*Deep Background Feature Transformation Learning*

\*Note: Sub-titles are not captured in Xplore and should not be used

line 1: 1st Given Name Surname   
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address

line 1: 2nd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address

line 1: 3rd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address

line 1: 4th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address

*Abstract*—One of the main challenges of foreground detection comes from the diversity and complexity of real-world scenes. While most of previous works in this field were proposed by designing an artificial model, we tend to present a self-adaption solution, by combining the fully convolutional networks (FCNs), which is named as the Deep Pixels Variation Learning model (DPVL). Our main job is to find a new representation of pixels historical observations in a new feature space. In this paper, videos are divided into fixed length image blocks, which are later transformed into the pixel’s observation patches as the input of a FCN for learning the variation of pixels. We conduct our training and prediction processes on block level. The architecture of the FCN is devised from the semantic segmentation problem. Experiment shows that our model has a strong learning ability to the patterns of those permutations of pixels’ variation. We compare our method with some of the state-of-art methods…

Keywords—Background, Deep learning, Feature, FCN

# Introduction

Background subtraction or foreground object detection, as a fundamental problem in computer vision, has been much discussed with the increasing number of outdoor cameras over the last few decades. It is widely used as the pre-processing step of video processing, which can help us efficiently mark the region of interest, e.g. vehicles and humans, thus saving us huge amount of computing resources. Typically, Background subtraction is a binary classification task that assigns each pixel in a video sequence with a label, for either belonging to the background or foreground scene.

With a huge number of works, existed algorithms have already achieved well performances in the scenes of low diversity or complexity, such as the indoor scenes. However, background Subtraction is still unsolved because of the diversity in background scenes and the changes originated from the camera itself. Scene variations can be in many forms such as, camera jitter, dynamic background, bad weather, illumination changes, intermittent object motion. This is principally because the major methods, in most cases, try to find a universal solution by generating some universal background models for later linear classification; for instance, the earlier Frame Difference Methods, trying to get a fixed image as background to classify each pixels. The fundamental problem lies in that linear classifier might not be powerful enough for the job, due to the complexity and the diversity of natural scenes.

In order to present a better and universal solution, we proposed the Deep Pixels’ Variation Learning (DPVL) model for Background Subtraction in diversely natural scenes. In our method, the main job is to transform the sequence of pixels into another sequence which is easy for classification.

sq1 = {p\_1,p\_2,...p\_n} -> sq2 = {f\_1,f\_2,...f\_n} (1)

For sq1, it is a sequence of pixels' intensity, which is hard to classify which entry is foreground or background.

But in sq2, it is more easy to classified, since the output of the network is the label of each entry in sequence.

In this paper, pixel’s observation patches are proposed to describe the historical observation of pixels. And a random initialized FCN network is trained to learn the historical observation of pixels and generate a new representation in a new feature space. We take advantage of the strong learning ability of FCN to learn a new representation of the Pixel’s observation patches in a new space where the pixels can be easily classified to background and foreground.

# RELATED WORK

Over the last few decades, Background subtraction has been well studied. Meanwhile, a huge number of methods were proposed. These methods can be broadly categorized into pixels-based, region-based, frame-based and Deep Learning.

## pixels-based methods

The most widely used algorithms in Background subtraction are pixels-based methods. And one of the famous method is Gaussian Mixture Model (GMM), in which a GMM is used to model the history over time of pixel’s intensity values. It is assumed that pixels are independent from their neighbors. Incoming pixels are labeled as background if there exists a Gaussian in the GMM, where the distance between its mean and the pixel lies within a certain bound. For learning the parameters, that maximize the likelihood, the authors proposed an online method that approximates the Expectation Maximization (EM) algorithm.

In XXX , Mingliang Chen et al. propose a background subtraction algorithm using hierarchical superpixels segmentation, spanning trees and optical flow. Their Background model combine the GMM with constrains of temporal and spatial from optical flow and superpixels.

Kim et al used a codebook to record the sampling background values at each pixel, which can be seemed as a compressed representation of background model. This allows them efficient in memory and speed compared with other background modeling techniques. The final foreground is detected by a distance measurement in a cylindrical color model.

In XXX, Zhi Zeng et al. proposed an equal-qualification updating strategy to replace the maximum-negative-run-length-based filtering strategy. Their experiments show that, the proposed method outperforms well, despite using only color information.

Elgammal et al. introduced a probabilistic non-parametric method to model the background. It is assumed that each background pixel is drawn from a PDF. The PDF for each pixel is estimated with Kernel Density Estimation (KDE).

## region-based approach

Identify applicable funding agency here. If none, delete this text box.

region-based approach assumpt that the neighbouring pixels have a similar variation as the pixel itself.

In xxx, Sriram Varadarajan et al. propose a region-based MoG to takes in which the updated mixtures represent the scene distribution in a neighbourhood region.

In (PCA), classification is done by comparing a block in current frame to its reconstruction from PCA coefficients and declaring it as background if the reconstruction is close.

A recently region based method is presented in XXX which used the statistical circular shift moments (SCSM) in image regions for change detection.

subspace learning method in xxx, is used to compress the background into the eigenbackground. For each video, the mean and the covariance matrix are calculated. After a PCA of the covariance matrix, a projection matrix is set up with M eigenvectors. Then, incoming images are compared with their projection onto the eigenvectors. Foreground labels are assigned to pixels with large distances, after calculating the distances between the image and the projection and comparing them with the corresponding threshold value.

Marghes et al. used a mixed method that combines a reconstructive method (PCA) with a discriminative one (LDA) to robustly model the background.

single Gaussians are employed for foreground modeling. By computing flux tensors, which depict variations of optical flow within a local 3D spatio-temporal volume, blob motion is detected. With the combination of the different information from blob motion, foreground models and background models, moving and static foreground objects can be spotted. Also, by applying edge matching, static foreground objects can be classified as ghosts or intermittent motions.

Frame-based background modeling via Principal Component Analysis (PCA) and low-rank/sparse decomposition approaches is a popular alternative to pixel-level modeling xx. These approaches are however not ideal for surveillance applications as most rely on batch or offline processing or suffer from scaling problems.

In XXX, ss et al. addressed scaling problems by reformulating principal component analysis for 2D images. Meanwhile their method takes much lower memory consumption and computational cost than others. Some online approaches have also been proposed recently, but they are still very computationally expensive.

## Deep learning methods

A novel approach for background subtraction with the use of CNN was proposed by Braham and Droogenbroeck. They used a fixed background model, which was generated from a temporal median operation over N video frames.

In XXX, Yi Wang et al. tried a CNN architecture combined with a Cascade model for segmentation in Background subtraction. Given 200 labelled images as training set, their model performed excellently in dataset2014.

Braham et al. present an Deep learning-based method. with the help of CNN, they generated a fixed background model from a temporal median operation over N video frames. Then, a scene-specific CNN is trained with corresponding image patches from the background image.

In xxx, M. Babaee et al. combine the segmentation mask from SuBSENSE algorithm and the output of Flux Tensor algorithm, which can dynamically change the parameters used in the background model based on the motion changes in the video frames. They also used spatial-median filtering as the post processing of the network outputs.

# Methodology

In this section, we introduce our DPVL model that consists of a pixels-based observation patches and a novel FCN networks for background subtraction. We explain the details of the procedures of capturing the observation patches and the architecture of network in our DPVL model. The complete system is illustrated in Figure XXX. We use a set of input and ground true images to generalize the Pixels’ Variation Permutations, and reshape them into fixed-size image patches as and fed into a FCN. After reassembling the patches into the complete output frame, it is post-processed, yielding the final segmentation of the respective video frame.

## Image blocks to observation patches

The historical observations of pixels which belong to the background are usually share a common variation pattern. In more specific terms, the variations of pixels in background are generally keeping in some pattern. And we want to use a FCN network to learn a new representation of pixels’ observations in a new feature space.

In order to get enough pixels’ observations, we conduct our experiment on block-level. In this paper, we divide each video into smaller image blocks, which share a common frame length of t. For each image blocks, it contains M\*N pixels’ historical observations in a temporal sequence of t frames. We need to take them out for later transformation as our network input.

Observation patch is purposed to take advantage of the strong learning ability of the FCNs. And what we need is an pixel-to-pixel transformation. More precisely, Every historical pixels’ observation in these observation patches has an corresponding transformation at the last layer of FCNs. Here is how we get these observation patches. First, we transform each image block into M\*N pixel’s observation series which are the single pixels’ historical observation series in a frame length of t. Next, we reshape the observation sequence into a rectangle image which size is \*. That rectangle image is what we called observation patch. The observation patches are taken as network training sample set.

Another conception about observation patches is interval sampling. For a sequence like this:

p\_t {p\_1,p\_2,...p\_n} ->{p\_1,p\_11,p\_21,….p\_n}

Where t denote the sequence number of frames. By this way, we can get an new sequence contains more temporary information than the continuous pixel sequences. This works well when it comes to some situation where the moving objects keep stationary for a long time.

**上面还没改**

## FCN-based Pixels Variation Learning

Convolutional Neural Networks(CNNs) are powerful models which keep giving state-of-the-art performances on recognition, detection and segmentation. However, there are some shortcomings of CNNs which should bring in consideration:

1). CNNs require a fixed-size input image due to the utilizing of fully connected layer. This requirement might impair the detection accuracy and bring challenges to the adaption of network.

2). Fully connected layers might also lead to more parameters and increasing of calculations, since each neuron in current layer has to connect to every neuron in previous layer.

3). In fully connected layers, feature maps are converted into vectors, which would result in loss of spatial information of inputs.

Based on above-mentioned factors, fully convolution neural network(FCN) [] is designed as the alternative network architecture which use convolutional layer with 1\*1 kernels to take the place of the fully connected layer.

Different with CNNs, FCN can take input of arbitrary size and produce correspondingly-sized output with efficient inference.

FCNs have already been used in many areas like sematic segmentation and so on. Researchers found FCN has a strong learning ability which won’t lost to the traditional ones, meanwhile, it also has a high efficient computation ability.

**我这里是想从CNN的不足引出我们为什么要用FCN，然后介绍FCN，后面的我就有点凌乱了。不知道接下来应该带着什么样的目的去写。要不你给点提示吧。或者给篇论文让我参考一下。**

%%%%%%%%%%%%%%%%%%%

The primary task of our network is to learn an optimal feature transform of pixel’s observation series, thus we can find a new representation which is easy for classification from raw RGB data. There needs to be a consistent one-to-one match between pixels’ observation sequence and our network output on the time series. In these circumstances, traditional convolutional network cannot meet our requirements.

Therefore, we take a fixed size of pixel’s observation patches as the input of our network. For the convenience of calculations, we transform observation sequence into observation patches.

And the structure of our fully convolutional network for background modeling is shown in Fig.3.

The architectures of our proposed FCN is illustrated in Figure xx, our network contains 5 convolutional layers, 2 pool layers and a convolutional layer which have a filter size of 1\*1. A short calculation revealed that the network output will less 10 pixels after forward calculating. In order to make output the same size as input, we borrowed ideas from Image semantic segmentation, which is doing zero-padding before the training. After the forward computing, we can get the new representation of input observation values in a new feature space. Then, after some thresholding calculations. Our experiment results show that a random initialized FCNs, trained end-to-end on feature learning can achieve the state-of-the-art without further machinery.

The image sequence is reshape. Moreover, the image blocks

In the experiment, we divide video into image blocks, each containing 100 frames. For each image block, we can obtain M\*N of pixel’s historical observation sequences, where the M and N denoted the size of images. After a simple deformation, pixel’s historical observation sequences are transformed into pixel’s observation patches. In training process, we take only one image block to generate the observation patches as our training sample set. Since most of the videos in dataset2014 contain over 1000 frames with groundtruth,

# E XPERIMENTAL R ESULTS

In this section, we ran comprehensive experiments to evaluate the performance of the proposed approach on the CDnet 2014 benchmark and CAMO-UOW. The CDnet is the largest dataset for background subtraction so far as we are aware, containing 11 categories with several complexly challenging scenes, such as Dynamic Background, Camera Jitter, Shadow, Night Videos, PTZ and so on. The CAMO-UOW is another challenging benchmark which contains 10 high resolution videos. For each video, one or two persons appear in the scene with the clothes in the similar color as the background.

The proposed approach is compared with several existing traditional state-of-the-art background subtraction algorithms, including the IUTIS-5[], the SuBSENSE[], the WeSamBE[], sharable GMM, the SharedModel[], word-dictionaries-based method[] and PAWCS[], the SemanticBGS[], the AAPSA[], etc. Moreover, two deep learning based algorithms are also compared with the proposed approach, which include DeepBS[], and DBMF[]. All the results of compared algorithms are provided by authors.

During the comparison, the F-measure(Fm) has been used for evaluation. The Fm is a general international standard in background subtraction which measures the segmentation accuracy by considering both the recall and the precision. The definition of Fm is shown as follows:

, (x)

Where TP, FP, and FN are true positives, false positives, and false negatives respectively, computed in pixels of all test frames for each video.

The quantitative and qualitative comparisons are shown in Table. I and Fig. X respectively. Due to the paper length, several typical videos are selected for the qualitative comparisons as well as the discussion. In the dynamic background scene, the video “canoe” is a typically challenging video which includes a large area of water rippling. The main challenge comes from the dynamic background, in which it is so hard to describe the background by a single image. In this condition, since the traditional background subtraction method such as the SharedModel and the WeSamBE do not have the enough ability to describe the complex dynamically background, they are fail to detect the people on the boat, as shown in the Fig. X. Besides, the detected moving objects of the SharedModel are not accurate in boundary due to the utilization of texture features. In contrast, benefited from the strong learning ability of Deep Learning network, the DeepBS successfully detected the people. Unfortunately, since the DeepBS ignores the fact that a single background is not enough to describe the dynamic background, even the deep learning based algorithm is suffering from the detection of the boat shape. In contrast, the proposed approach performed superior than others in this scene, since the essence of background subtraction is considered as a binary classification of pixels’ observation in time sequence. Based on this insight, the FCN network focuses on learning the patterns of the observation variation rather than a static background image, and proposed approach achieves promising performance in the canoe video.

As for the case of shadow scene, ‘peopleInShade’ is a typical example with prevalent hard and soft shadows. In the traditional approaches, these shadow regions are usually segmented as foreground since it is also moving with the objects. Therefore, traditional methods like the PAWCS, the WeSamBE and the SharedModel falsely segments part of shadows as moving objects. In addition, the foreground provided by the UBSS is incomplete on the part of the pedestrian’s body due to the interference of shades, which can be owing to the severe dependency of texture features. Whereas, the DeepBS performs well in this video benefited from the utilization of CNN. However, the shape of pedestrians are slightly deformed as the result of their patch-wise processing of CNN. In contrast, derived from the fact that our DPVL focus on learning the pattern of pixels’ variation in the shadow regions, proposed approach successfully segments the shadow part as the background and achieves the highest performance in the category of shadow scene.

In the video “corridor” among the Thermal scene, there is no color information since the videos are obtained through a Thermal camera. Moreover, the moving objects in these videos are exceedingly fuzzy and indistinct, which is the main challenge of this category. The WeSamBE, the SharedModel and the PAWCS successfully detect the target objects, owing to a stable background in this indoor video. However, they fail to remove the reflections since their modeling ability have already reached a limit under the extreme condition of thermal map. The DeepBS, by contrast, succeeds in eliminating most of the reflections. Meanwhile, the moving objects are also clearly divided from the background thanks to the strong modeling ability of CNN. However, due to the dependency of edge feature, a small object were missed in the detection result. Fortunately, the proposed approach focus on the pattern of pixels’ variation, which should be theoretically effective even in the observation without the color information. Consequently, our DPVL performed much better than compared algorithms, with the situation that most parts of shadow are removed and the segmentation results are more accurate.

The quantitative evaluation of proposed approach on CDnet 2014 is shown in the Table 1. It can be inferred that the proposed approach significantly outperformed all of the compared state-of-the-art algorithms in most of complex scenes and achieved 6% gain in FM over the second one on the whole dataset. Moreover, in order to compare proposed approach with the DBMF, which is also based on deep learning and only publish their results in several special vides. The proposed approach has also ran in these video and the results are shown in Table 2. Again, the proposed approach has noticeably better performance than the DBMF and some other classical background approaches.

As shown in the table 1 and table 2, previous deep learning based methods like the DeepBS and the DBMF achieve well performance. From our own perspective, that good performance should attribute to the stronger modeling ability and learning adaptation of CNN and FCN. However, the proposed approach focused on the pixels variation in temporal sequence rather than low-level static features such as color, edges and textures, which gives us the ability to avoid the shortcomings of the background models. Consequently, the proposed approach still get considerably better results, which over 10.46% in FM metrics compared with the DeepBS and over 6.38% compared with the DBMF.

The evaluation of proposed approach in CAMO-UOW dataset is shown in the Table 3. Unlike the CDnet dataset, the videos of CAMO-UOW dataset are specially proposed for the moving objects with camouflage, which is the main challenge of this dataset. As shown in the Table. 3 proposed approach achieves better performance compared to its competitions, with an average F-measure of 0.97, compared to values between 0.77 and 0.94 for the other methods. Therefore, it is fair to say that proposed approach performs better compared with their peers.

In this dataset, target objects have the similar color and textures with the background, which brings a lot of difficulties and obstacles to traditional methods. However, our FCN is a powerful Neural Network model which is good at capturing the non-linearities of the manifold of pixel variations.

All these experiments of the proposed method were implemented in matlab and ran on the computer with Nvidia tasela K80 GPU and all images are keep their original resolution. For each video in CDnet 2014, 100 training frames are extracted to produce the image patch. It should be noted that the 100 frames only accounts for less than 10% of total Groundtruth in CDnet 2014. In contrast, 90% of data were used as training samples in the DBMF, which suggest that the proposed approach achieves well performance with limited training frames. Considering that the videos in CAMO-UOW have fewer frames, we reduce the number of training frames to 49. During the experience, the training set and testing set are completely separated. More specifically, our FCN network is random initialized. We train the network with mini-batches of size 100, a learning rate α = 1 ∗ 10−3 over 20 epochs. The last threshold R is set to 0.6.

1. Table Type Styles
2. Sample of a Table footnote. (*Table footnote*)

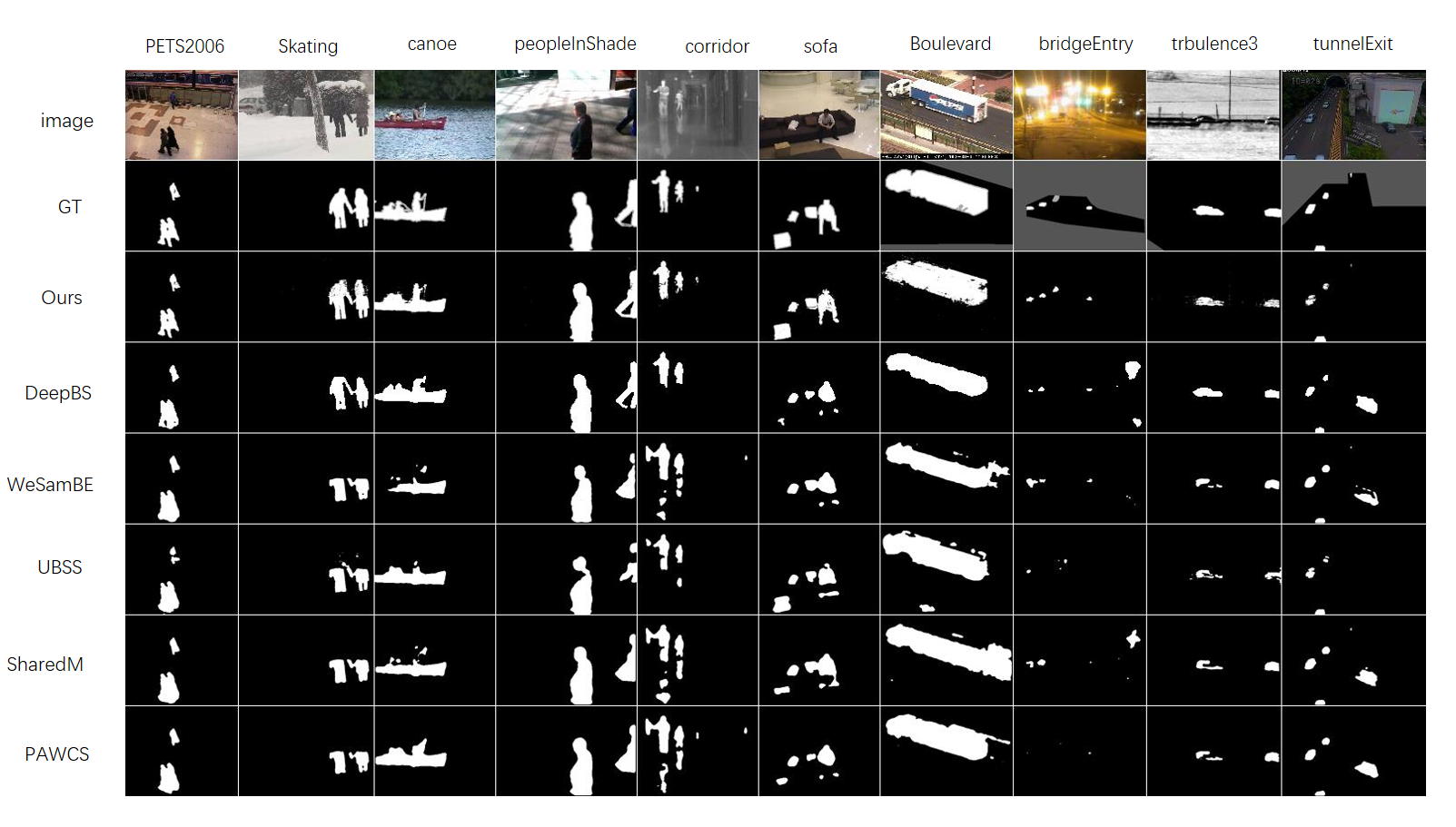
| Methods | baseline | dyna.bg | cam.jitter | int.obj.m | shadow | thermal | bad.weat | low f.rate | night vid. | PTZ | turbul. | overall |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DeepBS | 0.9580 | 0.8761 | **0.8990** | 0.6097 | 0.9304 | 0.7583 | 0.8647 | 0.5900 | 0.6359 | 0.3306 | **0.8993** | 0.7458 |
| IUTIS-5 | 0.9567 | 0.8902 | 0.8332 | 0.7296 | 0.9084 | 0.8303 | 0.8289 | **0.7911** | 0.5132 | 0.4703 | 0.8507 | 0.7717 |
| FTSG | 0.9330 | 0.8792 | 0.7513 | 0.7891 | 0.8832 | 0.7768 | 0.8228 | 0.6259 | 0.5130 | 0.3241 | 0.7127 | 0.7283 |
| AAPSA | 0.9183 | 0.6706 | 0.7207 | 0.5098 | 0.7953 | 0.7030 | 0.7742 | 0.4942 | 0.4161 | 0.3302 | 0.4643 | 0.6179 |
| CwisarDH | 0.9145 | 0.8274 | 0.7886 | 0.5753 | 0.8581 | 0.7866 | 0.6837 | 0.6406 | 0.3735 | 0.3218 | 0.7227 | 0.6812 |
| PAWCS | 0.9397 | 0.8938 | 0.8137 | 0.7764 | 0.8934 | 0.8324 | 0.8059 | 0.6433 | 0.4171 | 0.4450 | 0.7667 | 0.7403 |
| SuBSENSE | 0.9503 | 0.8177 | 0.8152 | 0.6569 | 0.8986 | 0.8171 | 0.8594 | 0.6594 | 0.4918 | 0.3894 | 0.8423 | 0.7408 |
| [SemanticBGS](http://jacarini.dinf.usherbrooke.ca/method/475/) | 0.9604 | **0.9489** | 0.8388 | 0.7878 | 0.9244 | 0.8219 | 0.8260 | 0.7888 | 0.5014 | 0.5673 | 0.6921 | 0.7892 |
| MBS | 0.9287 | 0.7915 | 0.8367 | 0.7568 | 0.8262 | 0.8194 | 0.7980 | 0.6350 | 0.5158 | 0.5520 | 0.5858 | 0.7288 |
| WeSamBE | 0.9413 | 0.7440 | 0.7976 | 0.7392 | 0.8999 | 0.7962 | 0.8608 | 0.6602 | 0.5929 | 0.3844 | 0.7737 | 0.7446 |
| ShareM | 0.9522 | 0.8222 | 0.8141 | 0.6727 | 0.8898 | 0.8319 | 0.8480 | 0.7286 | 0.5419 | 0.3860 | 0.7339 | 0.7474 |
| GMM | 0.8245 | 0.633 | 0.5969 | 0.5207 | 0.7370 | 0.6621 | 0.7380 | 0.5373 | 0.4097 | 0.1522 | 0.4663 | 0.5707 |
| RMoG | 0.7848 | 0.7352 | 0.7010 | 0.5431 | 0.7212 | 0.4788 | 0.6826 | 0.5312 | 0.4265 | 0.2470 | 0.4578 | 0.5735 |
| Ours | **0.9668** | 0.8639 | 0.8446 | **0.9181** | **0.9390** | **0.9268** | **0.9265** | 0.7338 | **0.7306** | **0.5845** | 0.8761 | **0.8504** |

1. Example of a figure caption. (*figure caption*)
2. Table Type Styles
3. Sample of a Table footnote. (*Table footnote*)

| Methods | highway | office | Pedestrians | PETS2006 | Fall | sofa | overall |
| --- | --- | --- | --- | --- | --- | --- | --- |
| GMM | 0.5788 | 0.2338 | 0.5202 | 0.6011 | 0.8026 | 0.5225 | 0.5432 |
| CodeBook | 0.8356 | 0.5939 | 0.7293 | 0.7808 | 0.3921 | 0.8149 | 0.6911 |
| ViBe | 0.7535 | 0.6676 | 0.8367 | 0.6668 | 0.6829 | 0.4298 | 0.6729 |
| PBAS | 0.8071 | 0.6839 | 0.7902 | 0.7280 | 0.3420 | 0.5768 | 0.6547 |
| [P2M](http://jacarini.dinf.usherbrooke.ca/method/475/) | 0.9160 | 0.3849 | 0.9121 | 0.7322 | 0.5819 | 0.4352 | 0.6604 |
| DBMF | 0.9412 | 0.9236 | 0.8394 | 0.9059 | 0.8203 | 0.8645 | 0.8824 |
| ours | **0.9775** | **0.9565** | **0.9652** | **0.9681** | **0.9157** | **0.8943** | **0.9462** |

1. Table Type Styles
2. Sample of a Table footnote. (*Table footnote*)

| Methods | MOG2 | FCI | LBA-SOM | PBAS | SuBSENSE | ML-BGS | DECOLOR | COROLA | FWFC | Ours |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Video 1 | 0.79 | 0.88 | 0.8 | 0.9 | 0.89 | 0.89 | 0.92 | 0.8 | **0.94** | **0.94** |
| Video 2 | 0.82 | 0.79 | 0.8 | 0.82 | 0.88 | 0.8 | 0.83 | 0.58 | 0.96 | **0.98** |
| Video 3 | 0.88 | 0.86 | 0.85 | 0.91 | 0.9 | 0.8 | 0.9 | 0.82 | **0.94** | **0.94** |
| Video 4 | 0.89 | 0.9 | 0.76 | 0.93 | 0.78 | 0.88 | 0.95 | 0.87 | 0.94 | **0.97** |
| Video 5 | 0.84 | 0.86 | 0.82 | 0.83 | 0.82 | 0.8 | 0.82 | 0.75 | 0.91 | **0.97** |
| Video 6 | 0.93 | 0.87 | 0.77 | 0.95 | 0.92 | 0.95 | **0.97** | 0.72 | 0.94 | 0.96 |
| Video 7 | 0.76 | 0.83 | 0.88 | 0.91 | 0.87 | 0.79 | 0.91 | 0.83 | 0.96 | **0.99** |
| Video 8 | 0.83 | 0.87 | 0.85 | 0.87 | 0.93 | 0.86 | 0.86 | 0.68 | 0.96 | **0.98** |
| Video 9 | 0.89 | 0.9 | 0.87 | 0.84 | 0.92 | 0.87 | 0.86 | 0.78 | 0.88 | **0.99** |
| Video 10 | 0.89 | 0.86 | 0.89 | 0.91 | 0.92 | 0.9 | 0.94 | 0.85 | 0.96 | **0.97** |
| average | 0.85 | 0.86 | 0.83 | 0.89 | 0.88 | 0.85 | 0.90 | 0.77 | 0.94 | **0.97** |



##### References

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
8. Spatiotemporal GMM for Background Subtraction with Super pixel Hierarchy Mingliang Chen
9. Region-based mixture of gaussians modelling for foreground detection in dynamic scenes Sriram Varadarajan
10. The SOBS algorithm: What are the limits?