MAPPING THE TEMPORAL DYNAMICS OF SLUMS FROM VHR IMAGERY
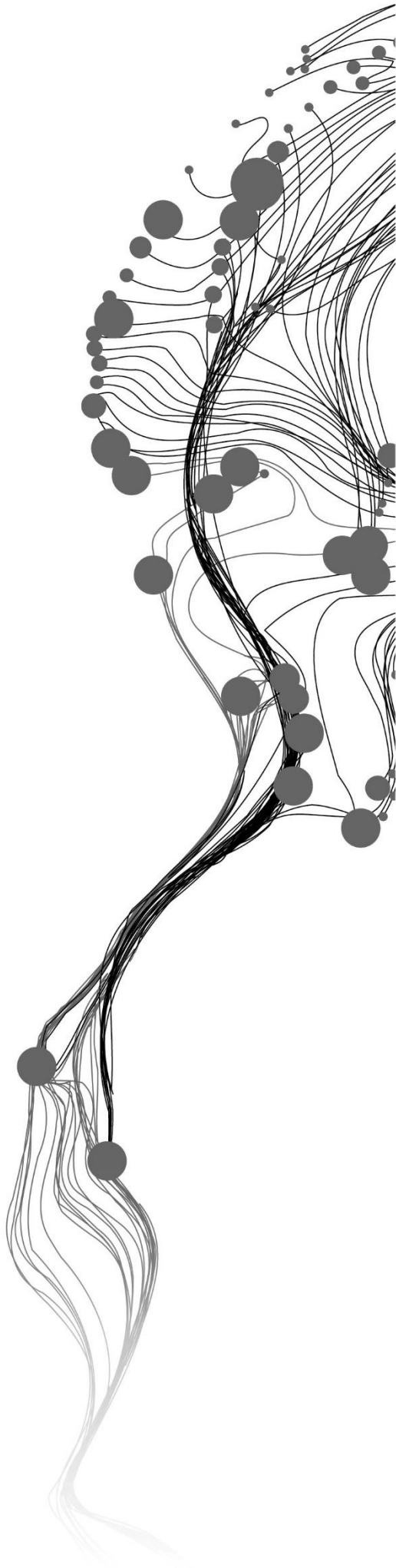
RUOYUN LIU
Enschede, The Netherlands, February 2018

SUPERVISORS:
Dr. M. Kuffer
Dr. C. Persello

MAPPING THE TEMPORAL DYNAMICS OF SLUMS FROM VHR IMAGERY

RUOYUN LIU
Enschede, The Netherlands, November 2018

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Urban Planning and Management

SUPERVISORS:
Dr. M. Kuffer
Dr. C. Persello

THESIS ASSESSMENT BOARD:
Prof. dr. R.V. Sliuzas (Chair)
Dr. M. Netzband (External Examiner, University of Wuerzburg)

ABSTRACT

As the urban population is increasing rapidly, the growth and persistence of slum settlements in the cities have become an important issue to be addressed. Remote sensing imagery is one of the common data sources for producing slum maps. Studies have been developed for the different purpose of slum mapping based on remote sensing method. However, only a few studies have analysed the temporal dynamics of slums. This study aims to explore the potential of using machine learning algorithm to analyse the temporal dynamics of temporary slums based on the very high resolution (VHR) imagery in Bangalore, India. The study proposes two fully convolutional networks (FCNs) based approaches to generate slum change maps and assesses their performance. The study takes the advantage of machine learning and develops two approaches applying FCNs architecture with dilated convolutions to classify the images. For one approach, the resulted slum maps from the land cover classification are used for post-classification change detection. For another approach, the FCNs is used to directly classify the changed slum areas in the image. The performance of 3×3 kernel and 5×5 kernel for the networks in both of the approaches are examined. After producing the change maps for temporary slums, the temporal dynamics are analysed. It is found that 7,173 m2 of land changed into temporary slums in our study area per year from 2012 to 2016, while 8,390 m2 of the existed temporary slums disappeared per year. Most of the slums appeared on the vacant land and disappeared into green land. The accuracies of the change maps are assessed by a confusion matrix and trajectory error matrix (TEM). The post-classification results obtained 53.80% for F1-score, while the change-detected networks results obtained 53.68%. For TEM, post-classification results scored higher for overall accuracy but lower for the accuracy difference of change trajectory than the change-detected networks results. The study concludes that FCNs-based slum classification can have high accuracy in the city of Bangalore. However, using these classification results for post-classification cannot generate very accurate change maps based on the assessment of the confusion matrix. The FCNs-based change-detected networks cannot produce accurate change maps in terms of the size as well. But, both of the two approaches give an accurate location of where the change is. This shows a potential of using machine learning algorithms to detect the change location of slums in VHR imagery.

Keyword: slum, fully convolutional networks (FCNs), very high resolution (VHR) imagery, change detection

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background justification

The developing world is experiencing rapid urbanization. In 2018, an estimated more than half of the world's population resided in urban settlements and by 2050 urban areas are expected to house 68% of people globally (UN-DESA, 2018). However, lack of cities' capacity to meet this sharply increasing housing demand coming together with the inability to provide infrastructure and basic service brings out the growth and persistence of slums (Kohli, Sliuzas, Kerle, & Stein, 2012). The definitions of slums vary across the world. As globally commonly used definition, UN-Habitat has defined that a slum is characterized by lack of one or more of the following: durable housing, sufficient living space, easy access to safe water, access to adequate sanitation and security of tenure (UN-Habitat, 2007). Upgrading slums to ensure access to adequate and affordable housing and basic services has become one of the targets to realize the Sustainable Development Goals (SDGs) by the United Nations (United Nations, 2015).

To address slum issues, slum maps provide information about spatial characteristics of slum locations, extents and structures. Assisted by a slum map, the government or local authority can improve the accessibility and availability of infrastructures in slums, e.g., some governments are not providing basic services and infrastructures as they have no awareness of the existence of slums (Mahabir et al., 2016), and even ignore the existence of slums (Beukes, 2015). It can also help to prioritize the areas which need to be upgraded (Kuffer, Pfeffer, & Sliuzas, 2016). With the development of the remote sensing technology, satellite imagery has become a common data sources for producing slum maps. However, most indicators of slums as defined by UN-Habitant cannot be mapped directly in the satellite image. Therefore, researchers have worked on the conceptualization of slums based on images, e.g., in form of the generic slum ontology (GSO)(Kohli et al., 2012). This proposed GSO provides a framework of identifying slums at three levels of the built environment morphology: the environs level, the settlement level and the object level.

Image based conceptualization of slums often refer to building characteristics, such as roof materials, shape and density (Kuffer, Pfeffer, & Sliuzas, 2016). Such characteristics can be used for slum identification from remote sensing imagery, while some other abstract variables are not directly reflected in the images, for instance, land-tenure rules, distribution of wealth and power, market mechanisms and social customs (Rindfuss & Stern, 1998). For instance, in Bangalore, slums are characterized by limited space between each shelter and a jumbled pattern of units (Krishna, Sriram, & Prakash, 2014). Slums in Sao Paulo State are featured with small roof size, high density and limited green space (Novack & Kux, 2010). With these physical characteristics, it is doable to detect, identify and even monitor slums from remote sensing imagery. This would complete slum information provided in the national census, knowing that this data is often very uncertain, e.g., they often cover only part of the slums (Ranguelova et al., 2018). Compared with the census method, remote sensing is less labour and time consuming. Moreover, remote sensing methods offer slum information at higher temporal resolution while the temporal gap between two census datasets is commonly 10 years, extending to several decades in some cases (Mahabir, Croitoru, Crooks, Agouris, & Stefanidis, 2018). Recently, an increasing number of very-high-resolution (VHR) sensors are more available, thus VHR imagery is becoming a new data source with the opportunity of slum identification at settlement as well as dwelling level.

There are three main study purposes of slum mapping based on remote sensing method: where, when and what (Kuffer, Pfeffer, & Sliuzas, 2016). "where" is about the location of the slums in urban region. "when" is to measure the temporal changes of slums. And "what" is related to the questions such as the population of slums (Kit, Lüdeke, & Reckien, 2013) and allocation of basic service in slum areas (Gruebner et al., 2014). Unlike the other two aspects, only a few studies have been performed to analyse the temporal dynamics of slums. Examples are the automated identification of change patterns of slums in Hyderabad (Kit & Lüdeke, 2013) and the change detection of Kibera informal settlements (Veljanovski, Kanjir, Pehani, Otir, & Kovai, 2012). One reason for the lack of studies is the availability of data and the required local knowledge (Kuffer, Pfeffer, & Sliuzas, 2016), but also the complexity to produce change detection results (Pratomo, Kuffer, Kohli, & Martinez, 2018). For example, the change captured might be the real change but the pixel differences caused by image conditions. A further issue relates to the transferability of mapping methods across multi-temporal images. Transferability is the ability to transfer the method or algorithm developed in one image to another image and achieving comparable mapping accuracies (Kohli, Warwadekar, Kerle, Sliuzas, & Stein, 2013). It is a key point, but also a main bottleneck, to realize the automated slum mapping globally (Sliuzas, Kuffer, Gevaert, Persello, & Pfeffer, 2017).

1.2   Research problem indentification

As mentioned above, not many studies have analysed the temporal dynamics of slums and none of them has used machine learning methods. This thesis will focus on developing a transferable slum mapping approach that allows mapping slums in multi-temporal VHR imageries.

Researchers have been working on various approaches for slum identification based on VHR imagery, including: texture analysis (Kuffer, Pfeffer, Sliuzas, & Baud, 2016); object-based image analysis (Hofmann, Strobl, Blaschke, & Kux, 2008); landscape analysis (H. Liu, Huang, Wen, & Li, 2017); and machine learning (Duque, Patino, & Betancourt, 2017).Convolutional Neural Networks (CNNs), which are specific technique in the machine learning field, have drawn increasing attention in solving remote sensing classification tasks and tended to have higher accuracy than other methods when aiming at extracting slum areas at the city scale (Kuffer, Pfeffer, & Sliuzas, 2016). CNNs can extract image features by itself instead of being provided by handcrafted features (Nielsen, 2015). Mboga, Persello, Bergado, & Stein (2017) presented that CNNs had a better performance than Support Vector Machine (SVM) algorithm with Grey-Level Co-Occurrence Matrix (GLCM) features in informal settlements identification. Fully Convolutional Networks (FCNs) for semantic image segmentation is a particular case of CNNs (W. Sun & Wang, 2018). By replacing the fully connected layers in a CNNs architecture into a convolution layer, FCNs maintain the structure of the original image (Fu, Liu, Zhou, Sun, & Zhang, 2017). Unlike CNNs, in which the output must be the same size as the input, FCNs allows taking images of any size as an input (Zhu et al., 2017). The study of Persello & Stein (2017) has shown that slums can be effectively detected in VHR images by FCNs technique. However, FCNs have not been used for analysing the temporal dynamics of slums.

This study intends to analyse the potential of transferring a FCN-based classifier trained to identify slums from time to time and also from image to image. Therefore, temporal dynamics and changes will be detected with the help of the developed approach.

1.3 Research objective

1.3.1 General objective

In this study, the main research objective is to develop a FCNs-based approach to map slums and analyse their temporal dynamics using VHR imagery.

1.3.2 Sub-objectives

i. To identify slum and non-slum-area from VHR imagery by applying fully convolutional networks (FCNs).
ii. To analyse the temporal dynamics based on the resulted slum maps.
iii. To evaluate the outcomes of change detection for temporal dynamics.

1.4 Research questions

1. To identify slums and non-slum areas from VHR imagery by applying convolutional neural networks method.
   - What are the physical and morphological characteristics of slums in Bangalore?
   - What is the best strategy to create samples for training, validation and testing?
   - What is the optimal FCN architecture to identify slum-areas in terms of accuracy and computational costs?
2. To analyse the temporal dynamics based on the resulted slum maps.
   - What is a suitable method to extract the temporal dynamics of slum?
   - What change characteristics can be observed from the slum maps?
3. To evaluate the outcomes of change detection for temporal dynamics.
   - What are the optional methods to assess the accuracy of multi-temporal change detection outcomes?
   - What is the assessed accuracy of the mapped temporary dynamics of slums?

2. LITERATURE REVIEW

This chapter reviews the thesis-related literature. The first section overviews the efforts in image-based slum mapping. In the second section, the basic concepts of CNNs and FCNs and their application in the urban remote sensing field are presented. The next sections provide a review of transfer learning and domain adaptation in satellite image classification. This chapter ends by the summarizing change detection methods for slum identification.

2.1 Image-based slum mapping

Many efforts have been made to establish a general objective measurement for slums, in practice, the definitions of slum vary from city to city globally. For examples, in Egypt, slums have been redefined by two distinctive terms: "Unsafe areas" and "Unplanned areas" (Khalifa, 2011). While the Egyptian "Unplanned areas" are characterized by its non-compliance to planning and building laws and regulations, the slums in Romania are often former worker's houses (Iacoboaea, 2009). The concept of "slum" can be regarded in a relative way. It can be viewed differently according social class, culture and ideology (Gilbert, 2007). Therefore, most studies of slums have three different lines of direction: social-economic and policy, physical characteristics using approaches such as remote sensing and slum modelling using approaches such as cellular automata (Mahabir et al., 2016). With the improved image data resolution and methodological advances, remote sensing studies are able to provide more information about slums. Compared with census-based data, remote sensing image data can provide a synoptic view with the ability to capture the situation on the ground (Mahabir et al., 2018). And recently, the increasing availability of high- and very high-resolution (H-/VH-R) imagery offers an opportunity to study slums with more spatial details, making the identification of slums from large settlement scale to small individual dwelling scale possible.

In literature, several methods are commonly used to identify slums areas from VHR imagery. Object-based image analysis is one of the commonly used methods. It partitions imageries into meaningful objects and then assess their characteristics. These objects are the generations of geographic information, and assess their characteristics (Blaschke et al., 2014). In object-based image analysis (OBIA), the image is treated as a set of objects rather than pixels. Apart from the original spectral information of image, other properties like the object size, shape, texture and the relationship with the neighbouring objectives (Giada, De Groeve, Ehrlich, & Soille, 2003) are also used. While pixel-based image classification assigns pixels with similar spectral reflectance into same class, object-based classification segments the image into a set of objects as a result of variations in physical characteristics of different classes. Some literatures have proved that this method provides several improvements over the pixel-based classification. For instance, Q. Yu et al. (2006) found OBIA overcame the salt-and-pepper effect problem in traditional pixel-based approach for vegetation classification in the study area of Northern California. OBIA can also emulate human interpretation and reflect the objects in real life better (Hay, Blaschke, Marceau, & Bouchard, 2003). In slum mapping, the accuracies of OBIA vary a lot. OBIA was found to have good performance when extracting objects, like roof and roads, while it has lower accuracies when the urban environment is complex and slums characteristics are hardly to be captured (Kuffer, Pfeffer, & Sliuzas, 2016).

As OBIA has difficulty in extracting slums from complicated urban environment, machine learning technique has been applied to slum mapping. It uses training samples from the images to learn how to identify different patterns in order to solve the classification problem (Richards & Jia, 2006). Various machine learning algorithms have been performed to identify slums, for instance, Random Forest (Wurm, Weigand, Schmitt, Gei, & Taubenbock, 2017) and Support Vector Machine (Leonita, Kuffer, Sliuzas, & Persello, 2018). Researchers aslo applied other machine learning based algorithms to address this problem. Markov Random Field is one of them (Graesser et al., 2012). Another machine learning algorithm, which is becoming increasingly popular, is Convolutional Neural Networks. The detail of this algorithm is going to be discussed in the next chapter.

2.2 FCNs based slum mapping

2.2.1 Background

The Convolutional Neural Networks belongs to Artificial Neural Networks (ANNs, which is an advanced algorithm in computer science inspired by the human biological neuron (Atkinson & Tatnall, 1997). An ANNs architecture usually has three main layers: input layer, hidden layer and output layer (*Figure 1*). Every neuron in each layer is connected to all neurons in the next layer. In the learning process, a weighted sum (y_i) of one neuron is calculated with the input (x_i), weight (w_i) and bias (b_i), explained in equation 1 (Stanford University, 2018).
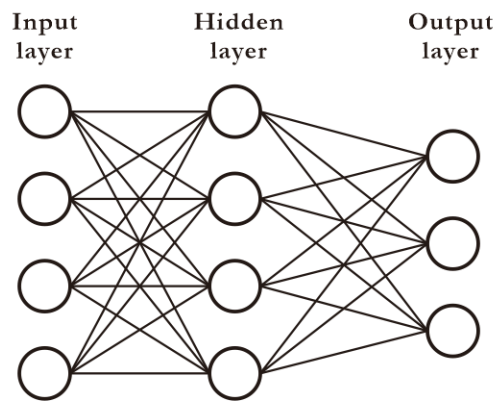
Figure 1: Simple ANNs architecture

$$y_i = \sum_{i}^{n} w_i \cdot x_i + b_i \qquad \text{(equation 1)}$$

The y_i of the neuron will be activated by an activation function and the most commonly used activation

Figure 2: Simple illustration of convolution

functions in ANNs are sigmoid, hyperbolic tangent function (tanh) and Rectified Linear Unit (ReLU) (Nielsen, 2015). Training the networks means tuning the weight and the bias for each neuron into a final result that the network can identify different classes.
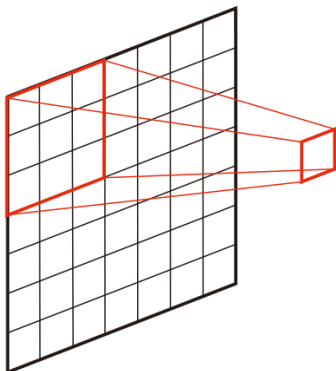
Deeper networks have several hidden layers in order to solve more complex problems. CNNs, which is a branch of deeper ANNs, employs two specific hidden layers: convolutional layer and fully-connected layer. During the convolutional operation in one convolutional layer, the input is downsampled by the filters (*Figure 2*), resulting in a reduction of the connection numbers as well as the parameter numbers. Therefore, the contextual information can be extracted through this process.

The standard CNNs classify images in a "patch-based" mode, labelling every central pixel in the patches extracted from the input (Bergado, Persello, & Gevaert, 2016). As CNNs generates the possibility distribution of different classes, in order to get a classification map with various classes, a large image is usually separated into small patches, where CNNs are applied to predict the class. However, as remote sensing images consist of a large amount of information, using CNNs to classify large remote sensing images will have a high computational cost because of the patch cropping. To address this issue, the Fully Convolutional Networks (FCNs), which are based on the standard CNNs, have been proposed and

applied in this field. In FCNs, the fully connected layers are replaced by the convolutional layers, which allow to use discretionary sized images as an input. By training the entire image instead of training the patches separated from the image, FCNs reduce the computation operations as well as the implementation complexity (Fu et al., 2017).

2.2.2   Application

A lot of complex Artificial Neural Networks have been designed in the field of computer vision and pattern recognition (CVPR) to solve different problems. Examples are AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), VGG (Chatfield, Simonyan, Vedaldi, & Zisserman, 2014) and GoogLeNet (Szegedy, Liu, et al., 2015). In the last decade, researchers started carrying out studies using CNNs in the analysis of remote sensing imagery. Castelluccio, Poggi, Sansone, & Verdoliva (2015) used pre-trained CNNs adopted from CaffeNet and GoogLeNet to classify land use classes. CNNs has also been used in the land cover classification research (X. Sun, Shen, Lin, & Hu, 2017). And for slum mapping, both CNNs (Mboga et al., 2017) and FCNs (Persello & Stein, 2017) showed promising results with overall accuracies over 80%.

2.3   Transfer learning and domain adaptation

The aim of transfer learning is to extract the knowledge learned from one or more source tasks and then applied it to a target task (Pan & Yang, 2010). Transfer learning techniques have been used in several studies about satellite image classification. Liu & Li (2014) proposed a model using old domain data to train a classifier for mapping the land use types of a target domain. Transfer learning has been used in the monitoring and analysis of urban villages in China with the use of landscape metrics (H. Liu et al., 2017). Although using trained CNNs for extracting features from high-resolution imagery via transfer learning is realized in the land-use classification field (Akram, Laurent, Naqvi, Alex, & Muhammad, 2018), the framework of using CNNs and transfer learning in slum mapping is still a gap. Besides, several transfer-learning problems have been considered in the literature, including domain adaptation, multitask learning, domain generalization, sample selection bias, and covariate shift (Pan & Yang, 2010).

Domain adaptation (DA) is a rising field of investigation in remote sensing. The purpose of DA is to overcome the shifts between input variables and the associated labels between the source and target domains (Matasci, Volpi, Kanevski, Bruzzone, & Tuia, 2015). In remote sensing field, when the source and target domain are related to two images acquired one the same geographical area at two different times, DA will be useful for image analysis (Persello & Bruzzone, 2012). It can reuse the available ground truth samples to classify new image that may be at different time instants and with different sensors (Tuia, Persello, & Bruzzone, 2016). With the help of it, we can use the scarce labelled data to classify multi-temporal images (Jean et al., 2016), providing the resulted base maps for further change detection analysis.

2.4   Change detection

Singh (1989) defined Change Detection as "the process of identifying the changes in remote sensing images that cover the same area of the earth surface in two different times". Many change detection methods have been performed in different studies. These methods are also been categorised in different ways by different researchers. Civco, Hurd, & Wilson (2002) identified four types for the method: 1) traditional post-classification; 2) cross-correlation analysis; 3) neural networks; 4) image segementation. Pacirici, Solimini, Del Frate, & Emery (2007) catogorised methods into two main group: unsupervised and supervised. Depending on the analysis unit, Tewkesbury, Comber, Tate, Lamb, & Fisher (2015) divided remote sensing change detection methods into six types: 1) layer arithmetic; 2) post-classification change; 3) direct classification; 4) transformation; 5) change vector analysis; and 6) hybrid change detection.

For VHR imagery, post-classification is one of the most established and wildly used change detection method (Tewkesbury et al., 2015). Hester, Nelson, Cakir, Khorram, & Cheshire (2010) generated post-classification land cover change maps in the study area of North Carolina from QuickBird images and presented a fuzzy framework for transforming map uncertainty into change analysis. Boldt, Thiele, & Schulz (2012) proposed a workflow using QuickBird images to detect urban change areas by the post-classification method. However, the biggest problems with post-classification method is the complete dependency on the input maps quality (Lu, Mausel, Brondízio, & Moran, 2004).

Direct classification method only requires one classification stage, as it directly identifies the changes occurring in the study area. Tewkesbury et al. (2015) suggested that direct classification is a cogent tool in the context of data mining problems and is an ideal scenario for machine learning algorithms. Some studies used this strategy to detect changes. For instance, Schneider (2012) presented an approach to capture urban changes from dense time stacks of imagery using boosted decision trees and support vector machine algorithms. Gao et al. (2012) also uses this strategy to map impervious surface expansion using the decision tree algorithm. In this study, we would also apply direct classification method based on the FCN algorithms to analyse the temporal dynamics of slums.

3.   STUDY AREA AND DATA DESCIRPTION

3.1   Study area

Bangalore is one of the biggest cities in India, holding more than 8 million population in the metropolitan area (Government of India, 2011). As more than 1300 ICT-companies (Information Technology and Communication) going about in the city (Dittrich, 2005), Bangalore has become the Silicon Valley of India. However, this development was mainly due to massive foreign investments, resulting in a highly competitive framework of inter-city (Dittrich, 2005). A highly fragmented and polarized urban society has been generated (Dittrich, 2005). The India census in 2011 reported that around 8.39% of total population in Bangalore city living in the slums (Census Organization of India, 2015). However, a recent research suggested that every fifth person in the city of Bangalore lives in a slum (Roy, Lees, Pfeffer, & Sloot, 2018). The difference is mainly caused by the different definitions of the slums, as well as their highly temporal dynamics. Besides, India also sets minimum settlement size for an area to be considered as a slum, requiring at least 3000 population or 60 households living in a settlement cluster.[1] Slum settlements are a big challenge which the city should address (Rains, Krishna, & Wibbels, 2017).

There are two types of officially identified slums: notified slums and non-notified slums (*Figure 3*). While notified slum dwellers do not merely survive but also invest in education and skill training, residents in non-notified slums are mostly unconnected to basic service and formal livelihood opportunities (Krishna et al., 2014). Krishna also categorized non-notified slums in Bangalore into three types: new migrants; very low-income settlements; and low-income settlements. In this hierarchy, "new migrants" is shelters typically characterized by blue plastic sheeting and small unite size. People live in these shelters require access to electricity, clean drinking water, livelihood and property security (Krishna et al., 2014).

(a) Notified slum   *(b) Non-notified slum ("new migrants")*

Figure 3: Slums in the city of Bangalore, Source: (Krishna et al., 2014)

In India, legal notification or designation is very important for the recognition of slums by the government, as this is the sign that government will afford the shelters rights to the provision of clean water and sanitation (Nolan, 2015). The first step in upgrading and transforming these shelters to areas of basic living conditions is the identification (Rains et al., 2017). Besides, these temporary slums have high temporal dynamics. An example shows in the *Figure 4*. A slum area can be seen from the satellite image on 2015.12.17. Within 100 days, this slum area decreased sharply, indicating that temporary slums in Bangalore can experience rapid change within few months, even weeks. Monitoring those slums with a

---

[1] http://nbo.nic.in/Images/PDF/SLUMS_IN_INDIA_Slum_Compendium_2015_English.pdf

high temporal granularity can help local planner understand their movements and hence provide help with target. Thus, this study will focus on Bangalore to explore potentials of using automated slum identification method, in order to analyse their temporal dynamics.

| (a) 2015.12.17 | (b) 2016.01.25 | (c) 2016.03.21 |

Figure 4: Example of one rapidly changing slum area, source: Google Earth

3.2 Data description

The basic data for this study is multi-temporal very-high-resolution imageries provided by the project Dynaslum (Netherlands eScience Center, 2018). All multispectral images from the WorldView satellites are with eight bands: Blue, Green, Red, Near Infrared 1, Coastal, Yellow, Red Edge and Near Infrared 2. Pan-sharpened images are used in this study. A summary of image dataset is shown in *Table 1*.

Table 1: Summary of image dataset used in this study

Satellite	Resolution	Band number	Time
Worldview 2	0.5 × 0.5 m (multispectral)	8 bands	2012. 12. 01
	2.0 × 2.0 m (panchromatic)		2013. 04. 24
Worldview 3	0.3 × 0.3 m (multispectral)	8 bands	2015. 02. 16
	1.2 × 1.2 m (panchromatic)		2016. 01. 06

Slum boundary data delineated by experts using visual interpretation and field verification in 2017 is also available in the study. However, these boundary data was generated for a specific date, not matching with the available image data.

4. METHODOLOGY

This chapter describes the methodology of this research. Experiments are carried out towards the sub-objectives in the study. The flowchart in *Figure 5* illustrates the general approach briefly.
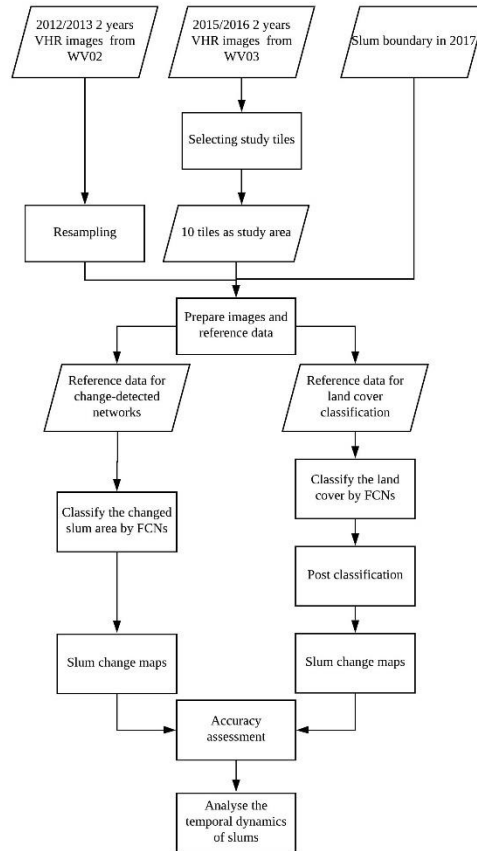
Figure 5: Flowchart of the methodology

4.1 Pre-processing of the data

4.1.1 Resampling

The pre-processing of the satellite images and reference data was performed at the beginning. As images from 2 satellites have different resolution, images from 2012 and 2013 with the multispectral resolution of 0.5×0.5 m FCNs optimization were resamples to 0.3×0.3 m, same as the multispectral resolution of the 2015 and 2016 images. Therefore, every pixel could represent the same geographical area in different images.

4.1.2 Selection of study tiles

Studies related to the extraction of slums often worked with smaller areas or tiles. Persello & Stein (2017) worked with five tiles of 2000×2000 pixels and Kit & Lüdeke (2013) also started with an urban subarea of 60×60 m (100×100 pixels). As the later classification process was performed in MATLAB, where the images were stored as numeric arrays. In consideration of the capacity of the machine which was used for this study,

10 specific tiles, each with 1000×1000 pixels, were selected (*Figure 6*, details in Annex 1). The selections were based on three rules:

(1) Tiles were covered by all image data. Due to the data limitation, the images of four years were not covering the exact same area. As this study intended to analyse the temporal dynamics of the slums in Bangalore from 2012 to 2016, the selected tiles should be covered by the images of four years.

(2) Slums existed in the selected tiles. The judgement of whether the tile had slums or not was made by the help of slum boundary data delineated by experts in 2017 as well as the visual check. The slum boundary data will firstly give the evidence of where may have slums. Then, a check of the existence of those slums in 2016 was carried out visually.

(3) Slums in the selected tiles had changes between 2012 and 2016. The temporal dynamics of slums can only be captured if changes happened to slums during the four years.

Figure 6: Distribution of study tiles

4.2 FCN-based land cover classification

4.2.1 Reference data preparation

As mentioned in chapter 3, the data of slum boundary delineated by experts was available. After a visual check of slum polygons on tip of the used images, it was found that most of the slum boundaries were not accurately showing the outlines of slums (*Figure 7*), which would cause problems to further steps. It is because the boundaries were delineated using different satellite imagery of different time. As reviewed in Chapter 3, slums could experience rapid changes. Therefore, visual interpretations were performed to each selected tile for four years in order to generate reference data. The reference maps contained five land cover classes, namely "temporary

Figure 7: Not matched Slum boundary data example

slum", "green land", "vacant land", "formal built-up" and "other" (shown in *Table 2*). Non-labelled cells are also included in each tile.

Class	Description	Label
Temporary slum	Tents with blue plastic sheeting and small unite size	1
Green land	Open land covered by vegetations	2
Vacant land	Bare soil land	3
Formal built-up	Formal buildings, roads	4
Other	Car park, water body…	5

Table 2: Land cover class for reference data

4.2.2 FCNs architecture – 5x5

The FCNs built in this study uses the architecture from Persello & Stein (2017) as the foundation. The architecture consists six convolutional layers, followed by a final classification layer with a 1×1 convolution layer and a softmax loss function. The structure of this architecture is shown in the *Table 3*.

Table 3: Structure of the 5×5 FCNs architecture

Layer	Module type	Dimension	Dilation	Stride	Pad
DK1	convolution lReLUs	$5 \times 5 \times 8 \times 16$	1	1	2
DK2	convolution lReLUs	$5 \times 5 \times 16 \times 32$	2	1	4
DK3	convolution lReLUs	$5 \times 5 \times 32 \times 32$	3	1	6
DK4	convolution lReLUs	$5 \times 5 \times 32 \times 32$	4	1	8
DK5	convolution lReLUs	$5 \times 5 \times 32 \times 32$	5	1	10
DK6	convolution lReLUs	$5 \times 5 \times 32 \times 32$	6	1	12
Class.	convolution softmax	$1 \times 1 \times 32 \times 5$	1	1	0

In this study, first, a network with the kernel size as 5×5 was trained and validated. Then, a deeper network with the 3×3 kernel size was used to see the result being improved or not.

The convolution layers in the architecture calculated the convolution of the input images of selected tiles, where the kernel size of the filter was 5×5. Stride was the spatial interval between the centre of convolutional calculation, while 1 meant there was no downsampling procedure. The number of Pad determined the number of zeros adding to the border of the image before performing the filter. The most important idea of this proposed architecture was the adoption of dilated kernels. It increased the receptive filed without increasing the learnable parameters in each layer (F. Yu & Koltun, 2015). A receptive field is the region in the input image that a neuron in the convolutional networks is looking at. Compared to normal kernels, dilated kernels inserted zeros between the elements in the filter. *Figure 8* illustrates how receptive field of a 3×3 filter increased with the increasing dilation factors: (a) a receptive field of 3×3 with dilation factor 1, which meant there was no dilation; (b) a receptive field of 7×7 with dilation factor 2; (c) a receptive field of 15×15 with dilation factor 3. Red circle represent learnable filter weights (Persello & Stein, 2017).

Leaky rectified linear units (lReLUs) was used as activations in the network (Maas, Hannun, & Ng, 2013).
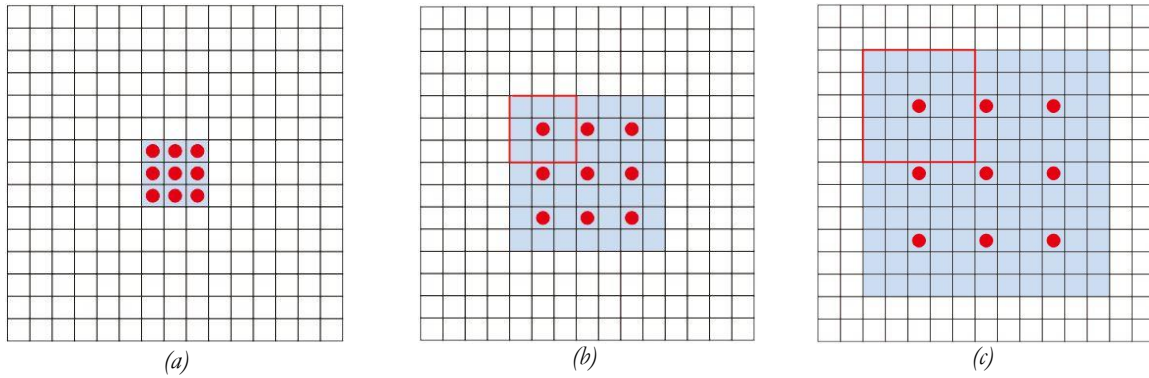
Figure 8: Kernels with increasing receptive field

4.2.3 FCNs architecture – 3x3

After training the network with 5×5 kernel size, a network with 3×3 sized filters was also performed. The structure is shown in the Table 4. In order to keep a same output spatial dimension as the previous network, each block of dilated convolution layers (DK) consisted of two convolution layers, each followed with an activation layer. The second 3×3 convolution layer was fully connected to the first 3×3 convolution, which had a receptive field same with the a 5×5 convolution (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2015). *Figure 9* shows how this work in a mini network. In (a), the first layer is a 3×3 convolution, followed by a fully connected convolution on top of the 3×3 output of the first layer, and at last the receptive field is as same as in the network from (b) with one 5×5 convolution. The setup of (a) leads to a high performance vision networks with a relatively modest computation cost compared to the setup of (b) (Szegedy, Vanhoucke, et al., 2015).
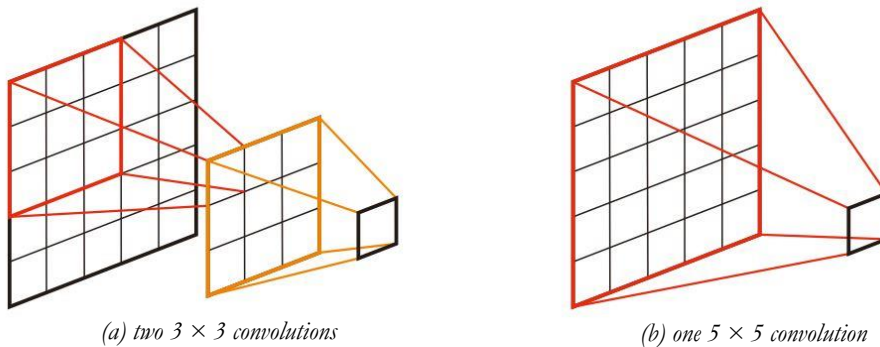
(a) two 3×3 convolutions   *(b) one 5×5 convolution*
Figure 9: Two 3×3 convolutions replacing one 5×5 convolution

Table 4: Structure of the 3 × 3 FCNs architecture

Layer	Module type	Dimension	Dilation	Stride	Pad
DK1	convolution	$3 \times 3 \times 8 \times 16$	1	1	1
	lReLUs				
	convolution	$3 \times 3 \times 16 \times 16$	1	1	1
	lReLUs				
DK2	convolution	$3 \times 3 \times 16 \times 32$	2	1	2
	lReLUs				
	convolution	$3 \times 3 \times 32 \times 32$	2	1	2
	lReLUs				
DK3	convolution	$3 \times 3 \times 32 \times 32$	3	1	3
	lReLUs				
	convolution	$3 \times 3 \times 32 \times 32$	3	1	3
	lReLUs				
DK4	convolution	$3 \times 3 \times 32 \times 32$	4	1	4
	lReLUs				
	convolution	$3 \times 3 \times 32 \times 32$	4	1	4
	lReLUs				
DK5	convolution	$3 \times 3 \times 32 \times 32$	5	1	5
	lReLUs				
	convolution	$3 \times 3 \times 32 \times 32$	5	1	5
	lReLUs				
DK6	convolution	$3 \times 3 \times 32 \times 32$	6	1	6
	lReLUs				
	convolution	$3 \times 3 \times 32 \times 32$	6	1	6
	lReLUs				
Class.	convolution	$1 \times 1 \times 32 \times 5$	1	1	0
	softmax				

4.2.4 Training the networks

Training of the network was accomplished with MATLAB. As mentioned in chapter 4.1.2, ten image patches of 1000 × 1000 pixels were selected as the study tiles. Among those, four tiles were used for training and the rest six for testing. The testing tiles were selected according to two rules:

(1) The training tiles covered all the land cover classes.
(2) Every slum change trajectory was included in the training tiles. This was the preparation for later change detection step.

In total, 40 images with 40 corresponded reference maps (4 images from different time for each tile) were the input data for the networks. 1000 labelled patch randomly picked from each training tile were used as the training set. The networks were trained with a learning rate of 10^{-4} for 100 epochs and a learning rate of 10^{-5} was used to train another 30 epochs. This two-stage training provided a substantial reduction in the training error at the first stage and a more stable training and validation with a lower learning rate at the second stage. Besides, the networks were trained using stochastic gradient descend with a momentum of 0.9.

All the trainings were performed on a desktop workstation with an Intel Xeon E5-2643 v3 CPU and a NVIDIA Quadro GPU.

4.3 Change detection

In this study, we carried out two change detection methods to analyse the temporal dynamics of slums. On the one hand, we used the land cover classification results to perform the post-classification method. On the other hand, we directly trained the FCNs to classify the changed areas in slum.

4.3.1 Post-classification change detection

Post-classification change detection method was employed after the independent land cover classification from FCNs. Each multi-temporal image for every tile was classified separately, with the same category label. Therefore, a land cover change would be detected as a change in the label between two images. In this study, for later analysis, the exact transformation patterns from temporary slums to other land cover class or from different land cover classes to temporary slum were expected. We first reclassified the classification results from different years (*Table 5*). By doing plus operation for raster calculation, every change trajectory would have a unique value. For instance, a pixel with a value of 1234 means that this pixel is classified as temporary slum in 2012, changing into green land in 2013. In 2015, this pixel is classified as vacant land and becomes a pixel of formal built-up in 2016.

Table 5: Land cover class label of classification map after reclassifying

| Year | Land cover class label | | | | |
	Temporary slum	Green land	Vacant land	Formal built-up	Other
2012	1	2	3	4	5
2013	10	20	30	40	50
2015	100	200	300	400	500
2016	1000	2000	3000	4000	5000

4.3.2 Change-detected network

4.3.2.1 Image preparation

Except the post-classification change detection method, we also applied an FCN-based network which directly detect the changed areas of slum. The input images to this network would be stacked images of different years. The images with n bands at one year and m bands at another year were combined into one image with $(n + m)$ bands. In this study, the 1st to 8th bands of the stacked image were from an earlier year image and the 9th to 16th bands were from a later year image at the same tile.

4.3.2.2 Reference data preparation

The reference data for change-detected net was based on the land cover reference data was prepared for all four years in chapter 4.2.1. The reference data consisted of four classes described in the *Table 6*.

Table 6: Class for change-detected net reference data

Class	Description	Land cover in T_1	Land cover in T_2	Label
Increased slum	Temporary slum did not exist in T_1 but appeared in T_2.	Green land Vacant land Formal built-up Other	Temporary slum	1
Decreased slum	Temporary slum existed in T_1 but disappeared in T_2.	Temporary slum	Green land Vacant land Formal built-up Other	2

Unchanged slum	Temporary slum stayed unchanged between T_1 and T_2	Temporary slum	Temporary slum	3
Other	Other land cover	Green land	Green land	4
		Vacant land	Vacant land	
		Formal built-up	Formal built-up	
		Other	Other	
			T_1: An earlier year T_2: A later year	

4.3.2.3   Training the network

The networks used to directly detect the changed slum areas shared a same architecture with the one proposed in the chapter 4.2. We used the same training and testing tiles for the change-detected networks, with newly generated images and reference data. Similarly, a 5 × 5 would be trained and validated at first, followed by a 3 × 3 network, to see the result being improved or not. As the image data became the stacked images with 16 bands, the dimension of the first convolution layer in the network was changed in to 5 × 5 × 16 × 16 (or 3 × 3 × 16 × 16). Besides, the number of classes in the change-detected network reference data was 4. The dimension of the last convolution layer was also changed from 1 × 1 × 32 × 5 into 1 × 1 × 32 × 4. The training was performed separately for every time period. For example, to capture the changed areas between 2012 and 2013, 10 stacked images from 2012 and 2013 and their corresponded reference maps were the input data for the networks.

4.4   Accuracy assessment

In this study, mainly two method to assess the accuracy of classification and change detection results have been applied. One is confusion matrix and another is trajectory error matrix (TEM).

4.4.1   Confusion matrix

The performance of the machine learning based classification results were evaluated by the quantitative indices from confusion matrix, comparing the classification result with the reference data. The Producer accuracy (PA) and User accuracy (UA) were included to reveal the wrong classification of each class. Producer accuracy (calculated using equation 2), which is also explained as precision by Radoux & Bogaer (2017), is the fraction of correctly classified pixels with regard to all pixels of that class in the reference map. The value illustrates how well the pixels in reference map are classified. User accuracy (calculated using equation 3) explained as recall, is the fraction of correctly classified pixels with regard to all pixels of that class in the classified map, illustrating the reliability of classed in the classification map. In these two equations, C_{ii} = number of pixels correctly classified by the class i, C_{+i} = column total of class i, C_{+i} = row total of class i.

$$\text{Producer accuracy (PA)} = \frac{C_{ii}}{C_{+i}} \cdot 100 \qquad \text{(equation 2)}$$

$$\text{User accuracy (PA)} = \frac{C_{ii}}{C_{i+}} \cdot 100 \qquad \text{(equation 3)}$$

In addition, F1-score of the classification result is calculated as well, in order to show a harmonic value balancing precision and recall. The equation of three values are shown in the equation 4.

$$\text{F1score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{PA \cdot UA}{PA + UA} \qquad \text{(equation 4)}$$

4.4.2 Trajectory error matrix

Trajectory error matrix was proposed by (Li & Zhou, 2009) to analyse multi-temporal images. One of the most important idea in this study is classifying the possible trajectory combinations of land cover change into six confusion sub-groups. Pratomo et al. (2018) used this framework to assess the temporal transferability of OBIA rulesets for slum detection. Based on these two studies, we determined our sub-groups in the TEM (*Table 7*). In S_1, both reference data and classification map agree that a sample stayed unchanged. In S_2, both reference data and classification map agree that a sample were changing with a same trajectory, i.e. changing from slum to non-slum and then becoming slum again. In S_3, both reference data and classification result tell that a sample did not changed while the classification result was wrong, i.e. staying unchanged as a non-slum area in reference data while in classification map it stayed unchanged as a slum area. In S_4, reference data suggests a sample as unchanged, but it is a changed area in classification map, while in S_5 is vice versa. Finally, in S_6, both reference data and classification map show changes, but the trajectory is different, i.e. the reference data suggested a sample changed from slum to non-slum and then stayed, while the classification map detected it as a slum changing to non-slum and then becoming slum again.

Table 7: Sub-groups in TEM

Groups	Classification result situation	Interpretations
S_1	Correct	Correctly detected as non-changed with correct classification
S_2		Correctly detected as changed slum with correct trajectory
S_3	Incorrect	Correctly detected as non-changed with incorrect classification
S_4		Incorrectly detected as changed slum
S_5		Incorrectly detected as non-changed
S_6		Correctly detected as non-changed with incorrect classification

After determining the sub-groups, the classification result of land cover was reclassified into binary images, combining the classes of Green land, Vacant land, Formal built-up and Other into a new class of "Non-slum". Similar to chapter 4.3.1, we also assign a unique class value to different years (*Table 8*). The binary classification maps for four years were stacked into one composite map. Therefore, every possible trajectory would have one unique value. For instance, a pixel of 2112 means that this pixel belonged to a non-slum area in 2012 and was classified as slum in 2013 and 2015, finally changed into non-slum in 2016.

Table 8: Land cover class label for TEM

Year	Label	
	Temporary	Non-slum
2012	1	2
2013	10	20
2015	100	200
2016	1000	2000

Then, we generated 500 random points for each tile, in total 5000 points, which were obtained with their corresponding classification and reference data. This information was used as the input to determine the change trajectory with *Table 8*. Based the result, Li & Zhou (2009) proposed two indices to measure overall accuracy: (1) overall accuracy (A_T); (2) change/no change accuracy ($A_{C/N}$), and three indices to measure accuracy difference: (1) overall accuracy difference (OAD); (2) accuracy difference of no change trajectory ($ADIC_N$); (3) accuracy difference of change trajectory ($ADIC_C$). These indices were calculated

using the equations below, where S_i means the number of sample points assigned to different sub-groups of TEM.

$$A_T = \frac{S_1 + S_2}{\sum_{i=1}^{6} S_i} \cdot 100 \qquad \text{(equation 5)}$$

$$A_{C/N} = \frac{S_1 + S_2 + S_3 + S_6}{\sum_{i=1}^{6} S_i} \cdot 100 \qquad \text{(equation 6)}$$

$$\text{OAD} = A_{C/N} - A_T \qquad \text{(equation 7)}$$

$$ADIC_N = \frac{S_1}{S_1 + S_3} \times 100 \qquad \text{(equation 8)}$$

$$ADIC_C = \frac{S_2}{S_2 + S_6} \times 100 \qquad \text{(equation 9)}$$

5. RESULTS

5.1 FCN-based land cover classification

5.1.1 Performance of 5 × 5 network and 3 × 3 network

We trained FCNs from a simple 5 × 5 networks to a deeper 3 × 3 networks. Images from 2012, 2013, 2015 and 2016 for each study tile were the dataset for training and validation (classification results are shown in Annex 2). *Table 9* shows the average F1-scores of the temporary slum class in testing tiles for two networks (accuracy for each tile in Annex 3). Both of the networks performed good when classifying temporary slums in the city, reaching a high accuracy of over 80%.

The best improvement happened to the 2016 classification, the 3 × 3 networks showed a higher accuracy of almost 5%. While in 2013 the 3 × 3 networks had a worse performance, but only 0.5%. On average, accuracy improved by 2% after applying the 3 × 3 networks. Thus, using this deeper network will boost the classification result. However, it requires a high computational ability and learned slower as layer increased to FCNs.

Table 9: F1-scores of temporary slum class, showing the accuracies of two networks

	5 × 5 networks			3 × 3 networks		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
2012	85.57%	97.04%	90.85%	85.79%	96.99%	90.95%
2013	84.20%	97.00%	90.03%	84.32%	96.02%	89.55%
2015	81.55%	85.76%	83.29%	84.41%	89.69%	86.82%
2016	74.40%	85.76%	81.97%	79.44%	89.69%	86.58%
In total	81.10%	93.19%	86.32%	83.30%	96.55%	88.38%
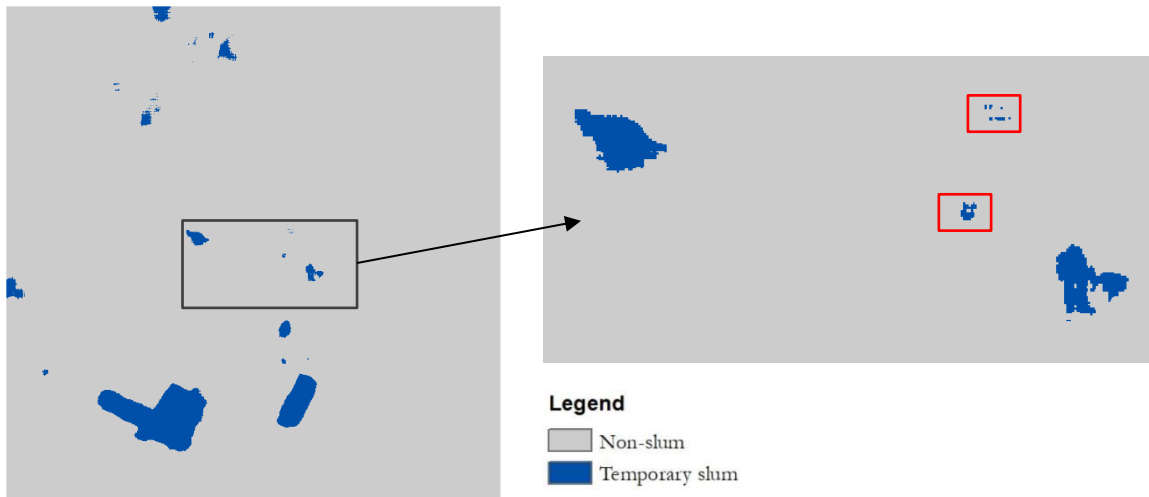
Figure 10: Classification map example of 2016, showing pixel islands
(reclassified from the original classification result)

One classification map is shown in the *Figure 10* as an example. It can be seen that there are some small pixel islands scattering in the map (i.e. the red square in *Figure 10*), which is not possibly existing in the real situation. These pixels or tiny patches were isolated in the image. As one individual temporary slum tent is about 21 × 21 pixels (determined by visual interpretation) on the image used in this study, a patch of

pixels which are smaller than this size would have a high possibility of being wrongly classified. Therefore, we considered removing those noise for further change detection process.

5.1.2 Noise reduction for land cover classification

To reduce the classification errors of pixel islands, we tried two related tools in ENVI: (1) majority analysis and (2) Classification clumping.

5.1.2.1 Majority Analysis

A majority analysis after classification has been performed to experiment its effect in noise reducing. It was done based on the "Majority/Minority Analysis" tool in ENVI. It allows the user to determine a certain kernel size for processing the whole image. The central pixel in the kernel would be replaced with the class value which made up the majority of the kernel. In this study, we set the kernel size as 21×21 pixels, since a patch smaller than this size would not be an individual temporary slum in reality. *Figure 11* illustrates examples of majority analysis for classification results. Compared with the original classification, the majority analysis successfully removed some pixel islands and smoothened the slum boundary as well.

(a) original classification   *(b)   after   majority*
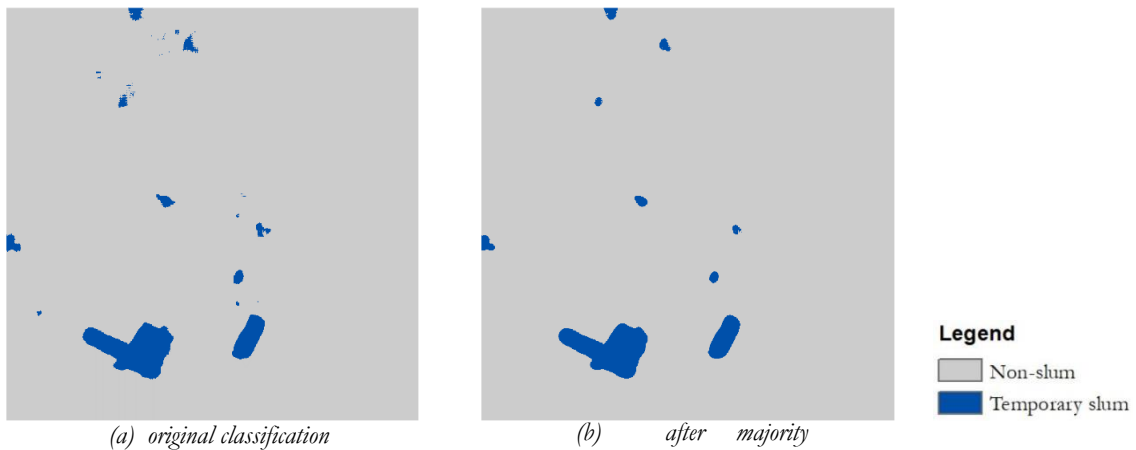
Figure 11: Comparison of the original classification and majority analysis result

5.1.2.2 Classification clumping

(a) original classification   *(b) after classification clumping*
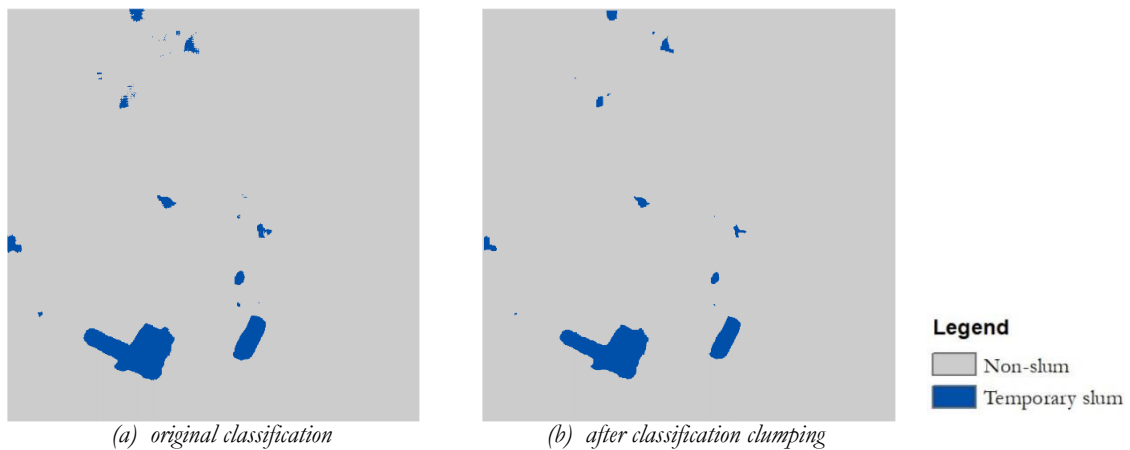
Figure 12: Comparison of the original classification and classification clumping result

Another method performed to reduce the noise was "Classification Clumping" which is also a post-classification tool in ENVI. Unlike majority analysis, classification clumping applied morphological operators to the classified areas. This tool would perform a morphological filter of dilating at first, followed by a morphological filter of eroding. The selected class would be clumped first by a dilate

operation and then an erode operation using specified kernel size for each operation. Here, we applied a 5 × 5 dilate operation followed by a 9 × 9 erode operation. The result after clumping is shown in *Figure 12*. Same as the majority analysis, pixels were removed and boundaries were smoothened.

5.1.2.3   Accuracy comparison

We also calculated the F1-scores of temporary slum class in the classification maps after operating these two methods (*Table 10*). By comparison, applying majority analysis showed a slightly higher accuracy than applying classification clumping. The reason why the accuracy was lower than the accuracy without noise reduction might be that although some inaccurate classification islands were removed, the boundaries of other big patches were smoothened. Therefore, those left out classified slum areas were somehow enlarged, leading to a decrease in the accuracy.

We used the classification maps with the majority analysis for the next change detection step.

Table 10: F1-scores showing the accuracies after noise reduction

	Majority analysis	**Classification clumping**
2012	89.38%	87.39%
2013	89.19%	86.43%
2015	88.03%	86.21%
2016	86.80%	84.23%
In total	88.35%	86.06%

5.2   Change detection result

5.2.1   Performance of 5 × 5 networks and 3 × 3 networks

We also trained FCNs from a 5 × 5 to 3 × 3 for the change-detected networks. Same with 3 × 3 networks showing a better accuracy in chapter 5.1.1, it also provided a more accurate result in change-detected network (*Table 11*). Although in the time period of 2012 to 2013, the 5 × 5 networks had a higher accuracy, it is a small improvement of 2%. The 3 × 3 networks performed more accurately in the other two time period analysis.

Table 11: F1-scores showing the accuracy of two networks, testing tiles

	5 × 5 networks			**3 × 3 networks**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
2012 - 2013	13.85%	42.26%	20.25%	12.75%	40.42%	18.31%
2013 - 2015	34.79%	42.31%	36.01%	31.87%	52.59%	37.88%
2015 - 2016	22.41%	47.46%	28.76%	31.52%	54.17%	36.49%
In total	23.68%	44.01%	28.34%	25.38%	49.06%	30.89%

5.2.2   Accuracy comparison

We assessed the accuracy of change detection result maps by both calculating the F1-scores from the confusion matrix and generating the TEM. The results for two change detection method performed in this study was compared in this chapter.

5.2.2.1   Confusion matrix

We calculated the F1-score for a new class of 'changed slum area', consisting the pixels with every slum change trajectory (chapter 4.3.1) from the land cover classification. For change-detected networks, the increased area and decreased area (chapter 4.3.2.2) were also merged together into one class as 'changed slum area'.

Table 12 showed the average F1-score of all the study tiles in each time period. Neither of the methods showed a significant advance over the other from the average accuracy in total. In the time period of 2012 to 2013, change-detected networks performed better than post-classification. But when analysing the change between 2015 and 2016, post-classification acted more accurately that the change-detected networks. Generally speaking, the lower accuracies were obtained in the analysis between 2012 and 2013 for both of the two methods and the higher accuracies were for the time period of 2013 to 2014.

Table 12: F1-scores of changed slum area in post-classification result

	Post-classification	Change-detected networks
2012 - 2013	43.69%	49.69%
2013 - 2015	61.52%	60.66%
2015 - 2016	55.95%	50.96%
In total	53.80%	53.68%

However, when looking into the individual accuracy of each tile, it could be seen that the accuracies varied a lot from tile to tile (*Table 13*). Higher accuracy could be over 90%, while the lowest accuracy was only 3.86%. In fact, the accuracies of land cover classification for this tile in 2015 and 2016 were 70.48% and 76.19% (3×3 network), which was also the lowest among all the tiles, resulting in the lowest accuracy among all the post-classification results as well. This might be ascribed to the images themselves. As the images were obtained in different time, the images were not exactly corresponded to each other due to the image registration problems resulting from the viewing angles, etc.

Table 13: F1-scores of change detection result for each tile

Tile	Post-classification			Change-detected networks		
	2012-2013	2013-2015	2015-2016	2012-2013	2013-2015	2015-2016
1	36.67%	38.42%	11.92%	22.69%	19.89%	3.86%
2	37.19%	55.15%	55.00%	19.31%	51.32%	40.37%
3	41.66%	70.22%	51.31%	78.46%	89.30%	73.27%
4	28.87%	63.24%	42.71%	17.79%	54.50%	24.37%
5	54.70%	73.69%	70.20%	91.54%	94.97%	91.29%
6	36.13%	57.11%	47.94%	23.66%	36.65%	39.20%
7	62.28%	82.82%	92.63%	91.29%	95.20%	96.93%
8	*	*	73.58%	*	*	48.65%
9	62.58%	72.98%	63.93%	84.70%	86.48%	75.51%
10	33.11%	40.03%	50.31%	17.78%	17.68%	16.12%
			Tile 3/5/7/9: Training tiles		* No changes in this tile	

Moreover, we calculated the average F1-scores for training tiles and testing tiles separately (*Table 14*). It is obvious that both of the two methods performed better in the training tiles than in the testing tiles. But the gap between the two groups was much bigger in the change-detected networks than in the post-classification results. Both of the methods had badly and well performed tiles. In general, post-classification generated a more balanced results with smaller gap between the highest and lowest, as well as the smaller gap between the training tiles and testing tiles.

All change maps are shown in Annex 4.

Table 14: F1-scores of the training and testing tiles

Tile	Method	2012-2013	2013-2015	2015-2016	In total
Training	**Post-classification**	55.30%	74.93%	69.52%	66.58%
	Change-detected networks	86.50%	91.49%	84.25%	87.41%
Testing	**Post-classification**	34.39%	50.79%	46.91%	44.21%
	Change-detected networks	20.25%	36.01%	28.76%	28.37%

5.2.2.2 Trajectory error matrix

In order to better understand the accuracy of change detection results, we also used TEM assess the change trajectories of temporary slums obtained by two methods. The classification maps for different year were stacked into one composite map (example in *Figure 14*). We generated 500 random points in two groups: 250 random points in the unchanged areas and 250 random points in the changed areas. This stratification is because of the limited changed areas in some tiles (example in *Table 14*). If positioning the points randomly in the whole image without considering about this, only few points would be located in the changed area. The indices became less affected by the small amount of the points, resulting in a incorrectly high value. This is a common problems in change analysis when most of the study areas did not change. Completely random samples did not work well in this situation.

Five indices were calculated based on the matrix (*Table 15*). For overall accuracies (A_T), we obtained about 76.36% for the post-classification result and 72.30% for the change-detected networks, meaning that 4% more sample had been classified correctly with the right trajectories. For two methods, the change/no change accuracies ($A_{C/N}$) were both higher than the A_T. This is because $A_{C/N}$ only considered whether the change maps detected

Figure 13: Example of stacked maps for change detection accuracy assessment
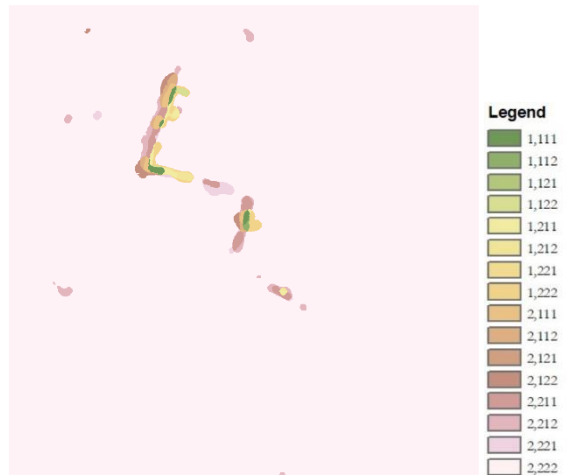
Figure 14: Example of limited changed areas

changes or not, without considering the correctness of trajectories. For OAD, the value is opposite, which means $A_{C/N}$ is higher than A_T, indicating some of the change trajectories did not match with reference data. And post-classification had more wrong trajectories. Besides, change-detected networks have a higher $ADIC_C$ than post-classification, suggesting that more sample points in the change-detected networks results could be identified with the correct change trajectories.

Table 15: TEM indices for two change detection methods

Indices	Post-classification	Change-detected networks
overall accuracy (A_T)	76.36%	72.30%
change/no change accuracy ($A_{C/N}$),	89.60%	80.12%
overall accuracy difference (OAD)	13.24%	7.82%
accuracy difference of no change trajectory ($ADIC_N$)	100.00%	100.00%
accuracy difference of change trajectory ($ADIC_C$)	67.18%	74.17%

5.2.3   Change-detection maps

Image of 2015

Image of 2016

Reference data of slum change between 2015 and 2016

Change map by post-classification change detection

Legend
- Increased slum
- Decreased slum
- Unchanged
- Other
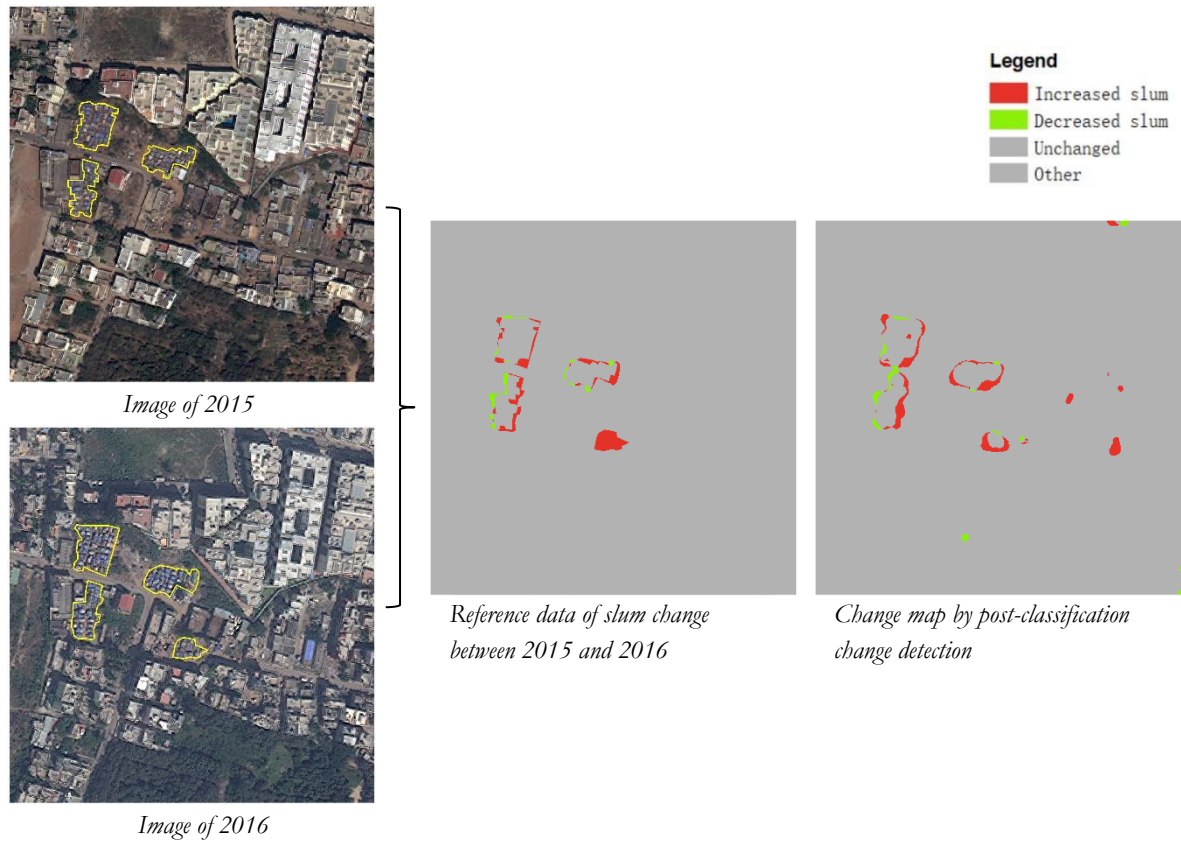
Figure 15: Example with low accuracy but correct location for change

After assessed the accuracy quantitatively, we also visually checked the change maps (Annex 4). Although the accuracy assessed in the previous chapter was relatively low for some maps, it showed the right location where changes happened to temporary slums. One example for this situation is shown in *Figure 15*. The post-classification change detection result of temporary slum from 2015 to 2016 for this tile obtained only 42.71% for F1-score based on confusion matrix. However, when looking at the map, we could say that it gave the correct result of where the change was and what type of change (increasing/decreasing) it was. Based on this strategy, the result could also be used to identify the slum change location.

6.   DISSCUSION AND LIMITATION

6.1   Temporal dynamics of slum in Bangalore in the study area

As mentioned in chapter 1.1, only a few studies have analysed the temporal changes of slums. For example, Kit & Lüdeke (2013) identified three trends of slum temporal changes: densification of slum settlements, slum growth in fringes and the spatial focus zones where had the most slum growth. In this study, we intended to understand how slum were changing between 2012 and 2016 in our study area. The changes were about two questions: how much slum area was changing and how the slum area was changing. The second questions would then be divided into two sub-questions: what land cover did change into slum and what land cover did appear when slums disappeared.

6.1.2   Area of slum changing

Before discussion, the area of changed slums were calculated for the result change maps with a comparison to the reference data (shown in *Table 16*). Here, 'Increase' and 'decrease' represent the changes from other land cover to temporary slums and from temporary slums to other land cover. The overall gap between reference data and post-classification was 13,579 m², while for change-detected networks was 20,579 m². Although change-detected networks showed a comparable accuracy in the assessments, it had a worse performance in the area aspect than the post-classification. As the gap between reference data and change-detection results could not be ignored, we used the value from reference data for discussion.

Table 16: Comparison of areas of changed slums

(m²)	2012		2013		2015	
	Increase	**Decrease**	**Increase**	**Decrease**	**Increase**	**Decrease**
Reference data	8873	4047	12614	9652	7203	19860
Post-classification	7981	6377	15205	12471	10030	21980
Change-detected networks	4826	2612	9313	13403	5654	13364

From 2012 to 2016, 12,012 m² of temporary slums appeared in the study area, while 1,7052 m² disappeared in this time period. There were also 11041 m² of slums stayed unchanged. On average, 7,173 m² of land changed into temporary slums in our study area per year, while 8,390 m² of the existed temporary slums disappeared, showing an entirely decreasing trend. A detailed changing pattern is shown in *Figure 16*.
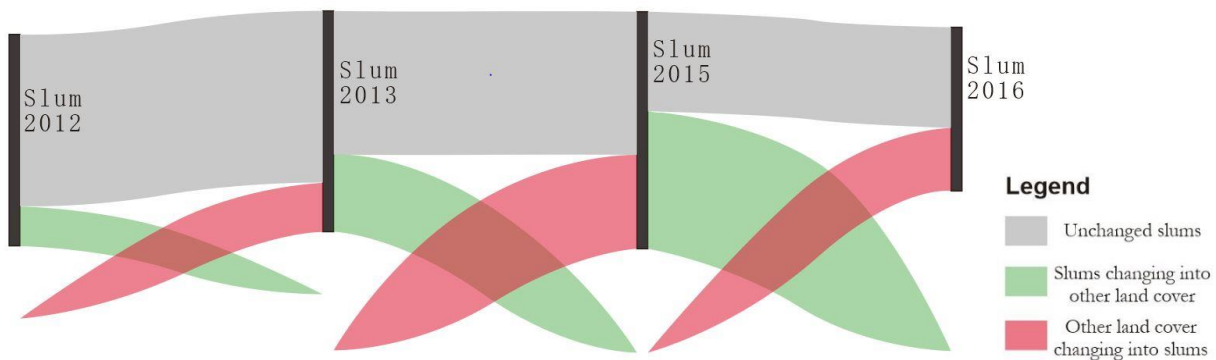
Figure 16: Diagram of temporary slum changing situation

The flow of grey colour presented how many slums remained unchanged in each time period. The flow of green colour was the area changing from slums to other land cover, while red colour stood for the area

becoming slums. It suggested that with the passage of time, the less slum area stayed unchanged and more and more slum areas were disappearing. The largest increase of temporary slums happened between 2013 to 2015, which is also the longest gap in our study period.

6.1.3 Pattern of slum changing

Table 17: Proportion of different temporal dynamics, 2012 to 2016

Increased		Decreased	
other → slum	0.64%	slum → green land	42.64%
formal built-up → slum	24.11%	slum → vacant land	36.71%
vacant land → slum	42.57%	slum → formal built-up	20.51%
green land → slum	32.68%	slum → other	0.14%

Table 18: Changing rate of different temporal dynamics, 2012 to 2016

Increased	m² /year	Decreased	m² /year
other → slum	22	slum → green land	2250
formal built-up → slum	819	slum → vacant land	1937
vacant land → slum	1447	slum → formal built-up	1083
green land → slum	1111	slum → other	7

Table 17 shows the proportion of different types of temporal dynamics from 2012 to 2016. And *Table 18* shows the changing rate of every temporal dynamic. The largest transition of the increased part was the changing from vacant land to the slums. About 42% of the new slums used to be vacant land, with a change rate of 1,447 m² per year. An example of this transition is in *Figure 17*. For the slum decreasing, most of the temporary slums changed into green land with a change rate of 2,250 m² per year, which is different from the increasing transition. A very specific example of this transition is in *Figure 18*. This transition was associated with some reforming programmes in this area. It could be seen that formal roads have been constructed in this area, with newly planted green land. We can infer that some reclaiming projects might be performed here.
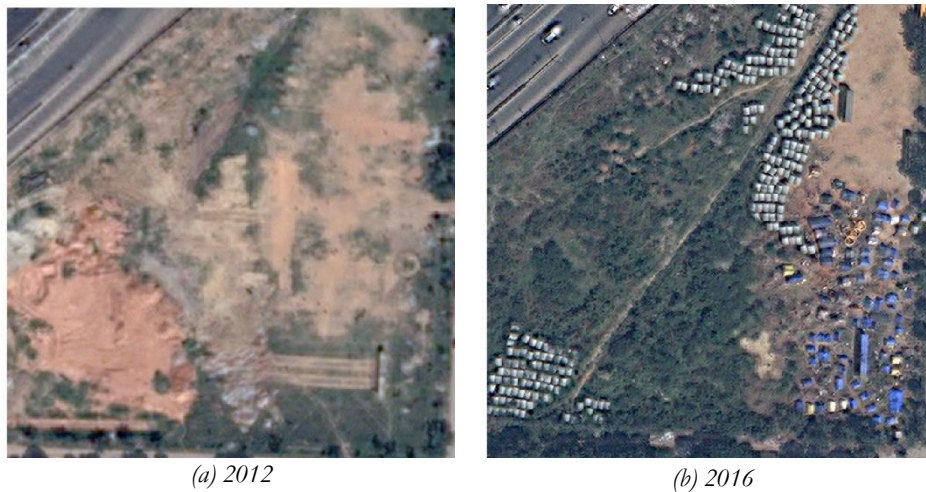
(a) 2012 *(b) 2016*
Figure 17: Example of vacant land changing into slums

(a) 2012   *(b) 2016*

Figure 18: Example of slums changing into green land

6.2   Methodological advantages and disadvantages

In this study, two change detection methods were employed to analyse the temporal dynamics of slums. Besides, two methods for accuracy assessment were employed. Discussion about these methods were also made in this chapter.

6.2.1   Post-classification change detection

For post-classification change detection, land cover classification maps were generated based on FCNs. The maps had a high accuracy of over 85%, indicating that using a deep learning algorithm to identify temporary slums from VHR imagery in urban areas is effective. This result also responds to the study from Persello & Stein (2017), showing their proposed framework to capture informal settlements working well from Dar es Salaam in Tanzania to Bangalore in India.

However, the post-classification results did not have a good performance based on the accuracy assessments. It could not show the exact size of areas where changes happened. This problem is associated with the uncertainty of slum boundary as well, as the reference data in this study were generated by visual interpretation based on the author's background and the understanding from the provided information. But the result change maps could tell the existence of slum changes. The changed slum areas in reference maps were also changing in the change-detection results, with only considering the location. Molenaar (2000) proposed two concepts of existential uncertainty and extensional uncertainty. Existential uncertainty means the uncertainty about the existence of a slum in reality. And extensional uncertainty implies the uncertainty of whether an area covered by a slum can be determined with limited certainty or not (Kohli, Stein, & Sliuzas, 2016). Based on these concepts, the post-classification method is beneficial in analysing the existence of changed slums, but not practical in analysing the size of changed slums.

6.2.2   Change-detected networks

On the other hand, a FCNs network with the same architecture as the one used land cover classification was employed to directly detect the changed slum areas. One of the problems for this method is that the accuracies for the training tiles were much higher than the testing tiles, indicating that the classifier learnt by the FCNs were not transferred well to the other images. This might also result from the reference data preparation. Except for the uncertainty of slum delineation which is the same in the post-classification process, another uncertainty is change trajectory. In this study, when selecting the training tiles, we only considered the trajectories between temporary slums and our determined land cover class. In fact, the

objects in one land cover class might be different from each other. For example, one training tile contained a trajectory from concrete buildings to temporary slums and taught the networks how to classify it. But in testing tiles, the trajectory was from brick buildings to temporary slums. Although both from formal built-up class to slums, the networks had no knowledge about this specific trajectory, hence leading to incorrect classification. The changed-detected networks had an 87% accuracy for the training tiles, indicating that it has a potential to detect the changes with precise reference data.

Besides, similar to post-classification, change-detected networks also performed well when capturing the existence of change.

6.2.3   Accuracy assessment

In this study, the confusion matrix and trajectory error matrix were employed to assess the accuracy of change detection results. Confusion matrix and related indices, like producer accuracy and user accuracy, are still a widely used method to assess the accuracy of deep learning algorithm based classification and change detection (Dai et al., 2018; Fu et al., 2017; Persello & Stein, 2017). Nevertheless, studies already stated the problems (Foody, 1992; Pontiuns, 2000). Foody (2002) also talked about the importance of reference data in the confusion matrix assessment. From this study, as discussed in the previous chapter, although the results were not performing well with low values for F1-scores, it did tell the correct location of where the changes happened. As confusion matrix provides a pixel-based, land cover with uncertain delineation will show a low accuracy around boundaries. This is the case in this study, without a standard definition of a slum, the boundary of changed areas is also fuzzy. Therefore, the confusion matrix cannot give a credible assessment with a consideration of the neighbourhood context.

Another assessment method employed in this study was the trajectory error matrix. While the confusion matrix provided an assessment of 'change/no change' status, TEM assessed the accuracy for 'from/to' changes. As mentioned in chapter 5.2.2.2, one shortcoming for TEM is that the random samples cannot be completely random, especially when the changed area only cover a small proportion of the whole region.

How to combine the assessment of change certainty and change trajectory will still be a topic for further study.

6.3   Limitations

First is about the image data. Although the images in this study were all from the Worldview satellites, two are from Worldview 02 and the other two are from Worldview 03. We resampled the images in order to keep a same cell size, which would also influence the image quality. Besides, the viewing-angles were different in these images, causing shadows as well as occlusion from high-rise buildings. Another problem is the registration of images. This pre-processing operation was performed manually. The images from different time were not able to exactly match with each other, which would influence the later change detection process.

During the FCNs based classification, we used visual interpretation to generate the reference data. Uncertainty were encountered with whether an area was a temporary slum or not. Another issue is related to the lack of an exact definition of the boundary of a slum (Kohli et al., 2016). Defining an exact delineation between a slum area and other land cover area is always a hard task. Therefore, both the location and boundary issues affected the accuracy of reference data, and therefore affected the quality of classification and change detection results.

For accuracy assessment, as we discussed in this chapter, the two methods which we used in this study could not assess the performance of our slum change maps well. Confusion matrix and its associated measurements could only tell the correctness of the classified label for pixel itself and could not deal with

fuzzy boundaries well. Trajectory error matrix had limited performance when the assessed area is small in the whole image.

7. CONCLUSION AND RECOMMADATION

7.1 Conclusion

The main objective of this study is to develop a FCNs-based approach to map slums and analyse their temporal dynamics using machine learning based on VHR imagery. To accomplish this objective, this study employed a FCNs-based method to map changes of slum areas in the city of Bangalore, with four Worldview images taken over the city from 2012 to 2016. After getting the change maps, we analysed the temporal dynamics of slums. Three sub-objectives were covered in order to accomplished the main goal by answering the related research questions.

The first sub-objective is to identify slum-area and non-slum from VHR imagery by applying fully convolutional networks (FCNs). Supported by recent literature and with the help of available slum boundary data, we understood the characteristics of slums in Bangalore. Temporary slums, also known as 'Blue tent' slums, were selected as the target slum type in this study. These slums are sheltered with blue plastic roof and usually cluster together. People living there commonly lack basic infrastructure. People living there were lack of basic infrastructure. These areas are commonly ignored by the government as they are not recognised slums. Therefore, investigating their dynamics is of particular interest as they do not appear on official maps.

A better understanding of these slums could provide information for the improvement of these areas. After learning about temporary slums, we selected ten tiles as our study area and then generated the reference data by visual interpretation. The reference data consisted four land cover class, which were related to the slum change trajectories in the study area. Among them, we selected four tiles, which had all the land cover classes in the reference data as well as all the slum change trajectories, as our training tiles. Then we applied a FCNs architecture with dilated convolutions to classify the images. We experimented the performance between a network with 5×5 kernel size and from a network with 3×3 kernel size. After assessing the accuracy of the classification maps, we found that the 3×3 network had an accuracy of 88.38%, which is higher than the 86.32% accuracy of the 5×5 network . Therefore, we applied the 3×3 network for later analysis.

The second sub-objective is to analyse the temporal dynamic. We performed two methods to accomplish this sub-objective. On the one hand, the resulted slum maps from the land cover classification were used for a post-classification change detection. On the other hand, we used the FCNs to directly classify the changed slum areas in the images. From the result, it could be observed that in general, there were 17,052 m^2 slum areas disappearing and 12,012 m^2 of new slums from 2016 to 2012 in our study area. Besides, from the change trajectory aspect, among all the trajectories of slum increasing, temporary slums were commonly appearing from the vacant land. For the decreasing trajectory, the majority of the transformation was from slum areas to green land.

The last sub-objective is to evaluated the outcomes of change detection for temporal dynamics. In this study, we applied two methods to assess the accuracy of the change detection outcomes. One is the confusion matrix. We assessed the performance of two change detection method results by calculating the F1-score from the confusion matrix. Both of the two methods obtained about 53% for the F1-score. Another assessment method is trajectory error matrix (TEM). Five indices were calculated to assess the outcomes. Post-classification had a better result in terms of overall accuracy (A_T), while change-detected networks having more sample points with correctly classified trajectories. We also visually compared the change maps and reference data, discovering that although the change maps did not have a good performance in pixel-based accuracy assessment, it could still give the right location of where the change

happened. Based on the discovery, we found that both of two methods did not give a good assessment on the results. The generated change maps did not give an accurate result for the size of changed slums, but offer a valuable information of the location of where the change happened.

7.2   Recommendations

For future works, we listed some possible directions as followed:

i.   Discovering a ruleset of defining the slum boundary.

In this study, we delineated the slum boundary by visual interpretation based on the author's understanding. It would be a direction to analyse the delineations from different experts to see their difference.

ii.   Exploring techniques to improve FCNs in slum classification field.

A FCNs architecture with dilated convolutions was used in this study. Other advanced convolution networks in computer vision field could also be experimented in slum classification to see if they were beneficial or not.

iii.   Proposing an accuracy assessment method for slum change

In our study, we found that the methods we used could not give a good assessment on our results. Therefore, a measurement needs to be proposed to better assess the change location rather than the change of pixel label.

LIST OF REFERENCES

Akram, T., Laurent, B., Naqvi, S. R., Alex, M. M., & Muhammad, N. (2018). A deep heterogeneous feature fusion approach for automatic land-use classification. *Information Sciences*, *467*, 199–218. https://doi.org/10.1016/j.ins.2018.07.074

Atkinson, P. M., & Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of Remote Sensing*, *18*(4), 699–709. https://doi.org/10.1080/014311697218700

Bergado, J. R., Persello, C., & Gevaert, C. (2016). A deep learning approach to the classification of sub-decimetre resolution aerial images. In *International Geoscience and Remote Sensing Symposium (IGARSS)* (Vol. 2016–Novem, pp. 1516–1519). IEEE. https://doi.org/10.1109/IGARSS.2016.7729387

Beukes, A. (2015). *Making the invisible visible Generating data on 'slums' at local, city and global scales*. *June* (Vol. 52). https://doi.org/978-1-78431-270-1

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., … Tiede, D. (2014). Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Boldt, M., Thiele, A., & Schulz, K. (2012). Object-based urban change detection analyzing high resolution optical satellite images. *Earth Resources and Environmental Remote Sensing/GIS Applications III*, *8538*, 85380E–1 to 85380E–9. https://doi.org/Artn 85380e 10.1117/12.973687

Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2015). Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. https://doi.org/10.1021/cr8002505

Census Organization of India. (2015). Bangalore (Bengaluru) City Population Census 2011-2019 | Karnataka. Retrieved February 14, 2019, from https://www.census2011.co.in/census/city/448-bangalore.html

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the Devil in the Details: Delving Deep into Convolutional Nets. https://doi.org/10.5244/C.28.6

Civco, D., Hurd, J., & Wilson, E. (2002). A comparison of land use and land cover change detection methods. *ASPRS-ACSM Annual …*, 12. https://doi.org/10.1007/s12274-017-1874-y

Dai, F., Wang, Q., Gong, Y., Chen, G., Zhang, X., & Zhu, K. (2018). Change detection based on Faster R-CNN for high-resolution remote sensing images. *Remote Sensing Letters*, *9*(10), 923–932. https://doi.org/10.1080/2150704x.2018.1492172

Dittrich, C. (2005). Bangalore: Divided City under the Impact of Globalization. *Asian Journal of Water, Environment and Pollution*, *2*(2), 22–30. Retrieved from https://content.iospress.com/download/asian-journal-of-water-environment-and-pollution/ajw2-2-05?id=asian-journal-of-water-environment-and-pollution%2Fajw2-2-05

Duque, J. C., Patino, J. E., & Betancourt, A. (2017). Exploring the potential of machine learning for automatic slum identification from VHR imagery. *Remote Sensing*, *9*(9). https://doi.org/10.3390/rs9090895

Foody, G. M. (1992). On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, *58*, 1459. Retrieved from https://www.asprs.org/wp-content/uploads/pers/1992journal/oct/1992_oct_1459-1460.pdf

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*, 185–201. Retrieved from www.elsevier.com/locate/rse

Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017). Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sensing*, *9*(5), 498. https://doi.org/10.3390/rs9050498

Gao, F., de Colstoun, E. B., Ma, R., Weng, Q., Masek, J. G., Chen, J., … Song, C. (2012). Mapping impervious surface expansion using medium-resolution satellite image time series: A case study in the Yangtze River Delta, China. *International Journal of Remote Sensing*, *33*(24), 7609–7628. https://doi.org/10.1080/01431161.2012.700424

Giada, S., De Groeve, T., Ehrlich, D., & Soille, P. (2003). Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. *International Journal of Remote Sensing*, *24*(22), 4251–4266. https://doi.org/10.1080/0143116021000035021

Gilbert, A. (2007). The return of the slum: Does language matter? *International Journal of Urban and Regional Research*, *31*(4), 697–713. https://doi.org/10.1111/j.1468-2427.2007.00754.x

Government of India. (2011). Census 2011 India. Retrieved August 23, 2018, from http://www.census2011.co.in/

Graesser, J., Cheriyadat, A., Vatsavai, R. R., Chandola, V., Long, J., & Bright, E. (2012). Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(4), 1164–1176. https://doi.org/10.1109/JSTARS.2012.2190383

Gruebner, O., Sachs, J., Nockert, A., Frings, M., Khan, M. M. H., Lakes, T., & Hostert, P. (2014). Mapping the Slums of Dhaka from 2006 to 2010. *Dataset Papers in Science*, *2014*, 1–7. https://doi.org/10.1155/2014/172182

Hay, G. J., Blaschke, T., Marceau, D. J., & Bouchard, A. (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(5–6), 327–345. https://doi.org/10.1016/S0924-2716(02)00162-4

Hester, D. B., Nelson, S. A. C., Cakir, H. I., Khorram, S., & Cheshire, H. (2010). High-resolution land cover change detection based on fuzzy uncertainty analysis and change reasoning. *International Journal of Remote Sensing*, *31*(2), 455–475. https://doi.org/10.1080/01431160902893493

Hofmann, P., Strobl, J., Blaschke, T., & Kux, H. (2008). Detecting informal settlements from QuickBird data in Rio de Janeiro using an object based approach. In *Object-Based Image Analysis* (pp. 531–553). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77058-9_29

Iacoboaea, C. (2009). SLUMS IN ROMANIA. *Theoretical and Empirical Researches in Urban Management Number*, *4*(10), 101–113. Retrieved from http://www.um.ase.ro/no10/9.pdf

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Khalifa, M. A. (2011). Redefining slums in Egypt: Unplanned versus unsafe areas. *Habitat International*, *35*(1), 40–49. https://doi.org/10.1016/j.habitatint.2010.03.004

Kit, O., & Lüdeke, M. (2013). Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *83*, 130–137. https://doi.org/10.1016/j.isprsjprs.2013.06.009

Kit, O., Lüdeke, M., & Reckien, D. (2013). Defining the bull's eye: Satellite imagery-assisted slum population assessment in Hyderabad, India. *Urban Geography*, *34*(3), 413–424. https://doi.org/10.1080/02723638.2013.778665

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, *36*(2), 154–163. https://doi.org/10.1016/j.compenvurbsys.2011.11.001

Kohli, D., Stein, A., & Sliuzas, R. (2016). Uncertainty analysis for image interpretations of urban slums. *Computers, Environment and Urban Systems*, *60*, 37–49. https://doi.org/10.1016/j.compenvurbsys.2016.07.010

Kohli, D., Warwadekar, P., Kerle, N., Sliuzas, R., & Stein, A. (2013). Transferability of object-oriented image analysis methods for slum identification. *Remote Sensing*, *5*(9), 4209–4228. https://doi.org/10.3390/rs5094209

Krishna, A., Sriram, M. S., & Prakash, P. (2014). Slum types and adaptation strategies: identifying policy-relevant differences in Bangalore. *Environment and Urbanization*, *26*(2), 568–585. https://doi.org/10.1177/0956247814537958

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, 1106–1114. https://doi.org/10.1145/3065386

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space-15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6). https://doi.org/10.3390/rs8060455

Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud, I. (2016). Extraction of Slum Areas From VHR Imagery Using GLCM Variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1830–1840. https://doi.org/10.1109/JSTARS.2016.2538563

Leonita, G., Kuffer, M., Sliuzas, R., & Persello, C. (2018). Machine learning-based slum mapping in support of slum upgrading programs: The case of Bandung City, Indonesia. *Remote Sensing*, *10*(10), 1522. https://doi.org/10.3390/rs10101522

Li, B., & Zhou, Q. (2009). Accuracy assessment on multi-temporal land-cover change detection using a trajectory error matrix. *International Journal of Remote Sensing*, *30*(5), 1283–1296. https://doi.org/10.1080/01431160802474022

Liu, H., Huang, X., Wen, D., & Li, J. (2017). The use of landscape metrics and transfer learning to explore urban villages in China. *Remote Sensing*, *9*(4), 365. https://doi.org/10.3390/rs9040365

Liu, Y., & Li, X. (2014). Domain adaptation for land use classification: A spatio-temporal knowledge reusing method. *ISPRS Journal of Photogrammetry and Remote Sensing*, *98*, 133–144. https://doi.org/10.1016/j.isprsjprs.2014.09.013

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, *25*(12), 2365–2407. https://doi.org/10.1080/0143116031000139863

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *ICML '13* (Vol. 28, p. 6). https://doi.org/10.1016/0010-0277(84)90022-2

Mahabir, R., Croitoru, A., Crooks, A., Agouris, P., & Stefanidis, A. (2018). A Critical Review of High and Very High-Resolution Remote Sensing Approaches for Detecting and Mapping Slums: Trends, Challenges and Emerging Opportunities. *Urban Science*, *2*(1), 8. https://doi.org/10.3390/urbansci2010008

Mahabir, R., Crooks, A., Croitoru, A., Agouris, P., Mahabir, R., Crooks, A., … The, P. A. (2016). The study of slums as social and physical constructs : challenges and emerging research opportunities. *Regional Studies, Regional Science*, *3*(1), 400–420. https://doi.org/10.1080/21681376.2016.1229130

Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L., & Tuia, D. (2015). Semisupervised Transfer Component Analysis for Domain Adaptation in Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(7), 3550–3564. https://doi.org/10.1109/TGRS.2014.2377785

Mboga, N., Persello, C., Bergado, J. R., & Stein, A. (2017). Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sensing*, *9*(11), 1106. https://doi.org/10.3390/rs9111106

Molenaar, M. (2000). Three conceptual uncertainty levels for spatial objects. *International Archives of Photogrammetry and Remote Sensing*, *33(B4/2; P*, 670–677. Retrieved from http://www.isprs.org/proceedings/XXXIII/congress/part4/670_XXXIII-part4.pdf

Netherlands eScience Center. (2018). DynaSlum. Retrieved August 23, 2018, from http://www.dynaslum.com/

Nielsen, M. A. (2015). Neural Networks and Deep Learning. Determination Press. Retrieved from http://neuralnetworksanddeeplearning.com/

Nolan, L. B. (2015). Slum Definitions in Urban India: Implications for the Measurement of Health Inequalities. *Population and Development Review*, *41*(1), 59–84. https://doi.org/10.1111/j.1728-4457.2015.00026.x

Novack, T., & Kux, H. J. H. (2010). Urban land cover and land use classification of an informal settlement

area using the open-source knowledge-based system InterIMAGE. *Journal of Spatial Science*, *55*(1), 23–41. https://doi.org/10.1080/14498596.2010.487640

Pacirici, F., Solimini, C., Del Frate, F., & Emery, W. J. (2007). An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(9), 2940–2952. https://doi.org/10.1109/TGRS.2007.902824

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Persello, C., & Bruzzone, L. (2012). Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(11 PART1), 4468–4483. https://doi.org/10.1109/TGRS.2012.2192740

Persello, C., & Stein, A. (2017). Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), 2325–2329. https://doi.org/10.1109/LGRS.2017.2763738

Pontiuns, R. G. (2000). Quantification Error Versus Location Error in Comparison of Categorical Maps. *Photogrammetric Engineering & Remote Sensing*, *66*(8), 1011–1016. Retrieved from https://pdfs.semanticscholar.org/ed7f/3f0896cda2b1852425f6aaef4f0ab44aba46.pdf?_ga=2.251385 54.1336279320.1551070275-676316023.1548332498

Pratomo, J., Kuffer, M., Kohli, D., & Martinez, J. (2018). Application of the trajectory error matrix for assessing the temporal transferability of OBIA for slum detection. *European Journal of Remote Sensing*, *51*(1), 838–849. https://doi.org/10.1080/22797254.2018.1496798

Rains, E., Krishna, A., & Wibbels, E. (2017). *SLUMMIER THAN OTHERS: A Continuum of Slums and Assortative Residential Selection*. Retrieved from http://urbanindiastories.com/wp-content/uploads/2014/08/Working-paper.pdf

Ranguelova, E., Weel, B., Roy, D., Kuffer, M., Pfeffer, K., & Lees, M. (2018). Image based classification of slums, built-up and non-built-up areas in Kalyan and Bangalore, India. *European Journal of Remote Sensing*, 1–22. https://doi.org/10.1080/22797254.2018.1535838

Richards, J. A., & Jia, X. (2006). *Remote sensing digital image analysis. Remote Sensing Digital Image Analysis: An Introduction*. Berlin/Heidelberg: Springer-Verlag. https://doi.org/10.1007/3-540-29711-1

Rindfuss, R. R., & Stern, P. C. (1998). Linking remote sensing and social science: The need and the challenges. In *People and pixels: Linking remote sensing and social science* (pp. 1–27). Washington, D.C.: National Academies Press. https://doi.org/10.17226/5963

Roy, D., Lees, M. H., Pfeffer, K., & Sloot, P. M. A. (2018). Spatial segregation, inequality, and opportunity bias in the slums of Bengaluru. *Cities*, *74*, 269–276. https://doi.org/10.1016/j.cities.2017.12.014

Schneider, A. (2012). Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sensing of Environment*, *124*, 689–704. https://doi.org/10.1016/j.rse.2012.06.006

Singh, A. (1989). Review Articlel: Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, *10*(6), 989–1003. https://doi.org/10.1080/01431168908903939

Sliuzas, R., Kuffer, M., Gevaert, C., Persello, C., & Pfeffer, K. (2017). Slum mapping. In *2017 Joint Urban Remote Sensing Event, JURSE 2017* (pp. 1–4). IEEE. https://doi.org/10.1109/JURSE.2017.7924589

Stanford University. (2018). CS231n Convolutional Neural Networks for Visual Recognition. Retrieved February 12, 2019, from http://cs231n.github.io/convolutional-networks/

Sun, W., & Wang, R. (2018). Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), 474–478. https://doi.org/10.1109/LGRS.2018.2795531

Sun, X., Shen, S., Lin, X., & Hu, Z. (2017). Semantic labeling of high-resolution aerial images using an ensemble of fully convolutional networks. *Journal of Applied Remote Sensing*, *11*(04), 1. https://doi.org/10.1117/1.JRS.11.042617

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 07–12–June, pp. 1–9). IEEE. https://doi.org/10.1109/CVPR.2015.7298594

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. https://doi.org/10.1109/CVPR.2016.308

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015, April 1). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*. Elsevier. https://doi.org/10.1016/j.rse.2015.01.006

Tuia, D., Persello, C., & Bruzzone, L. (2016, June). Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*. https://doi.org/10.1109/MGRS.2016.2548504

UN-DESA. (2018). *World Urbanization Prospects: The 2018 Revision. United Nations, Department of Economic and Social Affairs, Pupulation Division.* https://doi.org/(ST/ESA/SER.A/366)

UN-Habitat. (2007). Slums : Some Definitions. *State of the World's Cities 2006/7*, (2), 1–2. Retrieved from http://mirror.unhabitat.org/documents/media_centre/sowcr2006/SOWCR 5.pdf

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development. https://sustainabledevelopment.un.org/content/documents/7891Transforming%20Our%20World. pdf.* https://doi.org/10.1007/s13398-014-0173-7.2

Veljanovski, T., Kanjir, U., Pehani, P., Otir, K., & Kovai, P. (2012). Object-Based Image Analysis of VHR Satellite Imagery for Population Estimation in Informal Settlement Kibera-Nairobi, Kenya. In *Remote Sensing - Applications*. https://doi.org/10.5772/37869

Wurm, M., Weigand, M., Schmitt, A., Gei, C., & Taubenbock, H. (2017). Exploitation of textural and morphological image features in Sentinel-2A data for slum mapping. In *2017 Joint Urban Remote Sensing Event, JURSE 2017* (pp. 1–4). IEEE. https://doi.org/10.1109/JURSE.2017.7924586

Yu, F., & Koltun, V. (2015). Multi-Scale Context Aggregation by Dilated Convolutions. https://doi.org/10.16373/j.cnki.ahr.150049

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D. (2006). Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering & Remote Sensing, 72*(7), 799–811. https://doi.org/10.14358/PERS.72.7.799

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine, 5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

APPENDIX
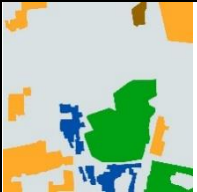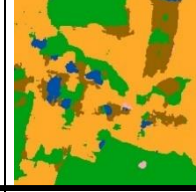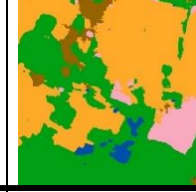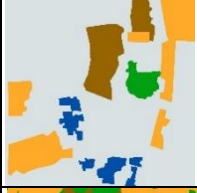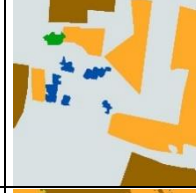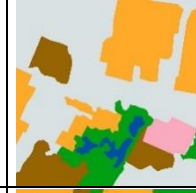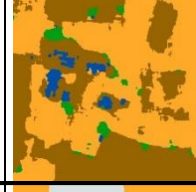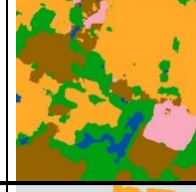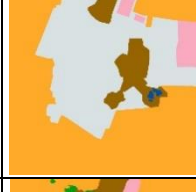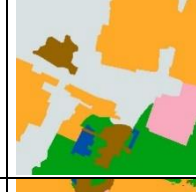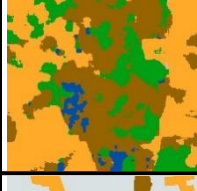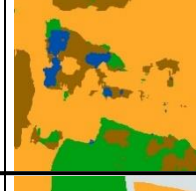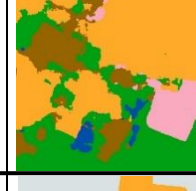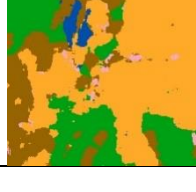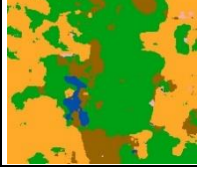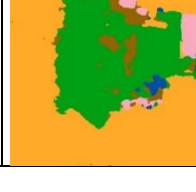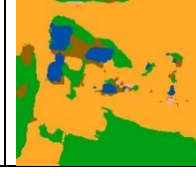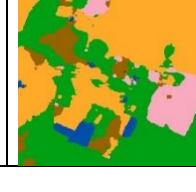
Annex 1: Images for the training and testing tiles

	2012	2013	2015	2016
Tile 1				
Tile 2				
Tile 3				
Tile 4				
Tile 5				
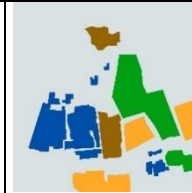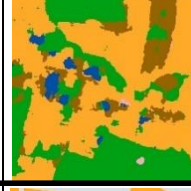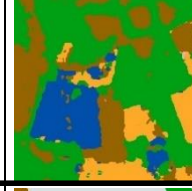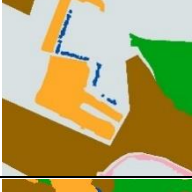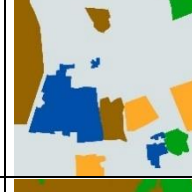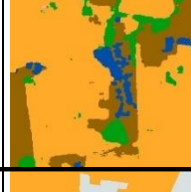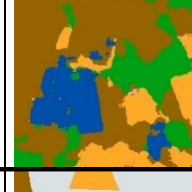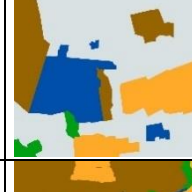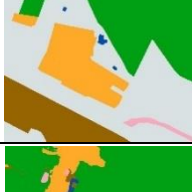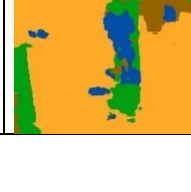
Annex 1-2: Images for the training and testing tiles

	2012	2013	2015	2016
Tile 6				
Tile 7				
Tile 8				
Tile 9				
Tile 10				

Annex 2-1: Land cover classification results from FCNs

	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5
Reference 2012					
Classification map 2012-2013					
Reference 2013					
Classification map 2013					
Reference 2015					
Classification map 2015					
Reference 2016					
Classification map 2016					

Annex 2-2: Land cover classification results from FCNs
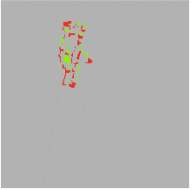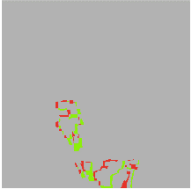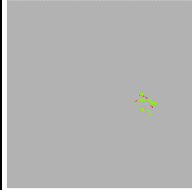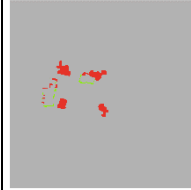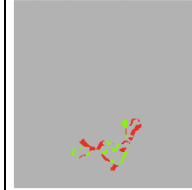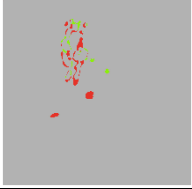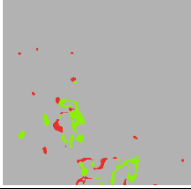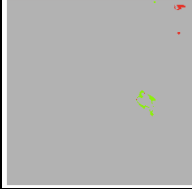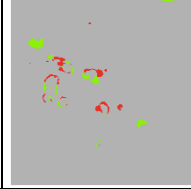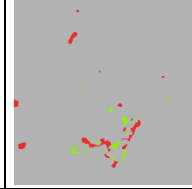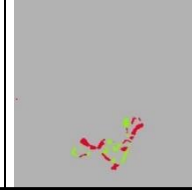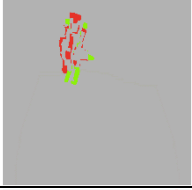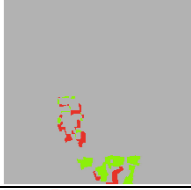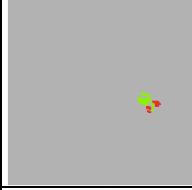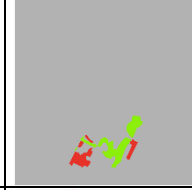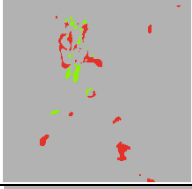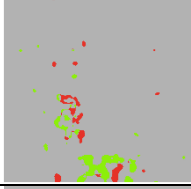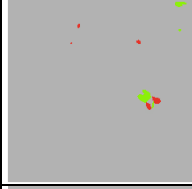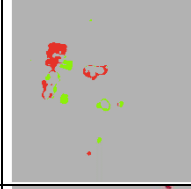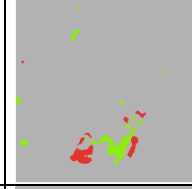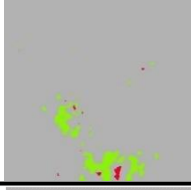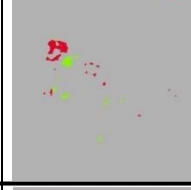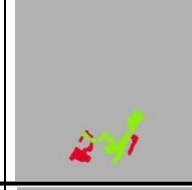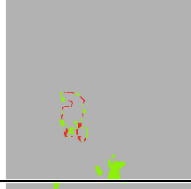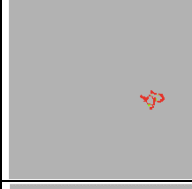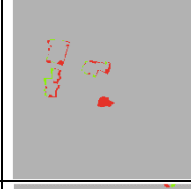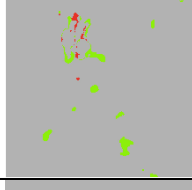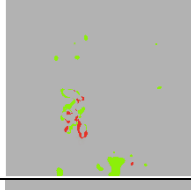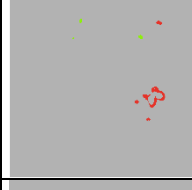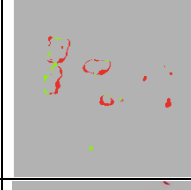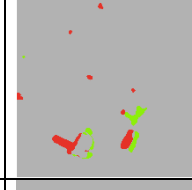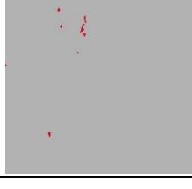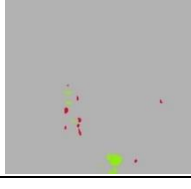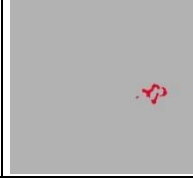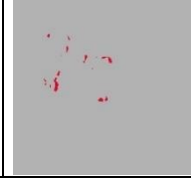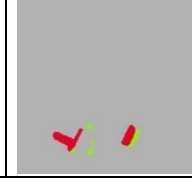
	Tile 6	Tile 7	Tile 8	Tile 9	Tile 10
Reference 2012					
Classification map 2012					
Reference 2013					
Classification map 2013					
Reference 2015					
Classification map 2015					
Reference 2016					
Classification map 2016					

Annex 3: F1-scores of slum classification results for each tile based on FCNs
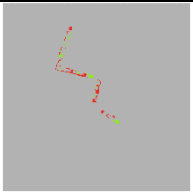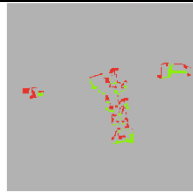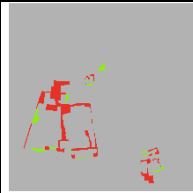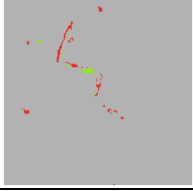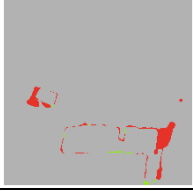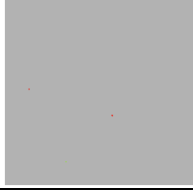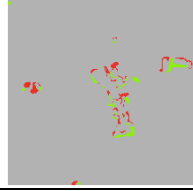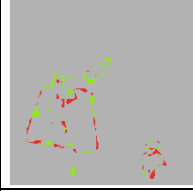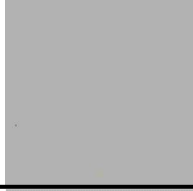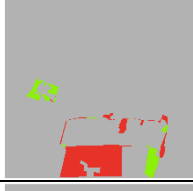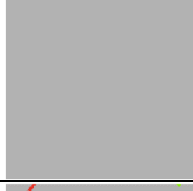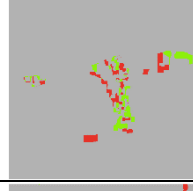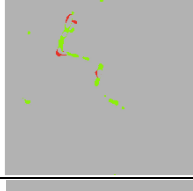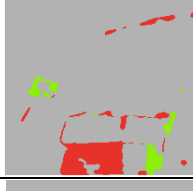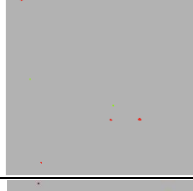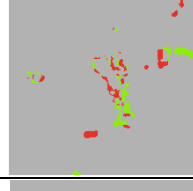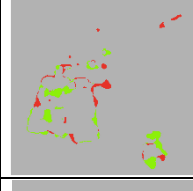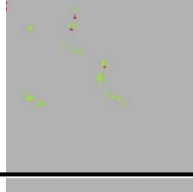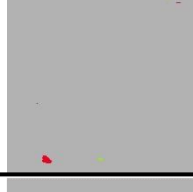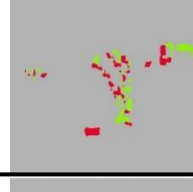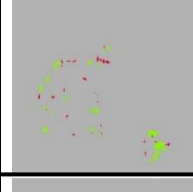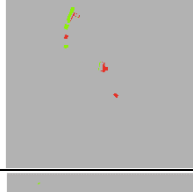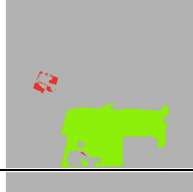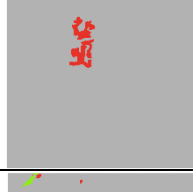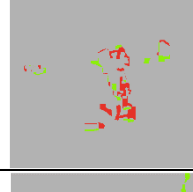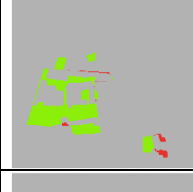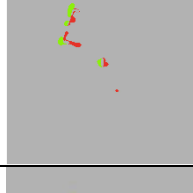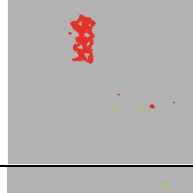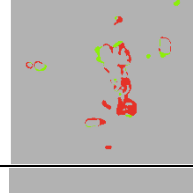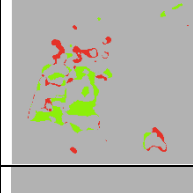
Tile	5 × 5 networks				3 × 3 networks			
	2012	2013	2015	2016	2012	2013	2015	2016
1	82.65%	85.10%	67.00%	67.92%	83.10%	86.12%	70.48%	76.19%
2	93.92%	87.50%	79.88%	76.14%	94.59%	82.12%	83.87%	79.62%
3	92.29%	89.12%	83.29%	86.74%	90.34%	90.11%	84.84%	78.36%
4	92.42%	89.46%	93.62%	93.20%	90.49%	89.70%	93.31%	94.12%
5	86.93%	91.16%	91.26%	95.97%	84.38%	87.11%	88.85%	95.80%
6	87.16%	89.73%	86.40%	81.99%	88.51%	91.85%	94.47%	90.45%
7	97.86%	97.82%	98.56%	94.13%	97.59%	97.40%	98.35%	91.56%
8	*	*	*	85.34%	*	*	*	92.05%
9	89.29%	91.48%	91.63%	94.75%	88.64%	90.61%	90.78%	93.80%
10	98.12%	98.36%	89.55%	87.24%	98.04%	97.97%	91.96%	87.02%
							* No slum	

Annex 4-1: Change maps of temporary slum -1

	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5
Reference 2012-2013					
Post-classification 2012-2013					
Change-detected net 2012-2013					
Reference 2013-2015					
Post-classification 2013-2015					
Change-detected net 2013-2015					
Reference 2015-2016					
Post-classification 2015-2016					
Change-detected net 2015-201					

Annex 4-1: Change maps of temporary slum -1

	Tile 6	Tile 7	Tile 8	Tile 9	Tile 10
Reference 2012-2013					
Post-classification 2012-2013					
Change-detected net 2012-2013					
Reference 2013-2015					
Post-classification 2013-2015					
Change-detected net 2013-2015					
Reference 2015-2016					
Post-classification 2015-2016					
Change-detected net 2015-201					