Remote Sensing Letters, Volume 11, Issue 12, 2020, pp. 1080-1089
DOI: 10.1080/2150704X.2020.1821120

LETTER

Fully convolutional neural nets *in-the-wild*

Daniel M. Simms

School of Water, Energy and Environment, Cranfield University, College Road, Cranfield,

MK43 0AL, UK;

ARTICLE HISTORY
Compiled September 4, 2020

ABSTRACT
The ground breaking performance of fully convolutional neural nets (FCNs) for semantic segmentation tasks has yet to be achieved for landcover classification, partly because a lack of suitable training data. Here the FCN8 model is trained and evaluated in real-world conditions, so called *in-the-wild*, for the classification of opium poppy and cereal crops at very high resolution (1 m). Densely labelled image samples from 74 Ikonos scenes were taken from 3 years of opium cultivation surveys for Helmand Province, Afghanistan. Models were trained using 1 km^2 samples, subsampled patches and transfer learning. Overall accuracy was 88% for a FCN8 model transfer-trained on all three years of data and complex features were successfully grouped into distinct field parcels from the training data alone. FCNs can be trained end-to-end using variable sized input images for pixel-level classification that combines the spatial and spectral properties of target objects in a single operation. Transfer learning improves classifier performance and can be used to share information between FCNs, demonstrating their potential to significantly improve land cover classification more generally.

KEYWORDS
Semantic Segmentation; Convnets; Landcover Classification; Opium Poppy; CNNs; FCN8

1. Introduction

Deep fully-convolutional neural networks (FCNs) are a relatively new type of convnet for image classification (semantic segmentation) that address the limitation of convolutional neural networks (convnets or CNNs), with one or more fully connected layer, that are trained with image subsets (patches) with a single label per sample. FCNs work end-to-end, with pixel-to-pixel training and inference on images of varying size (Shelhamer, Long, and Darrell 2017), encoding spatial and spectral information at the pixel level during training and without the extra processing required to reconstruct a pixel-by-pixel classification for a whole scene (Stoian et al. 2019).

A major barrier to the development of convnets for landcover classification in general is a lack of suitable data for training and validation (Zhu et al. 2017). While recent work has sought to increase the number of training images available (Helber et al. 2019) these datasets do not have the dense labelling required for end-to-end training of FCNs. Other work on FCNs for landcover classification in operational contexts

CONTACT D. M. Simms. Email: d.m.simms@cranfield.ac.uk

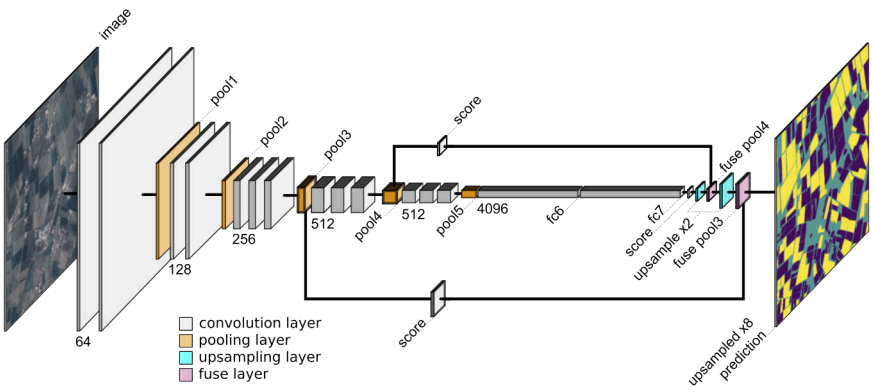


Figure 1. Fully convolutional deep neural network FCN8, showing relative sizes (2D) and labelled layer depth after max pooling, upsampling skip-architecture and scoring layers.

has shown improvements where texture and context add extra information for class separation, such as urban environments, at the resolution of Sentinel2 (10 m) (Stoian et al. 2019).

Combining the spectral and spatial properties of distinct groups of pixels into real-world objects becomes more important as resolution increases and class complexity increases (Blaschke et al. 2014). Studies using convnets with remote sensing imagery at very high resolution (VHR) are confined to a small number of datasets, such as UC Merced (Yang and Newsam 2010) and the Aerial Image Dataset (Xia et al. 2017), used for benchmarking. As such, there been limited research into how convnets perform on VHR images *in-the-wild*, a term that refers to testing in highly complex, time-varying environments closer to a real world application, instead of using highly controlled datasets (Zafeiriou et al. 2017). Realising the full potential of convets for landscape classification requires larger datasets that capture the full range of landscape variability (Krizhevsky, Sutskever, and Hinton 2012) and the variation in spatial and spectral resolution in remote sensing images.

This study trains and evaluates a FCN for opium poppy and cereal crop classification in a real world application, using data from operational opium-cultivation surveys. The large number of VHR Ikonos scenes were part of a statistically representative sample of crop variability. The accuracy of the dense labelling has been extensively quality checked and scrutinized, as these data were used to directly inform counter narcotics policy in Afghanistan (Taylor et al. 2010). FCNs were trained end-to-end for field parcel detection using 1 km^2 image regions, sub-sampled patches, and transfer learning across three years of data. The results are a first look at the performance of FCNs for landcover classification from VHR imagery *in-the-wild*.

2. Materials and methods

2.1. *Fully convolutional deep neural net FCN8*

The FCN8 model (Shelhamer, Long, and Darrell 2017) was chosen because of its performance on semantic segmentation and the relative simplicity of its architecture (figure 1). It is based on the VGG16 convnet (Simonyan and Zisserman 2015), with all fully-connected layers replaced with convolutional (conv) layers, ensuring that each pixel of the image is *path-connected* to the output.

Each conv layer in the model convolves the input layer with a set of kernal weights,

or *filters*, that produce an output of $h \times w \times d$, where h is the height, w is the width, and d is the depth or number of bands of the layer. The elements of the next layer (vector $\boldsymbol{y}$ at location i, j) are computed from the previous layer ($\boldsymbol{x}_{i,j}$) by

$$\boldsymbol{y}_{i,j} = f_{k,s}(\{\boldsymbol{x}_{si+\delta i, sj+\delta j}\}_{0 \leq \delta i, \delta j < k}) \tag{1}$$

where the size of the kernal (k) and the stride (s) map the spatial input region of the layer, known as the receptive field, at offsets $(\delta i, \delta j)$ such that $0 \leq \delta i, \delta j < k$. $f_{k,s}$ is the matrix multiplication for the conv layer. Pooling layers reduce the number of parameters in the spatial domain, downsampling the input layers to encode more complex spatial structure. The spatially dense output is rebuilt from the pooled layers by upsampling. Skips are used to *fuse* finer spatial detail from lower layers by summing with the upsampled layers, improving the pixel-wise accuracy. All layers are *scored* before fusing by matching the depth of the layer to the number of possible labels.

The FCN8 model code[1] was rewritten for TensorFlow (Abadi et al. 2016) with model weights initialized from VGG16 using the same procedure as in Shelhamer, Long, and Darrell (2017): reshaping the weights in fully-connected layers to match the dimensions of the replacement conv layers, and initializing the upsampling filters to bilinear interpolation. The optimizer used during training was changed to *Adam* (Kingma and Ba 2014) as it was found to be more stable than training with *Momentum*. The total loss for each training step was the cross entropy loss for each element in the dense output summed across the spatial dimensions (h and w).

2.2. *Labelled image data*

Labelled image data were taken from opium cultivation surveys conducted between 2007 and 2009 for the Afghanistan province of Helmand. These data were created by image-interpretation of very high resolution satellite imagery (VHR) as part of a statistical survey to calculate a yearly estimate of the area under opium cultivation. Each sample consisted of a 1 km×1 km subset of a pan-sharpened Ikonos image (4-band, 1 m resolution) and a manually digitized label of opium poppy and cereal fields on the same 1 m resolution pixel grid as the image. The locations were selected from an unaligned random sample within a 10 km sampling grid of Helmand Province (figure 2). Ikonos images were targeted within three week windows to coincide with the flowering period of poppy crops in each 10 km square to maximize the discrimination between classes (Simms et al. 2014). Image interpretation and field delineation were completed on-screen from true-colour and false-colour (near-infrared) Ikonos composites by trained interpreters with rigorous quality control.

The complete 2007 to 2009 dataset contained 281 samples, each of $\sim 1 \times 10^6$ pixels, digitized from 74 Ikonos scenes (table 1). The labels (or classes) were *no-data*, *unused*, *poppy*, *cereal*, and *other*. *No-data* areas were present in some samples located at the edges of images, *unused* areas were those that were excluded from the original survey because of cloud or haze. Each yearly dataset was split randomly into a 75% set for training and a 25% hold-out set for validation. The data were normalized by first scaling to 8bit and then subtracting the dataset mean from each band to be consistent with the approach used for training VGG16.

The 2008 data were used to create a set of 1,200 non-overlapping image patches to compare online training, using the whole sample, with mini-batches. A patch size

[1] https://github.com/dspix/deepjet

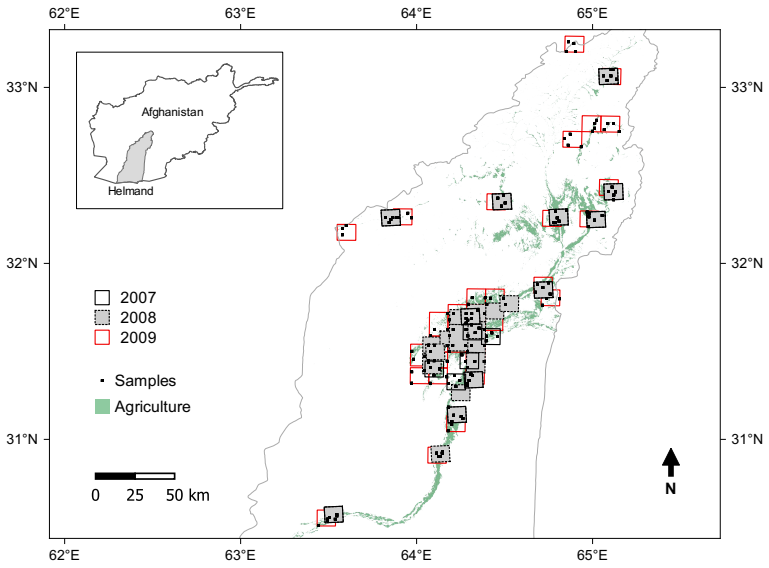


Figure 2. The 2008 agricultural area of Helmand Province, Afghanistan, showing Ikonos image collections (2007 to 2009) and sample distribution for all years.

Table 1. Number of Ikonos images and 1 km^2 samples used to train and evaluate the FCN8 models.

Year	No. images	No. samples
2007	17	83
2008	25	75
2009	32	123

of 227×227 pixels was chosen to match the input size of VGG16 and a batch size of 16 to match the size (in pixels) of each mini-batch to the size of an individual 1 km^2 sample.

2.3. *Model training and validation*

Separate FCN8 models were trained for each year using input images with 3 and 4 bands. For 3-band input a true-colour composite was chosen to match the colour channels of the original trained VGG16. The input depth was extended to 4 layers to test the performance on all bands of the Ikonos images, with initial weights for the increased depth taken from band 3 of VGG16. *Online* training was conducted for all layers of the model by feeding a single image per training step. Transfer learning across the 3 years of data was investigated by fine-tuning. All layers of the trained 2007 model were fine-tuned using 2008 data, with the resulting model further fine-tuned using the 2009 data.

A 2008 model was trained using image patches to test the effect of the number of training samples compared to the size of the labelled images. All training was carried out on a NVIDIA Quadro M5000 GPU over 1000 epochs with a learning rate of 10^{-5}, where an epoch is one complete pass through the training data. Total training time was 20, 17 and 24 hours for the 2007, 2008 and 2009 data respectively.

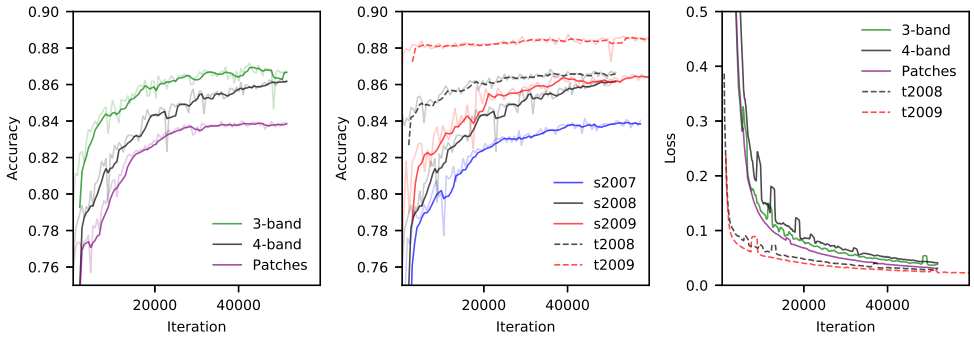
Output classifications for all FCN models were validated on the hold-out data. Overall pixel accuracy was calculated as the proportion of correct predictions, $\sum_i n_{ii} / \sum_i t_i$

where n_{ij} is the number of pixels of class i predicted as class j and t_i is the total number of pixels in class i. User accuracy, n_{ii}/t_i was used to assess the individual class agreement in the output predictions. The agreement between the segmentation in the prediction and the labels was measured using the frequency weighted intersection over union (fIoU), calculated as $(\sum_k t_k)^{-1} \sum_i t_i n_{ii}/(t_i + \sum_j n_{ji} - n_{ii})$ where k is the number of classes.

3. Results

3.1. *Model training*

Online learning using full-size samples was faster than batch training using patches and resulted in higher validation accuracy (figure 3(a)). Models with 4 and 3 spectral input bands had similar final validation accuracies (figure 3(a)). Transfer-trained models converged faster than models for individual years and achieved higher classification accuracies (figure 3(b), 3(c)).



(a) (b) (c)

Figure 3. Overall pixel accuracy during training for (a) images with 3 and 4 bands, image patches, (b) yearly FCN8 models from 2007 to 2009 (s2007 to s2009) and transfer-trained FCN8 models for 2008 and 2009 (t2008 and t2009), and (c) mean cross entropy loss. Each iteration is equivalent to a single sample (online) training step.

3.1.1. Classification accuracy

Overall pixel accuracies for the single-year classifications were between 73% and 87% for all validation datasets. The single year models performed 1% to 4% better on the validation dataset from the same year. FIoU was between 71% and 78%, with 2008 and 2009 models giving higher values for validation data from the same year. Poppy user accuracy ranged from 54% for the 2007 model using the 2009 validation data, to 80% for the 2009 model using the 2007 validation data. User accuracy for cereal followed a similar pattern while the other class had higher, or the same, values from validation on the same year as the model. Accuracy statistics for the three yearly models using all three validation datasets are shown in table 2.

The models trained by transfer-learning had improved overall pixel accuracy, FIoU, and user accuracy for all classes except poppy. The best performing model overall was 2009, transfer-trained from the 2007 and 2008 model using 2009 data, with an overall accuracy of 88%, an fIoU of 80%, and user accuracies of 67%, 86% and 92% for poppy, cereal and other classes respectively (table 2).

Table 2. User accuracy for poppy, cereal and other classes, total pixel accuracy and frequency weighted intersection over union (fIoU) for three years of validation data (07 to 09) using models trained for a single year (s) and transfer-trained models (t) starting with the 2007 single year model (s2007).

	Data	s2007	s2008	s2009	t2008	t2009
Poppy user	2007	0.77	0.73	0.80	0.73	0.82
	2008	0.68	0.77	0.77	0.74	0.77
	2009	0.54	0.6	0.65	0.57	0.67
Cereal user	2007	0.78	0.76	0.66	0.79	0.69
	2008	0.65	0.83	0.61	0.86	0.65
	2009	0.78	0.87	0.83	0.89	0.86
Other user	2007	0.89	0.72	0.78	0.79	0.82
	2008	0.89	0.89	0.89	0.91	0.92
	2009	0.89	0.86	0.90	0.89	0.92
Accuracy	2007	0.84	0.73	0.76	0.78	0.80
	2008	0.82	0.86	0.83	0.87	0.85
	2009	0.83	0.84	0.87	0.85	0.88
FIoU	2007	0.73	0.62	0.65	0.66	0.68
	2008	0.71	0.77	0.73	0.77	0.75
	2009	0.74	0.76	0.78	0.77	0.80

Segmented Field boundaries were consistent with those drawn for poppy and cereal during image-interpretation in areas with homogeneous crop cover (figure 4(a)) and also in more marginal crops, where the model was able to group areas of heterogeneous crop cover into field parcels (figure 4(b)). In late images, where poppy crops had started to senesce, the model was able to detect poppy fields despite their brown appearance although the field sizes were underestimated (figure 4(c)). The model was also able to detect small fields in remote areas dominated by the background class (figure 4(d)) across a range of landscapes. Significant errors in classification were found in samples containing shallow water with high sediment load (figure 4(e)).

The classified data was compared with the interpreter-digitized samples to test the use of FCN8 for poppy area estimation as part of a statistical survey. Comparison of the sample proportion of poppy between the classified and the validation samples using orthogonal regression showed a good fit (coefficient of determination $R^2 = 0.89$) with a small positive bias (figure 5). The root mean squared error (RMSE) was just below 2% (proportion 0.019)

4. Discussion

FCNs combine the spatial and spectral discrimination of landcover into a single operation. The receptive field of the network encodes spatial structure at the pixel level, which enables the FCN to join heterogeneous groups of pixels into meaningful objects based solely on the training data. The ability to transfer an interpretation key or classification schema directly to the FCN during training makes training very efficient. This is an advantage over an object based classifier, which requires a separate segmentation step before classifier training and groups segments into objects based on scale, similarity or through further modelling (Blaschke et al. 2014).

The variation in timing, landscape and crop management in the image dataset (74 VHR images) represented the highly complex environment of opium production in Helmand Province. The FCN8 output closely matched the interpreter drawn bound-

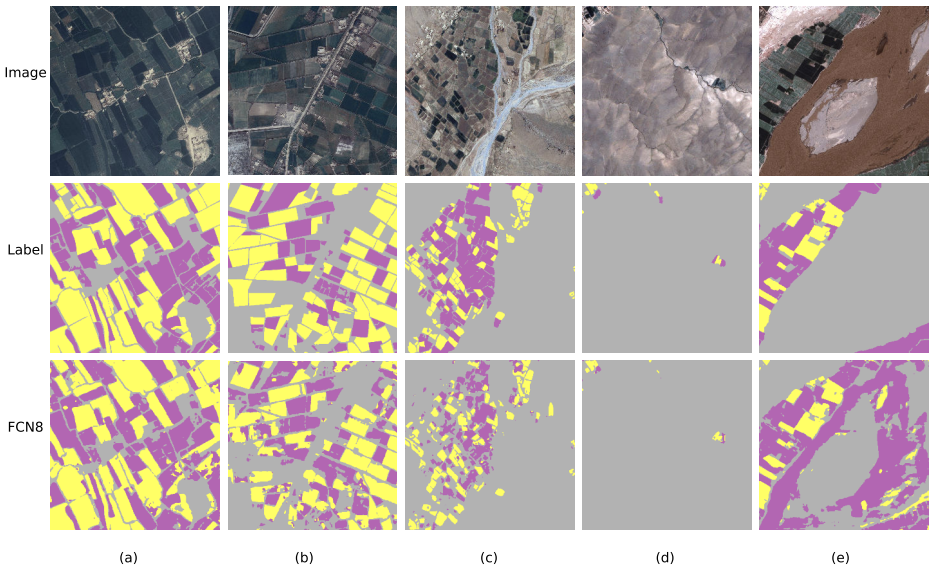


Figure 4. Classification of Ikonos validation-samples using FCN8, transfer-trained using three years of data (2007 to 2009). The top row is the true-colour pan-sharpened image, the centre row the labelled interpretation and the bottom row is the FCN8 classification; (a) area of high density uniform crops, (b) crops with variable establishment, (c) later image collection with fields senescing, (d) small isolated fields, and (e) misclassification of sediment loaded river in flood. Colour key: poppy is purple, cereal is yellow and other is gray.

aries for poppy and cereal fields of varying size and, in the case of poppy, was able to segment field parcels including sub-field areas of poor establishment or within field features that were smaller than the minimum mapping unit of the interpretation key. Isolated fields in remote locations, where images are dominated by the background class, were detected with few false positives showing the FCN is able to detect fields at a range of scales. One area of significant commission error was related to a flooded river with a high sediment load (figure 4) that was not present in any of the training samples. This highlights the requirement for a representative sample of landcover for supervised training, including the background class.

It is worth noting that poppy fields are complex objects that exhibit significant variation in both their spectral and spatial properties. Even skilled interpreters will encounter uncertain cases, particularly in more marginal areas or away from sources of irrigation, where crops canopies are sparse. As such there may be error within the classifier related to uncertainty in the training data or the images used for validation.

Improvements in FCN8 classification accuracy were gained by transfer learning across growing seasons, with models also able to generalize better to other years. Training was also much faster demonstrating a significant advantage of FCNs for classification by fine-tuning existing models using new data. However, model performance was still better for the specific year on which the were fine-tuned. There were some variations in the timing of images relative to the crop growth stage but the results show that crop appearance is related to the environmental variation found in a particular growing season. Adding further years of data would likely increase the variability in data used to train the model, improving the overall accuracy and the ability of the model to generalize.

Training with image patches in mini-batches showed no improvement in the speed of training, as also reported in Shelhamer, Long, and Darrell (2017). The lower final accuracy of the patch-trained FCN8 is thought to be related to the incomplete sub-

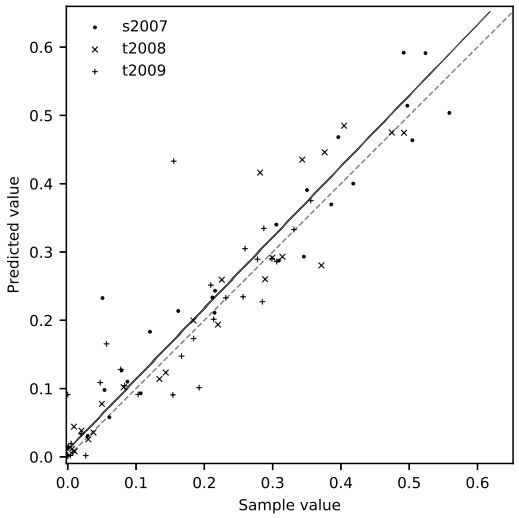


Figure 5. Proportion of poppy in each validation sample compared to predicted proportion of poppy from transfer-trained FC8 models. Orthogonal regression R^2 of 0.89 and an RMSE of 0.019, dotted line is one-to-one.

sampling of the dataset caused by the non-overlapping grid used to clip the patches. Subdividing image data to artificially create a greater number of samples to train an FCN was found to have no advantages. There are advantages to patch-wise training using mini-batches, such as correcting for class imbalance by sampling and selecting patches from across the total variance within each mini-batch. However, online training was found to be more efficient and faster than training with patches.

The use of online training using images of different dimensions is a significant advantage of an FCN over a convnet with one or more fully-connected layers, especially for remote sensing data, where images are large in terms of the number of pixels and where resampling to fixed input dimensions can create distortion. Training online using variable sized input removes the requirement to pre-process dense scenes into fixed-size inputs but another advantage is the ability to pass large input scenes to the FCN for classification, with the only limitation being the available computer memory. This means a complete 10×10 km Ikonos scene can be classified much faster using the FCN (inference time of less than 10 s) than by reconstructing from a moving window with fixed-size input, as in a convnet with fully-connected layers.

The increase in spectral information in the 4-band FCN8 did not improve the classification as might be expected considering the relationship between near-infrared reflectance and vegetation. This would suggest that the spatial characteristics are more important for the classification of the target fields. However the relative effect of the spatial versus the spectral information is unknown. The FCN8 architecture itself, being designed for images taken with consumer cameras, may be limiting the accuracy of the classification. The normalization step for input data is also unlikely to be optimal for remote sensing data, with its greater dynamic range, increased number of channels and constant scale compared to consumer camera images. FCN architectures specific to remote sensing imagery require further investigation and are an exciting new area of research.

The FCN outperforms a object-based classifier on similar data from 2009 (Simms et al. 2016) with higher classification accuracy (7% overall 8% for poppy) over a larger Ikonos image dataset (32 versus 14) with greater variation in crop growing

conditions. Part of the reason for this is the greater number of samples available through transfer learning. In the object-based approach each Ikonos image is classified separately, requiring a training dataset per image with the reuse of sample data limited to the determination of hyper-parameters for the segmentation algorithm used. In contrast the FCN can be trained across image datasets, either using all the available data or fine tuned to specific images, greatly increasing the amount of training data and making the process of supervised training much more efficient. Training the FCN also requires minimal pre-processing of the image and label data and while training can take hours it is largely free from user input.

Despite the improvement in accuracy and reduction in the bias (figure 5), FCN8 is not a suitable replacement to the manual interpretation for cultivation statistics without correction of the systematic error (Gallego 2004). However, the advantage for image classification is clear; whole scene classification with transfer training across multiple images means that there are many more examples per image and fewer processing steps compared with object-based classifiers and convnets with fully-connected layers. FCNs also show great potential for the classification of landcover more generally, as information can be shared between classifiers through transfer learning.

5. Conclusion

The overall pixel accuracy of the FCN8s tested was 84% to 88% across three years of data. Transfer trained models, using all previous years, performed better than models trained on a single year. Model generalisation across all three years improved with transfer learning but models performed better on validation data from the same year on which they were fine-tuned. Accuracies were found to be higher that object based classification (7%) for similar data with fewer processing steps and the ability to train across images datasets and across growing seasons.

FCNs were found to be efficient for pixel-wise classification of VHR satellite imagery with improvements over convnets with fully-connected layers. Training and classification using variable sized image subsets was better than using image patches. Segmented poppy and cereal fields were consistent with manual interpretation, with the FCN able to merge heterogeneous groups of pixels into complex objects based on training data alone, combing the spatial and spectral characteristics of target objects.

The results are significant as they evaluated *in-the-wild*, using a statistical sample of opium poppy and cereal cultivation from an extensive image datasets collected for operational cultivation surveys. They show the potential of FCNs for improving supervised landcover classification at very high resolution.

Data Availability Statement

Data sharing is not applicable to this letter as no new data were created or analyzed in this study. The code is available at https://github.com/dspix/deepjet

References

Abadi, Martin, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. “TensorFlow: A system for large-scale machine learning.” In

12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 265–283. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

Blaschke, Thomas, Geoffrey J. Hay, Maggi Kelly, Stefan Lang, Peter Hofmann, Elisabeth Addink, Raul Queiroz Feitosa, et al. 2014. "Geographic Object-Based Image Analysis - Towards a new paradigm." *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 180–191. http://dx.doi.org/10.1016/j.isprsjprs.2013.09.014.

Gallego, F J. 2004. "Remote sensing and land cover area estimation." *International Journal of Remote Sensing* 25 (15): 3019–3047. http://www.informaworld.com/10.1080/01431160310001619607.

Helber, Patrick, Benjamin Bischke, Andreas Dengel, and Damian Borth. 2019. "EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (7): 2217–2226. https://ieeexplore.ieee.org/document/8736785/.

Kingma, Diederik P., and Jimmy Ba. 2014. "Adam: A Method for Stochastic Optimization." Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015, http://arxiv.org/abs/1412.6980.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, USA, 1097–1105. Curran Associates Inc. http://dl.acm.org/citation.cfm?id=2999134.2999257.

Shelhamer, Evan, Jonathan Long, and Trevor Darrell. 2017. "Fully Convolutional Networks for Semantic Segmentation." *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (4): 640–651. https://doi.org/10.1109/TPAMI.2016.2572683.

Simms, Daniel M, Toby W Waine, John C Taylor, and Timothy R Brewer. 2016. "Image segmentation for improved consistency in image-interpretation of opium poppy." *International Journal of Remote Sensing* 1161 (February): 1243–1256. http://dx.doi.org/10.1080/01431161.2016.1148290.

Simms, Daniel M, Toby W Waine, John C Taylor, and Graham R Juniper. 2014. "The application of time-series MODIS NDVI profiles for the acquisition of crop information across Afghanistan." *International Journal of Remote Sensing* 35 (August 2014): 6234–6254.

Simonyan, Karen, and Andrew Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." In *International Conference on Learning Representations*, .

Stoian, Andrei, Vincent Poulain, Jordi Inglada, Victor Poughon, and Dawa Derksen. 2019. "Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems." *Remote Sensing* 11 (17): 1986.

Taylor, J.C., T.W. Waine, G.R. Juniper, D.M. Simms, and T.R. Brewer. 2010. "Survey and monitoring of opium poppy and wheat in Afghanistan: 2003–2009." *Remote Sensing Letters* 1 (3): 179–185. https://doi.org/10.1080/01431161003713028.

Xia, Gui Song, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. 2017. "AID: A benchmark data set for performance evaluation of aerial scene classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (7): 3965–3981.

Yang, Yi, and Shawn Newsam. 2010. "Bag-of-visual-words and spatial extensions for land-use classification." *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems* (January 2010): 270–279.

Zafeiriou, Stefanos, Dimitrios Kollias, Mihalis A. Nicolaou, Athanasios Papaioannou, Guoying Zhao, and Irene Kotsia. 2017. "Aff-Wild: Valence and Arousal 'In-the-Wild' Challenge." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2017-July (July): 1980–1987.

Zhu, Xiao Xiang, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. 2017. "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources." *IEEE Geoscience and Remote Sensing Magazine* 5 (4): 8–36. https://doi.org/10.1109/mgrs.2017.2762307.

Fully convolutional neural nets in-the-wild

Simms, Daniel M.

2020-10-20

Simms DM. (2020) Fully convolutional neural nets in-the-wild. Remote Sensing Letters, Volume 11, Issue 12, 2020, pp.1080-1089
https://doi.org/10.1080/2150704X.2020.1821120