Improving Gender Classification with Object Tracking

EECS 445 Final Project

**Max Mixikar, Spencer Buja,
Jesse Halpern, Austin Urlaub**

Problem

Our project aims to use an object tracking convolutional neural network to improve the accuracy of gender classification on video data. Our goal is to improve upon typical gender classification algorithms by taking a majority vote on sequential frames of output from an object tracker as opposed to a single frame.

Significance

Gender classification is significant because of its ubiquitous applications. First, an application could increase the effectiveness of targeted advertisements. Better advertisement targeting can increase sales, decrease the amount of money spent on ineffective ads , and increase customer satisfaction. For example, local businesses could use gender classifications to estimate the number of males and females among its local population and run targeted promotions. Second, gender classification could improve video search because a user could specify the gender of people appearing within a video.

In our project, we increase the accuracy of gender classification above our single-frame benchmark through aggregation of class labels over multiple frames. By increasing its accuracy, we increase its utility for future applications.

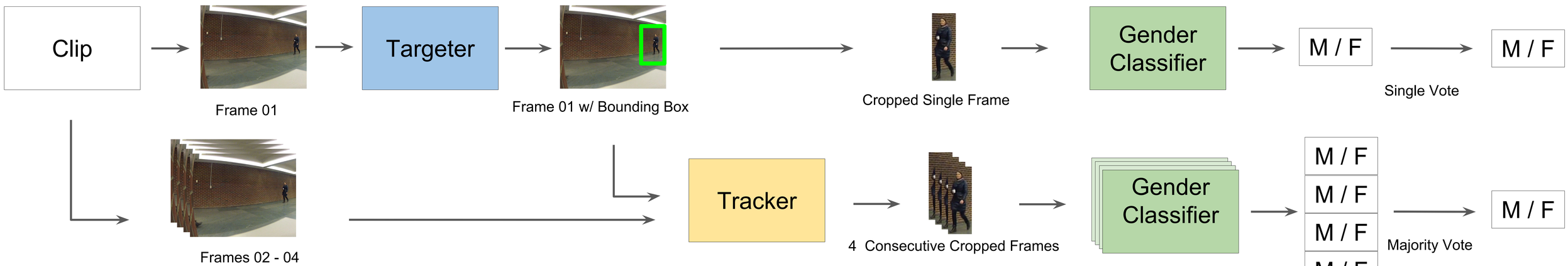

Figure I: The pipeline begins with a clip split into frames. The targeter creates a bounding box around Frame 01. Frames 01-04 are fed to the tracker. A CNN classifies gender of Cropped Frames 01-04. Single-frame classification gives the label of Cropped Frame 01. Multi-frame classification is the majority vote of Frames 01-04.

Pipeline

Targeter: The targeter determines the bounding box of all the people in the first frame. Both sections of the pipeline use the targeter, so we use an oracle targeter to prevent skewing results. All bounding boxes from the targeter were hand labeled.

Tracker: The tracker is responsible for tracking the targeted people in the first frame through the rest of the video's frames. The tracker does this by extracting features from the person using the VGG convolutional neural net, detecting the extracted features in the following frames, and using the position it found in the previous frame.

Gender classifier: The gender classifier uses a convolutional neural net designed by Gil Levi and Tal Hassner that has shown success in classifying the gender of human faces. We trained the convolutional neural net on our own set of training data consisting of full bodies images.

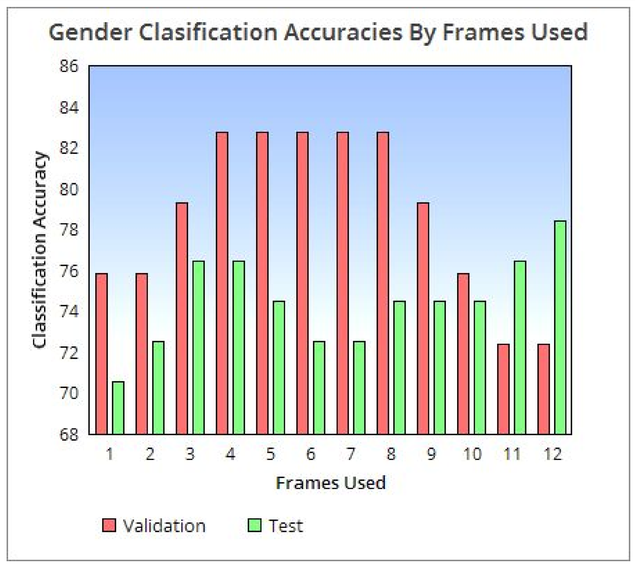

Figure II: Validation and test accuracies are higher for multi-frame gender classification than for single-frame gender classification using hold-out validation with 50 test examples, 68 training examples, and 28 validation examples, where examples are 20 frames long.

Gender classification

Single frame: For a single frame, the labelled gender is strictly determined by the output from the gender classifier.

Multi-frame: For multi-frames, we run the gender classifier over several cropped frames, outputed from the tracker. Then, we use majority voting to get the most common gender label. This is the final output from the multi-frame gender classifier.

Group's Contribution

- We built a novel application using convolutional neural networks to accomplish the tasks of object tracking and classification.