NATIONAL RESEARCH UNIVERSITY

MODERN TASKS OF COMPUTER VISION RESEARCH

Andrey V. Savchenko
Dr. of Sci., Prof.,
HSE University, N. Novgorod, Russia
Email: avsavchenko@hse.ru
URL: www.hse.ru/en/staff/avsavchenko

March 3, 2022

Outline

1. **Brief overview of 2D CV tasks**
2. **User Preference Prediction in Visual Data on Mobile Devices**
3. **Facial processing and emotion recognition**
4. **Food classification and restaurant recommendation**
5. **Understanding advertisements**

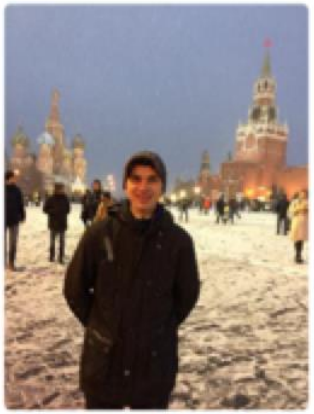
Brief overview of 2D CV tasks

Image recognition

(a) Siberian husky

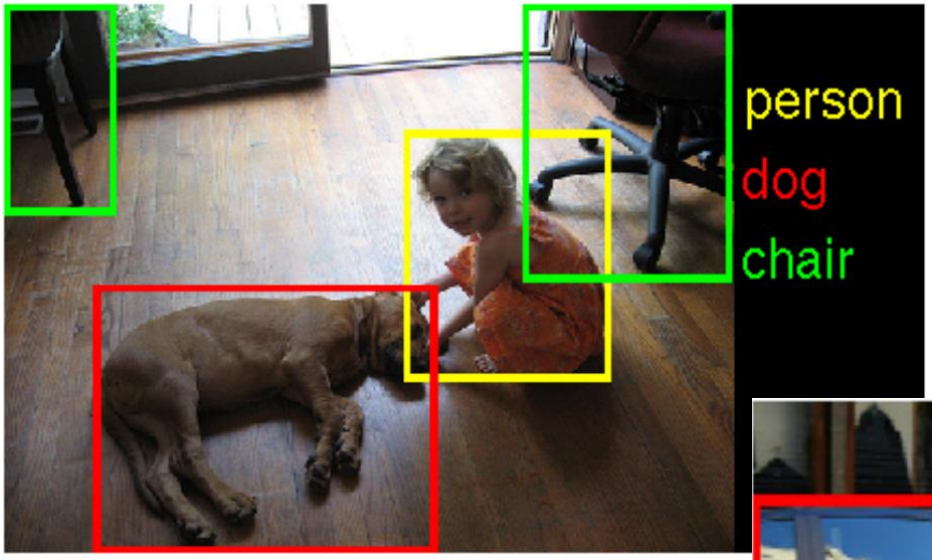
(b) Eskimo dog

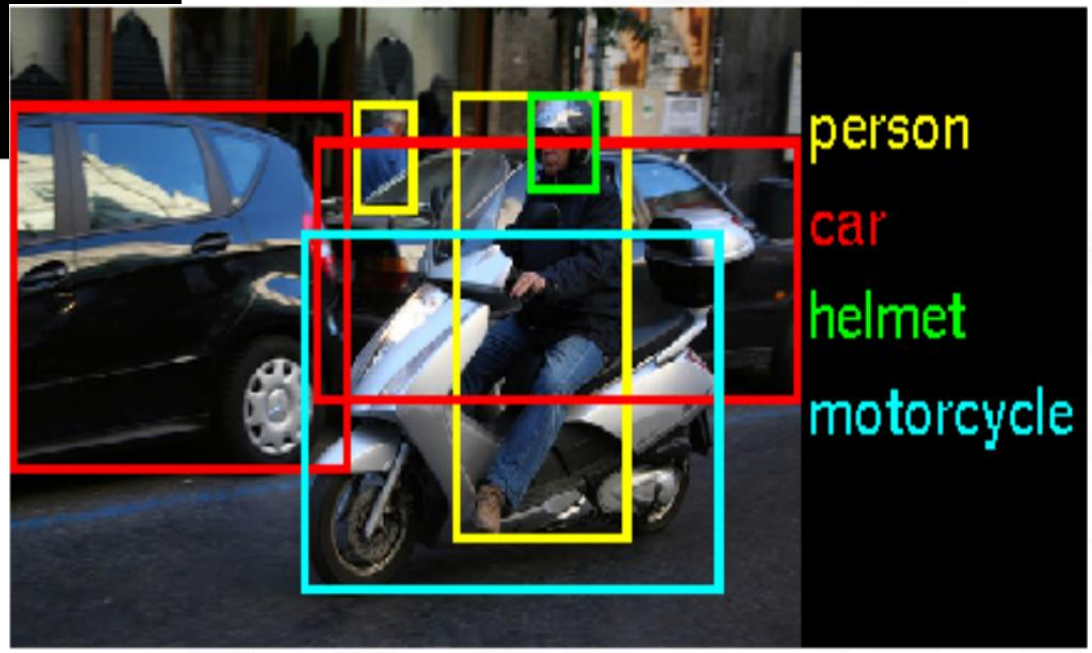
Wolfram ImageIdentify

ImageIdentify[]

bastion

Object detection

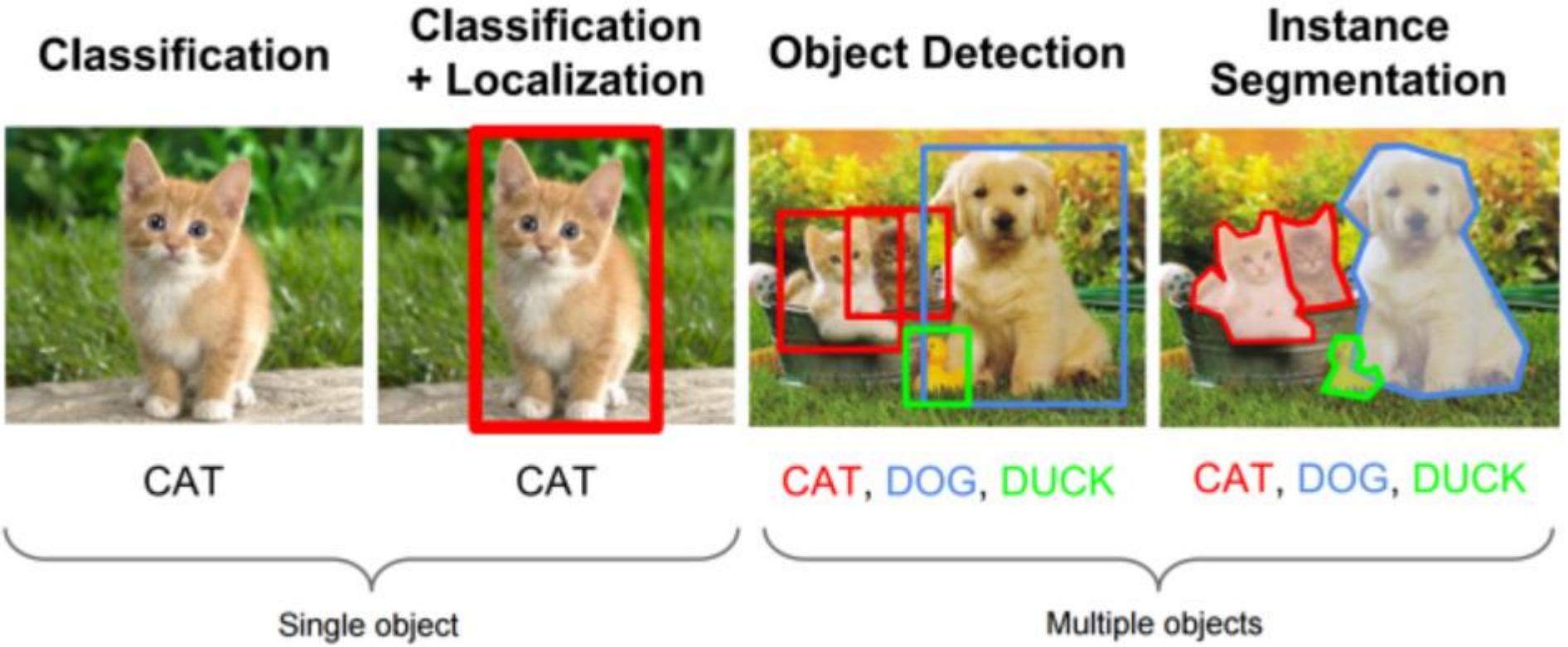




Semantic segmentation

object classes	building	grass	tree	cow	sheep	sky	airplane	water	face	car
bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat

Classification, detection, segmentation,…



https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/object_localization_and_detection.html

Image captioning

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 A person riding a motorcycle on a dirt road.	 Two dogs play in the grass.	 A skateboarder does a trick on a ramp.	 A dog is jumping to catch a frisbee.
 A group of young people playing a game of frisbee.	 Two hockey players are fighting over the puck.	 A little girl in a pink hat is blowing bubbles.	 A refrigerator filled with lots of food and drinks.
 A herd of elephants walking across a dry grass field.	 A close up of a cat laying on a couch.	 A red motorcycle parked on the side of the road.	 A yellow school bus parked in a parking lot.

http://arxiv.org/abs/1411.4555 “Show and Tell: A Neural Image Caption Generator”

Style transfer

https://arxiv.org/abs/1508.06576 “A Neural Algorithm of Artistic Style”

Prisma

Recognition of paintings

Predicting Image Memorability

- https://quickdraw.withgoogle.com/
- https://github.com/tensorflow/docs/blob/master/site/en/r1/tutorials/sequences/recurrent_quickdraw.md
- MemNet, https://arxiv.org/abs/1708.02209
- LaMem, http://memorability.csail.mit.edu/demo.html

User Preference Prediction in Visual Data on Mobile Devices

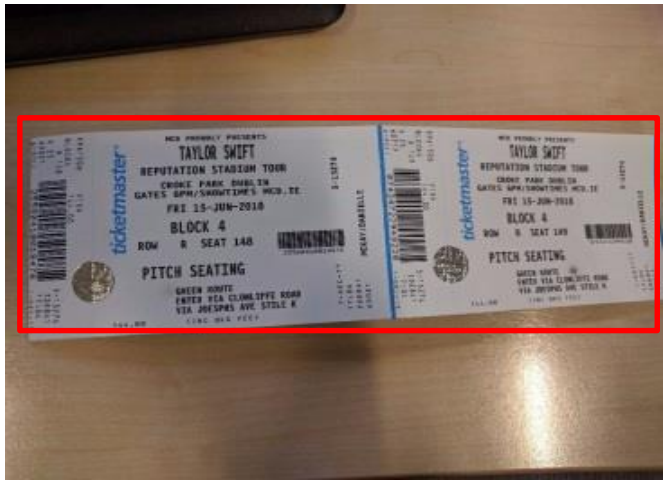
- Savchenko, Demochkin, Grechikhin, Pattern Recognition 2022
- Savchenko, Kopeykina, RusAutoCon 2019, ITNT 2020
- Savchenko, Rassadin, ISNN 2019
- Grechikhin, Savchenko, IbPRIA 2019
- Savchenko, Samsung US Patent 2022

Funded by

SAMSUNG

Deep understanding of user characteristics by analyzing user images and videos in a mobile device

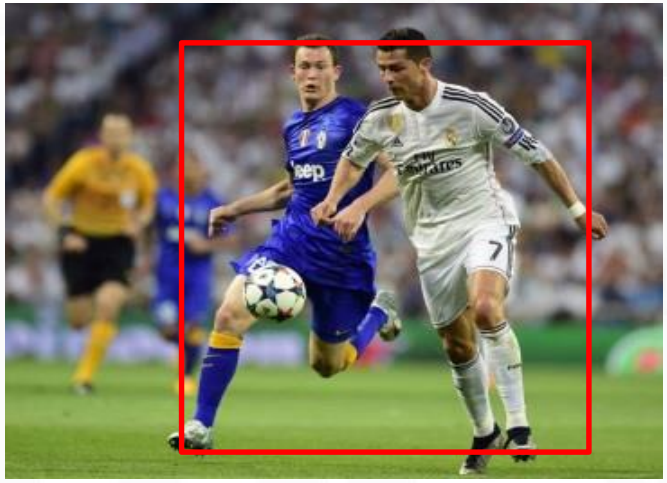
Singer : Taylor Swift



Game : Nintendo Switch

Food : Fried Chichen

Sports : Soccer



User's profile

Categorizing user's characteristics (taxonomy, classification, demographics, hobbies, occupation, lifestyle, etc.) → Generate user profile

Hobbies	Food code	Pets
Restaurant	Junk food	Dog
Beach	Health - salad and etc.	Cat
Tracks	Sandwich	Fish
Bar	Meat	Horse



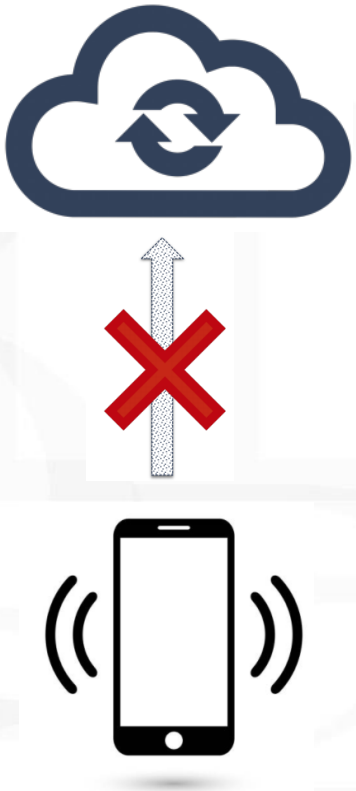
Sports	Household Income	Vacation
Fishing	Age	Ski
Golf	Gender	Museum
Diving	Car types	Tracks
Tennis	household	Monuments



Privacy detection

Public photos

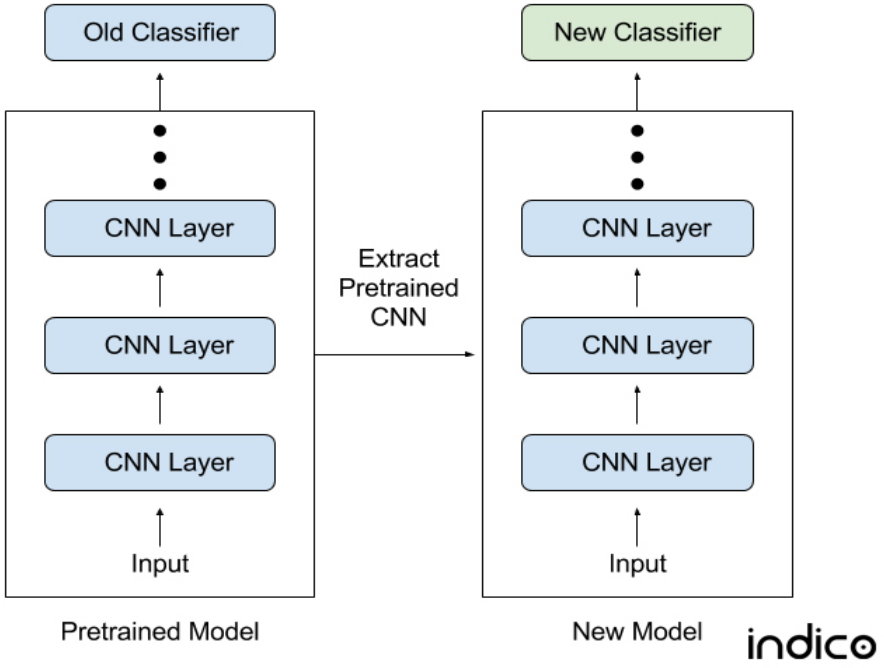
Private photos



It is required to assign an observed image X to one of C classes. Training set contains N reference images (examples) $\{X_n\}$, $n \in \{1,\dots N\}$, with known class label $c_n \in \{1,\dots C\}$

1. Fine-tune convolutional neural network (CNN) pre-trained on ImageNet, Places, etc.

2. Classify *embeddings (features)* from one of the last CNN's layers: D-dimensional feature vector $\mathbf{x}=[x_1,\dots, x_D]$. Training set is associated with embeddings $\{\mathbf{x}_n\}$



Classifier training using both visual features and results of object detection

Scene recognition

- Sport
- Nature
- Athletic field

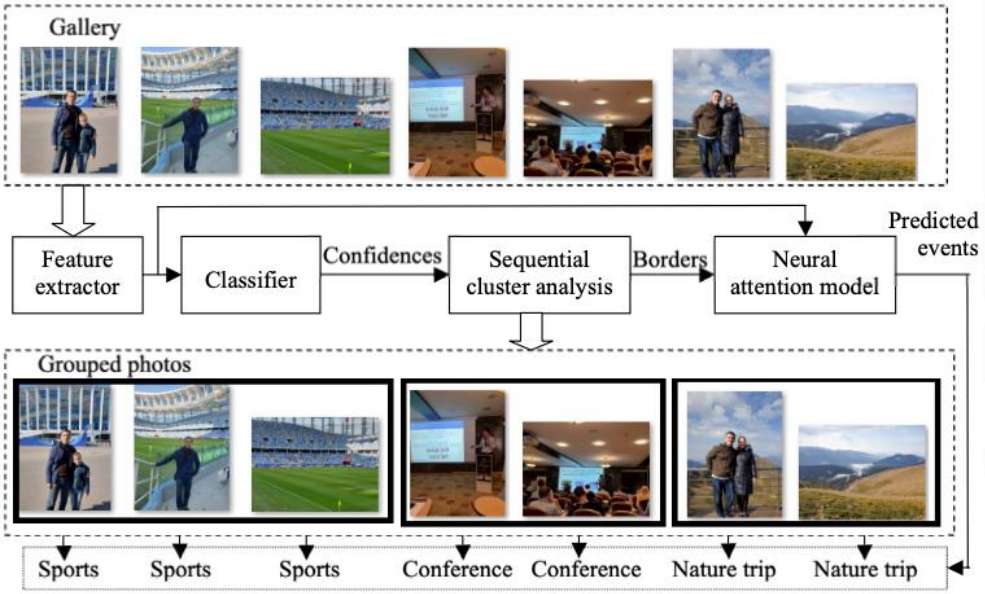
Embeddings

Object detection

- Person
- Shoes
- Tree

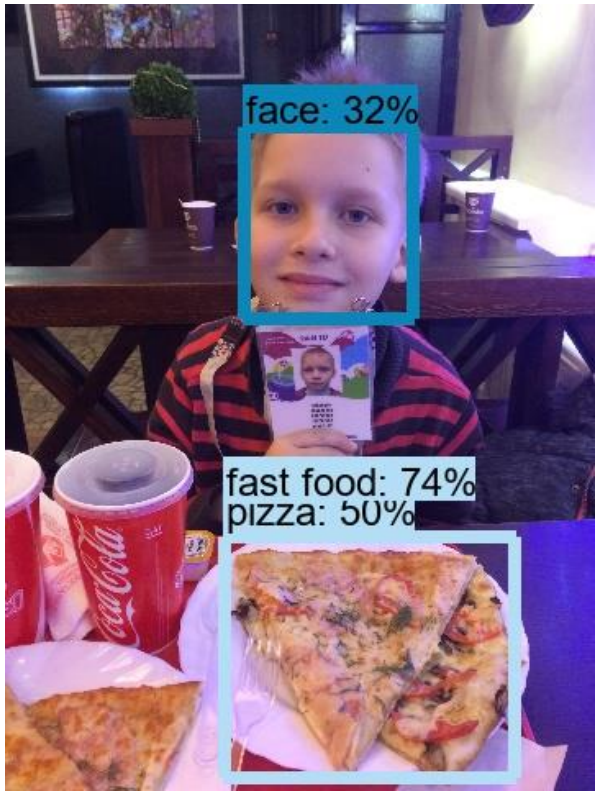
Scores

Ensemble (blending)

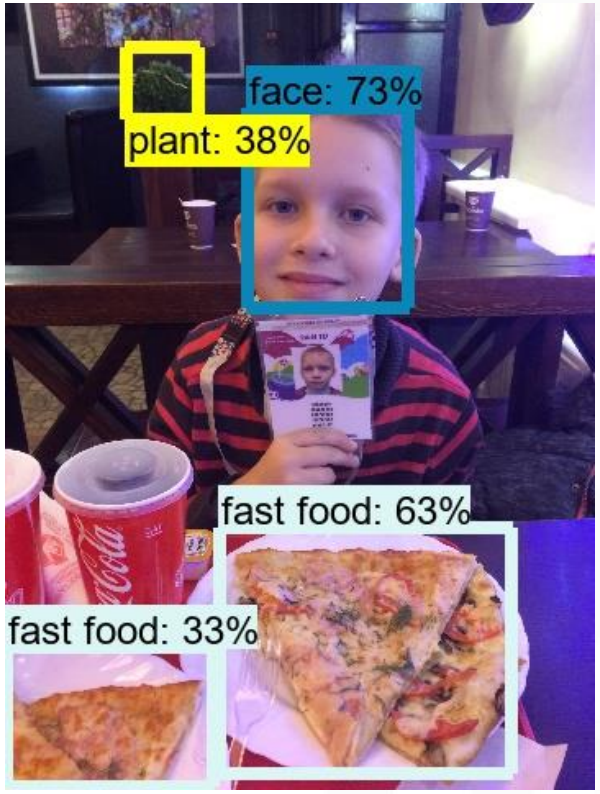


Example of object detection (1)

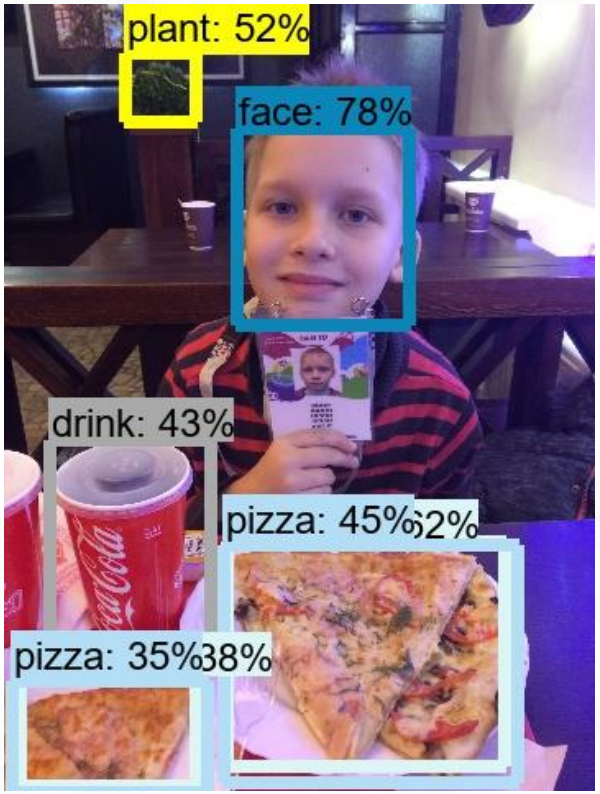
SSDLite+ MobileNet v2



Faster RCNN+ Inception v2



Faster RCNN+ InceptionResNet



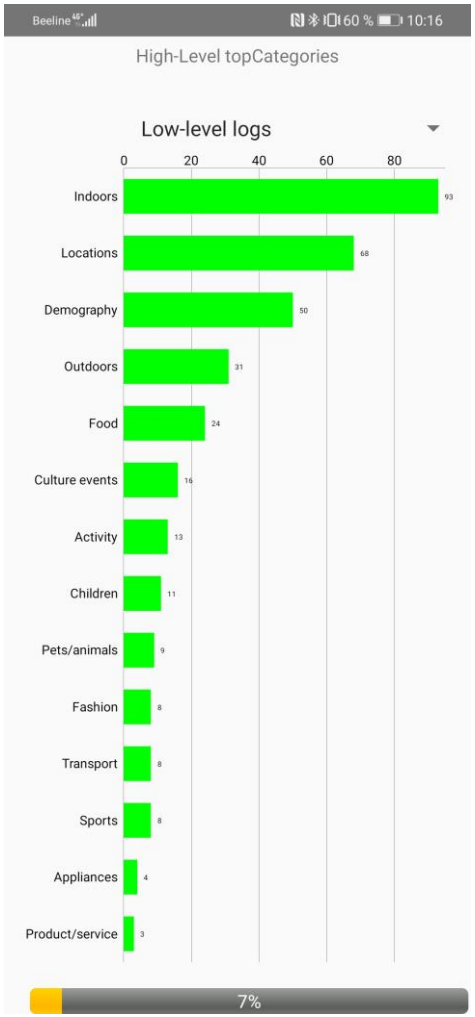
Example of object detection (2)

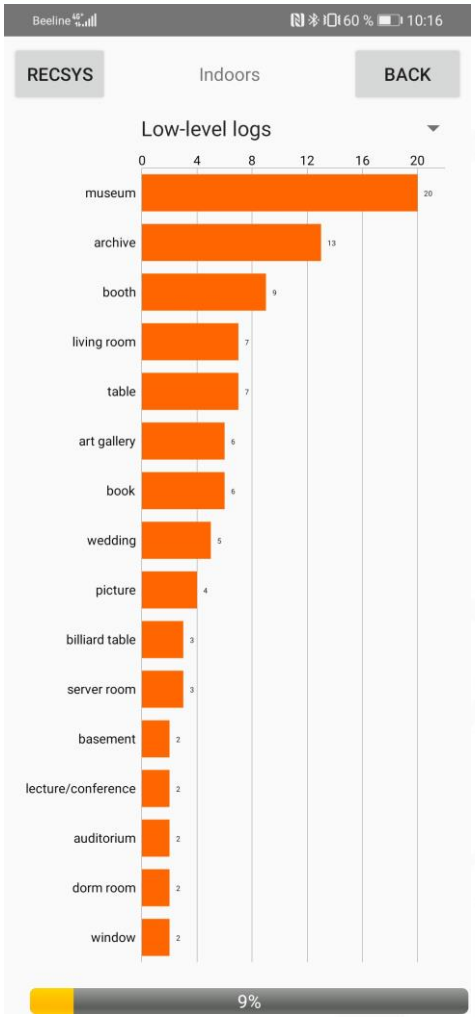
SSDLite+ MobileNet v2

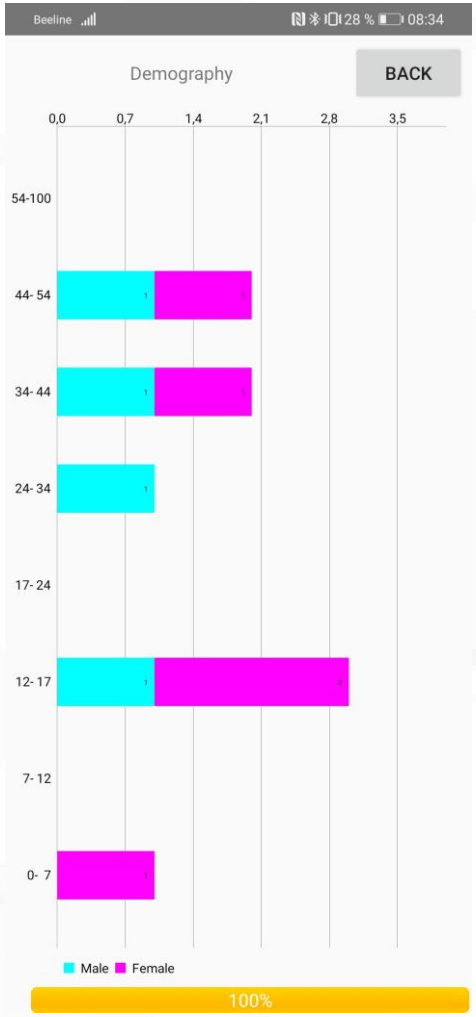
Faster RCNN+ Inception v2

Faster RCNN+ InceptionResNet

Mobile application (1)

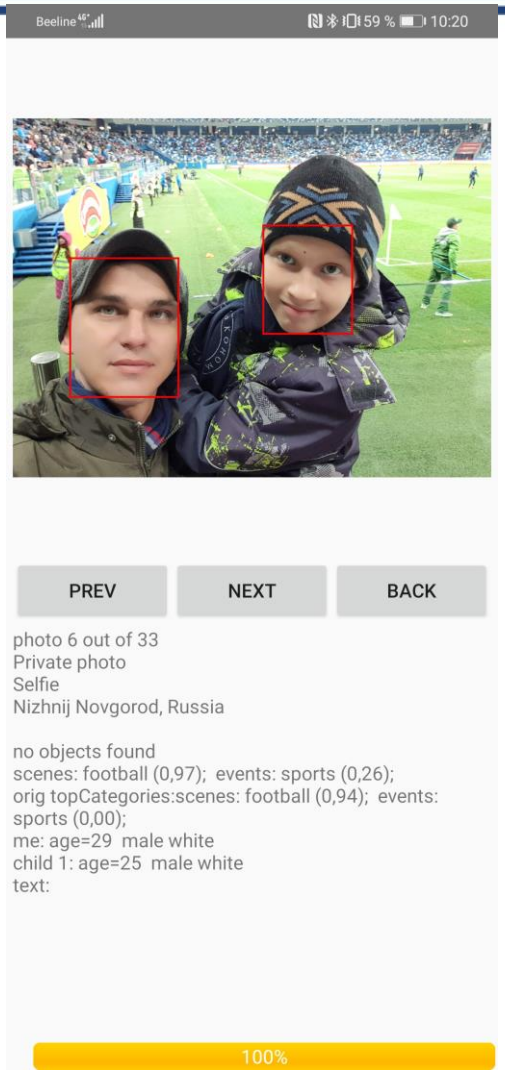


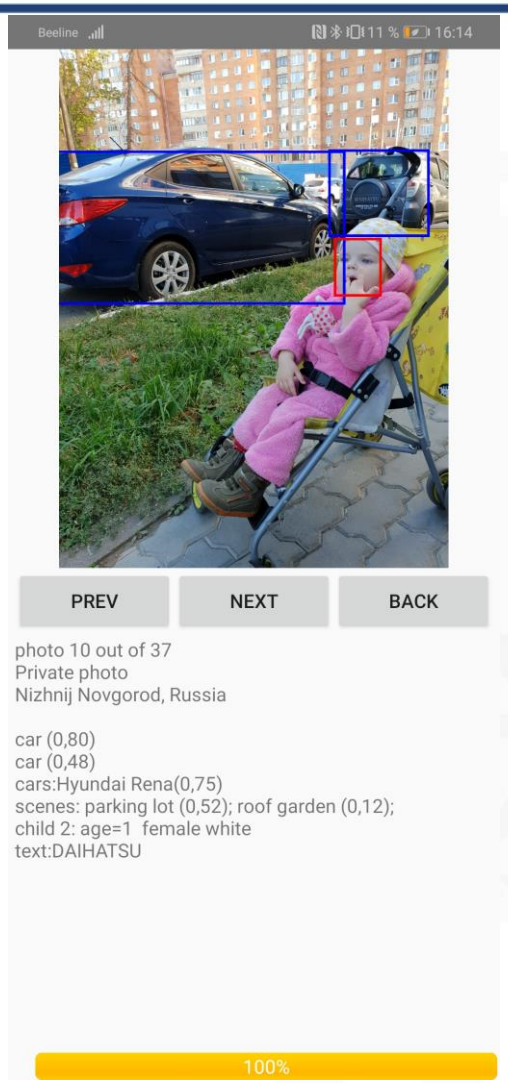


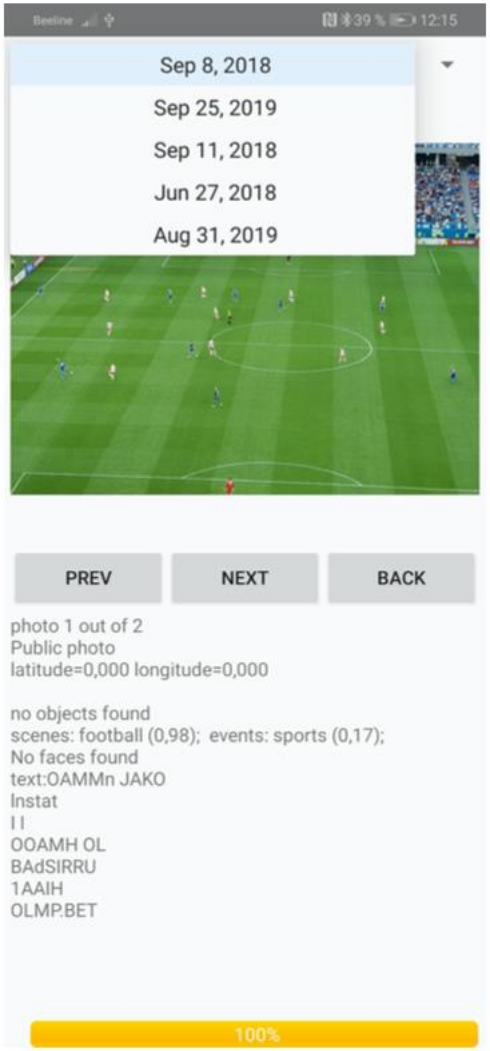


https://drive.google.com/drive/folders/1rQkJZifq_89pu0sT_UnYXziuxutTpNEN

Mobile application (2)

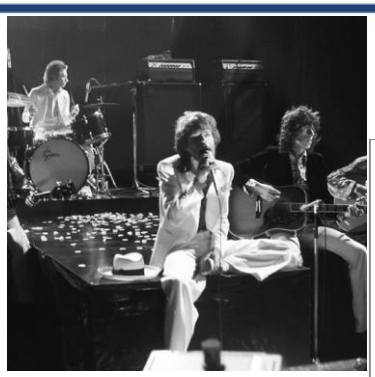




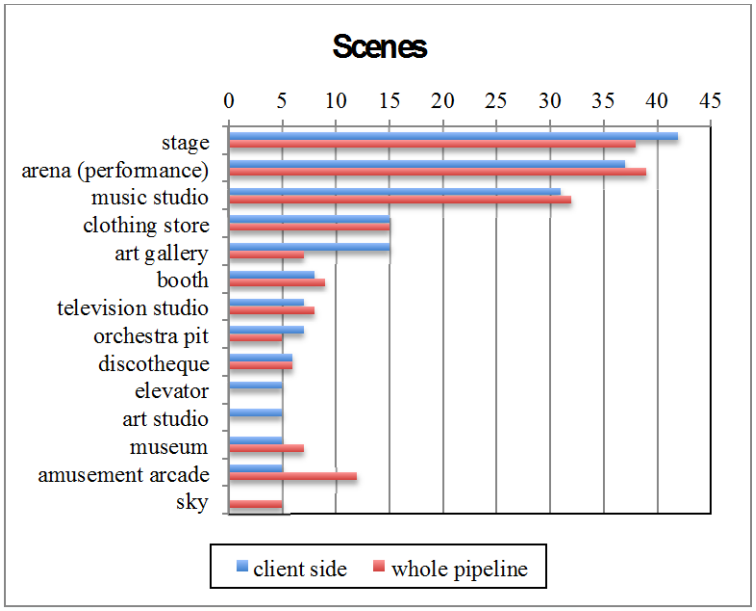


The Rolling Stones profile

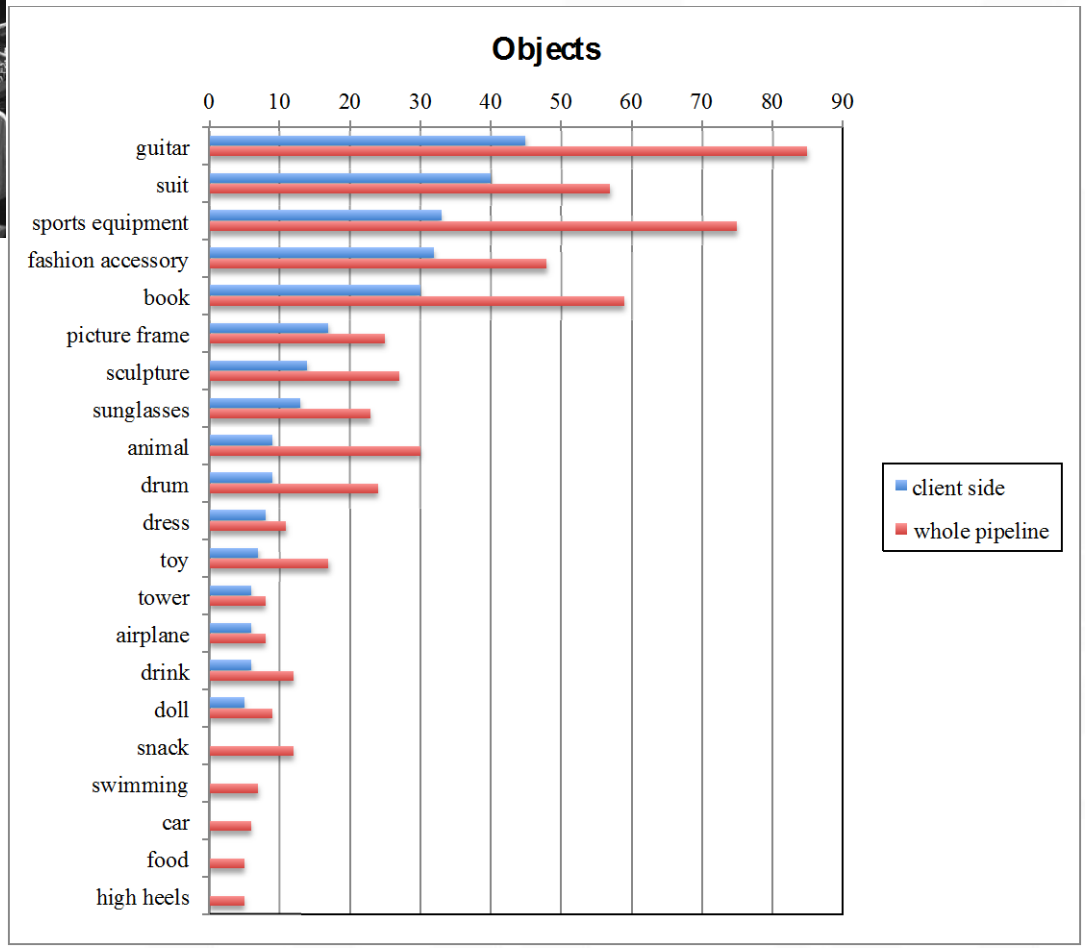
...



Scene recognition



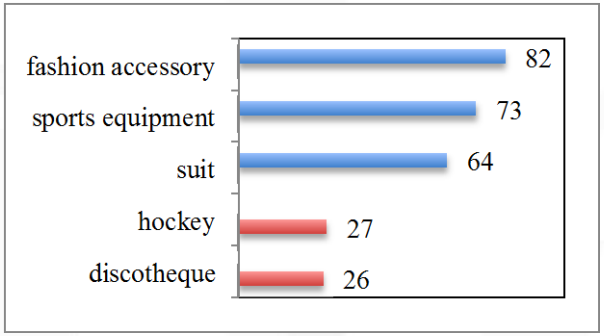
Object detection



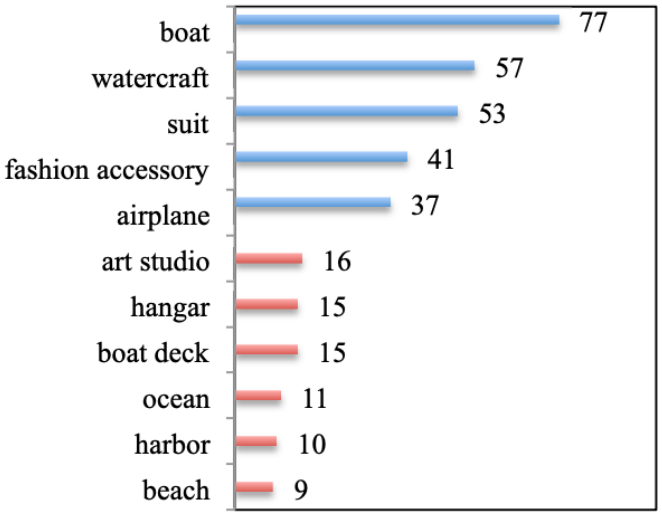
User's profiles from Instagram (1)

Alex Ovechkin

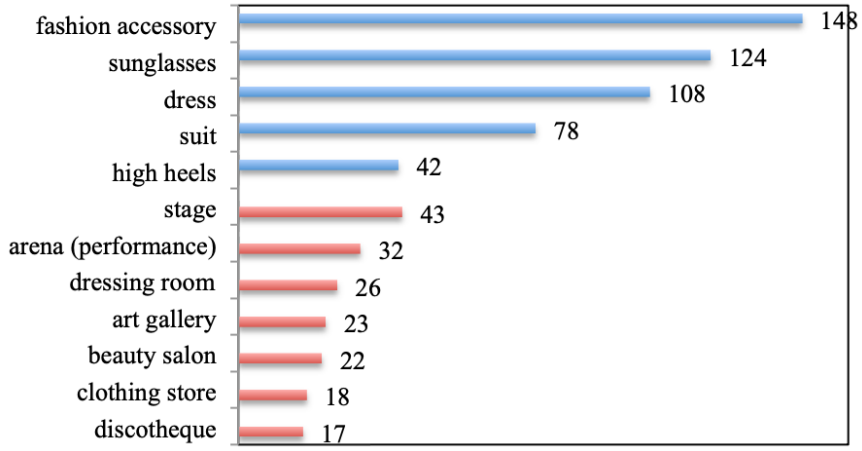
…



Fedor Konyukhov

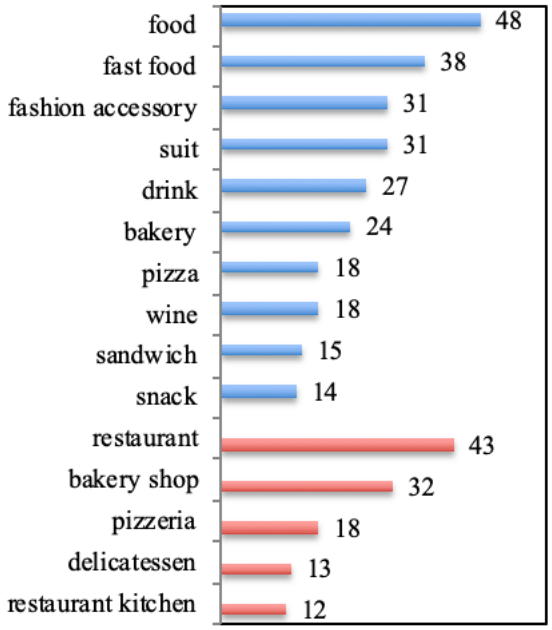


Beyonce

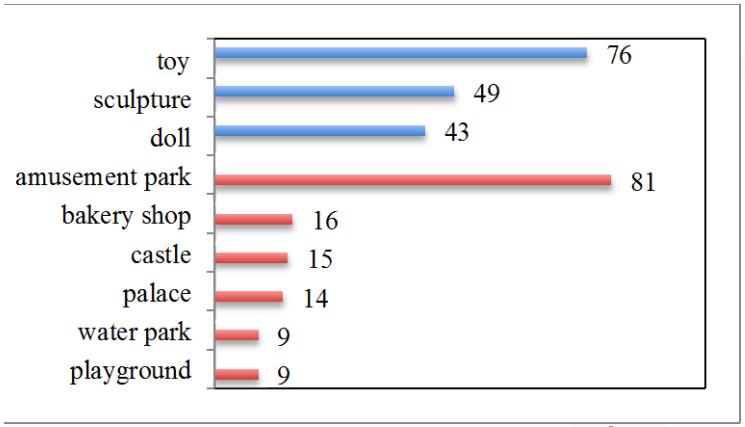


Profiles from Instagram (2)

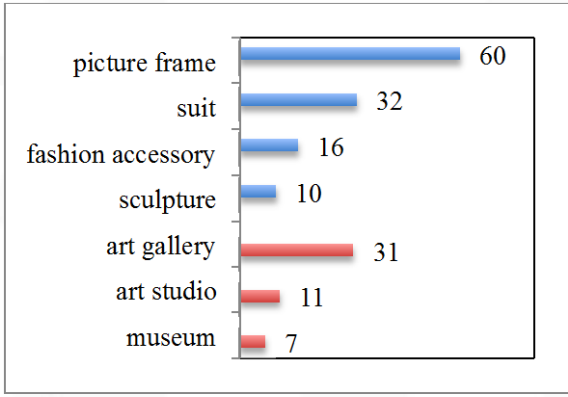
Gordon Ramsay



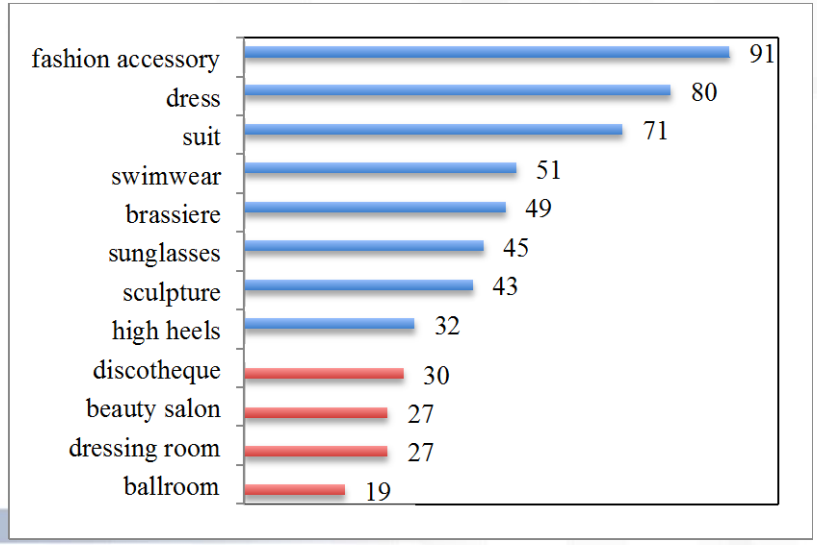
Disneyland



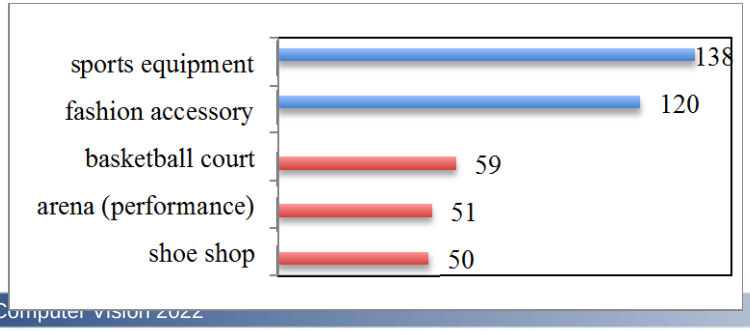
Kehinde Wiley



Kim Kardashian



LeBron James

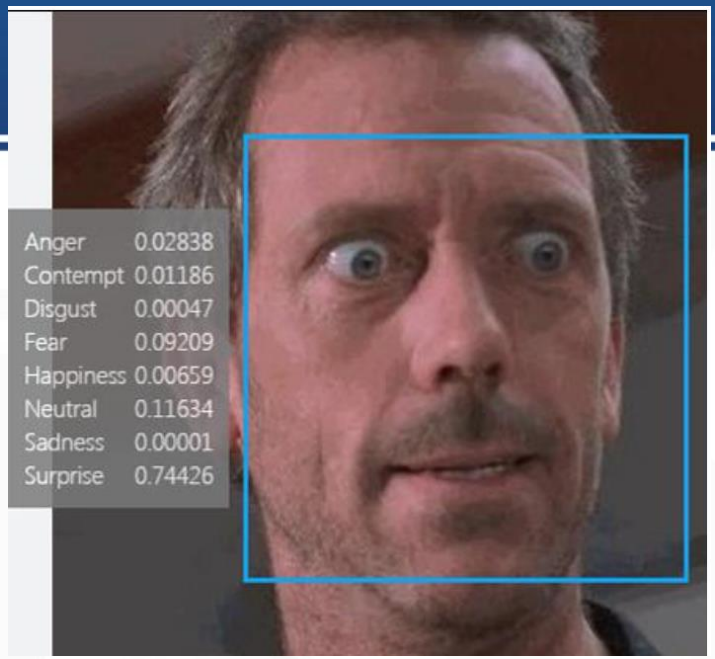


Facial processing and emotion recognition

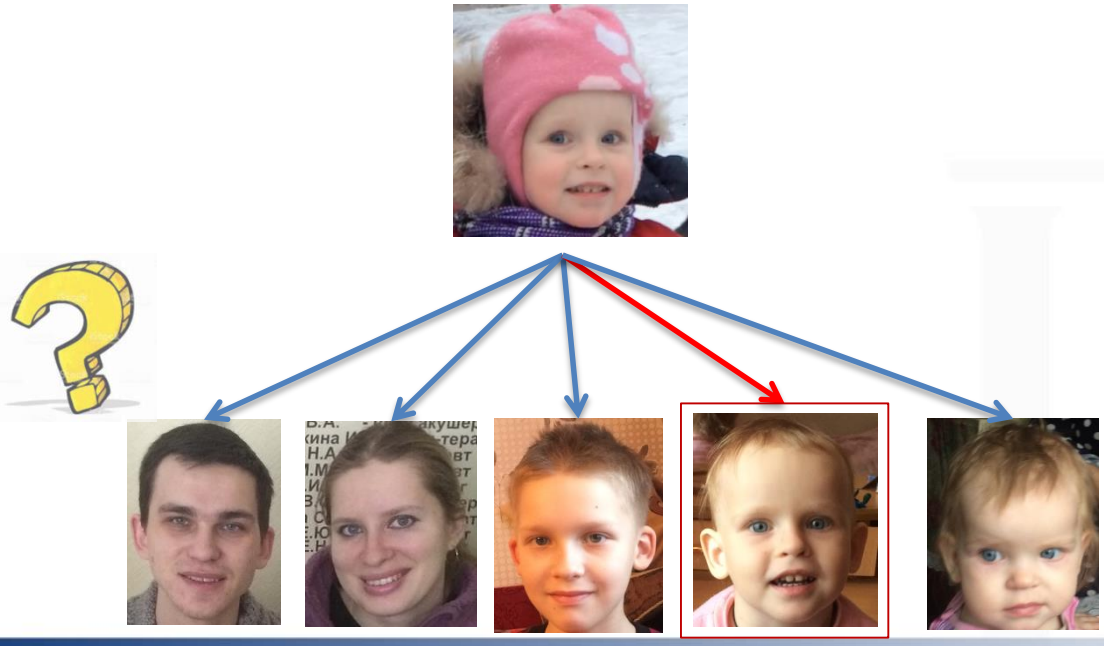
- Kharchevnikova, Savchenko, PeerJ Computer Science 2021
- Sokolova, Savchenko, AIST 2017, 2019, ITNT 2018, OMNN 2020
- Demochina, Savchenko, ICPR Workshop 2021
- Churaev, Savchenko 2021

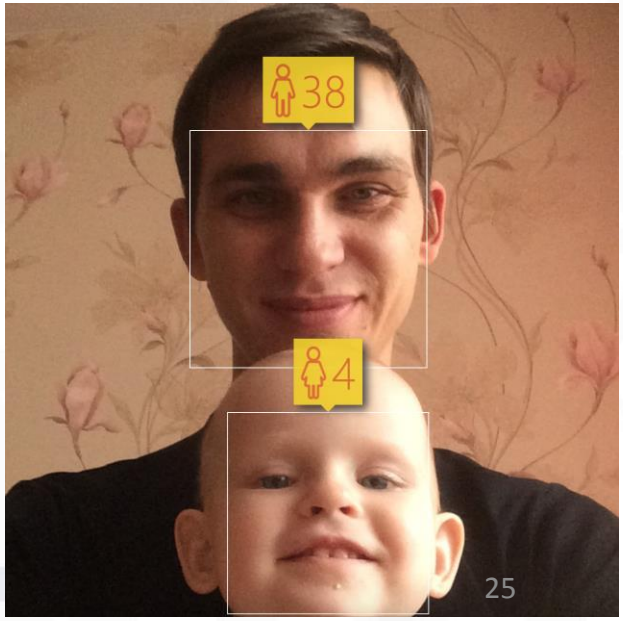
Facial processing tasks

- **Detection**
- **Verification**
- **Identification**
- **Recognition of emotions, age, gender, race,…**

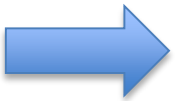


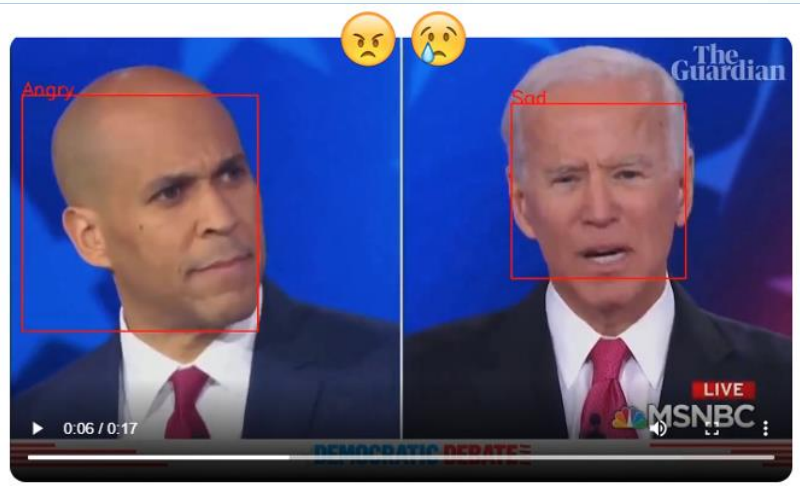
Input photo or video sequence

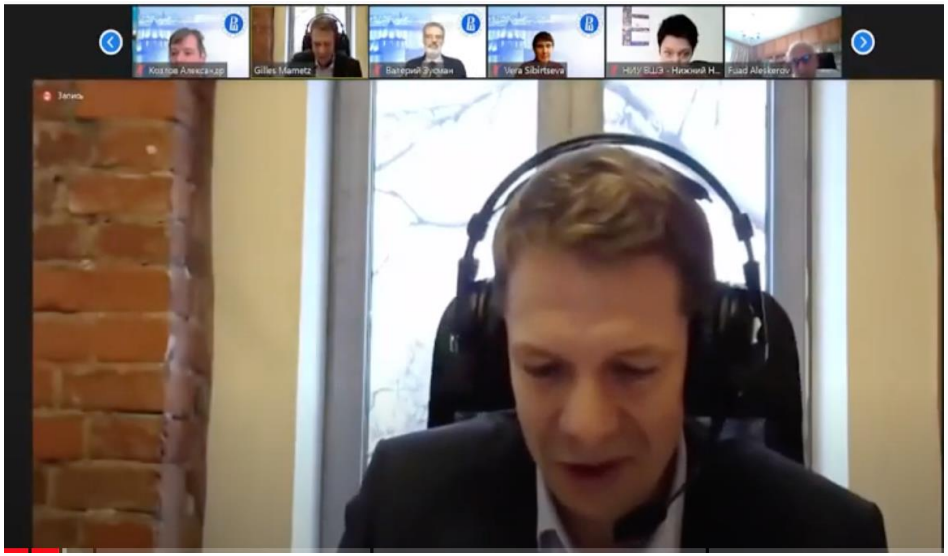




Emotion analysis and engagement prediction in videos







https://www.youtube.com/watch?v=kORwkI-Oq7o

AFEW (Acted Facial Expression In The Wild)




- # of emotions: 7 (Neutral, Happy, Sad, Surprise, Fear, Anger, Disgust)
- Training set: 773 short video clips
- Validation set: 383 videos

https://sites.google.com/view/emotiw2019

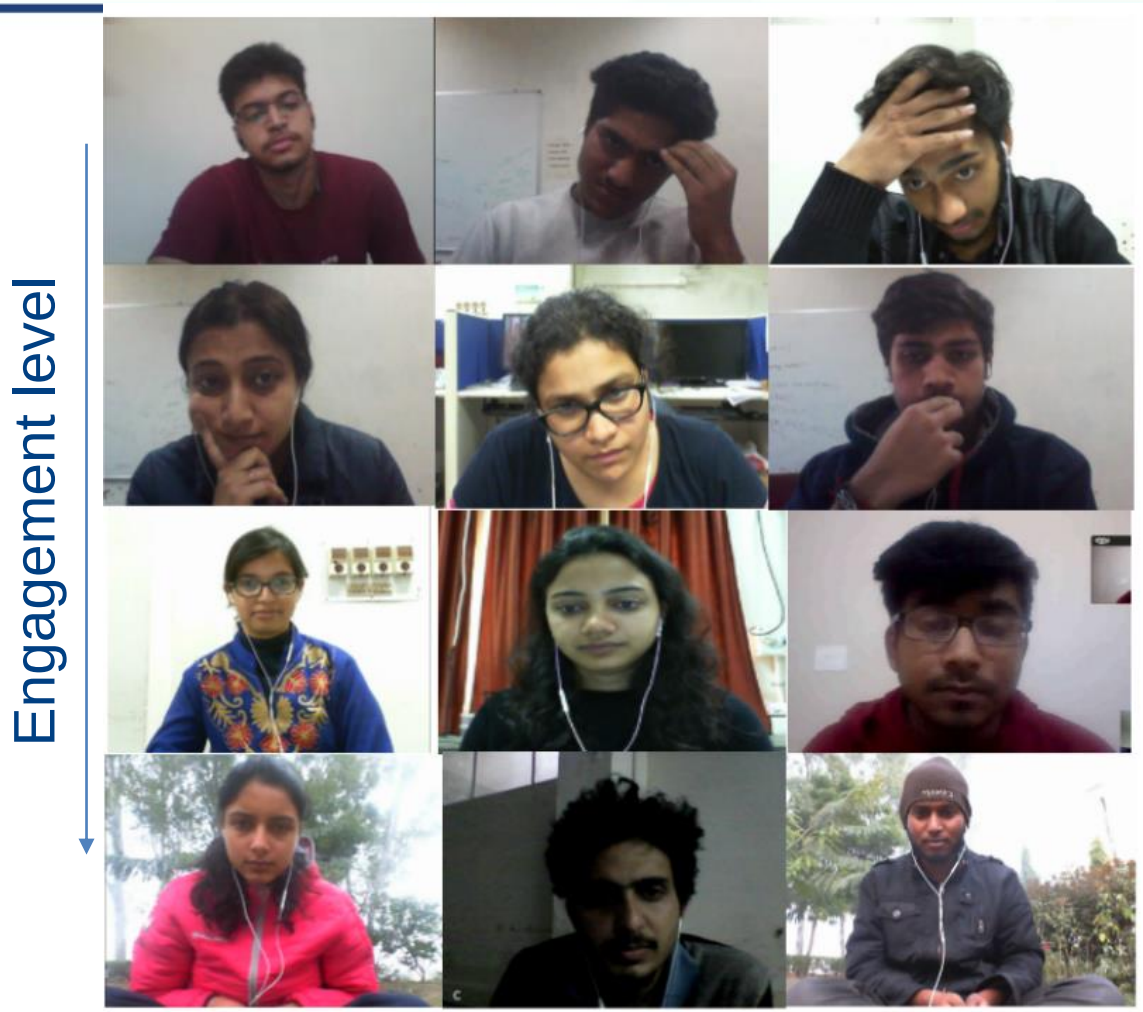
VGAF (Video-level Group AFfect)

- # of classes: 3 (Positive, Negative, Neutral)
- Training set: 2661 videos
- Validation set: 766 videos

https://sites.google.com/view/emotiw2020

EngageWild: student's engagement prediction

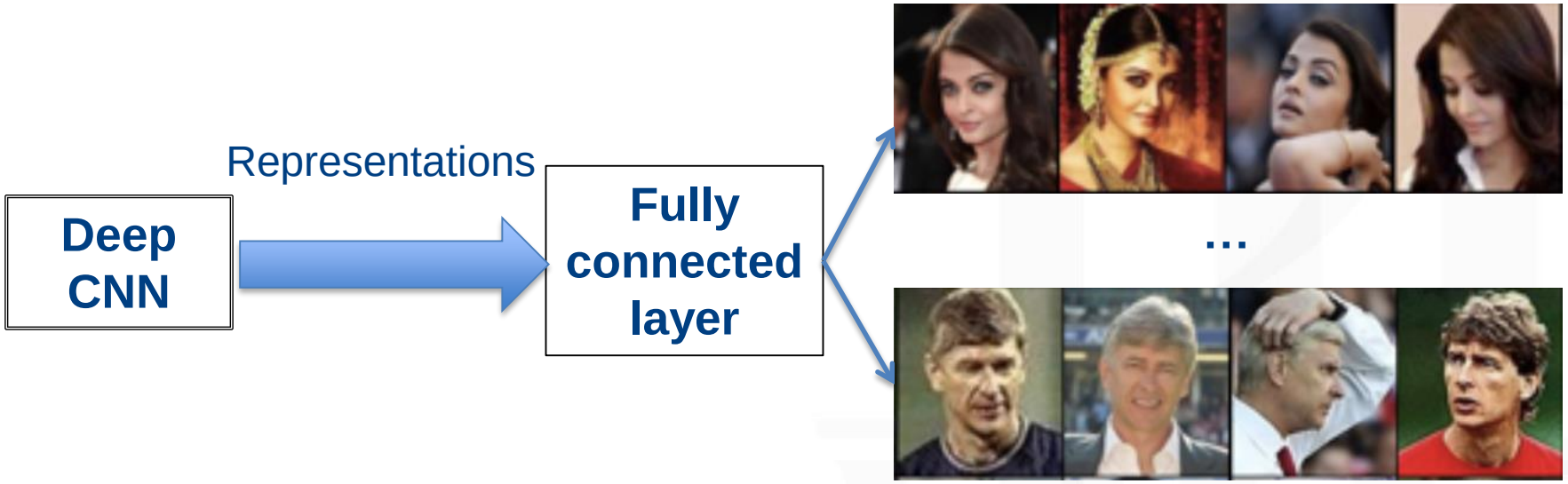
- # of engagement levels: 4 (0 – disengaged, 0.33, 0.66, 1 – highly engaged)
- Training set: 147 videos (~ 5 minutes)
- Validation set: 48 videos

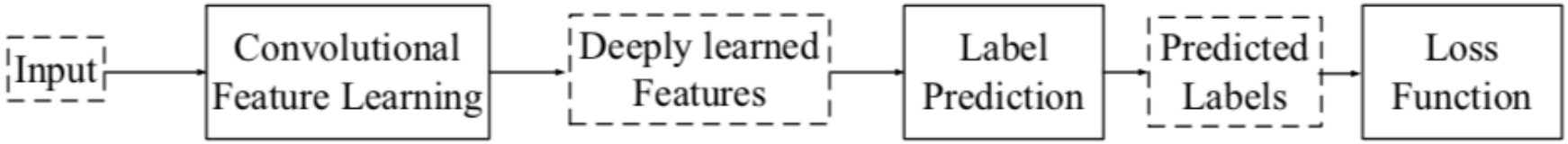


https://sites.google.com/view/emotiw2020

CNN (Convolutional Neural Networks) and domain adaptation: Deep faces

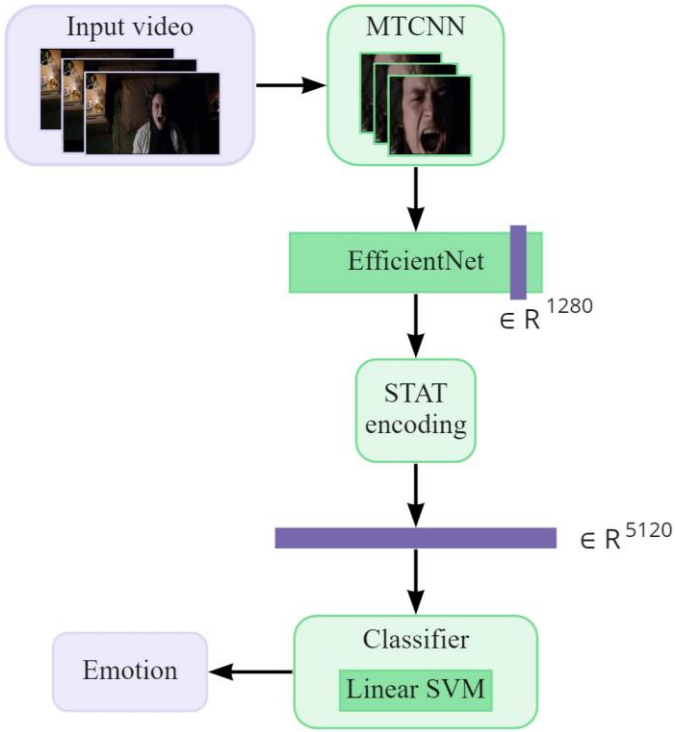
Large training set of celebrities (VGGFace2, CASIA-WebFace, MS-Celeb-1M,…)

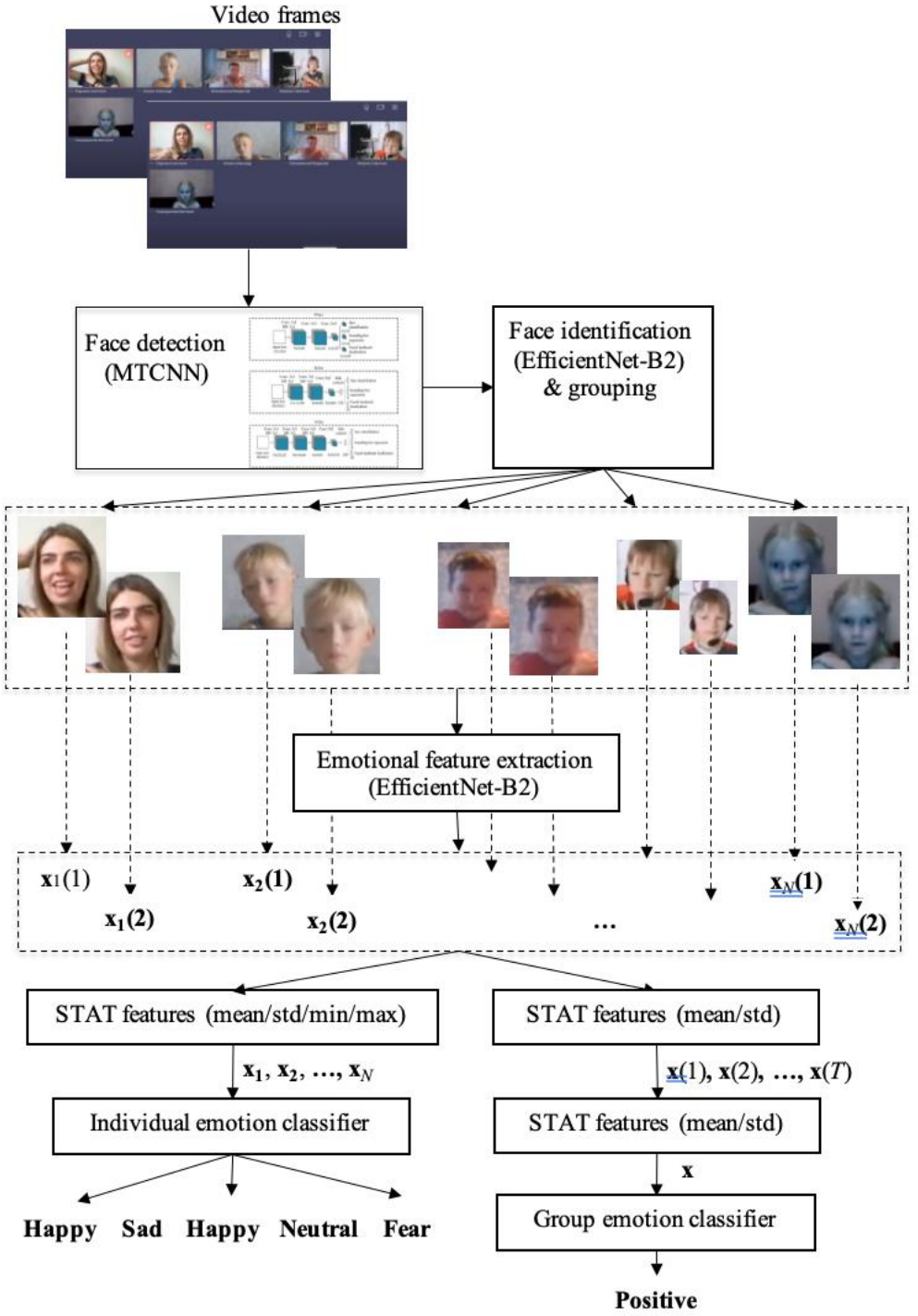




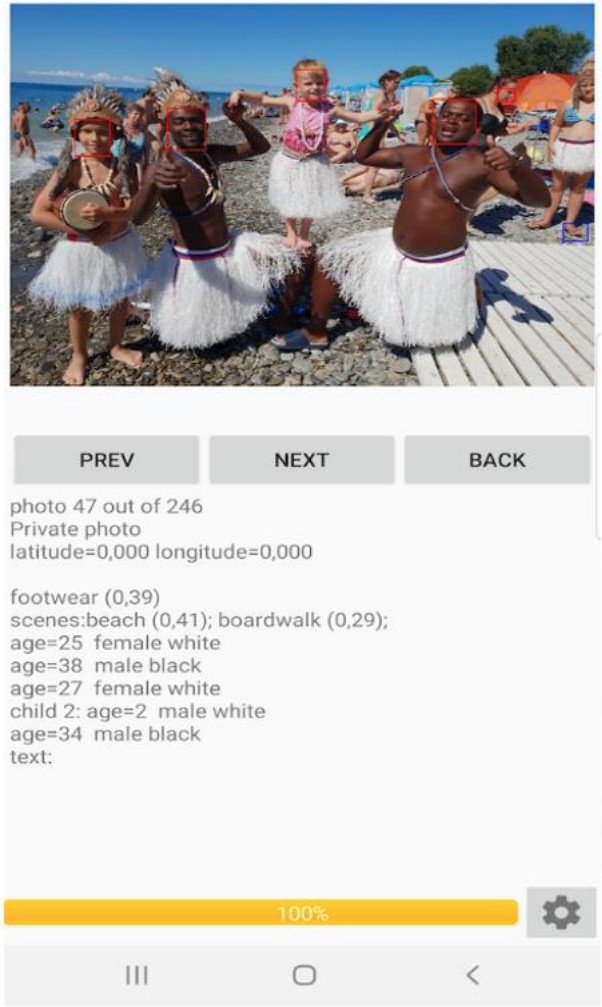
[Wen et al, ECCV 2016]

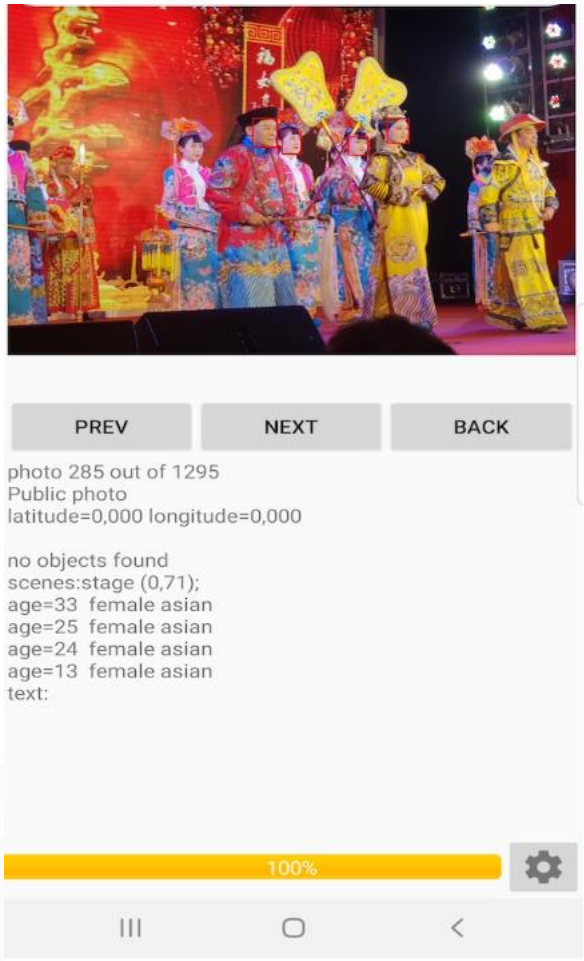
Analysis of video conference





Android demo app

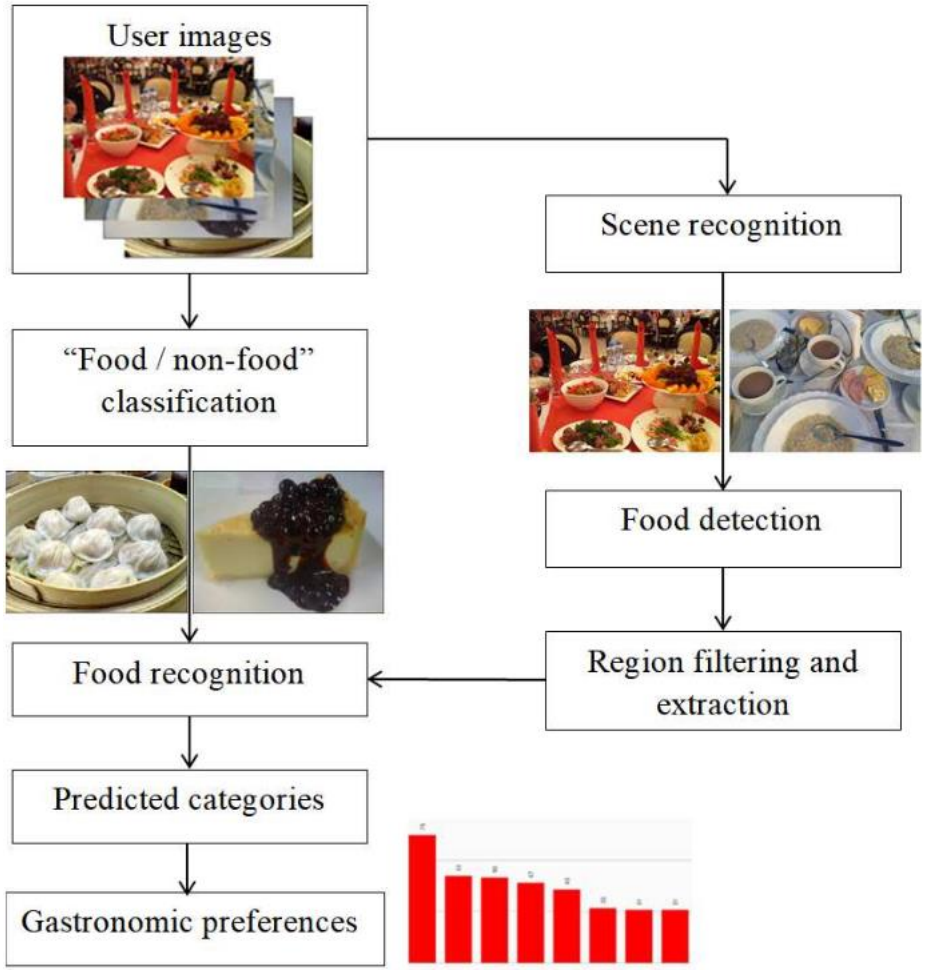




Food classification and restaurant recommendation

- Miasnikov, Savchenko, ICIAR 2020

Food detection and recognition in a gallery of mobile device



Restaurant recommendation

Scenes/Events
dining room (0.31) wedding (-0.51)

Objects
Food:0.768
Fast food:0.473
Fireplace:0.422
Fast food:0.407
Snack:0.398
Table:0.389
Snack:0.313
Snack:0.304

YELP label classification
food (0.86)

YELP photo classification
has_alcohol (0.93)

Select city for recommendation Las Vegas

Analyze

Scenes/Events
restaurant (0.62) birthday (-0.06)

Objects
Food:0.692
Furniture:0.604
Bottle:0.602
Plate:0.518
Chopsticks:0.480
Fast food:0.459
Table:0.412
Plate:0.393
Tableware:0.361

YELP restaurants
has alcohol (0.92)
has table service (0.97)
YELP cuisine
Chinese (0.31)
Japanese (0.22)
YELP labels
food (0.99)

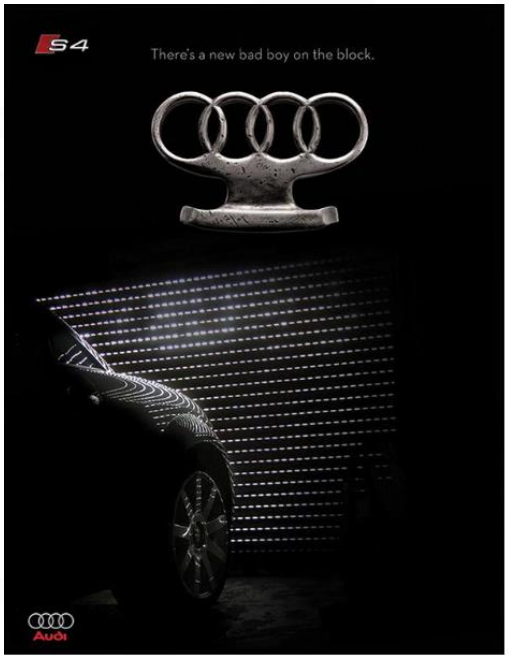
Recommended restaurants

Name	Cuisine	Stars
Red Plate	Chinese	5.0
Bar Sake & Robata Grill	Japanese,Sushi Bars	5.0
Ichi Belle	Japanese,Asian Fusion	5.0
Bar Charlie	Japanese	5.0
Tetsuro's Sayonara, Aloha, Going Away Uye At Japanese Curry n	Japanese	5.0

Understanding advertisements

- Savchenko, Alexeev, Nikolenko et al, COLING 2020

Automatic Understanding of Image Advertisements

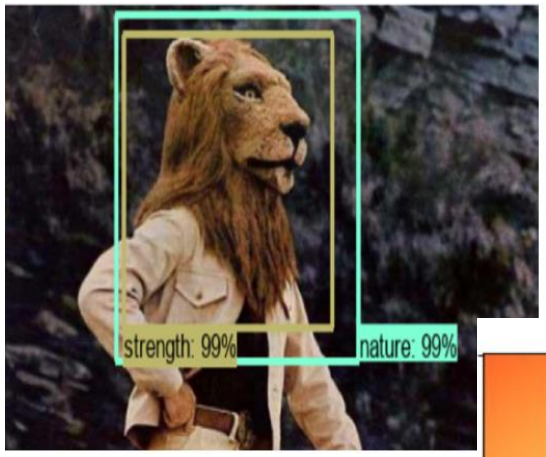


WWF anti-deforestation ad
Topic: environment.
Sentiment: alarmed.
Symbols: environment.

Audi ad: a new bad boy on the block
Topic: cars
Sentiment: inspired
Symbols: n/a.

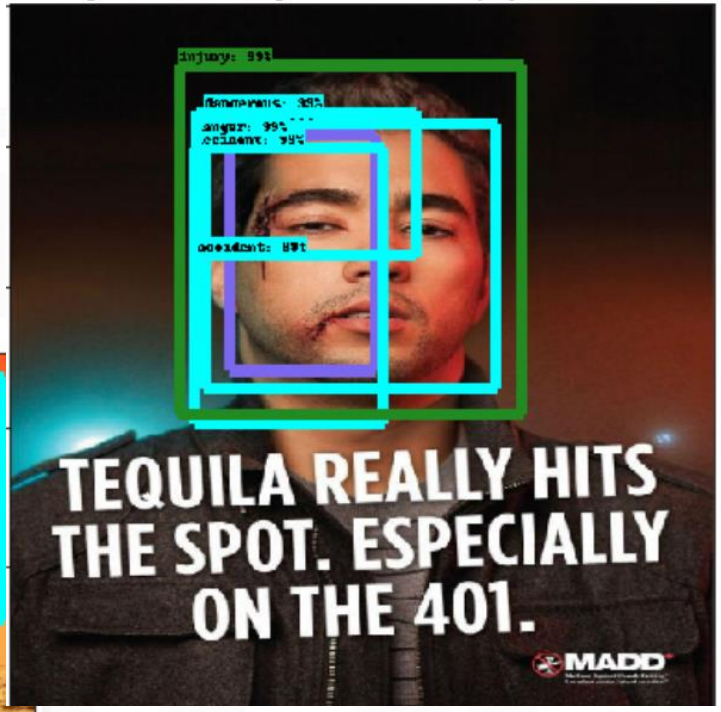
http://people.cs.pitt.edu/~kovashka/ads/

Symbolism in Image Advertisements

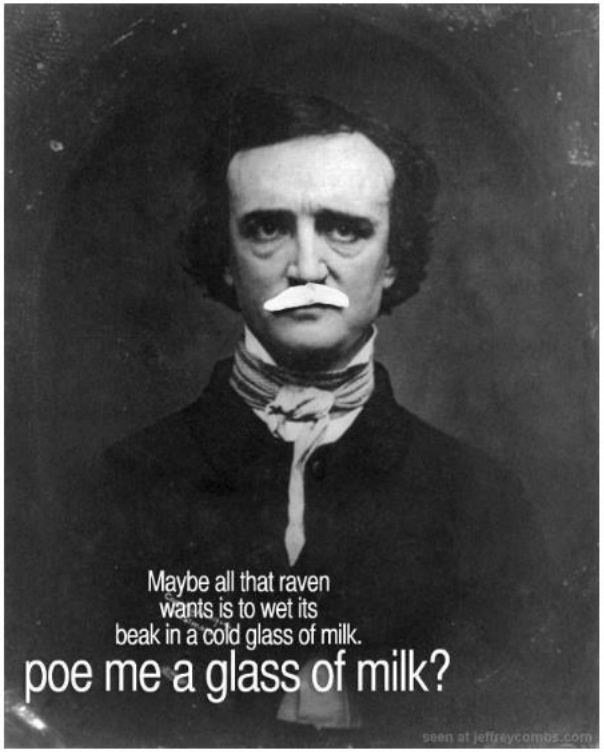


spicy alcohol destruction hot spicy spicy hot spicy

danger accident dangerous accident injury harm accident



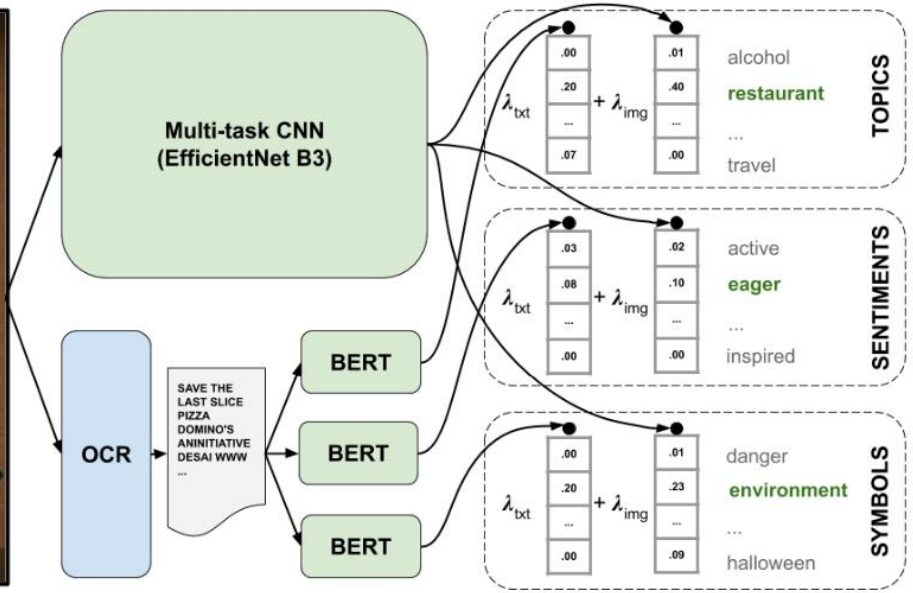
OCR (optical character recognition)

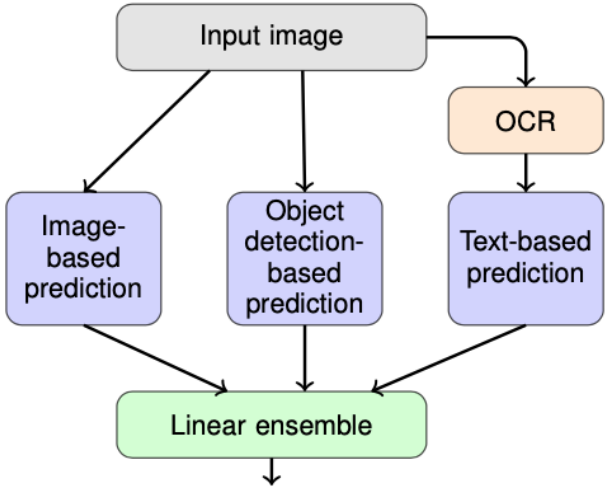


Tesseract Smith (2007)	Maybe all that raven wants is to wet its beak in a'cold glass of milk. poe mea lo EISsTo) aU eg
EAST+*Tesseract* Kopeykina and Savchenko (2019)	Maybe are Gein) eu SH Hostel S to wet its ake cold Me TS 10a milk. eee me glass 0) aa Penal see
PSENet Wang et al. (2019)	Elle that eV WM eMey iy is 10 Wed Mey in beak — Ey milk. of (eek glass — ranlll@a a me al eee) glass be at jeffreycombs-com
EasyOCR JaidedAI (2020)	Maybe all that raven poe me a glass ofmik? beak in a cold glass ofmik. wants is to wet its seen 3t jefieycomlzesolm
Charnet Xing et al. (2019)	MAYBE ALL THAT RAVEN WANTS WET ITS BEAKIN COLD GLASS MILK POE GLASS MILK? SEEN ATJ COM
CloudVision Otani et al. (2018)	Maybe all that raven wants is to wet its beak in a cold glass of milk. poe me a glass of milk? seen at jeffreycombs.com

Pipeline


SAVE THE LAST SLICE
An initiative by
Shefali Desai | www.designeratwork.net


Multi-task CNN
(EfficientNet B3)
OCR
SAVE THE LAST SLICE PIZZA DOMINO'S ANINITIATIVE DESAI WWW ...
BERT
BERT
BERT
λtxt
+ λimg
.00 .20 ... .07
.01 .40 ... .00
alcohol
restaurant
...
travel
TOPICS
.03 .08 ... .00
.02 .10 ... .00
active
eager
...
inspired
SENTIMENTS
.00 .20 ... .00
.01 .23 ... .09
danger
environment
...
halloween
SYMBOLS


Input image
OCR
Image-based prediction
Object detection-based prediction
Text-based prediction
Linear ensemble

NATIONAL RESEARCH
UNIVERSITY

Thank you!