ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Robust Feature Selection and Robust Training to Cope with Hyperspectral Sensor Shifts

Valerie Vaquet, Johannes Brinkrolf, and Barbara Hammer *

Machine Learning Group
Bielefeld University, Bielefeld - Germany

Abstract. Hyperspectral imaging is a suitable measurement tool across domains. However, when combined with machine learning techniques, frequently intensity and transversal shifts hinder the transfer between different sensors and settings. Established approaches focus on eliminating sensor shifts in the data or recalibrating sensors. In this contribution, we target the training procedure, propose robust training, and derive a robust feature selection strategy that can cope with multiple shift dynamics at the same time. We evaluate our approaches experimentally on artificial and real-world datasets.

1 Introduction

Hyperspectral imaging constitutes a powerful tool across various domains. For example, it is used in quality control in food production and pharmaceutical applications, precision agriculture, environmental analysis and earth observation, water resource management, medical diagnosis, and artwork and forensic document analysis [1, 2]. The usage of this technology will further expand with the increasing availability of low-cost sensors. Frequently, data collection by this technology is paired with machine learning techniques to make sense of the obtained high dimensional spectral signatures [3]. While machine learning is a successful tool in many use cases, applying it to hyperspectral data collected across different sensors poses significant challenges. Transferring a model trained on the data collected by one sensor to the data measured by another one, usually a strong decline in the accuracy of the model is observed. This is caused by distributional shifts between the sensor measurements which are commonly caused by slight differences in the physical components in the sensors, or due to aging components over time [4].

Recent work analyzed these sensor shifts and proposed a categorization into so-called intensity and transversal shifts where the intensity shift is an additive offset while in transversal shifts the measured spectral bands differ across sensors or over time [4]. These shifts can decrease the performance of machine learning models and – in extreme cases – render them useless. Basically, there are three main options to handle this problem: (i) retraining the ML model, (ii) eliminating the sensor shift in the data before training and inference, and (iii) developing robust models as we propose in this contribution. Retraining the model in case of sensor shifts requires ground truth information. Obtaining this kind of information for a new sensor which should be deployed is usually a costly and labor-intensive task if even possible and thus ruling out option (i). [4] proposed

*We gratefully acknowledge funding from the European Research Council (ERC) under the ERC Synergy Grant Water-Futures (Grant agreement No. 951424).

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

a set of strategies to eliminate intensity and transversal shifts from hyperspectral data. In contrast to the retraining strategy, no ground truth information is required for this realization of option (ii). However, the same class distribution across all sensor measurements is expected to obtain robust mean spectral signatures which are the basis for the proposed techniques. Besides, intensity and transversal shifts are considered individually which limits the practical application, as both types of shifts might be present in unknown magnitudes in real-world data.

In this work, we focus on option (iii) and target the training procedure. For this purpose, we first analyze the influence of intensity and transversal shifts on model robustness. Afterward, we propose a robust training strategy and derive a feature selection strategy from an analysis of the effects of transversal shift on the data measurements. While it has been shown that robust training can improve the transferability for instance in deep learning [5], to the best of our knowledge its effect and efficient realization has not been investigated in the setting of transversal and intensity shift in hyperspectral data. Our approaches are evaluated experimentally on artificial and real-world data.

2 Intensity and Transversal shift in Hyperspectral Data

We can describe the spectra in a functional form $X : \mathbb{R} \to \mathbb{R}$, where wavelengths l_j are mapped to intensities $X(l_j)$. As the spectra can only be observed at a finite number of bands, we obtain samples $x_i \in \mathbb{R}^d$ where components of x_i correspond to observations $x_i(L) := (x_i(l_1), \ldots, x_i(l_d))$ at a set of wavelengths $L = (l_1, \ldots, l_d)$. A sensor-specific number of wavelengths are measured resulting in data of dimensionality d. Prior work [4] showed, that in case of large d selecting a subset of these features as a preprocessing is suitable to reduce the curse of dimensionality [6].

As described before, distributional shifts over time or between sensors have to be expected. There are two types: *Intensity shift* corresponds to a function $S_v : \mathbb{R} \to \mathbb{R}$ which is added to each spectrum x_i, yielding signals of the form $x_i + S_v$ with finite-dimensional observation vector $(x_i + S_v)(L) = (x_i(l_1) + S_v(l_1), \ldots, x_i(l_d) + S_v(l_d)))$. *Transversal shift* corresponds to a strictly monotonically increasing function $S_h : \mathbb{R} \to \mathbb{R}$ which affects the wavelengths, i.e. the domain X, yielding the signal $S_h \circ x_i$ with finite-dimensional observation vector $S_h \circ x_i(L) = x_i(S_h(L)) = (x_i(S_h(l_1)), \ldots, x_i(S_h(l_d)))$ [4].

3 Datasets

In this contribution, we consider a dataset containing the hyperspectral signatures of Arabica, Robusta, and immature Arabica coffee beans. The dataset was collected by three similar but different sensors from the same manufacturer. The S2 and S3 are from the same production series. The models measure 256 to 288 spectral bands in the intensity of 950 nm to 2500 nm. In order to have data in the same dimensionality, the data has been linearly interpolated. The final dataset contains 75 954 samples with 1500 features each. The sensor-wise mean spectra of the dataset are visualized in Fig. 1. One can clearly see an

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

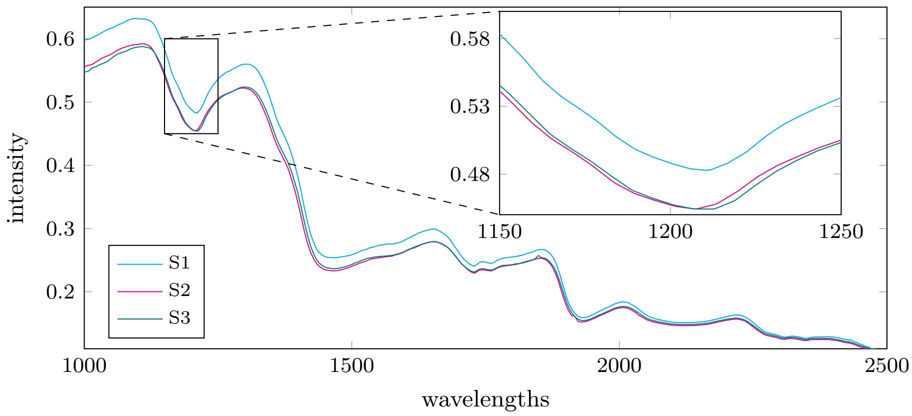


Fig. 1: Visualization of the sensor-wise mean spectra of the Coffee dataset

intensity shift between the measurements of the S1 and the other two sensors. Additionally, when zooming in, the presence of an additional transversal shift of the S2 in the data becomes apparent. For our experiments, a subsampling of the data to 50 equally spaced bands is performed as in [4].

To better analyze the influence of intensity and transversal shift, we additionally generate artificial datasets with isolated shifts of increasing magnitudes. The datasets are based on the original measurements of S2. For the intensity shift, we simulate a smooth additive offset in varying magnitudes. In contrast, for the transversal shift, we rely on reported transversal shifts and artificially increase their magnitudes. Finally, we consider combined shift by adding an intensity offset to the transversally shifted datasets.

4 Robustness under sensor shifts

Analyzing the effect of the intensity and transversal shift, we train our model on the original measurements of S2 and evaluate on the artificially shifted datasets. Based on prior experiments, we choose a logistic regression with L2 regularization[1]. The hyperparameters were tuned on the unshifted dataset. We perform 10-fold cross-validation. The results of our experiment are visualized in Fig. 2. One can clearly see that the model is robust to moderate intensity shifts. In contrast, it strongly suffers from transversal shifts.

Considering the real-world Coffee dataset, we train a model on each sensor and evaluate it separately on the data of each sensor. Summarizing the results in a matrix, we obtain a confusion matrix-like result representing the training in the rows and the sensor used for testing in the columns. The results are summarized in Fig. 3. As expected, we observe high accuracies on the diagonal where no transfer happens. Considering the transfer scenarios, we obtain a much better transfer between sensors S1 and S3 than between S2 and the other

[1]Note that other linear models performed very similarly but yielded slightly worse scores, as discussed in [4].

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

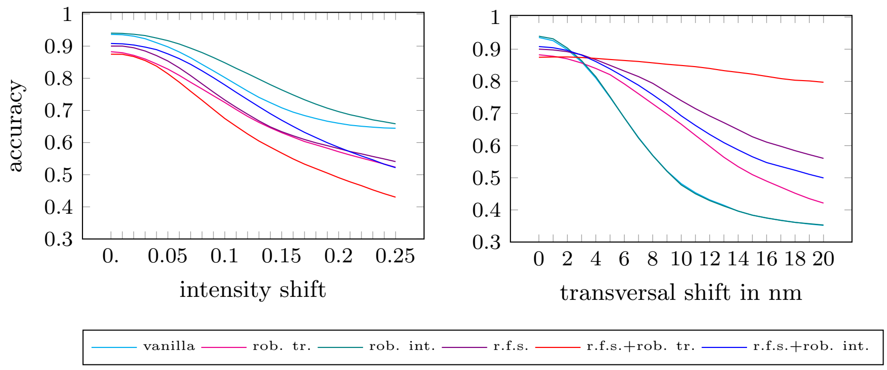
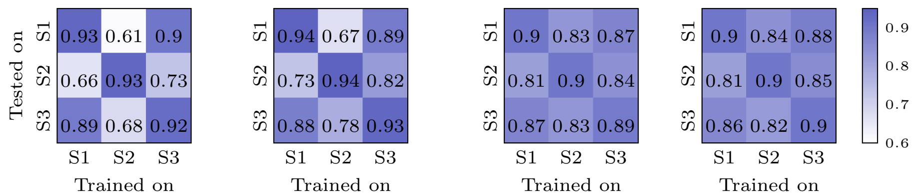


Fig. 2: Performance of classifiers for increasing shifts



(a) vanilla (b) rob. tr. (c) r.f.s. (d) r.f.s + rob. tr.

Fig. 3: Results of the experiments on the Coffee dataset

sensors. In contrast, transfer scenarios that include S2 result in worse accuracies. Reconsidering the shift analysis (see Fig. 1), this supports our previous finding that the considered model is more robust with respect to intensity than transversal shift. As we identified transversal shifts as the main challenge, we propose robust training procedures mainly targeting this type of shift in the next section.

5 Robustifying Training

Robust Training: Robust training is an established scheme in deep learning [7]. Usually, deep neural networks are trained with noisy data or adversarial examples to avoid overfitting and ensure robustness against different types of attacks. In our setting, the goal is to train a model that is robust in the presence of sensor shifts. Thus, we propose to enrich the training data with transversal shifts, whereby we rely on very simple data-independent instantiation of the shifts. We enrich the training set with transversally shifted versions of the spectra, e. g. each wavelength is shifted by the same factor: $S_h : \mathbb{R} \to \mathbb{R}, S_h(l_j) = l_j + \nu$. Thus, we obtain additional spectra $S_h \circ x_i(L) = x_i(S_h(L)) = (x_i(l_1+\nu), \ldots, x_i(l_d+\nu))$. An analysis of the Coffee dataset, further datasets measured with similar sen-

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

sors, and the recalibration reports indicate that average absolute transversal shifts of about $\nu = 3\,\text{nm}$ are expected. Therefore, for our training we choose $\nu \in \{-3, -2, -1, 1, 2, 3\}$. Similarly, there is the option to enrich the data with intensity shifts $S_v(l) = \delta$, where we choose $\delta = 0.05$ as we observed intensity shifts of this magnitude in real-world data (see Fig. 1). We refer to these strategies as rob. tr. and rob. int. in the experiments.

Robust feature selection: In addition to this robust training scheme, we propose a robust feature selection strategy that can be combined with the robust training strategies: On a data level, transversal shift results in changes in the intensity values. However, in contrast to relatively stable and smooth intensity shift, the magnitude of changes caused by transversal shift does not only depend on the size of the transversal shift but also on the spectra: as defined in Section 2, the intensity of x_i at measurement band j under transversal shifts can be formalized as follows: $x_i'(l_j) = x_i(S_h(l_j)$. Note that the change in the measured value depends both on the magnitude of the transversal shift $S_h(l_j)$ and the local properties of x_i around wavelength l_j. Considering the spectra plotted in Fig. 1, a transversal shift at band 2100 would lead to much less change in the intensity than a transversal shift of the same size at band 1450. We cannot control the transversal shift in the data, but we can control its impact by choosing our features accordingly. Bands with little local intensity differences in the measured material will cause smaller intensity changes given transversal shifts. To measure the local intensity differences on the discrete vectorial data representation, we rely on local differences $\Delta(l_j) = \frac{1}{2\eta}|\bar{x}(l_j - \eta) - \bar{x}(l_j + \eta)|$ with $\bar{x}$ being the datasets mean spectral signature and η the distance between measurements points in the given datasets. For the training, we only select $\{l_j \mid \Delta(l_j) < \frac{\alpha}{L}\sum_{j=1}^{L}\Delta(l_j)\}$ where α is a dataset specific hyperparameter. We refer to this strategy as r.f.s.

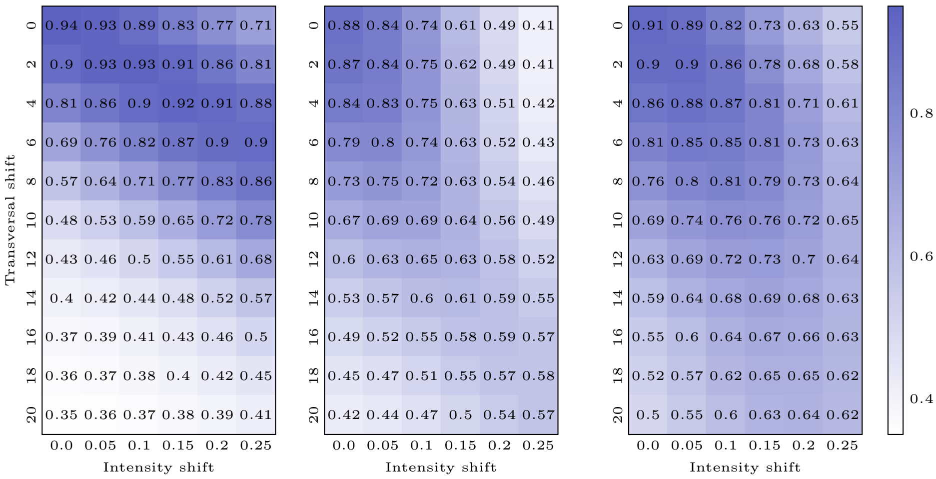
Results: We evaluate the proposed methods and their combinations on both artificial and real-world data[2]. The results of the isolated transversal and intensity shift are visualized in Fig. 2. While applying robust training, robust feature selection, and its combinations improve the vanilla performance under transversal shift, we observe a stronger decline for increasing intensity shifts. Especially the combination of feature selection and transversal robust training greatly improves the accuracy for large transversal shifts.

As we expect a combination of transversal and intensity shifts in many real-world applications, we additionally evaluate the methods on artificial datasets containing both shifts in varying magnitudes. The results are summarized in Fig. 4. In the vanilla setting, we observe a significant decline for increasing shift magnitudes. Transversal Robust training increases the accuracy mainly for scenarios with significant transversal shifts. In contrast, combining robust feature selection with intensity robust training yields a good compromise overall considered settings.

Finally, considering the real-world dataset, as summarized in Fig. 3, we can confirm the findings from the theoretical datasets. Again, the robust feature selection yields considerable accuracy increases of up to 0.2 for transversal shifts.

[2]Our code is available at `https://github.com/vvaquet/hyperspectral-sensor-shifts`

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.



(a) vanilla (b) rob. tr. (c) r.f.s. + rob. int.

Fig. 4: Accuracy scores of experiments on combined artificial shifts

6 Conclusion

In this contribution, we proposed algorithm-based approaches to cope with hyperspectral sensor shifts. In contrast to prior data-level approaches, our robust training methods do not pose any assumptions on the class distribution of the measurements. Besides, we demonstrated that they are suited for a combination of both transversal and intensity shifts.

References

[1] Muhammad Khan, Hamid Khan, Adeel Yousaf, Khurram Khurshid, and Asad Abbas. Modern trends in hyperspectral image analysis: A review. *IEEE Access*, 6:14118–14129, 03 2018.

[2] Clémence Goyens, Héloïse Lavigne, Antoine Dille, and Han Vervaeren. Using Hyperspectral Remote Sensing to Monitor Water Quality in Drinking Water Reservoirs. *Remote Sensing*, 14(21):5607, November 2022.

[3] Alberto Signoroni, Mattia Savardi, Annalisa Baronio, and Sergio Benini. Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review. *Journal of Imaging*, 5(5):52, May 2019.

[4] Valerie Vaquet, Patrick Menz, Udo Seiffert, and Barbara Hammer. Investigating intensity and transversal drift in hyperspectral imaging data. *Neurocomputing*, 505:68–79, September 2022.

[5] Hadi Salman, Andrew Ilyas, Logan Engstrom, Ashish Kapoor, and Aleksander Madry. Do adversarially robust ImageNet models transfer better? In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, pages 3533–3545, Red Hook, NY, USA, December 2020. Curran Associates Inc.

[6] John A. Lee and Michel Verleysen, editors. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer New York, New York, NY, 2007.

[7] Zhuang Qian, Kaizhu Huang, Qiu-Feng Wang, and Xu-Yao Zhang. A survey of robust adversarial training in pattern recognition: Fundamental, theory, and methodologies. *Pattern Recognition*, 131:108889, November 2022.