Supplementary materials - Accurate quantification of dislocation loops in complex functional alloys enabled by deep learning image analysis

---

S1 Model description

To perform instance segmentation, the *Detectron2* library, developed by the Facebook AI Research team, was used. Its implementation uses python and the deep learning backend is *PyTorch*. *Detectron2* is available in open access online, from a GitHub repository. The Mask R-CNN neural network was used for instance segmentation [1]. Its detailed architecture is represented in Fig. 1. It is composed of three parts :

- A feature pyramid network (FPN), using a ResNeXt-101 as a backbone, that takes a (3 channels RGB) color image as input and retrieves 6 sets of 256 feature maps, with dimensions ranging from a fourth to a sixty-fourth of the initial image dimensions. Its purpose is to extract important features in the image to localize and identify an object.
- A region proposal network (RPN), that takes the last sets of feature maps one by one to predict the regions of interest (ROI) of the image that is the most likely to present an object.
- The ROI heads, a first head predicts the class scores of the potential object present in the ROI and its corresponding localization, *i.e.* its box coordinates. A second head predicts the binary mask that defines the contour of the detected object.

To complete that description, one should note that a ROI Pooler stage crops the features map using the ROI coordinates provided by the RPN before the ROI Heads. A ROI Pooler also rescales the ROIs to 7×7 pixels for the first head and 14×14 for the second head. Even if the general architecture as presented here is well defined and implemented, some hyperparameters must be tuned to adapt the Mask R-CNN to the problem of the study. Those hyperparameters are presented now. The input image dimensions were 512×512 pixels. For the RPN anchors generation, the P2 set of images was used twice, to focus on small object detection. The P3 to P6 sets of images were used once. The anchor sizes, in pixels, were : 8 p and 16 p for P2 ; 32 for P3 ; 64 p for P4 ; 128 p for P5 ; and 256 p for P6. Seven anchor aspect ratios were used in order to detect the linear perfect loops on-edge as well as the elliptic open perfect and faulted ones : 0.1, 0.2, 0.5, 1, 2, 5, and 10. At the end of each Pn set crossing through the RPN, only the 5,000 regions that obtained the best scores were held. Finally, of the 30,000 proposed regions, only the 5,000 regions with the highest scores are selected as ROI for the ROI Heads networks.

For the training only, among the last 5,000 proposed regions, a batch sample of 100 proposed regions was kept to train the ROI Heads. Among those 100 regions, it is important for the heads to learn on both foreground, *i.e.* objects, and background, to differentiate them. A target fraction of foreground regions of 0.25 was applied. The distinction between foreground and background is defined by an intersection over union (IoU) above or below 0.5. The IoU is

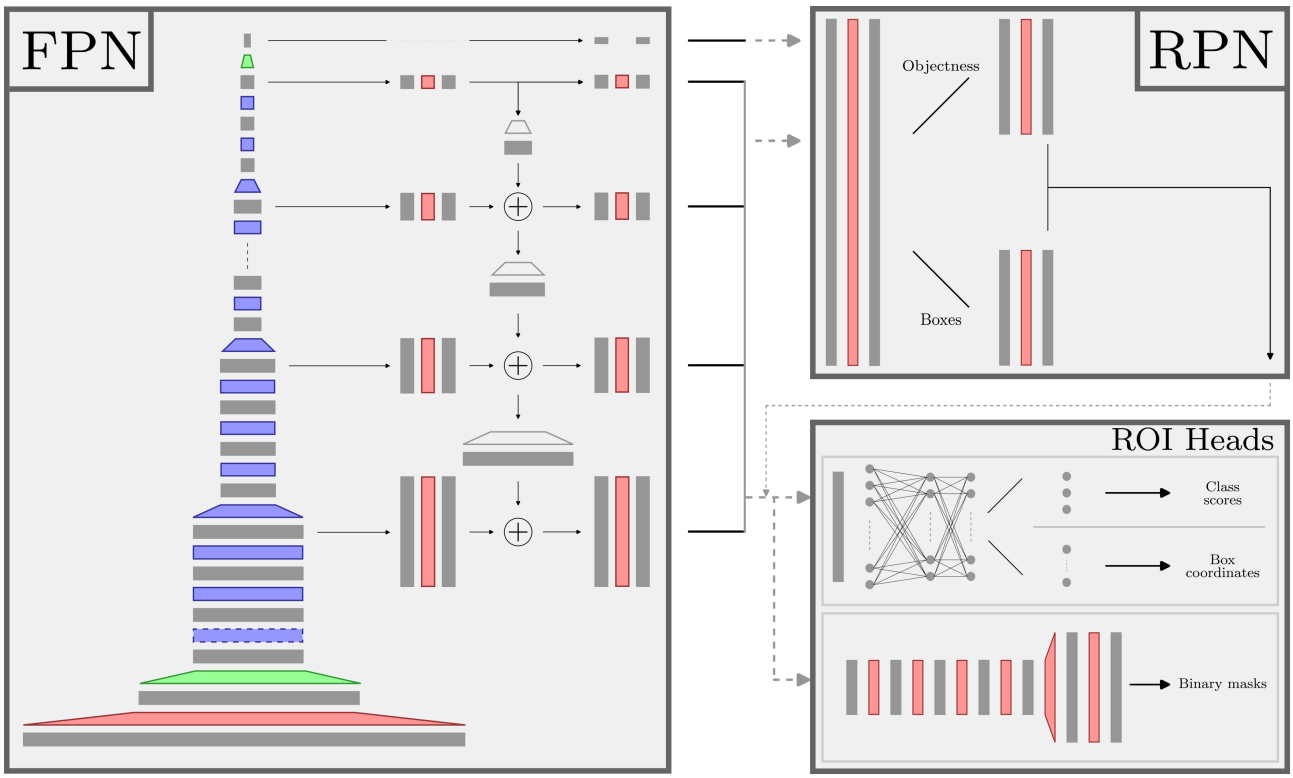

FIGURE 1 – Detailed architecture of a Mask R-CNN. Grey rectangles are a set of images, with the input image lying in the bottom-left corner of the feature pyramid network (FPN). Red rectangles and trapeze are convolutional layers. Green trapeze is max-pooling layers. The blue rectangles and trapeze are bottleneck blocks (a succession of convolutional layers). Layers represented as trapezes indicate a division of the dimensions of the images by a factor 2. Finally, the grey unfilled trapezes are simply nearest-neighbors up-sampling layers, to increase intermediary image dimensions.

simply the ratio of the intersection of the proposed and ground truth boxes over their union. Concretely, during training for each input image, 25 regions are used to train the ROI Head to recognize objects and 75 to recognize the background. During the inference phase, when using the trained neural network on a new image to get new object contours, all the 5,000 top-rated proposed regions go through the ROI Heads to detect potential objects. Only the predictions which get class scores higher than 0.8 are kept as valid predictions.

S2 Singular value decomposition on TEM micrographs

S2.1 Quick recall on SVD

Singular value decomposition (SVD) is a linear algebra process that allows to factorize a matrix. According to the spectral theorem, it is possible to diagonalize any $n \times n$ square matrix employing n eigenvectors, which form an orthonormal basis. The SVD is a generalization of this spectral theorem to a general $n \times m$ matrices.

Singular value decomposition factorizes a real matrix $\mathbb{X} \in \mathbb{R}^{m \times n}$ into three matrices : $\mathbb{X} \overset{\text{SVD}}{=} \mathbb{U} \cdot \mathbb{\Sigma} \cdot \mathbb{V}^\top$. Written in matrix representation :

$$\begin{pmatrix} x_{1,1} & \dots & x_{n,1} \\ \vdots & \ddots & \dots \\ x_{m,1} & \dots & x_{m,n} \end{pmatrix} = \begin{pmatrix} u_{1,1} & \dots & u_{n,1} \\ \vdots & \ddots & \dots \\ u_{m,1} & \dots & u_{m,m} \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & \dots & 0 & \\ \vdots & \ddots & \dots & 0 \\ 0 & \dots & \sigma_m & \\ & 0 & & \end{pmatrix} \cdot \begin{pmatrix} v_{1,1} & \dots & v_{1,1} \\ \vdots & \ddots & \dots \\ v_{n,1} & \dots & v_{n,n} \end{pmatrix}^\top \tag{1}$$

The size of the $\mathbb{\Sigma}$ matrix is the same as the size of the input matrix, *i.e.* $m \times n$. The bottom 0 in $\mathbb{\Sigma}$ indicates that when $m > n$, below σ_m the lines are full of 0. The right 0 indicates that when $n > m$, on the right of σ_m, the columns are full of zeros. Otherwise, the matrix is diagonal with the singular values sorted as $\sigma_i > \sigma_{i+1}$. $\mathbb{U}$ and $\mathbb{V}$ are respectively called left and right singular vectors of $\mathbb{X}$, as $\mathbb{X}$ may be expressed as a linear combination of $\mathbb{U}$ vectors as well as $\mathbb{V}$ ones. One should note that $\mathbb{U} \cdot \mathbb{U}^\top = \mathbb{1} = \mathbb{V} \cdot \mathbb{V}^\top$ Taking a 2×3 $\mathbb{X}$ matrix as an example :

$$\begin{aligned} \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{pmatrix} &\overset{\text{SVD}}{=} \begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_{1,1} & u_{2,1} & v_{3,1} \\ v_{1,2} & v_{2,2} & v_{3,2} \\ v_{1,3} & v_{2,3} & v_{3,3} \end{pmatrix} \\ &= \begin{pmatrix} u_{1,1}\sigma_1 v_{1,1} + u_{1,2}\sigma_2 v_{1,2} & u_{1,1}\sigma_1 v_{2,1} + u_{1,2}\sigma_2 v_{2,2} & u_{1,1}\sigma_1 v_{3,1} + u_{1,2}\sigma_2 v_{3,2} \\ u_{2,1}\sigma_1 v_{1,1} + u_{2,2}\sigma_2 v_{1,2} & u_{2,1}\sigma_1 v_{2,1} + u_{2,1}\sigma_2 v_{2,2} & u_{2,1}\sigma_1 v_{3,1} + u_{2,1}\sigma_2 v_{3,2} \end{pmatrix} \end{aligned}$$

Written that way, it appears that : i) the columns of $\mathbb{X}$ are linear combinations of the columns of $\mathbb{U}$; ii) the rows of $\mathbb{X}$ are linear combinations of the rows of $\mathbb{V}$ iii) the rank of the $\mathbb{X}$ is given by the number of non-zero values of σ's. Moreover, if we have a $\mathbb{X}$ matrix of rank r the best lower rank $p < r$ approximation of $\mathbb{X}$ matrix is given using the most important p left- and right- eigenvectors associated to the largest σ_p eigenvalues *i.e.* by truncating the $\mathbb{\Sigma}$ to the diagonal $\mathbb{\Sigma}_p \in \mathbb{R}^{p \times p}$ matrix having on diagonal the largest σ's, while left- and right- $\mathbb{U}$ and $\mathbb{V}$, respectively, are truncated to the first p columns, i.e. $\mathbb{U}_p \in \mathbb{R}^{m \times p}$ and $\mathbb{V}_p \in \mathbb{R}^{n \times p}$. The p-rank approximation of the $\mathbb{X}$ matrix becomes $\mathbb{X}_p = \mathbb{U}_p \mathbb{\Sigma}_p \mathbb{V}_p^\top$.

S2.2 Simple SVD on an image

The last conclusion means that the first components of the SVD reproduce the most global information of $\mathbb{X}$. In the case of image processing, a greyscale image is a simple $\mathbb{X}$ matrix. Those components rebuild the background in the image. The STEM-BF micrography of Fe^{2+}

irradiated HEA is once again presented in Fig. 2a, with the $\mathbb{U}$ matrix resulting from its singular value decomposition in Fig. 2b. That last appears mainly uniform, except in the few left columns of pixels, which encode the most information present in the micrograph.

The ten first columns are presented, broaden, in Fig. 2b-0) to Fig. 2b-9), the number corresponding to their column indexes. The first column, represented in Fig. 2b-0), appears mostly uniform compared to the other ones. As there are lot of dislocation loops in Fig. 2 micrograph, only that component is used to get the background. The linear combination of the other ones tends to reproduce the dislocation loops. This is illustrated in figures 3a, where are represented the reconstructed images using the first component, the two ones and ten ones respectively in a-0), a-1) and a-9).

Each columns of the image in Fig. 3a-0) is indeed the same vector, hence the horizontal lines, multiplied by a different factor to reproduce the best the mean vertical mean intensity, hence the vertical lines. Using two vectors complicate the result such that it is not possible anymore to recognize the same vector in each columns. Two vectors is too few to rebuild the image with the numerous dislocation loops. However, using the first ten components, an approximation of few objects is already found back, notably the dark faulted dislocation loops and the linear perfect on-edge ones.

The objective here is to remove the background. This means to reproduce the least as possible the dislocation loops, in order to keep a signal the closest to the experiments to analysis further. In other term, it is simply to remove the background while inducing the fewest artefacts as possible. The images presented in figures 3b correspond to the initial image subtracted by their left neighbors. It should be noted that from the first row, a linear artefact is observed in the treated images, due to the linear contrast of the background. A second effect is the apparition of bright areas in the images. A third one is the vanishing dark contrast inside the faulted dislocation loops when using more components, as the background estimation describe them better. These three effects get worse when considering more components for the background estimation. In that case with a lot of object it appears wiser to restrict the background estimation using only the first component of the SVD.

A first improvement is to convolve the background estimation with a gaussian kernel removing the final line artefacts and minimizing the two other ones. However a slightly more complex solution was developped and is discussed just next.

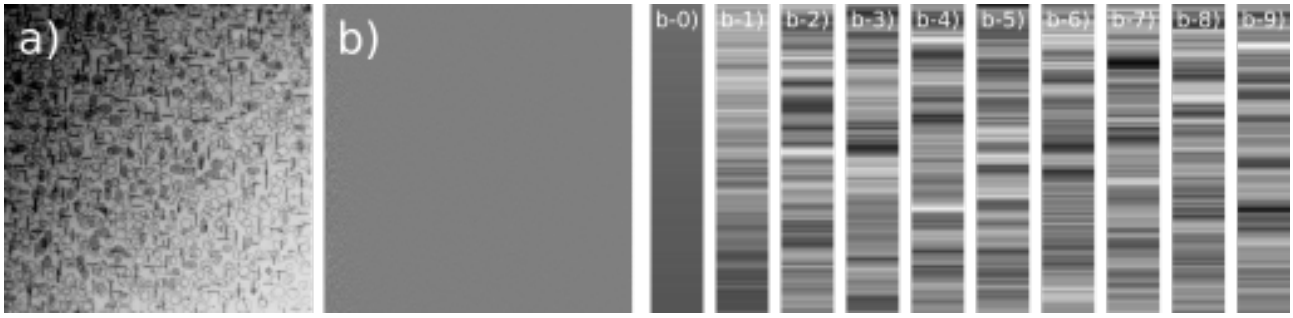

FIGURE 2 – (a) STEM-BF of an irradiated high entropy alloy. (b) $\mathbb{U}$ matrix resulting from the micrograph singular value decomposition. b-i) is a reproduction of the i^{th} column of $\mathbb{U}$.

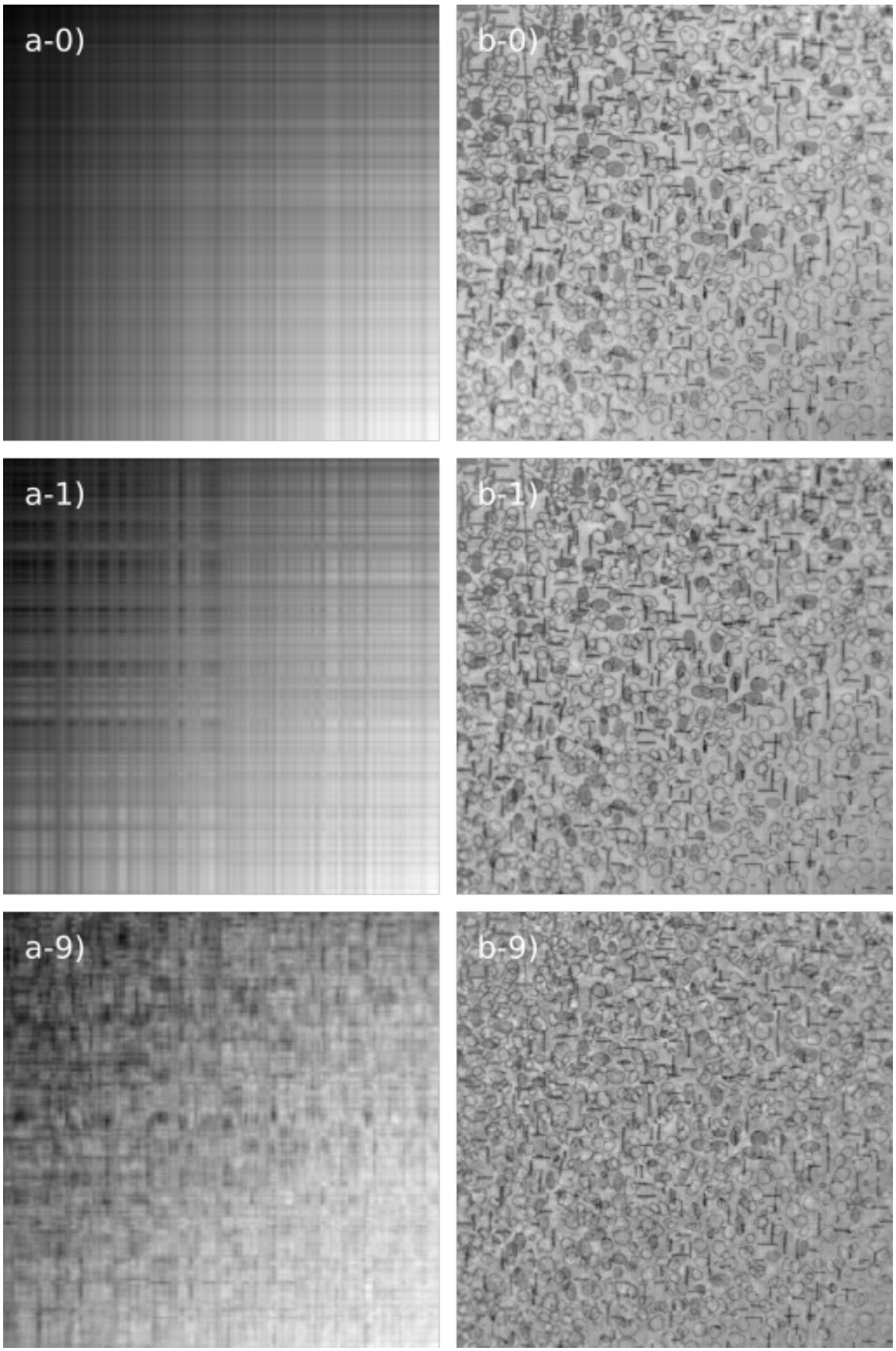

FIGURE 3 – (a-i) Reconstruction of the image using only the first i columns of $\mathbb{U}$, represented in the figures 2b-i. (b-i) Micrograph of Fig. 2(a) subtracted by (a-i).

S2.3 SVD treatment on a set of images

In the last treatment, the SVD finds patterns in the columns of pixels of the image. Its focus is to find a basis of vectors that allows to rebuild the image using linear combinations of the basis vectors. More over, the r first vectors of the basis must provide the best approximation of the original columns of pixels. In the case of the last micrography, it means that the SVD must find patterns in columns that are completely different and where pixels regarded as equivalent, by the SVD, because they have the same position are part of different defects. On one hand this way to proceed is a bit unsatisfying. On the other hand, luckily the background is linear in the micrography studied, which makes it possible to rebuild it with only one vector, and the method may experience difficulties regarding more complex cases.

Therefore, another treatment was investigated, considering $\mathbb{X}$ as a matrix of images, with its columns being flatted images instead of columns of an image. A question arises : how to use a set of images while only one micrograph present the observed gradient ? The initial image was convoluted with gaussian kernels, with different variances. Then they were flatten in vectors, which *in fine* define the $\mathbb{X}$ with a dimension $H \cdot W \times n$, H and W being the height and width of the image. Proceeding that way, the SVD process is more intuitive. Each elements of X represent the same defect as the elements of its row. Then the SVD tries to find patterns in equivalent pixels, in the sense that they have the same coordinate in the image and they are part of the same defect, for example. Compared to pixel-columns wise SVD, it is a huge progress, now SVD tries to find patterns in mostly unrelated vectors. An illustration of that process is represented in Fig. 4.

In a first place, the initial micrography $\mathbb{I}$ is blurred. Then the resulting images are flatten and putted into the matrix $\mathbb{X}$. The singular value decomposition is now applied on that large $\mathbb{X}$ matrix. That provide the basis vectors of $\mathbb{U}$, the *eigen-images*, then the decomposition α of the initial image on that basis is computed. The first eigen-image of $\mathbb{U}$, as well as the corresponding first decomposition factor a_0 of α are removed to form respectively $\mathbb{U}'$ and α'. Finally an approximation $\mathbb{I}'$ of $\mathbb{I}$ is reconstructed using $\mathbb{U}'$ and α'. In mathematic words, instead of reconstructing $\mathbb{I} = \mathbb{U}\mathbb{U}^{\mathrm{T}}\mathbb{I}$, because $\mathbb{U}\mathbb{U}^{\mathrm{T}}$, the image without background component is reconstructed : $\mathbb{I}' = \mathbb{U}'\mathbb{U}'^{\mathrm{T}}\mathbb{I}$.

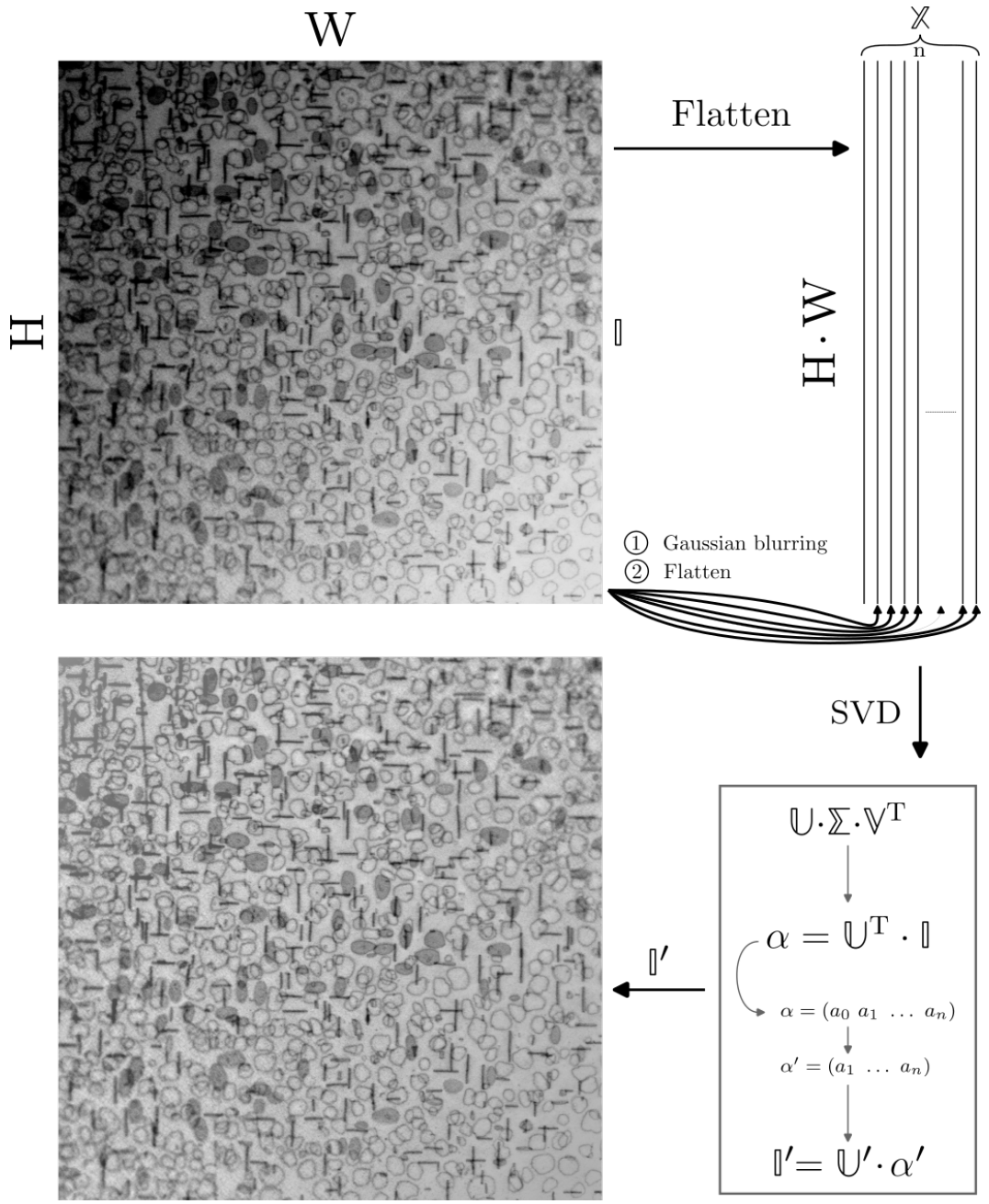

FIGURE 4 – Illustration of the second singular value decomposition process presented, the differents steps of the background removing are represented.

S3 Building a TEM learning database

Two approaches were identified and are discussed here to generate a database from experiments. The first method involves capturing micrographs under consistent illumination conditions. The second method entails using diverse illumination conditions as much as possible. The following section presents a discussion of the advantages and drawbacks of both methods.

S3.1 Strict illumination condition and data augmentation

Increasing the magnification yields improved object resolution and stable illumination conditions, facilitating object annotation. Additionally, it enables resolution reduction to mimic objects' appearance at lower magnifications. Another advantage is the lower number of objects in the micrograph, which requires fewer resources for processing. This aspect may be particularly relevant for samples with high defect densities. For example, for the micrograph depicted in Fig. 4a, a suitable approach is to capture 16 micrographs with four times higher magnification, resulting in approximately 50 objects per image, a reasonable number for the learning phase. The objects resolution would increase and be numerically downgradable to train a neural network with artificial different magnifications.

One of the key benefits of better annotation is that it can significantly improve the accuracy of model predictions. The person responsible for annotations carries the crucial responsibility of providing the neural network with ground truth, which is an absolute truth that the network can use as a reference. The quality of annotations directly impacts the inference performance of the model. In the case of dislocation loops, one could wonder where the dislocation loops points contour should be placed. Should they be placed around the dislocation line, strictly on it (i.e., the lowest intensity points of the contour), or inside the loop ? Applying the same criterion to most objects in Fig. 4a is feasible, but it becomes more complex for dislocation loops in the micrography of Fig. 5a. If the annotations are not accurate, it can result in a loss of database quality, ultimately affecting the model's detection and segmentation performance.

If we train a neural network solely on raw micrographs, it will only make accurate predictions for micrographs captured under similar conditions. However, the first method of preparing a database offers an additional advantage in that it simplifies the process of data augmentation

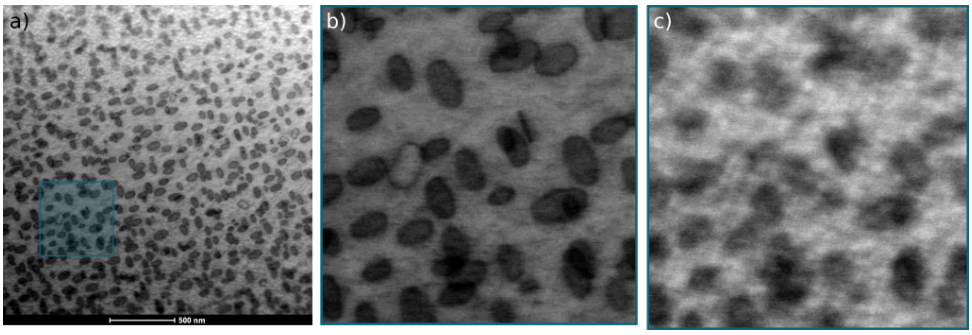

FIGURE 5 – (a) Typical STEM-BF micrograph of Y3 HEA exposed to 6 MeV Ni^{2+} ion irradiation. The micrographs cover different levels of brightness, contrast and noise. (b) and (c) are the same highlighted area from (a), but from two different STEM-BF acquisitions.

(DA). This involves modifying a micrograph by changing its resolution, brightness, contrast, or noise, then add it into the database. Data augmentation is performed to improve the transferability of the neural network. For instance, in Fig. 4a, the perfect on-edge loops are only visible in the horizontal and vertical directions. Since CNNs are not rotational invariants, inference on the same image rotated by an arbitrary angle, they won't detect those dislocation loops if the image is rotated by an arbitrary angle. Rotation augmentation resolve this issue and allow to detect them, whatever their orientation.

The main pitfall of the method is the difficulty to faithfully reproduce different TEM illumination conditions and TEM noises. The defects contrast may vary from one magnification to another, for example. But a stronger limitation is the TEM texture observed in Fig. 5c, which is particularly difficult to mimic. Moreover adding irrelevant micrographs far from any potential observations during augmentation process is counterproductive. The neural network learns on those, which wastes learning time on relevant images and decrease final performances.

S3.2 Probe experimental illumination conditions

To achieve better transferability, one could rely on more observations rather than augmentation. This can be accomplished by exploring as many illumination conditions as possible directly in the microscope. For example, in a low-magnification micrograph, the contrast, orientation, or z-position of the sample may vary locally. By doing this, a database with various conditions can be created quickly. However, at low magnification, more objects are recorded but with lower resolution, making the annotation task error-prone and likely to impact the learning phase and the final predictions' quality. In addition, in high magnification micrographs, high-resolution defects may not be detected. The micrograph of Fig. 5a is a typical example that carries different contrast and illumination conditions for the objects within. Fig. 5c shows a particular type of noise in the TEM, observed in the top-left and bottom-right areas of the micrographs. This noise is most likely caused by factors such as sample preparation, slight zone-axis shift, or low sample magnetism. It illustrates the difficulties could will face when trying to label the objects. These difficulties highlight the challenge of labeling objects with the same contour definition as such blurry objects.

An alternative approach is to capture high magnification micrographs under different illumination conditions, including one with clear defects imaging. High-quality labels can be created from images these, and can be used for the same objects in micrographs imaging the defects differently. High-quality labels can be created from images that are well-illuminated, and these labels can be used for the same objects in degraded micrographs. Augmentation is replaced with real micrograph acquisition, which reduces processing time but increases both TEM and annotation time. Particularly, creating a model with a good transferability requires a substantial amount of diligent TEM acquisitions. The idea of such a method is illustrated in Fig. 5. The same part of the sample in two different acquisition is represented in Fig. 5b and 5c. Object annotations are performed in the first image and used for both. This approach enables the model to detect objects that are difficult to discern with the naked eye, such as the perfect dislocation in Fig. 5b.

S4 Metrics details and augmentation insight

Initially, some datasets were built using only the micrograph presented in Fig. 4, with and without SVD, with and without partial (just rotational) or total (contrast, brightness) data augmentation. For the sake of clarity, the comparition of the learning metrics evolution of only two of those datasets are presented. That example illustrates that using micrographs with constant background as a base dataset, augmented further, it is possible to reach the performances of a dataset covering diverse experimental conditions.

The raw image was divided into 16 sub-images, 10 were kept for the learning, building a first dataset. A second dataset is obtained using the same sub-images, but after splitting the SVD treated micrograph. The 6 raw sub-images left were used to evaluate all models (validation sub-images). The quality of the models deduced is therefore quantified with raw transmission electron micrographs. Rotation augmentation was performed on the validation dataset to include on-edge dislocations in different orientations. A first dataset was created, named **NoSVD - DA**, with the 10 raw sub-images after the same rotation augmentation as validation dataset. A second one, named **SVD - FullDA**, was build based on the SVD treated images, but after rotation, contrast and brightness augmentations. Their performance comparison is presented in supplementary materials S4, along with a discussion on the learning metrics. This comparison confirms the interest of the data augmentation process.

Three metrics were followed during the learning phase. The learning loss, *i.e.* distance to the annotated ground truth of the training dataset, is evaluated at each learning step on sub-images used for the current step, and saved each 20 steps. The validation loss, is evaluated periodically each 100 iterations on the validation data, not used for training. At the same frequency, the mean average precision (mAP), discussed below, quantifies the quality of the predictions. The losses are calculated as the sum of the losses of : the RPN classification as foreground or background ; the RPN proposed ROI boxes coordinates ; the first ROI head classification ; its final box predictions ; and the second ROI head mask. For each of those losses, defined losses function is used, comparing the predictions with the annotated data.

The mAP is a slightly more complex notion. One should first define the precision and recall

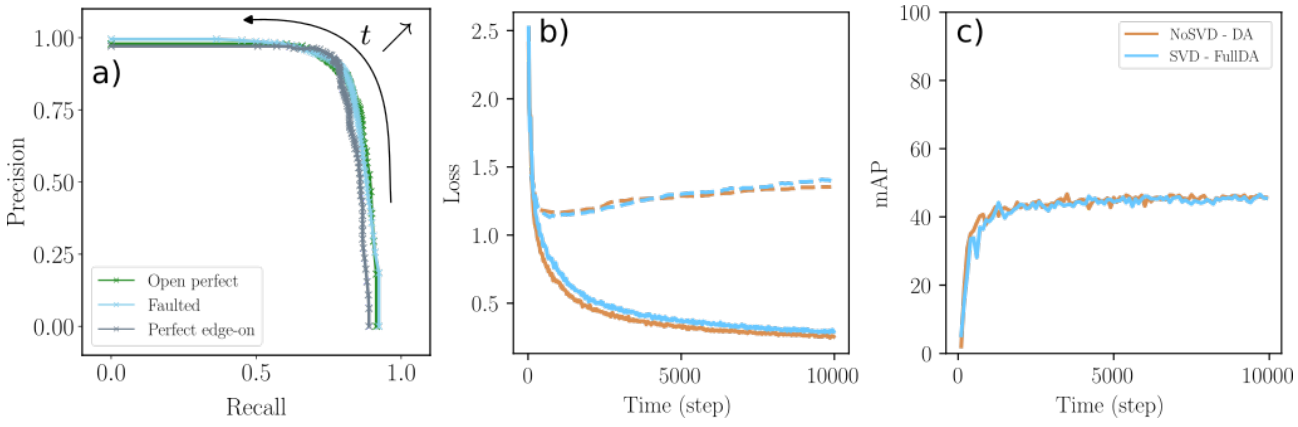

FIGURE 6 – (a) Precision-Recall curve, the integral under each curves gives the class AP, the mAP is the mean of these APs. Evolution of (b) the total loss (full lines), the validation loss (dashed lines) and (c) the mAP during the Mask R-CNN learning phase.

of an object class. The precision is the ratio of true positive predictions (TP) overall positive predictions (counting the false positive, FP), $P = {}^{\mathrm{TP}}/_{\mathrm{TP+FP}}$. It quantifies the confidence one can give to the predictions, the closer it gets to one, the more predictions will correspond to a ground truth annotation of the object and the fewer predictions correspond to other objects or backgrounds. Complementary, in order to quantify the number of objects that the neural network miss, the recall is defined as the ratio of true positive predictions over the total number of objects annotated (counting the false negative, FN), $R = {}^{\mathrm{TP}}/_{\mathrm{TP+FN}}$. The closer it gets to one, the greater the portion of annotated objects is predicted. Precision and recall depend both on the score threshold t that is considered to consider the prediction as positive, which is set to 0.8 as exposed on the last section. Precision-recall curves are then drawn for each category of objects, as depicted in Fig. 6a, varying that score threshold. A score threshold of 0.01 favors predictions, so it prevents false negatives but explodes the number of false positives. Therefore P tend to 0 and R to 1. In the opposite way, a threshold of 0.99 limits predictions. Then it prevents false positives, $P \to 1$, but it favors false negatives, $R \to 0$. In between those two limits, the values of P and R are between 0 and 1, as is the integral under the $P(R)$ curve, which defines the so-called average precision (AP) of the model on the object class. Finally, the mean average precision, mAP, is defined as the mean of the different AP.

Fig. 6b and 6c present the evolution of the last metrics. The total loss decrease indicates better performances thanks to the learning. Besides, after around 1000 steps, validation loss starts to increase instead of decreasing just like the total loss. This indicates most probably an undesirable overfitting on the learning database. Larger learning dataset minimizes that effect. The learning and performance curves of the models trained with the **NoSVD - DA** and **SVD - FullDA** overlap. It indicates that the artificial contrast and brightness variations in the second dataset include experimental variations observed in the raw micrograph of Fig. 2a. Moreover, thanks to augmentation, larger experimental conditions are covered, making the model transferable to illuminations not represented in the initial micrographs.

S5 Segmentation pitfalls - Limitations and solutions

This section aim is a discussion on the limitations encountered while developing the dataset and training the subsequent neural network, why these limitations arise and the different tracks to overcome them. The question of the detection of different objects size is discussed first. It is followed with a discussion on the detection of objects in different illumination conditions. Finally, the detection of dislocation loops in the noise.

S5.1 Size of the detects

The micrographs in Fig. 7a, 7b and 7c were taken in the Y3-Fe samples with a high magnification. As one could notice, the detection of defects imaged with a high resolution is far from expectations. This is mainly due to the fact that the dataset was build without such resoluted objects. The main objective of using a neural network here is to count and characterize a large amount of defects, in micrographs such as the one presented in Fig. 7d, or in Fig. 3. In such micrographs, the object resolution is significantly lower than the previous ones. Therefore false negatives may happen, as in Fig. 7a, wrong classifications as depicted in Fig. 7b, or both as illustrated in Fig. 7c.

As explained, this high objects pitfalls is due to lack of representative large defects in the

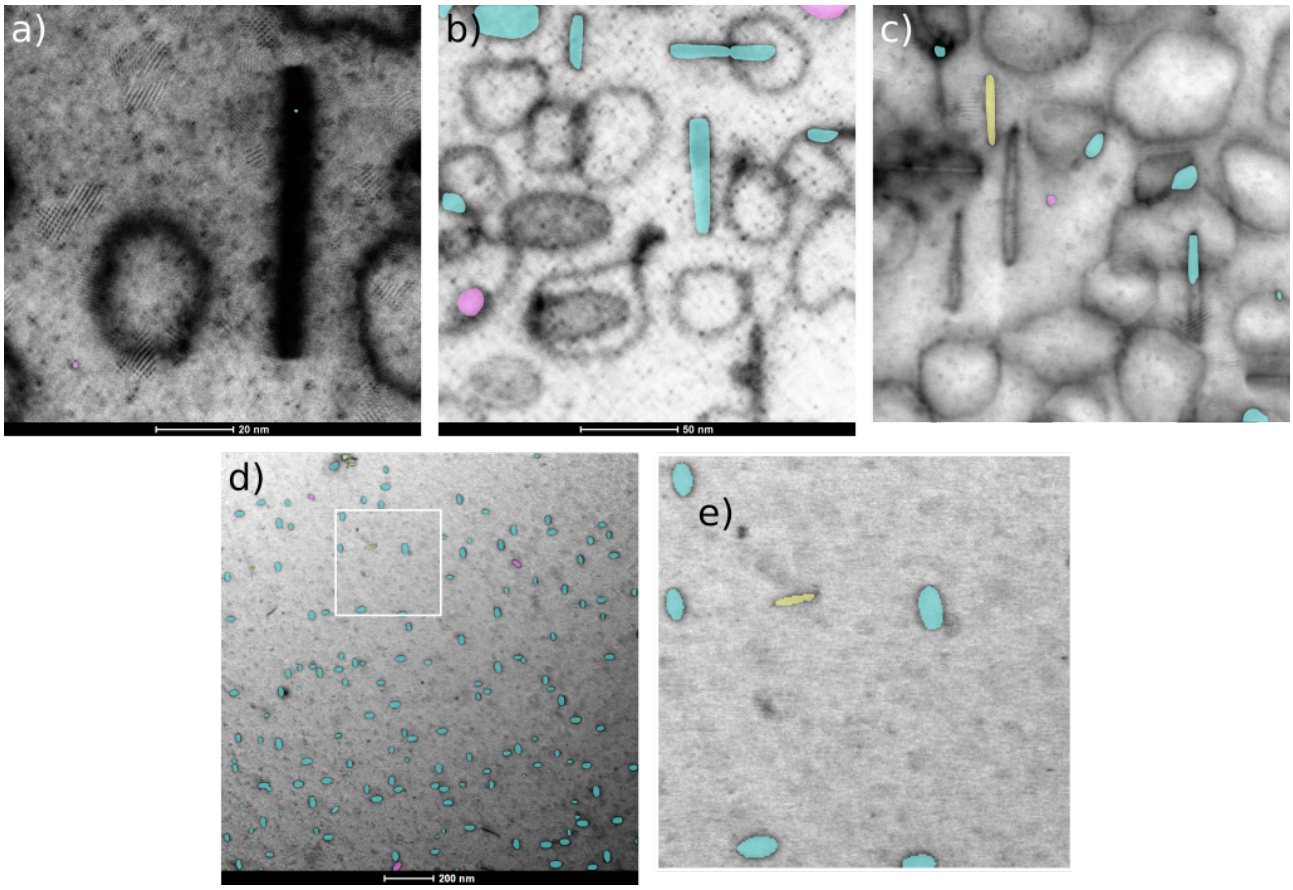

FIGURE 7 – (a-c) Dislocation loops in Y3-Fe samples imaged in STEM-BF at a high magnification. (d) STEM-BF micrograph of dislocation loops in ES1-Ni sample at a lower magnification. (e) Zoom on a small (less than 8×8 pixels) dislocation loop, undetected by the Mask R-CNN.

learning dataset. Regarding the architecture of the neural represented in figure 4 of the article, this lack of large objects representative data lead to a weak learning in the higher convolution layers of the FPN, in charge of extracting large objects features. Append the dataset with higher magnification micrographs would strengthen the learning in those layers and result in better detection for the presented micrographs. A second option presented in the paper consists of decreasing the image resolution, from 2048×2048 of a factor 4 to 8 to get objects resolution close to the ones of the dataset. The predicted masks can then be upgraded to fit in the original micrograph.

Finally, the micrograph of Fig. 7d presents a STEM-BF micrography of the ES1-Ni sample with good predictions for the small defects, as the learning dataset was build for it. A zoom in an area with an undetected dislocation loop is represented if Fig. 7e. This undetected defect is on the limit of the detection of the model, as it is seen as a 2×2 area in the RPN. From a more technical point of view, for the detection of both small and large defects, it should be noted that a specific attention must be adressed to the anchor sizes of different FPN output set of images.

S5.2 Illumination conditions

Fig. 8 illustrates the effect of a non-representative illumination diversity in the dataset when it comes to inference post-learning. Annotations were drawn on Fig. 8a and it was used as validation dataset during the learning phase of two models : i) one trained on a dataset is composed of sub-images of SVD treated Y3-Fe micrograph, see the bottom micrograph of Fig. 4 ; ii) the second trained on the dataset explicited in section 2.2 of the article, with illumination data augmentation on the previous SVD treated Y3-Fe micrograph. The subsequent model predictions on the micrograph are presented in Fig. 8b and 8c. In the first, predictions are only correct in the central area, where the illumination is close to the dataset one. Using a model trained on the SVD + DA treated dataset, the results are far better, with less defects

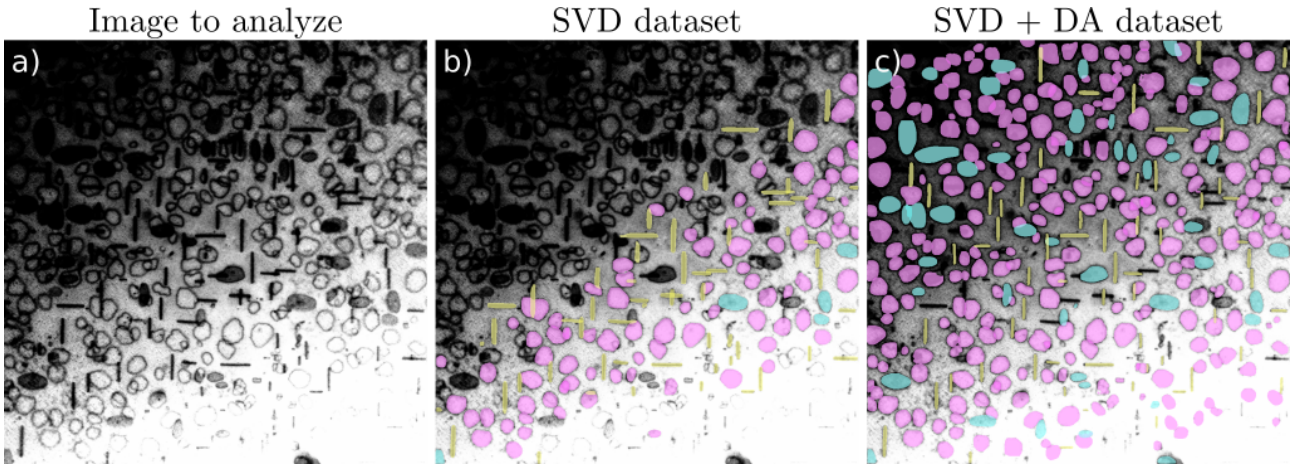

FIGURE 8 – (a) Raw micrograph useen by the neural network model during the learning phase. (b) Predictions on this micrograph, using a model trained only with the Y3-Fe SVD treated micrograph, in order to minimize background effect. (c) Predictions on the same micrograph, but using the final model, with more micrographs, and illumination data augmentation.

missed, even though one could point a low detection of on-edge DL, most probably to their under-representation in the dataset, or size effects. Therefore, this figure highlights the huge interest of the two step process : i) remove background fluctuation using a SVD treatment ; ii) artificially augment, in a controlled manner, illumination condition in the dataset, modifying brightness and contrast of the micrographs. One should note that to improve detection rate on different illumination micrographs, augmenting the dataset, either with new micrographs or by augmenting the acquired ones, is a good track to follow. The latter option was presented here for its accordance to the line of the paper, minimizing the human time to prepare a dataset, *i.e.* less acquisition and less further labeling needed.

S5.3 Noisy images

As a strange TEM noise was observed in Y3 STEM-BF micrographs, see figure 3 of the article or Fig. 9 below, a special effort was made to detect objects in those noisy parts of micrographs. Fig. 9a presents predictions of a Y3-Fe STEM-BF micrograph, a strong detection decrease is observed in the highlighted top-right corner. That kind of low detection rate lead us to consider micrographs such as the one presented in Fig. 9b, with a kind of noisy texture in the top and bottom zones. The initial variation of brightness when thickness increases lead us to remove the illumination data augmentation step. This is with that final dataset that we were able to detect dislocation loops such as in the Fig. 9c. Once again, augmenting the dataset with data representative of the detection loss lead to major improvement in noisy defect detection. It is however complicated to reproduce various noises that would occur in a TEM and on this specific part, data augmentation isn't of any help.

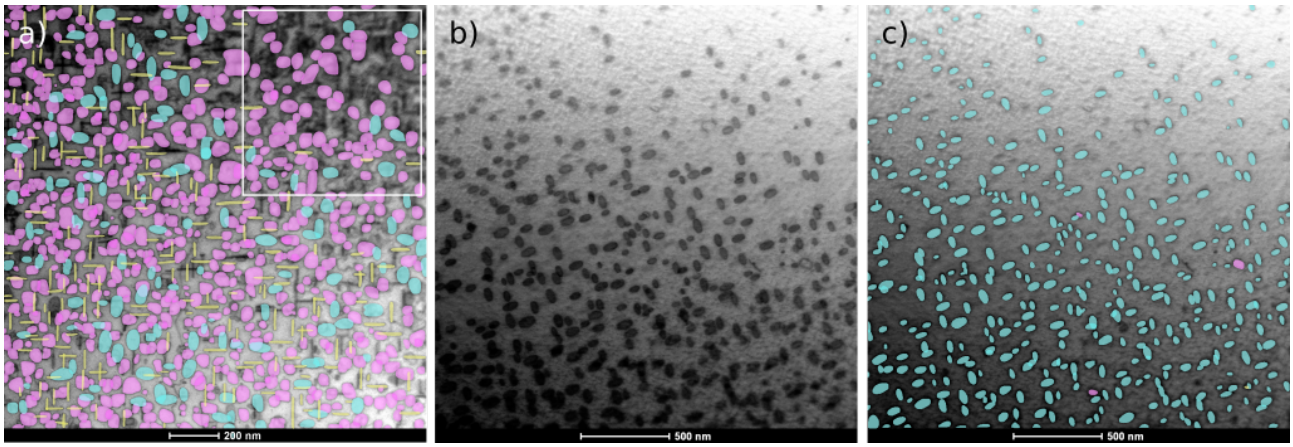

FIGURE 9 – (a) Illustration of a decrease in dislocation loop rate in a blurry area of a STEM-BF micrograph of Y3-Fe sample. (b) Initial STEM-BF micrograph taken on Y3-Ni, dislocation loops are imaged with a consequent noise in the top and bottom part of the image. (c) Prediction on such micrographs when using a dataset with close micrographs for the learning.

S6 Radial distribution function on a toy model

A simple modelization of the TEM lamella probed by the electron beam was performed by placing points inside a cube, using *Python*. Different geometries and distributions were tested. The distribution discussed in the maniscript is the one that makes the most sense to compare with the experiments. To recall, the points were randomly placed following : i) a Gaussian law distribution along the z axis ; ii) a minimal distance r between the points. An illustration of this random distribution following simple laws in a cube is presented in the top of Fig. 10. The bottom part represents the corresponding 2D projections, which is similar to a TEM micrograph. Finally, the gradient dark circle is represented to explain further the RDF calculation process.

In a distribution of points, the RDF of a specific point (the one on which the dark circle is centered in Fig. 10) is defined by drawing a set of consecutive rings. Their width was set to 1 pixels. For each of those rings, the number of objects within the ring is counted and divided by the area of the ring within the image, hence the cut on the border in Fig. 10. This defines an RDF of the distance to the specific point, which is the radius in axis of the Fig. 4 of the manuscript. The final RDF is then calculated as the mean of the RDFs calculated for all objects.

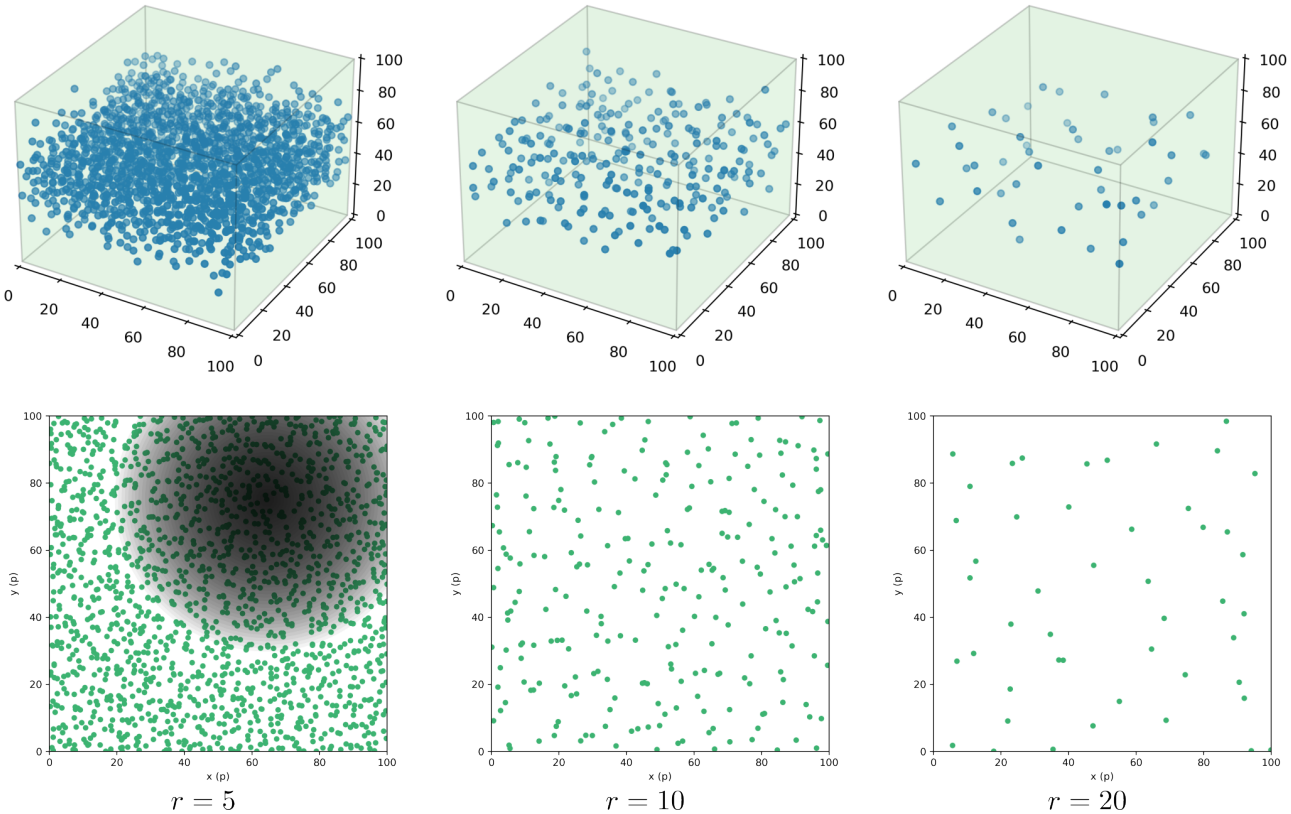

FIGURE 10 – Top : Boxes with points distributed with a minimal distance $r = 5, 10, 20$ in between them. Bottom : The corresponding two dimensional projection, equivalent to a TEM image. The gradient dark circle in the left represents the successives rings used to compute the RDF for a specific point.

S7 Ground truth labels vs model detections

The comparison between ground truth labels and model detections is presented in Fig. 11. The predicted masks quality is close to the labels, the shapes even appears more regular especially around on-edge dislocation loops.

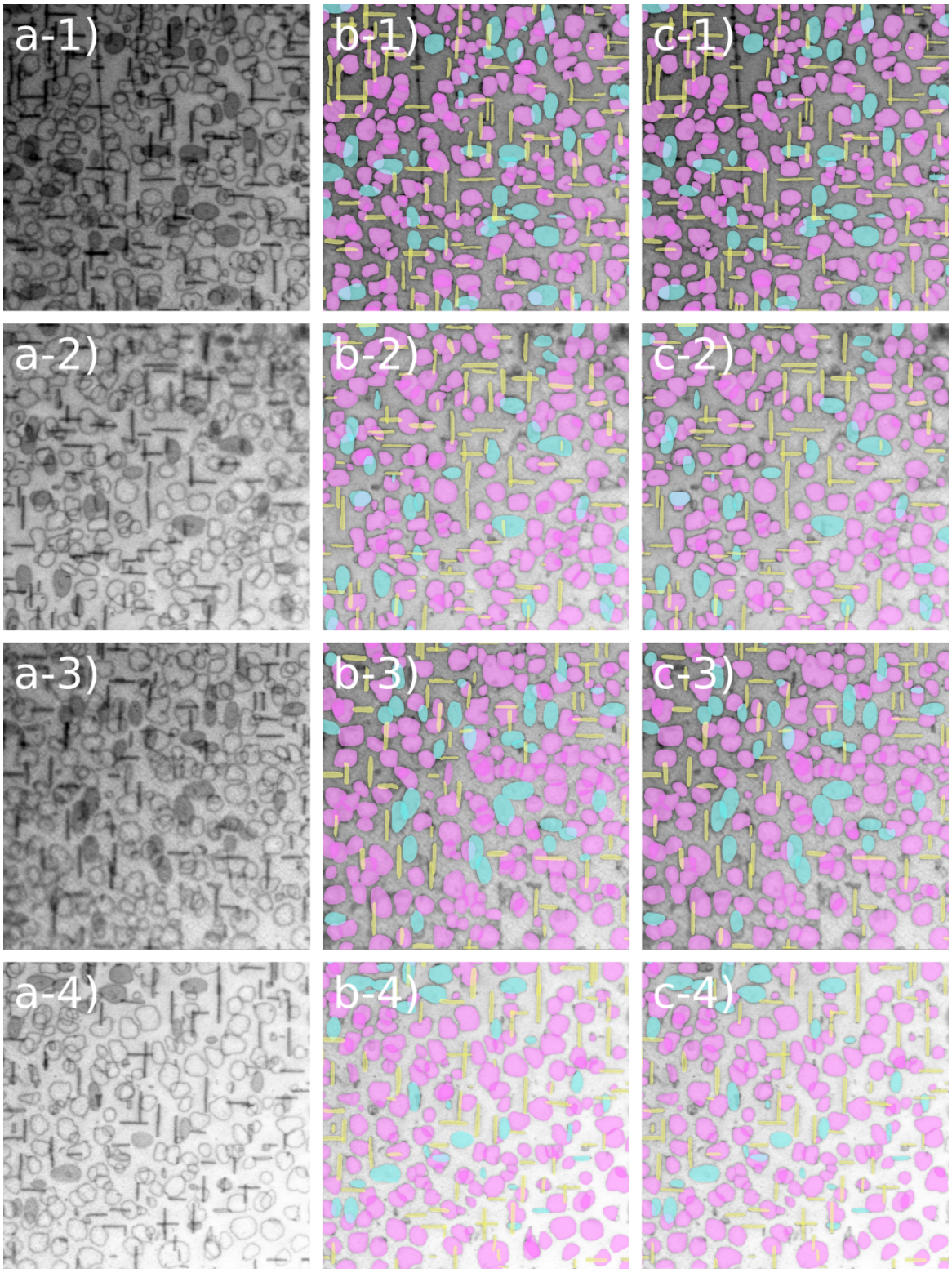

FIGURE 11 – Comparison of (a) subpart of an initial micrograph, (b) ground truth labels and (c) the model predictions.

Références

[1] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask r-cnn," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.