




OPEN

# Automatic detection of multilayer hexagonal boron nitride in optical images using deep learning-based computer vision

Fereshteh Ramezani<sup>1</sup>, Sheikh Parvez<sup>2,3</sup>, J. Pierce Fix<sup>2,3</sup>, Arthur Battaglin<sup>2</sup>, Seamus Whyte<sup>2</sup>, Nicholas J. Borys<sup>2,3,4</sup> & Bradley M. Whitaker<sup>1,4</sup>

Computer vision algorithms can quickly analyze numerous images and identify useful information with high accuracy. Recently, computer vision has been used to identify 2D materials in microscope images. 2D materials have important fundamental properties allowing for their use in many potential applications, including many in quantum information science and engineering. One such material is hexagonal boron nitride (hBN), an isomorph of graphene with a very indistinguishable layered structure. In order to use these materials for research and product development, the most effective method is mechanical exfoliation where single-layer 2D crystallites must be prepared through an exfoliation procedure and then identified using reflected light optical microscopy. Performing these searches manually is a time-consuming and tedious task. Deploying deep learning-based computer vision algorithms for 2D material search can automate the flake detection task with minimal need for human intervention. In this work, we have implemented a new deep learning pipeline to classify crystallites of hBN based on coarse thickness classifications in reflected-light optical micrographs. We have used DetectoRS as the object detector and trained it on 177 images containing hexagonal boron nitride (hBN) flakes of varying thickness. The trained model achieved a high detection accuracy for the rare category of thin flakes (< 50 atomic layers thick). Further analysis shows that our proposed pipeline could be generalized to various microscope settings and is robust against changes in color or substrate background.

**Computer vision.** Object detection is an important computer vision task that deals with detecting instances of visual objects of certain categories in digital images. The goal of object detection is to develop computational models and techniques that provide one of the most basic pieces of information needed by computer vision applications<sup>1</sup>. In other words, the goal of object detection is to determine whether there are any instances of objects from given categories in an image and, if present, to return the spatial location and extent of each object instance<sup>2</sup>. Object detection supports a wide range of applications, including robot vision, consumer markets, autonomous driving, human computer interaction, content based image retrieval, intelligent video surveillance, and augmented reality<sup>2,3</sup>.

Rapid progressions in deep learning and improvements in device capabilities including computing power, memory capacity, power consumption, image sensor resolution, and optics have improved the performance and cost-effectiveness and assisted the spread of vision-based applications. Compared to traditional computer vision techniques, which have a long trial-and-error process, deep learning enables end-to-end object detectors to achieve greater accuracy in tasks such as image classification, semantic segmentation, object detection and Simultaneous Localization and Mapping<sup>4</sup>.

At present, object detection based on deep learning (DL) frameworks fall into two main categories; two-stage detectors, such as the Region-based Convolutional Neural Network (R-CNN)<sup>5</sup> and its variants<sup>6–8</sup>, and one-stage detectors, such as You Only Look Once (YOLO)<sup>9</sup> and its variants<sup>10–13</sup>. On public benchmarks, two-stage detectors

<sup>1</sup>Electrical and Computer Engineering Department, Montana State University, Bozeman, USA. <sup>2</sup>Department of Physics, Montana State University, Bozeman, USA. <sup>3</sup>Materials Science Program, Montana State University, Bozeman, USA. <sup>4</sup>Optical Technology Center, Montana State University, Bozeman, USA. ✉email: fereshteh.ramezani@student.montana.edu

typically achieve higher accuracy, whereas one-stage detectors are significantly more time-efficient and more suited to real-time applications<sup>14</sup>.

**Two-dimensional materials.** Single and few-layers of two-dimensional (2D) materials provide many opportunities to explore quantum phenomena in systems with appealing features such as strong many-body interactions, pristine interfaces, and strong confinement in a single direction<sup>15–18</sup>. In bulk, 2D materials form van der Waals crystals where strong in-plane bonding within a single layer is complemented by significantly weaker van der Waals bonding between the layers<sup>19</sup>. The anisotropy in bonding strengths combined with a chemical structure where the bonds within a single layer are completely passivated enables individual layers of the 2D material to be isolated, forming chemically and structurally robust sheets of atoms that are atomically thick (i.e., 1–3 atomic layers thick). Although ultrathin, these 2D crystallites can have lateral sizes that are 10–100  $\mu\text{m}$  – 10,000–100,000  $\times$  larger than their thickness.

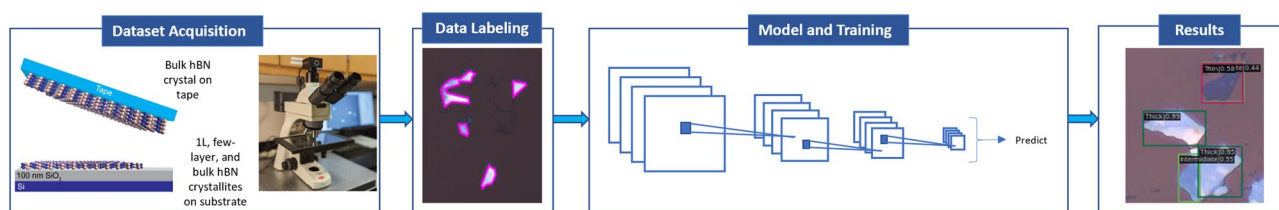
In single- and few-layer form, many 2D materials exhibit excellent optical, electronic, and/or magnetic phenomena that provide means to prepare, interact, and study quantum states in matter<sup>20–24</sup>. Different Van der Waals materials have been used to demonstrate appealing device applications such as emitting diodes. One such material is hexagonal boron nitride (hBN), an isomorph of graphene with a very indistinguishable layered structure, that holds interesting opto-electrical properties merged with mechanical robustness, thermal stability and chemical inertness<sup>25</sup>. With over 1000 known van der Waals 2D materials that span all functionalities (i.e., insulators, metals, semiconductors, ferromagnets, ferroelectrics, etc.), a prolific parameter space of systems and structure-property relationships for quantum, optoelectronic, and magnetic technologies exists<sup>26–28</sup>. The task of exploring the vast world of 2D materials requires the ability to efficiently and reliably produce single- and few-layer samples of 2D materials<sup>29,30</sup>. However, the most widely-used state-of-the-art method for preparing single- and few-layer samples of 2D materials relies on the manual exfoliation of a large population of smaller crystallites from a bulk crystal. The resulting crystallites have a range of thicknesses, and the entire population must be manually searched using optical reflected light microscopy to identify those that have a desired thickness, which is most often the thickness of a single layer. This labor-intensive method for collecting single- and few-layer samples strongly inhibits rapid discovery and exploration of these material systems.

**Automated 2D material detection.** Deep learning computer vision methods identify and analyze numerous images into useful information at a fast rate with a high accuracy. Deploying DL algorithms for 2D material search can automate the flake detection task with minimal need for human intervention. So far, there have been various attempts taken toward flake identification in images<sup>31–38</sup>. However, these methods failed to detect flakes within different microscope settings and image magnification unless the algorithms were retrained using the new settings. In this work, we propose a new detection pipeline which not only makes improvements in detection accuracy, but also is able to detect the rare class of thin flakes ( $< 50$  atomic layers) with a recall of 0.43 and precision of 0.43. We also applied our proposed pipeline to various microscope settings and showed that the algorithm is robust against changes in color settings and magnification.

## Methods

Our proposed deep learning method to detect hBN flakes in 2D material microscopic images is shown as a pipeline in Fig 1. This pipeline consists of three main steps which will be explained in detail.

**Data acquisition.** To prepare the dataset, 208 images were collected over 2 months. To do this, hexagonal boron nitride (hBN) samples were fabricated via mechanical exfoliation. Bulk hBN crystals were placed onto a piece of exfoliation tape and separated into thin layers by folding the tape together and peeling the tape apart five times. The exfoliation tape was then placed onto a  $1 \times 1 \text{ cm}^2$  silicon (Si) wafer with a 100 nm silicon oxide ( $\text{SiO}_2$ ) layer. To ensure the  $1 \times 1 \text{ cm}^2$  region of the tape was in contact with the Si/ $\text{SiO}_2$ , the tape was pressed down with a cotton swab across the full area of contact. The sample, Si/ $\text{SiO}_2$ /exfoliation tape, was then heated for 1–2 min at 90 °C and allowed to cool to room temperature, after which the exfoliation tape was removed. The hBN sample images were taken with a 20X objective using a Motic BA310MET-T Incident and Transmitted Metallurgical Microscope. The hBN images for the dataset were taken with the same camera settings and light intensity.



**Figure 1.** Proposed deep learning pipeline for 2D flake detection. First, a dataset of hBN samples are collected through mechanical exfoliation. Then as the second step, the dataset is annotated manually using an online annotation tool. During the third step, DetectoRS is chosen as the training algorithm, and for the last step, the trained model is applied to the test images to detect hBN samples.

A second dataset of hBN images was collected by our colleagues at the University of Arkansas (UA) following a similar procedure. The UA dataset consists of 10 images that we use exclusively for testing the data. Finally, we use a third dataset for testing purposes: 10 hBN images collected by Masubuchi et al.<sup>32</sup>.

**Data labeling.** Labeling the data is the preliminary step in a supervised learning object detection task, and the results of the model heavily depend on the accuracy of the annotations. To annotate our data, we used Roboflow (<https://roboflow.com>) which is an online annotation tool. For better accuracy of training, we annotated the flakes within each individual image manually and in three different categories as; Thick, Intermediate, and Thin. Table 1 shows the thickness and number of layers for each class. The table also indicates the number of images and the number of labeled instances of each flake type present in the full dataset.

In addition to annotating our images, we annotated the 10 test images in the UA dataset and re-annotated the 10 test images in the Masubuchi dataset to ensure uniformity in flake labels.

The annotations were then saved in the Microsoft common objects in context (COCO) JSON format. To train and evaluate the machine learning algorithm, the dataset was split into three subsets: training set, validation set and testing set. The training set consists of 177 images which contain 81% of the annotations; the validation set consists of 21 images which contain 14% of the annotations; and the testing set consists 10 images which contain 5% of the annotations.

**Model and training.** To detect the 2D objects, DetectoRS<sup>39</sup> was chosen as the training algorithm. DetectoRS is a multi-stage supervised object detection algorithm. This algorithm involves two major components: Recursive Feature Pyramid (RFP) and Switchable Atrous Convolution (SAC). RFP is employed at the macro level that incorporates extra feedback connections from Feature Pyramid Networks into the bottom-up backbone layers. At the micro level, SAC is utilized, which convolves the features with different atrous rates and gathers the results using switch functions. Combining them results in DetectoRS, which significantly improves the performance of object detection. DetectoRS achieves state-of-the-art performance on the COCO test-dev dataset. More information concerning this architecture can be found in the original paper<sup>39</sup>.

The performance of supervised machine learning models depends on the quantity of the dataset to avoid overfitting. However, most of the time collecting and labelling the data is a tedious time-consuming process. Data augmentation techniques are a solution to address this problem. The augmented data will represent a more comprehensive set of possible data points, thus minimizing the distance between the training and validation set, as well as any future testing sets<sup>40</sup>. To augment the data, we used various techniques such as rotating, flipping and color contrast twice; once within Roboflow before exporting the annotation and another time during the training procedure.

To improve the performance of the model, we employed transfer learning. Fine-tuning pre-trained deep networks is a practical way of benefiting from the representation learned on a large database while having relatively few examples to train a model<sup>41</sup>. We took the model which was previously trained on the COCO dataset<sup>42</sup> and used it as a starting point to retrain it on our own dataset. For training on 2D material images, we chose DetectoRS algorithm with Cascade+ResNet-50 as the backbone detector architecture implemented in MMDetection library<sup>43</sup>. The model was trained with the learning algorithm rate of 0.0025, using Cuda and for 20 epochs.

## Results

The inference results of the trained model when applied to one of the test images is shown in Fig 2. Each flake identification in the image consists of three components; a bounding box, a class label, and a confidence number or probability score which shows the certainty of the algorithm that the class is detected correctly. As can be seen, the algorithm was able to detect the Thin sample, which is the most important among the three classes as it possibly contains a true monolayer material.

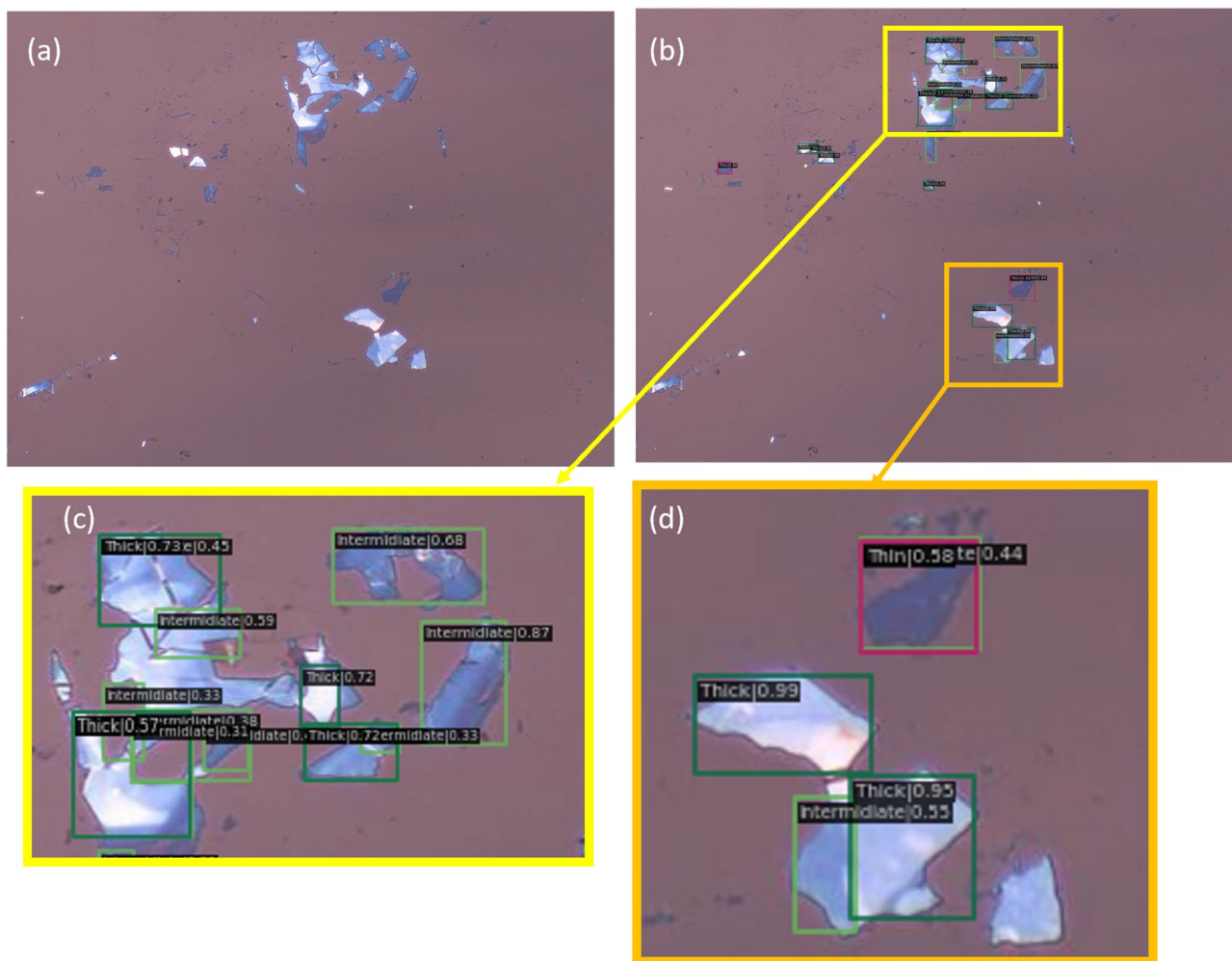
To measure the accuracy of the object detector across all test images, we used various evaluation metrics. Before examining these metrics, we present some basic concepts in object detection<sup>44</sup>:

- *True Positive (TP)* A correct detection of a ground-truth bounding box
- *False Positive (FP)* A detection of a nonexistent object or a misplaced detection of an existing object
- *False Negative (FN)* An undetected ground-truth bounding box

A correct or incorrect detection is determined based on Intersection over Union (IoU) which is a threshold. IoU measures how much of a projected object's bounding box overlaps with the bounding box surrounding the ground reference data.

Label	Thickness (nm)	Number of atomic layers	Number of annotations		
			Train	Validation	Test
Thick	≥ 50	≥ 100	996	240	73
Intermediate	25–50	50–100	465	14	12
Thin	≤ 25	≤ 50	140	23	7

**Table 1.** Dataset Labels and annotations.



**Figure 2.** Results obtained from using the proposed algorithm on one image from our dataset; Part (a) shows the raw image, part (b) shows the labels generated by the algorithm, and parts (c) and (d) are zoomed-in regions showing the flake type and confidence score reported by the algorithm. We manually identified 19 flakes in the image, and the algorithm correctly classified 14 of them.

It is also good to notice that a true negative result is when the algorithm detects the background correctly. Since labeled boundary boxes are only applied to objects, a true negative is not reported in the context of object detection.

Two common metrics to measure the accuracy of the object detection performance are precision and recall which are defined based on the TP, FP and FN: Precision =  $TP/(TP+FP)$  and Recall =  $TP/(TP+FN)$ .

In words, precision returns the ratio of correctly identified objects to all identified objects. Recall measures the proportion of the ground-truth objects that were identified by the algorithm. Mean Average Precision (mAP) is a measure that combines recall and precision for ranked retrieval results<sup>45</sup>.

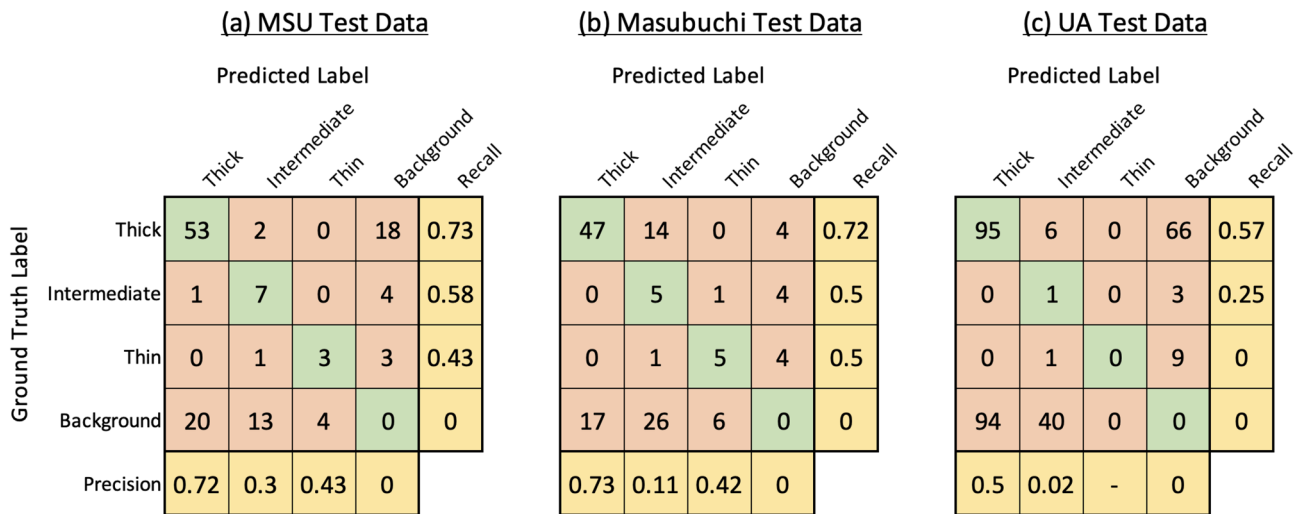
Another common tool used to measure the number (percentage) of correct and incorrect detections for each individual class based on the ground truth data is the confusion matrix. Figure 3 shows how a confusion matrix is defined for four generic classes. Rows represent the ground truth for each class and columns show the predicted results.

In our research, we have considered each flake within an image for evaluation. Following common practice in the field of object detection, we use an IoU of 0.5 to determine our numerical scores<sup>32,46</sup>. Therefore we consider a true positive example to be when individual flake was detected and classified correctly considering the IoU. Incorrect detections occur when a flake is not detected at all (i.e. it is considered by the algorithm to be part of the background substrate), when it is detected as an incorrect class, or when the background substrate is detected as a flake. Figure 4a shows the confusion matrix showing the classification results on the 10 images in the test dataset. It can be seen that our trained model was able to achieve a sufficiently high number of correct detections, more notably for the Thin.

To assess the generalization of our method, we employed our trained algorithm using on ten images taken from two other datasets: The Masubuchi dataset and the UA dataset. We also compared our algorithm to the pre-trained algorithm published by Masubuchi et al.<sup>32</sup>.

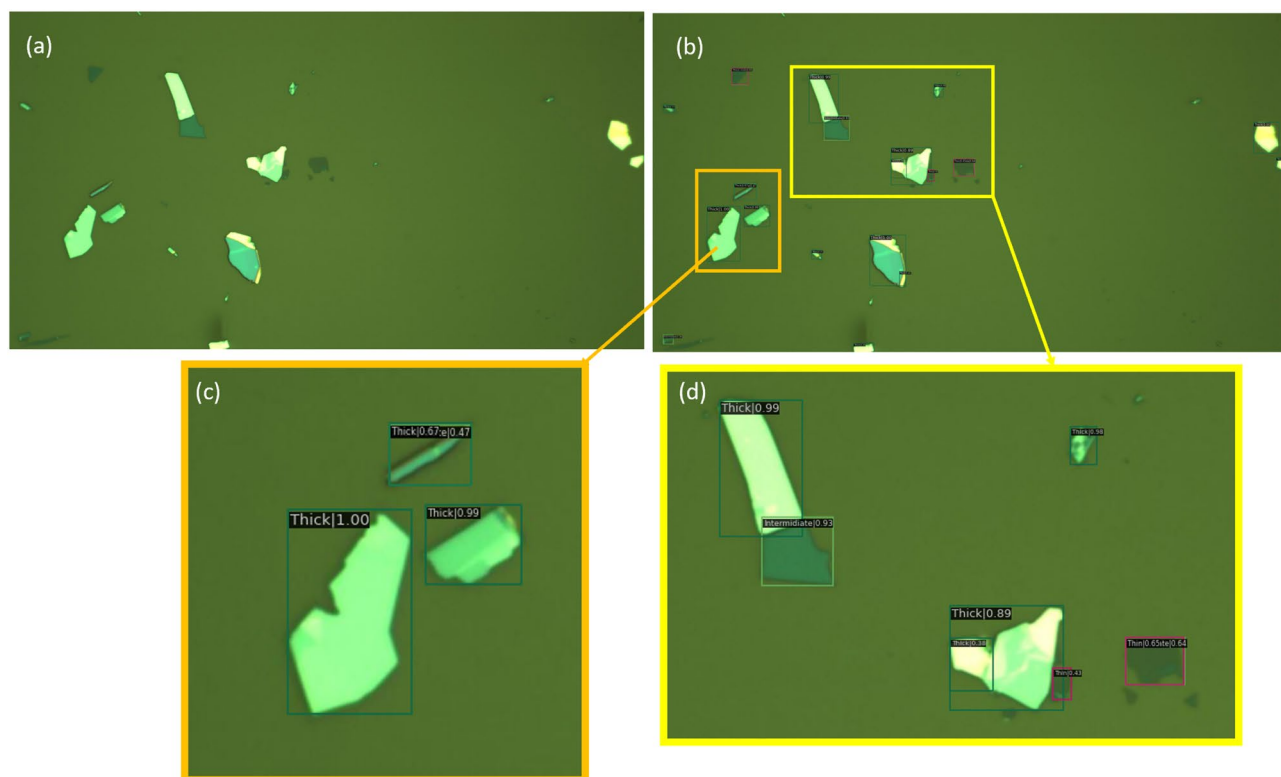
		Predicted Label				Recall
		Class A	Class B	Class C	Class D	
Ground Truth Label	Class A	TP <sub>A</sub>	FP <sub>B</sub> FN <sub>A</sub>	FP <sub>C</sub> FN <sub>A</sub>	FP <sub>D</sub> FN <sub>A</sub>	$\frac{TP_A}{(TP_A+FN_A)}$
	Class B	FP <sub>A</sub> FN <sub>B</sub>	TP <sub>B</sub>	FP <sub>C</sub> FN <sub>B</sub>	FP <sub>D</sub> FN <sub>B</sub>	$\frac{TP_B}{(TP_B+FN_B)}$
	Class C	FP <sub>A</sub> FN <sub>C</sub>	FP <sub>B</sub> FN <sub>C</sub>	TP <sub>C</sub>	FP <sub>D</sub> FN <sub>C</sub>	$\frac{TP_C}{(TP_C+FN_C)}$
	Class D	FP <sub>A</sub> FN <sub>D</sub>	FP <sub>B</sub> FN <sub>D</sub>	FP <sub>C</sub> FN <sub>D</sub>	TP <sub>D</sub>	$\frac{TP_D}{(TP_D+FN_D)}$
Precision		$\frac{TP_A}{(TP_A+FP_A)}$	$\frac{TP_B}{(TP_B+FP_B)}$	$\frac{TP_C}{(TP_C+FP_C)}$	$\frac{TP_D}{(TP_D+FP_D)}$	

**Figure 3.** Confusion matrix template for a generic multi-class classification problem, showing how precision and recall are calculated for each class.



**Figure 4.** Confusion matrix showing results obtained by applying the proposed algorithm to all three test datasets (MSU, Masubuchi, and UA).

To make a fair comparison between our method and Masubuchi method, we fed these unseen images from both datasets to our algorithm. Later we manually annotated the same images and labeled them as ground truth. Then we calculated the confusion matrix for both datasets based on the ground truths and detected results. The obtained confusion matrices when tested on 10 images from each dataset are shown in Fig. 4b,c.



**Figure 5.** Results obtained from using the proposed algorithm on one image from the Masubuchi dataset; Part (a) shows the raw image, part (b) shows the labels generated by the algorithm, and parts (c) and (d) are zoomed-in regions showing the flake type and confidence score reported by the algorithm. We manually identified 16 flakes in the image, and the algorithm correctly classified 12 of them (microscope images obtained from <https://doi.org/10.6084/m9.figshare.11881053> under the license CC BY 4.0).

Figures 5 and 6 show the detection results applying our proposed model to one image from the Masubuchi and UA datasets, respectively. In these images, our detector was able to identify 12 flakes correctly out of 16 annotated flakes within Masubuchi image, and 20 flakes correctly out of 32 annotated flakes for UA image.

## Discussion

Before training the DetectoRS model, we applied Masubuchi's pre-trained Mask-RCNN on our own dataset. We discovered this pre-trained model was not able to perform well on our dataset. For many test images, including the image shown in Fig. 2a, the pre-trained model was unable to identify any flakes. For other images, the pre-trained RCNN was able to identify only a small subset of the total number of flakes. Complete results of using the Masubuchi method on all three dataset are presented as confusion matrices in Fig. 7. We note that one cause of the poor performance of their algorithm is that we modified the labeling procedure and criteria. Future work will explore a robust, standardized method for data labeling so that algorithm performance can be compared more directly.

In addition, future work will explore integrating the proposed algorithm with an optical microscope to make the detection pipeline fully automatic. We also intend to collect and label data associated with additional 2D materials, expanding beyond hBN.

## Data availability

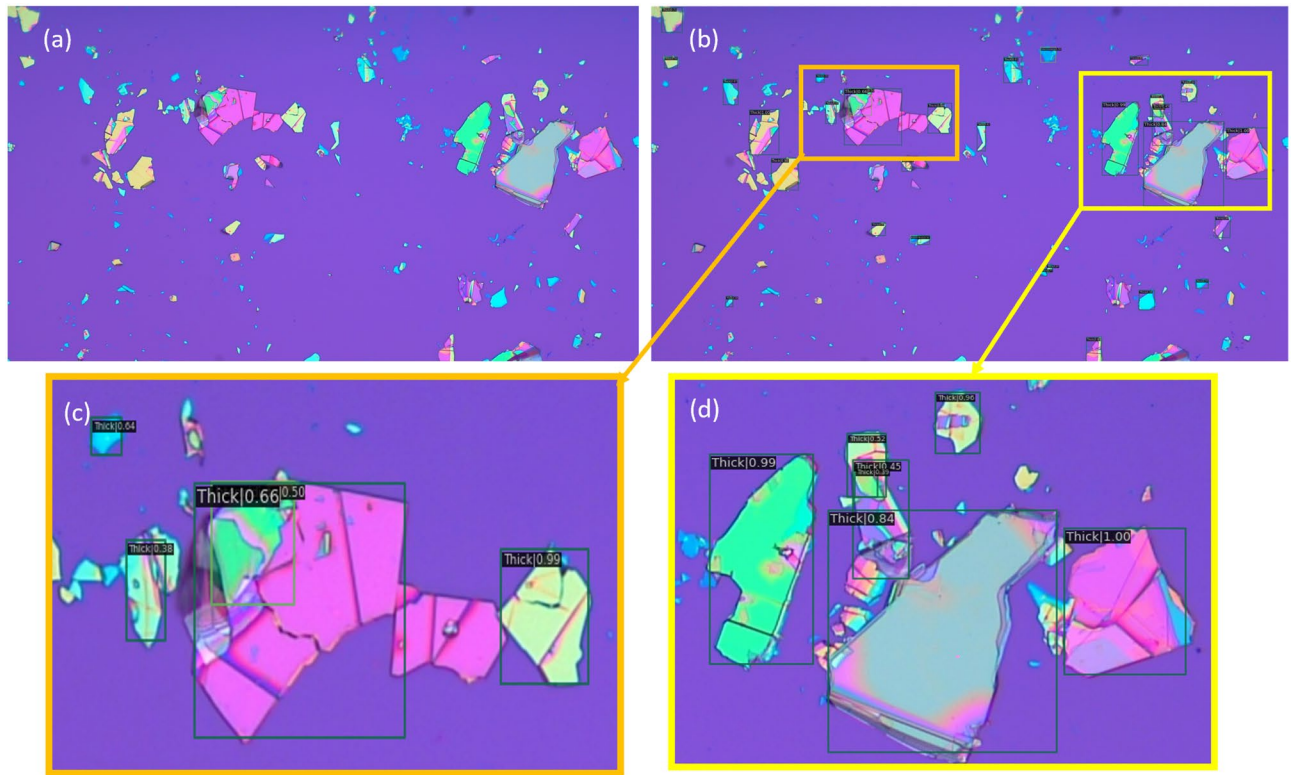
To help other researchers and scientists explore improved methods for solving 2D material object detection, the software and data used in this study is available on GitHub: [http://github.com/BMW-lab-MSU/hBN\\_Detection](http://github.com/BMW-lab-MSU/hBN_Detection). The software and data are published under the open-source BSD 3-Clause License. In addition, the repository has been archived on Zenodo and has received a digital object identifier (DOI): <https://doi.org/10.5281/zenodo.7576917>.

Received: 13 October 2022; Accepted: 23 January 2023

Published online: 28 January 2023

## References

1. Zou, Z., Shi, Z., Guo, Y., & Ye, J. Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055* (2019).
2. Liu, L. *et al.* Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **128**(2), 261–318 (2020).



**Figure 6.** Results obtained from using the proposed algorithm on one image from the UA dataset; Part (a) shows the raw image, part (b) shows the labels generated by the algorithm, and parts (c) and (d) are zoomed-in regions showing the flake type and confidence score reported by the algorithm. We manually identified 32 flakes in the image, and the algorithm correctly classified 20 of them. (Microscope images obtained by Hugh Churchill and used with permission.)

		<u>(a) MSU Test Data</u>					<u>(b) Masubuchi Test Data</u>					<u>(c) UA Test Data</u>				
		Predicted Label					Predicted Label					Predicted Label				
		Thick	Intermediate	Thin	Background	Recall	Thick	Intermediate	Thin	Background	Recall	Thick	Intermediate	Thin	Background	Recall
Ground Truth Label	Thick	3	0	0	70	0.04	16	0	0	49	0.25	9	0	0	158	0.05
	Intermediate	0	0	0	12	0	1	0	0	9	0	0	0	0	4	0
	Thin	0	0	0	7	0	5	0	0	5	0	0	0	0	10	0
	Background	0	0	0	0	-	0	0	1	0	0	1	0	0	0	0
	Precision	1	-	-	0		0.73	-	0	0		0.9	-	-	0	

**Figure 7.** Confusion matrices showing results when applying Masubuchi’s pre-trained Mask R-CNN to our all three test datasets (MSU, Masubuchi, UA). The pre-trained model was unable to identify most manually labeled flakes in the test images.

- Pathak, A. R., Pandey, M. & Rautaray, S. Application of deep learning for object detection. *Proc. Comput. Sci.* **132**, 1706–1717 (2018).
- O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. Deep learning vs. traditional computer vision. In *Science and information conference*, 128–144, Springer (2019).
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587 (2014).

6. Ren, S., He, K., Girshick, R., & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **28** (2015).
7. Girshick, R., Fast r-cnn. arXiv 2015. [arXiv preprint arXiv:1504.08083](https://arxiv.org/abs/1504.08083) (2015).
8. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125 (2017).
9. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788 (2016).
10. Redmon, J., & Farhadi, A. Yolo9000: Better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271 (2017).
11. Redmon, J., & Farhadi, A. Yolov3: An incremental improvement. [arXiv preprint arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018).
12. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. Ssd: Single shot multibox detector. In *European conference on computer vision*, 21–37, Springer (2016).
13. Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. [arXiv preprint arXiv:2004.10934](https://arxiv.org/abs/2004.10934) (2020).
14. Wu, X., Sahoo, D. & Hoi, S. C. Recent advances in deep learning for object detection. *Neurocomputing* **396**, 39–64 (2020).
15. Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**(11), 699–712 (2012).
16. Xu, X., Yao, W., Xiao, D. & Heinz, T. F. Spin and pseudospins in layered transition metal dichalcogenides. *Nat. Phys.* **10**(5), 343–350 (2014).
17. Lee, G.-H. *et al.* Flexible and transparent mos2 field-effect transistors on hexagonal boron nitride-graphene heterostructures. *ACS Nano* **7**(9), 7931–7936 (2013).
18. Shi, H. *et al.* Exciton dynamics in suspended monolayer and few-layer mos2 2d crystals. *ACS Nano* **7**(2), 1072–1080 (2013).
19. Lv, R. *et al.* Transition metal dichalcogenides and beyond: Synthesis, properties, and applications of single- and few-layer nanosheets. *Acc. Chem. Res.* **48**(1), 56–64 (2015).
20. Mak, K. F. & Shan, J. Photonics and optoelectronics of 2d semiconductor transition metal dichalcogenides. *Nat. Photonics* **10**(4), 216–226 (2016).
21. Mak, K. F., Lee, C., Hone, J., Shan, J. & Heinz, T. F. Atomically thin mos 2: A new direct-gap semiconductor. *Phys. Rev. Lett.* **105**(13), 136805 (2010).
22. Splendiani, A. *et al.* Emerging photoluminescence in monolayer mos2. *Nano Lett.* **10**(4), 1271–1275 (2010).
23. Darlington, T. P. *et al.* Imaging strain-localized excitons in nanoscale bubbles of monolayer wse2 at room temperature. *Nat. Nanotechnol.* **15**(10), 854–860 (2020).
24. Li, Y., Yang, B., Xu, S., Huang, B. & Duan, W. Emergent phenomena in magnetic two-dimensional materials and van der Waals heterostructures. *ACS Appl. Electron. Mater.* **4**(7), 3278–3302 (2022).
25. Zhang, K., Feng, Y., Wang, F., Yang, Z. & Wang, J. Two dimensional hexagonal boron nitride (2d-hbn): Synthesis, properties and applications. *J. Mater. Chem. C* **5**(46), 11992–12022 (2017).
26. Gupta, A., Sakthivel, T. & Seal, S. Recent development in 2d materials beyond graphene. *Prog. Mater. Sci.* **73**, 44–126 (2015).
27. Novoselov, K., Mishchenko, O. A., Carvalho, O. A. & Castro Neto, A. 2d materials and van der Waals heterostructures. *Science* **353**(6298), aac9439 (2016).
28. Lipatov, A. *et al.* Direct observation of ferroelectricity in two-dimensional mos2. *npj 2D Mater. Appl.* **6**(1), 1–9 (2022).
29. Zhang, X. *et al.* Advanced tape-exfoliated method for preparing large-area 2d monolayers: A review. *2D Mater.* **8**(3), 032002 (2021).
30. Guo, H.-W., Hu, Z., Liu, Z.-B. & Tian, J.-G. Stacking of 2d materials. *Adv. Func. Mater.* **31**(4), 2007810 (2021).
31. Han, B. *et al.* Deep-learning-enabled fast optical identification and characterization of 2d materials. *Adv. Mater.* **32**(29), 2000953 (2020).
32. Masubuchi, S. *et al.* Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials. *npj 2D Mater. Appl.* **4**(1), 1–9 (2020).
33. Lin, X. *et al.* Intelligent identification of two-dimensional nanostructures by machine-learning optical microscopy. *Nano Res.* **11**(12), 6316–6324 (2018).
34. Yang, J. & Yao, H. Automated identification and characterization of two-dimensional materials via machine learning-based processing of optical microscope images. *Extreme Mech. Lett.* **39**, 100771 (2020).
35. Li, Y. *et al.* Rapid identification of two-dimensional materials via machine learning assisted optical microscopy. *J. Materiomics* **5**(3), 413–421 (2019).
36. Masubuchi, S. & Machida, T. Classifying optical microscope images of exfoliated graphene flakes by data-driven machine learning. *npj 2D Mater. Appl.* **3**(1), 1–7 (2019).
37. Nguyen, X. B., Bisht, A., Churchill, H., & Luu, K. Two-dimensional quantum material identification via self-attention and soft-labeling in deep learning. [arXiv preprint arXiv:2205.15948](https://arxiv.org/abs/2205.15948) (2022).
38. Sanchez-Juarez, J., Granados-Baez, M., Aguilar-Lasserre, A. A. & Cardenas, J. Automated system for the detection of 2d materials using digital image processing and deep learning. *Opt. Mater. Express* **12**(5), 1856–1868 (2022).
39. Qiao, S., Chen, L.-C., & Yuille, A., “DetectorRS: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10213–10224 (2021).
40. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 1–48 (2019).
41. Li, X. *et al.* Transfer learning in computer vision tasks: Remember where you come from. *Image Vis. Comput.* **93**, 103853 (2020).
42. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755, Springer (2014).
43. Chen, K., *et al.* MMDetection: Open mmlab detection toolbox and benchmark. [arXiv preprint arXiv:1906.07155](https://arxiv.org/abs/1906.07155) (2019).
44. Padilla, R., Netto, S. L., & Da Silva, E. A. A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)*, 237–242, IEEE (2020).
45. Zhang, E., & Zhang, Y. Average precision. (2009).
46. Jiang, H., & Learned-Miller, E. Face detection with the faster r-cnn. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, 650–657, IEEE (2017).

## Acknowledgements

We acknowledge the MonArk NSF Quantum Foundry supported by the National Science Foundation Q-AMASE-i program under NSF award No. DMR-1906383. We also gratefully acknowledge our MonArk colleagues at the University of Arkansas: Hugh Churchill, Khoa Luu, Xuan Bac Nguyen, and Jeremy Choh. They were instrumental in creating, collecting, and sharing the 10 UA test images and allowing us to release those images with our annotations.



### Author contributions

F.R. designed and developed the algorithm, performed computational frameworks and wrote the manuscript with significant support and contribution from B.M.W., S.P., J.P.F., A.B. and S.W. contributed to data acquisition under the supervision of N.J.B. All authors, especially B.M.W. and N.J.B., provided critical feedback and helped shape the research, analysis and manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to F.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**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/>.

© The Author(s) 2023