A Review on Deep Learning Approaches for Spectral Imaging

Benedikt Fischer

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
benedikt.fischer@kit.edu

Abstract

Deep learning algorithms have revolutionized the computer vision field in the last decade. They can reduce tedious feature engineering and have opened new possibilities of automated visual inspection. With deep learning techniques, the availability of large amounts of qualitative labeled data became more important than ever. The main share of computer vision research focuses on RGB images. With the advances in sensor technologies multi- and hyperspectral cameras have become more cost effective and accessible in recent years, allowing this imaging technology to be applied to new fields of application.
This article gives an overview of approaches to apply deep learning techniques to multi- or hyperspectral data. Several state-of-the-art methods will be reviewed and problems and difficulties will be discussed. An overview of a selection of available datasets is presented. To give a broad and diverse insight, research from different fields of application are considered, namely the remote sensing domain, the agricultural domain and the food industry.

1 Introduction

Deep learning models have shown state-of-the-art performance in computer vision problems with RGB images. They are successfully applied for classification, object detection, semantic segmentation, anomaly detection problems and more. In computer vision, most Convolutional Neural Networks (CNNs) can be broken down into two parts. The first part is called backbone and usually consists of a series of convolutional layers and pooling layers that compress the input image data into high-level feature maps. These layers act as feature extractors, meaning that certain neurons in these layers will be activated if certain features are present in the input image. While the first layers can extract basic features like edges, deeper layers can extract increasingly complex features like faces or the shape of a specific object [23]. The features extracted by the backbone can then be used as input for the second part of the neural network, which is often called head. This head depends on the task and can, for example, learn to classify the image or its individual pixels or it can output bounding boxes that correspond to the position of a specific object. To solve the problem of sample inefficiency of deep neural networks, transfer learning approaches can be used. In transfer learning, a backbone pretrained on a large dataset is used as feature extractor and finetuned on a different dataset.

This paper gives an overview of state-of-the-art approaches from the literature that can be used to apply deep learning models to multi- and hyperspectral data. A selection of research papers from different fields of application will be discussed. Different fields of application are considered in this review to recognize similarities and differences between these different domains.

First, the following section gives an overview on the principles of spectral imaging. An overview over available multi- and hyperspectral datasets from different fields of application is given in Section 3. Section 4 discusses common approaches from the literature that can be used to apply deep learning methods to spectral imaging data. The article concludes with a recap and an outlook on promising future research directions.

2 Principles of spectral imaging

Spectral imaging is the general term for multi- and hyperspectral imaging. Spectral imaging is a non-destructive measurement technique that can record images with a different spectral wavelength range and/or spectral resolution than RGB cameras. Spectral images often go beyond the visible wavelength range into the near infrared (NIR) or even the short-wave infrared (SWIR) regime. These wavelength regions are especially interesting as they allow obtaining information about the chemical composition of a material that cannot be observed by RGB cameras or the human eye. Spectral images can be represented as 3D tensors of shape (W, H, C), with the spatial width W and height H of the image and the number of spectral channels or bands C. This tensor is often called a hypercube. The values in this hypercube represent the light intensities detected by the sensor, which usually corresponds to the light reflected by a scene. However, a hyperspectral image can also be obtained for a transmission setting.

Spectral imaging is becoming more and more popular. This can also partially be attributed to the technological improvements of the sensors. Multi- and hyperspectral cameras have become much smaller and more cost effective in recent years, which allows new possible applications. This trend will most likely continue and further increase the popularity and accessibility of spectral imaging.

3 Datasets

This section lists a selection of multi- and hyperspectral datasets that are freely accessible and have been used by various researchers to compare their models. The existing multi- and hyperspectral datasets are much more diverse than RGB datasets. They have different spatial and spectral resolution and cover different spectral ranges. This makes it more complicated to compare different datasets with each other or to train one model with different datasets. The main benefit of spectral images over standard RGB images is the possibility to detect properties that are invisible to the human eye and thus cannot be detected by using RGB images. However, this also makes the labeling much more complicated. This, and the fact that spectral cameras are much more expensive than RGB cameras contribute to the fact that the number of existing spectral datasets is much smaller than RGB datasets and also that the spectral datasets usually have fewer labeled samples. Spectral imaging is applied in many different domains like remote sensing, agricultural and food industry, medical technology [13] and recycling industry [19]. The objectives in the recycling domain are to distinguish different materials to be able to sort them. In the medical domain, disease diagnosis and image-guided surgery are relevant use cases. For example, the detection of cancer in tissues is an important topic where spectral imaging can provide added value. The remote sensing field and the agricultural and food domain are described in more detail below.

3.1 Remote Sensing

In remote sensing, hyperspectral images of the earth surface are acquired through satellites or aircrafts. These datasets are used in agriculture, environmental monitoring, urban planning and defense. Table 3.1 shows a selection of popular multi- and hyperspectral datasets that are publicly available and have been widely used by researchers in the remote sensing field. The Indian Pines (IP) [2], University of Pavia (UP), Salinas Valley (SV) and Kennedy Space Center (KSC) datasets are hyperspectral datasets with pixelwise annotations. They consist of a single image and the number of labels in table 3.1 refers to the number of pixels within that image for which a ground truth class is given. All remaining pixels are considered as background. The IP dataset was captured by the AVIRIS sensor [6] in Indiana in 1992. The KSC dataset and the SV dataset were also captured by the AVIRIS sensor in Florida and in California respectively. The UP dataset was recorded by the ROSIS sensor [10] over the campus of the University of Pavia and has the highest spatial resolution with 1.3 m per pixel. The IP and UP datasets are available online on the GRSS Data and Algorithm Standard Evaluation (DASE) website (http://dase.grss-ieee.org). More remote sensing dataset are summarized in the works of [1] and [15].

EuroSAT [9] is a multispectral dataset, created with the freely accessible Sentinel-2 satellite images. The dataset is patch-based, it consists of 27000 small image patches that contain one predominant class and thus have one ground truth class

Table 3.1: Details of five popular remote sensing datasets. The Indian Pines (IP), University of Pavia (UP), Salinas Valley (SV) and Kennedy Space Center (KSC) are hyperspectral datasets with pixelwise classification labels and the EuroSAT dataset is a multispectral dataset with imagewise classification labels.

Dataset Name	IP	UP	SV	KSC	EuroSAT
Pixels	145 $\times$145	610 $\times$340	512 $\times$217	512 $\times$614	64 $\times$64
Bands	224	103	227	176	13
Spectral Range in nm	400-2500	430-860	400-2500	400-2500	440-2200
Spatial Resolution in m	20	1.3	3.7	18	10
Classes	16	9	16	13	10
Labels	10,249	42,776	54,129	5,202	27,000

label each. Thus, this dataset does not allow evaluating per-pixel segmentation algorithms; however, this is not its intention. Since the Sentinel-2 satellite is scanning earth's surface repeatedly approx. every five days, a classifier trained on this data could be used to continuously monitor land surfaces and detect changes in land use.

3.2 Agriculture and food industry

The most common objectives of spectral data in the agricultural domain are monitoring the state of plants, crops, fruits and vegetables. Examples are the detection of diseases and weeds or the estimation of ripeness. In the food domain, common objectives are the detection of defects like bruises and the prediction of physical parameters like acid and sugar content in fruits and vegetables. In the food domain spectrometry has a long history, thus many spectral datasets exist of point measurements without spatial dimensions. Multi- and hyperspectral datasets have become more popular in recent years, but still many dataset of the food domain are not made publicly available. It is also noteworthy that there seem to be no popular benchmark datasets that are widely used by researchers like they exists in the remote sensing domain.

Varga et al. [21] recently published a dataset that contains hyperspectral images of avocados and kiwis and covers different ripening states from unripe to overripe.

The fruits were recorded with two cameras simultaneously: the Specim FX 10 with 224 channels from 400 to 1000 nm and an INNO-SPEC Redeye 1.7 with 252 channels from 950 to 1700 nm. The images were cropped to contain one fruit and have varying spatial dimensions of around 200 to 300 pixels in each dimension. In total the dataset contains 1038 recordings of avocados and 1522 recordings of kiwis. A subset of 180 avocado images and 262 kiwi images were annotated with the reference labels: weight, weight loss during storage, storage time, firmness determined with a penetrometer and ripeness level determined by appearance and taste.

The Ladybird Brassica dataset [3] contains image data, based on weekly scans of cauliflower and broccoli vegetables over a 10-week period from transplant to harvest. This multimodal dataset consists of stereo vision data, thermal images and hyperspectral images. The hyperspectral images were recorded with the Resonon Pika XC2 camera, with 447 channels in the range 400-1000 nm. The crops were annotated with bounding boxes.

4 Deep Learning Methods

Deep learning refers to models that use neural network with many layers. Deep learning methods and more specifically deep CNNs have shown state-of-the-art performance in computer vision problems like classification, object detection, semantic segmentation and anomaly detection. The convolutional layers consist of many filters that act as spatial-spectral feature extractors. In the early layers simple features like edges can be learned by the network, whereas deeper layers can extract more complex features like specific textures or geometries. Convolutional layers are so efficient for image data due to their translation invariance in the spatial dimension. A filter that learned to extract a specific feature will extract this feature regardless of its spatial location in the image. This idea is also called weight sharing and reduces the required weights dramatically compared to a fully connected layer. The convolutional filter exploits the local correlation in the spatial dimension of the image. Convolutional networks have been show to work well with grayscale images and 3-channel RGB images, but when working with multi-channel images the size of the filters in the first layer increases drastically if normal 2D-CNN filters are used. Just like adjacent pixels, adjacent spectral bands are also correlated. A 2D-CNN filter does not make full use of this spectral correlation. One possible solution to this problem are 3D-CNN filters, however they cannot detect long-range dependencies in the spectral dimension sufficiently. A possible solution to this problem might be the use of attention-based methods.

Another problem of all these models is sample efficiency. As discussed in Section 3, the available spectral datasets have fewer samples than popular RGB datasets like ImageNet. In computer vision with RGB images, transfer learning has shown to be a powerful tool to improve sample efficiency. Models that have been trained on datasets with millions of images can be finetuned on much smaller datasets and still achieve good results. This section presents some methods that try to make use of transfer learning to apply RGB pretrained models to spectral imaging data. Another promising method to improve sample efficiency are unsupervised learning approaches.

The selection of a suitable and efficient model for spectral images poses a challenge. This section shows a selection of different approaches to these problems from the literature and discusses their results.

4.1 Preprocessing

Unlike traditional machine learning models, neural networks are usually more robust with respect to data preprocessing. Thus, most works do not use preprocessing for the spectral imaging data, with the exception of normalization. Common image normalization techniques are to normalize all values to the range $[0, 1]$ or to normalize the first- and second-order moments to obtain a zero mean and unit variance [1]. This normalization can be done for each channel independently or for all channels globally.

4.2 2D CNN

To make standard 2D-CNN architectures work with spectral imaging data, that has more than 3 channels, either the number of channels of the input hypercube has to be reduced before it is fed to the network or the first layer of the network has to be modified. To reduce the number of channels, different methods can be used: representative wavelengths can be selected with feature selection algorithms, the spectral dimension can be compressed with dimensionality reduction techniques like Principal Component Analysis (PCA) or a compression layer with learnable weights can be added before the first layer.

4.2.1 Wavelength selection

Often the most useful wavelengths for the task at hand get selected with feature selection algorithms and then only those selected wavelengths are used as input for a model. This approach solves the problem of the high dimensional data by reducing it to a few wavelengths. However, this approach usually requires tedious feature engineering and does not generalize well as these wavelengths are chosen to work optimal for one specific problem. Gao et al. [5] recorded hyperspectral images of 120 strawberries and classified them into ripe and early ripe. They use a sequential feature selection algorithm to select a feature wavelength and input this as a grayscale image into an AlexNet Model.
Pang et al. [14] recorded 300 hyperspectral images of bruised apples with 256 wavelengths from 930 to 2548 nm. They use wavelength selection to compress the data to 3 channels and then apply a YOLOv3 object detection model. To extract the effective wavelengths they applied PCA to broad regions of the spectrum and visually selected the principle component (PC)-image with the most apparent contrast between sound and bruised tissue. Then they chose 3 wavelengths where the weighing coefficient curve of the PC-image had extreme values. They also compared the result with a traditional segmentation algorithm and found the deep learning approach with YOLOv3 is more robust.

A common option to reduce the number channels in spectral images is the use of PCA. The grayscale PC-images can be concatenated to one image. However, some research has found that this approach does not perform very well when used as input for CNNs. Varga et al. [21] tested different architectures with different input data: a full hyperspectral image, a pseudo-RGB image and PC-images of the full spectrum. They found that the PC-images do not perform as well as RGB images, even when using non-pretrained CNN architectures. They conclude that PCA might remove some necessary information that is still available in the pseudo-RGB images.

Zhao and Du [25] implemented a hybrid approach. They use PCA to reduce the spatial dimension of the UP dataset and feed this to a 2D-CNN to extract deep spatial features. They also implement a balanced local discriminant embedding (BLDE) in parallel to extract spectral features from the hyperspectral data. Finally, they stack both spectral and spatial features together and use a LR classifier to classify the pixels of the UP dataset with an accuracy of 96%.

4.2.2 Added trainable layers

Steinbrener et al. [20] added two 2D-CNN layers in front of a pretrained GoogLeNet network to reduce the number of channels from 16 to 3. They use a custom dataset with 2700 multispectral images of 13 different classes of fruits and vegetables with 16 spectral bands for their finetuning. This method shows better results for their dataset than using the pretrained GoogLeNet with pseudo-RGB images, which shows the added value of additional wavelengths. However, they do not compare the results with a non-pretrained GoogLeNet, thus the benefit of transfer learning cannot be evaluated with their paper.

Zhang et al. [24] utilized a VGG16 backbone, pretrained on ImageNet, to segment bruises in blueberries from hyperspectral transmittance images. Their dataset consists of 1200 hyperspectral images with pixelwise labels for the 4 classes bruised, unbruised, calyx and background. To use the hyperspectral images with a pretrained backbone, they added a convolutional layer before the first layer to reduce the number of channels from 87 to 3. They found that using the full spectrum with 87 channels achieved better accuracy than using only 3 or 9 selected wavelengths. They also compared the results of the pretrained backbone with a backbone that was trained from scratch with randomly initialized weights. For their dataset, the model that was trained from scratch performed better than the pretrained network. The reason for this might be that the output of the added first layer has a different distribution than the original input images of the pretrained model. Since the learned filters in deeper layers depend on the output feature maps of previous layers, those learned filters might be less useful if the input distribution changes. They conclude that a possible solution for this problem could be to train only the added layer and freeze the other layers to maximize similarity between the distribution of the added layer outputs and the original input images. Another possible reason for the poorer performance of the pretrained model could be that their use case of segmenting bruised regions within blueberries might differ too much from the objective of the pretrainig, which for ImageNet is classifying images based on more than 20000 categories. The blueberry bruises in this dataset apparently do not have distinct spatial patterns, unlike most categories in ImageNet. Thus, the model might benefit less from the pretraining. It would be interesting to test if a backbone pretrained on a segmentation dataset instead of a classification dataset would achieve better results.

4.2.3 Modified first layer

Wang et al. [22] use a modified ResNet architecture where the first 2D-CNN layer has been modified to work with input images that have 151 channels instead of 3. Their custom dataset contains 557 hyperspectral transmittance images of blueberries, which are classified into good and bruised. They resize the hyperspectral cubes from $(128, 128, 1002)$ to $(32, 32, 151)$ to reduce the computational complexity and feed this reduced hypercube to their model. They achieve an accuracy of of 88% in classifying the bruised samples, which was better than the result of traditional machine learning models like linear regression or random forest. However, they highlight the limitations of traditional 2D convolutional layers when working with multi-channel images. A 2D convolutional filter uses every channel of the input data, which does not fully exploit the local correlation between channels and introduces many unnecessary weights to be trained. This may lead to overfitting and harm the generalizing ability of the model. This is especially the case for small datasets, which are common for hyperspectral data.

4.3 2D + 1D CNN

The combination of 2D and 1D CNNs is also called depthwise separable convolution and can reduce the computational complexity and the number of weights, compared to pure 2D or 3D CNNs. The depthwise separable convolution consists of a depthwise (DW) convolution followed by a pointwise (PW) convolution (or PW followed by DW) [7]. The depthwise convolution allows to model spatial relationship by applying a 2D filter-kernel to each input channel. This allows learning different spatial features for different channels. The pointwise convolution is a 1×1 convolution that can model relationships across channels. A 1×1 convolution can also be used to reduce the number of channels of an input tensor, which will result in a reduction of the number of parameters in the following convolution layer. They can also be referred to as squeeze layers [18].

Depthwise separable convolution layers have been applied by Varga et al. [21] to classify the fruit ripening dataset that was described in section 3.2. They compared their model with a ResNet and an AlexNet architecture whose first 2D-CNN layers have been adapted to the size of the hyperspectral input data and found that the separable convolution outperformed the 2D-CNNs.

Senecal et al. [18] propose a SpectrumNet architecture to classify the EuroSat dataset. They also compared the use of depthwise separable convolutions with standard 2D-CNN. While the final classification accuracy of both CNNs was similar, the standard 2D-CNN was more sample efficient in their case. The decoupling of cross-channel correlations and spatial correlation seems to make the training more difficult. However, the use of the depthwise separable convolution reduced the computational requirements of the network significantly which could be a worthwhile trade-off.

4.4 3D CNN

While a 2D convolutional filter is sliding across the two spatial dimensions and produces a 2D feature map, a 3D convolutional kernel is sliding across all three dimensions of the hypercube and produces 3D feature cubes. Like with 2D-CNNs a layer can contain several filter kernels, in which case several feature cubes are created as output of that layer. Li et al. [11] and Chen et al. [4] achieved competitive results to state-of-the-art models on the IP and UP datasets, using 3D CNNs.

To reduce the computational complexity of standard 3D-CNNs, [17] proposes a hybrid spectral CNN (HybridSN) for HSI classification that achieves state-of-the-art performance on the IP dataset. The HybridSN is a 3D-CNN followed by 2D-CNN. The 3D-CNN can extract joint spatial-spectral features from the input image and the following 2D-CNN can learn more abstract spatial features.

4.5 Attention

One disadvantage of CNNs is that they are not good at modelling long-range dependencies. However, the spectra of hyperspectral images do contain long-range dependencies as the wavelengths are correlated and may contain hundreds of channels. To solve this problem, different attention-based models have been proposed recently [8], [16].

An Attention-Based Adaptive Spectral-Spatial Kernel (A2S2K) ResNet has been proposed by [16] very recently and has achieved state-of-the-art performance on the KSC dataset with an overall accuracy of 99.43% and competitive results to the state-of-the-art on the UP and IP datasets.

Zhu et al. [26] recently proposed a spectral-spatial dependent global learning (SSDGL) model that uses an attention mechanism as well as a convolution long short-term memory module. Their model achieved state-of-the-art performance on the UP dataset and competitive results to the state-of-the-art on the IP dataset.

4.6 Unsupervised Methods

Unsupervised learning is a promising approach to the problem of limited availability of labeled data in spectral imaging. Unsupervised methods can make use of unlabeled data and the amount of available unlabeled data is much higher than the amount of labeled data. A common unsupervised method are autoencoders. Autoencoders are composed of an encoder that compresses the input data into latent feature space and a decoder that gets the latent space as input and tries to reconstruct the original input data from it. The encoder can use different convolutional layers and pooling layers to compress the data. If the autoencoder is trained with the unlabeled data, the encoder can be used as a feature extractor that will ideally be able to extract meaningful spatial-spectral features. Such an autoencoder can be used to extract features, e.g. from a much smaller labeled dataset, which can then be used as input for a supervised classifier. This approach is called semi supervised learning. Liu et al. [12] use a similar approach to classify the UP dataset and achieves competitive results to state-of-the-art methods.

4.7 Perspectives

Many different deep learning architectures for spectral imaging have been published in recent years and have set a new state-of-the-art performance in the field. Especially in the remote sensing domain, a lot of research has been published. This may also be partially contributed to the availability of several widely used benchmark datasets. Such benchmark datasets allow researchers to compare their models in a competitive way. However, the current state-of-the-art models are reaching accuracies close to 100% on some of those datasets. This may indicate the need for new and potentially more diverse or more difficult datasets. To the knowledge of the author, no such widely used benchmark datasets exist in the agricultural or food domain. In fact, many researchers just use their own private datasets. These domains could benefit from more public benchmark datasets.

Apart from the data, there is also a lot of potential for improvement on the model side. For example, the sample efficiency and robustness of such models offers room for improvements. The existing hyperspectral datasets are very different in spatial size and resolution and in the spectral wavelength range and resolution. It could be beneficial to have a universal model that works for datasets with different resolution without the need to modify its architecture, similar to RGB models that work for different spatial resolutions. Such a model could be trained with a combination of multiple different datasets, which would massively increase the available data. Unsupervised models also have a great potential, since they do not need labeled data.

5 Conclusion

Deep learning models have proven to be efficient for multi- and hyperspectral data. Many different convolutional architectures have been proposed to process spectral imaging data. The 2D, 2D + 1D and 3D CNNs combine spatial and spectral information in an intuitive and efficient way. They show state-of-the-art performance on different multi- and hyperspectral datasets. CNNs that use an attention mechanism to be able to model long-range dependencies are becoming more popular lately, also showing state-of-the-art performance on the available datasets. One of the main remaining challenges is the scarce availability of large annotated datasets. More and bigger spectral imaging datasets, similar to the ImageNet and COCO datasets used in RGB imaging, would be beneficial for the research field. However, the labeling of such data is very time consuming. Thus, a promising direction is the development of unsupervised and semi-supervised approaches.

References

[1] Nicolas Audebert, Bertrand Saux, and Sébastien Lefèvre. "Deep Learning for Classification of Hyperspectral Data: A Comparative Review". In: *IEEE Geoscience and Remote Sensing Magazine* 7 (Apr. 2019). DOI: 10.1109/MGRS.2019.2912563.

[2] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*. Sept. 2015. DOI: doi:/10.4231/R7RX991C.

[3] Asher Bender, Brett Whelan, and Salah Sukkarieh. "A High-Resolution, Multimodal Data Set for Agricultural Robotics: A Ladybird's-Eye View of Brassica". In: *Journal of Field Robotics* 37.1 (2020), pp. 73–96. ISSN: 1556-4967. DOI: 10.1002/rob.21877.

[4] Yushi Chen et al. “Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* 54.10 (Oct. 2016), pp. 6232–6251. ISSN: 1558-0644. DOI: 10.1109/TGRS.2016.2584107.

[5] Zongmei Gao et al. “Real-Time Hyperspectral Imaging for the in-Field Estimation of Strawberry Ripeness with Deep Learning”. In: *Artificial Intelligence in Agriculture* 4 (Jan. 2020), pp. 31–38. ISSN: 2589-7217. DOI: 10.1016/j.aiia.2020.04.003.

[6] Robert O Green et al. “Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)”. In: *Remote Sensing of Environment* 65.3 (Sept. 1998), pp. 227–248. ISSN: 0034-4257. DOI: 10.1016/S0034-4257(98)00064-9.

[7] Jianbo Guo et al. “Network Decoupling: From Regular to Depthwise Separable Convolutions”. In: *arXiv:1808.05517 [cs]* (Aug. 2018). arXiv: 1808.05517 [cs].

[8] Xin He, Yushi Chen, and Zhouhan Lin. “Spatial-Spectral Transformer for Hyperspectral Image Classification”. In: *Remote Sensing* 13.3 (Jan. 2021), p. 498. DOI: 10.3390/rs13030498.

[9] Patrick Helber et al. “EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.7 (July 2019), pp. 2217–2226. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2019.2918242.

[10] B. Kunkel et al. “ROSIS (Reflective Optics System Imaging Spectrometer) - A Candidate Instrument For Polar Platform Missions”. In: *1987 Symposium on the Technologies for Optoelectronics*. Ed. by C. Stuart Bowyer and John S. Seeley. Cannes, France, Apr. 1988, p. 134. DOI: 10.1117/12.943611.

[11] Ying Li, Haokui Zhang, and Qiang Shen. “SpectralSpatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network”. In: *Remote Sensing* 9.1 (Jan. 2017), p. 67. DOI: 10.3390/rs9010067.

[12] Bing Liu et al. "A Semi-Supervised Convolutional Neural Network for Hyperspectral Image Classification". In: *Remote Sensing Letters* 8.9 (Sept. 2017), pp. 839–848. ISSN: 2150-704X. DOI: 10.1080/2150704X.2017.1331053.

[13] Guolan Lu and Baowei Fei. "Medical Hyperspectral Imaging: A Review". In: *Journal of Biomedical Optics* 19.1 (Jan. 2014), p. 010901. ISSN: 1083-3668, 1560-2281. DOI: 10.1117/1.JBO.19.1.010901.

[14] Qi Pang et al. "Detection of Early Bruises on Apples Using Hyperspectral Imaging Combining with YOLOv3 Deep Learning Algorithm". In: *Journal of Food Process Engineering* n/a.n/a (2021), e13952. ISSN: 1745-4530. DOI: 10.1111/jfpe.13952.

[15] M. E. Paoletti et al. "Deep Learning Classifiers for Hyperspectral Imaging: A Review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 158 (Dec. 2019), pp. 279–317. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2019.09.006.

[16] Swalpa Kumar Roy et al. "Attention-Based Adaptive SpectralSpatial Kernel ResNet for Hyperspectral Image Classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.9 (Sept. 2021), pp. 7831–7843. ISSN: 1558-0644. DOI: 10.1109/TGRS.2020.3043267.

[17] Swalpa Kumar Roy et al. "HybridSN: Exploring 3-D2-D CNN Feature Hierarchy for Hyperspectral Image Classification". In: *IEEE Geoscience and Remote Sensing Letters* 17.2 (Feb. 2020), pp. 277–281. ISSN: 1558-0571. DOI: 10.1109/LGRS.2019.2918719.

[18] Jacob J. Senecal, John W. Sheppard, and Joseph A. Shaw. "Efficient Convolutional Neural Networks for Multi-Spectral Image Classification". In: *2019 International Joint Conference on Neural Networks (IJCNN)*. Budapest, Hungary: IEEE, July 2019, pp. 1–8. ISBN: 978-1-72811-985-4. DOI: 10.1109/IJCNN.2019.8851840.

[19] Silvia Serranti, Roberta Palmieri, and Giuseppe Bonifazi. "Hyperspectral Imaging Applied to Demolition Waste Recycling: Innovative Approach for Product Quality Control". In: *Journal of Electronic Imaging* 24.4 (July 2015), p. 043003. ISSN: 1017-9909, 1560-229X. DOI: 10.1117/1.JEI.24.4.043003.

[20] Jan Steinbrener, Konstantin Posch, and Raimund Leitner. "Hyperspectral Fruit and Vegetable Classification Using Convolutional Neural Networks". In: *Computers and electronics in agriculture* (2019). ISSN: 0168-1699.

[21] Leon Amadeus Varga, Jan Makowski, and Andreas Zell. "Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning". In: *2021 International Joint Conference on Neural Networks (IJCNN)*. July 2021, pp. 1–8. DOI: 10.1109/IJCNN52387.2021.9533728.

[22] Zhaodi Wang, Meng-Han Hu, and Guangtao Zhai. "Application of Deep Learning Architectures for Accurate and Rapid Detection of Internal Mechanical Damage of Blueberry Using Hyperspectral Transmittance Data". In: *Sensors (Basel, Switzerland)* 18 (Apr. 2018). DOI: 10.3390/s18041126.

[23] Matthew D. Zeiler and Rob Fergus. "Visualizing and Understanding Convolutional Networks". In: *Computer Vision ECCV 2014*. Ed. by David Fleet et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 818–833. ISBN: 978-3-319-10590-1. DOI: 10.1007/978-3-319-10590-1_53.

[24] Mengyun Zhang et al. "Fully Convolutional Networks for Blueberry Bruising and Calyx Segmentation Using Hyperspectral Transmittance Imaging". In: *Biosystems Engineering* 192 (Apr. 2020), pp. 159–175. ISSN: 15375110. DOI: 10.1016/j.biosystemseng.2020.01.018.

[25] Wenzhi Zhao and Shihong Du. "SpectralSpatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.8 (Aug. 2016), pp. 4544–4554. ISSN: 1558-0644. DOI: 10.1109/TGRS.2016.2543748.

[26] Qiqi Zhu et al. "A Spectral-Spatial-Dependent Global Learning Framework for Insufficient and Imbalanced Hyperspectral Image Classification". In: *IEEE Transactions on Cybernetics* (2021), pp. 1–15. ISSN: 2168-2267, 2168-2275. DOI: 10.1109/TCYB.2021.3070577.