Convolutional Neural Network Approach for Mapping Arctic Vegetation using Multi-Sensor Remote Sensing

Zachary L. Langford[1], Jitendra Kumar[2,1], and Forrest M. Hoffman[2,3]

[1]Bredesen Center, University of Tennessee Knoxville; [2]Climate Change Science Institute, Oak Ridge National Laboratory; and [3]Dept. of Civil & Environmental Engineering, University of Tennessee Knoxville

November 18, 2017

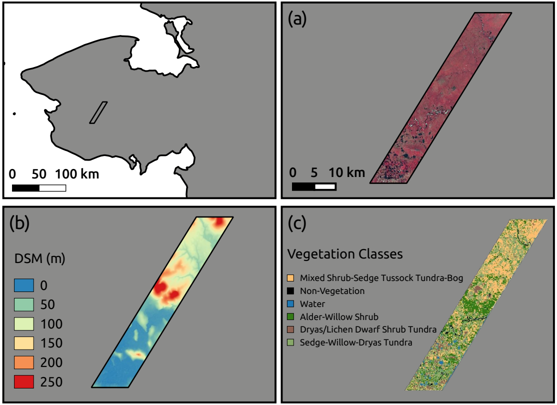
THE UNIVERSITY OF TENNESSEE KNOXVILLE

OAK RIDGE National Laboratory

CLIMATE CHANGE SCIENCE INSTITUTE
OAK RIDGE NATIONAL LABORATORY

Research Questions

- Can we provide accurate vegetation maps at high spatial resolutions (5×5 m) using public datasets and coarse/inaccurate vegetation maps?
- Does multi-sensor fusion (e.g., multi/hyper-spectral) increase performance?
- Can Convolutional Neural Networks (CNNs) provide an approach for learning vegetation characteristics based on multi-sensor datasets?

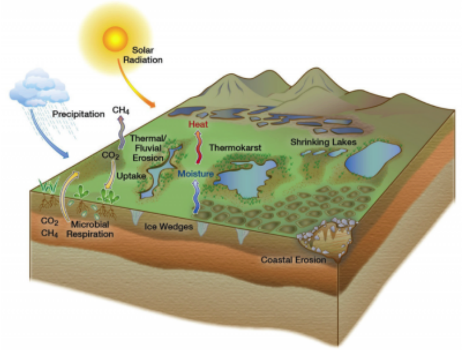


Overview

- Study area based in the Seward Peninsula, Alaska.
- (a) Bounded by the EO-1 Hyperion footprint.
- (b) Digital Surface Model (DSM) of the study region.
- (c) Vegetation dataset used to train our models, based on the Alaska Existing Vegetation Type (AKEVT).

Motivation

- Accurate and high-resolution maps of vegetation are critical for projects seeking to understand the terrestrial ecosystem processes and land-atmosphere interactions in Arctic ecosystems, such as the U.S. Department of Energy's Next Generation Ecosystem Experiment (NGEE) Arctic `https://ngee-arctic.ornl.gov/`.



- Increasing our confidence in climate projections for high-latitude regions of the world will require a coordinated set of investigations that target improved process understanding and model representation of important ecosystem-climate feedbacks.

Satellite Datasets

Dataset	Resolution	Acquisition Date	Bands
EO-1 Hyperion	30 m	24 June 2015	198
Landsat 8 OLI	30 m	17 August 2016	9
ALOS-1 PALSAR	12.5 m	29 August 2007	2
IfSAR DSM	5 m	Summer 2012	1
SPOT-5	2.5 m	Summer 2009–2012	3

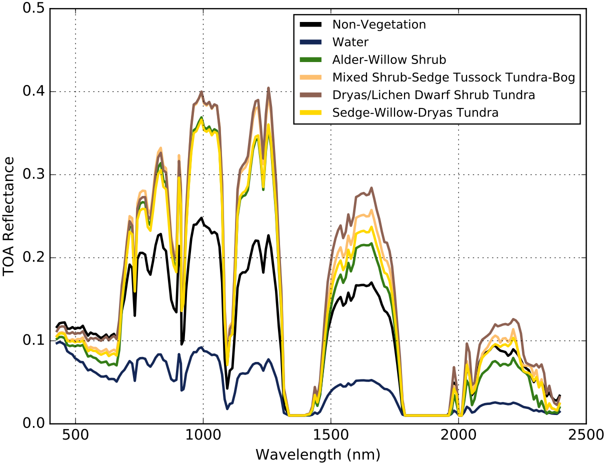
- EO-1 Hyperion hyper-spectral sensor that spans from visible to near infrared (NIR).
- Landsat 8 OLI multi-spectral sensors that spans from visible to short-wave infrared (SWIR).
- SPOT-5 multi-spectral with NIR, red, and green bands.
- ALOS-1 PALSAR's L-band synthetic aperture radar (SAR) that yields, all-weather, day-and-night observation.
- Interferometric synthetic aperture radar (IfSAR) data to generate Digital Surface Models (DSMs).

Combination of Satellite Datasets

Multi-sensor vegetation classification cases

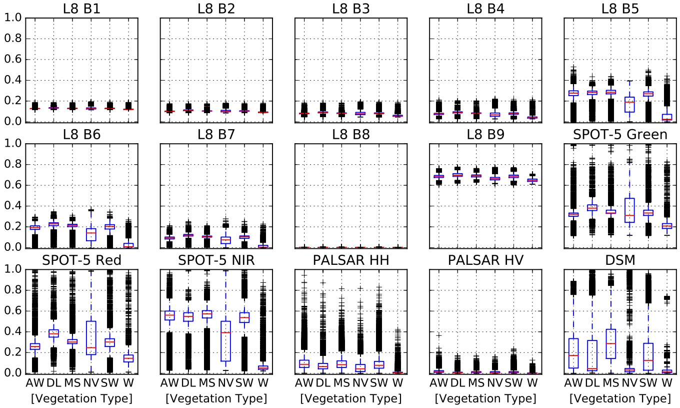
Case	Fusion Combinations of Remote Sensing Datasets	# of data layers
D1	EO-1, IfSAR	199
D2	EO-1, Landsat 8 OLI, IfSAR	208
D3	EO-1, ALOS-1 PALSAR, IfSAR	201
D4	EO-1, SPOT-5, IfSAR	202
D5	EO-1, ALOS-1 PALSAR, SPOT-5, IfSAR	204
D6	EO-1, ALOS-1 PALSAR, Landsat 8 OLI, IfSAR	210
D7	EO-1, SPOT-5, Landsat 8 OLI, IfSAR	211
D8	Landsat 8 OLI, IfSAR	10
D9	Landsat 8 OLI, ALOS-1 PALSAR, IfSAR	12
D10	Landsat 8 OLI, SPOT-5, IfSAR	13
D11	ALOS-1 PALSAR, IfSAR	3
D12	ALOS-1 PALSAR, SPOT-5, IfSAR	6
D13	SPOT-5, IfSAR	4
D14	EO-1, Landsat 8 OLI, ALOS-1 PALSAR, SPOT-5, IfSAR	213

Spectral Characteristics – EO-1 Hyperion



- Averaged spectra for the study region.
- Continuous signature of vegetation types show separability for certain wavelengths.

Spectral Characteristics – Other Datasets

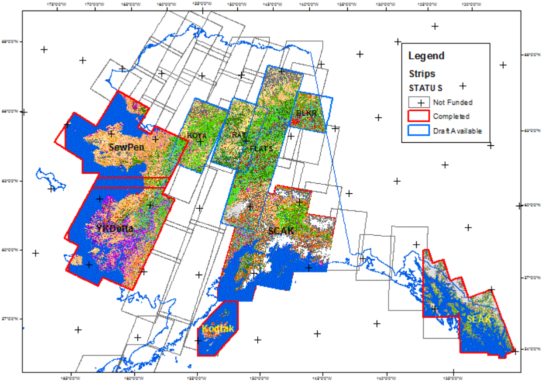


AW – Alder-Willow Shrub, DL – Dryas/Lichen Dwarf Shrub Tundra, MS – Mixed Shrub-Sedge Tussock Tundra-Bog, NV – Non-Vegetation, SW – Sedge-Willow-Dryas Tundra, W – Water

- Landsat 8 (L8) bands, B5 (Near Infrared (NIR)), B6 (Shortwave Infrared (SWIR) 1), and B7 (Shortwave Infrared (SWIR) 2) show greatest separability.
- SPOT-5 and DSM also show separability but with high ranges.
- High ranges and commonality could be to the noisy AKEVT dataset.

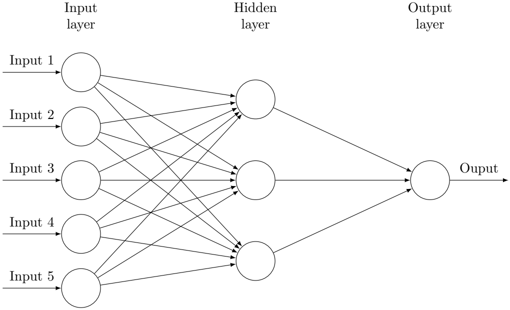
Vegetation Dataset

Class	Area (km^2)	Percentage
Mixed Shrub-Sedge Tussock Tundra-Bog	120.75	35.13%
Alder-Willow Shrub	74.33	21.63%
Dryas/Lichen Dwarf Shrub Tundra	20.52	5.97%
Sedge-Willow-Dryas Tundra	116.41	33.87%
Water	6.13	1.78%
Non-Vegetation	5.58	1.62%



- We used the Alaska Existing Vegetation Type, circa 2000 (AKEVT) represent the existing dominate vegetative species present, or the non-vegetative land cover.
- Based upon existing field plot data, Landsat 7 ETM+ spectral data, and topographic data.
- It can be difficult to accurately map large areas based on sparse field plots.

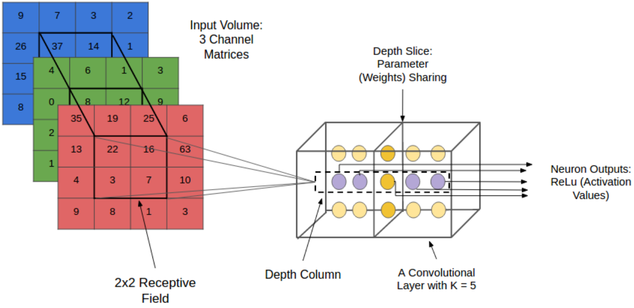
Neural Networks



- Neural networks are typically organized in layers, made up of a number of interconnected nodes connected in an acyclic graph, with the outputs of some neurons can become inputs to other neurons.
- Inspired by receptive fields in the brain.
- Deep learning refers to artificial neural networks that are composed of many layers (Goodfellow et al., 2016).

Convolutional Neural Networks

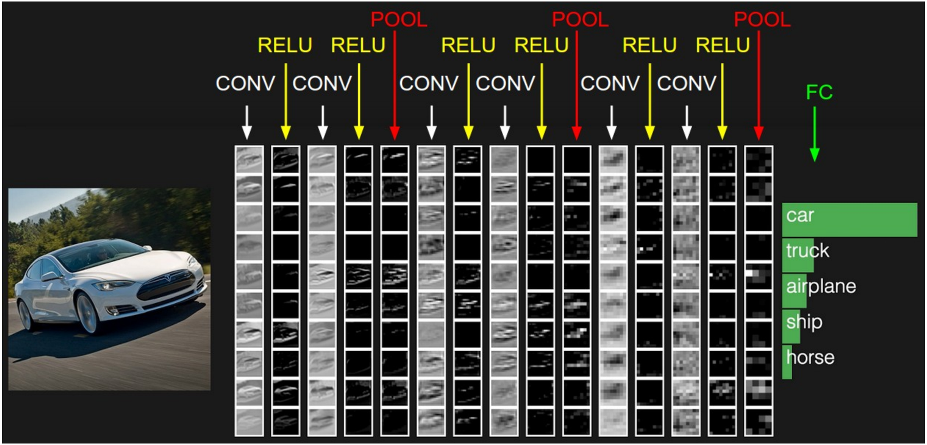
- The hidden layers of a Convolutional Neural Networks (CNNs) typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers.
- CNNs arranges its neurons in three dimensions (width, height, depth). In this example the depth would be 3 (Red, Green, Blue channels).
- The architecture of a CNN is designed to take advantage of the structure of an input image.



Source: `http://xrds.acm.org`

2D CNN Layers

Layer	Information
Convolution	Convolving a matrix with one single convolution kernel
Pooling	Smooths the data and reduces spatial resolution
Loss Layer	Objective function, required to compile the model
Activation	Defines the output of that node given an input or set of inputs (ReLU, logistic, etc.)
Fully Connected (FC) Layer	Full connections to all activations in the previous layer, as seen in regular Neural Networks

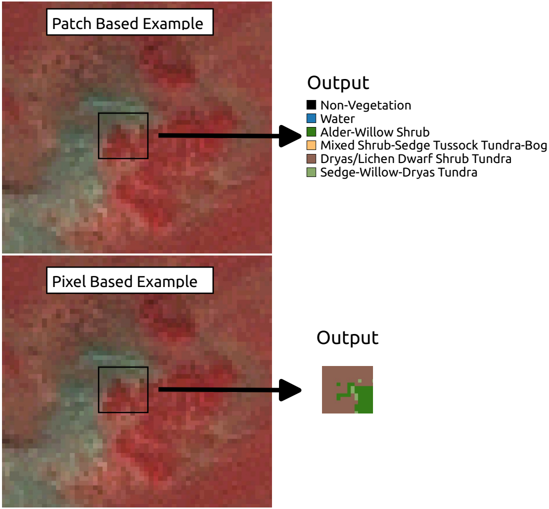


Source:`http://cs231n.github.io`

CNN Hyperparameters

- **Learning rate** – controls how much to update the weight in the optimization algorithm.
- **Number of epochs** – number of times the entire training set pass through the neural network.
- **Activation function** – introduces non-linearity to the model.
- **Number of hidden layers and units** – good to add more layers until the test error no longer improves.
- **Dropout** – preferable regularization technique to avoid overfitting.
- **Structure** – number of layers, size of filters, number of feature maps in convolution layers, etc.

CNN Architectures



- **Patch based** – break image up into patches (6×6, 12×12, and 16×16) and estimate the vegetation type for the patch.
- **Pixel based** – output image will have every pixel classified (i.e., image segmentation). We break up the images into patches (16×16, 32×32, and 64×64), for computational efficiency.

Unsupervised Clustering Vegetation Map (UCVM)

- Hargrove et al. (2006) developed a method for quantitatively comparing categorical maps that is independent of differences in resolution, independent of the number of categories in maps, and independent of the directionality of comparison.
- We first perform PCA and k-means clustering of the combined datasets from k=10, 25, and 50.
- We then reclass the k clusters image based on the Goodness of Fit (GOF) score, which is a unitless measure of spatial overlap between map categories.
- We call this method the Unsupervised Clustering Vegetation Map (UCVM).

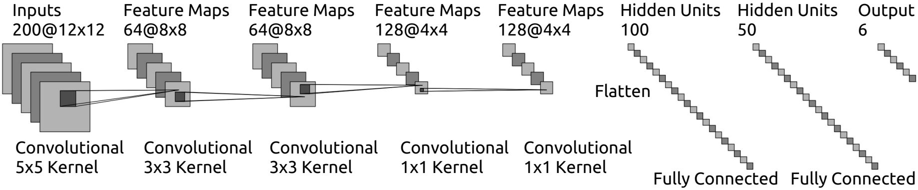
Training/Validation

- We use 90% of the patch sizes (e.g., 12×12) for training and 10% validation.
- Two training datasets for building CNN models.
  1. Alaska Existing Vegetation Type (AKEVT) map labels
  2. Unsupervised Clustering Vegetation Map (UCVM)

Example of UCVM method

(a) Clustering of D14 using k=25. (b) The AKEVT map and (c) the UCVM map, showing large differences. (d) SPOT-5 false color image.

CNN Networks – Patch Based

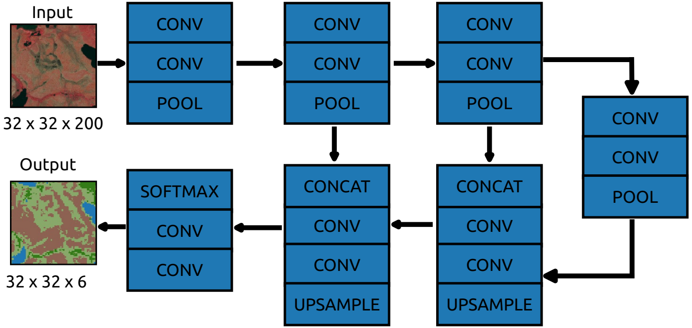


Layer	Output Shape	Param
Input	(None, 12, 12, 200)	800
conv2d	(None, 6, 6, 24)	43224
conv2d	(None, 3, 3, 36)	7812
conv2d	(None, 3, 3, 48)	15600
conv2d	(None, 3, 3, 64)	27712
flatten	(None, 576)	0
dense	(None, 100)	57700
dense	(None, 50)	5050
dense	(None, 6)	306

Total params: 158,204

- Example using D14 – 200 spectral bands.
- 4 convolutional layers.
- 2 hidden layers (100 and 50, respectively).
- Increase filter size as we downsample image.
- Output from patches will corresponds to a vegetation type.

CNN Networks – Pixel Based



- Inspired by the U-Net Algorithm (Ronneberger et al., 2015).
- Example of the pixel-based architecture from taking a 32×32 pixel patch with 200 bands (D14).
- Constant number of 64 filters throughout the network.
- Constant kernel size of 3×3.
- Pooling size of 2×2, used for downsampling images.
- Upsampling is performed with size of 2×2, which upsamples your image to a higher resolution.
- Perform filter concatenation based on previous layers.

Results – Patch Based

		AKEVT			UCVM		
		6×6	12×12	16×16	6×6	12×12	16×16
D1	5 m	0.64	0.65	0.63	0.94	0.93	0.91
	12.5 m	0.63	0.63	0.59	0.95	0.93	0.87
D2	5 m	0.66	0.66	0.66	0.94	0.92	0.91
	12.5 m	0.65	0.64	0.59	0.95	0.92	0.85
D3	5 m	0.64	0.64	0.64	0.94	0.93	0.90
	12.5 m	0.64	0.62	0.61	**0.96**	0.92	0.88
D4	5 m	0.66	0.67	0.66	0.95	0.93	0.91
	12.5 m	0.66	0.64	0.61	0.95	0.92	0.87
D5	5 m	0.67	0.67	0.66	0.95	0.93	0.91
	12.5 m	0.67	0.64	0.63	0.95	0.92	0.88
D6	5 m	0.66	0.67	0.66	0.94	0.92	0.90
	12.5 m	0.66	0.64	0.61	0.95	0.92	0.87
D7	5 m	0.67	0.68	0.67	0.95	0.93	0.91
	12.5 m	0.67	0.66	0.59	0.95	0.93	0.86
D8	5 m	0.63	0.64	0.65	0.78	0.77	0.78
	12.5 m	0.64	0.62	0.59	0.82	0.80	0.80
D9	5 m	0.63	0.64	0.63	0.78	0.77	0.77
	12.5 m	0.65	0.63	0.59	0.83	0.82	0.80
D10	5 m	0.66	0.67	0.67	0.80	0.79	0.78
	12.5 m	0.67	0.67	0.62	0.83	0.81	0.81
D11	5 m	0.52	0.53	0.53	0.67	0.66	0.67
	12.5 m	0.53	0.52	0.52	0.72	0.74	0.73
D12	5 m	0.62	0.63	0.63	0.76	0.76	0.76
	12.5 m	0.64	0.64	0.56	0.81	0.80	0.78
D13	5 m	0.62	0.63	0.64	0.76	0.75	0.76
	12.5 m	0.64	0.64	0.56	0.81	0.80	0.78
D14	5 m	0.67	0.68	0.67	**0.96**	0.93	0.91
	12.5 m	0.67	0.65	0.57	0.95	0.93	0.86

- Accuracy of patch sizes (i.e., 6×6, 12×12, and 16×16) at 5 m and 12.5 m resolution using AKEVT and UCVM maps for training.
- The CNN models trained using AKEVT had accuracies ranging from 52% to 68%.
- The hyper-spectral datasets performed the best with the AKEVT labels, ranging from 59% to 68% accurate, with patch size of 6×6 and 12×12 achieving the highest scores.
- Higher resolution pixels acheive better scores for large patch sizes (16×16 patch) for most datasets.
- UCVM vegetation map increased the accuracy for all datasets by a large margin.
- The highest scores were achieved for the hyper- spectral datasets, having 96% accuracy for D3 (EO-1, ALOS-1 PALSAR, IfSAR) for the 12.5 m dataset.
- The highest score from the 5 m datasets was achieved using D14 dataset with a 96% accuracy.

Results – Patch Based

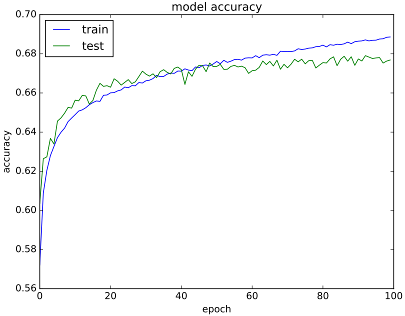


Figure: The model accuracy for D14 by varying the epoch to 100 using the **AKEVT** for training and patch size of 6×6.

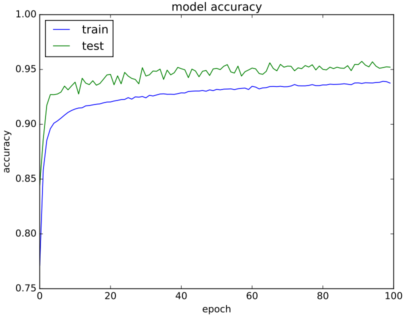


Figure: The model accuracy for D14 using the **UCVM** for training and varying the epoch to 100 for patch size of 6×6.

The green line represents the validation set (10% of data) and the blue line represents the training dataset (90% of data). AKEVT models overfit the data after a short number of epochs. UCVM models show optimal performance during training and validation.

Results – Pixel Based

		AKEVT			Mapcurves		
		16	32	64	16	32	64
D1	5 m	0.64	0.62	0.61	0.95	0.92	0.89
	12.5 m	0.61	0.59	0.61	0.94	0.90	0.86
D2	5 m	0.66	0.64	0.61	0.94	0.92	0.91
	12.5 m	0.62	0.60	0.61	0.94	0.92	0.87
D3	5 m	0.64	0.63	0.62	0.94	0.93	0.89
	12.5 m	0.62	0.60	0.60	0.94	0.90	0.86
D4	5 m	0.67	0.65	0.63	**0.97**	0.93	0.89
	12.5 m	0.64	0.60	0.62	0.94	0.89	0.86
D5	5 m	0.67	0.65	0.62	**0.97**	0.94	0.89
	12.5 m	0.64	0.62	0.62	**0.95**	0.90	0.86
D6	5 m	0.66	0.65	0.63	0.94	0.92	0.90
	12.5 m	0.64	0.60	0.63	0.94	0.91	0.87
D7	5 m	0.68	0.65	0.62	0.95	0.93	0.91
	12.5 m	0.64	0.62	0.62	0.94	0.91	0.86
D8	5 m	0.64	0.62	0.61	0.76	0.74	0.74
	12.5 m	0.61	0.60	0.56	0.80	0.79	0.80
D9	5 m	0.65	0.65	0.65	0.76	0.75	0.75
	12.5 m	0.65	0.62	0.63	0.82	0.79	0.80
D10	5 m	0.68	0.66	0.66	0.78	0.76	0.76
	12.5 m	0.65	0.64	0.65	0.83	0.80	0.82
D11	5 m	0.55	0.55	0.56	0.67	0.66	0.68
	12.5 m	0.55	0.54	0.55	0.74	0.71	0.74
D12	5 m	0.66	0.65	0.63	0.76	0.75	0.74
	12.5 m	0.64	0.63	0.63	0.81	0.79	0.81
D13	5 m	0.65	0.65	0.64	0.76	0.73	0.75
	12.5 m	0.65	0.62	0.63	0.81	0.78	0.81
D14	5 m	0.68	0.66	0.64	0.96	0.93	0.91
	12.5 m	0.64	0.62	0.63	0.94	0.90	0.87

- AKEVT labels had accuracies ranging from 54% to 68%.
- D11 (ALOS-1 PALSAR, IfSAR) performed the worst with the AKEVT labels, similar to the patch-level CNN architecture.
- D7 (EO-1, SPOT-5, Landsat 8 OLI, IfSAR), D10 (Landsat 8 OLI, SPOT- 5, IfSAR), and D14 (EO-1, Landsat 8 OLI, ALOS-1 PALSAR, SPOT-5, IfSAR) performed the best with the AKEVT at 5 m and 16×16 patch size.
- Increasing the resolution and having a patch size of 16×16 worked best for our segmentation architecture.
- The UCVM vegetation map increased the accuracy for all datasets by a large margin.
- Highest scores were achieved from the hyper-spectral datasets, having 97% accuracy for D4 (EO-1, SPOT-5, IfSAR) and D5 (EO-1, ALOS-1 PALSAR, SPOT-5, IfSAR) for the 5 m dataset.
- Highest score from the 12.5 m datasets were the hyper-spectral datasets (D1 D7 and D14) with a 94%-95% accuracy.
- The Landsat, SAR, and SPOT datasets (D8, D9, D10, D12, and D13) had accuracies from 0.77% to 0.83%.

Results – Pixel Based

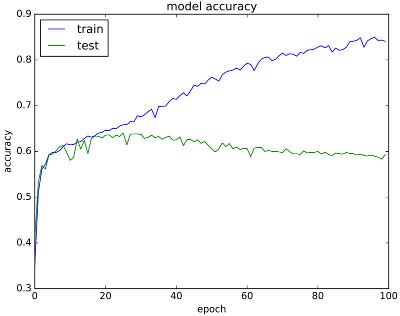


Figure: Model accuracy for D5 varying the epoch to 100 using the **AKEVT** for training and patch size of 16×16.

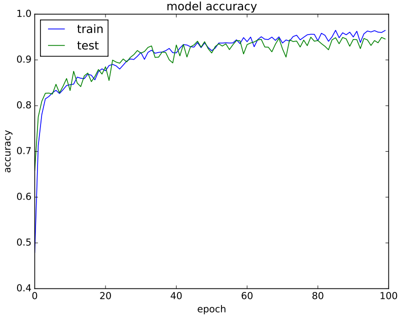
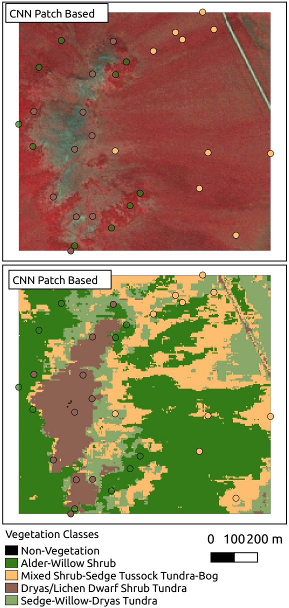


Figure: Model accuracy for D5 varying the epoch to 100 using the **UCVM** for training and patch size of 16×16.

The green line represents the validation set (10% of data) and the blue line represents the training dataset (90% of data). AKEVT models overfit the data after a short number of epochs. UCVM models show optimal performance during training and validation.

Results – Field Based Validation



Plot Number	Field Obs.	D7 CNN UCVM	D7 CNN AKEVT	AKEVT
1	MS	MS	MS	MS
2	MS	MS	AW	MS
3	AW	AW	AW	AW
4	DL	SW	DL	DL
5	AW	AW	AW	AW
6	MS	MS	MS	MS
7	MS	MS	SW	SW
8	AW	AW	AW	SW
9	DL	AW	AW	SW
10	AW	AW	AW	SW
11	AW	AW	AW	AW
12	DL	AW	MS	DL
13	DL	AW	AW	AW
14	MS	MS	MS	MS
15	MS	MS	MS	SW
16	MS	MS	MS	MS
17	MS	MS	MS	MS
18	MS	AW	AW	AW
19	MS	MS	SW	MS
20	AW	AW	AW	SW
21	AW	AW	AW	SW
22	AW	AW	AW	AW
23	DL	MS	MS	SW
24	DL	MS	MS	SW
25	DL	MS	DL	DL
26	AW	AW	AW	AW
27	AW	AW	AW	MS
28	DL	SW	DL	DL
29	DL	SW	DL	DL
30	DL	SW	DL	AW
Accuracy		0.63	0.70	0.53

Conclusions

- Both CNN model approaches (pixel and patch based) had similar accuracy metrics for all cases, with AKEVT CNN models ranging from 52% – 68% and UCVM CNN models ranging from 66% – 97%.
- Hyper-spectral datasets (D1–D7 and D14) performed the best, with an average of $\sim$95% accuracy.
- The UCVM dataset helped improve the accuracy scores compared to using the AKEVT map alone.
- D3 (EO-1, ALOS-1 PALSAR, IfSAR), D4 (EO-1, SPOT-5, IfSAR), and D5 (EO-1, ALOS-1 PALSAR, SPOT-5, IfSAR) are the optimal datasets.
- Future steps are to:
  - Improve CNN architectures that utilize hyper-spectral features.
  - Incorporate vegetation datasets into Earth system models (ESMs).

Acknowledgments

U.S. DEPARTMENT OF ENERGY

Office of Science

This research was supported by the Next-Generation Ecosystem Experiments (NGEE Arctic) and the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area (SFA), which are sponsored by the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the US Department of Energy Office of Science. Oak Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC, for the US Department of Energy under Contract No. DE-AC05-00OR22725.

References

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

W. W. Hargrove, F. M. Hoffman, and P. F. Hessburg. Mapcurves: A quantitative method for comparing categorical maps. *J. Geograph. Syst.*, 8(2):187–208, July 2006. doi:10.1007/s10109-006-0025-x.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.