Fractional Derivative Based Weighted Skip Connections for Satellite Image Road Segmentation

Sugandha Arora[a,*], Harsh Kumar Suman[b], Trilok Mathur[a], Hari Mohan Pandey[c], Kamlesh Tiwari[b]

[a]*Department of Mathematics, Birla Institute of Technology and Science Pilani, Rajasthan, INDIA, 333031*
[b]*Department of CSIS, Birla Institute of Technology and Science Pilani, Rajasthan, INDIA, 333031*
[c]*Data Science & Artificial Intelligence Department, Bournemouth University, Fern Barrow, Poole, Dorset, BH12 5BB, UK*

Abstract

Segmentation of a road portion from a satellite image is challenging due to its complex background, occlusion, shadows, clouds, and other optical artifacts. One must combine both local and global cues for an accurate and continuous/connected road network extraction. This paper proposes a model using fractional derivative-based weighted skip connections on a densely connected convolutional neural network for road segmentation. Weights corresponding to the skip connections are determined using Grunwald-Letnikov fractional derivative. Fractional derivatives being non-local in nature incorporates memory into the system and thereby combine both local and global features. Experiments have been performed on two open source widely used benchmark databases $viz.$ Massachusetts Road database (MRD) and Ottawa Road database (ORD). Both these datasets represent different road topography and network structure including varying road widths and complexities. Result reveals that the proposed system demonstrated better performance than the other state-of-the-art methods by achieving an F1-score of 0.748 and the mIoU of 0.787 at fractional order 0.4 on the MRD and a mIoU of 0.9062 at fractional order 0.5 on the ORD.

Keywords: Remote Sensing, Road Network Extraction, Image Segmentation, Fractional-Order Derivative

1. Introduction

Remote sensing images acquired through satellites are of high resolution and contain large coverage of the geographical region. These images provide accurate topographical information about the earth's surface [1, 2, 3]. Roads can be extracted from these images, which can help in urban planning, emergency rescue, autonomous driving $etc.$ Semantic segmentation is needed to extract roads from such images. It is equivalent to a classification setup that assigns a class of either being road or non-road to every pixel of the image. After extracting the pixels depicting roads, the extracted roads can be further used for extracting their centerline using morphological thinning-based algorithms [4]. This study is focused on extraction of all the pixels corresponding to roads. Traditional automatic road segmentation methods involve the usage of machine learning techniques like Bayesian [5] and heuristic methods [6, 7]. Heuristic methods semantically combine roads and group them using hypothesis and testing paradigm, while the

*Corresponding Author
Email addresses: p20180024@pilani.bits-pilani.ac.in (Sugandha Arora), f20190076@pilani.bits-pilani.ac.in (Harsh Kumar Suman), tmathur@pilani.bits-pilani.ac.in (Trilok Mathur), hpandey@bournemouth.ac.uk (Hari Mohan Pandey), kamlesh.tiwari@pilani.bits-pilani.ac.in (Kamlesh Tiwari)

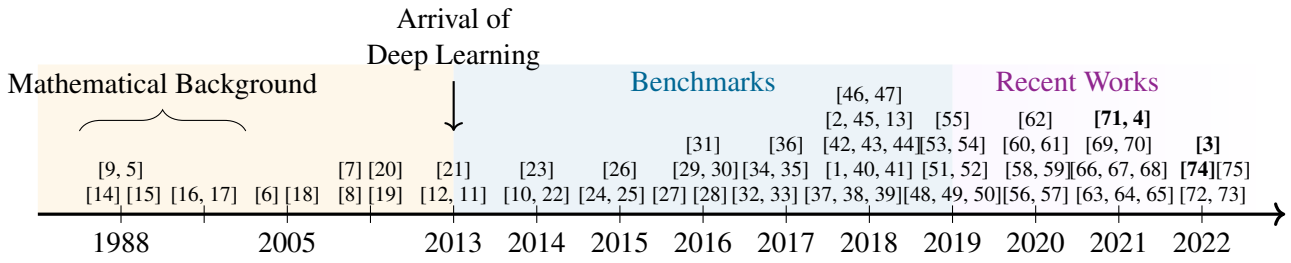


Figure 1: Visual representation of references arranged in chronological order that is helping in better understanding of the progress in road segmentation

Bayesian approach involves stochastic process models which extract roads by probabilistic modeling [8, 4]. Ambient challenges such as shadow, variable road width, complex surroundings, occlusion arising due to traffic and trees $etc.$ pose serious challenges to these methods and significantly reduces their accuracy [9, 10]. Most of the machine learning algorithms fail to achieve high precision in the road network segmentation due to their inability to handle multi-scale road sections, particularly narrow road segments with substantial width variability [11, 12]. Road extraction models based on road features, automation, problems arising while extraction and type of methods adopted, are comprehensively reviewed in [13].

To overcome the challenges of road network segmentation, several deep-learning approaches have been developed [29, 44, 69]. Convolutional neural networks (CNN) are popular filters for extracting useful features from an image. A network with multiple layers of CNNs called deep convolutional neural networks (DCNNs) can be devised to build a hierarchy of features from the multi-scale remote sensing data sets [57, 50]. The mainstream segmentation networks like Fully Convolutional Neural Networks (FCNN) [26], U-Net [24], U-Net++ [45], Deeplabv3 [34], deep residual U-Net [42] etc are successfully being used for the road extraction. The FCNNs with skip connection architecture can combine semantic information from deep, coarse layers and appearance information from shallow, fine layers to improve the segmentation performance. This fusion of fine layers and coarse layers intends to preserve global structure while making local predictions. Further, U-Net [24, 45] uses an end-to-end training approach by concatenating the pair of corresponding cropped feature maps from the encoder and decoder that prevents the loss of information and helps in the precise localization. ResNet [76, 42] uses the past feature maps using skip connections that help to prevent vanishing gradient problems and avoid any loss of information. DenseNets [49] also uses skip connections to get inputs from each preceding layer of a dense block, which increases the input variability and the network performs more accurately.

The architectures of networks mentioned above can efficiently perform the segmentation task, but complex backgrounds and occlusion due to buildings, cars, trees , $etc.$, overlapping, interlacing, and shadowing of the clouds, trees $etc.$ in satellite images affects the accuracy of road network extraction. Long-term feature interactions are required to extract roads from the images with complex backgrounds, since such interactions can preserve semantic/appearance information. Thus to improve the forward propagation of the features, it is better to reuse the feature extracted at preceding states. Several works related to combination of local and global information exist [77, 78]. For instance, a self-weighted technique of combining local and global motions has been proposed in [79] for crowd detection. The architecture of some DCNNs such as DenseNet [49], SDUNet [74], CondiNet++ [67], RoadVecNet [65] and multiview-based parameter free approach [79] have been designed to address the problem of long-term memory and preservation of global & local cues, but without any solid theoretical explanation of memory exploitation. These works extract global and local features separately, while fractional derivative based methods inherently compute global and local features. FOCNet [53] has been proposed in the

literature for image denoising to address the memory issue, where the network is designed by solving a system of fractional differential equations and, the global and local features are extracted inherently. Some relevant works involving road segmentation are compared in Table 1 on the basis of architecture and results. Figure 1 shows the arrangement of references in chronological order to show how over time, frontiers of mathematical background have developed and the work on the road network segmentation has gained research attention from the community. It can be noted that earlier fractional calculus was an abstract area of research with vigorous calculations and recently with the introduction of new definitions of fractional derivatives, the applications of fractional calculus is being seen in several domains of science and engineering including image processing. It can also be seen that the arrival of deep learning has motivated the growth of automatic road extraction techniques and researchers are engaged in overcoming the challenges through the development of new models.

The automatic way of finding the network parameters treats the system as a black box, which lacks interpretability [80, 81]. Due to this, understanding the conceptual basis, potential advantages, and limitations of methods become difficult [82]. Moreover, it is not feasible to tune the additional parameters when handling different databases. If the process of decision-making by the network is known, then this provides the flexibility of adjusting human preferences through parameters. Thus it is recommended to decide these parameters on a mathematical basis. Some recent studies have shown that the forward propagation in DCNNs can be treated as a dynamic system which can further be characterized by a differential equation [36, 15, 46]. The forward propagation of ResNet is shown to be equivalent to the forward Euler numerical discretization of an ordinary differential equation (ODE) in [46]. They also showed that other recent architectures like RevNet [35], and FractalNet [30] can be designed by the discretization of some ordinary/ partial differential equation [56]. The ordinary/partial differential equations systems are local, due to which only short-term feature interactions are possible, whereas fractional differential systems are non-local and have long-term memory. Numerous dynamic systems have been demonstrated to be better characterized by FODE than ODE [83]. Therefore, designing the road segmentation network using ODE is not enough. We have modeled the propagation of the deep road segmentation network as a dynamic system using fractional differential equations to enhance the memory of the network. Fractional differential equations are non-local and incorporate memory into the system by adding fractional weights to the skip connections between the layers. These fractional weights can be interpreted using the mathematical framework of Grunwald-Letnikov fractional derivatives [17]. The memory of a dynamic system obeys the power law [23], which means the recent stage will have more weightage than the older ones, and Grunwald Letnikov fractional derivatives also follow the power-law memory property [23, 47].

In this paper, the road segmentation network has been designed naturally by solving a fractional optimal control. The forward propagation of the proposed network is constructed by discretizing the system of Grunwald-Letnikov based FODEs with control variables. As a result, the output of each layer of the proposed network becomes dependent on the weighted sum of features of previous layers at the same level where weights are controlled by fractional order and state of the network. To summarize, the advantages of using FODE-based dense road segmentation networks over ODE-based dense segmentation networks are: 1) The power-law memory mode can be described by the FODE and has been proven effective in preserving memory. 2) The output of each layer of the proposed network is dependent on the weighted sum of features of previous layers at the same level instead of the concatenation of the features, which reduces the number of learning parameters. The main contributions of the manuscript are highlighted below:

1. This paper proposes a road segmentation network with weighted skip connections which is modelled as a dynamic system using G-L fractional derivative. This work has exploited the memory

property of fractional derivatives in designing the road extraction network.
2. The proposed fractional derivative-based dense network takes into account all past feature vectors and reuse them at forward states, but with a lesser number of parameters and with mathematical interpretations behind it . The weighted sum of outputs of the previous layer is computed in place of their concatenation to reduce the number of parameters. Weights corresponding to these skip connections are dependent on the chosen order of derivative .
3. Experiments have been carried out on two open source databases (Massachusetts Road Database (MRD) and Ottawa Road Database (ORD)) with different road structures and backgrounds. Results show that the proposed model has achieved state-of-the-art performance in terms of recall, F1-score, and mean IoU.

The manuscript has been structured as follows: Section 2 discusses some work related to the study including the basics of fractional calculus and the advantages of using it in the DCNNs. The proposed approach has been explained in Section 3, followed by the experimental results obtained in Section 4, and the conclusion and future scopes are presented in Section 5.

2. Background

This section briefly introduces fractional calculus by providing relevant definitions and the motivation behind its application in modelling real-world problems. It also explains the approach for dynamic system modelling of the forward propagation of Deep Convolutional Neural Networks (DCNNs).

*2.1. **Fractional Calculus***

Fractional calculus is a branch of mathematics that deals with derivatives and integrals of arbitrary order. Fractional calculus is as old as calculus, but due to the involvement of vigorous calculations, this area was considered as the area of theoretical research only [16]. With the introduction of new definitions of fractional derivatives and the availability of software, and computational support, fractional calculus is now being enormously used for modeling real-world problems [84, 85] and thus in Neural Networks [86], Image Processing [53] and various biological evolution [85, 87] based optimization techniques. There exist several ways of defining fractional derivative, but the Grunwald Letnikov derivative which is given by (1) has been used in this study due to its discrete nature.

Grunwald Letnikov derivative. Anton Karl Grünwald and Aleksey Vasilievich Letnikov has given the limit definition of fractional derivative in 1867 and 1868 respectively [17]. Without any assumptions on differentiability of the function for $\alpha \ : \ n-1 < \alpha < n > 0$, the GL derivative of order α for any function $g(t)$ is expressed as

$$ {}_0D_t^{\alpha} g(t) = \lim_{\substack{h \to 0 \\ nh=t}} h^{-\alpha} \sum_{r=0}^{n} (-1)^r \binom{\alpha}{r} g(t - rh) \tag{1} $$

where $\binom{\alpha}{r} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-r+1).\Gamma(r+1)}$, h is the step size and $\Gamma(.)$ is the Gamma function (extension of factorial function for non-integers).

Memory Property. It has been found that memory equipped systems are related with power-law memory, $i.e.$, while defining the current state t of the system, weights of previous states t_j of a system are directly proportional to $(t - t_j)^{\alpha-1}$ where $0 < \alpha < 1$. These power-law memory systems can be described with

Table 1: Comparison of State-of-the-art Road Segmentation Methods with Respect to Architecture and Results on ORD and MRD

SNo.	Methods	Architecture	Database	Results
1.	ResUnet: encoder decoder architecture [42] (2018)	U-Net like architecture with residual skip connections	MRD [21]	breaking point=0.9187 and outperformed SOTA methods : U-Net [24], Saito-CNN [27] and Mnih-CNN [19]
2.	D-linkNet : encoder decoder architecture [39] (2018)	pretained ResNet34 as encoder, and decoder of original linkNet with dilated convolutions	DGRD	Achieved best IoU scores 0.6466 and 0.6342 on the validation and the test set resp.
3.	GL-DenseUNet : encoder decoder architecture [40] (2018)	DenseUnet architecture with global attention unit	Google earth images	Outperformed U-Net [24], FCN [26] and DeepLabV3+ [38] with higher F1-scores
4.	End-to-End road centerline extraction using a confidence map [43] (2018)	The model adopted 13 VGG layers to create an encode-decode network architecture.The model generates multiple scale outputs.	MRD [21]	The model achieves 0.92 on completeness metrics and 0.87 on correctness matrics
5.	DenseUNet : encoder and decoder architecture [49] (2019)	UNet with dense blocks	MRD [21]	Outperformed UNet [24], SegNet [32], GL-DenseUNet [40] with F1-score and mIoU of 74.07% and 74.47% resp.
6.	A Two-Step Deep Convolution Neural Network [51] (2019)	Used two continuous Unet: high precision first Unet and then high recall second Unet.	MRD [21]	The model achieved significant results in terms of accuracy, precision, Recall and F1-score.
7.	VNet: Fully Convolutional Neural Network [60] (2020)	Similar to Unet model, but the left part of the network is split into various phases operating at different resolutions, with new CEDL loss function.	ORD MRD[21]	The model achieved high F1 scores on MRD and ORD.
8.	Generative Adversarial Network [64] (2021)	Used GAN based approach with Modified UNet as generative part	MRD	the approach achieved 91.54 precision, 92.92 recall, MCC of 91.13, mIOU of 87.43 and a F1-score of 92.20.
9.	RoadVecNet: encoder and decoder architecture [65] (2021)	Used two encoders, two decoders, basically two U-nets with dense connections	ORD & MRD	Achieved high F1-scores, MCC and mean IoU on both databases
10.	BMDANet: encoder decoder [71] (2021)	with multi-dimensional attention module	ORD	Outperformed S-O-A methods like UNet [24], D-linknet [39] with 0.9363 F1-score and 0.8802 mIoU.
11.	ConDinet++ : encoder and decoder architecture [67] (2022)	Pretrained VGG16 as encoder for the feature extraction. feature fusion in decoder part by adopting the conditional dilated convolution blocks (CDBs) with the joint loss of cross-entropy loss and Lovasz loss	MRD [21]	the model outperformed several SOTA with high precision, recall, mIOU and F1-Score
12.	SDUNet: encoder and decoder architecture [74] (2022)	DenseUNet with DULR Module, loss function as sum of binary cross entropy loss and dice coeff loss	MRD [21]	Outperformed SOTA like UNet [24], D-linknet [39] with 0.7410 F1-score & 0.7840 mIoU.

the help of FODE [23, 47]. It can be seen from the above expression (1) that the value of the function at all past points t_j are taken into account while calculating the fractional derivative at a point t. Hence, the fractional derivatives are non-local and incorporate memory into the system which is highly significant for a dynamic system.

2.2. Underlying Dynamics and Architecture of DCNN

The output of any layer of a CNN is dependent on the input of that layer and the parameters of the convolution kernel. Thus the propagation of a simple DCNN with T number of layers can be represented by the following evolution process:

$$z_{t+1} = \sigma(\theta_t * z_t) \quad t = 1, 2, \cdots T \tag{2}$$

where $z_t \in \mathbb{R}^n$ is the input of the t_{th} layer, $z_{t+1} \in \mathbb{R}^n$ is the output of the t_{th} layer of the network and $\theta_t \in \mathbb{R}^m$ is the matrix of parameters of the convolution kernel and $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is a non-linear activation function dependent on the input z_t and convolution parameters θ_t. The optimal parameters of the neural network are obtained by minimizing the sum of the loss function and the regularization function of the parameters over all layers. Mathematically, the optimal solution $\theta(t)$ is evaluated by solving the following expression

$$\min_{\{\theta_t\}_{t=1}^T} \sum_{t=1}^{T} R(\theta_t) + L(\Phi(z_T), \mathbf{u}), \tag{3}$$

where $R(\cdot)$ is the regularization function dependent on convolution parameters, $\Phi(z_T) : \mathbb{R}^n \to \mathbb{R}^d$ is the output of the final layer, u is the expected output of the network and $L(\cdot, \cdot) : \mathbb{R}^d \to \mathbb{R}$ is the loss function.

The forward propagation of a DCNN is equivalent to a dynamical system and a dynamic system can be described by an ODE [15, 36]. If $z(t)$ is the corresponding trajectory of the system, then the dynamics of the deep neural network can be represented by the following ODE:

$$\dot{z(t)} = f(z(t), \theta(t)), \qquad z(0) = z_0, \;\; t \in [0, T] \tag{4}$$

where z_0 is the initial condition of the dynamic system and $\theta(t)$ is the control parameter [18] and that system can be solved by the optimal control method. The target is to find the best control $\theta(t)$ for the system at time t. Thus for that the payoff P has to be maximized which is shown in the following equation :

$$P(\theta) = \int_0^T H(z(t), \theta(t)) + J(\Phi(\theta(t)))d\theta \tag{5}$$

where $H(\cdot)$ and $J(\cdot)$ are the running payoff and the terminal payoff respectively [14, 18, 20], which act like the loss and regularization function in (3).

3. Proposed Approach

This section describes the proposed methodology from a mathematical perspective. The proposed architecture has been designed by solving an optimal control problem comprising fractional differential equations. The fractional optimal problem view of road segmentation is presented in this section. Moreover, the architecture of the network has been well explained with the help of mathematical equations.

Table 2: Scalewise Architecture Details of the Proposed Road Segmentation Convolutional Network: The size of input features, number of layers and filters at each level are presented in this table.

Level	Input Size	Number of layers	Number of Filters
1	512×512	4 (both sides)	64
2	256×256	5 (both sides)	128
3	128×128	7 (both sides)	256
4	64×64	10 (both sides)	512
5	32×32	12	1024

3.1. Modeling as Fractional Optimal Control

In this manuscript, the dynamic system corresponding to the propagation in DCNN for road segmentation is represented with the help of a fractional differential system and the optimal control is to be evaluated after solving the corresponding system. Consider the following fractional differential system [16] describing the propagation of deep CNN in which the feature trajectory $z(t, s)$ is assumed to be continuous in time and space:

$$\min_{\theta(t)} \int_{\Omega} L(\Phi(z(T, s)), u(s))\, ds \tag{6}$$

s.t.

$${}_0D_t^{\alpha} z(t, s) = f(z(t, s), \theta(t)), \;\; z(0, s) = \Psi(I(s)), \;\; t \in [0, T] \tag{7}$$

where $s \in \Omega$ is the 2-D spatial position, $I(s)$ is the input road image, $u(s)$ is the ground truth segmented road image and Φ, Ψ are linear transformation , $e.g.$ convolution. The proposed network is designed by solving the above fractional optimal control problem as the fractional-order differential equation has long-term memory. The problem (7) aims to find the optimal control $\theta(t)$ such that the objective loss, $i.e.$, dice coefficient loss in (6) is minimized [31].

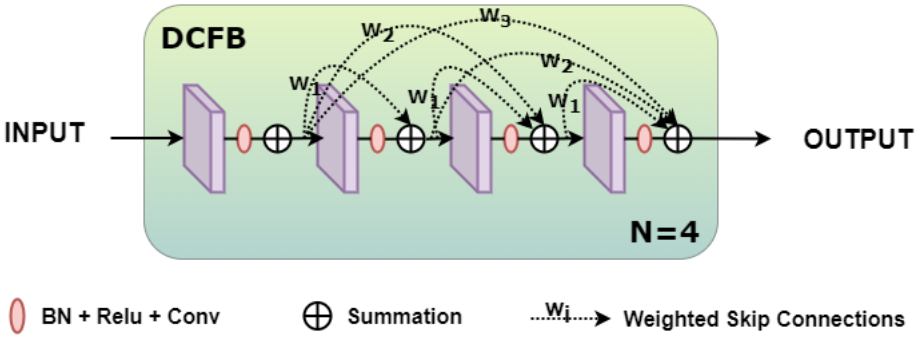


Figure 2: Forward Propagation of features in Densely Connected Fractional Block (DCFB) with 4 layers. The layers of the network are densely connected with the help of fractional-weighted skip connections. The general evolution of the network is represented by (14). BN stands for batch normalization and N stands for number of layers. The layer t is connected with $(t-1)^{th}$, $(t-2)^{th}$.. layers via skip connections with weights w_1, w_2.. respectively.

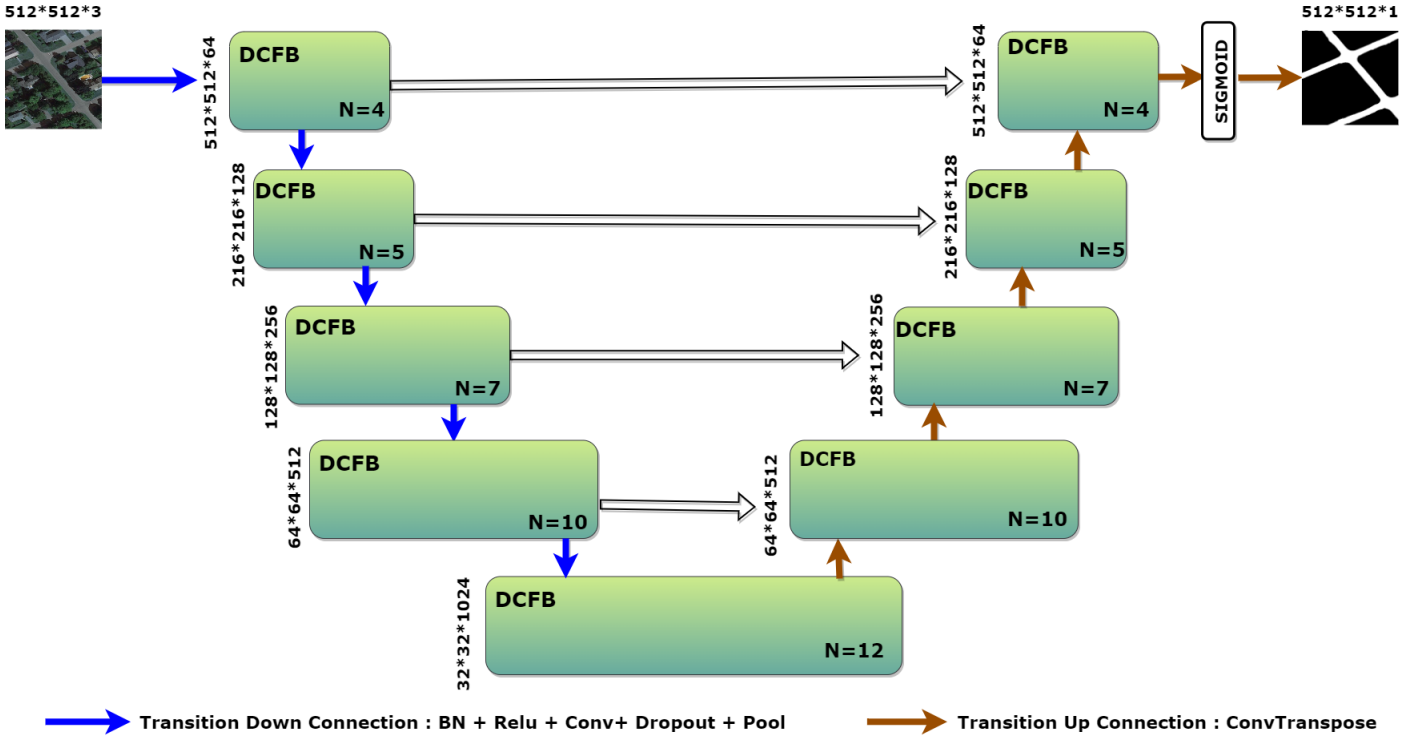


Figure 3: Diagram of the proposed architecture for road segmentation. The architecture is built from the DCFBs shown in Figure 2 with **N** number of layers . The network is composed of a downsampling path with 4 Transitions Down (TD) shown with blue arrows and an upsampling path with 4 Transitions Up (TU) shown with orange arrows. Black horizontal arrows represent skip connections, the feature maps from the downsampling path are added to the corresponding feature maps in the upsampling path. The propagation in between the layers at left are represented by (11) and by (12) for the layers at right with initial conditions given by (13).

3.2. Architecture

The architecture of the proposed model has been designed by discretizing the multi-scale system. In each transition, all the features given by the output of the historical states are being used at each state, and thus the features are being used for a long time. Due to the application of fractional derivatives in the evolution process, the network gets embedded with weighted skip connections and those weights are dependent on the fractional order of differentiation and the current state. The order α is positive and less than 1, due to which the weighted sum remains positive and doesn't explode. For discretization, the step size, h is set to be 1. Then definition (ref1) takes the following form for order $0 < \alpha < 1$:

$$ {}_0D_t^{\alpha} z(t,s) \approx \sum_{k=0}^{t} (-1)^k \binom{\alpha}{k} z(t-k,s) \tag{8} $$

Table 3: Layerwise Architecture Details of the Proposed Road Segmentation Convolutional Network: The network is U-shaped with four transitions down and correspondingly four transitions up. Layer at each level is addresses as Dense Fractional Block (DFB) layer.

Layer	**Description**	**Kernel Size**	**Stride Length**	**Regularizer**	**Padding**	**Dropout**	**Pool Size**	**Pool Stride**
Transition Up	Conv2DTranspose	2×2	2	L2	-	-	-	-
Transition Down	BN+Conv2D	1×1	1	-	-	-	-	-
	Dropout+MaxPool	-	-	-	-	0.2	2×2	2
DFB Layer	BN+Conv2D	3×3	1	-	same	-	-	-
	Dropout	-	-	-	-	0.2	-	-

Thus after combining the above expression (8) with ${}_0D_t^\alpha z(t,s) = f(z(t,s),\theta(t))$ of system (7), we get

$$z(t,s) = f(z(t,s),\theta(t)) - \sum_{k=1}^{t}(-1)^k \binom{\alpha}{k} z(t-k,s) \tag{9}$$

$$z_{t+1} = f(z_t,\theta_t) + \sum_{k=1}^{t}(-1)^{(t-k+1)} \binom{\alpha}{t-k+1} z_k \tag{10}$$

The proposed network is U-shaped, and thus it has four transitions down and correspondingly four transitions up. The detailed scale-wise architectural details $viz.$ number of layers and filters, and input size at a particular transition is presented in Table 2. The multi-scale representation of evolution process can be described by the following system:

$${}_0D_t^\alpha z(t,s,p_{l_i}) = f(z(t,s,p_{l_i})),\theta_{p_{l_i}}(t)) \tag{11}$$

$${}_0D_t^\alpha z(t,s,p_{r_j}) = f(z(t,s,p_{r_j})),\theta_{p_{r_j}}(t)) \tag{12}$$

for each $i \in [1,5]$, $j \in [1,4]$, $t \in [0,T]$ with the following initial conditions

$$z(0,s,p_{r_j}) = T_\uparrow z(n_{r_{j+1}},s,p_{r_{j+1}}) + z(n_{l_j},s,p_{l_j}),$$

$$z(0,s,p_{l_1}) = \Psi(I(s)), \;\; z(0,s,p_{l_i}) = T_\downarrow z(n_{l_{i-1}},s,p_{l_{i-1}}) \tag{13}$$

where $T_\downarrow$ and $T_\uparrow$ is denoting max pooling operation, $\theta_{p_{l_i}}$ is convolution kernel at scale i on the left and $\theta_{p_{l_i}}$ is convolution kernel at scale i on the right. Thus the evolution process of road segmentation network is represented by (14).

$$z_{t+1}^p = \sum_{k=0}^{t} w_k z_k^p + \sigma(\theta_t * (z_t^p))) \tag{14}$$

where σ is a non-linear unit denoting 'Convolution + Batch normalization + Relu', z_{t+1}^p is the output of t^{th} layer of scale p, $w_k = (-1)^{(t-k+1)}\binom{\alpha}{t-k+1}$, and thus $w_t = \alpha$. The forward propagation at each level of encoder/decoder of the network is shown in Figure 2. The architecture of the model can be seen in Figure 3 and the corresponding architectural details related to any layer of the network are presented in

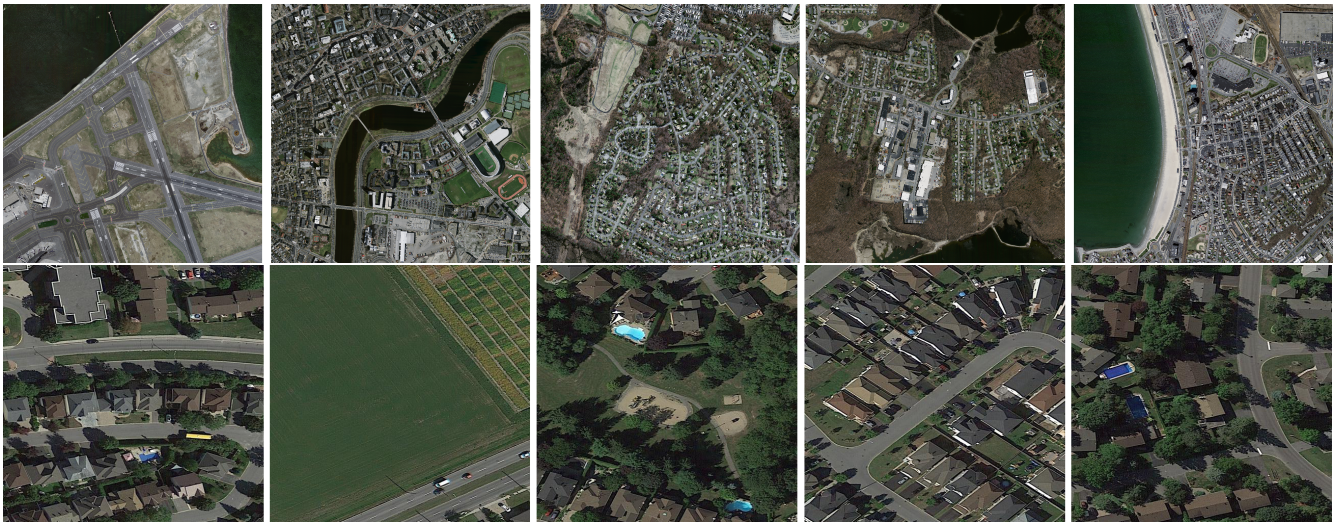


Figure 4: Sample Database Images. First row MRD, and second row ORD.

Table 3. If (x_i, y_i) represents the training set where x_i is the input satellite image and y_i is the corresponding ground truth segmented image. So, if $z_T^i(\Theta)$ is the final output and Θ represents the parameters of the network. Thus dice coefficient loss (DCL) for a particular image sample is given by:

$$DCL(z_T^i(\Theta), y_i) = 1 - DC(z_T^i(\Theta), y_i) = \ 1 - \ \left(\frac{2T_P^i}{2T_P^i + F_P^i + F_N^i}\right) \tag{15}$$

where T_P^i, F_P^i and F_N^i are true positive, false positive and false negative pixels respectively for i^{th} image sample.

The optimized value of network parameters are obtained after solving the above set of fractional differential equations that are representing the evolution process of the proposed network. At each level, the features of each layer are connected with each other with the help of weighted skip connections in accordance with the definition of G-L fractional derivative. This enhances the memory of the system and the features of the present state become more expressive due to the preservation of the semantic and appearance information of the road images.

4. Experimental Results

This section consists of descriptions of the database used, evaluation parameters chosen for assessing the performance of the proposed model, experimental settings, and the results obtained after all the experiments. To show the effectiveness of the proposed model in extracting the roads, experiments have been performed on two benchmark databases. An ablation study has been done by taking the proposed network with different fractional orders, and the performance of the proposed method is compared with some recent road segmentation networks (with the same hyperparameters).

*4.1. **Database***

Two remote sensing databases viz. MRD [21] and ORD [41] have been used to show the performance of the proposed model. Both the databases are open sourced and are being widely used. They have different topographical road maps with variable road-widths and complexities. The detailed information of

both databases is given below:

Massachusetts Road Database [21]. This database consists of 1171 aerial images with 0.5m spatial resolution of the state of Massachusetts[1]. Images cover a wide variety of urban, suburban, and rural regions. Each image is 1500×1500 pixels in size covering a total of 2.25 sq km area of Massachusetts with an approximate road width of 6-9 pixels. In this database, 1108 images are used for training, 49 for testing, and 14 for verification. This database split is kept same as in [74] and other SOTA methods for the comparison of segmentation results. The training images were further augmented by applying four different operations: horizontal/vertical flipping, grid distortion and rotation at 90°. The number of training images became five times of the number of images in the original training database split. The model was then trained using 5540 images. The first row of Figure 4 displays the sample images of this database.

Ottawa Road Database [41]. This database consists of 20 aerial images of several typical urban areas of the Ottawa state, obtained using Google Earth. The images are of 0.21m spatial resolution per pixel (zoom level 19) with variable size, some being of size 4500×5500 while others being of 1000×2000. To compare the segmentation results with the SOTA [71], the split of database is set to be as 14 images for training, 3 images for validating, and 3 for testing the model. For training the proposed network, the 14 training images were randomly segmented into patches of size 512×512, then the obtained patches were augmented by applying five different operations: horizontal / vertical flipping and rotation at 90°, 180° and 270°. On an average, 280 patches corresponding to each image of the training set were obtained and the model was then trained using 3946 images. This database is challenging as it has images with high complexity and covers several urban areas. The second row of Figure 4 shows some sample images of this database.

*4.2. **Evaluation Parameters***

Road segmentation can be viewed as a binary classification problem, i.e., pixels corresponding to roads are positive and the rest of the pixels are negative. According to this, the pixels of the predicted segmented output are divided into four categories in comparison with the actual segmented ground truth: true positive (T_P), false positive (F_P), true negative (T_N), and false negative (F_N). To assess the performance of the proposed model for road segmentation, the standard evaluation metrics have been chosen, which are briefly explained below.

Recall (R). It is the ratio of successfully retrieved relevant pixels to the total relevant pixels which can be expressed as

$$\text{Recall} = T_P/(T_P + F_N) \tag{16}$$

Precision (P). This metric is also based on relevance, which is given by the ratio of successfully retrieved relevant pixels to the total retrieved pixels. Thus it can be expressed as

$$\text{Precision} = T_P/(T_P + F_P) \tag{17}$$

Dice Coefficient/ F1-score. This metric is used for gauging the similarities between the output mask and the expected output, it is defined as the harmonic mean of precision and recall of the classifier. The output segmented road mask is a kind of Boolean data, thus dice coefficient can be expressed in the following

[1] [Online]. Available: http://www.cs.toronto.edu/vmnih/data/

Table 4: Comparison of Results Obtained by Proposed Road Segmentation Network with SOA methods on Massachusetts Road Database

Model	Precision	Recall	mIoU	F1-score
U-Net [24]	0.747	0.721	0.722	0.682
Dlinknet [39]	0.767	0.741	0.737	0.717
HsgNet [52]	0.769	0.752	0.749	0.720
Dense-UNet [74]	0.780	0.731	0.739	0.714
SUNet [74]	0.798	0.736	0.753	0.721
SDUNet [74]	**0.812**	0.757	0.784	0.741
Proposed (α =0.4)	0.698	**0.830**	**0.787**	**0.748**

Table 5: Comparison of Results Obtained by Proposed Road Segmentation Network with SOA methods on ORD

Model	Precision	Recall	mIoU	F1-score
DeeplabV3 [34]	-	-	0.8362	0.9146
MACUNet [63]	-	-	0.8419	0.9142
DeeplabV3+ [34]	-	-	0.8473	0.9174
D-LinkNet [39]	-	-	0.8360	0.9107
UNet++ [45]	-	-	0.8523	0.9203
BMDANet [71]	-	-	0.8802	0.9363
Proposed (α =0.5)	0.8867	0.9374	**0.9062**	0.9110

ways

$$\text{F1-score/ Dice Coefficient} = (2 \cdot P \cdot R + 1)/(P + R + 1) \; = \; 2T_P/(2T_P + F_P + F_N) \tag{18}$$

where P and R are precision and recall respectively.

Accuracy (Acc). It is a measure of the percentage of pixels predicted correctly.

Mean Intersection over Union (mIoU): The IoU is another metric like dice coefficient, that measures the percentage of overlap for the target mask and the predicted output. It can be evaluated by the following expression (19). In other words, it is the ratio of the number of pixels that are common between the expected output and the predicted masks, and the total pixels across both the masks.

$$\text{IoU} = (target \cap prediction)/(target \cup prediction) \tag{19}$$

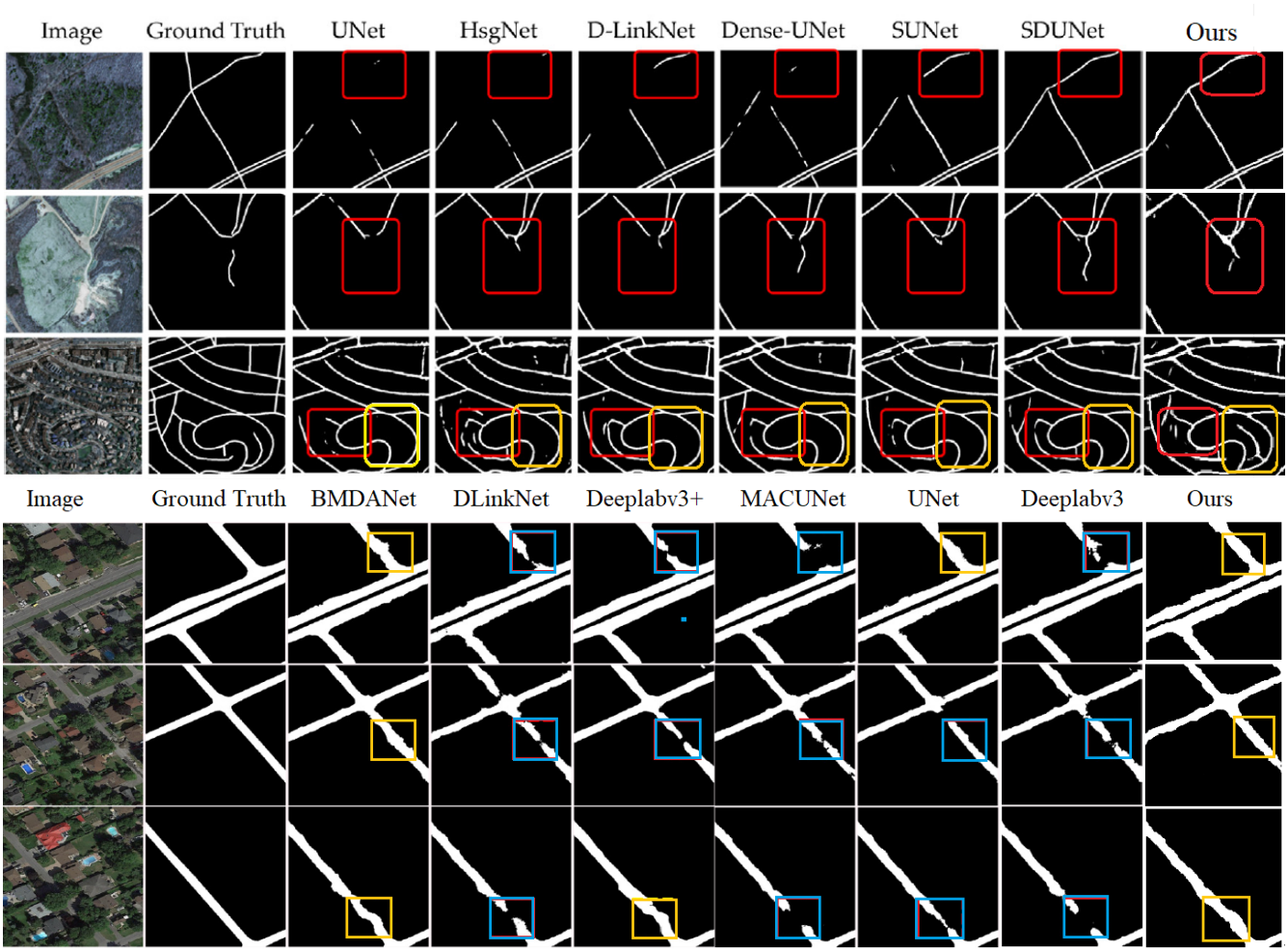


Figure 5: Visual results on Proposed Networks and other existing networks on MRD (top) and ORD (bottom): Yellow and Red boxes show good segmentation results whereas results in the blue boxes are not good for ORD and MRD.

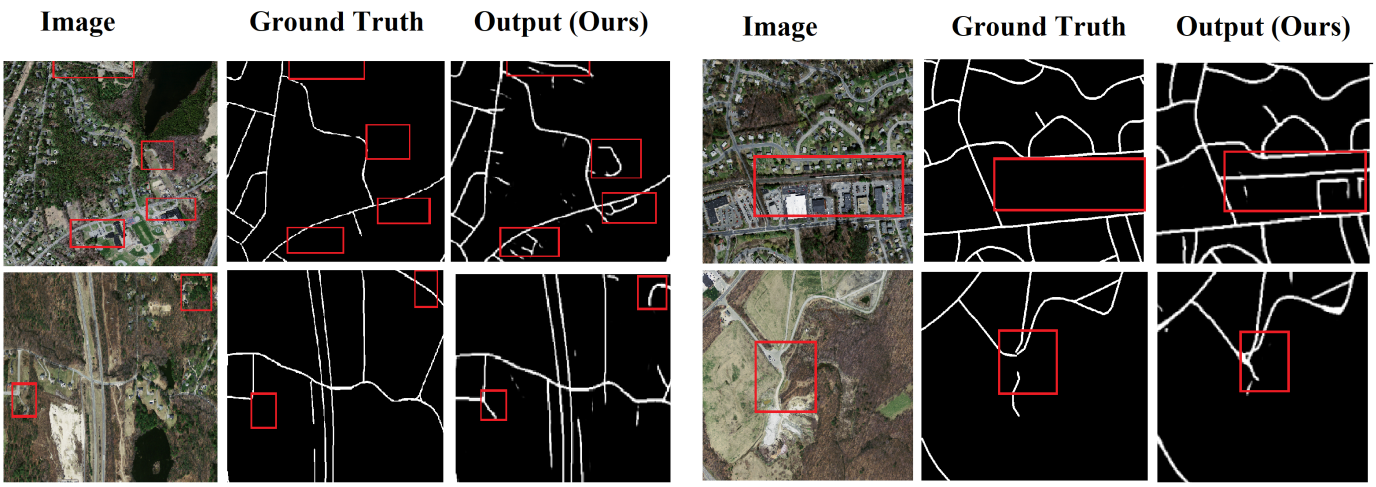


Figure 6: Visual results obtained by Proposed Network on some mislabeled data of MRD. The red boxes are draw over regions of interest. These areas are labelled as non-road in ground truth. The proposed network is correctly labelling these roads in the boxes.

*4.3. **Experimental Settings***

The performance of the model has been evaluated on two databases. The model has been trained from scratch without using any extra data, post-processing module, or pre-trained weights. For MRD, the training data has been generated by horizontal/vertical flipping and rotating the 1108 images and then resizing to 512×512. The actual images of ORD are quite large, thus augmentation techniques are applied after dividing the images into segments of size to 512×512. All the experiments have been performed on a Linux-based operating system with NVIDIA GeForce GTX 1080 Ti graphic card with graphic memory of 11 GB. The model has been initialized using the HeNormal initialization method [25] and the Adam optimizer [22] with default settings has been used for minimizing the loss function and find the optimal control parameters. The learning rate is set to be 0.0001 with a decay of 0.1 every 20 epochs. The validation set has also been set for early stopping and the drop-out rate is set to be 0.2 to avoid any over-fitting. The total number of epochs was set to be 200.

*4.4. **Results***

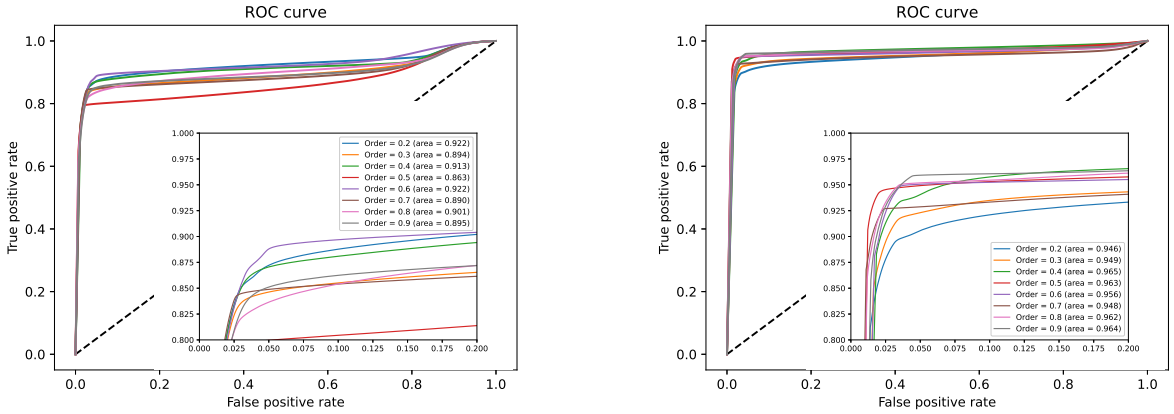
To evaluate the road segmentation performance of the proposed network, the ground truth segmented images are compared with the corresponding segmented outputs of the model. We have computed the F1-score, recall, precision, and mean IoU on both databases to assess the quantitative performance. The proposed network is dense, thus here, the comparison of the quantitative results has been made with state-of-the-art road segmentation dense networks. The proposed network ($\alpha = 0.4$) on MRD are compared with the benchmarking segmentation network U-Net [24], Dlinknet [39], HsgNet [52], and dense networks viz. Dense-UNet [74], SUNet [74], and SDUNet [74]. The values of evaluation parameters corresponding to these SOTA methods are extracted from [74]. It can be noted from the Table 4 that SDUNet model had achieved the maximum precision, recall, mean IoU and F1-score of 0.812,0.757, 0.784 and 0.741 respectively. The precision of the proposed network is found to be 0.698 which is comparatively lower for MRD due to the extraction of some false regions, but the overall performance is significantly good. The results in the Table 4 clearly show that the proposed network achieved 0.830 mean IoU, 0.787 F1-score and recall 0.748 and these metrics are higher than SDUNet [74] and hence than the other aforementioned methods. From the visual results in Figure 5 it can be seen that extracted road networks are

Table 6: Results of the proposed model with different fractional order on MRD and ORD

Model	Precision	Recall	F1-Score	Acc	mIoU	Precision	Recall	F1-score	Acc	mIoU
	Database: Massachusetts Road Database					Database: Ottawa Road Database				
α =0.2	0.686	0.810	0.741	0.971	0.776	0.836	0.881	0.859	0.962	0.851
α =0.3	0.682	0.802	0.737	0.971	0.764	0.829	0.904	0.867	0.963	0.864
α =0.4	0.698	**0.830**	**0.748**	**0.978**	**0.787**	0.822	0.926	0.873	0.964	0.876
α =0.5	**0.736**	0.764	0.741	0.973	0.772	**0.887**	0.937	**0.911**	**0.976**	**0.906**
α =0.6	0.675	0.811	0.735	0.970	0.758	0.829	0.936	0.882	0.966	0.879
α =0.7	0.679	0.821	0.721	0.970	0.783	0.868	0.921	0.864	0.971	0.889
α =0.8	0.699	0.771	0.729	0.970	0.758	0.833	**0.941**	0.875	0.967	0.882
α =0.9	0.657	0.799	0.715	0.968	0.754	0.796	0.951	0.855	0.962	0.874

mostly continuous with very few breaking points even in the presence of complex backgrounds. The third row of images clearly shows the superiorty of our model than other existing models. The presence of weighted skip connections in the proposed dense network has helped in gathering more global information as compared to other dense networks $viz.$ Dense-UNet [74], Dlinknet [39], SDUNet [74].
But the precision of the proposed model is less for MRD as compared other models. By lower precision, it means that the model returns comparatively more false positives than other models. But due to high recall, the model has low false negative rate. Thus our model is more capable to extract the pixels related to roads as compared to other models. But in the MR database, the driveways are of same color as roads. Secondly, the MR database contain some errors, and imprecision in ground truth where some of the road pixels are mislabeled as non-road pixels due to occlusion of trees, clouds and shadows. The mislabeled data affects the pixel classification accuracy [88, 89]. Our model intends to label these road pixels correctly. Thus these two issues lead to increase in false positives with respect to the ground truth. Hence Precision metric is lower in our model. Few samples where road pixels are mislabelled and the corresponding segmentation results obtained by the proposed model are shown in Figure 6.

The quantitative results of the final proposed network ($\alpha = 0.5$) on ORD are compared with other state-of-the-art road segmentation networks $viz.$ Deeplabv3 [34], MACUNet [63], DeeplabV3+ [34], D-LinkNet [39], UNet++ [45], and BMDANet [71] and the results are reported in the Table 5. The values of F1-score and mean IoU corresponding to the aforementioned methods are extracted from [71]. It can be observed from the table that the BMDANet [71] had achieved the maximum values of mean IoU and F1-score as 0.8802 and 0.9363 respectively. The results clearly show that the proposed network achieved a higher mean IoU of 0.9062 as compared to BMDANet [71] and hence than the other mentioned segmentation methods. The proposed method improved the segmentation results of SOTA for ORD by 3% in terms of mean IoU. The F1-score of the model is found to be 0.9110 which is comparable to other SOTA methods. The model achieved precision and recall of 0.8867 and 0.9374 respectively which depicts a significantly good performing binary classifier. From the visual results in Figure 5 it can be seen that extracted road networks are continuous and the edge information are preserved even in the presence of occlusions like trees and buildings. The effectiveness of using non-local fractional derivatives in the proposed network can be observed from the segmented outputs as there is less loss of information.
Ablation Study. The performance of the proposed network depends on the feature extracted and thus on the feature propagation between layers. The features obtained from each layer are being reused at future states with the help of fractional weighted skip connections. Therefore, an ablation study has been done



(a) ROC plots for MRD with zoomed view of upper left corner of the graphs: Largest area under the curve corresponding to order= 0.6

(b) ROC plots for ORD with zoomed view of upper left corner of the graphs: Largest area under the curve corresponding to order= 0.4

Figure 7: The receiver operating characteristic (ROC) curve shows the relationship between true positive rate and false positive rate. ROC curves for both databases have been plotted at different fractional orders (hence with different weighted connections). The curves are drawn and the corresponding area under the graphs is also mentioned. The more the area under the curves, the better the segmentation performance of the network.

to see the effect of varying the fractional order α on the performance of the model. Experiments have been carried on both databases for $\alpha = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ and 0.9. To analyze the performance of the network with different fractional weights, we have computed mean IoU, F1-score, recall, precision, and pixel-wise accuracy on the databases. The Table 6 shows the obtained results on MRD and ORD. The values of the evaluation parameters for MRD depict that the model performed well for fractional orders $\alpha : 0.2 \leq \alpha \leq 0.7$. It has been found that the model achieved highest precision of 0.736 at fractional order $\alpha = 0.5$ on MRD. However, the model achieved higher recall, F1-score, accuracy and mean IoU of 0.830, 0.748, 0.978 and 0.787 respectively on order $\alpha = 0.4$ for MRD. Furthermore, the model performed better at $\alpha = 0.5$ for ORD with higher precision, F1-score, accuracy, and mean IoU of 0.887, 0.911, 0.976 and 0.906 respectively. The recall value at $\alpha = 0.5$ is found to be 0.937 that is close to the highest recall value of 0.941 attained at $\alpha = 0.8$ for ORD. ROC curves shown in Figure 7 are also drawn to evaluate the performances of the models. It depicts the classification performance of the model at different thresholds for different fractional orders. It is a graph with the false positive rate on the x-axis and the true positive rate on the y-axis. Area under the curve are computed for each curve that signifies the degree of separability of road and non-road pixels in the database. The more the area under the curves, the better the segmentation (binary classifier). The curve corresponding to order $\alpha = 0.5$ acquired the highest area of 0.965 under it for ORD. Thus all the evaluation metrics are aligned with graphical results for ORD. The curves corresponding to order $\alpha = 0.2, 0.6$ acquired the highest area under them of 0.922 for MRD, but the rest of the evaluation metrics are found to be: precision 0.686, recall 0.810, F1-score 0.741, accuracy 0.971 and mean IoU 0.776 for $\alpha = 0.2$, and precision 0.675, recall 0.811, F1-score 0.735, accuracy 0.970 and mean IoU 0.758 for $\alpha = 0.6$, which are lesser for these orders. The curve corresponding to $\alpha = 0.4$ acquired the second highest area of 0.913 under it with the highest values of other evaluation metrics. Thus the model with order $\alpha = 0.4$ and $\alpha = 0.5$ are used throughout the experiments for MRD and ORD respectively.

5. Conclusion

In this manuscript, a dense network with weighted skip connections has been proposed for the segmentation of roads from high-resolution satellite images. The proposed network has been designed by solving a fractional optimal control problem. The proposed approach involves the weighted sum of the output of previous layers at a particular transition, unlike basic dense networks where the output of the previous layers are concatenated. Thus the proposed method is computationally efficient. Moreover, the performance of the proposed network was found to be better than other state-of-the-art methods in terms of F1-score and the mean IoU at fractional order 0.4 for MRD and at 0.5 for ORD than the state-of-the-art segmentation results. By exploiting the memory property of fractional derivatives, the forward propagation of features has strengthened and lesser loss of information has been seen. The results show that the extracted roads are more complete for both the databases. Moreover, the model has performed better than other SOTA methods on MRD in terms of higher recall, F1-score and mean IoU of 0.830, 0.748 and 0.787 respectively. In addition, the model achieved better mean IoU of 0.9062 that shows the improvement of around 3 % in segmentation results with respect to the SOTA method on ORD.

Two possible areas for improvement have been identified: 1) The driveways are of same color as that of the actual roads, due to which the pixels related to driveways are labelled as roads. This increases the number of false positive pixels. 2) Several trees and their shades falling on the roads cause occlusion. The proposed model capture local and global information simultaneously and is capable of extracting the pixels of roads occluded by the trees. But while solving the issue of occlusion, the model labels some pixels corresponding to trees nearby the roads as road-pixels and hence the sharpness of edges of the roads gets affected. This again increases the false positive pixels. More false negatives affects the segmentation performance of the model hence hampers the precision metric. In future, these problems can be targeted to obtain continuous and accurate road structures.

References

[1] G. P. Cardim, E. A. d. Silva, M. A. Dias, I. Bravo, A. Gardel, Statistical evaluation and analysis of road extraction methodologies using a unique dataset from remote sensing, Remote Sensing 10 (4) (2018). doi:10.3390/rs10040620.

[2] R. Kestur, S. Farooq, R. Abdal, E. Mehraj, O. S. Narasipura, M. Mudigere, UFCN: a fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle, Journal of Applied Remote Sensing 12 (1) (2018) 016020.

[3] H. Verma, Y. Sharma, S. Pasari, Synthetic aperture radar interferometry to measure earthquake-related deformation: A case study from nepal, in: Disaster Management in the Complex Himalayan Terrains, Springer, 2022, pp. 133–140.

[4] P. K. Soni, N. Rajpal, R. Mehta, Road centerline extraction from VHR images using SVM and multi-scale maximum response filter, Journal of the Indian Society of Remote Sensing 49 (7) (2021) 1519–1532.

[5] M. Barzohar, D. B. Cooper, Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (7) (1996) 707–721.

[6] S. Hinz, Automatic road extraction in urban scenes and beyond, International Archives of Photogrammetry and Remote Sensing 35 (PART B3) (2004) 349–355.

[7] A. Boyko, T. Funkhouser, Extracting roads from dense point clouds in large scale urban environment, ISPRS Journal of Photogrammetry and Remote Sensing 66 (6) (2011) S2–S12.

[8] Y. Wang, Y. Tian, X. Tai, L. Shu, Extraction of main urban roads from high resolution satellite images by machine learning, in: Asian Conference on Computer Vision, Springer, 2006, pp. 236–245.

[9] D. Geman, B. Jedynak, An active testing model for tracking roads in satellite images, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (1) (1996) 1–14.

[10] A. A. Matkan, M. Hajeb, S. Sadeghian, Road extraction from lidar data using support vector machine classification, Photogrammetric Engineering & Remote Sensing 80 (5) (2014) 409–422.

[11] P. P. Singh, R. Garg, Automatic road extraction from high resolution satellite image using adaptive global thresholding and morphological operations, Journal of the Indian Society of Remote Sensing 41 (3) (2013) 631–640.

[12] G. Fu, H. Zhao, C. Li, L. Shi, Road detection from optical remote sensing imagery using circular projection matching and tracking strategy, Journal of the Indian Society of Remote Sensing 41 (4) (2013) 819–831.

[13] P. K. Soni, N. Rajpal, R. Mehta, A comparison of road network extraction from high resolution images, in: 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), IEEE, 2018, pp. 525–531.

[14] L. C. Evans, An introduction to mathematical optimal control theory version 0.2, Lecture notes available at http://math. berkeley. edu/˜ evans/control. course. pdf (1983).

[15] F. J. Pineda, Dynamics and architecture for neural computation, Journal of Complexity 4 (3) (1988) 216–245.

[16] I. Podlubny, Fractional-order systems and $PI^{\lambda} D^{\mu}$-controllers, IEEE Transactions on automatic control 44 (1) (1999) 208–214.

[17] I. Podlubny, Fractional Differential Equations, Mathematics in Science and Engineering (1999).

[18] L. C. Evans, An introduction to mathematical optimal control theory, Lecture Notes, University of California, Department of Mathematics, Berkeley 3 (2005) 15–40.

[19] V. Mnih, G. E. Hinton, Learning to detect roads in high-resolution aerial images, in: European conference on computer vision, Springer, 2010, pp. 210–223.

[20] J. Macki, A. Strauss, Introduction to optimal control theory, Springer Science & Business Media, 2012.

[21] V. Mnih, Machine learning for aerial image labeling, University of Toronto (Canada), 2013.

[22] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[23] M. Edelman, Fractional maps as maps with power-law memory, in: Nonlinear dynamics and complexity, Springer, 2014, pp. 79–120.

[24] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[25] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026–1034.

[26] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[27] S. Saito, T. Yamashita, Y. Aoki, Multiple object extraction from aerial imagery with convolutional neural networks, Electronic Imaging 2016 (10) (2016) 1–9.

[28] Z. Zhong, J. Li, W. Cui, H. Jiang, Fully convolutional networks for building and road extraction: Preliminary results, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2016, pp. 1591–1594.

[29] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, P. Eklund, A review of road extraction from remote sensing images, Journal of traffic and transportation engineering (english edition) 3 (3) (2016) 271–282.

[30] G. Larsson, M. Maire, G. Shakhnarovich, FractalNet: ultra-deep neural networks without residuals, arXiv preprint arXiv:1605.07648 (2016).

[31] H. M. Ali, F. L. Pereira, S. M. Gama, A new approach to the pontryagin maximum principle for nonlinear fractional optimal control problems, Mathematical Methods in the Applied Sciences 39 (13) (2016) 3640–3649.

[32] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence 39 (12) (2017) 2481–2495.

[33] Y. Wei, Z. Wang, M. Xu, Road structure refined CNN for road extraction in aerial image, IEEE Geoscience and Remote Sensing Letters 14 (5) (2017) 709–713.

[34] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587 (2017).

[35] A. N. Gomez, M. Ren, R. Urtasun, R. B. Grosse, The reversible residual network: Backpropagation without storing activations, Advances in neural information processing systems 30 (2017).

[36] E. Weinan, A proposal on machine learning via dynamical systems, Communications in Mathematics and Statistics 1 (5) (2017) 1–11.

[37] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: learning where to look for the pancreas, arXiv preprint arXiv:1804.03999 (2018).

[38] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

[39] L. Zhou, C. Zhang, M. Wu, D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 182–186.

[40] Y. Xu, Z. Xie, Y. Feng, Z. Chen, Road extraction from high-resolution remote sensing imagery using deep learning, Remote Sensing 10 (9) (2018) 1461.

[41] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, Y. Liu, RoadNet: learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images, IEEE Transactions on Geoscience and Remote Sensing 57 (4) (2018) 2043–2056.

[42] Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual U-Net, IEEE Geoscience and Remote Sensing Letters 15 (5) (2018) 749–753.

[43] W. Yujun, H. Xiangyun, G. Jinqi, End-to-end road centerline extraction via learning a confidence map, in: 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), IEEE, 2018, pp. 1–5.

[44] A. Buslaev, S. Seferbekov, V. Iglovikov, A. Shvets, Fully convolutional network for automatic road extraction from satellite imagery, in: Computer Vision and Pattern Recognition workshops, IEEE, 2018, pp. 207–210.

[45] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A nested U-Net architecture for medical image segme in: Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 2018, pp. 3–11.

[46] Y. Lu, A. Zhong, Q. Li, B. Dong, Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations, in: International Conference on Machine Learning, PMLR, 2018, pp. 3276–3285.

[47] M. Edelman, Universality in systems with power-law memory and fractional dynamics, Chaotic, fractional, and complex dynamics: New insights and perspectives (2018) 147–171.

[48] B. Artacho, A. Savakis, Waterfall atrous spatial pooling architecture for efficient semantic segmentation, Sensors 19 (24) (2019) 5361.

[49] J. Xin, X. Zhang, Z. Zhang, W. Fang, Road extraction of high-resolution remote sensing images derived from DenseUNet, Remote Sensing 11 (21) (2019) 2499.

[50] Y. Li, L. Xu, J. Rao, L. Guo, Z. Yan, S. Jin, A Y-Net deep learning method for road segmentation using high-resolution visible remote sensing images, Remote sensing letters 10 (4) (2019) 381–390.

[51] P. Singh, R. Dash, A two-step deep convolution neural network for road extraction from aerial images, in: 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2019, pp. 660–664.

[52] Y. Xie, F. Miao, K. Zhou, J. Peng, HsgNet: a road extraction network based on global perception of high-order spatial information, ISPRS International Journal of Geo-Information 8 (12) (2019) 571.

[53] X. Jia, S. Liu, X. Feng, L. Zhang, FOCNet: a fractional optimal control network for image denoising, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6054–6063.

[54] J.-Y. Sun, S.-W. Kim, S.-W. Lee, Y.-W. Kim, S.-J. Ko, Reverse and boundary attention network for road segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, pp. 0–0.

[55] M. Yuan, Z. Liu, F. Wang, Using the wide-range attention u-net for road segmentation, Remote Sensing Letters 10 (5) (2019) 506–515.

[56] L. Ruthotto, E. Haber, Deep neural networks motivated by partial differential equations, Journal of Mathematical Imaging and Vision 62 (3) (2020) 352–364.

[57] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, A. Alamri, Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review, Remote Sensing 12 (9) (2020) 1444.

[58] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, UNet 3+: a full-scale connected unet for medical image segmentation, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 1055–1059.

[59] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, M. Jagersand, U2-Net: going deeper with nested u-structure for salient object detection, Pattern recognition 106 (2020) 107404.

[60] A. Abdollahi, B. Pradhan, A. Alamri, VNet: an end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data, IEEE Access 8 (2020) 179424–179436.

[61] M. Lan, Y. Zhang, L. Zhang, B. Du, Global context based automatic road segmentation via dilated convolutional neural network, Information Sciences 535 (2020) 156–171.

[62] P. Shamsolmoali, M. Zareapoor, H. Zhou, R. Wang, J. Yang, Road segmentation for remote sensing images using adversarial spatial pyramid networks, IEEE Transactions on Geoscience and Remote Sensing 59 (6) (2020) 4673–4688.

[63] R. Li, S. Zheng, C. Duan, J. Su, C. Zhang, Multistage attention ResU-Net for semantic segmentation of fine-resolution remote sensing images, IEEE Geoscience and Remote Sensing Letters 19 (2021) 1–5.

[64] A. Abdollahi, B. Pradhan, G. Sharma, K. N. A. Maulud, A. Alamri, Improving road semantic segmentation using generative adversarial network, IEEE Access 9 (2021) 64381–64392.

[65] A. Abdollahi, B. Pradhan, A. Alamri, RoadVecNet: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up, GIScience & Remote Sensing 58 (7) (2021) 1151–1174.

[66] A. Abdollahi, B. Pradhan, Integrated technique of segmentation and classification methods with connected components analysis for road extraction from orthophoto images, Expert Systems with Applications 176 (2021) 114908.

[67] K. Yang, J. Yi, A. Chen, J. Liu, W. Chen, ConDinet++: full-scale fusion network based on conditional dilated convolution to extract roads from remote sensing images, IEEE Geoscience and Remote Sensing Letters 19 (2021) 1–5.

[68] L. Zhang, M. Lan, J. Zhang, D. Tao, Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images, IEEE Transactions on Geoscience and Remote Sensing 60 (2021) 1–13.

[69] Z. Sun, H. Geng, Z. Lu, R. Scherer, M. Woźniak, Review of road segmentation for sar images, Remote Sensing 13 (5) (2021) 1011.

[70] C. Ding, L. Weng, M. Xia, H. Lin, Non-local feature search network for building and road segmentation of remote sensing image, ISPRS International Journal of Geo-Information 10 (4) (2021) 245.

[71] S. Dong, Z. Chen, Block multi-dimensional attention for road segmentation in remote sensing imagery, IEEE Geoscience and Remote Sensing Letters 19 (2021) 1–5.

[72] Z. Zhang, C. Miao, C. Liu, Q. Tian, DCS-TransUperNet: road segmentation network based on cswin transformer with dual resolution, Applied Sciences 12 (7) (2022) 3511.

[73] S. Gong, H. Zhou, F. Xue, C. Fang, Y. Li, Y. Zhou, FastRoadSeg: fast monocular road segmentation network, IEEE Transactions on Intelligent Transportation Systems (2022).

[74] M. Yang, Y. Yuan, G. Liu, SDUNet: road extraction via spatial enhanced and densely connected UNet, Pattern Recognition 126 (2022) 108549.

[75] S. Arora, T. Mathur, S. Agarwal, K. Tiwari, P. Gupta, Applications of fractional calculus in computer vision: A survey, Neurocomputing 489 (2022) 407–428.

[76] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[77] X. Pan, H.-B. Shen, Predicting rna–protein binding sites and motifs through combining local and global deep convolutional neural networks, Bioinformatics 34 (20) (2018) 3427–3436.

[78] C. R. Qi, H. Su, K. Mo, L. J. Guibas, PointNet: deep learning on point sets for 3D classification and segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652–660.

[79] X. Li, M. Chen, F. Nie, Q. Wang, A multiview-based parameter free framework for group detection, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[80] V. Buhrmester, D. Münch, M. Arens, Analysis of explainers of black box deep neural networks for computer vision: A survey, Machine Learning and Knowledge Extraction 3 (4) (2021) 966–989.

[81] R. Shwartz-Ziv, N. Tishby, Opening the black box of deep neural networks via information, arXiv preprint arXiv:1703.00810 (2017).

[82] Y.-h. Sheu, Illuminating the black box: interpreting deep neural network models for psychiatric research, Frontiers in Psychiatry 11 (2020) 551299.

[83] R. Caponetto, Fractional order systems: modeling and control applications, Vol. 72, World Scientific, 2010.

[84] Y. Luchko, Operational calculus for the general fractional derivative and its applications, Fractional Calculus and Applied Analysis 24 (2) (2021) 338–375.

[85] Y. Chen, F. Liu, Q. Yu, T. Li, Review of fractional epidemic models, Applied mathematical modelling 97 (2021) 281–307.

[86] N. Singh, S. Sugandha, T. Mathur, S. Agarwal, K. Tiwari, Stock price prediction using fractional gradient-based long short term memory, in: Journal of physics: Conference series, Vol. 1969, IOP Publishing, 2021, p. 012038.

[87] M. S. Couceiro, R. P. Rocha, N. Ferreira, J. Machado, Introducing the fractional-order Darwinian PSO, Signal, Image and Video Processing 6 (3) (2012) 343–350.

[88] M. J. Carlotto, Effect of errors in ground truth on classification accuracy, International Journal of Remote Sensing 30 (18) (2009) 4831–4849.

[89] A. E. Maas, F. Rottensteiner, C. Heipke, A label noise tolerant random forest for the classification of remote sensing data based on outdated maps for training, Computer Vision and Image Understanding 188 (2019) 102782.