AN INFORMAL ROAD DETECTION NEURAL NETWORK FOR SOCIETAL IMPACT IN DEVELOPING COUNTRIES

Inger Fabris-Rotelli[a][*]   Abraham Wannenburg[a]   Gao Maribe[a]   Renate Thiede[a]   Marie Vogel[a]
Mila Coetzee[a]   Kutloano Sethaelo[a]   Ephent Selahle[a]   Pravesh Debba[c,d]
Victoria Rautenbach[b]

[a]Department of Statistics, University of Pretoria, Pretoria, South Africa;
[b]Department of Geography, Geoinformatics and Meteorology, Univeristy of Pretoria, South Africa;
[c]Council for Scientific and Industrial Research, Pretoria, South Africa;
[d]Department of Statistics and Actuarial Science, University of Witwatersrand Johannesburg, South Africa

KEY WORDS: deep learning, informal roads, road extraction, neural networks, South Africa

ABSTRACT:

Roads found in informal settlements arise out of convenience, and are often not recorded or maintained by authorities. This complicates service delivery, sustainable development and crisis mitigation, including management and tracking of COVID-19. We, therefore, aim to extract informal roads in remote sensing images. Existing techniques aiming at the extraction of formal roads are not suitable for the problem due to the complex physical and spectral properties of informal roads. The only existing approaches for informal roads, namely (Nobrega et al., 2006, Thiede et al., 2020), do not consider neural networks as a solution. Neural networks show promise in overcoming these complexities. However, they require a large amount of data to learn, which is currently not available due to the expensive and time-consuming nature of collecting such data. This paper implements a neural network to extract informal roads from a data set digitised by this research group. Data quality is assessed by calculating validity completeness, homogeneity and the V-measure, a measure of consistency, in order to evaluate the overall usability of the dataset for neural network informal road detection. We implement the GANs-UNet model that obtained the highest F1-score in a 2020 review paper (Abdollahi et al., 2020) on the state-of-the-art deep learning models used to extract formal roads. The results indicate that the model is able to extract informal roads successfully in the presence of appropriate training data.

1. INTRODUCTION

The COVID-19 pandemic forced the world to take keen interest in high risk areas and how to prevent them from becoming hot spots (Corburn et al., 2020). Informal settlements carry high risk due to the close proximity of dwellings and overcrowding caused by the lack of adequate and secure housing (Corburn et al., 2020). Most residents have to travel out of the informal settlement for school, work or health services (Iimi et al., 2016). An estimated third of all city dwellers worldwide lived in informal settlements in 2007, and these numbers have been rising ever since due to rapid urbanisation (UN Habitat, 2006). Due to their unplanned nature basic services and infrastructure are often unavailable(UN Habitat, 2015). In order to ensure basic human dignity and stimulate economic activity, governments aim to upgrade and develop these areas. South Africa has several such development plans in accordance with the United Nations' Sustainable Development Goal 9 on Innovation, Industry and Infrastructure[1], for instance project Tirane in Tshwane[2].

Naturally, a need arises to expand and develop transportation infrastructure when there is an increase in population (Zhao et al., 2016, Rui, 2013). Governments invest heavily in the development and expansion of transportation infrastructure to prevent traffic congestion, overcrowding and improve the ease of transporting goods and services (Zhao et al., 2016). Informal settlements also face the problem of congestion and overcrowding when their population increases, forcing these settlements to grow and develop new transportation networks. This means that governments are not regularly investing into road networks for informal settlements. The road network in informal settlements are informal roads or a combination of formal and informal roads. Informal roads are created without government approval or knowledge and are not maintained by government, nor do they appear in any official databases or maps. Residents create these roads to accommodate their need for efficient transport by walking, cycling or driving to points of interest. This presents a problem for service delivery, medical services, criminal activity prevention and environmental emergencies, since the relevant teams will not know the most efficient route to the destination. Information on road networks in informal settlements will also make it easier to plan the best placement for emergency support facilities in informal settlements. In the South African context this is a great concern since the population in informal settlements is growing at a faster pace than elsewhere (Runsten et al., 2018). The pace of population growth also causes the informal roads to change more rapidly. Therefore, real-time road data is critical for service delivery, emergency responses and crisis mitigation measures. An automatic road extraction model from remote sensing imagery stands as a solution.

Neural networks show promise for application to informal road detection. Not only have neural networks been proved to be the most accurate method of road detection to date (Cheng et al., 2017, Kirthika and Mookambiga, 2011, Mendes et al., 2016),

---

but since formal and informal roads are both linear objects with shape-defining boundaries, the same method of object classification using neural networks can be used for informal road detection (Thiede et al., 2020). Object classification is a method of segmenting an input image and determining detailed features such as shape, context and spectral characteristics in order to assign a more informed classification, and it has already been documented to result in more accurate informal road detection than previous methods of per-pixel classification (Nobrega et al., 2006). Informal roads often contain contaminants such as littering, vegetative debris and construction waste that per-pixel classification can easily misclassify, due to both boundary and texture inconsistency. This results in the training dataset incorrectly teaching the neural network what should be labelled as a road (Nobrega et al., 2006). This effect is exacerbated by the presence of vegetation, shadows and bare soil. Bare soil and vegetation encroaching on road boundaries can reasonably be expected near informal roads that lack official maintenance. Bare soil of an unpaved informal road is difficult to distinguish from the bare soil of the adjacent plots of land. Object classification performs better with these complications since additional geometric, spectral and topological parameters help put road heterogeneity into context and reduce misclassifications (Nobrega et al., 2006). This mitigation of heterogeneity further justifies the use of neural networks as a solution for informal road detection.

This paper contributes the first digitised informal road remote sensing data set, as well as a trained neural network for road extraction, specifically improving informal road extraction. A dataset is created that consists of manually labelled aerial photographs in which informal roads are digitised and labelled using GIS software[3].

2. LITERATURE REVIEW

With the advancement of technology and specifically the increase of Volunteered Geographic Information (Goodchild, 2007), the need for quality-controlled data has become increasingly important. Open-source software like OpenStreetMap (Girres and Touya, 2010) allows anyone to create datasets relevant to previously overlooked problems like informal road detection. The volume of newly available data, however, makes diligent supervision and necessary data quality assessment difficult. In order to ensure standardisation and accuracy, crowd-sourced geographical data requires additional modelling and capture agreements (Goodchild, 2007) such as a "three recordings rule" which requires tracking data to first be provided by three independent sources before it is integrated into the dataset. Similar measures could be applied to manually digitalised datasets, but these datasets are still affected by the heterogeneity of geographical data as well as margins of error associated with human input (Goodchild, 2007). Informal road detection requires due diligence in both the data capturing and the data analysis stages. This is because very little ground truth data is available on informal road networks. Therefore, the accuracy of the dataset relies entirely on its adherence to strict standards of data quality.

Currently the mapping of informal roads is an expensive, time consuming manual process. Informal settlements and their associated informal road network can form very rapidly (UN Habitat, 2006). Google relies on official maps obtained from a variety of sources, such as governments and institutions which

are then combined with satellite imagery. Some of these images have spatial information such as roads already overlaid, for example those obtained from the TIGER geodatabase 4 produced by the USA's Census Bureau. TIGER files' spatial data is mapped manually. Google employs hundreds of map technicians to ensure location accuracy, add in missing roads on the images, extract additional information from Google street view images (obtained by driving cars equipped with cameras out to the actual location) and improve on inaccuracies reported by users. Much of the information obtained from Google street view is processed by image recognition algorithms and added to the map, but are still reviewed by humans and improved by user feedback. Some use is also made of anonymised movement data from Android device users to validate roads and traffic movement. Users can add in missing information or make corrections directly in Google mapmaker, but the developers explain that this must also be validated by a Google technician before it is published. In other words, even for a large company in the business of creating software and maps, there is still heavy reliance on manual labour in order to produce a finished product.

Data quality can broadly be defined as fitness for use (Wang and Strong, 1996). When data quality is assessed, we are therefore measuring whether the dataset is reliable enough to be of use in accordance with the criteria of a particular use case. A number of methods to measure accuracy are available. Completeness is a popular measure of data quality for road detection algorithms (Heipke et al., 1997) due to its ease of calculation. This measure was used by (Nobrega et al., 2006, Thiede et al., 2020, Thiede, 2019, Li et al., 2016). Other common performance metrics are F1-score, recall, precision, overall accuracy, correctness, Intersection over Union (IoU) and the kappa coefficient (Abdollahi et al., 2020). The V-measure, adapted by Nowosad and Stepinski (Nowosad and Stepinski, 2018) from the information-theoretical application, is a very useful method of comparing two categorical maps. The output is a single, easily comparable and interpretable value, which incorporates completeness and homogeneity of the categorical regions.

Previous methods on the extraction of informal roads (Nobrega et al., 2006, Thiede et al., 2020) have not used neural networks. This is the only literature available on informal road extraction. The biggest difficulty in using neural networks is the large amount of labelled data required for training. High resolution satellite images are very expensive potentially making deep learning methods impractical for developing countries that do not have their own satellites. Unmanned aerial vehicles (UAVs) have become more popular in image classification like forest mapping (Ruwaimana et al., 2018) and tomato detection (Senthilnath et al., 2016) since they are cheaper alternatives to high resolution satellite imagery. Drone images are cheaper over long-term use and work very well in humid climates (Ruwaimana et al., 2018). Satellites are dependent on external software when accumulating images, where UAVs use an embedded camera and a map of the environment making it a better solution when GPS fails (Mantelli et al., 2019). Aerial photography was made use of in our research.

Neural networks aim to simulate the visual capabilities humans have and teaching an algorithm to have these same capabilities. (Wang et al., 2016) reviewed 30 years' worth of road extraction methods. These methods were classified into mathematical morphology, active contour models, dynamic programming, classification-based methods and knowledge-based methods. The classification-based methods are further sub-divided

---

[3] The research was approved by under ethics number NAS200/2021.

into supervised and unsupervised learning, where neural networks (NNs) form part of the former. Previous works such as (Wang et al., 2015, Zhang et al., 2018, Xu et al., 2018) show that convolutional NNs (CNNs) (Buslaev et al., 2018, Wei et al., 2017) and generative adversarial networks (GANs) (Zhang et al., 2019) are popular methods for road extraction. (Abdollahi et al., 2020) systematically reviewed the four main groups of deep learning methods that are currently used most often for road extraction. These groups are GANs models, fully connected convolutional NNs, deconvolutional NNs like DenseNets, and patch-based convolutional NNs. The authors conclude that deep learning methods are the most effective in road extraction compared to regular approaches. Also, due to robust pre- and post-processing techniques these methods can be used on all types of roads, not just main and feeder roads.

The literature presented thus far only applies deep learning road extraction methods to formal roads, but none on informal roads. This paper presents a trained GANs-UNet model on a subset of a new manually digitised data set digitised by this research group. Both the model trained on our data set and the pretrained model presented in (Isola et al., 2017) are then tested on unseen data from our data set to see how a model designed and trained on formal road data sets performs on an informal road data set.

3. METHODOLOGY

3.1 The Data

Figure 1 shows a snapshot of an area near the Khayelitsha township in the Western Cape province in South Africa. Khayelitsha presents a useful area to train a neural network due to its wide variety of complex roads. The differences can clearly be seen between the urban roads on the left hand side, and informal roads on the right hand side of the image. The challenges informal roads pose to road extraction methods are listed and highlighted in the figure from A to D. A showcases how informal roads vary in length and width, B how informal roads do not appear in regular patterns, C shows how the colouring of informal roads are the same as the area around them and D shows how the colouring of informal roads is not constant.
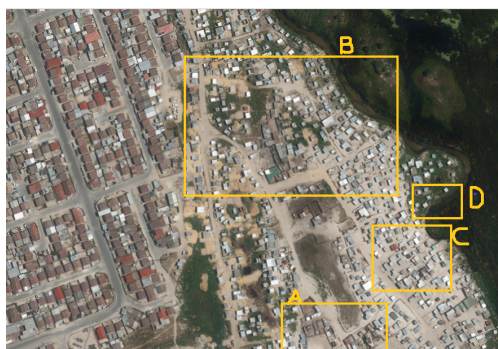

Figure 1. Aerial imagery of Khayelitsha Township in Western Cape, South Africa, obtained through a WMS service published by the City of Cape Town

Aerial images of informal roads that were manually digitised in ArcGIS by four digitisers to provide a training data set for the neural network, are of the informal settlement Khayelitsha in South Africa. This high resolution imagery is available from the City of Cape Town[4].

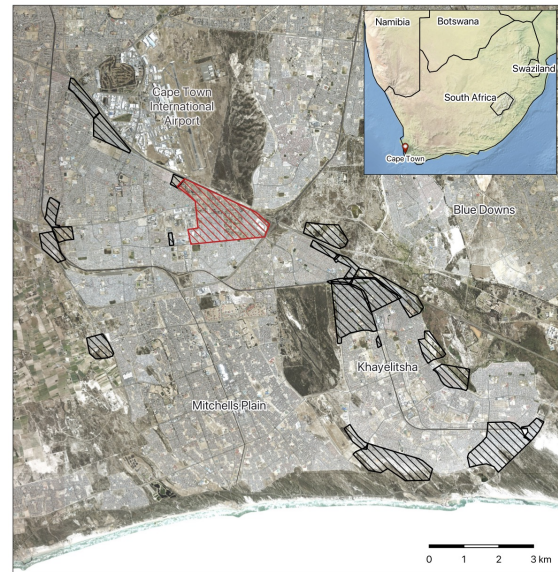[4] Imagery was obtained from the following website, https:

Figure 2. Aerial imagery indicating the region areas digitised in Khayelitsha, South Africa. (Polygon 11 shown in red.)

3.2 Digitisation

The project demarcated the area into 23 regions of 30cm resolution at a 1:500m scale representing informal roads of the area. These areas are shown in Figure 2. The digitising was done by four digitisers who digitised each of the polygons individually. The digitisers worked independently but collaborated throughout the project on any digitisation challenges and recommendations for improvement. Roads were digitised as polygons rather than road center lines to compensate for the irregularity of informal roads' width and shape. Three type of informal roads were defined, namely footpaths, vehicle roads and throughways. Footpaths are those that are used by pedestrians that are not broad enough for vehicles, but have the possibility to be broadened. Vehicle roads are broad enough for a vehicle to use and a throughway are those routes in-between houses that are too narrow to be used by a vehicle. In this paper all three attribute types were combined to test the improvement in road extraction of the neural network.

Logical rules and definitions were defined for the digitisers to reduce the variation between their data sets. The rules were specified as follows:

- Digitise up to the edges of where sandy soil is visible due to the movement of people and vehicles.

- Ignore footpaths or vehicle paths that are faint and not visibly used.

- Roads should be connected in some way to at least one other road.

- Vegetation and shadows are included in the road polygon if it is clear that the road continues underneath it.

- Digitise the 23 polygons in the same order: Polygon 11 was digitised first as a pilot digitising and the remaining 22 polygons are then digitised from 1 to 23. This allowed

---
//citymaps.capetown.gov.za/agsext1/rest/services/
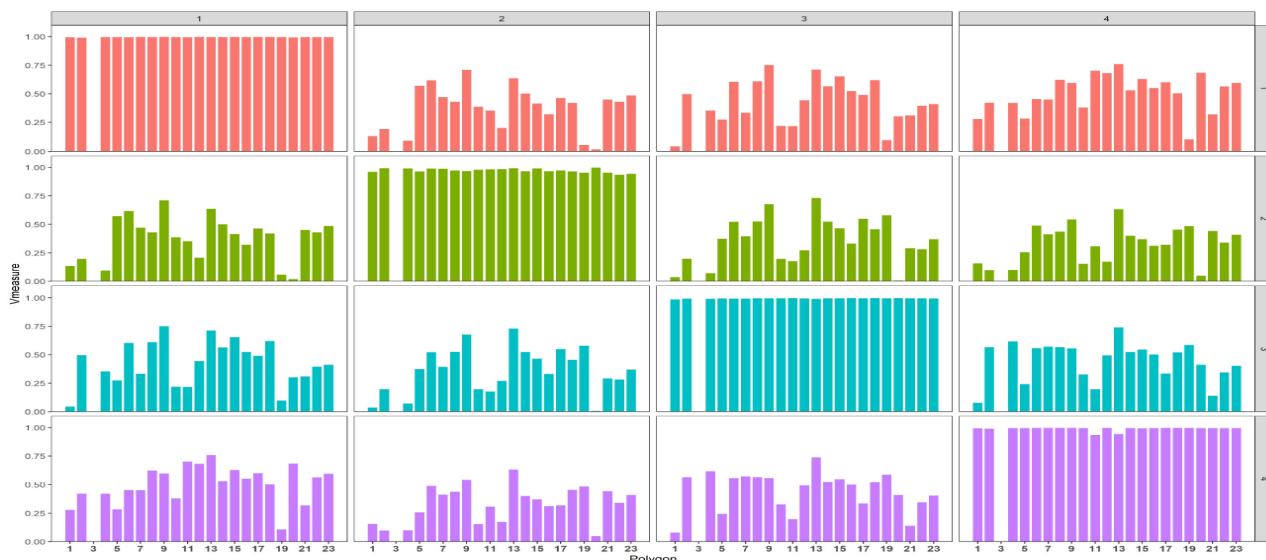Aerial_Photography_Cached/AP_2018_Feb/MapServer

Figure 3. Bar charts of V-measures for each area polygon, for every possible data set pair. V-measures are the same for data set combinations.
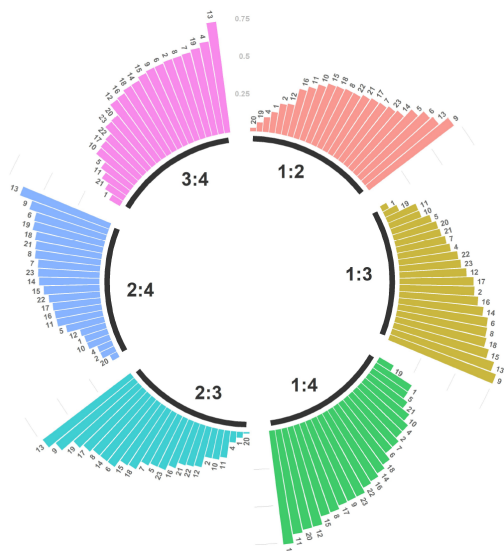

Figure 4. A circular bar chart of ordered V-measures for every unique data set combination, not including self-comparisons.

any necessary developments and rule amendments to be introduced in such a way that all still digitise according to the same standards.

3.3 Accuracy Metric

When doing pairwise map comparisons, accuracy can be easily summarised into a singe useful score, called the V-measure (Goodchild, 2007). The V-measure, proposed by (Nowosad and Stepinski, 2018) can be used when comparing the categorical regions of two maps of the same location. It is the optionally weighted harmonic mean of the completeness (c) and homogeneity (h) scores of the two maps,

$$V_\beta = \frac{(1+\beta)hc}{\beta h + c}. \qquad (1)$$

Two maps are placed on top of one another and the regions resulting from the intersections are then used to calculate the homogeneity and completeness, by their cluster theory definitions. The homogeneity is a function of how well the regions in the second map fit into those of the first, and the completeness is a function of the reverse. The V-measure can be calculated using the R-project[5] package sabre[6].

3.4 The GANs-UNet model

We implement one of the three models in (Abdollahi et al., 2020) to demonstrate the training using our data set. We implement the GANs-UNet model since it achieved the highest F1-score in (Abdollahi et al., 2020).

GANs models (Goodfellow et al., 2014) consist of two frameworks, namely the generator, G, and discriminator, D, networks. These two networks compete against each other where G learns how to generate data so that D will not be able to distinguish between the generated data and the actual data, while D learns how to identify data generated by G and the actual data (Goodfellow et al., 2014, Varia et al., 2018). The training of these two frameworks happens simultaneously where G tries to minimise the objective function while D tries to maximise the objective function (Goodfellow et al., 2014, Isola et al., 2017). Equation (2) shows the objective function $\mathcal{L}_{GAN}(G, D)$ being minimised and maximised, where $G(z; \theta_g)$ is a differentiable function represented as a multi-layer perceptron with parameters θ_g and represents the mapping from a random noise vector z to an output image y.

$$\min_G \max_D \mathcal{L}_{GAN}(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log(D(x)] \\ + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \qquad (2)$$

The function $D(x; \theta_d)$ is also a multi-layer perceptron that has a single scalar value as its output and $D(x)$ is the probability that x is from the data and not p_g. Note that $p_z(z)$ represents

---

[5] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[6] SABRE: Spatial Association Between REgionalizations. https://nowosad.github.io/sabre/articles/sabre.html Accessed 26 October 2021.

a prior defined so that the distribution of G, p_g, over data x can be learnt and $p_{data}(x)$ is the distribution of the data itself (Goodfellow et al., 2014).

The GANs-UNet model also learns a generative model like a normal GAN, but in a conditional way (Isola et al., 2017). The conditional GANs or cGANs learns a structured loss that only penalises "the joint configuration of the output", and learns to map an observed image x and a random noise vector z to y, $G : x, z \to y$. The expression of the objective function of a cGAN model is given by Equation (3). The goal of G is to minimise the objective function against D, while D tries to maximise the objective, which is shown in Equation (4) (Isola et al., 2017).

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log(D(x, y))]$$
$$+ \mathbb{E}_{x,z}(\log(1 - D(x, G(x, z))))] \quad (3)$$
$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) \quad (4)$$

The generator network follows a U-Net based architecture (Ronneberger et al., 2015) that consists of a contracting and a symmetric expanding path. The contracting path captures the context and the expanding path "enables precise localisation" (Ronneberger et al., 2015). A U-Net type of architecture network passes the input through layers that down sample until a bottleneck layer is encountered and the following layers up sample (Ronneberger et al., 2015, Isola et al., 2017). (Isola et al., 2017) point out that in most image translation problems, a lot of low-level information that is shared between input and output is desired to be directly fed across the network. Their example is that of an image colourisation task where the input and output shares the positions of the most notable edges.
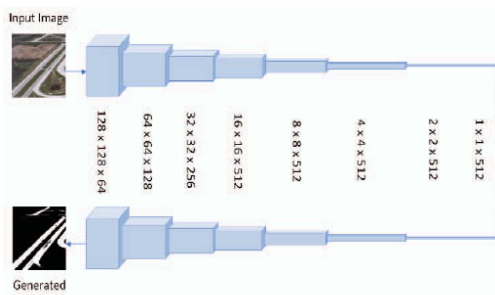

Figure 5. GANs-UNet generator function's architecture from (Isola et al., 2017).

Skip connections provide the solution of this problem as they feed low-level information directly to the decoder part of the network. Therefore skip connections were included between each layer i and layer $n - i$ with n being the total number of layers in the model. Every skip connection concatenates the channels that are at layer i with the channels that are at layer $n - i$. When (Varia et al., 2018) implement the model they repeat a down-sampling stack of two convolutional layers that uses filters of size 3×3, a ReLU layer and a maximum pooling layer that uses a stride of 2 before the bottleneck layer. The function for the ReLU function is given by ReLU : $f(x) = \max(0, x)$.

Figure 5 shows the architecture of the generator function proposed by (Isola et al., 2017) that (Varia et al., 2018) implemented on formal road data and we apply to informal road data.

The discriminator network utilises a PatchGAN to distinguish between a real and a fake image that was generated by the generator network that the research team of (Isola et al., 2017) de-
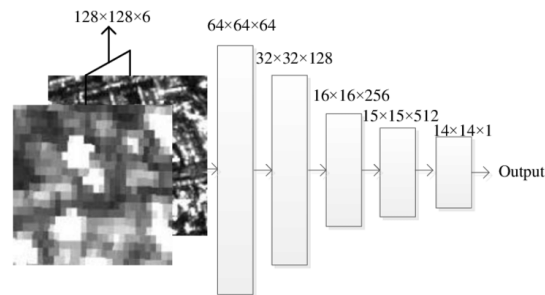

Figure 6. GANs-UNet discriminator function's architecture from (Ao et al., 2018).

signed. This architecture only focuses on the design of local image patches so that it penalises the structures at the scale of the patches. The discriminator's goal is to decide if each patch in an image is from the input data or from the generator. The output of the discriminator is the average of all of the responses after the discriminator is convolutionally run across the image. Figure 6 shows the discriminator function's architecture that (Isola et al., 2017) proposed and implemented by (Varia et al., 2018) to extract formal roads.

4. IMPLEMENTATION

4.1 Dataset

The data set used in this paper is the merger of the four digitised data sets of Polygon 11, considering human error by excluding digitised regions without significant consensus. This data intersection is combined with the City of Cape Town's official road center lines dataset, buffered according to the width attribute when available, otherwise a $5m$ buffering[7]. Polygon 11 has a 3.71 km^2 coverage at a resolution of $30cm$.

4.2 Training and testing

The high-resolution aerial imagery was split into 64 tiles of 600×600 pixels and their corresponding digitised segmentation masks. We further split the tiles by randomly assigning the tiles and their digitised maps into the following sets: 70% training, 15% validation and 15% testing. To overcome the issue of a limited data set and to reduce overfitting, we artificially enlarge the data set using label-preserving transformations. This is done by generating extra training data from the available tiles, through image flipping, cropping, rotations, and translations. The result is a total training set of 206 images. We employed a Linode cloud GPU using an NVIDIA Quadro RTX 6000 GPU (Graphics Processing Unit) to train our model. We use a batch size of 64 and a starting learning rate of 0.02 which is reduced with a factor of 0.01 when the loss continues to increase for more than 10 epochs up to a minimum of 0.0001 during training. The trained model and the pre-trained model were tested on 13 images. The GANs-UNet code that was used to train and to get a pre-trained model is available at `https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix` as the pix2pix model.

4.3 Model evaluation

The process of training a GAN model involves training two models concurrently namely; the generator and discriminator.

---

[7] `https://odp.capetown.gov.za/datasets/tct-road-centerlines/explore`

We therefore can not objectively assess the image quality or diversity produced by the generator exclusively from the loss function. Furthermore, it would be very time consuming to manually inspect all generated images to assess model performance. Fortunately there are a number of plausible quantitative and qualitative evaluation metrics used to assess GANs in literature. One such widely accepted metric which provides a robust assessment of GANs is the Fréchet Inception Distance (FID) score proposed by (Heusel et al., 2017).

The FID score gives the distance between the probability of observing real world data $p_w(.) \sim$ Gaussian with mean (m, C) and the the probability of generating model data $p(.) \sim$ Gaussian with mean (m_w, C_w). This is given by:

$$d^2((m, C), (m_w, C_w)) = \|m - m_w\|_2^2 + Tr(C + C_w - 2(CC_w)^{1/2}).$$
(5)

A high FID score indicates low-quality predicted image; while, a lower score indicates a high-quality image prediction. We measured FID on all resulting generated images as opposed to during training.
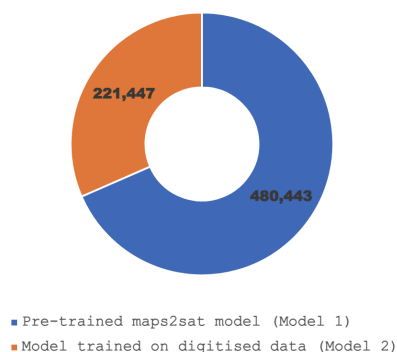

Figure 7. Model Evaluation: FID score

As can be seen in Figure 7, the model trained on the more accurately digitised images does a better job at generating informal roads images than the pre-trained model.

5. RESULTS

The 4 digitisations were analysed for accuracy using the V-measure. By definition, the V-measure is symmetrical. This means that it does not matter which map is used as first or second in the comparison. This is clearly seen in Figure 3, where the V-measures for the area polygons are plotted for every possible permutation of the four digitisations. Only the unique combinations (not including self-comparisons) need then be used when searching for area polygons that are consistently scoring low or high V-measures. These area polygons are easily spotted in Figure 4, a circular graph of the area polygon V-measures for the unique combinations. The area polygons are ordered according to V-measure size. Since the only possible range for the V-measure is zero to one, the three value markers on the plot serve as quartiles.

The hypothesis was that the polygons digitised last would prove more similar, have higher V-measures, than the polygons digitised first. The polygons were digitised in numerical order. However, this is clearly not the case. They seem to be scattered randomly among the ordered V-measures. Studying the lowest
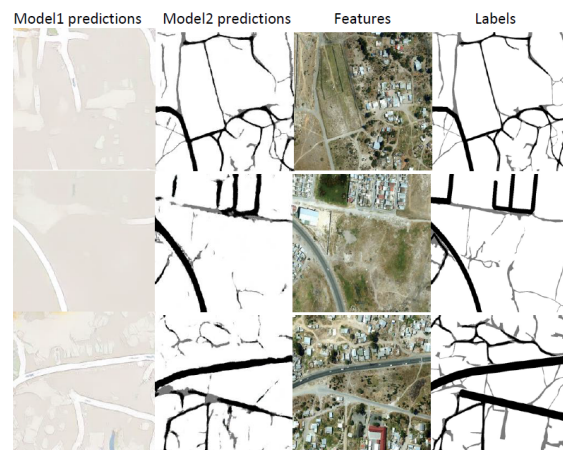

Figure 8. Results obtained from using the pre-trained maps2sat model (model 1) vs our trained model (model 2).

and highest scoring polygons give insight into the difficulties experienced by the digitisers and the reasons why data sets converged or diverged.

The highest scoring polygons are numbers 13 and 9. These were both small area polygons consisting of mostly informal housing and informal roads. The smaller size of the polygons makes it easier not to miss a road whilst digitising, and the homogeneous area type makes road classification less challenging. Road edges were easy to identify here, since buildings helped define them. Quite a number of area polygons have consistently low V-measures. These are numbers 1, 2, 4, 5, 10, 11, 19, 20, 21, all with poorly identifiable road edges, faded roads in rural areas, and a mixture of road and area types causing confusion about how roads should be digitised or classified.

Figure 8 shows the results obtained from the pre-trained model (model 1) proposed by (Isola et al., 2017) and the same model trained on our data set (model 2). In Figure 8, the fist column shows the predictions produced by the pre-trained model, the second column the predictions produced by the model trained on our data and the third and fourth columns show the features and labels respectively. Figure 9 illustrates a complex area with example. The model still experiences difficulties. Figure 10 shows the loss function produced when training the model on our data set.

Figure 9. The GANs-UNet model for a very complex area.

6. DISCUSSION

If we compare the labels and features columns in Figure 8 to both prediction columns, it is strikingly clear how well the model performs in extracting informal roads when it is trained on informal roads in comparison to the pre-trained model which did not have exposure to informal roads during training. The pre-trained model successfully extracted every paved road in the images, but struggled with any road that was not paved. Figure 8 also shows that model 2 can still improve extraction of

the finer informal roads. Figure 9 shows that there is still room for improvement in the model since it misses the complex informal roads, specifically the numerous small roads in between informal, haphazard houses. This is due to difficulty digitising these areas. Even in high resolution imagery it is difficult for the human eye to decide which paths are viable. Additional manpower is needed to digitise such areas. However, the trained neural network shows significant improvement over (Thiede, 2019, Thiede et al., 2020).

Figure 10 shows how the generator and discriminator losses (G-GAN) relate to the adversarial training of the auto-encoder portion of the network while the generator loss (G-L1) relates to the quality of the generated image. The L1 loss tends to improve as the number of training epochs increase. With more training samples and a longer training time, this could possibly be further improved. The discriminator losses also converge around [0, 1].

Although the digitisation accuracy indicates that improvement can be made, the trained neural network indicates significant results for informal road extraction none-the-less. Future research aims to manually recheck the digitisation and make improvements, incorporate the attribute types and the ability of a neural network to predict attribute type investigated. The need for a training data set of informal roads and pathways is important in order to progress towards informative extraction. This research is the first approach to training a road detection model using the concept of an informal road. The societal benefits extend into more rural areas across Africa as well, and can provide knowledge of a country's growth and needs. The full data set will be made freely available, in order to benefit other researchers and stakeholders.

7. CONCLUSION

In conclusion, the GANs-UNet model that was designed to extract formal roads also successfully extracts informal roads when trained on a data set that consists of informal and formal road labels. The informal roads training data developed here is the first available. Future work will extend this research by training and testing the other two strong deep learning models in (Abdollahi et al., 2020) on our data set, and will consider adaptations where needed as well as investigate transferability. In addition, the full data set will be available and consist of the 22 other polygons. Further research is underway on the best mechanisms to merge the digitisers' data sets towards achieving the best ground truth. In addition, the extraction of each type of road will be investigated as they serve different purposes and result in different stakeholder interventions.

ACKNOWLEDGEMENTS

REFERENCES

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., Alamri, A., 2020. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sensing*, 12(9).

Ao, D., Dumitru, C., Schwarz, G., Datcu, M., 2018. Dialectical GAN for SAR image translation: From Sentinel-1 to TerraSAR-X. *Remote Sensing*, 10(10), 1597.

Buslaev, A., Seferbekov, S., Iglovikov, V., Shvets, A., 2018. Fully convolutional network for automatic road extraction from satellite imagery. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 207–210.

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., , Pan, C., 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6), 3322–3337.

Corburn, J., Vlahov, D., Mberu, B., Riley, L., Caiaffa, W., Rashid, S., Ko, A., Patel, S., Jukur, S., Martínez-Herrera, E., 2020. Slum health: arresting COVID-19 and improving well-being in urban informal settlements. *Journal of Urban Health*, 97(3), 348–357.

Girres, J.-F., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459.

Goodchild, M., 2007. Citizen as voluntary sensors: Spatial data infrastructure in the world of web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2(1), 24–32.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Stat*, 1050, 10.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing*, 32(3 SECT 4W2), 151-160.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Iimi, A., Ahmed, F., Anderson, E., Diehl, A., Maiyo, L., Peralta-Quirós, T., Rao, K., 2016. New rural access index: main determinants and correlation to poverty. *World Bank Policy Research Working Paper*.

Isola, P., Zhu, J., Zhou, T., Efros, A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134.

Kirthika, A., Mookambiga, A., 2011. Automated road network extraction using artificial neural network. *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, IEEE.

Li, M., Stein, A., Bijker, W., Zhan, Q., 2016. Region-based urban road extraction from VHR satellite images using binary partition tree. *International Journal of Applied Earth Observation and Geoinformation*, 44, 217-225.

Mantelli, M., Pittol, D., Neuland, R., Ribacki, A., Maffei, R., Jorge, V., Prestes, E., Kolberg, M., 2019. A novel measurement model based on abBRIEF for global localization of a UAV over satellite images. *Robotics and Autonomous Systems*, 112, 304–319.

Mendes, C., Frémont, V., Wolf, D., 2016. Exploiting fully convolutional neural networks for fast road detection. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 3174–3179.
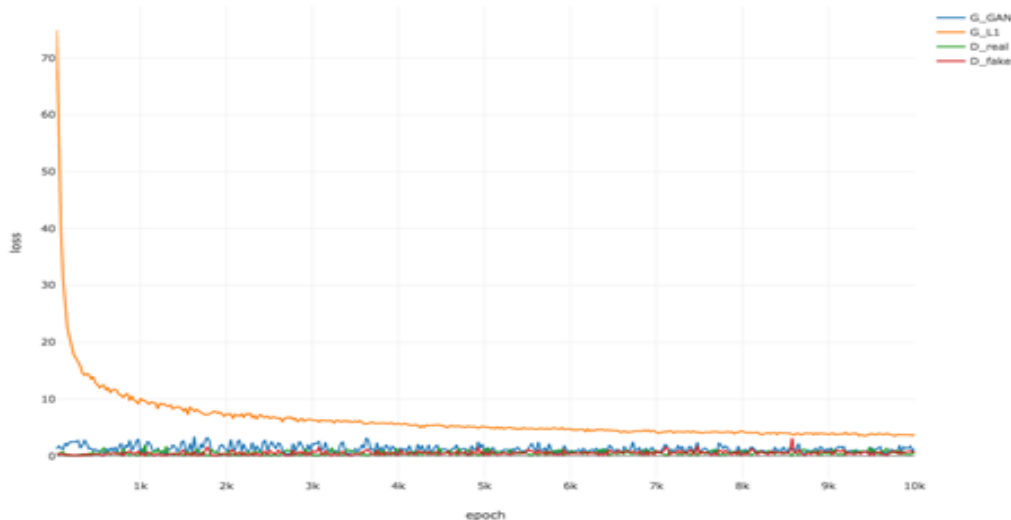
Figure 10. The loss function of the GANs-UNet, or pix2pix, trained on our data set.

Nobrega, R., O'Hara, C., Quintanilha, J., 2006. Detecting roads in informal settlements surrounding Sao Paulo city by using object-based classification. *Proceedings of the 1st International Conference on Object-based Image Analysis (OBIA 2006), Salzburg, Austria*, 4–5.

Nowosad, J., Stepinski, T. F., 2018. Spatial association between regionalizations using the information-theoretical V-measure. *International Journal of Geographical Information Science*, 32(12), 2386-2401.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net:Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 234–241.

Rui, Y., 2013. Urban growth modeling based on land-use changes and road network expansion. PhD thesis, KTH Royal Institute of Technology.

Runsten, S., Nerini, F., Tait, L., 2018. Energy provision in South African informal urban settlements- A multi-criteria sustainability analysis. *Energy Strategy Reviews*, 19, 76–84.

Ruwaimana, M., Satyanarayana, B., Otero, V., Muslim, A., Syafiq A., M., Ibrahim, S., Raymaekers, D., Koedam, N., Dahdouh-Guebas, F., 2018. The advantages of using drones over space-borne imagery in the mapping of mangrove forests. *PlOS ONE*, 13(7), e0200288.

Senthilnath, J., Dokania, A., Kandukuri, M., Ramesh, K., Anand, G., Omkar, S., 2016. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosystems Engineering*, 146, 16–32.

Thiede, R. N., 2019. Statistical accuracy of an extraction algorithm for linear image objects. Master's thesis, University of Pretoria.

Thiede, R. N., Fabris-Rotelli, I. N., Debba, P., Lib, M., A. Stein, A., 2020. Uncertainty quantification for the extraction of informal roads from remote sensing images of South Africa. *South African Geographical Journal*, 102(2), 249–272.

UN Habitat, 2006. State of the world's cities 2006/7. *New York: United Nations*, 292.

UN Habitat, 2015. Habitat III Issue Paper 22: informal settlements. *New York: United Nations*.

Varia, N., Dokania, A., Senthilnath, J., 2018. Deepext: A convolution neural network for road extraction using RGB images captured by UAV. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 1890–1895.

Wang, J., Song, J., Chen, M., Yang, Z., 2015. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing*, 36(12), 3144–3169.

Wang, R., Strong, D., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4).

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., Eklund, P., 2016. A review of road extraction from remote sensing images. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(3), 271–282.

Wei, Y., Wang, Z., Xu, M., 2017. Road structure refined CNN for road extraction in aerial image. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 709–713.

Xu, Y., Xie, Z., Feng, Y., Chen, Z., 2018. Road extraction from highresolution remote sensing imagery using deep learning. *Remote Sensing*, 10(9), 1461.

Zhang, X., Han, X., Li, C., Tang, X., Zhou, H., Jiao, L., 2019. Aerial image road extraction based on an improved generative adversarial network. *Remote Sensing*, 11(8), 930.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual U-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.

Zhao, F., Wu, J., Sun, H., Gao, Z., Liu, R., 2016. Population driven urban road evolution dynamic model. *Networks and Spatial Economics*, 16(4), 997–1018.