scans range from 5 to 15 minutes.

2D and 3D Image Representations

An image can be defined mathematically as a 2D spatial intensity function $I(i,j)$ or a 3D volumetric function $I(i,j,k)$, where $i = 0,\ldots,M-1$, $j = 0,\ldots,N-1$, and $k = 0,\ldots,D-1$ represent spatial coordinates.

The values $I(i,j)$ and $I(i,j,k)$ denote scalar intensity levels, typically rendered as grayscale values in MRI datasets. The image consists of a finite array of pixels in 2D space or voxels in 3D space. Each discrete element is uniquely identified by its intensity value and matrix coordinates (row index, column index, and slice index within the volumetric stack). Figure 1.6 illustrates this volumetric representation.

Voxel dimensions depend on acquisition parameters, magnetic field strength, maximum allowable scan duration, and hardware limits. In standard clinical MRI protocols, isotropic voxel sizes typically range between 1 mm and 2 mm.

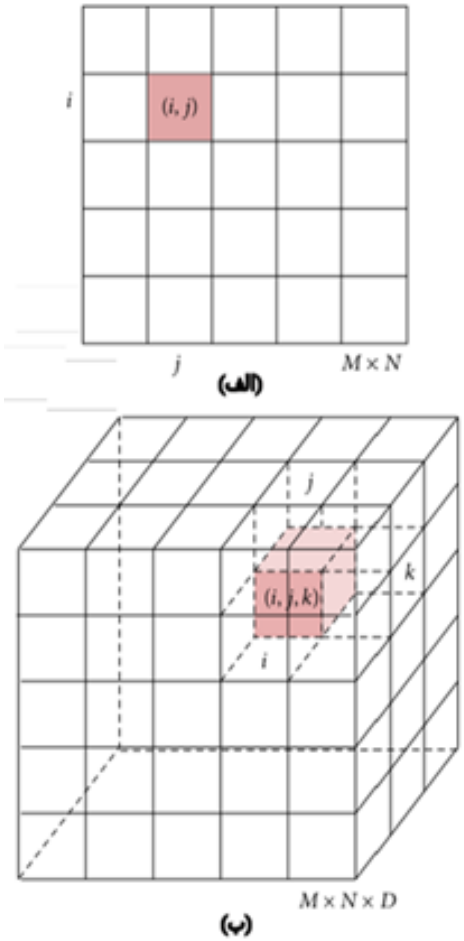

Fig. 1.6 Template matching using human image features

Intensity Distribution in MRI Images

Tissue intensity distribution serves as a primary feature for MRI image segmentation algorithms. However, raw intensity values are frequently corrupted by acquisition artifacts such as Rician noise, partial volume effects (PVE), and spatial magnetic field inhomogeneities (bias fields). Uncorrected artifacts degrade intensity-based segmentation performance. Consequently, standard preprocessing workflows—such as spatial co-registration and bias field correction—are essential prior to automated segmentation.

Standard MRI Pulse Sequences

MRI systems capture rich physical signals that yield valuable structural and functional information. Common imaging sequences include:

- **T_1-Weighted MRI**: A fundamental MRI sequence wherein fluid (water) appears hypointense (dark) relative to hyperintense (bright) adipose tissue.
- **T_2-Weighted MRI**: A primary sequence wherein fluid structures (e.g., cerebrospinal fluid) appear hyperintense (bright) while fat exhibits lower relative intensity.

Comparison Between MRI and Computed Tomography (CT)

Compared to CT imaging, MRI exhibits distinct diagnostic advantages and limitations:

Advantages of MRI over CT

- Superior soft-tissue contrast resolution
- Multiplanar imaging capabilities along arbitrary geometric planes
- Complete absence of ionizing radiation
- Freedom from beam-hardening artifacts typical of CT

Limitations of MRI relative to CT

- Higher operational costs, reduced availability, and greater technical complexity
- Longer acquisition times, increasing susceptibility to motion artifacts
- Lower spatial resolution for cortical bone and calcified structures
- Contraindications in patients with ferromagnetic metallic implants or pacemakers
- Sensitivity to involuntary motion (respiration, cardiac motion, peristalsis)
- Challenges in claustrophobic or dyspneic patients due to the enclosed scanner bore

Image Feature Extraction

Image features represent distinctive spatial and statistical characteristics used to segment target anatomical structures. Segmentation accuracy depends heavily on selecting and extracting highly informative feature sets [4].

Statistical approaches are widely used for feature extraction and pattern classification in MRI. Patterns are described using feature vectors defined in a multidimensional feature space. First-order statistical features are derived directly from the intensity histogram (e.g., mean, median, variance, skewness). Although first-order features omit spatial distribution patterns, they are frequently combined with second-order descriptors.

Second-order statistical features describe image texture and spatial inter-pixel relationships, typically computed via Gray-Level Co-occurrence Matrices (GLCM). First- and second-order metrics are collectively termed appearance features because they capture local intensity values and spatial interactions.

Pixel-level intensity segmentation (first-order features) is effective only when target structures exhibit distinct intensity separation from the background. In noisy or artifact-laden medical images, second-order spatial features are necessary to achieve robust segmentation performance.

1.4 Summary

In this chapter, we introduced the fundamental concepts surrounding the semantic segmentation of breast cancer pathology images and highlighted the core research challenges addressed in this thesis. We reviewed fundamental definitions of image segmentation, outlined modalities used in breast cancer imaging, and discussed the role of deep learning in medical image analysis. The theoretical concepts presented here serve as a foundational reference for the technical chapters that follow.

Chapter 2

Literature Review

2.1 Introduction

Numerous factors contribute to the onset of breast cancer, which can generally be categorized into three groups: unmodifiable factors, modifiable factors, and other risk factors [25]. Unmodifiable factors include age, gender, family history, late menopause, and genetic predisposition [14]. Beyond female gender, advancing age represents the most significant risk factor for breast cancer; its incidence increases significantly in women after age 35 and reaches a plateau after age 80 [3]. Breast cancer is rarely observed in women younger than 20 years, increases substantially from the third through the mid-eighth decades of life, and subsequently declines significantly. Women, like men with a family history of breast cancer—particularly those with first-degree relatives (mother, sister, daughter, father, brother)—face an elevated risk of developing breast cancer [11]. The risk of breast cancer for women with one first-degree female relative diagnosed with the disease increases by 1.8-fold compared to those without a family history, 3-fold with two first-degree relatives, and 4-fold with three or more first-degree relatives. It should be noted that the majority of women with one or more first-degree relatives never develop cancer, and most women who develop cancer report no family history of the disease. Furthermore, a family history of ovarian cancer is also associated with an increased risk of breast cancer in both men and women. Women with a family history of breast or ovarian cancer should consult their physician, as a genetic predisposition to cancer may be present.

In mammography-based breast cancer diagnosis, limitations such as imaging noise, low resolution during image acquisition, and the presence of minute cancerous patterns throughout the image hinder accurate detection [22]. Moreover, previous diagnostic methods suffer from uncertain inter-class boundaries, where most classification algorithm errors occur. By constraining these ambiguous decision boundaries, the detection rate of breast cancer can

be enhanced [13]. The objective of this research is to evaluate flexible tools for the early detection of breast cancer in mammographic images.

2.2 Cancer

Cancer is a general term for a group of diseases characterized by the uncontrolled division of abnormal cells that can invade surrounding tissues. Cancerous cells can spread to other parts of the body through the bloodstream and the lymphatic system. Cancer is not a single disease, but rather a collection of related diseases. More than one hundred distinct types of cancer exist. Most cancers are named after the organ or cell type in which they originate; for example, cancer originating in the large intestine is termed colorectal cancer.

2.2.1 Classification of Cancer

- **Carcinoma**: Cancers originating in the skin or in tissues that line or cover internal organs.
- **Sarcoma**: Cancers arising in bone, cartilage, fat, muscle, blood vessels, or other connective or supportive tissues.
- **Leukemia**: Cancers initiating in blood-forming tissues such as bone marrow, causing large numbers of abnormal blood cells to be produced and enter the bloodstream.
- **Lymphoma and Myeloma**: Cancers originating in the cells of the immune system.
- **Central Nervous System Cancers**: Cancers initiating in the tissues of the brain and spinal cord.

2.2.2 Origin and Pathophysiology of Cancer

All cancers originate at the cellular level. The human body is composed of trillions of cells that grow and divide in a controlled manner to produce new cells as needed for physiological maintenance. When cells become old or damaged, they undergo programmed cell death and are replaced by new cells. However, this orderly process can be disrupted. A cell's genetic material (DNA) may become damaged or mutated, altering normal cell growth and division. When this occurs, cells do not die when they should, and unnecessary new cells form continuously. These excess cells accumulate to form a tissue mass known as a tumor.

Not all tumors are cancerous; tumors can be classified as benign or malignant. Benign tumors are non-cancerous, can typically be removed surgically without recurring, and do not

invade adjacent tissues or metastasize to distant organs. Malignant tumors are cancerous; their constituent cells can invade surrounding tissues and metastasize throughout the body. The spread of cancer from a primary site to secondary distant sites is termed metastasis. Certain types of cancer, such as leukemias (cancers of the blood and bone marrow), do not form solid tumors.

The goal of cancer control programs is to reduce disease incidence and mortality while improving the quality of life for cancer patients [11]. Achieving this objective relies on prevention, early detection and screening, timely therapeutic intervention, and palliative or supportive care [25]. Etiological factors in cancer development fall into three primary categories: physical agents (e.g., ultraviolet radiation and ionizing solar radiation), chemical agents (e.g., asbestos, tobacco compounds, and aflatoxins in nuts), and biological agents (e.g., specific viruses, bacteria, and parasites). Age is also a major risk factor; as age increases, the probability of cancer incidence rises. Notable biological carcinogens include Hepatitis B and C viruses (hepatocellular carcinoma), Human Immunodeficiency Virus (Kaposi's sarcoma), Human Papillomavirus (cervical carcinoma), and *Helicobacter pylori* (gastric carcinoma).

2.3 Overview of Breast Anatomy and Physiology

The breasts are secondary sexual characteristics, and reproduction can occur independently of them. The non-lactating breast weighs an average of 150 to 250 grams, whereas the lactating breast weighs between 400 and 500 grams. The breast parenchyma consists of ductal, lobular, and epithelial structures. The breast stroma is composed of fibrous and adipose tissue [14]. The upper outer quadrant of the breast contains a higher proportion of glandular tissue and represents the most common site of tumor occurrence.

The breast is divided into 15 to 20 sections called lobes, each of which is further subdivided into smaller units called lobules. These lobes and lobules are interconnected by small conduits termed lactiferous ducts. Breast tissue comprises lobes, ducts, adipose tissue, connective tissue, muscle structures, and lymphatic networks. Malignant transformations can originate in any component of the breast tissue. Breast cancer refers to the uncontrolled growth and invasive proliferation of abnormal cells within breast tissue, most commonly in the lactiferous ducts (which convey milk to the nipple) or lobules (glandular milk-producing units) (Figure 2.1).

The nipple is situated at the fourth intercostal space and is encircled by a pigmented circular region known as the areola. Estrogen increases areolar pigmentation across all age groups. The areolar epithelium contains fine hairs, sebaceous glands, and sweat glands.

Sebaceous glands (glands of Montgomery) lubricate and enlarge the nipple during pregnancy and lactation.

The primary arterial blood supply to breast tissue is derived from the internal thoracic artery and the lateral thoracic artery, which together form an extensive anastomotic network. Venous drainage mirrors the arterial supply. Innervation of the breast arises from the anterior and lateral cutaneous branches of the fourth through sixth intercostal nerves.

Lymphatic pathways follow the venous drainage. Lymphatic drainage routes and lymph nodes are critically important in breast cancer diagnosis and treatment due to their role in mapping metastatic pathways. Three distinct types of lymphatic drainage exist in the breast:

- Superficial cutaneous drainage
- Subareolar lymphatic drainage (draining the nipple and areolar complex)
- Deep glandular lymphatic drainage (draining the underlying parenchyma)

The primary physiological function of the female breast is milk secretion for infant nutrition. Breast development typically initiates between 9 and 16 years of age and progresses over a period of 4 to 5 years. Estrogen is primarily responsible for ductal growth, whereas progesterone promotes lobular maturation.

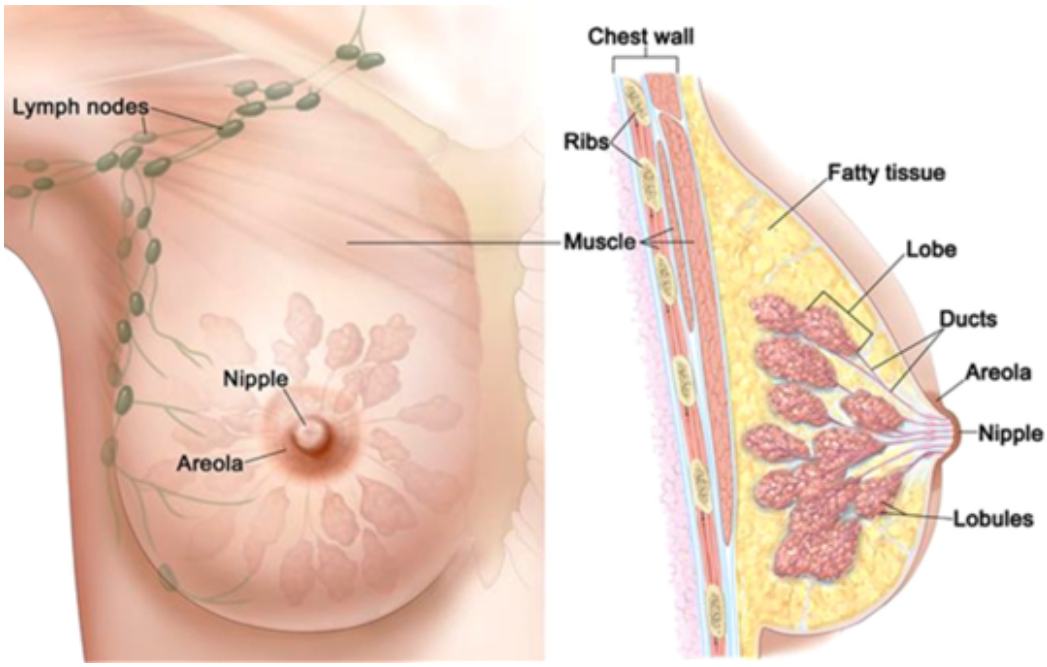

Fig. 2.1 Anatomy of the breast.

2.4 Breast Cancer

Breast cancer is among the most prevalent malignancies in women [3]. Understanding its key characteristics is essential for all women due to its high incidence and clinical significance [20]. Most detected breast masses are benign, and breast cancer treatment does not invariably necessitate total mastectomy; early-stage detection with modern therapeutic protocols offers high rates of complete recovery [11]. Breast cancer represents the leading cause of mortality among women aged 35 to 55 years, with 1 to 2 new cases diagnosed annually per 1,000 women. Breast tissue lies adjacent to surrounding pectoral musculature, and understanding the lymphatic system is essential for accurate staging and management. Within tumor pathology, the lymphatic network serves as a pathway for systemic dissemination. Internal and central regions of the breast are vascularized by specific arterial branches, and the organ possesses a defined nerve network.

2.4.1 Risk Factors in Breast Cancer

Breast cancer is the most common malignancy in women and accounts for more cancer-related deaths in females than any other tumor type [25]. Its lifetime incidence in the United States is approximately 1 in 9 among White women and 1 in 14 among Black women [3]. This disease is rare before age 30, but its incidence increases sharply post-menopause. Various histological subtypes of breast cancer exist, but all typically present initially as a palpable lump, localized induration, or structural breast asymmetry. Immediate medical evaluation is required upon identifying any such changes.

Following a physical examination, physicians routinely order mammography to help characterize the lesion [22]. In certain cases, fine-needle aspiration (FNA) or core biopsy is recommended. If a mass appears to be a simple cyst, it is often monitored conservatively; if it does not recur, no further intervention is required. Surgical biopsy is frequently advised, as it remains the gold standard for establishing a definitive histological diagnosis. At this stage, patient management follows either a one-stage approach (where the tumor is excised immediately if the biopsy reveals malignancy) or a two-stage approach (where biopsy is followed by definitive surgery a week or more later). Short delays in the two-stage protocol carry no known clinical disadvantage, allowing patients to consider treatment options and seek secondary surgical opinions. Most surgeons prefer the two-stage approach, as it allows time for supplementary diagnostic tests to assess disease extent and determine tumor receptor status for targeted therapies. Notably, approximately four-fifths of evaluated breast lumps prove to be non-cancerous (benign).

2.5 Symptoms of Breast Cancer

Breast cancer is a malignancy in which abnormal cells proliferate uncontrollably within one or both breasts. These cells can invade adjacent stroma and coalesce into a discrete mass or tumor. Furthermore, malignant cells can metastasize through blood and lymphatic channels to regional lymph nodes and distant organs. Common metastatic sites include the lungs, bones, liver, and adrenal glands; brain, spleen, and pituitary involvement occur less frequently, though no organ is exempt. Metastases may manifest years after primary tumor control. Breast cancer affects the left breast slightly more often than the right breast.

In most instances, the initial presenting symptom of breast cancer is a palpable mass within the breast, most commonly located in the upper outer quadrant (the region adjacent to the axilla) [14]. The mass is typically painless and may exhibit either rapid or indolent growth kinetics. Other clinical signs and symptoms of breast cancer include:

2.5.1 Breast Mass or Tumor

A palpable mass is the most common presenting sign of breast cancer, frequently discovered incidentally by the patient or during routine clinical examination by a physician [14]. Although usually painless, a mass may occasionally be tender. Because masses occur in both benign conditions and malignancies, any new palpable breast lump warrants prompt clinical evaluation. In certain presentations, breast cancer manifests as multiple discrete masses. On palpation, suspicious malignant masses typically feel firm or hard, ill-defined, immovable, and fixed to surrounding tissues. As noted previously, most suspicious masses present as solitary, painless lesions. Final judgment regarding benign versus malignant status rests with the treating physician following comprehensive diagnostic evaluation.

2.5.2 Nipple Discharge

Nipple discharge is most commonly associated with benign breast disorders. Although rarely indicative of underlying malignancy, detailed evaluation is necessary to rule out cancer. Bilateral discharge originating from multiple ducts typically reflects hormonal imbalances, pharmacological side effects, fibrocystic changes, or simple cysts.

Nipple discharge presents in various colors (white, brown, blue, gray, red, green, or clear). Discharge associated with breast cancer is characteristically frank bloody, serosanguinous, or water-clear. Spontaneous, unilateral, single-duct discharge—particularly when accompanied by a palpable mass—substantially raises the clinical suspicion of malignancy. Features of high-risk nipple discharge include:

- Bloody or water-clear consistency
- Spontaneous discharge occurring without manual expression
- Unilateral origin restricted to a single duct orifice
- Concomitant association with a palpable breast mass

2.5.3 Skin Changes Over the Breast

Pathological skin alterations manifest as localized retraction, dimpling, or skin tethering. Skin dimpling may be persistent or visible only during specific physical examination maneuvers. Although skin retraction is a significant finding, it is not pathognomonic for cancer and can occur secondary to benign breast conditions.

Adhesion of a malignant tumor to the overlying dermis causes structural breast distortion. Asymmetry or distortion of the nipple-areolar complex may similarly indicate an underlying malignant lesion. Tumor invasion into the skin can result in frank ulceration. Furthermore, cutaneous edema resulting in a classic "peau d'orange" (orange peel) appearance indicates advanced localized disease.

2.5.4 Nipple Changes

Nipple inversion or retraction may develop naturally following lactation or puberty in some individuals, often reversing spontaneously or with manual manipulation. In contrast, malignant nipple retraction is fixed and non-reducible, frequently associated with a palpable underlying subareolar mass.

Paget's disease of the breast is a specific form of ductal carcinoma that presents as eczematous skin changes of the nipple-areolar complex. Symptoms include persistent pruritus, burning, erythema, induration, and scaling of the nipple. Without treatment, lesions gradually spread to the surrounding areola and skin. Paget's disease is frequently misdiagnosed as benign eczema or dermatitis, leading to prolonged topical steroid treatment and delayed cancer management. Therefore, persistent cutaneous nipple changes must raise suspicion of malignancy until proven otherwise by biopsy.

2.5.5 Axillary Lymphadenopathy

Malignant cells possess the capacity to disseminate via lymphatic channels to regional nodes. The axillary lymph node basin represents the primary pathway for regional breast cancer metastasis. Axillary lymphadenopathy may present alongside a palpable primary breast

mass or as an isolated clinical finding without a palpable lesion. In occult cases, diagnostic mammography is instrumental in identifying the primary breast tumor responsible for axillary nodal enlargement [22]. Importantly, axillary lymphadenopathy can also result from benign conditions, localized inflammation, or upper extremity trauma; thus, while nodal enlargement mandates diagnostic workup, it is not definitively indicative of malignancy.

2.5.6 Asymmetry and Changes in Breast Size

Rapid enlargement of a malignant tumor or marked peritumoral edema can cause significant, newly acquired asymmetry between the breasts. While minor size variations between breasts are normal in healthy individuals, recent-onset unilateral breast enlargement warrants careful clinical evaluation to exclude underlying masses. Benign tumors, acute trauma, hematomas, lactational mastitis/abscesses, and large simple cysts can also cause acute breast asymmetry.

2.5.7 Distant Metastasis

In rare presentations, breast cancer manifests initially with systemic symptoms attributable to distant metastases without overt localized breast signs. These symptoms may include localized bone pain (skeletal metastases), dyspnea (pulmonary involvement), or jaundice and hepatomegaly (hepatic metastases). Comprehensive physical examination and diagnostic imaging are necessary in these scenarios to identify the occult primary breast lesion.

2.6 Differentiation Between Benign and Malignant Tumors

Both benign and malignant tumors exhibit abnormal, accelerated cellular proliferation. The fundamental distinction lies in their growth regulation: cellular proliferation in benign tumors is self-limiting and remains localized, whereas in malignant tumors (cancer), cell division proceeds unchecked and uninhibited. Left untreated, malignant growth invades local structures, metastasizes systemically, and ultimately compromises organ function. This destructive behavior never occurs with benign tumors.

Breast cancer is the most common malignancy affecting women worldwide [3]. In Western nations, it is predominantly diagnosed in women over 50 years of age, whereas in developing regions, patients often present at younger ages and at more advanced stages due to limited awareness and screening access [20]. Because early-stage detection significantly improves treatment efficacy and overall survival [11], public health education regarding breast cancer signs and screening guidelines is vital.

2.7 Early Detection of Breast Cancer

The World Health Organization (WHO) and the American Cancer Society (ACS) emphasize that early detection is the most effective strategy for controlling breast cancer [3]. Because breast cancer typically undergoes a prolonged preclinical phase (spanning 8 to 10 years), identifying non-palpable lesions during early stages can prevent premature mortality. In regions where up to 50% of breast cancer patients present with advanced-stage disease, implementing routine screening protocols and encouraging clinical participation are critical health priorities [20, 25].

2.8 Breast Cancer Screening Modalities

Screening refers to the systematic application of diagnostic tests across asymptomatic at-risk populations to detect subclinical disease prior to overt symptom onset. The ACS recommends the following diagnostic modalities for early detection of breast cancer in asymptomatic women [3]:

- Breast Self-Examination (BSE)
- Clinical Breast Examination (CBE)
- Mammography
- Ultrasonography
- Magnetic Resonance Imaging (MRI)

2.8.1 Breast Self-Examination (BSE)

Women should be familiar with the normal appearance and feel of their breasts to detect any unusual structural alterations and report them promptly to a healthcare provider. Although monthly BSE is no longer formally recommended by the ACS as a primary screening tool due to its limited impact on overall mortality [3], breast self-awareness remains beneficial. Palpable breast nodularity is not inherently pathological and may fluctuate naturally across the menstrual cycle.

2.8.2 Clinical Breast Examination (CBE)

For asymptomatic women in their 20s and 30s, a clinical breast exam is recommended at least once every three years as part of routine health maintenance. For women aged 40

and older, annual CBE serves as a valuable adjunct to screening mammography, helping identify mammographically occult cancers [3]. Ideally, CBE should be performed shortly before the annual mammogram. During CBE, the clinician systematically palpates the breast parenchyma and axillary vaults using the finger pads, assessing tissue consistency, focal nodularity, mobility, and depth attachment. Combining CBE with mammography enhances diagnostic sensitivity compared to mammography alone, albeit with a slight increase in false-positive rates.

2.8.3 Magnetic Resonance Imaging (MRI) in Breast Evaluation

Magnetic Resonance Imaging (MRI) utilizes strong magnetic fields and radiofrequency pulses to generate cross-sectional images based on nuclear magnetic resonance properties of hydrogen protons. Pulse sequence parameters can be tailored to yield specific soft-tissue contrasts. Spatial encoding along orthogonal axes is accomplished using gradient fields, and raw k-space signals are converted into spatial images via Fourier transformation.

In clinical oncology, dynamic contrast-enhanced MRI serves as a highly sensitive, non-invasive imaging modality for evaluating complex breast lesions [9]. Indicated scenarios for breast MRI include:

- Evaluation of breast architecture in patients with breast implants, where conventional mammography or ultrasound is limited
- Diagnostic assessment in post-surgical or post-radiation breasts where scarring reduces mammographic sensitivity
- Resolution of ambiguous or inconclusive findings on mammography and ultrasound [9]
- Surveillance for local recurrence in patients previously treated for breast cancer when conventional modalities are equivocal

Mass Detection in Mammograms

A mass is defined in mammographic imaging as a space-occupying lesion visible in at least two orthogonal projections [22]. Masses are characterized by their geometric shape and margin characteristics. Regularly shaped masses (e.g., round or oval) carry a higher probability of being benign, whereas irregular, spiculated masses strongly suggest malignancy [13, 12]. Automated mass detection frameworks typically execute two sequential computational phases:

1. Identification of suspicious Candidate Regions of Interest (ROIs)

2. Classification of candidate ROIs as true masses or normal parenchyma

Detection algorithms for candidate ROI extraction operate on either a pixel-level or region-level basis. In pixel-based approaches, feature vectors are computed for individual pixels and classified using statistical or machine learning models [1, 15]. For instance, Laws' texture energy measures and local edge descriptors extracted from ROIs have been combined with binary decision trees to classify spiculated lesions against normal tissue, achieving high sensitivity. Multiresolution wavelet frameworks decompose mammographic images across spatial frequency scales to extract multiscale pixel features for detecting subtle spiculated lesions within public benchmark databases [13].

Detection of Architectural Distortion

Architectural distortion is defined in the ACR BI-RADS lexicon as a condition where "the normal tissue architecture is distorted with no definite mass visible. This includes thin spicules radiating from a point and focal retraction or distortion at the edge of the parenchyma."

Architectural distortion is the third most common mammographic sign of non-palpable breast cancer; however, due to its subtle appearance, it is frequently missed during prospective screening, accounting for 12% to 45% of missed or misinterpreted breast cancers on mammograms [22]. While extensive literature focuses on microcalcifications and mass detection, fewer automated algorithms have been developed specifically to target architectural distortion.

Image Enhancement Techniques in Breast Cancer Detection

Mammographic images inherently exhibit low contrast and subtle detail visibility [22]. Image enhancement algorithms aim to improve image quality and visual contrast, transforming raw mammograms into representations better suited for diagnostic interpretation or automated analysis [12]. Enhancement methodologies fall into two categories: direct contrast enhancement and indirect contrast enhancement.

In direct contrast methods, contrast metrics are defined explicitly and manipulated directly. For example, local contrast enhancement algorithms define central and surrounding window neighborhoods around each pixel, computing local contrast as the intensity difference between these regions. Modified contrast values are then mapped back using specific nonlinear transformation functions to yield enhanced pixel values. Various spatial-domain and

transform-domain contrast enhancement algorithms have been applied to mammographic screening.

2.8.4 Ultrasonography

The primary clinical application of breast ultrasound is the diagnostic evaluation of suspicious lesions identified during screening mammography or physical examination [21]. Studies indicate that supplemental ultrasound combined with mammography improves cancer detection rates in women with dense breast tissue, albeit with an increased false-positive rate [9]. Ultrasound is not recommended as a standalone screening tool for average-risk populations.

2.9 Overview of Automated Mass Segmentation Techniques

In mammographic image analysis, a segmentation algorithm delineates target structures, such as the breast boundary, pectoral muscle, or focal abnormalities (microcalcifications and masses) [22]. Segmenting soft-tissue masses is technically more challenging than segmenting microcalcifications because masses often exhibit ill-defined boundaries that blend into surrounding dense parenchymal tissue [12].

Mass segmentation algorithms vary depending on whether they target specific morphologies (e.g., circumscribed vs. spiculated masses) or utilize multi-view information (e.g., comparing MLO and CC views, comparing bilateral left and right breasts, or analyzing temporal mammographic sequences). Image segmentation partitions an image into constituent regions to facilitate quantitative feature extraction. The accuracy of downstream feature extraction and classification depends directly on segmentation quality. Segmentation techniques generally follow two primary paradigms: region-homogeneity approaches (grouping pixels with similar intensity properties) and boundary-detection approaches (identifying intensity discontinuities across boundaries).

2.9.1 Region Growing Methods

Region growing is a region-based segmentation technique that exploits spatial homogeneity. The process begins by selecting one or more seed points within the target structure. Pixel similarity criteria—typically based on absolute gray-level intensity differences relative to a threshold—are defined. Neighboring pixels surrounding the seeds are evaluated iteratively; if a neighbor satisfies the similarity criterion, it is subsumed into the growing region. This process continues iteratively until no further pixels meet the inclusion criteria.

2.9.2 Two-Dimensional Discrete Wavelet Transform (2D DWT)

Wavelet analysis is a mathematical framework for multiresolution signal processing. Unlike Fourier analysis, which uses infinite sinusoidal basis functions, wavelet transforms expand signals using localized wavelets generated via translations and dilations of a mother wavelet (ψ). Wavelets provide simultaneous time-frequency (or spatial-frequency) localization, yielding improved numerical stability and multi-scale feature representation.

In medical image processing, the 2D Discrete Wavelet Transform (2D DWT) is widely used for noise reduction, texture analysis, and multi-scale feature extraction [13]. Applications in medical imaging include brain tissue segmentation in MRI, automated microcalcification cluster detection, dynamic MR spectroscopy analysis, and functional neuroimaging.

2.9.3 Watershed Segmentation Algorithm

Watershed segmentation models a grayscale image as a topographic surface, where pixel intensity values correspond to spatial elevations. Gradient magnitude images are frequently used as the topological surface, where high gradient values correspond to watershed lines (boundaries). Watershed algorithms segment images by identifying catchment basins and ridge lines, making them effective for separating touching or overlapping structures in medical images.

2.9.4 Linde-Buzo-Gray (LBG) Vector Quantization

The Linde-Buzo-Gray (LBG) algorithm is a vector quantization technique used for clustering multidimensional data. The algorithm iteratively partitions vector space into codebook vectors ($C_1, C_2, \ldots, C_k$). Initial centroids are split by applying a small perturbation vector (ε), generating new reference vectors (V_1, V_2). Input training vectors are assigned to the nearest centroid based on Euclidean distance metrics, and centroids are recomputed. This process repeats iteratively until the desired codebook size is achieved.

Figure 2.2 illustrates two-dimensional space partitioning using the LBG algorithm. When applied to medical image segmentation, vector quantization partitions image features into distinct clusters, producing segmented regions based on underlying intensity and texture characteristics [5].

2.9.5 Markov Random Fields (MRF)

Markov Random Field (MRF) models capture spatial contextual dependencies among neighboring pixels in image processing [5, 19]. By combining local neighborhood system configu-

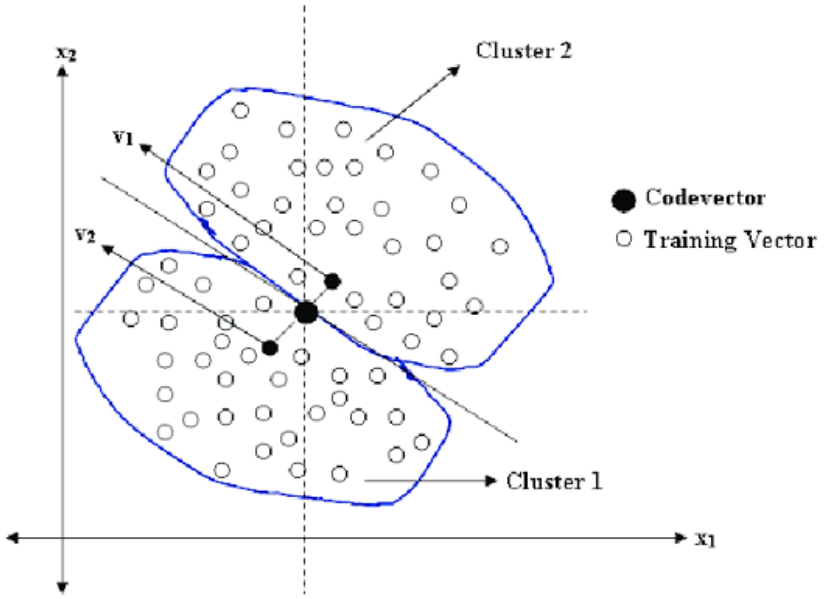

Fig. 2.2 Two-dimensional representation of the LBG algorithm.

rations with Gibbs energy distributions, MRF frameworks formulate image segmentation as an energy minimization problem.

A spatial random process X satisfies the Markov property if the conditional probability of a pixel state given all other pixels depends solely on its immediate local neighborhood system ($\mathcal{N}_i$):

$$P(X_i = x_i \mid X_{S\setminus\{i\}} = x_{S\setminus\{i\}}) = P(X_i = x_i \mid X_{\mathcal{N}_i} = x_{\mathcal{N}_i}) \tag{2.1}$$

Minimizing the associated posterior energy function using optimization algorithms (e.g., Simulated Annealing or Iterated Conditional Modes) yields optimal, contextually smooth image segmentations [19].

2.9.6 Gray-Level Co-occurrence Matrix (GLCM)

Haralick introduced the Gray-Level Co-occurrence Matrix (GLCM) to quantify second-order spatial texture characteristics [4]. GLCM elements $P(i, j \mid d, \theta)$ represent the joint probability of two pixels separated by a spatial displacement distance d along direction θ having gray-level intensities i and j, respectively.

GLCM matrices can be calculated across localized sliding windows to yield dense pixel-level texture feature maps. Common Haralick texture descriptors extracted from GLCM include Angular Second Moment (Energy), Contrast, Correlation, Variance, Inverse Difference Moment (Homogeneity), and Entropy [12]. These descriptors quantify textural properties in mammographic tissue, assisting in automated lesion discrimination [4, 12].

2.10 Application of Image Processing in Breast Cancer Detection

Computer-Aided Diagnosis (CAD) systems utilizing digital image processing and deep learning techniques provide objective secondary evaluations to assist radiologists in detecting breast lesions [22, 8]. Automated lesion detection algorithms typically follow a multi-stage pipeline [28, 16]:

1. Image Preprocessing (Denoising, contrast enhancement, and artifact removal)
2. Region of Interest (ROI) Segmentation (Delineation of candidate mass/lesion boundaries)
3. Feature Extraction and Classification (Discriminating between benign tissue, malignant lesions, and normal parenchyma) [27, 2]

Figure 2.3 illustrates a standard computer-aided image processing workflow for breast cancer detection.

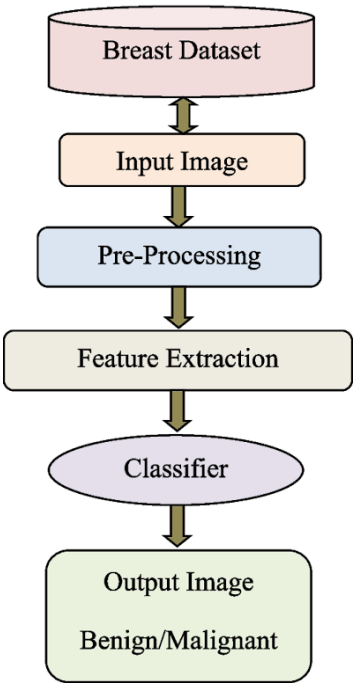

Fig. 2.3 General flowchart of using image processing in breast cancer detection.

2.11 Denoising Techniques for Medical Images

Medical imaging modalities capture internal anatomical structures in 2D and 3D representations. Raw acquired images frequently contain unwanted noise introduced by sensor hardware or transmission channels, requiring specialized pre-filtering prior to feature extraction and segmentation [12].

Standard digital image enhancement workflows encompass:

- Feature extraction and contrast enhancement for target lesion delineation [4]
- 3D volumetric reconstruction and multiplanar interpolation
- Denoising and artifact reduction to improve Signal-to-Noise Ratio (SNR)

Noise reduction filters must preserve structural edges and diagnostic detail. Linear low-pass filters reduce high-frequency noise but tend to blur critical anatomical boundaries. Non-linear spatial filters (e.g., median filtering and anisotropic diffusion) effectively suppress impulse noise while preserving edge sharpness [12]. However, when noise density exceeds high thresholds (e.g., >60%), standard median filters may struggle to differentiate noisy pixels from true signal, leading to loss of fine structural details.

2.11.1 Mean Filter

The spatial mean filter is a linear smoothing filter that replaces each pixel value with the arithmetic average of its local neighborhood. While effective at smoothing additive Gaussian noise, it introduces spatial blurring across image edges and high-frequency details. Because it applies uniform weighting across all pixels in the filter window regardless of whether a pixel is noisy or clean, image sharpness is degraded.

2.11.2 Median Filter

The median filter is a non-linear order-statistic filter. For each pixel, intensity values within an $m \times n$ neighborhood window are sorted in numerical order, and the median value is assigned to the center pixel. Median filtering effectively eliminates salt-and-pepper impulse noise while preserving sharp intensity step-edges.

In mammographic image processing, median filters with mask dimensions such as 5×5 are applied to suppress digital noise. Increasing the mask window size enhances noise suppression but can lead to loss of fine structural edges and subtle microcalcifications.

2.12 Summary

Early detection of breast cancer plays a vital role in reducing mortality rates among women [11, 3]. However, inherent limitations in mammographic images—such as low contrast, imaging artifacts, and dense parenchymal overlap—complicate automated lesion identification [22]. Over recent years, various image enhancement, segmentation, and machine learning techniques have been developed to address these diagnostic challenges [28, 16, 27]. This chapter presented a literature review of breast cancer risk factors, anatomical characteristics, screening modalities, and image processing algorithms used for automated mass detection and segmentation.

Chapter 3

Fundamentals of Deep Learning Concepts

3.1 Introduction

Deep learning refers to the construction of machine learning models designed to learn hierarchical representations of data [7, 6]. Deep neural networks comprise a broad family of neural network architectures with multiple layers, demonstrating how networks with a large number of layers can successfully construct the representational structures required for deep learning. Both supervised and unsupervised feature learning algorithms can be utilized to adjust the connection weights within these networks [18].

Indeed, various types of deep neural networks exist, including Convolutional Neural Networks (CNNs), Autoencoder Networks, Recurrent Neural Networks (RNNs), and Recursive Neural Networks [7]. Various combinations of these networks—such as combining convolutional neural networks with recurrent neural networks—can also be employed depending on the nature of the input space in the problem under investigation [6]. These networks are formed by stacking multiple non-linear transformations, aiming to achieve higher levels of abstraction and ultimately useful representations of existing data. As the number of layers in a neural network increases, the optimization problem becomes more complex. Consequently, one method for training these networks involves using unsupervised, layer-wise greedy pre-training algorithms [18, 6]. In this approach, each layer is initially trained individually, followed by fine-tuning across the entire integrated network.

3.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) were conceptualized and developed inspired by the information processing system of the biological brain [17]. Through a learning process and

using processing units called neurons, ANNs attempt to identify intrinsic relationships within data to establish a mapping between the input space (input layer) and the desired space (output layer). Hidden layer(s) process the information received from the input layer and pass it to the output layer. An example of a three-layer network architecture consisting of an input layer, a hidden layer, and an output layer is illustrated in Figure 3.1.

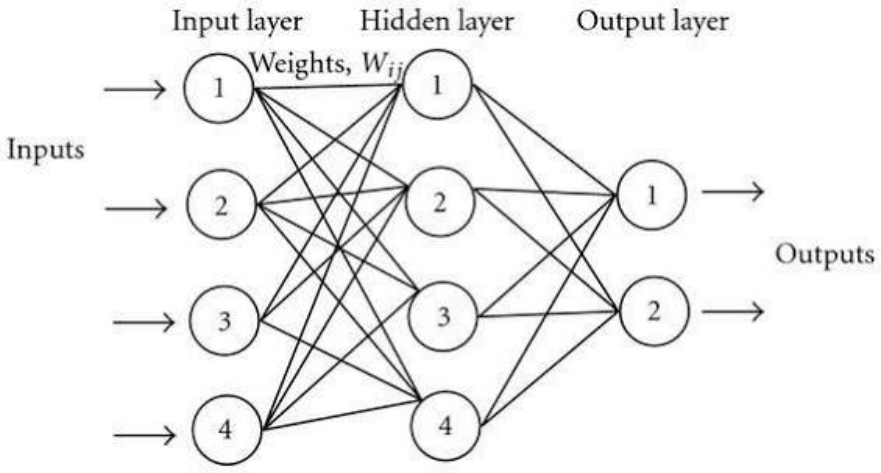

Fig. 3.1 Overview of a three-layer artificial neural network.

Based on the structure of artificial neural networks, their primary characteristics include high processing speed, pattern learning capability, pattern generalization ability after learning, robustness against unexpected errors, and the absence of significant disruption if a subset of connections fails—owing to the distributed nature of the network's weights [17]. The first practical application of artificial neural networks was realized with the introduction of Multilayer Perceptron (MLP) networks. The Backpropagation (BP) algorithm is commonly used to train these networks. The foundation of this algorithm rests on the error-correction learning rule, consisting of two main forward and backward passes. Parameter adjustments in a multilayer perceptron network are driven by the error signal and the input signal.

Determining the number of layers and the number of neurons within them represents one of the most critical challenges in artificial neural network simulation. Sigmoid and hyperbolic tangent activation functions are cited as the most common activation functions used in backpropagation networks. During the training of an MLP network using the BP learning algorithm, calculations initially proceed from the input toward the output layer; subsequently, calculated error values are propagated back through the preceding layers. Layer-by-layer output computation occurs first, where the output of each layer serves as the input to the subsequent layer. The training steps using this algorithm are as follows:

1. Assigning a weight matrix to each connection,
2. Selecting an input vector and its corresponding target output vector,
3. Computing neuron outputs in each layer, leading to the calculation of neuron outputs at the output layer,
4. Updating weights by propagating the network error back to previous layers, where said error stems from the discrepancy between the target output and the computed output,
5. Evaluating the performance of the trained network using defined metrics such as Root Mean Square Error (RMSE), and returning to step 2,
6. Terminating training upon convergence.

3.3 Deep Learning

Deep learning was first introduced as a major subfield of machine learning closely tied to deep neural networks [7]. Neural networks are capable of training on and classifying datasets in both supervised and unsupervised manners [6, 18]. Deep learning is a branch of machine learning built upon a foundational set of representation learning algorithms. Perhaps the single greatest advantage of deep learning lies in its capacity for automated feature extraction and feature learning [6]. Most successful deep learning methods—such as deep neural networks, convolutional neural networks, deep belief networks, recurrent neural networks, and stacked autoencoders—are built upon artificial neural networks [7]. In recent years, these methods have been continuously applied across diverse domains such as image processing, computer vision, natural language processing, and medical diagnosis, delivering exceptional results [23, 10, 26, 16]. Numerous studies demonstrate that deep learning significantly outperforms traditional methods in these domains [21, 8].

3.4 Distinction Between Deep Learning and Neural Networks

When using the term "deep learning," we generally refer to Deep Neural Networks (DNNs) [7]. To be more precise, the distinction between deep learning and traditional neural networks is that deep learning encompasses a broader scope, also incorporating deep reinforcement learning algorithms and unsupervised hierarchical feature extraction [6, 18]. Thus, the

distinction between neural networks and deep learning is clear, and they should not be confused. A fundamental question arises: what types of problems can be solved using deep learning? To answer this, two preliminary questions must be addressed:

1. What classifications matter to us? For instance, in email filtering, determining whether an email is spam or non-spam is essential. In customer relationship management systems, classifying customer satisfaction states (satisfied vs. dissatisfied) may serve as target classes [10, 26].

2. Do we possess labeled data to train the system? Specifically, are we capable of acquiring or constructing a labeled dataset? Yes, using current data collection and annotation techniques.

3.5 Deep Convolutional Neural Networks

Convolutional Neural Networks (CNNs) represent one of the most important deep learning methodologies, in which multiple layers are trained using powerful optimization algorithms [7, 24]. This approach is highly efficient and stands as one of the most prevalent techniques across various computer vision applications. The primary advantage of CNNs is their ability to extract spatial feature hierarchies from images automatically using deep learning principles [6]. Due to this crucial advantage, the application of CNNs for disease diagnosis in medical domain applications has attracted widespread research interest [28, 16, 21].

CNNs are specialized artificial neural network models composed of neurons, layers, and trainable weights, akin to conventional neural networks. The key distinction of CNNs lies in their specialized capability for deep spatial learning in tasks such as image recognition, audio analysis, and handwriting recognition [29]. Consequently, the input to these networks can consist of higher-dimensional data tensors.

A seminal study conducted in 1968 revealed that the biological visual cortex employs a complex spatial hierarchy to process visual information. Receptive fields within the visual cortex operate similarly to localized spatial filters applied over image data; simpler cells detect lower-level features such as edges, whereas complex cells recognize higher-level, task-specific features [6]. The identification of specific higher-level features occurs through the composition and combination of low-level features. This biological mechanism inspired modern deep neural networks.

The concept of a convolutional neural network was first proposed by Fukushima in 1980 under the name Neocognitron. In subsequent years, modern CNN structures were investigated for use in neural networks. However, due to the requirement for high-performance

hardware and Graphics Processing Units (GPUs), the widespread adoption of CNNs for image recognition was delayed until 2012, when specialized architectures demonstrated breakthrough performance [7].

3.5.1 Architecture of Convolutional Neural Networks

CNNs are multi-layered neural network models typically comprising convolutional layers, pooling layers, and fully connected layers [7]. In a conventional neural network, input data is fed through a series of hidden layers. Each hidden layer consists of a 1D vector of neurons, where every neuron is fully connected to all neurons in the preceding layer. Neurons within a layer operate independently and do not share connection weights [17]. The final layer is the output layer, responsible for predicting class probabilities.

In contrast, input to a CNN is typically a multi-channel matrix or volumetric tensor (such as an image), and the network layers consist of neurons arranged in three spatial dimensions (height, width, and depth/channel). In these networks, neurons in a given layer connect only to a small localized region of the preceding layer rather than establishing full connectivity across all units [29]. Figure 3.2 illustrates the distinction between a standard fully connected neural network and a convolutional neural network.

As shown in the figure, all neurons in a conventional neural network are fully interconnected, whereas in a CNN, neurons are arranged in three dimensions. In Figure 3.2, the input layer (highlighted in red) represents an input image possessing height, width, and depth (e.g., three color channels—Red, Green, Blue—for an RGB image).

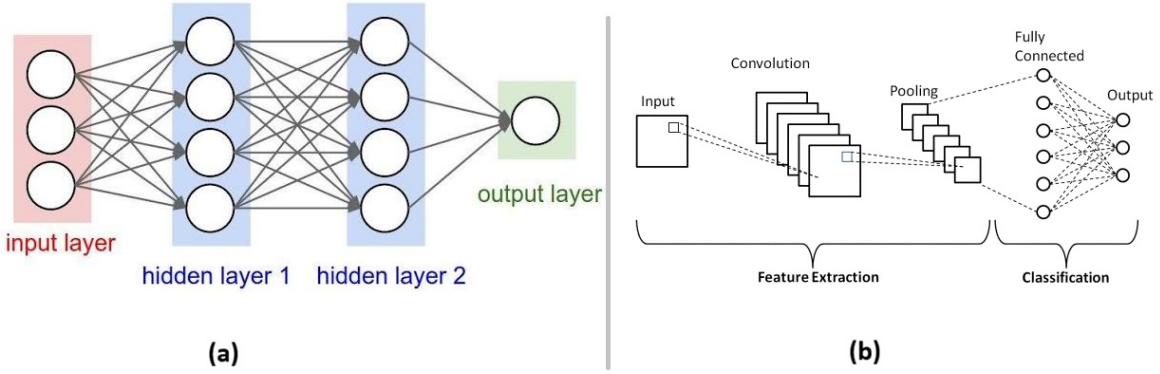

Fig. 3.2 Comparison between a standard neural network (a) and a convolutional neural network (b).

Convolutional Layer

The convolutional layer serves as the core building block of this network family, giving CNNs their name. Its primary function is feature extraction [6]. This layer performs discrete convolution operations over the input data, yielding output activation maps known as feature maps. Consequently, all neurons within a single feature map share an identical set of weights and biases (weight sharing), enabling spatial features to be detected regardless of their location in the image. Furthermore, weight sharing dramatically reduces the total number of trainable parameters. Connections in convolutional layers are constrained to small local spatial regions; that is, each neuron in the first hidden layer connects only to a small local receptive field of input units.

Pooling Layer

A pooling layer is routinely placed after one or more convolutional layers in a CNN architecture. This layer is crucial because it progressively reduces the spatial dimensions (width and height) of the feature representations, thereby decreasing the number of parameters and computational complexity in the network [7]. Additionally, by downsampling spatial dimensions, pooling helps control overfitting.

Pooling operates independently on each depth slice of the input tensor. Max pooling and average pooling represent the two most common pooling functions, with max pooling being the predominant choice in CNNs. Max pooling operates by selecting the maximum pixel intensity within each local sliding window. This window slides across the feature map from left to right and top to bottom using a specified stride, outputting the maximum value to the feature map. Because this operation is applied independently to each depth slice, the depth dimension of the output tensor remains identical to that of the input tensor.

The spatial dimensions of the output tensor $W_3 \times H_3 \times D_3$ derived from an input tensor $W_2 \times H_2 \times D_2$ are governed by the following equations:

$$W_3 = \frac{W_2 - F_p}{S_p} + 1 \tag{3.1}$$

$$H_3 = \frac{H_2 - F_p}{S_p} + 1 \tag{3.2}$$

$$D_3 = D_2 \tag{3.3}$$

In the above equations, F_p and S_p denote the receptive field (kernel) size and the stride of the pooling layer, respectively. Figure 3.3 illustrates the operation of a max pooling layer.

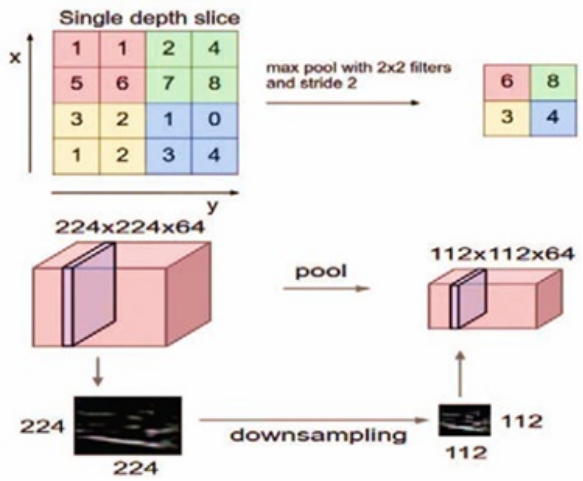

Fig. 3.3 Operation of a pooling layer with a 2×2 kernel to reduce trainable network parameters.

Fully Connected Layer

The fully connected (FC) layer forms the final stage of a convolutional neural network and establishes full connections to all activations in the preceding layer, similar to standard ANNs. This layer takes the flattened feature representations as input and outputs an N-dimensional vector, where N equals the number of target categories in the classification task.

A CNN is designed to yield an output vector of N numerical scores, where each value represents the predicted probability of belonging to a specific class. The class associated with the highest probability score represents the final prediction. In learning algorithms, a primary objective is defining a mathematical relation that measures the discrepancy between true target labels and network predictions. This relation—termed the loss function—must be minimized using optimization techniques [7].

Thus, the core concept in training a network involves optimizing the loss function via gradient steps to improve predictive accuracy. Important optimization algorithms for CNNs include Gradient Descent, Batch Gradient Descent, Stochastic Gradient Descent (SGD), Mini-batch Gradient Descent, Adagrad, RMSprop, and Adam [7].

Because CNNs possess deep architectures with large parameter spaces, they are susceptible to overfitting during training. Overfitting occurs when a network achieves high accuracy on training data but generalizes poorly to unseen test data. Numerous strategies have been introduced to mitigate overfitting. One common approach is data augmentation, while another involves inserting dropout layers within the fully connected stages [7].

Data augmentation expands the effective size of the training dataset by applying geometric and intensity transformations—such as random rotations, scaling, cropping, and horizontal

flipping—to input images. In the dropout technique, a random subset of neurons (along with their connections) is temporarily deactivated during each training iteration according to a predefined probability, forcing the remaining units to learn robust, non-redundant feature representations.

3.6 Review of Prominent Deep Learning Architectures

Deep learning is a subfield of machine learning that utilizes hierarchical architectures to learn multi-level abstractions and high-level feature representations directly from raw data [7, 6]. Deep Convolutional Neural Networks represent one of the most effective and widely adopted algorithmic paradigms in deep learning. As an evolving methodology, CNNs have been extensively deployed across diverse machine learning subdomains, most notably in computer vision. This section reviews several prominent deep CNN architectures.

3.6.1 AlexNet Architecture

AlexNet was the pioneering architecture that popularized deep convolutional neural networks within the computer vision community [7]. Developed by Krizhevsky, Sutskever, and Hinton in 2012, AlexNet features a relatively straightforward structure compared to modern networks, originally designed for 1,000-class image classification.

AlexNet consists of 5 convolutional layers followed by 3 fully connected layers. It accepts input images of dimension $227 \times 227 \times 3$ and processes them through sequential convolution, activation, and max pooling operations before passing the activations to fully connected layers. Trained on the ImageNet dataset using regularization techniques such as dropout and data augmentation, AlexNet won the ILSVRC-2012 competition by a large margin, triggering a resurgence of interest in deep learning.

Figure 3.4 illustrates the AlexNet architecture (the two-stream split in the diagram reflects parallel execution across two GPUs). While some diagrams label the input as $224 \times 224 \times 3$, the precise spatial dimensions used were $227 \times 227 \times 3$. The first convolutional layer applies 96 filters of size 11×11 with a stride of 4 and padding of 0, yielding a feature map tensor of size $55 \times 55 \times 96$. (The trainable parameter count for a convolutional layer equals: kernel width $\times$ kernel height $\times$ input depth $\times$ number of filters).

The output is then passed through a 3×3 max pooling layer with stride 2, reducing dimensions to $27 \times 27 \times 96$ (pooling layers contain zero trainable parameters). Next, a second convolutional layer applies 256 filters of size 5×5 with stride 1 and padding 2,

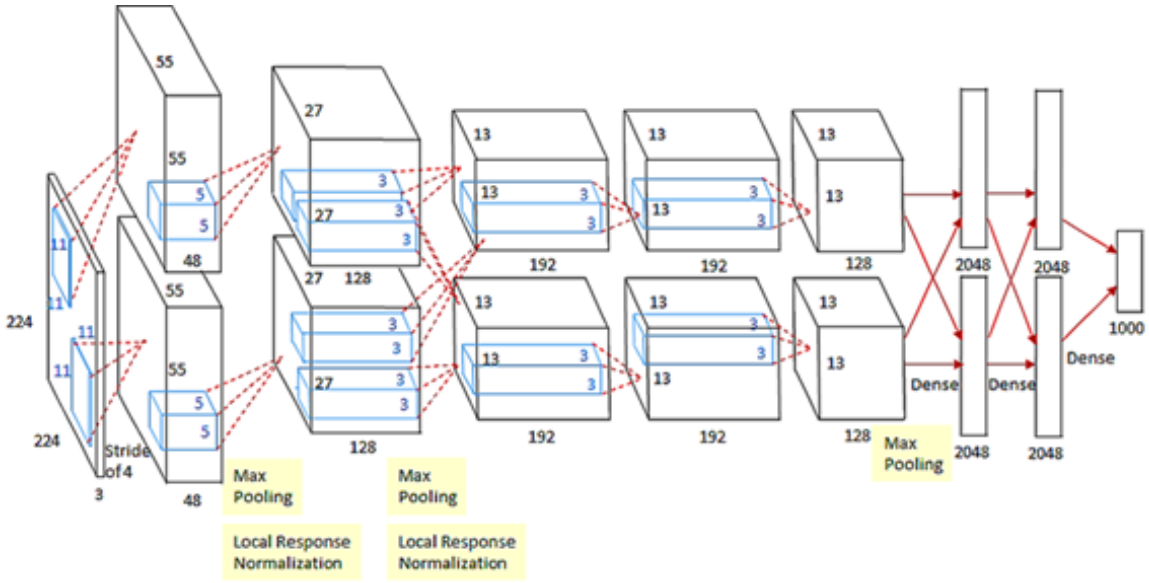

Fig. 3.4 The AlexNet architecture.

producing a $27 \times 27 \times 256$ output, which is again downsampled by a 3×3 max pooling layer (stride 2) to $13 \times 13 \times 256$.

The subsequent three convolutional layers utilize 3×3 kernels with stride 1 and padding 1, containing 384, 384, and 256 filters, respectively, resulting in a tensor of size $13 \times 13 \times 256$. A final 3×3 max pooling layer with stride 2 yields a tensor of size $6 \times 6 \times 256$. Finally, three fully connected layers (comprising 4,096, 4,096, and 1,000 units, respectively) process these features, culminating in a Softmax activation layer to compute class probabilities.

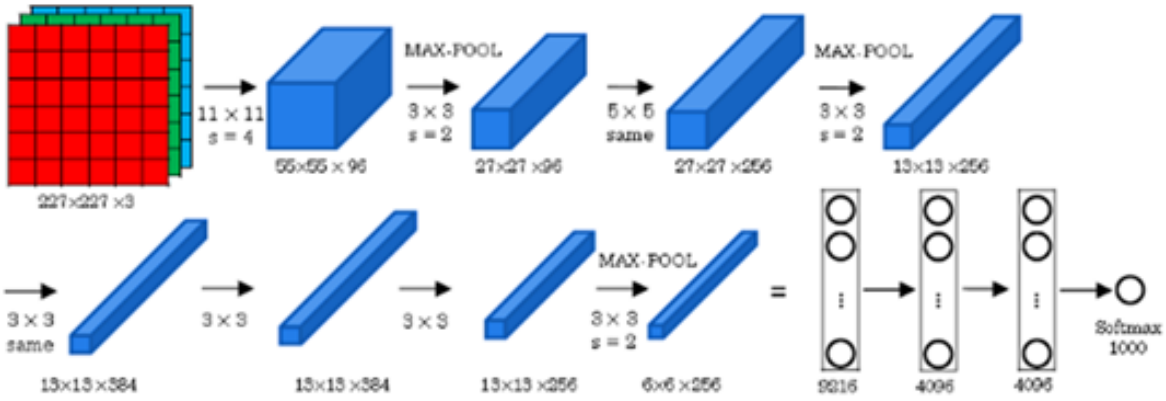

Fig. 3.5 An example of the AlexNet architecture.

3.6.2 GoogLeNet Architecture

GoogLeNet (Inception-v1) is a 22-layer deep network architecture that achieved a top-5 error rate of 6.7% in ILSVRC-2014 [24]. It departed from standard sequential paradigms where convolutional and pooling layers were stacked strictly one after another. GoogLeNet was designed with a focus on computational efficiency and memory optimization, addressing the issue that stacking numerous layers with large filters significantly increases parameter counts, memory footprint, and susceptibility to overfitting.

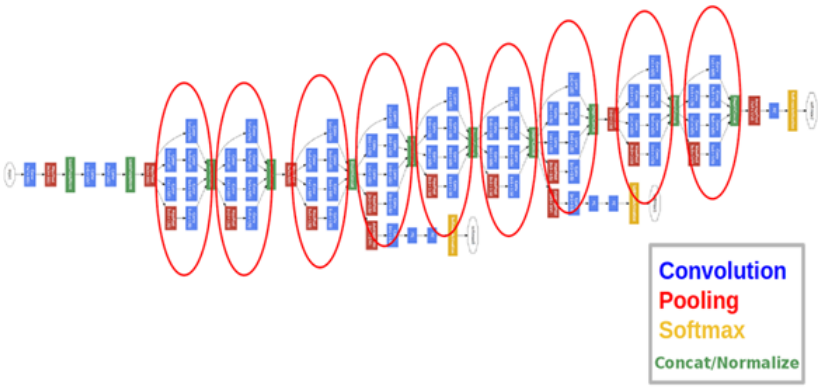

Fig. 3.6 The GoogLeNet architecture.

GoogLeNet introduces non-sequential, parallel multi-scale processing blocks known as *Inception modules* [24]. Rather than forcing a choice between specific filter sizes (1×1, 3×3, 5×5) or pooling operations at a given layer, the Inception module executes these operations in parallel and concatenates their output feature maps along the channel dimension.

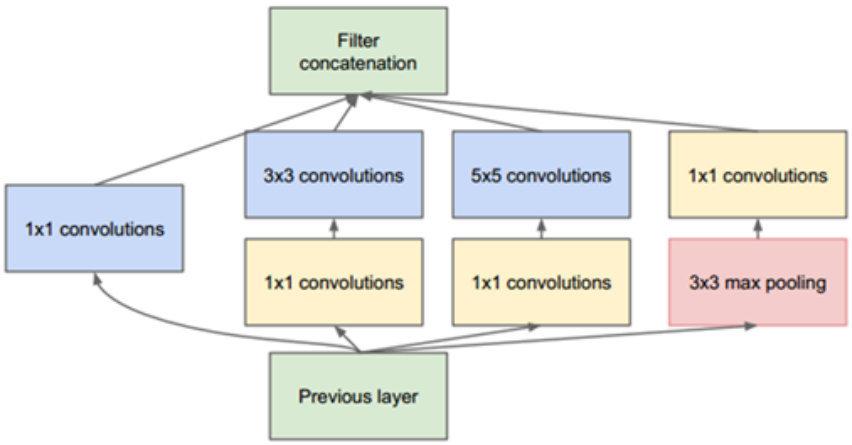

Fig. 3.7 The Inception layer module in the GoogLeNet architecture.

To prevent high computational costs from expanding channel depths, 1×1 convolutions are inserted prior to 3×3 and 5×5 convolutions as dimensionality reduction bottlenecks [24]. For example, if an input tensor has dimensions $100 \times 100 \times 60$, applying 20 filters of size 1×1 reduces its dimension to $100 \times 100 \times 20$, lowering the computational load for subsequent 3×3 and 5×5 filters.

By combining multi-scale receptive fields, 1×1 bottleneck convolutions, parallel max pooling, and non-linear ReLU activations, GoogLeNet captures spatial features across multiple scales while maintaining a low parameter footprint [24].

3.6.3 LeNet-5 Architecture

LeNet-5 is one of the earliest successful convolutional neural networks, introduced by LeCun et al. in their 1998 paper titled *"Gradient-based learning applied to document recognition"*. Named after lead author Yann LeCun, LeNet-5 established the foundational principles of modern CNNs—combining convolutional layers, subsampling (pooling) layers, and fully connected layers. It was originally designed to perform automated handwritten digit and zip code recognition on postal envelopes.

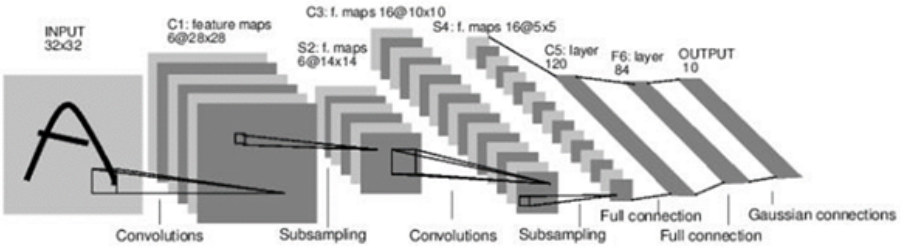

Fig. 3.8 The LeNet architecture.

LeNet-5 receives an input image of size 32×32 pixels and processes it through an initial convolutional layer with 6 feature maps using 5×5 filters with stride 1, yielding a tensor of size $28 \times 28 \times 6$. This is followed by a 2×2 average pooling layer with stride 2 (subsampling), which halves the spatial dimensions to $14 \times 14 \times 6$.

A second convolutional layer applies 16 filters of size 5×5 with stride 1, generating a tensor of size $10 \times 10 \times 16$, which is subsequently downsampled by another 2×2 average pooling layer to $5 \times 5 \times 16$. Finally, the spatial activations are flattened and passed through fully connected layers to generate the final classification outputs.

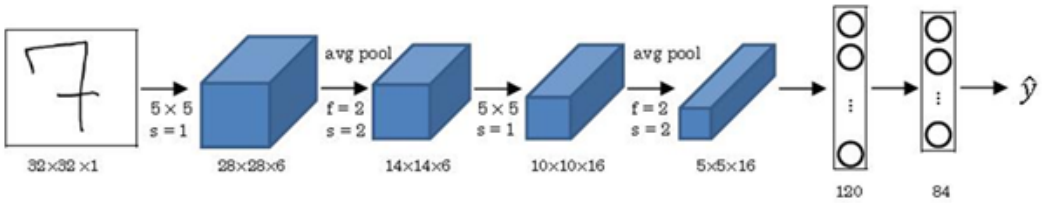

Fig. 3.9 An example of the LeNet-5 architecture.

3.6.4 VGGNet Architecture

VGGNet was developed by the Visual Geometry Group at the University of Oxford. It is recognized for its uniform architectural design, featuring a repeating hierarchy of small 3×3 convolutional filters stacked sequentially, interleaved with max pooling layers to progressively decrease spatial resolution while increasing channel depth.

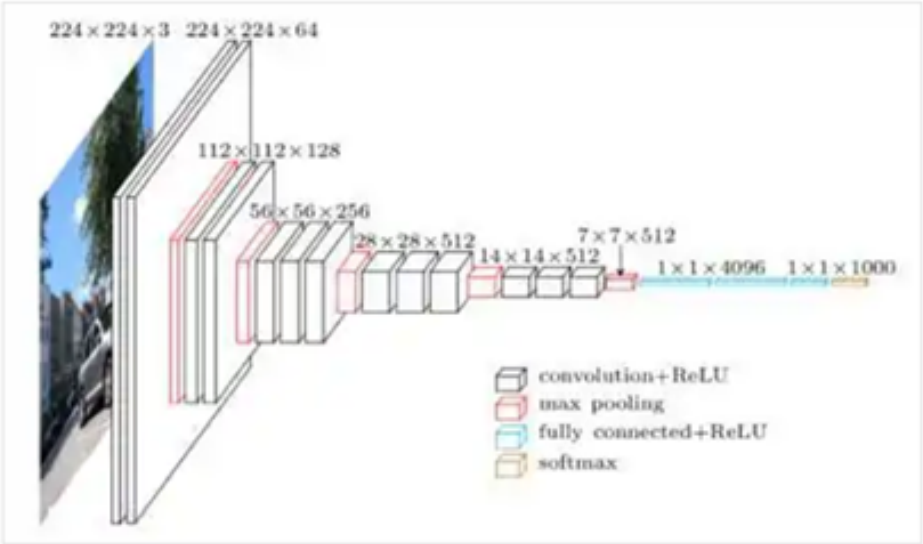

Fig. 3.10 The VGGNet architecture.

The VGG family includes several configurations (most notably VGG-16 and VGG-19), which vary according to network depth. Key advantages of VGGNet include:

- A clean, standardized architecture that serves as a robust benchmark for vision tasks,
- Highly effective transfer learning performance using publicly available pre-trained weights [16].

However, a main disadvantage of VGGNet is its large parameter count (over 138 million parameters in VGG-16), which leads to slow training times and high computational and memory requirements.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 3.11 The VGGNet architecture.

3.6.5 ResNet Architecture

Residual Networks (ResNet), introduced by He et al., represented a major breakthrough by enabling neural networks to be trained at extreme depths (e.g., 50, 101, or 152 layers) without performance degradation [7]. ResNet addresses the degradation problem—where extremely deep networks experience higher training error—by introducing shortcut (skip) connections that form *residual blocks*.

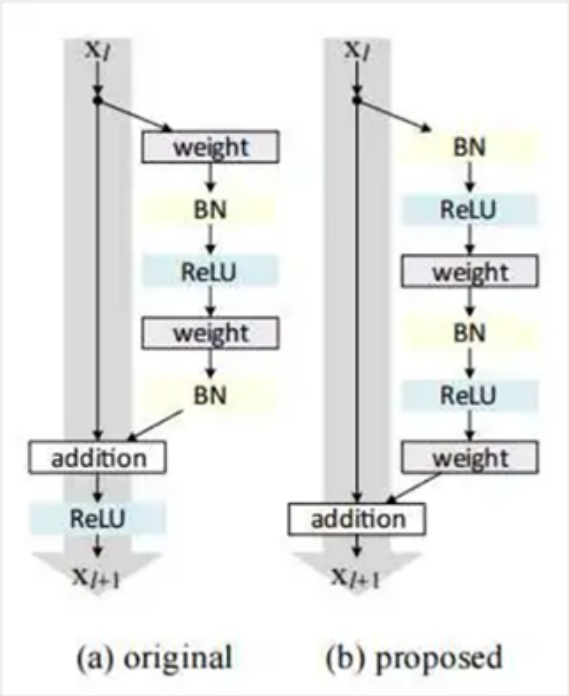

Fig. 3.12 The ResNet architecture.

A residual block allows inputs to bypass one or more convolutional layers via an identity mapping shortcut, adding the input x directly to the learned residual function $F(x)$:

$$y = F(x, \{W_i\}) + x \tag{3.4}$$

This formulation simplifies gradient flow during backpropagation, mitigating the vanishing gradient problem.

Additional design features introduced or popularized in ResNet include:

- Utilizing standard Stochastic Gradient Descent (SGD) with momentum alongside Batch Normalization to stabilize training dynamics,
- Replacing heavy fully connected layers at the network tail with a Global Average Pooling layer to reduce parameter counts.

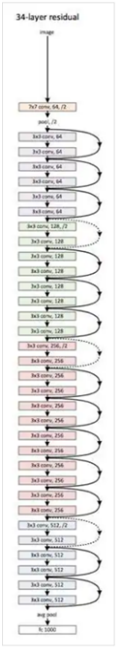

Fig. 3.13 The ResNet architecture.

3.6.6 ResNeXt Architecture

ResNeXt is an evolution of the ResNet architecture that incorporates multi-branch processing concepts inspired by GoogLeNet's Inception module [24]. It introduces a hyperparameter termed *cardinality*—defined as the number of parallel independent paths within a residual block. By repeating a simple, multi-branch bottleneck structure, ResNeXt increases model capacity and feature diversity without significantly increasing computational complexity or parameter counts.

3.6.7 Region-Based CNN (R-CNN)

Region-Based Convolutional Neural Networks (R-CNN) introduced a multi-stage framework for object detection tasks. R-CNN operates by generating approximately 2,000 category-independent region proposals per input image using a selective search algorithm. Each proposed candidate region is warped to a fixed spatial dimension and passed through a deep CNN to extract feature vectors. Finally, class-specific Support Vector Machines (SVMs) classify each region, while bounding-box regressors refine spatial localizations. Subsequent iterations—such as Fast R-CNN and Faster R-CNN—streamlined this pipeline by introducing region of interest (RoI) pooling and Region Proposal Networks (RPNs) for end-to-end execution.

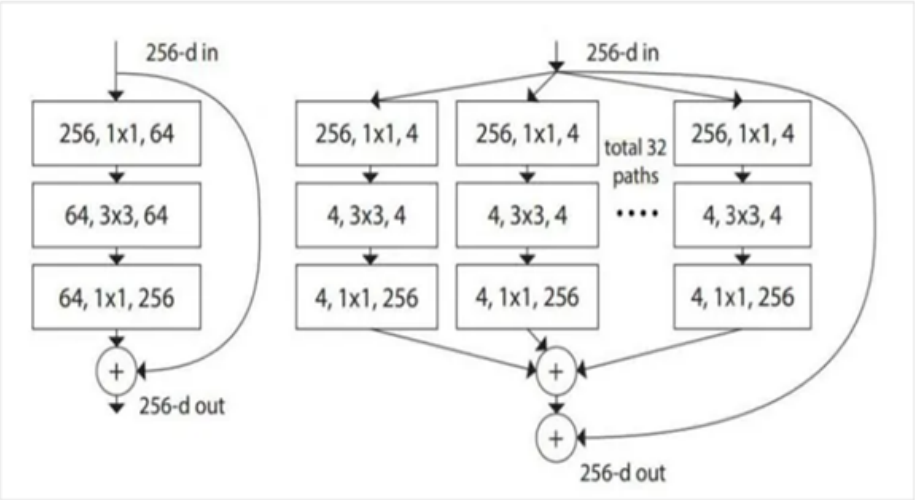

Fig. 3.14 The ResNeXt architecture.

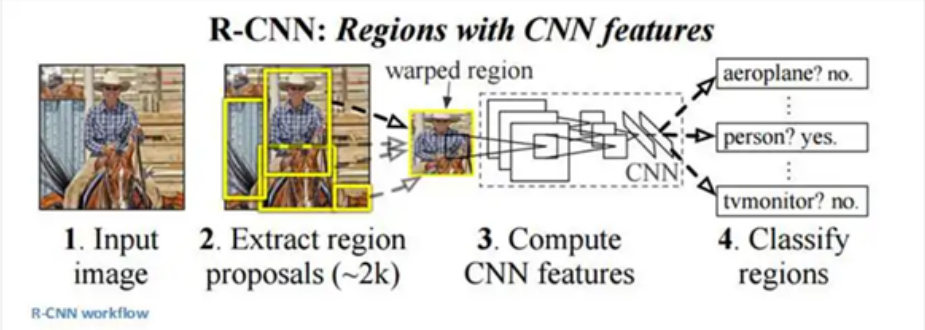

Fig. 3.15 The Region-Based CNN (R-CNN) architecture.

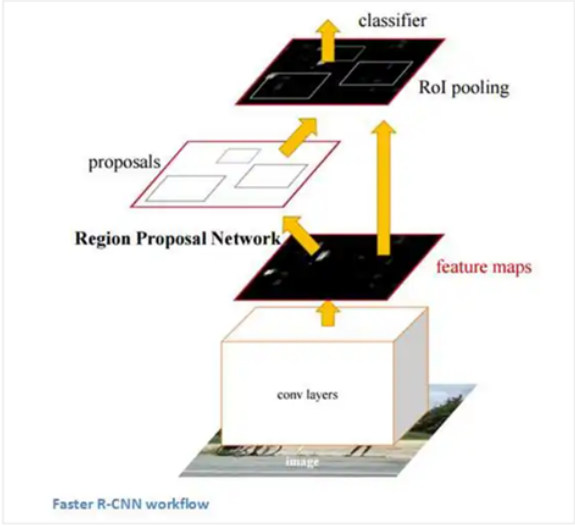

Fig. 3.16 The Region-Based CNN (R-CNN) architecture.

3.6.8 You Only Look Once (YOLO) Architecture

You Only Look Once (YOLO) re-framed object detection from a multi-stage pipeline into a single, real-time regression problem. Unlike region proposal-based frameworks, YOLO processes the entire image in a single forward pass.

Fig. 3.17 The You Only Look Once (YOLO) architecture.

The algorithm divides the input image into an $S \times S$ grid. If an object's center falls within a grid cell, that cell is responsible for predicting its bounding box coordinates, confidence scores, and class probabilities simultaneously. By evaluating spatial features across the entire image in parallel, YOLO achieves high inference speeds (processing up to 45+ frames per second), making it suitable for real-time video applications.

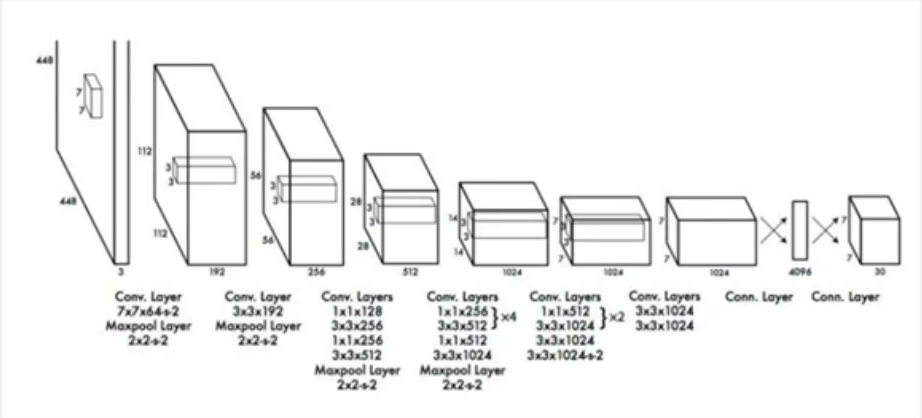

Fig. 3.18 The You Only Look Once (YOLO) architecture.

3.6.9 SqueezeNet Architecture

SqueezeNet was designed specifically for resource-constrained environments (such as mobile and embedded devices) where memory footprint and storage limitations are critical. SqueezeNet achieves AlexNet-level accuracy on ImageNet with a model size under 5 MB—a 50× reduction in parameter size compared to AlexNet's 240 MB footprint.

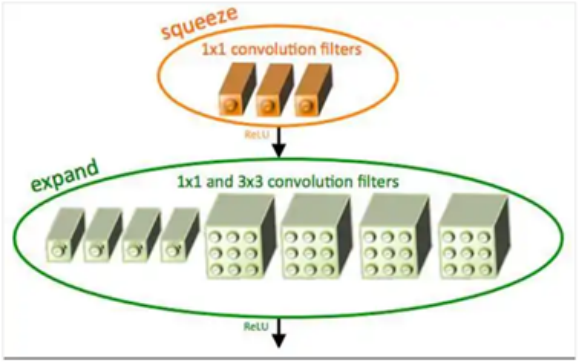

Fig. 3.19 The SqueezeNet architecture.

This compression is achieved using specialized building blocks called *Fire modules*. A Fire module consists of a *squeeze* convolutional layer (comprising 1×1 filters) fed into an *expand* layer (comprising a mix of 1×1 and 3×3 filters). By replacing 3×3 filters with 1×1 filters and decreasing input channels to 3×3 convolutions, SqueezeNet maintains high representation capacity while reducing total network parameters.

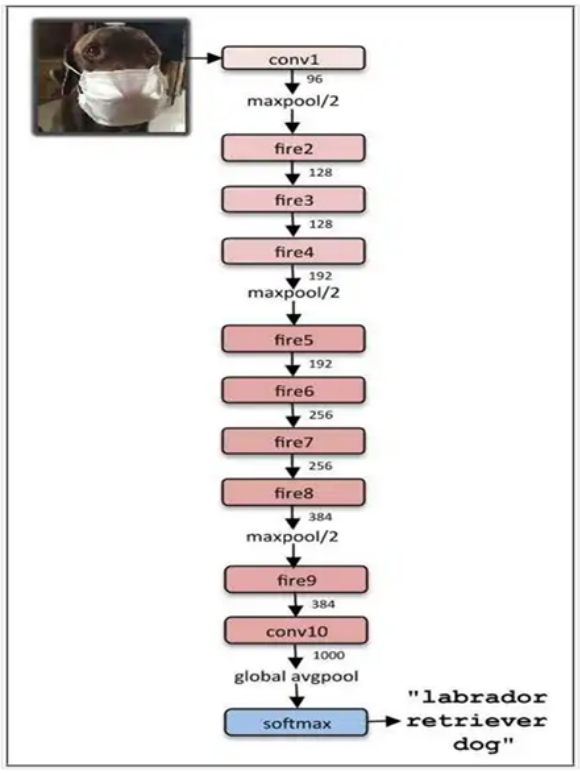

Fig. 3.20 The SqueezeNet architecture.

3.6.10 SegNet Architecture

SegNet is a deep fully convolutional encoder-decoder architecture designed for dense, pixel-wise semantic image segmentation. The encoder network consists of 13 convolutional layers corresponding to the topological structure of VGG-16, which extract spatial feature hierarchies. The decoder network mirrors the encoder structure and maps low-resolution feature representations back to full input resolution for pixel-level classification.

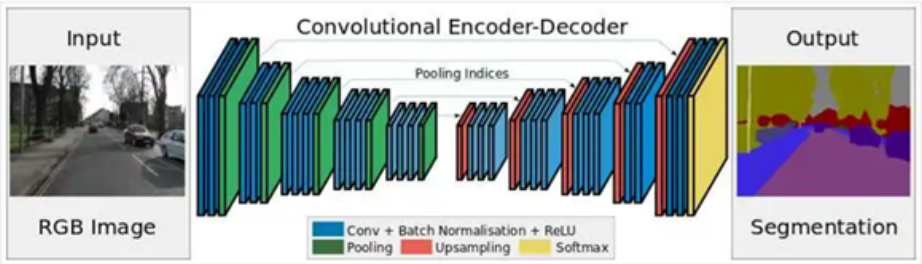

Fig. 3.21 The SegNet architecture.

A key technical feature of SegNet is its memory-efficient max-pooling indices transfer mechanism. During encoder max-pooling stages, the spatial locations (indices) of maximum feature activations within each pooling window are recorded. The decoder then utilizes these stored pooling indices to perform non-linear upsampling without requiring trainable parameters during upsampling. This preserves edge detail in the segmented output while maintaining computational efficiency.

3.6.11 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs), introduced by Goodfellow et al. in 2014, represent a generative modeling framework structured as a two-player adversarial game [7]. The architecture consists of two neural networks trained simultaneously:

1. **Generator** (G): Learns to capture the true data distribution and generate synthetic samples from random noise vectors (z), attempting to produce realistic samples.

2. **Discriminator** (D): Learns to discriminate between real samples drawn from the training dataset and synthetic samples synthesized by G.

The two networks are trained iteratively using a minimax objective function, driving the generator to produce highly realistic images that mimic the training distribution.

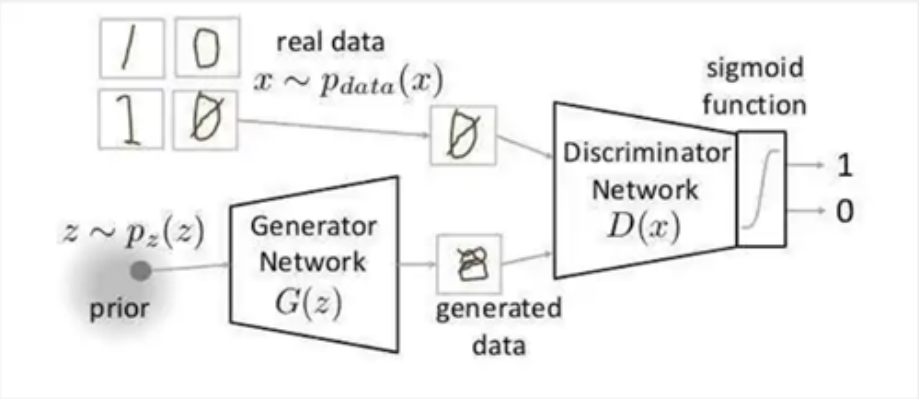

Fig. 3.22 The GAN architecture.

3.6.12 U-Net Architecture

U-Net is a specialized fully convolutional neural network architecture originally designed for biomedical image segmentation [21, 8]. The architecture derives its name from its symmetric, U-shaped structural layout. U-Net consists of two primary paths: a contracting path (encoder) on the left and an expansive path (decoder) on the right.

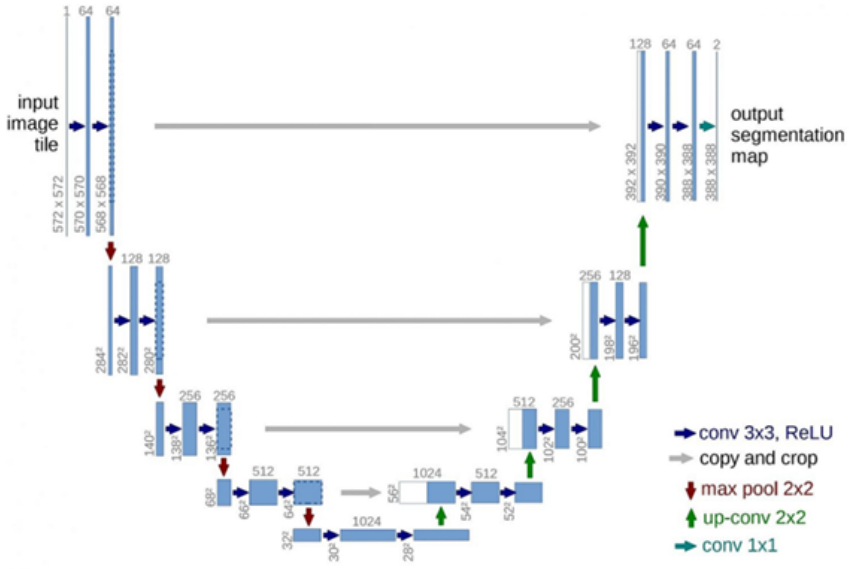

Fig. 3.23 The U-Net architecture (example for an image with the lowest possible resolution of 32×32 pixels). Each blue box corresponds to a multi-channel feature map. The number of output channels is specified on top of each box, and the dimensions are indicated at the lower-left corner. The white boxes represent copied feature maps, and each arrow denotes a specific operation.

The contracting path captures contextual information by applying repeated pairs of 3×3 convolutions, ReLU activations, and 2×2 max pooling operations to downsample spatial dimensions. The expansive path enables precise spatial localization by applying transposed convolutions (up-convolutions) to restore feature maps to full spatial resolution.

Crucially, U-Net incorporates high-resolution *skip connections* that concatenate feature maps directly from contracting layers to corresponding expansive layers [16]. This mechanism transfers fine-grained spatial and edge details lost during downsampling directly to the decoder, making U-Net particularly effective for segmenting complex anatomical structures in medical imaging datasets [28, 21].

3.7 Summary

Deep learning represents a key subfield of machine learning focused on learning hierarchical representations directly from raw data through multi-layered architectures [7, 6]. Deep Convolutional Neural Networks stand among the most prominent algorithms in deep learning, widely deployed across computer vision and medical image analysis tasks [29, 24, 16]. This chapter reviewed artificial neural network concepts, the core mechanics of convolutional networks, and several established CNN architectures used in automated image recognition and semantic segmentation.

Chapter 4

Breast Cancer Pathology Image Segmentation Using the U-Net Model

4.1 Google Colab

Google Colaboratory, commonly known as Google Colab, is a cloud-based Jupyter notebook environment provided with free hardware acceleration that runs entirely in the cloud. Colab was launched by Google to democratize working and conducting research in artificial intelligence. The Colab cloud service grants access to powerful Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) for executing Python-based artificial intelligence algorithms. With Colab, researchers and team members can collaborate on a project seamlessly without requiring high-performance local hardware. Working with Colab resembles Google Docs, where all operations are executed online. Many popular machine learning and computer vision libraries—such as TensorFlow, Keras, PyTorch, OpenCV, NumPy, and Scikit-learn—are pre-installed, allowing immediate and hassle-free usage.

4.1.1 Premium Service

During its initial rollout, Colab experienced various bugs and frequent session disconnections. However, Google continuously introduced platform enhancements, establishing it as a popular and widely adopted tool worldwide. In recent years, Google introduced Colab Pro and Colab Pro+, which offer expanded computing resources and high-priority access to premium GPUs for a low subscription fee.

4.1.2 Creating a Notebook in Google Colab

The initial phase in developing the reference implementation involves building a new notebook. The step-by-step procedure for creating a notebook file in Google Colab is outlined below:

Step 1: Navigate to the Colab website.

Step 2: Colab requires linking to a Google account. Click on the **Sign in** button (as shown in Figure 4.1) and enter your account credentials to connect Colab to your Google account.

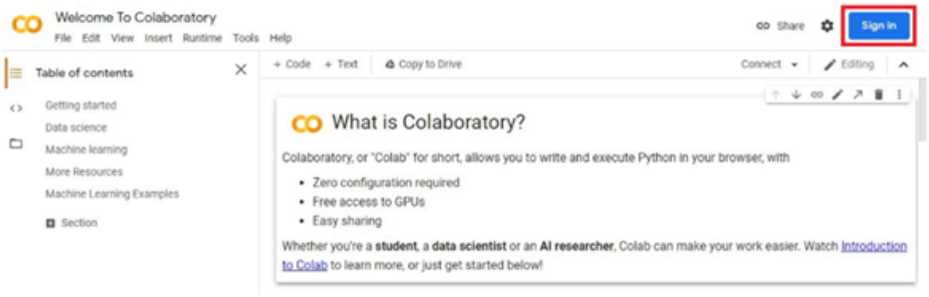

Fig. 4.1

Step 3: Once authenticated, a pop-up window containing template notebooks appears. Click on the **NEW NOTEBOOK** option to generate a fresh workspace.

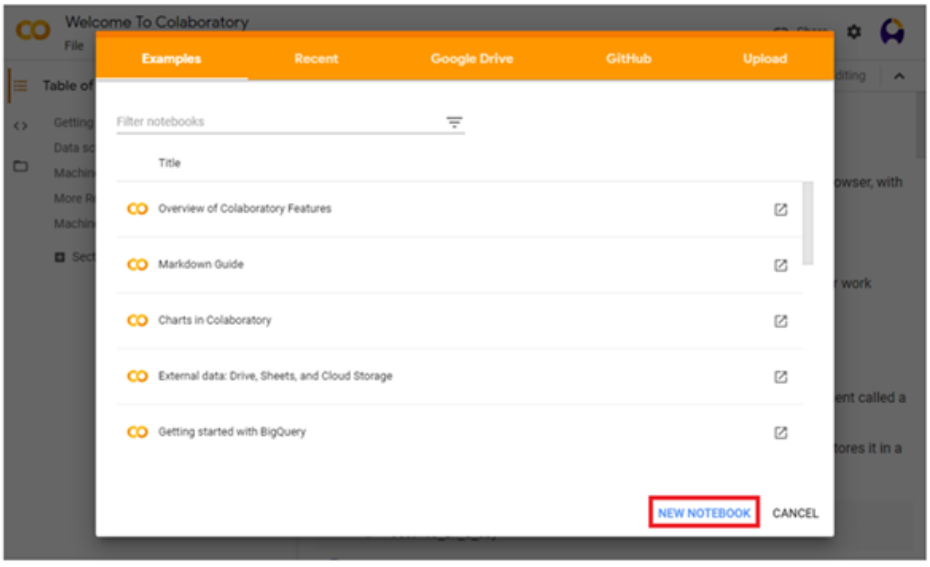

Fig. 4.2

Upon completing these steps, the notebook environment initializes, ready for code execution.

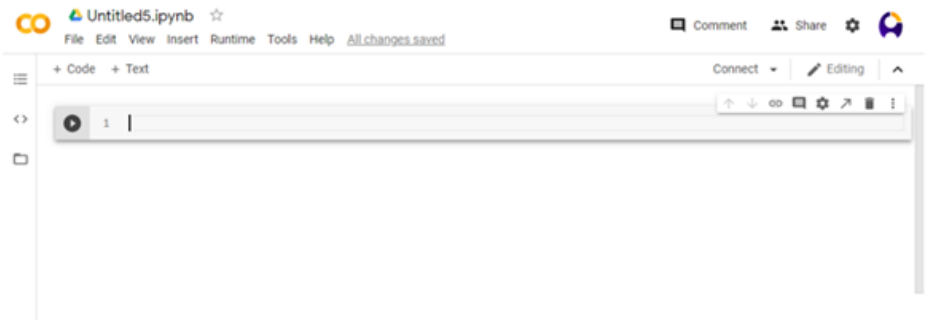

Fig. 4.3

4.2 Implementation of the U-Net Model

In this project, a U-Net architecture is implemented to perform semantic segmentation on breast cancer pathology images.

The overall methodology operates as follows: initially, a set of annotated training samples is provided to the network for model training. During this phase, the network learns the spatial mappings required to synthesize corresponding binary segmentation masks from input images.

During the testing phase, unseen test images are fed to the trained network. The system automatically predicts binary masks for each image, which are subsequently compared against expert-annotated ground truth masks. The quantitative evaluation metric used for mask comparison is Intersection over Union (IoU), as described in subsequent sections.

4.2.1 Mounting Google Drive

To access datasets stored on Google Drive within the Colab environment, the drive is mounted using the following Python code snippet:

```
from google.colab import drive
drive.mount('/content/gdrive')
```

Listing 4.1 Mounting Google Drive in Google Colab

4.2.2 Installing TensorFlow and Keras

To install specific target framework versions (TensorFlow version 1.14.0 and Keras version 2.0.0), the following commands are executed:

```
!pip install tensorflow==1.14.0
!pip install keras==2.0.0
```

Listing 4.2 Installing targeted framework versions

4.2.3 Importing Required Modules

The implementation relies on several numerical processing, deep learning, image processing, and visualization modules, imported as follows:

```
import os
import glob
import keras
import random
import numpy as np
import tensorflow as tf
from keras.layers import *
import keras.backend as K
from keras.models import *
from keras.optimizers import *
import matplotlib.pyplot as plt
from skimage.transform import resize
from skimage.io import imread, imshow, imsave
from keras.callbacks import ModelCheckpoint,
   LearningRateScheduler, EarlyStopping
```

Listing 4.3 Importing required Python libraries and modules

4.2.4 Defining Dataset Paths

The file paths for training images, training masks, testing images, and testing masks are defined as follows:

```
# Define Train & Test Paths (Images + Mask Paths)
TRAIN_IMAGE_PATH = '/content/gdrive/MyDrive/U_Net_New/
   Inputs_Train'
```

```
TRAIN_MASK_PATH = '/content/gdrive/MyDrive/U_Net_New/Masks_Train'
TEST_IMAGE_PATH = '/content/gdrive/MyDrive/U_Net_New/Inputs_Test'
TEST_MASK_PATH = '/content/gdrive/MyDrive/U_Net_New/Masks_Test'
```

Listing 4.4 Defining file paths for dataset directories

4.2.5 Initializing Image and Mask Dimensions

Target dimensions (height, width, and number of channels) for the image arrays are initialized as follows:

```
# Initialize Image and Mask Dimensions
IMG_HEIGHT, IMG_WIDTH, IMG_CHANNELS = 768, 896, 3
```

Listing 4.5 Setting spatial dimensions and image channels

4.2.6 Defining Data Preprocessing Functions

To perform Region of Interest (ROI) extraction, image loading, cropping, and spatial resizing, training and testing preprocessing functions are defined:

```
# Define Preprocessing Functions (Region of Interest Extraction - ROI)

Train_Mask_List = sorted(next(os.walk(TRAIN_MASK_PATH))[2])
Test_Mask_List = sorted(next(os.walk(TEST_MASK_PATH))[2])

def Data_Preprocessing_Train():
    Init_Image = np.zeros((len(Train_Mask_List), 768, 896, 3), dtype=np.uint8)
    Init_Mask = np.zeros((len(Train_Mask_List), 768, 896), dtype=bool)
    Train_X = np.zeros((len(Train_Mask_List), IMG_HEIGHT, IMG_WIDTH, IMG_CHANNELS), dtype=np.uint8)
    Train_Y = np.zeros((len(Train_Mask_List), IMG_HEIGHT, IMG_WIDTH, 1), dtype=bool)

    n = 0

```

```
    for mask_path in glob.glob('{}/*.TIF'.format(TRAIN_MASK_PATH)):
        base = os.path.basename(mask_path)
        image_ID, ext = os.path.splitext(base)
        image_path = '{}/{}_ccd.tif'.format(TRAIN_IMAGE_PATH, image_ID)
        mask = imread(mask_path)
        image = imread(image_path)

        y_coord, x_coord = np.where(mask == 255)

        y_min = min(y_coord)
        y_max = max(y_coord)
        x_min = min(x_coord)
        x_max = max(x_coord)

        cropped_image = image[y_min:y_max, x_min:x_max]
        cropped_mask = mask[y_min:y_max, x_min:x_max]

        Train_X[n] = resize(cropped_image[:, :, :IMG_CHANNELS],
                            (IMG_HEIGHT, IMG_WIDTH, IMG_CHANNELS),
                            mode='constant',
                            anti_aliasing=True,
                            preserve_range=True)

        Train_Y[n] = np.expand_dims(resize(cropped_mask,
                                           (IMG_HEIGHT, IMG_WIDTH),
                                           mode='constant',
                                           anti_aliasing=True,
                                           preserve_range=True), axis=-1)

        Init_Image[n] = image
        Init_Mask[n] = mask

        n += 1

```

```
48      return Train_X, Train_Y, Init_Image, Init_Mask
49
50  Train_Inputs, Train_Masks, Init_Image, Init_Mask =
        Data_Preprocessing_Train()
51
52
53  def Data_Preprocessing_Test():
54      Test_X = np.zeros((len(Test_Mask_List), IMG_HEIGHT,
        IMG_WIDTH, IMG_CHANNELS), dtype=np.uint8)
55      Test_Y = np.zeros((len(Test_Mask_List), IMG_HEIGHT,
        IMG_WIDTH, 1), dtype=bool)
56
57      n = 0
58
59      for mask_path in glob.glob('{}/*.TIF'.format(TEST_MASK_PATH)
        ):
60          base = os.path.basename(mask_path)
61          image_ID, ext = os.path.splitext(base)
62          image_path = '{}/{}_ccd.tif'.format(TEST_IMAGE_PATH,
        image_ID)
63          mask = imread(mask_path)
64          image = imread(image_path)
65
66          y_coord, x_coord = np.where(mask == 255)
67
68          y_min = min(y_coord)
69          y_max = max(y_coord)
70          x_min = min(x_coord)
71          x_max = max(x_coord)
72
73          cropped_image = image[y_min:y_max, x_min:x_max]
74          cropped_mask = mask[y_min:y_max, x_min:x_max]
75
76          Test_X[n] = resize(cropped_image[:, :, :IMG_CHANNELS],
77                             (IMG_HEIGHT, IMG_WIDTH, IMG_CHANNELS)
        ,
78                             mode='constant',
79                             anti_aliasing=True,
80                             preserve_range=True)
```

```
81
82          Test_Y[n] = np.expand_dims(resize(cropped_mask,
83                                            (IMG_HEIGHT, IMG_WIDTH
    ),
84                                            mode='constant',
85                                            anti_aliasing=True,
86                                            preserve_range=True),
    axis=-1)
87
88          n += 1
89
90      return Test_X, Test_Y
91
92  Test_Inputs, Test_Masks = Data_Preprocessing_Test()
```

Listing 4.6 Defining training and testing data preprocessing pipeline

4.2.7 Visualizing Preprocessing Results

The raw input images, ground truth masks, and extracted ROI regions are visualized using Matplotlib:

```
1  # Show Results in Preprocessing Stage
2
3  print('Original_Image')
4  imshow(Init_Image[0])
5  plt.show()
6
7  print('Original_Mask')
8  imshow(Init_Mask[0])
9  plt.show()
10
11 print('Region_of_Interest_Image')
12 imshow(Train_Inputs[0])
13 plt.show()
14
15 print('Region_of_Interest_Mask')
16 imshow(np.squeeze(Train_Masks[0]))
17 plt.show()
18
```

```
rows = 1
columns = 4
Figure = plt.figure(figsize=(15, 15))
Image_List = [Init_Image[0], Init_Mask[0], Train_Inputs[0],
    Train_Masks[0]]

for i in range(1, rows * columns + 1):
    Image = Image_List[i - 1]
    Sub_Plot_Image = Figure.add_subplot(rows, columns, i)
    Sub_Plot_Image.imshow(np.squeeze(Image))
plt.show()
```

Listing 4.7 Visualization of original and preprocessed ROI masks

4.2.8 U-Net Architecture Implementation

The full U-Net architecture for semantic segmentation is defined using Keras functional API:

```
# Implementation of U-Net Model for Semantic Segmentation

def U_Net_Segmentation(input_size=(IMG_HEIGHT, IMG_WIDTH,
    IMG_CHANNELS)):

    inputs = Input(input_size)
    n = Lambda(lambda x: x / 255)(inputs)

    # Contracting Path (Encoder)
    c1 = Conv2D(16, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(n)
    c1 = Dropout(0.1)(c1)
    c1 = Conv2D(16, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(c1)
    p1 = MaxPooling2D((2, 2))(c1)

    c2 = Conv2D(32, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(p1)
    c2 = Dropout(0.1)(c2)
    c2 = Conv2D(32, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(c2)
    p2 = MaxPooling2D((2, 2))(c2)
```

```
18
19     c3 = Conv2D(64, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(p2)
20     c3 = Dropout(0.2)(c3)
21     c3 = Conv2D(64, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(c3)
22     p3 = MaxPooling2D((2, 2))(c3)
23
24     c4 = Conv2D(128, (3, 3), activation='elu',
    kernel_initializer='he_normal', padding='same')(p3)
25     c4 = Dropout(0.2)(c4)
26     c4 = Conv2D(128, (3, 3), activation='elu',
    kernel_initializer='he_normal', padding='same')(c4)
27     p4 = MaxPooling2D((2, 2))(c4)
28
29     # Bottleneck
30     c5 = Conv2D(256, (3, 3), activation='elu',
    kernel_initializer='he_normal', padding='same')(p4)
31     c5 = Dropout(0.3)(c5)
32     c5 = Conv2D(256, (3, 3), activation='elu',
    kernel_initializer='he_normal', padding='same')(c5)
33
34     # Expansive Path (Decoder)
35     u6 = Conv2DTranspose(128, (2, 2), strides=(2, 2), padding='
    same')(c5)
36     u6 = concatenate([u6, c4])
37     c6 = Conv2D(128, (3, 3), activation='elu',
    kernel_initializer='he_normal', padding='same')(u6)
38     c6 = Dropout(0.2)(c6)
39     c6 = Conv2D(128, (3, 3), activation='elu',
    kernel_initializer='he_normal', padding='same')(c6)
40
41     u7 = Conv2DTranspose(64, (2, 2), strides=(2, 2), padding='
    same')(c6)
42     u7 = concatenate([u7, c3])
43     c7 = Conv2D(64, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(u7)
44     c7 = Dropout(0.2)(c7)
```

```
45      c7 = Conv2D(64, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(c7)
46
47      u8 = Conv2DTranspose(32, (2, 2), strides=(2, 2), padding='
    same')(c7)
48      u8 = concatenate([u8, c2])
49      c8 = Conv2D(32, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(u8)
50      c8 = Dropout(0.1)(c8)
51      c8 = Conv2D(32, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(c8)
52
53      u9 = Conv2DTranspose(16, (2, 2), strides=(2, 2), padding='
    same')(c8)
54      u9 = concatenate([u9, c1], axis=3)
55      c9 = Conv2D(16, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(u9)
56      c9 = Dropout(0.1)(c9)
57      c9 = Conv2D(16, (3, 3), activation='elu', kernel_initializer
    ='he_normal', padding='same')(c9)
58
59      outputs = Conv2D(1, (1, 1), activation='sigmoid')(c9)
60
61      model = Model(inputs=[inputs], outputs=[outputs])
62      model.compile(optimizer='adam', loss='binary_crossentropy',
    metrics=[Mean_IOU_Evaluator])
63      model.summary()
64      return model
```

Listing 4.8 Definition and compilation of the U-Net architecture

4.2.9 Loss Functions in Medical Image Segmentation

Loss functions represent a pivotal component in deep learning-based medical image segmentation frameworks. Over recent years, more than twenty distinct loss functions have been formulated to address challenges in automated image segmentation.

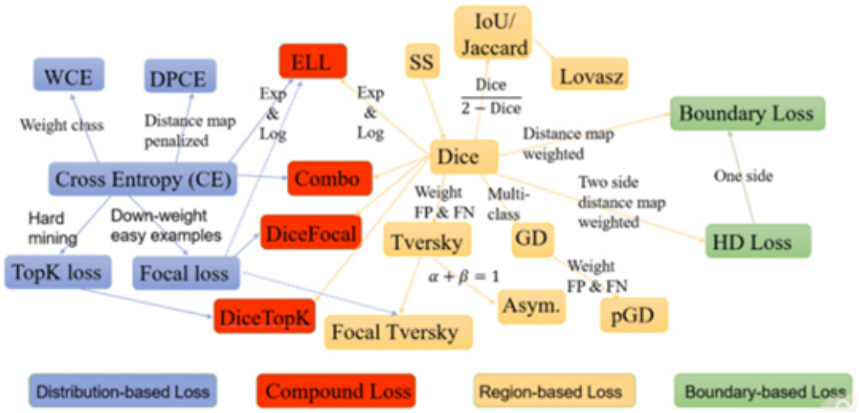

Fig. 4.4

Distribution-Based Loss Functions

Cross-Entropy (CE) Loss: Derived from the Kullback-Leibler (KL) divergence, Cross-Entropy quantifies dissimilarity between two probability distributions. In machine learning tasks, the empirical target distribution is defined by ground-truth labels and remains fixed. Consequently, minimizing CE loss is mathematically equivalent to minimizing KL divergence.

$$D_{KL}(p \parallel q) = \underbrace{H(p,q)}_{\text{Cross entropy}} - \overbrace{H(p)}^{\text{Entropy}}$$

Fig. 4.5

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} g_i^c \log s_i^c$$

Fig. 4.6

Weighted Cross-Entropy (WCE) Loss: A generalization of standard CE loss wherein class-specific weights are assigned to penalize underrepresented target categories. Higher weights are typically assigned to sparse target classes.

TopK Loss: Designed to force the network to focus training updates on hard, high-error samples during backpropagation.

Focal Loss: Modifies standard CE loss to address extreme class imbalance between background pixels and foreground target structures by suppressing loss contributions from well-classified easy samples.

Distance-Penalized Cross-Entropy Loss: Incorporates distance-map penalty matrices derived from binary ground truth masks into weighted CE loss, guiding model attention toward difficult boundary contours.

Region-Based Loss Functions

Region-based loss functions aim to minimize structural misalignment or maximize spatial overlap between ground-truth and predicted segmentation regions:

Sensitivity-Specificity (SS) Loss: Computes the weighted mean squared discrepancy between sensitivity and specificity, assigning higher weight to specificity to counteract severe class imbalance.

Dice Loss: Directly optimizes the Dice Similarity Coefficient (DSC), the most widely used metric for medical image segmentation evaluation.

IoU (Jaccard) Loss: Optimizes the Intersection over Union metric directly during backpropagation to improve spatial alignment.

Tversky Loss: Generalizes Dice loss by assigning asymmetric weights (α and β) to False Positives (FP) and False Negatives (FN), allowing tuning for sensitivity over precision.

Generalized Dice Loss (GDL): A multi-class extension of Dice loss where class weights are set inversely proportional to the square of their label frequencies.

Focal Tversky Loss: Applies a focal exponent parameter to the Tversky index to force network updates toward hard, low-confidence predictions.

Penalty Loss: Extends Generalized Dice loss by introducing explicit penalty constraints for misclassified pixels.

Boundary-Based Loss Functions

Boundary-based loss functions formulate optimization based on spatial distances between predicted boundaries and target contours, frequently combined with region-based losses to enhance boundary fidelity:

Boundary Loss: Computes contour distances across regional boundaries using distance transform maps derived from ground-truth masks:

$$\mathscr{L}_{\text{Boundary}}(\theta) = \int_{\Omega} \phi_G(q) s_\theta(q)\, dq \tag{4.1}$$

where $\phi_G(q)$ represents the distance map derived from ground-truth boundary G, and $s_\theta(q)$ denotes continuous probability outputs produced by the network.

Hausdorff Distance (HD) Loss: Approximates the Hausdorff distance metric directly on soft probability outputs using distance transform functions, training the network to minimize maximum boundary distances.

Compound Loss Functions

By combining complementary loss formulations, compound loss functions—such as Dice-CE, Dice-TopK, or Focal-Dice—can be constructed to simultaneously optimize regional overlap and distribution statistics.

4.2.10 Intersection over Union (IoU) Metric

The Intersection over Union (IoU) metric, also known as the Jaccard Index, measures spatial overlap between two defined regions. Mathematically, it is defined as the area of intersection divided by the area of union:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{4.2}$$

where A and B represent predicted and ground-truth binary masks, respectively. IoU measures prediction accuracy relative to ground-truth annotations, providing a robust index of segmentation performance.

Chapter 5

Conclusions and Future Work

5.1 Conclusion

Based on the experimental evaluations conducted in this thesis, the quantitative performance metric of the developed semantic segmentation system achieved an evaluation score ranging approximately between 60% and 65%. Figure 5.1 illustrates the training and validation loss curves observed across training epochs.

Fig. 5.1

Fig. 5.2

5.2 Future Directions and Recommendations

To further extend this research and improve model performance, the following recommendations are proposed for future study:

1. **Dataset Scaling**: Evaluating and benchmarking the proposed pipeline on a substantially larger dataset comprising over 5,000 images to assess generalization capabilities and comparative performance.

2. **Data Augmentation Techniques**: Applying spatial transformation strategies such as geometric rotations, horizontal/vertical flipping, and affine transformations simultaneously to both input images and their corresponding ground-truth binary masks to artificially expand dataset size and mitigate overfitting.

3. **Comparative Benchmarking**: Conducting comprehensive benchmark comparisons against state-of-the-art baselines and prior literature to rigorously assess system accuracy and diagnostic utility.

4. **Imaging Device Stratification**: Analyzing image acquisitions based on specific imaging equipment parameters and scanner modalities to account for device-specific noise characteristics and intensity variations.

5. **Architectural Justification and Baseline Comparisons**: Evaluating traditional machine learning algorithms (such as Support Vector Machines) alongside deep convolutional models to systematically justify the selection of the U-Net architecture and compare overall classification/segmentation performance.

References

[1] Alias, A. and Paulchamy, B. (2014). Detection of breast cancer using artificial neural networks. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(3).

[2] Alickovic, E. and Subasi, A. (2019). Normalized neural networks for breast cancer classification. In *International Conference on Medical and Biological Engineering*, pages 519–524, Cham. Springer.

[3] American Cancer Society (2013). Breast cancer facts & figures 2013-2014. Accessed: March 9, 2014.

[4] Araújo, M. C. (2014). Interval symbolic feature extraction for thermography breast cancer detection. *Expert Systems with Applications*, 20:305–311.

[5] Baraldi, A. (1999). A survey of fuzzy clustering algorithms for pattern recognition–Part I and II. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6).

[6] Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

[7] Bengio, Y., Goodfellow, I., and Courville, A. (2017). *Deep Learning*, volume 1. MIT Press.

[8] Benzebouchi, N. E., Azizi, N., and Ayadi, K. (2019). A computer-aided diagnosis system for breast cancer using deep convolutional neural networks. In *Computational Intelligence in Data Mining*, pages 583–593. Springer, Singapore.

[9] Berg, W. A., Zhang, Z., Lehrer, D., Jong, R. A., Pisano, E. D., Barr, R. G., et al. (2012). Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA*, 307(13):1394–1404.

[10] Day, M. (2016). Deep learning for financial sentiment analysis on finance news providers.

[11] Early Breast Cancer Trialists' Collaborative Group (1998). Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet*, 351(9114):1451–1467.

[12] Gherghout, Y., Tlili, Y., and Souici, L. (2019). Classification of breast mass in mammography using anisotropic diffusion filter by selecting and aggregating morphological and textural features. *Evolving Systems*, pages 1–30.

[13] Görğel, P. and Kiliç, N. (2009). Mammographic mass classification using wavelet based support vector machine. *Electrical & Electronics Engineering*, 9(1):867–875.

[14] Kaviani, A. (2016). *What is Breast Cancer and What Are Its Symptoms?* Breast Diseases Institute, Tehran, Iran.

[15] Kaymak, S., Helwan, A., and Uzun, D. (2017). Breast cancer image classification using artificial neural networks. *Procedia Computer Science*, 120:126–131.

[16] Khan, S., Islam, N., Jan, Z., Din, I. U., and Rodrigues, J. J. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*.

[17] Khanna, T. (1990). *Foundation of Neural Networks*. Addison-Wesley Series in New Horizons in Technology. Addison-Wesley, New York.

[18] Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8595–8598. IEEE.

[19] Liu, S., Zeng, J., Gong, H., Yang, H., Zhai, J., Cao, Y., and Ding, X. (2018). Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach. *Computers in Biology and Medicine*, 92:168–175.

[20] Moodi, M., Rezaeian, M., Mostafavi, F., and Sharifirad, G. R. (2012). Determinants of mammography screening behavior in Iranian women: A population-based study. *Journal of Research in Medical Sciences*, 17(8):750–759.

[21] Qi, X., Zhang, L., Chen, Y., Pi, Y., Chen, Y., Lv, Q., and Yi, Z. (2019). Automated diagnosis of breast ultrasonography images using deep neural networks. *Medical Image Analysis*, 52:185–198.

[22] Rangayyan, R. M. and Ayres, F. J. (2007). A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin Institute*, 344(3–4):312–348.

[23] Socher, R., Pennington, J., Huang, E. H., Ng, A. Y., and Manning, C. D. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 151–161. Association for Computational Linguistics.

[24] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9.

[25] Tahergorabi, Z. (2014). Breast cancer, a preventable disease. *Journal of Birjand University of Medical Sciences*, 21:126–141.

[26] Vateekul, P. and Koomsubha, T. (2016). A study of sentiment analysis using deep learning techniques on Thai Twitter data. In *Proceedings of the International Conference*, Khon Kaen.

[27] Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., and Xin, J. (2019). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*.

[28] Xiao, Y., Wu, J., Lin, Z., and Zhao, X. (2018). Breast cancer diagnosis using an unsupervised feature extraction algorithm based on deep learning. In *2018 37th Chinese Control Conference (CCC)*, pages 9428–9433. IEEE.

[29] Zeiler, M., Ranzato, M., Monga, R., Mao, M., Yang, K., Le, Q., Nguyen, P., Senior, A., Vanhoucke, V., Dean, J., and Hinton, G. (2013). On rectified linear units for speech processing. In *38th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada.

Appendix A

Source Code for Model Evaluation, Training, and Visualization

This appendix presents the supplementary Python code implementations for model evaluation metrics, custom callbacks, training routines, and performance plotting functions used in this thesis.

A.1 Mean Intersection over Union (IoU) Metric Evaluator

The Jaccard Index, commonly known as Intersection over Union (IoU), quantifies the spatial overlap between predicted segmentation masks ($\hat{Y}$) and ground-truth expert annotations (Y). For a binary mask, IoU is defined mathematically as:

$$\text{IoU}(Y, \hat{Y}) = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{A.1}$$

where TP, FP, and FN represent the number of True Positive, False Positive, and False Negative pixels, respectively.

To evaluate model robustly across varying classification confidence levels, a multi-threshold Mean IoU evaluator is constructed. This function iterates through target decision thresholds $t \in [0.50, 0.95]$ with a step increment of 0.05. For each threshold, continuous sigmoid predictions are binarized and passed to TensorFlow's native streaming metric framework (‘tf.metrics.mean_iou‘). Local session variables are initialized within the TensorFlow execution graph to dynamically accumulate pixel confusion matrices.

```
# Define U-NET Model Evaluator (Intersection Over Union - IOU)
def Mean_IOU_Evaluator(y_true, y_pred):
    prec = []
```

```
4
5       # Iterate across confidence binarization thresholds from
    0.50 to 0.95
6       for t in np.arange(0.5, 1, 0.05):
7           y_pred_ = tf.to_int32(y_pred > t)
8           score, up_opt = tf.metrics.mean_iou(y_true, y_pred_, 2)
9
10          # Initialize local execution graph variables for metric
    accumulation
11          K.get_session().run(tf.local_variables_initializer())
12
13          with tf.control_dependencies([up_opt]):
14              score = tf.identity(score)
15
16          prec.append(score)
17
18      # Return mean metric score across all threshold steps
19      return K.mean(K.stack(prec), axis=0)
20
21  # Instantiate the U-Net architecture
22  model = U_Net_Segmentation()
```

Listing A.1 Python implementation of the multi-threshold Mean IoU metric evaluator

A.2 Epoch Loss History Callback and Model Checkpointing

Monitoring qualitative segmentation progress during gradient descent optimization is vital for detecting early overfitting and vanishing gradient issues. A custom Keras callback, named ‘loss_history‘, is implemented by overriding the ‘on_epoch_begin‘ lifecycle method. At the start of each training epoch, the callback performs forward inference on a fixed validation sample index ($x = 4$) and plots three distinct images:

1. The original input histopathology tissue image.
2. The corresponding expert-annotated binary mask.
3. The continuous sigmoid probability mask generated by the model in its current state.

Furthermore, automated model preservation and early regularization are controlled through ‘ModelCheckpoint‘ and ‘EarlyStopping‘ callbacks. The checkpointer automatically serialized model weight matrices to disk whenever validation loss achieves a new minimum, while the early stopper halts training if validation loss fails to improve over 7 consecutive epochs.

```
# Custom Callback to Visualize Intermediate Predictions per Epoch
class loss_history(keras.callbacks.Callback):
    def __init__(self, x=4):
        super(loss_history, self).__init__()
        self.x = x

    def on_epoch_begin(self, epoch, logs={}):
        # Display raw tissue image sample
        imshow(Train_Inputs[self.x])
        plt.show()

        # Display target ground-truth mask
        imshow(np.squeeze(Train_Masks[self.x]))
        plt.show()

        # Perform forward inference on sample image and display predicted mask
        preds_train = self.model.predict(np.expand_dims(Train_Inputs[self.x], axis=0))
        imshow(np.squeeze(preds_train[0]))
        plt.show()

# Configure naming convention and file paths for model serialization
imageset = 'BCC'
backbone = 'UNET'
version = 'v1.0'
model_h5 = 'model-{imageset}-{backbone}-{version}.h5'.format(
    imageset=imageset, backbone=backbone, version=version
)
model_h5_checkpoint = '{model_h5}.checkpoint'.format(model_h5=model_h5)
```

```
29
30 % Initialize regularization callbacks
31 earlystopper = EarlyStopping(patience=7, verbose=1)
32 checkpointer = ModelCheckpoint(model_h5_checkpoint, verbose=1,
      save_best_only=True)
```

Listing A.2 Implementation of visual callback and model checkpoint configuration

A.3 Model Training and Inference Execution

The optimization phase is executed using the Adam optimizer with binary cross-entropy loss over a maximum of 50 epochs. A validation split ratio of 0.10 (10%) is withheld from training samples to evaluate validation performance and compute early stopping triggers after each epoch. A batch size of 2 is chosen to fit within available GPU memory constraints given the high input dimension of $768 \times 896 \times 3$ pixels.

Following optimization, full-dataset inference is executed on both training and test data arrays. continuous probability predictions are thresholded at $p > 0.5$ to generate binary decision masks.

```
1  # Execute U-Net Model Training Loop
2  results = model.fit(
3      Train_Inputs,
4      Train_Masks,
5      validation_split=0.1,
6      batch_size=2,
7      epochs=50,
8      callbacks=[earlystopper, checkpointer, loss_history()]
9  )
10
11 # Perform Inference on Training Dataset
12 preds_train = model.predict(Train_Inputs, verbose=1)
13 preds_train_t = (preds_train > 0.5).astype(np.uint8)
14
15 # Perform Inference on Unseen Test Dataset
16 preds_test = model.predict(Test_Inputs, verbose=1)
17 preds_test_t = (preds_test > 0.5).astype(np.uint8)
```

Listing A.3 Model optimization and full-dataset inference execution

A.4 Qualitative Result Visualization Code

To perform qualitative validation, random indexing scripts are used to pull sample cases from both training and unseen testing subsets. For each selected index, the script renders a side-by-side comparison of the input image, ground-truth mask, and model-predicted mask.

```
# Select Random Sample Index from Training Set
ix = random.randint(0, len(Train_Inputs) - 1)
print(f"Selected Training Sample Index: {ix}")

print('Train Image (Input):')
imshow(Train_Inputs[ix])
plt.show()

print('Train Mask (Ground Truth):')
imshow(np.squeeze(Train_Masks[ix]))
plt.show()

print('Predicted Segmented Mask:')
imshow(np.squeeze(preds_train[ix]))
plt.show()

# Select Random Sample Index from Unseen Testing Set
iix = random.randint(0, len(Test_Inputs) - 1)
print(f"Selected Test Sample Index: {iix}")

print('Test Image (Input):')
imshow(Test_Inputs[iix])
plt.show()

print('Test Mask (Ground Truth):')
imshow(np.squeeze(Test_Masks[iix]))
plt.show()

print('Predicted Segmented Test Mask:')
imshow(np.squeeze(preds_test[iix]))
plt.show()
```

Listing A.4 Qualitative visualization of training and testing prediction outputs

A.5 Training Progress Curve Plotting

To visualize convergence characteristics and monitor potential overfitting, the history logs stored during model fitting (‘results.history‘) are extracted and plotted using Matplotlib.

```
# Plot Loss Convergence Curves (Training vs. Validation)
plt.figure(figsize=(8, 6))
plt.plot(results.history['loss'], label='Training Loss',
    linewidth=2)
plt.plot(results.history['val_loss'], label='Validation Loss',
    linewidth=2)
plt.title('U-Net Model Loss Progression', fontsize=14)
plt.ylabel('Loss Value (Binary Cross-Entropy)', fontsize=12)
plt.xlabel('Epochs', fontsize=12)
plt.legend(loc='upper right', fontsize=10)
plt.grid(True, linestyle='--', alpha=0.6)
plt.show()

# Plot Mean IoU Accuracy Curves (Training vs. Validation)
plt.figure(figsize=(8, 6))
plt.plot(results.history['Mean_IOU_Evaluator'], label='Training
    Mean IoU', linewidth=2)
plt.plot(results.history['val_Mean_IOU_Evaluator'], label='
    Validation Mean IoU', linewidth=2)
plt.title('Intersection Over Union Progression', fontsize=14)
plt.ylabel('Mean IoU Metric Score', fontsize=12)
plt.xlabel('Epochs', fontsize=12)
plt.legend(loc='lower right', fontsize=10)
plt.grid(True, linestyle='--', alpha=0.6)
plt.show()
```

Listing A.5 Plotting loss and IoU metric curves across training epochs