Region-based Appearance and Flow Characteristics for Anomaly Detection in Infrared Surveillance Imagery

Yona Falinie A. Gaus[1], Neelanjan Bhowmik[1], Brian K. S. Isaac-Medina[1],
Amir Atapour-Abarghouei[1], Hubert P. H. Shum[1], Toby P. Breckon[1,2]
Department of {Computer Science[1], Engineering[2]}, Durham University, Durham, UK

Abstract

Anomaly detection is a classical problem within automated visual surveillance, namely the determination of the normal from the abnormal when operational data availability is highly biased towards one class (normal) due to both insufficient sample size, and inadequate distribution coverage for the other class (abnormal). In this work, we propose the dual use of both visual appearance and localized motion characteristics, derived from optic flow, applied on a per-region basis to facilitate object-wise anomaly detection within this context. Leveraging established object localization techniques from a region proposal network, optic flow is extracted from each object region and combined with appearance in the far infrared (thermal) band to give a 3-channel spatiotemporal tensor representation for each object (1 $\times$ thermal - spatial appearance; 2 $\times$ optic flow magnitude as x and y components - temporal motion). This formulation is used as the basis for training contemporary semi-supervised anomaly detection approaches in a region-based manner such that anomalous objects can be detected as a combination of appearance and/or motion within the scene. Evaluation is performed using the Long-Term infrared (thermal) Imaging (LTD) benchmark dataset against which successful detection of both anomalous object appearance and motion characteristics are demonstrated using a range of semi-supervised anomaly detection approaches.

1. Introduction

Automated video surveillance has become increasingly prevalent in society for the security of various public facilities, transportation systems and national infrastructure alike [4]. An important operational aspect of this type of monitoring is the detection of anomalous [24, 24, 46] or unusual events [29, 30, 44], which is an area within which algorithmic solutions currently lag behind the increasingly conventional use of object detection and tracking for automated video surveillance [13].

An anomaly, in this context, refers to behaviour or appearance that deviates from normal or expected patterns for that locale. In many deployment scenarios, the use of conventional object detection or tracking solutions results in a high level of system-generated alarms for benign occurrences that fall within the regular Pattern-of-Life (PoL) of activity for that location [1].

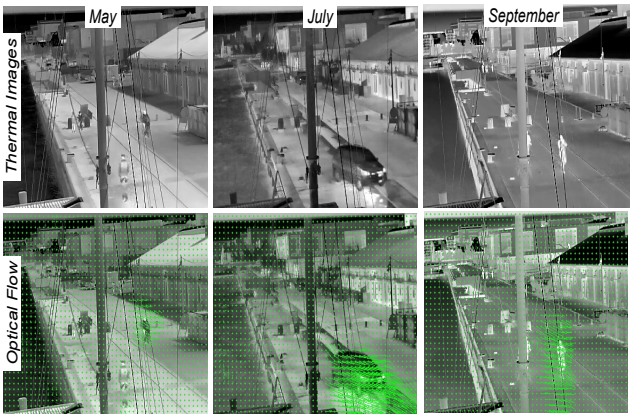

Figure 1. Exemplar infrared (thermal) imagery from the LTD dataset (upper) and corresponding motion information (lower).

In general, the anomaly detection problem addresses an aspect of the open set problem in computer vision - whilst normality in terms of the appearance and behaviour of objects within the scene can be bounded, conversely the set of possible anomalous occurrences is unbounded. To this end, anomalous events rarely occur as compared to normal activities, which in itself results in the commonplace dataset challenge of anomaly detection - whilst *normal* data samples may be abundant, the availability of *abnormal* anomalous samples is limited in both volume and variety. A common approach is to learn a model of the normal (non-anomalous) data distribution from the abundance of normal sample training data available and then detect anomalies as outliers in a semi-supervised manner [3, 5]. However, within the context of visual surveillance, this requires an understanding of complex visual patterns, and some patterns can only be detected when long-term temporal relationships and causal reasoning are learned in the model [55], such as traffic accidents, crimes or illegal activity.

Whilst there is an abundance of prior work in anomaly detection within the context of visual surveillance, they largely consider fairly basic anomalous object occurrences over datasets with a very limited timeframe [24, 29, 30, 44] such as *UCSD Ped1/2* [24], *Avenue* [29] or *ShanghaiTech* [30]. Although larger datasets such as *UCF-Crime* [44] offer more scope, their non-uniform (internet-based) curation makes their use in the evaluation of anomaly detection for fixed camera surveillance deployment challenging. In addition, all datasets largely focus on visible-band (colour) imagery, with only very few containing both visible and infrared (thermal) imagery [9, 17] despite the increasing prevalence of infrared (thermal) imagery within the operational surveillance context [6, 13, 21].

An additional challenge is the varying environmental conditions, which in itself affects both visible-band [24, 29, 30, 44] and infrared-band anomaly detection alike [17]. This issue is illustrated in Figure 1, where four normal harbour scenes exhibit varying benign changes over time in terms of contrast, illumination, foreground water ripples and other environmental factors that can then differ significantly against the *a priori* (non-anomalous) data distribution for the scene [33].

To overcome these issues, in this paper, we investigate the use of short-duration, spatiotemporal signatures as a means to in-scene object-wise anomaly detection and apply these over infrared (thermal) imagery captured under varying environmental conditions.

To this end, we propose the dual use of both visual appearance and localized motion characteristics, derived from optic flow, applied on a region-based basis to facilitate object-wise anomaly detection within this context. Leveraging established object localization from a region proposal network, optic flow is extracted from each object region and combined with the appearance in the far infrared (thermal) band to give a 3-channel spatiotemporal tensor representation for each object (1 $\times$ thermal - spatial appearance; 2 $\times$ optic flow magnitude as x and y components - temporal motion). This formulation is used as the basis for training contemporary semi-supervised anomaly detection approaches in an object-wise manner such that anomalous objects can be detected as a combination of appearance and/or motion within the scene.

The main contributions of this work are as follows:

- an extension to prior work in region-based anomaly detection [1] to jointly consider the use of both visual appearance and localized motion characteristics for in-scene objects.
- the evaluation of five semi-supervised anomaly detection approaches within this context that differ in formulation - i.e., classification-based (DFKDE [2]), reconstruction-based (FastFlow [51], GANomaly [3]) and student-teacher pair (RD [10], STFPM [48]).
- an illustration of region-based (per object) anomaly detection in the context of automated visual surveillance applied to far infrared (thermal) band imagery with quantitative and qualitative performance reported on a per object basis over the varying environment conditions of the Long-Term infrared (thermal) Imaging (LTD) benchmark dataset [33].

2. Literature Review

Anomaly detection via automated video surveillance has been intensively studied because of its potential for use in autonomous surveillance systems [44] [50] [53]. In this context, most research addresses the problem under the assumption that the operational data availability is highly biased towards one class (normal) due to both insufficient sample size, and inadequate distribution coverage for the other class (abnormal). The process is often carried out via the following steps. In the training phase, features of normal training samples are extracted. A reference model is then fitted on these features. During the testing phase, if features of the input data cannot fit the reference model well, they are considered as anomalies [52] [29] [23] [8] [16] [20].

Recently, with the great success of deep learning, contemporary approaches use the features from trained deep neural networks [11] [47] [34] [43]. Alternatively, some deep learning approaches depend on data reconstruction methods. This relies on using generative models to learn the representations of normal samples in video clips by minimising the reconstruction error [18] [28] [31] [32] [35] [42]. During inference, it is assumed that unseen anomalous video clips often cannot be reconstructed well and samples with high reconstruction errors are considered anomalies.

Whilst previous approaches mainly analyse video clips on a frame-wise basis, another approach is to classify normal or abnormal by modelling object trajectories. For instance, Li et al. [22] modelled the trajectories of normal events via sparse reconstruction analysis, then detect any abnormal trajectories as outliers. In [39], a deep autoencoder is trained to model normal trajectory, whilst the follow-up work [40] incorporates a GAN in which the discriminator is trained to distinguish normal and abnormal trajectory reconstruction errors given by a deep autoencoder [39]. While object trajectories manage to capture long-term object-level patterns, this may fail iin crowded or cluttered scenarios.

An alternative approach is to model the visual appearance using low-level features extracted from local regions. Hinami et al. [15] propose joint detection and recounting of abnormal events via multi-task Fast RCNN [14]. Whilst [15] use geodesic object proposal [19] and moving object proposals [12] to extract local regions, Adey et al. [1] incorporate the same method but take advantage of the state-of-the-art Faster-RCNN [38] to extract potential local regions, to be fed into Kernel Density Estimation (KDE) for

classification purposes. In another approach, more modern object detectors such as Single Stage Detector (SSD) [25] and CenterNet [54] are used in [18] and [41] respectively to detect local regions. In [18], the local region is then fed into SVM classifiers for anomaly classification, while in [41], the local region is trained in an adversarial manner for anomaly classification.

Another approach is to incorporate motion characteristics as an indicator of anomaly [28] [37] [32] [49]. Liu et al. [28] add an optical flow loss as the motion constraint during training time, whilst the work in [37] [32] attempts to learn motion by predicting the optical flow of the current frame. On the other hand, the work in [49] leverages the optical flow information by guiding the frame prediction, where they predict normal frames with high quality and abnormal frames with low quality.

While these efforts have shown good detection accuracy in anomaly detection tasks, most of the methods mentioned above concentrate on anomaly detection on visible-band (colour) imagery [55] [24] [29] [30] [44] and grayscale imagery [15] [1] [18] [41]. On the other hand, all aforementioned work [28] [37] [32] [49] focuses on how to directly predict future frames via optical flow. Existing work that uses region-based approaches such as [1] [15] specifically ignore the motion characteristic. Meanwhile, [18] incorporates expensive motion information such as a motion convolutional autoencoder, while [41] only relies on past spatial gradient as motion information for anomaly detection.

By contrast, we propose the dual use of both region-based appearance and flow characteristics to facilitate object-wise anomaly detection in infrared surveillance imagery. Inspired by region-based object localization based on a region proposal network [1], and computationally less expensive optical flow method, we combine the infrared (thermal) appearance with optic flow resulting 3-channel spatiotemporal tensor representation for each region. This will be used as the basis for training contemporary semi-supervised anomaly detection approaches in an object-wise manner such that anomalous objects can be detected as a combination of appearance and/or motion within the scene.

3. Methodology

Figure 2 illustrates the overall architecture of the proposed method, which consists of a two-stage approach that separates the object and anomaly detection tasks.

3.1. Object Detection and Optical Flow

An object detector is trained to predict a set of bounding boxes surrounding objects belonging to a set of classes C given the i-th thermal image $I_i \in [0,1]^{H\times W}$, where H and W are the dimensions of the image, from a sequence of N images $\mathcal{I} = \{I_i\}_{i=1}^{N}$. A prediction consists of a box represented as $\mathbf{b} = (x_c, y_c, w', h', c)$, where x_c, y_c are the centre of the box, w', h' are the width and height and $c \in C$ is the category. In parallel, the optical flow of I_i is estimated, resulting in a flow $\phi_i \in \mathbb{R}^{H\times W\times 2}$ that describes the pixel-wise displacements in the x and y directions (encoded in the last 2 channels of ϕ_i). The i-th optical flow is estimated from images I_i and I_{i-1} using a pre-trained PWC-Net [45], due to its compact model size. Subsequently, a patch $p_t \in [0,1]^{h'\times w'}$ with associated category c is extracted from the thermal image I_i given the prediction $\mathbf{b}$, and it is aggregated with an optical flow patch $p_\phi \in \mathbb{R}^{h'\times w'\times 2}$ from ϕ_i at the same spatial location defined by $\mathbf{b}$. The final object representation $p \in \mathbb{R}^{h'\times w'\times 3}$ is finally obtained from the concatenation of p_t and p_ϕ .

3.2. Anomaly Detection

We present a semi-supervised anomaly detection approach, where we train based solely on normal data samples comprising object regions within the scene (from first-stage object detection). Given the challenges of comprehensive anomalous data collection within the context of infrared (thermal) video surveillance, here, we leverage existing anomaly detection methods by down-selecting five second-stage anomaly detection approaches that do not require anomalous training examples.

DFKDE [2]: Deep Feature Kernel Density Estimation (DFKDE) is a fast one-class anomaly classification algorithm that consists of a deep neural network-based feature extraction stage followed by an anomaly classification stage consisting of principal component analysis (PCA) and Gaussian Kernel Density Estimation (KDE). In the first stage of anomaly classification, the features are reduced to the first 16 principal components via principal component analysis (PCA). In the second stage of anomaly classification, Gaussian Kernel Density estimation (KDE) is applied to the principal component features. The idea of KDE is that the training datasets follow some arbitrary distribution, and the distribution can be modelled by employing kernel density estimation. At the inference phase, if a lower probability density is observed below the threshold, which is determined by the training dataset, this indicates the presence of an anomaly against the data distribution learned from the training data examples.

FastFlow [51]: Unsupervised anomaly detection and localization via 2D normalizing flows (FastFlow) consists of 2D normalising flow for anomaly detection with a fully convolutional neural network architecture. The visual features are first extracted by a deep feature extractor and subsequently fed into the normalising flow component to estimate the probability density. In the training phase, FastFlow learns to transform the original distribution of the features into a tractable distribution in a 2D manner via a normalising flow methodology. In the inference phase, when the normal images and abnormal images simultaneously occur, the

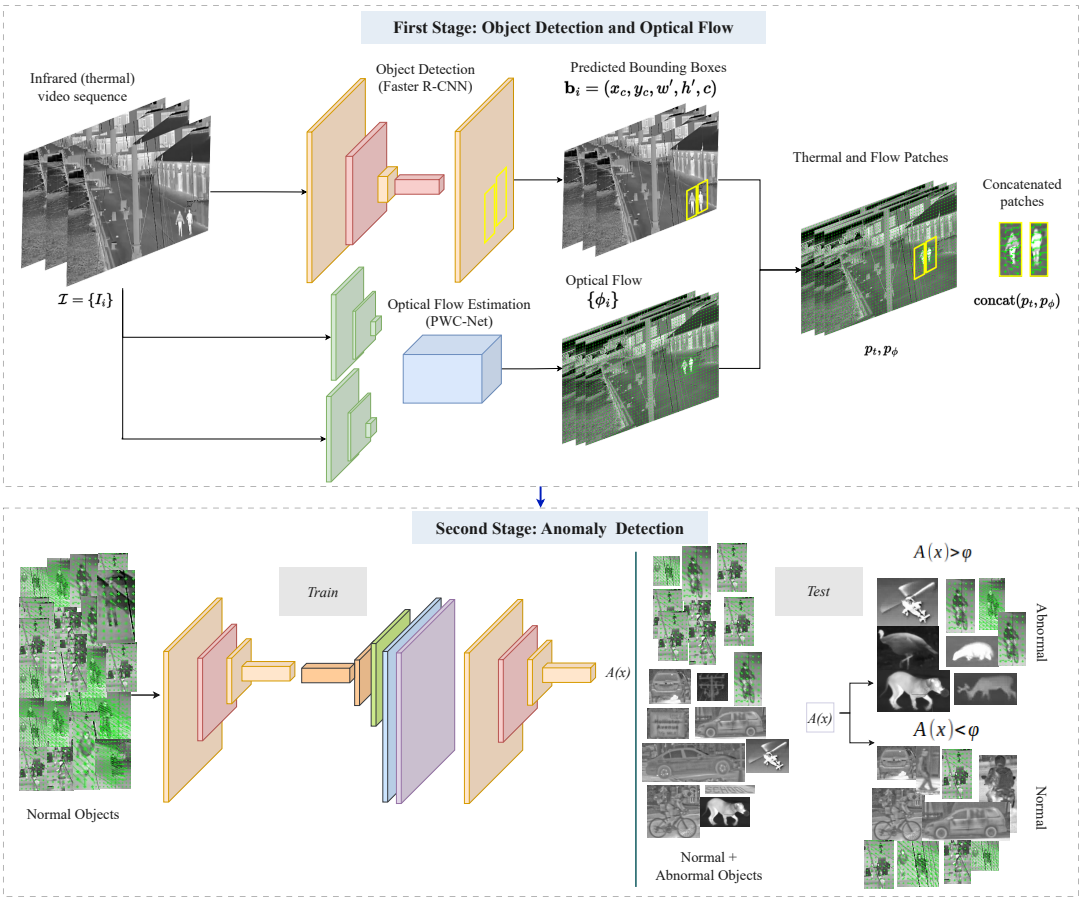

Figure 2. The proposed architecture for anomaly detection.

features of normal images will project within the centre of the distribution, while abnormal image features will project far from the centre of the distribution, indicating their abnormality. In simple terms, the probability value of each position on the 2D feature manifold is directly employed as the anomaly score.

RD [10]: Anomaly detection via reverse distillation from one-class embedding (RD) is based on a pre-trained teacher network and a trainable student network and consists of three sub-networks. The first is a pre-trained feature extractor (E). The next two are a one-class bottleneck embedding (OCBE) and the student decoder network (D). During the feature extraction stage, as the student network is trained on a normal image dataset, its feature representation of image anomalies is expected to be distinct from that of the teacher network. During training, it forces the output to be similar to the corresponding feature extractor layers by using cosine distance as the inter-feature loss metric. In this way, it is able to improve the similarity of student-teacher (S-T) representations on normal images, whilst at the same time being capable of differentiating anomalous image examples. Finally, in the inference phase, when both normal images and abnormal images occur, the cosine distance between the resultant feature maps can be used to indicate the presence of anomalies.

GANomaly [3]: Semi-supervised anomaly detection via adversarial training (GANomaly) calculates the reconstruction error based on the latent feature representation (z) and reconstructed latent feature representation (z') by adding an additional encoder structure to a conventional Generative Adversarial Network (GAN) architecture. During training, the model aims to learn the distribution of the normal (non-anomalous) data set by minimizing the difference between the two latent feature representations. Subsequently, at inference time, an anomaly score is derived from the L2 distance between the two latent feature representations.

STFPM [48]: Student-teacher feature pyramid matching for unsupervised anomaly detection (STFPM) consists of a pre-trained teacher network and a student network with an identical neural network architecture. The student network learns by making feature maps similar to those in the teacher network. Since training is performed only on normal images, the student network can only output the features of normal regions. In order to detect anomalies, STFPM uses the difference between feature maps at three different scales in the student and teacher networks. Subsequently, at the inference phase, the final anomaly score is calculated by multiplying the three maps with

the aggregation of anomalies at differing scales readily contributing to the accurate detection of anomalies of various sizes.

Each anomaly detection approach is trained over a set of normal samples, $\{p \mid p \in \mathbb{R}^{h' \times w' \times 3}\}$ constructed as a spatiotemporal object representation, as per Section 3.1.

4. Evaluation

This section presents the dataset used for evaluation, implementation details and final experimental results.

4.1. Evaluation Dataset

The LTD dataset [33] consists of infrared (thermal) surveillance imagery spanning 188 days in the period of 14^{th} May 2020 to 30^{th} of April 2021, with a total of 1689 two minutes clips sampled at 1 fps with associated bounding box annotations for 4 classes {*Human, Bicycle, Motorcycle, Vehicle*}. The images are captured at a resolution of 288×384 captured through a period of 8 months using a Hikvision DS-2TD2235D-25/50 infrared (thermal) camera (long wavelength infrared (LWIR): $8 - 14\ \mu m$). The dataset spans all hours of the day in a wide array of weather conditions overlooking the harbour front of Aalborg, Denmark depicting drastic changes in both object and scene appearance due to seasonal conditions within a static security monitoring context. Normal training samples are extracted via object detection from infrared (thermal) images. Since the LTD dataset does not provide information on which object is an anomaly, we construct our own anomaly thermal dataset by manually cropping anomalous objects from a video surveillance scenario dataset [27], for validation purposes. In total, we make use of 17,109 normal objects as training data for the anomaly detection step.

4.2. Implementation Details

We follow the dataset protocol from [33] by choosing the coldest day of the month (February) as the training set, which exists in three variants: coldest day 13^{th} of February, the corresponding week $13 - 20$ of February, and the entirety of February. In this experiment, we use the infrared (thermal) video imagery from the coldest day of the month, day 13th of February as the training dataset. We then incorporate optical flow by combining it with appearance in the infrared (thermal) imagery, resulting in a 3-channel spatiotemporal tensor representation for each object (i.e. $1 \times$ thermal channel for spatial appearance; $2 \times$ optic flow channels for horizontal and vertical motion magnitudes).

Since our anomaly detection is based on object-wise image regions, we first extract a large number of object regions by training a Faster R-CNN [38], pre-trained on MS-COCO [26], over our set of normal objects using [7], and therein only retain the bounding box localisation information and discard the classification labels. We employ the SGD optimiser and set the learning rate as 252×10^{-3}, the momentum as 0.9 and the weight decay as 1×10^{-4}. Finally, we train for 100 epochs with a batch size of 16.

We construct our training dataset using the cropped object-wise image regions (i.e. bounding boxes) obtained via the earlier Faster R-CNN approach, and combine each with the corresponding optic flow for that image region, to form the input set to our proposed anomaly detection model training using [2].

All implementations and visualisation are conducted in PyTorch [36] with a single NVIDIA 1080Ti GPU. All CNNs used in the experiments were pretrained with ImageNet. For fair comparison and consistency, we used the same parameters for all experiments; the parameters follow the defaults used in [2] or within the original work.

4.3. Quantitative Evaluation

The model performance is evaluated quantitatively through the area under ROC (AUROC), Accuracy, F1 score, Precision and Recall at object level. We compare the use of anomaly detection via classification-based (DFKDE [2]), reconstruction-based (FastFlow [51], GANomaly [3]) and student-teacher paired (RD [10], STFPM [48]) methods.

Anomalies are detected when the model output exceeds a given threshold A(x) $> \phi$. When performing anomaly detection, image regions with scores below the calculated threshold are considered normal and whilst regions with scores above the threshold are considered as anomalies [2].

Table 1 compares anomaly detection performance based on object appearance in infrared (thermal) imagery (only) and when also combined with short-term object motion characteristics (via optical flow, Section 3.2). In the first experiment, anomaly detection performance in infrared (thermal) imagery (*IR*) is used to provide benchmark performance. The combination of infrared appearance and flow-based motion characteristics (*IR+Flow*) produces superior anomaly detection results than infrared appearance (*IR*) alone. Statistically, we observe a significant increase in performance given by student-teacher pair approach, both RD method (0.468 $\Rightarrow$ 0.912) and STFPM method (0.467 $\Rightarrow$ 1.000) respectively in AUROC. Whilst for Accuracy and F1 score, the highest performance jump is given by reconstruction based approach, GANomaly (0.853 $\Rightarrow$ 0.999 and 0.914 $\Rightarrow$ 0.999) respectively. Meanwhile, in Precision and Recall, a significant gain in performance is observed for RD (0.889 $\Rightarrow$ 0.972) and GANomaly (0.873 $\Rightarrow$ 0.998) respectively.

Table 2 shows the comparison of mean anomaly score on infrared appearance (*IR*) and combined appearance and motion characteristics (*IR+Flow*) for threshold $\phi = 0.5$. An increase can be observed for FastFlow, GANomaly, RD and whilst a comparable mean value for DFKDE is observed potentially demonstrating a more pronounced separation between anomalous and non-anomalous samples with the use of motion characteristics (*IR+Flow*). However, a performance drop is observed for RD (0.859 $\Rightarrow$ 0.709), despite

Table 1. LTD dataset: object-level performance of anomaly detection using only infrared object appearance (*IR*) and combined infrared appearance and motion characteristics (*IR+Flow*) information.

	DFKDE		FastFlow		GANomaly		RD		STFPM	
	IR	*IR+Flow*	*IR*	*IR+Flow*	*IR*	*IR+Flow*	*IR*	*IR+Flow*	*IR*	*IR+Flow*
AUROC ↑	0.980	0.998	0.984	0.999	0.788	1.000	0.468	0.912	0.467	1.000
Accuracy ↑	0.924	0.988	0.944	0.999	0.853	0.999	0.867	0.975	0.826	1.000
F1 score ↑	0.955	0.993	0.968	0.999	0.914	0.999	0.928	0.986	0.904	1.000
Precision ↑	0.995	0.995	0.993	1.000	0.952	1.000	0.889	0.972	0.886	1.000
Recall ↑	0.918	0.991	0.944	0.998	0.873	0.998	0.972	1.000	0.923	1.000

Table 2. LTD dataset: mean anomaly score of anomalous samples using *IR* and *IR+Flow* information (at $\phi = 0.5$).

Model	Anomaly Score	
	IR	*IR+Flow*
DFKDE	0.901	0.895
FastFlow	0.634	0.669
GANomaly	0.691	0.924
RD	0.859	0.709
STFPM	0.566	0.607

having performance increase in AUROC (0.468 ⇒ 0.912, Table 1).

4.4. Qualitative Evaluation

Qualitative results are provided for all five of the selected anomaly detection approaches, where normal object occurrences are shown in *green* and anomalous objects (e.g. large vehicles, based on our training regime) are shown in *red* with an associated anomaly score (in *blue*, normalised to range 0→1).

Detected anomalous objects, not present in the training set, generally correspond to the appearance of non-pedestrian objects (e.g. large vehicles) or unusual object motion in the scene. For example, an unusual pedestrian motion could be people walking haphazardly, sudden changes in walking speed or an unusual direction of motion within the scene.

In Figure 3, we can observe that: *1)* normal object occurrences such as pedestrians (Figure 3 - first row); and *2)* anomalous object occurrences, such as large vehicles (Figure 3 - first, second and third row); are all detected well. The large delivery vehicle (Figure 3 - second row) and the large construction vehicle (Figure 3 - third row) were detected as anomaly instances since these are new objects observed by the anomaly detection model. In addition, anomaly detection is observed to perform well on the anomalous objects that contain normal pixels within their bounds. For example, the overlap between the large delivery vehicle and the normal walking pedestrian (Figure 3 - second row), contributes to higher performance in F1-score across all models in Table 1.

Figure 4 shows consecutive frames from one of the test video sequences within the infrared (thermal) imagery LTD dataset [33]. In the figure (Figure 4, upper) we observe that anomaly detection via DFKDE and RD detect the large vehicle object as anomalous whilst the pedestrian objects remain labeled as normal objects (denoted with *green* and *red* bounding box annotation, respectively). However, in the next consecutive frame (Figure 4, lower), only the DFKDE approaches detect the large vehicle object as anomalous when the portion of the vehicle within the frame is occluded (as it departs the scene), whilst RD detects the same objects as normal. Such factors contribute to the lower AUROC performance for RD, as shown in Table 1.

Qualitative evaluation results are shown in Figure 5 using the combination of infrared (thermal) imagery and optical flow information from the LTD dataset [33]. A visualisation of the visual appearance of the infrared (thermal) imagery (*IR*) and its corresponding motion characteristics, obtained from optic flow (*IR+FLOW*), are additionally presented in Figure 5. Overall, the qualitative results in Figure 5 demonstrate that the combination of infrared (thermal) imagery and optical flow magnitude components, as a measure of short-term object motion characteristics, reduce the false positive rate for anomaly detection (as also reflected in Table 1). In the first example (Figure 5, A and D), the truck and construction plant vehicle are only detected as anomalous when motion information included via optical flow. In this example, the flow pattern on the large vehicle (*IR+FLOW*) suggested that the vehicle is moving towards the camera. In the second example (Figure 5, B), the appearance (only) of pedestrian walking is detected as normal. However, with motion information (*IR+FLOW*), it manages to detect pedestrian walking in an unusual scene direction as an anomaly. In the same example, there are anomalous flow patterns drawn on the image (*IR+FLOW*), suggesting that the pedestrian is walking at a high velocity. In the third example (Figure 5, C), the truck appearance is detected as normal whilst entering the frame in the infrared (thermal) imagery (*IR*). However, with motion information added (+*Flow*), it is correctly detected as an anomaly. In the fourth example, it shows that the model still works well in a crowded scene with pedestrians and an anomalous object having similar intensities with the background (Figure 5, E). In the fifth example, we can see that the model also gener-

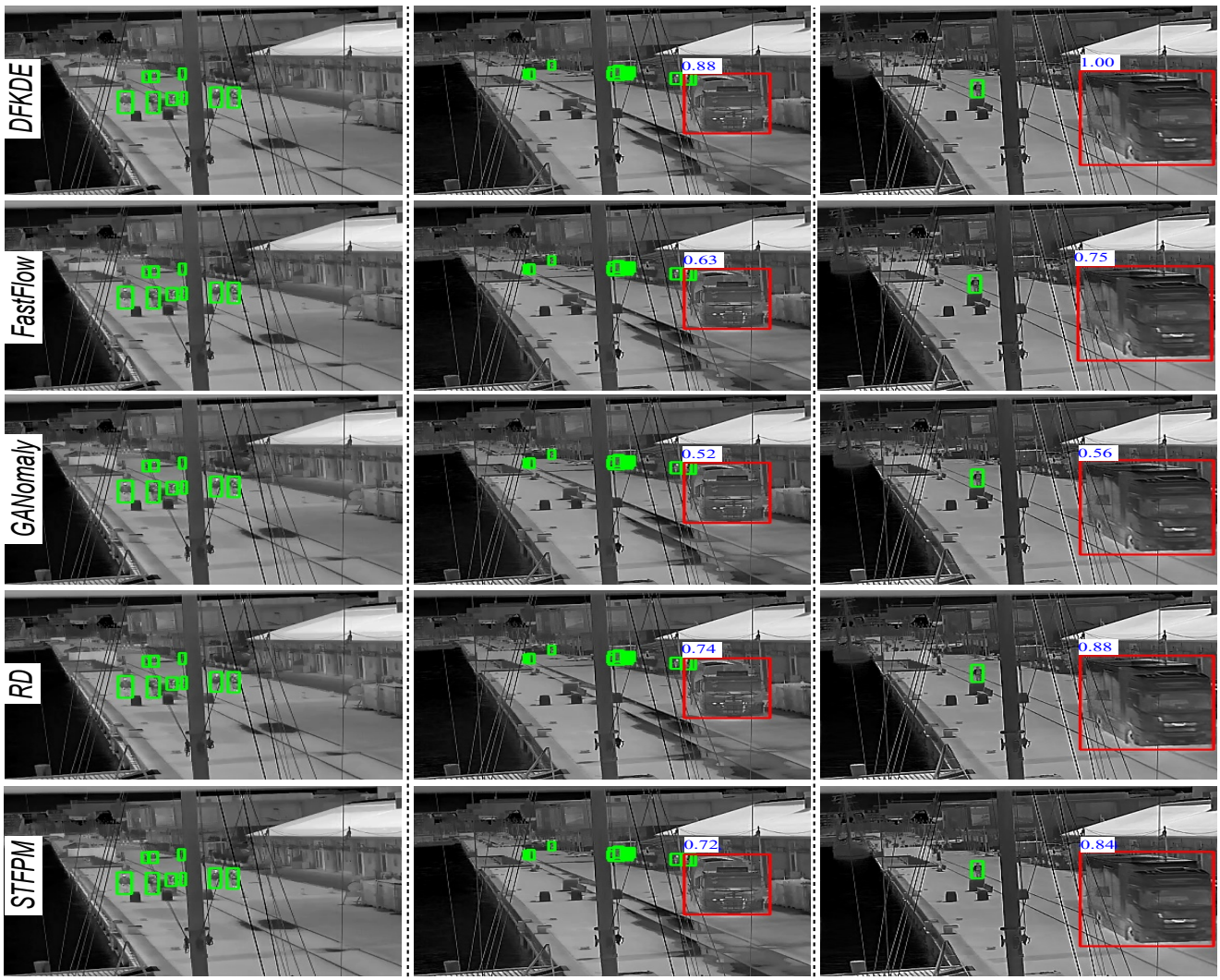

Figure 3. Detection of of normal object [*left*] denoted as green and anomalous object [*middle,left*] denoted as red.

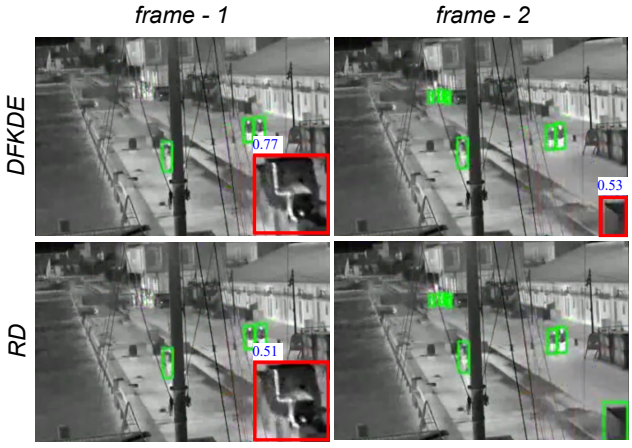

Figure 4. Detection in consecutive frames where a construction plant vehicle (red box) is consistently detected as anomalous with DFKDE (upper) but not with RD (lower) due to occlusion.

ates false alarms, by detecting normal pedestrian walking as an anomaly (Figure 5,F).

By taking a closer look at such false positives, we can observe that even normal pedestrian walking may generate high velocity motion, due to periodic movement of their limbs, although it is restricted to small regions of the body. Therefore it can be determine that some average walking motions may be detected as anomlous under such conditions, giving rise to such false positive results. Overall, most genuine anomalous objects appear to generate associated anomalous flow patterns across the entire object surface, as shown in all anomalous objects in Figure 5.

Overall, these examples (Figures 3 / 5) illustrate the performance of region-based anomaly detection in the context of infrared (thermal) surveillance imagery and the performance benefits from the dual use of both object appearance and (short-term) motion characteristics.

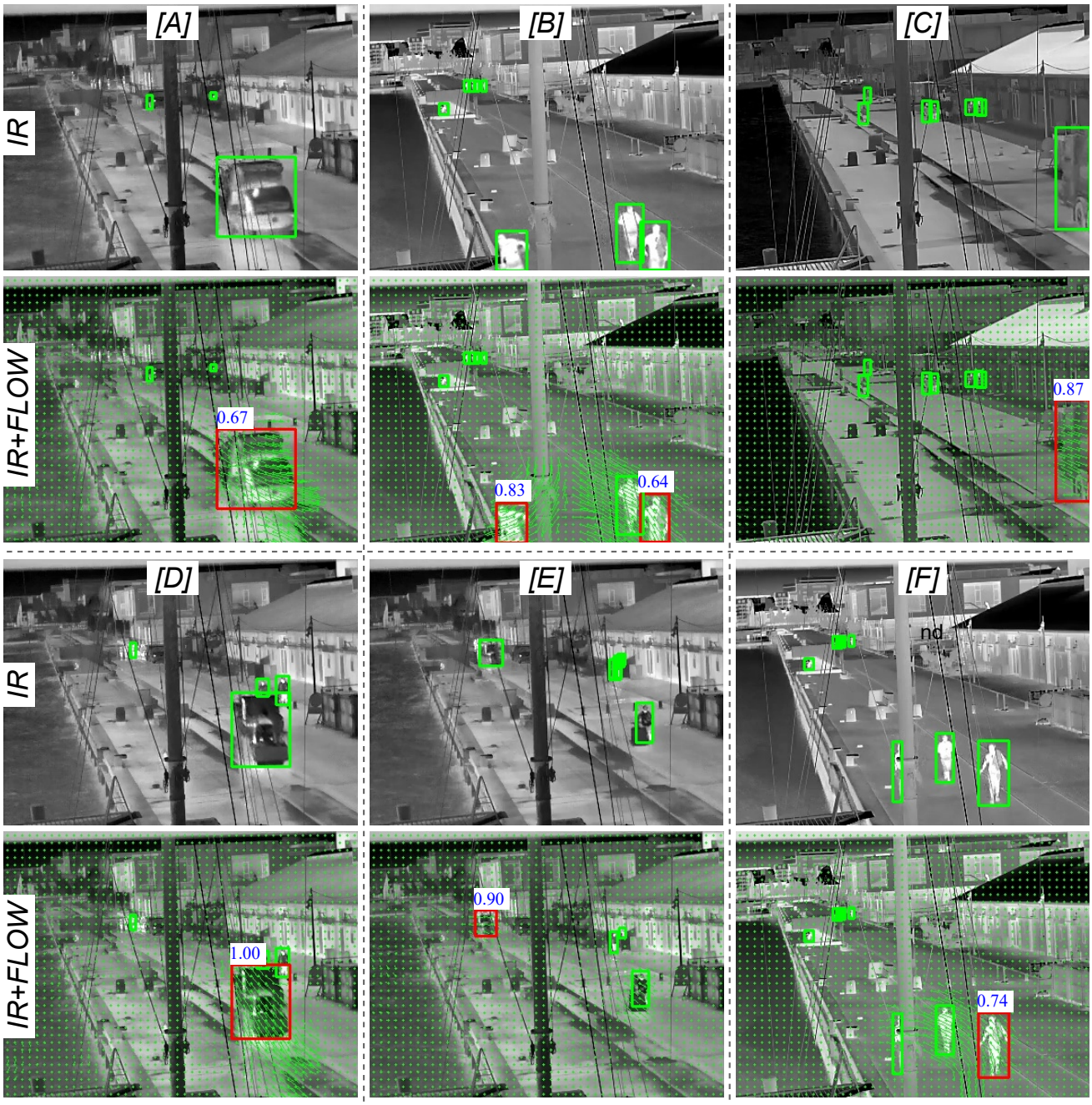

Figure 5. Detection of normal and anomalous object in infrared (thermal) imagery (*IR*) and its corresponding optical flow (*IR+Flow*)

5. Conclusion

In this work we present region-based anomaly detection in an automated visual surveillance context, by proposing the dual use of both appearance and short-term motion characteristics across infrared (thermal) imagery. We evaluate the performance of five semi-supervised anomaly detection approaches spanning classification based (DFKDE), reconstruction based (FastFlow, GANomaly) and student-teacher pair (RD, STFPM) paradigms. Within this study, we observe that the combination of infrared (thermal) object appearance and short-term motion characteristics, recovered from optic flow, result in a notable improvement in anomalous object detection performance then compared to using infrared appearance alone.. We also qualitatively demonstrate temporally consistent anomaly detection on a per-object basis. Future work includes the use of temporal scene analysis to expand this use of automatic anomaly detection towards anomalous behaviour detection within infrared (thermal) surveillance imagery.

Acknowledgment

This work was funded through the Defence and Security Accelerator on behalf of the Nuclear Decommissioning Authority.

References

[1] P. Adey, M. Bordewich, O.K. Hamilton, and T.P Breckon. Region based anomaly detection with real-time training and analysis. In *Proc. Int. Conf. on Machine Learning Applications*, pages 495–499. IEEE, December 2019. 1, 2, 3

[2] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh Ahuja, and Utku Genc. Anomalib: A deep learning library for anomaly detection, 2022. 2, 3, 5

[3] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer, 2018. 1, 2, 4, 5

[4] Ernesto L Andrade, Scott Blunsden, and Robert B Fisher. Modelling crowd scenes for event detection. In *18th international conference on pattern recognition (ICPR'06)*, volume 1, pages 175–178. IEEE, 2006. 1

[5] J.W. Barker and T.P. Breckon. Panda: Perceptually aware neural detection of anomalies. In *Proc. Int. Joint Conference on Neural Networks*, pages 1–8. IEEE, July 2021. 1

[6] Toby P Breckon, Anna Gaszczak, Jiwan Han, Marcin L Eichner, and Stuart E Barnes. Multi-modal target detection for autonomous wide area search and surveillance. In *Emerging Technologies in Security and Defence; and Quantum Security II; and Unmanned Sensor Systems X*, volume 8899, pages 172–191. SPIE, 2013. 2

[7] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 5

[8] Kai-Wen Cheng, Yie-Tarng Chen, and Wen-Hsien Fang. Video anomaly detection and localization using hierarchical feature representation and gaussian process regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2909–2917, 2015. 2

[9] Kellie Corona, Katie Osterdahl, Roderic Collins, and Anthony Hoogs. Meva: A large-scale multiview, multimodal video dataset for activity detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1060–1068, 2021. 2

[10] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9737–9746, 2022. 2, 4, 5

[11] Zhiwen Fang, Jiafei Liang, Joey Tianyi Zhou, Yang Xiao, and Feng Yang. Anomaly detection with bidirectional consistency in videos. *IEEE Transactions on Neural Networks and Learning Systems*, 2020. 2

[12] Katerina Fragkiadaki, Pablo Arbelaez, Panna Felsen, and Jitendra Malik. Learning to segment moving objects in videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4083–4090, 2015. 2

[13] Yona Falinie A Gaus, Neelanjan Bhowmik, Brian KS Isaac-Medina, and Toby P Breckon. Visible to infrared transfer learning as a paradigm for accessible real-time object detection and classification in infrared imagery. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies IV*, volume 11542, pages 13–27. SPIE, 2020. 1, 2

[14] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 2

[15] Ryota Hinami, Tao Mei, and Shin'ichi Satoh. Joint detection and recounting of abnormal events by learning deep generic knowledge. In *Proceedings of the IEEE international conference on computer vision*, pages 3619–3627, 2017. 2, 3

[16] Timothy Hospedales, Shaogang Gong, and Tao Xiang. A markov clustering topic model for mining behaviour in video. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1165–1172. IEEE, 2009. 2

[17] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1037–1045, 2015. 2

[18] Radu Tudor Ionescu, Fahad Shahbaz Khan, Mariana-Iuliana Georgescu, and Ling Shao. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7842–7851, 2019. 2, 3

[19] Philipp Krähenbühl and Vladlen Koltun. Geodesic object proposals. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 725–739. Springer, 2014. 2

[20] Louis Kratz and Ko Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *2009 IEEE conference on computer vision and pattern recognition*, pages 1446–1453. IEEE, 2009. 2

[21] M.E. Kundegorski, S. Akcay, G. Payen de La Garanderie, and T.P. Breckon. Real-time classification of vehicle types within infra-red imagery. In *Proc. SPIE Optics and Photonics for Counterterrorism, Crime Fighting and Defence*, volume 9995, pages 1–16. SPIE, September 2016. 2

[22] Ce Li, Zhenjun Han, Qixiang Ye, and Jianbin Jiao. Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. *Neurocomputing*, 119:94–100, 2013. 2

[23] Nannan Li, Xinyu Wu, Huiwen Guo, Dan Xu, Yongsheng Ou, and Yen-Lun Chen. Anomaly detection in video surveillance via gaussian process. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(06):1555011, 2015. 2

[24] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE transactions on pattern analysis and machine intelligence*, 36(1):18–32, 2013. 1, 2, 3

[25] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 3

[26] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 5

[27] Qiao Liu, Xin Li, Zhenyu He, Chenglong Li, Jun Li, Zikun Zhou, Di Yuan, Jing Li, Kai Yang, Nana Fan, et al. Lsotb-tir: A large-scale high-diversity thermal infrared object tracking benchmark. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3847–3856, 2020. 5

[28] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6536–6545, 2018. 2, 3

[29] 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*, pages 2720–2727, 2013. 1, 2, 3

[30] Weixin Luo, Wen Liu, and Shenghua Gao. A revisit of sparse coding based anomaly detection in stacked rnn framework. In *Proceedings of the IEEE international conference on computer vision*, pages 341–349, 2017. 1, 2, 3

[31] Romero Morais, Vuong Le, Truyen Tran, Budhaditya Saha, Moussa Mansour, and Svetha Venkatesh. Learning regularity in skeleton trajectories for anomaly detection in videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11996–12004, 2019. 2

[32] Trong-Nguyen Nguyen and Jean Meunier. Anomaly detection in video sequence with appearance-motion correspondence. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1273–1283, 2019. 2, 3

[33] Ivan Adriyanov Nikolov, Mark Philip Philipsen, Jinsong Liu, Jacob Velling Dueholm, Anders Skaarup Johansen, Kamal Nasrollahi, and Thomas B Moeslund. Seasons in drift: A long-term thermal imaging dataset for studying concept drift. In *Thirty-fifth Conference on Neural Information Processing Systems*, 2021. 2, 5, 6

[34] Guansong Pang, Cheng Yan, Chunhua Shen, Anton van den Hengel, and Xiao Bai. Self-trained deep ordinal regression for end-to-end video anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12173–12182, 2020. 2

[35] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14372–14381, 2020. 2

[36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 5

[37] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe. Abnormal event detection in videos using generative adversarial nets. In *2017 IEEE international conference on image processing (ICIP)*, pages 1577–1581. IEEE, 2017. 3

[38] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 2015. 2, 5

[39] Pankaj Raj Roy and Guillaume-Alexandre Bilodeau. Road user abnormal trajectory detection using a deep autoencoder. In *Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13*, pages 748–757. Springer, 2018. 2

[40] Pankaj Raj Roy and Guillaume-Alexandre Bilodeau. Adversarially learned abnormal trajectory classifier. In *2019 16th Conference on Computer and Robot Vision (CRV)*, pages 65–72. IEEE, 2019. 2

[41] Pankaj Raj Roy, Guillaume-Alexandre Bilodeau, and Lama Seoud. Local anomaly detection in videos using object-centric adversarial learning. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV*, pages 219–234. Springer, 2021. 3

[42] Mohammad Sabokrou, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3379–3388, 2018. 2

[43] Sorina Smeureanu, Radu Tudor Ionescu, Marius Popescu, and Bogdan Alexe. Deep appearance features for abnormal behavior detection in video. In *International Conference on Image Analysis and Processing*, pages 779–789. Springer, 2017. 2

[44] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6479–6488, 2018. 1, 2, 3

[45] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8934–8943, 2018. 3

[46] Yipeng Sun et al. Anomaly detection by principal component analysis and autoencoder approach. 2021. 1

[47] Radu Tudor Ionescu, Sorina Smeureanu, Bogdan Alexe, and Marius Popescu. Unmasking the abnormal events in video. In *Proceedings of the IEEE international conference on computer vision*, pages 2895–2903, 2017. 2

[48] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for unsupervised anomaly detection. *arXiv preprint arXiv:2103.04257*, 2021. 2, 4, 5

[49] Hongyong Wang, Xinjian Zhang, Su Yang, and Weishan Zhang. Video anomaly detection by the duality of normality-granted optical flow. *arXiv preprint arXiv:2105.04302*, 2021. 3

[50] Peng Wu, Jing Liu, Yujia Shi, Yujia Sun, Fangtao Shao, Zhaoyang Wu, and Zhiwei Yang. Not only look, but also listen: Learning multimodal violence detection under weak supervision. In *European conference on computer vision*, pages 322–339. Springer, 2020. 2

[51] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021. 2, 3, 5

[52] Bin Zhao, Li Fei-Fei, and Eric P Xing. Online detection of unusual events in videos via dynamic sparse coding. In *CVPR 2011*, pages 3313–3320. IEEE, 2011. 2

[53] Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H Li, and Ge Li. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1237–1246, 2019. 2

[54] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019. 3

[55] Sijie Zhu, Chen Chen, and Waqas Sultani. Video anomaly detection for smart surveillance. In *Computer Vision: A Reference Guide*, pages 1–8. Springer, 2020. 1, 3