FINAL EVALUATION PRESENTATION

Object Detection, Tracking and Suspicious Activity Recognition for Maritime Surveillance using Thermal Vision

Department of Electronic and Telecommunication Engineering

University of Moratuwa

Group Members

160005C

K.G. Abeywardena

160243D

H.L.S.H. Jayasundara

160285G

G.K.S.R. Karunasena

160616B

S.K. Sumanthiran

Supervisors

Dr. Peshala Jayasekara

Dr. Ranga Rodrigo

Motivation

CNA938 | CNA Lifestyle | CNA Insider | Singapore | Asia | World | Business | Sport | Commentary | News Clips | Video on Demand | Podc | All Sections

05 Mar 2020 06:34PM

Bookmark

Asia

Sri Lanka seizes record US$33 million drugs haul at sea

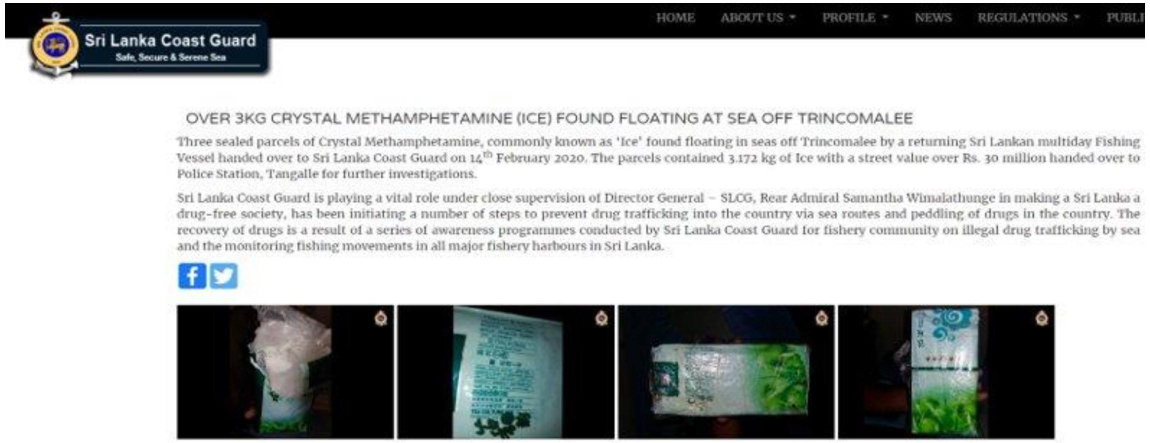
ඇඳිරි නීතිය නොතකා පුත්තලමේ සිට මුහුදු මාර්ගයෙන් මන්නාරමට යන්න ගිය 20ක් අත්අඩංගුවට

Friday, 20 March 2020 - 18:27

mnews.lk www.sooriyanfmnews.lk

Trending News

01 April 2020



Sri Lanka Coast Guard
Safe, Secure & Serene Sea

HOME | ABOUT US | PROFILE | NEWS | REGULATIONS | PUBLI

OVER 3KG CRYSTAL METHAMPHETAMINE (ICE) FOUND FLOATING AT SEA OFF TRINCOMALEE

Three sealed parcels of Crystal Methamphetamine, commonly known as 'Ice' found floating in seas off Trincomalee by a returning Sri Lankan multiday Fishing Vessel handed over to Sri Lanka Coast Guard on 14th February 2020. The parcels contained 3.172 kg of Ice with a street value over Rs. 30 million handed over to Police Station, Tangalle for further investigations.

Sri Lanka Coast Guard is playing a vital role under close supervision of Director General – SLCG, Rear Admiral Samantha Wimalathunge in making a Sri Lanka a drug-free society, has been initiating a number of steps to prevent drug trafficking into the country via sea routes and peddling of drugs in the country. The recovery of drugs is a result of a series of awareness programmes conducted by Sri Lanka Coast Guard for fishery community on illegal drug trafficking by sea and the monitoring fishing movements in all major fishery harbours in Sri Lanka.

Event News >> Thirty (30) illegal Sri Lankan immigrants held by Navy in southern seas

Thirty (30) illegal Sri Lankan immigrants held by Navy in southern seas

Naval personnel attached to Fast Attack Craft (FACs) belonging to the Southern Naval Command, deployed on patrol waters, intercepted a suspicious trawler plying in southern seas this morning (07th March) and held 30 suspects onboa

Having spotted a suspicious boat movement at sea about 80 nautical miles off the Galle Lighthouse, two Fast Attack C directed to the location of the suspicious trawler. Accordingly, the naval personnel held 30 illegal Sri Lankan immigra are due to be handed over to the Galle Harbour Police after a medical examination and preliminary naval investigatio

The Navy urges the general public not to involve in high risk sea-borne migration to overseas countries based on false reiterates that such attempts would finally end up behind bars.

Further the Navy reminds of its strong network of intelligence and regular patrols in place to nip such illegal migratio strict measures to curb such attempts.

Final Deliverables

1

Functioning Object Detection algorithm

- For objects related to maritime environment such as boats, swimmers etc. in RGB domain.
- For objects related to urban driving such as cars, pedestrians, bicycles etc. in Thermal environments.
- Competitive mAP score on predicted bounding boxes

2

Functioning Object Tracking algorithm

- Capable of tracking detected objects by the object tracker in challenging environments.

3

Functioning Activity Detection algorithm

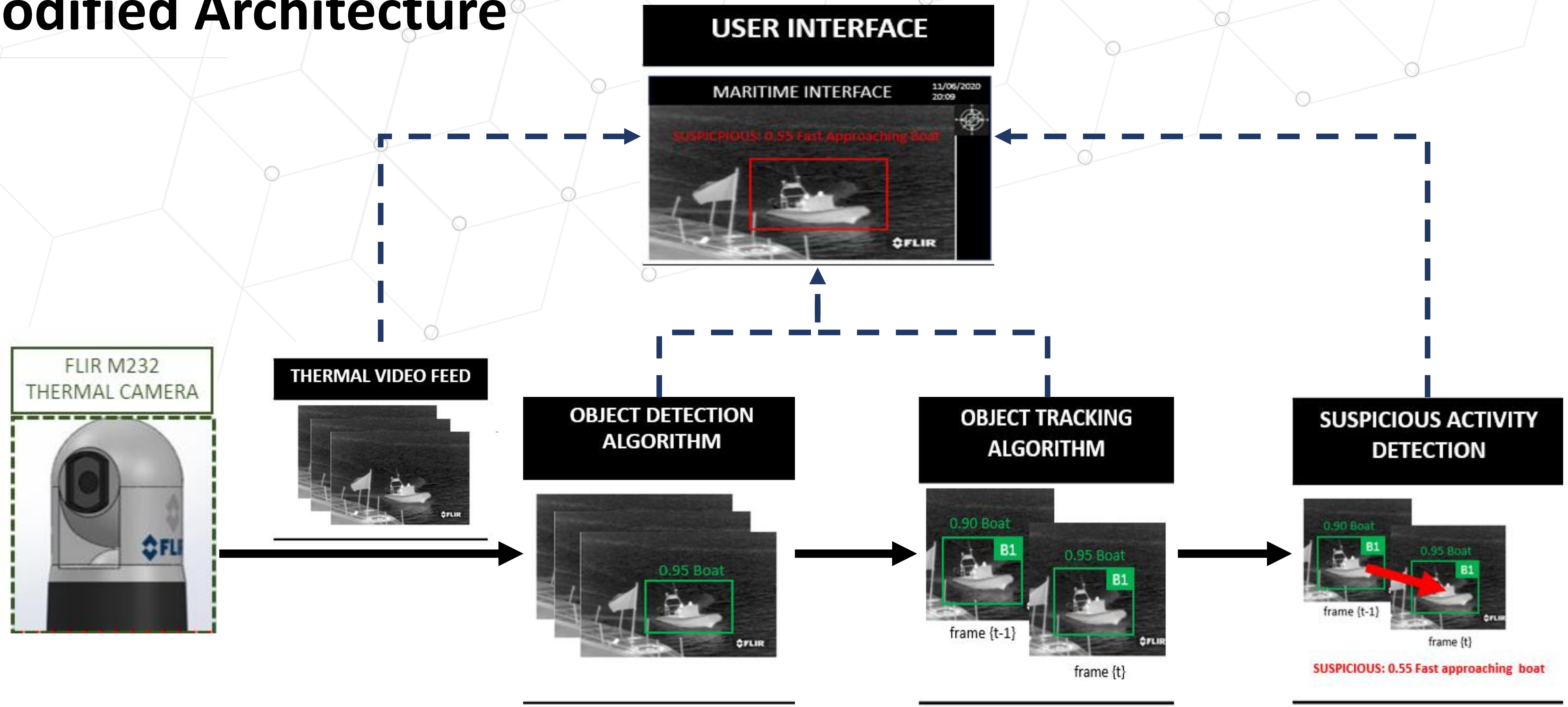
- Identifying set of pre-defined activities in UCF 101-24 and J-HMDB-21 dataset.
- With competitive f-mAP, v-mAP and FPS score.

4

Well documented software

- User-friendly

Modified Architecture



May include:
- Dumping objects
- Unauthorized fishing
- Loitering
- Evasive actions

Content

Object Detection

RGB, Thermal, objects in Maritime, General environments.

Object Tracking

Tracking of detected objects.

User Interface

An interface to showcase the detected and tracking of objects.

Action Detection

Online Real Time Spatio Temporal Action Detection

Research Paper

Publication in a well-known conference.

Maritime Object Detection

Alternative Datasets for Maritime environments

Datasets	Link	Images/ Video type	Description
Singapore Maritime Dataset	https://sites.google.com/site/dilipprasad/home/singapore-maritime-dataset	RGB and Near IR	• RGB (Onshore and Onboard) • Near IR (Onshore) • NIR videos were captured using Canon 70D camera with hot mirror removed and Near-IR Bandpass filter (Different than actual thermal images)
SeaShips	http://www.lmars.whu.edu.cn/prof_web/shaozhenfeng/datasets/SeaShips(7000).zip	Only RGB	• Contains 31455 images (Only 7000 images publicly available) • Annotations provided
IPATCH	http://ftp.pets.rdg.ac.uk/pub/PETS2016/MidHighLevelChallengeData/IPATCH/	Both RGB and Thermal	• Contains a set of fourteen multi camera recordings (visible, thermal) collected off the coast of Brest, France • No annotations provided/ The categories of the objects

Maritime Object Detection

Alternative Frameworks for Maritime Object Detection

Dataset	Evaluation Criterion	Framework		
		SSD[1]	CornerNet-Lite (Squeeze)[2]	CenterNet[3]
SeaShips	mAP % @ IoU 0.5	28.4	59	81.8
	FPS	19	60	52
Singapore Maritime Dataset	mAP % @ IoU 0.5	27	55.3	60.7
	FPS	19	60	52

*Algorithms are trained under both SeaShips and Singapore Maritime dataset.

Maritime Object Detection

RGB version of Singapore Maritime Dataset and SeaShips Dataset

Video from an Onboard camera -SMD (Inferenced using CenterNet)

Video from an Onshore camera -Seaships (Inferenced using CenterNet)

Maritime Object Detection

Near IR version of Singapore Maritime Dataset

Video from an Onshore NIR camera (Without Haze) -SMD (Inferenced using CenterNet)

Video from an Onshore NIR camera (With Haze) -SMD (Inferenced using CenterNet)

Thermal Object Detection

Alternative Frameworks for Thermal Object Detection

Evaluation Criterion	Framework		
	SSD[1]	CornerNet-Lite (Squeeze)[2]	CenterNet [3]
mAP % @ IoU 0.5	28.8	81	88
FPS	19	60	52

Video from FLIR ADAS Thermal Dataset (Inferenced using CenterNet)

*Algorithms are trained under FLIR dataset.

Camera Set-up

Initial Testing – Setting up the camera

Initial Testing – Obtaining real-time video

Plate to mount on tripod

To be connected to an automobile auxiliary power outlet

Mounted thermal camera

Inference Video Collection

Collecting videos for urban driving data in both daytime and nighttime.

Testing the thermal camera.

Collecting urban driving video during nighttime.

Collecting urban driving video during daytime.

Thermal Object Detection

An urban driving dataset in Colombo suburbs collected using FLIR M232 Thermal Camera.

Video from urban driving dataset during nighttime - Inferenced using CenterNet

Video from urban driving dataset during daytime - Inferenced using CenterNet

Location

- Nighttime - Baseline Road, Colombo 05
- Daytime – Colombo 06

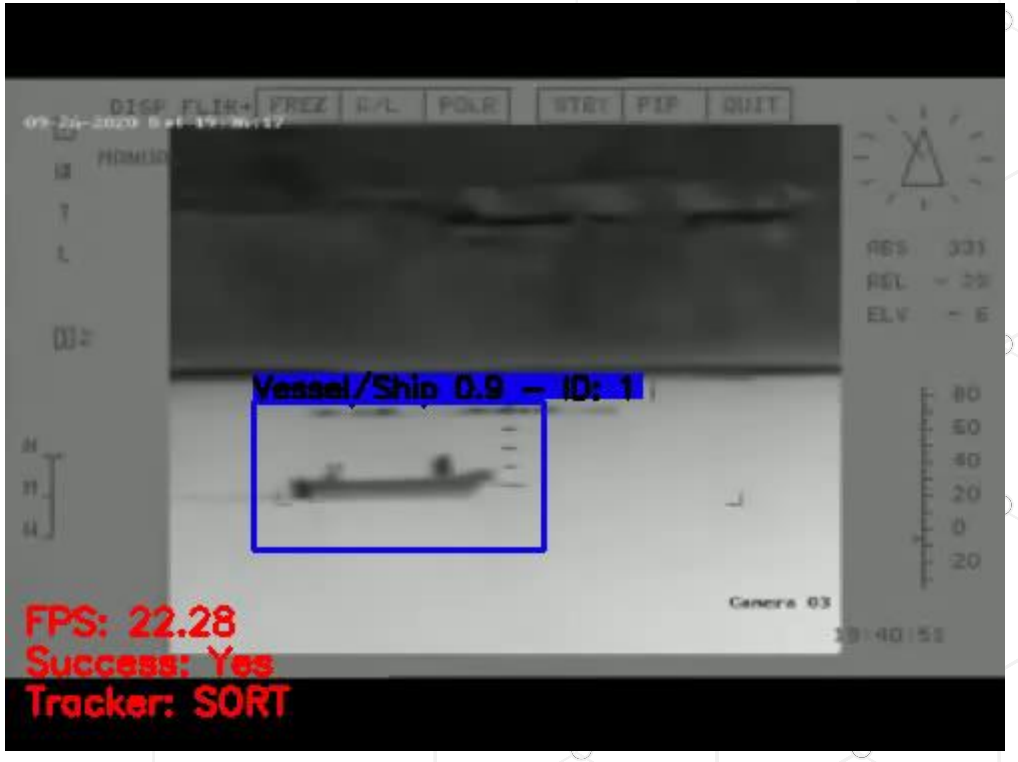
Object Tracking

Alternative Algorithms

Tracker	Metric					
	MOTA ↑	MOTP ↑	FAF ↓	MT ↑	FP ↓	ML ↓
RMOT	18.6	69.6	2.20%	5.30%	12473	53.30%
TC_ODAL	15.1	70.5	2.20%	3.20%	12970	55.80%
TDAM	33	72.8	1.70%	13.30%	10064	39.10%
MDP	30.3	71.3	1.70%	13.00%	9717	38.40%
SORT	**33.4**	**72.1**	**1.30%**	**11.70%**	**7318**	**30.90%**

MOTA – Multi-Object Tracking Accuracy
FAF – Number of false alarms per frame
FP – False Positives

MOTP – Multi-Object Tracking Precision
MT – Number of mostly tracked trajectories
ML – Number of mostly lost trajectories



Video from SL Navy (Tracker - SORT)

Evaluation results taken from : *Simple Online and Real-Time Tracking – Bewley et. al.*

Object Tracking

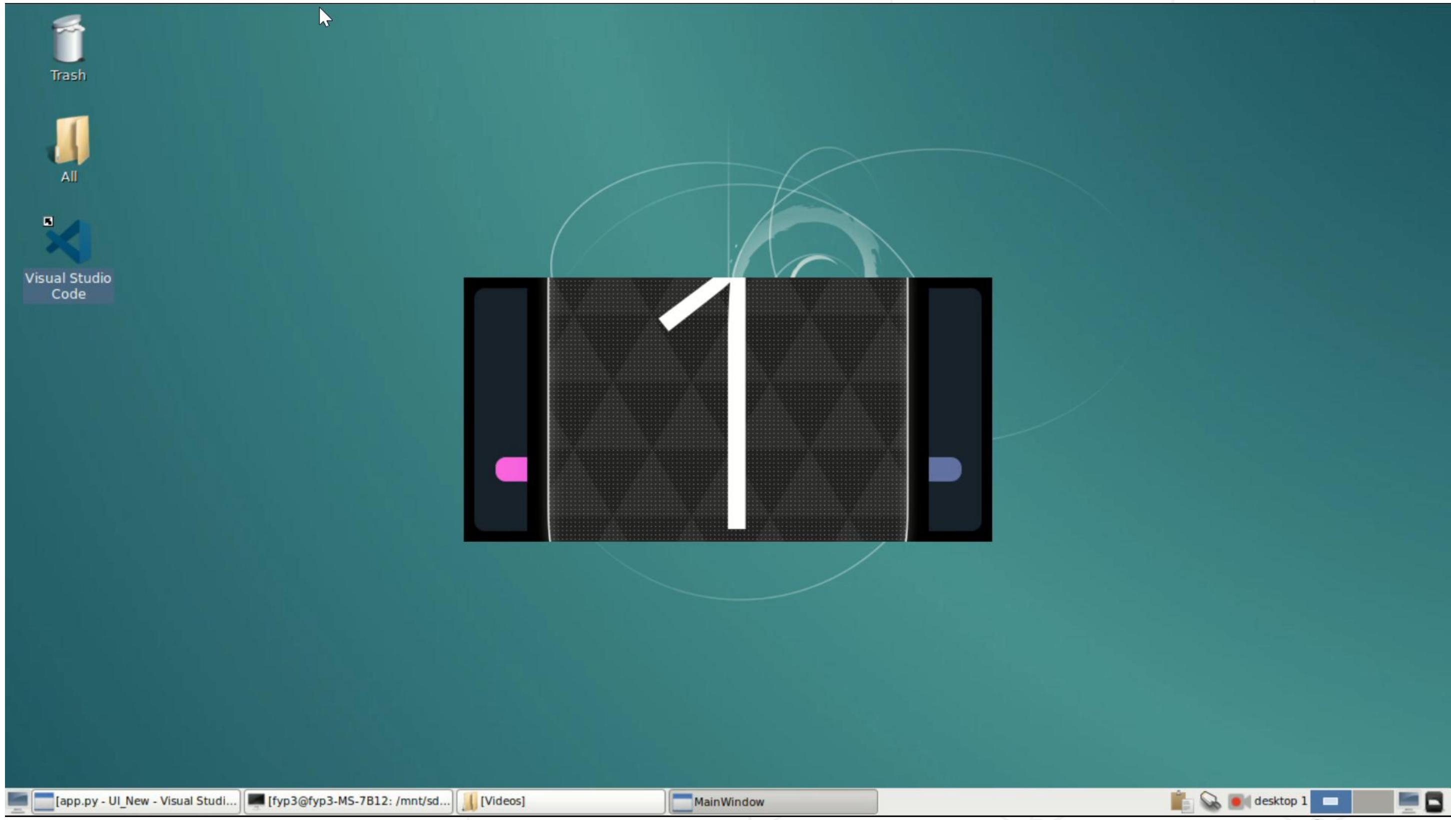
Singapore Maritime Dataset and Urban Driving Dataset

Video from NIR-SMD tracked using SORT Tracker.

Video from SMD tracked using SORT Tracker.

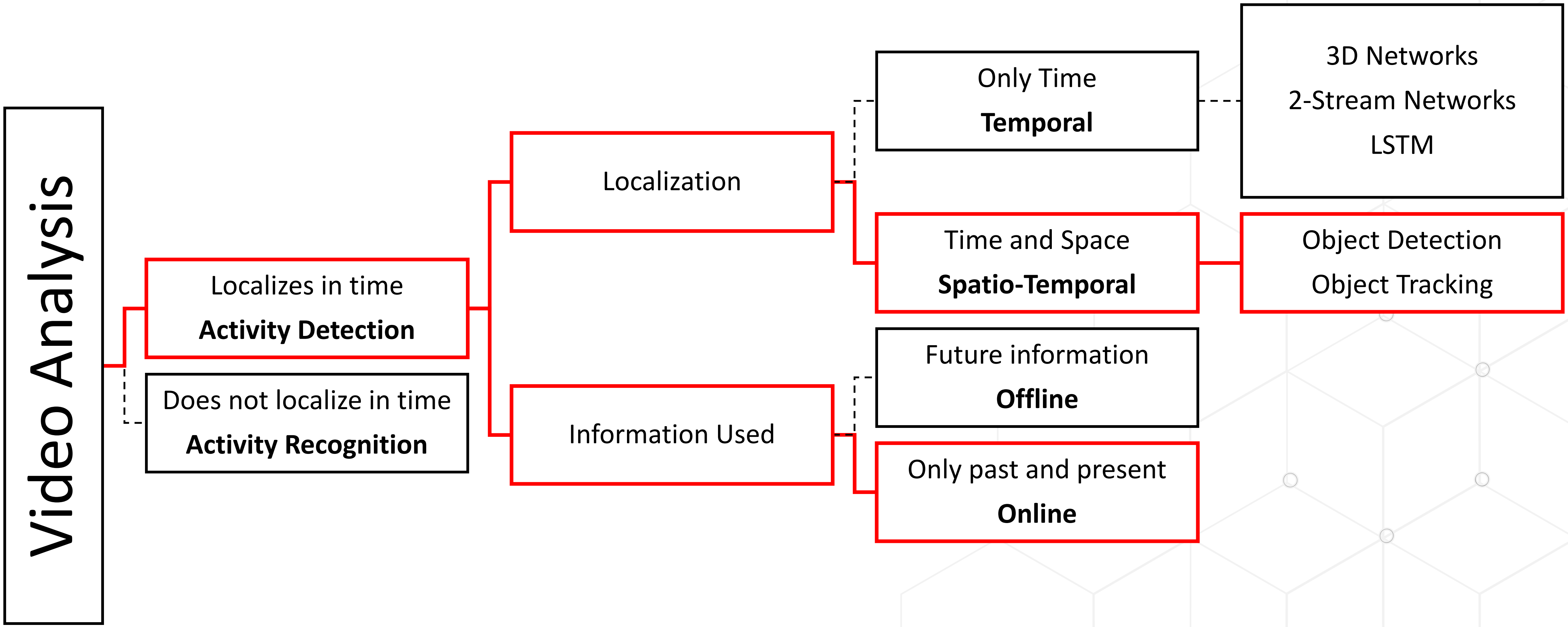
User Interface

- Developing using PyQt5
- Python based interface with the ability to display detected objects, activities and track those in real-time
- Displays detected objects on the side



Interface created using PyQt5

Action Detection



Action Detection

Alternative Datasets

Datasets	Link	Images/ Video type	Description
UCF 101 - 24	https://www.crcv.ucf.edu/data/UCF101/UCF101.rar	• RGB • Sports	• Extracted from a large dataset (UCF 101) - 101 action categories • 13320 videos • Annotations provided • UCF 24 – Only sports with 24 categories
J-HMDB 21	https://serre-lab.clps.brown.edu/wp-content/uploads/2013/10/hmdb51_org.rar	• RGB • Facial actions and Body movements	• Extracted from a large dataset (HMDB 51) - 51 action categories • 928 videos • Annotations provided

Action Detection

Thermal dataset creation for Action Detection

- Experimental Technique
- Used Generative Adversarial Network (GAN) based frameworks to generate synthetic images
- Used Pix2Pix framework

- Challenges :
    1. High Noise
    2. Different thermal signatures

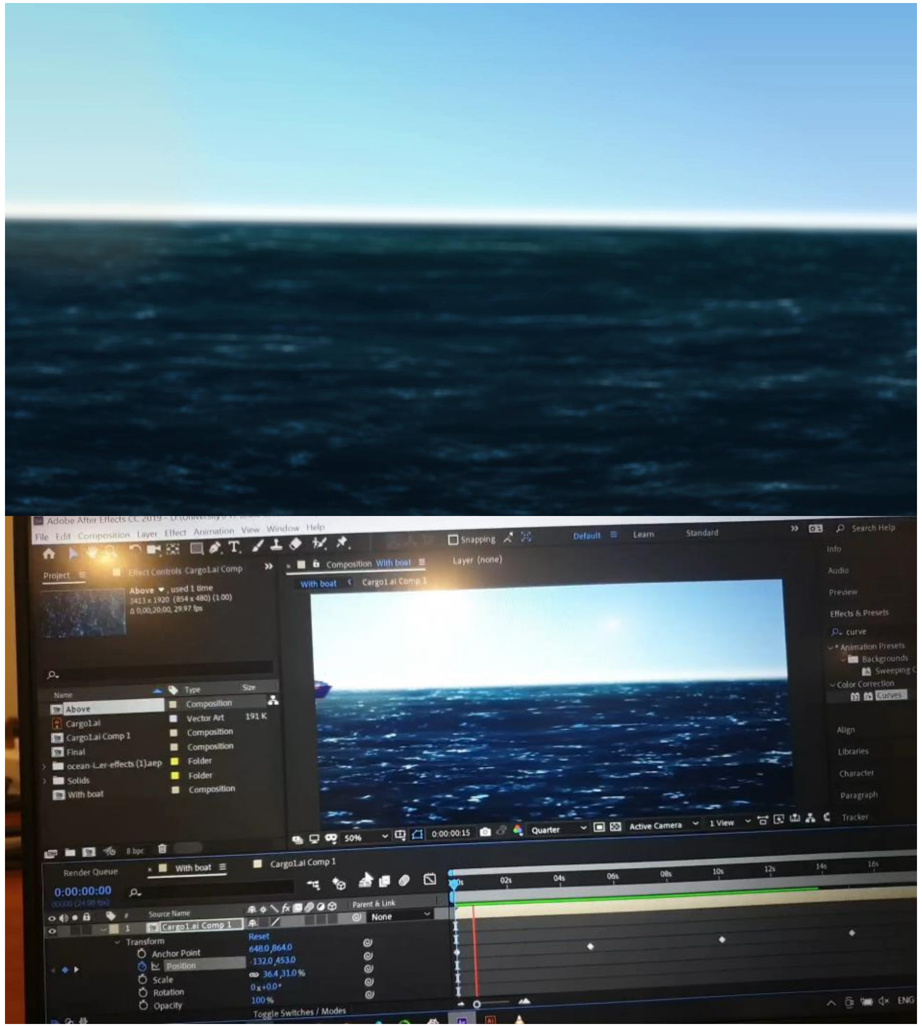
Converted RGB video using Pix2Pix

Action Detection

Maritime dataset creation for Action Detection

- Experimental Technique with the intention of creating a maritime dataset with action instances.
- Used Adobe After Effects to develop videos from scratch.

- Challenges:
  1. High rendering/creating time for each video.
  2. Repetition of patterns in the sea waves, which might affect the deep learning algorithms.



Custom made maritime action classification data using After Effects.

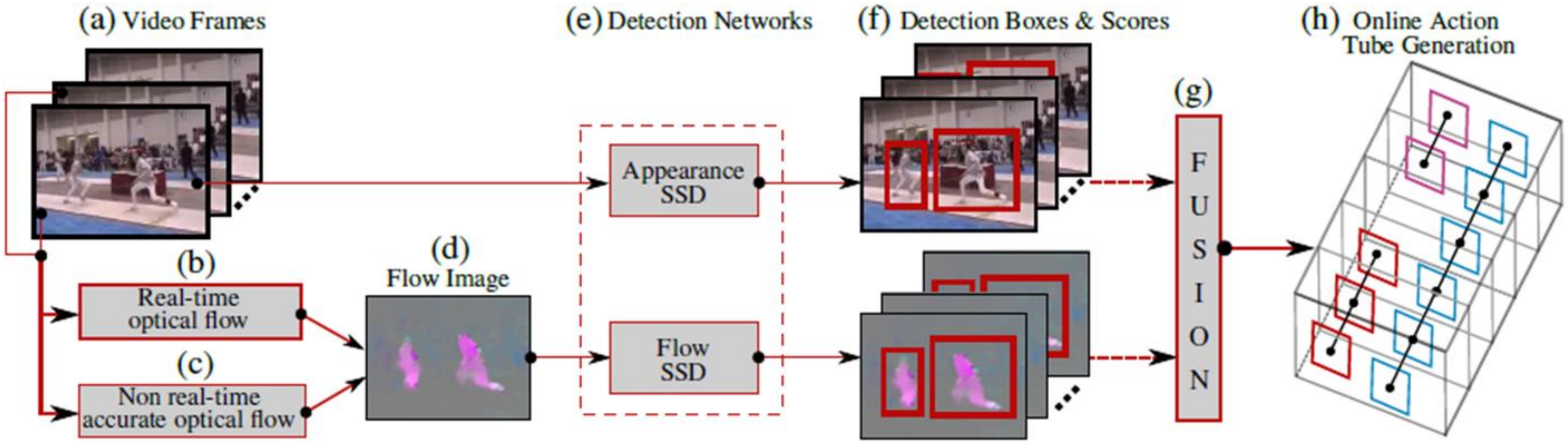
Action Detection

Evaluation of alternatives – Action Detection Algorithms

	Temporal Recurrent Networks (TRN)[11]	Information Discrimination Unit (IDU) [12]	Spatio-Temporal and Motion Encoding (STM) [13]	A structured Model for Action Detection[14]	Spatio-Temporal Progressive Learning(STEP) [15]	Online Real-time Multiple Spatio-temporal Action Localization and Prediction (ROAD) [16]
Temporal/Spatio-temporal	Temporal	Temporal	Temporal	Spatio-Temporal	Spatio-Temporal	Spatio-Temporal
Backbone	VGG-16 / ResNet-200	VGG-16 / ResNet-200	ResNet-50	ResNet-50/Mask-RCNN	VGG-16	VGG-16
Online/Offline	Online	Online	Offline	Offline	Offline	Online
FPS	24	24	-	12	21	28
Dataset	THUMOS'14	THUMOS'14	UCF101	UCF101	UCF101	UCF101
mAP score	47.2	60.3	96.0	77.9	75.0	43.0

Action Detection

ROAD Architecture



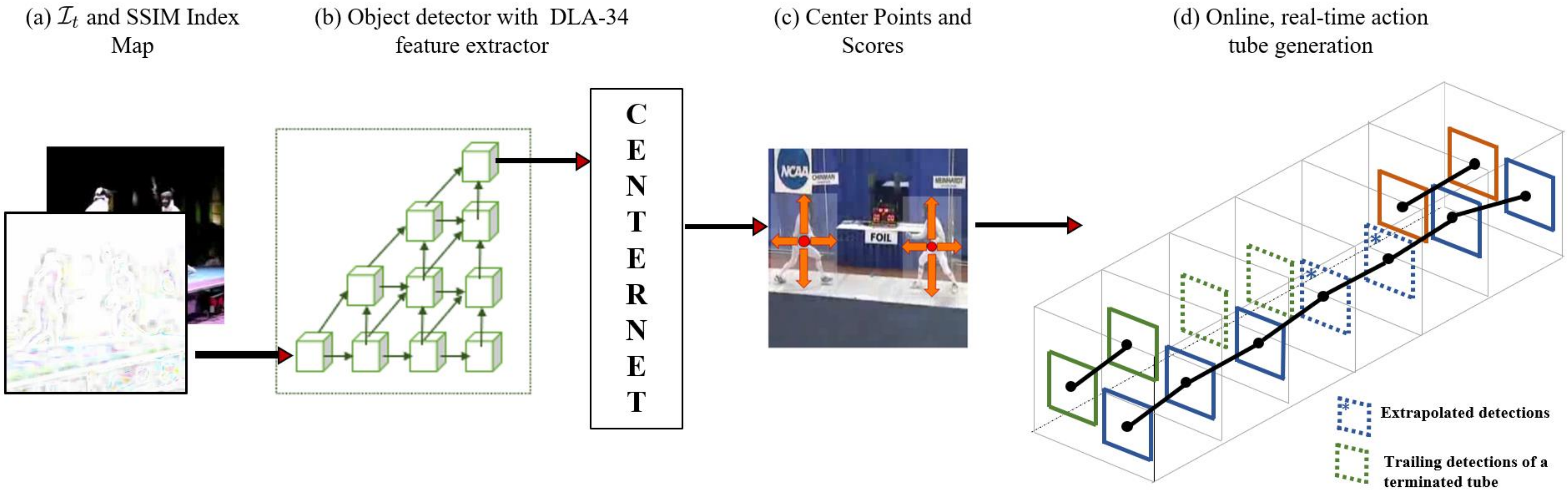
Issues	Solutions Worked-On
No implementation of linking algorithm in python	Implementing the linking algorithm in python
Lack of end-to-end pipeline from action localizations to action linking	Building the end-to-end pipeline in python (100% implementation with SSD detector)
Duplication of results for Fast Optical Flow implementation is not provided	

Action Detection – Our Approach

Problems in Current Methods	*Proposed Solutions*
Anchor box based Spatio-Temporal (ST) Action Localization	A key-point based Spatio-Temporal (ST) Action Localization
Computationally expensive optical flow based inter-frame temporal information extraction	Structural Similarity (SSIM) index map based inter-frame temporal information extraction
Two-stream 2D CNN architecture	Cascaded spatial and temporal information based single input to a single feature extractor
Interpolation-based tube linking algorithm	Extrapolation-based improved tube-linking algorithm

Action Detection – Our Approach

Proposed ST Action Localization Architecture



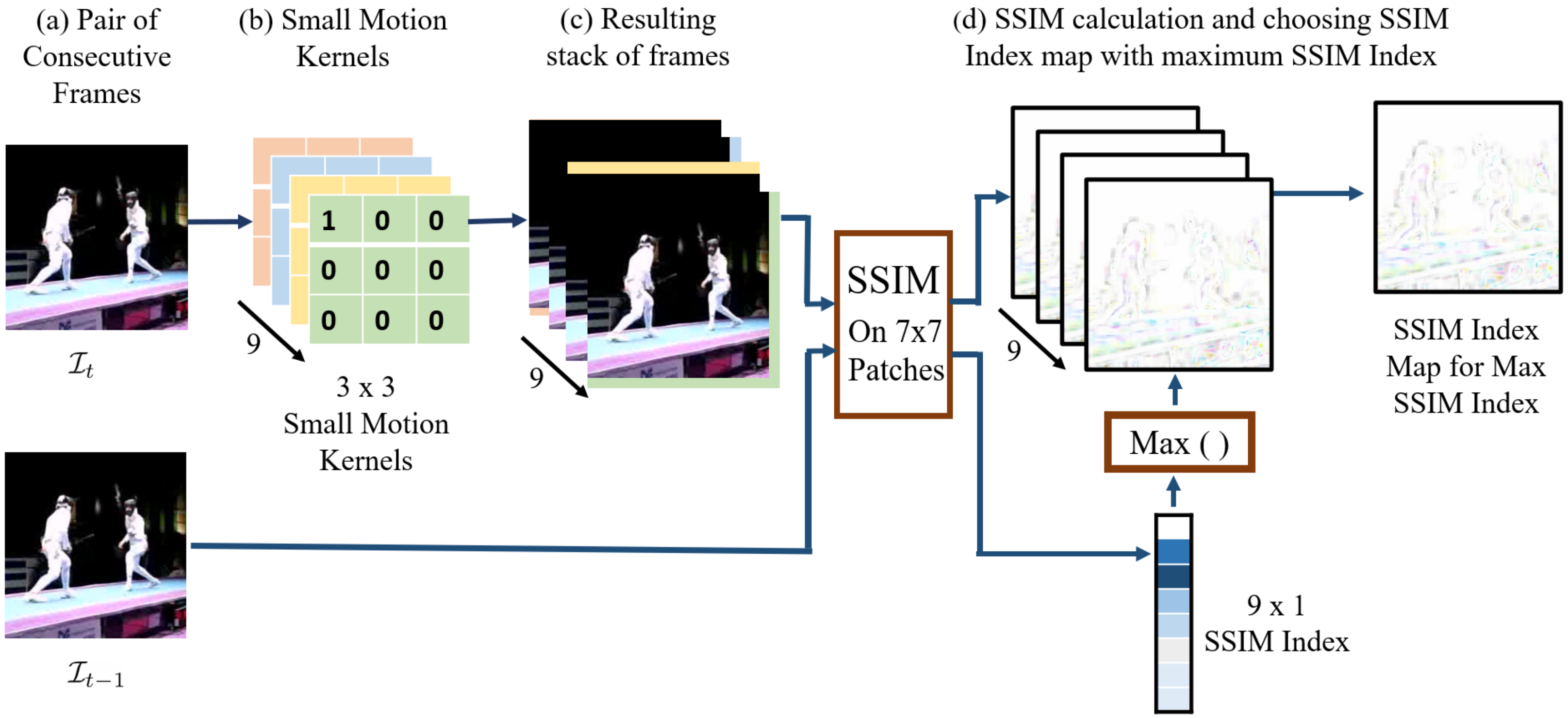
Proposed ST Action Localization Approach

Action Detection – Our Approach

Novelty in our approach

1. **Structural Similarity (SSIM) index map based temporal information extraction**

Objective – To replace computationally expensive optical flow calculation with an alternative that capture the temporal information efficiently



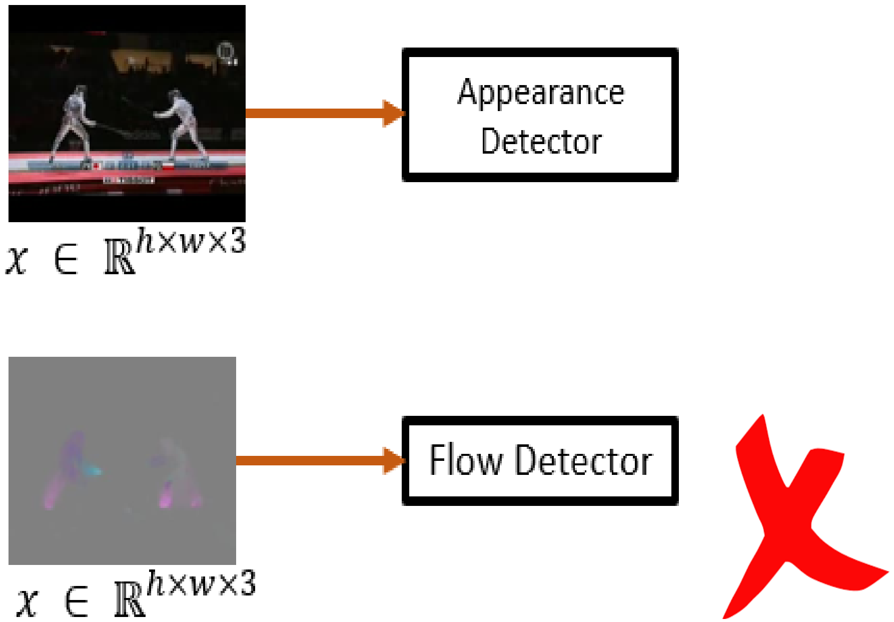
Temporal Information Extraction using SSIM index map

Action Detection – Our Approach

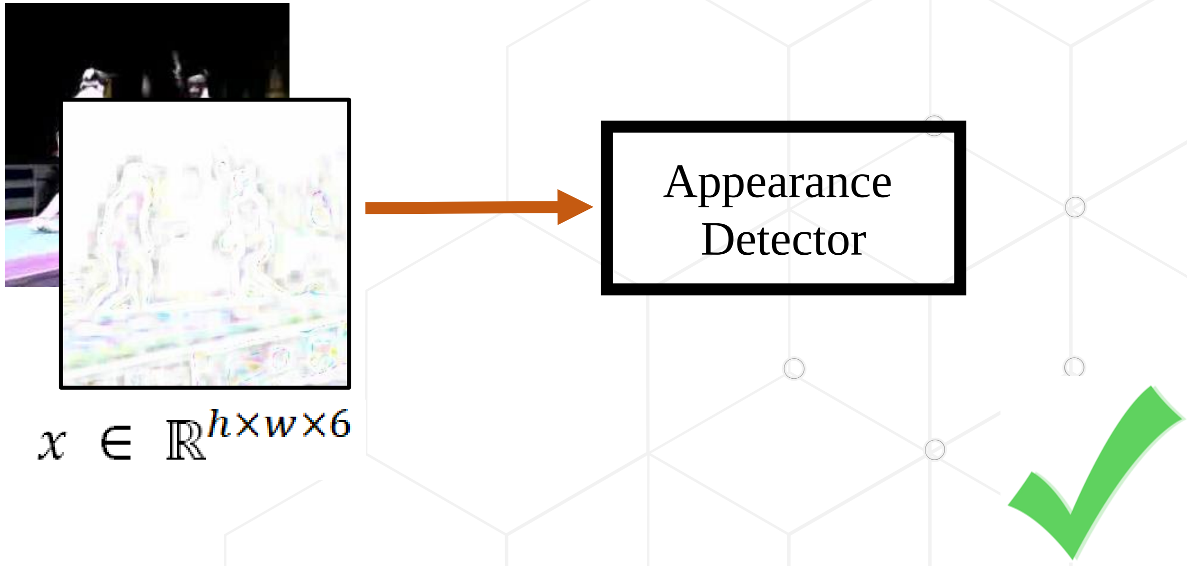
Novelty in our approach

2. Cascaded spatial and temporal information based single input to a single feature extractor for discriminative learning

Objective – To replace the redundant two-stream architectures by providing both spatial and temporal information together



Two Stream RGB and Optical Flow Input



Proposed Cascaded Input

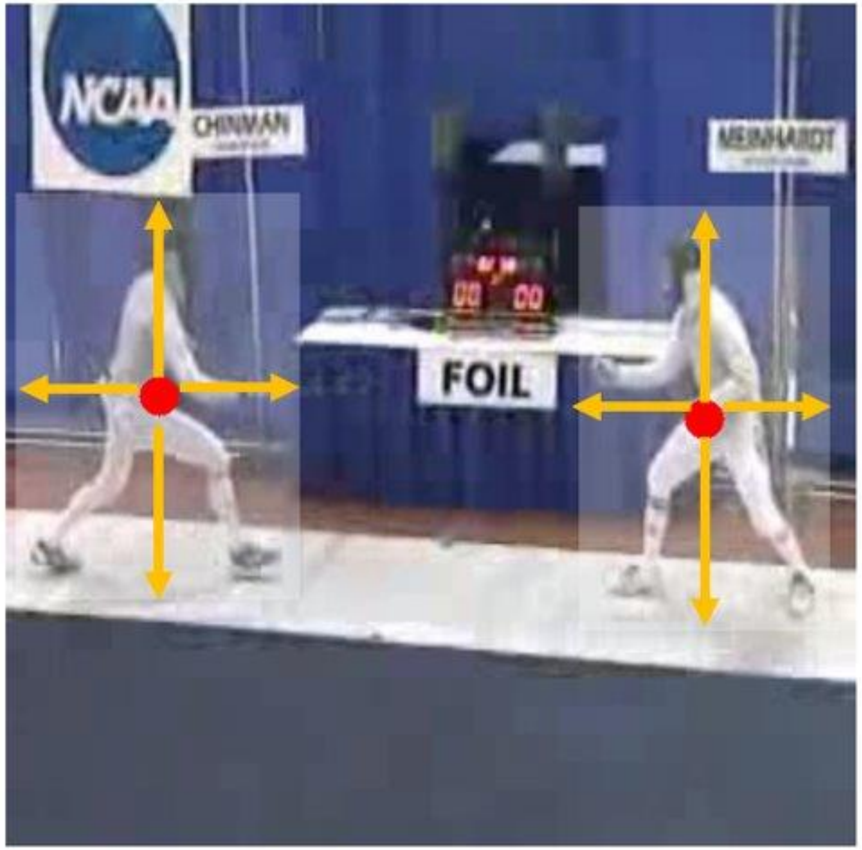
Action Detection – Our Approach

Novelty in our approach

3. Key points for the Action Detection

Key-point based Action Localization has not being exploited in any of the past work.

Objective – To reduce the complexity and to improve inference time with key point detection using CenterNet with DLA-34.



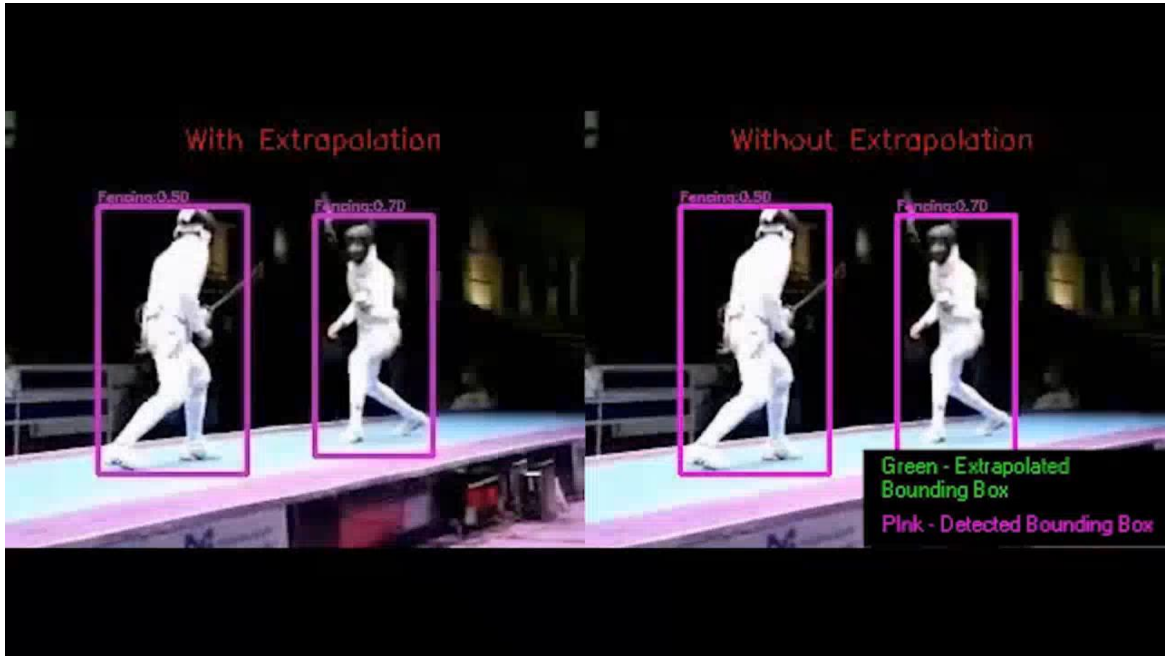
Key Point Detection of Action Instances

Action Detection – Our Approach

Novelty in our approach

4. Improved Tube Linking Algorithm

Objective – To link the actions localized in the video frames and reduce the miss rate through extrapolation based on movement of the bounding boxes of past detections and their class scores



Demonstration of the Proposed Linking Algorithm

Algorithm 1: Online tube generation

Input: $\mathcal{T}^{t-1}, \mathcal{D}^{t}, c, \lambda, k$
Output: $\mathcal{T}^{t}$
for $T_j^{t-1} \in \mathcal{T}^{t-1}$ **do**
  $s \leftarrow 0;\ m \leftarrow 0;$
  for $D_i^t \in \mathcal{D}^t$ **do**
    if $\mathtt{IoU}\,(b^t_{D_i},\ \ b^{t-1}_{T_j}) \geq \lambda$ **and** $s < \boldsymbol{s}^t_{D_i}(c)$ **then** $b^t_{T_j} \leftarrow b^t_{D_i};\ \tau \leftarrow 0\ \ s \leftarrow \boldsymbol{s}^t_{D_i};\ m \leftarrow i;$
  end
  if $m = 0$ **and** $\tau < k$ **then**
    if $box\_pred = True$ **then** $b^t_{T_j} \leftarrow \mathtt{predict\_bbox}\,(b^{t-1}_{T_j},\ \ b^{t-2}_{T_j})$;
    else $b^t_{T_j} \leftarrow b^{t-1}_{T_j}$;
    $\tau \leftarrow \tau + 1;$
  end
  $s^t_{T_j}, c_{T_j} \leftarrow \mathtt{update\_label}\,(s^{t-1}_{T_j},\ \ \boldsymbol{s}^t_{D_m})$;
end

Proposed Linking Algorithm

Action Detection – Our Approach

Results Comparison

Experiments were done on two datasets:

1. UCF-101-24 dataset : Challenging dataset with multiple action instances per video
2. J-HMBD-21 : Challenging dataset with single action instance per video

Table 1: ST action localization results (v-mAP) on UCF-101-24 and J-HMDB-21 dataset.

Method	UCF-101-24					J-HMDB-21					FPS
	f-mAP @0.5	v-mAP 0.2	v-mAP 0.5	v-mAP 0.75	v-mAP 0.5:0.95	f-mAP @0.5	v-mAP 0.2	v-mAP 0.5	v-mAP 0.75	v-mAP 0.5:0.95	
Saha *et al.* [15]◇	-	66.6	36.4	7.9	14.4	-	72.6	71.5	43.3	40.0	4
Peng *et al.* [13]◇	65.7	72.9	-	-	-	58.5	74.3	73.1	-	-	-
Zhang *et al.*[28]◇⋆	67.7	74.8	46.6	16.7	21.9	37.4	-	-	-	-	37.8
ROAD+AF [16]‡	-	73.5	46.3	15.0	20.4	-	70.8	70.1	43.7	39.7	7
ROAD+RTF [16]‡⋆	-	70.2	43.0	14.5	19.2	-	66.0	63.9	35.1	34.4	28
ROAD (A) [16]†⋆	-	69.8	40.9	15.5	18.7	-	60.8	59.7	37.5	33.9	40
Ours (A)†⋆	71.8	70.2	**44.3**	16.6	**20.6**	**51.2**	59.3	59.2	48.2	**41.2**	**52.9**
Ours†⋆	**74.7**	**72.7**	**43.1**	**16.8**	**20.2**	50.5	58.9	58.4	**49.5**	40.6	**41.8**

◇ Offline ⋆ Real-time † Online with no OF ‡ Online with OF

Analysis on Proposed Action Detection Architecture

Impact on Inference Time

Framework Module	Ours	A + DSIM	A + $\mathcal{I}_{t-1}$	A	RTF	A + AF
Temporal INFO EXT (ms)	5.0	5.0	-	-	7.0	110.0
Detection network (ms)	16.4	16.4	16.4	16.4	16.4	16.4
Tube generation time (ms)	2.5	2.5	2.5	2.5	3.0	3.0
Overall (ms)	23.9	23.9	18.9	18.9	26.4	129.4

Ours – Cascading current frame with SSIM index map

A + DSIM – Cascading current frame with Structural dissimilarity index map

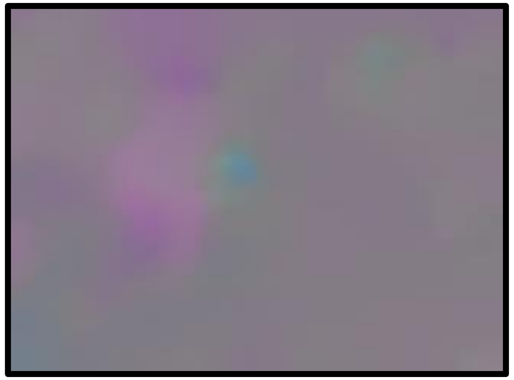
A + $\mathcal{I}_{t-1}$ – Cascading current frame with previous frame

A – Current Frame only

RTF – Two-Stream Architecture with Real-Time OF

A + AF - Two-Stream Architecture with Accurate OF

AF



RTF

SS-map

DSIM

Analysis on Proposed Action Detection Architecture

Impact on performance by temporal information representation methods

Objective – Evaluating different temporal representation methods to be concatenated with the current frame as the cascaded input

Candidate	UCF-101-24					J-HMDB-21				
	f-mAP	v-mAP				f-mAP	v-mAP			
	@0.5	0.2	0.5	0.75	0.5:0.95	@0.5	0.2	0.5	0.75	0.5:0.95
$\mathcal{I}_{t-1}$	74.4	71.6	44.1	17.0	20.7	47.9	57.2	55.9	48.1	39.9
SS-map	74.7	72.4	43.0	16.6	20.2	50.5	58.9	58.4	49.4	40.5
DSIM index map	74.5	73.4	44.9	16.4	20.7	49.9	56.4	55.9	49.2	39.9

$\mathcal{I}_{t-1}$ - Cascading with previous frame
SS-map – Cascading with SSIM index map
DSIM index map – Cascading with Structural Dissimilarity index map

Analysis on Proposed Action Detection Architecture

Impact on performance by linking algorithm variations

Objective – Evaluating the improvements made on the temporal linking algorithm against the original implementation on the detections obtained through key-point detector

Linking Algorithm	Improvement		UCF-101-24				J-HMDB-21			
	EXPLT	BOXP	v-mAP				v-mAP			
			0.2	0.5	0.75	0.5:0.95	0.2	0.5	0.75	0.5:0.95
Original			72.6	43.4	16.8	20.3	58.8	58.3	49.4	40.5
Ours	✓		72.7	43.1	16.8	20.2	58.9	58.4	49.4	40.6
Ours	✓	✓	72.4	43.0	16.6	20.2	58.9	58.4	49.4	40.5

EXPLT – With Extrapolation
BOXP – With Box Prediction

Paper Submission on Action Detection Domain

British Machine Vision Conference - 2021

Paper Title: Key-point Detection based Online Real-Time Spatio-temporal Action Localization

Submission Summary

Conference Name
British Machine Vision Conference 2021

Paper ID	1556
Paper Title	Key-point detection based Online Real-Time Spatio-Temporal Action Localization
Abstract	Real-time and online action localization in a video is a critical yet highly challenging problem. Accurate action localization requires utilization of both temporal and spatial information. Recent attempts achieve this by using computationally intensive 3D CNN architectures or highly redundant two-stream architectures with optical flow, making them both unsuitable for real-time, online applications. To accomplish activity localization under highly challenging real-time constraints, we propose utilizing fast and efficient key-point based bounding box prediction to spatially localize actions. We then introduce a tube-linking algorithm that maintains the continuity of action tubes temporally in the presence of occlusions. Further, we eliminate the need for a two-stream architecture by combining temporal and spatial information into a cascaded input to a single network, allowing the network to learn from both types of information. Temporal information is efficiently extracted using a structural similarity index map as opposed to computationally intensive optical flow. Despite the simplicity of our approach, our lightweight end-to-end architecture achieves state-of-the-art frame-mAP of 74.7% on the challenging UCF101-24 dataset, demonstrating a performance gain of 6.4% over the previous best online methods. We also achieve state-of-the-art video-mAP results compared to both online and offline methods. Moreover, our model achieves a frame rate of 41.8 FPS, which is a 10.7% improvement over contemporary real-time methods.
Created on	6/19/2021, 12:13:10 AM
Last Modified	6/26/2021, 12:45:15 AM

Percentage Completed

Object Detection

100%

Thermal Object Detection

100%

Maritime Object Detection

100%

Action Detection

85%

Online Real-Time ST Action Localization

100%

Thermal Action Detection

0%

Maritime Action Detection

0%

Object Tracking

100%

Thermal Object Tracking

100%

Maritime Object Tracking

100%

Paper Submission

100%

BMVC 2021

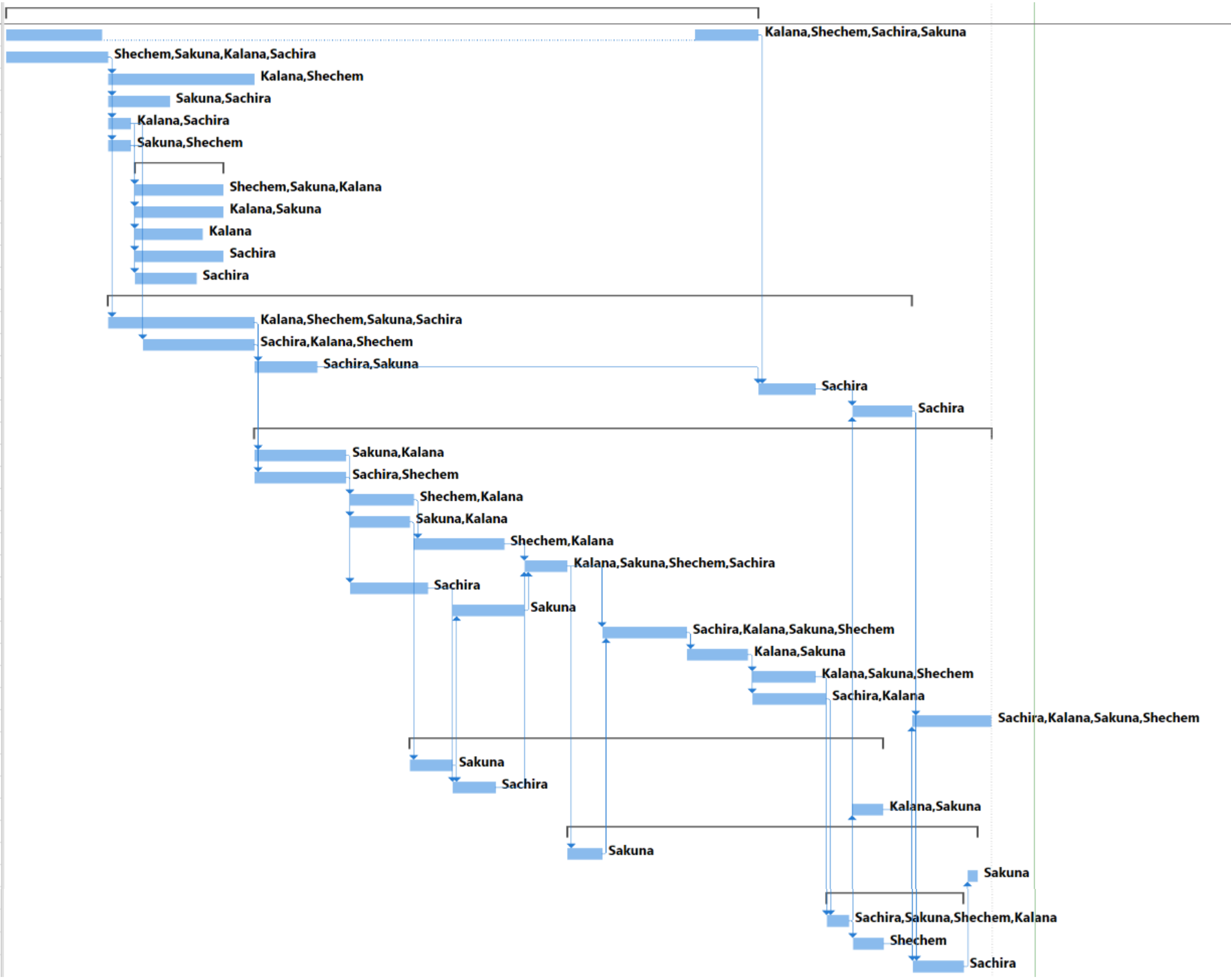
100%

User Interface

100%

Work Delegation

1	**Literature Review**
1.1	Datasets
1.2	Object Detection
1.3	Activity Detection
1.4	Suspicious Action Detection
1.5	Platform Design
1.6	Controller Design
2	**Platform Design and Controller Design**
2.1	Control system design based on IMU data
2.2	Control system design based on Stewart platform
2.3	Deriving the mathematical models for the platform design
2.4	Designing the platform using SolidWorks
2.5	Running simulations on SolidWorks on the built platform
3	**Object Detection**
3.1	Testing the object detection framework
3.2	Training FLIR Dataset on SSD, YOLOv3, CornerNet-Lite
3.3	Training SMD Dataset on CenterNet
3.4	Training SeaShips Dataset on CenterNet
3.5	Training FLIR Dataset on CenterNet
4	**Activity Detection**
4.1	Conversion of Matlab files to Python for tube linking algorithm
4.2	Replacing the SSD OD with CornerNet-Lite/CenterNet OD in ROAD architecture
4.3	Initial research on STM Net and custom block for action detection
4.4	Development of data loaders
4.5	Developing the STM Net block using Tensorflow
4.6	Developing the STM Net block using PyTorch
4.7	Maritime action data creation through AfterEffects
4.8	Thermal data generation through GAN
4.9	ICIP Paper Submission
4.10	Small motion pre-processing
4.12	Tube Linking algorithm
4.13	System evaluation for performance
4.11	BMVC Paper Submission
5	**Object Tracking**
5.1	Initial testing on OpenCV trackers
5.2	Object tracking on thermal video given by SL Navy using OpenCV tracking
5.3	Testing on SORT and DeepSORT trackers
6	**User Interface Development**
6.1	Initial design of UI
6.2	Merging the OD with tracker in UI
7	**Thermal Camera**
7.1	Setting up the thermal camera
7.2	Urban Driving dataset collection
7.3	Object detection and Inference video creation on urban driving dataset



Thank you.

References

[1] SSD: Single Shot MultiBox Detector - Liu et al.

[2] CornerNet-Lite: Efficient Keypoint Based Object Detection - Law et al.

[3] Objects as Points – Zhou et al.

[4] Real-time tracking via on-line boosting – Grabner et al.

[5] High-speed tracking with kernelized correlation filters – Henriques et al.

[6] Forward-backward error: Automatic detection of tracking failures – Kalal et al.

[7] Visual tracking with online multiple instance learning – Babenko et al.

[8] Visual object tracking using adaptive correlation filters – Bolme et al.

[9] Tracking-learning-detection – Kalal et al.

[10] Discriminative correlation filter with channel and spatial reliability – Lukezic et al.

[11] Temporal Recurrent Networks for Online Action Detection – Xu et al.

[12] Learning to Discriminate Information for Online Action Detection – Eun et al.

[13] STM: SpatioTemporal and Motion Encoding for Action Recognition – Jiang et al.

[14] A Structured Model For Action Detection – Zhang et al.

[15] STEP: Spatio-Temporal Progressive Learning for Video Action Detection – Yang et al.

[16] Online Real-time Multiple Spatiotemporal Action Localisation and Prediction – Singh et al.