SSRG International Journal of Computer Science and Engineering
ISSN: 2348 – 8387 / https://doi.org/10.14445/23488387/IJCSE-V9I3P103

Volume 9 Issue 3, 21-28, March 2022

Original Article

Urban Remote Sensing Image Segmentation using Dense U-Net+

Keerti Maithil[1], Tasneem Bano Rehman[2]

Computer Science and Engineering, SAGE School of Advanced Computing, SAGE University Bhopal, Madhya Pradesh, India.

Received: 13 February 2022 Revised: 22 March 2022 Accepted: 27 March 2022 Published: 31 March 2022

Abstract *- For a long time, man has been dreaming that we should make such a machine with human-like intelligence, the power to understand like a human and can think like a human. One of the fascinating ideas was to give computers the ability to see and interpret the world around them. The concept of computer vision is based on training a computer, which processes an image to understand and analyze it at a pixel level. Technically, machines attempt to retrieve visual information, handle it, and interpret results through special software algorithms. An important subject within computer vision is image segmentation. Image Segmentation is a process of identifying objects or boundaries to simplify an image and efficiently analyzing it by dividing the image into different regions based on the characteristics of pixels. The existing U-net-based segmentation model and its other variants are the deep learning module design, especially for biomedical image segmentation; initially, it was proposed for cell segmentation. This work finds a new application area: Urban Remote Sensing Image Segmentation using the Dense U-Net+ model. DenseU-Net+ is a powerful form of the U–net architecture inspired by DenseNet. The imbalance is a serious problem in the remote sensing image segmentation class. Another one is that segmentation of large objects in the image is easy, but for small objects, segmentation causes difficulties.*

Keywords *- Image segmentation, Computer vision, DenseNet, U-Net, DenseU-Net.*

1. Introduction

The main goal of image segmentation is that the representation of an image should be simplified or changed so that it is more meaningful and easier to analyze. The image segmentation process is used to locate boundaries and objects in the image. The set of segments of contours extracted from the image that collectively cover the entire image is the result of image segmentation [1]. DenseNet is a neural network that is used for the classification of an image. Performing the segmentation class imbalance is a serious problem that afflicts the image segmentation task in urban remote sensing images; where we talk about the segmentation of large object classes then, it is not a difficult thing to solve, but the same we talk about small object classes becomes a very tough task to segmentation[2]. Remote sensing is a technique of obtaining the physical properties of an area without being there and making measurements of the earth using sensors and satellites. One of the main applications of remote sensing image data is the 3D geographical information system that represents digital models of urban areas: earth's surface, vegetation, buildings, infrastructure, and other related objects belonging to urban areas. Deep learning models have played an important role in the task of segmentation of images. The convolutional neural network is used for image processing. The advanced version of the convolutional neural network, the u-net, is used to perform the segmentation task of Biomedical images and Smartphone wound images [3]-[5].

2. Literature Survey

So far, there has been a lot of research related to semantic image segmentation based on different deep-learning neural network architectures. A deep guidance network to segment the biomedical image. The proposed network consists of a guided image filter module to restore the structure information through the guidance image. The method enables end-to-end training and fast inference (43ms for one image) [1]. A model DenseUNet-Based Semantic Segmentation of Small Objects in Urban Remote Sensing Images. The main idea of the DenseU-Net is to connect convolutional neural network features through cascade operations and use its symmetrical structure to fuse the detail features in shallow layers and the abstract semantic features in deep layers. The experiments were based on the 2016 ISPRS Vaihingen 2D semantic labelling dataset and demonstrated the following outcomes. When boundary pixels were considered (GT), MFB_Focalloss achieved a good overall segmentation performance using the same U-Net model. The F1-score of the small object class ``car'' was improved by 9.28% compared with the cross-entropy loss function [2]. A Deep network for brain tumour image segmentation, an ensemble of two segmentation networks: a 3D CNN and a U-Net, in a significant yet straightforward combinative technique that results in better and more accurate predictions. Both models were trained separately on the BraTS-19 challenge dataset and evaluated to yield segmentation maps that differed considerably in segmented tumour sub-regions and were assembled variably to achieve the final prediction[3]. Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi. The Newly-devised 2D and 3D-Net approaches were more effective than a commercial auto-segmentation tool. Even the relatively shallow2DU-Net, which does not require high-performance computational resources, was effective enough for lung segmentation. Semantic segmentation using deep learning was useful in radiation treatment planning for lung cancers [4].A Double U-Nets for image segmentation by integrating the region and boundary information. The proposed network consisted of a down-sampling path and two symmetric up-sampling paths. The down-sampling path learned the low-level features of regions and boundaries, and the two up-sampling paths learned the high-level features of regions and boundaries, respectively. The outputs from the down-sampling path were concatenated with the corresponding ones from two up-sampling paths by skip connections. The outputs of double U-Nets were the predicted probability images of object regions and boundaries, and they were integrated to calculate the dice loss with the corresponding labels. The proposed double U-Nets were evaluated on two datasets: 247 radiographs for the segmentation of lungs, hearts, and clavicles and 284 radiographs for the segmentation of pelvises [5].Semantic Image Segmentation using Deep Convolutional Neural Networks. In this paper, deep convolutional neural networks with multiple layers are projected multiple layers work to build an improved feature space. This deep convolutional neural network is given a solution in the form of colour segmentation. The deep CNN-based segmentation model shows 93% accuracy on the BSDS300 dataset.[6]. Semantic segmentation of Smartphone wound images: Comparative analysis of AHRF and CNN-Based approaches.This paper compares AHRF and CNN approaches (FCN, U-Net, DeepLabV3) using various metrics, including segmentation accuracy (dice score), inference time, required training data, and performance on diverse wound sizes and tissue types. Improvements possible using various image pre-and post-processing techniques are also explored. As access to adequate medical images/data is a common constraint, we explore the sensitivity of the approaches to the size of the wound dataset. We found that for small datasets (<300 images), AHRF is more accurate than U-Net but not as accurate as FCN and DeepLabV3. AHRF is also over 1000x slower. For larger datasets (>300 images), AHRF saturates quickly, and all CNN approaches (FCN, U-Net and DeepLabV3) are significantly more accurate than AHRF[7]. In computed tomography images, a U-Net Plus model for deep semantic segmentation for Esophagus and Esophageal cancer. U-Net Plus is proposed to segment oesophagus and oesophagal cancer from a 2-D CT slice. In the new network architecture, two blocks enhance the feature extraction performance of complex abstract information, effectively resolving irregular and vague boundaries [8]. A study on iris Segmentation algorithm based on Dense U-Net in which iris is segmented by taking advantage of a dense U-Net network, which is narrower and has fewer parameters, and taking advantage of U-Net in semantic segmentation. A densely connected path is derived from a densely connected network (Dense U-Net), in which improved information and parameters are helpful to reduce the training difficulty of deep networks [9]. A convolutional network architecture originally proposed for the semantic segmentation of biomedical images, the proposed method uses image conversion by a U-Net framework. The proposed method does not use any information from mathematical and linguistic grammar. It can be a supplemental bypass in the conventional mathematical optical character recognition (OCR) process pipeline. The evaluation experiments confirmed that (1) the performance of mathematical symbol and expression detection by the proposed method is superior to that of InftyReader, which is state-of-the-art software for mathematical OCR; (2) the coverage of the training dataset to the variation of the document style is important; and (3) retraining with small additional training samples will be effective to improve the performance [10]. U-Net and Its Variants for Medical Image Segmentation, U-net is a technique developed primarily for image segmentation tasks. These traits provide U-net with a high utility within the medical imaging community and have resulted in extensive adoption of U-net as the primary tool for segmentation tasks in medical imaging. The success of U-net is evident in its widespread use in nearly all major image modalities, from CT scans and MRI to X-rays and microscopy [11]. HA U-Net, A improved model for building extraction from high-resolution remote sensing imagery.The designed HA U-Net performs well on WHU Building Dataset and Urban 3d Challenge dataset and achieves 9.31%, 2.17% better F1-score and 10.78%, and 1.77% better IOU than the standard U-Net, respectively. The experimental results indicate that the proposed method can solve the building adhesion problem well. The research can serve as updating geographic databases[12]. The U-Net++ model for automatic brain tumour segmentation variation of the U-Net++ model, an adaptation of U-Net, and evaluating its brain tumour segmentation capabilities. The proposed approach obtained Dice Coefficient scores of 0.7192, 0.8712, and 0.7817 for the Enhancing Tumor, Whole Tumor and Tumor Core classes of the BraTS 2019 challenge Validation Dataset [13]. Automatic building extraction on high-resolution remote sensing imagery using deep Convolutional Encoder-Decoder with spatial pyramid pooling [14].Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with a dense connection. A U-Net-based deep convolutional neural network, TL-DenseUNet, for the semantic segmentation of remote sensing images. The proposed TL-DenseU-Net contains two subnetworks. To extract multi-level semantic features, the encoder subnetwork uses a transferring DenseNet pretrained on three-band image Net images. The decoder subnetwork adopts a dense connection to fuse the multiscale information in each layer, which can strengthen the expressive capability of the features [15]. A Transformer-Based Decoder Designs for Semantic Segmentation on Remotely Sensed Images. A deep-learning (DL) model can improve the semantic segmentation network in two ways. First, utilizing the pre-training Swim Transformer (SwinTF) under Vision Transformer (ViT) as a backbone, the model weights downstream tasks by joining task layers upon the pretrained encoder. Secondly, decoder designs are applied to our DL network with three decoder designs, U-Net, pyramid scene parsing (PSP), and feature pyramid network (FPN), to perform pixel-level segmentation. The results are compared with other image labellings state-of-the-art (SOTA) methods, such as global convolutional network (GCN) and ViT. Extensive experiments show that our Swin Transformer (SwinTF) with decoder designs reached a new state of the art on the Thailand Isan Landsat-8 corpus (89.8% F1 score), Thailand North Landsat-8 corpus (63.12% F1 score), and competitive results on ISPRS Vaihingen. Moreover, our best-proposed methods (SwinTF-PSP and SwinTF-FPN) outperformed SwinTF with supervised pre-training ViT on the ImageNet-1K in Thailand, Landsat-8, and ISPRS Vaihingen corpora[16]. A U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery the U-Net convolutional net for land-cover classification is based on a multispectral Unmanned aerial vehicle (UAV) image in a mining area of Daknong province, Vietnam.The nal U-Net model can interpret six land cover types: (1) open-case mining lands, (2) old permanent croplands, (3) young permanent croplands, (4) grasslands, (5) bare soils, and (6) water bodies. As a result, two models using Nadam and Adadelta optimizer functions can classify six land cover types with accuracy higher than 83%, especially in open-case mining lands and polluted streams owed out from the mining areas [17]. A MACU-Net for Semantic Segmentation of Fine-Resolution Remotely Sensed Images, a deep encoder-decoder architecture, has been used frequently for image segmentation with high accuracy. In this letter, we incorporate multiscale features generated by different layers of U-Net and design a multiscale skip-connected and asymmetric-convolution-based U-Net (MACU-Net) for segmentation using fine-resolution remotely sensed images [18]. An Object-Level Remote Sensing Image Augmentation Using U-Net-Based Generative Adversarial Networks, an object-level remote sensing image augmentation approach leveraging the U-Net-based generative adversarial networks. Specifically, our proposed approach consists of the semantic tag image generator and the U-Net GAN-based translator. To evaluate the effectiveness of the proposed approach, comprehensive experiments are conducted on a public dataset HRSC2016. State-of-the-art generative models, DCGAN, WGAN, and CycleGAN, are selected as baselines. According to the experimental results, our proposed approach significantly outperforms the baselines in terms of drawing the outlines of target objects and capturing their meaningful details[19]. An End to End Segmentation of Canola Field Images Using Dilated U-Net, Maximum Likelihood Classification (MLC) and image processing techniques to labelled images in three classes; background, crop, and weeds. This data is processed through our modified U-Net, improving semantic accuracy with reduced memory cost. We train our model with DICE loss and compare the results with state of art. We achieve 89.12% mean Intersection Over Union (mIOU) with 86.11%, 82.99%, and 98.23% individual IOU for the crop, weeds, and background. Our proposed model uses only 15M parameters which are 57M less than the state-of-the-art models with a compromise of a 1% mIOU score [20]. A model which performs Automatic Segmentation of Human Placenta Images With U-Net. An automatic segmentation method of the human placenta reduces manual intervention and greatly speeds up the segmentation, making large-scale segmentation possible. The human placenta data set we used was labelled by experts and obtained from prenatal examinations of 11 pregnant women, about 1,110 images. It was a comprehensive and clinically significant data set. By training the network with such a data set, the robustness of the model will be better. After testing the data set, the segmentation effect is consistent with the manual segmentation effect [21]. Study recent progress in semantic image segmentation, and divide semantic image segmentation methods into two categories: traditional and recent DNN methods. Firstly, we briefly summarize the traditional method and datasets released for segmentation. We comprehensively investigate recent methods based on DNN, which are described in the eight aspects: fully convolutional network, upsample ways, FCN joint with CRF methods, dilated convolution approaches, progresses in the backbone network, pyramid methods, Multi-level feature and multi-stage method, supervised, weakly-supervised and unsupervised methods [22]. In a survey providing a comprehensive review of the literature at the time of this writing, covering a broad spectrum of pioneering works for semantic and instance-level segmentation, including fully convolutional pixel-labelling networks, encoder-decoder architectures, multiscale and pyramid-based approaches, recurrent networks, visual attention models, and generative models in adversarial settings. We investigate these deep learning models' similarities, strengths and challenges, examine the most widely used datasets, report performances, and discuss promising future research directions in this area[23].Semantic Image Segmentation for Building Detection in Urban Areas with Aerial Photograph Images using U-Net Models. Semantic image segmentation aims to build detection in urban areas using a deep learning model. Each image pixel is categorized into either building or non-building labels. Based on experimentation using aerial photograph imagery of Pasar Minggu Sub-District, South Jakarta City District, DKI. The Jakarta Province and UNet model achieved 0.83 average training accuracy and 0,87 testing accuracy [24]. A Complex Network Classification with Convolutional Neural Network, a novel framework of Complex Network Classifier (CNC) by integrating network embedding and convolutional neural network to tackle the problem of network classification. By training the classifier on synthetic complex network data, we show CNC can classify networks with high accuracy and robustness and extract the features of the networks automatically. We also compare our CNC with baseline methods on benchmark datasets, which shows that our method performs well on large-scale networks [25].

3. Methodology

The combination makes the model that we will use here two DenseU-Net models. First of all, the DenseNet layer, a kind of neural network, will be doing the image classification, and after that, the segmentation process will occur. A max-pooling layer will be used to collect the features from an image. After the feature collection process, the convolution layer will do the feature extraction process. Both the max-pooling and convolution layers work in the model's down-sampling path, also known as the contracting path. Then next, there will be an up-sampling path where the up-convolution layer is present for recreating the feature coming from the down-sampling path. A 1x1 convolution layer will be used to reduce the feature map, and the expansion path and a concatenation layer are present here, which will connect the down-sampling and up-sampling blocks.

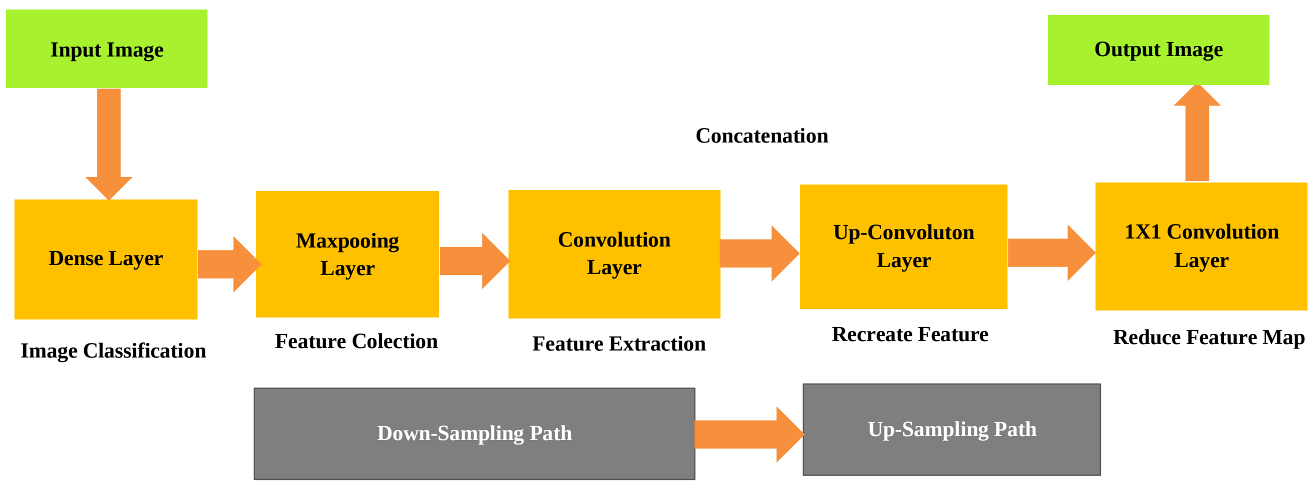

Fig. 1 Flow Diagram of DenseU-Net+ Model

3.1. DenseNet

DenseNet is the standard convolutional neural network used for image classification; the input image goes through the multiple convolution layers and gets the high-level features in the model. In DenseNet, each layer gets the additional input from all previous layers and passes on its output to all the next layers. As shown in fig.2, Dance blocks have multiple convolution blocks. DenseNet provides a versatile and effective method for medical image classification tasks, but they require large amounts of data with labels and involve complex and time-consuming training processes.

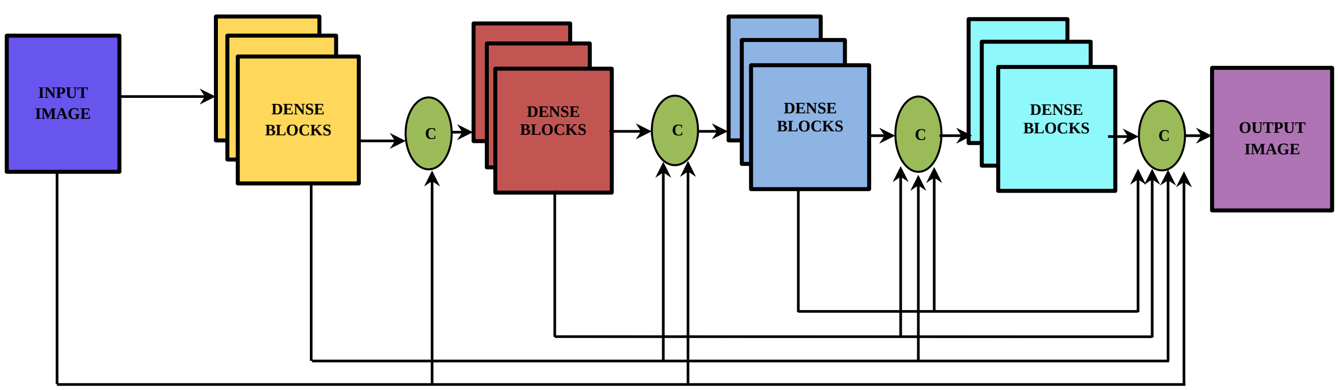

Fig. 2 DenseNet

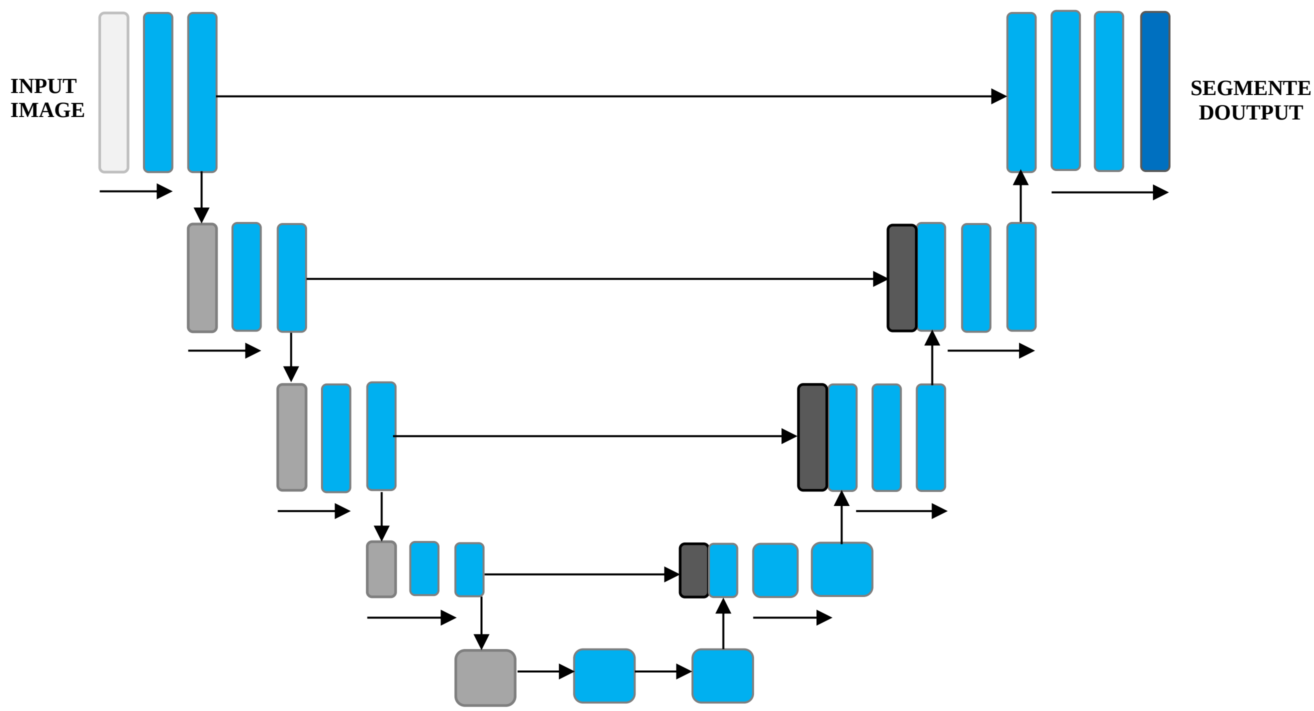

Fig. 3 U-Net Architecture

3.2. U-NET

The u-Net architecture was proposed by Ronne Berger et al., motivated by the promising achievement of its semantic segmentation of biomedical images. U-Net is an FCN architecture proposed for the segmentation of biomedical images. Introducing skip connections between corresponding layers in the encoder and decoder successfully preserves the high-frequency components in the converted output images [5]. The encoder is responsible for projecting the input feature vectors into a low-dimensional space where similar features lie close together.

The decoder network takes features from this low-dimensional space as input and attempts to recreate the original input features [7]. The first is the contracting path that uses a typical CNN architecture. Each block in the contracting path consists of two successive 3X3 convolutions followed by a ReLU activation unit and a max-pooling layer. This arrangement is repeated several times. The novelty of the U-net comes in the second part, called the expansion path, in which each stage up-samples the feature map using 2X2 up-convolution. Then, the feature map from the corresponding layer in the contracting path is cropped and concatenated onto the upsampled feature map[11]. It is followed by two successive 3X3 convolutions and ReLU activation. At the final stage, an additional 1X1 convolution is applied to reduce the feature map to the required number of channels and produce the segmented image [11]. Fig. 3 illustrates the overall U-net Architecture. The energy function for the network is given :

$$\mathrm{E}{=}\sum w(x) log(p_{k(x)}(x)) \tag{1}$$

Where is the pixel-wise SoftMax function applied over the final feature map, defined as:

$$p_k{=}\exp(a_k(\mathrm{x}))/\sum_{k'=1}^{K} exp\ (a_k(x)') \tag{2}$$

and $a_k$denotes the activation in channel k.

3.3. DenseU-Net

In the traditional neural network model, each network layer only gets the input signal from the upper layer and then transmits the extracted features to the next layer. In fig.4, Dense U-Net is different: first, there is a direct connection between any two layers of the network; that is, the input of each layer of the network is a union of the outputs of all the previous layers. Each layer of Dense U-Net learns only a very limited number of features, which are passed directly to all the layers behind it as input To reduce redundancy [4]. The output of each dense layer is described as follows:

$$x_l = H_l([x_{0,}x_{1,}x_{2,....,}x_{l-1}]) \tag{3}$$

DenseNet preserves all identity maps from prior layers and significantly promotes gradient propagation. The implication is that each layer can have fewer channels, as information is more easily preserved between layers, resulting in higher accuracy with fewer computations, allowing deep learning models with greater layers.

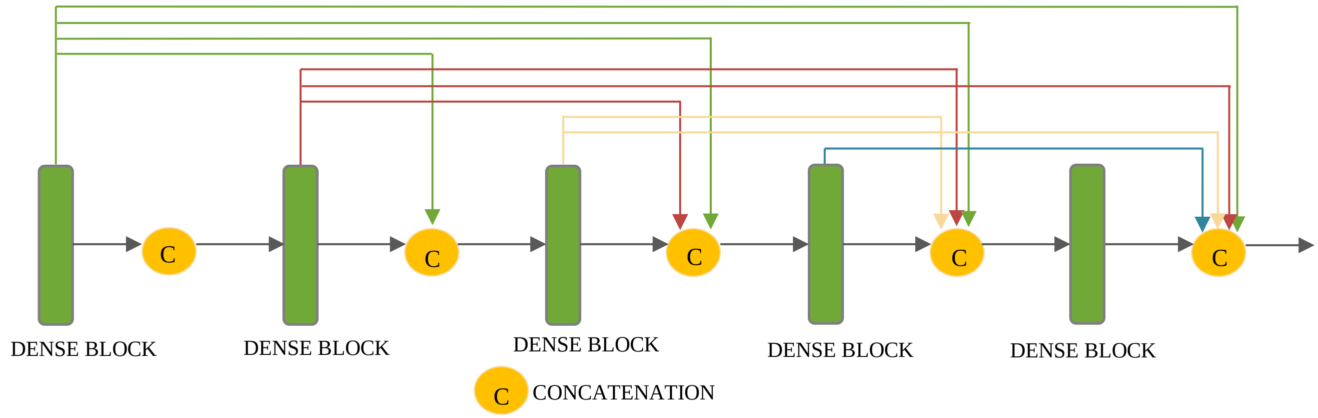

Fig. 4 A five-layer DenseU-Net

The concatenation unit receives the feature map from all previous layers and passes it onto the next layer. It ensures that any given layer has contextual information from the previous layers in the block [11].

3.4. DENSEU-NET+

DenseU-Net+ is a powerful form of U-net architecture inspired by DenseNet [11]. It uses a dense network of skip connections as an intermediary grid between the contracting and expansive paths [2]. It aids the network by propagating more semantic information between the two paths, thereby enabling it to segment images more accurately. As shown in fig.5, each square denotes a convolutional block. Unlike the base U-net, which has a single direct concatenation from the contracting path to the expansive path, DenseU-Net+ has a series of intermediary convolutional blocks between the two paths. Each intermediary and expansive block receives the concatenated feature maps from the previous blocks at the same level and the upsampled feature map from the block immediately below it [11]. In traditional U-net, the feature maps of the contracting path are directly concatenated onto the corresponding layers in the expansive path. DenseU-Net+ has several skip connection nodes between each corresponding layer, as represented in Fig. 6. Each skip connection unit receives all of the feature maps from all previous units at the same level and an upsampled feature map from its immediate lower unit. Therefore, each level is equivalent to a dense block. This arrangement minimizes the loss of semantic information between the two paths. The operation of the skip connection unit in which x represents the feature map and i and j correspond to the index down the contracting path and across the skip connections, respectively [11]. The operation of the skip connection unit in which x represents the feature map and i and j correspond to the index down the contracting path and across the skip connections, respectively, is defined

$$\boldsymbol{x}^{i,j} = \begin{cases} H(x^{i-1,j}), & j = 0 \\ H\left(\left[[x^{i,k}]_{k=0}^{j-1}, U(x^{i+1,j-1})\right]\right), & j > 1 \end{cases} \quad \text{........ (4)}$$

Here, H(.) denotes convolution, and the activation operation, U(.) represents the upsampling operation, and [] signifies a concatenation. The number of intermediary skip connection units depends on the layer number and decreases linearly when traversing the contracting path.

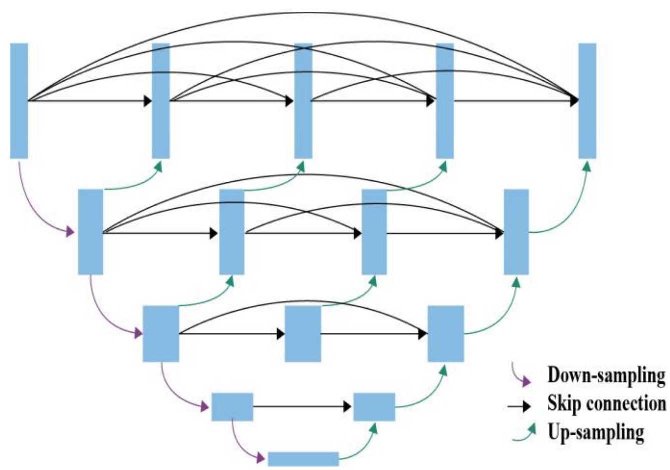

Fig. 5 DenseU-Net+ schematic representation

Some advantages are using the DenseU-Net+ model:

1. DenseU-Net+ is built on the network architecture of DenseNet and FCN. The network is modified by combining two DenseU-Nets to get very accurate segmentation results with very few training images.
2. Dense net is encouraged to reuse features and substantially reduce the number of features.
3. Add the upper sampling phase and a lot of feature channels, allowing more of the original image texture information in the high-resolution layers for dissemination.
4. U-net has no FC layer and uses valid convolution. It ensures that the segmentation result is based on no missing context feature, so the size of the input and output images is not quite the same for large input images; you can use the overlapping strategy for seamless image output.

5. To predict the edge of the input image, it is possible to extrapolate the missing context information by mirroring the input image. It is also possible to input a large image, but this strategy is based on the case of insufficient GPU memory [9].

4. Conclusion

Some older variants of the U-Net could not solve the problem of class imbalance and the problem of detecting small objects with high accuracy, So keeping the same problem in mind, I will work with this new variant of u-net. The DenseU-Net+ model will address the class imbalance problem and for automatic segmentation with high performance due to several factors such as low contrast of image and similarity between the shapes of nearby objects or semantic segmentation of urban remote sensing image. To eliminate the problem of class imbalance and also want to remove the problem of not correctly identifying the small objects in images with similarities between their shapes.

References

[1] Peng Shuai Yin et al., "Deep Guidance Network for Biomedical Image Segmentation," *IEEE Access*, vol. 8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Rongsheng Dong, Xiaoquan Pan, and Fengying Li, "Dense U-Net-Based Semantic Segmentation of Small Objects in Urban Remote Sensing Images," *IEEE Access*, vol. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Mahnoor Ali et al., "Brain Tumor Image Segmentation using Deep Networks," *IEEE Access*, vol. 8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Takafumi Nemoto et al., "Efficacy Evaluation of 2D, 3D U-Net Semantic Segmentation and Atlas-Based Segmentation of Normal Lungs Excluding the Trachea and Main Bronchi," *Journal of Radiation Research*, vol. 61, no. 2, pp. 257–264, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Wei Guo et al., "Double U-Nets for Image Segmentation by Integrating the Region and Boundary Information," *IEEE Access*, vol. 9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Santosh Jangid, and P S Bhatnagar, "Semantic Image Segmentation Using Deep Convolutional Neural Networks and Super-Pixels," *International Journal of Applied Engineering Research*, vol. 13, no. 20, pp. 14657-14663, 2018. [Google Scholar] [Publisher Link]

[7] Ameya Wagh et al., "Semantic Segmentation of Smartphone Wound Images: Comparative Analysis of AHRF and CNN-Based Approaches," *IEEE Access*,vol. 8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Shuchao Chen et al., "U-Net Plus: Deep Semantic Segmentation for Esophagus and Esophageal Cancer in Computed Tomography Images," *IEEE Access*, vol. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Xiaoqiang WU, and Lng Zhao, "Study on Iris Segmentation Algorithm Based on Dense U-Net," *IEEE Access*, vol. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Wataru Ohyama, Masakazu Suzuki, and Seiichi Uchida, "Detecting Mathematical Expressions in Scientific Document Images Using a U-Net Trained on A Diverse Dataset," *IEEE Access*, vol. 7, pp. 144030 – 144042, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Nhian Siddique et al., "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access*, vol. 9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Leilei Xu et al., "HA U-Net: Improved Model for Building Extraction From High-Resolution Remote Sensing Imagery," *IEEE Access*, vol. 9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Neil Micallef, Dylan Seychell, and Claude J. Bajada, "Exploring the U-Net++ Model for Automatic Brain Tumor Segmentation," *IEEE Access*, vol. 9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yaohui Liu et al., "Automatic Building Extraction on High-Resolution Remote Sensing Imagery Using Deep Convolutional Encoder-Decoder With Spatial Pyramid Pooling," *IEEE Access*, vol. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Binge Cui, Xin Chen, and Yan Lu, "Semantic Segmentation of Remote Sensing Images Using Transfer Learning and Deep Convolutional Neural Network with Dense Connection," *IEEE Access*, vol. 8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Teerapong Panboonyue et al., "Transformer-Based Decoder Designs for Semantic Segmentation on Remotely Sensed Images," *Remote Sensing*, vol. 13, pp. 5100, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Tuan Linh Giang et al., "U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery," *IEEE Access*, vol. 8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Rui Li et al., "MACU-Net for Semantic Segmentation of Fine-Resolution Remotely Sensed Images," *Electrical Engineering and Systems Science*, vol. 19, 2022. [CrossRef] [Publisher Link]

[19] Jian Huang et al., "Object-Level Remote Sensing Image Augmentation Using U-Net-Based Generative Adversarial Networks," *Wireless Communications and Mobile Computing*, vol. 2021, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Hafiz Sami Ullah, Muhammad Hamza Asad, and Abdul Bais, "End to End Segmentation of Canola Field Images Using Dilated U-Net," *IEEE Access*, vol. 9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Mo Han et al., "Automatic Segmentation of Human Placenta Images with U-Net," *IEEE Access*, vol. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Xiaolong Liu, Zhidong Deng, and Yuhan Yangn, "Recent Progress in Semantic Image Segmentation," Artificial Intelligence Review, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Shervin Minaee et al., "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Edy Irwansyah, and Yaya Heryadi, "Semantic Image Segmentation for Building Detection in Urban Areas with Aerial Photograph Image Using U-Net Models," *2020 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Ruyue Xin, Jiang Zhang, and Yitong Shao, "Complex Network Classification with Convolutional Neural Network," *Tsinghua Science and Technology*, vol. 25, no. 4, 2020. [CrossRef] [Google Scholar] [Publisher Link]