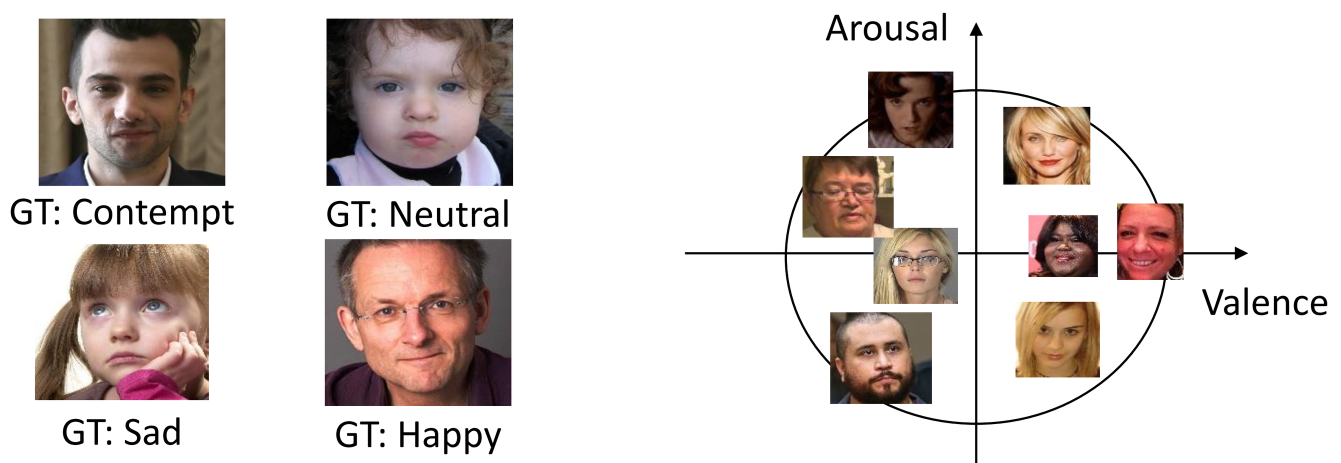
Contrastive Adversarial Learning for Person Independent Facial Emotion Recognition

Dae Ha Kim, Byung Cheol Song
Computer Vision and Image Processing Lab. (CVIP)
Inha University, Republic of Korea

Facial emotion recognition (FER)

- Definition: To grasp a person's emotions using various facial factors such as eyes, mouth, action unit (AC), etc.
- Two approaches: Discrete domain FER & continuous domain FER



Discrete domain FER **Continuous domain FER**

Limitation and our approach

- Limitation: One-to-one mapping btw. Input and label supervision
  - Tend to be biased towards given data → Person-dependent learning
- Solution: Generative network, i.e., generative adversarial network (GAN)
  - Consider label supervision as well as <u>latent features</u>
  - Two inputs for GAN are defined by <u>emotion grouping</u>

Theoretical analysis

- Contrastive adversarial loss using φ-divergence
  - Induce adversarial structure between $\mathbb{P}$ and $\mathbb{Q}$
  - Guarantee generalization bound

$$d_\varphi(\mathbb{P}||\mathbb{Q}) = \int_{\mathcal{Z}} \mathbb{Q}(\mathbf{z})\varphi\left(\frac{\mathbb{P}(\mathbf{z})}{\mathbb{Q}(\mathbf{z})}\right)d\mathbf{z}$$
$$\geq \sup_{f\in\mathcal{F}} \mathbb{E}_{\mathbf{z}\sim\mathbb{P}} f(\mathbf{z}) - \mathbb{E}_{\mathbf{z}\sim\mathbb{Q}}\varphi^*\big(f(\mathbf{z})\big) \quad \textbf{[1]}$$
$$\geq \sup_{f\in\mathcal{F}} \mathbb{E}_{\mathbf{z}\sim\mathbb{P}} f(\mathbf{z}) - \frac{1}{2}\mathbb{E}_{\mathbf{z}\sim(\mathbb{P}+\mathbb{Q})} f(\mathbf{z}) - \frac{1}{4}\mathbb{E}_{\mathbf{z}\sim\frac{\mathbb{P}+\mathbb{Q}}{2}} f^2(\mathbf{z}) \quad \textbf{[2]}$$

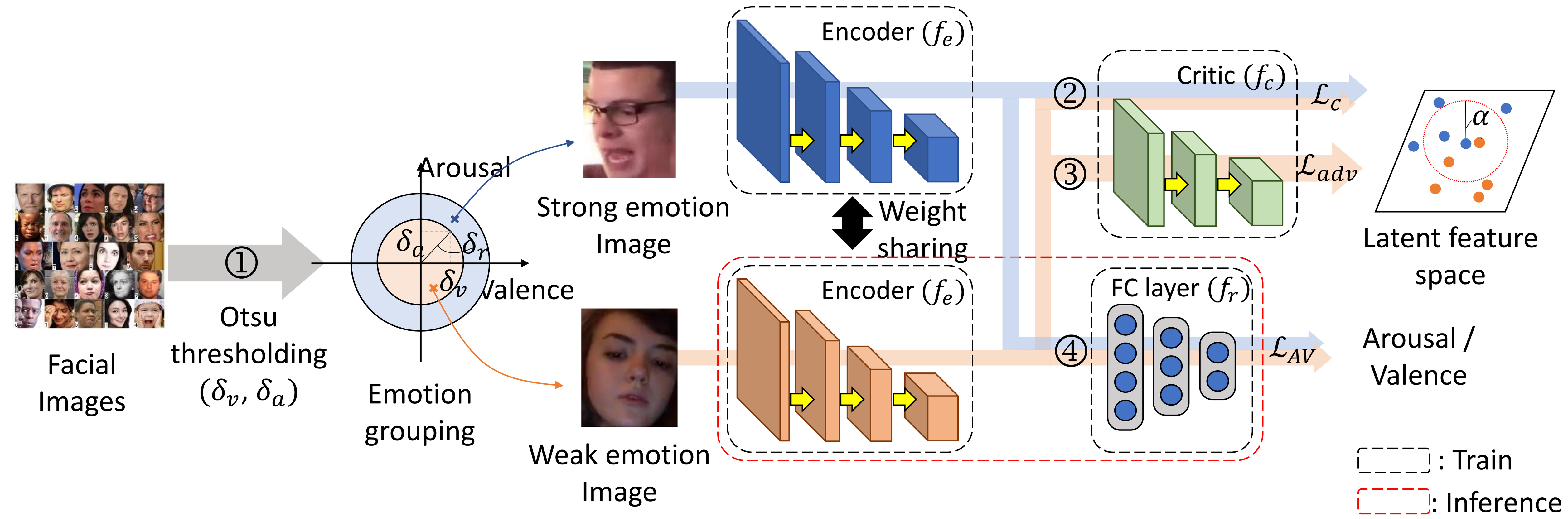
where $\varphi: \mathbb{R}^+ \to \mathbb{R}$ is a convex, lower semi-continuous function satisfying $\varphi(1) = 0$ and φ^* is Fenchel conjugate of φ.

- If f is set to pointwise hinge function as f_{cont}, then pointwise metric learning can be performed.

$$f_{cont} = -\xi_{i,j}\max\{0, \xi_{i,j}\left(D^2(\mathbf{z}_i, \mathbf{z}_j) - \alpha\right)\}$$

Overall framework

① Emotion grouping via Otsu thresholding
② Discriminative learning of critic network
③ Adversarial learning of encoder network
④ AV emotion learning of FC layer



Experiments

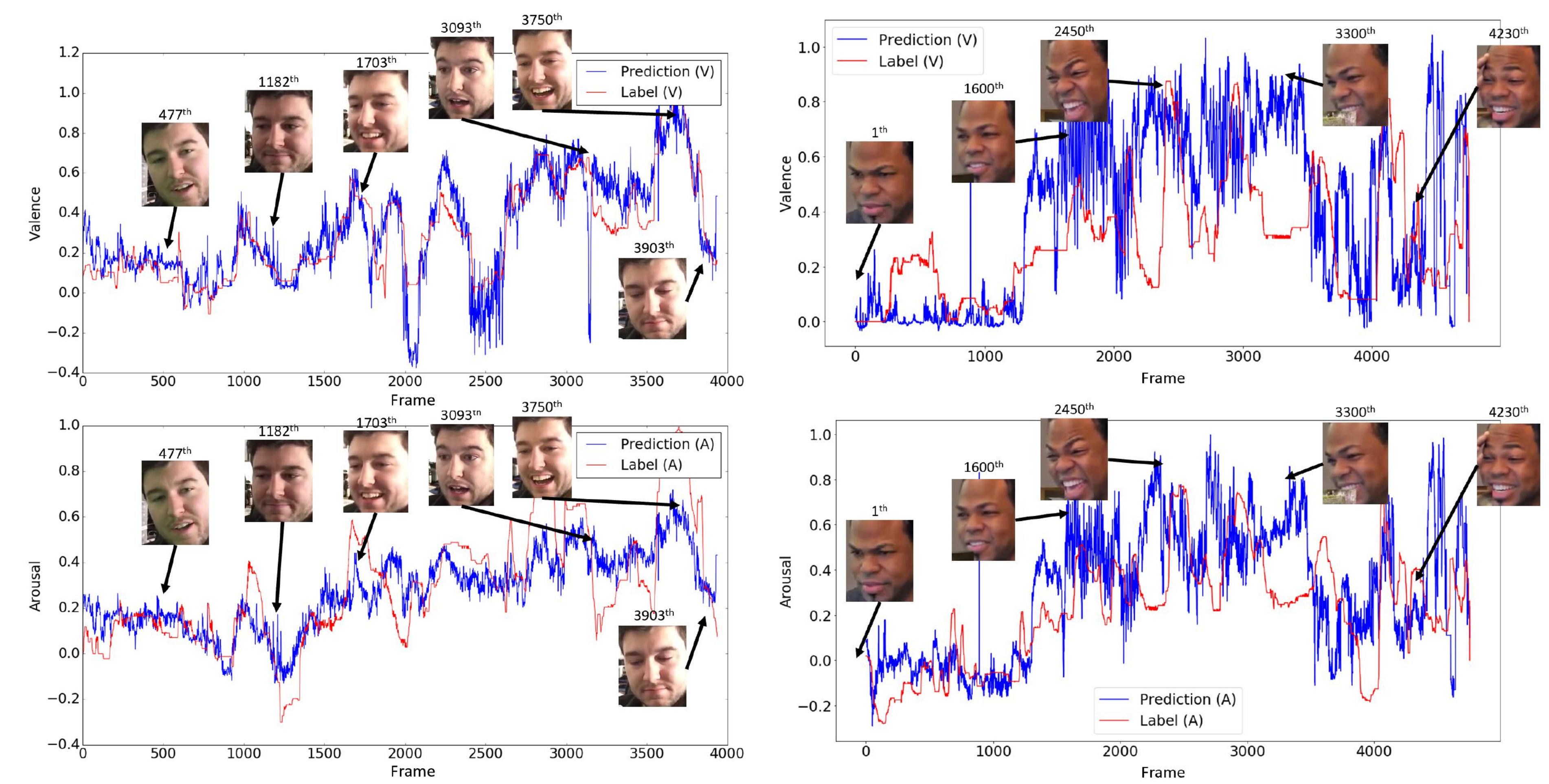
Quantitative results on AffectNet dataset

Methods	Backbone	Params.	RMSE		PCC		CCC	
			(V)	(A)	(V)	(A)	(V)	(A)
(Mollahosseini, Hasani, and Mahoor 2017)	AlexNet	61M	0.37	0.41	0.66	0.54	0.60	0.34
(Jang, Gunes, and Patras 2019)	SSD w/ VGG16	-	0.44	0.39	0.58	0.50	0.57	0.47
(Kollias et al. 2018)	VGG16	-	0.37	0.39	0.66	0.55	0.62	0.54
(Barros, Parisi, and Wermter 2019)	AlexNet	-	-	-	-	-	0.67	0.38
(Kossaifi et al. 2020)	ResNet18	-	0.35	0.32	0.71	0.63	0.71	0.63
(Hasani, Negi, and Mahoor 2020)	ResNeXt50	3.1M	0.2668	0.2482	0.78	0.86	0.74	0.85
Ours	ResNet18	11M	**0.2186**	**0.1873**	**0.86**	0.85	**0.83**	0.84
	AlexNet (tuned)	3.6M	0.2216	0.1916	0.81	**0.86**	0.80	**0.85**

Efficiency of adaptive margin using DML techniques

Methods	CUB200-2011				Cars196			
	R@1	R@2	R@4	R@8	R@1	R@2	R@4	R@8
Constrastive (Hadsell, Chopra, and LeCun 2006)	55.2	68.2	77.9	85.0	64.2	74.7	82.2	88.4
Contrastive w/ $\widehat{\alpha}$	**57.1**	**68.9**	**78.4**	**85.6**	**71.7**	**81.5**	**88.4**	**93.2**
Trip-semi (Schroff, Kalenichenko, and Philbin 2015)	57.5	68.8	78.3	85.4	65.5	76.9	85.2	90.4
Trip-semi w/ $\widehat{\alpha}$	**60.2**	**71.5**	**80.0**	**87.4**	**74.0**	**83.3**	**89.1**	**93.5**
Margin (Wu et al. 2017)	63.6	74.4	83.1	90.0	79.6	86.5	91.9	95.1
Margin w/ $\widehat{\alpha}$	**63.9**	74.4	**83.7**	**90.3**	**80.1**	**88.6**	**92.3**	**95.4**
DSML (Tri) (Yuan et al. 2019)	63.8	74.6	83.4	90.4	80.9	88.5	92.6	95.9
DSML (Tri) $\widehat{\alpha}$	**63.9**	**74.7**	**83.7**	**90.5**	**81.1**	**88.6**	92.6	95.9
Proxy-Anchor (Kim et al. 2020)	68.4	79.2	86.8	91.6	86.1	91.7	95.0	97.3
Proxy-Anchor $\widehat{\alpha}$	**69.7**	**79.8**	**87.0**	**92.1**	**87.4**	**92.0**	**95.2**	**97.4**

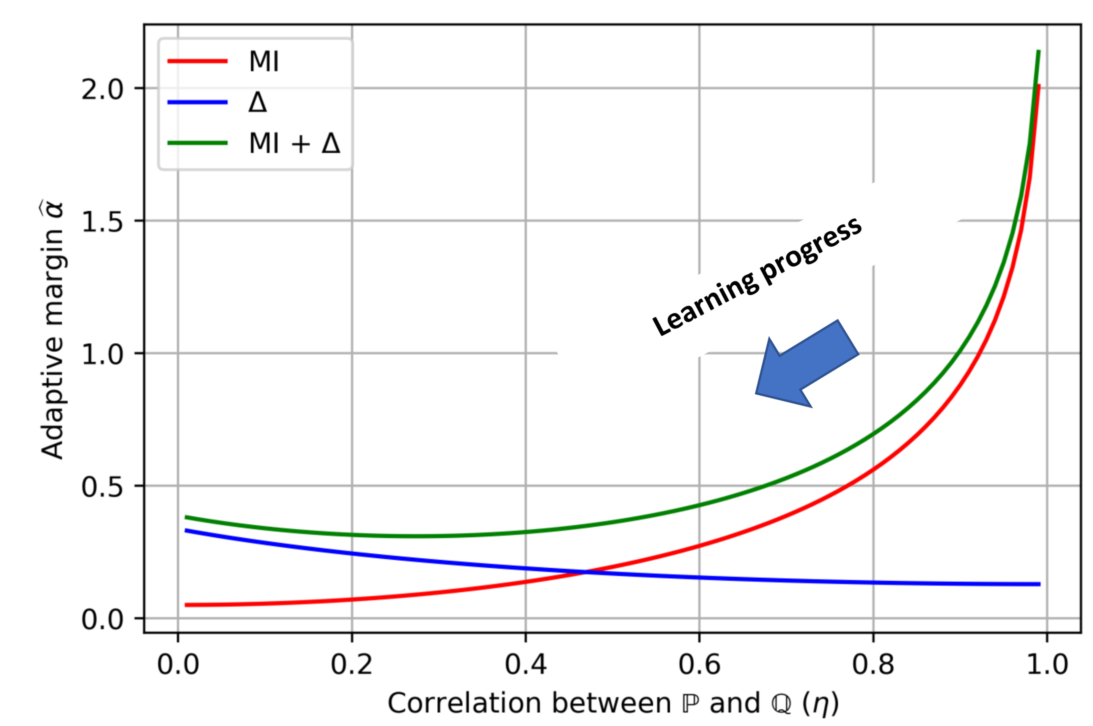
Qualitative results on Aff-Wild dataset



- Adaptive margin
  - Problem of static margin (α): overfitting phenomenon (chronic problem of DML)
  - Solution: using mutual information (MI) and confidence interval (Δ) **[3]**

$$\hat{\alpha} = \underbrace{-\frac{1}{2}\log(1-\eta^2)}_{\text{Mutual information}} + \underbrace{\log\left(\frac{N}{c}\right)^{\frac{1}{2K}}}_{\text{Confidence interval}} = \log(1-\eta^2)^{-\frac{1}{2}}\left(\frac{N}{c}\right)^{\frac{1}{2K}}$$

  - η indicates the correlation coefficient of inputs of MI
  - c, K, and N are hyper-parameters and batch size, respectively
- Lift the lower bound of margin value $\hat{\alpha}$ when MI is close to 0.



Code & Demo

Recommended papers

[1] Nguyen, X. et al. (2009). On surrogate loss functions and f-divergences. *The Annals of Statistics*, *37*(2), 876-904.
[2] Mroueh, Y., and Sercu, T. (2017). Fisher gan. In *Advances in Neural Information Processing Systems* (pp. 2513-2523).
[3] Balsubramani, A. et al. (2019). An adaptive nearest neighbor rule for classification. In *NeurIPS* (pp. 7579-7588).