Learning to Detect Aircraft for Long Range, Vision-Based Sense and Avoid Systems

Jasmin James, Jason J. Ford, and Timothy L. Molloy

***Abstract*—The commercial use of unmanned aerial vehicles (UAVs) would be enhanced by an ability to sense and avoid potential mid-air collision threats. In this paper we propose a new approach to aircraft detection for long range, vision-based sense and avoid. We first train a deep convolutional neural network to learn aircraft visual features using flight data of mid-air, head-on near collision course encounters between two fixed-wing aircraft. We then propose an approach that fuses these learnt aircraft features with hand-crafted features that are used by the current state of the art. Finally, we evaluate the performance of our proposed approach on real flight data captured from a UAV where it achieves a mean detection range of 2527m and a mean detection range improvement of 299m (or 13.4%) compared to the current state of the art with no additional false alarms.**

***Index Terms*—Aerial Systems: Perception and Autonomy, Computer Vision for Automation, Deep Learning in Robotics and Automation**

I. Introduction

THE emerging global market for commercial unmanned aerial vehicle (UAV) services is anticipated to reach $\$21.47$ billion by 2021 with the potential for UAVs to be used in many important sectors including infrastructure, agriculture, transport, security, entertainment and media, insurance, telecommunication and mining [1]. There have been significant efforts to safely integrate routine UAV operations into the national airspace so that they do not compromise the existing safety levels [2]. One of the most significant risks that UAVs face, and pose, is mid-air collision.

The national airspace is heavily regulated with strict rules and safety layers designed to mitigate the risk of mid-air collisions. The first few safety layers involve operational procedures, air traffic management, and cooperative collision avoidance systems (for aircraft that make their presence known). The final safety layer is for potential mid-air collisions that are not successfully managed by the other layers. For manned aircraft, this final safety layer corresponds to human pilots using their eyes and judgment to see and avoid potential mid-air collision threats.

Sense and avoid (SAA) refers to the implied regulatory requirement that UAVs be capable of sensing and avoiding potential non-cooperative mid-air collision threats with a proficiency matching or exceeding that of human pilots. Human pilots that are alerted to the presence of potential collision threats are reported to be able to detect them with an 86% success rate at a median range of 2593m [3]. The development of systems capable of matching (and exceeding) the reported performance of human pilots and meeting the implied SAA regulatory requirement is one of the key technical challenges hindering the routine, standard and flexible operation of UAVs in the national airspace [2].

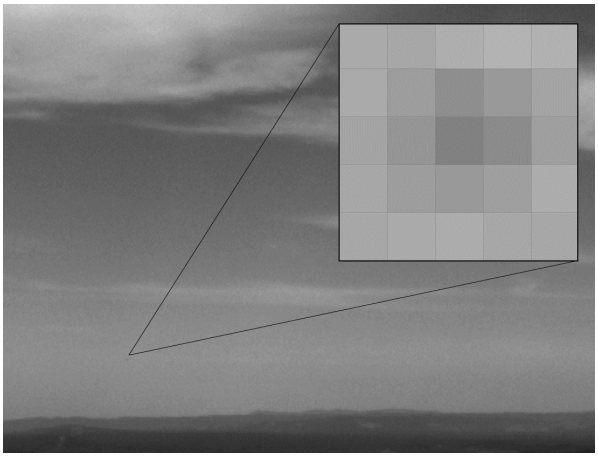

Fig. 1: An illustration of an aircraft at a long range where it visually appears as a small number locally dim pixels.

Machine vision has recently been identified as a promising technology for detecting potential collision threats as vision sensors have size, weight, power and cost advantages over other sensing approaches such as radar for small to medium sized UAVs [4]. To ensure sufficient time to perform avoidance manoeuvers, it is desirable to detect aircraft at ranges where they appear in the raw images from vision sensors as a small number of slow-moving locally dim pixels that contrast poorly with the background (as shown in Figure 1). The signal to noise ratios (SNRs) of raw images containing distant aircraft are therefore very low, and only increase significantly when the aircraft are in close proximity (e.g., SNRs or local detectability values of less than -3dB or 0.5 for ranges greater than 2000m are reported in [5]). In order to detect aircraft at long ranges with very low SNRs, in this paper we propose a new approach to vision-based aircraft detection which uses deep learning to exploit learnt aircraft features.

The key contributions of this paper are:

i) The proposal of a new aircraft detection system which

Manuscript received: May, 3, 2018; Revised July, 24, 2018; Accepted August, 16, 2018.

This paper was recommended for publication by Editor Jonathan Roberts upon evaluation of the Associate Editor and Reviewers' comments.

The authors are with the School of Electrical Engineering and Computer Science, Queensland University of Technology, 2 George St, Brisbane QLD, 4000 Australia. jasmin.james@connect.qut.edu.au, j2.ford@qut.edu.au, t.molloy@qut.edu.au. This work was supported by funding from the Australian Research Council Centre of Excellence CE140100016 in Robotic Vision.

Digital Object Identifier (DOI): see top of this page.

fuses learnt and hand-crafted visual features to boost the SNR of long range aircraft for vision-based SAA.

ii) An experimental evaluation of our proposed system, which demonstrates improvement in detection ranges and false alarm rates relative to the current state of the art.

iii) A comparison of our proposed fused system against several other candidate feature detectors which have previously been used in sense and avoid.

More broadly, the problem of vision-based aircraft detection at long range may be viewed as an instance of dim target detection. The approach and results of this paper provide one of the first compelling applications of fusing learnt and handcrafted features in a dim target detection pipeline. Finally, we note that whilst we will focus on the "sense" aspect of SAA in this paper, specifically vision-based aircraft detection, a variety of automated collision avoidance strategies suitable for use with our approach and other vision-based aircraft detection approaches have previously been proposed to address the "avoid" aspect of SAA (see [4], [6] and references therein).

The rest of this paper is structured as follows. In Section II we discuss the related work in vision-based aircraft detection. In Section III we describe our approach for learning aircraft features and present our proposed vision-based aircraft detection system. In Section IV we experimentally investigate the performance of our proposed system. In Section V we discuss our proposed system and its limitations. Finally, we provide some conclusions in Section VI.

II. Related Work

Numerous approaches to vision-based aircraft detection have been presented in the literature and have exploited advances in the more general field of dim target detection [7]–[10]. The most effective approaches for long range, fixed-wing, vision-based aircraft detection with low false alarms utilise a multi-stage detection pipeline [5], [11]–[14]. Additionally, with advances in hardware over the last decade, there have been several investigations into the use of machine learning [15]–[19] and deep learning [20], [21]. We highlight that the performance of different vision-based aircraft detection systems is characterised by detection rates, detection ranges and false alarms.

A. Dim Target Detection

The problem of dim target detection arises in many applications including space surveillance systems, object tracking systems and many more [9]. Due to their small size and operation in cluttered and complex environments (land, sea, air etc.), it is common for targets to have a low SNR [9]. Most existing methods of dim target detection (in IR and other sensing modalities) appear to have been based on multistage detection pipelines with handcrafted features [7]–[10]. In this paper we exploit this multistage pipeline, as is common for vision based aircraft detection, however in contrast to hand crafted features we explore using learnt features.

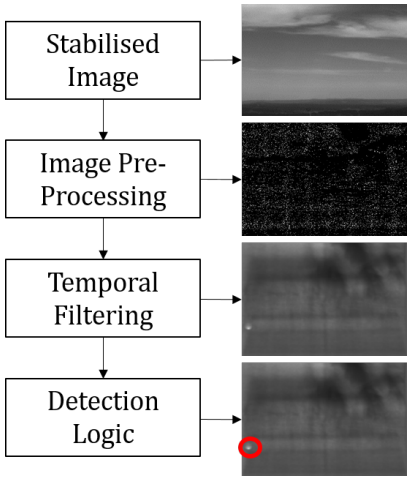

Fig. 2: Key stages of the vision-based aircraft detection pipeline.

B. Multi-Stage Detection Pipeline

There are several variations of multi-stage detection pipelines presented in the literature. As shown in Figure 2, the key stages include image pre-processing, temporal filtering and detection logic [5], [11]–[14].

The goal of the image pre-processing stage is to suppress background clutter and highlight small pixel sized aircraft. Some popular approaches that are proposed in the literature to accomplish this include morphology [5], [11], [12], [14], image frame differencing [12], [13], and machine learning [15].

Due to the low SNR of small pixel sized aircraft, the temporal filtering stage is required to emphasize and extract features that possess aircraft-like dynamics. In [11] a Viterbi-based filtering approach is proposed. The authors report detection ranges of around 6km, however they do not report the false alarm rate. An extended Kalman filter is proposed in [12]; a "valid track" of the aircraft (where the aircraft is consistently detected) is declared at an average detection range of 1747m with an average of 4 false alarms over their tested image sequences. In [5] hidden Markov model (HMM) filtering is used for detection and the authors report detection ranges of at least 1540m with no false alarms.

The detection logic stage aims to utilise the information available from the image pre-processing and temporal filtering stages in order to declare whether an aircraft is present or not. An exponentially weighted moving average test statistic based on the HMM output is used in [5], [13]. As an improvement, [14] proposes a new test statistic which is related to the probability that an aircraft is present in the image. The authors report detection ranges of 2227m with no false alarms.

This multi-stage pipeline of [14] which uses morphological processing, HMM filtering and the aircraft probability test statistic is the current state of the art.

C. Machine Learning

There are several approaches to vision-based aircraft detection that use machine learning. In [15] an approach based on

the Viola and Jones framework is proposed, where a trained AdaBoost cascade using Haar features is used. The authors report a successful detection in around 80% of the tested images with aircraft present. In [16] an approach using spatio-temporal image cubes for classification is proposed with regression-based motion stabilization of local image patches and an Adaboost cascade. The authors report an average precision of 75% on their UAV dataset and 79% on their aircraft dataset. In [17] an AdaBoost-based approach is also proposed however the authors are interested in detecting rotorcraft at close range.

In [18], [19] a multi-stage detection pipeline is proposed which used a support vector machine (SVM) to exploit aircraft visual features. The first stage utilises morphological filtering to highlight potential aircraft in the image. In the next stage, shape descriptors and an SVM based classifier are used to reduce false positives. The final stage involves tracking the remaining potential aircraft over time in order to eliminate remaining false alarms. In this approach an average detection rate of 98% of the tested images with aircraft present out to 8km is achieved with a false alarm rate of 1 every 50 frames.

Despite their potential, prior machine learning approaches have under achieved in detection ranges, false alarm rates or both compared to the state of the art multi-stage detection pipeline of [14].

D. Deep Learning

More recently some deep learning approaches have been explored in the vision-based aircraft detection application. In [20] a deep convolutional neural network (CNN) and a deep belief network are trained to detect larger aircraft in satellite imagery. In [21] the authors propose a deep CNN which is able to detect aircraft in complex backgrounds. The authors achieved a 83% detection rate on the tested images with aircraft present. These deep learning approaches for vision-based aircraft detection have not yet been utilised in a multi-stage detection pipeline.

In this paper we propose using deep learning in a multi-stage detection pipeline to increase detection ranges whilst maintaining low false alarms. We aim to implement a new image pre-processing stage that incorporates learnt aircraft features building on the state of the art multi-stage pipeline of [14].

III. Approach

In this section we present a variant of the SegNet architecture [22] for semantic pixel-wise segmentation applied to vision-based aircraft detection. We first describe our training data and labeling approach. We then outline our network architecture and training regime and describe our testing methodology. Finally, we present our proposed vision-based aircraft detection system which involves fusing our learnt aircraft features with morphological processing.

A. Training Data and Labeling

For vision-based, long range sense and avoid we aim to detect low SNR, long range aircraft and therefore require a very specific dataset for training. We used the aircraft data presented in [5] which consisted of 12 head-on and tail chase encounters between a Cessna 172 (camera aircraft) and a Cessna 182 (target aircraft) captured with a 5mm lens. We note that while the system examined in [5] only used 7 cases, we were able to use the additional 5 cases where the stabilisation was poor as stabilised data is not necessary for training purposes. The data encompassed a range of different cloud conditions from blue sky environments to very textured clouds as seen in Figure 3. In total we used 15874 images of target aircraft data. In addition to this we used 2925 images of ground clutter with no aircraft present. We note that all our images were greyscale.

Fig. 3: A example of the background variation in different cases used in the training data. The left image is blue sky and the right image is cloudy sky.

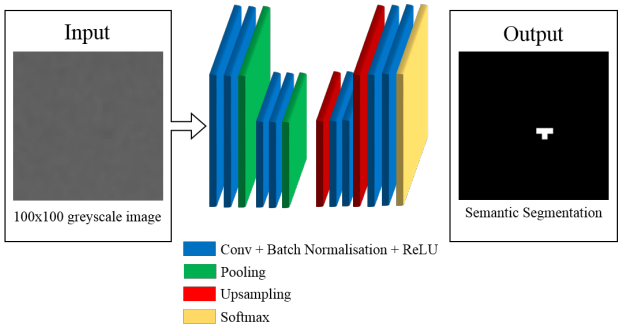

Fig. 4: An illustration of the network architecture for the SegNet variation used. The input to the system is a greyscale image, there are two encoders and two decoders which are fed into a softmax layer for pixel-wise classification.

For all images, each pixel was classified as either aircraft or non-aircraft. The images were manually labeled by a human and pixels were classified as an aircraft if they were visually distinguishable from the background pixels. This labeling was subject to human error. We decided on an input image size of $100 \times 100 \times 1$ and randomly cropped around the aircraft so that the aircraft was present in various locations in the images.

B. Network Architecture and Training

Our network is a variation of the SegNet architecture proposed in [22] for pixel-wise segmentation. The architecture consists of an encoder network, a corresponding decoder network and a pixel-wise classification layer. The architecture is fully convolutional (i.e., it has no fully connected layers) which allows us to efficiently train on our dataset of cropped

TABLE I: The training options for our network.

Momentum	0.9
Initial Learn Rate	0.001
L2 Regularisation	0.0005
Max Epochs	500

$100 \times 100 \times 1$ images whilst testing on datasets with larger images.

We found that the SegNet proposed in [22] over-fit to our training dataset, so we modified the network to have 2 encoders as seen in Figure 4. Each encoder layer performs 64 $3 \times 3 \times 1$ convolutions with stride $[1\ 1]$. These are then batch normalised and an element-wise rectified-linear non-linearity (ReLu) is applied. The corresponding decoder upsamples its input feature maps using the memorized max pooling indices, see [22] for more details. The output of the final decoder is fed to a softmax layer which classifies each pixel independently into 2 classes: aircraft and non-aircraft. For an input image Y, the output of our network is a binary image $S(Y)$ with the ith pixel value given by

$$S(Y)^i = \begin{cases} 1 & \text{if} \quad \text{aircraft} \\ 0 & \text{if} \quad \text{non-aircraft.} \end{cases} \tag{1}$$

To train our network, we weighted the loss function to balance the aircraft and non-aircraft classes. The weights were initialised using the common 'MSRA' weight initialization method [23]. The optimisation algorithm used for training was stochastic gradient descent with momentum and the training options are presented in Table I. Additionally, we used data augmentation to provide more examples to the network, we used random left/right reflection and random X/Y translation of ± 10 pixels.

C. Testing

We test on two different datasets that are described below: a previously published Project ResQu dataset [14], [24] (dissimilar to our training data), and a new unpublished Project Smart Skies dataset (similar to our training data). This allows us to evaluate the performance of our proposed system across data captured with different aircraft, camera lens, and weather conditions and potentially gives some insight into the sensitivity of our proposed system.

1) ResQu dataset (Captured from a UAV): This dataset consists of $1024 \times 768 \times 1$ pixel image sequences captured with a 5mm lens, collected during 15 mid-air head-on near collision course encounters between two fixed-wing aircraft: a ScanEagle UAV (camera aircraft) and a Cessna 172 (target aircraft). The data was captured, processed and stabilised using a GPS/INS attitude solution at 9Hz on a NVIDIA GeForce 9400M GPU with 16 graphical cores. See [24] for more details of the flight experiments. We note that we maintain the same numbering conventions as in [14], [24] for comparison purposes.

We highlight that this is a different dataset to our training data as: it was captured with a different lens (5mm lens with a $2\times$ extender on the training dataset and 5mm lens on this testing dataset); had different flight paths (the training dataset has head on encounters and tail chase encounters whilst this testing dataset only has head on); and was captured in different conditions (some cases in this testing dataset were captured in the rain).

TABLE II: The Smart Skies dataset collection times and geometry

Label	Collection Date (AEST)	Geometry
S1	February 28 2012 (10:48:07)	Head-on
S2	February 28 2012 (11:14:09)	Head-on
S3	February 28 2012 (11:27:24)	Head-on
S4	February 28 2012 (11:41:20)	Head-on

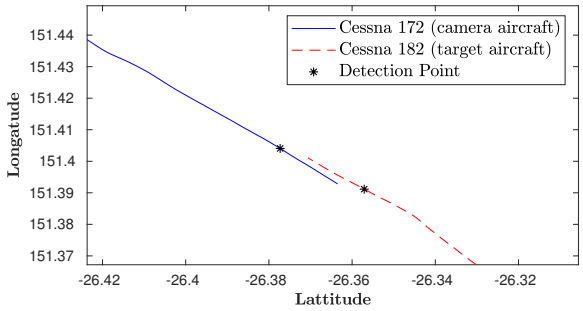

Fig. 5: The flight path of the camera and target aircraft for case S1 of the Smart Skies dataset. The target aircraft is detected at a range of 2588m. We note that there is an altitude separation between the aircraft of approximately 150m (no lateral maneuver was required).

2) Smart Skies dataset: This dataset consists of $1024 \times 768 \times 1$ pixel image sequences captured with a 5mm lens with a $2\times$ extender collected during 4 mid-air head-on near collision course encounters between two fixed-wing aircraft: a Cessna 172 (camera aircraft) and a Cessna 182 (target aircraft). The data was captured at 15Hz and stabilised using a GPS-INS attitude solution. See [5] for more details of the data collection equipment equipment.

This data was captured over the skies of Kingaroy in Queensland in Australia (the approximate GPS location of the testing site is $26°22.4700$ S, $151°24.1000$ E). The collection times and geometry of the flight experiments are described in Table II, and an example of the flight path for case S1 is presented in Figure 5.

We highlight that this is a similar dataset to our training data. It is captured with the same lens and on the same aircraft.

3) SegNet Computation and Performance: We implemented our network in MATLAB using the function "segnetLayers". The implementation processed approximately 2 frames per second on a PC running Ubuntu 14.04 with a 3.4GHz Intel Core i7-6700 CPU and NVIDIA GeForce GTX 1070 GPU. With a suitably chosen threshold value of 0.99, our network was successfully able to extract aircraft features from the image sequences with a low number of false positives.

D. Proposed System: Fusion of SegNet with Morphology

When an aircraft first emerges in an image sequence it is typically small in size and does not exhibit significant visual features beyond a few pixels. Morphological processing is the current state of the art at extracting small pixel-sized aircraft [12], [14], however it also extracts a high number of false positives (on the order of thousands per frame). Morphological processing therefore typically offers very high recall but very low precision. In contrast, our SegNet implementation typically offers high precision but low recall for small pixel-sized aircraft (i.e., SegNet reliably extracts larger sized aircraft but intermittently misses small pixel-sized aircraft).

To exploit the complementary precision-recall properties of morphological processing and our SegNet implementation, we propose fusing our SegNet implementation with bottom-hat (BH) morphological processing [5] by performing a pixel-wise addition of our SegNet output $S(Y)$ and the BH processed image $BH(Y)$. That is, the ith pixel of our proposed fused output $F(Y)$ is given by

$$F(Y)^i = BH(Y)^i + bS(Y)^i, \tag{2}$$

where $b = 3$ is a boosting factor experimentally selected; we term this SegNet+BH. When the aircraft is present in SegNet, the magnitude of the fused SegNet+BH output will be greater than the outputs of BH and SegNet. In a multi-stage detection pipeline (see Figure 2), the temporal filter accumulates the image pre-processing output. Due to the intermittent increases in the magnitude of the fused SegNet+BH output, we propose using this SegNet+BH fusion as a new image pre-processing stage for use in a multi-stage detection pipeline.

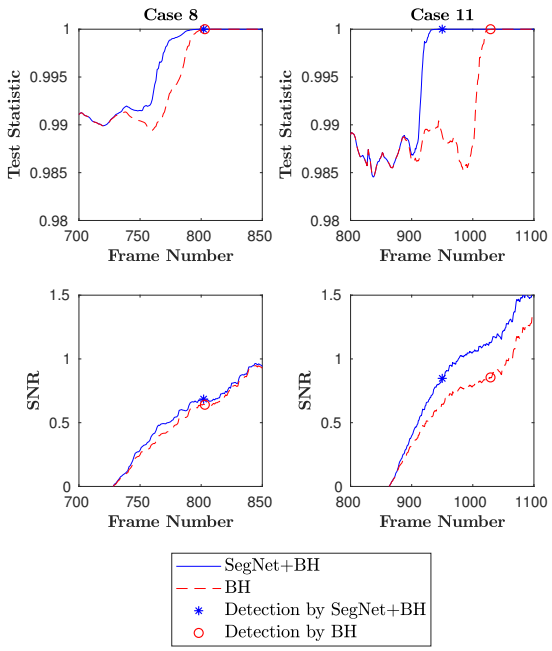

Fig. 6: Illustrative examples of the test statistic (top) and the SNR (bottom) when an aircraft is approaching. In Case 8 our proposed fused SegNet+BH system detects at 2606m (frame 802) and the baseline BH system detects at a similar 2597m (frame 803). In Case 11 our proposed fused SegNet+BH system detects at 2849m (frame 950) while the baseline BH system detects at 2253m (frame 1029).

IV. Results

In this section we evaluate the performance of our proposed fused SegNet+BH image pre-processing stage in a multi-stage detection pipeline. We also evaluate the performance of SegNet fused with several feature detectors previously proposed for SAA [25] as well as an unfused SegNet system.

We note that detection range and false alarm performance varies with the choice of the threshold parameters. For comparison purposes we will identify the lowest threshold for each system that achieves zero false alarms (ZFAs) for our testing data and compare the resulting ZFA detection ranges. We highlight that these thresholds are calculated separately for the two datasets. In practice, detection thresholds could be adaptively selected on the basis of scene difficulty such as proposed in [26].

We first do an extensive evaluation on our (larger) ResQu testing dataset. We then evaluate the performance on our (smaller) Smart Skies testing dataset.

A. Performance Study of Proposed System

We investigate the performance of our multi-stage pipeline with our proposed fused SegNet+BH image pre-processing stage. As seen in Figure 2, the input to our system is a stabilised image. For our image pre-processing stage, we use our proposed SegNet+BH. For the temporal filtering stage and detection logic we follow [14] and implement a HMM filter to calculate the aircraft probability test statistic.

We compare its performance to a state of the art baseline system developed in [14] that utilises a BH image pre-processing stage. Note that both systems use the same temporal filtering stage and detection logic.

We investigate the detection ranges and the SNR at the output of the image pre-processing stage for both systems. We define this aircraft SNR as the maximum intensity (the pixel with the highest value) within a 10×10 pixel grid around the ground truth aircraft location divided by the maximum intensity of the non-aircraft pixels in the output of the image pre-processing stage (everywhere outside the 10×10 pixel grid).

1) Illustrative example: Figure 6 presents an illustrative example of the test statistic and the SNR at the output of the image pre-processing stage as an aircraft is approaching in two cases from our ResQu testing dataset: Case 8 and Case 11. In Case 8 there is minimal improvement by our proposed fused SegNet+BH system which detects at 2606m, while the baseline BH system detects at a similar 2597m. We note that there is only a small boost from the SNR in this case and the detection occurs at similar SNR for both systems. In Case 11 there is significant improvement by our proposed fused SegNet+BH system which detects at 2849m, while the baseline BH system detects later at 2253m.

There is an obvious boost in the SNR from the proposed

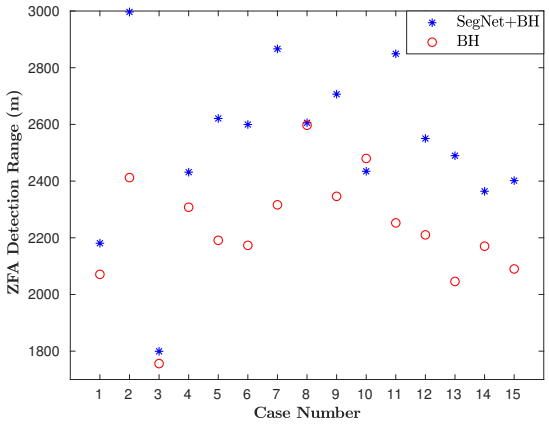

Fig. 7: A comparison of the ZFA detection ranges of our fused SegNet+BH system, and the baseline BH system. The mean ZFA detection ranges and SEOM are 2527m and 75m for the proposed fused SegNet+BH system and 2228m and 52m for the baseline BH system.

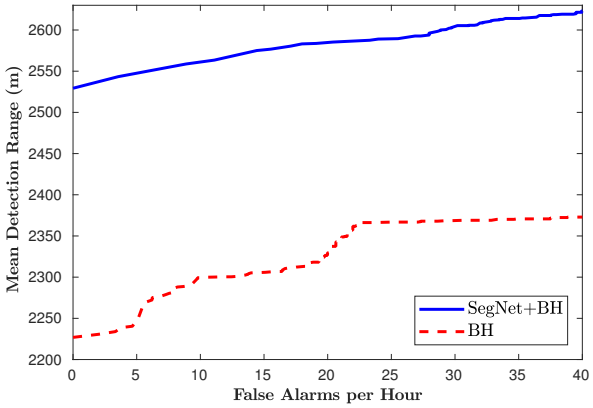

Fig. 8: The mean detection range and false alarm rate of our proposed SegNet+BH system and the baseline BH system for a range of different thresholds. The maximum SEOM is 83m for all tested thresholds.

fused system. We note that even though detection is occurring at different ranges it still occurs at a similar SNR. Notably, detection is occurring when the SNR is less that 1, highlighting the importance of the temporal filtering stage.

2) ZFA detection ranges: We ran both algorithms on the ResQu dataset and the ZFA detection ranges are presented in Figure 7. The mean ZFA detection ranges and standard error of mean (SEOM) are 2527m and 75m for the proposed fused SegNet+BH system and 2228m and 52m for the baseline BH system. Our proposed fused SegNet+BH system improves the mean ZFA detection range by 299m (13.4%) relative to the current state of the art. Additionally, the mean SNR of our proposed fused SegNet+BH system is 1.4361, the SNR of the BH system is 1.3276. On average, our proposed fused SegNet+BH system has boosted the SNR by 0.1085 (8.2%). Note that Case 3 is significantly worse for both algorithms. As reported in [24] this case involved significant flight turbulence and platform motion that was not inertially sensed accurately (degrading the stabilisation of the image sequence).

3) System operating characteristic curves: To examine the detection range and false alarm performance for our ResQu testing dataset we composed system operating characteristic (SOC) curves. A SOC curve is commonly used to evaluate the performance of a vision-based aircraft detection system [5], [13], [14], [18].

Figure 8 presents the mean detection range versus the mean false alarms per hour for a range of different thresholds. The maximum SEOM of the mean detection ranges is 83m for our proposed fused SegNet+BH and system 71m for the baseline BH system. Figure 8 illustrates longer detection ranges for our proposed image pre-processing stage whilst maintaining lower false alarm rates across all tested thresholds.

B. Fusion of SegNet with Feature Detectors

For comparison purposes we compare our proposed SegNet+BH image pre-processing stage to the fusion of SegNet with several candidate features extraction techniques previously used in SAA [25].

1) Candidate feature detectors and adjustable parameters: We compare our proposed SegNet+BH with a pixel-wise addition of SegNet with: features from accelerated segment test (FAST) [27], Harris-Stephens (H-S) corner detector [28], and Shi and Tomasi (S&T) corner detector [29].

We now briefly outline the adjustable parameters for each of the candidate approaches. For FAST the minimum accepted quality of the corners was set to 0.01 and the minimum intensity difference between the corner and surrounding regions was set to 0.01. For H-S and S&T the minimum accepted quality of the corners was set to 0.01, and the Gaussian filter dimension was set to 3. For implementation purposes, we rescaled the outputs of FAST, H-S and S&T to integers between 0 and 5.

2) Fusion results and discussion: For each of the fusion approaches we compare the ZFA detection range where the system correctly declared a detection.

The ZFA detection ranges for the candidate fusion approaches are presented in Table III. Cases with missed detections were marked as "X". The fusion of SegNet+BH is the best performing; it has no missed detections and the best average detection range. The SEOM is similar for all algorithms. It is perhaps unsurprising that SegNet+BH is the best performing as the current state of the art [14] already utilises a BH image pre-processing stage.

C. SegNet Without Fusion

We also investigated the performance of SegNet without fusion in a multi-stage detection pipeline. The resulting ZFA detection ranges are compared to our SegNet+BH system in Figure 9. We highlight that, to ensure a fair comparison, we have reselected the threshold for each system to obtain the ZFA detection ranges.

For the ResQu dataset the mean ZFA detection range and standard error for the 15 cases is 2299m and 74m respectively (recall the mean ZFA detection ranges and SEOM are 2527m

TABLE III: The detection ranges, mean and SEOM of the candidate systems. Missed detections are marked with an X. The SegNet+BH has the best mean detection range and no missed detections.

	SegNet+			
	BH	FAST	H-S	S&T
Case No.	Range (m)	Range (m)	Range (m)	Range (m)
1	2181	2029	1969	2021
2	2997	2820	2554	2804
3	1799	1808	1817	1817
4	2432	2324	2316	2333
5	2621	2512	2436	2478
6	2600	2437	2395	2437
7	2867	2679	2665	2672
8	2606	X	X	X
9	2707	2624	2616	2646
10	2435	X	X	X
11	2849	2712	2629	2705
12	2550	2560	2550	2560
13	2490	2516	2496	2503
14	2364	2204	2120	2170
15	2401	2375	2382	2388
Mean	**2527**	2431	2380	2426
SEOM	76	79	73	79

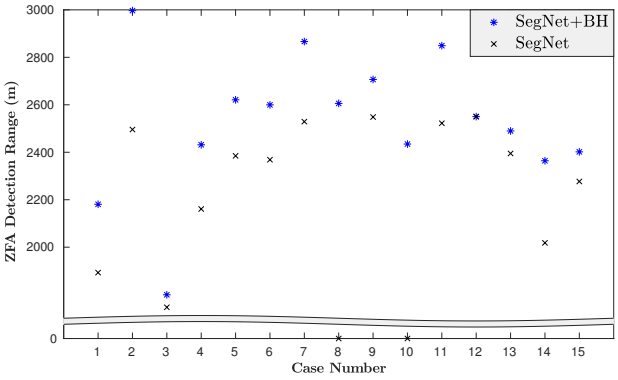

Fig. 9: A comparison of the ZFA detection ranges of our fused SegNet+BH system and an unfused SegNet system. Case 8 and 10 are missed detections for the unfused SegNet system.

and 75m for our proposed fused SegNet+BH system). There were 2 missed detections in Case 8 and Case 10. In these cases we observed that the aircraft pixels were not classified by the unfused SegNet system until the aircraft was around 1500m away. In this situation there is insufficient frames for the HMM filter to declare a detection alert.

D. Smart Skies Data Comparison

We now evaluate the performance of our proposed fused system on our smaller Smart Skies dataset. As this dataset is similar to our training data (same target aircraft and same camera lens), we can potentially gain some insight into the sensitivity of our proposed fused SegNet+BH system. We ran our proposed fused SegNet+BH system, the baseline BH system and an unfused SegNet system and the ZFA detection ranges are presented in Figure 10. Our proposed fused system is able to detect equal to or better than the baseline BH and the

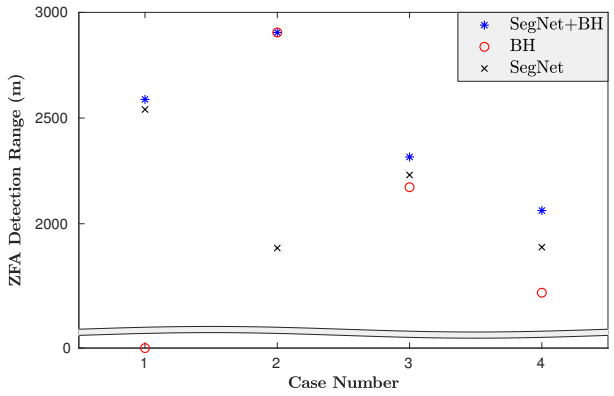

Fig. 10: A comparison of the ZFA detection ranges of our proposed fused SegNet+BH, the baseline BH system and an unfused SegNet system. Case 1 is a missed detections for the baseline BH system.

unfused SegNet system on all cases. Interestingly, in Case 1, our proposed fused system is still able to boost the detection range even when there is a missed detection by the BH system. In contrast, in Case 2, the unfused SegNet system does not detect until 1885m and there is no boost in detection range from our fused SegNet+BH system.

We note that, opposite to the ResQu dataset, the baseline BH system has a missed detection and the unfused SegNet system has no missed detections. We suspect this is due to the Smart Skies being similar to our training data.

V. Discussion

In this section we discuss our proposed fused SegNet+BH system, its limitations and potential future work.

To detect long range aircraft we designed a complementary system which fused our SegNet implementation with BH morphological processing to propose a new image pre-processing stage for use in a multi-stage detection pipeline. Our proposed SegNet+BH system was able to improve detection range in both our datasets despite being captured on different platforms, with different lenses and in different conditions. On our ResQu dataset, the mean detection ranges of our proposed system (cf. Figure 8) are comparable to the medium range of 2593m reported for alerted pilots with an 86% success rate [3].

We found that our SegNet output would detect an aircraft with high precision but not until the aircraft had some discernible features (late in an image sequence when the aircraft was closer). From the literature we knew that BH was able to detect pixel-sized aircraft (early in an image sequence), but with a very low precision. By choosing a pixel-wise addition fusion we are able to exploit the high precision of our SegNet output to intermittently increase the aircraft signal.

A key limitation of our proposed approach is the computational burden of the SegNet compared to a pure morphological approach. Prior SAA systems with pure morphological image pre-processing stages and implemented in Nvidia CUDA/C++ on GPUs performed at 15 frames per second [5] and 9 frames per second [24]. Our MATLAB implementation runs at 2

frames per second. We expect the computational performance of our approach to improve with a specialised implementation.

A fundamental limitation in this application is the lack of available data. Image sequences depicting aircraft on collision course are very limited and expensive to capture due to the risk of flying aircraft on converging paths [30]. Ideally we would have more training and testing data (potentially color data although our intuition is that color features may be less important for detection of aircraft at a long range).

VI. Conclusion

In this paper we trained a deep CNN to learn aircraft features using a variation of the SegNet architecture for semantic pixelwise segmentation. We evaluated the performance of our learnt features fused with BH morphological processing as a new image pre-processing stage in a multi-stage detection pipeline. Our proposed fused system offers an increase in mean detection range of 299m (13.4%) to 2527m with no additional false alarms and improves the SNR of the image pre-processing stage by 0.1085 (8.2%) on the ResQu dataset relative to the current state of the art [14]. We also compared the performance to several other candidate feature extractors as well as the performance of the SegNet without fusion.

VII. Acknowledgements

The authors would like to acknowledge Sean McMahon and Jake Bruce for the numerous deep learning discussions.

The collection of the Smart Skies dataset was supported by the Australian Research Councils Linkage Projects funding scheme (project number LP100100302) and the Smart Skies Project, which is funded, in part, by the Queensland State Government Smart State Funding Scheme.

References

[1] Researchandmarkets.com, "Global Unmanned Aerial Vehicle (UAV) Market Value and Volume: Focus on Class (small and large), Component, and End-User Analysis and Forecast (2017-2021)," Tech. Rep., 2018. [Online]. Available: https://www.researchandmarkets.com/research/tr34m4/

[2] R. A. Clothier, B. P. Williams, and N. L. Fulton, "Structuring the safety case for unmanned aircraft system operations in non-segregated airspace," *Safety Science*, vol. 79, pp. 213–228, Nov 2015.

[3] J. Andrews, "Modeling of air-to-air visual acquisition," *The Lincoln Laboratory Journal*, vol. 2, no. 3, pp. 475–482, 1989.

[4] A. Mcfadyen and L. Mejias, "A survey of autonomous vision-based See and Avoid for Unmanned Aircraft Systems," *Progress in Aerospace Sciences*, vol. 80, pp. 1–17, 2016.

[5] J. Lai, J. J. Ford, L. Mejias, and P. O'Shea, "Characterization of sky-region morphological-temporal airborne collision detection," *Journal of Field Robotics*, vol. 30, no. 2, pp. 171–193, Mar 2013.

[6] D. Gunasinghe and M. Srinivasan, "A strategy for mid-air collision avoidance: Speed modulation to increase minimum separation using a mutually independent and mutually beneficial technique," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2017, pp. 1936–1943.

[7] M. G. Bruno and J. M. Moura, "Multiframe detector/tracker: optimal performance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 925–945, 2001.

[8] M. A. Zaveri, S. N. Merchant, and U. B. Desai, "Wavelet-based detection and its application to tracking in an ir sequence," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1269–1286, Nov 2007.

[9] C. Gao, L. Wang, Y. Xiao, Q. Zhao, and D. Meng, "Infrared small-dim target detection based on markov random field guided noise modeling," *Pattern Recognition*, vol. 76, pp. 463 – 475, 2018.

[10] X. Bai and Y. Bi, "Derivative entropy-based contrast measure for infrared small-target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2452–2466, Apr 2018.

[11] R. Carnie, R. Walker, and P. Corke, "Image processing algorithms for UAV "sense and avoid"," in *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2006. IEEE, 2006, pp. 2848–2853.

[12] A. Nussberger, H. Grabner, and L. Van Gool, "Aerial object tracking from an airborne platform," in *2014 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, May 2014, pp. 1284–1293.

[13] T. L. Molloy, J. J. Ford, and L. Mejias, "Adaptive detection threshold selection for vision-based sense and avoid," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS 2017)*. Miami, FL: IEEE, Jun 2017, pp. 893–901.

[14] J. James, J. J. Ford, and T. L. Molloy, "Quickest Detection of Intermittent Signals With Application to Vision Based Aircraft Detection," Apr 2018. [Online]. Available: http://arxiv.org/abs/1804.09846

[15] S. Petridis, C. Geyer, and S. Singh, "Learning to Detect Aircraft at Low Resolutions," in *Computer Vision Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 474–483.

[16] A. Rozantsev, V. Lepetit, and P. Fua, "Flying objects detection from a single moving camera," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 4128–4136.

[17] K. R. Sapkota, S. Roelofsen, A. Rozantsev, V. Lepetit, D. Gillet, P. Fua, and A. Martinoli, "Vision-based Unmanned Aerial Vehicle detection and tracking for sense and avoid systems," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct 2016, pp. 1556–1561.

[18] D. Dey, C. Geyer, S. Singh, and M. Digioia, "Passive, Long-Range Detection of Aircraft: Towards a Field Deployable Sense and Avoid System," *Field and Service Robotics*, vol. 7, no. Jul 2009, pp. 113–123, 2010.

[19] ——, "A cascaded method to detect aircraft in video imagery," *The International Journal of Robotics Research*, vol. 30, no. 12, pp. 1527–1540, Oct 2011.

[20] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Aircraft detection by deep convolutional neural networks," vol. 7, pp. 10–17, 01 2014.

[21] S. Hwang, J. Lee, H. Shin, S. Cho, and D. H. Shim, "Aircraft Detection using Deep Convolutional Neural Network in Small Unmanned Aircraft Systems," in *2018 AIAA Information Systems-AIAA Infotech @ Aerospace*. Reston, Virginia: American Institute of Aeronautics and Astronautics, Jan 2018.

[22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, Dec 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1026–1034.

[24] D. Bratanov, L. Mejias, and J. J. Ford, "A vision-based sense-and-avoid system tested on a ScanEagle UAV," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, Jun 2017, pp. 1134–1142.

[25] D. Tulpan, N. Belacel, F. Famili, and K. Ellis, "Experimental evaluation of four feature detection methods for close range and distant airborne targets for unmanned aircraft systems applications," in *2014 International Conference on Unmanned Aircraft Systems (ICUAS)*, May 2014, pp. 1267–1273.

[26] T. L. Molloy, J. J. Ford, and L. Mejias, "Detection of aircraft below the horizon for vision-based detect and avoid in unmanned aircraft systems," *Journal of Field Robotics*, vol. 34, no. 7, pp. 1378–1391, 2017.

[27] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, Oct 2005, pp. 1508–1515 Vol. 2.

[28] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[29] J. Shi and C. Tomasi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Jun 1994, pp. 593–600.

[30] J. Lai, L. Mejias, and J. J. Ford, "Airborne vision-based collision-detection system," *Journal of Field Robotics*, vol. 28, no. 2, pp. 137–157, Mar 2011.