Visual Attributes in Videos

Marielle Morris
with Mahdi Kalayeh
Update Week 8

Video Attributes



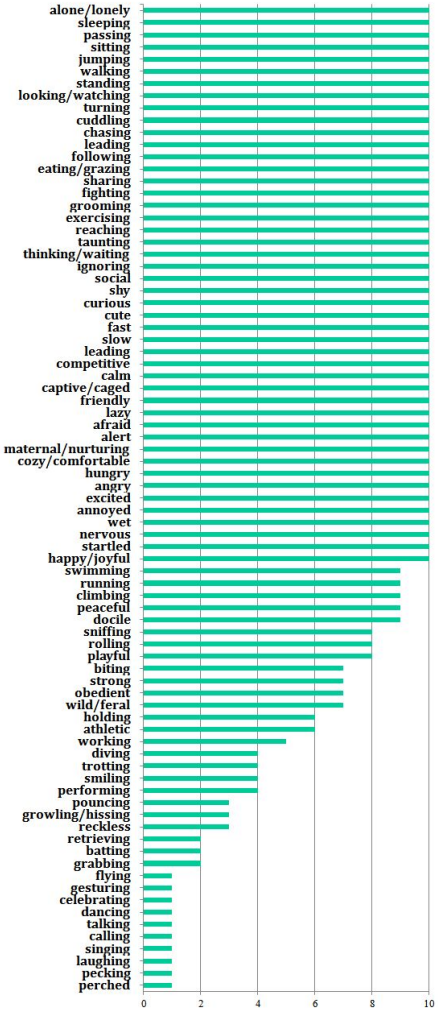


Project Goals

- Create large-scale, modern video attributes dataset
- Build off of YouTube-BoundingBoxes Dataset
  - Short video clips
  - Large-scale, publicly accessible
  - Human-annotated bounding boxes
- Design and train Deep Learning model to predict attributes

Attribute List

happy/joyful	alert	angry	excited
leading	sniffing	fighting	cozy/comfortable
following	trotting	gesturing	docile
hungry	slow	walking	calm
eating/grazing	fast	looking	playful



Creating the YouTube Attribute Dataset

- 20,000 video clips, average 17s
- 10 classes: *bear, bird, cat, cow, dog, elephant, giraffe, horse, person, zebra*
- 85 total attributes
  - Average 60 applicable per class
  - Separated into: actions, descriptives, emotions
  - Designed to be general
- Amazon Mechanical Turk for crowd-sourced annotation

Model: C3D

- 3D Conv Nets for feature extraction
- Use as baseline for attribute prediction
- Can be modified - potentially TCNN

