

Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods

Rose Rustowicz, Robin Cheong, Lijing Wang, Stefano Ermon, Marshall Burke, David Lobell
Sustainability and Artificial Intelligence Lab, Stanford University

roserust@alumni.stanford.edu, {robinc20, lijing52, ermon, mburke, dlobell}@stanford.edu

Abstract

Automatic, accurate crop type maps can provide unprecedented information for understanding food systems, especially in developing countries where ground surveys are infrequent. However, little work has applied existing methods to these data scarce environments, which also have unique challenges of irregularly shaped fields, frequent cloud coverage, small plots, and a severe lack of training data. To address this gap in the literature, we provide the first crop type semantic segmentation dataset of small holder farms, specifically in Ghana and South Sudan. We are also the first to utilize high resolution, high frequency satellite data in segmenting small holder farms. Despite the challenges, we achieve an average F1 score and overall accuracy of 57.3 and 60.9% in Ghana and 69.7 and 85.3% in South Sudan. Additionally, our approach outperforms the state-of-the-art method in a data-rich setting of Germany by over 8 points in F1 and 6 points in accuracy. Code and a link to the dataset are publicly available at <https://github.com/roserustowicz/crop-type-mapping>.

1. Introduction

According to the UN, approximately 815 million people around the world are undernourished [4]. In particular, some countries in Sub-Saharan Africa suffer extreme food insecurity and malnutrition. Within some districts in Ghana, for example, 33-40% of people face chronic malnutrition [25], 74-82% of children suffer from anaemia [25], and economically, Ghana loses 6.4% of its GDP to child under-nutrition [1]. Furthermore, Ghana's employment is heavily dependent on agriculture, where 90% of households in northern Ghana depend on agricultural livelihoods [25]. South Sudan faces similar problems – after years of civil conflict, nearly 60% of the population or a record 6.1 million people needed food assistance in the lean season between July and August, 2018 [32].

Although agriculture plays a critical role in these regions, current food systems are poorly understood. Ground

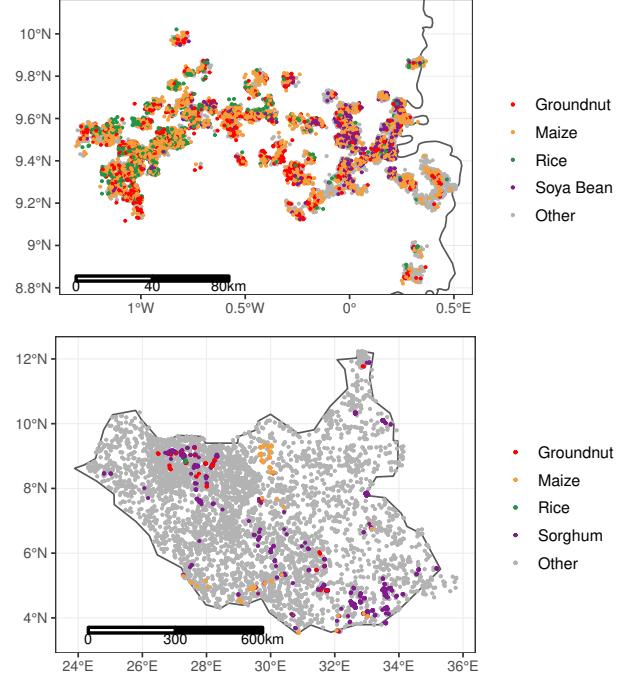


Figure 1. Distribution of crop type in the Ghana (top) and South Sudan (bottom) datasets. Up to 3km of random jitter have been added to locations for privacy. We use a subset of the datasets which includes all fields shown in color. We disregard non-crop labels and non-top crops (both in gray), but release the full datasets to enable further exploration of other cover types

surveys are conducted to collect information from farmers on the types of crops they are growing and with what yields, but are infrequent and expensive to acquire. At the same time, there has been a surge in satellite imagery collection. From 2014 to 2016, the Copernicus program deployed Sentinel-1 (S1) and Sentinel-2 (S2) satellites with six to twelve day revisit rates and ten meter spatial resolution. Meanwhile, companies such as Planet Labs and Digital Globe capture several terabytes of earth imagery every

day with higher spatial resolution. Combined with developments in computer vision, there is an unprecedented opportunity to understand issues in food security through satellite imagery. Accurate crop type segmentation could help in understanding how farmers decide what crops to grow, provide insight into interactions of crop types with environmental factors, give information on crop diversity and nutrition outcomes, and facilitate crop monitoring and yield estimation.

In this work, we apply deep-learning based semantic segmentation models to remotely sensed data in order to map crop type from space. Specifically, given a temporal sequence of satellite imagery over an agricultural area, we classify each pixel as one of several different crop types. As described above, we explore crop type classification in Ghana and South Sudan, where this problem is particularly relevant [25, 1, 4]. Automatic crop type classification in developing countries has unique challenges in that small-holder farms such as those in Africa tend to have smaller fields and sparser ground truth labels as compared to larger studies conducted in places such as the United States and Europe. Smaller fields give less pixels of information, while sparse labels introduce missing data gaps. Additionally, the growing season in our study area is dominated by rain and cloud cover, leading to low visibility in optical imagery.

In this paper, we make the following contributions:

- We release a novel dataset for crop type segmentation of small holder farms in Africa.
- We develop an approach achieving state-of-the art performance on a large crop type dataset in Germany (a data rich regime).
- To the best of our knowledge, we give the first experimental evaluation of deep learning segmentation methods for crop type mapping of small holder farms.
- We demonstrate our system achieves an average F1 score and overall accuracy of 57.3 and 60.9% in Ghana and 69.7 and 85.3% in South Sudan.

2. Related Work

We explore supervised machine learning techniques for land cover classification of agricultural crop types. Land cover classification algorithms are used to predict labels of surface types, where each pixel is given a class label. Class labels may include cover types such as forest, urban, water, agriculture, etc. Historically, these methods have used information from one satellite image to predict land cover types, while recent work also incorporates temporal observations [14]. Inputs include spectral features collected by the satellite, which often extend beyond the red, green, and blue features in a typical color image. Other bands based on texture or vegetation indices may be constructed, and predictions are usually made on a pixel-by-pixel basis without the use of contextual information.

| Region | Fields | Cover Types | Source |
|-----------------|--------|-------------|--------|
| S. Sudan subset | 837 | 4 | |
| S. Sudan all | 5,604 | 38 | [36] |
| Ghana subset | 4,439 | 4 | |
| Ghana all | 8,937 | 24 | [2] |
| Germany | 137k | 17 | [27] |

Table 1. Dataset Statistics

Cropland classification studies often use many temporal observations as input, since the spectral properties of crops change throughout a growing season [28, 23, 12]. In addition, a combination of both optical and radar data has often lead to improved results for land cover and crop type classification [15, 34, 16].

In the general computer vision community, we relate our task with semantic segmentation for video, as well as action recognition in which several temporal observations are used to make predictions. Some of the first work in exploring CNNs for video classification began with [19], in which video frames were concatenated and input into a single stream model. This work was extended to use two data streams to model temporal features by pre-computing optical flow vectors [30], further extended for longer time-range modeling [11, 35] and with learnable flow vectors [39]. [22] and [9] instead build upon work from RNNs and use a CNN + LSTM network to incorporate spatial and temporal information. A similar work [33] uses a CNN and convolutional RNN to predict pixel level labels of objects vs background for semantic segmentation in video. 3D convolutions are applied to video volumes in [31], and attention was incorporated with a 3D CNN network in [37]. [7] also incorporate 3D models into two stream networks, further extended in [8].

In recent years, deep learning methods have also gained popularity for crop classification [17]. Studies have used 1D CNN [6] [24] [38] and 2D CNN architectures [20] [18], RNNs [26] [21], convolutional RNNs [27], and 3D CNNs [10]. Additionally, most datasets used in crop classifications fall into two extremes – small, limited datasets that may be easy to overfit on a local region but fail to generalize across a wider area, and large datasets often set in the United States or Europe with large field sizes and dense labels. Among the limited works that study crop type classification in Africa, we note that incorporating both radar and optical information often improves performance, but the often small size of the dataset makes generalization questionable [13].

3. Dataset and Features

Locations and labels: Our dataset is made up of sparse ground truth labels of crop fields in South Sudan and northern Ghana. Ground truth labels consist of geo-referenced

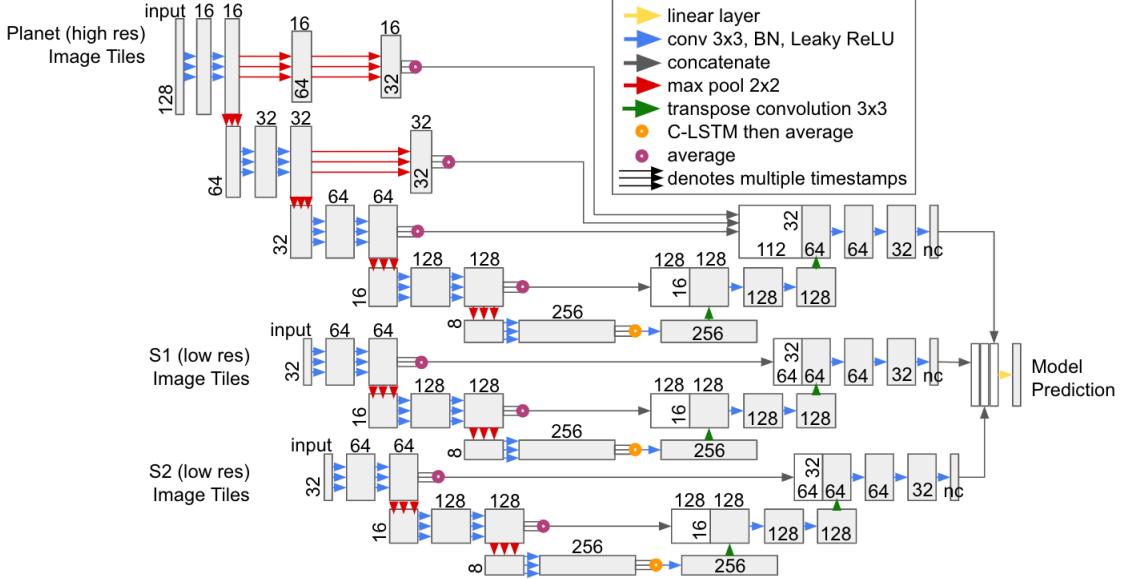


Figure 2. The 2D U-Net + CLSTM model architecture used in this study; “nc” denotes number of output classes

polygons, where each polygon represents an agricultural field boundary with a crop type label. Plots of the provided polygon locations for both countries overlaid on maps are shown in Figure 1. Locations on this visual map have been randomly jittered by up to 3km in order to preserve anonymity of field locations.

The Ghana and South Sudan subset datasets in Table 1 are used in this study. Both datasets use the top four crop types in the respective regions, which make up more than 90% of the available crop data. In Ghana, we focus on Maize (51%), Groundnut (15%), Rice (14%), and Soya Bean (10%). In South Sudan, we focus on Sorghum (67%), Maize (10%), Rice (9%), and Groundnut (7%). Note that we release the full datasets in addition to only the fields and cover types used in our subset, which may provide the chance to further expand on our study. The full South Sudan dataset also includes a majority of non-crop classes which may be further leveraged for a crop vs. non-crop classification task.

Input features: We create inputs to the model by mapping S1, S2, and Planet satellite imagery to the labelled locations within our dataset. Both Sentinel satellites have a 10m spatial resolution and a temporal revisit rate of 6-12 days. Planet’s PlanetScope imagery has 3m spatial resolution and a temporal revisit rate of 1-2 days. With high cloud cover and small field sizes, we believe incorporating Planet imagery will be beneficial. The number of satellite observations for each scene varies from 25 to 100+ observations, as we filter out all scenes with less than 25 observations. Ground truth data was collected in 2016 for Ghana and in 2017 for South Sudan. As an exception, we use Planet imagery from 2017 rather than 2016 in Ghana due to higher

data availability, with the assumption that most fields do not change crop from year to year. Table 1 provides further dataset statistics as well as attribution to the sources of ground truth labels.

We normalize all input bands to zero mean and unit variance based on statistics from the training set. We use random data augmentation of both rotation and flips. As input features, we use ten S2 bands (blue, green, red, near infrared (NIR), four red edge bands, and two short wave infrared (SWIR) bands), both S1 bands (vertical-vertical (VV) and vertical-horizontal (VH) polarizations), and all four Planet bands (blue, green, red, NIR). We also include day of year as an input band, and construct additional bands commonly used in remote sensing. For example, for Planet and S2, we explore normalized difference vegetation index (NDVI) and green chlorophyll vegetation index (GCVI) vegetation indices. For S1, we use a ratio of VH/VV as an additional input.

Splits: We subdivide our area of interest into 32 x 32 pixel grids in Ghana and South Sudan, and use the provided 48 x 48 pixel grids in Germany. We split according to a 80 / 10 / 10 split for train, validation, and test. Our splitting algorithm attempts to best preserve the relative percentages of all crops, allowing for consistent class balances in all splits.

4. Methods

In this work, we explore two approaches to crop type segmentation: a 2D U-Net + CLSTM approach, and a 3D CNN. We chose these two architectures as both are still being explored in this domain and differences in behavior between the two models has yet to be understood, especially

| Crop Type | | 2D U-Net + | | |
|--------------------|-------------|-------------|-------------|------|
| | RF | CLSTM | 3D U-Net | [27] |
| Germany | | | | |
| Sugar Beet | 79.6 | 94.7 | 95.5 | 85.3 |
| Summer Oat | 40.7 | 86.5 | 84.1 | 75.8 |
| Meadow | 83.2 | 90.7 | 91.5 | 88.2 |
| Rapeseed | 95.9 | 98.2 | 97.7 | 92.6 |
| Hop | 74.9 | 95.2 | 92.9 | 91.7 |
| Winter Spelt | 7.7 | 77.4 | 76.8 | 65.6 |
| Winter Triticale | 18.0 | 78.3 | 72.8 | 61.8 |
| Beans | 75.2 | 94.4 | 93.5 | 89.6 |
| Peas | 34.2 | 93.4 | 87.2 | 80.4 |
| Potato | 85.9 | 94.8 | 95.5 | 89.5 |
| Soybeans | 40.7 | 91.1 | 93.5 | 87.7 |
| Asparagus | 65.9 | 89.5 | 86.9 | 83.7 |
| Winter Wheat | 87.8 | 97.1 | 96.4 | 90.3 |
| Winter Barley | 87.0 | 97.0 | 96.0 | 91.0 |
| Winter Rye | 16.9 | 81.7 | 79.0 | 60.7 |
| Summer Barley | 74.3 | 95.3 | 93.9 | 85.4 |
| Maize | 92.5 | 98.1 | 97.8 | 93.9 |
| Macro Avg F1 | 62.4 | 91.4 | 90.0 | 83.1 |
| Overall Accuracy | 86.2 | 95.8 | 95.2 | 89.7 |
| Kappa Coefficient | .821 | .947 | .940 | .870 |
| Ghana | | | | |
| Groundnut | 8.5 | 51.2 | 36.3 | – |
| Maize | 73.1 | 59.5 | 67.8 | – |
| Rice | 60.6 | 78.0 | 70.0 | – |
| Soy bean | 19.6 | 48.1 | 46.1 | – |
| Macro Avg F1 | 40.5 | 57.3 | 55.1 | – |
| Overall Accuracy | 61.1 | 59.9 | 60.9 | – |
| Kappa Coefficient | .234 | .396 | .373 | – |
| South Sudan | | | | |
| Sorghum | 93.5 | 89.4 | 91.2 | – |
| Maize | 71.0 | 59.6 | 53.0 | – |
| Rice | 100 | 100 | 100 | – |
| Groundnut | 2.1 | 24.4 | 34.4 | – |
| Macro Avg F1 | 66.6 | 68.4 | 69.7 | – |
| Overall Accuracy | 88.7 | 82.6 | 85.3 | – |
| Kappa Coefficient | .572 | .454 | .493 | – |

Table 2. Comparison of test results from the models which performed the best from our ablation studies, given by the bold “Val F1” scores in Tables 4 and 5. We show per class F1 score and overall metrics for all regions.

on the same task and dataset [27, 3]. As a baseline, we compare with Random Forest (RF), which is commonly used within the land cover classification community.

4.1. 2D U-Net + CLSTM

In previous work, [27] used temporal satellite observations as input to a C-LSTM [29] to predict crop type. We extend this work by incorporating an encoder network before the C-LSTM and use these extracted features as input rather than the satellite observations themselves. Figure 2 shows our model architecture. Both the encoder and decoder net-

works are trained to share weights between all input temporal observations per satellite source thus serving as a general image feature extractor. Each satellite source uses its own encoder-CLSTM-decoder network, and predictions are aggregated at the output with a final linear layer. All convolutional layers use 3x3 kernels. The model is trained end-to-end with a weighted cross entropy loss function, where weights are chosen as a function of class balance. To handle sparse labels, we only calculate loss on image regions that have valid labels. We mask all unknown pixels and set the loss at these locations to zero.

4.2. 3D U-Net

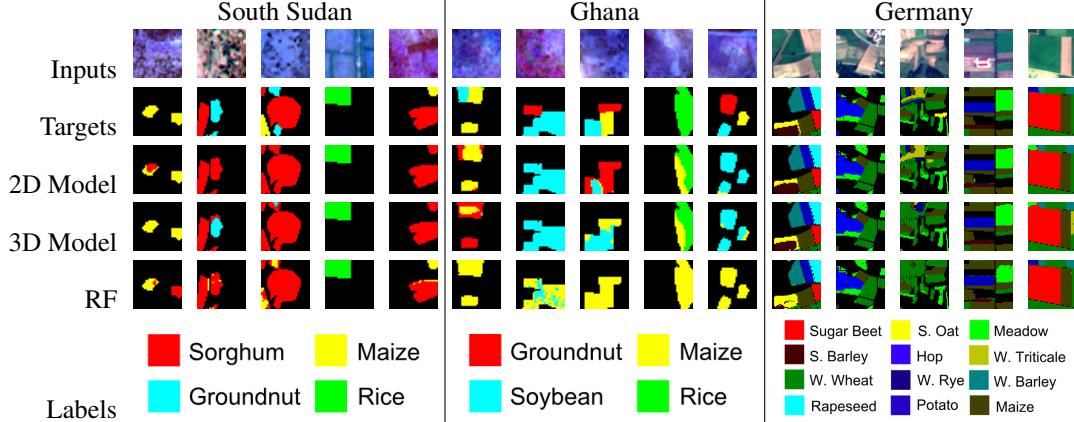
The 3D U-Net used in this study is relatively simple in design to provide a fair comparison against the 2D U-Net + CLSTM model. We define a 3D convolutional block as a 3D convolution with 3x3 kernel followed by batch norm and a leaky ReLU activation. We have five of these blocks form the first half of the U-Net architecture with spatial down-sampling occurring after the 2nd and 4th block. To form the upsampling portion of the U-Net, we also use five convolutional blocks. However, instead of downsampling spatially every two blocks, we upsample using a transpose convolution. Following the traditional U-Net architecture, we concatenate the features from the encoder branch. The model framework is nearly identical to one of the low resolution modules in Figure 2, except that the input is processed as a sequence, all layers are the 3D analogs, and there is no CLSTM or averaging in the encoder, since the temporal inputs are processed as a whole rather than in parallel.

5. Experiments

5.1. Experimental Details

We do a hyperparameter search across optimization method (Adam vs. SGD), learning rate, weight decay, number of timestamps, and using loss weight per class. For the 2D U-Net + CLSTM model, we use Adam with a learning rate of 0.003, weight decay of 0, and weighted cross entropy loss. In addition, due to GPU memory constraints, we take at most 40 samples from all the observations over a given grid in a year. We find these settings produce reasonable results across all countries.

For the 3D U-Net, we use a learning rate of 0.0003, weight decay of 0, and weighted loss. For the baseline Random Forest model, we group observations from a satellite into bi-weekly buckets and take the average of each bucket for S1 imagery, the median for cloud-filtered planet imagery, and the minimum for unfiltered S2 imagery. These aggregated observations, as well as a day of year band, are then used as input.



| | 2D U-Net + CLSTM Experiments | | | | | | | | | 3D U-Net Experiments | | | | | |
|------------------|------------------------------|------|-------------|------|------|------|------|------|-------------|----------------------|------|------|------|-------------|-------------|
| C-LSTM | x | | | | | | | | | x | x | x | x | x | x |
| 2D CNN + CLSTM | | x | x | x | x | x | x | x | x | | | | | | |
| 3D U-Net | | | | | | | | | | x | x | x | x | x | x |
| Aggregate | | | | | | | | | | x | x | x | x | x | x |
| Veg Indices | | | x | x | x | x | x | x | x | | x | x | x | x | x |
| Day of Year | | | | x | | | | | | | x | x | x | x | x |
| Cloud Band | | | | | x | x | | | | | x | | | | |
| Sample w/ Clouds | | | | | | x | | | | | x | | | | |
| Use S1 | | | | | | | x | x | x | | | | | x | |
| Low Res Planet | | | | | | | | x | | | | | | | x |
| High Res Planet | | | | | | | | | x | | | | | | |
| Best Epoch | 24 | 106 | 61 | 63 | 56 | 24 | 72 | 52 | 104 | 90 | 18 | 73 | 102 | 127 | 124 |
| Val F1 | 56.7 | 57.9 | 58.4 | 57.0 | 53.3 | 57.7 | 54.4 | 50.2 | 55.8 | 57.1 | 50.3 | 57.3 | 57.4 | 58.9 | 57.5 |
| Val Accuracy | 58.4 | 60.6 | 60.7 | 59.6 | 53.3 | 62.3 | 57.4 | 61.2 | 65.7 | 61.9 | 54.4 | 62.9 | 61.0 | 62.8 | 63.5 |

Table 4. Ablation Experiments in Ghana

5.2. Evaluation Metrics

Since our datasets have a strong class imbalance, reported accuracy results may be biased toward dominating crops. To account for this and to give an equal treatment to classification importance across all relevant classes, we compute the F1 score for each class, and then average across all classes to give the reported macro average F1 score. To compare against previous works, we report per class accuracy as well as overall accuracy.

5.3. Results

Table 2 gives a quantitative overview of our model results in Ghana, South Sudan, and Germany.

5.3.1 Germany

Notably, we achieve state-of-the-art results on the Germany dataset [27], outperforming the previous work by 8.3 F1 points and 6.7 accuracy points. We note, however, our Ger-

many results are reported on a custom dataset split that preserves class balance, and was based on available 2016 data provided from [27]. Reported results for [27] were taken directly from their paper.

5.3.2 Small Holder Farms in Africa

Unsurprisingly, models trained on the data-rich Germany dataset outperformed models in South Sudan and Ghana. These smallholder datasets have smaller training sets, high cloud cover, and complex landscapes in the smallholder setting. Yet as shown in the second half of Table 2, our models performed reasonably well in both Ghana and South Sudan, with the following notable trends:

- The 2D CNN + CLSTM model generally outperforms the 3D U-Net model.
- In Ghana the 2D CNN + CLSTM and 3D U-Net significantly outperform random forest in terms of F1 score and achieve comparable accuracy.

| | 2D U-Net + CLSTM Experiments | | | | | | | | | 3D U-Net Experiments | | | | | |
|-----------------|------------------------------|------|------|------|------|------|------|------|------|----------------------|------|------|------|-------------|-------------|
| Use S2 | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| C-LSTM | x | | | | | | | | | | | | | | |
| 2D CNN + CLSTM | | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| 3D U-Net | | | | | | | | | | x | x | x | x | x | x |
| Aggregate | | | | | | | | | | | | | | | |
| Veg Indices | | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Day of Year | | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Cloud Band | | | x | x | | | | | | | x | | x | x | x |
| Sample w/ Cloud | | | x | x | | | | | | | x | | x | x | x |
| Use S1 | | | | | x | | | | | | | x | | x | x |
| Low Res Planet | | | | | | x | x | x | x | | | | x | x | x |
| High Res Planet | | | | | | | | | x | | | | | | |
| Best Epoch | 43 | 116 | 100 | 83 | 57 | 138 | 127 | 123 | 83 | 126 | 121 | 88 | 102 | 118 | 61 |
| Val F1 | 81.2 | 74.7 | 76.8 | 74.9 | 59.8 | 75.4 | 76.2 | 75.6 | 76.7 | 75.4 | 72.3 | 77.4 | 75.5 | 77.1 | 79.8 |
| Val Accuracy | 88.7 | 82.7 | 86.4 | 86.3 | 69.8 | 85.0 | 88.5 | 86.8 | 88.5 | 85.1 | 89.3 | 89.5 | 86.9 | 90.0 | 88.3 |

Table 5. Ablation Experiments in South Sudan

- We are able to achieve significantly higher performance in South Sudan than in Ghana, likely because Ghana is a much cloudier region.
- Aside from the most prevalent crop, rice tends to be the easiest to distinguish and groundnut the hardest. We attribute this to the temporal spectral features of these crops. Upon visual inspection, rice appears to differentiate itself the most, while groundnut the least.
- Random Forest performs quite well in South Sudan, achieving high performance on all crops except groundnut. We refer back to Figure 1 and note that crops are rather localized within the South Sudan dataset, especially rice and maize. We hypothesize that the limited number of fields gives similar spectral and temporal characteristics within crop type, making classification significantly easier.

5.4. Ablation Studies

In addition, we provide ablation studies in Tables 4 and 5. The following results are most worth emphasizing:

- Contrary to related work, we find that including S1 does not always improve performance. S1 features are indicative of surface scattering, and it may be that crop types within small fields do not have enough differentiating signal in these bands.
- Our proposed model changes to the CLSTM structure do improve performance in Ghana though not in South Sudan. This is likely because South Sudan has much less data, and the proposed changes introduce many more parameters, causing the model to overfit.
- We find that including high resolution, high frequency Planet data improves accuracy for the 2D CNN + CLSTM model and generally improves the performance of the 3D U-Net model.
- We find including additional indices such as NDVI and

GCVI marginally improve performance. However, including the day of the year the observation was taken inconsistently improves performance across countries.

- We find explicitly including information about clouds does not improve maximum model performance, though in some cases, such as in Ghana, including this information does seem to improve convergence speed.

6. Conclusion

Motivated to better understand cropping systems for applications in food security and other sustainable development goals, we set out to map crop type from space. A recent surge in satellite data collection, as well as computational advances in storage and compute allow us to explore the intersection of remote sensing datasets with deep learning methods for semantic segmentation of crop type.

We compare performance between a 3D U-Net and a model that incorporates both CNNs and RNNs for semantic segmentation of multi-temporal, multi-spatial satellite images. To gain further insight into the sequence models and contributing attributes, we explore ablation studies and compare with a random forest baseline. We predict crop type with reasonable performance in Ghana and South Sudan where data is limited and of poor quality due to high cloud cover, class imbalance, and lack of labels. When applied on a large dataset in Germany, we surpass state-of-the-art performance on this task. We release the full datasets and code repository and hope to encourage the development of crop type segmentation systems for small holder farms.

Acknowledgements: Many thanks to Chris Udry and the World Food Programme for providing ground truth polygons for Ghana and South Sudan, respectively, and for allowing us to release the datasets derived from them. We thank Zhongyi Tang for help in exporting imagery and George Azzari for providing the tools [5] to do so.

References

- [1] The cost of hunger in africa. Technical report, UN Economic Commission for Africa, 2014. 1, 2
- [2] Disseminating innovative resources and technologies to smallholders (dirts) in northern region, ghana, 2016. 2
- [3] 3d convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, 10(1), 2018. 4
- [4] The sustainable development goals report 2018. Technical report, United Nations Department of Economic and Social Affairs, New York, June 2018. 1, 2
- [5] George Azzari. Gee tools. Available at https://github.com/george-azzari/gee_tools. 6
- [6] Yaping Cai, Kaiyu Guan, Jian Peng, Shaowen Wang, Christopher Seifert, Brian Wardlow, and Zhan Li. A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach. *Remote Sensing of Environment*, 210:35 – 47, 2018. 2
- [7] J Carreira and A Zisserman. Quo vaids, action recognition? a new model and the kinetics dataset. *CVPR*, 2017. 2
- [8] A Diba, M Fayyaz, V Sharma, A Karami, M Arzani, R Yousefzadeh, and L Gool. Temporal 3d convnets: New architecture and transfer learning for video classification. *arXiv*, 2017. 2
- [9] J Donahue, L A Hendricks, M Rohrbach, S Venugopalan, S Guadarrama, K Saenko, and T Darrell. Long-term recurrent convolutional networks for visual recognition and description. *CVPR*, 2015. 2
- [10] Shunping Ji et al. 3d convolutional neural networks for crop classification with multi-temporal remote sensing images. 2018. 2
- [11] C Feichtenhofer, A Pinz, and A Zisserman. Convolutional two-stream network fusion for video action recognition. *arXiv*, 2016. 2
- [12] S Foerster, K Kaden, and Trevor Darrell. Crop type mapping using spectral-temporal profiles and phenological information. *Computers and Electronics in Agriculture*, 89:30–40, 2012. 2
- [13] Gerald Forkuor, Christopher Conrad, Michael Thiel, Tobias Ullmann, and Evence Zoungrana. Integration of optical and synthetic aperture radar imagery for improving crop mapping in northwestern benin, west africa. *Remote Sensing*, 6(7):6472–6499, 2014. 2
- [14] C Gomez, White J, and M. Wulder. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116:55–72, 2016. 2
- [15] J Ingлада, A Vincent, M Arias, and C Marais-Sicre. Improved early crop type identification by joint use of high temporal resolution sar and optical image time series. *Remote Sensing*, 8(362). 2
- [16] N Joshi, M Baumann, A Ehammer, R Fensholt, K Grogan, P Hostert, M Rudbeck Jepsen, T Kuemmerle, P Meyfroidt, E Mitchard, J Reiche, C Ryan, and B Waske. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1). 2
- [17] K Kamilaris and F. X. Prenafeta-Boldu. A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, pages 1–11, 2018. 2
- [18] C Karakizi, K Karantzalos, M Vakalopoulou, and G Antoniou. Detailed land cover mapping from multitemporal landsat-8 data of different cloud cover. *Remote Sensing*, 2018. 2
- [19] A Karpathy, G Toderici, S Shetty, T Leung, R Sukthankar, and L Fei-Fei. Large-scale video classification with convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 2
- [20] N Kussul, M Lavreniuk, S Skakun, and A Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing*, 14, 2018. 2
- [21] E Ndikumana, D Minh, N Baghdadi, D Courault, and L Hosssard. Deep recurrent neural network for agriculture classification using multitemporal sar sentinel-1 for camargue, france. *Remote Sensing*, 2018. 2
- [22] J Ng, M Hausknecht, S Vijayanarasimhan, O Vinyals, R Monga, and G Toderici. Beyond short snippets: Deep networks for video classification. *arXiv*, 2015. 2
- [23] Hao P, L Wang, and Z Niu. Comparison of hybrid classifiers for crop classification using normalized difference vegetation index time series: A case study for major crops in north xinjiang, china. *PLoS ONE*, 10(9). 2
- [24] C Pelletier, G Webb, and F Petitjean. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 2019. 2
- [25] World Food Programme. Ghana, 2018. 1, 2
- [26] Marc Rußwurm and Marco Körner. Multi-temporal land cover classification with long short-term memory neural networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1/W1:551–558, 2017. 2
- [27] Marc Rußwurm and Marco Körner. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7(4), 2018. 2, 4, 5
- [28] B Schultz, M Immitzer, A Roberto Formaggio, I Del'Arco Sanches, A Jose Barreto Luiz, and C Atzberger. Self-guided segmentation and classification of multi-temporal landsat 8 images for crop type mapping in southeastern brazil. *Remote Sensing*, 7:14482–14508, 2015. 2
- [29] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *CoRR*, abs/1506.04214, 2015. 4
- [30] K Simonyan and A Zisserman. Two-stream convolutional networks for action recognition in videos. *NIPS*, 2014. 2
- [31] D Tran, L Bourdev, Fergus R, L Torresani, and M Paluri. Learning spatiotemporal features with 3d convolutional networks. *ICCV*, 2015. 2
- [32] USAID. Food assistance fact sheet - south sudan, 2018. 1

- [33] S Valipour, M Siam, M Jagersand, and N Ray. Recurrent fully convolutional networks for video segmentation. *arXiv*, 2016. [2](#)
- [34] Kristof Van Tricht, Anne Gobin, Sven Gilliams, and Isabelle Piccard. Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: A case study for belgium. *Remote Sensing*, 10(10), 2018. [2](#)
- [35] L Wang, Y Xiong, Z Wang, Y Qiao, D Lin, X Tang, and Gool L V. Temporal segment networks towards good practices for deep action recognition. *ECCV*, 2016. [2](#)
- [36] World Food Programme, Analysis and Trends Service. South sudan land cover dataset, 2017. [2](#)
- [37] L Yao, A Torabi, K Cho, N Ballas, C Pal, H Larochelle, and A Courville. Describing videos by exploiting temporal structure. *ICCV*, 2015. [2](#)
- [38] L Zhong, L Hu, and H Zhou. Deep learning based multi-temporal crop classification. *Remote Sensing*, 2018. [2](#)
- [39] Y Zhu, Z Lan, S Newsam, and A Hauptmann. Hidden two-stream convolutional networks for action recognition. *ACCV*, 2018. [2](#)