Patch-Based Approach to Land-Classification of Sentinel-2 Satellite Imagery

David Smith

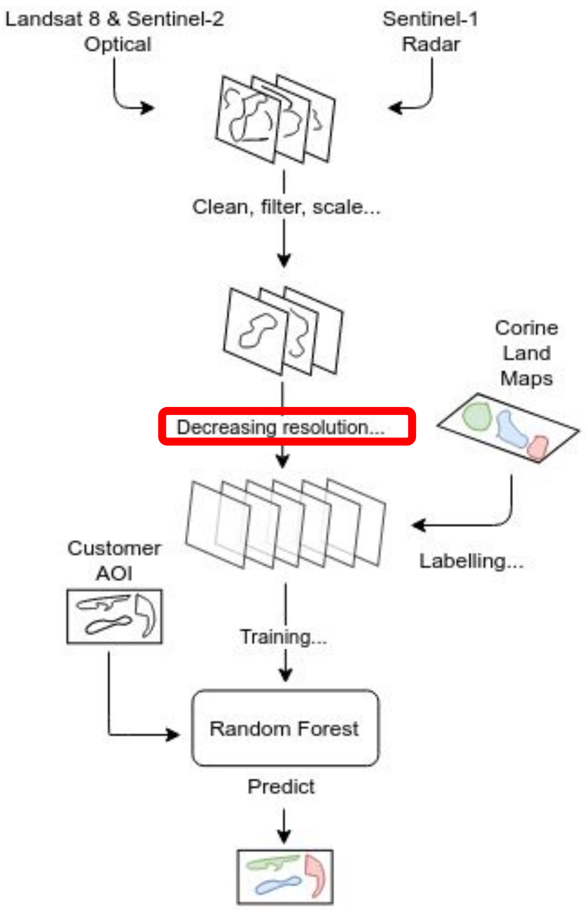
Project Outline

- Explore alternative methods to the current GSI land-classification methodology.
- Evaluate these methods over a shared area of interest.
- Assess how useful they could be for GSI.

Current Methodology

- After retrieving imagery, it is downsampled to match CORINE ground-truth.
- 10m resolution degraded to 400m.
- Gives an informative average of local area.

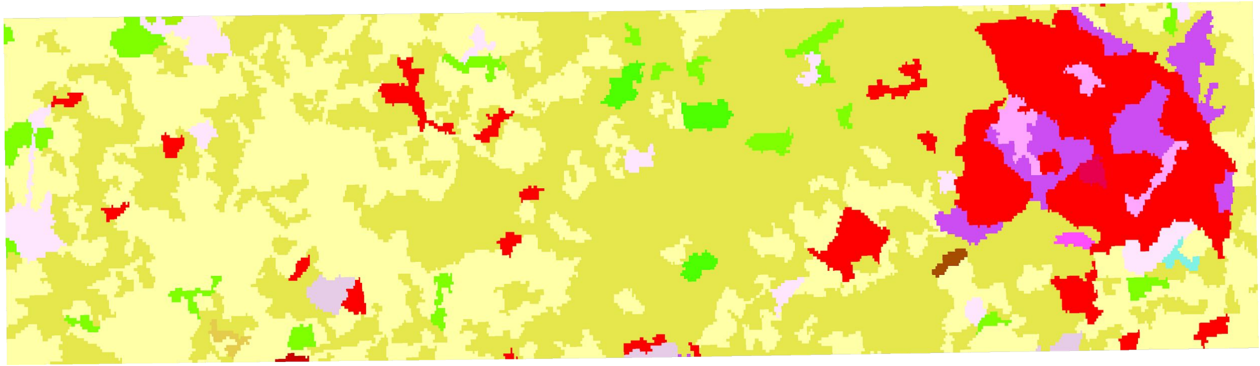
- High risk of information loss.
- 10m resolution imagery has high spatial content over 400m region.



Project Objectives

- This project was exploring alternatives to downsampling.
- Develop methods incorporating the spatial and textural information within image regions, instead of just smoothing the region with an average.
- Test a predictive model trained on the processed image.
- Evaluate the model performance vs the current methodology.

Area of Interest



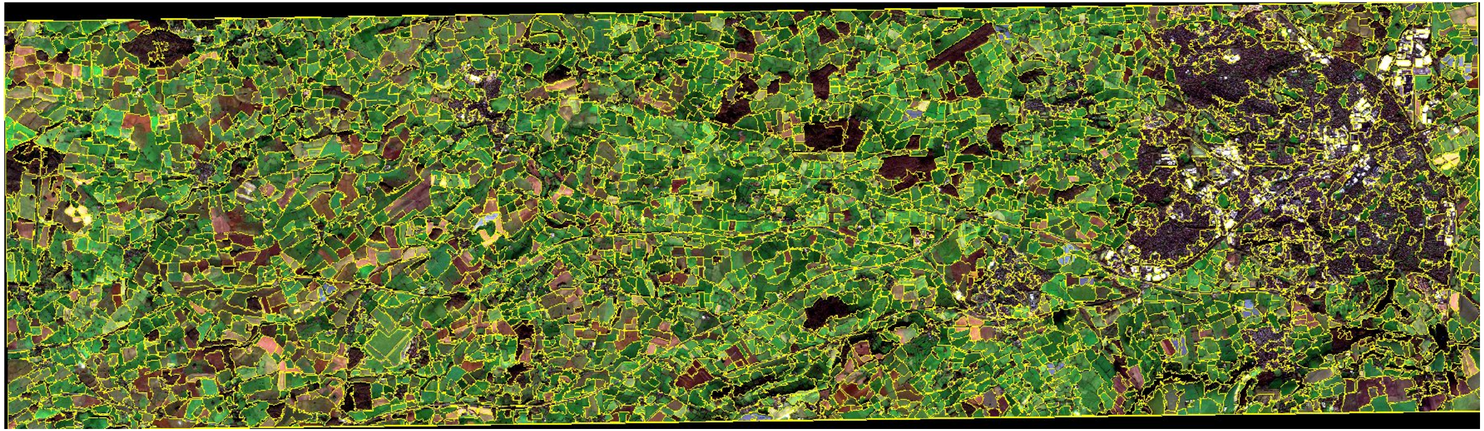
Data Considerations

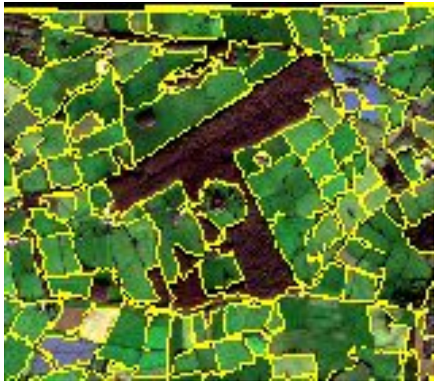
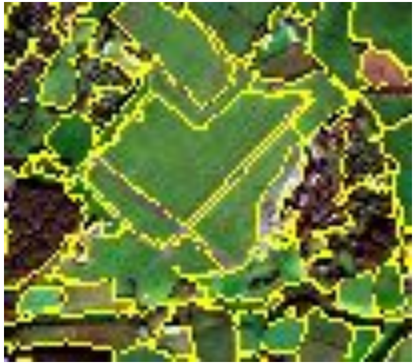
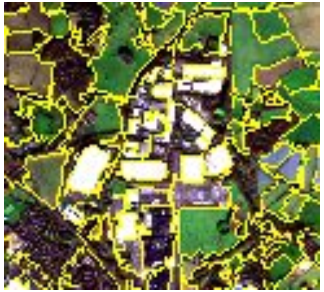
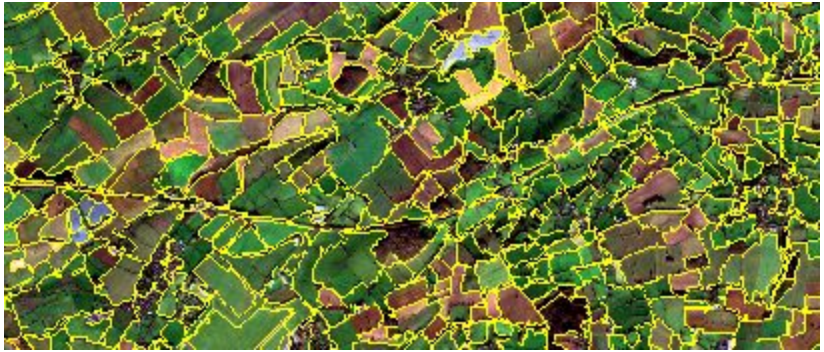
- We were only considering bands 2,3,4 and 8.
- Calculated NDVI index.
- Predicting 17 classes present in the AOI.

Methodologies

Segmentation Method

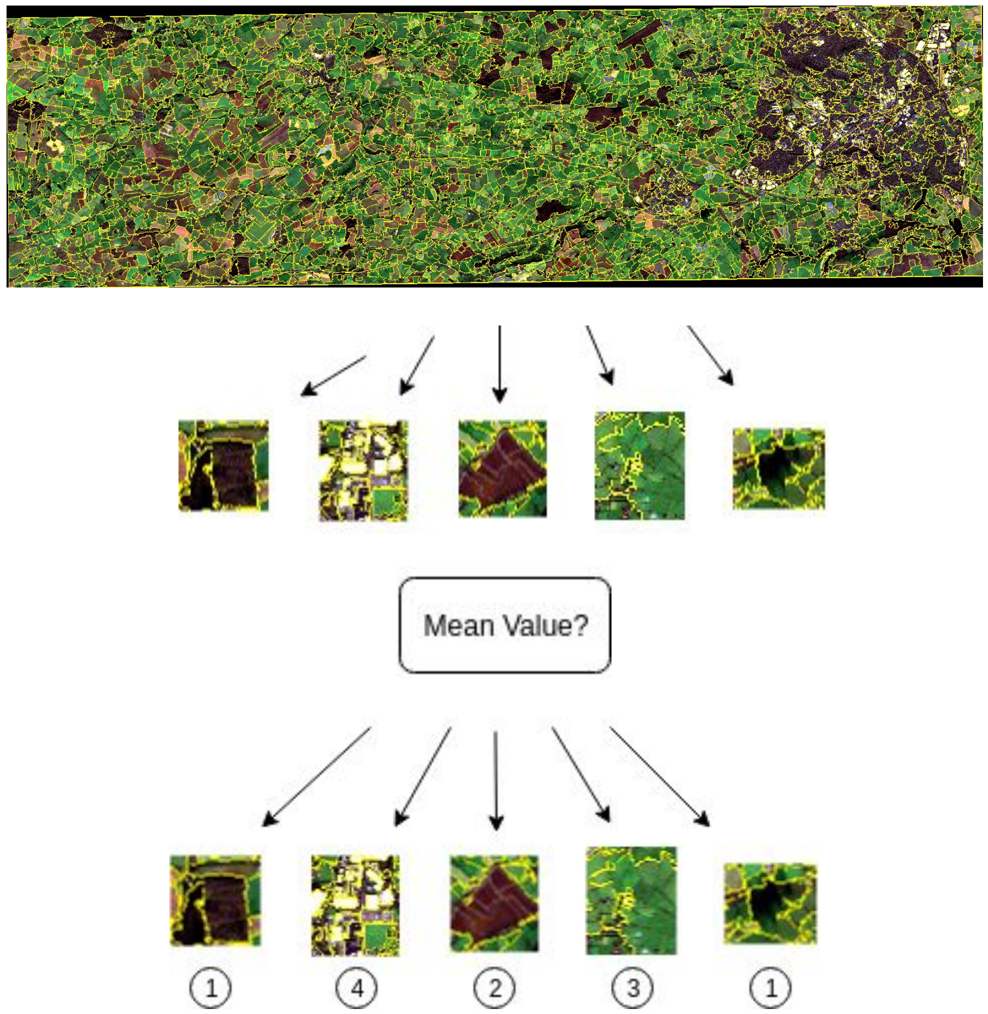
- First step to exploring the texture within the image.
- Take advantage of obvious colour similarities in the AOI.
- Graph-based segmentation.
- Plots each pixel as a connected node on a graph.
- Nodes with strong connections grouped together.



Approach

- Quantify segments based on mean value.
- Create a hierarchy based on these values.
- Apply this information to the pixels in the AOI.

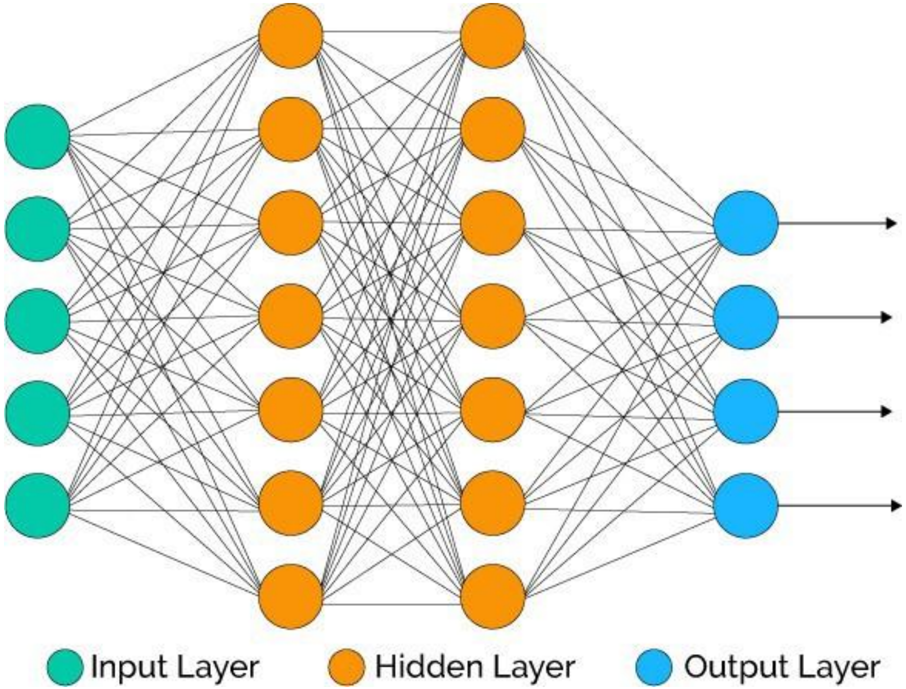

Mean Value?
1
4
2
3
1

Training

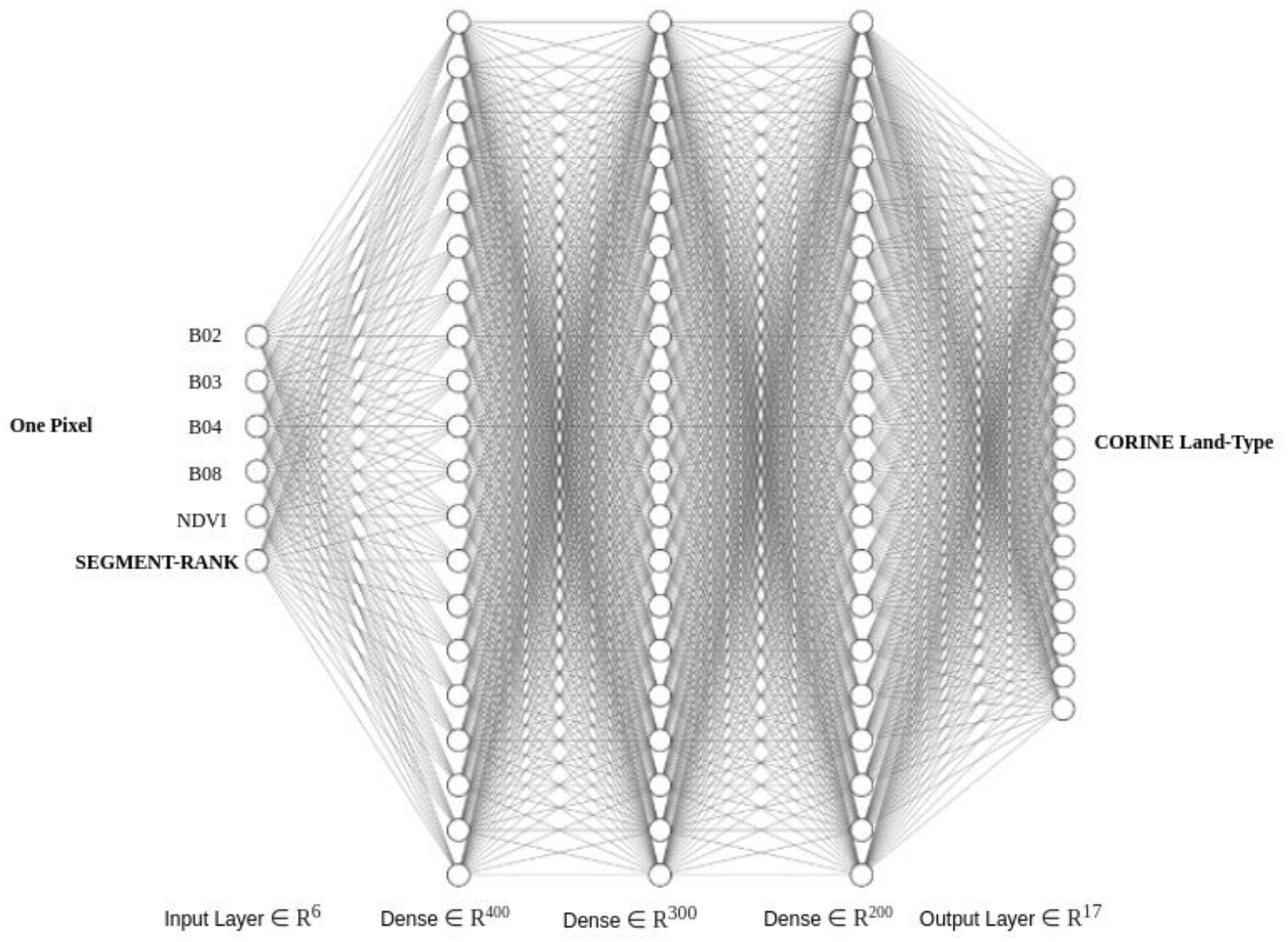
- Randomly extracted single pixels.
- For every pixel, we attached a “segment-rank” variable based on the segment that contained it.
- Training set consisted of a dataframe of pixel band values, NDVI index values, segment-rank values and CORINE labels.
- A single row represented one pixel.

Algorithm Choice

- Deep learning approach
- Neural networks
- Basic architecture:



- Multi-Layer Perceptron (MLP)



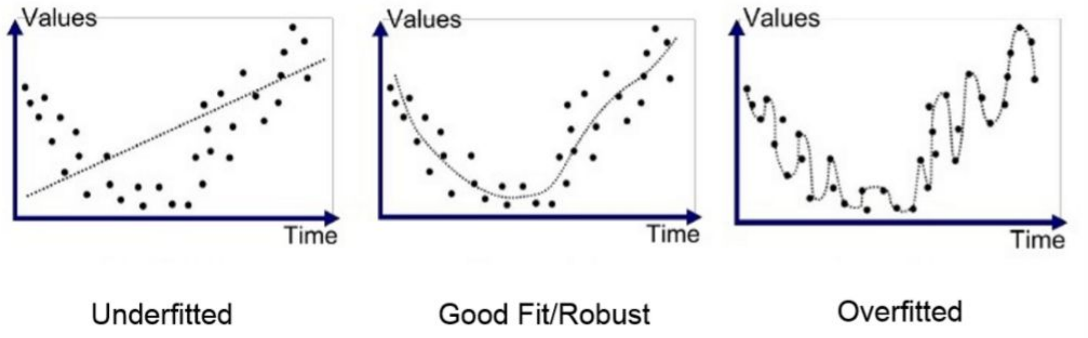
One Pixel
B02
B03
B04
B08
NDVI
SEGMENT-RANK
CORINE Land-Type
Input Layer ∈ R6
Dense ∈ R400
Dense ∈ R300
Dense ∈ R200
Output Layer ∈ R17

Baseline

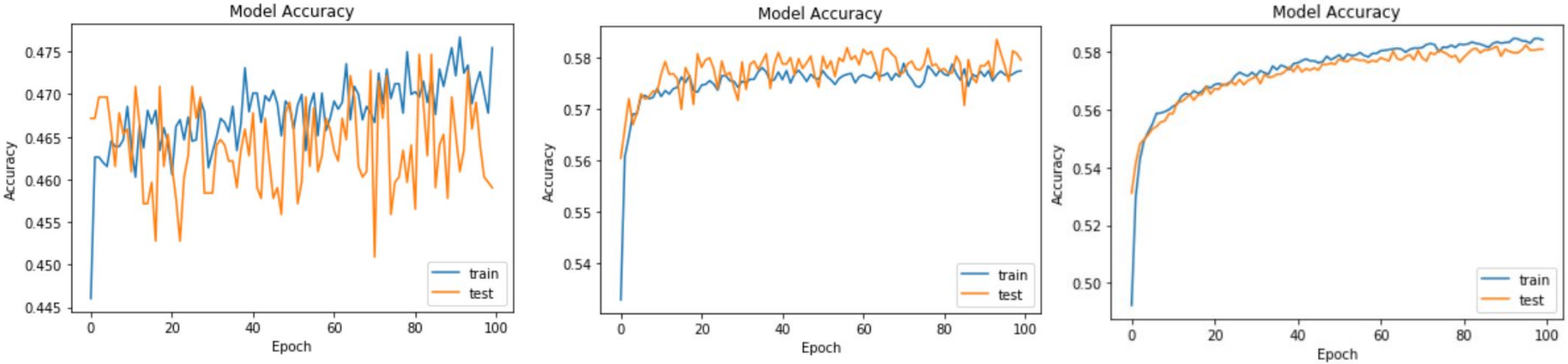
- MLP-Pixel
- Random Forest Regressor
- GSI-RF metrics from AOI test

Training and Validation

- Trained on between 10,000 and 250,000 pixels.
- Tweaked model parameters and architecture.
- Batch-size, learning rate, epochs and regularization.



- 60/20/20 training, validation and testing.


Model Accuracy
Accuracy
Epoch
train
test
Model Accuracy
Accuracy
Epoch
train
test
Model Accuracy
Accuracy
Epoch
train
test

Testing Results

Classifier	Training Set	Valid Acc (%)	Test Acc (%)
	10000	54.53	51.84
MLP	100000	58.5	56.19
	250000	59.21	58.3
	10000	55.64	51.75
MLP-Pixel	100000	56.32	55.53
	250000	57.23	56.87
	10000	24.25	21.52
Random Forest	100000	47.85	44.56
	250000	51.36	47.52

Prediction

- Wanted to recreate the testing conditions from the GSI metrics:

	Land-Type	Precision	Recall	F1-Score	Support
GSI-RF	*continuous urban*	0.05	0.92	0.09	4042
	discontinuous urban	0.42	0.52	0.46	39915
	industrial/commercial	0.64	0.94	0.76	75389
	road and rail networks	0.00	0.00	0.00	6459
	airports	0.21	1.00	0.35	51498
	green urban areas	0.00	0.00	0.00	0
	sport/leisure facilities	0.00	0.00	0.00	1540
	non-irrigated arable land	0.22	0.24	0.23	51650
	pastures	0.98	0.14	0.25	28582
	broad-leaved forest	0.39	0.45	0.42	51063
	mixed forest	0.08	0.85	0.15	78961

	Correct Pred.	Incorrect Pred.	Overall Accuracy
GSI-RF	145479	254521	37.00%

Land-Type	Precision	Recall	F1-Score	Support
continuous urban fabric	0.04	0.05	0.04	97
discontinuous urban fabric	0.63	0.40	0.49	567
industrial or commercial units	0.59	0.32	0.41	42588
road and rail networks land	0.01	0.00	0.00	10741
airports	0.28	0.02	0.04	2000
dump sites	0.00	0.00	0.00	429
construction sites	0.00	0.00	0.00	422
green urban areas	0.10	0.00	0.00	2968
sport and leisure facilities	0.02	0.00	0.00	9025
non-irrigated arable land	0.63	0.07	0.12	134621
pastures	0.73	0.75	0.74	181995
agriculture/vegetation	0.00	0.00	0.00	514
broad-leaved forest	0.47	0.27	0.34	51063
mixed forest	0.57	0.29	0.38	9102
transitional woodland-shrub	0.00	0.00	0.00	183
water bodies	0.61	0.16	0.26	454

	Correct Pred.	Incorrect Pred.	Overall Accuracy
MLP	116999	283001	34.19%

Analysis

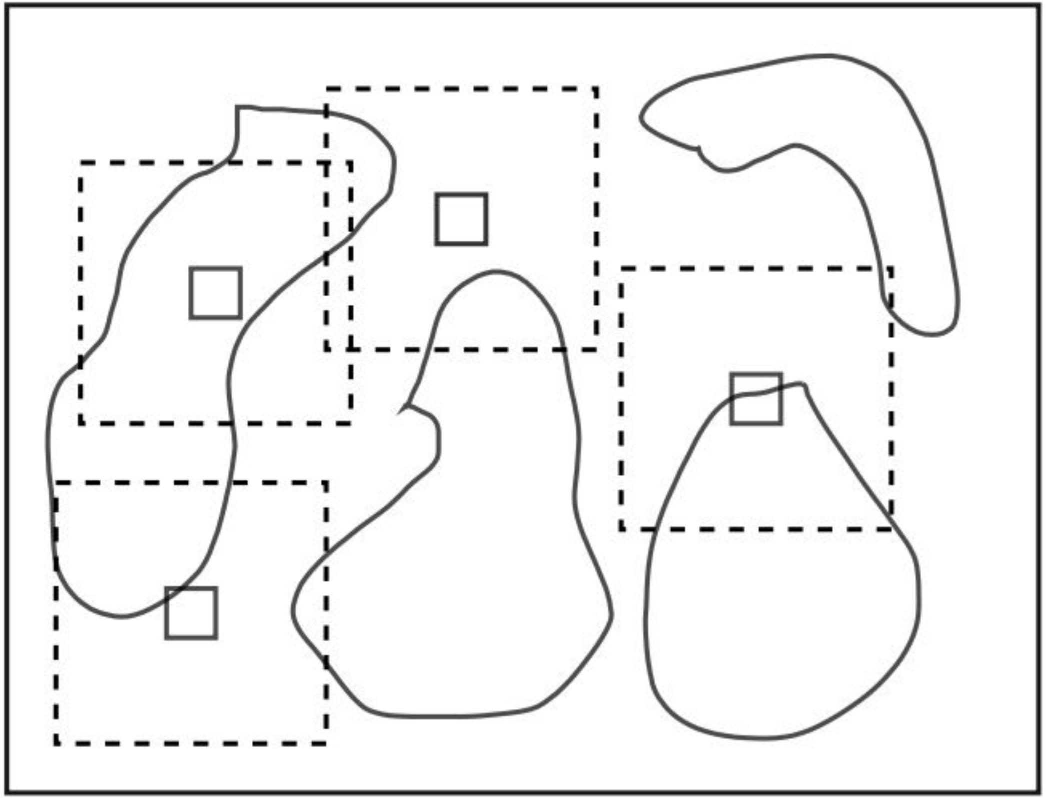
- Varying levels of support between classes.
- Disparity between validation and prediction sets.
- Class-by-class performance.
- Pixel-wise training.

Patch Method

- Training was still constrained to one pixel at a time.
- Wanted to expand our inputs to multiple pixels.
- Gain knowledge of spatial content within an input.
- Could achieve this by extracting clusters, or patches, of pixels.

Approach

- By randomly sampling pixels in the same way, we could extract a buffer-zone around each, forming a patch of pixels.
- We still wanted to provide predictions for pixels, not patches.
- Predicted the centre pixel label based on its surrounding patch.



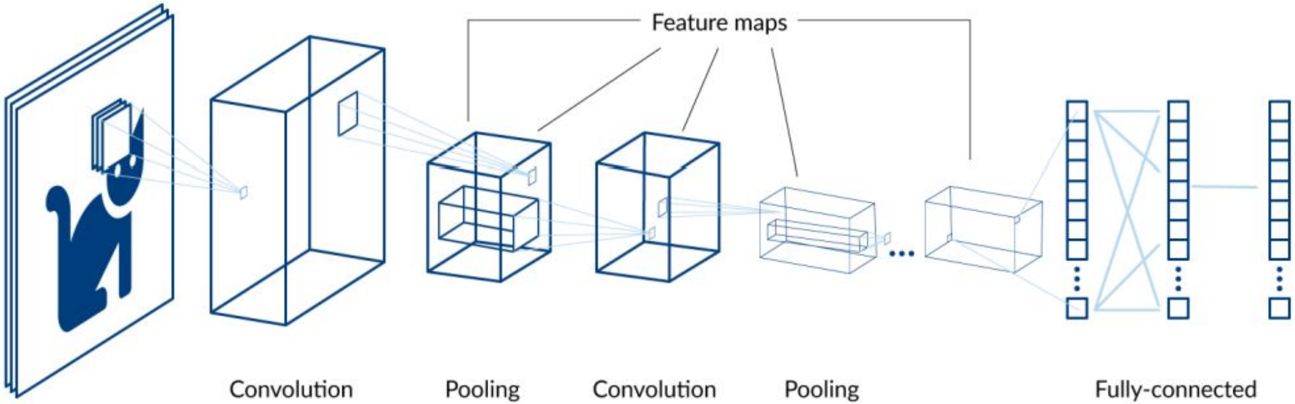
- By tweaking the buffer size, we could construct datasets of differently sized patches:

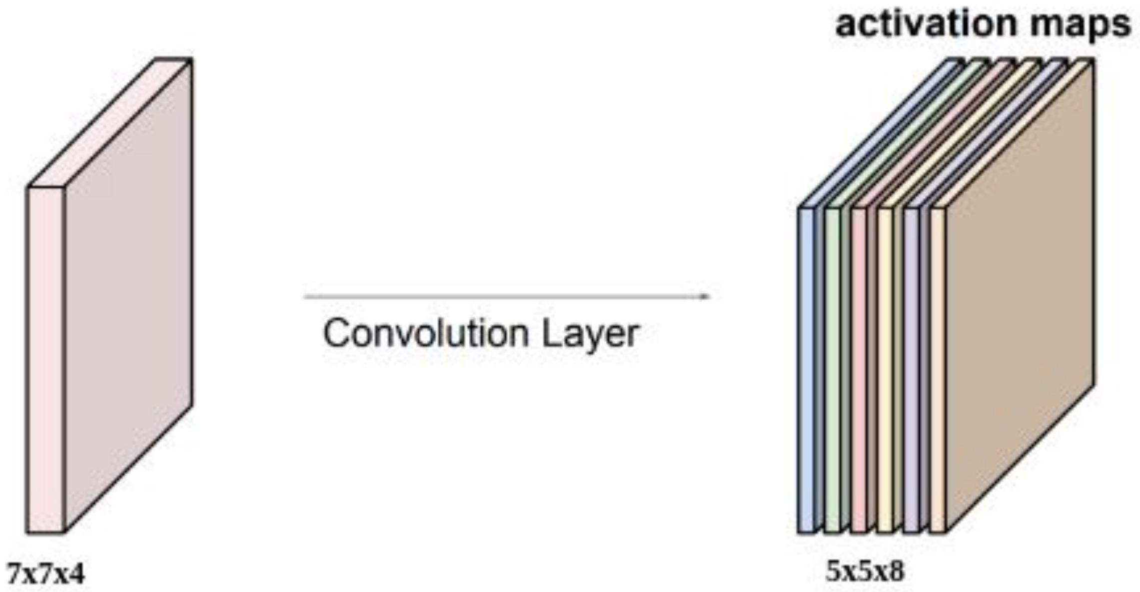
- Patches were three dimensional.
- One layer for each band.
- Dataframe consisted of band value and label for every pixel in patch.

3x3 Patch | 7x7 Patch | 21x21 Patch

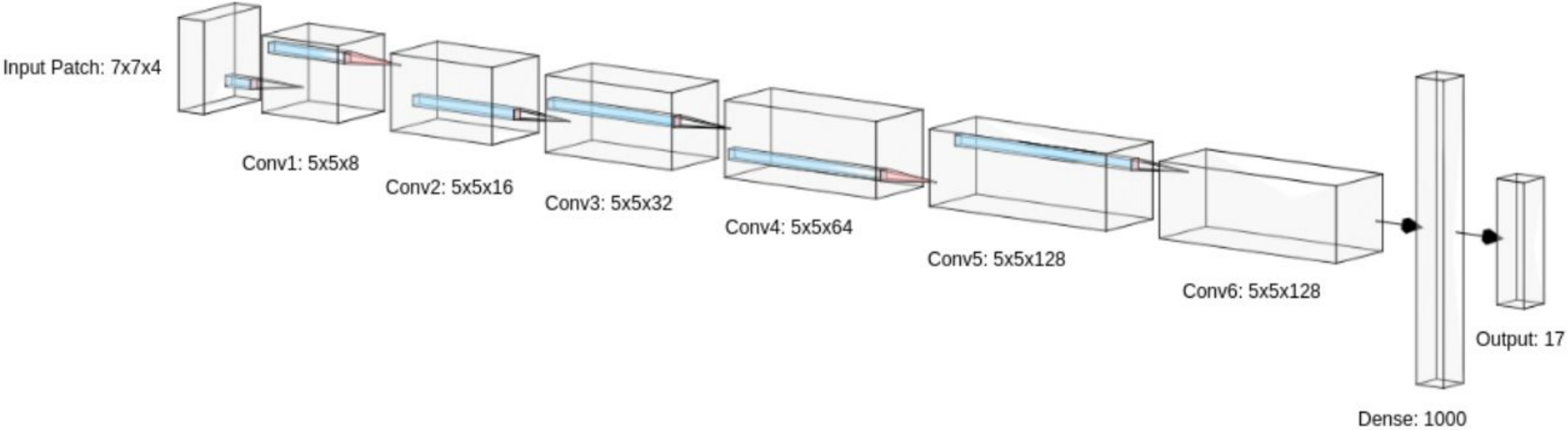
Training

- Extracted between 10,000 and 500,000 patches.
- Learning relationship between the patch values and the centre label.
- This time, one individual represented an image, instead of a pixel.
- Convolutional Neural Networks (CNNs):




activation maps
Convolution Layer
7x7x4
5x5x8

Model


Input Patch: 7x7x4
Conv1: 5x5x8
Conv2: 5x5x16
Conv3: 5x5x32
Conv4: 5x5x64
Conv5: 5x5x128
Conv6: 5x5x128
Dense: 1000
Output: 17

Training and Validation

- 60/20/20 split.
- Training varied depending patch size:

Patch Dimensions	Input Values	Validation Acc (%)
3x3x4	36	58.32
5x5x4	100	59.66
7x7x4	196	60.27
11x11x4	484	49.06
21x21x4	1764	35.78
61x61x4	14884	41.32

Prediction

- GSI metrics
- Scene-wide prediction

	Land-Type	Precision	Recall	F1-Score	Support
GSI-RF	*continuous urban*	0.05	0.92	0.09	4042
	discontinuous urban	0.42	0.52	0.46	39915
	industrial/commercial	0.64	0.94	0.76	75389
	road and rail networks	0.00	0.00	0.00	6459
	airports	0.21	1.00	0.35	51498
	green urban areas	0.00	0.00	0.00	0
	sport/leisure facilities	0.00	0.00	0.00	1540
	non-irrigated arable land	0.22	0.24	0.23	51650
	pastures	0.98	0.14	0.25	28582
	broad-leaved forest	0.39	0.45	0.42	51063
	mixed forest	0.08	0.85	0.15	78961

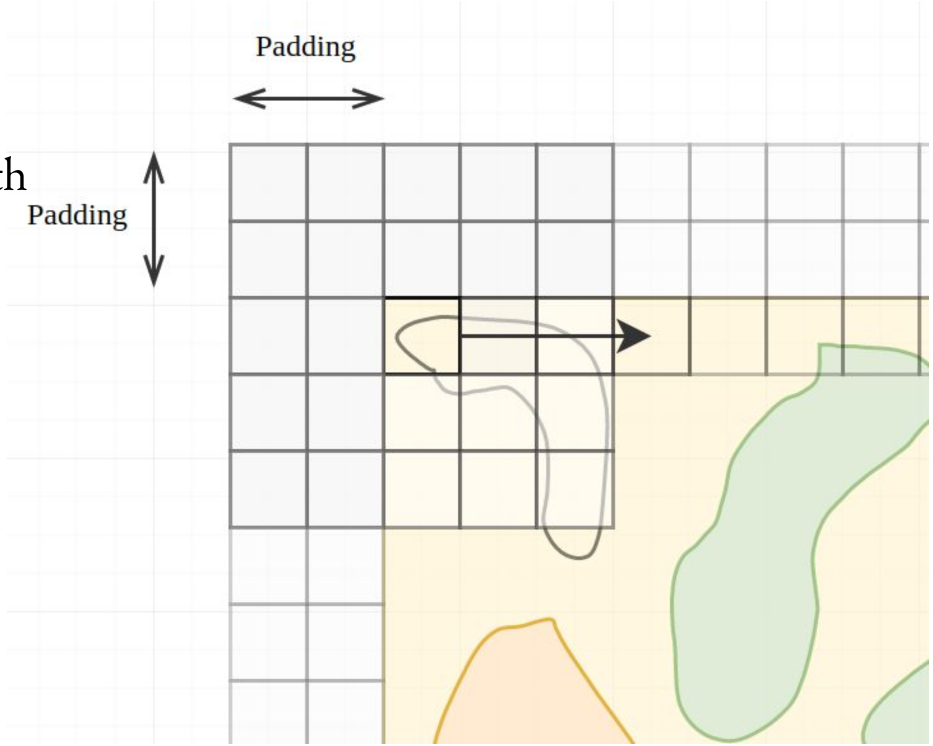
	Correct Pred.	Incorrect Pred.	Overall Accuracy
GSI-RF	145479	254521	37.00%

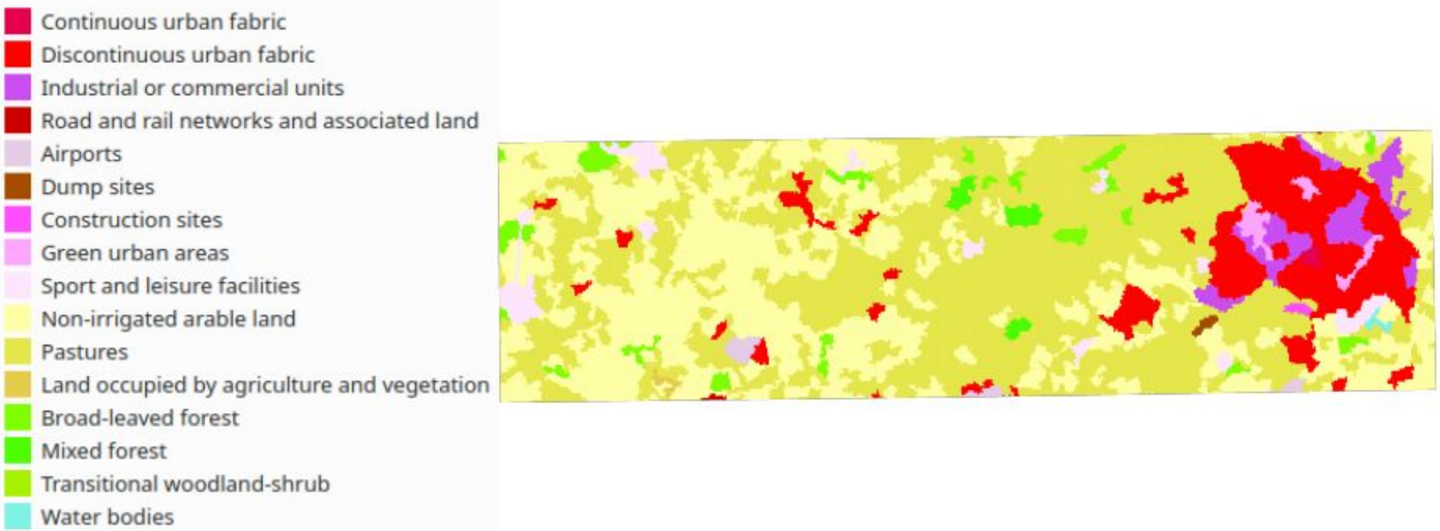
Land-Type	**Precision**	**Recall**	**F1-Score**	**Support**
continuous urban fabric	0.33	0.05	0.09	97
discontinuous urban fabric	0.53	0.70	0.66	567
industrial or commercial units	0.58	0.32	0.41	42588
road and rail networks land	0.55	0.38	0.00	10741
airports	0.30	0.10	0.15	2000
dump sites	0.22	0.04	0.00	429
construction sites	0.44	0.11	0.00	422
green urban areas	0.58	0.10	0.00	2968
sport and leisure facilities	0.47	0.39	0.00	9025
non-irrigated arable land	0.55	0.70	0.62	134621
pastures	0.75	0.55	0.63	181995
agriculture/vegetation	0.42	0.00	0.00	514
broad-leaved forest	0.44	0.27	0.33	51063
mixed forest	0.57	0.29	0.38	9102
transitional woodland-shrub	0.33	0.00	0.00	183
water bodies	0.41	0.15	0.22	454

	Correct Pred.	**Incorrect Pred.**	**Overall Accuracy**
CNN-7x7	187520	212480	43.94%

Full AOI Prediction

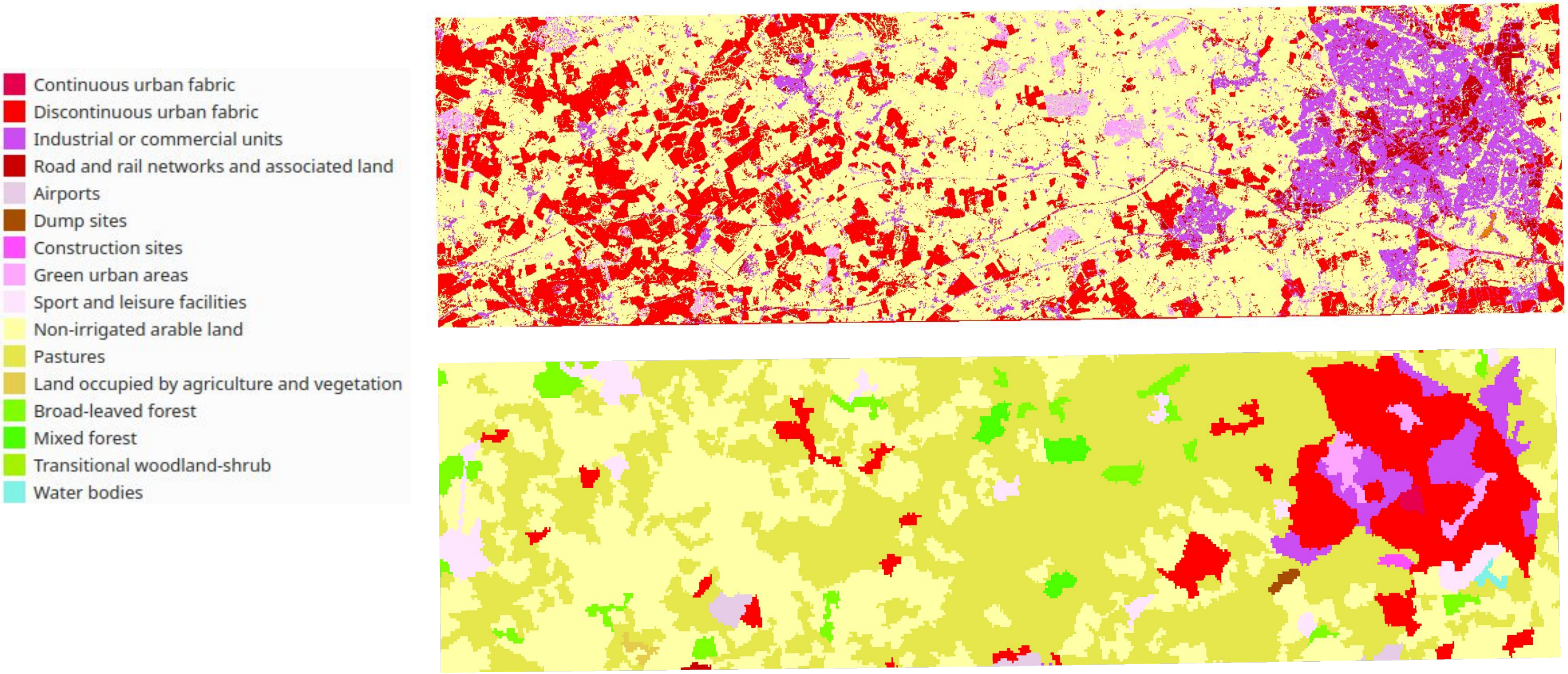
- Wanted to provide pixel-level predictions over the entire AOI.
- Had to partition into patches.
- Sliding patch-window mechanism.
- Padded the boundaries of the scene with empty pixels.
- Cropped the image, predicted centre pixel, and continue.
- Predicted with CNNs trained on two different patch sizes.



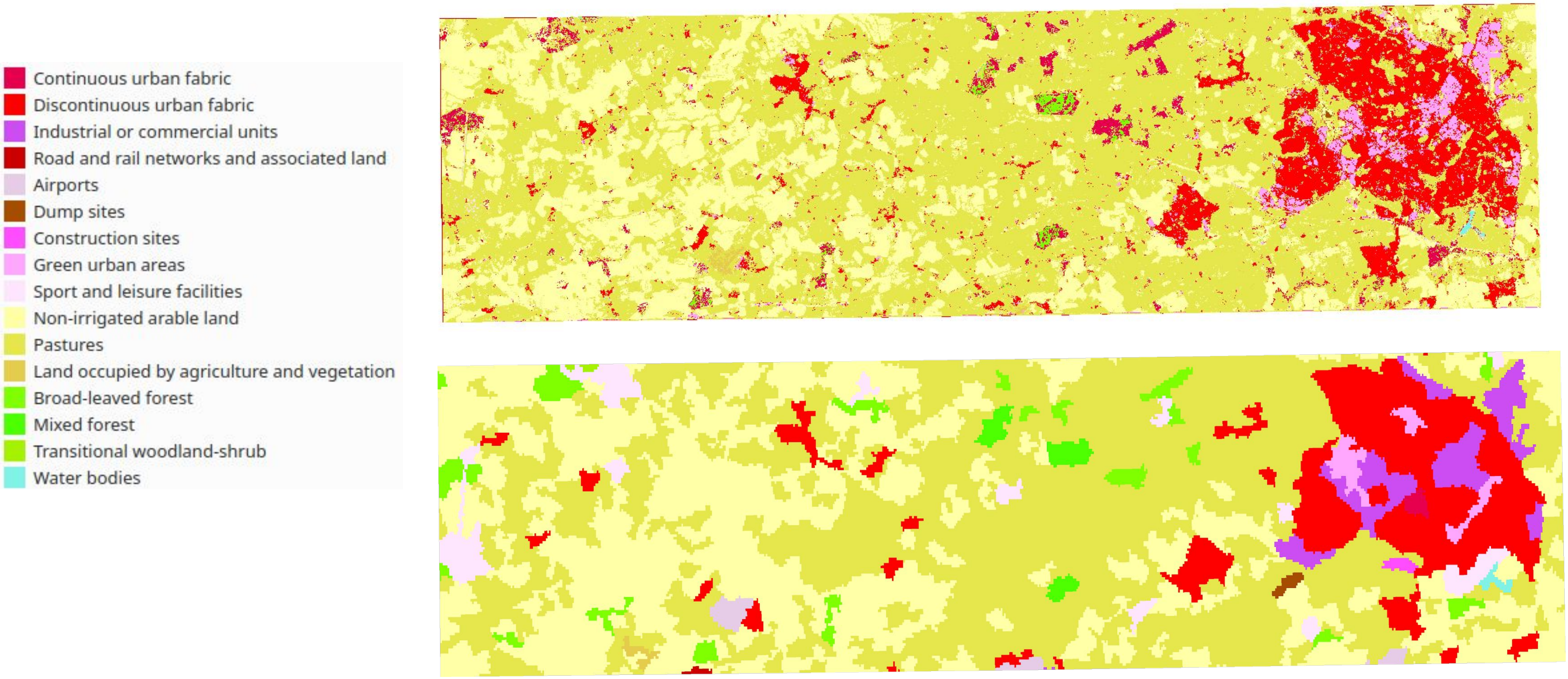


Continuous urban fabric
Discontinuous urban fabric
Industrial or commercial units
Road and rail networks and associated land
Airports
Dump sites
Construction sites
Green urban areas
Sport and leisure facilities
Non-irrigated arable land
Pastures
Land occupied by agriculture and vegetation
Broad-leaved forest
Mixed forest
Transitional woodland-shrub
Water bodies

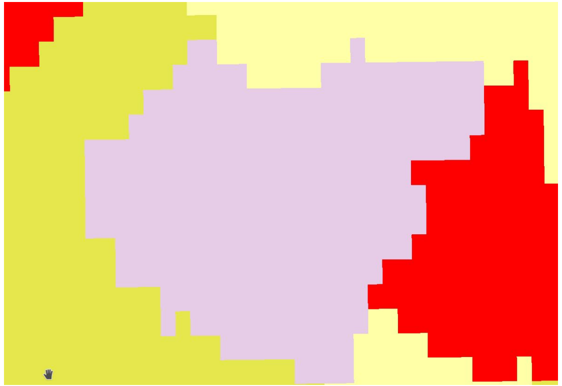
CNN trained on 3x3 patches



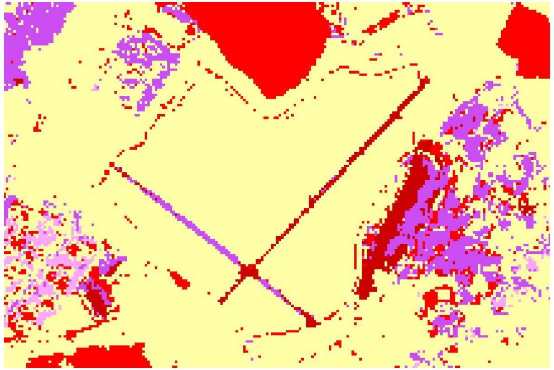
CNN trained on 7x7 patches



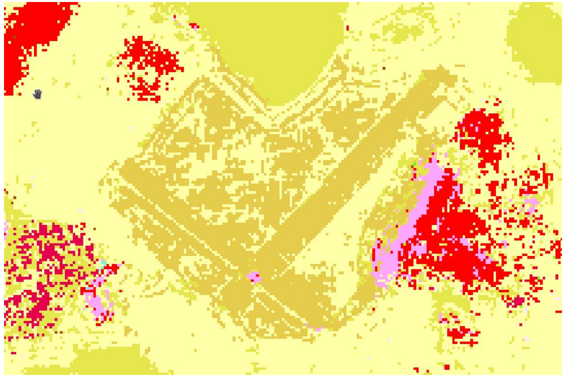
True Colour



CORINE

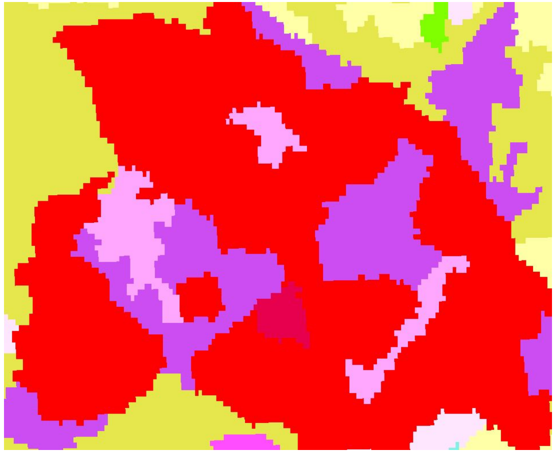


CNN-3x3 Predictions

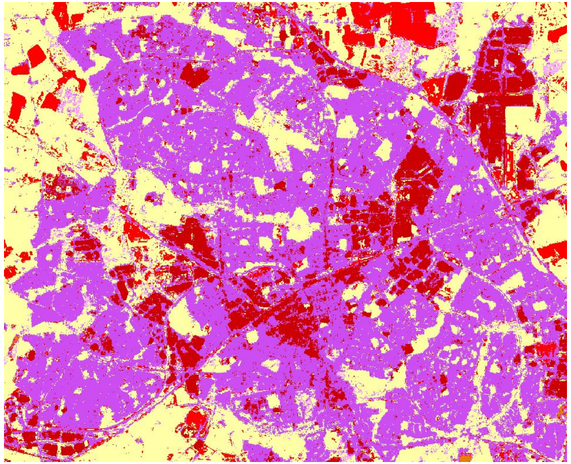


CNN-7x7 Predictions

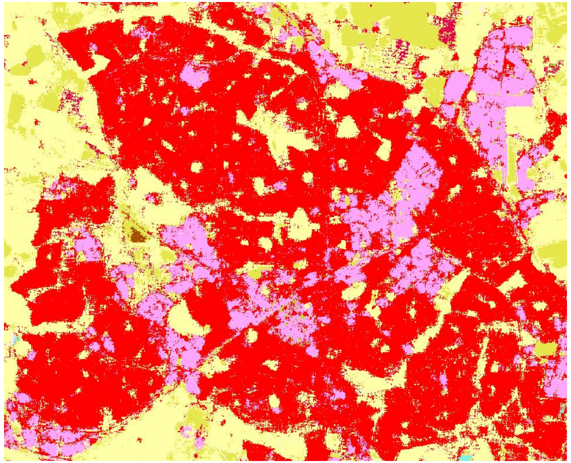
True Colour



CORINE

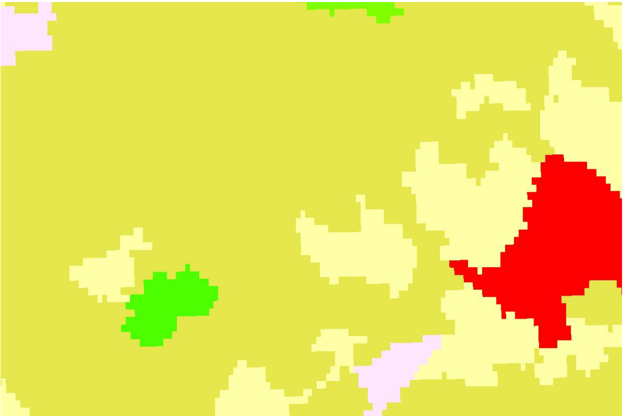


CNN-3x3 Predictions

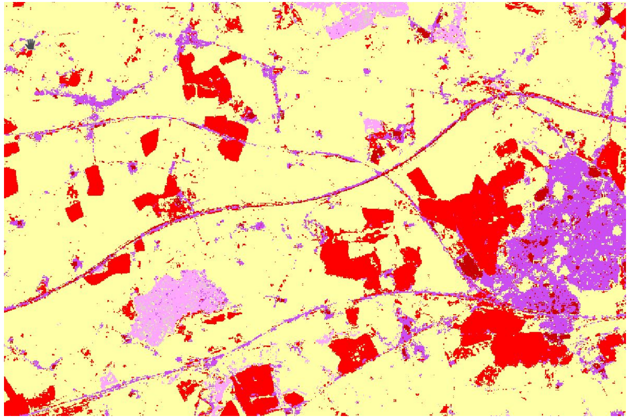


CNN-7x7 Predictions

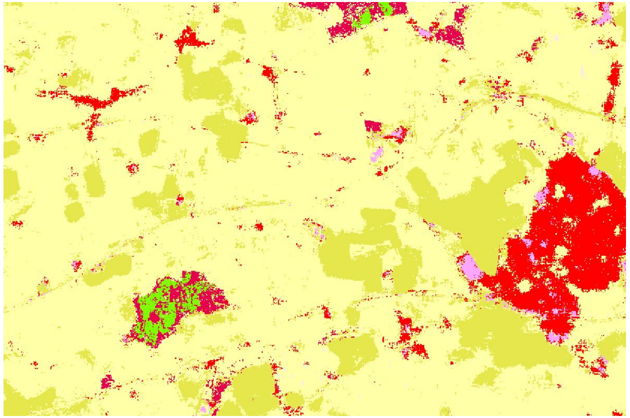
True Colour



CORINE



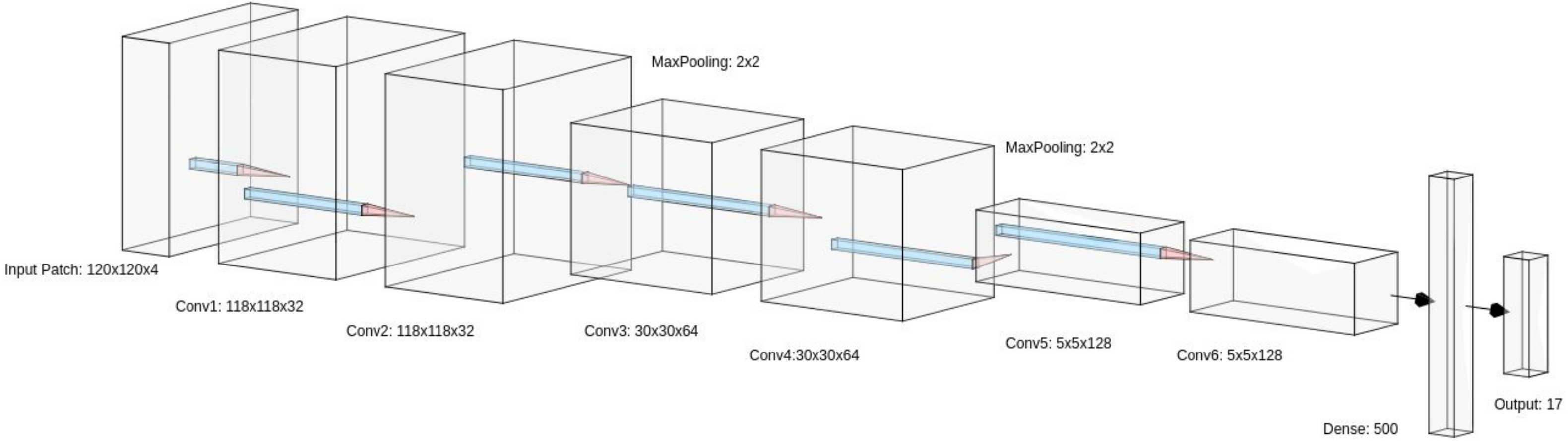
CNN-3x3 Predictions



CNN-7x7 Predictions

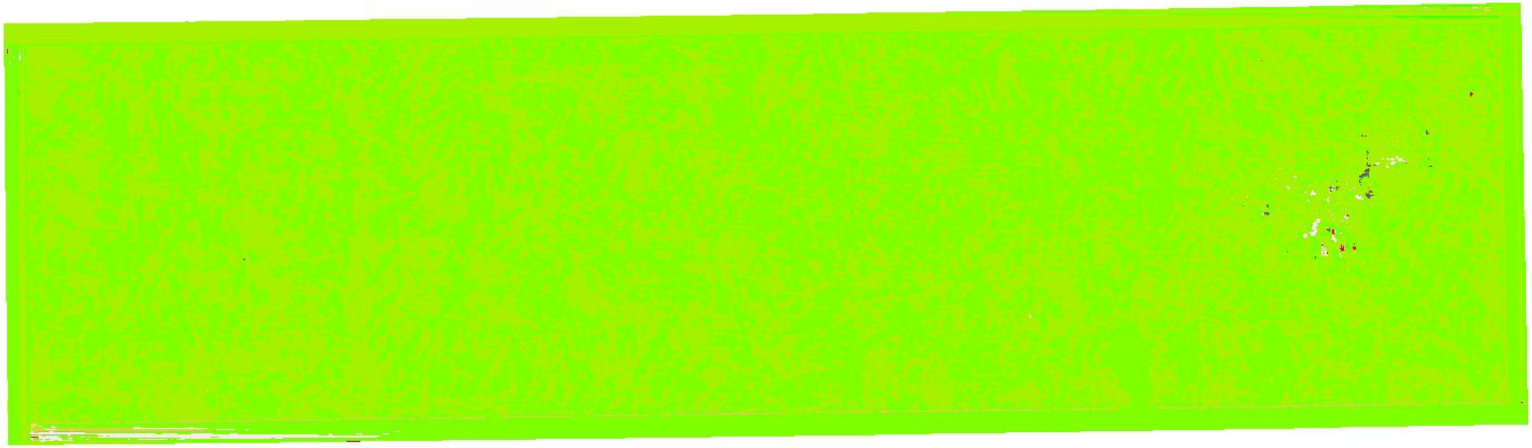
Transfer Learning Approach

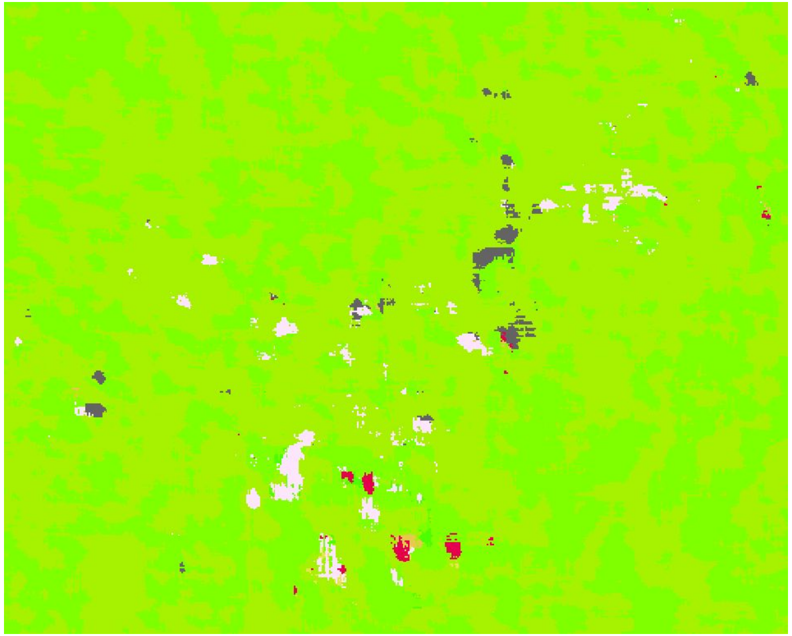
- BigEarthNet: largest archive of high-res Sentinel-2 imagery available.
- Images in the form of 120x120 patches.
- Trained on subset of the overall dataset.
- Used a similar CNN model.



Performance

- Had to severely cut down the dataset to fit our project constraints.
- Struggled to achieve >33% validation accuracy.
- Scene-wide predictions:



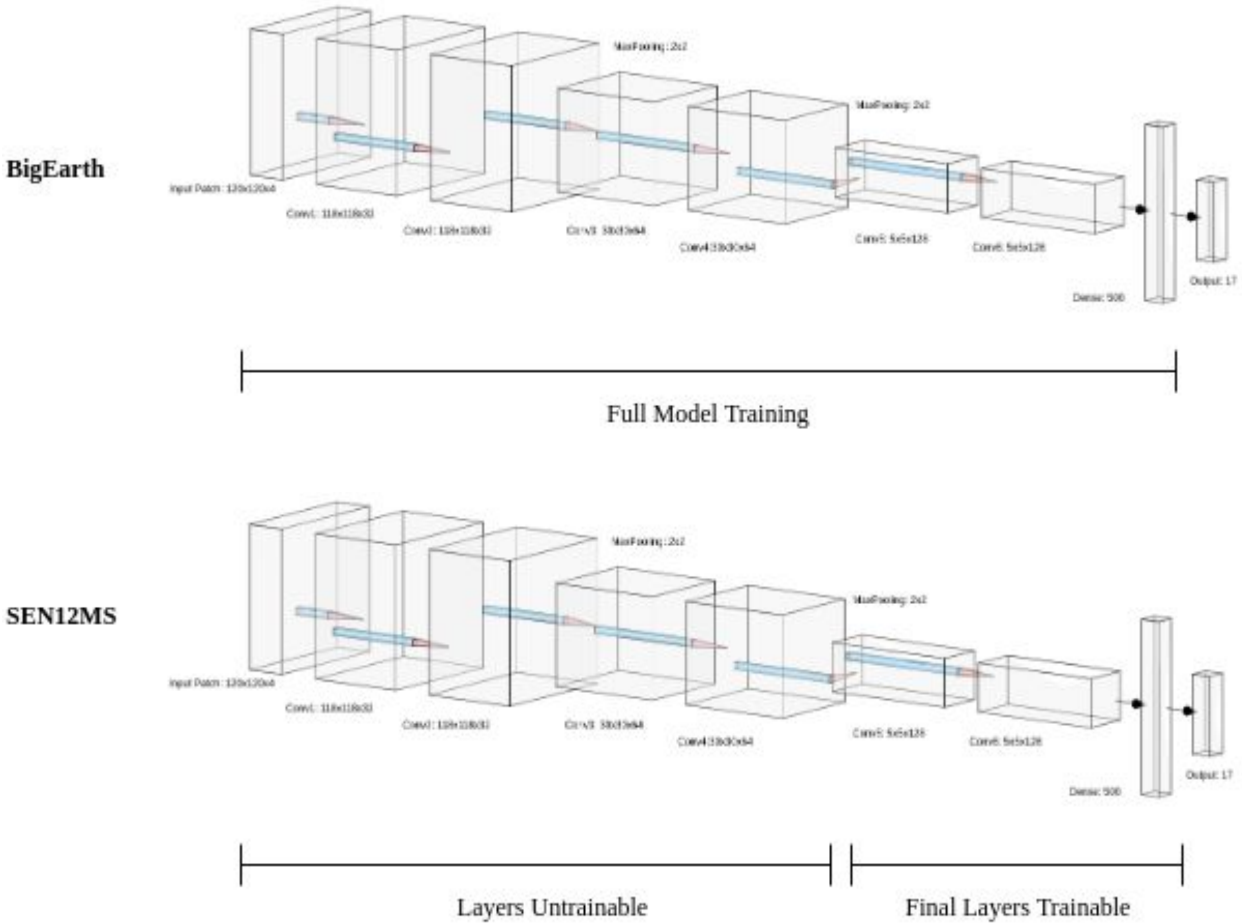



Analysis

- Excellent shape detection using patches.
- Patch-size important predictor.
- Smaller the patch better at detecting finer details, larger patch better at getting actual class.
- Limits to patch size.
- Promising potential.

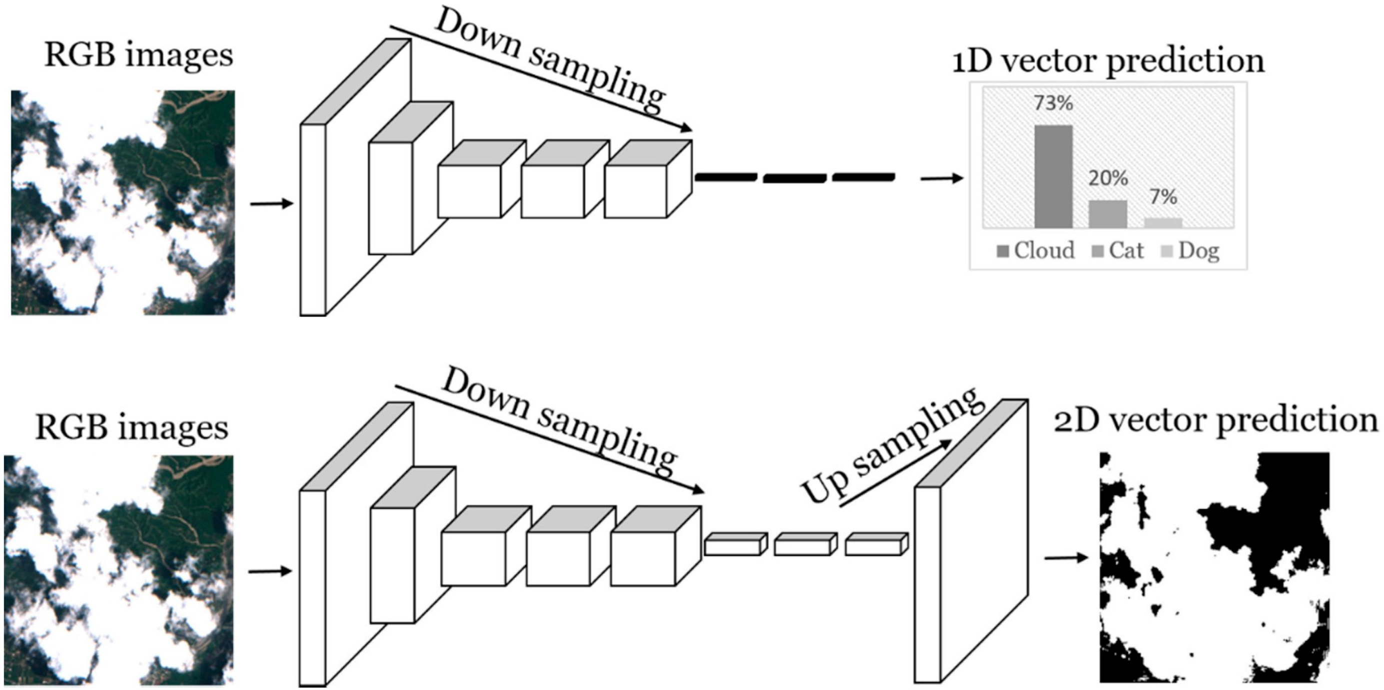
Future Work

- Further explore Transfer Learning
  - BigEarth + SEN12MS = > 1 million Sentinel-2 patches
  - Train one model on multiple datasets:

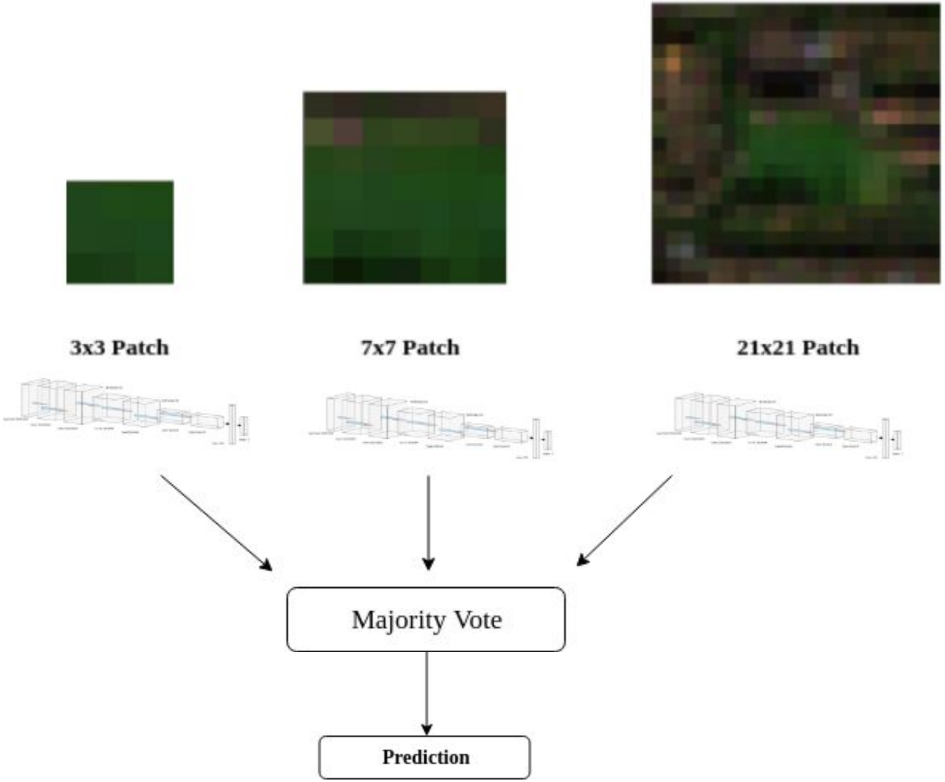


BigEarth
MaxPooling: 2x2
MaxPooling: 2x2
Conv5: 5x5x128
Conv6: 5x5x128
Dense: 500
Output: 17
Full Model Training
SEN12MS
MaxPooling: 2x2
MaxPooling: 2x2
Conv5: 5x5x128
Conv6: 5x5x128
Dense: 500
Output: 17
Layers Untrainable
Final Layers Trainable

- Semantic Segmentation
  - Shrink patch, predict, resize it back to original dimensions
  - Output segmented heatmap with pixel-wise predictions:



- Ensemble Learning:
    - Run multiple models in parallel.
    - Combine the range of information from different patch sizes:



Conclusion

- Considering local image regions can have a positive impact on accuracy.
- Pick up shapes and details overlooked by CORINE.
- The size and content of these regions are critical.
- Combining these techniques with the use of outside datasets and model ensembles.

Thanks for listening! Any questions?

- Thanks for listening!
- Any questions?