# Deep Facial Animation Retargeting from Humans to 3D Stylized Characters

**Facial Expression Research Group (FERG), UW CSE**

Deepali Aneja[1], Bindita Chaudhuri[1], Linda Shapiro[1], Barbara Mones[1],
Alex Colburn[2], Gary Faigin[3]

[1]Paul G. Allen School of Computer Science and Engineering,
University of Washington, Seattle WA, USA
[2] Zillow Group, Seattle WA, USA
[3] Gage Academy of Art, Seattle WA, USA

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

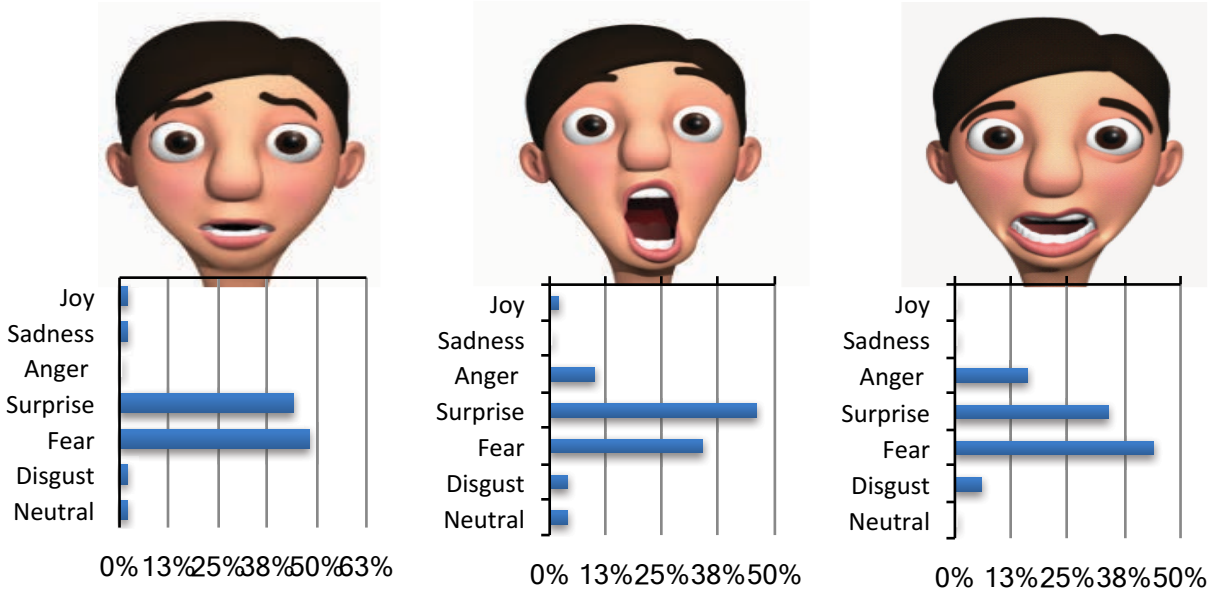
# Facial expressions : The art of non-verbal communication

Source: https://www.globalintelconsultants.com/02-13-16-facial-expressions--the-art-of-non-verbal-communication.html

# Creating recognizable expressions

- Accurate facial expression depiction is critical and **difficult** for storytelling.

- We asked professional animators to make this character look surprised. None of the expressions achieved greater than 50% recognition on Mechanical Turk.



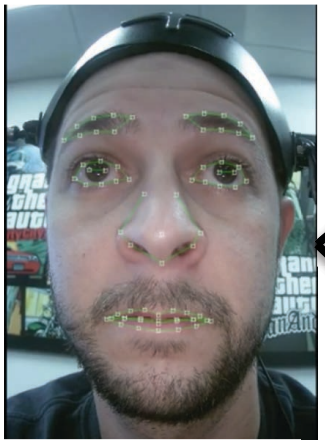
# Introduction

- **Goal:** *accurate and perceptually correct facial expression transfer*
  - Capture human facial expression from 2D images
  - Transfer the expression to multiple 3D stylized characters

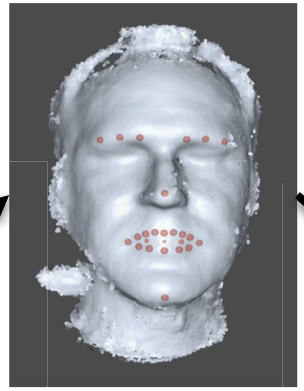
- **Related Work**:
  - Landmark mapping (manual labor intensive)
  - Blendshape interpolation (memory intensive)

- We propose a generative method for expression transfer that ensures:
  - *Perceptually validity* by using deep neural network features
  - *Geometric consistency* by using geometric features

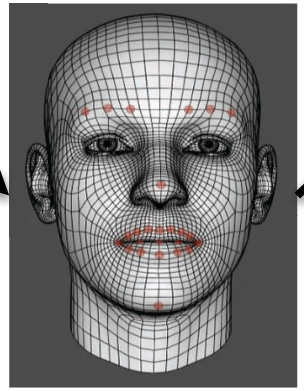
# Related work: Landmark Mapping



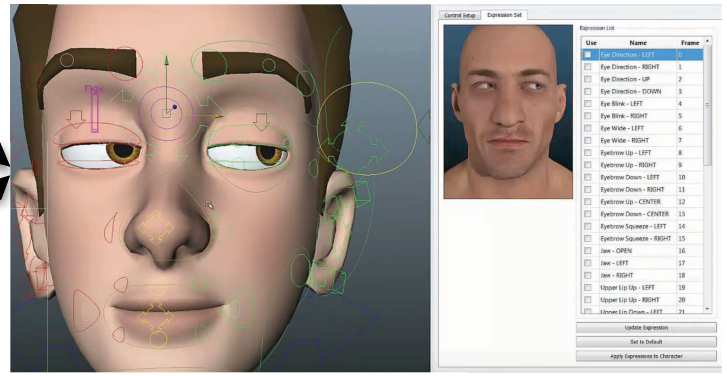
Human input followed by landmark detection



New 3D model generation

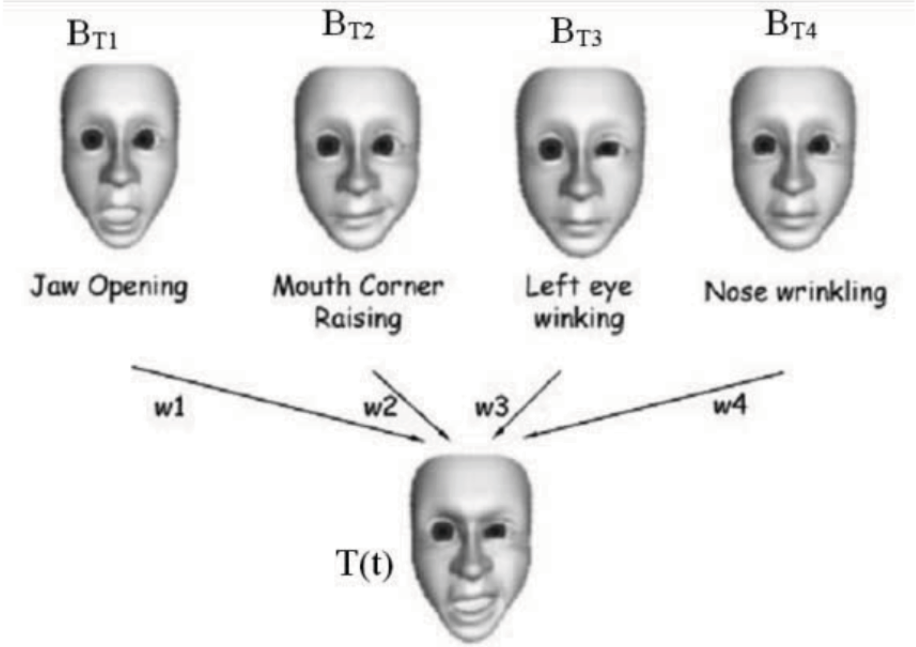


Existing 3D model morphing



Associate character rig controls to expression sets/landmark groups

# Related work: Blendshape Interpolation




Performance capture using blendshapes;
Regression-based estimation of weights

$$\mathbf{S}(t) = \sum_{i=1}^{k} \omega_i(t).B_{si}$$

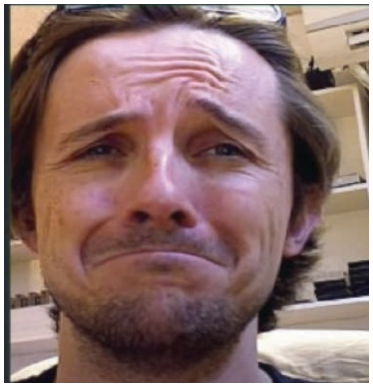
$$\mathbf{T}(t) = \sum_{i=1}^{k} \omega_i(t).B_{Ti}$$

# Drawbacks of existing methods

Geometric features transfer, expressions often do not!

Actual : **Disgust**

Perceived : **Angry**

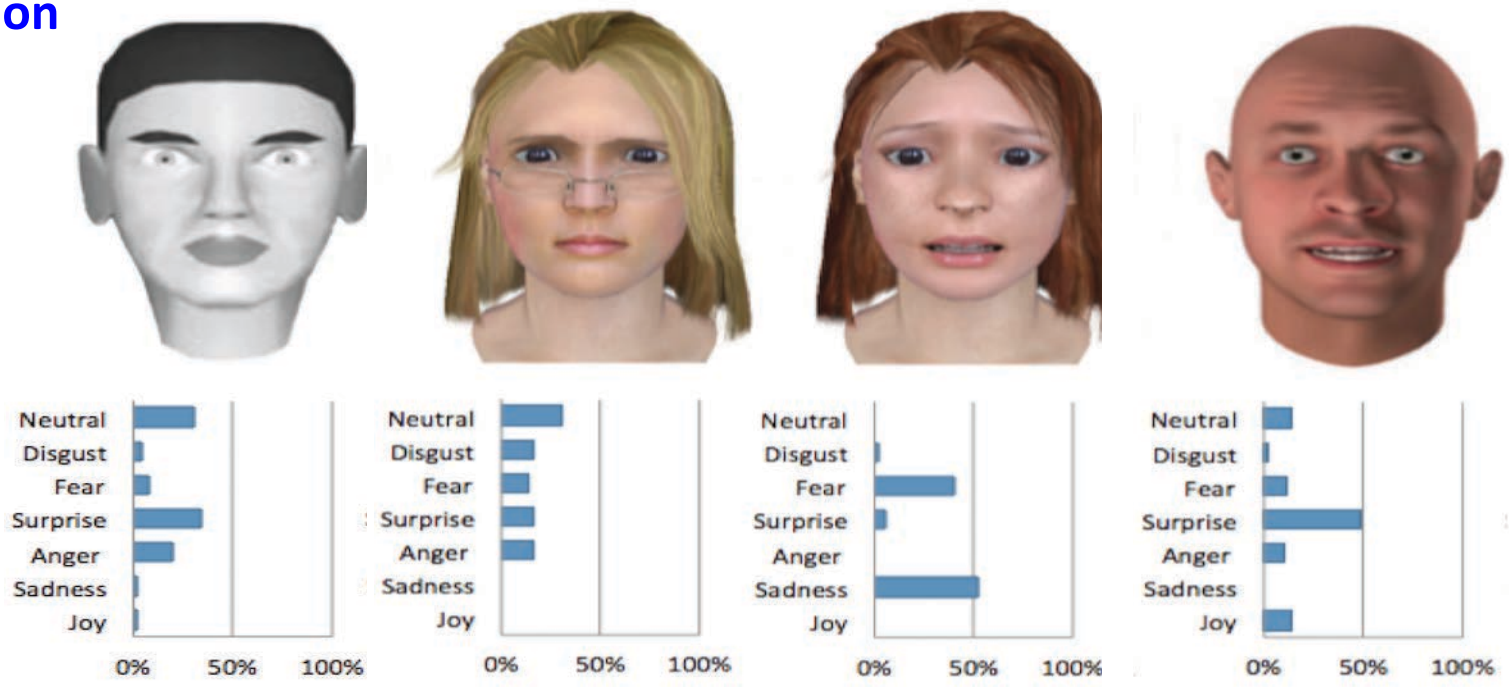


Actual : **Sad**

Perceived : **Confused**

# Drawbacks of existing methods

| Intended Expression | MPEG – 4 | HapFACS | HapFACS | FACSGen |
|---|---|---|---|---|
| | Anger | Anger | Fear | Fear |

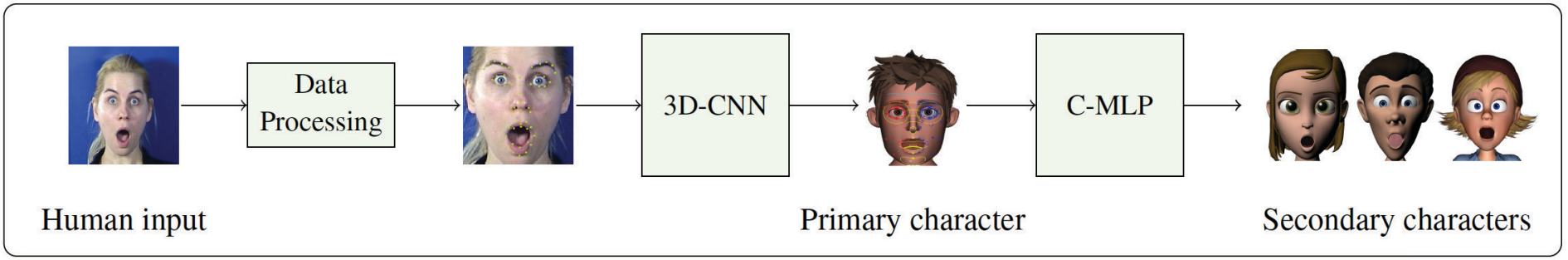



MPEG-4 : Pereira, F.C., Ebrahimi, T.: The MPEG-4 Book. Prentice Hall PTR, Upper Saddle River, NJ, USA (2002)
HapFACS : Amini, R., Lisetti, C.: HapFACS: an open source API/Software to generate FACS- Based expressions for ECAs animation (ACII). (2013) 270–275
FACSGen: Roesch, E.B., Tamarit, L., Reveret, L., Grandjean, D., Sander, D., Scherer, K.R.: FACSGen: a tool to synthesize emotional facial expressions through systematic manipulation of facial action units. Journal of Nonverbal Behavior (2011) 1–16

# Contributions

- A novel **perceptually valid** method to map 2D human face images to 3D stylized character rig controls.
- The ability to utilize this mapping to **generate 3D characters** with clear unambiguous facial expressions.
- A semi-supervised method to enable expression transfer between **multiple characters**.

# Our Approach




**3D-CNN**: Deep convolutional neural network for human to character transfer

**C-MLP**: Lightweight multi-layer perceptrons for character to character transfer

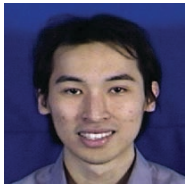
*Advantage*: No retraining of 3D-CNN required for a new character

*Challenge*: No existing dataset of matching 2D image – 3D character model pairs

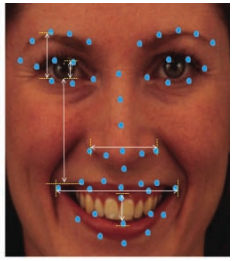
# Dataset and Preprocessing

Expression classes : Anger, Disgust, Fear, Neutral, Joy, Sadness, Surprise





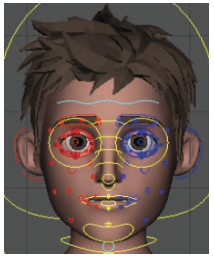
SFEW, CK+, MMI, KDEF and DISFA
~100k images

Bounding box and landmark detection; frontalization and geometry extraction



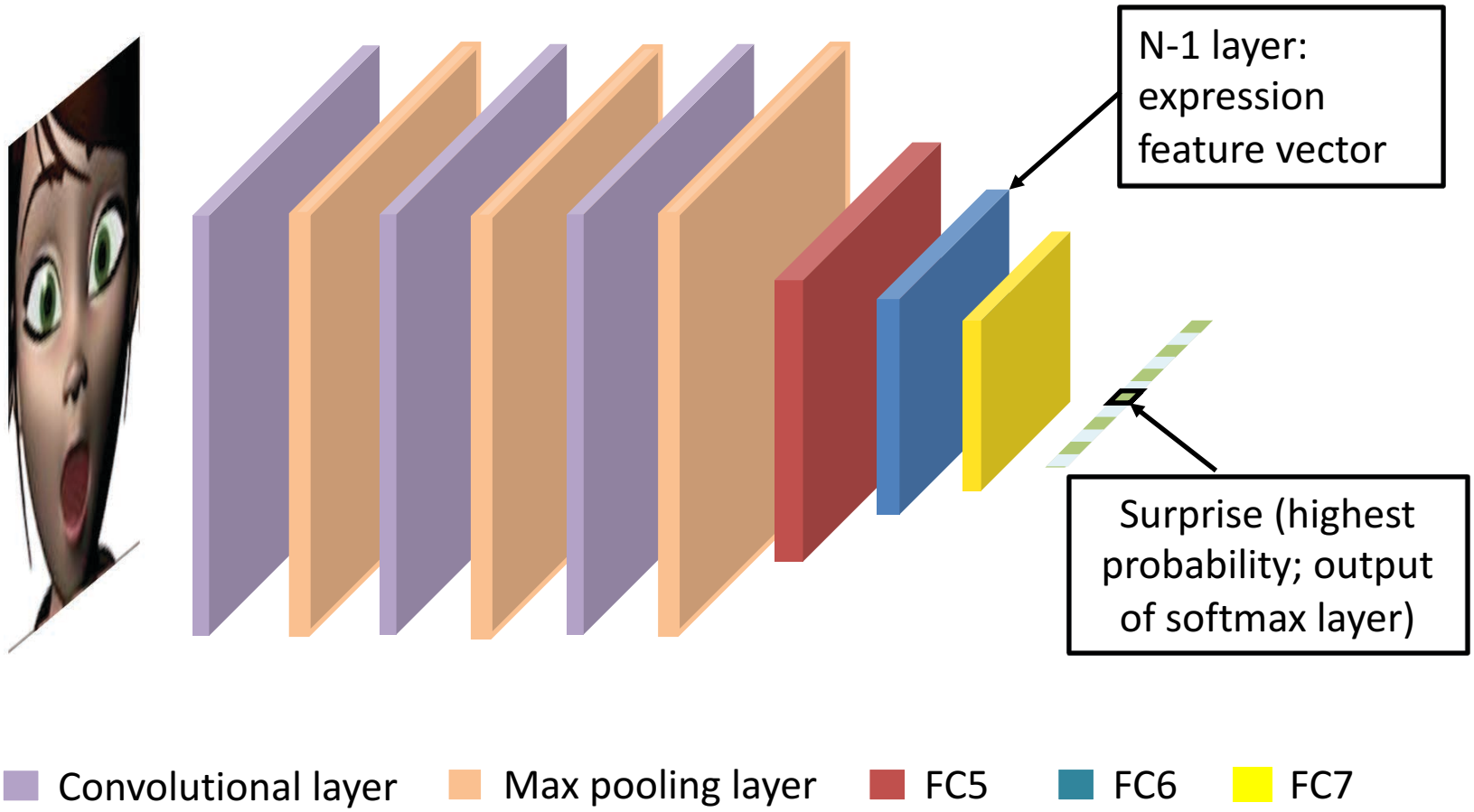
FERG-DB + 3 new characters
~55K images



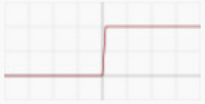
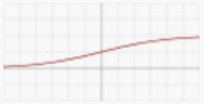
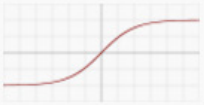
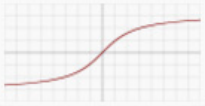
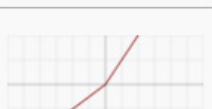

3D rig parameters; geometry extraction

*Aneja et al.,* Modeling Stylized Character Expressions via Deep Learning, *ACCV 2016*

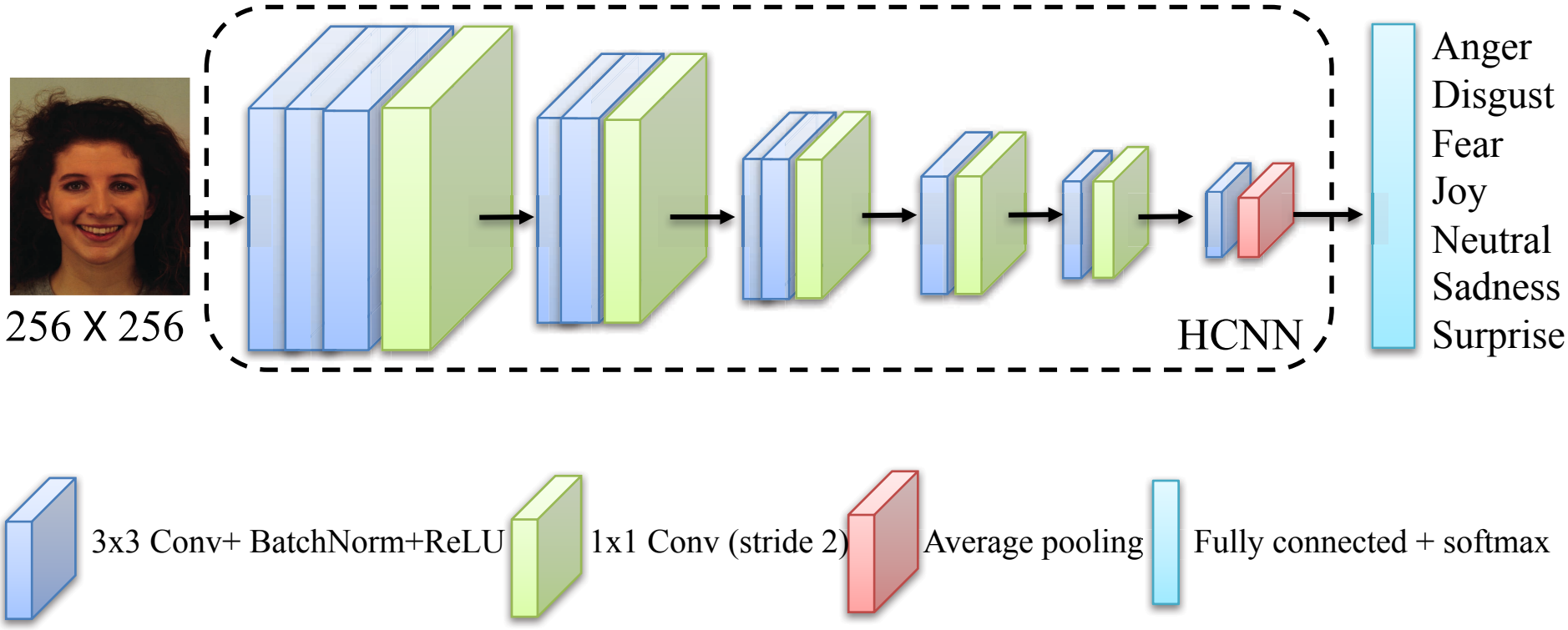
# CNN for expression recognition



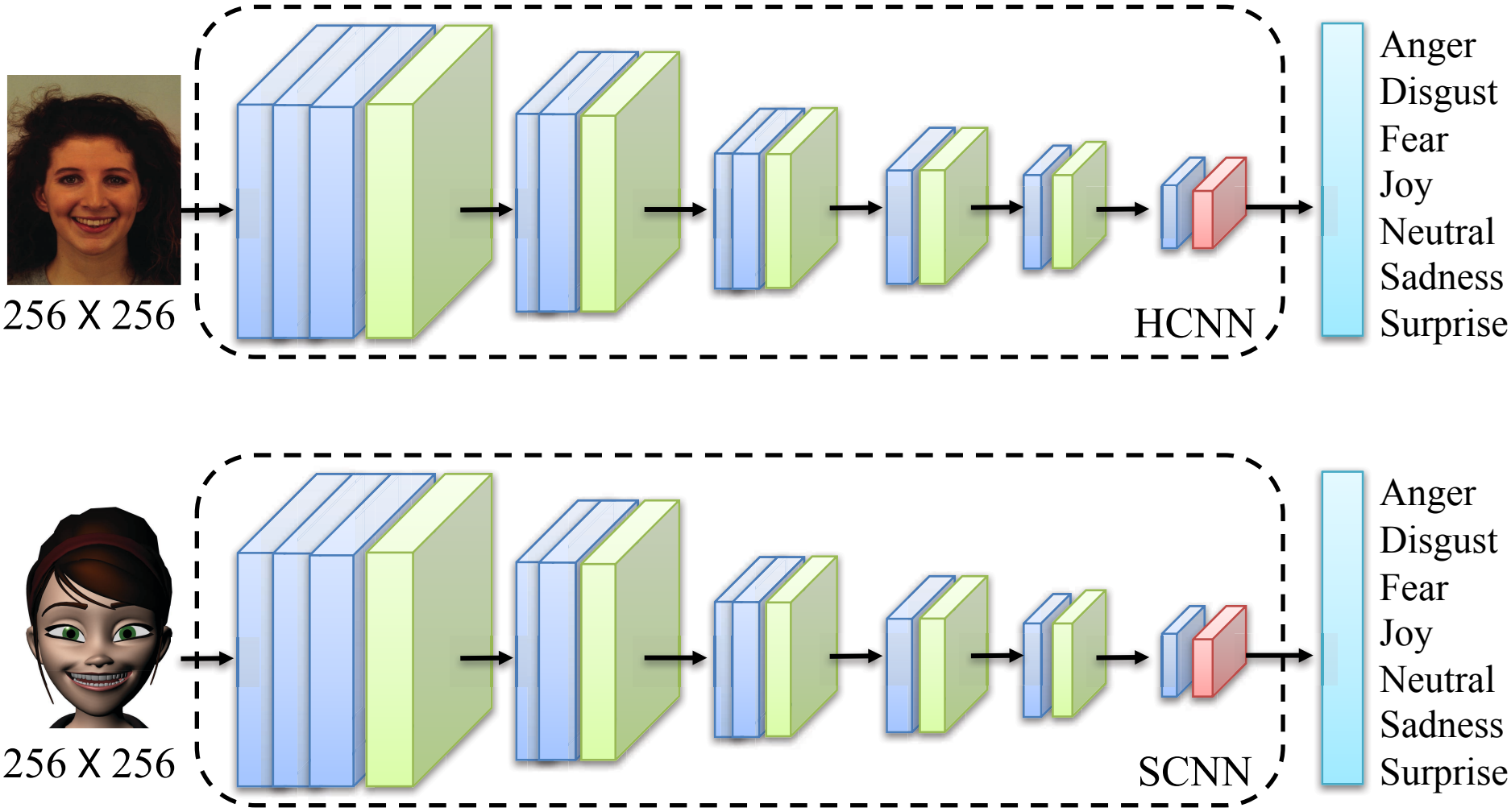
# Different activation functions in CNN

| Name | Plot | Equation | Derivative |
|---|---|---|---|
| Identity |  | $f(x) = x$ | $f'(x) = 1$ |
| Binary step |  | $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$ | $f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$ |
| Logistic (a.k.a Soft step) |  | $f(x) = \frac{1}{1+e^{-x}}$ | $f'(x) = f(x)(1-f(x))$ |
| TanH |  | $f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$ | $f'(x) = 1 - f(x)^2$ |
| ArcTan |  | $f(x) = \tan^{-1}(x)$ | $f'(x) = \frac{1}{x^2+1}$ |
| Rectified Linear Unit (ReLU) |  | $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$ | $f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$ |
| Parameteric Rectified Linear Unit (PReLU)[2] |  | $f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$ | $f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$ |
| Exponential Linear Unit (ELU)[3] |  | $f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$ | $f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$ |
| SoftPlus |  | $f(x) = \log_e(1+e^x)$ | $f'(x) = \frac{1}{1+e^{-x}}$ |

# Network Architecture



# Network Architecture



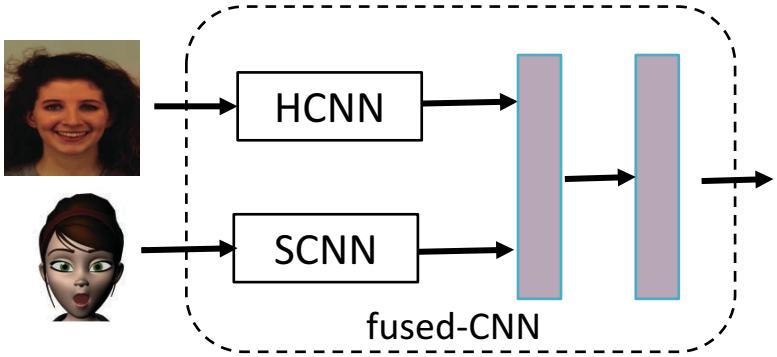
# Network Architecture

Distance metric

$$\phi_d = \alpha\,|\text{JS Distance}| + \beta\,|\text{Geometric Distance}|$$

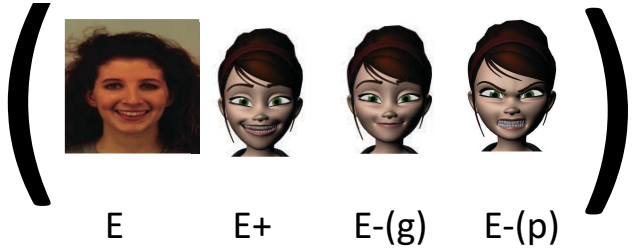
Expression feature vectors

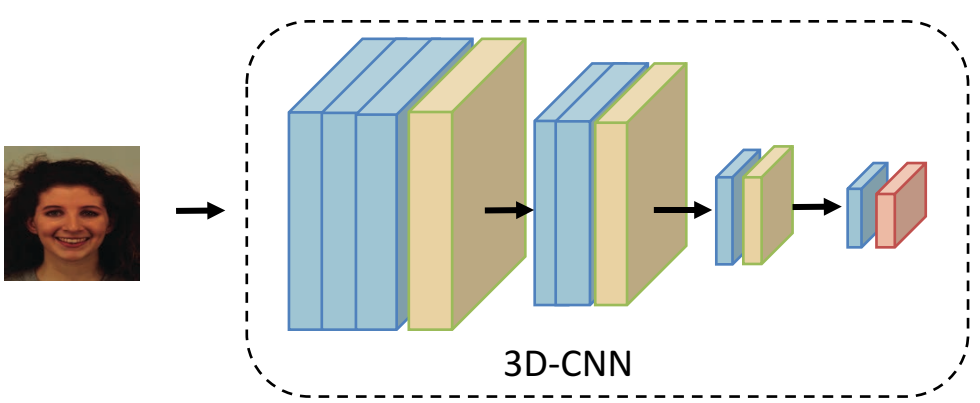
Geometry feature vectors




Similarity

Triplet generation based on geometry




Primary character

C-MLP

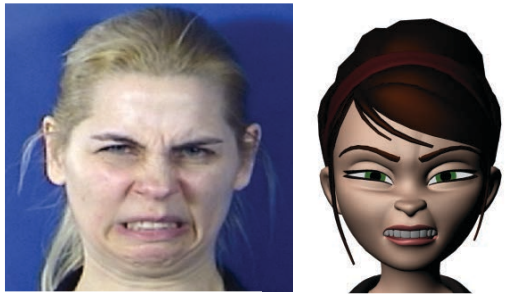
Secondary characters

# Results
# Human input - Character output

Expression: Sadness

Expression: Surprise

Expression: Disgust

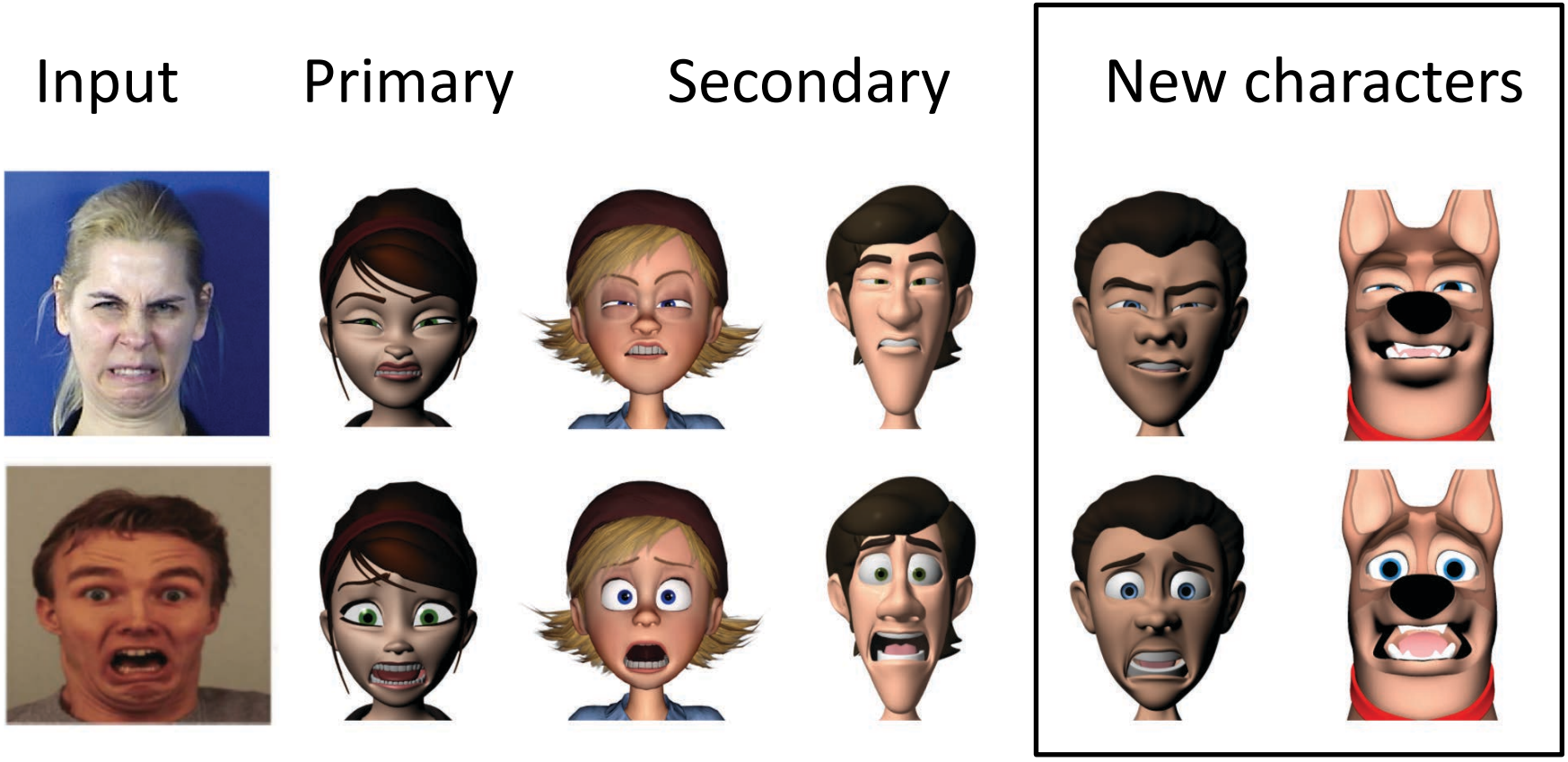


Expression: Anger

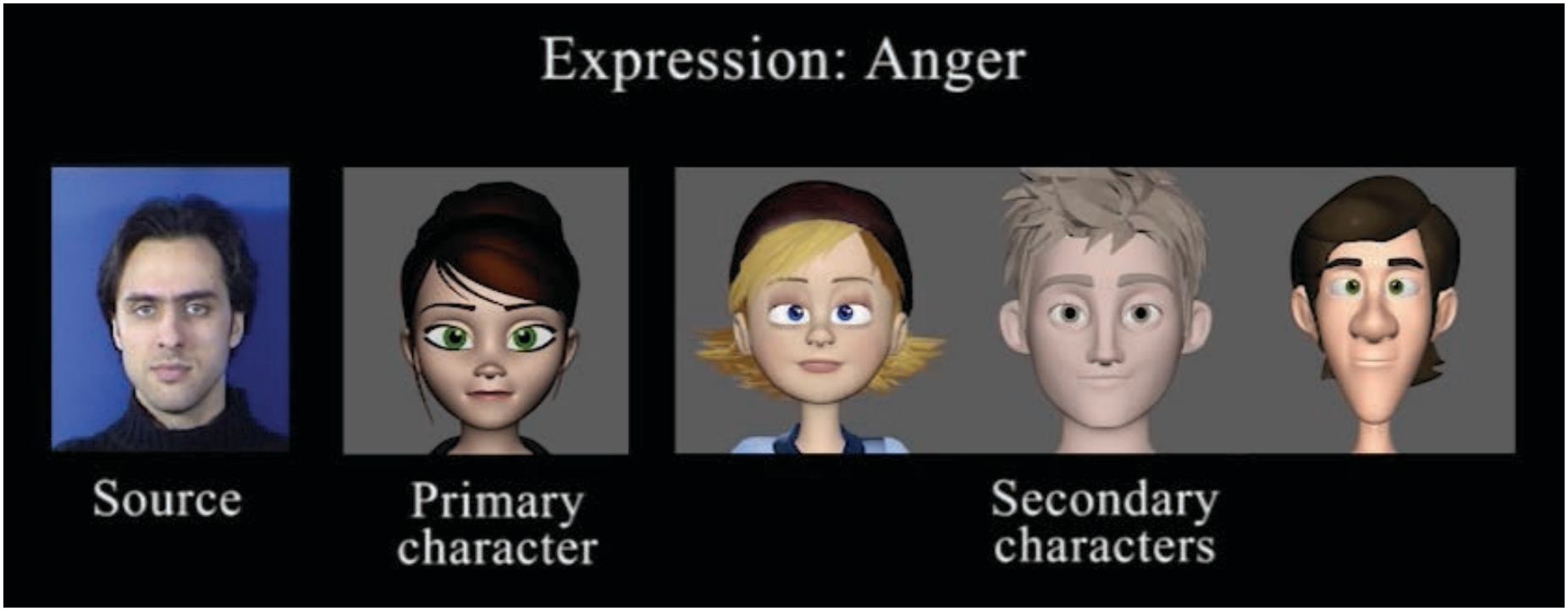
Expression: Fear

Expression: Joy

# Single Human to Multiple Characters



# Results: Video frame sequences

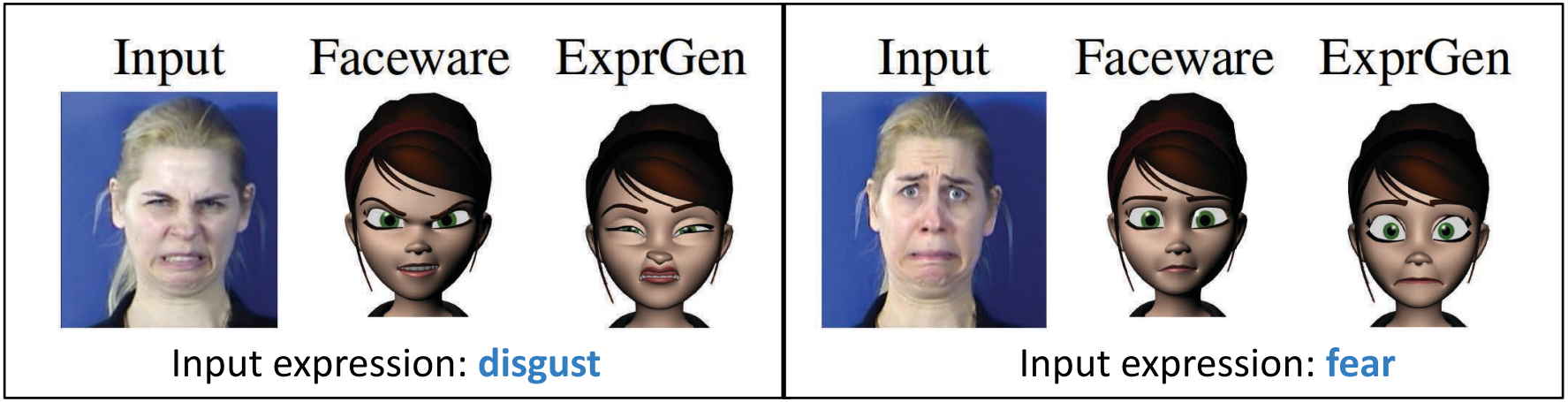


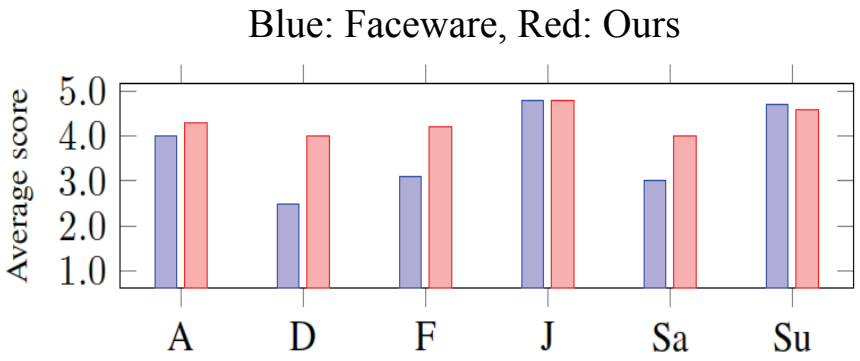
# Evaluation

| Perceived human expression (%) | Perceived character expression (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | A | D | F | J | N | Sa | Su |
| A | 71.32 | 16.28 | 5.43 | 1.55 | 3.10 | 0.78 | 1.55 |
| D | 14.29 | 67.35 | 4.08 | 1.02 | 4.08 | 8.16 | 1.02 |
| F | 2.88 | 6.47 | 64.03 | 2.16 | 3.60 | 3.60 | 17.27 |
| J | 0.92 | 1.83 | 0.92 | 90.83 | 1.83 | 0.92 | 2.75 |
| N | 1.09 | 3.26 | 2.17 | 4.35 | 76.09 | 10.87 | 2.17 |
| Sa | 1.80 | 3.60 | 2.70 | 1.80 | 18.02 | 71.17 | 0.90 |
| Su | 0.52 | 1.04 | 7.77 | 1.55 | 0.52 | 0.52 | 88.08 |

Confusion matrix of expression recognition for 1000 test cases with 30 MT subjects

# Evaluation




Quantitative comparison of expression transfer results
Faceware (blue bars) and ExprGen (red bars).

# Applications

- Efficient visual storytelling:

video games


motion capture films

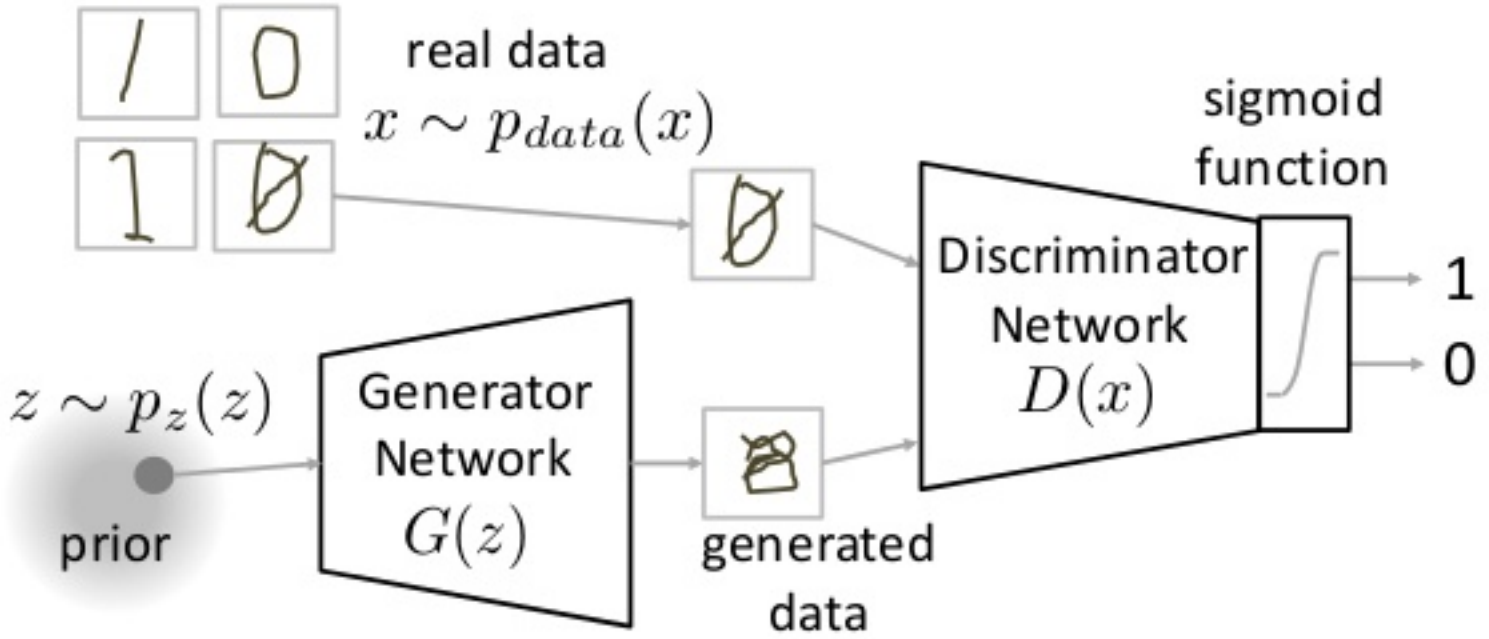

social VR experience

- Human-computer or human-robot interaction
- Helping autistic kids to recognize and convey emotions

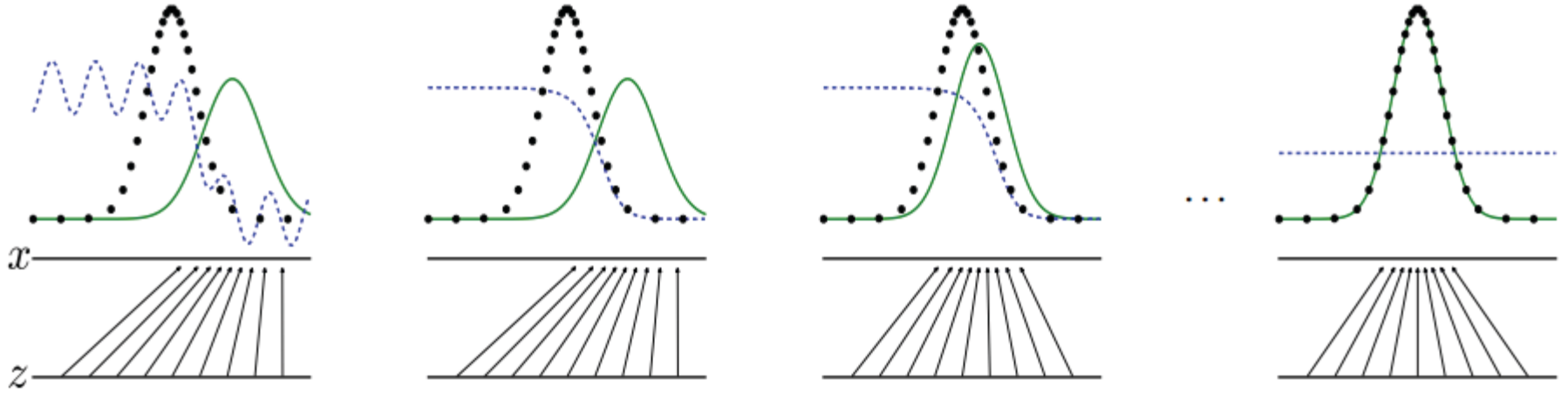
# Generative Adversarial Networks

$$\min_G \max_D V(D,G)$$

$$V(D,G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z))]$$

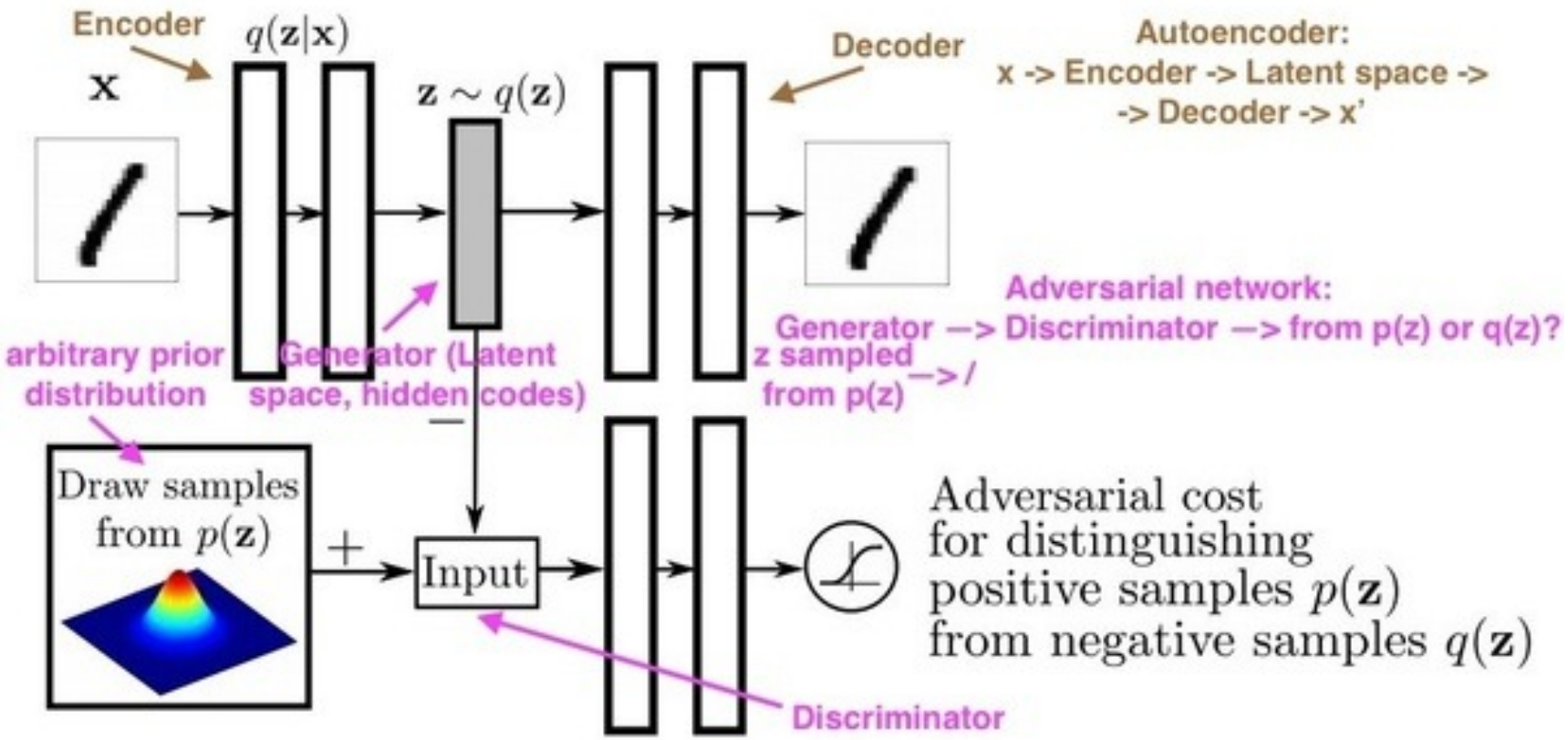


# GAN training

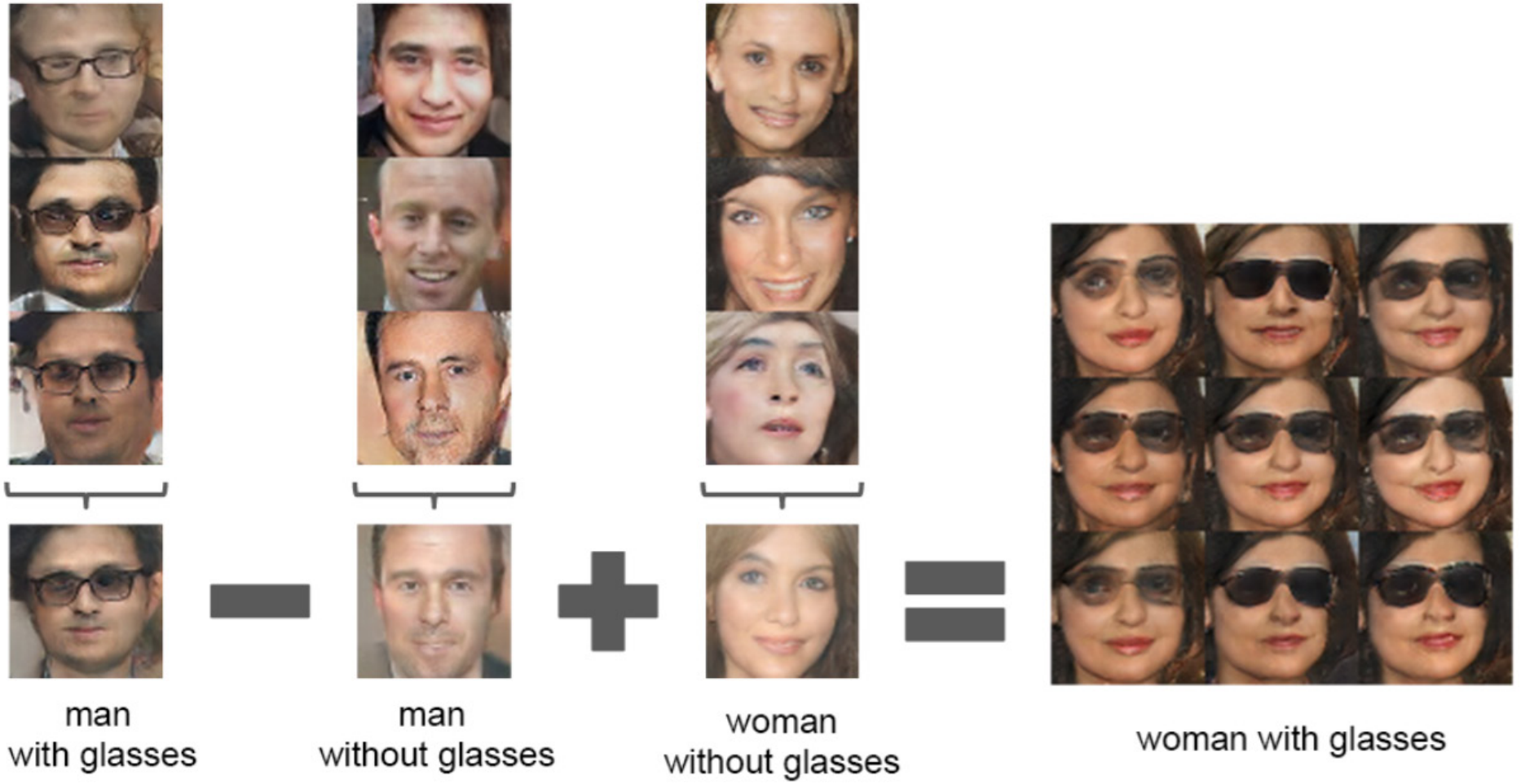



x: actual distribution of training data (black dotted line)
z: random noise input
green solid line: distribution of generated samples
blue dotted line: discriminator loss

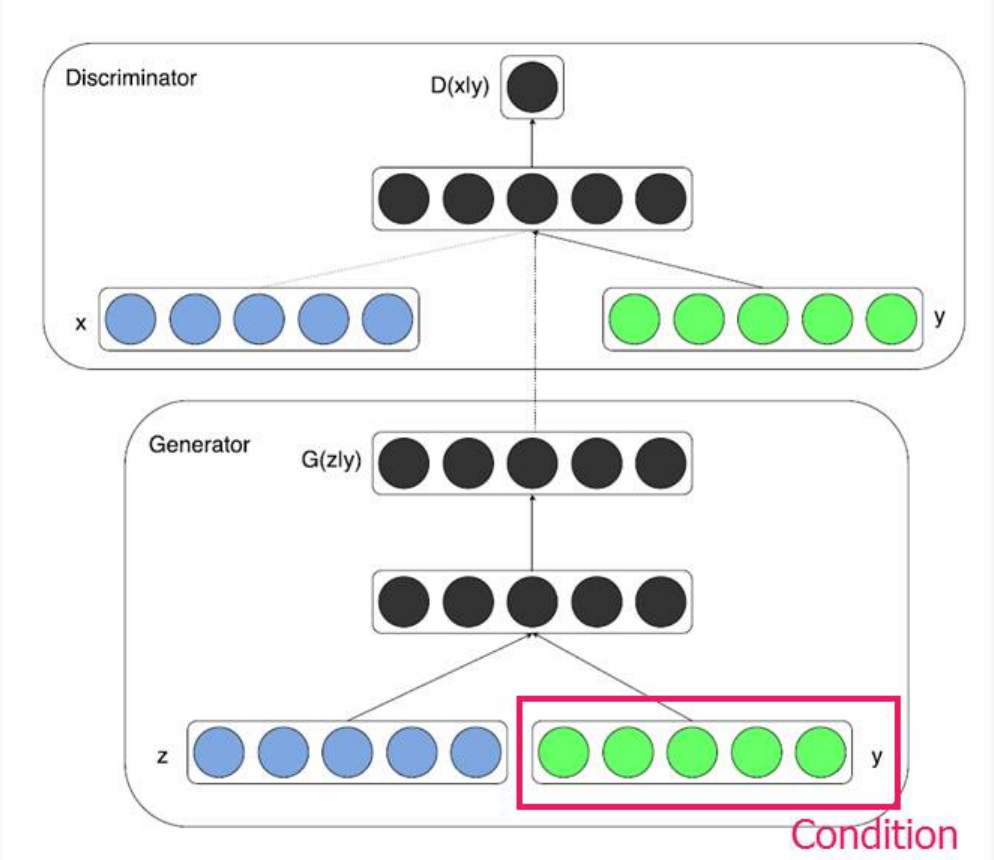
# GAN alternative; use of Autoencoder



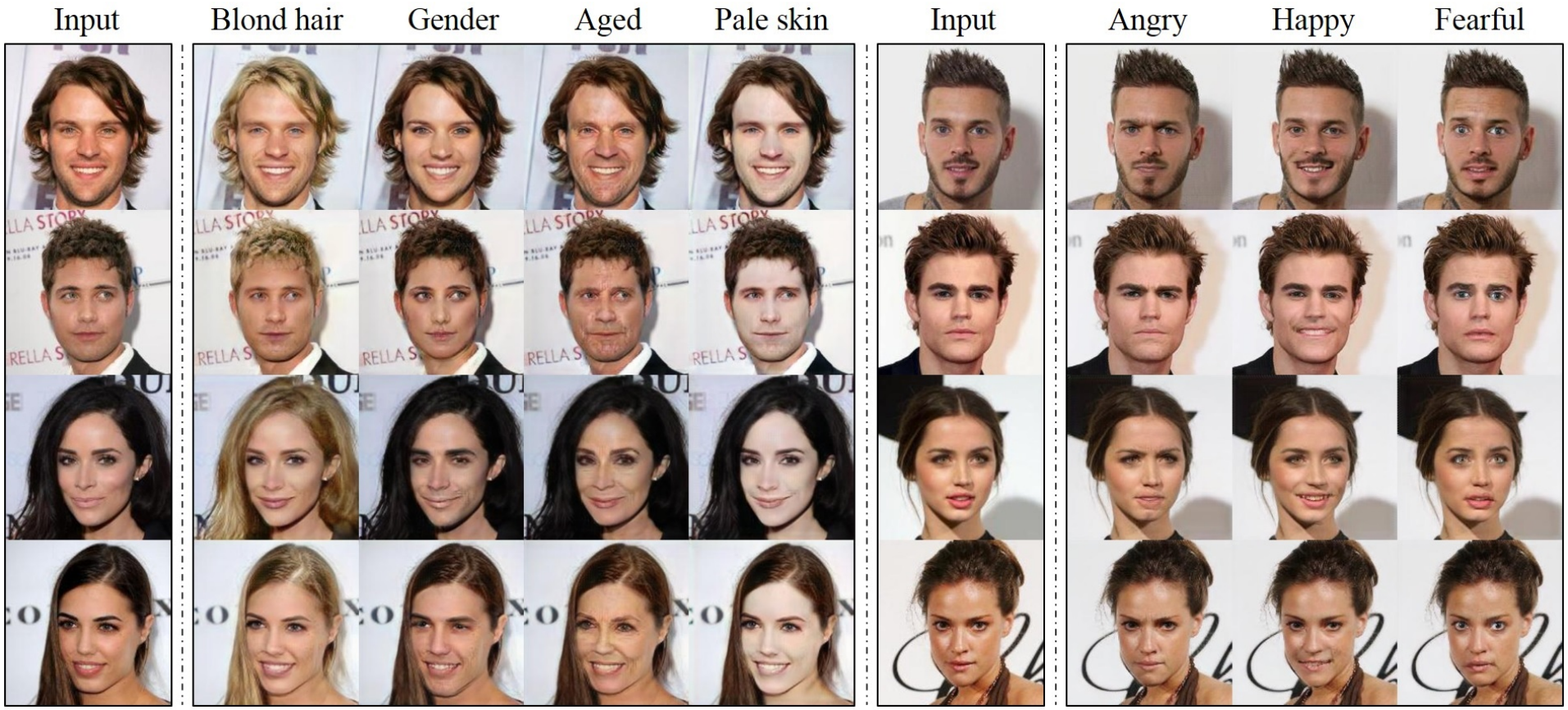
# Vector arithmetic with latent space in GAN



# Conditional Generative Adversarial Nets



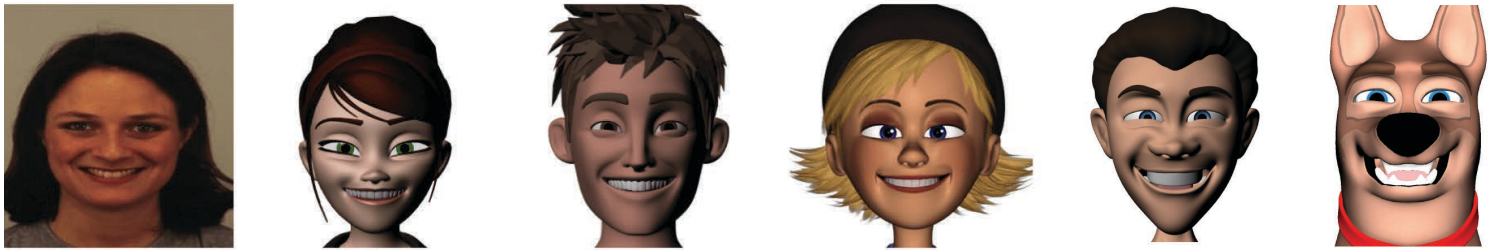
# Outputs of GAN conditioned on attributes



# References


- [***Our ACCV'16 paper***]: http://grail.cs.washington.edu/projects/deepexpr/deepali_accv2016.pdf
- [***Our WACV'18 paper***]: https://homes.cs.washington.edu/~bindita/2Dto3Dexpr_WACV.pdf
- [***Faceware***]: http://facewaretech.com/ ;geometry mapping technique based software
- [***Blendshape interpolation***]: https://www.researchgate.net/publication/320472350_Semi_Automatic_Retargeting_for_Facial_Expressions_of_3D_Characters_with_Fuzzy_logic_Based_on_Blendshape_Interpolation
- [***List of GAN papers***]: https://github.com/nightrome/really-awesome-gan
- [***Facial Animation survey***]: http://freesouls.github.io/2015/04/16/3d-facial-animation/index.html
- [***Details about CNNs***]: http://cs231n.github.io/convolutional-networks/

# Thank you!



**Questions?**