SwimSafe: A Computer Vision Pool Alarm

Luke Hansen

Computer Science Department, Stanford University

Stanford Computer Science

Introduction

- Children ages 1–4 are more likely to die from drowning than any other cause except birth defects
- 87% of these drownings occur in home pools or hot tubs.
- Many children who drown gain unsupervised access to a pool
- Current surveillance techniques are unreliable

Problem Statement

Better surveillance of home pools could save the lives of many children by alerting unaware parents that a child has entered their pool.

This project presents SwimSafe: a binary classifier which identifies an image as having an occupied pool or an unoccupied pool. Such a classifier could be implemented to monitor a video stream of a pool and alert pool owners if someone has entered the pool.

Dataset

676 images were gathered from:

1. Swimm400 dataset (images taken during swim races)
2. Images collected from the Internet (to increase pool diversity)

Data augmentation: random rotations, random horizontal flips, and some images were cropped.

Methods

Two different approaches to creating the binary classifier were attempted:

1. Fine-tune a pretrained network using training data (called SwimSafe). Pretrained network backbones that were experimented with: VGG-19; ResNet-50; MobileNetv2; NasNetMobile
2. Implement an algorithm which uses the output of an off-shelf image segmentation model (which identifies pools and swimmers in images) to classify the image (called SegSwimSafe).

Psuedocode for SegSwimSafe classifier:

1. Run the Image Segmentation model (pretrained weights provided by MIT)
2. Identify all the distinct humans in the output through the contagious areas of "person" pixels
3. For each person, count the number of "water" pixels directly underneath them, and divide that value by the total number of the pixels for that person
4. If this ratio is greater than a hyperparameter called the sensitivity (determined through experimentation), classify the person as in the water

Experiments

- Each SwimSafe model with a pretrained backbone was trained for 25 epochs with a learning rate of 0.0001 and a batch size of 4 using Adam optimization
- These hyperparameters were determined after trying many learning rates and batch sizes
- SegSwimSafe only had one hyperparameter to choose: the sensitivity
- This value was determined by running the model on a subset of the training data with various candidate values and selecting the one with the best performance

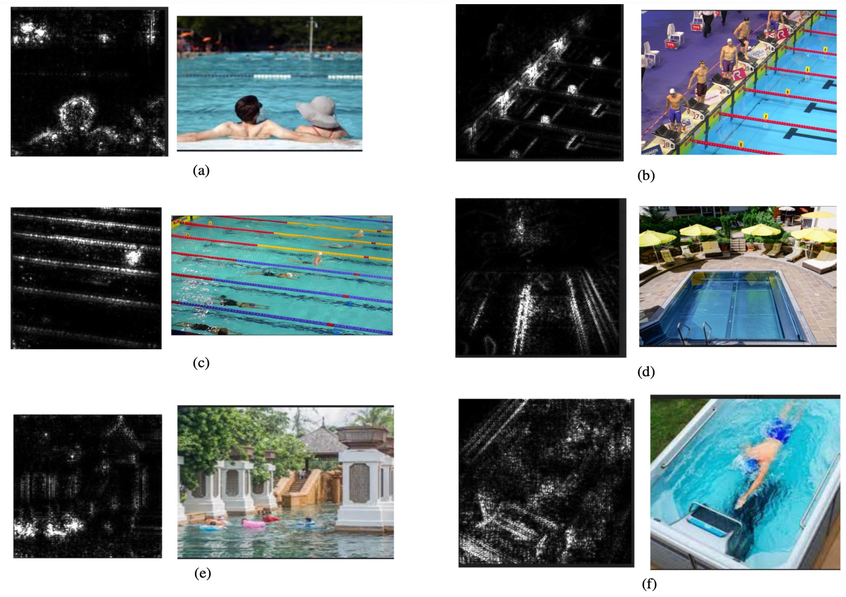

Figure 1. Correct Predictions (Images a, b, c, and d) and Incorrect Predictions (Images e and f) with Saliency Maps

Model	Val Set		Test Set	
	Accuracy	F1	Accuracy	F1
VGG-19 SwimSafe	0.970	0.978	0.942	0.956
ResNet-50 SwimSafe	0.978	0.983	0.949	0.960
MobileNetv2 SwimSafe	0.955	0.966	0.956	0.966
NasNetMobile SwimSafe	0.955	0.966	0.942	6.956
SegSwimSafe	0.955	0.966	0.912	0.934

Table 1. Comparison of Model performance on Dev and Test Sets

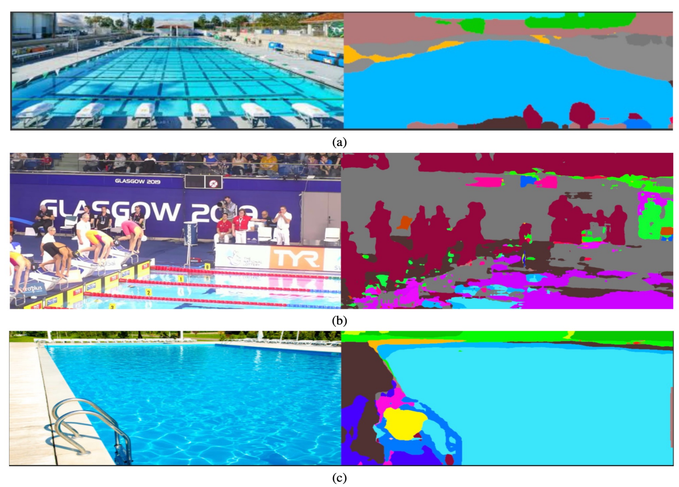

Figure 2. Misclassified Images of SegSwimSafe with Pixel Classifications (Pixels Colors: Maroon = Person; Dark Blue = Pool; Light Blue = Water)

Results

SwimSafe Models

- All of the models with a pretrained backbone performed similarly well, with only slight differences in their accuracies and F1 scores (as seen in Table 1)
- ResNet-50 backbone performed the best on the val set (accuracy .978; F1: .983)
- MobileNetv2 backbone performed the best on the test set (accuracy .956; F1: .966)
- Qualitative Analysis:
  - In Figure 1 Images a and c, the people in the image were correctly identified as being important to classifying the pool as being occupied
  - The model identified features of the pool as being relevant when it was unoccupied as can be seen in Figure 1 Images b and d
  - Figure 1 Image b contained people, but the algorithm did not flag them as being relevant to the classification--it was not a people detector
  - In Image e (Incorrectly classified), the model identified the people as being important, but did not recognize that they were in the pool. (Probably due to the abnormal properties of the pool, having a very different color and shape than other samples)
  - In Image f (Incorrectly classified), the model seemed to have identified the pool as being relevant, but not the person. (Probably due to the image having an over-the-head angle of the person in the pool. Viewing people at such an angle was not very well represented in the training set)

SegSwimSafe Model

- SegSwimSafe performed as well as some of pretrained SwimSafe algorithms on the dev set (accuracy .955; F1: 0.967)
- It was slightly less effective than the SwimSafe algorithms on the test set having the lowest metrics of all the algorithms (accuracy 0.912; F1: 0.934)
- Qualitative Analysis:
  - Generally, the mistakes made were easier to understand
  - Some mistakes were due to glaring mistakes by the segmentation model. In Figure 3 Image a, a diving board was identified as human
  - In Figure 3 Image b, the segmentation model classified the swimmer's reflection in the water as "person" pixels, causing the image to be misclassified.
  - The classification algorithm was fragile to small, random mistakes made by the segmentation model. In Figure 3 Image c, a very small portion of the image (in the nook of the pool ladder) was categorized as a human, causing the image to be misclassified.

Conclusions

1. Consistent with other research, transfer learning can be used to create effective classification models for tasks without significant amounts of training data
2. General methods which leverage computation outperformed a hand-crafted algorithm infused with human knowledge. In this work, learning based methods were the most effective, but at the cost of some interpretability of the model output
3. Computer vision techniques show promise in helping reduce the number of drownings in unsupervised home pools.

More data would allow the exploration of more sophisticated computer vision techniques such as Video Action Understanding, instead of approaching the problem as a frame-by-frame binary classification task.