Predicting Protein Content in Grain Using Hyperspectral Deep Learning

Ole-Christian Galbo Engstrøm[1,2], Erik Schou Dreier[1,2,3], and Kim Steenstrup Pedersen[1,4]
[1]: Department of Computer Science (DIKU), University of Copenhagen, Denmark
[2]: FOSS Analytical A/S, Denmark
[3]: Department of Food Science (UCPH FOOD), University of Copenhagen, Denmark
[4]: Natural History Museum of Denmark (NHMD), University of Copenhagen, Denmark
ocge@foss.dk, erik.dreier@food.ku.dk, kimstp@di.ku.dk

Abstract

We assess the possibility of performing regression analysis on hyperspectral images utilizing the entire spatio-spectral data cube in convolutional neural networks using protein regression analysis of bulk wheat grain kernels as a test case. By introducing novel modifications of the well-known convolutional neural network, ResNet-18, we are able to significantly increase its performance on hyperspectral images. Our modifications consist of firstly applying a 3D convolution layer enabling learning of spectral derivatives that 2D spatial convolution is unable to learn, and secondly, the application of a (1×1) *2D convolution layer that downsamples the spectral dimension. Analysis of the responses learned by the convolution kernels in our modifications reveals meaningful representations of the input data cube that reduce noise and enable the subsequent ResNet-18 to perform more accurate regression analysis.*

1. Introduction

Recently, near-infrared (NIR) hyperspectral imaging (HSI), combining spectroscopy with image analysis, has been applied for grain quality assessment. Caporaso *et al*. [8] highlights that NIR-HSI combines the advantages of spectroscopy - the rapidity and non-destructive nature - with the possibility of analyzing the spatial distribution of grain quality parameters. Feng *et al*. [13] mention that deep learning has great potential of using HSI to assess seed quality. Several parameters for grain quality exist, such as Hagberg Falling Number (HFN) [14], grain sprouting, kernel hardness, moisture content, mycotoxin levels, and protein content.

Various studies have shown how NIR-HSI can be applied to solve different grain quality assessment problems. Serranti *et al*. [24] use NIR-HSI to distinguish between oat and groat kernels and achieve an accuracy of almost 100%. Yu *et al*. [28] combine NIR-HSI and deep learning employing stacked autoencoders (SAE) and a fully connected network (FCN) to predict the Nitrogen content of oilseed rape leaves. Nitrogen compounds are related to protein content [17, 27]. They achieve an RMSEP of 0.31% on a dataset with ground truth values ranging from $3\% - 6\%$ and conclude that deep learning is feasible for HSI. Caporaso *et al*. [6] use NIR-HSI for HFN regression analysis on batches of wheat grain kernels achieving an RMSEP of 62 s (HFN is measured in seconds) with a dataset with ground truth values ranging from 62 s $-$ 318 s. Apan *et al*. [2] apply partial least squares (PLS) regression on wheat crops and achieve an RMSEP of 0.50% on a dataset with ground truth values ranging from $9.4\% - 16.0\%$. Caporaso *et al*. [9] use PLS regression on single wheat kernels to determine the protein content and achieve an RMSEP of 0.94% on a dataset with ground truth values ranging from $6.2\% - 19.8\%$. Pedersen *et al*. [21] apply PLS regression on single wheat kernels and achieve an RMSEP of 0.47% on a dataset with ground truth values ranging from $6.8\% - 17.0\%$.

We analyze the possibility of combining deep convolutional neural networks (CNNs) with NIR-HSI for regression analysis within bulk grain sample quality assessment. The combination of NIR-HSI and CNN makes it possible to simultaneously utilize spatial and spectral information important for grain quality parameters which are correlated with both chemical and physical changes in grain kernels. In this paper, we use protein content of bulk wheat grains as a test case to examine possible methods for employing regression analysis with CNNs directly on the NIR-HSI data. Thus, the main focus of this article is to extend the current knowledge on deep learning for regression analysis on hyperspectral images of bulk grain kernels. Previous studies have performed regression analysis using CNNs on one-dimensional spectral data [18]. To the best of our knowledge, we are, however, the first to apply end-to-end trained CNNs for regression analysis on hyperspectral images utilizing the complete spatio-spectral data cube.

We present novel architectural modifications to the well-known CNN, ResNet-18 [15], that allows the model to better utilize the spatio-spectral features of hyperspectral images than its original design which was made for RGB images. The modifications are primarily inspired by those from Dreier *et al*. [12]. We use PLS on the spectral dimension as a baseline for comparison with the original ResNet-18 and our extensions. Additionally, as PLS uses purely spectral information, we also include a comparison with ResNet-18 using purely spatial information by averaging the spectral dimension, yielding a grayscale NIR image.

In this paper we first present, in Sec. 2, the dataset and preprocessing used for our experiments. In Sec. 3, we present the different deep learning models that we investigate in Sec. 4. In Sec. 5, we analyze and provide some interpretation to what the deep learning models learn from the dataset. Finally, we provide a discussion of our results in Sec. 6 and conclusions in Sec. 7.

2. Dataset

2.1. Acquisition

For our study, we use wheat grain obtained from reference samples of harvests from The FOSS European Grain Network (EGN) and The FOSS World Grain Network (WGN). The samples are made available for our study by Foss Analytical A/S [3]. The harvests are divided by year, with between 6 and 10 harvests from different parts of the world belonging to each year. A total of 63 samples from harvests divided non-uniformly over seven years have been processed.

From each harvest, between 100 and 130 grams of grain kernels were randomly sampled without replacement and all sampled kernels were imaged twice in two sequences. A sequence was constructed in the following way: From each grain sample, 4 images of sparsely packed grain were taken with the remaining grain from that sample being used for images of densely packed grain. This typically yielded between 5 and 7 images of densely packed grain per sample for each sequence. The samples were imaged in a random order for each sequence in order to eliminate any systematic bias that might be introduced in the hyperspectral images due to changes in camera or light source characteristics during use. Examples of densely and sparsely packed images can be seen in Figure 1.

The hyperspectral images were acquired with a *Specim FX17* [26] line-scan camera with a spatial sampling of 640 pixels and 224 spectral channels in the near-infrared spectrum in the 900 nm - 1700 nm range. The grains samples were presented to the camera in a 3D-printed sample tray that was placed on a conveyor belt. The sample was illuminated using 6 halogen light bulbs from which light was reflected onto the camera's line of focus by two mirrors, yielding a uniform and very bright illumination of the focus line. The setup can be seen in Figure 2.

The reference protein contents for each grain sample were measured with the FOSS Infratec™Nova [4]. It uses NIR transmission spectroscopy to measure the mean protein content of bulk grain samples which we use as the ground truth value for all hyperspectral image crops from that sample.

2.2. Preprocessing and data augmentation

To account for electronic noise, variations in illumination intensity, and detector response in the camera, the measured intensity spectrum $I_{i,j}$ for image pixel (i, j) was transformed into an absorbance spectrum

$$A_{i,j} = -\log\left(\frac{I_{i,j} - D_i}{W_i - D_i}\right), \tag{1}$$

where W_i is the mean spectrum of the measured reflectance of the white PTFE foil in the front and back of the sample tray by camera spatial detector pixel i. D_i is the spectrum of the reflectance of a dark image taken with the camera's shutter closed as measured by camera spatial detector pixel i. D_i was measured before imaging of each sample.

The hyperspectral images are cropped using a window size of 128×128 pixels with a stride of 64×64 pixels. Following this, we sanitize the dataset to contain only those image crops containing at least 10% grain pixels. This is determined by a segmentation mask computed by an Otsu threshold [20] on the mean spectrum of each uncropped hyperspectral image and yields a threshold on the mean spectrum that separates background from grain. This procedure yields $17,754$ hyperspectral image crops from 13 harvests for test dataset, and $69,706$ crops from 50 harvests used for 5-fold cross-validation with each validation split containing samples from 10 different harvests. The test split contains samples with a range of protein contents between 9.975% and 16.390%. The training/validation set contains samples with a range of protein contents between 8.660% and 17.780%. A histogram of the sanitized crops and the distribution of ground truths for each validation and test split can be seen in Figure 3.

Signoroni *et al*. [25] offer a review on deep learning for HSI and mention that the main problem for the interpretation of hyperspectral data comes from the curse of dimensionality. In this paper, this issue is addressed by reducing the number of spectral channels in the dataset. The first and last 10 spectral channels are discarded as the *Specim FX17* camera has reduced sensitivity near the limits of its spectrum. After this, the spectral channels are downsampled by averaging the closest pairs of channels, which finally reduces the spectral dimension to 102 channels. The

reduction is formalized by

$$A^{\text{reduced}}_{i,j,c} = \frac{A_{i,j,2c+10} + A_{i,j,2c+11}}{2} \text{ for } c = 0, 1, ..., 101 \ . \tag{2}$$

The results spectral channel width covers a wavelength range of 7.1 nm similar to the 8 nm average spectral resolution of the *Specim FX17* camera [26].

For the CNN models, we apply centering and scaling per spectral channel and independently for each different hyperspectral image crop by applying

$$A^{\text{final}}_{i,j,c} = \frac{A^{\text{reduced}}_{i,j,c} - \mu_c}{\sigma_c} \text{ for } c = 0, 1, ..., 101 \ , \tag{3}$$

where the channel mean and standard deviation is computed by

$$\mu_c = \frac{1}{n} \sum_{i=0}^{127} \sum_{j=0}^{127} \mathbb{1} \left(A_{i,j} = \text{grain} \right) A_{i,j,c} \quad \text{for } c = 0, 1, ..., 101 \ , \tag{4}$$

$$\sigma_c = \sqrt{\frac{\sum_{i=0}^{127} \sum_{j=0}^{127} \mathbb{1} \left(A_{i,j} = \text{grain} \right) \left(A_{i,j,c} - \mu_c \right)^2}{n}} \quad \text{for } c = 0, 1, ..., 101 \ , \tag{5}$$

where $n = \sum_{i=0}^{127} \sum_{j=0}^{127} \mathbb{1} \left(A_{i,j} = \text{grain} \right)$ and $\mathbb{1} \left(A_{i,j} = \text{grain} \right)$ is the segmentation mask for the image crop as computed by the Otsu threshold. Thus, the centering and scaling is done relative to the grain spectra regardless of the amount of background in any given image crop.

The PLS models predict the protein content of each image crop based on the average spectra of pixels containing grain. Prior to averaging the spectra of each image crop, a standard normal variate (SNV) [5] standardization of each pixel spectrum was applied, followed by a transformation of the spectrum into a derivative spectrum using a Savitzky–Golay (SG) smoothing [23]. The pre-processing is applied to mitigate the effect of light scattering artifacts, primarily from Rayleigh scattering occurring when the radiation interacts with particles that are small compared with the wavelength of the light [22]. The best prediction result was found using a 2nd derivative SG transformation with a 3rd order polynomial and kernel width of 7 reduced spectral channels.

We apply augmentation to the training set by rotating each hyperspectral image crop by $\{0, 90, 180, 270\}$ degrees and flipping them left-right. This increases the variability of the training data and thus increases the robustness of the model. The augmentation is done uniformly at random and independently for each hyperspectral image crop and each epoch.

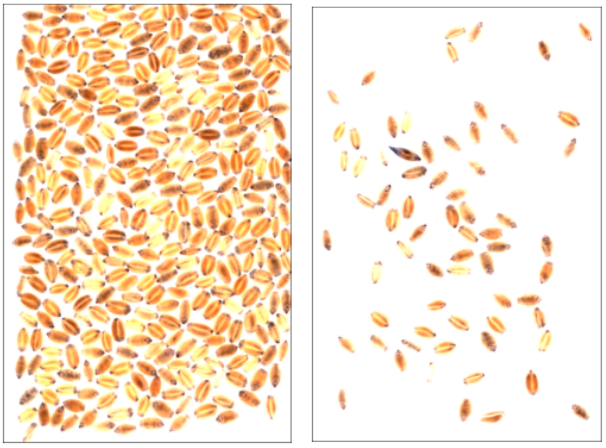


Figure 1: Examples of densely (left) and sparsely (right) packed grain images.

3. Models

All models are modifications of the ResNet-18 CNN model from the well-known family of ResNet models [15]. The modifications consist of adding new layers before the original ResNet-18; changing the ordering of rectified linear units (ReLUs) and batch normalization (BN) layers such that BN layers come after ReLU instead of before to avoid having the weights updated in a suboptimal way as argued by Chen *et al.* [10]; and changing the final layer to be a fully connected layer with a single neuron using linear activation, enabling interpretation of the output as a regression prediction. We compare the raw ResNet-18 model with models containing different 1×1 2D convolution layers, called downsamplers, that serve to downsample the spectral dimension. For some models, we also add a 3D convolution layer before the downsamplers to facilitate the learning of spectral gradients such as those computed by SG derivative filters [23] which are paramount to apply for the success of PLS [22]. In the case that we apply 3D convolution as the initial layer, the output is 4D. However, as the downsamplers expect 3D input, we concatenate the features yielded by each of the 3D convolution filters to reduce the 4D tensor to 3D. The downsamplers and 3D convolution layers have linear activation and a BN layer is applied after each for training stability. The 3D convolution layers use linear activation to mimic the linearity of the SG derivative filters. We experimented with using ReLU as activation for the downsamplers but that did not lead to increased performance.

We compare these spatio-spectral models with models utilizing purely spatial or purely spectral information. For using purely spatial information, we average the spectral dimension of the hyperspectral image crop, yielding a grayscale image crop, and feed it to the ResNet-18 skeleton. For using purely spectral information, we perform PLS

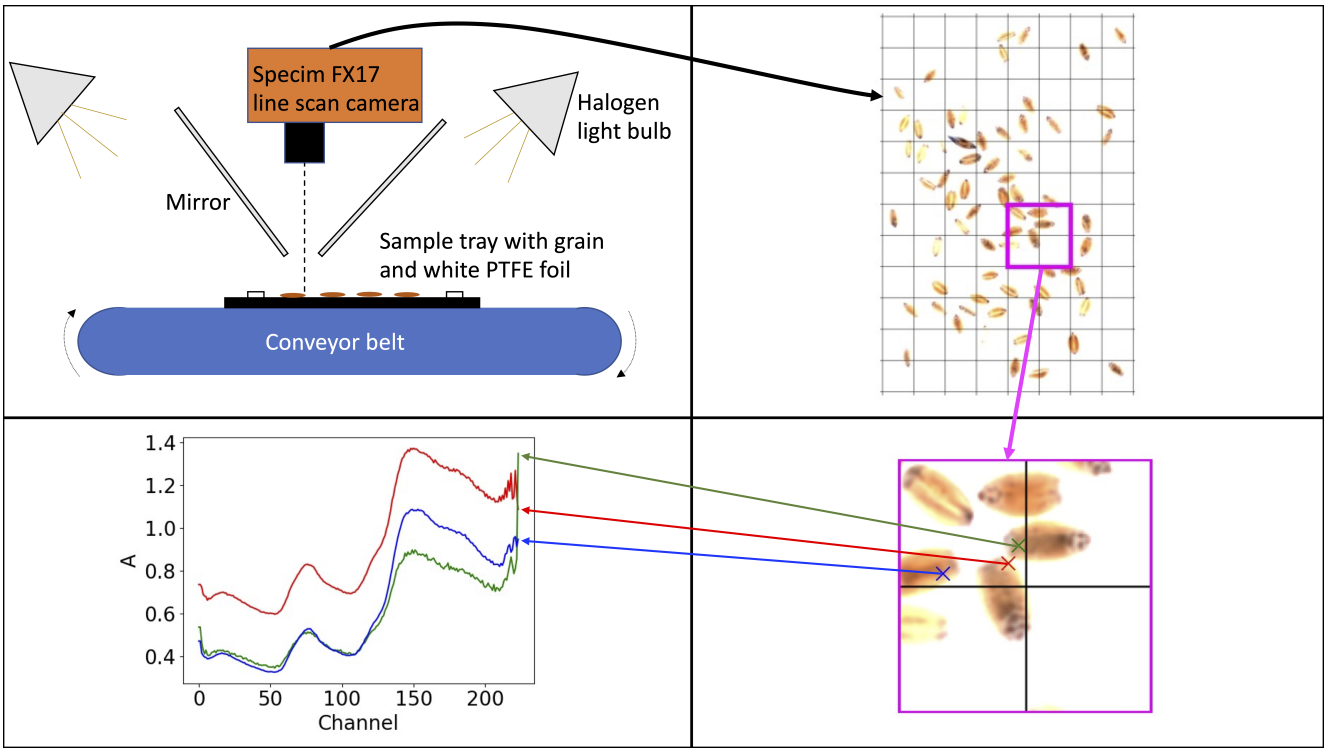


Figure 2: Top left: The imaging setup. Top right: The raw output hyperspectral image. Bottom right: A single hyperspectral image crop. Bottom left: Spectra for each of the spatial pixels marked in the hyperspectral image crop.

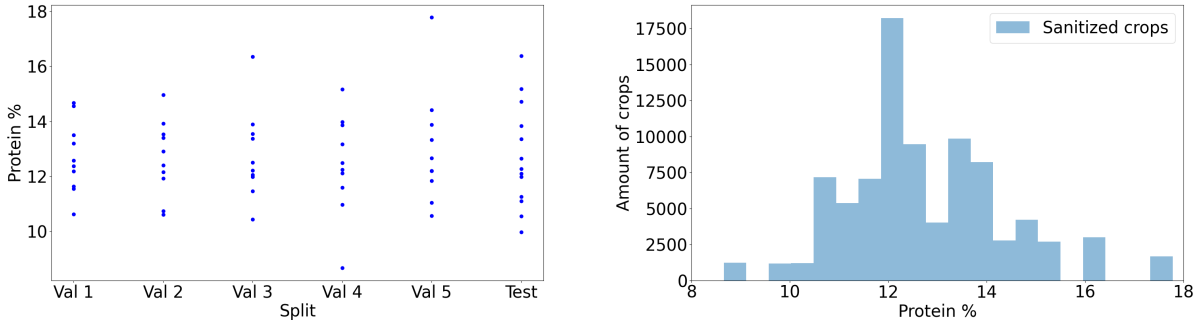


Figure 3: (Left) The protein contents of the samples in the validation and test splits. (Right) A histogram of the crops of the sanitized dataset. Most of the hyperspectral image crops are yielded from samples with a protein content close to the mean protein content.

regression on the mean grain spectrum of the hyperspectral image crop using 3 components which compares to using 3 downsamplers in the CNNs. An overview of the neural models used in this paper is shown in Table 1.

3.1. Model implementation

The CNN models are implemented using the deep learning libraries, Keras [11] and TensorFlow [1]. Similar to the original implementation of ResNet [15], we use stochastic gradient descent (SGD) with a momentum of 0.9, L2-regularization with a regularization parameter of 10^{-4}, and initialize all weights with the Kaiming He Normal Distribution [16]. The loss function is root mean squared error (RMSE). The learning rate is multiplied by a factor of 0.1 if validation loss plateaus for 4 epochs. Training halts if the validation loss plateaus for 12 epochs or if the number of epochs exceeds 50 - in practice, all training runs were halted by the plateau criterion. We train with a mini-batch

Model	Conv3D filters, kernel size	Downsampler filters
Model 1 (Grayscale input)	None	None
Model 2	None	None
Model 3	None	3
Model 4	1, $(1 \times 1 \times 7)$	3
Model 5	3, $(1 \times 1 \times 7)$	3
Model 6	1, $(7 \times 7 \times 7)$	3
Model 7	3, $(7 \times 7 \times 7)$	3

Table 1: Overview of the layers applied before the ResNet-18 skeleton. Conv3D filters are always applied before downsamplers if they are both used.

Model	Validation [%]	Test [%]	Test [%] (best on val)
Model 1	1.23 ± 0.07	1.44 ± 0.04	1.45
Model 2	1.09 ± 0.11	1.21 ± 0.08	0.97
Model 3	0.98 ± 0.09	1.05 ± 0.07	0.98
Model 4	1.91 ± 0.56	434 ± 384	2152
Model 5	$\mathbf{0.85} \pm 0.06$	$\mathbf{0.90} \pm 0.05$	**0.83**
Model 6	0.89 ± 0.06	0.99 ± 0.03	1.03
Model 7	$\mathbf{0.85} \pm 0.06$	$\mathbf{0.92} \pm 0.04$	**0.82**
PLS	$\mathbf{0.76} \pm 0.02$	$\mathbf{0.75} \pm 0.01$	**0.77**

Table 2: Mean RMSE $\pm$ standard error on the mean on the validation and test sets for each of the models. (best on val) is defined as the model which achieved the lowest RMSE on its validation set.

size of 64. All models are trained and validated on each of the five folds independently to achieve an accurate estimate of the models' true performances which includes estimates of the uncertainty of the trained models.

4. Results

We show the results obtained on the validation sets and the test set for all models in Table 2. It is clear that PLS outperforms all other models and that Model 1 (using grayscale input) performs the worst of all models that converged.

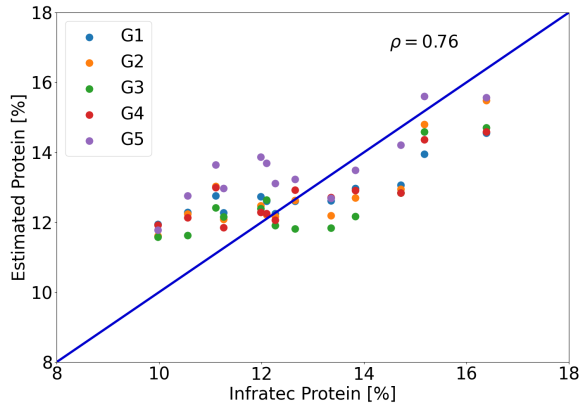
We did not succeed in making Model 4 converge even after several independent trials. Inspection showed that the single 3D convolution filter it applies has learned responses that are orders of magnitude larger than those learned by our other models and apply no meaningful transformation of the input.

Models 5-7 apply an initial 3D convolution layer and perform better than the other CNN models. Especially Model 5 and Model 7 are very close in performance and perform well in comparison with the other CNN models. In particular, if we consider only the test performance of models that performed best on their respective validation splits, Model 5 and Model 7 perform significantly better and move close to the performance of PLS. In Figure 4 we show predictions on the samples of each harvest in the test set for Models 1, 2, 5, and PLS. For visualization purposes, the predictions shown in Figure 4 are, for each Model trained on each training split, its average prediction on the hyperspectral image crops of each test harvest.

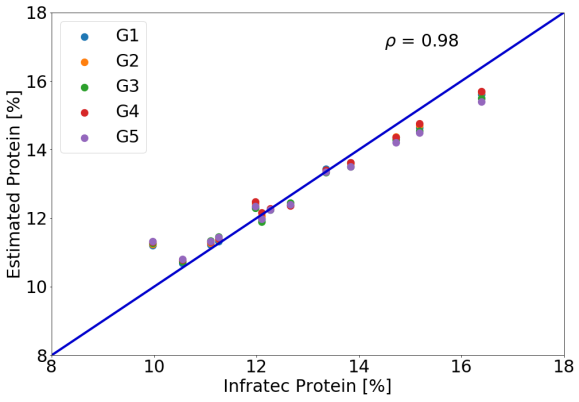
5. Analysis

It is evident in Table 2 and Figure 4 that the CNN models employing downsamplers perform better than those that do not. Furthermore, applying a 3D convolution filter before the downsamplers significantly increases the performance. Here, we analyze both of these layers. In Figure 5, we show the responses learned by the three $(1 \times 1 \times 7)$ convolution filters from Model 5 validated on split 4. The filters seem to approximate spectral gradient filters, which implies that the model is able to learn representations of the input hyperspectral image crop that is similar to the SG derivative filter applied in the preprocessing for the PLS models.

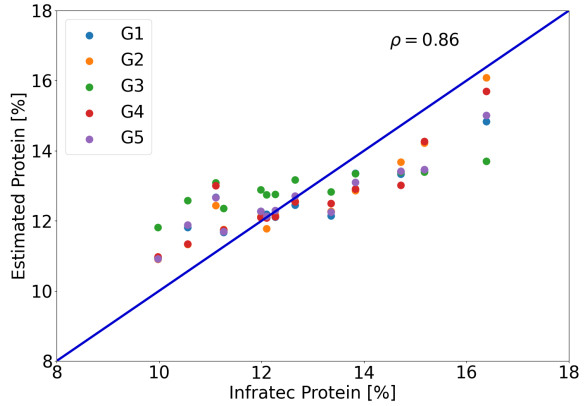
The responses learned by the three downsampling filters applied by the same model following the 3D convolution filters are shown in Figure 6. As each 3D convolution filter yields an output that consists of a learned representation of the entire input hyperspectral image, the concatenation of these outputs will have three consecutive representations of the input hyperspectral image. Each (1×1) kernel in the downsamplers is applied to a specific wavelength offset, consisting of the response learned by the $(1 \times 1 \times 7)$ 3D convolution filter applied at that wavelength. Nitrogen (N) compounds are related to protein content [17, 27]. Therefore, in Figure 7 we zoom in on the last third of the downsampling filters from Figure 6. These parts of the downsampling filters operate on the representation of the hyperspectral image as output by Filter 3 in Figure 5. Here we have plotted the expected wavelength centers of light absorption caused by NH and CH overtones in the 938 nm - 1662 nm range (this is the spectrum remaining after we cut off the first and last 10 channels of the *Specim FX17*) [19] along with the responses learned by the three downsampling filters. The downsamplers learn spikes in their responses near these expected wavelength positions of NH overtones. However, they also learn spikes in their responses at other wavelengths, noticeably, where CH overtone bands are located (between 1100 nm and 1250 nm, between 1350 nm and 1500 nm, and above 1600 nm) [19]. Similarly, absorbance values located at wavelengths related to CH overtones are also found to be of importance in the PLS model used for purely spectral prediction in this paper as well as previous attempts of predicting protein content with NIR



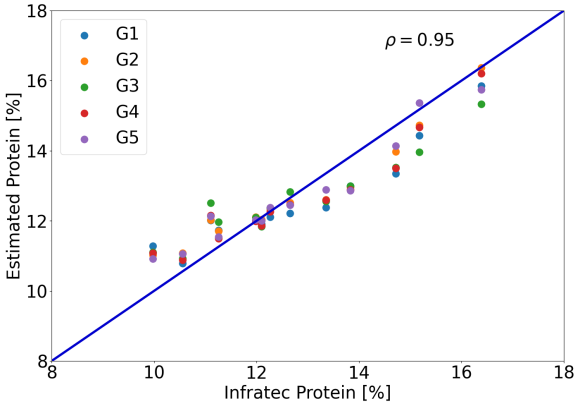
(a) Model 1 (grayscale input).



(b) PLS.



(c) Model 2 (pure ResNet).



(d) Model 5 (Conv3D 3, $(1 \times 1 \times 7)$ and 3 downsampling filters).

Figure 4: Average predictions on the test set as functions of the protein % measured by the Infratec™Nova by purely spatial based ResNet-18 (Model 1), purely spectral based PLS (PLS), spatio-spectral based ResNet-18 (Model 2), and spatio-spectral based modified ResNet-18 with downsampler and 3D convolution layer (Model 5). Gi is the model validated on validation split i. ρ is the Pearson correlation coefficient between the plotted average predictions of all Gi and the ground truths.

spectroscopy and PLS [7], indicating that CH content correlates with protein content.

We examine the representation of the hyperspectral image crops from training split 4 as output by Model 5 (trained and validated on split 4) after its downsamplers in Figure 6. We use the data from the model's training to analyze if the model is able to learn to distinguish protein content at this early stage. As the downsampling layer utilizes three (1×1) convolution filters, the output will have each spatial pixel represented by a 3D vector, that is a reduced representation of the 102 input wavelength channels. In Figure 8a we perform a PCA on this representation on a representative subset of training split 4 to reduce the representation of each spatial pixel to two dimensions. Here, we color black the spatial pixels containing background and color the spatial pixels containing grain according to their protein content. Clearly, the variance between background and foreground dominates the principal components. Therefore, we apply PCA only to the subset spatial pixels containing grain of the same data and plot this in Figure 8b. We observe that this representation shows that there is a gradient from low protein contents (blue) near the edges of the cluster through middle protein contents (red) to high protein contents (green) in the center.

6. Discussion

Considering the results shown in Table 2 and Figure 4 it is evident that the modified ResNet does not achieve a better RMSE than PLS. While the predictions of Model 1 in Figure 4a (pure ResNet-18 with grayscale input) reveal

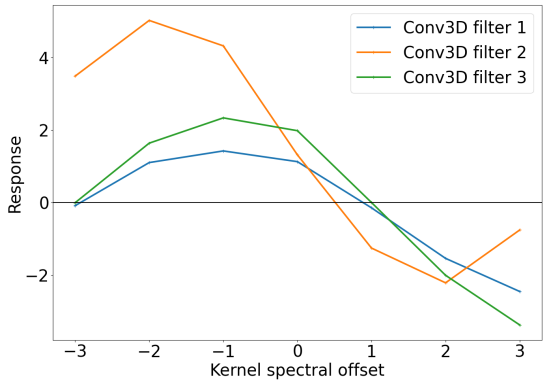


Figure 5: The three 3D convolution filters for Model 5 validated on split 4.

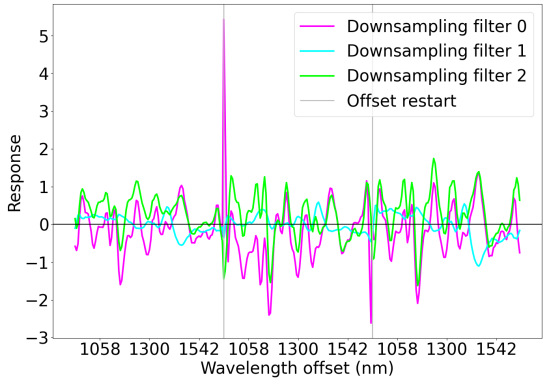


Figure 6: The three downsampling filters following the 3D convolution filter in Figure 5. The output of each 3D convolution filter consists of its own representation of the entire input hyperspectral image crop. Thus, each channel in the output is represented by the wavelength offset on which the 3D convolution filter was placed. The offset restarts indicate the downsampling filter channels at which the outputs of the previous 3D convolution filters are concatenated, restarting the wavelength at which the downsampling filter operates.

that there is some correlation between spatial feature distribution and protein content indicated by the Pearson correlation coefficient of $\rho = 0.76$, the PLS model using only spectral information vastly outperforms Model 1. Model 2 (pure ResNet-18 with spatio-spectral input) outperforms Model 1 with a large margin but is outperformed with an even larger margin by our modified ResNet-18 from Model 5, indicating that our modifications to ResNet-18 are highly favorable.

From the distribution of the average predictions shown

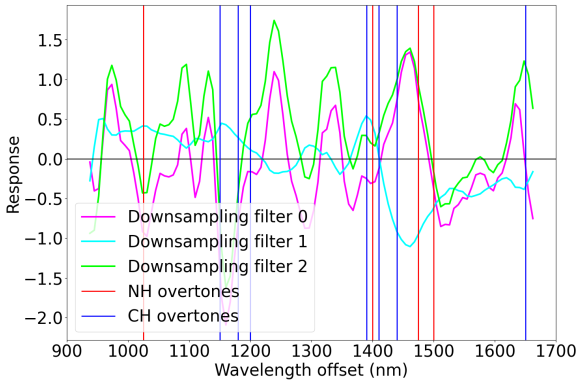


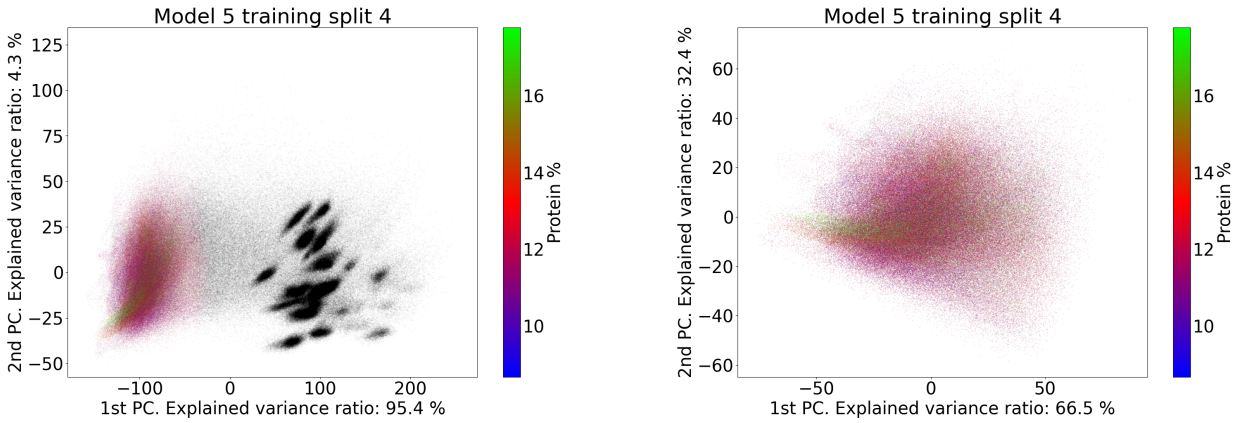
Figure 7: Zoom-in of the last third of the downsamplers shown in Figure 6 along with expected wavelength centers of light absorption caused by NH and CH overtones.

in Figure 4 it is evident that the variance in the mean predictions of the same model validated on different splits is relatively higher for the CNN models when compared to PLS. This indicates that the CNNs will benefit from a larger dataset and have not yet reached their full potential.

For PLS, we apply the segmentation mask yielded by the Otsu threshold in order to determine which pixels contain grain and which contain background. For the CNN models, we do not apply the segmentation mask. This implies that the CNNs have to learn to distinguish foreground from background. We experimented with applying the mask to nullify the spectra of any spatial pixels containing background. This did, however, not lead to increased performance, but warrants more future investigation.

Indeed, analyzing our modifications to ResNet-18 revealed that they were able to autonomously learn highly relevant responses to the problem on which they were applied including the segmentation of foreground and background but also the learning of spectral derivatives which have to be manually applied for PLS. Thus, from an end-to-end learning perspective, our modified ResNet-18 is able to extract relevant spatio-spectral features for the application and solve it to a reasonable performance level without the need for application of any advanced pre-processing or prior knowledge of the problem. Additionally, this indicates that it may not be necessary to swap the model or the simple minimal preprocessing that we apply in order to perform regression analysis on a different parameter than protein content, although it may be necessary to fine-tune the number of downsamplers used.

Although our modified ResNet-18 provides a clear advantage over the pure ResNet-18, we expect that its full potential is yet to be uncovered on other problems where there is a higher correlation between spatial features and the tar-



(a) (b)

Figure 8: PCA reductions of the output of the downsamplers shown in Figure 6 using pixels with grain and background (a) and using only pixels with grain (b). The protein content in % is measured by the Infratec™Nova.

get parameter of interest than for protein content.

7. Conclusion

The work presented in this paper shows that CNNs can be applied to hyperspectral images for regression analysis on bulk wheat grain kernels and achieve reasonable results. While seemingly not as relevant as the spectral information, we show that there is some correlation between the spatial features of wheat grain kernels and their protein content. We applied ResNet-18, which is known to perform well on RGB images, to hyperspectral images. By incorporating two concrete extensions of ResNet-18 presented in this work, we are able to significantly boost its performance to a level that is close to the performance achieved by PLS which is known to solve the problem of protein content regression analysis well. We show that our extensions of ResNet-18 are able to learn representations of the input data that resemble the preprocessing that humans with extensive knowledge of the problem would apply before applying PLS to perform regression analysis. Furthermore, we show that our extensions are able to learn representations of the hyperspectral data that are highly relevant to the task of protein regression analysis on which they are applied. This end-to-end approach indicates that using our modifications, CNNs should be applied to other regression problems of interest within NIR-HSI, especially where both spatial and spectral information is relevant.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 4

[2] Armando Apan, Rob Kelly, Stuart Phinn, Wayne Strong, David Lester, David Butler, and Andrew Robson. Predicting grain protein content in wheat using hyperspectral sensing of in-season crop canopies and partial least squares regression. *International Journal of Geoinformatics*, 2(1):93–108, 2006. 1

[3] Foss Analytical A/S. https://www.fossanalytics.com/. 2

[4] Foss Analytical A/S. Infratec™nova, https://www.fossanalytics.com/en/products/infratec-nova. 2

[5] R. J. Barnes, M. S. Dhanoa, and Susan J. Lister. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5):772–777, jul 1989. 3

[6] Nicola Caporaso, Martin B Whitworth, and Ian D Fisk. Application of calibrations to hyperspectral images of food grains: Example for wheat falling number. *Journal of Spectral Imaging*, 6, 2017. 1

[7] Nicola Caporaso, Martin B. Whitworth, and Ian D. Fisk. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chemistry*, 240:32–42, 2017. 6

[8] Nicola Caporaso, Martin B Whitworth, and Ian D Fisk. Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Applied spectroscopy reviews*, 53(8):667–687, 2018. 1

[9] Nicola Caporaso, Martin B Whitworth, and Ian D Fisk. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food chemistry*, 240:32–42, 2018. 1

[10] Guangyong Chen, Pengfei Chen, Yujun Shi, Chang-Yu Hsieh, Benben Liao, and Shengyu Zhang. Rethinking the usage of batch normalization and dropout in the training of deep neural networks. *CoRR*, abs/1905.05928, 2019. 3

[11] François Chollet et al. Keras. https://keras.io, 2015. 4

[12] Erik Schou Dreier, Klavs Martin Sørensen, Kim Steenstrup Pedersen, Toke Lund-Hansen, and Birthe P. Møller Jespersen. Study of hyperspectral imaging for bulk grain kernel classification with convolutional neural networks. *Journal of Spectroscopy*, 2021. Under review. 2

[13] Lei Feng, Susu Zhu, Fei Liu, Yong He, Yidan Bao, and Chu Zhang. Hyperspectral imaging for seed quality and safety inspection: A review. *Plant methods*, 15(1):1–25, 2019. 1

[14] Sven Hagberg. A rapid method for determining alpha-amylase activity. *Cereal Chem*, 37(2):218–222, 1960. 1

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 2, 3, 4

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 4

[17] ICR Holford, AD Doyle, and CC Leckie. Nitrogen response characteristics of wheat protein in relation to yield responses and their interactions with phosphorus. *Australian Journal of Agricultural Research*, 43(5):969–986, 1992. 1, 5

[18] Ine L Jernelv, Dag Roar Hjelme, Yuji Matsuura, and Astrid Aksnes. Convolutional neural networks for classification and regression analysis of one-dimensional spectral data. *arXiv preprint arXiv:2005.07530*, 2020. 1

[19] Metrohm NIRSystems. A guide to near-infrared spectroscopic analysis of industrial manufacturing processes. *United States: Silver Spring*, 2002. 5

[20] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transaction on Systems, Man and Cybernetics*, SMC-9(1):62–66, 1979. 2

[21] Dorthe Kjær Pedersen, Harald Martens, Jesper Pram Nielsen, and Søren Balling Engelsen. Near-infrared absorption and scattering separated by extended inverted signal correction (eisc): Analysis of near-infrared transmittance spectra of single wheat seeds. *Applied spectroscopy*, 56(9):1206–1214, 2002. 1

[22] Åsmund Rinnan, Frans van den Berg, and Søren Balling Engelsen. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10):1201–1222, nov 2009. 3

[23] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964. 3

[24] Silvia Serranti, Daniela Cesare, Federico Marini, and Giuseppe Bonifazi. Classification of oat and groat kernels using nir hyperspectral imaging. *Talanta*, 103:276–284, 2013. 1

[25] Alberto Signoroni, Mattia Savardi, Annalisa Baronio, and Sergio Benini. Deep learning meets hyperspectral image analysis: A multidisciplinary review. *Journal of Imaging*, 5(5), 2019. 2

[26] Specim. Specim fx17, https://www.specim.fi/products/specim-fx17/. 2, 3

[27] WM Strong. Nitrogen requirements of irrigated wheat on the darling downs. *Australian Journal of Experimental Agriculture*, 21(111):424–431, 1981. 1, 5

[28] Xinjie Yu, Huanda Lu, and Qiyu Liu. Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (brassica napus l.) leaf. *Chemometrics and Intelligent Laboratory Systems*, 172:188–193, 2018. 1