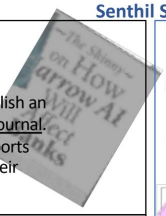
Customer Analytics: Identification and Classification using YOLO

Senthil Selvaraj / sen07 Yan Liu / liuyan84 Chu-Chi Liu / Taiwan

CS230 Stanford

Motivation



- Ng: "*AI is the new electricity*"…
- CS230 and this project inspired us to publish an article for Banks to adopt AI in a premier journal.
- Our project is to help banks, malls or airports identify people, detect face and classify their gender / age.

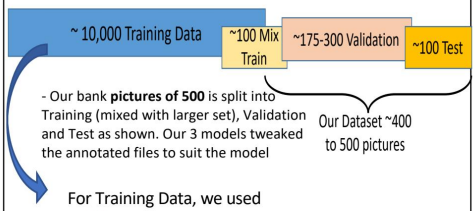
Related Work

- Redmon et al.'s YOLO real-time object detection network achieves high performance with prediction of bounding box and class probabilities in boxes.
- Hassner et. al.'s model for gender / age classification and uses 3 deep CNN for face shape, viewpoint and expressions

We use **3 models:**

#1 YOLO v2/9000 for identification
#2 YOLO v3 for face detection
#3 CNN for gender/age classification

Data Strategy

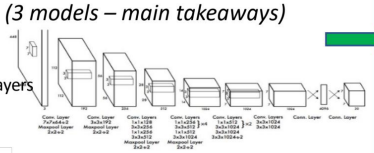


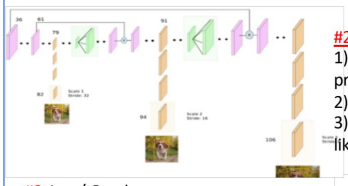
- Our bank **pictures of 500** is split into Training (mixed with larger set), Validation and Test as shown. Our 3 models tweaked the annotated files to suit the model

For Training Data, we used
- #1 PASCAL VOC for YOLO v2
- #2 WIDERFace for YOLO v3
- #3 Adience for Age/Gender

Architecture / Methods *(3 models – main takeaways)*

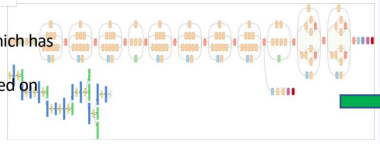
#1 - YOLO v2 for Person Id:
1) Tiny YOLO uses 9 conv and 6 pool layers
2) We modified the last Conv and Regions layers to have 2 classes (person and pets).
3) Achieved max 0.47% accuracy





#2 - YOLO v3 for Face Detection:
1) Divides the input image into small regions and predicts the b-boxes as well as the probabilities.
2) YOLOv3 has increased number of layers to 106 .
3) Due to improvements in productivity features like b-box, we used Alexab's version of YOLOV3.

#3 Age / Gender:
1) Initially, a shallow layer model (Levi) which has 70+/80+ percent accuracy) is used.
2) We then adopted transfer learning based on Inception v3 (48 layers).



Experiments / Tuning

Minibatch steps	Detections count	Unique truth count	AP	Precision Thresh 0.25	Recall Thresh 0.25	F1-score Thresh 0.25	average IoU	mAP @0.50
601K	36629	39708	52.73%	0.92	0.46	0.62	69.39%	52.73%
600K	29910	39708	58.78%	0.95	0.38	0.54	72.68%	50.79%
500K	38959	39708	50.79%	0.91	0.49	0.64	69.27%	58.78%
400K	30327	39708	51.46%	0.95	0.38	0.55	72.54%	51.46%

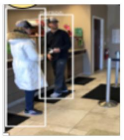
#2 YOLOv3: Achieved 58.78% with mv avg. loss < 1.0 at 600K iterations

#3 Age & Gender accuracy improved with Inception v3

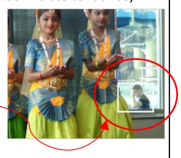
Model	Accuracy	# of Epoch	Batch Size
Levi/hassener's age	0.863905325443787	337	128
Levi/hassener's gender	0.727810650887574	315	128
Inception v3 age	0.9526627218934911	40	32
Inception v3 gender	0.757396449704142	39	32

Results

#1 - YOLO v2: As shown below, prediction is less accurate in totally different Test sets (ex. school kids in classical dance).

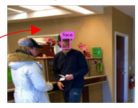


See b-box on a picture of a man but not on kids



#2 – YOLO v3 - Training on WiderFace requires modification of the dataset label format model. Due to this bounding box setup, faces that a really small were not been able to detect. Person needs to be camera-facing .

Only male has a face prediction



#3 – Age / Gender - Picture on left is face-cropped manually, so male classified in age (38-43). Female predicted as "male" in both pic as she is not facing the camera.

Age predicted in 15-20 group with full body

Conclusion / Future Work

1. Improve accuracy with more resources, data and GPU
2. Pipeline to crop pictures from live / rec videos
3. Bank product - Wait time analysis and so on.

Reference / Links

Thanks to Hojat Ghorbani and all CS230 Teaching Staff
Final Presentation 3 min video - Link
AI paper for banks - Link