A spatiotemporal deep learning solution for Automatic Micro-Expressions Recognition From Local Facial Regions

Authors :

Mouath Aouayeb
Wassim Hamidouche
Kidiyo Kpalma
Amel Benazza-Benyahia

IEEE MLSP 2019

Pittsburgh, PA, USA

SUP'COM
Higher School of Communication of Tunis

INSA
INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

IETR
INSTITUT D'ÉLECTRONIQUE ET DE TÉLÉCOMMUNICATIONS DE RENNES

Outline

I. Introduction

II. State-of-the-art

III. Proposed solution

IV. Experiments and results

V. Conclusion and perspectives

Outline

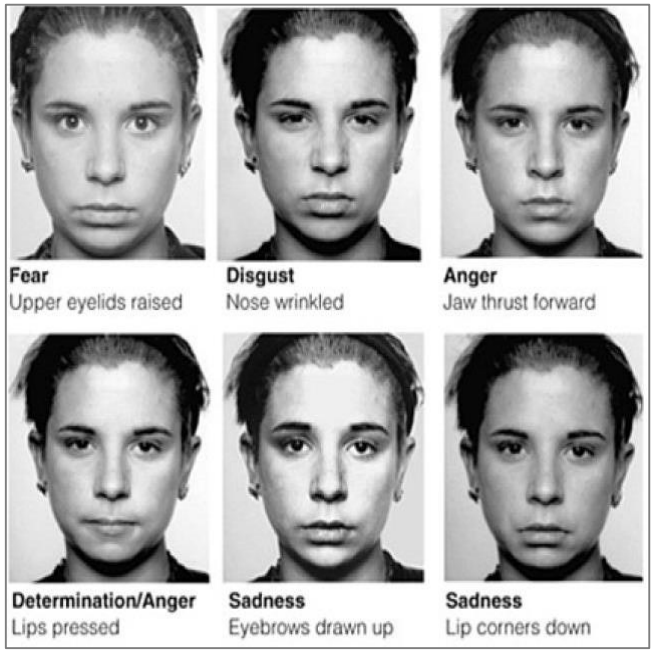
Police Interrogation

Psycho treatment

Driven warning

Macro-Expressions

- Obvions / Intense
- Global Face reaction
- > 1/2 s
- Real / Fake
- Gender, ethnicity, age, ...



Micro-Expressions

- Low intensity
- Local Face reaction
- < 1/2 s (to 1/25 s)
- Spontaneous (Real)
- Universal

Truths VS **Lies**

↓

Macro VS **Micro**

But ...

&

Challenging task

→ **Deep Learning**

Outline

I. Introduction

II. State-of-the-art
   I. Handcrafted Approach
   II. DL Approach
   III. Hybrid Approach
   IV. Region based Approach

III. Proposed solution

IV. Experiments and results

V. Conclusion and perspectives

- Handcrafted Approach
- Hybrid Approach
- DL Approach
- Region based Approach

Handcrafted Solutions

Deep Learning Solutions

Hybrid Solutions

Region based Solutions

Handcrafted Solutions :

- LBP-TOP : Local Binary Pattern on Three Orthogonal Planes
- 3D-HOG : 3D-Gradients orientation Histogram
- Bi-WOOF : Bi-Weighted Oriented Optical Flow
- HOOF : Histogram of Oriented Optical Flow

Deep Learning Solutions :

- CNN + LSTM
- 3D-CNN: 3D Spatiotemporal CNN
- LEARNet : Lateral Accretive Hybrid Network
- CapsuleNet

Hybrid Solutions :

- ELRCN : Enriched Long-term Recurrent CNN
- Off-ApexNet : Optical Flow Features from Apex frame Network
- STSTNet : Shallow Triple Stream Three-dimensional CNN
- STRCN : Spatiotemporal Recurrent Convolution Network

Region based Solutions :

- NMPs: Necessary Morphological Patches
- Improved version of NMPs
- MicroExpFuseNet

Handcrafted Solutions :

- LBP-TOP : Local Binary Pattern on Three Orthogonal Planes
- 3D-HOG : 3D-Gradients orientation Histogram
- Bi-WOOF : Bi-Weighted Oriented Optical Flow
- HOOF : Histogram of Oriented Optical Flow

Deep Learning Solutions :

- CNN + LSTM
- 3D-CNN: 3D Spatiotemporal CNN
- LEARNet : Lateral Accretive Hybrid Network
- CapsuleNet

Hybrid Solutions :

- ELRCN : Enriched Long-term Recurrent CNN
- Off-ApexNet : Optical Flow Features from Apex frame Network
- STSTNet : Shallow Triple Stream Three-dimensional CNN
- STRCN : Spatiotemporal Recurrent Convolution Network

Region based Solutions :

- NMPs: Necessary Morphological Patches
- Improved version of NMPs
- MicroExpFuseNet

Handcrafted Solutions :

- LBP-TOP : Local Binary Pattern on Three Orthogonal Planes
- 3D-HOG : 3D-Gradients orientation Histogram
- Bi-WOOF : Bi-Weighted Oriented Optical Flow
- HOOF : Histogram of Oriented Optical Flow

Deep Learning Solutions :

- CNN + LSTM
- 3D-CNN: 3D Spatiotemporal CNN
- LEARNet : Lateral Accretive Hybrid Network
- CapsuleNet

Hybrid Solutions :

- ELRCN : Enriched Long-term Recurrent CNN
- Off-ApexNet : Optical Flow Features from Apex frame Network
- STSTNet : Shallow Triple Stream Three-dimensional CNN
- STRCN : Spatiotemporal Recurrent Convolution Network

Region based Solutions :

- NMPs: Necessary Morphological Patches
- Improved version of NMPs
- MicroExpFuseNet

Handcrafted Solutions :

- LBP-TOP : Local Binary Pattern on Three Orthogonal Planes
- 3D-HOG : 3D-Gradients orientation Histogram
- Bi-WOOF : Bi-Weighted Oriented Optical Flow
- HOOF : Histogram of Oriented Optical Flow

Deep Learning Solutions :

- CNN + LSTM
- 3D-CNN: 3D Spatiotemporal CNN
- LEARNet : Lateral Accretive Hybrid Network
- CapsuleNet

Hybrid Solutions :

- ELRCN : Enriched Long-term Recurrent CNN
- Off-ApexNet : Optical Flow Features from Apex frame Network
- STSTNet : Shallow Triple Stream Three-dimensional CNN
- STRCN : Spatiotemporal Recurrent Convolution Network

Region based Solutions :

- NMPs: Necessary Morphological Patches
- Improved version of NMPs
- MicroExpFuseNet

Outline

I. Introduction

II. State-of-the-art

III. Proposed solution
   I. Idea and added value
   II. Overview
   III. CNN
   IV. LSTM

IV. Experiments and results

V. Conclusion and perspectives

MiEs : Catch me If you can :p

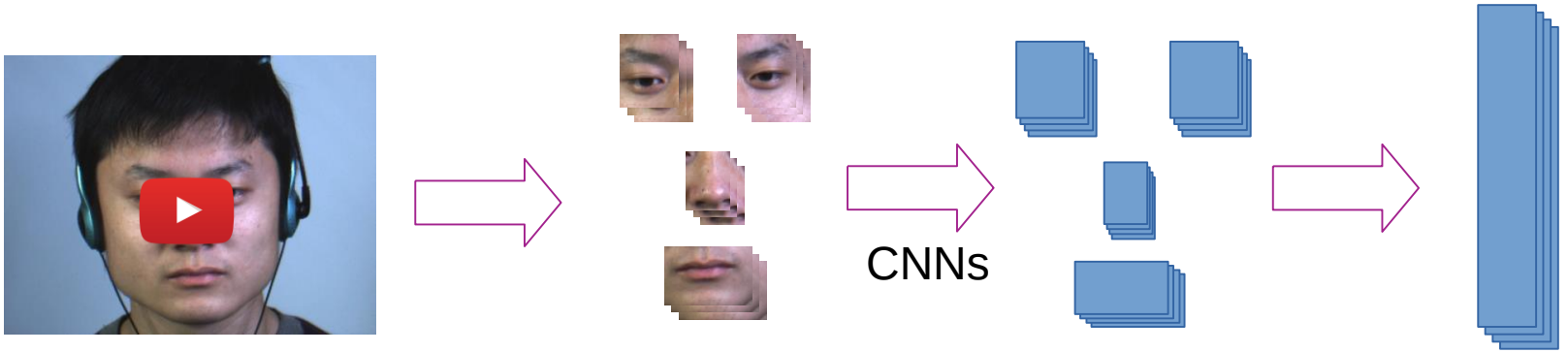
Me : Hey ! First of all, are you local or global reaction of the face

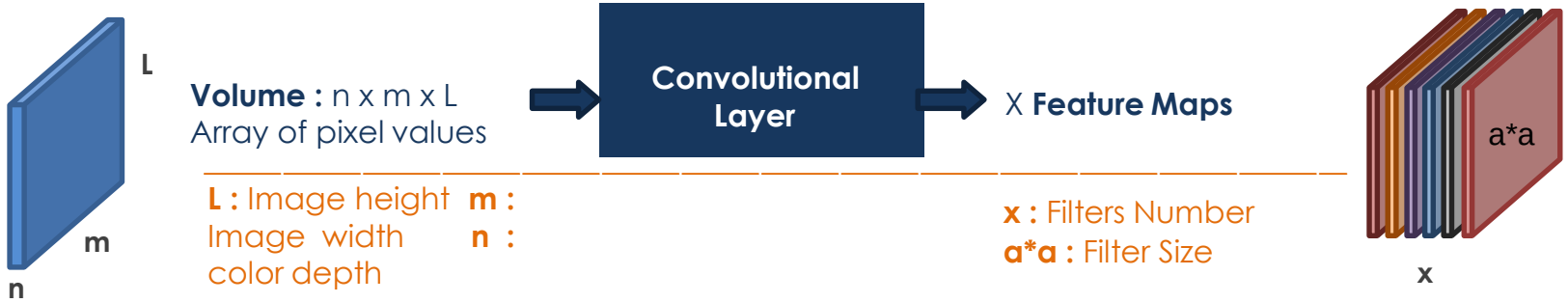
MiEs : Hmmmm I really don't know, but you can ask Paul Ekman

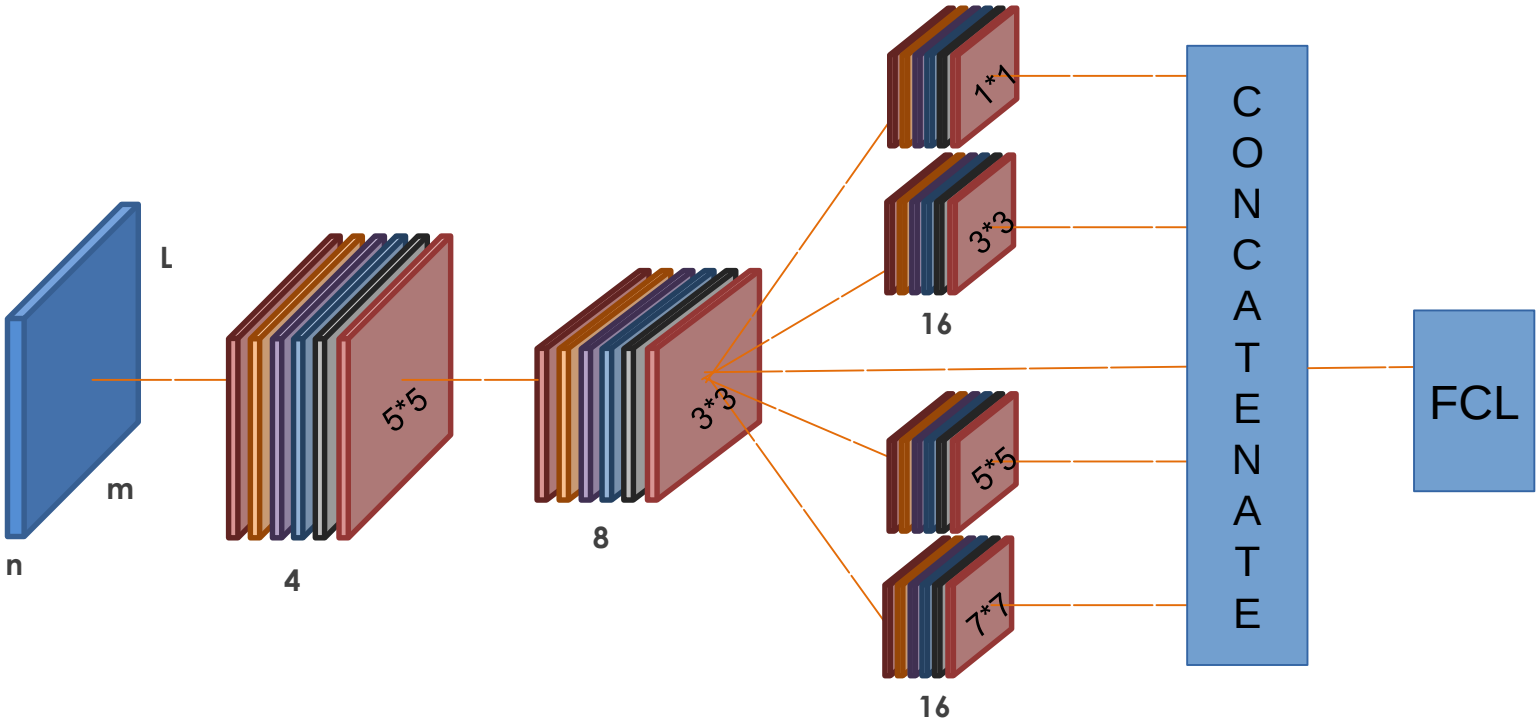
P. Ekman : yeah !! you can say that

Zaho et al. : we confirm, we recently did a research on that and guess what! we've got more than 20% precision higher with only traditional method. Get yourself ready MiEs :p :p

Me : Hmmmm OK, so the less parts I use the more relevant spatio-temporal features I got and with DL I expect to get a better result ….. good, thanks


CNNs
LSTM
Spatio-temporal features
FCL
MiE Recognition
Happy
Sad
Angry
Surprise
Disgust


L
m
n
Volume : n x m x L
Array of pixel values
Convolutional Layer
X Feature Maps
L : Image height m : Image width n : color depth
x : Filters Number
a*a : Filter Size
a*a
x


L
m
n
5*5
4
3*3
8
1*1
3*3
16
5*5
7*7
16
CONCATENATE
FCL

LSTM: Long Short Term Memory

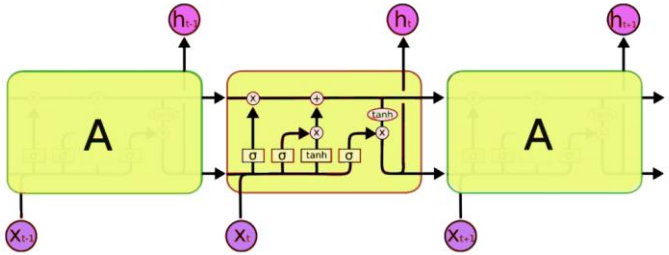
RNN: Recurrent Neural Network

Sequences

Evoloution of Filters through Time

Words in paragraph

Video etc ...



LSTM architecture

Outline

I. Introduction

II. State-of-the-art

III. Proposed solution

IV. Experiments and results
   I. Data
   II. Experimental Setup
   III. Evaluation Metric
   IV. Results & Discussion

V. Conclusion and perspectives

Provided Database : Spontaneous Micro-Expressions

		SMIC	CASME II	SAMM
Participants		16	24	28
Frame rate (*fps*)		100	200	200
Avg. frame number		34	68	74
Avg. video duration (*s*)		0.34	0.34	0.37
Ground-truth (index)	Onset	Yes	Yes	Yes
	Offset	Yes	Yes	Yes
	Apex	No	Yes	Yes
Number of classe		3	5	7
Number of samples		164	255	159

- CASME I
- **CASME II**
- CAS(ME)^2
- **SAMM**
- SMIC-SUB
- **SMIC**
- Polikovsky's
- USF-HD
- MEVIEW
- YorkDDT
- ...

Provided Database :

3 Classes : Emotions : - / + / s

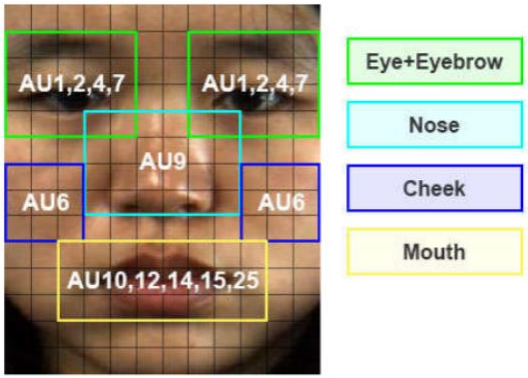
Emotion Class	SMIC	CASME II	SAMM	3DB-combined
Negative	70	88†	92‡	250
Positive	51	32	26	109
Surprise	43	25	15	83
TOTAL	164	145	133	442

MEGC 2019

- Negative class of CASMEII consists of samples from its original emotions class of Disgust and Repression
- Negative class of SAMM consists of samples from original emotions class of Anger, Contempt, Disgust, Fear and Sadness

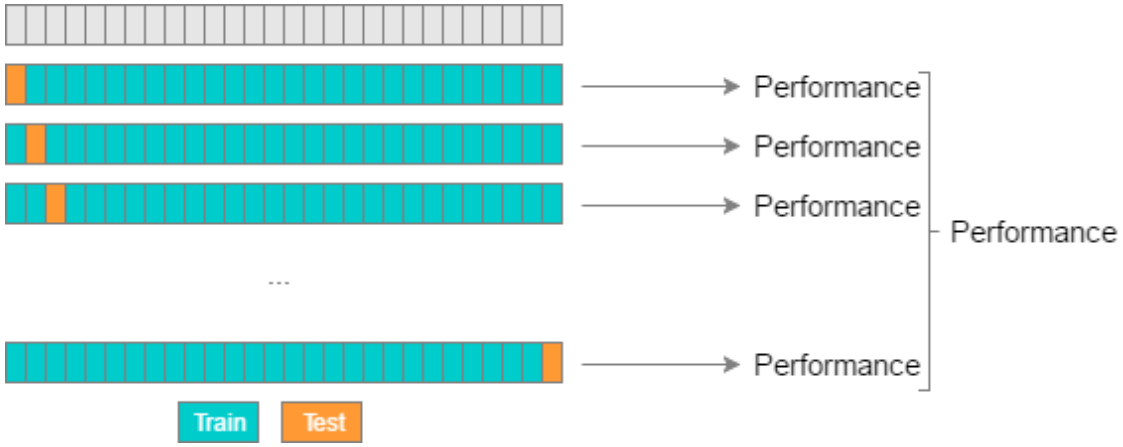
Network Settings :

- **Regions crop : Dlib library , 68 landmarks**
- **Different regions, different labels**
- CNNs trained with 64 batch size & 100 epochs
- LSTM+FCL Network trained with 244 batch size and 60 epoch
- Ubuntu 18.04.2 LTS, python3.6, keras-gpu2.2.4, tensorflow-gpu1.12.0, Geforce GTX 1080Ti GPU (32 GB memory) and Intel Xeon Processor



Training Protocol :

LOSO-CV Protocole



UF1 score

$$F1_c = \frac{2TP_c}{2TP_c + FP_c + FN_c},$$

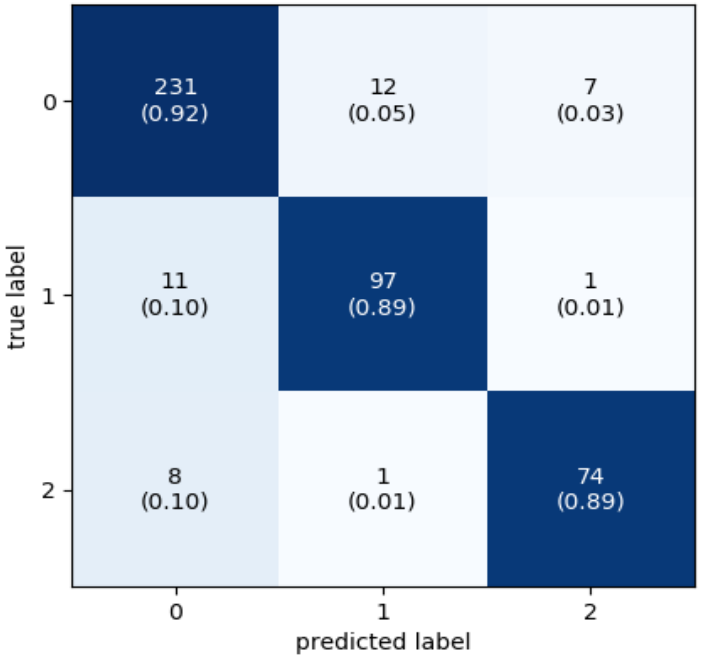
$$UF1 = \frac{F1_c}{C},$$

UAR score

$$UAR = \frac{1}{C}\sum_{c=1}^{C} ACC_c$$

$$ACC_c = \frac{TP_c}{N_c}$$

Results :



Confusion Matrix , FULL Database (CASMEII, SAMM, SMIc) ,MEGC 2019 Conditions, (0 : Negative, 1:Positive, 2 : Surprise)

Accuracy : 0.9095

UAR : 0.9018

UF1 : 0.9022

MEGC 2019

Comparison :

Models	FULL		SMIC		CASAME II		SAMM	
	UF1	UAR	UF1	UAR	UF1	UAR	UF1	UAR
LBP-TOP [22]$^{\diamond}$	0.5882	0.5785	0.2000	0.5280	0.7026	0.7429	0.3954	0.4102
Bi-WOOF [5]$^{\diamond}$	0.6296	0.6227	0.5727	0.5829	0.7805	0.8026	0.5211	0.5139
OFF-ApexNet [7]†	0.7196	0.7096	0.6817	0.6695	0.8764	0.8681	0.5409	0.5392
Micro-Attention [13]$^{\oplus}$	0.5080	0.4930	0.4730	0.4660	0.5390	0.5170	0.4030	0.3400
ATNet *(Fusion)* [26]$^{\oplus}$	0.6310	0.6130	0.5530	0.5430	0.7980	0.7750	0.4960	0.4820
Quang *et al.* [12]$^{*\oplus}$	0.6520	0.6506	0.5820	0.5877	0.7068	0.7018	0.5882	0.5989
Zhou *et al.* [27]*†	0.7322	0.7278	0.6645	0.6726	0.8621	0.8560	0.5868	0.5663
Liong *et al.* [8]*†	0.7353	0.7605	0.6801	0.7013	0.8382	0.8686	0.6588	0.6810
Liu e*et al.* [28]*†	0.7885	0.7824	0.7461	0.7530	0.8293	0.8209	0.7754	0.7152
Our proposed method $^{\oplus}$	**0.9022**	**0.9018**	**0.8886**	**0.8828**	**0.9857**	**0.9857**	**0.7855**	**0.8103**

$^{\diamond}$ handcrafted approach, † hybrid approach, $^{\oplus}$ deep learning approach.

Outline

I. Introduction

II. State-of-the-art

III. Proposed solution

IV. Experiments and results

V. Conclusion and perspectives

I. Conclusion
II. Perspectives

- Less is More
- Deep Learning : CNN (inception block) + LSTM

Accuracy :
More than 90 %

- Data Augmentation
- Adaptive analysis of MiE for Medical use case
- Complexe Micro-Expressions

A spatiotemporal deep learning solution for Automatic Micro-Expressions Recognition From Local Facial Regions