ACS NANO

www.acsnano.org

ARTICLE

3D Deep Learning Enables Accurate Layer Mapping of 2D Materials

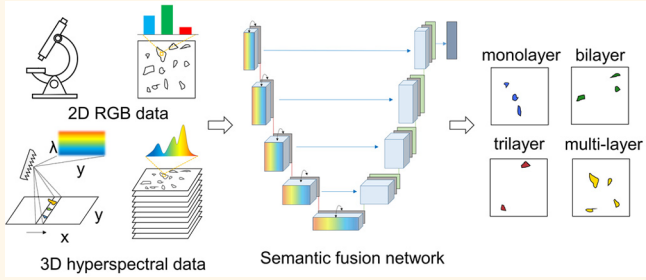
Xingchen Dong,*,# Hongwei Li,# Zhutong Jiang,# Theresa Grünleitner, İnci Güler, Jie Dong, Kun Wang, Michael H. Köhler, Martin Jakobi, Bjoern H. Menze, Ali K. Yetisen, Ian D. Sharp, Andreas V. Stier, Jonathan J. Finley, and Alexander W. Koch

Cite This: *ACS Nano* 2021, 15, 3139–3151

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Layered, two-dimensional (2D) materials are promising for next-generation photonics devices. Typically, the thickness of mechanically cleaved flakes and chemical vapor deposited thin films is distributed randomly over a large area, where accurate identification of atomic layer numbers is time-consuming. Hyperspectral imaging microscopy yields spectral information that can be used to distinguish the spectral differences of varying thickness specimens. However, its spatial resolution is relatively low due to the spectral imaging nature. In this work, we present a 3D deep learning solution called DALM (deep-learning-enabled atomic layer mapping) to merge hyperspectral reflection images (high spectral resolution) and RGB images (high spatial resolution) for the identification and segmentation of MoS_2 flakes with mono-, bi-, tri-, and multilayer thicknesses. DALM is trained on a small set of labeled images, automatically predicts layer distributions and segments individual layers with high accuracy, and shows robustness to illumination and contrast variations. Further, we show its advantageous performance over the state-of-the-art model that is solely based on RGB microscope images. This AI-supported technique with high speed, spatial resolution, and accuracy allows for reliable computer-aided identification of atomically thin materials.



KEYWORDS: *deep learning, 2D materials, layer number identification, hyperspectral imaging microscopy, semantic segmentation*

Atomically thin 2D materials and their heterostructures are prominent to promote the development of next-generation optics, optoelectronics, and quantum devices, due to their unique optical, mechanical, and electrical properties.[1−5] The fabrication of 2D materials is typically achieved by either growth (*e.g.*, chemical vapor deposition) or mechanical exfoliation, where 2D flakes with varying layer numbers are deposited randomly on a substrate.[6−8] However, it is important to know the exact layer numbers of 2D materials to make use of their layer-dependent properties for various applications. For example, the direct to indirect bandgap change of transition metal dichalcogenides (TMDs) enables high-efficiency optical emission from monolayer materials, whereas few-layer materials exhibit no luminescence but desirable electronic transport characteristics.[9−11] Specifically in the range from monolayer to few-layer materials, absorption spectra change significantly with the number of layers.[12,13] Due to these thickness-dependent properties, significant effort is currently devoted to identifying spatial regions with specific layer numbers. Likewise, optimization of wafer-scale mono- or bilayer deposition strategies requires robust measurement methods over large areas with high accuracy. To this end, atomic force microscopy (AFM), Raman, and photoluminescence spectroscopy techniques have been utilized for layer number identification,[14−18] but they are time-consuming when applied to flake search and layer number identification within large areas.[19] Optical microscopy has been widely implemented to search large-area samples and distinguish 2D flakes from monolayer to few-layer based on the optical contrast.[20−24] However, optical contrast experiments are not reliable specifically for the correct identification of few-layer samples and may depend on a calibrated illumination method.

Deep learning techniques, especially convolutional neural networks,[25] have shown significant advantages in computer vision tasks such as image segmentation and object classification.[26−29] Application of such networks for spatial

Received: November 18, 2020
Accepted: January 14, 2021
Published: January 19, 2021

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139–3151

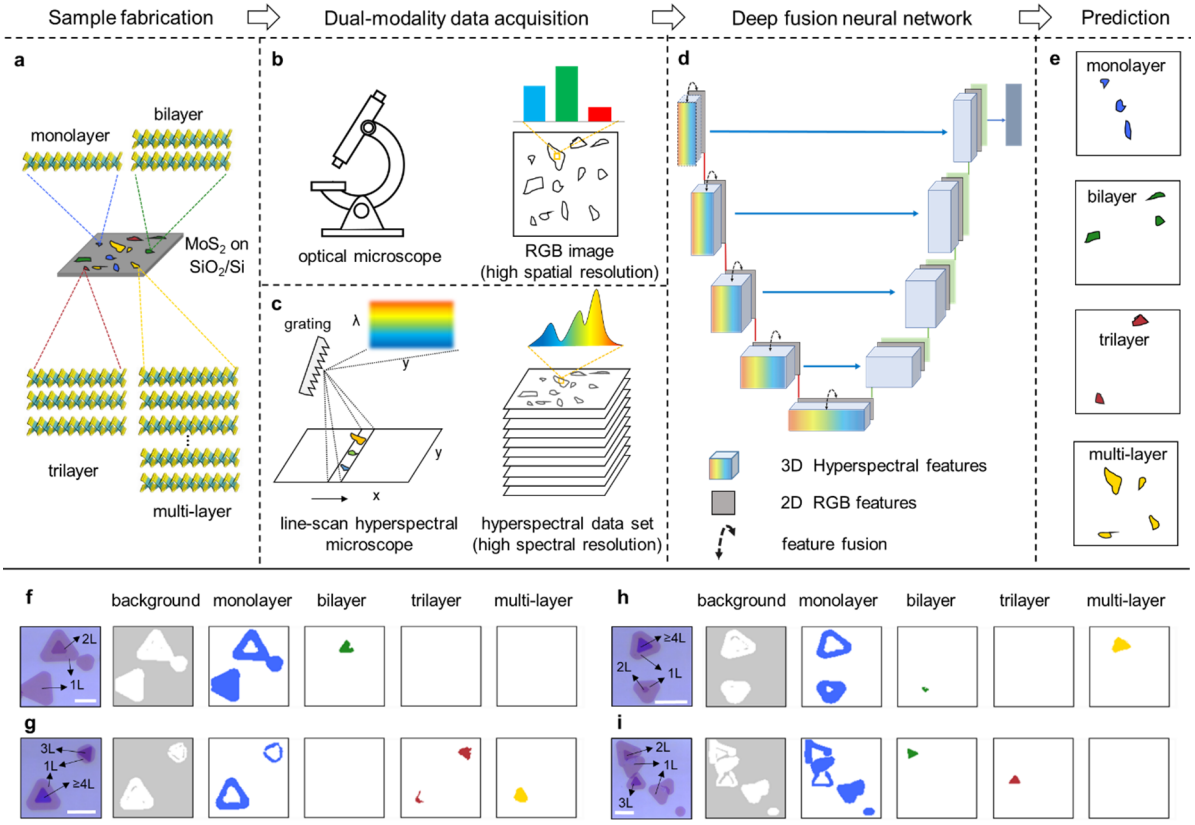


Figure 1. Process of the atomic layer mapping of 2D materials using DALM. (a–e) Workflow of 2D materials fabrication, dual-modality data acquisition, deep neural network training, and network prediction of layer number maps. (f–i) Predicted layer number maps of four testing regions with flake combinations of different layer numbers. Scale bar = 20 μm.

identification of 2D materials has recently attracted interest due to the outstanding prospects of such systems and the pressing need for intelligent image analysis to advance both fundamental research and device development. State-of-the-art approaches are commonly based on optical microscopy RGB images, to address pixel-wise segmentation,[30] thickness identification,[31] and flake quality classification.[32] There is existing work to distinguish monolayer and bilayer (together as one category) MoS_2 and graphene flakes using U-Net[33] and to segment optical images of 2D crystals (graphene, TMDs, and hBN) with rough categories (mono-, few-, and multilayer).[34,35] However, accurate layer mapping with only RGB images is extremely difficult because the image appearance of few-layer flakes can be visually similar. Making use of other imaging modalities, such as hyperspectral imaging, is not yet explored with 3D deep learning-based approaches. In this work, we show that the combination of hyperspectral imaging and deep learning enables accurate identification of specifically few-layer samples.

Hyperspectral imaging microscopy combines both spectroscopy and imaging techniques, providing spatial and spectral information on the measured region.[36,37] The abundant spectral information potentially allows for higher accuracy of layer number identification as compared to contrast-based microscope imaging. However, the spatial resolution of hyperspectral imaging microscopy is lower than conventional optical microscopy, thus resulting in less accurate profile outputs on the pixel-wise level. Hence, merging the semantics of both RGB images (accurate profile information) and hyperspectral reflection images (abundant spectral information) can potentially enable accurate atomic layer mapping of 2D materials. To achieve this, there are two main challenges: (a) hyperspectral imaging data are three-dimensional with two spatial and one spectral domain. Learning 3D semantics is challenging and often computationally expensive. (b) RGB data are two-dimensional. Effectively merging 3D hyperspectral and 2D RGB images into an end-to-end training process is not yet explored. State-of-the-art approaches fusing complementary information from different modalities are promising. For example, in the remote sensing field, the fusion of hyperspectral and LiDAR data is helpful for multiobject classification.[38] However, existing fusion methods can only handle data of the same dimensions.[39−41] Thus, we aim to extract useful hyperspectral information with 3D deep learning techniques[42] and combine it with 2D RGB information.

In this work, a 3D convolution neural network termed DALM was developed for automated characterization and layer number identification of MoS_2 flakes, which was achieved by learning the semantics from both 3D hyperspectral reflection microscope images and 2D RGB microscope images. DALM can distinguish layer numbers and generate accurate profiles of mono-, bi-, tri-, and multilayer flakes, by retaining the high spatial resolution of RGB images and the high spectral resolution of hyperspectral data sets. The image segmentation performance of DALM was quantitatively analyzed and fairly compared with that of a single-modal U-Net network that used only RGB microscope images for training and testing. Furthermore, the robustness of DALM to different illumination conditions was investigated. We have made DALM an

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

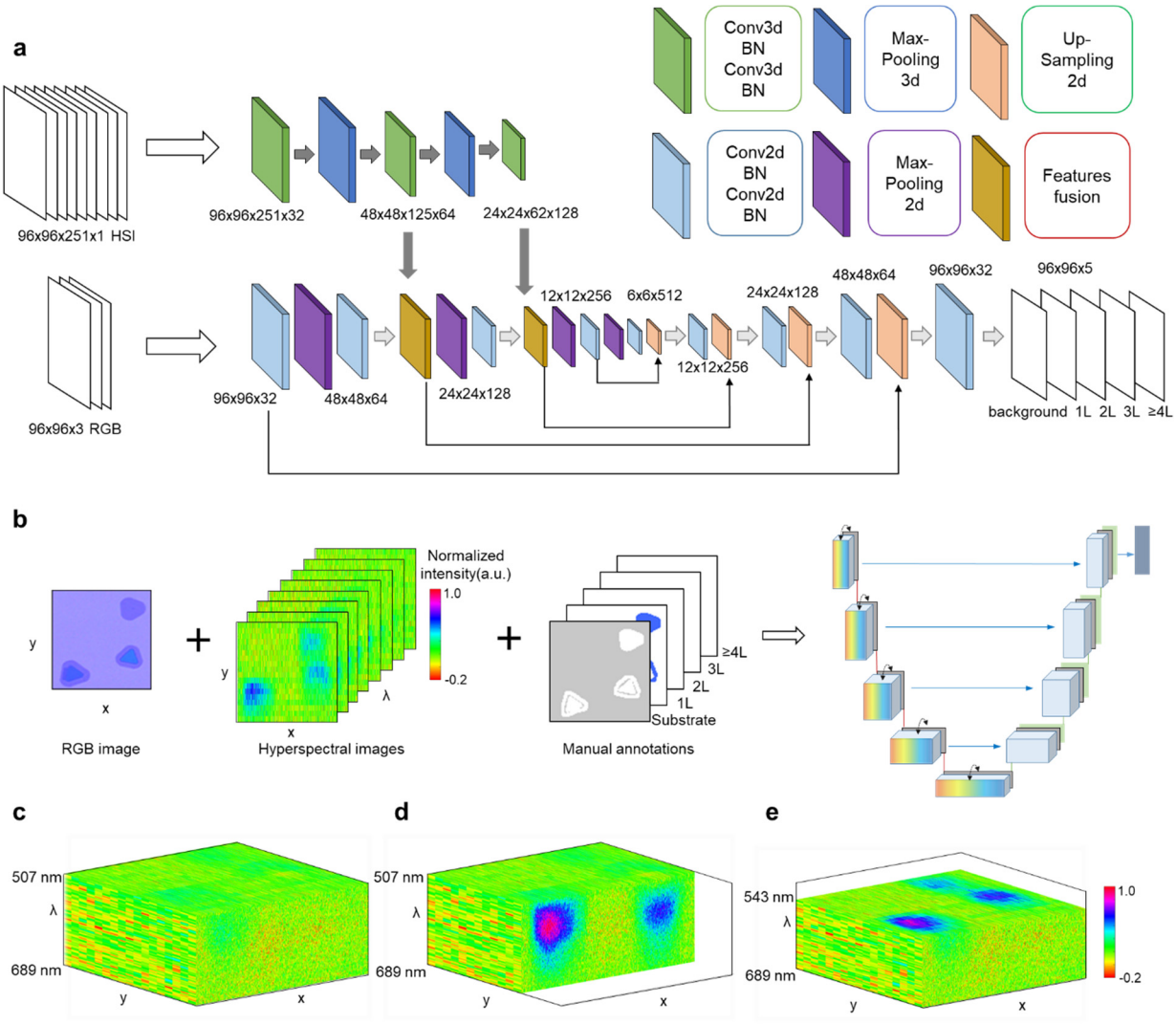


Figure 2. Detailed network architecture in DALM. (a) Our dual-stream network architecture. We treat hyperspectral reflection contrast images as 3D volumes and utilize 3D deep learning technique to extract semantic features. (b) Schematic view of the training stage. (c–e) Hyperspectral data sets of one training set at different scales. 3D hyperspectral images at the wavelength range of 507–689 nm contained a reflection contrast spectrum of each pixel within the measured region. Pixels belonging to flakes of different thicknesses showed different reflection contrast spectra. The reflection intensities of hyperspectral images were normalized to [0, 1] using the min–max intensity normalization (see Methods and Supporting Information Figure S9).

open-source repository in GitHub (https://github.com/hongweilibran/DALM) and a tool for research use.

RESULTS AND DISCUSSION

The Principle of DALM for Multimodal Multiclass Segmentation. Figure 1a illustrates the pipeline of atomic layer mapping of 2D flakes using the dual-stream deep convolutional neural network for fusing two data acquisition modalities. For sample preparation, MoS_2 flakes were fabricated on 270 nm SiO_2/Si substrates by chemical vapor deposition (CVD), with growth conditions selected to yield varying atomic thicknesses distributed over the substrate. An optical microscope (Leica) and a custom-built line-scan hyperspectral reflection microscope (100× objective, NA 0.85) were integrated to acquire the RGB and hyperspectral reflection contrast images, respectively. Considering that the characteristic reflection contrast spectra of MoS_2 flakes are dependent on the atomic layer number, thickness of the SiO_2 on the Si substrate, illumination conditions, and the numerical aperture of microscope objectives, the microscope objective and the substrate thickness were kept unchanged during measurement (see Methods and Supporting Information Figure S1).

Before the acquired data were fed into the deep-learning-based system, several simple preprocessing steps including data smoothing, background subtraction, and channel selection were applied to the hyperspectral data set to reduce the noise, correct the inhomogeneity of the illumination, and extract the key information from redundant hyperspectral raw data (see Methods). To transform the RGB images (3 channels) and the hyperspectral images (251 channels after dimension reduction) into one coordinate system, the two modalities were linearly co-registered (see Methods and Supporting Information Figure S2). The two types of images had different value ranges due to different camera sensors. Min–max data normalization was used to normalize the intensity range (see Methods). The training of the network required both registered data pairs and corresponding manual annotations. The data labeling was

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139–3151

conducted manually, and four layer categories along with the background were assigned to five subclasses, *i.e.*, background, mono-, bi-, tri-, and multilayers (see Methods and Supporting Information Figures S3−S5). Figure S6 illustrates changes of data dimension from raw data to data pairs suitable for network training. Data augmentation was adopted to generate more data based on the original data set and enlarge the training set. Paired hyperspectral and RGB images were augmented with random rotations, random flipping, and optical contrast changes to simulate varying experimental conditions (see Methods and Supporting Information Figure S7).

The registered RGB images and hyperspectral images were resampled to the same resolution and employed as the inputs of a 3D deep neural network. Since the hyperspectral images are in 3D while the RGB are in 2D, the fusion of the two distinct modalities is challenging. We proposed a method to combine the hyperspectral images and RGB images in multiple feature levels that were explicitly learned by state-of-the-art U-shape architecture.[43] In the training process, the segmentation masks (expected outputs) that indicated the correct layer maps were fed jointly with the RGB images and hyperspectral images into DALM. After data augmentation (*i.e.*, random rotation, random flipping), the training data sets included 800 image pairs (covering over 3000 flakes with varying layer numbers), each pair containing 3-channel RGB images and 251-channel hyperspectral images and covering a sample region of scales from 50 × 50 μm^2 to 80 × 80 μm^2. The network parameters were optimized to minimize the differences between the predictions and the manual annotations. In the testing process, data pairs of co-registered hyperspectral images and RGB images containing more than 20 MoS_2 flakes were employed for the network prediction of layer distributions. Through this deep fusion network, each pixel of the imaged area was classified into a subclass (background, monolayer, bilayer, trilayer, or multilayer). For fair comparison with the state-of-the-art, a single-stream U-Net (S-U-Net) model in parallel was built, trained, and tested using only the same RGB images from the paired data. The segmentation performance metrics (Dice similarity coefficient, Hausdorff distance, and confusion matrix) of both models were quantitatively compared. To assess the network's ability to generalize to the independent data set, the leave-one-sample-out methodology was conducted for cross-validation, which supported a more comprehensive evaluation of the network.

Figure 1f−i show the layer number maps of MoS_2 flakes from monolayer to multilayer using test data pairs, which did not appear in the training process. Each test data pair represented a region with a size from 50 × 50 μm^2 to 80 × 80 μm^2, where at least two types of MoS_2 flakes existed. To show the identification ability of the network, regions containing various layer numbers were selected. For example, Figure 1f contains mono- and bilayer flakes, while Figure 1g contains mono-, tri-, and multilayer flakes. Our DALM approach precisely identified layer numbers and accurate spatial profiles.

Deep Neural Network Architecture, Training, Validation, and Testing. DALM was based on the fully convolutional network that extracted the hierarchical image features of data sets and achieved pixel-wise semantic segmentation. Figure 2a illustrates the structure of the two-stream deep neural network, which consisted of 6 blocks including the 3D convolution layer "conv3d" with batch normalization (BN), the 3D max-pooling layer "maxpooling3d", the 2D up-sampling layer "up-sampling2d", the 2D convolutional layer "conv2d" with BN operations, the 2D max-pooling layer "maxpooling2d", and the features fusion block. The network included a contraction part (left) and an expansion part (right). The contraction part had a two-stream structure, where RGB images (2D) and hyperspectral images (3D) were fed jointly into the network. The 2D and 3D convolutions were operated upon separately in the two streams. In 2D convolutions, filters (3 × 3) moved in both spatial dimensions (x, y) to calculate low dimensional features and output a 2D matrix, while for 3D convolutions, 3D filters (3 × 3 × 1) were applied to the data set and the filters moved in all spatial and spectral dimensions (x, y, z) to calculate the low-level feature representations and output a 3D patch. Both the 2D and 3D convolution blocks conducted the down-sampling operation and extracted 2D and 3D feature maps from RGB and hyperspectral images, separately. The 2D and 3D features were combined through the feature fusion, which integrated the hyperspectral information on the 3D data set with the spatial information on 2D images using an optimal dimension fusion strategy (see Methods and Supporting Information Figure S8). The 2D spatial information on hyperspectral images was also extracted. The convolution layers (contraction part) were repeated, with a rectified linear unit (ReLu) activation function[44] and a max-pooling operation for each network layer. The expansion (decoder) part comprised the same structure as the encoder part for 2D RGB images (Figure 2a). DALM outputs a 2D predicted image, of which the number of channels corresponded to the number of segmented classes (background, monolayer, bilayer, trilayer, and multilayer).

To determine the error between the network output and the manual annotations, the Dice loss function was used to measure the overlap of two images. In this work, predicted probabilistic maps and manual annotation target masks had five categories, and each category represented a prediction class. Therefore, Dice loss was calculated class-by-class, where y_{label} was the manual annotation target mask and y_{pred} was the predicted probabilistic map. The Dice loss function can be expressed as

$$\text{Dice Loss} = 1 - \frac{2\sum_{pixels} |y_{label} \circ y_{pred}| + s}{\sum_{pixels} (|y_{label}| + |y_{pred}|) + s} \quad (1)$$

where ∘ represented the entrywise product of two matrices. "s" was set as 1 to avoid division by 0 (without s, if both y_{label} and y_{pred} were zero, the fraction would be undefined).

The framework for the training stage of DALM is illustrated in Figure 2b. In the network, batch normalization was used after each convolutional layer to accelerate training and improve the performance (see Methods). Ten percent of the training samples were pooled as a validation set to optimize the hyperparameters (see Methods for network training resources and parameters). Figure 2c−e show the hyperspectral data sets of the training set in Figure 2b at different spatial and spectral scales for demonstration. The reflection contrast spectra of different flakes after intensity normalization in the hyperspectral data set are shown in Figure S9. The learning curves of both DALM and S-U-Net are shown in Figure S10. Both plots showed the good fit learning curves, which meant the models were well trained during the process.

Prediction Performances and Evaluation of DALM. To evaluate the prediction performance of DALM, five pairs of

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

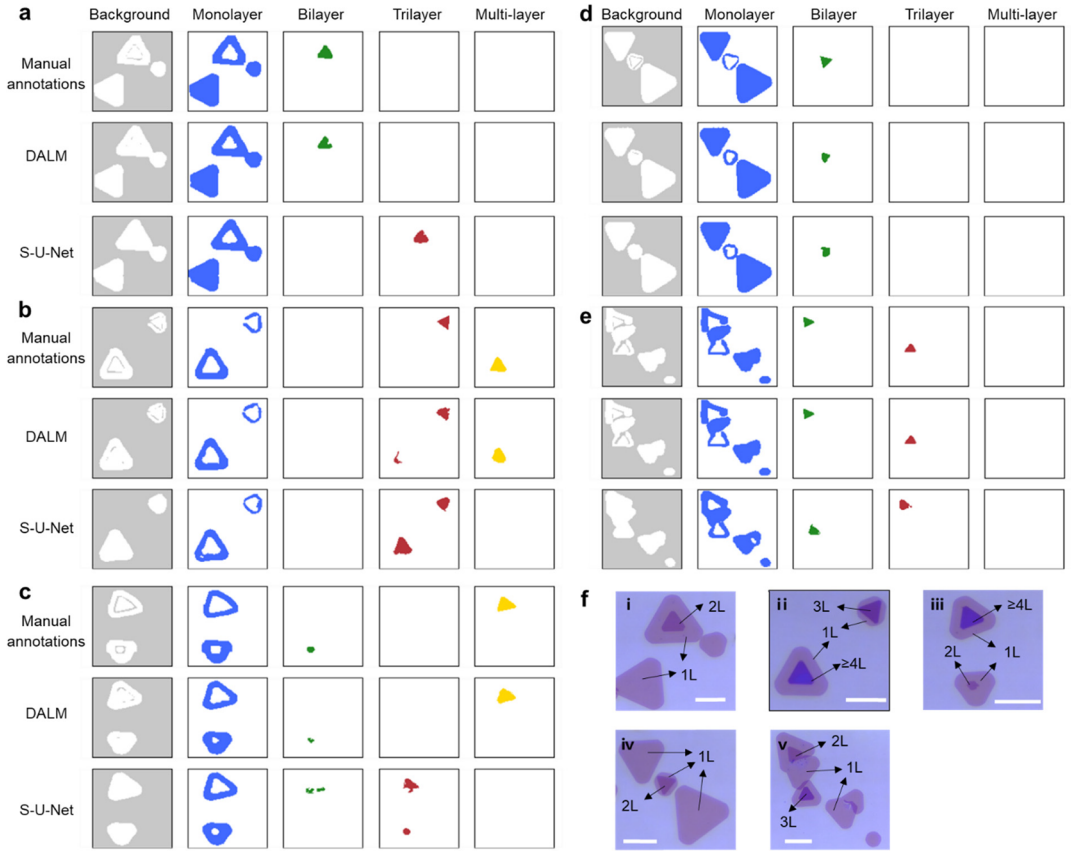


Figure 3. Prediction results of the distribution maps of substrate, monolayer, bilayer, trilayer, and multilayer flakes by the deep fusion network and the single-stream U-Net. (a−e) Segmentation of five subclasses of both DALM and S-U-Net, with manual annotations as references. (f) Optical images of these five test regions. (i)−(v) are the optical images of (a)−(e) regions, respectively. Scale bar = 20 μm.

hyperspectral and RGB images, captured at five regions containing more than 20 MoS_2 flakes, were used as a demonstration. The prediction results of S-U-Net that was trained and tested using only RGB images are demonstrated for comparison. The test was designed as a cross-evaluation using regions containing MoS_2 flakes with different layer number combinations. Each test pair contained three to five subclasses (background, mono-, bi-, tri-, and multilayer). Figure 3 illustrates the segmentation results of the test data using DALM and S-U-Net after eliminating the noisy points (see Methods). Figure 3a−e show the manual annotations, the prediction results by DALM, and the S-U-Net prediction results of each subclass. In Figure 3a, background and monolayer regions were correctly identified by both DALM and S-U-Net. However, the bilayer region, which was distinguished by DALM, was misidentified as trilayer by S-U-Net. In Figure 3b, mono-, tri-, and multilayer regions could be successfully identified by DALM, while S-U-Net took tri- and multilayer as one subclass. In Figure 3c, DALM output the correct layer number identification and accurate flake profiles of bilayer and multilayer regions, while S-U-Net was not able to distinguish between bi-, tri-, and multilayer features. In Figure 3d, both networks could identify monolayer and bilayer regions. Compared to Figure 3a, the S-U-Net showed randomness to correctly identify bilayer regions, which was unreliable in practice. Figure 3e shows the better performances of DALM to identify mono-, bi-, and trilayer regions over S-U-Net. The optical images of these five test regions are shown in Figure 3f with subclass labels as references.

Both DALM and S-U-Net can identify background and monolayer regions with high accuracy. DALM achieved a high success rate when predicting bi-, tri-, and multilayer regions, while S-U-Net was confused regarding these three subclasses. The prediction took several seconds for one testing data pair, which enabled measurement with small latency. In principle, the minimum flake size identified by the DALM is diffraction limited. In practice, to eliminate noise, a lateral threshold size of roughly double the diffraction limit is realized. Therefore, in Figure 3e, a flake smaller than ~1 × 1 μm^2 will be removed because of the image smoothing process.

To quantitatively analyze the prediction performances of both models, Dice similarity coefficient (DSC), Hausdorff distance, and confusion matrix were computed and employed to evaluate the layer number identification and flake region segmentation results. DSC and Hausdorff distance were calculated to assess the overlapping and localization accuracy. The DSC is defined as

$$\mathrm{DSC} = \frac{2(G \cap P)}{|G| + |P|} \tag{2}$$

where G is the labeled pixels of each subclass (manual annotation) and P is the predicted pixels of each subclass. The closer the score is to 1, the higher the successful prediction rate (Figure 4a). Figure 4b shows the calculated DSC values of DALM and S-U-Net. Both models had high DSC values (>90%) in the segmentation of background and monolayer subclasses. However, S-U-Net achieves 22%, 20%, and 4% in segmentation of bi-, tri-, and multilayer subclasses, respectively.

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

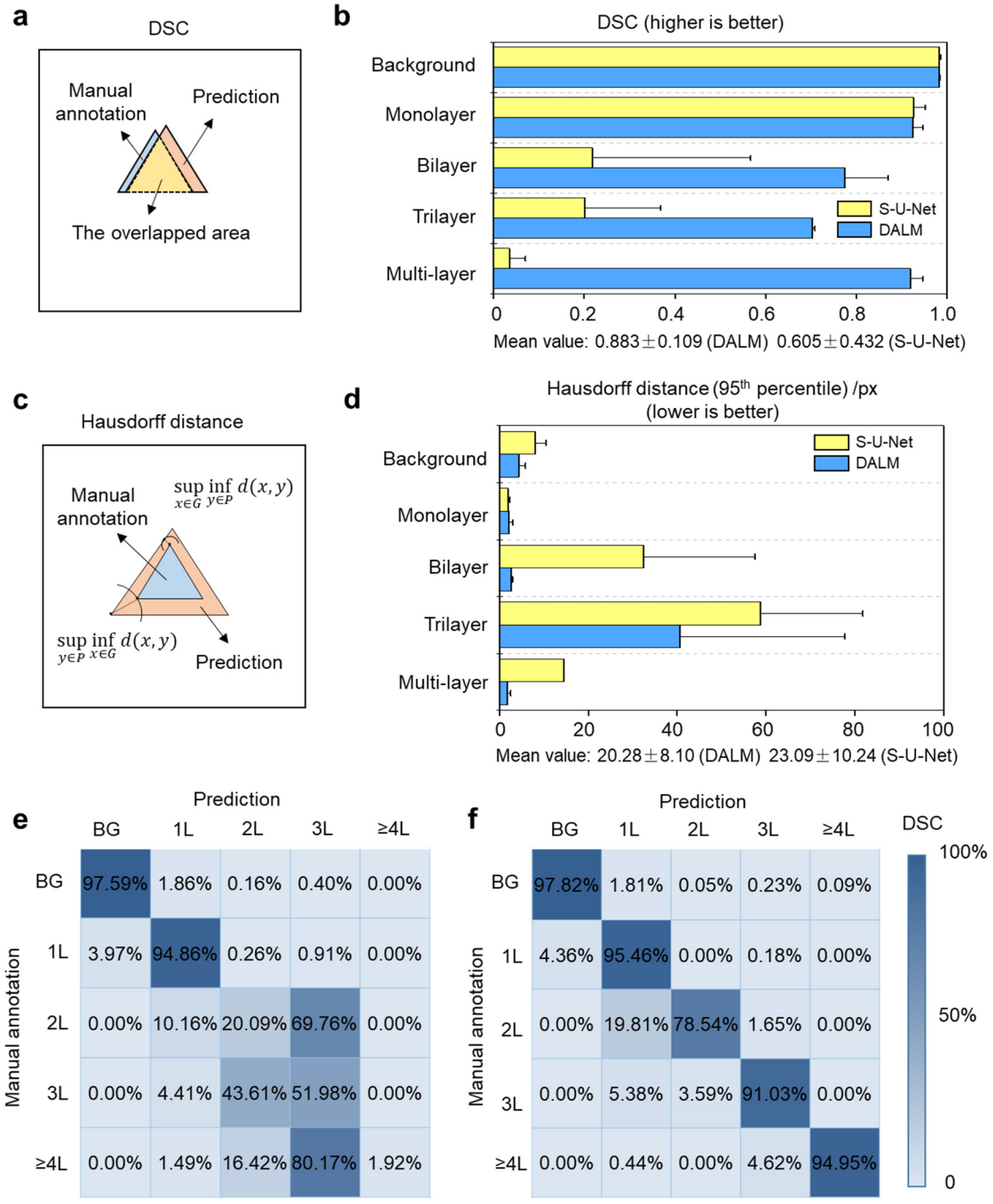


Figure 4. DSC, HD95, and confusion matrices for atomic layer prediction using the test data sets. (a) Diagram of DSC. (b) Calculated DSC values of both models. (c) Diagram of Hausdorff distance. (d) Evaluation results of HD95. Confusion matrices of S-U-Net (e) and DALM (f) using averaged values of testing samples. BG, background; 1L, monolayer; 2L, bilayer; 3L, trilayer; ≥4L, multilayer. The row indices are the manual annotation subclasses, and the column indices are the prediction subclasses. For example, the value 4.36% (row 2, column 1) in (f) means 4.36% of the monolayer manual annotations are predicted as the background category, while 19.81% (row 3, column 2) in (f) means 19.81% of the bilayer manual annotations are predicted as the monolayer category. The values on the diagonal are the DSC values of successful predictions, while the values off the diagonal are the DSC values of misclassified results.

The predicted distribution maps of each subclass (bi-, tri-, multilayer) had limited overlapping regions with the manual annotation, which was mainly due to the similar color contrast among these layers. DALM achieved 77.5%, 70.3%, and 91.9% accuracy when classifying bi-, tri-, and multilayer pixels into the correct subclasses. With the addition of hyperspectral images, which provided abundant spectral information, DALM had a higher performance to distinguish different layers, and the RGB images ensured the accurate localization of each subclass. The mean DSC values of DALM and S-U-Net were 88.3% and 60.5%, respectively, which indicated that the overall performance increased by 28% using imagery fusion.

Hausdorff distance was calculated to assess the localization accuracy of the prediction. Hausdorff distance is defined as the longest distance between one point of a set to all the points of the other set:

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

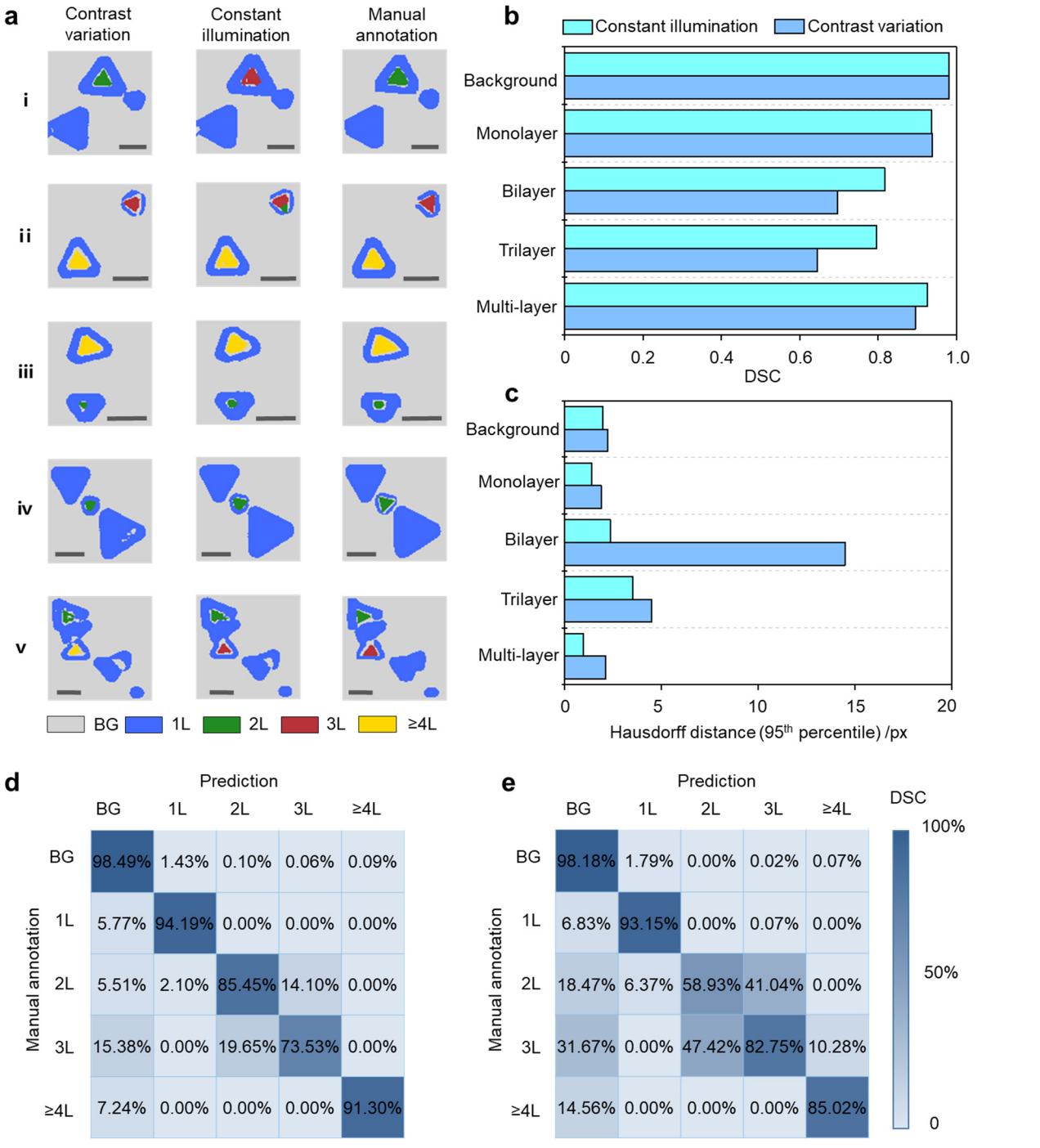


Figure 5. Leave-one-sample-out evaluation. (a) Segmentation results of the same five regions in Figure 3 for comparison, under conditions of contrast variation (10%) and constant illumination. Scale bar = 20 μm. (b) DSC values and (c) HD95, under both constant illumination and 10% contrast variation conditions. Confusion matrices of constant illumination (d) and 10% contrast variation (e) using median values of testing samples.

$$H(G, P) = \max\{\sup_{x\in G}\inf_{y\in P} d(x, y), \sup_{y\in P}\inf_{x\in G} d(x, y)\} \quad (3)$$

where G is the manual annotation set and P is the prediction set, while x and y are the points from both sets, respectively.

In practice, to eliminate the influence of outliners, the 95th percentile of the Hausdorff distance (HD95) was used instead of the maximum (100th percentile). As the HD95 represented the absolute distance between two point sets, large values indicated that the predicted flake area was far from the labeled area. Figure 4c shows a diagram of Hausdorff distance, and Figure 4d illustrates the calculated values of both models. For S-U-Net, the values of background and monolayer classes were small, which implied a good prediction compared with the target data. However, the distances for bi-, tri-, and multilayer predictions had large values of 32.411, 58.776, and 14.488 px, respectively. For DALM, the distance values for background, monolayer, bilayer, and multilayer subclasses were 4.353, 2.051, 2.562, and 1.751 px, respectively. Only the trilayer subclass had a large value of 40.708 px, which was mainly due

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

to one wrongly predicted area in Figure 3b. The overall mean HD95 of DALM was less than S-U-Net, indicating a higher localization accuracy of DALM when classifying pixels.

To understand how the pixels of one subclass were misclassified to another one, the confusion matrices of both models were calculated to visualize the performance of classification. Each row of the matrix represented the instances in a manually annotated class, while each column indicated the instances in a predicted class. The confusion matrix allows for visualizing whether the network is confusing about two or more subclasses. Since both models performed well in background and monolayer segmentation, bi-, tri-, and multilayer subclasses were studied. In the confusion matrix of S-U-Net (Figure 4e), only 20.09% of bilayer pixels were classified correctly, and 69.76% of bilayer pixels were misclassified as trilayer. In addition, 43.61% of trilayer pixels were misclassified as the bilayer subclass and 80.17% of multilayer pixels were incorrectly classified as the trilayer subclass. From the confusion matrix of DALM (Figure 4f), the misclassified pixels only account for a small proportion (around 5%), except one case in which approximately 20% bilayer pixels were misclassified as the monolayer subclass. The S-U-Net model exhibited confusion among bi-, tri-, and multilayer subclasses, while DALM was more intelligent in accurately discerning these subclasses. It is worthwhile to compare the confusion matrix of S-U-Net to previous reports of layer identification (we note that the confusion matrix of DALM was not compared because no reports have been published): in a recent report where flakes with 2−6 layers were grouped into one subclass, the VGG16 network reached 99% and 74% accuracy in segmenting background and monolayer subclasses, respectively, and 61% in segmenting the few-layer (2−6 layers) subclass.[35] In the present work, if 2 or 3 layers are combined as one subclass, a significantly higher accuracy of ≥90% can be achieved (DSC values of 97.82%, 95.46%, 87.41%, and 94.95% for background, monolayer, 2 or 3 layers, and ≥4 layers subclasses; see Figure S11).

Model Generalizability Analysis. To evaluate the predictive performance of DALM to generalize to illumination and contrast variation, which is essential for practical applications, we perform leave-one-sample-out for cross-validation. Specifically, we used the sample IDs to split the data set into training, validation, and test sets. There were 13 different samples available. In each split, we used data from 10 samples for training, 2 samples for validation, and the data from the remaining sample for testing. This procedure was repeated until all of the samples were used for testing. We did leave-one-sample-out in two scenarios: (a) 20% hyperspectral illumination variation, constant RGB illumination; (b) 20% hyperspectral illumination, 10% contrast variation in RGB images.

For comparison, only the five regions that were previously employed as testing pairs (Figure 3) are shown in Figure 5a. The image segmentation results of the monolayer were mostly consistent with the manual annotation, while mistaken predictions happened for the bi-, tri-, and multilayer in both leave-one-sample-out predictions. For example, the bilayer region (green) in Figure 5a(i) was misidentified as trilayer in the first leave-one-sample-out operation (20% hyperspectral illumination variation, constant RGB illumination) and was correctly identified in the second leave-one-sample-out operation (20% hyperspectral illumination variation, 10% RGB contrast variation). The five shown images were chosen as a comparison but are not necessarily representative of all the other samples. Therefore, the statistical results based on all of the samples are illustrated in Figure 5b,c. In the constant RGB illumination case, the median DSC was 98.1% (substrate), 93.7% (monolayer), 81.7% (bilayer), 79.6% (trilayer), and 92.6% (multilayer) (Figure 5b), showing the predictive stability compared to the previous values (blue in Figure 4b). At the same time, in the 10% RGB contrast variation case, the DSC values for substrate (98.1%) and monolayer (93.9%) were close to those of the constant illumination case, and a decrease in bilayer (69.7%), trilayer (64.5%), and multilayer (89.4%) occurred (Figure 5b). HD95 (blue in Figure 4d) was more stable in the leave-one-sample-out evaluation, while contrast variation (10%) increased the deviation when classifying bilayer pixels (Figure 5c). The confusion matrices of both leave-one-sample-out calculations were acquired using the median values. The predictions of the constant illumination case (Figure 5d) were highly consistent with the previous results (Figure 4f), while for the 10% contrast variation case, the uncertainty increased when identifying bilayer and trilayer regions (Figure 5e), indicating that bilayer and trilayer were the most confusing subclasses. Through this analysis, cross-validation showed that DALM had a stable statistical performance; a constant illumination condition when capturing RGB microscope images was important to ensure a higher rate of accurate identification, especially for bilayer and trilayer flakes.

By leveraging both 3D hyperspectral microscope images and 2D RGB microscope images, DALM distinguishes and segments MoS_2 flakes with monolayer to multilayer thickness with high accuracy, which is difficult when solely using RGB microscope images. DALM possessed both advantages of two inputs with high spatial and spectral resolution. Although the RGB-based approach had a low performance for bi-, tri-, and multilayer identification, it could be useful for specific applications in a first screening process. For example, this network can be used for a rough classification of flakes with three categories including monolayer, few-layer (2−10 layers), and bulk flakes. As only 2D images were involved, the acquisition of data sets and the model training can be advantageous over DALM. For applications where accurate atomic layer numbers of flakes are desired, DALM outperforms the RGB-based approach by a large margin. The performances of DALM are based on our in-house hyperspectral imaging system. MoS_2 flakes on 270 nm SiO_2/Si substrates were used for the proof-of-concept demonstration. DALM can be further implemented to other 2D material compositions, different substrates or SiO_2 thicknesses, and different preparation methods including mechanical exfoliation. However, each type will require corresponding training sets with manual labeling. Transfer learning techniques that can make use of existing pretrained models and reduce the effort of data labeling would be promising for creating a generalized intelligence for the identification of 2D flakes with single-atomic-layer accuracy across different samples and substrate types.[45] Taking $MoSe_2$ as an example, only a small amount of training data sets of CVD-deposited $MoSe_2$ is required to fine-tune the parameters of the DALM model which was trained on CVD-deposited MoS_2 samples. Similarly, models can also be developed for MoS_2 samples on the SiO_2/Si substrate with different oxidation thicknesses. DALM can be potentially trained and developed as a tool for materials identification based on the dramatic spectral changes of different materials as

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

a function of thickness. Furthermore, transfer learning provides the possibility for research groups to make use of the shared pretrained models from different institutes and modalities. We have made DALM an open-source tool for research use in the community.

The robustness evaluation of the deep fusion network includes two matches: First, in the dual-modality data acquisition process, hyperspectral microscope images were obtained with 20% illumination variations, where RGB images were acquired under constant illumination conditions. The DSC performance of DALM (88.3%) was higher than S-U-Net (60.5%). Second, in the cross-validation process (leave-one-sample-out), to test the DALM ability to generalize to independent data, RGB images with 10% contrast variations were employed. The overall accuracy decreased from 89.1% (constant illumination) to 83.1% (10% contrast variation), mainly due to the misclassification of bilayer and trilayer pixels.

The current dual-modality setup can be further improved. The hyperspectral data sets and RGB images were acquired from different modalities. The dual-modality data acquisition step is time-consuming, although the network can predict layer maps in a few seconds after training. Therefore, dual modalities can be combined as a hybrid system with two cameras capturing the high-spatial-resolution color images and high-spectral-resolution hyperspectral images simultaneously. In this case, the time for data acquisition and data preprocessing can be largely reduced. The additional measurement time of hyperspectral data sets was 30 s for a region of 80 × 200 μm^2. In this work, 251 channels of hyperspectral images covering a wavelength range of 507−689 nm were extracted and employed for layer number identification since exciton peaks existed in this range and changed with layer numbers, which provided abundant information for layer number identification by the network. The computational time may be further reduced without degrading performance if the hyperspectral channels were well constrained by either reducing the wavelength range while retaining the spectral resolution (*e.g.*, a wavelength range only covering the A exciton peak) or reducing the spectral resolution while retaining the wavelength range (*e.g.*, dimensionality reduction using principal component analysis[46]).

CONCLUSION

To conclude, DALM, as a multimodal multiclass segmentation model, was developed to fuse RGB images (high spatial resolution) and hyperspectral images (high spectral resolution) for the identification and segmentation of MoS_2 flakes with mono-, bi-, tri-, and multilayer thickness. It reached an overall high accuracy (>80%) compared to the state-of-the-art RGB-based approach (~60%) for the identification of all layer categories. Evaluation of robustness and generalization of DALM showed a stable predictive performance to illumination (20%) and contrast (10%) variations. However, constant illumination conditions in the dual-modality data acquisition process were essential to ensure a high success rate, especially for bilayer and trilayer identification.

METHODS

Optical and Hyperspectral Data Acquisition. The samples were measured by both the optical microscope (Leica) and the custom-built line-scan hyperspectral microscope.[36] The region of interest was selected under the optical microscope, and the RGB images were captured. The same region of interest was positioned under the hyperspectral reflection microscope, and the hyperspectral contrast data sets were acquired by automatically scanning the region. The hyperspectral microscope uses a broadband LED light source working at 470−850 nm. To achieve high-speed scanning, the hyperspectral microscope was designed to work in line-scan mode, which means that the spectrum of all the pixels along the line-shaped area (5 × 80 μm^2) can be captured at one frame. The control software of the hyperspectral system was based on MATLAB. In this work, the parameters were controlled as follows: 100 μm/s scanning speed, 5 μm step size, and 0.1 s stage waiting time for camera capture. The scanning time for a region of 80 × 200 μm^2 was 30 s, with a high spectral resolution of 0.728 nm/px. To reduce the dimensions of the data set and extract the important spectral information, 251 channels (507−689 nm) were extracted from the original 1004 channels (325−1056 nm). The dimension reduction of the original hyperspectral data set required less computation resources. The photographs of the optical modalities for data acquisition can be found in Figure S1, and the required components are summarized in Table S1. To enrich the acquired data set, the illumination intensity had a 20% variation during hyperspectral measurements.

Data Smoothing, Background Subtraction, and Wavelength Selection of Hyperspectral Data Sets. A hyperspectral imaging microscopy data set can be expressed by

$$p_k = [x_k, y_k, I_k(\lambda_1, \lambda_2, ..., \lambda_n)] \quad (4)$$

where x_k and y_k are the positions of one pixel and I_k is the radiation intensity with the variation of wavelength λ. Gaussian smoothing was applied to the hyperspectral data set to lower spectrum noise. The influence of inhomogeneous illumination was reduced by background subtraction using the following operation:

$$p_k^{\text{intensity}}(x, y, \lambda) = p_k(x, y, \lambda) - p_k^0(x, y, \lambda) \quad (5)$$

where $p_k^{\text{intensity}}(x,y,\lambda)$ is the reflected intensity with the subtraction of inhomogeneous intensity distribution of illumination light. $p_k(x,y,\lambda)$ is the measured spectrum from the sample. $p_k^0(x,y,\lambda)$ is the averaged reflection spectrum of the bare substrate under inhomogeneous illumination.

Registration of RGB and Hyperspectral Images. In this work, feature-based algorithms were employed for dual-modality image registration based on the correspondence between image features such as points, lines, and contours. Due to the high spatial resolution of the optical microscope, RGB images were referred to as the fixed image (reference image), while hyperspectral images were referred to as the moving image. The geometric transformation was applied to hyperspectral images to be aligned. The whole image registration (in the MATLAB environment) can be divided into three steps. First, the corresponding points between moving and fixed images were selected. Second, the transformation was determined according to the corresponding points. Third, the geometrical transformation was applied to the moving images. By an image processing toolbox, the corresponding points between two images were selected through the control point selection tool (Figure S2). The geometric transformation was defined by a rule where the point with Cartesian coordinates (x, y) was mapped to another point with Cartesian coordinates (u, v). The affine transformation used in this work can be described as

$$[u \quad v] = [x \quad y \quad 1]T \quad (6)$$

where T is 3-by-3 transformation matrix, which can be determined through the selected points.

Data Labeling. Figure S3 shows the graphical user interface (GUI) for data labeling, where the regions of four subclasses (mono-, bi-, tri-, and multilayer) are shown in different colors in the optical images. The regions without labeling were set as the fifth subclass of substrate. The pixels belonging to different subclasses were labeled with different values. For example, the background pixels were labeled with 0, the monolayer pixels were labeled with 1, the bilayer pixels were labeled with 2, and so forth. Figure S4 illustrates a diagram of optical RGB images and the pixel values after labeling. After this

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

process, pixels of all five subclasses had labeled values as a classification. The labels were converted into one-hot images to obtain the binary target masks in the network (Figure S5). Each subclass had a corresponding one-hot image after this conversion.

Min−Max Intensity Normalization. The intensity value of 8-bit RGB images ranges from 0 to 255, while the intensity of hyperspectral data ranges from 0 to 20. To fuse the features from both inputs and balance the attributions from small and large values, min−max normalization was used to rescale the intensity range of both modalities to [0, 1]. The formulation was as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{7}$$

where x is the original value of one pixel, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of all the pixel values in one image, and x' is the normalized image matrix.

Data Augmentation. Data augmentation techniques included flip, rotation, scaling, cropping, translation, and contrast changes of the original data sets. At the beginning of the training, the augmented data were considered as different images, which could be used to improve the robustness of the network and prevent overfitting during the training process. In this work, operations including flipping and rotation were conducted for both hyperspectral and RGB images to enlarge the training data sets in the training stage. In the cross-validation process (leave-one-sample-out), contrast variation (10%) of RGB images was employed to equip the network with the desired invariance (Figure S7).

Feature Fusion Strategy. Due to dimension differences between 3D and 2D features, a feature fusion block was used to combine them optimally. The core of the feature fusion block was to reduce the dimension of 3D features to 2D and to fuse them in a squeeze-and-excitation (SE) block. The dimensions of 3D and 2D features could be expressed as $H \times W \times D \times C$ and $H \times W \times C$, where $H \times W$ represents the feature dimensions of height and width, D represents the depth of 3D features, and C represents feature channels. To transform the dimension of 3D features into 2D, channel C of 3D features was compressed into 1 using the 3D convolution ($1 \times 1 \times 1$). The dimension of channel-compressed 3D features was further processed through squeezing and the 2D convolution (3×3) to $H \times W \times C$, which was consistent with the dimension of 2D features (Figure S8a).

Before the feature concatenation, a SE block acquired the importance of respective feature channels, enhanced the useful features, and suppressed the less important ones (Figure S8b). The global pooling layer was used as the squeeze operation. Features were compressed along the spatial dimension into $1 \times 1 \times C$. That means each 2D feature channel was turned into a real number, which to some extent had a global receptive field. To reduce parameters and computational complexity, the feature channel was reduced from C to C/r, where r denotes the reduction ratio. After a ReLu activation layer and a fully connected layer, the feature channel could be ascended to C again. A sigmoid layer was used for the excitation operation. The weight could be generated for each feature channel through a parameter w, which was learned to describe the correlation between feature channels. Through sigmoid, the normalized weights between 0 and 1 could be obtained. Finally, through a scale operation, the normalized weights could be weighted to the features of each channel.[47]

Batch Normalization. In this deep fusion network, the batch normalization was added after convolution operations to normalize the values in the hidden layers. The whole batch normalization operation can be divided into four steps. The input of a batch in a layer was $X = [x_1, x_2, \cdots, x_n]$, where x_i represents one sample and n is batch size. First, the mean value μ_B of the elements in the minibatch can be calculated by

$$\mu_B = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

The variance σ_B^2 of the mini-batch was calculated by

$$\sigma_B^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_B)^2 \tag{9}$$

Each element can be normalized by

$$x_i' = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{10}$$

where ε, an arbitrarily small constant, was added in the denominator for numerical stability.

Finally, the data can be rescaled and shifted to recover the original distribution. The transformation step of batch normalization followed as

$$y_i = \gamma_i x_i' + \beta_i \tag{11}$$

where γ and β are two trainable parameters to each layer, and therefore, the normalized output (y) was multiplied by a standard deviation (γ) with an addition of a mean value (β).

Computation Complexity and Network Training. All of the experiments are conducted on a GNU/Linux server running Ubuntu 18.04, with 64 GB RAM memory. The number of trainable parameters in DALM with dual-modalilty input is 8 815 495. The algorithm was trained on a single Nvidia RTX Titan GPU with 24 GB memory. It takes around 3 h to train a single model for 100 epochs on a training set containing 800 data pairs after data augmentation. In this test stage, it takes only two seconds to generate a segmentation mask of one sample with the same GPU, while it costs 30 s when using an Intel Xeon CPU, which is practically useful when GPU is not available. When training the single-modal U-Net with 7 766 117 parameters after excluding the 3D network, it takes 2 h for training on the same RGB data. An appropriate parameter setting is crucial to the successful training of DALM. We stop the training by observing the training loss and the Dice score on a validation set over epochs. Thus, we choose a number of 100 epochs to avoid overfitting and to keep a low computational cost. The batch size was empirically set to 4 considering the GPU memory and the learning rate was set to 0.0002 throughout all of the experiments by observing the training stability.

Data Postprocessing. False predictions contained noisy points, which could influence the network performances such as DSC, Hausdorff distance, and confusion matrix. To eliminate the noise points, postprocessing was carried out to improve the overall performance of the network. Connected component analysis, which is usually used in computer vision to detect and count the number of connected regions,[48] was employed to eliminate the noise points of prediction results.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsnano.0c09685.

Photographs of the optical microscope and the hyperspectral imaging microscope for data acquisition; GUI for dual-modality image registration and layer subclass labeling; labeled data and pixels with different values showing different layer numbers; principle of image conversion from labeled data to one-hot images of different layer numbers; data dimension variation from raw data to data pairs suitable for network training; data augmentation of RGB images; dimension fusion strategy of DALM; optical RGB image, normalized hyperspectral images, and normalized spectra of one data pair for demonstration; training curves of both models; confusion matrix of the network prediction using a rough classification of different subclasses; component list of the hyperspectral imaging microscope (PDF)

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

AUTHOR INFORMATION

Corresponding Author

Xingchen Dong − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany;* orcid.org/0000-0001-6734-7568; Email: xingchen.dong@tum.de

Authors

Hongwei Li − *Department of Computer Science, Technical University of Munich, 85748 Garching, Germany*

Zhutong Jiang − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Theresa Grünleitner − *Walter Schottky Institut and Physik Department, Technische Universität München, 85748 Garching, Germany*

İnci Güler − *Walter Schottky Institut and Physik Department, Technische Universität München, 85748 Garching, Germany*

Jie Dong − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Kun Wang − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Michael H. Köhler − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Martin Jakobi − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Bjoern H. Menze − *Department of Computer Science, Technical University of Munich, 85748 Garching, Germany*

Ali K. Yetisen − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Ian D. Sharp − *Walter Schottky Institut and Physik Department, Technische Universität München, 85748 Garching, Germany;* orcid.org/0000-0001-5238-7487

Andreas V. Stier − *Walter Schottky Institut and Physik Department, Technische Universität München, 85748 Garching, Germany*

Jonathan J. Finley − *Walter Schottky Institut and Physik Department, Technische Universität München, 85748 Garching, Germany*

Alexander W. Koch − *Institute for Measurement Systems and Sensor Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsnano.0c09685

Author Contributions

[#]X.D., H.L., and Z.J. contributed equally to this work.

Author Contributions

X.D. conceived the idea. X.D. conducted the experiments with samples fabricated by T.G. H.L. and Z.J. developed the neural network architecture, training, testing, and analysis. X.D. and Z.J. wrote the manuscript supported by H.L. İ.G. and A.V.S. participated in the discussion of the project progress. All authors commented on the manuscript. X.D., H.L., and Z.J. contributed equally to this work.

Notes

The authors declare no competing financial interest.

The data and codes that support the conclusions of this study are available on https://github.com/hongweilibran/DALM.

ACKNOWLEDGMENTS

X.D., J.D., and K.W. acknowledge support of China Scholarship Council (CSC) (201706050026, 201706030161, and 201808340074). J.J.F., I.D.S., A.V.S., and T.G. gratefully acknowledge the DFG for funding *via* the Cluster of Excellence e-conversion (EXC2089) and the TUM International Graduate School of Science and Engineering (IGSSE). J.J.F. and A.V.S. also acknowledge support of DFG *via* the Cluster of Excellence MCQST (EXC2111), as well as the projects SPP2244, FI947/8, and INST95-1496-1. Furthermore, we gratefully acknowledge the BMBF for financial support *via* the projects Q.LinkX and MARQUAND.

REFERENCES

(1) Geim, A. K.; Grigorieva, I. V. van der Waals Heterostructures. *Nature* **2013**, *499*, 419−425.

(2) Novoselov, K. S.; Mishchenko, O. A.; Carvalho, O. A.; Neto, A. C. 2D Materials and van der Waals Heterostructures. *Science* **2016**, *353*, 9439.

(3) Mak, K. F.; Lee, C.; Hone, J.; Shan, J.; Heinz, T. F. Atomically Thin MoS_2: A New Direct-Gap Semiconductor. *Phys. Rev. Lett.* **2010**, *105*, 136805.

(4) Xia, F.; Wang, H.; Xiao, D.; Dubey, M.; Ramasubramaniam, A. Two-Dimensional Material Nanophotonics. *Nat. Photonics* **2014**, *8*, 899−907.

(5) 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.* **2012**, *7*, 699−712.

(6) Frisenda, R.; Navarro-Moratalla, E.; Gant, P.; De Lara, D. P.; Jarillo-Herrero, P.; Gorbachev, R. V.; Castellanos-Gomez, A. Recent Progress in the Assembly of Nanodevices and van der Waals Heterostructures by Deterministic Placement of 2D Materials. *Chem. Soc. Rev.* **2018**, *47*, 53−68.

(7) Kum, H.; Lee, D.; Kong, W.; Kim, H.; Park, Y.; Kim, Y.; Baek, Y.; Bae, S. H.; Lee, K.; Kim, J. Epitaxial Growth and Layer-Transfer Techniques for Heterogeneous Integration of Materials for Electronic and Photonic Devices. *Nat. Electron.* **2019**, *2*, 439−450.

(8) Kim, Y.; Cruz, S. S.; Lee, K.; Alawode, B. O.; Choi, C.; Song, Y.; Johnson, J. M.; Heidelberger, C.; Kong, W.; Choi, S.; Qiao, K. Remote Epitaxy through Graphene Enables Two-Dimensional Material-Based Layer Transfer. *Nature* **2017**, *544*, 340−343.

(9) Tonndorf, P.; Schmidt, R.; Böttger, P.; Zhang, X.; Börner, J.; Liebig, A.; Albrecht, M.; Kloc, C.; Gordan, O.; Zahn, D. R.; de Vasconcellos, S. M. Photoluminescence Emission and Raman Response of Monolayer MoS_2, $MoSe_2$, and WSe_2. *Opt. Express* **2013**, *21*, 4908−4916.

(10) Tonndorf, P.; Schmidt, R.; Schneider, R.; Kern, J.; Buscema, M.; Steele, G. A.; Castellanos-Gomez, A.; van der Zant, H. S.; de Vasconcellos, S. M.; Bratschitsch, R. Single-Photon Emission From Localized Excitons in an Atomically Thin Semiconductor. *Optica* **2015**, *2*, 347−352.

(11) Arora, A.; Noky, J.; Drüppel, M.; Jariwala, B.; Deilmann, T.; Schneider, R.; Schmidt, R.; Del Pozo-Zamudio, O.; Stiehm, T.; Bhattacharya, A.; Krüger, P. Highly Anisotropic In-Plane Excitons in Atomically Thin and Bulklike 1 T′-$ReSe_2$. *Nano Lett.* **2017**, *17*, 3202−3207.
(12) Hsu, C.; Frisenda, R.; Schmidt, R.; Arora, A.; De Vasconcellos, S. M.; Bratschitsch, R.; van der Zant, H. S.; Castellanos-Gomez, A. Thickness-Dependent Refractive Index of 1L, 2L, and 3L MoS_2, $MoSe_2$, WS_2, and WSe_2. *Adv. Opt. Mater.* **2019**, *7*, 1900239.
(13) Li, X. L.; Han, W. P.; Wu, J. B.; Qiao, X. F.; Zhang, J.; Tan, P. H. Layer-Number Dependent Optical Properties of 2D Materials and Their Application for Thickness Determination. *Adv. Funct. Mater.* **2017**, *27*, 1604468.
(14) Nolen, C. M.; Denina, G.; Teweldebrhan, D.; Bhanu, B.; Balandin, A. A. High-Throughput Large-Area Automated Identification and Quality Control of Graphene and Few-Layer Graphene Films. *ACS Nano* **2011**, *5*, 914−922.
(15) Berkdemir, A.; Gutiérrez, H. R.; Botello-Méndez, A. R.; Perea-López, N.; Elías, A. L.; Chia, C. I.; Wang, B.; Crespi, V. H.; López-Urías, F.; Charlier, J. C.; Terrones, H. Identification of Individual and Few Layers of WS_2 Using Raman Spectroscopy. *Sci. Rep.* **2013**, *3*, 1−8.
(16) Dhakal, K. P.; Duong, D. L.; Lee, J.; Nam, H.; Kim, M.; Kan, M.; Lee, Y. H.; Kim, J. Confocal Absorption Spectral Imaging of MoS_2: Optical Transitions Depending on the Atomic Thickness of Intrinsic and Chemically Doped MoS_2. *Nanoscale* **2014**, *6*, 13028−13035.
(17) Ferrari, A. C.; Meyer, J. C.; Scardaci, V.; Casiraghi, C.; Lazzeri, M.; Mauri, F.; Piscanec, S.; Jiang, D.; Novoselov, K. S.; Roth, S.; Geim, A. K. Raman Spectrum of Graphene and Graphene Layers. *Phys. Rev. Lett.* **2006**, *97*, 187401.
(18) Millard, T. S.; Genco, A.; Alexeev, E. M.; Randerson, S.; Ahn, S.; Jang, A. R.; Shin, H. S.; Tartakovskii, A. I. Large Area Chemical Vapour Deposition Grown Transition Metal Dichalcogenide Monolayers Automatically Characterized through Photoluminescence Imaging. *npj 2D Mater. Appl.* **2020**, *4*, 1−9.
(19) Crovetto, A.; Whelan, P. R.; Wang, R.; Galbiati, M.; Hofmann, S.; Camilli, L. Nondestructive Thickness Mapping of Wafer-Scale Hexagonal Boron Nitride down to a Monolayer. *ACS Appl. Mater. Interfaces* **2018**, *10*, 25804−25810.
(20) Blake, P.; Hill, E. W.; Castro Neto, A. H.; Novoselov, K. S.; Jiang, D.; Yang, R.; Booth, T. J.; Geim, A. K. Making Graphene Visible. *Appl. Phys. Lett.* **2007**, *91*, 063124.
(21) Li, H.; Wu, J.; Huang, X.; Lu, G.; Yang, J.; Lu, X.; Xiong, Q.; Zhang, H. Rapid and Reliable Thickness Identification of Two-Dimensional Nanosheets Using Optical Microscopy. *ACS Nano* **2013**, *7*, 10344−10353.
(22) Ouyang, W.; Liu, X. Z.; Li, Q.; Zhang, Y.; Yang, J.; Zheng, Q. S. Optical Methods for Determining Thicknesses of Few-Layer Graphene Flakes. *Nanotechnology* **2013**, *24*, 505701.
(23) Ni, Z. H.; Wang, H. M.; Kasim, J.; Fan, H. M.; Yu, T.; Wu, Y. H.; Feng, Y. P.; Shen, Z. X. Graphene Thickness Determination Using Reflection and Contrast Spectroscopy. *Nano Lett.* **2007**, *7*, 2758−2763.
(24) Golla, D.; Chattrakun, K.; Watanabe, K.; Taniguchi, T.; LeRoy, B. J.; Sandhu, A. Optical Thickness Determination of Hexagonal Boron Nitride Flakes. *Appl. Phys. Lett.* **2013**, *102*, 161906.
(25) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84−90.
(26) Moen, E.; Bannon, D.; Kudo, T.; Graf, W.; Covert, M.; Van Valen, D. Deep Learning for Cellular Image Analysis. *Nat. Methods* **2019**, *16*, 1233−1246.
(27) 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*, IEEE, June 24−27, 2014; Columbus, OH, USA: 2014; pp 581−587.
(28) Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation. *Appl. Soft Comput.* **2018**, *70*, 41−65.
(29) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436−444.
(30) Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, June 8−10, 2015; Boston, MA, USA, 2015; pp 3431−3440.
(31) Masubuchi, S.; Machida, T. Classifying Optical Microscope Images of Exfoliated Graphene Flakes by Data-Driven Machine Learning. *npj 2D Mater. Appl.* **2019**, *3*, 1−7.
(32) Greplova, E.; Gold, C.; Kratochwil, B.; Davatz, T.; Pisoni, R.; Kurzmann, A.; Rickhaus, P.; Fischer, M. H.; Ihn, T.; Huber, S. D. Fully Automated Identification of Two-Dimensional Material Samples. *Phys. Rev. Appl.* **2020**, *13*, 064017.
(33) Saito, Y.; Shin, K.; Terayama, K.; Desai, S.; Onga, M.; Nakagawa, Y.; Itahashi, Y. M.; Iwasa, Y.; Yamada, M.; Tsuda, K. Deep-Learning-Based Quality Filtering of Mechanically Exfoliated 2D Crystals. *npj Comput. Mater.* **2019**, *5*, 1−6.
(34) Masubuchi, S.; Watanabe, E.; Seo, Y.; Okazaki, S.; Sasagawa, T.; Watanabe, K.; Taniguchi, T.; Machida, T. Deep-Learning-Based Image Segmentation Integrated with Optical Microscopy for Automatically Searching for Two-Dimensional Materials. *npj 2D Mater. Appl.* **2020**, *4*, 1−9.
(35) Han, B.; Lin, Y.; Yang, Y.; Mao, N.; Li, W.; Wang, H.; Yasuda, K.; Wang, X.; Fatemi, V.; Zhou, L.; Wang, J. I. J. Deep-Learning-Enabled Fast Optical Identification and Characterization of 2D Materials. *Adv. Mater.* **2020**, *32*, 2000953.
(36) Dong, X.; Yetisen, A. K.; Tian, H.; Güler, I.; Stier, A. V.; Li, Z.; Köhler, M. H.; Dong, J.; Jakobi, M.; Finley, J. J.; Koch, A. W. Line-Scan Hyperspectral Imaging Microscopy with Linear Unmixing for Automated Two-Dimensional Crystals Identification. *ACS Photonics* **2020**, *7*, 1216−1225.
(37) Dong, X.; Dong, J.; Yetisen, A. K.; Köhler, M. H.; Wang, S.; Jakobi, M.; Koch, A. W. Characterization and Layer Thickness Mapping of Two-Dimensional MoS_2 Flakes via Hyperspectral Line-Scanning Microscopy. *Appl. Phys. Express* **2019**, *12*, 102004.
(38) Ghamisi, P.; Höfle, B.; Zhu, X. X. Hyperspectral and LiDAR Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3011−3024.
(39) Palsson, F.; Sveinsson, J. R.; Ulfarsson, M. O. Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 639−643.
(40) Guo, Z.; Li, X.; Huang, H.; Guo, N.; Li, Q. Deep Learning-Based Image Segmentation on Multimodal Medical Imaging. *IEEE Trans. Radiat. Plasma Med. Sci.* **2019**, *3*, 162−169.
(41) Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C. W.; Heng, P. A. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Trans. Med. Imag.* **2018**, *37*, 2663−2674.
(42) Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features with 3D Convolutional Networks. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE, Dec 7−13, 2015; Santiago, Chile, 2015; pp 4489−4497.
(43) Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Oct 5−9, 2015; Munich, Germany, 2015; pp 234−241.
(44) Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, PMLR, Apr 11−13, 2011; Gordon, G., Dunson, D., Dudík, M., Eds.; Fort Lauderdale, FL, USA, 2011; pp 315−323.
(45) Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In *International Conference on*

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151

Artificial Neural Networks; Springer, Oct 4−7, 2018; Rhodes, Greece, 2018; pp 270−279.
(46) Du, Q.; Fowler, J. E. Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis. *IEEE Geosci. Remote. Sens. Lett.* **2007**, *4*, 201−205.
(47) Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE, June 18−22, 2018; Salt Lake City, UT, USA, 2018; pp 7132−7141.
(48) Dillencourt, M. B.; Samet, H.; Tamminen, M. A General Approach to Connected-Component Labeling for Arbitrary Image Representations. *J. Assoc. Comput. Mach.* **1992**, *39*, 253−280.

https://dx.doi.org/10.1021/acsnano.0c09685
ACS Nano 2021, 15, 3139−3151