Two-Dimensional Quantum Material Identification via Self-Attention and Soft-Labeling in Deep Learning

XUAN BAC NGUYEN\textsuperscript{1,2}, (Graduate Student Member, IEEE), APOORVA BISHT\textsuperscript{1,2}, BENJAMIN THOMPSON\textsuperscript{1}, HUGH CHURCHILL\textsuperscript{2,3}, KHOA LUU\textsuperscript{1,2}, AND SAMEE U. KHAN\textsuperscript{4}, (Senior Member, IEEE)

\textsuperscript{1}Department of Electrical Engineering and Computer Science, University of Arkansas, Fayetteville, AR 72701, USA
\textsuperscript{2}MonArk NSF Quantum Foundry
\textsuperscript{3}Department of Physics, University of Arkansas, Fayetteville, AR 72701, USA
\textsuperscript{4}Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762, USA

Corresponding author: Khoa Luu (khoaluu@uark.edu)

This work was supported in part by MonArk NSF Quantum Foundry through the National Science Foundation Q-AMASE-i Program under NSF Award DMR-1906383.

ABSTRACT Detecting two-dimensional (2D) materials in silicon chips presents a significant challenge in the field of quantum machines due to the difficulty of data collection. Specifically, among thousands of flakes, not all flakes are useful or well-annotated, resulting in noisy and hard samples within the dataset, which challenges the deep neural network (DNN) to learn. To address this problem, we propose a novel method for identifying quantum 2D flakes even when there is a high rate of missing annotations in the input images. In particular, we first propose a new mechanism for automatically detecting false negative flakes that are missing annotations. Second, we introduce an attention-based loss function to mitigate the negative impact of these unannotated flakes on the DNNs. The experimental results demonstrate that our method outperforms previous approaches.

INDEX TERMS Quantum material, 2D flake detection, deep learning, computer vision, identification.

I. INTRODUCTION

Two-dimensional (2D) materials hold immense potential for studying optical, electrical, thermal, and magnetic properties of materials and exploiting these properties at the micro-level. In particular, two-dimensional (2D) materials present a unique opportunity to engineer van der Waals (vdW) heterostructures for diverse applications and to explore fundamental material properties [1]. These vdW heterostructures can vary from monolayers encapsulated in hexagonal boron nitride (hBN) to enhance electronic [2], [3], [4], [5], [6], and optical properties and ensure spatial homogeneity, to stack with up to ten layers, including various semiconductors, graphitic gates, and hBN dielectrics and spacer layers [7], [8], [9]. Despite advances in the growth of 2D materials and their large-scale fabrication into vdW heterostructures [10], [11], exfoliation and mechanical stacking of 2D crystals remain the preferred methods [12], [13] for basic research and proof-of-principle devices [14].

Till now, many physical methods [15] like Atomic Force Microscopy [16], High-Resolution Transmission Electron Microscopy, Raman Spectroscopy [17], and White Light Contrast Spectroscopy [18] have been used to characterize the dimensions of 2D flakes. However, this is a tedious process, and the experimenter has to observe the flakes to determine their characteristics manually. It requires resources, skill, and time. Recently, deep learning based approaches to identify and characterize 2D flakes have been proposed and implemented [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. Depending on the thickness of the flakes deposited on the dielectric layers, the optical images observed under the microscope show a gradient of colors. These colors are
characteristic of the flake thickness and depend on the material used. For example, hBN (hexagonal boron nitride) flakes have a distinct color profile; similarly, graphene has its own color profile. It is this relationship among thickness, material, and color that the classification hinges on. The bottleneck is that an extensive data set is required to implement a fully functional and accurate deep learning based application, and the data set consisting of images of flakes needs to be fully annotated. It means all image flakes should be identified, segmented, and annotated to train the model without false information. In addition to identifying the thickness of the flake at specific pixels, the color gradient can be used to determine the quality of the flake due to changes in thickness. With more characteristics of 2D flakes identified based on microscope images, a deep-learning model can be trained to characterize not just the thickness but also the grade and other optical properties of the 2D flakes. It holds great promise to increase the efficiency and accuracy of the 2D flake “hunting” process.

Recently, several studies have introduced machine learning techniques such as K-Means and Support Vector Machines [28], [29], as well as deep learning models [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], to characterize flakes. However, these approaches are still limited by the challenges related to flake data. Unlike typical computer vision problems involving common objects, such as those in ImageNet [30] and COCO [31], flake detection data can contain thousands of flakes in a single image. Labeling all these flakes is time-consuming, labor-intensive, and costly. As a result, some useful flakes might remain unannotated, as illustrated in Fig 2 from the database [19]. Consequently, deep neural networks (DNNs) may be confused during training, as some flakes are currently treated as background due to missing annotations. In this paper, we propose a novel framework of instance segmentation for 2D quantum material detection using optical microscopy images. This framework aims to address the problem of sparse or missing annotations existing in the dataset, resulting in lower recall and overall performance. Our contributions can be summarized as follows:

- We introduce a new mechanism based on attention for automatically detecting potential objects that do not have annotations. This method is trained end-to-end together with the full deep learning-based detection framework.
- We propose an attention-based loss strategy that can assign soft-labels for the objects detected above. This approach helps reduce the gradient impact of these objects contributing to the whole framework, thus increasing the recall and overall performance.

II. RELATED WORKS
A. PRIOR WORKS IN 2D QUANTUM MATERIAL IDENTIFICATION

Recently, material scientists have been using various methods to characterize the flakes such as location and physical thickness. Some of these methods are summarized below.

1) AFM (ATOMIC FORCE MICROSCOPY)
Although an AFM can provide resolutions as fine as a few fractions of a nanometer, it requires minute calibrations and has a limited scanning area of 150 by 150 micrometers [16]. It also has a slow throughput and can damage the crystal lattice during measurement [18].

2) TEM (TRANSmission ELECTRON MICROSCOPY)
TEM utilizes a high-voltage beam of electrons to generate a highly magnified sample image. Properties of materials, like the thickness and defects, can be detected and imaged. However, the cost of a typical transmission electron microscope can be as high as $10 million [32].

3) RAMAN SPECTROSCOPY
Raman spectroscopy uses Raman signals from the materials to determine the thickness. Although a common method to determine the thickness of 2D materials, it has limitations. Raman spectroscopy gives the thickness of the materials at the location where the light is focused. Thus, many Raman measurements need to be taken to determine the thickness at multiple locations. It is time-consuming and often hard to map the entire flake [33].

4) WHITE LIGHT CONTRAST SPECTROSCOPY
This method increases the contrast between the 2D flake and the substrate by utilizing the reflection spectrum from normal white light [34].

B. DEEP LEARNING IN 2D QUANTUM MATERIALS

2D quantum material research topics have gained attention from researchers in recent years. In particular, Bingnan et al. [20] proposed a deep learning-based optical identification method to recognize 13 types of 2D materials from optical microscopic images. The method consists of several convolutions, batch norm [35], ReLU, [36] followed by a softmax layer at the end. The method is inspired by UNet [37] architecture to segment the flakes with different thickness values. Satoru et al. [19] presented an end-to-end pipeline for 2D flake identification. The method uses optical microscopy images as the input and outputs each flake’s location and material type. The authors leverage Mask-RCNN [38] to predict the bounding box and segmentation mask of different types of flake. This method achieves a breakthrough in the field of 2D quantum material detection.

III. THE PROPOSED METHOD

Our proposed method employs Mask-RCNN [38] as the baseline architecture. It contains three main components: the Backbone (B), the Region Proposal Network (RPN), and the Region of Interest Head (ROI).

Let $I \in \mathbb{R}^{H \times W \times C}$ be the input image, where $H, W, C$ are height, width, and number of channels accordingly. The backbone $B$ extracts deep feature maps of $I$ denoted as
Let $na$ be the number of anchors, then the total number of proposals that can be generated is $|\mathcal{P}_s| = H_s \times W_s \times na$. $\mathcal{P}_s$ is split into two subsets: $\mathcal{P}_{pos}$ for positive proposals and $\mathcal{P}_{neg}$ for negative proposals.

$$\mathcal{P}_s = \mathcal{P}_{pos} + \mathcal{P}_{neg}$$

(2)

The objectives of RPN are: (1). Predicting if a proposal is foreground (1) or background (0) and (2). Estimating anchor delta.

$$L_{RPN} = L_{cls} + L_{reg}$$

(3)

$$= \frac{1}{|\mathcal{P}_s|} \sum_{p_i \in \mathcal{P}_s} L_{cls}(\hat{p}_i, p_i)$$

$$+ \frac{1}{|\mathcal{P}_{neg}|} \sum_{p_i \in \mathcal{P}_{neg}} L_{reg}(\hat{t}_i, t_i)$$

where $L_{cls}$ and $L_{reg}$ are the loss functions. $\hat{p}_i$ is the probability the anchor $i^{th}$ is foreground while $p_i$ is the ground truth. Similarly, $\hat{t}_i, t_i$ are predictions and ground truth of anchor size respectively. It is noted that $\mathcal{P}_s$ includes both negative and positive proposals. Thus, in the Eq. (8), $L_{reg}$ encounters the positive proposals only, while $L_{cls}$ involves all type of proposals. The $L_{cls}$ can be reformulated as follows,

$$L_{cls} = L_{pos} + L_{neg}$$

(4)

$$= \frac{1}{|\mathcal{P}_{pos}|} \sum_{p_i \in \mathcal{P}_{pos}} L_{cls}(\hat{p}_i, p_i)$$

$$+ \frac{1}{|\mathcal{P}_{neg}|} \sum_{p_i \in \mathcal{P}_{neg}} L_{cls}(\hat{p}_j, p_j)$$

It is clear that $\forall \hat{p}_i \in \mathcal{P}_{pos}, p_i = 1$ and $\forall \hat{p}_j \in \mathcal{P}_{neg}, p_j = 0$. However, along with missing annotations, it leads to $\forall \hat{p}_j \in \mathcal{P}_{neg}$ such that $p_j = 1$ and $\hat{p}_j$ in other words, false negative samples. Thus, we propose a mechanism to detect false negative samples automatically by use of an alternative loss function that uses a soft label for these samples to help reduce gradient impacts on the overall architecture. As in Eq. (1), $\mathcal{F}_s$
We measure the self-attention between two sets of proposals \( A \) or background. It is noted that the probability of every anchor belonging to the foreground denotes the embedding feature.

First, we use a \( G_{\text{share}} \) to create a \( F_e \) feature map. It is used to estimate anchor delta values of each anchor by passing to a convolution \( G_{\text{reg}} \). Additionally, \( F_e \) is also used to create embedding features \( F_c \) for each anchor using \( G_e \), where \( G_e \) denotes the embedding feature.

It is followed by \( G_e \) is \( G_{\text{cls}} \), which is used to estimate the probability of every anchor belonging to the foreground or background. It is noted that \( F_e \in \mathbb{R}^{H_e \times W_e \times (na \times D)} \) and \( F_{\text{cls}} \in \mathbb{R}^{H_{\text{cls}} \times W_{\text{cls}} \times na} \), where \( D \) is the dimension of embedding feature and \( na \) is the number of anchors. \( G_{\text{cls}} \) is designed as a convolution with kernel size and stride as 1 to make sure that:

\[
H_e = W_{\text{cls}}, H_w = H_{\text{cls}} \text{ and every anchor has its corresponding embedding.}
\]

The proposal set \( \mathcal{P} \) is generated from \( F_{\text{cls}} \), and proposals can be identified as either positive or negative ones. We measure the self-attention between two sets of proposal \( \mathcal{P}_{\text{pos}} \) and \( \mathcal{P}_{\text{neg}} \) as follows,

\[
A = \text{softmax}(\text{Norm}(F_{\text{neg}}) \odot \text{Norm}(F_{\text{pos}}))^\dagger),
\]

where \( F_{\text{pos}} \in \mathbb{R}^{N_{\text{pos}} \times D}, F_{\text{neg}} \in \mathbb{R}^{N_{\text{neg}} \times D} \) are feature sets of all positive and negative proposals. The \( \dagger \) denotes for transpose matrix. \( \text{Norm} \) is denoted as the normalization function and \( \odot \) is matrix multiplication. \( A \in \mathbb{R}^{N_{\text{neg}} \times N_{\text{pos}}} \) is the attention map, \( A_{ij} \) represents for the similar score between \( i^{\text{th}} \) negatives and \( j^{\text{th}} \) positive sample and \( \sum_j A_{ij} = 1 \). We expect a high \( A_{ij} \) score for the false negative proposals and a low score for the true negative one. We define the threshold \( t \) to determine which sample is false negative if \( A_{ij} > t \). The loss function for negative proposals is reformulated as follows,

\[
L_{\text{neg}} = \frac{1}{|\mathcal{P}_{\text{neg}}|} \sum_{\hat{p}_i \in \mathcal{P}_{\text{neg}}} L_{\text{cls}}(\hat{p}_i, 1_{\text{max}(A_i) > t, \text{max}(A_i)})
\]

Now, the RPN loss as defined in the Eqn (8) now can be reformulated as follow

\[
L_{\text{RPN}} = L_{\text{cls}} + L_{\text{neg}}
= \frac{1}{|\mathcal{P}_s|} \sum_{p_i \in \mathcal{P}_s} L_{\text{cls}}(\hat{p}_i, p_i)
+ \frac{1}{|\mathcal{P}_{\text{neg}}|} \sum_{\hat{p}_i \in \mathcal{P}_{\text{neg}}} L_{\text{cls}}(\hat{p}_i, 1_{\text{max}(A_i) > t, \text{max}(A_i)})
\]

IV. DATASETS AND EXPERIMENT SETTINGS
A. 2D QUANTUM MATERIAL IDENTIFICATION DATASETS
1) 2D QUANTUM MATERIAL DETECTION DATASET
We use the quantum image dataset published in [19]. The dataset contains four types of materials: hBN, Graphene, MoS2, and WTe2. The number of corresponding images for each material is 353, 862, 569, and 318 images, and the number of objects for each material type is 456, 4805, 839, and 1053, respectively. Fig. 2 shows a sample from the dataset.

2) 2D QUANTUM MATERIAL THICKNESS ESTIMATION DATASET
In this work, we propose to develop an extensive training data set with thickness measurements of various flakes, ranging from around 5 nm to 100 nm, using a combination of optical microscopy and AFM. We start with silicon dioxide/silicon substrates already mounted with hexagonal boron nitride flakes. Since the substrate material and thickness alter the reflected light captured by the microscope, it was important...
to use consistent substrates. Switching substrates required recalibration to maintain the correspondence between color and thickness. hBN allows light to pass through it, and the reflected light from the substrate has a different spectrum based on the thickness of hBN.

We use an optical microscope to capture these images. The images are calibrated using calibration slides so that images captured across different times of the day and different microscopes can be compiled for use in the same data set.

We then use Atomic Force Microscopy (AFM) to measure the thickness of each flake captured via optical microscopy. Thus, we get color (RGB) information via the microscope image and thickness information via the AFM. It can be combined to create a data set that establishes a color-to-thickness relation. To do this, we create two types of data sets:

1) In the first version, we search for uniform-looking flakes and measure their thickness via the AFM. Since the variation in thickness is estimated to be below the target error margin of 5 nm, an average thickness is determined from various random spots on the flake, and a thickness value is assigned to it. Flake thickness is then annotated in the 20x optical microscope images. It means that every selected flake in the optical microscope image has associated metadata referring to the average thickness of the flake.

2) In the second version, we aim to generate aligned AFM and optical microscope images of individual flakes. The two images perfectly superimpose each other, and information from one image serves as metadata for the other image (thickness information from AFM and RGB values from optical microscope images). It helps us discard the bias towards uniform thickness flakes, which means that the dataset can accommodate an increase in the variety of thickness information. However, we need higher-resolution AFM and microscope images to enable this superposition between each pixel (or group of pixels). Thus, in version #2, we use 256 × 256 resolution AFM scans (over a region spanning 20µm² to 40µm²) and 50x optical microscope images.

During data collection, the following are the constraints to collect data.

3) SUBSTRATE THICKNESS
Our current data set uses a uniform-thickness silicon dioxide (SiO₂) substrate. However, different applications and groups use varying thicknesses, which results in variations in the apparent color of the 2D material. Since the model recognizes the thickness based on color, such drivers (like substrate thickness) can change the wavelength of the light reflected, creating an issue. Thus, we need proper calibration for every configuration.

4) MICROSCOPE CALIBRATION
Different research groups utilize microscopes, light sources, exposure times, RGB, and white balance. It means the same flake can have different apparent colors under different microscopes. It, again, poses an issue and hinders the goal of consistency. We use the same calibration across all samples and images to develop our data set. To make the application available to multiple research institutes, a one-time built-in calibration step should allow users to calibrate the application once and use it for their purpose.

5) MATERIALS
Various 2D materials are utilized in various research labs, and all have different thickness-dependent color spectrums based on their lattice structure. It means that training the model for one material implies that the application can be reliably used for that model. Thus, so far, to use the application for multiple materials, the model needs to be separately trained.

6) PROCESS TO COLLECT DATA
First, the optical microscope had to be calibrated, and we had to save a calibration picture. Then, the chip was loaded into the optical microscope, and we had to hunt for flakes to measure the thickness. After taking individual pictures of each flake, we took a stitched picture of all flakes to determine the chip’s position. Second, we loaded the chip into the AFM. After calibrating many parts of the AFM, including the (voltage?) and (Z-value?), we could hunt for the same flakes we imaged in the optical microscope. After finding a flake, we had to approach it with the microscopic tip of the AFM, which vibrates over the sample and collects the thickness data. The AFM then begins measuring the flake’s thickness, and even for small flakes, this takes at least 30 minutes. Upon obtaining this thickness measurement, we have to use software to remove noise from the image and make it cleaner. Finally, we have to annotate the image from the optical microscope. Additionally, it’s worth noting that a common issue with the process was a broken or blunted tip in the AFM, which resulted in corrupted and sometimes nonsensical thickness measurements. To fix this, the tip had to be replaced, consuming more unnecessary time and making the solution in sight seem all the more optimal. Since the time to collect data is very expensive, i.e., seven samples/week, we could collect 101 samples of hBN and 30 samples of Graphene, including their annotated pictures on the optical microscope and their thickness measurements on the AFM. They were labeled in CVAT.

B. EXPERIMENT SETTINGS
In our experiments, we use the detectron2 framework based on Pytorch. We design the RPN module as presented in Fig 1. In particular, \( G_{share} \) is a convolution that has kernel size, stride, and padding are 3, 1, and 1, respectively. For \( G_e \) and \( G_{reg} \), we employ a convolution with a kernel size of 1, and strike and padding are 3 and 1, correspondingly. Finally, the \( G_{cls} \) is designed as a fully connected layer. We follow the protocol and use the same hyperparameters as described in [40]. More specially, we employ Resnet-101 [41] as the backbone and FPN [42] for multilevel features enhancement. The number of
anchors is \( na = 3 \), and the feature dimension of the anchor’s embedding is \( D = 256 \). We train the network with a learning rate of 0.02 in 10000 iterations and gradually decrease ten times at the milestone of 5000 and 8000 iterations. The batch size is set to 48. The model is trained on a machine of 4GPUs with 48G each, and training time takes 2 hours to complete. Furthermore, we understand that the baseline in [19] uses Keras/Tensorflow implementation, so the model is retrained using detectron2 as a baseline.

We evaluate our method with the previous works [40], [43], [44], [45]. It is important to note that these methods are designed to solve the sparse label problem. For a fair comparison, we do not address the imbalance issue of the dataset in this work. Instead, we leave it for future work.

### C. THICKNESS ESTIMATION

Along with the flake detection, we also built a model to estimate the thickness of the detected flakes. This model is a linear regression machine learning algorithm implemented with Sklearn. While it is intended to be used on flakes that are uniform in color/thickness, it can also be used on less consistent flakes to predict the average thickness.

Furthermore, this module takes the image masks as predicted by the detection module as input. For each mask, \( n \) crops are taken randomly from inside the mask. Then the crops undergo a series of transformations, i.e., horizontal flips, vertical flips, and \( 90^{\circ} \) rotations, each with a probability of 50%.

Each crop, \( C \), was then flattened from an RGB image into a 1D array, \( x \).

\[
C \in \mathbb{R}^{H \times W \times C} \rightarrow x \in \mathbb{R} \tag{9}
\]

and the input array \( X \) was constructed by appending all flattened crops

\[
X = \{x_i\}_{i=0}^{nk-1} \tag{10}
\]

, where \( k \) is the number of flakes in the training set. The input array \( y \) is constructed by appending the thickness of each crop such that each thickness is included \( n \) times. It is demonstrated in Equation 11, where \( T_i \) represents the thickness of the flattened crop \( x_i \).

\[
y = \{T_i\}_{i=0}^{nk-1} \tag{11}
\]

The inputs \( X \) and \( y \) are then passed into the regression algorithm. This algorithm plots the dependent variable (thickness) against the independent variable (color) to find the linear function that best represents the relationship between the two.

Using this function, the algorithm predicts the thickness for each pixel within a crop. Each of these predictions is averaged together to get the thickness of that crop. This step is repeated \( n \) times for each of the \( n \) crops, and each of these is averaged again for the final thickness prediction of the flake.

### TABLE 1. Performance on different types of 2D quantum materials. G: Graphene, BN: hBN, M: MoS2, W: WTe2.

| Type | Methods | Bounding Box | Segmentation |
|------|---------|--------------|--------------|
| Baseline | 86.3 | 78.8 | 72.4 | 84.1 | 71.2 | 63.2 |
| BRL [40] | 86.5 | 79.8 | 73.1 | 84.5 | 71.4 | 63.7 |
| Co-mining [43] | 86.8 | 79.8 | 73.6 | 85.1 | 71.5 | 63.8 |
| Sparse R-CNN [44] | 86.2 | 79.2 | 73.3 | 85.2 | 72.8 | 64.4 |
| Ours (t = 0.8) | 87.8 | 80.8 | 75.7 | 85.7 | 73.8 | 65.4 |

| Type | Methods | Bounding Box | Segmentation |
|------|---------|--------------|--------------|
| Baseline | 73.2 | 62.0 | 59.0 | 73.4 | 72.3 | 62.9 |
| BRL [40] | 71.4 | 71.0 | 61.8 | 74.0 | 62.0 | 61.4 |
| Co-mining [43] | 71.6 | 71.4 | 62.8 | 71.0 | 69.8 | 60.8 |
| Sparse R-CNN [44] | 71.5 | 71.8 | 63.0 | 71.7 | 70.7 | 61.5 |
| Ours (t = 0.8) | 73.4 | 73.1 | 64.4 | 73.1 | 72.9 | 63.0 |

### TABLE 2. Performance on different types of 2D quantum materials using different training data. An asterisk denotes the use of \( k \)-fold cross-validation.

| Train | Test | Error |
|-------|------|-------|
| hBN | hBN* | 4.0 nm |
| hBN | Graphine | 8.15 nm |
| Graphine | Graphine* | 7.39 nm |
| hBN & Graphine | hBN & Graphine* | 5.6 nm |

### VI. ABLATION STUDIES

In this section, we further analyze the impact of choosing threshold \( t \) as in the Eqn (7). Additionally, we study the efficiencies of our method while it deals with different sparse label ratios.
A. THE CHOICE OF FALSE NEGATIVE DETECTION THRESHOLD

We select different threshold values $t$ to determine false negative samples and evaluate the performance of our proposed methods. In particular, we select $t \in [0.6, 0.8, 0.9]$. The performance is presented in the Table 3. Overall, the threshold $t = 0.8$ works best for Graphene, hBN, and MoS2 materials. Meanwhile, there is a mixed result with WeT2 material since each threshold has its own good performance in different metrics such as $AP_{50}$, $AP_{75}$, $AP$, etc. However, the differences in the performances are not much different in this case. For that reason, after considering the performance, stability, and tradeoff between thresholds, we suggest selecting $t = 0.8$ to determine false negative samples.

B. THE EFFICIENCY OF THE PROPOSED METHOD IN DIFFERENT SPARSE LABEL RATIOS

In this section, we evaluate the efficiency of our method under different sparse label ratios. Specifically, we retain 50% and 75% of the labels, i.e., bounding boxes and masks in the dataset, and drop the rest, resulting in sparse label ratios. We select Sparse R-CNN [44] for comparison, as this method achieved the second-best result, as shown in the previous section. The performance results are reported in Table 4. Performance drops dramatically when 50% of the labels are missing. For Sparse R-CNN [44], we observe an approximate drop of 13% in $AP_{box}$ and 6% in $AP_{mask}$. However, our proposed method only has a 10% drop in $AP_{box}$ and a 5% drop in $AP_{mask}$. Similarly, when 75% of the labels are retained, the performance drops are comparable to those observed with 50% label retention. Our method consistently outperforms Sparse R-CNN in scenarios with sparse label ratios. For other materials, i.e., hBN, MoS2, and WeT2, we observe similar results as Graphene: our method performs well compared to the Sparse R-CNN [44].

C. RECALL PERFORMANCE

In this section, we analyze how the proposed method helps to improve the recall performance. We report the recall performance of various methods on different types of material as in Fig 3. This demonstrates that our proposed method has higher recall than previous approaches in graphene, hBN, MoS2, and WeT2. This means our method tends to detect more flakes than others in the case of sparse labels.

VII. CONCLUSION

In this paper, we present a novel attention-based mechanism for automatically detecting potential flakes that lack annotations. Upon identifying these unannotated flakes, we introduce a new loss strategy that assigns soft labels to mitigate their

---

### Table 3. Performance of our method on different thresholds. G: Graphene, BN: hBN, M: MoS2, W: WeT2.

| Type | Threshold | Bounding Box | Segmentation |
|------|-----------|--------------|--------------|
|      |           | AP$_{50}$   | AP$_{75}$   | AP | AP$_{50}$ | AP$_{75}$ | AP |
| G    | $t = 0.6$ | 87.1        | 80.3        | 74.6 | 85.0 | 72.7 | 64.6 |
|      | $t = 0.8$ | 87.8        | 80.8        | 75.7 | 85.7 | 73.8 | 65.4 |
|      | $t = 0.9$ | 87.1        | 80.2        | 74.1 | 84.5 | 72.4 | 64.3 |
| BN   | $t = 0.6$ | 72.1        | 72.1        | 63.7 | 72.1 | 71.3 | 62.7 |
|      | $t = 0.8$ | 73.4        | 73.1        | 64.4 | 73.1 | 72.9 | 63.0 |
|      | $t = 0.9$ | 70.3        | 70.3        | 62.1 | 70.3 | 70.3 | 60.3 |
| M    | $t = 0.6$ | 64.0        | 57.8        | 54.2 | 63.3 | 57.5 | 50.9 |
|      | $t = 0.8$ | 65.6        | 60.3        | 55.1 | 64.9 | 58.1 | 51.9 |
|      | $t = 0.9$ | 64.1        | 57.9        | 52.0 | 63.9 | 60.2 | 51.8 |
| W    | $t = 0.6$ | 78.3        | 65.3        | 61.4 | 77.0 | 64.6 | 55.9 |
|      | $t = 0.8$ | 78.8        | 67.6        | 61.8 | 77.3 | 63.2 | 55.3 |
|      | $t = 0.9$ | 79.0        | 68.5        | 61.2 | 77.4 | 65.0 | 55.4 |

### Table 4. Performance of our method on different sparse label ratios. G: Graphene, BN: hBN, M: MoS2, W: WeT2.

| Type | Threshold | 50% labels | 75% labels |
|------|-----------|------------|------------|
|      |           | AP$_{box}$ | AP$_{mask}$ | AP$_{box}$ | AP$_{mask}$ |
| G    | Ours $t = 0.8$ | 62.6 | 58.4 | 70.8 | 62.3 |
| BN   | Ours $t = 0.8$ | 55.1 | 59.5 | 60.7 | 58.4 |
| M    | Ours $t = 0.8$ | 47.7 | 45.8 | 50.2 | 48.4 |
| W    | Ours $t = 0.8$ | 50.6 | 48.3 | 52.8 | 50.1 |

### Figure 3. Recall the performance of our method on different material types. G: Graphene, BN: hBN, M: MoS2, W: WeT2.

### Figure 4. Qualitative results. The first row is the input images, the second row is our results, and the third row is the results of the baseline.
impact on the overall network. This approach enhances both performance and recall. Extensive experiments show that our method significantly surpasses previous approaches in 2D quantum material identification.

ACKNOWLEDGMENT
The authors acknowledges the Arkansas High-Performance Computing Center for providing GPUs.

REFERENCES
[1] A. K. Geim and I. V. Grigorieva, “Van der Waals heterostructures,” Nature, vol. 499, no. 7459, pp. 415–421, Jul. 2013.
[2] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrey, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, and J. Hone, “Boron nitride substrates for high-quality graphene electronics,” Nature Nanotechnol., vol. 5, no. 10, pp. 722–726, Oct. 2010.
[3] C. Neumann, S. Reichardt, P. Venezuela, M. Drügeler, L. Banszerus, M. Schmitz, K. Watanabe, T. Taniguchi, F. Mauri, B. Beschoten, S. Y. Won, and C. Stumper, “Raman spectroscopy as probe of nanomaterial scale strain variations in graphene,” Nature Commun., vol. 6, no. 1, p. 8429, Sep. 2015.
[4] L. Banszerus, M. Schmitz, S. Engels, J. Dauber, M. Oellers, F. Hautp, K. Watanabe, T. Taniguchi, B. Beschoten, and C. Stumper, “Ultrahigh mobility graphene devices from chemical vapor deposition on reusable copper,” Sci. Adv., vol. 1, no. 7, Jul. 2015, Art. no. e1500222.
[5] P. Cadiz, E. Courtade, C. Oliver, H. Cai, T. Taniguchi, K. Watanabe, H. Carrere, D. Lagarde, M. Manca, T. Amand, P. Renucci, S. Tongay, X. Marie, and B. Ursäkse, “Excitonic linewidth approaching the homogeneous limit in MoS2 -Based van der Waals heterostructures,” Phys. Rev. X, vol. 7, no. 2, May 2017, Art. no. 021026.
[6] A. Raja, L. Waldecker, Z. Zipf, Y. Cho, S. Brem, J. D. Ziegler, M. Kulig, T. Taniguchi, K. Watanabe, E. Malic, T. F. Heinz, T. C. Berkelbach, and A. Chernikov, “Defect disorder in two-dimensional materials,” Nature Nanotechnol., vol. 14, no. 9, pp. 832–837, 2019.
[7] Y. Xu, S. Liu, D. A. Rhodes, K. Watanabe, T. Taniguchi, J. Hone, V. Elser, K. F. Mak, and J. Shan, “Correlated insulating states at fractional fillings of moiré superlattices,” Nature, vol. 587, no. 7833, pp. 214–218, Nov. 2020.
[8] J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, “Flavour Hund’s coupling, Chern gaps and charge diffusivity in moiré superlattices,” Nature Phys., vol. 59, no. 7852, pp. 43–48, Apr. 2021.
[9] X. Liu, Z. Wang, K. Watanabe, T. Taniguchi, O. Vafek, and J. J. A. Li, “Tuning electron correlation in magic-angle twisted bilayer graphene using Coulomb screening,” Science, vol. 371, no. 6535, pp. 1261–1265, Mar. 2021.
[10] C. Backes et al., “Production and processing of graphene and related materials,” 2D Mater., vol. 7, no. 2, 2020, Art. no. 022001.
[11] O. J. Burton, Z. Winter, K. Watanabe, T. Taniguchi, B. Beschoten, C. Stumper, and S. Hofmann, “Putting high-index cu on the map for and H. N. Alshareef, “Growth of 2D materials at the wafer scale,” Adv. Mater., vol. 33, no. 12, pp. 1–9, 2021.
[12] L. Liu and T. Zhai, “Wafer-scale vertical van der Waals heterostructures,” Infomat, vol. 3, no. 1, pp. 3–21, 2021.
[13] X. Xu, T. Guo, H. Kim, M. K. Hota, R. S. Alsaadi, M. Lanza, X. Zhang, and H. N. Alshareef, “Growth of 2D materials at the wafer scale,” Adv. Mater., vol. 34, no. 14, Apr. 2022, Art. no. 2108258.
[14] Y. Liu, N. O. Weiss, X. Duan, H.-C. Cheng, Y. Huang, and X. Duan, “Van der Waals heterostructures and devices,” Nature Rev. Mater., vol. 1, no. 9, pp. 1–17, Jul. 2016.
[15] Y. Y. Wang, R. X. Gao, Z. H. Ni, H. He, S. P. Guo, H. P. Yang, C. X. Cong, and T. Yu, “Thickness identification of two-dimensional materials by optical imaging,” Nanotechnology, vol. 23, no. 49, Nov. 2012, Art. no. 495713.
[16] S. S. Ray, “Techniques for characterizing the structure and properties of polymer nanocomposites,” in Environmentally Friendly Polymer Nanocomposites (Composites Science and Engineering), S. S. Ray, Ed. Sawston, U.K.: Woodhead Publishing, 2013, ch. 4, pp. 74–88.
[17] A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, and A. K. Geim, “Raman spectrum of graphene and graphene layers,” Phys. Rev. Lett., vol. 97, no. 18, Oct. 2006, Art. no. 187401.
[40] H. Zhang, F. Chen, Z. Shen, Q. Hao, C. Zhu, and M. Savvides, “Solving missing-annotation object detection with background recalibration loss,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2020, pp. 1888–1892.

[41] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[42] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.

[43] T. Wang, T. Yang, J. Cao, and X. Zhang, “Co-mining: Self-supervised learning for sparsely annotated object detection,” in Proc. AAAI Conf. Artif. Intell., 2021, pp. 1–9.

[44] P. Sun, R. Zhang, Y. Jiang, C. Xu, W. Zhan, M. Tomizuka, L. Li, Z. Yuan, C. Wang, and P. Luo, “Sparse R-CNN: End-to-end object detection with learnable proposals,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 14454–14463.

[45] A. Shrivastava, A. Gupta, and R. Girshick, “Training region-based object detectors with online hard example mining,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 761–769.

XUAN BAC NGUYEN (Graduate Student Member, IEEE) received the B.Sc. degree in electronics and telecommunications from the VNU University of Engineering and Technology, in 2015, and the M.Sc. degree in computer science from the Department of Electrical and Computer Engineering, Chonnam National University, South Korea, in 2020. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Computer Engineering, University of Arkansas. In 2016, he was a Software Engineer in Yokohama, Japan. His research interests include quantum machine learning, face recognition, facial expression, and medical image processing.

APOORVA BISH was received the dual B.Sc. degree (Hons.) in physics and in computer science from the University of Arkansas, in 2023. She is currently pursuing the Ph.D. degree in AMO physics with the University of Colorado at Boulder, Boulder, CO, USA. As an Undergraduate Research Assistant with the Department of Physics and Computer Science, her research projects dealt with non-linear and quantum optics and using AI to characterize 2D materials.

BENJAMIN THOMPSON is currently pursuing the degree in mathematics and computer science. He is also an Undergraduate Research Assistant with the Computer Vision and Image Understanding Laboratory, University of Arkansas. His research interests include computer vision, machine learning, quantum computing, and optimization.

HUGH CHURCHILL received the Ph.D. degree in physics from Harvard University, in 2012. He was a Pappalardo Postdoctoral Fellow in physics at MIT. He is currently a Professor and the 21st Century Chair of Nanophysics with the Department of Physics, University of Arkansas (UA). He is also the Associate Director of Operations at MonArk NSF Quantum Foundry. Since 2015, he has led the Quantum Device Laboratory, UA. His research interests include quantum materials and devices, low-dimensional materials, quantum transport, optoelectronics, and automation of experiments.

KHOA LUU was the Research Project Director at the CyLab Biometrics Center, Carnegie Mellon University (CMU), Pittsburgh, PA, USA. He is currently an Assistant Professor and the Director of the Computer Vision and Image Understanding (CVIU) Laboratory, Department of Electrical Engineering and Computer Science (EECS), University of Arkansas (UA), Fayetteville, AR, USA. He is also with the MonARK NSF Quantum Foundry. He has received eight patents and three best paper awards and co-authored more than 120 papers in conferences, technical reports, and journals. His research interests include quantum machine learning, biometrics, smart health, and precision agriculture.

SAMEE U. KHAN (Senior Member, IEEE) received the Ph.D. degree from The University of Texas at Arlington, Arlington, TX, USA, in 2007. From August 2020 to July 2024, he was the Department Head. He is currently a Professor of electrical and computer engineering at Mississippi State University (MSU). Before arriving at MSU, he was the Cluster Lead (2016–2020) of Computer Systems Research at the National Science Foundation and the Walter B. Booth Professor at North Dakota State University. His work has appeared in more than 450 publications. His research interests include optimization, robustness, and security of computer systems. He is a fellow of the IET and BCS. He has won several awards, including the Best Paper Award (Systems Track), the IEEE Cloud Summit, in 2024; the IEEE R3 Outstanding Engineer Award, in 2024; the IEEE Computer Society Distinguished Contributor Award, in 2022 (inducted in the inaugural class); the IEEE ComSoc Technical Committee on Big Data Best Journal Paper Award, in 2019; the IEEE-USA Professional Achievement Award, in 2016; the IEEE Golden Core Member Award, in 2016; the IEEE TCSC Award for Excellence in Scalable Computing Research (Middle Career Researcher), in 2016; the IEEE Computer Society Meritorious Service Certificate, in 2016; the Tapestry of Diverse Talents Award, North Dakota State University (NDSU), Fargo, ND, USA, in 2016; an Exemplary Editor of IEEE TRANSACTIONS ON CLOUD COMPUTING AND DISTRIBUTED COMPUTING, and the Journal of Parallel and Distributed Computing. He is a Distinguished Member of the ACM.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.