2019 International Conference on Unmanned Aircraft Systems (ICUAS)
Atlanta, GA, USA, June 11-14, 2019

Anomaly Detection and Cognizant Path Planning for Surveillance Operations using Aerial Robots

Tung Dang, Shehryar Khattak, Christos Papachristos, and Kostas Alexis

***Abstract*— In this paper we address the problem of unsupervised anomaly detection and cognizant path planning for surveillance operations using aerial robots. Through one–class classification exploiting deep learned features on image data and a Bayesian technique to fuse, encode and update anomaly information on a real–time reconstructed occupancy map, the robot becomes capable of detecting and localizing anomalies in its environment. Provided this information, path planning for autonomous exploration of unknown areas and simultaneous maximization of the entropy of sensor observations over abnormal regions is developed. The method is verified experimentally through field deployments above a desert-like environment and in a parking lot. Furthermore, analysis results on the suitability of different deep learning–based and hand–engineered features for anomaly detection tasks are presented.**

I. INTRODUCTION

Recent advancements in autonomous aerial robotics research have enabled their wide utilization in a multitude of critical applications including inspection and surveillance missions, search and rescue operations, precision agriculture, and more [1–9]. In such applications, the essential role of the aerial robot is that of information gathering, either through a prescribed inspection path or based on online exploration path planning algorithms. Very often the objective is that of mapping and reconstructing a representation of the environment or detecting objects of interest belonging to predefined classes. However, in a significant set of important applications, especially those related with the tasks of surveillance, our prior knowledge of the environment is such that although it permits an understanding of what is expected to be perceived under normal conditions, knowledge of what types of objects or events may constitute "anomalies of interest" is either sparse or absent altogether.

In response to this fact, this work focuses on the problem of unsupervised anomaly detection and cognizant exploration path planning using aerial robots. Considering the broad family of surveillance, monitoring and exploration missions, the first question asked is how can an aerial robot autonomously detect "anomalies" in its sensor stimuli given only knowledge of example "normal data". As "normal" we refer to data collected from *analogous* environments but not from the exact area in which an anomaly detection mission takes place. As such, the problem of anomaly detection is an unsupervised one and belongs to the broader case of one–class classification learning. Provided the ability of the robot to efficiently detect anomalies in sensor stimuli, the second goal in this work is that of fusing this information at the map level in order to not only exploit multi–view consensus and robustify the anomaly detection but also to enable path planning that ensures environment exploration and optimized information gathering in the abnormal areas.

Fig. 1. Anomaly detection and cognizant path planning for aerial robotic surveillance missions. Two missions are conducted, namely above a Desert-like environment and inside a Parking lot. Indicative results are shown at https://youtu.be/Bim-wIiTKe8.

More specifically, in this paper anomaly is detected in an online manner using camera data onboard the aerial robot. Normal camera data are collected a priori (through previous robotic deployments) from areas similar but not identical to those of the exploration or surveillance mission and the anomaly detection classifier is trained using deep learning–based feature extraction and one–class Support Vector Machines (SVM). During the robot's mission, real–time camera data are tested regarding the presence of anomalies. When an abnormal camera frame then object proposals based on visual saliency [10–14] are utilized to attempt to localize the anomaly in the frame. Provided this relative localization of the anomaly and the simultaneous ability of the robot to build 3D maps of its environment, anomaly is then "encoded" in an occupancy map. Exploiting a Bayesian update on the map level, parts of the environment are annotated as abnormal only when there is a consensus among different viewpoints. Finally, a two–step path planning algorithm is designed to guarantee efficient environment exploration, while simultaneously ensuring the re–observation of areas presenting anomaly from robot configurations that maximize entropy compared to prior viewpoints in terms of the distance and the orientation of a viewpoint to an abnormal region.

The proposed strategy for unsupervised anomaly detection and cognizant information gathering using aerial robots is verified in field experiments that include a) surveillance over a desert-like environment in Northern Nevada, as well as b) a fully autonomous exploration mission inside a parking lot. In both cases the data used for the training of the anomaly detection framework were not sampled from the place of the missions. Furthermore, comparison of the suitability of

This material is based upon work supported by the Department of Energy under Award Number [DE-EM0004478].

The authors are with the Autonomous Robots Lab, University of Nevada, Reno, 1664 N. Virginia, 89557, Reno, NV, USA tung.dang@nevada.unr.edu

different deep learned features for one–class classification using SVMs is also presented. An open dataset that contains the training and test data is released [15]. The algorithm implementation will be released as an open source code package prior to the possible publication of our paper.

This paper is structured as follows: The considered problem is defined in II and the proposed approach is detailed in Section III. Evaluation studies are presented in Section IV, followed by conclusions in Section V.

II. PROBLEM STATEMENT

The first problem considered in this work is that of being given a set of "normal" images that correspond to the training data $\mathcal{C}^T$ for one–class classification, and addressing the challenge of detecting anomaly in camera images $\mathcal{C}^r$ captured by the robot during its operation while also encoding this information on a real–time reconstructed 3D map $\mathcal{M}$. Each image $c_i^r \in \mathcal{C}^r$ is automatically classified as being "normal" or presenting "anomaly" given the above mentioned unsupervised learning process. More formally the problem as considered in this work is formulated as follows:

***Problem** 1 (Unsupervised Anomaly Detection)* Let $\mathcal{C}^T$ be a training set of N_T collected images $c_i^T \in \mathcal{C}^T$ that correspond to "normal" training data. The problem of "Unsupervised Anomaly Detection" (UAD) as considered in this paper is that of training a One-class SVM such that a decision (sign) function $f(c_i^r)$ is derived and can decide if a camera frame c_i^r collected by the robot during its mission belongs to the type of "normal" data provided for training or if it is as an outlier for this population and therefore an "anomaly".

Provided the successful detection of anomaly in camera data $\mathcal{C}^r$, the task of identifying the image patches $c_i^{r,a}$ that correspond to the object or entity that leads to the detection of anomaly is then resolved, and the corresponding pixels are projected on an online reconstructed occupancy map $\mathcal{M}$ which consists of voxels m with each of them being characterized as "normal" or "abnormal". Voxel anomaly encoding should account for multi-view consensus.

Given the solution to UAD and the capability to annotate an online reconstructed map with classification information for anomaly, the second problem considered is that of anomaly–aware planning for exploration. This problem is divided into two subproblems, namely that of ensuring full exploration of an initially unknown area, and that of ensuring high–entropy observations in the area of anomaly such that anomaly detection is robustified.

***Problem** 2 (Volumetric Exploration)* Given a bounded volume V^E, find a collision free path σ starting at an initial configuration $\xi_{init} \in \Xi$ that leads to identifying the free and occupied parts V^E_{free} and V^E_{occ}, such that there does not exist any collision free configuration from which any piece of $V^E \setminus \{V^E_{free}, V^E_{occ}\}$ could be perceived ($V^E_{free} \cup V^E_{occ} = V^E \setminus V^E_{res}$). Feasible paths σ are subject to the limited Field of View (FoV) of the sensor and its effective sensing distance.

Volumetric exploration is considered as an extrinsic objective for the robot to be optimized with priority. Maximization of information sampling over abnormal areas can then be formulated as a local path optimization problem given the initial and final robot configurations ξ_{init}, ξ_{final} corresponding to sequential viewpoints of the exploration path.

***Problem** 3 (Anomaly-aware Planning)* Given a bounded volume $V^A \subseteq V^E$ enclosing ξ_{init} and ξ_{final}, find a collision free path σ^A connecting $\xi_{init} \in \Xi$ and ξ_{final}, and maximizing an entropy–based information sampling objective over areas of the environment that present anomaly, given a classifier of anomaly $f(c_i^r)$ on images c_i^r and from that to occupancy maps $\mathcal{A}(\mathcal{M})$. Feasible paths σ^A_f are subject to a) the sensor FoV and effective sensing distance, b) robot dynamic constraints, and c) a defined time budget.

III. PROPOSED APPROACH

In this section, the proposed anomaly detection, Bayesian fusion and map encoding scheme, and anomaly–aware exploration path planning are described. Figure 2 provides an overview of the algorithmic architecture detailed below.

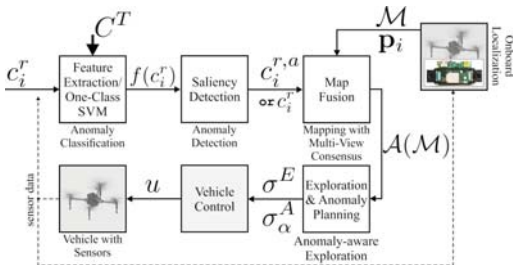

Fig. 2. Overview of the architecture for anomaly classification, detection, map fusion, exploration and anomaly–aware cognizant path planning.

A. Anomaly Detection and Bayesian Map Encoding

Unsupervised anomaly detection on camera data is achieved through the appropriate selection of reliable feature extraction and efficient training of a single–class SVM as detailed below. This approach differs from other works such as [16] using a mixture of dynamic texture model and visual saliency to detect anomaly based on temporal and spatial information, [17] using Markov Random Fields to model the video frames and detecting anomaly based on probabilistic principal component analysis, or the work in [18] employing autoencoders and generative models to detect anomaly based on the reconstruction error of image frames.

1) Feature Extraction to Train for Anomaly Detection: To enable learning from camera images $\mathcal{C}^T$ and respectively testing during the operation of the vehicle, a feature extraction scheme is employed. More specifically, the pre–trained Convolutional Neural Network (CNN) in [19] (AlexNet) was utilized to extract a set of features ϕ_i per camera frame i during the training (c_i^T) and test steps (c_i^r). This network is pretrained using images from the ImageNet dataset and is known to achieve high-quality results on object recognition.

In this work, the first stage of the AlexNet is employed as a feature extractor for the one–class classification problem. As such, a 4096–dimensional vector extracted from the first

fully connected layer (FC-6) in the last three layers of the AlexNet was used to provide features for the subsequent one–class SVM classification step. This approach follows the intuition that the first stage of this network, which includes convolutional layers, aims to extract good features from images and then in the last stage, fully connected layers are used to perform the classification task as in traditional neural nets. Moreover, given the fact that AlexNet was trained in a very large dataset, its first stage is expected to perform well as a feature extractor for such tasks. As detailed in the experimental analysis results presented in Section IV-A, AlexNet was found to provide superior accuracy compared to a variety of deep learning–based features (e.g., based on VGG16 [20], VGG19 [20], ResNet50 [21], InceptionResNet [22]) and traditional schemes (e.g., SIFT [23]), while simultaneously having a sufficiently small computational cost for its onboard and online utilization.

2) One–class SVM for Anomaly Detection: Consider the unlabelled training data of camera images $\mathcal{C}^T$ and the associated set of sets of extracted features Λ:

$$\Lambda = \{\phi_1, ..., \phi_m\} \in \mathcal{L} \tag{1}$$

where $m \in \mathbb{N}$ is the number of sets of extracted features for each of the camera frames used for training. Let Φ be a feature map $\mathcal{L} \rightarrow \mathcal{H}$, where $\mathcal{H}$ is a dot product space such that the dot product in the image of Φ can be computed by evaluating a Gaussian kernel:

$$k(\phi_i, \phi_j) = e^{-\left\|\phi_i - \phi_j\right\|^2/c}, \ i, j \in 1, ..., m \tag{2}$$

The One–class SVM method allows to train a function f that takes the value $+1$ in a "small" region capturing most of the data points of the features extracted from the normal training data, and -1 elsewhere. This is achieved by mapping the training data through the Gaussian kernel and then separating them from the origin with maximum margin as described in [24]. For a new set of image features ϕ', the value $f(\phi')$ is determined by evaluating which side of the hyperplane it falls on. To separate the provided training dataset from the origin, the following quadratic program is solved:

$$\min_{\mathbf{w} \in \mathcal{H}, \xi \in \mathbb{R}^m, \rho \in \mathbb{R}} \quad \frac{1}{2}\left\|w\right\|^2 + \frac{1}{\nu m}\sum_i \xi_i - \rho \tag{3}$$

$$\text{s.t.} \quad \langle \mathbf{w}, \Phi(\phi_i) \rangle \geq \rho - \xi_i, \ \xi_i \geq 0 \tag{4}$$

where $\nu \in (0, 1]$ is a parameter that tends to be equal both to the fraction of support vectors and the fraction of outliers, and ξ_i represents slack variables. As nonzero slack variables ξ_i are penalized, then if $\mathbf{w}$ and ρ correspond to the solution of this problem, then the decision function takes the form:

$$f(\phi_i) = \text{sgn}(\langle \mathbf{w}, \Phi(\phi) \rangle - \rho) \tag{5}$$

and will be equal to 1 for most training examples ϕ_i (features extracted from "normal" camera frames $c_i^T \in \mathcal{C}^T$). Solving this problem is achieved by introducing the Lagrangian:

$$L(\mathbf{w}, \boldsymbol{\xi}, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\left\|\mathbf{w}^2\right\| + \frac{1}{\nu m}\sum_i \xi_i - \rho \tag{6}$$

$$- \sum_i \alpha_i (\langle \mathbf{w}, \Phi(\phi_i) \rangle - \rho + \xi_i) - \sum_i \beta_i \xi_i, \quad \alpha_i, \beta_i \geq 0$$

Setting the derivatives with respect to $\mathbf{w}, \boldsymbol{\xi}, \rho$ to zero:

$$\mathbf{w} = \sum_i \alpha_i \Phi(\phi_i), \quad \alpha_i = \frac{1}{\nu m} - \beta_i \leq \frac{1}{\nu m}, \ \sum_i \alpha_i = 1 \tag{7}$$

This transforms the decision function into a kernel expansion $f(\phi) = \text{sgn}\left(\alpha_i k(\phi_i, \phi) - \rho\right)$. Finding the parameters α_i is achieved through the dual one–class quadratic program:

$$\min_{\mathbf{a} \in \mathbb{R}^m} \quad \frac{1}{2}\sum_{ij} \alpha_i \alpha_j k(\phi_i, \phi_j) \tag{8}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq \frac{1}{\nu m}, \ \sum_i \alpha_i = 1$$

while the last required parameter ρ can be recovered by $\rho = \langle \mathbf{w}, \Phi(\phi_i) \rangle = \sum_j \alpha_j k(\phi_j, \phi_i)$.

3) Bayesian Encoding of Anomaly Detection on Occupancy Maps: Exploiting the capacity of the robot to use multi–view consensus by observing the same region from different perspectives and in order to enable anomaly–aware path planning, the detected anomalies per image c_i^r are utilized to derive an Anomaly–annotated 3D occupancy map $\mathcal{A}(\mathcal{M})$ [25] by appending camera anomaly detections to the map voxels $m \in \mathcal{M}$. To conduct this step, first anomaly is localized in the image under the assumption that it is due to the presence of an object in the environment which is then detected through a saliency proto object-segmentation algorithm [10, 26]. Provided the anomaly detections $f(c_i^r)$ derived per image, the robot pose $\mathbf{p}_i$ at the corresponding viewpoint, the transformation between the camera coordinate frame $\mathcal{K}$ and the body–fixed frame $\mathcal{B}$, the camera model and the occupancy map $\mathcal{M}$, geometric projection allows to derive a measurement z_t about the subset of pixels $c_i^{r,a}$ of the anomaly/saliency map A_i that refers to each of the visible leaf voxels of the occupancy map from $\mathbf{p}_i$. At every associated i–th update of the map, each voxel maintains a confidence value for it being "normal" or "abnormal".

To update this state of each voxel, the binary Bayes filter [27] is utilized to calculate the respective probability. More specifically, the belief update step in the binary Bayesian filter can be simplified using *logOdds* (L) and updated recursively. The *logOdds* function takes the form:

$$\ell(m) = \log \frac{p(m)}{1 - p(m)} \tag{9}$$

where $p(m)$ is the probability of a voxel m being abnormal. The belief is then recovered as:

$$bel(m) = \frac{1}{1 + e^{-\ell(m)}} \tag{10}$$

Then through the Bayesian formulation, a recursive update rule for the *logOdds* function is derived:

$$\begin{aligned} \ell_t(m) &= \ell_{t-1}(m) + \log \frac{p(m|z_t)}{1 - p(m|z_t)} - \log \frac{p(m)}{1 - p(m)} \\ &= \ell_{t-1}(m) + \ell(m|z_t) - \ell(m) \end{aligned} \tag{11}$$

where the conditional probability $p(m|z_t)$ is called the inverse sensor model since we aim to infer the outcome from one measurement z_t (of a map area corresponding to an abnormal voxel). Similarly, the probability $p(m)$ represents our prior guess for the state of voxel m.

B. Anomaly–aware Exploration Path Planning

The online anomaly detection scheme described in this work was further merged with the autonomous planning and information gathering capabilities of aerial robots. A 2–step path planning scheme [26, 28] is employed in order to ensure full exploration of an initially unknown space and maximum entropy of re–observation of areas of anomaly. During every iteration, the path planner simultaneously optimizes for exploration, as well as for sampling of viewpoints re–observing abnormal voxels from robot configurations that present maximum entropy - in terms of distance and heading to an abnormal voxel - compared to prior steps. To achieve this goal, the planner first samples a random tree (using RRT [29]) of admissible robot position and heading configurations ($\boldsymbol{\xi} = [x, y, z, \psi]$) and identifies the branch that optimizes for exploration of new volume:

$$\mathbf{ExplorationGain}(\sigma^E) = \sum_k \left(\mathbf{Visible}(\mathcal{M}, \xi_k) e^{-\lambda c(\sigma^E_{k-1,k})} \right) \tag{12}$$

where $\sigma^E_{k-1,k}$ are all the path segments of σ^E, $c(\cdot)$ its length, and $\lambda > 0$. The first viewpoint of this branch and the associated robot state $\boldsymbol{\xi}_\mu$ is selected but the way towards it is derived by a second planning step that samples a new random tree connecting the current robot configuration $\boldsymbol{\xi}_{\mu-1}$ with $\boldsymbol{\xi}_\mu$. The branch of this tree that maximizes the following objective related to the entropy of abnormal voxels re–observation is then considered as optimal.

$$\begin{aligned} \mathbf{EntropyGain}(\sigma^A_\alpha) &= \sum_{m \in \mathbf{V}(\sigma^A_\alpha)} \Delta E \left(\log 1/p_m \right) \\ &= \sum_{m \in \mathbf{V}(\sigma^A_\alpha)} \left[\sum_i^{k_m} \left(-p^i_m \log p^i_m \right) - e_m \right] \end{aligned} \tag{13}$$

where $\mathbf{V}(\sigma^A_\alpha)$ represents the set of abnormal voxels visible from σ^A_α, k_m corresponds to the number of observations of voxel m from σ^A_α, p^i_m represents the probability in a 2D histogram of viewing distances and headings of voxel m contributed by a single observation (vehicle viewpoint), $E(\log 1/p^i_m)$ represents the expected value of information (where as information bin, the value $1/p^i_m$ is considered) and ΔE is the gain in information for a certain voxel given the entropy of prior observations of that voxel e_m, while $\sum_{m \in \mathbf{V}(\sigma^A_\alpha)}$ indicates that this calculation takes place for all the voxels that are perceived by the sampled path σ^A_α. To calculate this value, the camera FoV and the effective perception distance are accounted for, alongside calculation of the distance $\zeta_{n_{\alpha,i},m}$ between every sampled vertex to each perceivable voxel. Among the admissible branches, the one that maximizes this entropy gain (relating to the distance and heading of abnormal regions reobservation) is selected and executed by the robot. The process is then repeated in a receding horizon fashion. Importantly, for a branch to be admissible it must further satisfy the dynamic and time budget constraints of the robot as discussed below.

1) Planning subject to Dynamic and Time Budget Constraints: In the second planning step, the objective is that of maximizing the entropy of (re–)observation for abnormal voxels in the map. However, the planned robot paths must not only be collision–free but also feasible with respect to the constraints imposed by the dynamics (maximum yaw rate) and the time budget of the robot. To efficiently address this problem, in this work we propose a sampling policy over the $[x, y, z, \psi]$–space that accounts for the vehicle constraints. It is facilitated through two steps, namely one for sampling positions and one for sampling heading angles.

The first step is responsible to *sample robot positions* for paths that connect the initial and final robot configurations from the first step of the planner, while simultaneously satisfying a time budget t_{allow}. To limit the search space, the locus of all feasible points can be shown to be a 3D space enclosed by an ellipsoid. Given that A and B are positions of the initial and final robot configuration, r is the maximum extent ratio that rules how much the robot can diverge from the shortest path $\overrightarrow{AB}$ (subject to t_{allow} and constant velocity assumption), the locus of possible sampling points M is:

$$\left\| \overrightarrow{MA} \right\| + \left\| \overrightarrow{MB} \right\| \leq r \left\| \overrightarrow{AB} \right\| \tag{14}$$

To simplify, let us define a new coordinate system $\mathcal{C}'$ with $A' = [0, 0, 0]$ and $B' = [d, 0, 0]$, where $d = \left\| \overrightarrow{AB} \right\|$. Solving in $\mathcal{C}'$, the updated constraint $\left\| \overrightarrow{M'A'} \right\| + \left\| \overrightarrow{M'B'} \right\| \leq r \left\| \overrightarrow{A'B'} \right\|$ takes the form:

$$\sqrt{p^T_{M'} p_{M'}} + \sqrt{(p_{M'} - p_{B'})^T (p_{M'} - p_{B'})} \leq r \sqrt{(p_{A'} - p_{B'})^T (p_{A'} - p_{B'})} \tag{15}$$

where $p_\star$ here represents point coordinates. Through algebraic manipulation, the ellipsoid equation is derived:

$$\frac{4}{r^2 d^2} (x_{\mathcal{C}'} - \frac{d}{2})^2 + \frac{4}{d^2 (r^2 - 1)} y^2_{\mathcal{C}'} + \frac{4}{d^2 (r^2 - 1)} z^2_{\mathcal{C}'} \leq 1 \tag{16}$$

where $[x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C}]$ is the coordinate of M' in $\mathcal{C}'$. In spherical coordinates we have

$$x_{\mathcal{C}'} = a \cos(\gamma) \cos(\delta),\ y_{\mathcal{C}'} = b \cos(\gamma) \sin(\delta),\ z_{\mathcal{C}'} = c \sin(\gamma) \tag{17}$$

where $|\gamma| \leq \pi/2$, $|\delta| \leq \pi$, $a = rd/2$, $b = c = d\sqrt{r^2 - 1}/2$.

The second part of the proposed approach to satisfy the constraints of the robot related to accounting for a *maximum yaw rate* $\dot{\psi}_{\max}$ represented as a maximum change in yaw $\Delta\psi_{\max}$ over certain time, and a maximum linear velocity $v_{\max}$. Given a path consisting of $n + 1$ vertices, with the first newly sampled node having a heading angle ψ_1 and the last node a heading ψ_n, alongside the respective edge lengths $d_1, d_2, ..., d_n$ then the goal is to sample the admissible heading changes $\Delta\psi_1, \Delta\psi_2, ..., \Delta\psi_n$. The first condition takes the form:

$$\psi_n = \psi_0 + \sum_{i=1}^{n} \Delta\psi_i \Rightarrow \sum_{i=1}^{n} \Delta\psi_i = \psi_n - \psi_0 \tag{18}$$

The second condition that further accounts for an overall allowed time t_{allow}, takes the form:

$$\sum_{i=1}^{n} t_i \leq t_{allow}, \quad t_i = \max\{ \frac{d_i}{v_{max}}, \frac{|\Delta\psi_i|}{\dot{\psi}_{max}} \} \tag{19}$$

Since the function $t_i = g(\psi_i)$ is U-shaped and symmetric, a parabola $t_i = a_i \Delta\psi^2 + b_i$ may be utilized to simplify it:

$$\sum_{i=1}^{n} a_i \Delta\psi^2 \leq t_{allow} - \sum_{i=1}^{n} b_i \tag{20}$$

From Equations (18) and (20) it is derived that admissible solutions lie on the intersection of a hyperplane with an ellipsoid which can be calculated using analytic methods [30].

IV. EXPERIMENTAL EVALUATION

To evaluate the proposed approach for aerial robotic anomaly detection and anomaly–aware path planning, two experimental studies were conducted. The first study related to outdoors surveillance and anomaly detection over a desert-like environment ("Desert" test) by following a pre–defined path (evaluating purely the anomaly detection scheme). In this mission, all training and test data are collected by a nadir facing camera–equipped DJI Matrice 100. The second study related to autonomous anomaly detection and anomaly–aware exploration inside a parking lot ("Parking Lot" test). In this mission, all training and test data are collected by an oblique–view camera onboard a GPS–denied navigation-capable hexarotor. For both experiments the training data refer to environments similar but not identical to those of the test phase. Prior to presenting the results for each mission, a data–driven analysis regarding the selection of AlexNet–based features as compared to other options is discussed.

A. Feature Selection for Anomaly Detection

To support the design process underlying this research, a comparison of the performance achieved by using different features for one–class classification using SVMs was conducted. The comparison took place using the camera data from the Desert and Parking Lot datasets. Training data are collected from analogous landscapes as the Desert mission and similar parking lots as the Parking Lot mission (but not identical ones). For this evaluation, deep learning–based features extracted based on AlexNet [19], VGG16 [20], VGG19 [20], ResNet50 [21], and InceptionResNet [22]) pre-trained using ImageNet dataset [31] were utilized, alongside a visual bag of SIFT features [23]. Table I details the layer in the relevant deep networks from which the features are extracted from, while the SVM training parameters were kept identical across all this analysis. The selection of the layer to consider for feature extraction relates to the role of each in the classification process and the need to maintain a tractable dimensionality for SVM training. Evaluating these features against the collected Desert and Parking Lot datasets and training the associated one–class SVMs it was found that AlexNet provided superior performance with less false negatives and false positives, while its computational cost - lower than other deep learning–based solutions such as using VGG19 - allowed online operation. Figure 3 present the relevant results. Beyond the poor performance from the Bag-of-Features based on SIFT, it was further observed that results from certain state-of-the-art models including ResNet50 and InspectionResNet were also inadequate and inferior compared to those of using AlexNet. This result is aligned with a finding from [32] in which the authors explained that the performance of image classification tasks using transfer learning methods might be degraded due to representation specificity. In further detail, the average pooling layer used to extract features from ResNet50 and IncepetionResnet in this study is preceding the last layer of those networks and therefore features learned from this layer appear to be less generalizable to images collected from dissimilar environments (e.g., our desert or parking lot environments versus object–centric images from ImageNet). On the contrary, features extracted from the fully connected layer FC6 of AlexNet, VGG16, and VGG19 generalize better since FC6 is three (3) layers ahead from the last layer. The above analysis guided the selection of AlexNet–based features for both of the experimental studies in this paper.

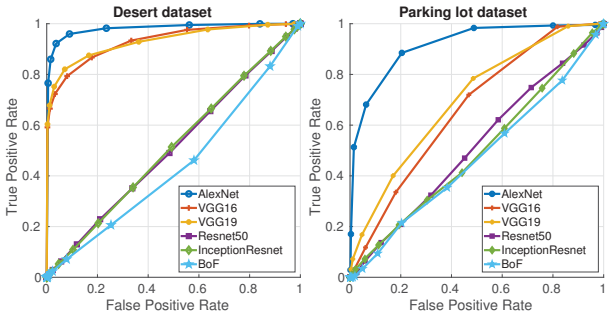

Fig. 3. Receiver Operating Characteristic (ROC) curves from the proposed one-class SVM classifier with respect to different feature extraction algorithms. Pretrained CNN models including AlexNet, VGG16, VGG19, ResNet50, and InceptionResNet were employed to extract features from color images. The visual bag-of-features method using the SIFT descriptor was also evaluated due to its relatively lightweight computation. The left and right plot show ROC curves for the desert and the parking lot datasets respectively. The y-axis of the plots represents the *True Positive Rate* (TPR) value which is the percentage of anomaly correctly identified. The x-axis indicates the *False Positive Rate* (FPR) which is the percentage of normal data incorrectly classified as anomaly. As shown, the SVM model trained with features extracted from the AlexNet FC6 layer outperforms other models. In particular, in the desert dataset, the TPR of the classifier using AlexNet–based features achieves a value equal 95% at FPR equal 9%, while TPR in the parking lot dataset is 88% and FPR 20%.

TABLE I
DETAILS ON EVALUATED FEATURES.

Pretrained Model	Feature Layer	Feature Size
AlexNet	FC6	4096
VGG16 & VGG19	FC6	4096
ResNet50	Avg Pool	2048
InceptionResNet	Avg Pool	1536
SIFT/BoF	N/A	500

B. Outdoors Surveillance and Anomaly Detection

For this experiment, a nadir facing camera–equipped DJI Matrice 100 micro aerial vehicle was used to first collect training data over a desert-like environment in Northern Nevada. The one–class classification technique described in Section III was utilized to detect anomaly during two test missions where the vehicle performed a predefined inspection trajectory over a *different* desert–like area. Indicative

camera frame results on anomaly detection are shown in Figure 4. Statistical results for the overall classification task are shown in Figure 5 and demonstrate the accuracy of the proposed method for anomaly detection. Since the system was trained over images from the nadir facing camera observing the desert environment, as anomaly we introduced a car, a blanket, and humans. This test serves to verify the accuracy of anomaly detection on aerial robotic image data.

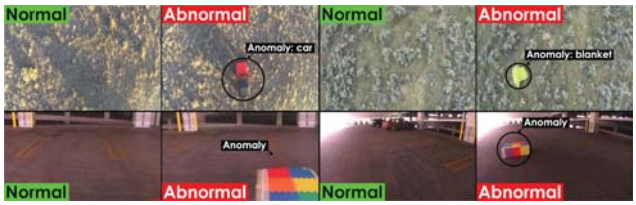

Fig. 4. Indicative frames from a) the two surveillance operations of an aerial robot above a desert-like environment, and b) from the exploration mission inside a parking lot. Within the desert dataset, camera frames capturing mostly soil and bushes are correctly detected as normal, while when the frame contains objects not present in the training data (such as cars, or a blanket) it is considered abnormal. In the parking lot dataset, most frames not containing the colorful box are detected as normal, while those containing are detected as abnormal. Map fusion and cognizant viewpoint planning contribute significantly to filter out any false detections.

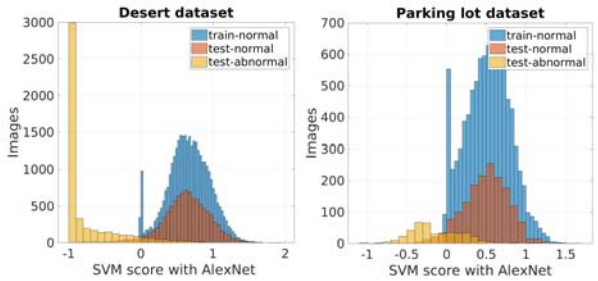

Fig. 5. Classification scores of the proposed anomaly detection algorithm. Threshold for the SVM's decision function was set at zero. Hence, negative score leads the SVM to report abnormal ("-1") and positive score means normal ("+1"). The left part of the figure shows the results on the desert dataset. This dataset includes a training set of 35033 normal images and a test set with 14770 images (10079 normal images and 4691 abnormal images). The detection accuracy for the normal and abnormal data in the test set is 95.67% and 93.43% respectively. The right part of the figure presents results on the parking lot dataset. This set consists of 9249 normal images for training and 2070 test images (1653 normal and 417 abnormal images). The accuracy achieved in this set is 95.28% for normal images and 66.19% for abnormal images. The test sets in both cases included images collected environments similar but different to those of the training data.

C. Autonomous Anomaly Search inside a Parking Lot

A more challenging and holistic test was subsequently conducted and referred to the autonomous exploration of a parking lot environment, while simultaneously detecting anomaly, encoding this information on the online reconstructed occupancy map, and performing anomaly–aware planning to maximize the entropy of observations over the abnormal areas. For this test a hexarotor aerial robot was employed and all anomaly detection, GPS–denied localization and mapping, anomaly encoding on the map, path planning and control operations were running onboard the robot based on a NUC5i7RYH processing module. A linear MPC controller is employed for low-level position control [33, 34]. Localization and mapping relied on a stereo visual–inertial solution [35]. The anomaly detection algorithm is developed based on LibSVM [36] and a pretrained model of AlexNet in TensorFlow [37]. On average, the computation time for feature extraction based on AlexNet using the onboard CPU is 500ms and another 6ms is spent for the inference step of the one-class SVM. Furthermore, the complexity analysis for the proposed planning algorithm in this work is similar to those described in [2, 26] with a dominant term primarily depending on the collision checking step $\mathcal{O}(N_{\mathbb{T}} \times d_{avg}/r \times \log(V^{\mathcal{A}}/r^3))$; where $N_{\mathbb{T}}$ is the number of nodes of the random tree, d_{avg} is an average length of each edge, and r is voxel resolution of the occupancy map $\mathcal{A}(\mathcal{M})$.

Training data in this experiment were collected from different parking lots but not from the one within which the test took place. Naturally, cars in the environment were also different from those in the training data. Indicative anomaly classification results are shown in Figure 4. Figure 6 presents instances of the anomaly–aware cognizant exploration path planning. As shown, the robot identifies anomaly–aware planning steps within which the entropy of observation of abnormal voxels is maximized (through variation in viewpoint distance and heading). For this experiment, the 2D histogram of distances and orientations from the abnormal voxels uses bins of 1m distance, and 22.5° orientation step size. Statistical results are also shown on Figure 5.

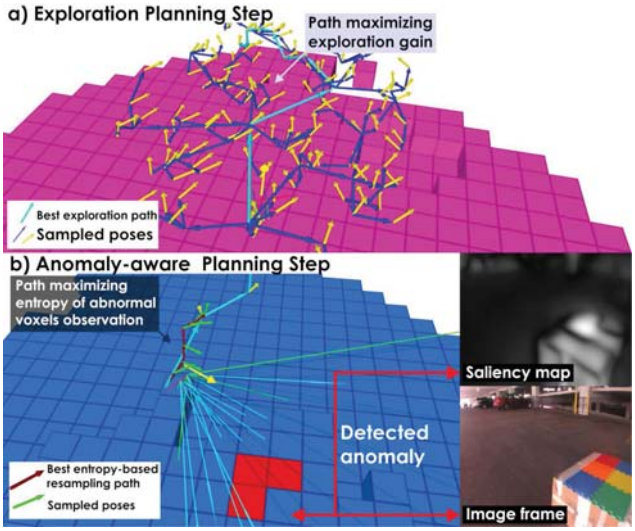

Fig. 6. Indicative results from the anomaly-aware exploration mission in the parking lot. Figure a) is an instance of the first sampling based planning step in a height map where the main objective is exploring towards unknown areas. Figure b) presents an optimal resampling path to maximize the entropy gain observed over abnormal voxels. As shown in the figure, red occupied voxels (abnormal regions) were reobserved during the second planning step.

As depicted, the accuracy in this type of more complex environment is reduced compared to that of the desert test. It was found that intense variations in light (i.e., looking

directly towards the sun from the sides of the parking lot that were open) occasionally led to false positives, while false negative is rather sparse. This is mainly because images acquired in such real-life settings contain extremely bright regions from the sunlight which are not always present in the training data. It is also not trivial to introduce relevant data with intense light conditions in the training phase since it will add undesirably noisy samples to the one-class SVM model. Hence, the filter step on the map level using Bayesian update is necessary and essential to ensure a consistent anomaly detection map for the second planning step.

V. CONCLUSIONS

A methodology for unsupervised anomaly detection, map encoding and anomaly–aware path planning using aerial robots was presented. The method relies on deep learned features, one–class SVMs, multi-view anomaly consensus through Bayesian fusion on the map, and a 2–step path planner that ensures unknown environment exploration and simultaneous high–entropy reobservation of abnormal regions. Experimental studies related to aerial robotic surveillance above a desert and autonomous exploration inside a parking lot are presented to evaluate the method and its performance.

REFERENCES

[1] C. Papachristos, S. Khattak and K. Alexis, "Autonomous exploration of visually-degraded environments using aerial robots," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2017.

[2] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova and R. Siegwart, "Receding horizon "next-best-view" planner for 3d exploration," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2016. [Online]. Available: https://github.com/ethz-asl/nbvplanner

[3] S. Nuske, S. Choudhury, S. Jain, A. Chambers, L. Yoder, S. Scherer, L. Chamberlain, H. Cover, and S. Singh, "Autonomous exploration and motion planning for an unmanned aerial vehicle navigating rivers," *Journal of Field Robotics*, vol. 32, no. 8, pp. 1141–1162, 2015.

[4] B. Grocholsky, J. Keller, V. Kumar, and G. Pappas, "Cooperative air and ground surveillance," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 16–25, 2006.

[5] M. Popovic, T. Vidal-Calleja, G. Hitz, I. Sa, R. Siegwart, and J. Nieto, "Multiresolution mapping and informative path planning for uav-based terrain monitoring," *arXiv preprint arXiv:1703.02854*, 2017.

[6] C. Papachristos, M. Kamel, M. Popović, S. Khattak, A. Bircher, H. Oleynikova, T. Dang, F. Mascarich, K. Alexis, and R. Siegwart, "Autonomous exploration and inspection path planning for aerial robots using the robot operating system," in *Robot Operating System (ROS)*. Springer, 2019, pp. 67–111.

[7] S. Khattak, C. Papachristos and K. Alexis, "Keyframe-based direct thermal–inertial odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2019.

[8] S. Khattak, C. Papachristos, and K. Alexis, "Visual–thermal landmarks and inertial fusion for navigation in degraded visual environments," in *2019 IEEE Aerospace Conference*. IEEE, 2019.

[9] T. Dang, C. Papachristos, and K. Alexis, "Autonomous exploration and simultaneous object search using aerial robots," in *2018 IEEE Aerospace Conference*. IEEE, 2018, pp. 1–7.

[10] S. Frintrop, T. Werner, and G. Martin Garcia, "Traditional saliency reloaded: A good old model in new shape," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, 1998.

[12] J. K. Tsotsos, *A computational perspective on visual attention*. MIT Press, 2011.

[13] J. Gottlieb, "Attention, learning, and the value of information," *Neuron*, vol. 76, no. 2, pp. 281–295, 2012.

[14] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 5, pp. 802–817, 2006.

[15] T. Dang, C. Papachristos, and K. Alexis, "Anomaly Detection using Aerial Robots Open Dataset." [Online]. Available: https://www.autonomousrobotslab.com/uad-dataset.html

[16] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010.

[17] J. Kim and K. Grauman, "Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2921–2928.

[18] A. Dimokranitou, "Adversarial autoencoders for anomalous event detection in images," Ph.D. dissertation, Purdue University, 2017.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, 2014.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning." in *AAAI*, vol. 4, 2017, p. 12.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, 2004.

[24] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[25] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013.

[26] T. Dang, C. Papachristos, and K. Alexis, "Visual saliency-aware receding horizon autonomous exploration with application to aerial robotics," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2018.

[27] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[28] C. Papachristos, S. Khattak, and K. Alexis, "Uncertainty–aware receding horizon exploration and mapping using aerial robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2017.

[29] S. LaValle and J. Kuffner, J.J., "Randomized kinodynamic planning," in *IEEE ICRA*, 1999, pp. 473–479 vol.1.

[30] Alex A. Kurzhanskiy, and Pravin Varaiya, "Technical Report No. UCB/EECS-2006-46, Ellipsoidal Toolbox."

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. Ieee, 2009, pp. 248–255.

[32] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in neural information processing systems*, 2014, pp. 3320–3328.

[33] M. Kamel, T. Stastny, K. Alexis, and R. Siegwart, "Model predictive control for trajectory tracking of unmanned aerial vehicles using ros," *Springer Book on Robot Operating System (ROS)*.

[34] C. Papachristos, T. Dang, S. Khattak, F. Mascarich, N. Khedekar, K. Alexis *et al.*, "Modeling, control, state estimation and path planning methods for autonomous multirotor aerial robots," *Foundations and Trends® in Robotics*, vol. 7, no. 3, pp. 180–250, 2018.

[35] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 298–304.

[36] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[37] "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/