A Supplementary Materials for Constructing a Visual Dataset to Study the Effects of Spatial Apartheid in South Africa

In our supplementary materials, we describe our experiments and results in more detail. All related code and instructions on how to download the dataset will be found at: `https://github.com/sefalab/Spatial_Project`

A.1 Data Annotation Procedure

A.1.1 Data subset used to iterate on the dataset labeling process

While constructing and refining the dataset, we iterated on 19; $21,688 \times 21,688$ pixels of satellite images from Gauteng, Limpopo, North West, Free State and Mpumalanga provinces (Figure 1). Although we had planned to perform our experiments only on the Gauteng province, Gauteng has a non-rectangular shape which means that the satellite images also cover parts of the neighboring provinces (Limpopo province, North West province, Free state province and Mpumalanga province) as depicted in Figure 1. The test set covers the South West of the dataset while the validation set covers the North East and Southern parts of the dataset sample. Table 1 shows the number of images in each split.

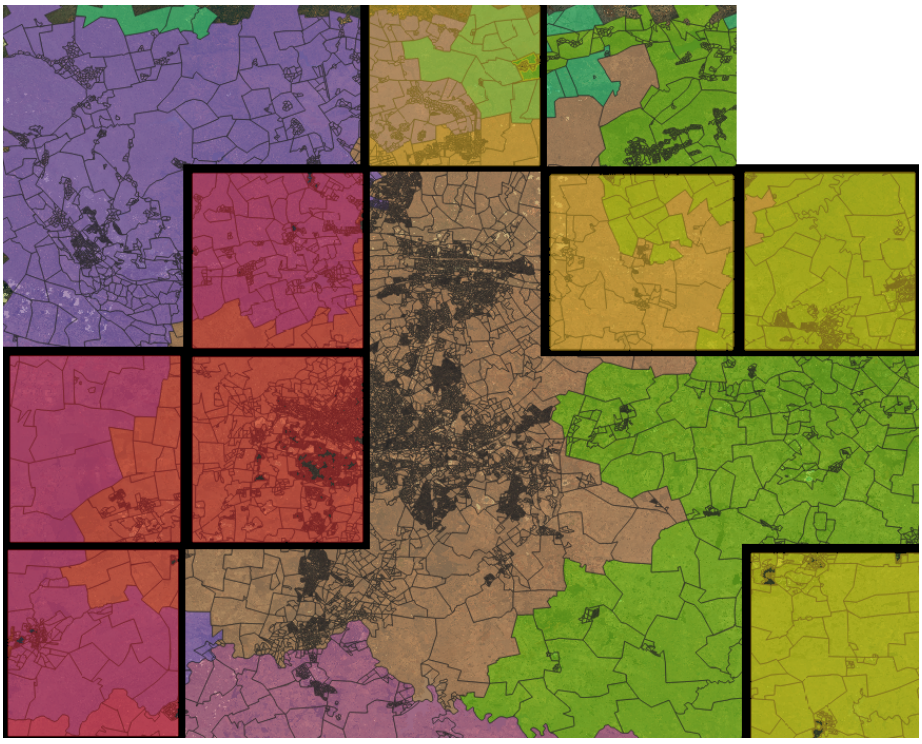


Figure 1: The subset of data chosen for creating the final dataset for training a neighborhood classification model. Each block represents one $21,688 \times 21,688$ pixel satellite image and there is a total of 19 images in this subset. The blocks in red represent images in the test set, those in yellow represent images in the validation set and the rest of the data is in the training set.

Dataset	**Number of images** ($21,688 \times 21,688$ **pixels)**	**Number of images** (256×256 **pixels)**
Training set	11	77,616
Validation set	4	28,224
Test set	4	28,224
Total	19	134,064

Table 1: Splitting the data to create models before scaling to the entire country.

A.1.2 Training procedure for U-Net model used to refine dataset

To perform the experiments below, we first represented each satellite image of size $21,688 \times 21,688$ by tiles of images of size 256×256. Since $21,688 \times 21,688$ is not a factor of 256×256, we could not cover the entirety of a satellite image by 256×256 images, but a portion of the image which is of size $21,504 \times 21,504$–a factor of 256×256.

Our U-Net model was first trained with the following hyper-parameters as defined in [56]. It has 23 convolution layers, with the Relu activation function between the layers and the Softmax activation function on the last layer. The model was trained with the Adam optimizer at a learning rate of $1e-4$ [29], and we applied batch normalization with batch sizes of 32. We used the categorical cross-entropy loss function since there are multiple classes and each pixel can only belong to one of the classes [31]. A model with these specifications was used to guide us in the dataset creation process and as a baseline model for experiments on the final dataset.

We divided the training data into 32 mini-batches and trained the U-Net model for 30 epochs and $2,425$ steps per epoch (length of training data divided by the batch size)–these hyperparameters were found through a grid search on our validation set. The error and validation accuracy did not change significantly for values higher than these numbers. The 100 steps per epoch hyper-parameter is used to define how many mini-batches of samples to use in one epoch. We divided our dataset into mini-batches because our experiments use large datasets and using mini-batches reduces training time given that each mini-batch can be processed in parallel on the GPU. Thus, the number of steps per epoch is calculated based on the size of the dataset and the size of the mini-batches.

Balancing training data: Without any sampling, 81% of pixels in our training data consist of the background class. To enable the model to classify non-vacant land, we balanced the training and validation data by reducing the number of vacant land pixels so that the model can learn to distinguish the other classes. We balanced the training data by only keeping images which do not have vacant land pixels, or if they do, ensuring that these pixels cover at most 70% of the image. This method eliminates many of the images consisting of buildings like farm-houses and those representing parks and recreational land. This is because buildings like farm-houses are usually surrounded by farms/vacant land. Thus, over 70% of the pixels in an image of a farm-house will consist of vacant land resulting in the image being discarded from the training data.

A.1.3 Confusion between vacant land and undeveloped land using solely the EA dataset

Farmland that has not been cultivated looks like vacant land. Since the EA labels only specify the designated land use; vacant, farm, commercial and industrial lands can be confused for each other because undeveloped land often looks like vacant land. It is very difficult even for humans to distinguish farmland from smallholding land in this image on Figure 2 for example, and many of these neighborhoods have open spaces which look like farmland/vacant land. Using solely the EA data as labels, there is no way of tracking which land is developed and which is not. The model trained with solely the EA dataset as labels was also unable to distinguish between wealthy neighborhoods with large yet undeveloped yards, a patch of which could look like a farm (column 2 of Figure 3), and farmland.

A.1.4 More details on Preparing EA+Building point data

Buffing points into polygons: As mentioned in Section 4, our first step was to use the buffer algorithm to transform each building point (a single latitude and longitude coordinate pair) into a circular polygon of a specific radius. In our case, we inflated the points by a distance of 0.0007 decimal degrees.

Each circle can either be too big in informal areas where there are small compounds or too small for the building as in the case of the industrial neighborhood class. A polygon that is slightly too large for the compound will still allow our neighborhoods to be appropriately labeled as in row 2, column 2 of Figure 4. But a circle around a larger compound, e.g. in smallholdings or industrial areas, only usually covers the building as shown in row 1, column 2 of Figure 4. We assume that a semantic segmentation model such as U-Net that has access to images at different scales should be able to learn visual characteristics distinguishing these classes given enough examples.

Larger figures illustrating the steps described in Section 4 of the paper are reproduced in Figure 6.

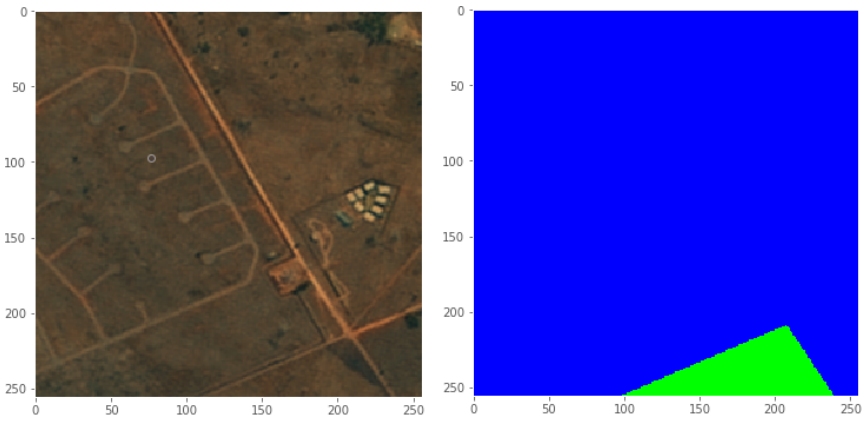


Figure 2: Example of an image with an associated label of what the land should be used for, while the land is not yet developed for the intended use. In this case the blue and green denote small-holding and farm ground-truth labels respectively. The land is designated to be used for small-holdings but is still vacant.

Split	Satellite	Tiled
Training set	60	1,121,904
Validation set	20	373,968
Testing set	20	373,968
Total	100	1,869,840

Table 2: Reproduced from Section 5. The number of images in each split for our baseline experiments. Satellite refers to the $21,760 \times 21,760$ resolution satellite images and tiled refers to the 256×256 images.

A.1.5 More samples of images and masks

Figure 7 reproduces images in Section 4 for ease of viewing, and Figure 8 shows examples of our dataset with 12 (uncollapsed) class labels.

A.2 Experiments

A.2.1 Sampling data for experiments

There are 550 satellite images covering the entire country. However, most of these images are of farmland/non-built up land. To perform our baseline experiments, we sampled from regions with more residential neighborhoods. In all our experiments, we tile each satellite image of size $21,688 \times 21,688$ to many images of size 256×256 pixels. As 256×256 is not a factor of $21,688 \times 21,688$, we upsampled the satellite images from $21,688 \times 21,688$ to $21,760 \times 21,760$ so that we could cover one satellite image in its entirety with tiles of 256×256 images. We performed this upsampling using the GDAL library[1]. After this process, each upsampled satellite image is covered by $7,225$ images of size 256×256, giving us a total of 3,973,750 256×256 images for 2011.

In total, the sampled dataset consists of 100 images of size $21,760 \times 21,760$ pixels from the year 2011. As illustrated in Table 2 this equates to approximately $1,869,840$ images of size 256×256. Table 3 shows the distribution of these pixels per class per province.

A.2.2 Sampled dataset train/validation/test split

The training data consists of provinces in northern South Africa (which includes all the images we used to create the dataset), and the KwaZulu-Natal province located on the eastern coast of South Africa. This is 60 images of size $21,760 \times 21,760$ pixels from the following 6 provinces: Gauteng, North West, Limpopo, Free State, Mpumalanga and Kwa-Zulu Natal. The test set consists of 20 images of size $21,760 \times 21,760$ pixels from both the Western Cape and Northern Cape provinces.

[1] `https://gdal.org/`

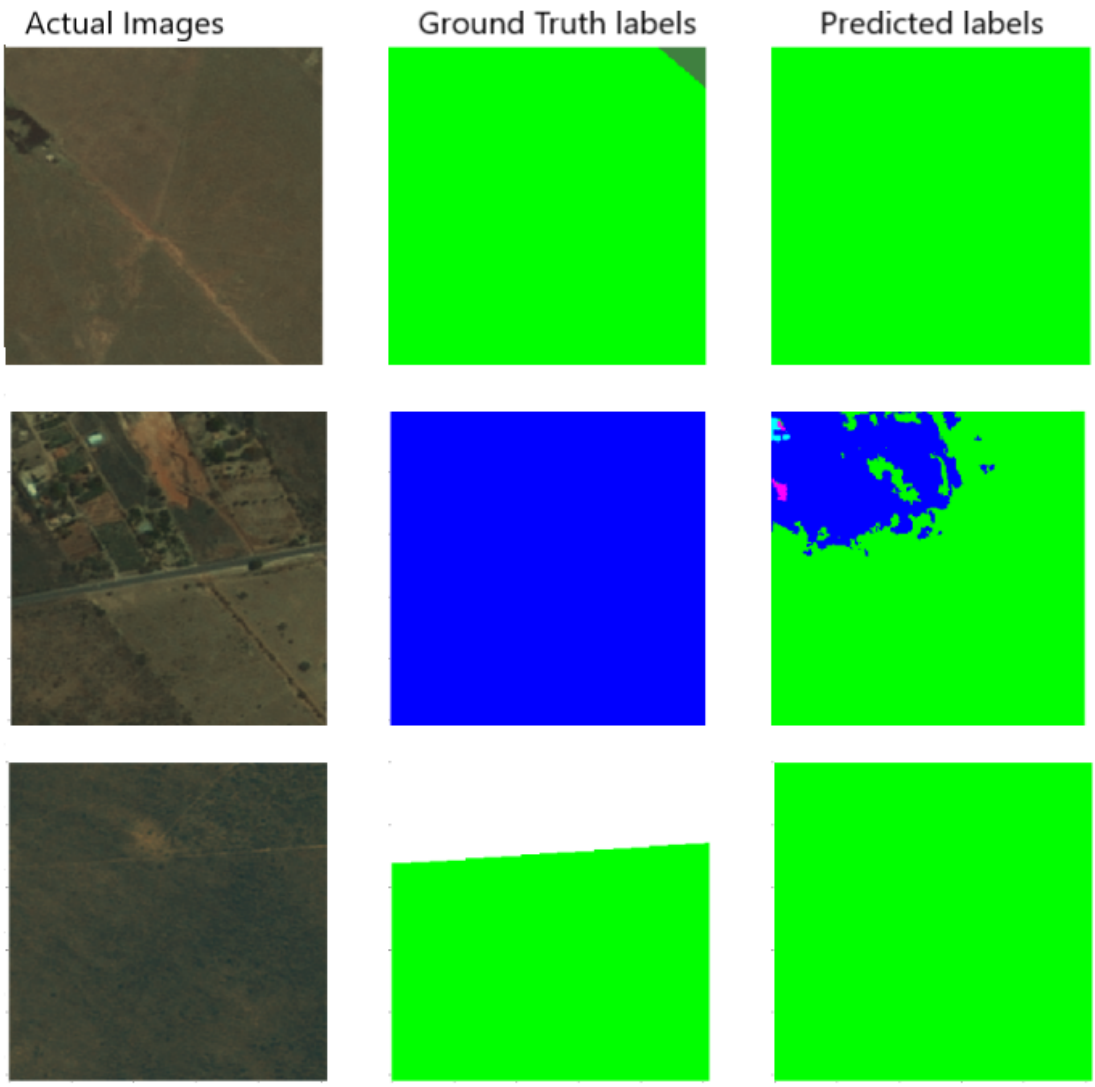


Figure 3: Examples of images from our 12 class dataset created using EA ground truth labels. The first column shows the input images, the second column depicts the ground truth labels, and the third column shows our model predictions. Yellow: Township, Light blue: Suburb, Pink: Industrial area, Olive green: Commercial land, Red: Informal area, Light Green: Farm, Light grey: Collective living Quarters, Dark Grey: Village, Blue: Smallholdings, Magenta: Parks and Recreational sites, White: Background.

And our validation set consists of 20 images of size $21,760 \times 21,760$ pixels from the Eastern Cape province.

While creating the train/validation/test splits, we ensured that a single neighborhood did not span multiple splits. This was done by having each split sampled by province, i.e. the validation set only spans the Eastern Cape province. We further illustrate this visually in Figure 9 where we show exactly where the samples are on the map.

During the compilation of the training dataset specifically, we picked images which cover different sceneries in the dataset such as densely/sparsely populated areas, mainland/coastal land and different ecosystems like forests/grassland. This is because we would like to train a neighborhood segmentation model that generalizes across the whole country.

Table 3 shows the class distribution by province.

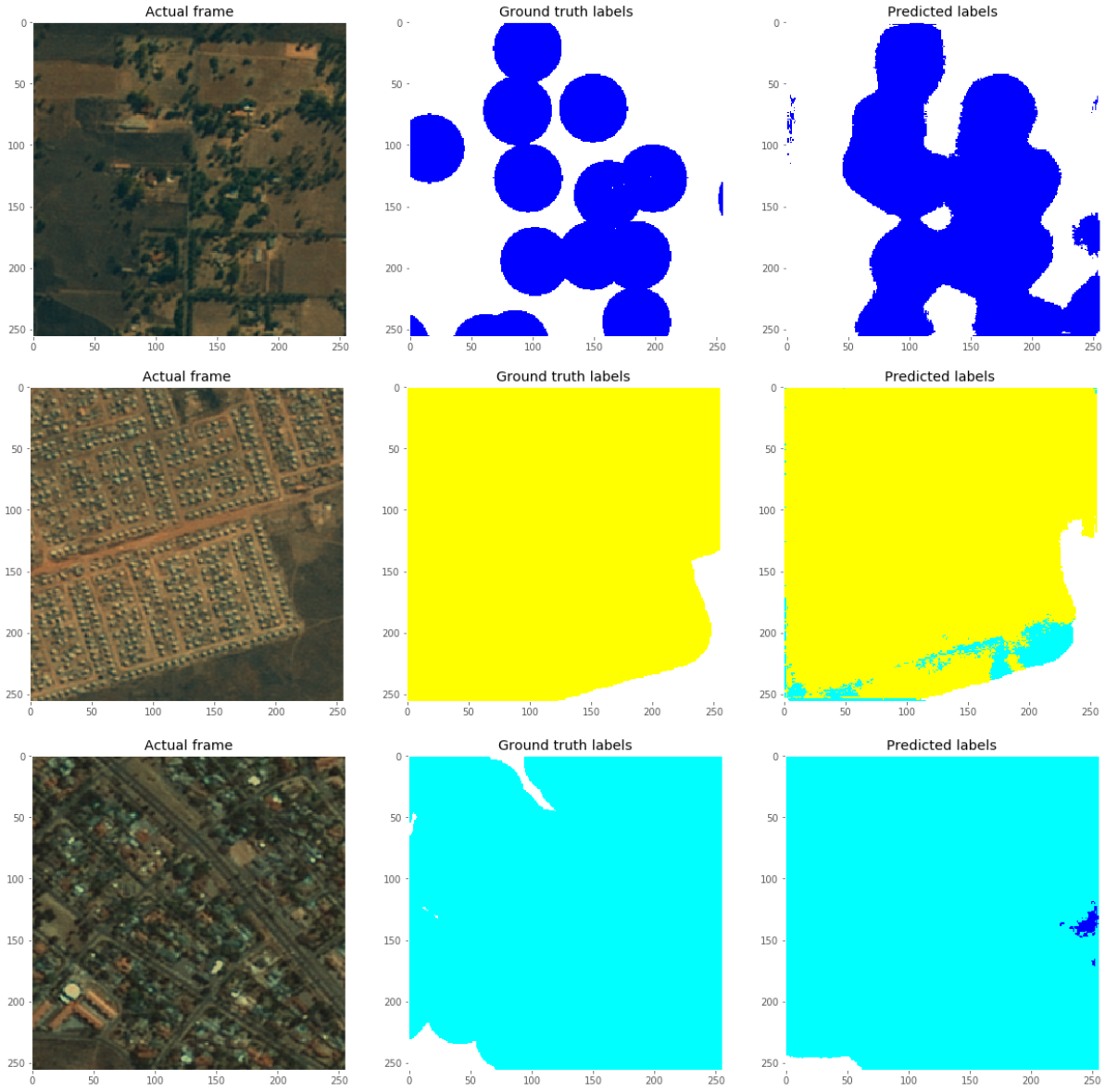


Figure 4: Results from training the U-Net model using labels from the EA and buildings datasets as Ground-truth. Column 1 shows the images, column 2 shows the ground-truth labels and column 3 shows the model predictions. Yellow: township, light blue: suburb, pink: industrial area, olive green: commercial land, red: informal area, light green: farm, light grey: collective living quarters, dark grey: village, blue: smallholdings, magenta: parks and recreational sites, white: background.

A.2.3 U-Net and DeepLabV3+ training details

For both baseline experiments in section 5.1, we used a similar hyper-parameter setup like the one described in Section A.1.2–using a grid search on the validation set to tune the hyper-parameters. The U-Net architecture has 42 layers with the same setup as in [56]. For DeepLabV3+, we use the same architecture as [8]. We used these hyper-parameters for both models: batch sizes of 256 and a weighted (see section 5.1) categorical cross-entropy loss function with the Adam optimizer at a learning rate of $1e-4$. A grid search on our validation set found that with this larger dataset, increasing the number of epochs to 150 gave the best validation accuracy and similarly, increasing the number of steps per epoch to $4,382$ (length of training data divided by batch size) gave better results. Our grid search showed that training past 200 epochs results in overfitting. While training

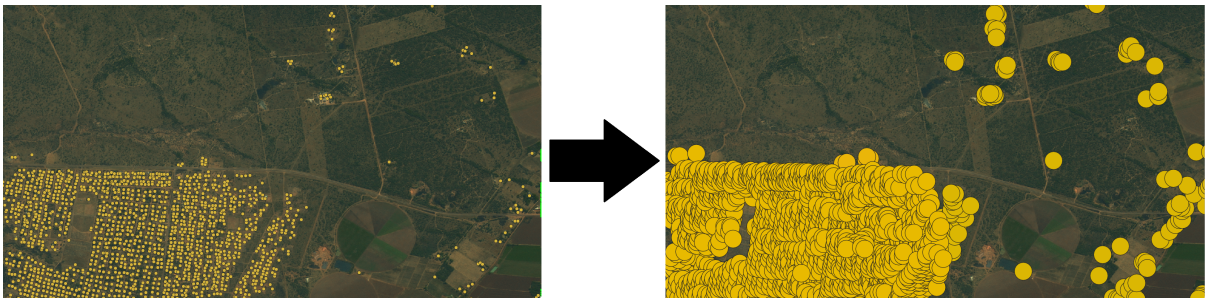


Figure 5: Converting building point data to polygons using a buffer algorithm so that we can approximate the space covered by the building. The images illustrated here are at a resolution of 10m per pixel and size of 786 x 386 pixels.

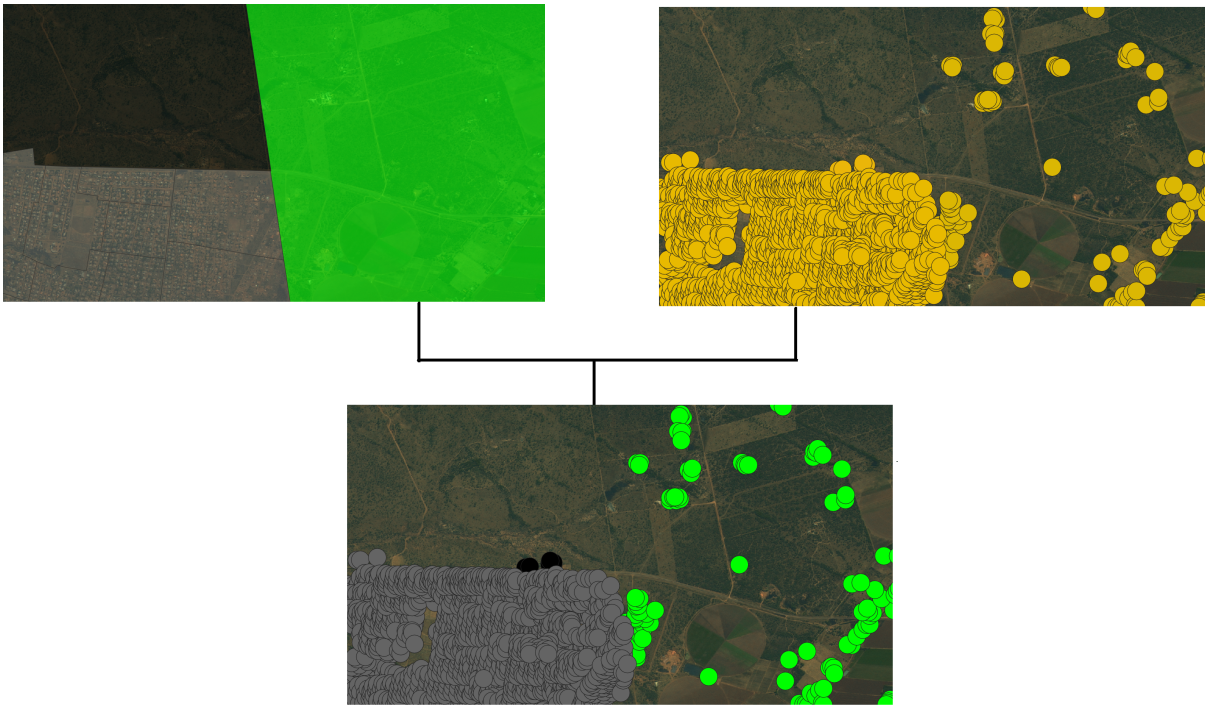


Figure 6: Computing the spatial intersection between the land use labels from the EA dataset and the buffed building polygons so that we can know the neighborhoods in which these houses belong.

Provinces	Pixels (#)				Pixels (%)			
	W	NW	NR	B	W	NW	NR	B
Gauteng	887300609	1034046417	232687267	13763611529	5.574	6.496	1.461	86.467
Limpopo	853566179	655643962	102129867	15802033942	4.902	3.765	0.586	90.746
Mpumalanga	955626063	501826466	227261556	10169240987	8.062	4.233	1.917	85.787
KwaZulu-Natal	1441281799	1603498365	244896953	17969414851	6.779	7.542	1.152	84.526
Free State	782864423	375528191	144284589	13167081773	5.410	2.595	0.997	90.997
North West	1287159471	475138083	147799837	16629643793	6.943	2.562	0.797	89.697
Northern Cape	827708243	984847240	113583209	17882772666	4.178	4.972	0.573	90.276
Eastern Cape	135792755	1244310434	212633563	21192407038	5.656	5.183	0.885	88.274
Western Cape	733576689	111824658	76129426	13791891051	4.986	0.760	0.517	93.736

Table 3: The number and percentage of pixels per class for the sampled dataset. NW=non wealthy, W=wealthy, NR=non residential,B=background.

for 150 epochs in all our experiments, we use the early stopping technique which allows us to stop training if there is no change in the validation loss, and save the best weights for the network.

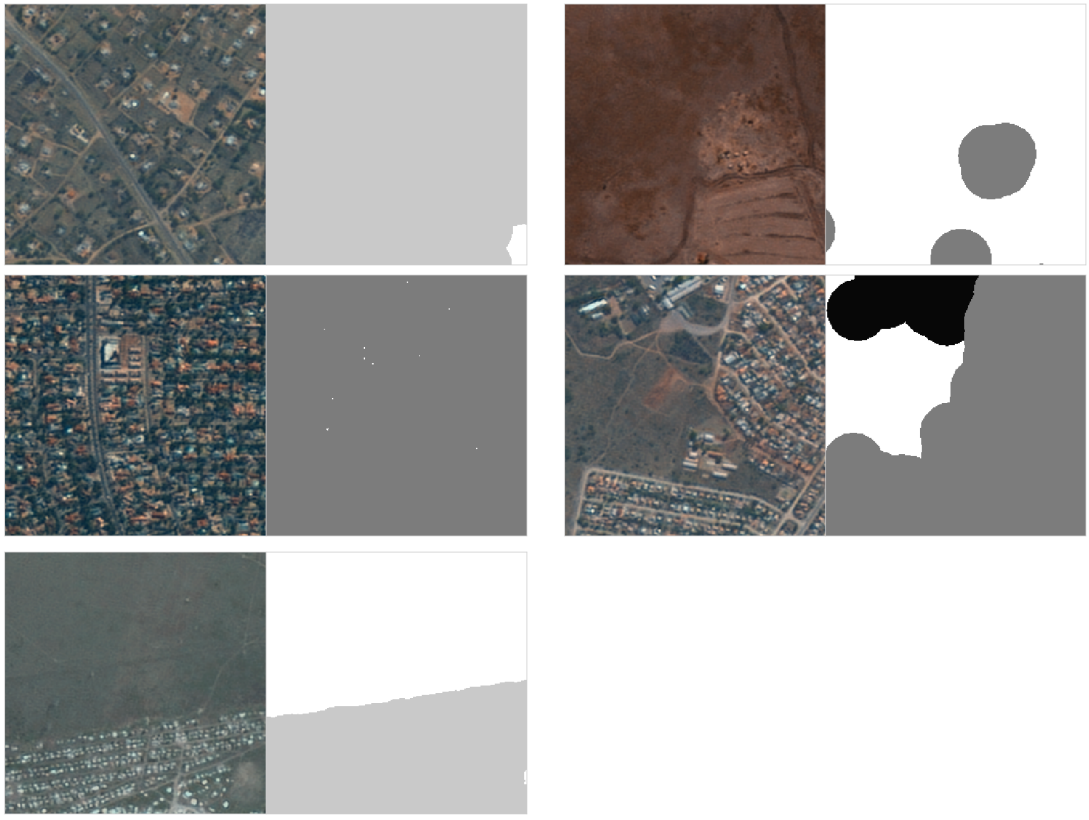


Figure 7: Samples of image mask pairs from our dataset. White: background, black: nonresidential neighborhood, light gray: non wealthy neighborhood, dark gray: wealthy neighborhood.

Training this model took approximately 6 hours using a V3-8 Google Tensor Processing Unit (TPU) on a machine with 36GB of RAM. One significant difference from the testing phase of the data creation step with this technical setup however was performing inference during the testing phase. Semantic segmentation requires inference on individual pixels. $373,168$ images with 256×256 pixels each result in approximately 24.5 billion pixels that have to be forward passed through the model to obtain predictions. This is 13 times the number of pixels we used during the dataset tuning process. Storing the prediction versus ground truth arrays of length 24.5 billion is infeasible using our computational resources. We decided to store predictions as images and store information about things like calculated metrics in comma-separated values (CSV) file formats on disk. We also stored filenames with interesting model predictions such as images with 99% IOU for each model into CSVs on disk for easy retrieval during analysis. We did this for all experiments.

Balancing training data: Even in our sampled dataset, over 92% of our pixels consist of the background class. Thus we balance the training data using the same procedure described in A.1.2. This is done before weighting each class as discussed in Section 5.1.

A.2.4 Additional Confusion Matrices from Experiments Across Provinces (Section 5.1)

Figures 10, 11 12, 13, 14, 15 show additional confusion matrices for experiments training our models on 8 provinces and testing them on the 9^{th}.

A.2.5 Additional Confusion Matrices from Experiments on Sampled Dataset (Section 5.1)

Figures 16, 17 show confusion matrices for models trained on our sampled dataset. Results are reported on the test set.

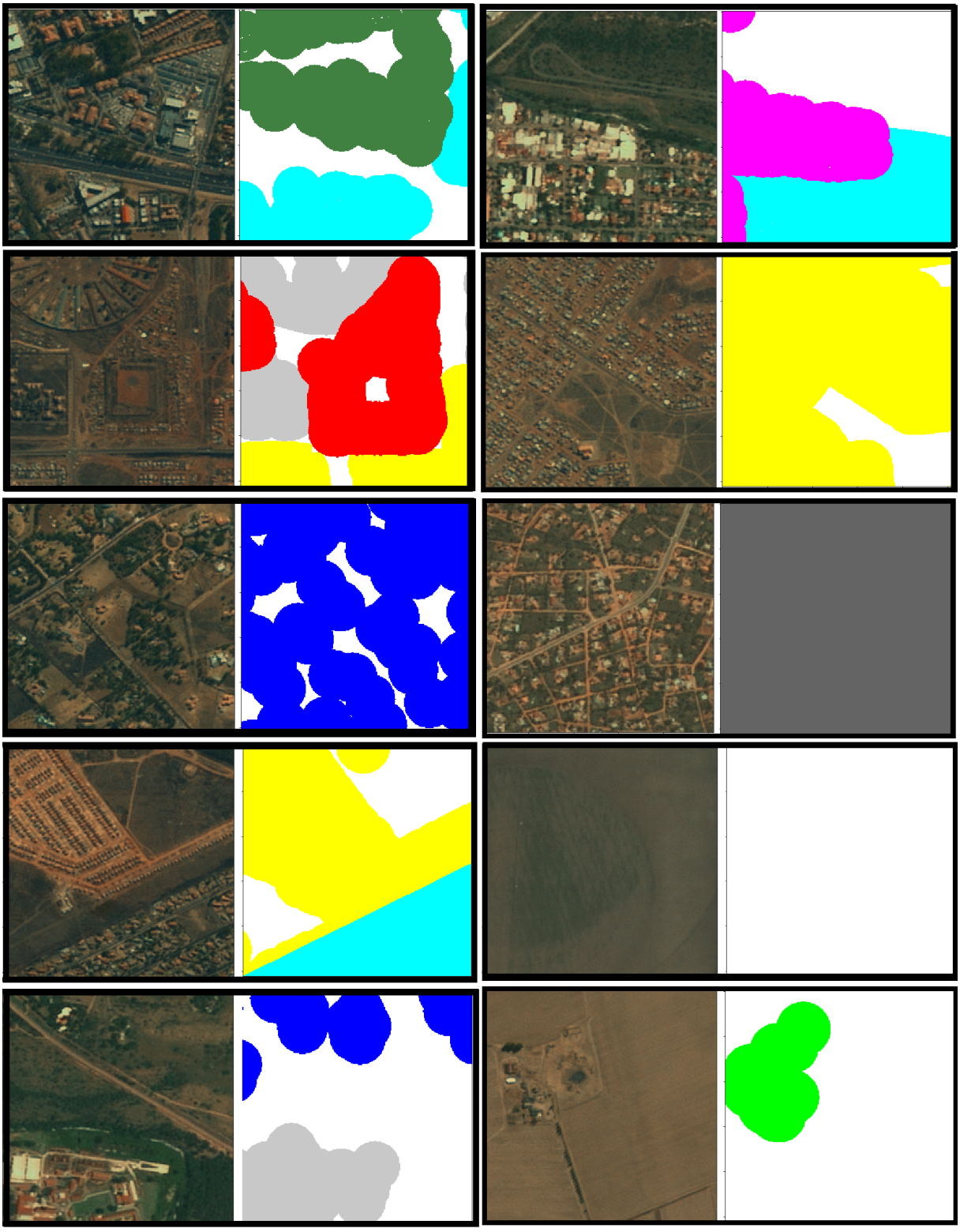


Figure 8: Samples from the final dataset with 12 classes which consists of image and mask pairs. Yellow: township, light blue: suburb, pink: industrial area, olive green: commercial land, red: informal area, light green: farm, light grey: collective living quarters, dark grey: village, blue: smallholdings, magenta: parks and recreational sites, white: background.

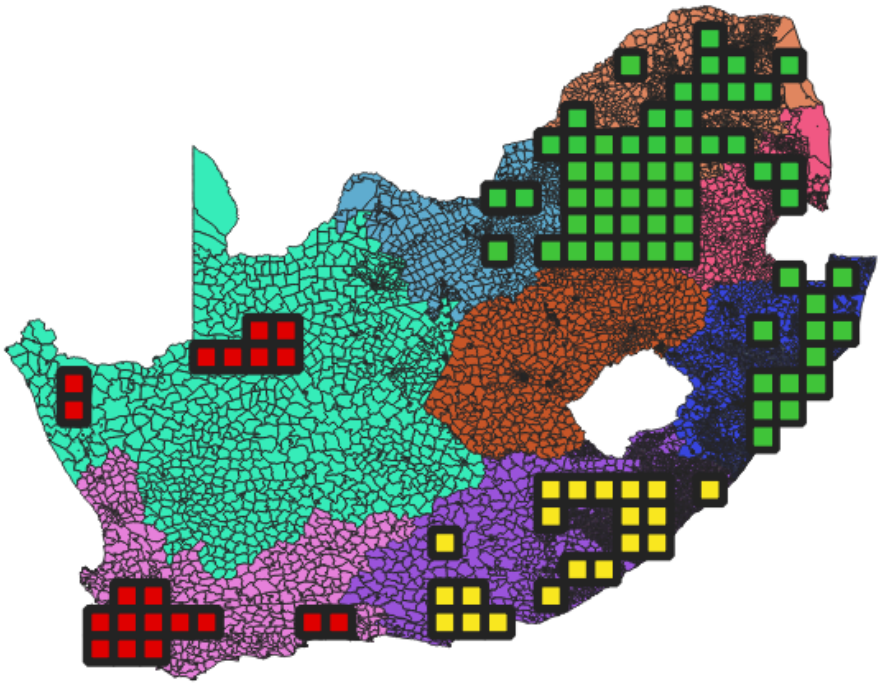


Figure 9: The sampled dataset for training a neighborhood classification model. Each block represents one $21,760 \times 21,760$ satellite image and there is a total of 100 images in this subset. The blocks in red represent images in the test set, those in yellow represent images in the validation set and those in green represent the images in the training set.

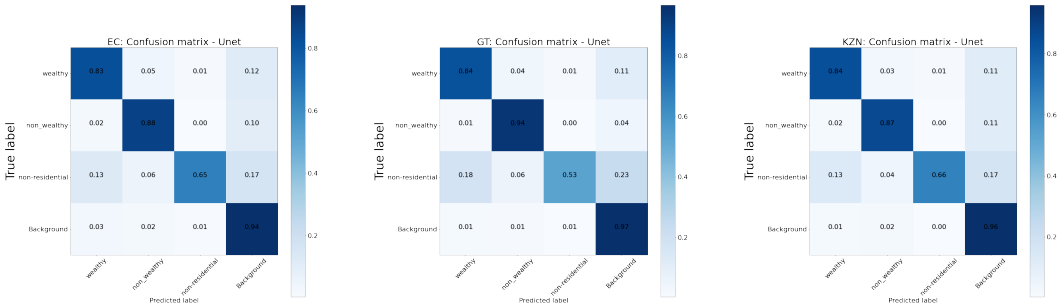


Figure 10: U-Net Confusion matrices per province

A.2.6 Additional Analyses: Differences Between Neighborhood Patterns

Segmenting south African neighborhoods is not an easy task. From visual inspection, we see that the same neighborhood class can have different visual characteristics depending on the province. For instance, Figure 18 shows images from Western Cape and Northern Cape provinces. The neighborhood types enclosed in the blue boundaries are all suburbs. However, the suburbs in the two larger boundaries towards the right have different characteristics from those on the left. The suburb clusters on the left look more like townships because of their smaller yards. And the large yards make those on the right look more like small-holdings (the neighborhood surrounding the suburbs with large yards) or even villages. We have seen this occur in other classes too: some townships in other provinces have even smaller yards and may look like informal settlements. Or they may have larger yards which make them look like suburbs from afar. This variance in the visual characteristics of neighborhoods across the country shows that careful consideration must be given to the construction of the train and test splits of the dataset. Our future work plans to explore visual similarity metrics to inspect which cities have similar visual characteristics.

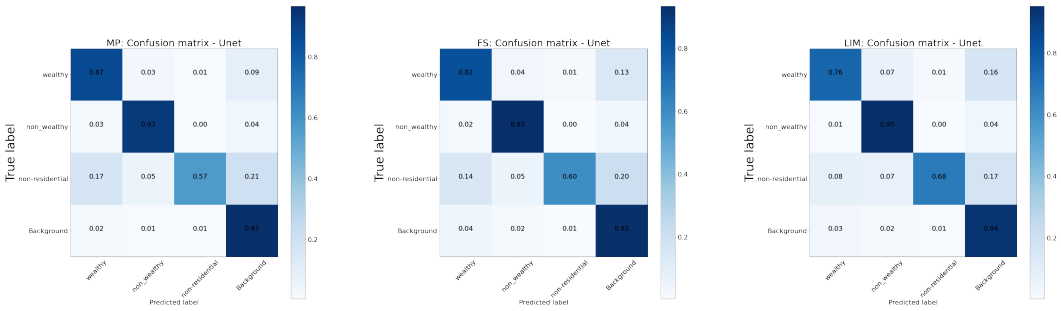


Figure 11: U-Net Confusion matrices per province

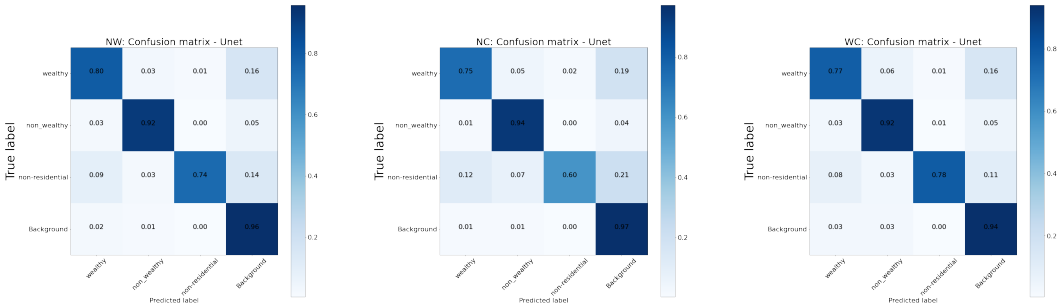


Figure 12: U-Net Confusion matrices per province

From the confusion matrices, we can see that the DeepLabV3+ model, in particular, misclassifies a high number of background images to other classes. Figure 19 shows examples of cases where the U-Net model correctly classifies images depicting vacant land, while the DeepLabV3+ model was unable to do so. Looking at its predictions, it seems to be sensitive to edges that may look like road networks or patterns related to villages.

Figure 20 shows other failure cases. The first row, for instance, shows that remote non-wealthy neighborhoods (a village in this case) with a sparse building density could be confused with wealthy neighborhoods such as farmlands/smallholdings because of the open land surrounding the buildings. From row 2 and 3 we can see that some farmlands with large roofs can be confused with industrial zones, and industrial zones surrounded by open farm-like land can be confused with farm yards. This is inevitable given that industrial zones are typically away from residential places and they are typically on large land plot sizes. The last two rows show examples of false positives mainly for industrial-like activities or non-wealthy neighborhoods. For row 4, mines are also included as industrial zones, and as such open-cast mining activities without buildings would also be expected to register as a positive by the model.

A.2.7 Additional Details on Methodology: Changes in South African Neighborhoods

Here we give additional details of our methodology to compute differences in image masks between our estimates for neighborhood types in 2017 and groundtruth for 2011.

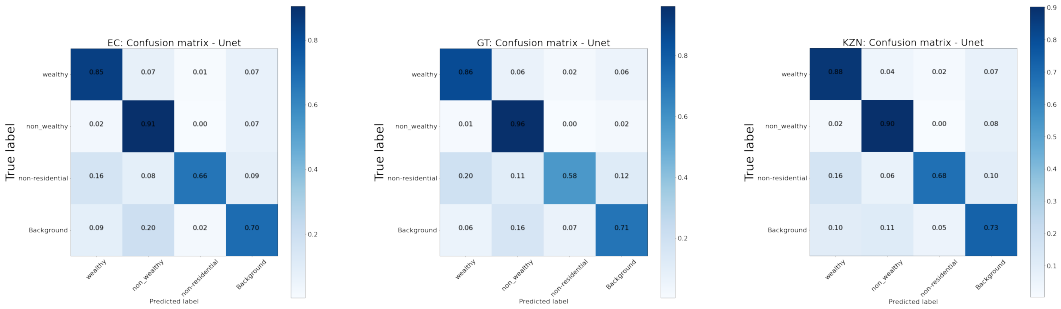


Figure 13: DeepLabV3+ Confusion matrices per province

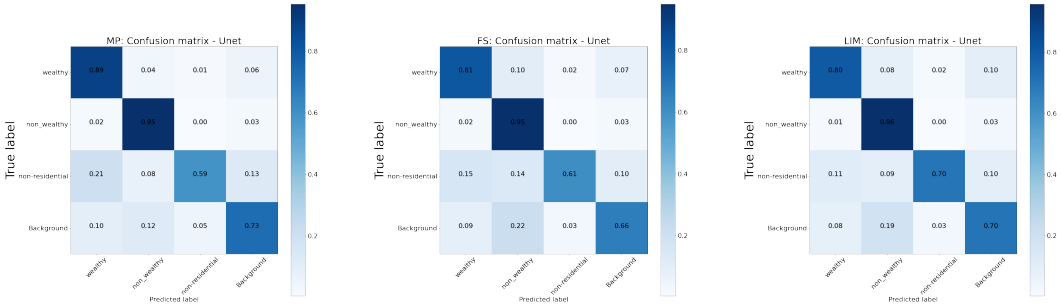


Figure 14: DeepLabV3+ Confusion matrices per province

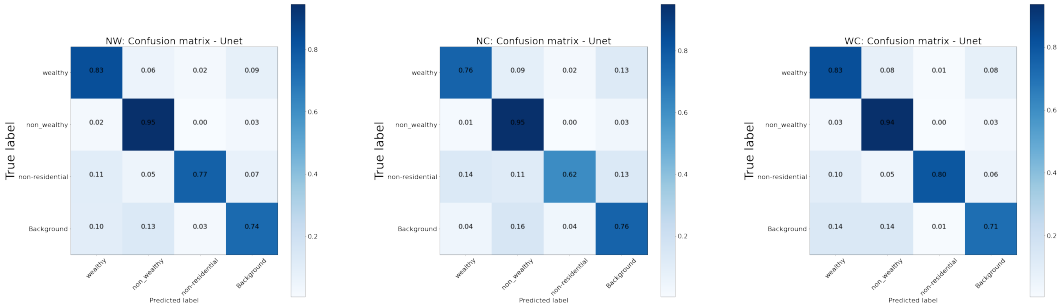


Figure 15: DeepLabV3+ Confusion matrices per province

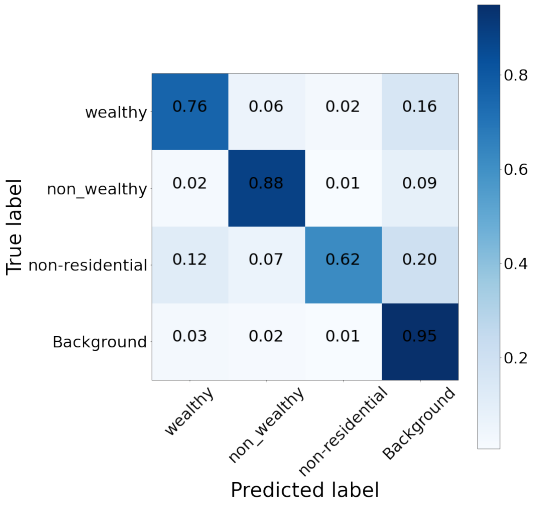


Figure 16: Confusion matrix for the U-Net model trained on the sampled dataset with a 60-20-20 train-test-validation split.

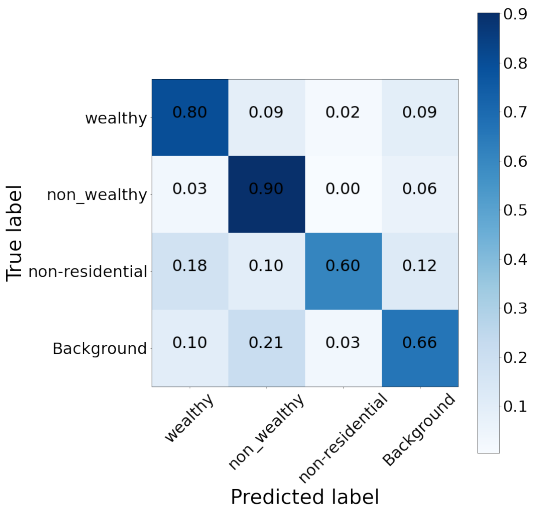


Figure 17: Confusion matrix for the DeepLabV3+ model trained on the sampled dataset with the 60-20-20 train-test-validation split.

Figure 18: The suburb neighborhoods enclosed in the blue boundary show how these neighborhood categories can have varying sub-types.

Once we convert our masks to grayscale, each class label is associated with a unique value as shown in Table 4. Similarly, differences between each type of class also result in unique values (Table 5).

Label	**RGB**	**Grayscale**	**Colour**
Wealthy(W)	[124,124,124]	124	Dark grey
Non wealthy(NW)	[201,201,201]	201	Light grey
nonresidential(NR)	[7,7,7]	7	Black
Background(B)	[255,255,255]	255	White

Table 4: Pixel values per label at RGB and Grayscale.

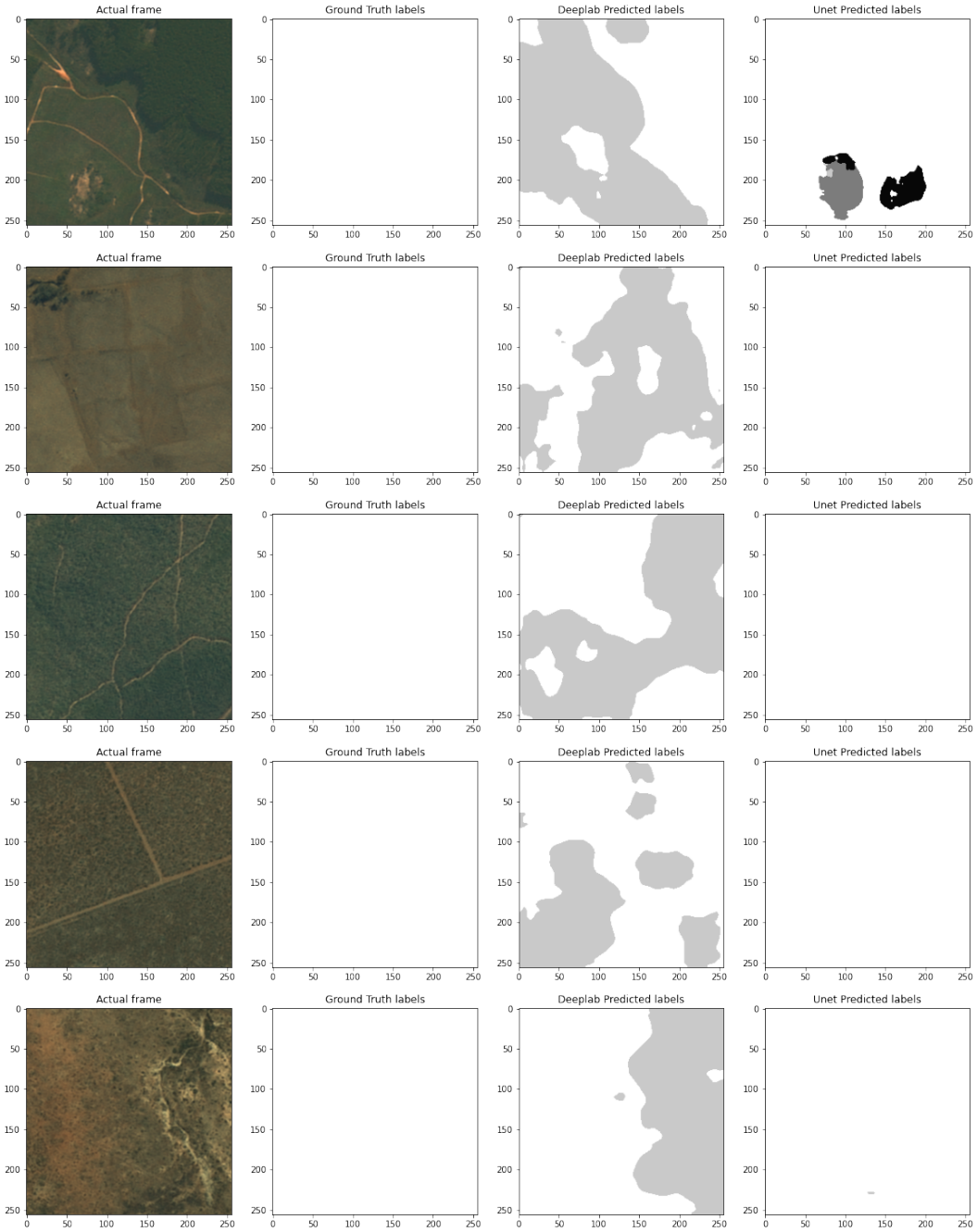


Figure 19: Some common DeepLabV3+ failure cases. Left to right: Input image, ground truth, DeeplabV3+ predicted labels, U-Net predicted labels. These results show that the DeepLabV3+ based model could be sensitive to things that look like edges and associates them with non-wealthy neighborhoods, perhaps because sparsely populated villages tend to have road networks. Light gray: Non wealthy neighborhoods, Dark gray: wealthy neighborhoods, black:nonresidential neighborhoods, white: background.

Once we take an image difference between the 2017 and 2011 masks, we blur the resulting difference image using a Gaussian kernel of size (5x5), then threshold the output with a threshold of 125 to reduce noise. The threshold is determined through trial and error.

After these steps, we store cluster centroids and the corresponding area of each cluster. A cluster is a collection of pixels that are adjacent to each other with the same color. For each cluster of pixels in

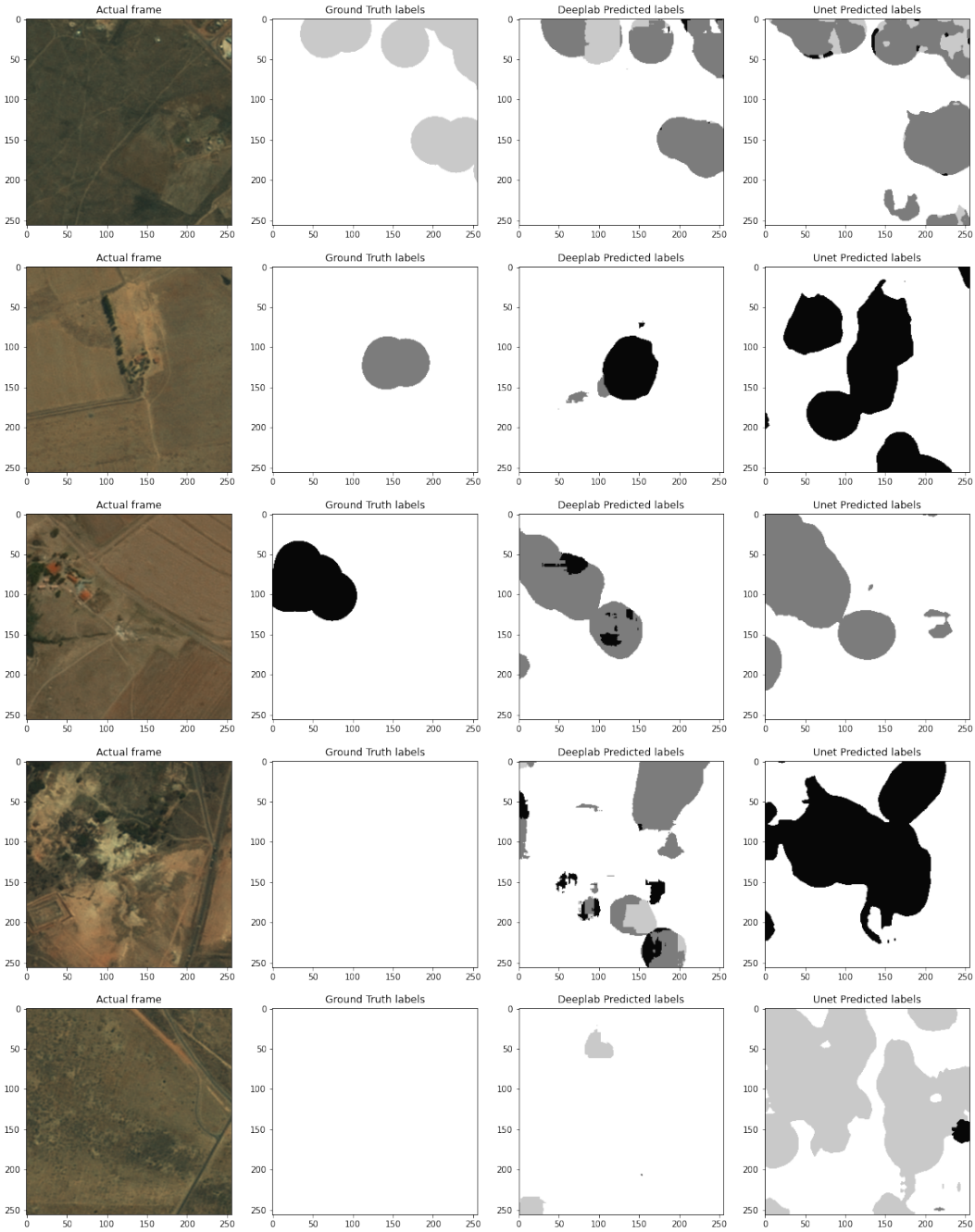


Figure 20: Additional examples of failure cases. Left to right: Input image, ground truth, DeeplabV3+ predicted labels, U-Net predicted labels. Light gray: non wealthy neighborhoods, dark gray: wealthy neighborhoods, black:nonresidential neighborhoods, white: background.

	W(124)	NW(201)	NR(7)	B(255)
W(124)	0	179	117	125
NW(201)	77	0	194	202
NR(7)	139	62	0	8
B(255)	131	54	248	0

Table 5: Image difference lookup table

the difference image, we find the cluster centroid and the corresponding area of pixels per cluster. We then store these attributes in a table.

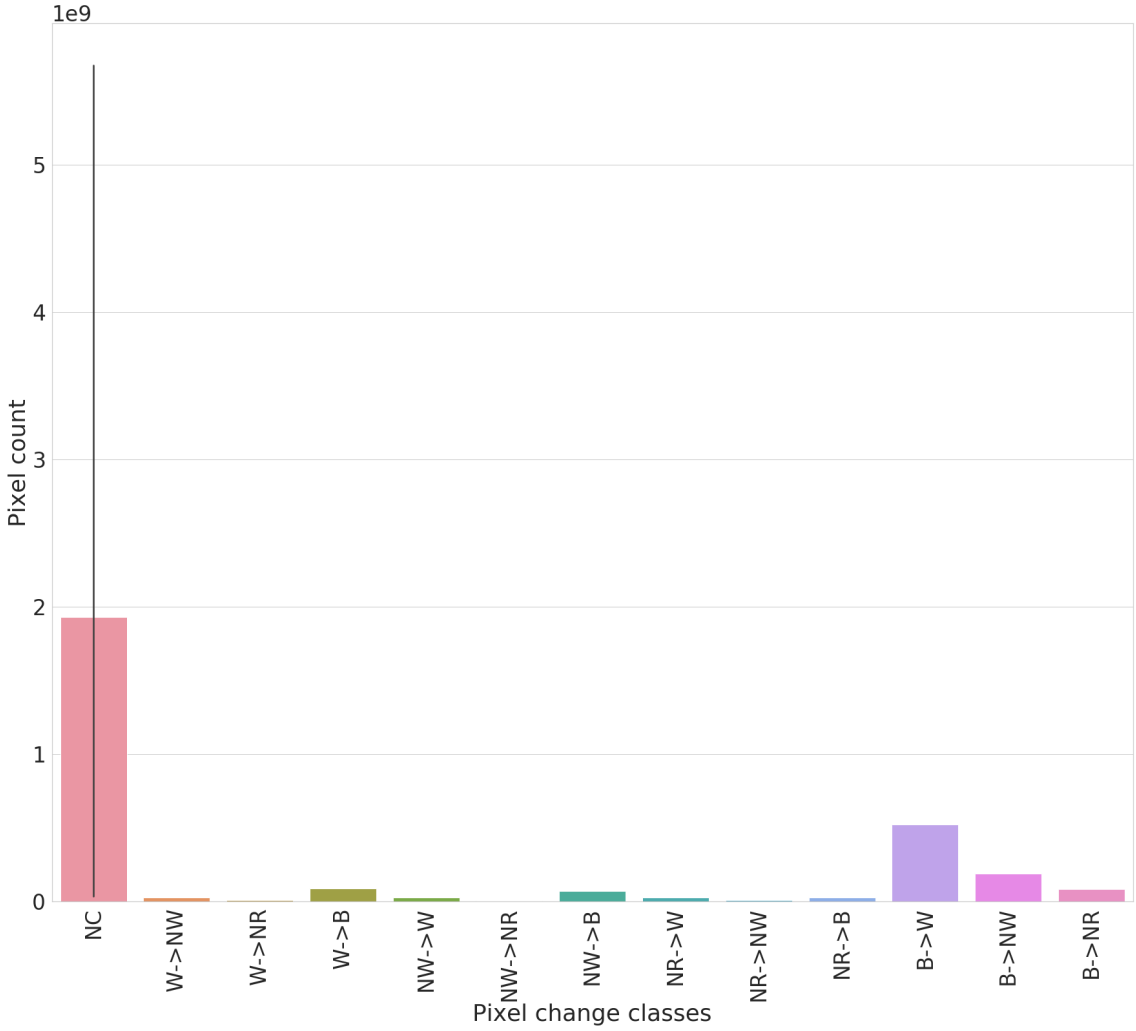


Figure 21: Histogram of pixels and the estimated class change they underwent between 2011 and 2017. NC=No change, W=wealthy, NW=non wealthy, NR=nonresidential B=background. Arrows represent direction of change (e.g. W->NW shows the number of pixels that depicted wealthy neighborhoods in 2011 but we estimate to have changed to non wealthy neighborhoods in 2017.

Finally, we have a table with these attributes: index of cluster, x-y coordinates of cluster centroid, area size of cluster, the grey scale color of cluster and the image filename the cluster belongs to. Given that all the images are geo-referenced (they are smaller tiles of larger satellite images), we are able to overlay the EA dataset from Section 3 and then associate the changes with specific neighborhoods, municipalities, main places, provinces and districts from 2011.

A.2.8 Additional Results: Changes in South African Neighborhoods

Our goal, moving forward, is to perform more experiments using our dataset across multiple years to understand how neighborhoods are changing and perform this analysis across the entire country of South Africa. Here, we discuss a few preliminary results. Figures 21 and 22 show histograms of the calculated change in the number of pixels representing each neighborhood type. According to our estimate, the overwhelming number of pixels are associated with the same neighborhood in 2017 as they were in 2011. After that, the 2^{nd} largest change is in the construction of wealthy neighborhoods (B->W or pixels representing the background class changing to wealthy neighborhoods).

Figure 23 shows time-lapse images of a wealthy neighborhood next to Johannesburg from our satellite image repository. The neighborhood circled in red looks like a smallholding neighborhood type because of the sparsity of buildings and large plot sizes. However, this neighborhood seems to change

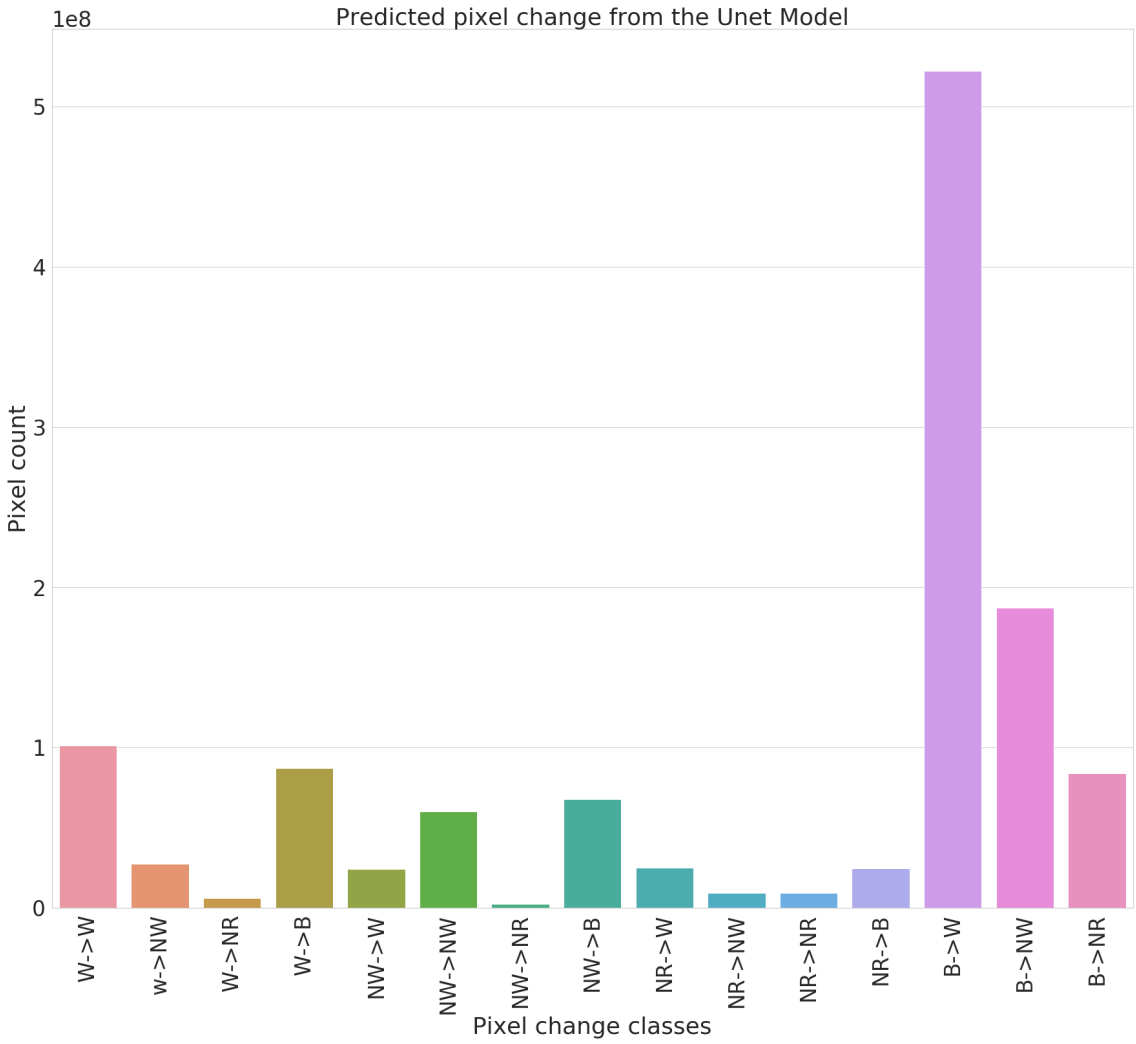


Figure 22: Histogram of pixels and the estimated class change they underwent between 2011 and 2017, excluding the pixels that underwent no change. NC=No change, W=wealthy, NW=non wealthy, NR=nonresidential B=background. Arrows represent direction of change (e.g. W->NW shows the number of pixels that depicted wealthy neighborhoods in 2011 but we estimate to have changed to non wealthy neighborhoods in 2017.

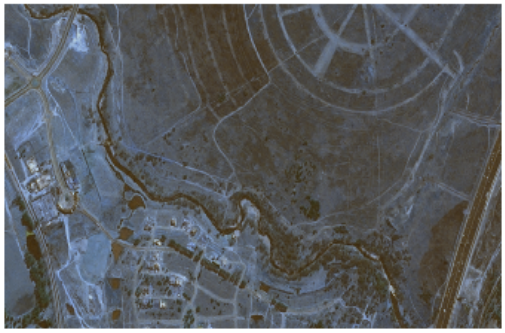
into what appears to be a suburb over time with a higher density of houses and smaller plot sizes. As shown in Figure 24, our U-Net model trained on Gauteng was able to correctly identify the growth of this neighborhood between 2011 and 2017. Figure 26 shows additional examples of changes that have been detected.

Figure 23: These are time-lapsed images from our satellite image repository of a wealthy neighborhood next to Johannesburg. The structure at the top-right of the image developing over time is the biggest mall in Africa (Mall of Africa) and around it developing is a wealthy neighborhood. The neighborhood circled in red looks like a smallholding neighborhood type because of the sparsity of buildings and bigger plot sizes, however, this neighborhood seems to change into what appears to be a suburb over time with a higher density of houses and smaller plot sizes.

A.2.9 Sources of Error: Changes in South African Neighborhoods

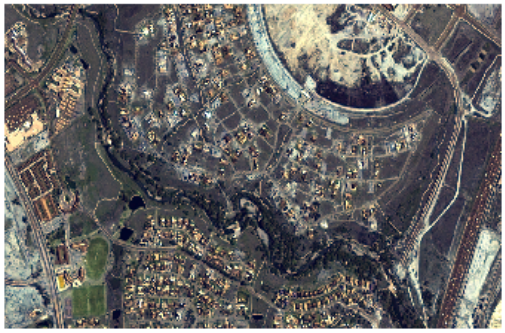
While some of the overall trends we have estimated are supported by other studies, we also see failure cases. Querying the growth of Soweto, one of South Africa's largest townships, using our method, we see results like those in Figure 25. The model is able to detect the extent of growth relatively well but it is unclear if the classification of Soweto as a wealthy neighborhood is accurate. While we have labeled Soweto as a non wealthy neighborhood in our 2011 dataset, the township has started to have many wealthy households, as discussed in [2]. We plan to investigate these results further. In

[2] https://www.nytimes.com/2019/09/29/world/africa/soweto-south-africa-inequality.html

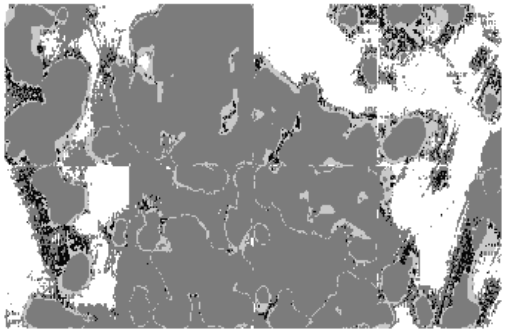


(a) 2011 Image

(b) 2011 Ground Truth



(c) 2017 Image



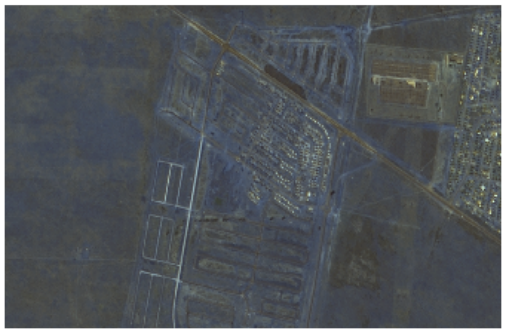
(d) 2017 Predictions

Figure 24: From (a) - (d): 2011 satellite image, 2011 ground truth, 2017 satellite image, 2017 prediction from U-Net model. Examples of the change detected on the 2017 images in a wealthy neighborhood near a big mall. Light gray: non wealthy neighborhoods, dark gray: wealthy neighborhoods, black:nonresidential neighborhoods, white: background.

addition, the difference in satellite image resolution between the 2011 and 2017 images is also a potential source of error. Even after downsampling, the 2017 images appear clearer than the 2011 images which the model was trained on.

Other sources of error are confusions between neighborhood types (e.g. nonresidential neighborhoods and wealthy neighborhoods), that we have seen, and the various sources of error discussed in Sections 3 and 4, pertaining to the dataset construction process.

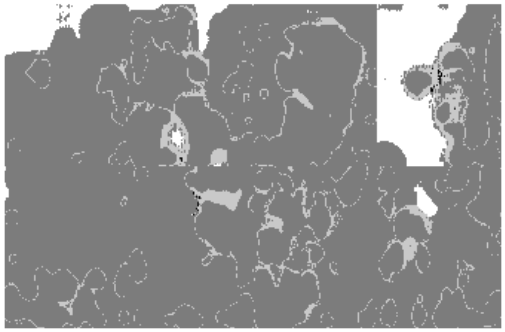
Thus, much more work remains in quantifying our sources of error and refining our methodology to analyze South African neighborhoods using satellite imagery.



(a) 2011 Image

(b) 2011 Ground Truth

(c) 2017 Image



(d) 2017 Predictions

Figure 25: From (a) - (d): 2011 satellite image, 2011 ground truth, 2017 satellite image, 2017 prediction from U-Net model. Examples of the change detected between 2011 and 2017 images of Soweto, one of South Africa's oldest townships. Light gray: non wealthy neighborhoods, dark gray: wealthy neighborhoods, black: nonresidential neighborhoods, white: background.

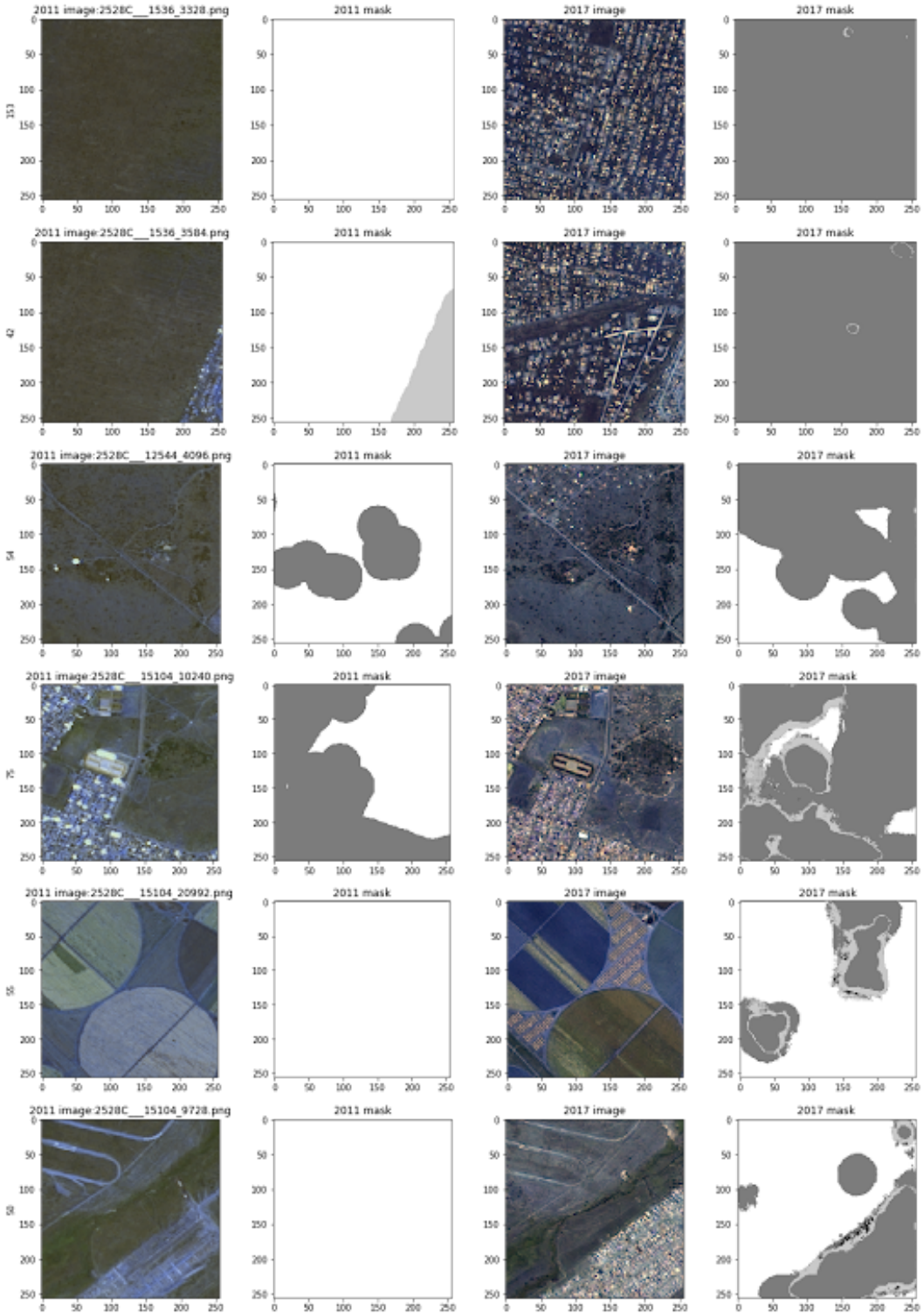


Figure 26: From left-right: 2011 satellite image, 2011 ground truth, 2017 satellite image, 2017 prediction from U-Net model. Examples of the change detected between 2011 and 2017 images.

Datasheet for Visual Dataset to Study the Effects of Spatial Apartheid in South Africa

Motivation

1. **For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.**

The dataset was created for Classifying neighbourhoods in South Africa According to 4 neighbourhood types: wealthy (a combination of suburbs, smallholdings, farms), non wealthy (combination of townships, informal areas, collective living quarters, villages), non residential areas (combination of i industrial areas, commercial lands, parks and recreational areas, vacant land) and background (all land that does not have a building on it). The disaggregated labels (12 classes rather than 4) are also available with the dataset. The specific application the authors created this dataset for is to enable researchers and policymakers to quantify the effects of spatial apartheid over time, for the specific purpose of helping to uncover and working to reverse its effects.

2. **Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**

Raesetje Sefala (University of the Witwatersrand), Timnit Gebru (DAIR), Nyalleng Moorosi (Google) and Richard Klein (University of the Witwatersrand) created this dataset. The project was conceived in 2017 when Nyalleng Moorosi was at the Council for Scientific and Industrial Research (CSIR), South Africa and Timnit Gebru was at Microsoft Research in the USA. The dataset was created using a combination of other pre-existing datasets: The Enumeration Areas (EAs) dataset created in 2011 by Statistics South Africa (Stats SA)--a government agency responsible for conducting the census; Geographically referenced (Geo-referenced) buildings dataset created by Eskom (a South African electricity public utility company) in partnership with the Council for Scientific and Industrial Research (CSIR) consisting of building count data in South Africa from 2006 to 2016; and satellite images from 2006-2017 from the South African National Space Agency (SANSA).

3. **Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.**

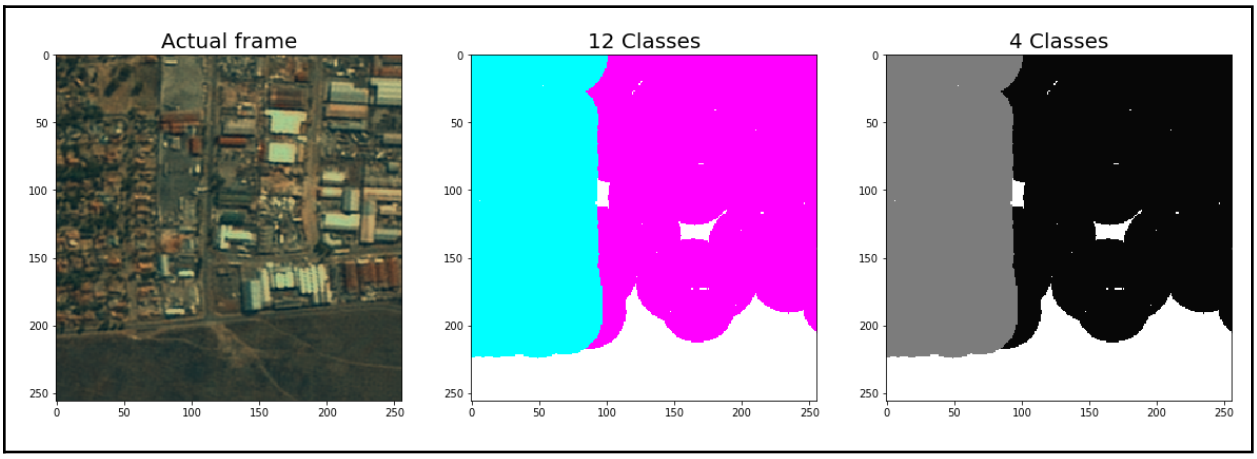
Funding was provided through five sources: The South African Department of Science and Technology and CSIR (as a masters scholarship award to Raesetje Sefala), Google (research award to Raesetje Sefala and compute credits), the Deep learning Indaba and Nvidia (Nvidia Titan V prize for best poster presentation at the 2018 Deep Learning Indaba summer school), and the DAIR institute.

4. Any other comments?

Composition

5. What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

Each instance is a 256x256 satellite image of South Africa from 2011 paired with masks of the neighbourhood type each building cluster represents. We also include satellite images from other years (2006-2017) without masks. The 256x256 images were obtained by tiling each high-resolution satellite image for ease of processing. Images are taken from the SPOT sensor with varying resolutions in different years: one pixel represents 10 meters on the ground for 2011 and each satellite image consists of 21,688 x 21,688 pixels. Before tiling the satellite images by images of 256x256, we upsampled them to 21,760x21,760 using the GDAL library (https://gdal.org/). This was done in order to make sure that a single satellite image can be fully covered by an integer number of 256x256 images.
Table 1 shows the resolution of satellite images for each year and the number of images per year. There are two sets of masks, one set has 12 neighbourhood classes and the other set has 4 neighbourhood classes. Five example instances are shown in figure 1 below. The 12 and 4 classes are listed in answering question 1.



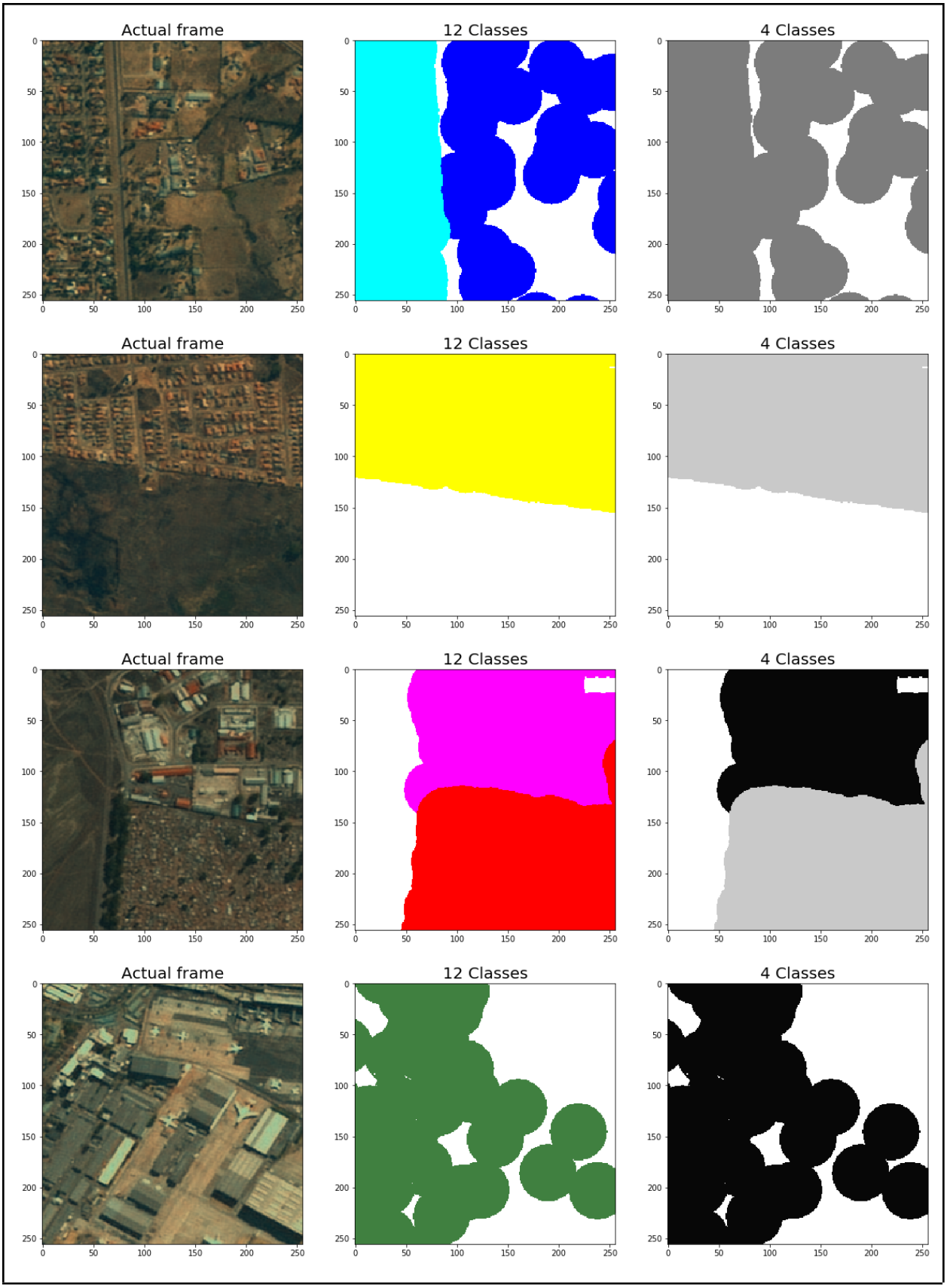

Actual frame
12 Classes
4 Classes
Actual frame
12 Classes
4 Classes
Actual frame
12 Classes
4 Classes
Actual frame
12 Classes
4 Classes

Figure 1: Samples from the final dataset which consists of row instances of satellite image 12 class mask and 4 class mask. For the 12 classes - yellow: township, light blue: suburb, pink: industrial area, olive green: commercial land, red: informal area, lightGreen: farm, light grey: collective living quarters, dark grey: village, blue: smallholdings, white: background. For the 4 classes - black: nonresidential, dark gray: wealthy residential neighbourhoods, light gray: non wealthy residential neighbourhoods and white: background.

6. How many instances are there in total (of each type, if appropriate)?

There are 3,973,750 256x256 satellite images in total from 20211 with associated masks. This corresponds to the 550 satellite images in total, which were originally of resolution 21,688X21,688 and which we upsampled to 21,760x21,760. We also include satellite images from 2006-2017 without labels (a total of 6,218 satellite images).

7. Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

The dataset is not sampled (represents all images of South Africa in 2011). However, given the highly imbalanced nature, and a large amount of vacant land, we perform experiments on a sampled version of our dataset (see associated paper for details).

8. What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)or features? In either case, please provide a description.

Each instance consists of a 256x256 subset of a raw satellite image, a 12-class mask of features and a 4-class mask of features. The satellite images have 3 channels (RGB) and were taken using the SPOT5 Satellite sensor. The 12-class masks are PNG images with 3 channels (RGB), with 12 distinct colours to represent the 12 individual classes (e.g. green[0,255,0] - farm, yellow[255,255,0] - township, white[255,255,255] - background). The 4 class masks are also PNG images with 3 channels and 4 distinct colours to represent the 4 individual classes.

9. Is there a label or target associated with each instance? If so, please provide a description.

For each satellite image, corresponding masks are labelled at the pixel level to represent the neighbourhood type for each pixel on the satellite image.

10. **Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.**

Everything is included. However, if there are missing instances in our source datasets due to errors (e.g. building dataset), our dataset will also miss these instances.

11. **Are relationships between individual instances made explicitly (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.**

Each instance is associated with a latitude and longitude. This allows us to know which physical location each instance represents, and how instances are spatially related to each other.

12. **Are there recommended data splits (e.g., training, development/validation, testing)? If so, please describe these splits, explaining the rationale behind them.**

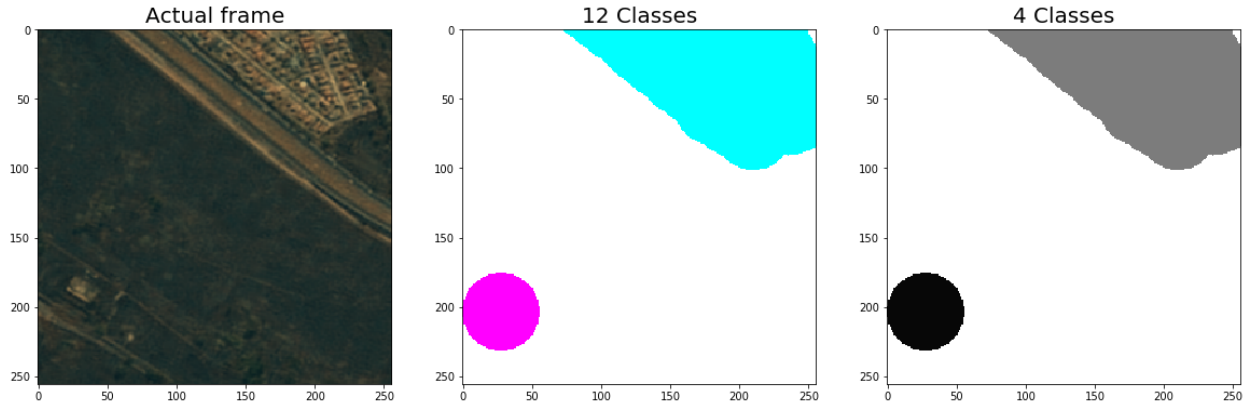
We have performed experiments using a subset of our data and report the train/validation/test split we used in these experiments as part of the dataset metadata. What splits people should use depends on the specific task/application they work on. We recommend ensuring that all images representing the same neighbourhood are in the same split.

13. **Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.**

Yes.

- The masks are created in such a way that each house is represented using a circle with a diameter of 0.0007 decimal degrees irrespective of the type/size of the building. We did that because we did not have a source that captures the exact sizes of the buildings throughout the country consistently across all neighbourhood types (available datasets do not capture informal settlements and some villages consistently). The pink polygon in figure 2 represents an industrial building on the satellite image, the representation is not a tight bound around the extent of the building but instead a circular polygon of diameter 0.0007 decimal degrees over the centroid of

the building. Similarly, that is how we created the rest of the dataset.



- The building dataset has other potential sources of error especially around the labelling of informal settlements. Individual buildings in these neighbourhoods are usually very difficult to distinguish from satellite images. Small buildings camouflaged by the surrounding environment can also be difficult to detect.
- Although we have taken steps to verify that the information is correct, another potential source of error is how we labelled the townships in our dataset. We followed the procedure in section 3 of our associated paper to distinguish suburbs from townships, given that they are both labelled as formal residential neighbourhoods in the EA dataset. As noted in the paper, our labeling process could result in some townships that are labeled as suburbs or vice versa. It is much more likely that we have misclassified some townships as suburbs as Wikipedia may not have all the labels of townships in 2011 and if 2 people label something as a township and we cannot find the name listed as a township in other sources, we classify it as a suburb.
- Additionally, the manner in which we demarcated wealthy and non wealthy neighborhoods can be a source of error. Some collective living quarters that are close to places of economic activity may be wealthy neighborhoods. In addition, while townships were allocated very low budgets during apartheid, there are now wealthy households within townships.

14. Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

Yes, the dataset is self-contained. Although users can substitute the satellite images for other types (more/fewer channels)/pixel resolutions. They will have to match the geographical referencing so that the masks can align properly. The satellite images included in our dataset are from the South African

National Space Agency, and they have permitted us to release the dataset for research purposes only. We have similar permission from Eskom to release the building count dataset for research purposes. The EA dataset is publicly available as part of the South African census data.

12. Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.

Our dataset does not include confidential data and only provides masks at the neighbourhood level. However, the building count dataset which is used in the construction of our data locates each building in South Africa. If an individual lives in a particular building and does not want their house/building to be located with this building dataset, it is unclear if they can do so.

13. Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

No, it does not.

14. Does the dataset relate to people? If not, you may skip the remaining questions in this section.

Yes. While the dataset does not have people, it depicts the types of neighbourhoods in which people live.

15. Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

The dataset indirectly identifies subpopulations by neighbourhood types which can roughly approximate standards of living. The labels were created by Statistics South Africa (a government entity responsible for the census) for the census. We overlaid their dataset with ours so that we can label the buildings according to these neighbourhood types.

16. Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.

No, it is not possible to identify individuals. This dataset labels clusters of buildings according to their type.

17. Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of

government identification, such as social security numbers; criminal history)? If so, please provide a description.

The enumeration area (EA) dataset from the census, which is public, contains demographic data for each Main Place (A Main place is a group of EAs) but we do not include this data as part of the data we are releasing, it was released as part of the 2011 South African census. While the EA dataset already labels EAs by associating them with 12 types of neighborhoods according to the land's intended use, our dataset further approximates the location of building clusters within the EA polygons to distinguish between intended and actual land use.

18. Any other comments?

Collection Process

18. How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

The dataset was created using 3 other datasets as input.

- Satellite images: Captured using the SPOT5 Satellite sensor.
- Enumeration Area dataset: Directly observed and captured during the time they made the 2011 South African census
- Building dataset: systematically collected and cleaned by a group of subjects, this dataset went through various stages of verification, more information about the dataset can be found at https://www.ee.co.za/wp-content/uploads/legacy/PositionIT%202009/PositionIT%202010/SPOT.pdf

Building clusters were inferred using a buffing algorithm of diameter 0.007 degrees. This process can introduce noise as mentioned in answering question 13.

19. What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?

- Satellite images: Captured using the SPOT5 Satellite sensor.
- Enumeration Area dataset: Unknown
- Building dataset: Unknown
- In our dataset processing phase, we used the QGIS software for tasks requiring spatial processing (projections, overlays etc).

20. If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

The sample is of satellite images of all of South Africa.

21. Who was involved in the data collection process (e.g., students, crowd workers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

Source datasets:

- Satellite images: The South African National Space Agency bought the images from a third-party source
- Enumeration Area dataset: We are not sure
- Building dataset: Contractors, we do not know if they were compensated.

In addition to those listed in question 2, we recruited 10 volunteer students who grew up in townships to aid in labeling townships.

22. Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., a recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

Source datasets were collected during these timeframes:

- Satellite images: 2006--2017
- Enumeration Area dataset: The labels are for 2011 but the census verification process lasted until 2013.
- Building dataset: 2006-2017

Our ground truth data creation process was done between 2018 and 2020.

23. Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

There were no review processes conducted by a review board. However, see answer to Q29 on an analysis of potential risks and harms.

24. Does the dataset relate to people? If not, you may skip the remaining questions in this section.

The data relates to people in that it captures the characteristics of neighbourhoods (clusters of buildings) in which people live. Our generated masks are masks of neighborhoods in aggregate.

25. Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

The dataset was obtained via other sources listed in question 2 (South African Space Agency, Eskom, and Statistics South Africa).

26. Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

For the following source datasets used in our annotations,

- Satellite images: No: the data was collected using satellites.
- Enumeration Area dataset: Created under the South African 2011 census project.
- Building dataset: To our knowledge building occupants were not notified that a building dataset of all buildings in South Africa was constructed and that the building they occupy is in the dataset.

27. Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

Yes, for the EA dataset as this was done under the South African census project and is publicly available data. Stats SA is mandated to provide the state with information about the economic, demographic, social and environmental situation in the country. This is in line with the Statistics Act, (Act No. 6 of 1999), and the fundamental principles of official statistics of the United Nations. Legally, Section 16 of the Statistics Act (Act 6 of 1999) obliges a respondent to answer all questions put to them by an officer of Statistics South Africa. Section 17 of the Statistics Act guarantees the confidentiality of your information. The data collected is used for statistical purposes only and no-one can access data on an individual level.

28. If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

See answer for Question 27. Beyond that we do not know if individuals have a mechanism to revoke their consent for the collection of the EA dataset. The data collected is used for statistical purposes only and no-one can access data on an individual level.

29. Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

Our analysis consisted of speaking to various stakeholders and incorporating their feedback. Some of the researchers working on this dataset also grew up in townships and have seen various manners in which data driven systems can marginalize people in South Africa. This led us to believe that our dataset should only be available for research, rather than commercial, purposes. Our dataset is also only available through requests, rather than on a website where it can be downloaded by anyone. This allows us to grant access only to those who request it for uses endorsed by us.

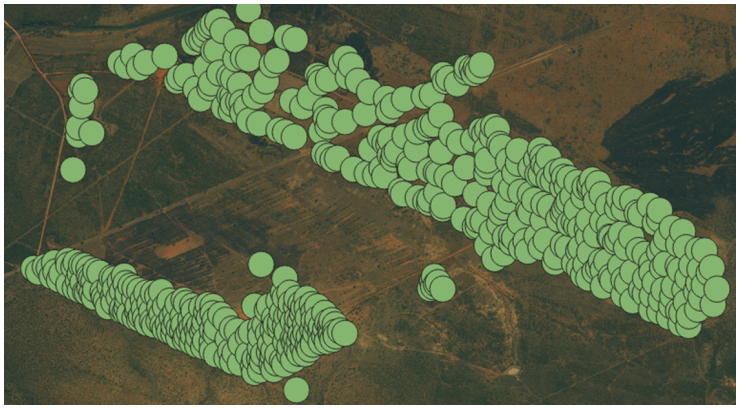
30. Any other comments?

Preprocessing/cleaning/labeling

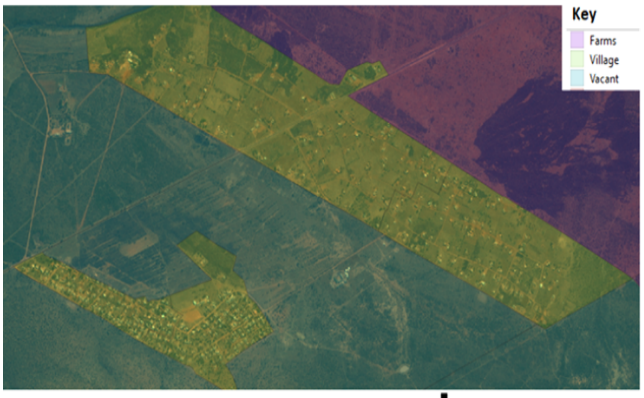
31. Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remainder of the questions in this section.

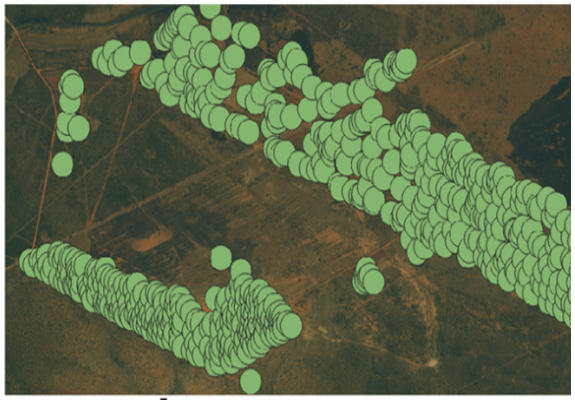
Our associated paper and supplementary materials describe in detail how the dataset was acquired and processed. In short, we started with centroids of building locations, polygons with labels denoting land use as mandated by the government, and satellite images of South Africa, we performed the following steps;

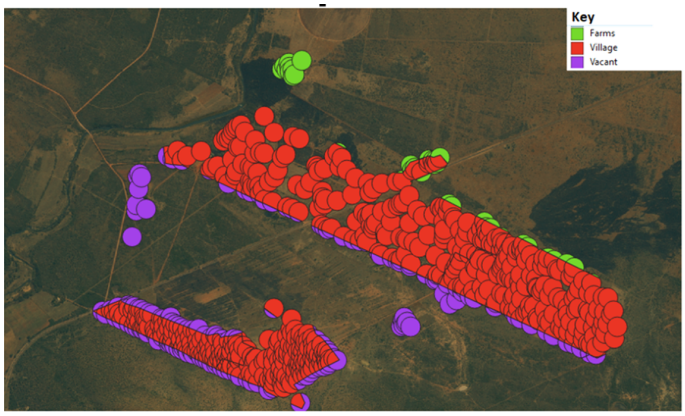
- We inflated the building centroids into polygons as shown in figure 3 and 4 to cover the houses.



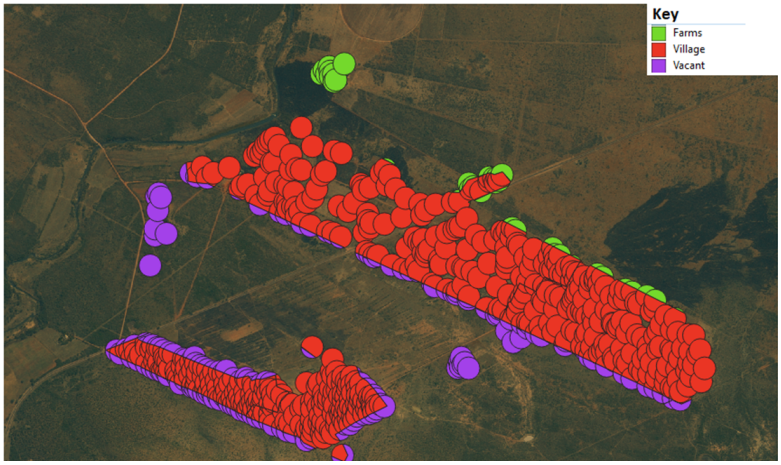
- Intersect the inflated building polygon data with the polygons denoting land use

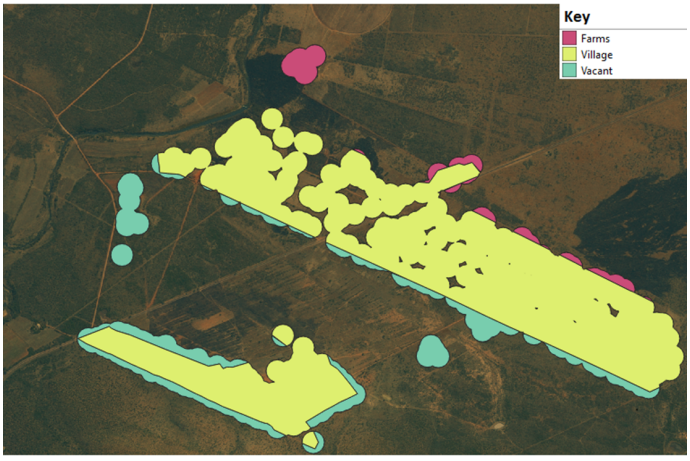






- And smoothed overlapping building polygons by neighbourhood type





Given the way we created the dataset, any building that was not captured (missing data) in the building dataset will al
not be represented in our dataset.

32. Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

Yes. Raw data from all sources (building count data, satellite images, EA dataset, has been saved).

33. Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point.

We are making the code we used to process the data available with the dataset.

34. Any other comments?

Uses

35. Has the dataset been used for any tasks already? If so, please provide a description.

The dataset has been used to perform experiments in section 5 of the associated paper related to neighborhood classification in South Africa including:

- Training a model on 8 provinces and testing on the 9th to investigate the visual similarity between provinces.
- Investigating if our dataset can be used to detect the evolution of neighborhoods in South Africa between 2011 and 2017. For instance, what types of neighborhoods were built on vacant land by 2017 that did not exist in 2011?

36. Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

We plan to create such a repository and reach out to people who request the dataset to update us with the paper/task they have used it for. We will update the datasheet with the location of the repository.

37. What (other) tasks could the dataset be used for?

The dataset can be used to experiment with semantic segmentation more generally. It can also be merged with other existing datasets to make inferences about the standard of living in various South African neighborhoods, and the characteristics of people who live in these neighborhoods (using census data). Insurance, bank and other types of companies have, in the past, used these types of datasets to help them make predictions about the type of loans people can receive or the types of insurance groups can receive. Many of these practices have been found to be discriminatory so we do not allow for our dataset to be used in those scenarios.

38. Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks). If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?

There are entities that discriminate against people based on their zip codes in many countries including South Africa. Also see answers to question 39 below.

39. Are there tasks for which the dataset should not be used? If so, please provide a description.

While our dataset does not identify individuals, we know that there are cases of entities using information about neighborhoods, and linking them to data they have on individuals to make inferences about them in ways that are discriminatory. We plan to screen for uses of our dataset by asking people to fill out a request form to obtain it, including what they plan to do with the dataset, asking for an update on what the dataset was used for and tracking it in our repository. We will not accept use cases that:

- Enable harassment, threatening, intimidating, predatory or stalking conduct;
- Determine financial consequences such as interest rates, insurance prices or loans;
- Pertain to the military or aid in drone targeting;
- Have commercial deployment. This dataset is to be used for strictly research purposes.

40. Any other comments?

Distribution

41. Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

The dataset will only be available for academic research use. It will be available via this request form.

42. How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

The dataset will be hosted on a Google Cloud Platform in a Bucket and will be available based on requests on https://forms.gle/x6YmS96VVPgsUSiQ6. The dataset will be associated with a DOI upon release which will be added to the datasheet.

43. When will the dataset be distributed?

The dataset will be available for release on December 1, 2021.

44. Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

This dataset is freely available for academic and non-academic entities to use for non-commercial purposes such as academic research, teaching, scientific publications, or personal experimentation. Permission is granted to use the data given that users agree the terms below.

1. Users should include a reference to the dataset in any work that makes use of the dataset. For research papers, cite our preferred publication as listed on our website; for other media cite our preferred publication as listed on our website or link to the dataset website.
2. Subject to compliance with these terms, users are granted a limited, non-exclusive, non-transferable, non-sublicensable, revocable license to access and use the dataset.
3. Users should not distribute this dataset or modified versions. It is permissible to distribute derivative works in as far as they are abstract representations of this dataset (such as models trained on it or additional annotations that do not directly include any of our data), given that the use cases are in line with those listed in this datasheet.
4. The dataset or any derivative work may not be used for commercial or military purposes as, for example, licensing or selling the data, or using the data with a purpose to procure a commercial or military gain.
5. That all rights not expressly granted to the users are reserved by us (Authors).

45. Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

The South African National Space Agency and Eskom have given us permission to distribute the satellite and building count datasets respectively, for research use.

46. Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

No.

47. Any other comments?

Maintenance

48. Who is supporting/hosting/maintaining the dataset?

Raesetje Sefala is supporting/maintaining the dataset.

49. How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

Email them at sa.spatialproject@gmail.com.

50. Is there an erratum? If so, please provide a link or other access point.

This is the first version of the dataset release. Any changes/updates will be posted on the dataset's official website (link to be added to the datasheet with official release). The changes/updates will also be emailed to those who received the dataset through an official datast request.

51. Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances')? If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)?

If we do find errors or other information that should be corrected, we will update the dataset accordingly and post the update on the dataset webpage as well as email registered users of the dataset.

52. If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.

N/A.

53. Will older versions of the dataset continue to be supported/hosted/maintained?If so, please describe how. If not, please describe how its obsolescence will be communicated to users.

Older versions will be kept and maintained for consistency even if newer versions are released.

54. If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description.

Those who would like to make contributions can request to use the dataset like others and explain what they plan to do. We plan to regularly poll users to update our repository of dataset use. If they release derivative datasets which require accessing our dataset, we require that our dataset still be accessed via our form so that we can track what it is used for.

55. Any other comments?