IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 November 2022, accepted 11 December 2022, date of publication 20 December 2022,
date of current version 30 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3230983

TOPICAL REVIEW

A Systematic Literature Review on Machine Learning and Deep Learning Methods for Semantic Segmentation

ALI SOHAIL[1], NAEEM A. NAWAZ[2], ASGHAR ALI SHAH[3], SAIM RASHEED[4], SHEEBA ILYAS[1], AND MUHAMMAD KHURRAM EHSAN[5]

[1]Department of Computer Science, Minhaj University Lahore, Lahore 54782, Pakistan
[2]College of Computer and Information Systems, Umm Al-Qura University, Makkah Al-Mukarramah 24381, Saudi Arabia
[3]Department of Computer Science, Bahria University, Lahore Campus, Lahore 54782, Pakistan
[4]Department of Information Technology, Faculty of Computing and IT, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[5]Faculty of Engineering Sciences, Bahria University, Lahore Campus, Lahore 54782, Pakistan

Corresponding author: Asghar Ali Shah (asgharali.bulc@bahria.edu.com)

ABSTRACT Machine learning and deep learning algorithms are widely used in computer science domains. These algorithms are mostly used for classification and regression problems in almost every field of life. Semantic segmentation is an instantly growing research topic in the last few decades that refers to the association of each pixel in the image to the class it belongs. This paper illustrates the systematic survey of advanced research in the field of semantic segmentation till date. This study provides the brief knowledge about the latest proposed methods in the domain of semantic segmentation. The proposed study comprehends the concepts, techniques, tool, and results of different research frameworks proposed in the context of semantic segmentation. This study discusses the latest research papers in which machine learning and deep learning techniques are exploited for semantic segmentation and published between 2016 and 2021. The systematic literature review collected from seven different article libraries including ACM digital Library, Google Scholar, IEEE Xplore, Science Direct, Google Books, Refseek and Worldwide Science. For assuring the quality of the paper those papers are selected which have several citations on standardized platforms. Most of the studies used COCO, PASCAL, Cityscapes and CamVid dataset for training and validation of the machine learning and deep learning models. The results of the selected research articles are collected in the form of accuracy, mIoU value, F1 score, precision, and recall. In this study, we also conclude that most of the semantic segmentation studies use ResNet as the backbone of the architecture and none of the researchers used ensemble learning methods for semantic segmentation that is the loophole of the selected studies.

INDEX TERMS Semantic segmentation, PASCAL, COCO, cityscapes, CamVid, ResNet, deep learning, machine learning.

I. INTRODUCTION

An image is the collection of rectangular sequentially arranged pixels. Segmentation refers to the classification of the pixels in accordance with the class they belong. The main goal of the image segmentation is to cluster all the pixels of an image [1]. Semantic segmentation recognizes and analyze the image on pixel level and assign each pixel a specific class where it belongs. The process refers to the labeling of image pixels belonging to the class of the image. This method has various real-life applications includes medical image analysis, facial recognition, self-driving cars, virtual fitting rooms, and industrial inspections etc. Several techniques proposed by the researcher for the semantic segmentation. Machine learning and deep learning algorithms are widely used in the field of bioscience, cancer identification [2], supply chain

The associate editor coordinating the review of this manuscript and approving it for publication was Antonio J. R. Neves.

management [3] computer vision, NLP and semantic analysis [4], [5]. The process of semantic segmentation followed by three steps as classifying, localizing and segmentation.

Classifying separate the different classes of the pixels from each other in an image. Localization finds the objects and draw the boundaries around these selected pixel classes. Segmentation creates the segmented mask. This process is different from the process of classification. The classification process just assign a single class to the whole image while semantic segmentation assign every pixel to a specific classes where it belongs [6]. In machine learning approaches the algorithm takes the image of $W \times H$ and generate the pixel wise $W \times H$ matrix of that image. Different form of objects can be detected in semantic segmentations. If we have the outdoor images the segmentation may predict ground, people, tree, sky etc. The most popular datasets used for the process of semantic segmentations are COCO, Cityscapes, CamVid and PASCAL datasets containing thousands of image segmentation sequences. The process of segmentation can be useful in various fields including medical field, autonomous vehicles, image canalization using satellites etc.

The proposed study is going to elaborate the latest proposed deep learning and machine learning techniques for semantic segmentation. The aim of this study is to create a base paper that is beneficial for the researchers to learn the all the different methods proposed in context of semantic segmentation deeply in one paper, so that they can learn advantages and disadvantages of latest proposed methods. This study will create a benchmark for the researchers to learn different machine learning and deep learning techniques, their pros and cons, accuracies, results, and flaws of semantic segmentation papers proposed using deep learning and machine learning in one paper. This study is discussing 44 latest research articles collected from seven different articles libraries including ACM digital Library, Google Scholar, IEEE Xplore, Science Direct, Google Books, Refseek and Worldwide Science. The detail of the working methods, application and results are explained in the next sections. The list of abbreviations used in the study are defined in Table 1.

TABLE 1. List of abbreviations.

Words	Abbreviation
ANN	Asymmetric Non-local Neural Network
BiSeNet	Bilateral Segmentation Network
COCO	Common Objects in Context
Dnlnet	Disentangled Non-local Neural Networks
DaNet	Dual Attention Network
ENet	Efficient Neural Network
ENCNet	Encoding Network
EMA	Expectation-Maximization Attention
FCN	Fully Convolutional Network
gscnn	Gated shape cnns for semantic segmentation
GiNet	Graph interaction network
GloRe	Global Reasoning
HRNet	High Resolution Network
HardNet	Harmonic Densely connected network
isanet	Interlaced Sparse Self-Attention Network
mIoU	Mean Intersection Over Union
ocrnet	Object-Contextual Representations for Semantic Segmentation
pfpn	Panoptic Feature Pyramid Networks
ResNet	Residual Neural Network
sfnet	Semantic Flow Network
SAM	Self-attention mechanism
SPPM	Simple Pyramid Pooling Module

II. SYSTEMATIC REVIEW METHODOLOGY

The proposed study is inspired by the systematic literature review method proposed by Brereton et al. [7]. This systematic review process provides an appropriate, efficient, and reliable approach for the literature review. The process of systematic literature review for semantic segmentation based on five steps [8].

I **Planning of the review process**: This step includes to specify the requirements for the review process. Form the questions that are requires for the study.

II **Find the relevant work:** This phase includes the selection of relevant review studies.

III **Assessing the research quality:** This phase includes to describe the minimum acceptance criteria for the study.

IV **Writing up the document:** This step includes documenting the research on a paper.

V **Analyze the results:** In the last step the gathered results are analyzed to create a valid report.

III. RESEARCH PLANNING

The research planning involves the identification of research questions, identification of related keywords, the criteria for including or excluding the research article, and measurement of the quality of the research paper. Figure 1 explains the criteria of systematic literature review process [9].

The main object of this study is to develop a review paper that determine the advantages and disadvantages of all the latest techniques proposed for semantic segmentation. The research questions used for the current systematic literature review process are

RQ1: What is semantic segmentation an overview?

RQ2: How semantic segmentation is useful in different contexts?

RQ3: What is the research proposed for the identification of semantic segmentation?

RQ4: What is the machine learning or deep learning approaches used in these studies?

RQ5: What are the results generated through these research papers?

RQ6: What are the research possibilities?

RQ7: How these studies are compared with each other?

RQ8: What are the pros and cons of the selected studies?

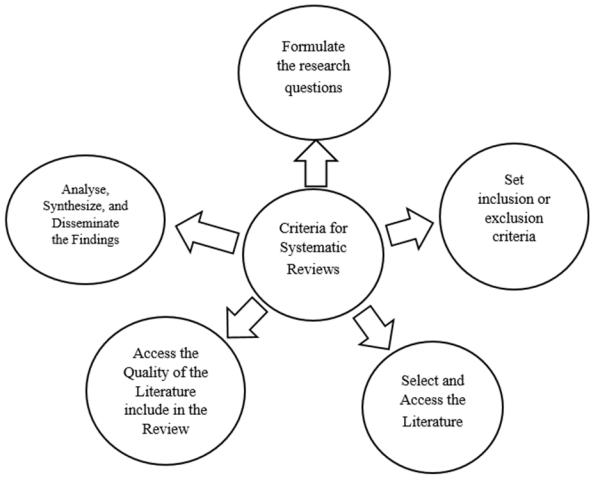


FIGURE 1. Selected criteria for systematic literature review.

IV. FINDINGS AND RELEVENT WORK

Different articles are selected from different websites based on the titles of the various studies, useful keywords, the abstract and conclusion. The articles are selected from the following database sources

1) ACM digital Library
2) Google Scholar
3) IEEE Xplore
4) Science Direct
5) Google Books
6) Refseek
7) Worldwide Science

In the proposed study different combinations of keywords word include, ''Semantic segmentation, Machine learning for semantic segmentation, segmentation, Deep learning for semantic segmentation, Accurate segmentation, Real time semantic, image segmentation, Fast semantic segmentation, Semantic representation, Context encoding, Neural network for semantic segmentation'' are used for searching different articles from the above-mentioned websites. The include exclude criteria for the paper selection in this research is as follows.

Include:

- The papers in English.
- Journal papers, Conference papers, Books etc.
- The papers of semantic segmentation using Machine learning and deep learning Methodologies
- Papers after year 2015

This paper focused only on the machine learning and deep learning papers published between 2015 to 2021. Only Journal papers, conference paper and books are selected for the review. The irrelevant papers that are published before 2015, published in any other language than English, and use any other computational technique rather than machine learning and deep learning for semantic segmentation is excluded.

Exclude:

- Irrelevant papers
- Papers before 2015
- Paper which did not use machine learning or deep learning methods.
- Paper other than English language.

In the proposed study we select 1,172 research articles from seven different articles websites. The study is going to review the articles that are published after the year 2015. So, from all 1,172 articles 659 articles are rejected due to their publication date 2015 or before 2015. The further selected articles are forwarded for the next selection. 352 articles are rejected based on their keywords, title and abstract that were not according to the required condition. This systematic literature review paper is purely illustrating the machine learning and deep learning process for the semantic segmentation, so the articles that uses the techniques rather than the machine learning and deep learning frameworks are also rejected.

The overall articles that are selected after filtering from these phases are 83. Out of these articles 39 are rejected due to their low quality. These are the papers that did not show efficient results in semantic segmentation. The final articles used for the review process were 44. Figure 2 explains the study flow of different phases using Preferred Reporting Items for Systematic Reviews and Meta-Analyze (PRISMA) [10] diagram for the article selection.

V. LITERATURE REVIEW

Abhishek Chaurasia, Sangpil Kim and Eugenio Culurciello [11] proposed a study presents a deep neural network technique ENet (Efficient Neural Network) study based on pixel wise semantic segmentation for mobile application. The process is carried out in different phases. The image of 512×512 resolution passed from three convolutional layer from each block that reduce the dimensions of the images. Normalization and PRELU activation function is applied in between each convolutional layer. Figure 3 explains the block diagram for the prosed methodology [11].

ENet provides an efficient large scale computational of high-resolution images on high end GPUs in much faster and better way with large gains in tasks.

Down sampling of images cause the elimination of feature map and spatial information for overcoming this the model use FCN (Fully Convolutional Network) [12] and SegNet [13]. SegNet deep Encoder-Decoder architecture is used for image segmentation. The study uses CamVid, SUN RGB-D and Cityscapes dataset. Cityspace dataset is widely used benchmark dataset contain the images from 27 cities of Germany and neighbor country having 500 validation, 2975 training and 1525 test images [14]. Torch7 and cuDNN libraries are used for performance evaluation. IOU matric is used for the result comparison. The network is trained on Adam optimization algorithm. SegNet is efficient for both computational time and memory of the system as compared to other proposed models. The global average result for SegNet

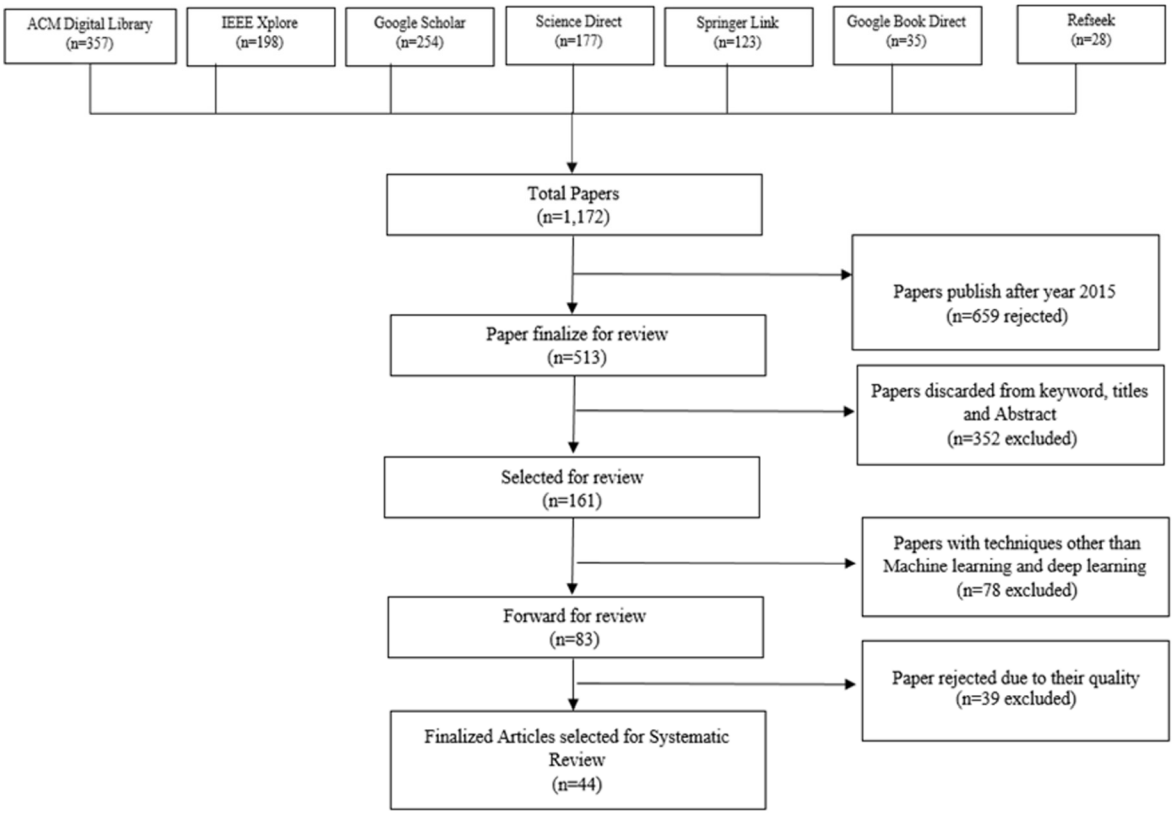


FIGURE 2. Article selection criteria for systematic literature review.

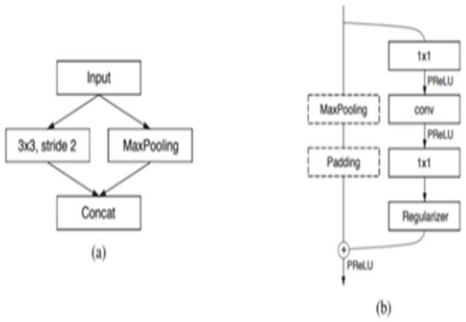


FIGURE 3. ENet architecture with bottleneck module.

was 70.3% and ENet was 59.5%. The end-to-end deep segmentation learning is hard in SegNet model.

Hang Zhang et al. [15] introduce Context encoding Module for the identification of contextual information. The extracted image first passed from CNN layers for feature extraction. Then the Context Encoding module capture the encoded segments and predict the scaling. Semantic Encoding Loss (SE-loss) is used to predict the different scene categories. At last fully connected layer for per pixel prediction [15]. SE-Loss learned from each pixel in the semantic segmentation. Pascal context and ADE20K dataset is used for the study. Mean Intersection of Union (mIoU) and pixel accuracy (pixAcc) evaluation matrix are used for evaluation matrix. The study shows 85.9% mIoU on PASCAL dataset.

In another research Spatial pyramid pooling module with deep learning method is used for semantic segmentation task. This study was using the combine methods of multiple effective fields-of-view and multi-scale contextual information for encoding the features [16]. This study was also developed on PASCAL VOC 2012 [17] dataset along with city spaces dataset. The accuracy for this method was 89%. DeepLabv3 is used for semantic segmentation. Atrous convolutional tool is used as encoder decoder. DeepLabv3+ efficiently encode the segments and a decoder method is used for getting the object boundaries. Xception model also used for making the performance of the system faster. The methodology of this model is explained in Figure 4 [18].

Changqian Yu et al. [19] use BiseNet process for increasing the speed and accuracy of semantic segmentation. In this research Spatial path is designed for extraction and preservation of the high resolution features of the images and Context path is used for down sampling [20]. BiseNet provides the real-time efficiency in speed and accuracy of semantic segmentation. Cityscapes, COCO Stuff and CamVid dataset is used for training and testing the study [14]. The image resolution of $2,048 \times 1,024$ is used with each pixel carried

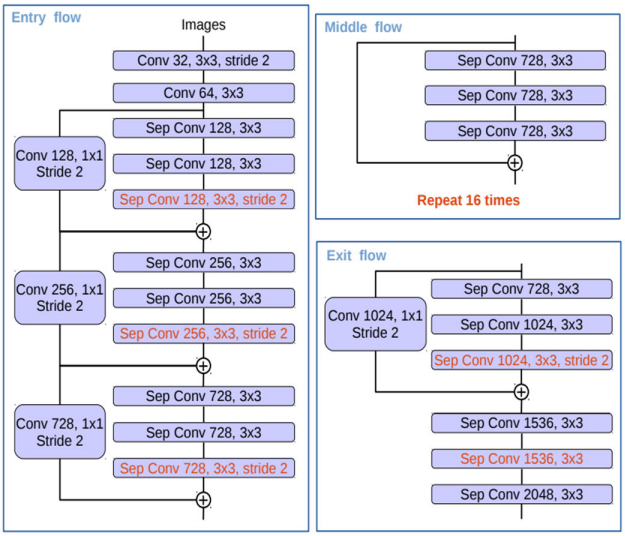


FIGURE 4. DeepLabV3 Segmentation using Atrous convolutional.

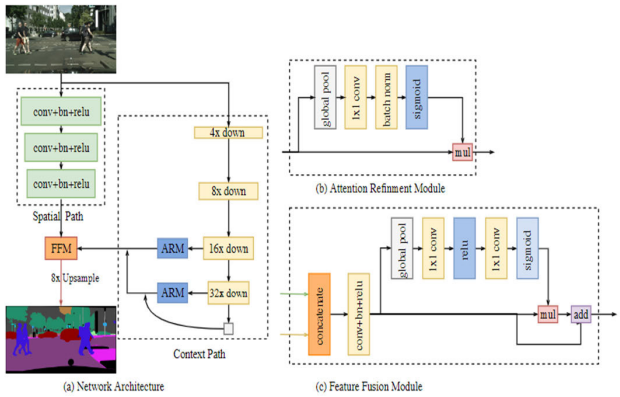


FIGURE 5. BiseNet model of Semantic Segmentation.

19 classes that are pre-defined. Figure 5 gives the overview of BiseNet.

Xception model is used along with the context path and three convolutional layers are used with the spatial path. Auxiliary loss function is used in the back propagation to overcome the gradient vanishing problem [21]. The model gives the 74.8% mIoU value along with 105.8 FPS and 65.55 pixel accuracy.

A study proposed [22] use the Attention gates for automatically focus the medical images with different sizes, edges and shapes. CT Abdominal was the benchmark dataset for the study. This technique is used first time in medical imaging task. AGs are integrated with U-Net CNN model for increasing the accuracy of the result. CNN model is used to extract high dimensional images from the medical imaging. Attention gates efficiently down sample the images. The study shows the best results for organs/tissues localization and identification. Song-Hai Zhanga, Xin Donga, Jia Lib, Ruilong Lia and Yong-Liang Yang [23] develop a real time mobile application for detection of portrait segmentation called PotraNet. This application was also based on U-Net architecture along with two auxiliary losses functions at training stage for increasing the speed of the model. The study shows that PotraNet model provides a lightweight real-time portrait segmentation using mobile phones.

This application increases the robustness of the pixels in the complex lightning environment. EG1800 and Supervise-Portrait datasets are used for PotraNet framework. The application was able to process the images of 224 $\times$ 224 RGB with the FPS of 30. An Encode is used for extracting the main features of the picture from RGB image. And a Decoder use to create feature map. Deformation and texture augmentation is used for data augmentation process. The experiment is carried out on Pytorch framework along with NVDIA graphic card. The accuracy of PotraNet for EG1800 dataset was 96.62% and 93.4% for Supervise-Portrait dataset.

Alexander Kirillov, Ross Girshick, Kaiming He and Piotr Dollar [24] use Panoptic Features for both instance and semantic segmentation. The study uses Mask R-CNN model with feature Pyramid Network (FPN). The dataset is taken from COCO and city spaces dataset.

ResNet serves as the backbone for the study. The model gives the value of 79.1 mIoU. Object Contextual representation are also used in the studies for semantic segmentation. The process in carried out in three steps [25].

- In the first step the ground truth segmentation is used to learn the object region (Object region learning).
- In the second step Object region representation is computed from the pixel representation (Object Region Representation).
- In the last step the relation between the pixel and object region calculated and augment each pixel with object-contextual representation (Output of the representation).

The dataset used for the study include ADE20K, Cityscapes, LIP, COCO-Stuff and PASCAL-Context. Res-Net 101 [26] serve as the backbone of this architecture. Lang Huang et al. [27] presents the interlaced sparse self-attention technique for the improvement of semantic segmentation process using self-attention process [28]. The main features of the study are to factorize dense affinity matrix as the product of two sparse affinity matrices. The process is carried out in two modules. First module estimates the similarities for subset of position with long spatial interval distances and the second module is used calculate similarities in the subset of positions having short spatial interval distances. Permute and reshape functions are used for dividing the short range and long-range attention using Pytorch. The study works on six datasets. Polyak's learning rate is applied for all the semantic segmentations [29]. Auxiliary loss is also applied in the feature map. The process gives 81.4% mIoU for Cityspace dataset.

In all the studies discussed in the above section used ResNet, DenseNet or MobileNet. In the study proposed by Ping Chao et al. used HarDNet (Harmonic Densely connected network) to predict accurate inference time in low memory traffic and MAC's. The study proposed that memory traffic is

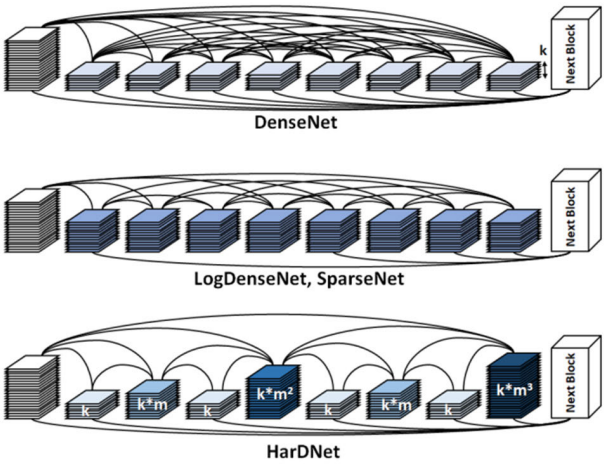


FIGURE 6. HarDNet model.

highly considerable factor while designing a NN for high resolution networks. Figure 6 compare the working of DenseNet, LongDenseNet and HarDNet model [30].

The layers of convolutional make a harmonic dense block. In HarDNet the gradient from the depth layer L will pass over most log L layers. DenseNet used as bottleneck to improve the efficiency of parameters. To bypass all the input parameters of Harmonic Dense Block, HarDNet use six models. Max-Pooling used as down-sampling and BN-Relu serve for batch normalization. The results show the HarDNet reduce the GPU interface reduction time by 35% and DRAM traffic about 40% as compared to ResNet and DenseNet.

Gated stream architecture (Regular stream and shape stream) use gatian process to connect the layers of CNN for semantic segmentation [31]. The process uses low level activations for shape stream and High-level activation for classical or regular stream. Regular stream is any feed forward CNN backbone architecture used with shape stream that produce the semantic boundaries of the input shape. ResNet and WideNet mostly used in the regular stream of gated architecture. The main objective of the shape stream is to process the shapes based on GCL, Residual blocks and supervisions. Fusion Module use Atrous Spatial Pyramid Pooling module (ASPP) [32] for combining the information gain from two streams and make high quality boundaries through Dual Task Regularizer (DTR) [33]. IoU, Boundary Matrix and Distance based evaluation matrixes for gathering the results [34]. The process gives the mean result of 82.5% for all objects detection.

Graph-Based Global Reasoning Networks is an approach in which a set of features are globally collected and then anticipated to an interaction space where relational reasoning is computed. Global Reasoning (GloRe) unit implements the coordinate-interaction space mapping by weighted broadcasting, weighted global pooling and the relation reasoning via graph convolution. GloRe is used to enhance the performance of ResNet, DPN, SE-net etc [35]. ResNet is used for image classification. FCN is used for semantic segmentation.

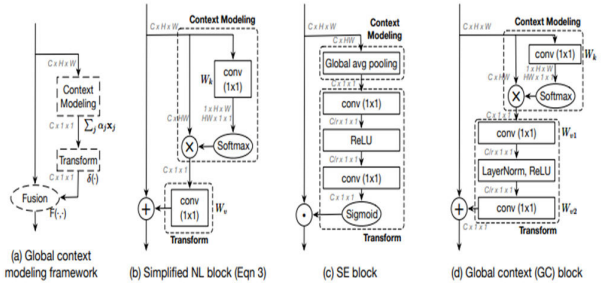


FIGURE 7. GCNet layers.

GloRe is a lightweight approach give 79.8% result on RGB video. Another light weighted framework a global context network (GCNet) [36] is developed by the research based on query independent formulation for global context modelling for capturing long distance dependencies. The long-range dependency modelling based on self-attention and modelling of e query-independent global context. COCO and Kinetics dataset is used for the study. The study simplifies the non-local networks and use this version in a global context modelling framework. The GC block is used for effectively model long-range dependency. GCNet model is constructed while combining GC blocks to multiple layers of CNN. Figure 7 explains the architecture of main block of GCNet.

Rudra PK Poudel and Stephan Liwicki introduce a Fast-SCNN model for the semantic segmentation of high-quality image of 1024 $\times$ 2048px on real time scenarios. The process illustrated that the large sample pre- training is unnecessary on an additional auxiliary task for low-capacity network. The model is inspired by Encoder Decoder method with skip connections. Skip connections shown efficient results for spatial details recovery. The system is based on feature extractor model that is used for creating the global context of image segmentation, fusion module that add some features for ensuring the efficiency and classifier. The accuracy of this model was 68.0% with 123.5 FPS value [37].

Huikai Wu, Junge Zhang and Kaiqi Huang [38] works on FastFCN approach for overcoming the memory and computational complexity problem by using Joint Pyramid Up sampling (JPU) approach. JPU show best result for reducing the computational complexity of the model with respect to unsampling scenario. DilatedFCN [39] removes the last two down samples of FCN that increase the computational complexity. FastFCN approach overcome the threshold of DilatedFCN by removing all the dilated convolutions layers and replacing them with regular convolutional layers and introducing GPU. As compared to DilatedFCN and ResNet, FastFCN use 4 time less computational resources and memory. The results shows that the JPU model is better than other unsampling models. The state of the art performance of the model shows 53.13% mIoU for Pascal context dataset.

Self-attention mechanism(SAM) has a complex computational consuming [40]. To overcome the computational complexity problem of Self-attention semantic

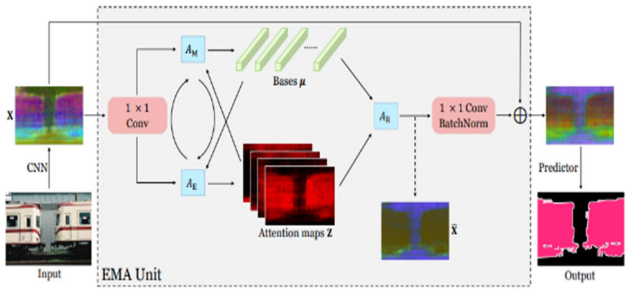


FIGURE 8. Structure of EMA unit.

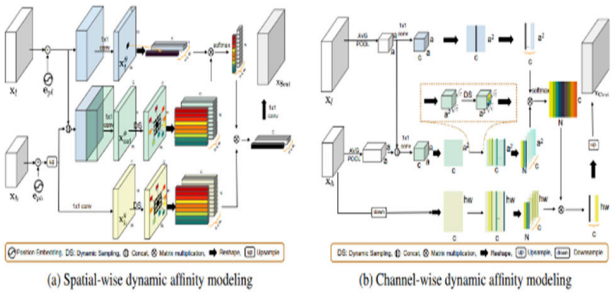


FIGURE 9. a) Spatial wise Dynamic Affinity Modelling b) Channel wise dynamic affinity modelling.

segmentation Xia Li, Zhisheng Zhong, Jianlong Wu, Yibo Yang, Zhouchen Lin and Hong Liu1 [41] proposed Expectation-Maximization Attention (EMA) method which is an augmented version of SAM. Exception maximizing is an iterative process for finding the maximum posteriori for the estimation of parameters in statistical models. Unlike other models of selecting all data points as base Expectation-Maximization find the compressed basis set. The structure of EMA unit is explained in Figure 8 [42].

EMA-net gives the mIoU value of 88.2% for PASCAL VOC dataset. Junjun He, Zhongying Deng and Yu Qiao [43] proposed Dynamic Multi-scale Network (DMNet) based on different filters for semantic segmentation. DMNet is collection of different Dynamic Convolutional Modules that are parallel connected with each other serve as the backbone of the model. The output from these DCM's is further pass to final segmentation. To handle the scale variations of objects DCM use context-aware filters. These filters are enthusiastically generated from input image features and inserted with high-level semantics, which then capture more details. The model gives the maximum of 84.4% mIoU on PASCAL VOC 2012 dataset. Another research work on Dual Attention Network (DANet) for scene segmentation with the integration of local features with global independencies.

Two type of attention modules position attention module and channel attention are used with Dilated FCN in this process to enhance the feature representation ability of different scenes. Position attention module learn the spatial dependencies of features and a channel attention module learn channel interdependencies for improving the segmentation. The study gives best mIoU results of 81.5% for Cityscapes dataset [44].

Asymmetric Non-local Neural Network is also proposed by the researcher for semantics segmentation. APNB is used for reducing the memory consumption and computational loss while ANFB is used for the fusion of different features in this scenario. Both of APNB and ANFB serve as sampling pyramid for this network [45]. The model is trained and test on cityscapes dataset and give the mIoU value of 81.3. Encode-Decoder technique is used for reducing the feature maps. CRF in machine learning used for contextual information. The model works on two supervisions one is at the end where the model gets the output and the other one was the stage 4 of output layer. The code used for the study is based on semantic segmentation using Pytorch. Stochastic Gradient Descent serve as optimizer [46]. The testing is done by left-right flip testing. The sampling methods plays an important role in the system performance. The sampling is based on random, max, and average selection.

Chen Shi, Xiangtai Li, Yanran Wu, Yunhai Tong and Yi Xu [47] used Dynamic Duel Sampling Module (DDSM) for sematic segmentation. This study focusses on finding the relationship between local detail layers and semantic context. DDSM model contains dynamic affinity models for spatial and channel wise. The DDSM model works in two modules.

1) Spatial wise Dynamic Affinity Modelling (SDAM)
2) Channel-wise Dynamic Affinity Modelling (CDAM)

In SDAM the features are samples from the higher layer to the lower layer. CDAM finds the dependencies among different channels. Figure 9 explains the working process of both of these processes [47].

The method is implemented in Pytorch framework. Cityscapes and CamVid dataset are used in the study. The evaluation matrixes mIoU and F-Score is used in the results.

The process gives the mIoU value of 81.7% for cityscapes dataset. Peng et al. proposed PP-LiteSeg model for semantic segmentation. This model is a light weighted approach use three modules as Flexible and Lightweight Decoder (FLD), Unified Attention Fusion Module (UAFM) and Simple Pyramid Pooling Module (SPPM).

The purpose of using these are

- FLD is used in LiteSeg for real time semantic segmentation.
- UAFM is used for feature representations. It uses channel and spatial attentions for features weight fusing.
- SPPM is used for the purpose of aggregation of global contexts.

The process is inspired by encoder-decoder framework. The working process of PP-LiteSeg is explained in Figure 10 [48].

The main contribution of this study in semantic segmentation includes.

- The process mitigates the redundancy of the decoder and balance the cost.
- Lessen the extra inference time and increase the segmentation accuracy.

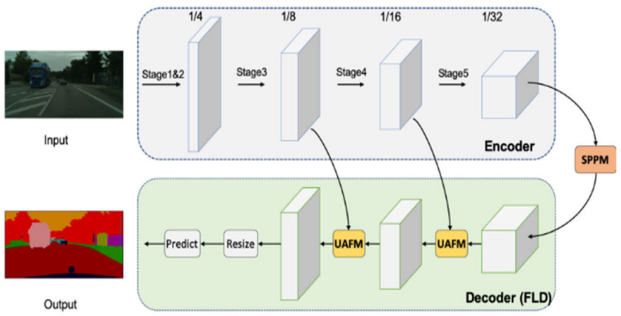


FIGURE 10. PP-LiteSeg working framework.

- Use the spatial feature for strengthening the feature representation.
- State of the Art performance in terms of accuracy and speed for semantic segmentation.

This model achieves 72.0% mIoU and 273.6 FPS on cityscapes dataset. A study was proposed for removing the background from the real time video of the conferencing participants using PP-HumanSeg [49]. The dataset used for the study is taken from 291 videos of different conferences along Semantic Connectivity-aware Learning (SCL) for portrait segmentation using deep learning algorithms. The pixel level labelling has two ambiguous events i.e., Handheld items and people at the back of the video. The SC loss function is used for modelling the components of portrait segmentation and measure the inconsistency using ground truth of the labels and predictions. This model shows the mIoU value of 94.6 under different weight coefficients.

Alexander Kirillov Yuxin Wu Kaiming He and Ross Girshick [50] presents efficient rendering method for image segmentation. The process used PointRend neural network. PointRend is a machine learning algorithm that is based on iterative algorithm that use point base segmentations for prediction. This model is fit for the prediction of semantic and instance segmentation. The model is implemented on two mostly used semantic segmentation dataset COCO and Cityscapes. The rendering algorithm gives the regular pixel grids taken from random instances. The process of PointRend consists of three components.

- Point selection process that selects different randomly points from the output grid.
- Feature representation process that is used to represents point wise features extracted from point selection.
- Prediction of labels from the feature selection ResNet serves as the backbone of the PointRend. FCN with four convolutional layers is used for masking. The model gives the mIoU value of 78.6 for Cityscapes dataset.

In the recently developed researches Class-wise dynamic graph convolution network (CC) is also used for semantic segmentation [51]. This network is based on two main parts

1) CDGC Module
2) Segmentation Network

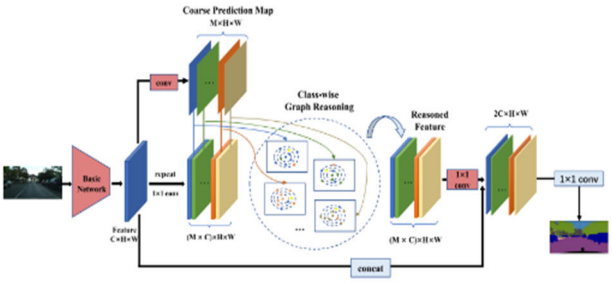


FIGURE 11. Overview of CDGCNet.

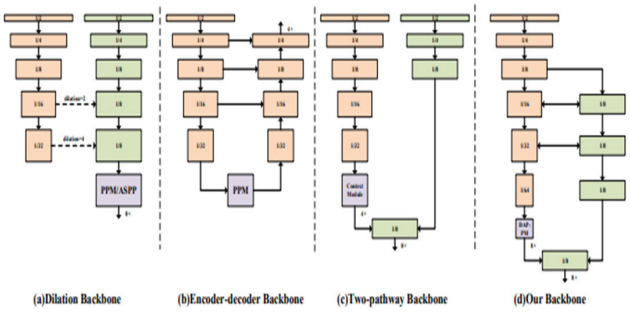


FIGURE 12. DDRnets architecture.

This study is developed on three datasets COCO, PASCAL, and Cityscapes. The study considered class wise learning for contextual learning. Hard negative and positive information are samples for segmentation results which benefits the feature learning. Figure 11 explains the overview of the network [51].

CDGC Module use ResNet 101 as backbone of the architecture along atrous spatial pyramid pooling and provide class wise graph reasoning. The graph is constructed based on the similarities between the features of different nodes. The model demonstrated on the dataset and give the mIoU value of 81.45 in ablation study [52]. GiNet also provides Graph Interaction unit AND SC loss for context reasoning. GI unit is very useful for enhancing the feature representation and promote contextual information. SC loss used by GiNet for improvement of semantic representation.

Deep learning method of high resolution representation is also an essential techniques for the identification of objects, human pose and semantic segmentation. In the other processes the image gets low resolution after encoding, while the HRNet maintain the high resolution of the image throughout the process. The features of HRNet are [53] includes Exchange of the information through the resolution and Parallel multi resolution.

The Multi resolution streams add high to low streams one by one forming a parallel stage. HRNet for human pose estimation locates the points from the parts of body include elbow, chin, wrist etc. 2D Gaussian filters are applied on the images for generating the ground truth heat map. COCO dataset is used in the study along Object Key point Similarity

TABLE 2. Literature review table.

Paper Title	Publication year	Dataset used by the study	Method	Result
ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation	2016	CamVid, SUN, RGB-D and Cityscapes	Deep learning	SegNet 70.3% Enet 59.5%.
Context Encoding for Semantic Segmentation	2018	Pascal context and ADE20K	Context Encoding Module	85.9% mIoU
Encoder-decoder with atrous separable convolution for semantic image segmentation.	2018	PASCAL VOC	DeepLabv3	89% object detection
BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation	2018	Cityscapes, COCO Stuff and CamVid	CNN BiSeNet	74.8% mIoU
Attention u-net: Learning where to look for the pancreas.	2018	CT Abdominal	Ags with U-Net	0.835 precision 0.824 Recall
PortraitNet: Real-time portrait segmentation network for mobile device	2018	EG1800 Supervise-Portrait dataset	U-Net Portrait segmentation	EG1800 96.62% Supervise-Portrait 93.4%
Panoptic Feature Pyramid Networks	2018	COCO and city spaces	R-CNN with FPN	79.1% mIoU
Object-contextual representations for semantic segmentation	2019	ADE20K, Cityscapes, LIP, COCO-Stuff and PASCAL-Context	Object Contextual representation	81.8% object detection
Interlaced Sparse Self-Attention for Semantic Segmentation	2019	Cityscapes ADE20K LIP PASCAL VOC 2012. COCO-Stuff Network	self-attention method	81.4% mIoU for Cityspace dataset.
Hardnet: A low memory traffic network.	2019	CamVid ImageNet PASCAL VOC COCO	Harmonic Densely connected network	GPU interface reduction time by 35% and DRAM traffic about 40%
Gated-scnn: Gated shape cnns for semantic segmentation.	2019	Cityscapes	Gated-SCNN	82.5% object detection
Graph-based global reasoning networks	2019	ImageNet Cityscapes Kinetics	Graph-Based Global Reasoning Networks	79.8%
GCNet: Non-local networks meet squeeze-excitation networks and beyond	2019	COCO Kinetics	Global context network (GCNet)	
Fast-scnn: Fast semantic segmentation network.	2019	Cityscapes	Fast-SCNN	68.0% Accuracy 123.5 FPS value.
FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation.	2019	Pascal Context ADE20K	FastFCN using JPU	53.13% mIoU for Pascal
Expectation-Maximization Attention Networks for Semantic Segmentation	2019	PASCAL VOC, PASCAL Context COCO Stuff	Expectation-Maximization Attention (EMA)	88.2% mIoU for PASCAL VOC
Dynamic Multi-scale Filters for Semantic Segmentation	2019	PASCAL VOC 2012, Pascal-Context, ADE20K.	Dynamic Multi-scale Network	84.4% mIoU
CCNet: Criss-cross attention for semantic segmentation	2019	Cityscapes, PASCAL Context and COCO Stuff dataset	Dual Attention Network (DANet)	81.5% mIoU
Asymmetric Non-local Neural Networks for Semantic Segmentation	2019	Cityscapes ADE20K PASCAL Context	Asymmetric Non-local Neural Network	81.3% mIoU
DYNAMIC DUAL SAMPLING MODULE FOR FINE-GRAINED SEMANTIC SEGMENTATION	2020	Cityscapes and CamVid datasets	Dynamic Duel Sampling Module	81.7 mIoU

TABLE 2. *(Continued.)* **Literature review table.**

PP-LiteSeg: A Superior Real-Time Semantic Segmentation Model	2020	Cityscapes NVIDIA GTX	Flexible and Lightweight Decoder with Simple Pyramid Pooling Module (SPPM)	77.5% mIoU
PP-HumanSeg: Connectivity-Aware Portrait Segmentation with a Large-Scale Teleconferencing Video Dataset	2020	EG1800 FVS AISeg	Semantic Connectivity-aware Learning (SCL)	94.6% mIoU
PointRend: Image Segmentation as Rendering	2020	COCO Cityscapes	PointRend neural network	78.6% mIoU
Class-Wise Dynamic Graph Convolution for Semantic Segmentation	2020	COCO, PASCAL and Cityscapes	CDGCNet	81.4% mIoU
Deep High-Resolution Representation Learning for Visual Recognition	2020	COCO	High resolution representation Network (HRNet)	81.6% mIoU
Disentangled Non-Local Neural Networks	2020	Cityscapes, PASCAL Context, ADE20K	DnlNet	83.0% mIoU
Improving semantic segmentation via decoupled body and edge supervision	2020	Cityscapes, KIITI, CamVid BDD datasets	decoupled_segnet	83.7% mIoU
Rethinking BiSeNet For Real-time Semantic Segmentation	2021	Cityscapes NVIDIA	STDC Network	79.1% mIoU
Exploring Cross-Image Pixel Contrast for Semantic Segmentation	2021	Cityscapes Pascal CamVid	pixel-wise contrastive algorithm	83.5% mIoU
Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes	2021	Cityscapes CamVid	Deep dual-resolution networks	77.4% mIoU

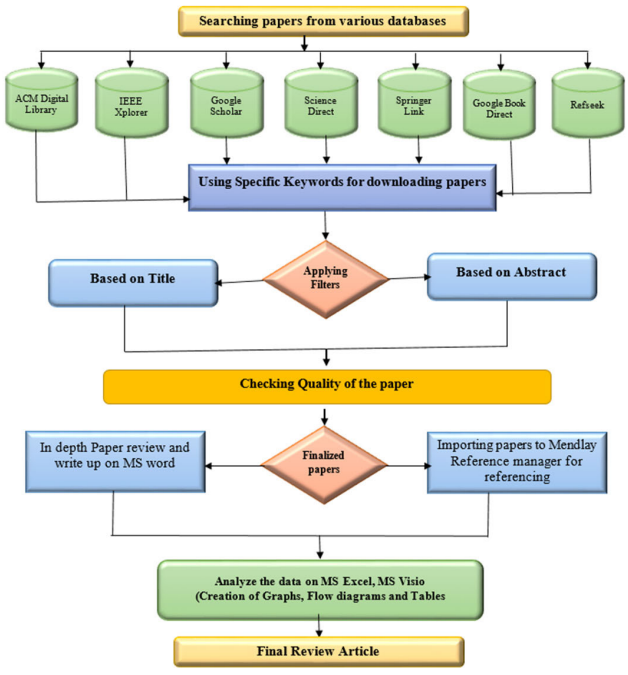


FIGURE 13. **Working method for the proposed systematic literature review.**

(OKS) evaluation matrix [54]. The key different between the proposed model and the previous models was.

- HRNet connects high and low resolution convolutions in parallel while the old models kept them in series.
- HRNet always Maintain high resolution of the image throughout the whole process while other models recover the high resolution of the images from the low resolutions.
- Fuse multi-resolution representations repeatedly.

HRNet has a very small inference time but takes large training time. The HRNet give the mIoU of 81.65 for COCO dataset. Li et al. improve the process of semantic segmentation using edge supervision [55]. The model is trained on Cityscapes, KIITI, CamVid and BDD datasets. The model achieves a best mIoU value of 83.7%. A research also proposed short term dense connected network for reducing the dimension features of the images in the network and produces a final segmentation with the fusion of deep features and low level features [56].

Wang et al. [57] proposed pixel wise contrastive representation for semantic segmenting, by enhancing the pixel wise matrix learning. This study uses the old methods with ResNet and HRNet as backbone and apply pixel wise contrastive representation for generating the better results. In one of the latest study the researchers used DDRnets with multiple bilateral fusions that receive 77.95 mIoU on cityscapes dataset. The architecture of the model is explained in Figure 12 [58].

TABLE 3. Research papers with maximum Citations.

Authors	Paper title	Year of publication	Citations	Technique/method
Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam [18]	Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation	2018	7263	DeepLabV3 Segmentation using Atrous convolutional
Jun Fu; Jing Liu; Haijie Tian; Yong Li; Yongjun Bao; Zhiwei Fang; Hanqing Lu [44]	Dual Attention Network for Scene Segmentation	2019	2858	Dilated FCN
Ozan Oktay, Jo Schlemper , Loic Le Folgoc , Matthew Lee , Mattias Heinrich , Kazunari Misawa , Kensaku Mori , Steven McDonagh , Nils Y Hammerla , Bernhard Kainz , Ben Glocker , and Daniel Rueckert [22]	Attention U-Net: Learning Where to Look for the Pancreas	2018	1924	U-Net CNN model
Abhishek Chaurasia, Sangpil Kim, Eugenio Culurciello [11]	ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation	2016	1648	Efficient Neural Network
Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao [53]	Deep High-Resolution Representation Learning for Visual Recognition	2020	1233	High resolution representation Network (HRNet)
Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang [19]	BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation	2018	1163	Bilateral Segmentation Network
Hang Zhang , Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, Amit Agrawal [15]	Context Encoding for Semantic Segmentation	2018	962	Context Encoding Module
Yue Cao, Jiarui Xu, Stephen Lin , Fangyun Wei3 , Han Hu [36]	GCNet: Non-local Networks Meet Squeeze-Excitation Networks and Beyond	2019	900	Global Context Network
Alexander Kirillov, Ross Girshick, Kaiming He, Piotr Dollar [24]	Panoptic Feature Pyramid Networks	2019	633	Panoptic Feature Pyramid Networks
Towaki Takikawa, David Acuna, Varun Jampani1 Sanja Fidler [31]	Gated-SCNN: Gated Shape CNNs for Semantic Segmentation	2019	411	Gated Stream Architecture
Zhen Zhu , Mengde Xu1 , Song Bai , Tengteng Huang , Xiang Bai1 [45]	Asymmetric Non-local Neural Networks for Semantic Segmentation	2019	375	Asymmetric Non-local Neural Networks

All the papers selected for the study proposed novel approaches for the semantic segmentation scenario. All the studies have their own benefits and limitations explained in the above section. Most of the studies use mobile camera for efficiently detecting real time segmentation. The results of the studies vary according to the available dataset, proposed model, and hardware. The results of the selected research articles are collected in the form of accuracy, mIoU value, and F1 score and precision. The comparison of the results and techniques used in these papers are presented in Table 2.

VI. ANALYSIS AND DISCUSSION

In this section of the proposed study the results of the systematic literature review are discussed. The overall working of the whole scenario of the review paper is explained in Figure 13.

Figure 13 explains that the research articles for the review are selected from seven different research articles database. From these research articles the high quality articles are selected for the review purpose after applying filters. After extracting the 44 research papers, these papers are reviewed deeply, their results are analyzed, and a final draft of the review study is generated. All the papers are selected between the years 2016 to 2021. The frequency distribution of the papers is shown in Figure 14.

The heat map [55] diagram for all the selected years is shown in Figure 15.

All the papers selected for the study are collected from high source journals and conferences with maximum citations are used in this study for maintaining the quality of the research. The 11 most cited papers for the study are explained in Table 3. Machine learning and deep learning are also used in other areas as discussed in [59], [60], [61], and [62]

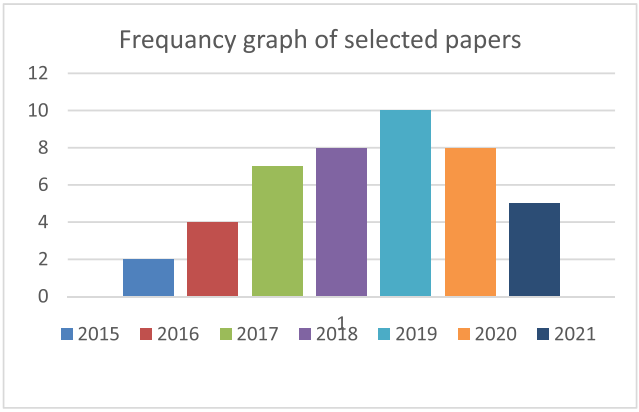


FIGURE 14. Frequency graph of selected papers.

	2016	2017	2018	2019	2020	2021
Enet	1					
ecnet			1			
Deeplabv3p			1			
BiseNet		1			1	1
Attensio U-Net			1			
Potranet		1			1	
pfpn				1		
ocrnet				1		
isnet				1		
gscnn				1		
Glore					1	
gcnet			1		1	
fastscnn				1		
Danet		1		1		
CCNet	1				1	
ANN		1		1	1	1
SFNET					1	
PP_liteseg			1			1
ginet						1
fcn						1
dnlnet					1	
DDRnet						1
Hardnet	1			1		1
amanet				1		
Dmnet		1		1		

FIGURE 15. Heat map diagram of survey papers.

VII. CONCLUSION

The aim of the proposed study is to make state of the art baseline for the researchers to have the comparative knowledge of selected machine learning and deep learning methods for semantic segmentation. This study selects 44 research articles collected from seven different research databases. For assuring the quality of the paper those papers are selected which have several citations on standardized platforms. The proposed study comprehends the concepts, techniques, tool and results of different research frameworks proposed in the context of semantic segmentation.

All the papers selected for the study proposed novel approaches for the semantic segmentation. All the studies have their own benefits and limitations, and both are explained in the literature review segment. Most of the studies efficiently detect the real time segmentation using mobile phone camera. This study mainly considers on deep learning and machine the techniques that are used for the semantic segmentations. The results of the studies are according to the available dataset, proposed model, and available hardware resources for processing. The comparison of the techniques used in these papers and results are presented in Table 2. It is concluded that most of the semantic segmentation studies use ResNet as the backbone of the architecture. Almost all the studies use COCO, PASCAL, CamVid, Cityscapes and ADK dataset to train and test the proposed model. The results of the selected research articles are collected in the form of accuracy, mIoU value, and F1 score, precision, and recall. The best mIoU of 96.6% is obtained from PotraNet 96.62% [23] Semantic Connectivity-aware Learning (SCL) in PP-HumanSeg gives the mIoU value of 94.6% [49], Encoder decoder method gives the object detection of 89% [37], and the mIoU of 88.2% is obtained by EMA-net [43].

VIII. FUTURE WORK

As the loophole identified by the proposed study, none of the article use deep ensemble learning model for semantic segmentation. Different versions of RNN model along with ensemble learning approach can show excellent results for semantic segmentation. In future we are going to use deep ensemble techniques with RNN for the semantic segmentation of images.

REFERENCES

[1] D. J. Fleet, *Example Segmentations: Simple Scenes*. 2007, pp. 1–34.
[2] A. A. Shah, H. A. M. Malik, A. Mohammad, Y. D. Khan, and A. Alourani, "Machine learning techniques for identification of carcinogenic mutations, which cause breast adenocarcinoma," *Sci. Rep.*, vol. 12, no. 1, pp. 1–15, Jul. 2022, doi: 10.1038/s41598-022-15533-8.
[3] S. Ilyas, A. A. Shah, and A. Sohail, "Order management system for time and quantity saving of recipes ingredients using GPS tracking systems," *IEEE Access*, vol. 9, pp. 100490–100497, 2021, doi: 10.1109/ACCESS.2021.3090808.
[4] Y. H. Bhosale and K. S. Patnaik, "Application of deep learning techniques in diagnosis of COVID-19 (Coronavirus): A systematic review," *Neural Process. Lett.*, to be published, doi: 10.1007/s11063-022-11023-0.
[5] A. A. Shah, M. K. Ehsan, A. Sohail, and S. Ilyas, "Analysis of machine learning techniques for identification of post translation modification in protein sequencing: A review," in *Proc. Int. Conf. Innov. Comput. (ICIC)*, Nov. 2021, pp. 1–6, doi: 10.1109/ICIC53490.2021.9693020.
[6] C. Sutton and A. Mccallum, "An Introduction to Conditional Random Fields for Relational Learning," in *Introduction to Statistical Relational Learning*. 2019, doi: 10.7551/mitpress/7432.003.0006.
[7] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007, doi: 10.1016/j.jss.2006.07.009.
[8] K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, "Five steps to conducting a systematic review," *J. Roy. Soc. Med.*, vol. 96, no. 3, pp. 118–121, Mar. 2003, doi: 10.1258/jrsm.96.3.118.
[9] A. Ramdhani, M. A. Ramdhani, and A. S. Amin, "Writing a literature review research paper: A step-by-step approach," *Int. J. Basic Appl. Sci.*, vol. 3, no. 1, pp. 47–56, 2014.
[10] L. A. Kahale, R. Elkhoury, I. El Mikati, H. Pardo-Hernandez, A. M. Khamis, H. J. Schünemann, N. R. Haddaway, and E. A. Akl, "PRISMA flow diagrams for living systematic reviews: A methodological survey and a proposal," *FResearch*, vol. 10, p. 192, Mar. 2021, doi: 10.12688/f1000research.51723.1.

[11] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*.

[12] J. Zhuang, J. Yang, L. Gu, and N. Dvornek, "ShelfNet for fast semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 847–856, doi: 10.1109/ICCVW.2019.00113.

[13] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for robust semantic pixel-wise labelling," 2015, *arXiv:1505.07293*.

[14] 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," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223, doi: 10.1109/CVPR.2016.350.

[15] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7151–7160, doi: 10.1109/CVPR.2018.00747.

[16] X. Wang, R. Lv, Y. Zhao, T. Yang, and Q. Ruan, "Multi-scale context aggregation network with attention-guided for crowd counting," in *Proc. 15th IEEE Int. Conf. Signal Process. (ICSP)*, Dec. 2020, pp. 240–245, doi: 10.1109/ICSP48669.2020.9321067.

[17] A. Z. M. Everingham, L. V. Gool, C. Williams, and J. Winn, "The PASCAL visual object classes challenge 2012 (VOC2012) results," Tech. Rep., 2012, pp. 1–32.

[18] M. Firdaus-Nawi, O. Noraini, M. Y. Sabri, A. Siti-Zahrah, M. Zamri-Saad, and H. Latifah, "DeepLabv3+_encoder–decoder with atrous separable convolution for semantic image segmentation," *Pertanika J. Trop. Agric. Sci.*, vol. 34, no. 1, pp. 137–143, 2011.

[19] C. Yu, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 325–341.

[20] K. Sumiya and K. Tanaka, "Summarization and presentation for web documents using context paths," in *Proc. ISDB*, Jan. 2002, pp. 1–9.

[21] S. Hui, S. Zhou, Y. Deng, W. Huang, and J. Wang, "Auxiliary loss reweighting for image inpainting," 2021, *arXiv:2111.07279*.

[22] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[23] S.-H. Zhang, X. Dong, H. Li, R. Li, and Y.-L. Yang, "PortraitNet: Real-time portrait segmentation network for mobile device," *Comput. Graph.*, vol. 80, pp. 104–113, May 2019, doi: 10.1016/j.cag.2019.03.007.

[24] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6392–6401, doi: 10.1109/CVPR.2019.00656.

[25] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Lecture Notes Computer Science (Including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 12351. 2020, pp. 173–190, doi: 10.1007/978-3-030-58539-6_11.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[27] L. Huang, Y. Yuan, J. Guo, C. Zhang, X. Chen, and J. Wang, "Interlaced sparse self-attention for semantic segmentation," 2019, *arXiv:1907.12273*.

[28] G. Xie, Q. Li, Y. Jiang, T. Dai, G. Shen, and R. Li, "SAM: Self-attention based deep learning method for online traffic classification," *Proc. Work. Netw. Meets AI ML (NetAI)*, pp. 14–20, Aug. 2020, doi: 10.1145/3405671.3405811.

[29] M. Prazeres and A. M. Oberman, "Stochastic gradient descent with Polyak's learning rate," *J. Sci. Comput.*, vol. 89, no. 1, p. 25, 2021, doi: 10.1007/s10915-021-01628-3.

[30] P. Chao, C.-Y. Kao, Y. Ruan, C.-H. Huang, and Y.-L. Lin, "HarDNet: A low memory traffic network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3551–3560, doi: 10.1109/ICCV.2019.00365.

[31] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5228–5237, doi: 10.1109/ICCV.2019.00533.

[32] X. Lian, Y. Pang, J. Han, and J. Pan, "Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107622, doi: 10.1016/j.patcog.2020.107622.

[33] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proc. 35th AAAI Conf. Artif. Intell.*, vol. 10, 2021, pp. 8801–8809.

[34] A. Ahmadzadeh, D. J. Kempton, Y. Chen, and R. A. Angryk, "Multi-scale IOU: A metric for evaluation of salient object detection with fine structures," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 684–688, doi: 10.1109/ICIP42928.2021.9506337.

[35] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 433–442, doi: 10.1109/CVPR.2019.00052.

[36] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1971–1980, doi: 10.1109/ICCVW.2019.00246.

[37] R. P. K. Poudel, S. Liwicki, and R. Cipolla, "Fast-SCNN: Fast semantic segmentation network," in *Proc. 30th Brit. Mach. Vis. Conf.*, 2019.

[38] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation," 2019, *arXiv:1903.11816*.

[39] S. Gong, Z. Wang, T. Sun, Y. Zhang, C. D. Smith, L. Xu, and J. Liu, "Dilated FCN: Listening longer to hear better," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2019, pp. 254–258, doi: 10.1109/WASPAA.2019.8937212.

[40] Z. Zhou, Y. Zhou, D. Wang, J. Mu, and H. Zhou, "Self-attention feature fusion network for semantic segmentation," *Neurocomputing*, vol. 453, pp. 50–59, Sep. 2021, doi: 10.1016/j.neucom.2021.04.106.

[41] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, and H. Liu, "Expectation-maximization attention networks for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9166–9175, doi: 10.1109/ICCV.2019.00926.

[42] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1997, doi: 10.1109/79.543975.

[43] J. He, Z. Deng, and Y. Qiao, "Dynamic multi-scale filters for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3561–3571, doi: 10.1109/ICCV.2019.00366.

[44] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149, doi: 10.1109/CVPR.2019.00326.

[45] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 593–602, doi: 10.1109/ICCV.2019.00068.

[46] A. Tato and R. Nkambou, "Improving ADAM optimizer," in *Proc. Work. Track-ICLR*, 2018, pp. 1–4. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[47] C. Shi, X. Li, Y. Wu, Y. Tong, and Y. Xu, "Dynamic dual sampling module for fine-grained semantic segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021.

[48] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, and G. Chen, "PP-LiteSeg: A superior real-time semantic segmentation model," 2022, *arXiv:2204.02681*.

[49] Z. Chen, B. Online, N. Technology, H. Xiong, B. Online, and N. Technology, "PP-HumanSeg: Connectivity-aware portrait segmentation," Tech. Rep., Dec. 2021.

[50] A. Kirillov, Y. Wu, K. He, and R. Girshick, "PointRend: Image segmentation as rendering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 9796–9805, doi: 10.1109/CVPR42600.2020.00982.

[51] F. Yu, M. Kumar, and K. Reddy, *Computer Vision–(ECCV)*, vol. 12350. 2020, doi: 10.1007/978-3-030-58558-7.

[52] R. Meyes, M. Lu, C. Waubert de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks," 2019, *arXiv:1901.08644*.

[53] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, and D. Liu, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021, doi: 10.1109/TPAMI.2020.2983686.

[54] M. R. Ronchi and P. Perona, "Benchmarking and error diagnosis in multi-instance pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 369–378, doi: 10.1109/ICCV.2017.48.

[55] X. Li, X. Li, L. Zhang, G. Cheng, J. Shi, Z. Lin, and S. Tan, "Improving Semantic Segmentation via Decoupled Body and Edge Supervision," in *Lecture Notes Computer Science (Including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 12362. 2020, pp. 435–452, doi: 10.1007/978-3-030-58520-4_26.
[56] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, and X. Wei, "Rethinking BiSeNet for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9711–9720, doi: 10.1109/CVPR46437.2021.00959.
[57] W. Wang, T. Zhou, F. Yu, J. Dai, E. Konukoglu, and L. V. Gool, "Exploring cross-image pixel contrast for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7283–7293, doi: 10.1109/ICCV48922.2021.00721.
[58] Y. Hong, H. Pan, W. Sun, and Y. Jia, "Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes," 2021, *arXiv:2101.06085*.
[59] A. A. Shah, F. Alturise, T. Alkhalifah, and Y. D. Khan, "Deep learning approaches for detection of breast adenocarcinoma causing carcinogenic mutations," *Int. J. Mol. Sci.*, vol. 23, no. 19, p. 11539, Sep. 2022.
[60] I. Qadeer and M. Khurram Ehsan, "Improved channel reciprocity for secure communication in next generation wireless systems," *Comput., Mater. Continua*, vol. 67, no. 2, pp. 2619–2630, 2021.
[61] A. A. Shah, F. Alturise, T. Alkhalifah, and Y. D. Khan, "Evaluation of deep learning techniques for identification of sarcoma-causing carcinogenic mutations," in *Digital Health*. London, U.K.: Sage, 2022.
[62] S. Saeed, A. A. Shah, M. K. Ehsan, M. R. Amirzada, A. Mahmood, and T. Mezgebo, "Automated facial expression recognition framework using deep learning," *J. Healthcare Eng.*, vol. 2022, pp. 1–11, Mar. 2022.

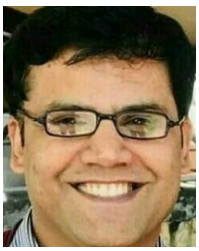
ALI SOHAIL received the M.Phil. degree in computer science from Minhaj University, Lahore. His research interest includes machine learning.

NAEEM A. NAWAZ received the B.Sc. degree from the University of Punjab, Pakistan, in 1997, the M.Sc. degree in computer science from Hamdard University, Pakistan, in 2000, the M.S. degree in computer engineering from Mid Sweden University, Sweden, in 2008, and the Ph.D. degree in computer science from International Islamic University Malaysia, in 2018. From 2009 to 2011, he worked as a Lecturer with the College of Computer Science, King Khalid University, Saudi Arabia. From 2001 to 2005, he worked as a Teaching Assistant, a Technical Instructor, and a Teaching Fellow with the University of Management and Technology, Pakistan. Currently, he is with the College of Computer and Information Systems, Umm Al-Qura University, Makkah Al-Mukarramah, Saudi Arabia, where he has been, since 2011. He has published research articles in renowned journals and conferences. He is also serving as a reviewer for many journals, conferences, and books. His teaching and research interests include WSN, the IoT, crowd management, computer networks, and programing languages. He received different Diploma (DCS) and Certifications.

ASGHAR ALI SHAH received the M.S. degree in IT from IM|Sciences Peshawar, Pakistan. He is currently an Assistant Professor with the Department of Computer Sciences, Bahria University, Lahore Campus. He is also a Ph.D.—CS Scholar. He has been teaching and research experience in computer sciences, since 2004. He has 30 research articles in reputed journals. His research interests include machine learning, deep learning, computer network security, and bioinformatics.

SAIM RASHEED received the Graduate degree in computer science from the University of Milan, Italy. He is currently working as an Associate Professor at the Department of Information Technology, King Abdulaziz University, Jeddah. His research interests include variety of different areas, such as image processing, computer vision, computer graphics, HCI, electroencephalograph, and brain–computer interaction.

SHEEBA ILYAS received the M.Phil. degree in computer science from Minhaj University, Lahore. Her research interest includes machine learning.



MUHAMMAD KHURRAM EHSAN received the M.S. degree in electrical communication engineering and the Ph.D. degree in engineering with specialization in statistical signal processing from the University of Kassel, Germany, in 2010 and 2016, respectively. He has been working as an Associate Professor with the Faculty of Engineering, Bahria University, Pakistan, since July 2022. He has also been working as a Visiting Lecturer with the Faculty of Electrical Engineering and Computer Science, University of Kassel, since July 2017. As a Reviewer, he is closely working with *Journal of Network and Computer Applications* (Elsevier) and with IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATIONS and IEEE SYSTEMS JOURNAL. His research interests include statistical modeling, data analysis, and cognitive radio enabled systems, including wireless sensor networks and the Internet of Things (IoT).