Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 1577–1584
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.7123

Using Densenet121 to extract roads from satellite images

Hussein Ali Al-Iedane

Basra University for Oil and Gas, Basra, Iraq

(Communicated by Javad Vahidi)

Abstract

The most significant issue in remote sensing is the extraction of roads from very high-resolution satellite (VHR) photos. The authors of this research provide an effective Densenet121 block that can understand relationships between global features. As a result, road segmentation is more precise due to the ability of each geographical version available as a pointer to other data that is collected. In the specific, our single model outperformed every other contemporary aggregation model that has been presented in the official Densenet121, our suggested model offers a shorter training convergence time, fewer parameters, and fewer Giga floating-point operations per second (GFLOPs). The authors also provide empirical evaluations on how non-local blocks should be used appropriately for the base model. Theoretically, the applied methodology provided a DenseNet-121 and an unpublished best alternative identification method DenseNet-121, a really sizable dataset provided by the CNN model's training phase. Additionally, because the implemented architecture takes advantage of data augmentation, no custom data extraction method is necessary.

Keywords: Densenet121, Road Extraction, image satellite, convolutional neural networks (CNNs)
2020 MSC: 68T07, 92B20

1 Introduction

Remote sensing applications heavily rely on Very high resolution (VHR) satellite images are used for ROAD extraction. The derived road network contains useful geographic information that is highly accessible. It can therefore be utilized for a variety of activities like automated map updates, unmanned vehicles, road navigation, urban planning, supporting disaster relief missions, or focusing on road safety hazards.

Road extraction techniques have been studied for many years. The Hough transform [9], and Canny edge detector [3] were used to start off the traditional line segmentation process. Eventually, all lines, including a specific amount of edge points, were extracted. These techniques took a lot of time and often produced false positives.

Unsupervised techniques utilizing clustering algorithms came next. Automatic approaches for road extraction were made possible by support vector machine (SVM) [16] and Markov Random Field (MRF) [17].

Convolutional neural networks (CNN) have made significant progress in recent years at identifying roads in VHR satellite pictures. For instance, by using end-to-end learning, [13, 21, 18, 12] and U-Net [23, 4, 25], which won the DeepGlobe 2018 Road Extraction Challenge [8], obtained significant gains in scene parsing.

However, it has been thought that CNNs' limited receptive fields present a difficulty. The problem is resolved by suggesting a novel approach of increasing the convolution's kernel size, which allows the dilated convolutions [6, 5, 7] to be somewhat enhanced with the bigger receptive field [24, 19] and [10].

Email address: hussein.aliedane@buog.edu.iq (Hussein Ali Al-Iedane)

Received: April 2022 *Accepted:* August 2022

The fundamental problem was that, even with dilated convolution, it was challenging to capture long-range correlations. Since local characteristics may be hidden by extra obstructions like shadows, clouds, trees, or buildings because satellite data is collected from above, the single convolution layer, depending on its kernel size, only considers a small number of neighbourhoods. Additionally, a deeper CNN design does not necessarily ensure a larger effective receptive field [14].

In order to address this issue, we suggest the Densenet121 network in this research. Recent research on the non-local neural network [20] and its applications for picture restoration [22] strongly influenced our point of view. The values of feature maps are computed by non-local neural processes as a weighted sum of the features across all places. Consequently, it makes it possible for the model to effectively account for long-range dependencies and capture distant information. The three primary contributions of this study are summarized above.

1. When taking part in the DeepGlobe Challenge, our single model outperformed any public ensemble model that has previously been released with no sophisticated post-processing. Densenet121 even surpassed within the official DeepGlobe Challenge, solution [25] was top-ranked [8] while utilizing 43 percent less parameters, fewer GFLOPs, and a quicker training convergence time.
2. The author provides empirical evaluations of road extraction activities. For every set of unique pairwise functions and non-local block locations, Densenet121 outperformed the baseline with more encouraging findings. In comparison to the baseline, block models that are single and double non-local performed better.
3. The pioneering work that used neural processes to extract roads from VHR satellite photos The procedures give the model the ability to record long-range dependencies to address geographic limits of roadways that are only partially covered with different elements, such as shadows, clouds, structures, or trees.

The structure of the paper is as follows: the first section is an introduction of the problem and contributions. The second part is the overview of the Densenet121. The detail of the densenet121 applied for road extraction in the third section. The experimental result and discussion and displayed in fourth section. The last section is the conclusion.

2 DenseNet Background Overview

In order to train robots or computers, Deep learning allows computers to classify and recognize data or images as the human brain does by learning from experience using complicated algorithms and artificial neural networks. Deep learning has become a popular method for analyzing huge data. A CNN, also known as an Among artificial neural network, convolutional neural networks are one type used in deep learning and are frequently employed for object and picture recognition. By employing a CNN, Deep Learning can therefore identify items in an image. CNNs are being used extensively in a wide range of activities and functions, including voice recognition in NLP and challenges with image processing, computer vision tasks such as localization and segmentation, video analysis, and identifying impediments for self-driving cars.

Due to their major contribution to these rapidly developing and expanding fields, CNNs are widely used in deep learning. The DenseNet (Dense Convolutional Network) design tries to increase the depth of deep learning networks while simultaneously enhancing training effectiveness by utilizing shorter connections between the layers. Every layer of a convolutional neural network called DenseNet is connected to every layer beneath it. Five levels of the DenseNet are depicted in figure 1.

CNN have issues when they delve further. This is due to the fact that the distance between the input layer and the output layer (and the gradient in the opposite direction) increases to the point where information may become lost before crossing to the other side.

DenseNets streamline the layer connectivity pattern introduced in prior architectures:

Networks of Highways [9]

Fractal networks [3]

residual networks [16]

By assuring maximum information (and gradient) flow, the authors are able to fix the issue. To do that, they merely link each layer directly to the next. DenseNets utilize the network's potential by recycling features as opposed to getting representational strength from extremely deep or broad designs.

Due to the previously described, It was difficult to train very deep networks because of information flow and gradients. This issue is handled by DenseNets since each layer has direct access to the gradients from the loss function and the original input image.

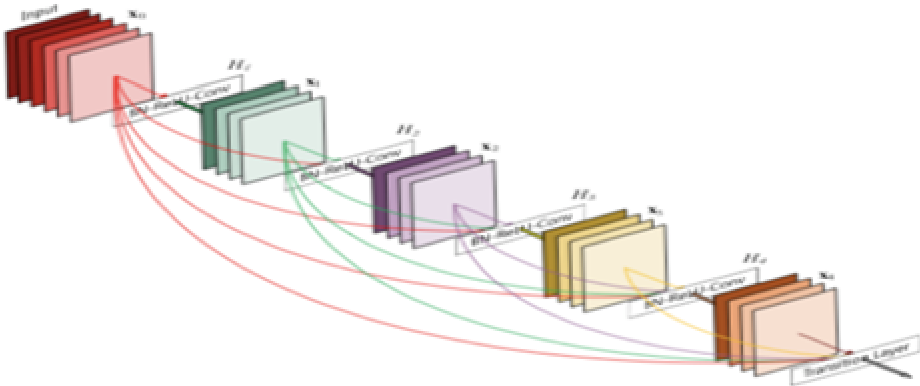


Figure 1: DenseNet with 5 layers with the expansion of 4. [8]

Following the application of traditional feed-forward neural networks connect the output of the layer to the following layer through a series of processes. However, the visualization becomes slightly more complicated than it was for VGG and ResNets due to the extremely dense amount of connections on DenseNets. Figure 2 depicts a very basic diagram of the DenseNet-121's design, which is the DenseNet that will be the subject of this paper. This is due to the fact that it is the most straightforward DenseNet created using the ImageNet dataset.

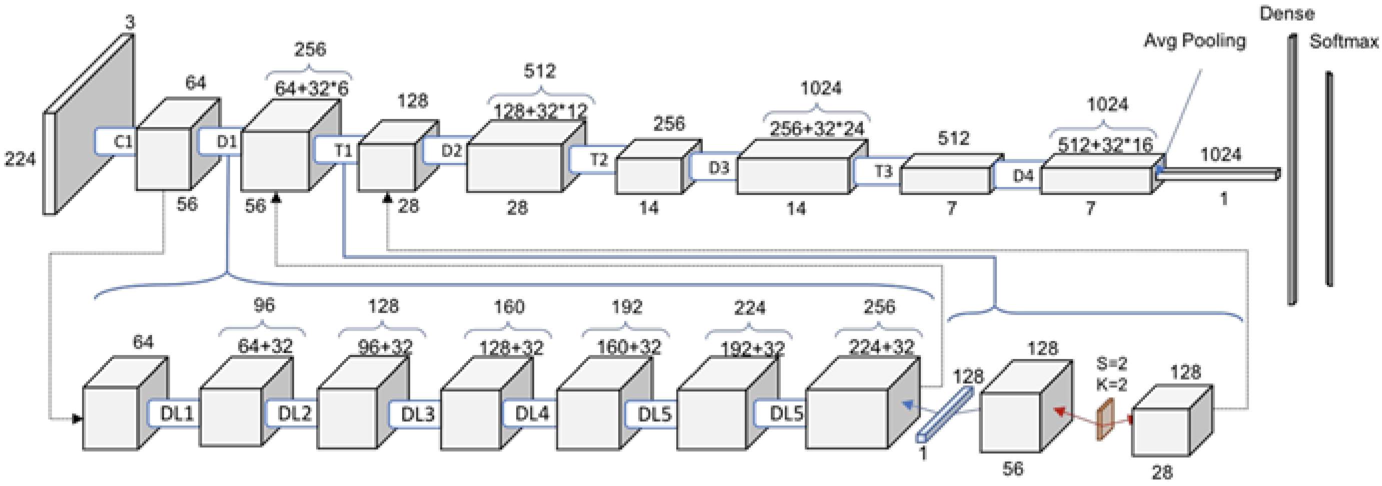


Figure 2: Dense Layer x (DLx): Dense Block and Transition Block. An analysis of DenseNet-121

View how this behavior of adding 32 times as many layers is actually carried out in the new deeper level, which represents the first Dense Layer inside the first Dense Block. Used to perform the authors' suggested 1x1 convolution using 128 filters to condense the image features, and then perform a more expensive 3x3 convolution with the 32 feature maps we've selected for the growth rate (Don't forget to apply padding to ensure that the dimensions stay the same.). The input volume and the output of the two procedures, which are identical for every Dense Layer inside of every Dense Block, are then concatenated in order to update the network's collective understanding.

3 EXTRACTION ROAD BY DENSENET121 NETWORK

Satellite images are typically obstructed by other objects since they are captured overhead at great heights. The likelihood that a route will be obstructed by another obstruction, such as the shadows cast by high buildings or roads covered by trees, is highest. Capturing long-range relationships is crucial for overcoming this problem, but almost all CNN approaches have trouble doing so. It is challenging to refer to distant information when using a convolution technique because it only refers to local information through a small kernel. Another local method, the recurrent operation, solely uses features from the past and present. Local approaches regularly generate these actions, but they have a number of disadvantages, including a lower memory and computation efficiency for the model optimization.

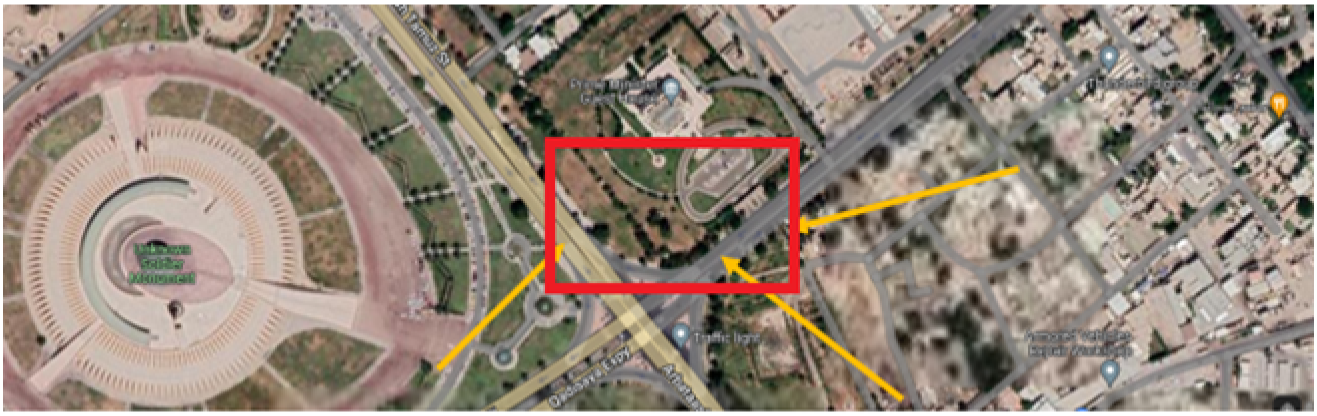
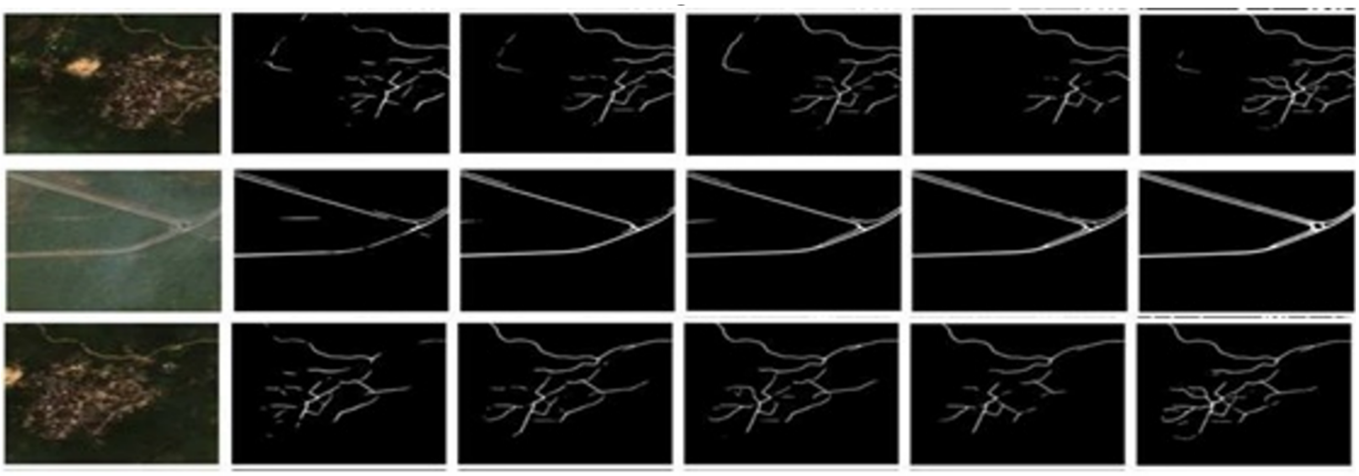


Figure 3: Road extraction from a VHR satellite image is shown in action. DenseNet operation must make reference to other roads in orange boxes in order to identify the road in the red box that is blocked by trees.

The long-range dependencies of the DenseNet operation [20] model were thus implemented. Fig. 3's centre road (red box) is hidden by trees, Nevertheless, a non-local operation calculates a depth map as the weighting factor of all pixels, allowing the model to recognize distant dependencies and information. The connections between important highways cannot be detected using local approaches. Non-local networks, on the other hand, make reference to (orange boxes) close roadways and appropriately extract the covered roads.



(a) Input image. (b) leyer1. (c) layer2. (d) layer3. (e) layer4 (f) layer5

Figure 4: Qualitative of the output layers in DenseNet121

3 transition layers (6,12,24), 1 classification layer (16), 5 convolution and pooling layers, and 2 dense Block layers make up the Densenet 121's architecture (1x1 and 3x3 conv). The DenseNet is separated into Dense Blocks, each of which has the same dimensions but differs in a number of filters. It is a crucial step in CNN that Transition Layer applies batch normalization via down-sampling. First, after BN algorithm and pre-activation structure of activation function, convolutional layer is simpler to train and performs better in terms of generalization. To achieve super-resolution images, a dense jump-connected network was used. Large sensory fields were generated by the dense unit using 3x3 convolution, and the network was deepened using 1x1 convolution to learn reliable feature representations. Finally, to maximize the utilization of the hierarchical feature deepening structure, simple filters with few parameters were utilized to deepen the network. A new method with robust fitting capabilities was outperformed by the network. Second, the convolutional layer does not employ pre-activated dense units prior to the activation and BN processes. Third, before activation function, BN, and ReLU may control the level of activation saturation, convolutional layer

sends BN-normalized data to the non-saturated zone. However, the outcomes might alter if the BN and ReLU are applied in a different order. Some BN neurons will become inactive as a result of ReLU, which causes BN instability and reduces the model's effectiveness. In reality, numerous application scenarios will take on diverse responsibilities within the hierarchy. Fourth, the convolutional layer comes before the activation function, and BN is more suited for denoising, but this produces a long super resolution convergence and considerable loss function volatility, increasing the memory and processing requirements.

4 EXPERIMENTS RESULTS AND DISCUSSIONS

4.1 Dataset and Evaluation Metric

They conducted the tests using the Road Extraction Challenge dataset by DeepGlobe 2018 [8], which allowed us to evaluate the effectiveness of road extraction. The dataset contains 512x512 pixels wide and tall photos. The satellite of DigitalGlobe acquired each image as an RGB image with a 0.7 ground sampling distance (GSD). The mask is an identically sized grayscale binary image as the input image. In the mask illustration, the white represents the highways and the black represents the backdrops.

Totaling 8,200 images, the dataset comprises of 6,560 photos for training, 820 images for validation, and 820 images for testing. Mean intersection-over-union (mIOU) was utilized by the author as the evaluation metric. The same evaluation metrics, including accuracy and mean square error, were used in DeepGlobe 2018 Road Extraction. The labeling of pixels for roads and non-roads is categorized to 255 or 0, respectively, in order to evaluate the models appropriately.

4.2 Pre-processing

The goal of the image pre-processing stage is to remove any undesirable twists from the image and shrink and normalize it for subsequent processing. According to the need for model development, there are many picture pre-processing techniques found in the prior literature. The most often utilized methods among these are image scaling, image normalization, and covert level to categorical. Using the "Pillow 2.7+ + " Footnote3 Python library, photos were resized for this investigation to guarantee the same size and pixel. In this study, image dimensions of (512 x 512) are taken into account. Additionally, we can alter the pixel intensity with image normalization to make the image appear more realistic. Typically, the majority of image pixels integrate values in the range of 0 to 255. However, because of how networks are built, it is preferable to execute all values between 0 and 1, since this will be a good fit for model creation. By using this method, the computational complexity of the model training process is reduced.

Table 1: Model Summary of DenseNet

Layer (type)	**Output shape**	**Param**
Model: "model_1"		
input_2 (Input Layer)	(None, 512, 64, 3)	0
conv2d_1 (Conv 2D)	(None, 512, 64, 3)	84
Densenet-121 (Model)	Multiple	6,560
global_average_pooling-2d_1	(None, 1024)	0
batch_normalization_1	(None, 1024)	4096
dropout_1 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 256)	262,400
batch_normalization_2 (Batch	(None, 256)	1024
dropout_2 (Dropout)	(None, 256)	0
root (Dense)	(None, 2)	524

Images were adjusted, though, using Eq. (4.1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4.1}$$

where $\mathrm{X_{min}}$ and $\mathrm{X_{max}}$ refer to the minimum and maximum pixel values.

4.3 Implementation details

Python 3.7 software Footnote4 and associated packages were used to implement the suggested DenseNet classification model. It utilised a 64-bit Windows operating system and an Intel(R) Core(TM) i7-4600M CPU @ 2.90GHz processor with 16 GB of main memory and 4 GB of NVIDIA GeForce 940MX graphics. Finding out whether or not an image is labeled as a road is the goal of road extraction identification. Data were divided into three categories before model building: training (80%), validation (10%), and testing (10 percent). Table 1 shows the model history, which was trained across 50 epochs. Fig. 3 proposes the used mechanism for route prediction. A confusion matrix, illustrated in table 2, can be used to summarize the results of samples of each category that were correctly and wrongly identified. Based on the confusion matrix, they can determine the accuracy (Eq. 4.2). Precision, recall, F1-measure, and G-Mean were used to calculate the performance of the classification model. The proportion of occasions for which a prediction was made accurately is known as accuracy. Despite this, accuracy, especially for the positive class in classification problems, cannot distinguish between the quantities of correctly categorized samples of each class. Positivity classes may have been misclassified as negativity by a highly trustworthy classifier. Therefore, precision isn't always enough to gauge a model's effectiveness in categorization challenges.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.2}$$

$$F_{Measure} = \frac{2\ *\ Precision * recall}{Precision + Recall} \tag{4.3}$$

$$G_{mean} = \sqrt{\frac{TP}{TP+FP}} + \sqrt{\frac{TP}{TP+FN}} = \sqrt{Recall} + \sqrt{Precision} \tag{4.4}$$

4.4 Performance Evaluation

For 50 epochs, the CNN model built on DenseNet-121 was trained to categorize the satellite photos. On grayscale photos, each pre-trained model was honed. Fig. 44 and b display the DenseNet CNN model accuracy and loss graph, respectively. Grayscale test images were used to assess DenseNet-121's performance. table **??** displays the test data's overall correctness. For a thorough performance analysis, table **??** shows the DenseNet-121 values for precision, recall, f1-score, and G-Mean using the grayscale test dataset.

While the applied DenseNet-121 model employed a rather large dataset, the vast majority of current techniques used a tiny dataset. By using several state-of-the-art methods, many images and procedures for validation have been developed. table **??** lists the sample size and validation strategy that each author employed. table **??** shows that a sizable number of models were developed, tested, and found to successfully categorize traffic occurrences. Some models for multi-class classification achieved accuracy of up to 95 percent or more [20], despite being more complex and expensive in terms of computation. A meaningful comparison of performance evaluation results and validation procedures is not possible due to the diversity of data sources. However, it is important to note that the 8200 road and nonroad picture dataset used to demonstrate the efficiency of the employed technique. The alternative methods employ a little bit less road photos. In an unbalanced dataset where only 219 and 127 photos were used for model development, the methods of Ouchicha, Ammor [20] and Narayan Das, Kumar achieved 97.2 percent and 97.4 percent overall testing accuracy, respectively. Pathak, Shukla [11] and Shaban, Rabie [15] employed CT scans simultaneously and attained 93 percent accuracy using a limited picture dataset. Other research [19, 10, 14, 22, 2, 1] proposed various methods for early identification of images revealing an accuracy rate of less than or equal to 90%. In this work, a CNN built on the DenseNet-121 architecture was employed to effectively detect using a total of 8200 images (4200 road and 4000 non-road). With 94 percent sensitivity (recall), 94 percent precision, 94 percent F1-score, and 94 percent G-Mean, the DenseNet-121 has produced 94 percent total accuracy. CNN built on DenseNet-121 performed better than previous experiments in the literature. Instead of using the same dataset as earlier studies, this one used data augmentation to provide superior results. For instance, GooLeNet offers accuracy of 80%. DenseNet-121's ability to identify COVID-19 cases with 92 percent accuracy, a low computational cost, and greater dependability than the traditional RT-PCR testing method is another advantage of using it.

1. Road recognition using satellite image classification outperforms using other photos, such as camera images. In comparison to prior studies, the DenseNet-121 model shows higher classification accuracy.
2. The applied architecture doesn't require any special extraction methods built by hand.

3. To sum up, our work demonstrated the potential of using deep learning to assist governments or engineers in automatically diagnosing road photographs.

Table 2: DenseNet Convolutional Neural Networks Application for Extracting Road Using satellite Image

	Precision	**Recall**	**f1-score**	**G-Mean**
Road	0.96	0.93	0.94	0.94
Not Road	0.94	0.95	0.93	0.93
				Overall accuracy
	0.94	0.94	0.94	0.94

5 CONCLUSIONS

In this research, we proposed the Densenet for satellite picture road extraction. All of the features in the satellite photos are referenced in the operation, which records remote information. More accurately than any other publication, our Densenet121 won the official DeepGlobe 2018 Road Extraction Challenge. With fewer parameters, fewer GFLOPs, and a quicker training convergence time, it also performed better than the top-ranked solution. By changing the positions and pairwise functions, the investigated DenseNet blocks. In the processing of satellite pictures, automated image classification performed by Computer-assisted systems are essential. It takes a lot of time and effort to analyze satellite images. This study determined that deep learning-based road extraction from satellite images beat accuracy in terms of performance. Recall is 94 percent accurate overall, and accuracy is 94 percent. The DenseNet-121 model's calculation time average is 195.35 seconds. For picture categorization, we used the DenseNet-121 deep learning architecture. For instance, a number of categorization models were used to support our assertion.

In order to obtain a true image of the infection rate, which is highly correlated with the daily incidence of infections, it would thus be beneficial to increase dependability as reflected by the computational approach for testing that has been provided. The used methodology provided a potent machine learning-based strategy to reduce manual judgment errors made by human specialists and offer a quicker manner of generating time and resource savings. Our current investigation, however, still contains a lot of flaws. First, future network architecture and optimization improvements might enhance categorization accuracy. Second, the lack of training data during the early stages of the extraction of the road is an inherent problem faced by some places. For accurate prediction, several hundred photos might not be enough. The absence of Grad Cam representations of the used DenseNet-121 model, which could have improved readability, is another flaw in this study. To increase road extraction's precision from satellite photos, further deep learning techniques or modified deep learning techniques may be incorporated and tested in future research.

References

[1] O.M. Amin Ali, S. Wahhab Kareem, and A.S. Mohammed, *Evaluation of electrocardiogram signals classification using CNN, SVM, and LSTM algorithm: A review*, 8th Int. Engin. Conf. Sustain. Technol. Dev. (IEC) (Erbil, Iraq), IEEE, February 2022, pp. 185–191.

[2] H.Q. Awla, A. Rahman Mirza, and S.W. Kareem, *An automated CAPTCHA for website protection based on user behavioral model*, 8th Int. Engin. Conf. Sustain. Technol. Dev. (IEC) (Erbil, Iraq), IEEE, February 2022, pp. 161–167.

[3] J. Canny, *A computational approach to edge detection*, IEEE Trans. Pattern Anal. Machine Intell. (1986), no. 6, 679–698.

[4] A. Chaurasia and E. Culurciello, *Linknet: Exploiting encoder representations for efficient semantic segmentation*, IEEE, 2017, pp. 1–4.

[5] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, *Encoder-decoder with atrous separable convolution for semantic image segmentation*, Proc. Eur. Conf. Comput. Vision (ECCV), 2018, pp. 801–818.

[6] L.C. Chen, G. Papandreou, F. Schroff, and H. Adam, *Rethinking atrous convolution for semantic image segmentation*, December 2017.

[7] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, *Deformable convolutional networks*, June 2017.

[8] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, *Deepglobe 2018: A challenge to parse the earth through satellite images*, Proc. IEEE Conf. Comput. Vision Pattern Recogn. Workshops, 2018, pp. 172–181.

[9] R.O. Duda and P.E. Hart, *Use of the Hough transformation to detect lines and curves in pictures*, Commun. ACM **15** (1972), no. 1, 11–15.

[10] C. Han, F. Shen, L. Liu, Y. Yang, and H.T. Shen, *Visual spatial attention network for relationship detection*, Proc. 26th ACM Int. Conf. Multimedia, 2018, pp. 510–518.

[11] R.Sc. Hawezi, F.S. Khoshaba, and S.W. Kareem, *A comparison of automated classification techniques for image processing in video internet of things*, Comput. Electric. Engin. **101** (2022), 108074.

[12] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, *The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation*, Proc. IEEE Conf. Comput. Vision Pattern Recogn. Workshops, 2017, pp. 11–19.

[13] J. Long, E. Shelhamer, and T. Darrell, *Fully convolutional networks for semantic segmentation*, Proc. IEEE Conf. Comput. Vision Pattern Recogn., 2015, pp. 3431–3440.

[14] W. Luo, Y. Li, R. Urtasun, and R. Zemel, *Understanding the effective receptive field in deep convolutional neural networks*, Adv. Neural Inf. Process. Syst. **29** (2016).

[15] H.A. Muhamad, S.W. Kareem, and A.S. Mohammed, *A comparative evaluation of deep learning methods in automated classification of white blood cell images*, 8th Int. Engin. Conf. Sustain. Technol. Dev. (IEC) (Erbil, Iraq), IEEE, February 2022, pp. 205–211.

[16] M. Song and D. Civco, *Road extraction using SVM and image segmentation*, Photogramm. Engin. Remote Sens. **70** (2004), no. 12, 1365–1371.

[17] F. Tupin, H. Maitre, J.-F. Mangin, J.-M. Nicolas, and E. Pechersky, *Detection of linear features in SAR images: Application to road network extraction*, IEEE Trans. Geosci. Remote Sens. **36** (1998), no. 2, 434–453.

[18] M. Ullah, A. Mohammed, and F. Alaya Cheikh, *PedNet: A spatio-temporal deep convolutional neural network for pedestrian segmentation*, J. Imag. **4** (2018), no. 9, 107.

[19] H. Wang, Y. Fan, Z. Wang, L. Jiao, and B. Schiele, *Parameter-free spatial attention network for person re-identification*, arXiv:1811.12150 [cs] (2018).

[20] X. Wang, R. Girshick, A. Gupta, and K. He, *Non-local neural networks*, Proc. IEEE Conf. Comput. Vision Pattern Recogn., 2018, pp. 7794–7803.

[21] Y. Wei, Z. Wang, and M. Xu, *Road structure refined CNN for road extraction in aerial image*, IEEE Geosci. Remote Sens. Let. **14** (2017), no. 5, 709–713.

[22] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, *Residual non-local attention networks for image restoration*, arXiv preprint arXiv:1903.10082 (2019).

[23] Z. Zhang, Q. Liu, and Y. Wang, *Road extraction by deep residual u-net*, IEEE Geosci. Remote Sens. Let. **15** (2018), no. 5, 749–753.

[24] H. Zhao, Y. Zhang, S. Liu, J. Shi, C.C. Loy, D. Lin, and J. Jia, *Psanet: Point-wise spatial attention network for scene parsing*, Proc. Eur. Conf. Comput. Vision (ECCV), 2018, pp. 267–283.

[25] L. Zhou, C. Zhang, and M. Wu, *D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction*, Proc. IEEE Conf. Comput. Vision Pattern Recogn. Workshops, 2018, pp. 182–186.