Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis BSc^2 in Econometrics and Economics

Seeing the Bigger Picture: Forecasting Economic Growth Using Satellite Imagery and Neural Networks

Jesse Dijkstra (542226)

Supervisor:	dr. A. A. Naghi
Second assessor:	S. J. Koobs
Date final version:	July 8, 2023

Abstract

This thesis examines whether Landsat 7 daytime satellite imagery of US Census blocks in combination with machine learning, specifically Convolutional Neural Networks (CNNs), can effectively help to predict GDP, population, GDP growth, and population growth. In line with research conducted by Khachiyan et al. (2022b), it is evident that machine learning holds significant promise: for GDP and population levels, the CNNs achieve R^2 values that are greater than or equal to 0.844. For the GDP and population difference models, these values are substantially lower at a maximum of 0.552. Against research previously done using a combination of nighttime satellite imagery and standard OLS, this is a great improvement, with R^2 increases of at least 76 percent. Since the difference models perform rather poorly, this thesis trains a new structure that adds to the research done by Khachiyan et al. (2022b), to see whether the models for GDP growth rates can be improved. The findings show that the new structure does not seem to perform significantly better than the original model. However, limited computational resources have complicated the extensive training of the models. Future research should hence focus on testing whether the altered structure indeed proves to be an improvement after training for a longer period of time.

Contents

1 Introduction

In recent years, machine learning has emerged as a powerful tool transforming the field of economics. By using the power of these advanced techniques, economists are able to uncover hidden patterns, analyze complex economic systems, and make more accurate predictions than ever before. Traditional econometric methods often struggle to handle the large volume and complexity of data, limiting their ability to capture the complex dynamics of economic phenomena. In contrast, machine learning algorithms are highly effective at extracting valuable insights from large and varied datasets, allowing economists to gain a deeper understanding of how the economy behaves and evolves over time.

One example of data that normal econometric models struggle with is satellite imagery. It poses a significant challenge for traditional econometric methods, as they are not designed to directly process visual data. Converting images into a format suitable for conventional econometric analysis requires additional difficult transformations. In the past, papers have used nighttime satellite images and converted those into scores that correspond to the degree of nightlight intensity for different regions. Using this score in basic linear models has revealed that nightlight intensity is a good proxy for economic development (Doll, Muller, & Morley, 2006). However, satellite images contain much more hidden information that is lost due to this procedure. Machine learning offers the solution.

More specifically, (Convolutional) Neural Networks have recently been employed more and more in the field of remote sensing and computer science to predict outcomes from daytime satellite images (Khachiyan et al., 2022b). One example of this is in the paper by Khachiyan et al. (2022b). The authors use satellite images and Convolutional Neural Networks to predict income and population levels, as well as income and population growth. They find that their machine learning approach greatly outperforms the traditional Ordinary Least Squares approach.

This thesis is meant to replicate and expand upon their research. The trained models by the authors are used to obtain estimates for log income, log population, log income change, and log population change. Furthermore, these estimates are compared with those derived from simple econometric models to reexamine whether machine learning indeed holds as much promise. Additionally, multiple new models will be trained to see whether the architecture for the log income change can be improved. The research question can hence be formulated as follows:

Does a different model architecture lead to better predictions, when looking at the R^2*,*

than those provided by the model by Khachiyan et al. (2022b) that aims to predict economic growth among US Census blocks?

This thesis hence contributes to the existing literature in multiple ways. It examines whether a different model architecture can improve predictions in the field of economics. If this is the case, officials and policymakers can update their models to more accurately predict economic growth. Scientifically this is relevant, because it might shine more light on the workings of the Convolutional Neural Network and how its architecture can be altered to obtain better predictions. Socially, the relevance concerns the way policymakers can use these types of models to be able to more directly target regions where economic development is low, and where the impact of different policies such as infrastructure projects can be more clearly evaluated on the regional level.

1.1 Economic Relevance

Economically however, this research can add a lot more to the literature. Being able to more correctly predict income and income growth is important in many cases. The first one being the fact that Census data is only collected once every couple of years. For example, in the US, the United States Census Bureau collects data every ten years. Government officials in need of information on the state of the economy at the regional level therefore often have to resort to estimates. The better the estimates, the better informed politicians are in making policy regarding infrastructure development, resource allocation, and targeted interventions to foster economic progress in specific regions. Second, they assist investors and businesses in making informed decisions regarding market entry, expansion, or diversification. Accurate predictions identify regions with high growth potential, enabling strategic resource allocation and capitalization on emerging opportunities. For instance, investors can invest money in markets where the growth model sees the most upside in income development, increasing profits. Third, improved predictions of income and income growth contribute to socio-economic development efforts. In poor and emerging countries with underdeveloped information collection systems, obtaining accurate economic development information is challenging. Increased model accuracy allows governments, international organizations, and NGOs to design effective poverty reduction strategies, allocate resources to underserved areas, and address inequalities by targeting regions most in need. For example, improved growth models can better detect where certain areas are expected to see a large drop in income before Census data could provide this information. Policymakers can then better and earlier redistribute resources to make sure people are supported. Lastly, accurate predictions of economic growth play a crucial role in risk assessment

and environmental monitoring, particularly in the context of climate change. Climate-related hazards such as heatwaves, floods, hurricanes, and droughts pose significant risks to economic stability and growth (European Central Bank, 2022). Accurate predictions help assess the vulnerability of regions to these hazards, facilitating the development of effective adaptation and mitigation strategies. Incorporating climate risk projections into economic growth models helps to identify high-risk areas and take proactive measures to strengthen infrastructure, improve disaster preparedness, and allocate resources for post-disaster recovery. This can greatly reduce the impact disasters have on vulnerable communities and potentially save lives. Furthermore, accurate predictions inform policymakers and stakeholders about the potential environmental impacts of economic activities. Neural Networks, with their ability to analyze satellite imagery, can identify regions that contribute disproportionately to greenhouse gas emissions or environmental damage. This knowledge helps policymakers to create targeted policies and regulations that promote sustainability.

1.2 Thesis Structure

This thesis is structured in the following way. First, in the Literature Review section, the paper dives deep into what has been done before in the economic community surrounding satellite imagery and economic forecasting. It also includes an extensive review of a model used by the World Bank that forecasts GDP growth in the long run. Second, the data used for this research are discussed in the Data section. Moreover, this section provides more information on how this data is split up in order to train and evaluate the models. Third, in the Methodology section, all of the methods and models that are used will be explained from a theoretical standpoint, as well as how the models are evaluated. Fourth, the Results section outlines the results and compares the different models. Finally, the results are put into context and further discussed in the Conclusion & Discussion, where the thesis will also restate some of the relevance the results provide. The section concludes with suggestions for future research.

2 Literature Review

Khachiyan et al. (2022b) were not the first to connect satellite data to GDP (growth). Nighttime satellite images have been used extensively in research about estimating economic variables. Doll et al. (2006) analyzed nighttime radiance data along with regional economic productivity in Europe and the United States. They found a strong relationship between the degree of nighttime radiance and GDP across many spatial scales. In numerous Western European countries and the United States, a strong correlation of more than 0.9 was observed at the subnational level.

However, the authors noted that there was a problem related to economic activity not corresponding to additional light emissions, such as agricultural and fishing activities. Using daylight imagery can solve this issue, since these developments are clearly visible in images captured during the daytime. Similarly, Sutton, Roberts, Elvidge, and Baugh (2001) took nighttime data to estimate the global human population. They first estimated the population of around 23,000 urban clusters using the satellite imagery. Combining these estimates with the known percentage of the population living in urban centers, they obtained an estimate of the population for every nation. By aggregating those values, an estimation of the global population was derived. They arrived at an estimate of around 6.3 billion, which was very close to the officially accepted estimate at the time of around 5.9 billion.

The two papers show the power satellite imagery has in improving models to predict variables such as GDP and population but they fail to show the potential of deep learning. Deep learning is a branch of machine learning that utilizes Neural Networks with multiple layers to process large datasets and automatically discover intricate patterns, enabling advanced pattern recognition. In a paper that is much closer to the paper by Khachiyan et al. (2022b), Perez et al. (2017) analyzed Landsat 7 imagery using machine learning to predict poverty in different countries in Africa. They trained multiple Convolutional Neural Network (CNN) structures to see whether daylight satellite images are useful in this context. They found that using daytime images in combination with a Neural Network significantly improves the predictions compared to using only nighttime images in regression. Moreover, they found that the predictions from the CNNs were much more accurate among countries in the poorer segment of the wealth distribution. Therefore, it is possible that the transition from extreme poverty to less poverty translates more easily to tangible development on the ground. In a very similar paper, Okaidat, Melhem, Alenezi, and Duwairi (2021) aimed to predict the distribution of poverty in Ethiopia, Malawi and Nigeria using deep learning and satellite images. They compared multiple architectures of CNNs, among which the VGG-16 model that also inspired the models trained by Khachiyan et al. (2022b). The researchers found that their best performing CNN architecture achieves a test accuracy of nearly eighty percent. The original paper by Khachiyan et al. (2022b) found even better accuracy for the models that estimate GDP and population levels, with R^2 values substantially higher than 0.8. For the difference models, that estimate the growth rates of these variables, the accuracy was significantly lower, with R^2 values of below 0.5.

They were not the first researchers to use a CNN for looking at changes in satellite images

between two time periods. Xu, Lu, Li, Khaitan, and Zaytseva (2019), for example, compared the performance of multiple CNN architectures in detecting damaged buildings after the 2010 Haiti earthquake. They found that all four models perform very well, with areas under the receiver operating characteristic curve (AUCs) of above 0.8. The AUC is a metric used to evaluate the performance of a binary classification model, where a higher AUC value indicates better discrimination between positive and negative classes. Moreover, they found that their trained model still performs reasonably well (with accuracies of just below 0.7) when tested on different datasets, such as Mexico and Indonesia, showing that the model generalizes well to different disasters and to new regions with similar levels of economic development.

2.1 Extensive Review of Economic Model Predicting GDP Growth

Next to the machine learning method Convolutional Neural Networks used in this thesis, other methods have been employed to predict GDP growth, both in the short term and long term. This section outlines a major economic model, employed by the World Bank, that aims to predict economic growth in the long run. This section first provides an overview of the model's structure and subsequently examines two papers that implement the model in real-world applications.

2.1.1 Long Term Growth Model

World Bank Group (2023) have used the so-called Long Term Growth Model (LTGM) in over 40 growth analyses and country reports. This model is based on the Solow-Swan Growth Model. The Solow-Swan Growth Model is a neoclassical economic model that aims to explain long-run economic growth by considering the accumulation of capital and technological progress (Solow, 1956). The model provides insights into the factors that drive economic growth. One of the fundamental assumptions of the Solow-Swan model is that savings and investments are key to economic growth. Moreover, production exhibits diminishing returns to capital and labor inputs. This implies that as more capital and labor are used in the production process, the additional output generated from each additional unit of capital or labor decreases. This assumption reflects the idea that resources become less productive as they become more abundant. The building blocks of the LTGM are given by the following three equations (Pennings, 2017):

$$Y_t(GDP) = A_t K_t^{1-\beta} (h_t L_t)^{\beta}, \tag{1}$$

where $Y_t(GDP)$ denotes the GDP of a country at time t, and A_t is the Total Factor Productivity (TFP) which reflects the extent to which the inputs (capital and labor) are efficiently utilized in the production process, capturing the overall efficiency of an economy beyond what can be

attributed solely to changes in input quantities. K_t is the amount of capital used in the inputs at time t, h_t denotes the human capital per worker, L_t the number of workers, and β the share of labor used as inputs. The equation represents the production function and implies that the level of GDP is determined by the inputs and the productive capacity of an economy, which is influenced by the efficiency of resource allocation, technological progress, and the skill level of the workforce. The second equation is given by:

$$K_{t+1} = (1-\delta)K_t + I_t, \tag{2}$$

where δ denotes the depreciation rate of (physical) capital, and I_t is the amount of investments at time t. The equation represents the capital accumulation process in the economy, where the amount of capital in the next period is equal to whatever is left from the current period plus the additional investments done. The third equation is given by:

$$\begin{aligned} y_t^{PC}(GDP\ per\ capita) &= \frac{Y_t}{N_t} \\ &= \frac{Y_t}{L_t}\frac{L_t}{W_t}\frac{W_t}{N_t} \\ &= \frac{A_t K_t^{1-\beta}(h_t L_t)^{\beta}}{L_t}\rho_t\omega_t \\ &= A_t\left(\frac{K_t}{L_t}\right)^{1-\beta} h_t^{\beta}\rho_t\omega_t \\ &= A_t k_t^{1-\beta} h_t^{\beta}\rho_t\omega_t, \end{aligned} \tag{3}$$

where N_t is the total population, W_t the total working-age population, $\rho_t := \frac{L_t}{W_t}$ the participation rate, $\omega_t := \frac{W_t}{N_t}$ the working-age-population to population ratio, and $k_t := \frac{K_t}{L_t}$ the capital per worker ratio.

Using Equation (3) yields:

$$g_{y,t+1} \approx g_{A,t+1} + \beta(g_{h,t+1} + g_{\omega,t+1} + g_{N,t+1} + g_{\rho,t+1}) + \left(\frac{1-\beta}{K_t/Y_t}\right)\frac{I_t}{Y_t} - (1-\beta)\delta, \tag{4}$$

where the g denotes the growth rate of the respective variables. It shows that GDP growth is fuelled by boosting productivity, human capital, the labor force, and the level of investments. According to Pennings (2017), in order to obtain a sustainable level of income growth, governments should fund the necessary level of investments either by increasing the savings rate or by attracting a constant flow of foreign direct investment.

To solve the model, only a minimal set of data is required, which is limited to three parameters: the labor share (β), the depreciation rate (δ), and the initial capital-to-output ratio $\left(\frac{K_0}{Y_0}\right)$. Moreover, researchers should make assumptions about the expected growth rates of TFP, human capital per worker, population, working-age-population ratio, participation rates, and the level of investment share of GDP.

Next to the basic model, researchers from the World Bank Group (2023) have introduced extensions to incorporate additional factors to ensure that the model better captures the complexities of real-world economic dynamics. For example, public capital, and a country's scores for innovation, infrastructure, institutions, education, and health are all included in the extended models. These augmented models have been employed in multiple World Bank reports, analyzing the economic development paths of various countries. For example, Jeong (2017) analyzed the rapid growth South Korea experienced in the last six decades using the LTGM. The study highlighted that the accumulation of human capital and advancements in productivity served as the primary drivers of sustained GDP growth, rather than labor or capital investments. By calibrating the model and creating counterfactual scenarios, the model was able to effectively explain the actual growth path of the Korean economy. These findings demonstrate the valuable insights and practical tools the model offers policymakers when designing policies to boost economic growth. Another example of the use of the LTGM to study economic growth in a developing country is a paper by Devadas, Guzman, Kim, Loayza, and Pennings (2019). The researchers examined economic growth in Malaysia with the aim of assessing the potential for the country to attain the status of a high-income country. As Malaysia's economy is slowly maturing, the expected GDP growth is expected to fall from 4.5 to 2.0 percent over the next 30 years. According to the researchers, implementing strong reforms targeting human capital, TFP, and female labor force participation is crucial to propel the economy towards achieving the high-income country status. The adoption of such measures has the potential to significantly increase Malaysia's future GDP growth rate, reaching 3.6 percent instead of a mere 2.0 in 2050. Again, this shows that the model can effectively help policymakers in deciding on their strategy to improve GDP growth.

As shown, the LTGM is able to deliver promising results in the analysis of long-term income growth paths for multiple countries, and is able to aid policymakers in focusing on important growth drivers. However, it is clear that since the model is meant for long-term analysis, it does not perform very well over a shorter forecast horizon. This underlines the importance for other tools, such as CNNs, to obtain a more complete picture of the trends in economic development

for multiple countries.

3 Data

The satellite imagery that is used in this research is taken from the dataset provided by Khachiyan et al. (2022b) and was captured by the US Geological Survey (USGS) Landsat 7 satellite. Launched in 1999, the Landsat 7 satellite remains active and covers extensive areas within the United States (US Geological Survey, n.d.). The satellite can capture seven different spectral bands: three are the visible bands (red, blue, and green), two are near-infrared, one is thermal, and one is mid-infrared. The spatial resolution of these images is 30m, i.e., each pixel corresponds to an area on the ground that is 30 meters by 30 meters. Khachiyan et al. (2022b) had access to these pictures and extracted them using the Google Earth Engine (GEE). In order to ensure that images were clear and free from snow or clouds, the authors constructed annual composites of cloud-free images for the years 2000 and 2010 taken during the summer months (May to August). Figure 1 shows three examples of satellite images captured by the Landsat 7 satellite. Moreover, to make sure that the dataset was not filled with satellite images

Figure 1: Landsat satellite imagery at 30m (low) resolution (80 $\times$ 80 pixels, or 2.4 km^2)

capturing uninhabited areas, Khachiyan et al. (2022b) only included pictures covering urbanized US census block groups, which have at least 600 inhabitants. Using this rule, the authors managed to retain a dataset with 338,796 pictures that cover 93 percent of the population in the continental US in 2000. To match the picture with the correct outcome variables values (so income and population), Khachiyan et al. (2022b) constructed labels based on US Census data for 2000 and 2010, extracted from Manson, Schroeder, Van Riper, Kugler, and Ruggles (2020). Lastly, in addition to satellite images, the trained models also have controls for local economic characteristics as inputs, such as the percentage of women, employment in business

services, and employment in construction. A complete overview of these variables is shown in the Appendix in Table A1. The authors collected these variables from the LODES database, which is administered by the US Census Bureau and provides information on employment and commuting patterns for census blocks (U.S. Census Bureau, 2020).

3.1 Training, Testing, and Validation Split

As is common practice in machine learning, the data is split up into three disjoint parts: a training, a testing, and a validation set. The training set contains the observations the model uses to train its weights, or to estimate the necessary parameters. The models used in this research also contain hyperparameters, which are independent parameters that determine aspects such as the model's architecture or learning rate of the optimizer. They are not influenced by the training set used during the training stage but are rather set a priori by the modeler. For (Convolutional) Neural Networks these are for example the number of layers or the number of nodes. In order to select the model's architecture that performs the best, the validation set is used. For a given set of hyperparameters, the mean squared prediction error (MSPE) is determined on the validation set, after training the model on the training set. The final trained model is then obtained by taking the model with the set of hyperparameters that has the lowest MSPE. Finally, since machine learning models, and especially Neural Networks, are prone to overfitting, they tend to learn the training data exceptionally well, yielding minimal errors after extended training. However, this implies that the same model can perform very poorly when generating out-of-sample predictions. To evaluate the true effectiveness of the model, the testing set is therefore used to assess its performance on unseen data. This approach provides a more comprehensive understanding of the model's actual performance rather than solely evaluating its ability to learn the training data. There exists no general rule for how to exactly split the dataset into three parts. However, a typical split might be 50 percent for training the model and then 25 percent each for validation and testing (Hastie, Tibshirani, & Friedman, 2017). Khachiyan et al. (2022b) use a total of 112,932 spatially unique images for each subset, i.e., they use an equal split among the three parts.

4 Methodology

This section explains the methodology used in this paper. It dives deep into what models are used, how they function, and how they are trained. First, Neural Networks are discussed, forming the basis for Convolutional Neural Networks. Convolutional Neural Networks are the main type of model used in this research. After that, the exact architectures of the different

models are discussed, as well as the extension this paper provides to the original research done by Khachiyan et al. (2022b). Finally, the section discusses how the models' performances are evaluated, including against a benchmark model based on nighttime satellite imagery.

4.1 Neural Networks

In order to understand the bulk of the methodology, which is comprised of Convolutional Neural Networks, one must first understand standard Neural Networks (NNs). The idea of (feedforward) NNs is that linear combinations of the input features are extracted as derived features, after which the output is calculated as a nonlinear function of these features (Hastie et al., 2017). This all happens through a network of (hidden) layers and nodes, as shown in Figure 2. The first layer is the input layer. This contains all the features the modeler deems relevant. The features are then all fed into a network of hidden layers, each with a number of nodes, that transform the inputs in a nonlinear way. This transformation is done using the so-called activation function. More specifically, following the notation of Gu, Kelly, and Xiu (2020), let $K^{(l)}$ denote the

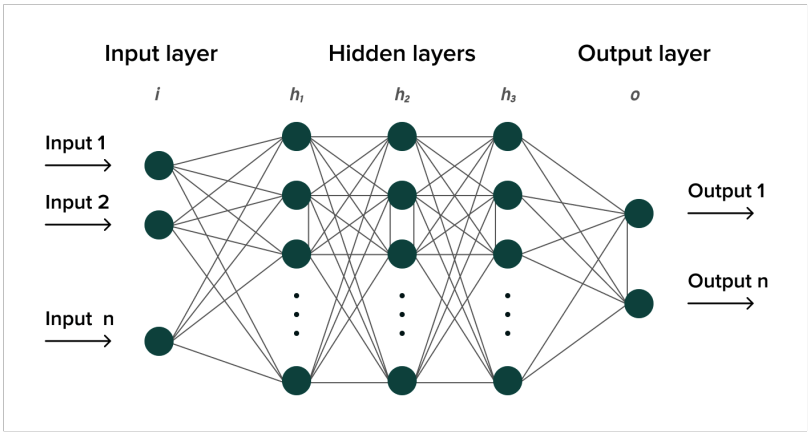


Figure 2: An example of a Neural Network with multiple hidden layers. *Retrieved from* https://serokell.io/blog/introduction-to-Convolutional-neural-networks

number of nodes or neurons in layer l and $x_k^{(l)}$ the output of neuron k in layer l. Moreover, let $\theta_k^{(l)}$ denote the weight assigned to output $x_k^{(l)}$. Combining the outputs in a vector and including a 1 to account for a constant term, also called bias, yields $x^{(l)} = (1, x_1^{(l)}, ..., x_{K^{(l)}}^{(l)})'$. A neuron's output is then calculated as $x_k^{(l)} = f_l(x^{(l-1)'}\theta^{(l-1)})$, where f_l is the activation function used in layer l. Essentially what happens is that the neuron takes a weighted sum of the previous outputs, adds a bias term, and then transforms it using the activation function. There are many different types of activation functions that are used in the literature. Examples of some functions include the sigmoid function, $\sigma(x) = (1 + \exp(-x))^{-1}$, which forces the output to be between zero and one, the hyperbolic tangent, $f(x) = \tanh(x)$, which forces the output to be between

minus one and one, and the Rectified Linear Unit (ReLU) activation function, which is given by $\text{ReLU}(x) = \max(0, x)$. Training the model implies finding the correct weights and biases that optimize a certain metric, such as the mean absolute error, on the training set. This happens during a process called backpropagation, where the gradient of the loss function is calculated with respect to each weight starting at the output node and moving in the opposite direction. Using stochastic gradient descent (SGD), a subset of training examples (mini-batches) is randomly selected in each iteration to compute the gradient and update the model's parameters with a scaled learning rate. The learning rate determines the step size for parameter updates and the optimal rate is found using hyperparameter tuning. The number of training examples in the batch is fixed at 16, following Khachiyan et al. (2022b). Furthermore, the training process is typically performed over multiple iterations known as epochs. Each epoch involves going through the entire training dataset using mini-batches and updating the model's parameters based on the computed gradients. By repeating this process for multiple epochs, the model has the opportunity to refine its weights and biases, progressively improving its performance on the training data.

4.2 Convolutional Neural Networks (CNNs)

In theory, when trying to analyze satellite imagery using Neural Networks, one could take black-and-white satellite pictures, transform each pixel into a score that represents the degree to which the pixel is black and create an input vector with all the pixel scores that is fed into a regular Neural Network. However, this approach does not take into account that there is a correlation between how black one pixel is compared to the one next to it, depending on the objects on the image. Moreover, for large resolutions, the number of nodes in the input layer would be very large, making it a computationally very demanding approach. A slightly modified type of network specifically for multidimensional data such as satellite images is called the Convolutional Neural Network (CNN). CNNs can take into account the spatial dependencies on the image by processing different parts of the image and extracting features from it that are eventually fed into the regular Neural Network, also called the dense network. Again, there are an input layer, hidden layers and an output layer, but the hidden layers now contain one or more Convolutional layers and pooling layers. The basic structure is shown in Figure 3. The Convolutional layers make sure that the most important features of the previous output are extracted. This is done using so-called filters or kernels. The filters are essentially matrices that glide over the pixels of the picture. The layer then calculates the Frobenius inner product of the pixel values and the filter at each area. The Frobenius inner product is calculated as the sum of the element-wise

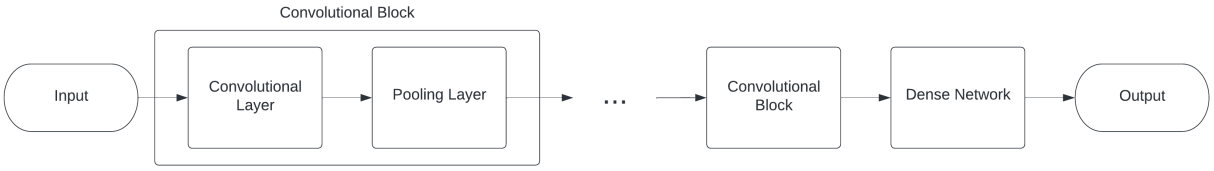
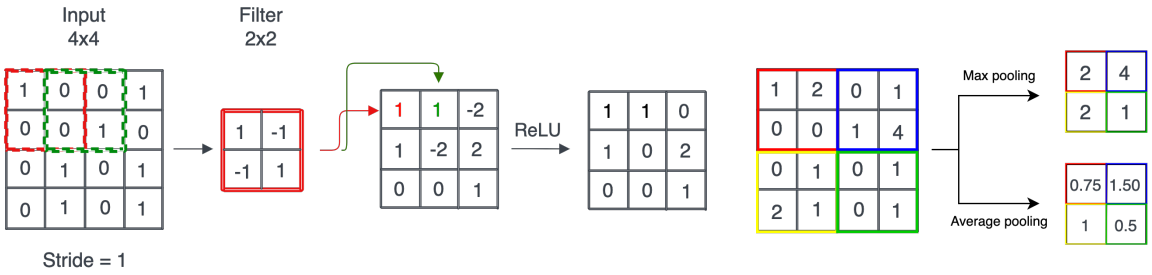


Figure 3: An example of a basic CNN structure

multiplication of the values in both matrices, which have to be of equal size. The output of a Convolutional layer is then again a number of grid-like matrices, the same as the number of filters, with values equal to the inner products that were calculated. The number of values that are calculated depends on the stride of the filter, i.e., the number of pixels the filter shifts after each calculation. Finally, the values are transformed by adding a potential bias term and putting it through an activation function. The conventional activation function used is the ReLU function (He, Zhang, Ren, & Sun, 2015). Training the model implies that the exact weights in the filters and the bias term need to be calculated. The optimal number of filters can be found using hyperparameter tuning. See Figure 4a for a visualization of a Convolutional layer. After one or more Convolutional layers, the output is fed into a pooling layer. This layer reduces the dimensionality by again sliding over the matrix and calculating a value for each area. There are two types of pooling, Average pooling and Max pooling. Average pooling takes the average value of the values in an area, whereas Max pooling takes the maximum value. This step is visualized in Figure 4b. After multiple Convolutional blocks, i.e., combinations of



(a) A Convolutional layer (b) Two types of pooling operations

Figure 4: Structure of a Convolutional block

Convolutional and pooling layers, the output is vectorized and fed into a regular Neural Network. As discussed in the previous subsection, this network consists of hidden layers of neurons.

4.3 Exact CNN Structures Used in Research

For the replication part of this thesis, the exact same code will be used as done in the paper by Khachiyan et al. (2022b). The code contains specifics on the exact architecture of the CNNs used by the authors. More specifically, a version of the VGG16 network model, widely employed in the computer vision community, was used (Khachiyan et al., 2022b). The input of the model are the satellite images. In particular, the inputs are the different channels of the image. For the Landsat 7 imagery, there are seven channels, or spectral bands: the visible bands (red, blue, and green; RGB), two near-infrared, one thermal, and one mid-infrared. Since the 2.4 km^2 images have a resolution of 80 by 80 pixels, the inputs of the models are seven 80×80 pictures. The model has three Convolutional blocks, each consisting of three Convolutional layers and a Max pooling layer, a so-called 'flatten' layer that vectorizes the output of the Convolutional layers, and a dense network. The dense network consists of three fully connected layers, i.e., all neurons in the next layer are connected to the neurons in the previous layer. The number of neurons in each layer depends on the number of filters used in the Convolutional layers. If the first Convolutional block outputs n filters, the first and second hidden layers in the dense network use $16n$ nodes, and the third layer $8n$. The number of filters is multiplied by a factor of two between each Convolutional block. Using hyperparameter tuning, for all the models that the authors consider, the optimal number of filters in the first Convolutional block is found to be equal to $n = 32$. An exact overview of the hyperparameters used in the authors' models can be found in the Appendix in Table A2. The following subsections dive deeper into the models architectures for each type of model.

4.3.1 Levels Models

The levels models aim to predict the level of log income or level of log population. The architecture is as described in the main body of the subsection just above. Weights in the layers are initialized by using the normal distribution. The Convolutional layers use a stride of one and kernel size of three, each with ReLU activation functions. For models that also take into account initial conditions, the vector of features is concatenated to the vector output of the Convolutional blocks and then fed into the dense network. See Figure 5. To prevent overfitting, the authors use multiple regularization techniques. For example, the Convolutional kernels are regularized using an L2 norm penalty, imposing a penalty term that increases with the magnitude of the squared weights in the filter. This forces the coefficients in the filters to not become too large or small, encouraging the model to distribute its learning across multiple kernels rather than relying heavily on a few specific kernels. It prevents the network from becoming too sensitive to

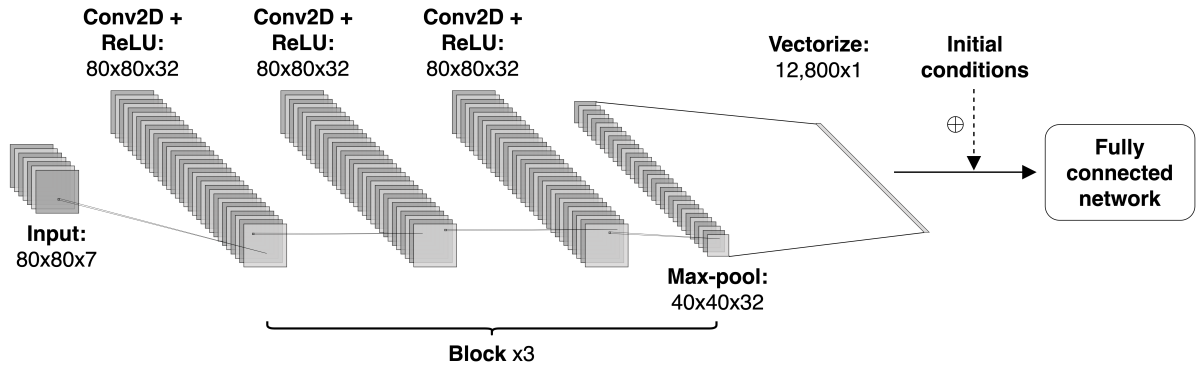


Figure 5: The specific architecture of the CNNs used for the Landsat imagery

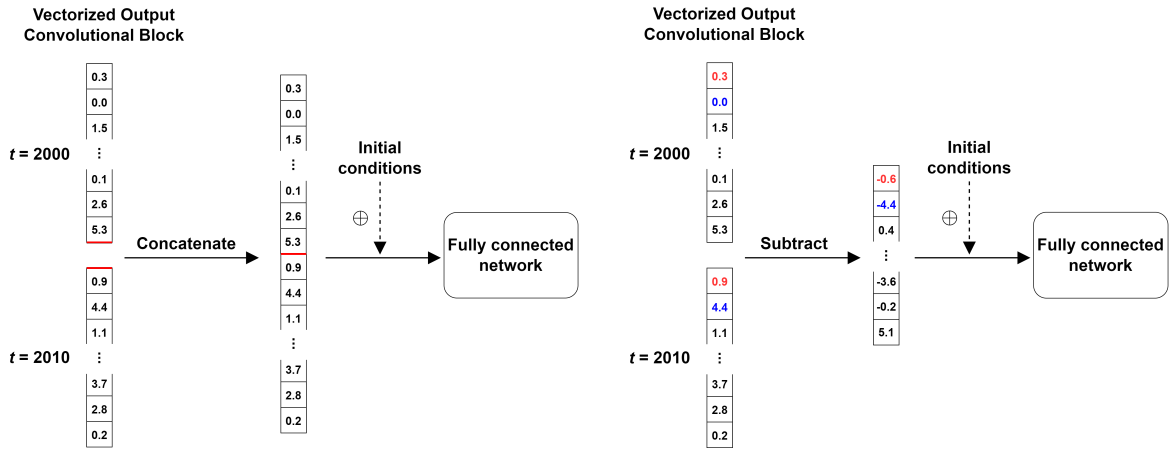
individual kernels, increasing robustness. Furthermore, the three connected layers in the dense network are all separated by a dropout layer. This layer randomly 'turns off' nodes with a certain probability. This prevents individual neurons from relying too heavily on specific input features or other neurons, again making the model more robust. The dropout probability, or rate, is fixed at 0.5. Lastly, all layers in the dense block use ReLU activations and are regularized using the L2 norm penalty. The degree of regularization is found using cross-validation. Cross-validation is a resampling technique that evaluates the performance and generalization ability of a predictive model through multiple iterations. It involves dividing the dataset into subsets and using each subset for both training and validation, allowing for reliable assessment of the model's performance across different data partitions. The used hyperparameter values are listed in Table A2.

4.3.2 Differences Models

The differences models aim to predict changes in log income, log population or log income per capita. The structure of the models hence needs to be altered to accommodate the combination of two images. In particular, a model takes a pair of images of the same area but in different years, feeds them separately through the Convolutional blocks used in the levels models, and then concatenates the two flattened vectorized outputs. This concatenated vector is then passed to the dense block, optionally with the concatenated initial conditions, just like before. The output layer finally yields an estimate of the growth in log income, log population or log income per capita. The regularization techniques and structure of the dense network are identical to the levels models.

4.3.3 Augmented Differences Model

This paper expands upon the research done by Khachiyan et al. (2022b), by focusing on the improvement of the differences models. As will be shown in the Results section, the models for both income differences and population differences perform much worse than the levels models. One reason could be that, with the architecture specified above, the models fail to accurately capture differences in the satellite images between two periods. Hence, this paper aims to train a model using a different structure to see whether the performance can be significantly increased. In their paper on detecting damaged buildings after natural disasters using satellite images, Xu et al. (2019) test different structures of the CNN to check which performs best. The authors call the structure used by Khachiyan et al. (2022b) the Twin-tower Concatenate (TTC) model, where images from different periods are put through the Convolutional blocks separately after which the features are concatenated and fed into the dense network. Next to this model, the Twin-tower Subtract (TTS) model is trained. This architecture combines the features not by concatenating them but by subtracting them element-wise. Figure 6 visualizes the difference.



(a) Part of the Twin-tower Concatenate architecture (b) Part of the Twin-tower Subtract architecture

Figure 6: The two different Twin-tower structures after the Convolutional blocks

Xu et al. (2019) show a significant improvement in performance when using the TTS architecture compared to the TTC strucure. They hypothesize that this is because the TTS structure ensures that differences in satellite images are more directly captured. For the extension, this paper will therefore train three additional models, specifically for income. The first one is a levels model, after which a TTC model and a TTS model are trained that use the levels model. Apart from the difference outlined before, the rest of the architectures and the hyperparameters are all the same for comparability's sake. Khachiyan et al. (2022b) mention on their project GitHub page

that they "would estimate that re-training of all of our models would take 1 month or more if run on a modern super-computer level cluster" (Thomas, n.d.). Hence, all models use a simplified structure and use a random subset of just under 5 percent of the original training data, due to the limited computational resources that were available. More specifically, the number of nodes in the dense network is quartered in each layer. Furthermore, there is no hyperparameter tuning — the same values are taken as used by Khachiyan et al. (2022b) — and a total of 25 epochs were executed, compared to 200 for the levels model and 100 for the differences models in the authors' research. An overview of the hyperparameters is shown in Table A3.

4.4 Evaluation of Predictions

Khachiyan et al. (2022b) evaluate the different models by simply considering the standard R^2. In the case of highly non-linear CNN models, the authors use the general formula given by $R^2 = 1 - \frac{SSR}{TSS}$. SSR is the sum of squared residuals and TSS is the total sum of squares, given by the sum of squared differences between the true value and the mean true value. The R^2 captures the degree to which variation in the inputs, in this case the vectorized extracted features and the initial conditions, can explain the variation in the observed outcome variable.

4.5 RGB Only Models

Since the training of the Neural Networks takes a considerable amount of time. Khachiyan et al. (2022b) consider the performance of the CNNs when using not only the full seven spectral bands of each satellite image, but also the three visible ones (red, green, and blue; or RGB). This greatly lowers the number of input nodes that are fed into the dense network, and hence the model complexity. The authors examine whether the added spectral information of the non-RGB data justifies its extended computational running time. Hence, also this thesis will look at whether the models trained on RGB only images perform reasonably well.

4.6 Comparison of Daytime Vs Nighttime Satellite Imagery

As mentioned in the Literature Review section, the field of economics has mainly focused on using nighttime satellite imagery in the prediction of GDP, since the degree of nighttime lights has been found to be a good proxy for economic activity. To show the potential of machine learning, Khachiyan et al. (2022b) compare their own CNNs, which use daytime satellite images, to linear models based on nighttime images. For this, they construct a linear model based on the testing data using Ordinary Least Squares (OLS) and then compare the R^2 values. More specifically,

the structure of the linear model is given by:

$$y_{i,j} = \alpha_{i,j} + \beta_{i,j} x_i + \varepsilon_{i,j}, \quad j \in \{INC,\ POP\}, \tag{5}$$

where $y_{i,j}$ denotes either log income ($j = INC$) or log population ($j = POP$) of image i, and x_i denotes the log night light intensity of image i. The night light intensity of an image is calculated using data collected by the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS). This is a sensor that is onboard a series of satellites that captures information on the intensity of lights emitted during the night across different regions. More specifically, x_i is taken as the log of the spatial sum of DMSP-OLS average visible light in both 2000 and 2010, i.e., the summation of the average visible light values of 2000 and 2010 for each pixel. The higher x_i is, the more light is emitted in the picture during the nighttime.

5 Results

This section provides an overview of the results obtained through the estimation of the different models described in the Methodology section. Initially, the outcomes derived from the predictions based on the trained models developed by Khachiyan et al. (2022b) are presented. Subsequently, a comparative analysis is conducted between these results and those obtained from the implementation of linear models based on nighttime satellite images. Lastly, the results from the extension part of this thesis are shown, where the newly trained models are evaluated against the benchmark models trained by the authors in the original research.

5.1 Replication Part

Table 1 presents R^2 values for the different models trained by Khachiyan et al. (2022b). More specifically, there are sixteen different models that are shown: one levels model and one difference model for each subcategory (income or population) and also including or excluding initial conditions. As a robustness check, the authors also trained models based on images with only three channels, the red, green and blue spectral bands. This doubles the number of trained models and are presented in panel B. Starting at panel A, the levels models are first considered. For income and population, and including initial conditions, the R^2 values in the test set are on average around 0.911 and 0.933, respectively. For the images from the year 2000, the performance is slightly better, especially when including initial conditions. This is possibly due to the fact that these features are measured for the year 2000 and hence more directly capture the local economic conditions at the time. Without initial conditions, the average R^2 values

Table 1: R^2 values for baseline models of large (2.4 km^2) images in the test set

	Levels		Difference
	2000	2010	2000 to 2010
Panel A. All seven bands included			
Income			
With initial conditions	0.913	0.908	0.441
Without initial conditions	0.850	0.844	0.433
Population			
With initial conditions	0.933	0.932	0.501
Without initial conditions	0.889	0.891	0.552
Panel B. RGB bands included			
Income			
With initial conditions	0.867	0.862	0.403
Without initial conditions	0.748	0.750	0.332
Population			
With initial conditions	0.894	0.895	0.441
Without initial conditions	0.798	0.815	0.384

drop to 0.847 and 0.890, corresponding to a decrease of 0.064 and 0.043, respectively. Including local economic characteristics hence seems to significantly improve the predictive performance of the models. Comparing these values to those in panel B, it is clear that having more complete spectral imagery has a sizeable positive impact on model performance. The average R^2 values for income and population levels including initial conditions are now 0.865 and 0.895, indicating a decrease of 0.046 and 0.038, respectively. This reduction is even larger when omitting initial conditions, namely 0.098 and 0.088.

Shifting the focus to the predictions for changes over the 2000 to 2010 time period, the R^2 values for income and population growth in the test set, when accounting for initial conditions, are 0.441 and 0.501, respectively. Conversely, without the inclusion of initial conditions, the R^2 values become 0.433 and 0.552. It is now less clear whether initial economic conditions really enhance the predictive performance, considering the R^2 for population growth is higher when omitting them. To get a sense of whether this difference is actually significant, the same models should be trained for many different seeds. However, due to the size of the dataset and the complexity of the CNNs, training all of the different models many times was practically infeasible. Comparing these values to those in panel B, a substantial decrease in performance can again be observed for models based on RGB only pictures. The R^2 values are now equal

to 0.403 and 0.441 when including initial conditions, and 0.332 and 0.384 without. This is a decline of 0.038 and 0.060 with conditions, and a striking 0.101 and 0.17 without. This shows that, for the RGB only models, the exclusion of initial conditions seems to significantly deteriorate model performance, whereas for the models based on all seven bands, this does not seem to be the case. This might be due to the fact that for images with only three channels, the initial conditions add relatively a lot more information than for the models based on all seven channels. The additional bands may have already absorbed a large part of the information that is given by the initial conditions, decreasing the need for the conditions in those models.

To assess the accuracy of the models' predictions, Figure 8 presents scatterplots illustrating the relationship between the model's predicted values and the actual values for log income and population levels in 2010, as well as for time differences. The better the predictions, the closer the data will lie around the 45 degree line. For the levels models, the data points are tightly

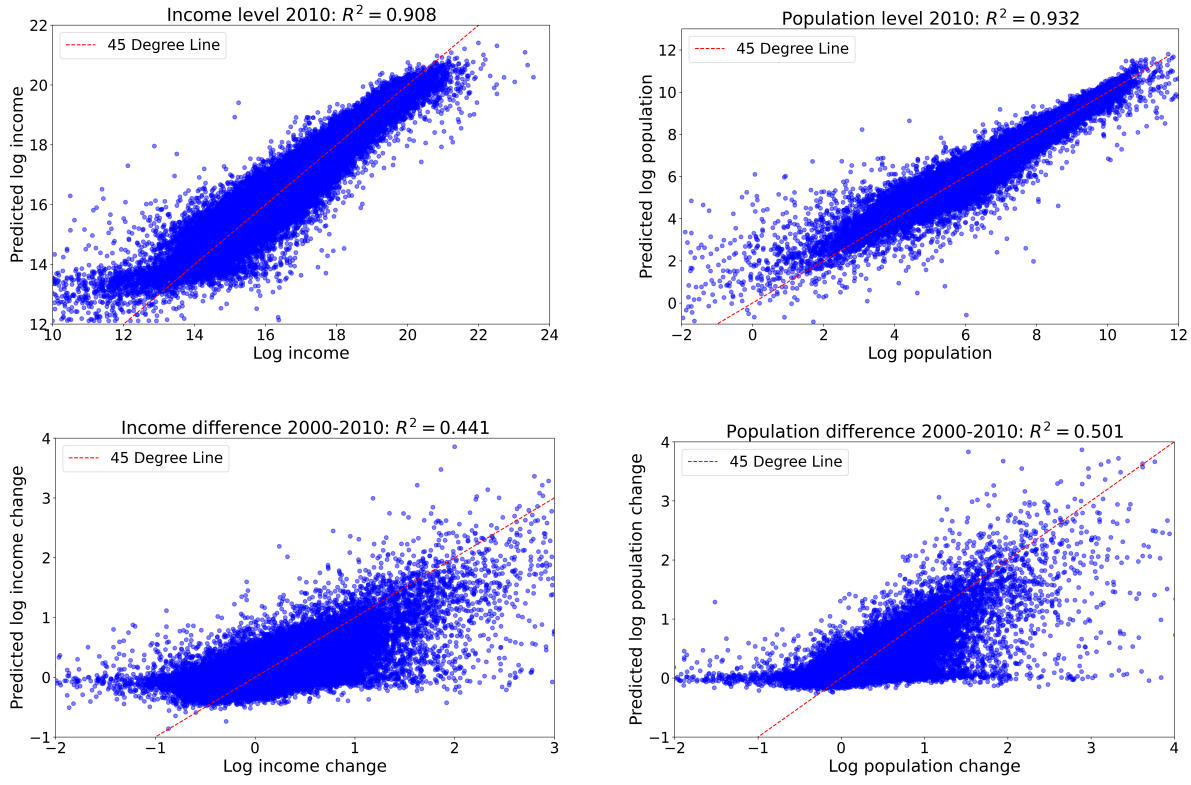


Figure 7: Model predictions against actual values

scattered around the 45 degree line, suggesting a high level of accuracy in capturing the log income and population. The most variation is observed for lower values of log income or population, where predicted values seem to differ from the actual values more than on average.

This observation could be attributed to the fact that the CNN encounters greater challenges in predicting log income and population when analyzing images with relatively minimal urban development. The limited visual cues available in such scenarios may complicate the prediction process.

For the difference models, there is much more variation in the data points. The models find it especially difficult to estimate decreases in log income or population. For actual values of log income or population that are negative, both models tend to give predictions that are around zero or higher. According to Khachiyan et al. (2022b), this phenomenon could be attributed to the gradual depreciation of physical capital. In expanding regions, income growth often translates into direct investments in new construction projects. However, for declining regions, the loss of income may trigger a slower process of structural changes or even the removal of buildings only in the long run.

5.1.1 Comparison of Daytime Vs Nighttime Satellite Imagery

Figure 8 shows the R^2 values for the CNNs based on daytime satellite imagery, as also presented in Table 1, as well as for the linear models based on nighttime imagery. The figure strikingly

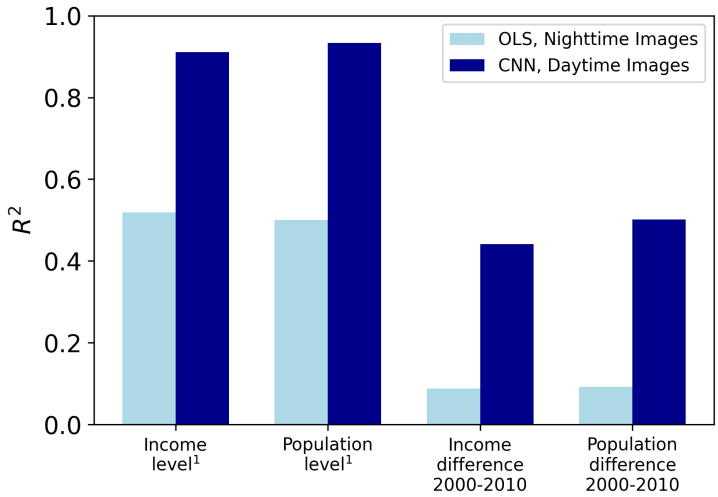


Figure 8: R^2 values of the OLS models based on nighttime satellite images and the CNNs based on daytime satellite images and including initial conditions, side by side

[1]Pooled for the years 2000 and 2010.

shows the benefit of using Neural Networks. For the level models, the R^2 for log income and population increase from 0.519 and 0.500, respectively, to 0.911 and 0.933 when using the models including initial conditions. That is an improvement of 76 and 87 percent, respectively.

For the difference models the relative improvement is even greater. The R^2 values of the linear models are 0.088 and 0.092, respectively, and increase to 0.441 and 0.501. This corresponds to an improvement of 301 and 345 percent.

The CNNs hence seem to be able to more accurately capture patterns on the satellite imagery that are indicative of economic development. Especially for smaller geographies, where luminosity is less accurately captured due to light pollution surrounding very bright city centers, this new approach seems to be promising (Khachiyan et al., 2022b).

5.2 Extension Part

Turning to the results of the extension component of this research, the R^2 values of the newly trained models are presented in Table 2. Again, only the models incorporating initial conditions were considered for analysis. What becomes obvious is that the R^2 of the newly trained models are substantially lower than those trained by Khachiyan et al. (2022b). This makes sense due

Table 2: R^2 values for the different income models, all including initial conditions, based on (2.4 km^2) images in the test set

		Extension		
	Benchmark (Khachiyan et al.)	Levels Model	TTC Architecture	TTS Architecture
Income Level[1]	0.911	0.789		
Income Difference 2000-2010	0.441		0.382	0.369

[1]Pooled for the years 2000 and 2010.

to the lower complexity of the newer models, the smaller dataset they are trained on, and the reduced number of epochs that were performed. The R^2 value of the level model for log income is about 13 percent lower than the model used in the original research. Similarly, the R^2 value of the TTC model is also 13 percent lower than the difference model trained by Khachiyan et al. (2022b). For the augmented difference model, which uses the TTS structure, the R^2 is even 16 percent lower. Hence, the hypothesis that the TTS architecture would perform better than the TTC architecture also used by the authors cannot be confirmed. On the contrary, the observed difference of 0.013 in the R^2 values implies the opposite, namely that the TTC architecture is optimal. However, it is important to note that the models were trained on a limited subset of the

complete dataset, and the number of epochs was restricted to 25. Thus, it might very well be that a properly trained TTS model could outperform the TTC model. During the training process, the mean absolute squared error (MSE) was still showing significant signs of improvement when training was halted after 25 epochs for all three models. Increasing the number of epochs to 100 or 200 might have shown a big difference in R^2 of the two architectures. Unfortunately, limited computational resources, even after paying for a faster GPU through an online Python environment, held back the possibility to perform a higher number of epochs and to empirically verify the hypothesis. Furthermore, there was no hyperparameter tuning performed. This might have greatly boosted the performance of the models and might have altered the ranking of the R^2 values.

6 Conclusion & Discussion

This research has explored the potential of machine learning, specifically Convolutional Neural Networks, in predicting log income, log population, log income growth, and log population growth at a microspatial level. It expanded upon the research done by Khachiyan et al. (2022b) by examining a different structure for the model predicting log income growth. Through the replication and extension of their research, this thesis has demonstrated the value and relevance of incorporating advanced algorithms and large-scale datasets in economic analysis. The findings of this research show that satellite imagery is a wonderful source of economic information that can enhance, together with the power of Neural Networks, economic forecasting. The main shortcoming of the research done is that the computational power needed to properly train the models was lacking. Hence, the actual performance of the newly trained models was rather poor, with exception of the levels model, and the hypothesis that a TTS structure would outperform the TTC structure in forecasting economic growth could not be substantiated. If the models' hyperparameters were tuned, the number of epochs were increased, and the number of nodes in each layer in the dense network were increased, the performance of the three models might have been substantially better. Moreover, the hypothesis that the TTS architecture holds much promise could have been more rigorously evaluated. An encouraging aspect is that despite the simplified structure and training process of the models, they still significantly outperform the simple linear regression models based on nighttime satellite images. This again underscores the great benefits machine learning can offer in the field of economic forecasting.

6.1 Suggestions for Future Research

As mentioned before, the limited computational capacity hindered a thorough comparison between the TTC and TTS architectures for the differences models. In light of this, future research should prioritize the thorough training of both the TTC and TTS architectures, as well as potentially exploring alternative architectures, to determine the optimal model for GDP forecasting. In contrast to the previous approach, an alternative strategy would be to prioritize limiting the dataset size for efficient training within a limited timeframe. This can for example be achieved by exclusively using satellite images from a specific state. This thesis took a subsample of the full dataset covering the entire United States but did not specifically take images from one specific region. By reducing the sample size while focusing on a particular state for example, the Neural Network can potentially adapt better to the specific characteristics of that area, leading to better results. Moreover, expanding the scope or type of data, for example by looking at different countries, may yield interesting results. It might be worthwhile looking at whether the models trained on US data perform just as well on satellite imagery from different countries, especially for emerging countries. This is because the need for accurate economic forecasting is the greatest in these regions. Other applications where researchers could expand on this thesis could be in the evaluation of policies aiming to improve economic conditions in undeveloped regions, the evaluation of transport infrastructure, and the analysis of natural disasters (Khachiyan et al., 2022b). The latter becomes more and more relevant in light of increased frequency of extreme weather events due to climate change. Leveraging the CNNs used in this paper may help localize regions where a natural disaster has the most detrimental effect on income, and which regions should therefore be targeted by policymakers. Similarly, researchers in the finance industry can examine whether a portfolio with assets from certain markets deemed attractive by the improved model, as also mentioned in the Literature Review, can outperform other portfolios in out-of-sample periods.

In conclusion, it is evident that there are several promising avenues for future research that could deepen our understanding of CNNs, and uncover novel, economic insights by implementing the models in reality. By addressing the outlined suggestions, future researchers can continue to contribute to the body of the rapidly growing field of machine learning, which holds great potential for the benefit of society.

References

Devadas, S., Guzman, J. P., Kim, Y. E., Loayza, N. V., & Pennings, S. M. (2019). *Malaysia's Economic Growth and Transition to High Income: An Application of the World Bank Long Term Growth Model (LTGM) (English)* (Tech. Rep. No. WPS 9278). Washington, D.C.: World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/844761592242576533/Malaysias-Economic-Growth-and-Transition-to-High-Income-An-Application-of-the-World-Bank-Long-Term-Growth-Model-LTGM

Doll, C. N., Muller, J.-P., & Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, *57*(1), 75-92. Retrieved from https://www.sciencedirect.com/science/article/pii/S0921800905001254

European Central Bank. (2022, May). Climate-related risks to financial stability. *Financial Stability Review*. Retrieved from https://www.ecb.europa.eu/pub/financial-stability/fsr/special/html/ecb.fsrart202205_01~9d4ae00a92.en.html

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223-2273. Retrieved from https://doi.org/10.1093/rfs/hhaa009

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning*. New York, NY, USA: Springer. Retrieved from https://link.springer.com/book/10.1007/978-0-387-84858-7

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Jeong, H. (2017). *Korea's Growth Experience and Long-Term Growth Model (English)* (Tech. Rep. No. WPS 8240). Washington, D.C.: World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/721471510252769303/Koreas-growth-experience-and-long-term-growth-model

Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A., Rosing, T., & Khandelwal, A. (2022a). *Data and Code for: Using Neural Networks to Predict Micro-Spatial Economic Growth.* Nashville, TN: American Economic Association [publisher], 2022. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-11-21. https://doi.org/10.3886/E158002V1.

Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A., Rosing, T., & Khandelwal, A. K. (2022b). Using Neural Networks to Predict Microspatial Economic Growth. *American Economic Review: Insights*, *4*(4), 491–506.

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., & Ruggles, S. (2020). *IPUMS National Historical Geographic Information System: Version 15.0 [Database].* http://doi.org/10.18128/D050.V15.0. Minneapolis, MN: IPUMS.

Okaidat, A., Melhem, S., Alenezi, H., & Duwairi, R. (2021). Using Convolutional Neural Networks on Satellite Images to Predict Poverty. In *2021 12th International Conference on Information and Communication Systems (ICICS)* (p. 164-170). doi: 10.1109/ICICS52457.2021.9464598

Pennings, S. (2017). *Long Term Growth Model (LTGM v4. 1)-Model Description.* World Bank Group. Retrieved from https://thedocs.worldbank.org/en/doc/133191589476085869-0050022020/original/ModelOutlineV43.pdf

Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D., & Ermon, S. (2017). Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning. In *31st Conference on Neural Information Processing Systems.*

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, *70*(1), 65–94.

Sutton, P., Roberts, D., Elvidge, C., & Baugh, K. (2001). Census from Heaven: An estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, *22*(16), 3061-3076. Retrieved from https://doi.org/10.1080/01431160010007015 doi: 10.1080/01431160010007015

Thomas, A. (n.d.). *Estimate of model re-training time.* GitHub. (Retrieved from https://github.com/thomas9t/spatial-econ-cnn.git)

U.S. Census Bureau. (2020). *Longitudinal Employer-Household Dynamics (LODES) [Database].* [Database]. Retrieved from https://lehd.ces.census.gov/data/lodes/LODES7/

US Geological Survey. (n.d.). *Landsat 7.* https://www.usgs.gov/landsat-missions/landsat-7.

World Bank Group. (2023, April). *The Long Term Growth Model - World Bank Group.* World Bank Group. Retrieved from https://www.worldbank.org/LTGM

Xu, J. Z., Lu, W., Li, Z., Khaitan, P., & Zaytseva, V. (2019). Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks. In *33rd Conference on Neural Information Processing Systems.*

A Appendix

A.1 Tables

Table A1: Local economic characteristics included in some of the models

Variable	Description
Female	The share of females among the residents living within the boundaries of the image
Emp in Business Services	The share of people employed in the business services
Emp in Accommodation & Food Services	The share of people employed in accommodation and food services
Emp in Wholesale Trade	The share of people employed in wholesale trade
Log Income, County	The log income of the county shown in the image
Emp in Administrative/Support/ Waste/Remediation Services	The share of people employed in administrative, support, waste, or remediation services
Emp in Production, County	The share of people of the county shown in the image employed in production
White	The share of people that is white
Emp in Non-Business Services	The share of people employed in non-business services
Log Population, County	The log population of the county that is partly shown in the image
Hispanic	The share of people that is hispanic
Emp in Business Services, County	The share of people of the county in the image employed in business services
Emp in Construction	The share of people employed in construction
Emp in Professional/Scientific/ Technical Services	The share of people employed in professional, scientific, or technical services
Emp in Real Estate, Rental & Leasing	The share of people employed in real estate, rental and leasing
Emp in Production	The share of people employed in production
Emp in Finance & Insurance	The share of people employed in finance and insurance
Emp in Non-Business Services, County	The share of people of the county shown in the image employed in the non-business services
Emp in Public Administration	The share of people employed in public administration
Black	The share of people that is black
Emp in Information	The share of people employed in information
Emp in Transportation and Warehousing	The share of people employed in transportation and warehousing
Group Quarters	The share of people living in shared spaces
Emp in Mining/Quarrying & Oil/Gas Extraction	The share of people employed in mining, quarrying and oil and gas extraction
Emp in Manufacturing	The share of people employed in manufacturing
Emp in Retail Trade	The share of people employed in retail trade
Emp in Agriculture, Forestry, Fishing, & Hunting	The share of people employed in agriculture, forestry, fishing, and hunting
Emp in Arts, Entertainment & Recreation	The share of people employed in arts, entertainment, and recreation
Emp in Health Care & Social Assistance	The share of people employed in health care and social assistance
Emp in Management	The share of people employed in management
Emp in Utilities	The share of people employed in utilities
Emp in Educational Services	The share of people employed in educational services
Emp in Other Services	The share of people employed in other services
Working Age	The share of people being part of the labor force

Note. All features in the dataset capture the initial values, with employment data measured at the image-level in 2004, and the remaining features measured at the image-level in 2000.

Table A2: Hyperparameters used in the different models by Khachiyan et al. (2022b)

	Learning Rate	L2 Regularization	Batch Size	Number of Epochs	Number of Filters	Dropout Rate
Panel A. All seven bands included						
Income level						
With initial conditions	1e-4	1e-6	16	200	32	0.5
Without initial conditions	1e-4	1e-8	16	200	32	0.5
Income growth						
With initial conditions	1e-5	1e-8	16	100	32	0.5
Without initial conditions	1e-4	1e-6	16	100	32	0.5
Population level						
With initial conditions	1e-4	1e-7	16	200	32	0.5
Without initial conditions	1e-4	1e-7	16	200	32	0.5
Population growth						
With initial conditions	1e-5	1e-8	16	100	32	0.5
Without initial conditions	1e-4	1e-6	16	100	32	0.5
Panel B. RGB bands included						
Income level						
With initial conditions	1e-4	1e-6	16	200	32	0.5
Without initial conditions	1e-4	1e-8	16	200	32	0.5
Income growth						
With initial conditions	1e-4	1e-8	16	100	32	0.5
Without initial conditions	1e-5	1e-8	16	100	32	0.5
Population level						
With initial conditions	1e-4	1e-7	16	200	32	0.5
Without initial conditions	1e-4	1e-7	16	200	32	0.5
Population growth						
With initial conditions	1e-4	1e-8	16	100	32	0.5
Without initial conditions	1e-5	1e-8	16	100	32	0.5

Table A3: Hyperparameters used in the trained models

	Learning Rate	L2 Regularization	Batch Size	Number of Epochs	Number of Filters	Dropout Rate
Income level model, TTC model, TTS model	1e-4	1e-6	16	25	32	0.5

A.2 Data and Code

Next to the code retrieved from Khachiyan et al. (2022a) and the code specifically written for this thesis, an explanation to the code can be read in the readme.rtf file also provided in the zip folder.