복잡한 영상에서 객체 분할을 위한 적대적 비지도학습

마리암 술타나*, 정순기*
*경북대학교 컴퓨터학부
e-mail : skjung@knu.ac.kr

Unsupervised Adversarial Learning for Object Segmentation in Complex Scenes

Maryam Sultana*, Soon Ki Jung*
*Dept of Computer Science and Engineering, Kyungpook National University

Abstract

The object segmentation algorithms suffer performance degradation in complex scenes because of presence of various challenges such as dynamic backgrounds, camouflage and sudden illumination changes. To handle these challenges, we present an object segmentation method, based on generative adversarial network (GAN). Our aim is to segment objects in the presence of major challenges in background scenes in real environments. Our proposed GAN is trained on background information, after that for testing the GAN has to generate the similar background information as test input via back-propagation technique. The generated background is then subtracted from the given test input to segment objects. The comparison of our proposed method with five state-of-the-art methods highlights the strength of our algorithm for object segmentation in the presence of various challenging conditions in complex scenes.

1. Introduction

The fundamental steps in many computer vision and artificial intelligence applications involves background estimation and object segmentation for the purpose of moving object detection. Object segmentation has further applications such as salient motion detection, video surveillance and visual object tracking. Background modeling is a crucial process, which describes the scene without the presence of any moving objects. However, object segmentation is a process for extracting moving objects with prior knowledge of background scene. Object segmentation is significantly affected by various challenges in background scene information, for instance, camera jitters, dynamic background, and sudden illumination variations.

2. Related Work

Many inclusive studies have been conducted to address the problem of object segmentation [1],[2]. A well-known method for object segmentation is *Gaussian Mixture Model (GMM)* [3]. GMM use probability density functions on the basis of mixture of Gaussians to model intensity variations in color at pixel level. Another very well-known technique for object segmentation is *Robust Principal Component Analysis (RPCA).* Many techniques have been proposed based on RPCA [4], [5], [6], [7], [8] for object segmentation. However, RPCA based techniques are mostly offline methods with high computational complexity and global optimization, which is a great challenge in these techniques.

3. Proposed Method

The proposed object segmentation technique called "OS_GAN", has two phases as shown in figure 1. Phase 1.) Training of the OS_GAN with background sequences estimated by Optical Flow [9], Phase 2.) Testing of OS_GAN with video sequences including moving objects.

Phase 1: OS_GAN Training

A GAN model has two neural networks, a generator **G** and discriminator **D**. The objective of generator **G** is to learn a distribution over input data $\boldsymbol{X_t}$ by mapping $\mathbf{z}$ samples through $\mathbf{G(z)}$. This mapping facilitates the **1d** vectors of input noise which is

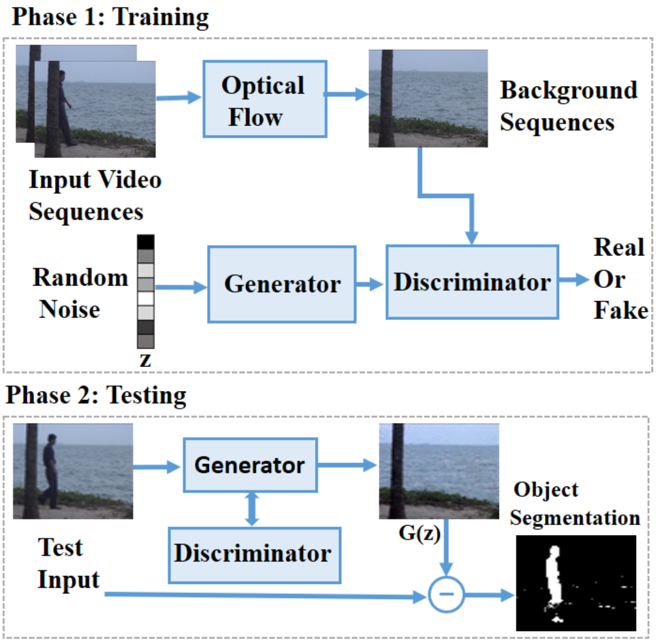

Figure 1. Workflow of OS_GAN

uniformly distributed and sampled from latent space Z to the create a 2d image representation. In a GAN architecture the discriminator, is a CNN model that maps a 2d image representation to a single value D(.). This single value D(.) is considered as a probability that whether the input given to the discriminator was a fake image generated by the generator or a real image X sampled from training data $\boldsymbol{X_t}$. The discriminator and the generator are simultaneously optimized via cross entropy loss in a following two-player minimax game:

$$\min_{G}\max_{D} \Upsilon(D,G) = \mathbb{E}_{X_t \sim \mathcal{P}_{data}(X_t)}\left[log\left(D(X_t)\right)\right] + \mathbb{E}_{z \sim \mathcal{P}_z(z)}\left[log\left(1 - D\left(G(z)\right)\right)\right].$$

Phase 2: OS_GAN Testing

Since during the phase: 1 training, the generator learns the mapping from latent space representation to more realistic images but in testing we need to perform inverse of this process. To achieve inverse mapping, a back-propagation method is applied to input data. Following are loss functions to achieve back-propagation:

Object Segmentation Loss:

$$\Gamma_o(z_\beta) = \sum |X_i - G(z_\beta)|,$$

where X_i is input test data and $G(z_\beta)$ is image generated by generator with β back-propagation steps

Background Matching Loss:

$$\Gamma_B(z_\beta) = log\left(1 - D\left(G(z)\right)\right)$$

Final Loss:

$$\Gamma(z_\beta) = (1-\alpha)\Gamma_o(z_\beta) + \alpha\Gamma_B(z_\beta)$$

The back-propagation is only applied on the coefficients of **z**, other hyper parameters of already trained OS_GAN remains unchanged during testing.

4. Experiments

We have presented the comparison of OS_GAN with 5 state-of-the-art methods for object segmentation. The state-of-the-art methods includes GRASTA [4], DECOLOR [5], 3TD [6], RMAMR [7] and TVRPCA [8] evaluated on two benchmark datasets. These datasets are Wallflower [10] and I2R [11] with challenges like dynamic background changes, illuminations conditions, and camouflage objects. We used F-score for the quantitative evaluation of the object segmentation, which is calculated as follows:

$$R = \frac{TP}{TP + FN}\ .$$

$$P = \frac{TP}{TP + FP}\ .$$

$$F = 2\frac{P \times R}{P + R}\ ,$$

where **TN** is True Negatives, **FN** is False Negatives, **TP** is True Positives, **FP** is False Positives, **R** is Recall, **P** is Precision and **F** is F-score.

It can be seen in Table 1 that our OS_GAN on average has achieved the highest F-score in both datasets and qualitative results are shown in figure 2.

Table 1. Comparison of proposed OS_GAN by using F-score on two datasets. The first highest and the second highest scores for each dataset is shown in red and blue color respectively.

Datasets/ Methods	Wallflower[10]	I2R[11]
OS_GAN	0.89	0.80
GRASTA[4]	0.33	0.54
DECOLOR[5]	0.59	0.74
3TD[6]	0.75	0.72
RMAMR[7]	0.80	0.75
TVRPCA[8]	0.61	0.69

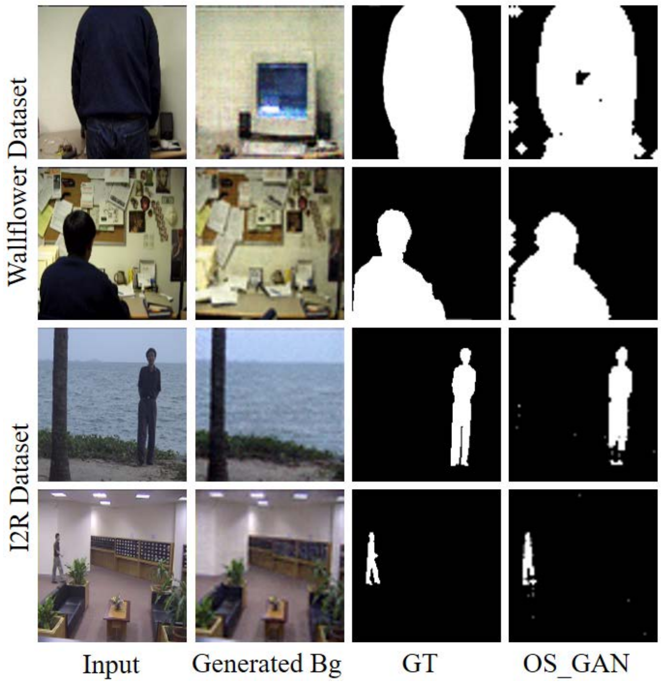

Figure 2. Performance comparison on two datasets based on backgrounds generated by OS_GAN along with object segmentation. Row: (1) video sequence "Camouflage", row: (2) video sequence "Foreground Aperture", from Wallflower dataset. Row: (3) video sequence "WaterSurface", row: (4) video sequence "Lobby" both from I2R dataset.

5. Conclusion

In this study, we present the object segmentation based on Generative Adversarial Network (GAN). Our goal is to segment objects in the presence of major challenges in complex background scenes. Our presented solution based on GAN works on the principle of generating background image samples similar to test data. The comparison of our proposed method with five state-of-the-art methods highlights the strength of our algorithm for object segmentation in the presence of challenging illumination conditions and dynamic background scenario.

6. Acknowledgements

This research was supported by Development project of leading technology for future vehicle of the business of Daegu metropolitan city (No. 20171105).

References

[1] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung, “Unsupervised deep context prediction for background estimation and foreground segmentation,” Machine Vision and Applications, pp. 1-21, 2018.

[2] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, “Spatiotemporal low-rank modeling for complex scene background initialization,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 6, pp. 1315-1329, 2018.

[3] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., vol. 2. IEEE, 1999, pp. 246-252.

[4] J. He, L. Balzano, and A. Szlam, “Incremental gradient on the grassmannian for online foreground and background separation in subsampled video,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 1568-1575

[5] X. Zhou, C. Yang, and W. Yu, “Moving object detection by detecting contiguous outliers in the low-rank representation,” IEEE T-PAMI, vol. 35, no. 3, pp. 597-610, 2013.

[6] O. Oreifej, X. Li, and M. Shah, “Simultaneous video stabilization and moving object detection in turbulence,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 2, pp. 450-462, 2013.

[7] X. Ye, J. Yang, X. Sun, K. Li, C. Hou, and Y. Wang, “Foreground-background separation from video clips via motion-assisted matrix restoration,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 11, pp. 1721-1734, 2015.

[8] X. Cao, L. Yang, and X. Guo, “Total variation regularized rpca for irregularly moving object detection under dynamic background,” IEEE Transactions on cybernetics, vol. 46, no. 4, pp. 1014-1027, 2016.

[9] C. Liu et al., “Beyond pixels: exploring new representations and applications for motion analysis,” Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[10] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: Principles and practice of background maintenance,” in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 1. IEEE, 1999, pp. 255-261.

[11] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, “Statistical modeling of complex backgrounds for foreground object detection,” IEEE Transactions on Image Processing, vol. 13, no. 11, pp. 1459-1472, 2004.