Human Activity Recognition From First-Person Dog-Centric Videos

Master's in Machine Learning and AI - LJMU
C9 Cohort

Saiyana Ramisetty (ID: 974823)

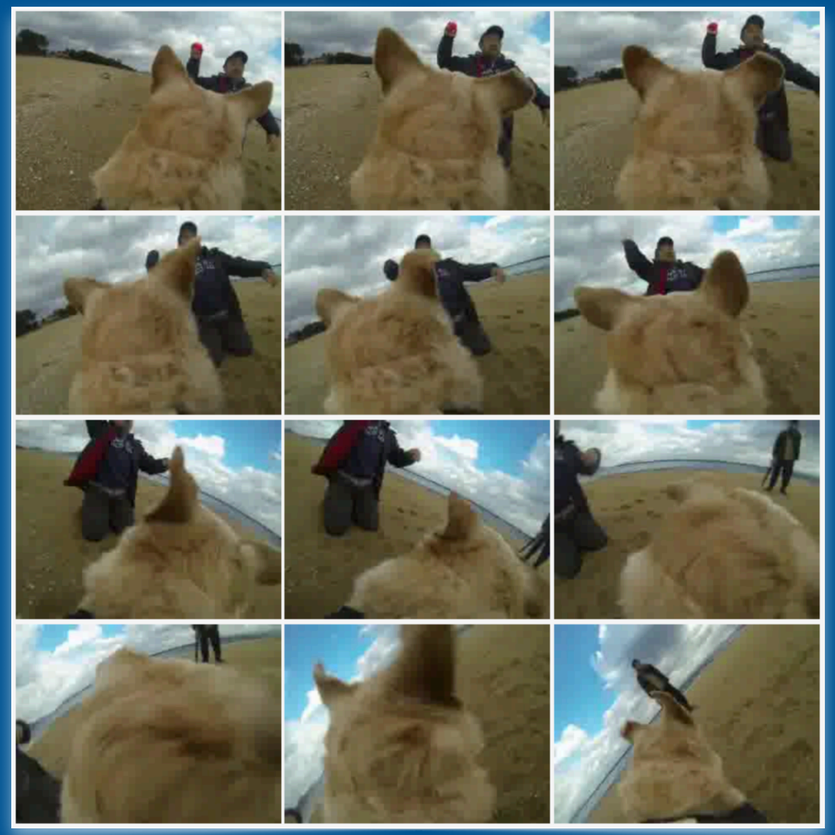
April 2022

Overview

- Background
- Problem Statement
- Literature Review
- Methodology
- Results and Discussion
- Conclusion and Future Work

Background

- First-person activity recognition
  - Camera Wearer
  - Interaction between first-person and second-person
- Animal ego-centric first-person activity recognition
- Limited data availability
- Difference between human vs animal ego-centric
  - Quadruped movement
  - Dynamic and unsteady movement
  - Data view (Height)



Problem Statement

- Current work on detecting animal ego-motion
- Minimal research on human activity detection in dog-centric data

- Detection of human activity from dog-centric first-person videos (DCAD) by extracting spatial and temporal features to classify into one of the three output classes (Pet, Play, and Feed)
- Scope: Human-animal interactions but not animal ego-actions

- Dataset availability
  - DCAD (DogCentric Activity Dataset)
  - DECADE (Dog Ego-Centric Activity Dataset)
  - Rescue Dog

Pet

Play

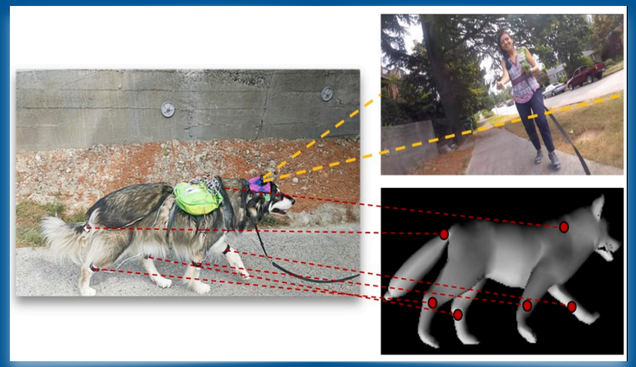
Feed

Literature Review

Dog-Centric First-Person Datasets

DogCentric Activity Dataset (DCAD)

Dog Ego-Centric Activity Dataset (DECADE)



Rescue Dog Dataset (Cyber-Rescue)

Literature Review

Related Works

- First-Person Animal Activity Recognition from Egocentric Videos (DCAD)

  Yumi Iwashita, Asamichi Takamine, Ryo Kurazume, M.S. Ryoo

- Who Let The Dogs Out? Modelling Dog Behaviour From Visual Data (DECADE)

  Diana Ehsani, Hess Bagherinezhad, Joseph Redmon, Roozbeh Mottaghi, Ali Farhadi

- Dog-Centric Activity Recognition by Integrating Appearance, Motion, and Sound (Cyber-Rescue)

  Tsuyohito Araki, Ryunosuke Hamada, Kazunori Ohno, Keiji Yanai

Methodology

DogCentric Activity Dataset (DCAD)

	Dog A	Dog B	Dog C	Dog D	Total (category)
Ball play	6	5	3	0	14
Car	7	1	14	4	26
Drink	5	2	2	1	10
Feed	7	3	8	7	25
Turn head (left)	8	4	3	6	21
Turn head (right)	6	3	4	5	18
Pet	8	4	8	5	25
Body shake	9	2	3	5	19
Sniff	8	7	7	5	27
Walk	7	4	7	7	25
Total (dog)	71	35	59	45	**210**

Number of videos of all activities in DCAD

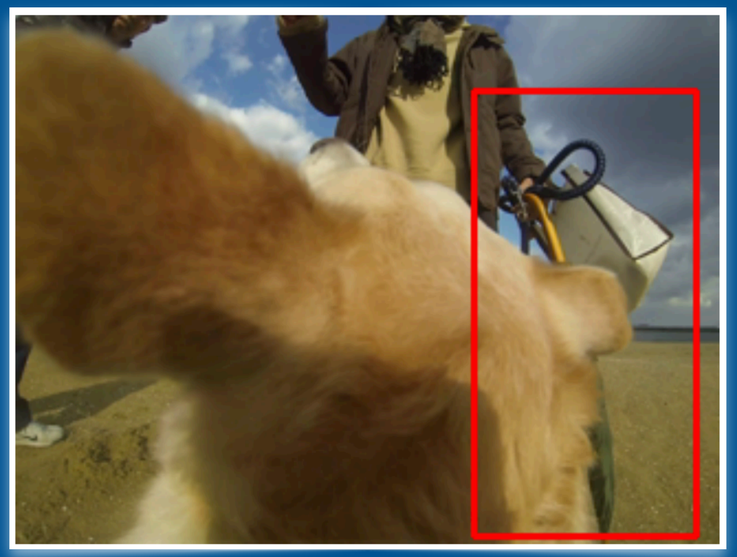
Dog vs Output Class	Pet	Play	Feed
Hime	8	3	8
Ku	8	6	7
Ringo	5	0	7
Ryu	4	4	3
Total number of videos	25	13	25

Number of videos selected for each output class in current research

Methodology

Human Appearance Detection

- Height and position of the camera
- Human appearance distortion due to heavy and dynamic movement
- Human body masking or cropping

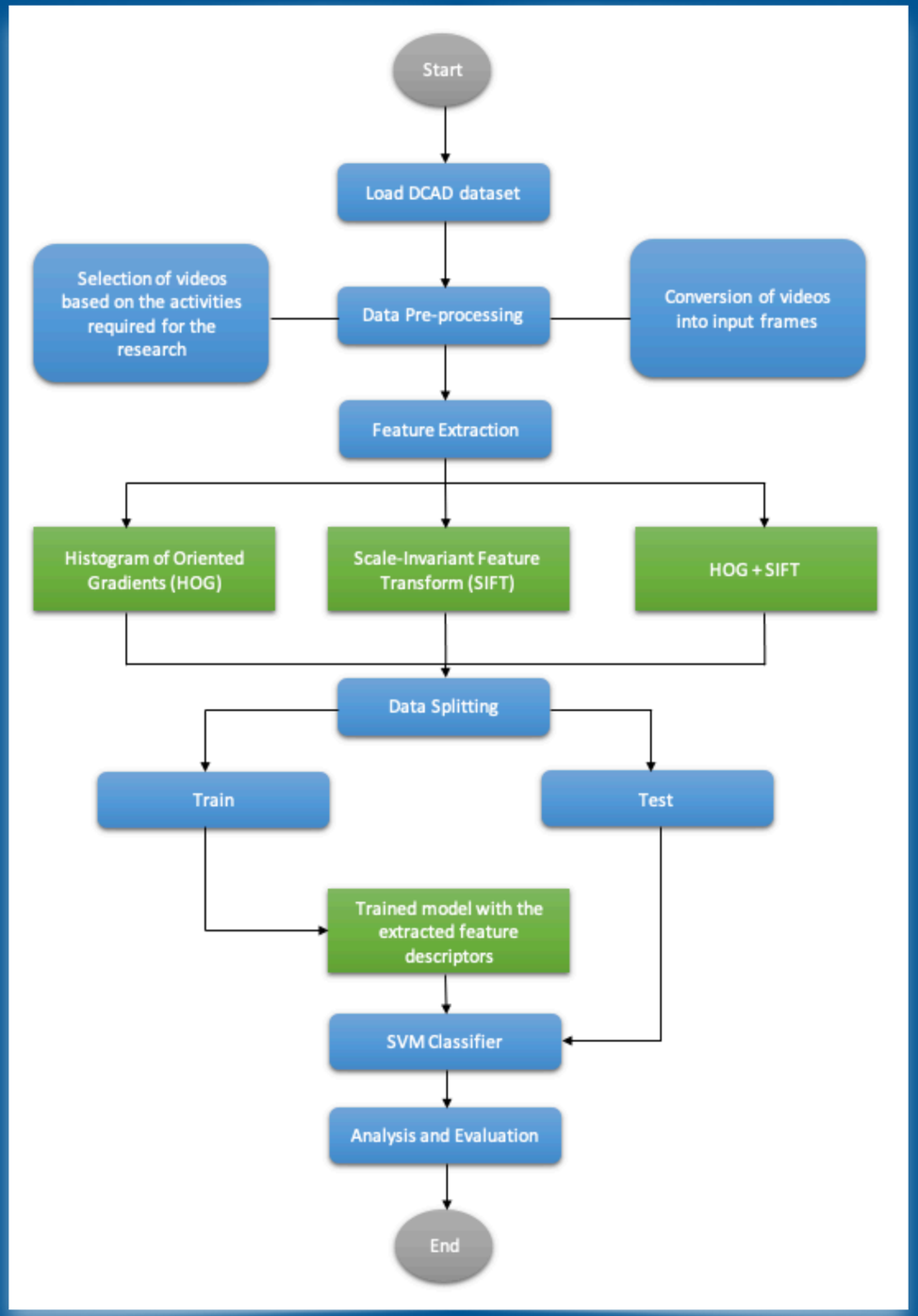


Methodology

Process Flow

Steps:

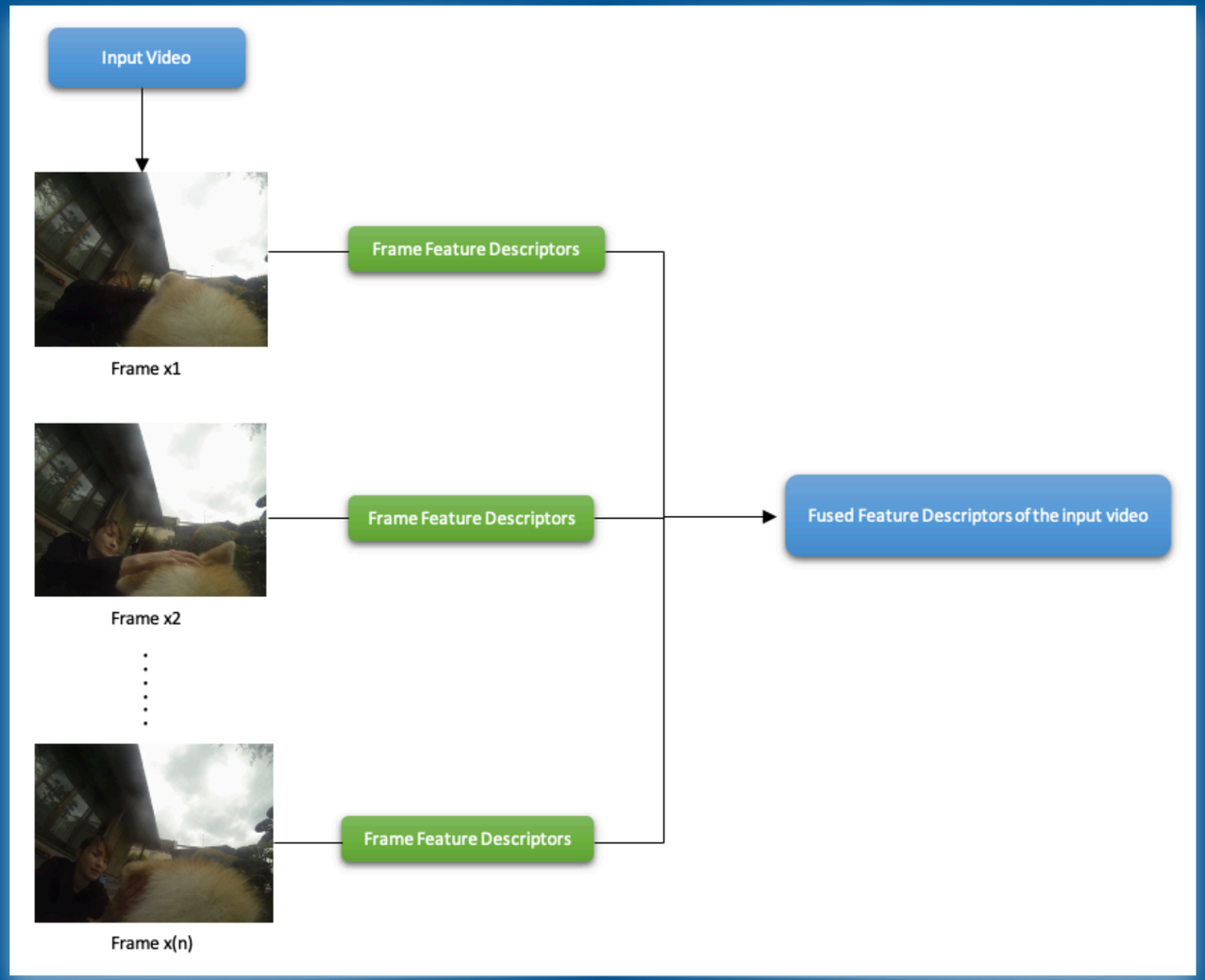
- Pre-process DCAD dataset
- Extract HOG and SIFT feature descriptors
  - Histogram of Oriented Gradients (HOG) – Temporal
  - Scale-Invariant Feature Transform (SIFT) – Spatial
- Fusion of temporal and spatial features
- Classification using Support Vector Machines (SVM) – (Pet, Play, and Feed)
- Validate effectiveness of the model on DCAD



Methodology

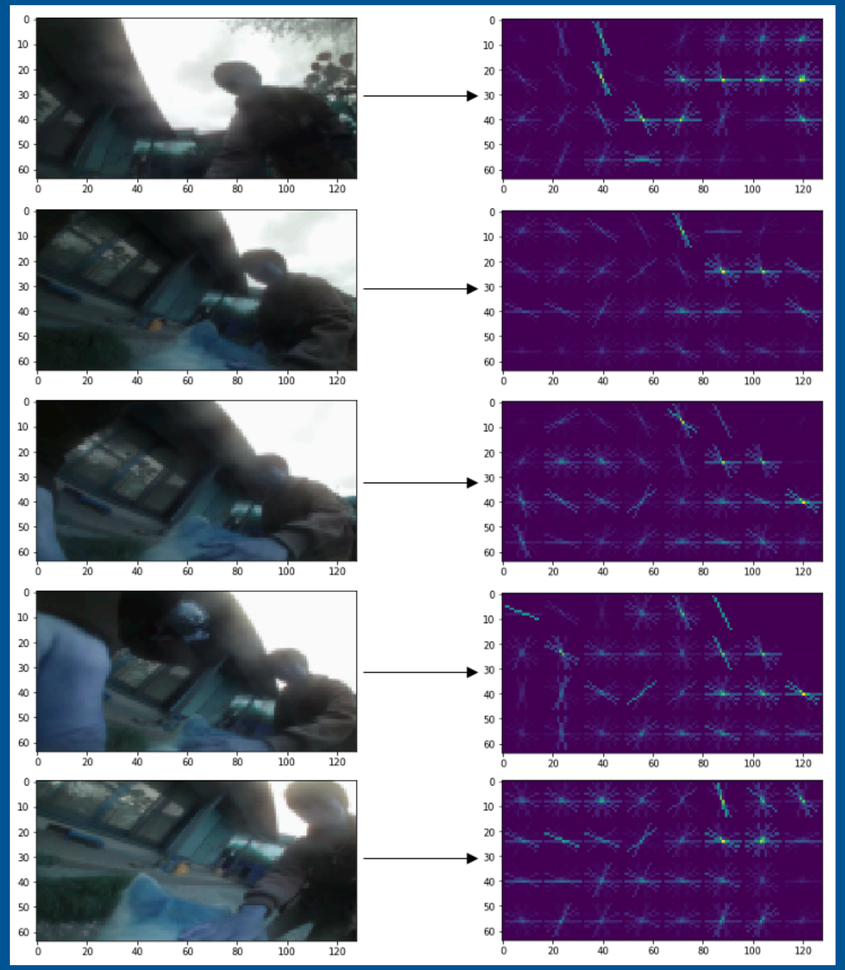
Feature Descriptors Extraction

- HOG:
  - Each frame 1x288 feature vector
  - Each video 5x288 feature vector
- SIFT:
  - Detecting 5 best key points
  - Each frame 5x128 feature vector
  - Each video 5x640 feature vector
- HOG + SIFT:
  - HOG – 5x288
  - SIFT – 5x640
  - Concatenation



Methodology

Feature Descriptors Extraction - HOG and SIFT



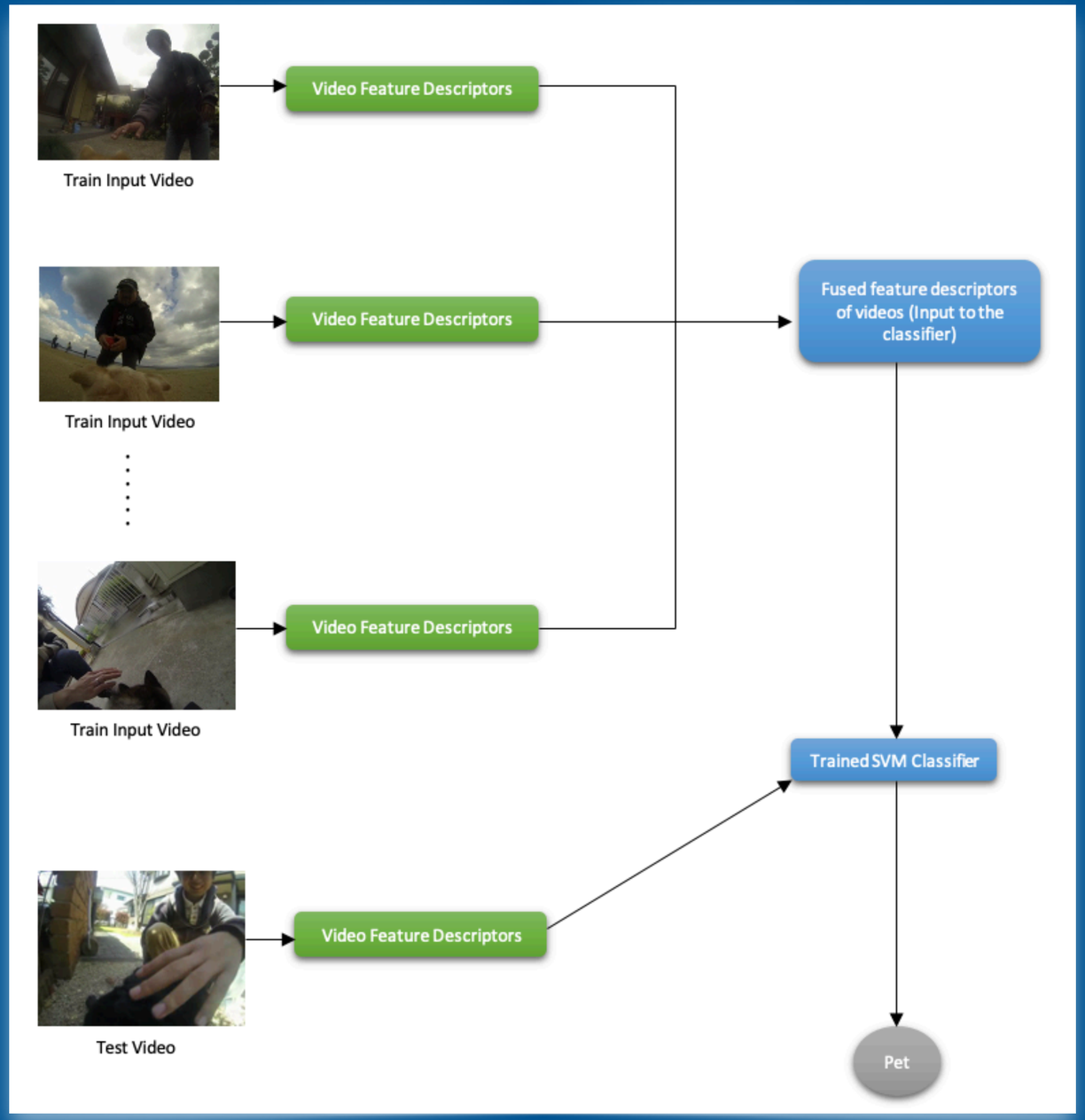
HOG

SIFT

Methodology

SVM Classification

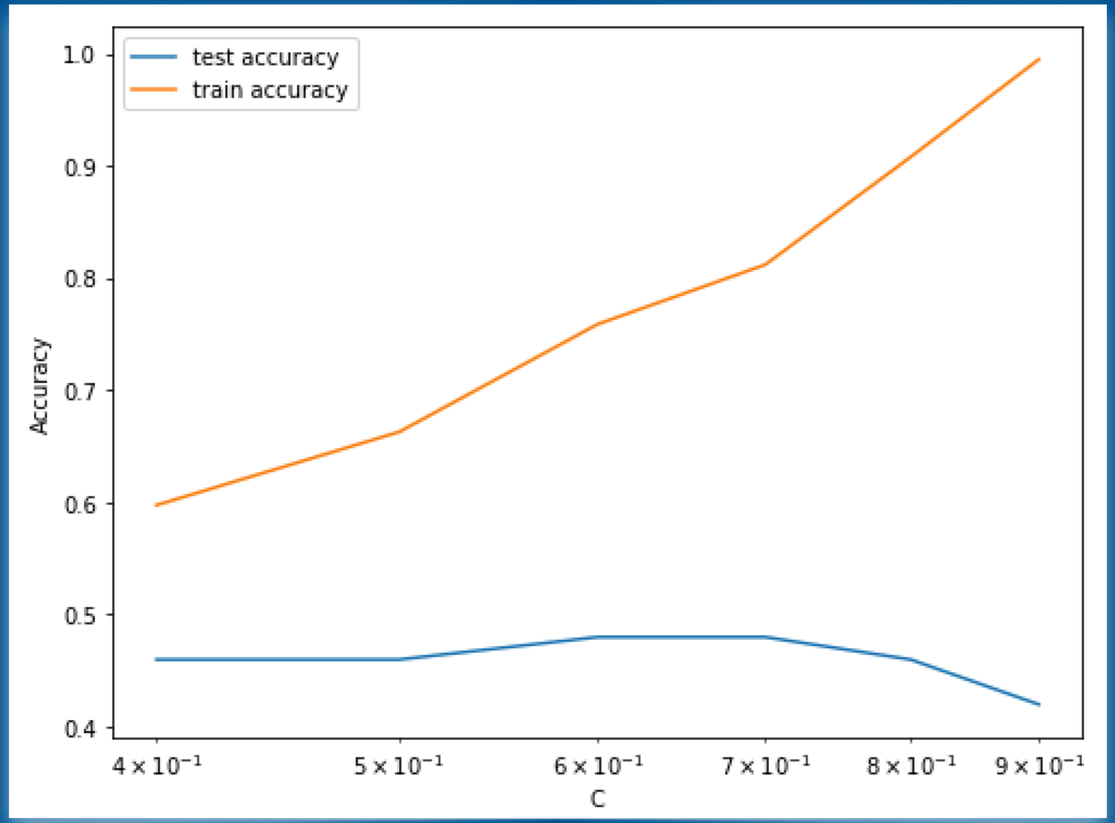
- HOG:
  - Each video 5x288 feature vector
  - 63x1440 + class_label
- SIFT:
  - Each video 5x640 feature vector
  - 63x3200 + class_label
- HOG + SIFT:
  - HOG – 1x1440
  - SIFT – 1x3200
  - 63x4640 + class_label
- Leave-One-Out Cross-Validation
- GridSearchCV to find optimal ‘C’ parameter of SVM



Results and Discussion

Histogram of Oriented Gradients

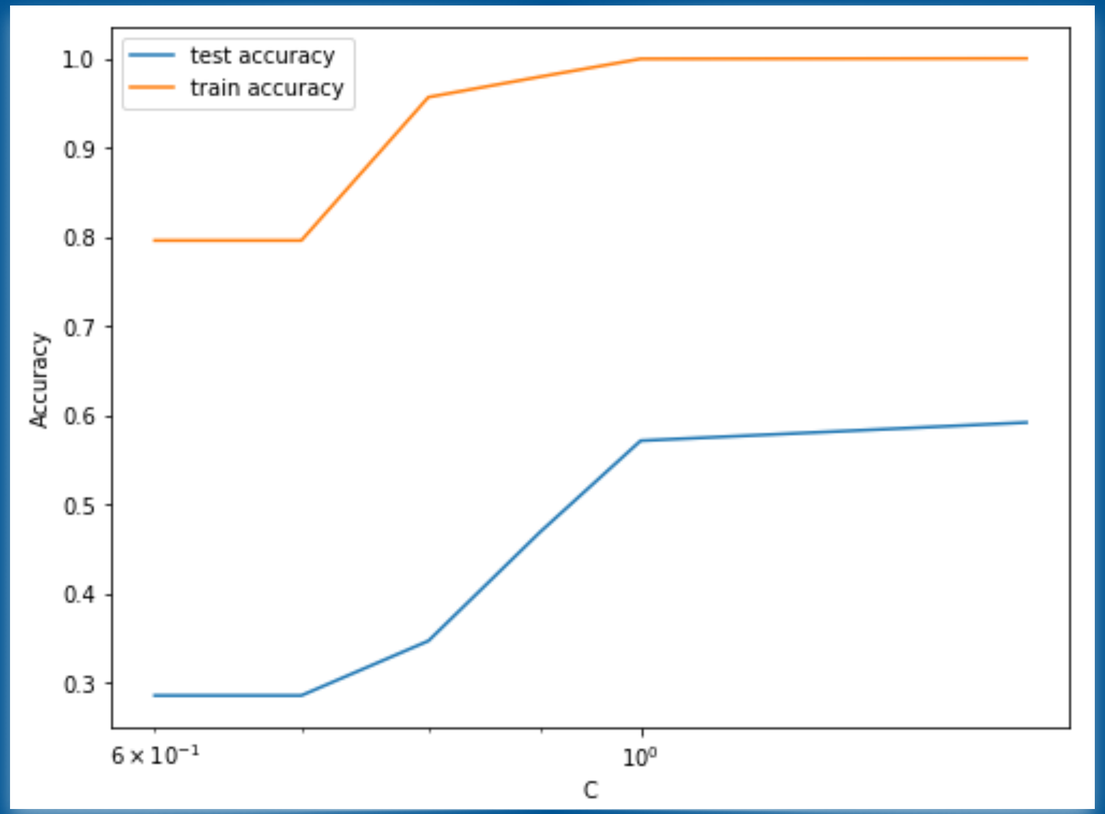
- SVM Penalty Parameter (C=0.6)
- Test Accuracy = 48%
- Train Accuracy = 75.9%
- Majority of the videos were classified as ‘Pet’ class



Results and Discussion

Scale-Invariant Feature Transform

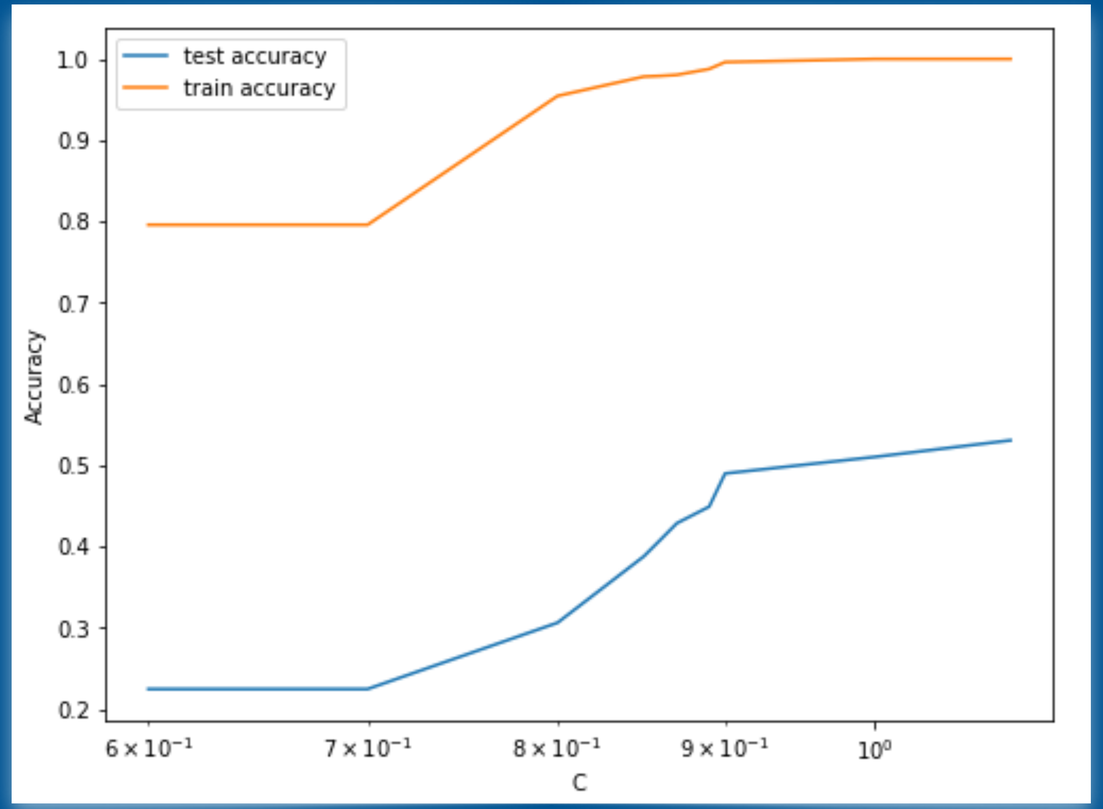
- SVM Penalty Parameter (C=0.9)
- Test Accuracy = 46.9%
- Train Accuracy = 98%
- Majority of the videos were classified as ‘Feed’ class



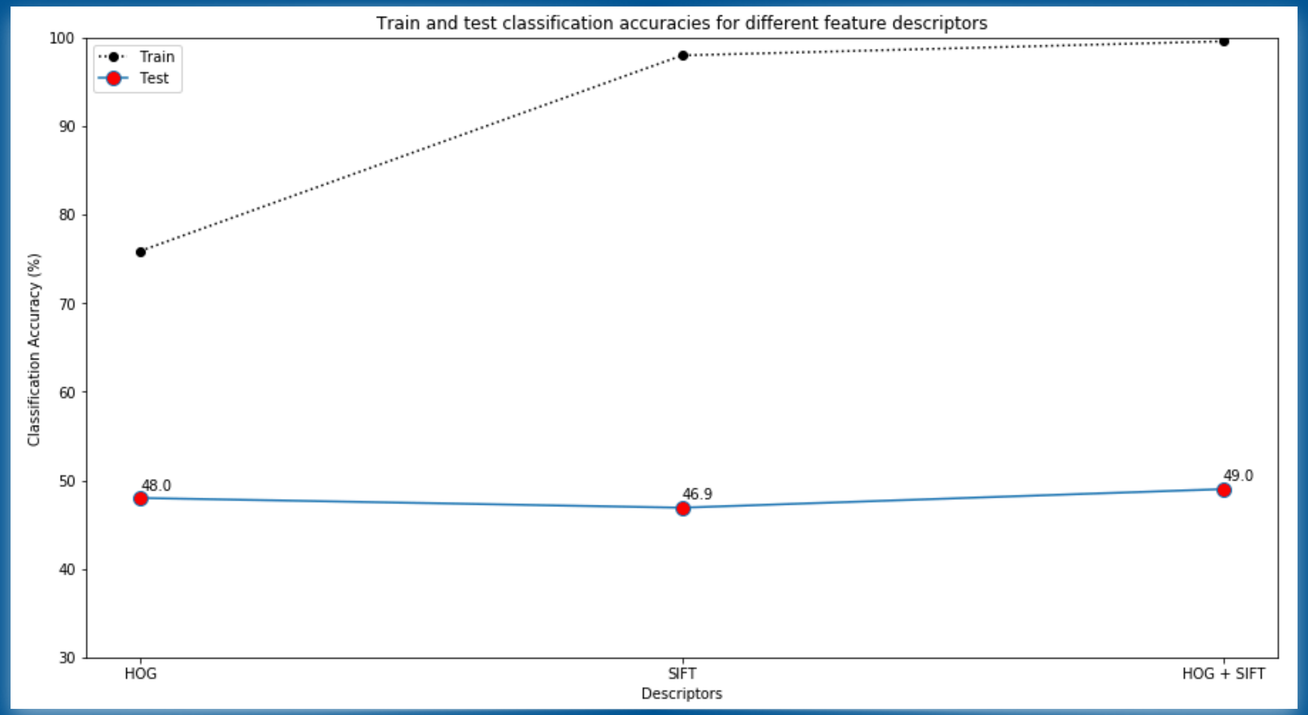
Results and Discussion

HOG + SIFT

- SVM Penalty Parameter (C=0.9)
- Test Accuracy = 49%
- Train Accuracy = 98.3%
- Classification split across output classes



Results and Discussion



Feature Extraction Method	Test Accuracy	Train Accuracy
HOG	48	75.9
SIFT	46.9	98
HOG + SIFT	49	98.3

Conclusion and Future Work

- Baseline algorithm – (49% Classification Accuracy)

- Challenges:
  - Dataset View point height
  - Dynamic and Unsteady Movement
  - Human Appearance Detection

- Dataset enrichment
- Camera position on the dog’s body
- Extraction of more features
- Comparison between DCAD and DECADE datasets

Thank you!

Image Credits: Dog collars harness and leashes

Dedicated to all dogs!