RCDFNN: ROBUST CHANGE DETECTION BASED ON CONVOLUTIONAL FUSION NEURAL NETWORK

Chunlei Cai[1], Li Chen[1], LeiZhou[2], Xiaoyun Zhang[1], Zhiyong Gao[1]
[1]Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai, China
[2]University of shanghai for science and technology, Shanghai, China
{caichunlei, hilichen, xiaoyun.zhang, zhiyong.gao}@sjtu.edu.cn zmbhou@163.com

Introduction

- The detection of changes within video streams usually comes first in the queue of computer vision and video processing, including visual surveillance, video retrieval and smart environments.
- Although there are numerous work that performs well in some types of videos, there is no single traditional algorithm which can simultaneously address all the key challenges in real-world videos.
- A robust change detection scheme based on convolutional fusion neural network (RCDFNN) is proposed to automatically integrate the traditional methods into a robust one with improved performance.

Analysis

- Most traditional change detection algorithms depend on hand-crafted features and are designed for some types of scenes with specific challenges. Table I shows the performance of 4 recent algorithms in terms of f-measure for different challenge types.
- The results show that different algorithms have different application scenarios and some of the algorithms are complementary. For example, PWACS performs the best for dynamic background but can not handle night videos well while EFIC is just the reverse.
- We attempt to design a new framework to integrate the complementary strength of existing methods, which perform well for some scenes individually, into a robuster one based on the features extracted from video content.

Method	Dyn. bg	Night. videos	Cam. jitter	Turnu-lence	Ther-mal
PWACS	**0.894**	0.415	0.814	0.645	0.828
EFIC	0.578	**0.655**	0.713	0.671	**0.849**
SharedModel	0.822	0.542	**0.814**	0.734	0.832
WeSamBE	0.744	0.593	0.798	**0.774**	0.796

Table I. The performance of four recent change detection methods for different challenges in terms of f-measure. The challenges include dynamic background, night videos, camera jitter, turbulence and thermal. **Red** color indicates the best performance and green color indicates the worst performance.

Methods

a. Framework of the proposed CNN based change detection scheme

- There are 3 components in this framework. 1). **Basic detectors**. 2). **Feature extraction network** (**NET**^{e}). 3). **Fusion network** (**NET**^{f}).
- In the procedure of the proposed method, basic detection methods work simultaneously to generate raw results for the current frame. Meanwhile, **NET**^{e} takes adjacent frames as input to extract features of the video content. Then, the features are fed into **NET**^{f} to decide how to integrate the basic results into an optimal one. The details of the two networks are described in the next sub-sections.

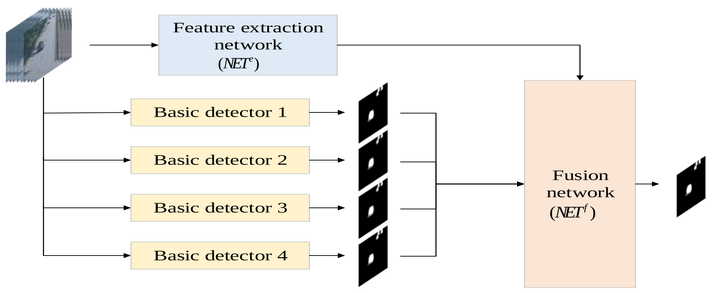

Figure 1. The framework of the proposed method.

Figure 2. The architecture of the feature extraction network.

b. Feature extraction network

- This paper utilizes CNN to extract high-level features to describe the characteristics of video frames. This paper applies a VGG-16, which is already well trained for image classification, as the feature extractor. However, VGG-16 is originally designed for processing a single image rather than video frames. In this paper we design a network for video feature extraction.

c. Fusion network

- For a given video with some types of challenges, the fusion network is designed to obtain an optimal fusion strategy to combine these raw results of basic detectors into a better one. The strategy uses a linear combination of raw results to produce an optimal one as follows. Obviously, function F for m is the key of obtaining the optimal prediction of **p**. In this paper, a CNN is trained to approximate F for its great nonlinear expression ability.

$$\mathbf{s} = \sum_{c=1}^{C} \mathbf{p}(c) \odot \mathbf{r}(c) \quad (1) \qquad \mathbf{p}(c) = \frac{e^{\mathbf{m}(c)}}{\sum_{k=1}^{C} e^{\mathbf{m}(k)}} \quad (2) \qquad \mathbf{m} = F(f;\omega), \mathbf{m} \epsilon \mathbb{R}^{H \times W \times C} \quad (3)$$

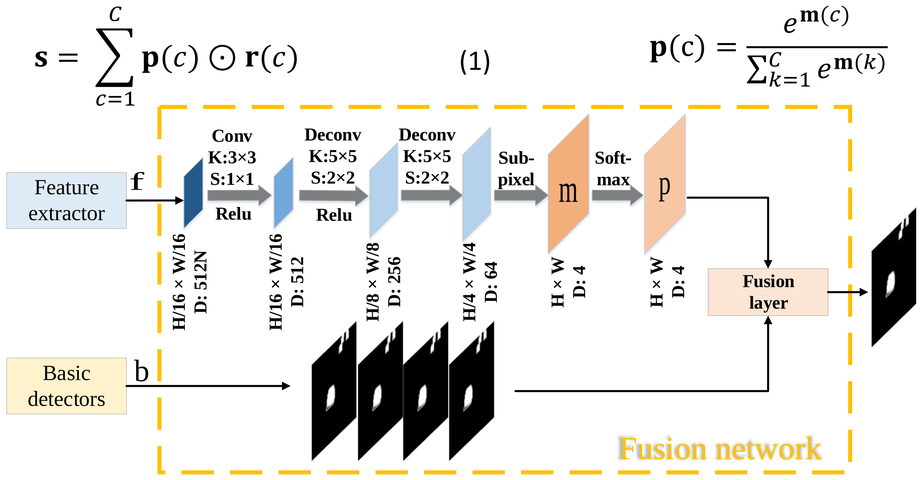

Figure 3. The architecture of the fusion network.

Training the networks

To train the networks effectively, the loss function is defined as the mean square error between the fusion result **s** *and the ground truth map* **g**.

$$L = \frac{1}{H \times W} \sum_{i=1}^{H \times W} (\mathbf{g}(i) - \mathbf{s}(i))^2 \quad (4)$$

Method	*FPR*↓	*FNR*↓	*PWC*↓	*Re*↑	*Pr*↑	*FM*↑	Average rank
RCDFNN	0.0054	**0.2432**	1.4955%	**0.7568**	0.8543	**0.8026**	**2.33**
IUTIS-5	0.0040	0.2764	**1.4951%**	0.7236	0.8832	0.7954	2.50
PWACS	0.0037	0.3009	1.5673%	0.6991	**0.8868**	0.7819	3.17
FTSG	0.0054	0.2830	1.6555%	0.7170	0.8475	0.7768	4.17
SharedModel	0.0080	0.2698	1.8105%	0.7402	0.7950	0.7666	5.17
CwisarDRP	**0.0032**	0.4124	1.9642%	0.5876	0.8848	0.7062	6.50
WeSamBE	0.0059	0.3418	1.9423%	0.6582	0.8228	0.7314	6.83
SubSENSE	0.0076	0.3026	1.9450%	0.6974	0.7935	0.7423	7.00
DeepBS	0.0060	0.4047	2.1983%	0.5953	0.8068	0.6851	8.33
EFIC	0.0134	0.3181	2.5639%	0.6819	0.6812	0.6812	9.00

Table II. The performance of the proposed RCDFNN method and 9 recent outstanding change detection methods in terms of 6 metrics and the average ranking.

Conclusion

- In this paper, an integration scheme for change detection is proposed. Although no single change detection method can address all challenges, they are complementary for different scenes. Thus the proposed method fuses the results of several existent methods into a robust one which is more adaptive to different challenges.
- The fusion weights is obtained based on the features extracted from the video content. Two CNNs are applied as the feature extractor and fusion network. The proposed networks can be trained in an end-to-end way. Comparison with recent methods shows that the proposed method has achieved state-of-the-art performance.