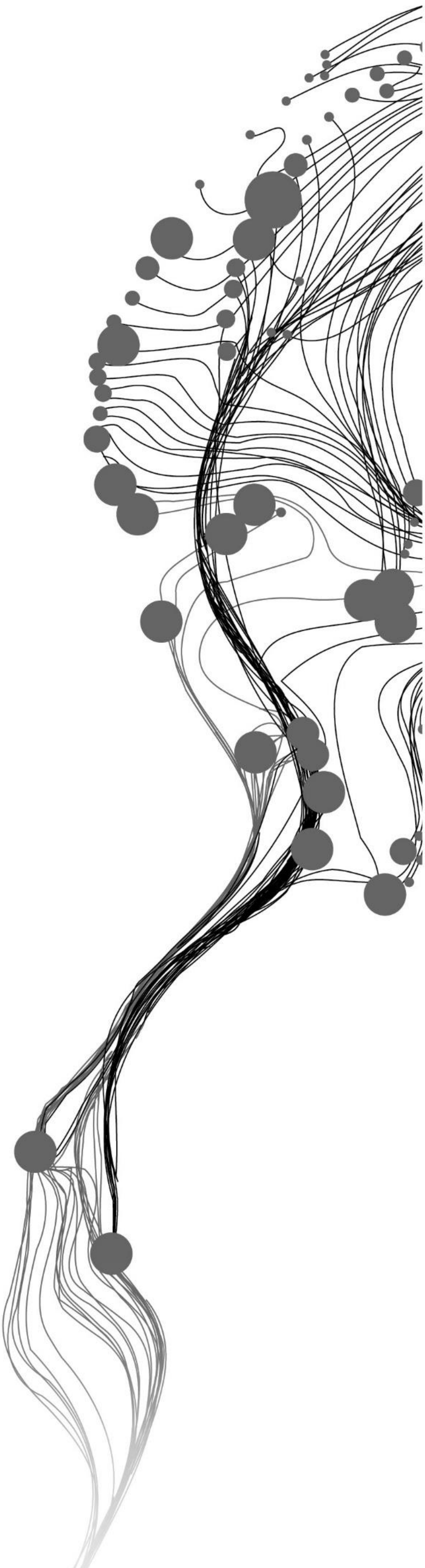
ASSESSING THE SPATIAL TRANSFERABILITY OF FULLY CONVOLUTIONAL NETWORKS FOR SLUM MAPPING

YUNYA GAO
June, 2020

SUPERVISORS:
Dr. M. Belgiu
Dr. M. Kuffer

ADVISORS:
Dr. D. Kohli



ASSESSING THE SPATIAL TRANSFERABILITY OF FULLY CONVOLUTIONAL NETWORKS FOR SLUM MAPPING

YUNYA GAO
Enschede, The Netherlands, June, 2020

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: Spatial Engineering

SUPERVISORS:
Dr. M. Belgiu
Dr. M. Kuffer

ADVISORS:
Dr. D. Kohli

THESIS ASSESSMENT BOARD:
Dr. A. Stein (Chair)
Dr. G. Yeboah (External Examiner, University of Warwick)

ABSTRACT

To fight against poverty, the United Nations have made slum upgrading an important task within the Sustainable Development Goals 11. In support of this target, slum maps providing information about slum spatial location and extent are, thus, significant. In the past few decades, remote sensing (RS) based slum mapping approaches have been developed fast. However, due to the complexity of slums in terms of morphological characteristics, definitions, dynamics and the existence of multiple satellite sensors, transferability has become one of the biggest challenges for RS based slum maping approaches. Based on existing researches, Fully Convolutional Networks (FCNs) have been proved to produce relatively high accuracies for slum mapping compared to other approaches. Existing studies have shown that FCNs are capable of transferring learnt features across different sensors and perform well when tested on the same place at different periods. However, very few studies tested the performance of FCNs when applied to different geographic contexts. For this reason, this research aims to assess the spatial transferability of FCNs for slum mapping.

This research selected Mumbai, Nairobi and Rio de Janeiro (Rio) as study areas whose slums are various in terms of morphological characteristics and conceptualizations. This research designed a systematic assessment framework for the spatial transferability of FCNs for slum mapping. The framework includes three dimensions: (i) what are the differences in selection of FCN architecture and hyperparameter setting to reach optimal performance for different spatial contexts, (ii) whether the model trained on data from one source study area and tested on the corresponding study area performs similarly, and (iii) whether the performance of the model pre-trained on data from one source study area and tested on data from a different place is similar. The selection process of hyperparameter setting for a certain FCN architecture is time-consuming and complex. Due to time limitation, this research explored the second and third dimension of the spatial transferability. Furthermore, this research analysed the influences of three adaptations in training strategies on the performance of the FCN model. Adaptation 1 applies fine-tuning before using the model trained on one source study area to predict slums from a different study area. Adaptation 2 uses training data from multiple study areas rather than only one source study area to train the FCN model. Adpation 3 applies fine-tuning before using the model trained on data from multiple source study areas to predict slums in a different study or one of the selected study areas for training.

The results revealed that the second dimension of the spatial transferability is low. The performance of the FCN model varied when applied to Mumbai (IoU (Intersection over Union) =65.09%), Nairobi (IoU=43.39%) and Rio (IoU=31.42%). The differences in accuracies are mainly caused by different levels of diversity within slums and similarity between slums and non-slums in different study areas. Slums in Mumbai are more homogenous and distinctive from non-slums compared to Nairobi and Rio. Slums in Rio are more heterogeneous and similar to non-slums. Besides, different reference data collections approaches may also influence the performance of FCN models. Slum reference data for Mumbai and Nairobi are image-based, which means slums are mainly determined based on morphological characteristics reflected on satellite imagery and thus may be easier to be detected by RS based approaches. Slum reference data for Rio are ground-based, where slums are determined by morphological, social and economical characteristics. This means it is harder to detect slums in Rio by RS-based appraoches due to incapability of recognizing social and economical characteristics of slums directly from satellite imagery. Besides, slum reference data for Rio includes some non-slums due to data aggragation. For these reasons, the model trained on Mumbai data performs best while the model trained on Rio data performs worst.

For the third dimension, spatial transferability of FCN models is also low. This is mainly because slum morphological characteristics in Mumbai, Nairobi and Rio are different. Therefore, learnt features from the model trained on data from one of the cities are not effective for detecting slums in other cities.

Adaptation 1 makes the FCN model trained on data from one source study area perform similarly when predicting slums in a different study area compared to the model trained on the predicted study area with lower computational cost. It can help improve the third dimension of the spatial transferability but cannot improve the second dimension. The performance of adaptation 3 is similar to the adaptation 1. Both adaptation 1 and 3 perform a bit worse than the adaptation 2. Adaptation 2 helps the FCN model perform better than the model trained on one source study area and tested on the same study area. The results indicate that combining training data from multiple source study areas with different slum characteristics can help improve the performance of the FCN model in both the second and the third dimensions of the spatial transferability. Therefore, adaptation 2 may have potentials to help FCN models map slums at large scales and produce comparable or even higher accuracies as FCN models trained on only one source study area and tested on the same study area.

Key words: slum, fully convolutional networks (FCNs), spatial transferability

ACKNOWLEDGEMENTS

I sincerely appreciate people who provided me help to finish this two-year master degree and this thesis. First of all, I would like to show my sincere gratitude to my great supervisors, Dr. Belgiu and Dr. Kuffer. Thank you so much for your support, guidance and giving me much freedom to implement different experiments, make various expected and unexpected errors and finally make some clear findings from these trials. You are always motivated, hopeful and provided me courage to keep going on when I feel a bit depressed. Without your support, I cannot carry on when the relationships between the model trained and tested on data from Mumbai, Nairobi and Rio were so entangled for me. I really enjoy the meetings with you and talking on different aspects in life, like pretty Alexia, day care, how we changed our physical or human geography background to remote sensing and machine learning background. The thesis is not easy to be done especially in this special period. Thanks so much.

Secondly, I would like to express my sincere appreciation to my parents and my sister. This is the seventh year that I leave our hometown for studying. During these years, I have experienced a lot of amazing things and met many great people. I know you also want to see this world when you were young. But due to many reasons, you gave up and decied to settle down. Now I can understand the depression you experienced after I experienced a lot of dilemma. It takes courage to go out but even more to settle down. Without your support, I cannot have chance to experience these things. Thanks so much for your love and understanding even if I know you didn't hope me to do so actually. I really love you, more than I expected.

Thirdly, I would like to thank all of people in Spatial Engineering, especially Muhammud, Keke, Rui and Kas. You helped me a lot and make me feel at home when we eat and talk together. Without you, I would feel more lonely. Besides, I hope to show my thanks to Tiny and Wietske. You are so kind to help me no matter in life or on study. There are many other people I do hope to show my appreciation to. Thanks so much.

Forthly, I hope to thank dear Yuwen, Yan, Tao, Na and Yu. You bring a lot of joy to my life. We talk a lot on researches, gossips, future, and many other interesting things. Besides, I do hope to thank Wufan. You give me a lot of useful suggestions when I felt lost in various FCN models. Further, I'd like to express special appreciation to Dr. Claudio Persecllo. You helped me a lot and answered my stupid questions patiently. The codes from your course Advanced Image Analysis really helped me a lot.

Last but not the least, I would like to express my appreciation to Spatial Engineering, ITC faculty, the University of Twente, the Netherlands. Thanks so much for providing me enough scholarship to study here. You let me realize personal development not only relies on personal efforts but also support from the society. This kind of support is from the efforts of many generations whom I may never meet. I do hope that I can help people one day as you do. Thanks so much.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Key concepts

1.1.1. Slums

Slums are a complex phenomenon (Kuffer, Pfeffer, & Sliuzas, 2016). Slums are usually regarded as a manifestation of urban poverty and inequality. They develop usually due to the collective effects of fast rural-urban migration and inability of providing sufficient affordable housing by government. (United Nations Human Settlements Programme, 2003)

The main features of slums include poor built-environment (e.g. construction materials, lay-out of buildings), inadequate public service (e.g. water, sanitation, electricity, transportation, schools, solid waste management), social-economic exclusion (e.g. poverty level, security of tenure, crime and safety) and bad ecological conditions (e.g. green space, harzards) (Lilford et al., 2019).

Many efforts have been made to conceptualize slums around the world, such as expert meetings (Sliuzas, Mboup, & de Sherbinin, 2008; UN-Habitat, 2017) and published conceptualization frameworks (Lilford et al., 2019). However, there is still no universal definition of slums due to their spatial diversity and temporal dynamics. In addition, the features mentioned can also occur in non-slum areas, which makes it more complicated to define slums (Lilford et al., 2019). For example, although slums and urban poverty are usually co-located, not all slum dwellers are classified as the poor (United Nations Human Settlements Programme, 2003).

One of the widely accepted definitions at household level is from UN-Habitat which defines slums or informal settlements as urban areas where the majority of households face one or more of the following challenges: (1) lack of durable housing; (2) insufficient living space; (3) difficult access to safe water; (4) demanding access to enough sanitation; (5) tenure insecurity (Lilford et al., 2019; UN Habitat, 2007).

Satellite imagery have been used for slum mapping in the past few decades. Morphological characteristics (e.g. density, size, pattern) of slums which are different from non-slums can be helpful for remote sensing (RS) based slum mapping approaches (Kuffer, Pfeffer, & Sliuzas, 2016). However, social-economic characteristics of slums cannot be reflected from satellite imagery directly (Mahabir, Croitoru, Crooks, Agouris, & Stefanidis, 2018). This incapability constrains the performance of RS-based approaches.
This research focuses on RS-based slum mapping approaches. For this reason, slum characteristics to be analyzed in this research expressly include the morphological characteristics of slums, such as shape, size, roof materials, which usually can be detected from satellite imagery (Kuffer, Pfeffer, & Sliuzas, 2016). Figure 1.1 shows the scope of slum features to be focused in this research.

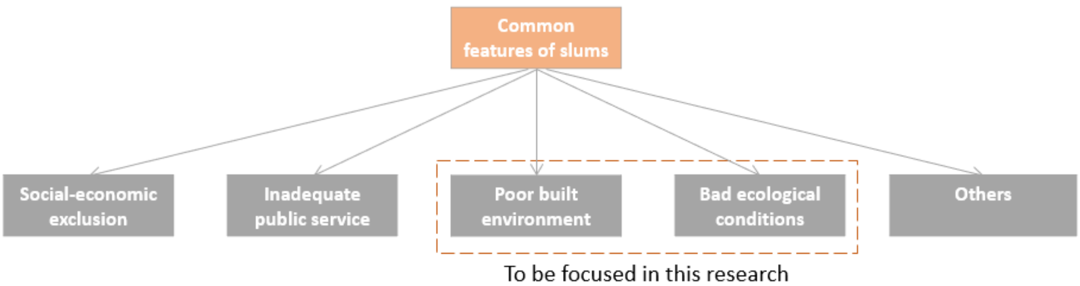


Figure 1.1 Common features of slums (Lilford et al., 2019) and the features to be focused in this research

1.1.2. Spatial transferability

Transferability is defined as "the quality of being transferable" (Vocabulary.com, n.d.). The word "transferable" has multiple meanings. One of them to be used in this research is described as "suitable for different situations or uses" (Cambridge Business English Dictionary, n.d.). Hence, in general, transferability refers to the an ability to be suitable for different situations.

Pratomo, Kuffer, Martinez, & Kohli (2016) define the transferability as an ability of an approach to perform similarly with minimum changes, when applied to various situations. Based on this definition, transferability has two aspects. First, the selected approach can perform similarly or comparably when applied to different situations. Second, it is not necessary to make big changes in the approach to achieve comparable performance. The changes can be different for different approaches. Some Object-Based Image Analysis (OBIA) approaches, for example, involves the definition of classification rulesets to identify target classes. Hence, to assess their transferability, it is essential to figure out how the rulesets need to be changed to perform similarly (Pratomo, Kuffer, Kohli, & Martinez, 2018). For data-driven approaches such as most deep learning approaches, it is essential to test how learnt information from one dataset can be helpful and what techniques can be adopted to transfer the learnt information when applied to another task. Figure 1.2 summaries the above description of the transferability of an approach.

Based on the above description of transferability, spatial transferability in this research is defined as the ability of an approach to perform similarly with minimum changes, when applied to different geographic contexts. The process of assessing spatial transferability includes applying the selected approach after some changes to different geographic contexts (Place1, Place2…PlaceN) and then comparing the conducted performance (Performance1, Performance2…PerformanceN). Figure 1.3 shows the process of assessing spatial transferability.

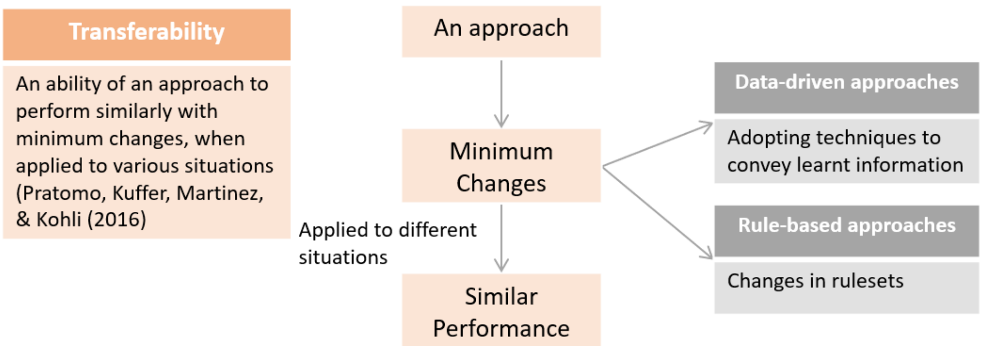


Figure 1.2 The definition of transferability and potential "minimum changes" for two types of approaches

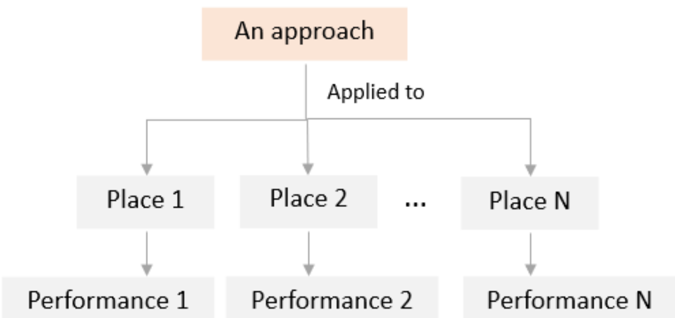


Figure 1.3 The process of assessing spatial transferability

1.2. Justification of the research topic

Up to now, around a quarter of the global urban population lives in slums. The number of slum dwellers has reached more than 1 billion and is estimated to reach 1.5 billion in 2025 (Willis, 2019). These slum dwellers have experienced, are experiencing, or will experience different degrees of deprivation (Ajami, Kuffer, Persello, & Pfeffer, 2019). Adequate housing is one of the human rights recorded in the Universal Declaration of Human Rights (De Schutter, 2014). Inadequate housing such as commonly found in slums can lead to urban inequity, exclusion, unsafety, unfair livelihood opportunities and other problems (Willis, 2019). Thus, the Sustainable Development Goal (SDG) 11 - "Make cities and human settlements inclusive, safe, resilient and sustainable", includes tasks of monitoring the right to adequate housing. The "proportion of urban population living in slums" is selected as SDG indicator 11.1.1 (UN Habitat, 2019, p.2).

Slum upgrading is a major task of SDG 11, which means raising slum households' living standards and helping them get rid of slum-like conditions (UN Habitat, 2019). It is a wicked process that involves diverse stakeholders and social, economic, environmental, financial, governance issues (Willis, 2019). It is essential for these stakeholders to make decisions based on a good understanding of the spatial distribution and characteristics of slums. Slum maps, as a medium of slum spatial information, are, thus, important for upgrading slums. Nowadays, many cities have not mapped slums because some governments neglect them due to their informality. For many other cities, available slum maps are commonly incomplete or outdated (Kohli, Warwadekar, Kerle, Sliuzas, & Stein, 2013). Therefore, it is necessary to develop effective and efficient approaches to map slums more accurately for slum upgrading.

Up to now, there are mainly three categories of slum mapping approaches, including survey-based approaches, participatory approaches and RS-based approaches (Mahabir et al., 2018). Surveys are used for collecting slum data in many countries every ten years. Due to fast changes within urban areas, data from surveys are usually outdated. Furthermore, slums are often ignored in these formal surveys (Joshi, Sen, & Hobson, 2002). Participatory approaches need the participation of local people. Hence, it takes a lot of time and financial resources to implement them (Kohli et al., 2013). RS-based approaches are usually less time-consuming, less resource-intensive and can help update slum maps more frequently. Thus, they have received a lot of attention from researchers (Kuffer, Pfeffer, & Sliuzas, 2016).

With the development of RS technologies in the past few decades, especially increasing availability of very high resolution (VHR) RS imagery, RS-based slum mapping approaches have been developed fast with the efforts of many researchers (Kuffer, Pfeffer, & Sliuzas, 2016). The primary task of most approaches is to select or design efficient texture and spectral features or rulesets, which can differentiate between slums and non-slum areas from satellite imagery (Mahabir et al., 2018). These methods are, however, challenged by the fact that the slums have different definitions and characteristics in different places and at different moments (Kuffer, Pfeffer, & Sliuzas, 2016). Furthermore, the appearances of the same slums can vary across various sensors (Wurm, Stark, Zhu, Weigand, & Taubenböck, 2019). This complexity makes features or rulesets designed for slums mapping suitable for one situation but may perform poorly in other circumstances. Therefore, transferability has become one of the biggest challenges in the RS-based slum mapping research domain (Kuffer, Pfeffer, & Sliuzas, 2016; Mahabir et al., 2018). In the case of traditional machine learning methods such as Support Vector Machine (SVM) and Random Forest (RF), their performance for slum mapping is relatively high. However, they rely on feature selection which requires a clear understanding of slum characteristics (Leonita, Kuffer, Sliuzas, & Persello, 2018). As mentioned above, it is hard to conceptualize a universal concept of slums. Consequently, these methods face transferability issues in slum mapping.

Compared to traditional machine learning approaches, Fully Convolutional Networks (FCNs) do not require feature selection. Instead, FCNs can make use of features determined from imagery data to detect slums and usually achieve higher accuracies than traditional machine learning approaches (Persello & Stein, 2017). Thus, FCNs are a promising approach for slum mapping. However, only a reduced number of studies are dedicated to investigating the transferability of FCNs in slum mapping. Wurm, Stark, Zhu, Weigand, Taubenböck (2019) investigated the transferability of FCN-VGG19 across different sensors. FCN-VGG19 is adapted from VGG19, which is a classic Convolutional Neural Networks (CNNs) architecture designed by the Visual Geometry Group of Oxford University (Simonyan & Zisserman, 2015). Wurm, et al. (2019) found out that the accuracy of the model trained on QuickBird imagery (IoU(Intersection of Union)=77%) is higher than that on Sentinel-2 imagery (IoU=36%). Then, they applied transfer learning to keep learnt features from QuickBird model and then trained on Sentinel-2 imagery. It turned out that the new model performed better than the previous Sentinel-2 model with an IoU of 51% (Wurm et al., 2019). Transfer learning is an approach to repurpose a model trained on one dataset to other classification tasks (Verma, Jana, & Ramamritham, 2019). The result suggests that transfer learning can help transfer knowledge from one task to another task, in general, and also for slum mapping in terms of satellite sensors. Liu, Kuffer and Persello (2019) applied FCN-DK6 to detect small and temporal slums in Bangalore from VHR imagery. They found out that the prediction results of imagery from different years are high (average F1 score=88.4%). Stark, Wurm, Taubenock and Zhu (2019) used FCN-VGG19 to detect slums in Mumbai and Delhi and found out that IoU of the model trained on Mumbai data is 66%. IoU of the model trained on Delhi data is 49%. This result indicates that FCNs may produce different accuracies when mapping slums from different geographic contexts.

Based on the previous researches, FCNs have been proved to be transferable across different sensors and time periods. Yet, FCNs face some challenges in the case of spatial transferability. Spatial transferability is regarded as one of the most major bottlenecks for slum mapping researches. Besides, spatial transferability of approaches are especially important for regions with sparse reference data of slums (Kuffer, Pfeffer, & Sliuzas, 2016). However, studies dedicated to assessing the spatial transferability of FCNs for slum mapping are missing. Therefore, this research aims to fill this gap.

In conclusion, RS-based slum mapping approaches have developed a lot in the past few decades. However, transferability remains an open issue. The primary reason is that slums are too complicated in terms of conceptualizations and characteristics. FCNs have been proved to perform better than other traditional slums mapping methods in most cases. In addition, FCNs do not require pre-designed features or rulesets. Therefore, they may have high spatial transferability for slum mapping. Therefore, this research aims to assess the spatial transferability of FCNs for slum mapping to promote the development of slum mapping approaches in support of slum upgrading projects globally.

1.3. Research gap and innovation points

Up to now, very few studies research on assessing the spatial transferability of FCNs. Stark (2018) tested the spatial transferability of FCN-VGG19. However, he only tested on two study areas: Mumbai (IoU=66%) and Delhi (IoU=49%). Both two cities are from India. The findings may not be universal due to the variety of slums globally. Furthermore, a clear definition of the meaning of spatial transferability of FCNs for slum mapping and a systematic assessment framework are missing. Therefore, it is necessary to design a systematic assessment framework for spatial transferability of FCNs for slum mapping.

The primary innovation point of this research is assessing the spatial transferability of FCNs for slum mapping by testing on data from three cities (Mumbai, Nairobi and Rio de Janeiro) with different characteristics of slums under different social-economic conditions based on a systematic assessment framework.

1.4. Research objectives and questions

1.4.1. Main objective

The primary research objective of this research is to systematically assess the spatial transferability of FCNs for slum mapping.

1.4.2. Sub-objective

To achieve the main objective, three sub-objectives have been made:

(1) To design a systematic assessment framework for spatial transferability of FCNs for slum mapping;
(2) To design suitable experiments for the spatial transferability assessment based on the framework in (1);
(3) To assess the spatial transferability of FCNs based on results from the experiments designed in (2);

1.4.3. Research questions

The following research questions are addressed in this research:

(1) To design a systematic assessment framework for spatial transferability of FCNs for slum mapping;
    1) How to measure the spatial transferability of FCNs for slum mapping?
    2) What adaptations are required to improve the spatial transferability of FCNs for slum mapping?
(2) To design suitable experiments for the spatial transferability assessment;
    1) What experiments are essential for the assessment based on the framework from (1)?
    2) What is the optimal FCN architecture for slum mapping?
    3) What are the proper hyperparameters of the selected FCN architecture for slum mapping?
(3) To assess the spatial transferability of FCNs based on results from the experiments designed in (2)
    1) What is the performance of FCNs in terms of spatial transferability based on the results from the experiments in (2)?
    2) What are the effects of the adaptations from (1) 2) on the spatial transferability of FCNs?

1.5. Relations to Spatial Engineering

Spatial engineering aims to help students cultivate a capability to solve wicked problems in reality by multidisciplinary solutions and broad thinking. Wicked problems usually happen when there is no consensus among stakeholders. The dissensus usually is caused by knowledge gaps, disagreement on goals (or values), and selection of technologies (Hoppe, 2018).

Slum upgrading can be regarded as a wicked problem because this task involves many stakeholders with different and dynamic interests (UN-Habitat, 2020). These stakeholders include slum dwellers, NGOs (Non-Government Organizations), various government ministries, local real estate associations, researchers etc. Participation of the involved stakeholders is important to reach a commitment to improving the living conditions of slum dwellers. It is essential to combine ideas and inputs of these stakeholders to strengthen the links between public services, transportation, infrastructures, etc. (UN-Habitat, 2020) Before integrating multiple views from stakeholders, one of the most important pre-conditions for the participation is that they have a common understanding of slum situations. Slum maps can help provide spatial information of slums and be used for scenario analysis (Carr-Hill, 2013). Thus, they can be helpful for these stakeholders to share their ideas and analyse the effects of proposed plans on a common base. However, there are still many areas without accuracte slum maps (van Steensel, 2016).

For these reasons, spatial information of slums (slum maps) becomes a knowledge gap during the process of designing slum upgrading plans. This research aims to assess the spatial transferability of FCNs for slum mapping. If the spatial transferability of the FCN model is high, then it is possible to map slums for areas without slum maps by using the model trained on data from other places. In other words, this research tries to help narrow down knowledge gaps in terms of slum spatial information for data sparse areas by assessing the spatial transferability of FCNs for slum mapping.

2. LITERATUE REVIEW

This chapter reviews research papers related to this thesis. Section 2.1 introduces the complexity of slums. The complexity is one of the primary reasons for transferability issues faced by existing RS-based slum mapping approaches. Section 2.2 summarizes the advantages and disadvantages of some RS-based slum mapping approaches, including visual interpretation, OBIA and machine learning (ML) approaches. Section 2.3 introduces deep learning approaches and their applications in slum mapping. Section 2.4 summarizes the main findings about the transferability of FCNs. Section 2.5 presents applied accuracy indicators in slum mapping researches.

2.1. Complexity of slums

This section mainly introduces the complexity of slums in terms of conceptualizations, characteristics, causes and influences on society.

Slums have various definitions in different regions based on different standards at different periods (Verma et al., 2019). It is hard to make a universal definition of slums. The definitions are mainly dependent on local-level political decisions. Different local authorities emphasize on different components for slums, such as construction materials, land tenure security, health and hygiene, crowding, basic services and infrastructures (e.g. water, sanitation, electricity), low-income, crime and violence (United Nations Human Settlements Programme, 2003). For example, the government of Bangkok accounts for health and hygiene, crime and violence, crowding and environment in the definition of slums. In Jakarta, land legality and low-income are two most important components for slums (United Nations Human Settlements Programme, 2003). For most definitions of slums from different places, poor construction materials and land tenure security are the two most commonly adopted components. Besides, it is essential to point out that not all slum dwellers are classified as the poor (United Nations Human Settlements Programme, 2003).

Lilford et al. (2019) introduced two basic approaches to conceptualize slums. The first approach is "feature first" (or called as bottom-up). An area is labelled as slums or non-slums by taking into account the observed features as mentioned in the last paragraph and standards of local authorities. This approach usually relies on surveys at the household level. The second approach is "space first" (or called as top-down). In this approach, an area is selected and classified as slums or non-slums at first. Then, the household is classified as slum if it is identified as originating in a slum.

Different terms are usually used to refer to slums. These terms and slum have been used in literature interchangeably (Kuffer, Pfeffer, & Sliuzas, 2016). For example, terms with descript tion like "informal", "illegal" or "squatter" emphasize the insecure status of land tenure. "Unplanned" is usually related to planning context. "spontaneous" or "irregular" highlighting the dynamics of slums. "Deprived," "shantytown" and "sub-standard" are related to physical and socio-economic conditions (Kuffer, Pfeffer, & Sliuzas, 2016).

Slums have various physical characteristics, such as higher roof coverage densities, more organic patterns, and smaller building sizes compared to non-slum built-up areas (Kuffer, Pfeffer, & Sliuzas, 2016). Some features, such as density, can be clearly defined, while others, such as land tenure, are more difficult to be identified. However, even if the definitions of features are clear, their measurement using RS-based approaches can be problematic (Pratomo et al., 2018). Due to lacking of the local knowledge on the characteristics and concepts of slums, there are some fuzzy classes such as semi-informal settlement that has morphological characteristics similar to slums but are historic areas (Kuffer, Pfeffer, & Sliuzas, 2016). The similarity in physical appearances of slums and these non-slum areas from RS imagery hence brings more uncertainty to RS-based approaches.

The formation of slums usually originated from fast urban population expansion. The expansion is often triggered by natural population growth, rural-urban migration, population displacement due to conflicts and or violence (United Nations Human Settlements Programme, 2003). The increase of population together with a poor governance and lack of land lead to the development of slums (UN-Habitat, 2013).

Slums bring many problems such as poverty, inequality, and diseases, which affect the sustainable development of cities. However, slums are usually the only affordable housing option for slum dwellers. They usually face challenges of discrimination and spatial-economic exclusion (UN-Habitat, 2016). Given that many slums are built on hazardous regions, they are vulnerable to natural disasters. Due to poor living conditions, it is highly possible to spread diseases within slum areas (UN-Habitat, OHCHR, & UNOPS, 2016).

2.2. RS-based slum mapping

RS-based slum mapping approaches have developed fast in the past several decades, especially with the advent of VHR imagery (Mahabir et al., 2018). However, there is still no universal approach for slum mapping (Kuffer, Pfeffer, & Sliuzas, 2016).

Existing approaches could be mainly divided into several types: visual image interpretation, OBIA-based approaches, ML-based approaches (Mahabir et al., 2018).

Visual image interpretation can produce slum maps with high accuracy (Wurm and Taubenböck, 2018; Taubenböck et al., 2018). Other slum mapping approaches usually make use of the results by this approach as reference data. However, it is time-consuming, and it also has uncertainties such as fuzziness of boundaries mainly caused by multiple perceptions of slums (Pratomo et al., 2018).

OBIA is one of the most frequently applied techniques for slum mapping (Kuffer, Pfeffer, & Sliuzas, 2016). Urban objects are often complicated. They consist usually of multiple heterogeneous parts. For instance, a building can have several parts made up of different materials with different spectral characteristics. Consequently, pixel-based approaches usually face some challenges, such as salt-and-pepper effect, mainly because they rely on spectral information solely (Kohli et al., 2013). OBIA approaches have the potential to combine spatial, spectral, and contextual properties of the target objects for classification purposes. Besides, they can make use of physical proxies, e.g. grey-level co-occurrence matrix, to determine the characteristics of the objects of interest. Hence, it usually performs better than traditional pixel-based approaches (Kohli et al., 2013). Kohli, Sliuzas and Stein (2016) tested the accuracy of OBIA by using data from different areas in Ahmedabad. The accuracies range from 47% to 68%. OBIA have promoted the development of slum mapping approaches significantly (Kuffer, Pfeffer, & Sliuzas, 2016). However, it is essential to clarify the concepts of slums when designing an effective ruleset to detect slums (Kohli et al., 2013). Due to the complexity of slums, the transferability of the developed rulesets is one of the biggest obstacles faced by OBIA approaches (Kohli et al., 2013). Though the rulesets of OBIA can also be learnt by combining with traditional ML approaches such as SVM in a data-driven way, feature selection of the traditional ML approaches also faces transferability issues (Zahidi, Yusuf, Hamedianfar, Shafri, & Mohamed, 2015).

ML-based approaches are also frequently applied to slum mapping. They are data-driven approaches which learn the characteristcis of the target class from training data repeatedly (Baud, Kuffer, Pfeffer, Sliuzas, & Karuppannan, 2010). Therefore, they can perform well if there is a large size of training data. Leonita, Kuffer, Sliuzas, and Persello (2018) explored the performance of SVM and RF for slum mapping. They found out that SVM achieved an F1 score ranging from 0.73 to 0.92, and RF yields an F1 score ranging from 0.72 to 0.94. In general, previous studies found ML-based approaches are superior to many other slum mapping approaches. However, the performance of these traditional ML approaches relies heavily on feature selection, which requires a clear understanding of slum characteristics (Leonita et al., 2018). As mentioned above, it is hard to make a universal quantitative measurement of slums. Consequently, these approaches also face transferability issues.

Recently, deep learning has received an increasing attention for slum mapping (Persello & Stein, 2017). So far, only a few studies used CNNs and FCNs for slum mapping. This supervised learning process of deep learning approaches can help learn weights and bias of models to reduce prediction errors (Persello & Stein, 2017). Hence, it is unnecessary to pre-design rulesets or select features for these approaches, which makes them less dependent on the conceptualization of slums. Therefore, deep learning approaches may have high transferability for slum mapping. Section 2.3 introduces more details for these approaches.

2.3. Deep learning-based slum mapping

Deep learning algorithms try to learn multiple features from input training data, which do not require manually designed features. They usually consist of more than two hidden layers (Zhu et al., 2017). CNNs are important image classification approaches in deep learning. Section 2.3.1 and 2.3.2 introduce the basics of CNNs and FCNs (adapted from classic CNNs) and their applications in slum mapping. Section 2.3.3 demonstrates general training strategies for training neural networks. In this research, CNNs refer to neural networks for patch-based image classification and FCNs refer to that for pixel-based image classification.

2.3.1. CNNs-based slum mapping

Classic CNNs such as VGG19 are patch-based approaches. The output of CNN models is a classification label for the central pixel of the whole input image (Jean & Luo, 2016). Figure 2.1 shows a simplified CNN architecture adapted from O'Shea and Nash (2015). In general, CNNs have multiple layers to implement three types of operations, namely convolutions, non-linear activations and pooling. A standard CNN architecture consists of several convolutional layers and fully connected (FC) layers. Convolutional layers can help learn features from input data. FC layers are one-dimensional vectors flatten by learnt features from convolutional layers. They are responsible for learning classification rules (Persello & Stein, 2017). Architectures of CNNs include both feature extraction processes and classification processes. Thus, they are trained in an end-to-end way.

Many studies have proved that CNNs can outperform many other approaches based on hand-made features (Persello & Stein, 2017). Vermaa, Janaa, and Ramamritham (2019) applied a CNN model to map slums for Mumbaiand. They obtained an IoU of 0.58 when using VHR imagery as input data and a IoU of 0.43 when using MR imagery. Xie, Jean, Burke, Lobell, and Ermon(2016) applied CNNs to make global poverty maps by means of nightlight satellite imagery. One primary obstacle faced by CNNs when applied to large satellite imagery for slum mapping is the high computational cost (Persello & Stein, 2017). The number of learnable parameters in FC layers is much more than that of convolutional layers. FCNs, adapted from CNNs, may help solve this problem.

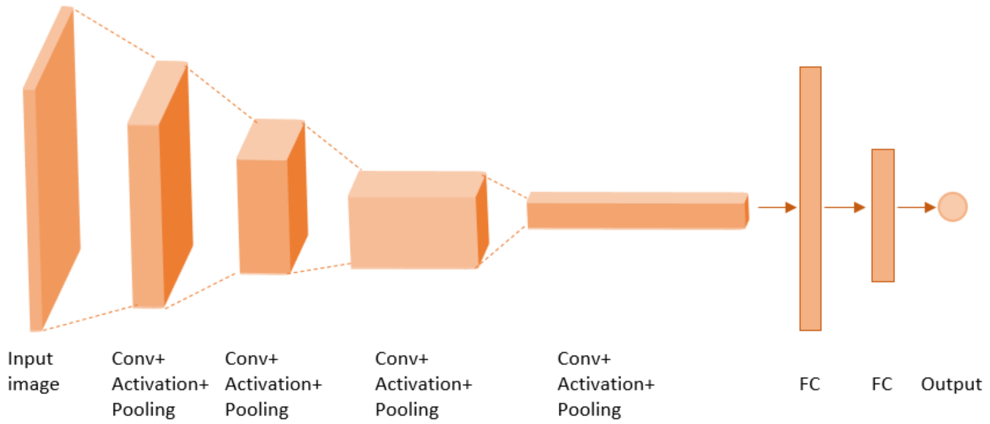


Figure 2.1 A simplified architecture of CNNs adapted from O'Shea and Nash (2015)

2.3.2. FCNs-based slum mapping

FCNs are pixel-based classification approaches and are also trained in an end-to-end way. They are also known as semantic segmentation. FCNs delete FC layers and adopt a convolution-deconvolution (encoder-decoder) strategy or dilated convolutions to keep the size and resolution of output prediction maps the same as input images (Long, Shelhamer, & Darrell, 2015; Persello & Stein, 2017; Wurm et al., 2019). Hence, FCNs requires less computational resources than classic CNNs. Table 2.1 summaries applied FCN architecture, study areas, satellite imagery, accuracies of existing related researches for slum mapping. For most studies reported in this table, there are usually more than one accuracy values due to multiple experiments. Only maximum accuracy values are depicted here.

Table 2.1 Applied FCN architecture, study areas, satellite imagery, accuracies of existing related researches in slum mapping. (IoU: Intersection over Union; PA: Producer Accuracy; TL: Transfer Learning)

Research	**Architecture**	**City (Country)**	**Satellite imagery**	**Accuracy indicator**	
				Variable	**Value**
Persello & Stein (2017)	FCN-DKs	Dar es Salaam (Tanzania)	Quickbird	FCN-DK3	PA = 58.29%
				FCN-DK4	PA = 58.16%
				FCN-DK5	PA = 62.09%
				FCN-DK6	PA = 65.58%
Stark, (2018)	FCN-VGG19	Mumbai (India)	QuickBird	Mumbai	IoU = 66.12%
		Delhi (India)		Delhi	IoU = 48.85%
Wurm, Stark, Zhu, Weigand, & Taubenböck, (2019)	FCN-VGG19	Mumbai (India)	QuickBird	QuickBird	IoU = 77.02%
			Sentinel-2	Sentinel-2	IoU = 35.51%
			QuickBird - TL - Sentinel-2	QuickBird - TL - Sentinel-2	IoU = 51.23%
Stark, Wurm, Taubenbock, and Zhu (2019)	FCN-VGG19	Mumbai (India)	QuickBird	Mumbai - TL - Delhi	IoU = 59%
		Delhi (India)		Delhi - TL - Mumbai	IoU = 34%
Liu, Kuffer, & Persello, (2019)	FCN-DK6	Bangalore (India)	WorldView	F1 score = 88.38%	

Encoder-decoder FCN architectures usually include the following parts: (1) convolutional layers; (2) no-linear activation functions (e.g. leaky Rectified Linear Unit (lReLU)); (3) pooling (e.g. max pooling); (4) deconvolutional layers; (5) classification layers (e.g. Softmax). Deconvolution is usually realized by skipped connections which are significant for improving performances and avoiding overfitting for FCNs (Panboonyuen, Jitkajornwanich, Lawawirojwong, Srestasathiern, & Vateekul, 2017). Skipped connections refer to additional connections between nodes in different layers within a neural network. The connections skip several layers of nonlinear activation (Graesser et al., 2012). Similar to CNNs, convolutional layers are used for learning features of input imagery and encoding location. Deconvolutional layers have the same functions. Classification layers are for learning prediction rules (Zhang et al., 2018). Figure 2.2 shows a simplified architecture of encoder-decoder FCN models modified from Peng, Zhang, and Guan (2019).

Some studies have used encoder-decoder FCNs for slum mapping. Wurm et al. (2019) used FCN-VGG19 to detect slums in Mumbai and to test the transferability of FCNs across different sensors for slum mapping. The authors found out that the model trained on QuickBird imagery reaches an IoU of 77%. Stark, Wurm, Taubenbock, and Zhu (2019) used FCN-VGG19 to detect slums in Mumbai and Delhi and tested the influences of the proportion of slum labelled data in training data on the yielded prediction accuracy. They obtained an IoU of 72% for Mumbai and 69% for Delhi by using pre-trained weights from ImageNet. Besides, both of these studies proved that transfer learning is useful in transferring knowledge of slums from one task to another task.

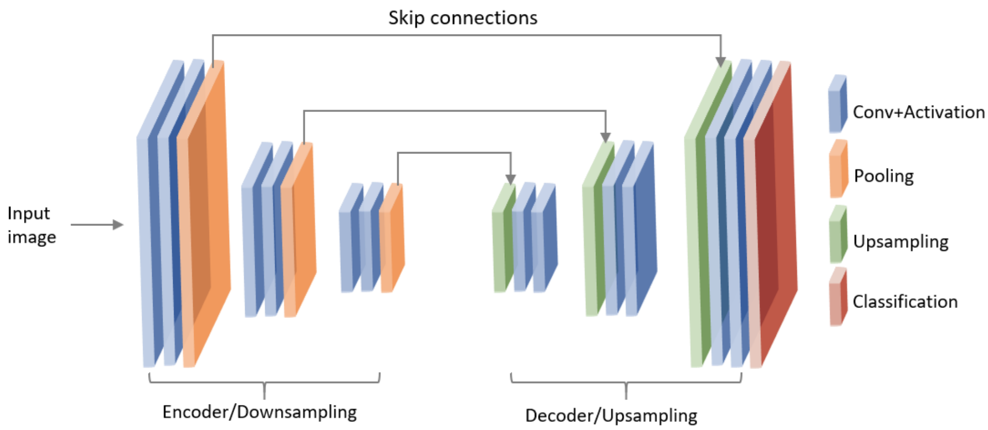


Figure 2.2 A simplified architecture of encoder-decoder FCNs based on Peng et al. (2019)

FCN architectures with dilated convolutions make use of dilated kernels (DKs) to increase the sizes of the receptive fields (RFs) and, thus, these architecture do not need upsampling layers. This type of FCN architectures is called as FCN-DKs (Persello & Stein, 2017).

FCN-DKs can reduce both the number of parameters to avoid overfitting and computational cost. In general, these architectures consist of several blocks. Each block usually includes zero-padding layers, convolutional layers with different dilated rates in different blocks, activation layers and pooling layers. At the end of the model, there is a classification layer. Zero paddings are important for FCN-DKs to keep the size of output predictions as input images (Persello & Stein, 2017). Figure 2.3 presents a simplified architecture of FCN-DKs adapted from Persello and Stein (2017).

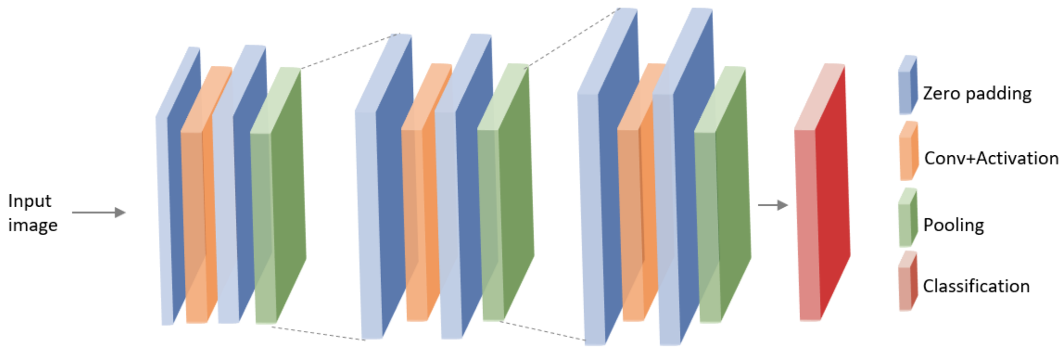


Figure 2.3 A simplified architecture of FCNs with dilated kernels modified from Persello and Stein (2017)

Several studies have applied FCN-DKs for slum mapping. Persello and Stein (2017), for example, used patch-based CNN, SVM, FCN-DK3, FCN-DK4, FCN-DK5 and FCN-DK6 to map slums in Dar es Salaam by using QuickBird satellite imagery. They found that FCN-DK6 outperformed the other evaluated FCN-DKs, obtaining an overall accuracy (OA) of around 84% (Persello & Stein, 2017).

2.3.3. Training strategies

There are mainly two ways to train deep learning models. The first way is to apply fine-tuning or transfer learning to adapt a pre-trained model to meet the requirements of target tasks with less labelled training data and less computational cost. For fine-tuning, we can decide to freeze some layers in a pre-trained model or make all layers trainable before using it for another task by training it on another dataset. Transfer learning takes layers of a pre-trained model, freeze these layers, then add new trainable layers after the frozen layers and, in the end, train the new trainable layers with new datasets. The second way is to train deep learning models from scratch. Generally, the accuracy of the second way is lower than the first way. Training from scratch is usually more computationally expensive. Stark et al. (2019) compared the accuracies of FCN-VGG19 under two training strategies. The first model is fine-tuned from VGG19 pre-trained on ImageNet. The second model is trained from scratch. They found out that IoU of the first model is 0.69, while the second model yielded an IoU of 0.34.

2.4. Transferability of RS-based slum mapping approaches

The transferability of RS-based slum mapping approaches contains four aspects: conceptual transferability, spatial transferability, temporal transferability and transferability across RS sensors (Kuffer, Pfeffer, & Sliuzas, 2016; R. Liu, 2018; Stark, 2018; Stark, Wurm, Taubenbock, & Zhu, 2019). Figure 2.4 shows the four transferability dimensions of slums.

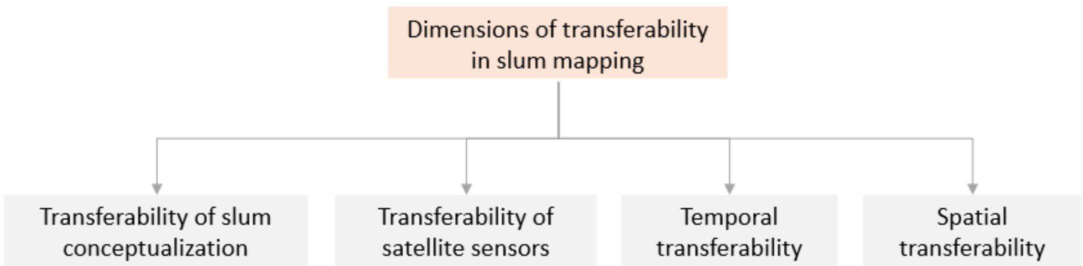


Figure 2.4 Dimensions of transferability of slum mapping approaches

Several studies focused on the transferability of OBIA (Kohli et al., 2013; Pratomo, Kuffer, Martinez, & Kohli, 2016; Pratomo, 2016; Pratomo et al., 2018; Hofmann, Blaschke, & Strobl, 2011). Kohli, Warwadekar, Kerle, Sliuzas, and Stein (2013) used OBIA to detect slums from RS imagery of different areas in Ahmedabad. They defined transferability of OBIA as "the degree to which a particular method is capable of providing comparable results for other images". Pratomo, Kuffer, Kohli, and Martinez (2018) used the trajectory error matrix (TEM) to measure the temporal transferability of an OBIA ruleset for slum mapping in Jakarta. They emphasized two primary reasons accounting for low transferability of the ruleset: (1) uncertainty of fuzzy boundaries of reference data of slum and non-slum areas; and (2) different viewing angles of the input images. Pratomo, Kuffer, Martineza, Kohli (2016) analysed spatial and temporal transferability of Generic Slum Ontology (GSO) and Local Slum Ontology (LSO) for OBIA. The authors concluded that GSO performs better than LSO in spatial transferability. Yet, LSO performs better than GSO for temporal transferability.

Only a few studies have focused on the transferability of FCNs. Wurm et al. (2019) evaluated the transferability of FCN-VGG19 on imagery from different sensors. They concluded that the accuracy of the model trained on QuickBird imagery (IoU=77.02%) is higher than the model trained on Sentinel-2 imagery (IoU=35.51%). Then, they applied transfer learning to train the model pre-trained on QuickBird imagery by using Sentinel-2 imagery. IoU of this new model reaches 51%. Liu, Kuffer, and Persello (2019) applied FCN-DK6 to detect small and temporal slums of Bangalore by using VHR imagery. They found out that FCNs can perform well when using imagery from the same place but at different time. Yet, the accuracy of change detection through FCNs drops. Stark (2018) used FCN-VGG19 to detect slums in Mumbai and Delhi. The model trained on Mumbai data reaches an IoU of 66.12% and an IoU of 48.85% when trained on Delhi data. The author adopted the same training strategy for two cities. Furthermore, this study evaluated the effects of fine-tuning and combination of training data from two cities on the performance of FCNs. The study showed the IoU of the model trained on combined data from Mumbai and Delhi increases to around 67.65%. This accuracy is higher than that of the models trained on data from Mumbai or Delhi individually. The improvement indicates that combining training data from multiple geographic contexts may improve prediction accuracies of FCN models.

To conclude, existing researches dedicated to assessing the transferability of FCNs explored spatial, temporal and sensor dimensions of transferability. However, there are three limitations of these studies:

(1) For the transferability of optical sensors, only two sensors have been compared. Given that there are much more optical sensors with different spectral characteristics, more researches are required;
(2) For temporal transferability, FCNs perform well on imagery from the same area but at different time periods. However, FCNs perform worse in change detection;
(3) For spatial transferability, existing researches only tested it on two cities. Both of them are from India. It is not enough due to variety of slums globally. More study areas with different characteristics under various cultural and social background should be explored in the spatial transferability.

2.5. Accuracy indicators applied for slum mapping approaches

To compare the performance of different slum mapping approaches, it is essential to apply common accuracy indicators. However, in existing related researches, accuracy indicators are various. To make this research easily compared to other researches, this section will review accuracy indicators applied in slum mapping researches to assess the results of FCNs and other approaches in recent years. Table 2.2 shows numerous accuracy indicators which have been applied in existing studies of slum mapping by using FCNs. Table 2.3 shows various accuracy indicators applied by some other researches recently. It can be found that precision, recall, IoU (also known as Jaccard Index), F1-score, Overall Accuracy (OA) are frequently applied indicators.

Table 2.2 Accuracy indicators applied in researches of slum mapping by FCNs (OA: Overall Accuracy; PA: Producer Accuracy; IoU: Intersection over Union; PPV: Positive Prediction Value)

Research	Accuracy indicators
Persello & Stein (2017)	OA, Recall
Stark, (2018)	OA, IoU, Kappa estimate, Precision, Recall
(Wurm, Stark, Zhu, Weigand, & Taubenböck, 2019)	PPV, IoU
Stark, Wurm, Taubenbock, and Zhu (2019)	IoU
Liu, Kuffer, & Persello, (2019)	Precision, Recall, F1-Score

Table 2.3 Accuracy indicators applied in researches of slum mapping in recent years (OA: Overall Accuracy; IoU: Intersection over Union)

Research	Approaches	Accuracy indicators
Kohli, Sliuzas, & Stein(2016)	Grey-Level Co-occurrence Matrix	OA, Precision, Recall
Maiya & Babu(2018)	Mask R-CNN	IoU, Recall
Leonita et al.(2018)	Machine learning (SVM, RF)	OA, Kappa estimate, F1-Score
Verma et al.(2019)	Typical CNN	OA, Kappa estimate, IoU
(Ranguelova et al., 2019)	Bag of Visual Words framework and Speeded-Up Robust Features	OA, Precision, Recall, F1-Score

3. STUDY AREAS AND DATA

3.1. Study areas

This research selected three cities, Mumbai, Nairobi and Rio, as study areas. These cities have been selected because they are from different countries under different cultural background and have different morphological characteristics reflected from satellite imagery. Besides, we had access to reference data for all of them. As seen from the description in the following sub-sections, not only slums are different in these three cities, but also the similarity of slums and non-slums is different.

3.1.1. Mumbai

Mumbai, also called as Bombay, is the capital city of the Indian, and is a densely built-up megacity. It has experienced rapid growth over the past 20 years in terms of population and economy. The growth is mainly caused by the millions of migrants who moved here from other areas in India due to the business and work opportunities. In 1991, the census of India population showed that around 9.9 million people lived in Mumbai. Up to now, it is estimated that around 20 million people live in the metropolitan area. This tremendous increase has led to around 40% of residents (around 9 million) living in slums. Dharavi whose area is only 2.17 km^2, is the largest slum in Mumbai and the second largest slum in Asia. Approximately one million people live there (The Census Organization of India, 2011).

The physical characteristics of slums in Mumbai mainly include high densities, clustering of small buildings, and a rather organic morphology. Their roofing materials are mostly iron and asbestos sheets (Kuffer, Pfeffer, Sliuzas, & Baud, 2016). Figure 3.1 shows several slum clusters in Mumbai on PlanetScope imagery. The areas within red boundaries are slums.

Figure 3.1 Slums in Mumbai. The areas within red boundaries are slums (Hannes Taubenböck & Wurm, 2015)

3.1.2. Nairobi

Nairobi is Kenya's capital and has around 4.4 million people (Kenya National Bureau of Statistics, 2019). Around 60% of the population (2.5 million) has settled in over 100 slums and squatter settlements on only 6% of the land (United Nations Human Settlements Programme, 2003). The rural-urban migration in Kenya is the main reason for massive population growth in Nairobi (Dögg & Pétursdóttir, 2011).

The housing of slums is usually poor. Most of the people live on the muddy ground in shanties made of tin walls and tin roofs or other available materials. Most families live in a one-bedroom shack with no electricity and no access to clean water. Sewage runs above-ground between the houses since access to latrines is rare

(Dögg & Pétursdóttir, 2011). Figure 3.2 shows several slum clusters in Nairobi using PlanetScope imagery as background.

Figure 3.2 Slums in Nairobi. The areas within red boundaries are slums (Njoroge, 2016)

3.1.3. Rio de Janeiro

Rio de Janeiro, also called Rio, is the capital of the state of Rio de Janeiro, Brazil's third-most populous state. More than 1.5 million people live in more than 700 slums, which is around 20% of Rio's total population. 95% of the population living in slums are poor. UN-Habitat identified four different types of slums in Rio de Janeiro: Favelas, Loteamentons, Invasoes, and Cortiços (UN-Habitat, 2003). The first three types of slums lack basic infrastructure and services (Fricke, 2015). The last type could be regarded as social housing. Most of them were built illegally on hazardous spots without any formal urban planning and usually located on the eastern part of Rio (Fricke, 2015). Figure 3.3 shows several slum clusters in Rio on PlanetScope imagery.

Figure 3.3 Slums in Rio. The areas within red boundaries are slums (Data.rio, 2018)

3.2. Data

3.2.1. Satellite imagery data

This research uses PlanetScope imagery with a spatial resolution 3 m. There are mainly three reasons for selecting this imagery:

(1) Most FCNs studies for slum mapping either choose VHR satellite imagery, such as QuickBird imagery, or medium resolution satellite (MR) imagery, such as Sentinel-2 (Ajami, Kuffer, Persello, & Pfeffer, 2019; Verma et al., 2019; Wurm et al., 2019). However, higher spatial resolution of satellite imagery does not guarantee better classification results (Huang & Zhang, 2013). Because too much unnecessary information such as shadows may cause noise due to very high spatial resolution (Wang, Kuffer, & Pfeffer, 2019). The optimal characteristic scales for slums in Dar es Salaam, Bangalore and Pune are 3.39 m, 1.72 m and 4.29 m respectively (Wang et al., 2019). Characteristic scale is defined as "the scale at which the dominant pattern emerges" (Padt & Arts, 2014). The spatial resolution of PlanetScope imagery (3 m) is closer to optimal characteristic scales for slums in the three cities mentioned compared to that of VHR imagery or MR imagery. Though slums in the above three cities may have different morphological characteristics compared to the studies areas in this research, it is still worthy to try PlanetScope imagery. Because slums in the above three cities also have different characteristics but averagely the optimal characteristic scale is around 3 m. Up to now, no related studies have applied PlanetScope imagery for slum mapping by FCNs;

(2) Given that this research focuses on spatial transferability, to avoid the influences of sensors on prediction results, this research uses PlanetScope imagery for all experiments;

(3) PlanetScope imagery is open for researchers without costs, which makes it more accessible than commercial VHR imagery (Planet Labs, 2017). Besides, it may perform better than Sentinel-2 imagery which is open for users because it is closer to optimal characteristic scales and contain more detailed information of slums.

PlanetScope imagery has four bands, namely red, green, blue and near-infrared. It adopts Transverse Mercator projection. Table 3.1 shows the retrieved date of imagery for each city.

Table 3.1 Retrieved date of imagery from Planet Scope

City	Retrieved date
Mumbai	2019-10-11
Nairobi	2018-05-28, 2019-03-17
Rio	2016-08-29

For satellite imagery data, there are mainly four limitations which may bring uncertainties to prediction results. They are:

(1) though all of them are from PlanetScope, the viewing angles are different. This may cause some slums occluded by shadows from high buildings in different ways;

(2) the retrieved dates are different. This may cause different land cover situations, especially for vegetation. As mentioned in section 2.1, green space is considered as one of the common features of slums. The differences in vegetation caused by different retrieved dates may bring some errors;

(3) due to lack of data from the same retrieved date, satellite imagery for Nairobi were made up by imagery from two retrieved dates. Similar to (2), different spectral characteristics of slums and non-slums caused by imagery from different retrieved dates may bring some uncertainties to the results;

(4) the retrieved dates of satellite imagery are different from the retrieved year of reference data. Due to fast dynamics of slums, the realistic extent of slums in satellite imagery may be different from the slum extent in applied slum reference data;

3.2.2. Slum reference data

Slum reference data are from multiple sources. Table 3.2 shows the sources of reference data for each city and the year when they were generated. For Mumbai and Nairobi, the slum reference data are made by visual interpretation based on very high resolution imagery such as Google earth imagery (Njoroge, 2016). For Rio, the slum reference data are made by municipality in a "space first" way, based more on social and economic characteristics on the ground. At the same time, the Rio reference data include some non-slums due to data aggregation which makes slum follows administrative boundearies and merges areas where the majority of households are slums.

Table 3.2 Sources of slum reference data for Mumbai, Nairobi and Rio

City	Source	Retrieved year
Mumbai	Hannes Taubenböck & Wurm (2015)	2015
Nairobi	Njoroge (2016)	2015
Rio	Data.rio(2018)	2018

For reference data for Mumbai and Nairobi, there are mainly two limitations which may bring errors to the FCN models. They are:

(1) some locations in satellite imagery are hard to be classified as slums or non-slums from imagery (Njoroge, 2016);

(2) it is hard to determine boundaries of slums due to lack of an exact definition of slum boundaries. Besides, local people and slum experts may have different opinions towards slum delineation (Njoroge, 2016);

For reference data for Rio, slums were determined by ground information such as social and economic conditions. Besides, there are non-slums in reference data due to data aggregation. In addition to the nature of slums in Rio, consequently, slums in the reference data have very similar morphological characteristics as non-slums. RS-based slum mapping approaches are usually unable to recognize social and economic characteristics of slums from satellite imagery directly, which can make the performance of RS-based approaches low in detecting slums in Rio from satellite imagery.

4. METHODOLOGY

4.1. Assessment framework for the spatial transferability of FCNs for slum mapping

4.1.1. Definitions of spatial transferability of FCNs

Based on the existing studies, spatial transferability has mainly two meanings. First, it is defined as the capability of methods to produce comparable classification accuracies with minimum changes in parameter settings when applied to different geographic contexts (Pratomo et al., 2018). Second, it represents the capability of producing comparable accuracies when applying a model trained on data from a specific region referred to as source study area to another geographic area referred to as target study area (Agyemang-Duah & Hall, 1997).

Based on the above two definitions, there are three dimensions of spatial transferability of FCNs for slum mapping. Figure 4.1 shows the process of evaluating the second dimension of spatial transferability. Figure 4.2 shows the process of evaluating the third dimension of spatial transferability. For clarity, only two places were added in both figure. In application, more than two study areas can be assessed.

(1) only minimum changes are required to determine FCN architectures, hyperparameters and training strategies to reach optimal performance when training models on data from different study areas;
(2) the accuracy of a model trained on data from one source study area and tested on data from the corresponding study area is similar to that of a model trained on data from a different source study area and tested on data from the corresponding study area;
(3) when tested by the same pre-trained model, accuracies of the model tested on data from the same source study area and tested on data from different target study areas are similar;

To clarify the second and third dimensions, the following matrix was made. Acc_{i_j} represents the accuracy of a model trained on data from SA_i (SA: source study area) and tested by data from TA_j (TA: target study area). For the second dimension, the diagonal entries, $Acc_{1\_1}, Acc_{2\_2} \ldots Acc_{n\_n}$ (Acc: Accuracy) should be similar. For the third dimension, the entries of each row, $Acc_{i\_1}, Acc_{i\_2} \ldots Acc_{i\_n}$, should be similar. Standard deviation (SD) is frequently used to show the variance of datasets. Hence, the values of SD can be used for assessing the similarity of results. In addition to deviation, accuracy (the selection of accuracy indicators can be found in section 4.3.4) is also important to evaluate the performance of models. Hence, the mean values of the entries are also considered in the assessment. To summary, we can use SD and mean values of diagonal entries, SD_{Diag}, M_{Diag}, to reflect the performance of FCNs in spatial transferability for the second dimension; use SD and mean values of each row, $SD_1, SD_2 \ldots SD_n$, $M_1, M_2 \ldots M_n$ to represent the third dimension of spatial transferability.

$$\begin{bmatrix} Acc_{1\_1} & Acc_{1\_2} \ldots & Acc_{1\_n} \\ Acc_{2\_1} & Acc_{2\_2} \ldots & Acc_{2\_n} \\ \vdots & \vdots & \vdots \\ Acc_{n\_1} & Acc_{n\_2} \ldots & Acc_{n\_n} \end{bmatrix} \rightarrow \begin{matrix} SD_1 \\ SD_2 \\ \vdots \\ SD_n \\ SD_{Diag} \end{matrix} , \begin{matrix} M_1 \\ M_2 \\ \vdots \\ M_n \\ M_{Diag} \end{matrix}$$

In short, if an FCN model has high spatial transferability, then in terms of the second dimension, SD_{Diag} is low and M_{Diag} is high; in terms of the third dimension, SD_1, SD_2 ... are low and M_1, M_2 ... are high.

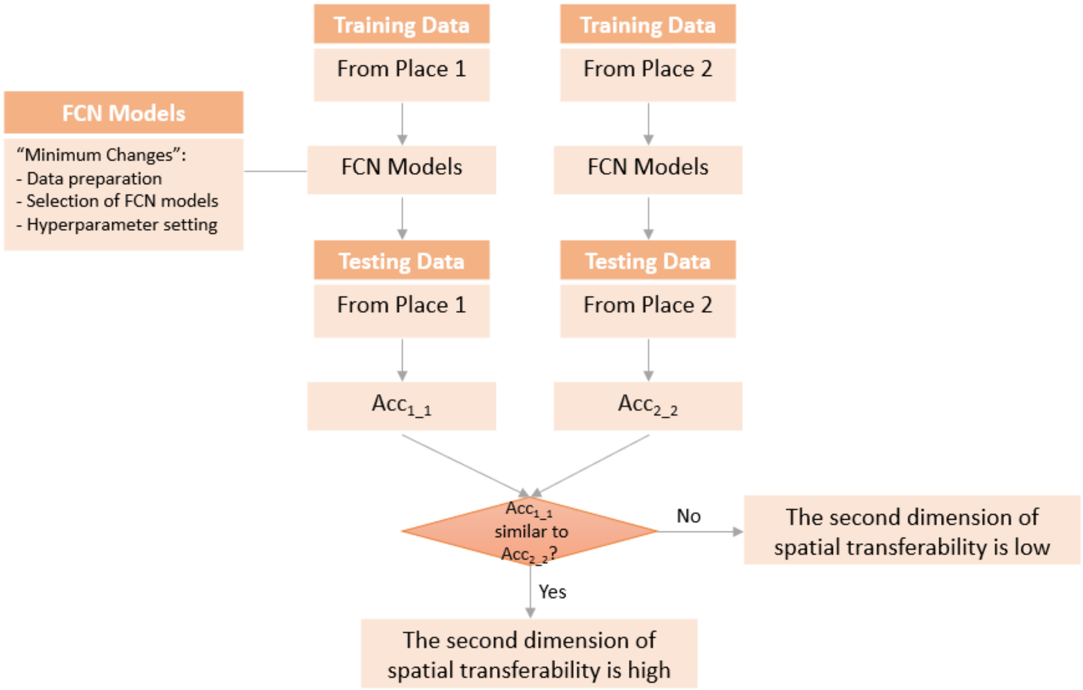


Figure 4.1 The process of evaluating the second dimension of the spatial transferability

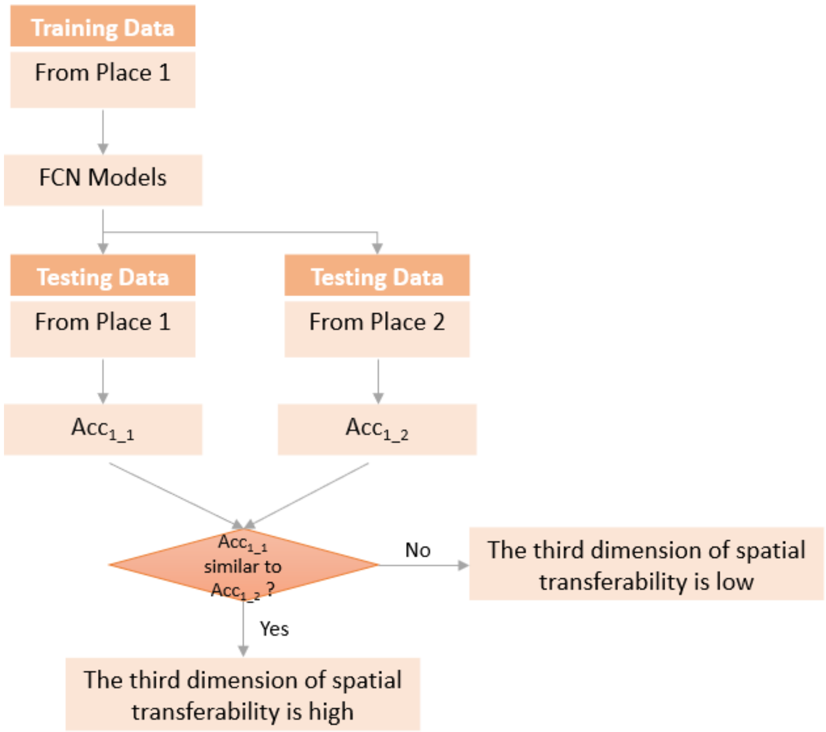


Figure 4.2 The process of evaluating the third dimension of the spatial transferability

4.1.2. Adaptations in training strategies for improving spatial transferability

For FCNs, there are potentially three adaptations in training strategies to improve the spatial transferability. Figure 4.3 shows how these adaptations work to improve the spatial transferability of FCNs. $Acc_{1\_ft\_2}$ represents the accuracy of the model trained on data from place 1 then fine-tuned by data from place 2 and finally tested on data from place 2. Acc_{12_1} represents the accuracy of the model trained data from place 1 and place 2 and tested on data from place 1. Other annotations can be understood in this way.

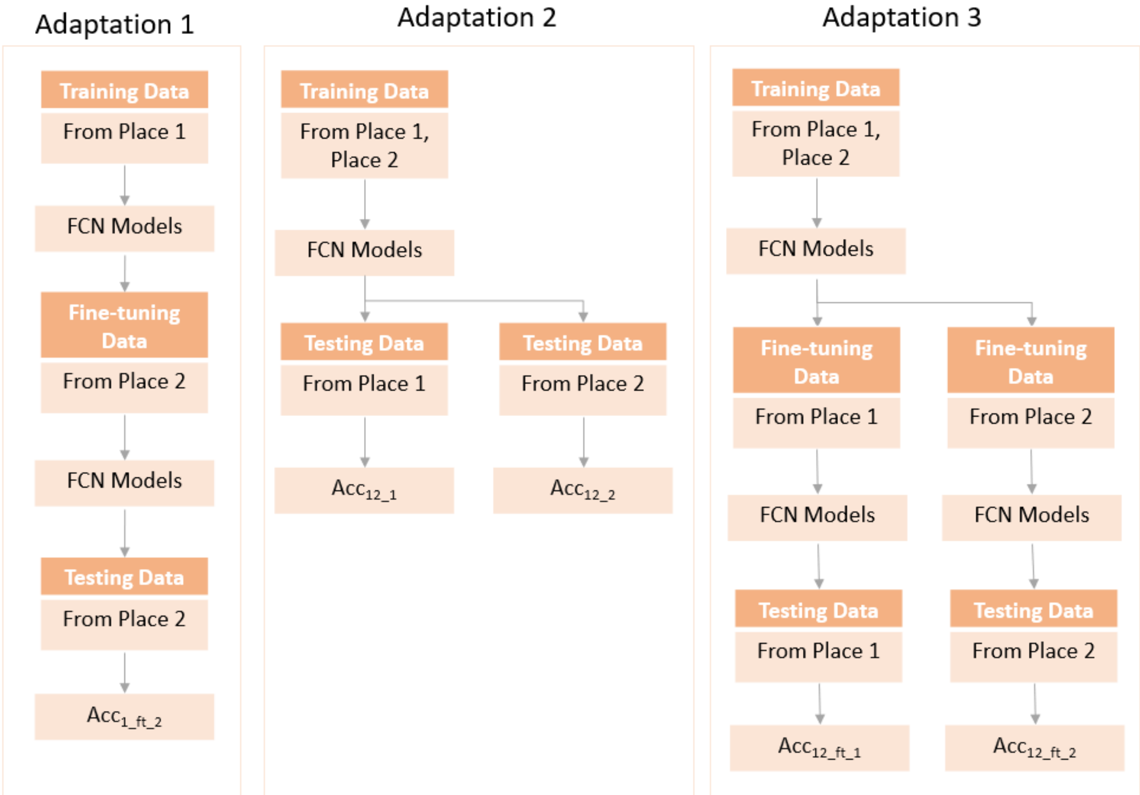


Figure 4.3 Three potential adaptations to improve the spatial transferability of FCNs

The first adaptation is applying fine-tuning the model trained on one source study area with data from a different study area before predicting slums in the different study area. It makes use of learnt features from other datasets with similar data distribution for the target task. If this adaptation works, SD of Acc_{1_1}, $Acc_{1\_ft\_2}$ and $Acc_{1\_ft\_3}$ will be lower than SD of Acc_{1_1}, Acc_{1_2} and Acc_{1_3}; Mean values of Acc_{1_1}, $Acc_{1\_ft\_2}$ and $Acc_{1\_ft\_3}$ will be higher than mean of Acc_{1_1}, Acc_{1_2} and Acc_{1_3}.

The second adaptation is combining training data from multiple source study areas. This strategy may help reduce differences among accuracies of the model tested on data from these study areas. If this adaptation works, for example, SD of Acc_{12_1} and Acc_{12_2} will be lower than SD of (Acc_{1_1}, Acc_{1_2}) and (Acc_{2_1}, Acc_{2_2}); Mean of Acc_{12_1} and Acc_{12_2} will be higher than mean of (Acc_{1_1}, Acc_{1_2}) and (Acc_{2_1}, Acc_{2_2})..

The third way is training FCN model by using combined training data from multiple source study areas, then fine-tuning the model with data from one of these source study areas or other target study areas and finally testing the fine-tuned model on the data for fine-tuning. The targeted area can be the same as or different from source study areas. If this adaptation works, for example, SD of $Acc_{12\_ft\_1}$, $Acc_{12\_ft\_2}$ and $Acc_{12\_ft\_3}$ will be lower than that of ($Acc_{1\_1}, Acc_{1\_2}, Acc_{1\_3}$) and ($Acc_{2\_1}, Acc_{2\_2}, Acc_{3\_3}$); Mean of $Acc_{12\_ft\_1}$, $Acc_{12\_ft\_2}$ and $Acc_{12\_ft\_3}$ will be higher than that of ($Acc_{1\_1}, Acc_{1\_2}, Acc_{1\_3}$) and ($Acc_{2\_1}, Acc_{2\_2}, Acc_{3\_3}$).

4.2. Experiments for assessing the spatial transferability

The above sections describe three dimensions of spatial transferability of FCNs for slum mapping and three adaptations in training strategies which may improve the spatial transferability. For time limitation of this research and due to complex process of selecting hyperparameter setting for FCN models, this research does not explore the influences of different selection of FCN architectures on the performance of the FCN models. Instead, this research focuses on the second and third dimensions of the spatial transferability for FCNs and how three adaptations can help improve the spatial transferability. Specifically, this research is designed to focus on the following parts:

(1) selecting a suitable FCN architecture and its hyperparameter setting for the architecture trained for different source study areas;
(2) exploring the second dimension of the spatial transferability;
(3) exploring the third dimension of the spatial transferability;
(4) exploring how three adaptations in training strategies influence the spatial transferability. of FCNs for slum mapping:
   1) Adaptation 1 – Fine-tuning the model pre-trained on one source study area with data from a different target study area;
   2) Adaptation 2 – Training model with datasets combined with multiple source study areas;
   3) Adaptation 3 – Fine-tuning the model pre-trained on datasets from multiple source study areas;

Table 4.1 Expected results of experiments (Acc: Accuracy; M: Mumbai; N: Nairobi; R: Rio; MN: Mumbai and Nairobi; MR: Mumbai and Rio; NR: Nairobi and Rio; M_M: the model trained on Mumbai dataset and tested on Mumbai dataset (similar meanings for other codes))

Experiment	Expected results
The second dimension	$(Acc_{M\_M}, Acc_{N\_N}, Acc_{R\_R})$
The third dimension	$(Acc_{M\_M}, Acc_{M\_N}, Acc_{M\_R})$, $(Acc_{N\_M}, Acc_{N\_N}, Acc_{N\_R})$, $(Acc_{R\_M}, Acc_{R\_N}, Acc_{R\_R})$
Adaptation 1	$(Acc_{M\_M}, Acc_{M\_ft\_N}, Acc_{M\_ft\_R})$
Adaptation 2	$(Acc_{MN\_M}, Acc_{MN\_N}, Acc_{MN\_R})$, $(Acc_{MR\_M}, Acc_{MR\_N}, Acc_{MR\_R})$, $(Acc_{NR\_M}, Acc_{NR\_N}, Acc_{NR\_R})$, $(Acc_{MNR\_M}, Acc_{MNR\_N}, Acc_{MNR\_R})$
Adaptation 3	$(Acc_{MNR\_ft\_M}, Acc_{MNR\_ft\_N}, Acc_{MNR\_ft\_R})$

4.3. Experiment setting-up

4.3.1. Data preparation

For each city, 5 tiles of 1000 x 1000 pixels covering an area on the ground of 3 km x 3 km have been selected for training tiles and 2 tiles with the same size have been chosen for testing tiles. The preparation of training data and testing data is based on the data preparation strategy in previous researches (Liu, Kuffer, & Persello, 2019; Persello & Stein, 2017) and a lot of trials. Figure 4.4, 4.5, 4.6 show slum maps and spatial distribution of the selected training and testing tiles for Mumbai, Nairobi and Rio respectively. Only two classes are available from the reference data of each city, namely "slums" and "non-slums". Slum reference data were presented as polygons in shapefile originally. They were converted to a raster format with the same spatial resolution as imagery data. The selection of the tiles was based on two rules:

(1) 20%-50% of tile area is slums based on slum reference data;
(2) The selected tiles distribute in different parts of each city;

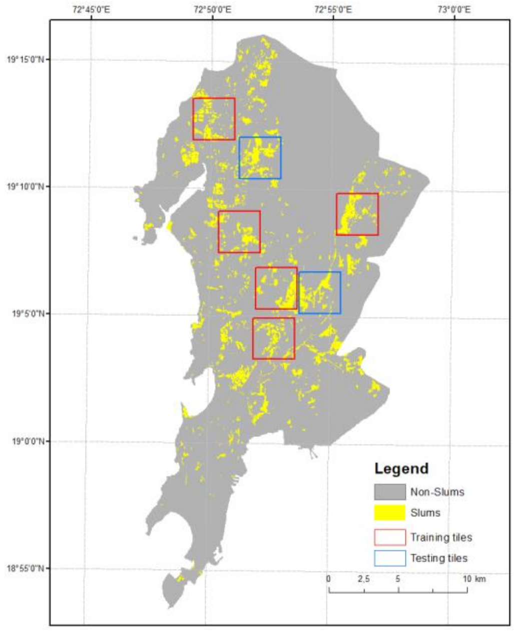


Figure 4.4 Slum map of Mumbai and spatial distribution of training and testing tiles for Mumbai

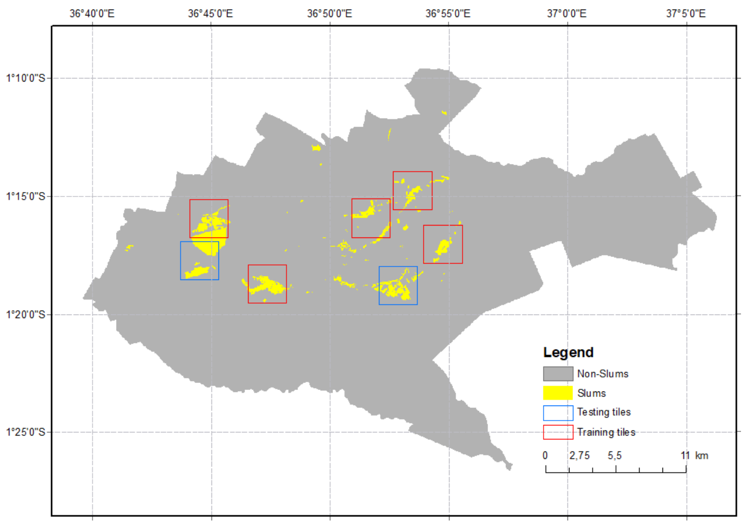


Figure 4.5 Slum map of Nairobi and spatial distribution of training and testing tiles for Nairobi

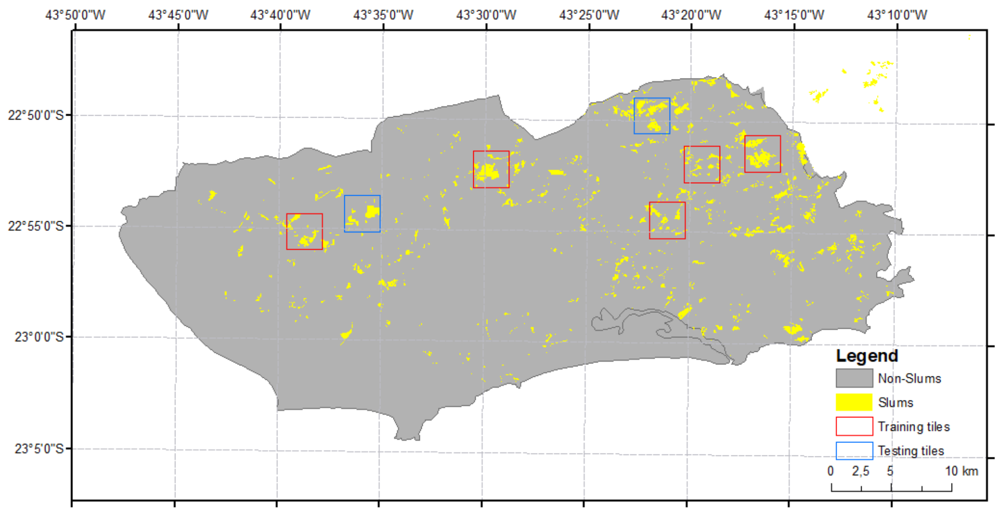


Figure 4.6 Slum map of Rio and spatial distribution of training and testing tiles for Rio

4.3.2. FCN architecture

Previous researches about the spatial transferability of FCNs selected FCN-VGG19 as the architecture for experiments (Stark et al., 2019; Wurm et al., 2019, 2019). One limitation of FCN-VGG19 is that only three bands of input imagery data can be used for training (Long et al., 2015).

Persello and Stein (2017) applied another FCN architecture for slum mapping, which makes use of dilated kernels (DKs) to keep the shape of prediction results the same as input data without deconvolution, and supports any number of bands of input imagery data for training. DKs are realized by inserting zeros between original elements of convolution filters. A dilation factor (d) means adding (d-1) zeros between each element of filters. After one operation of dilated convolution, the size of the kernel turns to $H' \times W' = [d \times (H-1)+1] \times [d \times (W-1)+1]$ (H and W represents height and width of the original kernel). By this way, the receptive field can be enlarged exponentially and do not require more learnable parameters. Figure 4.4 shows the receptive fields of dilated kernels when d=1 and d=2.

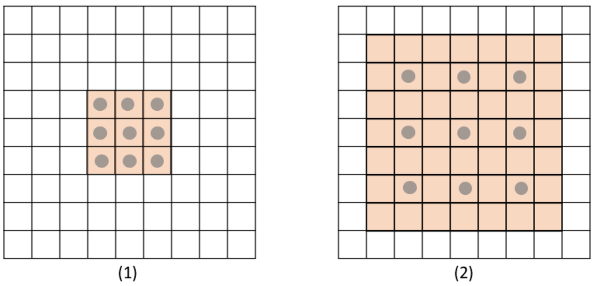


Figure 4.7 Receptive fields of a 3x3 kernel when d=1 and 2. Orange grids represent the receptive field of the kernel. Gray circles mean weights to be learnt. (1) d=1, receptive field=(3,3); (2) d=2, receptive field=(7,7).

The original architecture from Persello and Stein (2017) consists of 6 blocks. Hence, it is called as FCN-DK6. Each block consists of one zero padding layer, one convolutional layer, one batch normalization layer, one leaky Rectified Linear Units (lReLU) layer, another zero padding layer, and one max pooling layer. After 6 blocks, one dropout layer, one classification layer and one loss function are added at the end of the networks. Zero paddings are used for keeping the output feature maps with the same shape as the input imagery data. Hence, FCN-DKs can classify images at arbitrarily sizes and produce outputs with the corresponsive size (Persello & Stein, 2017). Batch normalization is used for normalizing each input minibatch to avoid internal covariate shift issue during training the model. Internal covariate shift issue usually occurs when the inputs to each layer are calculated cumulatively by all preceding layers. Hence, the distribution of each layer's input will be changed when learnable parameters from previous layers change (Ioffe & Szegedy, 2015). Leaky ReLU is the activation function for the architecture, which can help determine whether one pixel belongs to slums or non-slums (Xia, 2019).

Liu et al. (2019) modified the architecture by replacing one convolution layer with 5 x 5 kernel size in each block with two convolution layers with 3 x 3 kernel size. In this way, the receptive field of convolution layers keeps the same and the learnable parameters are reduced. Besides, it turns out that the performance of the modified FCN-DK6 is better than the original architecture for slum mapping. Therefore, the FCN architecture proposed for this research is adapted from the modified FCN-DK6 from Liu et al. (2019). Instead of using 6 blocks in the architecture, the proposed FCN-DK architecture consists 5 blocks, which is called as FCN-DK5. Because by trial and error, FCN-DK5 shows highest stability in terms of hyperparameter setting and training data from different study areas among FCN-DK1, FCN-DK2, FCN-DK3, FCN-DK4, FCN-DK5 and FCN-DK6 for experiments in this research. Figure 4.8 shows the architecture of the proposed FCN-DK5 for this research. Table 4.2 presents the structure of the FCN-DK5.

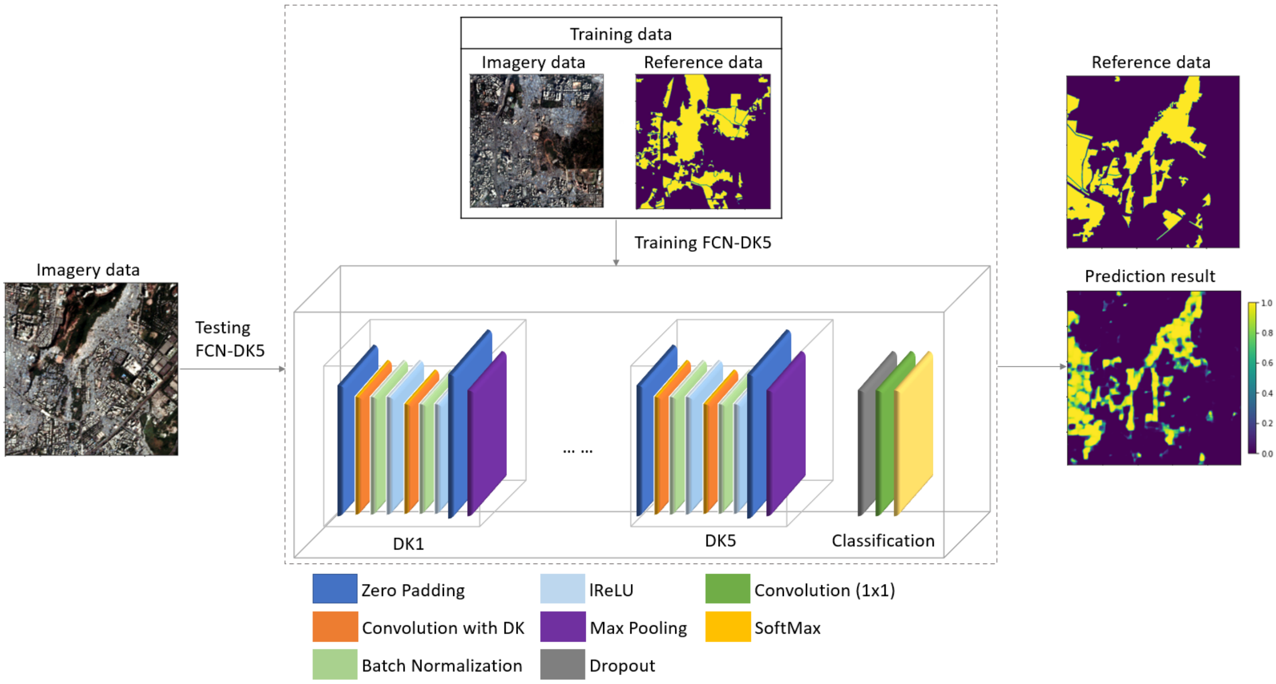


Figure 4.8 Architecture of the proposed FCN-DK5 for this research

Table 4.2 The structure of the proposed FCN-DK5 for this research

Layer	Module type	Parameters
DK1	ZeroPadding	(2, 2)
	Convolution	kernel size=3x3; number of filters=16; dilation rate=(1,1)
	BatchNormalization	---
	lReLU	0.1
	Convolution	kernel size=3x3; number of filters=16; dilation rate=(1,1)
	BatchNormalization	---
	lReLU	0.1
	ZeroPadding	(2, 2)

	MaxPooling	pool size=(5,5); strides=1
DK2	ZeroPadding	(4, 4)
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(2,2)
	BatchNormalization	---
	lReLU	0.1
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(2,2)
	BatchNormalization	---
	lReLU	0.1
	ZeroPadding	(4, 4)
	MaxPooling	pool size=(9,9); strides=1
DK3	ZeroPadding	(6, 6)
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(3,3)
	BatchNormalization	---
	lReLU	0.1
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(3,3)
	BatchNormalization	---
	lReLU	0.1
	ZeroPadding	(6, 6)
	MaxPooling	pool size=(13,13); strides=1
DK4	ZeroPadding	(8, 8)
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(4,4)
	BatchNormalization	---
	lReLU	0.1
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(4,4)
	BatchNormalization	---
	lReLU	0.1
	ZeroPadding	(8, 8)
	MaxPooling	pool size=(17,17); strides=1
DK5	ZeroPadding	(10, 10)
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(5,5)
	BatchNormalization	---
	lReLU	0.1
	Convolution	kernel size=3x3; number of filters=32; dilation rate=(5,5)
	BatchNormalization	---
	lReLU	0.1
	ZeroPadding	(10, 10)
	MaxPooling	pool size=(21,21); strides=1
Classification	Dropout	0.2
	Convolution	kernel size=(1, 1); number of filters=2(number of classes)
	SoftMax	---

4.3.3. Training the networks

The training patches for the networks are overlapped by 32 pixels for data augmentation. For each training tile, 784 patches have been generated. Table 4.3 shows the numbers of training patches of different datasets. The shape of patches for this research is 128 x 128 pixels. The training patches for each dataset will be shuffled randomly before training the model.

Table 4.3 Number of training patches for each dataset

Datasets	Number of training patches
Mumbai (M)	3920
Nairobi (N)	3920
Rio (R)	3920
Mumbai and Nairobi (MN)	7840
Mumbai and Rio (MR)	7840
Nairobi and Rio (NR)	7840
Mumbai, Nairobi and Rio (MNR)	11760

The networks are trained by stochastic gradient descent (SDG) with a momentum of 0.9. Minibatches of 32 samples were chosen for training. Batch normalization was added after every convolution layer as shown in table 4.2. A dropout rate of 0.2 was applied before the classification layer. The FCN-DK5 was trained on MNR dataset by initializing weights randomly. The model was trained for 100 epochs at a learning rate of 10^{-2} with a learning rate decay of 10^{-4}. The trained weights ($W_{MNR\_Initial}$) are used as initialized weights for experiments for the second and third dimension of spatial transferability of FCNs and for the adaptation 2. Table 4.4 presents tasks for each experiment, initialized weights, parameters and expected output weights for each task.

Table 4.4 Initialized weights, learning rate, decay, epochs, expected output weights of tasks for each experiment

Experiment	Tasks	Initialized weights	Parameters	Output weights
The second dimension	(1) Train the model on M dataset; (2) Train the model on N dataset; (3) Train the model on R dataset;	$W_{MNR\_Initial}$	learning rate=10^{-2} decay=10^{-4} epochs=100	(1)W_M (2)W_N (3)W_R
The third dimension	Use the trained models from the experiments for the second dimension			
Adaptation 1	(1) Fine-tuning the pre-trained model on M dataset with N dataset; (2) Fine-tuning the pre-trained model on M dataset with R dataset;	W_M	learning rate=10^{-4} epochs=50	(1)$W_{M\_ft\_N}$ (2)$W_{M\ ft\_R}$
Adaptation 2	(1) Train the model on MN dataset; (2) Train the model on MR dataset; (3) Train the model on NR dataset;	$W_{MNR\_Initial}$	learning rate=10^{-2} decay=10^{-4} epochs=100	(1)W_{MN} (2)W_{MR} (3)W_{NR}
	(4) Train the model on MNR dataset;	$W_{MNR\_Initial}$	learning rate=10^{-4} epochs=50	(4) W_{MNR}

Adaptation 3	(1) Fine-tuning the pre-trained model on MN dataset with M dataset; (2) Fine-tuning the pre-trained model on MN dataset with N dataset; (3) Fine-tuning the pre-trained model on MN dataset with R dataset;	W_{MN}	learning rate=10^{-4} epochs=50	(1)$W_{MN\_ft\_M}$ (2)$W_{MN\_ft\_N}$ (3)$W_{MN\_ft\_R}$
	(4) Fine-tuning the pre-trained model on MNR dataset with M dataset; (5) Fine-tuning the pre-trained model on MNR dataset with N dataset; (6) Fine-tuning the pre-trained model on MNR dataset with R dataset;	W_{MNR}	learning rate=10^{-4} epochs=20	(1)$W_{MNR\_ft\_M}$ (2)$W_{MNR\_ft\_N}$ (3)$W_{MNR\_ft\_R}$

4.3.4. Accuracy assessment

Applying commonly accepted accuracy measurements for describing the quality of prediction results is essential for comparing the performance of different experiments and other researches. In this research, accuracies of the models are evaluated by prediction accuracies on testing tiles.

As mentioned in the section 2.5, researches related to slum mapping have adopted many accuracy indicators such as OA (overall accuracy), precision, recall, IoU, F1 score and Kappa estimate. A recent research proves that Kappa estimate is not suitable for image classification tasks (Hossin & Sulaiman, 2015). Similarly, OA can mislead the prediction results, especially when applied to imbalanced images (Hossin & Sulaiman, 2015). High OA does not guarantee good performance of models for predicting minority class such as slums. For these reasons, the evaluation metrics in this research resorted to precision, recall, F1 score and IoU.

Precision (P) is the ratio of correctly predicted slum pixels to the total predicted slum pixels. Recall (R) is the ratio of correctly predicted non-slum pixels to the total predicted non-slum pixels (Hossin & Sulaiman, 2015). F1 score is defined as the harmonic average of Precision and Recall and is considered more useful than OA.(Hossin & Sulaiman, 2015). IoU is the ratio of correctly predicted slum pixels to the total predicted slum pixels and wrongly predicted slum pixels. It is the standard metric to assess the PASCAL VOC challenges, which are famous computer vision contests (Liu, Salberg, & Jenssen, 2018). Table 4.5 presents the confusion matrix for binary classification. Formula 4.1, 4.2, 4.3 and 4.4 show the calculation of four selected measurements.

Table 4.5 Confusion matrix for binary classification

	Positive Prediction	**Negative Prediction**
Positive Class	True Positive (TP)	False Negative (FN)
Negative Class	False Positive (FP)	True Negative (TN)

$$P = \frac{TP}{TP + FP} \ (4.1)$$

$$R = \frac{TP}{TP + FN} \ (4.2)$$

$$F1\ score = \frac{P \times R}{P + R} \times 2 \ (4.3)$$

$$IoU = \frac{TP}{TP + FP + FN} \ (4.4)$$

4.3.5. Software and Platform

ENVI 5.5 was used for mosaicking PlanetScope imagery for each city. ArcGIS 10.7 was used for generating reference data from polygon to raster with the same spatial resolution as imagery data. Python is used for implementing FCNs and ploting prediction results. The implementation of the model is based on the TensorFlow framework and Keras libraries.

The proposed FCN model is trained on Google CoLab, which provides a single 12GB NVIDIA Tesla K80 GPU. The GPU is free for users and sufficient for this research. Besides, it can connect to Google Drive which can provide large storage for satellite imagery. The scripts and data stored on Google CoLab and Google Drive can be conveyed to other researchers more conveniently.

5. RESULTS AND DISCUSSION

This chapter analyses the results of the conducted experiments and assesses the spatial transferability of FCNs for slum mapping. For clarity, not all results but only key results are displayed in the chapter. Complete results can be found in table A.1, table A.2 and table A.3 in appendix, which include precision, recall, F1 score and IoU of two testing tiles (TS1 and TS2) and their average accuracies for all experiments for Mumbai, Nairobi and Rio.

5.1. Assessment of the second dimension of the spatial transferability of FCNs

The second dimension of the spatial transferability of FCNs for slum mapping is measured by the differences in the accuracies of a model trained on data from one source study area and tested on data from the same study area, and a model trained on data from a different source study area and tested on data from the corresponding study area. Figure 5.1 shows the experiments implemented for the second dimension. Acc_{M_M} means the accuracy of the model trained on Mumbai data and tested on Mumbai data. Acc_{N_N} and Acc_{R_R} can be understood in the way as the meaning of Acc_{M_M}.

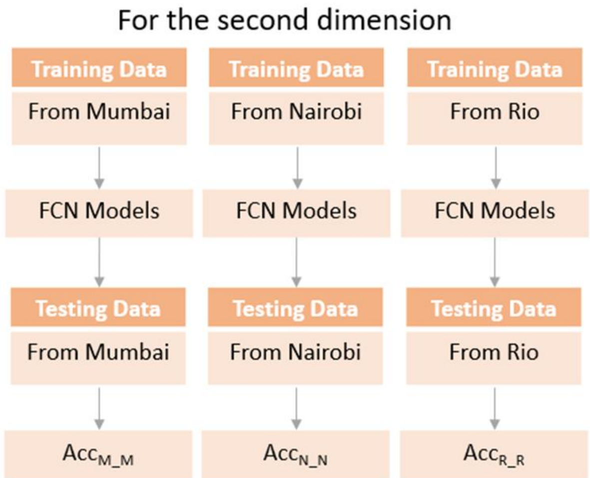


Figure 5.1 The process of implementing experiments for the second dimension of the spatial transferability of FCN models in this research. "Acc" represents accuracy. M_M means the model trained on Mumbai data and tested on Mumbai data. N_N and R_R have similar meanings.

The precision, recall, F1 score, and IoU of M_M, N_N and R_R are shown in table 5.1. F1 score and IoU indicators evaluate the performance of models in an integrated way. Based on these results, it can be concluded that the FCN model M_M performs best (IoU=65.09%), whereas the FCN model R_R performs worst (IoU=31.42%). The standard deviation (SD_{D2}) and the mean value (M_{D2}) of the second dimension of spatial transferability are 0.171 and 46.63% (based on IoU). Figure 5.5, 5.6, 5.7 show the prediction maps of two testing tiles (TS1 (left) and TS2 (right)) for Mumbai, Nairobi and Rio.

Table 5.1 Average precision, recall, F1 score and IoU of M_M, N_N and R_R

Experiment	Precision	Recall	F1 score	IoU
M_M	84.77%	73.95%	79.00%	65.09%
N_N	66.88%	55.62%	60.73%	43.39%
R_R	51.92%	44.67%	48.02%	31.42%

In these experiments, the recall refers to the ratio of pixels truly predicted as slums to all pixels predicted as slumsand precision refers to the ratio of pixels truly predicted as slums to all pixels labelled as slums. Low recall and precision values indicate that the trained FCN model is unable to distinguish slums and non-slums from satellite imagery. Figure 5.2 explains the inference of precision and recall values for slum mapping. In short, low recall indicates that the similarity between slums and non-slums in the study areas is large and thus many non-slum pixels have been predicted as slums. On the other hand, low precision indicates that the diversity within slums in this same study area is large and thus slums in testing tiles cannot be predicted as slums. Besides, these two aspects can influence =each other. For example, if the similarity between slums and non-slums is too large, then the learnt features might be too weak to detect slums from satellite imagery and, thus, both precision and recall would be very low.

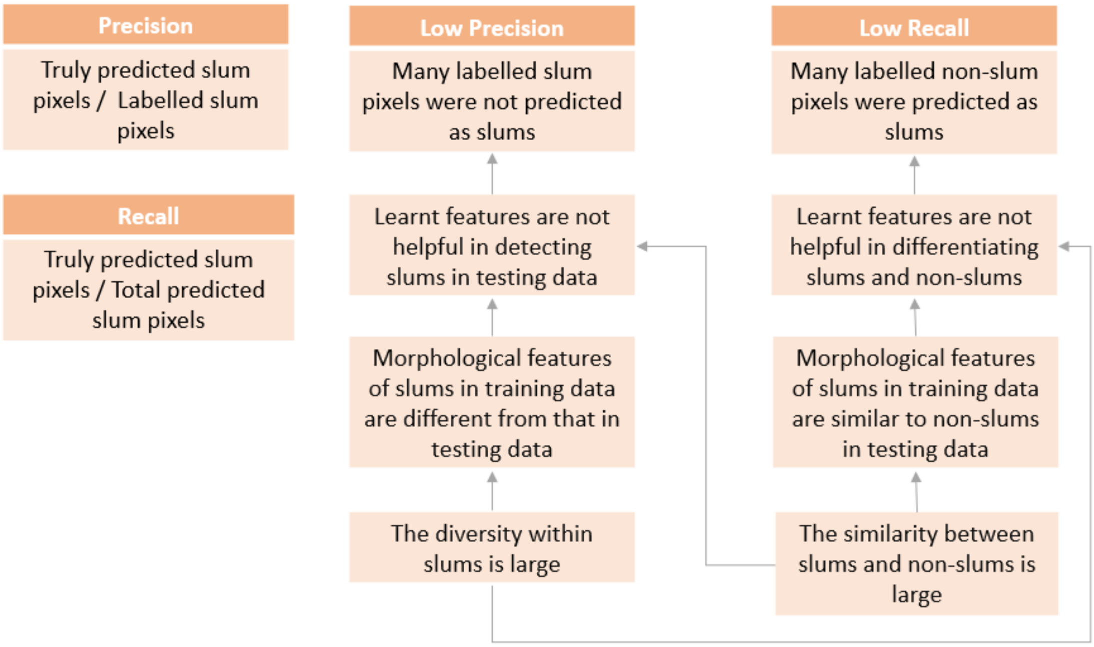
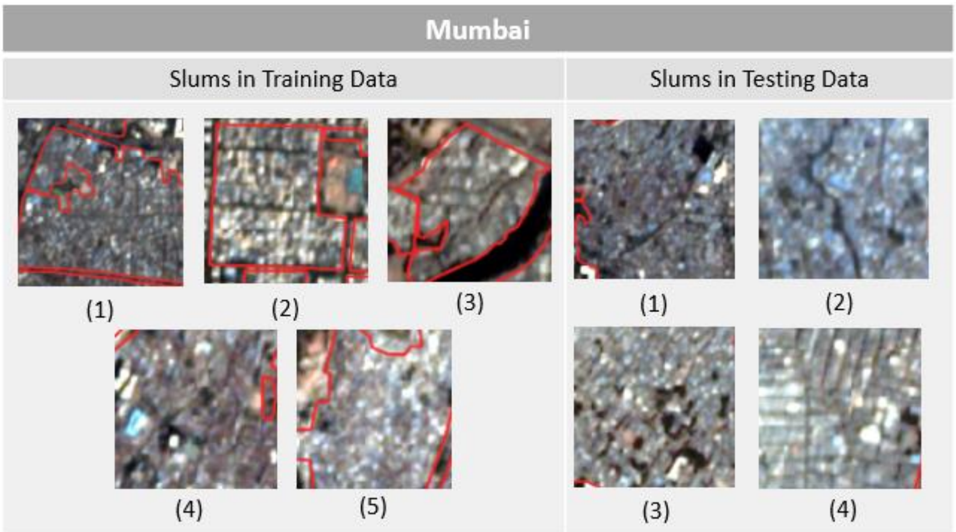


Figure 5.2 Description of the precision and recall metrics for assessing slum mapping results

In these experiments, the highest recall was produced for M_M (84.77%) and R_R (51.92%) =. Satellite imagery together with slum reference data of the investigated three cities (figure 5.3), revealed that the discrimination between slums and non-slums in terms of morphological characteristics in Mumbai is much more obvious than that in Nairobi or Rio.

Figure 5.3 Appearances of slums and surrounding non-slums in Mumbai (up), Nairobi (middle) and Rio (down). Slums are delineated within red boundaries.

Low precision values indicate that the diversity of slums is large. Hence, the trained model is unable to detect slums from testing tiles based on the learnt features from training tiles. Figure 5.4 shows typical morphological characteristics of slums in training and testing tiles of Mumbai, Nairobi and Rio. It can be found that slums in Mumbai and Nairobi are more diverse than those in Rio. However, due to large similarity in terms of morphological characteristics between slums and non-slums in Rio, both of the precision and recall values of R_R are the lowest. Besides, it can be found that the precision values of all three experiments are higher than recall values. This may indicate that the similarity between slums and non-slums brings more errors to the model compared to the diversity within slums in the same city.



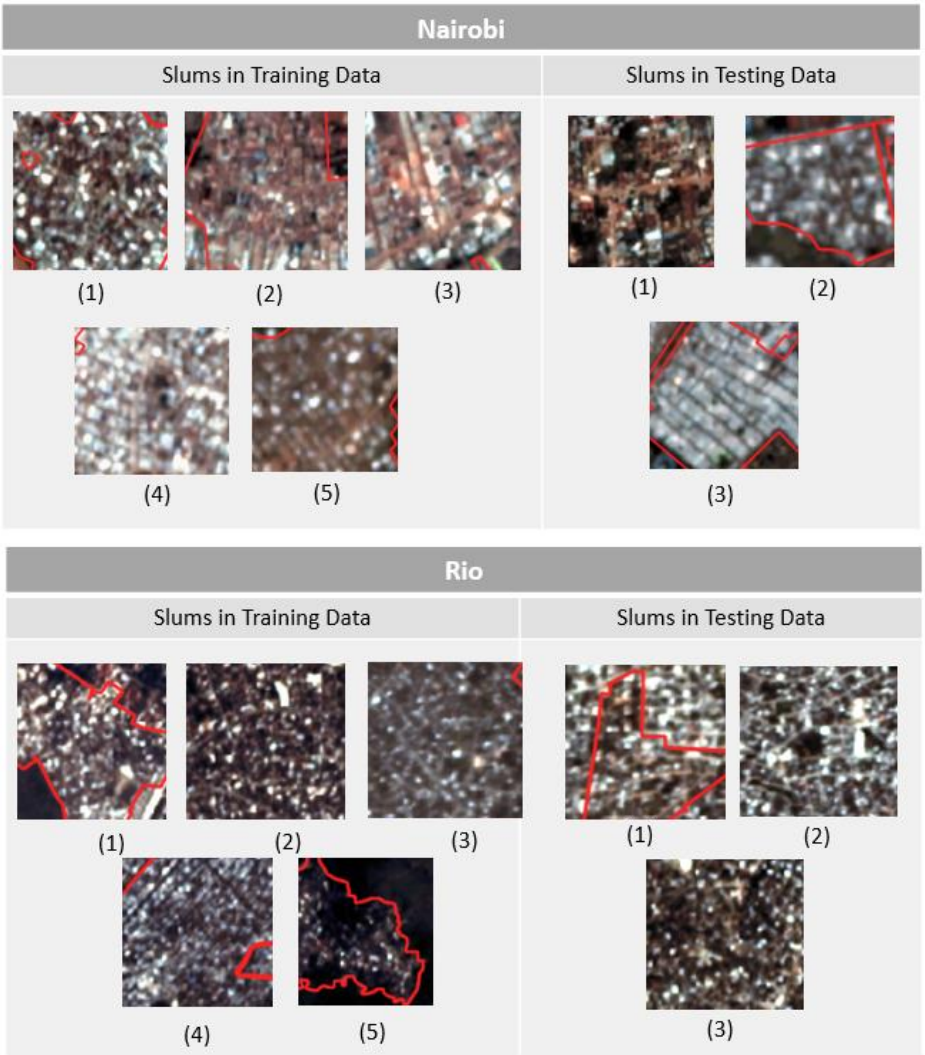


Figure 5.4 Typical morphological features of slums in training and testing data of Mumbai, Nairobi, Rio

The lowest precision and recall indicate that slums in Rio are hard to be detected. The low performances of R_R is mainly caused by two reasons. The first reason is related to the nature of slum situation in Rio. Multiple types of slums exist in Rio and some of them are similar to non-slums in terms of morphological characteristics as mentioned in section 3.1.3. Favela and cortiço are two main types of slums in Rio. Favelas are usually agglomerated by dense dwellings and built with informal materials such as cardboard, tin and old wood. They are distributed irregularly and often do not have adequate basic services and infrastructure. A cortiço consists of several buildings built on an urban lot and it is always subdivided into rent units with multiple functions in the same room. The physical appearances of cortiço are similar to formal settlements in Rio, which means it is hard to differentiate them from other formal settlements from satellite imagery. The ratio of favela dwellers to cortiço dwellers is round 3:1. (Ferreira, Machado, Franco, & de Mello Franco, 2020) The second reason is related to the collection of slum reference data. As mentioned in section 3.2.2, slums in the reference data for Rio were determined by ground information such as social and economic conditions rather than only morphological characteristics. Besides, there are non-slums in reference data due to data aggregation according to the administrative boundaries. In consequence, the mixture of favela, cortiço and other types of slums and collection approaches of slum reference data in Rio result in large similarity between cortiço and non-slums and, consequently, make prediction accuracies of R_R much lower compared to M_M and N_N.

As for Mumbai, Taubenböck & Kraff (2014) compared the differences of building density, size and height across slum areas at block level. They found out there is no considerable difference in building density of slums. The differentiation between slums and non-slums in building size is clear. Compared to non-slum areas, slum areas are more homogenous. Building height across slum areas is also homogenous, which is typically identified as one- or two-floor height. However, slums in Mumbai are also heterogeneous at different levels. They developed the "heterogeneity index"(HI) integrated from building density, size and height to evaluate the heterogeneity of slums. According to the reported results, HI of Dharavi is higher than that of Santosh Nagar and Bharat Nagar. This research helps explain why the model M_M performs best among three cities but does not achieve 100% accuracy. For these reasons, many related studies used Mumbai as a benchmark for slum mapping (Ibrahim, Titheridge, Cheng, & Haworth, 2019; Kuffer, Pfeffer, Sliuzas, et al., 2016; Kuffer et al., 2018; Stark, 2018; Stark et al., 2019; Verma et al., 2019; Wurm et al., 2019).

In terms of the second dimension, spatial transferability of FCNs for slum mapping is low given the large differences in accuracies of M_M, N_N and R_R. This finding is also supported by the research of Stark (2018) where IoU of FCN-VGG19 trained and tested on Mumbai data (QuickBird imagery) is 66.12% and IoU for Delhi data is 48.85%. The author adopted a different FCN architecture and different RS satellite imagery but obtained similar findings. These findings indicate that FCNs achieves different accuracy for slum mapping when applied in different study areas. This happens because of the different levels of heterogeneity of slums and similarity between slums and non-slums in these three cities. M_M performs best because the diversity within slums in Mumbai is low and the similarity between slums and non-slums is also low. Besides, slum reference Reference data collection approaches may also explain the low spatial transferability of FCNs. Reference data for Mumbai and Nairobi are generated through visual interpretation of slums from satellite images , whereas those for Rio are ground-based collected data . This means that slums in Mumbai and Nairobi are easier to be detected from satellite images than that in Rio. In addition to data aggregation, some non-slums were included in the reference data for Rio. For these reasons, the performance of R_R is the worst among the three designed and implemented experiments.

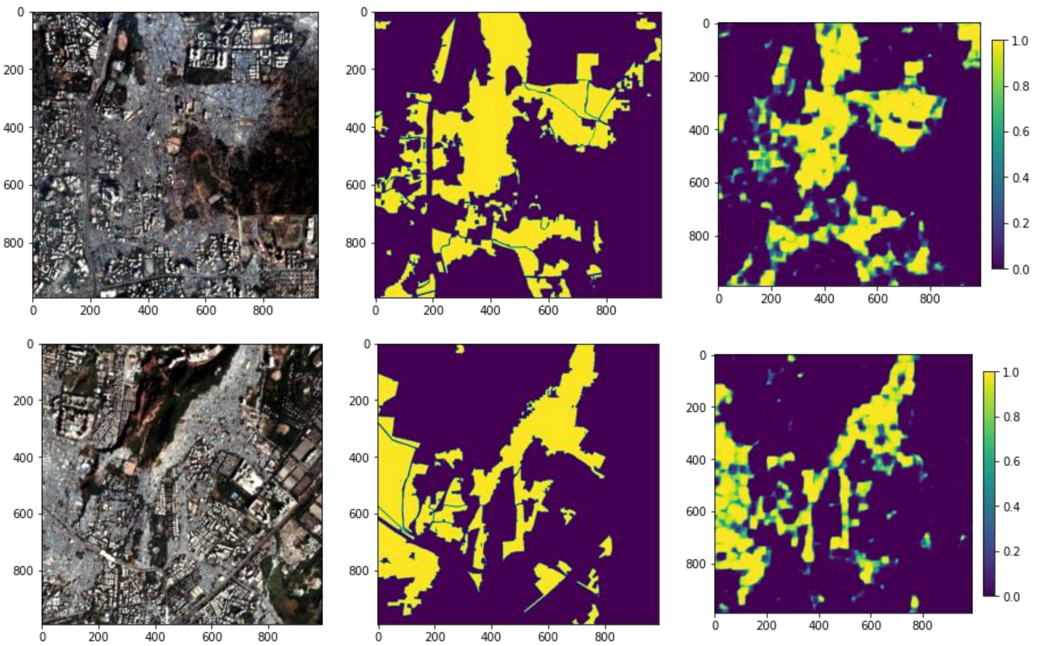


Figure 5.5 PlanetScope imagery (left), slum reference maps (middle), probability maps (right) of TS1 (up, IoU=61.63%) and TS2 (down, IoU=69.65%) of Mumbai (Experiment: M_M)

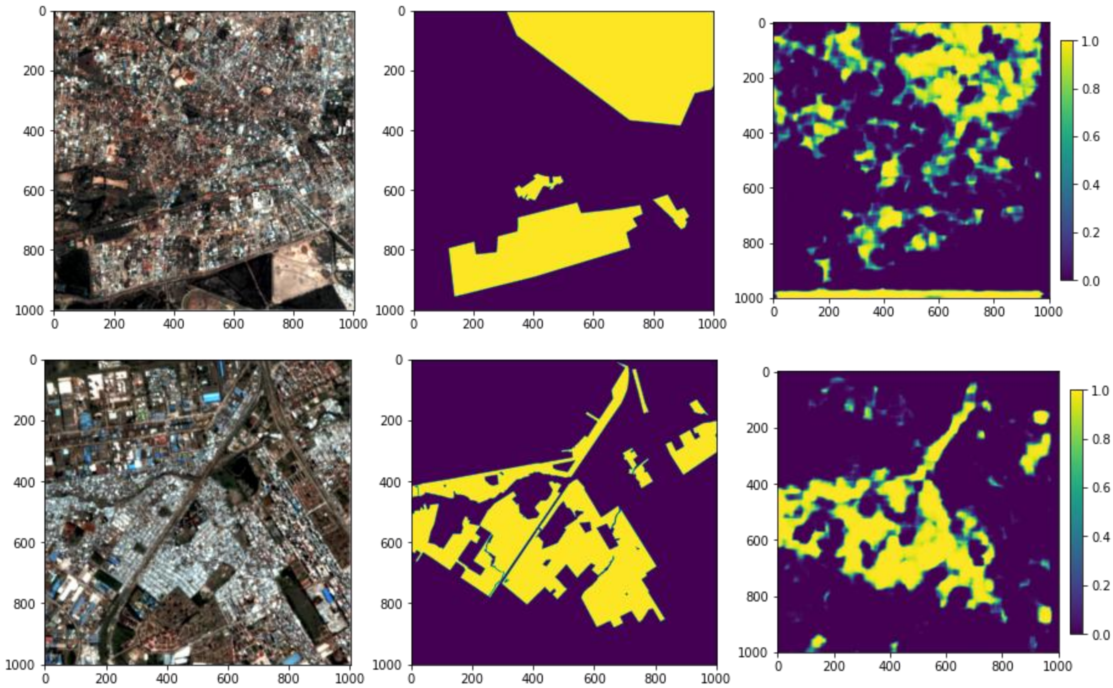


Figure 5.6 PlanetScope imagery (left), slum reference maps (middle), probability maps (right) of TS1 (up, IoU=33.92%) and TS2 (down, IoU=56.86%) of Nairobi (Experiment: N_N)

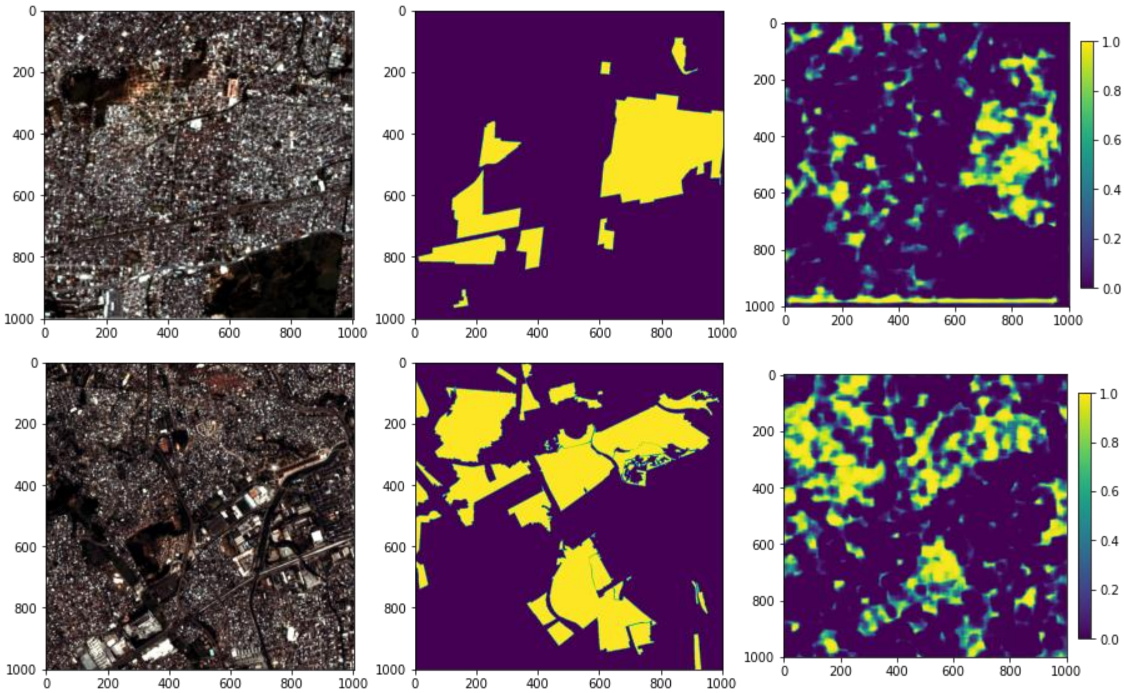


Figure 5.7 PlanetScope imagery (left), slum reference map (middle), classification maps (right) of TS1 (up, IoU=27.97%) and TS2 (down, IoU=33.79%) of Rio (Experiment: R_R)

5.2. Assessment of the third dimension of the spatial transferability of FCNs

The third dimension of the spatial transferability of FCNs can be assessed by the accuracies of the same pre-trained model tested on data from the same source study area and tested on data from different target study areas. Figure 5.8 shows the process of implementing the experiments for the third dimension of the spatial transferability of FCNs for slum mapping in this research.

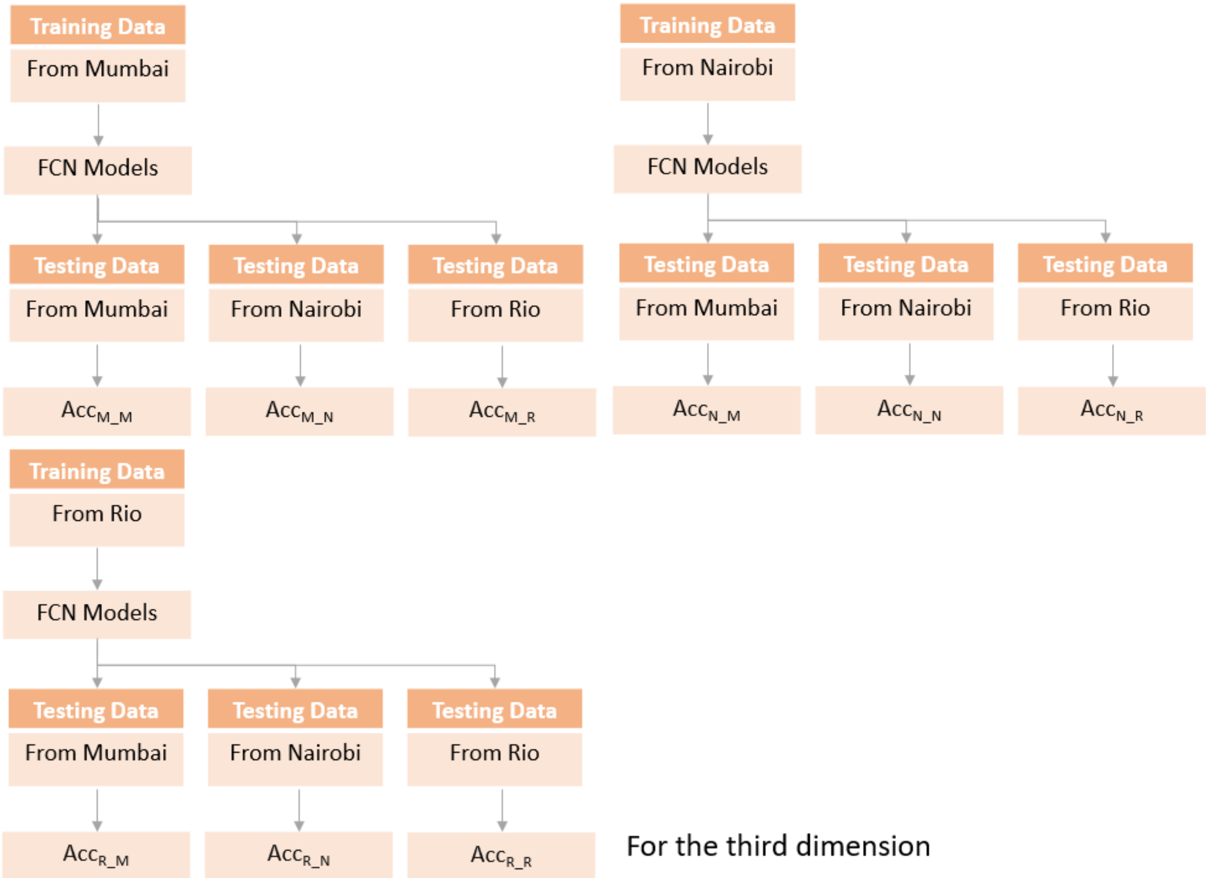


Figure 5.8 The process of implementing experiments for the second dimension of the spatial transferability of FCN models in this research

IoU, F1 score together with SD and mean values obtained for the experiments dedicated to assessing the third dimension of the spatial transferability are shown in table 5.2 and table 5.3. The results revealed that the accuracy of the model trained on one source study area and tested on a different target study area is much lower than tested on the corresponding source study area. For instance, the IoU of M_M is 65.09%, but the IoU of M_N and M_R is only 0.34% and 10.58% respectively. Similar observations can be found from experiments related to Nairobi and Rio. The SD values of (Acc_{M_M}, Acc_{M_N}, Acc_{M_R}) and (Acc_{N_M}, Acc_{N_N}, Acc_{N_R}) are higher than those of SD_{D2} (0.171). This means that pre-trained model] applied directly on data from different target study areas produces less comparable accuracies. The SD of (Acc_{R_M}, Acc_{R_N}, Acc_{R_R}) is lower than SD_{D2} but its mean IoU (14.77%) is much lower than that of M_{D2} (46.63%).

As seen in figure 5.4, the different morphological characteristics of slums and non-slums in the three investigated cities reflected on satellite imagery mainly caused the low spatial transferability of FCNs in terms of the third dimension. These differences made learnt features from data in one source study area unable to directly detect slums in other target areas with different slum characteristics.

In summary, in terms of the third dimension, spatial transferability of FCNs for slum mapping is low when a model trained on a source study area is tested on data from other different target study areas with different morphological characteristics of slums reflected on satellite imagery.

Table 5.2 IoU, SD and mean of the experiments for the third dimension

Training data	Testing data			SD	Mean
	M	N	R		
M	65.09%	0.34%	10.58%	0.348	25.34%
N	1.31%	43.39%	0.62%	0.245	15.10%
R	0.03%	12.87%	31.42%	0.158	14.77%

Table 5.3 F1 score, SD and mean of the experiments for the third dimension

Training data	Testing data			SD	Mean
	M	N	R		
M	79.00%	0.68%	19.16%	0.409	32.94%
N	2.60%	60.73%	1.23%	0.34	21.52%
R	0.07%	22.88%	48.02%	0.24	23.66%

5.3. Assessment of the influences of three adaptations on the spatial transferability of FCNs

In this section, the results of the proposed three adaptations will be analysed to see how they can influence the spatial transferability of FCNs for slum mapping. These adaptations are considered to improve the spatial transferability if they decrease SD values and increase mean values of accuracy indicators produced by the trained model compared to the results of the experiments for assessing the second dimension (M_M, N_N, R_R) and the third dimension of the spatial transferability (M_M, M_N, M_R). This process can be seen in figure 5.9. In this case, IoU is adopted as the indicator to measure SD and mean values.

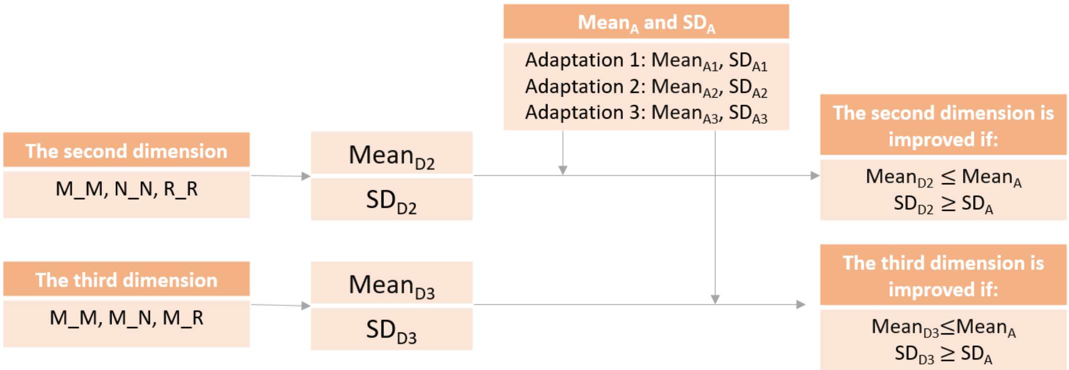


Figure 5.9 The process of judging whether the proposed adaptations improve the second and third dimension of the spatial transferability of FCNs. A means Adaptation. D means Dimensions.

Figure 5.10 depicts the process of implementing the experiments for the three adaptations for assessing the spatial transferability of FCNs for slum mapping. These three adaptations are:

1) Adaptation 1 – Fine-tuning the model pre-trained on data from a source study area with data from a different target study area before predicting slums in this different target study area;
2) Adaptation 2 – Training the model with datasets combined with multiple source study areas;
3) Adaptation 3 – Fine-tuning the model pre-trained on datasets from multiple source study areas with data from a different target study area before predicting slums in different target study area;

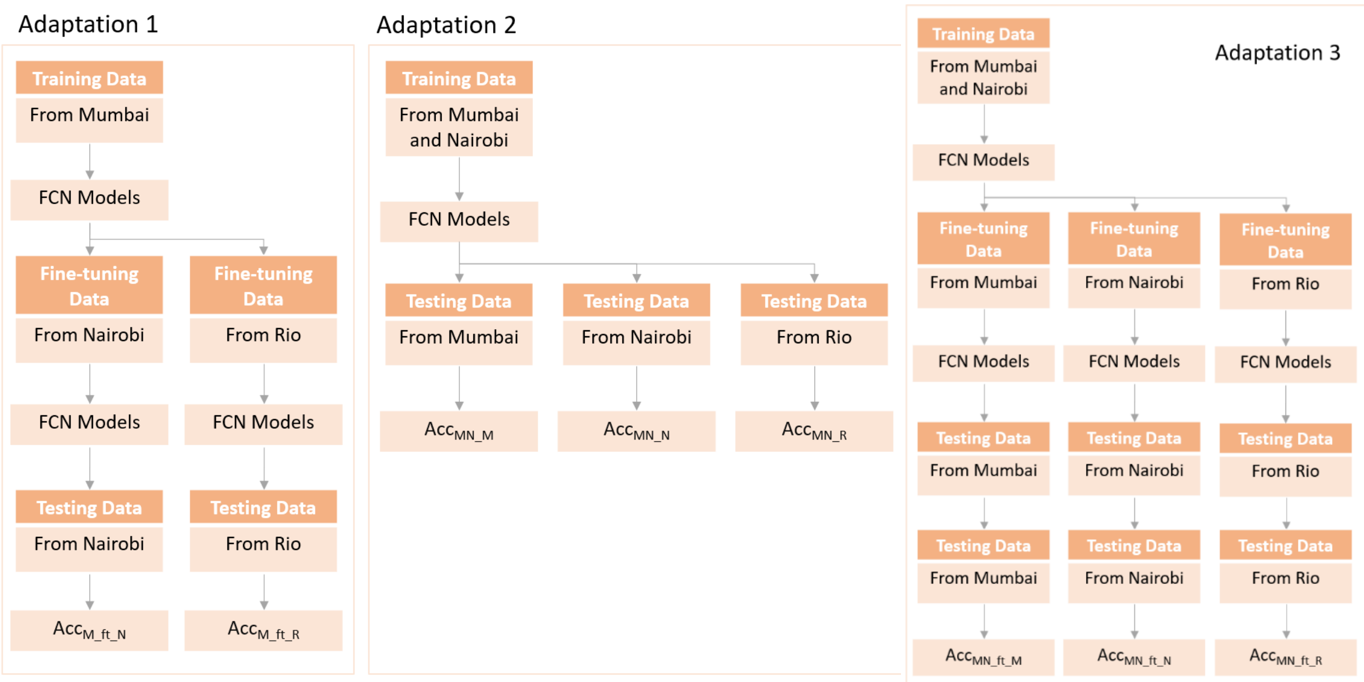


Figure 5.10 The process of implementing experiments for the three adaptations for assessing the spatial transferability of FCNs for slum mapping

5.3.1. Adaptation 1 – Fine tuning

Table 5.4 shows IoU and F1 score of M_M, M_ft_N and M_ft_R and comparison with that of the experiments for the second and the third dimension of the spatial transferability. It can be observed that IoU of M_ft_N (43.61%) is similar to that of N_N (43.39%) and much higher than that of M_N (0.34%). IoU of M_ft_R (29.76%) is a bit lower than that of R_R (31.42%) but much higher than that of M_ft_R (10.58%). Besides, SD_{A1} (SD of the IoU of experiments for adaptation 1) (0.178) is also closer to SD_{D2} (0.171) and is much lower than SD_{D3} (0.348). Figure 5.10 and 5.11 show probability maps and IoU of testing tiles for Nairobi and Rio under the experiment M_ft_N and M_ft_R respectively.

The results indicate that fine-tuning can help effectively transfer learnt features of slums in one source study area to a task for predicting slums in a different target study area. Compared with the results from M_M, M_N and M_R, adaptation 1 improves the third dimension of the spatial transferability of FCNs for slum mapping. Compared with the results from M_M, N_N and R_R, adaptation 1 does not improve the second dimension. This happens because the SD_{A1} is lower than SD_{D3} and higher than SD_{D2}; $Mean_{A1}$ is higher than $Mean_{D3}$ but lower than $Mean_{D2}$. This means the results after fine-tuning are less comparable and lower in prediction accuracies compared to the model trained and tested on data from the same study area. However, they are more comparable and have better prediction accuracies compared to the model trained on data from one source study but tested on data from a different target study area. In other words, adaptation 1 makes prediction accuracies close to accuracies of M_M, N_N and R_R with lower computational cost and help to improve the third dimension of the spatial transferability but cannot improve the second dimension.

Table 5.4 F1 score, IoU of experiments for adaptation 1 and comparison with that of other experiments

Experiment	F1	IoU	Experiment	F1	IoU	Experiment	F1	IoU
M_M	79.00%	65.09%	M_M	79.00%	65.09%	M_M	79.00%	65.09%
M_ft_N	60.96%	43.61%	N_N	60.73%	43.39%	M_N	0.68%	0.34%
M_ft_R	46.11%	29.76%	R_R	48.02%	31.42%	M_R	19.16%	10.58%
SD_{A1}	0.165	0.178	SD_{D2}	0.156	0.171	SD_{D3}	0.409	0.348
$Mean_{A1}$	62.02%	46.16%	$Mean_{D2}$	62.58%	46.63%	$Mean_{D3}$	32.94%	25.34%

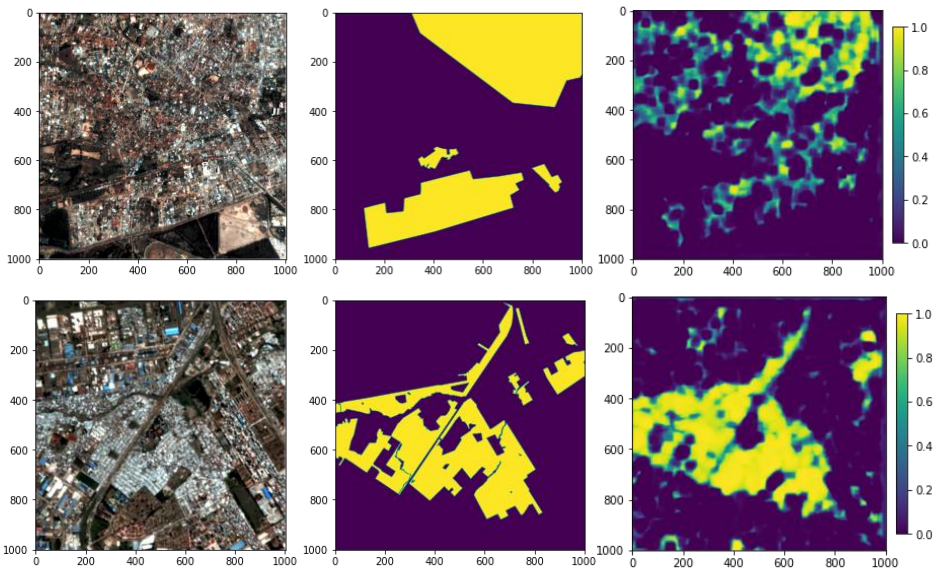


Figure 5.11 PlanetScope imagery (left), slum reference map (middle), classification maps (right) of TS1 (up, IoU=29.43%) and TS2 (down, IoU=60.05%) of Nairobi (Experiment: M_ft_N)

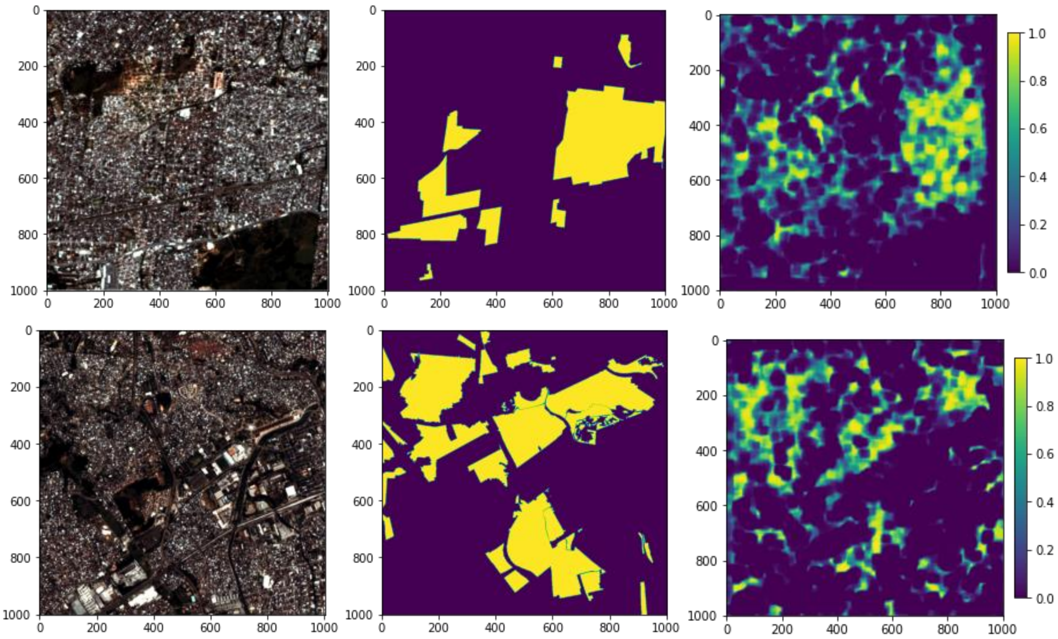


Figure 5.12 PlanetScope imagery (left), slum reference map (middle), classification maps (right) of TS1 (up, IoU=30.90.97%) and TS2 (down, IoU=28.86.79%) of Rio (Experiment: M_ft_R)

5.3.2. Adaptation 2 – Training datasets from multiple study areas

The IoU results of the experiments using combined datasets from multiple source study areas are shown in table 5.5. The SD and mean values of the experiments and other experiments are shown in table 5.6. Figure 5.12, 5.13 and 5.14 display probability maps and IoU of two testing tiles (TS1(left), TS2(right))of Mumbai under the experiment MN_M, MR_M and MNR_M.

The IoU of MN_M (IoU=66.38%) is higher than that of M_M (IoU=65.09%). The IoU of MN_N (IoU=44.89%) is higher than that of N_N (IoU=43.39%). The combination of datasets from Mumbai and Nairobi can help improve the performance of the model when tested on data from these two cities. The IoU of MR_M (62.58%) drops compared to that of M_M. While the IoU of MR_R (IoU=31.66%) slightly increases compared to that of R_R (IoU=31.42%). Combining datasets from Mumbai and Rio can slightly improve the prediction accuracy of testing data from Rio while reduces the accuracy for Mumbai.

As mentioned in section 5.1, due to ground-based data collection approach together with data aggragation and the complex nature of slums in Rio, the similarity between slums and non-slums in terms of morphological characteristics is quite large reflected on satellite imagery, which results in low IoU of R_R. Slums in Mumbai are more distinctive from non-slums in terms of morphological characteristics. Besides, slum reference data in Mumbai is mainly based on physical appearances reflected on VHR satellite imagery. Therefore, for Mumbai, combining datasets from Rio may increase confusion to the classification rules of the model. However, when adding datasets from Mumbai, the performance of the Rio model was improved. Reference data of slums in Nairobi also includes some slums with similar appearances as non-slum areas. Therefore, adding datasets from Mumbai decreases confusion in training data and increase the amount of training data for the model. The improvement in accuracies for Mumbai may be caused by more training data for the model. When combining dataset from Rio, the performance of related models decreases. This might happen because the training datasets from Rio confuses the model.

The IoU of MN_R (0.53%), MR_N (0.16%) in table 5.5 revealed that the prediction accuracy is low when the model is trained on data from one or more source study areas and tested on data from a different target study area directly. The high IoU of NR_M (28.61%) maybe because slums in Mumbai are relatively easily recognized.

The results turn out that training the FCN model with data from multiple source study areas can improve the second and the third dimensions of the spatial transferability of FCNs. This finding can be proved by the lower SD and higher mean values of MN_M and MN_N than that of M_M and N_N and that of M_M and M_N; the lower SD and higher mean values of MNR_M and MNR_R than that of M_M and R_R and that of M_M and M_R. This finding indicates combining slum reference data from multiple study areas can help increase the third and second dimensions of the spatial transferability of FCN models. By this way, it only requires changing testing data when applied to different study areas and thus also meets the requirement of the first dimension of the spatial transferability.

Table 5.5 IoU, F1 score of experiments using training data from multiple source study areas

Training data	Testing data					
	IoU			F1 score		
	M	N	R	M	N	R
MN	66.38%	44.89%	0.53%	77.68%	51.64%	0.53%
MR	62.58%	0.16%	31.66%	77.58%	0.00%	31.66%
NR	28.61%	42.95%	18.96%	59.69%	69.61%	18.96%
MNR	66.27%	43.75%	33.45%	72.68%	73.33%	48.22%

Table 5.6 Standard deviation (SD) and mean values of IoU under the experiments (D/A: Dimension or Adaptation)

D/A	Experiments	SD	Mean	Experiments	SD	Mean
D2	M_M, N_N	0.153	54.24%	M_M, R_R	0.238	48.26%
D3	M_M, M_N	0.458	32.72%	M_M, M_R	0.385	37.84%
A2	MN_M, MN_N	0.152	55.64%	MR_M, MR_R	0.219	47.12%
A2	MNR_M, MNR_N	0.159	55.01%	MNR_M, MNR_R	0.232	49.86%

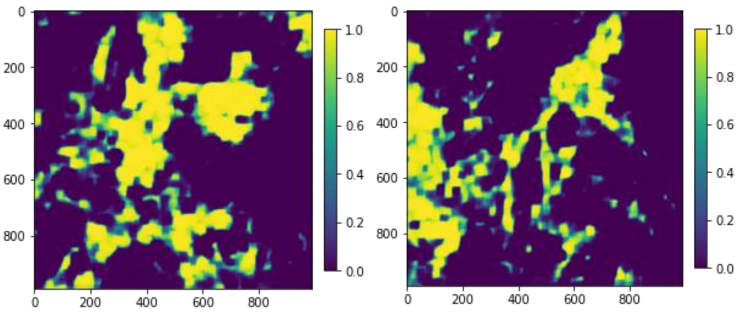


Figure 5.13 Probability maps of TS1 (IoU=63.35%) and TS2 (IoU=70.28%) for Mumbai (Experiment: MN_M)

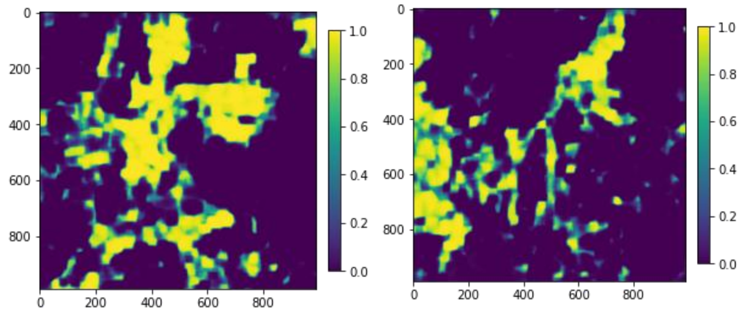


Figure 5.14 Probability maps of TS1 (IoU=63.19%) (left) and TS2 (IoU=61.79%) (right) for Mumbai (Experiment: MR_M)

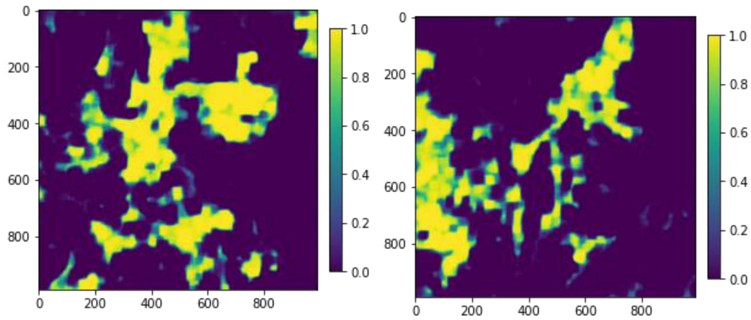


Figure 5.15 Probability maps of TS1 (IoU=60.51%) (left) and TS2 (IoU=69.58%) (right) for Mumbai (Experiment: MNR_M)

5.3.3. Adaptation 3 – Fine tuning the model trained on datasets from multiple study areas

Table 5.7 shows the results of experiments for adaption 3 and comparison with other experimentsThese results showed that the model trained on the dataset MNR performs best among the selected four types of experiments. While the model pre-trained on the dataset MNR and fine-tuned on the dataset of one target study area performs worst. The results of M_ft_N and M_ft_R are slightly better than the results of MNR_ft_N and MNR_ft_R in both SD and mean values. The slight decrease in performance shows that the effects of fine-tuning a model pre-trained on multiple source study areas cannot help to improve the second dimension of the spatial transferability.

As mentioned in section 5.3.2, adding slum reference data of Rio may reduce the performance of the FCN model. Hence, IoU of all of the experiments tested on Rio data is much lower than that tested on Mumbai or Nairobi data. Given the fact that slums in Mumbai and Nairobi are more distinctive from non-slums in terms of morphological characteristics reflected on satellite imagery, adding reference data from Mumbai and Nairobi to Rio data can help the model recognize slums in Rio more correctly. Thus, IoU of MNR_R is even higher than that of R_R. After fine-tuning IoU of MNR_ft_R became lower than that of MNR_R. This happens possibly because more effective learnt features from MNR dataset gradually shifted to less effective features from Rio data. Similar observations can be found on MNR_ft_M and MNR_ft_N. This output may indicate that learnt features from MNR dataset can perform better than learnt features from data of one individual city. In this case, fine tuning does not improve the performance of the model. This finding indicates that increasing slum reference data from multiple study areas to train FCN models may be helpful to predict slums better.

In terms of spatial transferability, similar to the above two adaptations, the results revealed that adaptation 3 can make prediction accuracies close to accuracies of M_M, N_N and R_R and thus improve the third dimension of the spatial transferability ($SD_{A3} < SD_{D3}$; $Mean_{A3} > Mean_{D3}$). However, it performs worse than adaptation 1 and adaptation 2 and thus does not improve the second dimension of the spatial transferability

($SD_{D2} < SD_{A3}$; $Mean_{D2} > Mean_{A3}$). Similar to the other two adaptations, it cannot make prediction accuracies for different cities more comparable. Therefore, to make prediction accuracy closer to M_M, N_N and R_R, adaptation 1 and adaptation 2 are preferred over adaptation 3, which does not improve accuracies but is more time and computational consuming. Figure 5.11, 5.12, 5.13 presents the prediction maps of two testing tiles of Mumbai, Nairobi and Rio under the experiments for the adaptation 3.

Table 5.7 IoU, SD and mean of MNR_ft_M, MNR_ft_N and MNR_ft_R and comparison with other experiments

Experiment	**IoU**	**Experiment**	**IoU**	**Experiment**	**IoU**	**Experiment**	**IoU**
M_M	65.09%	MNR_M	66.27%	M_M	65.09%	MNR_ft_M	65.04%
N_N	43.39%	MNR_N	43.75%	M_ft_N	43.61%	MNR_ft_N	43.16%
R_R	31.42%	MNR_R	33.45%	M_ft_R	29.76%	MNR_ft_R	29.16%
SD_{D2}	0.171	SD_{A2}	0.168	SD_{A1}	0.178	SD_{A3}	0.181
$Mean_{D2}$	46.63%	$Mean_{A2}$	47.82%	$Mean_{A1}$	46.16%	$Mean_{A3}$	45.79%

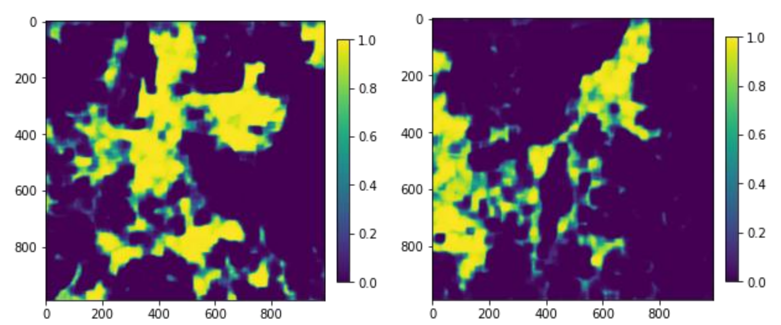


Figure 5.16 Probability maps of testing tiles in Mumbai. TS1 (left, IoU=62.28%) and TS2 (right, IoU=66.82%) under (Experiment: MNR_ft_M)

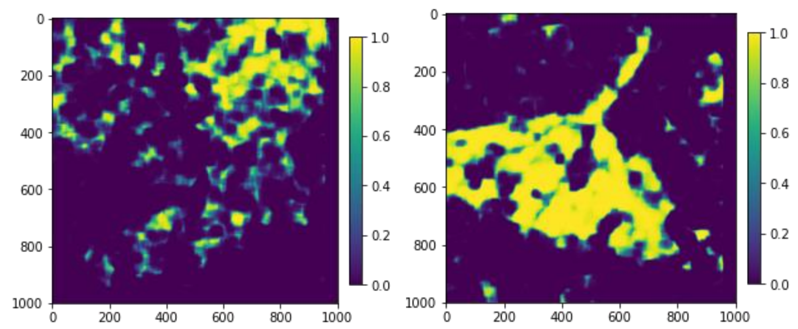


Figure 5.17 Probability maps of testing tiles in Mumbai. TS1 (left, IoU=62.28%) and TS2 (right, IoU=66.82%) under (Experiment: MNR_ft_N)

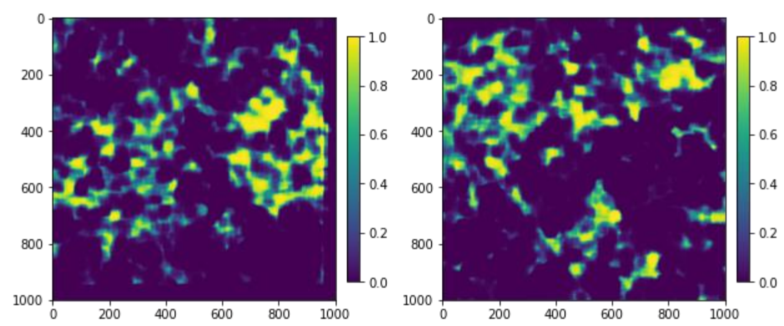


Figure 5.18 Probability maps of TS1 (IoU=22.63%) (left) and TS2 (IoU=29.45%) (right) for Rio (Experiment: MNR_ft_R)

5.4. Performance of PlanetScope imagery for slum mapping

In this research, the FCN-DK5 model trained on PlanetScope imagery data of Mumbai (spatial resolution=3m) reaches an IoU of 65.09% (experiment: M_M). Wurm et al. (2019) applied FCN-VGG19 for slum mapping in Mumbai. In their study, the IoU of the FCN-VGG19 model trained on QuickBird imagery (spatial resolution=0.5m) reaches an IoU of 77.02%. The same model trained on Sentinel-2 data (spatial resolution=10m) reaches an IoU of 35.51%. Regardless of the influences of the selection of FCN models, hyperparameter setting, training and testing data on prediction results, it can be found that PlanetScope imagery performs much better than Sentinel-2 imagery but still worse than QuickBird imagery. Therefore, PlanetScope as open-source data for researchers may be a better substitute for VHR satellite imagery compared to Sentinel-2 if researches or projects have a limited financial budget.

Verma, Jana and Ramamritham (2019) applied a classic patch-based CNN model for slum mapping in Mumbai. The CNN model trained on VHR Pleiades-1A imagery (spatial resolution=0.5m) reaches an IoU of 58.3% and the same model trained on Sentinel-2 data reaches an IoU of 33.2%. From this output, it can be found that FCN models performed better than patch-based CNN models when tested on data from Mumbai. Even the spatial resolution of PlanetScope imagery is lower than Pleiades-1A imagery, the IoU is still higher than that of patch-based CNN model trained on Pleiades-1A imagery. However, the differences in the yielded IoU values can also be caused by a different selection of training strategies and testing data.

5.5. Limitations of this reseach

There are two primary limitations in the results of all the experiments in this research. The first is from the selection of FCN model and hyperparameter setting. The second is from the selected PlanetScope imagery and slum reference data.

FCN models are sensitive to the hyperparameter setting. The process of selecting proper hyperparameter setting for the selected FCN architectures is complicated (Choi, Cho, & Rhee, 2019). The hyperparameters adopted in this research were chosen based on previous research dedicated to using FCN-DKs for slum mapping (Liu et al., 2019; Persello & Stein, 2017) and a lot of trials including comparing the performance of the FCN-DK5 under different learning rate, training epochs and convolutional kernel size. However, it is still possible that the optimal hyperparameter setting for one study area is different from that for another study area. This is the content of the first dimension of the spatial transferability of FCNs for slum mapping, which aims to find out the optimal FCN architecture and hyperparameter setting for slum mapping= in different study areas. Due to time limitations, this research did not fully explore this dimension.

The reference data of Mumbai and Nairobi used in these experiments are generated by visual interpretation (Njoroge, 2016; Taubenböck & Wurm, 2015). It is possible that local stakeholders have different perceptions upon slum boundaries (Pratomo et al., 2016). The reference data of Rio was downloaded from related websites, which was collected by municipality. It is worthy to exlore the influences of different data collection approaches on the performance of FCN models. Besides, the retrieved dates of the reference data are older than PlanetScope imagery. Due to the fast dynamics of slums, it is possible that the status of some pixels have been changed. For these two reasons, ground truth data may be not totally true.

As mentioned in section 3.2.3, though all of satellite imagery are from PlanetScope, the viewing angles may be different. This may cause some slums occluded by shadows from high buildings. Due to lack of data from the same retrieved date, satellite imagery for Nairobi were made up by imagery from two retrieved dates. Different spectral characteristics of slums and non-slums caused by imagery from different retrieved dates may bring some uncertainties to the results.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

This research assessed the spatial transferability of FCNs for slum mapping by using PlanetScope imagery. Spatial transferability was defined as the ability of a mapping approach to perform similarly with minimum changes, when applied to different geographic contexts.

This research is the first attempt dedicated to developing a systematic framework for assessing the spatial transferability of FCNs for slum mapping. The framework contains three dimensions of the spatial transferability for FCN models and three adaptations that have the potential to improve the spatial transferability. The specific content of the dimensions and adaptations can be found in section 4.1 and 4.2 respectively. This framework can be used as a guideline to assess the different aspects of the spatial transferability of FCN models for slum mapping and help evaluate the effects of the adaptations on it. For time limitation, this research only assessed the second and third dimensions of the spatial transferability.

The results revealed that the second dimension of the spatial transferability of FCNs for slum mapping is bad. This was reflected by the results that the FCN model cannot produce comparable accuracies when applied to Mumbai, Nairobi and Rio. This is emphasized by the big differences in IoU (or F1 score) of M_M (65.09%), N_N (43.39%) and R_R (31.42%) as shown in table 5.1.

There are mainly two reasons accounting for the low performance of the second dimension of the spatial transferability. The first reason lies in the different nature of slums in the investigated study areas located in different continents. From the morphological characteristics of slums reflected from satellite imagery, the diversity within slums in Mumbai and Nairobi are a bit larger than that in Rio. However, the similarity between slums and non-slums in Rio is higher than that in Mumbai and Nairobi. This similarity is mainly caused by the existence of cortiço whose appearances are similar to formal settlements. The second reason is due to different reference data collection approaches. The reference data of Mumbai and Nairobi were collected by visual interpretation based on very high-resolution satellite imagery. Thus, slums determined in the reference data were mainly dependent on their morphological characteristics reflected on satellite imagery. The reference data of Rio were collected mainly based on ground information which combines a lot of social and economic factors. These factors are hard to be detected from satellite imagery directly, and, thus, slums in the reference data for Rio can be hard to be detected by RS-based approaches. Besides, due to data aggregation of slums in Rio reference data, some non-slums have been included in the reference data, which increases the similarity between slums and non-slums and thus reduces the performance of the FCN models. For these two reasons, the performance of FCN models varied when applied to different selected study areas and the model trained and tested on Rio data performs worst.

This research found out that the third dimension of the spatial transferability is bad. The conclusion can be supported by the results shown in big differences in values of IoU and F1 score (e.g. IoU of N_M is 1.31%, IoU of N_N reaches 42.39%) in table 5.2 and table 5.3. The reason accounting for the low performance of the third dimension of the spatial transferability can be attributed to the different morphological characteristics of slums in Mumbai, Nairobi and Rio. Therefore, learnt features from the model trained on data from one of the cities were not effective for detecting slums in other cities. However, it also indicated that pre-trained FCN model may help detect slums with similar characteristics to that of slums in the source study area for training the model.

Adaptation 1 can help improve the third dimension of the spatial transferability. Yet it isunable to improve the second dimension. This is supported by $SD_{D2} < SD_{A1} < SD_{D3}$ and $Mean_{D2} > Mean_{A1} > Mean_{D3}$ as shown in table 5.4. Adaptation 3 performs similarly to adaptation 1. It can improve the third dimension of the spatial transferability while cannot improve the second dimension. Similar results can be seen in table 5.6.

Adaptation 2 can help improve both the second and third dimension of the spatial transferability. This is supported by $SD_{A2} < SD_{D2} < SD_{D3}$, $Mean_{A2} > Mean_{D2} > Mean_{D3}$. This improvement may be caused by the increased amount of training data for all three cities. This finding indicated that combining slum data from multiple study areas for training can help increase the third and second dimensions of the spatial transferability of the FCN models. In this way, it is possible to map slums at large scale and produce comparable or even better accuracies compared to the FCN models trained and tested on data from only one study area. It only requires changing testing data when applied to different study areas and thus also easier in application.

Based on the results, it can be concluded that both the second dimension and the third dimension of the spatial transferability of FCN models for slum mapping are low. This is caused by the diverse nature of slums and different reference data collection approaches. Adaptation 1 and adaptation 3 can improve the third dimension of the spatial transferability of FCNs while cannot improve the second dimension. Adaptation 2 can improve both the second and third dimensions of the spatial transferability. The finding from adaptation 2 indicates the potentials of combining training datasets from various study areas with different characteristics of slums for mapping slums at large scale by only using one FCN model which may produce comparable or even higher accuracies compared to the FCN model trained and tested on the same study area. This can be an interesting future direction for the slum mapping researches related to the spatial transferability of FCN models.

Besides, this research applied PlanetScope as input satellite imagery for training and testing the FCN model. The results reveal that the FCN model trained and tested on Mumbai data (M_M) reaches an IoU of 65.09%. Wurm et al. (2019) applied FCN-VGG19 for slum mapping in Mumbai. The IoU of the FCN-VGG19 model trained on QuickBird imagery reaches an IoU of 77.02%. The same model trained on Sentinel-2 data reaches an IoU of 35.51%. Regardless of the influences of the selection of FCN models, hyperparameter setting, training and testing data on prediction results, it can be found that PlanetScope imagery performs much better than Sentinel-2 imagery but still worse than QuickBird imagery. Therefore, PlanetScope as open-source data for researchers may be a better substitute for VHR satellite imagery compared to Sentinel-2 if researches or projects have limited financial budget.

6.2. Recommendations

Based on the content of this research, there are two recommendations for future researches:

(1) It is recommended to explore the first dimension of the spatial transferability. In other words, to explore how the selection of FCN architectures and hyperparameter setting influence on the prediction accuracies for different study areas;

(2) It is recommended to try to combine data from more study areas with various slum characteristics for training FCN models. The higher accuracies of MNR_M, MNR_N and MNR_R compared to M_M, N_N and R_R indicate that combining training data from multiple study areas may work better in slum mapping by FCNs;

6.3. Reflections to Spatial Engineering

This research tries to explore the spatial transferability of FCNs for slum mapping to fill knowledge gaps in spatial information of slums to help solve the wicked problem – slum upgrading. However, the results turn out that the spatial transferability of FCNs for slum mapping is low in terms of the second and third dimensions. The low performance is determined by the diversity within slums, similarity between slums and non-slums and various reference data collection approaches. However, the results from the adaptation 2 indicated that it may be possible to combine slum reference data from multiple study areas together as training data for FCN models and help detect slums for resource constraint areas. FCNs may be a good approach to map slums but it is still a long way to go before it can produce high quality slum maps for application.

APPENDIX

Annex 1: Training and testing tiles for Mumbai, Nairobi and Rio

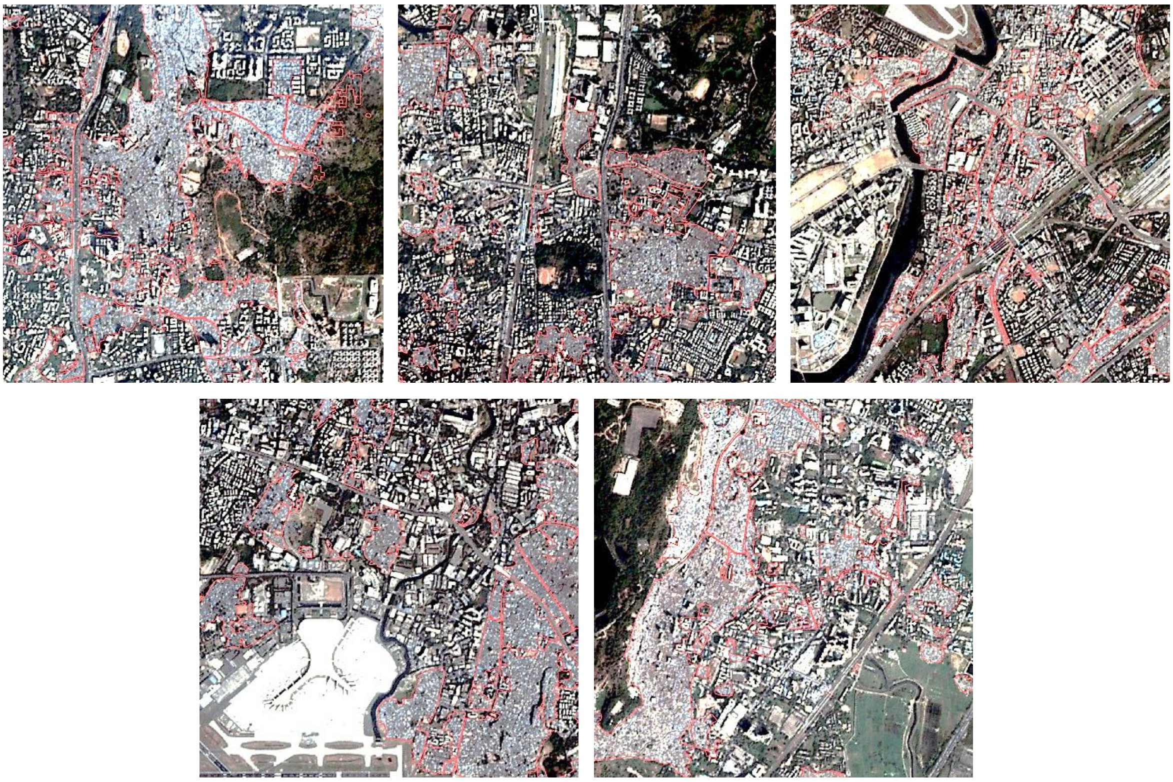


Figure A.1 Training tiles of PlanetScope imagery for Mumbai (Red boundary: the boundary of slums)

Figure A.2 Testing tiles of PlanetScope imagery for Mumbai

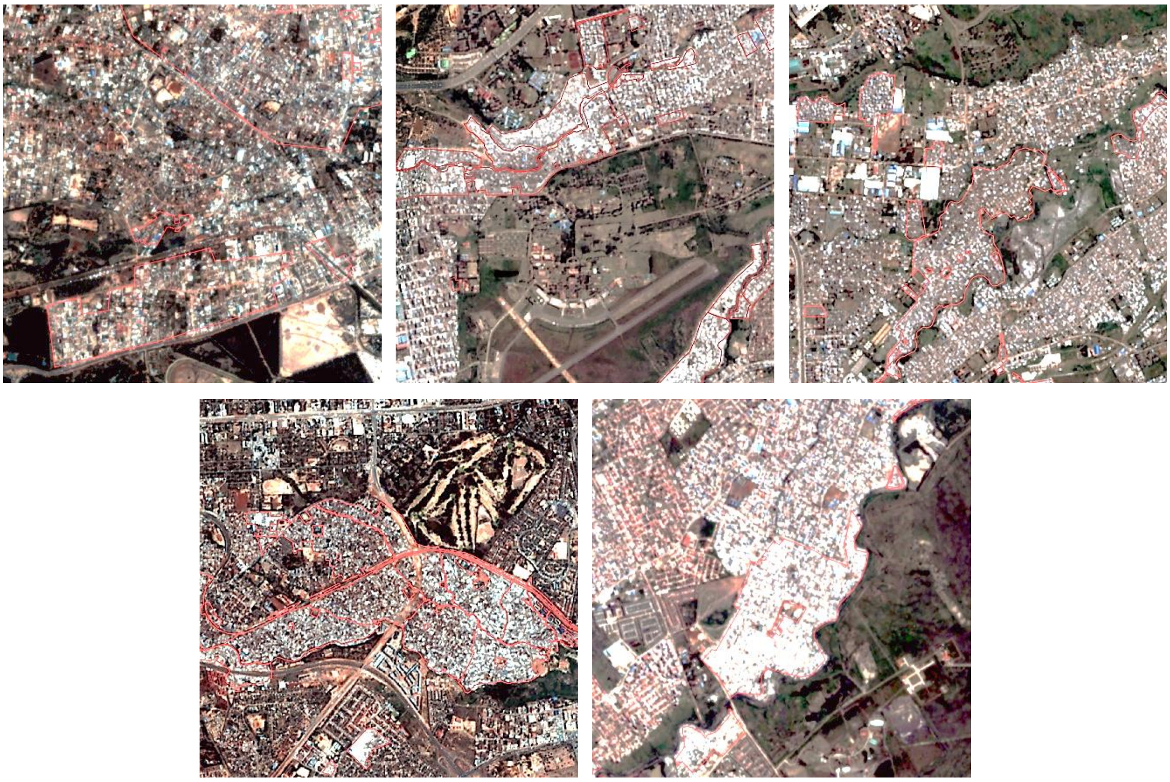


Figure A.3 Training tiles of PlanetScope imagery for Nairobi (Red boundary: the boundary of slums)

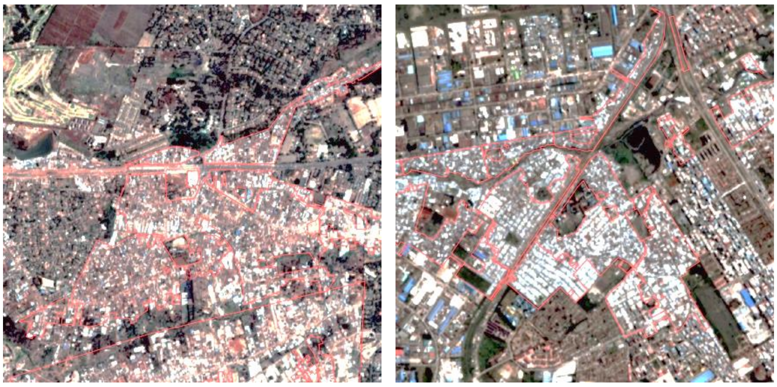


Figure A.4 Testing tiles of PlanetScope imagery for Nairobi

Figure A.5 Training tiles of PlanetScope imagery for Rio (Red boundary: the boundary of slums)

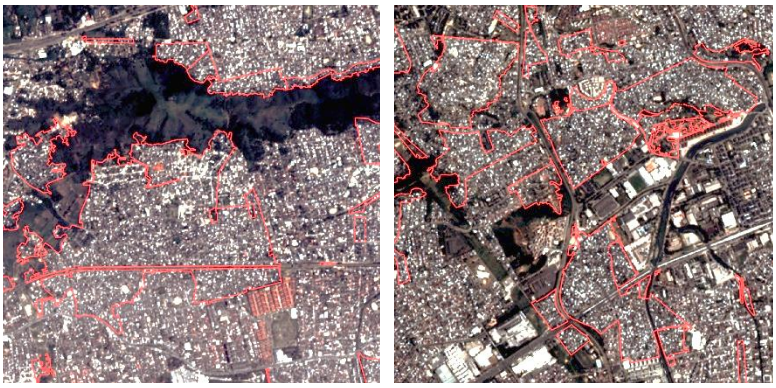


Figure A.6 Testing tiles of PlanetScope imagery for Rio

Annex 2: Precision, recall, F1 score and IoU of two testing tiles under all experiments for Mumbai, Nairobi and Rio in this research

Table A.1 Precision, recall, F1 score and IoU of two testing tiles (TS1 and TS2) and their overall accuracies under all experiments for Mumbai

Experiments	Mumbai											
	TS1				TS2				Average			
	precision	recall	F1	IoU	precision	recall	F1	IoU	precision	recall	F1	IoU
M	81.87%	71.57%	76.37%	61.63%	88.43%	76.95%	82.29%	69.65%	84.77%	73.95%	79.00%	65.09%
N	4.63%	0.64%	1.13%	0.57%	0.95%	0.03%	0.06%	0.03%	4.09%	0.37%	0.68%	0.34%
R	10.69%	11.20%	10.94%	5.78%	18.37%	42.06%	25.57%	14.64%	15.57%	24.89%	19.16%	10.58%
M_ft_N	37.60%	0.96%	1.87%	0.94%	16.85%	0.03%	0.07%	0.03%	36.37%	0.55%	1.08%	0.54%
M_ft_R	29.60%	100.00%	45.68%	29.60%	24.90%	99.52%	39.83%	24.87%	27.32%	99.79%	42.89%	27.30%
MN	81.72%	74.02%	77.68%	63.35%	85.69%	79.97%	82.73%	70.28%	83.51%	76.66%	79.94%	66.38%
MR	84.07%	72.03%	77.58%	63.19%	87.42%	68.11%	76.56%	61.79%	85.48%	70.29%	77.14%	62.58%
NR	48.15%	78.51%	59.69%	42.47%	18.15%	11.58%	14.14%	7.59%	41.01%	48.81%	44.57%	28.61%
MN_ft_M	82.49%	70.54%	76.05%	61.19%	91.83%	62.67%	74.50%	59.12%	86.12%	67.05%	75.40%	60.32%
MN_ft_N	73.44%	19.85%	31.25%	18.49%	88.59%	19.78%	32.34%	19.23%	79.46%	19.82%	31.73%	18.81%
MN_ft_R	34.78%	98.85%	51.45%	34.64%	29.82%	94.26%	45.30%	29.26%	32.45%	96.81%	48.60%	32.09%
MNR	82.29%	72.68%	72.68%	62.68%	89.35%	77.93%	83.25%	71.04%	85.40%	75.01%	79.87%	66.27%
MNR_ft_M	80.56%	73.50%	76.87%	62.28%	89.20%	73.00%	80.29%	66.82%	84.17%	73.28%	78.34%	65.04%

Table A.2 Precision, recall, F1 score and IoU of two testing tiles (TS1 and TS2) and their overall accuracies under all experiments for Nairobi

Experiments	Nairobi											
	TS1				TS2				Average			
	precision	recall	F1	IoU	precision	recall	F1	IoU	precision	recall	F1	IoU
M	71.89%	1.40%	2.75%	1.38%	12.99%	1.34%	2.43%	1.22%	24.25%	1.37%	2.60%	1.31%
N	56.85%	45.91%	50.80%	33.92%	78.67%	67.80%	72.83%	56.86%	66.88%	55.62%	60.73%	43.39%
R	13.43%	1.16%	2.14%	1.07%	0.00%	0.00%	0.00%	0.00%	13.37%	0.65%	1.23%	0.62%
M_ft_N	64.25%	35.37%	45.63%	29.43%	72.81%	78.07%	75.35%	60.05%	69.45%	54.32%	60.96%	43.61%
M_ft_R	23.81%	1.41%	2.65%	1.34%	0.00%	0.00%	0.00%	0.00%	23.80%	0.78%	1.51%	0.76%
MN	66.52%	34.88%	45.76%	29.53%	79.63%	38.21%	51.64%	34.54%	72.05%	36.36%	48.33%	31.67%
MR	77.67%	0.30%	0.59%	0.30%	0.00%	0.00%	0.00%	0.00%	77.67%	0.17%	0.33%	0.16%
NR	58.97%	46.28%	51.86%	34.89%	68.83%	70.40%	69.61%	53.05%	64.00%	56.98%	60.29%	42.95%
MN_ft_M	69.09%	37.30%	48.44%	31.82%	73.51%	43.42%	54.59%	37.27%	71.15%	40.01%	51.22%	34.23%
MN_ft_N	60.12%	35.55%	44.68%	28.64%	79.07%	69.36%	73.90%	58.19%	70.39%	50.55%	58.84%	41.46%
MN_ft_R	13.59%	0.01%	0.03%	0.01%	64.92%	0.14%	0.28%	0.14%	45.70%	0.07%	0.14%	0.07%
MNR	80.13%	15.88%	26.50%	15.19%	80.53%	67.31%	73.33%	57.47%	80.44%	38.70%	52.26%	35.15%
MNR_ft_N	66.54%	32.06%	43.27%	27.48%	77.41%	76.21%	76.80%	61.92%	73.28%	51.65%	60.59%	43.16%

Table A.3 Precision, recall, F1 score and IoU of two testing tiles (TS1 and TS2) and their overall accuracies under all experiments for Rio

Experiments	Rio											
	TS1				TS2				Average			
	precision	recall	F1	IoU	precision	recall	F1	IoU	precision	recall	F1	IoU
M	17.89%	0.09%	0.18%	0.09%	0.00%	0.00%	0.01%	0.00%	15.05%	0.04%	0.07%	0.03%
N	23.09%	47.69%	31.12%	18.38%	39.86%	4.74%	9.09%	4.39%	24.48%	21.48%	22.88%	12.87%
R	46.60%	41.65%	43.99%	27.97%	55.55%	46.59%	49.89%	33.79%	51.92%	44.67%	48.02%	31.42%
M_ft_N	30.00%	4.77%	8.24%	4.25%	0.00%	0.00%	0.90%	0.00%	29.41%	1.86%	3.50%	1.77%
M_ft_R	47.93%	47.06%	47.49%	30.90%	61.40%	35.52%	43.18%	28.86%	54.40%	40.02%	46.11%	29.76%
MN	10.56%	1.44%	2.54%	1.28%	0.06%	0.00%	0.27%	0.00%	9.88%	0.56%	1.07%	0.53%
MR	44.27%	26.26%	32.97%	19.56%	58.23%	54.68%	53.74%	39.08%	54.21%	43.60%	48.33%	31.66%
NR	33.06%	9.13%	14.31%	7.63%	68.74%	30.38%	41.53%	26.53%	58.56%	22.10%	32.09%	18.96%
MN_ft_M	16.36%	3.97%	6.39%	3.28%	32.70%	9.06%	18.37%	7.60%	26.84%	7.08%	11.20%	5.90%
MN_ft_N	5.86%	1.71%	2.64%	1.33%	20.46%	0.19%	3.14%	0.19%	6.55%	0.78%	1.39%	0.70%
MN_ft_R	59.64%	21.13%	31.20%	18.29%	57.89%	21.92%	34.33%	18.80%	58.55%	21.61%	31.57%	18.60%
MNR	48.22%	37.19%	41.99%	26.35%	64.29%	49.24%	55.77%	38.45%	58.00%	50.39%	33.45%	48.22%
MNR_ft_R	41.67%	33.52%	37.16%	22.63%	64.47%	35.36%	45.36%	29.44%	44.54%	34.65%	42.04%	29.16%

LIST OF REFERENCES

Agyemang-Duah, K., & Hall, F. L. (1997). Spatial transferability of an ordered response model of trip generation. *Transportation Research Part A: Policy and Practice*, *31*(5), 389–402. Retrieved from https://doi.org/10.1016/S0965-8564(96)00035-3

Ajami, A., Kuffer, M., Persello, C., & Pfeffer, K. (2019). Identifying a slums' degree of deprivation from VHR images using convolutional neural networks. *Remote Sensing*, *11*(11). https://doi.org/10.3390/rs11111282

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan india: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation*, *12*(5), 359–374. https://doi.org/10.1016/j.jag.2010.04.008

Cambridge Business English Dictionary. (n.d.). TRANSFERABLE | meaning in the Cambridge English Dictionary. Retrieved June 11, 2020, from https://dictionary.cambridge.org/dictionary/english/transferable

Carr-Hill, R. (2013). Missing Millions and Measuring Development Progress. *World Development*. https://doi.org/10.1016/j.worlddev.2012.12.017

Choi, D., Cho, H., & Rhee, W. (2019). On the Difficulty of DNN Hyperparameter Optimization Using Learning Curve Prediction. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, *2018-Octob*(October), 651–656. https://doi.org/10.1109/TENCON.2018.8650070

Data.rio. (2018). Limite Favelas.

Dögg, S., & Pétursdóttir, D. (2011). Technology Enabled Citizen Participation in Nairobi Slum Upgrades by Technology Enabled Citizen Participation in Nairobi Slum Upgrades. *Civil Engineering*, (December).

Ferreira, M. A., Machado, H. A., Franco, & de Mello Franco, F. (2020). São Paulo, Brazil. *Parking*, 35–60. https://doi.org/10.1016/b978-0-12-815265-2.00003-0

Fricke, J. (2015). Slums in Rio de Janeiro Spatial and morphologic analyses of slums derived from remote sensing data based on visual image interpretation. Retrieved from http://elib.dlr.de/99728/1/Fricke_BA.pdf

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

Hoppe, R. (2018). Rules-of-thumb for problem-structuring policy design. *Policy Design and Practice*, *1*(1), 12–29. https://doi.org/10.1080/25741292.2018.1427419

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations, *5*(2), 1–11.

Huang, X., & Zhang, L. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 257–272. https://doi.org/10.1109/TGRS.2012.2202912

Ibrahim, M. R., Titheridge, H., Cheng, T., & Haworth, J. (2019). predictSLUMS: A new model for identifying and predicting informal settlements and slums in cities from street intersections using machine learning. *Computers, Environment and Urban Systems*, *76*(July 2018), 31–56. https://doi.org/10.1016/j.compenvurbsys.2019.03.005

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*, *1*, 448–456.

Jean, N., & Luo, R. (2016). Finding Poverty in Satellite Images. *CS229 Stanford*, 2–6. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjkro72sObiAhVJ3qQKHc0dBV4QFjAAegQIABAC&url=https%3A%2F%2Fpdfs.semanticscholar.org%2F896f%2F58a44b6e01529b4feff5c447223cb3cc7273.pdf&usg=AOvVaw1ViF_kHsKgXiz5YkvyuUy-

Joshi, P., Sen, S., & Hobson, J. (2002). Experiences with surveying and mapping Pune and Sangli slums on a geographical information system (GIS). *Environment and Urbanization*, *14*(2), 225–240. https://doi.org/10.1177/095624780201400218

Kenya National Bureau of Statistics. (2019). *2019 Kenya population and housing census Volume IV. 2019 Kenya* *population and housing census* (Vol. I).

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

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

Kuffer, M., Pfeffer, K., Sliuzas, R., Taubenbock, H., Baud, I., & Van Maarseveen, M. (2018). Capturing the Urban Divide in Nighttime Light Images from the International Space Station. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(8), 2578–2586. https://doi.org/10.1109/JSTARS.2018.2828340

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). https://doi.org/10.3390/rs10101522

Lilford, R., Kyobutungi, C., Ndugwa, R., Sartori, J., Watson, S. I., Sliuzas, R., … Ezeh, A. (2019). Because space matters: Conceptual framework to help distinguish slum from non-slum urban areas. *BMJ Global Health*, *4*(2). https://doi.org/10.1136/bmjgh-2018-001267

Liu, Q., Salberg, A. B., & Jenssen, R. (2018). A comparison of deep learning architectures for semantic mapping of very high resolution images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2018-July*, 6943–6946. https://doi.org/10.1109/IGARSS.2018.8518533

Liu, R. (2018). Mapping the Temporal Dynamics of Slums From Vhr Imagery Mapping the Temporal Dynamics of Slums From. *Thesis*, (February).

Liu, R., Kuffer, M., & Persello, C. (2019). The temporal dynamics of slums employing a CNN-based change detection approach. *Remote Sensing*, *11*(23). https://doi.org/10.3390/rs11232844

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

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

Njoroge, D. . (2016). Growth and Eviction of Informal Settlements in Nairobi, 103. Retrieved from http://www.itc.nl/library/papers_2016/msc/upm/githira.pdf

Padt, F., & Arts, B. (2014). Concepts of Scale. *Scale-Sensitive Governance of the Environment*, *2001*(Chapter 2), 1–16. https://doi.org/10.1002/9781118567135.ch1

Panboonyuen, T., Jitkajornwanich, K., Lawawirojwong, S., Srestasathiern, P., & Vateekul, P. (2017). Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields. *Remote Sensing*, *9*(7), 1–19. https://doi.org/10.3390/rs9070680

Peng, D., Zhang, Y., & Guan, H. (2019). End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sensing*, *11*(11), 20–22. https://doi.org/10.3390/rs11111382

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

Planet Labs. (2017). *PLANET IMAGERY PRODUCT SPECIFICATION*. Retrieved from https://www.planet.com/products/satellite-imagery/files/1610.06_Spec Sheet_Combined_Imagery_Product_Letter_ENGv1.pdf

Pratomo, J., Kuffer, M., Martinez, J., & Kohli, D. (2016). Uncertainties in analyzing the transferability of the generic slum ontology, (2012). https://doi.org/10.3990/2.428

Pratomo, Jati. (2016). Transferability of The Generic and Local Ontology of Slum in Multi-temporal Imagery, Case Study: Jakarta, (March), 90. https://doi.org/10.13140/RG.2.2.18849.81764

Pratomo, Jati, 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

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.

Sliuzas, R., Mboup, G., & de Sherbinin, A. (2008). Report of the expert group meeting on slum identification and mapping. *Report by CIESIN, UN-Habitat, ITC*, (September 2008), 36.

Stark, T. (2018). Using Deep Convolutional Neural Networks for the Identification of Informal Settlements to Improve a Sustainable Development in Urban Environments. Retrieved from https://elib.dlr.de/119019/1/Stark_MA.pdf

Stark, T., Wurm, M., Taubenbock, H., & Zhu, X. X. (2019). Slum mapping in imbalanced remote sensing datasets using transfer learned deep features. *2019 Joint Urban Remote Sensing Event, JURSE 2019*, 1–4. https://doi.org/10.1109/JURSE.2019.8808965

Taubenböck, H., & Kraff, N. J. (2014). The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data. *Journal of Housing and the Built Environment*, *29*(1), 15–38. https://doi.org/10.1007/s10901-013-9333-x

Taubenböck, Hannes, & Wurm, M. (2015). Ich weiß, dass ich nichts weiß – Bevölkerungsschätzung in der Megacity Mumbai (pp. 171–178). https://doi.org/10.1007/978-3-662-44841-0_18

The Census Organization of India. (2011). Mumbai (Greater Mumbai) City Census 2011 data. Retrieved May 7, 2020, from http://www.census2011.co.in/census/city/365-mumbai.html

Thomson, D. R., Ku, M., Boo, G., Hati, B., Grippa, T., Elsey, H., … Sliuzas, R. (2020). Need for an Integrated Deprived Area " Slum " Mapping System (IDEAMAPS) in Low- and Middle-Income Countries (LMICs). *Social Sciences*, *9*(80).

UN-Habitat-b. (2016). Slum Almanac 2015–2016: Tracking Improvement in the Lives of Slum Dwellers. *Participatory Slum Upgrading Programme.*, *s4-XII*(308), 413–413. https://doi.org/10.1093/nq/s4-xii.308.413b

UN-Habitat. (2003). *The Challenge of Slums. Global Report on Human Settlements 2003.* London.

UN-Habitat. (2013). STATE OF THE WORLD'S CITIES 2012/2013 Prosperity of Cities.

UN-Habitat. (2017). Distinguishing slum from non-slum areas to identify occupants' issues. Retrieved June 9, 2020, from https://unhabitat.org/distinguishing-slum-from-non-slum-areas-to-identify-occupants-issues

UN-Habitat, OHCHR, & UNOPS. (2016). Habitat III – Informal Settlements, *22*(May), 1–9. https://doi.org/http://dx.doi.org/10.3402/gha.v5i0.19065

UN-Habitat, P. (2020). Participation for inclusive, city-wide slum upgrading.

UN Habitat. (2007). Slums : Some Definitions. *State of the World's Cities 2006/7*, (2), 1–2.

UN Habitat. (2019). The Urban SDG Monitoring Series, (1), 1–10.

United Nations Human Settlements Programme. (2003). *THE CHALLENGE OF SLUMS GLOBAL REPORT ON HUMAN SETTLEMENTS*. *Earthscan Publications Ltd* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

van Steensel, M. A. M. (2016). Making the invisible visible - Generating data on 'slums' at local, city and global scales. *Seminars in Cell and Developmental Biology*, *52*(December), 58–65. https://doi.org/10.1016/j.semcdb.2016.02.013

Verma, D., Jana, A., & Ramamritham, K. (2019). Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat International*, *88*(July 2018), 0–1. https://doi.org/10.1016/j.habitatint.2019.04.008

Vocabulary.com. (n.d.). transferability - Dictionary Definition. Retrieved June 12, 2020, from https://www.vocabulary.com/dictionary/transferability

Wang, J., Kuffer, M., & Pfeffer, K. (2019). The role of spatial heterogeneity in detecting urban slums. *Computers, Environment and Urban Systems*, *73*(April 2018), 95–107. https://doi.org/10.1016/j.compenvurbsys.2018.08.007

Willis, K. (2019). The Sustainable Development Goals. *The Routledge Handbook of Latin American Development*, 121–131. https://doi.org/10.4324/9781315162935-11

Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*(May 2018), 59–69. https://doi.org/10.1016/j.isprsjprs.2019.02.006

Xia, X. (2019). Extracting Cadastral Boundaries From Uav Images Using Fully Convolutional Networks. Retrieved from https://library.itc.utwente.nl/papers_2019/msc/upm/xia.pdf

Zahidi, I., Yusuf, B., Hamedianfar, A., Shafri, H. Z. M., & Mohamed, T. A. (2015). Object-based classification of QuickBird image and low point density LIDAR for tropical trees and shrubs mapping. *European Journal of Remote Sensing*, *48*, 423–446. https://doi.org/10.5721/EuJRS20154824

Zhang, W., Huang, H., Schmitz, M., Sun, X., Wang, H., & Mayer, H. (2018). Effective fusion of multi-modal remote sensing data in a Fully convolutional network for semantic labeling. *Remote Sensing*, *10*(1). https://doi.org/10.3390/rs10010052

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