1298

Knowledge Augmented Deep Neural Networks for Joint Facial Expression and Action Unit Recognition

Zijun Cui(presenter)
Rensselaer Polytechnic Institute
cuiz3@rpi.edu

Tengfei Song
Southeast University
songtf@seu.edu.cn

Yuru Wang
Northeast Normal University
wangyr@nenu.edu.cn

Qiang Ji
Rensselaer Polytechnic Institute
qji@ecse.rpi.edu

OVERVIEW

- ***Tasks***:
  - Facial Expression Recognition(FER)
  - Action Unit(AU) Detection

- ***Motivations***:
  - Facial expression and AUs are strongly correlated
  - Generic knowledge on expression-AUs relationships is available

- ***Contributions***:
  - A *knowledge model* encoding the generic knowledge systematically
  - A deep learning framework for *joint* facial expression and AU recognition

GENERIC KNOWLEDGE

-- on expression-AUs probabilistic relationships

- *Expression-dependent single AU probabilities*
  - AU4 is a primary AU given Anger expression

$$p(\mathrm{AU4} = \mathrm{ON} | \mathrm{Anger}) > p(\mathrm{AU4} = \mathrm{OFF} | \mathrm{Anger})$$

- *Expression-dependent joint AU probabilities*
  - AU6 and AU12 are positively correlated given Happy expression

$$p(\mathrm{AU6} = \mathrm{ON} | \mathrm{AU12} = \mathrm{ON}, \mathrm{Happy}) > p(\mathrm{AU6} = \mathrm{OFF} | \mathrm{AU12} = 1, \mathrm{Happy})$$
$$p(\mathrm{AU6} = \mathrm{ON} | \mathrm{AU12} = \mathrm{ON}, \mathrm{Happy}) > p(\mathrm{AU6} = \mathrm{ON} | \mathrm{AU12} = \mathrm{OFF}, \mathrm{Happy})$$

- *Expression-independent joint AU probabilities*
  - AU1 and AU2 are negatively correlated

$$p(\mathrm{AU1} = \mathrm{ON} | \mathrm{AU2} = \mathrm{ON}) > p(\mathrm{AU1} = \mathrm{OFF} | \mathrm{AU2} = \mathrm{ON})$$
$$p(\mathrm{AU1} = \mathrm{ON} | \mathrm{AU2} = \mathrm{ON}) > p(\mathrm{AU1} = \mathrm{ON} | \mathrm{AU2} = \mathrm{OFF})$$

ENCODING OF GENERIC KNOWLEDGE

-- Bayesian Network(BN) Learning with Probability Constraints

❑ Regression Bayesian Network

$$p(X_i = k|\pi(X_i)) = \sigma_M(\sum_{j=1}^{J} w_{ijk}\pi_j(X_i) + b_{ik})$$

where weights $\boldsymbol{w} = \{w_{ijk}\}$ and bias $\boldsymbol{b} = \{b_{ik}\}$ are to be learned. And $A(\boldsymbol{w})$ defines the structure.

- The constraint of Directed Acyclic Graph(DAG): $\mathrm{tr}\left(e^{A(\boldsymbol{w})\circ A(\boldsymbol{w})}\right) = 0$

❑ A penalty function $f(\boldsymbol{w}, \boldsymbol{b}; \boldsymbol{s})$ measures the violation of constraints given weights $\boldsymbol{w}$, bias $\boldsymbol{b}$ and margins $\boldsymbol{s}$

❑ A *Constraint Optimization Approach* for BN learning

$$\boldsymbol{w}^*, \boldsymbol{b}^*, \boldsymbol{s}^* = \arg\min_{w,b,s} f(\boldsymbol{w}, \boldsymbol{b}; \boldsymbol{s}) + \gamma||\boldsymbol{w}||_1 - \mu||\boldsymbol{s}||_2^2$$
$$\text{s. t. } \mathrm{tr}\left(e^{A(\boldsymbol{w})\circ A(\boldsymbol{w})}\right) = 0$$

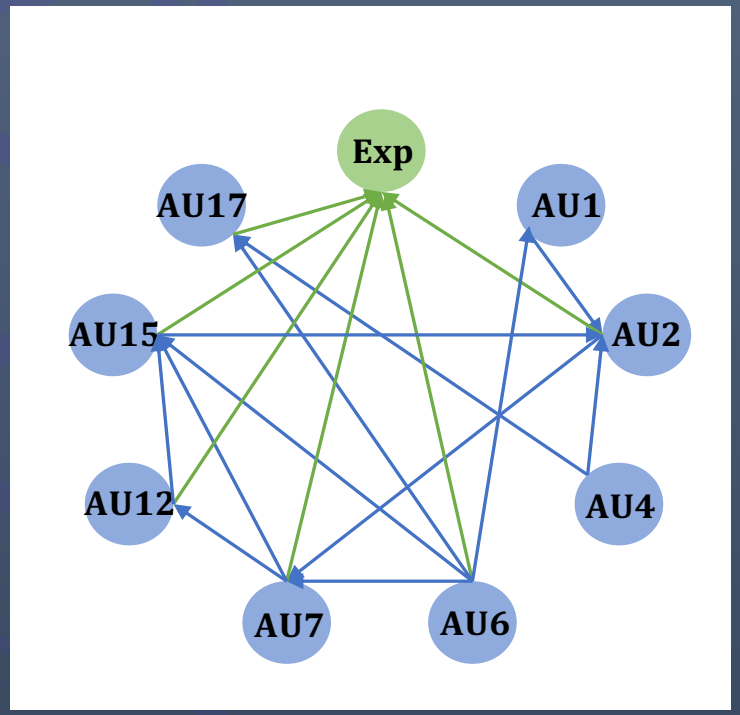
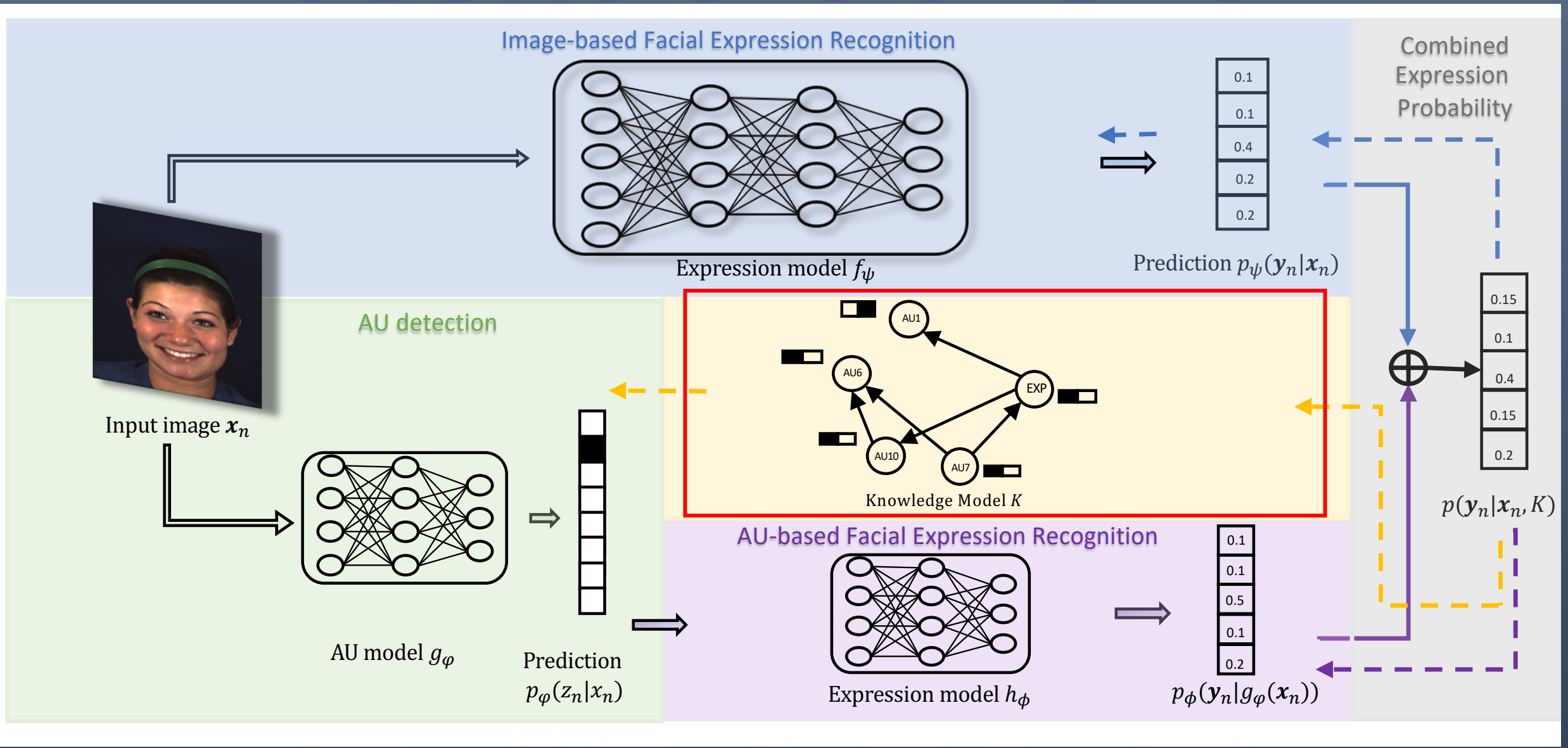


Figure: the learned structure given constraints

The Joint Learning Framework



EXPERIMENTS

-- compare to SOAs

❑ Action Units detection

Table 6: Comparison to the SoAs on AU detection.

Supervision	Method	BP4D	CK+	MMI
Supervised	HRBM[47]	.67	.79	.56
	MC-LVM[8]	-	**.80***	-
	JPML[56]	**.68***	.78*	-
	AU R-CNN[30]	.63*	-	-
Weakly-supervised	HTL[40]	.50	.66	.42
	LP-SM[54]	.55	.72*	.50
	TCAE[22]	.56*	-	-
	AUD-BN(baseline)	.56	.69	.47
	AUD-EA(gBN)	**.57**	**.74**	**.58**

❑ Facial Expression Recognition

Table 8: Comparison with SoA FER methods

Methods	BP4D	CK+	MMI	EmotioNet
STM-Explet[27]	-	94.19*	75.12*	-
DTAGN(Joint)[12]	-	97.25*	70.24*	-
DeRL[50]	-	97.30*	73.23*	-
ILCNN[3]	-	94.35*	70.67*	-
DAM-CNN[49]	-	95.88*	-	-
FMPN-FER[4]	60.16	96.53	82.74*	84.88
DeepEmotion[32]	79.54	95.23	72.66	81.51
FER-I(baseline)	61.68	94.29	67.35	80.85
FER-IK(gBN)	**83.82**	**97.59**	**84.90**	**95.55**

Thank you!