Modeling Stylized Character Expressions via Deep Learning

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

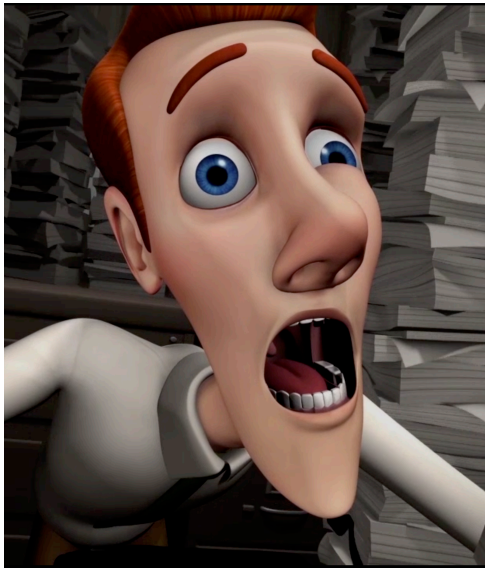
[1] Department of Computer Science and Engineering,
University of Washington, Seattle WA, USA
[2] Zillow Group, Seattle WA, USA
[3] Gage Academy of Art, Seattle WA, USA

ACCV '16

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

Stylized character expressions

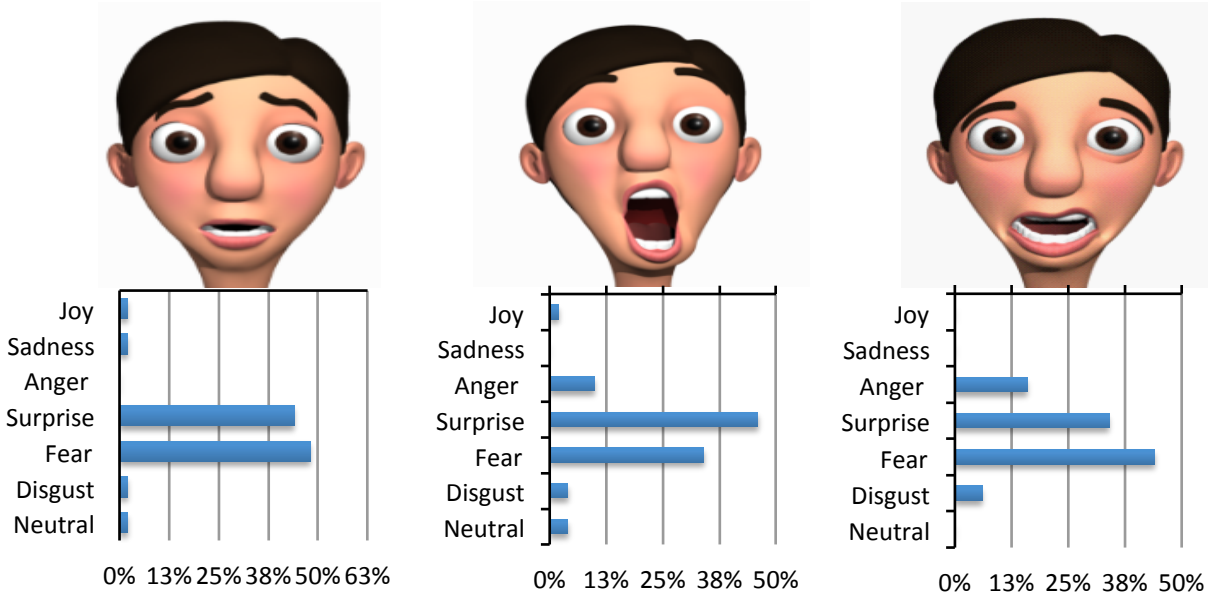


Animated Shorts from Animation Research Labs, University of Washington.

Creating recognizable expressions

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

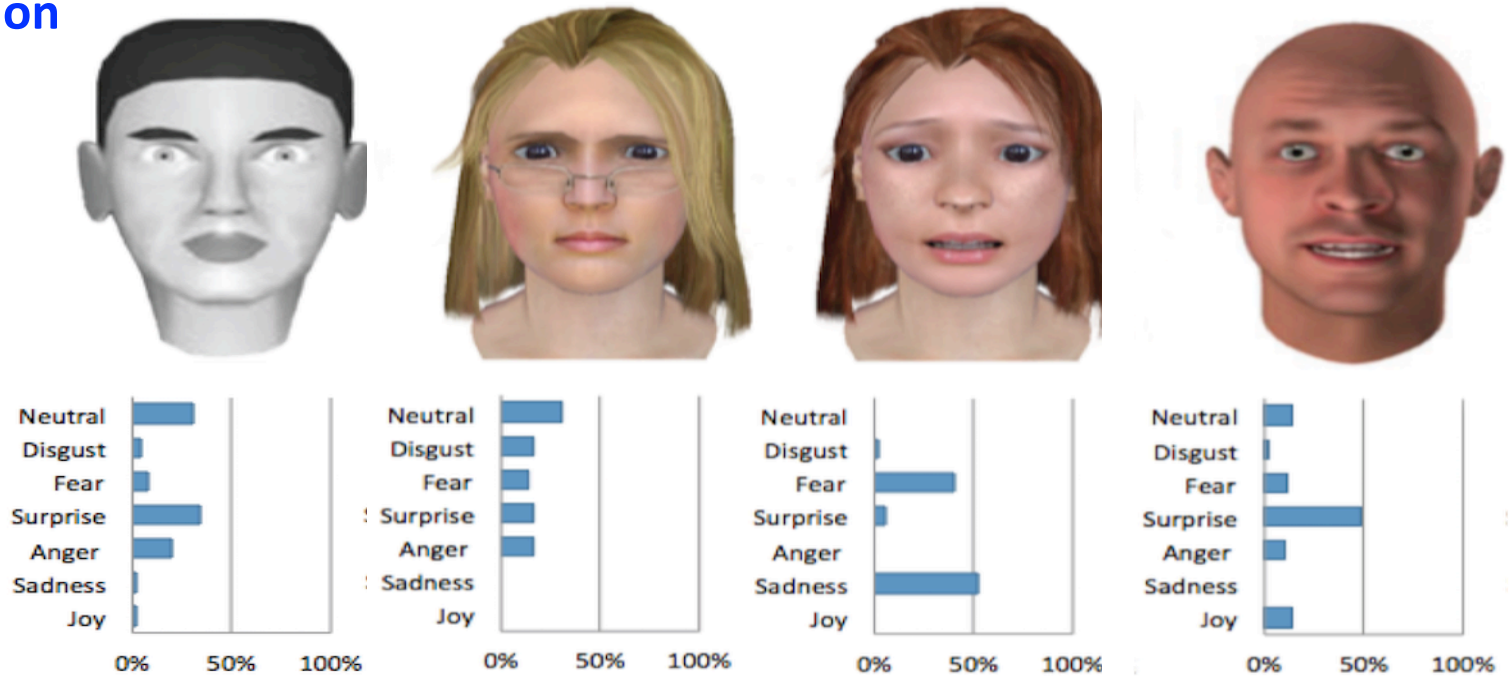
- We asked professional animators to make this character look surprised.
  - None of the expressions achieved above 50% recognition on Mechanical Turk.



Geometric Mapping

Intended Expression

MPEG – 4	HapFACS	HapFACS	FACSGen
Anger	Anger	Fear	Fear

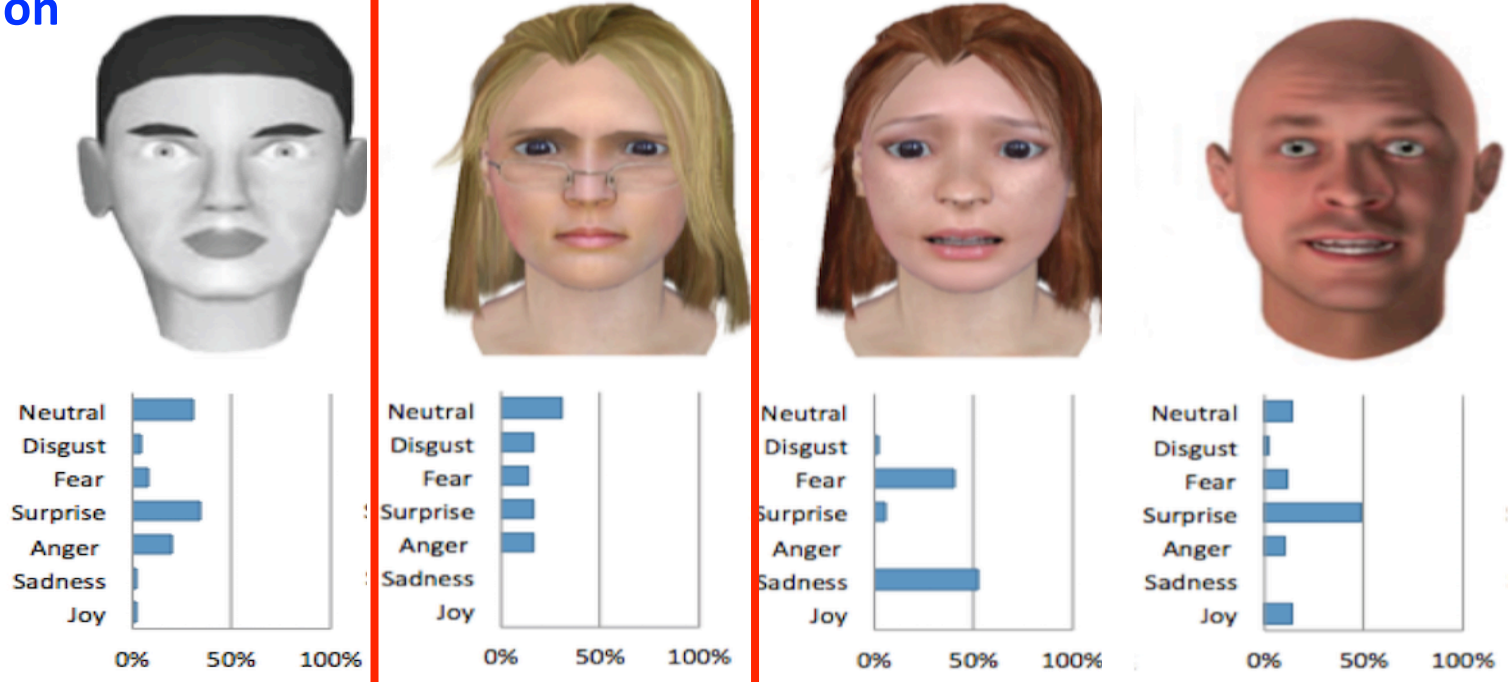


MPEG-4 : Pereira, F.C., Ebrahimi, T.: The MPEG-4 Book. Prentice Hall PTR, Upper Saddle River, NJ, USA (2002)
HapFACS : Amini, R., Lisetti, C.: HapFACS: an open source API/Software to generate FACS- Based expressions for ECAs animation (ACII). (2013) 270–275
FACSGen: Roesch, E.B., Tamarit, L., Reveret, 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

Geometric Mapping

Intended Expression

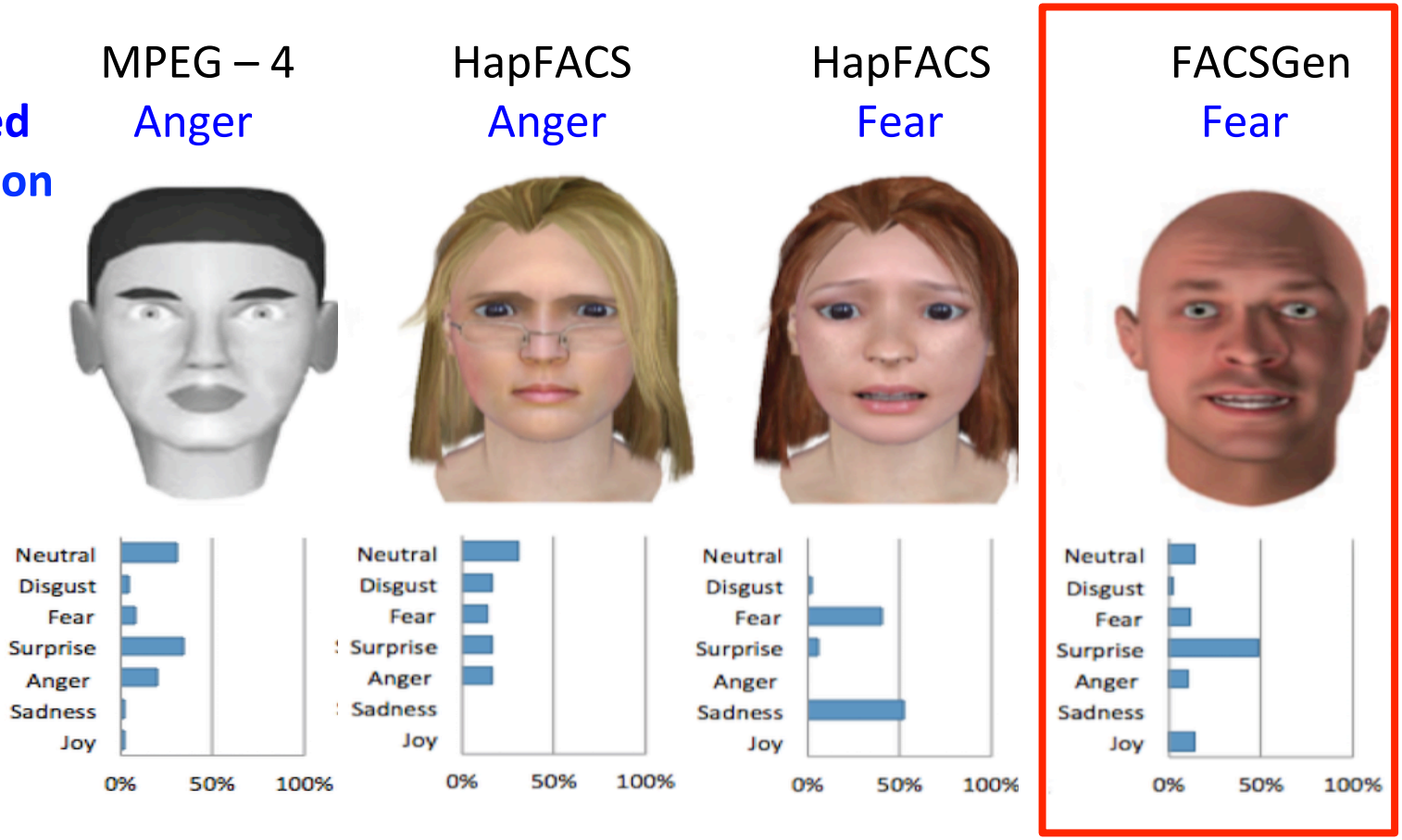
MPEG – 4	HapFACS	HapFACS	FACSGen
Anger	Anger	Fear	Fear



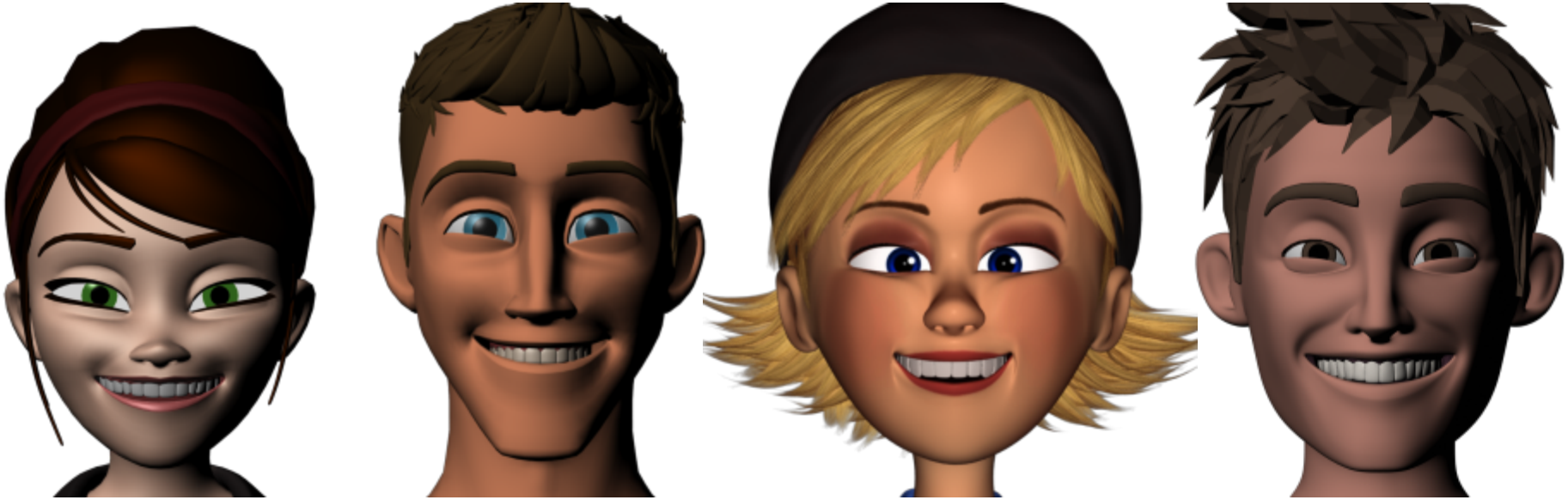
MPEG-4 : Pereira, F.C., Ebrahimi, T.: The MPEG-4 Book. Prentice Hall PTR, Upper Saddle River, NJ, USA (2002)
HapFACS : Amini, R., Lisetti, C.: HapFACS: an open source API/Software to generate FACS- Based expressions for ECAs animation (ACII). (2013) 270–275
FACSGen: Roesch, E.B., Tamarit, L., Reveret, 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

Geometric Mapping

Intended Expression



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 data-driven **perceptual** model of facial expressions.
- A novel stylized character data set **(FERG-DB)** with cardinal expression annotations.
- A mechanism to accurately **retrieve plausible character expressions** from human expression queries.
  - Validated the results (Expert and Mechanical Turk)

Our Approach

Retrieve characters using

perceptual model mapping
and human geometry

- Use deep learning to learn mappings between
  - **Human expressions and characters expressions**
  - Humans and humans
  - Characters and characters

- This is not only geometric mapping
  - It is perceptual modelling of expressions!

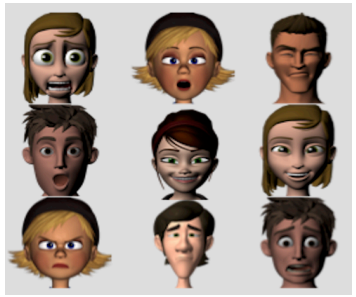
Expression Retrieval

Step 1

Use deep learning to create a perceptual model of human expressions

Human feature space

$f($  $)$



Step 2

Learn analogous character model

Character feature space

$f'($  $)$

$f($  $)$

Step 3

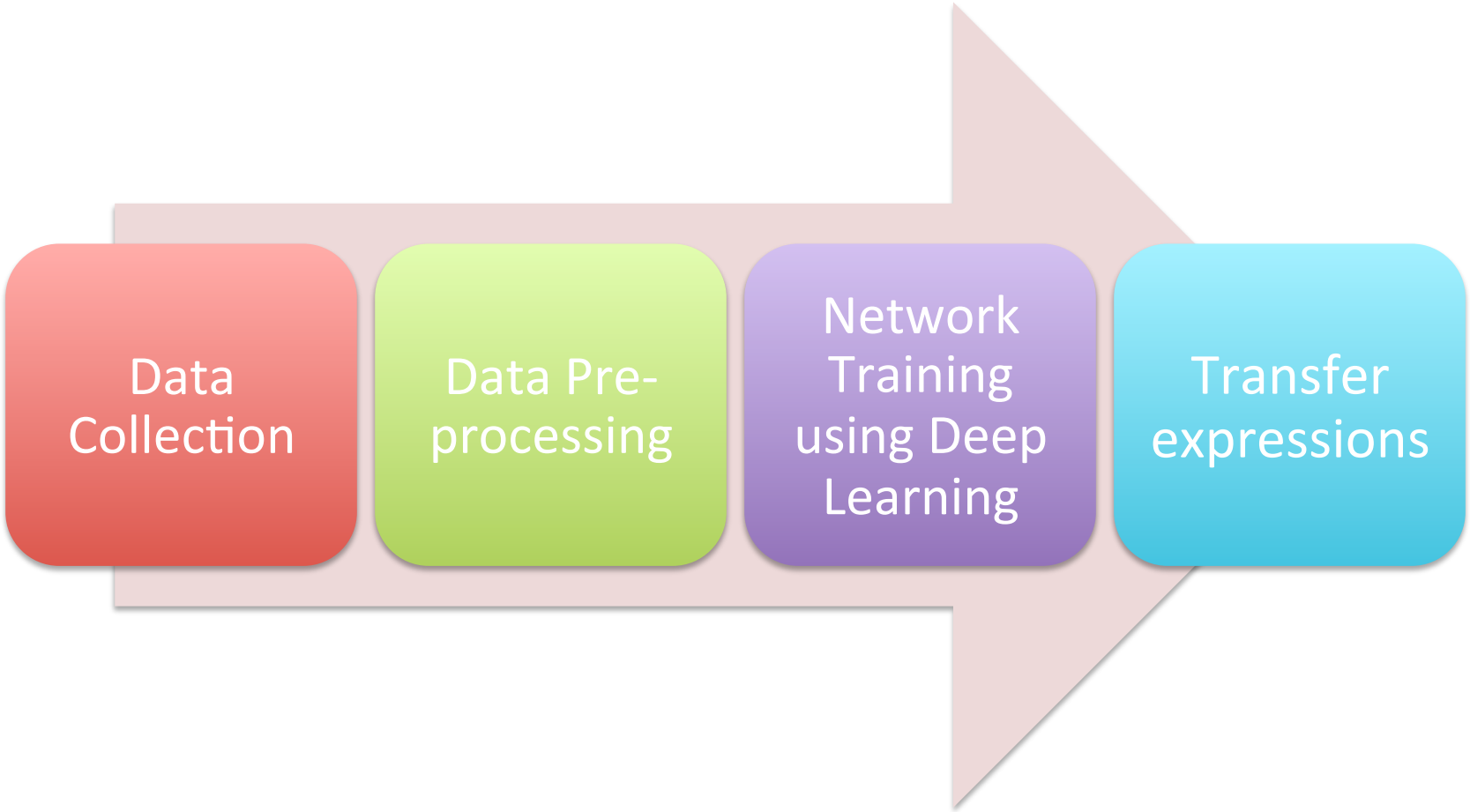
Learn Mapping

$f'($  $)$

Step 4

Retrieve characters using perceptual model mapping and human geometry

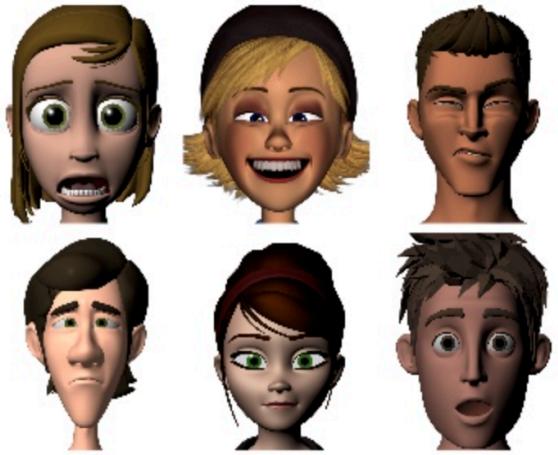
Steps


Data Collection
Data Pre-processing
Network Training using Deep Learning
Transfer expressions

Training Data

- Seven classes : Anger, Disgust, Fear, Joy, Neutral, Sad and Surprise

- Stylized Characters expression database
  - Total of 70K images
  - **Facial Expression Research Group (FERG-DB) is publicly available.**



- Human expression database
  - Total of 75K images

FERG-DB Stylized character expression database: http://grail.cs.washington.edu/projects/deepexpr/ferg-db.html

Human Database

- CK+: The Extended Cohn-Kanade -309 images
- DISFA: Denver Intensity of Spontaneous Facial Actions 60,000 images
- KDEF: The Karolinska Directed Emotional Faces 4900 images
- MMI: 10,000 images

- Total of 75K images - We balanced out the final number of the samples for training our network to avoid any bias towards any particular expression.

Stylized Character Database

- Six stylized characters (adding two more characters soon!)
  - The animator created the key poses for each expression and labeled via MT to populate the database initially
  - Key poses having **70%** MT test agreement and then interpolated between the key poses
- We only used the expression key poses having **70%** MT test agreement among 50 Turkers for the same pose. Interpolating between the key poses resulted in **70K images** (around 8,000 images per character).

Data Pre-processing

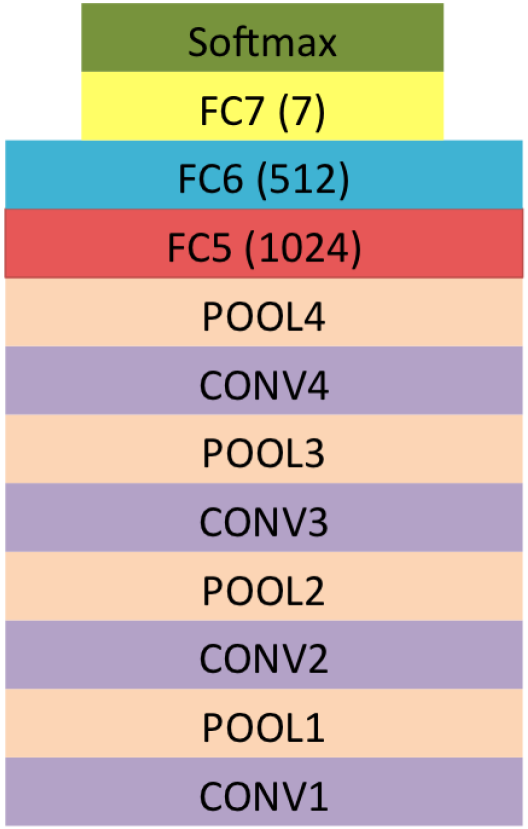
Extract Face 49 landmarks (Intraface)

Register faces to an average frontal face via an **affine transformation**

Face bounding box selection

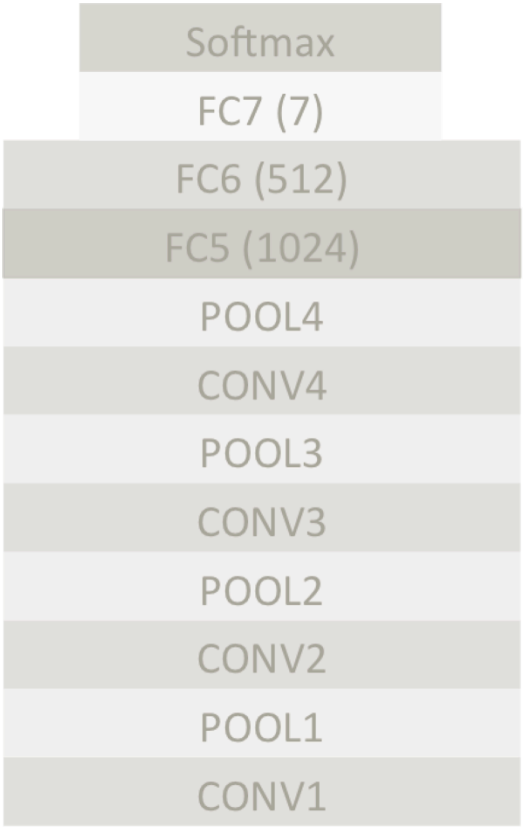
Re-size to 256x256 pixels for analysis

Network Architecture

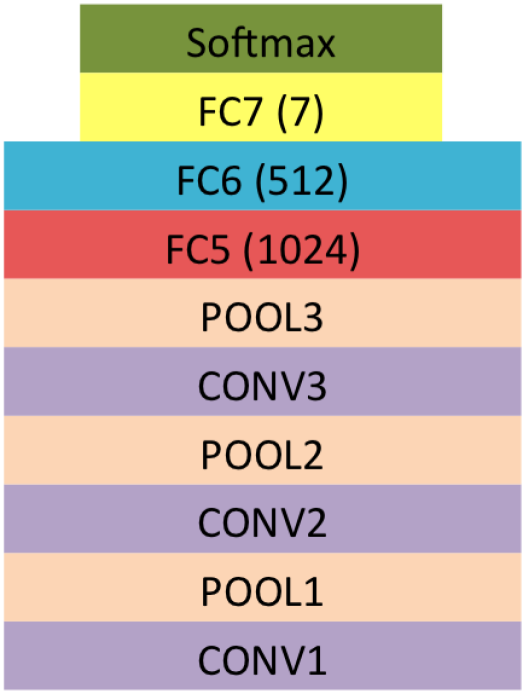


Human CNN (HCNN)

Network Architecture

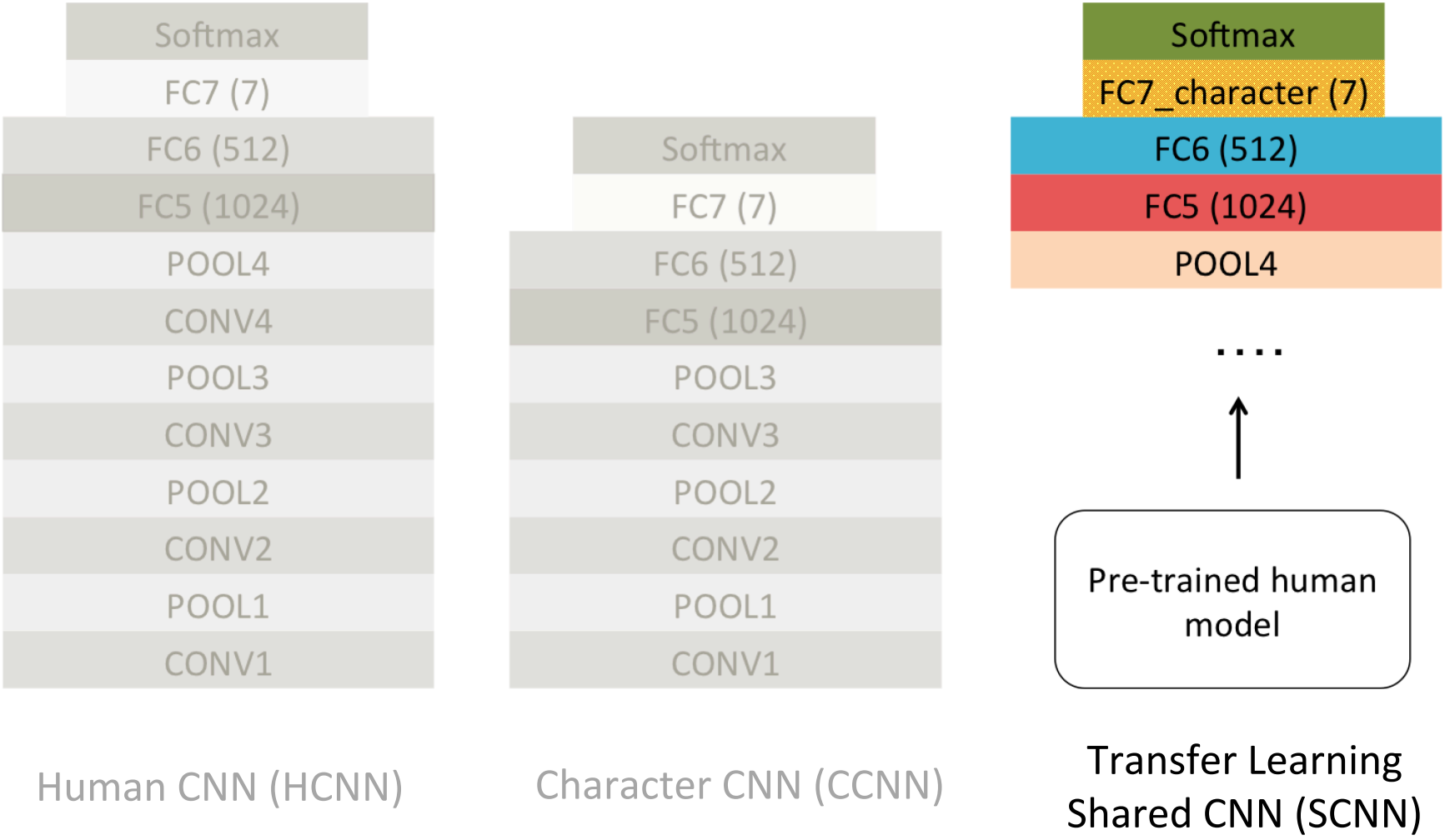


Human CNN (HCNN)

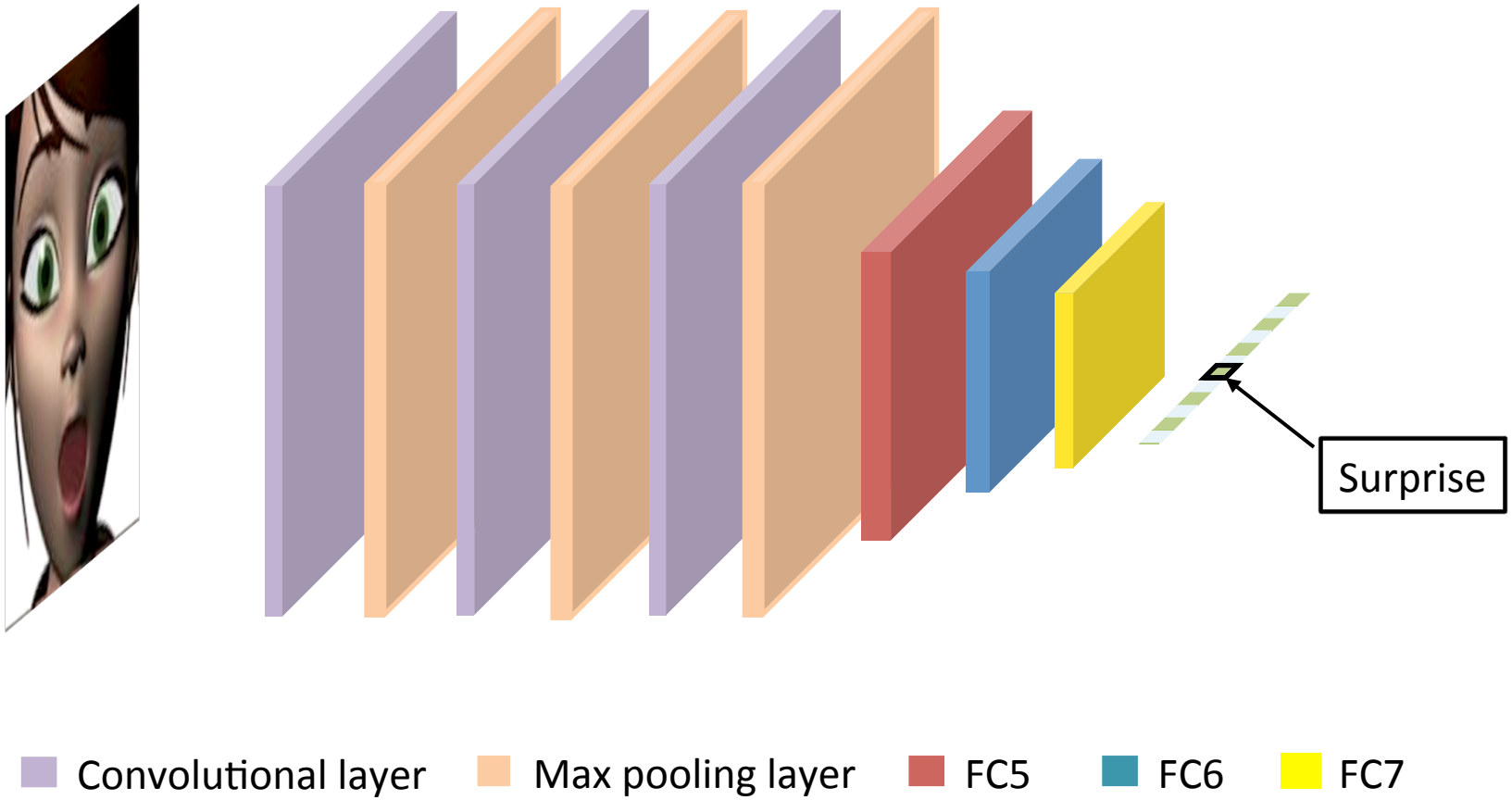


Character CNN (CCNN)

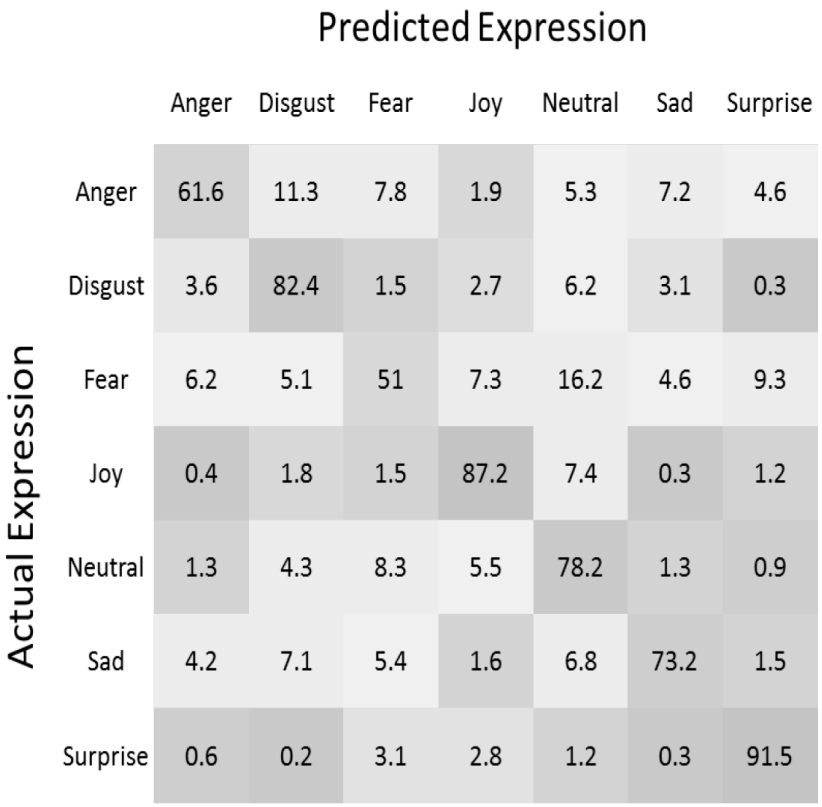
Network Architecture



Network prediction

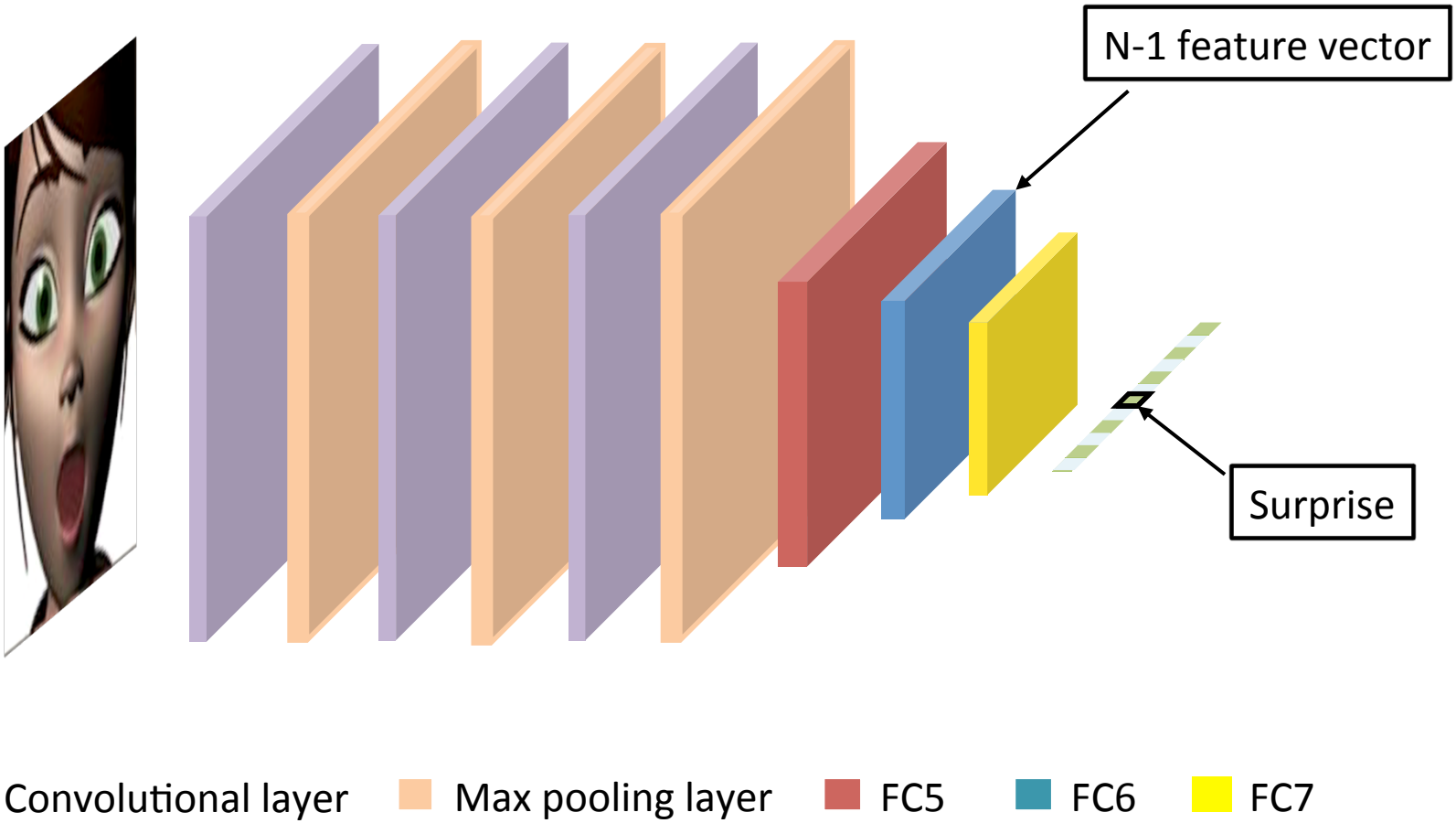


Expression Recognition Accuracy



- Accuracy of Human CNN - 85.27%
- Accuracy of Character CNN - 89.02%

Retrieval


N-1 feature vector
Surprise
Convolutional layer
Max pooling layer
FC5
FC6
FC7

Distance Metrics

- Extracted features from the **last fully connected layer (FC6)** of both the models: HCNN and SCNN and normalized the feature vectors

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

Expression feature vectors
(N-1) Layer features

Geometry feature vectors

Lin, J.: Divergence measures based on the shannon entropy. Information Theory, IEEE Transactions on 37 (1991) 145–151 44.

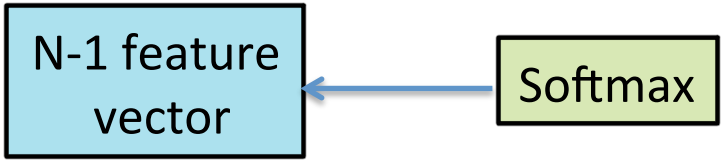
Distance Metrics

- Extracted features from the **last fully connected layer (FC6)** of both the models: HCNN and SCNN and normalized the feature vectors

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

Expression feature vectors
(N-1) Layer features

Geometry feature vectors



Back propagation

Lin, J.: Divergence measures based on the shannon entropy. Information Theory, IEEE Transactions on 37 (1991) 145–151 44.

Distance Metrics

- Extracted features from the **last fully connected layer (FC6)** of both the models: HCNN and SCNN and normalized the feature vectors

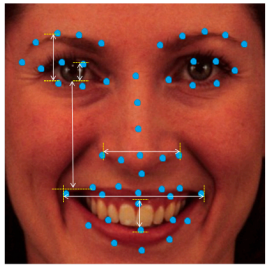
$$\phi_d = \alpha\,|\text{JS Distance}| + \beta\,|\text{Geometric Distance}|$$

Expression feature vectors
(N-1) Layer features

Geometry feature vectors

N-1 feature vector ← Softmax

Back propagation



Lin, J.: Divergence measures based on the shannon entropy. Information Theory, IEEE Transactions on 37 (1991) 145–151 44.

Character Retrieval

Query

Multiple retrieval results for the joy query image

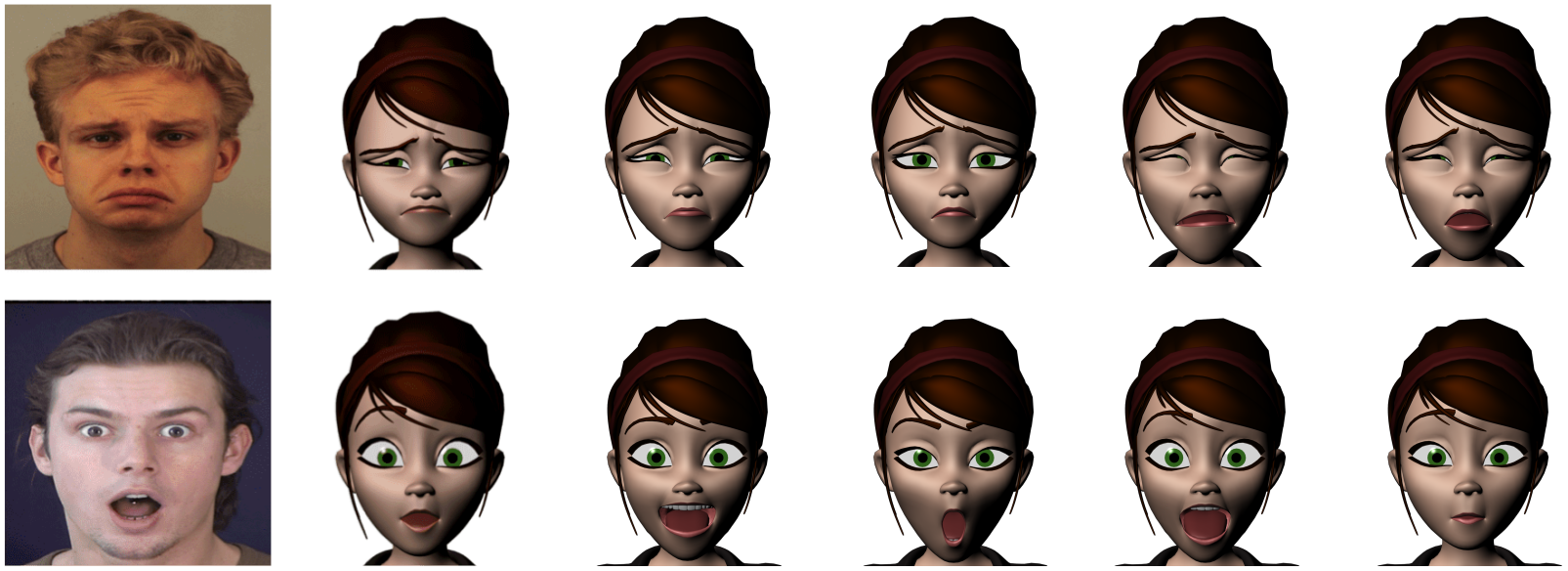
Query

$$\phi_d = \alpha \left| \text{JS Distance} \right|$$

Character Retrieval

Query

Character retrievals sorted by geometry



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

DeepExpr Results

Query

Top matches of Character retrievals

Average Retrieval Score
(for each expression across all characters)

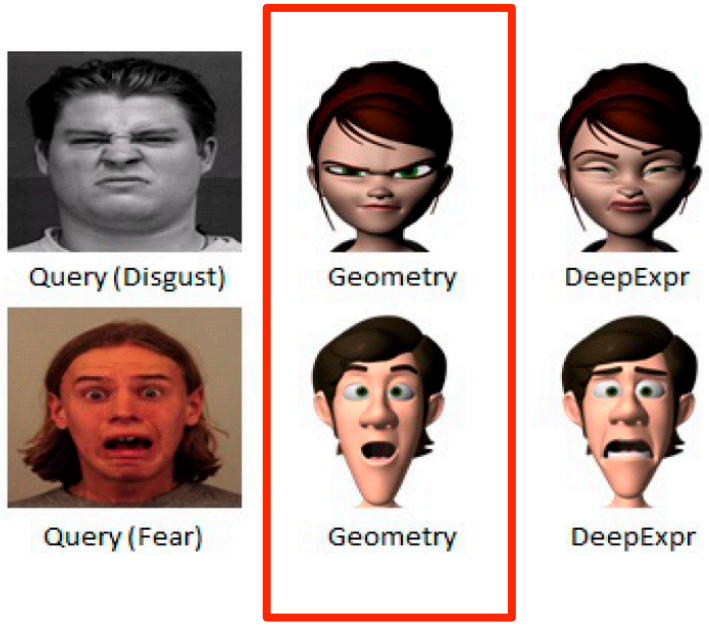
$$score(q) = \frac{1}{1 - N \cdot N_{rel}} \left(\sum_{k=1}^{N_{rel}} R_k - \frac{N_{rel}(N_{rel}+1)}{2} \right)$$

Expression	Geometry	DeepExpr
Anger	0.384	0.213
Disgust	0.386	0.171
Fear	0.419	0.228
Joy	0.276	0.106
Neutral	0.429	0.314
Sad	0.271	0.149
Surprise	0.322	0.125

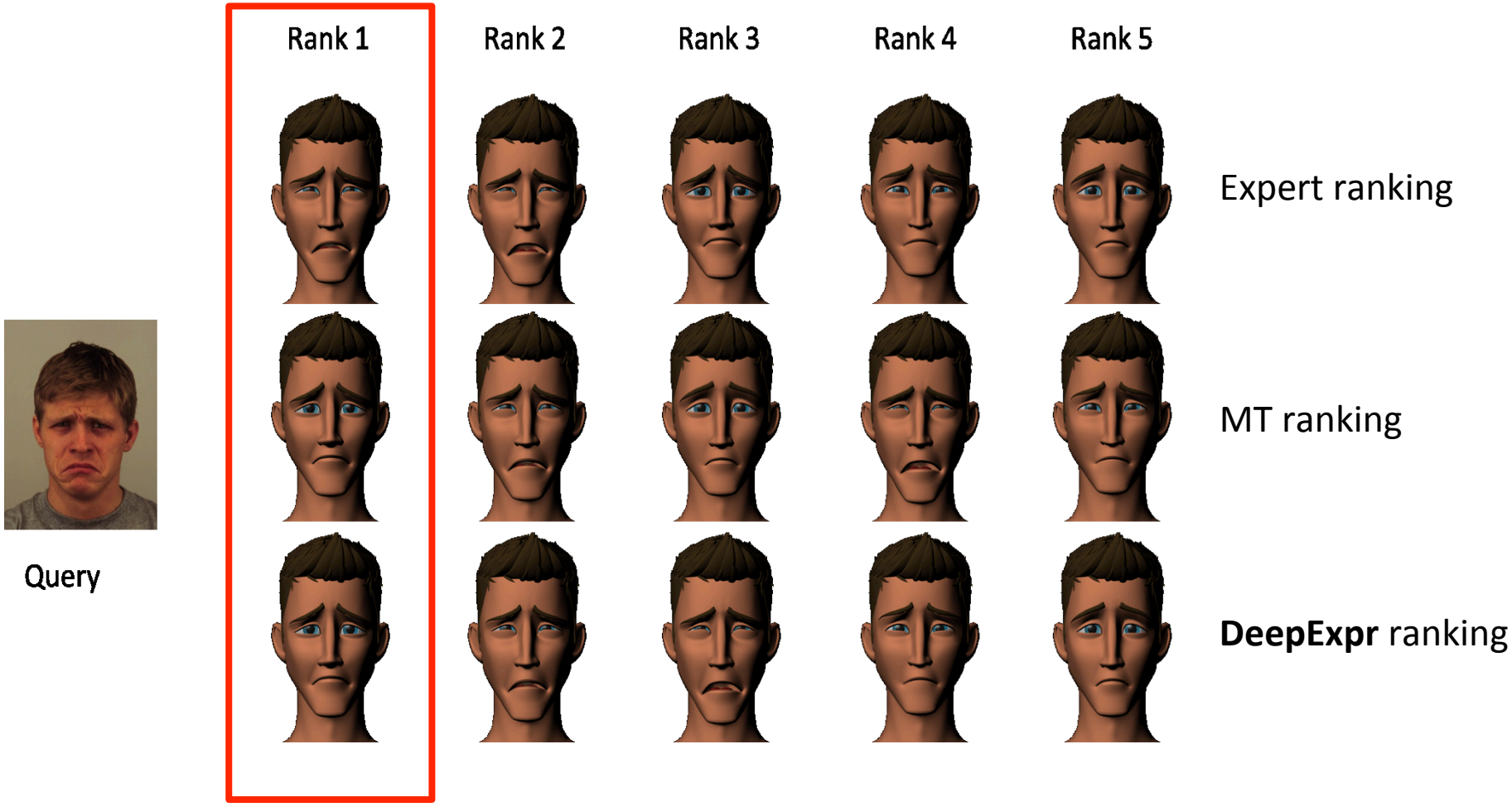
Average Retrieval Score
(for each expression across all characters)

$$score(q) = \frac{1}{1 - N \cdot N_{rel}} \left(\sum_{k=1}^{N_{rel}} R_k - \frac{N_{rel}(N_{rel}+1)}{2} \right)$$

Expression	Geometry	DeepExpr
Anger	0.384	0.213
Disgust	0.386	0.171
Fear	0.419	0.228
Joy	0.276	0.106
Neutral	0.429	0.314
Sad	0.271	0.149
Surprise	0.322	0.125

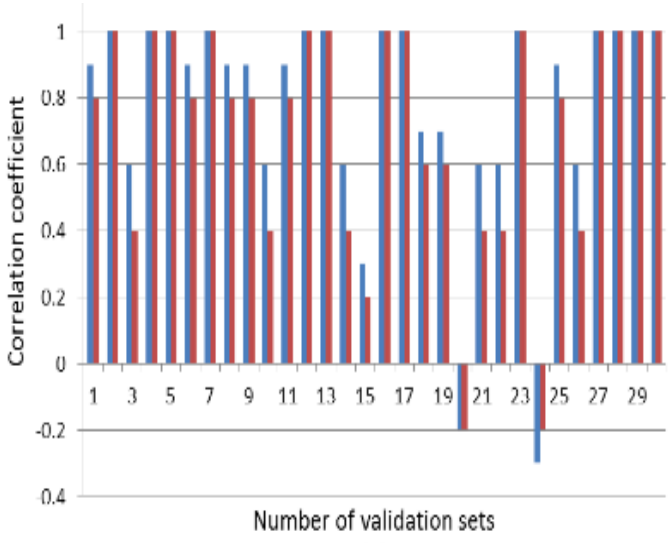


Comparison tests



Correlation

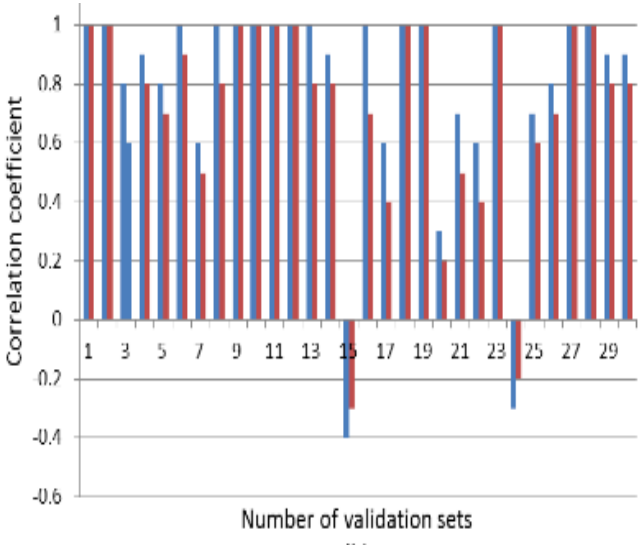
Correlation with Expert



Spearman

Kendall test

Correlation with MT subjects



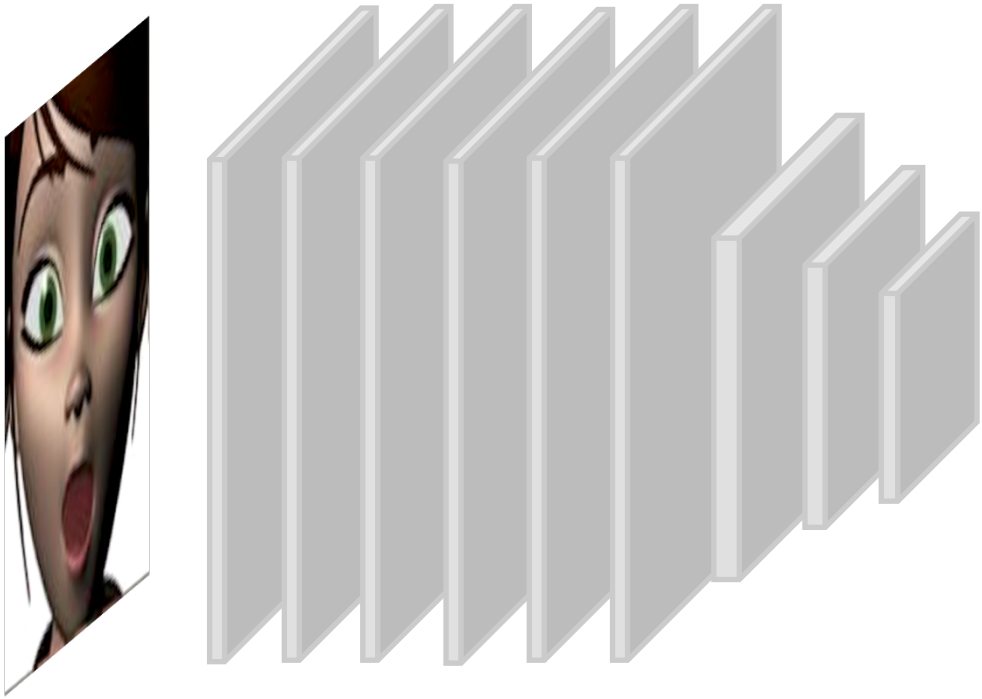
- Spearman correlation with expert best rank is **0.934** and with MT best rank is **0.942**
- Kendall correlation with expert best rank is **0.910** and with MT best rank is **0.927**

Conclusions

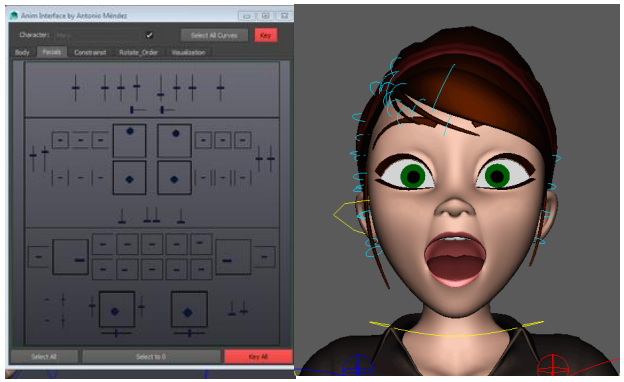
- **Perceptual** model of facial expressions
- **FERG-DB** with cardinal expression annotations
- **Plausible character expression** retrieval

- Improve visual storytelling applications:
  - Animated films
  - Gaming
  - Online marketing
  - VR/AR experiences
  - Robotics

Future work



2D feature vectors



3D Maya parameters

Thank you!

Project webpage
http://grail.cs.washington.edu/projects/deepexpr/

Stylized Character expression Database download
http://grail.cs.washington.edu/projects/deepexpr/ferg-db.html

Acknowledgements – Jamie Austad (animator), Zillow Group, the creators of the rigs we used in our project: Mery (www.meryproject.com), Ray (CGTarian Online School), Malcolm (www.animSchool.com), Aia & Jules (www.animationmentor. com), and Bonnie (Josh Sobel Rigs).