Engineering Applications of Artificial Intelligence 119 (2023) 105743

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Identification and classification of exfoliated graphene flakes from microscopy images using a hierarchical deep convolutional neural network

Soroush Mahjoubi [a], Fan Ye [b], Yi Bao [a,*], Weina Meng [a,*], Xian Zhang [c,*]

[a] *Department of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA*
[b] *Department of Electrical and Computer Engineering, The University of Massachusetts Amherst, Amherst, MA 01003, USA*
[c] *Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA*

ARTICLE INFO

Keywords:
Deep convolutional neural network
Machine learning
Nanomaterials
Optimized adaptive gamma correction
Semantic segmentation
Two-dimensional (2D) material

ABSTRACT

Identification of exfoliated graphene flakes and classification of the thickness are important in the nanomanufacturing of advanced materials and devices. This paper presents a deep learning method to automatically identify and classify exfoliated graphene flakes on Si/SiO_2 substrates from optical microscope images. The presented framework uses a hierarchical deep convolutional neural network that is capable of learning new images while preserving the knowledge from previous images. The deep learning model was trained and used to classify exfoliated graphene flakes into monolayer, bi-layer, tri-layer, four-to-six-layer, seven-to-ten-layer, and bulk categories. Compared with existing machine learning methods, the presented method showed high accuracy and efficiency as well as robustness to the background and resolution of images. The results indicated that the pixel-wise accuracy of the trained deep learning model was 99% in identifying and classifying exfoliated graphene flakes. This research will facilitate scaled-up manufacturing and characterization of graphene for advanced materials and devices.

1. Introduction

Two-dimensional (2D) materials, endowed with unique electrical (Radisavljevic et al., 2011), ferromagnetic (Huang et al., 2017), semiconducting (Xu et al., 2014), superconducting (Xi et al., 2016), thermal (Zhang et al., 2015), and optical properties (Li et al., 2014) are ideal candidates for mechanical, photonic, and electronic devices. Fabrication of most 2D devices starts with mechanical exfoliation (Novoselov et al., 2004, 2005), which ensures the most pristine properties and conditions of 2D materials. Mechanical exfoliation methods yield flakes of 2D materials with random thicknesses and shapes on the substrate, followed by a manual search under an optical microscope. For scientific and device development purposes, the ideal flakes have a uniform thickness, and many interesting phenomena occur only in atomically thin samples. In current practices, flake thickness is mainly evaluated by two methods: (1) manual examination of the color and contrast of flakes from microscope images, and (2) material characterization equipment, such as Raman microscope and atomic force microscope (Li et al., 2013; Lin et al., 2018; Masubuchi and Machida, 2019; Ni et al., 2007; Nolen et al., 2011; Blake et al., 2007). The two prevalent methods have disadvantages: Manual examination inevitably yields errors due to different colors and background conditions regarding color and brightness, and characterization of 2D flakes is costly, inefficient, and time-consuming. The two methods suffer from a trade-off between efficiency and accuracy. Efficient and accurate methods for identifying thickness are highly needed (Shelhamer et al., 2017; Chen et al., 2018b; Ronneberger et al., 2015; Badrinarayanan et al., 2017).

Deep learning is a subset of machine learning based on artificial neural networks that can solve sophisticated machine learning problems (Mahjoubi et al., 2023). Previous studies show that deep learning methods can produce results comparable or better than human experts (Chong et al., 2020; Allugunti, 2022). Deep learning is represented by a variety of architectures. Deep neural network is the most popular architecture that has been applied to various problems such as computer vision (Hassaballah and Awad, 2020), natural language processing (Otter et al., 2020), material discovery (Mahjoubi et al., 2022), and drug discovery (Chen et al., 2018a). Convolutional neural networks (Gu et al., 2018) are a type of deep neural network that is designed for processing multidimensional data such as images and has been widely applied for computer vision tasks, such as image classification (Wang et al., 2016), object detection (Ren et al., 2015), and pixel-wise semantic segmentation (Hazirbas et al., 2017).

A few methods based on convolutional neural networks were reported to identify 2D flakes from microscopy images (Masubuchi and Machida, 2019; Masubuchi et al., 2018; Han et al., 2020; Saito et al., 2019; Greplova et al., 2020; Masubuchi et al., 2020). Five major limitations were identified from the existing methods: (1) The existing

---

* Corresponding authors.
E-mail addresses: yi.bao@stevens.edu (Y. Bao), wmeng3@stevens.edu (W. Meng), xzhang4@stevens.edu (X. Zhang).

https://doi.org/10.1016/j.engappai.2022.105743
Received 23 September 2022; Received in revised form 22 November 2022; Accepted 15 December 2022

Table 1
Statistics of class weights.

Designation of class	Description of class	Mean class weight	Median class weight	Maximum class weight	Images with zero-class weight
0	Background	91.99%	95.87%	99.83%	0.00%
1	1L	0.59%	0.00%	19.72%	75.63%
2	2L	1.77%	0.00%	62.62%	55.51%
3	3L	0.70%	0.00%	15.46%	55.15%
4	4–6L	0.68%	0.00%	12.17%	61.76%
5	7–10L	0.47%	0.00%	27.51%	79.04%
6	Bulk	3.81%	0.00%	40.61%	51.84%

Note: "L" stands for layer(s).

methods were designed for specific microscope conditions such as background color and light brightness, compromising the applicability under different conditions. (2) The existing machine learning models are incapable of learning new micrographs. Re-training the existing models using new data leads to catastrophic loss of the knowledge learned from old data (Kirkpatrick et al., 2017). (3) Optical microscope images suffer from artifacts such as poor contrast, vignetting, and overexposure. Multiple methods have been developed to enhance microscope images for human vision (Chen et al., 2010; Leong et al., 2003), but not applicable to machine vision. (4) Imbalanced class distribution in training images compromises the evaluation of under-represented classes (Wu et al., 2019). The areas of thin flakes are relatively small compared with the areas of the substrate (background) and bulk flakes. (5) The classification precision of graphene thickness was rough. The flakes were categorized into three classes (Masubuchi et al., 2018), which are insufficient in many applications.

The motivation of this study is to address the identified limitations through developing a novel hierarchical deep learning method based on an unsupervised classification model and multiple semantic segmentation models and evaluate the performance of the method in automatically identifying exfoliated graphene flakes and determining thicknesses from optical microscope images. The flakes are categorized into six classes: monolayer (1L), bi-layer (2L), tri-layer (3L), four-to-six-layer (4–6L), seven-to-ten-layer (7–10L), and bulk. Novel computer vision techniques are presented to improve microscope images using an image quality metric and improved adaptive gamma correction (Cao et al., 2018). The proposed method leverages weak learning, data augmentation, iterative stratification (Sechidis et al., 2011), and weighted cross-entropy loss to improve performance in terms of accuracy and generalizability. The effects of resolution and background conditions of images are considered to improve the robustness of the deep learning model. The proposed method will promote the capability of processing and manufacturing 2D materials and devices by improving efficiency and accuracy while minimizing human intervention.

The present study is novel in three aspects: (1) A robust framework based on unsupervised learning and supervised deep learning is presented to identify the thickness of 2D flakes from microscopy images. The presented framework classifies 2D flakes into six categories, is robust to microscopy conditions, and has reasonable accuracy for any of the considered categories. (2) A novel computer vision method, named optimized adaptive gamma correction, is developed based on optimization to improve the quality of microscopy images for machine vision. (3) weak learning and iterative stratification are proposed in this study to improve the generalization performance of deep learning models trained on imbalanced datasets.

The remainder of the paper is organized as follows: Section 2 discusses the methodology of this study; Section 3 discusses the results obtained by the proposed framework; and Section 4 discusses the conclusions and future prospects.

2. Methodology

This section elaborates the hierarchical framework (Fig. 1) which leverages supervised and unsupervised machine learning. The hierarchical framework has five main steps: (1) Dataset development: Optical microscopy images are captured to develop a dataset for training a machine learning model (see Section 2.1). (2) Dataset enhancement: The quality of images is improved using multiple novel methods developed in this study (see Section 2.2). (3) Dataset standardization and categorization: The images are standardized and categorized using a novel unsupervised machine learning model (see Section 2.3). (4) Semantic segmentation: A deep convolutional neural network is trained to identify exfoliated graphene flakes and quantify layer numbers (see Section 2.4). (5) Performance evaluation: The performance of the proposed framework is evaluated using performance metrics (see Section 2.5). Fig. 2 shows the pseudocode of the proposed hierarchical framework.

2.1. Dataset

A high-quality dataset is essential for training a machine learning model. The dataset of microscopy images of exfoliated graphene flakes was developed in three steps: (1) An optical microscope (Nikon Eclipse LV150N) was used to capture RGB images of graphene flakes that were mechanically exfoliated from bulk graphite and transferred to SiO_2/Si substrates. A total of 273 images were captured. (2) An atomic force microscope (Bruker BioScope Resolve) and Raman spectroscope (Renishaw inVia Confocal) were used to evaluate the thickness of each flake. (3) Pixel-wise ground truth labels were annotated with the thickness evaluation results. The pixels were categorized into seven classes: background, 1L, 2L, 3L, 4–6L, 7–10L, and bulk.

Statistical analysis was conducted using the ground truth labels to evaluate the distributions of the seven classes. The weight of the ith class was defined as the number of pixels of the ith class divided by the total number of pixels. Table 1 lists the results of the mean, median, and maximum class weights of the seven classes, as well as the percentage of images with a zero-class weight. The mean class weights of classes 1, 3, 4, and 5 were less than 1%. The median class weights of classes 1 to 6 were zero, while the median class weight of background was higher than 95%. The percentages of images with zero-class weight for classes 1 (1L) and 5 (7–10L) were higher than 75%. The statistical results indicated that the different classes in the dataset were highly imbalanced. The interested classes 1 to 6 were under-represented, while the background class was over-represented. The statistical analysis was extended to the colors of images. It was found that different images had different background colors and brightness (see (Fig. A.1). Due to the variation of pixel color and intensity, it is unsuitable to use color-based segmentation methods (Masubuchi and Machida, 2019). In summary, the dataset of images introduces challenges due to the imbalanced classes and scattering of background colors and brightness.

Microscopy images introduce challenges due to overexposure, specular reflection, vignetting, dirt, and out-of-focus problems: (1) Overexposure: When excessive light is received by the microscope, overexposure may occur and cause loss of detail in images. (2) Specular

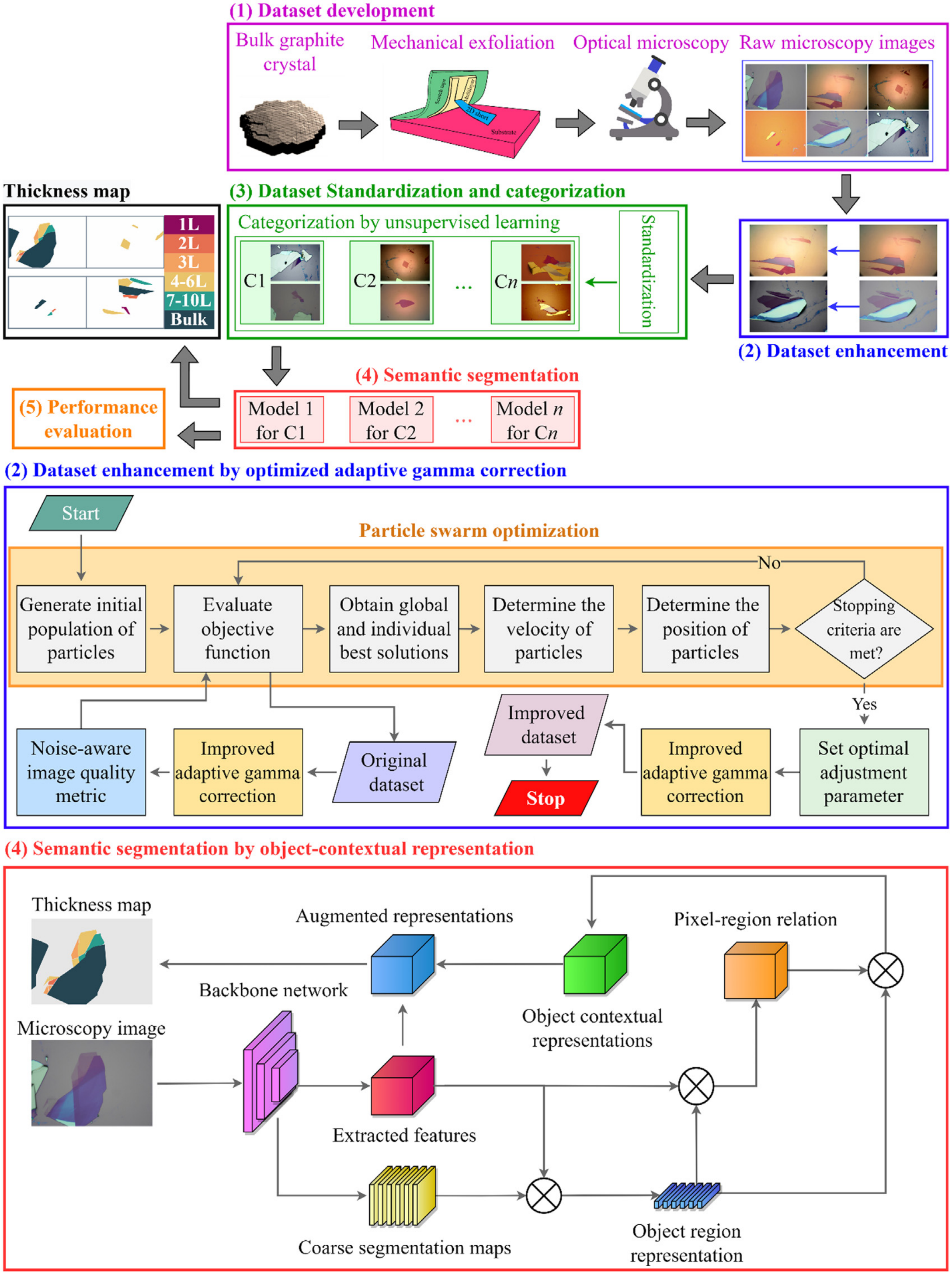


Fig. 1. Flowchart of the robust hierarchical framework for classification of 2D flakes.

reflection: Overhead lighting may cause specular reflection of graphene flakes and lead to oversaturation and loss of detail. (3) Vignetting: Vignetting appears as a radial darkening toward the corners of an image. (4) Dirt: Dirt and debris cause blur images. (5) Out-of-focus: Improper focus of the microscope causes blur images and loss of details.

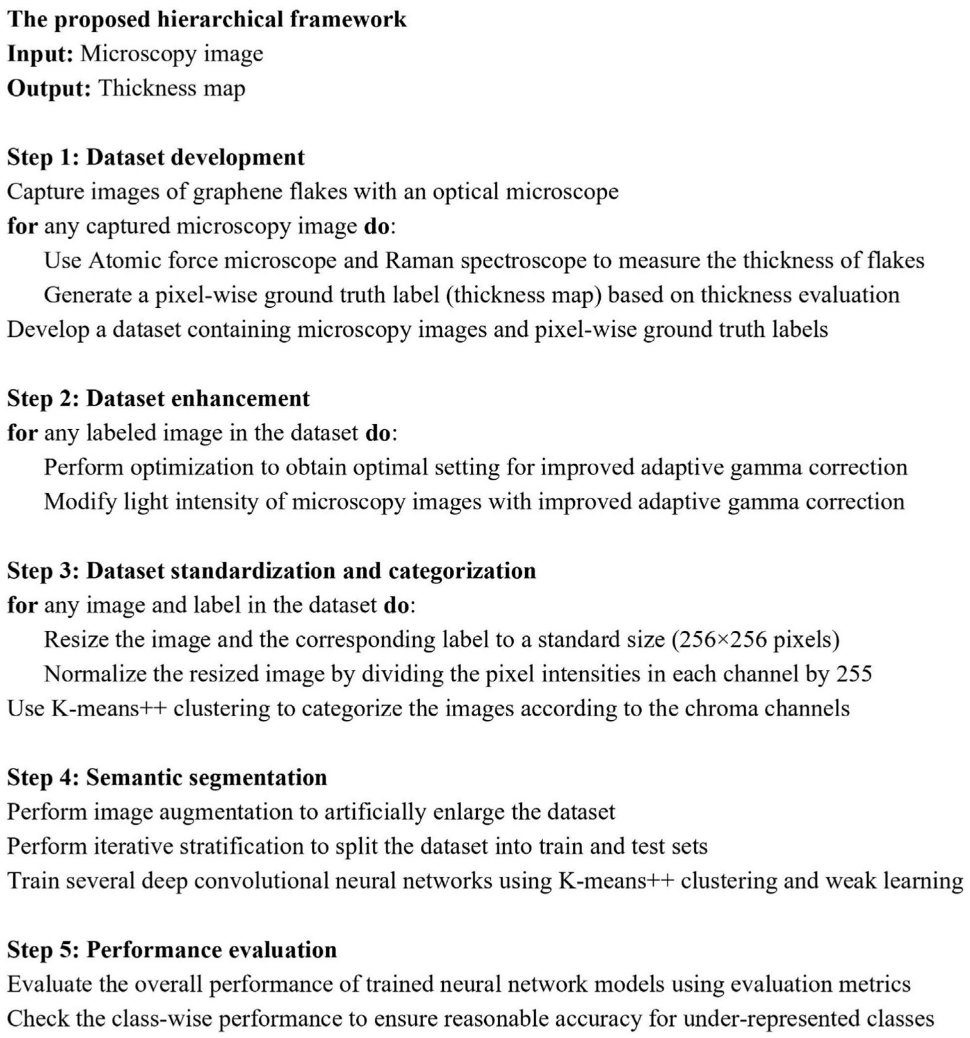
2.2. Optimized adaptive gamma correction

To improve the quality of images, we present an optimized adaptive gamma correction method (Fig. 1), through integrating the improved adaptive gamma correction method (Cao et al., 2018), noise-aware image quality metrics (Shin et al., 2019), and particle swarm optimization (Kennedy and Eberhart, 1995). The development and integration of these methods are elaborated in the following subsections.

2.2.1. Improved adaptive gamma correction

The improved adaptive gamma correction method modifies light intensity while preserving the color properties of images with overexposure problems in the following eight steps (Cao et al., 2018):

(1) RGB images are converted to YC_bC_r color space, where Y is the luma component for light intensity; C_b and C_r are the blue-difference and red-difference chroma components, respectively.



```
The proposed hierarchical framework
Input: Microscopy image
Output: Thickness map

Step 1: Dataset development
Capture images of graphene flakes with an optical microscope
for any captured microscopy image do:
    Use Atomic force microscope and Raman spectroscope to measure the thickness of flakes
    Generate a pixel-wise ground truth label (thickness map) based on thickness evaluation
Develop a dataset containing microscopy images and pixel-wise ground truth labels

Step 2: Dataset enhancement
for any labeled image in the dataset do:
    Perform optimization to obtain optimal setting for improved adaptive gamma correction
    Modify light intensity of microscopy images with improved adaptive gamma correction

Step 3: Dataset standardization and categorization
for any image and label in the dataset do:
    Resize the image and the corresponding label to a standard size (256×256 pixels)
    Normalize the resized image by dividing the pixel intensities in each channel by 255
Use K-means++ clustering to categorize the images according to the chroma channels

Step 4: Semantic segmentation
Perform image augmentation to artificially enlarge the dataset
Perform iterative stratification to split the dataset into train and test sets
Train several deep convolutional neural networks using K-means++ clustering and weak learning

Step 5: Performance evaluation
Evaluate the overall performance of trained neural network models using evaluation metrics
Check the class-wise performance to ensure reasonable accuracy for under-represented classes
```

Fig. 2. Pseudocode of the robust hierarchical framework for classification of 2D flakes.

(2) The luma component is used to convert an image into a negative image:

$$Y(x,y)_N = 255 - Y(x,y) \tag{1}$$

where $Y(x,y)_N$ is the negative image; $Y(x,y)$ is the luma component of the input image; x and y are the coordinates of images: $x = 1,2,\ldots,M$, and $y = 1,2,\ldots,N$, where M and N are the width and height of an image in pixel.

(3) The probability density function of pixels with an intensity level l is expressed as:

$$PDF(l) = \frac{n_l}{MN} \tag{2}$$

where n_l is the number of pixels with an intensity level l.

(4) The weighting distribution function is expressed as:

$$W(l) = PDF_{max}\left[\frac{PDF(l) - PDF_{min}}{PDF_{max} - PDF_{min}}\right]^{\alpha} \tag{3}$$

where PDF_{max} and PDF_{min} are the maximum and minimum probability density functions of pixels with an intensity level l, respectively; α is the adjusted parameter; and l_{max} is the maximum intensity level of the luma component.

(5) The cumulative distribution function is expressed as:

$$CDF(l) = \frac{\sum_{i=1}^{l} W(i)}{\sum_{i=1}^{l_{max}} W(i)} \tag{4}$$

(6) The modified luma component of the negative image is obtained from the following transformation function:

$$I'(x,y)_N = l_{max}\left(\frac{l}{l_{max}}\right)^{1-CDF(l)} \tag{5}$$

(7) The modified luma component of the input image is determined as:

$$Y'(x,y) = 255 - I'(x,y)_N \tag{6}$$

(8) The modified luma component and chroma components of the image are converted to an RGB image, as the improved image.

2.2.2. Noise-aware image quality metric

A noise-aware image quality metric was used to assess image quality based on the gradient, entropy, and noise of an image (Shin et al., 2019). We introduced three metrics based on gradient, entropy, and noise. The gradient-based metric ($M_{gradient}$) was used to evaluate the edge information.

$$\tilde{g}_i = \frac{1}{N_g}\left[\log \lambda(g_i - \gamma) + 1\right] \tag{7a}$$

$$N_g = \log\left[\lambda(1-\gamma) + 1\right] \tag{7b}$$

$$G_j = \sum_{i \in c_j} \tilde{g}_i, \quad \mathrm{j} = 1,2,\ldots,N_{\mathrm{c}} \tag{7c}$$

$$M_{gradient} = \frac{\mathrm{K}_G E(G)}{S(G)} \tag{7d}$$

where $\tilde{g}_i$ is the amount of gradient information at pixel i; g_i is the gradient magnitude at pixel i, which is estimated using the Sobel operator; λ is the control parameter; γ is the activation threshold; N_g and K_G are normalization factors; G_j is the grid cell; N_c is the number of grid cells; E and S are the mean and standard deviation operators, respectively.

The entropy-based metric was used to estimate the information in an image:

$$M_{entropy} = -k_e \sum_{i=0}^{255} P(i) \log_2 P(i) \tag{8}$$

where $P(i)$ is the probability of intensity level i in a grayscale image; k_e is a normalization factor.

The noise-based metric is defined as follows:

$$M_{noise} = \sum_{j=1}^{3} \sqrt{\frac{\pi}{2}} \frac{1}{N_p} \sum_{i} H(i) U(i) \left| I_j * K \right| (i) \tag{9}$$

where N_p is the number of pixels whose pixel value is 1 in $H.U$, where "." is the elementwise multiplication operator; $H(i)$ is the homogeneous region mask for pixel i; $U(i)$ is the overexposed–underexposed region mask for pixel i; I_j is the pixel value of the input image in channel j; and "$*$" is the convolution operator; K is a noise estimation kernel that was used to evaluate the noise level of an image with Gaussian noise, as defined in Eq. (10); the inhomogeneous, overexposed, and underexposed regions of images were removed using a homogeneous mask and overexposed–underexposed mask defined in Eqs. (11) and (12), since the noise estimation kernel overestimates noises in those regions.

$$K = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \tag{10}$$

$$H(i) = \begin{cases} 1, & g_i \le \delta \\ 0, & g_i > \delta \end{cases} \tag{11}$$

$$U(i) = \begin{cases} 1, & \tau_l \le I_i \le \tau_u \\ 0, & otherwise \end{cases} \tag{12}$$

where δ is the adaptive threshold; τ_l and τ_u are the lower and upper bounds, respectively; and I_i is the pixel value of pixel i. Finally, the image quality metric is derived by combining gradient, entropy, and noise-based metrics:

$$f(I) = AM_{gradient} + BM_{entropy} - CM_{noise} \tag{13}$$

where *A, B,* and *C* are user parameters, which are set to 0.4, 0.6, and 0.4 according to (Shin et al., 2019). A high $f(I)$ implies higher edge and texture details and lower noise. Therefore, it is promising to maximize the quality metric to obtain a well-exposed microscopy image.

2.2.3. Particle swarm optimization

With the noise-aware image quality metric, the particle swarm optimization algorithm was used to optimize the adjusted parameter, which is the only control parameter of the improved adaptive gamma correction method. The adjusted parameter affects the weighting distribution function, elaborated in Eq. (3). The image quality metric, defined in Eq. (13), was used as the objective function of optimization. We used 20 search agents that moved in the search space following Eq. (14a) to seek the optimal solutions in 30 iterations for five independent runs (Kennedy and Eberhart, 1995):

$$\phi_i' = \phi_i + V_i' \tag{14a}$$

$$V_i' = \omega V_i + c_1 r_1 \left(pbest_i - \phi_i \right) + c_2 r_2 \left(gbest - \phi_i \right) \tag{14b}$$

$$\omega = \left(\frac{itermax - iter}{itermax - 1} \right) \left(\omega_{max} - \omega_{min} \right) + \omega_{min} \tag{14c}$$

where ϕ_i' and ϕ_i are the new and current positions of the ith search agent; V_i' is the current velocity of the ith agent; V_i is the previous velocity of the ith agent; $pbest_i$ is the best solution obtained by the ith agent; $gbest$ is the global best solution; ω is the inertia weight; ω_{max} and ω_{min} are the maximum and minimum inertia weights, respectively; c_1 and c_2 are the acceleration coefficients; r_1 and r_2 are random numbers in the range of 0 to 1.

2.3. Dataset standardization and categorization

The images with different resolutions were resized to a standard size (256 × 256 pixels). Next, the resized images were normalized by dividing the pixel intensities in each channel by 255, so that the pixel intensity values were in the range of 0 to 1.

With the different distributions of colors in the images, an unsupervised classification model was developed based the k-means++ clustering with squared Euclidean distance (Arthur and Vassilvitskii, 2006) to categorize the images into different groups according to the chroma channels (C_b and C_r, see Fig. A.2).

2.4. Semantic segmentation

A semantic segmentation model was developed based on an object-contextual representation network with a high-resolution network (Yuan et al., 2020; Wang et al., 2020). The model consisted of multiple convolutional neural networks for pixel-wise classification, and a convolutional neural network was trained for each class. A weak learning method was presented to enable the neural network to learn from new images while preserving the knowledge learned from previous images. Three methods were used for the imbalanced dataset: (1) A multi-class cross-entropy loss function was presented to mitigate the extent of imbalance. (2) An iterative stratification method was proposed to split the highly imbalanced dataset into train and test sets. (3) Image augmentation was performed to artificially enlarge the dataset. These methods are elaborated in the following subsections.

2.4.1. Object-contextual representation

Object-contextual representation network (Yuan et al., 2020) is a recently developed convolutional neural network originally developed for autonomous driving. The neural network was recently used for various practical applications, such as traffic monitoring (He et al., 2022), roof defect detection based on aerial images (Yudin et al., 2021), and pedestrian lane detection for vision-impaired individuals (Lei et al., 2022). This study uses object-contextual representation network in materials science for the first time. The object-contextual representation network has six main steps (Fig. 1):

(1) Feature extraction: The HRNetV2-W18 is used as the feature extractor for the neural network with 20 million parameters and pre-trained using ImageNet for faster convergence (Yuan et al., 2020).

(2) Coarse segmentation: With the extracted features, a coarse segmentation map is generated for each class. The pixel value indicates the class of a pixel. The map is obtained according to the output of the HRNetV2-W18 and a 1 × 1 convolution unit (Yuan et al., 2020).

(3) Object region representation: The object region representation for class k is defined as:

$$f_k = \sum_{i \in I} \overline{m}_{k,i} x_i \tag{15a}$$

$$\overline{m}_{k,i} = softmax \left(m_{k,i} \right) = \frac{e^{m_{k,i}}}{\sum_{j \in I} e^{m_{k,j}}} \tag{15b}$$

where f_k is the object region representation for class k; x_i is the output of the backbone network for pixel i; $\overline{m}_{k,i}$ is the normalized value of $m_{k,i}$; $m_{k,i}$ is the value of pixel i in coarse segmentation map for class k; the softmax function is used to normalize the coarse segmentation maps.

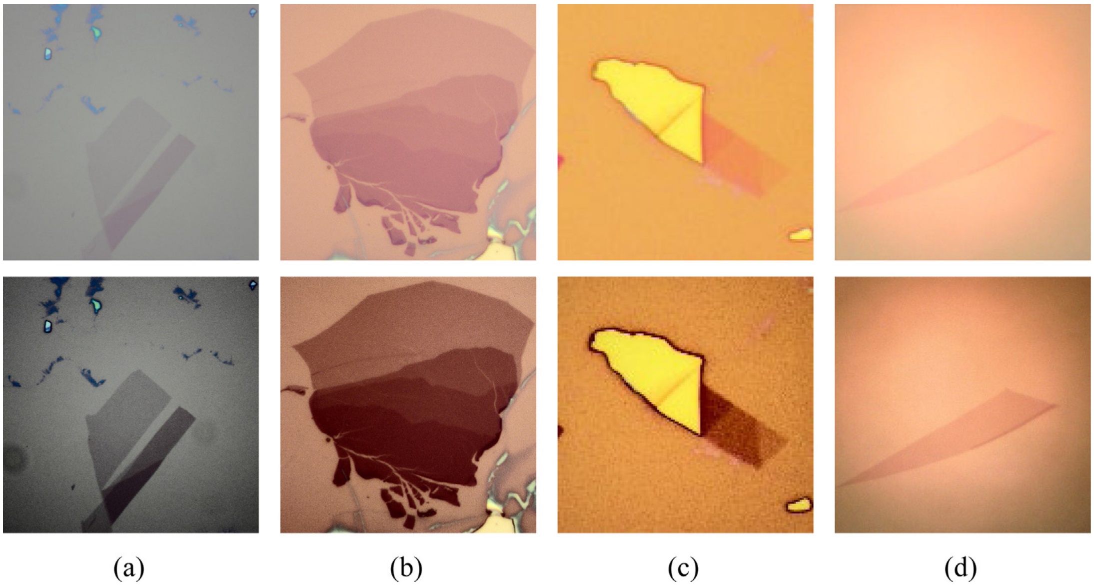


Fig. 3. The effect of optimized adaptive gamma correction on the microscopy images. The upper row are the original images, and the lower row of images are the enhanced images by the optimized adaptive gamma correction.

(4) Pixel-region relation: The relation between pixels with each class is defined as:

$$\omega_{k,i} = \frac{e^{\psi(x_i).\psi(f_k)}}{\sum_{j=1}^{K} \psi\left(x_i\right).\psi(f_j)} \tag{16}$$

where K is the number of classes; ψ is the transformation function consisting of 1×1 convolution, batch normalization, and ReLU activation function.

(5) Object-contextual representation: The object-contextual representation is computed by aggregating the pixel-region relation with all coarse segmentation maps:

$$y_i = \psi\left[\sum_{k=1}^{K} \omega_{k,i}\psi\left(f_k\right)\right] \tag{17}$$

where y_i is the object-contextual representation for pixel i; f_j is the object region representation for class k; K is the number of classes; and $\omega_{k,i}$ is the pixel-region relation for class k and pixel i.

(6) Augmented representation: The final semantic segmentation map is obtained using the transformation function, the output of the backbone network, and object-contextual representation:

$$z_i = \psi([x_i y_i]^T) \tag{18}$$

where z_i is the final segmentation solution for pixel i; x_i is the output of the backbone network for pixel i; and y_i is the object-contextual representation for pixel i.

2.4.2. Weak learning

Weak learning was proposed to address two challenges in training convolutional neural networks: (1) The dataset is limited and becomes smaller when it is divided into different groups. (2) It is time-consuming to repeatedly train the convolutional neural network when new images are added to the dataset. Weak learning was performed in two steps: (1) A convolutional neural network was trained using the entire dataset. (2) Transfer learning was performed.

The neural network was retained for each group of images while the learning rate was near zero. The learning rate is a hyperparameter of a convolutional neural network and controls the changes in the weights in training. Since the learning rate was low, the change of weights was small, so the knowledge learned from previous images was reserved. Meanwhile, the convolutional neural network learned new images and slightly modified the weights to achieve high performance for the new images.

2.4.3. Multi-class cross-entropy loss

Cross-entropy was applied to obtain the discrepancy between the predicted distribution mask $P(x)$ and the true distribution mask $D(x)$:

$$CE(P, D) = \sum_{x\epsilon\Omega} D(x) \cdot \log[P(x)] \tag{19}$$

where Ω is the mask region. The multi-class cross-entropy loss function is expressed as:

$$Loss_C = \sum_{i=1}^{K} \sum_{x\epsilon\Omega} d_{x,i} \log(p_{x,i}) \tag{20}$$

where K is the number of classes; $d_{x,i}$ is the true one-hot distribution probability of pixel x on class i; and $p_{x,i}$ is the predicted probability for pixel x and class i.

Since the annotated pixels of under-represented classes were scarce, the effect of the under-represented classes on the loss function was limited. Thus, the neural network tends to maximize the prediction accuracy for over-represented classes while ignoring the under-presented classes. This research assigned sample weights to mitigate the under-representation problem:

$$Loss_W = \sum_{i=1}^{K} \sum_{x\epsilon\Omega} w_i d_{x,i} \log(p_{x,i}) \tag{21}$$

where w_i is the sample weight for class i.

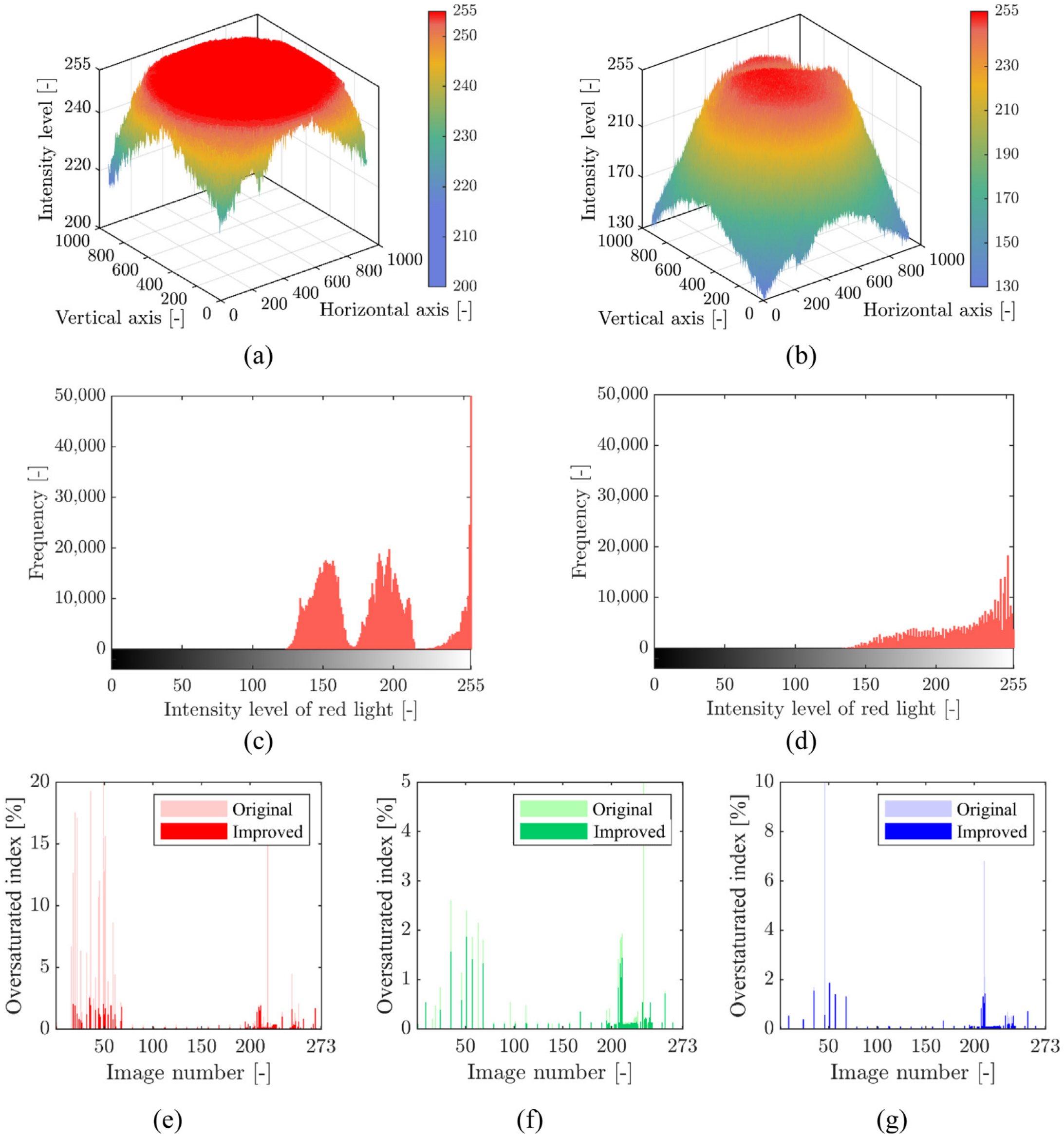
There was no definitive advice on how to set the sample weights. We proposed to calculate the sample weights using Eq. (22):

$$w_i = \left(\frac{1}{\mu_i}\right)^{\beta} \tag{22}$$

where μ_i is the class weight for class i; and β is an adjustment factor. The adjustment factor was optimized by the particle swarm optimization to minimize the image quality metric.

2.4.4. Iterative stratification

When the dataset is small and highly imbalanced, it is inappropriate to randomly split the dataset into train and test sets, because under-represented classes may be missing in one of the sets and thus the trained model fails to represent the whole dataset. We proposed to employ an iterative stratification method for multi-label data (Sechidis et al., 2011). The whole dataset was divided into more than two subsets while the class weights of the sets were almost the same.



(a) (b) (c) (d) (e) (f) (g)

Fig. 4. An investigation on an overexposed image: (a) the original image, (b) the improved image, (c) the pixel intensity histogram of the red channel of the original image, (d) the pixel intensity histogram of the red channel of the improved image, (e) oversaturated index of the red channel, (f) oversaturated index of the green channel, and (g) oversaturated index of the blue channel. In (a) and (b), pixels with intensity levels over 253 represent oversaturation and are shown in red color.

Table 2
Results of the performance metrics.

Methods	Pixel accuracy	Mean accuracy	F1	mIoU	Precision	Recall
The existing methods						
UNet++	96.7%	61.7%	56.0	56.4	60.9	51.9
PSPNet	95.5%	64.6%	56.0	56.5	59.5	52.8
DeepLabv3+	96.9%	70.9%	62.8	58.5	66.7	59.3
The proposed methods						
Baseline (HRNet+OCR)	97.1%	71.7%	65.9	59.0	67.9	63.9
Baseline+Loss $_{w}$	98.3%	73.6%	72.9	61.5	77.2	69.1
Baseline+Loss $_{w}$+OAGC	98.8%	75.2%	87.2	66.7	89.7	85.1
Baseline+Loss $_{w}$+OAGC+WL	99.0%	81.6%	91.8	71.7	93.5	90.2

Note: “Loss $_{w}$” is weighted cross-entropy loss; “OAGC” is optimized adaptive gamma correction; and “WL” is weak learning.

2.4.5. Image augmentation

Augmentation was performed to increase the dataset size and improve the generalizability of semantic segmentation models (Shorten and Khoshgoftaar, 2019). The training dataset was enlarged using four strategies: (1) Resizing: Images were resized from 256 × 256 pixels to 320 × 256 pixels. (2) Random cropping: Images were randomly cropped to 256 × 256 pixels. (3) Random flipping: Images were randomly flipped horizontally and vertically with a probability of 50%. (4)

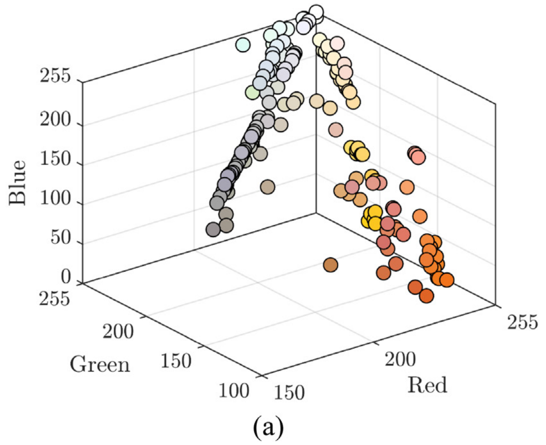


Fig. A.1. Scatter plots of pixel intensity levels: (a) the pixel intensity of the background, and (b) the average pixel intensity of images. Each dot represents a microscope image. The color of dot represents the calculated color value.

Photometric distortion: brightness, contrast, saturation, and hue were randomly modified with a probability of 50%.

2.5. Performance metrics

Six statistical performance metrics were used to evaluate the semantic segmentation model:

$$Pixel\ accuracy = \frac{\sum_{k=1}^{K}(TP_k + TN_k)}{\sum_{k=1}^{K}(TP_k + FP_k + TN_k + FN_k)} \tag{23a}$$

$$Mean\ accuracy = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TP_k + TN_k}{TP_k + FP_k + TN_k + FN_k}\right) \tag{23b}$$

$$mIoU = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TP_k}{TP_k + FP_k + FN_k}\right) \tag{23c}$$

$$Precision = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TP_k}{TP_k + FP_k}\right) \tag{23d}$$

$$Recall = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TP_k}{TP_k + FN_k}\right) \tag{23e}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{23f}$$

where TP_k and TN_k are respectively the true positives and true negatives corresponding to class k; FP_k and FN_k are respectively the false positives and false negatives corresponding to class k; K is the number of classes; and $mIoU$ is mean intersection-over-union.

3. Results

Four pairs of representative images before and after applying the optimized adaptive gamma correction are compared in Fig. 3. The comparison shows that the optimized adaptive gamma correction method improves the visibility of the microscope images of graphene regardless of the background color and brightness. The texture and edges of the improved images become easier to distinguish from the background. The optimal adjusted parameter obtained by the particle swarm optimization to minimize the image quality metric was 0.561. The lowest image quality metric value was 1.48.

The pixel intensity levels of the fourth pair of images are compared in Fig. 4. The comparison shows that the oversaturation and oversaturated regions are greatly reduced by impeding the optimized adaptive gamma correction. An index defined by Eq. (24) is proposed to indicate the effect of the optimized adaptive gamma correction for improvement of the images:

$$\delta = \frac{A_o}{A_T} \tag{24}$$

where δ is the oversaturated index; A_o is the number of pixels with an intensity level over 253; and A_T is the total number of pixels. The oversaturated indices of original and modified images for the red, green, and blue channels are highly reduced as shown in Fig. 4(e) to 4(g).

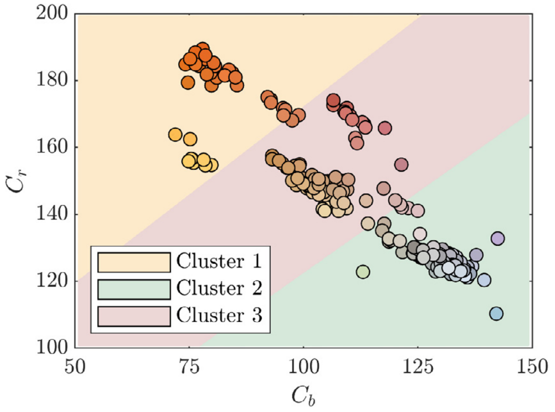


Fig. A.2. The geography of groups in the C_b and C_r space. Each dot represents a microscope image. The color of each dot represents the mean pixel intensity. The microscope images are grouped by k-means clustering based on the blue-difference and red-difference chroma components.

The proposed method was compared with three state-of-the-art convolutional neural networks, which are UNet++ (Zhou et al., 2018), PSPNet (Zhao et al., 2017), and DeepLabv3+ (Chen et al., 2018). The proposed models were trained by stochastic gradient descent. The learning rate was 0.1. The momentum was 0.9. The weight decay was 0.0005. The batch size was eight. The maximum number of iterations was 10,000. Conventional cross-entropy loss was used as the loss function. The deep learning models were trained using the labeled images. That is, the microscopy images are the inputs while the thickness map of exfoliated graphene flakes are the outputs. The performance metrics of the different methods are compared in Table 2. The proposed method (Baseline+Lossw+OAGC+WL) shows the highest accuracy.

Four HRNet+OCR models established based on the presented methods were compared to test the effects of the weighted cross-entropy

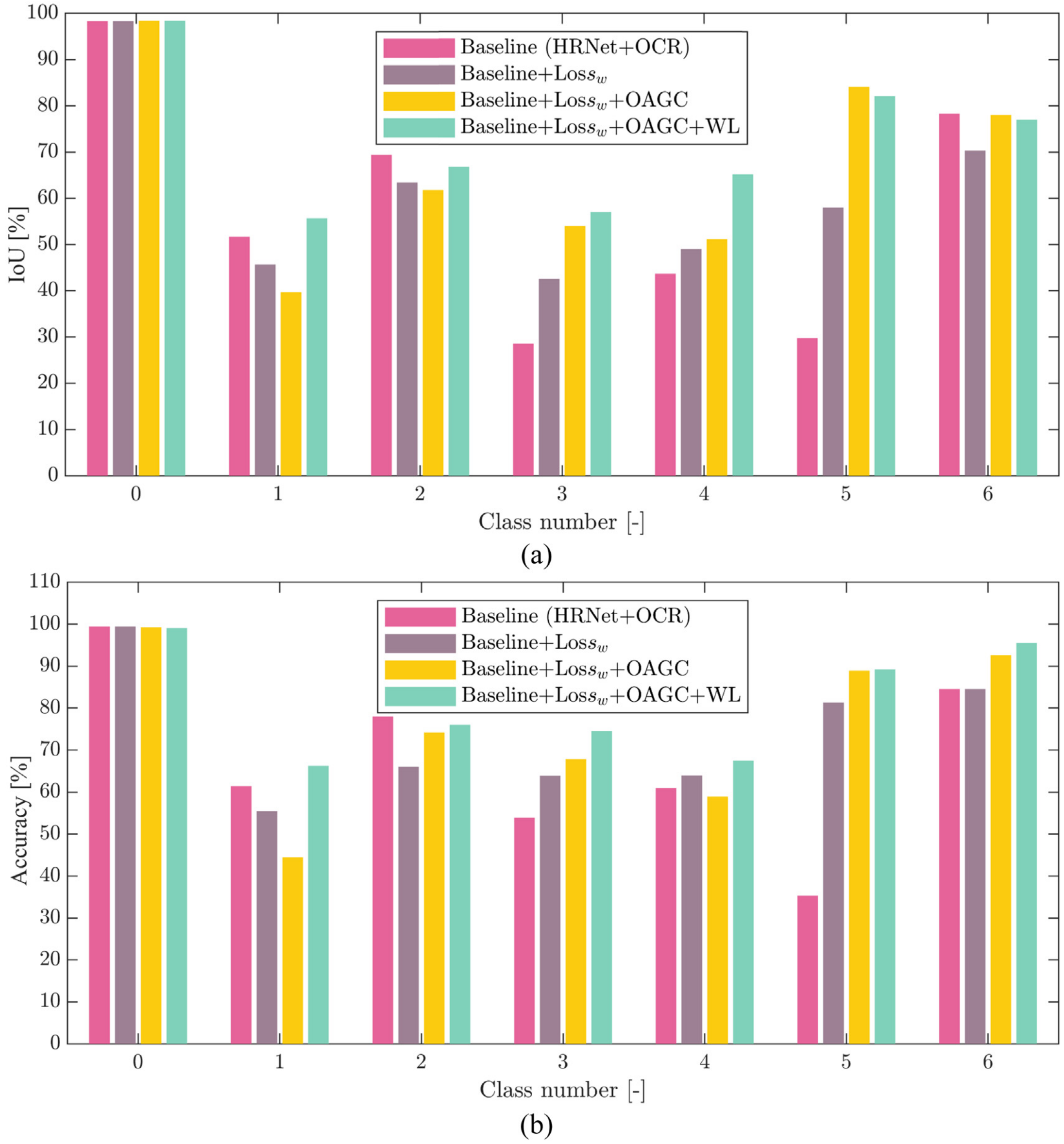


Fig. A.3. The class-wise performance of the applied segmentation methods: (a) Mean IoU (mIoU), (b) accuracy. Loss$_w$, OAGC, WL stand for the weighted cross-entropy loss, optimized adaptive gamma correction, and weak learning, respectively.

loss, optimized adaptive gamma correction, and weak learning on the accuracy. The comparison results show that the accuracy is significantly improved by applying the weighted cross-entropy loss, optimized adaptive gamma correction, and weak learning in the final model (HRNet+OCR+Loss$_w$+OAGC+WL). The pixel accuracy of the final model is higher than 99%. The class-wise intersection-over-union (IoU) and accuracy of the four HRNet+OCR models are shown in Fig. A.3. The IoU and accuracy of the highly under-represented classes 1L, 3L, and 6–10L are significantly increased by the weighted cross-entropy loss, optimized adaptive gamma correction, and weak learning. The minimum IoU of the final model is higher than 56%. Fig. A.4 shows representative optical microscope images in the test set and the corresponding prediction results.

4. Conclusions

This study develops a deep learning method for automatic identification and classification of exfoliated graphene flakes from optical microscope images. Novel methods are presented to improve the quality of optical microscope images and address important challenges associated with the imbalanced classes, limited dataset size, different background conditions, and different resolutions of the optical microscope images. Based on the above investigations, the following findings are summarized:

- The presented machine learning framework based on object-contextual representations is promising for automatically identifying 2D material flakes and classifying the thickness of the flakes based on microscope images. The pixel accuracy of the trained model is higher than 99%, outperforming the state-of-the-art deep convolutional neural networks DeepLabv3+, PSPNet, and UNet ++.
- The proposed method is robust to the background color of microscope images because the proposed machine learning method can automatically identify the background color of the images and categorize the images into appropriate groups. The proposed method is also robust to the brightness and resolution of images because the optimized adaptive gamma correction method can effectively improve the quality of overexposed microscopy images.

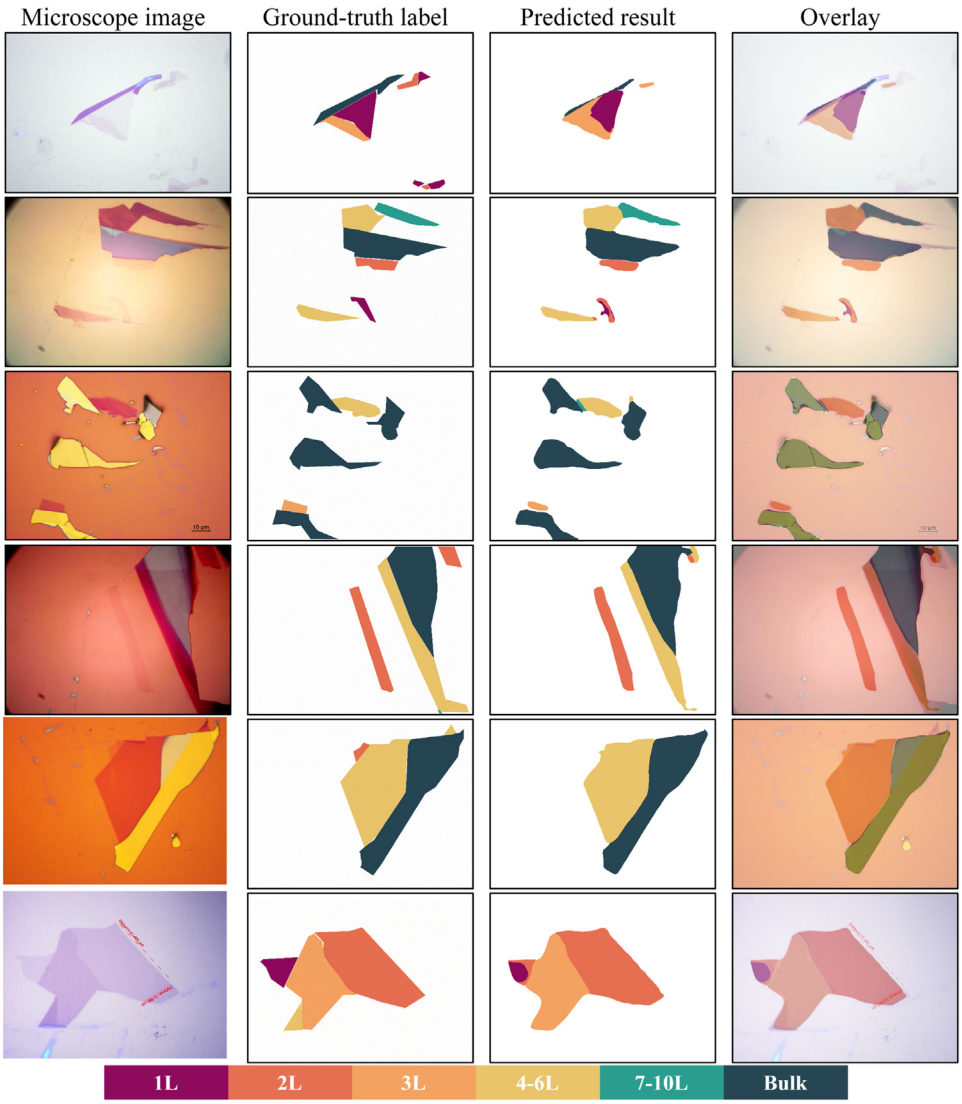


Fig. A.4. Representative optical microscope images and the predicted results using HRNet+OCR.

- The presented machine learning framework can learn characterization from new images while preserving the knowledge learned from old images. The weak learning method enables us to retrain convolutional neural networks while preserving the learned knowledge by the network. The method is useful especially when the training set is small but there is a convolutional neural network trained on a similar task.
- The presented method can handle imbalanced datasets. The IoU of different classes is in the range of 57% and 99%, and the mIoU of all the classes is 59%. This study shows that image augmentation, iterative stratification, and weighted cross-entropy loss can significantly enhance the predictive performance of convolutional neural networks on under-represented classes.

Further research can be conducted from the following aspects:

- It is promising to test the applicability of the proposed hierarchical deep learning model to other 2D materials through transfer learning, which will address limited data issues and accelerate the training process of new predictive models for classification of 2D materials.
- Although the proposed model shows higher accuracy in identifying exfoliated graphene flakes compared with the competing models, it is unclear whether the method can be used under other microscopy conditions. It is necessary to examine the proposed method using other datasets to evaluate the generalization performance in future research.
- One of the challenges in performing microscopy imaging experiments is to configure the microscope based on room setup and lighting conditions to obtain reasonable resolution. It is desirable to develop a machine vision method based on the proposed noise-aware image quality metric and reinforcement learning for automated microscopy configuration, aiming to maximize the spatial resolution of microscopy images as well as distinction between various 2D flakes.

CRediT authorship contribution statement

Soroush Mahjoubi: Coding, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. **Fan Ye:** Raw image annotation, Writing – review & editing. **Yi Bao:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing. **Weina Meng:** Funding acquisition, Methodology, Resources, Writing – review & editing. **Xian Zhang:** Raw image generation, Writing review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Weina Meng received financial supports from the National Science Foundation of the US.

Data availability

Data will be made available on request.

Acknowledgment

This research was funded by the National Science Foundation, USA [grant No. CMMI-2046407].

Appendix

See Figs. A.1–A.4.

References

Allugunti, V.R., 2022. A machine learning model for skin disease classification using convolution neural network. Int. J. Comput. Program. Database Manag. 3, 141–147.

Arthur, D., Vassilvitskii, S., 2006. K-Means++: The Advantages of Careful Seeding. Stanford.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A deep convolutional encoder–decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495.

Blake, P., Hill, E.W., Castro Neto, A.H., Novoselov, K.S., Jiang, D., Yang, R., Booth, T.J., Geim, A.K., 2007. Making graphene visible. Appl. Phys. Lett. 91, 063124.

Cao, G., Huang, L., Tian, H., Huang, X., Wang, Y., Zhi, R., 2018. Contrast enhancement of brightness-distorted images by improved adaptive gamma correction. Comput. Electr. Eng. 66, 569–582.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T., 2018a. The rise of deep learning in drug discovery. Drug Discov. Today 23, 1241–1250.

Chen, W., Mao, X., Ma, H., 2010. Low-contrast microscopic image enhancement based on multi-technology fusion. In: IEEE International Conference on Intelligent Computing and Intelligent Systems. pp. 891-895.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018b. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder–decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801-818.

Chong, E., Clark-Whitney, E., Southerland, A., Stubbs, E., Miller, C., Ajodan, E.L., Silverman, M.R., Lord, C., Rozga, A., Jones, R.M., Rehg, J.M., 2020. Detection of eye contact with deep neural networks is as accurate as human experts. Nature Commun. 11, 6386.

Greplova, E., Gold, C., Kratochwil, B., Davatz, T., Pisoni, R., Kurzmann, A., Rickhaus, P., Fischer, M.H., Ihn, T., Huber, S.D., 2020. Fully automated identification of two-dimensional material samples. Phys. Rev. A 13, 064017.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recognit. 77, 354–377.

Han, B., Lin, Y., Yang, Y., Mao, N., Li, W., Wang, H., Yasuda, K., Wang, X., Fatemi, V., Zhou, L., Wang, J.I., Ma, Q., Cao, Y., Rodan-Legrain, D., Bie, Y.-Q., Navarro-Moratalla, E., Klein, D., MacNeill, D., Wu, S., Kitadai, H., Ling, X., Jarillo-Herrero, P., Kong, J., Yin, J., Palacios, T., 2020. Deep-learning-enabled fast optical identification and characterization of 2D materials. Adv. Mater. 32, 2000953.

Hassaballah, M., Awad, A.I., 2020. Deep Learning in Computer Vision: Principles and Applications. CRC Press.

Hazirbas, C., Ma, L., Domokos, C., Cremers, D., 2017. FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture. In: Lai, S.-H., Lepetit, V., Nishino, K., Sato, Y. (Eds.), Computer Vision – ACCV 2016. Springer International Publishing, pp. 213–228.

He, P., Jiao, L., Shang, R., Wang, S., Liu, X., Quan, D., Yang, K., Zhao, D., 2022. MANet: Multi-scale aware-relation network for semantic segmentation in aerial scenes. In: IEEE Trans. Geosci. Remote Sens..

Huang, B., Clark, G., Navarro-Moratalla, E., Klein, D.R., Cheng, R., Seyler, K.L., Zhong, D., Schmidgall, E., McGuire, M.A., Cobden, D.H., Yao, W., Xiao, D., Jarillo-Herrero, P., Xu, X., 2017. Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit. Nature 546, 270–273.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of ICNN'95-International Conference on Neural Networks, pp. 1942-1948.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., 2017. Overcoming catastrophic forgetting in neural networks. Proc. Natl. Acad. Sci. 114, 3521–3526.

Lei, Y., Phung, S., Bouzerdoum, A., Le, H.T., Luu, K., 2022. Pedestrian lane detection for assistive navigation of vision-impaired people: Survey and experimental evaluation. IEEE Access 10, 101071-101089.

Leong, F.W., Brady, M., McGee, J.O., 2003. Correction of uneven illumination (vignetting) in digital microscopy images. J. Clin. Pathol. 56, 619–621.

Li, Y., Chernikov, A., Zhang, X., Rigosi, A., Hill, H.M., van der Zande, A.M., Chenet, D.A., Shih, E.-M., Hone, J., Heinz, T.F., 2014. Measurement of the optical dielectric function of monolayer transition-metal dichalcogenides: MoS2, MoSe2, WS2, and WSe2. Phys. Rev. B 90, 205422.

Li, H., Wu, J., Huang, X., Lu, G., Yang, J., Lu, X., Xiong, Q., Zhang, H., 2013. Rapid and reliable thickness identification of two-dimensional nanosheets using optical microscopy. ACS Nano 7, 10344–10353.

Lin, X., Si, Z., Fu, W., Yang, J., Guo, S., Cao, Y., Zhang, J., Wang, X., Liu, P., Jiang, K., Zhao, W., 2018. Intelligent identification of two-dimensional nanostructures by machine-learning optical microscopy. Nano Res. 11, 6316–6324.

Mahjoubi, S., Barhemat, R., Meng, W., Bao, Y., 2022. Deep Learning from Physicochemical Information of Concrete with an Artificial Language for Property Prediction and Reaction Discovery. SSRN, http://dx.doi.org/10.2139/ssrn.4143089.

Mahjoubi, S., Barhemat, R., Meng, W., Bao, Y., 2023. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). Resour. Conserv. Recy. 189, 106741.

Masubuchi, S., Machida, T., 2019. Classifying optical microscope images of exfoliated graphene flakes by data-driven machine learning. Npj 2D Mater. Appl. 3, 4.

Masubuchi, S., Morimoto, M., Morikawa, S., Onodera, M., Asakawa, Y., Watanabe, K., Taniguchi, T., Machida, T., 2018. Autonomous robotic searching and assembly of two-dimensional crystals to build van der Waals superlattices. Nature Commun. 9, 1413.

Masubuchi, S., Watanabe, E., Seo, Y., Okazaki, S., Sasagawa, T., Watanabe, K., Taniguchi, T., Machida, T., 2020. Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials. NPJ 2D Mater. Appl. 4, 3.

Ni, Z.H., Wang, H.M., Kasim, J., Fan, H.M., Yu, T., Wu, Y.H., Feng, Y.P., Shen, Z.X., 2007. Graphene thickness determination using reflection and contrast spectroscopy. Nano Lett. 7, 2758–2763.

Nolen, C.M., Denina, G., Teweldebrhan, D., Bhanu, B., Balandin, A.A., 2011. High-throughput large-area automated identification and quality control of graphene and few-layer graphene films. ACS Nano 5, 914–922.

Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., Firsov, A.A., 2004. Electric field effect in atomically thin carbon films. Science 306, 666.

Novoselov, K.S., Jiang, D., Schedin, F., Booth, T.J., Khotkevich, V.V., Morozov, S.V., Geim, A.K., 2005. Two-dimensional atomic crystals. Proc. Natl. Acad. Sci. 102, 10451.

Otter, D.W., Medina, J.R., Kalita, J.K., 2020. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. 32, 604–624.

Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V., Kis, A., 2011. Single-layer MoS2 transistors. Nature Nanotechnol. 6, 147–150.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems 28 (NIPS 2015).

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, pp. 234–241.

Saito, Y., Shin, K., Terayama, K., Desai, S., Onga, M., Nakagawa, Y., Itahashi, Y.M., Iwasa, Y., Yamada, M., Tsuda, K., 2019. Deep-learning-based quality filtering of mechanically exfoliated 2D crystals. NPJ Comput. Mater. 5, 124.

Sechidis, K., Tsoumakas, G., Vlahavas, I., 2011. On the stratification of multi-label data. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 145-158.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 640–651.

Shin, U., Park, J., Shim, G., Rameau, F., Kweon, I., 2019. Camera exposure control for robust robot vision with noise-aware image quality assessment. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1165-1172.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6, 60.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., 2020. Deep high-resolution representation learning for visual recognition. In: IEEE Trans. Pattern Anal. Mach. Intell.. vol. 43, (10), pp. 3349–3364.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., Xu, W., 2016. CNN-RNN: A unified framework for multi-label image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2285-2294.

Wu, Z., Gao, Y., Li, L., Xue, J., Li, Y., 2019. Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. Connect. Sci. 31, 169–184.

Xi, X., Wang, Z., Zhao, W., Park, J.-H., Law, K.T., Berger, H., Forró, L., Shan, J., Mak, K.F., 2016. Ising pairing in superconducting NbSe2 atomic layers. Nat. Phys. 12, 139–143.

Xu, X., Yao, W., Xiao, D., Heinz, T.F., 2014. Spin and pseudospins in layered transition metal dichalcogenides. Nat. Phys. 10, 343–350.

Yuan, Y., Chen, X., Wang, J., 2020. Object-contextual representations for semantic segmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), Computer Vision – ECCV 2020. pp. 173–190.

Yudin, D.A., Adeshkin, V., Dolzhenko, A.V., Polyakov, A., Naumov, A.E., 2021. Roof defect segmentation on aerial images using neural networks. In: Kryzhanovsky, B., Dunin-Barkowski, W., Redko, V., Tiumentsev, Y. (Eds.), Advances in Neural Computation, Machine Learning, and Cognitive Research IV. pp. 175–183.

Zhang, X., Sun, D., Li, Y., Lee, G.-H., Cui, X., Chenet, D., You, Y., Heinz, T.F., Hone, J., 2015. Measurement of lateral and interfacial thermal conductivity of single- and bilayer MoS_2 and $MoSe_2$ using optothermal Raman technique. ACS Appl. Mater. Interfaces 7, 25923–25929.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881-2890.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested U-net architecture for medical image segmentation. In: Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R.S., Bradley, A., Papa, J.P., Belagiannis, V., Nascimento, J.C., Lu, Z., Conjeti, S., Moradi, M., Greenspan, H., Madabhushi, A. (Eds.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 3–11.