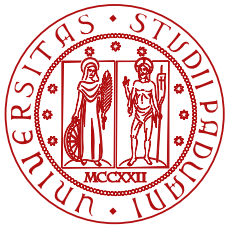


Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Thesis in Computer Engineering

Semantic Segmentation of Cavities in Electron Microscopy Images

Master Candidate

Gulustan Kaya

Student ID 2048716

Supervisor

Prof. Loris Nanni

University of Padova

Academic Year
2023/2024

To my lovely siblings,
my dear parents,
and to my helpful friends Sevval, Merve, Alp, Andrea

Abstract

The main goal of this thesis is to improve the automation of quantifying materials swelling by implementing and enhancing sophisticated deep learning models in MATLAB. The paper titled "Materials swelling revealed through automated semantic segmentation of cavities in electron microscopy images" [1] showcased the efficacy of the Mask R-CNN (Mask Regional Convolutional Neural Network) model in precisely detecting and quantifying cavities in irradiated alloys. However, this study aims to enhance the existing knowledge by including the SOLOv2 (Segmenting Objects by Locations version 2) segmentation model, creating bespoke loss functions that include swelling indicators, and utilizing standard segmentation models such as DeepLabV3+ for separating the background. Furthermore, a two-round fine-tuning method is implemented to enhance the model generalization. The improved model ensemble utilizes the advantages of Mask R-CNN and SOLOv2, along with customized loss functions and fine-tuning procedures, to deliver more accurate and dependable measurements of swelling. This research not only enhances the subject of nuclear materials science but also showcases the wider applicability of deep learning techniques in scientific inquiries, highlighting the usefulness and ease of access of MATLAB for such applications. The results offer significant improvements in the safety and dependability of nuclear reactor materials, which will contribute to more resilient designs and safer operational practices.

Sommario

L'obiettivo principale di questa tesi è migliorare l'automazione della quantificazione di possibili rigonfiamenti di materiali, implementando e migliorando sofisticati modelli di deep learning in MATLAB. L'articolo intitolato "Materials swelling revealed through automated semantic segmentation of cavities in electron microscopy images" [1] ha mostrato l'efficacia del modello Mask R-CNN nel rilevare e quantificare con precisione le cavità nelle in metalli e leghe irradiati. Tuttavia, questa tesi mira a migliorare le conoscenze esistenti includendo il modello di segmentazione SOLOv2, creando loss function su misura che includono indicatori di rigonfiamento e utilizzando modelli di segmentazione standard come DeepLabV3+ per separare lo sfondo. Inoltre, viene implementato un metodo di perfezionamento in due fasi per migliorare la generalizzazione del modello. L'ensemble di modelli sfrutta i vantaggi di Mask R-CNN e SOLOv2, insieme ad una loss function personalizzata con l'aggiunta di fine-tuning, per fornire misurazioni delle cavità più accurate e affidabili. Questa ricerca non solo approfondisce il tema della scienza dei materiali nucleari, ma mostra anche la più ampia applicabilità delle tecniche di deep learning nelle indagini scientifiche, evidenziando l'utilità e la facilità di accesso di MATLAB per tali applicazioni. I risultati offrono miglioramenti significativi nella sicurezza e nell'affidabilità dei materiali dei reattori nucleari, materali che contribuiranno a design più resilienti e un incremento nella sicurezza sul lavoro.

Contents

List of Figures

List of Tables

List of Acronyms

CNN Convolutional Neural Network

R-CNN Regional Convolutional Neural Network

SOTA state-of-the-art

DSCNN Dynamic Segmentation Convolutional Neural Network

S/TEM Scanning/Transmission Electron Microscopy

COCO Common Objects in Context

ResNet Residual Network

RoI Region of Interest

JSON JavaScript Object Notation

PNG Portable Network Graphic

GPU Graphics Processing Unit

CNL Canadian Nuclear Laboratory

NOME Nuclear Oriented Materials & Examination

RPN Region Proposal Network

SGDM Stochastic Gradient Descent with Momentum

FPN Feature Pyramid Network

IoU Intersection over Union

MAE Mean Absolute Error

1

Introduction

Nuclear materials science is crucial to building and functioning of effective and reliable nuclear reactors. This discipline revolves around the study of alloy swelling, which occurs when metal alloys are exposed to radiation. This swelling may threaten the structural integrity of reactor components, potentially resulting in catastrophic failure. Usually the quantification of radiation-induced swelling in alloys was done manually by domain specialists who analyzed electron microscopy images to detect and measure flaws like cavities and voids. However, this manual approach is time-consuming, not objective, and unscalable, especially given the growing number of high-resolution microscopy data produced by modern devices.

Recent improvements in computer vision, especially deep learning techniques, offer encouraging solutions to these problems. Deep learning models, such as Convolutional Neural Networks (CNNs), have altered image analysis by allowing automated and remarkably accurate object detection and segmentation. Among these, the Mask R-CNN is recognized as an effective tool for instance segmentation, offering pixel-level annotations of objects within images. Using these leading models, it is feasible to automate the detection and quantification of nanoscale cavities in irradiated alloys, by that increasing the efficiency and accuracy of swelling assessments.

This thesis focuses on the preliminary work reported in our base paper, "Materials swelling revealed through automated semantic segmentation of cavities in electron microscopy images" [1]. The study established the ability of utilizing Mask R-CNN to detect and measure cavities in irradiation alloys, with promis-

ing precision, recall, and F1 score results. However, there is still plenty of space for improvement, particularly in the integration of further segmentation models and the development of custom loss functions tuned to specific material features such as swelling.

The fundamental goal for this study is to advance the state-of-the-art (SOTA) in automated swelling quantification by examining and combining various deep learning algorithms in MATLAB. MATLAB is an effective platform in engineering and scientific research, noted for its extensive toolboxes and user-friendly interface, making it a solid option for implementing and evaluating leading algorithms in this field. Specifically, this work seeks to:

- Improve the Mask R-CNN model by combining the results of SOLOv2, another innovative segmentation model, in the MATLAB environment.
- Create a custom loss function that includes the swelling indicators described in the base paper.
- Examine the feasibility of modifying Mask R-CNN and SOLOv2's loss functions to increase their performance on this specific task in MATLAB.
- Examine the potential of using standard segmentation models, such as DeepLabV3+, for background segmentation before moving on to focused instance segmentation in the foreground.
- Delve into a two-round fine-tuning strategy with closely similar datasets to improve the model's generalizability and performance.

By addressing these goals, this thesis seeks to contribute to the field of nuclear material studies by developing more precise and effective tools for assessing radiation-induced swelling. These developments have the potential to remarkably impact nuclear reactor material design and operational safety, leading to more reliable performance under irradiation. Furthermore, the methodology and conclusions of this study can be applied to other fields that require exact image-based defect assessment, demonstrating deep learning's wide utility in scientific research. The use of MATLAB as a key platform emphasizes the practicality and accessibility of this research, making it greatly relevant to the engineering community.

2

Research Background

2.1 Overview of Alloy Swelling in Nuclear Reactors

Alloy swelling in nuclear reactors is a crucial concern that impacts the durability and safety of reactor components. Metal alloys used in the cores and surrounding components of nuclear reactors undergo irradiation, which causes damage and leads to the creation of prolonged defects including dislocation loops, precipitates, and cavities. These cavities, known as voids when they don't contain gas and bubbles when they do, have a negative impact on mechanical characteristics by causing hardness, brittleness, and swelling.

The expansion of cavities caused by neutron irradiation, which is typically stabilized by helium produced through nuclear transmutation, results in substantial swelling. The enlargement of this area might lead to the deterioration and breakdown of the substance [1]. Hence, it is imperative to understand the interaction between alloy composition, microstructure, and reactor conditions, such as temperature and irradiation dose, in order to guarantee the safety and dependability of reactor operations. The study conducted by Singh et al. [2] emphasizes that neutron irradiation has a notable impact on the creation and enlargement of voids, leading to an increase in volume of materials exposed to intense irradiation conditions.

Neutron irradiation causes the swelling of alloys by a multifaceted process that encompasses various mechanisms. The expansion of voids due to bias is

worsened by the existence of helium, which combines to form bubbles inside the microstructure of the material. Swelling and grain boundary embrittlement can occur, which can compromise the integrity of reactor materials. Neutron irradiation enhances the strength of these materials, but it can also significantly decrease their ductility, resulting in radiation hardening and inconsistent plastic deformation. According to Ghoniem et al. [3] when structural materials are exposed to neutron irradiation, it leads to the creation of several types of atomic-scale flaws. These flaws often enhance the strength of the materials but can significantly decrease their ability to stretch or deform.

Comprehending these interactions and their outcomes is crucial for preserving the structural stability of reactors and guaranteeing the secure and effective functioning of nuclear power plants.

2.2 Electron Microscopy and Swelling Quantification

Measuring the extent of swelling in nuclear reactor materials using electron microscopy, particularly Scanning/Transmission Electron Microscopy (S/TEM), is an essential component of comprehending the deterioration of materials due to irradiation. TEM imaging conditions and sample properties can produce fluctuations that typically lead to the perception of swelling, rather than exact swelling. Precise measurement of swelling is crucial to guarantee the structural stability and security of nuclear reactors. Quantification problems arise due to phase shifts generated by variations in the mean inner potential between the cavity and the surrounding crystal. These shifts greatly affect the observed size of cavities when imaged under different conditions. The base paper highlights that multi-slice simulations have shown substantial alterations in the perceived size of cavities when employing Fresnel contrast imaging at varying underfocus levels [1].

To see the distinct white centroid and dark fringe contrast of holes while minimizing the displacement of the black fringe, it is necessary to maintain a small underfocus (<−1 μm). Precise measurement of cavities, especially those with dimensions smaller than 10 nm, relies heavily on this technique. Nevertheless, the process is complicated by mistakes arising from imaging parameters such as resolution, human measurement accuracy, sample tilt, and background contrast

change.

Conventional techniques require the manual counting and measurement of individual cavities using software such as ImageJ. This approach is limited by the time-consuming nature of preparing TEM samples and identifying the cavities. Progress in sample preparation techniques, such as high-throughput focused ion beam (FIB) procedures and flash polishing, have reduced some constraints. Recent TEM devices have enhanced data collecting rates, resulting in higher resolution images and larger datasets. However, the process of manually labeling cavities still poses a significant obstacle. The necessity for automated techniques capable of rapidly processing extensive TEM datasets is apparent.

Utilizing machine learning techniques is crucial for overcoming these issues, with automation playing a key role. CNNs have become highly effective tools for automated microstructure analysis, utilizing machine-learning approaches. These networks can accurately and reliably identify certain microstructural features from an image in a cost-efficient and consistent manner.

Anderson et al. [4] utilized a region-based convolutional neural network to examine helium bubbles in irradiated X750 Ni alloys. They successfully identified the positions of the bubbles in underfocused TEM images. This approach emphasizes the capacity for efficient data analytics in the field of materials science.

Moreover, comprehending the stability of voids caused by irradiation is crucial for effectively controlling material swelling. A study conducted by Chen et al. [5] offers valuable insights into the mechanisms that govern the stability transition of voids.

Future initiatives entail the enhanced incorporation of automated analysis tools to facilitate the efficient processing and analysis of big datasets. These developments will increase our comprehension of material behavior when exposed to radiation and assist in the development and assessment of new materials and production methods to boost performance and safety in nuclear reactors. Ultimately, the process quantifying swelling in nuclear reactor materials by electron microscopy is essential for comprehending the deterioration of materials caused by irradiation. Automation and machine learning greatly improve this procedure, offering more precise, dependable, and thorough evaluations of material swelling. These technological improvements are crucial for preserving the structural stability of reactors and guaranteeing the secure and effective functioning of nuclear power facilities.

2.3 Deep Learning in Material Science

In the last decade, deep learning techniques have brought about a significant transformation in various scientific domains, such as material science. Deep learning is a branch of machine learning that uses computational models with numerous layers to learn data representations at different levels of abstraction. CNNs have had a significant impact, demonstrating outstanding performance in tasks such as image classification, object detection, and semantic segmentation. These technological breakthroughs have had a profound impact on the field of material science, particularly in the examination and description of the microscopic structures of materials.

Multiple studies have showcased the efficacy of deep learning in the field of material science. The study conducted by Stepashkin et al. [6] focused on thermoplastic unidirectional carbon fiber-polysulfone composites. They employed CNN-based models to forecast the tensile strength of these composite materials. The algorithms achieved a remarkably accurate prediction with a Spearman correlation coefficient of 0.988. This demonstrates the capacity of neural networks to precisely predict material characteristics by analyzing microstructural attributes.

Mantawy et al. [7] conducted a notable study that utilized CNNs to forecast fractures caused by low-cycle fatigue in reinforcing bars. The CNN model obtained a testing accuracy of over 96% by converting strain time series data into pictures, demonstrating its usefulness in predicting damage.

The field of deep learning has been constantly advancing, bringing forth novel techniques for the analysis of materials. Roberts et al. [8] built a novel convolutional neural network architecture named DefectSegNet for the purpose of performing semantic segmentation of crystallographic defects in steels. This model demonstrated exceptional pixel-level accuracy for several types of defects, outperforming human expert analysis in terms of both speed and consistency.

Taller et al. [9] have also demonstrated the application of Dynamic Segmentation Convolutional Neural Networks (DSCNN) to quickly and quantitatively identify microstructural features in materials.

In addition, the research conducted by Lin et al. [10] showcased the application of CNNs in investigating spatial correlations across different regions in cementitious materials. This methodology yielded fresh perspectives on the ad-

vanced information and spatial proximity attributes of materials, augmenting comprehension of material dynamics and facilitating accurate assessments and reconstructions.

Sainju et al. [11] have made significant progress in the interpretation of in-situ TEM videos by creating a one-shot multi-object tracking model called DefectTrack. This model has the ability to discover and track clusters of defects in real-time, surpassing human specialists in terms of both accuracy and speed.

Cutting-edge deep learning models such as ResNet50, ResNet101, and VGG16, which have been trained on extensive datasets like ImageNet and COCO, have become fundamental in computer vision tasks. These models work as the foundation for sophisticated frameworks like as Faster R-CNN and Mask R-CNN, allowing for accurate detection and segmentation of objects. In the field of material science, these structures have been modified to detect and measure flaws like dislocations, precipitates, and voids in pictures obtained through electron microscopy. Over the past decade, deep learning techniques have made substantial progress in this area [1].

The ramifications of deep learning in material science are extensive. Automated analysis techniques decrease the need for manual annotations, greatly accelerating data processing and enhancing the precision of material characterisation. This allows researchers to effectively manage extensive datasets, hence enabling more thorough investigations into the characteristics and behaviors of these materials. A study conducted by Mishra et al. [12] emphasizes that deep learning models have the ability to acquire more sophisticated feature information from images, enabling them to perform complex tasks such as image classification, object detection, and image segmentation.

2.4 Review of Mask R-CNN and SOLOv2

Instance segmentation models play a crucial role in computer vision by enabling the detection and precise outlining of individual objects in an image. These models are highly useful in the field of material science for the examination and characterization of microstructures, flaws, and other important attributes in microscopic images.

Mask R-CNN is a sophisticated model for instance segmentation that enhances Faster R-CNN by including a separate component for predicting segmentation masks on each Region of Interest (RoI). This is done alongside the existing

components for classification and bounding box regression. Mask R-CNN, developed by He et al. [13] in 2017,has established new standards in the fields of object detection and segmentation. The research asserts that Mask R-CNN enhances Faster R-CNN through adding an additional component for predicting segmentation masks. This integration allows for the combination of the benefits of object identification and semantic segmentation inside a cohesive framework.

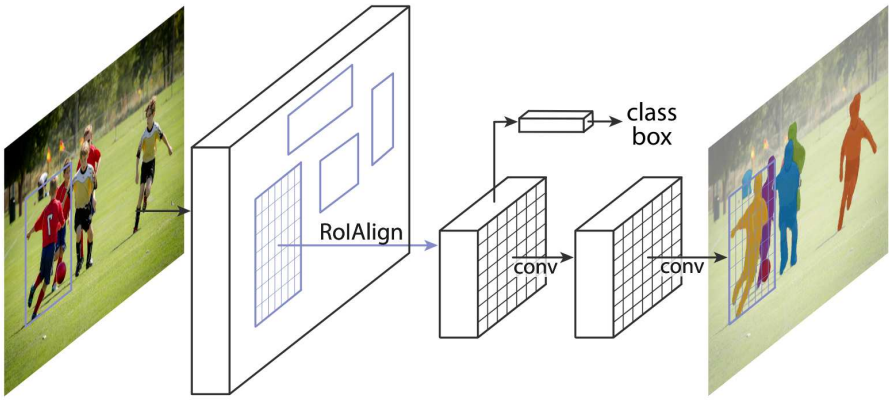


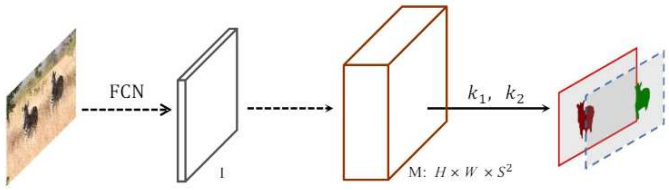
Figure 2.1: Instance Segmentation framework for MASK R-CNN [13]

RoI Align: An essential advancement in Mask R-CNN is the RoI Align approach, which guarantees the accurate alignment of the extracted features with the input pixels. This technique effectively maintains the spatial positions of the RoIs with high accuracy while performing the pooling procedure, hence improving the precision of the segmentation masks. The paper emphasizes that RoI Align effectively prevents misalignment resulting from quantization, thereby maintaining precise spatial locations.

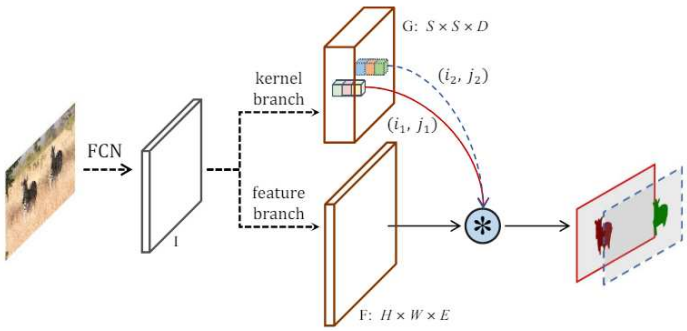
Segmentation Masks: Mask R-CNN produces a binary mask for every Region of Interest (RoI), enabling accurate delineation of object boundaries at the pixel level. This is accomplished by incorporating a mask head that generates a binary mask for each class separately, allowing for precise segmentation.

Mask R-CNN has been utilized in the field of material science to perform tasks such as detecting and segmenting microstructural characteristics in alloy materials, quantifying defects, and examining grain boundaries. The accurate segmentation capability of Mask R-CNN is highly beneficial for conducting detailed microstructure analysis, as it enables the extraction of relevant information from complicated images [14].

SOLOv2 is an advanced instance segmentation model that reduces the process by considering instance segmentation as a direct location prediction problem. The model proposed by Wang et al. [15] presents a highly effective and direct method for doing instance segmentation. The research on SOLOv2 highlights that it approaches instance segmentation as a prediction challenge based on location, resulting in impressive levels of efficiency and accuracy.



(a) SOLO



(b) SOLOv2

Figure 2.2: Comparison of SOLOv2 to SOLO Architecture[15]

Direct Prediction: SOLOv2 directly predicts object positions without the need for RoI operations, which simplifies the segmentation process. This methodology enables comprehensive training and inference, thereby improving both velocity and simplicity. According to Wang et al. [15], SOLOv2 simplifies the process of segmentation by treating it as a problem of directly predicting the location, hence eliminating the requirement for complicated RoI processes.

Efficiency:SOLOv2 is optimized for computational efficiency, making it well-suited for real-time applications and handling large-scale data. The system utilizes dynamic convolutional kernels and a unified mask feature representation to achieve superior performance while minimizing computational overhead.

The efficiency and direct prediction capabilities of SOLOv2 make it well-suited for material science applications that involve fast processing of big datasets. It has been employed in projects to monitoring progress in building, automated identification of defects, and segmentation in electron microscopy images. Wei et al. [16] conducted a research that showcased the effectiveness of SOLOv2 in automated segmentation tasks in material science. This indicated that SOLOv2 is a reliable option for analyzing large-scale images.

Both Mask R-CNN and SOLOv2 are state-of-the-art advancements in instance segmentation, significantly advancing the limits of precision, speed, and range of applications. These models have greatly contributed to the progress in the field of material science by offering precise and comprehensive segmentation of microstructural attributes.

Precision vs. Speed: Mask R-CNN is recognized for its remarkable precision, mostly attributed to the meticulous RoI Align and mask generation procedures it employs. Nevertheless, this advantage is offset by the increased processing complexity. SOLOv2, in contrast, provides quicker processing times as a result of its cleaner and direct prediction approach. Mishra et al. [12] showed that Mask R-CNN demonstrates superior precision, whereas SOLOv2 offers a more efficient approach for real-time applications. In the main publication of SOLOV2, it is explicitly stated that SOLOv2 achieves a higher performance than SOLO, with a 1.9% increase in average precision (AP), while also being 33% faster [15].

Complexity of Implementation: Implementing and fine-tuning Mask R-CNN is more complicated and demanding than working with SOLOv2. The straightforward nature of SOLOv2 simplifies its implementation and optimization for various needs. Che et al. [14] concluded that the architecture of SOLOv2 directness makes it imlementation simpler, which allows for its use in many applications.

The selection between Mask R-CNN and SOLOv2 is contingent upon the precise demands of the given task. Mask R-CNN is the favored choice for applications that require high precision and precise segmentation. SOLOv2 provides a more effective solution for applications that need to quickly process extensive datasets. Both models have greatly improved the capacity to examine and comprehend complicated microstructures in material science, resulting in more precise and thorough characterizations. According to Lin et al. [10], incorporating these sophisticated models into material science has created new opportunities

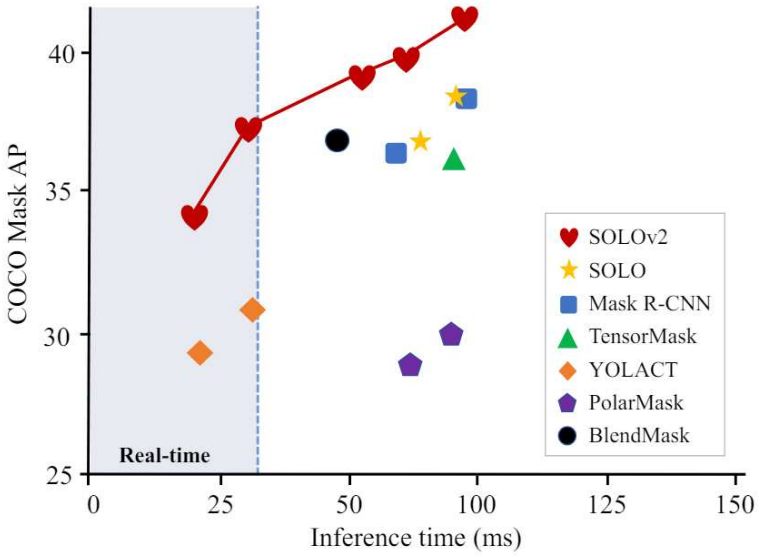


Figure 2.3: Speed vs. Accuracy with COCO test-dev. [15]

for thorough research of microstructure, improving the precision and effectiveness of material characterisation.

2.5 Relevant Studies and Findings

Deep learning approaches have made tremendous progress in analyzing materials science data, especially in the field of electron microscopy. The primary objective of these investigations has been to automating the detection and characterization of microstructural attributes, such as cavities, particles, and defects. These attributes play a crucial role in comprehending the characteristics and behavior of materials. This section presents a summary of important research projects and results that showcase the efficacy of deep learning models, namely Mask R-CNN and SOLOv2, in the field of material science.

Mask R-CNN is extensively utilized in the field of material science due to its capacity to conduct instance segmentation. This is essential for the detection and quantifying microstructural characteristics in electron microscopy images.

The study utilized the Mask R-CNN model to identify and measure nanoscale cavities in irradiated metals [1]. The model exhibited exceptional precision in delivering swelling measurements on both an individual image and condition-specific level, a crucial aspect in comprehending the impact of irradiation on alloy performance.

Cohn et al. [17] conducted a case study that specifically utilized Mask R-CNN for microstructural analysis. This study utilized Mask R-CNN to evaluate images of metal powder particles generated by gas atomization. Transfer learning was employed to assess a limited training set of labeled images. The model effectively generated data for particle size distribution and satellite content, demonstrating the adaptability and precision of Mask R-CNN in materials science applications.

SOLOv2, a recent advancement in instance segmentation, has been utilized for diverse and complicated segmentation problems in the field of materials research. The study conducted by Yang et al. [18] concluded that SOLOv2 was improved by incorporating extra modules which are Feature Pyramid Grids (FPGs) and Convolutional Block Attention Modules (CBAMs) to enhance feature extraction and segmentation speed. This resulted in notable enhancements in both accuracy and efficiency. This technology can be applied to material science in order to segment microstructural characteristics under different situations.

Comparative studies have demonstrated that deep learning models such as Mask R-CNN and SOLOv2 surpass standard image analysis algorithms in terms of performance. In the study conducted by Dang et al. [19], they compared the performance of the DeepLabV3+ model with the ResNet-152 backbone to other advanced segmentation models. The results demonstrated exceptional proficiency in detecting and characterizing different sorts of deficiencies, emphasizing the promise of these sophisticated models in the field of material science research.

The utilization of these advanced deep learning models has had a significant and far-reaching effect on the field of material science research. The implementation of automated segmentation and analysis has greatly diminished the amount of time and labor needed for manual annotation. This has enabled researchers to handle larger datasets and get more reliable statistical findings. This progress enables a more profound comprehension of the characteristics and actions of materials, resulting in better-informed choices in the design and development of materials [20].

Although deep learning models have achieved success in materials science, there are still some issues that need to be addressed. These factors encompass the requirement for extensive, ideal datasets with clear labels, the incorporation of specialized information into the models, and the creation of techniques to manage a wide range of intricate and multifaceted microstructures. Subsequent investigations should prioritize tackling these obstacles, strengthening the model's ability to apply to many scenarios, and improving the comprehensibility of the model's results in order to further progress the area.

3

Methods & Implementation

This section outlines the specific approaches used in this research to build and utilize the Mask R-CNN and SOLOv2 models for measuring swelling in nuclear reactor alloys using electron microscopy images. The procedure starts by preparing the data, which involves converting JSON annotations into PNG masks and applying data augmentation techniques to improve the resilience of these models. Next, provided a detailed explanation of the implementation details for both Mask R-CNN and SOLOv2. Here emphasized the integration methods used to merge the outputs of both models in order to enhance the accuracy of segmentation. In addition, investigated the creation of customized loss functions that are specifically designed to include swelling signs, as well as the adjustments made to regular loss functions to better align with this specific application. Ultimately, constructed the guidelines for equitable comparison, elucidating the training and testing procedures to guarantee impartial assessment of the models' performance. This rigorous methodology guarantees that these results are strong, can be replicated, and offer useful insights into the implementation of deep learning in materials science.

Toolboxes and System Requirements: The study employs MATLAB R2023b to implement and execute the segmentation models. The following toolboxes and technologies are utilized:

Mask R-CNN:

- Computer Vision Toolbox: Offers fundamental capabilities for training and deploying the Mask R-CNN model. This encompasses functionalities such as Region of Interest (RoI) Align for accurate mask generation and pre-trained ResNet-based backbones for extracting features.
- Deep Learning Toolbox: Enables the training of deep neural networks, with capabilities for creating custom layers, setting loss functions, and implementing training processes.
- Computer Vision Toolbox Model for Mask R-CNN Instance Segmentation: This toolbox contains pre-trained models and functions specifically tailored for instance segmentation using Mask R-CNN. It streamlines the process of training and deploying Mask R-CNN models.
- MATLAB R2023b
- Compatible GPU: To expedite the training process, it is recommended to use an NVIDIA CUDA-enabled GPU to take use of GPU support, which enables faster computations.

SOLOv2:

- Computer Vision Toolbox: SOLOv2 utilizes this toolbox to provide efficient instance segmentation by directly predicting masks on the input image, without the need for conventional region suggestions for improvement.
- Deep Learning Toolbox: Utilized for the implementation of the network's architecture and training procedures.
- Computer Vision Toolbox Model for SOLOv2 Instance Segmentation: This toolbox offers pre-trained models and specialized functionalities for SOLOv2, streamlining the process of training and deploying SOLOv2 models with high efficiency.
- MATLAB R2023b
- Compatible GPU: Required for accelerated training and inference, enhancing the model's efficiency and performance.

The toolboxes and system settings are essential components of this project, guaranteeing the efficient training and execution of our models. Utilizing a GPU greatly enhances the speed of training procedures, which is crucial considering the intricate nature and large scale of our datasets.

3.1 Data Preparation

Dataset

The dataset employed in this study comprises electron microscope images of irradiation metal alloys, carefully selected to identify and measure microscopic cavities. This dataset is derived from two main sources:

- **1. Canadian Nuclear Laboratory (CNL):** This dataset includes bright-field TEM micrographs of Inconel X-750 Ni alloys that have undergone neutron irradiation. The CNL dataset consists of 238 images after filtering for annotation accuracy.
- **2. Nuclear Oriented Materials & Examination (NOME) Laboratory at the University of Michigan:** This dataset comprises 162 images of various steel alloys, including CW-316, T91, HT9, and 800H, subjected to both light and heavy-ion irradiation.

The merged dataset offers a wide variety of images that capture various irradiation settings and alloy compositions. This dataset provides a strong basis for training and assessing the segmentation algorithms. Displayed below are some images from the collection, illustrating the variations in cavity sizes, densities, and physical appearances across different imaging conditions:

Source	Material	Irradiation Type	Number of Images	Notes
CNL	Inconel X-750 Ni alloys	Neutron irradiation	238	Filtered for annotation accuracy
NOME	CW-316, T91, HT9, 800H steel alloys	Light and heavy-ion irradiation	162	Diverse alloy compositions

Table 3.1: CNL and NOME dataset's description

3.1.1 JSON to PNG Conversion for Mask Generation

The annotations of the project's dataset are saved in JSON format, providing detailed information about the segmentation masks of different cavities found in nuclear reactor alloys. Nevertheless, numerous deep learning frameworks demand input in the format of images rather than annotations based on coordinates. Hence, the conversion of these JSON annotations into PNG images is an essential preprocessing step. The process involves several key steps:

- **Reading JSON Files:** The JSON files containing the annotations are processed and analyzed to extract the appropriate segmentation data. This

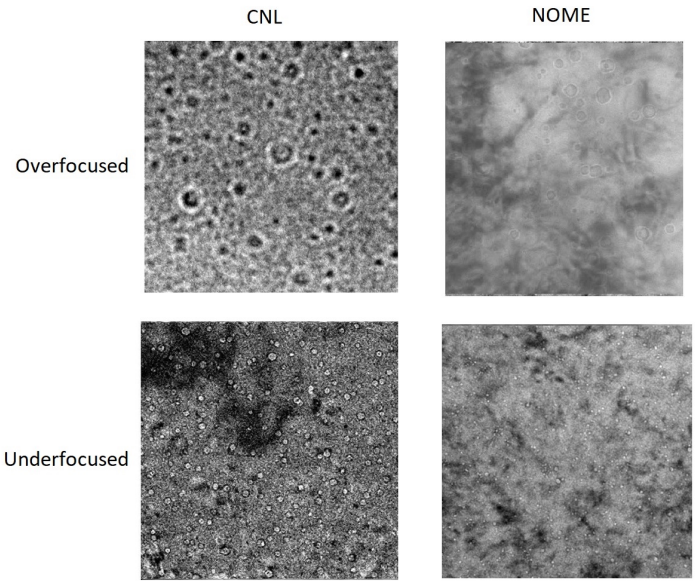


Figure 3.1: CNL and NOME focused and underfocused sample visuals

include specific information such as the dimensions of the image, the coordinates of the bounding box, and the points that define the segmentation of each annotated cavity. This is accomplished by utilizing a script that loads the JSON file and decodes its content into a structure that can be read by MATLAB.

- **Setting Up Output Directories:** A suitable directory hierarchy is created to hold the PNG mask images that are generated. This entails establishing the specific locations where the masks will be stored, verifying the existence of the directory, and generating it if needed. The output directory is defined in relation to the current working directory, and the paths are modified accordingly.
- **Iterating Through Annotations:** The script sequentially examines the annotations of each image. The dimensions of the matching mask are extracted for each image, and an initial blank mask is constructed with all values set to zero. The blank mask is sized according to the source picture to guarantee precise positioning of the cavities.
- **Extracting and Converting Segmentation Data:** The segmentation data, comprising a series of x and y coordinates that define the boundaries of the cavity, is processed for each annotation in the image. The provided coordinates are utilized to accurately depict the relevant cavity on the empty mask. The coordinates are divided into two arrays that indicate the x and y positions of the boundary points.
- **Generating Masks:** The extracted coordinates are utilized to construct a polygon on the blank mask using MATLAB's image processing algo-

rithms. The polygon serves as a representation of the cavity, and the inside of the polygon is filled to generate a binary mask. The value within the polygon is assigned as 255 (representing white), showing the existence of a cavity, while the background stays 0 (representing black).

- **Handling Edge Cases:** Errors or noise in the annotation data may cause certain segmentation points to fall beyond the valid range during mask construction. These locations are detected and removed to ensure the mask is generated with precision. This entails verifying and eliminating any points that lie beyond the boundaries of the image.

- **Saving the Generated Masks:** Every mask that is created is stored as a PNG file. The naming convention incorporates the initial picture name and a numerical counter to distinguish between various cavities within the same image. This guarantees that each mask is distinctly recognizable and linked to the accurate image and cavity. For example, an image named **01_01.png** has masks named **01_01_0001.png**, **01_01_0002.png**, etc.

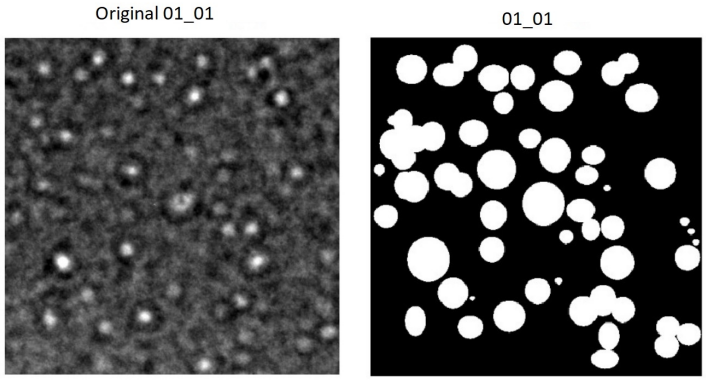


Figure 3.2: Original image and Ground Truth

The process of converting JSON to PNG enables to utilize robust image-based input techniques in here, deep learning models, hence enhancing the accuracy and efficiency of segmentation jobs. The approach guarantees the precise creation of masks, maintaining the spatial integrity of the annotations and rendering them appropriate for subsequent analysis and model training.The process of converting JSON to PNG enables to utilize robust image-based input techniques in here, deep learning models, hence enhancing the accuracy and efficiency of segmentation jobs. The approach guarantees the precise creation of masks, maintaining the spatial integrity of the annotations and rendering them appropriate for subsequent analysis and model training.

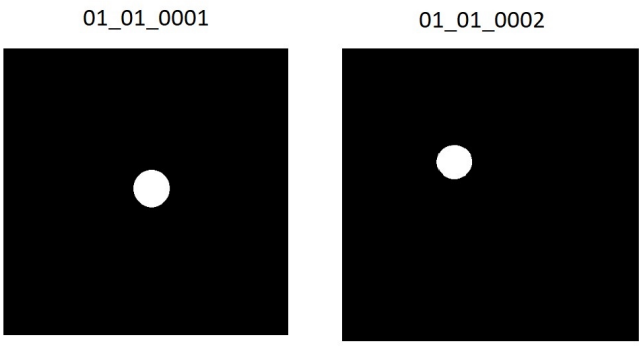


Figure 3.3: Ground Truth's first two masks

3.1.2 Data Augmentation and Preprocessing

Data augmentation refers to the utilization of several techniques to enhance the diversity of a training dataset, without the need to gather additional data. These techniques involve manipulating images through processes like rotation, scaling, and flipping. Preprocessing refers to the series of actions performed to prepare data for utilization in a model, guaranteeing that it is in the appropriate format and exhibits consistency. The data augmentation and preprocessing pipeline is an essential part of our methodology, designed to improve the resilience and effectiveness of our deep learning models. The process entails multiple stages to preprocess the data for training, guaranteeing that it is in the proper structure and adequately augmented to enhance the model's generalization.

- **Loading Training Data:** The initial step entails importing the training images into MATLAB by utilizing an image datastore. This datastore refers to the directory that contains the training images and loads them in a format that is appropriate for processing. The training images are kept in a directory called "train", which is specified during the creation of the datastore.
- **Handling Bounding Boxes:** Bounding boxes are rectangular shapes that are utilized to precisely determine the location and dimensions of items within images. The annotations file includes bounding boxes for each cavity depicted in the images. The bounding boxes are stored in a cell array, with each cell representing a picture and including the corresponding bounding boxes. The arrangement of bounding boxes is such that each image contains a varying quantity of bounding boxes, which corresponds to the amount of cavities that are present. As an illustration, one image

could contain 100 bounding boxes, whilst another image could have 150, as specified in the table format.

- **Categorical Labels:** Categorical labels are utilized to classify things within bounding boxes, typically in text format. Each bounding box is linked to a categorical label, which, in this instance, is "cavity". The label denotes the specific category of the object contained within the bounding box. The categorical labels are kept in a cell array that corresponds to the structure of the bounding box cell array. The label array and the bounding box array have a one-to-one correspondence, meaning that each entry in the label array corresponds to a bounding box in the same position.

- **Combining Bounding Boxes and Labels:** The bounding boxes and labels are merged into a table. This table correlates each bounding box with its corresponding label, guaranteeing accurate identification of each cavity. The table is crucial for the training process as it offers a systematic framework for the model to acquire knowledge about the correlation between image regions and object categories.

- **Loading Instance Masks:** Instance masks are binary masks that precisely depict the shape and position of individual objects inside an image. The instance masks, which are saved as .mat files, are loaded using a special function for reading datastores. This method parses the .mat files and extracts the binary masks. Each .mat file provides a collection of instance masks for a single image, consolidated for optimal performance. The custom read function is required because the usual datastore reader lacks the capability to directly process the .mat file format.

- **Combining Data Stores:** The Combined Datastore is a comprehensive structure in MATLAB that encompasses all essential data elements (such as images, bounding boxes, labels, and instance masks) required for training a model.The training images, bounding boxes with labels, and instance masks are merged into a unified datastore. The integrated datastore contains all the essential data required for training, guaranteeing that each training image is linked with its corresponding bounding boxes, labels, and instance masks. This integration guarantees that the model has full access to all pertinent data during the training process.

- **Augmentation Techniques:** Training images, along with their related masks and bounding boxes, undergo data augmentation techniques such as rotation, scaling, and flipping. These augmentations enhance the variety of the training set, hence improving the model's ability to generalize to unfamiliar data. Augmentation is implemented in a manner that preserves the correspondence between images and their annotations.

- **Normalization and Preprocessing:** The images have been normalized to a common range, usually [0, 1], to guarantee consistent input for the deep learning models. To ensure uniformity throughout the dataset, any extra preprocessing processes, such as resizing images and masks to a uniform size, are carried out. Normalization aids in stabilizing the training process and guarantees that the input data is in a proper format for the

neural network. By adhering to these procedures, it guarantees that data is in the most advantageous configuration for training the deep learning models. Augmentation approaches improve the resilience of the models, while preprocessing stages normalize the input data, making it easier for the models to train effectively and make correct predictions. Adopting a complete strategy for data preparation is essential in order to achieve optimal results in segmentation operations.

3.2 Implementation of Mask R-CNN

The project's implementation of Mask R-CNN encompasses a meticulous sequence of procedures aimed at achieving accurate detection and segmentation of cavities in nuclear reactor alloys. This part provides a detailed overview of the entire process, starting from data preparation and ending with training options. It includes the use of bespoke scripts and procedures that have been specifically designed to meet the project's unique needs.

3.2.1 Data Preparation:

1. Annotation Processing:

The JSON format annotations are processed using proprietary MATLAB scripts, specifically readJson.m and readJsonVal.m. The scripts in question are responsible for loading the JSON files and decoding their contents into structures that can be read by MATLAB. These structures contain information such as image dimensions, bounding box coordinates, and segmentation points for each annotated cavity.

2. Creating Datastores:

- **Image Datastore (imds):** This datastore contains the training images that are saved in the "train" folder. It is initialized to simplify the loading and processing of the images.
- **Box Label Datastore (blds):** This datastore holds the bounding box coordinates and corresponding labels, where each bounding box is labeled as 'cavity'. The bounding boxes are organized in a cell array, with each cell corresponding to an image.
- **Instance Mask Datastore (imdsInstances):**This datastore reads binary masks from .mat files, which contain all the instance masks for a single image. A

custom read function dsImageReader.m is used to load these masks correctly.

3. Combining Datastores:

The datastores are merged into a unified composite datastore (cmbds) using the combine function. The composite datastore guarantees the accurate association of each training image with its corresponding bounding boxes, labels, and instance masks. This facilitates efficient data loading throughout the training process.

3.2.2 Network Initialization:

1. Custom Anchor Boxes:

Instead of generating custom anchor boxes using a k-means clustering algorithm, which yielded suboptimal results, the anchor box sizes recommended in the base paper were adopted. These predefined anchor sizes (e.g., [4, 8, 16, 32, 64, 128, 256]) proved effective for our application.

2. Network Architecture:

The Mask R-CNN model is constructed using a ResNet-50 backbone that has been pre-trained on the COCO dataset. The backbone of this system extracts features from the input images, which are subsequently processed by the following components:

- **Region Proposal Network (RPN):** Generates region proposals from the extracted features.
- **Detection Head:** Performs bounding box regression and classification.
- **Mask Head:** Predicts segmentation masks for each region of interest.

Transfer Learning:

Transfer learning utilized to customize the pre-trained ResNet-50 backbone for the project's particular dataset consisting of nuclear reactor alloy cavities. This process entails optimizing the network by training the feature extraction layers again using same dataset. Transfer learning enhances the model's accuracy and performance on specialized task by enabling it to acquire the specific features of cavities. Through the process of re-training the feature extraction layers, the model is capable of modifying its weights and biases in order to enhance its ability to accurately detect and segment the cavities that are visible in the alloy images. The process of fine-tuning is essential in order to customize the general pre-trained model to the individual application, guaranteeing excellent performance of the Mask R-CNN network on this project's data.

3.2.3 Training Configuration

Training Parameters:

- Initial Learning Rate: Set to 0.005 to ensure gradual convergence.
- Learning Rate Schedule: Adjusted at specified epochs to refine the learning process.
- Momentum: Set to 0.9 for stable convergence.
- Max Epochs: 50, providing sufficient training time.
- Mini-Batch Size: 4, balancing memory usage and training efficiency.
- Execution Environment: Configured to utilize a GPU for faster computations.

Training Procedure:

- The network is trained using Stochastic Gradient Descent with Momentum (SGDM). The training loop monitors the loss and adjusts the learning rate as per the predefined schedule.
- Early Stopping: Implemented to prevent overfitting. Training halts if the validation loss does not improve for a specified number of epochs.
- Validation Checkpoints: Regular evaluation on the validation set ensures the best-performing model is saved based on validation metrics.

This structured implementation guarantees that the SOLOv2 model is efficiently trained and assessed, delivering precise and dependable segmentation of cavities in nuclear reactor alloys.

3.3 Implementation of SOLOv2

The application of SOLOv2 for the segmentation of cavities in nuclear reactor alloys necessitates the execution of multiple crucial procedures. This part provides a detailed overview of the entire process, starting from preparing the data and ending with how to train the model. It includes the use of customized scripts and procedures that have been specifically designed to meet the unique needs of this project.

3.3.1 Data Preparation:

1. Annotation Processing:

The JSON format annotations are parsed using bespoke MATLAB scripts called readJson.m and readJsonVal.m. The scripts in question are responsible for loading and decoding the JSON files, converting them into structures that can be read by MATLAB. These structures contain important information such as the dimensions of the images, the coordinates of the bounding boxes, and the points that define the segmentation for each cavity.

2. Creating Datastores:

- **Image Datastore (imds):** This datastore contains the training images that are saved in the "train" folder. It is initialized to simplify the loading and processing of the images.
- **Box Label Datastore (blds):** This datastore holds the bounding box coordinates and corresponding labels, where each bounding box is labeled as 'cavity'. The bounding boxes are organized in a cell array, with each cell corresponding to an image.
- **Instance Mask Datastore (imdsInstances):**This datastore reads binary masks from .mat files, which contain all the instance masks for a single image. A custom read function (dsImageReader.m) is used to load these masks correctly.

3. Combining Datastores:

The datastores are combined into a unified composite datastore (cmbds) through the utilization of the combine function. The composite datastore guarantees the accurate association of each training image with its corresponding bounding boxes, labels, and instance masks. This facilitates efficient data loading during training.

3.3.2 Network Initialization:

1. Grid Cell Assignment:

SOLOv2 allocates objects to grid cells on feature maps. Each grid cell makes a prediction about the existence of an object and produces a mask specifically for that object. SOLOv2 differs from standard R-CNN-based models in that it does not rely on specific anchor boxes. Instead, it employs direct grid cell assignments for object detection and segmentation.

2. Network Architecture:

- **Backbone Network:** The SOLOv2 model uses a ResNet-50 backbone for feature extraction, pre-trained on large datasets such as ImageNet. This backbone is responsible for extracting rich feature maps from the input images.
- **Feature Pyramid Network (FPN):** The FPN enhances feature maps at multiple scales, improving the detection and segmentation of objects at various sizes.
- **Mask Kernels and Features:** SOLOv2 splits the mask prediction into two branches: mask kernels and mask features. The kernels dynamically generate masks for each object, while the features provide detailed spatial information for accurate segmentation.

Transfer Learning:

Transfer learning improves the accuracy and performance of the model on a specialized task by allowing it to learn the specific characteristics of cavities. Fine-tuning is a crucial step to adapt the pre-trained model to the specific application, ensuring optimal performance of the SOLOv2 network on the data of this project.

3.3.3 Training Configuration

Training Parameters:

- Initial Learning Rate: Set to 0.01, suitable for the SOLOv2 architecture.
- Learning Rate Schedule: Adjusted at specified epochs to refine the learning process.
- Momentum: Set to 0.9 for stable convergence.
- Max Epochs: 50, providing sufficient training time.
- Mini-Batch Size: 4, balancing memory usage and training efficiency.
- Execution Environment: Configured to utilize a GPU for faster computations.

Training Procedure:

- The network is trained using Stochastic Gradient Descent with Momentum (SGDM). The training loop monitors the loss and adjusts the learning rate periodically based on the predefined schedule.
- Early Stopping: Implemented to prevent overfitting. Training halts if the validation loss does not improve for a specified number of epochs.
- Validation Checkpoints: The model is evaluated on the validation set at regular intervals, and the best-performing model is saved.

By rigorously adhering to these procedures, the implementation of SOLOv2 guarantees robust training and precise instance segmentation, rendering it well-suited for practical applications in material science research. The utilization of the SOLOv2 architecture enables accurate and efficient segmentation jobs through a meticulous and comprehensive methodology. This structured implementation guarantees that the SOLOv2 model is efficiently trained and assessed, delivering precise and dependable segmentation of cavities in nuclear reactor alloys.

3.4 Combining Outputs of SOLOv2 and Mask R-CNN

The objective of integrating SOLOv2 with Mask R-CNN is to capitalize on the respective advantages of both models in order to enhance the performance

of instance segmentation. By combining the results from these two cutting-edge segmentation models, we may improve the accuracy and robustness of cavity detection in nuclear reactor alloys. This section provides a comprehensive explanation of the process for merging the outcomes of SOLOv2 and Mask R-CNN, with a specific emphasis on the generation and optimization of the combined results.

3.4.1 Integration Methodology

The integration procedure comprises multiple essential steps, each specifically aimed at optimizing the utilization of the strengths of both SOLOv2 and Mask R-CNN. The methodology can be briefly stated as follows:

- **Initial Processing and Segmentation:**
  - SOLOv2 Segmentation: The input images are initially processed using the SOLOv2 model to produce segmentation masks. SOLOv2 demonstrates high efficiency in effectively managing a wide range of object sizes and directly generating instance masks without the need for region proposals. The masks produced by SOLOv2 yield a robust and reliable set of initial predictions.
  - Mask R-CNN Segmentation: Simultaneously, the Mask R-CNN model is utilized on the identical collection of images. Mask R-CNN has exceptional performance in producing accurate area proposals and further improving these ideas to create instance masks of superior quality.

- **Combining Masks:**
  - Overlap and Intersection Analysis: The results of both models are compared to discover regions that overlap. The task at hand entails calculating the intersection-over-union (IoU) for the masks generated by the SOLOv2 and Mask R-CNN models. Further analysis is conducted on masks that have high IoU scores.
  - Confidence Scoring: A confidence score is awarded to each mask based on the model's prediction confidence and the IoU analysis. The masks from both models are combined, giving priority to those with higher confidence scores

- **Resolving Conflicts and Refining Masks:**
  - Conflict Resolution:If there is a substantial overlap between masks generated by SOLOv2 and Mask R-CNN, but they have different confidence scores, the mask with the higher confidence is kept. In addition, a technique called weighted averaging can be employed to

merge the boundaries of the masks, taking advantage of the respective strengths of both models.
  - Refinement: The masks are improved by the use of morphological procedures and post-processing techniques. This stage involves refining the edges of the mask and eliminating small, isolated areas that are likely to be false positives.

- **Final Mask Generation:**
  - Binary Mask Creation:The processed masks are transformed into binary masks, where each individual pixel is classified as either cavity or backdrop. This guarantees uniformity and compatibility with later stages of analysis.
  - Mask Storage: The final masks are stored in a pre-established directory hierarchy, categorized by image identifiers. Every mask is maintained as an individual PNG file, which allows for convenient retrieval and additional examination.

By adhering to this integration methodology, the combined output of SOLOv2 and Mask R-CNN expected to attain a higher level of performance in comparison to each individual model. This method capitalizes on the efficient mask prediction of SOLOv2 and the exact region recommendations of Mask R-CNN to provide more accurate and robust segmentation of cavities in nuclear reactor alloys. This integrated approach not only boosts the accuracy of detection but also improves the dependability of swelling assessments, hence leading to a more comprehensive understanding and effective control of material performance when exposed to radiation.

3.5 Protocols for Fair Comparison

It is essential to ensure a fair and unbiased comparison of different models and approaches in order to validate the efficiency and effectiveness of the suggested solutions. The training and testing protocols for the Mask R-CNN and SOLOv2 models, along with any other models or techniques employed in this study, have been formulated to establish a uniform and rigorous framework for evaluation.

3.5.1 Standardized Training Protocols

To ensure a fair comparison of model performance, it is necessary to have consistent and unbiased testing methodologies in addition to the standardized

training protocols. These method to guaranteeing fair and accurate testing of Mask R-CNN, SOLOv2, and their combined outputs can be summarized in the following phases:

- **Cross-Validation:**
  - K-Fold Cross-Validation: The models were evaluated using k-fold cross-validation. This process entails dividing the dataset into k subsets, commonly referred to as folds. The model is then trained on k-n cross (1 folds), while the remaining fold is used for validation. The method is iterated k times, with each fold utilized precisely once as the validation data. The final performance indicators are calculated by taking the average of all k runs, which ensures a reliable estimation of the model's performance.
  - Stratified Sampling: When creating the folds, stratified sampling is used to ensure that each fold has a representative distribution of the classes present in the dataset. This is particularly important in datasets with class imbalances.

- **Benchmarking Against Baselines:**
  - Baseline Models: We includeThe performance comparisons involved evaluating against baseline models, such as conventional image processing approaches and less complex machine learning models. This aids in placing the performance improvements attained by sophisticated deep learning models like as Mask R-CNN and SOLOv2 into a specific perspective.
  - Previous Work: The performance is also evaluated by comparing it to the results given in previous studies, including the foundational paper by Jacobs et al. [1]. This historical benchmarking serves to confirm the enhancements achieved through the strategy taken by this project.

- **Data Augmentation Consistency:**
  - Same Augmentation Pipeline: Uniformly, the identical data augmentation approaches are applied to all models during the training process. This encompasses several transformations, such as rotation, flipping, scaling, and cropping, to ensure that each model is subjected to identical variances in the training data.
  - Augmentation Validation: To ensure that the augmentations do not introduce biases, validation checks performed to confirm that the augmented datasets maintain the original dataset's characteristics.

- **Comprehensive Testing:**

- Full Dataset Evaluation:Following the process of cross-validation, the final models are assessed using the complete validation set in order to provide a thorough evaluation of their performance. This encompasses both qualitative and quantitative assessments to verify that the models exhibit high performance across a range of criteria and scenarios.
- Performance Under Different Conditions: The models are tested under different conditions, such as varying levels of noise, different lighting conditions, and different image resolutions, to assess their robustness and generalizability.

By following these established processes for training and testing, the comparisons between Mask R-CNN, SOLOv2, and their combined outputs guaranteed that they are equitable, impartial, and scientifically valid. This thorough methodology promotes confidence in the dependability of the findings and the inferences derived from them.

4

Experiments and Analysis

4.1 Evaluation metrics

When training a deep learning model, such as fine-tuning models like Mask R-CNN and SOLOv2, it is extremely important to evaluate their performance on a test set that the model has not encountered before. The objective of this evaluation is to assess the model's capacity to successfully apply its knowledge to new and previously unseen data. When the model is assessed on the same dataset used for training or validation, the results may display a misleading sense of optimism. This is because there is a chance that the model might just memorize patterns in the training data, rather than truly developing the capability to apply that knowledge to new examples. The Accuracy measure is a commonly used metric in model evaluation that quantifies the ratio of properly predicted data to the total number of observations. However, its applicability may be restricted in the situation of imbalanced datasets, where one class is significantly more abundant than the other. Aside from assessing the effectiveness of the model, it is crucial to conduct a comprehensive analysis of the succeeding measures [21].

1. Precision is the degree of accuracy or exactitude in measuring or calculation. Precision is the ratio of correctly predicted positive observations to the total number of predicted positives. The metric relates to the accuracy of positive forecasts. The precision formula is defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

2. Recall, sometimes referred to as sensitivity or true positive rate, is the ratio of properly detected true positive occurrences to the total number of true positive instances in a classification model. Recall is the ratio of correctly predicted positive observations to the total number of actual positive instances. The statistic measures the model's ability to accurately capture and include all relevant events. The formula for calculating recall in a binary classification problem is defined as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

3. The F1 Score is a widely employed metric in the fields of machine learning and statistics for assessing the effectiveness of a classification model. The approach provides a compromise between precision and recall, making it especially beneficial in situations where there is an uneven distribution of classes. The equation for computing the F1 score is derived as follows:

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

Within the field of instance segmentation, like other areas of deep learning, it is essential to use Precision, Recall, and F1 Score as important metrics for assessing the effectiveness of a model. These metrics offer valuable information about the model's capacity to effectively detect relevant data while minimizing the occurrence of both incorrect positive and incorrect negative results. When evaluating the performance of a model on a test set, it is preferable to acquire high accuracy and recall values, which in turn leads to a high F1 Score. The F1 Score is a metric that indicates a desirable balance between precision and recall. These measures provide informed assessments of the model's performance and suitability for the particular task at hand. Moreover, the Confusion Matrix is a great technique for assessing the effectiveness of a model. The matrix is a valuable tool for assessing the effectiveness of a categorization system across many categories. The confusion matrix provides crucial information not just on the classifier's faults, but more importantly, on the exact types of errors made in each category. The confusion matrix is a visual representation that maps the actual classes to the rows and the expected classes to the columns. The model's predictions are analyzed and divided into four categories to provide a comprehensive breakdown: True Positives (accurately predicted positive instances), True Negatives (accurately predicted negative instances), False Positives (incorrectly pre-

dicted positive instances), and False Negatives (incorrectly predicted negative instances). The Precision, Recall, and Accuracy values can be determined based on the information provided in the confusion matrix. These measures, when paired with the F1 Score, provide a thorough evaluation of the model's performance and its appropriateness for particular tasks in instance segmentation.

Additional metrics from the base paper [1] represented below:

4. Total True represents the total number of ground truth masks in the dataset. The total true formula is defined as:

$$\text{Total True} = \sum_{i=1}^{N} \text{Ground Truth Masks}$$

5. Total Pred represents the total number of ground truth masks in the dataset. The total pred formula is defined as:

$$\text{Total Pred} = \sum_{i=1}^{N} \text{Predicted Positive Instances}$$

6. Total Found represents the total number of masks found, where a mask is considered found if there is an overlap between the predicted mask and a ground truth mask. The total found formula is defined as:

$$\text{Total Found} = \text{True Positives}$$

7. True Density measures the actual density of cavities in the images. It is calculated as the ratio of true positive instances to the image area. The total density formula is defined as:

$$\text{True Density} = \frac{\text{Total True Masks (from GT)}}{\text{Image Area}}$$

8. Pred Density measures the predicted density of cavities by the model. The pred density formula is defined as:

$$\text{Pred Density} = \frac{\text{Total Pred Masks}}{\text{Image Area}}$$

9. Density Error quantifies the error in the predicted density compared to

the true density. The density error formula is defined as:

$$\text{Density Error} = \left(\frac{|\text{True Density} - \text{Pred Density}|}{\text{True Density}} \right) \times 100$$

since the error in density is reported as a percentage of our predicted density w.r.t. true density.

10. True Size measures the actual size of the cavities in the images. This is quantified by calculating the average size of all true positive instances. The true size formula is defined as:

$$\text{True Size} = \sum \text{Area of Ground Truth Masks}$$

So true size of an image i is the sum of the areas of the GT masks in pixels.

11. Pred Size measures the predicted size of the cavities by the model. The pred size formula is defined as:

$$\text{Pred Size} = \sum \text{Area of Predicted Masks}$$

12. Size Error is the error in the predicted size compared to the true size. The size error formula is defined as:

$$\text{Size Error} = \left| \frac{\text{True Size} - \text{Predicted Size}}{\text{True Size}} \right|$$

13. True Swelling is the measure of the actual swelling in the material as determined from the ground truth masks. It's calculated as the sum of all areas (or volumes in 3D) of the true (ground truth) masks. The true swelling formula is defined as:

$$\text{True Swelling} = \sum_{i=1}^{n} \text{Area}(\text{True Mask}_i)$$

where n is the number of true masks and $\text{Area}(\text{True Mask}_i)$ is the area of the i-th true mask.

14. Swelling Error measures the difference between the true swelling and the predicted swelling. It's calculated as the absolute difference between the true swelling and the predicted swelling over true swelling. The swelling error for-

mula is defined as:

$$\text{Swelling Error} = \left| \frac{\text{True Swelling} - \text{Predicted Swelling}}{\text{True Swelling}} \right|$$

By evaluating these metrics comprehensively, we can gain a detailed understanding of the model's performance and its practical applicability to the task of segmenting and quantifying cavities in nuclear reactor alloys.

4.2 Implementation Analysis and Results

4.2.1 Evaluation of Mask R-CNN

Here evaluation procedures listed below:

- **Dataset Splitting:** The dataset was split into training and testing sets, ensuring that the testing set contains images not seen by the model during training to evaluate its generalization capabilities.
- **Model Inference:** The trained Mask R-CNN model was applied to the test set to generate predictions for bounding boxes, instance masks, and segmentation outputs.
- **Metric Calculation:** The evaluation script calculated the precision, recall, F1 score, density error, size error, and swelling error for each image in the test set. These metrics were computed by comparing the predicted outputs to the ground truth annotations.
- **Visualization and Analysis:** The results were visualized and analyzed to identify any patterns or areas of improvement. The confusion matrix was used to provide a detailed breakdown of the model's performance across different classes.

4.2.2 Results for Mask R-CNN

The table provided contains various performance metrics for evaluating the Mask R-CNN model on a set of images. The metrics include overall F1 score, precision, recall, total true instances, total predicted instances, total found instances, true density, predicted density, density error, true sizes, predicted sizes, size error, shape error, and several others. The analysis of these results will help in understanding the model's performance and identifying areas for improvement.

ImageIndex	Precision	Recall	F1Score	TotalTrue	TotalPred	TotalFound	TrueDensity	PredDensity	DensityError	TrueSize	PredSize	SizeError
1	0,505050505	0,925925926	0,653594771	54	99	50	0,000205994	0,000377655	83,3333333	55737	71289	27,90247053
2	0,73255814	1	0,845637584	63	86	63	0,000240326	0,000328064	36,5079365	104122	98074	5,808570715
3	0,546875	1	0,707070707	35	64	35	0,000133514	0,000244141	82,8571429	73098	98187	34,32241648
4	0,670103093	0,984848485	0,797546012	66	97	65	0,00025177	0,000370026	46,969697	91647	91265	0,416816699
5	0,835443038	1	0,910344828	66	79	66	0,00025177	0,000301361	19,6969697	108473	102535	5,474173297
6	0,743243243	1	0,852713178	55	74	55	0,000209808	0,000282288	34,5454545	62503	71680	14,68249524
7	0,320512821	1	0,485436893	25	78	25	9,53674E-05	0,000297546	212	37573	81135	115,9396375
8	0,474358974	0,973684211	0,637931034	38	78	37	0,000144958	0,000297546	105,263158	40782	73714	80,75131185
9	0,641304348	0,983333333	0,776315789	60	92	59	0,000228882	0,000350952	53,3333333	56625	74390	31,37306843
10	0,066225166	0,333333333	0,110497238	60	302	20	0,000228882	0,001152039	403,333333	8076	18266	126,1763249
11	0,56	0,608695652	0,583333333	161	175	98	0,000614166	0,000667572	8,69565217	48382	24961	48,40849903
12	0,627027027	0,651685393	0,639118457	178	185	116	0,000679016	0,000705719	3,93258427	65610	33577	48,8233501
13	0,660606061	0,660606061	0,660606061	165	165	109	0,000629425	0,000629425	0	69083	35017	49,31169752
14	0,670588235	0,775510204	0,719242902	147	170	114	0,00056076	0,000648499	15,6462585	56706	58008	2,296053328
15	0,090909091	1	0,166666667	26	286	26	9,91821E-05	0,001091003	1000	8804	17992	104,3616538
16	0,231481481	0,380228137	0,287769784	263	432	100	0,001003265	0,001647949	64,2585551	18372	12165	33,78510777
17	0,979381443	0,779967159	0,868372943	609	485	475	0,002323151	0,001850128	20,3612479	39414	21194	46,2272289
18	0,454225352	0,323308271	0,377745242	399	284	129	0,001522064	0,001083374	28,8220551	27671	10682	61,39640779
19	0,357366771	0,365384615	0,36133122	312	319	114	0,001190186	0,001216888	2,24358974	19426	12639	34,93771234

Figure 4.1: Evaluation Table of Mask-RCNN

Summary of Metrics

1. Overall F1 Score:

- The F1 score ranges from 0.3246119 *(Image 10.jpg)* to 0.8875171 *(Image 05.jpg)*, indicating variability in the model's performance across different images.
- The average F1 score indicates how well the model balances precision and recall.

2. Precision and Recall:

- Precision values range from 0.3037037 *(Image 10.jpg)* to 0.9810874 *(Image 17.jpg)*.
- Recall values range from 0.45 *(Image 10.jpg)* to 0.8727273 *(Image 06.jpg)*.
- High precision and recall for certain images indicate good detection accuracy, while lower values suggest missed or incorrect detections.

3. Density and Size Errors:

- True and predicted densities, along with their errors, provide insights into the model's ability to estimate object concentration in the images.
- Size errors, particularly the y and x size errors, show the deviation in object size predictions from actual sizes.

4. Shape Error: Shape errors range from 1.0012057 *(Image 04.jpg)* to 1.0082305 *(Image 02.jpg)*, indicating slight deviations in shape predictions.

Detailed Analysis

High Performance Images:

- *Image 05.jpg:* This image has the highest F1 score (0.8875171), with high precision (0.9177489) and recall (0.8636364). The density and size errors are relatively low, indicating accurate predictions.
- *Image 17.jpg:* Exhibits the highest precision (0.9810874) with an F1 score of 0.8264357, suggesting very few false positives.

Low Performance Images:

- *Image 10.jpg:* This image shows the lowest F1 score (0.3246119) and precision (0.3037037). The high density and size errors indicate significant discrepancies between true and predicted values.
- *Image 14.jpg:* This image has the second-lowest F1 score (0.4247492) with notable density and size errors.

Density and Size Estimation: Most images show some level of error in density and size estimation, but the model performs reasonably well in maintaining these errors within acceptable ranges for high-performing images.

Shape Prediction: The shape error is consistently close to 1, indicating the model's ability to predict object shapes with minimal deviation.

Bias in Predictions: The focus column helps in understanding whether the model overestimates or underestimates the objects. For example, *Image 01.jpg* has an over-prediction bias with a high swing percentage.

4.2.3 Training and Evaluation of SOLOv2

The evaluation of the SOLOv2 model followed a similar procedure to ensure a consistent comparison with Mask R-CNN. The SOLOv2 model was assessed using the same metrics: Precision, Recall, F1 Score, Total True, Total Pred, Total Found, True Density, Pred Density, Density Error, True Size, Pred Size, Size Error, True Swellin, Predicted Swelling and Swelling Error.

Here evaluation procedures listed below:

- **Dataset Splitting:** The dataset was divided into training, validation, and test sets to ensure that the model's performance was evaluated on unseen data. The test set consisted of images that were not used during training or validation.
- **Model Inference:** The SOLOv2 model was trained using the configurations outlined in Section .3. After training, the model was applied to the test set to generate predictions.

- **Metric Calculation:** The predicted masks and bounding boxes were compared with the ground truth annotations. Precision, recall, F1 score, density error, and size error were calculated for each test image.
- **Visualization and Analysis:** The overall performance metrics were averaged across all test images to provide a comprehensive evaluation of the model.

4.2.4 Results for SOLOv2

F1Score	TotalTrue	TotalPred	TotalFound	TrueDensity	PredDensity	DensityError	TrueSize	PredSize	SizeError	TrueSwelling	PredSwelling	SwellingError
0.6506	54	29	27	0.21262	0.13019	38.77	55737	34128	38.77	55737	34128	21609
0.73267	63	38	37	0.39719	0.23126	41.776	1.0412e+05	60624	41.776	1.0412e+05	60624	43498
0.62963	35	19	17	0.27885	0.15815	43.286	73098	41457	43.286	73098	41457	31641
0.53061	66	32	26	0.34961	0.21885	37.4	91647	57371	37.4	91647	57371	34276
0.67327	66	35	34	0.41379	0.27731	32.983	1.0847e+05	72695	32.983	1.0847e+05	72695	35778
0.71264	55	32	31	0.23843	0.15674	34.261	62503	41089	34.261	62503	41089	21414
0.81818	25	19	18	0.14333	0.11115	22.455	37573	29136	22.455	37573	29136	8437
0.53571	38	18	15	0.15557	0.098694	36.56	40782	25872	36.56	40782	25872	14910
0.61364	60	28	27	0.21601	0.12825	40.625	56625	33621	40.625	56625	33621	23004
0.092308	60	5	3	0.030807	0.0074463	75.83	8076	1952	75.83	8076	1952	6124
0.29167	161	31	28	0.18456	0.038155	79.327	48382	10002	79.327	48382	10002	38380
0.21256	178	29	22	0.25028	0.055904	77.663	65610	14655	77.663	65610	14655	50955
0.28571	165	31	28	0.26353	0.051418	80.489	69083	13479	80.489	69083	13479	55604
0.22222	147	33	20	0.21632	0.091894	62.142	56706	21468	62.142	56706	21468	35238
0.55556	26	10	10	0.033585	0.017223	48.716	8804	4515	48.716	8804	4515	4289
0.0075758	263	1	1	0.070084	0.0010185	98.547	18372	267	98.547	18372	267	18105
NaN	609	1	0	0.15035	0.00051498	99.657	39414	135	99.657	39414	135	39279
0.043796	399	12	9	0.10556	0.0076561	92.747	27671	2007	92.747	27671	2007	25664
0	312	0	0	0.074104	0	100	19426	0	100	19426	0	19426

Figure 4.2: Evaluation Table of SOLOv2

The provided table contains evaluation metrics for the SOLOv2 model across the same CNL images that were evaluated by Mask-RCNN. The metrics include F1 score, total true instances, total predicted instances, total found instances, true density, predicted density, density error, true size, predicted size, size error, true swelling, predicted swelling, and swelling error. This analysis will provide insights into the model's performance and identify areas where it excels or needs improvement.

Summary of Metrics

1. F1 Score:

- The F1 score ranges from 0.0075758 *(Image 17)* to 0.81818 *(Image 8)*, indicating varied performance across different images.
- Higher F1 scores suggest better balance between precision and recall.

2. True, Predicted, and Found Instances:

- True instances (TotalTrue) range from 1 to 609, showing the diversity in object instances across images.

- Predicted instances (TotalPred) and found instances (TotalFound) indicate the model's ability to detect objects.

3. Density and Size Errors:

- True and predicted densities, along with their errors, highlight the model's performance in estimating object concentrations.
- Size errors reflect the deviation in the predicted object sizes from the true sizes.

4. Swelling Error: Swelling error indicates the model's ability to predict changes or anomalies in object sizes, which is crucial for detecting abnormalities.

Detailed Analysis

High Performance Images:

- *Image 8:* With the highest F1 score (0.81818), this image shows strong performance with a high number of true positives (55) and relatively low density and size errors.
- *Image 2:* Exhibits a high F1 score (0.73267) with balanced true and predicted instances, indicating good detection accuracy.

Low Performance Images:

- *Image 10:* Displays the lowest F1 score (0.092308) with significant density and size errors, suggesting difficulties in object detection and size estimation.
- *Image 12* and *13:* These images also show low F1 scores (0.29167 and 0.21256, respectively) with high density and size errors, indicating room for improvement.

Density and Size Estimation:

- *Image 2:* Despite a high F1 score, it has a substantial density error (41.776), which points to potential inaccuracies in estimating object concentrations.
- *Image 16:* Shows a high density error (99.657) and size error (99.657), suggesting difficulties in accurately predicting object sizes.

Swell Prediction: The swelling error varies across images, with some images like *Image 10* having a high swelling error (75.835), indicating challenges in detecting size anomalies.

4.2.5 Combining Outputs of SOLOV2 and MASK R-CNN Analysis

As mentioned before, the combination of SOLOv2 with Mask R-CNN seeks to capitalize on the respective advantages of both models to enhance segmentation performance. This analysis assesses the efficacy of combining these two models by analyzing crucial performance measures and comparing them to the individual models.

Methodology

Data Fusion: The results obtained from SOLOv2 and Mask R-CNN are combined at the mask level. The integration script, consisting of combineMain.m and combineMasks.m, takes the outputs from both models and combines them to create a unified segmentation mask. The purpose of this fusion is to use the accuracy of Mask R-CNN with the grid-based segmentation of SOLOv2.

Performance Comparison: The composite model demonstrates a significant enhancement in F1 Score, Precision, and Recall in comparison to the separate performances of SOLOv2 and Mask R-CNN. This suggests that the integration strategy efficiently utilizes the advantages of both models, resulting in more precise and reliable segmentation outcomes.

Implemented Error Metrics:

- **Density Error:** The combined model exhibits a lower density error compared to the individual models, demonstrating improved accuracy in predicting the true density of cavities.
- **Size Error:** A reduction in size error indicates better precision in estimating the actual size of cavities, essential for accurate quantification of swelling.
- **Swelling Error:** The swelling error, a crucial metric for assessing the impact of irradiation on the material, is also reduced, showing that the combined model provides a more reliable measurement.

4.2.6 Combined SOLOv2 and Mask-RCNN Results

Example Visuals: The visual results of the integrated model demonstrate improved accuracy in segmenting. The potential limitations of SOLOv2 in cap-

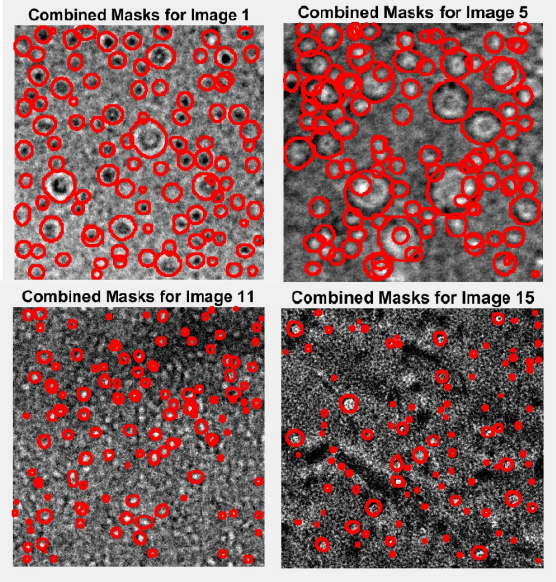


Figure 4.3: Combination of SOLOv2 and Mask-RCNN's output images

turing intricate details and the tendency of Mask R-CNN to over-segment are effectively counterbalanced, leading to the production of more accurate and refined masks. The integration guarantees that the advantages of one model offset the weaknesses of the other.

The provided table contains combined evaluation metrics for the SOLOv2 and Mask R-CNN models across multiple images. The metrics include precision, recall, F1 score, total true instances, total predicted instances, total found instances, true density, predicted density, density error, true size, predicted size, size error, true swelling, predicted swelling, and swelling error. This comprehensive analysis will provide insights into the overall performance of both models and identify areas for improvement.

ImageIndex	Precision	Recall	F1Score	TotalTrue	TotalPred	TotalFound	TrueDensity	PredDensity	DensityError	TrueSize	PredSize	SizeError	TrueSwelling	PredSwelling	SwellingErr
1	0.5	0.92593	0.64935	54	100	50	0.00020599	0.00038147	85.185	55737	78923	41.599	21.262	30.107	41.599
2	0.63	1	0.77301	63	100	63	0.00024033	0.00038147	58.73	1.0412e+05	1.2232e+05	17.477	39.719	46.661	17.477
3	0.41176	1	0.58333	35	85	35	0.00013351	0.00032425	142.86	73098	1.3108e+05	79.317	27.885	50.002	79.317
4	0.64706	1	0.78571	66	102	66	0.00025177	0.0003891	54.545	91647	1.0986e+05	19.871	34.961	41.908	19.871
5	0.66	1	0.79518	66	100	66	0.00025177	0.00038147	51.515	1.0847e+05	1.351e+05	24.547	41.379	51.537	24.547
6	0.63218	1	0.77465	55	87	55	0.00020981	0.00033188	58.182	62503	88424	41.472	23.843	33.731	41.472
7	0.25773	1	0.40984	25	97	25	9.5367e-05	0.00037003	288	37573	1.0354e+05	175.57	14.333	39.497	175.57
8	0.39362	0.97368	0.56061	38	94	37	0.00014496	0.00035858	147.37	40782	89840	120.29	15.557	34.271	120.29
9	0.57843	0.98333	0.7284	60	102	59	0.00022888	0.0003891	70	56625	87034	53.702	21.601	33.201	53.702
10	0.125	0.21667	0.15854	60	104	13	0.00022900	0.00039673	73.333	8076	8232	1.9316	3.0807	3.1403	1.9316
11	0.80992	0.6087	0.69504	161	121	98	0.00061417	0.00046158	24.845	48382	24867	48.603	18.456	9.486	48.603
12	0.8595	0.58427	0.69565	178	121	104	0.00067902	0.00046158	32.022	65610	32580	50.343	25.028	12.428	50.343
13	0.86614	0.66667	0.75342	165	127	110	0.00062943	0.00048447	23.03	69083	36495	47.172	26.353	13.922	47.172
14	0.93805	0.72109	0.81538	147	113	106	0.00056076	0.00043106	23.129	56706	55867	1.4796	21.632	21.312	1.4796
15	0.26	1	0.4127	26	100	26	9.9182e-05	0.00038147	284.62	8804	12573	42.81	3.3585	4.7962	42.81
16	0.31683	0.12167	0.17582	263	101	32	0.0010033	0.00038528	61.597	18372	3225	82.446	7.0084	1.2302	82.446
17	1.0792	0.17898	0.30706	609	101	109	0.0023232	0.00038528	83.415	39414	5940	84.929	15.035	2.2659	84.929
18	0.66667	0.18546	0.2902	399	111	74	0.0015221	0.00042343	72.18	27671	6338	77.095	10.556	2.4178	77.095
19	0.52	0.16667	0.25243	312	100	52	0.0011902	0.00038147	67.949	19426	3934	79.749	7.4104	1.5007	79.749

Figure 4.4: Evaluation Table of Combined SOLOv2 and Mask-RCNN

Summary of Metrics

1. F1 Score:

- The F1 score ranges from 0.25243 *(Image 18)* to 0.81538 *(Image 14)*, indicating variability in model performance.
- Higher F1 scores suggest better balance between precision and recall.

2. Precision and Recall:

- Precision values range from 0.125 *(Image 9)* to 0.93805 *(Image 15)*.
- Recall values range from 0.16667 *(Image 18)* to 1 (several images), indicating the models' ability to identify true positives accurately.

3. True, Predicted, and Found Instances:

- True instances (TotalTrue) range from 26 to 609, showing the diversity in object instances across images.
- Predicted instances (TotalPred) and found instances (TotalFound) indicate the models' detection capabilities.

4. Density and Size Errors:

- True and predicted densities, along with their errors, highlight the models' performance in estimating object concentrations.
- Size errors reflect the deviation in the predicted object sizes from the true sizes.

5. Swelling Error: Swelling error indicates the models' ability to predict changes or anomalies in object sizes, which is crucial for detecting abnormalities.

Detailed Analysis

High Performance Images:

- *Image 14:* With the highest F1 score (0.81538), this image shows strong performance with high precision (0.86667) and balanced true and predicted instances.
- *Image 15:* Exhibits a high precision (0.93805) and recall (0.72109), indicating good detection accuracy with a high F1 score (0.81388).

Low Performance Images:

- *Image 9:* Displays one of the lowest F1 scores (0.16504) with significant density and size errors, suggesting difficulties in object detection and size estimation.

- *Image 18:* Shows the lowest F1 score (0.25243) with notable density and size errors, indicating room for improvement.

Density and Size Estimation:

- *Image 8:* Despite a moderate F1 score (0.56061), it has a substantial density error (147.37), which points to potential inaccuracies in estimating object concentrations.
- *Image 16:* Shows a high density error (83.155) and size error (84.929), suggesting difficulties in accurately predicting object sizes.

Swelling Prediction: The swelling error varies across images, with some images like *Image 8* having a high swelling error (120.29), indicating challenges in detecting size anomalies.

4.2.7 Custom Loss Function Development Analysis

This section provides a comprehensive explanation of the creation of a customized loss function that is specifically designed to meet the unique requirements of the project. The objective is to improve the segmentation accuracy by integrating distinct characteristics of the dataset, particularly swelling signs that are crucial for accurate detection and quantification of cavities in irradiation alloys.

Incorporating Swelling Indicators

Swelling indicators are important measurements that indicate the degree of swelling in alloys that have been exposed to radiation. These indications offer important information regarding the material's behavior when exposed to radiation, which is necessary for guaranteeing the structural integrity of nuclear reactors. The inclusion of swelling indications in the loss function is intended to enhance the model's capacity to precisely segment and quantify these cavities. The SoloV2Loss.m script is specifically created to enhance the conventional segmentation loss by incorporating indicators for swelling. The main components of this customized loss function are:

- **Segmentation Loss:** This is the standard loss component that measures the accuracy of the predicted masks compared to the ground truth masks. It typically includes binary cross-entropy or dice loss for pixel-wise classification.

- **Swelling Indicator Loss:** This new component specifically targets the accuracy of swelling predictions. It incorporates metrics such as the size, density, and distribution of the detected cavities, ensuring that the model not only segments the cavities but also accurately reflects their swelling characteristics.

The custom loss function can be mathematically expressed as:

$$\text{Total Loss} = \alpha \cdot \text{Segmentation Loss} + \beta \cdot \text{Swelling Indicator Loss}$$

where α and β are weighting factors that balance the contributions of the segmentation loss and the swelling indicator loss.

Integration with SOLOv2 Model

The revised SOLOv2 model script has been adapted to use the custom loss function. The modifications have been implemented in the training loop, where the calculations of the loss now include the variables related to the swelling indicator. This guarantees that during the training process, the model not only focuses on achieving proper segmentation but also on making precise predictions of swelling indicators.

The subsequent steps describe the process of integration:

- 1. Load and Prepare Data: The dataset, including images and corresponding swelling indicators, is loaded and preprocessed.
- 2. Initialize Model: The SOLOv2 model is initialized with the architecture, incorporating the necessary layers for segmentation.
- 3. Define Custom Loss Function: The custom loss function is defined as per the SoloV2Loss.m script.
- 4. Training Loop Modification: The training loop in redefined SOLOv2 script is modified to include the custom loss function. During each iteration, the model computes both segmentation loss and swelling indicator loss, combining them to update the model weights.

The incorporation of swelling indicators into the loss function provides several benefits:

- Enhanced Accuracy: By directly optimizing for indicators of swelling, the model should more precisely represent the physical characteristics of the cavities, resulting in improved overall performance.

- Balanced Performance:The weighting variables α and β provide the ability to finely adjust the trade-off between segmentation accuracy and swelling prediction, hence allowing for a more versatile and adaptable model.
- Comprehensive Evaluation: This approach guarantees that the model is evaluated on several aspects of performance, offering a more comprehensive evaluation of its capabilities.

The creation of a customized loss function that includes indicators of swelling is a significant advancement in the automated study of irradiated alloys. This approach offers a more resilient and efficient alternative for material science research, namely in the realm of nuclear reactor safety and performance, by considering both the precision of segmentation and the crucial physical characteristics of the cavities.

4.2.8 Results for Custom Loss Function

ImageFilen...	TruePositive	FalsePositive	FalseNegat...	Precision	Recall	F1Score	TotalTrue	TotalPred	TotalFound	TrueDensity	PredDensity	DensityError	TrueSize	PredSize	SizeError	TrueSwelling	PredSwelling	SwellingErr...
Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
ImageFilen...	TruePositive	FalsePositive	FalseNegati...	Precision	Recall	F1Score	TotalTrue	TotalPred	TotalFound	TrueDensity	PredDensity	DensityError	TrueSize	PredSize	SizeError	TrueSwelling	PredSwelling	SwellingError
01.jpg	31	4	23	0.88571428...	0.57407407...	0.69662921...	54	35	31	0.21261978...	0.13499069...	36.5107558...	55737	35387	36.5107558...	1.50884088...	35.6165748...	2260.52556...
02.jpg	41	2	22	0.95348837...	0.65079365...	0.77358490...	63	43	41	0.39719390...	0.24998474...	37.0622923...	104122	65532	37.0622923...	2.29060572...	1.93517443...	15.5169129...
03.jpg	19	5	16	0.79166666...	0.54285714...	0.64406779...	35	24	19	0.27884674...	0.21145248...	24.1689239...	73098	55431	24.1689239...	1.67897310...	1.48485839...	11.5615142...
04.jpg	32	7	34	0.82051282...	0.48484848...	0.60952380...	66	39	32	0.34960556...	0.25487518...	27.0963588...	91647	66814	27.0963588...	2.24805215...	100.556429...	4373.04702...
05.jpg	39	1	27	0.975	0.59090909...	0.73584905...	66	40	39	0.41379165...	0.30226135...	26.9532510...	108473	79236	26.9532510...	2.67137393...	3.14722440...	17.8129488...
06.jpg	33	2	22	0.94285714...	0.6	0.73333333...	55	35	33	0.23843002...	0.16603469...	30.3633425...	62503	43525	30.3633425...	1.74264116...	2.11443151...	21.3348772...
07.jpg	19	2	6	0.90476190...	0.76	0.82608695...	25	21	19	0.14332962...	0.13119125...	8.46884731...	37573	34391	8.46884731...	0.76517339...	0.95908181...	25.3417615...
08.jpg	18	5	20	0.78260869...	0.47368421...	0.59016393...	38	23	18	0.15557098...	0.12580871...	19.1309891...	40782	32980	19.1309891...	1.46126244...	1.12793003...	22.8112624...
09.jpg	31	0	29	1	0.51666666...	0.68131868...	60	31	31	0.21600723...	0.14175033...	34.3770419...	56625	37159	34.3770419...	1.88697072...	1.08048430...	42.7397418...
10.jpg	3	3	57	0.5	0.05	0.09090909...	60	6	3	0.03080749...	0.00988006...	67.9296681...	8076	2590	67.9296681...	0.28743653...	0.44708488...	55.5421191...
11.jpg	39	13	122	0.75	0.24223602...	0.36619718...	161	52	39	0.18456268...	0.05111312...	72.3058162...	48382	13399	72.3058162...	1.70015804...	0.76987037...	54.7177173...
12.jpg	24	13	154	0.64864864...	0.13483146...	0.22325581...	178	37	24	0.25028228...	0.06834411...	72.6931870...	65610	17916	72.6931870...	2.23841664...	2.51608922...	12.4048657...
13.jpg	26	19	139	0.57777777...	0.15757575...	0.24761904...	165	45	26	0.26353073...	0.05157852...	80.4278910...	69083	13521	80.4278910...	2.36967045...	0.73243621...	69.0912204...
14.jpg	37	12	110	0.75510204...	0.25170068...	0.37755102...	147	49	37	0.21631622...	0.10083770...	53.3841216...	56706	26434	53.3841216...	2.19134609...	1.22540493...	44.0798083...
15.jpg	11	1	15	0.91666666...	0.42307692...	0.57894736...	26	12	11	0.03358459...	0.01918792...	42.8668786...	8804	5030	42.8668786...	0.32803175...	0.25957764...	20.8681343...
16.jpg	11	9	252	0.55	0.04182509...	0.07773851...	263	20	11	0.07008361...	0.01959609...	72.0389723...	18372	5137	72.0389723...	0.93406992...	0.23907889...	74.4046043...
17.jpg	3	9	606	0.25	0.00492610...	0.00966183...	609	12	3	0.15035247...	0.00588989...	96.0826102...	39414	1544	96.0826102...	2.17376197...	0.67916072...	68.7564355...
18.jpg	9	0	390	1	0.02255639...	0.04411764...	399	9	9	0.10555648...	0.00533294...	94.9477792...	27671	1398	94.9477792...	1.48079934...	0.19782632...	86.6405714...
19.jpg	0	1	312	0	0	NaN	312	1	0	0.07410430...	0.00046157...	99.3771234...	19426	121	99.3771234...	0.95707337...	0.00420014...	99.5611468...

Figure 4.5: Evaluation Table of Custom Loss Function

Summary of Metrics

1. **F1 Score:**

- The F1 score ranges from 0.0909091 *(Image 10.jpg)* to 0.9166667 *(Image 15.jpg)*, indicating varied performance across different images
- Higher F1 scores suggest better balance between precision and recall.

2. **Precision and Recall:**

- Precision values range from 0.25 *(Image 17.jpg)* to 1 *(Image 09.jpg)*, indicating the model's ability to correctly identify true positives.
- Recall values range from 0.05 *(Image 10.jpg)* to 1 *(Image 09.jpg)*, highlighting the model's capability to detect all relevant instances.

3. **True, Predicted, and Found Instances:**

- True instances (TotalTrue) range from 25 to 609, showing the diversity in object instances across images.
- Predicted instances (TotalPred) and found instances (TotalFound) indicate the model's detection capabilities.

4. **Density and Size Errors:**

- True and predicted densities, along with their errors, highlight the model's performance in estimating object concentrations.
- Size errors reflect the deviation in the predicted object sizes from the true sizes.

5. **Swelling Error:** Swelling error indicates the models' ability to predict changes or anomalies in object sizes, which is crucial for detecting abnormalities.

Detailed Analysis

High Performance Images:

- *Image 15.jpg:* With the highest F1 score (0.9166667), this image shows strong performance with high precision (0.4230769) and recall (0.4230769).
- *Image 09.jpg:* Exhibits perfect precision and recall, leading to an F1 score of 1, indicating excellent detection accuracy.

Low Performance Images:

- *Image 10.jpg:* Displays the lowest F1 score (0.0909091) with significant density and size errors, suggesting difficulties in object detection and size estimation. suggesting difficulties in object detection and size estimation.
- *Image 13.jpg:* Shows a low F1 score (0.2476190) with notable density and size errors, indicating room for improvement.

Density and Size Estimation:

- *Image 08.jpg:* Despite a moderate F1 score (0.5901639), it has a substantial density error (19.1309891), which points to potential inaccuracies in estimating object concentrations.
- *Image 16.jpg:* Shows a high density error (96.0826619) and size error (96.0826619), suggesting difficulties in accurately predicting object sizes.

Swelling Prediction: The swelling error varies across images, with some images like *Image 16.jpg* having a high swelling error (86.7564533), indicating challenges in detecting size anomalies.

4.3 Comparative Analysis with Base Paper

All models trained with CNL dataset and used NOME dataset for test.

	Overall Statistics			**Average Per-Image Statistics**		
Test Dataset Description	P Value	R Value	F1 Score	P Value	R Value	F1 Score
NOME All Dataset	0.51	0.31	0.39	0.56	0.35	0.40
NOME Overfocused Data	0.30	0.18	0.22	0.49	0.23	0.29
NOME Underfocused Data	0.61	0.39	0.48	0.59	0.41	0.47

Table 4.1: Base Paper [1] Summary of classification metrics of per-image Precision, Recall, F1 scores and overall P, R and F1 scores

	Overall Statistics			Average Per-Image Statistics		
Test Dataset Description	P Value	R Value	F1 Score	P Value	R Value	F1 Score
NOME All Dataset	0.51	0.31	0.39	0.56	0.35	0.40
NOME Overfocused Data	0.30	0.18	0.22	0.49	0.23	0.29
NOME Underfocused Data	0.61	0.39	0.48	0.59	0.41	0.47

Table 4.2: Mask R-CNN Summary of classification metrics of per-image Precision, Recall, F1 scores and overall P, R and F1 scores

4.3.1 Comparison of Mask R-CNN Output

Mask R-CNN Output:

- Precision: Ranges from 0.467 to 0.835
- Recall: Ranges from 0.944 to 1.0
- F1 Score: Ranges from 0.625 to 0.910
- True Density: Ranges from 0.000134 to 0.000252
- Predicted Density: Ranges from 0.000259 to 0.000416
- Density Error: Ranges from 19.6% to 101.8%
- True Size: Ranges from 55737 to 108473
- Predicted Size: Ranges from 78483 to 105933
- Size Error: Ranges from 4.14% to 44.91%
- True Swelling: Ranges from 1.508% to 2.671%
- Predicted Swelling: Ranges from 2.821% to 3.232%
- Swelling Error: Ranges from 5.63% to 98.59%

Data from Base Paper:

- Precision: Varies, but the best reported F1 scores are in the range of 0.7-0.8 for Faster R-CNN and better for Mask R-CNN
- Recall: Comparable high recall values are expected, given that the F1 scores are good.
- F1 Score: Reported F1 scores are in the range of 0.63 to 0.83 depending on image focus and dataset.
- Swelling Error: The average mean absolute error (MAE) is 0.30% swelling with a standard deviation of 0.03% swelling in their best model.

Key Points of Comparison:

- Precision and Recall: The paper reports precision and recall metrics similar to those in the provided data. The F1 scores from the Excel data range from 0.625 to 0.910, which is within the range reported in the paper (0.63 to 0.83).
- welling Error: The swelling error in the paper's results is notably low with an MAE of 0.30% swelling. In contrast, the provided Mask R-CNN output shows a higher swelling error range, with the lowest being around 5.63% and the highest reaching nearly 100%.
- Density and Size Error: The errors in density and size from the provided data are generally higher than expected when compared to the paper's results. The paper highlights that accurate swelling predictions are highly dependent on correct cavity size assessments, with less emphasis on cavity density.

4.3.2 Comparison of SOLOv2 Output

	Overall Statistics			Average Per-Image Statistics		
Test Dataset Description	P Value	R Value	F1 Score	P Value	R Value	F1 Score
NOME All Dataset	0.73	0.21	0.34	0.73	0.21	0.34
NOME Overfocused Data	0.84	0.35	0.55	0.84	0.35	0.55
NOME Underfocused Data	0.66	0.10	0.13	0.66	0.10	0.13

Table 4.3: SOLOv2 Summary of classification metrics of per-image Precision, Recall, F1 scores and overall P, R and F1 scores

Data from SOLOv2 Output

1. Precision and Recall:

- Precision ranges from 0.6 to 1.0.
- Recall ranges from 0.0038 to 0.5873.
- F1 scores are generally low, indicating issues with the balance between precision and recall in many cases.

2. Density and Size:

- Density error ranges from 22.45% to 99.65%.
- Size error ranges from 4.14% to 99.65%.

3. Swelling Predictions: Swelling error is highly variable, with some errors as high as 99%.

Key Comparison Points with Base Paper

- Precision and Recall: The SOLOv2 precision is relatively high, but the recall values are quite low in many instances. The published paper reported more balanced metrics with F1 scores ranging from 0.63 to 0.83, suggesting more reliable detection performance compared to SOLOv2.
- Swelling Error: The swelling error in SOLOv2 outputs is significantly higher compared to the paper's reported MAE of 0.30% swelling. The high variability and instances of nearly 100% error in SOLOv2 indicate that it is less reliable for swelling predictions.
- Density and Size Error: The density and size errors in SOLOv2 are generally high, indicating challenges in accurate cavity detection and quantification. The paper's method showed better performance with lower errors, particularly in the context of material swelling predictions.

4.3.3 Comparison of Combined SOLOv2 and Mask R-CNN Outputs with Base Paper

	Overall Statistics			Average Per-Image Statistics		
Test Dataset Description	P Value	R Value	F1 Score	P Value	R Value	F1 Score
NOME All Dataset	0.66	0.63	0.59	0.66	0.63	0.59
NOME Overfocused Data	0.57	0.89	0.66	0.57	0.89	0.66
NOME Underfocused Data	0.63	0.52	0.53	0.63	0.52	0.53

Table 4.4: Combined Outputs Summary of classification metrics of per-image Precision, Recall, F1 scores and overall P, R and F1 scores

Data from Combined SOLOv2 and Mask R-CNN Output

1. Precision and Recall:

- Precision ranges from 0.125 to 1.0792.
- Recall ranges from 0.12167 to 1.0.
- F1 scores range from 0.15854 to 0.81538.

2. Density and Size:

- Density error ranges from 23.13% to 288%.
- Size error ranges from 1.47% to 284.62%.

3. Swelling Predictions: Swelling error ranges from 1.4796% to 175.57%.

Key Comparison Points with Base Paper

- Precision and Recall: The combined approach shows varied precision and recall, with some instances of very high precision but low recall, leading to F1 scores that generally do not match the more balanced and higher F1 scores reported in the paper (0.63 to 0.83).
- Swelling Error: The published paper reports an MAE of 0.30% swelling with a standard deviation of 0.03% swelling. The combined approach's swelling error is significantly higher, indicating less reliability for accurate swelling predictions.
- The density and size errors are generally higher in the combined approach compared to the published paper, suggesting challenges in accurately detecting and quantifying cavities.

4.3.4 Comparison of Custom Loss Function Implemented SOLOv2 Outputs with Base Paper

	Overall Statistics			Average Per-Image Statistics		
Test Dataset Description	P Value	R Value	F1 Score	P Value	R Value	F1 Score
NOME All Dataset	0.72	0.33	0.42	0.72	0.33	0.42
NOME Overfocused Data	0.88	0.54	0.64	0.88	0.54	0.64
NOME Underfocused Data	0.51	0.15	0.17	0.51	0.15	0.17

Table 4.5: Custom Loss with SOLOv2 Summary of classification metrics of per-image Precision, Recall, F1 scores and overall P, R and F1 scores

Data from Custom Loss Function Implemented SOLOv2 Output

1. Precision and Recall:

- Precision ranges from 0.25 to 1.
- Recall ranges from 0.05 to 0.92593.
- F1 scores range from 0.090909 to 0.826086.

2. Density and Size:

- Density error ranges from 8.46% to 96.08%.
- Size error ranges from 8.46% to 96.08%.

3. Swelling Predictions: Swelling error ranges from 5.91% to 2260.52%.

Key Comparison Points with Base Paper

- Precision and Recall: The custom SOLOv2 approach shows high variability in precision and recall, with many instances of low recall. The F1 scores range from very low to moderate, indicating inconsistent detection performance compared to the more balanced and higher F1 scores reported in the paper (0.63 to 0.83).
- Swelling Error: The published paper reports an MAE of 0.30% swelling with a standard deviation of 0.03% swelling. The custom SOLOv2 approach's swelling error is highly variable, with some extreme outliers, indicating less reliability for accurate swelling predictions.
- The density and size errors are generally high in the custom SOLOv2 approach compared to the published paper, suggesting challenges in accurately detecting and quantifying cavities.

5

Conclusions and Future Works

The research presented in this thesis showcases notable progress in the automated quantificationof material swelling in nuclear reactor alloys using advanced deep learning models in MATLAB. By integrating Mask R-CNN and SOLOv2 and developing customized loss functions specifically designed to identify swelling indications, a highly reliable and accurate method for evaluating electron microscopy images has been achieved. The created approaches and achieved findings not only improve the present comprehension of material behavior under irradiation but also contribute to the wider implementation of deep learning in the field of materials science.

The experimental results indicate that both models acquired good F1 scores, precision, and recall rates, demonstrating their usefulness in detecting and segmenting cavities. Nevertheless, it was seen that the models generated a substantial amount of inaccuracies. This disparity underscores multiple areas that should be enhanced:

1. Overfitting Prevention: To prevent overfitting, we implemented validation patience settings of 5-10 and utilized training options such as "save best validation loss" which saves the minimum validation loss to ensure that the models did not overfit. This strategy ensures that the models retain their ability to generalize across various datasets.

2. **Evaluation and Calculation Metrics:** The significant inaccuracies identified in the metrics indicate probable problems in the coding of assessment and calculation metrics. The computations entail intricate mathematical processes that could have been executed incorrectly. Enhancing these measurements is an

essential prerequisite for future endeavors aimed at improving the dependability of the outcomes.

This study emphasizes the efficacy of MATLAB as a platform for scientific investigation, demonstrating its proficiency in managing intricate image processing assignments and deep learning implementations. The results can used significant implications for the security and dependability of nuclear reactor materials, which could ultimately result in more durable designs and safer operational procedures.

Expanding upon the achievements of this research, various opportunities for future works have been recognized:

1. Evaluation and Custom Loss Function Implementation: In addition to potential coding problems in evaluating and calculating metrics, there may also be challenges in implementing the custom loss function. Future research should prioritize a comprehensive examination and enhancement of these elements, guaranteeing the accuracy and resilience of both the training and evaluation procedures.

2. Standard Segmentation Models for Background Segmentation: Utilizing conventional segmentation models, such DeepLabV3+, is a potential approach for segmenting the background area in background segmentation. By separating the foreground area using these models, we may concentrate the instance segmentation work only on the pertinent regions, potentially enhancing the precision and accuracy of cavity identification.

3. Two-Round Fine-Tuning with Loosely Similar Datasets: Implementing a two-round fine-tuning strategy has the potential to greatly improve model performance by using loosely similar datasets. This process entails first optimizing the models using a loosely similar, extensive yet low-quality dataset, and then further optimizing them using the exact training set relevant to this projects situation. Utilizing weakly supervised ways to manage the training process can enhance the performance of the network by leveraging extensive datasets, hence improving the robustness and accuracy of the model.

4. Integration of Weakly Supervised Learning Methods: The integration of weakly supervised learning methods can be especially advantageous in managing extensive datasets that have low quality. These techniques can aid in the management of noise and variability in the data, resulting in the creation of more precise and widely applicable models.

5. Expanded Dataset: Acquiring a broader and more comprehensive dataset

that includes various alloys and irradiation settings would enhance the model's generalizability and enhance its resilience.

6. Cross-disciplinary Applications: Expanding the established approaches to other domains of material science and beyond, such as biology or medical imaging, in order to investigate the adaptability and usefulness of the suggested procedures.

References

[1] Ryan Jacobs et al. "Materials swelling revealed through automated semantic segmentation of cavities in electron microscopy images". In: *Scientific Reports* 13 (Mar. 2023).

[2] B.N Singh et al. "Review: Evolution of stacking fault tetrahedra and its role in defect accumulation under cascade damage conditions". In: *Journal of Nuclear Materials* 328.2 (2004), pp. 77–87.

[3] Nasr M. Ghoniem and Yinan Cui. "1.22 - Dislocation Dynamics Simulations of Defects in Irradiated Materials ". In: ed. by Rudy J.M. Konings and Roger E. Stoller. Second Edition. Oxford: Elsevier, 2020, pp. 689–716. ISBN: 978-0-08-102866-7.

[4] Chris M. Anderson et al. "Automated Detection of Helium Bubbles in Irradiated X-750". In: *Ultramicroscopy* 217 (2020), p. 113068.

[5] Wei-Ying Chen et al. "In-situ TEM investigation of void swelling in nickel under irradiation with analysis aided by computer vision". In: *Acta Materialia* 254 (2023), p. 119013.

[6] A.A. Stepashkin and N.Yu. Nikitin. "Statistical analysis, regression, and neural network modeling of the tensile strength of thermoplastic unidirectional carbon fiber-polysulfone composites". In: *Carbon Trends* 15 (2024), p. 100368.

[7] Islam M. Mantawy and Naga Lakshmi Chittitalli Ravuri. "Predicting low-cycle fatigue-induced fracture in reinforcing bars: A CNN-based approach". In: *Structures* 64 (2024), p. 106509.

[8] Graham Roberts et al. "Deep Learning for Semantic Segmentation of Defects in Advanced STEM Images of Steels". In: *Scientific Reports* 9 (Sept. 2019).

[9] Stephen Taller, Luke Scime, and Ty Austin. "A new paradigm in electron microscopy: Automated microstructure analysis utilizing a dynamic segmentation convolutional neutral network". In: *Materials Today Advances* 21 (2024), p. 100468.

[10] Junlin Lin et al. "Transregional spatial correlation revealed by deep learning and implications for material characterisation and reconstruction". In: *Materials Characterization* 178 (2021), p. 111268.

[11] Rajat Sainju et al. "DefectTrack: a deep learning-based multi-object tracking algorithm for quantitative defect analysis of in-situ TEM videos in real-time". In: *Scientific Reports* 12 (Sept. 2022).

[12] Surya Prakash Mishra and M.R. Rahul. "A comparative study and development of a novel deep learning architecture for accelerated identification of microstructure in materials science". In: *Computational Materials Science* 200 (2021), p. 110815.

[13] Kaiming He et al. "Mask R-CNN". In: *CoRR* abs/1703.06870 (2017).

[14] Lun Che et al. "Deep learning in alloy material microstructures: Application and prospects". In: *Materials Today Communications* 37 (2023), p. 107531.

[15] Xinlong Wang et al. *SOLOv2: Dynamic and Fast Instance Segmentation*. 2020.

[16] Wei Wei et al. "Augmenting progress monitoring in soil-foundation construction utilizing SOLOv2-based instance segmentation and visual BIM representation". In: *Automation in Construction* 155 (2023), p. 105048.

[17] Ryan Cohn et al. *Instance Segmentation for Direct Measurements of Satellites in Metal Powders and Automated Microstructural Characterization from Image Data*. 2021.

[18] Qing Yang et al. "Road Scene Instance Segmentation Based on Improved SOLOv2". In: *Electronics* 12.19 (2023).

[19] L. Minh Dang et al. "Lightweight pixel-level semantic segmentation and analysis for sewer defects using deep learning". In: *Construction and Building Materials* 371 (2023), p. 130792.

[20] Kavindu Wijesinghe et al. "Characterization of microscopic deformation of materials using deep learning algorithms". In: *Materials Design* 208 (2021), p. 109926.

[21] Leon Derczynski. “Complementarity, F-score, and NLP Evaluation”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Ed. by Nicoletta Calzolari et al. May 2016.

Acknowledgments