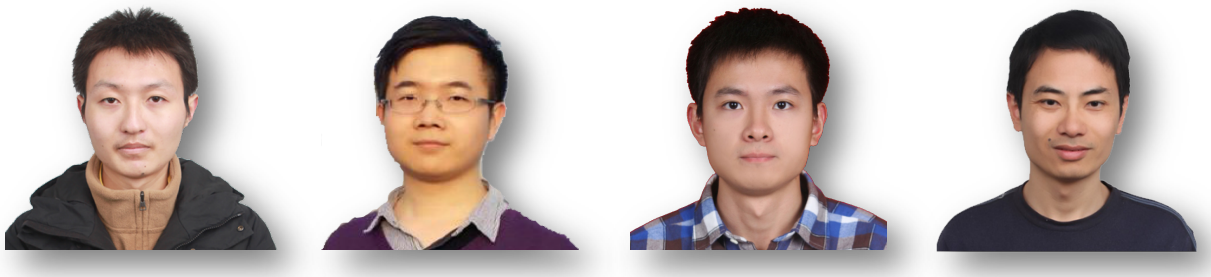
Deep Bimodal Regression for Apparent Personality Analysis

Chen-Lin Zhang, Hao Zhang, Xiu-Shen Wei and Jianxin Wu
National Key Laboratory for Novel Software Technology, Nanjing University



Oct. 9, 2016 Amsterdam, Netherlands

ECCV'16
EUROPEAN CONFERENCE ON COMPUTER VISION

Outline

LAMDA
Learning And Mining from DatA

- Backgrounds
- Deep Bimodal Regression (DBR)
- Implementation details
- Experimental results

Backgrounds

LAMDA
Learning And Mining from DatA

Apparent personality analysis (APA):

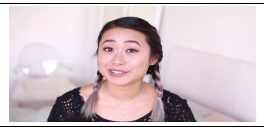
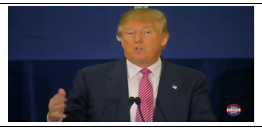
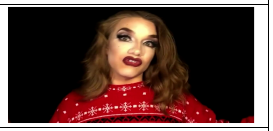
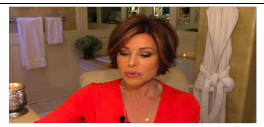
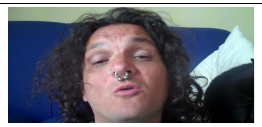
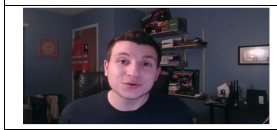
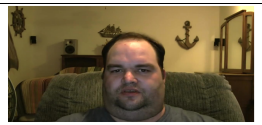
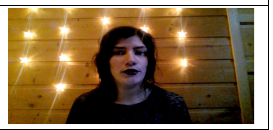
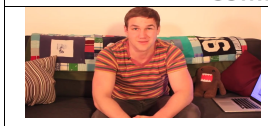
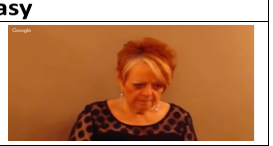
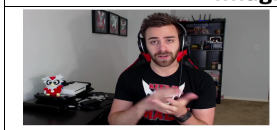
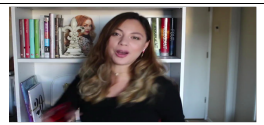
Agreeableness			
Authentic		Self-interested	
			
0.9230	0.9340	0.1098	0.0879
Conscientiousness			
Organized		Sloppy	
			
0.9708	0.9514	0.0873	0.1068
Extraversion			
Friendly		Reserved	
			
0.9158	0.9252	0.0521	0.0933
Neuroticism			
Comfortable		Uneasy	
			
0.9585	0.9791	0.1005	0.0872
Openness			
Imaginative		Practical	
			
0.9777	0.9582	0.0549	0.1113

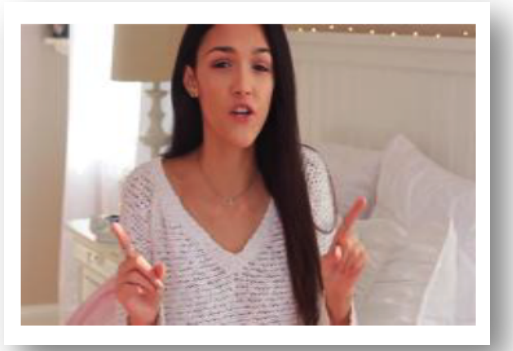
Figure courtesy of [V. Ponce-López et. al., ECCV'16 LAP Workshop]

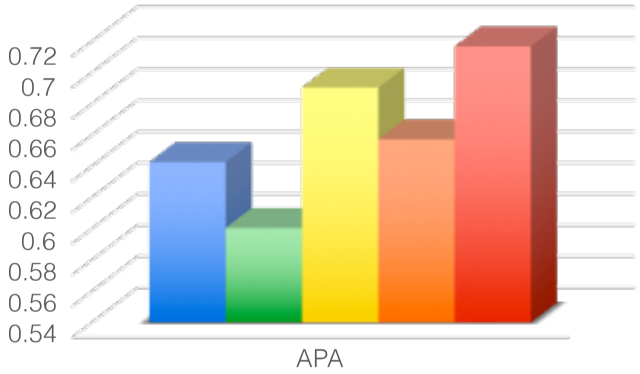
Backgrounds (con't)

LAMDA
Learning And Mining from DatA

Apparent personality analysis:

Single image:





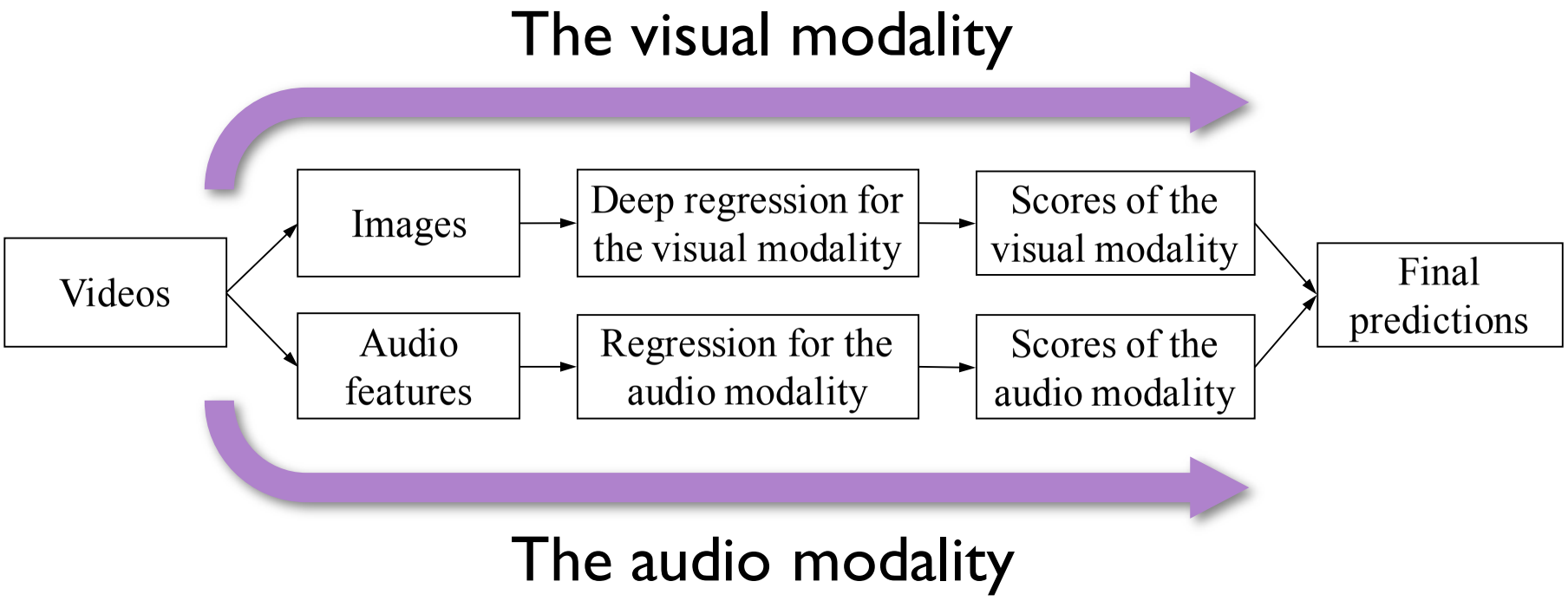
Videos:

Figures courtesy of [V. Ponce-López et. al., ECCV'16 LAP Workshop]

Deep Bimodal Regression (DBR)

Learning And Mining from DatA

The proposed DBR framework:



DBR (con't)

The visual modality:

- Res-Net
- Descriptor Aggregation Network (DAN)
- DAN^{+}

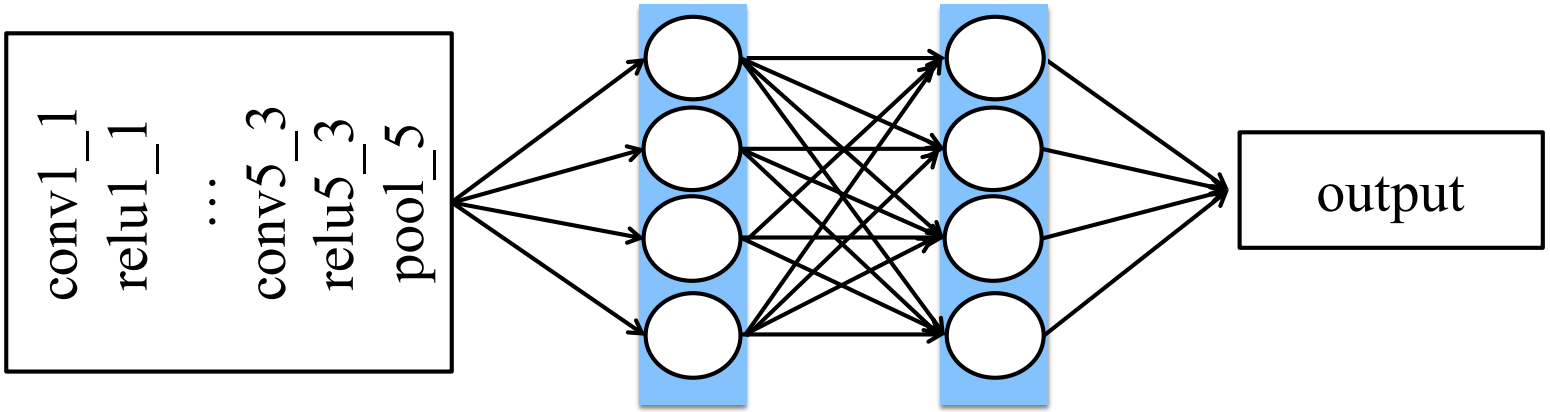
The audio modality:

- The Logfbank feature
- Linear regression

Descriptor Aggregation Network (DAN)

LAMDA
Learning And Mining from DatA

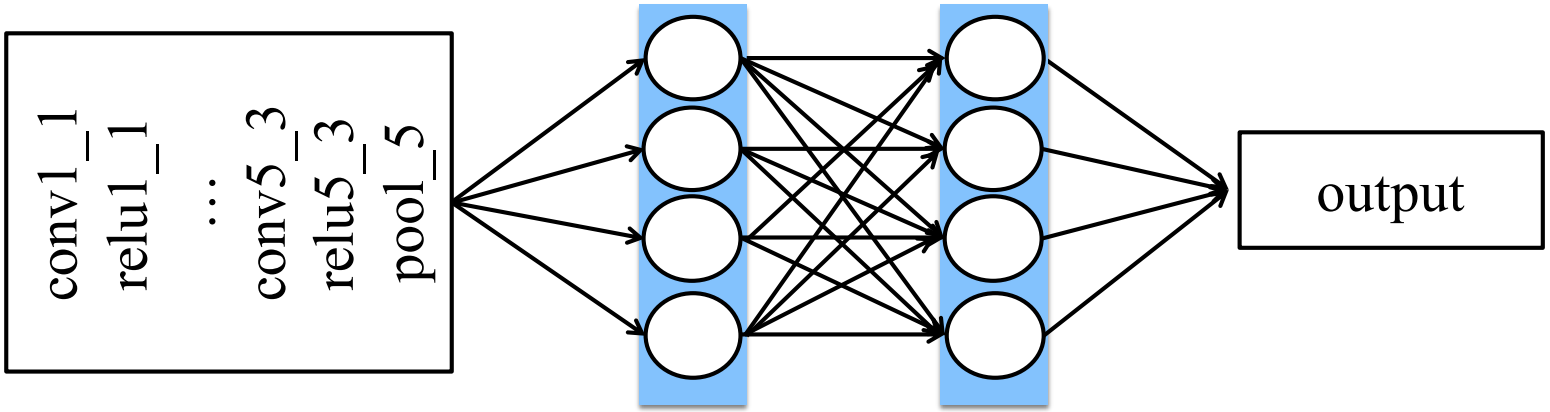
Traditional CNN (VGG-16):



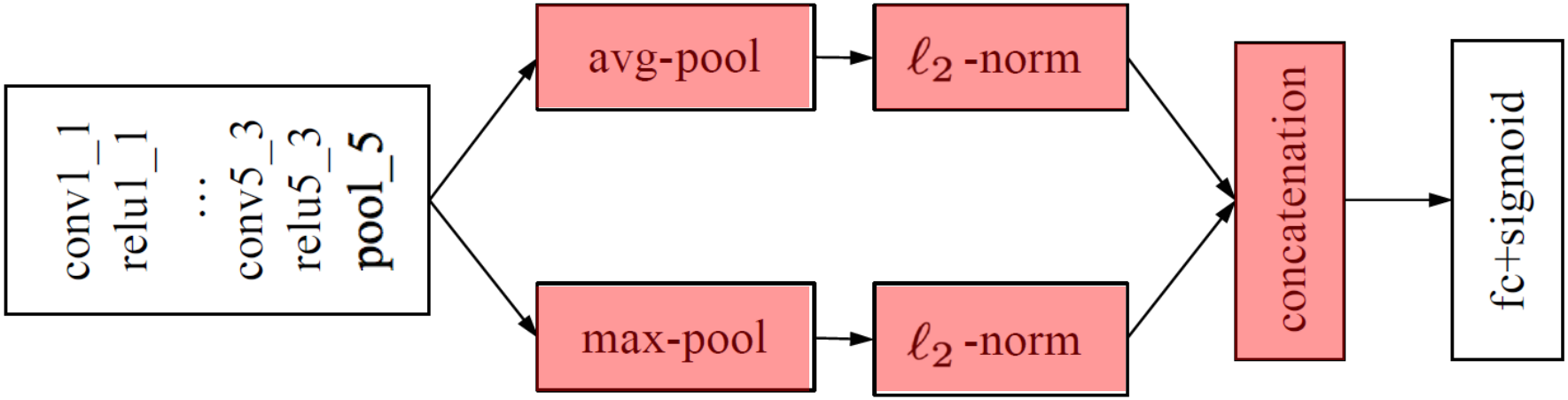
Descriptor Aggregation Network (DAN)

Learning And Mining from DatA

Traditional CNN (VGG-16):



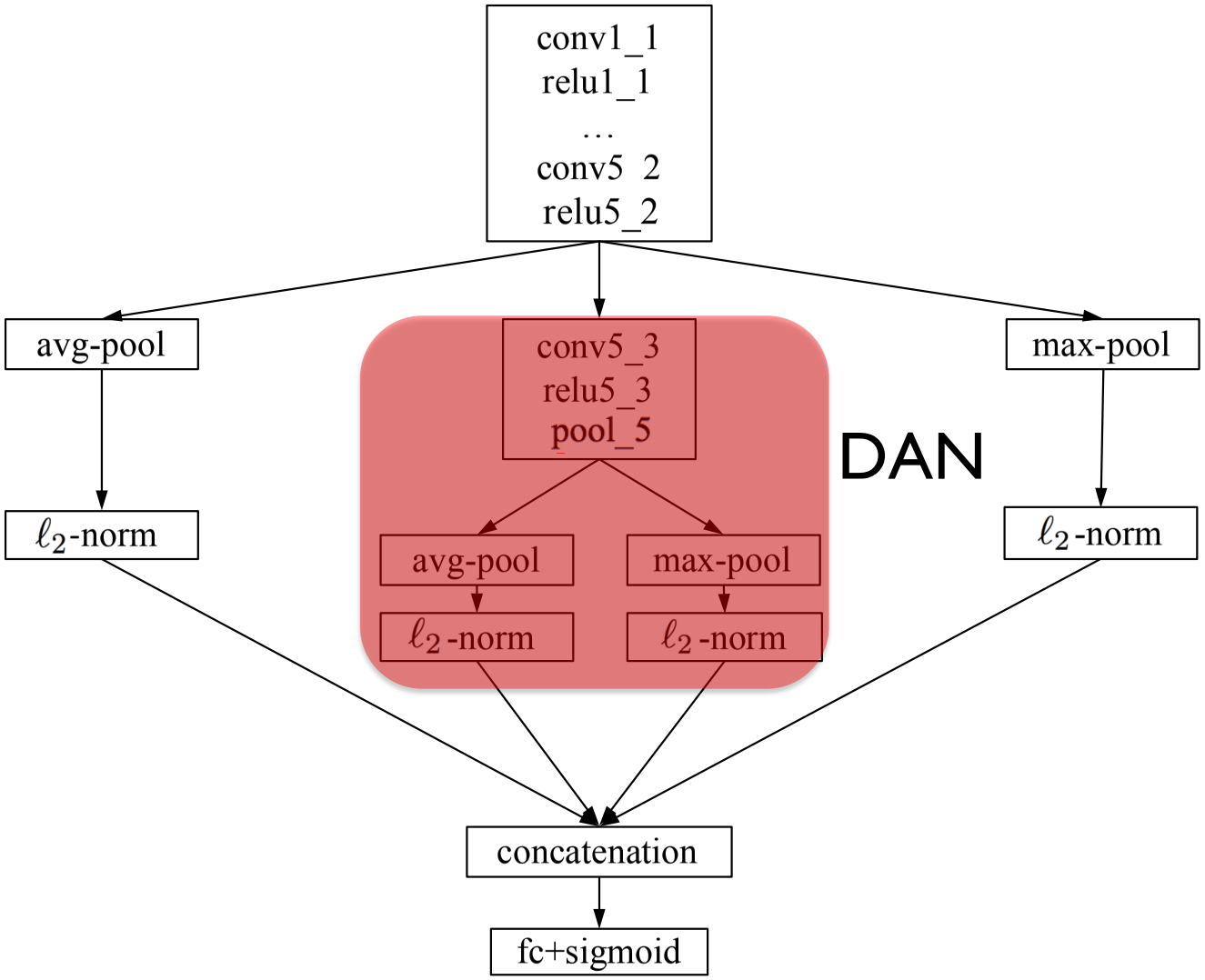
DAN:



DAN and DAN+

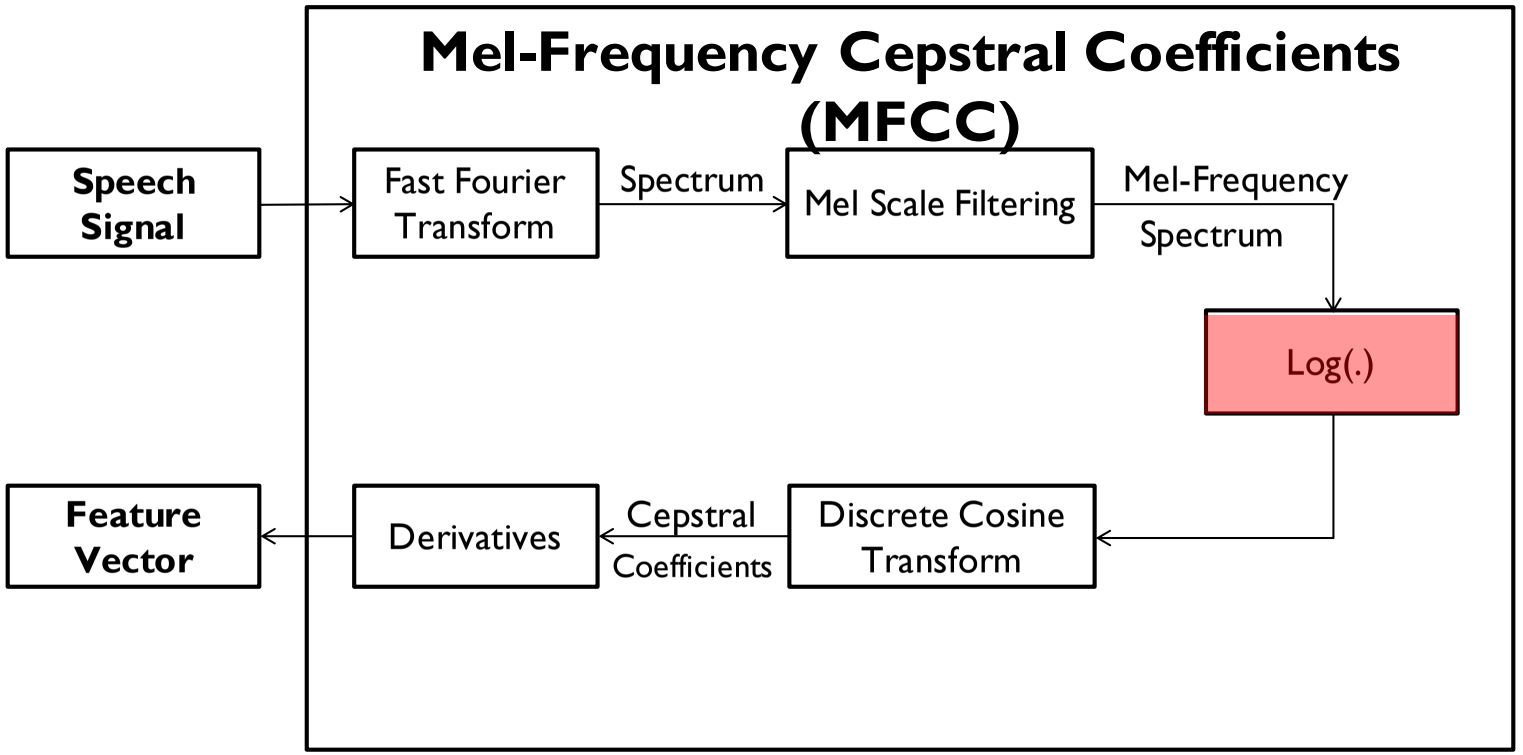
LAMDA
Learning And Mining from DatA

DAN+:



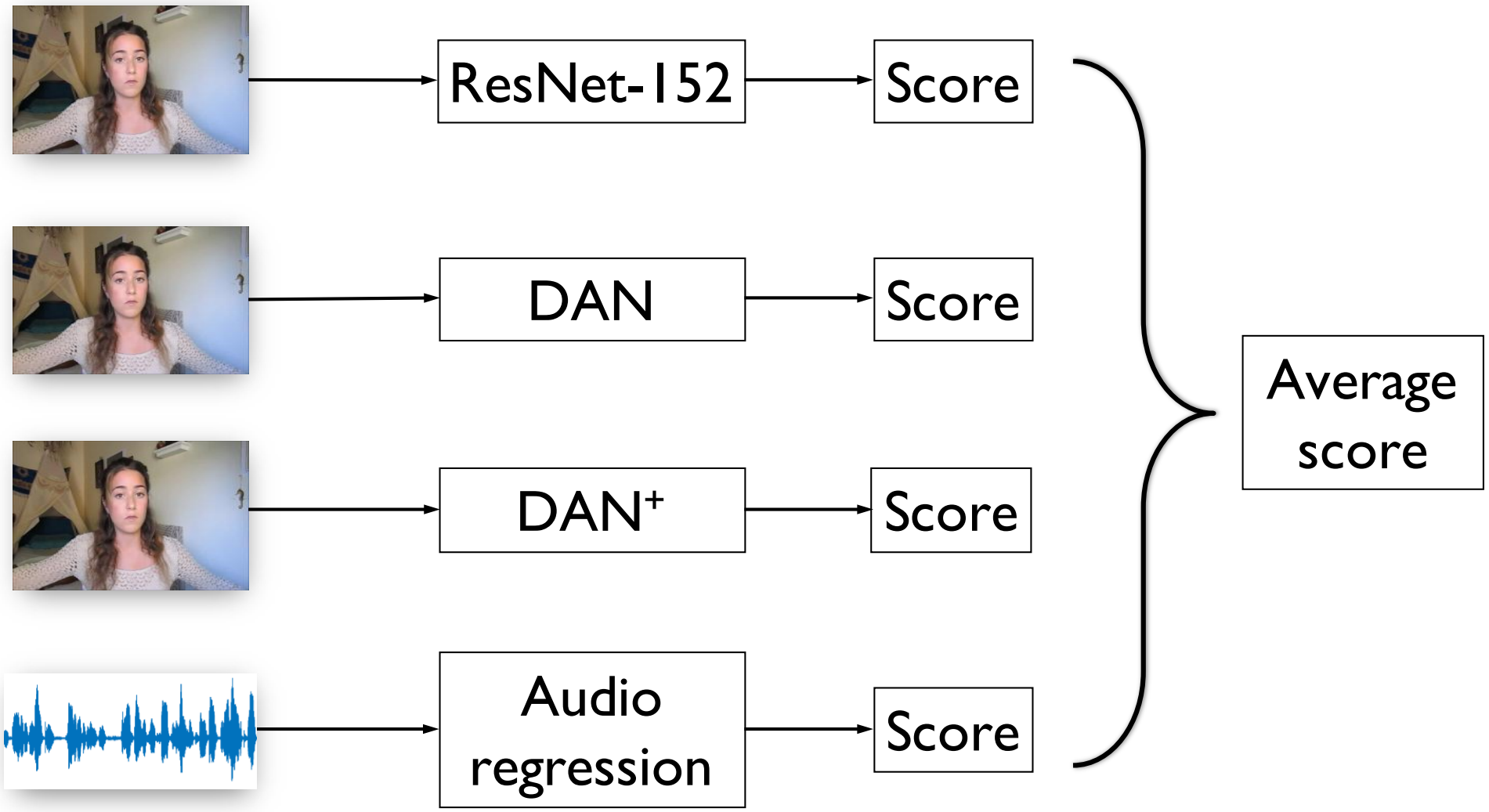
Audio modality

- The MFCC feature
- The logfbank feature



- Linear regression

Modality ensemble



Implementation details

LAMDA
Learning And Mining from DatA

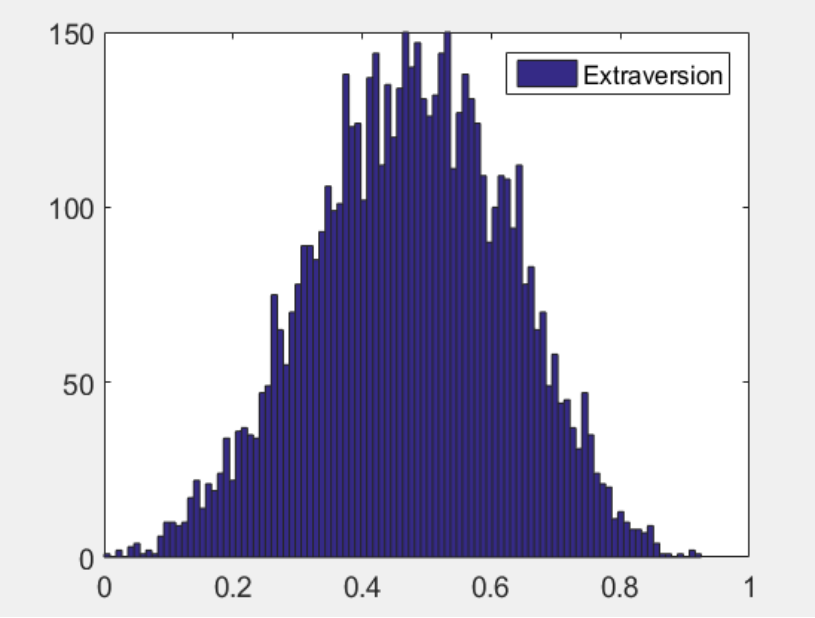
Extracting images from the original videos:

Implementation details (con't)

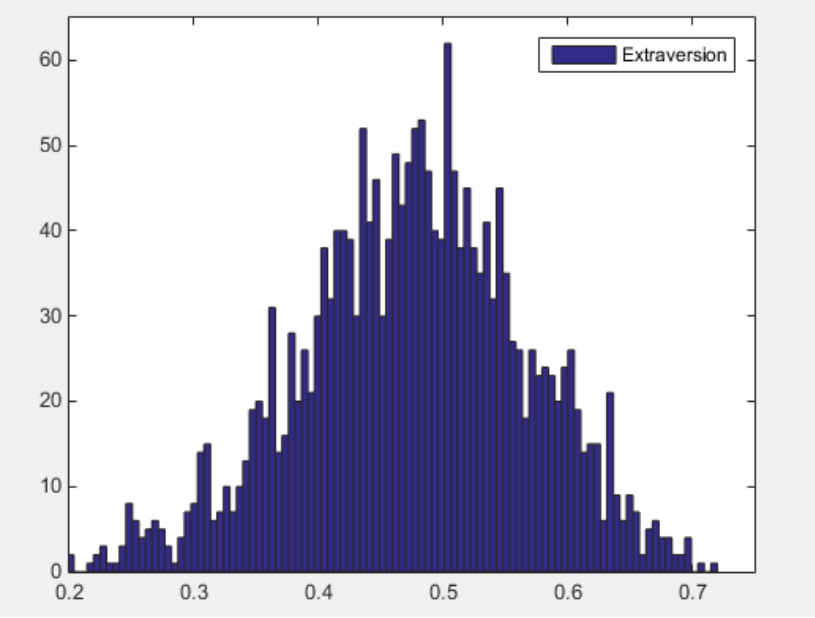
Learning And Mining from DatA

The distributions of video labels:

Training



Testing



Experiment results

LAMDA
Learning And Mining from DatA

Single model comparison at the training stage:

Modality	Model	♯ Para.	Dim.	*Epoch 1*	*Epoch 2*	*Epoch Fusion*
Visual	VGG-Face	134.28M	4,096	0.9065	0.9060	0.9072
	ResNet	58.31M	512	0.9072	0.9063	0.9080
	DAN	14.71M	1,024	0.9082	0.9080	0.9100
	DAN$^+$	14.72M	2,048	0.9100	0.9103	0.9111
Audio	Linear regressor	0.40M	79,534	0.8900	–	0.8900

Model speed testing on a GTX 1080 GPU:

Inference speed	Single image	Whole video
VGG-Face	15ms	400ms
Res-Net	149ms	450ms
DAN	12ms	389ms
DAN$^+$	13ms	390ms

Experiment results (con't)

Fusion accuracy at the final testing stage:

Rank	Team name	*Mean Acc.*	*Extra.*	*Agree.*	*Consc.*	*Neuro.*	*Open.*
1	**Ours**	**0.9130**	0.9133	**0.9126**	**0.9166**	**0.9100**	**0.9123**
2	evolgen	0.9121	0.9150	0.9119	0.9119	0.9099	0.9117
3	DCC	0.9109	0.9107	0.9102	0.9138	0.9089	0.9111
4	ucas	0.9098	0.9129	0.9091	0.9107	0.9064	0.9099
5	BU-NKU	0.9094	**0.9161**	0.9070	0.9133	0.9021	0.9084

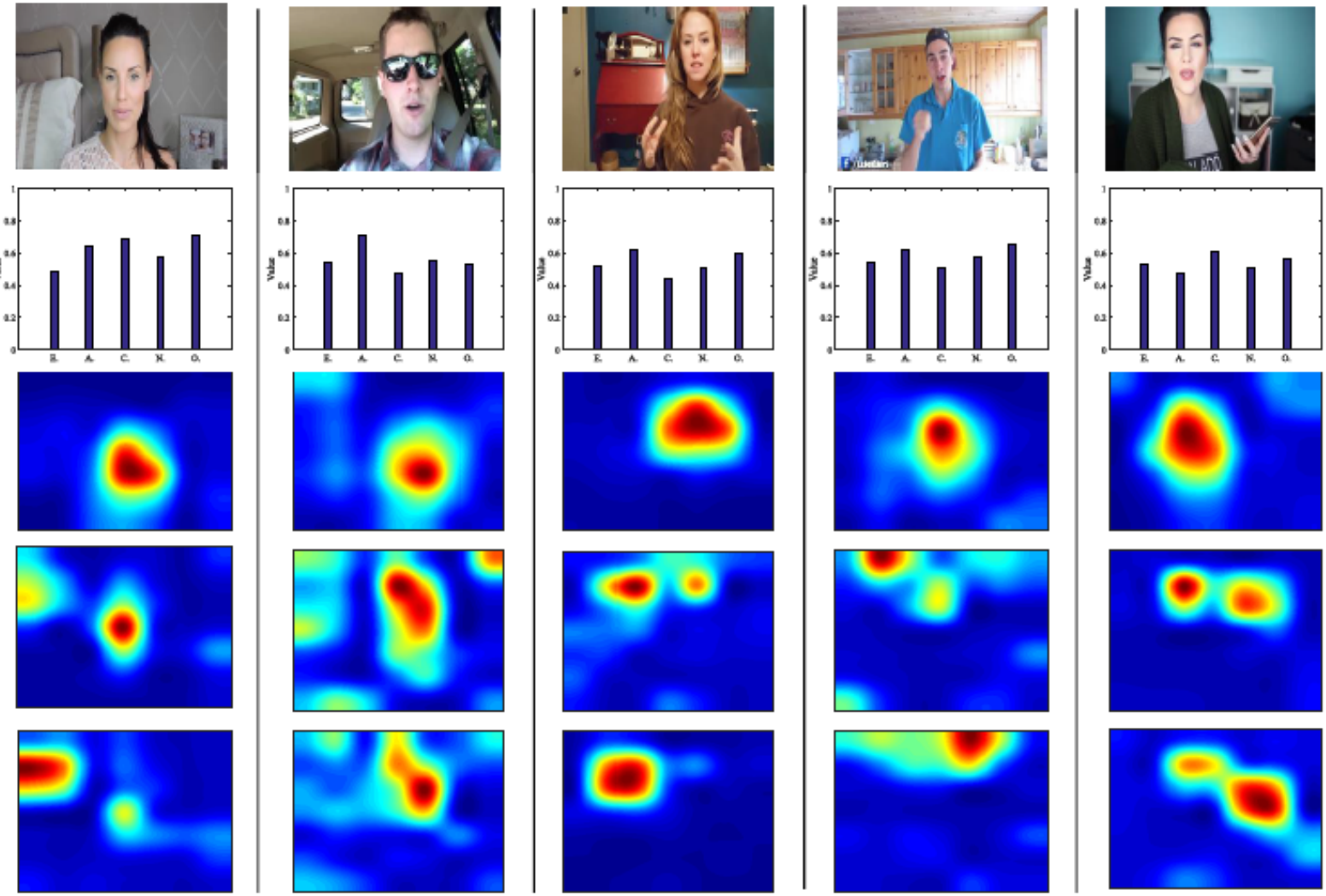
Visualization

Learning And Mining from DatA

ResNet-152

DAN

DAN+



Future works

- Train more discriminative deep audio representations;
- Try to use the time series information for further improving the visual modality's accuracy

LAMDA
Learning And Mining from DatA

Thank you!

Our source codes can be found via:
https://github.com/tzzcl/ChaLearn-APA-Code