Conference papers

School of Computer Sciences

2021-07-14

TRANSFER LEARNING PERFORMANCE FOR REMOTE PASTURELAND TRAIT ESTIMATION IN REAL-TIME FARM MONITORING

Patricia O'Byrne
Technological University Dublin, patricia.obyrne@tudublin.ie

Patrick Jackman Dr.
Technological University Dublin, patrick.jackman@tudublin.ie

Damon Dr. Berry Dr.
Technological University Dublin, damon.berry@tudublin.ie

See next page for additional authors
Follow this and additional works at: https://arrow.tudublin.ie/scschcomcon

Recommended Citation

Authors

Patricia O'Byrne, Patrick Jackman Dr., Damon Dr. Berry Dr., Hector-Hugo Franco-Penya Dr., Michael French, and Robert J. Ross Dr

TRANSFER LEARNING PERFORMANCE FOR REMOTE PASTURELAND TRAIT ESTIMATION IN REAL-TIME FARM MONITORING

Patricia O'Byrne*, Patrick Jackman, Damon Berry, Hector-Hugo Franco-Peña, Michael French[†], Robert J. Ross[‡]
Technological University of Dublin, Ireland

ABSTRACT

In precision agriculture, having knowledge of pastureland forage biomass and moisture content prior to an ensiling process enables pastoralists to enhance silage production.While traditional trait measurement estimation methods relied on hand-crafted vegetation indices, manual measurements, or even destructive methods, remote sensing technology coupled with state-of-the-art deep learning algorithms can enable estimation using a broader spectrum of data, but generally require large volumes of labelled data, which is lacking in this domain. This work investigates the performance of a range of deep learning algorithms on a small dataset for biomass and moisture estimation that was collected with a compact remote sensing system designed to work in real time. Our results showed that applying transfer learning to Inception ResNet improved minimum mean average percentage error from 45.58% on a basic CNN, to 28.07% on biomass, and from 29.33% to 8.03% on moisture content. From scratch models and models optimised for mobile remote sensing applications (MobileNet) failed to produce the same level of improvement.

Index Terms— grassland biomass, Inception ResNet, MobileNet, proximal sensing, transfer learning.

1. INTRODUCTION

Remote sensing has long had applications to agriculture. However in recent years the potential for agricultural applications of remote sensing have grown many fold thanks to advances in miniaturisation and computational power. Such advances allow applications to challenges like the real time monitoring of vegetation traits during a treatment or cutting phase. In our case we consider this with application to silage production where farm machinery and the farmer can benefit from knowing the moisture and biomass levels of grass during a cutting process just prior to ensiling.

Taking dry matter as an example, the dry matter yield can be estimated approximately by an experienced farmer, or calculated using contact sensors such as rising plate meters or sensors such as spectrometers that can be used to calculate vegetation indices. Red and near-infrared bands are used by the well-known Normalized Difference Vegetation Index (NDVI), but there are a huge number of vegetation indices [1]. Vegetation indices are adequate but based on limited data, so require extensive calibration in relation to crop phenology, site, date and sensors. Satellite and aerial sensors are calibrated and can be used to provide summary biomass information over a wide area, with coarse granularity. Although satellite coverage is becoming more frequent, it is not on-demand. Vegetation indices can also be used by proximal hand-held sensors that often need to make contact with the substance to be assessed. Such sensors are useful in that they can be deployed when necessary, but are typically only useful in providing spot estimations.

Sensors, mounted on farming equipment, can provide a wider range of more fine-grained data, specifically targeting the farm site and can be used during day-to-day farming activities. We view these sensor systems as remote in the sense that unlike many vegetation index based instruments, almost direct contact with the target for analysis is not required. Advances in technology and edge computing allow multi-spectral and multi-modal sensing devices to be deployed to farm machinery, but challenges around the accuracy and applicability of results remain. Without large volumes of training data and robust modelling approaches, re-calibration of devices may be required if devices are used across multiple sites and this is far from trivial [2]. To provide a reusable method of estimation that is appropriate for small farms over

a wide area, more research is required [3].

Machine learning and specifically deep learning give us the potential to learn complex models for the estimation of traits such as biomass and moisture content, but these have been traditionally seen as being dependent on having a wide range of labelled data available. In an ideal scenario, a very large dataset of pasture samples, made up of labelled data from inexpensive sensors from diverse sites is needed. However, the collection and labelling of such samples is time-consuming, costly and dependent on weather conditions, access to harvest sites and labelling equipment. The authors are not aware of such a dataset being available publicly at the time of writing. Transfer learning based around large labelled datasets, such as ImageNet [4], enable the development of deep learning algorithms for classification and estimation [5, 6].

This paper is part of the investigation into the following hypothesis: can a dataset consisting of full RGB and near infrared (NIR) images, along with plant height, date, location, and labels such as biomass and moisture content be collected with inexpensive sensors to develop a model to estimate targets with minimal calibration. In the absence of a very large dataset, we investigate how successful a deep learning model can be, using a limited dataset but through taking advantage of transfer learning where appropriate within the input pipeline.

For processing image data, the most popular family of network types are the convolutional neural network (CNN), which takes advantage of local features that are invariant across a full image. These vary from shallow CNNs, which usually use a small number of convolutional layers and are typically well suited to small domains, through to highly sophisticated and very deep networks that apply multiple layers of processing, along with tweaks such as skip layers, to ensure that training is successful even for very deep networks [7]. The models chosen for this paper are a shallow CNN, Inception Resnet V2 and MobileNet, as described in Section 3. The methodology used was to design and deploy a data collection trolley and protocol, to collect data and use it to iteratively experiment with deep learning models. This paper is organized into sections, where Section 2 describes data collection, Section 3 describes the experiments that took place and Section 4 gives detailed results. The final section 5 gives the conclusion and intended further work.

2. DATA

Our study commenced by building a dataset for biomass and moisture estimation of grassland, prior to silage production.

An experimental platform was designed and deployed, housing a network of sensors that could move across grassland. The sensor array included a four-channel JAI AD-130 GE [1] camera, mounted approximately 150cm above ground,

pointing downwards. This takes two simultaneous images, an RGB image (964 x 1296 x 3) and a near infrared (NIR) image (966 x 1296). These combined images made up VIS-NIR data. A LiDAR-Lite v3HP [2] mounted beside the camera, recorded canopy height. Canopy height readings were also taken manually for validation purposes. To establish target values for biomass and moisture content, a $10cm^2$ square from the sensed area was harvested, weighed for biomass and dried in a HPP260 Memmert [3] oven at 60 °C for 24 hours, before being re-weighed to determine moisture content as a percentage. This target value will henceforth be known as dryness. Forty-three usable samples were collected across three different sites. Data for each sample included RGB and NIR images, LiDAR height, biomass and dryness labels. Biomass values ranged from 5,144 to 60,391Kg/Ha, with an average of 21,129.22Kg/Ha and a standard deviation of 15,532.58Kg/Ha. Dryness values ranged from 14.3% to 43.6% with an average of 26.7% and a standard deviation of 9.6%. Each image was checked to ensure all components were present and then split into 48 patches of 156 x 156 pixels. To enable cross validation, the dataset was divided into five sets, with all patches from a single sample assigned to the same set. The sets were combined into five folds, each with a training set (1156 patches) and a validation set (288 patches).

3. EXPERIMENTS

A number of analysis models were constructed to estimate biomass and moisture content, varying the model, the data input and the use of pre-trained weights. Specifically, we designed a baseline shallow convolutional neural network architecture, an Inception ResNet implementation and a MobileNet implementation, and where possible, compared both pre-trained and trained-from-scratch variants of these models.

Our baseline CNN takes in LiDAR and VIS-NIR patch data. The VIS-NIR data is fed through four convolutional layers, each with a 3x3 kernel size with filters of 32, 32, 64 and 64 respectively. Max-pooling (2,2) is applied after the second and fourth convolutional layer. For biomass training, the output of the second pooling layer is flattened and subsequently concatenated with the scalar value from the LiDAR sensor, before being fed through two fully connected layers. No LiDAR information is used for dryness training. A number of variants on this baseline CNN network were investigated, that varied on the basis of the data supplied. In addition to making full use of LiDAR and VIS-NIR data (BASIC model), variants were also constructed and trained that omitted LiDAR (VISNIR model) and used only RGB data (RGB model). Each of these models was run for 1,000 epochs and trained for targets biomass (B) and dryness (D) simultaneously.

While our shallow models provided a useful baseline, they do not facilitate the true power of deep networks, so a number of alternate training models were constructed, that build on state-of-the-art deep neural network architectures. Inception Resnet V2 [6] aims to overcome issues in scale invariance, through the application of heterogeneous kernel architectures and factorisation of large networks to produce predictors that give accurate estimations. As more convolutional layers are added to a network, the output becomes sparser and the ability to back-propagate error signals becomes more difficult, resulting in diminishing performance. To overcome this issue, residual networks reinforce learning by using residual layers, or skip connections, where the output from an earlier layer is added to the output of a later layer, reinforcing the input to further layers [8]. Inception-ResNet v2 [6] includes these skip connections. We used InceptionResNetV2 [4], an implementation of this model that has been widely used across the image processing community. Inception Resnet v2 was run on three-channel data, from scratch (IR-RGB) and with weights pre-trained on ImageNet (IR-RGB-PT) and also on four-channel data (IR-VISNIR) from scratch. Both targets were trained for independently.

One challenge with deploying models such as Inception ResNet in fields such as agricultural machinery, is that they require significant computational resources. Given such challenges, MobileNet V2 was designed for use on mobile devices, specifically for embedded computer vision applications using RGB data. MobileNet uses a combination of multiple depth-wise and point-wise convolution layers to replace fewer, more resource-hungry convolutional layers. MobileNet V2 introduced residual connections to reinforce feature maps, and bottleneck layers to compress the data [9]. We apply MobileNet V2[5] to assess the quality of low-resource optimised networks for our purposes, separately for both targets, from scratch (MobileNet model) and with pre-trained weights from ImageNet (Mobile-PT model).

Due to the improvement in learning pattern observed and corresponding slowness in training IncResNetV2 and MobileNetV2 were run for 300 epochs. All models were implemented on a Keras front-end to Tensorflow and trained on Nvidia K40 GPUs. Due to the size of our dataset it was not possible to partition out a holdout test dataset. However, five-fold cross validation and averaging of results was applied to all models, using the fold sets described in Section 2, allowing us to generalise our results somewhat. To check the accuracy of our model, we calculated a precision error PE_t for each target t, which was the average fractional difference between each validation target value V_t and the average training value $\overline{T_t}$. $PE_t = \frac{1}{n} \sum_{i=1}^{n} \frac{|\overline{T_t} - V_t|}{\overline{T_t}}$ where n is the number of test cases. The average overall precision error was PE_B

---

[4]https://www.tensorflow.org/api_docs/python/tf/keras/applications/inception_resnet_v2
[5]https://www.tensorflow.org/api_docs/python/tf/keras/applications/mobilenet

	RGB	NIR	LiDAR	Weights	Target	Training	Validation
					Inputs		
BASIC	Y	Y	Y	N	B	3.03%	45.48%
VISNIR	Y	Y	N	N	B	0.12%	55.93%
RGB	Y	N	N	N	B	0.07%	46.16%
BASIC	Y	Y	Y	N	D	13.15%	29.33%
VISNIR	Y	Y	N	N	D	5.21%	31.51%
RGB	Y	N	N	N	D	6.56%	30.54%
IR-RGB	Y	N	N	N	B	14.87%	28.35%
IR-VISNIR	Y	Y	N	N	B	10.36%	31.44%
IR-RGB-PT	Y	N	N	Y	B	4.65%	**28.07%**
IR-RGB	Y	N	N	N	D	4.89%	9.42%
IR-VISNIR	Y	Y	N	N	D	3.62%	**8.03%**
IR-RGB-PT	Y	N	N	Y	D	2.80%	8.83%
MobileNet	Y	N	N	N	B	10.13%	41.95%
Mobile-PT	Y	N	N	Y	B	7.70%	38.05%
MobileNet	Y	N	N	N	D	4.73%	12.35%
Mobile-PT	Y	N	N	Y	D	3.39%	9.16%

Table 1. Cross-validated MAPE results for Biomass (B) and Dryness (D)

= 56.6% for biomass and PE_D = 31% for dryness. Mean Square Error was used as the loss function when training all models.

4. RESULTS

Results of the application of these models to the biomass and dryness prediction tasks are presented in terms of Mean Average Precision Error (MAPE), for both training and validation data. This is a popular metric for regression problems, allowing a comparison of model performance across different target variables by normalising for the scale of the target. MAPE values shown are based on average minima per fold. Table 1 presents results from applying each variant of all models to the prepared folds. The baseline CNN model variants are the BASIC model, the VISNIR model and the RGB model. The validation results show marginal differences between the three architectures, with the VISNIR performing worst and BASIC performing best. However, there is strong evidence of overfitting, as training errors are very low, suggesting a memorisation effect on the training data, rather than a generalised model. This is more prominent in the VISNIR and RGB models than in BASIC, suggesting that the introduction of the LiDAR data allowed some degree of stronger generalisation than was possible for the purely image-based processing approach. These results represent a moderate improvement on our baseline PE_B, but only a marginal improvement in baseline PE_D. The Inception ResNetV2 model variants

are IR-RGB, IR-VISNIR and IR-RGB-PT. The improvement in performance over the set of shallow baseline CNN models is quite stark. After just 300 epochs, the biomass validation MAPE has improved by over 14% and dryness by over 20%. Performance was boosted further, by priming the model with pre-trained weights. Whilst adding NIR data resulted is a slight improvement in model performance for dryness, it had no positive effect on results for biomass. The use of pre-trained weights gave improvements for both dryness and biomass over the baseline RGB model – though the improvement for dryness results was not as large as with the application of NIR data.

The MobileNet model variants are MobileNet and Mobile-PT. As MobileNet is optimised for low-resourced devices, the results seen for this model are not as strong as for the IncResNet model. However, MobileNet does outperform the baseline CNN for both biomass and dryness and its performance improves again with the use of pre-trained weights.

5. CONCLUSION AND FURTHER WORK

Given the cost of data collection in farming domains, it is useful to know whether typically data hungry methods such as deep learning can still be beneficial in developing intelligent remote sensing applications. Despite the very low volumes in our dataset, our results demonstrate that a variety of models do provide a clear improvement over baseline mean based estimators (PE_t) and that very deep models outperform more basic architectures or architectures that are optimised to low resource devices. We observed that performance consistently improved when using pre-trained ImageNet weights, despite the fact that ImageNet images are generally at a different scale than our data. This suggests the usefulness of applying pre-trained models, in keeping with the assumptions of transfer learning, where very generic features can be learned. When NIR data was omitted there was also a decrease in performance. Transfer learning in this domain is hampered by the lack of readily available pre-packaged pre-trained models for processing multi-channel data. It is interesting to consider the performance of the MobileNet model with respect to both our baseline CNN and the IncResNet model. Despite not matching the performance of IncResNet, the MobileNet model provided an 18.5% absolute performance increase (32.7% relative improvement) over PE_B and an increased performance of 21.9% absolute (70.6% relative) over PE_D. Therefore, the usefulness of such models should not be discounted.

Finally, in our basic CNN implementation we found the addition of LiDAR information resulted in the best performing model, even though training MAPE values were higher than in the RGB and VIS-NIR variants. We have not expanded the analysis of LiDAR information to IncResNet and MobileNet models at this point. However, it is very interesting to note how additional information provided to our models may have had a regularising effect on the training process.

The experiments shown in this paper are essentially a starting point and leave a lot of questions unanswered. Firstly, it would be worthwhile developing a model that combined LiDAR with the deeper models. GPS data and collection date could also be added, thereby broadening the opportunities for learning. Secondly, although weights are not readily available to implement transfer learning on VIS-NIR data, if RGB data is put through Inception Resnet V2, NIR data could be put through another model with weights for grey-scale images and the results combined. Finally, it would be well worth investigating how collection of a lot more data over diverse and distributed sites would affect the estimation process.

6. REFERENCES

[1] Jinru Xue and Baofeng Su, "Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications," *Journal of Sensors*, 2017.

[2] Abdou Bannari, D Morin, F Bonn, and Alfredo Huete, "A review of vegetation indices," *Remote Sensing Reviews*, vol. 13, pp. 95–120, Jan. 1996.

[3] Alem Gebremedhin, Pieter E. Badenhorst, Junping Wang, German C. Spangenberg, and Kevin F. Smith, "Prospects for Measurement of Dry Matter Yield in Forage Breeding Programs Using Sensor Technologies," *Agronomy*, vol. 9, no. 2, pp. 65, Feb. 2019, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 2009, pp. 248–255, IEEE.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[6] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, Adaptive Computation and Machine Learning. The MIT Press, Nov. 2016.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 2016, pp. 770–778, 2016, arXiv: 1512.03385.

[9] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2 Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 4510–4520, IEEE.