ChaLearn Looking at People: Face Analysis Challenges

Joint Challenge on Dominant and Complementary Emotion Recognition Using Micro Emotion Features and Head-Pose Estimation – DCER&HPE. Washington 2017

Xavier Baró – Universitat Oberta de Catalunya

Sergio Escalera – Universitat de Barcelona

Hugo J. Escalante – INAOE

Isabelle Guyon - University Paris-Saclay and Chalearn

Outline

1. Introduction
2. Designing a Challenge
3. Face Analysis Challenges
4. Conclusions and future challenge

Outline

1. **Introduction**
2. Designing a Challenge
3. Face Analysis Challenges
4. Conclusions and future challenge

Introduction

- ChaLearn is a non-profit organization focusing on challenges organization in Machine Learning
- Organize challenges to stimulate research in machine learning.
- Past challenges remain open:

http://chalearn.org

Introduction

- ChaLearn is organized in different work groups:
  - **CiML:** Fundamentals in challenge organization.
  - **AutoML:** Automatic machine learning.
  - **LAP:** Looking at people challenges.
  - **Causality and feature selection:** Forecasting, spatiotemporal analysis and causality

Introduction

- ChaLearn LAP

Isabelle Guyon
University Paris-Saclay

Xavier Baró
Universitat Oberta de Catalunya

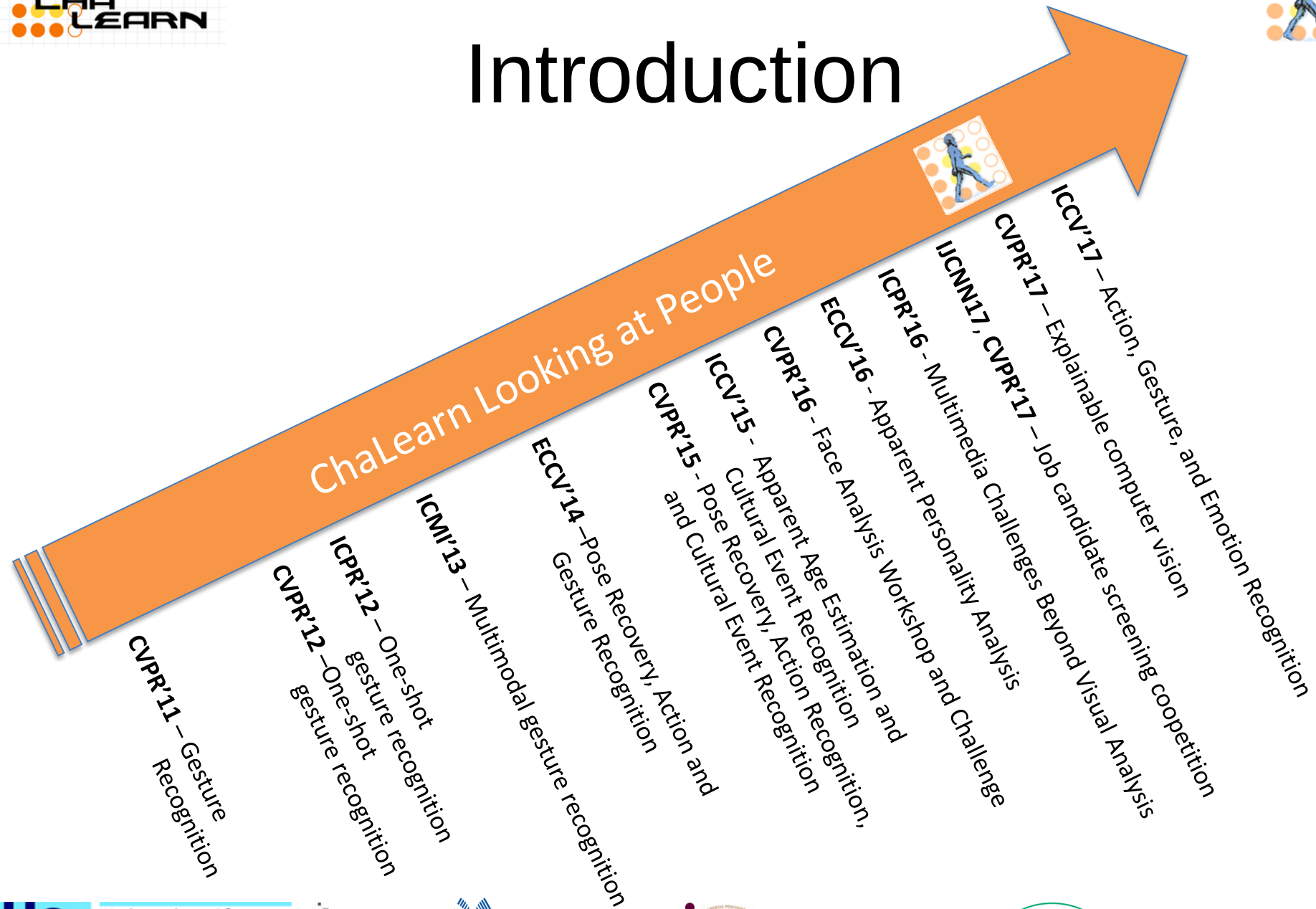
Hugo Jair Escalante
INAOE

Sergio Escalera
Universitat de Barcelona

Introduction

- ChaLearn LAP
  - The goal is to promote the advance on human-centric challenges
    - Human body analysis (pose, gesture)
    - Face analysis (age, accessories, emotions)
    - Context analysis (cultural events)
    - Multimodality (RGB, depth, audio)

Introduction



Introduction

Associated special issues:

- Journal of Machine Learning Research: SI-Gesture recognition 2012-2014
- International Journal of Computer Vision: SI-Looking at People 2014-2016
- IEEE Trans. on PAMI: SI-Multi-modal Human Pose Recovery and Behavior 2016-2017
- IEEE Trans. on Affective Computing: SI-Personality Analysis 2016-2017
- IEEE Trans. on PAMI: SI-The Computational Face 2017

Introduction

Sponsors:

Outline

1. Introduction
2. **Designing a Challenge**
3. Face Analysis Challenges
4. Conclusions and future challenge

Designing a Challenge

1) The Problem.

- Easy to understand
- Interesting from a research point of view
- Feasible in time and resources, but not trivial
- Active community on the field

Designing a Challenge

2) The Data.

- Think on which data you need
  - Modalities
  - Labels
  - Quality

- Is there any already existing data that you can use?
  - Check copyrights
  - Check available ground-truth information
  - Format of the data
- Otherwise, you will need to create or complement the data.
  - Collect data
    - Copyright
    - Data privacy, ethical and legal issues
  - Label the data
    - Undergraduate and PhD students can help, but usually it is not enough
    - Crowd-sourcing platforms (Amazon Mechanical Turk, custom, …)

Designing a Challenge

3) The Dissemination.

- Attract participants to your challenge
  - Distribution lists
  - Events in conferences
- Share the results with the community
  - Workshops
  - Special Issues in Journals

Designing a Challenge

4) The Competition.

- The sponsors
  - Prices for the competition (money, devices, …)
  - Travel grants to associated events
  - Costs of the events itself
- The platform
  - Many options, but not all will adapt to your needs
- The evaluation metric
  - Should be clear and fair
  - As standard as possible
- The results
  - Leaderboard
  - Baselines
  - Phases and tracks
- The schedule
- Communication channels

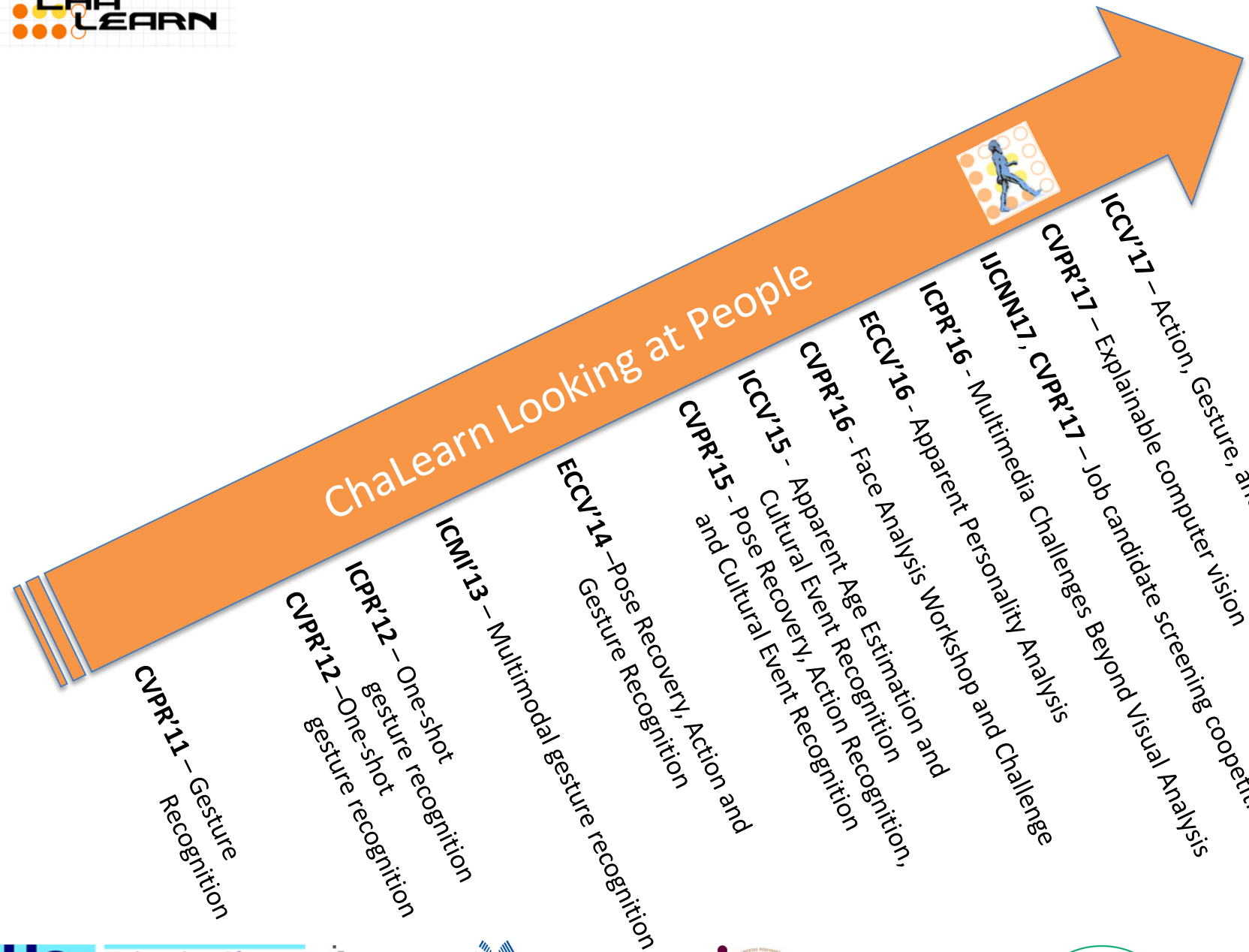
Designing a Challenge

As a conclusion:

- Have a good team
- The most difficult task is to have the data
- A good dissemination strategy can attract more participants than the prices
- Take into account the context

Outline

1. Introduction
2. Designing a Challenge
3. **Face Analysis Challenges**
4. Conclusions and future challenge

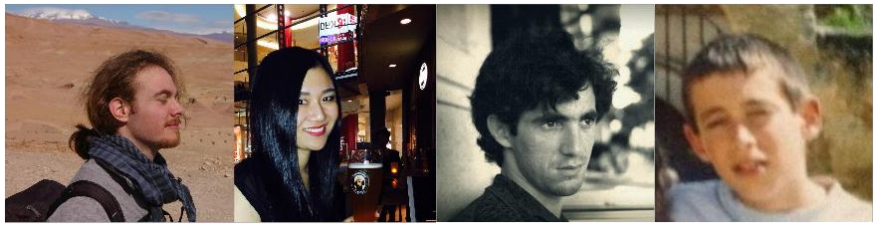

ChaLearn Looking at People
CVPR'11 – Gesture Recognition
CVPR'12 –One-shot gesture recognition
ICPR'12 – One-shot gesture recognition
ICMI'13 – Multimodal gesture recognition
ECCV'14 –Pose Recovery, Action and Gesture Recognition
CVPR'15 - Pose Recovery, Action and and Cultural Event Recognition
ICCV'15 - Apparent Age Estimation and Cultural Event Recognition,
CVPR'16 - Face Analysis Workshop and Challenge
ECCV'16 - Apparent Personality Analysis
ICPR'16 - Multimedia Challenges Beyond Visual Analysis
IJCNN17, CVPR'17 – Job candidate screening coopetition
CVPR'17 – Explainable computer vision
ICCV'17 – Action, Gesture, and Emotion Recognition


ChaLearn Looking at People
ICCV'15 - Apparent Age Estimation and Cultural Event Recognition
CVPR'16 - Face Analysis Workshop and Challenge
ECCV'16 - Apparent Personality Analysis
IJCNN17, CVPR'17 – Job candidate screening coopetition
ICCV'17 – Action, Gesture, and Emotion Recognition

Face Analysis Challenges

Apparent Age Estimation

Automatic apparent age estimation from RGB face still images. The Age Estimation challenge aims to investigate the performance of estimation methods on apparent age rather than real 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.

Face Analysis Challenges

Apparent Age Estimation

- The Data (ICCV 2015)

The dataset consisted of 4691 face images

- **Train Set**: 2476 images
- **Validation Set**: 1136 images
- **Test Set**: 1079 images

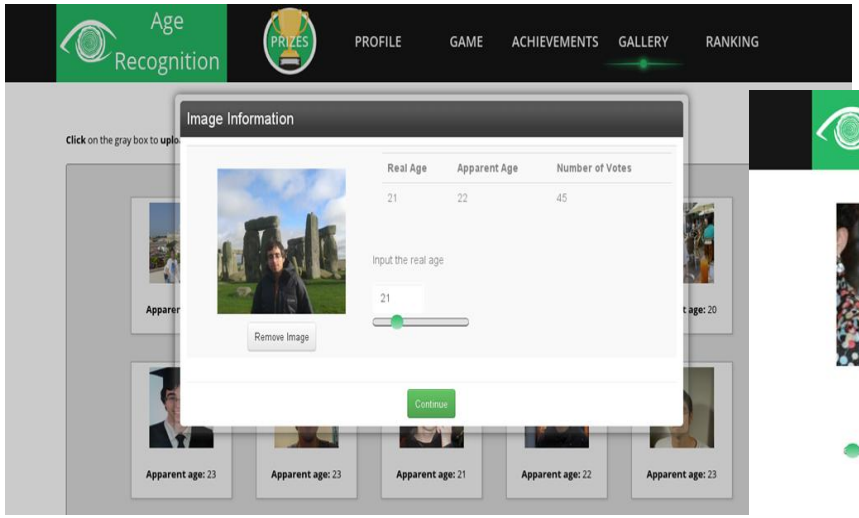
For each image, the mean and standard distribution on the guessed age by all the labelers was provided.

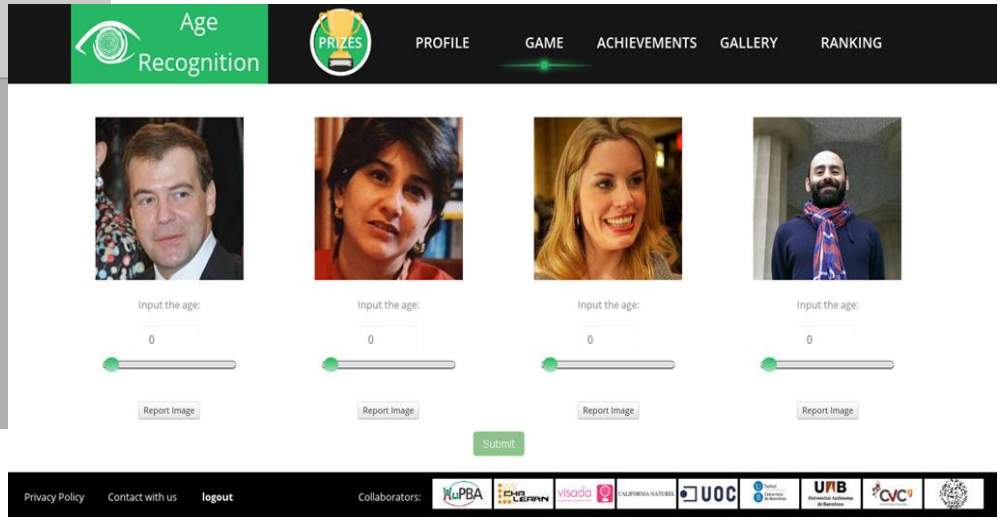
Sergio Escalera, Jordi Gonzàlez, Xavier Baró, Pablo Pardo, Junior Fabian, Marc Oliu, Hugo J. Escalante, Ivan Huerta, Isabelle Guyon, ChaLearn Looking at People 2015 new competitions: Age Estimation and Cultural Event Recognition, IJCNN 2015.

Face Analysis Challenges
Apparent Age Estimation

- The Data (ICCV 2015)

All images were labeled using a custom application based on Facebook, with gamification strategy. We also included images from AgeGess platform (www.ageguess.org).

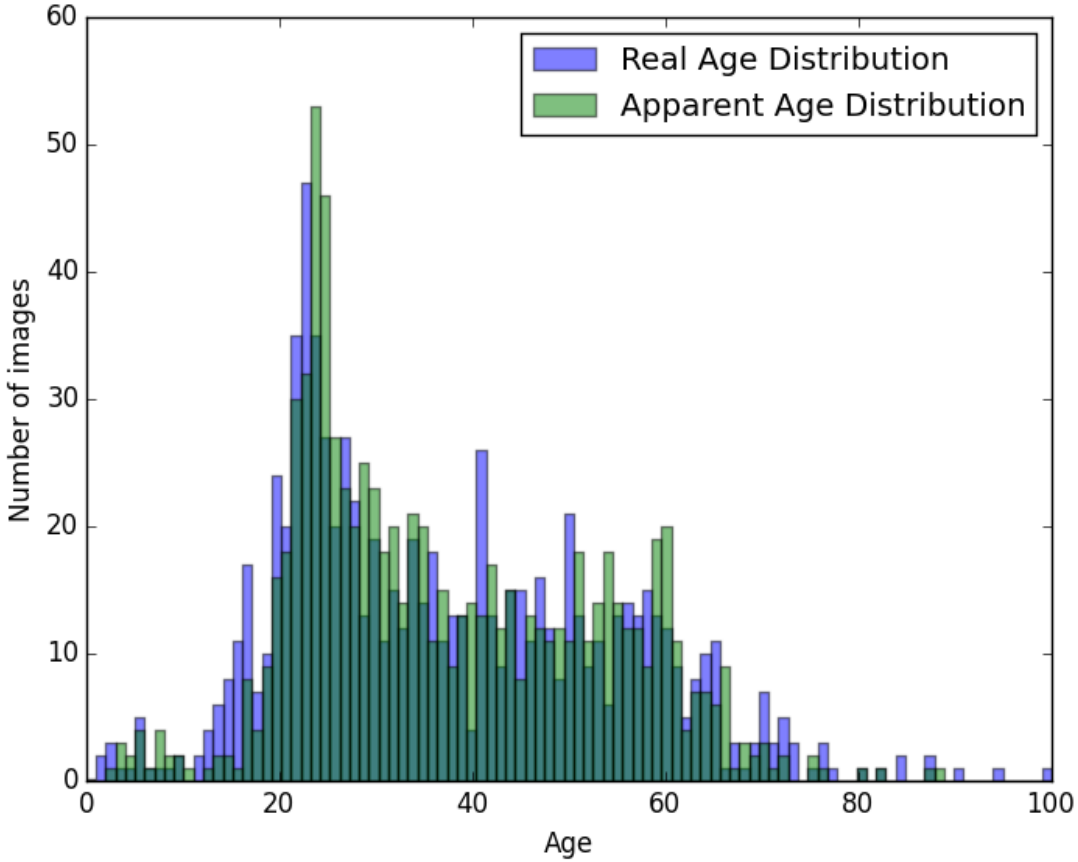




Face Analysis Challenges

Apparent Age Estimation

- The Data (ICCV 2015)



Face Analysis Challenges

Apparent Age Estimation

- **Schedule (ICCV 2015)**
  - **June 15, 2015:** Beginning of the competition, release of development (Training and validation data).
  - **August 30, 2015:** Release of the encrypted final evaluation data and validation labels. Participants can start training their methods with the whole data set.
  - **September 7, 2015:** Release of the decryption key for the final evaluation data. Participants can start predicting the results on the final evaluation labels. Deadline for code submission.
  - **September 12, 2015:** End of the competition. Deadline for submitting the predictions over the final evaluation data. The organisers start the code verification by running it on the final evaluation data.
  - **September 15, 2015:** Deadline for submitting the fact sheets.
  - **September 19, 2015:** Release of the verification results to the participants for review.

Face Analysis Challenges

Apparent Age Estimation

- The Results (ICCV 2015)

Rank	Team	Development	Test
1	**CVL ETHZ**	**0.295116**	**0.264975**
2	**ICT-VIPL**	**0.292297**	**0.270685**
3	AgeSeer	0.327321	0.287266
3	**WVU CVL**	**0.316289**	**0.294835**
4	SEU-NJU	0.380615	0.305763
5	UMD	-	0.373352
6	Enjuto	0.370656	0.37439
7	Sungbin Choi	-	0.420554
8	Lab219A	0.477079	0.499181
9	Bogazici	0.483337	0.524055
10	Notts CVLab	-	0.594248

Face Analysis Challenges

Apparent Age Estimation

- The Data (CVPR 2016)

  The previous dataset was increased up to near 8000 face images. AMT used to label new images

  - **Train Set**: 4113 images
  - **Validation Set**: 1500 images
  - **Test Set**: 1978 images

  For each image, the mean and standard distribution on the guessed age by all the labelers was provided.

Sergio Escalera, Mercedes Torres-Torres, Brais Martinez, Xavier Baró, Hugo Jair Escalante, Isabelle Guyon, Georgios Tzimiropoulos, Ciprian Corneou, Marc Oliu Simón, Mohammad Ali Bagheri, Michel Valstar, ChaLearn Looking at People and Faces of the World: Face AnalysisWorkshop and Challenge 2016, CVPRW, CVPR, 2016

Face Analysis Challenges

Apparent Age Estimation

- The Results (CVPR 2016)
  - Nearly 100 participants registered for the competition
  - The results on the final phase were:

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

Face Analysis Challenges

Face Analysis: Accessories

The aim was to detect the presence of a set of complements and accessories worn by the subjects.

Performance was evaluated using the Mean Square Error between participant's predictions and the ground-truth.

Face Analysis Challenges

Face Analysis: Accessories

- The Data (Faces of the World)

  Over 25,000 images downloaded from Flickr. Zooniverse Citizen Science crowd-sourcing platform used for annotations. Finally face bounding boxes were detected.

Sergio Escalera, Mercedes Torres-Torres, Brais Martinez, Xavier Baró, Hugo Jair Escalante, Isabelle Guyon, Georgios Tzimiropoulos, Ciprian Corneou, Marc Oliu Simón, Mohammad Ali Bagheri, Michel Valstar, ChaLearn Looking at People and Faces of the World: Face AnalysisWorkshop and Challenge 2016, CVPRW, CVPR, 2016

Face Analysis Challenges

Face Analysis: Accessories

- The Data (Faces of the World)

  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

Face Analysis Challenges

Face Analysis: Accessories

- The Data (Faces of the World)

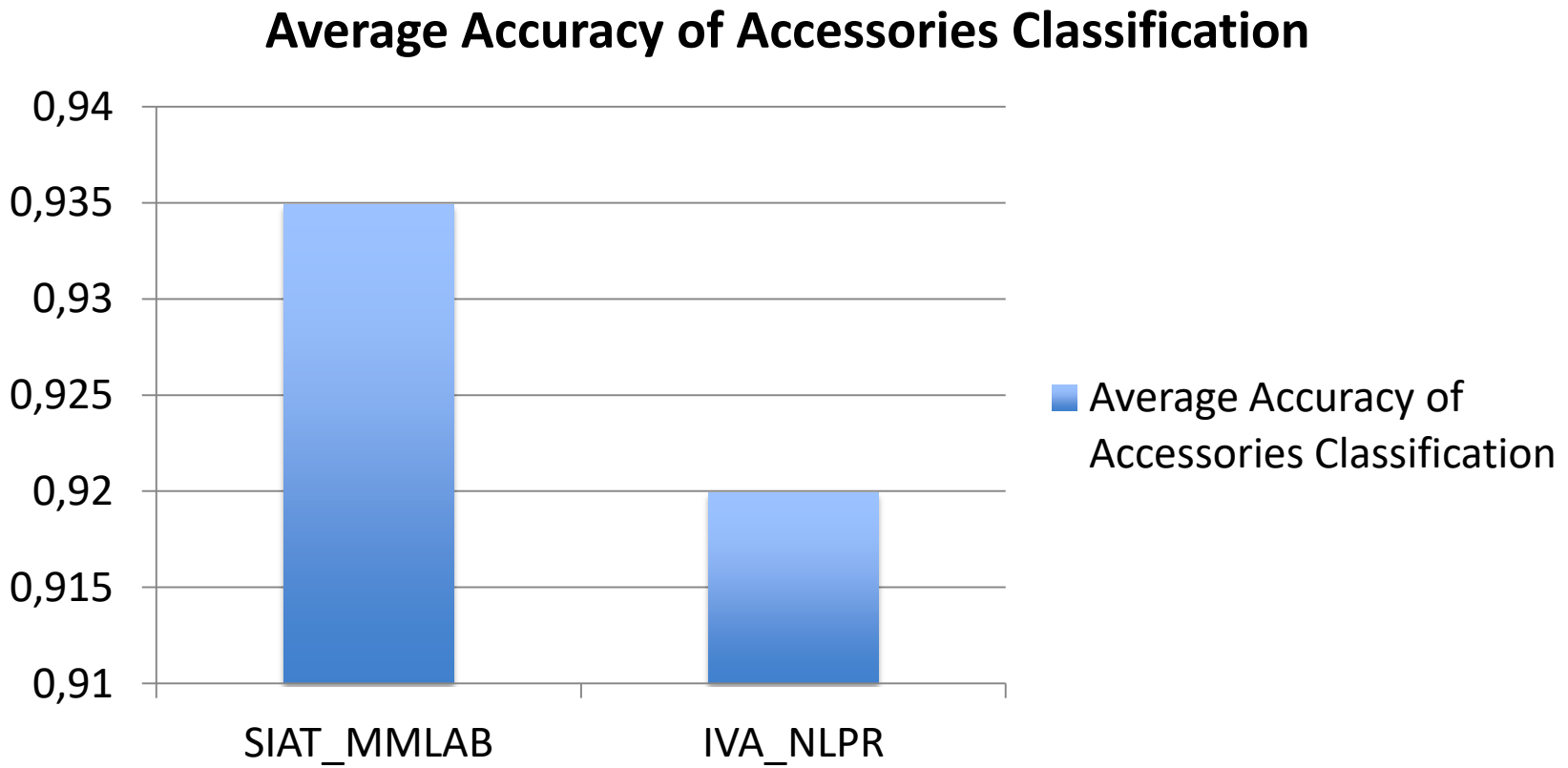
  The dataset consisted of near 12.000 face images
  - **Train Set**: 5,651 images
  - **Validation Set**: 2,826 images
  - **Test Set**: 4,086 images

Accessory	Train	Val	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

Face Analysis Challenges

Face Analysis: Accessories

- The Results (CVPR2016)



Face Analysis Challenges

Face Analysis: Smile and Gender

- The Problem

The aim of this competition is to classify face images in terms of the gender of the person appearing and if is smiling or not.

The evaluation metric is based on classification accuracy.

Face Analysis Challenges

Face Analysis: Smile and Gender

- The Data (Faces of the World)

  Over 25,000 images downloaded from Flickr. Zooniverse, a citizenscience platform, was used to annotate all the images. Finally face bounding boxes were detected.

Sergio Escalera, Mercedes Torres-Torres, Brais Martinez, Xavier Baró, Hugo Jair Escalante, Isabelle Guyon, Georgios Tzimiropoulos, Ciprian Corneou, Marc Oliu Simón, Mohammad Ali Bagheri, Michel Valstar, ChaLearn Looking at People and Faces of the World: Face AnalysisWorkshop and Challenge 2016, CVPRW, CVPR, 2016

Face Analysis Challenges

Face Analysis: Smile and Gender

- The Data (Faces of the World)

  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

Face Analysis Challenges

Face Analysis: Smile and Gender

- The Data (Faces of the World)

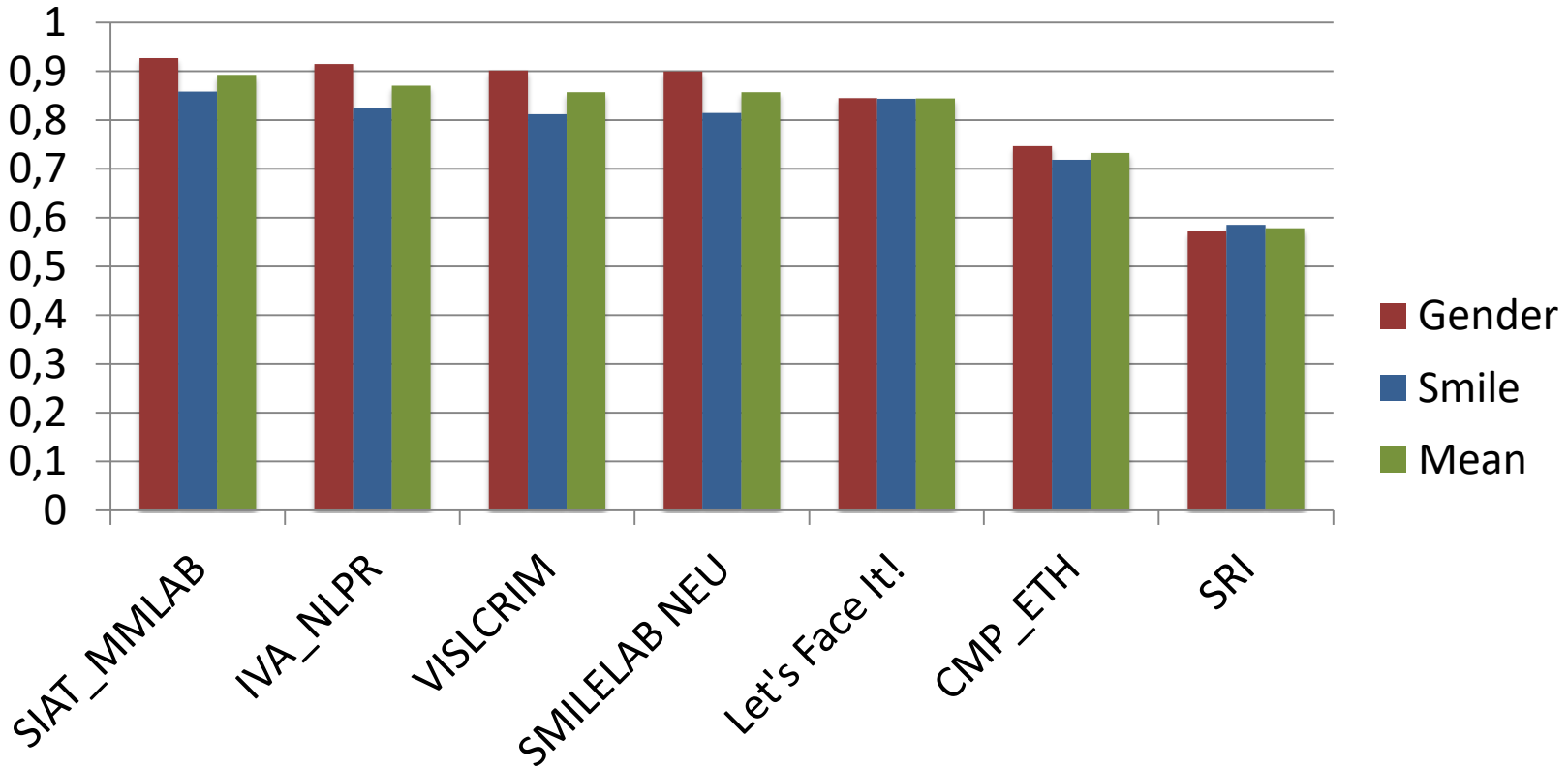
  The dataset consisted of near 18.000 face images
  - **Train Set**: 6,171 images
  - **Validation Set**: 3,086 images
  - **Test Set**: 8,505 images

Attribute	Train	Val	Test
Male	2946	1691	4614
Female	3318	1361	3799
Not sure	93	34	92
Smile	2234	1969	4411
No smile	3937	1117	3849

Face Analysis Challenges

Face Analysis: Smile and Gender

- The Results (CVPR2016)



Face Analysis Challenges

Apparent Personality Analysis

The goal of this competition was to automatically evaluate apparent personality traits from videos of subjects speaking in front of a camera. The traits follow the Big Give model: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism.

The evaluation metric was the mean accuracy over all traits and videos. Accuracy for each trait was computed as:

$$A = 1 - \frac{1}{N_t}\sum_{i=1}^{Nt}|g_i - p_i| / \sum_{i=1}^{Nt}|g_i - \overline{g}|$$

Face Analysis Challenges
Apparent Personality Analysis

- The Data

  A total of 10,000 short clips (15 seconds) were collected from YouTube videos:
  - 41.6 hours of total footage
  - Aprox. 4.5 millions frames

Victor Ponce Lopez, et al. ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results. ChaLearn Looking at People Workshop on Apparent Personality Analysis, ECCV Workshop proceedings, LNCS, Springer, 2016.

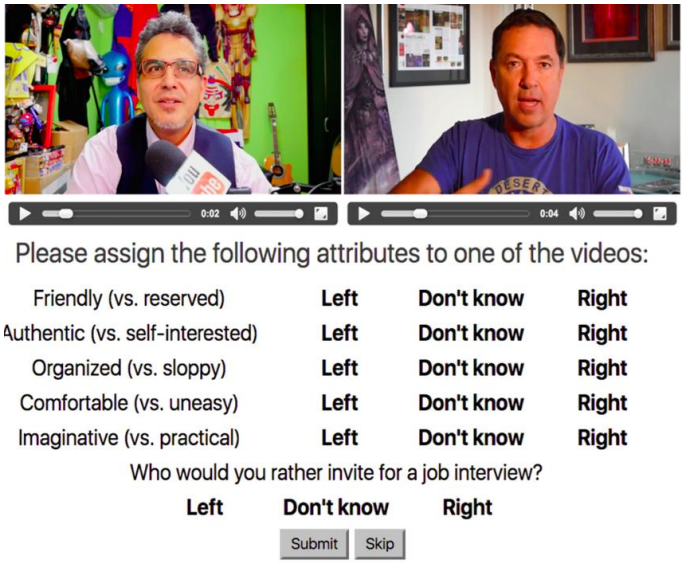
H. J. Escalante, V. P. López, J. Wan, M. Riegler, B. Chen, A. Clapés, S. Escalera, I. Guyon, X. Baro, P. Halvorsen et al. ChaLearn Joint Contest on Multimedia Challenges Beyond Visual Analysis: An Overview. Proc. ICPRW 2016

Face Analysis Challenges

Apparent Personality Analysis

- The Data

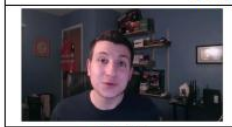
  Annotations where done using AMT. As annotate a personality trait is complex, an interface was developed to make pairwise annotations. Finally we get 321,684 pairwise video annotations. Cardinal scores were obtained by pairwise fitting a Bradley-Terry-Luce (BTL) model.

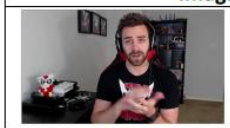
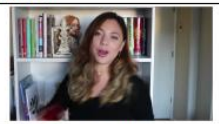


Face Analysis Challenges

Apparent Personality Analysis

- The Data

Agreeableness			
Authentic		Self-interested	
			
0.9230	0.9340	0.1098	0.0879
Conscientiousness			
Organized		Sloppy	
			
0.9708	0.9514	0.0873	0.1068
Extraversion			
Friendly		Reserved	
			
0.9158	0.9252	0.0521	0.0933

Neuroticism			
Comfortable		Uneasy	
			
0.9585	0.9791	0.1005	0.0872
Openness			
Imaginative		Practical	
			
0.9777	0.9582	0.0549	0.1113

Face Analysis Challenges

Apparent Personality Analysis

- The Results (ECCV 2016)

Accuracy score (normalized)							
Rank	Team Name	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Average
	Median Pred.	0.4188 ± 0.0132	0.3179 ± 0.0148	0.4193 ± 0.0097	0.3892 ± 0.0121	0.3749 ± 0.0116	0.3840 ± 0.0123
1	NJU-LAMDA	0.4215 ± 0.0146	**0.3450** ± 0.0210	**0.4497** ± 0.0145	0.4087 ± 0.0171	0.3876 ± 0.0171	**0.4025** ± 0.0169
2	evolgen	0.4358 ± 0.0164	0.3318 ± 0.0178	0.4295 ± 0.0126	0.4069 ± 0.0238	**0.3920** ± 0.0181	0.3992 ± 0.0178
3	DCC	0.3987 ± 0.0217	0.3236 ± 0.0157	0.4310 ± 0.0153	**0.4091** ± 0.0116	0.3740 ± 0.0184	0.3873 ± 0.0165
4	ucas	0.4180 ± 0.0129	0.3123 ± 0.0111	0.4128 ± 0.0168	0.3891 ± 0.0134	0.3811 ± 0.0118	0.3827 ± 0.0132
5	BU-NKU	**0.4416** ± 0.0188	0.2990 ± 0.0175	0.4324 ± 0.0217	0.3586 ± 0.0156	0.3651 ± 0.0162	0.3794 ± 0.0180
6	pandora	0.3771 ± 0.0150	0.3008 ± 0.0187	0.3770 ± 0.0156	0.3767 ± 0.0211	0.3670 ± 0.0200	0.3597 ± 0.0181
7	Pilab	0.2825 ± 0.0142	0.2464 ± 0.0214	0.2581 ± 0.0124	0.2897 ± 0.0142	0.2977 ± 0.0166	0.2749 ± 0.0158
8	Kaizoku	0.1620 ± 0.0314	0.1848 ± 0.0242	0.2183 ± 0.0299	0.1885 ± 0.0313	0.2353 ± 0.0179	0.1978 ± 0.0270
9	ITU-SiMiT	0.1847 ± 0.0067	0.1953 ± 0.0106	0.1750 ± 0.0082	0.1990 ± 0.0099	0.1915 ± 0.0091	0.1891 ± 0.0089
	Random Guess	0.0697 ± 0.0423	0.1253 ± 0.0456	0.0865 ± 0.0512	0.1039 ± 0.0383	0.0799 ± 0.0490	0.0931 ± 0.0453

[1] p_i are the predicted scores, g_i are the ground truth scores, with the sum running over the N_t test videos, and $\overline{g}$ is the average ground truth score over all videos

Face Analysis Challenges

Job candidate screening coopetition

Predicting whether a job seeker would be invited to a job interview.

Produce a human readable explanation of the recommendations.

Face Analysis Challenges

Job candidate screening coopetition

- The Data

  A total of 10,000 short clips (15 seconds) were collected from YouTube videos:

  – 41.6 hours of total footage

  – Aprox. 4.5 millions frames

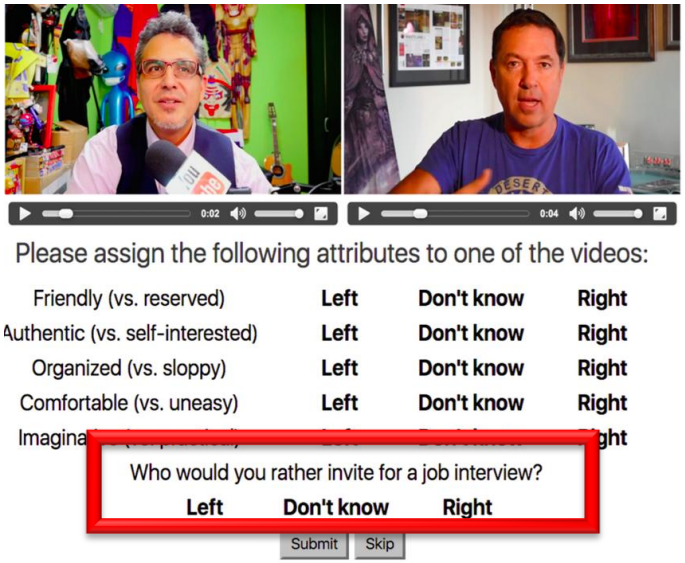
Hugo Jair Escalante, Isabelle Guyon, Sergio Escalera, Julio Jacques Jr., Meysam Madadi, Xavier Baro, Stephane Ayache, Evelyne Viegas, Yagmur Gucluturk, Umut Guclu, Marcel van Gerven, Rob van Lier. Design of an Explainable Machine Learning Challenge for Video Interviews. Proceedings of the The 2017 International Joint Conference on Neural Networks (IJCNN 2017), IEEE, 2017.

Face Analysis Challenges

Job candidate screening coopetition

- The Data

  Annotations where done using AMT. As annotate a personality trait is complex, an interface was developed to make pairwise annotations. Finally we get 321,684 pairwise video annotations. Cardinal scores were obtained by pairwise fitting a Bradley-Terry-Luce (BTL) model.



Face Analysis Challenges

Job candidate screening coopetition

- The Results (CVPR 2017)

Rank	Team	Invite-Interview *	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
1	BU-NKU	0.920916 (1)	0.913731 (1)	0.919769 (1)	0.921289 (1)	0.914613 (1)	0.917014 (1)
-	baseline[46]	0.916202 (2)	0.911230 (2)	0.915228 (2)	0.911220 (3)	0.910378 (2)	0.911123 (2)
2	PML	0.915746 (3)	0.910312 (3)	0.913775 (3)	0.915510 (2)	0.908297 (3)	0.910078 (3)
3	ROCHCI	0.901859 (4)	0.903216 (4)	0.894914 (4)	0.902660 (4)	0.901147 (4)	0.904709 (4)
4	FDMB	0.872129 (5)	0.891004 (5)	0.865975 (5)	0.878842 (5)	0.863237 (5)	0.874761 (5)

Participant	Clarity	Explainability	Soundness	Interpretability	Creativity	Mean_ score
Team A	4.31±0.54	3.58±0.64	3.4±0.66	3.83±0.69	2.67±0.75	3.56
Team B	3.33±1.43	3.23±0.87	3.43±0.9	2.4±1.02	3.4±0.8	3.16

Face Analysis Challenges

Emotion Recognition

The aim of this competition is to being able to recognize when an emotion is real or fake.

The evaluation metric is based on classification accuracy.

Face Analysis Challenges

Emotion Recognition

- The Data

The emotion challenge dataset contains video set of 50 subjects.

12 videos per subject representing 6 basic emotions (angry, happy, sad, disgust, contempt, surprise) for real and fake expressions.

Each video was recorded with a high resolution camera with 100 frames per second and is about 3-4 seconds.

In each video, subjects started from a neutral emotion and the length of this neutral emotion is not predefined.

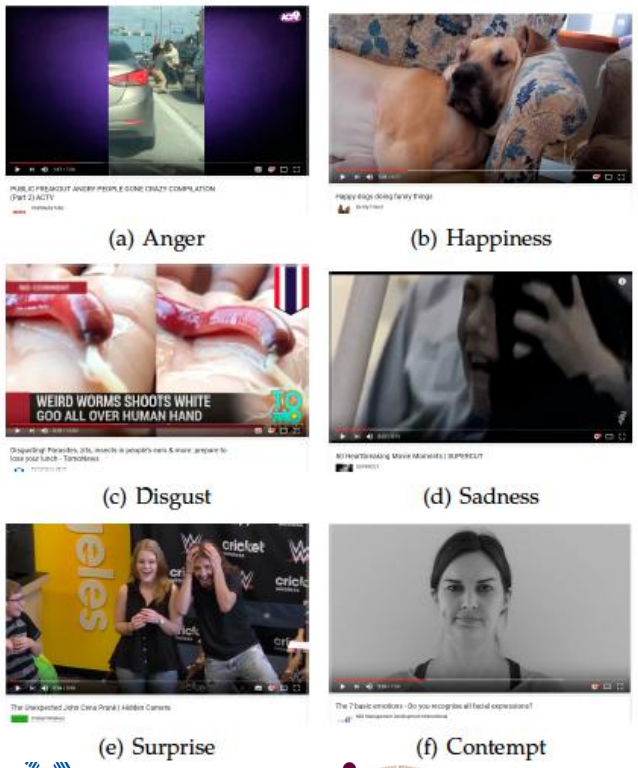
Sets	# of Lables	# of Videos	# of Actors	Labels provided
Training	12	480	40	Yes
Validation	12	60	5	No
Testing	12	60	5	No

Face Analysis Challenges

Emotion Recognition

- The Data

In order for the subjects to express these emotions, they were shown videos which are meant to induce these emotions and were acted accordingly.



(a) Anger (b) Happiness

(c) Disgust (d) Sadness

(e) Surprise (f) Contempt

Face Analysis Challenges

Emotion Recognition

- The Results (ICCV2017)

Results

#	User	Team Name	Perfomance rate ▲
1	bnulee	BNU_CIST	65.000 (1)
2	xjc_faceall	faceall_Xlabs	55.000 (2)
3	Arulkumar	evolgen	55.000 (2)
4	icv	ICV Team	53.000 (3)
5	philippwerner		46.000 (4)

Outline

1. Introduction
2. Designing a Challenge
3. Face Analysis Challenges
4. **Conclusions and future challenge**

Conclusions

- Competitions help boosting performance on specific problems
- Published datasets and results are useful for future publications as a benchmark

Conclusions

- Since 2015 most of the participants based on deep learning strategies
- VGG was the most considered pre-trained model on faces to be fine-tuned on the different competitions
- Face alignment was considered by most of the participants before CNN training
- Ensembles and fusion of networks use to boost final performance

Future

- We are working on a challenge on Generative LAP
  - Data augmentation
  - Occlusion recovery
  - Image photography

Dissemination

Springer
the language of science

springer.com

~ Call for Book Proposals ~

Springer Series on Challenges in Machine Learning

Series Editors from ChaLearn (www.chalearn.org):

Hugo Jair Escalante, ChaLearn, USA & INAOE, Mexico
Isabelle Guyon, ChaLearn, US & Université Paris-Saclay, France
Sergio Escalera, ChaLearn, USA & University of Barcelona, Spain

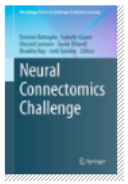
Gesture Recognition

Series: » The Springer Series on Challenges in Machine Learning

Escalera, Sergio, Guyon, Isabelle, Athitsos, Vassilis (Eds.)
2017

Available Formats: eBook Hardcover

More Information



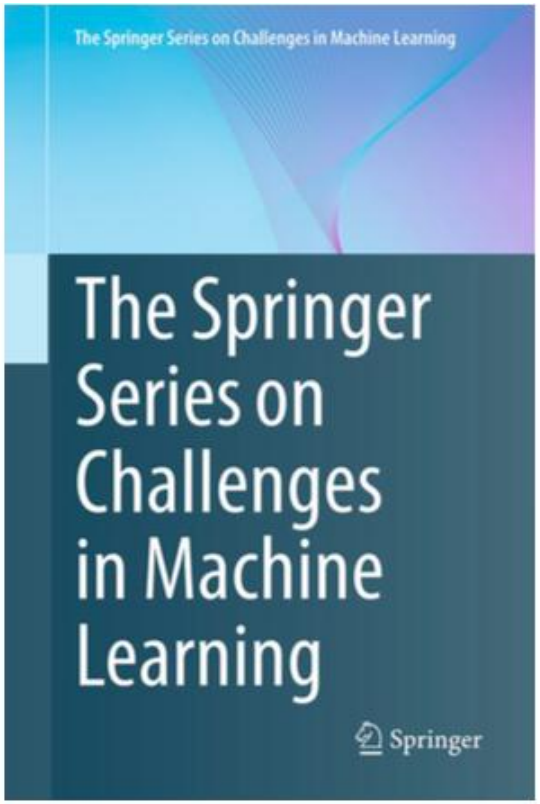
Neural Connectomics Challenge

Series: » The Springer Series on Challenges in Machine Learning

Battaglia, D. (et al.) (Eds.)
2017

Available Formats: eBook Hardcover

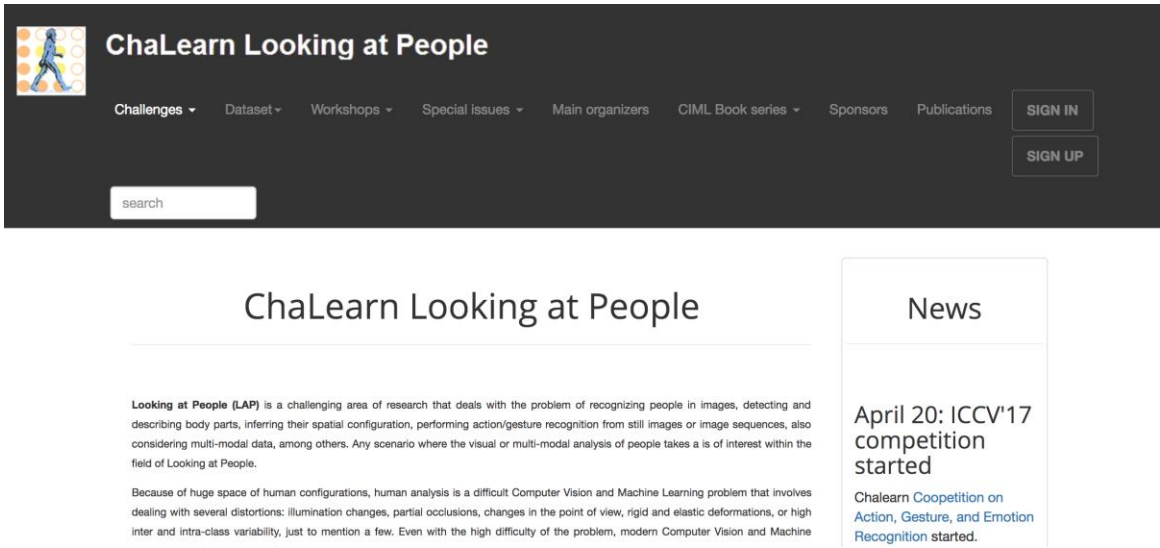
More Information



http://www.springer.com/series/15602

Dissemination

- ChalearnLAP webpage
  - Information of past Challenges
  - Datasets
  - Workshops
  - Special Issues



http://chalearnlap.cvc.uab.es

@ChaLearn_org

ChaLearn Looking at People: Face Analysis Challenges

Joint Challenge on Dominant and Complementary Emotion Recognition Using Micro Emotion Features and Head-Pose Estimation – DCER&HPE. Washington 2017

Xavier Baró – Universitat Oberta de Catalunya

Sergio Escalera – Universitat de Barcelona

Hugo J. Escalante – INAOE

Isabelle Guyon - University Paris-Saclay and Chalearn