

Photo-Realistic Facial Emotion Synthesis Using Multi-level Critic Networks with Multi-level Generative Model

Minho Park, Hak Gu Kim and Yong Man Ro

Image and Video Systems (IVY) Lab.,

<http://ivylab.kaist.ac.kr>

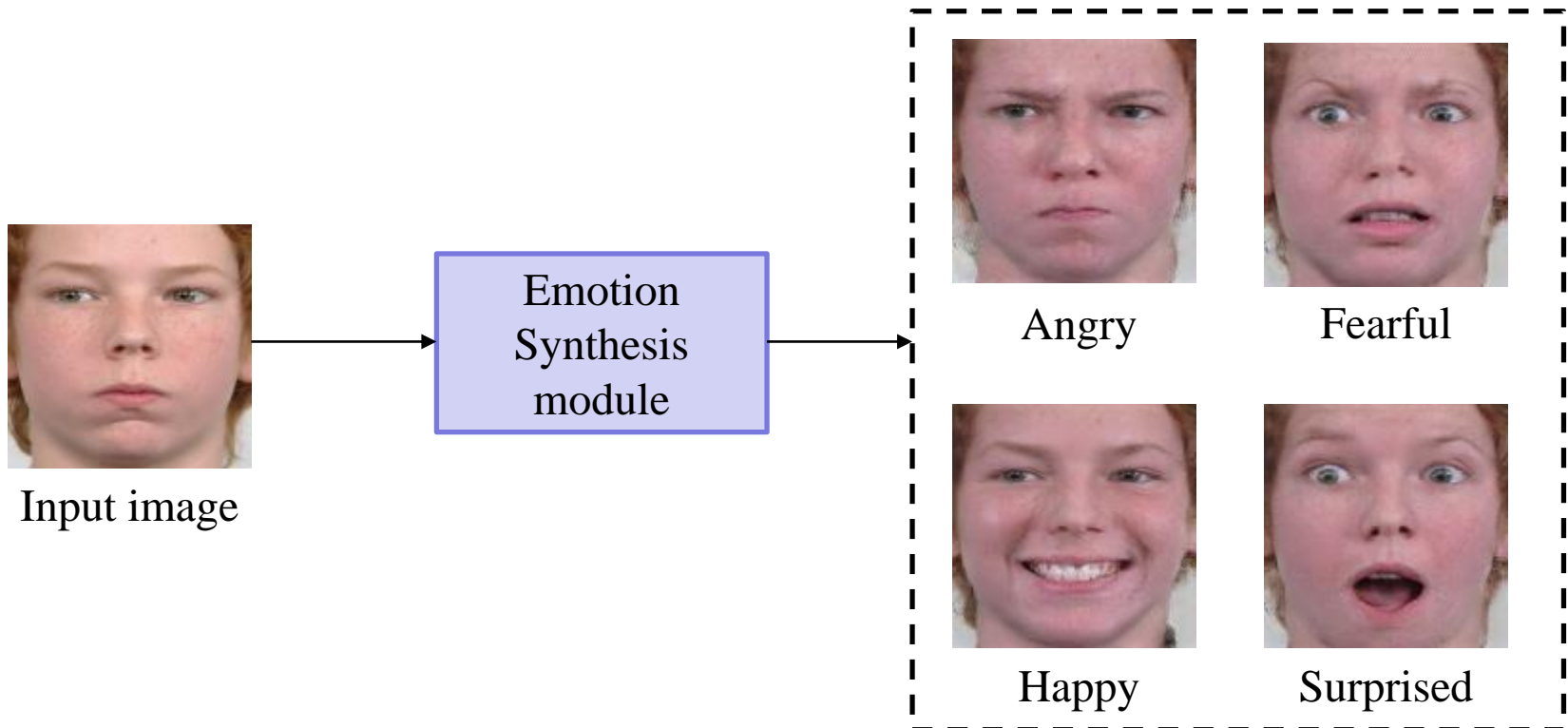
School of Electrical Engineering,

Korea Advanced Institute of Science and Technology (KAIST)

What is Facial Emotion Synthesis?

❖ Facial emotion synthesis

- Maintaining identity information + Changing facial expression

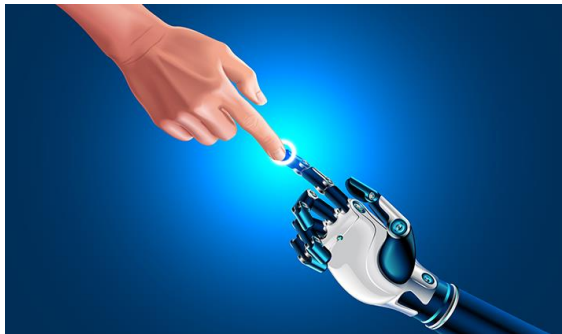


What is Facial Emotion Synthesis?

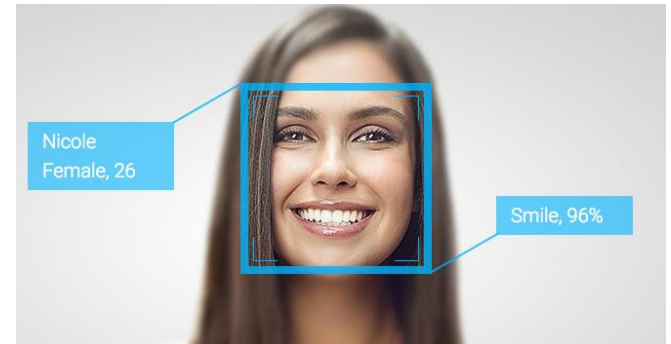
❖ Emotion synthesis related applications



Facial Animation



Human-Computer Interaction



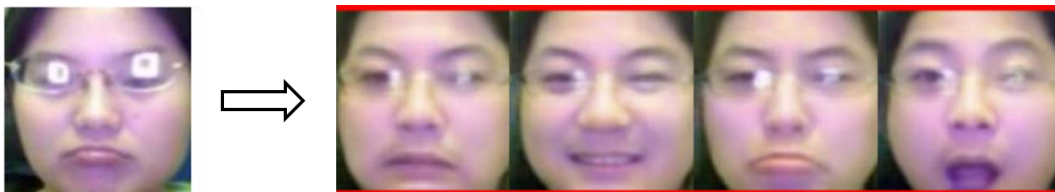
Facial Expression Recognition

Previous Works

❖ Previous works for generating facial expression

- ExprGAN^[1]

- ✓ Facial expression synthesis model using expression classifier



- StarGAN^[2]

- ✓ Facial image generating model with several attribute



The **variation** of expression is **large**



The expression **can't** be generated well

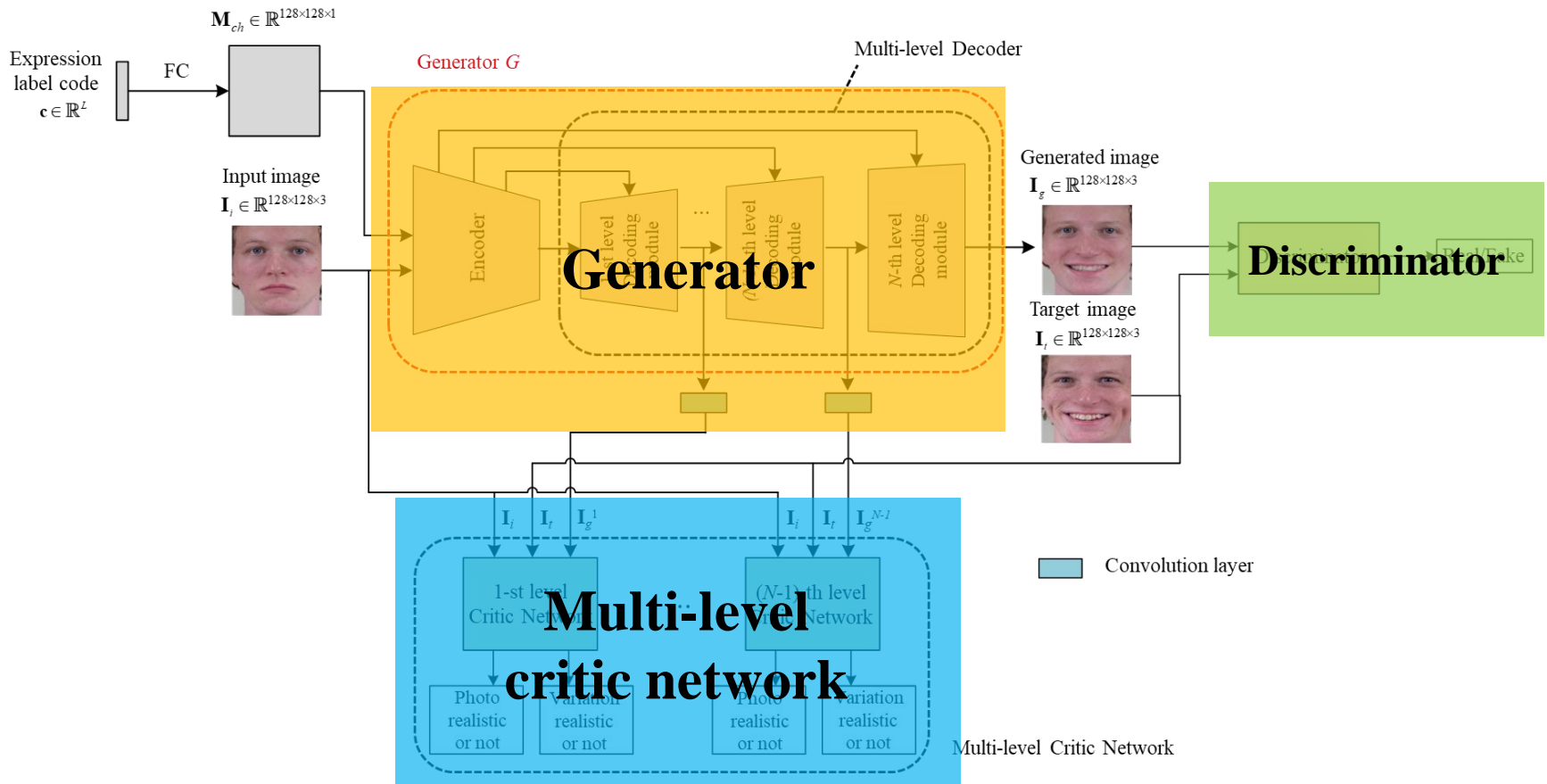
[1] Ding, H., Sricharan, K., Chellappa, R.: Exprgan: Facial expression editing with controllable expression intensity. In: The Association for the Advance of Artificial Intelligence (AAAI). (2018)

[2] Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: unified generative adversarial networks for multi-domain image-to-image translation. arXiv preprint arXiv: 1711.09020

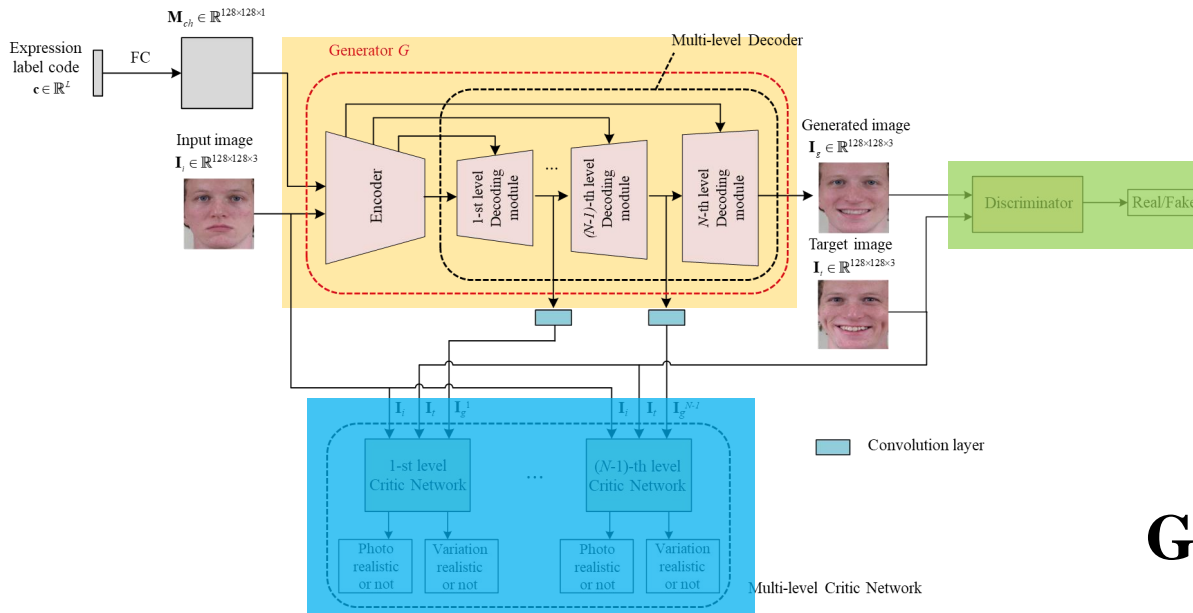
Contributions

- ❖ We proposed a novel **multi-level critic network** for generating photo-realistic facial image with desired expression which has **large variation**.
- ❖ With qualitative and quantitative evaluations, we show the effectiveness of the **enhanced feature** guided by the proposed **multi-level critic network**.

Overall proposed framework



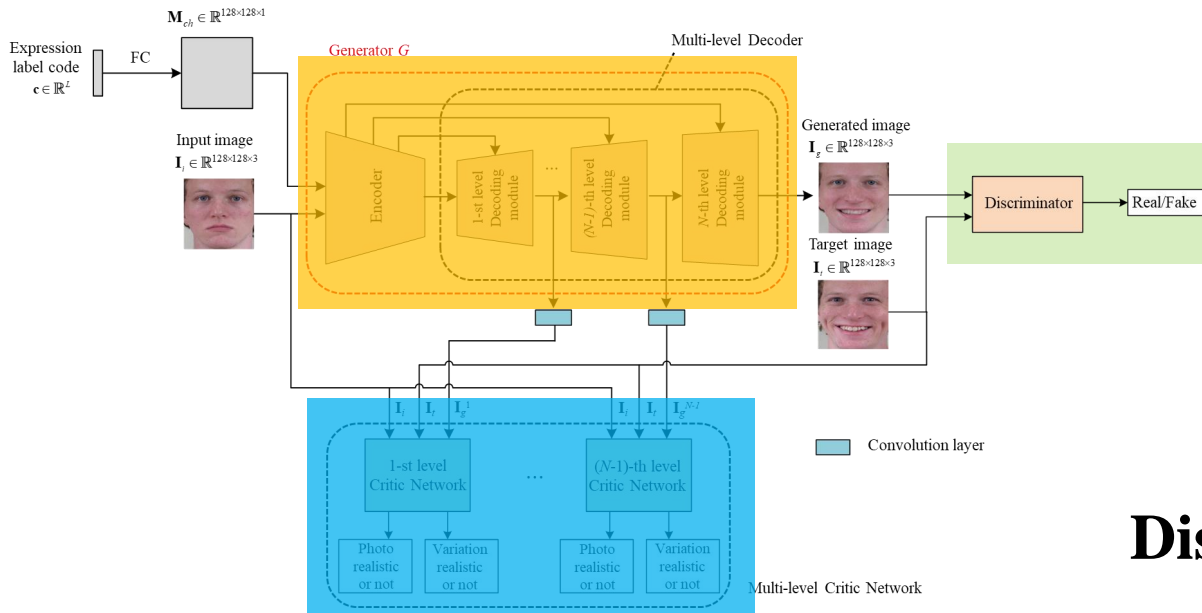
Overall proposed framework



Generator

Encoder, Multi-level decoder
 Convolutional neural network
 Generating facial image with desired expression

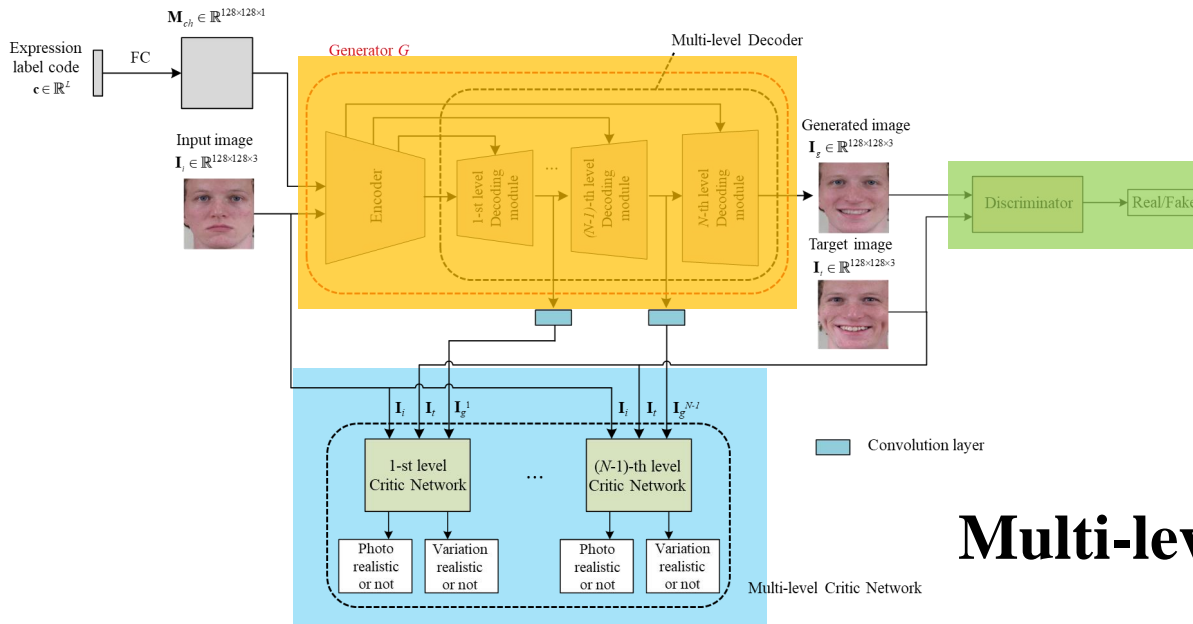
Overall proposed framework



Discriminator

Convolutional neural network
Distinguishing between generated image and
ground-truth target image

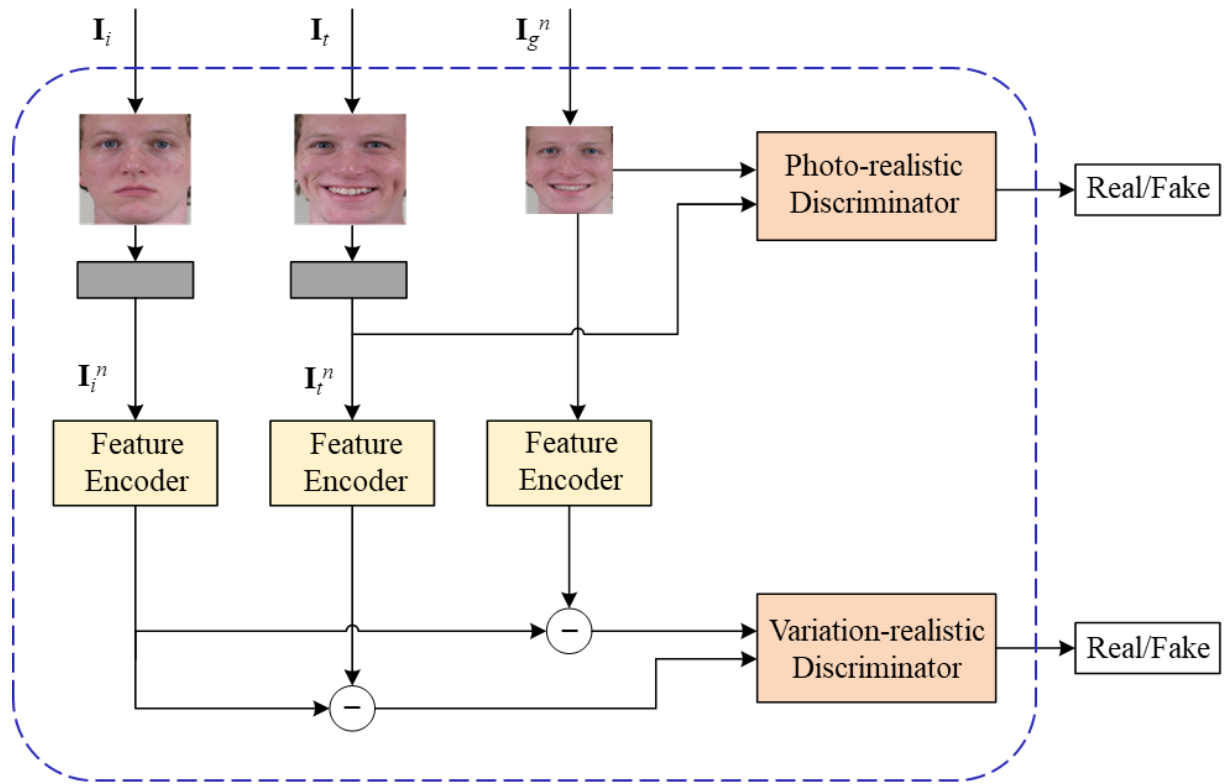
Overall proposed framework



Multi-level Critic Network

Photo-realistic discriminator,
 Variation-realistic discriminator
 Determining whether the low-resolution
 generated image is real or not
 Refining inter-features of the decoder

n -th level Critic Network



n -th Multi-level Critic Network

 Down sampling

n -th level Critic Network

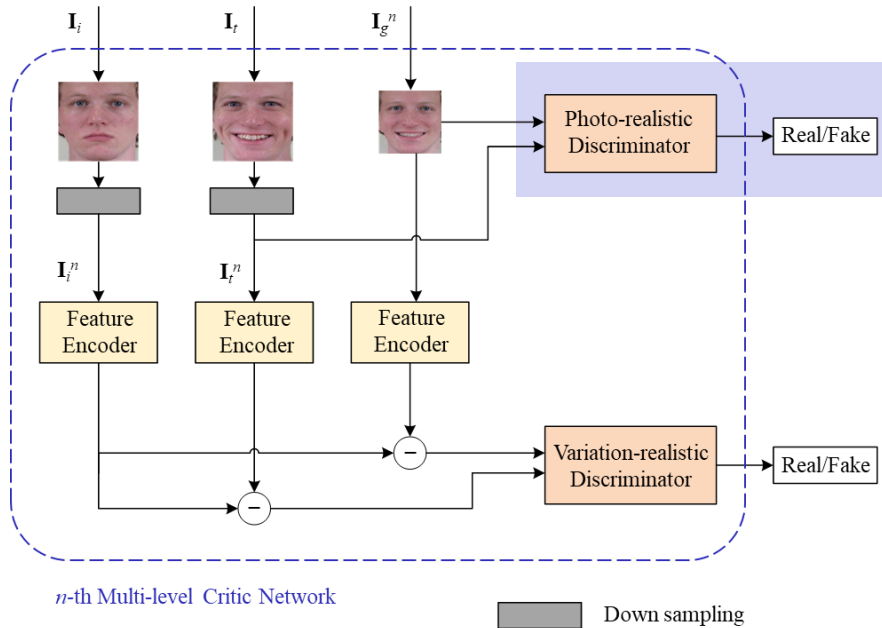
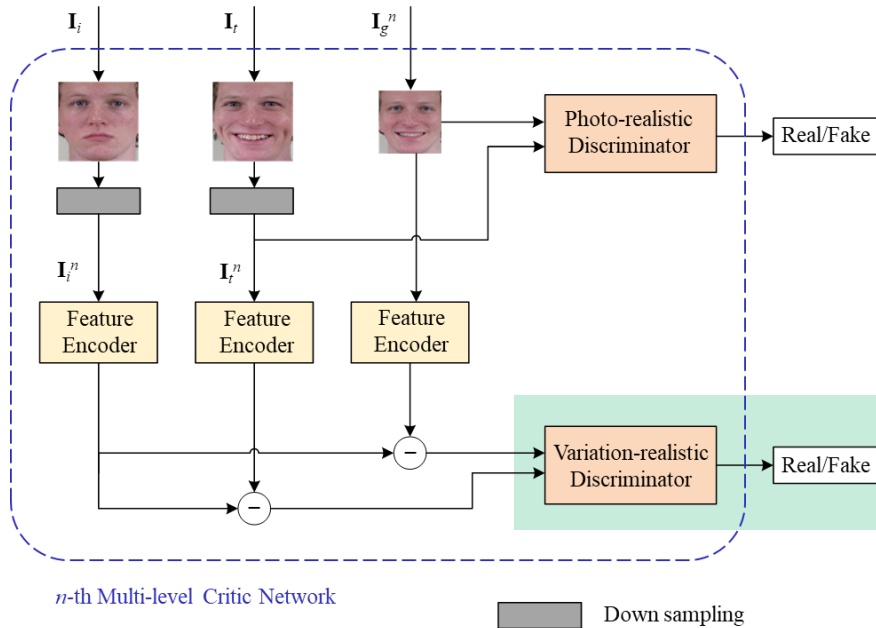


Photo-realistic Discriminator

Convolutional neural network
Distinguishing between generated image
and target image

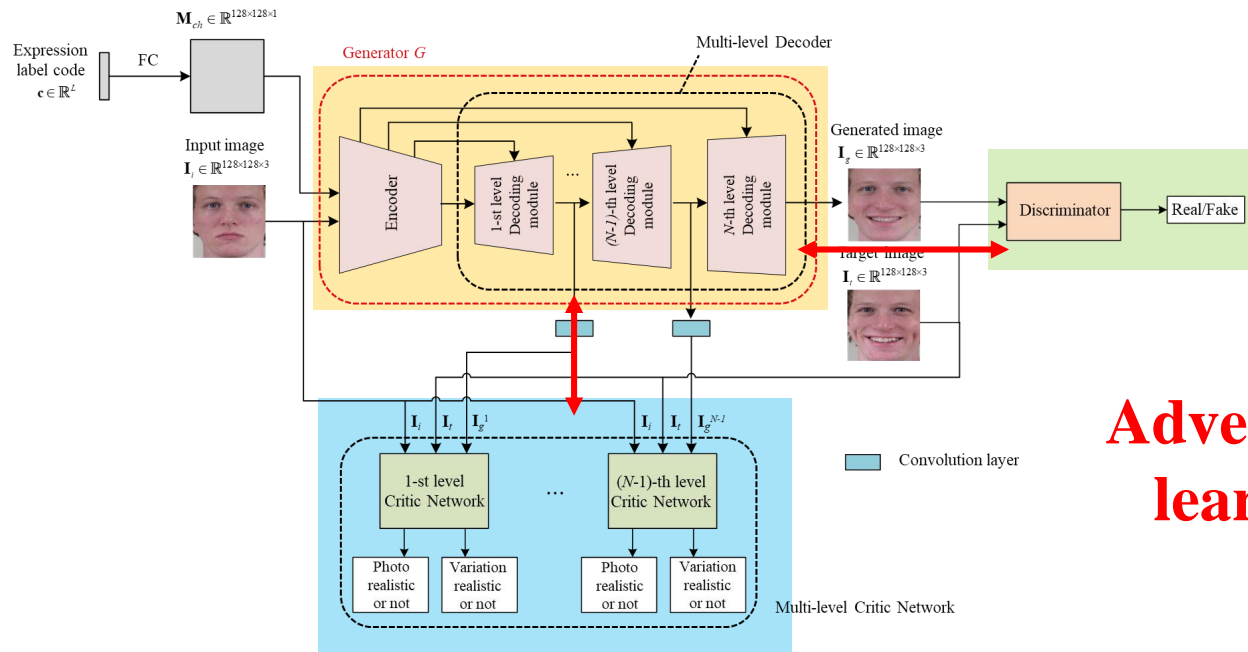
n -th level Critic Network



Variation-realistic Discriminator

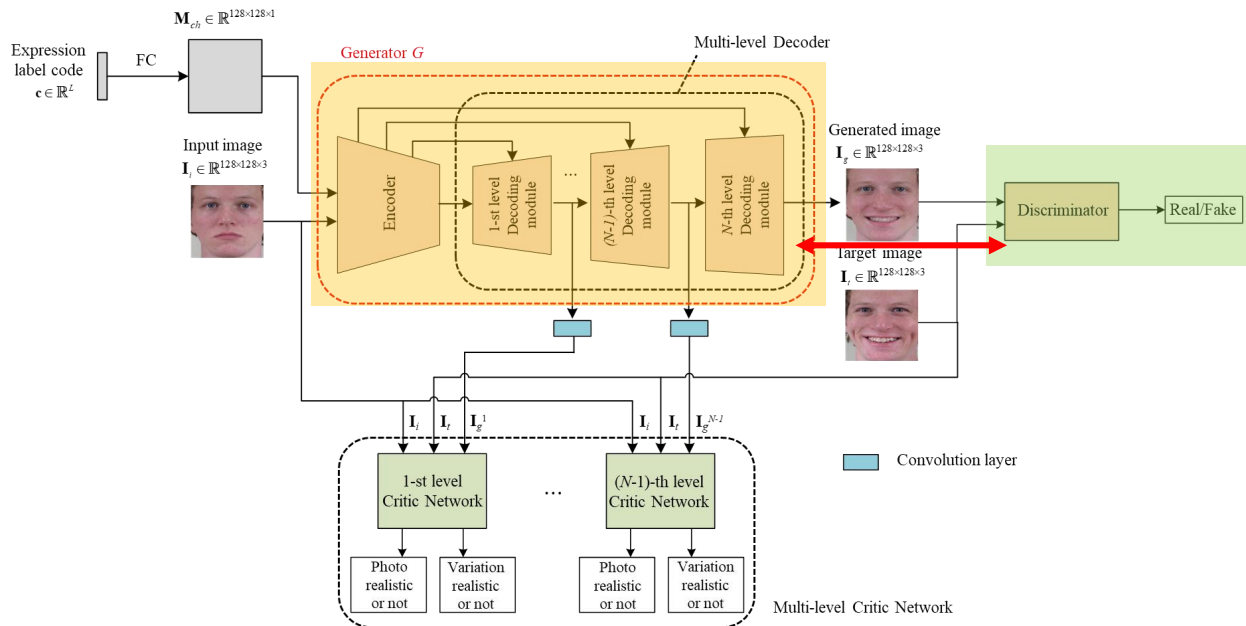
Encoding three images and
subtracting each other
Convolutional neural network
Determining whether the extracted feature is
real or not

Training



Adversarial learning

Objective function



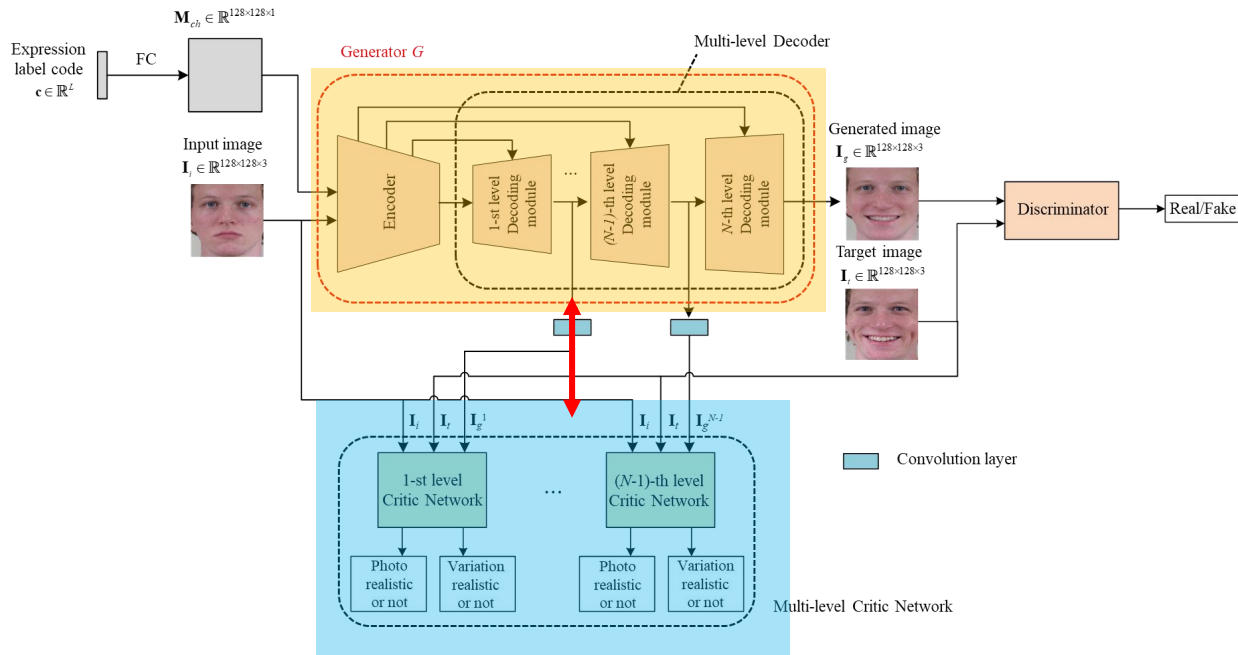
Discriminator loss :

$$L_D = E_{x \sim P_{data}} [\log D(x)] + E_{x \sim P_{data}} [\log (1 - D(G(x, c)))]$$

Generator loss :

$$L_G = E_{x \sim P_{data}} [\log (D(G(x, c)))]$$

Objective function

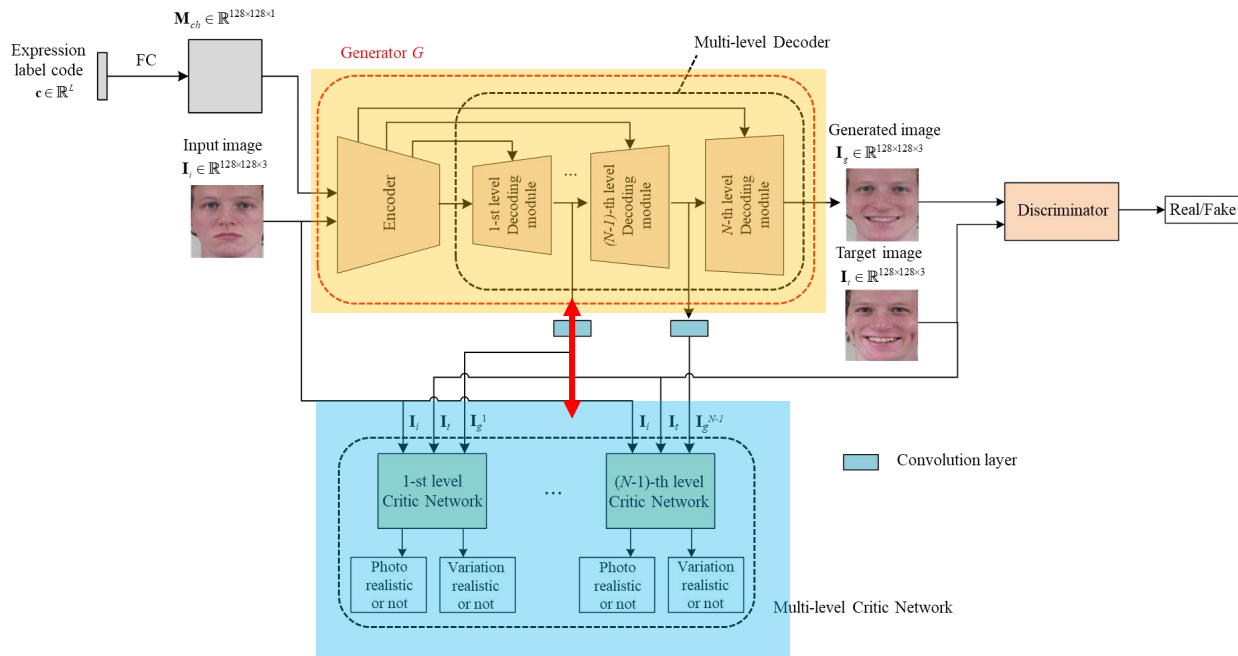


Multi-level critic loss

Photo-realistic :
$$L_P = E_{x \sim P_{data}} [\log D(x)] + E_{x \sim P_{data}} \left[\log \left(1 - D(G(x, c)) \right) \right]$$

Variation-realistic :
$$L_V = E_{x \sim P_{data}} [\log D(f(x_t) - f(x))] + E_{x \sim P_{data}} \left[\log \left(1 - D(f(G(x, c)) - f(x)) \right) \right]$$

Objective function



Generator loss

$$L_G = \sum E_{x \sim P_{data}} \left[\log \left(D(G(x, c)) \right) \right] + \sum E_{x \sim P_{data}} \left[\log \left(D(f(G(x, c)) - f(x)) \right) \right]$$

Experiments and Results

❖ Dataset

▪ Training

- ✓ PNAS dataset^[1]: total 1,610 images, 230 human subjects and 7 expressions (neutral, anger, disgusted, fearful, happy, sad and surprised), horizontally flip for data augmentation

▪ Testing

- ✓ MMI dataset^[2]: total 30 images, 30 human subjects and neutral expression, were used.
- ✓ RaFD dataset^[3]: total 201 images, 67 human subjects, neutral expression and 3 eye direction

[1] Du, S., Tao, Y., Martinez, A.M.: Compound facial expressions of emotion. Proc. Natl. Acad. Sci. 111(15), E1454–E1462 (2014)

[2] Valstar, M., Pantic, M.: Induced disgust, happiness and surprise: an addition to the MMI facial expression database. In: Proceedings of 3rd International Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect, p. 65 (2010)

[3] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A.: Presentation and validation of the radboud faces database. Cogn. Emot. 24(8), 1377–1388 (2010)

Experiments and Results

❖ Comparison with existing models



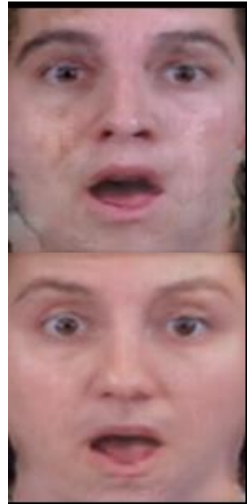
MMI Dataset

Experiments and Results

❖ Comparison with existing models



Pix2pix



StarGAN



Ours



MMI Dataset

Experiments and Results

❖ Effects of refining multi-level features



(a)



(b)



(c)



(a) : Use only 1 critic network at (32x32) features

(b) : Use only 1 critic network at (64x64) features

(c) : Use 2 critic networks at (32x32) and (64x64) features

RaFD Dataset

Experiments and Results

❖ Facial expression recognition

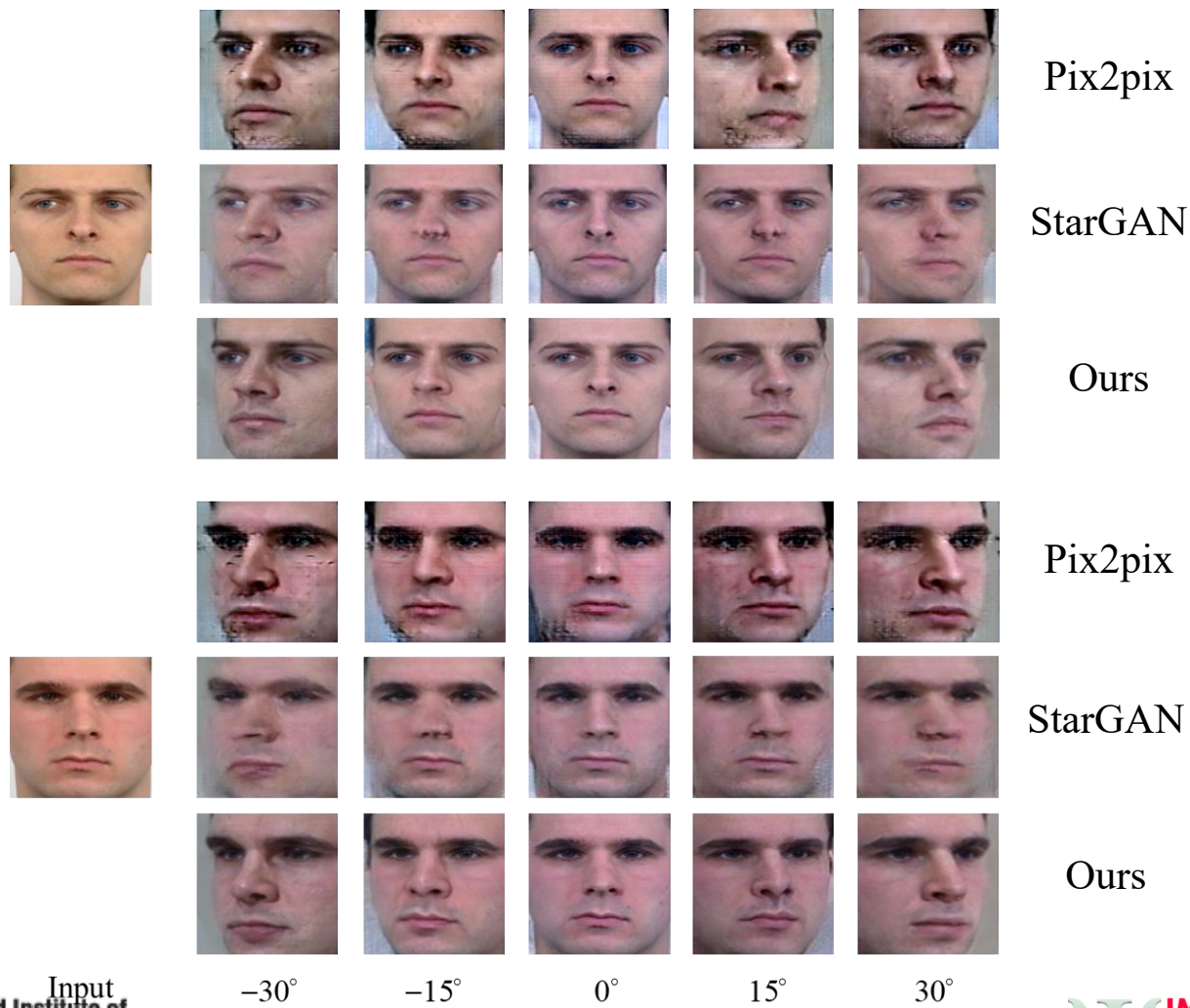
Method	Classification error (%)
Pix2pix	10.73
StarGAN	26.37
Ours	6.11
Real images (RaFD validation set)	2.86

- The AlexNet^[1] architecture was used for the classifier

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

Experiments and Results

❖ Qualitative of head-pose generation



Conclusion

- ❖ **We proposed the novel facial expression synthesis model by using novel multi-level decoder and multi-level critic network.**
- ❖ **Multi-level critic networks were trained together with the generator and the discriminator in the adversarial manner.**
- ❖ **Comparative experiments have been conducted to demonstrate the effectiveness of the proposed method.**

Thank you