ChaLearn Looking at People and Faces of the World: Face Analysis Workshop and Challenge 2016

http://gesture.chalearn.org/

Sergio Escalera
Mercedes Torres
Brais Martínez
Xavier Baró
Hugo J. Escalante
Isabelle Guyon
Georgios Tzimiropoulos
Ciprian Corneanu
Marc Oliu
Mohammad Ali Bagheri
Michel Valstar

CVPR 2016

Context of the Challenge

ChaLearn LAP Challenges, Workshops, and Special Issues

CVPR 2011 - Workshop and Challenge on Gesture Recognition

CVPR 2012 - Workshop and Challenge on Gesture Recognition

ICPR 2012 - Workshop and Challenge on Gesture Recognition

ICMI 2013 - Workshop and Challenge on Gesture Recognition

ECCV 2014 – Workshop and Challenge on Human Pose, Action and Gesture Recognition

CVPR 2015- ChaLearn Looking at People 2015 - Action spotting and cultural event recognition

ICCV 2015 - ChaLearn Looking at People 2015 – Apparent age and cultural event Recognition

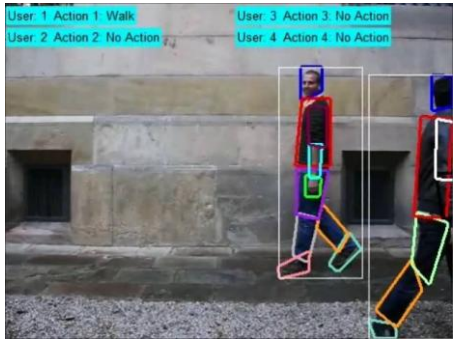
CVPR 2016 – Face analysis challenges and workshop

ECCV 2016 – Apparent personality analysis: first impressions

ICPR 2016 - Apparent personality analysis: first impressions, gesture recognition, context experience

JMLR special issue 2011-2014, TPAMI special issue 2015-2016, IJCV special issue on LAP 2016, TAC

Apparent Personality Analysis – Deadline 15 October 2016



http://gesture.chalearn.org/

Context of the Challenge

CVPR 2016: FOCUS ON FACE ANALYSIS

[Track 1:](#) Apparent Age Estimation

[Track 2:](#) Accessories Classification

[Track 3:](#) Smile and Gender Classification

http://gesture.chalearn.org/

Competitions Schedule

January 25, 2016: Beginning of competition for track 1, release of development and validation data.

February 10, 2016: Beginning of competition tracks 2 and 3, release of development and validation data.

March 1, 2016: Release of encrypted final evaluation data and validation labels. Participants can start training their methods with the whole data set.

March 8, 2016: Release of final evaluation data decryption key. Participants start predicting the results on the final evaluation data.

March 15, 2016: End of the quantitative competition for track 1. Deadline for submitting the predictions over the final evaluation data. Deadline for code submission. The organizers start the code verification by running it on the final evaluation data.

March 18, 2016: Deadline for submitting the fact sheets.

March 22, 2016: Release of the verification results to the participants for review.

March 30, 2016: End of the quantitative competition for tracks 2 and 3. Deadline for submitting the predictions over the final evaluation data. Deadline for code submission. The organizers start the code verification by running it on the final evaluation data. Winners of each track will be announced during the workshop.

Codalab

The CodaLab Team

Percy Liang is an assistant professor of Computer Science at Stanford University. His primary research areas are machine learning and natural language processing. He leads the development of CodaLab in close collaboration with Microsoft Research and the rest of the community.

Isabelle Guyon is an independent consultant, specializing in statistical data analysis, pattern recognition and machine learning. Isabelle served as an advisor in the development of the CodaLab competition platform and pioneered the implementation of several challenges on Codalab.

Sergio Escalera leads the Human Analysis group (HuPBA) at the University of Barcelona and the Computer Vision Center. He is one of the directors of ChaLearn. He pioneered the implementation of several Computer Vision challenges on Codalab within the field of Looking at People.

Evelyne Viegas is a Director at Microsoft Research responsible for the outreach artificial intelligence program. She leads the CodaLab project working in collaboration with Isabelle Guyon, Percy Liang and the machine learning and artificial intelligence communities.

https://competitions.codalab.org/

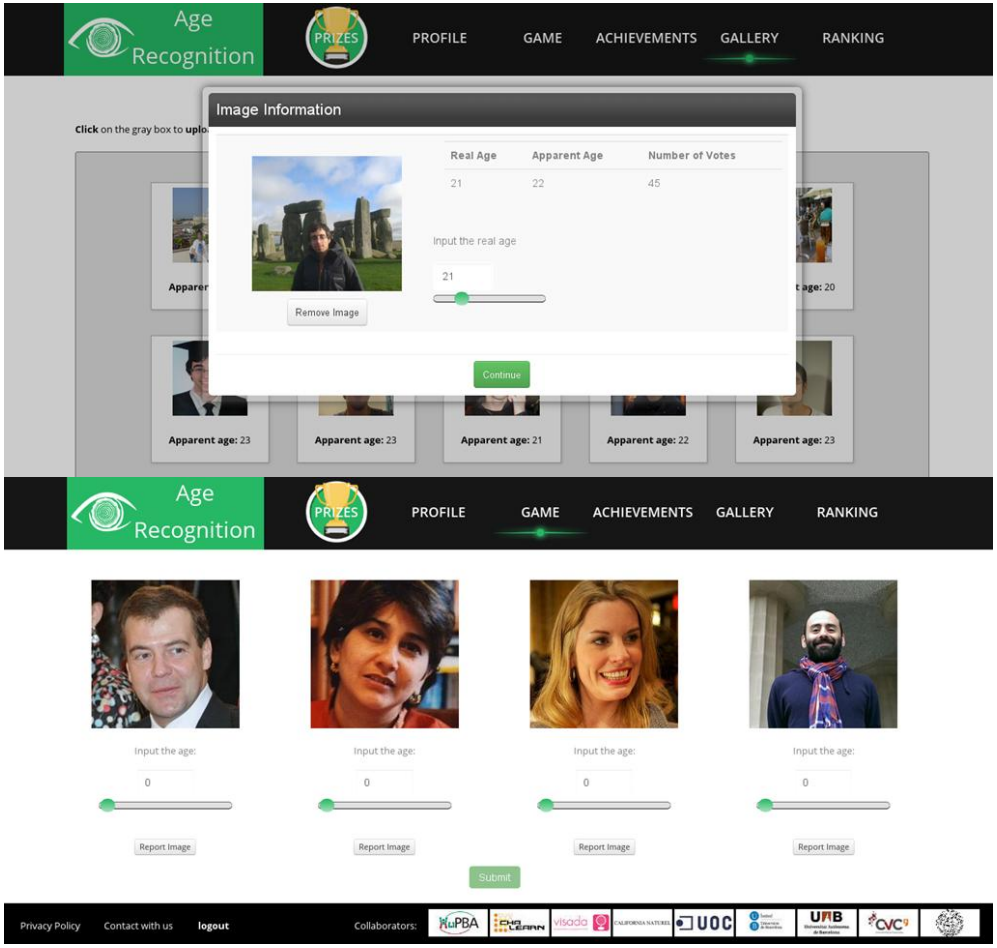
https://github.com/codalab/

http://gesture.chalearn.org/

Apparent Age Estimation Track

Age estimation is a difficult task which requires the automatic detection and interpretation of facial features. We have designed an application using the Facebook API for the collaborative sharing of images and labeling by the community in a gamified fashion.

(Also complemented via AMT voting using public images)



Dataset and evaluation metric

The dataset consists of 7,591 images and nearly 300,000 votes (almost double size in comparison with 2015 first round competition)

- **Train Set**: 4,113 images
- **Validation Set**: 1,500 images
- **Test Set**: 1,978 images

The images were labeled by hundreds of people with the apparent age, so for each image we have the mean and standard deviation of its apparent age.

The evaluation metric was the following:

$$\epsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where x is the prediction and μ and σ are the mean and std. of the apparent age.

Challenge results (track 1)

- Nearly 100 participants registered for the competition.
- Results on final evaluation data:

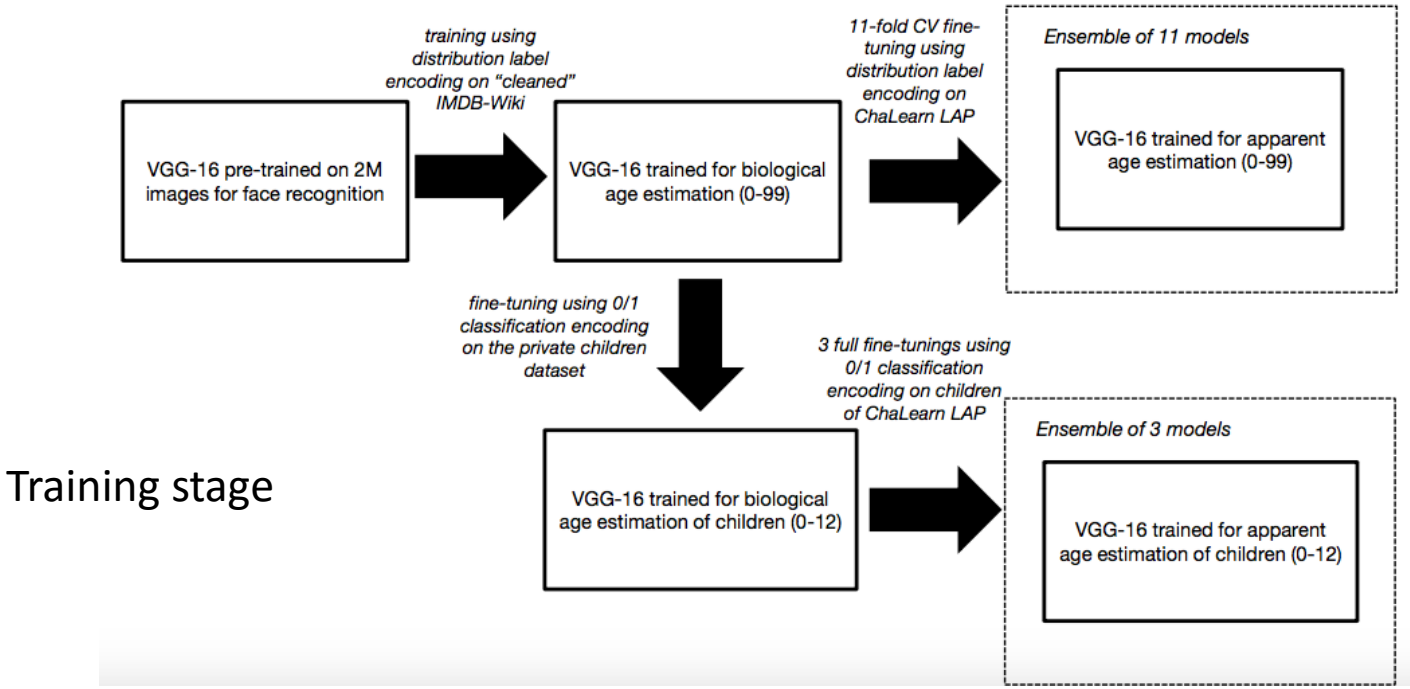
Position	Team	Test error
1	**OrangeLabs**	**0.2411**
2	palm_seu	0.3214
3	cmp+ETH	0.3361
4	WYU_CVL	0.3405
5	ITU_SiMiT	0.3668
6	Bogazici	0.3740
7	MIPAL_SNU	0.4569
8	DeepAge	0.4573

Challenge results (track 1)

Rank	Team	Test Error	General Idea	Pre-trained models	Preprocessing	Fusion	Additional data used
1	OrangeLabs	0.2411	Two-phase learning by an ensemble of several CNN models	VGG-16	face detection / pose estimation / alignment	Score-level fusion	IMDB Wiki Children images from web
2	palm_seu	0.3214	An ensemble of four fine-tuned CNN models	VGG-16	face detection pose estimation alignment	Score-level fusion	IMDB-WIKI
3	cmp+ETH	0.3361	Ensemble of 8 SO-SVM classifiers learned on the features from the last layer of VGG-16 network	VGG-16	face detection	Score-level fusion	IMDB-WIKI
4	WYU_CVL	0.3405	Multiple models using grouped deep age networks and random forest regressor	None	face detection / image augmentation	Score-level fusion	WebFace Morph, CACD, FG-Net
5	ITU_SiMiT	0.3668	An ensemble of 3 CNN models originated from VGG-16 and fine-tuned on the challenge data	VGG-16	face detection / face cropping	Score-level fusion	IMDB-WIKI
6	Bogazici	0.374	A two part model: classification into overlapping age groups and regression among each group	VGG-16	face detection, intensity averaging	Score-level fusion	None
7	MIPAL_SNU	0.4569	An ensemble of 3 CNNs with different loss functions	ImageNet Pre-trained CNN models with ILSVRC and CACD data	face detection / face cropping	Score-level fusion	ILSVRC 2015, CACD
8	DeepAge	0.4573	Deep Label Distribution Learning (DLDL) framework	VGG-16	face detection / face cropping	None	1) 53,969 web face images 2) LAP Age Estimation 2015

Methods Used by the Winners – 1st place

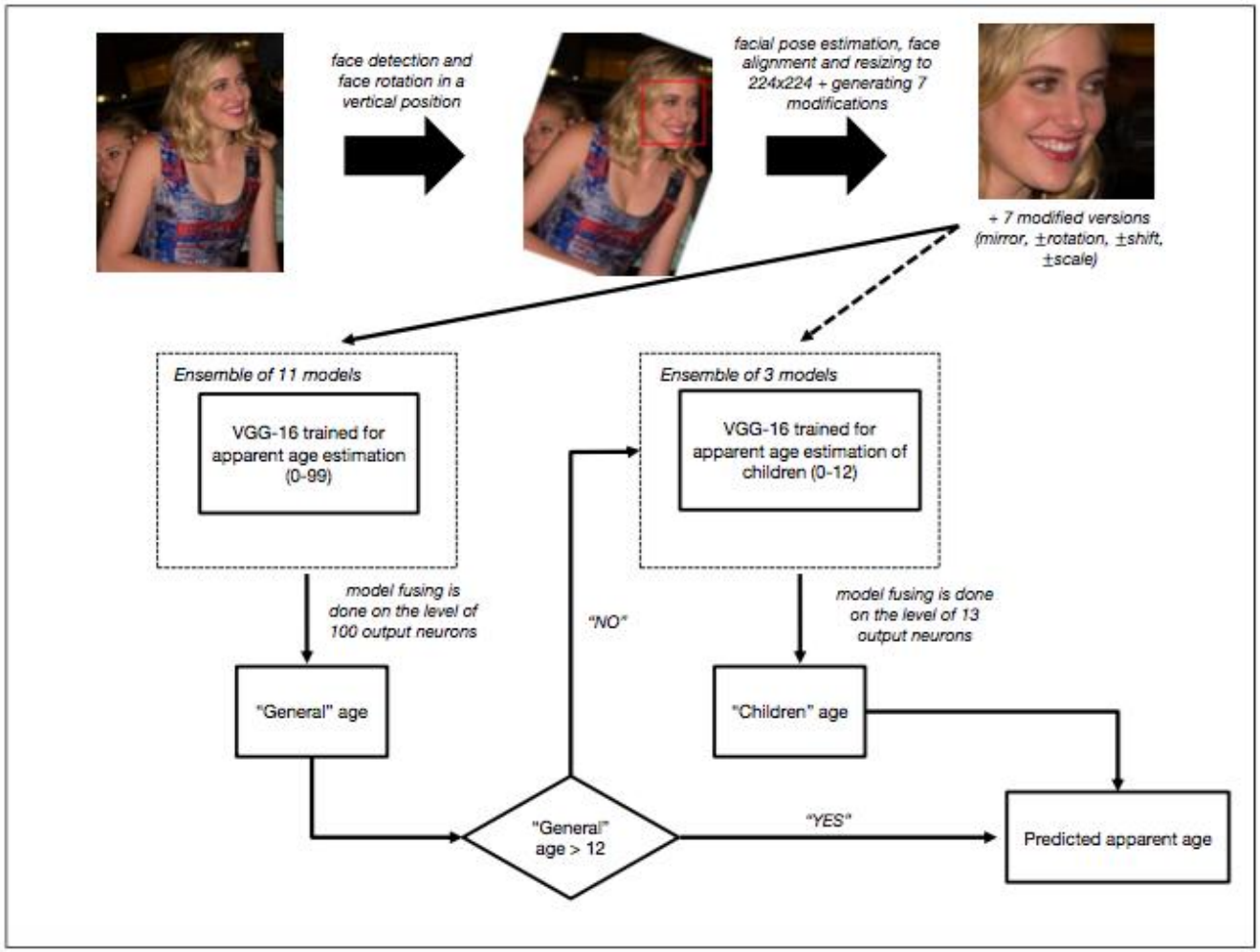
- The winners of the apparent age competition (Orange labs team) used the VGG-16 Convolutional Neural Network (CNN) trained for face recognition on 2 million faces [33] as the starting point. An ensemble of 11 models is used for general age estimation. An additional model for children age estimation was proposed.



Training stage

Apparent age

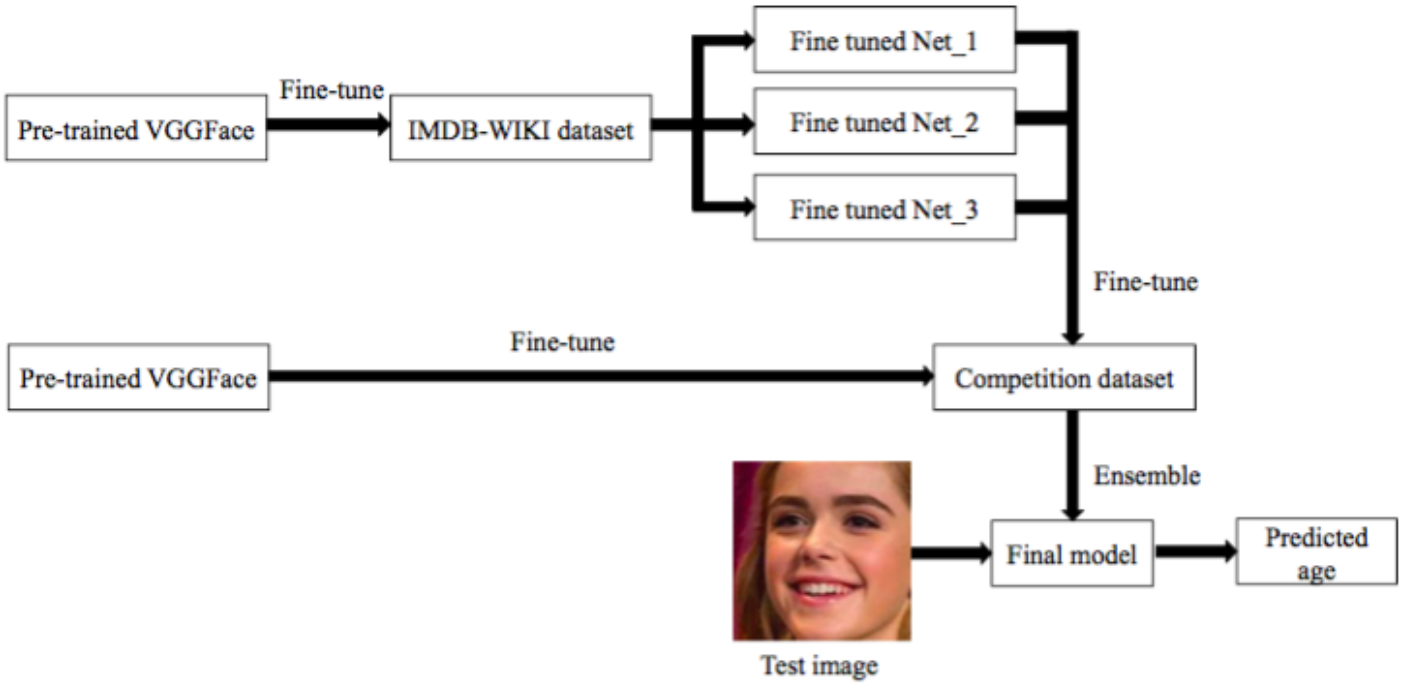
Methods Used by the Winners – 1st place



Test stage

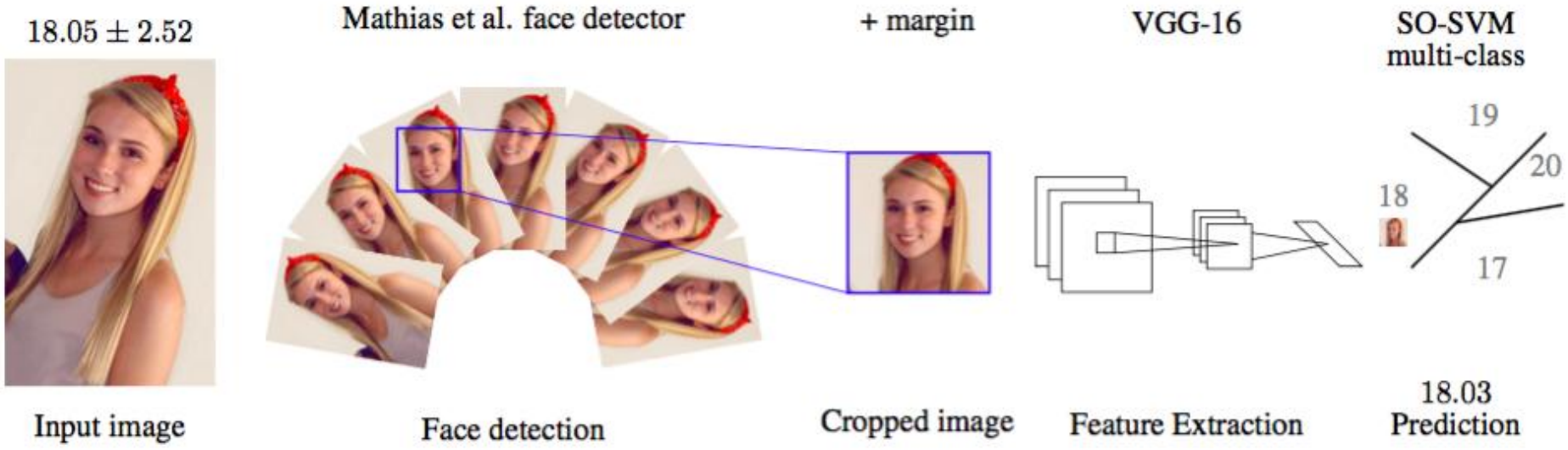
Methods Used by the Winners – 2nd place

- The second place of the apparent age competition was obtained by Palm_seu team. The method is a CNN based on VGG-16, but trained to learn the probability distribution of age (mean and standard deviation)



Methods Used by the Winners – 3rd place

- The third place of this track (ETH, first round winners) adopted a similar approach as round 1, except that the predictive model was a structured output SVM



Faces of the World Dataset

- Addresses the problem of over-simplified datasets
- 25,000 images collected from flickr
- Balanced data:
  - 25% of each of the Asian, Black, Hispanic, and White demographic groups
  - 50% male/female
  - Attention was paid to achieve a near-uniform age distribution
- Images were annotated using the Zooniverse Citizen Science crowd-sourcing platform
  - > 500 people annotated data
  - > 28,000 annotations collected

Faces of the World Dataset

- Annotations undergo a quality-control process:
  - removed images with ambiguous annotations, e.g. faces with very little visibility/ large occlusions
  - Annotations of remaining images were verified visually
- Dataset was split in three: Training, Validation, and Test
- Participants are free to report any study on the Validation set
- But have to submit their code/program for the organisers to return their scores – Test set remains held back

Examples of the Faces of the World Dataset

Pre-processing

- As it would be too difficult for people to also do face detection in such challenging conditions, images were cropped around the target face

Tasks on the FOTW dataset

- Track 2: Accessory Classification
  - Goal is to detect the presence/absence of the following attributes:

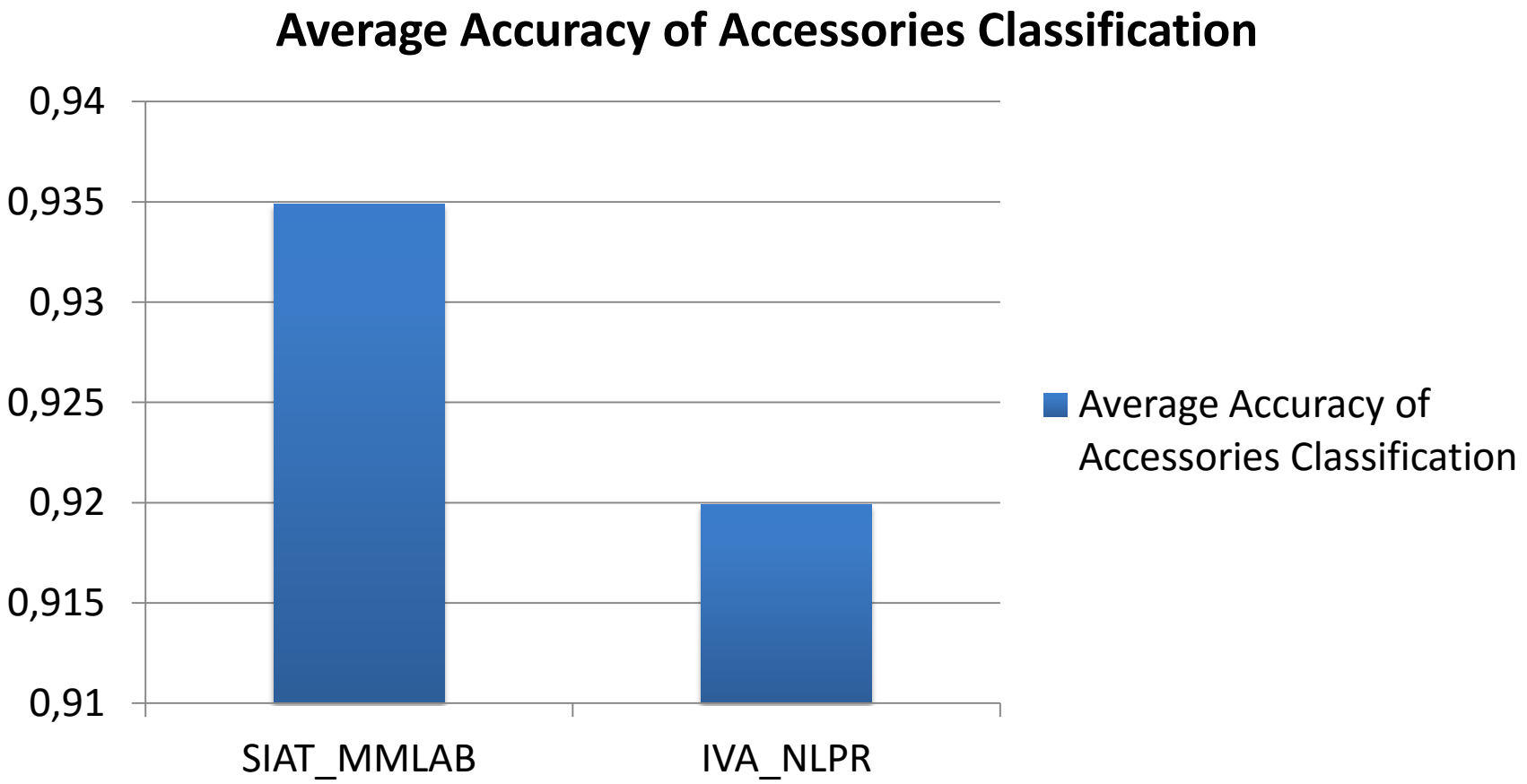
Accessory	Train	Validation	Test
Hat	1151	608	869
Headband	243	109	193
Glasses	1232	614	828
Earrings	770	389	592
Necklace	615	300	559
Tie	151	72	220
Scarf	256	137	256

Tasks on the FOTW dataset

- Track 3: Smile and Gender Classification
  - Goal is to determine whether someone is smiling, and what their gender is

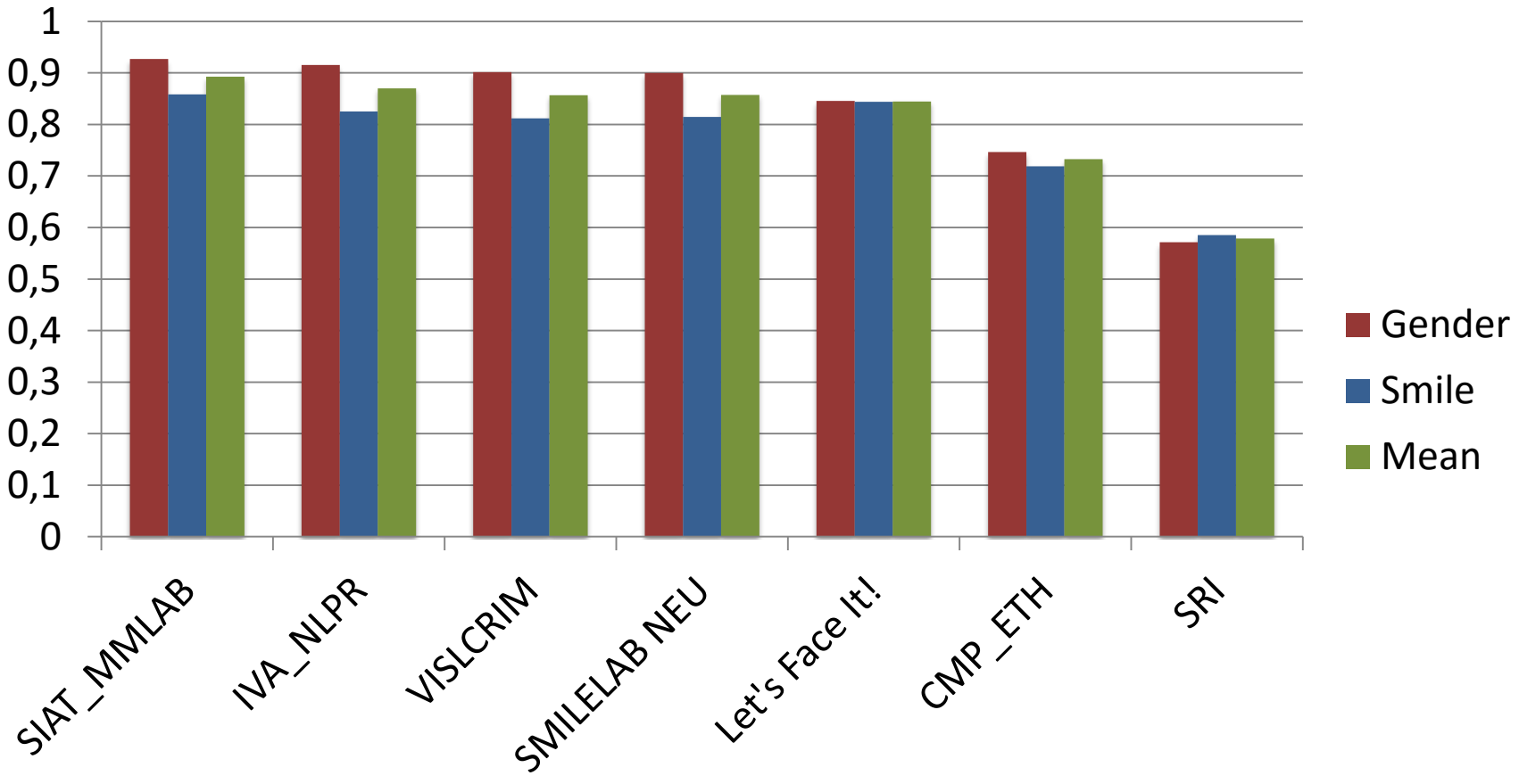
Attribute	Train	Validation	Test
Male	2946	1691	4614
Female	3318	1361	3799
Smile	2234	1969	4411
No Smile	3937	1117	3849

Results of Track 2 – Accessories Classification



Smile and Gender Classification

Smile and Gender Classification Track

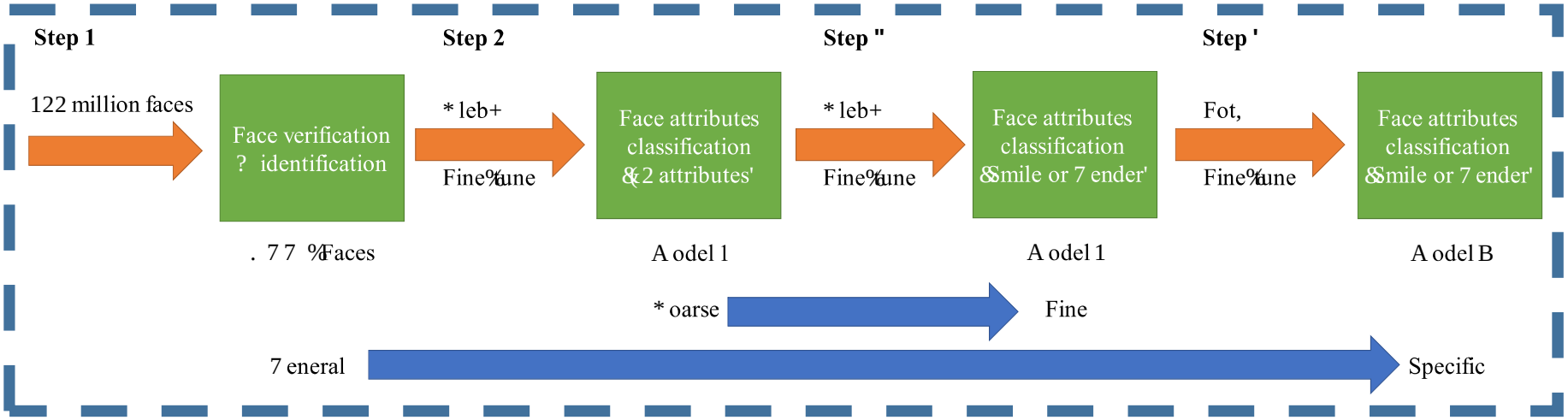


Accessories Classification

Methods Used by the Winners

Kaipeng Zhang, Liangzhi Tan, Zhifeng Li, Yu Qiao, *'Gender and Smile Classification using Deep Convolutional Neural Networks'*

- Multi-task approach (identification, attributes, smile/gender)
- General to specific fine-tuning scheme



ChaLearn Looking at People and Faces of the World: Face Analysis Workshop and Challenge 2016

http://gesture.chalearn.org/

Thank you!

CVPR 2016

Challenge winners

Apparent age

Position	Team	Test error
1	**OrangeLabs**	**0.2411**
2	palm_seu	0.3214
3	cmp+ETH	0.3361
4	WYU_CVL	0.3405
5	ITU_SiMiT	0.3668
6	Bogazici	0.3740
7	MIPAL_SNU	0.4569
8	DeepAge	0.4573

Accessories Classification

Position	Team	Accuracy
1	**SIAT-MMLAB**	**0.9349**
2	IVA-NLPR	0.9199

Smile and gender classification

Position	Team	Accuracy
1	**SIAT-MMLAB**	**0.8926**
2	IVA-NLPR	0.8702
3	VISLCRIM	0.8568
4	SMILELAB NEU	0.8574
5	Let's Face It!	0.8447
6	CMP_ETH	0.7327
7	SRI	0.5785

Best workshop paper award

Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, Jean-Luc Dugelay, *Apparent Age Estimation from Face Images Combining General and Children-Specialized Deep Learning Models.*