IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

A Comparative Study of Real-Time Semantic Segmentation for Autonomous Driving

Anush Gupta
Department of Computer Science and Engineering
Dronacharya College of Engineering, Gurgaon, Haryana, India

Abstract: *Semantic Segmentation (SS) is the task to assign a semantic label to each pixel of the observed images, which is of crucial significance for autonomous vehicles, navigation assistance systems for the visually impaired, and augmented reality devices. However, there is still a long way for SS to be put into practice as there are two essential challenges that need to be addressed: efficiency and evaluation criterions for practical application. For specific application scenarios, different criterions need to be adopted. Recall rate is an important criterion for many tasks like autonomous vehicles. For autonomous vehicles, we need to focus on the detection of the traffic objects like cars, buses, and pedestrians, which should be detected with high recall rates. In other words, it is preferable to detect it wrongly than miss it, because the other traffic objects will be dangerous if the algorithm miss them and segment them as safe roadways. In this paper, our main goal is to explore possible methods to attain high recall rate. Firstly, we propose a real-time SS network named Swift Factorized Network (SFN). The proposed network is adapted from SwiftNet, whose structure is a typical U-shape structure with lateral connections. Inspired by ERFNet and Global Convolution Networks (GCNet), we propose two different blocks to enlarge valid receptive field. They do not take up too much calculation resources, but significantly enhance the performance compared with the baseline network. Secondly, we explore three ways to achieve higher recall rate, i.e loss function, classifier and decision rules. We perform a comprehensive set of experiments on state-of-the-art datasets including CamVid and Cityscapes. We demonstrate that our SS convolutional neural networks reach excellent performance. Furthermore, we make a detailed analysis and comparison of the three proposed methods on the promotion of recall rate.*

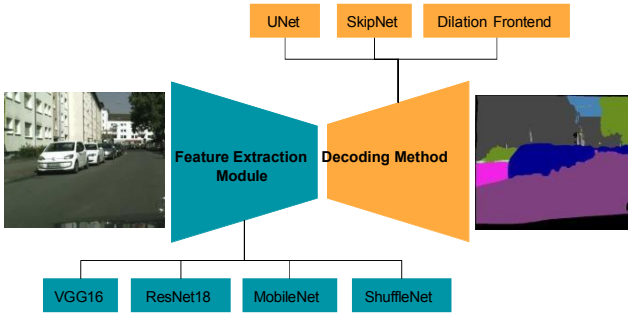
Keywords: Semantic Segmentation

I. INTRODUCTION

Semantic Segmentation (SS) is the task to assign a semantic label to each pixel of the observed images, which is of crucial significance for autonomous vehicles, navigation assistance systems for the visually impaired, and augmented reality devices 1.2 It has gained great development from traditional method into methods based on deep Convolutional Neural Networks (CNNs) since the milestone created by Fully Convolutional Networks (FCN).3 However, there is still a long way for SS to be put into practice as there are two essential challenges that need to be addressed. Firstly, the inference efficiency is paramount for real-time applications with limited computational resources, which should be assured. Secondly, different evaluation criterions have been used in the literature, without consideration of the target application scenario. To address these issues, we aim to lay a good balance between inference speed and performance measured using certain criterion suitable for certain applications instead of merely depending on mean Intersection over Union (mIoU). In this paper, we propose a real-time SS network named Swift Factorized Network (SFN) based on ResNet18,4 a light-weight convolutional neural network specially designed for classification tasks. The proposed network is adapted from SwiftNet,5 which can reach state-of the art performance at a fast inference speed. The structure of the network is a typical U-shape structure with lateral connections. Inspired by ERFNet67 and Globa However, some aspects for semantic segmentation such as computational efficiency has not been thoroughly studied in the literature. Although, when it comes to applications such as autonomous driving this would have tremendous impact. There is little work which address the segment Convolutional Network, We have realized the importance of the receptive field. In order to extract better features and enlarge valid receptive field, we propose two different blocks. The first method is to add a series of factorized convolution blocks with different dilation rate, and the other one is to insert global

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

convolutional blocks with diverse kernel sizes to refine features. They won't take up too much calculation resources, but significantly enhance the performance compared with the baseline network. On the other hand, in order to cope with specific application scenarios, we need to adopt different criterions.



Recall rate is an important criterion for many tasks. Taking autonomous driving as an example, we need to focus on the detection of the traffic objects like cars, buses, and pedestrians, which should be detected with high recall rates. In other words, it is preferable to detect it wrongly than miss it, because the other traffic objects will be dangerous if the algorithm miss them and segment them as safe roadways.

Moreover, for autonomous driving, different objects should be coped with different rank of importance, the traffic light, cars, riders are much more important than road, sky, and etc. If the algorithm does not differentiate the different priorities for certain classes, the module will reach a normal performance for all classes. In order to promote the certain classes' recall rates, we explore three different methods as shown in Fig. 1: (1) Loss Function: Loss function is of great significance for training a model. In order to reach high recall rate, we utilize importance- aware loss.9 (2) Classifier: Graph Convolution Networks (GCN) are very useful and popular to cope with the graph data. It can also be utilized as a classifier, such that the result is based on the pre-defined graph. (3) Decision Rules: The normal decision rule in classification or semantic segmentation tasks is Bayes rule. In our experiment, we also adopt Maximum Likelihood (ML) rule which is promising to boost the recall rate.

II. SEMANTIC SEGMENTATION

The milestone created by FCN results in the prosper for method based on Deep Learning to SS task. The fundamental framework of SS Neural Networks is an Encoder-Decoder structure. Typical representative networks like UNet,12 PSPNet,13 DeepLab V314 and ACNet15 can reach brilliant performance but can not strike a balance between efficiency and accuracy. Therefore, they can not be directly applied into application. In order to put SS into practice, many real-time SS networks have been proposed in recent years. ENet16 is the pioneer in the era of real-time SS networks, which is adapted from ResNet structure4 but discards the last stage of the model to raise efficiency. Following that, many real-time SS Network appear. ERFNet67 and ERF- PSPNet1718 utilize residual factorized module to reduce parameters and keep fine performance. ICNet19 and BiSeNet20 are multiplypath structures using immense calculations on small feature maps for excavating context information and a few calculations on big feature maps to keep spatial information so that they can keep high efficiency and favourable performance. SwiftNet5 and many other real-time SS networks are based on the Encoder-Decoder structure with light- weight base networks like ShuffleNet,21 MobileNet22 and etc., whose performance vary considerably depending on their decoder structures and lateral connections.

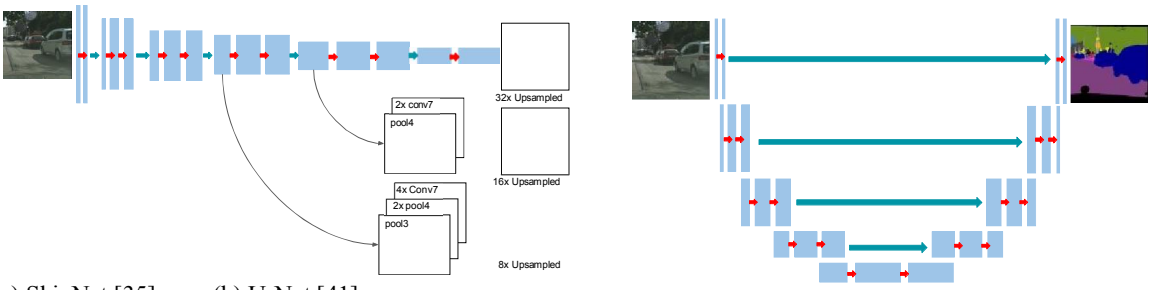
2.1. Fully Convolutional Networks (FCN)

The initial direction in semantic segmentation using con- volutional neural networks was towards patch-wise training [14, 19, 2] to yield the final segmentation. Grangier et al.

[19] proposed a multi-patch training strategy for convolutional neural networks to perform segmentation. Farabet et al. [14, 15] proposed a multi-scale dense feature extrac- tor. The method used a Laplacian pyramid of the image, where each scale is forwarded through a 3-stage network to extract hierarchical features. For each pixel the features are encoded from a contextual patch around the pixel. The scene is then over-segmented into super pixels and condi- tional random fields over the super pixels are used. Bell et al. [2] proposed a method to utilize convolutional neural networks to classify each patch in a sliding window fashion.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

The dominant direction in deep semantic segmentation is to learn pixel-wise classification in an end-to-end man- ner [35, 38, 1]. Long et al. [35] started with proposing fully convolutional networks(FCN). The network learned heatmaps that were then upsampled with-in the network us- ing transposed convolution to get dense predictions. Unlike patch-wise training methods this method uses the full im- age to infer dense predictions. The SkipNet architecture was utilized to refine the segmentation using higher res- olution feature maps. Noh et al. [38] proposed a deeper decoder network, in which stacked transposed convolution and unpooling layers are used. Badrinarayanan et al. [1] proposed SegNet which is an encoder-decoder architecture. The decoder network upsampled the feature maps by keep- ing the maxpooling indices from the corresponding encoder layer. Kendall et al. [28] followed that work by proposing Bayesian SegNet, which incorporates uncertainties in the predictions using dropout during inference. Ronneberger et al. [41] proposed a u-shaped architecture network where feature maps from different encoding layers are concatenated with the upsampled feature maps from the corre- sponding decoding layers. Paszke et al. [39] proposed the use of bottleneck modules for a computationally efficient solution that is denoted as ENet. Figure 3 shows the archi- tecture for FCN8s [35] and U-Net [41].



(a) SkipNet [35] (b) U-Net [41]

Figure 3: Different Decoding methods for fully convolutional networks. Figure reproduced from [35, 41]

2.2. Context Aware Models

Refinements on fully convolutional networks was intro- duced to improve the segmentation accuracy by incorpo- rating context. In this section we consider only the spatial context that does not include any temporal information. The methods to enforce models to become context aware are mainly categorized into multi-scale support, utilizing con- ditional random fields, or recurrent neural networks. Fara- bet et al. [14] handled the scale by introducing multiple rescaled versions of the image to the network. However with the emergence of end-to-end pixel-wise training, Long et al. [35] proposed the skip architecture to merge heat- maps from different resolutions. Since these architectures include pooling layers to increase the receptive field, this leads to the downsampling of the image with a loss in the resolution.

Yu et al. [60] introduced dilated or atrous convolutions, which expanded the receptive field without losing resolu- tion based on the dilation factor. Thus it provided a better solution for handling multiple scales. Wu et al. [59] pro- posed a shallower network using residual connections that included dilated convolution and outperformed deeper mod- els. Chen et al. [7] proposed DeepLab that uses atrous spa- tial pyramid pooling (ASPP) for multi-scale support. This idea builds on utilizing the dilated convolutions. Figure 4 shows dilated convolutions and spatial pyramid pooling as separate methods that can be used to incorporate multi-scale support. Zhao et al. [64] proposed to incorporate global context features from previous layers into the next layers. Chen et al. [8] refined further the DeepLab method by in- corporating global context features. Chen et al. [9] provided a way for handling scale by using attention models that pro- vides a mean to focus on the most relevant features. This attention model is able to learn a weight map, that weighs feature maps pixel-by-pixel from different scales. Eigen et al. [13] proposed a method to sequentially utilize multiple scales to refine the prediction of depth, surface normals, and semantic segmentation.

One of the commonly used models to incorporate con- text is conditional random field (CRF). Chen et al. [7] uti- lized the fully connected conditional random fields as a post processing. The unary potentials of the CRF are set to the probabilities from their convolutional network, while pair- wise potentials are gaussian kernels based on the spatial and color features. Lin et al [34] proposed a method to use pair- wise potentials based on convolutional neural networks fea- ture maps. In contrast to the previous work that uses condi- tional random fields as post processing refinement step, this work went further in integrating CNNs and CRFs. Zheng et al. [65] formulated the mean field CRF inference algorithm

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

as a recurrent network. Thus, the proposed method enabled the end-to-end training of the model.

Another way to incorporate context is using recurrent neural networks (RNN) to capture the long range depen- dencies of various regions. Visin et al. [57] used a recur- rent layer to sweep the image horizontally and vertically, which ensures the usage of contextual information for a bet- ter segmentation. One of the main bottlenecks in vanilla RNN is the vanishing gradients problem, gated recurrent architectures such as LSTMs [23] and GRUs [10] alleviate this problem. Byeon et al. [5] proposed a segmentation method that splits the image into non overlapping regions, then incorporates context using four separate LSTM blocks. Li et al. [32] proposed a method for context fusion using LSTMs. In their work both RGB and depth information were utilized, and the global context was modeled verti- cally on both, followed by the horizontal fusion. Another bottleneck in vanilla recurrent networks is that it could lead to the loss in spatial relationships. Shuai et al. [46] utilized directed acyclic graph RNN to incorporate long range de- pendencies. This directed acyclic graph maintains spatial relations-ships unlike using chained RNNs. Finally, Qi etnated with the upsampled feature maps from the corre- sponding decoding layers. Paszke et al. [39] proposed the use of bottleneck modules for a computationally efficient solution that is denoted as ENet. Figure 3 shows the archi- tecture for FCN8s [35] and U-Net [41].

2.3. Context Aware Models

Refinements on fully convolutional networks was intro- duced to improve the segmentation accuracy by incorpo- rating context. In this section we consider only the spatial context that does not include any temporal information. The methods to enforce models to become context aware are mainly categorized into multi-scale support, utilizing con- ditional random fields, or recurrent neural networks. Fara- bet et al. [14] handled the scale by introducing multiple rescaled versions of the image to the network. However with the emergence of end-to-end pixel-wise training, Long et al. [35] proposed the skip architecture to merge heat- maps from different resolutions. Since these architectures include pooling layers to increase the receptive field, this leads to the downsampling of the image with a loss in the resolution.

Yu et al. [60] introduced dilated or atrous convolutions, which expanded the receptive field without losing resolu- tion based on the dilation factor. Thus it provided a better solution for handling multiple scales. Wu et al. [59] pro- posed a shallower network using residual connections that included dilated convolution and outperformed deeper mod- els. Chen et al. [7] proposed DeepLab that uses atrous spa- tial pyramid pooling (ASPP) for multi-scale support. This idea builds on utilizing the dilated convolutions. Figure 4 shows dilated convolutions and spatial pyramid pooling as separate methods that can be used to incorporate multi-scale support. Zhao et al. [64] proposed to incorporate global context features from previous layers into the next layers. Chen et al. [8] refined further the DeepLab method by in- corporating global context features. Chen et al. [9] provided a way for handling scale by using attention models that pro- vides a mean to focus on the most relevant features. This attention model is able to learn a weight map, that weighs feature maps pixel-by-pixel from different scales. Eigen et al. [13] proposed a method to sequentially utilize multiple scales to refine the prediction of depth, surface normals, and semantic segmentation.

One of the commonly used models to incorporate con- text is conditional random field (CRF). Chen et al. [7] uti- lized the fully connected conditional random fields as a post processing. The unary potentials of the CRF are set to the probabilities from their convolutional network, while pair- wise potentials are gaussian kernels based on the spatial and color features. Lin et al [34] proposed a method to use pair- wise potentials based on convolutional neural networks fea- ture maps. In contrast to the previous work that uses condi- tional random fields as post processing refinement step, this work went further in integrating CNNs and CRFs. Zheng et al. [65] formulated the mean field CRF inference algorithm as a recurrent network. Thus, the proposed method enabled the end-to-end training of the model.

Another way to incorporate context is using recurrent neural networks (RNN) to capture the long range depen- dencies of various regions. Visin et al. [57] used a recurrent layer to sweep the image horizontally and vertically, which ensures the usage of contextual information for a bet- ter segmentation. One of the main bottlenecks in vanilla RNN is the vanishing gradients problem, gated recurrent architectures such as LSTMs [23] and GRUs [10] alleviate this problem. Byeon et al. [5] proposed a segmentation method that splits the image into non overlapping regions, then incorporates context using four separate LSTM blocks. Li et al. [32] proposed a method for context fusion using LSTMs. In their work both RGB and depth information were utilized, and the global context was modeled vertically on both, followed by the horizontal fusion. Another bottleneck in vanilla recurrent networks is that it could lead to the loss in spatial

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

relationships. Shuai et al. [46] utilized directed acyclic graph RNN to incorporate long range de- pendencies. This directed acyclic graph maintains spatial relations-ships unlike using chained RNNs. Finally, Qi et al [40] proposed hierarchically gated deep network, which is a multi-scale deep network that incorporates context at various scales. Multiple LSTM memory cells are used in the network between convolutional layers, to learn whether to incorporate spatial context from the lower layer into the higher one.

Figure 4: Atrous Convolution and Spatial Pyramid Pooling for Multi-scale support. Figure reproduced from [8, 60].

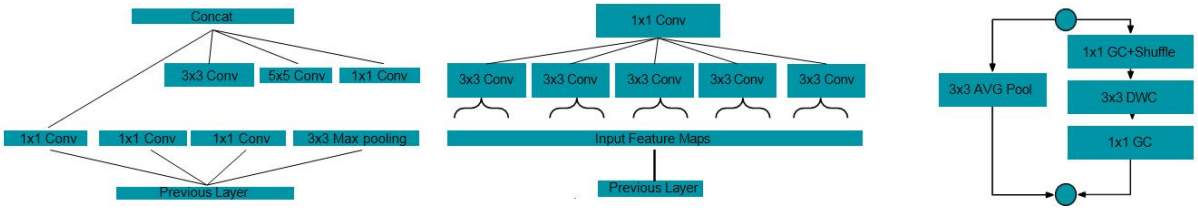
III. REAL-TIME CNNS

In recent years there has been an increasing need for running deep neural networks real-time on embedded plat-forms, in various applications. Two main categories in the work of efficient CNNs are discussed: (1) Efficient CNN models that introduce different layers and modules to im- prove its computational efficiency. (2) Model compression and pruning. Other approaches such as model quantization and hardware acceleration are out of the scope of this paper.

3.1. Efficient CNN Models

Convolutional layers are required to learn cross chan- nel and spatial correlations. This process can be performedin an efficient manner by separating both. Szegedy et al. [50, 51, 49] introduced the inception module and utilized it in Inception V1, V2 and further refined it in Inception V3 [51] and Inception-ResNet [49]. The main purpose of the inception module is to decouple the cross channel and spatial convolution. This separation is performed using 1x1 for the cross channel convolution that maps to 3 or 4 sep- arate spaces. This is followed by 3x3 and/or 5x5 convolu- tion for the spatial correlations. The extreme case of the inception module with one spatial convolution per channel is what is termed as depthwise separable convolution. Figure 5 shows the inception module [50], and depthwise separable convolution which is kind of an extreme case of inception. Howard et al. presented depth-wise separable convolutions as a mean to improve efficiency [24] in what is known as MobileNets. Zhang et al. developed a generalized form of separable convolution denoted as grouped convolution, while utilizing channel shuffle to ensure the input-output connectivity between different groups [62]. Figure 5 shows the shufflenet unit utilized in their model.

Huang et al. [25] proposed training a densely connected network with sparsified connections denoted as Con- denseNet. The connectivity pattern is implemented efficiently using grouped convolutions. This method is considered also as a network pruning method. Most of the research conducted in efficient convolutional networks is directed to- wards classification and detection. Little attention is given to the computational efficiency of deep neural networks for semantic segmentation. When it comes to applications such as autonomous driving this consideration is extremely important. Some studies such as the work by Paszke et al. [39] tried to address the issue of segmentation efficiency.



(a) Inception Module [50] (b) Depthwise separable Convolution [24] (c) ShuffleNet Unit [62]

Figure 5: Differences between Computationally Efficient Modules for Convolution. GC: Grouped Convolution. DWC: Depth-Wise Convolution

Sandler et al. [43] proposed inverted residual module with linear bottleneck. This takes low dimensional representa- tion as input then expands it to a higher dimensional space applies convolution then maps it back. The convolution op- eration is performed using the efficient depthwise separable convolutions. This work proposed an efficient segmentation method as well.

3.2 Model Compression and Pruning

There are two main pruning techniques for model com- pression namely weight pruning and filter pruning. Han et al. proposed DeepCompression framework [21] which learns both weights and connections in a three steps pro- cess. They make use of a regularization loss which pushes parameters towards zero and thus reducing the number of parameters of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

AlexNet by a factor of 9. Sparsity can lead to inefficient parallelism. To alleviate sparsity constraint, Han et al. [20] presented an efficient inference engine rely- ing on sparse matrix-vector multiplication with weight shar- ing. The resulting computation speed achieves x189 and x13 gain when compared to CPU and GPU implementa- tions of the same DNN without compression. Model com- pression also enables networks to fit in the on-chip SRAM which reduces energy consumption per memory read by a factor x120 compared from fetching weights from DRAM. Filter pruning is a similar approach like weight pruning.

While weight pruning results in sparse connectivity pattern, removing the entire filter and their associated feature maps preserve dense connectivity. Consequently computational cost reduction does not rely on sparse convolution libraries or dedicated hardware and existing efficient BLAS libraries for dense matrix multiplication can be further used. Wen et al. [58] proposed filter pruning using model structure learning and group lasso which is an efficient regularization to learn sparse structures. Their method is even more gen- eral than filter regularization since the Structured Sparsity Learning (SSL) method can regularize any structure (fil- ters, channels, filter shapes, and layer depth) of CNNs. This learning technique acts like a compression method to learn a smaller model from a larger one reducing the computational cost. Li et al. [31] presented another pruning approach which is not based on filter magnitude. The method relied on reinforcement Learning to train a pruning agent which made a set of binary actions to decide to remove or not each filter. It maximized a reward function which combined two terms, the accuracy term and the efficiency term. The ac- curacy term ensured the performance drop is bounded, and the efficiency term encouraged to prune more filters away.

IV. METHODOLOGY

In this section a detailed description of the benchmark- ing framework is presented. We implemented a generic framework through the decoupled encoder-decoder design. This allows the extensibility for more encoding and decoding methods. It also allows principled comparison between different design choices that can aid practitioners.

4.1. Meta Architectures

Three meta-architectures are integrated in our bench- marking software: (1) SkipNet meta-architecture [35]. (2) U-Net meta-architecture [41]. (3) Dilation Frontend meta- architecture [60]. The meta-architectures for semantic seg- mentation identify the decoding method for in the network upsampling. All of the network architectures share the same down-sampling factor of 32. The downsampling is achieved either by utilizing pooling layers, or strides in the convolu- tional layers. This ensures that different meta architectures have a unified down-sampling factor to assess the effect of the decoding method only.

SkipNet architecture denotes a similar architecture to FCN8s [35]. The main idea of the skip architecture is to benefit from feature maps from higher resolution to improve the output segmentation. SkipNet applies transposed convo- lution on heatmaps in the label space instead of performing it on the feature space. This entails a more computationally efficient decoding method than others. Feature extraction networks have the same downsampling factor of 32, so they follow the 8 stride version of skip architecture. Higher resolution feature maps are followed by 1x1 convolution to map from feature space to label space that produces heatmaps

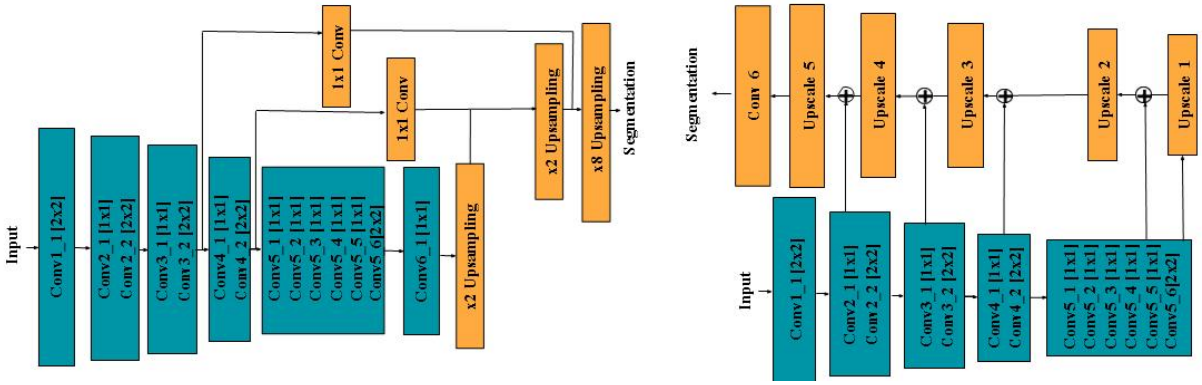


Figure 6: Different Meta Architectures using MobileNet as the feature extraction network. a) SkipNet architecture. b) UNet.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

corresponding to each class. The final heatmap with down- sampling factor of 32 is followed by transposed convolution with stride 2. Elementwise addition between this upsam- pled heatmaps and the higher resolution heatmaps is performed. Finally, the final output heat maps are followed by a transposed convolution for up-sampling with stride 8. Figure 6(a) shows the SkipNet architecture utilizing a Mo- bileMet encoder.

U-Net architecture denotes the method of decoding that up-samples features using transposed convolution corresponding to each downsampling stage [41]. The up- sampled features are fused with the corresponding features maps from the encoder with the same resolution. The stage- wise upsampling provides higher accuracy than one shot 8x upsampling. The current fusion method used in the frame- work is element-wise addition. Concatenation as a fusion method can provide better accuracy, as it enables the net- work to learn the weighted fusion of features. Nonetheless, it increases the computational cost, as it is directly affected by the number of channels. The upsampled features are then followed by 1x1 convolution to output the final pixel-wise classification. Figure 6(b) shows the UNet architecture using MobileNet as a feature extraction network.

Dilation Frontend architecture utilizes dilated convolu- tion [60] instead of downsampling the feature maps. Dilated convolution enables the network to maintain an adequate receptive field, but without degrading the resolution from pooling or strided convolution. However, a side-effect of this method is that computational cost increases, since the operations are performed on larger resolution feature maps. The encoder network is modified to incorporate a downsampling factor of 8 instead of 32. The decrease of the down- sampling is performed by either removing pooling layers or converting stride 2 convolution to stride 1. The pooling or strided convolutions are then replaced with two dilated convolutions [60] with dilation factor 2 and 4 respectively.

4.2. Feature Extraction Architectures

In order to achieve real-time performance multiple net- work architectures are integrated in the benchmarking framework. The framework includes four state of the art real-time network architectures for feature extraction. These are: (1) VGG16 [48]. (2) ResNet18 [22]. (3) MobileNet [24]. (4) ShuffleNet [62]. The reason for using VGG16 is to act as a baseline method to compare against as it was used in [35]. The other architectures have been used in real-time systems for detection and classification. ResNet18 incorporates the usage of residual blocks that di- rects the network toward learning the residual representa- tion on identity mapping.

MobileNet network architecture is based on depthwise separable convolution [24]. It is considered the extreme case of the inception module, where separate spatial con- volution for each channel is applied denoted as depthwise convolutions. Then 1x1 convolution is used and denoted as pointwise convolutions. The separation in depthwise and pointwise convolution improve the computational efficiency on one hand. On the other hand it improves the accuracy as the cross channel and spatial correlations mapping are learned separately.

ShuffleNet encoder is based on grouped convolution that is a generalization of depthwise separable convolution [62]. It uses channel shuffling to ensure the connectivity between input and output channels. This eliminates connectivity restrictions posed by the grouped convolutions.

V. EXPERIMENTS

In this section experimental setup, detailed ablation study and results in comparison to the state of the art are reported.

Table 1: Comparison of different encoders and decoders on Cityscapes validation set. GFLOPs are measured on image size 1024x512.

Encoder	Decoder	GFLOPs	mIoU	Road	Sidewalk	Building	Sign	Sky	Person	Car
SkipNet	MobileNet	13.8	61.3	95.9	73.6	86.9	57.6	91.2	66.4	89.0
SkipNet	ShuffleNet	4.63	55.5	94.8	68.6	83.9	50.5	88.6	60.8	86.5
UNet	ResNet18	43.9	57.9	95.8	73.2	85.8	57.5	91.0	66.0	88.6
UNet	MobileNet	55.9	61.0	95.2	71.3	86.8	60.9	92.8	68.1	88.8
UNet	ShuffleNet	17.9	57.0	95.1	69.5	83.7	54.3	89.0	61.7	87.8
Dilation	MobileNet	150	57.8	95.6	72.3	85.9	57.0	91.4	64.9	87.8
Dilation	ShuffleNet	71.6	53.9	95.2	68.5	84.1	57.3	90.3	62.9	86.6

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Table 2: Comparison of different encoders and decoders on Cityscapes validation set with Coarse annotations pre-training then using fine annotations.

Encoder	Decoder	mIoU	Road	Sidewalk	Building	Sign	Sky	Person	Car
SkipNet	MobileNet	62.4	95.4	73.9	86.6	57.4	91.1	65.7	88.4
SkipNet	ShuffleNet	59.3	94.6	70.5	85.5	54.9	90.8	60.2	87.5

5.1 Experimental Setup

Through all of our experiments, weighted cross entropy loss from [39] is used, to overcome the class imbalance. The class weight is computed as wclass Adam optimizer [29] learning rate is set to 1e−4. Batch normaliza- tion [26] after all convolutional or transposed convolution layers is incorporated. L2 regularization with weight decay rate of 5e−4 is utilized to avoid over-fitting. The feature ex- tractor part of the network is initialized with the pre-trained corresponding encoder trained on Imagenet. A width multi- plier of 1 for MobileNet to include all the feature channels is performed through all experiments. The number of groups used in ShuffleNet is 3. Based on previous [62] results on classification and detection three groups provided adequate accuracy. Results are reported on Cityscapes dataset [12] which contains 5000 images with fine annotation, with 20 classes including the ignored class. Another section of the dataset contains coarse annotations with 20,000 labeled images. These are used in the case of Coarse pre-training that proved to improve the results of the segmentation. Experiments are conducted on images with resolution of 512x1024.

5.2 Ablation Study

Semantic segmentation is evaluated using mean intersec- tion over union (mIoU), per-class IoU, and per-category IoU. Table 1 shows the results for the ablation study on different encoders-decoders with mIoU and GFLOPs to demonstrate the accuracy and computations trade-off. The main insight gained from our experiments is that, UNet de- coding method provides more accurate segmentation results than Dilation Frontend. This is mainly due to the transposed convolution by x8 in the end of the Dilation Frontend, unlike the UNet stage-wise upsampling method. The Skip- Net architecture provides on par results with UNet decoding method. In some architectures such as SkipNet-ShuffleNet it is less accurate than UNet counter part by 1.5%. the network provide the best in terms of accuracy. How- ever, SkipNet architecture is more computationally efficient with x4 reduction in GFLOPs. This is explained by the fact that transposed convolutions in UNet are applied in the fea- ture space unlike in SkipNet that are applied in label space. Table 2 shows that pre-training with cityscapes coarse an- notation, then finetuning on the fine annotation improves the segmentation in terms of mIoU with 1-4%. The un- derrepresented classes are the ones that often benefit from pre-training.

5.3. Embedded Vision Experiments

Experimental results on the cityscapes test set are shown in Table 3. ENet [39] is compared to SkipNet-ShuffleNet and SkipNet-MobileNet in terms of accuracy and computa- tional cost. SkipNet-ShuffleNet outperforms ENet in terms of GFLOPs, yet it maintains on par mIoU. Both SkipNet- ShuffleNet and SkipNet-MobileNet outperform SegNet [1] in terms of computational cost and accuracy with reduction up to x143 in GFLOPs. SkipNet-ShuffleNet was deployed on a Jetson TX2 that delivered real-time performance in 15 frames per second on image resolution 640x360. Fig- ure 8 shows the comparison between different image reso- lution versus frame-rate and running time in milliseconds. These were measured on the Jetson TX2 for the SkipNet- ShuffleNet architecture. Figure 7 shows qualitative results for different encoders including MobileNet, ShuffleNet and ResNet18 segmentation results than the later two.

Table 3: Comparison to the state of the art segmentation networks on Cityscapes test set. GFLOPs is computed on image resolution 640x360.

Model	GFLOPs	Class IoU	Class iIoU	Category IoU	Category iIoU
SegNet[1]	286.03	56.1	34.2	79.8	66.4
ENet[39]	3.83	58.3	24.4	80.4	64.0
SkipNet-VGG16[35]	445.9	65.3	41.7	85.7	70.1
SkipNet-ShuffleNet	2.0	58.3	32.4	80.2	62.2
SkipNet-MobileNet	6.2	61.5	35.2	82.0	63.0

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Figure 7: Qualitative Results on CityScapes. (a) Original Image. (b) SkipNet-MobileNet pretrained with Coarse Annotations. UNet-Resnet18. (d) SkipNet-ShuffleNet pretrained with Coarse Annotations

Figure 8: Running Time in milliseconds and Frames per second versus the different image resolution. Measured on Jetson TX2

VI. CONCLUSION

In this paper, we present the first principled approach for benchmarking real-time segmentation networks. The decoupled design of the framework separates encoder and de- coder modules and allows for systematic comparison. The first module is comprised of the feature extraction network architecture and the second module is the meta-architecture that provides the decoding method. This generic meta- architecture allows for extensibility further on to other en- coders and decoding methods. Detailed analysis of different image resolutions versus frame-rate on Jetson TX2 is presented. Our benchmarking framework provides researchers and practitioners a mechanism to systematically evaluate new encoders and decoders. New computationally efficient models for segmentation emerged that outperform the state of the art in terms of GFLOPs, while maintaining on par accuracy. It enabled one of the models to run real-time at ∼16 fps on a Jetson TX2. Future work is to mathematically formalize the meta-architecture to enable automated topology exploration using meta-learning.

REFERENCES

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla. Seg- net: A deep convolutional encoder-decoder archi- tecture for image segmentation. arXiv preprint arXiv:1511.00561, 2015.
[2] S. Bell, P. Upchurch, N. Snavely, and K. Bala. Mate- rial recognition in the wild with the materials in con- text database. In Computer Vision and Pattern Recog- nition (CVPR). IEEE, 2015.
[3] T. M. Bonanni, A. Pennisi, D. Bloisi, L. Iocchi, and D. Nardi. Human-robot collaboration for semantic la- beling of

the environment. In Proceedings of the 3rd Workshop on Semantic Perception, Mapping and Ex- ploration, 2013.
[4] G. J. Brostow, J. Fauqueur, and R. Cipolla. Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 30(2):88–97, 2009.
[5] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki. Scene labeling with lstm recurrent neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3547–3555, 2015.
[6] J. Carreira and A. Zisserman. Quo vadis, action recog- nition? a new model and the kinetics dataset. In 2017 IEEE Conference on Computer Vision and Pat- tern Recognition (CVPR), pages 4724–4733. IEEE, 2017.
[7] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image seg- mentation with deep convolutional nets, atrous con- volution, and fully connected crfs. arXiv preprint arXiv:1606.00915, 2016.
[8] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.
[9] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille. Attention to scale: Scale-aware semantic image seg- mentation. arXiv preprint arXiv:1511.03339, 2015.
[10] K. Cho, B. Van Merrie¨nboer, D. Bahdanau, and
Y. Bengio. On the properties of neural machine trans- lation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.
[11] O¨ . C¸ ic¸ek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense vol- umetric segmentation from sparse annotation. In In- ternational Conference on Medical Image Computing and Computer-Assisted Intervention, pages 424–432. Springer, 2016.
[12] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. En- zweiler, R. Benenson, U. Franke, S. Roth, and
B. Schiele. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Con- ference on Computer Vision and Pattern Recognition, pages 3213–3223, 2016.
[13] D. Eigen and R. Fergus. Predicting depth, surface nor- mals and semantic labels with a common multi-scale convolutional architecture. In Proceedings of the IEEE International Conference on Computer Vision, pages 2650–2658, 2015.
[14] C. Farabet, C. Couprie, L. Najman, and Y. Le- Cun. Learning hierarchical features for scene label- ing. IEEE transactions on pattern analysis and ma- chine intelligence, 35(8):1915–1929, 2013.
[15] C. Farabet, N. EDU, C. Couprie, L. Najman, and Y. LeCun. Scene parsing with multiscale feature learn- ing, purity trees, and optimal covers.
[16] M. Fayyaz, M. H. Saffar, M. Sabokrou, M. Fathy, and R. Klette. STFCN: spatio-temporal FCN for semantic video segmentation. CoRR, abs/1608.05971, 2016.
[17] R. Gadde, V. Jampani, and P. V. Gehler. Semantic video cnns through representation warping. CoRR, abs/1708.03088, 2017.
[18] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A re- view on deep learning techniques applied to seman- tic segmentation. arXiv preprint arXiv:1704.06857, 2017.
[19] D. Grangier, L. Bottou, and R. Collobert. Deep con- volutional networks for scene parsing. In ICML 2009 Deep Learning Workshop, volume 3. Citeseer, 2009.
[20] S. Han, X. Liu, H. Mao, J. Pu, A. Pedram, M. A. Horowitz, and W. J. Dally. Eie: efficient inference en- gine on compressed deep neural network. In Proceed- ings of the 43rd International Symposium on Com- puter Architecture, pages 243–254. IEEE Press, 2016.
[21] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In Advances in Neural Information Processing Sys- tems (NIPS), pages 1135–1143, 2015.
[22] K. He, X. Zhang, S. Ren, and J. Sun. Deep resid- ual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
[23] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.
[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.
[25] G. Huang, S. Liu, L. van der Maaten, and K. Q. Weinberger. Condensenet: An efficient densenet using learned group convolutions. arXiv preprint arXiv:1711.09224, 2017.

[26] S. Ioffe and C. Szegedy. Batch normalization: Accel- erating deep network training by reducing internal co- variate shift. In International Conference on Machine Learning, pages 448–456, 2015.
[27] S. D. Jain, B. Xiong, and K. Grauman. Fusionseg: Learning to combine motion and appearance for fully automatic segmention of generic objects in videos. arXiv preprint arXiv:1701.05384, 2(3):6, 2017.
[28] A. Kendall, V. Badrinarayanan, and R. Cipolla. Bayesian segnet: Model uncertainty in deep convolu- tional encoder-decoder architectures for scene under- standing. arXiv preprint arXiv:1511.02680, 2015.
[29] D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[30] A. Kundu, Y. Li, F. Dellaert, F. Li, and J. M. Rehg. Joint semantic segmentation and 3d reconstruction from monocular video. In European Conference on Computer Vision, pages 703–718. Springer, 2014.
[31] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf. Pruning filters for efficient convnets. arXiv preprint arXiv:1608.08710, 2016.
[32] Z. Li, Y. Gan, X. Liang, Y. Yu, H. Cheng, and L. Lin. Lstm-cf: Unifying context modeling and fusion with lstms for rgb-d scene labeling. In European Confer- ence on Computer Vision, pages 541–557. Springer, 2016.
[33] G. Lin, C. Shen, A. v. d. Hengel, and I. Reid. Explor- ing context with deep structured models for seman- tic segmentation. arXiv preprint arXiv:1603.03183, 2016.
[34] G. Lin, C. Shen, I. Reid, et al. Efficient piecewise training of deep structured models for semantic seg- mentation. arXiv preprint arXiv:1504.01013, 2015.
[35] J. Long, E. Shelhamer, and T. Darrell. Fully convo- lutional networks for semantic segmentation. In Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3431–3440, 2015.
[36] O. Miksik, V. Vineet, M. Lidegaard, R. Prasaath, M. Nießner, S. Golodetz, S. L. Hicks, P. Pe´rez, S. Izadi, and P. H. Torr. The semantic paintbrush: Interactive 3d mapping and recognition in large out- door spaces. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 3317–3326. ACM, 2015.
[37] D. Nilsson and C. Sminchisescu. Semantic video seg- mentation by gated recurrent flow propagation. arXiv preprint arXiv:1612.08871, 2016.
[38] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vi- sion, pages 1520–1528, 2015.
[39] A. Paszke, A. Chaurasia, S. Kim, and E. Culur- ciello. Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147, 2016.
[40] G.-J. Qi. Hierarchically gated deep networks for se- mantic segmentation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
[41] O. Ronneberger, P. Fischer, and T. Brox. U-net: Con- volutional networks for biomedical image segmenta- tion. In International Conference on Medical Im- age Computing and Computer-Assisted Intervention, pages 234–241. Springer, 2015.
[42] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of ur- ban scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3234–3243, 2016.
[43] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and seg- mentation. arXiv preprint arXiv:1801.04381, 2018.
[44] E. Shelhamer, K. Rakelly, J. Hoffman, and T. Dar- rell. Clockwork convnets for video semantic segmen- tation. In Computer Vision–ECCV 2016 Workshops, pages 852–868. Springer, 2016.
[45] E. Shelhamer, K. Rakelly, J. Hoffman, and T. Darrell. Clockwork convnets for video semantic segmentation. CoRR, abs/1608.03609, 2016.
[46] B. Shuai, Z. Zuo, B. Wang, and G. Wang. Dag- recurrent neural networks for scene labeling. In Pro- ceedings of the IEEE conference on computer vision and pattern recognition, pages 3620–3629, 2016.
[47] M. Siam, S. Valipour, M. Jagersand, and N. Ray. Con- volutional gated recurrent networks for video segmen- tation. arXiv preprint arXiv:1611.05435, 2016.
[48] K. Simonyan and A. Zisserman. Very deep convo- lutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[49] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of resid- ual

connections on learning. In AAAI, volume 4, page 12, 2017.
[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al. Going deeper with convolutions. Cvpr, 2015.
[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Con- ference on Computer Vision and Pattern Recognition, pages 2818–2826, 2016.
[52] P. Tokmakov, K. Alahari, and C. Schmid. Learning motion patterns in videos. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 531–539. IEEE, 2017.
[53] P. Tokmakov, K. Alahari, and C. Schmid. Learning video object segmentation with visual memory. arXiv preprint arXiv:1704.05737, 2017.
[54] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Deep end2end voxel2voxel prediction. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on, pages 402–409. IEEE, 2016.
[55] A. Valada, G. L. Oliveira, T. Brox, and W. Burgard. Deep multispectral semantic scene understanding of forested environments using multimodal fusion. In The 2016 International Symposium on Experimental Robotics (ISER 2016), 2016.
[56] V. Vineet, O. Miksik, M. Lidegaard, M. Nießner, S. Golodetz, V. A. Prisacariu, O. Ka¨hler, D. W. Mur- ray, S. Izadi, P. Perez, and P. H. S. Torr. Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction. In IEEE International Confer- ence on Robotics and Automation (ICRA), 2015.
[57] 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 seman- tic segmentation. In Proceedings of the IEEE Con- ference on Computer Vision and Pattern Recognition Workshops, pages 41–48, 2016.
[58] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li. Learn- ing structured sparsity in deep neural networks. In Advances in Neural Information Processing Systems, pages 2074–2082, 2016.
[59] Z. Wu, C. Shen, and A. v. d. Hengel. Wider or deeper: Revisiting the resnet model for visual recog- nition. arXiv preprint arXiv:1611.10080, 2016.
[60] F. Yu and V. Koltun. Multi-scale context ag- gregation by dilated convolutions. arXiv preprint arXiv:1511.07122, 2015.
[61] H. Zhang, A. Geiger, and R. Urtasun. Understanding high-level semantics by modeling traffic patterns. In Proceedings of the IEEE International Conference on Computer Vision, pages 3056–3063, 2013.
[62] X. Zhang, X. Zhou, M. Lin, and J. Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. arXiv preprint arXiv:1707.01083, 2017.
[63] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia. Icnet for real-time semantic segmentation on high-resolution images. arXiv preprint arXiv:1704.08545, 2017.
[64] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 2881– 2890, 2017.
[65] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vi- neet, Z. Su, D. Du, C. Huang, and P. H. Torr. Condi- tional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision, pages 1529–1537, 2015.
[66] W. Zhu and X. Xie. Adversarial deep structural net- works for mammographic mass segmentation. arXiv preprint arXiv:1612.05970, 2016.

Smart Bins contribute to a cleaner, safer, and more sanitary environment, as well as improved operating efficiency, while

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

lowering management costs, resources, and roadside radiation. Campuses, theme parks, airports, railway stations, and shopping malls are all good candidates for the Smart Bin. Smart bins are a trash management device that is clever. They include wireless ultrasonic fill-level sensors built within that detect how full the bin is and send the data to a cloud-based monitoring and analytics platform via the Internet of Things. A dustbin is a garbage receptacle constructed of metal, plastic, or any other hard-to-store waste material that is used for temporarily keeping trash. They store energy in a variety of renewable and non-renewable materials and contribute to environmental protection clean. When the trashcan receives the signal, it automatically opens and closes its hatch. A level measuring ultrasonic sensor is also included in the dustbin, which continuously measures the amount of waste in the bin and automatically identifies when it is about to fill up. Smart waste management is characterized by the application of technology to improve waste management efficiency. This not only allows trash collectors to design more effective routes for emptying the bins, but it also reduces the likelihood of any bin remaining full for more than a week. As the world around us transforms and becomes more linked and digitalized, public transportation is expected to provide in every way. How far are we from developing a complete Smart Rail system, from cutting-edge technology to comfort, safety, reliability, and sustainability? We remark that cleanliness of the surrounding space in the train is also vital. We don't know how much scrap is collected in the dustbins because so many people go by train, and we don't utilize enough manpower for each train, so we apply this system and establish a green and clean city. The smart railway dustbin locator is primarily focused on human health issues, as various variant viruses, such as Covid-19, are having an impact on human bodies and reducing human life spans. As a result, human power will be reduced by using this system, which will automatically notify you on your screen where dustbins need to be replaced and where ordnance is detected

Shikha Parashar and colleagues [1] Waste management, from collection to dumping and interruption, has become one of the most difficult and time-consuming tasks for municipal organizations all over the world. A new concept of Smart Dustbin has been considered for Smart buildings, hospitals, schools, and train stations to make this arduous choreeasier. The Smart Garbage Collector concept is an evolution of the standard garbage collector that incorporates sensors and some type of logic to make it smart. This smart collector is a ground-breaking concept that uses a line-following garbage vehicle and a pole-mounted rubbish component to follow a pre-planned railway pattern. The stationary bin uses ultrasonic sensors to detect garbage levels and uses an RF module to communicate the bin's current level to the garbage car. As a result, this is a fully automated system that contributes to the Clean India, Green India theme.

According to M.Ashwin et al. [2] trash management is a big issue all over the world. The clever is a cutting-edge automated gadget that collects old absorbent care products separately. In smart cities, the Automatic Intelligence Smart Bin assists in resolving trash management issues. In smart cities, the smart bin will play a vital role. The smart lifestyle starts with a clean environment in the city, which starts with a smart bin. The HC-SR04 Ultrasonic Sensor is used to recognize humans, and the TowerPro SG90 Servo Motor is used to automatically open and close the smart bin lid. The smart bin's overall functioning was controlled by an Arduino Uno microcontroller. The smart bin uses an IoT sensor to differentiate dry and wet waste collection. When the smart bin is 80 percent full, the microcontroller automatically sends a warning message to the garbage collector.

M. Dhaifallah and colleagues [3] on a micro grid model of a hospital, hybrid optimization of diverse energy resources was undertaken to evaluate the capability of a standalone energy system and simultaneous waste mitigation. The study'smain goals were to gather renewable energy resource data for a hybrid hospital, use the average amount of hospital waste from the literature and NASA surface meteorology, as well as the solar energy database from HOMER Pro software, to build a hybrid model for a conceptual hospital in Saudi Arabia's new green city, NEOM. Biogas cofire and diesel generators, as well as a PV solar array and batteries, made up the hybrid model. The load requirements of a freestanding hospital were analyzed using simulations. Tesla batteries were used to design the energy storage system.

One of the most significant applications of the Internet of Things (IoT) in this digital era, according to Jacob John et al. [4], is the development of smart cities. Smart objects (devices) are connected to each other via the internet as a backbone in IoT-based smart cities. Using multi hop connectivity, the smart objects' detected data is sent to the sink for further processing. Smart cities use studied data to improve their infrastructure, public utilities, and services by utilizing IoT technology for the betterment of the general population's well-being. Waste collection is a major issue for governments that want to establish a clean environment in IoT-based smart cities. As the population of metropolitan areas grows, so does the amount of waste produced.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

D. Krishnakumar and colleagues [5] On India, solid waste management has been a big issue in railway waggons for decades. It is common knowledge that if solid waste is not adequately managed, it will have a significant negative impact on the environment. The goal of this project is to collect solid garbage that passengers throw out the window and recycle or reuse it. Between two parallel window frames, two conveyor belts are placed one above the other. Four collection tanks are also installed, one pair at each end of one conveyor belt. A level sensor (ultrasonic sensor) and a GSM module are installed in the tanks. The garbage is directed onto the conveyor via two exit pipes connected from the interior. Each coop is equipped with a switch that, when pressed, activates the conveyor belt, preventing it from running indefinitely without any work to be done. The entire system, which consists of two sets of conveyor belts, is housed in a casing between the window's parallel frames.

According to Dr.T.M.N.Vamsi et al. [6,] appropriate waste management through the use of technology is a serious worry of the hour. Monitoring and disposal of waste are currently done by people, which is inconvenient; however, by augmenting the traditional system with the flavour of IoT, monitoring of garbage bins will be simple. This benefits those who work in the traditional garbage collection system, and it is the most practical solution. The SGMDSS is a cutting-edge information management control system that aids metros, cities, and villages in maintaining hygiene and cleanliness through improved waste disposal. This technology employs a cutting-edge approach that automates garbage monitoring and disposal. SGMDSS monitors garbage bins at various locations and notifies cleaning employees of the level of waste gathered in the bins via an android mobile application for disposal, as well as providing the shortest path to the garbage bin site that is almost full. This data is also transferred to a webpage, and the full database is saved and retrieved via the cloud. In addition, the worker receives an alarm message.

According to Ashi Goel et al. [7], the centrality of a town is determined by the quality of its air, the cleanliness of its roads and highways, and its overall atmosphere. One of the most pressing concerns as we work toward our wonderful vision 2020 aim of becoming a developed and affluent nation is hygiene. Our mission is 'Swachh Bharat Abhiyan,' hence we invented a 'Smart Toilet' and a 'Smart Dustbin.' Those who live in the city must be forced to suffer from a variety of ailments if cleanliness is not maintained. Suddenly, a plethora of new diseases appeared. There are numerous rubbish bins available, as well as many public toilets being built by the government, but most people have no knowledge where they are or where they are located when they walk out in new places. The method's architecture collects this information and transmits it via a wireless network. This paper may be useful in encouraging the majority to support the Clean India effort. It will be able to demonstrate the emerging role in the Clean India scheme in the future. Sensors are used to detect the level of rubbish in the dustbins and public bathrooms, and the information is transferred to the official mobile station via GPS module.

Mudike Koushal Yadav and colleagues [8] Modern garbage management practices, from collection to dumping and disruption, have become a difficult and time-consuming task for municipal organizations all over the world. A novel concept of Smart Garbage Collection System has been considered for Smart buildings, hospitals, schools, and railway Stations to make this tedious work easier. The Smart Garbage Collector concept is a modernization of the traditional garbage collector that incorporates sensors and electronics to make it smart. This smart garbage collector is a ground-breaking concept that uses a line-following garbage vehicle and a pole-mounted rubbish component to follow a predetermined train path. The fixed bin uses ultrasonic sensors to indicate waste level and uses an RF Module to update the volume level of the bin to the garbage car. As a result, this gadget is a completely automated system that contributes significantly to the Clean India, Green India initiative.

According to Murali Krishna Thirumalakonda et al. [9], waste collection in public spaces and communities is a big problem today. Unsanitary conditions cause a variety of diseases and harm to the environment. This may be avoided by placing a smart dustbin in the vicinity. This smart garbage management system is an advance over a regular bin, and it uses ultrasonic sensor systems to check the garbage level above the dustbin. An ultrasonic sensor is a device that measures the distance between an object and the user. The bin may be opened and closed automatically using a sensor. When the bin is full, the buzzer will sound. Then, when the trashcan is full, it will transmit an alert message through GSM module, which is attached to the circuit. As a result, the system is useful in waste management when it is monitored and informed on a regular basis. This results in a cleaner city for a higher quality of life. The smartphone app is developed as a graphical representation of daily updates to ensure a greener environment and support for Swachh Bharat for cleanliness. It is improved by the use of two bins, one for wet garbage and the other for dry waste. Wet trash decomposes

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

quickly to produce biogas, which can be used in the home.

According to Rishabh Kumar Singhvi et al. [10], a smart system that monitors the trashcan and provides real- time status is required for the development of smart cities. Municipal corporations in India do not yet have access to real-time information about trash cans. In order to address this issue, we are building an Internet of Things (IoT)-based system that can send a notification to a company about the overflow and toxicity level of the dustbins. A website is also being created to keep track of the data related to the trash cans. The dustbin status is updated on the website and a message is sent to the mobile phone through GSM module. Citizens can also file complaints about trash cans or waste management on this website. Arduino is utilized as a microcontroller in the recommended system to interface between the GSM/GPRS module and the sensors. Dustbin level and toxicity are measured using an ultrasonic sensor and a gas sensor, respectively.

III. PROBLEM STATEMENT

Railway workers, like humans, are on the front lines of the fight against contagious diseases. They are regularly in contact with diseased people and also with health in the surrounding railway boogie.

IV. PROPOSED SYSTEM

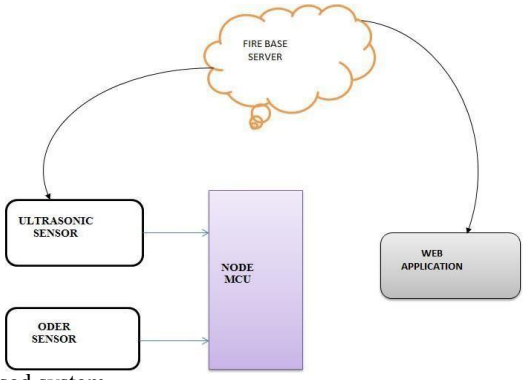


Figure 4.1: Architecture of proposed system

4.1 Working of Proposed System

A. Ultrasonic Sensor

Ultrasonic sensors work by emitting a sound wave that is above the human hearing range. The sensor's transducer functions as a microphone, receiving and transmitting ultrasonic sound. To deliver a pulse and receive the echo, our ultrasonic sensors, like many others, use a single transducer. The sensor measures the time between sending and receiving an ultrasonic pulse to determine the distance to a target.

Figure 4.2: Transmit and receive signal

1. Ultrasonic sensors send sound waves at a target and measure the time it takes for the reflected waves to return to the receiver to determine its distance.
2. This sensor is an electronic device that uses ultrasonic sound waves to detect the distance to a target and then converts the reflected sound into an electrical signal.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

B. Oder Sensor

Fig 4.3 Oder Sensor

When molecules of any chemical element are placed on the surface of a sensor, the e-nose operates. When a sensor is exposed to scents, the change in resistance is detected. The outcome is a pattern that is unique to that element. Our technology is built around an Oder sensor that detects where Oder is present in the trash can. To keep the railway bogie's environment clean. Aside from that, there is a stench there, and there isn't a lot of scrap. As a result, we'll need another sensor to detect odor.

C. NODE MCU

The ESP-12E module on the NodeMCU ESP8266 development board contains an ESP8266 chip with a Tensilica Xtensa 32-bit LX106 RISC microprocessor. This microprocessor runs on a configurable clock frequency of 80MHz to 160MHz and supports RTOS. To store data and programmer, the NodeMCU contains 128 KB of RAM and 4MB of Flash memory. It is perfect for IoT projects due to its high processing power, built-in Wi-Fi / Bluetooth, and Deep Sleep Operating capabilities. A Micro USB jack and VIN pin can be used to power NodeMCU (External Supply Pin). It has interfaces for UART, SPI, and I2C. The NODE MCU is the system's heart in our design. This is used to control all operations by connecting two sensors, one Ultrasonic sensor and the other an Oder sensor

Figure 4.4: Node MCU

4.2 Server / Web Application

In this system, data can be predicted by ultrasonic sensor and order sensor, which is nothing more than an ultrasonic sensor that detects the level of scrap in the dustbin and whether or not there is a smell present. This data is sent to the server, where we register/ login with our login id and password and then open one sheet with content ID, level, order, and comment, which is nothing more than whether or not the dustbin needs to be replaced. And the above technology provides excellent information regarding dustbin levels, resulting in a reduction in human manpower and a significant reduction in other infections infecting humans.

Login ID	Level	Oder	Comment
12546	5inch	Present	Need to replace
25345	8inch	No	Need to replace

Login /Register Output:

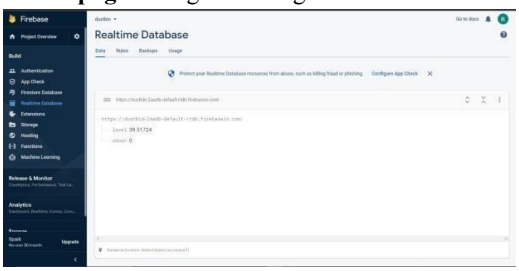
The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

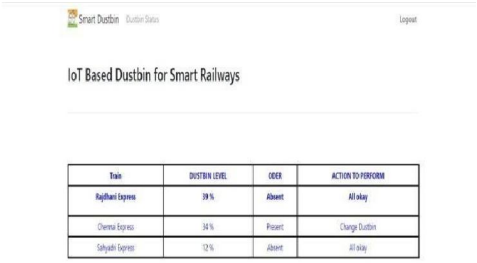
V. RESULT AND DISCUSSION

The below result page shows the construction of a login and registration page, which has we give a signup and login page with a username and password to present the real-time database of design project results.

Result page 1. Login and Registration



Result Page 2. Fetching Hardware components to take real time database.



Result page -3 Experimental result of smart Railway dustbin.

Above result page which will show Real time database from server to this page project is designed based on hardware module which is mechanically connected with dustbin and Node MCU ESP8266 is Wi-Fi compatible is heart of system which takes real time analog value to send server and design and developed web application to show all the notifications of the dustbin related where is level, Oder is present or not, and what will be action to take.

Above result page shows real-time experimental results of smart railway dustbin, which clearly displays train position such as rajdhani express to display dustbin level 39 percent and order is absent, action performance is all right. At the Chennai express location, the second row displays 34% of the dustbin level and an order is present, therefore action will be taken to change the trash.

VI. CONCLUSION AND FUTURE WORK

Smart Bins contribute to a cleaner, safer, and more sanitary environment, as well as improved operating efficiency, while lowering management costs, resources, and roadside emissions. The Smart Bin is suitable for high- traffic areas. The smart bin transmits information about fill levels and guarantees that the bin is only collected when it is completely full. As a result, the streets will be cleaner and safer. Because of the pattern of the presence of keen innovation, the advancement of clever frameworks, notably in the improvement of clever dustbins, would in general increase. The experimental results show that the smaller-than-normal and super-brilliant dustbins framework works and performs as expected. The results of the evaluation of the application for savvy dustbins revealed that the presence of shrewd dustbins spread throughout the room unequivocally consented to provide benefits and drew in extremely exorbitant interest in familiarity with discarding waste perfectly positioned, notwithstanding, it is important to further develop the brilliant dustbin framework to further develop its presentation just as it is important to further develop the brilliant dustbin

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

framework to further develop its presentation just as it is important to further develop the brilliant dustbin

FUTURE WORK

Using photoelectric, methane, and smell sensors, separate garbage (wet and dry) based on moisture content. To sort garbage into hazardous and non-hazardous categories. To dispose of the garbage created in an efficient manner.

REFERENCES

[1]. Shikha Parashar, Pankaj Tomar "Waste Management by a Robot- A Smart and Autonomous Technique" 2018, DOI: 10.9790/2834- 1303023136.

[2]. Dhaifallah M.Alotaibi Mohammad AkramiMahdieh DibajAkbar A.Javadi "Smart energy solution for an optimised sustainable hospital in the green city of NEOM" October 2019.

[3]. Chunsheng Zhu; Huan Zhou; Victor C. M. Leung; Kun Wang; Yan Zhang; Laurence T. Yang "Toward Big Data in Green City 'November 2017

[4]. Jacob John "'Smart Prediction and Monitoring of Waste Disposal System Using IoT and Cloud for IoT Based Smart Cities" 08 august 2021

[5]. D. Krishnakumar,Soumik Chakraborty,Amit R. Yadav,Aditya Jaideep,Surya Parashar "Solid waste management in railway wagon" 2019, ; International Journal of Advance Research, Ideas and Innovations in Technology.

[6]. Dr.T.M.N.Vamsi#1, Mr.G.Kalyan Chakravarthi "An IoT Based Smart Garbage Monitoring and Disposal Support System" 2021.

[7]. Ashi Goel, Esha Bansal, Tripti Gupta, "Smart City Hygiene Management using Android Application" International Journal of Progressive Research in Science and Engineering,Volume-1, Issue-3, June-2020

[8]. Mudike Koushal Yadav1, Rohith Mutyala2, Surgu Rahul Goud3 "SMART BOT WITH AUTOMATIC GARBAGE COLLECTING SYSTEM" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 07 Issue: 05 | May 2020.

[9]. Murali Krishna Thirumalakonda, K. Khaja Baseer, V. C. Praveena,V. Varsha,Abbas Ali Poralla "Smart Garbage Monitoring System using IoT" 2020.

[10]. Rishabh Kumar Singhvi, Roshan Lal Lohar, Ashok Kumar, Ranjeet Sharma, Lakhan Dev Sharma, Ritesh Kumar Saraswat." IoT Based Smart Waste Management System: India prospective" 2019.