iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XI **Month of publication:** November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32280

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Poverty Prediction of Nigeria by using Convolutional Neural Network with Combination of Satellite Image

Mustapha Ibrahim Danbirni[1], Ren Dongxiao[2]
[1, 2]School of Applied Statistics, Zhejiang University of Science and Technology

Abstract: While it is important that local poverty has been targeted to help predictionand its policies have been established in emerging economies, this paper assesses the potential of features of local satellite images of high resolution that accurately presents poverty and economic well-being, with a combination of convolutional neural network (CNN). The properties of items and clothing are disbursed from Nigerian satellite images, which are used to assess poverty rates in Abuja and other regions. It includes the properties and density of buildings, shadow areas, which are the type of building height, spread, number of cars, density and length of roads, agricultural land and roofing materials. Buildings, shadows and road features have a strong relationship with poverty. Both application examples estimate estimates of the adjoining areas and estimate poverty in local areas using an artificially low census, confirming sample prediction capabilities other than. We have indicated that high resolution local images have the ability to change the estimate of poverty in small spaces, which have the potential for better design of surveys, and use of acquired features presents significant benefits for the methods that use satellite images to predict poverty. In this article, cconvolutional neural networks (CNN) directly assess poverty in high- and medium-resolution satellite images. We have come to the conclusion that CNN's estimated poverty can be created end-to-end in satellite images, but much work needs to be done to understand how the educational process affects audit patterns.
Keywords: convolutional neural network, poverty prediction, CNN, satellite image, machine learning

I. INTRODUCTION

In recent years, thanks to the continuous development and progress of convolutional neural network (CNN) technology, more and more scholars have begun to use CNN to conduct research on target detection methods and have made considerable progress. The current CNN-based target detection methods are mainly divided into two categories. One is regression-based target detection methods, represented by algorithms such as [1], etc. The detection speed of one class of methods is very fast, but the detection accuracy is average; the other is a target detection method based on candidate regions, with R-CNN, Fast R-CNN, R-FCN, Faster Representatives such as R-CNN have high detection accuracy but slow detection speed [2].

As the number of layers of the convolutional neural network increases, the model fitting ability is gradually enhanced, and the recognition effect is better [3]. But the deeper the model, the more difficult it is for training to converge, and the training efficiency is low; conversely, when the number of network model layers is relatively small, the model's fitting ability will decrease and the recognition rate will also decrease [4]. In addition, for complex models, in addition to extremely low training efficiency, model theoretical analysis is more difficult. Training often relies on rich parameter tuning skills and experience to conduct multiple experiments to obtain results. In the task of processing image classification with convolutional neural network, the normal operation is to preprocess the image first, and then input the entire image into the convolutional neural network model and then output the classification. However, because the category targets of different images may have regional differences, and different categories may have extremely similar feature information, the accuracy of image classification only through global features is not high [5].

Deep learning [6] is a new method derived from the study of artificial neural networks in machine learning algorithms. Its concept was proposed in 2006. It contains a variety of algorithms, such as automatic encoder, deep belief network [7] and convolutional neural network. Convolutional neural network classification and recognition algorithm was first proposed by [8] and applied to handwritten font recognition. Convolutional neural networks mainly use the idea of weight sharing to reduce the complexity of network learning.

In recent years, significant advances in computer vision research and the increasing availability of geo-measurement have enabled new methods of estimating socio-economic indicators. To address the problem of poverty prediction we proposed using convolutional neural network techniques with satellite image as fast, low-cost, and scalable means of providing granular poverty

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

estimates in Nigeria [9]. In this study, we look at the extent to which geospatial data, including day lights, daytime satellite images, human installation data and crowd information, can be used to thrive socio-economic prosperity in the Nigeria.

II. PRINCIPLE OF CONVOLUTIONAL NEURAL NETWORK

In recent years, with the unremitting development of deep learning algorithms, convolutional neural networks has engrossed the attention of researchers and scientists. The concept of Convolutional Neural Networks (CNN) is a unique network structure discovered, when they were studying neurons in the 1960s . Its unique features is that the convolutional neural network can effectively reduce the complexity of the network [10]. Convolutional neural networks can well extract the key features of input data through convolution and pooling operations. Therefore, convolutional neural networks have received extensive attention and applications in the field of pattern classification. The mathematical expression of the CNN convolution process is as follows:

$$(i,j) = (W \times X)(i,j) + b = \sum_{k-1}^{n\_in} X_k \times W_k)(i,j) + b \tag{1}$$

In the above formula; (n_in) is the number of matrices composed of input data, and (X_k) is the k-th input matrix. (W_k) is the k-th sub-convolution kernel matrix in the convolution kernel we selected during the convolution process. s(i, j) It represents the value of the element at the corresponding position of the convolution kernel (W) its output matrix [11].

The basic structure of convolutional neural network includes convolutional layer, pooling layer and fully connected layer. The convolutional layer mainly performs convolution operations.

The methods used are mainly local connections and weight sharing methods, mainly to simulate cells with local receptive fields in the brain, so as to extract some information from the obtained information. The pooling layer mainly performs down-sampling operations, including methods such as maximum pooling and average pooling.

After the input data is down-sampled by the pooling layer, the output data matrix will become smaller, but the number remains unchanged, so the pooling layer can compress the data output from the convolutional layer of the previous layer, so Reduce the complexity of calculation, so that the number of learning parameters is reduced and at the same time can effectively prevent over-fitting problems.

In the convolutional neural network model [12], the last layer or layers is the fully connected layer, The main function of the fully connected layer is to perform a weighted summation of the features extracted by the previous convolution and pooling operations to ensure that the input data is in The few data features retained after the pooling operation can reproduce the original input data as much as possible [13].

In Fig. 1 'input' is the data input into the convolutional neural network in the form of a matrix. The input data is subjected to convolution operation to obtain the C_1 layer and the data of the C_1 layer is down-sampled to obtain the S_1 layer. The convolutional neural network is completed. The first convolution and downsampling operation, the output of the first convolution down-sampling operation is used as the input of the next convolution operation, and the second convolution and down-sampling operations are performed and so on, until the last convolution operation is completed. After convolution and down sampling, the obtained features are fully connected and used as the output of the convolutional neural network to extract the features and the output data is input into the classifier for classification and the classification results, we need are obtained [14].

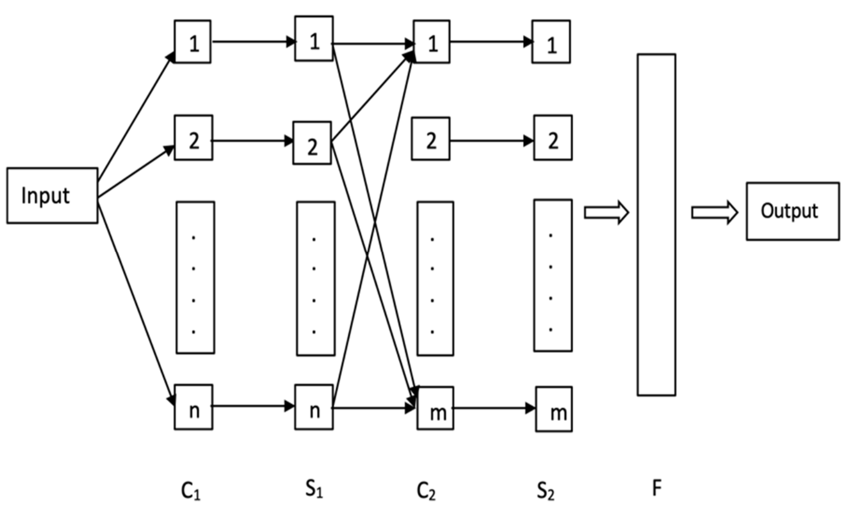

Fig. 1 Block diagram of principle of convolutional neural network

A. Multi-Region Deep Convolutional Neural Network

A deep convolutional neural network (DCNN) is generally composed of three neural network layers: convolutional layer, pooling layer, and fully connected layer, and an output layer [15]. The image is generally convolved through a convolutional layer, and several feature maps are obtained, and then the feature maps are compressed by a pooling layer, and finally output through a fully connected layer to identify image features, and then classify the image with a classifier. Although the traditional convolutional neural network uses parallel calculation between the convolution and pooling of each feature map, the entire network is completed in a single path, and the network only expands in depth, which is likely to cause gradient disappearance and overfitting [16].

1) *Guidelines for Regional Interception:* The original image is equally divided into four blocks, which are respectively labeled as the upper left block (LT), the upper right block (RT), the lower left block (LB), and the lower right block (RB) of the original image. Divide the original image into four evenly, which can focus on the local area of the image, and then better extract the local features of the original image, and improve the recognition of the local detail information of the image.
2) *Central Area Interception Standard:* Since most of the information of a general image is mainly concentrated in the center of the image, in addition to the average cropping of the original image into four parts to extract more local details of the image, the central area of an image is extracted to make the local information of the image The combination is more complete. At the same time, the central area has the same size as the sub-areas in the four directions.
3) *Model Design:* First, input the original image into the MR-CNN module and perform convolution operation with filters of different sizes. Each block image is input into MR-CNN module 2 for convolution operation. The size of the filter is 3×3 first because the size of the block image is smaller [17]. Using a smaller size filter can better extract each block The context feature information of the image, and the use of a 1×1 filter can greatly increase the nonlinear characteristics while keeping the scale of the feature map unchanged (that is, without loss of resolution), so that the network can extract more robust features. Then cascade the original image and the feature map after each block convolution, and then input it into MR-CNN module 3, which learns the contextual interaction features of the image in a way of information supplementation, and module 3 also uses two layers of 3×3 The filter adds a layer of 1×1 filter. This design is because the cascaded feature map is small and the number of channels is large. It is necessary to use a smaller size filter convolution to extract the features. The 1×1 filter can not only increase the nonlinearity of the feature, but also The role of dimensionality reduction. Finally, input the global average pooling layer and output the classification.

III. OVERALL STRUCTURE

This article is a target detection method based on poverty regions. The main feature extraction network first uses the convolutional layer and depth separable convolutional layer of the residual network ResNet-101 [18] to extract the feature map of the input image. The specific situation of each convolutional layer of ResNet As shown in Table 1 . The feature map generated by the last shared convolutional layer (Conv4) of ResNet-101 is input into the region extraction network RPN (Region Proposal Network) to obtain the feature map of the candidate region with classification labels and regression boxes, and then the RPN network the output feature map and the feature map obtained from the depth separable convolution layer are input to the PS-RoI Align layer for further pooling operation. At this time, the network will be divided into two parts, one is used to determine the category of the target object, and the other is used to determine the location of the target object.

Table I

Convolutional layer of ResNet

Layer name	Output size	101-layer
Conv1	112 × 112	7 × 7, 64, stride = 2 3 × 3 max pool, stride = 2
Conv2_x	56 × 56	$\begin{bmatrix} 1 & 1 & 64 \\ 3 & 3 & 64 \\ 1 & 1 & 256 \end{bmatrix} \times 3$
Conv3_x	28 × 28	$\begin{bmatrix} 1 & 1 & 128 \\ 3 & 3 & 128 \\ 1 & 1 & 512 \end{bmatrix} \times 4$
Conv4_x	14 × 14	$\begin{bmatrix} 1 & 1 & 256 \\ 3 & 3 & 256 \\ 1 & 1 & 1024 \end{bmatrix} \times 23$
Conv5_x	7 × 7	$\begin{bmatrix} 1 & 1 & 512 \\ 3 & 3 & 512 \\ 1 & 1 & 2048 \end{bmatrix} \times 3$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

A. Residual Network Structure

Residual network is a convolutional neural network model proposed by He Kaiming and others in 2015. It won the first place in image classification at the ILSVRC2015 competition. Compared with the traditional convolutional neural network, the residual network not only increases the depth of the network, but also effectively avoids the problems of gradient disappearance, gradient explosion and network generalization, and has a stronger generalization ability. In the field of target recognition, as the depth of the network increases [19]. Deep networks will be more difficult to train, and model accuracy will be affected. To solve this problem, the residual network model introduces residual structure and short-circuit connection, such asAs shown in Fig. 2 .

The characteristics of the residual network are asAs shown in Fig. 2 , a jump connection is added for every two layers. If the output is H(x), H(x) = F(x) + x. x is the identity mapping, and F(x) is the residual mapping. These two mappings provide two options to solve the problem of the decrease in model accuracy caused by the deepening of the network. When the accuracy of the model has reached the optimal value, the residual mapping F(x) will be set to 0, then the output of the network is only the identity mapping x, after which the network will maintain the optimal state, thus ensuring the accuracy of the network The rate will not decrease as the network depth increases. Using ResNet-101 can take advantage of its high network depth and extract rich feature information [20]. The depth of the separable convolutional layer can greatly reduce the dimensionality of the feature map, greatly reduce the amount of calculation, and reduce the memory footprint, thereby increasing the calculation speed while ensuring accuracy.

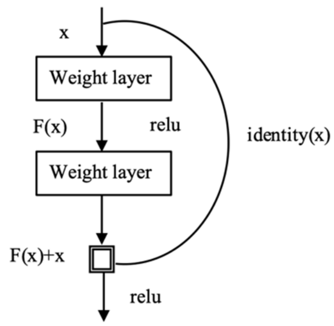

Fig. 2 Residual structure model

IV.METHOD

This article introduce the ability to capture a series of data representation often uses of convolutional neural network learning approaches in visual work. CNN includes complex layers in the neural network and their translational diversity and local co-dishonesty allow you to solve a variety of problems such as video validation and analysis [21]. CNN is a feed-forward artificial neural network that usually consists of four main types of layers: convoluous layer, RELU layer, pooling layer and fully assembled layer. Like standard neural networks, neurons are arranged in three dimensions in each layer of CNN: width, height and depth. In addition, the neurons of the twisted layer are not connected to all neurons in the old layer; Instead, each neuron is connected only to local areas, a com-produced mode point between the weights and the area to which they are connected. A ReLU set is put on a reluctance after a convolutional set, bringing the threshold for all activations to zero. a pair of partially common shallow samples that activate amid diversity and height, compressing space hardware. After dry repetition of [Conv-ReLU-Pool] layers, CNN models sometimes use fully connected layers as the final layers of the network. A fully connected layer calculates class points as in ordinary neural networks [22]. An example of a common Conv-Net architecture is as follows:

$$\text{INPUT} \rightarrow [[\text{CONV} \rightarrow \text{RELU}]*\text{N}] \rightarrow \text{POOL?}]*\text{M} \rightarrow [\text{FC} \rightarrow \text{RELU}]*\text{K} \rightarrow \text{FC}$$

His layered architecture allows CNNs to learn the different functions of images during exercise, low-level features in the earlier layers of the network and high-level functions later layers. The results of the last attached layer can be explained as the score type for each class, and the output of the second in the last attached layer can be used as a vector element for any other interest-bearing task.

A. Satellite Imagery

We used satellite image imparted by google maps live satellite imagery, examples of images are shown in Fig. 3. Given the corresponding daytime satellite imagery, we want to accurately predict nighttime light [23]. Static Map API provides images on a

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

purpose; we used 15, 10, level of zoom and have pixel level of resolution of about 1 meter. In this zoom level, 400-of-400 pixel images compatored with the 0.5km solution for daylight lights data. An example of daylight satellite image as shown in Fig. 3.

Fig. 3 Daylight satellite image of Abuja Nigeria

V. RESULTS

The task presented in this article proves that CNN applies satellite images of liberal resolutions available to predict poverty in Nigeria. The discovery reflects the amount of spatial information to predict development metrics, especially when processed by CNN-based strategies. While the use of the procedures presented in this article requires further research to predict non-independent environments and related policy applications, our results show that a possible CNN-based approach to applications is being applied beyond existing literature findings. As shown in previous computer vision work, visual functions can be used to predict socio-economic indicators. Here we try to predict changing income at different levels of aggression in different parts of Nigeria. One of the most striking discoveries of this task is the limited impact of different network parameters, not the modification of samples to training data. Due to the impact of sampling methods on the overall performance of the network, an increase in the range of studies comparing these techniques may lead to significant improvements. The use of repetitive training data is defined as the most effective training method, but it can provide the same benefits by increasing the number of delegates proportionally during practice. Comparing overlaps at different levels can provide insight into the impact on each strategy as the number of delegates increases proportionately. To find out how urban environment features affect final estimates, we can create and validate our models on different subsets of images, whether they belong to urban areas or not. In this section, each image is assigned to the urban area if its center falls in the shape of the area, but we noticed that even images with marginal overlap with the planting aftermath for urban areas (a less restrictive statement respecting the above) do not significantly change the results.

The first experiment relates to earnings estimates at the same city level, including two importantly-pects: which complex model gives more informative features than images, and how important the level of urbanization is in terms of regression performance.

These figures make sense, as day-to-day satellite images of two squares have relatively different characteristics and most images rank in the lowest class. Most mistakes occur when the real class is of the middle class. It also makes sense: the middle class has more features that overlap with each other class, and there are two directions in which an error (either low or high) can be created instead of one. For mistakes made in the lowest and highest sections, CNN generally seems to predict the middle class rather than the class at the other end of the spectrum. These figures seem reasonable again, as the middle class has pre-features in common with the lower and upper classes compared with each other.

Whether or not our transfer training model improves the use of night light to assess life, we have conducted several separate tests for each area of the country and for integrated models whenever the expected strength of our transfer training model is comparable to night light only.

We're also looking at whether our approach improves other, simpler ways to extract configuration from daytime images and predict economic results using available survey data. We find that our CNN feature extractor is absent, it performs common use general image features such as color terraces and ori-ented gradient pedestals. The Committee on Our Approach also leads to an intuitive or better approach than using data from previous surveys to predict the results of more recent surveys.

Finally, capitalizing on our survey-based measures of consumption and assets in multiple part of country, we study the extent to which a model trained using data and satellite image features from one country can estimate livelihoods in other countries. Examining whether a particular model generalizes across borders is useful for understanding whether accurate predictions can be made from imagery alone in areas with no survey data, an important practical concern given the paucity of existing survey data in many regions of Nigeria. As well as for gaining insighting commonalities in the determinants of livelihoods across country.

VI. CONCLUSION

In this work, we asked if the CNN model could assess poverty indication from urban housing and economic and social data, especially in the context of developed and non-developed regions of Nigeria. Our results show that CNN is optimistic about this work. In the current study, CNN samples created with layer depth were optimized for a set of hyperparameters, which offered the best predictive accuracy for our relatively small training data. Building codes taken from satellite images also proved to be the right input resources to assess poverty prediction using CNN, despite promising results achieved with shallow structures. But adequate training data finds itself a factor. Therefore, the accuracy of predicting multiple architectural scales with large training data is interesting. We have also proposed an alternative and cost-effective approach to poverty prediction, using free and publicly available information on local information related to the public. Our results include satellite images of daylight and night images of Abuja Nigeria.

REFERENCES

[1] A. Shafiee et al., "ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars," Proc. - 2016 43rd Int. Symp. Comput. Archit. ISCA 2016, pp. 14–26, 2016.

[2] Y. Zhang, J. Gao, and H. Zhou, "Breeds Classification with Deep Convolutional Neural Network," in ACM International Conference Proceeding Series, 2020, pp. 145–151.

[3] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017, 2018, vol. 2018-Janua, pp. 1–6.

[4] C. Zhao, B. Ni, J. Zhang, Q. Zhao, W. Zhang, and Q. Tian, "Variational convolutional neural network pruning," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, vol. 2019-June, pp. 2775–2784.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[6] Z. Gao and X. Wang, "Deep learning," in EEG Signal Processing and Feature Extraction, 2019, pp. 325–333.

[7] R. E. Neapolitan and R. E. Neapolitan, "Neural Networks and Deep Learning," in Artificial Intelligence, 2018, pp. 389–411.

[8] S. S. Mousavi, M. Schukat, and E. Howley, "Deep Reinforcement Learning: An Overview," in Lecture Notes in Networks and Systems, vol. 16, 2018, pp. 426–440.

[9] E. O. Akeredolu-Ale, "Poverty as a Social Issue: A Theoretical Note, in Poverty in Nigeria," Proc. Annu. Conf. Niger. Econ. Soc. (NES)., no. 06, 1975.

[10] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," in 30th AAAI Conference on Artificial Intelligence, AAAI 2016, 2016, pp. 3929–3935.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[12] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," IEEE Trans. Image Process., vol. 26, no. 9, pp. 4509–4522, 2017.

[13] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 9906 LNCS, pp. 391–407.

[14] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in Advances in Neural Information Processing Systems, 2014, vol. 2, no. January, pp. 1790–1798.

[15] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," Neurocomputing, vol. 323, pp. 37–51, 2019.

[16] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, vol. 07-12-June-2015, pp. 3367–3375.

[17] S. Goodman, A. BenYishay, and D. Runfola, "A convolutional neural network approach to predict non-permissive environments from moderate-resolution imagery," Trans. GIS, 2020.

[18] N. Pokhriyal and D. C. Jacques, "Combining disparate data sources for improved poverty prediction and mapping," Proc. Natl. Acad. Sci. U. S. A., vol. 114, no. 46, pp. E9783–E9792, 2017.

[19] S. Hallegatte et al., Shock Waves: Managing the Impacts of Climate Change on Poverty. 2016.

[20] A. Perez, C. Yeh, G. Azzari, M. Burke, D. Lobell, and S. Ermon, "Poverty prediction with public landsat 7 satellite imagery and machine learning," arXiv. 2017.

[21] B. Babenko, J. Hersh, D. Newhouse, A. Ramakrishnan, and T. Swartz, "Poverty mapping using convolutional neural networks trained on high and medium resolution satellite images, with an application in Mexico," arXiv. 2017.

[22] R. S. Thorat, P. C. Shetiye, and A. K. Gulve, "Poverty prediction using satellite imagery and machine learning," Int. J. Sci. Technol. Res., vol. 9, no. 2, pp. 4785–4789, 2020.

[23] P. Agarwal, N. Garg, and P. Singh, "Predicting poverty index using deep learning on remote sensing and household data," Int. J. Recent Technol. Eng., vol. 8, no. 3, pp. 164–168, 2019.