Estimating Poverty Using Aerial Images: South African Application

Vongani H. Maluleke[1,2,*], Sebnem Er[1], Quentin R. Williams[2]
[1]*PD Hahn Building (South Entrance) Level 5, Upper Campus, Statistical Sciences Department,*
University of Cape Town, 7700, Cape Town, South Africa
[2]*Meraka, CSIR, Meiring Naude Road, 0001, Pretoria, South Africa*

***Abstract*—Policy makers and the government rely heavily on survey data when making policy-related decisions. Survey data is labour intensive, costly and time consuming, hence it cannot be frequently or extensively collected. The main aim of this research is to demonstrate how deep learning in computer vision coupled with statistical regression modelling can be used to estimate poverty on aerial images supplemented with national household survey data. This is executed in two phases; aerial classification and detection phase and poverty modelling phase. The poverty measure estimated in this paper is the Sen-Shorrocks-Thon index (SST). The models in phase one performed relatively well with the classification model having an accuracy rate of 90.85% and a log loss of 0.5783 while the instance segmentation model has a log-loss of 0.839. The ridge model in phase two also performed well with an R^2 of 0.708, a root mean square error (RMSE) of 0.081, and a strong positive correlation between the estimated SST and actual SST of 0.838 which indicates the strong ability of the model to estimate the SST from the geo-type and dwelling type of an area.**

***Keywords*— Poverty, Computer vision, statistical regression modelling, Aerial Image(s).**

I. INTRODUCTION

Poverty is a complex, multi-dimensional phenomenon that constitutes of multiple deprivation aspects [1]. It requires extensive research to comprehend the severity and impact it has in a community. The complexity of poverty is derived from the many factors that it is influenced by, in which the effect of some factors is still unknown.

The National Development Plan (NDP) has been compiled by the South African government to combat poverty and reduce inequality in South Africa by 2030 [2]. Survey data is the main information source to monitor the development and growth of the country to meet the set NDP targets. However, due to limited resources, the collection of survey data is not always conducted frequently nor extensively to a level where the data satisfactory reflects the true livelihood of individuals and household in South Africa.

Currently the South African government is using poverty lines to set and monitor poverty alleviation targets. Poverty line is an index used to divide the population into two groups based on a socio-economic measure such as welfare [3]. Households or individuals that fall above the poverty line are classified as not poor, while those that fall below the poverty line are classified poor.

There has been a growing interest in the usage of aerial images supplemented with publicly available data to monitor, assess and measure socio-economic indicators, especially with the recent advancements in computer vision. This has also ignited researches in the use of deep learning to address socio-economic related problems.

Some researchers have successfully introduced an approach of using night-time light intensities as a proxy to estimate and map poverty at a country and continental level [4]. They trained a fully convolutional neural network (CNN) to estimate night-time light intensity from day-time aerial images and simultaneously trained another model that captures the effect of the features on the aerial images to estimate poverty. They obtained that day-time aerial images can be utilised to make relatively accurate spatial socio-economic status estimations across Nigeria, Uganda, Malawi, and Tanzania [4].

Recently, [5] further investigated the use of aerial images to estimate poverty and they found that built-up area and roof type have a strong correlation with welfare. They used a deep learning CNN to detect and classify the built-up area and roof type. This paper will demonstrate the use of deep learning in computer vision coupled with statistical regression modelling to estimate poverty in South Africa using aerial images supplemented with longitudinal survey data, with the purpose of developing a tool that will assist the government to combat poverty.

In this paper, the poverty rate is estimated from aerial images by training a deep learning convolutional neural network to detect dwelling types (brick house, traditional house, and informal settlement) and classify geo-types (urban, farm, and rural) from aerial images. This is then followed by training a statistical regression model to estimate the poverty rate, Sen-Shorrocks-Thon index (SST), from the detected dwelling types and geo-types in the aerial images.

II. BACKGROUND

This section will give a foundational understanding of the measures and techniques that will be used to estimate poverty.

Manuscript received December 06, 2018; accepted December 22, 2018.
*Corresponding author: malulekevon@gmail.com

A. Poverty Measures

Poverty measures are guiding tools of how poverty is monitored in a country. This is because policy makers construct poverty reduction policies based on how poverty is defined and measured. For instance, in South Africa the government structures policies that will result in more individuals and households living above the poverty line. They further use it to assess the progress they've made by implementing a certain policy.

South Africa has three national poverty lines that the government uses to monitor and measure poverty [6]. The poverty line that is used in this paper is called the Food Poverty Line (FPL), which is the rand value below which individuals or households cannot afford purchasing adequate food to obtain minimum per capita per day energy required for adequate health [6]. These poverty lines were introduced by Statistics South Africa with the intension of enabling the country to measure and monitor poverty on a money-metric or expenditure-based dimension [6].

The poverty line is an imperfect construct that makes the crude assumption of defining poverty by using some minimum level in which people continue to survive below it. However, when making analysis, an index ought to be used to further build understanding of the nature of poverty as argued by [3].

There are many different poverty measures that one can use. The most commonly used measures are namely head count index, denoted as P_0, and poverty gap index, denoted as $P^P{}_1$ [7]. Headcount index represent the incidence of poverty as it measures the proportion of households or individuals who fall below the poverty line. Poverty gap index represents the depth of poverty as it measures the extent of which households and individuals lie below the poverty line [7].

These poverty measures are popularly used because they are easy to compute and understand [7]. However, they only capture one aspect of poverty, which can result in misleading output and hence skewed conclusions. For a robust view of poverty, composite poverty measures are highly recommended [8].

Sen [9] proposed a composite poverty measure called the Sen index, denoted as P_S. This measure considers the population size of the poor, welfare (income or expenditure) shortfall relative to the poverty line and the degree of inequality [9]. The Sen index is defined as:

$$P_S = P_0(P^P{}_1 + (1 - P^P{}_1)G^P) \qquad (1)$$

where $P_{0,}$ $P^P{}_1$, G^P is the headcount index, poor population poverty gap and poor population Gini coefficient of the poverty gap ratio. The Gini coefficient is a measure of inequality for a given distribution based on the Lorenz curve [7].

Shorrocks [14] later proposed a modified version of the Sen index which was then later modified by Thon to become the Sen-Shorrocks-Thon (SST) index, denoted as P_{SST}, and defined as:

$$P_{SST} = P_0 P^P{}_1\left(1 + \hat{G}^P\right) \qquad (1)$$

SST index takes into account the depth, incidence and inequality of poverty [7]. It is a more appealing version of the Sen index because the decomposing SST allows for analysis of source of changes in poverty overtime [7]. This decomposition enables us to determine the change in the number of poor people, the change in the level of poor people being poor and the change in the inequality among poor people [7]. In this paper, the SST index is used as the poverty measure and is computed using the survey data.

B. Convolutional Neural Networks

Convolutional neural networks (CNN) is a specialised kind of neural network that uses a mathematical operation called convolution, denoted by *, in at least one layer in replacement of the normal neural network matrix multiplication [10]. The convolution operation is defined as:

$$S(i,j) = (I * F)(i,j) = \sum_m \sum_n I(i+m, j+n) * F(m,n) \qquad (3)$$

where *I( )* is the input signal and *F( )* is the filter.

A standard CNN contains three types of layers, namely; convolutional, pooling and fully connected layers, where two broad processes happen across the entire network [10]. Collectively the convolutional and pooling layers perform feature extraction, where an image input is passed through the first layers of the network to extract low-level features (edges and corners) and then the last layers extract high-level features (texture and objects) [11].
The convolutional layer is the core layer in CNN, which involves simultaneously performing several convolutions to produce a set of linear activations defined by:

$$a_l = W_l * x + b_l \qquad (4)$$

Where W_l is the weights of the *l-th* convolutional layer filter used to perform convolution (*) on the image or feature map, denoted *x*, offset by the *l-th* layer bias (b_l). The linear activation is then passed through a non-linear function ($f_l(a_l)$). Rectified linear unit (ReLU) is popularly used as the non-linear function in classification tasks because it has low computational complexity and it has an ability to maintain the scale of output values [10].

In the pooling layer, a pooling function is applied to the convolutional layer output. This layer is used to reduce the size of the image to speed the computation of the model. Max pooling and average pooling are popularly used in CNN to down sample the image spatially.

The final layer of a standard CNN is a fully connected layer that uses the final output of the feature extraction process, called a feature map, to perform classification. This is executed by matrix multiplication of the neurons with an offset of the bias to output a K dimensional vector using soft-max. Where K is the number of classes and each element in the output is the probability of the image input belonging to the k-th class.

C. *Mask R-CNN*

Mask R-CNN is a type of instance segmentation CNN model that is used to perform instance object detection by identifying the outlines of objects at pixel level [12]. It is essentially a combination of object detection and semantic segmentation, where an additional branch in Faster R-CNN is added to generate a segmentation output as seen in Fig. 1. According to[12], it currently outperforms all existing single-model entries on every task. Furthermore, it is easy to train and to generalise to other tasks [12].

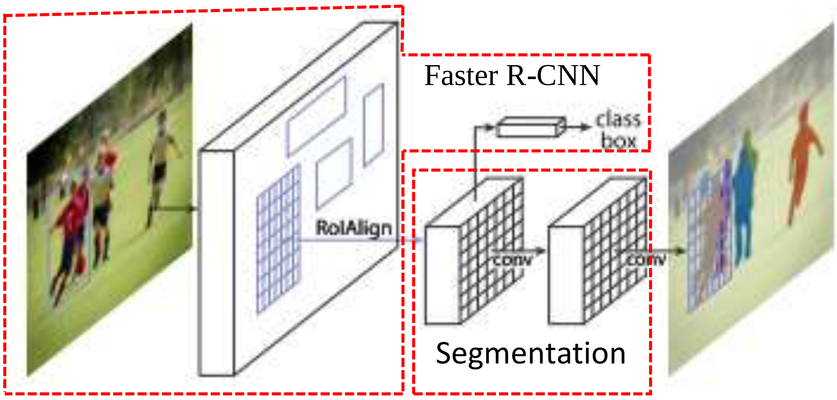


Fig. 1. Mask R-CNN architecture. Extension of Faster R-CNN with a segmentation branch and RoIAlign for accurate mapping between original image and proposal [12].

Firstly, an image is passed through a CNN backbone model where low-level and high-level features are extracted. This backbone network model outputs a feature map which becomes the input for proceeding stages. The feature map from the backbone model is then passed through a region proposal network (RPN) where the network uses a sliding window to scan the feature map with the purpose of finding regions that are likely to contain an object called proposals (regions of interest). These regions that are scanned to find proposals are called anchors, which are overlapping boxes of different aspect ratios and sizes distributed all over the feature map.

Once the proposals have been generated by the RPN, each proposal is passed into a Faster R-CNN to perform classification and localisation to output a class label and a four-valued bounding box tuple (x, y, w, h), respectively. Simultaneously, a Fully Convolutional Network (FCN) is applied to each proposal to predict the mask of the detected object instance. This mask is a binary encoding of the spatial layout of the object, where pixels that belong to the object are encoded with the value one and zero otherwise [12].

RoIAlign was introduced in Mask R-CNN to avoid misalignment between the original image and feature map caused by quantisation in RoIPool, which is observed in Faster R-CNN when a feature map is resized into a fixed size [12]. RoIAlign allows for accurate mapping of the proposals from the original image onto a feature map.

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{bbox} + \mathcal{L}_{mask}$$

The loss function of the Mask R-CNN is defined on each proposal during training as the combination of the log loss of classification $(\mathcal{L}_{cls})$, localisation $(\mathcal{L}_{bbox})$ and segmentation mask $(\mathcal{L}_{mask})$ and is given by:

where $\mathcal{L}_{cls}$ is multi-class cross entropy for the class labels, $\mathcal{L}_{bbox}$ is the smoothed *L1* loss for the bounding box and $\mathcal{L}_{mask}$ is the average cross entropy loss for the masks. A model that performs well has a log loss that is close to zero.

A. *Statistical Regression Modelling: Ridge Regression*

Statistical regression is a tool used for the identification and characterisation of the functional relationships between the dependent variable and *p* independent variables [13]. In this paper, the dependent variable is the SST and the independent variables are the number of detected dwelling types found within each geo-type.

Ridge regression is an extension of multiple linear regression that is used for analysing multiple regression data suffering multicollinearity[13]. Ridge regression was first introduced by [14] to improve the instability of the ordinary least squares (OLS) estimators and has since been popularly used to get improved prediction results.

Ridge regression is also known as a shrinkage method that shrinks parameters zero by imposing an L2 penalty ($\sum_{j=1}^{p} \beta^2_j$) on their size [13]. It assigns low weights to variables that explain less variance [13]. Parameters in ridge regression are estimated by minimising the penalised residual sum of squares (RSS) defined as:

$$\hat{\beta}^{ridge} = \underset{\beta}{argmin} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j}^{p} \beta_j x_{ij})^2 = \underset{\beta}{argmin} \, RSS(\beta) \tag{5}$$

$$Subject\ to: \sum_{j}^{p} \beta_j^2 \leq t \tag{6}$$

which can also be represented in Lagrangian form as:

$$\mathcal{L}(\boldsymbol{\beta}, \lambda) = \underset{\beta}{argmin} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j}^{p} \beta_j^2 \tag{7}$$

$$\mathcal{L}(\boldsymbol{\beta},\lambda) = \underset{\beta}{argmin} RSS(\beta) + \lambda \sum_{j}^{p} \beta_j^2 \tag{8}$$

where λ ≥ 0 is a tuning parameter, which controls the amount of shrinkage observed. Different values of λ result in different ridge regression coefficients and cross validation is used to obtain an optimal λ that minimises the prediction error [13].

Ridge regression models have a low prediction error [13]. However, shrinking parameters by a penalty introduces bias into the model but reduces the sample variance and consequently resulting in a smaller mean squared error (MSE) than OLS estimators [13].

III. DATA

Two different and closely concurrent data sources were used to estimate poverty.

A. *National Income Dynamic Survey (NIDS)*

The survey data used in this paper is National Income Dynamic Survey (NIDS) wave 4 (2014-2015) [15]. NIDS is an income, expenditure, household and individual longitudinal survey conducted nationally in South Africa. The lowest geographic aggregation level is district municipality [15]. The survey data was collected via face-to-face questionnaires and was designed with the sole purpose of analysing the effect of the dimensions of South Africans' well-being over a time.

Households differ in size and composition of demographics which makes it hard to make straight forward poverty comparisons [3]. This hindrance can be resolved by normalising the household expenditure using the number of household members.

A new variable is derived from the household monthly food expenditure variable that represents household food expenditure per household member. This new variable will be used in the computation of SST. The variables that were used from NIDS data is given in Table I. The province, district, latitude and longitude variables are used as location variables when computing the SST.

TABLE I. NATIONAL INCOME DYNAMIC SURVEY (NIDS) VARIABLE DESCRIPTION OF THE USED.

Variable	***Description***	
w4_expf	Wave 4 household monthly food expenditure	Numeric
w4_h_dwltyp	Wave 4 household dwelling type	Categorical
w4_geo2011	Wave 4 household geo type	Categorical
w4_hhsizer	Wave 4 number of household members	Numeric
w4_prov2011	Wave 4 household province	Categorical
w4_dc2011	Wave 4 household district	Categorical
w4_gps_e	Wave 4 household longitude	Numeric
w4_gps_s	Wave 4 household latitude	Numeric

The survey data performs three roles in this paper, which are to:

- assist in manually labelling dwelling types and geo-types in the aerial images;
- compute poverty index, Sen-Shorrocks-Thon-Index, for data preparation; and
- train a statistical regression model to estimate SST.

For dealing with missing data, only successfully interviewed households were considered and the rest were discarded. In wave 4 data, 9620 (80.9%) of household interviews were successful and 2275 (19.1%) were unsuccessful and hence discarded. The training and testing set was obtained by using topological codes, where 70% of topological codes was randomly selected for training and 30% for testing. Topological codes are unique alphanumeric identifiers that are used to reference a map series location coverage on map sheet. An example of a map sheet with topological codes overlaid on it is shown in Fig.2, where each code references a map series location coverage.

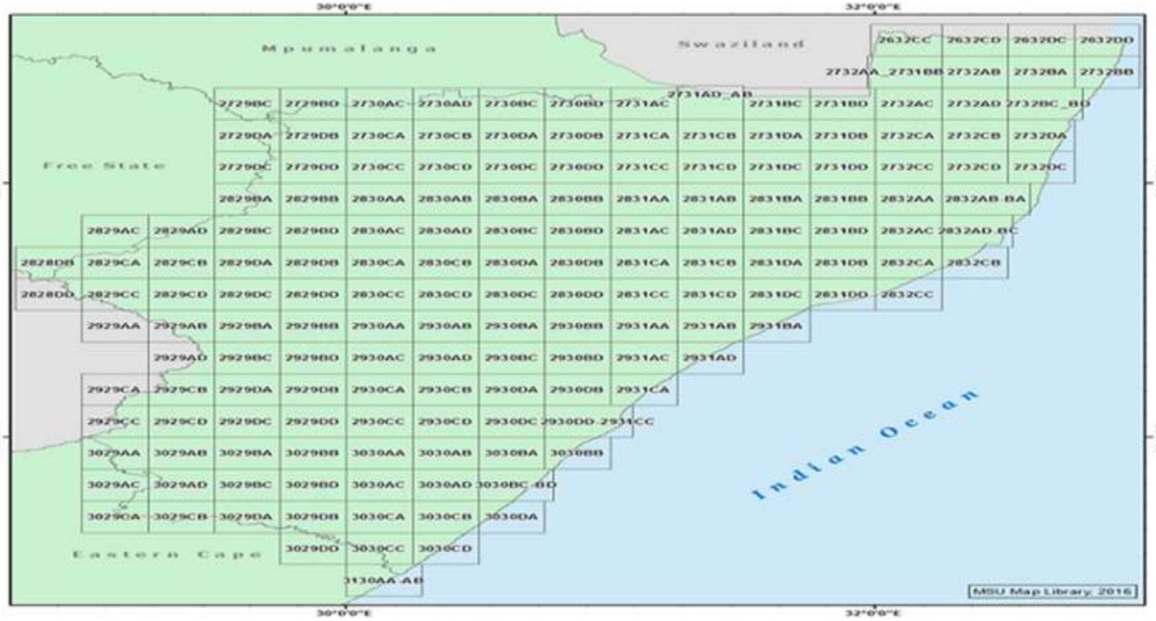
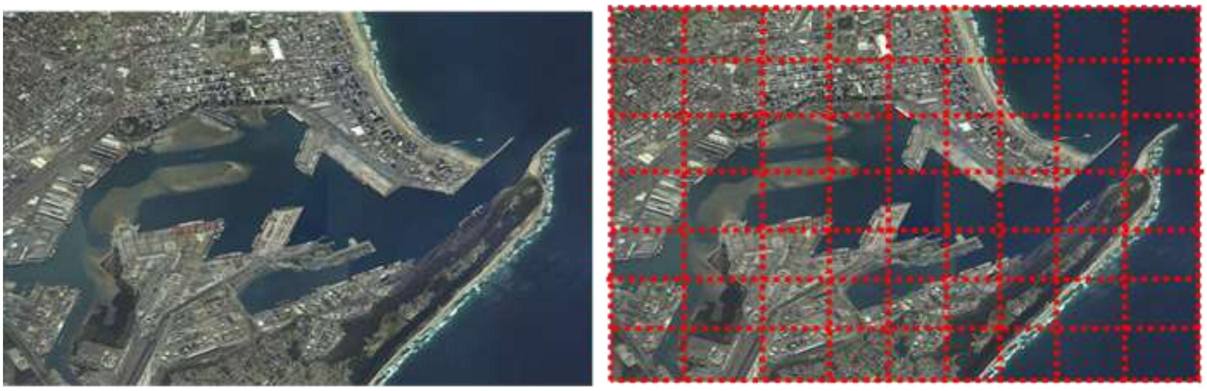

30°0'0"E
32°0'0"E
Mpumalanga
Swaziland
Free State
Eastern Cape
Indian Ocean
2632CC 2632CD 2632DC 2632DD
2732AA_2731BB 2732AB 2732BA 2732BB
2729BC 2729BD 2730AC 2730AD 2730BC 2730BD 2731AC 2731AD_AB 2731BC 2731BD 2732AC 2732AD 2732BC_BD
2729DA 2729DB 2730CA 2730CB 2730DA 2730DB 2731CA 2731CB 2731DA 2731DB 2732CA 2732CB 2732DA
2729DC 2729DD 2730CC 2730CD 2730DC 2730DD 2731CC 2731CD 2731DC 2731DD 2732CC 2732CD 2732DC
2829BA 2829BB 2830AA 2830AB 2830BA 2830BB 2831AA 2831AB 2831BA 2831BB 2832AA 2832AB BA
2829AC 2829AD 2829BC 2829BD 2830AC 2830AD 2830BC 2830BD 2831AC 2831AD 2831BC 2831BD 2832AC 2832AD BC
2828DB 2829CA 2829CB 2829DA 2829DB 2830CA 2830CB 2830DA 2830DB 2831CA 2831CB 2831DA 2831DB 2832CA 2832CB
2828DD 2829CC 2829CD 2829DC 2829DD 2830CC 2830CD 2830DC 2830DD 2831CC 2831CD 2831DC 2831DD 2832CC
2929AA 2929AB 2929BA 2929BB 2930AA 2930AB 2930BA 2930BB 2931AA 2931AB 2931BA
2929AD 2929BC 2929BD 2930AC 2930AD 2930BC 2930BD 2931AC 2931AD
2929CA 2929CB 2929DA 2929DB 2930CA 2930CB 2930DA 2930DB 2931CA
2929CC 2929CD 2929DC 2929DD 2930CC 2930CD 2930DC 2930DD 2931CC
3029AA 3029AB 3029BA 3029BB 3030AA 3030AB 3030BA 3030BB
3029AC 3029AD 3029BC 3029BD 3030AC 3030AD 3030BC BD
3029CA 3029CB 3029DA 3029DB 3030CA 3030CB 3030DA
3029DD 3030CC 3030CD
3130AA AB
MSU Map Library 2016

Fig. 2. Map sheet of the province KwaZulu-Natal, South Africa with topological code overlaid [16].

A. *National Geo-Spatial Information (NGI) Aerial Images*

The aerial images were provided by the National Geo-Spatial Information (NGI) and are concurrent with the NIDS data. The images provided have a resolution of 34000x34000 and for classifying and detecting dwelling types and geo-types, these images were cropped into 300x300 resolution images. This was implemented by overlaying a grid with 300x300 sized cells as shown in Fig. 3. A python package called labelme.py was used to manually label the aerial images [17]. For accurate and consistent labelling, NIDS data was used to assist with the labelling.



(a) **(b)**

Fig. 3. a) Original image b) Grid Overlaid image, with 2931CC topological code.

I. METHODOLOGY

Estimating poverty is executed in two phases namely the aerial classification and detection phase and the poverty modelling phase.

A. *Phase1: Aerial classification and detection*

This phase firstly involves classifying three classes of geo-types (urban, rural and farm) using a CNN model with three convolutional layers, each layer followed by a max-pooling layer. This classification model architecture is depicted in Fig. 4, where an image is passed through the convolutional layers and max-pooling layers to perform feature extraction and then through a fully connected layer which uses a soft-max to output the probability of the image belonging to the three broad geo-type classes namely; urban, rural, and farm.

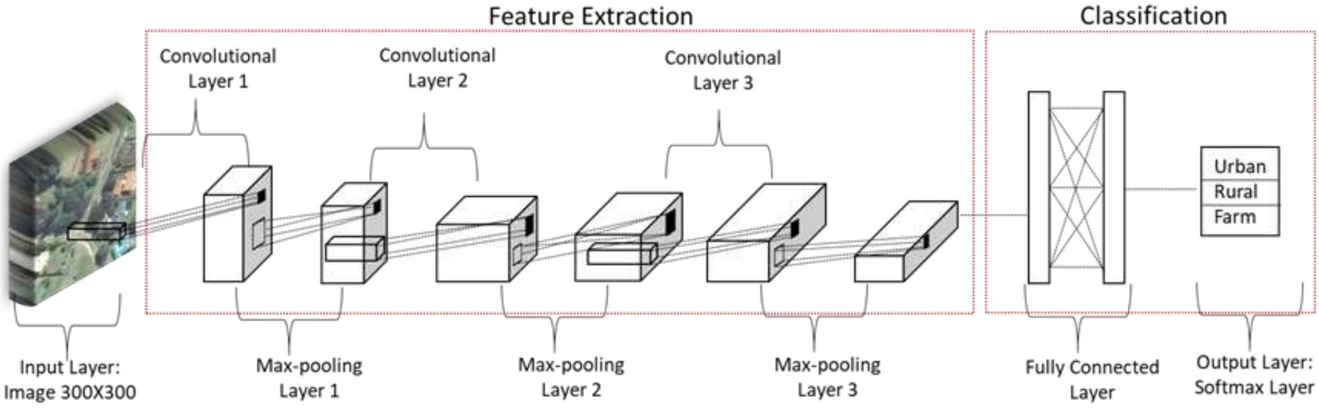


Fig. 4. Convolutional Neural Networks (CNN) model architecture used to perform geo-type (urban, rural, and farm) classification on aerial images.

Next, a Mask R-CNN model is trained on the same labelled images to detect three broad classes of dwelling types namely; brick house, traditional house, and informal settlement. Fig. 5 shows the detection model architecture, where the training data images are passed through a ResNet101 CNN, which is the backbone CNN of this model, to perform feature extraction and output a feature map that is used to generate proposals by the RPN. For each proposal made, three tasks are simultaneously performed to output the class label, bounding box coordinates and the mask. Different mask colours are used to show different object instances detected as seen in Fig. 5 output.

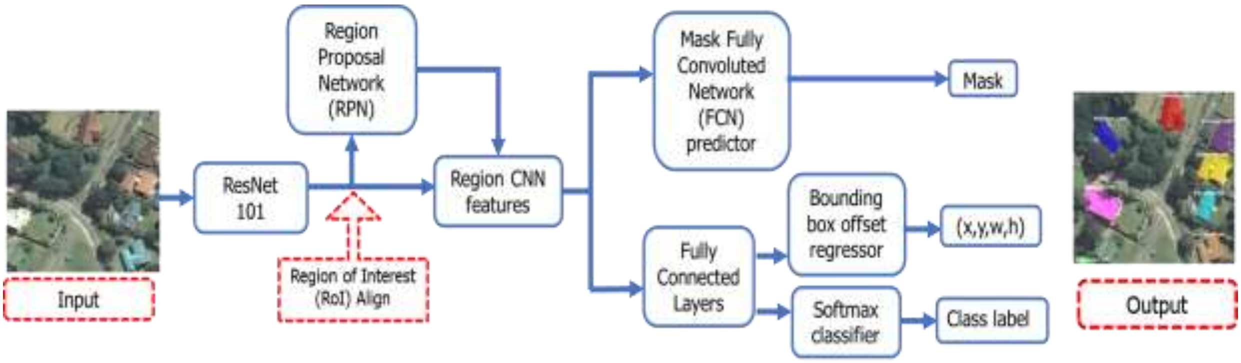

Input
ResNet 101
Region Proposal Network (RPN)
Region of Interest (RoI) Align
Region CNN features
Mask Fully Convoluted Network (FCN) predictor
Mask
Fully Connected Layers
Bounding box offset regressor
(x,y,w,h)
Softmax classifier
Class label
Output

Fig. 5. Mask Region-based Convolutional Neural Networks (Mask R-CNN) model architecture used to perform dwelling-type (brick house, traditional house, and informal settlement) detection on aerial images.

A. *Phase2: Poverty Modelling*

In this phase, the NIDS data was used to compute the SST and the corresponding vectorised cross Table of the geo-type and dwelling type for each topological code. This was then used to train a ridge regression model where the computed SST is the dependent variable and the items in the vectorised cross table are independent variables.

B. *Final Model*

The architecture of the final poverty estimation model constitutes of phase one and phase two trained models, as shown in Fig. 6. To estimate poverty and create a poverty distribution map, a set aerial images for a certain location (topological code, district, or province) is used as input and classification of geo-type and detection of dwelling type is performed individually on each image. The classified geo-types and detected dwelling types are then aggregated and used as input in phase 2 to estimate the poverty rate, SST, using a trained ridge regression.

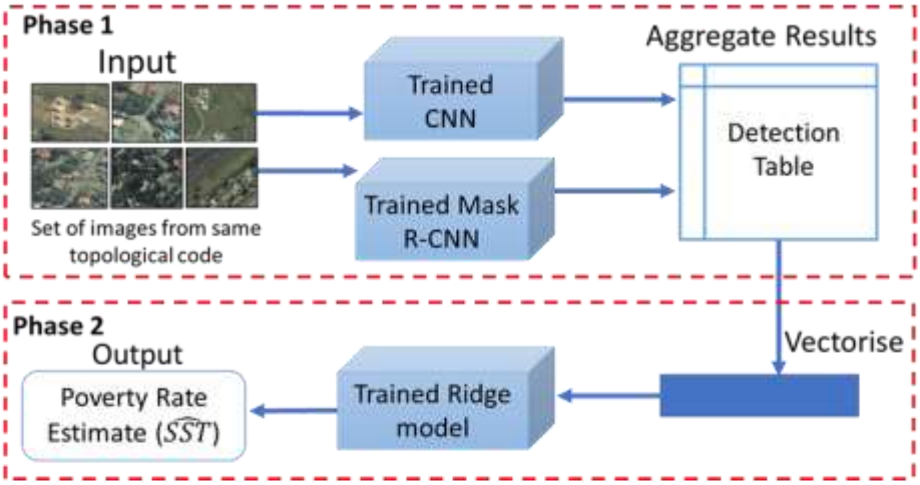


Fig. 6. Architecture of the final poverty estimation model.

I. RESULTS

A. *Phase 1 results*

The models in phase one perform significantly well, Table II and Table III summaries phase one result. The geo-type classification model classifying images into three geo-type classes has a high accuracy rate of 0.9085 and a relatively low log-loss of 0.5783. The accuracy rate indicates that the classification model has correctly classified 90.85% of the images. The log-loss captures the uncertainty of the classifications made based on the amount of divergence between the predicted probabilities of class from the actual class labels. The low log-loss of 0.5783 indicates that there is minimal divergence between predicted classification probabilities made by the geo-type classification model and the actual class labels.

The dwelling type detection model has a log-loss of 0.839 which is the sum of the class label loss, bounding box loss and mask loss. From Table III, the class label loss is the lowest loss with a loss of 0.219, followed by the bounding box and mask loss with a loss of 0.220 and 0.401, respectively. These three losses each indicate that there is minimal divergence between the predicted probabilities and actual values of the class labels, bounding box tuple and binary mask, respectively.

TABLE II. PERFORMANCE RESULTS OF THE GEO-TYPE CLASSIFICATION MODEL

Metric	**Geo-type Classification Model**
Accuracy rate	0.9085
Log loss	0.5783

TABLE III. PERFORMANCE RESULTS OF THE DWELLING-TYPE DETECTION MODEL

Metric	**Dwelling Type Detection Model**
Log loss	0.839
Class label loss	0.219
Bounding box loss	0.220
Mask loss	0.401

B. *Phase 2 results*

The performance of the ridge regression model has been summarised in Table IV. The optimal tuning parameter, λ = 0.032, was obtained by using ten-fold cross-validation. This optimal tuning parameter corresponds to the minimum cross-validation error of the model. The resultant ridge regression model with the optimal λ produces an R^2 of 0.708 which means that the geo-type dwelling type variables explains 70.80% of the variation in the SST poverty rate. Furthermore, the root mean square error (RMSE) of the ridge model is 0.081 which indicates that the ridge model fits the data well.

There is a positive strong correlation (r) of 0.838 between the estimated SST and the observed SST, indicating the strong ability of the model to estimate SST from the classification of geo-types and detection of dwelling types from aerial images. This can also be seen in Fig. 7, where the points lie relatively close to the red 45° line. Observations that lie on the red line have been accurately estimated. From Fig. 7, it is observed that the model tends to over-estimate (data points lie above the red line) the SST value when there is low poverty (low SST) and under-estimate estimate (data points lie below the red line) when there is high poverty (high SST).

TABLE IV. PERFORMANCE RESULTS OF THE RIDGE REGRESSION MODEL.

Metric	**Ridge Regression Model (λ = 0.032)**
R^2	0.708
r	0.219
RMSE	0.220

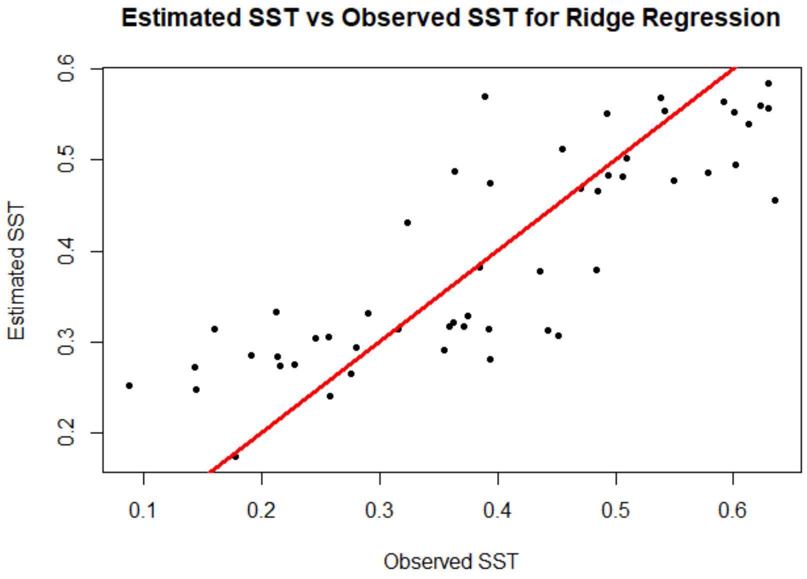


Fig. 7. Estimated SST vs Observed SST for NIDS wave 4.

I. CONCLUSION

This paper introduced an approach to estimate poverty from aerial images by using deep learning computer vision models coupled with a statistical regression model. This approach involves detecting the geo-types and dwelling-types from aerial images and then aggregating the results to estimate the poverty index, SST, which is a composite poverty measure that captures a robust view of poverty. The application of this approach was demonstrated by using aerial images of KwaZulu-Natal, South Africa. This approach displays great potential in becoming a tool that the government can use to efficiently measure, monitor and analyse the poverty rate as they implement policies targeted to alleviate poverty. Considering other features that can be detected or identified using computer vision such as mines, police stations, hospitals, water sources, road type, etc. can be used to improve the model to estimate poverty.

II. ACKNOWLEDGEMENTS

We would like to acknowledge DataFirst and National Geo-spatial information (NGI) for the provision of survey data and aerial images respectively and for the data support from Lynn WoolFrey and Alex Montgomery.

REFERENCES

[1] D. Narayan, R. Patel, K. Schafft, A. Rademacher, and S. Koch-Schulte. *Voices of the poor: can anyone hear us?*. New York: Oxford University Press. 2000. [Online]. Available: http://documents.worldbank.org/curated/en/131441468779067441/Voices-of-the-poor-can-anyone-hear-us.

[2] National Planning Commission. *Our future-make it works*. 2011.

[3] H. Bhorat, M. Leibbrandt, M. Maziya, S. van der Berg, and I. Woolard. *Fighting Poverty-Labour Markets and Inequality in South Africa*. UCT Press. 2001

[4] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon. "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping." 2015. CoRR abs/1510.00098. http://arxiv.org/abs/1510.00098.

[5] R. Engstrom, S.H. Jonathan, and D.L. Newhouse. "Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being". Washington, D.C: World Bank Group. 2017. [Online]. Available:

http://documents.worldbank.org/curated/en/610771513691888412/Poverty-from-space-using-high-resolution-satellite-imagery-for-estimating-economic-well-being.

[6] Statistics South Africa. *National Poverty Lines*. Technical report, Pretoria: Statistics South Africa. 2018.

[7] J. Haughton, and R. Khandker. *Handbook on Poverty and Inequality*. Washington, DC: The World Bank. 2009.

[8] P. Govender, N. Kambaran, N. Patchett, A. Ruddle, G. Torr, and N. Zyl. “Poverty and Inequality in South Africa and the world.”. doi:10.4314/saaj.v7i1.24511. 2007.

[9] A. Sen. “Poverty: An Ordinal Aproach to Measurement.” Econometrica, 219-231. 1976.

[10] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press. 2016. [Online]. Available: http://www.deeplearningbook.org.

[11] M.D. Zeiler, and R. Fergus. “Visualizing and Understanding Convolutional Networks.” CoRR abs/1311.2901. 2013. [Online]. Available: http://arxiv.org/abs/1311.2901.

[12] K. He, G. Georgia, P. Dollar, and R. Girshick “Mask R-CNN.” CoRR abs/1703.06870. . 2017. [Online]. Available: http://arxiv.org/abs/1703.06870.

[13] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer New York Heidelberg Dordrecht. 2013.

[14] A.E. Hoerl, and R.W. Kennard. “Ridge Regression: Biased Estimation for Problems Nonorthogonal.” Technometrics 55-67.

[15] Southern Africa Labour and Development Research Unit. 2016. “National Income Dynamics Study 2014 - 2015, Wave 4 [dataset]. Version 1.1.” Cape Town: Southern Africa Labour and Development Research Unit [producer]. (Cape Town: DataFirst [distributor], 2016. Pretoria: Department of Planning Monitoring and Evaluation [commissioner], 2014). 1970.

[16] University Michigan State. n.d. *South Africa 1:50,000*. [Online]. Available: https://lib.msu.edu/branches/map/findingaids/SouthAfrica50k_KZN/.

[17] Tzutalin. labelimg. 2015.

[18] V. Kshirsagar, J. Wieczorek, S. Ramanathan, and R. Wells. “Household poverty classification in data-scarce environments: a machine learning approach.” 2017. arXiv preprint arXiv:1711.06813.