A versatile deep learning architecture for classification and label-free prediction of hyperspectral images

Bryce Manifold, Shuaiqian Men, Ruoqian Hu and Dan Fu

Hyperspectral imaging is a technique that provides rich chemical or compositional information not regularly available to traditional imaging modalities such as intensity imaging or colour imaging based on the reflection, transmission or emission of light. Analysis of hyperspectral imaging often relies on machine learning methods to extract information. Here we present a new flexible architecture—the U-within-U-Net—that can perform classification, segmentation and prediction of orthogonal imaging modalities on a variety of hyperspectral imaging techniques. Specifically, we demonstrate feature segmentation and classification on the Indian Pines hyperspectral dataset and simultaneous location prediction of multiple drugs in mass spectrometry imaging of rat liver tissue. We further demonstrate label-free fluorescence image prediction from hyperspectral stimulated Raman scattering microscopy images. The applicability of the U-within-U-Net architecture on diverse datasets with widely varying input and output dimensions and data sources suggest that it has great potential in advancing the use of hyperspectral imaging across many different application areas ranging from remote sensing, to medical imaging, to microscopy.

Computer vision techniques based on deep learning have recently demonstrated myriad novel applications in many disciplines. With the continuous improvement and availability of advanced computing hardware and open-source methods, deep learning is finding broader use in a wide variety of imaging, sensing and biophotonics research. The flexibility of deep learning for image processing enables facile adoption of existing frameworks for many different imaging modalities such as transmitted-light microscopy, fluorescence microscopy, X-ray imaging, magnetic resonance imaging and many more. The images from such techniques are often passed to a deep learning algorithm to perform tasks such as classifying diseases, segmenting spatial features, improving image quality or predicting alternate imaging modalities; however, the majority of work done so far performs deep learning on monospectral images. Such monospectral images contain only a single intensity value at each pixel; that is, there is no spectral information inherent to the imaging technique such as in black-and-white photography, X-ray imaging or magnetic resonance imaging. Contrary to monospectral images are multispectral and hyperspectral images, where multiple spectral components of a field of view can be depicted in their own image. We take multispectral to be a subset of hyperspectral, specifically pertaining to images that contain relatively few spectral channels (for example, RGB imaging). Hyperspectral imaging combines spectroscopy and imaging such that each pixel of the image contains a wide spectral profile that allows for detailed characterization.

Linear decomposition, phasor analysis, support vector machines and other machine learning methods have indeed been used to analyse hyperspectral imaging datasets. Although many of these techniques have demonstrated promising results, such methods may suffer from limited generalizability or information loss, limiting their ultimate performance. Deep learning, by contrast, potentially offers a method for learning on the basis of both spectral and spatial signatures—as well as their nonlinear interplay—which allows for improved performance in a variety of hyperspectral imaging analysis tasks; however, techniques for analysing these hyperspectral stacks face unique challenges in computer vision research. For example, standard deep learning architectures that work for monospectral images (consisting of two or three spatial dimensions) may not work for hyperspectral stacks due to the extra dimension needed for spectral information. Frameworks such as Mayerich and colleagues’s stainless staining or Behrmann and co-workers’s work in mass spectrometry imaging (MSI) address this by interpreting the spectra at individual pixels of hyperspectral images to produce excellent results in label-free prediction and classification; however, they may be missing contextual information from spatial convolutions of the whole image. A recently published work by Zhang and colleagues bypasses the need for spectral deep learning by using machine learning to interpret spectral information and create truth maps to which spatial deep learning of images can be trained. Other frameworks for hyperspectral deep learning based on spectral–spatial convolutions also exist but are often rigid, only performing a particular task such as binary pixel or multiclass label classification. Moreover, to the best of our knowledge, a convolutional framework for predicting entirely alternate imaging modalities (where the final number of spectral channels is unlikely to match the input, but spatial resolution is maintained) from hyperspectral images has not yet been reported. We thus present a new architecture, the U-within-U-Net (UwU-Net), to address these current shortcomings in hyperspectral deep learning and improve the utility of hyperspectral imaging techniques.

The UwU-Net architecture presented here is based on the U-Net architecture originally developed by Ronneberger and colleagues, where a specialized autoencoder encodes and decodes spatial feature information in an input image to reconstruct some new output image. The U-Net separates itself from a traditional autoencoder in the recontextualization of information through concatenations at equivalent encode–decode levels (marked as dark blue arrows in Fig. 1a). This eliminates the discarding of information, as in a traditional autoencoder. Although the original work was concerned...
with image segmentation, the U-Net has seen use in a variety of applications including segmentation, label-free prediction and denoising\(^{30-34}\); however, most works that utilize the U-Net in this way are not concerned with images that contain multiple spectral channels. The original U-Net is generally not applicable to hyperspectral images as the architecture is dedicated to encoding multiple spatial feature channels starting from a single spatial channel image (see Fig. 1a). The typical two-dimensional kernel of a U-Net is thus not well suited for hyperspectral stacks, which have a third tensor dimension dedicated to spectral channels. A three-dimensional kernel could potentially be used, but then the spatial and spectral information are mixed during the feature encoding in a problematic fashion for image reconstruction\(^{35}\). Modification of input and output layers to match spectral dimensions is often useful in the multispectral regime but may be too facile of a change to adequately handle spectrally complex hyperspectral images. Although other recently reported modifications to the U-Net have also shown improvements with respect to the original U-Net on semantic segmentation and classification of remote sensing datasets (some of which involve multispectral datasets)\(^{36-38}\), we report a robust architecture for multiple hyperspectral imaging tasks.

To create a hyperspectral deep learning architecture with the robustness and features of the traditional U-Net, we have amended the U-Net architecture such that spectral channel information is handled by a separate U-structure outside of an arbitrary number of traditional spatial U-Nets (see Fig. 1b). This UwU-Net architecture allows dedication of tunable free parameters to both spectral information (outer U) and spatial information (inner Us). The architecture's parameters can be empirically tuned to change the spectral layer depth, number of spatial Us at the centre or output spectral size based on the dataset. Here we demonstrate the utility of this new architecture in three different tasks on three different types of hyperspectral imaging: feature segmentation and classification on the high-altitude hyperspectral imaging Indian Pines dataset; prediction of monoisotopic drug location within rat livers using mass spectrometry images; and label-free prediction of cellular organelle fluorescence in stimulated Raman scattering (SRS) microscopy.

The first task concerns segmentation and classification of the Indian Pines dataset, which depicts a scene of farmland in north-west Indiana across a large range of wavelengths spanning the ultraviolet to short infrared region (400–2,500 nm)\(^{39}\). The publicly available dataset was acquired by the airborne visible/infrared imaging spectrometer and provides a model task for hyperspectral deep learning, that is, the segmentation and classification of various crop and foliage types. The broad spectrum and spatial heterogeneity of the scene demonstrates a deep learning algorithm's ability to correctly identify and segment features on the basis of both spectral signatures and spatial positions. Moreover, the use of this dataset by previous work in hyperspectral deep learning allows for comparison of our proposed architecture\(^{40-42}\).

The second task concerns predicting drug location in a model rat liver tissue sample from MSI, a powerful technique that provides spatially resolved, highly specific chemical information in the form of molecular ion masses. Where most deep learning computer vision work is centred around interpretation of optical images, MSI is particularly interesting to approach with deep learning as it has an enormous spectral dimension that provides highly specific, but difficult to interpret in situ chemical information\(^{43,44}\). Most MSI work follows from traditional linear decomposition and analysis that is well-developed and ubiquitous in mass spectrometry\(^{45-49}\). Deep learning has been demonstrated for MSI datasets\(^{46,47,50,51}\), but has been chiefly used for spectral dimensionality reduction or interpretation. To the best of our knowledge, the simultaneous interpretation of spatial and spectral information using convolutional deep learning in MSI has yet to be reported. We demonstrate one way the UwU-Net architecture could be used in MSI by simultaneously
predicting the highly specific monoisotopic peak locations of twelve drugs from low-mass-resolution binned images.

Finally, the third task demonstrates the capability of the UwU-Net to perform label-free prediction of fluorescence images from SRS microscopy images. Stimulated Raman scattering microscopy is a hyperspectral imaging technique where molecular vibrational bonds are coherently interrogated by two ultrashort laser pulses. Although the vibrational information afforded by SRS microscopy can be specific to a given molecule, there are often many overlapping contributions to vibrational signals that confound image interpretation. In this work, we show that the specificity of SRS microscopy can be improved by deep learning to predict fluorescence images that are highly specific to an organelle. We also show that the trained algorithms can be multiplexed to create label-free cell organelle images in live cells.

Indian Pines classification
To demonstrate the flexibility of the UwU-Net and to validate its capability to classify an arbitrary number of features from hyperspectral images, a 1-U UwU-Net (where there is 1 spatial U-Net at the centre of the architecture) and 17-U UwU-Net (where there are 17 spatial U-Net's at the centre) were trained to classify the Indian Pines AVARIS dataset. The hyperspectral images consist of 200 spectral channels (where 20 of the original 220 bands have been removed due to water absorption) across a broad range of wavelength lengths (400–2,500 nm) with 144×144 pixel images (cropped from 145×145 to be compatible with the spatial U-Nets) at each wavelength. The images contain a high-altitude two mile by two mile field of view of farmland in northwest Indiana. The ground-truth images consist of non-mutually exclusive hand-drawn maps of the various crops and foliage depicted in the field of view. In total, there are 16 classifications shown in Fig. 1c and listed in Table 1. Here the UwU-Net is trained to predict a 17×144×144 image stack (16 classifications plus an unused background) from the 200×144×144 input image stack. The initial 200 channels are first reduced via convolution to 100 channels and then to the final 1 (for the 1-U UwU-Net) or 17 (for the 17-U UwU-Net) before spatial learning. The output predicted images are thresholded to create a binary map to compare against the ground-truth image. Looking at the results in Table 1, the 17-U UwU-Net performs well with nearly all classifications exceeding 99% accuracy. The exception is the classification of an unutilled corn field in the upper left of the field of view that is instead identified as a mixture of the three soybean classifications. We also note the prediction of crops at the top-middle, top-right and bottom of the field of view. Although these areas contribute to the error, we note that crops do exist in these parts of the hyperspectral images (as seen in the composite image in Fig. 1c) but are unidentified in the hand-drawn truth maps. To better reflect the model's performance, especially in these cases, counts of false-positive and negative pixels, and the intersection over union (IOU) for each class, are provided in Extended Data Fig. 6. The overall accuracy (99.48±0.50%), however, is in concordance with state-of-the-art architectures for hyperspectral classification on the Indian Pines dataset. The classification accuracies for three of these architectures—ResNet, Multi-Path ResNet (MPRN) and Auxiliary Capsule GAN (AU-Caps-GAN)—are shown in Table 1 for comparison with the 17-U UwU-Net, which demonstrates the highest accuracy. We note that the 1-U UwU-Net (with its more modest modifications to the original U-Net) performs worse than the other models, which suggests that the extra spatial parameters afforded by the parallel U-Nets at the centre of the UwU-Net contribute towards a more accurate model. For further comparison, a basic U-Net (where the initial and final layers have been simply adjusted to accommodate the desired input/output channel number) was also trained; however, it was unable to classify any of the labels properly, which suggests that UwU-Net spectral layers are critical for proper identifications. A representative example of one of the basic U-Net's errant classifications is shown in Extended Data Fig. 1. These results demonstrate the UwU-Net's ability to simultaneously segment and classify features from hyperspectral images with high accuracy; however, the UwU-Net is not limited to a binary pixel classification, like some hyperspectral architectures.

Table 1 | Classification accuracy of the Indian Pines dataset

| Label                        | UwU-Net (1-U) | ResNet41 | MPRN41 | AU-Caps-GAN42 | UwU-Net (17-U) |
|------------------------------|--------------|----------|--------|---------------|----------------|
| Alfalfa                      | 97.40        | 98.33    | 98.89  | 99.15         | 99.96          |
| Corn (untilled)              | 93.66        | 99.28    | 99.51  | 99.50         | 98.57          |
| Corn (minimal till)          | 95.98        | 98.80    | 98.92  | 99.12         | 99.19          |
| Corn                         | 98.83        | 98.20    | 98.52  | 98.34         | 99.78          |
| Grass (pasture)              | 97.60        | 97.97    | 97.92  | 98.70         | 99.48          |
| Grass (trees)                | 98.26        | 98.80    | 99.08  | 99.42         | 99.80          |
| Grass (mowed pasture)        | 99.98        | 100      | 98.18  | 98.74         | 99.98          |
| Hay (windrowed)              | 97.65        | 100      | 100    | 99.27         | 99.91          |
| Oats                         | 99.90        | 97.50    | 97.50  | 98.68         | 99.98          |
| Soybeans (untilled)          | 96.35        | 97.99    | 98.14  | 98.45         | 99.18          |
| Soybeans (minimal till)      | 79.37        | 99.27    | 99.38  | 99.12         | 98.49          |
| Soybeans (clean till)        | 97.15        | 98.35    | 98.69  | 98.34         | 99.23          |
| Wheat                        | 99.50        | 99.14    | 98.90  | 98.69         | 99.93          |
| Woods                        | 94.87        | 99.88    | 99.98  | 99.33         | 99.18          |
| Buildings (grass/trees/drives)| 98.07        | 99.55    | 99.68  | 99.41         | 99.18          |
| Stone-steel tower            | 99.74        | 94.52    | 96.44  | 98.94         | 99.91          |
| Overall accuracy              | 96.52±4.7    | 99.01    | 99.16  | 99.12±0.25    | 99.48±0.50     |

The individual and overall classification accuracies of the Indian Pines dataset from various hyperspectral deep learning models are shown, as well as the presented UwU-Net models. Note that the ResNet, MPRN and AU-Caps-GAN models are reported as produced in their respective references; the ResNet and MPRN classifications were reported without uncertainties. Reported uncertainties refer to the s.d. among the n=16 classifications.
mixtures of diluted drugs, where each mixture contains some components to be seen at a time, whereas binning sacrifices the hallmark for interpretability. Windowing allows for only a few mass combinations at a time (a window containing all monoisotopic drug peaks) to be viewable. Note that both windowing and binning sacrifice information for interpretable information. Windowing allows for only a few mass combinations to be seen at a time, whereas binning sacrifices the hallmark specificity of mass spectrometry. Analysis of these large datasets can also be cumbersome, taking potentially hours or longer to interpret per dataset.

The work we present here demonstrates a potential solution to this information trade-off issue by predicting high-mass-resolution drug location images (corresponding to each drug’s monoisotopic peak) from a window of hyperspectral low-resolution binned mass images of the spiked rat liver tissue. Specifically, the region of 330–630 m/z (a window containing all monoisotopic drug peaks) was binned into 1 m/z images and concatenated into a hyperspectral image stack. The 0.001 m/z resolution images corresponding to the monoisotopic peaks of the twelve drugs (as determined in the previous publication) were then isolated from the raw MSI data and concatenated to produce a stack where each image corresponds to a specific drug. The UwU-Net architecture was trained to predict twelve drug images from the 300-channel hyperspectral images. Figure 2 shows the results of these predictions and the corresponding 1 m/z bin image that contains the monoisotopic peak. Although some of these low-mass-resolution bins are already highly correlated with the specific monoisotopic peak (for example, ipratropium and vatalanib in Fig. 2a and Fig. 2b, respectively), other images have strong background contributions and conflicting drug spot signal due to fragment peaks from other drugs (for example, erlotinib and gefitinib in Fig. 2c and Fig. 2f, respectively). From Fig. 2, it is apparent that the deep learning algorithm is able to reliably predict each drug’s location from the low-resolution hyperspectral data even when there are conflicting background/fragment peaks or when the drug concentration is low (as with lapatinib and trametinib in Fig. 2k and Fig. 2l, respectively). Even in trametinib, where the drug

**Fig. 2 | Mass spectrometry images of drug-spikes rat liver slice.** a–l. Each row shows (from left to right) a 1 m/z bin image from the input 300 image hyperspectral stack that contains a given drug’s monoisotopic peak, the 5-U UwU-Net predicted 0.001 m/z bin image of the drug and the 0.001 m/z bin image specific to that drug’s monoisotopic peak. The following drugs are depicted in their respective panels: ipratropium (a), vatalanib (b), erlotinib (c), sunitinib (d), pazopanib (e), gefitinib (f), sorafenib (g), dasatinib (h), imatinib (i), dabrafenib (j), lapatinib (k) and trametinib (l). Scale bar, 4 mm.
### Table 2 | Quality metric values for the MSi dataset predictions

| Drug (m/z) | t/m_bin | U-Net (1-U, non-HS) | U-Net (1-U, only drug bins) | U-Net (1-U) | U-Net (5-U) | UwU-Net (12-U) | UwU-Net (1-U) | PCC | NrMSE |
|-----------|---------|--------------------|-----------------------------|-------------|-------------|----------------|----------------|------|-------|
| 894.177   | 0.014   | 0.99              | 0.13                        | 0.78        | 0.57        | 0.67           | 0.44           | 0.78 | 0.76  |
| 899.220   | 0.009   | 0.97              | 0.08                        | 0.67        | 0.39        | 0.53           | 0.44           | 0.78 | 0.76  |
| 938.171   | 0.025   | 0.94              | 0.25                        | 0.68        | 0.32        | 0.44           | 0.44           | 0.78 | 0.76  |
| 1047.160  | 0.023   | 0.99              | 0.80                        | 0.44        | 0.24        | 0.36           | 0.44           | 0.78 | 0.76  |
| 1052.143  | 0.022   | 0.96              | 0.25                        | 0.68        | 0.32        | 0.44           | 0.44           | 0.78 | 0.76  |
| 1265.086  | 0.011   | 0.99              | 0.25                        | 0.68        | 0.32        | 0.44           | 0.44           | 0.78 | 0.76  |

To better understand the role of spectral versus spatial learning and their interplay on model accuracy, multiple basic U-Nets were trained on a single drug at a time. Here the single 1/m/z bin image and corresponding high-mass-resolution peak image were used for training. Although some of the drugs are correctly identified and predicted (suggesting spatial learning of a single image from the hyperspectral stack may drive some predictions of drugs), many of the drugs (sunitinib, gefitinib, sorafenib, dabrafenib and trametinib) go partially or entirely unpredicted. A single basic U-Net modified to accept 300 channels and output twelve channels again produces unacceptable results (Extended Data Fig. 1). The use of a UwU-Net with a single spatial U-Net at its centre (denoted as 1-U in Table 2) allows for spectral learning of the data in addition to spatial learning. When a stack of just the twelve drug 1/m/z bins is used for training (1-U, only drug bins category in Table 2), only gefitinib, dabrafenib and trametinib were unidentified. The use of the full 300 hyperspectral stack in the 1-U UwU-Net shows further improvement leaving only one spot of dabrafenib unpredicted. This suggests that additional spectral information improves the accuracy of the model in drugs where spatial information from the principal bins is insufficient for prediction. The use of a 12-U UwU-Net on the full hyperspectral data eliminates any unidentified drug spots, but errantly predicts spots in sunitinib and initinib that do not exist in the respective truth images. A 5-U UwU-Net demonstrates the most accurate prediction of drug spots with no missing or errantly predicted spots for any of the twelve drugs (as seen in Fig. 2). This analysis and comparison suggest that—similar to depth in a traditional U-Net or ResNet—architecture parameters such as the spectral depth or number of spatial U-Nets at centre can be empirically tuned to improve model accuracy.

These results highlight the capability of the UwU-Net to mine MSI datasets for relevant features from both spatial and rich spectral features afforded in MSI in a convolutional manner. One way this is potentially useful for MSI is in the design and execution of experiments. If a priori ground-truth information is available (in this case, the masses of the drug molecules sought, their locations and their concentrations), a UwU-Net model can be trained and utilized in other similar experiments to vastly improve analysis speed. For example, although the training of this algorithm took ~8h, the final prediction of all images shown took only ~1s. This upfront single-time investment of training then affords analysis of further samples to be performed extremely quickly in comparison with costly linear analysis of each dataset. The specific demonstration presented here could also be highly useful for the miniaturization of MSI systems for in situ use where the trade-off of reduced mass-to-charge resolution would be mitigated by a pretrained algorithm. We also note the possibility of combining MSI with an orthogonal method such as fluorescence or Raman imaging, to predict alternate imaging modalities using the UwU-Net as we demonstrate below.

### Label-free organelle prediction from SRS microscopy images

Label-free prediction via deep learning has been a recent area of interest for augmenting the information acquired from a given microscopy modality40. The label-free prediction usually involves a microscopy image such as transmitted-light or autofluorescence...
microscopy being converted to an image that mimics a more complex label-requisite modality such as fluorescent or histologically stained images [15,16]. The value of this type of work is clear due to the elimination of staining protocols and the disadvantages associated with labelling the sample (photobleaching, toxicity, disruption of biological structures or functions and so on); however, the quality of label-free prediction depends heavily on the information present in the input images [17]. For example, although transmitted-light microscopy is relatively simple to perform, it only reveals information based on light scattering due to differences in refractive index. In the context of cells and their organelles, there may not be a substantial enough difference between an organelle and cytosol to produce relevant information for a deep learning algorithm to reliably predict a corresponding organelle’s fluorescence.

In comparison with simple bright field or autofluorescence imaging, Raman imaging is a much more information-rich, label-free alternative. The Raman spectrum of a sample reflects specific molecular vibrations quantitatively associated with the molecules within. Hyperspectral Raman imaging improves conventional Raman imaging by significantly speeding up the image acquisition by three to four orders of magnitude [18,19]. Regardless of the acquisition method, for biological samples, Raman spectra are often congested and highly convolved due to the overlapping Raman bands from many different molecules. Principle component analysis and phasor analysis have been used to extract individual organelles from the myriad vibrational signatures in a cell [15,16]; however, the subtle variations of Raman spectra for individual organelles present significant challenges to the analysis of smaller structures such as mitochondria and endoplasmic reticulum. Previous attempts to produce label-free staining based on hyperspectral Raman imaging have shown promising results for some organelles but not as rich of predictions for smaller ones [19]. The architecture we present here shows improved fluorescence prediction across three organelles. Deep learning using the rich spectral and spatial information afforded by hyperspectral Raman microscopy also outperforms previous work of label-free prediction from transmitted-light microscopy [15]. As shown in Fig. 3a–c, we create label-free prediction algorithms for nuclei, mitochondria and endoplasmic reticuli fluorescence in fixed lung cancer cells (A549, from ATCC). The accuracy of the predictions is quantified in Table 3 by Pearson’s correlation coefficient (PCC), the normalized root mean squared error (NRMSE) and feature similarity index (FSIM) [20,21]. Across all computed quality metrics, we find high correlation and acceptably low error between predicted images and their respective truths. Previous work reported PCC values of 0.58, 0.69 and 0.70 for DNA (nuclei), mitochondria and endoplasmic reticuli, respectively [15]. We thus see a considerable improvement in label-free organelle prediction with the information-rich hyperspectral SRS microscopy in comparison with bright field microscopy. A basic U-Net was again trained for comparison, as seen in Extended Data Fig. 1. Although this task was more successful than in the previous demonstrations, unacceptable residual SRS features were also present in the image. For additional comparisons with another modern architecture used for image reconstructions, a U-Net utilizing ResNet Blocks [22,23] was also trained to predict the organelles (Extended Data Figs. 2 and 7). Although the Res-U-Net showed slightly improved organelle predictions in comparison with previously reported results, the UwU-Net predictions still outperformed across all organelles and metrics.

The utilization of both spectral and spatial information is paramount towards demonstrating the utility of this architecture. This is most clearly demonstrated in the mitochondria prediction model by the differentiation of the organelle from lipid droplets in the cell. In SRS images, lipid droplets appear as bright dots typically ~1 μm in size. This means that they have a similar size and

**Fig. 3 | Predicted organelle fluorescence from hyperspectral SRS microscopy images.** All SRS images shown depict only the peak signal image from the hyperspectral stack. a–c, Input SRS (left), ground-truth fluorescence (middle) and predicted fluorescence (right) are shown for nuclei (a), mitochondria (b) and endoplasmic reticuli (c). d, A typical cellular SRS spectrum (black) and the ten vibrational transitions imaged and used for prediction (red) are shown. Note that the transitions marked in red represent the centre of a band of probed transitions with a resolution of 19 cm⁻¹. The 15 cm⁻¹ steps between each spectral image means the entire CH vibrational region is effectively probed during hyperspectral imaging. e, An SRS image of live cells (left) that contain no dye, each algorithm’s predicted fluorescence (right, top row) and fluorescence images taken after the cells are stained (right, bottom row). f, An overlaid combination of each organelle prediction (left) and the same group of cells after staining (right). Scale bar, 25 μm.
**Table 3 | Quality metric values for the label-free prediction of organelle fluorescence**

| Organelle model         | PCC   | NRMSE | FSIM  |
|-------------------------|-------|-------|-------|
| Nucleus                 | 0.92 ± 0.03 | 0.047 ± 0.022 | 0.89 ± 0.04 |
| Mitochondria            | 0.84 ± 0.05 | 0.059 ± 0.019 | 0.93 ± 0.02 |
| Endoplasmic Reticulum   | 0.94 ± 0.02 | 0.038 ± 0.016 | 0.92 ± 0.03 |

The table shows PCC, NRMSE and FSIM values for the three organelles predicted from hyperspectral SRS images. The values shown are based on the average of all withheld test images (nine, nine and seven images for the nuclei, mitochondria and endoplasmic reticuli, respectively) of 512×512 pixels. Uncertainty refers to the s.d. among the withheld test images.

Discussion

In this work we have presented UwU-Net, a new architecture for deep learning using hyperspectral images. The architecture is highly flexible in both the types of tasks it can perform (for example, classification, segmentation, label-free prediction) and the types of hyperspectral images with which it is compatible (for example, remote sensing, MSI and SRS microscopy). Specifically, we show excellent performance of Indian Pines classification with 99.48% overall accuracy for all classifications. We also demonstrate successful drug location prediction in fixed tissue from MSI data from windowed and binned images. This highlights the capability to mine spectrally dense MSI datasets using both spectral and spatial information and offers new possibilities for deep learning in MSI. Finally, we show improved label-free prediction of organelle fluorescence by using hyperspectral SRS microscopy. We note a significant improvement in nuclear, mitochondrial and endoplasmic reticulum prediction correlation with respect to previous work by the use of the UwU-Net to interpret spectral and spatial information.

We further note that although all models were trained using randomized starting parameters and stochastic gradient descent to minimize mean squared error between output and truth images, the architecture is easily amenable to transfer learning methods and more complex error functions for particular tasks. We also note that the UwU-Net architecture can potentially be used in a generative adversarial network framework to perform an even broader class of tasks. However, generative adversarial network training of a UwU-Net is not feasible currently given memory constraints.

Finally, although only a subset of tasks and imaging techniques are demonstrated here, we expect the UwU-Net to be broadly applicable or adaptable to any reasonably designed computer vision task involving a hyperspectral imaging technique with potential use in medical imaging, microscopy and remote sensing.

Methods

The following are the methods for the label-free fluorescence prediction demonstration experiments and utilization of the UwU-Net algorithm. The methods for the publicly available datasets (Indian Pines and the MSI of Spiked Rat Liver) are briefly discussed above and details of their experimental parameters can be found in their respective original publications.

Cell sample preparation. A549 cells were cultured in ATCC F-12K medium with 10% foetal bovine serum at 37 °C with 5% CO₂ atmosphere. Cells were seeded on coverslips 24 h before imaging. Fixed cells were first dyed then fixed using 2% paraformaldehyde. Live cells were first mounted, imaged with SRS and then stained for fluorescence imaging. The fluorescent dyes used were Hoechst 33342, MitoTracker Red CMXRos and ER-Tracker Green for nucleus, mitochondria and endoplasmic reticulum respectively. All dye protocols were based on the provided instructions from the manufacturer.

Simultaneous SRS and fluorescence microscopy. Stimulated Raman scattering microscopy was performed on a homebuilt SRS microscope as described previously. Briefly, an Insight DeepSee+ provides synchronized 799 nm and 1040 nm laser pulses that are passed through high-density glass and a grating stretcher pair, respectively, to control pulse chirp. The 1040 nm beam is modulated by an electro-optical modulator and polarizing beam splitter to operate in the stimulated Raman loss mode. A time delay of the 1040 nm beam was controlled by a computer-controlled Z8006CA X-PMQ2P DE-522-KX1A-DM delay stage. Both pulses are combined on a dichroic mirror before being directed through the microscope by a pair of scanning galvo mirrors. The microscope is a Nikon Eclipse FN1 equipped with a 40× 1.15 numerical aperture (NA) objective. The 800 and 1040 nm laser powers were set to 20 mW at focus for both beams in all experiments. Light passed through the sample is collected by a 1.4 NA condenser, filtered by a 700 nm longpass filter (to remove fluorescence light) and 1000 nm shortpass filter (to remove the 1040 nm light) and finally collected on a homebuilt photodiode connected to a Zurich Instruments HF2LI lock-in amplifier. Two-photon fluorescence signal from the photomultiplier tube was collected simultaneously using ScanImage. Images were acquired with 512×512 pixels and a pixel dwell time of 8 μs at each of the ten vibrational transitions as noted in Fig. 3d. It is noted that the transitions noted in Fig. 3d represent only the center of the probe band with 19 cm⁻¹ spectral resolution. This means that at the step size of ~15 cm⁻¹ per image in the stack, the full CH region is probed during hyperspectral imaging.

UwU-Net functional description. An input hyperspectral stack of dimensions (L, X, Y) is first passed to the architecture. Here, L represents the number of input channels of the hyperspectral stack (for example, 200 for Indian Pines, 300 for MSI drug location prediction or ten for SRS images) and X and Y are the number of spatial pixels in the image (in all cases here X = Y). The stack is first reduced to (X, Y) in the channel dimension, where L > M, with a 3×3 kernel convolution of stride 1 over all L channels followed by a batch normalization and rectified linear unit activation function. The new stack is then reduced once more in the channel dimension by the same process to a stack of (N, X, Y) where N is the desired final number of spatial tuning channels. The stack is then split at the channel dimension (if N > 1) such that there are now N number of (X, Y) images. Each of these images is passed to its own U-Net for spatial feature learning as described previously.

The resulting N number of images from each spatial U-Net are then reconcatenated...
in the channel dimension to reform a \((N, X, Y)\) stack. This \((N, X, Y)\) stack is then concatenated in the channel dimension to the \((N, X, Y)\) stack from before splitting (mimicking the recovery of information as in the traditional U-Net) to form a stack \((N \times 2, X, Y)\). This stack is reduced to \((O, X, Y)\) by a 3 x 3 kernel convolution of stride 1 over the 2N channels followed by a batch normalization and rectified linear unit activation function. This predicted stack is then compared to the truth stack (also of dimension \([O, X, Y]\)) for a mean squared error which is calculated for all channels and parameters are tuned in a backpropagating fashion.

Training parameters, data preparation and hardware. The models trained and shown in this paper were developed and built using the pytorch-fnet framework originally developed by Ounkomol and co-workers\(^{33}\). All models were trained using the pytorch\_fnet default parameters with a few exceptions. The models were trained using randomized starting parameters on batches of randomized patches from the entire dataset. The model parameters are tuned per resolution and are kept as a randomized descent manner based on minimization of mean squared error. The pytorch-fnet framework utilizes an Adam optimizer with a 0.001 learning rate and beta values of 0.5 and 0.999. The rat liver drug prediction model was trained only for 23,000 iterations due to the satisfactory prediction accuracy and long training iterations. The original dataset was split into six training datasets and two test datasets. Final predictions were recoloured by rotations and flips with the original dataset withheld for testing. This equated to six training datasets and two test datasets. Final predictions were recoloured for each channel and then overlaid into the shown prediction image (Fig. 1c). The UwU-Net reports in Table 1 use 1 (U-1) or 17 (17-U) spatial U-Nets at their centre during training.

The rat liver MSI dataset was first prepared by saving the 330–630 m/z window at 1 m/z bins from the raw data using Databox Explorer. All 300 images were concatenated into a TIF stack using Fiji. The monoisotopic images at 0.001 m/z resolution were then saved for each drug using Databox Explorer following the m/z peaks and appropriate full-width at half-maximum bins as noted by Eriksson and colleagues\(^{36}\). The twelve drug peak images were concatenated into a TIF stack using Fiji. Both stacks were padded with zeros in Fiji from their native 247 x 181 pixel size to 256 x 256 pixels to be compatible with the spatial U-Nets within the UwU-Net architecture. Training data here were also augmented by rotations and flips with the original dataset withheld for testing; there were seven datasets used for training. The shown 1 m/z bin, truth peak and predicted peak images for the drugs were normalized, contrast adjusted to the same level and coloured using the Red Hot Fiji lookup table. The UwU-Net reports in Table 1 use one (U-1), five (5-U) or 17 (17-U) spatial U-Nets at their centre.

The simultaneously collected SRS and fluorescence images were first separated into two respective TIF stacks. The SRS stack was used as is for training and prediction. The fluorescence stacks were averaged to a single image and used as the truth for training and prediction. The fixed cell nucleus, mitochondria and endoplasmic reticulum models utilized 43, 46 and 35 images, respectively, with a random 80/20% train-test split for each model. Images predicted by the model were normalized, contrast adjusted to the same level, then coloured using the mpl-inferno, Cyan, Green and Magenta Fiji lookup tables for SRS, nucleus, mitochondria and endoplasmic reticulum, respectively. All model development, training and prediction as well as image processing was performed on a homebuilt machine running Ubuntu 18.04. The machine is equipped with an AMD 2950X processor, Nvidia Titan RTX graphics processing unit, 64 GB memory and a 2 TB solid state drive. All dependency software versions were based on the pytorch\_fnet requirements. On our machine, trainings for Indian Pines, rat liver drug and organelle models took ~4, ~8 and ~3h, respectively. In all models, prediction of individual test images took 1s or less.

Quantitative metrics. Prediction quality was assessed by overall accuracy, IOU, PCC, NRMSE and FSIM.

Overall accuracy is used to evaluate the binary pixel values assigned for each classification. Here, the number of errantly predicted pixels are counted, subtracted from the total number of pixels and then divided by the total number of pixels. A percentage score is reported here where accuracy closer to 100% indicates a more accurate prediction. Intersection over union also measures the segmentation and classification accuracy by taking the ratio of the intersection between predicted pixels and true pixels (that is, true positives plus false positives). The resulting ratio indicates how accurately the model segments and classifies areas where values closer to one indicate more accurate prediction.

PCC, NRMSE and FSIM. The PCC correlation coefficient is used to correlate the pixels of the truth and predicted images. The PCC is defined as the number of true pixels minus the sum of the absolute difference between truth and predicted pixels. The PCC is used to express the degree of linear correlation between two variables. PCC values range from -1 to 1, with 1 indicating a perfect positive linear correlation, 0 indicating no linear correlation, and -1 indicating a perfect negative linear correlation.

NRMSE is used to express the accuracy of a predicted pixel versus the same pixel in the truth image. Here a value closer to 0 indicates a more accurate prediction model.

The FSIM is calculated using the MATLAB code provided by Zhang and colleagues\(^{30}\), following the prescribed instructions.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Indian Pines dataset used can be found at https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html. The MSI dataset used can be found at https://www.ebi.ac.uk/pride/archive/projects/PXD016146. The Hyperspectral SRS and Fluorescence data used can be found at: https://doi.org/10.6894/m9.lighshare.13497138. Source data are provided with this paper.

Code availability

The original pytorch-fnet framework with traditional U-Net is available for download at https://github.com/AllenCellModeling/pytorch_fnet/tree/release_1. The code for the UwU-Net along with instructions for training can be found at https://github.com/B-Manifold/pytorch_fnet_UwUnet/tree/v1.0.0 (https://doi.org/10.5281/zenodo.4396327).

Received: 10 June 2020; Accepted: 27 January 2021; Published online: 11 March 2021

References

1. Litjens, G. et al. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–66 (2018).
2. Yuan, H. et al. Computational modeling of cellular structures using conditional deep generative networks. Bioinformatics 35, 2141–2149 (2019).
3. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. 25, 44–56 (2019).
4. Mittal, S., Stoean, C., Kajdacsy-Balla, A. & Bhargava, R. Digital assessment of stained breast tissue images for comprehensive tumor and microenvironment analysis. Front. Bioeng. Biotechnol. 7, 246 (2019).
5. Mukherjee, P. et al. A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets. Nat. Mach. Intell. 2, 274–282 (2020).
6. Pianykh, O. S. et al. Improving healthcare operations management with machine learning. Nat. Mach. Intell. 2, 266–273 (2020).
7. Varma, M. et al. Automated abnormality detection in lower extremity radiographs using deep learning. Nat. Mach. Intell. 1, 578–583 (2019).
8. Zhang, L. et al. Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy. Theranostics 9, 2541–2554 (2019).
9. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. Nat. Methods 15, 1090–1097 (2018).
10. Rana, A. et al. Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. JAMA Netw. Open 3, e2005111 (2020).
11. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. Cell 173, 792–803.e19 (2018).
12. Rajan, S., Ghosh, J. & Crawford, M. M. An active learning approach to hyperspectral data classification. IEEE Trans. Geosci. Remote Sens. 46, 1231–1242 (2008).
13. Melgani, F. & Bruzzone, L. Support vector machines for classification of hyperspectral remote-sensing images. In IEEE International Geoscience and Remote Sensing Symposium Vol. 1, 506–508 (IEEE, 2002).
14. Jahr, W., Schmid, B., Schmied, C., Fahrbach, F. O. & Huisken, J. Hyperspectral light sheet microscopy. Nat. Commun. 6, 7990 (2015).
ARTICLES
69. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. OnLine* **2**, 13 (2003).

**Acknowledgements**

We kindly thank G. Johnson and C. Ounlomol for the development and guidance on the pytorch_fnet framework. We also thank P. Horvatovich for his correspondence and public release of the MSI dataset used here. Finally, we thank B. Tardif, A. Hammon and A. Rokem for their helpful discussions. The work is supported by NSF CAREER 1846503 (D.F.), the Beckman Young Investigator Award (D.F.) and the NIH R35GM133435 (D.F).

**Author contributions**

B.M. was responsible for the conception, development, and utilization of the UwU-Net architecture. B.M. and D.F. conceived the demonstrations of the UwU-Net architecture. S.M. and R.H. were equally responsible for care and preparation of the cells used for imaging. B.M. performed the imaging experiments. B.M. and D.F. prepared the manuscript with contributions from all authors. D.F. supervised the research.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s42256-021-00309-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42256-021-00309-y.

Correspondence and requests for materials should be addressed to B.M. or D.F.

Peer review information *Nature Machine Intelligence* thanks Rohit Bhargava and the other, anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021
Extended Data Fig. 1 | Representative predictions facile U-Nets for Hyperspectral images. Panel a shows the Grass (Mowed Pasture) Indian Pines classification prediction with no thresholding. Panel b shows the prediction of Ipratropium from the MSI dataset. Panel c shows prediction of nuclear fluorescence from SRS images with contrast values set to mimic the images shown in Fig. 3. Panel d shows the same image as Panel c with higher contrast to demonstrate the U-Net’s inability to remove non-nucleus features. Panel e shows the UwU-Net prediction from Fig. 3a with high contrast demonstrating superior non-nuclear feature removal.
Extended Data Fig. 2 | Fluorescence predictions the Modified U-Net with ResNet blocks. Panel a shows nucleus fluorescence prediction. Panel b shows mitochondrial prediction. Panel c shows endoplasmic reticulum prediction. All truth fields of view are the same as in Fig. 3.
Extended Data Fig. 3 | Predicted Organelle fluorescence using traditional U-Net. Panel a shows prediction of nucleus fluorescence. Panel b shows prediction of mitochondrial fluorescence. Panel c shows prediction of endoplasmic reticulum fluorescence. We note the improper inclusion of lipid droplets in the mitochondria model and off nucleoli in both the mitochondria and endoplasmic reticulum models. The comparison between lipid droplets and mitochondria is further depicted in Extended Data Figure 4.
Extended Data Fig. 4 | Comparison of mitochondria prediction between UwU-Net and traditional U-Net. Panel a shows a zoomed in field of view from Fig. 1b where a UwU-Net is trained to predict mitochondrial fluorescence from a hyperspectral SRS stack. The shown input SRS only corresponds to the brightest image out of the 10-image hyperspectral stack. Normalized pixel values are plotted below each image corresponding to the drawn dashed lines. In the SRS image, a strong lipid droplet is found at ~1.4 μm but is properly removed during prediction of the mitochondria at ~1.8 μm and ~3 μm. Panel b shows a zoomed in field of view from Extended Data Figure 3b where a traditional U-Net is trained to predict mitochondrial fluorescence from a single SRS image. The normalized pixel value plots beneath each zoomed-in field of view show a marked difference in how lipid droplets are handled. Here the lipid droplets at ~0.8 μm and ~1.8 μm are not removed during prediction.
Extended Data Fig. 5 | UwU-Net predicted fluorescence in live-cell SRS imaging. Panel a shows prediction of nucleus fluorescence. Panel b shows prediction of mitochondrial fluorescence. Panel c shows prediction of endoplasmic reticulum fluorescence.
Extended Data Fig. 6 | The count of false positive, false negative pixels, and intersection over union (IOU) per class in the UwU-Net (17-U) Indian Pines model.
Extended Data Fig. 7 | Quality Metrics for Res-U-Net. PCC, NRMSE, and FSIM values for the Res-U-Net trained as in Extended Data Figure 2. The number of images for used for each calculation is the same as in Table 3. Uncertainty refers to standard deviation.

| Organelle Model          | PCC    | NRMSE  | FSIM   |
|--------------------------|--------|--------|--------|
| Nucleus                  | 0.74 ± 0.04 | 0.379 ± 0.016 | 0.76 ± 0.05 |
| Mitochondria             | 0.75 ± 0.15 | 0.172 ± 0.053 | 0.77 ± 0.06 |
| Endoplasmic Reticulum    | 0.72 ± 0.13 | 0.112 ± 0.023 | 0.78 ± 0.03 |
Extended Data Fig. 8 | Quality Metrics for Traditional U-Net Fluorescence Prediction. PCC metrics for the organelle fluorescence prediction models trained with a traditional U-Net using a single SRS image. While still highly correlated, we note the errant prediction of spurious features in Supplementary Figs 1 and 2.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about: availability of computer code

Data collection
- MATLAB 2016a
- Scanimage 2019a
- ZIControl 15.11.34730

Data analysis
- Fiji using ImageJ 1.52p
- Datacube Explorer Version 2.3
- FeaturesSIMM calculator from: DOI: 10.1109/TIP.2011.2109730
- U-Net pytorch/fnet DOI: 10.1038/s41592-018-0111-2
- UwU-Net code [DOI:10.5281/zenodo.4396327]

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about: availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The Indian Pines dataset used can be found at:
https://engineering.purdue.edu/~biehl/MultSpec/hyperspectral.html
The MSI dataset used can be found at:
https://www.ebi.ac.uk/pride/archive/projects/PXD016146
The Hyperspectral SRS and fluorescence data used can be found at:
DOI: 10.6084/m9.figshare.13497138

Field-specific reporting
Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.
☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design
All studies must disclose on these points even when the disclosure is negative.

| Sample size | 43, 46, and 35 images were used for the training of the nucleus, mitochondria, and ER in the fixed cell models. 49, 47, and 51 images were used for the same respective organelles in the live-cell models. These numbers were chosen to be above the previously reported >30 image diminishing return using the traditional U-Net for fluorescence prediction. |
| Data exclusions | Images were excluded from training if they contained no cells or cells that were out of focus. No images from the test dataset predictions were excluded during the shown quantitative analyses |
| Replication | In the nucleus and mitochondria models, datasets were from two different experimental days separated by a period of ~6 weeks where images appeared spectrally and visually the same indicating sufficient reproducibility in our cell preparation and SRS imaging setup. |
| Randomization | During training for the Indian Pines and mass spectrometry imaging datasets, the original orientation was withheld while the flips and rotations were used for training as described in the main manuscript. For the training of the organelle predictions, the datasets were randomly divided into 80%/20% train/test subsets that were used for training and testing respectively. For all datasets during training however, randomized patches of specified size are pulled from the training images meaning the training sessions are fully randomized to this degree |
| Blinding | For the Indian Pines and mass spectrometry imaging datasets, the small sample size and necessity to demonstrate prediction on the appropriately oriented images means a macro-level knowledge of the datasets used for training/testing. However the randomized patches pulled from these images mean there is blinding as to what the model is being trained to see in any given iteration. This is compounded in the organelle models where the picking of images for training/testing is also blinded. With respect to training, the algorithm does not utilize the withheld test images to influence training. |

Reporting for specific materials, systems and methods
We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a | n/a |
| ☒ Antibodies | ☐ ChiP-seq |
| ☒ Eukaryotic cell lines | ☐ Flow cytometry |
| ☐ Palaeontology and archaeology | ☐ MRI-based neuroimaging |
| ☒ Animals and other organisms |   |
| ☐ Human research participants |   |
| ☒ Clinical data |   |
| ☒ Dual use research of concern |   |

Eukaryotic cell lines
Policy Information about cell lines

Cell line source(s) | ATCC
Authentication | Cell lines were not authenticated
Mycoplasma contamination | Cell lines were not tested for mycoplasma contamination
No commonly misidentified cell lines were used in this study.