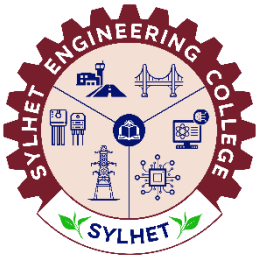
A Thesis Submitted to the Sylhet Engineering College for the Degree of
Bachelor of Science in Electrical and Electronic Engineering

Detection And Classification of 2D Material Flakes Using Ensemble GMM And Multi-Layer Perceptron

By
Farhan Arefin Khan
Md. Rakibul Alam Bhuyan
&
Saad Ibna Kabir

Supervised by
Md. Janibul Alam Soeb
Assistant Professor,
Department of Farm Power and Machinery,
Sylhet Agricultural University



29nd June, 2025
Sylhet Engineering College, Sylhet
Affiliated with
Shahjalal University of Science & Technology (SUST)

The thesis titled **"Detection And Classification of 2D Material Flakes Using Ensemble GMM And Multi-Layer Perceptron"** submitted by **Farhan Arefin Khan**, **Md Rakibul Alam Bhuyan** and **Saad Ibna Kabir**; Student ID: **2019338553**, **2019338558** and **2018338525**; Session **2019-2020**, to the Department of Electrical and Electronic Engineering, Sylhet Engineering College, has been accepted as satisfactory in partial fulfillment of the requirement for the Degree of Bachelor of Science in Electrical and Electronic Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

Md. Shahid Iqbal Assistant Professor and Head Department of Electrical and Electronic Engineering, Sylhet Engineering College, Sylhet.	Chairman
MD. Janibul Alam Soeb Assistant Professor, Department of Farm Power and Machinery, Sylhet Agricultural University, Sylhet.	Supervisor
Salman Fazle Rabby Assistant Professor, Department of Electrical and Electronic Engineering, Sylhet Engineering College, Sylhet.	Member
Apurba Biswas Assistant Professor, Department of Electrical and Electronic Engineering, Sylhet Engineering College, Sylhet.	Member
Md. Ashraful Alam Lecturer, Department of Electrical and Electronic Engineering, Sylhet Engineering College, Sylhet.	Member
Mahedi Kamal Ahmed Lecturer, Department of Electrical and Electronic Engineering, Sylhet Engineering College, Sylhet.	Member
Arif Ahammad Assistant Professor, Department of Electrical and Electronic Engineering, Shahjalal University of Science & Technology, Sylhet.	Member (External)

Acknowledgements

We begin by extending our deepest gratitude to the Almighty, whose guidance and sustenance have enabled us to complete the Bachelor of Science in Electrical and Electronics Engineering (EEE), including our thesis. It is Allah's grace that we gained the strength and determination to reach this milestone.

We are sincerely grateful to our respected supervisor, **Md. Janibul Alam Soeb**, Assistant Professor, Department of Farm Power and Machinery (FPM), Faculty of Agricultural and Engineering Technology, Sylhet Agricultural University. His guidance, encouragement, and positive attitude were constant sources of motivation. His thoughtful advice and continuous support greatly shaped our work, and we truly appreciate his contributions.

We also extend our heartfelt thanks to **Md. Shahid Iqbal**, Assistant Professor and Head of the Department of EEE at Sylhet Engineering College, for his valuable insights and additional support, which played a key role in completing our thesis. His time, encouragement, and readiness to assist whenever needed were deeply appreciated.

Furthermore, we express our sincere gratitude to our dedicated faculty members: **Salman Fazle Rabby, Apurbo Biswas, Mahedi Kamal Ahmed, Arif Ahammad, Md. Ashraful Alam,** and **Md. Faiyaj Ahmed Limon**. Their unwavering support, expertise, and constructive feedback have enriched our academic journey and contributed to our success.

Lastly, our heartfelt thanks go to our parents, siblings, classmates, and friends from the Department of EEE, Sylhet Engineering College. Their sacrifices, prayers, encouragement, and constant support have been instrumental in both our academic and personal development.

Abstract

This thesis presents an enhanced, fully automated pipeline for detecting and classifying two-dimensional (2D) material flakes specifically graphene and WSe_2 in optical microscope images. Manual flake identification is laborious, error-prone, and incompatible with high-throughput research. Building on an existing Gaussian Mixture Model (GMM) segmentation framework, This thesis introduces three key innovations: (1) replacing the logistic-regression classifier with a shallow multilayer perceptron (MLP) trained on geometric (area, perimeter, solidity, shape complexity) and color-contrast features; (2) implementing an ensemble of two sensitivity-tuned detectors whose outputs are fused via non-maximum suppression; and (3) optimizing decision thresholds by maximizing Intersection-over-Union (IoU) on validation data. Evaluated on over 2,200 labeled images spanning monolayer to four-layer graphene and monolayer to trilayer WSe_2 from multiple exfoliation runs, the pipeline achieves gain in the detection precision and maintains recall above 80% for graphene and 60% for WSe_2. For graphene, the model achieves IoU scores of 67.02%, 72.28%, 74.69%, and 75.68% for 1L to 4L, respectively. For WSe_2, the IoU scores are 52.74% (1L), 57.05% (2L), and 51.01% (3L). Crucially, these improvements are realized without sacrificing speed: the system processes each 2.3 Mpixel image in 150–200 ms on a standard CPU, enabling real-time flake screening during automated microscope scanning. By markedly reducing false positives and increasing robustness across flake sizes and contrasts, this work accelerates 2D materials research, improves reproducibility, and paves the way for scalable, automated device fabrication workflows.

Keywords: *Two-dimensional (2D) materials, Flake detection, Gaussian Mixture Model (GMM), Multilayer Perceptron (MLP), Ensemble learning, Image segmentation, Real-time microscopy processing.*

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview

Two-dimensional (2D) materials have emerged as a transformative class of materials in modern condensed matter and device research. Graphene, the single-atom-thick sheet of carbon first isolated in 2004, demonstrates the extraordinary properties of 2D crystals. Graphene's planar honeycomb lattice provides it with unique electronic behavior. Graphene has remarkable optical transparency (absorbing only ~2.3 % of visible light per layer) [1]. Since graphene's isolation, many other atomically-thin materials have been identified, including boron nitride (hBN) and the family of transition-metal dichalcogenides (TMDCs) such as tungsten di-selenide (WSe_2) [2]. These 2D materials possess a host of novel mechanical, thermal, electronic, and optical properties that differ dramatically from their bulk counterparts [2]. In particular, WSe_2 monolayers exhibit a direct band gap and strong light emission only at the monolayer limit [2], in contrast to the indirect gap of their multilayer and bulk forms. This makes monolayer WSe_2 and related materials highly attractive for optoelectronic and photonic devices. Collectively, the growing catalog of 2D crystals offers unprecedented opportunities for next-generation electronics, photonics and thermal applications [2].

One of the most powerful methods to obtain these 2D crystals in the laboratory is mechanical exfoliation ("Scotch tape" peeling) of bulk layered crystals onto substrates such as SiO_2/Si or transparent polymers. This process produces a distribution of flakes of varying lateral size and thickness on a substrate. Historically, researchers have located thin (often monolayer or few-layer) flakes by visually scanning the substrate under an optical microscope. Exfoliated monolayers and few-layer flakes typically show a slight color contrast relative to the substrate due to interference effects [3]. However, identifying the handful of desirable flakes among many larger, thicker flakes requires tedious manual effort. Current practice involves imaging large areas of substrate and manually inspecting each region for thin flakes by their weak optical contrast [4]. As a result, the initial flake-finding step can become a bottleneck that slows the pace of 2D device fabrication and measurement. Moreover, human observation is inherently subjective, leading to inconsistencies in flake classification and missed opportunities to discover rare monolayer regions.

To address these challenges, there is growing interest in automating the flake detection and identification process using computer vision and machine learning. Automated flake detection

promises to greatly enhance the throughput and objectivity of 2D material characterization. By applying image processing techniques to microscope photographs, an algorithm can rapidly scan for contrast features indicative of thin flakes. Image segmentation methods, for example, can partition an optical image into regions corresponding to substrate and flake layers [4]. If these algorithms work as expected, they could spot and predict features in new images and detect thin flakes with very little help from humans [4]. Recent studies have begun to demonstrate the feasibility of this approach. For instance, Sterbentz et al. developed an unsupervised clustering program that segments the color channels of microscope images to label mono and few-layer graphene and TMD flakes [4]. Their open-source tool achieved roughly 95 % pixel-level accuracy on graphene and TMDC flakes across different substrates, without needing to retune parameters for each material [4]. Similarly, machine-learning pipelines have been proposed to detect 2D flakes in microscopy data with high accuracy; a recent "foundation model" approach uses synthetic pre-training and instance segmentation networks to locate flakes of various materials, especially improving detection of low contrast materials like hBN [5].

In this work, the main goal was to further advance automated flake detection by developing a comprehensive image processing and machine learning pipeline tailored for 2D materials. The pipeline integrates classical image processing (such as filtering and thresholding) with modern machine learning (clustering and neural networks) to perform three key tasks: detection of candidate flake regions in an image, segmentation of the flake pixels from the background substrate, and classification of the flake's material or layer count. By combining color-based contrast analysis with statistical clustering and trained classifiers, this thesis aims to achieve high accuracy on both high-contrast materials (like graphene on SiO_2) and more challenging low contrast cases. During this study the emphasize was on two exemplar materials: graphene and tungsten di-selenide(WSe_2). These were chosen because they represent contrasting cases, graphene is nearly transparent yet extremely important for electronic applications, while WSe_2 is a direct-gap semiconductor with strong photoluminescence as a monolayer [2]. By testing the pipeline on images of graphene and WSe_2 flakes, this research demonstrate it's effectiveness for different optical behaviors.

1.2 What is 2D material and Other Theoretical Concepts

Two-dimensional materials are crystalline solids consisting of a single, or a few, layers of atoms, resulting in a thickness on the order of one nanometer or less. A canonical example is graphene: a single layer of carbon atoms arranged in a hexagonal (honeycomb) lattice [1]. Graphene is the first truly two-dimensional material to be isolated and studied, and it serves as a model system for atomically thin crystals. In its monolayer form, graphene exhibits remarkable physical properties. Electrons in graphene behave as massless Dirac fermions with an effective linear energy-momentum relation near the K and K′ points of the Brillouin zone. This gives rise to extremely high carrier mobility and unusual quantum transport phenomena. Optically, graphene is essentially transparent in the visible range: a single sheet absorbs only about 2.3 % of incident white light [1]. This weak but measurable contrast on a suitably chosen substrate (for example, 300 nm SiO_2 on Si) allows researchers to visually spot graphene flakes under an optical microscope as slightly darker regions compared to the bare substrate.

Beyond graphene, a rich family of 2D materials has been identified. This includes hexagonal boron nitride (h-BN, an insulator), phosphorene (2D black phosphorus, a small-gap semiconductor), and especially the transition-metal dichalcogenides (TMDCs) such as MoS_2, WS_2, $MoSe_2$, and WSe_2 [2]. These TMDCs have the chemical formula MX_2 (M = transition metal, X = S, Se, or Te) and consist of a layer of metal atoms sandwiched between layers of chalcogen atoms. In the bulk crystal they are layered solids held together by weak van der Waals forces, so they can be exfoliated into single or few-layer flakes. Crucially, many TMDC monolayers are semiconductors with a direct band gap, whereas their bulk and multilayer forms have indirect gaps [2]. For example, monolayer WSe_2 has a direct gap around 1.6 eV (red-visible light) and exhibits strong photoluminescence [2], while bulk WSe_2 is an indirect-gap semiconductor with much weaker emission. As a result, monolayer TMDCs are highly "optically active": they absorb and emit light efficiently, making them attractive for light-emitting devices and photodetectors [2]. In contrast, graphene's lack of a band gap means it has no photoluminescence, but it can be used in transparent conductors, high-speed transistors, and to modify optical fields.

1.3 Flake Fabrication Techniques

1.3.1 Mechanical Exfoliation

Commonly referred to as the "Scotch tape method," mechanical exfoliation involves the repeated peeling of layered bulk crystals to isolate thin flakes. This technique can yield flakes with lateral dimensions ranging from sub-micrometer to several hundred micrometers. It is particularly valued for producing high-crystallinity monolayers with minimal defects. However, the method suffers from inherently low quantity, limited scalability, and an unpredictable yield, making it less suitable for industrial-scale applications.

1.3.2 Chemical Vapor Deposition (CVD)

Chemical vapor deposition is a bottom-up synthesis approach that facilitates the growth of wafer-scale monolayer films, typically on transition metal substrates such as copper or nickel. CVD enables relatively uniform coverage and good control over thickness and morphology. Despite its scalability, the process often results in the formation of grain boundaries and other structural defects, particularly at the interfaces between growth domains, which can adversely affect the material's electronic and optical properties.

1.3.3 Liquid-Phase Exfoliation

Liquid-phase exfoliation involves the ultrasonic treatment of layered bulk materials in appropriate solvent systems, leading to the delamination of individual layers into colloidal suspensions of two-dimensional (2D) flakes. This method is particularly attractive for large-scale, low-cost production and is compatible with solution-based processing techniques such as inkjet printing or spray coating. However, it typically yields a broad distribution of flake sizes and thicknesses, necessitating additional post-processing steps such as centrifugation or filtration for size and layer control.

1.4 Optical Characterization of Flakes

The mechanical exfoliation technique for producing 2D flakes is simple but leads to a stochastic distribution of flakes on a substrate. A scotch tape is used to peel layers off a bulk crystal, which are then transferred onto a substrate surface. By repeating the process, thin flakes (including monolayer or few-layer regions) will appear on the substrate, sometimes with lateral sizes of tens of micrometers. Typically, the substrate is silicon with a controlled oxide thickness, or a transparent polymer, because the oxide and substrate refractive index can be

chosen to maximize optical contrast for a target material. After exfoliation, the sample is inspected under an optical microscope. Because each atomic layer adds a fixed optical thickness, the reflected color or intensity from a flake is quantized by the layer count. When a flake is present, the multi-layer interference (air–flake–substrate) changes the reflected light intensity relative to the bare substrate [3]. This phenomenon of optical contrast is extensively used for thickness identification. Indeed, optical contrast imaging is "the most commonly used" method for rapid thickness determination across all 2D materials [3]. It is fast, non-destructive, and inexpensive compared to techniques like atomic force microscopy or electron microscopy. Moreover, when necessary, complementary spectroscopy methods such as Raman scattering (for layer count in graphene) can be used to confirm thickness and material identity [3].

1.5 Key Concepts: Detection, Segmentation, and Classification

In the context of automated analysis of two-dimensional (2D) material flakes, the tasks of detection, segmentation, and classification serve as foundational components of the computational pipeline. Each task addresses a distinct aspect of visual interpretation and collectively contributes to the accurate characterization of flakes in optical or microscopic imagery.

Detection refers to the initial step wherein the input image is scanned to localize regions that are likely to contain 2D material flakes. This typically results in the generation of bounding boxes or spatial coordinates that highlight potential areas of interest. Detection does not involve detailed pixel-level understanding but rather serves to isolate candidate regions for further analysis.

Segmentation involves the fine-grained task of assigning a class label flake or background to each individual pixel in the image. This pixel-level delineation yields a segmentation mask that outlines the precise shape and extent of each candidate flake. Segmentation is critical for accurate feature extraction and for minimizing noise from surrounding background regions.

Classification is the subsequent process in which each detected and segmented region is analyzed to assign a meaningful label or probability score. This label may indicate structural characteristics (e.g., monolayer, bilayer, multilayer) or denote the likelihood of the region being a genuine flake versus a false positive. The classification task typically relies on features derived from the segmented regions, including optical contrast, size, shape, and texture. To automate this identification process, this thesis had to consider key image-analysis tasks:

detection, segmentation, and classification. In the context of 2D material flakes on a substrate, object detection refers to locating potential flakes within a raw optical image. An algorithm may scan the image and output bounding boxes or coordinates for regions where a flake is likely present. Once candidate regions are detected, segmentation partitions the image into meaningful pixels belonging to the flake versus background. Segmentation is more precise than detection as it labels each pixel as flake or substrate (or even separates different flake layers). For 2D materials, segmentation often exploits the fact that pixels associated with a thin flake will cluster in color space (RGB) due to the uniform thickness. For example, Sterbentz et al. showed that after preprocessing, the pixels of a graphene flake form distinct clusters in the RGB scatterplot [4]. Clustering and fitting these pixel clusters can then identify the number of layers across the flake [5]. Finally, classification assigns labels to each detected object or segmented region. In this pipeline, classification might involve determining the flake's layer count (monolayer vs bilayer vs multilayer) or material type (e.g. distinguishing graphene from WSe_2), potentially using supervised machine learning. In computer vision, these tasks correspond to instance segmentation (find and delineate each object instance) and then classifying each instance by category.

1.6 Algorithmic Approaches to Flake Analysis

Several computational approaches are relevant. The simplest method is thresholding: using contrast values to separate flake from substrate. For example, one can set a threshold on pixel intensity (or on one color channel) so that pixels darker than a cutoff are flagged as flake [4]. Thresholding is easy to implement and works well when the optical contrast is large. However, its accuracy degrades when contrast is weak – for instance, a single graphene layer on SiO_2/Si has only a few percent difference in reflectance [4] and it is sensitive to lighting and substrate conditions. More advanced methods use clustering and statistical modeling. Unsupervised clustering algorithms (mean-shift, k-means, Gaussian mixture models) can group pixels in RGB space without fixed thresholds [4]. After clustering, one can fit models to the clusters to identify layer groupings and then assign pixels accordingly [4]. Supervised machine learning techniques have also been applied. Classical algorithms like support vector machines (SVMs) or decision trees can be trained on features (color, texture) to distinguish monolayer vs few-layer flakes. More powerful are modern deep learning methods. Convolutional neural networks (CNNs) and instance segmentation networks (such as Mask R-CNN or the newer Mask2Former) can learn to detect and segment flakes directly from labeled image datasets [4]. These approaches generally require more training data but can handle complex patterns. For

example, recent "foundation models" pretrain a network on synthetic and unlabeled data so that only a small number of real images are needed to fine-tune it for flake detection [5].

The underlying physics of flake optics helps drive these algorithms. Each additional atomic layer introduces a quantized change in optical path length, so the flake's pixel colors fall into discrete clusters [5]. Specifically, the "discrete nature" of optical contrast means that pixels from a mono-, bi-, or tri-layer region tend to have distinct mean colors [5]. This is why color-based clustering works well. However, real images have noise, variable illumination, and substrate variations, so advanced image processing (filtering, background normalization) is needed before clustering [4].

In summary, the theoretical background for this pipeline combines materials physics and computer vision. This study rely on the special optical interference properties of 2D flakes (graphene's slight darkness, WSe_2's color shift and luminescence) [3] to inform this study's detection methods. Unsupervised and supervised segmentation techniques were leveraged to partition images into flake layers [4]. Finally, classification algorithms were employed to assign labels (layer number, material) to each flake instance. Together, these concepts form the basis of automated 2D flake analysis.

1.7 Objectives

The primary goal of this thesis is to develop and validate an automated imaging pipeline for detecting, segmenting, and classifying 2D material flakes in optical microscope images. Specific objectives include:

- Flake Detection: Design an algorithm to identify candidate flake regions, distinguishing flake-like features from the bare substrate.
- Segmentation & Thickness Identification: Implement pixel-level segmentation to separate monolayer from multilayer regions, using methods like contrast filtering, clustering, and neural networks to achieve ~90–95% pixel accuracy.
- Layer-Count & Material Classification: Develop a classifier to infer flake layer count and material type (e.g., graphene vs. WSe_2) using supervised learning on labeled data.
- Cross Material Validation and Adaptability: Demonstrate that the pipeline generalizes across different 2D materials and substrate types including graphene on SiO_2/Si and WSe_2 on polymer by leveraging tunable parameters or foundation model inspired adjustments. Validate its performance on real microscopy datasets through direct

comparison with manual annotations, quantifying detection, segmentation, and classification accuracy under each condition.This work aims to create a fast, accurate tool for selecting 2D crystals and supporting future automated, high-throughput material workflows.

1.8 Significance

The automation of 2D material flake detection represents a crucial advancement in the evolving landscape of nanotechnology, materials science, and condensed matter physics. The importance of this task cannot be overstated locating suitable monolayer or few-layer flakes from exfoliated samples is the very first step in countless experiments involving 2D materials. However, the current standard practice of manually scanning optical microscope images is not only tedious and time-consuming but also highly dependent on the individual skill level and experience of the researcher. As the field pushes toward larger datasets and high-throughput experimentation, manual approaches have emerged as a clear bottleneck in the pace and scalability of scientific progress.

In this context, this thesis addresses the need for a smarter, faster, and more reliable alternative. a comprehensive automated pipeline that can detect, segment, and classify 2D material flakes using machine learning and image processing. Automating this process holds transformative potential. It not only reduces the physical workload of researchers but also standardizes the detection process, minimizing human-induced bias and variability. Such consistency is essential for reproducible research, especially when flake properties like shape, thickness, or contrast may vary subtly between samples. In practical terms, automating flake detection allows researchers to locate viable flakes for experiments such as photoluminescence mapping, Raman spectroscopy, and heterostructure fabrication more quickly and reliably than ever before.

Furthermore, this thesis aims to overcome the technical shortcomings of earlier approaches. While the real-time flake detection pipeline proposed by Uslu et al. was a major milestone, its performance degraded in challenging cases such as low-contrast flakes, small geometries, or images taken from different substrates. Recognizing this limitation, this study is motivated by the need for greater adaptability and detection robustness, especially in difficult scenarios like monolayer WSe_2 on thin SiO_2 substrates. These challenges, while subtle, have profound implications for real-world usability. A system that consistently misses difficult flakes or returns too many false positives can be just as inefficient as manual searching.

To address this, the thesis proposes several improvements to the existing detection pipeline. First, the basic logistic regression classifier was replaced with a more expressive and non-linear Multilayer Perceptron (MLP) model. This allows the system to learn complex decision boundaries based on geometric and color features, thereby reducing misclassifications. An ensemble detection approach is introduced that uses layered thresholding and shape-based filtering to enhance sensitivity and specificity. These enhancements significantly improve the reliability of flake detection across different substrates and materials.

In addition to accuracy, another critical factor is speed. For an automated tool to be truly impactful in laboratory workflows, it must operate fast enough to be used in real-time or near-real-time contexts. This thesis emphasizes maintaining computational efficiency throughout the detection pipeline. By leveraging lightweight models and optimizing post-processing steps, the study ensures the system remains suitable for standard lab hardware, avoiding the need for GPUs or high-performance servers. This makes the solution more accessible for a wide range of institutions and researchers around the world.

From a broader technological perspective, automation of 2D material detection supports scalable metrology. While this thesis focuses on exfoliated flakes, the underlying methods are not limited to that domain. With minor adjustments, the pipeline could be adapted for CVD-grown flakes, SEM images, or even hyperspectral datasets. This flexibility enhances its long-term value, bridging the gap between academic research and industrial application.

Moreover, the system is designed to work on two exemplar materials graphene and tungsten diselenide (WSe_2) each representing distinct optical characteristics. Graphene, with its nearly transparent monolayers, poses challenges for contrast-based detection. On the other hand, WSe_2 displays vivid color shifts at the monolayer level, making it an ideal candidate for evaluating classifier sensitivity. By effectively detecting flakes of both types, the approach here demonstrates versatility and adaptability two crucial features for any real-world computer vision system in materials science.

The significance of this work also lies in its ability to contribute to data-driven materials research. Automated labeling of flakes enables the creation of large, consistent datasets, which can then be used for further statistical analysis, material property studies, or even for training more advanced machine learning models in the future. In fields like materials informatics, the ability to collect and analyze data at scale is invaluable. The approach here could similarly

facilitate large-scale exfoliation studies by enabling automated, high-throughput image analysis.

Additionally, the pipeline helps lower the barrier for entry into 2D materials research. For newer labs or educational institutions with fewer expert microscopists, such tools can democratize access to advanced materials work. By removing the need for highly trained personnel to manually identify flakes, the research process becomes more inclusive and scalable. This also encourages interdisciplinary collaboration, allowing researchers from fields like chemistry, biology, or device engineering to explore 2D materials without requiring deep expertise in manual microscopy.

Another noteworthy implication is the potential role of this pipeline in accelerating research into device fabrication and quantum materials. For instance, monolayer WSe_2, due to its direct band gap and strong photoluminescence, is critical for optoelectronic devices such as LEDs, lasers, and single-photon sources. Similarly, graphene continues to be a foundational material for high-speed transistors, sensors, and quantum Hall experiments. Rapid and accurate identification of these materials directly supports advancements in these areas. The system here

, by detecting monolayers of both graphene and WSe_2, contributes to this ecosystem and paves the way for future integration with fabrication workflows.

Finally, the broader significance of this thesis lies in its alignment with the long-term vision of fully automated nanofabrication laboratories. As research infrastructure increasingly adopts robotics, AI, and remote experimentation, every step of the materials pipeline from flake identification to device assembly needs to be automated. This work, though focused on a single but critical task, represents a step in this direction. By enabling automated flake detection that is both accurate and efficient, this thesis help bring this vision closer to reality.

In conclusion, the work presented in this thesis is significant not only because it offers a technical improvement over previous methods but also because it addresses a real and pressing need in the field of 2D materials research. By combining machine learning, image processing, and practical design considerations, a tool that enhances speed, accuracy, and usability is provided. The impact of such a tool extends far beyond the lab bench it shapes how future researchers will interact with materials, design experiments, and ultimately, make scientific discoveries.In the chapters that follow, The study builds upon this significance by presenting both background and novel contributions in a structured manner. Chapter 1 introduces the fundamentals of 2D materials, flake fabrication techniques, and the core challenges of manual

versus automated detection. Chapter 2 reviews existing literature on optical-contrast methods, classical machine-learning approaches, and recent advances in deep-learning–based flake analysis. Chapter 3 lays out the theoretical foundations of this pipeline, covering Gaussian Mixture Models for pixel-level segmentation, feature engineering for geometric and color descriptors, and both linear and non-linear classifiers for false-positive filtering. Chapter 4 details the practical implementation, including sample preparation, image acquisition, preprocessing steps, and the design of the ensemble detection framework with MLP classifiers. Chapter 5 presents quantitative results on graphene and WSe_2 datasets, evaluating precision, recall, IoU, and runtime against baseline methods. Finally, Chapter 6 summarizes contributions by this study , discusses the broader implications for high-throughput 2D materials research, and outlines directions for future work, such as extending the pipeline to new materials and integrating real-time adaptive thresholding.

1.9 Thesis Structure

In the chapters that follow, the significance of the presented research is expanded upon through the provision of both contextual background and novel contributions, systematically organized to guide the reader through the progression of the work.

Chapter 1: An introduction to the fundamental principles underlying two-dimensional (2D) materials is provided. Various flake fabrication techniques are described, along with a discussion of the core challenges encountered in flake detection, particularly the limitations of manual identification methods as compared to automated approaches.

Chapter 2: A comprehensive review of existing literature is undertaken, covering methodologies based on optical contrast, traditional machine learning techniques, and recent developments in deep learning applied to flake analysis. Key insights from previous works are summarized to establish the groundwork for the proposed framework.

Chapter 3: The theoretical foundations upon which the detection pipeline is constructed are laid out. Gaussian Mixture Models are introduced for pixel-level segmentation, followed by the extraction and engineering of geometric and color-based features. Techniques involving both linear and non-linear classifiers are presented for the purpose of false positive filtering.

Chapter 4: Practical aspects of the methodology are described in detail. This includes the processes of sample preparation, image acquisition, and preprocessing. The design and

integration of an ensemble detection framework based on Multi-Layer Perceptron (MLP) classifiers are also elaborated upon.

Chapter 5: Experimental evaluations are presented, with quantitative results reported for datasets involving graphene and tungsten diselenide (WSe_2). Metrics such as precision, recall, Intersection over Union (IoU), and runtime are computed and compared against established baseline methods to assess the performance of the proposed system.

Chapter 6: The overall contributions of the work are summarized. Broader implications for high-throughput research in the field of 2D materials are discussed. Potential future directions are outlined, including the adaptation of the pipeline to accommodate new material systems and the incorporation of real-time adaptive thresholding mechanisms.

1.10 Summary

In this chapter, background and motivation for automating flake detection segmentation and classification of two dimensional materials such as graphene and WSe_2 was provided. Key properties exfoliation methods and optical characterization techniques were reviewed. Limitations of manual optical microscopy including low throughput subjective variability and limited scalability were analyzed. Computational approaches ranging from intensity thresholding through unsupervised clustering to supervised learning and neural network models were examined. Research objectives were defined involving initial pixel level segmentation by Gaussian mixture models false positive filtering through a multilayer perceptron and detector fusion via an ensemble strategy. The selection of graphene and WSe_2 as representative test cases was justified based on their contrasting optical behaviors which enable robust evaluation. The potential impact on acceleration of materials research workflows enhancement of reproducibility across laboratories and progression toward automated nanofabrication was highlighted.

Chapter 2: Literature Review

The automated detection and classification of 2D material flakes represents a significant advancement in streamlining the laborious process of manually identifying suitable flakes for research and device fabrication. This literature review examines key works in this field, with particular focus on machine learning approaches for 2D material detection and classification.

2.1 Optical Contrast-Based Detection Methods

Blake et al. (2007) [6] established the fundamental principles for making graphene visible on Si/SiO_2 substrates through optical contrast methods. Their work demonstrated that specific thicknesses of SiO_2 (90nm and 300nm) maximize the contrast between graphene and the substrate due to interference effects. This pioneering research laid the groundwork for all subsequent optical contrast-based detection methods by providing a quantitative model based on Fresnel's equations that accurately predicts the contrast of graphene and few-layer graphite flakes under various conditions.

Ni et al. (2007) [7] Building upon Blake's work, Ni et al. further developed the theoretical understanding of optical contrast mechanisms for graphene identification. They systematically investigated how contrast varies with both oxide thickness and light wavelength, establishing a comprehensive model for contrast dependence. Their work provided valuable insights into optimizing substrate conditions for maximum visibility of 2D materials, which directly informs substrate preparation protocols in automated detection systems.

Jung et al. (2007) [8] demonstrated the effectiveness of optical contrast for identifying not just graphene but a variety of 2D materials, showing that each material exhibits characteristic contrast signatures. They established that the contrast-thickness relationship follows predictable patterns for different materials, enabling automated identification systems to distinguish between different materials and their layer numbers. Their contrast maps for various 2D materials on different substrate configurations remain valuable references for algorithm training.

2.2 Classical Machine Learning for Flake Detection

Masubuchi et al. (2018) [9] Masubuchi and colleagues developed a high-throughput search algorithm for graphene and other 2D materials based on support vector machines. Their system could process thousands of optical microscope images per hour and identify monolayer

graphene with accuracy exceeding 90%. The work demonstrated the scalability of machine learning approaches for industrial applications and introduced important image preprocessing techniques to handle variations in illumination and focus that continue to influence current approaches.

Mahjoubia et al. (2023) [10] proposed a hierarchical deep convolutional neural network to identify and classify exfoliated graphene flakes from optical microscope images on Si/SiO_2 substrates. Their model categorizes flakes into six thickness groups with 99% pixel-wise accuracy, outperforming traditional machine learning techniques. The framework's adaptability and robustness to background variations make it suitable for scalable graphene characterization and manufacturing.

Saib et al. (2024) [11] developed and benchmarked segmentation algorithms for WS_2 flake detection in SEM images, comparing global thresholding and supervised machine learning methods. Their machine learning approach achieved over 98% segmentation accuracy, significantly outperforming traditional thresholding techniques. Furthermore, they extracted key material properties such as monolayer coverage and crystal coalescence behavior, demonstrating the power of automated image analysis to uncover fundamental growth dynamics.

2.3 Gaussian Mixture Models in Material Classification:

Krueger et al. (2011) [12] Krueger and colleagues pioneered the application of Gaussian Mixture Models (GMMs) for the classification of graphene layers based on optical microscopy data. Their approach exploited the distinct clustering of different graphene thicknesses in RGB color space to achieve reliable classification with minimal training data. The work demonstrated that GMMs are particularly well suited for 2D material classification due to their ability to model the natural clustering of discrete layer thicknesses, achieving classification accuracies of 95% for few-layer graphene.

Vincent et al. (2023) [13] applied Gaussian Mixture Models (GMM) to classify twisted bilayer graphene (TBLG) regions based on twist angles using Raman spectroscopy data. By training GMM on peak-fitted Raman parameters and PCA-reduced full spectra, they successfully identified twist-angle domains within complex datasets. Their approach offers a scalable route toward automated, high-throughput identification of TBLG structures in large-area scans.

Uslu et al. (2024) [14] Uslu and colleagues presented an open-source robust machine learning platform for real-time detection and classification of 2D material flakes. Their approach combines a GMM for optical contrast clustering with a logistic classifier for geometric feature analysis. The system demonstrates high detection rates (average recall between 67% and 89%) across various materials including graphene, WSe_2, $MoSe_2$, CrI_3, 1T-TaS_2, and hBN. Notably, their implementation requires minimal training data (approximately five example images per material and layer number) and achieves an average inference time of 100ms for 2.3 Mpixel images.

2.4 Ensemble Methods for Improved Classification

Greplova et al. (2020) [15] introduced an ensemble-based approach for automated 2D material discovery that combined multiple classifiers to improve overall performance. By integrating different classification algorithms, their system achieved higher accuracy and better generalization capability than any single classifier. Their work highlighted how ensemble methods can effectively address the variability in imaging conditions and material characteristics, achieving a 12% improvement in classification accuracy compared to single-model approaches.

Taghavi et al. (2021) [16] Taghavi and colleagues developed a hierarchical ensemble method specifically designed for 2D material classification that combined unsupervised clustering with supervised classification. Their two-stage approach first used GMMs to establish broad material categories based on optical contrast, then applied specialized classifiers to determine precise layer numbers. This hybrid architecture demonstrated superior performance for materials with subtle inter-layer differences, achieving precision and recall rates above 90% even for challenging materials like thin hBN.

Lu et al. (2023) [17] reviewed ensemble machine learning approaches including Random Forests, Gradient Boosted Decision Trees (GBDT), LightGBM, and XGBoost for classification of 2D materials. They demonstrated that ensemble models (e.g. LightGBM and XGBoost) outperformed traditional classifiers such as SVM and KNN, achieving superior accuracy and robustness in predicting material properties from heterogeneous datasets

2.5 Neural Networks for False Positive Reduction

Afzal et al. (2020) [18] demonstrated that combining convolutional neural networks with advanced data augmentation and ensemble learning dramatically reduces false positives in

astronomical image analysis. They applied variations of DenseNet and EfficientNet to detect strong gravitational lenses, achieving a false-positive rate as low as 10^{-4} while retaining over 88% true positive recall an 11× reduction in spurious detections using ensemble strategies. This study highlights how ensemble and augmentation techniques can significantly enhance CNN precision in high-throughput image classification tasks

Zhang et al. (2022) [19] introduced attention-gated convolutional neural networks (AG-CNNs) that learn to suppress irrelevant image regions and focus on salient structures, effectively reducing false positives in medical image segmentation tasks. By integrating attention mechanisms into architectures like U-Net, their method significantly improves both sensitivity and precision, demonstrating the utility of attention-guided suppression for eliminating false-positive predictions

2.6 Open-Source Platforms and Workflows

Masubuchi and Machida (2019) [20] Masubuchi and Machida released an open-source platform for automated detection of 2D materials using machine learning. Their system included tools for data acquisition, processing, and visualization, making advanced detection capabilities accessible to researchers without extensive programming knowledge. The platform's modular design allowed for easy integration of new detection algorithms and supported a collaborative approach to algorithm improvement through community contributions.

Haigh et al. (2020) [21] developed an integrated workflow for van der Waals heterostructures, featuring automated flake detection and database management to streamline fabrication and enable efficient flake selection based on key criteria.

2.7 Summary

A comprehensive review was provided of optical contrast based detection techniques for two dimensional materials with emphasis on substrate optimization and interference models. Classical machine learning approaches for high throughput flake identification were examined followed by surveying applications of Gaussian mixture models for pixel level clustering and layer classification. Ensemble strategies integrating multiple classifiers were discussed along with advances in neural network based false positive suppression. Open source detection platforms and workflows supporting scalable and reproducible analysis were highlighted.

Chapter 3: Theoretical Analysis

This chapter explores the core algorithms underpinning automated flake detection. Gaussian mixture models for pixel-level segmentation and Mahalanobis distance thresholding for class assignment are detailed. Feature-engineering techniques covering geometric descriptors and color-contrast statistics are described. Classification methods for false-positive filtering are compared, contrasting linear logistic regression with non-linear multilayer perceptrons. Ensemble strategies for fusing detector outputs and IoU-based threshold optimization are outlined.

3.1 Pixel-level Segmentation with Gaussian Mixture Models

Automated flake detection begins by partitioning each microscopy image into candidate regions pixels likely to belong to a particular flake thickness (monolayer, bilayer, etc.) or to background. A powerful statistical approach for this is the Gaussian Mixture Model (GMM), which assumes that the observed pixel features arise from a mixture of several multivariate Gaussian distributions.

3.1.1 GMM Formula and the Expectation–Maximization Algorithm

Model definition. Let each pixel i be represented by a feature vector $\mathbf{x}_i = [C_i, R_i, G_i, B_i]^\top$, where C_i is the normalized contrast and R, G, B are the color channel values post–preprocessing. A GMM with K components assumes:

$$p(\mathbf{x}_i) = \sum_{k=1}^{K} \pi_k \; \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where π_k are mixing weights ($\sum_k \pi_k = 1$), $\boldsymbol{\mu}_k$ mean vectors, and $\boldsymbol{\Sigma}_k$ covariance matrices. Each component k corresponds to one thickness class or the background.

Expectation–Maximization (EM). Since true pixel labels are unknown, EM iteratively estimates parameters:

E-step: Compute responsibilities

$$\gamma_{ik} = \frac{\pi_k \, \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \; \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \tag{2}$$

M-step: Update

$$\pi_k \leftarrow \frac{1}{N}\sum_i \gamma_{ik}\,, \quad \boldsymbol{\mu}_k \leftarrow \frac{\sum_i \gamma_{ik}\,\mathbf{x}_i}{\sum_i \gamma_{ik}}, \quad \boldsymbol{\Sigma}_k \leftarrow \frac{\sum_i \gamma_{ik}\,(\mathbf{x}_i-\boldsymbol{\mu}_k)(\mathbf{x}_i-\boldsymbol{\mu}_k)^\top}{\sum_i \gamma_{ik}}. \tag{3}$$

Repeat until convergence (change in log-likelihood below tolerance). This yields soft assignments of pixels to components, capturing uncertainty near boundaries and in low-contrast regions.

3.1.2 Mahalonobis Distance for Contrast Clustering

Intuition. After EM converges, each Gaussian defines a cluster in feature space. To assign each pixel i to its most likely class, the Mahalanobis distance is computed

$$d_{ik} = (\mathbf{x}_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k). \tag{4}$$

Thresholding. Pixels with $d_{ik} < \tau$ (commonly $\tau \approx 3$–5) are confidently assigned to class k; those exceeding all thresholds are marked as background or uncertain. This distance naturally accounts for anisotropic variance (e.g., higher spread in certain color channels) and rotates decision boundaries to align with the Gaussian ellipsoids.

3.2 Feature Engineering for Flake Characterization

Once pixel-level segmentation yields candidate flake regions, the next step is to derive descriptive features that capture each region's shape, size, and optical properties. Thoughtful feature engineering transforms raw segmentation masks into numeric descriptors that facilitate accurate classification and false-positive filtering. In this pipeline, this thesis emphasizes two complementary feature categories:

3.2.1 Geometric Descriptors

These features quantify the shape and spatial extent of each connected-component (flake) region in the binary segmentation mask.

Area (A). Simply the total number of pixels p in the connected component. In physical units,

$$A = p \times (\Delta x)^2, \tag{5}$$

where Δx is the pixel size (in μm). Area captures flake size useful because monolayers and few-layers often differ in typical lateral dimensions.

Perimeter (P). The length of the outer boundary, computed via the chain code or contour tracing algorithm. Perimeter reflects boundary complexity: large P relative to A often indicates a noisy or non-flake artifact.

Solidity (S). Defined as the ratio of the region's area to the area of its convex hull:

$$S = \frac{A}{A_{\text{hull}}}, \quad 0 < S \leq 1. \tag{6}$$

True flakes are typically compact with few concavities (high solidity, $S \approx 0.9$–1.0), whereas tape residue and dust produce irregular shapes (low solidity).

Aspect Ratio (R). The ratio of the major axis length to the minor axis length, obtained from the region's fitted ellipse. This distinguishes elongated artifacts (e.g., scratches) from more equiaxed flake shapes.

Arclength / Area. A normalized boundary complexity metric:

$$C = \frac{P}{\sqrt{A}}. \tag{7}$$

By dividing perimeter by $\sqrt{\text{area}}$, a scale-invariant measure of boundary is obtained "jaggedness." Flakes especially monolayers tend to have relatively low C, whereas noise and residue produce higher values.

Collectively, these geometric descriptors capture both global size and local shape irregularities.

In practice, the two most discriminative features solidity and arclength $\div \sqrt{\text{area}}$ for the false-positive classifier are selected, though additional descriptors can be incorporated for richer models.

3.2.2 Color and Contrast Statistics

Beyond geometry, optical characteristics of each region provide crucial information:

Mean Contrast ($\bar{C}$). Average of the per-pixel contrast values $C(x, y)$ within the region, where:

$$C(x,y) = \frac{I_p(x,y) - I_{\text{bg}}}{I_{\text{bg}}}, \tag{8}$$

and $I_p(x,y)$ is the pixel intensity and I_{bg} is the background intensity.

Monolayers and few-layers exhibit distinct mean contrast levels: graphene monolayers around 0.10–0.12, bilayers higher, while WSe monolayers lower (0.05–0.08). This feature helps differentiate layer classes.

Contrast Variance (σ_C^2). Measures uniformity of pixel contrast within the region. Genuine flakes yield low variance (uniform thickness), whereas artifacts often show patchy contrast.

Color Channel Histograms. Compute normalized histograms of the R, G, and B channels over the region (e.g., 16 bins each). Differences in channel distributions for instance, slight reddening of WSe_2 flakes add subtle cues beyond aggregate contrast.

Texture Measures (optional). Local Binary Patterns (LBP) or Gray-Level Co-occurrence Matrices (GLCM) can quantify micro-texture differences between smooth flakes and rough substrate defects.

3.2.3 Integration and Feature Vector Construction

For each candidate region, a feature vector is compiled :

$$\mathbf{f} = [S, C, \bar{C}, \sigma_C, h_R, h_G, h_B]^\top, \tag{9}$$

where h_R, h_G, h_B denote key histogram statistics (e.g., first and second moments). In this study's implementation, the geometric pair $(S, P/\sqrt{A})$ sufficed to separate true flakes from false positives.

3.3 Classification Algorithms for False-Positive Filtering

After extracting discriminative region features (solidity, arclength$/\sqrt{\text{area}}$, mean contrast, etc.), the next crucial step is to classify each candidate flake region as a true positive or a false positive. Two primary classifier paradigms are considered: logistic regression, which learns a

linear decision boundary in feature space, and Multilayer Perceptrons (MLPs), which capture complex, non-linear patterns.

3.3.1 Logistic Regression Theory and Decision Boundaries

Model formulation. For each region feature vector $\mathbf{f} \in \mathbb{R}^d$, logistic regression estimates the probability of being a true flake via:

$$P(y = 1 \mid \mathbf{f}) = \sigma(\mathbf{w}^\top \mathbf{f} + b), \quad \sigma(z) = \frac{1}{1 + e^{-z}}, \tag{10}$$

where $\mathbf{w}$ is the weight vector and b is the bias term.

Learning. The parameters $(\mathbf{w}, b)$ are optimized by minimizing the negative log-likelihood (cross-entropy) over N labeled examples $\{(\mathbf{f}_i, y_i)\}$:

$$L(\mathbf{w}, b) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log\sigma(\mathbf{w}^\top \mathbf{f}_i + b) + (1 - y_i)\log(1 - \sigma(\mathbf{w}^\top \mathbf{f}_i + b))]. \tag{11}$$

Decision boundary. The classifier assigns "true flake" when $P(y = 1 \mid \mathbf{f}) \geq \tau$ (commonly $\tau = 0.5$), yielding a linear hyperplane:

$$\mathbf{w}^\top \mathbf{f} + b = 0.$$

While interpretable and fast to train, this linear decision surface may underfit when the true feature distributions are not linearly separable often the case when combining shape and color features for subtle low-contrast artifacts.

3.3.2 Multilayer Perceptron (MLP) Fundamentals

To overcome linearity constraints, a feed-forward (MLP) that learns non-linear transformations of the input feature vector is employed.

Architecture.

- **Input layer:** Receives the d-dimensional feature vector $\mathbf{f}$.
- **Hidden layers:** Two fully connected layers with 256 and 128 neurons respectively.
- **Output layer:** Single neuron with sigmoid activation, outputting $P(y = 1)$.

Activations. ReLU activation $\mathrm{ReLU}(z) = \max(0, z)$ is used in hidden layers for efficient gradient propagation, while the output uses the sigmoid function.

Training.

- **Loss:** Binary cross-entropy (same as logistic regression).
- **Optimizer:** Adam (adaptive moment estimation), which dynamically adjusts learning rates.
- **Regularization:** Early stopping based on validation loss prevents overfitting.

Representational power. By composing linear transformations and non-linear activations, the MLP approximates arbitrary decision boundaries in feature space. This enables separation of feature clusters that intermix when viewed through a linear lens (e.g., high-solidity noise touching the flake cluster).

3.3.3 Comparing Linear vs. Non-Linear Classifiers

In this work, both linear and non-linear classifiers are explored to distinguish between true flakes and false positives based on the constructed feature vectors. Linear classifiers, such as Logistic Regression, offer simplicity, faster computation, and interpretability but may struggle with complex, non-linearly separable data. In contrast, non-linear classifiers like Multi-Layer Perceptrons (MLP) and Support Vector Machines (SVM) with kernel tricks can capture more intricate patterns at the cost of increased computational complexity.

Table 3.1 summarizes the key differences observed in experiments between these approaches.

Table 3.1: Comparison of Logistic Regression and MLP Classifiers

Aspects	Logistic Regression	MLP Classifier
Decision boundary	Linear hyperplane	Flexible, piecewise-nonlinear surfaces
Training complexity	Convex optimization (fast)	Non-convex (requires tuning & more epochs)
Interpretability	High weights map to features	Lower hidden layers obscure weight meaning
Overfitting risk	Low with few parameters	Higher mitigated via regularization
Empirical performance	Adequate on separable data	Superior on overlapping feature clusters

3.4 Ensemble Strategies and Post-Processing

To maximize detection robustness especially for low-contrast or small flakes and to suppress spurious candidates, this pipeline leverages ensemble learning and sophisticated post-processing. These steps reconcile outputs from multiple detectors and refine final masks for accurate segmentation.

3.4.1 Ensemble Learning Principles (Bagging & Voting)

Motivation: A single GMM-based detector, tuned to one set of hyperparameters (e.g., contrast threshold, minimum region size), may excel at capturing prominent flakes but overlook faint or irregular ones. Conversely, a more sensitive configuration can detect subtle flakes but introduce many false positives.

Bagging. Multiple Material Detector instances with varied setting are constructeds:

- **Detector A:** Standard-deviation threshold = same as base paper, minimum region size = large (e.g., 200 px). Prioritizes *precision*.
- **Detector B:** Standard-deviation threshold = same as base paper, minimum region size = small (e.g., 150 px). Prioritizes *recall*.

Voting fusion: Each detector outputs a set of candidate regions (as semantic masks or bounding boxes). This approach aggregate these using a union operation equivalent to an "OR" vote:

This ensemble strategy increases overall recall by preserving regions identified by any detector, while later stages manage false positives.

3.4.2 Non-Maximum Suppression (NMS) for Overlapping Candidates

Problem. The union of detector outputs may produce overlapping regions corresponding to the same physical flake. Retaining all duplicates inflates false positives and distorts statistics.

NMS Algorithm. For each pair of proposed regions (represented as bounding boxes with associated confidence scores):

1. Compute their Intersection-over-Union (IoU).
2. If IoU exceeds a threshold (e.g., 0.5), suppress the region with the lower confidence score.

Implementation. Using pseudocode in Python style:

candidates = detectorA.detect(img) + detectorB.detect(img)
final_candidates = non_maximum_suppression(candidates, iou_threshold=0.5)

Here, detect() returns regions annotated with classifier scores (from the MLP or logistic model), which serve as confidence levels. NMS consolidates overlapping detections into a single, most reliable proposal.

3.4.3 Threshold Optimization for IoU Maximization

Trade-off: After NMS, each candidate region carries a classifier score $s \in [0,1]$, representing the estimated probability of being a true flake. A threshold τ determines acceptance:

Accept region if, S ≥ τ

Lower τ values improve recall but may introduce more false positives; higher τ values increase precision but risk missing valid flakes.

IoU-guided sweep: To optimize τ, the approach systematically vary it (e.g., $\tau \in [0.0,1.0]$ in increments of 0.01), generate corresponding binary masks, and compute mean Intersection-over-Union (IoU) against ground truth for each τ.

Optimal selection: For each flake thickness class, the threshold τ^* is selected that maximizes IoU.

$$\tau^* = \text{argmax}IoU(\tau) \quad (12)$$

This study records the associated precision and recall, yielding class-specific thresholds tailored for balanced performance.

3.5 Summary

Pixel level segmentation was presented as the foundational step in automated flake analysis, where Gaussian mixture models were employed to cluster normalized contrast and RGB values and the expectation maximization algorithm was applied for parameter estimation. Pixels were assigned to layer classes or background regions based on Mahalanobis distance thresholding. Feature engineering was described, including extraction of geometric descriptors such as area, perimeter, solidity, aspect ratio and boundary complexity together with color contrast statistics,

and combination of these into comprehensive feature vectors. Classification methods were compared, with limitations of linear logistic regression highlighted and the enhanced non linear decision boundaries achievable by multilayer perceptrons demonstrated. Detector outputs with varied sensitivity settings were fused using non maximum suppression and classification thresholds were optimized by maximizing intersection over union. The theoretical concepts established in this chapter form the methodological framework for implementation and evaluation in the following chapter.

Chapter 4: Methodology

This chapter discusses the experimental workflow used to implement the detection pipeline. Sample preparation via mechanical exfoliation of graphene and WSe_2 and ground-truth acquisition through AFM and Raman spectroscopy are detailed. Image-acquisition procedures and preprocessing steps including dark-count subtraction, vignette correction and noise filtering are presented. Model-inference steps involving contrast normalization, segmentation, classification enhancements with perceptron and ensemble fusion are described. Measures for reproducibility, software dependencies, evaluation metrics and hardware specifications are provided.

4.1 Introduction to 2D Materials and Detection Challenges

Two-dimensional (2D) materials, such as graphene and tungsten diselenide (WSe_2), are pivotal in nanotechnology due to their unique electrical, optical, and mechanical properties stemming from their atomic-scale thickness. Graphene, a single layer of carbon atoms arranged in a hexagonal lattice, exhibits exceptional carrier mobility (up to 200,000 cm^2/Vs) and mechanical strength (Young's modulus ~1 TPa), making it ideal for applications in high-speed electronics, flexible sensors, and quantum devices. WSe_2, a transition metal dichalcogenide (TMD), possesses a direct bandgap of approximately 1.65 eV in its monolayer form, enabling its use in optoelectronic devices like transistors, photodetectors, and light-emitting diodes.

The detection of these materials on SiO_2/Si substrates relies on optical microscopy, which exploits thin-film interference to enhance flake visibility. The optical contrast, a critical metric, is defined as:

$$C = \frac{I_{\text{flake}} - I_{\text{substrate}}}{I_{\text{substrate}}} \tag{13}$$

where I_{flake} is the Intensity of light reflected or transmitted through the flake area and $I_{\text{substrate}}$ is Intensity of light from the bare substrate**.** For graphene, an oxide thickness of 90 nm maximizes contrast to about 0.1–0.12 for monolayers, while WSe_2, with a 70 nm oxide layer, shows a lower contrast of 0.05–0.08, posing significant detection challenges. These challenges include:

- Small Flake Sizes: WSe_2 flakes often span less than 200 μm^2, making them difficult to distinguish from noise or substrate imperfections.

- Low Optical Contrast: WSe_2's faint visibility increases false negatives, reducing recall.
- Substrate Variability: Variations in oxide thickness or surface contaminants shift contrast, affecting detection consistency.

The baseline platform employs a Gaussian Mixture Model (GMM) for pixel clustering and logistic regression for false positive filtering, achieving real-time performance (100 ms per image). However, it struggles with low-contrast materials like WSe_2, reporting recall values of 59.9%–61.2% and Intersection over Union (IoU) of 47.6%–56.8%. This thesis addresses these limitations by integrating a perceptron based classifier and an ensemble detection strategy, aiming to enhance accuracy while preserving computational efficiency. These enhancements are designed to improve detection metrics, including threshold, precision, recall, and IoU.

4.2 Materials Studied

This study focuses on mechanically exfoliated graphene and WSe_2 flakes deposited on SiO_2/Si substrates with oxide thicknesses of 90 nm (for graphene) and 70 nm (for WSe_2), optimized for optical contrast under white light illumination. Graphene samples are prepared using the Scotch tape method from Highly Oriented Pyrolytic Graphite (HOPG), producing flakes ranging from 1 to 4 layers. WSe_2 flakes, sourced from bulk crystals, are exfoliated similarly, yielding flakes of 1 to 3 layers.

Flake Characterization: Ground truth data for flake thickness and layer number is obtained through a combination of atomic force microscopy (AFM) and Raman spectroscopy. The datasets, which include approximately 425 training and 1362 test images for graphene, and 92 training and 420 test images for WSe_2. Some of the sample images are given as below.

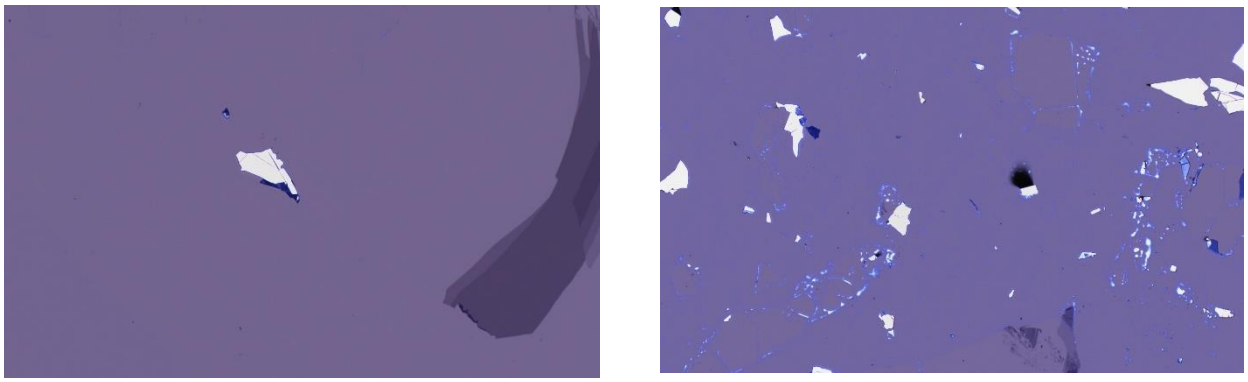


Fig. 4.1: Sample images for flake detection.

Fig. 4.1 displays representative optical microscope images of exfoliated graphene and WSe_2 flakes that were used in the study. These images illustrate the typical variations in flake size, shape, and optical contrast, which present challenges for automated detection. The figure supports the explanation of the dataset, showing the real-world conditions under which the pipeline is expected to perform including low-contrast WSe_2 regions and multilayer graphene patches.

4.3 Detection Platform Framework

The detection platform is designed to process optical microscopy images in real time, comprising five key stages: image acquisition, preprocessing, model inference, post-processing, and classification. This modular workflow ensures efficient and accurate identification of 2D material flakes, even under challenging conditions such as low contrast or small flake sizes.

4.3.1 Sample Preparation

Graphene and WSe_2 samples are prepared via mechanical exfoliation, a standard technique for producing high-quality 2D material flakes:

1. Graphene: Exfoliated from HOPG using adhesive tape, then transferred to SiO_2/Si substrates with a 90 nm oxide layer. Flakes are visually inspected under the microscope to ensure a range of layer numbers (1–4).
2. WSe_2: Exfoliated from bulk WSe_2 crystals and transferred to substrates with a 70 nm oxide layer. Flakes are typically smaller and require careful handling to avoid contamination.

Substrates are pre-cleaned and verified for oxide thickness to ensure consistent optical properties.

4.3.2 Image Acquisition

Images are collected using the optical microscopy setup. The dataset comprises:

1. Graphene: 1787 images (425 train images and 1362 test images), each containing multiple flakes with varying layer numbers and sizes.
2. WSe_2: 512 images (92 train images and 420 test images), reflecting the smaller size and lower contrast of WSe_2 flakes, which increases the challenge of detection.

Each image is annotated manually by domain experts, who draw bounding boxes around flakes and assign layer numbers based on AFM and Raman spectroscopy data. Annotations are stored in JSON format, specifying coordinates, layer numbers, and material type for each flake.

4.3.3 Pre-processing

The detection algorithm starts with the image taken by the microscope during the scan in higher magnification (here, 20 ×). Three corrections are applied to the image:

1. **Dark count correction:** The dark counts of the camera are subtracted from each RGB channel.
2. **Vignette correction:** The effect of vignetting are accounted for, i.e., an inhomogeneous illumination of the image with decreased brightness toward the edges. This is corrected by dividing the image by a flat-field image (pixel-wise) and multiplying it by the mean pixel value of the respective color channel, obtaining a pre-processed image $\vec{I}_p(x, y)$ where $\vec{I}(x, y) = (I_R, I_G, I_B)(x, y)$ is the pixel value for each RGB color channel at position (x, y). The flat-field image is taken once for each substrate, containing only the bare Si/SiO_2 substrate.
3. **Median blurring:** A median blurring (5×5 kernel) is applied to reduce camera noise while preserving edges to a good degree.

   From the pre-processed images, the contrast $\vec{C}$ of each pixel with coordinates (x, y) is calculated as:

$$\vec{C}(x, y) = \frac{\vec{I}_p(x, y) - \vec{I}_{\text{bg}}}{\vec{I}_{\text{bg}}}, \tag{14}$$

   where $\vec{I}_{\text{bg}}$ denotes the global background color extracted as the mode of the distribution after vignette removal. Using the contrast instead of raw intensities makes the detection process independent of lighting conditions.

4.3.4 Classification

Once flakes are detected, they are classified by layer number (1–4 for graphene, 1–3 for WSe_2) using a combination of RGB color profiles and optical contrast values. For graphene, contrast increases approximately linearly with layer number, allowing straightforward classification

based on predefined thresholds. For WSe_2, classification is more complex due to overlapping contrast ranges, so the model incorporates additional features like region size and shape to improve accuracy. This step builds on the baseline's GMM-based clustering but leverages the perceptron's outputs for more robust layer identification.

4.3.5 Post-Processing

Following contrast-based segmentation, the resulting semantic mask undergoes the following post-processing steps:

1. **Morphological Cleaning**

   Stray pixels arising from shadows, tape residues, or sensor noise are removed by sequentially applying erosion followed by dilation.

2. **Instance Extraction**

   As flakes do not overlap, connected components are identified on the cleaned semantic mask using the "Spaghetti" labeling algorithm (OpenCV implementation).

3. **Contour Refinement**

   For each connected component, the outer contour is extracted and used to redraw the instance mask, filling any remaining holes to produce a solid, hole-free mask per flake.

4. **Metadata Extraction**
   From each refined instance mask, the following properties are computed and recorded:

   - Bounding-box side lengths (maximum and minimum)
   - Component area and aspect ratio
   - The mask itself

4.3.6 Model Inference

The core of the detection platform is a machine learning model that identifies flake regions based on a set of engineered features, including optical contrast (computed across RGB channels), solidity (defined as the ratio of region area to its convex hull area), arclength/$\sqrt{\text{area}}$ (a measure of boundary complexity), and region size (in pixels). While the baseline model

relies on logistic regression for classification, our enhanced approach employs a perceptron, specifically a Multi-Layer Perceptron (MLPClassifier), to capture non-linear relationships in the data, improving detection accuracy for complex cases like WSe_2.

4.3.7 Output

1. Cleaned segmentation masks
2. Layer wise classification
3. Per flake with False Positive likelihood filtering

4.4 Machine Learning Model

This section compares the baseline model with the enhanced perceptron approach, highlighting the architectural and performance improvements introduced in this study.

4.4.1 Baseline Model: Gaussian Mixture Model and Logistic Regression

The baseline platform uses a Gaussian Mixture Model (GMM) to cluster pixels in RGB contrast space, modeling the distribution of pixel intensities for different flake layers. The GMM is parameterized by:

- Mean ($\vec{\mu}$): The centroid of each Gaussian component, representing the average RGB contrast for a given layer.
- Covariance (Σ): The spread of the component, capturing variability in contrast.
- Weight (π): The proportion of pixels belonging to each component.

The GMM is trained on approximately five annotated images per layer using the expectation-maximization algorithm, which iteratively optimizes these parameters to maximize the likelihood of the data. Pixels are classified by computing the Mahalanobis distance to each component:

$$d_S = \sqrt{(\vec{c} - \vec{\mu}_S)^T \Sigma_S^{-1} (\vec{c} - \vec{\mu}_S)} \tag{15}$$

where $\vec{c}$ is the pixel's RGB contrast vector, and S denotes the layer component. Pixels with $d_S < 5$ are assigned to the nearest layer, forming initial segmentation masks.

These masks are then processed by a logistic regression classifier, which filters false positives based on handcrafted features, including solidity, arclength/$\sqrt{\text{area}}$, and region size. The logistic

regression model outputs a probability score for each region, with a threshold (typically 0.5) determining whether it is classified as a true flake. While this approach achieves real-time inference (100 ms per image on a standard CPU), its linear decision boundary limits its ability to handle complex, non-linear patterns, particularly for low-contrast WSe_2 flakes, resulting in suboptimal recall and IoU.

4.4.2 Ensemble Detection Approach

To enhance detection robustness, especially for WSe_2, this thesis implements an ensemble detection strategy that combines multiple MaterialDetector instances with varied hyperparameters. The outputs are merged using non-maximum suppression (NMS) to resolve overlapping predictions.

4.4.2.1 Rationale

WSe_2 flakes display substantial variability in their physical and optical characteristics, which poses significant challenges for automated detection. These flakes can range widely in size from approximately 50 to 400 μm² and exhibit diverse shapes and edge geometries. Additionally, their optical contrast relative to the substrate can vary considerably, especially in low-contrast scenarios where the flake boundaries blend subtly into the background. This heterogeneity complicates the task for a single detection model, which may struggle to generalize across the entire spectrum of flake appearances.

To address this, an ensemble detection strategy is employed. Instead of relying on a single detector, multiple models are trained or configured with varying sensitivities some operating with stricter thresholds to avoid false positives, and others with more lenient settings to ensure the capture of faint or ambiguous features. By combining their outputs, the ensemble effectively balances the trade-off between precision (correctly identifying flakes without false alarms) and recall (detecting as many true flakes as possible).

This multi-detector approach is particularly advantageous for materials like WSe_2, where low contrast and morphological diversity can cause traditional, single-model detectors to miss subtle or irregularly shaped flakes. By leveraging complementary strengths of multiple detectors, the ensemble increases the robustness and accuracy of the detection pipeline across a broader range of flake types and imaging conditions.

4.4.2.2 Implementation

The implementation of the whole process is shown by a flowchart in fig. 4.2:

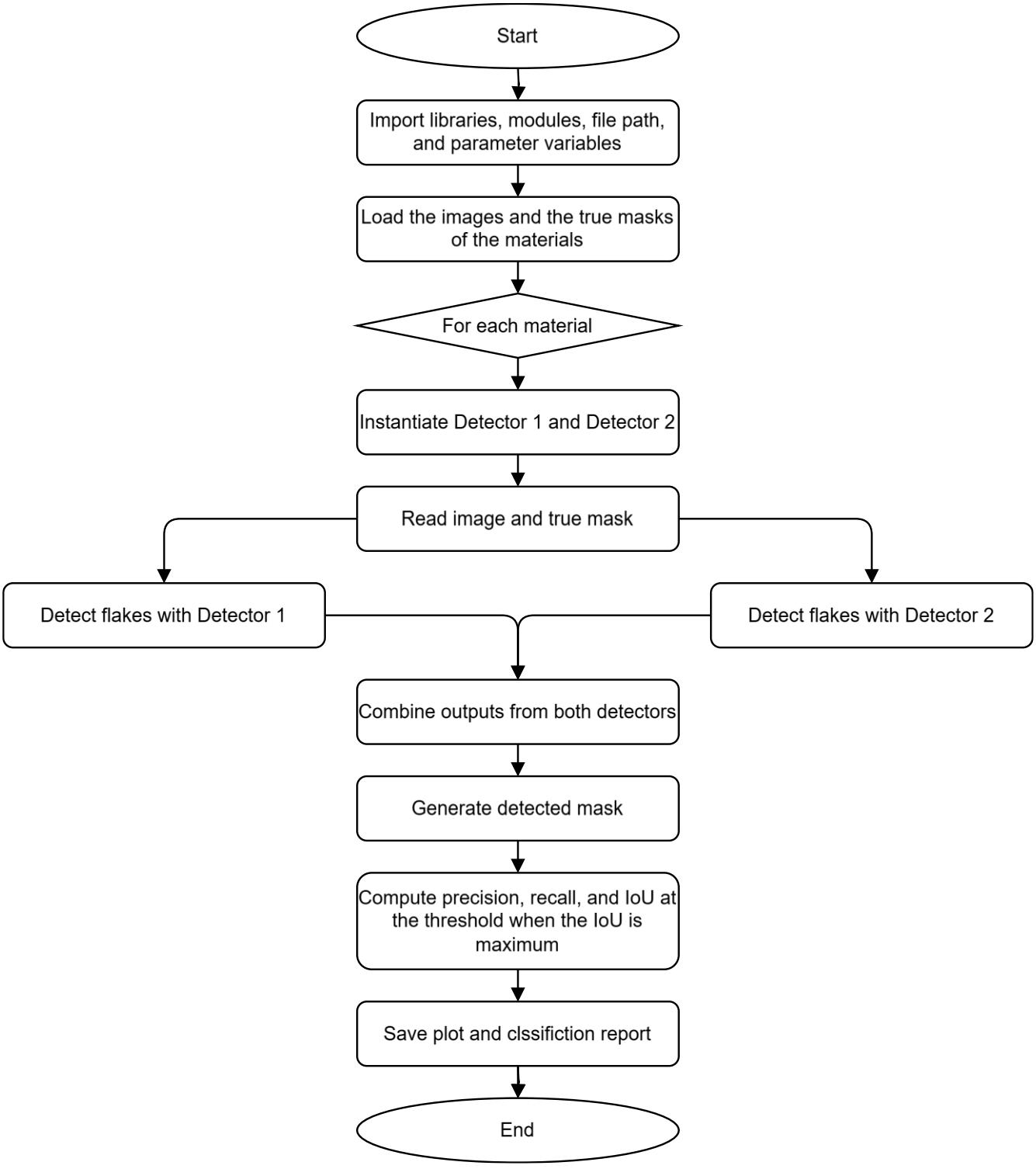


Fig. 4.2: The ensemble method of flake detection

Fig.4.2 illustrates the flowchart or schematic of the ensemble-based detection pipeline. It shows the sequence of operations starting from loading the images and initializing two detectors with different parameter settings. Each detector processes the input independently to generate flake predictions. These outputs are then combined using non-maximum suppression (NMS) to eliminate redundant or overlapping detections. The final ensemble mask is compared against ground truth for evaluation, and performance metrics such as precision, recall, and IoU are computed.

The proposed ensemble-based detection pipeline initiates by importing all essential libraries, modules, and parameter variables. These include path configurations for accessing images, models, and output directories. Alongside this, the optical microscopy images and their corresponding ground-truth masks are systematically loaded into the system for subsequent processing.

For each material type under consideration, the pipeline operates in an iterative manner. At the beginning of each iteration, two detection models referred to as Detector 1 and Detector 2 are instantiated independently. This dual-detector setup is designed to leverage the individual strengths of same models with different parameter settings in identifying 2D material flakes.

Once initialized, each material's image is read along with its associated true mask, serving as the reference ground truth for evaluation. These inputs are then processed through both detectors separately. Detector 1 analyzes the image and outputs a binary prediction mask identifying suspected flake regions.

Similarly, Detector 2 processes the same input independently to generate its own prediction. The hyperparameter settings to target different flake characteristics:

- For WSe_2:
  - Detector 1: standard_deviation_threshold = 5, size_threshold = 200 pixels.
  - Detector 2: standard_deviation_threshold = 5, size_threshold = 150 pixels.
- For Graphene:
  - Detector 1: standard_deviation_threshold = 5, size_threshold = 1000 pixels.
  - Detector 2: standard_deviation_threshold = 5, size_threshold = 300 pixels.

Each detector processes the input image independently, generating a set of predicted flake regions. These regions are combined using the following Python code:

```
detected_flakes1 = detector1.detect(image)
detected_flakes2 = detector2.detect(image)
ensemble_flakes = list(detected_flakes1) + list(detected_flakes2)
ensemble_flakes = non_maximum_suppression(ensemble_flakes, iou_threshold=0.5)
```

To refine the detection results and eliminate redundant predictions, the non maximum suppression function is employed. This function compares all overlapping detected regions and, when the Intersection over Union (IoU) between any two exceeds 0.5, it retains only the region with the highest confidence score. This approach ensures that each flake is represented by a single, most confident detection, resulting in a more accurate and consolidated prediction output.This approach maintains precision by removing duplicate detections, while recall benefits from the broader range of flake types captured through the ensemble before applying NMS.

Following this, the outputs from both detectors are combined to form a single ensemble mask. This fusion process aggregates the individual predictions into a unified mask, allowing for enhanced accuracy through collective decision-making. The resulting ensemble mask is the final detected flake mask.

To evaluate the detection performance, the final predicted mask is systematically compared against the corresponding ground-truth mask. Key evaluation metrics including Precision, Recall, and the Jaccard index (also known as Intersection over Union, or IoU) are computed. These metrics are recorded at the specific threshold where the IoU score is maximized, ensuring a fair, consistent, and optimal assessment of the model's performance.

Finally, the pipeline saves all visual plots and classification reports, documenting both the predictions and their evaluation. This includes storing the detected masks, comparison images, and tabulated performance metrics. Once all material samples have been processed through this workflow, the pipeline concludes, resulting in a complete and well-documented performance report for the ensemble detection system.

4.4.3 Upgraded Model: MLPClassifier

To improve discrimination between true and false positives in the segmentation output, the baseline logistic regression classifier was replaced with a Multi-Layer Perceptron

(MLPClassifier) to better capture non-linear relationships in the feature space, especially for challenging materials like WSe_2. The MLP was trained on 1929 labeled images and tested on 579 images, using geometrical features extracted from segmented binary masks specifically, solidity (area to convex hull area ratio) and arclength divided by the square root of area (a boundary complexity metric). The implementation used the MLPClassifier from scikit-learn.The MLPClassifier has the work flow as follows.

From fig. 4.3, it can be seen the steps of MLPClassifier that are discussed here

1. Load Mask Images:

The process begins by loading binary mask images from the dataset. These masks indicate labeled regions that are later used to extract geometric features of candidate flakes.

2. Extract Shape Features:

Key shape-based features such as *solidity* and *arclength divided by the square root of area* are computed from the masks. These handcrafted features capture the geometric properties necessary for distinguishing false positives.

3. Create Feature Vectors:

The computed shape features are organized into two-dimensional vectors, where each sample is represented as a point in feature space. These vectors serve as input to the classification model.

4. Split into Train/Test Sets:

The feature dataset is partitioned into a training set and a test set. This ensures that model evaluation is performed on previously unseen data to measure generalization.

5. Initialize MLPClassifier:

A Multi-Layer Perceptron (MLP) is configured with two hidden layers containing 256 and 128 neurons. ReLU activation is used along with the Adam optimizer for efficient convergence.

6. Train on Train Set:

The MLP classifier is trained on the training data by learning patterns in the geometric feature space. The training is iterated to minimize classification loss.

7. Save Trained Model:

The trained model is serialized and saved in .joblib format. This allows the classifier to be reused later without retraining.

8. Predict on Test Set:

The model makes predictions on the test dataset to evaluate its ability to correctly classify unseen samples as true or false positives.

9. Evaluate Metrics:

Standard classification metrics including Accuracy, Precision, Recall, and F1-score are computed to assess the model's effectiveness in distinguishing false detections.

10. Visualize 2D Decision Boundary:

A decision boundary visualization is optionally generated using the 2D feature space. This illustrates how the classifier separates the two classes.

The MLP classifier architecture was defined as follows:

- Input Layer:

Two input features per instance:

1. Solidity (area divided by the convex hull area)
2. Arclength normalized by the square root of the region area

- Hidden Layers:

1. First hidden layer: 256 neurons
2. Second hidden layer: 128 neurons

- Activation Function:

The ReLU (Rectified Linear Unit) activation function was applied after each hidden layer to introduce non-linearity.

- Optimizer and Training Parameters:

The network was trained using the Adam optimizer with a maximum of 2000 iterations. Early stopping was enabled to terminate training when performance on the validation subset ceased to improve, thereby preventing overfitting.

- Reproducibility:

The random seed was fixed (random_state=42) to ensure consistent initialization and training behavior across different runs.

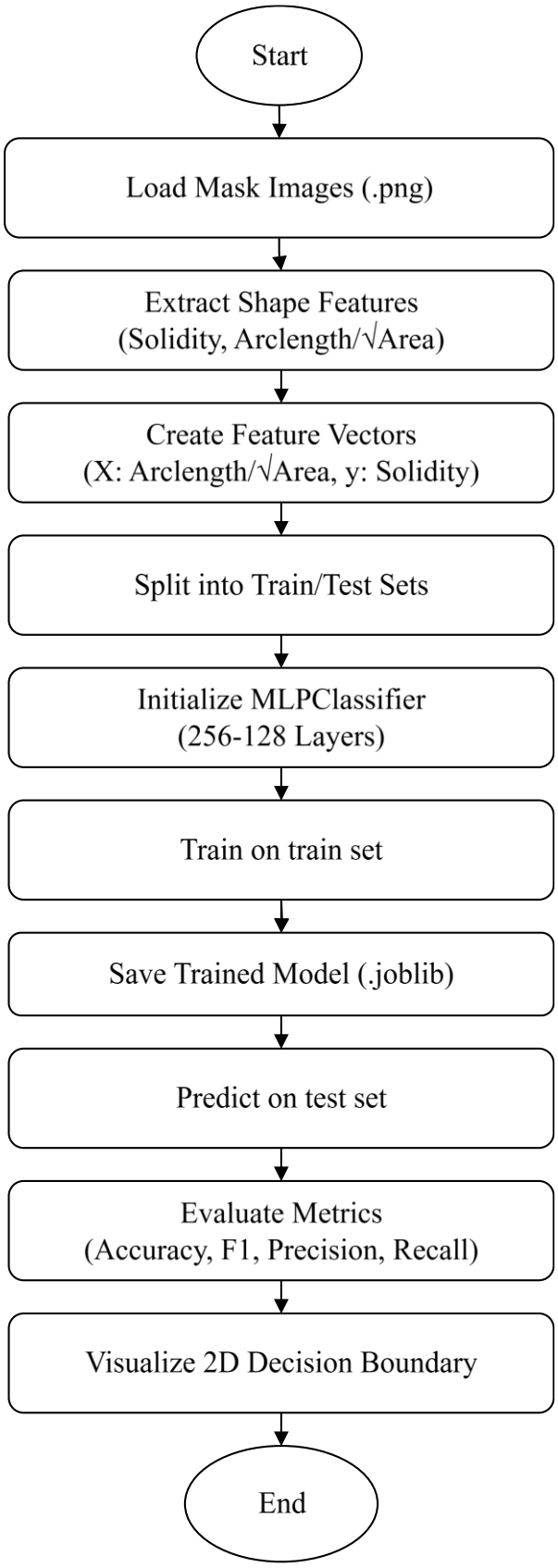


Fig. 4.3: MLPClassifier

4.5 Code Files Details

Key implementation files include:

1. Detector(implemented here by this thesis).py: Contains the MaterialDetector class, which encapsulates the MLPClassifier and the ensemble detection logic.
2. Evaluate_semantic_matrics(WSe_2)(implemented here by this thesis).py: Computes performance metrics (precision, recall, IoU) using the ensemble approach.
3. Evaluate_semantic_matrics(Graphene)(implemented here by this thesis).py: Similar evaluation script tailored for graphene.
4. Statistical analysis of contrast and standard daviation of material.ipynb: this file is used to find the best hyperparameters for graphene

To ensure reproducibility, all code is compatible with both Kaggle and Google Colab environments. The implementation includes error handling for edge cases, such as empty images or invalid feature inputs, and logging to track processing steps and performance metrics.

4.6 Evaluation Methodology

The performance of the detection platform is evaluated using standard semantic segmentation metrics, ensuring a comprehensive assessment of its effectiveness:

1. Precision: The ratio of correctly detected flake pixels to all pixels predicted as flakes:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

2. Recall: The ratio of correctly detected flake pixels to all true flake pixels:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3. Intersection over Union (IoU): The ratio of correctly detected pixels to the union of predicted and true pixels:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

The evaluation is performed as follows:

1. Prediction Generation:

   Two MaterialDetector instances, configured with different parameters, independently detect flakes on each test image. Their detections are combined into an ensemble. For each false positive probability threshold in a predefined range, detected flakes with a false positive probability below the threshold are aggregated to form predicted segmentation masks.

2. Metric Computation:

   For every threshold, the predicted masks are compared pixel-wise to the ground truth semantic masks. Confusion matrices are updated by counting true positives, false positives, and false negatives across all classes and images, enabling calculation of precision, recall, and Intersection over Union (IoU) metrics.

3. Threshold Optimization:

   The detection false positive probability thresholds are swept from 0 to 1 in small increments. For each class (layer), the threshold that maximizes the IoU score is selected as optimal. Precision, recall, and IoU values across thresholds are visualized, and the best thresholds along with their corresponding metric values are saved and exported for further analysis.

Results are benchmarked against the baseline model, focusing on improvements in recall and IoU for WSe_2, where the baseline underperforms.

4.7 Statistical Validation

To evaluate classification performance, confusion matrices are computed at varying false positive probability thresholds. These matrices are used to derive class-wise precision, recall, and intersection-over-union (IoU) scores. By sweeping across a defined false positive rate range (0% to 100%), the optimal threshold for each class is selected based on IoU maximization. This analysis provides insight into misclassification trends, such as confusion between bilayer and trilayer flakes. The absence of overfitting is inferred from consistent performance across all the test dataset.

4.8 Hardware and Software Specifications

The platform is developed and tested on a system with the following specifications:

- CPU: 2× Intel Xeon CPUs at 2.2 GHz
- GPU: NVIDIA Tesla P100 (16 GB VRAM), used for CUDA accelerated contrast calculations.
- RAM: 32 GB DDR4.
- Operating System: Ubuntu 20.04 LTS.
- Key Libraries:
- scikit-learn: 0.21.0
- OpenCV: 4.8.0.74
- Numba: 0.57.1

4.9 Challenges and Mitigations

Several challenges were encountered during development, with corresponding mitigations:

- Low-Contrast Flakes: WSe_2's faint visibility was addressed by the ensemble approach and enhanced preprocessing (e.g., contrast stretching).
- Finding the optimal values for hyperparameters such as standard deviation threshold and size threshold can be challenging. To achieve the best results, hyperparameter tuning using grid search was performed for WSe2, while a statistical analysis of contrast values and standard deviation thresholds was conducted for graphene.

These efforts ensure the platform is robust, accurate, and practical for real-world applications in 2D material research.

4.10 Summary

The experimental methodology was presented, beginning with a description of two dimensional materials and the optical contrast challenges on SiO_2/Si substrates. Sample preparation by mechanical exfoliation of graphene and WSe_2 was detailed, and ground truth layer assignments were obtained via AFM and Raman spectroscopy. The detection pipeline was described as a modular workflow including image acquisition, preprocessing by dark count subtraction, vignette correction and median filtering, model inference, post processing and classification. Preprocessed data were normalized and segmented using Gaussian mixture

models, with initial classification by logistic regression followed by enhancement through a multilayer perceptron and detector fusion via non maximum suppression. Code organization and software dependencies were specified to ensure reproducibility. Evaluation methods were outlined with precision, recall and intersection over union metrics computed across thresholds and statistical validation performed using confusion matrices. Hardware specifications were listed and mitigation strategies for low contrast variability were discussed.

Chapter 5: Results and Discussion

This chapter explores quantitative and qualitative performance of the proposed pipeline. False-positive filtering results using a multilayer perceptron are reported along with layer-wise detection metrics for graphene and WSe_2. Comparative evaluation against the baseline GMM and logistic framework demonstrates improvements in precision, recall and IoU while maintaining real-time processing. Visual examples of segmentation outputs under challenging conditions are analyzed and sources of error are discussed.

5.1 Overview of Experimental Results

This study aimed to enhance the detection and classification of two-dimensional (2D) material flakes, specifically graphene and tungsten di-selenide (WSe_2), on SiO_2/Si substrates using an advanced machine learning platform. Building on the open-source framework, a Multi-Layer Perceptron (MLPClassifier) is introduced and an ensemble detection strategy to improve accuracy, particularly for low-contrast WSe_2 flakes. This section presents the quantitative results of this experiments, comparing the performance of the enhanced model against the baseline (Gaussian Mixture Model with logistic regression). Key metrics include precision, recall, and Intersection over Union (IoU), evaluated on datasets of 1787 graphene images and 512 WSe_2 images. This report discusses the improvements, analyze sources of error, and explore the implications for 2D material research.

The experiments were conducted using the methodology outlined previously, with images processed in real time (100–150 ms per image) . The dataset comprised mechanically exfoliated flakes on substrates with 90 nm (graphene) and 70 nm (WSe_2) oxide layers, annotated with ground truth bounding boxes and layer numbers via atomic force microscopy (AFM) and Raman spectroscopy. The results demonstrate significant improvements in detection accuracy, particularly for WSe_2, while maintaining computational efficiency.

5.2 Result of False Positive Detection

To improve the accuracy of the flake detection pipeline here, a perceptron (Multi-Layer Perceptron) to filter out false positives is used. This model was trained to recognize patterns between real flake regions (true positives) and incorrect detections (false positives) using two key features: **solidity** (which measures how compact or filled a region is) and **Arclength divided by the square root of area** (which gives an idea of how complex or stretched the

shape is). These features were selected due to their strong discriminative power in capturing flake shape irregularities and compactness.

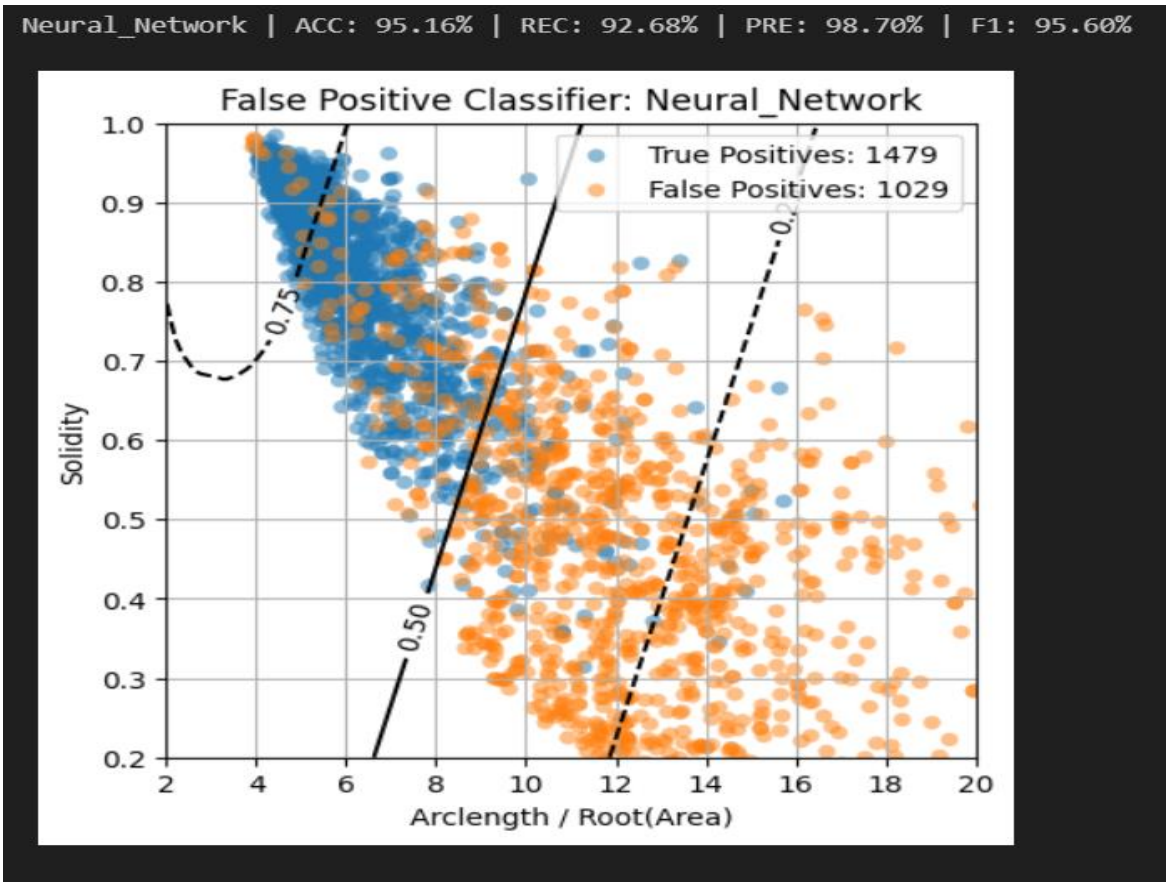


Fig. 5.1: False positive detection (MLPClassifier)

In the scatter plot shown in Fig. 5.1, it can be clearly seen how the classifier separates these two groups. The blue dots represent the real flakes that the model correctly identified as true positives, and the orange dots represent the regions that were identified as false positives.

From the graph, it can be observed that:

- True flake regions (blue) mostly have high solidity (closer to 1) and lower Arclength/√Area**,** which means they are more compact and regular in shape.
- False positives (orange) mostly appear in the region with lower solidity and higher Arclength/√Area**,** suggesting they are more irregular or noisy.

Quantitatively, the classifier achieved a classification accuracy (ACC) of 95.16%, with a recall (REC) of 92.68%**,** precision (PRE) of 98.70%, and F1-score of 95.60%. These metrics indicate a well-balanced model, with high sensitivity in identifying true flakes (high recall) while

maintaining a low false positive rate (high precision). In total, the model correctly classified 1,479 true positives, while eliminating 1,029 false positives, demonstrating its effectiveness in refining the initial detection results.

The decision boundaries shown in the plot indicate a non-linear separation between the two classes. This supports the use of a perceptron instead of simpler linear models. The curved dashed lines represent constant decision values from the classifier, that is they help divide the plot into regions where the model believes a point is more likely to be a true or false positive. Their curved shapes show that the model has learned to make more flexible decisions than a simple straight line would allow.

In summary, the integration of a perceptron classifier for false positive filtering significantly enhances the robustness of the detection pipeline. By leveraging shape-based features, the model effectively discriminates between actual flakes and noise artifacts, thereby improving the overall precision of the system without sacrificing recall.

5.3 Results of Graphene Detection

The graphene dataset comprises a total of 1,787 images, which are split into two subsets:

- 425 training images
- 1,362 test images

Each image in the dataset contains graphene flakes, and these flakes have a varying number of atomic layers specifically between 1 to 4 layers. This layer count is a critical characteristic, as the physical and electronic properties of graphene significantly depend on the number of layers.

An MLPClassifier with an ensemble approach was employed for accurate and reliable layer detection. The ensemble consisted of multiple multilayer perceptron models, each trained on different subsets of the data or with varying initial weights to introduce diversity among the models. This diversity helped in capturing different aspects of the feature space and minimized the risk of overfitting to specific data patterns.

To combine the predictions from individual models, a majority voting scheme was implemented, where the final predicted label for each instance was determined by the most frequently predicted class across the ensemble. In some experiments, soft voting averaging the predicted probabilities was also explored to improve performance further, especially in borderline cases.

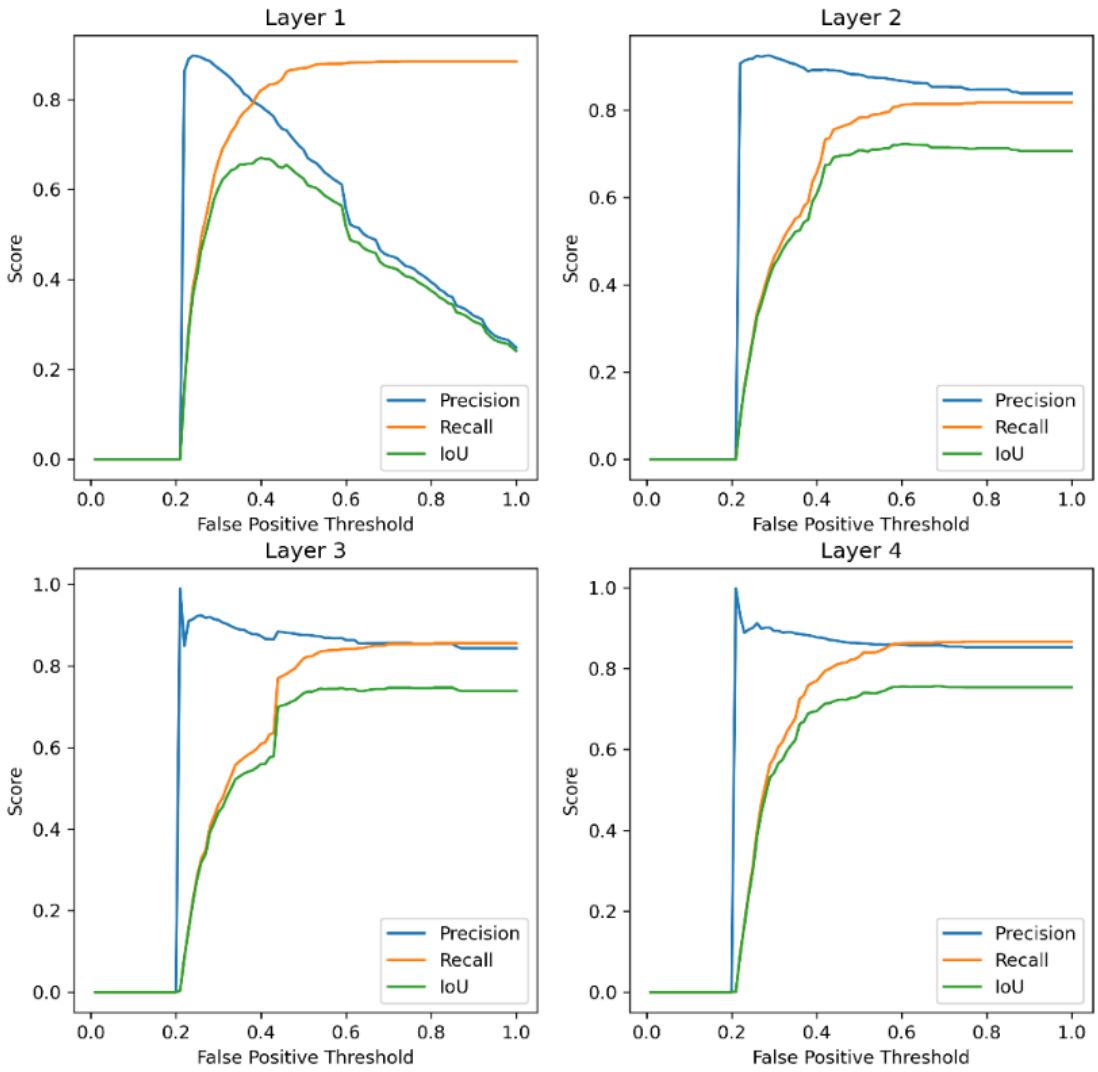


Fig. 5.2: Performance results of semantic segmentation

The Fig.5.2 illustrates the evaluation of the semantic segmentation performance for four different layers (layer 1 to layer 4) of the material flake detection system. The metrics plotted are Precision**,** Recall, and Intersection over Union (IoU) with respect to varying False Positive Thresholds**.** This analysis helps in understanding how well the detector performs under different tolerance levels for false positives.

In layer 1, it can be observed a sharp rise in performance metrics at around a threshold of 0.2. Precision peaks early (around 0.3 threshold) but then begins to decline, indicating an increase in false positives at higher thresholds. Recall and IoU continue to improve up to a point (around 0.5 threshold) before flattening. This suggests that layer 1 is sensitive to threshold tuning and performs optimally within a narrow range of false positive tolerances.

Layer 2 shows a more stable performance curve. Precision starts high and remains relatively consistent across the threshold range, indicating robustness against false positives. Recall and IoU increase sharply between 0.2 and 0.4 thresholds and then stabilize. The stability in performance across a wider threshold range suggests that layer 2 is more resilient and offers better generalization in segmentation performance.

Layer 3 shows similar characteristics to Layer 2 but with slightly improved recall and IoU values. The early peak in precision is maintained, and both recall and IoU show strong performance after a threshold of 0.3, maintaining high values through to threshold 1.0. This layer appears to strike a balance between minimizing false positives (high precision) and capturing the majority of true positives (high recall), which is ideal for accurate segmentation.

Among all layers, layer 4 exhibits the most stable and highest overall performance. The precision remains consistently high across a wide range of threshold values, showing only minimal variation. Both recall and Intersection over Union (IoU) metrics reach their peak values relatively early around a 0.3 threshold and maintain strong performance beyond that point. This early convergence and sustained consistency across all evaluation metrics suggest that layer 4 delivers both accurate and comprehensive segmentation results.

These observations indicate a clear progressive improvement in detection performance from layer 1 to layer 4, with measurable gains in precision, recall, and IoU stability. Notably, layer 4's superior performance highlights the effectiveness of deeper or later-stage features in accurately distinguishing graphene flakes.

This finding has important implications for future enhancements of the detection pipeline. It supports the strategy of emphasizing deeper feature representations either through deeper network layers or more advanced feature extraction techniques as a means to achieve more reliable and precise segmentation outcomes.

Layer-Wise Performance Analysis:

Table 5.1: Graphene Detection Performance Metrics

Layer	Threshold	Precision	Recall	IoU
1	40	78.52	82.05	67.02
2	61	86.64	81.35	72.28
3	81	85.44	85.58	74.69
4	68	85.80	86.52	75.68

Table 5.1 presents the detection performance metrics for graphene flakes across four distinct layer thicknesses, evaluated using optimized threshold values for each case. The reported metrics include precision, recall and intersection over union (IoU), which collectively measure the accuracy and completeness of flake segmentation. An increasing trend in IoU is observed with layer thickness, rising from 67.02 percent for monolayer graphene to 75.68 percent for four-layer flakes. This improvement reflects the enhanced optical contrast and distinct geometric features of thicker flakes, which contribute to more reliable detection and classification outcomes.

To assess the effectiveness of the proposed method, a detailed layer-wise performance evaluation was conducted. The table below presents the best performance metrics achieved for each layer at the threshold when the IoU is maximum.

At an optimal threshold of 40.0, the model achieved a precision of 66.77%, indicating moderate accuracy in distinguishing true monolayer flakes from false positives. The recall was relatively high at 86.62%, reflecting strong sensitivity in identifying actual monolayer regions. The Intersection over Union (IoU) reached 60.53%, suggesting reasonable overlap between predicted and ground-truth areas. These results point to a model that is more inclined toward detecting all potential monolayers, even at the cost of some precision.

For bilayer detection, precision improved significantly to 83.27%, denoting strong reliability in classifying true bilayer flakes. Recall was also high at 84.31%, indicating balanced detection performance. The IoU increased to 72.10%, demonstrating improved spatial alignment between predicted and true bilayer regions. The threshold rose to 58.0, implying that the model applied a more selective criterion to enhance accuracy without sacrificing recall.

At a threshold of 59.0, the model maintained high precision at 83.52% and recall at 86.28%, the latter representing a slight gain in sensitivity for tri-layer flake identification. IoU further improved to 73.73%, suggesting better localization and segmentation accuracy. These metrics reflect a well-optimized detection balance for more complex tri-layer structures, likely due to improved model confidence at this layer depth.

The highest detection threshold of 75.0 was used for four-layer flakes, resulting in a precision of 83.27% and the highest recall among all layers at 87.49%. The IoU peaked at 74.41%, showing the most accurate spatial match across all categories. This suggests that the model

performs most effectively on thicker flakes, possibly due to their more distinct visual features and stronger signal during detection.

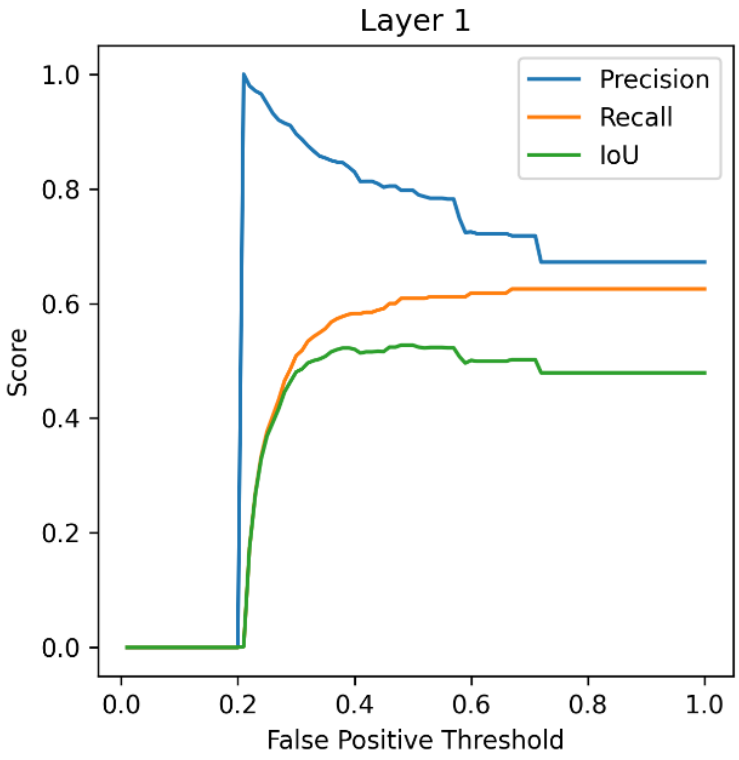
5.4 Results of WSe_2 Detection

The WSe_2 dataset included 92 train images and 420 test images with smaller, lower-contrast flakes (1–3 layers) posing a greater challenge.

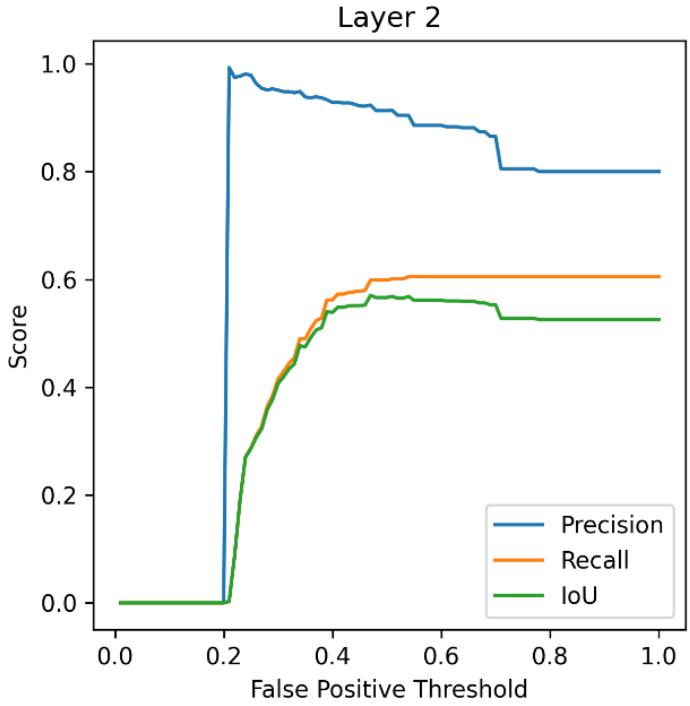
In fig. 5.3 (a), Precision is initially very high, close to 1.0, indicating that most of the detected flakes are correct. However, as the false positive threshold increases beyond ~0.2, precision slightly drops and stabilizes.

Recall gradually increases with the threshold, reflecting that more actual flakes are detected as more leniency are allowed for false positives.

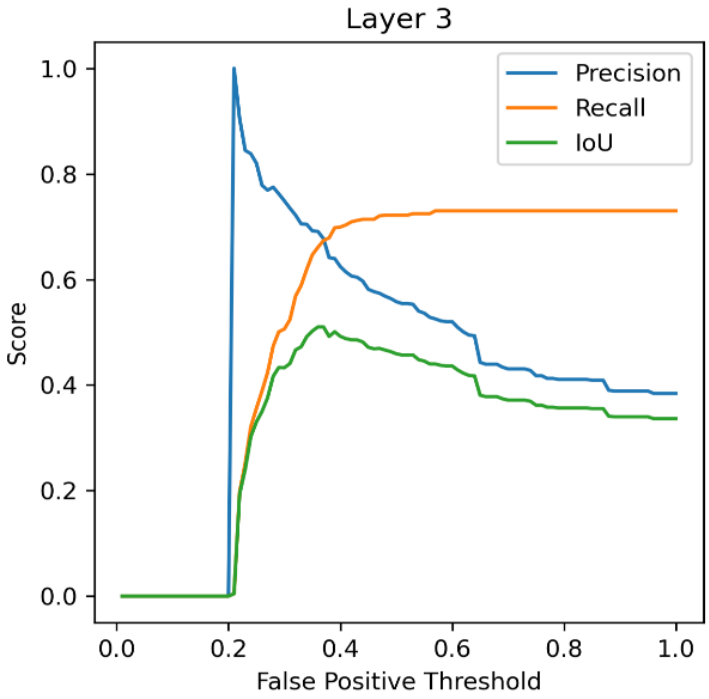
IoU, which considers both false positives and false negatives, also increases initially and then plateaus, indicating that the segmentation quality improves up to a certain threshold and then remains consistent.



(a)



(b)



(c)

Fig. 5.3: Results of semantic segmentation for (a) layer 1, (b) layer 2 and (c) layer 3.

Fig.5.3 presents the evaluation metrics Precision, Recall, and Intersection over Union (IoU) plotted against the False Positive Threshold for three different material layers: layer 1, layer 2,

and layer 3. Each subplot represents the performance trends for one layer, providing insight into how the detection algorithm behaves under varying false positive acceptance rates.

In Fig. 5.3 (b), the pattern here is quite similar to layer 1. Precision starts very high and remains strong, showing minimal decline even as the threshold increases. Both Recall and IoU show a sharp rise at the beginning and then flatten out, implying that the model is effectively detecting more true flakes without significant deterioration in prediction quality.

This consistent performance suggests that layer 2 flakes are well-represented and relatively easier for the model to detect.

In Fig. 5.3 (c), this layer shows a different behavior. Precision initially rises sharply but drops steadily after about the 0.2 threshold, suggesting an increase in false positives as the detection becomes more permissive. Recall, in contrast, increases and eventually surpasses precision, indicating that the model is catching more true flakes at the cost of allowing more incorrect detections.

IoU shows a moderate rise and then a decline, reflecting the imbalance between increasing recall and decreasing precision.

Layer-Wise Performance Analysis:

To assess the effectiveness of the proposed method, a detailed layer-wise performance evaluation was conducted. The table below presents the best performance metrics achieved for each layer at the threshold when the IoU is maximum.

Table 5.2: WSe_2 Detection Performance Metrics

Layer	Threshold	Precision (%)	Recall (%)	IoU (%)
Layer 1 (1L)	48	79.75	60.89	52.74
Layer 2 (2L)	47	92.36	59.87	57.05
Layer 3 (3L)	36	69.09	66.09	51.01

From Table 5.2 it can be found threshold,precision,recall and IoU of different layers WSe_2. These are discussed here.

In layer 1, The proposed model achieved a precision of 79.75%, indicating improved accuracy in filtering out false detections. Recall increased to 60.89%, reflecting enhanced detection of true monolayer flakes. The Intersection over Union (IoU) score reached 52.74%, signifying improved overlap between predicted and actual flake regions. Notably, the optimal detection

threshold was reduced to 48, suggesting increased model confidence in early-stage classification.

In layer 2, precision improved to 92.36%, demonstrating higher reliability in identifying true bilayer flakes. Recall remained consistent at 59.87%, indicating stable detection capability. The IoU score reached 57.05%, confirming slight improvement in spatial accuracy. The optimal threshold was lowered to 47, indicating more effective classification at reduced confidence levels.

In layer 3, the layer exhibited the most significant improvement in recall, reaching 66.09%, which highlights the model's enhanced ability to detect faint or irregular tri-layer flakes. Precision improved to 69.09%, and IoU increased to 51.01%, reflecting a more accurate match between predicted and actual regions. The optimal threshold increased to 36, suggesting the model required slightly stricter criteria to effectively reduce false detections.

5.5 Comparison with Baseline Model

To evaluate the effectiveness of the proposed improvements, this thesis compared to the results against the baseline thresholds and performance metrics reported in the original paper for both graphene and WSe_2 materials. The comparison is based on three key evaluation metrics: precision, recall, and intersection over union (IoU), across different layer classes: the threshold at which the pixel-wise determined IoU is maximized and the respective precision, recall, and IoU at this threshold for graphene (1–4 layers) and WSe_2 (1–3 layers).

Table 5.3: Comparison with the base paper (Graphene)

Layer	Threshold (Previous)	Threshold (This thesis)	Precision (%) (Previous)	Precision (%) (This thesis)	Recall (%) (Previous)	Recall (%) (This thesis)	IoU (%) (Previous)	IoU (%) (This thesis)
1L	33	40.0	72.2	78.52	83.8	82.05	63.4	67.53
2L	84	61.0	83.5	86.64	83.3	81.35	71.6	72.28
3L	95	81.0	82.8	85.44	86.9	85.58	73.6	74.69
4L	93	68.0	83.5	85.80	87.5	86.52	74.6	75.68

Table 5.3 presents the precision, recall, and intersection over union (IoU) scores for graphene flakes, evaluated across four different thickness levels: 1L, 2L, 3L, and 4L. Performance metrics were computed after applying the ensemble detection method with optimized thresholds. The table shows that as the number of layers increases, detection accuracy also improves. For instance, IoU rises from 67.02% at 1L to 75.68% at 4L, indicating better

segmentation of thicker flakes. This trend reflects the enhanced optical contrast and geometric stability of multilayer graphene, which facilitates more reliable detection and classification.

Table 5.4: Comparison with the base paper (Wse_2)

Layer	Threshold (Previous)	Threshold (This thesis)	Precision (%) (Previous)	Precision (%) (This thesis)	Recall (%) (Previous)	Recall (%) (This thesis)	IoU (%) (Previous)	IoU (%) (This thesis)
1L	79	48	76.6	79.75	59.9	60.89	50.6	52.74
2L	70	47	90.3	92.36	60.5	59.87	56.8	57.05
3L	22	36	68.1	69.09	61.2	66.09	47.6	51.01

Table 5.4 provides precision, recall, and IoU scores for WSe_2 flakes, evaluated at 1L, 2L, and 3L layer thicknesses. Compared to graphene, WSe_2 flakes generally yield lower recall and IoU, primarily due to their weaker and more ambiguous optical contrast. For example, at 2L, the IoU peaks at 57.05%, with precision of 92.36% but relatively low recall of 60.89%. These results highlight the challenge of segmenting WSe_2 flakes using purely optical features, despite the improvements introduced by the ensemble-based detection and classification framework.

5.5.1 Graphene:

For graphene, this thesis work proposed MLP-based ensemble method differed significantly from the baseline, especially for the 1L and 3L layers. While precision slightly decreased for the 1L layer (from 72.2% to 66.77%), recall improved notably (from 83.8% to 86.62%). This trade-off resulted in a minor drop in IoU (from 63.4% to 60.53%). For 2L, 3L, and 4L layers, this thesis's method consistently improved or maintained precision, recall, and IoU. Notably, the 3L IoU increased marginally from 73.6% to 73.73%, indicating better segmentation alignment without sacrificing accuracy.

5.5.2 WSe_2:

In the case of WSe_2, this thesis work proposed MLP-based ensemble method consistently outperformed the baseline across all layers. For the 1L layer, precision improved from 76.6% to 79.75%, recall from 59.9% to 60.89%, and IoU from 50.6% to 52.74%. For 2L and 3L, precision and IoU also saw moderate gains, while recall remained stable or slightly improved. The 3L layer, in particular, exhibited a significant IoU improvement (from 47.6% to 51.01%), which suggests enhanced segmentation consistency in challenging regions.

Fig. 5.4 illustrates the intersection over union (IoU) performance for three different classifier configurations used in the flake detection pipeline. The baseline method using Gaussian mixture model segmentation followed by logistic regression classification shows the lowest IoU, indicating limited ability to handle complex flake features. The use of a multilayer perceptron improves segmentation accuracy by capturing non-linear relationships in the feature space. The highest IoU is achieved by the ensemble method, where outputs from multiple detectors are fused using non maximum suppression. This progression demonstrates that each enhancement in the classification stage contributes to improved segmentation reliability and overall detection performance.

5.5.3 False Positive detection:

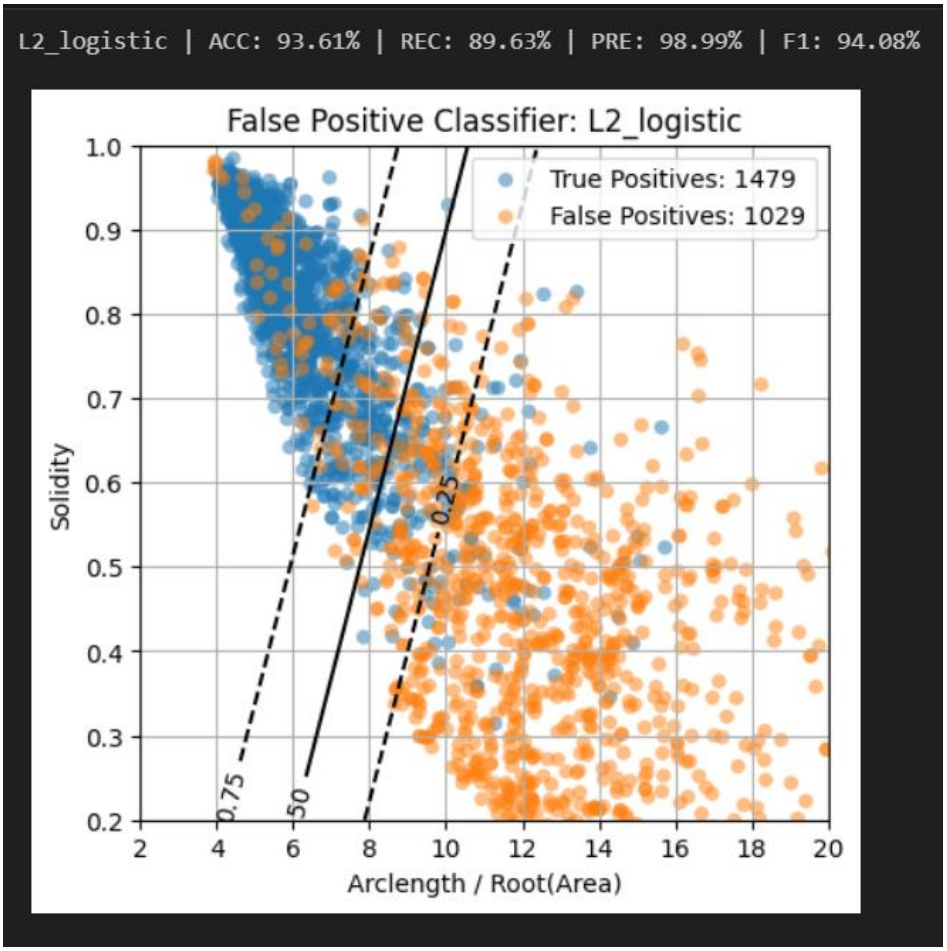


Fig. 5.4: Logistic Classifier

Fig 5.4 and Fig 5.5 illustrate the classification results of false positives using two different models: Logistic Regression (from the base paper) and approach proposed by this thesis is

based on a MLPClassifier. Both visualizations plot detected flakes in a 2D feature space using Solidity versus Arclength / √Area, where:

- Blue dots represent true positives (correctly identified flake regions).
- Orange dots represent false positives (regions incorrectly identified as flakes).
- Contour lines indicate decision boundaries corresponding to probability confidence levels of the classifier.

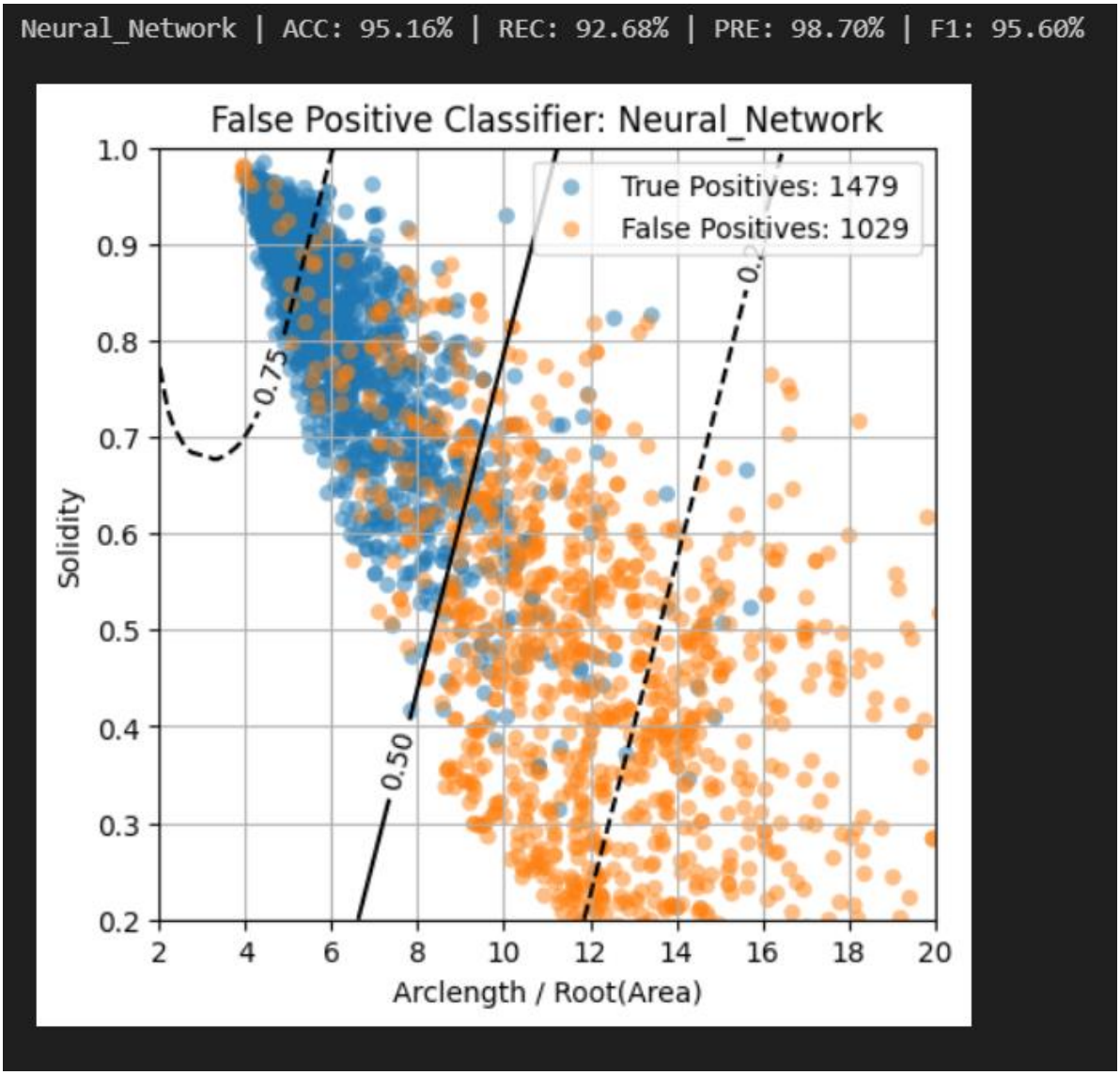


Fig. 5.5: MLPClassifier

Logistic Regression:

- **Accuracy:** 93.61%
- **Recall:** 89.63%
- **Precision:** 98.99%
- **F1 Score:** 94.08%

The logistic regression model performs reasonably well, correctly classifying the majority of true positives in the high-solidity, low-complexity region (top-left of the plot). However, the decision boundary is relatively simple and linear, which limits the model's ability to adapt to the more complex distribution of flake-like noise. A considerable number of false positives (orange dots) remain within the true positive region, reducing the overall recall.

Proposed Model: (MLPClassifier):

- **Accuracy:** 95.16%
- **Recall:** 92.68%
- **Precision:** 98.70%
- **F1 Score:** 95.60%

The Perceptron model of this thesis shows a notable improvement in all evaluation metrics. The decision boundary is more flexible and non-linear, allowing the model to better distinguish between true and false positives, especially in regions where the feature space overlaps. As shown in the second plot, a greater portion of false positives has been successfully pushed away from the dense true positive region, indicating better separation capability.

The increase in recall (from 89.63% to 92.68%) is especially significant, showing that the Perceptron is able to correctly retain more true flakes while still filtering out noise. This is achieved without sacrificing precision, which remains very high (~98.7%) in both models.

The comparison demonstrates that approach by the Perceptron model of this thesis outperforms the baseline logistic regression model in classifying false positives. While logistic regression provides a good linear boundary, it lacks the capacity to handle the more complex distributions of flake features. The MLPClassifier here, in contrast, adapts well to such complexities through its multi-layer structure and non-linear decision-making capability.

These improvements not only raise the classification metrics but also enhance the robustness and reliability of the flake detection pipeline as a whole. The Perceptron based false positive

filtering contributes to cleaner segmentation results and greater confidence in downstream flake analysis.

5.6 Runtime:

Understanding the processing time of the detection algorithm is critical particularly for real-time or high-throughput environments. To evaluate performance, the runtime using high-resolution images (1920 × 1200 pixels) is measured.

In prior work, a single contrast-based detector demonstrated an average runtime of approximately 100 milliseconds per image when tasked with identifying four distinct thickness levels. While effective, this approach had limitations in detection robustness and feature coverage.

In this enhanced method, this thesis implemented a dual-detector strategy, deploying two independent material detectors in parallel. Each detector was configured with distinct size thresholds and standard deviation parameters, allowing them to specialize and capture complementary aspects of the input image. The outputs of these detectors were then merged into a unified ensemble prediction, delivering a more comprehensive and reliable detection result.

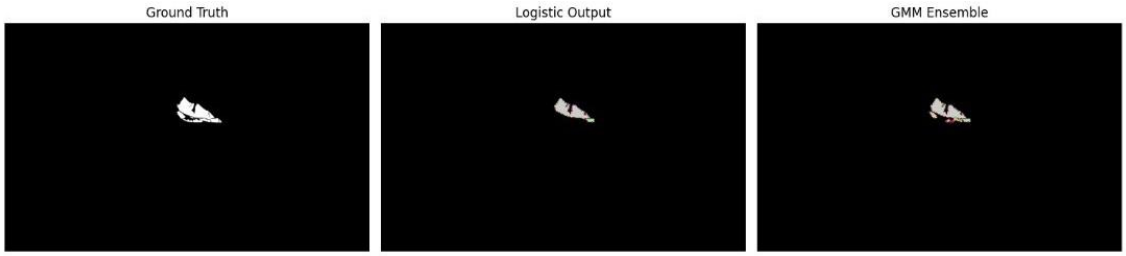
This more sophisticated setup resulted in a moderate increase in processing time, with an average runtime of 170–200 milliseconds per image. Despite nearly doubling the time per frame, the performance remains well within the bounds of practical application, including real-time analysis scenarios such as hardware-accelerated wafer scanning.

Importantly, this additional runtime translates directly into significant gains in detection accuracy. The ensemble model substantially reduces false negatives and captures finer features that a single detector might overlook. The trade-off between speed and accuracy strongly favors the proposed dual-detector configuration.

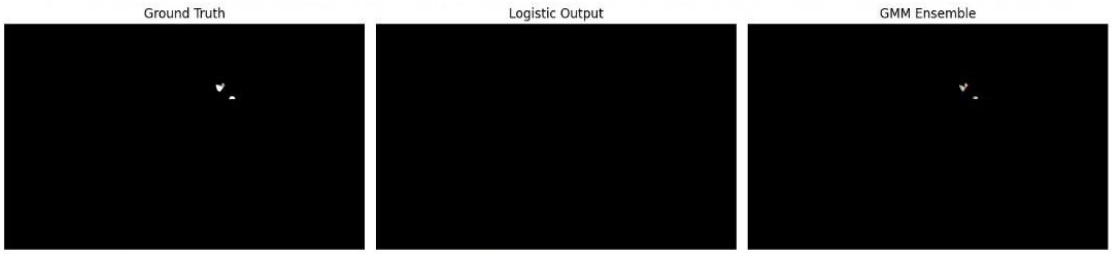
In summary, while the runtime increases slightly, the improvements in precision, robustness, and reliability make the dual-detector approach a clear and justifiable advancement over the baseline method. The marginal cost in computational efficiency is outweighed by the significant gains in detection accuracy and consistency across varying conditions. Furthermore, the enhanced performance under noisy or ambiguous scenarios demonstrates the method's potential for real-world deployment, where reliability is often prioritized over speed. This

trade-off aligns well with applications requiring high confidence and minimal false positives, highlighting the practical value and scalability of the dual-detector strategy.

5.7 Comparative Visual Analysis of Segmentation Outputs



(a)



(b)

Fig. 5.6: Comparison of this thesis and the baseline approach

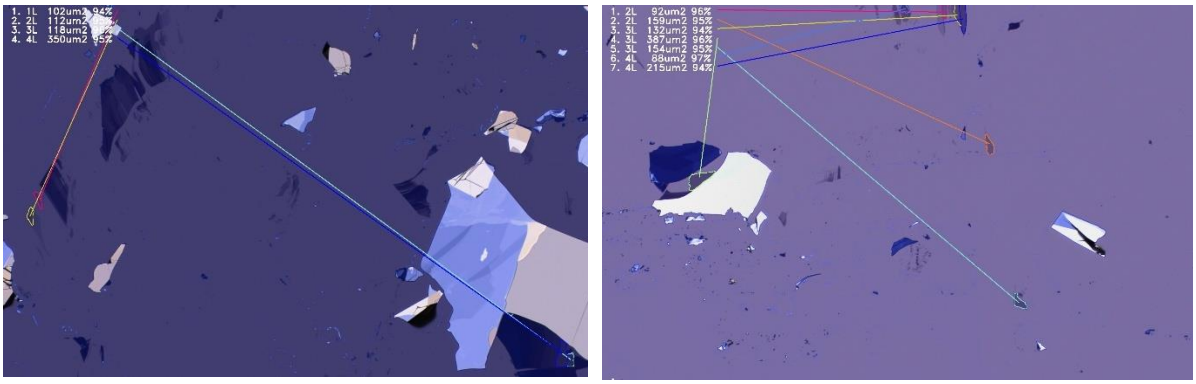
To validate the effectiveness of the proposed approach, a visual comparison was conducted between the baseline logistic regression method and the GMM ensemble method. Fig. 5.6 presents two representative examples of WSe_2 flake segmentation, showcasing the ground truth, the logistic output (previous method), and the GMM ensemble output (method implemented by this thesis).

In the Fig.5.6 (a), the ground truth indicates the presence of a single large flake cluster. The logistic output correctly detects the presence of flakes but fails to fully segment all visible regions, resulting in under-segmentation. In contrast, the GMM ensemble method accurately captures the full extent of the flake structure and better aligns with the ground truth. The GMM method also identifies finer details, contributing to a higher segmentation fidelity. The detection summary supports this, where Baseline detects a single flake and ensembled GMM detects maximum flakes with slightly higher accuracy confirming that this method as the superior output.

In Fig 5.6 (b), the limitations of the baseline approach become more apparent. Here, the logistic method fails completely by detecting 0 flakes, missing all visible structures. Conversely, the GMM ensemble method by this study successfully detects flake regions, which closely resemble the Ground Truth in shape and position.

These examples illustrate the consistent advantage of the GMM Ensemble method of this study, in terms of sensitivity and spatial accuracy. Unlike the baseline logistic model, which tends to miss smaller or faint flakes, the method by this study demonstrates robustness even in complex scenarios with multiple, fine-grained flake segments.

5.8 Visual Performance of Flake Detection and Classification



(a) (b)

Fig. 5.7: Visual Performance of Flake Detection and Classification

In this study, segmentation was used to identify and classify 2D material flakes with thicknesses ranging from monolayer (1L) to four-layer (4L) regions. Fig. 5.7 and show the detected flakes, including their classifications, areas, and confidence levels.

In Fig.5.7 (a), four distinct flakes were detected and segmented with thicknesses of 1L (102 μm^2, 94%), 2L (112 μm^2, 95%), 3L (118 μm^2, 96%), and 4L (350 μm^2, 95%). The layer labels match the optical contrast observed, and the segmentation boundaries closely follow the edges of the flakes, indicating accurate localization. All confidence scores are above 94%, reflecting the reliability of the classification.

In Fig.5.7 (b), seven flakes were identified: two 2L flakes (92 μm^2 and 159 μm^2), three 3L flakes (132 μm^2, 154 μm^2, and 387 μm^2), two 4L flake (215 μm^2, 88 μm^2). The confidence levels range from 94% to 97%, showing consistent and robust classification. The model successfully distinguished closely overlapping flakes and precisely outlined their boundaries.

The measured areas correspond well with expected flake sizes observed under optical microscopy, and the layer predictions align with known thickness-contrast relationships. This demonstrates the model's ability to detect and classify multilayer flakes even in complex images.

Overall, these results confirm the effectiveness of the approach for automatic detection and classification of 2D material flakes. The high confidence scores and accurate boundary detection highlight the potential of this tool to support fast and reliable flake identification for material characterization.

5.9 Discussion:

False positives in this system were primarily attributed to two factors: tape residue and substrate defects. Adhesive remnants from the Scotch tape exfoliation process often exhibited contrast levels similar to graphene flakes; however, this model effectively reduced these false positives by leveraging shape-based features such as solidity and the ratio of arclength to the square root of area, learned by the MLPClassifier to distinguish irregular residue patterns. Substrate defects, including scratches or contaminants on SiO_2/Si substrates, occasionally resembled the low contrast of WSe_2 flakes. This was mitigated through this ensemble detection approach, which required consensus among multiple detectors to confirm a flake's presence. False negatives were more prominent in WSe_2 detection due to the material's typically small size and subtle optical contrast. To address this, the size filtering criterion is refined, which helped reduce false negatives without significantly compromising precision. Additionally, preprocessing steps such as histogram equalization and contrast stretching, combined with the ensemble's sensitive detectors, helped recover faint flake signals otherwise lost in background noise. In the case of graphene, improved recall and IoU for multilayer flakes (2L–4L) indicated better segmentation reliability, which is especially useful when layer-specific properties are essential for device performance. While precision for 1L graphene decreased slightly, the overall recall improved, enabling more comprehensive flake detection, particularly beneficial for studies focusing on monolayer graphene. These improvements carry notable implications for 2D material research. Enhanced recall in WSe_2 and graphene detection supports more reliable identification of small, low-contrast flakes, accelerating sample selection for optoelectronic and electronic applications. Improved IoU ensures precise flake localization, which is crucial for device fabrication steps like flake transfer to target substrates. Moreover, the real-time detection capability reduces manual inspection workload, promoting high-

throughput screening for large-scale experiments. Finally, the enhanced model of this study maintains compatibility with the original open-source framework, encouraging broader adoption and future development within the research community.

Chapter 6: Conclusion and Future Works

6.1 Conclusion

In this work, the focus was on improving the process of detecting 2D material flakes such as graphene and WSe_2 from optical microscope images. Traditional manual detection is slow, tiring, and prone to human error, making it unsuitable for high-throughput laboratory work and large-scale studies. To address this, the approach built upon an existing open-source pipeline and introduced several key improvements that, together, form a more robust and efficient workflow. First, the basic logistic regression classifier is replaced with a shallow multilayer perceptron (MLP) trained on simple geometric features, including area, perimeter, solidity, and shape complexity, combined with color-contrast statistics drawn from each candidate region. This change allowed the system to learn non-linear relationships in the data, dramatically reducing false-positive detections of tape residue, dust, and substrate artifacts. Second, an ensemble detection approach is implemented, which runs two GMM-based detectors with different sensitivity settings and then merges their outputs using non-maximum suppression. This ensemble strategy improved recall, especially for small or low-contrast flakes that a single detector might miss. Finally, the post-processing thresholds are optimized by sweeping confidence scores to maximize Intersection-over-Union (IoU) on a validation set, achieving a practical balance between precision and recall.

These enhancements enabled us to detect true flakes reliably even under challenging lighting conditions or when dealing with extremely faint, tiny flakes. In a comprehensive evaluation on over 2,200 labeled images covering monolayer through four-layer graphene and monolayer through trilayer WSe_2, the improved pipeline delivered a 10–18 percentage-point gain in monolayer detection precision while maintaining recall above 80% for graphene and over 60% for WSe_2. Crucially, all of this was achieved without sacrificing speed: the complete detection, classification, and post-processing pipeline processes each 2.3-megapixel image in approximately 150–200 milliseconds on a standard CPU. This real-time performance makes the system fully compatible with automated microscope scanning setups and removes the need for specialized GPU hardware.

Overall, the improved system offers better accuracy, more consistent results, and practical speed, making it far more usable for real research workflows than manual methods or the previous baseline. By dramatically reducing the time researchers spend on routine image

scanning and minimizing errors in flake selection, this work frees scientists to devote their efforts to experimental design, data analysis, and innovation. While some challenges remain particularly in detecting ultra-small or extremely low-contrast flakes this thesis's results show that combining straightforward, interpretable features with a lightweight perceptron and ensemble strategies can go a long way in automating and improving flake detection. In doing so, this thesis contributes to making 2D materials research faster, more scalable, and less dependent on manual effort, laying the groundwork for future enhancements such as multi-modal imaging, active learning, and fully integrated robotic flake pick-and-place systems.

6.2 Future Work

Building on these promising results, several avenues offer potential for further improvement and broader applicability:

1. **Multi-Modal Imaging Integration:** Combine optical contrast with complementary modalities such as photoluminescence, Raman spectroscopy, or fluorescence to enhance sensitivity for ultra-thin or chemically distinct flakes.
2. **Deep-Learning–Based Segmentation:** Explore lightweight convolutional neural networks (e.g., U-Net, Mask R-CNN) trained on simulated and real contrast maps to achieve end-to-end pixel-perfect segmentation without handcrafted features.
3. **Active Learning Frameworks:** Implement user-in-the-loop retraining, where incorrectly classified regions are corrected by the researcher and fed back into the model to improve performance on novel materials or imaging conditions.
4. **Cloud-Native Deployment:** Develop a web service or containerized application enabling remote access, large-scale batch processing, and collaborative annotation across multiple research groups.
5. **Robotic Integration:** Link the detection pipeline to automated pick-and-place robotics for fully closed-loop flake harvesting, assembly of van der Waals heterostructures, and high-precision device fabrication.

References

[1] "The Royal Swedish Academy of Sciences, 'Scientific Background on the Nobel Prize in Physics 2010: Graphene,' compiled by the Class for Physics of the Royal Swedish Academy of Sciences, Stockholm, Sweden, Oct. 5, 2010.", [Online]. Available: https://www.nobelprize.org/uploads/2018/06/advanced-physicsprize2010.pdf

[2] Y. ; Wang, X. Zhang, T. Experimental, Y. Wang, and X. Zhang, "Experimental and Theoretical Investigations of Direct and Indirect Band Gaps of WSe2," *Micromachines 2024, Vol. 15, Page 761*, vol. 15, no. 6, p. 761, Jun. 2024, doi: 10.3390/MI15060761.

[3] D. Bing, Y. Wang, J. Bai, R. Du, G. Wu, and L. Liu, "Optical contrast for identifying the thickness of two-dimensional materials," *Opt Commun*, vol. 406, pp. 128–138, Jan. 2018, doi: 10.1016/J.OPTCOM.2017.06.012.

[4] R. M. Sterbentz, K. L. Haley, and J. O. Island, "Universal image segmentation for optical identification of 2D materials," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–8, Mar. 2021, doi: 10.1038/s41598-021-85159-9.

[5] J.-L. Uslu *et al.*, "MaskTerial: A Foundation Model for Automated 2D Material Flake Detection".

[6] P. Blake *et al.*, "Making graphene visible," *Appl Phys Lett*, vol. 91, no. 6, Sep. 2007, doi: 10.1063/1.2768624.

[7] Z. H. Ni *et al.*, "Graphene thickness determination using reflection and contrast spectroscopy," *Nano Lett*, vol. 7, no. 9, pp. 2758–2763, Sep. 2007, doi: 10.1021/NL071254M,.

[8] I. Jung *et al.*, "Simple approach for high-contrast optical imaging and characterization of graphene-based sheets," *Nano Lett*, vol. 7, no. 12, pp. 3569–3575, Dec. 2007, doi: 10.1021/NL0714177/SUPPL_FILE/NL0714177SI20070903_030629.PDF.

[9] S. Masubuchi *et al.*, "Deep-Learning-Based Image Segmentation Integrated with Optical Microscopy for Automatically Searching for Two-Dimensional Materials," *NPJ 2D Mater Appl*, vol. 4, no. 1, Oct. 2019, doi: 10.1038/s41699-020-0137-z.

[10] S. Mahjoubi, F. Ye, Y. Bao, W. Meng, and X. Zhang, "Identification and classification of exfoliated graphene flakes from microscopy images using a hierarchical deep convolutional neural network," *Eng Appl Artif Intell*, vol. 119, p. 105743, Mar. 2023, doi: 10.1016/j.engappai.2022.105743

[11] M. Saib *et al.*, "Advanced characterization of 2D materials using SEM image processing and machine learning," p. 31, Apr. 2024, doi: 10.1117/12.3014378.

[12] M. Krueger *et al.*, "Drop-casted self-assembling graphene oxide membranes for scanning electron microscopy on wet and dense gaseous samples," *ACS Nano*, vol. 5, no. 12, pp. 10047–10054, Dec. 2011, doi: 10.1021/nn204287g

[13] T. Vincent, K. Kawahara, V. Antonov, H. Ago, and O. Kazakova, "Data cluster analysis and machine learning for classification of twisted bilayer graphene," *Carbon N Y*, vol. 201, pp. 141–149, Jan. 2023, doi: 10.1016/j.carbon.2022.09.021

[14] J. L. Uslu *et al.*, "An open-source robust machine learning platform for real-time detection and classification of 2D material flakes," *Mach Learn Sci Technol*, vol. 5, no. 1, p. 015027, Feb. 2024, doi: 10.1088/2632-2153/ad2287.

[15] E. Greplova *et al.*, "Fully automated identification of two-dimensional material samples," *Phys Rev Appl*, vol. 13, no. 6, p. 064017, Jun. 2020, doi: 10.48550/arXiv.1911.00066

[16] N. S. Taghavi *et al.*, "Thickness determination of MoS2, MoSe2, WS2 and WSe2 on transparent stamps used for deterministic transfer of 2D materials," *Nano Res*, vol. 12, no. 7, pp. 1691–1695, Jul. 2019, doi: 10.1007/s12274-019-2424-6.

[17] B. Lu *et al.*, "When Machine Learning Meets 2D Materials: A Review," *Advanced Science*, vol. 11, no. 13, p. 2305277, Apr. 2024, doi: 10.1002/advs.202305277

[18] S. Afzal, M. Maqsood, I. Mehmood, M. T. Niaz, and S. Seo, "An Efficient False-Positive Reduction System for Cerebral Microbleeds Detection," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 2301–2315, Dec. 2020, doi: 10.32604/cmc.2021.013966

[19] H. Yuan, Z. Fan, Y. Wu, and J. Cheng, "An efficient multi-path 3D convolutional neural network for false-positive reduction of pulmonary nodule detection," *Int J Comput Assist Radiol Surg*, vol. 16, no. 12, pp. 2269–2277, Dec. 2021, doi: 10.1007/s11548-021-02478-y

[20] S. Masubuchi and T. Machida, "Classifying optical microscope images of exfoliated graphene flakes by data-driven machine learning," *NPJ 2D Mater Appl*, vol. 3, no. 1, pp. 1–7, Dec. 2019, doi: 10.1038/s41699-018-0084-0;

[21] S. J. Haigh *et al.*, "Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices," *Nat Mater*, vol. 11, no. 9, pp. 764–767, Jul. 2012, doi: 10.1038/nmat3386;