PawPal - Your Dog's Best Friend

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

EECS 504, Group 1: Aman Kumar Jha, Eric Hofesmann, Preet Gill, Vihang Agarwal

Executive Overview

In today's world it is difficult to find time to watch over your dog all day. As the world's first virtual dog sitter, PawPal tracks the activities of your dog and autonomously reinforces their behavior. PawPal strives to create a safe and loving home for both you and your dog.

Our system requires just a camera which detects your dog and the objects in your home in real time by capturing video sequences. The system learns to recognize its activities and interactions using object detection and activity classification deep networks. By classifying the activity as good or bad, the system provides audio based reinforcement to the dog in the owner's voice.

PawPal allows you to leave your dog at home with complete peace of mind!

Background and Impact

No matter how much you love your dog, it is unfeasible to spend every moment with them. In 2017, a total of 89.7 million dogs lived as pets in US households. $5.41 billion was spent on dog sitting in the US in 2015 alone[1]. However even with huge demand, current technological alternatives[2] in this growing market do not autonomously interact with your pet and require active monitoring.

Alternatives such as crating dogs is becoming increasingly unpopular as studies have shown negative physiological effects of canine crating. Our solution : PawPal not only makes the house a nurturing place for pet, it also ensures pets safety and most importantly protects your house from pet destruction while cultivating good behavior.

Method

- ➢ <u>The detection algorithm</u> (YOLO[4] - You Only Look Once)
  - Bounding boxes are obtained for pet-specific classes such as dogs and cats, and static objects like couch, pillow, sofa, tables etc.
  - After considering all state of the art detection algorithms - YOLO, SSD500, SSD300 and RetinaNet, YOLO made sense for a balance between real time implementation and accuracy.
- ➢ We define and maintain a visual relationship between animate and inanimate classes.
- ➢ <u>The classification algorithm</u> (C3D[5] - A 3D CNN architecture for video classification)
  - Process video sequences in buffers of 16 frames each.
  - Activities are recognized through spatio-temporal visual features on a discriminative video classification task.
- ➢ An ensemble of visual relationships and activity recognition is used to predict if the dog is on a piece of furniture or if it is biting objects

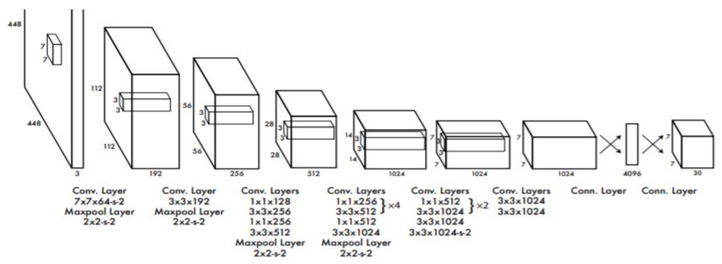

Figure 2: Network architecture for YOLO.

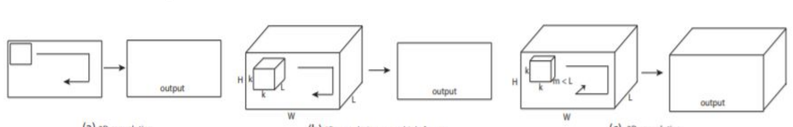

Figure 3: Schematic for C3D implementation

Prototype

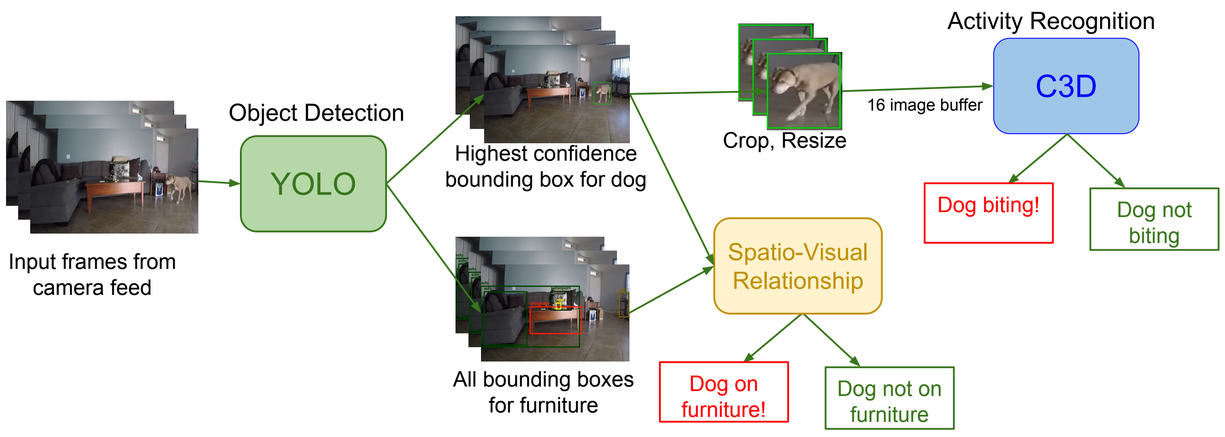

Figure 4: Information flow and schematic of PawPal Pipeline

<u>Pipeline Features</u>:

- ➢ Uses linear interpolation to obtain bounding boxes in frames with missing detections, when is dog detected with high confidence in each clip.
- ➢ Algorithm to classify if dog is on furniture based on spatial proximity of bounding box coordinates
- ➢ C3D was trained using Tensorflow to perform binary classification between 'dog biting' and 'dog not biting'
- ➢ Methods adopted to improve C3D prediction accuracy
  - Froze all layers except the last two 3D Conv and Fully Connected layers
  - Increased the dropout rate in the FC layers to avoid overfitting on limited data
  - Expanded the bounding box detections to encompass more information

<u>Dataset:</u>

- ➢ YOLO was used with pre-trained weights on PASCAL-VOC dataset.
- ➢ We designed our own dataset to train C3D. It is a collection of 862 16-frame clips extracted from 200 unique four second videos from YouTube.
- ➢ In total, 337 'biting' clips and 525 'non-biting' clips

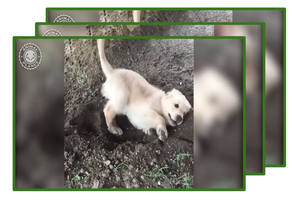

Figure 5: 'Dog not biting' dataset

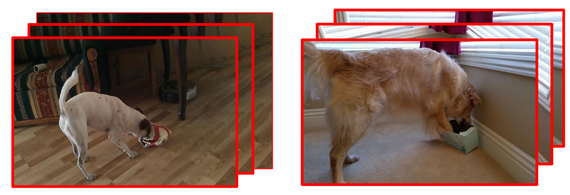

Figure 6: 'Dog biting' dataset

Results

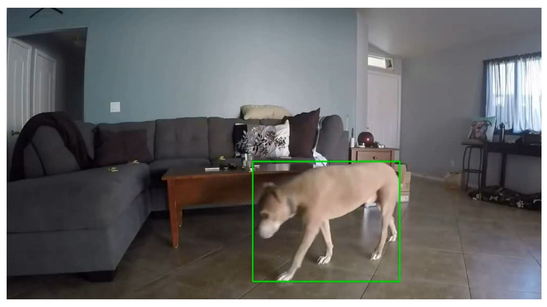

Figure 7: Frames classified as <u>dog not on furniture</u>

Frames classified as <u>dog on furniture</u>

Figure 8: Frame of the video Classified as <u>Biting</u>

Frames of the video classified as <u>Not Biting</u>

C3D Biting Recognition Accuracy		
Mean	Standard Deviation	Random Chance
68.41%	6.10%	50.00%

Table 1: Performance of C3D on the task of dog biting recognition. This is a binary classification task where videos are either labeled as dogs biting or dogs not biting. These statistics are over 5 training and testing splits from 862 16-frame clips extracted from 200 unique videos.

Conclusions & Future Work

- ➢ As a prototype, we have developed a system which can recognize pets climbing on top of pieces of furniture and biting objects by processing video feed in an online manner.
- ➢ We will provide more extensive tracking capabilities by extending pet activity classification to multiple classes including walking, running, playing etc.
- ➢ We will use robust bounding box coordinate fitting methods such as RANSAC to allow for accurate detection of each pet for households with multiple pets.
- ➢ The pipeline will be optimized to run in real time given a stream of video data.
- ➢ The hardware will include a speaker mounted on the pet which will speak to the dog in the owner's voice.

References

1. https://www.petsit.com/site/get.php?id=95691
2. https://petcube.com/
3. https://github.com/ehofesmann/PawPal
4. Redmon, J. et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
5. Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.