ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

Augmenting a convolutional neural network with local histograms - A case study in crop classification from high-resolution UAV imagery

Julien Rebetez[1a], Héctor F. Satizábal[1a], Matteo Mota[1c], Dorothea Noll[1c], Lucie Büchi[2], Marina Wendling[2], Bertrand Cannelle[1b], Andres Perez-Uribe[1a] and Stéphane Burgos[3]

1- University of Applied Sciences Western Switzerland (HES-SO)
a) HEIG-VD - IICT - Intelligent Data Analysis Group - Yverdon-les-Bains
b) HEIG-VD - G2C - Yverdon-les-Bains
c) Changins, Viticulture and Oenology - Nyon
Switzerland

2- Agroscope
Institute for Plant Production Sciences - Nyon
Switzerland

3- Bern University of Applied Sciences
School of Agricultural, Forest and Food Sciences HAFL - Zollikofen
Switzerland

Abstract. The advent of affordable drones capable of taking high resolution images of agricultural fields creates new challenges and opportunities in aerial scene understanding. This paper tackles the problem of recognizing crop types from aerial imagery and proposes a new hybrid neural network architecture which combines histograms and convolutional units. We evaluate the performance of the hybrid model on a 23-class classification task and compare it to convolutional and histogram-based models. The result is an improvement of the classification performance.

1 Introduction

In the past few years, the UAV[1] industry has grown from a niche market to mainstream availability, lowering the cost of aerial imagery acquisition and opening the way to many interesting applications. In agriculture, those new data sources can be used to help farmers and decision makers better understand and manage crops.

An automatic crop classification system leveraging UAV imagery would be useful for a number of studies. For example, in erosion risk assessment, an overview of the whole landscape upon several growers is necessary because the flow of water depends on soil coverage and annual changes (crop rotation). Other domains such as watershed management and crop yield estimation could benefit from such a classification too.

There are a number of previous works that use UAV imagery in an agricultural context. In [1], the authors used a neural network to classify different crops using remote sensed images to help administrations evaluate and target

[1]Unmanned Aerial Vehicle, or Drone

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

their agricultural subsidies programs. In [2], the authors designed a procedure to quantify the ground coverage of weed from UAV imagery to better target herbicides. In [3], the authors evaluate different vegetation indices to quantify vegetation coverage from UAV imagery[2]. Moreover, [4] and [5] explore the problem of segmenting fields from aerial imagery.

In the field of texture classification, numerous approaches use hand-picked filters to extract features which are then used by a classifier. In [6], the author proposed an improved Local Binary Patterns descriptor and in [7], a new histogram-based, rotation invariant approach is explored. In [8], the author explored the use of random projection to extract texture descriptors.

In this paper, we explore the use of a hybrid deep neural network which combines convolutional layers [9] with per-window histograms to increase crop classification performance. We show that the hybrid system performs better than either model individually and that the resulting classification map is of high quality.

2 Dataset

The dataset we use in this paper was built from aerial images of experimental farm fields issued from a series of experiments conducted by the Swiss Confederation's Agroscope research center. The particular image we used is shown in Figure 1 and covers a small zone ($\sim 100 \times 60$ m) in which different plant species were sown to perform agronomic research.

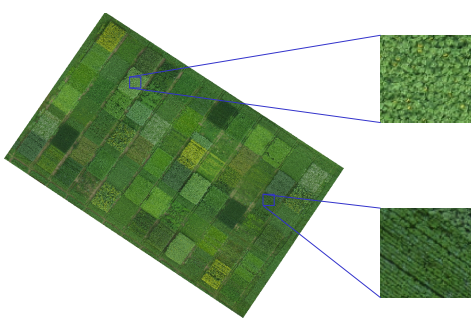


Fig. 1: The aerial image used in our experiments. Some zoomed regions are shown to highlight the different textures that appear depending on the crops.

Figure 2 shows the spatial distribution of the 22 different crops (23 if we count bare soil) that can be found in the area under study. The RGB image has a size of 2425×2175 pixels with a ground resolution of 5 cm. The crops are divided in small parcels of about 6×8 m, with 3 repetitions for each crop.

The dataset was split into a training part and a test part following two different policies. Figure 2 shows the distribution of the resulting folds in both

[2]The UAV used in this work is a Swinglet CAM from SenseFly. The camera is a compact Canon IXUS 220 HS with a CMOS de 12.1 MP sensor and a 24 mm equivalent focal length. Each image has a resolution of 3000 x 4000 pixels in RGB. The final mosaic is constituted of 100 assembled images.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

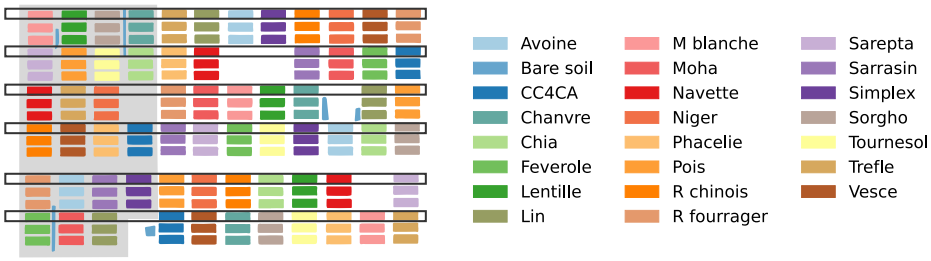


Fig. 2: Labels of the different crops present in the area under study. The black boxes show the test fold for experiment 0 and the gray area the test fold for experiment 1.

cases. The parcels in the black rectangles were used to build the test fold of experiment 0 (exp. 0) whereas experiment 1 (exp. 1) was built from the parcels in the light gray area. For some crops, the parcel in the lower-right part of the image is different from the parcel in the upper-left part, making experiment 1 more difficult.

3 Models

In order to infer the class of a pixel, we provide the models with the information of a window of size 21×21 of the pixel's neighbourhood[3].

The image shown in Figure 1 suggests that both color and texture seem to be important characteristics to discriminate between different crop classes. Hence, we used a deep neural network which consists of a convolutional side (CNN) which uses the raw pixel values and a dense side which uses RGB histograms (HistNN). The output of both networks was merged by a final layer which predicts the class of each pixel.

The input image is centered so that all channels are in the $[-0.5, 0.5]$ interval. The HistNN consists of two 32 units dense layers and is fed with three 20-bin histograms (one per channel). The CNN consists of two convolutional layers: the first with 48 11×11 filters and the second with 48 3×3 filters. Each one is followed by a 2×2 max-pooling layer.

The outputs of the *CNN* and the *HistNN* are merged into a dense layer of size 128 which is then used to predict a class probability amongst the 23 target classes using a softmax layer. All inner layers in the network have a rectified linear activation function.

To train our networks, we used the *Adam* [10] stochastic optimization implementation included in the *Keras* [11] deep learning library[4]. We kept the recommended parameters for the learning rate α ($\alpha = 0.001$), β_1 ($\beta_1 = 0.9$) and β_2 ($\beta_2 = 0.999$), and we used the multiclass log loss [5] objective function. We ran

[3] 21×21 pixels represents a square of 1 meter side on the ground.
[4] We used a *NVIDIA Tesla M2075* to train our models.
[5] Also known as categorical cross-entropy

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

the training for 60 epochs with a batch size of 256 samples and with additional dropout layers to reduce over-fitting. We observed that those settings led to good convergence. For the training phase, each window was rotated by 0°, 90°, 180° and 270° to force the learnt filters to be rotation invariant, resulting in a total of *160 000* examples.

We first trained the *CNN* and *HistNN* models separately. Then, we added the merging layer to obtain the *Merged* model which was then fine-tuned. Although we did not explicitly forbid updates to the *CNN* and *HistNN* layers during fine-tuning, we observed that fine-tuning mostly affected the merging layer.

4 Results

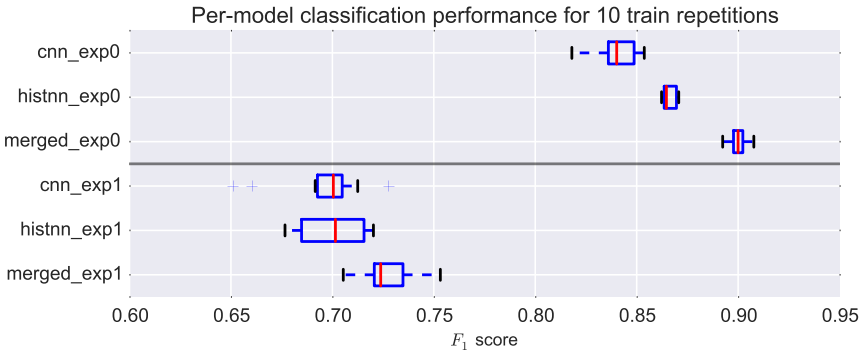


Fig. 3: F_1-score for different models in our two settings

Figure 3 shows the classification performance for the three models we tested. We see that experiment 1 is harder, due to the folding structure. In both cases, our hybrid CNN-HistNN network performs better than either model alone. In the resulting classification maps shown in Figure 4, one can see that HistNN makes coarser predictions while CNN looks noisier. Merging the two provides a homogeneous classification map inside the parcels. Although it exhibits better performance, one disadvantage of the Merged network is the training time. For one epoch, the CNN takes 53 seconds, the Merged network 57 seconds but the HistNN only 6 seconds.

Figure 5 shows the test F_1-score for each class and each model, averaged over 10 runs. For most classes, the HistNN performs as well or better than the CNN. For some classes such as *Lin* and *Niger*, the CNN performs better than the HistNN. In Figure 6, we compare an example window of the *Lin* class with a window of the *Simplex* class. We can see that the *Lin* window is wrongly classified as *Simplex* by the HistNN. Indeed, the histograms are very similar and therefore, it is not possible to distinguish those two classes based solely on color. For each class, we plot the output of the first convolution layer of the

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

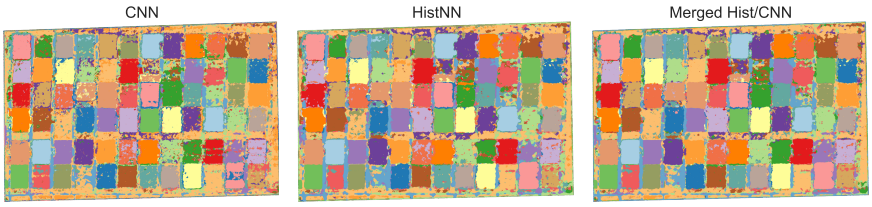


Fig. 4: Classification maps for experiment 0

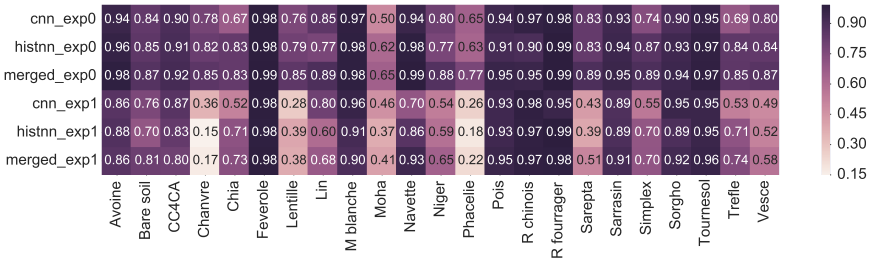
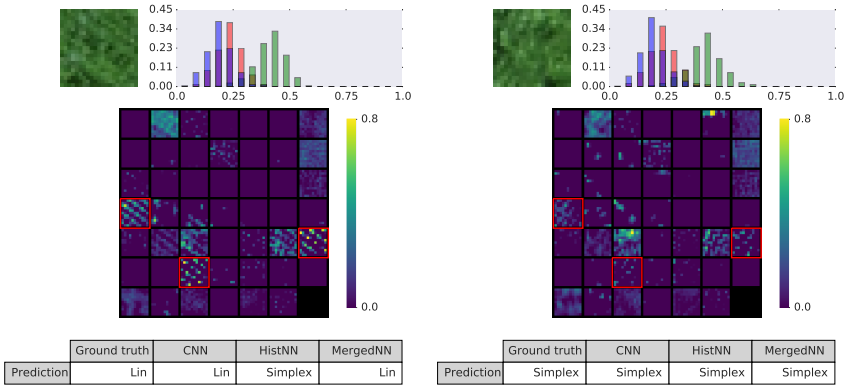


Fig. 5: Test F_1-scores for each model/experiment.

CNN applied to the window. We can see that for the *Lin*, the CNN is able to extract the diagonal structure of the image (highlighted by red squares in Figure 6). Similar observations can be made on other examples of the *Lin* class and this strengthen our intuition that both texture and color are important for classifying the crops in our dataset.



(a) Lin (b) Simplex

Fig. 6: Comparison of models on Lin and Simplex. Each plot shows the input RGB window, the corresponding histogram, the output of the first convolutional layer of the CNN and the predicted labels for each model.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

5 Conclusion

In this paper, we studied the classification of crops based on UAV aerial imagery. We proposed a hybrid CNN - HistNN deep neural network that is capable of using both color distribution and texture patterns to successfully classify a wide variety of crops. Our model exhibited good performance under two different folding policies, which shows the robustness of the approach. Further work should explore many parameters of the model, such as the number of filters in the CNN and the number of layers. It would be interesting to see if we can transfer our model to a new area with different fields, by only fine-tuning the last layer. Another interesting question would be to evaluate the ability of our model to classify the growth stage, which could be interesting for yield prediction.

Acknowledgments

This project has been funded by the University of Applied Sciences Western Switzerland (HES-SO), grant IA-INTERDISC13-04.

References

[1] Manuel Cruz-Ramírez et al. A multi-objective neural network based method for cover crop identification from remote sensed data. *Expert Systems with Applications*, 39(11):10038–10048, 2012.

[2] José Manuel Peña et al. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (uav) images. *PLoS One*, 8(10):e77151, 2013.

[3] J Torres-Sánchez, JM Peña, AI De Castro, and F López-Granados. Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from uav. *Computers and Electronics in Agriculture*, 103:104–113, 2014.

[4] Matthias Butenuth, Bernd-Michael Straub, and Christian Heipke. Automatic extraction of field boundaries from aerial imagery. In *KDNet Symposium on Knowledge-Based Services for the Public Sector*, pages 14–25, 2004.

[5] Carolyn Evans, Ronald Jones, Imants Svalbe, and Mark Berman. Segmenting multispectral landsat tm images into field units. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(5):1054–1064, 2002.

[6] Li Liu, Lingjun Zhao, Yunli Long, Gangyao Kuang, and Paul Fieguth. Extended local binary patterns for texture classification. *Image and Vision Computing*, 30(2):86–99, 2012.

[7] Li Liu, Yunli Long, Paul W Fieguth, Songyang Lao, and Guoying Zhao. Brint: binary rotation invariant and noise tolerant texture classification. *Image Processing, IEEE Transactions on*, 23(7):3071–3084, 2014.

[8] Li Liu and Paul W Fieguth. Texture classification from random features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(3):574–586, 2012.

[9] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[10] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[11] Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.