ANOMALOUS MOTION DETECTION ON HIGHWAY USING DEEP LEARNING

Harpreet Singh    Emily M. Hand    Kostas Alexis

University of Nevada, Reno, USA
Department of Computer Science and Engineering

ABSTRACT
Research in visual anomaly detection draws much interest due to its applications in surveillance. Common datasets for evaluation are constructed using a stationary camera overlooking a region of interest. Previous research has shown promising results in detecting spatial as well as temporal anomalies in these settings. The advent of self-driving cars provides an opportunity to apply visual anomaly detection in a more dynamic application yet no dataset exists in this type of environment. This paper presents a new anomaly detection dataset – the Highway Traffic Anomaly (HTA) dataset – for the problem of detecting anomalous traffic patterns from dash cam videos of vehicles on highways. We evaluate state-of-the-art deep learning anomaly detection models and propose novel variations to these methods. Our results show that state-of-the-art models built for settings with a stationary camera do not translate well to a more dynamic environment. The proposed variations to these SoTA methods show promising results on the new HTA dataset.

Index Terms— anomaly detection, deep learning, one-class classification

1. INTRODUCTION
Anomaly detection is an unsupervised one-class classification problem with the goal of learning the normal state of data during training and then detecting aberrations without any provided labels. Many anomaly detection applications involve visual data, such as images or videos, and are motivated by interest in surveillance. Datasets commonly evaluated for visual anomaly detection are constructed with a stationary camera observing a region in which the background environment is relatively static while foreground objects such as pedestrians and vehicles are in motion. Anomalies are defined as appearance or motion deviations from normal data in the foreground objects, shown in Figure 1. [1] provides a comprehensive list of video anomaly detection datasets, but they all maintain the same characteristics. As autonomous robots become increasingly common, there is a need to perform anomaly detection while an agent is moving within an environment yet no such dataset exists.

This paper presents a new anomaly detection dataset, the Highway Traffic Anomaly (HTA) dataset, to evaluate anomaly detection methods with a moving agent. Specifically, the dataset consists of dash cam videos captured from vehicles driving on the highway. The goal is to learn normal driving motion of traffic in the camera’s field of view and then detect conditions in which other vehicles are moving abnormally. In this dataset, not only is the camera moving but other vehicles and background features are also in motion relative to the camera. The dataset consists of five types of anomalies: speeding vehicle, speeding motorcycle, vehicle accident, close merging vehicle and halted vehicle. Three state-of-the-art deep learning based anomaly detection models are evaluated and two variations, specifically for the problem of detecting anomalous highway traffic motion, are proposed. Code for the HTA dataset and the evaluated models is available at [2].

2. RELATED WORK
2.1. Anomaly Detection Datasets
Frequently evaluated visual anomaly datasets consist of a static background, moving foreground objects, and a stationary camera. The UCSD Pedestrian dataset [3] consists of videos of pedestrians on a university campus. Normal data corresponds to pedestrians walking on pathways. Abnormal data consists of small vehicles or cyclists that are different in appearance and motion from usual pedestrian traffic. The CUHK [4] Avenue dataset also consists of videos of a
crowded walkway with anomalies defined as unusual behavior, such as throwing an object. Similarly, the UMN Unusual Crowd Activity dataset [5] consists of videos of unusual behavior in crowds. While not exhaustive, this list summarizes the typical visual anomaly datasets previously studied. To the best of our knowledge, there is no open-source autonomous driving dataset specifically for the task of anomaly detection.

2.2. Anomaly Detection Methods

Initial work in anomaly detection relied on hand-engineered features (e.g., HOG) to create meaningful representations of normal image data. The features were then used to fit a statistical model such as an SVDD [6]. Deep learning models have recently shown promising results in anomaly detection. The deep learning models evaluated in this study can be grouped using the taxonomy proposed in [7]: generative and predictive learning models.

Generative models attempt to estimate the probability distribution of the training data (i.e., the normal motion data). After training, test videos are evaluated by some form of a reconstruction error. [8] proposes a method to use adversarial autoencoders for anomaly detection by relying on the generator’s ability to model only the normal data distribution. More recently, a conditional generative adversarial network (CGAN) is used by [9]. The authors train two independent CGANs to learn an image-to-flow and flow-to-image transformation. We evaluate this model on the HTA dataset, providing more details in section 3. Predicting Optical flow with CNNs for visual understanding has been shown to be effective, such as in [10] for action classification.

Predictive networks seek to model sequential data. This approach is effective in learning temporal patterns in videos; given a sequence of \( N \) video frames, predict the \( N + 1 \) frame. Anomaly detection is performed by first training only with normal data to accurately predict images from normal motion sequences. After training, predicting a future frame of an abnormal sequence will result in a larger error. Various architectures have been proposed [11] [12] [13], they all share the same anomaly detection mechanism.

3. METHOD

3.1. Highway Traffic Anomaly Dataset

The HTA dataset was curated from the Berkeley DeepDrive dataset [14] that consists of 100k high resolution (1280 × 720, 30 FPS) dash cam videos collected from cars in New York and the Bay Area. We sifted through the entire dataset, selecting only highway driving videos. From that subset, videos in visually degraded conditions were removed. In summary, the HTA dataset consists of highway videos: during clear lighting conditions; clear, partly cloudy, or overcast weather conditions; minimally occluded from large vehicles; and contain at least some traffic in motion. Due to the imperfect nature of the data collection process, the HTA dataset is not without noise. For instance, bumps and cracks on the highway cause transient shaking in the videos. These characteristics make the HTA dataset more challenging and realistic.

Normal driving conditions in this dataset are defined by the motion of vehicles that does not perturb the motion of the dash cam vehicle or the motion of other vehicles stays relatively self-similar. The training set consists of 286 videos of normal traffic conditions, a total of 322,202 video frames and an average duration of 40 seconds. Figure 2 shows a sample of video frames from the training set.

The test set contains a total of 103 videos, 78 normal traffic videos not in the training set and 25 abnormal traffic videos. Abnormal traffic videos consist of five types of anomalous motion: speeding vehicle (4 videos), close merge (13 videos), halted vehicle (1 video), vehicle accident (5 videos), and speeding motorcycle (2 videos). Each case represents a situation in which a human driver will practice caution. Vehicle accident anomalies were downloaded from YouTube. Example image sequences of each abnormality are shown in Figure 3. Abnormal motion is manually annotated at the frame level since only short sequences contain abnormal motion. A frame was labelled anomalous if the motion from the previous frame to the current was part of an anomalous motion. There are a total of 1531 frames labeled as anomalous, making up 6% of the abnormal test set.

3.2. Generative Models

Since an anomaly is defined as irregular motion, generative models can learn to predict dense optical flow to model normal motion. For training, ground truth optical flow is computed using OpenCV’s dense optical flow implementation. The first generative model evaluated is the conditional GAN (CGAN) proposed in [9]. Figure 4. The CGAN is trained to predict the optical flow between a pair of sequential frames. The generator’s input are two RGB images, concatenated depth-wise and it predicts the corresponding optical flow. The discriminator classifies patches in the input as real or fake, producing a 2D-one channel output in the
Fig. 3. Examples of abnormal images from the training set. Top to bottom row: speeding motorcycle, halted vehicle, close merge, speeding vehicle. Vehicle accident not shown but YouTube links will be provided.

Fig. 4. The CGAN’s generator learns an image-to-optical flow transformation. The discriminator processes the input and either the ground truth or generator’s output concatenated and classify it as real or fake.

Due to the high capacity of GANs, they may estimate anomalous data as well as normal data [16]. In order to evaluate the effectiveness of the CGAN, we also evaluate the state-of-the-art deep learning dense optical flow prediction model: FlowNet [17]. Unlike the CGAN, FlowNet is not trained adversarially.

3.3. Predictive Model

Predictive models can learn sequential data such as videos. By training on sequences of \( N \) normal traffic frames, a look back of \( N \), a predictive model will then predict the \( N + 1 \) frame. The Predictive Coding Network (PredNet) proposed in [18] showed promising results in predicting future frames in the KITTI dataset. PredNet, shown in Figure 5, is constructed with a series of modules that perform local predictions and then propagates only the errors between the predicted and actual to the subsequent layers. The evaluated model was constructed using the recommendations in [18]. Anomalies are detected using the pixel-level reconstruction error between the predicted \((N + 1)\)th frame and the ground truth frame, employing the averaging sliding window approach to compute pixel-level difference.

3.4. Proposed Variations

The first modification we propose is related to the detection mechanism that relies on the reconstruction error. The pixel-level reconstruction error averaged with a sliding window can be improved by using a variable window size. Since vehicles in the center of the image tend to be further from the camera, anomalous motion near the center of the image will displace a small number of pixels. In order to account for depth, we propose to use a smaller 3x3 averaging window size near the center and a larger 9x9 averaging window size around the edges.

Another characteristic, specific to highway anomaly detection, is the motion of the vehicles is relative due to ego-motion; vehicles moving at speeds similar to the camera will seem to displace very few pixels, if any. When using the PredNet model to predict future frames, rather than just one frame into the future, the model can predict the 6th future frame instead. [18] tests the PredNet’s capabilities of predicting frames further into the future by using PredNet’s first predicted frame as input for the next prediction. Five frames further into the future are shown to maintain reasonable accuracy. We evaluated this method using \( N = 1, 2, 3, \ldots \) and found that \( N = 6 \) gave the best results.

4. EXPERIMENTS

A total of four models are evaluated with the HTA dataset: CGAN, FlowNet, PredNet \( N + 1 \), and PredNet \( N + 6 \). The performance of each model is measured using an AUC score.
Table 1. AUC scores computed from the all the evaluated models. PredNet(N+6) performs the best across the board.

| Anomaly Type       | CGAN  | FlowNet | PredNet(N+1) | PredNet(N+6) |
|--------------------|-------|---------|--------------|--------------|
| Speeding Vehicle   | 0.608 | 0.623   | 0.497        | 0.614        |
| Accident           | 0.607 | 0.657   | 0.559        | 0.601        |
| Speeding Motorcycle| 0.580 | 0.593   | 0.581        | 0.828        |
| Close Merge        | 0.422 | 0.531   | 0.619        | 0.643        |
| Halted Vehicle     | 0.337 | 0.216   | 0.554        | 0.236        |

Fig. 6. The CGAN discriminators output (right) is unable to identify the anomalous motion of the motorcycle.

The thresholds in each experiment range from 0.002 to 1.0 in increments of 0.002. Five AUC scores are reported for each model, one for each type of abnormal motion, Table I.

4.1. Conditional GAN

The CGAN is trained on pairs of RGB images. Input images are cropped from the bottom to remove the visible hood of the vehicle as well as from the top to reduce the amount of sky/background in each frame. The final input size is 128 \( \times \) 512 \( \times \) 6. The model is trained for 40 epochs.

The first approach to detect anomalies is by using the discriminator’s 2D - one channel output as a heat map for patches containing an anomaly. Figure 6 shows promising results using this approach but the discriminator’s output trained on the HTA dataset does not produce the same results, Figure 6. The salient characteristic that may cause this discrepancy in the results is that, unlike the UCSD Pedestrian dataset, the HTA dataset does not maintain a static background.

The second approach to detect anomalies with the CGAN is to use the reconstruction error from the generator’s output. The AUC scores of the CGAN in Table I show that all anomaly types are near 0.5, meaning that the reconstruction error from generator’s output has minimal discriminative capabilities in classifying abnormal motion.

4.2. FlowNet

FlowNet is also trained on pairs of RGB images from the training dataset. The input RGB images for FlowNet are cropped from the bottom, making the input size 2 \( \times \) 256 \( \times \) 512 \( \times \) 3. The model is trained for 100 epochs. Anomalies are detected in the test set using the same approach as CGAN, the reconstruction error between the predicted and ground truth optical flow. The AUC scores for the FlowNet are only slightly better than the CGAN results, Table I.

Both generative models estimate optical flow of abnormal motion just as well as normal motion; AUC scores of both models are near 0.5, indicating that they are unable to discern abnormal motion from normal motion. It seems that both models learn to estimate optical flow better than learning distinguishing patterns in normal motion.

4.3. Predictive Coding Network

The PredNet model is trained on a look back of 10 consecutive RGB images as suggested by the original work. As with the CGAN, each image is cropped from the bottom as well as the top with the final input size 64 \( \times \) 256 \( \times \) 3 \( \times \) 10. The model was trained for 100 epochs. Anomalies are detected using the reconstruction error between the model’s predicted future frame and the ground truth frame. Using this mechanism, AUC scores for PredNet \( N + 1 \) are provided in Table I. The AUC scores of PredNet \( N + 1 \) show that close merge and speeding motorcycle anomalies perform relatively better than others and achieves the highest AUC score for the halted vehicle anomaly albeit still near 0.5.

By setting a look back of two, we extrapolate four frames into the future, testing the sixth frame for anomalous motion, PredNet \( N + 6 \). Figure 7 shows a sample output and ground truth frame. Anomalies are detected in the same manner as with PredNet \( N + 1 \). AUC scores for PredNet \( N + 6 \) are shown in Table I and show significant improvement in detecting the speeding motorcycle anomaly. PredNet \( N + 6 \) achieves the highest or close to the highest AUC score for each anomaly type except for halted vehicle.

5. CONCLUSION

This paper presents a new anomaly detection dataset, the Highway Traffic Anomaly (HTA) dataset. It differs from existing anomaly detection datasets in many ways that make it more challenging. To the best of our knowledge, this is the first anomaly detection dataset for autonomous driving. Four state-of-the-art deep learning models were evaluated with a proposed heuristic to improve the reconstruction error for anomaly detection tailored for the HTA dataset. The results indicate that state-of-the-art models do not perform well on the HTA dataset. Our proposed variation of the PredNet model to predict the sixth future frame shows promising results on the speeding motorcycle anomaly and relatively better in all anomalies.
6. REFERENCES

[1] N. Patil and Prabir Kumar Biswas, “A survey of video datasets for anomaly detection in automated surveillance,” 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), pp. 43–48, 2016.

[2] Harpreet Singh, “Highway traffic anomaly repository,” [https://github.com/harpreets652/highway-traffic-anomaly](https://github.com/harpreets652/highway-traffic-anomaly), 2019.

[3] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos, “Anomaly detection in crowded scenes,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010, pp. 1975–1981.

[4] Cewu Lu, Jianping Shi, and Jiaya Jia, “Abnormal event detection at 150 fps in matlab,” in Proceedings of the IEEE international conference on computer vision, 2013, pp. 2720–2727.

[5] Nikos Papanikolopoulos, “Detection of unusual crowd activity,” Unusual crowd activity dataset.

[6] David MJ Tax and Robert PW Duin, “Support vector data description,” Machine learning, vol. 54, no. 1, pp. 45–66, 2004.

[7] B Kiran, Dilip Thomas, and Ranjith Parakkal, “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos,” Journal of Imaging, vol. 4, no. 2, pp. 36, 2018.

[8] Asimenia Dimokranitou, Adversarial autoencoders for anomalous event detection in images, Ph.D. thesis, 2017.

[9] Mahdyar Ravanbakhsh, Enver Sangineto, Moin Nabi, and Nicu Sebe, “Training adversarial discriminators for cross-channel abnormal event detection in crowds,” in 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019, pp. 1896–1904.

[10] Pierre-Etienne Martin, Jenny Benois-Pineau, Renaud Pêteri, and Julien Morlier, “Optimal choice of motion estimation methods for fine-grained action classification with 3d convolutional networks,” in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 554–558.

[11] Jefferson Ryan Medel and Andreas Savakis, “Anomaly detection in video using predictive convolutional long short-term memory networks,” arXiv preprint arXiv:1612.00390, 2016.

[12] Yong Shean Chong and Yong Haur Tay, “Abnormal event detection in videos using spatiotemporal autoencoder,” in International Symposium on Neural Networks. Springer, 2017, pp. 189–196.

[13] Weixin Luo, Wen Liu, and Shenghua Gao, “Remembering history with convolutional lstm for anomaly detection,” in 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2017, pp. 439–444.

[14] Fisher Yu, Wenhui Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell, “Bdd100k: A diverse driving video database with scalable annotation tooling,” arXiv preprint arXiv:1805.04687, 2018.

[15] G. Bradski, “The OpenCV Library,” Dr. Dobb’s Journal of Software Tools, 2000.

[16] B Kiran, Dilip Thomas, and Ranjith Parakkal, “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos,” Journal of Imaging, vol. 4, no. 2, pp. 36, 2018.

[17] Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazırbas, Vladimir Golkov, Patrick Van der Smagt, Daniel Cremers, and Thomas Brox, “Flownet: Learning optical flow with convolutional networks,” arXiv preprint arXiv:1504.06852, 2015.

[18] William Lotter, Gabriel Kreiman, and David Cox, “Deep predictive coding networks for video prediction and unsupervised learning,” arXiv preprint arXiv:1605.08104, 2016.