Int. J. Nonlinear Anal. Appl. 15 (2024) 3, 125–140
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2023.30232.4371

Improving image segmentation using artificial neural networks and evolutionary algorithms

Mohammadreza Fadavi Amiri[a,*], Maral Hosseinzadeh[a], Seyyed Mohammad Reza Hashemi[b]

[a]Faculty of Computer Engineering, Shomal University, Amol 46161-84596, Mazandaran, Iran

[b]Faculty of Computer Engineering Department, Shahrood University of Technology, Shahrood, Semnan, Iran

(Communicated by Seyed Hossein Siadati)

Abstract

Image segmentation can be used in object recognition systems. Today, it is considered in most branches of science and industry, and in many of these branches the identification of the main components of the image is very important. For example, automatic detection and tracking of moving targets in military applications and segregation of different products in industrial applications, identification of road signs, segmentation of colonies, land use and land cover classification. It is also widely used in medicine, such as diagnosing brain and tumors and self-driving. In this study, image sections are performed by a feature extraction process using a neural network. In the process of applying the neural network method, optimization was performed using the ant colony algorithm. The results show that the identification of image segments using the neural network has an accuracy of 87

Keywords: Image segmentation, Neural network, Ant colony optimization algorithm
2020 MSC: 68T07

1 Introduction

Image segmentation is the process of dividing and segmenting an image into several parts. In this process, each pixel of the image is assigned to an object. CNN[1] is a popular deep learning architecture that automatically learns useful feature representations directly from image data [14, 15]. It encodes a set of image features in a compact representation suitable for image classification and image retrieval. Pattern matching uses a small image or pattern to find matching regions in a larger image [1, 25].

Image segmentation is the first step and the most critical step of image analysis, the purpose of which is to extract information inside the image (such as edges, views, and the identity of each area) [40]. Which through description, prepares the obtained regions to reduce them in a form suitable for computer processing and recognition of each region. The accuracy of image segmentation has a direct effect on the efficiency of the entire system, so that it can determine the possible success or failure of the final analysis of the image.

Nowadays, semantic segmentation - applied to still 2D images, videos, and even 3D or volumetric data is one of the key problems in the field of computers vision. Looking at the big picture, semantic segmentation is one of the high-level

*Corresponding author
Email addresses: fadavi@shomal.ac.ir (Mohammadreza Fadavi Amiri), maral.hsz95@gmail.com (Maral Hosseinzadeh), smr.hashemi@shahroodut.ac.ir (Seyyed Mohammad Reza Hashemi)

[1]Convolutional Neural Network

Received: January 2023 *Accepted:* April 2023

tasks that paves the way towards complete scene understanding. The importance of the scene understanding as a core computer vision problem is highlighted by the fact that an increasing number of applications nourish from inferring knowledge from imagery. Some of those applications include autonomous driving [8, 10, 7], human-machine interaction [22], computational photography [35], image search engines [33], and augmented reality to name a few. Such problem has been addressed in the past using various traditional computer vision and machine learning techniques. Despite the popularity of those kind of methods, the deep learning revolution has turned the tables so that many computer vision problems – semantic segmentation among them – are being tackled using deep architectures, usually Convolutional Neural Networks (CNNs) [21, 6, 9, 13, 12], which are surpassing other approaches by a large margin in terms of accuracy and sometimes even efficiency.

One of the main challenges of artificial neural networks is how to update its weights during training. In order to increase the accuracy of learning in neural networks, this network can be combined with meta-innovative or evolutionary algorithms. In order to increase the accuracy of learning in neural networks, this network can be combined with meta-innovative or evolutionary algorithms that is for the weighting of different existing layers that are related to ridges, the values of weights or ridges can be compared with algorithms determined meta-initiative.

Using the combination of artificial neural network and evolutionary algorithms that can improve the efficiency of image segmentation. Neural network design involves choosing an optimal set of design parameters to achieve fast convergence during training and the required accuracy during recall. Using evolutionary algorithms, we try to determine the best parameters for the neural network in each problem and improve the efficiency of the neural network in comparison with other optimization methods.

2 Previous works

Due to its excellent structure, convolutional neural network has achieved remarkable results in image classification, segmentation and recognition. The following section presents recent developments using convolutional neural network in semantic image segmentation.

2.1 Fully convolutional neural networks

The paper [19] is the first work that introduces ANNFCN to image segmentation area. The main insight is the replacement of fully connected layer by fully convolutional layer. The last fully connected layer was replaced with a fully convolutional layer (Fig. 1). This major improvement allows the network to have a dense pixel-wise prediction. To achieve better localization performance, high resolution activation maps are combined with up sampled outputs and passed to the convolution layers to assemble more accurate output. The main architecture is shown in Figure 1.

This improvement enables the FCN to have pixel-wise predictions from the full-sized image instead of a patch-wise prediction and is also able to perform the prediction for the whole image in just one forward pass.

2.2 U-Net architecture

The U-Net architecture is a convolutional neural network, originally designed for image segmentation in the medical field. The icon of this architecture is similar to the English letter U, and for this reason it is called U-Net or U-shaped network. The architecture of this network consists of two parts; the left part is the compression path and the right part is the expansion path. The purpose of the compression path is to understand the content of the image and the role of the expansion path is to help in the process of accurately locating objects. The U-Net architecture has an expansion path on the right and a compression path on the left. The compression path consists of two layers of three by three twists. Each of these convolutional layers has a Relo activation function and a two-by-two max-pooling algorithm for down sampling feature mapping.

One of the most well-known structures for medical image segmentation is U-Net, at first, proposed by Ronneberger et al. [26] using the concept of deconvolution introduced by [38]. This model is built on the elegant architecture of fully convolutional neural network .This novel structure has attracted a lot of attention in medical image segmentation and based on which many variations have been developed. For instance, Gordienko et al. [11] explored lung segmentation in X-ray scans with a U-Net structure-based network. The obtained results have demonstrated that U-Net is capable of fast and precise image segmentation.

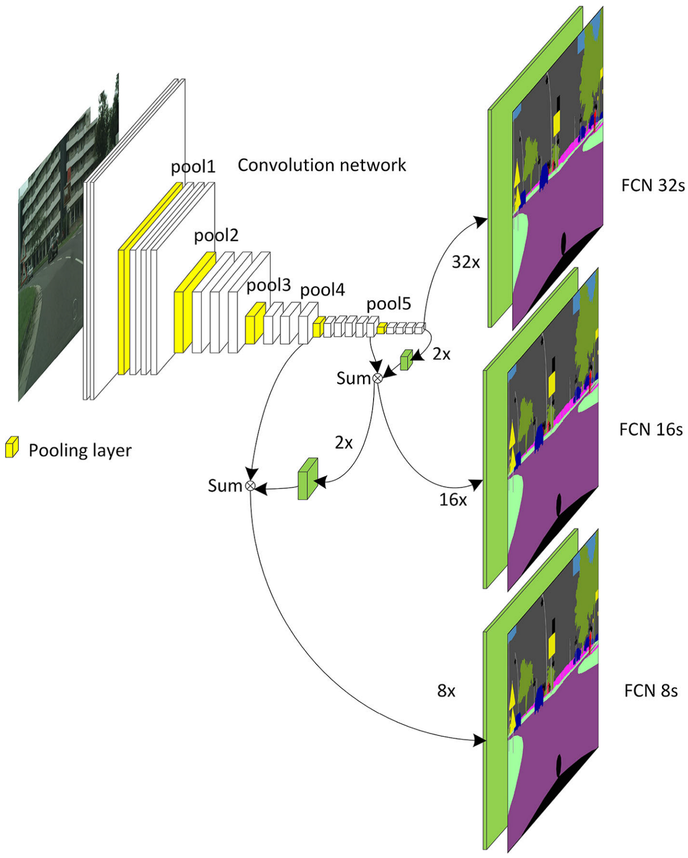

Figure 1: Fully convolutional network (FCN) architecture

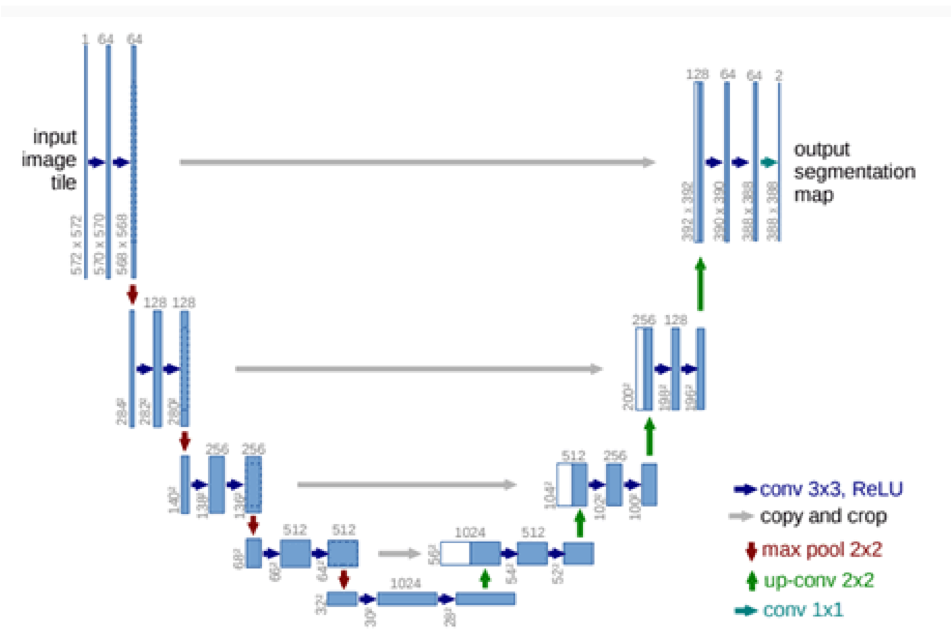

Figure 2: The structure of the U-Net [26]

2.3 FCN joint with CRF and other traditional methods

According to the research of Deeplab, the responses at the final layer of Deep Convolutional Neural Networks (DCNNs) are not sufficiently localized for accurate object segmentation [4]. They overcome this poor localization property by combining a fully connected Conditional Random Field (CRF) at the final DCNN layer. Their method reaches 71.6% IOU accuracy in the test set at the PASCAL VOC-2012 image semantic segmentation task. After this work, they carry out another segmentation architecture by combining domain transform (DT) with DCNN [5] because dense CRF inference is computationally expensive. DT refers to a modern edge-preserving filtering method, in which the amount of smoothing is controlled by a reference edge map. Domain transform filtering is faster than dense CRF inference. Lastly, through experiments, it not only yields comparable semantic segmentation results but also accurately captures the object boundaries. Researchers also exploit segmentation by using super-pixels [20].

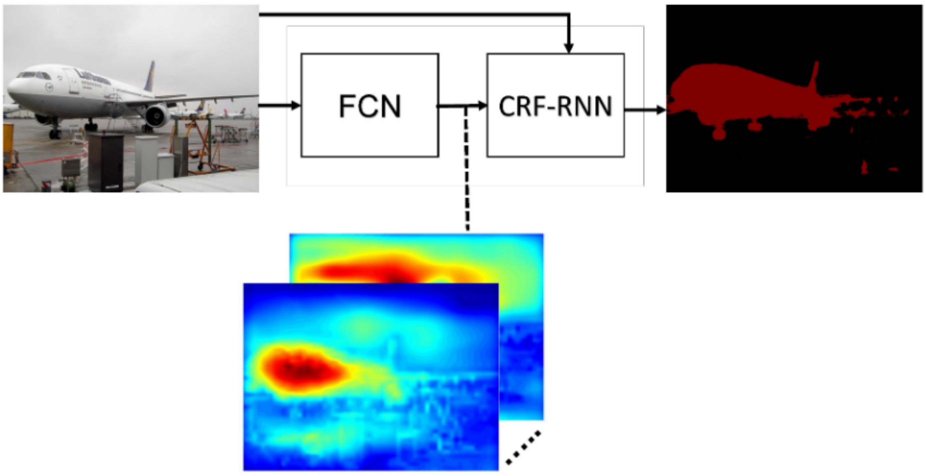

Figure 3: FCN joint with CRF architecture [42]

2.4 AlexNet architecture

AlexNet was the first deep CNN that won the ILSVRC-2012 with a TOP-5 test accuracy of 84.6% while the closest competitor, which made use of traditional techniques instead of deep architectures, achieved a 73.8% accuracy in the same challenge. The architecture presented by [17] was relatively simple. It consists of five convolutional layers, max-pooling ones, Rectified Linear Units (ReLUs) as non-linearity, three fully-connected layers, and dropout. Figure 4 shows that CNN architecture.

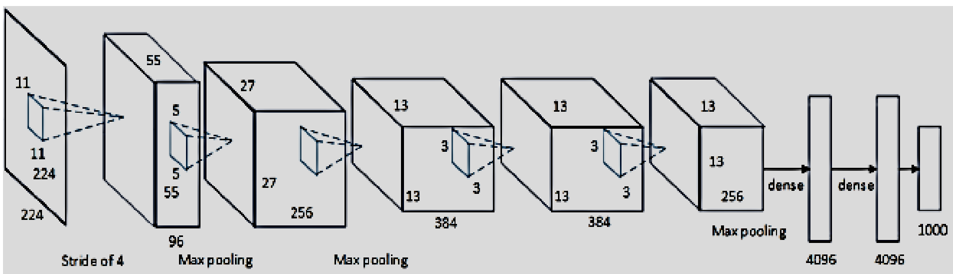

Figure 4: AlexNet Convolutional Neural Network architecture [17].

2.5 VGG Architecture

Visual Geometry Group (VGG) is a CNN model introduced by the Visual Geometry Group (VGG) from the University of Oxford. They proposed various models and configurations of deep CNNs [27], one of them was submitted to the Image Net Large Scale Visual Recognition Challenge (ILSVRC)-2013. That model, also known as VGG-16 due to the fact that it is composed by 16 weight layers, became popular to its achievement of 92.7% TOP-5 test accuracy. Figure 5 shows the configuration of VGG-16. The main difference between VGG-16 and its predecessors is the use of a stack of convolution layers with small receptive fields in the first layers instead of few layers with big 3 receptive fields. This leads to less parameters and more nonlinearities in between layer, thus making the decision function more discriminative and the model easier to train.

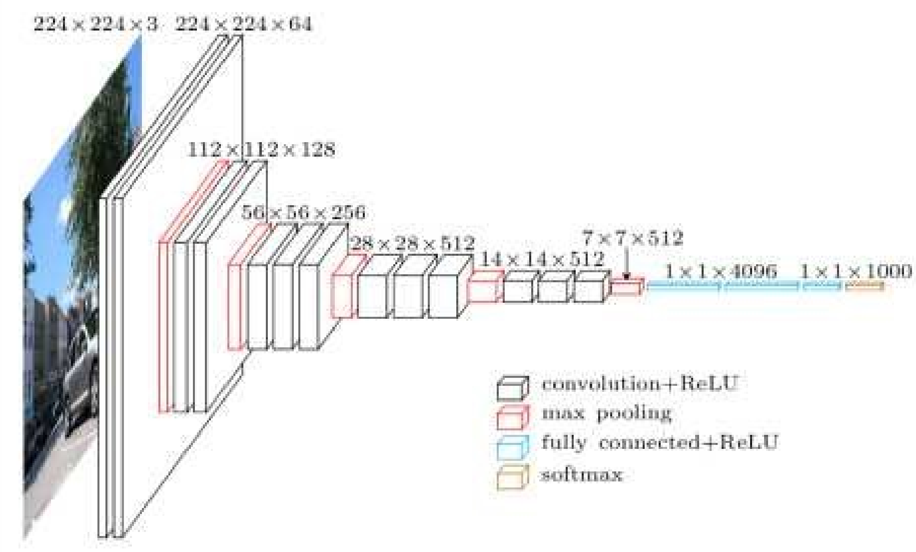

Figure 5: VGG-16 CNN architecture

2.6 GoogLeNet architecture

GoogLeNet is a network introduced by [30] which won the ILSVRC-2014 challenge with a TOP-5 test accuracy of 93.3%. This CNN architecture is characterized by its complexity, emphasized by the fact that it is composed by 22 layers and a newly introduced architecture block called inception module (see Figure 6). This new approach proved that CNN layers could be stacked in more ways than a typical sequential manner. In fact, those modules consist of a Network in Network (NiN) layer, a pooling operation, a large-sized convolution layer, and small-sized convolution layer. All of them are computed in parallel and followed by 1×1 convolution operations to reduce dimensionality. By using of those modules, this network puts special consideration on memory and computational cost by significantly reducing the number of parameters and operations.

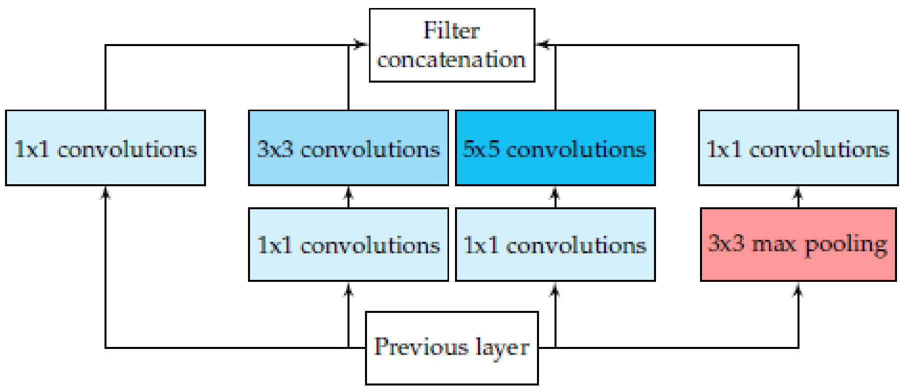

Figure 6: Inception module with dimensionality reduction from the GoogLeNet architecture [30].

2.7 ResNet

Microsoft's ResNet [16] is specially remarkable because of winning ILSVRC-2016 with 96.4% accuracy. The network is well-known due to its depth (152 layers) and the introduction of residual blocks (see Figure 7). The residual blocks address the problem of training a really deep architecture by introducing identity skip connections so that layers can copy their inputs to the next layer.

The intuitive idea behind this approach is that it ensures that the next layer learns something new and different from what the input has already encoded (since it is provided with both the output of the previous layer and its unchanged input). In addition, this kind of connections help overcoming the vanishing gradients problem.

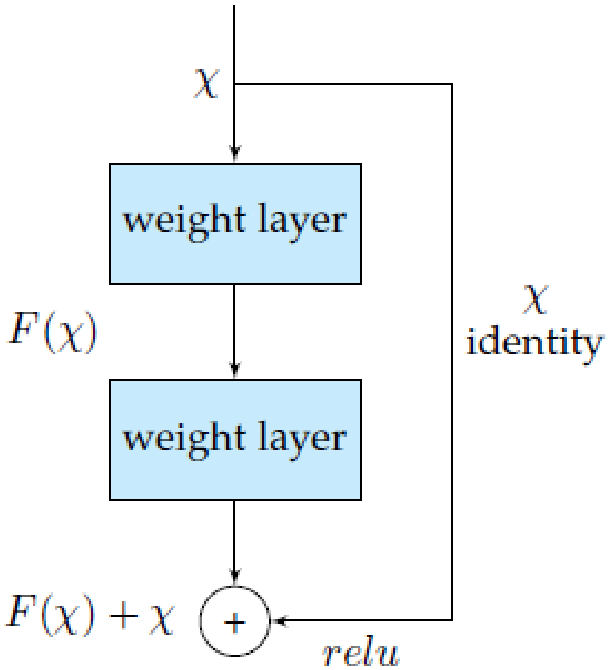

Figure 7: Residual block from the ResNet architecture [16].

In previous works, most of the attempts were made to use neural networks alone, but in this research, an attempt is made to use a combination of colony algorithm and neural network in image segmentation in order to improve the performance and efficiency of the neural network and to achieve faster convergence during training.

3 Method of segmentation of proposed concepts

In general, the methodological steps can be shown in the flowchart in Figure 8. The method in this research is developed using neural network and optimization of network weights using ant colony and image segmentation using deep learning as well as data collection.

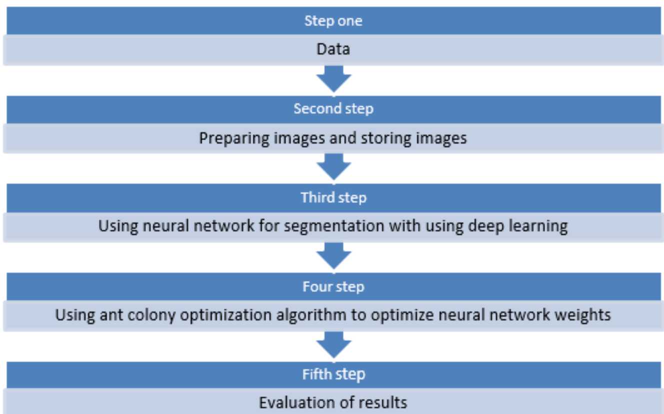

Figure 8: Methodological steps in the flow diagram

3.1 Database

This research uses the CamVid dataset from the university of cambridge for training. This dataset is a collection of images containing street level views obtained while driving. The dataset provides pixel-level labels for 32 semantic classes, including vehicle, pedestrian, and road.

CamVid is a road/driving scene perception database originally captured as five video sequences with a 720*960 resolution camera mounted on a car dashboard. Those sequences were sampled to 701 frames. These images were manually annotated with 32 classes: empty space, building, wall, tree, vegetation, fence, sidewalk, parking block, pillar/pole, traffic cone, bridge, sign, miscellaneous text, traffic light, sky, tunnel , archway, road, road shoulder, line (driving), line (non-driving), animal, pedestrian, child, luggage cart, cyclist, motorcycle, car, SUV/pickup/truck, truck/bus, train, and other moving equipment.

In the proposed method, out of these 701 data, we randomly selected 60% for training data, 20% for validation, and 20% for test data. We divided the existing 32 semantic classes into 11 classes that use a subset of the tags of those 32 semantic classes, which are: building, tree, sky, car, sign, road, pedestrian, fence, pole, sidewalk, and cyclist.

3.2 Using neural network for segmentation using deep learning

To illustrate the training process, this research trains Deeplab $v3+$, a type of convolutional neural network designed for semantic image segmentation. Other types of networks for semantic segmentation include fully convolutional networks, SegNet, and U-Net. The training method shown here can be applied to those networks as well.

This research builds a pre-trained Deeplab $v3+$ network with initial weights from a resnet-18 network optimized with the ant colony algorithm. ResNet-18 is an efficient network suitable for applications with limited processing resources. Other pre-trained networks such as MobileNet v2 or ResNet-50 can also be used depending on the application requirements.

To perform semantic segmentation of images, we used DeepLabv3+ deep learning network, and this network is trained on Pascal VOC dataset and it can segment 20 different object classes, including airplanes, buses, cars, trains, people, and horses.

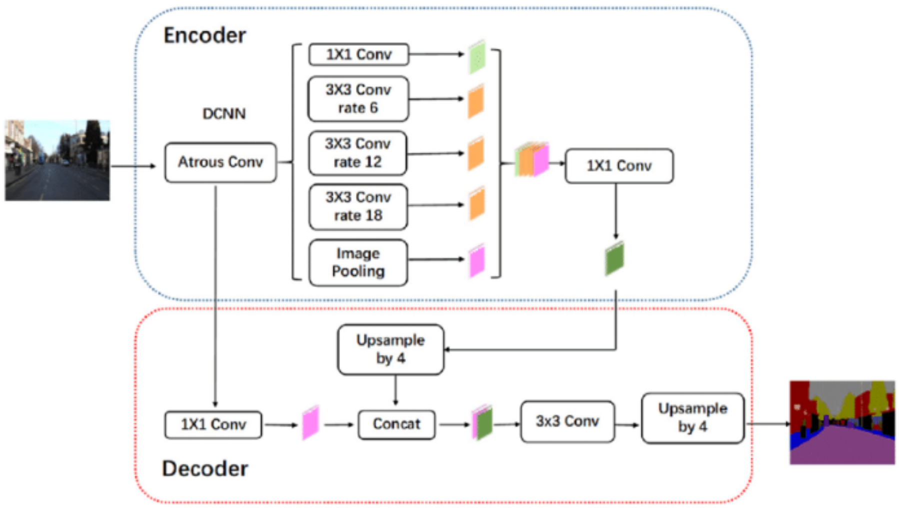

Figure 9: DeepLab V3+ network structure [5]

3.3 Using the ant colony optimization algorithm to optimize the neural network weights

In this research, the ant colony evolutionary algorithm was used to optimize the weights of the neural network. The ant colony algorithm, inspired by the foraging behavior of ants, is an evolutionary algorithm and works well in discrete optimization (in this research, it is used in image segmentation).

In the proposed method, the population number is 30 and the number of repetitions is 100. Optimization is done on the third layer, which is the first convolution layer of the neural network. This layer has about 9408 weights, which we are trying to optimize these 9408 weights by the ant colony algorithm.

3.3.1 Normalization

A segmentation problem in machine learning is where we are given some input (independent variables) and need to predict a target. It is likely that the distribution of values will be very different. Because of this difference in each class, the algorithms tend to be biased towards the available majority values and do not perform well on the minority values. This difference in class frequencies affects the overall predictive ability of the model.

Most machine learning algorithms assume that data is evenly distributed across classes. In the case of class imbalance problems, the broad issue is that the algorithm will be more biased towards the prediction of the majority class. The algorithm will not have enough data to learn the patterns in the minority class, which is why there will be high misclassification errors for the minority class.

To solve this problem, we do something similar to normalization so that the majority class and the minority class have the same effect on our diagnosis. For this, we used the following formula for normalization:

$$Weighted\ average\ of\ classes = \frac{Average\ frequency\ of\ images}{Frequency\ of\ each\ image} \tag{1}$$

The range of values between zero and one is considered.

3.3.2 The process of performing evolutionary algorithm of ants to optimize weights

At first, an initial matrix is created randomly from the responses between the min and max ranges, which are -1 and 1. The number of rows of the constructed matrix is equal to the number of the population and the number of its columns is equal to the total number of weights on which the optimization is to be performed. Then, line by line from the population matrix is entered into the objective function. By using the equations and the process of solving the ant colony algorithm in solving the problem, the matrix numbers created in the previous steps will be improved, which will continue until the end of the number of iterations of the algorithm. Finally, by examining the objective function, the optimal weight will be determined among the population of the algorithm, which are considered the most optimal weights.

In the objective function, we are looking for the minimum error, which means that in each iteration, the average accuracy increases and the error is minimized. The variable that has less error than the others is selected as the optimal weight.

3-3-2-1 Cost Function

The objective function in this research is to minimize the error by improving the weights of the deep neural network in diagnosis. For this purpose, by classifying the data into 11 data classes in the dataset, it is possible to obtain the accuracy measure (it has a value between zero and one) and finally the error value using formulas (2) and (3). In this case, the error value should be close to the zero value in order to obtain the best possible state of the accuracy criterion.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Error = 1 - Accuracy \tag{3}$$

3-3-2-2 The process of solving the ant colony algorithm in problem solving

The Ant algorithm was used to solve problems with discrete nature, and so far, few applications of the Ant algorithm in the field of continuous optimization problems have been reported. ACOR is actually an extension of the ACO method for continuous optimization problems. In this algorithm, the selection probability is based on the amount of pheromone and exploration information. In this regard, ACOR is extended to continuous space using a probability density function. For this purpose, one of the most common functions that creates such a structure is the Gaussian kernel function. For example, for the i^{th} decision variable, there will be a kernel Gaussian function $G_i(x)$, which is shown below [28]:

$$(x) = \sum_{l=1}^{k} w_l g(x) \tag{4}$$

$$\sum_{l=1}^{k} w_l \frac{1}{\sigma_l^i \sqrt{2\pi}} e^{-\frac{(s_l^i - \mu_l^i)^2}{2\sigma_l^{i^2}}} \tag{5}$$

where parameter W is a function of the weight assigned to each answer and σ_l^i is the standard deviation and K is the number of ants.

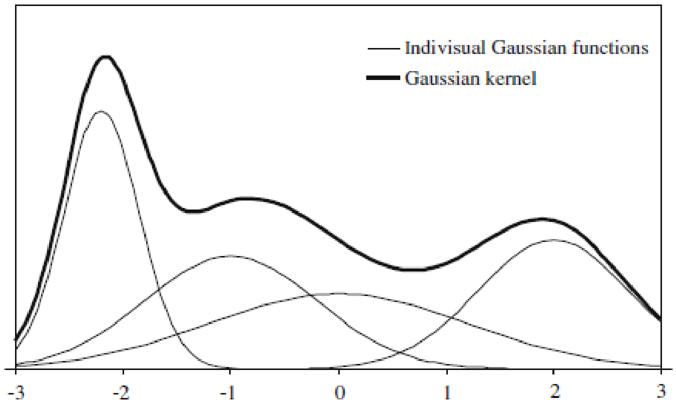

Figure 10: An example of a Gaussian "kernel", consisting of four Gaussian functions

In ACOR, an archive is used to store the set of answers. For this purpose, there will be k number of individual Gaussian functions for each decision variable in the archive, and by selecting each of them and producing a new answer, a situation equivalent to the kernel Gaussian function will be created. The number of answers stored in the archive is equal to k, the l^{th} answer stored in the archive is marked with $s1$, and the decision variables related to the l^{th} answer are denoted by s_1^1 and s_1^2 and The same way up to the nth decision variable is shown by s_1^n. These decision variables are entered into the objective function and its value f(sl) is calculated. The answers in the archive are stored based on their quality For this purpose, to minimize a problem $f(s1) \leq f(s2) \leq ... \leq f(sl) \leq ... \leq f(sk)$, a weight w must be determined for each solution sl which is related to the quality of the solution be relevant $(w1 \geq w2 \geq ... \geq wl \geq ... \geq wk)$ where the value of w is determined according to relation (6) [28].

$$\omega_1 = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(l-1)^2}{2q^2k^2}}, l \in [1:k] \tag{6}$$

Where q is the coefficient, For a fixed k, a small value of q means that only the probability density function of the best solution is used to generate a new solution, while for a large value of q, a more uniform probability is obtained. With a large q, the speed of obtaining the final result slows down.

The parameter I in equation (7), for example, for solution l and the ith decision variable in the solution archive is equal to s1. The determination of the third parameter σ_i is more complicated than the previous two parameters, so to clarify the value of this parameter, it is assumed It is possible for an ant to choose one of these answers, such as sl, using a probabilistic method such as a spinning wheel; By choosing this answer, a new answer should represent the average in a Gaussian function. In addition, the standard deviation between all of them should be generated as a normal distribution using the Gaussian function. In the Gaussian function, the mean and standard deviation must be specified, and as mentioned above, the parameter μ_i represents the mean in a Gaussian function. In addition, the standard deviation between all the values of the i-th decision variable (k values for the i-th decision variable) with respect to the decision variable sl should be calculated, which is denoted by $\sigma 1$. Accordingly, a random number normal to the mean of sl and the standard deviation of I for The i-th decision variable is generated, and this process will continue until the n-th decision variable, and finally, the values assigned to the decision variables are entered into the objective function, which by repeating this step for each ant, finally reaches the number of ants. A new answer is

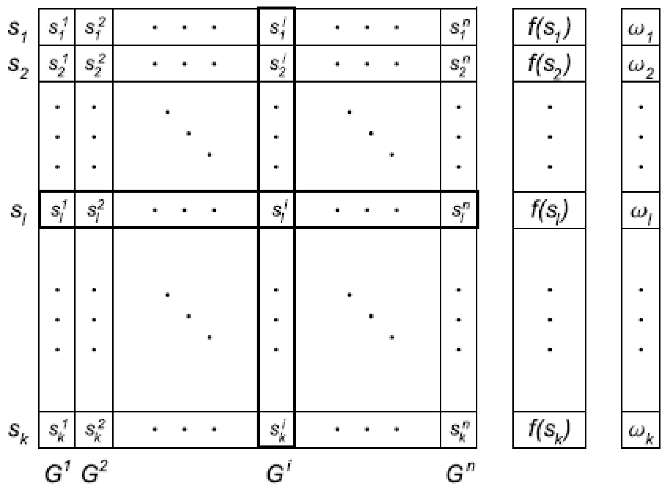

Figure 11: Archive of solutions maintained by ACOR. The solutions are sorted by their quality in the archive

generated and added to the archive and next, the amount of pheromone should be modified, which will be done by sorting all the answers and saving the k best answers. In this regard, the standard deviation is calculated as follows [28]:

$$\sigma_l^i = \xi \sum_{e=1}^{k} \frac{|s_e^i - s_l^i|}{k-l} \tag{7}$$

Where $I \in [1:n]$, and ξ is the determining coefficient of pheromone evaporation ($x > 0$).

ACOR Algorithm

1. For k- ants, solutions $s^1 ... s^n$ are randomly obtained.
2. Sorted by the value of the objective function, where the rank of the solution $l = 1$ is the best.
3. Calculate the weight ω_l for each solution

$$\omega_1 = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(l-1)^2}{2q^2k^2}}, l \in [1:k] \tag{8}$$

   Where q is the coefficient. For a fixed k, a small value of $q(\sim 0)$ means that only the probability density function of the best solution will be used to create a new solution, while for a large value of q , a more uniform probability is obtained. With a large q, decreases the speed of obtaining the final result.
4. The probability of each decision is calculated

$$p_1 = \frac{\omega_1}{\sum_{r=1}^{k} \omega_r}, l \in [1:k] \tag{9}$$

5. By the roulette method , one solution is randomly selected - Sl, using the calculated probability.
6. It is believed that the mathematical expectation μ_l^i is equal to s_l^i.
7. Calculate the variance (deviation from s_l^i.) in the i^{th} dimension by the formula

$$\sigma_l^i = \xi \sum_{e=1}^{k} \frac{|s_e^i - s_l^i|}{k-l} \tag{10}$$

   Where $i \in [1:n]$, and ξ is the coefficient that determines the evaporation of pheromones ($\xi > 0$).
8. Get a solution using a random number generator and a probability distribution obtained using the Gaussian "core".
9. The values of the objective function of each solution are calculated.
10. Add the resulting solutions to the T solutions archive.
11. Streamline the resulting decisions.
12. Save k solutions in the T archive.
13. If the best solution meets the search criteria, complete the search, otherwise go to the third step.

4 Examining the results of applying the ant colony optimization algorithm on image segmentation

In the proposed method, to improve the training process in solving the problem of semantic image segmentation, a new deep learning structure based on convolutional layers and ant colony algorithm was presented to strengthen learning and improve neural network weights. The proposed structure has been used with a database containing 701 images in 11 different classes. These images are a collection of Camvid database images that are used to train and test the proposed segmentation. In the following, the results obtained from the proposed method are given in the form of figures and tables.

The figure below shows the frequency of each class in the CamVid dataset, which frequency each class assigne in the dataset images.

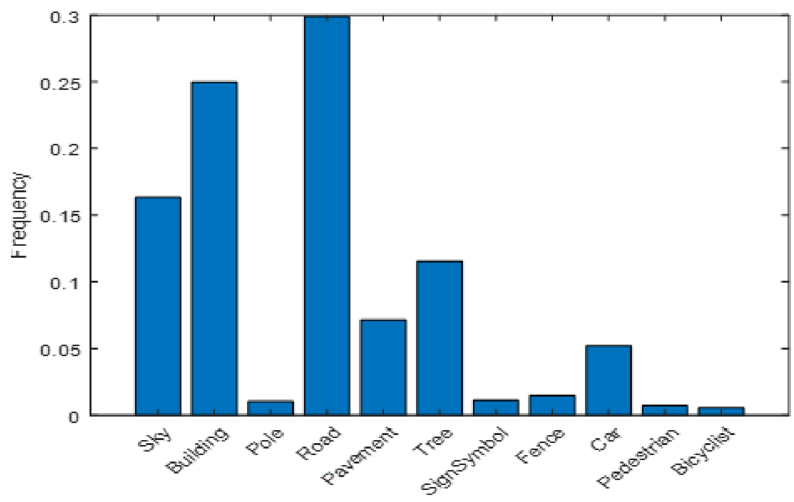

Figure 12: Detection frequency of each class

Figure 13 shows the implementation of the semantic segmentation network in the proposed method and the SegNet method for images from the CamVid database. As it is clear from Figure 13, the proposed method shows superior performance compared to the SegNet method give especially with its ability to determine boundaries, it has done better segmentation than SegNet. Qualitative results show the ability of the proposed architecture to segment smaller classes in road scenes while creating a smooth segmentation of the overall scene.

Also, according to figure 14, it can be seen that image segmentation using neural network and using ant colony algorithm is better than image segmentation using the same neural network without using ant colony algorithm has done. That this also shows the effectiveness of the proposed method (improving neural network weights with ant colony algorithm).

Table 1 shows a quantitative comparison between the proposed method and other methods and shows the detection percentage of each class for each method. According to table1, the proposed method outperforms all other methods, including methods that use depth, video and/or CRF in most classes. Compared with other methods, the predictions of the proposed method are more accurate in 8 classes out of 11 classes.

4.1 Comparing the results of different semantic image segmentation methods with CamVid dataset

In this section, the proposed method is compared with different methods. The mentioned methods all used the CamVid dataset, which makes the comparison easier.

Tables 2 and 3 show the results of image segmentation using a neural network and using the ant colony algorithm versus image segmentation using the same neural network without using the ant colony algorithm and also in table 4 the results related to the base article method (SegNet) are given.

Tables 5 and 6 show a comparison between some common methods and the proposed method. According to these tables, the proposed method shows better performance than other methods and has higher accuracy than other methods.

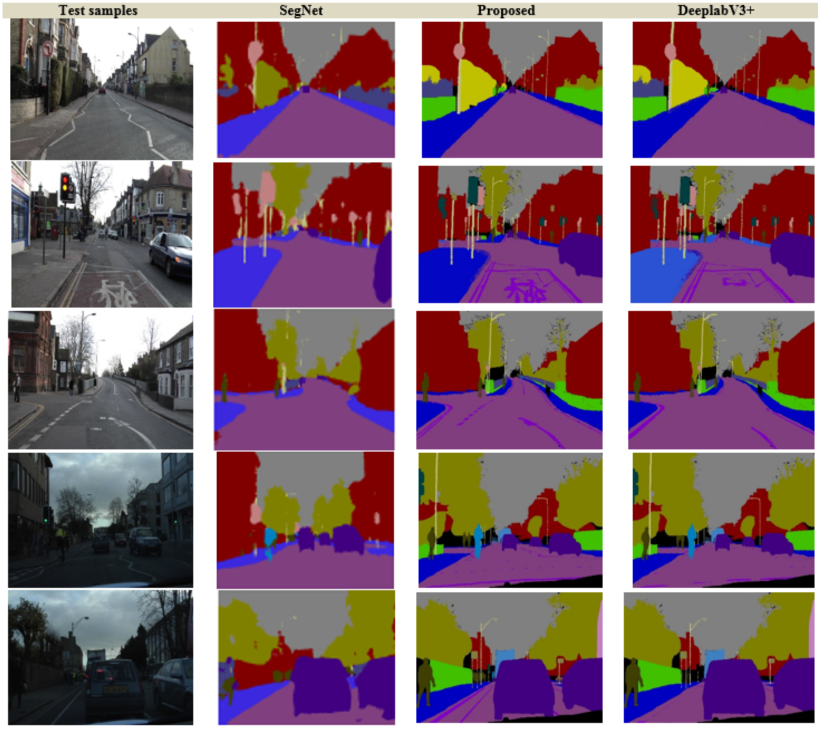

Figure 13: Results of CamVid database test samples

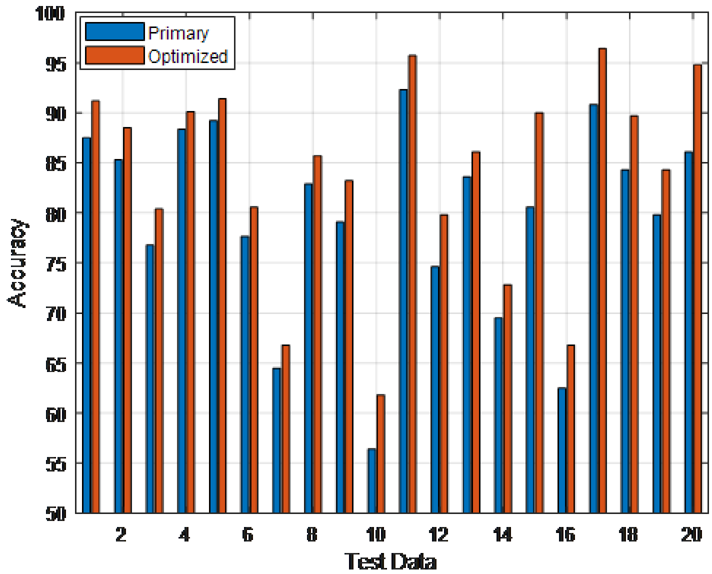

Figure 14: Comparison of detection accuracy in the initial algorithm without ACO and the optimized algorithm with ACO in several sample image

Table 1: Estimation of the research model

Method	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Side-walk	Bicyclist	Class avg.
SfM+Appearance [3]	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	53.0
Boosting [29]	61.9	67.3	91.1	71.1	58.5	92.9	49.5	37.6	25.8	77.8	24.7	59.8
Dense Depth Maps [39]	85.3	57.3	95.4	69.2	46.5	98.5	23.8	44.3	22.0	38.1	28.7	55.4
Local Label Descriptors [34]	80.7	61.5	88.8	16.4	n/a	98.0	1.09	0.05	4.13	12.4	0.07	36.3
Super Parsing [31]	87.0	67.1	96.9	62.7	30.1	95.9	14.7	17.9	1.7	70.0	19.4	51.2
SegNet (3.5K dataset training - 140K) [2]	89.6	83.4	96.1	87.7	52.7	96.4	62.2	53.45	32.1	93.3	36.5	71.20
Proposed Method	90.2	86.1	93.8	89	56.25	95.9	64.6	54.9	37.1	90.8	48.2	73.35

Table 2: Accuracy detection for Deeplab$V3+$ algorithm

Global Accuracy	**Mean Accuracy**	**Mean IoU**	**Weight edIoU**	**Mean BF Score**
0.87695	0.85392	0.6302	0.80851	0.65051

Table 3: Accuracy detection for Deeplab$V3+$ algorithm optimized with ACO

Global Accuracy	**Mean Accuracy**	**Mean IoU**	**Weight edIoU**	**Mean BF Score**
0.92748	0.89561	0.7204	0.86926	0.65051

Table 4: Accuracy detection for SegNet algorithm

Global Accuracy	**Mean Accuracy**	**Mean IoU**	**Weight edIoU**	**Mean BF Score**
0.98960	-	0.6010	-	0.4684

Table 5: Accuracy detection for SegNet algorithm

Models	Mean IOU	Global Accuracy	Epochs
Small U-Net + ReLU	52.74 %	88.83 %	77
Small U-Net + Leaky ReLU	54.38 %	89.58 %	72
U-Net + ReLU	52.60%	89.70%	81
Extended U-Net + ReLU	54.50%	89.91%	84
Extended U-Net + ReLU + Dropout	54.32%	89.60%	57
FC-DenseNet103	66.9%	91.5%	-
EDANet	66.4%	90.8%	-
Proposed	72.04%	92.74%	100

Table 6: Comparison of image segmentation results in the CamVid test dataset

Method	Accuracy (IoU)
CNN-SPP [33]	63.1
DPN [37]	60.1
DeepLab [5]	61.6
ENet [24]	51.3
ICNet [41]	67.1
BiSeNet1 [36]	65.6
BiSeNet2 [36]	68.7
DFANet A [18]	64.7
DFANet B [18]	59.3
SwiftNet pyr [23]	72.85
SwiftNet [23]	73.86
SegNet [2]	90.4
ReSeg [32]	88.7
Proposed Method	92.74

5 Conclusion

In this research, by using the capabilities of the ant colony algorithm and combining it with the deep convolutional neural network, a powerful solution for a semantic image segmentation process that can be used for other data sources with various applications proposed. The proposed structure was evaluated with a CamVid database containing 701 images in 11 classes. As seen, the results obtained in the fourth chapter show the success of the proposed method in segmenting semantic images.

The image segmentation recognition system developed in this research uses a neural network and ant colony optimization, which has resulted in an accuracy of 92%, the accuracy before optimization is 87%, and the accuracy of our proposed method is 90.4% compared to the SegNet method and other methods that did the segmentation with Camvid image sets were higher and performed better. In Table 1, we made a comparison between the detection percentage of each class for the proposed method and other methods and it was observed that compared to other methods, the predictions of the proposed method are more accurate in 8 classes out of 11 classes.

So it can be concluded that using the combination of artificial neural network and evolutionary algorithms improves the efficiency of image segmentation. At the same time, it has been able to show better performance in terms of segmentation accuracy, which the results at hand express this claim.

These results show an increase in the accuracy of the neural network method after optimization using ant colony optimization, which is expected to help identify more accurate image segmentation so that it can be used to make better decisions in all tasks where accuracy is very important.

References

[1] M. Analoui and M. Fadavi Amiri, *Feature reduction of nearest neighbor classifiers using genetic algorithm*, Proc. World Acad. Sci. Engin. Technol. **17** (2006), 36–39.

[2] V. Badrinarayanan, A. Kendall and R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017), no. 12, 2481–2495.

[3] G.J. Brostow, J. Shotton, J. Fauqueur and R. Cipolla, *Segmentation and Recognition Using Structure from Motion Point Clouds*, Forsyth, D., Torr, P., Zisserman, A. (eds) Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol 5302. Springer, Berlin, Heidelberg, 2008.

[4] L. Chen, J. Barron, G. Papandreou, K. Murphy and A. Yuille, *Semantic image segmentation with taskspecific edge detection using CNNS and a discriminatively trained domain transform*, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2016, pp. 4545–4554.

[5] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. Yuille, *Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs*, IEEE Trans. Pattern Anal. Machine Intell. **40** (2017), no. 4, 834–848.

[6] D. Ciresan, A. Giusti, L.M. Gambardella and J. Schmidhuber, *Deep neural networks segment neuronal membranes in electron microscopy images*, Adv. Neural Inf. Process. Syst. **25** (2012), 2843–2851.

[7] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth and B. Schiele, *The cityscapes dataset for semantic urban scene understanding*, Proc. IEEE Conf. Comput. Vision Pattern Recog. (CVPR), 2016, pp. 3213–3223.

[8] A. Ess, T. Muller, G. Grabner and L.J. Van Gool, *Segmentation-based urban traffic scene understanding*, BMVC **1** (2009).

[9] C. Farabet, C. Couprie, L. Najman and Y. LeCun, *Learning hierarchical features for scene labeling*, IEEE Trans. Pattern Anal. Machine Intell. **35** (2013), no. 8, 1915–1929.

[10] A. Geiger, P. Lenz and R. Urtasun, *Are we ready for autonomous driving? the Kitti vision benchmark suite*, IEEE Conf. Comput. Vision Pattern Recog., 2012, pp. 3354–3361.

[11] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokovyi and S. Stirenko, *Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer*, Int. Conf. Theory Appl. Fuzzy Syst. Soft Comput., Springer, 2018, pp. 638–647.

[12] S. Gupta, R. Girshick, P. Arbeláez and J. Malik, *Learning rich features from RGB-D images for object detection*

and segmentation, Computer Vision–ECCV 2014: 13th Eur. Conf., Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13, Springer International Publishing, 2014, pp. 345–360.

[13] P. Hariharan, P. Arbelae, R. Girshick and J. Malik, *Simultaneous detection and segmentation*, Eur. Conf. Comput. Vision, Springer, 2014, pp. 297–312.

[14] M. Hashemi, H. Hassanpour, E. Kozegar and T. Tan, *Cystoscopy image classification using deep convolutional neural networks*, Int. J. Nonlinear Anal. Appl. **10** (2019), no. 1, 193–215.

[15] M. Hashemi, H. Hassanpour, E. Kozegar and T. Tan, *Cystoscopic image classification based on combining MLP and GA*, Int. J. Nonlinear Anal. Appl. **11** (2019), no. 1, 93–105.

[16] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2016, pp. 770–778.

[17] A. Krizhevsky, I. Sutskever and G.E. Hinton, *ImageNet classification with deep convolutional neural networks*, Adv. Neural Inf. Process. Syst. **25** (2012), 1097–1105.

[18] H. Li, P. Xiong, H. Fan and J. Sun, Dfanet: Deep feature aggregation for real-time semantic segmentation, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2019, pp. 9522–9531.

[19] J. Long, E. Shelhamer and T. Darrell, *Fully convolutional networks for semantic segmentation*, IEEE Trans. Pattern Anal. Mach. Intell. **79** (2014), no. 10, 1337–1342.

[20] M. Mostajabi, P. Yadollahpour, G. Shakhnarovich and A. Feedforward, *Semantic segmentation with zoom-out features*, Proc. Comput. Vision Pattern Recog., 2015, pp. 3376–3385.

[21] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou and E.P. Barbano, *Toward automatic phenotyping of developing embryos from videos*, IEEE Trans. Image Process. **14** (2005), no. 9, 1360–1371.

[22] M. Oberweger, P. Wohlhart and V. Lepetit, *Hands deep in deep learning for hand pose estimation*, arXiv preprint arXiv: 1502.06807. (2015).

[23] M. Orsi, I. Kreso, P. Bevandic and S. Segvic, *In defence of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images*, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2019, pp. 12607–12616.

[24] A. Paszke, A. Chaurasia, S. Kim and E. Culurciello, *ENet: A deep Neural Network Architecture for Real-Time Semantic Segmentation*, 7 Jun 2016, in Computer Vision and Pattern Recognition (cs.CV) arXiv:1606.02147 [cs.CV]. 2016.

[25] P. Pirozmand, M. Fadavi Amiri, F. Kashanchi and N. Yugeeta Layne, *Age estimation, a Gabor PCA-LDA approach*, J. Math. Comput. Sci. **2** (2011), no. 2, 233–240.

[26] O. Ronneberger, P. Fischer and T. Brox, *U-net: convolutional networks for biomedical image segmentation*, Int. Conf. Med. Image Comput. Comput.-Assis. Interven., Springer, 2015, pp. 234–41.

[27] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv: 1409.1556, 2014.

[28] K. Socha and M. Dorigo, *Ant colony optimization for continuous domains*, Eur. J. Oper. Res. **185** (2005), no. 3, 1155–1173.

[29] P. Sturgess, K. Alahari, L. Ladicky and P.H.S. Torr, Combining appearance and structure from motion features for road scene understanding, BMVC-British Machine Vision Conf., BMVA, 2009.

[30] W. Szegedy, Y. Liu, P. Jia, S. Sermanet, D. Reed, D. Anguelov, V. Erhan, V. Vanhoucke and A. Rabinovich, *Going deeper with convolutions*, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2015, pp. 1–9.

[31] J. Tighe and S. Lazebnik, *SuperParsing: Scalable nonparametric image parsing with superpixels*, Daniilidis, K., Maragos, P., Paragios, N. (eds) Computer Vision – ECCV 2010. ECCV 2010, Lecture Notes in Computer Science, vol 6315, Springer, Berlin, Heidelberg, 2010.

[32] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci and A. Courville, ReSeg: A recurrent neural network-based model for semantic segmentation, Proc. IEEE Conf. Comput. Vision Pattern Recog. Workshops, 2016, pp. 41–48.

[33] J. Wan, D. Wang, S.C.H. Hoi, P. Wu, J. Zhu, Y. Zhang and J. Li, *Deep learning for content-based image retrieval: A comprehensive study*, Proc. 22nd ACM Int. Conf. Multimedia, ACM, 2014, pp. 157—166.

[34] Y. Yang, Z. Li, L. Zhang, C. Murphy, J. Ver Hoeve and H. Jiang, *Local Label Descriptor for Example Based Semantic Image Labeling*, Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds) Computer Vision – ECCV 2012. ECCV 2012, Lecture Notes in Computer Science, vol 7578, Springer, Berlin, Heidelberg, 2012.

[35] Y. Yoon, H.G. Jeon, D. Yoo, J.Y. Lee and I. So Kweon, *Learning a deep convolutional network for light-field image super resolution*, Proc. IEEE Int. Conf. Comput. Vision Workshops, 2015, pp. 24–32.

[36] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu and N. Sang, *Bisenet: Bilateral segmentation network for real-time semantic segmentation*, Proc. Eur. Conf. Comput. Vision (ECCV), 2018a, pp. 325–341.

[37] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu and N. Sang, *Learning a discriminative feature network for semantic segmentation*, Proc. IEEE Conf. Comput. Vision Pattern Recognition, 2018b, pp. 1857–1866.

[38] M.D. Zeiler and R. Fergus, *Visualizing and understanding convolutional networks*, Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014, Lecture Notes in Computer Science, vol 8689. Springer, Cham, 2014.

[39] C. Zhang, L. Wang and R. Yang, *Semantic segmentation of urban scenes using dense depth maps*, Computer Vision–ECCV 2010: 11th Euro. Conf. Comput. Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11. Springer Berlin Heidelberg, 2010, pp. 708–721.

[40] Y.J. Zhang, *Influence of segmentation over feature measurement*, Pattern Recog. Lett. **16** (1995), no. 2, 201–206.

[41] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, *Pyramid scene parsing network*, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2017, pp. 2881–2890.

[42] Sh. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Zh. Su, D. Du, Ch. Huang, and Ph.H.S. Torr, *Conditional random fields as recurrent neural networks*, On 11 Feb 2015 arXiv:1502.03240 [cs.CV].