



# Image segmentation and classification for fission track analysis for nuclear forensics using U-net model

Noam Elgad<sup>1,2</sup> · Rami Babayew<sup>1,2</sup> · Mark Last<sup>3</sup> · Aryeh Weiss<sup>4</sup> · Erez Gilad<sup>2</sup> · Galit Katarivas Levy<sup>5</sup> · Itzhak Halevy<sup>1,2</sup>

Received: 20 September 2023 / Accepted: 6 March 2024 / Published online: 3 April 2024  
© The Author(s) 2024

## Abstract

This study introduces a novel methodology for the detection and classification of fission track (FT) clusters in microscope images, employing state-of-the-art deep learning techniques for segmentation and classification (Elgad in nuclear forensics—fission track analysis—star segmentation and classification using deep learning, Ben-Gurion University, 2022). The U-Net model, a fully convolutional network, was used to carry out the segmentation of various star-like patterns in both single-class and multi-class scenarios.

**Keywords** Nuclear forensics · Fission track analysis · Holmeland security · Safeguards investigations · U-Net · Computer vision

## Introduction

Nuclear Forensics (NF) is a research discipline that focuses on the study of radioactive materials. For more than 50 years, international safeguards have been employed to confirm that nuclear materials disclosed by a state to the IAEA are used for peaceful purposes [2]. One of the safeguards' regime main strategies is based on environmental sampling and analysis to discover radioactive fingerprints of undeclared activity. Environmental sampling as a routine for

safeguards inspections was approved in 1995 [3]. Environmental sample analysis provides information about nuclear activity in terms of form, size and surface structure of uranium microparticles of industrial dust in nuclear plants, often using scanning electron microscopy [4].

Particle isotopic analysis in NF has grown quickly in the past two decades due to advances in technologies for particle isotopic composition determination [5–8]. Forensic measurements in this field support action against illicit use of materials, guiding nations to improve their nuclear defense capabilities using scientific methods, evidence, and additional sources, including law enforcement and intelligence agencies [9].

NF uses analytical techniques for measurements of isotopic, chemical, and physical properties of nuclear material, as well as traditional forensic evidence like DNA samples, hair, fingerprints, tool marks, and gunpowder residue, to identify fissile radioactive material and establish connections to individuals, locations, and incidents. Fission Track Analysis (FTA) detects fissile compounds. Radioactive samples from nuclear incident sites are collected for study. After collecting samples on dedicated sampling paper, they are sandwiched between two polycarbonate Lexan resin detector sheets and bombarded with thermal neutron flux to create fission tracks. Traces form star-like formations (fission track clusters) show nuclear fission products' trajectories. Those remnants can reveal interesting details about the measured substance's origin. Taken together with advanced statistical modelling

---

✉ Noam Elgad  
n.elgad@gmail.com

✉ Itzhak Halevy  
halevy.itzhak.dr@gmail.com

<sup>1</sup> Engineering and Physics Departments, Nuclear Research Center Negev, Beer-Sheva, Israel

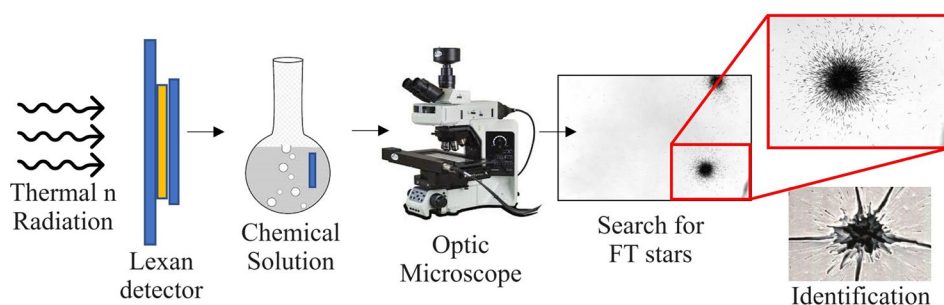
<sup>2</sup> Unit of Nuclear Engineering, Faculty of Engineering Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

<sup>3</sup> Department of Software and Data Engineering, Faculty of Engineering Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

<sup>4</sup> Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel

<sup>5</sup> Department of Biomedical Engineering, Faculty of Engineering Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

**Fig. 1** Schematic flow of process of creating fission track cluster



methods to extract additional information from radiological NF investigations [4], FTA can advance the utility of NF measurements.

Machine Learning (ML) is a branch of AI that enables systems to learn from experience without explicit programming [10–12]. The software learn patterns in a sample set to decide how to handle new information. Recently, machine learning (ML) and its derivatives, particularly deep learning (DL), have evolved considerably due to the exponential development of information capture, storage, and processing capacities. [13]. This approach is compatible with computer vision (CV) application [14, 15], and integrates across various academic and engineering fields [16–19]. In recent years, FT cluster identification has moved from manual procedures on a database of thousands of images to automation, easing researchers' labor and providing more robust and accurate analysis. This study aims to improve field automation and propose an innovative DL-based solution for FTA using segmentation and classification of nuclear fission track shapes in microscopic images as part of our efforts to improve FTA as applied to NF.

## Theory

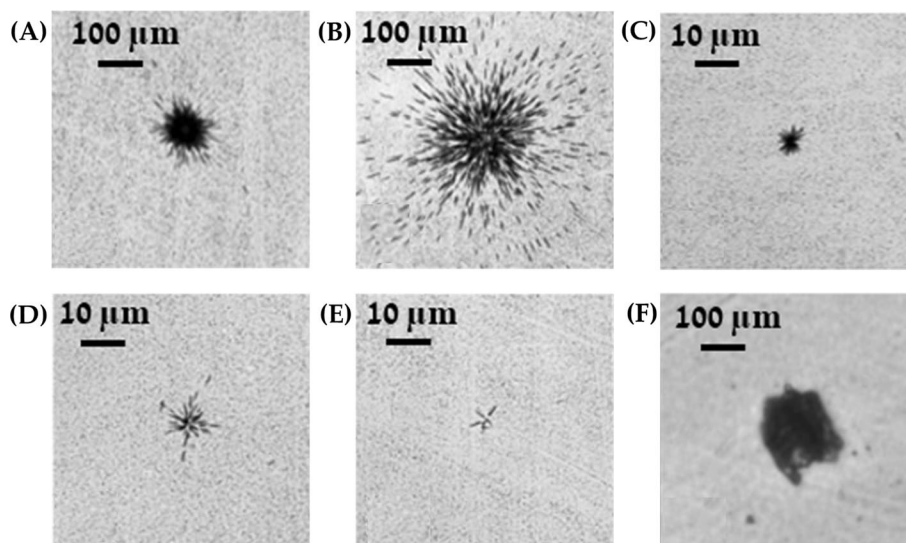
This section provides an overview of the FTA process and presents a broad perspective on the selected technologies used for identification purposes in this study.

### FTA

One advantage of the FTA method is its ability to identify fissile materials from background elements. One particle in 1,000,000 is fissile and significant. Sample collection is carried out using suitable safety and contamination control procedures. Potentially radioactive samples are collected on particle-free paper from items in the prescribed area. These particles are sealed in Lexan foil. To simplify collection, a mini-bulk model was built. This approach determines the average enrichment and concentration of the material,

guiding the micro-bulk transfer quantity. The mean quantity of particles can indicate the existence of artificial chemicals, such as uranium isotopes, providing an initial forensic investigative trail. Mini-bulk sampling involves sampling a reference item and a suspected contaminated item. To dissolve the particles, the pieces are cut into 1 cm<sup>2</sup> squares and soaked in Heptane liquid (C<sub>7</sub>H<sub>14</sub>) in a test tube. The sample preparation method follows these three steps: (1) completion of a mini-bulk preparation for ICP-MS testing, (2) preparation of a 300–500 picogram sample, and (3) attachment of the sample to a 31 × 30 mm Lexan surface to construct a Lexan detector. After sample preparation, a reactor irradiates the detector, which is a capture sandwich of several detectors, with thermal neutrons. This irradiation exposes the detector to 5 × 10<sup>13</sup> n/cm<sup>2</sup> s thermal neutrons for 20 min. The detector undergoes forced fissions due to irradiation. After irradiation, the detector sandwich cools for three days before being returned to the lab. After cooling, markers are securely mounted to the top and lower detectors at three locations. This creates a coordinate system for navigation between the detectors. Then, fission tracks from the material under test are developed and imaged in a lab. The particle's spatial placement determines its identification, and several chemical methods will test it. There are several star configurations that represent the orbits of nuclear fission products [20]. About 700 images are acquired using a transmitted light microscope equipped on a motorized XY scanning stage and automated Z-focus. There is a list of landmarks. The shapes are cut to verify the precision of preparation, casting, screening, and cutting. If a FT cluster is visible in the configuration, the procedure was correct and the substance is real. The FTA method can identify fissile radioactive material via spontaneous or induced fissions. Detector grooves show fission trail signatures and indicate the fission site. Since fission is isentropic, the fission product will be a star with a circular border. FTA offers exact location data on FT location, so that the sample associated with any FT cluster can be excised and sent for ICP-MS investigation. If uranium is present and the enrichment level is abnormal, forensic proof can be obtained. Generally accepted guidelines for

**Fig. 2** Types of stars: **A** black center with a long halo **B** huge without a black center **C** small black center **D** rich **E** poor **F** artifact



enrichment levels are as follows: natural: 0.72%, commercial power reactors: 4–5%, research reactor: 20%, and nuclear weapons: 90% and higher. Figure 1 illustrates a schematic representation of the procedure.

### Identification processes

Until 2017, FT cluster detection was done manually, like "finding a needle in a haystack". Recently, automated algorithms have been developed for FT cluster identification [21, 22]. Image processing software with a suitable user interface was used for identification. This technology overcomes biases associated with manual identification. For example, when people know the source of a forensic sample, they tend to work harder to find FT clusters. In a study [23], a 10% increase in cluster identification (FP) was observed when the examiner understood the relevance of the sample. Deviation between testers with similar training was found to be 15% [23]. AI-based systems may have advantages over previous technology. Autonomous systems' objectivity is a major asset. Two staff members and two days are needed for manual scanning and detection of FT clusters. Modern image processing methods can detect FT clusters from microns to roughly 300  $\mu\text{m}$  in size. Depending on the importance of the sample, repeated testing by the same examiner yielded 90% accuracy. Using different testers reduced accuracy 85%. The false positive rate rose by 5% [23]. Figure 2 shows an example artifact and a variety of stars observed in our lab. The tested material is natural sand containing uranium [24]. The stars are located and the quantity of tracks characterizes the material as fissile. It is important to note that a large natural uranium particle can form a star like a small highly enriched

particle. The shape of a star only shows the amount of fissile material present, not its enrichment.

This study used a U-Net neural network architecture, based on a Fully Convolutional Network (FCN) [25], to segment and classify fission track clusters in microscope images. Neural networks, also known as artificial neural networks, are computational mathematical models inspired by biological neural networks used in machine learning. The wide range of shapes and sizes they can take in an image in addition to the possibility of contaminants being mistaken for FT clusters make textural star shapes difficult to recognize. Noise and blur are common problems with microscope images. Typical image processing models are excessively sensitive to these problems. Clusters have very varied shapes, and occasionally they overlap each other, causing standard segmentation and classification procedures to lose accuracy. Neural network-based semantic segmentation algorithms for image processing have advanced recently, so that these tools are now available to address the shortcomings of classical image processing workflows. Semantic segmentation labels each image pixel (e.g. person, road, sky, ocean, car).

Semantic segmentation is utilized in autonomous cars, industrial inspections, satellite image-based terrain classifications, and medical imaging analysis [26–29]. Currently, FCN algorithms are considered the 'State of The Art' in practical neural network performance [25]. These algorithms have yielded positive results in many segmentation challenges [30, 31]. FCN architecture has only locally connected layers such as convolution, pooling, and un-sampling. Avoiding dense layers reduces parameters, making network training faster. U-Net networks [32] were developed

by adapting the FCN architecture for biomedical image segmentation. This adaptation allows training with fewer images and improves segmentation accuracy. For segmentation and classification purposes, an algorithm based on U-Net network could be a viable option, since U-Net network captures the data's different shapes and textures, enabling accurate predictions and labeling of the stars. U-Net topology has a skip link between the Encoder and Decoder, two layers with ReLU activation functions and normalization, and a Maxpooling layer. Convolution creates image feature maps. Kernels, or quadratic two-dimensional filters, are applied to input data at fixed spatial locations during convolution. Convolution between the Kernel and the input produces the output.

The selection of the U-Net architecture was motivated by its known efficacy and resilience in handling diverse spatial variations in object size, texture, and shape [33]. A model can be assessed using accuracy (precision) and recall (sensitivity). Accuracy measures the proportion of relevant examples among the recovered instances, while recall measures the proportion of relevant instances that were successfully retrieved, where:

$$\text{Recall} = \frac{TP}{TP + FN}; \text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

Contemporary cluster detection methods have 80% precision. However, manual scanning causes inconsistency. Thus, professional quality control is required, and this is time-consuming. Cluster detection and classification is difficult due to their varied forms and sizes. This makes it difficult to identify stars from other non-star phenomena like scratches, dirt, shadows, etc., resulting in many false positives. The study seeks a solution to these issues.

Our current focus in forensic segmentation and identification is demonstrating FTA segmentation capacity. An image processing workflow created using the Fiji distribution of ImageJ [35] was developed as part of an ongoing research collaboration between the International Atomic Energy Agency and the Israel Atomic Energy Commission. This algorithm finds three types of prominent FT clusters [22, 36]. The primary challenge lies in identifying sparse FT clusters. We will use DL methods to improve the existing approach by reclassifying sparse clusters as a discrete category rather than noise. In the beginning of our inquiry, we found dense clusters utilizing only 30 observations. With sparse clusters, adding data improved performance. The ultimate goal is to recognize roughly 7 classes of clusters, including floating, asymmetric, and complex clusters. The idea is to automate FT cluster identification for about ten categories. No methodology currently supports this.

The U-Net technique was chosen for its robust performance on image segmentation problems, with relatively

small datasets. In this study, the dataset had over 100 labeled observations per category.

## Experimental

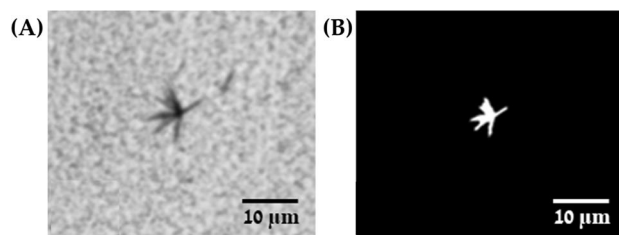
The dataset acquisition process had two stages: Fission stars were first collected using a nuclear fission detector, Lexan. Second, an optical microscope generated 700 images with a 10% overlap. Each image was then divided into  $250 \times 250 \mu\text{m}$  pieces for further processing.

## Methodology

Intensity thresholding together with morphological criteria were used to divide an image into foreground and background components to retrieve stars. High-intensity areas in the binary image were classified as potential FT clusters. Measurements were based on star detection thresholds. Detection thresholds were set considering characteristics such as cluster size, criterion to distinguish clusters from scratches or dirt, ratio of pixels in the dark area to the bounding box (fill ratio), Circularity and Eccentricity.

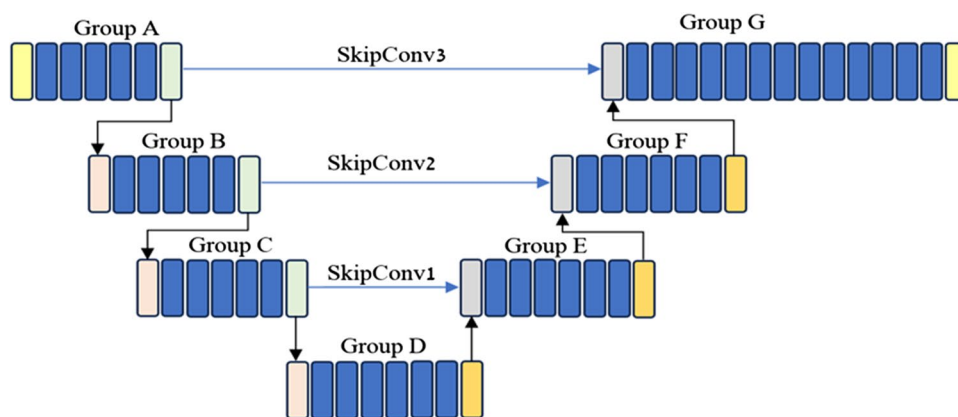
## Data preparation

Cluster categories were defined to standardize the nomenclature for description of objects of interest. The categorization parameters were set by these criteria, in order to create a complete cluster catalog. The study had four classification criteria: size (large/small), richness (poor/rich), center (black center/no black center), and halo length. Asymmetric, complex, and hovering stars were also identified. The star categorization symbol is  $w, x, y,$  and  $z$ . The letters denote binary values of 0 or 1 for each condition. The magnitude, richness, core black region, and halo extent are represented by  $w, x, y,$  and  $z$ . Tables 5, 6, 7 present the binary notation for cluster classification. Typical images and color setting for cluster classes and types are shown. All three tables appear in Appendix.



**Fig. 3** **A** Original image containing a poor rose with dirty background. **B** The image after manual editing in preparation for training used as the tagged image

**Fig. 4** Schematic diagram of the network layers in a U structure



## Machine learning

Machine learning algorithms use mathematical models to learn and predict from input data. The model input data is split into different datasets. Model development often uses three datasets: training, validation, and test. The model is trained using the training set to optimize parameters using known cases. For each input vector in the training set, the model generates an output and compares it to the ground truth. Model parameters are modified based on comparison results and learning technique. A validation set dataset uses the fitted model to estimate performance on data not in the training set. The validation set impartially evaluates a model's performance during training, and is used to detect overfitting. A naive test set is then used to test the model.

Cross-validation was used for multiple model testing and training iterations. This method optimizes data use across arrays. In this study fivefold Cross Validation technique was employed. This method partitioned the data into five subsets, with 60% for training, 20% for validation, and 20% for testing. A full data replacement process was designed to assure the independence of the validation and test sets. This was repeated five times, changing the training data composition.

Data must be carefully labeled at the pixel level to provide the ground truth for DL training. The raw image has a light background and dark objects. The fission track clusters need to be labeled. The images include artifacts such as shadows, stones, dirt, scrapes, and other factors. To train, each pixel must be classified correctly. The FT clusters were labeled with white pixels (255), and all background and artifacts were black (zero). Figure 3 shows a source image on the left and its corresponding ground truth. All dataset images undergo this treatment. Source image resolution was reduced from 20 to 4 MB during training data preparation, due to information storage and training run session limits.

## Architecture

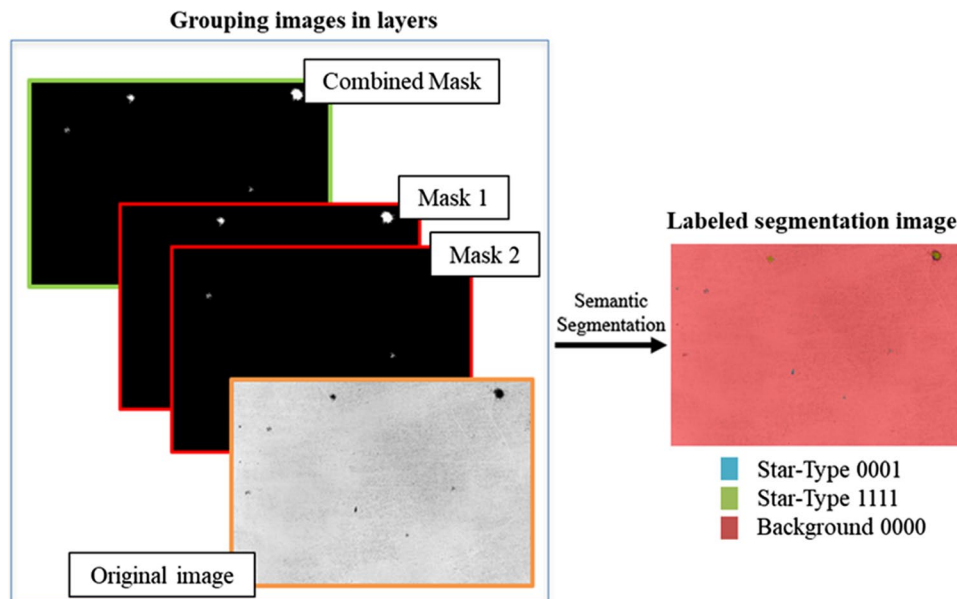
The tasks of data preparation, running training runs, completing segmentation, and other operations were executed using MATLAB software. As previously stated, this study's model uses U-Net neural network architecture. The name "U-Net" describes the network's architecture. The network has a connected path and a broad path, forming a U shape. Convolutional networks use iterative convolutions, ReLU activations, and pooling operations in the connection pathway.

Contraction reduces spatial information and increases feature information. A series of convolutions with high-resolution features from the shrinking path merges feature and spatial information in the expansive path. This study's U-Net architecture is shown in Fig. 4. Table 1 details our model's network layers after this depiction. The model has three jump points and uses convolution.

One of the limitations associated with the utilization of deep learning techniques is the requirement for a substantial volume of data. For deep learning networks to achieve optimal performance, a substantial volume of data is required, typically in the order of millions of records. Furthermore, it is observed that the quality of training outcomes improves as the quantity of available information increases. As previously said, the U-Net model has an inherent advantage in its ability to effectively utilize small amounts of data. However, it is important to note that a substantial amount of data is still required. The process of generating data, particularly when it involves tagging information, presents a formidable undertaking that necessitates significant allocation of resources. The generation of artificial information, often known as Augmentation, has the potential to substantially enlarge the available data.

**Table 1** Network layers arranged by groups

Layer Group	A	B	C	D	E	F	G
Layers	Input image	Maxpool1	Maxpool2	Maxpool3	Add1	Add2	Add3
	Conv1	Conv3	Conv5	Conv7	Conv9	Conv11	Conv13
	Relu1	BN3	BN5	BN7	BN9	BN11	BN13
	Conv2	Relu3	Relu5	Relu7	Relu9	Relu11	Relu13
	BN2	Conv4	Conv6	Conv8	Conv10	Conv12	Conv14
	Relu2	BN4	BN6	BN8	BN10	BN12	BN14
		Relu4	Relu6	Relu8	Relu10	Deconv3	Relu14
				Deconv1	Deconv2		Conv15
							BN15
							Relu15
							Softmax
							Class Layer

**Fig. 5** An example of creating a tagged image using semantic segmentation

## Image labeling

As part of this work, we created a new classification system. A custom MATLAB application was used for image labeling. Sequential steps were taken: We start with source images for analysis. We then created labeling categories. Within these categories, we established a reference label. This label can be used on clusters of various categories and non-cluster objects like dirt or scratches. The label highlights star features. Finally, we compiled and exported all of the data to a file for analysis.<sup>1</sup>

<sup>1</sup> The thorough method development is not covered in this paper and is subject to separate publication.

## Experimental—segmentation processes

Once the data are ready, training starts. This study's system trained for single- and multi-class classification. Effective training produces a model that can segment images and detect stars. The semantic segmentation procedure and evaluation are covered in this section.

### Semantic segmentation

Semantic segmentation labels each visual pixel as animal, human, road, sky, car, etc. The aim is to recognize certain features in an image, not categorize it. This approach

**Table 2** Confusion matrix (in Parenthesis: our case)

Classification	Actual Existence		
		Exist (ST)	Not Exist (Dirt/ Stone etc.)
	Positive (ST)	True Positive	False Positive
Negative (Dirt/ Stone etc.)	False Negative	True Negative	

In Parenthesis: confusion matrix related to our case

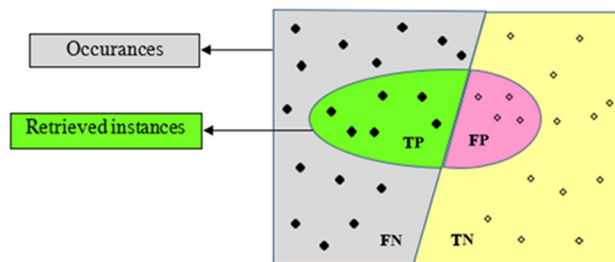
distinguishes visual entities. In our work, segmentation requires labeling each pixel for different types of clusters, non-cluster objects, and the background. A MATLAB function named `semanticseg` was used. For segmentation, the function takes the trained neural network and an image. It then semantically segments the input image and assigns classification scores to each classified label. An array of scores is extracted and grouped by pixel or voxel in the supplied image. After this phase, the layers are blended using Label Overlay to create a composite image with several labels. The Image Scaled Color technique generates a categorization results map. Figure 5 shows how semantic segmentation creates a multi-layer tagged image from a source image. This generation is done through training with an integrated mask image (Ground Truth). Each mask in the Ground Truth image represents a cluster type.

### Segmentation evaluation

We created a model that maximizes ROC curve area. We will explain the ROC curve, its ideas, and how to generate it in this section. We will also examine how to test our model against this curve.

The curve structure: the ROC curve shows binary classifier performance at different decision levels. A curve is created by plotting the true positive rate (TPR) and false positive rate (FPR) across acceptance levels. The sensitivity, or true positive rate, is a machine learning coverage measure. The false positive rate is 1 minus the specificity, or leakage or significance threshold. The receiver operating characteristic (ROC) curve shows how significance level affects sensitivity. Binary classification divides the results into positive and negative groups. True positive (TP), false positive (FP), true negative (TN), and false negative (FN) are the categorization outcomes. Table 2 uses the Confusion Matrix to display these values. Our matrix is formed after training using segmentation (during the activation of the trained model on the test set) and can be used to calculate ROC values.

Visually, as depicted in Fig. 6, the diagonal line serves as a demarcation between positive things on the left and negative items on the right. The classifier is represented by an



**Fig. 6** True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN)

egg-shaped boundary, encompassing positive objects inside its confines and excluding negative items outside of it.

When considering errors in fission track analysis, it is important not to miss actual FT clusters (which would be a false negative), but marking non-stars as clusters (false positives) is a less serious error. This is due the fact that the meaning of FN is a radioactive material that has been missed, as opposed to FP which means a non-radioactive item who may be marked falsely as radioactive.

### Automation

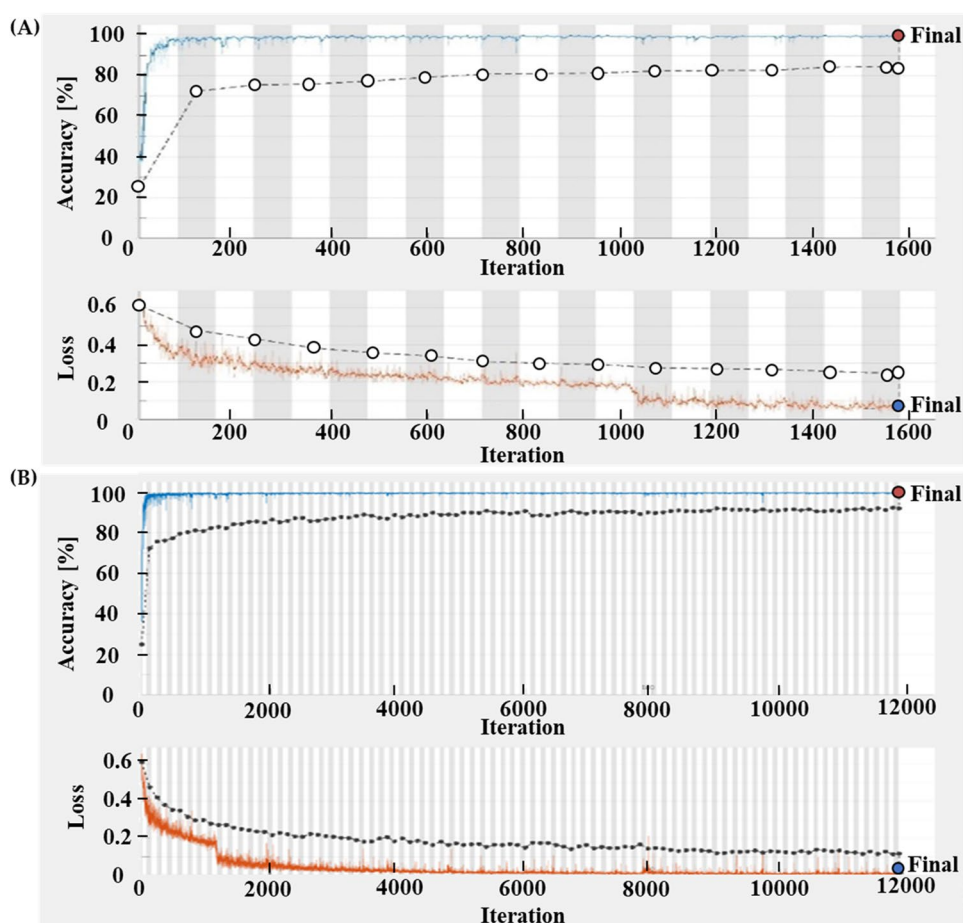
The above steps for labeling training, and application of a DL model were developed into a set of computer codes that can be used to automate the processing of large numbers of images and FT clusters.<sup>2</sup>

### Results and discussion

This section provides an overview of the training and segmentation workflows for both single class and multi class classification.

<sup>2</sup> The in-depth automation development is not detailed in this paper and is planned to be published in a separate publication.

**Fig. 7** Training results of single poor star model in 20 (A) and 150 (B) epochs



**Table 3** Training results repetitions: 20, 50, 100, 150 epochs

Epochs	Time elapsed (hh:mm:ss)	Mini-batch accuracy	Validation accuracy (%)	Mini-batch loss	Validation loss
20	01:34:02	NA	84.55	NA	0.2470
50	02:15:04	99.79%	90.12	0.0174	0.1627
100	03:13:33	99.83%	90.61	0.0175	0.1238
150	07:52:39	99.76%	<b>92.04</b>	0.0048	<b>0.1168</b>

Bold values indicate the best results received for the two most significant columns (best validation accuracy result was 92.04% and best validation loss results was 0.1168)

### Single class training—poor & rich stars

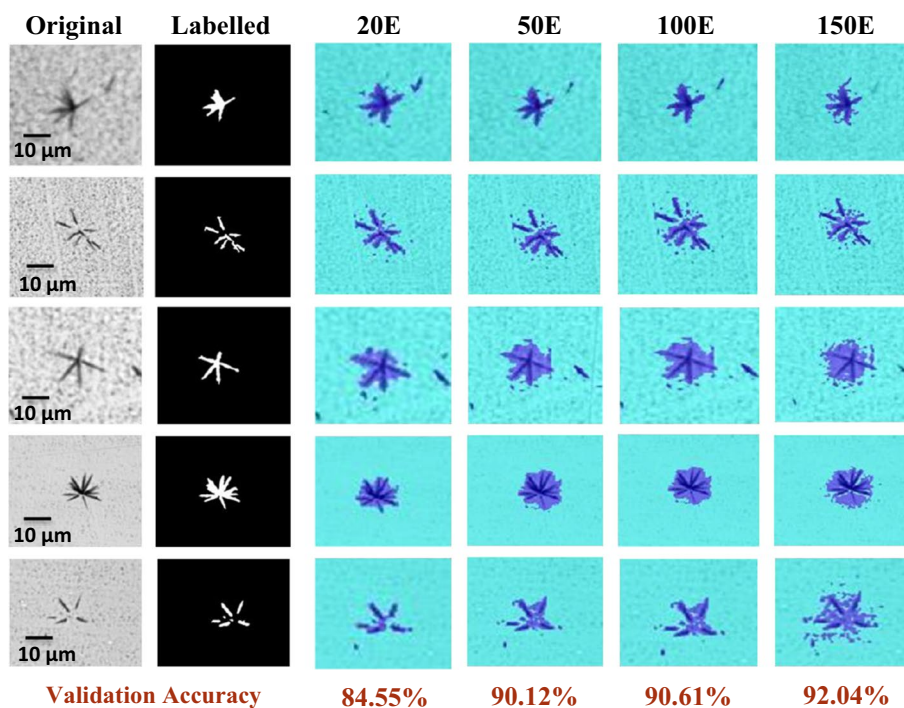
Figure 7 shows two training outcomes for a low-quality star detection model. The primary training consisted of either 20 epochs (94 min), 50 epochs (135 min), 100 epochs (193 min), or 150 epochs (492 min). Accuracy increases with repetitions and run duration. The figure shows two runs: 20 epochs and 1580 iterations (top image) and 150 epochs and 11,850 iterations (bottom image). Both sets use 120 iterations for validation. The accuracy chart at the top of each image shows the Mini-Batch Accuracy graph as a blue line and the Validation Accuracy as black dots. The

Mini-Batch Loss (brown) and Validation Loss (black dots) are shown in the lower part of the figure.

Table 3 displays the outcomes of the four training sessions.

In comparison to prior investigations [23], it has been observed that by carrying out training for a duration of 20 epochs, it is feasible to get similar outcomes of 84.55% accuracy (in comparison to about 85%) as reported by several human evaluators. The application of 50 epochs of training enhances the accuracy level, enabling comparable results to a single tester (90%). Subsequently, employing 150 epochs of training further enhances the accuracy level, resulting in

**Fig. 8** Results of stars segmentation after training with 20, 50, 100 and 150 epochs



an accuracy rate of 92.04%. This achievement surpasses the highest level attained by other methods thus far by approximately 2% (90%). Figure 8 displays a collection of segmented FT clusters.

Some examples of artifact objects are shown in Fig. 9. The blue parts on the turquoise background symbolize segmentation with different epochs (20, 150). Five images were segmented after two classifiers were trained. The first three images show stones and dirt, the fourth an unclassified star object, and the fifth a scratch. The objective is for the blue marking (FP) to be minimal in all of these objects. As the number of epochs increase, blue marking intensity decreases, meaning that FP pixels are reduced.

Table 4 presents a collection of focused segmentation images, illustrating a comparative analysis of outcomes achieved for objects classified as FT clusters and those classified as artifacts. For FT clusters, a higher degree of blue marking indicates a. Conversely, for other objects, the objective is to minimize the blue marking and achieve a greater resemblance to the background hue.

Figure 10 shows an image that has undergone segmentation, with the resulting segmentation image displayed on the left side, and a corresponding classification map on the right side. The visual attention is directed towards a star, which is distinctly highlighted in green, as well as an item that is not classified as a star, which is noted in red. Table 4 presents a shared view of these two objects.

This work produced two ROC curves using segmented images from a single-classifier model of a black-centered rich cluster, as shown in Fig. 11. These curves also have

AUC values calculated. The ROC curve shows how a binary classifier performs based on its decision threshold. At the middle, the graph shows the TPR-FPR relationship across acceptance criteria.

The AUC calculation shows the graph's area when a normal result is 0.5 to 1.  $AUC = 0.5$  implies a random prediction. Thus, the model does not improve star categorization. A model with a higher AUC value identifies and classifies better. Any outcome above 0.5 is better than random selection for maximizing value. The plots show a reasonable  $AUC = 0.68$  and a superior one at  $AUC = 0.84$  in our model.

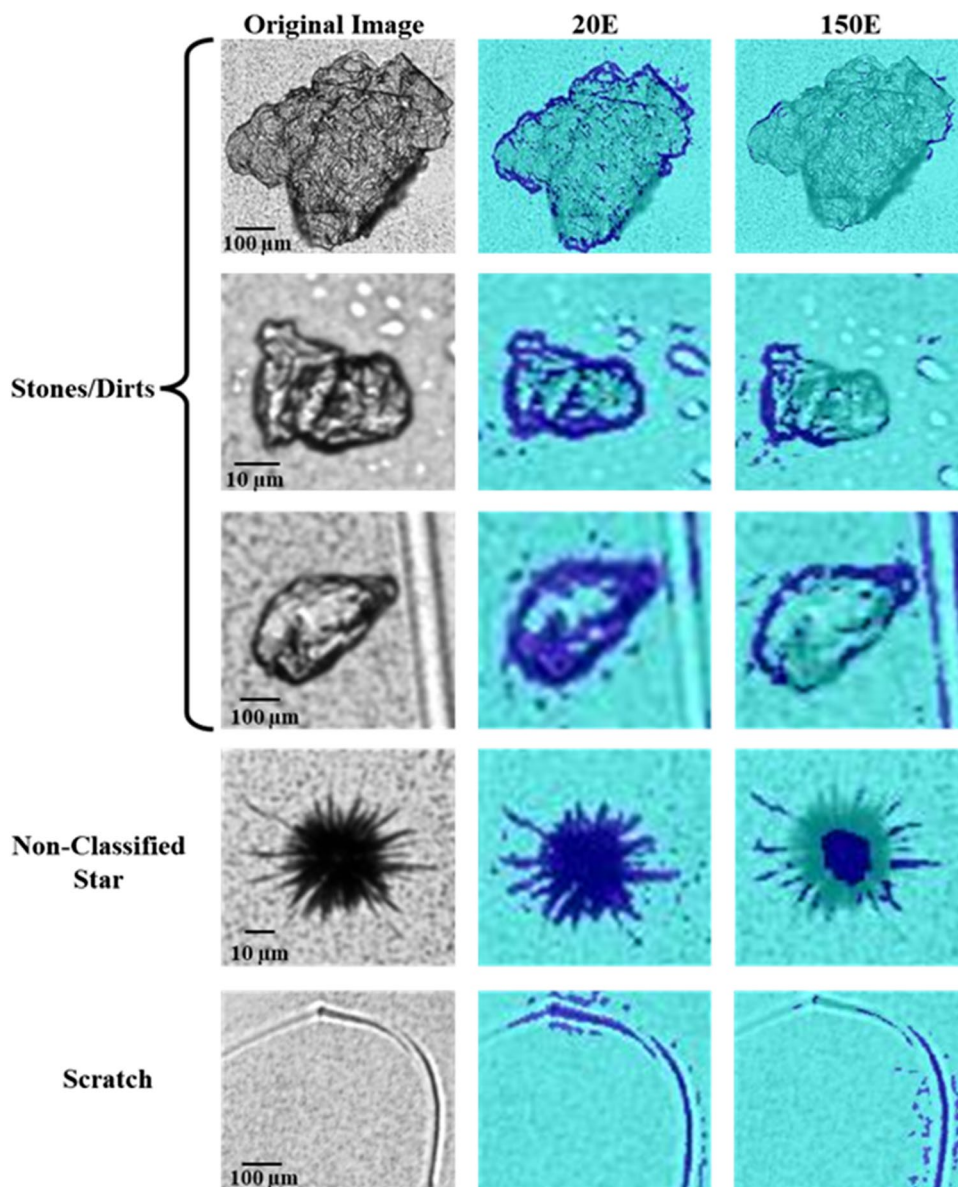
Table 8 in the appendix summarizes one example of results of 4 images and shows pixel intensity, TP, FN, FP, recall and precision of each image. The model has a fair TP and FP and is biased toward zero FN in the specified images as these images went through an adaptive threshold algorithm which was developed for background cleaning.<sup>3</sup>

### Multi class training—poor & rich FT clusters

After training and segmenting rich and poor single cluster images, we trained a multi-class model for both, so that the model identified poor (type 0001) and rich (type 1111) clusters. The training had 20 epochs, 4880 iterations, and 250 validation iterations. Training results: Mini-Batch Accuracy: 96.43%, Validation Accuracy: 86.30%, Losses: 0.1461,

<sup>3</sup> This development is not detailed in this paper as it issued to be published separately.

**Fig. 9** Non-star objects segmentation results after training with 20 and 150 epochs



1.4974. Training outcomes are good for Validation Accuracy. Loss performance is good until epoch = 12. However, the results oscillate after this. Several factors may explain this. One possibility is that pre-processing techniques, such as zero-meaning and normalization, were applied to the training set but not to the validation set, or vice versa. Another explanation could be that the observed discrepancy is due to overfitting, where the model has learned to perform well on the training data but fails to generalize effectively to the validation data. Due to the unchanged code execution mechanism, the first choice is less relevant. The second possibility is more likely, so models with more data were

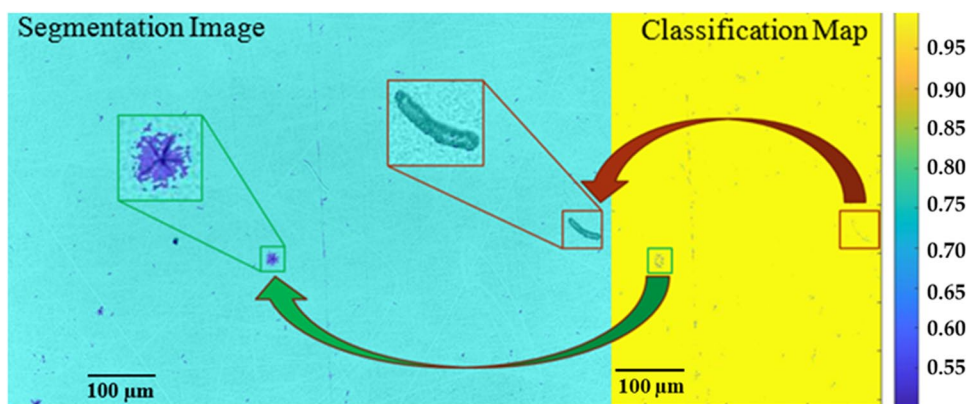
created. The dataset was enhanced via augmentation, and nodes were randomly excluded during training using drop-out. In addition, noise reduction and Cross Validation were used to improve data quality. This algorithm improved the accuracy rate to 86.30%, compared to two persons' 85%. Compared to a manual tester's 90% recognition rate, the model's accuracy is still lower.

Figure 12 shows segmentation with classifiers: ST\_0001, ST\_1111, and background. The left half of the illustration shows images of ST\_0001 stars with blue markings on yellow backgrounds. The right side shows ST\_1111 type stars with green markings on a yellow background. The file names

**Table 4** Comparison of segmentation between stars and non-star objects

Sample ID	Stars (The bluer the merrier)			Non-Stars (The bluer the gloomier)	
	Original	Ground Truth	Segmented	Original	Segmented
tile_x004_y008					
tile_x005_y009					
tile_x004_y010					
tile_x004_y021					

**Fig. 10** Comparison between star and non-star object—a segmentation image and a classification map



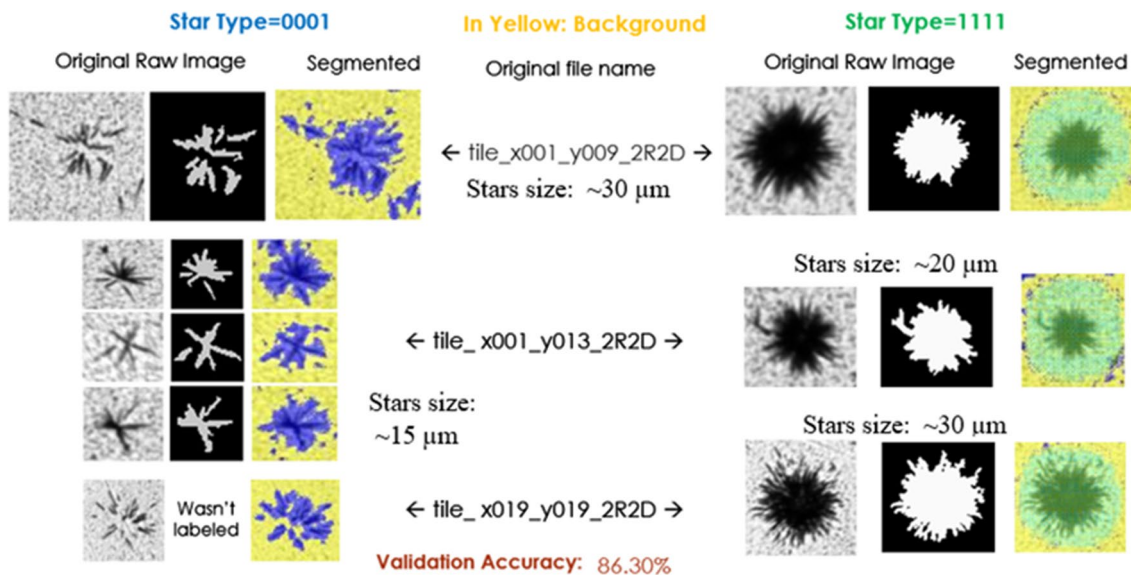
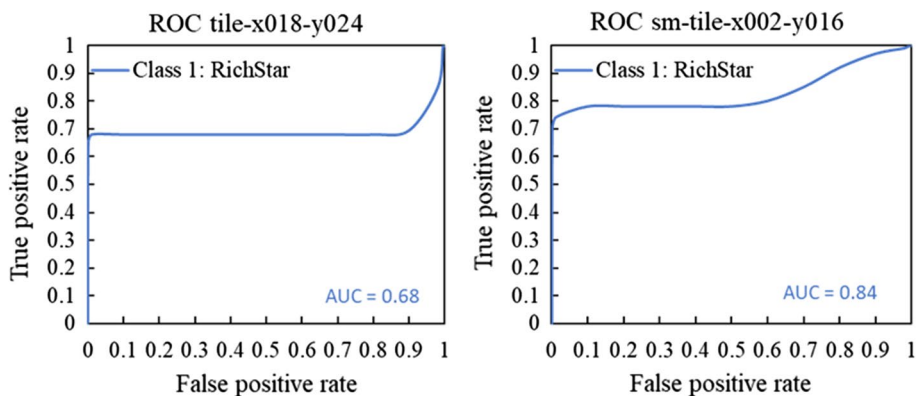
are centrally placed, indicating that the left and right images came from a single file.

Two ROC charts show the performance of several classifiers, including AUC values, in Fig. 13. Each chart shows multiple classifier graphs. Source and labeled images were used to generate segmentation images and charts. The chart below shows the results of two star categories in a trained model image. These categories include improved results

with AUC values of 0.99 and 0.96. An alternate model is used to explore type 0001 and type 1111 stars. Type 1111 stars have an AUC value of 0.66, as indicated in the chart. It shows less satisfactory results for type 0001 stars, with an AUC of 0.24. The value of 1111 is appropriate given its attributes. Its magnitude is less than 0.7. In contrast, it exceeds 0.5, the chance of a random guess. Trial 0001's results are below 0.5, indicating a technical problem.

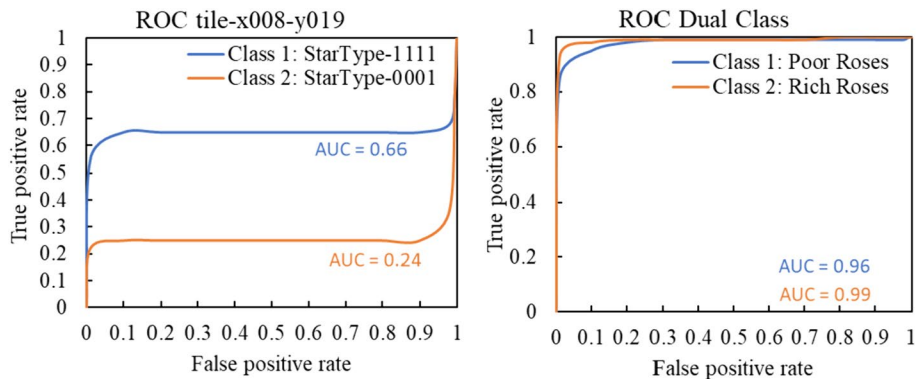
We found that the threshold vector from the MATLAB perfcurv function is cut off at 0.33 instead of decreasing

**Fig. 11** ROC curves produced after completing model training for a single classifier

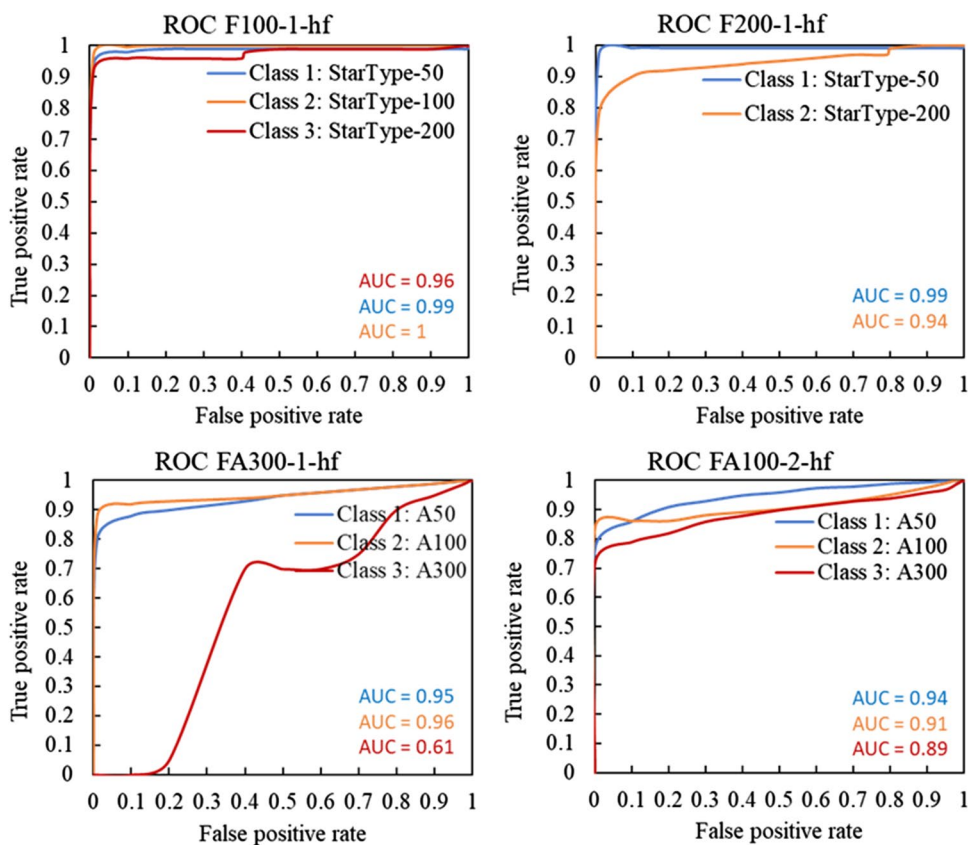


**Fig. 12** Results of segmentation and identification of two types of stars after training

**Fig. 13** ROC charts obtained after training different models with two classifiers



**Fig. 14** ROC charts obtained for segmentation images from three-classifier models



to 0, which appears to be a software fault. It is necessary to carry out new training with segmentation and ROC calculations after training and study the results to observe the root cause of these false results.

### Synthesized FT clusters

After performing multi class training and segmenting, we trained a model with FT clusters that were generated using a specialized software tool designed for the purpose of creating synthetic FT clusters [34]. The process created 50 clusters with 50 leaves, 50 clusters with 100 leaves, and 15 clusters with 300 leaves per cluster. The results of synthesized cluster training model utilizing a retrain procedure consists of 40 epochs in the first stage, 10 in the second, and 40 in the third: Mini-Batch Accuracy of 74.04%, Validation Accuracy of 73.65%, Mini-Batch Loss of 0.1893, and Validation Loss of 1.9106.

Figure 14 shows ROC charts and AUC computations of segmentation images from multiple models. The top-charting model was trained on synthesized StarType<sub>50</sub>,

StarType<sub>100</sub>, and StarType<sub>200</sub> clusters. StarType<sub>50</sub> and StarType<sub>200</sub> performance is shown in the chart. AUC scores were 0.99 for the former and 0.94 for the latter. Another illustration of the same model with all three classes is on the left. A model trained with A50, A100, A300 stars is shown in the bottom plots. The right graph provides good AUC scores for the 3 classifiers: 0.94, 0.91, and 0.89. The left graphic provides AUC for the three classifications: 0.95, 0.96, and 0.61. All results are good, except for the A300 classification, which is only a little better than random guessing (better than 0.5).

### Conclusions

This paper has introduced a novel methodology for detecting FT clusters in microscope images using advanced deep learning segmentation and classification. A U-Net FCN model was developed to accurately segment star-like patterns in single-class and multi-class scenarios. The methodologies described in this study have been effectively

employed as a novel analytical technique for the identification of fission track clusters.

A model was created for clusters with a diameter of less than 60 microns (200 pixels), less than 10 leaves, and no black center (type 0001). The model had a 0.84 ROC area. A computational model was also created to evaluate type 1111 stars of higher magnitude and brightness. This model had an outstanding ROC curve area of 0.90. Early models were also designed to detect cluster shape and size variations simultaneously. It was found that as epochs increase, model accuracy increases. After 20 epochs, the ML model attained ~85% accuracy, comparable to human manual identification. After 50 epochs, the model attained 90% accuracy, surpassing manual identification. After 150 epochs, it reached 92%.

The research included creating a new FT cluster database, classifying cluster types, characterizing the model, designing an architecture, and optimizing. The research also established segmentation measures, optimized the number of epochs and validation frequency, and investigated background noise filtering thresholds. In addition, detection capabilities and adaptive threshold definitions were improved and an Image Labeler tool was created to generate labeling information semi-automatically, followed by Auto Labeling automation. However, the labeling tools are not covered in this paper. A software tool to create synthetic FT clusters was created in the project's latter phases, and used to train a DL model. The initial outcome of the Retrain approach in this collaboration yielded a classification accuracy of 73.65%. Three different types of clusters were included in the images which were used to train a DL model.

We are currently working to extend this work. First improvement in classical methods for processing the ground truth images to improve their SNR will assist in creating larger numbers of labeled images for training, validation and testing. Similarly, improvement in the FT cluster simulator will allow the creation of large simulated datasets for training. Synthetic data usage for AI projects is largely investigated in the recent years.<sup>4</sup>

Other DL architectures can be tested, with different optimizers, etc., in order to optimize the resulting DL network that is to be deployed.

It is also feasible to expand from a segmentation model to a detection model by adding bounding boxes to the ground truth. It is possible to extend the investigation to identify the precise coordinates of the detected clusters, as well as investigate cluster class frequencies and cluster densities per unit area. A correlation test can be used to compare laboratory data with MS-ICP particle observations of particle size, isotopic composition, and cluster shape. Another option is to apply this method to a parallel study of Single Track detection in a high-flux gamma detector based on uranium foil that converts the gamma flux into fissions ( $fy$ ).

## Appendix

See Tables 5, 6, 7, 8.

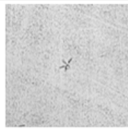
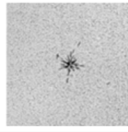
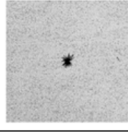
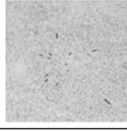
**Table 5** Binary notation for star classification criteria

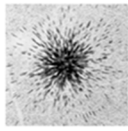
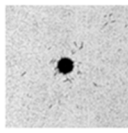
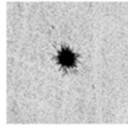
Attr	Large	Small	Rich	Poor	With black center	No black center	Long halo	Short halo
Mark	1	0	1	0	1	0	1	0
Star Size	> 200px i.e > 60 $\mu\text{m}$	$\leq$ 200px i.e $\leq$ 60 $\mu\text{m}$	> 15 leaves	$\leq$ 15 leaves	> 70px i.e > 20 $\mu\text{m}$	$\leq$ 70px i.e $\leq$ 20 $\mu\text{m}$	> 30px i.e > 10 $\mu\text{m}$	$\leq$ 30px i.e $\leq$ 10 $\mu\text{m}$

px pixels

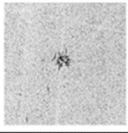
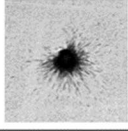
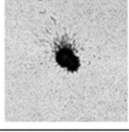
<sup>4</sup> In the context of synthetic data, since this is an area that is particularly relevant for AI systems, for which there is an ever-increasing need for data, there is a rapid wild race in the academic and commercial world in which prominent universities [37], large and research-intensive companies such as Google [38] and young and fast start-up companies [39, 40]. Gartner predicts that by the year 2024 synthetic information will make up 60% of the data used for the development of AI and analytical projects [41, 42].

**Table 6** Typical images and color setting for labeling star types

Star Description	Typical Shape	Code	Size
background			
non star		0000	
very poor star < 10 leaves small without black center		0001	< 200 pixels
non poor star without black center		0101	< 300 pixels
small with black center		0111	< 200 pixels
huge but poor star without black center		1001	> 600 pixels

Star Description	Typical Shape	Code	Size
huge and rich star with black center		1101	> 600 pixels
big star, with black center and short halo		1110	> 300 pixels
big star, with black center and long halo		1111	> 300 pixels

**Table 7** Unique star types

Star Name	Description	Typical Shape
Hovered star	A star composed of fission tracks in an undefined and ununiform direction	
Asymmetric star	A star composed of fission tracks in a defined asymmetrical direction	
Complex stars	Two stars covering each other	

**Table 8** An example of a results table with Recall and Precision calculations for 4 images in a single Rich Star classifier model

Image name	Class	Pixel intensity	TP	FN	FP	Recall	Precision
tile_x001_y008	RichRose	255	2097	0	53	1.000	0.975
tile_x001_y009	RichRose	255	2079	0	72	1.000	0.967
tile_x001_y008	RichRose	255	2059	0	68	1.000	0.968
tile_x001_y009	RichRose	255	2007	0	232	1.000	0.896

**Acknowledgements** We would like to express our appreciation to Eli Boblil to the success of this project. We wish to thank NRCN organization for funding the project. All data for this project is available and can be reached by email request: n.elgad@gmail.com.

**Funding** Open access funding provided by Ben-Gurion University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Elgad N (2022) Nuclear forensics—fission track analysis—star segmentation and classification using deep learning. In: M.Sc. Thesis. Ben Gurion University
- Donohue DL (1998) Strengthening IAEA safeguards through environmental sampling and analysis. *J Alloy Compd* 271–273:11–18
- Bush W et al. (2001) IAEA experience with environmental sampling at gas centrifuge enrichment plants in the European Union. In: Proceedings of the Symposium on International Safeguards: Verification and Nuclear Material Security
- Stebelkov V (2005) Informativeness of morphology analysis of uranium microparticles from industrial dust at nuclear plants. In: GLOBAL 2005: Proceedings of the international conference on nuclear energy systems for future generation and global sustainability. Japan: Atomic Energy Society of Japan
- Lee C-G et al (2006) Improved method of fission track sample preparation for detecting particles containing fissile materials in safeguards environmental samples. *Jpn J Appl Phys* 45(3L):L294
- Inglis JD et al (2023) An introduction to the significance of sample size in particle analyses for nuclear forensics and radiological investigations. *J Radioanal Nucl Chem* 332(8):3371–3381
- Salbu B, Lind OC (2020) Analytical techniques for characterizing radioactive particles deposited in the environment. *J Environ Radioact* 211:106078
- Park R et al (2022) Quantitative and isotopic analysis of single micrometer-sized uranium particles using multiple mass spectrometric techniques. *J Radioanal Nucl Chem* 332(7):2833–2840
- Nuclear Forensics (Retrieved on July 14, 2020) IAEA org
- Michie D et al (1994) Machine learning, neural and statistical classification. Technometrics. Ellis Horwood, New York
- Mitchel T (1997) Machine learning. McGraw-Hill, Boston
- Courville A, Bengio Y, Goodfellow I (2016) Deep learning. MIT Press, Cambridge
- Amodei D, Hernandez D (2018) AI and Compute. OpenAI
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
- Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. *Comput Intell Neurosci*. <https://doi.org/10.1155/2018/7068349>
- (Ret. 2020–07–15) Google AI algorithm masters ancient game of Go. *Nature News & Comment*
- Zhang B, Liu S, Shin YC (2019) In-Process monitoring of porosity during laser additive manufacturing process. *Addit Manuf* 28:497–505
- Rosanovsky M, Oren G, Ifergane S, Beeri O (2019) MLography: An Automated Quantitative Metallography Model for Impurities Anomaly Detection using Novel Data Mining and Deep Learning Approach. [arXiv:2003.04226v1](https://arxiv.org/abs/2003.04226v1) [eess.SP]
- Shen D, Wu G, Suk HI (2016) Deep learning in medical image analysis. *Ann Rev Biomed Eng* 19:221–248
- Stetzer O, Betti M, van Geel J, Erdmann N, Kratz JV, Schenkel R, Trautmann N (2004) Determination of the <sup>235</sup>U content in uranium oxide particles by fission track analysis. *Nucl Inst Methods Phys Res A* 525:582–592
- Halevy I, Admon U, Chinea-Cano E, Weiss AM, Dziga N, Boblil E, Dagan M, Orion I, Radus R (2018) Advances in fission-track detection and analysis for nuclear forensics and safeguards investigations. *Progress Nuclear Sci Technol* 5:175–178. <https://doi.org/10.15669/pnst.5.175>
- Weiss AM, Halevy I, Dziga N, Chinea-Cano E, Admon U (2017) Fission track detection using automated microscopy. *J Nuclear Eng Radia Sci*. <https://doi.org/10.1115/1.4036434>
- Halevy I (2022) Private communication
- Madbouly A, Nassaf MH, Diab AM, El-Mongy SA (2009) A Comparative analysis of uranium ore using laser fluorimetric and gamma spectrometry techniques. *J Nuclear Radia Phys* 4(2):74–81
- Shelhamer E, Long J, Darrel T (2014) Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(4):640–651
- (Retrieved on May 13, 2022) Self-Driving Car in Carla. [roboticsclub.mnrit.ac.in](https://roboticsclub.mnrit.ac.in)
- Nakazawa T, Kulkarni DV (2019) anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder-decoder neural network architectures in semiconductor manufacturing. *IEEE Trans Semicon Manufact* 32(2):250–256
- Liu ZY-C, Chamberlin AJ, Tallam K, Jones IJ, Lamore LL, Bauer J, Bresciani M, Wolfe CM, Casagrandi R, Mari L, Gatto M, Diongue AK, Toure L, Rohr JR, Riveau G, Jouanard N, Wood CL, Sokolow SH, Mandle L, Daily G, Lambin EF, De Leo GA (2022) Deep learning segmentation of satellite imagery identifies aquatic vegetation associated with snail intermediate hosts of schistosomiasis in senegal, Africa. *Remote Sens* 14(6):1345
- Valizadeh A, Shariatee M (2021) The progress of medical image semantic segmentation methods for application in COVID-19 detection. *Comput Intell Neurosci*. <https://doi.org/10.1155/2021/7265644>
- Yang L, Zhang Y, Guldner IH, Zhang S, Chen DZ (2016) 3d segmentation of glial cells using fully convolutional networks and k-terminal cut. In: MICCAI, pp 658–666

31. Aydin AS et al (2017) CNN Based yeast cell segmentation in mutli-modal fluorescent microscopy data. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) pp753–759
32. Ronneberg O, Fischer P, Brox T (2015) U-NET: Convolutional Networks for Biomedical Image Segmentation. [arXiv:1505.04597](https://arxiv.org/abs/1505.04597)
33. Zhang M, Li X, Xu M, Li O (2017) Image Segmentation and Classification for Sickle Cell Disease Using Deformable U-NET. [arXiv:1710.08149v3](https://arxiv.org/abs/1710.08149v3)
34. Halevy I, Cerna-Maskova S, Babayew R, Elgad N, Lorincik J, MicheliD (2022) Nuclear forensics—fission track analysis—simulation and image analysis
35. Rasband WS (1997–2018) Image J US National Institutes of Health, Bethesda, Maryland, USA
36. Admon U, China-Cano E, Dzigal N, Halevy I, Boblil E, Elkayam T, Weiss A and Vogt S (2014) Advancements in Particle Analysis Procedures and Their Application in Characterizing Reference Materials for Safeguards. In: Symposium on International Safeguards: Linking Strategy, Implementation and People. Vienna, Austria, pp 20–24
37. Jahanian A, Puig X, Tian Y, Isola P (2021) Generative models as data source for multiview representation learning. In: MIT Computer Science and Artificial Intelligence Laboratory
38. Bileschi M (2022) Using deep leaning to annotate the protein universe. Google ai blog—the latest from google research. Nature Biotechn 40(6):932–937
39. Wiggers K (2022) Synthesis AI raises \$17M to generate synthetic data for computer vision. <https://techcrunch.com/2022/04/28/synthesis-ai-raises-17m-to-generate-synthetic-data-for-computer-vision/>
40. (2021) 7 Synthetic Data Companies for Training AI Models. nanalyze. <https://www.nanalyze.com/2021/08/synthetic-data-companies-ai-models/>
41. White A (2021) By 2024, 60% of the data used for the development of AI and analytics projects will be synthetically generated. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2023-08-01-gartner-identifies-top-trends-shaping-future-of-data-science-and-machine-learning>
42. (2021) Maverick\* Research: Forget About Your Real Data — Synthetic Data Is the Future of AI. Gartner Research. <https://www.gartner.com/en/documents/4002912>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.