Tech Science Press

Realtime Object Detection Through M-ResNet in Video Surveillance System

S. Prabu[1,*] and J. M. Gnanasekar[2]

[1]Department of Computer Science & Engineering, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu, Tamilnadu, 603308, India
[2]Department of Computer Science & Engineering, Sri Venkateshwara College of Engineering & Technology, Chennai, 602001, Tamilnadu, India
*Corresponding Author: S. Prabu. Email: sprabumkm@gmail.com

Abstract: Object detection plays a vital role in the video surveillance systems. To enhance security, surveillance cameras are now installed in public areas such as traffic signals, roadways, retail malls, train stations, and banks. However, monitoring the video continually at a quicker pace is a challenging job. As a consequence, security cameras are useless and need human monitoring. The primary difficulty with video surveillance is identifying abnormalities such as thefts, accidents, crimes, or other unlawful actions. The anomalous action does not occur at a higher rate than usual occurrences. To detect the object in a video, first we analyze the images pixel by pixel. In digital image processing, segmentation is the process of segregating the individual image parts into pixels. The performance of segmentation is affected by irregular illumination and/or low illumination. These factors highly affect the real-time object detection process in the video surveillance system. In this paper, a modified ResNet model (M-Resnet) is proposed to enhance the image which is affected by insufficient light. Experimental results provide the comparison of existing method output and modification architecture of the ResNet model shows the considerable amount improvement in detection objects in the video stream. The proposed model shows better results in the metrics like precision, recall, pixel accuracy, etc., and finds a reasonable improvement in the object detection.

Keywords: Object detection; ResNet; video survilence; image processing; object quality

1 Introduction

Image processing in a low light environment is a challenging task, particularly in video surveillance systems. Due to the lack of light conditions the quality of the image is automatically downgraded. A lot of image information is deformed which affects the image processing applications like object detection, object tracking, segmentation, etc. In video surveillance systems, object detection is a major part due to the automatic identification of anomaly events. Detecting objects from a video stream needs more calculation due to a large amount of data and grabbed of image from the video sequence must be in high

quality. In a real-world environment, the lighting condition is uneven particularly, at night time. Even though the video camera quality is high, the extracted images from the video stream may be low quality of images. In image data processing, image segmentation is the key step in analyzing image data [1]. Dividing the image into different parts of semantic regions is the main task of image segmentation. Low-quality images, particularly taken from the captured videos at night time, will affect the accuracy of the image understanding. To enhance the visualization, different kinds of methods can be used. The non-uniform illumination prior model [2] is proposed for identifying the illumination parts in segmented scenes. Convolution neural networks, it is a mathematical models that consists of a large number of processing units, working concurrently on multiple sets of data [3]. Convolutional Neural Network (CNN) has the number of layers which automatically detects the important features of image data without any human supervision [4]. The more number of layers leads to more accuracy of the output. When increasing deeper networks, the performance increases by a particular amount, then it degrades rapidly [5]. In other words, adding more layers leads to more training errors.

Fig. 1 depicts, 56 layers of CNN has the more error rate on both testing and training dataset, than 20 layer CNN architecture. ResNet (Residual Network) is one kind of deep learning model [6], which is made up of residual blocks. In which there is a direct connection that skips some layers of the model. Residual Block is a set of layers, in which the output of the layer is added with another layer in the block [7]. That by pass connection is called skip connection or shortcut connection shown in Fig. 2. These set of residual blocks combined together formed residual networks.
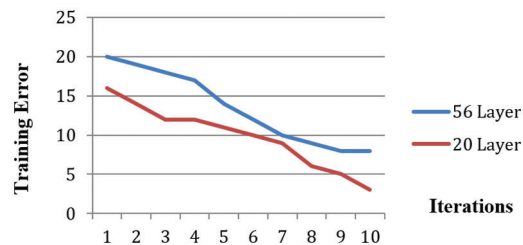

Figure 1: Training error in 56 and 20 layer deeper network
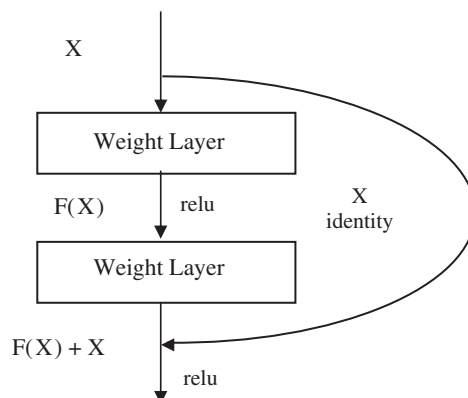

Figure 2: Basic building block of residual learning

The reason for using the residual networks are training phase becomes less complex compared to normal deeper neural networks and adding to more layers leads more training errors. Residual networks (ResNets) are used in the deeper networks to reduce training errors. ResNet is an effective way of identifying the

features in the salient area [8]. Two different Residual Networks are used to improve the high-level and low-level semantic features. ResNet has shortcut connections, which can bypass the deficient training layers using mapping between high-level features to low-level features. We need to enhance the image from poor quality due to insufficient light into preferable quality so that to process the image to identify the objects efficiently. Various image enhancement methods have been developed already which can be classified into two parts, called statistical-based and decomposition-based approaches [9]. By using volume-based subspace analysis, the illumination part is identified and the noise can be reduced.

We summarize the main contributions of this paper as follows:

We introduce the new convolution layers inside the existing ResNet CNN architecture called Modified ResNet (M-ResNet) which includes bilateral and adaptive super sampling operations.

- The proposed operation consists of residual units which enhance the image quality from the low illumination to normal illumination.
- If the image has already a good lighting environment, the proposed residual units are skipped, so that we avoid unnecessary computations.
- The introduced residual unit removes the noise as well as strengthens the edges of each object present in the captured image.
- The performance of the proposed method has experimented on different data sets. The preferable improvements are achieved from the proposed method while comparing existing methods.

The rest of the paper is structured as follows: Section 2 describes the object detection process in various methodologies like low light environment, Gaussian distribution, Probabilistic model, background subtraction, and Graph-based methods. Section 3 describes the proposed work; it shows how the image can be enhanced from the low light by superimposing large-scale and small-scale components of the bilateral operations. In Section 4, the proposed model is validated with different types of datasets and compared with related data. Section 5 summarizes the actual work of this paper and proposes possible future directions.

2 Related Work

2.1 Object Detection in Low Light Environment

Detecting the objects in the low light environment is a challenging task in a surveillance system. The captured images are having a lot of dark areas and noise in the picture information, due to insufficient light. The existing deep learning methods cannot perform well in the low light environment. In this situation, the classification of the objects is difficult due to the irregular distribution of brightness. To get accurate object detection, the captured image from a low light environment must be enhanced. The following disadvantages exist, in the process of enhancing low light images using deep learning methods,

   (i) Need complex structure and a huge number of parameters.
  (ii) Need additional layers and more computation
 (iii) The training needs paired dataset, but in practice paired images are difficult to obtain.

The above issues lead to the object detection systems becoming low performance and consuming more computing power. Hence simple mechanism needs to solve these issues. The first and second issues are solved by the volume-based subspace [10,11] to segregate the illumination area and noise in the image. The illumination component can be identified and segregated by principle energy analysis in the subspace [12]. The noise can be suppressed by using an adaptive truncation scheme, in the volume-based subspace [13]. The deep noise suppression [14] and regularized illumination an optimization method [15] are also used for suppressing noise. The combination of refined illumination and optimized reflection map can be

applied to enhance the image from the low light environment. Another method deals with a Night Vision Detector (NVD) based on Receptive Field Block (RFB)-Net [16,17], which uses both the context fusion and feature pyramid network. Here various illumination data can be modeled separately even though they are interfering with each other during the training period. The third issue can be solved by the new model called Retinex Generative Adversarial Network (GAN) and EnlightenGAN [18]. In which the training phase uses unpaired datasets and this model uses the simple generative adversarial network. Another method constructed by the pipelined convolution neural network with Gaussian kernel is based on multi-scale Retinex and discrete wavelet transformation [19], performs denoising net as well as image enhancement net. This architecture learns the image enhancement function from dark and bright image pairs.

2.2 Object Detection Using Gaussian Distribution

Noise corruptions and illumination change heavily affect the performance of the change detection algorithms. The existing noise severely affects the relationship between the neighborhood pixels. The salience enhancement approach [20,21] improves the saliency weight information of objects. By suppressing background information with the Gaussian mixture model [22], that makes to identify the feature difference between changed and unchanged regions. Another classic techniques are denoising algorithms [23], the blind universal image fusion denoiser network [24], which derive the optimal fusion denoising function integrated with Fusion Net. It evaluates the optimality of the noise level of the training network and unseen Gaussian noise levels using the Bayesian solution. The Gaussian Process Morphable Model (GPMM) [25] effectively defines the covariance function compared to Point Distribution Model. This GPMM integrates different kernel functions in recent registration schemes. Since using the Gaussian process, shape variations can be effectively approximated. The multivariate Gaussian approach [26] can model the normal data distributed in deep feature representations. Transferring the learned representations from large datasets like ImageNet into small datasets is also efficient. The existing Gaussian Mixture Modeling (GMM) for the background subtraction method is highly affected by noise and dynamic background [27]. By using modified GMM background subtraction [28]. First, the background model is reconstructed by averaging image blocks, then eliminating the noise information, and finally updates the background information for effective results.

2.3 Object Detection Using Probabilistic Methods

Medical images are characterized by dynamic image intensity and variable boundary information, for this reason, common image segmentation methods are not fit for the UltraSound images. To resolve this issue, Bayesian CNN [29] can be applied to Ultra Sound images to generate random predictions from probability distributions. In this approach, the probability can be evaluated by the combination of Magnetic Resonance Imaging (MRI) volumes and femoral cartilage contours. For modeling brain images expectation, the maximization (EM) technique is usually applied. These methods also need some unique denoising methods [30] for any level of noise. Another method [31], extends the probabilistic atlas which provides the healthy tissue information, by latent atlas which provides the lesion information. This generative probabilistic model and discriminative extensions provide semantic meaning to the tissues. To make a more accurate segmentation result, the author [32] makes a new algorithm by using the probability distribution of both object and background. The proposed framework provides the maximization of the distance between the background and Gaussian mixture distributions. This probability-based model is applied to different imaging modalities like dermoscopy and chromoendoscopy and MRI. In remote sensing image processing, due to the presence of thin clouds, can cause the effectiveness of cloud detection algorithms. It is necessary to remove the cloud content before processing remote sensing images. The author [33] proposes a deep learning cloud detection algorithm based on the combination of attention mechanism and probability upsampling. The algorithm focuses on

the relationship between the spatial dimension and multispectral image of spectral segments. The single label retrieval is updated into the multi-label by a fully convolution network [34] and choosing the right sampling pattern is important to reconstruct high-quality images [35]. The probability mass function-based sampling approach can dynamically adapt the sampling rate based on the data measured in advance. This static incremental sampling technique with probability mass function avoid the sampling delay so that the reconstruction of high quality images.

2.4 Object Detection Using Background Subtraction Methods

The background subtraction method is critical for distinguishing static and moving objects. Dynamic changes in the background will exacerbate the complexity of this procedure and result in erroneous results. As a result, the dynamic Auto Regressive Moving Average (ARMA) model [36] is used, which makes use of the spatial and temporal correlations between the input images to create an appropriate model for the background image. An adaptive least mean square technique can be used to update the dynamic features of the background. The fuzzy histogram describes the temporal properties of the pixels by utilising the fuzzy C means clustering with fuzzy nearness degree (FCFN) [37] background subtraction method. It overcomes categorization challenges by classifying things in the background and foreground. Due to the huge number of bands in Hyper Spectral Images (HSI), dimension reduction is required prior to processing. After dimensionality reduction with HSI, a hyperspectral visual attention model (HVAM) [38] is used to detect anomalies. Remove the noise with a curvature filter and then use the background subtraction method to acquire the first result. To acquire a final result, the given partial result might be submitted to the adaptive weight approach. Additionally, fast and slow illumination changes have an effect on the background subtraction models. The adaptive local median texture feature [32] is introduced to address this issue. It computes the adaptive threshold value for foreground pixels. By utilising ALMT characteristics in foreground pixels, the background model samples are compared to the image sequence from the video. To get the optimum object detection performance in low-light conditions, it is vital to choose the appropriate background removal methods and associated parameters. The author [39] conducted an examination of several background subtraction algorithm settings in order to develop an optimal background subtraction method with the required parameters for detecting nighttime falls.

2.5 Object Detection Using Graph Based Network

Because the input flow between various neurons in a convolutional neural network may be viewed as a graph, developing graph-based convolutional neural networks (GCNNs) is a rising technique in image processing. GCNN can be classified into two categories based on the filters used: spatial-based techniques and spectral-based approaches. The spatial-based technique is based on the aggregation of neighbouring pixels, whereas the spectral-based technique is based on the undirected graph. The learning process is harmed significantly as a result of the graphs' lack of direction. The directed graph convolution network is constructed using a fast localised convolution operator that scales well to huge graphs. There is a possibility that information about the item's boundaries is lost in video salient object detection models. The author [40] blends the advantages of the graph model and the deep neural network. For video SOD, the proposed solution utilises a unified multi-stream architecture. This architecture operates inside the context of GCN, which provides a mechanism for effectively grouping the common super pixels. The author [41] proposes a new attention module for superpixel encoding. Finally, smoothness awareness regularisation is used to assure the homogeneity of critical items. Skeleton-based action recognition systems typically employ hierarchical GCN, which may result in the loss of information on joint properties after extended diffusion. To increase the local context information of joints, the author [42] suggests a multi-scale mixed dense graph CNN. Two modules, spatial and attention, are used to fine-tune

the spatial-temporal characteristics. This suggested model has distinct kernel sizes for each layer, resulting in a highly flexible temporal graph.

Few modifications are required to boost the efficiency of image processing while dealing with denoising challenges. A graph convolution layer can be added into a trainable neural network design [43–46], which discovers the relationship between the network's hidden features, hence enhancing the network's robust learning power. Each pixel is represented as a vertex in a graph convolution network, and dynamically determined similarities are represented as edges. The advantages of incorporating graph convolution into a current CNN are that neighbourhood graphs are calculated dynamically, the constructed non-local filters aggregate the weights of the features, and predefined parameter operations are avoided. The architecture makes use of both local and non-local similarities to provide adaptable functionality. By combining the advantages of GNN and CNN, it is possible to solve the knowledge base completeness problem using things that are not part of the knowledge base [47]. To transfer knowledge for entities that are not in the knowledge base, a new method is proposed that utilises the weight matrix to describe the relations in the KBC model. After learning the information between nodes in this design, transition matrices are used to build more expressive embeddings. The suggested transition-based knowledge graph model solves knowledge base completion tasks using these parameter values [48,49].

3 Proposed Work

3.1 Resnet Architecture

By adding more layers to deep learning architecture, we can solve the complex tasks in the image processing operations like classification and recognition of particular objects. But the addition of more layers in the neural network turns into accuracy loss and a challenging training phase. The residual blocks in the Resnet architecture has overcome this issue. The Resnet architecture contains 34 Layers and has shortcut connections between the layers. These shortcut connections are called residual blocks. The overview of Resnet architecture is illustrated in Fig. 3.

Figure 3: 34 layer residual network

By adding the new layers that enhance the image by means of avoiding low and improper lighting effects and increasing the accuracy of object detection. The addition of new layer operations is Bilateral, Adaptive supersampling, and symmetric local binary pattern, shown in Fig. 4. Bilateral filtering operation produces noise-reducing smoothing operation at the same time edges are preserved. So the skeleton of the objects present in the video frames can be maintained, and hence the objects are identified accurately.

3.2 Proposed Architecture

Bilateral Filtering

The input image is treated to a non-linear bilateral filtering procedure using a video sequence as the source. This method improves the smoothness of the image while retaining the edge information. This technique calculates the average of the adjacent pixels, which can be substituted by the original pixel. Thus, the weighted average of pixels is another name for this bilateral process.

$$\text{BFI}_p = \frac{1}{\underset{\downarrow}{Wp}} \sum_{q \in S} \underbrace{G\sigma s((\|p - q\|))}_{\text{Space Weight}} \underbrace{G\sigma r(|\text{Ip}-Ip|)}_{\text{Range Weight}} \text{I}_q$$
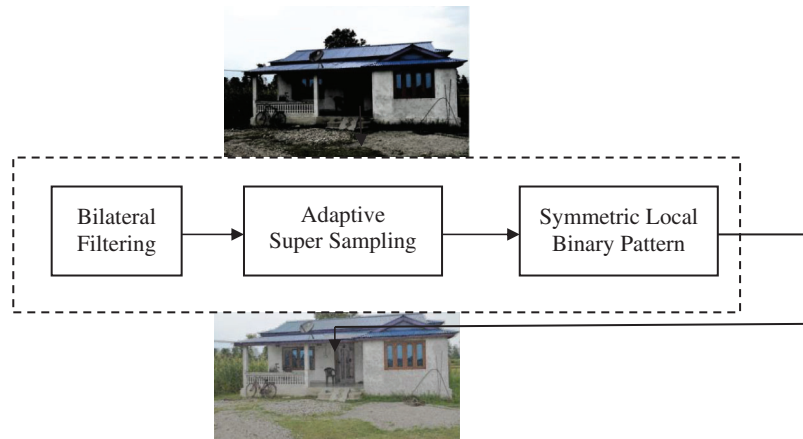
Normalization
Factor

Figure 4: Additional layer operation

This process provides that the two pixels are similar to each other means, not only for the adjacent regions, but also they are having some similar features. The bilateral filter operation mentioned by,

$$\text{BF}[\text{I}_{\text{input}}]_p = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) F_r(\|I(x_i) - I(x)\|) g_s(|\, |x_i - x|\,|)$$

$$\text{W}_p = \sum_{x_i \in \Omega} F_r(\|\text{I}(x_i) - \text{I}(x_i)\|) g_s(\|x_i - x\|)$$

As illustrated in Fig. 5, the input image can be divided into two layers: a smoothed version referred to as a large scale component and a residual version referred to as a small scale component. These remaining parts contain noise and provide insight into the structure of the input image, which is beneficial throughout the denoising operation. Bilateral filtering combines domain and range filters. It calculates the mean of a pixel's similar and neighbouring pixel values and replaces it. To apply this proposed work to the video surveillance system, sample photographs in proper lighting circumstances are acquired in advance. During surveillance time, particularly at night, the suggested system compares the current and sample image frames to the image at a specific time interval. The sample image's small scale component can be superimposed on the present image frame's big scale component to create an enhanced image that accurately depicts all of the object's details, allowing the object identification process to proceed efficiently. Fig. 6 illustrates the full operation.

3.3 Adaptive Super Sampling

The result of bilateral operation may contain pixelated edges, which contribute to the picture data's aliasing effect. Aliasing happens as a result of smooth curves and lines that continue indefinitely. A few samples are taken for each pixel; if the samples are comparable, the output pixel value is determined; if the samples are dissimilar, additional samples are taken to establish the target pixel value. As a result,

more samples are not required at all times shown in Fig. 7. Thus, adaptive supersampling supersamples only the pixels on the edges of objects, thereby preserving the objects' edges. This operation approximates the integral of a function f as the average of the function evaluated at a set of points $x_1, \ldots, x_N$:

Large Scale Components   Small Scale Components

Figure 5: Output of bilateral filter

This can be calculated by aggregating the image function p(x, y), which can represent the radiance of the particular point (x, y) in the image pixels. The Radiance L can be calculated by,

$$L = {}^n_A p(x, \ y)f(x, \ y)dx \ dy$$

Here, f(x, y) is a anti-aliasing filter, A is a supporting area of the filter. The Random samples based on the Monte Carlo method [11], X_i, i = 1,....n:

$$L = \frac{1}{n}\sum_{i}^{n} Xi$$

The samples are disseminated to corresponding kernel filters.

Symmetric Local Binary Pattern

It labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number. The LBP operator in a video surveillance application can find the variations during illumination changes. The value of the LBP code of a pixel is calculated by

LPB can be calculated by identifying the difference between the intensities of pixels of neighbourhood pixels. Let I_0 represent the intensity of a particular pixel, and the neighbours are represented as In, where n represents the position of the neighbour. Fig. 8 represents the size of n is 8. If the neighboring pixel value is equal or greater, the value is set into one, otherwise, it is zero.

$$LBP \ \ = \ \sum_{i=0}^{p-1} s(n_i - G_c)2^i$$

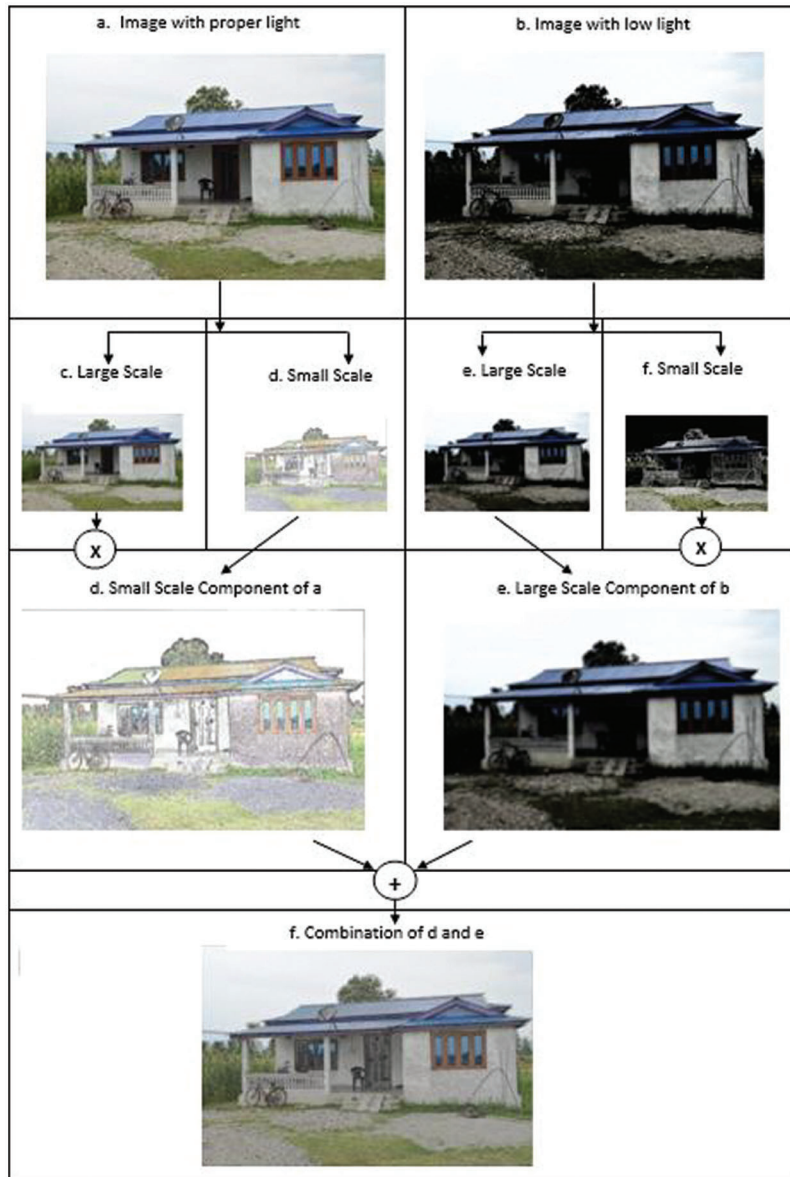$$s(x) = \begin{cases} 1, & if \ x > 0 \\ 0, & otherwise \end{cases}$$

Figure 6: Modified image generation

Experiment & Results

In order to illustrate the difference between the existing ResNet architecture and modified ResNet (M-ResNet) architecture during the processing of low illumination images in video surveillance system, we selected more images from the three different datasets coco, CIFAR, and wild tract, for testing to check the effective improvements after the modification in the ResNet architecture. The test results are compared with the existing methods. In the training phase there are 5000 images are used in the coco dataset. In Fig. 9, the three different images having the bad light conditions can be applied to existing resnet architecture and modified resnet architecture. In the coco dataset, two variations of the same image can be taken for the experiment which can be represented in Fig. 9. Here the input picture is pre-processed with the normal lighting condition (Fig. 9a). This image has proper lighting condition and three

zebras are identified, and the probabilities of the three objects are listed as 99.9, 99.8, and 99.4 respectively. The same image that does not having proper lighting conditions, can be mentioned in Fig. 9b. The second image is now tested with the resnet model and produce the result as four zebras and one horse are found. Due to the bad lighting environment, this erroneous output can be obtained.
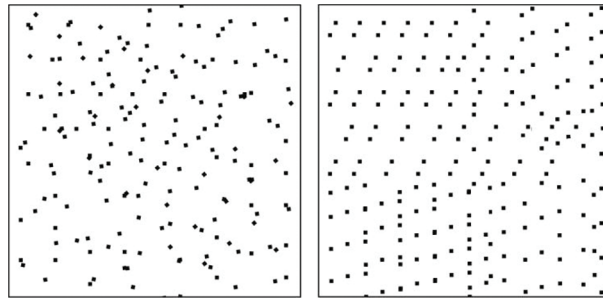

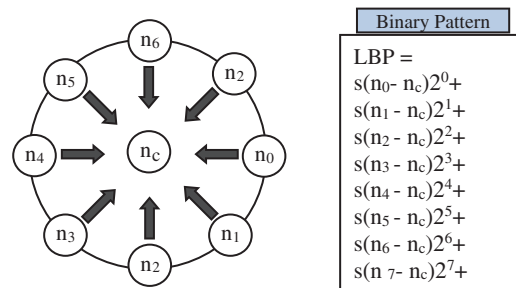**Figure 7:** Output of adaptive super sampling

Figure 8: Local binary pattern

After applying the mentioned modifications described in the proposed architecture to the existing model, the challenging light condition image can be subjected to bilateral filtering and adaptive sampling process that can increase the atmospheric light environment in the image. The new image can be subjected to the convolution process, and the outputs can be recorded. The first and third outputs are almost the same and less error. That can be depicted in the following Fig. 10.

Fig. 10 clearly shows the improvement of the detection process after the modification in the deep learning network. Similarly, the different images are subjected to the proposed convolution architecture. These images are taken from the Wildtract seven camera hd dataset (1920×1080 resolution) and real-time capturing images.
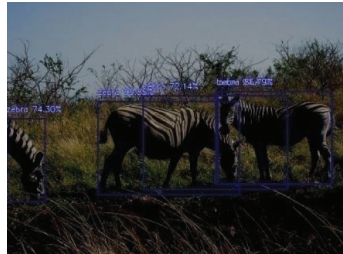
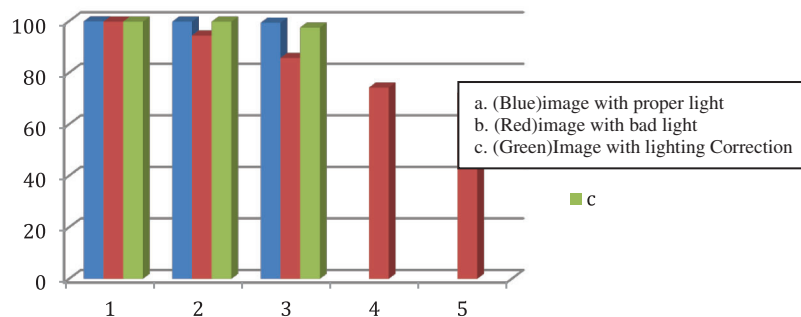|		
a.1 Image with proper light	b.1.Image with bad light	c.1.after correction
a.2.object detection	b.2.object detection	c.2.object detection
zebra : 99.92891550 064087 zebra : 99.82913136482239 zebra : 98.41715693473816	zebra : 99.84723329544067 horse : 94.41177248954773 zebra : 85.79044938087463 zebra : 74.2976725101471 zebra : 72.1422553062439	zebra : 99.88270401954651 zebra : 99.811416864395 14 zebra : 97.57338762283325
a.3 Probability of detection	b.3 Probability of detection	c.3 Probability of detection

Figure 9: Testing the dataset images

Figure 10: Analysis of existing and modified architecture results

4 Performance Evaluation

The performance can be evaluated by different parameters like recall, precision, and F1 score, pixel accuracy, intersection over union, and mean intersection over union, in the coco, WildTrack, and CIFAR data sets. The recall provides the completeness of the obtained predictions to the ground truth. The precision illustrates, how the positive detections are relative to the ground truth. The pixel accuracy is the percentage of pixels in the image that are classified correctly. The intersection over union (IoU) is also

called the Jaccard index, to provide the percentage overlap between target and predicted output. The mean IoU metric is measured by the average of all semantic class intersection over union values.

Sn	Method	Precision	Recall	F1-score	Pixel accuracy	IoU	Mean IoU
1	Generative probablistic model	64	81	75	81	80	71.3
2	ACNN model	64	91	74	90	84	79.5
3	Faster R-CNN	52	81	72	79	69	58.8
4	Gaussian mixture model	72	86	74	80	81	75.4
5	DCGAN model	66	83	72	85	79	68.9
6	Active contour mode	73	88	80	91	82	80.2
7	Dilated convolution	62	85	79	80	78	70.5
8	Edge_Constraint U-Net	56	80	72	83	74	58.4
9	CSCC algorithm	67	81	76	81	77	68.4
10	Proposed model	75	84	83	90	86	84.3

It can be concluded from the above graphs, the performance of M-ResNet can be much more improved when compared to existing methods.

5 Conclusion

In this article, an improved ResNet model is proposed to avoid the fault detection of objects due to insufficient light in the video surveillance system. Resnet architecture has the skip connection to avoid the problems due to vanishing exploding gradient. To include the new layers with skip connection in existing Resnet architecture will provide better results on the low illumination images in video surveillance system, without affecting the performance. The new layer operations include the enhance of lighting conditions using bilateral filtering, avoiding the anti-aliasing effect using adaptive sampling, improve the quality of image using the local binary patterns. These three operations give more clear information for further analyzing the image for a better object detection process. This modified resnet architecture is compared against the different image quality parameters with various datasets. By comparing existing methods, the proposed method shows better results. There is a limitation in this article is, if the processed image has very low illumination, it takes more time to process the data for real-time images. The future work will explore how to improve the image illumination data without using both low light and normal light images in the video surveillance object detection process, which provides both the better performance as well as less processing time.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Ding, L. Wang and L. Cong, "Super-pixel image segmentation algorithm based on adaptive equalisation feature parameters," *IET Image Processing*, vol. 14, no. 17, pp. 4461–4467, 2020.

[2] Y. Wu, J. Zheng, W. Song and F. Liu, "Low light image enhancement based on non-uniform illumination prior model," *IET Image Processing*, vol. 13, no. 13, pp. 2448–2456, 2019.

[3] J. Zhang, Y. Wang, H. Wang, J. Wu and Y. Li, "CNN cloud detection algorithm based on channel and spatial attention and probabilistic upsampling for remote sensing image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[4] D. Valsesia, G. Fracastoro and E. Magli, "Deep graph-convolutional image denoising," *IEEE Transactions on Image Processing*, vol. 29, pp. 8226–8237, 2020.

[5] Q. Guo, C. Zhang, Y. Zhang and H. Liu, "An efficient SVD-based method for image denoising," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, 2015.

[6] T. Ma, M. Guo, Z. Yu, Y. Chen, X. Ren *et al.,* "RetinexGAN: Unsupervised low-light enhancement with two-layer convolutional decomposition networks," *IEEE Access*, vol. 9, pp. 56539–56550, 2021.

[7] B. Li and Y. He, "An improved ResNet based on the adjustable shortcut connections," *IEEE Access*, vol. 6, pp. 18967–18974, 2018.

[8] D. Kang, S. Park and J. Paik, "SdBAN: Salient object detection using bilateral attention network with dice coefficient loss," *IEEE Access*, vol. 8, pp. 104357–104370, 2020.

[9] W. Kim, R. Lee, M. Park, S. H. Lee and M. S. Choi, "Low-light image enhancement using volume-based subspace analysis," *IEEE Access*, vol. 8, pp. 118370–118379, 2020.

[10] H. Lee, K. Sohn and D. Min, "Unsupervised low-light image enhancement using bright channel prior," *IEEE Signal Processing Letters*, vol. 27, pp. 251–255, 2020.

[11] L. Sun, B. Jeon, B. N. Soomro, Y. Zheng, Z. Wu *et al.,* "Fast superpixel based subspace low rank learning method for hyperspectral denoising," *IEEE Access*, vol. 6, pp. 12031–12043, 2018.

[12] Q. Lv, W. Feng, Y. Quan, G. Dauphin, L. Gao *et al.,* "Enhanced-random-feature-subspace-based ensemble cnn for the imbalanced hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3988–3999, 2021.

[13] C. Li, X. Qin, X. Xu, D. Yang and G. Wei, "Scalable graph convolutional networks with fast localized spectral filter for directed graphs," *IEEE Access*, vol. 8, pp. 105634–105644, 2020.

[14] H. Su and C. Jung, "Perceptual enhancement of low light images based on two-step noise suppression," *IEEE Access*, vol. 6, pp. 7005–7018, 2018.

[15] Y. Guo, Y. Lu, R. W. Liu, M. Yang and K. T. Chui, "Low-light image enhancement with regularized illumination optimization and deep noise suppression," *IEEE Access*, vol. 8, pp. 145297–145315, 2020.

[16] Y. Xiao, A. Jiang, J. Ye and M. W. Wang, "Making of night vision: Object detection under low-illumination," *IEEE Access*, vol. 8, pp. 123075–123086, 2020.

[17] G. Chen, L. Li, W. Jin and S. Li, "High-dynamic range, night vision, image-fusion algorithm based on a decomposition convolution neural network," *IEEE Access*, vol. 7, pp. 169762–169772, 2019.

[18] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang *et al.,* "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[19] Y. Guo, X. Ke, J. Ma and J. Zhang, "A pipeline neural network for low-light image enhancement," *IEEE Access*, vol. 7, pp. 13737–13744, 2019.

[20] D. Xue, T. Lei, X. Jia, X. Wang, T. Chen *et al.,* "Unsupervised change detection using multiscale and multiresolution Gaussian-mixture-model guided by saliency enhancement," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1796–1809, 2020.

[21] L. Han, X. Li and Y. Dong, "Convolutional edge constraint-based U-net for salient object detection," *IEEE Access*, vol. 7, pp. 48890–48900, 2019.

[22] F. Riaz, S. Rehman, M. Ajmal, R. Hafiz, A. Hassan *et al.,* "Gaussian mixture model based probabilistic modeling of images for medical image segmentation," *IEEE Access*, vol. 8, pp. 16846–16856, 2020.

[23] G. Ghimpețeanu, T. Batard, M. Bertalmío and S. Levine, "A decomposition framework for image denoising algorithms," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 388–399, 2016.

[24] M. E. Helou and S. Susstrunk, "Blind universal Bayesian image denoising with Gaussian noise level learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 4885–4897, 2020.

[25] M. Luthi, T. Gerig, C. Jud and T. Vetter, "Gaussian process morphable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1860–1873, 2017.

[26] O. Rippel, P. Mertens, E. Konig and D. Merhof, "Gaussian anomaly detection by modeling the distribution of normal data in pretrained deep features," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[27] M. Antico, F. Sasazawa, Y. Takeda, A. T. Jaiprakash, M. L. Wille *et al.,* "Bayesian CNN for segmentation uncertainty inference on 4D ultrasound images of the femoral cartilage for guidance in robotic knee arthroscopy," *IEEE Access*, vol. 8, pp. 223961–223975, 2020.

[28] K. Wang and M. Z. Liu, "Object recognition at night scene based on DCGAN and faster R-CNN," *IEEE Access*, vol. 8, pp. 193168–193182, 2020.

[29] T. Kim, J. Lee and Y. Choe, "Bayesian optimization-based global optimal rank selection for compression of convolutional neural networks," *IEEE Access*, vol. 7, pp. 17605–17618, 2020.

[30] B. H. Menze, K. V. Leemput, D. Lashkari, T. Riklin-Raviv, E. Geremia *et al.,* "A generative probabilistic model and discriminative extensions for brain lesion segmentation—with application to tumor and stroke," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 933–946, 2015.

[31] Y. Zhang, W. Zheng, K. Leng and H. Li, "Background subtraction using an adaptive local median texture feature in illumination changes urban traffic scenes," *IEEE Access*, vol. 8, pp. 130367–130378, 2020.

[32] E. Iqbal, A. Niaz, A. A. Memon, U. Asim and K. N. Choi, "Saliency-driven active contour model for image segmentation," *IEEE Access*, vol. 8, pp. 208978–208991, 2020.

[33] Z. Shao, W. Zhou, X. Deng, M. Zhang and Q. Cheng, "Multilabel remote sensing image retrieval based on fully convolutional network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 318–328, 2020.

[34] S. Grosche, M. Koller, J. Seiler and A. Kaup, "Dynamic image sampling using a novel variance based probability mass function," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1440–1450, 2020.

[35] J. Li, Z. M. Pan, Z. H. Zhang and H. Zhang, "Dynamic ARMA-based background subtraction for moving objects detection," *IEEE Access*, vol. 7, pp. 128659–128668, 2019.

[36] T. Yu, J. Yang and W. Lu, "Dynamic background subtraction using histograms based on fuzzy c-means clustering and fuzzy nearness degree," *IEEE Access*, vol. 7, pp. 14671–14679, 2019.

[37] P. Xiang, J. Song, H. Qin, W. Tan, H. Li *et al.,* "Visual attention and background subtraction with adaptive weight for hyperspectral anomaly detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2270–2283, 2021.

[38] M. Alonso, A. Brunete , M. Hernando and A. Gambao, "Background-subtraction algorithm optimization for home camera-based night-visionfall detectors," *IEEE Access*, vol. 7, pp. 152399–152411, 2019.

[39] Y. Xu, Y. Feng, Z. Xie, M. Xie and W. Luo, "Action recognition using high temporal resolution 3D neural network based on dilated convolution," *IEEE Access*, vol. 8, pp. 165365–165372, 2020.

[40] M. Xu, P. Fu, B. Liu and J. Li, "Multi-stream attention-aware graph convolution network for video salient object detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 4183–4197, 2021.

[41] Z. Bi, T. Zhang, P. Zhou and Y. Li, "Knowledge transfer for out-of-knowledge-base entities: Improving graph-neural-network-based embedding using convolutional layers," *IEEE Access*, vol. 8, pp. 159039–159049, 2020.

[42] H. Xia and X. Gao, "Multi-scale mixed dense graph convolution network for skeleton-based action recognition," *IEEE Access*, vol. 9, pp. 36475–36484, 2021.

[43] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai *et al.,* "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2017.

[44] Y. Lin, X. Sun, Z. Xie, J. Yi and Y. Zhong, "Semantic segmentation with oblique convolution for object detection," *IEEE Access*, vol. 8, pp. 25326–25334, 2020.

[45] Q. Chen, T. Liu, Y. Shang, Z. Shao and H. Ding, "Salient object detection: integrate salient features in the deep learning framework," *IEEE Access*, vol. 7, pp. 152483–152492, 2019.

[46] S. Fan, H. Yu, D. Lu, S. Jiao, W. Xu *et al.,* "CSCC: Convolution split compression calculation algorithm for deep neural network," *IEEE Access*, vol. 7, pp. 71607–71615, 2019.

[47] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan *et al.,* "An improved faster R-CNN for small object detection," *IEEE Access*, vol. 7, pp. 106838–106846, 2019.

[48] W. Sun, G. Z. Dai, X. R. Zhang, X. Z. He and X. Chen, "TBE-net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, pp. 1–13, 2021.

[49] W. Sun, L. Dai, X. R. Zhang, P. S. Chang and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 1, pp. 1–16, 2021.