# Shadow Detection from surveillance videos

Prof. Manish Khare
*Dhirubhai Ambani Institute of Information and Communication Technology*
Gandhinagar, Gujarat, India
manish_khare@daiict.ac.in

Nishtha Chaudhary
*Dhirubhai Ambani Institute of Information and Communication Technology*
Gandhinagar, Gujarat, India
nishthakc@gmail.com

Khushali Shah
*Dhirubhai Ambani Institute of Information and Communication Technology*
Gandhinagar, Gujarat, India
khushali9930@gmail.com

***Abstract*—Shadow detection and removal are fundamental and summoning tasks for scene understanding. This paper uses Effective-Context Augmentation to learn about neighborhood contexts for the robust detection of shadows from the keyframes extracted from videos. Further, we propose an encoder-decoder type of shadow detection method that acts as the principal building block for the extraction of the strong feature representations for the encoder and the process of classification for the decoder. Additionally, the networks are optimized for fast and easy training with only one loss. The code is available: https://github.com/khushali77/ShadowDetection**

## I. Introduction

A shadow is a dark area that appears on a surface when light from a source (or multiple sources) is hindered by a non-transparent object. There are two types of shadows: (1)self-shadow, an object cast a shadow on itself, and (2) cast shadow, an object cast a shadow on another surface. [1] Although shadows can provide valuable cues about the physical properties of an object, obtaining illumination conditions [2], finding the geometry of the object casting the shadow [3] and many more, and many more, their presence may cause hindrance in various video keyframe and video processing tasks, such as video keyframe segmentation, object detection, and tracking, and video surveillance. Hence, shadow detection and removal have become an essential task for improving the quality of the scene. Furthermore, for the flawless execution of the tasks mentioned above. Shadow detection and removal is mainly incorporated into the preprocessing stage of these applications, because they are aimed at enhancing an video keyframe or video to make it suitable for a computer vision task.

Many early works in shadow detection and removal have been developed in order to analyze the demoraphy of the color and illumination by implementing physical models with traditional [4] or hand-crafted [5] features, or by learning discriminative features via neural network (CNN) [6], [7].

## II. Related work

Related work can be divided into traditional methods or deep learning methods conforming to whether they adopt the conventional feature-based or deep network-based ideas.

Subsequently, handcrafted features were adopted based on the information from the color [8], edge [9], texture [10], etc. In contrast, the usual methods were based on region classification [11]. Deep learning-based approaches have been popular lately due to their high performance from feature extraction via deep networks. The compelling multi-scale features for robust shadow detections [7], [12] can be extracted via CNN. For example, patch-based CNN was adopted by Vicente et al. [13] for the detection of shadows via video keyframe-level prior and video keyframe patches. Those rich shadow characteristics were not captured via CNN-based frameworks.

Few authors [14] outstretched CNN by scouting more contextual cues, whereas few of them [15] made use of generative adversarial networks (GAN) for working around with the finite training data [16] and to improve the power of discrimination. Further, Zheng et al. [17], proposed another idea in which he considered distraction areas to be the fake detection regions and coalesced several existing results with the ground truths to obtain labels for robust discrimination. Hu et al. [18] presented the detail enhancement module (DEM) for complex shadows. The existing shadow detection methods based on deep learning uses the regular features from the layer-by-layer convolution, that can be erratic, taking into consideration the extensive losses while convolution.

Few [19], [20] anticipate on various restrictions might require additional computational resources by computing numerous loss functions and also, due to imprecise intermediate features. However, our model includes ECA that boosts the discriminative multi-scale features. It can augment effective object detection contexts compared to the prevailing methods. Furthermore, it is easy to train with a less computational load, leading to improved stability since it uses only one loss function.

Aforementioned structure cannot discover a compelling context for object discrimination that is precisely the aim of ECA. Thus, it only needs to compute the convolutions once. While, other combines them via the deep global features in order to boost the effective object contexts.

## III. Proposed Methodology

Our proposed method consists of a framework having two steps as shown in Fig. 1. First, key frame extraction from videos via HDBscan [21] clustering. Second, shadow detection from the video keyframes obtained via keyframe extraction using ECA-N framework.

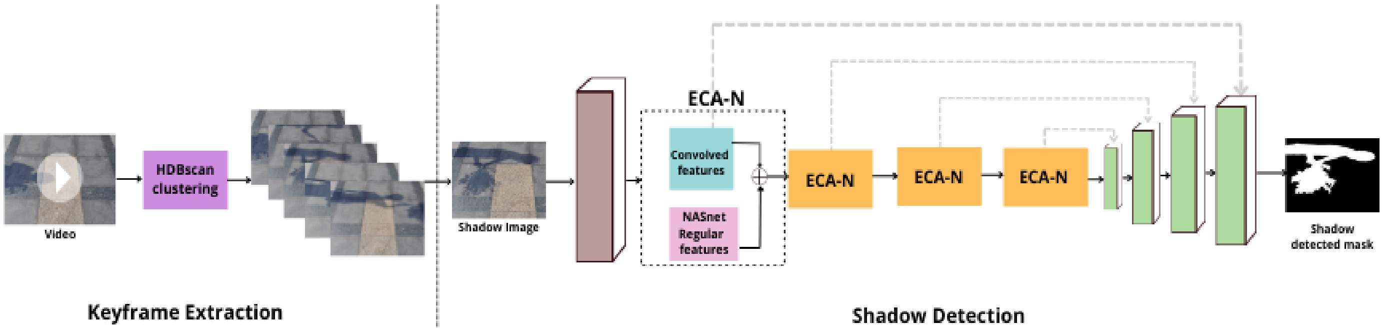

Fig. 1. The architecture of the proposed method

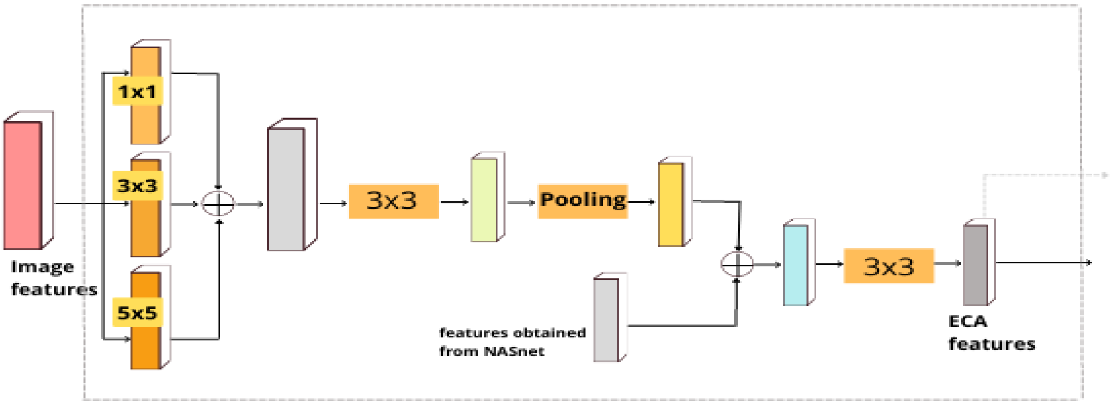

Fig. 2. The structure of ECA-N Module

### A. *Keyframe extraction*

The proposed keyframe extraction algorithm initially extracts candidate frames from videos using the local maxima information that provides us with images having strongest change from its vicinity of frames, where vicinity is defined using window length. Now, the HDBscan clustering algorithm is being used to obtain clusters from the histograms of the previously obtained candidate frames, which represents the global representation of the image appearance after which the keyframes are obtained via selecting images from each clusters having low blurr(high laplacian) score (eq.(1)).

$$Laplacian(C) = \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \tag{1}$$

### B. *Shadow Detection*

The shadow detection part consists of an encoder-decoder-based framework for shadow detection, which aims at discovering an effective object context from keyframes extracted in the previous step.

We know that the humans take the neighbouring distributions as allusions to affirm whether the dark areas are shadows or objects. Thus, the information of object contexts can be attained via Effective Context Augmentation (ECA) [22] module by taking the global cues as the deep features. As shown in Fig. 2, the convolved features are obtained by performing convolution of the input feature with $1 \times 1$, $3 \times 3$ and $5 \times 5$ convolutions concurrently. Hence, there will be three parallel channels and, therefore, a preliminary fusion is then required to reduce parameters. Additionally, the effective context boosting was carried out via the concatenation of the max pooled features; obtained via pooling the convolved features, with the regular deep features from ResNet-101 [23]. In order to improve the accuracy of the detected shadow mask, we use NASnet [24] to provide general object information (e.g., position, silhouettes) as global guidance to augment object discrimination with effective contexts. Using NASnet turned out to be better than ResNet as the SDR improved by 1.4% which is shown in Table I.

$$F_{ECA-N}(V) = f_{(3,3)}(C(P(F_{fuse}(V)), F_{NAS})) \tag{2}$$

where $F_{ECA-N}$ are the ECA-N features, $F_{fuse}$ are the features obtained via preliminary fusion and $F_{NAS}$ are the

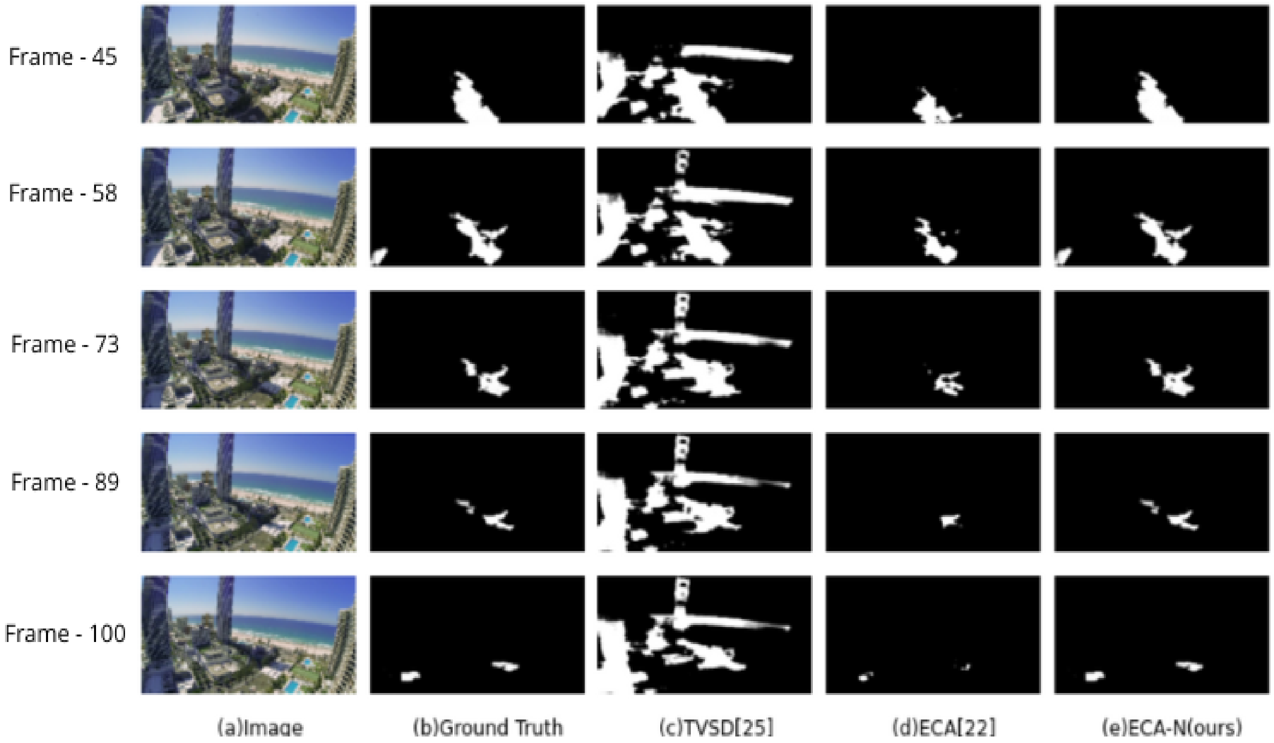

Fig. 3. Shadow detection results

regular features obtained from NASnet [24].

In case of the complete framework used, the encoder section processes through layer-by-layer processing with the help of ECA-N module in each layer to obtain its representation in the reduced form. After which, the decoder is put in to map the scale-by-scale representation for the generation of its eventual shadow distribution. In each layer of the encoder, discriminative contexts features are obtained using ECA-N module. Then the next ECA-N layer takes this feature as an input to continue this abstraction process for discriminative contexts. While decoding, effective-contexts guided generation can be obtained in the decoder where the ECA-N features (eq.(2)) are combined again with the feature map. Finally, a convolution layer of size 1 × 1 is being applied and the sigmoid function is used for thresholding to get the final predicted output.

## IV. Training and Testing

Our current method is implemented via Tensorflow and PyTorch. In order to supply the regular features, NASnet is pre-trained by video keyframeNet.

### A. Loss functions

We acquire the notion of binary cross-entropy and take into consideration a weighted cross-entropy in order to stabilize the contributions to detect the shadow from positive and negative samples which is as follow:

$$L = -\sum_{i}(\lambda y_i^{lab} \log y_i^{pre} + (1-\lambda)(1-y_i^{lab}) \log(1-y_i^{pre})) \tag{3}$$

where $y_i^{lab}$ and $y_i^{pre}$ indicates the ground truth and the model prediction of the i-th pixel's shadow class respectively; whereas $\lambda$ represents the weight which is set to be 0.65.

### B. Dataset Details

Training is done with a single Cuda GPU 11.0. ViSha [25] and SBU Timelapsed [26] datasets are taken as the training sets. We have resized all the samples' size as 256 × 256.

## V. Experimental Results

We considered our experimental comaprisons via SBU Timelapse and ViSha datasets. The number of video keyframes were 50 videos and 120 videos respectively which were further divided into training and testing video keyframes.

Several state-of-the-art shadow detection and removal methods are considered for the performance comparison, for instance, TVSD [25] and ECA [22].

### A. Qualitative Results

The detection results of keyframes obtained from videos can be observed in Fig. 3 among the existing methods and ours. It can be said that our method is immune to the dark surfaces and good at the light shaded shadows and obtains better performance as compared to other methods.

### B. Quantitative Results

Considering, BER, Balance Error Rate is taken into consideration to assess the ability in obtaining balanced results,

$$BER = 1 - \frac{1}{2}(\frac{T_P}{T_P + F_N} + \frac{T_N}{T_N + F_P}) \qquad (4)$$

while SDR, Shadow Detection Rate is used to evaluate shadow occupied regions via:

$$SDR = \frac{T_P}{T_P + F_N} \qquad (5)$$

where $T_P$, $F_P$ , $T_N$ and $F_N$ represents the number of true positives, false positives, true negatives and false negatives, respectively. Lower the BER is, better the performance is. Whereas more the SDR, the better the performance. Table I shows the statistical comparison results in terms of SDR. For SBU and Visha datset, ECA-N outperforms with the high SDR of 97.26% and 95.01%, respectively. Whereas cross comaparision of BER scores is shown in Table II, in which our framework obtains 7.82% and 8.9% lower in BER, for SBU and Visha dataset respectively, than TVSD algorithm.

TABLE I
CROSS COMPARISON OF SDR SCORE FOR SHADOW DETECTION

| **Training** | **SBU Timelapse** | | **ViSha** | |
|---|---|---|---|---|
| **Testing** | **SBU Timelapse** | **ViSha** | **SBU Timelapse** | **ViSha** |
| TVSD [25] | 96.91 | 87.53 | 83.48 | 95.14 |
| ECA [22] | 97.06 | 94.64 | 83.25 | 91.61 |
| **ECA-N(Ours)** | **97.26** | **94.67** | **84.74** | **95.01** |

TABLE II
CROSS COMPARISON OF BER SCORE FOR SHADOW DETECTION

| **Training** | **SBU Timelapse** | | **ViSha** | |
|---|---|---|---|---|
| **Testing** | **SBU Timelapse** | **ViSha** | **SBU Timelapse** | **ViSha** |
| TVSD [25] | 12.35 | 23.77 | 37.23 | 20.47 |
| ECA [22] | 2.94 | 5.36 | 16.75 | 8.39 |
| **ECA-N(Ours)** | **2.71** | **5.33** | **15.26** | **4.99** |

## VI. CONCLUSION

In order to efficiently conduct the task of shadow detection, keyframe extraction from videos using HDBscan algorithm is being carried out. ECA-N(effective context augmentation - NASnet) module was exploited in this paper, which combines the discriminative features obtained, with the regular deep features and also for effective object detection, augments the appropriate object contexts.

## ACKNOWLEDGMENT

We would like to express our thanks of gratitude to Prof. Manish Khare for giving us the opportunity to do this wonderful project on the topic ”Shadow Detection and Removal from Surveillance videos”.

## REFERENCES

[1] Murali, S., Govindan, V. K., and Kalady, S. (2018). ”A survey on shadow detection techniques in a single video keyframe. Information Technology And Control”, 47(1). https://doi.org/10.5755/j01.itc.47.1.15012

[2] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, “Simultaneous cast shadows, illumination and geometry inference using hypergraphs,” IEEE PAMI, 2013.

[3] I. Okabe T Sato and Y. Sato, “Attached shadow coding: estimating surface normals from shadows under unknown reflectance and lighting conditions,” in Proc. ECCV, 2009.

[4] F. Liu and M. Gleicher, “Texture-consistent shadow removal,” in European Conference on Computer Vision, 2008, pp. 437–450.

[5] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. 2018. ”Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection”. In ECCV. 122–137. https://doi.org/10.1007/978-3-030-01231-1_8

[6] Tomás F. Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. 2016. ”Large-Scale Training of Shadow Detectors with Noisily-Annotated Shadow Examples”. In ECCV. 816–832. https://doi.org/10.1007/978-3-319-46466-4_49

[7] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. 2014. ”Automatic Feature Learning for Robust Shadow Detection”. In CVPR. 1939–1946. https://doi.org/10.1109/cvpr.2014.249

[8] G.D. Finlayson, S.D. Hordley, Cheng Lu, and M.S. Drew. 2006. ”On the removal of shadows from video keyframes.” IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 1 (2006), 59–68. https://doi.org/10.1109/tpami.2006.18

[9] Xiang Huang, Gang Hua, Jack Tumblin, and Lance Williams. 2011. ”What characterizes a shadow boundary under the sun and sky?” In ICCV. 898–905. https://doi.org/10.1109/iccv.2011.6126331

[10] Jiejie Zhu, Kegan G. G. Samuel, Syed Z. Masood, and Marshall F. Tappen. 2010. ”Learning to recognize shadows in monochromatic natural video keyframes”. In CVPR. 223–230. https://doi.org/10.1109/cvpr.2010.5540209

[11] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. 2011. ”Single-video keyframe shadow detection and removal using paired regions”. In CVPR. 2033–2040. https://doi.org/10.1109/cvpr.2011.5995725

[12] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. ”Fully convolutional networks for semantic segmentation”. In CVPR. 640–651. https://doi.org/10.1109/cvpr.2015.7298965

[13] Tomás F. Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. 2016. ”Large-Scale Training of Shadow Detectors with Noisily-Annotated Shadow Examples”. In ECCV. 816–832. https://doi.org/10.1007/978-3-319-46466-4_49

[14] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. 2018. ”Direction-Aware Spatial Context Features for Shadow Detection”. In CVPR. 2795–2808. https://doi.org/10.1109/cvpr.2018.00778

[15] Hieu Le, Tomas F. Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. 2018. ”A+D Net: Training a Shadow Detector with Adversarial Shadow Attenuation”. In ECCV. 680–696. https://doi.org/10.1007/978-3-030-01216-8_41

[16] Vu Nguyen, Tomas F. Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. 2017. ”Shadow Detection with Conditional Generative Adversarial Networks”. In ICCV. 4510–4518. https://doi.org/10.1109/iccv.2017.483

[17] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. 2019. ”Distraction-Aware Shadow Detection”. In CVPR. 5167–5176. https://doi.org/ 10.1109/cvpr.2019.00531

[18] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. 2021. ”Revisiting shadow detection: A new benchmark dataset for complex world”. IEEE Transactions on video keyframe Processing 30 (2021), 1925–1934

[19] Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, and Huchuan Lu. 2017. ”A Stagewise Refinement Model for Detecting Salient Objects in video keyframes”. In ICCV. 4039–4048. https://doi.org/10.1109/iccv.2017.433

[20] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. 2019. ”Distraction-Aware Shadow Detection”. In CVPR. 5167–5176. https://doi.org/10.1109/cvpr.2019.00531

[21] R. Campello, D. Moulavi, and J. Sander, ”Density-Based Clustering Based on Hierarchical Density Estimates” In: Advances in Knowledge Discovery and Data Mining, Springer, pp 160-172. 2013

[22] Fang, Xianyong, et al. ”Robust Shadow Detection by Exploring Effective Shadow Contexts.” Proceedings of the 29th ACM International Conference on Multimedia. 2021.

[23] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, and Kaiming He. 2017. ”Aggregated Residual Transformations for Deep Neural Networks”. In CVPR. 5987– 5995. https://doi.org/10.1109/cvpr.2017.634

[24] Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2017). ”Learning Transferable Architectures for Scalable video keyframe Recognition”. In arXiv [cs.CV]. http://arxiv.org/abs/1707.07012

[25] Chen, Z., Wan, L., Zhu, L., Shen, J., Fu, H., Liu, W., and Qin, J. (2021). ”Triple-cooperative video shadow detection”. ArXiv [Cs.CV]. https://doi.org/10.48550/ARXIV.2103.06533

[26] Le, H., and Samaras, D. (2021). ”Physics-based shadow video keyframe decomposition for shadow removal”. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP, 1–1. https://doi.org/10.1109/TPAMI.2021.3124934