Deep Learning Based Land Cover and Crop Type Classification: A Comparative Study

Asim Hameed Khan, Muhammad Moazam Fraz and Muhammad Shahzad
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
{akhan.mscs19seecs, moazam.fraz, muhammad.shehzad} @ seecs.edu.pk

2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2) | 978-1-6654-1285-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICoDT252288.2021.9441483

***Abstract*—Remote sensing data is available free of cost with an ever-increase in the number of satellites. This satellite imagery can be used as raw input from which cultivated/non-cultivated and crop fields can be mapped. Previous trends included the use of traditional ML techniques and standard CNN, RNN for such mappings. In this paper, we investigate the segmentation models for the task of Landcover and Crop type Classification. We investigate the UNet, SegNet, and DeepLabv3+ in the data-rich states of Nebraska, Mid-West, United States. We acquire dataset from Cropland data Layer provided by USDA National Agricultural Statistics Service. Our Experimental results show that cultivated and non-cultivated landcover is classified with an accuracy of 90% and crop types are classified around 70% ensuring the models trained on one geographical area can be used for accurate classification in other geographical areas, which makes it more reliable for real-time application in agricultural business. [GitHub]**

Keywords—semantic segmentation, remote sensing, satellite imagery, convolutional neural networks, deep Learning, machine learning, landsat8, crop land data layer, classification.

I. Introduction

Automated pixel-wise classification has found its application in many fields of science like medical imaging, remote sensing, agriculture, etc. In computer vision, the task of pixel-wise classification (also called semantic segmentation), being tackled by the convolutional neural networks (CNNs) and their advanced modifications to achieve great benchmarks results. The use of the advanced architecture of CNNs as compared to conventional machine learning (ML) techniques i.e., support vector machines (SVM), logistic regression, and random forest take pixels as input and perform the task of classification. In this context, we describe a comparative study for landcover and crop type classification.

Traditionally, the crop level statistics were gathered by conducting consensus or surveys by interviewing the farmers or by personally visiting the fields and acquiring. With the advances in the agricultural field, we were able to acquire the image-level labels from the database provided by USDA NASS). In today's era, there is now a huge amount of satellite imagery available from platforms like Google Earth Engine (GEE), the burden of acquiring satellite imagery became very easy and researchers are now able to access the data with little effort. The access of data has now enabled the researchers to perform research and recently many studies have been proposed using the traditional ML approaches.

In this study, we investigate state-of-the-art semantic segmentation, architectures like UNET [1], SegNet [2], and modified Deeplabv3+ [3]. These architectures have proven their worth in the medical sciences etc. and are being used to solve many real-life problems. We explore these architectures to further solve the problems of remote sensing. The study includes the comparisons of deep learning methods with direct supervision of crop type labels. We train the models on a specific geographical location and validate on other geographical location making thus making sure the models are portable for solving real-world problems.

We explore the task of segmenting crop fields using the Landsat8 satellite imagery provided by the GEE. We acquire the labels of fields provided by the cropland data layer which is annually maintained by the USDA. The labels are provided for each state and each county. We are motivated to locate each of the crops based on remote sensing imagery to further assess real-life problems like crop monitoring, yield estimation, and future agricultural planning for a study area.

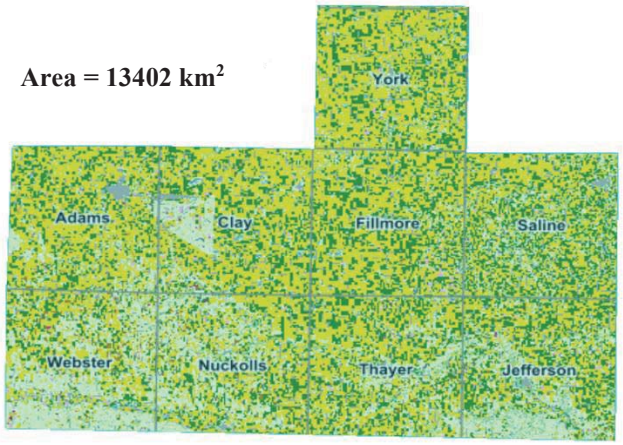

Fig. 1. Crop type labels of selected Nebraska's Counties

Our Key contributions can be summarized as follows:

1. Preparation of remote sensing dataset based on cultivated/non-cultivated land maps and crop type labels provided by cropland layer (CDL) that is annually maintained by USDA.
2. We compare three supervised deep learning architectures to classify land pixels and map crop fields using UNet, SegNet, and modified Deeplabv3+. We demonstrate how CNN outperforms traditional ML approaches with direct supervision of crop type labels with direct supervision from CDL.
3. We demonstrate how our study can be further used to locate the chosen crop classes in other geographics parts where crop type labels are available.

II. Related Work

In this section, we explore the studies that were built on traditional ML algorithms like SVM, logistic regression, and random forests methods and further progress towards studies that are based on deep learning algorithms conducted on satellite imagery and data collection methods like surveys and

978-1-6654-1285-8/20/$31.00 ©2021 IEEE

other data sources like cropland data layer for the task of the land map and crop type classification.

A. Land Cover Classification

Land cover classification can help to benefit from data that can be provided as input for modeling. Different studies have been proposed for accurate assessment of land use. The classification of land maps and usage was conducted using a non-parametric rule using the kappa coefficient [4]. Landsat imagery provides multi-spectral thermal images for both single and time-series data using random forest and k-Nearest Neighbors using the classification criteria from coordination of information on Environment which concludes that near-infrared band can alone be sufficient for accurately mapping of land and its combination with thermal imagery can be used to improve the accuracies [5]. The combination of Generative algorithm with Support vector machines [6] for feature selection and classification of land map respectively.

CNNs have played a major role in recent history for the task of image classification hence have been studied for remote sensing data collections [7]. With the advancement of satellite imagery, high spatial resolution data has been brought into research along with data disturbances caused by environmental conditions. The methods [8],[9],[10] relied on deep learning for exploring situational information in various type of land maps using transferability (using ImageNet pre-trained models) of convolutional neural networks (CNN) by first training on an annotated dataset and then used prediction with high confidence on the unlabeled dataset as a pseudo-supervised approach. Another study exploring the lack of precise labels using a Deep convolutional network [11] using fine-tuning and data augmentation specifically for remote sensing images. Using the sentinel-2 imagery a patch-based approach based on the dataset generated using 13 spectral bands for ten classes providing benchmarks using CNN [12].

The availability of satellite imagery for geographical areas through time span of days i.e. Landsat8 provides images after every 8 days, thus time series data has been made available since the last couple of decades. Recurrent Neural Networks (RNN) [13],[14] and Long Short Term Memory (LSTM) [14],[15] has been known for a contribution towards time-series data and many studies have been conducted for their application in remote sensing applications.

The method [16] investigates replacing the use of a single image as input to a CNN with the time series input for an RNN and LSTM and utilizes pixel-based and object-based data for the task of classification. The phenological changes are always expected for a certain area over time. Phenological indicators [17] were used for Normalized vegetation index (NDVI) and Enhanced vegetation index (EVI) to precisely estimate land cover changes over time. Bi-directional LSTM [18] for the land cover change using longitude and latitude along with nighttime data were used as features for time series classification.

B. Crop Type Classification

Land cover classification seems to be an easy task as compared to identifying crops in any geographical area because a particular cover of a class for land use would have more pixels in an area. The task of crop type classification would become more complex as a crop field would occupy less acreage as compared to land cover. These crop mapping had success with machine learning algorithms like SVM [19], Random forests [20], and neural networks.

Temporal data from multiple satellite sources to analyze the spectral difference in different vegetation indices in crop growth to identify crop types using SVM, neural networks, and random forests [21]. Another study trained random forest for each of the six bio-variable bands from Landsat8 for different seasons of the year for land cover and crop type classification [22]. Summer crops were classified using the EVI time series using 1-D CNN and LSTM [23].

Cropland percentages have been estimated using global distribution maps using grid cells at regional, provincial, and country levels [24],[25],[26]. Mapping of crop type at regional and country-wide from MODIS, Landsat8, and Sentinel at moderate resolution includes regional and country-wide cropland mapping of United states [27], Africa and Europe [28], Canada [29], India [30], and Brazil [31]. Very High-resolution cropland mapping with the ability to capture field-level labels but is challenging due to its computational cost [32].

USDA's cropland data layer (CDL) is free of cost dataset available for researchers and many studies have been conducted on crop type classification. The method investigates the use of random forest over the geographic changes and time in the Midwest US [33]. For temporal data observations from Landsat and CDL labels have been used for temporal classification using RNN and LSTM [34]. Transfer learning can be made useful for crop classification in the areas where the same type of crop is growing. The approach was tested on Alberta (Canada), Heng Shui (China), and Nebraska (USA) [35]. Corn and Soybean Data layer maps corn and soya crop using Landsat8 in Midwest US [36]. A method for early prediction of crops to be produced using CDL with satisfactory accuracy [37].

III. Dataset

In this section, we emphasize how we select a dataset and how we sample it for the task of semantic segmentation.

A. Study Area

To thoroughly test our supervised learning approach to classifying crop type labels, we evaluate our methods, with data-rich Nebraska state of the United States. The counties of Nebraska we chose are following Adams, Clay, Filmore, Nuckolls, Saline, Webster, Thayer, Jefferson, and York (an illustration of the study area can be seen in Figure 1). The satellite images for a total of eight counties of Nebraska state are from the year 2015 to 2019. We opted for these counties because this geographic area contains a wide variety of crop production and the variety of production is reasonable enough to analyze the generalizability of our supervised learning approach.

We approach our classification paradigm in two ways: (1) Cultivated/non-cultivated landcover classification (2) Crop type classification out of cultivated areas. Much of the chosen area produces corn and soybean which together make up more than 80% of the cultivated area and the overlapping crop region can be a good approach for studying the generalizability of the semantic segmentation models. The crop categories are the following that we consider classifying: (1) Corn (2) Soybean (3) Winter Wheat (4)

Alfalfa Hay (5) Others. The detailed map of counties of Nebraska can be seen in Figure 1.

B. Remote Sensing Data

The series of Earth satellite imagery is observed by the Landsat cooperatively managed by USGS and NASA being observed since 1972. We obtained the remote sensing imagery from Landsat8 which provides images of modest resolution of 30m once with time span of two weeks. The categories of the data from each satellite are Tier 1, Tier 2, and Real-Time (RT). We chose the Tier 1 surface reflectance imagery which is atmospherically corrected using OLI/TIRS sensors from Landsat 8. It consists of 7 seven reflectance bands i.e. Ultra-blue, blue, green, red, near-infrared, and shortwave 1 and 2 [38].

We obtained the polygons i.e. shapefiles for each county from the CropScape platform provided by NASS [27]. We used the GEE to acquire the imagery for our study region provided by Landsat 8 Tier 1 surface reflectance. The time-period we chose for our dataset was from 2015-2019. As per our knowledge, no latest studies were proposed with these years. So, the motivation is to study the remote area that can be used to further assess the crop yield and monitoring process for future use. The windows span we chose for each year is from January 1 to December 31. We acquire all the images for each year using this windows span. Although all the surface reflectance bands can be used to study the area, but we chose the red (B4), green (B3), and blue (B2) bands of Landsat 8 and obtained an RGB dataset so that minimal information can be used to generate desired maps.

To construct a single image per year for each county, we created a composite image by taking the median of each pixel in Image Collection for every band using the median function provided by the GEE. Thus, the median Image contains all the information of the time series pixel in a single image.

There is a high probability because of atmospheric conditions that may affect the study region. Those atmospheric conditions could be different types of contaminations like shadows, cloud covers, and snowfall. Thus, to overcome these issues, we used the Quality Assessment band (QA) provided with Landsat 8 before computing the median of the images. Thus, the final median images are cloud and shadows-free. These images are available at 30m resolution, and we downloaded each image at a spatial resolution of 1700x1700 images. As we cannot feed a CNN, the huge size of the image so, we divided each image into a tile 256x256. This will be later covered in detail in the methodology section.

C. Pixel Level Labels

The Cropland data layer (CDL) is a geo-referenced raster layer composed by USDA for the entire United States continent [27]. Just like Landsat 8, it is offered at 30m resolution which could be used to map on spectral bands retrieved from Landsat 8 imagery. Overall, CDL contains 255 classes out of which we chose major crops as mentioned Study area section. Thus, constructing a pixel-level label for the corresponding satellite image region. Pixel-wise crop distribution can be seen in Figure 2.

Since our end task is to generate segmented labels from CNN model on image-level labels, we prepare image-level labels using 256x256 patches with 50% overlap across them. We computed a binary label i.e. {0,1} for cultivated and non-cultivated land cover maps and for mapping crop types, categorical labels i.e. {0,1,2,3,4} based on CDL ground truth.

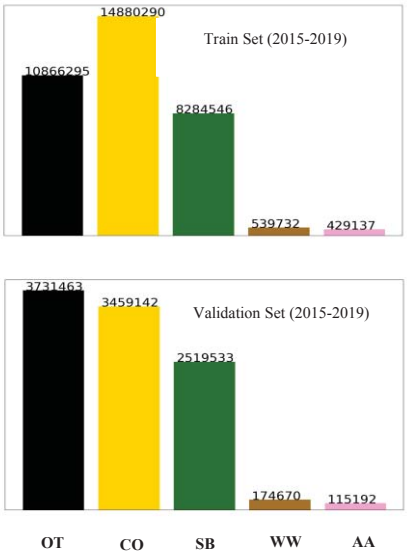

Fig. 2. Pixel-wise crop distribution in the dataset (CO as Corn, SB as SoyBean, WW as Winter Wheat, AA as Alfalfa Hay, and OT as Others)

IV. Methodology

In this section, we explain the approaches used to train segmentation models for (1) landcover classification i.e. cultivated and non-cultivated lands (2) out of cultivated lands we identify crop type with segmentation models.

A. Data Augmentation

During networks training, we employed augmentation techniques to avoid our networks from overfitting, making sure that segmentation labels and images are invariant to our techniques. To make our dataset more effective, we rotated images at an angle of 90 degrees, flipped left, right, up, and down. We do not employ any other techniques because that would cause disruption in the geographical area i.e., we do not need to use brightness variation because Landsat8 imagery has already taken care of it. During the experiments, it was observed that it affected the generalizability of the models.

B. Implementation Details

The deep learning framework we chose for training is Keras 2.4 (TensorFlow v1 Backend). The training is performed on a single GPU for 30 epochs for cultivated/non-cultivated classification pipeline and for crop type mapping we ran our networks for 50 epochs. We have initialized the learning rate at 10^{-2} with exponential decay with $\boldsymbol{k}$ as a hyperparameter for stepwise decay. All the networks are given an input image of size 256x256x3 in batches of 8.

C. Network Architectures

We tested three major types of segmentation models to classify land cover and crop types. The first model in this study is UNet [1], inspired from the original paper, containing two paths: (1) contraction (Encoder). An Input image (HxWxC) is encoded into a low dimensional feature vector contained high semantic information. In Encoder, the size of the input image reduces from 256x256x3 into 16x16x256. The other half of the network (2) expansion path (decoder),

upsamples the representation into original input form with SoftMax probabilities of classified pixels. In the decoder, 16x16x256 is gradually upsampled to 256x256x1 for landcover and 256x256x5 for crop type mapping. We employ ReLU activation, Batch Normalization, and Dropout of 0.05 after each convolutional layer. SegNet [2] architectures are majorly different from UNet where pooling indices (caches maximum features value for each encoder level) are transferred instead of entire feature maps.

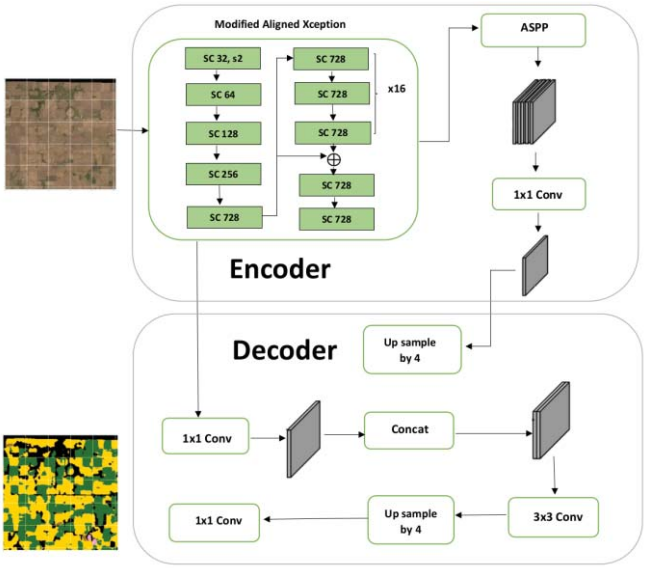

Fig. 3. Deeplabv3+ architecture [3].

Much focused in architecture for our study is Deeplabv3+ [3]. In its implementation we used Xception [39] for feature extraction. The architecture as shown in original paper uses Deeplabv3 as encoder [40]. In the Encoder, all maximum pooling operations are replaced with Depthwise Separable convolution (SC), comprising of depth-wise and point-wise convolutional followed by Batch Normalization and ReLU activation, useful to reduce the computational complexities by factorizing operations into multiple steps. This improved Xception is the main body of our encoder differentiating it from the originally proposed DeepLabv3. This aftereffect into the extraction of important feature information of remote sensing imagery for accurate segmentation of crops, thus preserving low-level semantics being invariant to the geometric transformation of land cover. In the later part of Encoder Atrous Spatial Pyramid Pooling (ASPP) is applied with different atrous rates, combining multi-level features. SC in ASPP component captures global context by using Global average pooling, thus, allowing the fields and crops of varying acreage and outline are segmented more accurately. Furthermore, the aggregation of multi-level features provides a more diversified representation of remote sensing imagery.

It is crucial to integrate the low-level semantics for an accurate depiction of cropland boundaries in remote sensing imagery. We up-sample the low-level information in the Decoder block, steadily by the factor of 4, whereas as it can be seen in Figure 3, the low-level features are concatenated with the initial layers of the encoder. 1x1 convolution is applied to reduce the number of channels before concatenation. The reason for this concatenation is that the initial layer has low-level information which is very useful for accurate pixel-wise segmentation. Furthermore, 3x3 convolution followed by up-sampling by a factor of 4 helps refine the feature map. A succeeding 1x1 convolution squeezes the resulting feature map into the dimensions of the input feature map, thus, finally generating the mask.

D. Loss Function

To estimate the generalizability of UNet, SegNet, and Deeplabv3+ we used two loss functions as per the class distribution of two datasets. For the task of Landcover classification, the loss function for classifying cultivated and non-cultivated land cover we used Binary cross-entropy, and for crop type classification categorical cross-entropy is used.

V. Results

Sample visual results of land cover classification and crop type classification can be seen in Figure 4 where the original image of RGB channels generated from GEE Landsat8 spectral bands i.e., B4, B3, and B2 as input, ground truth from CDL, and predicted mask from selected networks are shown.

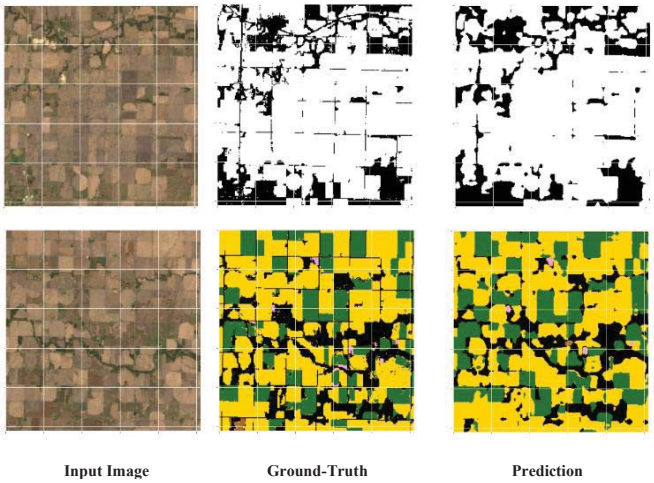

Fig. 4. Visual Results generated by Deeplabv3+. The above Images show cultivated (black) and non-cultivated (black) for land maps while the image below show results for crop type classification (corn as yellow, soybean as green, alfalfa as pink, wheat as orange, and other as black).

It can be observed that segmentation networks had a hard time predicting the boundaries of the fields but accurate in the sense that it can precisely estimate crop type location on geographical areas.

The quantitative results of UNet, SegNet, and DeepLabv3+ can be seen in Table 1 for Land cover maps and Table 2 for crop type maps. The performance measures we use for the evaluation of the models are Accuracy and Dice Similarity Co-efficient (DSC). Accuracy is the percentage of pixels that are correctly classified in Landsat8 imagery while the DSC is a measure of an overlap between two images i.e. predicted and target masks. The comparable studies graph of training and validation losses for segmentation networks can be seen in Figure 5. From the evaluation results, we state that studied segmentation models can be used in remote sensing applications that identify cultivated/non-cultivated landcover and crop types in the United States. Our method is more generalizable because there is no data leakage between any geographical area much unlike studies based on random splits between train and validation sets [41].

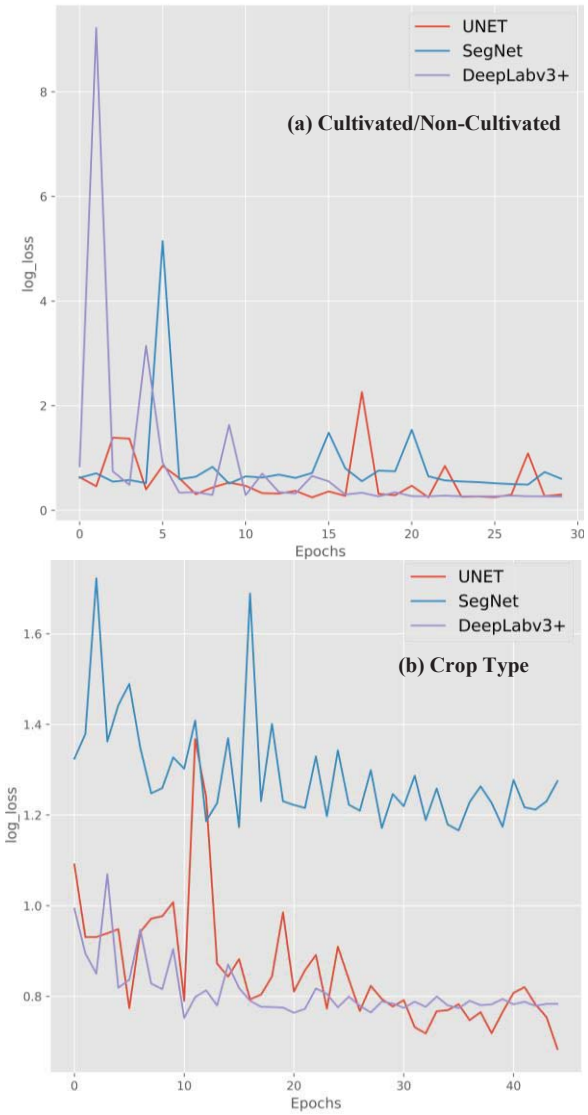

Fig. 5. A comparative analysis of Validation loss vs the number of epochs for UNet, SegNet, and DeepLabv3+.

TABLE I. PERFORMANCE MEASURES

Sr. No.	Cultivated/Non-Cultivated Land maps		
	Algorithm	*Accuracy*	*DSC*
1.	UNet	89.5	89.2
2.	SegNet	74.69	73.43
3.	DeepLabv3+	89.13	88.6

TABLE II. PERFORMANCE MEASURES

Sr. No.	Crop type Maps		
	Algorithm	*Accuracy*	*DSC*
1.	UNet	67.3	57.1
2.	SegNet	49.5	37.0
3.	DeepLabv3+	69.7	62.02

The purpose is to evaluate the model which is more effective in segmenting cultivated/non-cultivated and crop map given an RGB Landsat8 image. Figure 5a illustrates the log loss for a total of 30 epochs for cultivated and non-cultivated landcover for fully supervised UNet, SegNet, and DeepLabv3+. It can be observed that for all the models, the validation loss remains high in early epochs but gradually decreases with time. SegNet validation loss has multiple fluctuations as compared to others but it can be observed around 15 epochs that DeepLabv3+ has acquired maximum generalizability. In comparison, UNet requires more time to train. For the task of distinguishing crop types from cultivated landcover, the log loss is illustrated in Figure 5b. SegNet performance is relatively poor, loss throughout the training remains high. DeepLabv3+ outperforms UNet by a noticeable margin in terms of accuracy and DSC. Similar to landcover classification, DeepLabv3+ reaches the lowest around 10 epochs which gradually decreases in further epochs. Thus, it can be inferred from the qualitative and quantitative measures DeepLabv3+ shows better generalizability both in training time and achieving maximum performance measures for classifying cultivated land and distinguishing specified crop types out of that particular landcover.

VI. CONCLUSION

In this paper, we showed that state-of-the-art models for end-to-end segmentation, can accurately segment the cultivated/non-cultivated land maps and out of the cultivated area, the mapping of crop type was done in the state of Nebraska. We make sure that there is no data leakage among training and validation samples which makes the study more reliable for solving real-world problems. We showed that DeepLabv3+ among all the models can be most reliable in landcover and crop type classification from Landsat8 imagery with only 3 spectral bands. Since our dataset is comprised of an annual median composite of chosen geographical area, our contribution does not leverage the temporal nature of remote sensing imagery which would be further helpful in assessing the growing patterns of crops in future work.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "2015-U-Net," *arXiv*, pp. 1–8, 2015.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[3] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11211 LNCS, pp. 833–851.

[4] S. S. Rwanga and J. M. Ndambuki, "Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS," *Int. J. Geosci.*, vol. 08, no. 04, pp. 611–622, 2017.

[5] L. Sun and K. Schulz, "The Improvement of Land Cover Classification by Thermal Remote Sensing," *Remote Sens.*, vol. 7, no. 7, pp. 8368–8390, Jun. 2015.

[6] C. Sukawattanavijit, J. Chen, and H. Zhang, "GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.

[7] R. Pires de Lima and K. Marfurt, "Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis," *Remote Sens.*, vol. 12, no. 1, p. 86, Dec. 2019.

[8] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using ImageNet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, 2016.

[9] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land Use Classification in Remote Sensing Images by Convolutional Neural Networks," Aug. 2015.

[10] X. Y. Tong *et al.*, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, p. 111322, Feb. 2020.

[11] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.

[12] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.

[13] G. Petneházi, "Recurrent Neural Networks for Time Series Forecasting," Dec. 2018.

[14] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network," Aug. 2018.

[15] Y. L. Kong, Q. Huang, C. Wang, J. Chen, J. Chen, and D. He, "Long short-term memory neural networks for online disturbance detection in satellite image time series," *Remote Sens.*, vol. 10, no. 3, 2018.

[16] Di. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, Oct. 2017.

[17] Z. Xue, P. Du, and L. Feng, "Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 4, pp. 1142–1156, 2014.

[18] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

[19] J. K. Gilbertson, J. Kemp, and A. van Niekerk, "Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques," *Comput. Electron. Agric.*, vol. 134, pp. 151–159, Mar. 2017.

[20] P. Hao, Y. Zhan, L. Wang, Z. Niu, and M. Shakir, "Feature Selection of Time Series MODIS Data for Early Crop Classification Using Random Forest: A Case Study in Kansas, USA," *Remote Sens.*, vol. 7, no. 5, pp. 5347–5369, Apr. 2015.

[21] and J. P. C. Sun, Y. Bian, T. Zhou, "Using of Multi-Source and Multi-Temporal Remote," *Sensors*, vol. 19, pp. 1–23, 2019.

[22] F. Waldner *et al.*, "Land cover and crop type classification along the season based on biophysical variables retrieved from multi-sensor high-resolution time series," *Remote Sens.*, vol. 7, no. 8, pp. 10400–10424, 2015.

[23] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, no. December 2018, pp. 430–443, 2019.

[24] F. T. Portmann, S. Siebert, and P. Döll, "MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling," *Global Biogeochem. Cycles*, vol. 24, no. 1, p. n/a-n/a, Mar. 2010.

[25] L. You, S. Wood, U. Wood-Sichra, and W. Wu, "Generating global crop distribution maps: From census to grid," *Agric. Syst.*, vol. 127, pp. 53–60, May 2014.

[26] B. Wu *et al.*, "Global crop monitoring: A satellite-based hierarchical approach," *Remote Sens.*, vol. 7, no. 4, pp. 3907–3933, 2015.

[27] "USDA National Agricultural Statistics Service Cropland Data Layer." [Online]. Available: https://nassgeodata.gmu.edu/CropScape/. [Accessed: 08-Feb-2021].

[28] J. Inglada *et al.*, "Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery," *Remote Sens.*, vol. 7, no. 9, pp. 12356–12379, 2015.

[29] T. Fisette *et al.*, "AAFC annual crop inventory: Status and challenges," in *2013 2nd International Conference on Agro-Geoinformatics: Information for Sustainable Agriculture, Agro-Geoinformatics 2013*, 2013, pp. 270–274.

[30] E. Heller *et al.*, "Mapping crop types, irrigated areas, and cropping intensities in heterogeneous landscapes of southern india using multi-temporal medium-resolution imagery: Implications for assessing water use in agriculture," *Photogramm. Eng. Remote Sensing*, vol. 78, no. 8, pp. 815–827, 2012.

[31] A. S. Cohn, L. K. Vanwey, S. A. Spera, and J. F. Mustard, "Cropping frequency and area response to climate variability can exceed yield response," *Nat. Clim. Chang.*, vol. 6, no. 6, pp. 601–604, May 2016.

[32] J. L. McCarty, C. S. R. Neigh, M. L. Carroll, and M. R. Wooten, "Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery," *Remote Sens. Environ.*, vol. 202, pp. 142–151, Dec. 2017.

[33] S. Wang, G. Azzari, and D. B. Lobell, "Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques," *Remote Sens. Environ.*, vol. 222, no. January, pp. 303–317, 2019.

[34] Z. Sun, L. Di, and H. Fang, "Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series," *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 593–614, Jan. 2019.

[35] P. Hao, L. Di, C. Zhang, and L. Guo, "Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples," *Sci. Total Environ.*, vol. 733, p. 138869, Sep. 2020.

[36] S. Wang, S. Di Tommaso, J. M. Deines, and D. B. Lobell, "Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive," *Sci. Data*, vol. 7, no. 1, pp. 1–14, Dec. 2020.

[37] R. Yaramasu, V. Bandaru, and K. Pnvr, "Pre-season crop type mapping using deep neural networks," *Comput. Electron. Agric.*, vol. 176, p. 105664, Sep. 2020.

[38] "Google Earth Engine." [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_SR. [Accessed: 07-Feb-2021].

[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, pp. 1800–1807.

[40] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *arXiv*, Jun. 2017.

[41] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery," no. Section 2, pp. 1–25, 2020.