Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning

Leon Amadeus Varga
Cognitive Systems Group
University of Tuebingen
Tübingen, Germany
leon.varga@uni-tuebingen.de

Jan Makowski
LuxFlux GmbH
Reutlingen, Germany
j.makowski@luxflux.de

Andreas Zell
Cognitive Systems Group
University of Tuebingen
Tübingen, Germany
andreas.zell@uni-tuebingen.de

Abstract—We present a system to measure the ripeness of fruit with a hyperspectral camera and a suitable deep neural network architecture. This architecture did outperform competitive baseline models on the prediction of the ripeness state of fruit. For this, we recorded a data set of ripening avocados and kiwis, which we make public. We also describe the process of data collection in a manner that the adaption for other fruit is easy. The trained network is validated empirically, and we investigate the trained features. Furthermore, a technique is introduced to visualize the ripening process.

Index Terms—hyperspectral, deep learning, convolutional neural network, ripening fruit

I. INTRODUCTION

In the fruit industry, one of the goals is to determine how ripe a fruit is. Furthermore, it is helpful for supermarkets to know the ripeness level of fruit, in order not to sell far overripe fruit or give significant discounts shortly before. For fruit like bananas, the ripeness can easily be inferred from the skin color. For others like avocados, mangos, and kiwis, this is not trivial. The fruit industry mostly uses destructive indicator measurements. So only random samples are possible here. To give a solution, we verify whether hyperspectral imaging and deep neural networks could predict the ripeness level of fruit. With this work, we contribute a hyperspectral data set and tested different models on this, thereby showing the advantage of a small neuronal network.

II. BACKGROUND AND RELATED WORK

This work covers the idea of determining the ripeness level of fruit by using hyperspectral recordings. Other works already showed that it is possible to predict the ripeness of fruit by this kind of data. Pinto et al. [1] and Olarewaju et al. [2] used hyperspectral imaging to determine the ripeness level of avocados. Zhu et al. [3] predicted the firmness and the soluble solids content of kiwis with hyperspectral recordings. In these three works, the authors use approaches without neural networks. So far, most fruit classification data analysis has been done with classical machine learning algorithms, which were often supported by small data sets. In contrast to these works, we concentrate on deep learning approaches.

The combination of hyperspectral data and deep learning was already heavily examined in the area of remote sensing. Chen et al. [4] introduced deep learning into hyperspectral remote

(a) Day 3   (b) Day 5   (c) Day 6
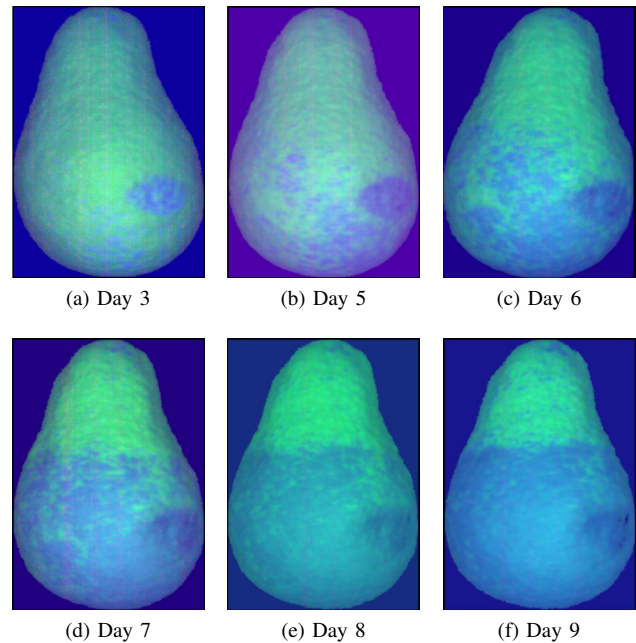
(d) Day 7   (e) Day 8   (f) Day 9

Fig. 1. Visualization of the ripening process of an avocado

sensing. In [5], a convolutional neural network outperformed SVM approaches in the classification of hyperspectral remote sensing data. Ma et al. [6] used contextual deep learning for feature mining. In [7], the HTD-Net framework was presented, which focuses on target detection with hyperspectral data. Here an autoencoder enhances the training data to produce more reliable predictions.

However, the use-cases of remote sensing differ widely from the classification task of fruit, and a direct comparison is not possible.

Mollazade et al. [8] showed the prediction capability of a simple neural network for the moisture content of tomatoes. Gao et al. [9] could predict the ripeness state of strawberries with hyperspectral imaging and a pretrained AlexNet, which is a deep convolutional neural network [10]. The ideas of both works are very similar to ours. In contrast to them, we focus on two new fruits, avocados and kiwis. For both it was already validated, that a prediction with hyperspectral data is possible

Name	Ultra-violet (UV)	Visible (VIS)	Near-infrared (NIR)
Wavelength	100-380 nm	380-740 nm	740-2500 nm

(a) Avocados   (b) Kiwis

Fig. 2. Two of the fruit crates at day 1 of the first test series.

[1] [2] [3].In contrast to the mentioned works we used a larger variety of models and recorded a large data set, which we make public. We further analyzed if hyperspectral data is necessary for this task or if pure color images are sufficient. The other works missed this validation.

A. Hyperspectral Imaging (HSI)

Hyperspectral imaging is a non-destructive measurement technique that has become increasingly popular in recent years. It is based on camera recordings that have a spectrum beyond the visible light. In contrast to standard RGB color images with three color channels, a hyperspectral image has significantly more channels [11], usually more than 100. Each channel represents the intensity for a specific wavelength. Hyperspectral imaging can be seen as a spatial spectroscopy. The wavelengths around the visible range are divided into subcategories. The relevant ranges for the most frequently used hyperspectral cameras are named in Table I. The ranges of the wavelengths reveal different chemical properties of the inspected substance. For example, the NIR range shows the presence of hydroxyl groups [12]. Hydroxyl groups are an essential part of organic chemistry. With this in mind, it is obvious why the NIR range is vital for fruit inspection. The application of hyperspectral imaging is very variable. Besides the named fruit inspection and remote sensing, for example, medical technology [13] or the recycling industry [14] uses hyperspectral measurements. All applications have in common that wavelengths outside the visible range can give valuable information.

In this work, hyperspectral imaging is used to predict the ripeness level of avocados and kiwis in a non-destructive way.

B. Fruit ripening

Here we give a short overview of the ripening process of fruit. There is a distinction between non-climacteric and climacteric fruit. Non-climacteric fruit do not ripen after harvesting [15]. Therefore, the focus in the following is on climacteric fruit. The chemical ripening process highly depends on the fruit type. The three main processes are [16]:

- Deconstruction of the cell walls, so the fruit becomes softer.
- Starch hydrolyzes to sugar, which leads to sweetness.
- Deconstruction of chlorophyll and synthesis of other pigments leads to a color change.

With the following indicators, the ripeness level of fruit are commonly measured:

- Soluble Solids Content (SSC) is based on the creation of sugar while ripening. Sugars are the majority of soluble solids in most fruit.

- Fruit flesh firmness shows the degeneration of the cell walls.
- Starch content indicates the degeneration of the starch.

Especially SSC and the fruit flesh firmness are widely used because they are reliable indicators for many fruit types. Nowadays, their measurement is destructive. So, it is only possible to measure random samples. Other works already showed that it is possible to predict the ripeness level by using hyperspectral imaging for some fruit [1] [2] [3]. In this work, the focus lies on two types of fruit. Avocados and kiwis are both fruit with a critical ripening process. The time window between unripe and overripe is small for both. Accordingly, the focus of this work lies at the end of the ripening process. Our goal is to predict the perfect consumption date.

1) Avocado: The avocado is the berry of an evergreen laurel plant. There are more than 400 different types of avocado, Hass and Fuerte being the most common. Through this broad diversity of species, the appearance of avocados may vary widely. Avocados only ripe after harvesting because the tree produces an inhibitor that prevents the fruit from ripening [17]. Besides the small consumption window, the avocado was chosen because of its relatively high price. Pinto et al. and Olarewaju et al. showed that it is possible to conclude the ripeness level with hyperspectral imaging [1] [2]. Nevertheless, the most common ripeness measurement technique for avocados is the firmness of the fruit flesh.

2) Kiwi: Like the avocado, the looping berry fruit plant kiwi has many subspecies. The best known are probably Actinidia deliciosa and Actinidia chinensis. Their appearance is very similar, only the color of the fruit flesh differs. Hyperspectral imaging for the ripeness determination of kiwis is uncommon. Useful indicators for the ripeness of kiwis are the SSC and the firmness of the fruit flesh [18].

III. DATA SET

We now describe the measurement setup, so it is possible to reproduce the data. It is also possible to adapt the procedure for other fruit. Our hyperspectral recordings are available under https://github.com/cogsys-tuebingen/deephs_fruit. This
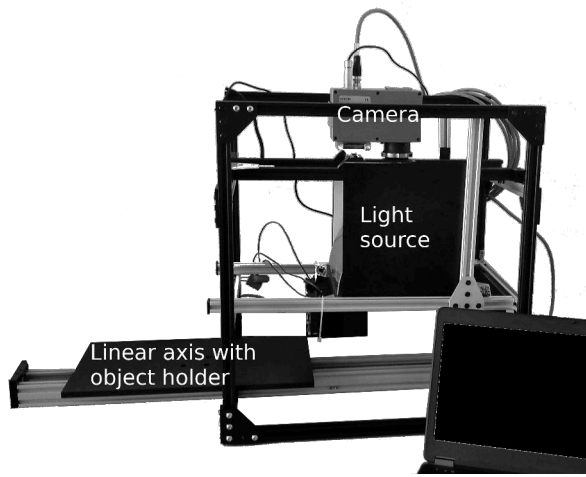
Fig. 3. The recording system. With the object holder and linear axis, the light source and the camera.

data set is used in the further analysis. The data set contains 1038 recordings of avocados and 1522 recordings of kiwis. It covers the ripening process from unripe to overripe for both fruit. Because of the destructive manner of the labeling process, only 180 avocado recordings and 262 kiwis recordings are labeled by indicator measurements. The data set was recorded in two separate measurement series. We applied a division into training set ($\frac{3}{4}$), validation set ($\frac{1}{8}$) and test set ($\frac{1}{8}$), evenly distributed among the different states of ripening.

A. Measurement setup

There are three main components, which are visible in Figure 3. The first component is the object holder, which is moved by a linear actuator. The linear actuator is necessary for the line scan operation mode of the hyperspectral cameras. The linear axis would not be needed for one-shot or mosaic hyperspectral cameras instead of line scan cameras. The latter, however, still seem to have better sensitivity.

The second component is the light source. For hyperspectral imaging, a sufficiently bright and homogeneous light source is indispensable. We used halogen lamps and LED lamps in combination to cover a broad spectrum. In addition, we used a polytetrafluoroethylene curvature reflector to create diffuse light, which is preferable.

The last component is the camera. We used two different cameras to allow a better validation of the results and cover various wavelength ranges. We used a Specim FX 10 and an INNO-SPEC Redeye 1.7. Both cameras operate in the line scan mode. For the second measurement series, only the Specim FX 10 was used. The Specim FX 10 has 224 channels, and a spectral range from 400 to 1000 nm. This range holds the VIS range with the addition of the lower NIR range. The INNO-SPEC Redeye 1.7 records 252 channels. Their spectral range run from 950 to 1700 nm.

Aside from that, we used a refractometer to measure the soluble solids content. A refractometer can indicate the concentration of a certain substance in the sample. For the fruit flesh

firmness, we used a penetrometer. A penetrometer can measure penetration resistance. Both techniques are destructive, so only random samples as labels were possible.

B. Data acquisition

The described setup was used for two measurement series covering a total of 28 days in the years 2019 and 2020. The design of the two measuring systems used was identical and followed the described setup. We acquired fresh avocados and kiwis for the two series from a supermarket which was supporting our measurement plans. Each day the following procedure was followed:

1) Record the temperature
2) Start the measurement setup for the warm up of the lamps
3) Calibrate the linear actuator
4) For both cameras:
   a) Adjust the focus of the camera on the surface of a reference object
   b) Record a white reference (average of 10 measurements)
   c) Record a dark reference (average of 10 measurements)
   d) Record the front and the back of each fruit to double the data without much effort
5) Select fruit for destructive indicator measurement
   a) Weigh the fruit
   b) Determine the fruit flesh firmness with a penetrometer
   c) (Only for kiwis:) Measure the sugar content via the refractometer
   d) Record the overall ripeness level of the fruit by appearance and taste

The number of destructively measured fruit was adapted to the ripening progress each day. The output of the test series is a collection of hyperspectral recordings of kiwis and avocados. Each recording contains only one fruit.

C. Data preparation

To improve the quality of the recorded data, we used background extraction. We excluded the background with a simple pixel-based neural network that we trained to differentiate between background and fruit. Further, the smallest possible rectangle around the fruit was extracted from the recordings to remove most of the background. We observed that the results are better if the intensity of the remaining background is forced to zero. Therefore, the results are the smallest possible recordings of the fruit with an empty background.

For the labels, we defined categories. Our goal was to classify whether the fruit is unripe, ripe, or overripe. Consequently, a regression problem is not necessary, and we reduced the complexity to three classes for the firmness, the sweetness, and the overall ripeness level. For the category firmness, the classes were based on the penetrometer measurements. The sweetness category, which is only useful for kiwis, was based
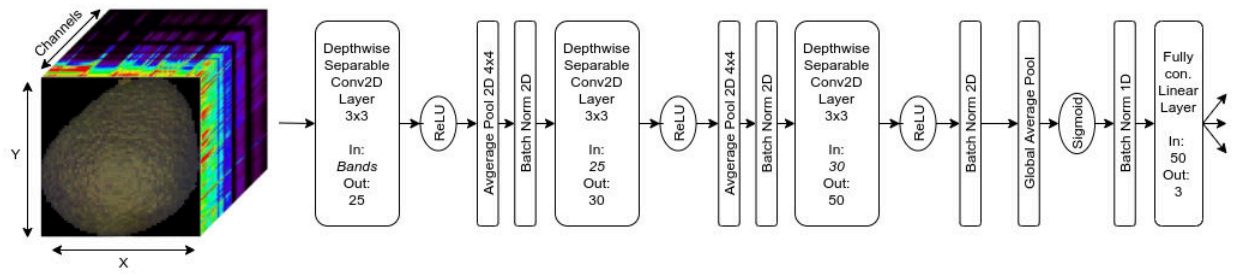
Fig. 4. Architecture of our hyperspectral convolutional neural network. The image of the input cube is a adapted version of [19].

TABLE II
CLASSES FOR AVOCADOS

	Too hard	Perfect	Too soft
Firmness	$>1200 \frac{g}{cm^2}$	Perfect	$<900 \frac{g}{cm^2}$
Ripeness	Unripe	Perfect	Overripe

TABLE III
CLASSES FOR KIWIS

	Too hard	Perfect	Too soft
Firmness	$>1500 \frac{g}{cm^2}$	Perfect	$<1000 \frac{g}{cm^2}$
Sweetness	Not sweet $<15.5\ °$Brix	Perfect	Too sweet $>17\ °$Brix
Ripeness	Unripe	Perfect	Overripe

on the refractometer tests. The last category, the ripeness, was based on the appearance and the taste. The class assignments are visible in Table II and III.

IV. EXPERIMENT

In this section we compare different models on our data set. First, we describe a simple neural network, which was designed for this application. The focus was here to reduce the chance of overfitting, so a tiny convolutional neural network was the goal. Afterwards, the training and test process is specified. Then we compare the different models. Our implementations can be found at https://github.com/cogsys-tuebingen/deephs_fruit.

A. Our Hyperspectral Convolutional Neural Network

Our Hyperspectral Convolutional Neural Network is a small neural network, which was specialized for the application of ripening fruits. We try to give reasons for some architecture decisions and explain why their usage is beneficial for hyperspectral data.

An RGB color image is a cuboid with two spatial dimensions and one channel dimension with the three channels red, green, blue. A hyperspectral image has significantly more channels than a color image, so the input data is much larger for the same spatial resolution. For computational reasons, it is essential to extract the necessary information at an early step. In many approaches this is done by a preprocessing step, where the most important bands are extracted and used for the further

inspection. We wanted to give the network the option to select the most informative bands on its own.

Besides the large data size of individual images, a further problem of hyperspectral data is often a small data set in comparison to common image classification, which can lead to overfitting. Unlike for standard color images, there are no commonly usable large data sets available to train hyperspectral models.

In Figure 4, the architecture of our HS-CNN for fruit classification is presented. The whole network is designed to keep it as simple and small as possible. The input is a hyperspectral recording of a fruit. The recording consists of two spatial dimensions and the channel dimension. Three convolutional layers extract feature maps from the input. The convolutions are separated into two smaller separable convolutions to reduce the number of parameters [20]. Instead of the frequently used max-pooling layer, we used average-pooling layers because they gave empirically better results in our experiments. An explanation might be that for this task, the winner-takes-all strategy of max-pooling layers is counter-productive. Furthermore, batch normalization was used to speed up the training process [21]. The final classification happens in the head of the CNN, consisting of a global average pooling layer and a fully connected layer. The global average pooling layer reduces the number of parameters massively and leads to more stable predictions compared to a fully connected head of similar size [22]. In the present case, the network classified three different categories visible in the output of the final layer. We developed this architecture for hyperspectral recordings with around 200 channels of wavelengths. If the number of channels differs, adaptions of the hidden layers are necessary.

B. Training

For training, the size of the classes in the categories was balanced. Thus, there was no bias towards one class. We used rotation, flipping, random noise and random cut as data augmentation techniques, as each of these doesn't change the label. The neural networks were optimized with Adabound using 1×10^{-2} as learning rate [23]. Focal loss was used as loss function [24]. We used early stopping based on the validation loss to prevent over-fitting [25]. For training, we used a batch size of 32. The hyperspectral images were resized to 64x64 pixels.

Fruit		Avocado				Kiwi					
Category		Firmness		Ripeness		Firmness		Sweetness		Ripeness	
Camera		INNO-SPEC Redeye	Specim FX 10	INNO-SPEC Redeye	Specim FX 10	INNO-SPEC Redeye	Specim FX 10	INNO-SPEC Redeye	Specim FX 10	INNO-SPEC Redeye	Specim FX 10
SVM		77.8%	73.3%	44.4%	66.7%	44.4%	60.9%	44.4%	**82.6%**	33.3%	45.8%
kNN		73.3%	77.8%	**88.9%**	60.0%	**55.6%**	60.9%	22.2%	73.9%	55.6%	50.0%
ResNet-18	RGB	66.7%	66.7%	66.7%	53.3%	44.4%	56.5%	55.6%	47.8%	44.4%	54.2%
	PCA	44.4%	53.3%	44.4%	60.0%	33.3%	60.9%	44.4%	47.8%	66.7%	33.3%
11M parameters	Full	66.7%	80.0%	33.3%	80.0%	**55.6%**	60.9%	**66.7%**	47.8%	66.7%	58.3%
AlexNet	RGB	44.4%	33.3%	33.3%	33.3%	33.3%	52.2%	44.4%	47.8%	33.3%	33.3%
	PCA	44.4%	33.3%	33.3%	33.3%	33.3%	52.2%	44.4%	47.8%	33.3%	33.3%
58M parameters	Full	44.4%	33.3%	33.3%	60.0%	33.3%	52.2%	44.4%	47.8%	66.7%	33.3%
HS-CNN (our)	RGB	77.8%	53.3%	55.6%	40.0%	44.4%	65.2%	55.6%	60.9%	44.4%	62.5%
	PCA	44.4%	80.0%	44.4%	66.7%	44.4%	34.78%	44.4%	47.8%	33.3%	33.3%
32K parameters	Full	**88.9%**	**93.3%**	**88.9%**	**93.3%**	44.4%	**69.57%**	**66.7%**	**82.6%**	**77.8%**	**66.7%**

C. Test

We tested five models on our data set. The models were a Support Vector Machine (SVM) with a radial basis kernel [26], a k-nearest neighbor classifier (kNN) [27] and a ResNet-18, a convolutional neural network architecture with identity shortcut connections and 18 layers [28]. Further an AlexNet, which performed well for strawberries [9], and our Hyperspectral Convolutional Neural Network (HS-CNN) were used. The ResNet-18 was used, because a larger represenantive of the ResNet-family would more likely tend to over-fitting. For the ResNet-18 and the AlexNet, the first layer of the network was adapted to the hyperspectral images as input. The parameter C of the SVM was evaluated by grid search with cross validation on the training set. The same applies to the parameter k of the kNN.

The test-set was $\frac{1}{8}$ of the labeled hyperspectral recordings. For the evaluation test time augmentation [29] was used. The test results are given in Table IV. For each neural network three values are given. The *Full* value gives the accuracy when the network has access to the whole hyperspectral recording. In the *RGB* case, the hyperspectral recordings were reduced to color images in a preprocessing step. And for the *PCA* case, a Principal Component Analysis (PCA) was used to reduce the channel size of the hyperspectral recordings to five. The PCA technique is often used for hyperspectral recordings to extract only the necessary information in an early step.

Our model outperformed the reference models in most cases. Moreover, it produced the most stable results. With our model, it was possible to predict the firmness of avocados with an accuracy of over 93.33 % and further predict the ripeness level in 3 categories with over 90 %. The prediction of the ripeness level of the kiwis is much harder than for the avocados. Thus, the prediction accuracy for them was significantly lower for all models. However, our model could still predict the firmness of untested kiwis with an accuracy of nearly 70% and the ripeness with nearly 80%. Further the *Full* use case was in most cases better than the reduced use cases (*RGB* or *PCA*). In the *Full* case the network could select the most influential

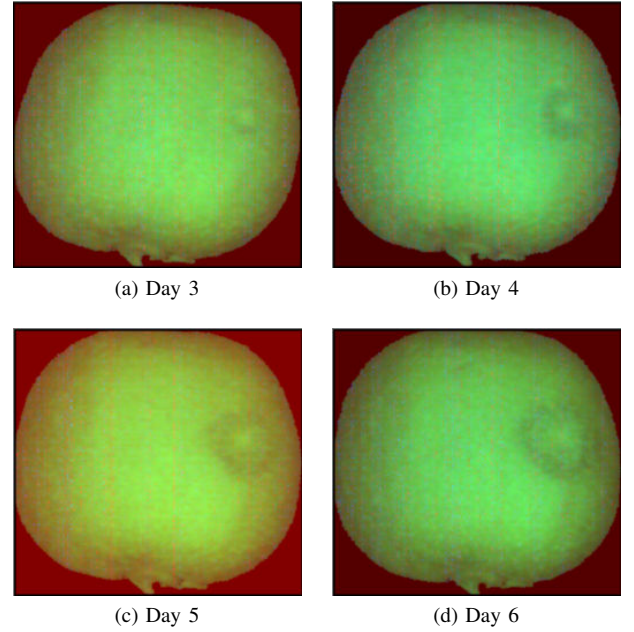

(a) Day 3   (b) Day 4

(c) Day 5   (d) Day 6

Fig. 5. Visualization of the firmness distribution of a kiwi

bands. *RGB* was in some cases better than the *PCA* approach. The *RGB* reduction doesn't use the largest variance in contrast to *PCA*. Instead, it uses the CIE color matching functions to calculate the impact of each wavelength. Most likely *PCA* removes some necessary information by the reduction, which are still available in the *RGB* reduction.

V. ABLATION STUDY

By removing or replacing components of our Hyperspectral Convolutional Neural Network, we study the impact of the different parts. In the following, the test accuracy for the prediction of the avocado firmness is given.

A. Augmentation

The influence of the different augmentation techniques is visible in this table. Random cut and test time augmentation
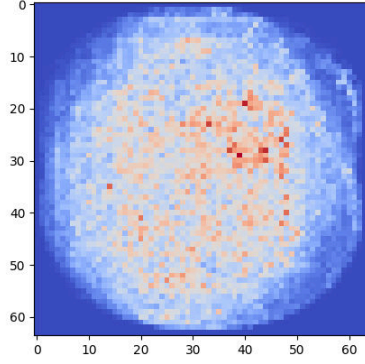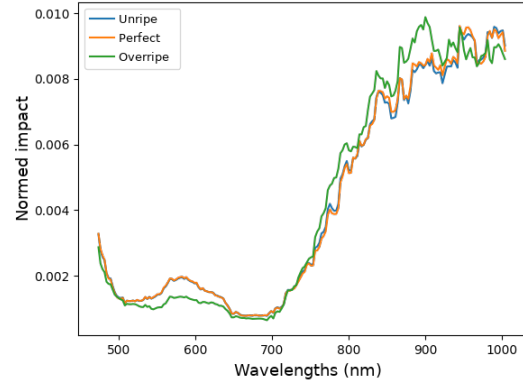
(a) Spatial based   (b) Wavelength based

Fig. 6. The impact of the input on the decision of the class for an avocado recorded with the Specim FX 10.

seems essential in this scenario. On the other hand, the effect of the transformation augmentations is smaller, so fruit alignment seems to be less of an issue in this data set.

Augmentation variant	Accuracy
Full augmentation	*93.3 %*
Without test time augmentation	70.8 %
Without random noise	73.3 %
Without random cut	69.3 %
No transformation augmentation	80.0 %

B. Depth-wise separable convolution (DSCNV)

The idea behind depth-wise separable convolution [20] is to split up the normal convolution into the spatial and a depth-wise convolution, which corresponds to the channel dimension. With this technique, the number of parameters is reduced, which can prevent overfitting.

Convolution type	Normal convolution	*DSCNV*
Accuracy	80.0%	*93.3 %*

C. Head

The head of the network uses the feature map of the convolutional part to determine the classification result. We inspected three head architectures. A fully connected head, a Global Average Pooling [22] head and a head based on Global Average Pooling with an additional linear layer. The Global Average Pooling reduces the number of parameters, which prevents overfitting. Still an additional linear layer is useful in this case.

Head architecture	Accuracy
Global Average Pooling with additional layer [22]	*93.3 %*
Global Average Pooling [22]	80.0 %
Fully connected layers	86.7 %

D. Loss function

The Focal loss is a Cross entropy loss which weights the impact of a sample corresponding to their classification error. This improves the behavior with unbalanced classes [24]. Although we were careful to avoid class bias, the Focal loss still improves the result.

Loss function	Cross entropy loss	*Focal loss*
Accuracy	80.0%	*93.3 %*

E. Optimizer

We tested different optimizers for the training process. In our case, the Adabound optimizer with a learning rate of 0.01 worked best.

Optimizer	Accuracy
Stochastic gradient descent [30]	80.0 %
Adam [31]	80.0 %
Adabound with default parameters [23]	80.0 %
Adabound with learning rate 0.01 [23]	*93.3 %*

F. Pooling layers

We compared max pooling layers with average pooling layers. The results with average pooling layers were minimally better. For this problem, it seems to be more important not to consider only the extreme value.

Pooling	Max pooling	*Average pooling*
Accuracy	86.7%	*93.3%*

VI. INVESTIGATION OF THE LEARNED CNN FEATURES

Besides the ablation study we want to show that the trained HS-CNN network learns meaningful features for the classification, which validates the correctness of the prediction. We used Integrated gradient [32] to see what parts of the hyperspectral recording are important to determine the state of the fruit. This technique can show the influence of neurons on the decision of the network. It is possible to validate the decision process

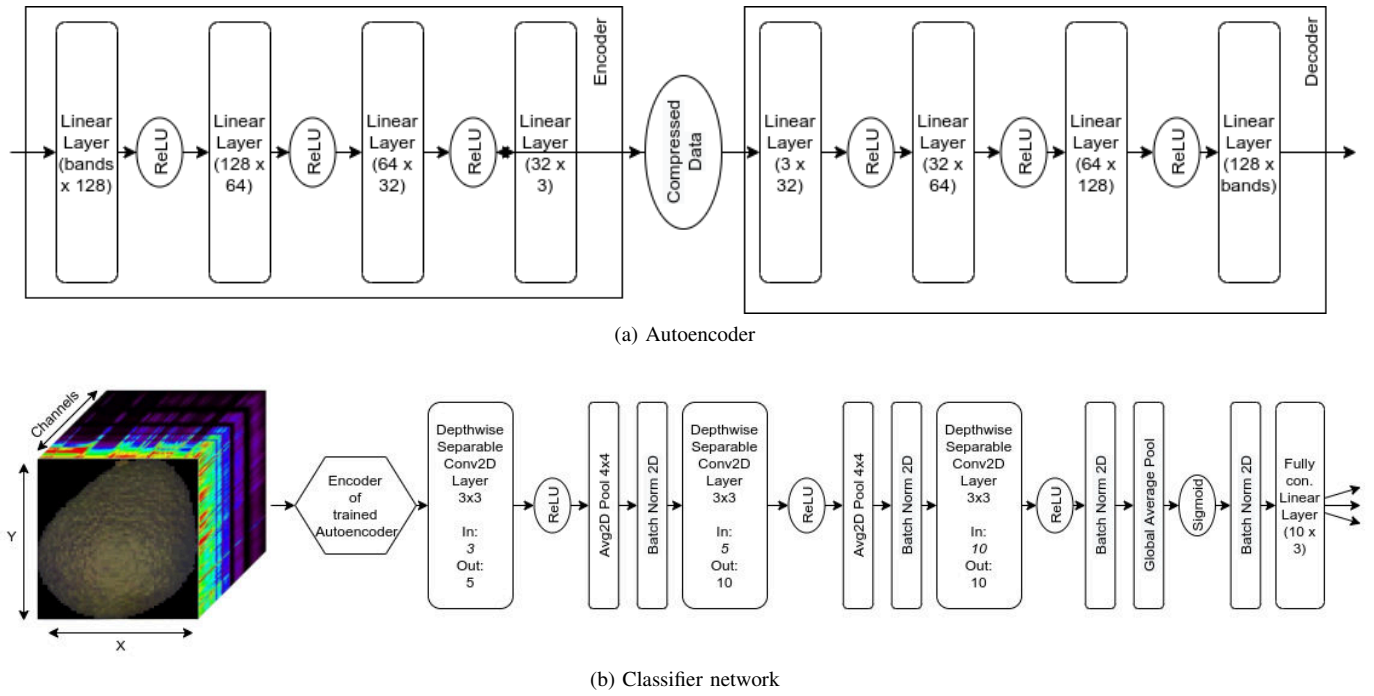(a) Autoencoder

(b) Classifier network

Fig. 7. The architecture of the Pretrained approach. The image of the input cube is a adapted version of [19].

of the neural network to a certain extent.

In Figure 6a the spatial distribution of the impact for the avocado ripeness prediction is presented. The impact is evenly distributed over the whole fruit. In Figure 6b the wavelength-based impact is visualized. The main decision happens over 800 nm. This discovery fits with the findings of Pinto et al. [1]. Additionally, to a small extent, the range of the visible light between 520 nm and 650 nm was used by the network to differentiate between unripe and perfect fruit. This range matches the visible change of the avocados. Overall the features learned by the convolutional neural network seem plausible.

VII. Visualization of the ripening process

Furthermore, we introduce a technique to generate false-color images of hyperspectral recordings for specific tasks. For this, we used a two-stage training process and a two-level classifier, presented in Figure 7. In the first step, we trained a pixel-based autoencoder (Figure 7a) to encode and decode hyperspectral images of fruit. The unlabeled data can also be used here. We used the mean-squared error for training. The latent space had a size of three, so the interpretation as a color image is possible. In the second step, we used the encoder's embedding as the input for a classifier network (Figure 7b) and trained the classifier to differ between ripeness levels. Here a Focal loss was used [24]. For the second step the labeled data is necessary. The weights of the encoder were not fixed in the second step. So, the embedding representation was adapted to fit better to the classification task. As a result, we got an encoder specialized in encoding information to differentiate ripeness levels.

An encoder we have trained in this way can produce false-color images that visualize the ripening process.

For avocados, an example is visible in Figure 1. The ripe parts are growing from the bottom to the top of the fruit. Another example is visible in Figure 5. Here the encoder was specialized for firmness prediction. The output visualizes the firmness distribution of a kiwi. A damaged part slowly grows over the fruit.

A big advantage of this technique is, that it can benefit from the large amount of unlabeled data.

VIII. Conclusion

In this work, we showed that convolutional neural networks may be used on hyperspectral data to classify exotic fruit into three classes (unripe, ripe, and overripe). We published a data set of ripening avocados and kiwis. Our HS-CNN classifier network shows superb performance in the classification of ripeness states for avocados and good performance for kiwis. We could validate the results by a more in-depth look into the trained features. Moreover, we described how to record further data. Besides that, we presented a technique to produce false-color images for specific use-cases with a pretrained autoencoder.

Semi-supervised approaches are particularly promising for further research, as they can also use unlabeled data sets.

Acknowledgment

REFERENCES

[1] J. Pinto, H. Rueda-Chacón, and H. Arguello, "Classification of Hass avocado (persea americana mill) in terms of its ripening via hyperspectral images," *TecnoLógicas*, vol. 22, no. 45, pp. 109–128, 5 2019.

[2] O. O. Olarewaju, I. Bertling, and L. S. Magwaza, "Non-destructive evaluation of avocado fruit maturity using near infrared spectroscopy and PLS regression models," *Scientia Horticulturae*, vol. 199, pp. 229–236, 2 2016.

[3] H. Zhu, B. Chu, Y. Fan, X. Tao, W. Yin, and Y. He, "Hyperspectral Imaging for Predicting the Internal Quality of Kiwifruits Based on Variable Selection Algorithms and Chemometric Models," *Scientific Reports*, vol. 7, no. 1, pp. 1–13, 12 2017.

[4] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[5] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2015-Novem. Institute of Electrical and Electronics Engineers Inc., 11 2015, pp. 4959–4962.

[6] X. Ma, J. Geng, and H. Wang, "Hyperspectral image classification via contextual deep learning," *Eurasip Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–12, 12 2015.

[7] G. Zhang, S. Zhao, W. Li, Q. Du, Q. Ran, and R. Tao, "HTD-Net: A deep convolutional neural network for target detection in hyperspectral imagery," *Remote Sensing*, vol. 12, no. 9, 5 2020.

[8] K. Mollazade, M. Omid, F. A. Tab, S. Mohtasebi, and M. Sasse-Zude, "Spatial mapping of moisture content in tomato fruits using hyperspectral imaging and artificial neural networks," in *International workshop on Computer Image Analysis in Agriculture*, 2012.

[9] Z. Gao, Y. Shao, G. Xuan, Y. Wang, Y. Liu, and X. Han, "Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning," *Artificial Intelligence in Agriculture*, vol. 4, pp. 31–38, 1 2020.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. [Online]. Available: http://code.google.com/p/cuda-convnet/

[11] D. W. Sun, *Hyperspectral Imaging for Food Quality Analysis and Control*. Elsevier Science, 2010. [Online]. Available: https://books.google.de/books?id=FVTbineZq54C

[12] K. Mitsui, T. Inagaki, and S. Tsuchikawa, "Monitoring of hydroxyl groups in wood during heat treatment using NIR spectroscopy," *Biomacromolecules*, vol. 9, no. 1, pp. 286–288, 1 2008.

[13] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, 1 2014.

[14] S. Serranti, R. Palmieri, and G. Bonifazi, "Hyperspectral imaging applied to demolition waste recycling: innovative approach for product quality control," *Journal of Electronic Imaging*, vol. 24, no. 4, p. 043003, 7 2015.

[15] L. Alexander and D. Grierson, "Ethylene biosynthesis and action in tomato: A model for climacteric fruit ripening," *Journal of Experimental Botany*, vol. 53, no. 377, pp. 2039–2055, 2002. [Online]. Available: https://academic.oup.com/jxb/article-abstract/53/377/2039/497226

[16] P. M. Toivonen and D. A. Brummell, "Biochemical bases of appearance and texture changes in fresh-cut fruit and vegetables," pp. 1–14, 4 2008.

[17] C. E. Lewis, "The maturity of avocados—a general review," *Journal of the Science of Food and Agriculture*, vol. 29, no. 10, pp. 857–866, 10 1978. [Online]. Available: http://doi.wiley.com/10.1002/jsfa.2740291007

[18] P. Martinsen and P. Schaare, "Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy," *Postharvest Biology and Technology*, vol. 14, no. 3, pp. 271–281, 11 1998.

[19] Arbeck, "Mono, Multi and Hyperspectral Cube and corresponding Spectral Signatures," 2013. [Online]. Available: https://bit.ly/38lJTY0

[20] J. Guo, Y. Li, W. Lin, Y. Chen, and J. Li, "Network decoupling: From regular to depthwise separable convolutions," in *arXiv*. BMVA Press, 8 2018. [Online]. Available: http://arxiv.org/abs/1808.05517

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, vol. 1. International Machine Learning Society (IMLS), 2 2015, pp. 448–456.

[22] M. Lin, Q. Chen, and S. Yan, "Network in network," in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 12 2014, p. 10. [Online]. Available: http://arxiv.org/abs/1312.4400

[23] L. Luo, Y. Xiong, Y. Liu, and X. Sun, "Adaptive gradient methods with dynamic bound of learning rate," in *arXiv*. International Conference on Learning Representations, ICLR, 2 2019. [Online]. Available: http://arxiv.org/abs/1902.09843

[24] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," pp. 2980–2988, 2017.

[25] L. Prechelt, "Automatic early stopping using cross validation: Quantifying the criteria," *Neural Networks*, vol. 11, no. 4, pp. 761–767, 6 1998.

[26] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 3 2000.

[27] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, p. 238, 12 1989.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem. IEEE Computer Society, 12 2016, pp. 770–778.

[29] A. G. Howard, "Some improvements on deep convolutional neural network based image classification," in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014. [Online]. Available: http://code.google.com/p/cuda-convnet

[30] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462–466, 9 1952. [Online]. Available: https://projecteuclid.org/euclid.aoms/1177729392

[31] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[32] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *34th International Conference on Machine Learning, ICML 2017*, vol. 7. International Machine Learning Society (IMLS), 3 2017, pp. 5109–5118. [Online]. Available: http://arxiv.org/abs/1703.01365