Face Landmarks

and stuff you can do with them

CS448V — Computational Video Manipulation

May 2019

Constrained Local Models

Improving Portraits

Editing Video

Constrained Local Models

Constrained Local Models

Deformable Model Fitting by Regularized Landmark Mean-Shift

Jason M. Saragih · Simon Lucey · Jeffrey F. Cohn

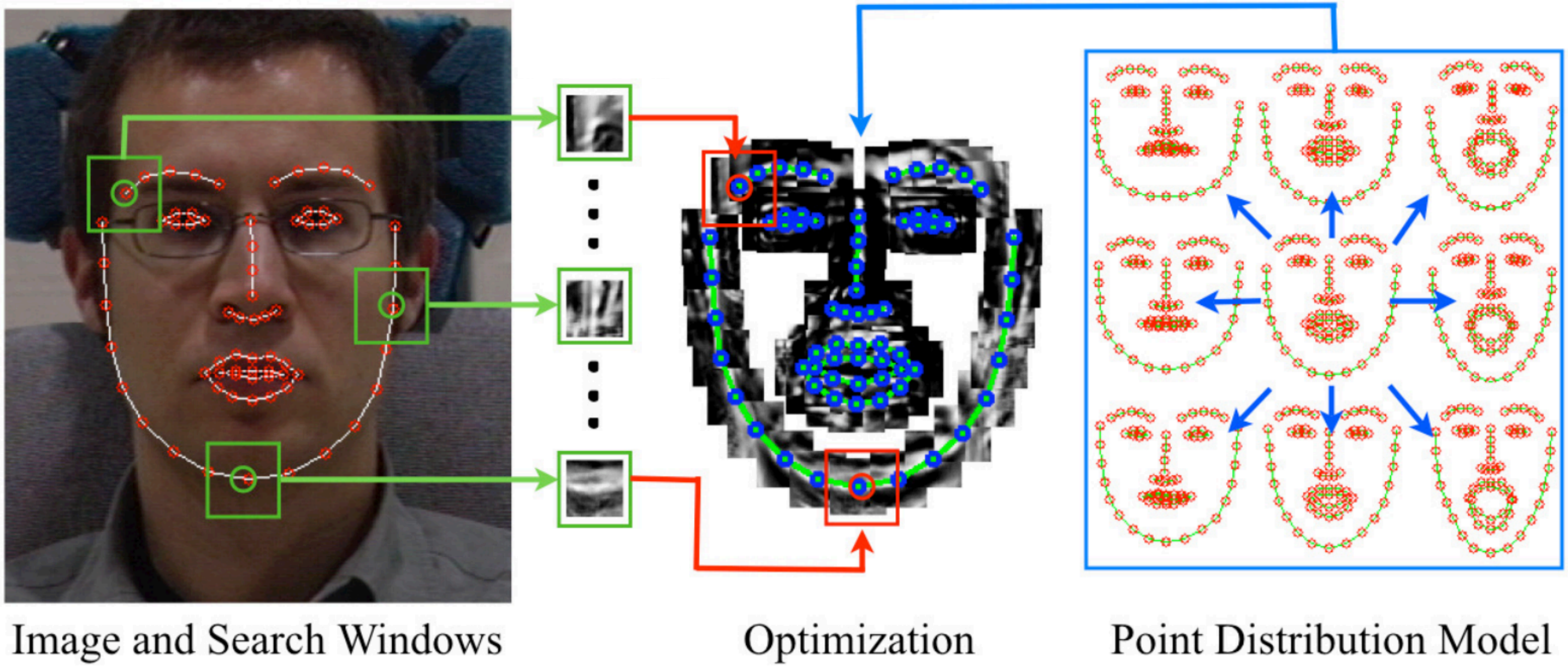
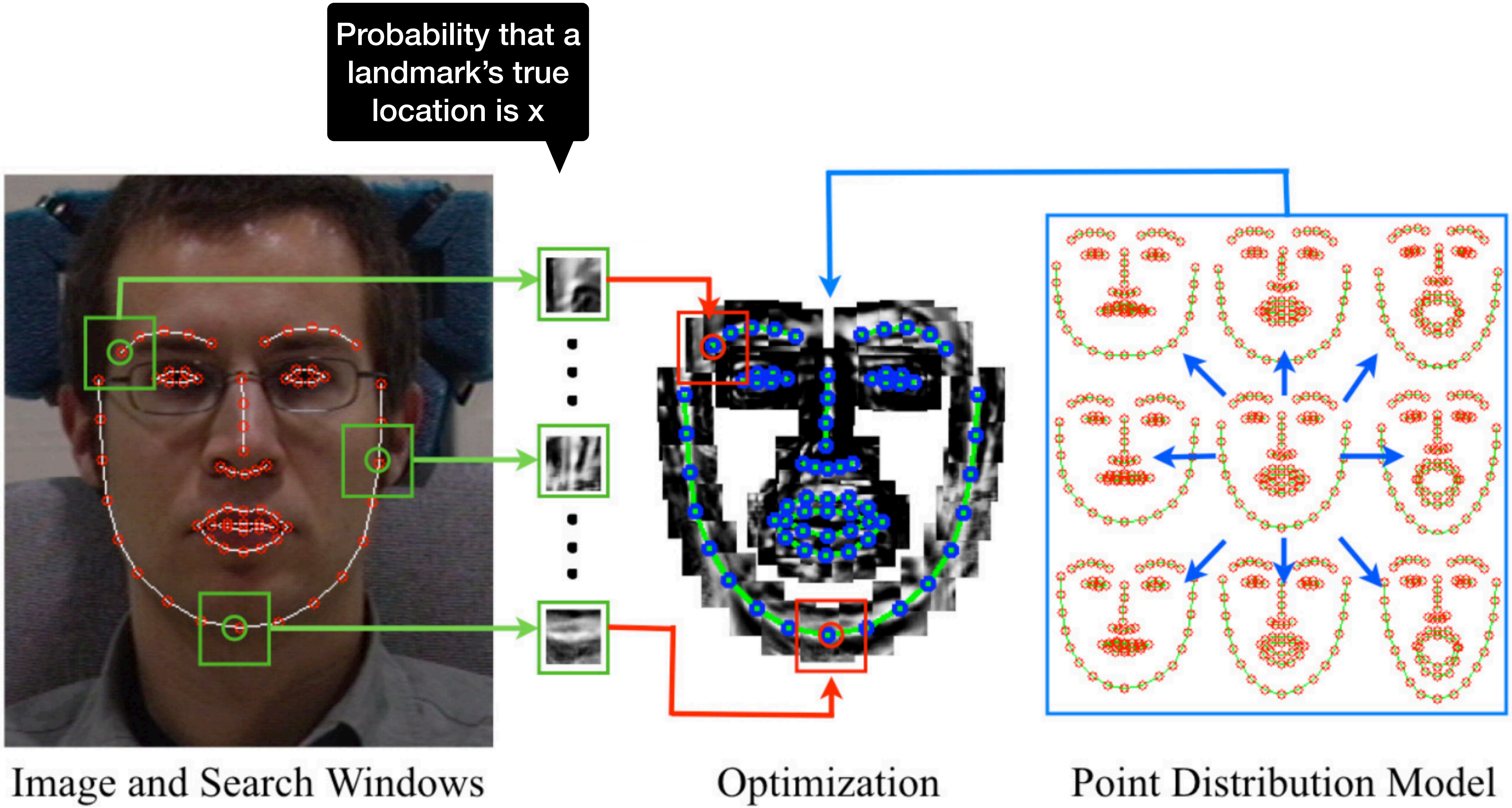


Image and Search Windows

Optimization

Point Distribution Model


Probability that a landmark's true location is x
Image and Search Windows
Optimization
Point Distribution Model

Probability that a landmark's true location is x

Change current (independent) estimation for landmark location

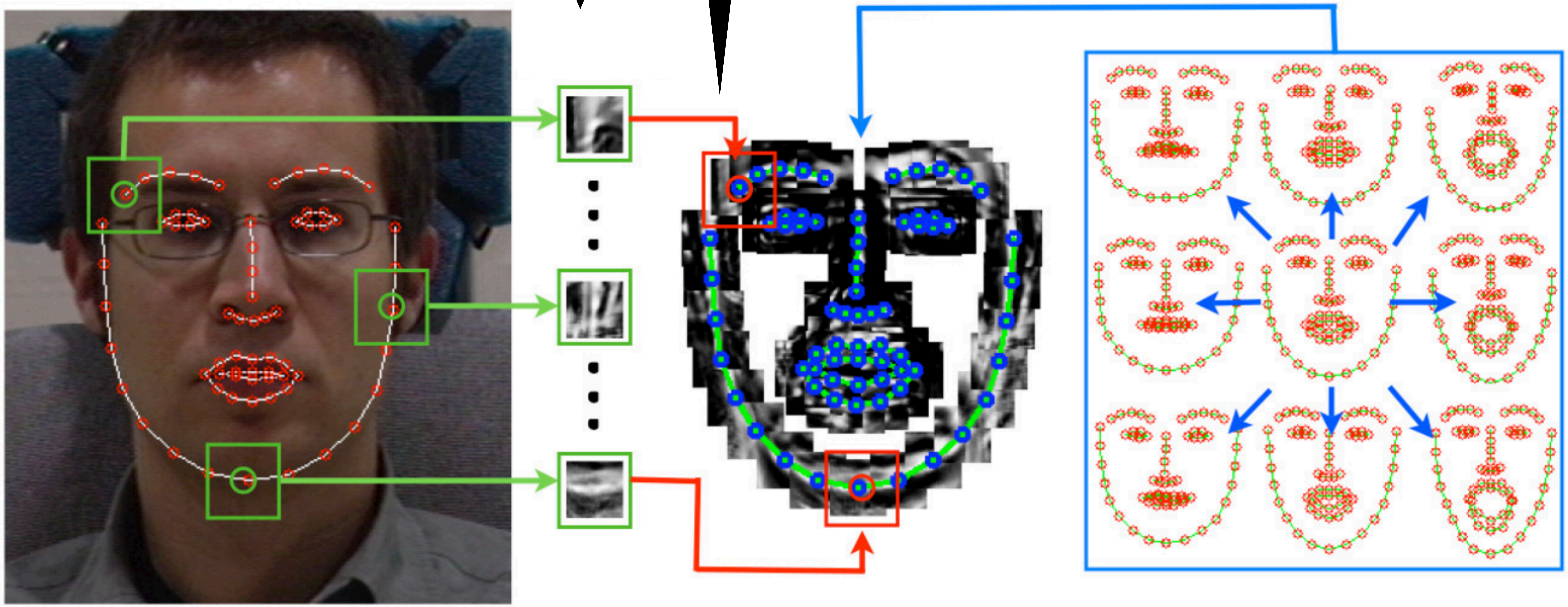


Image and Search Windows

Optimization

Point Distribution Model

Probability that a landmark’s true location is x

Change current (independent) estimation for landmark location

Change current (global) estimation for parameterized landmark model

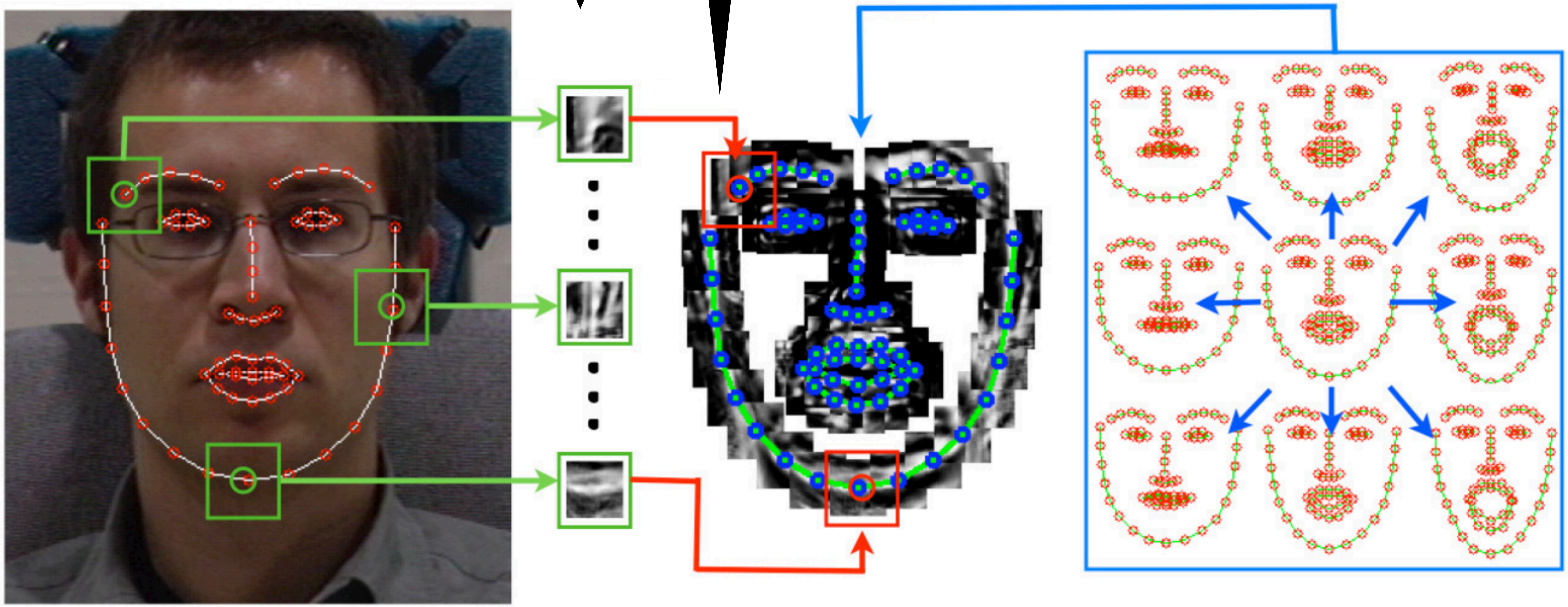
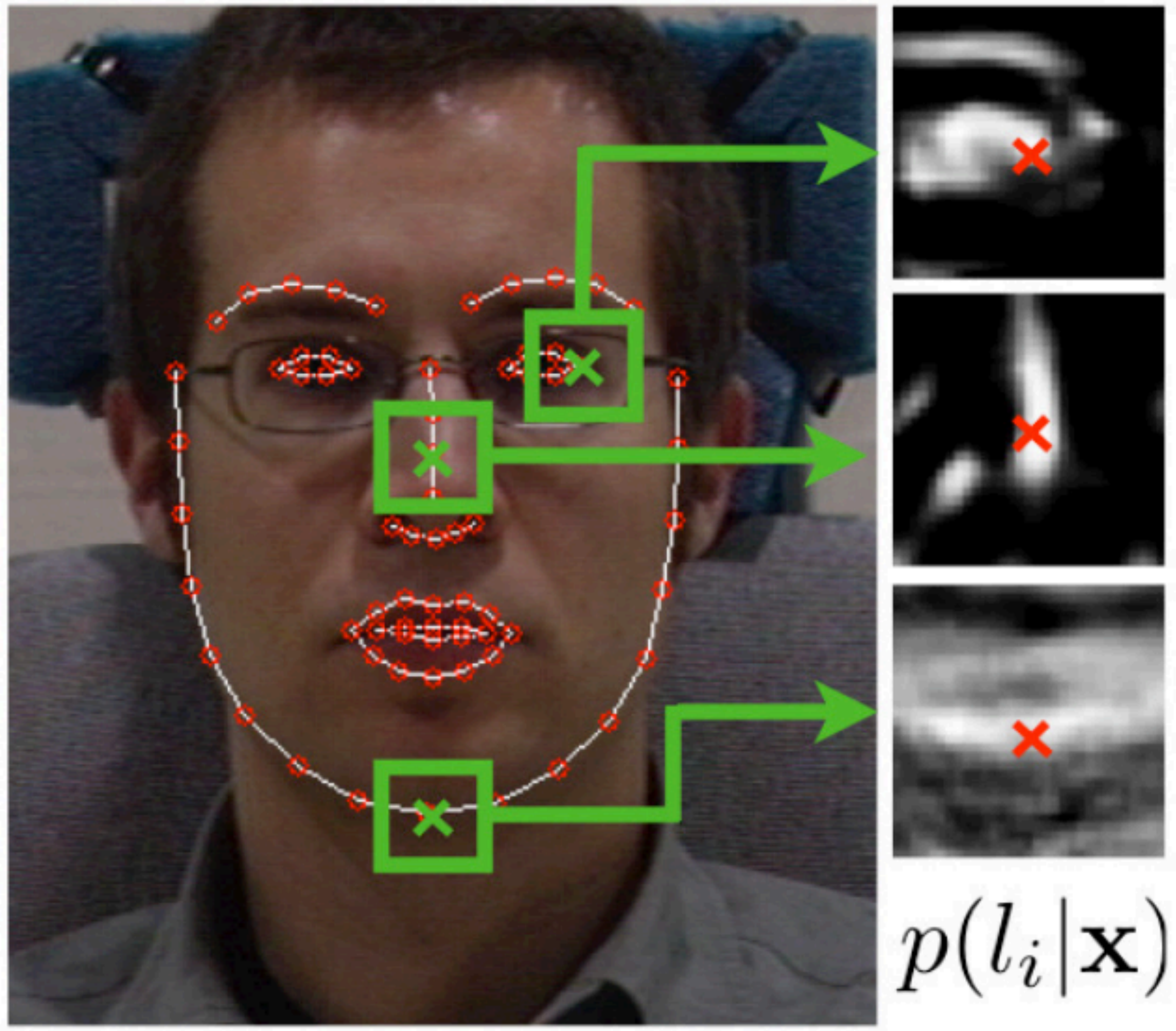


Image and Search Windows

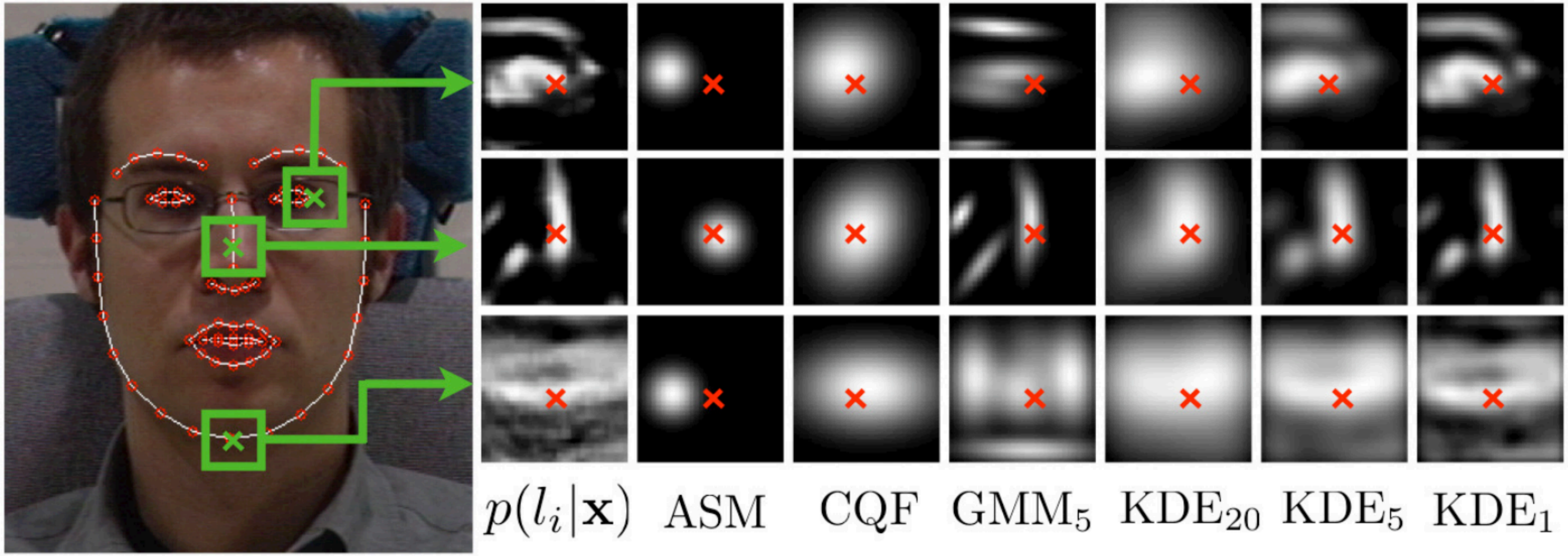
Optimization

Point Distribution Model

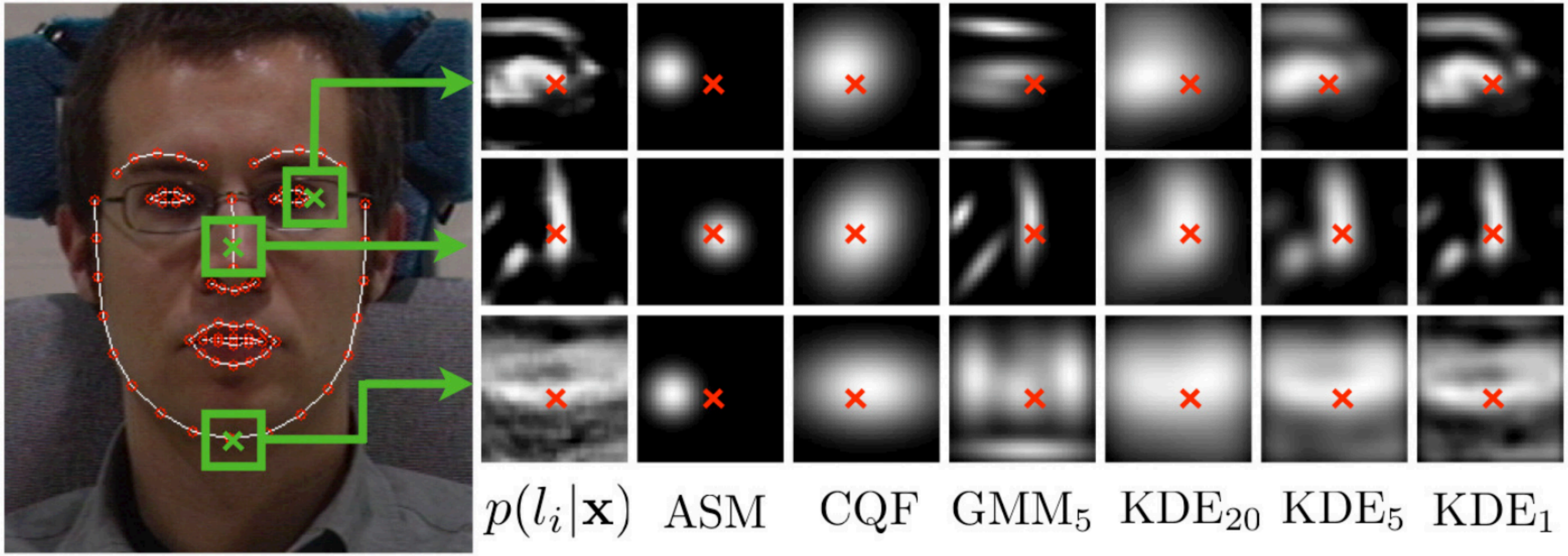
Landmark probabilities

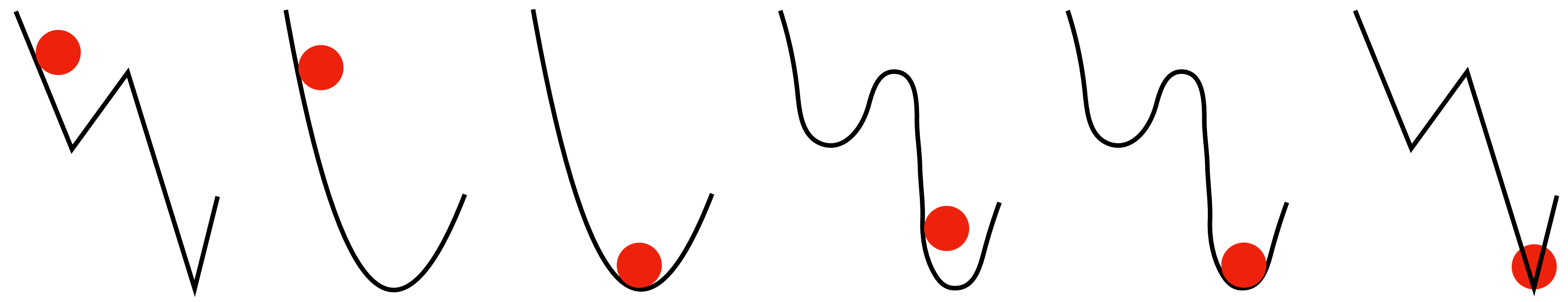


Landmark probabilities

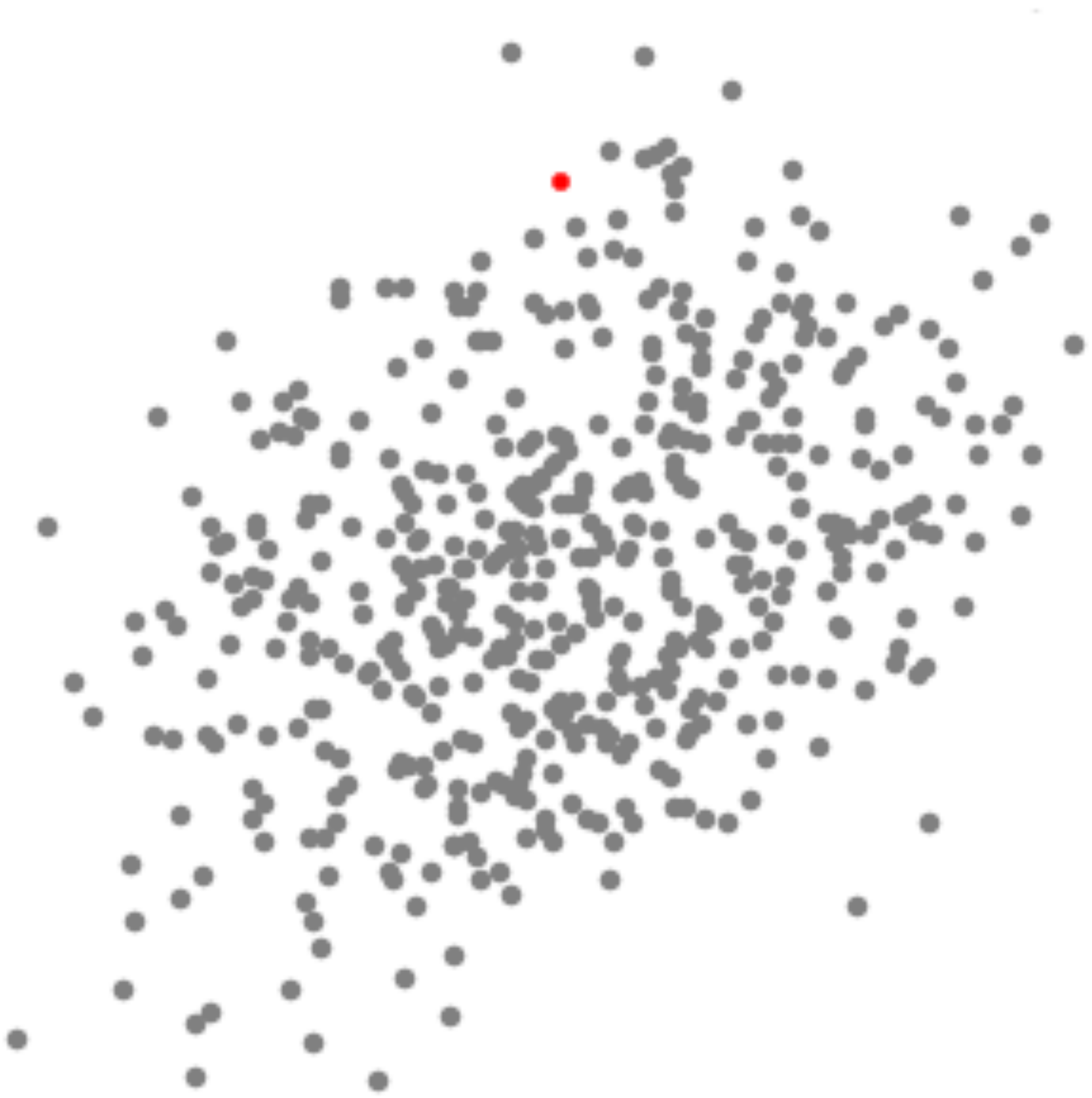


Landmark probabilities



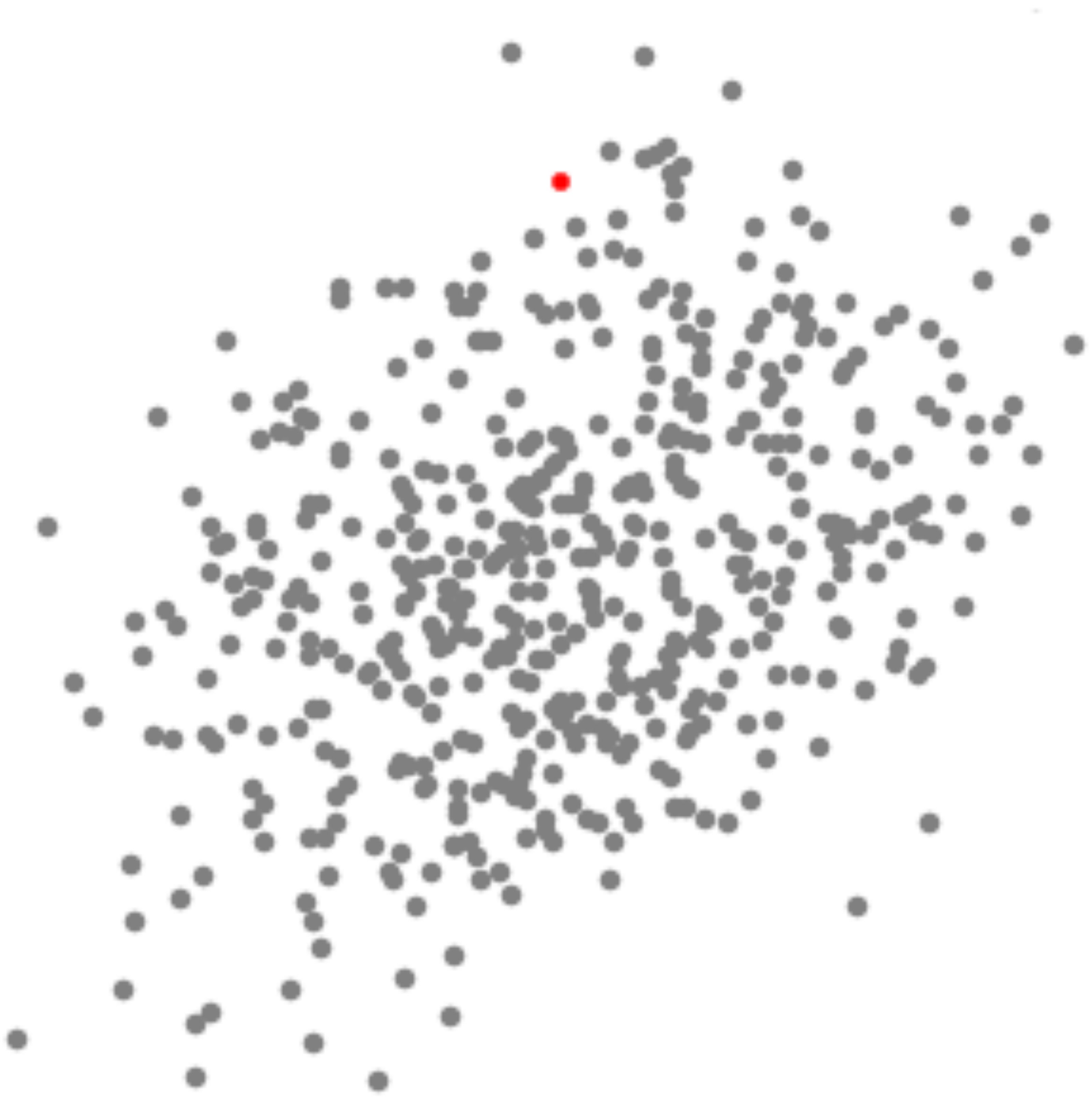


Mean shift



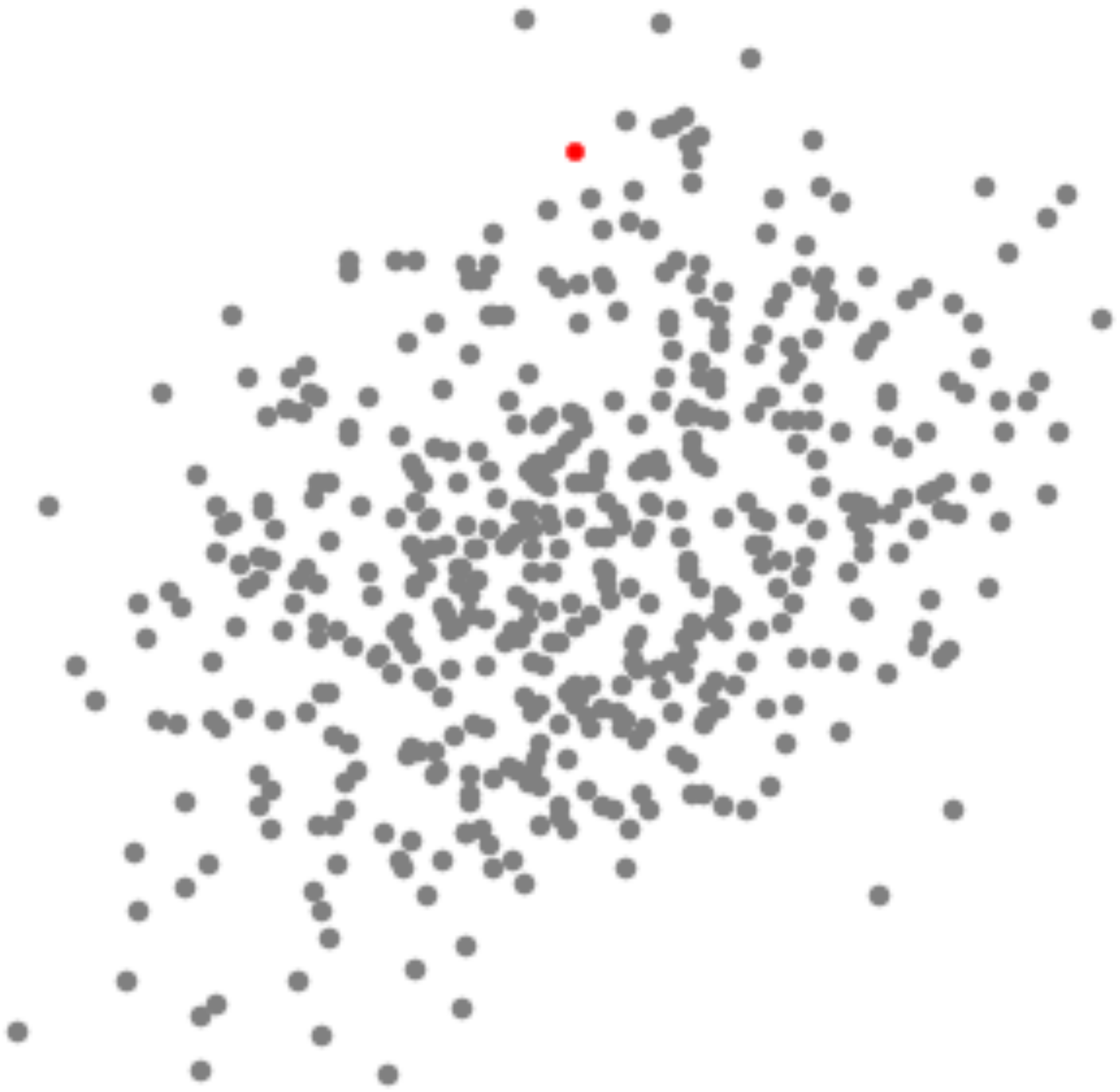
GIF by David Sheehan

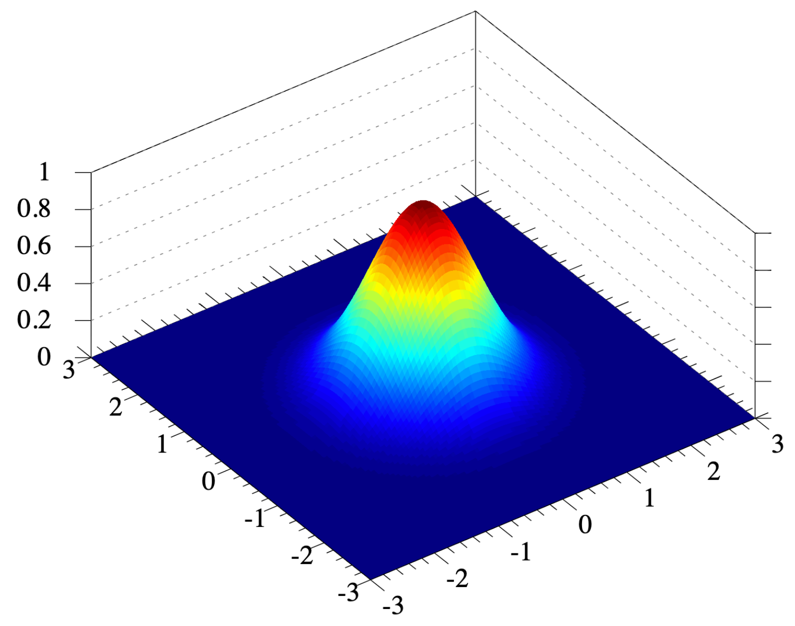
Mean shift



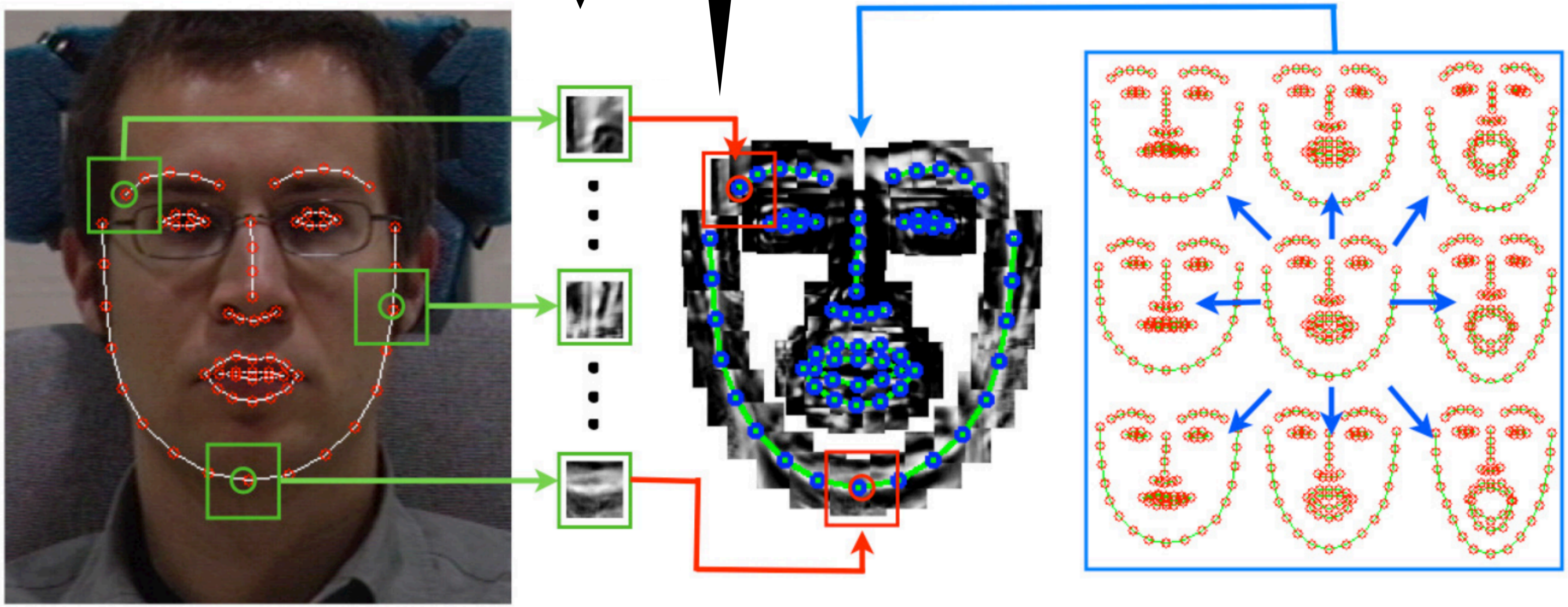
GIF by David Sheehan

Mean shift





GIF by David Sheehan



Probability that a landmark's true location is x

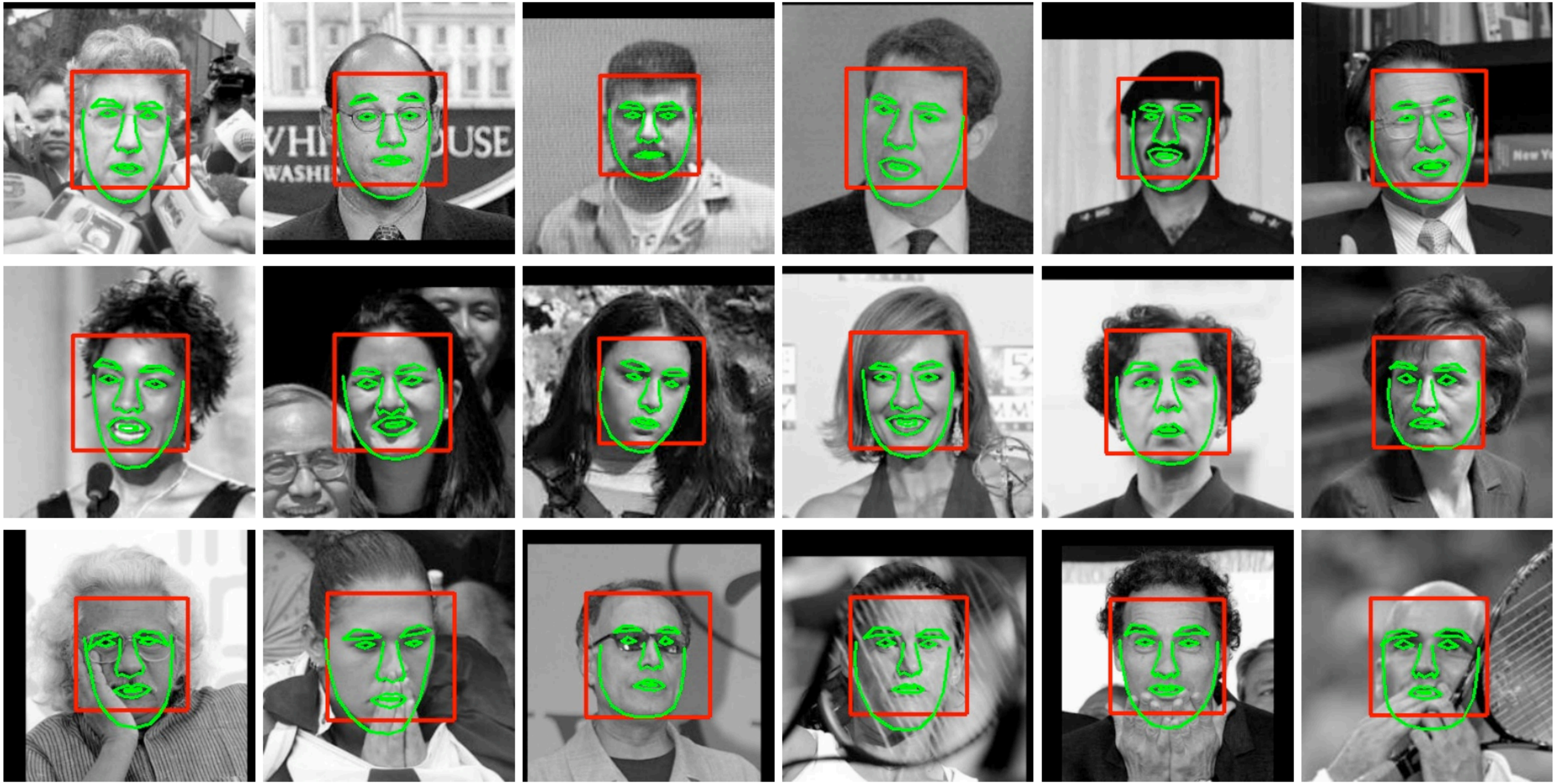
Change current (independent) estimation for landmark location

Change current (global) estimation for parameterized landmark model

Image and Search Windows

Optimization

Point Distribution Model


WHI
USE
WASHI

What can you do with sparse face landmarks?

What can you do with sparse face landmarks?

What can you do with sparse face landmarks?

Bootstrap a better model

What can you do with sparse face landmarks?

Bootstrap a better model

dense

What can you do with sparse face landmarks?

Bootstrap a better model

dense

3D

What can you do with sparse face landmarks?

Bootstrap a better model

dense

3D

richer (albedo, lighting, ...)

What can you do with sparse face landmarks?

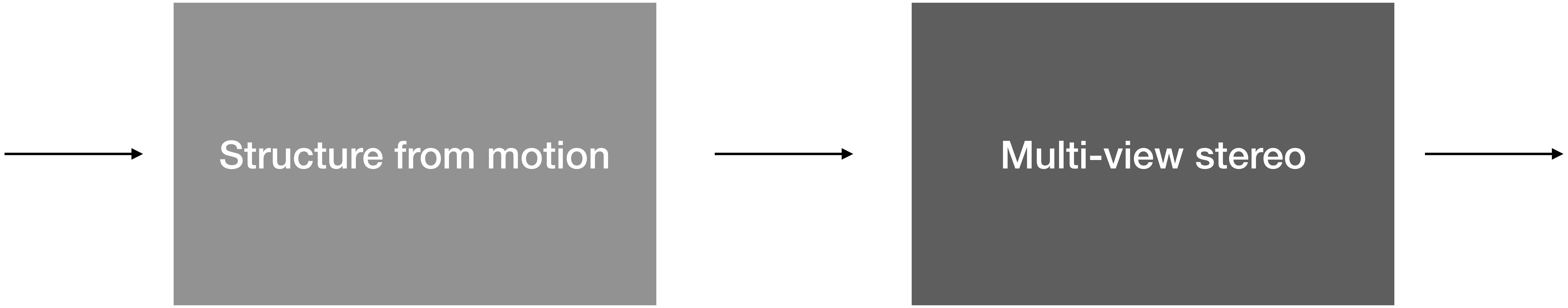
Bootstrap a better model

dense

3D

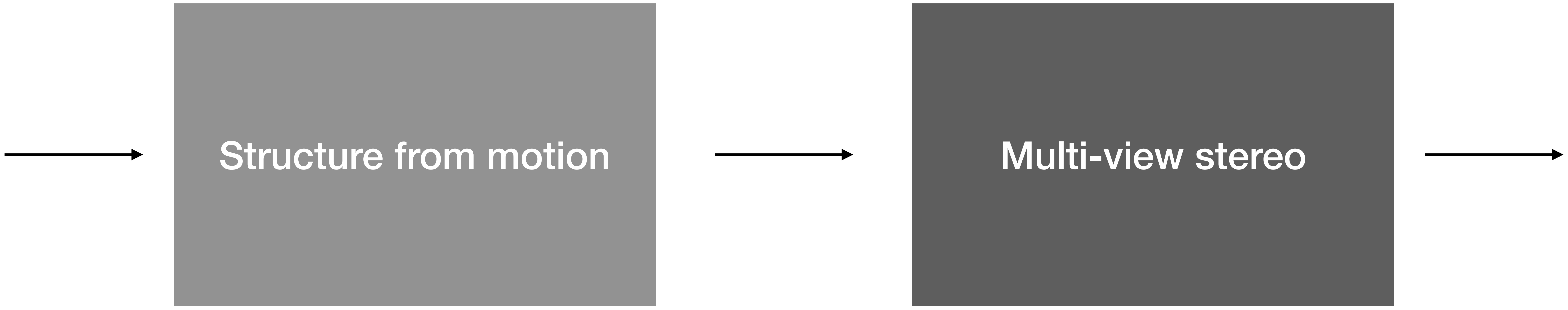
richer (albedo, lighting, ...)

Sounds familiar?



Structure from motion

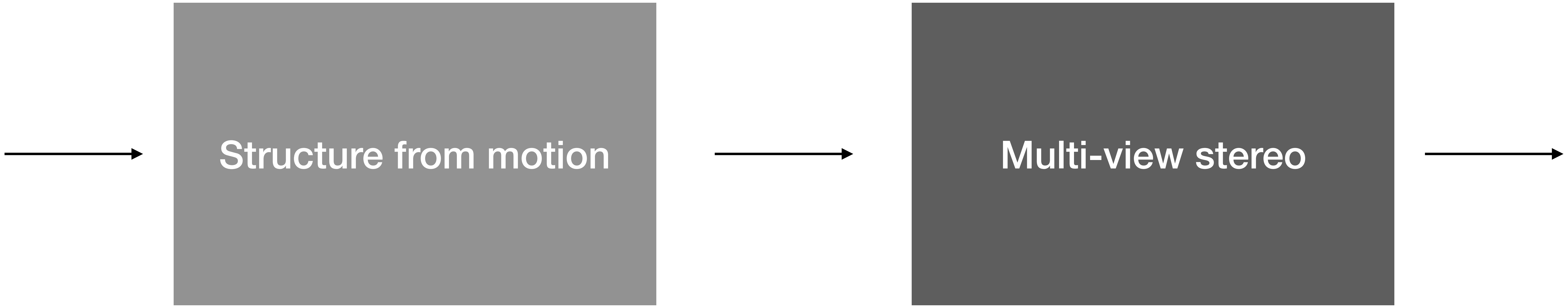
Multi-view stereo



Structure from motion

Multi-view stereo

Sparse points that we "trust"

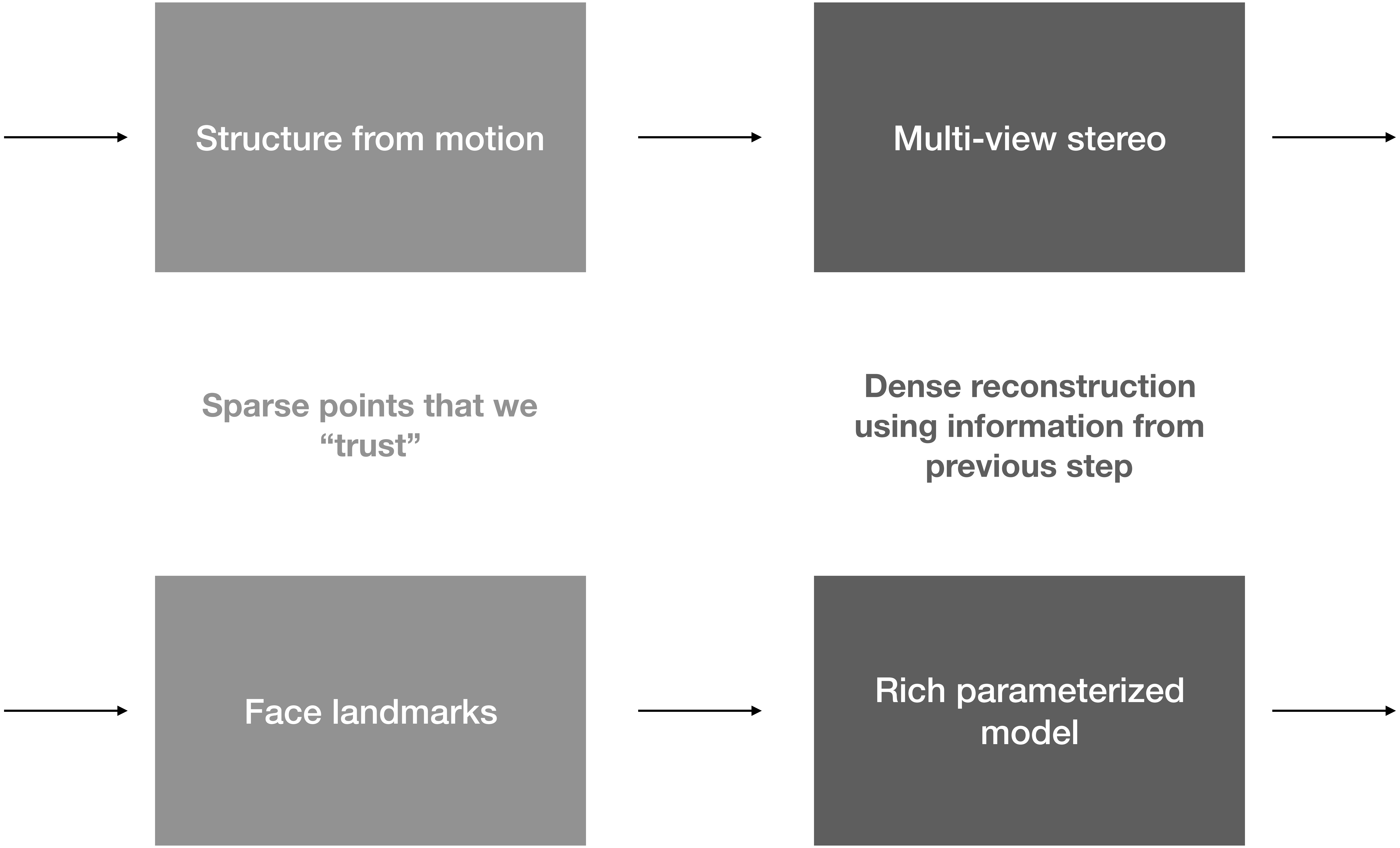


Structure from motion

Multi-view stereo

Sparse points that we "trust"

Dense reconstruction using information from previous step



Structure from motion

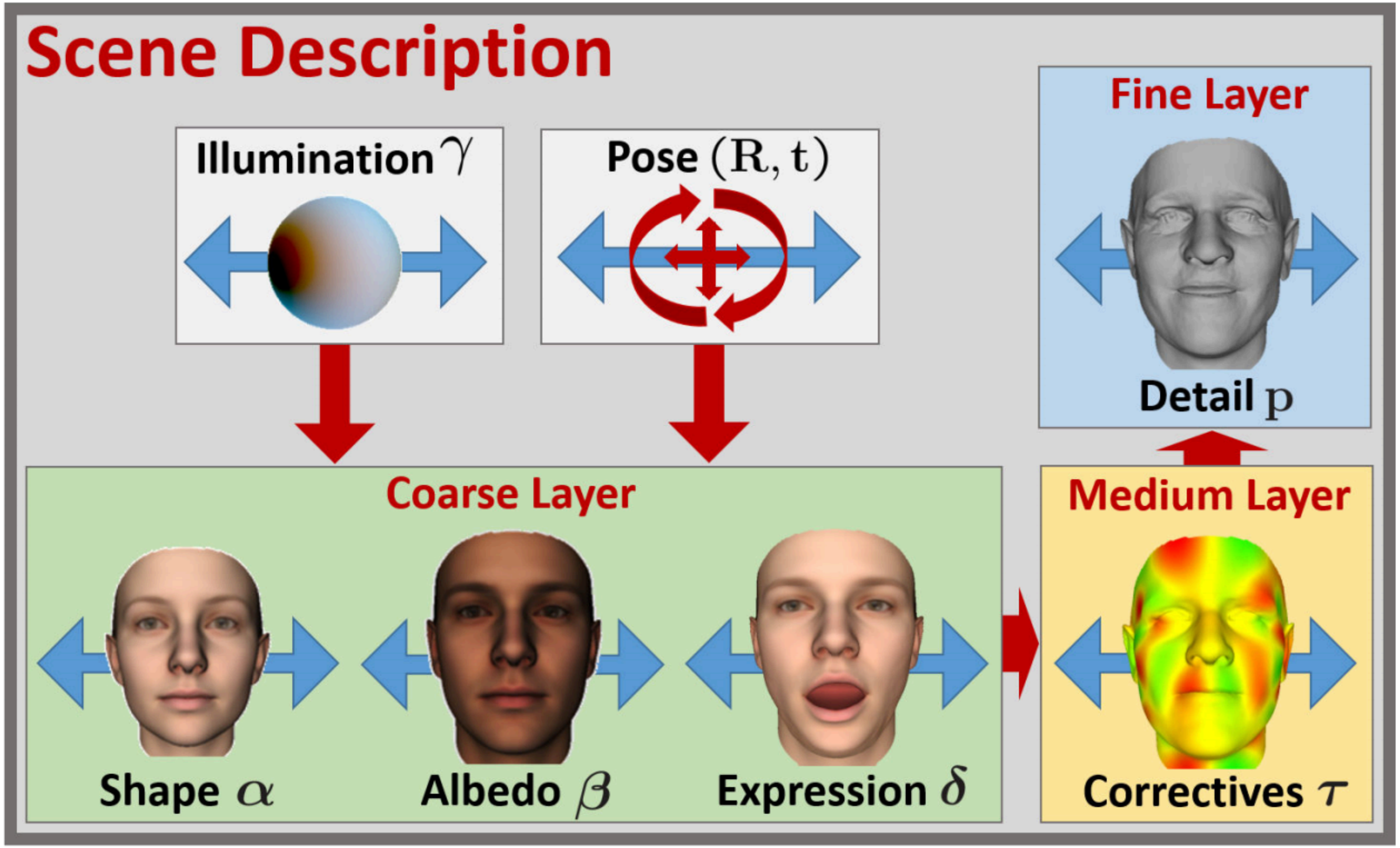
Multi-view stereo

Sparse points that we “trust”

Dense reconstruction using information from previous step

Face landmarks

Rich parameterized model



[Garrido et al. 2015]

You can also use deep learning…

Constrained Local Models

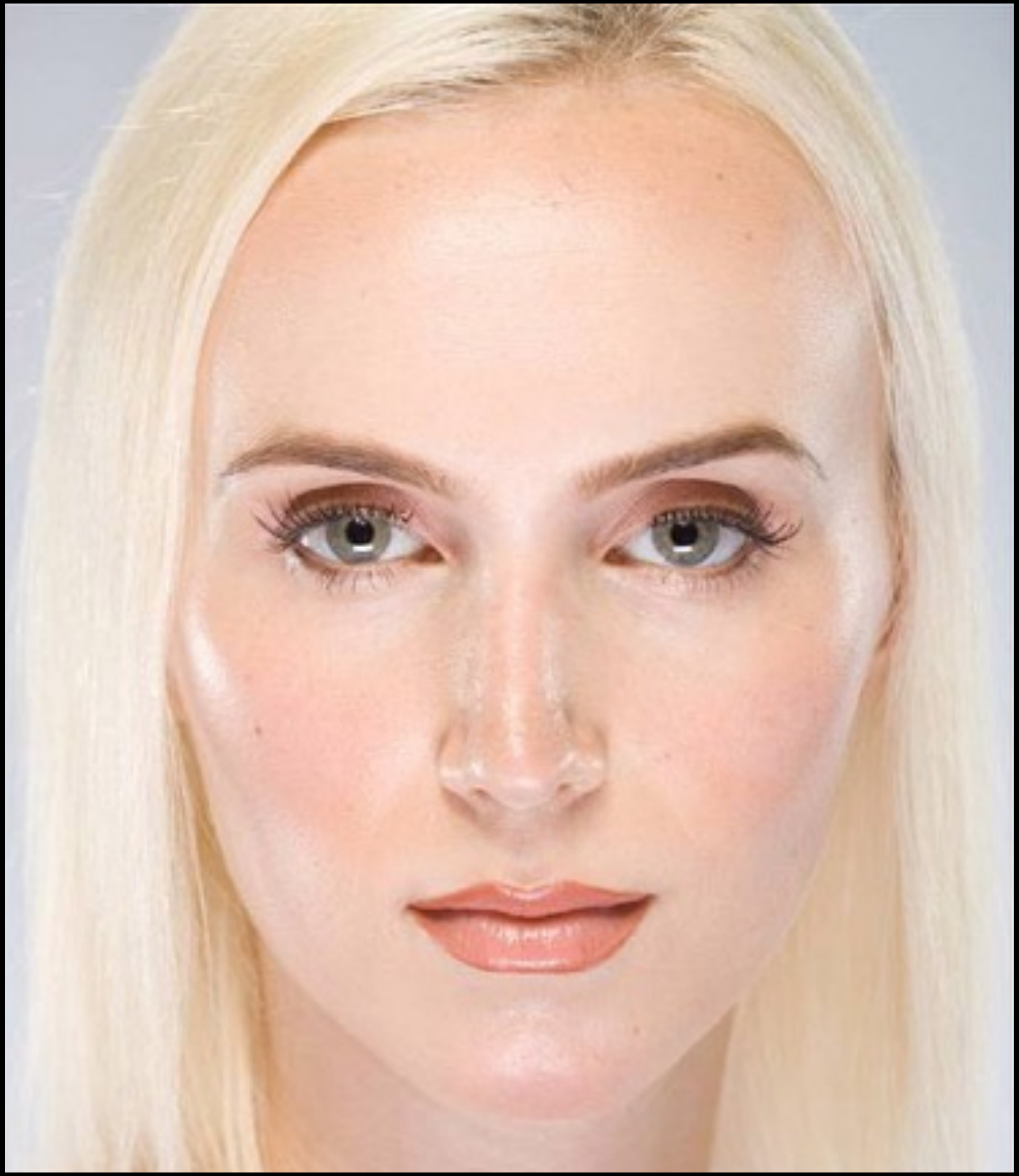
Improving Portraits

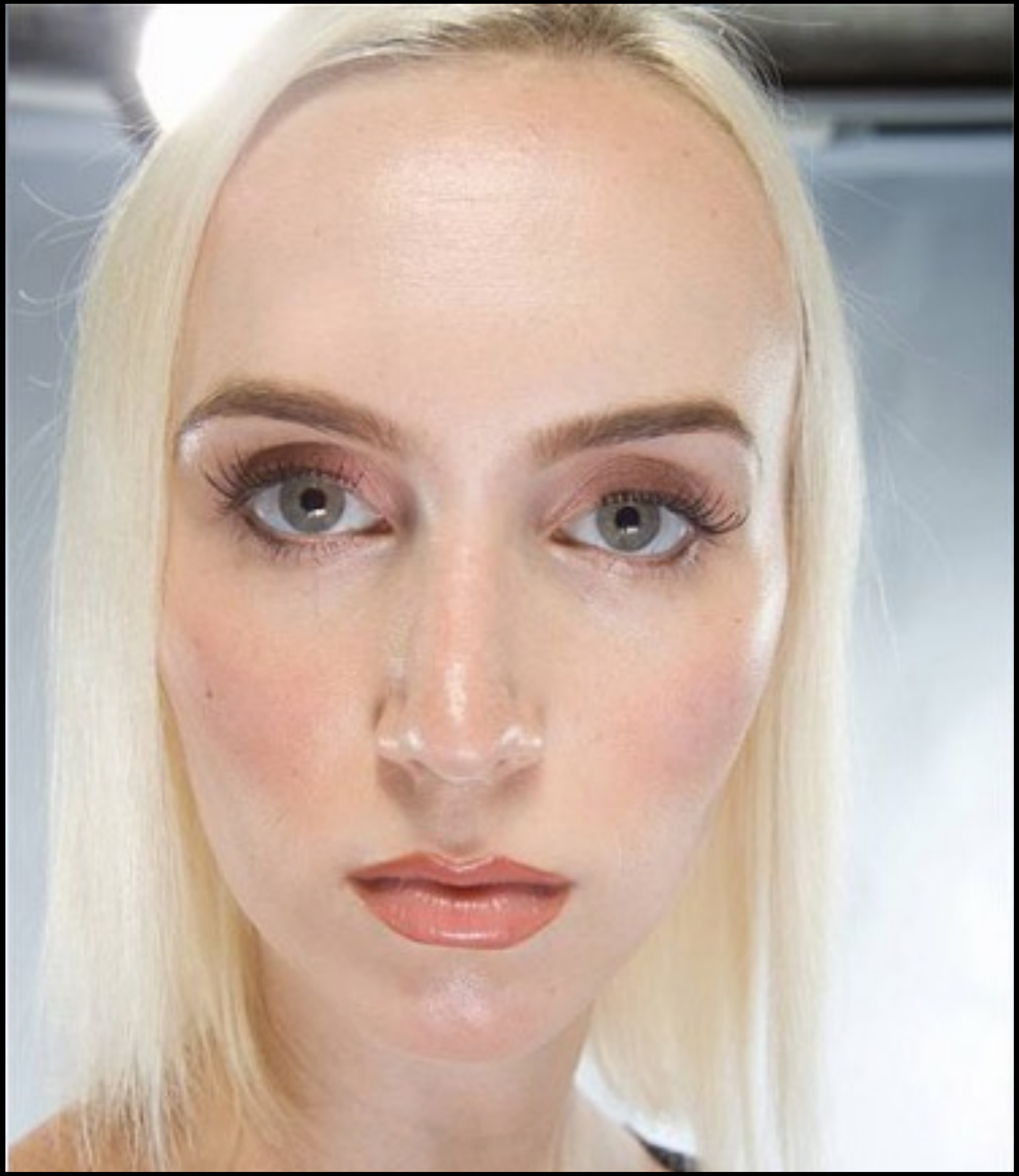
Editing Video

Improving Portraits

Improving Portraits

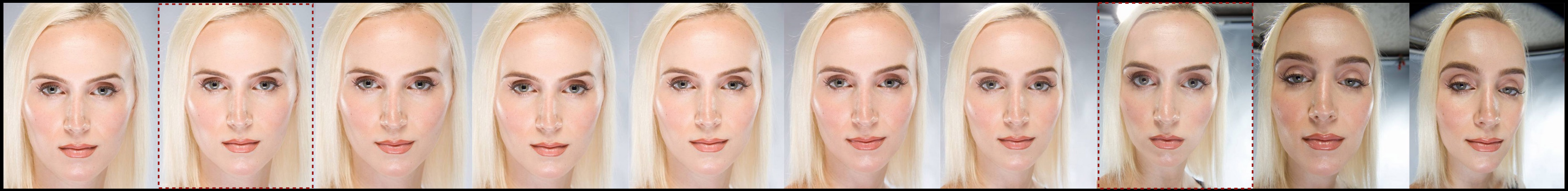
Distance Matters





photos from http://stepheneastwood.com/

Distance Matters



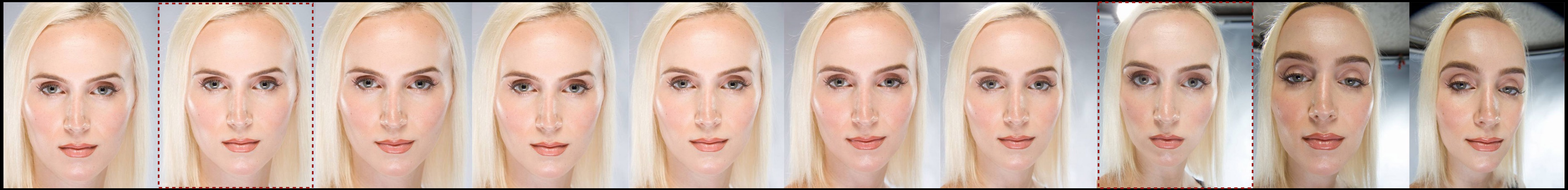
← Pro Photographer →

Uses telephoto lens

← Selfie →

Limited by arm length

Distance Matters



Close → “approachable”

Far → “impressive”

[Perona 2007]

Camera placement is hard

Camera placement is hard

- Lack of equipment

© Juza - www.juzaphoto.com

Camera placement is hard

- Lack of equipment

- Lack of expertise

Camera placement is hard

- Lack of equipment

- Lack of expertise
- Ephemeral moment

Camera placement is hard

- Lack of equipment

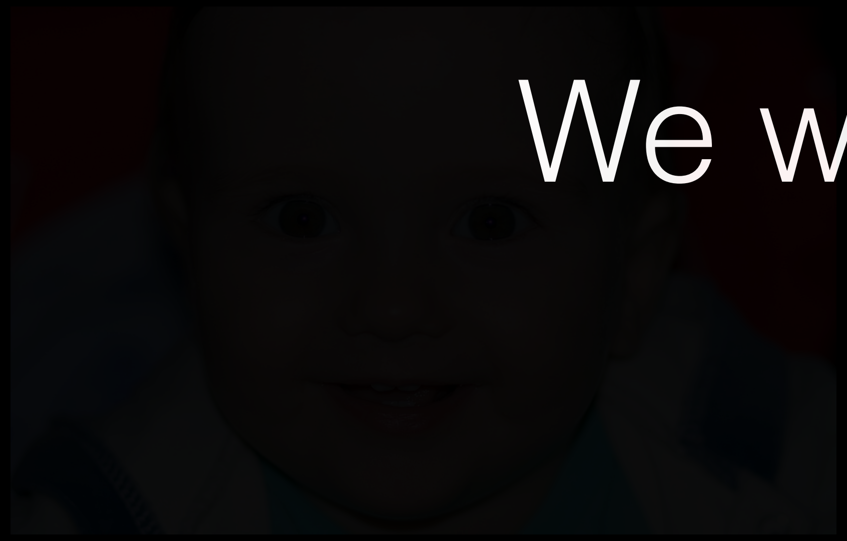
- Lack of expertise

- Ephemeral moment

- Physical constraints

It is hard to capture the perfect photo…

- Lack of equipment



- Momentary event

- Passersby
- Lack of knowledge

We would like to move the camera in
post processing

Demo

Goals

Virtually move perspective camera

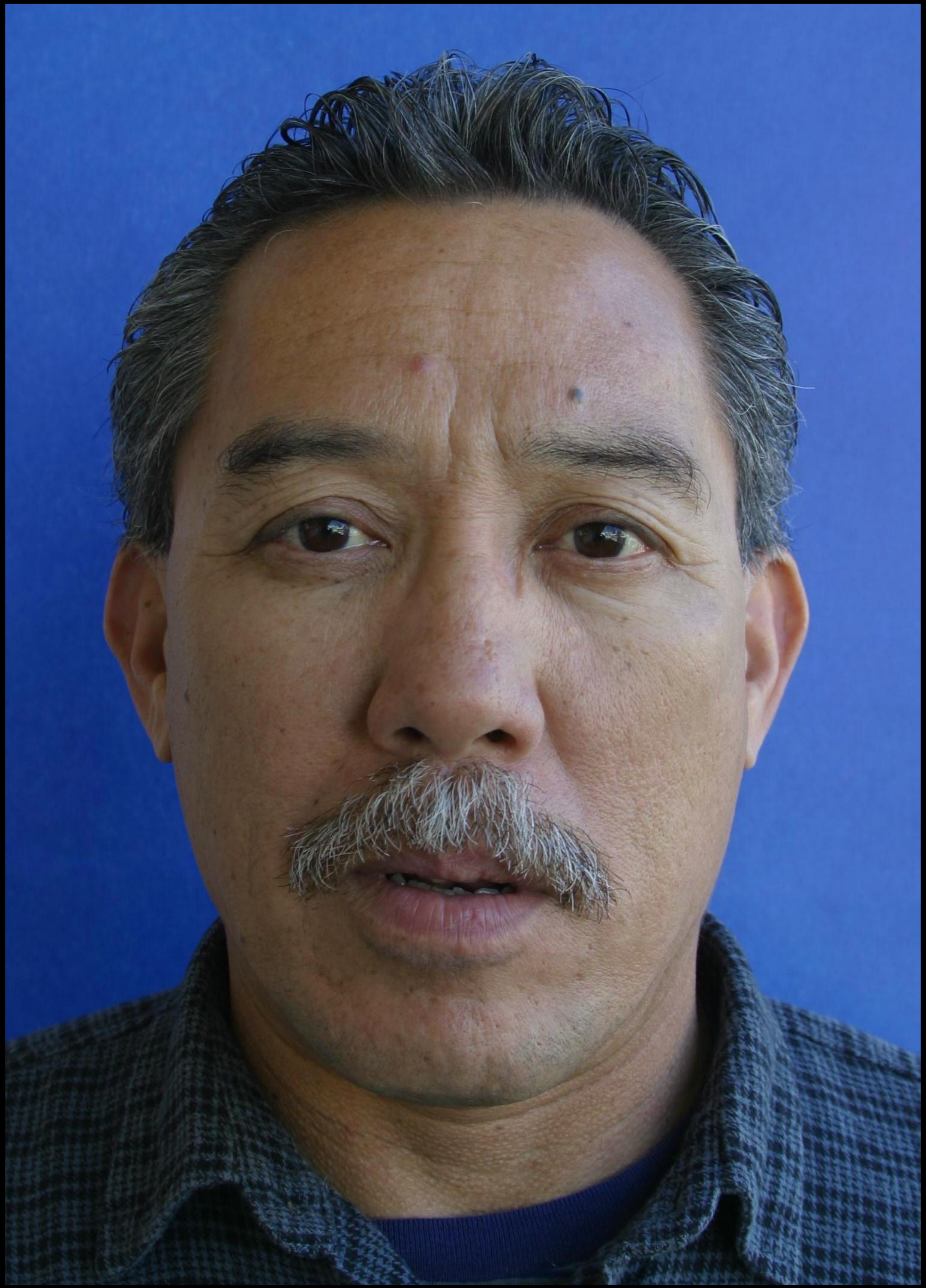
- From single image
- Results similar to ground truth
- No background artifacts
- No seams

Goals

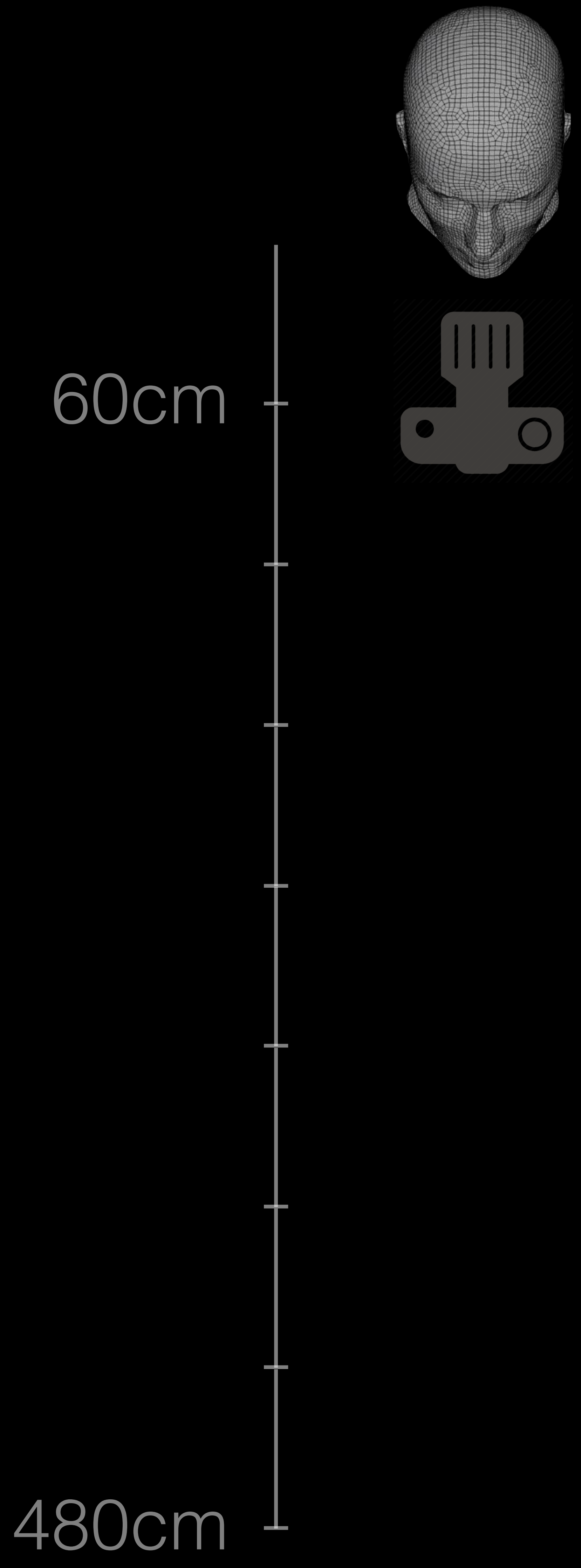
Virtually move perspective camera

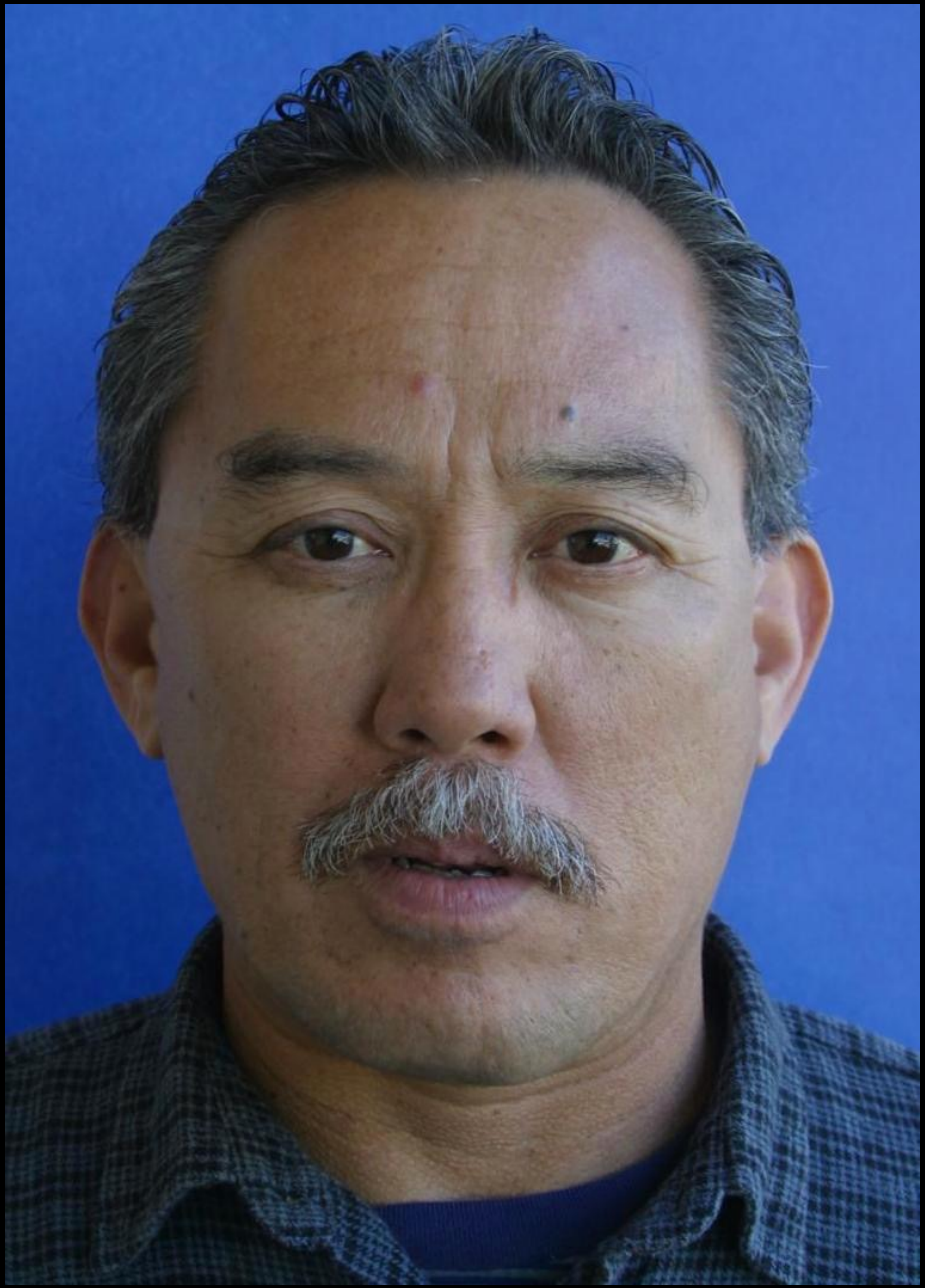
- From single image
- Results similar to ground truth
- No background artifacts
- No seams

Dataset:
[Burgos-Artizzu et al. 2014]

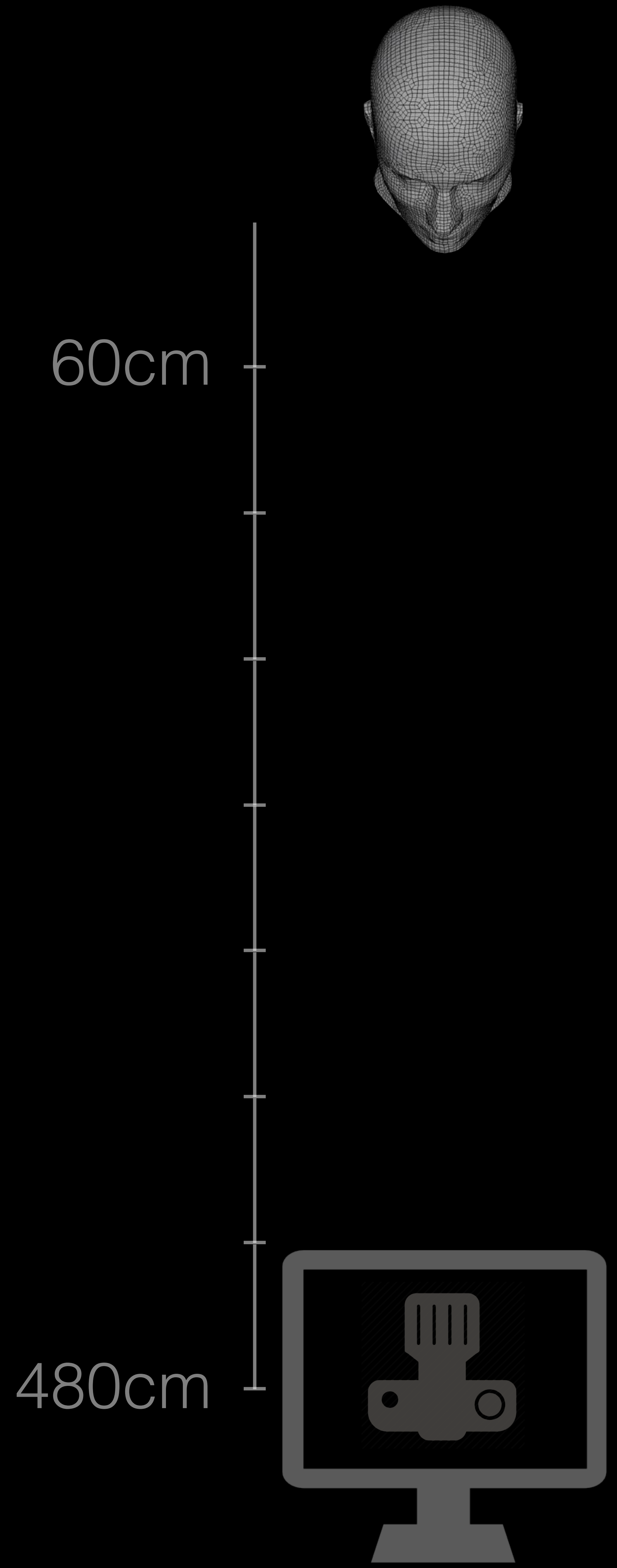


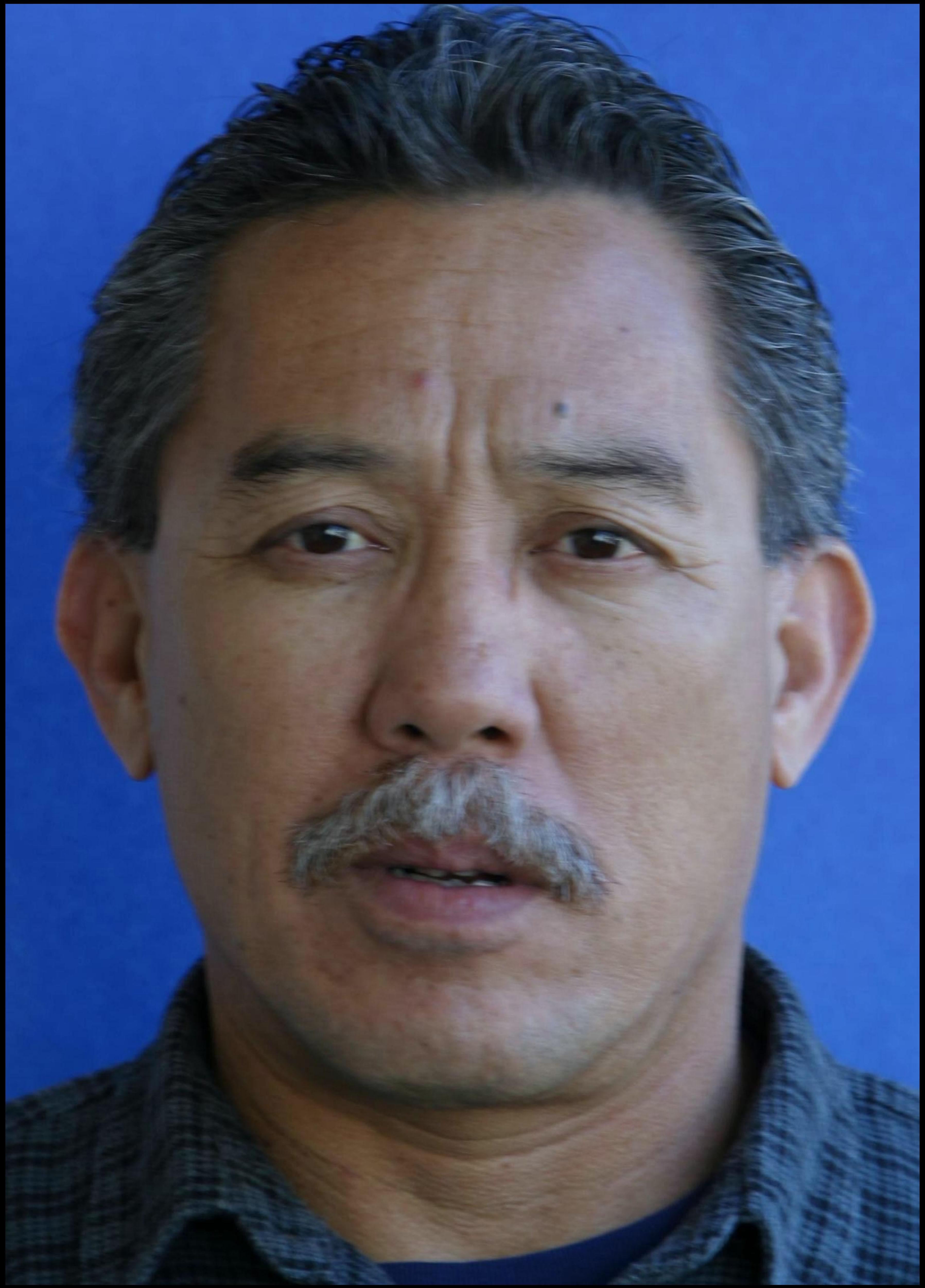
Input 60cm



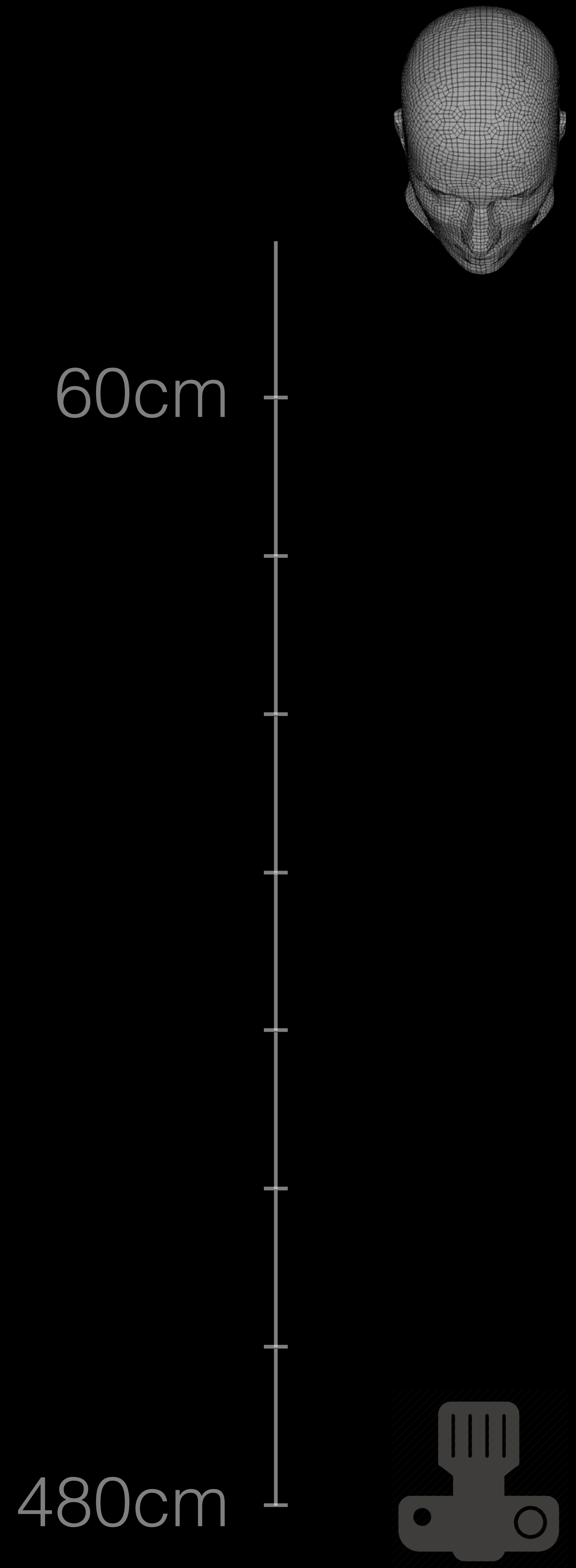


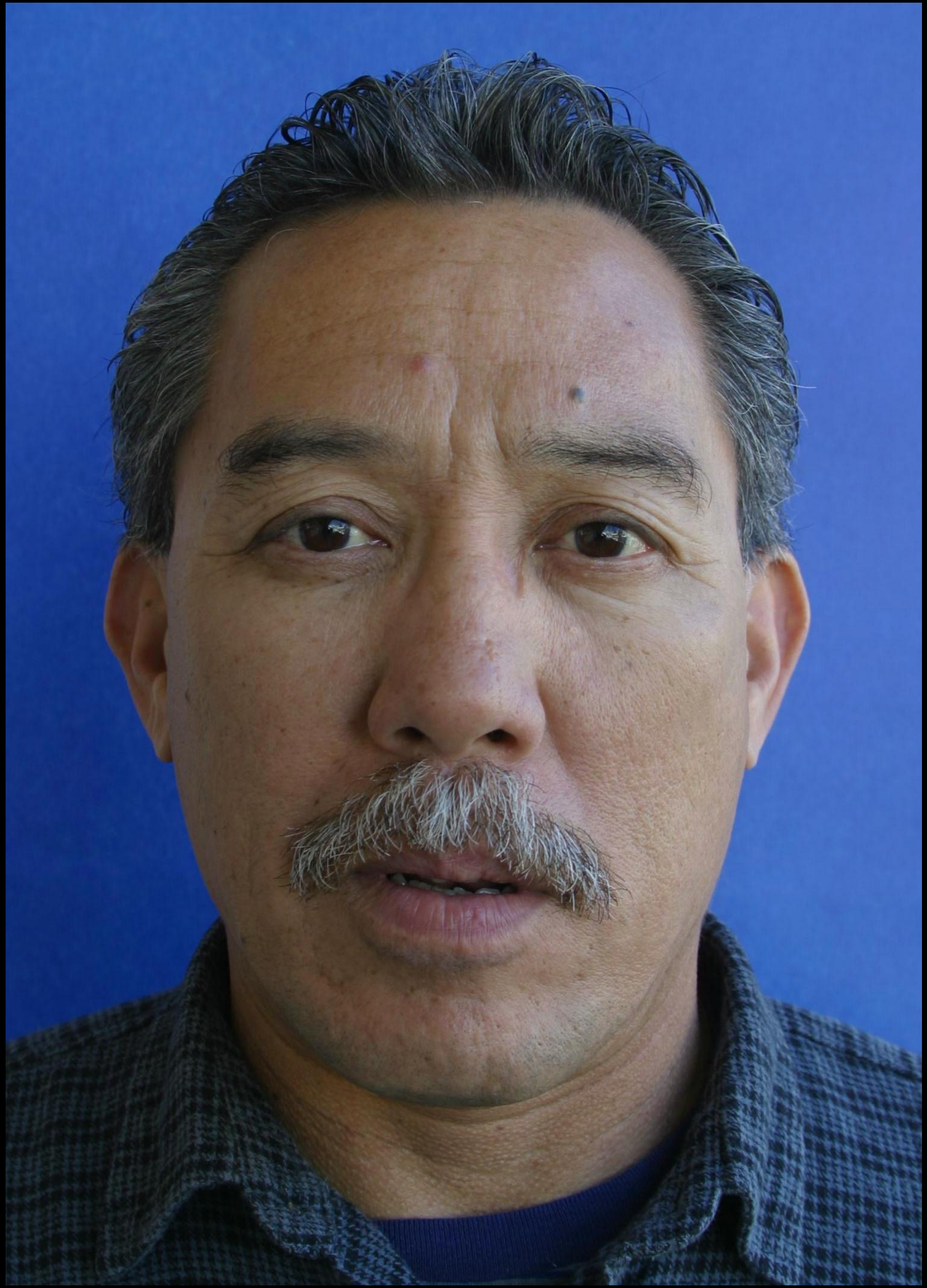
Simulated 480cm



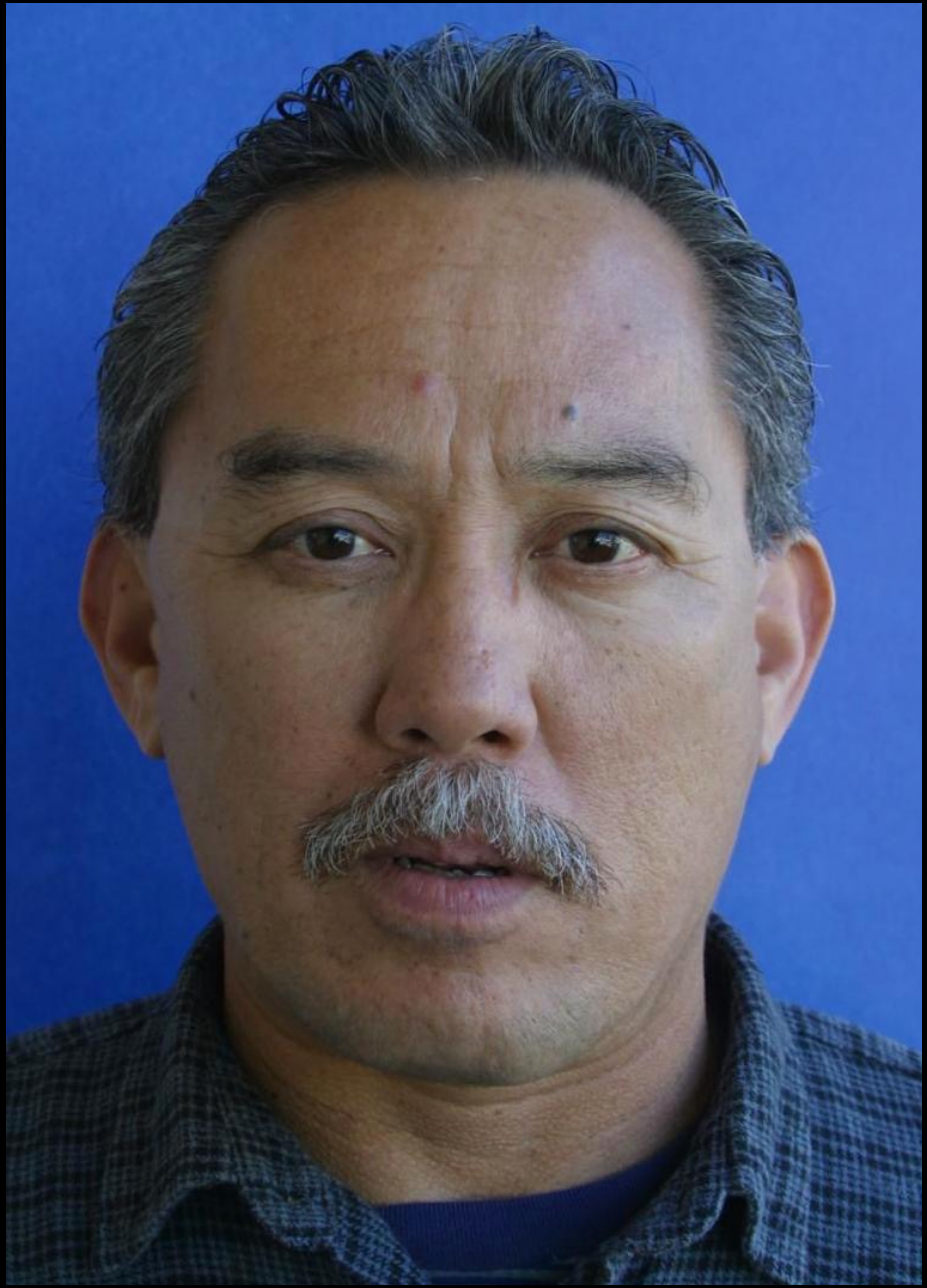


Ground Truth 480cm

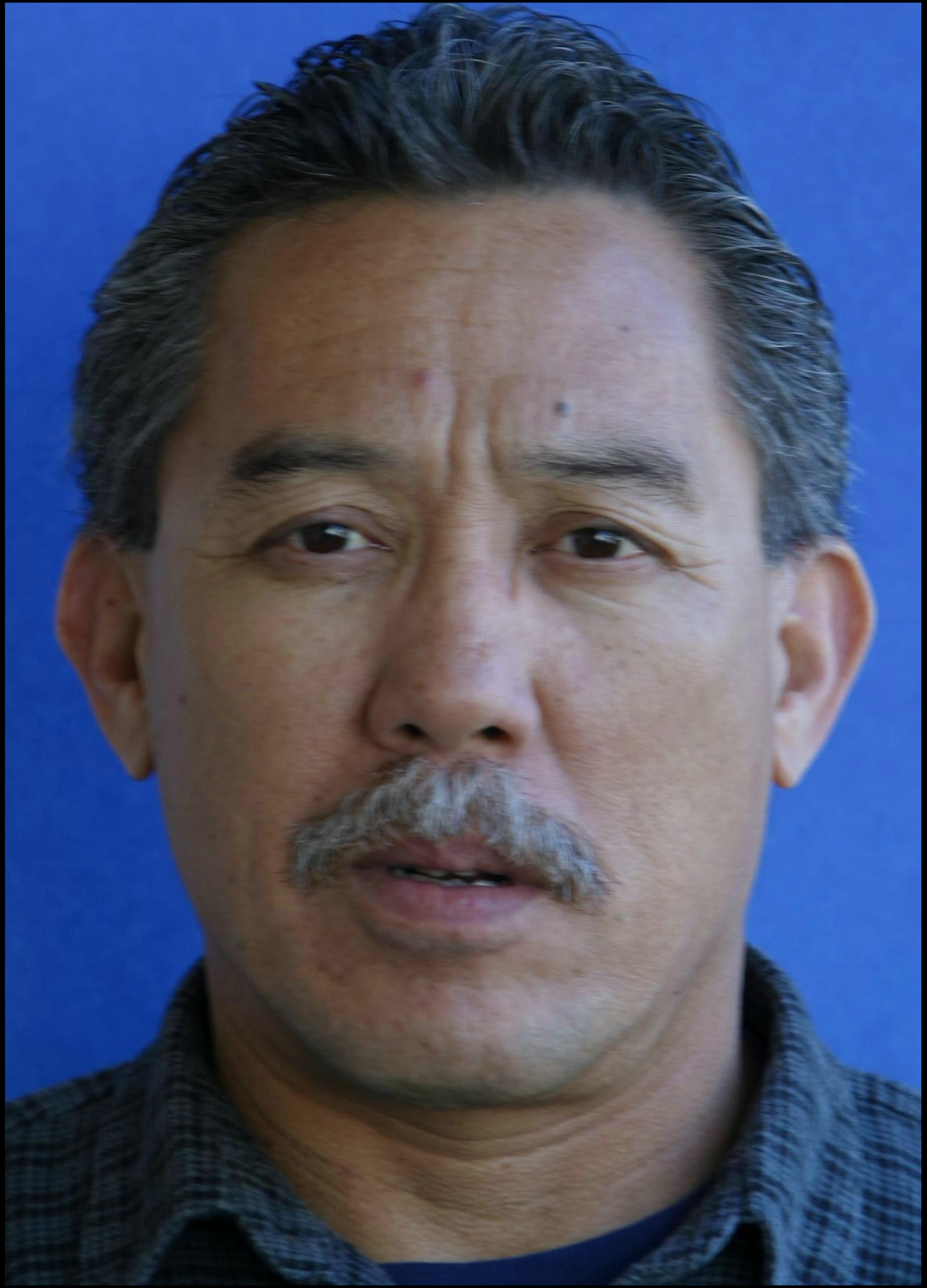




Input 60cm



Simulated 480cm



Ground Truth 480cm

Key Contributions

Key Contributions

- New application
    - Timely: “year of the selfie”

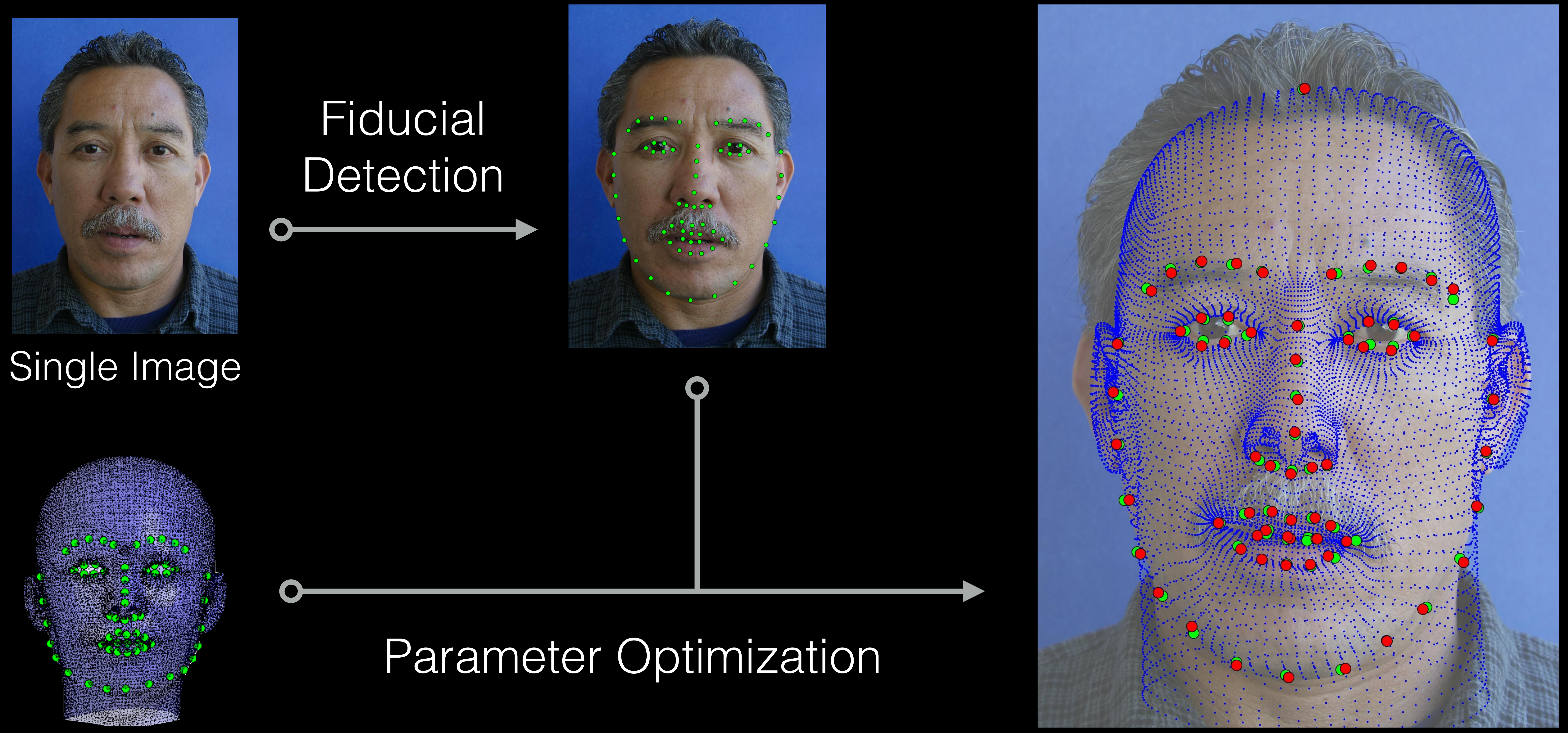
Key Contributions

- New application
    - Timely: “year of the selfie”
- New optimization framework
    - Full perspective model
    - Robust for single input photo

Key Contributions

- New application
  - Timely: “year of the selfie”
- New optimization framework
  - Full perspective model
  - Robust for single input photo
- Fit in 3D, warp in 2D (following, e.g. [Yang ’11])
  - Warp field works for perspective model

Part I: Fitting



Input

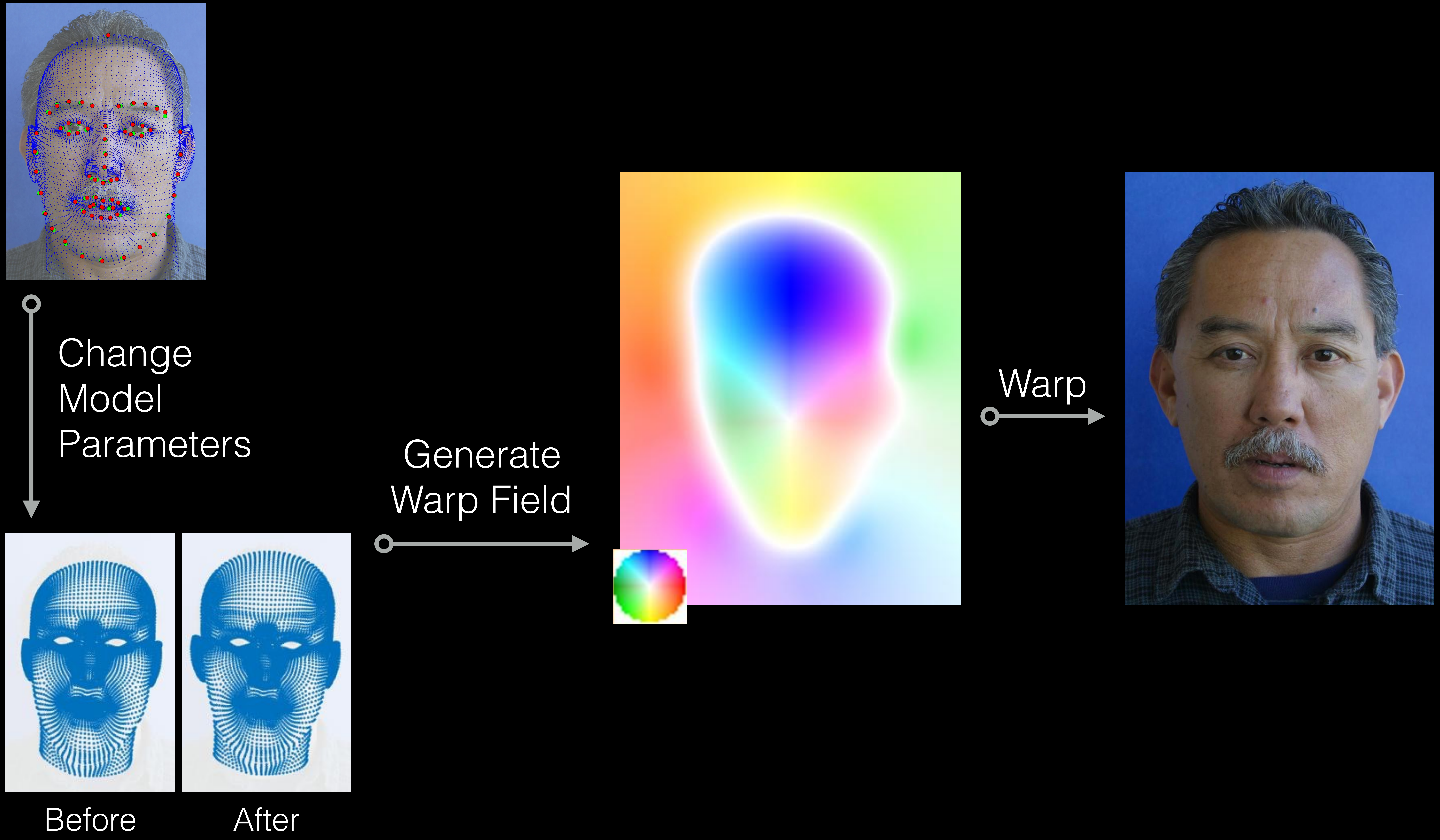
Single Image

Fiducial Detection

Head Model

Parameter Optimization

Part II: Warping



Change Model Parameters

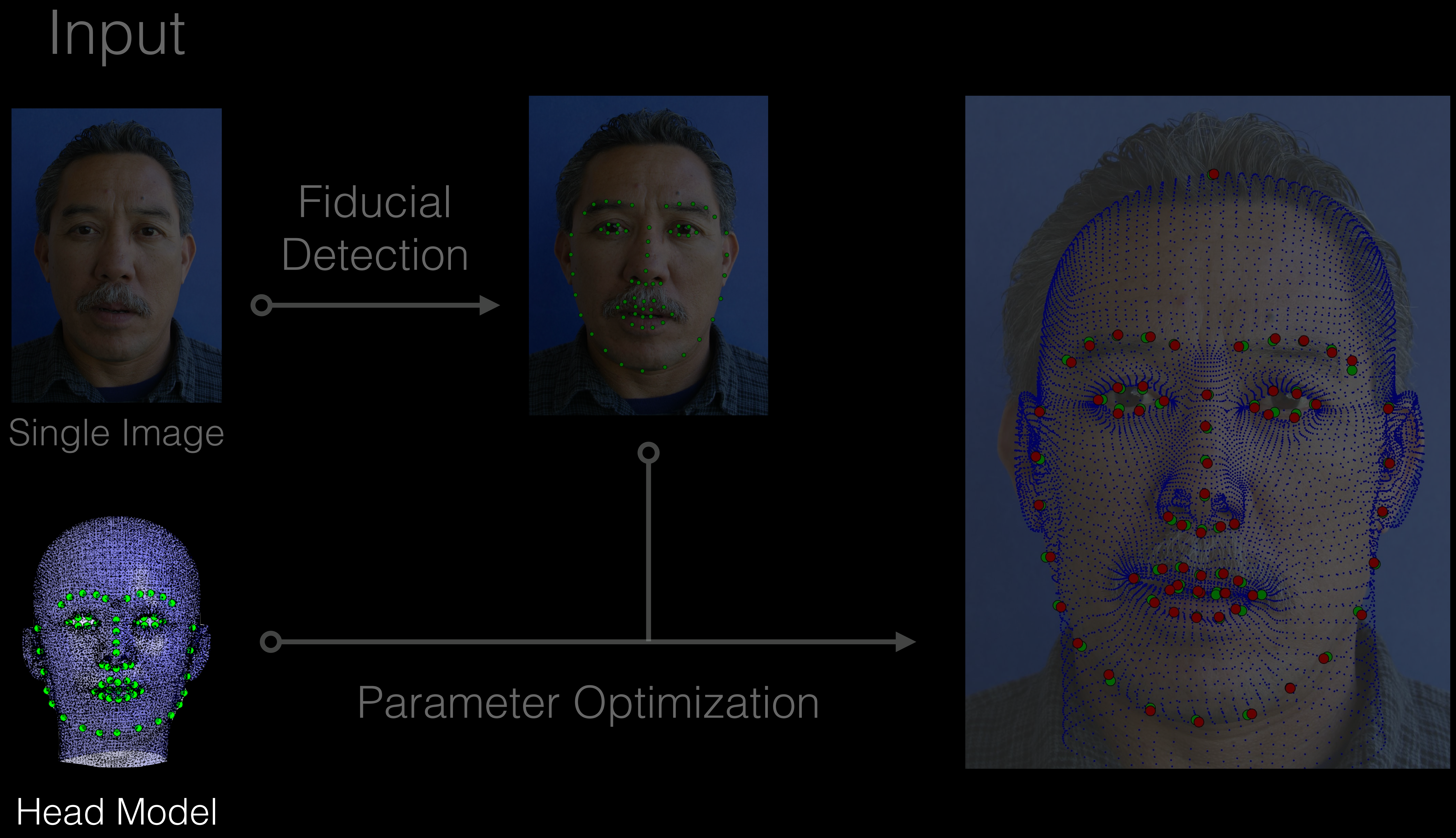
Before

After

Generate Warp Field

Warp

Part I: Fitting



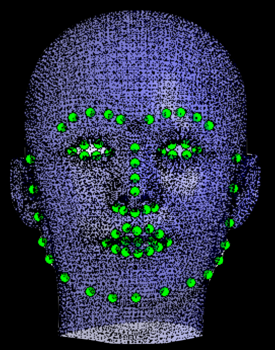
Input

Fiducial Detection

Single Image

Head Model

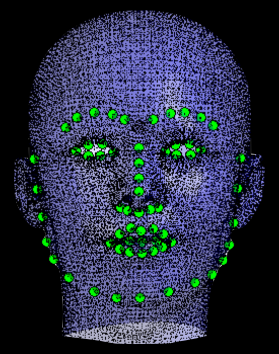
Parameter Optimization



Head Model

We want to support **diverse input photos**

We need an **expressive model**

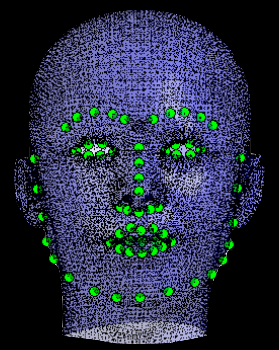


Head Model

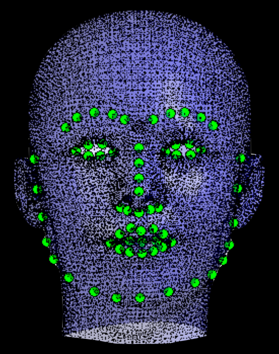
We want to support **diverse input photos**

We need an **expressive model**

What should it include?

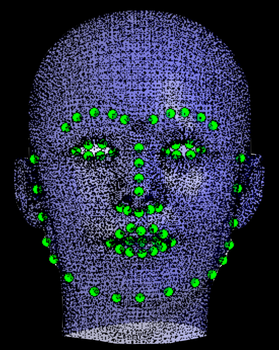


Head Model



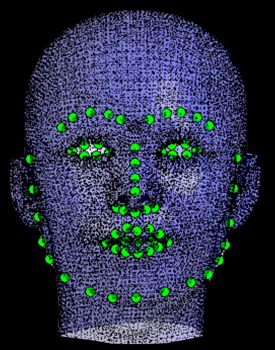
Head Model

- Head shape (“identity”)



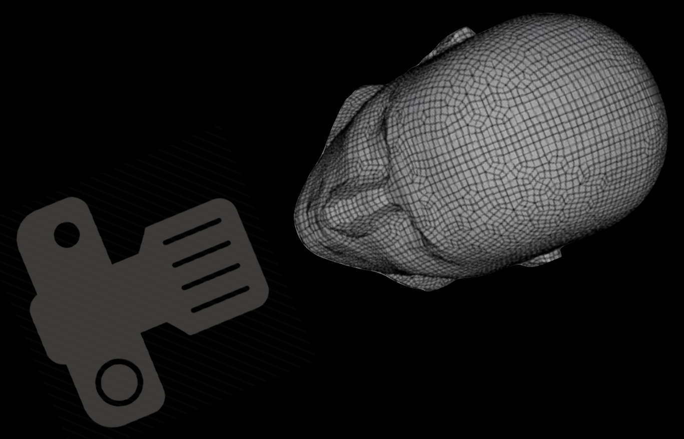
Head Model

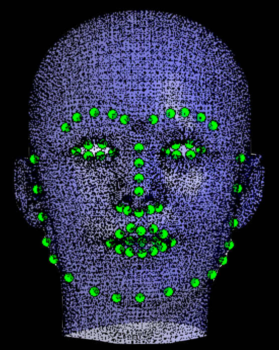
- Head shape (“identity”)
- Bone/muscle layout (“expression”)



Head Model

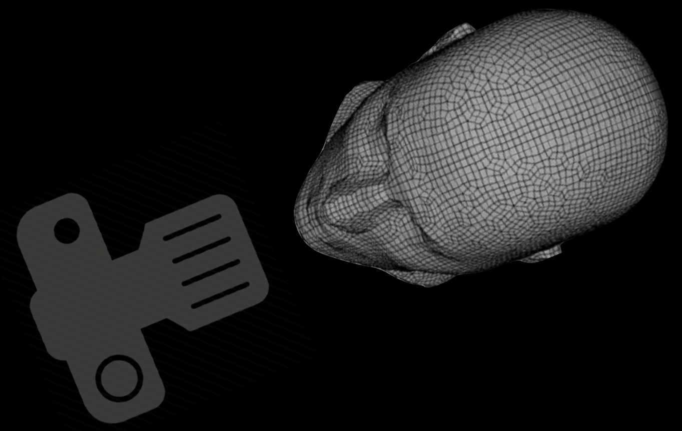
- Head shape (“identity”)
- Bone/muscle layout (“expression”)
- Location and pose, relative to camera





Head Model

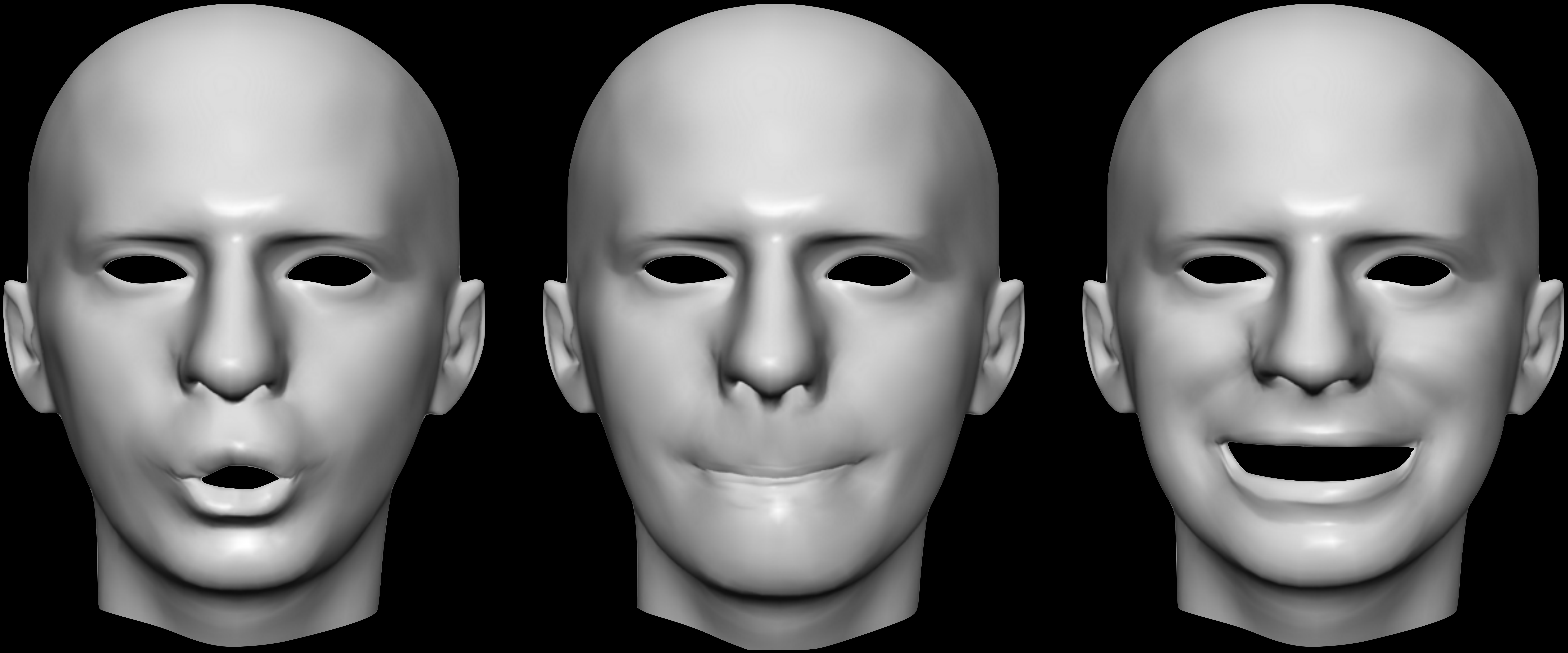
- Head shape (“identity”)
- Bone/muscle layout (“expression”)
- Location and pose, relative to camera
- Internal camera parameters



FaceWarehouse

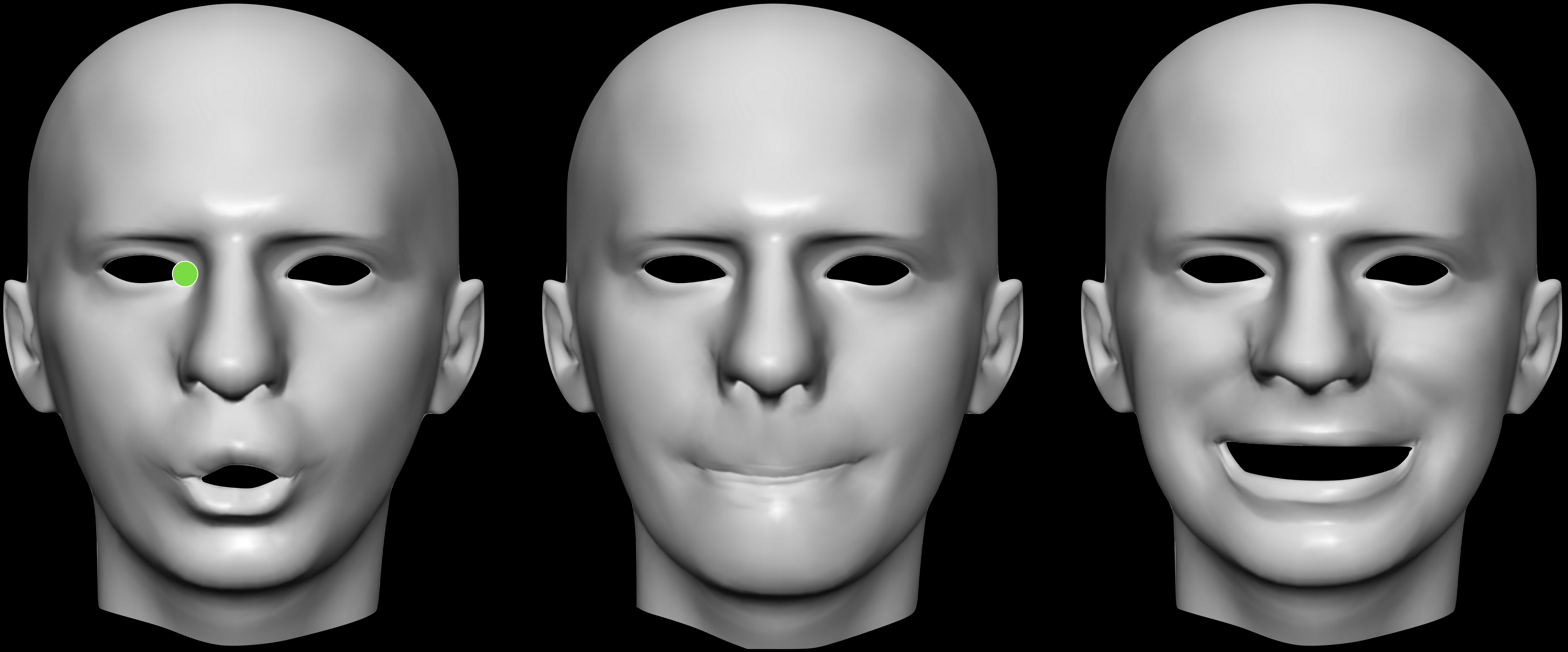
[Cao et al. 2014]

Aligned models



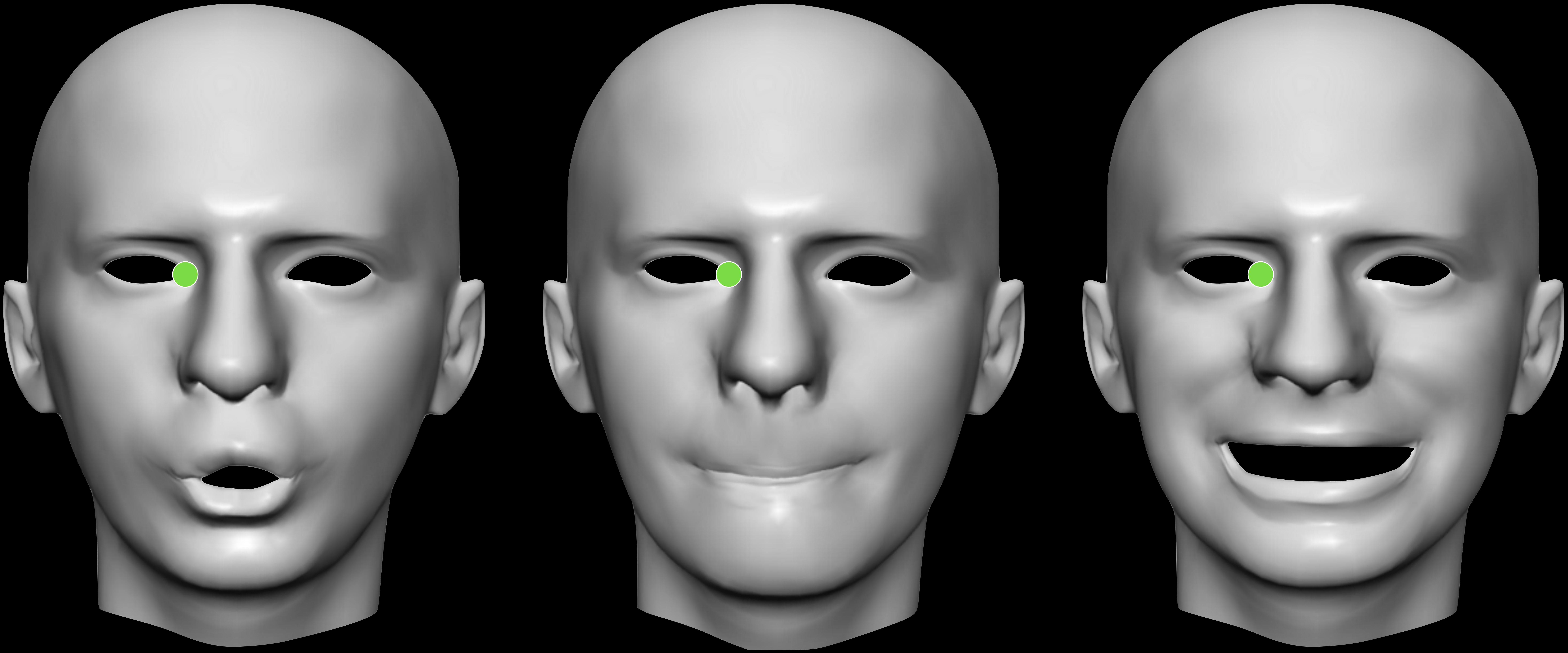
[Cao et al. 2014]

Aligned models



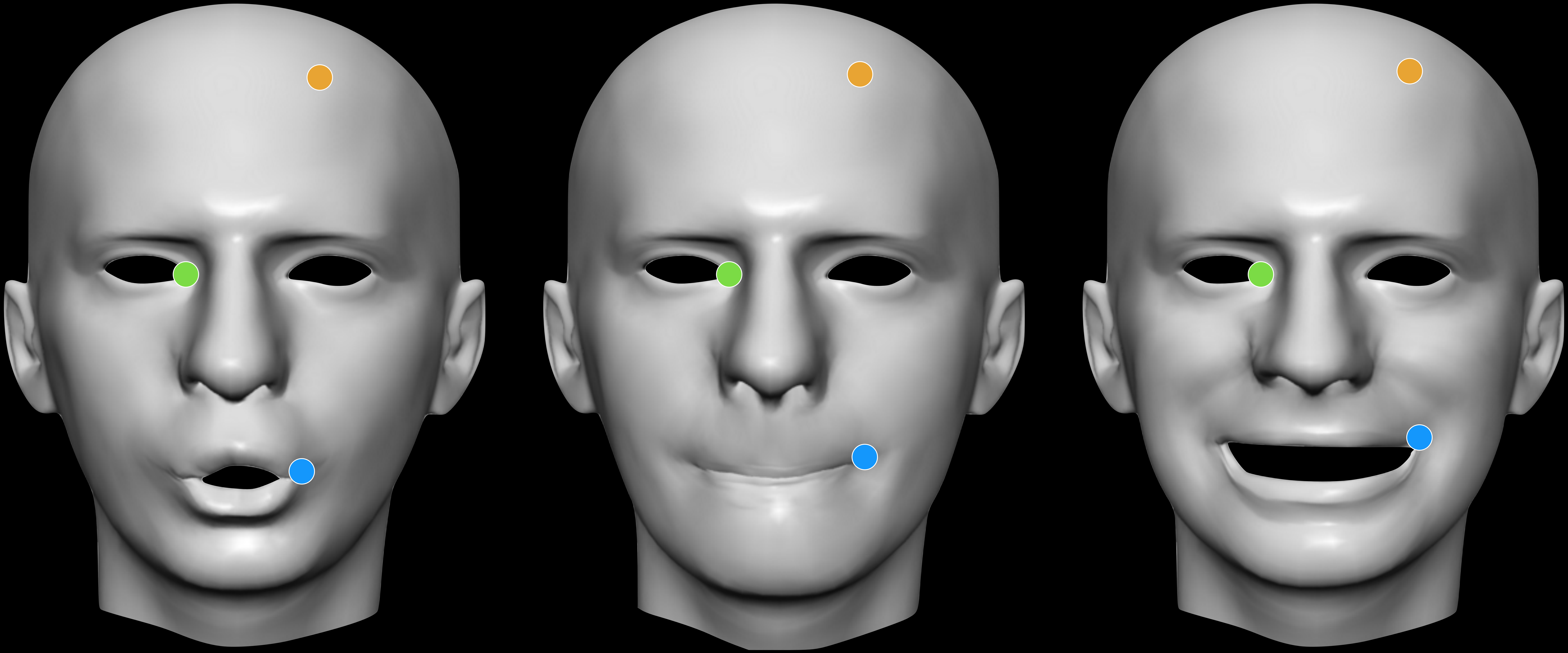
[Cao et al. 2014]

Aligned models

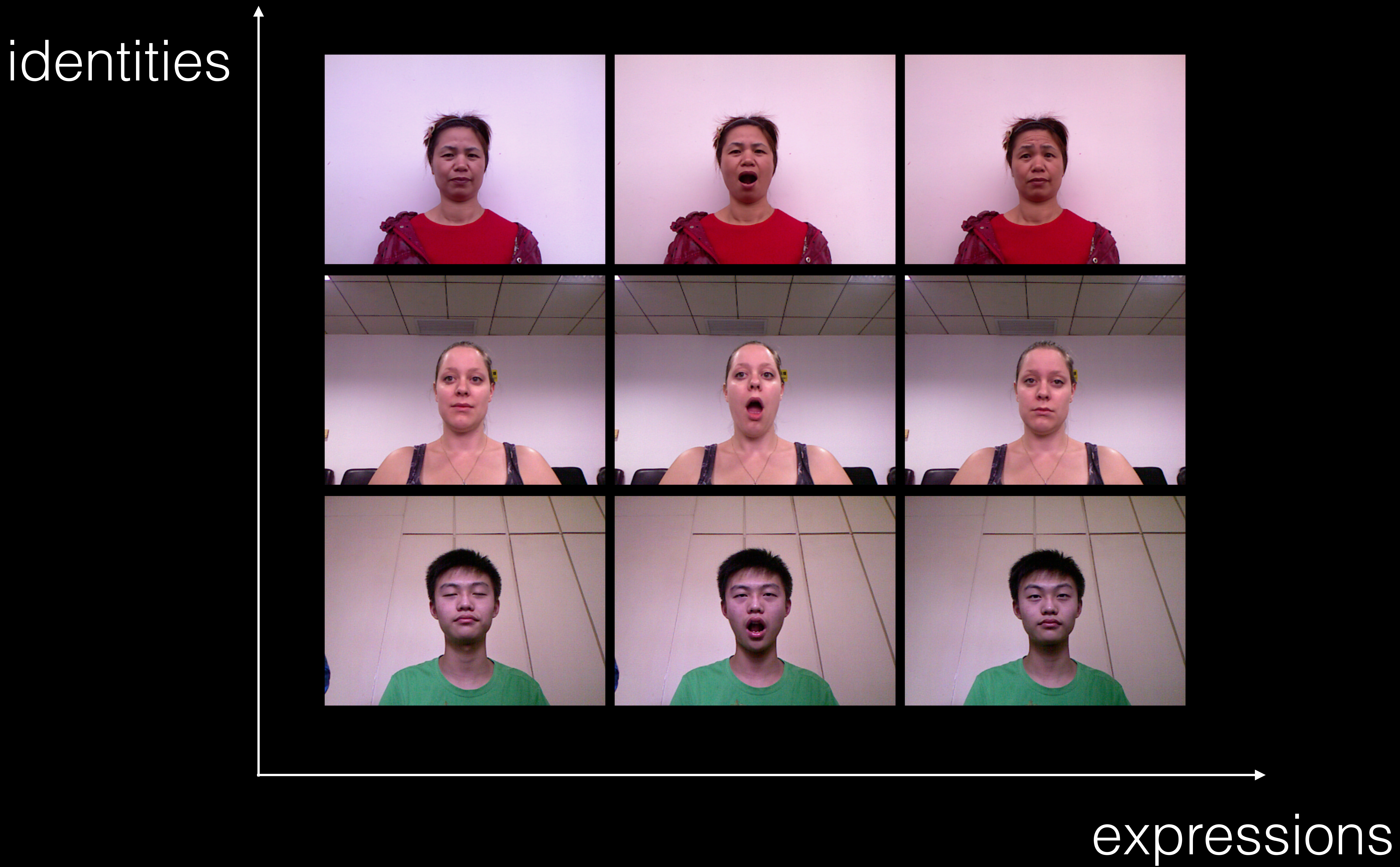


[Cao et al. 2014]

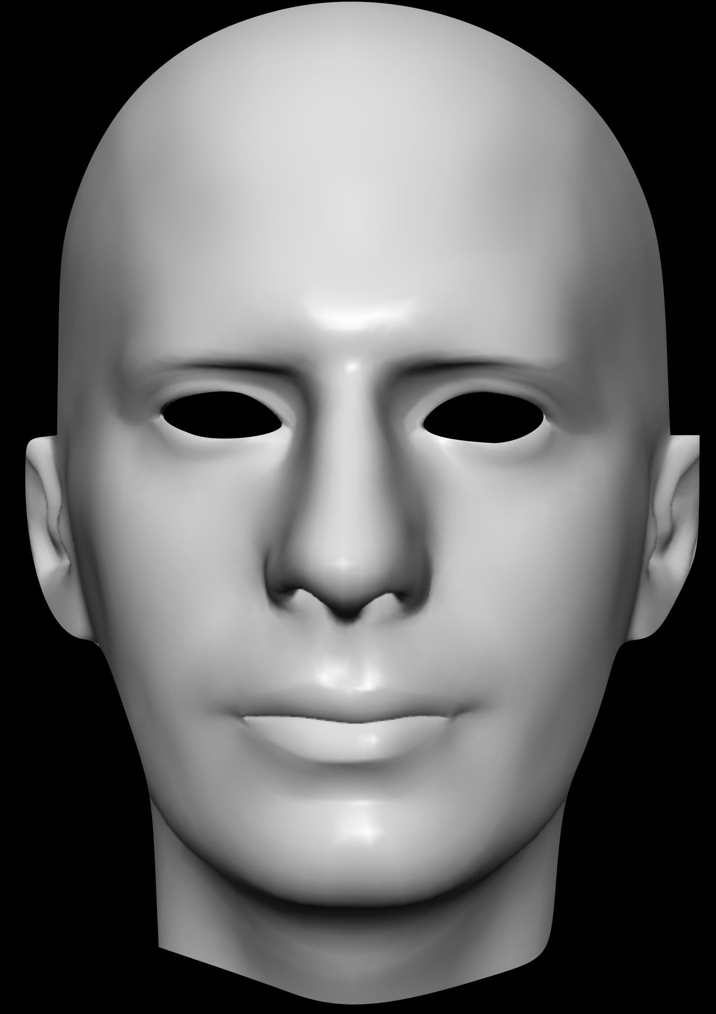
Aligned models



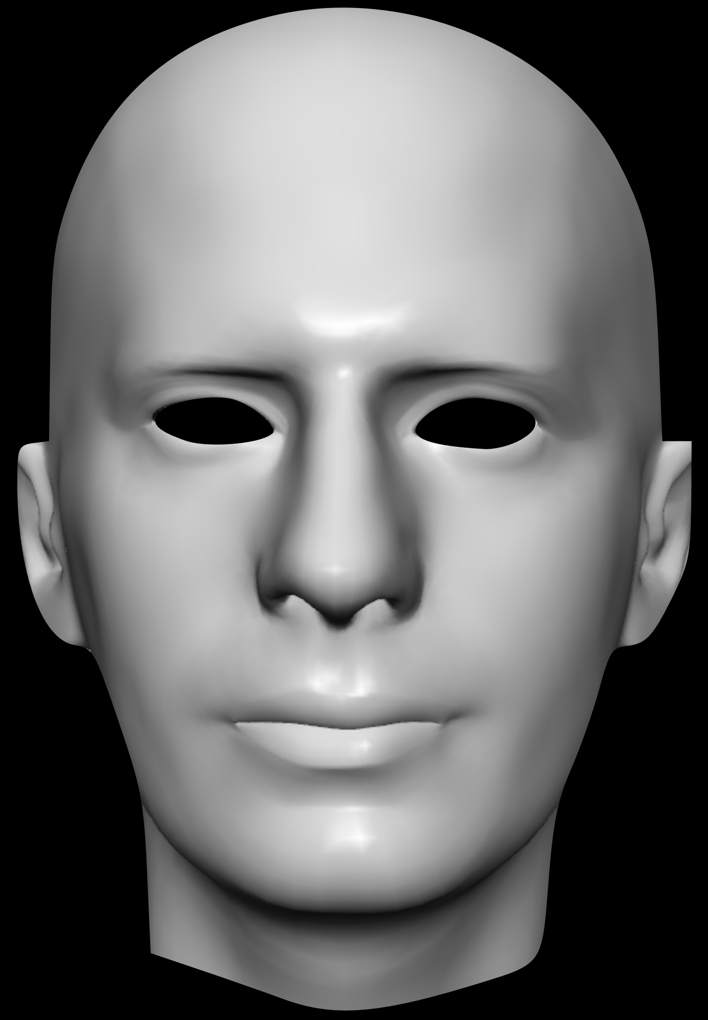
[Cao et al. 2014]



[Cao et al. 2014]

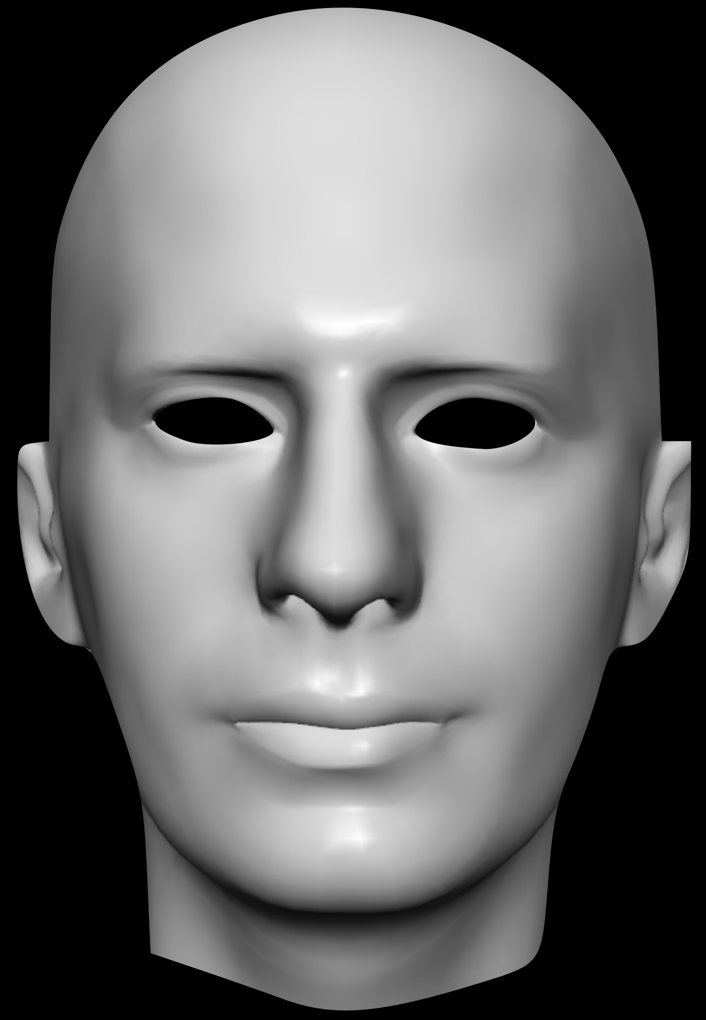


11K vertices
x
150 identities
x
47 expressions



11K vertices
x
150 identities
x
47 expressions

We can combine these
to create new heads!



11K vertices
x
150 identities
x
47 expressions

We can combine these
to create new heads!

HOSVD ⬟

Short detour — SVD

$$M_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}$$

$$M_{3\times 4} = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \end{bmatrix} \begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ - & v_3 & - \\ - & v_4 & - \end{bmatrix}$$

Short detour — SVD

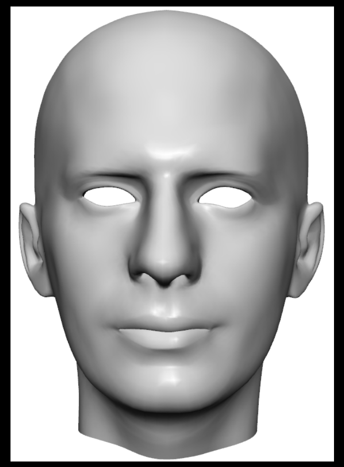
$$M_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}$$

$$M_{3\times 4} = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \cancel{\sigma_3} & 0 \end{bmatrix} \begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ - & v_3 & - \\ - & v_4 & - \end{bmatrix}$$

Short detour — SVD

$$M_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}$$

$$M_{3\times 4} = \begin{bmatrix} | & | & \cancel{|} \\ u_1 & u_2 & \cancel{u_3} \\ | & | & \cancel{|} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \cancel{\sigma_3} & 0 \end{bmatrix} \begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ \cancel{-} & \cancel{v_3} & \cancel{-} \\ \cancel{-} & \cancel{v_4} & \cancel{-} \end{bmatrix}$$



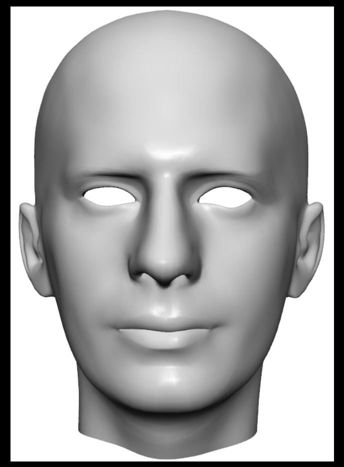
11K vertices
150 identities
47 expressions

34530x150x47

HOSVD

Core Tensor $\otimes_1$ $A_{34530x40}$ $\otimes_2$ B_{150x50} $\otimes_3$ Γ_{47x25}

40 x 50 x 25



11K vertices
150 identities
47 expressions

34530x150x47

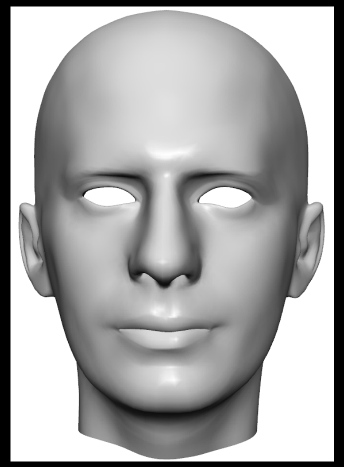
HOSVD

Low rank approximation:

- ✓ less space
- ✓ less noise

Core Tensor $\otimes_1$ $A_{34530x40}$ $\otimes_2$ B_{150x50} $\otimes_3$ Γ_{47x25}

40 x 50 x 25



11K vertices
150 identities
47 expressions

34530x150x47

HOSVD

Low rank approximation:

✓ less space

✓ less noise

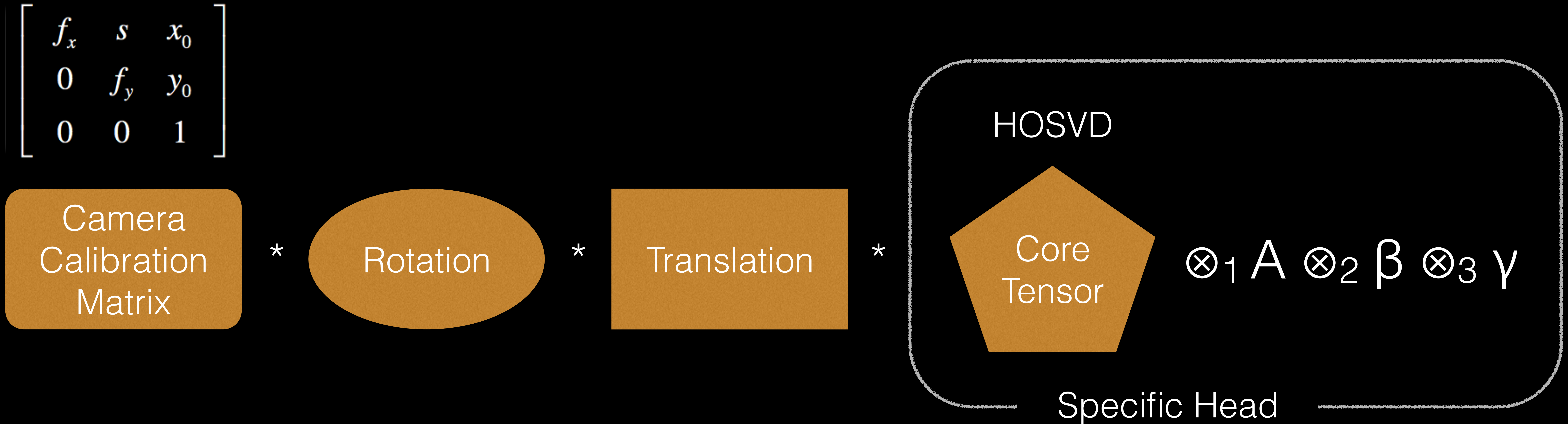
Core Tensor $\otimes_1$ $A_{34530x40}$ $\otimes_2$ B_{150x50} $\otimes_3$ Γ_{47x25}

40 x 50 x 25

And for a single head:

Core Tensor $\otimes_1$ $A_{34530x40}$ $\otimes_2$ β_{1x50} $\otimes_3$ γ_{1x25}

40 x 50 x 25



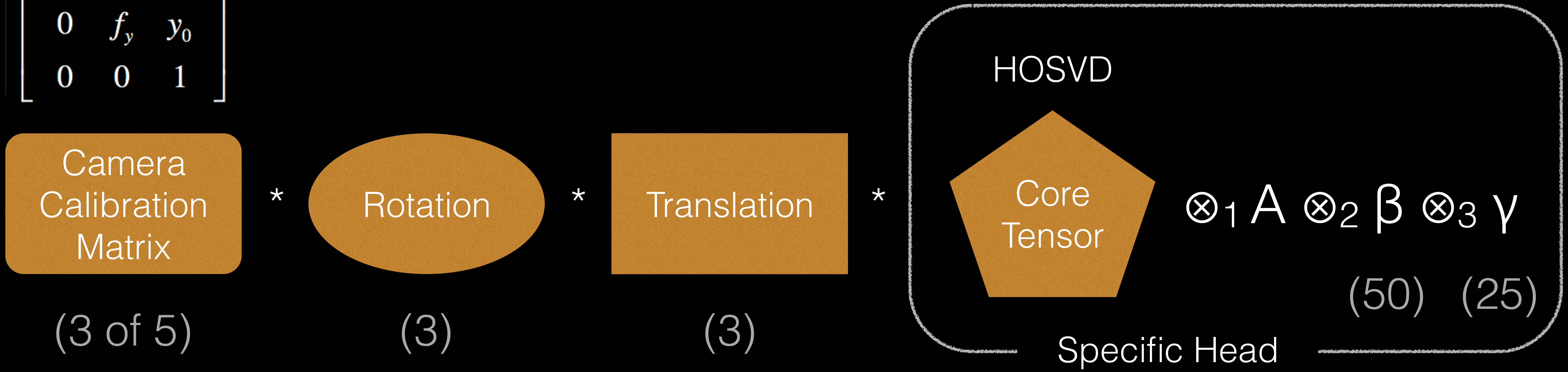
$$\begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Camera Calibration Matrix * Rotation * Translation * Core Tensor $\otimes_1 A \otimes_2 \beta \otimes_3 \gamma$

HOSVD

Specific Head

$$\begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$



Camera Calibration Matrix * Rotation * Translation * Core Tensor $\otimes_1 A \otimes_2 \beta \otimes_3 \gamma$

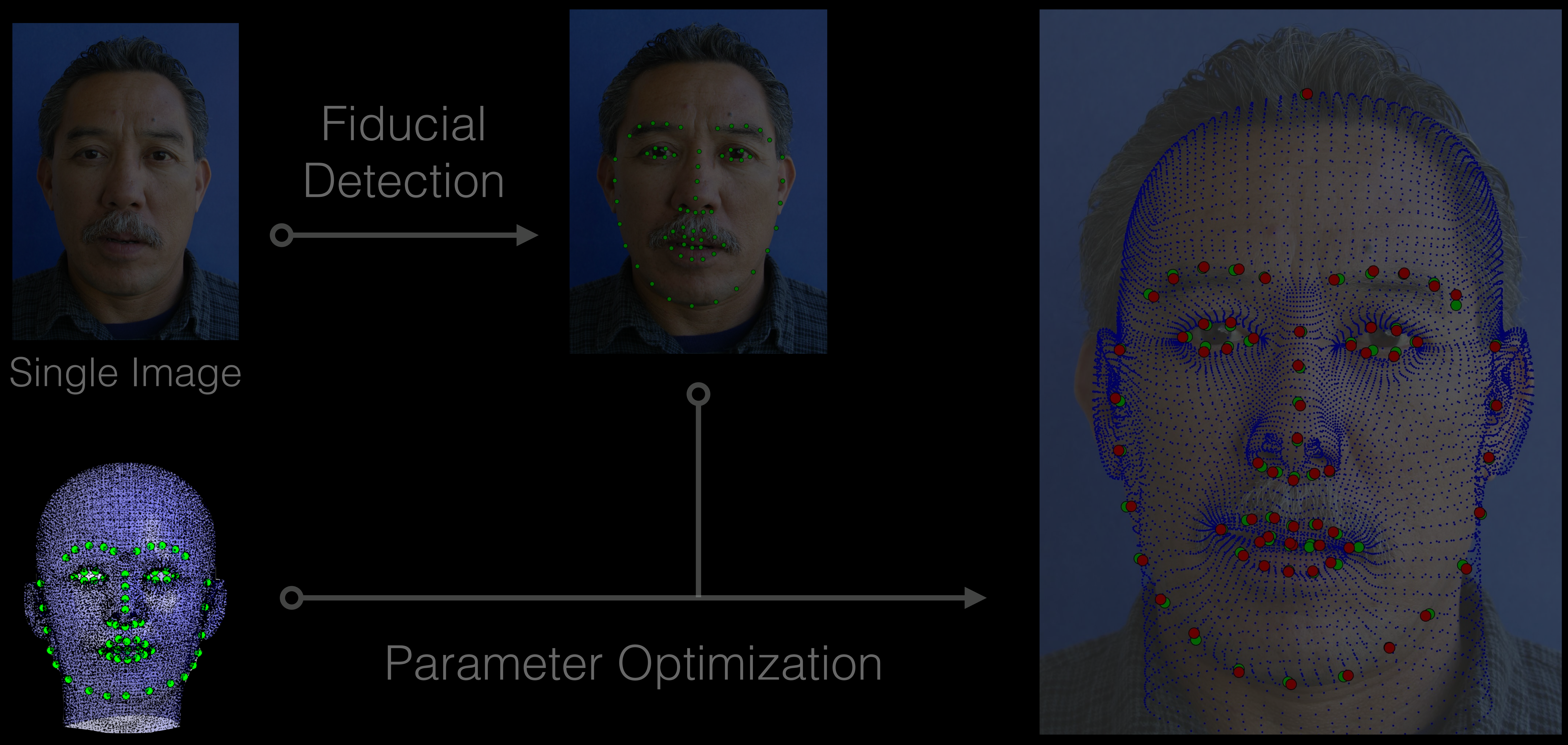
(3 of 5) (3) (3) (50) (25)

HOSVD

Specific Head

3 + 3 + 3 + 50 + 25 = 84 parameters

Fitting



Input

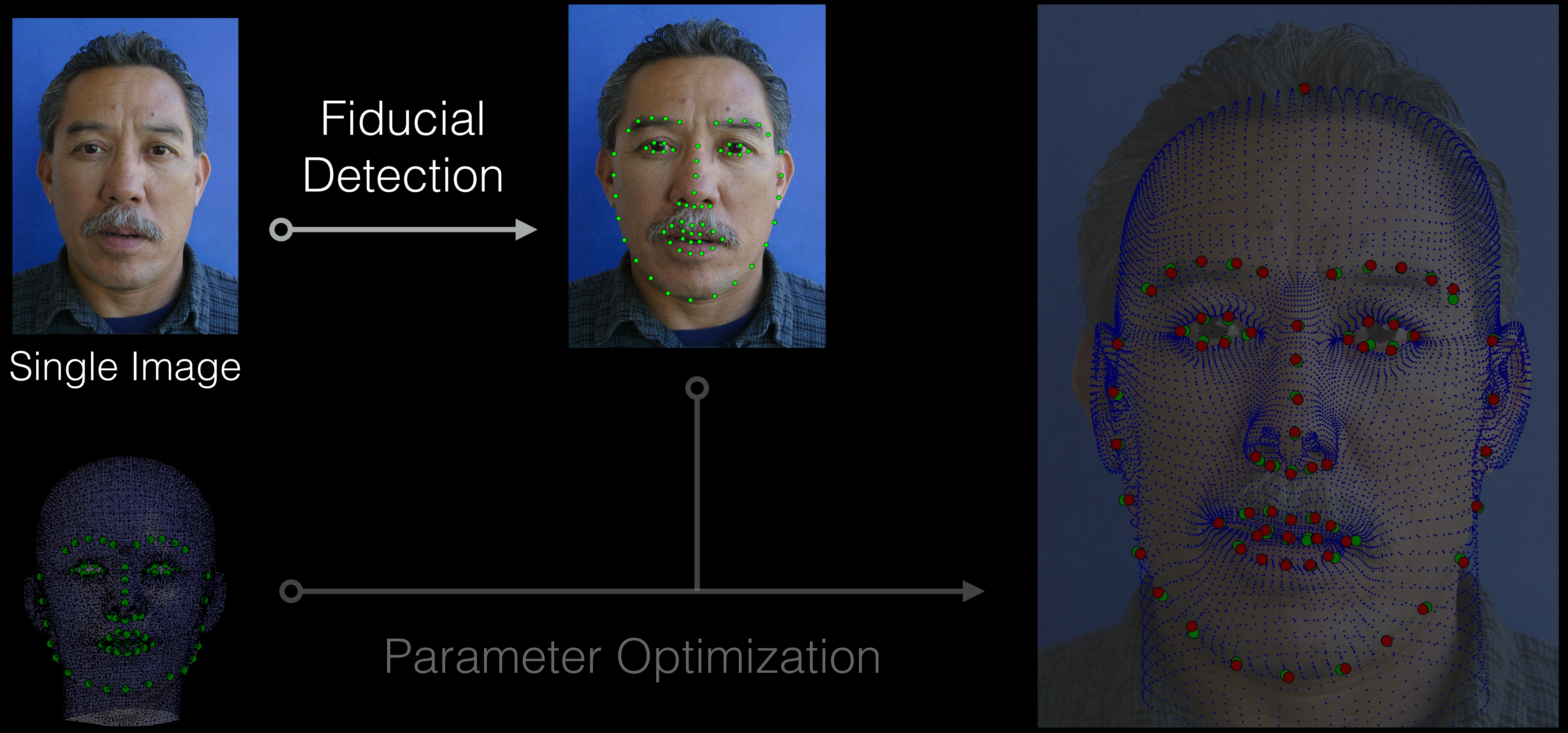
Single Image

Fiducial Detection

Head Model

Parameter Optimization

Fitting



Input

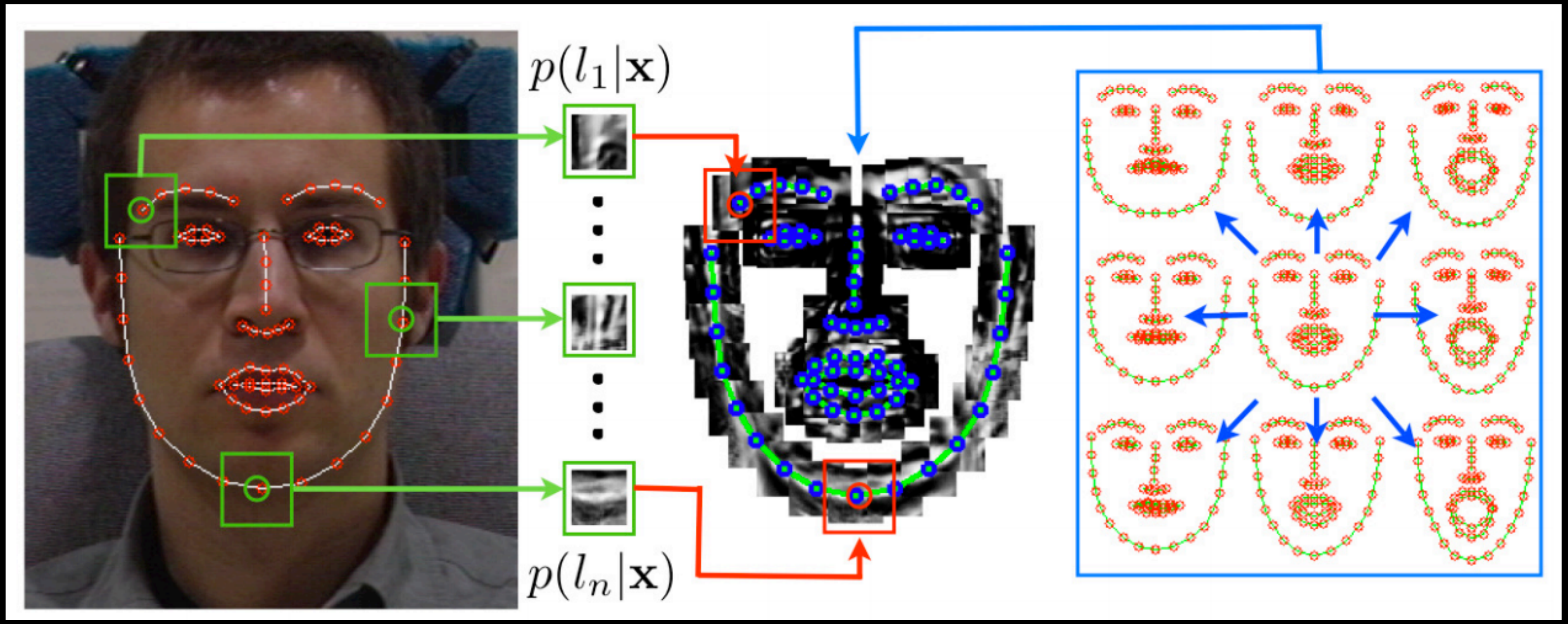
Single Image

Fiducial Detection

Head Model

Parameter Optimization

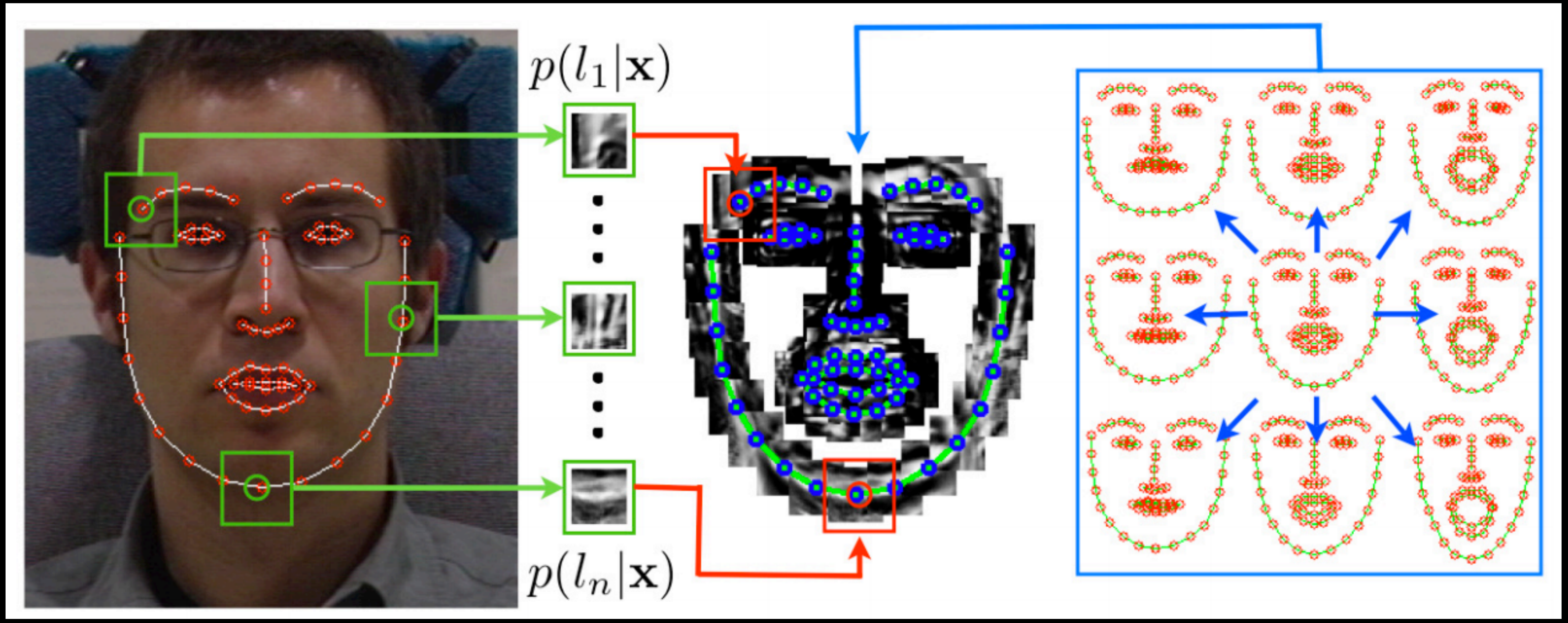
Fiducial points



66 points

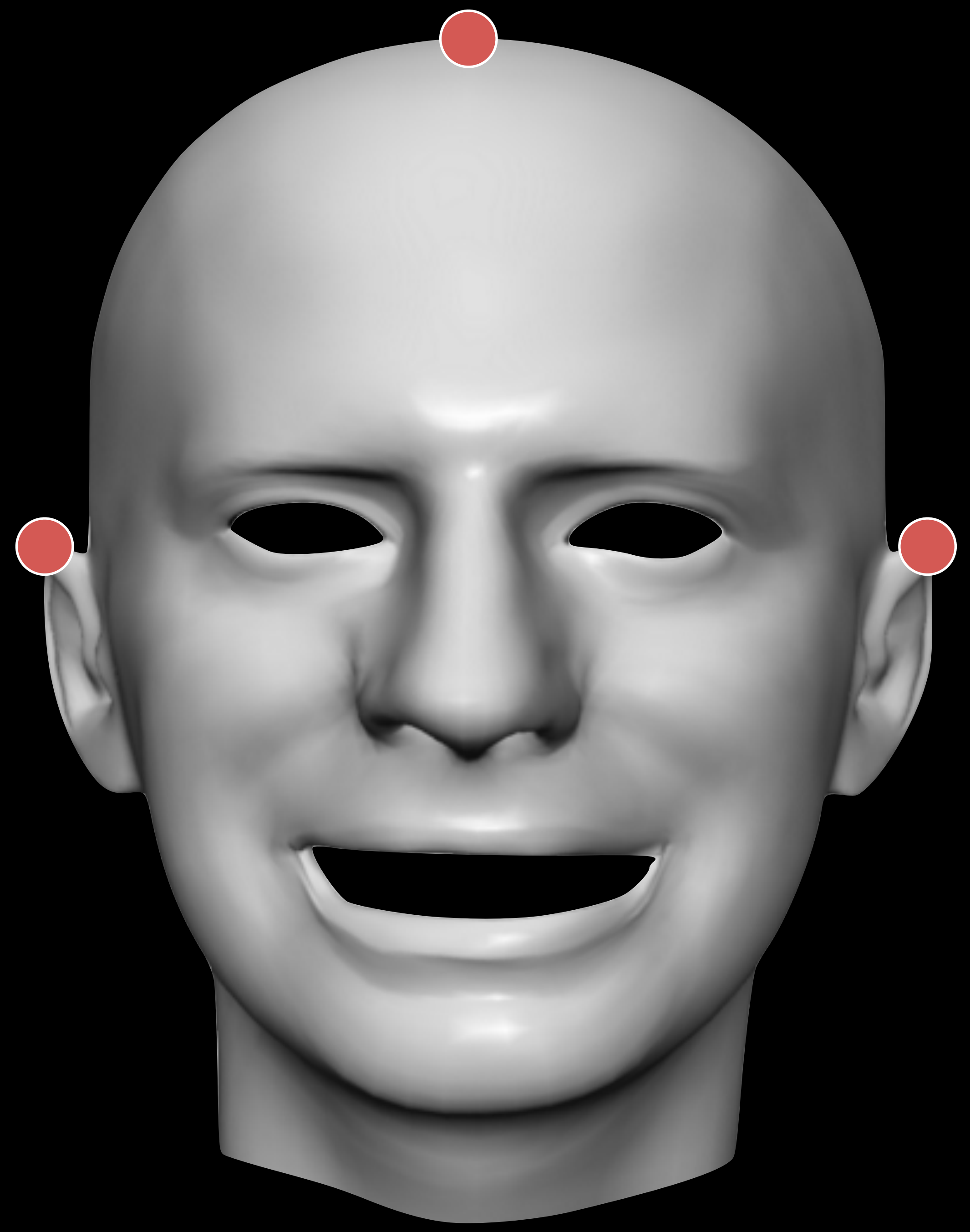
[Saragih et al. 2009]

Fiducial points



66 points

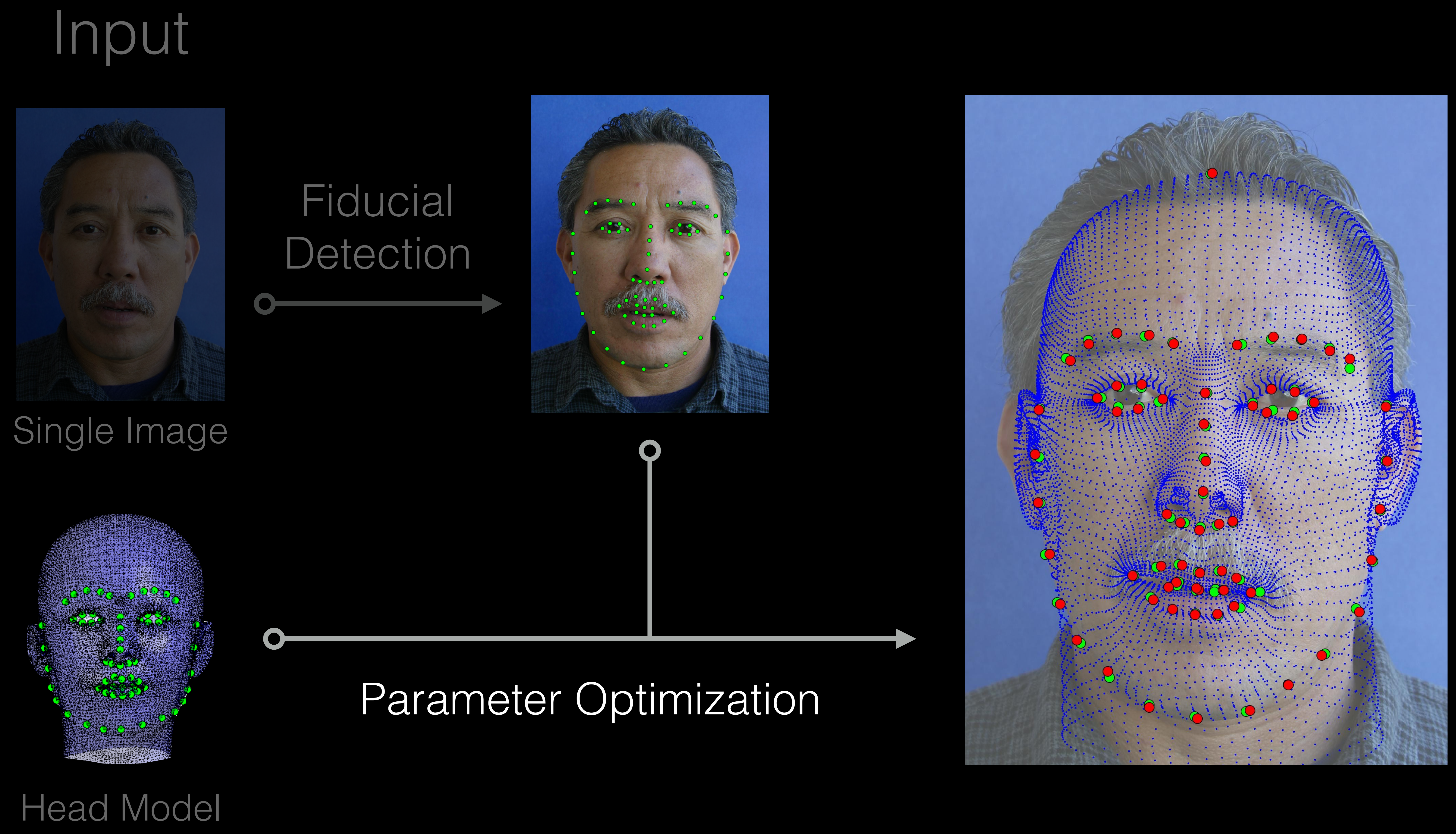
[Saragih et al. 2009]



3 points

(manually marked)

Fitting



Input

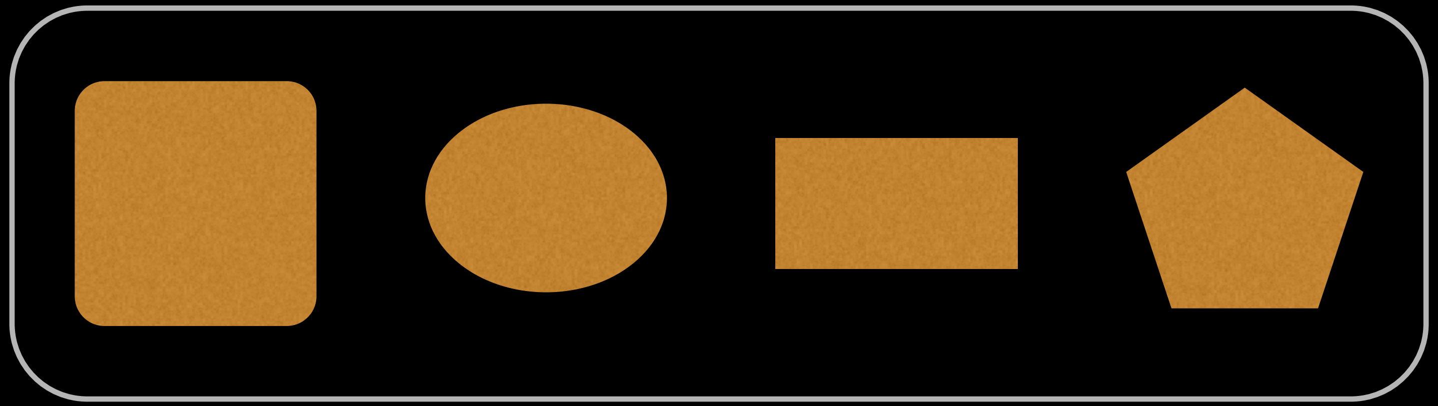
Single Image

Fiducial Detection

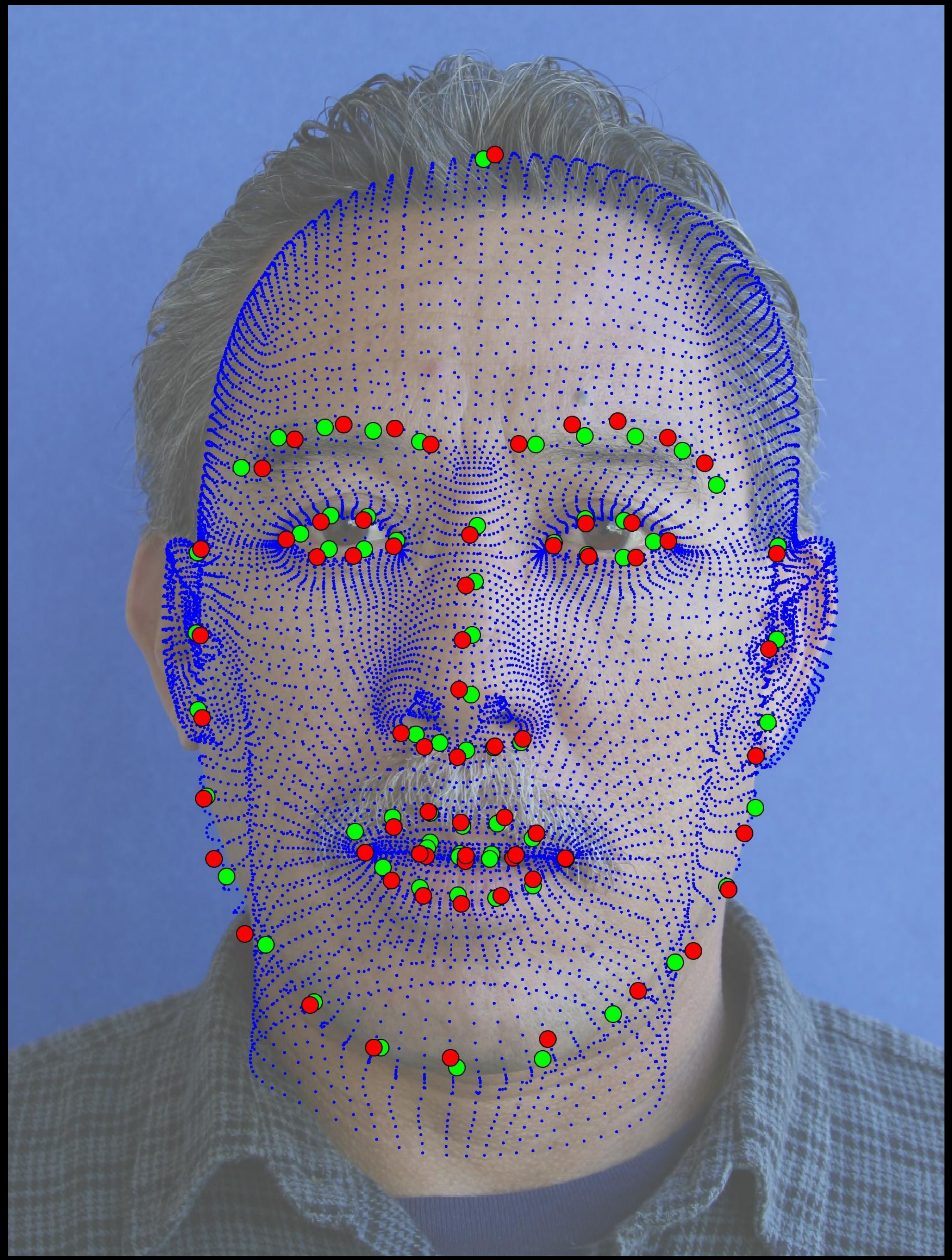
Head Model

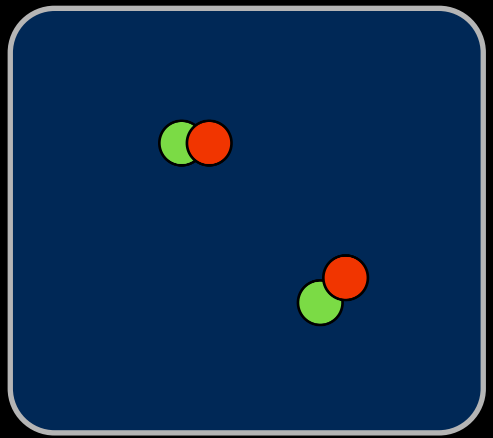
Parameter Optimization

Fitting

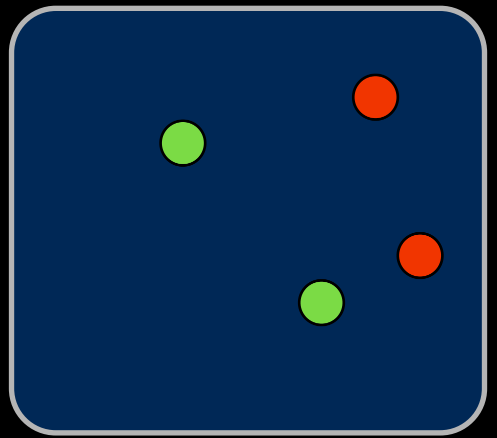


Plug in parameters



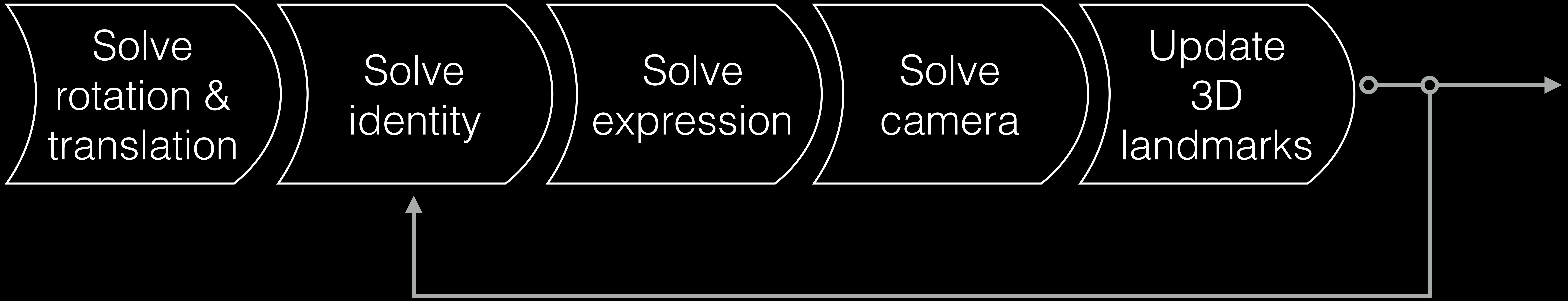


good



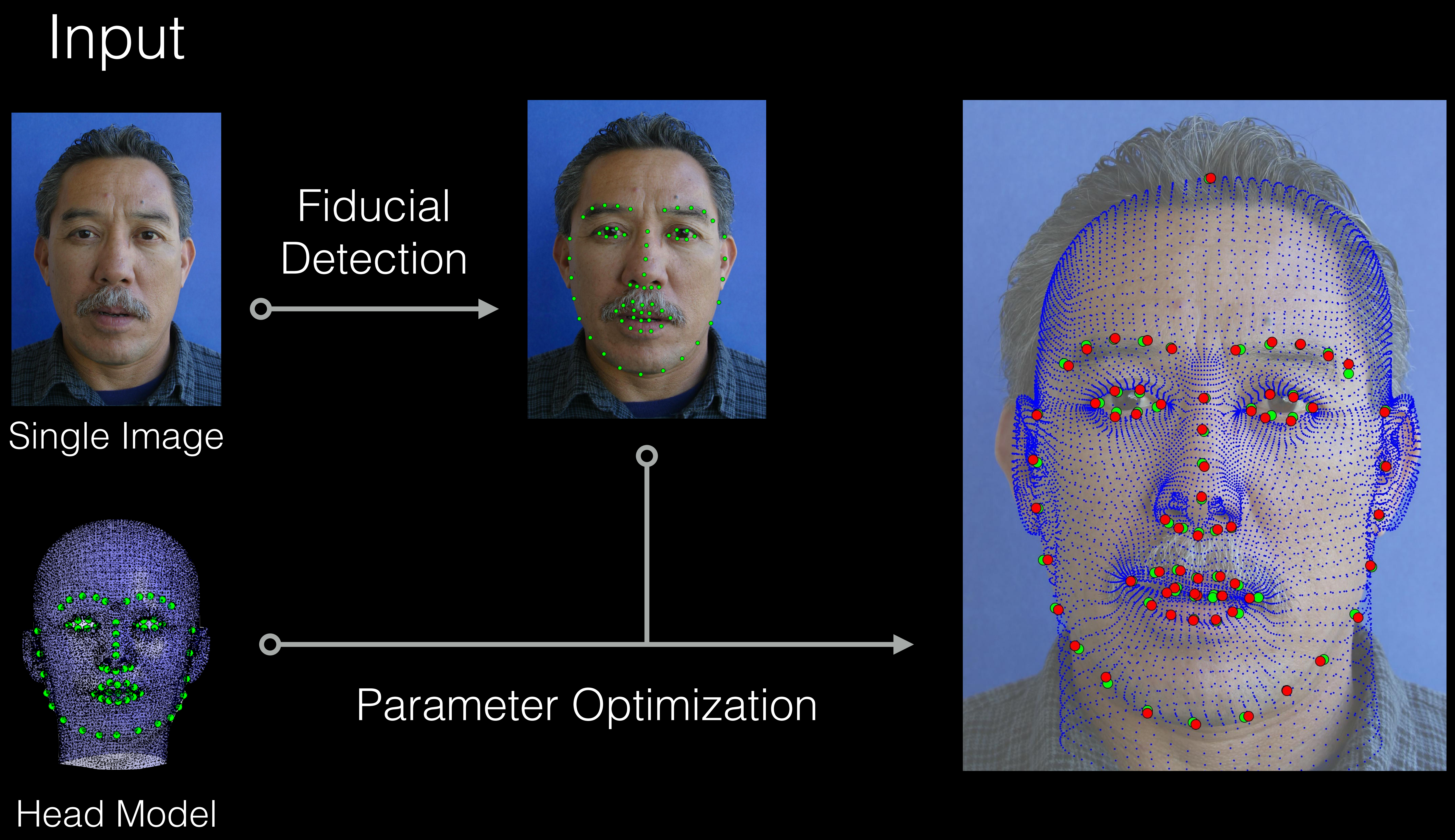
bad

Optimization



Coordinate descent

Fitting



Input

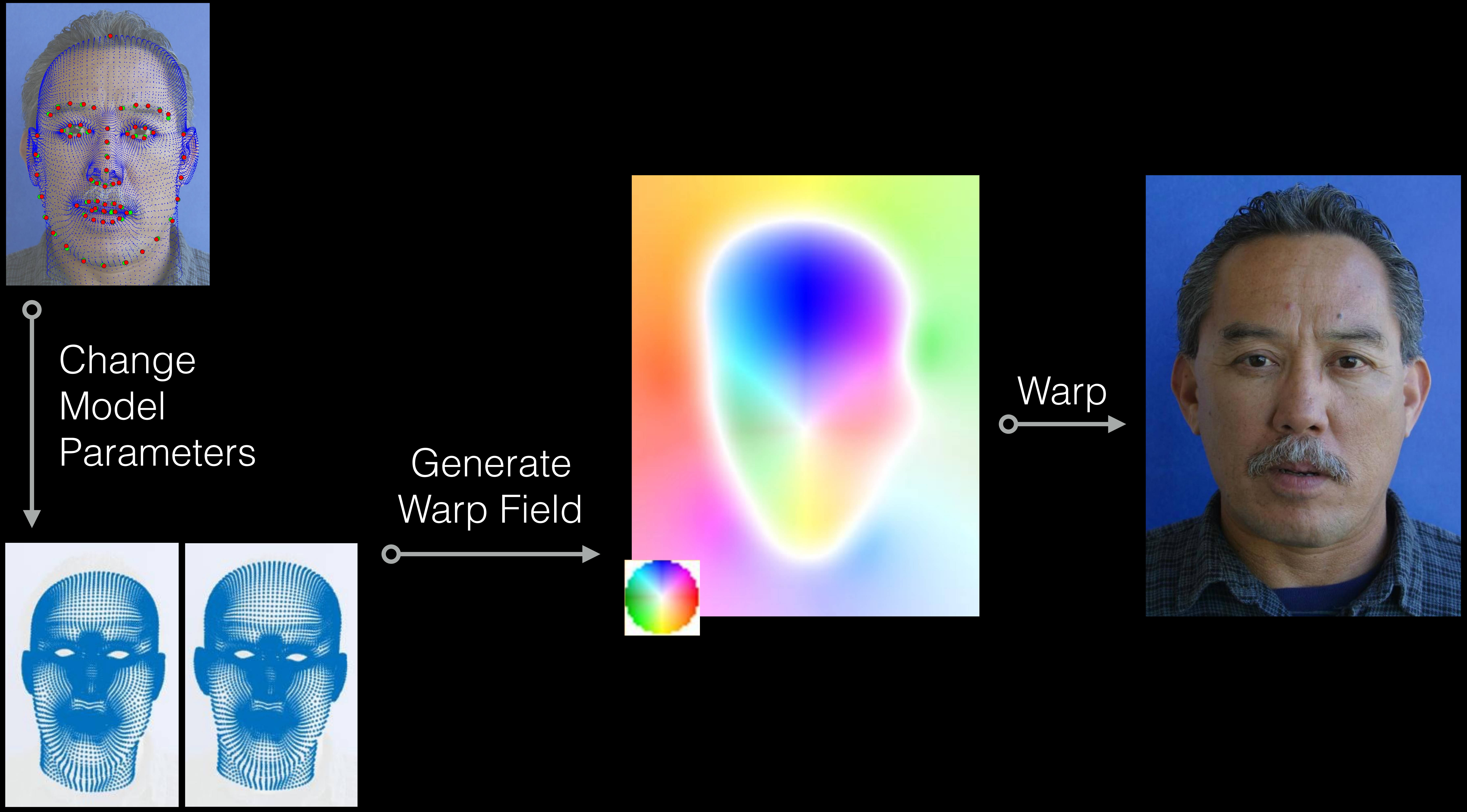
Single Image

Fiducial Detection

Head Model

Parameter Optimization

Part II: Warping



Change Model Parameters

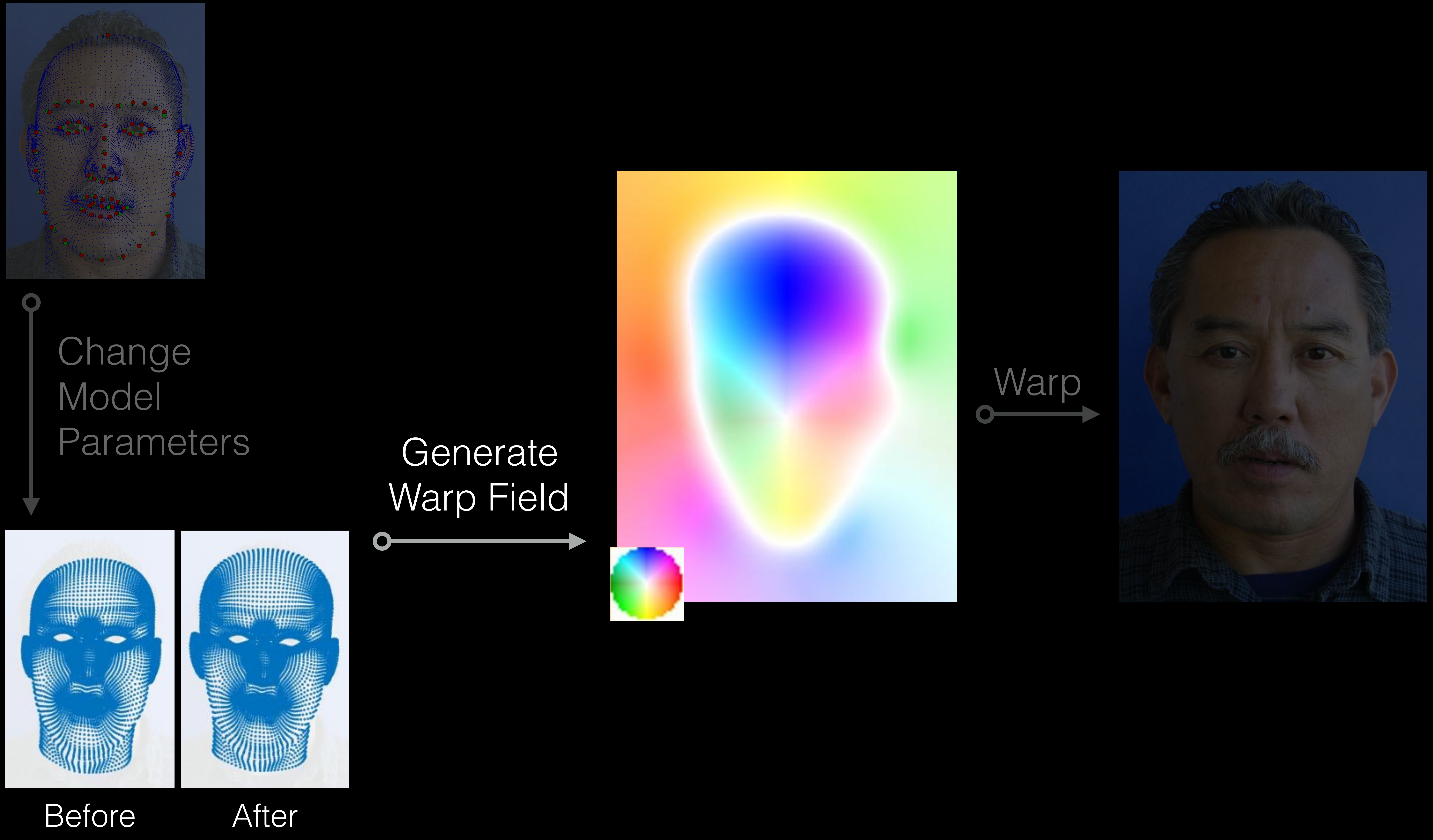
Generate Warp Field

Warp

Before

After

Warping



Change Model Parameters

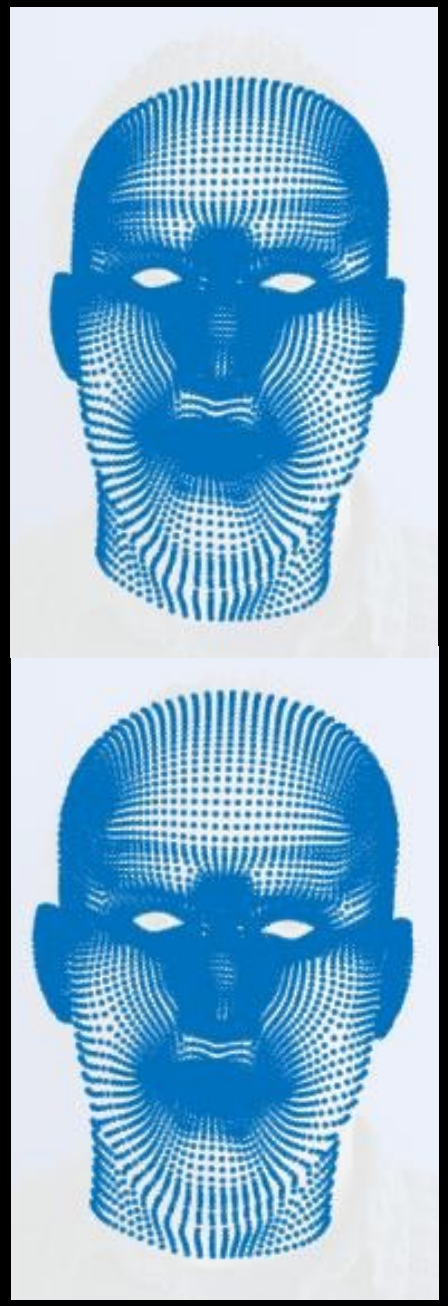
Before

After

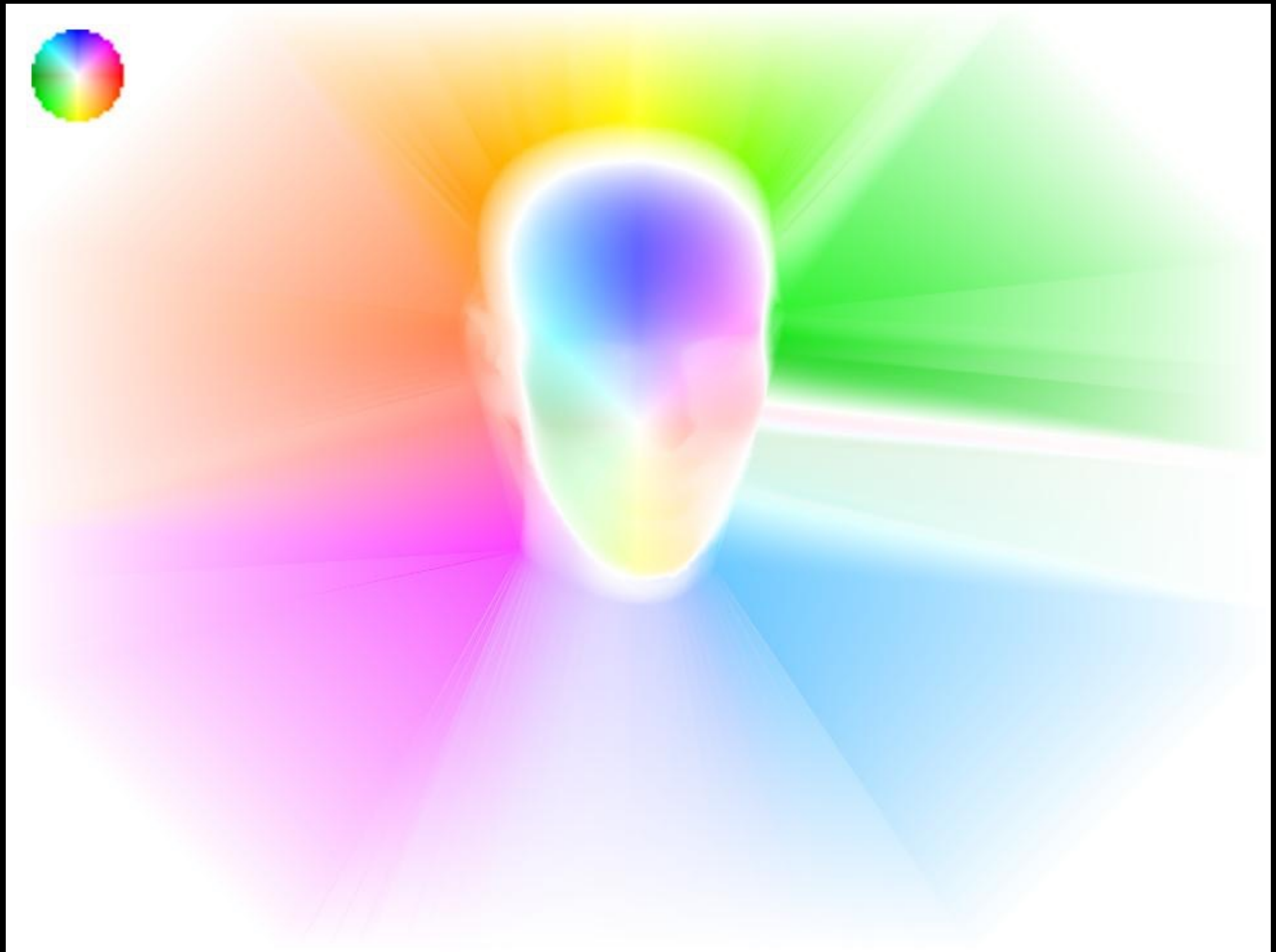
Generate Warp Field

Warp

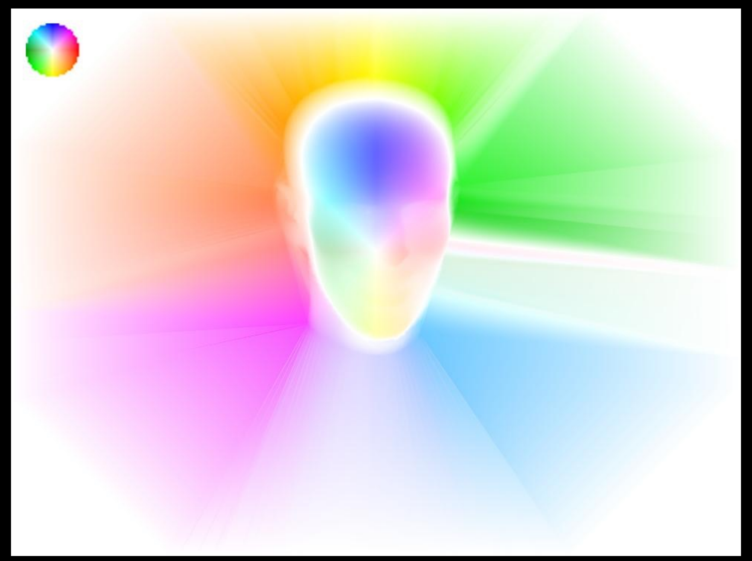
Warp field

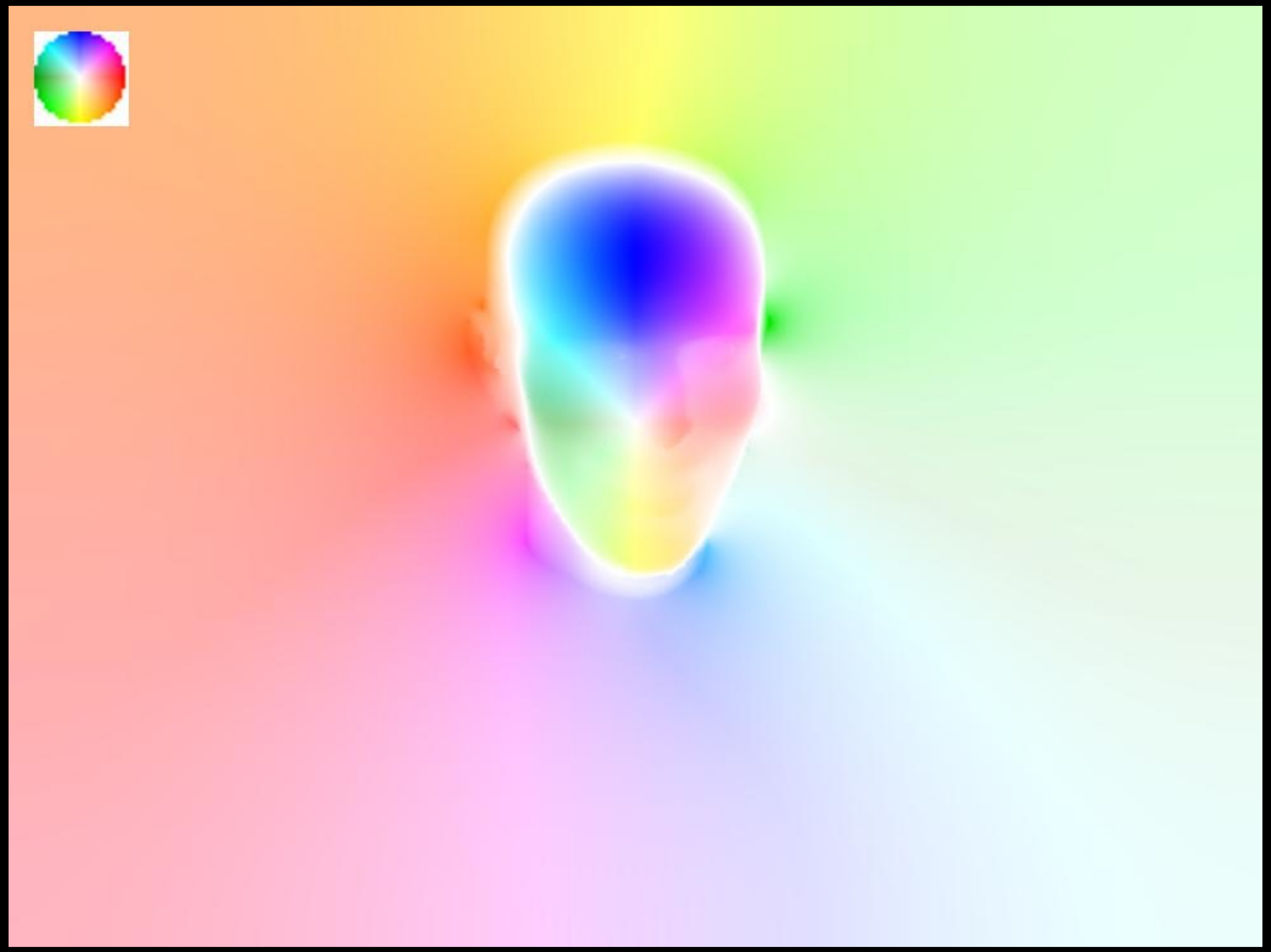


Triangulation based interpolation

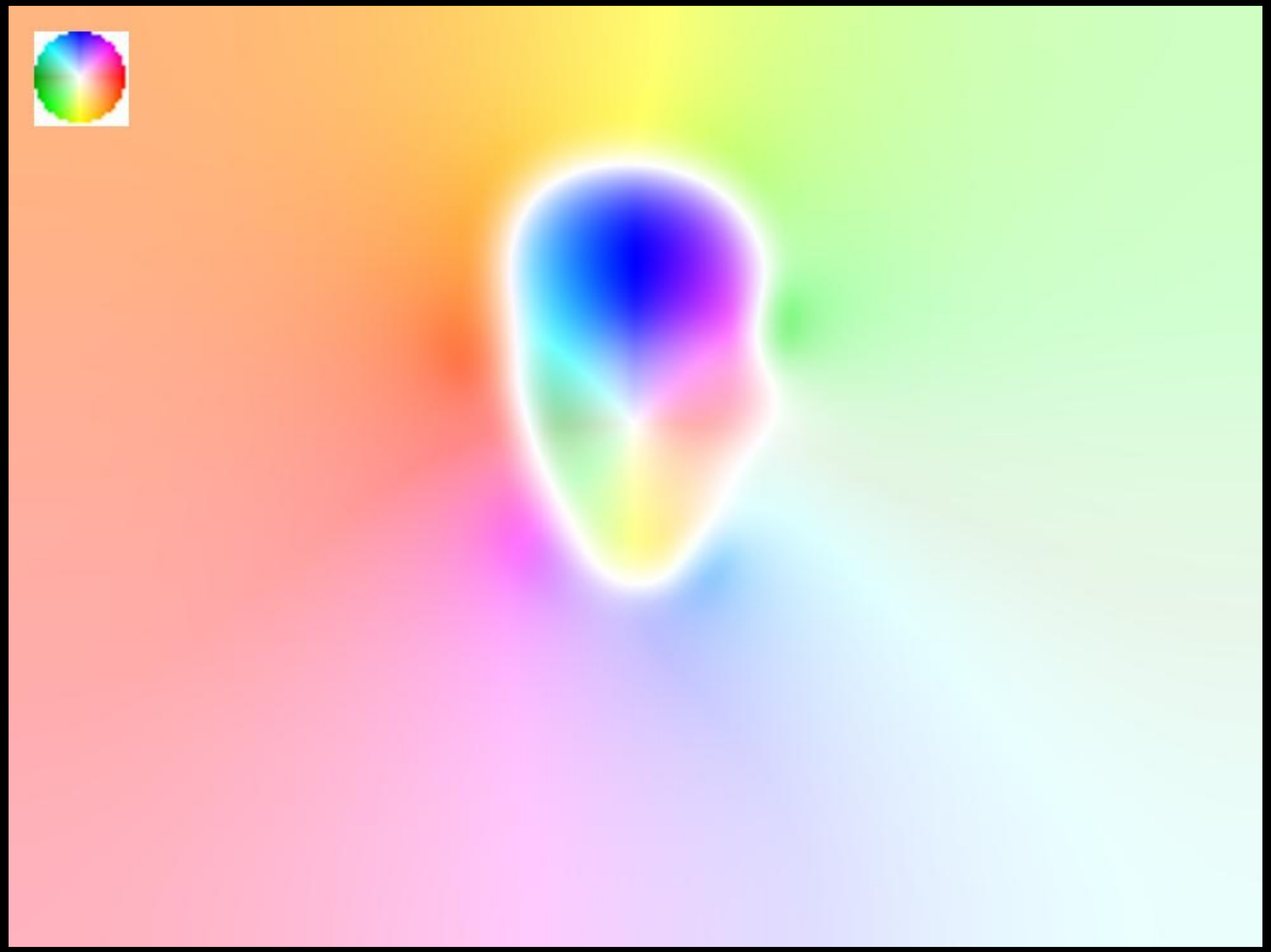


Warp field



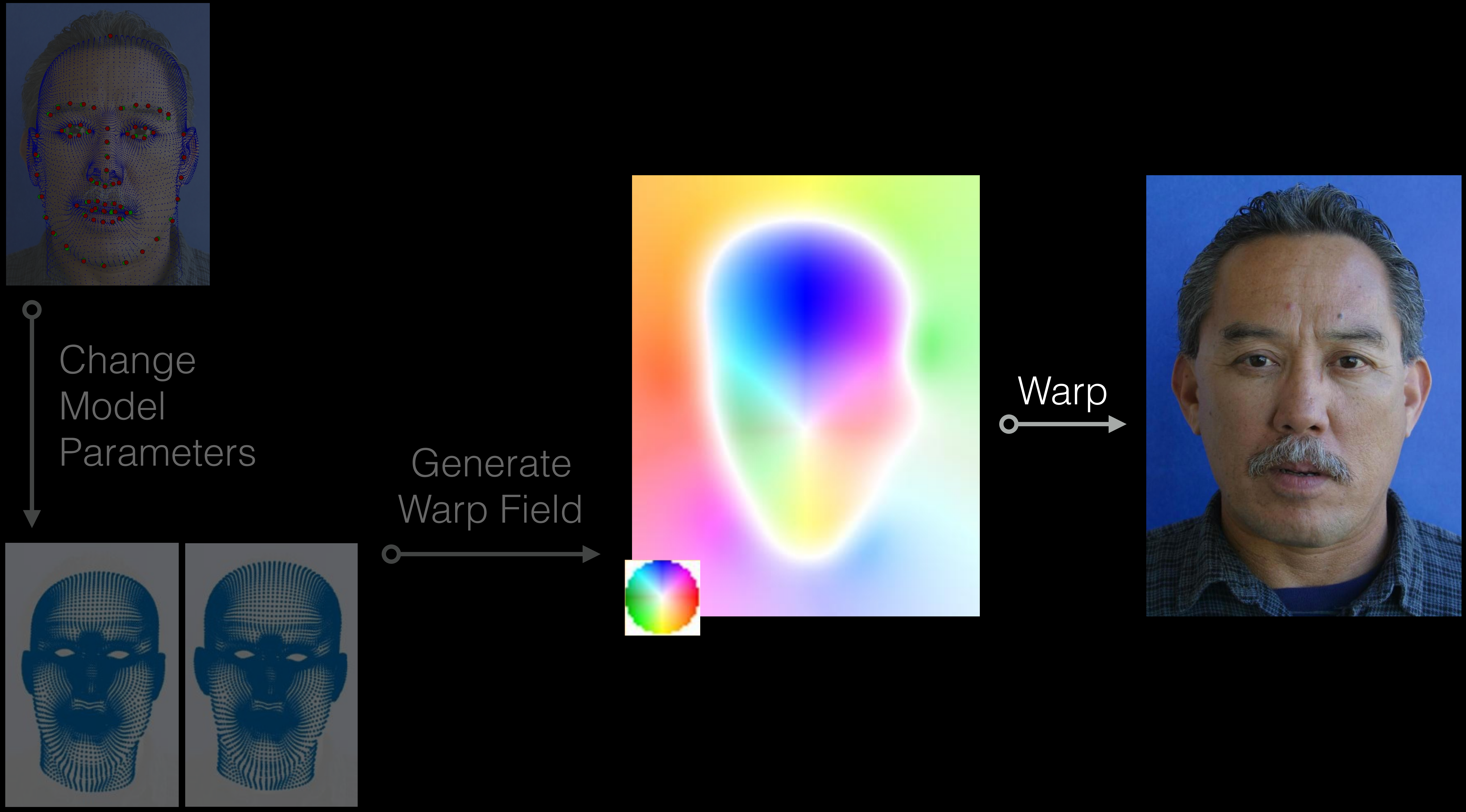


Discrete Laplace operator



Averaging filter

Warping



Change Model Parameters

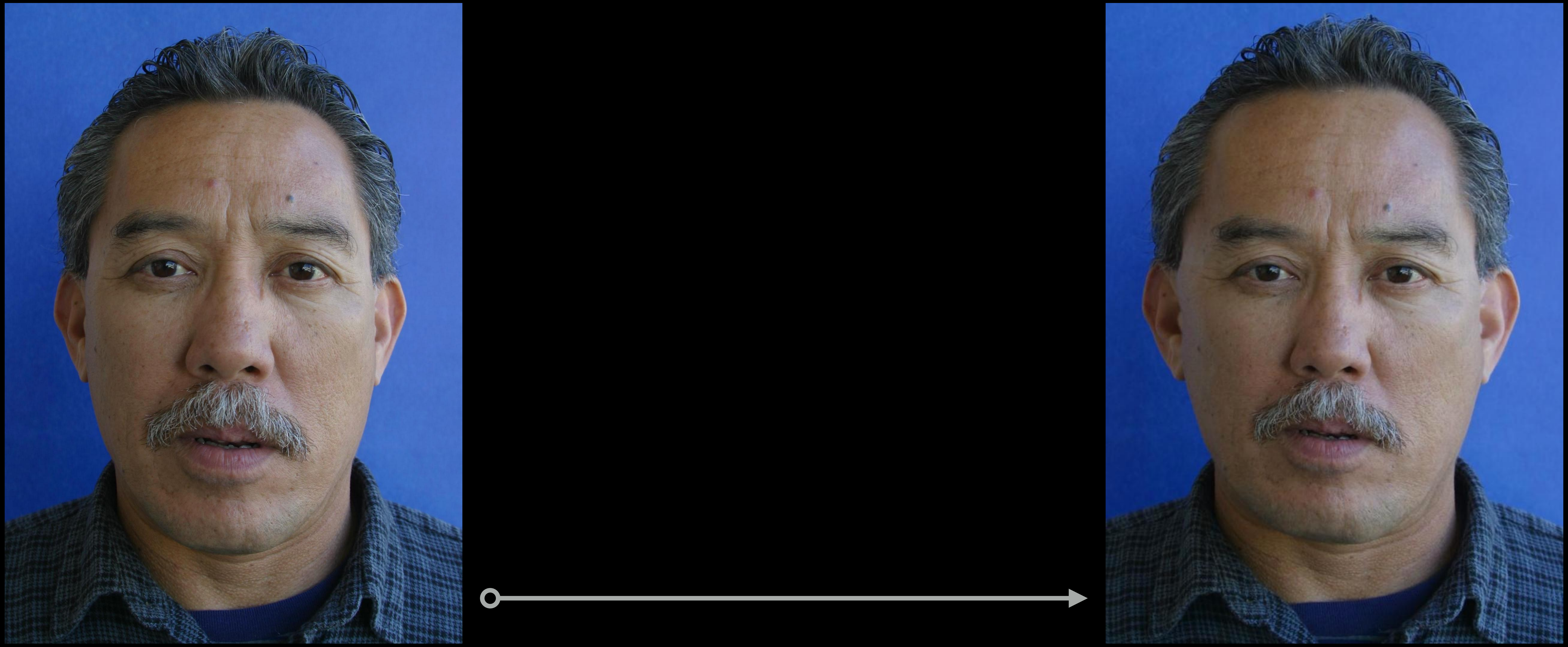
Before

After

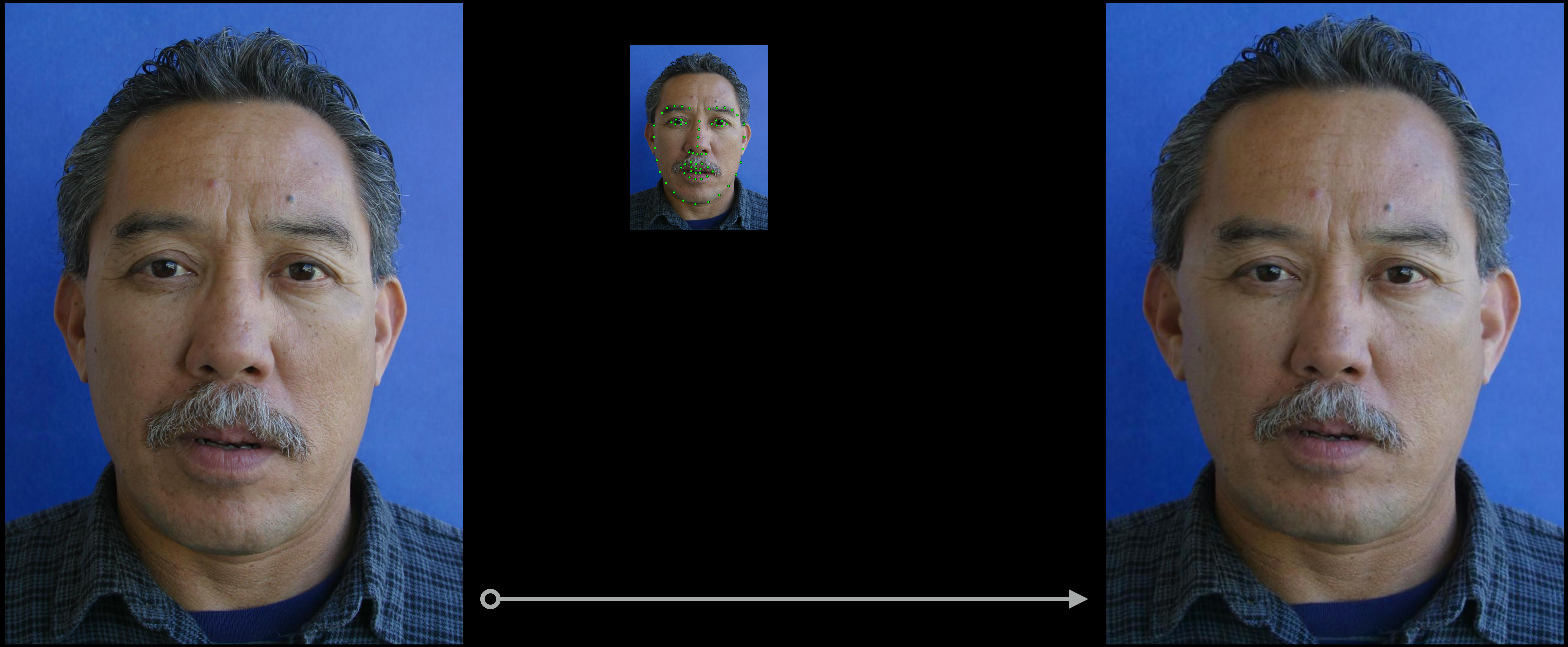
Generate Warp Field

Warp

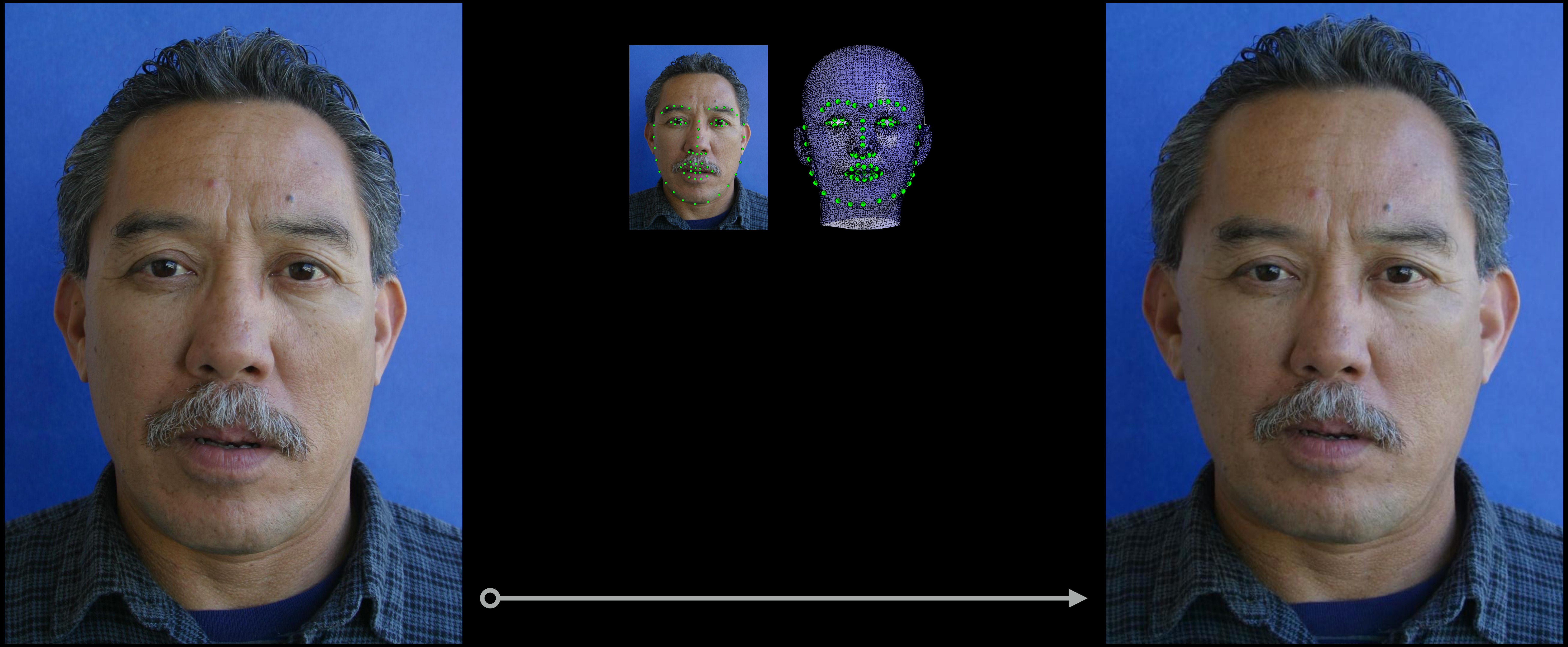
Recap



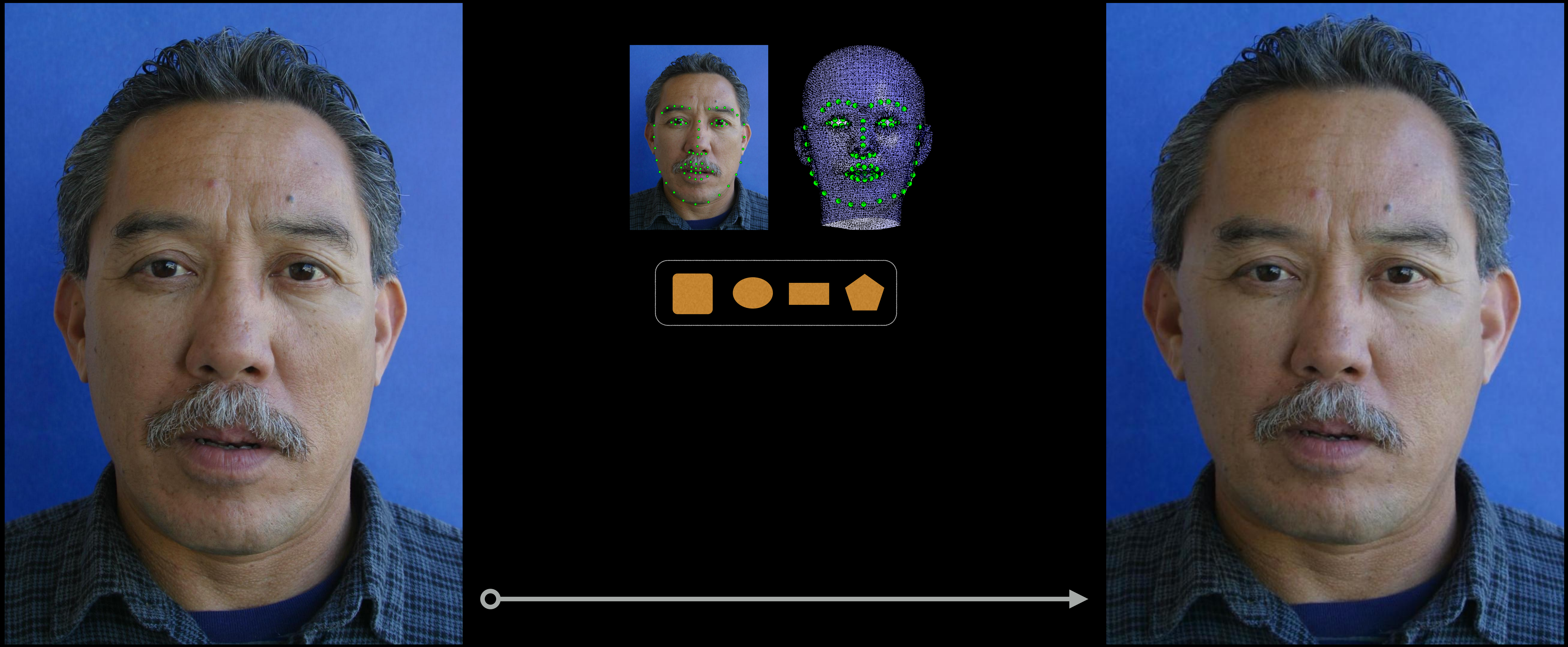
Recap



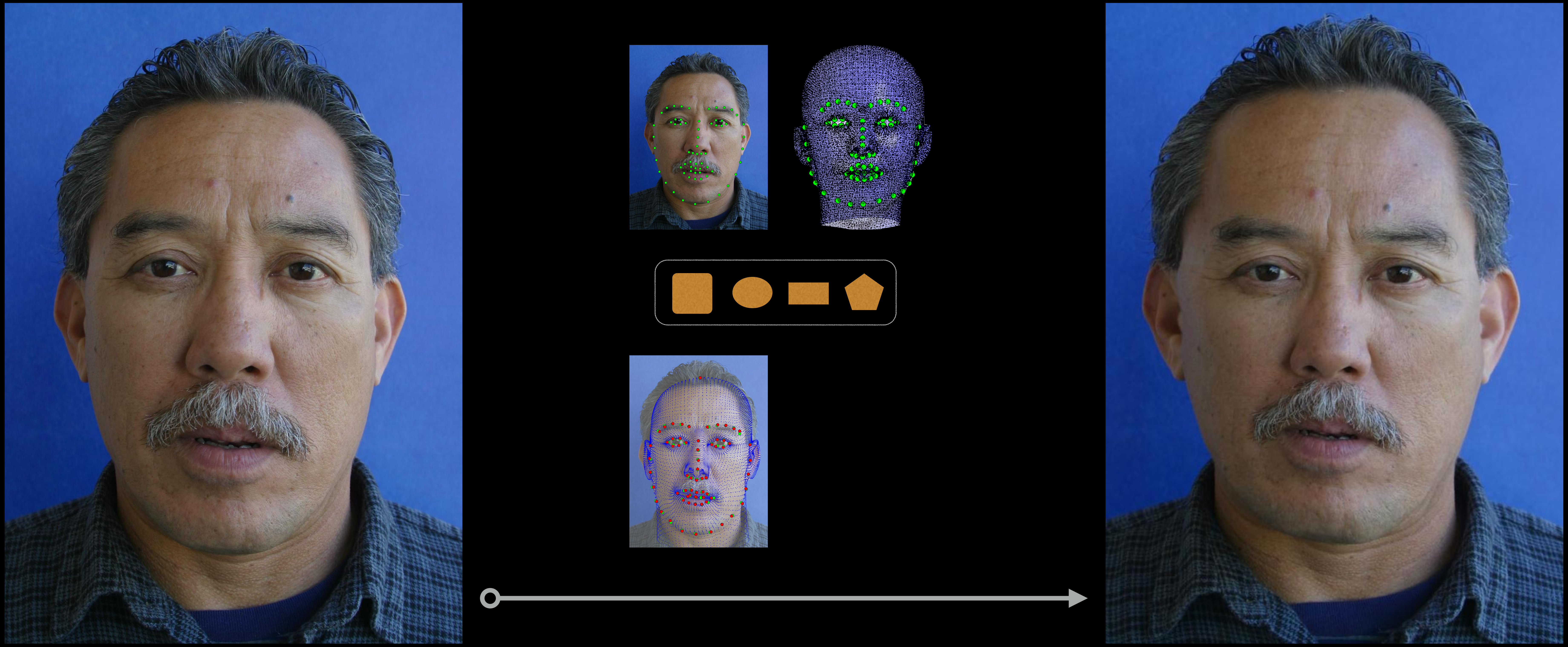
Recap



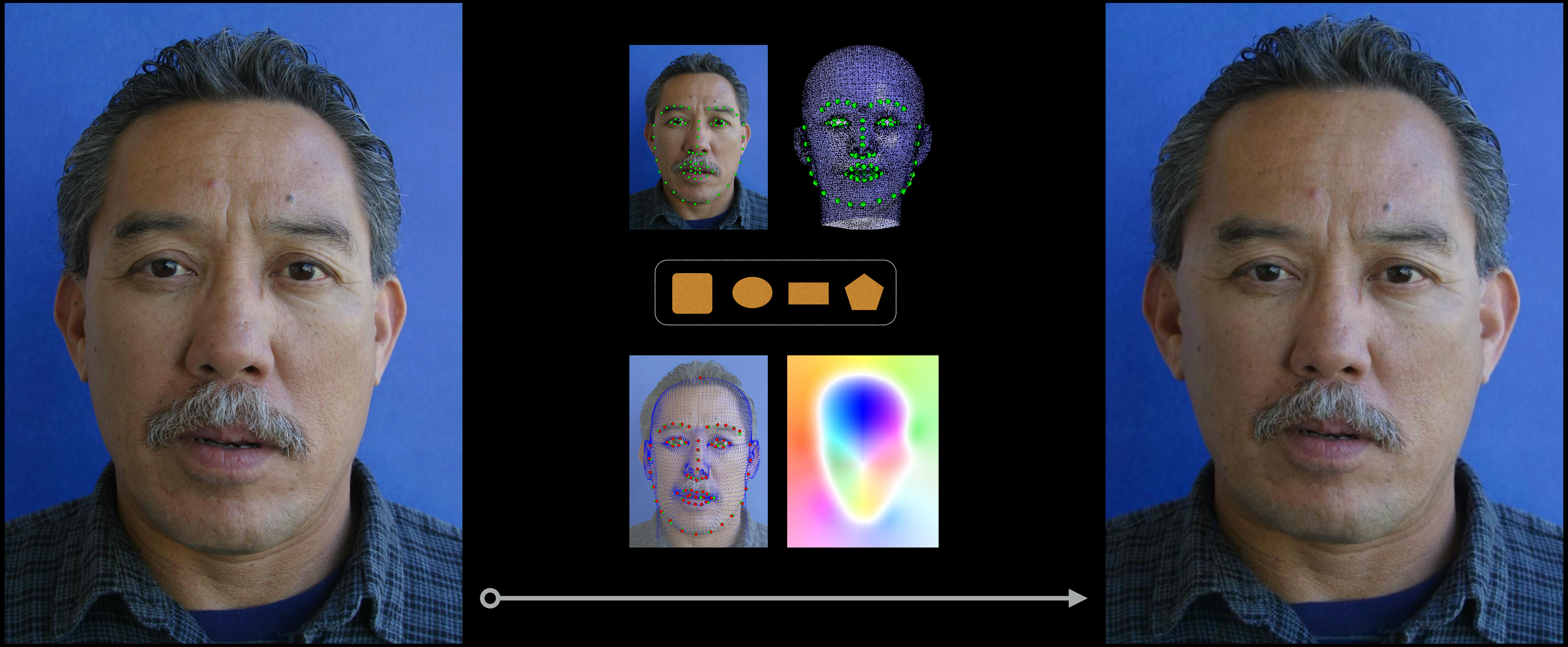
Recap



Recap



Recap



Evaluation

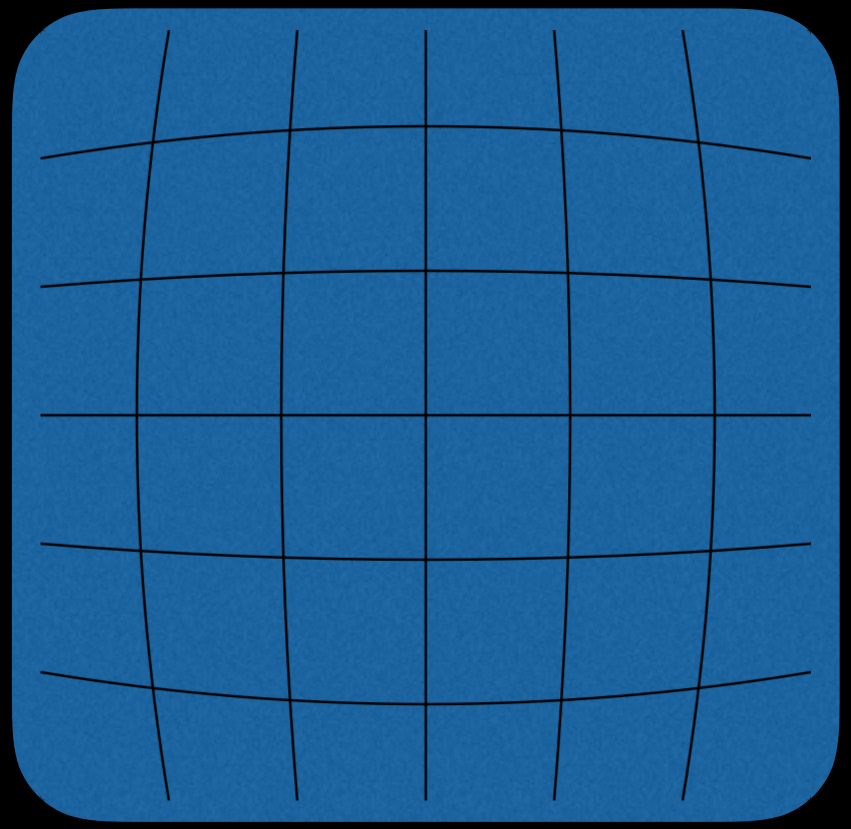
Radial

Face only

Mean head

Ours

Radial

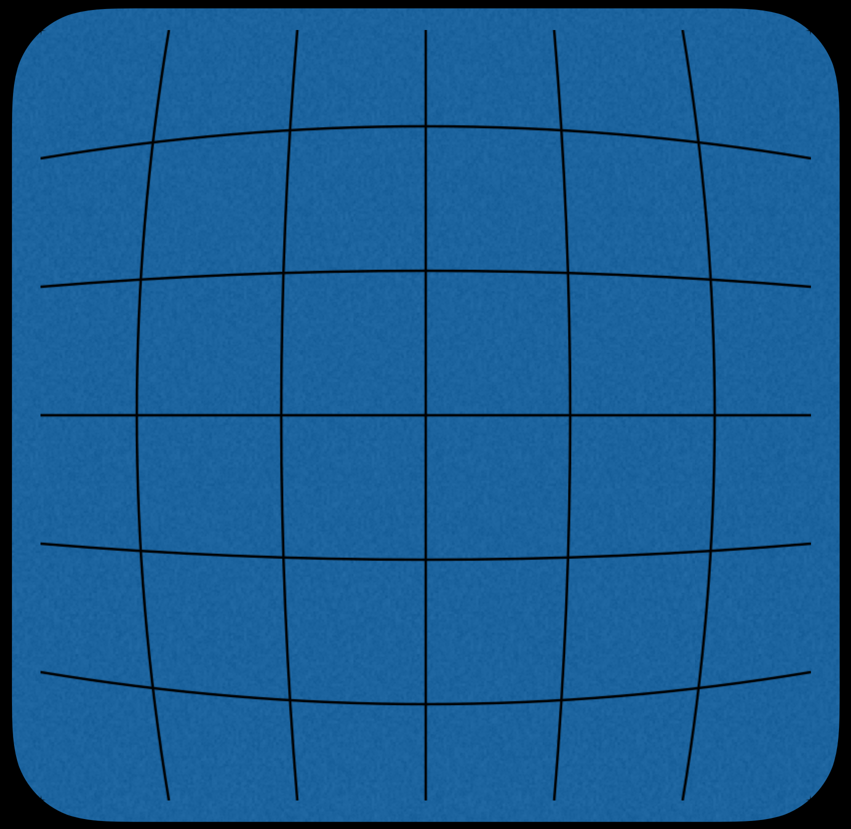


Face only

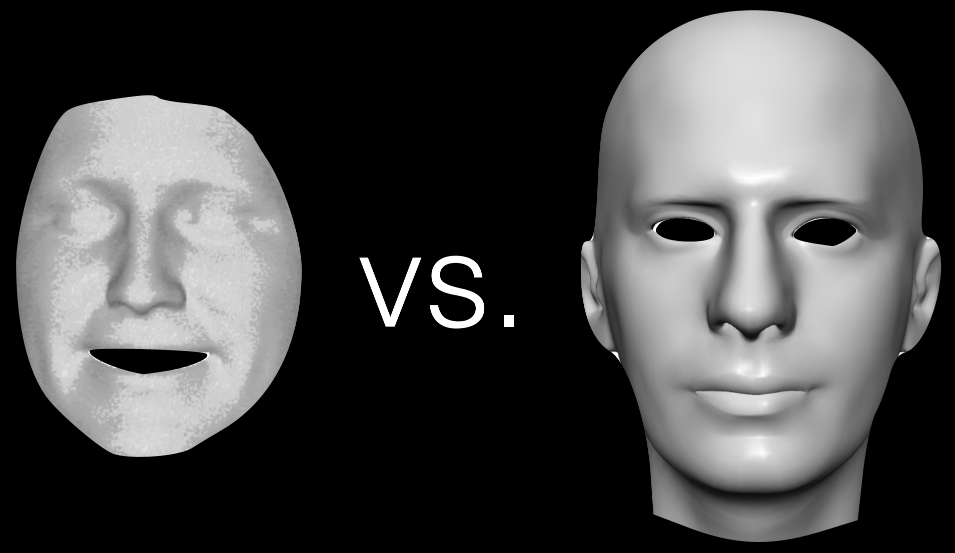
Mean head

Ours

Radial



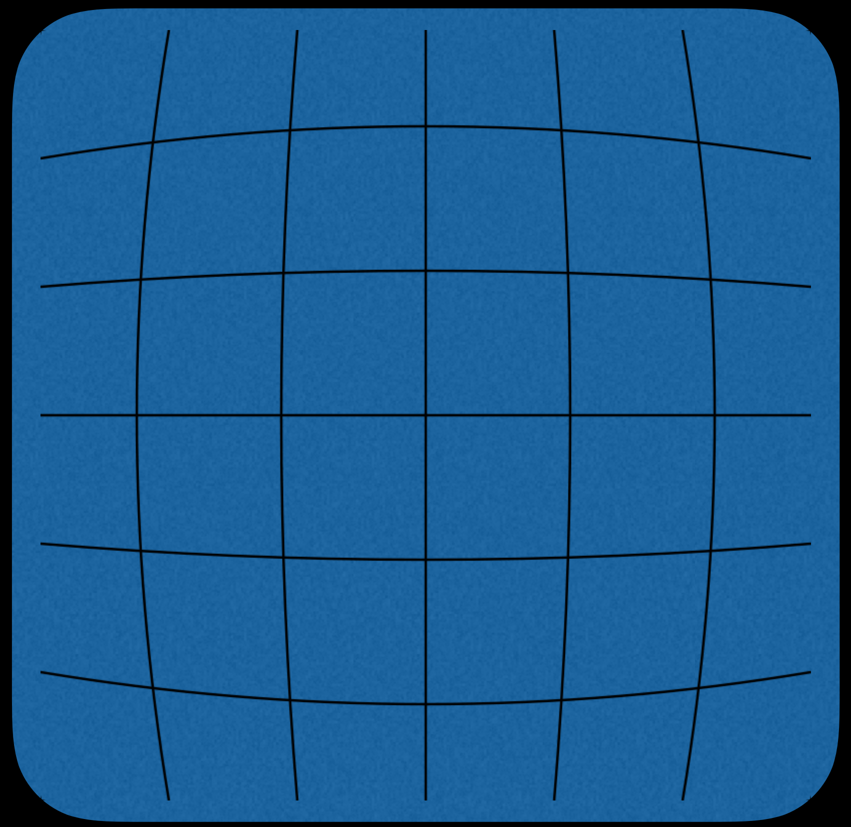
Face only



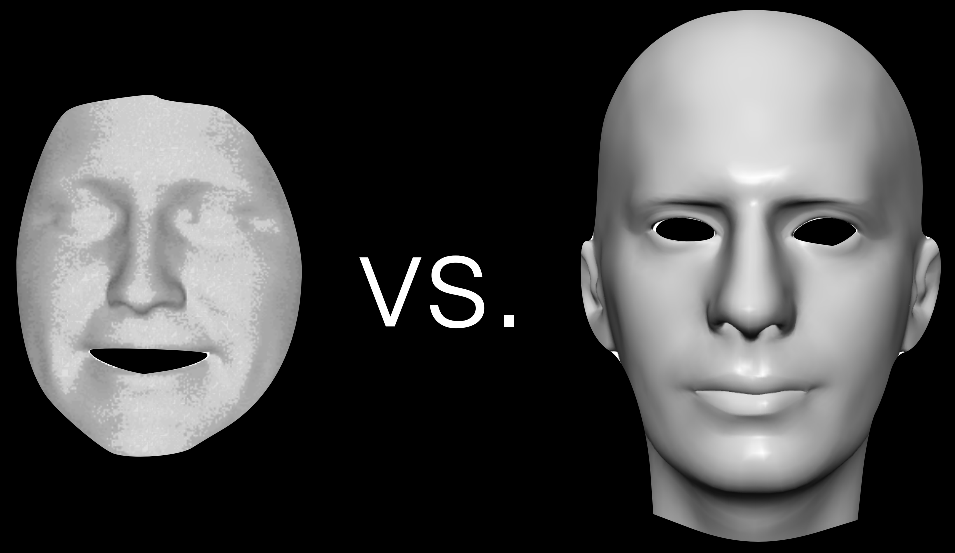
Mean head

Ours

Radial



Face only



[Vlasic et al. 2005]

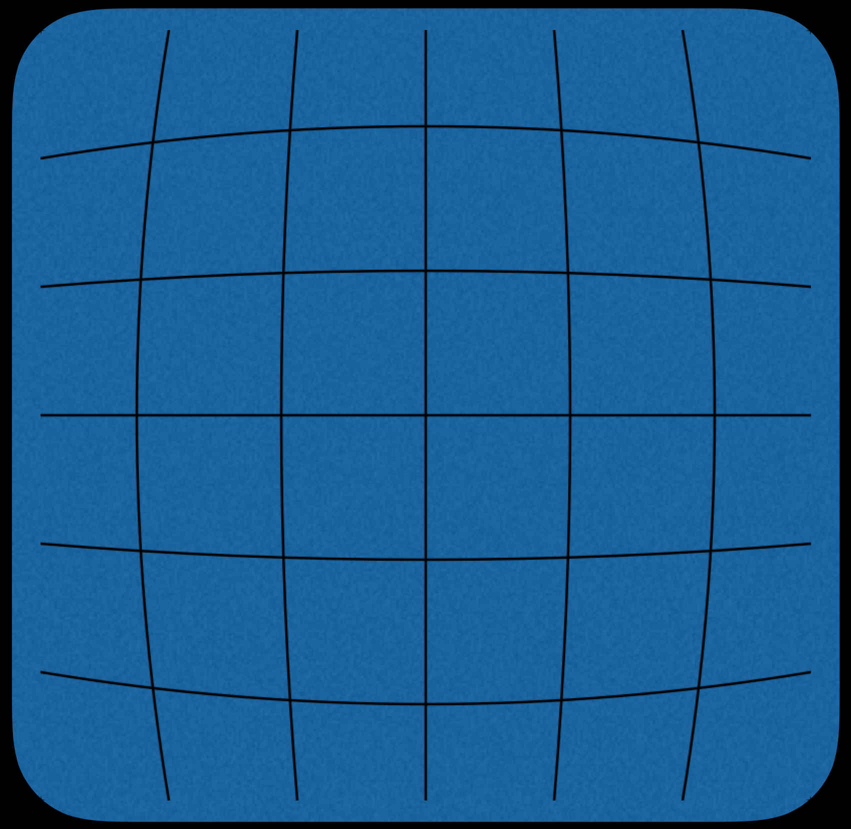
[Yang et al. 2011]

...

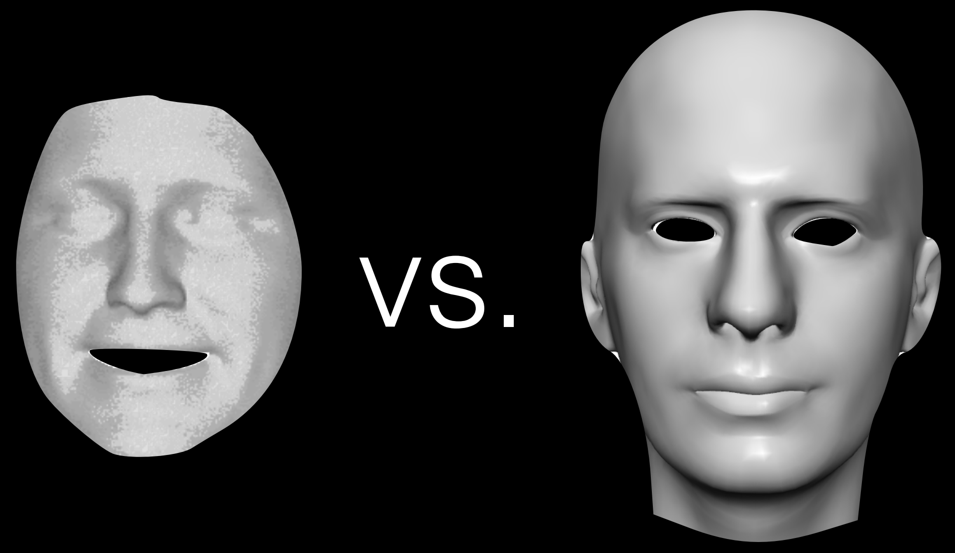
Mean head

Ours

Radial



Face only



[Vlasic et al. 2005]

[Yang et al. 2011]

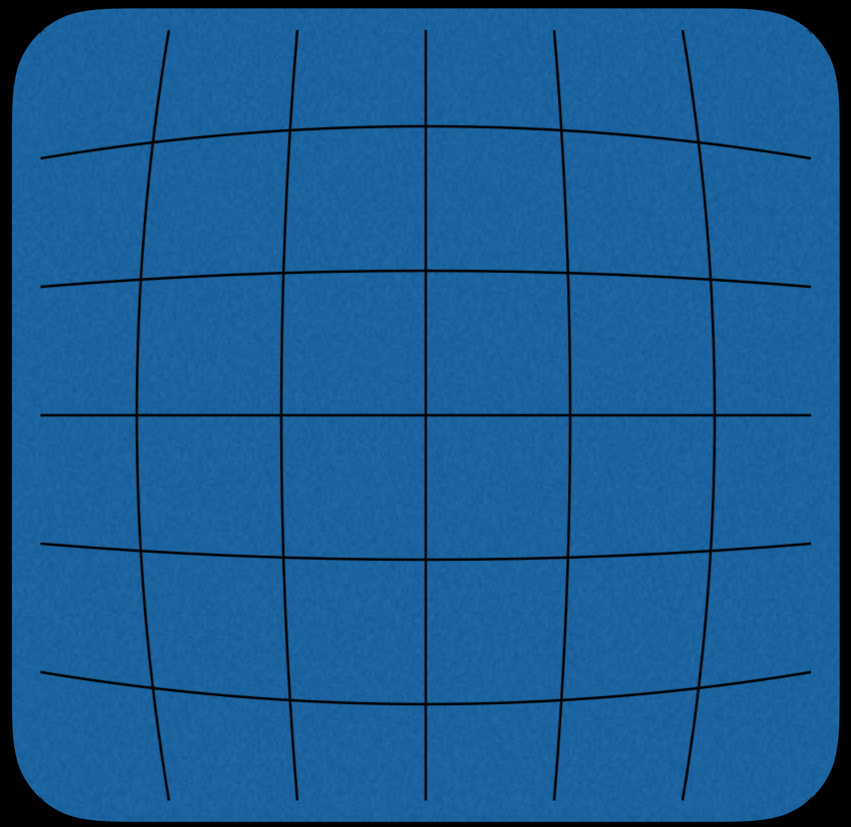
…

Mean head

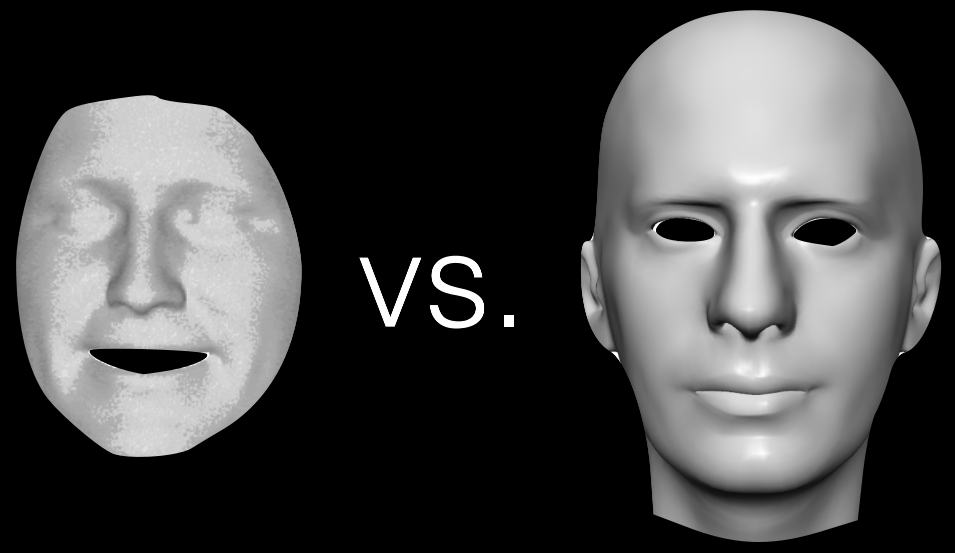
no expression
or identity

Ours

Radial



Face only



[Vlasic et al. 2005]

[Yang et al. 2011]

...

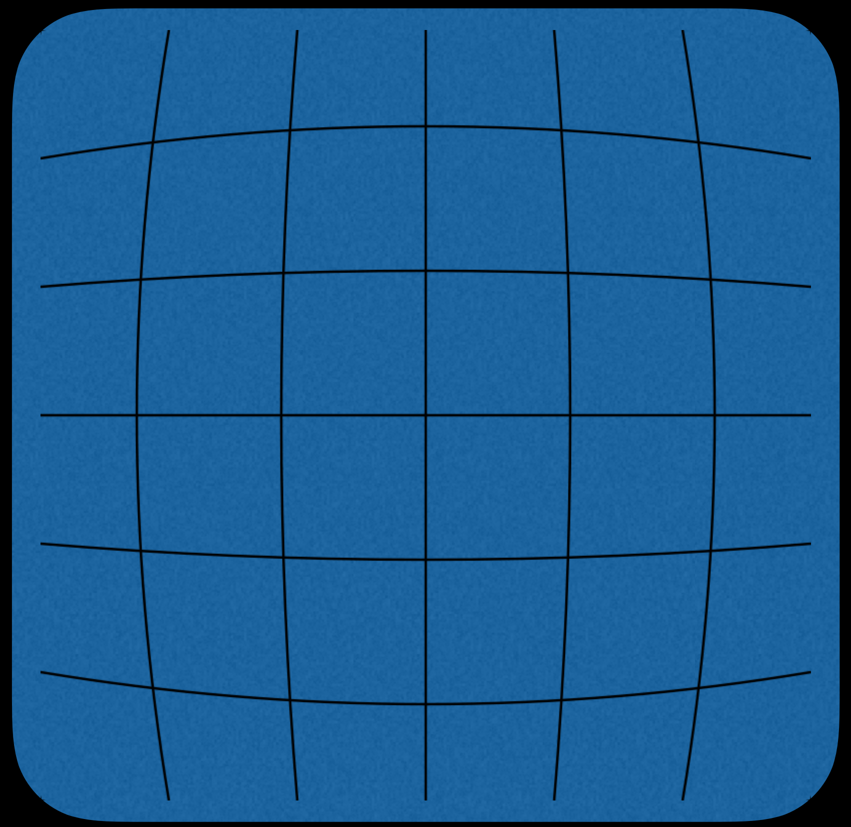
Mean head

no expression
or identity

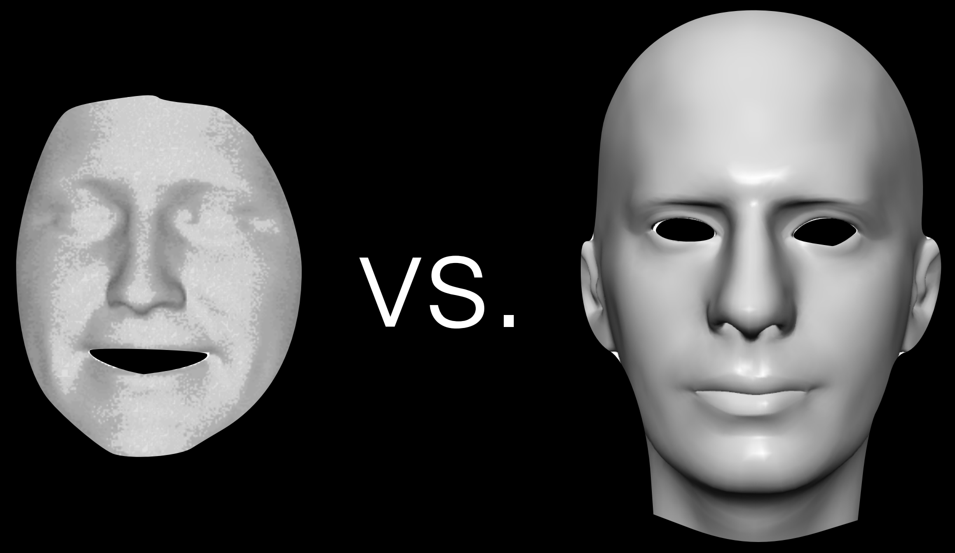
[Hassner et al. 2015]

Ours

Radial



Face only



[Vlasic et al. 2005]

[Yang et al. 2011]

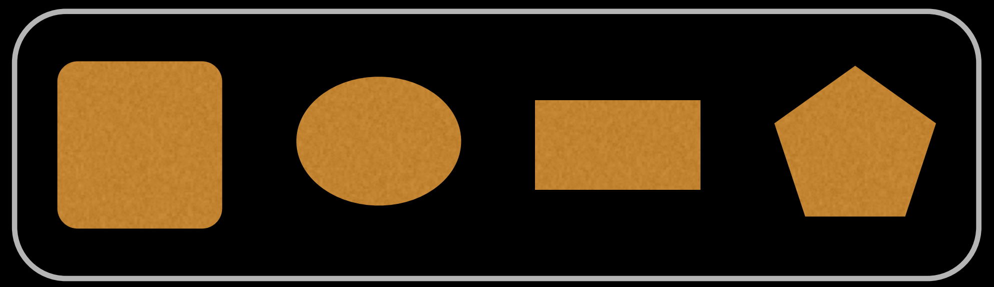
…

Mean head

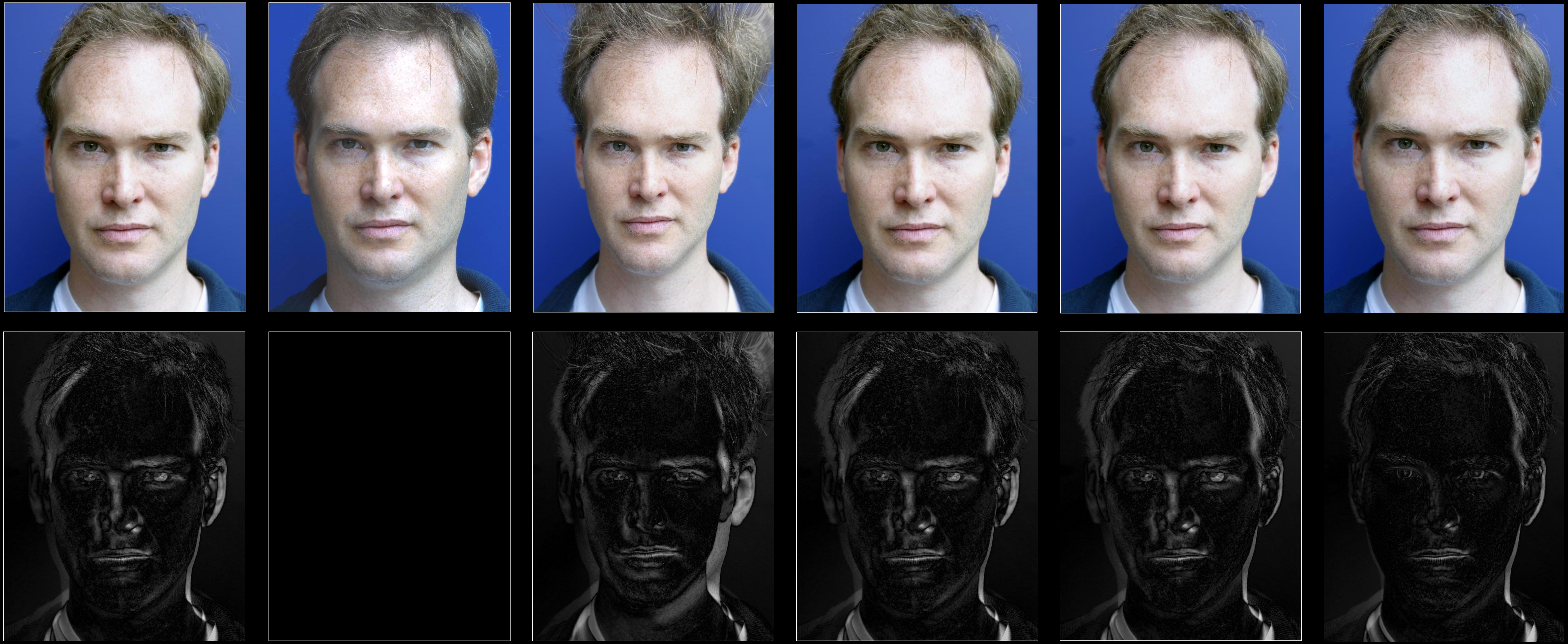
no expression
or identity

[Hassner et al. 2015]

Ours

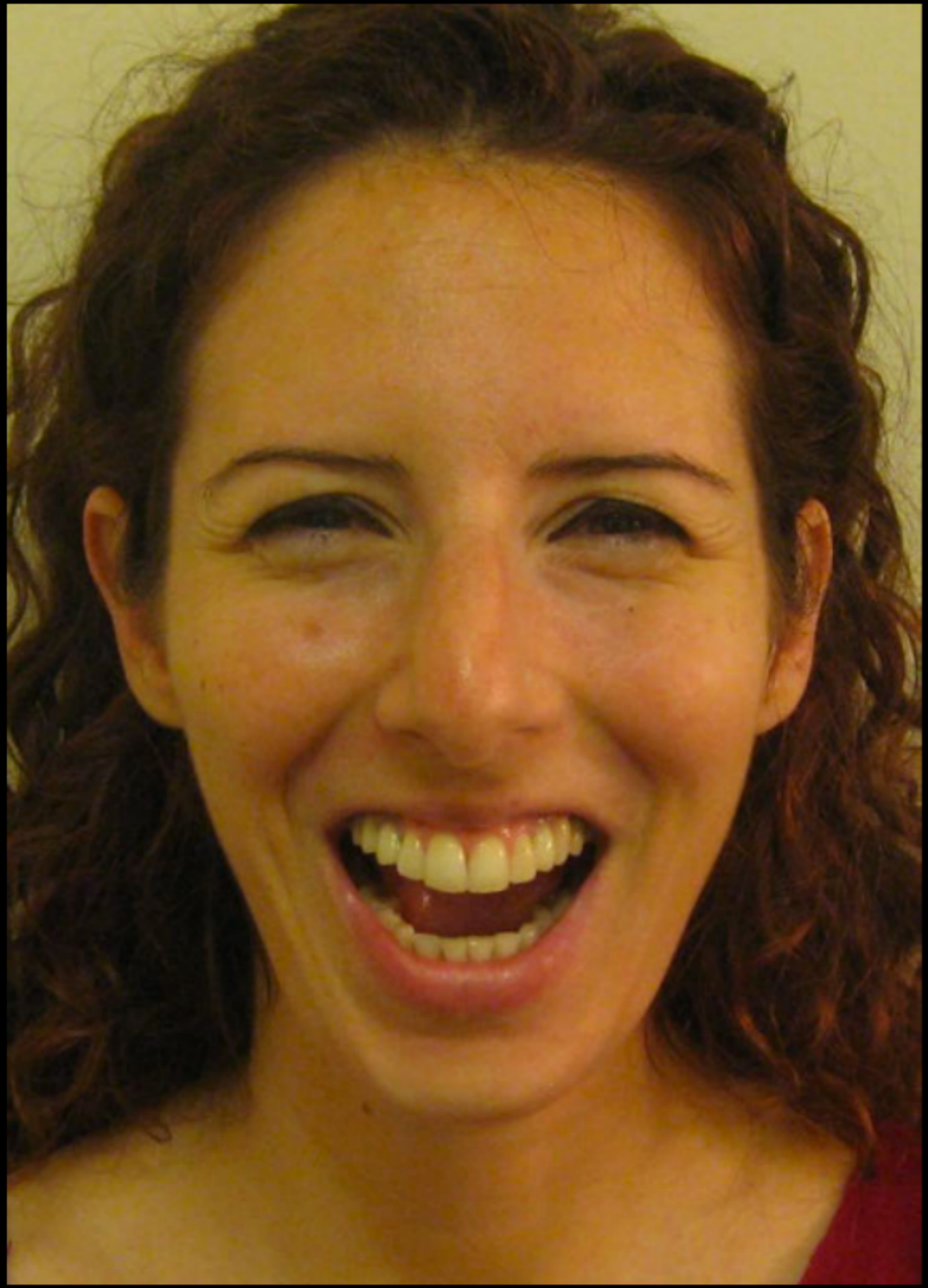


Full model

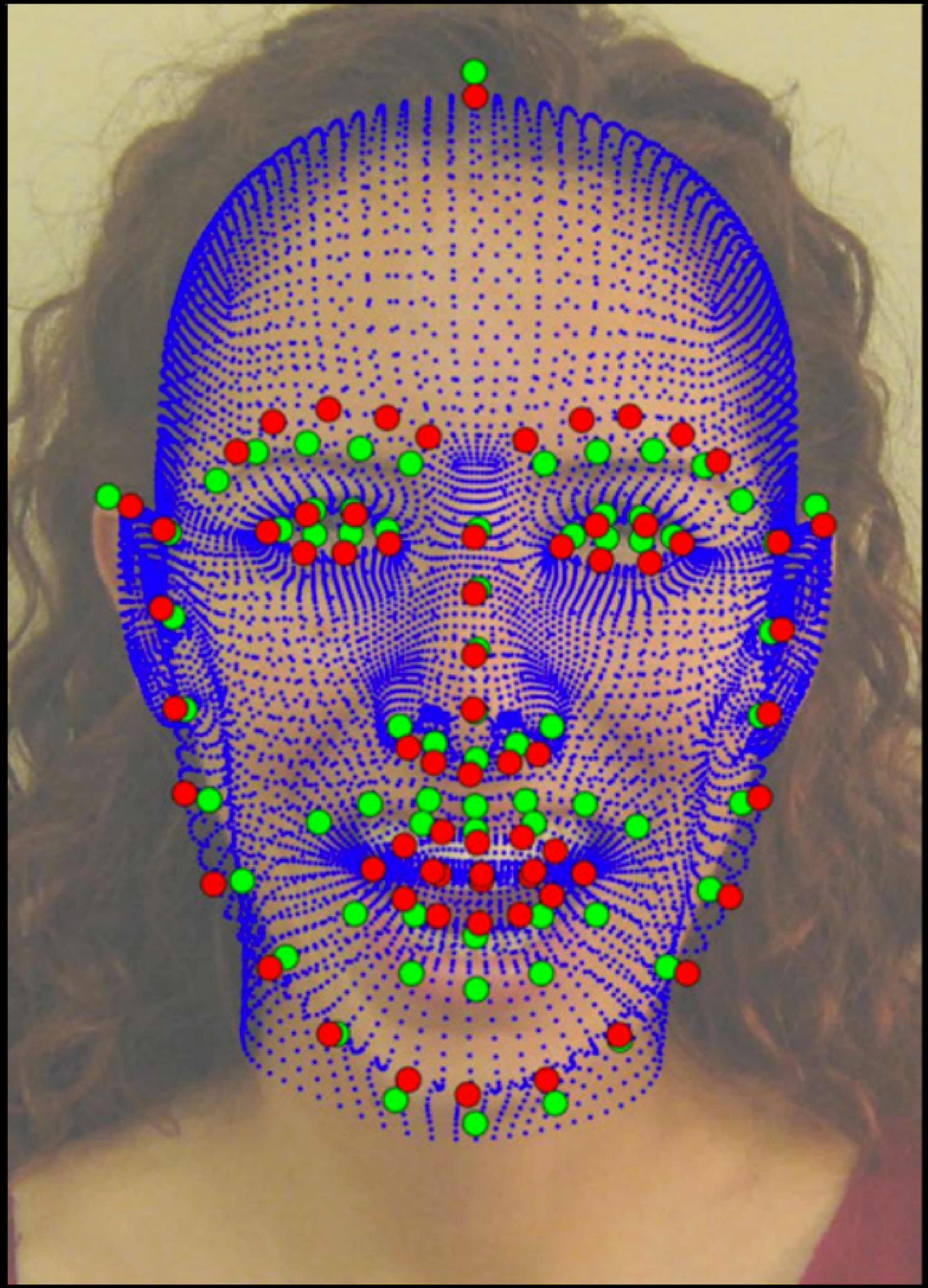


Input Ground truth Radial Face only Mean head Ours

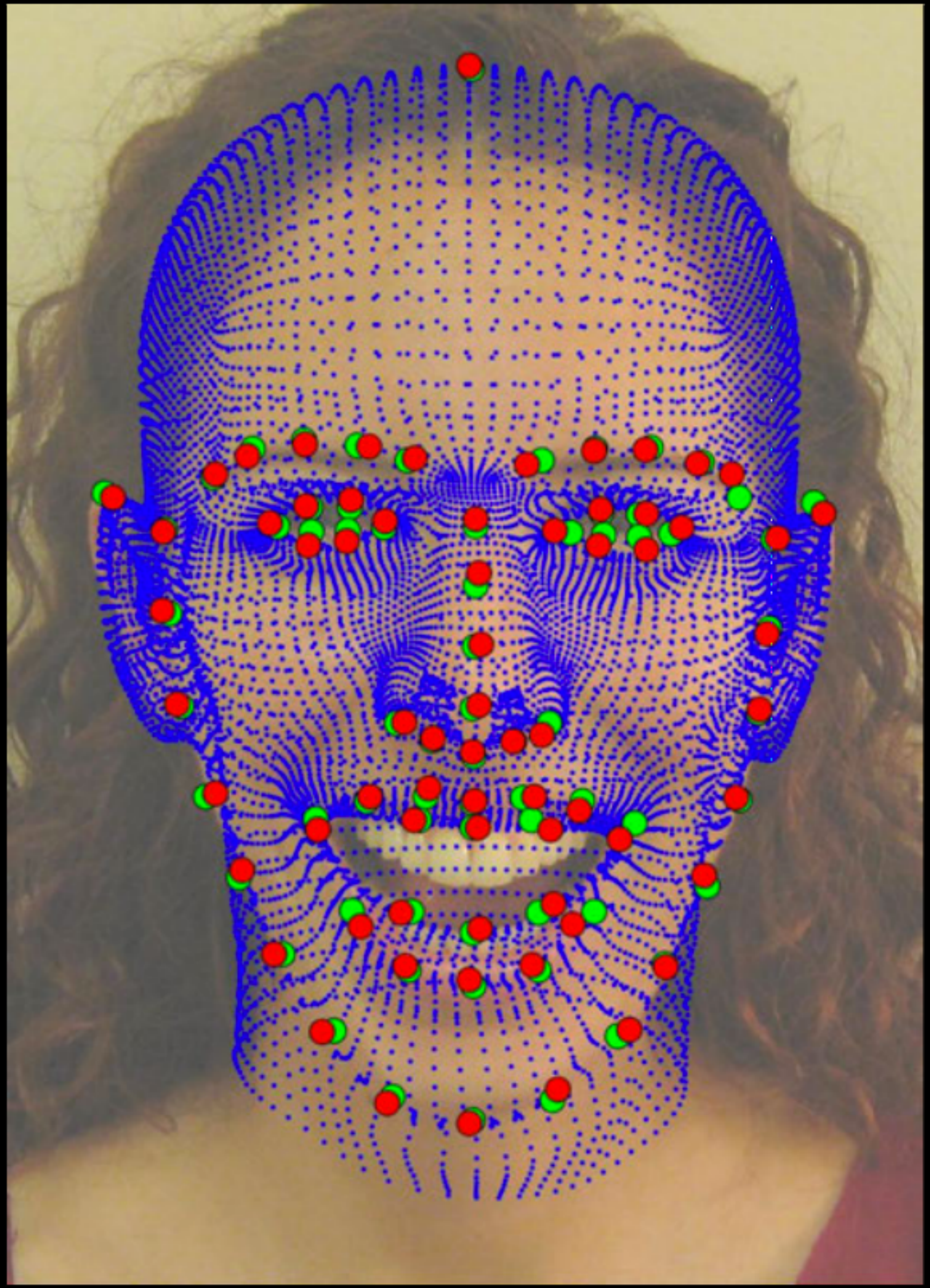
□ no error ■ high error



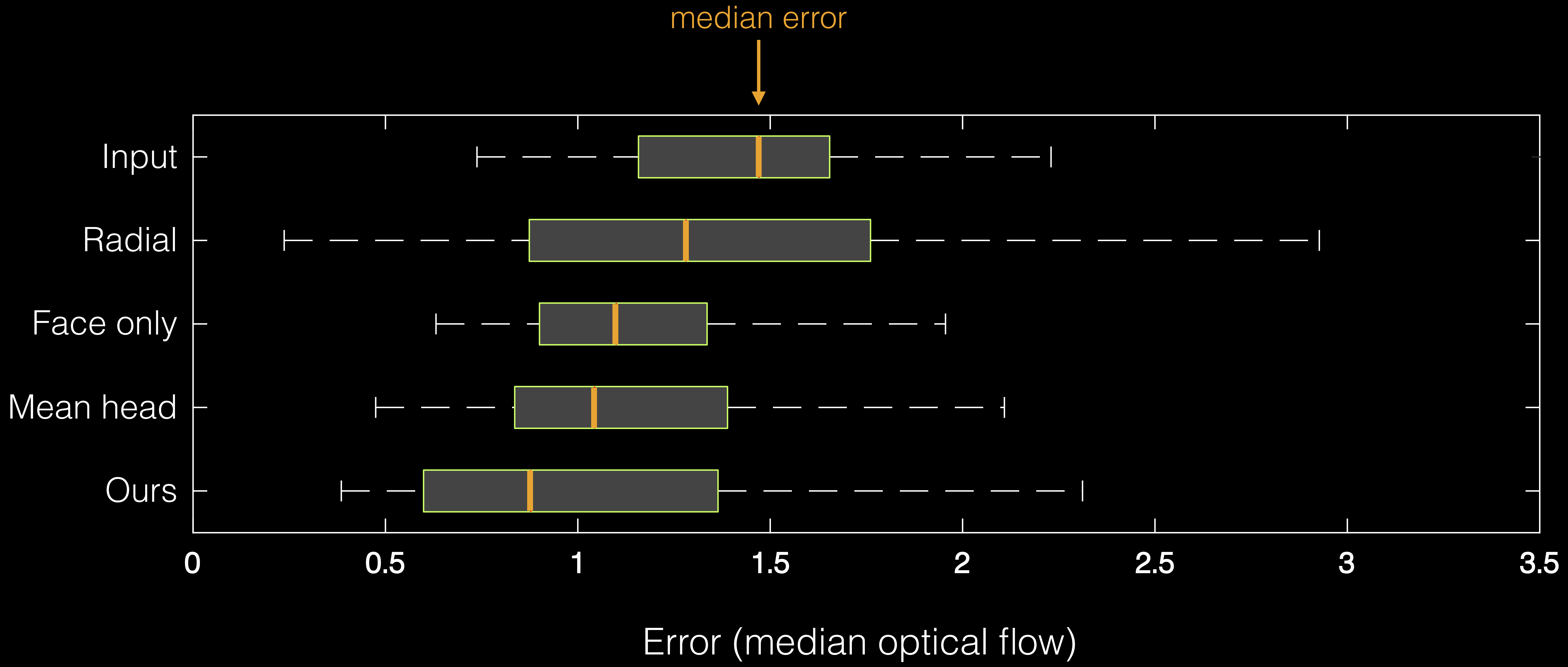
Input



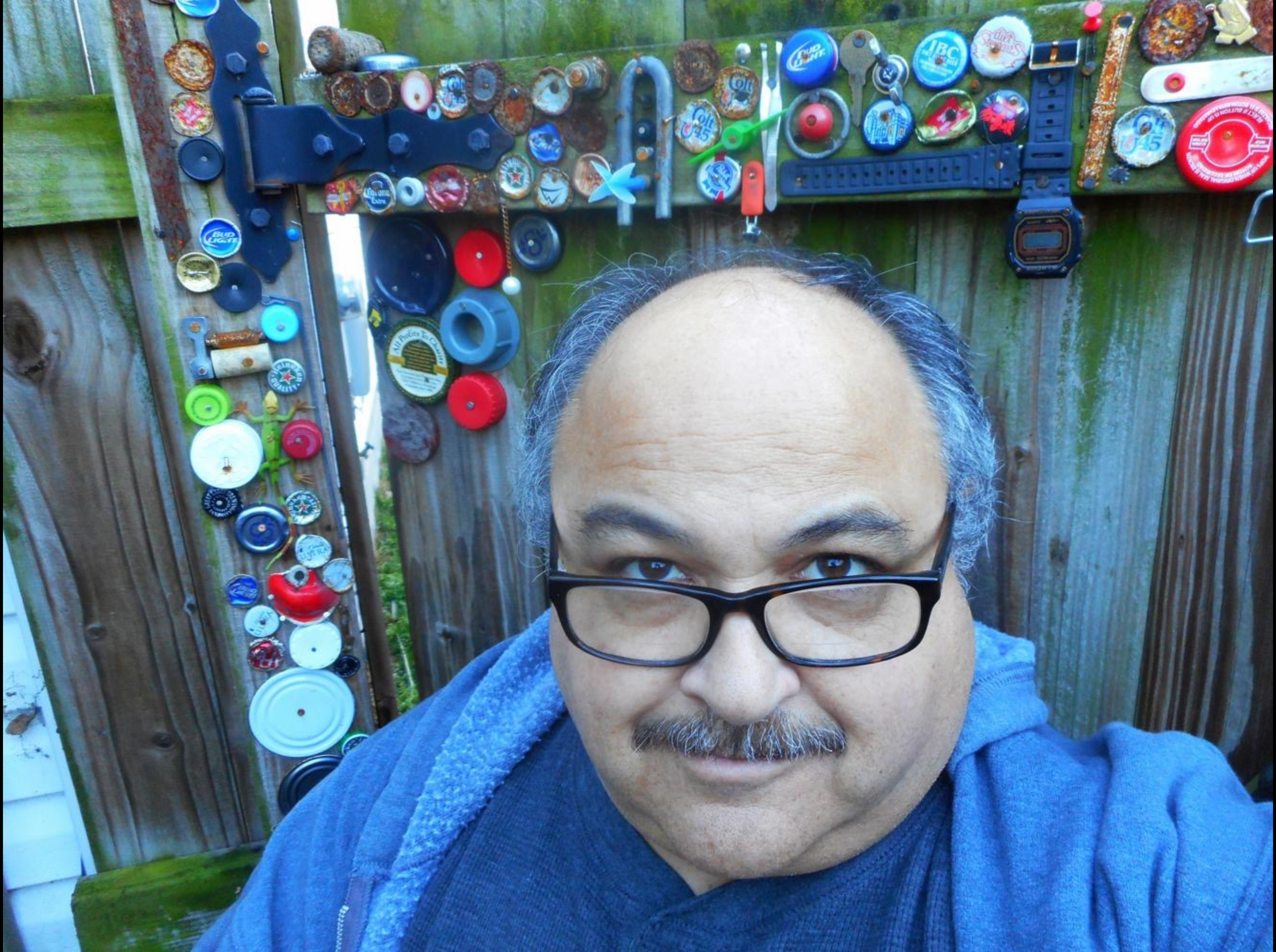
Mean head



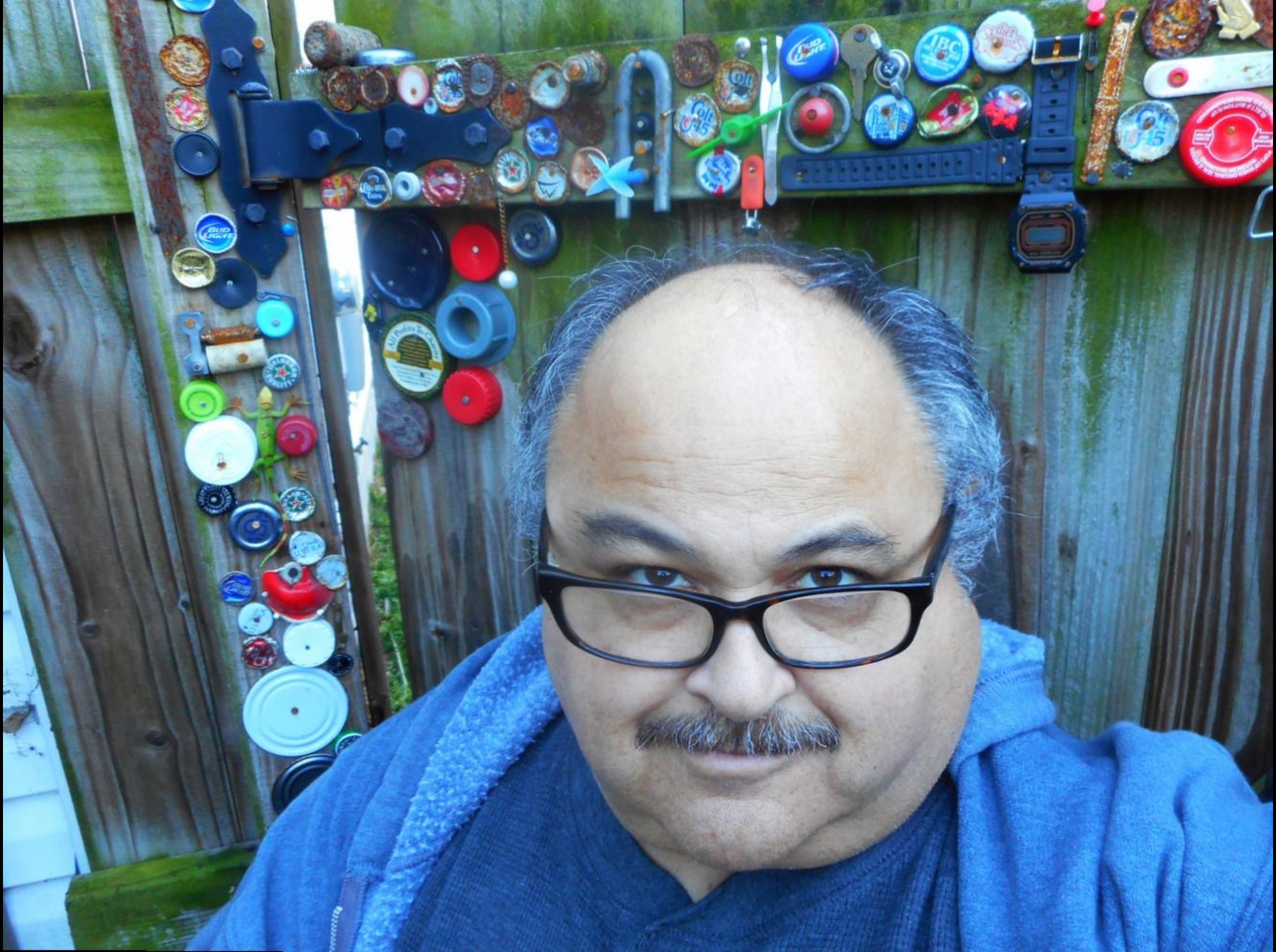
Ours

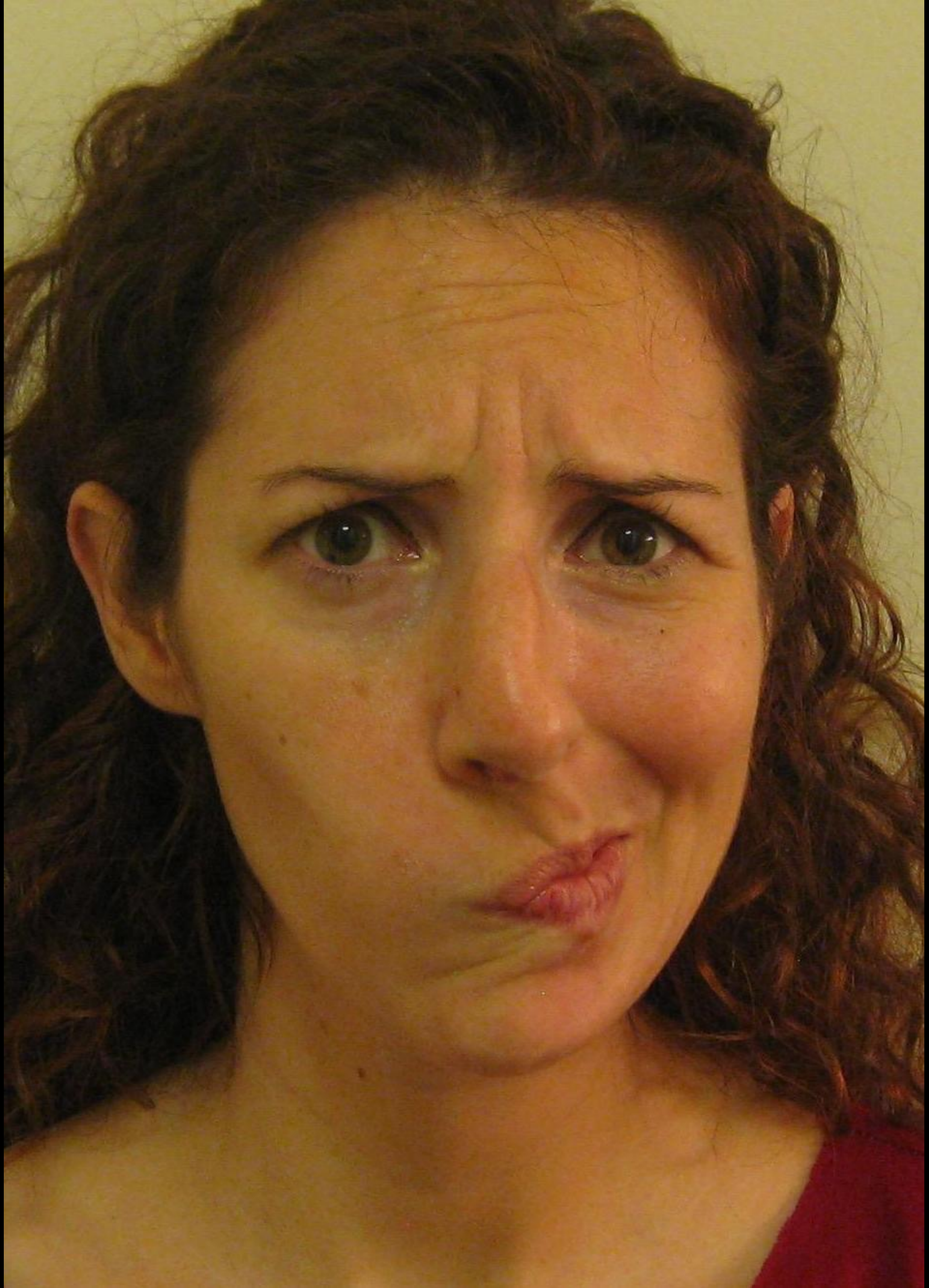

median error
Input
Radial
Face only
Mean head
Ours
0
0.5
1
1.5
2
2.5
3
3.5
Error (median optical flow)

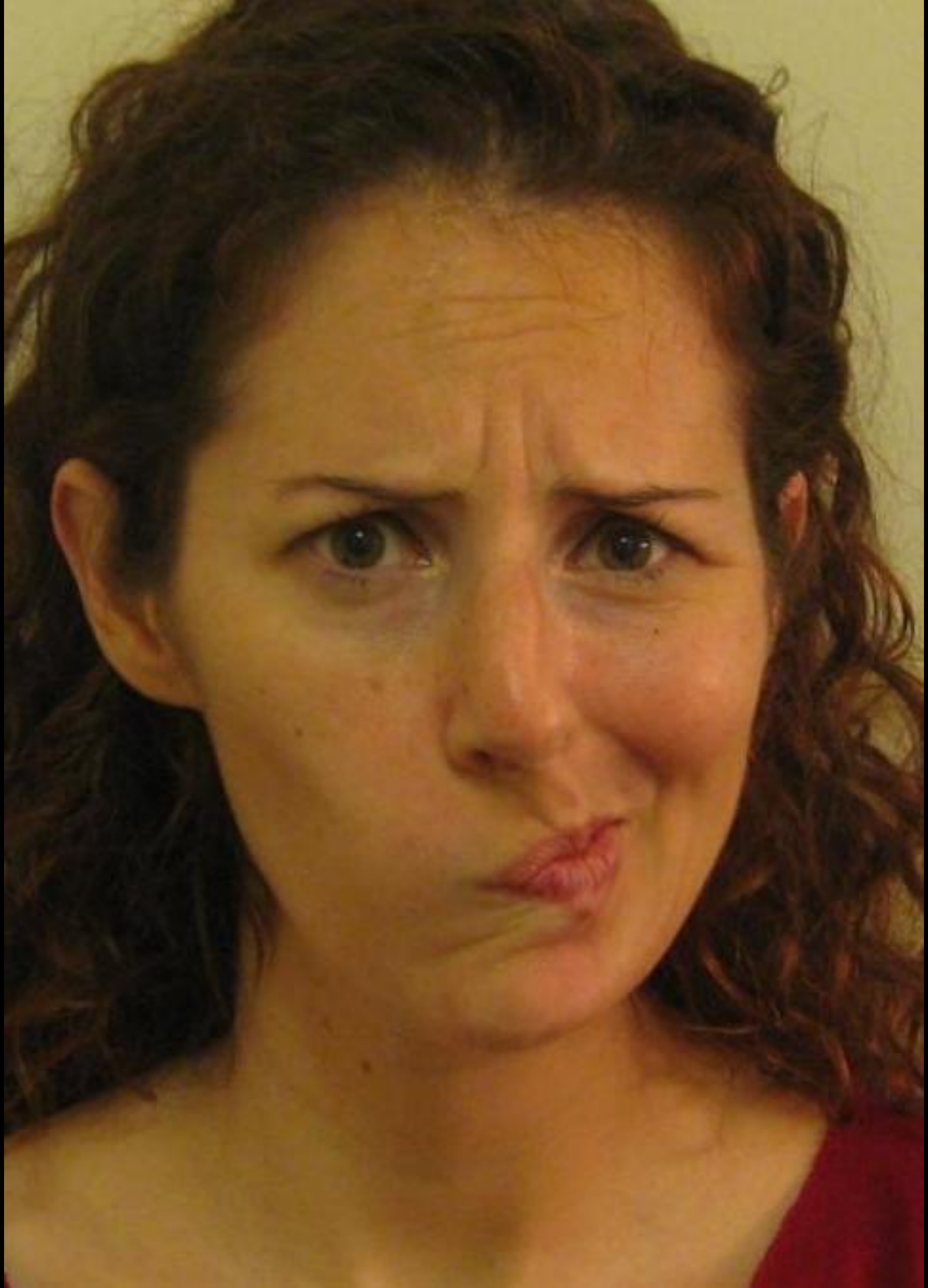
MASS



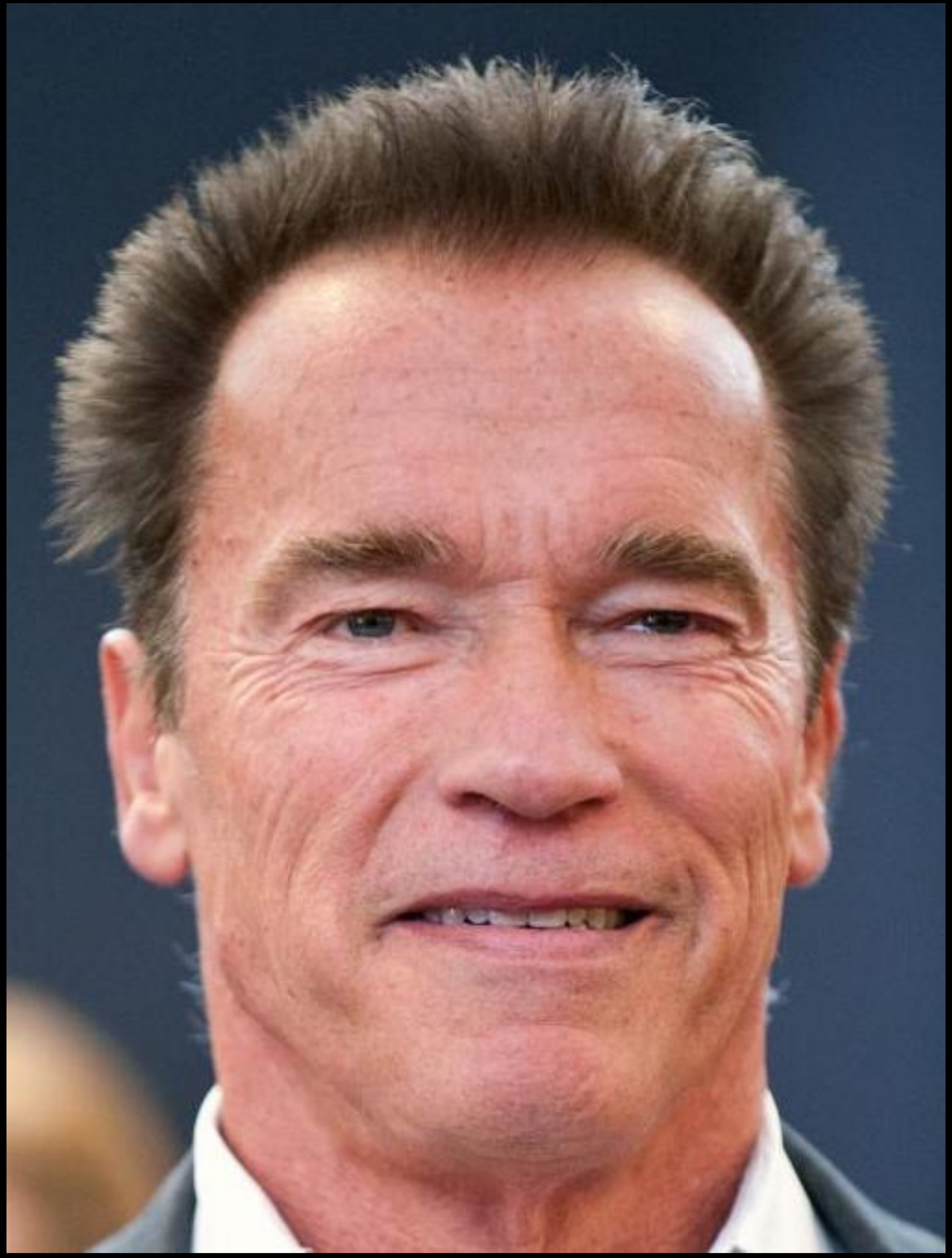
BUD LIGHT
Colt 45
IBC
Colt 45
All Profits To Charity


BUD LIGHT
Colt 45
JBC
BUD LIGHT
Colt 45
All Profits To Charity



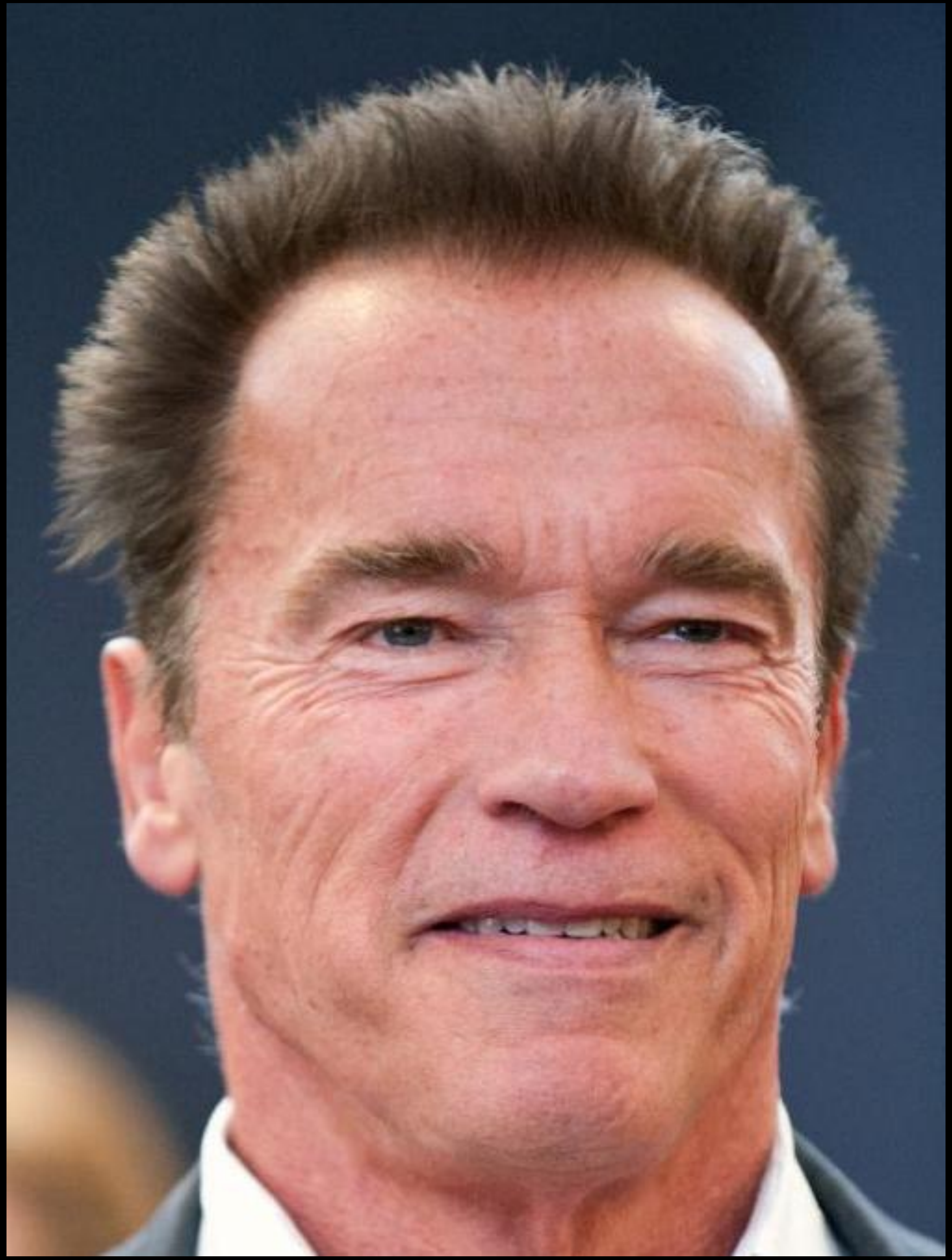


Rotation



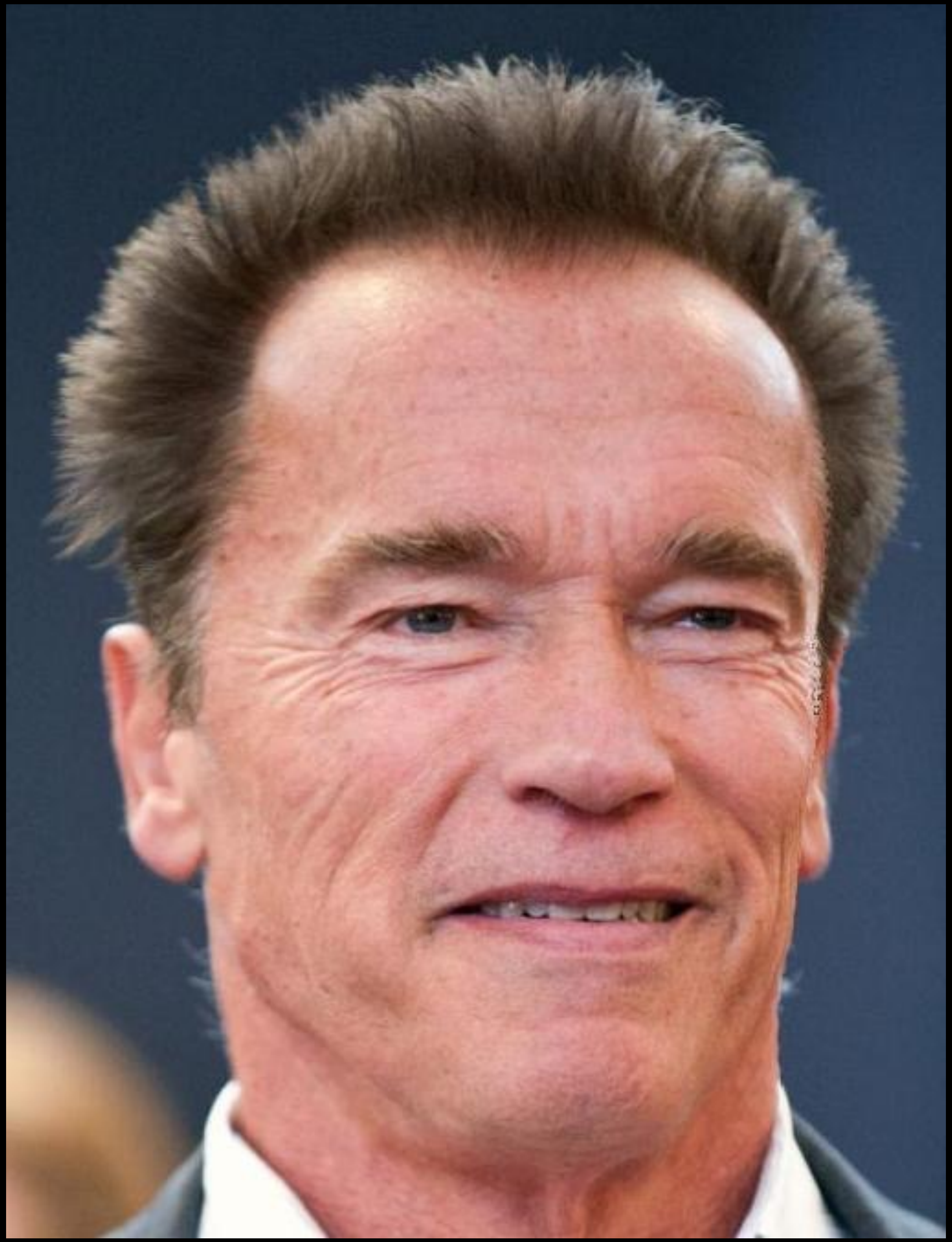
Input

Rotation



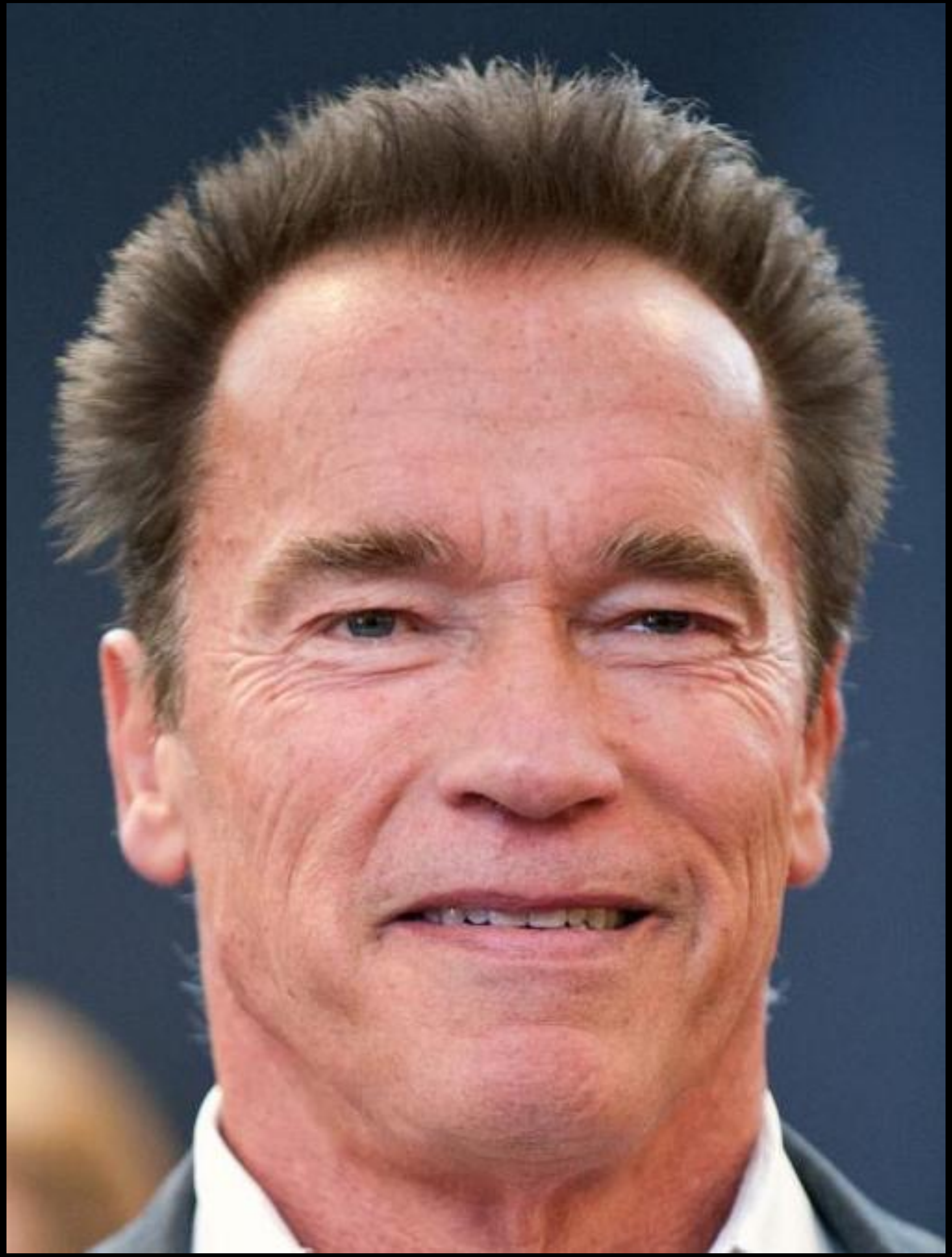
Left 4°

Rotation



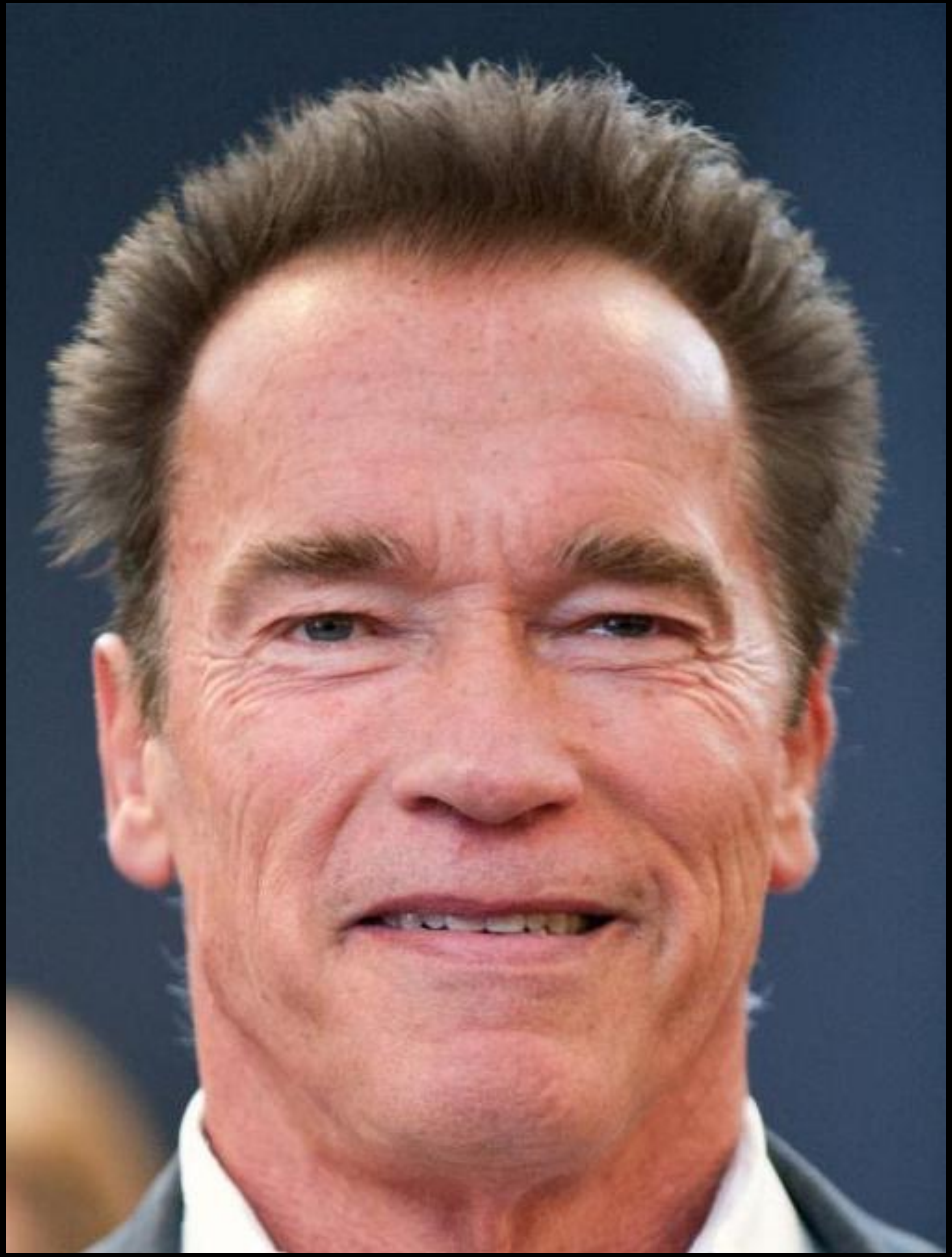
Left 7°

Rotation



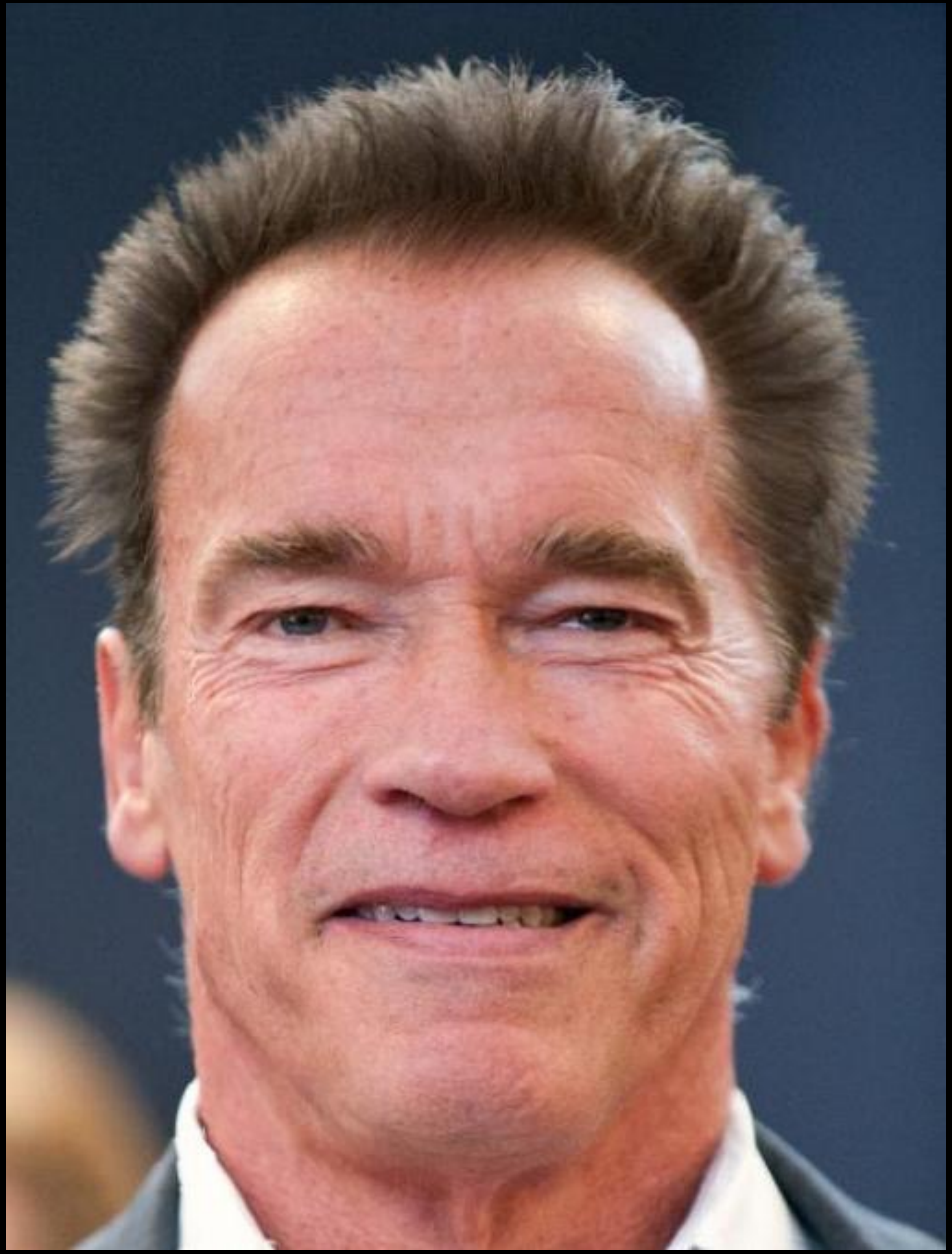
Input

Rotation



Right 4°

Rotation



Right 7°

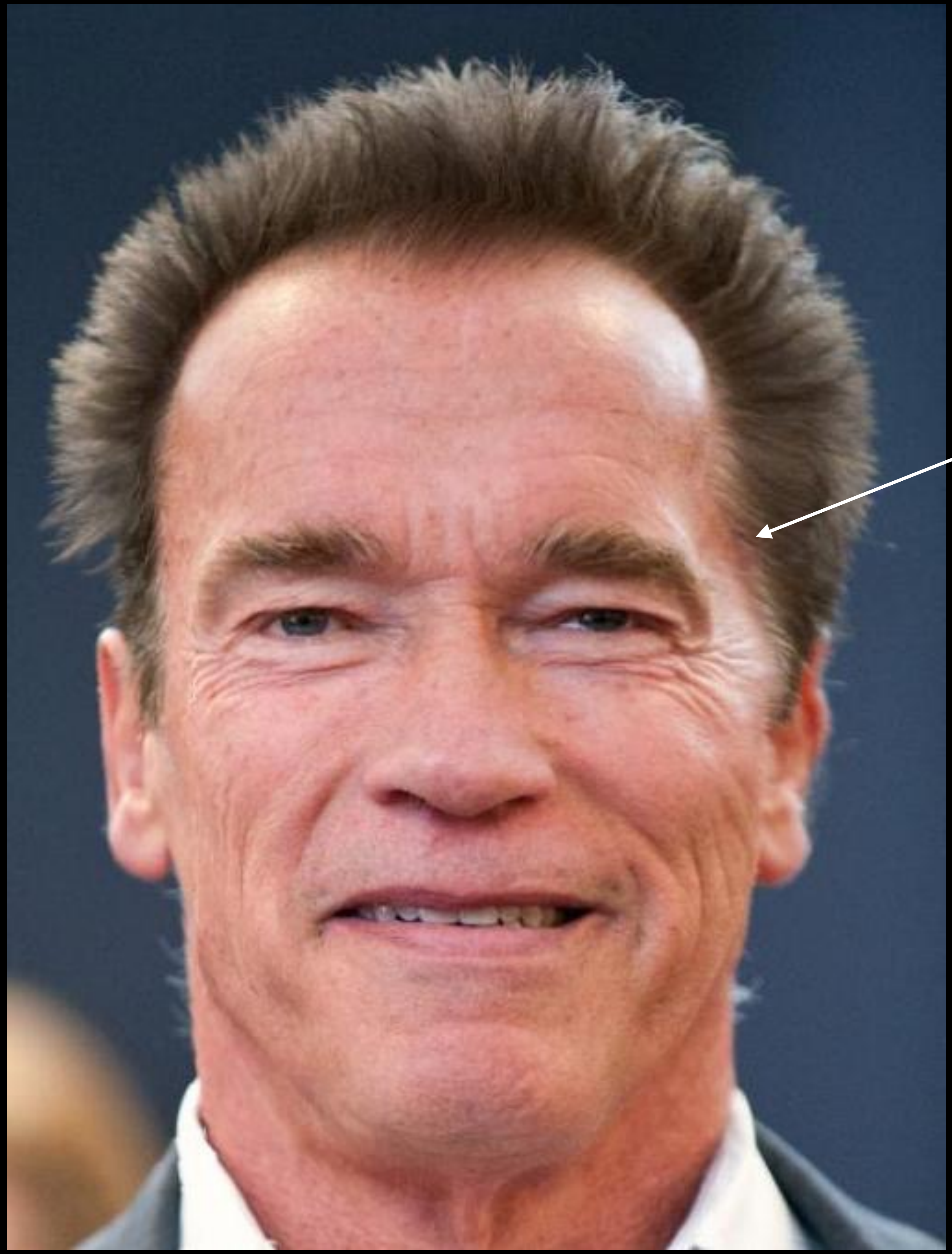
Rotation

Useful for Face Identification!

(e.g. DeepFace ’14)

Right 7°

Rotation



Not Perfect

Right 7°

Useful for Face Identification!

(e.g. DeepFace ’14)

Limitations & Future Work

Limitations & Future Work

- Fitting may fail

Limitations & Future Work

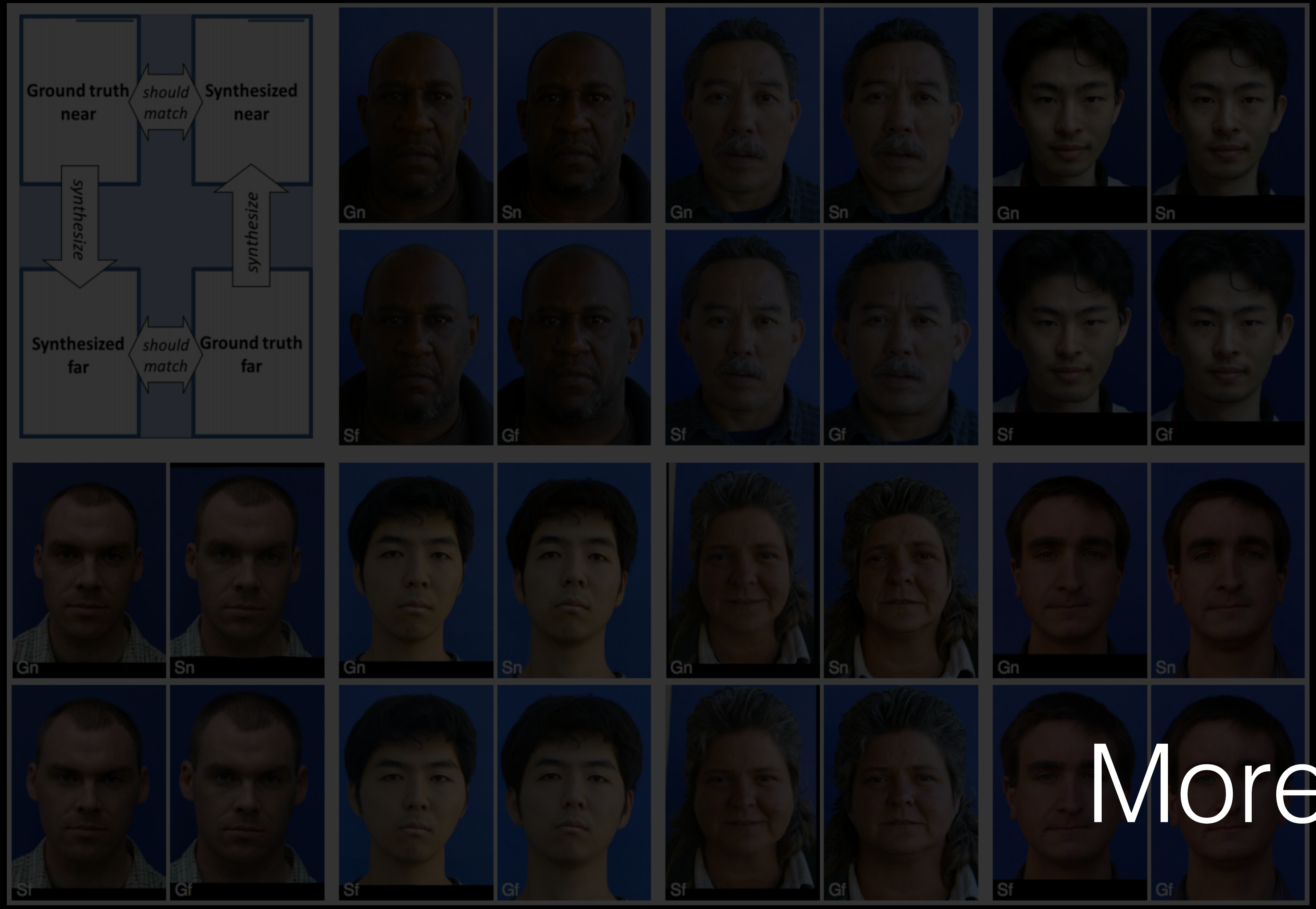
- Fitting may fail
    - Use more data — RGB values

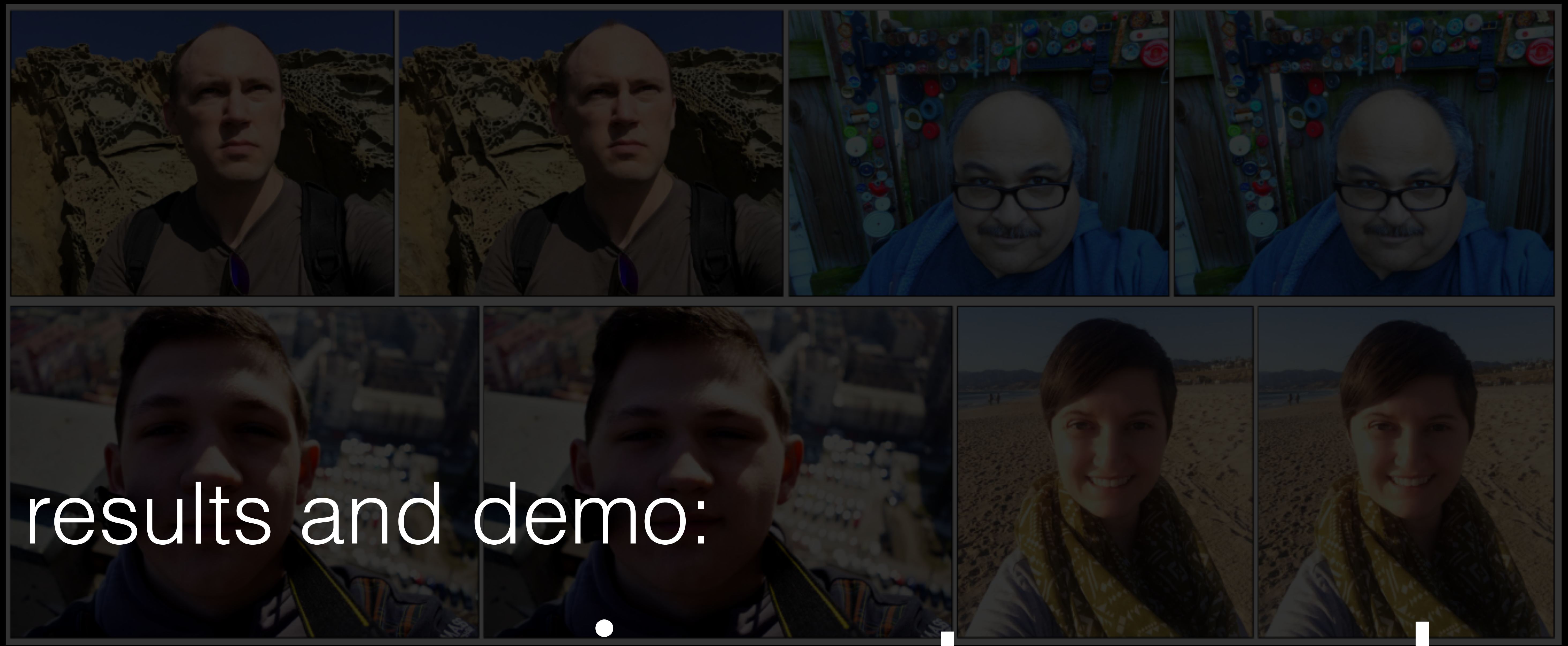
Limitations & Future Work

- Fitting may fail
  - Use more data — RGB values
- Does not support extreme rotations

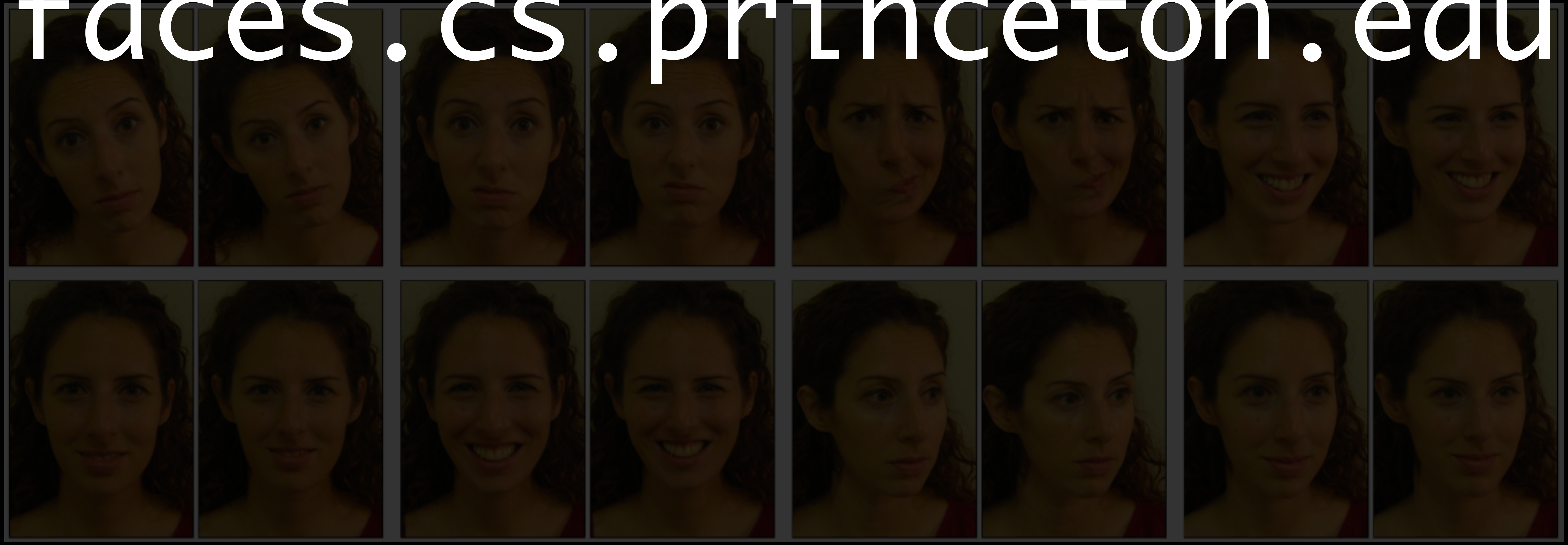
Limitations & Future Work

- Fitting may fail
  - Use more data — RGB values
- Does not support extreme rotations
  - Hallucinate missing features





More results and demo:



http://faces.cs.princeton.edu

Constrained Local Models

Improving Portraits

Editing Video

Editing Video

Editing
Video

Text-based Editing of Talking-head Video

Ohad Fried · Ayush Tewari · Michael Zollhöfer · Adam Finkelstein · Eli Shechtman ·
Dan Goldman · Kyle Genova · Zeyu Jin · Christian Theobalt · Maneesh Agrawala

Video Blogs

Interviews

Online Courses

Speeches

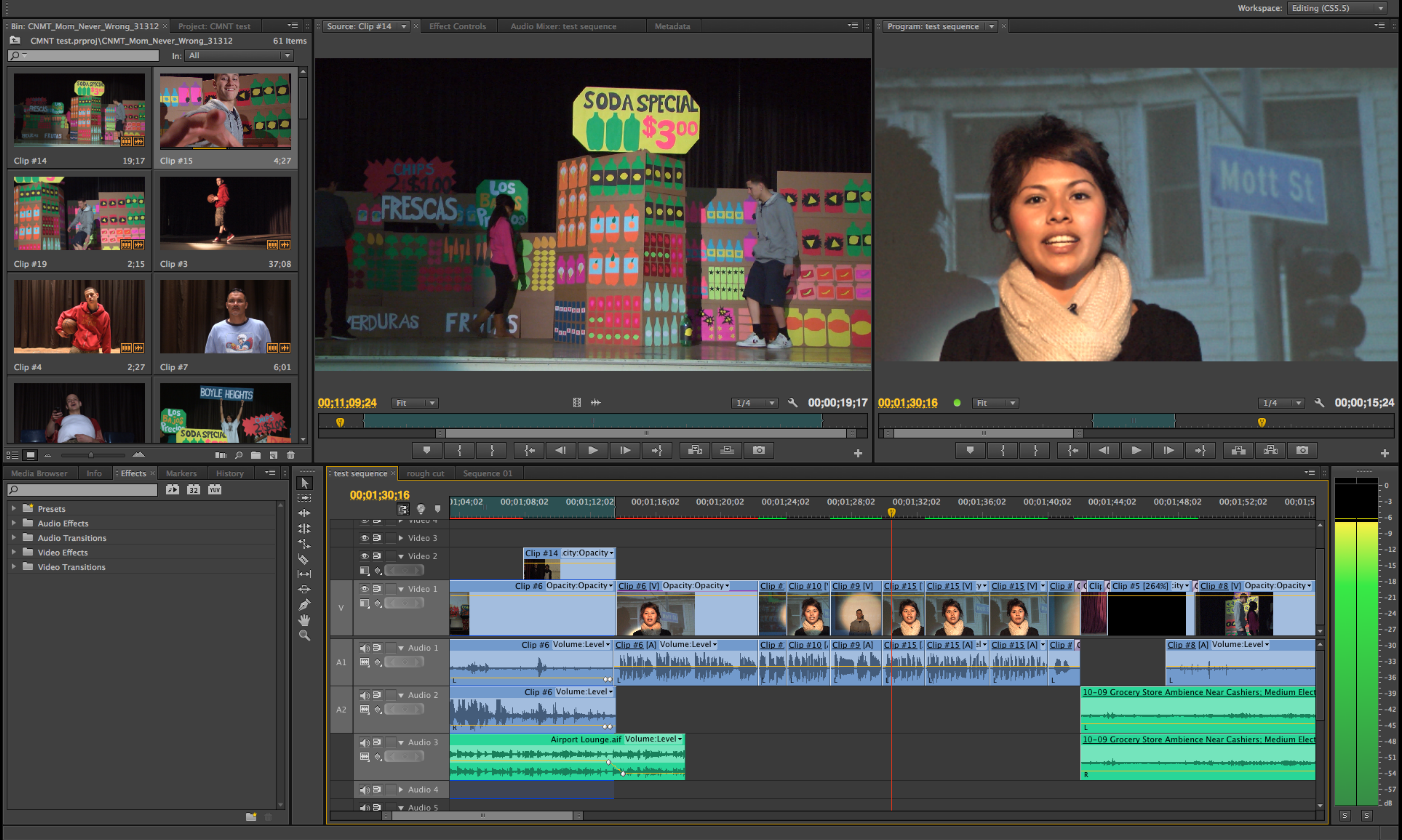
Commercials

…

In these videos we care mostly about the spoken transcript

In these videos we care mostly about the spoken transcript

But we edit them just like any other video…


Workspace: Editing (CS5.5)

Task	Current	Proposed

Task	Current	Proposed
Compose a sentence from multiple takes		

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions
Delete words		

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions
Delete words	Jump cuts	

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions
Delete words	Jump cuts	Seamless transitions

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions
Delete words	Jump cuts	Seamless transitions
Change or add words		

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions
Delete words	Jump cuts	Seamless transitions
Change or add words	New recording session	

Task	Current	Proposed
Compose a sentence from multiple takes	Jump cuts	Seamless transitions
Delete words	Jump cuts	Seamless transitions
Change or add words	New recording session	Synthesize new video

MEAS
PUBL
PERC

MEAS
PUBL
PERC

“The market closed today with apple stock price at
one hundred and ninety one point four five dollars per share”

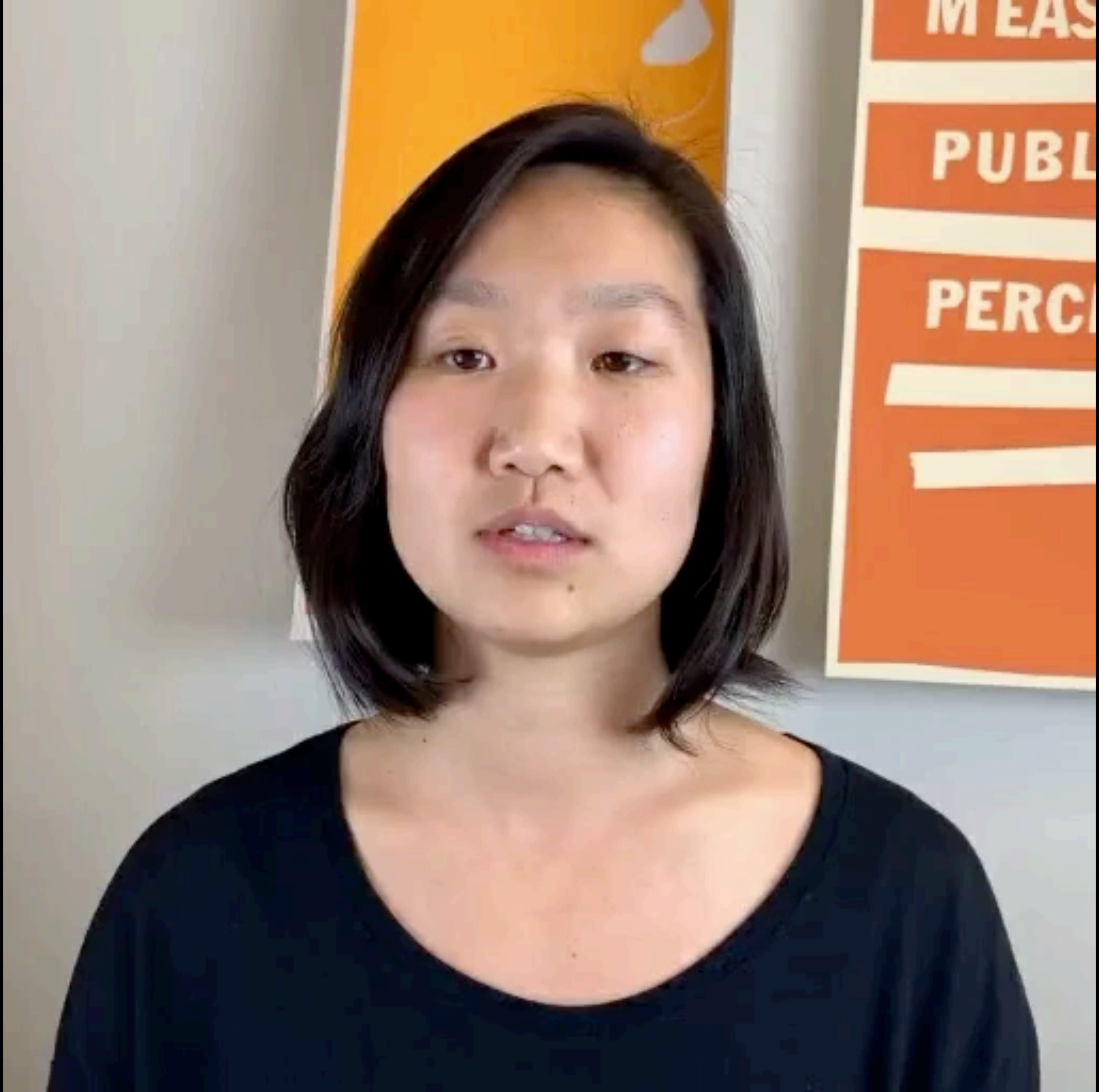
“The market closed today with apple stock price at one hundred and ~~ninety one point four~~ five dollars per share”

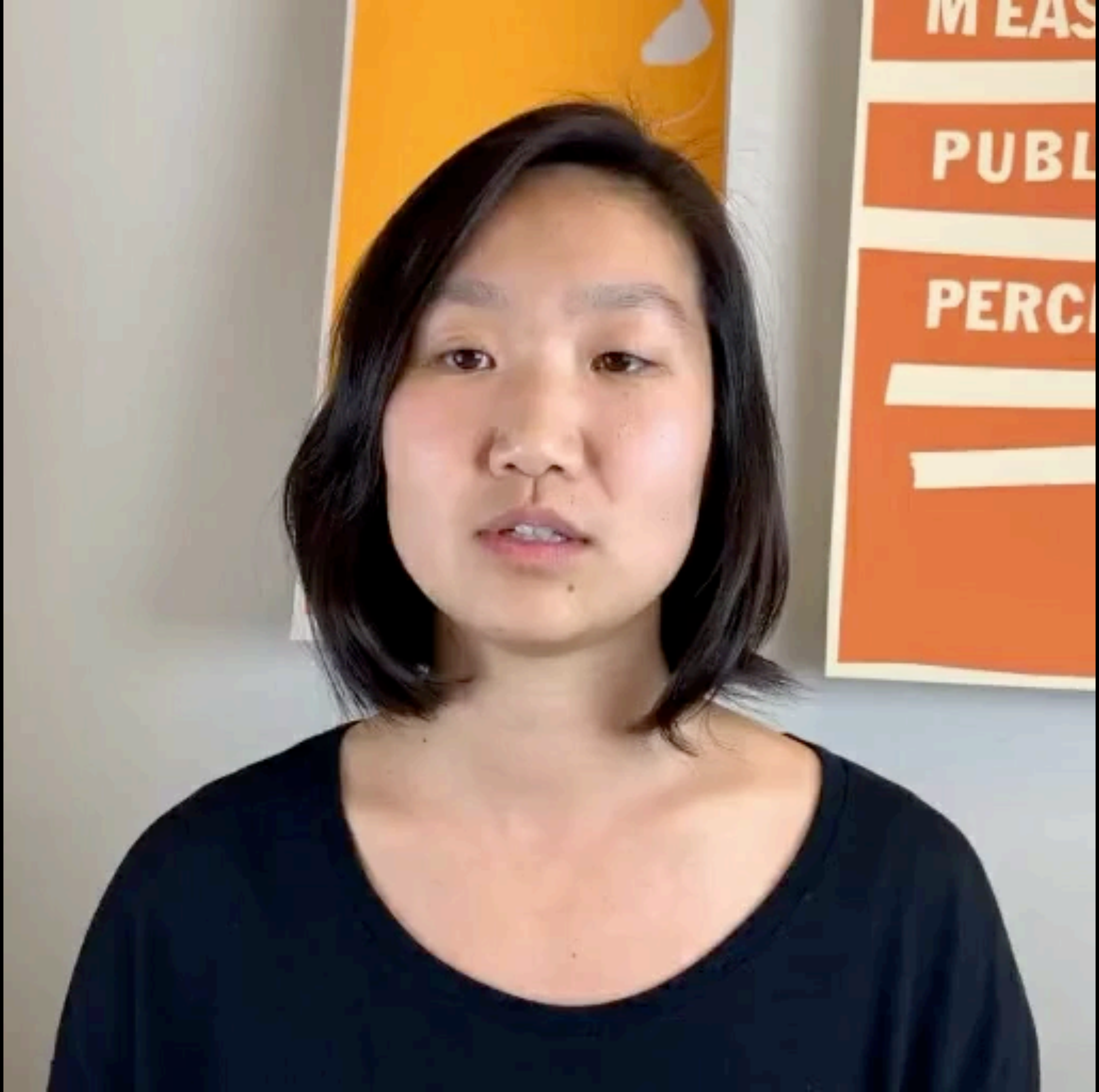
eighty two point two

MEAS
PUBL
PERC

MEAS
PUBL
PERC

Idea: use existing snippets to construct new words


MEAS
PUBL
PERC



MEAS
PUBL
PERC

Reframed as a selective interpolation problem

Reframed as a selective interpolation problem

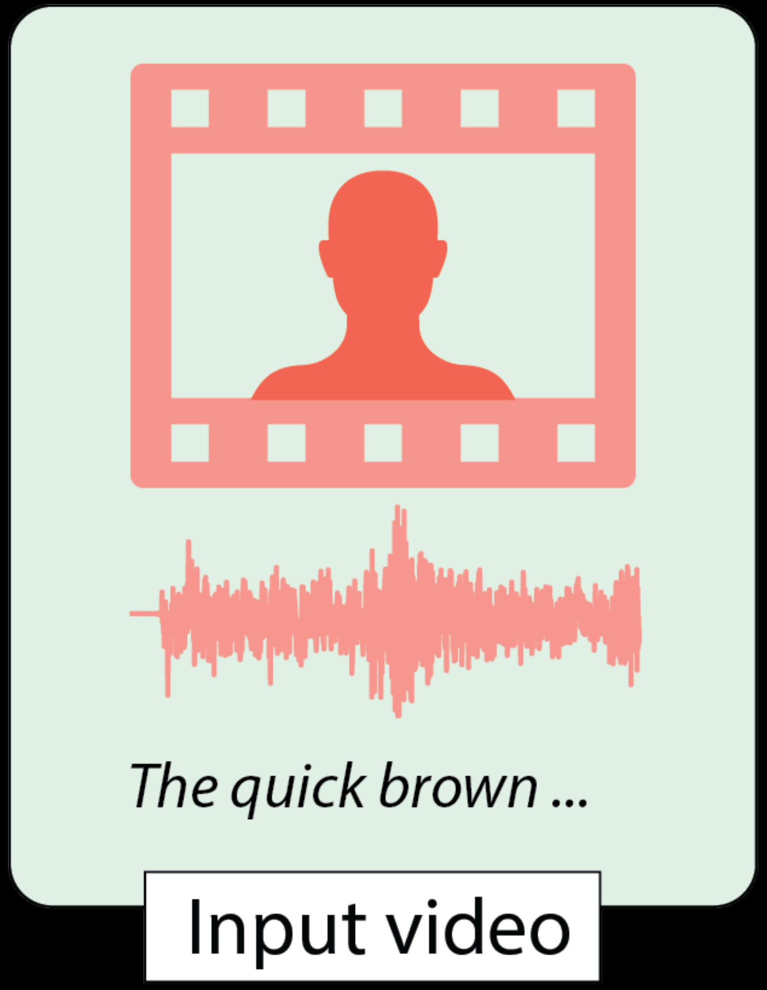
Only use expressions

Only use expressions

Reframed as a selective interpolation problem

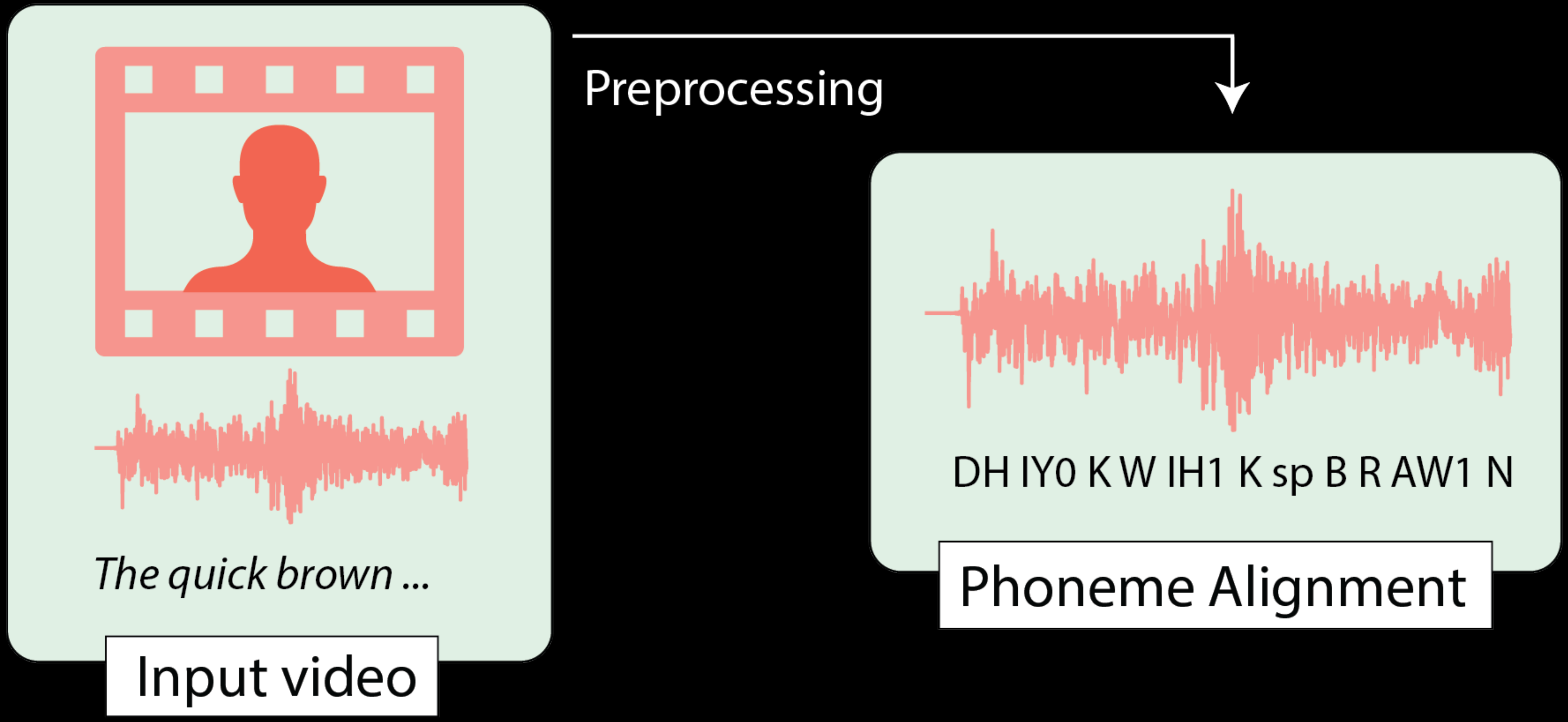
Seamless transitions

Method overview



The quick brown ...

Input video



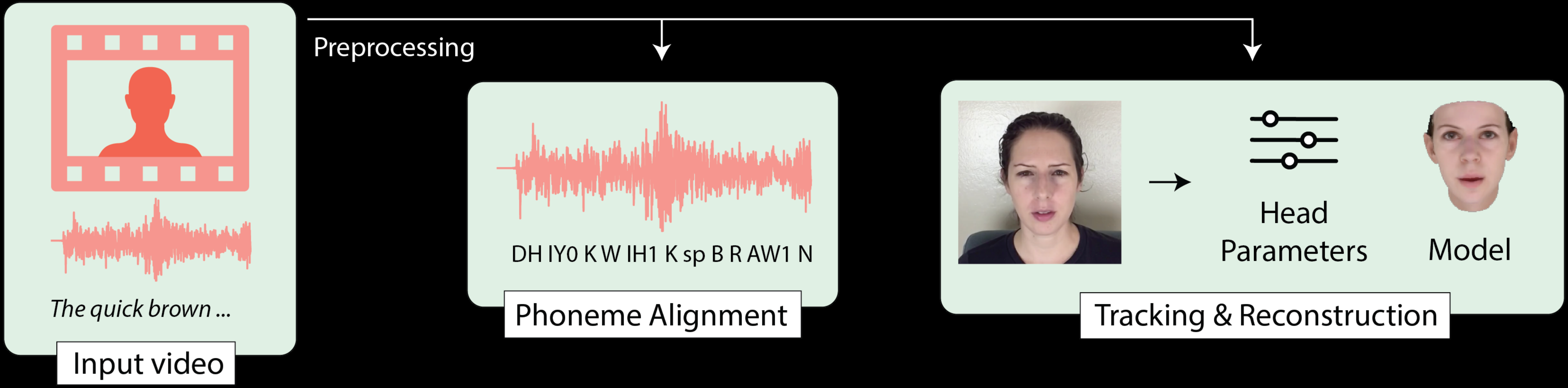
Preprocessing

The quick brown ...

Input video

DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment



Preprocessing

The quick brown ...

Input video

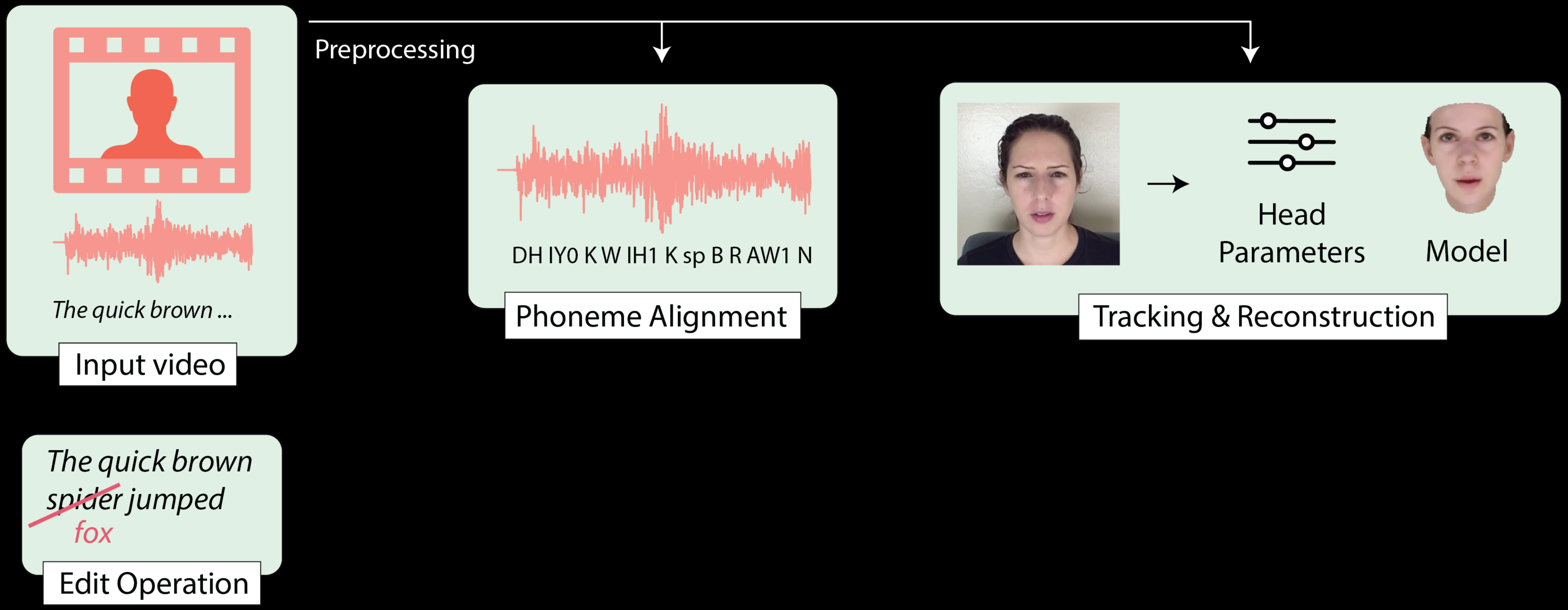
DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment

Head Parameters

Model

Tracking & Reconstruction



Input video

The quick brown ...

Preprocessing

Phoneme Alignment

DH IY0 K W IH1 K sp B R AW1 N

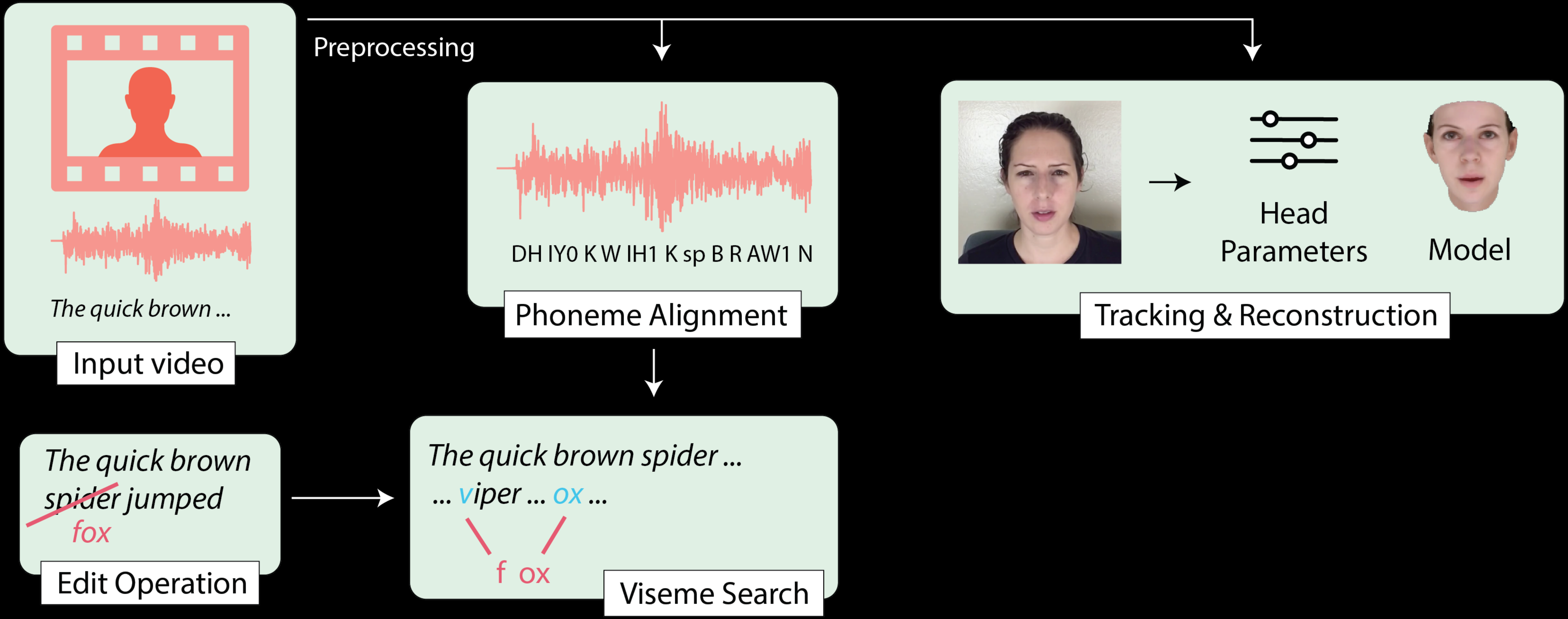
Tracking & Reconstruction

Head Parameters

Model

Edit Operation

The quick brown ~~spider~~ jumped
fox



The quick brown ...

Input video

Preprocessing

DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment

Head Parameters

Model

Tracking & Reconstruction

The quick brown ~~spider~~ jumped

fox

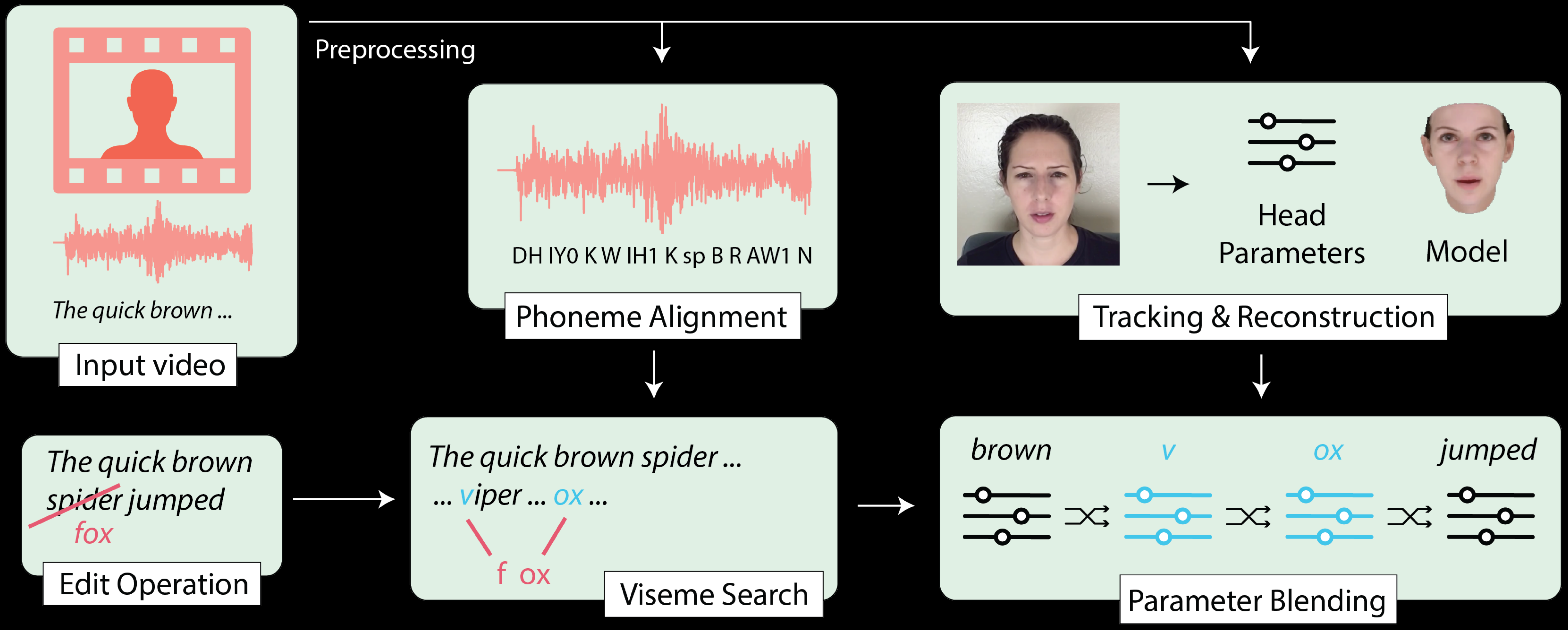
Edit Operation

The quick brown spider ...

... viper ... ox ...

f ox

Viseme Search



Input video

The quick brown ...

Preprocessing

Phoneme Alignment

DH IY0 K W IH1 K sp B R AW1 N

Tracking & Reconstruction

Head Parameters

Model

Edit Operation

The quick brown ~~spider~~ jumped

fox

Viseme Search

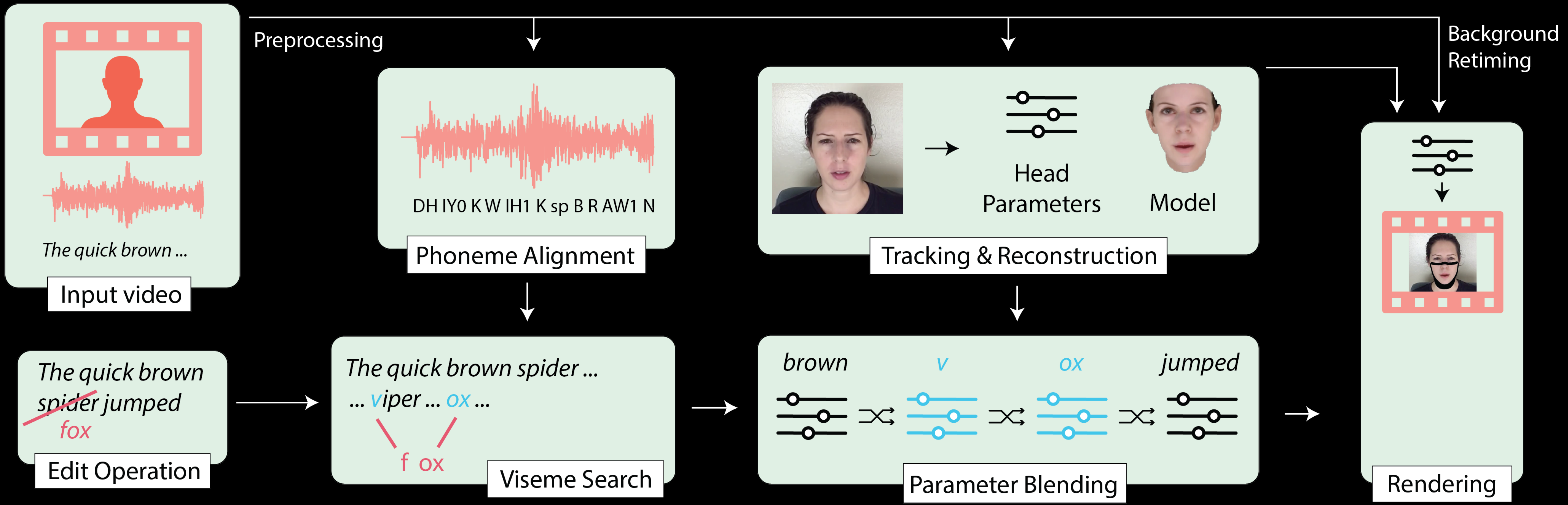
The quick brown spider ...

... viper ... ox ...

f ox

Parameter Blending

brown v ox jumped



The quick brown ...

Input video

Preprocessing

DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment

Head Parameters

Model

Tracking & Reconstruction

Background Retiming

The quick brown ~~spider~~ jumped
fox

Edit Operation

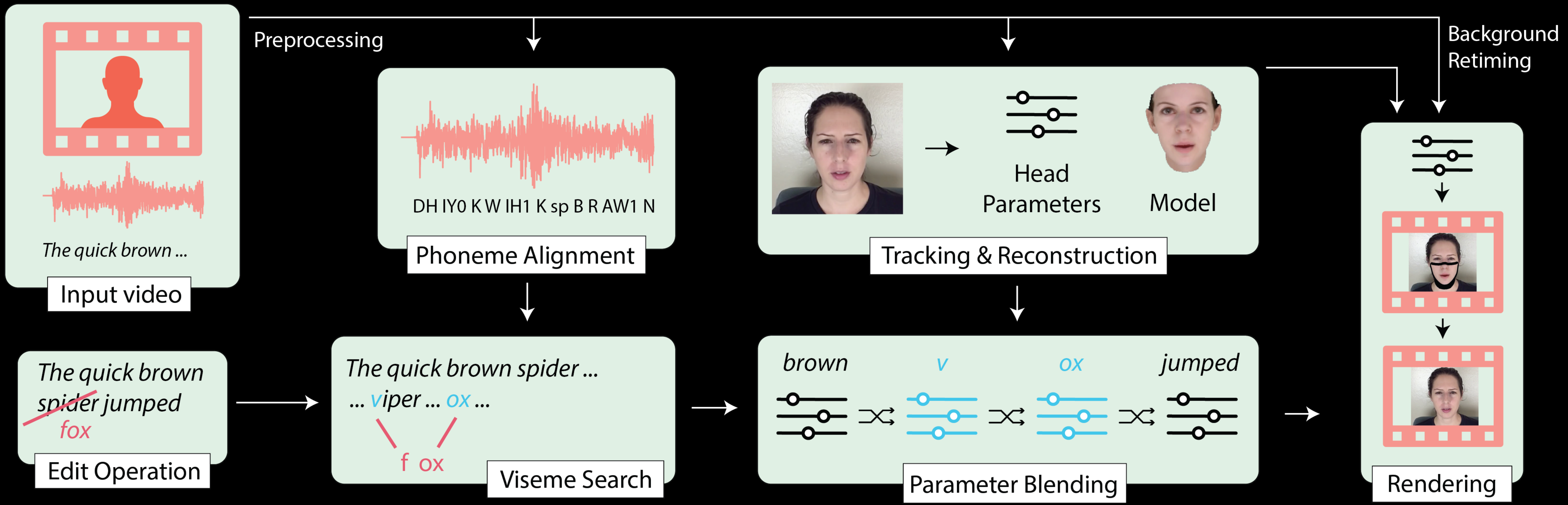
The quick brown spider ...
... viper ... ox ...
f ox

Viseme Search

brown v ox jumped

Parameter Blending

Rendering



The quick brown ...

Input video

Preprocessing

DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment

Head Parameters

Model

Tracking & Reconstruction

Background Retiming

The quick brown ~~spider~~ jumped
fox

Edit Operation

The quick brown spider ...
... viper ... ox ...
f ox

Viseme Search

brown *v* *ox* *jumped*

Parameter Blending

Rendering

Preprocessing

Background Retiming

The quick brown ...

Input video

DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment

Head Parameters

Model

Tracking & Reconstruction

The quick brown ~~spider~~ jumped

fox

Edit Operation

The quick brown spider ...

... viper ... ox ...

f ox

Viseme Search

brown

v

ox

jumped

Parameter Blending

Rendering

New stuff

Viseme search

Viseme search

- Identical phonemes are likely to be visually similar

Viseme search

- Identical phonemes are likely to be visually similar
- Same for visemes (but less so)

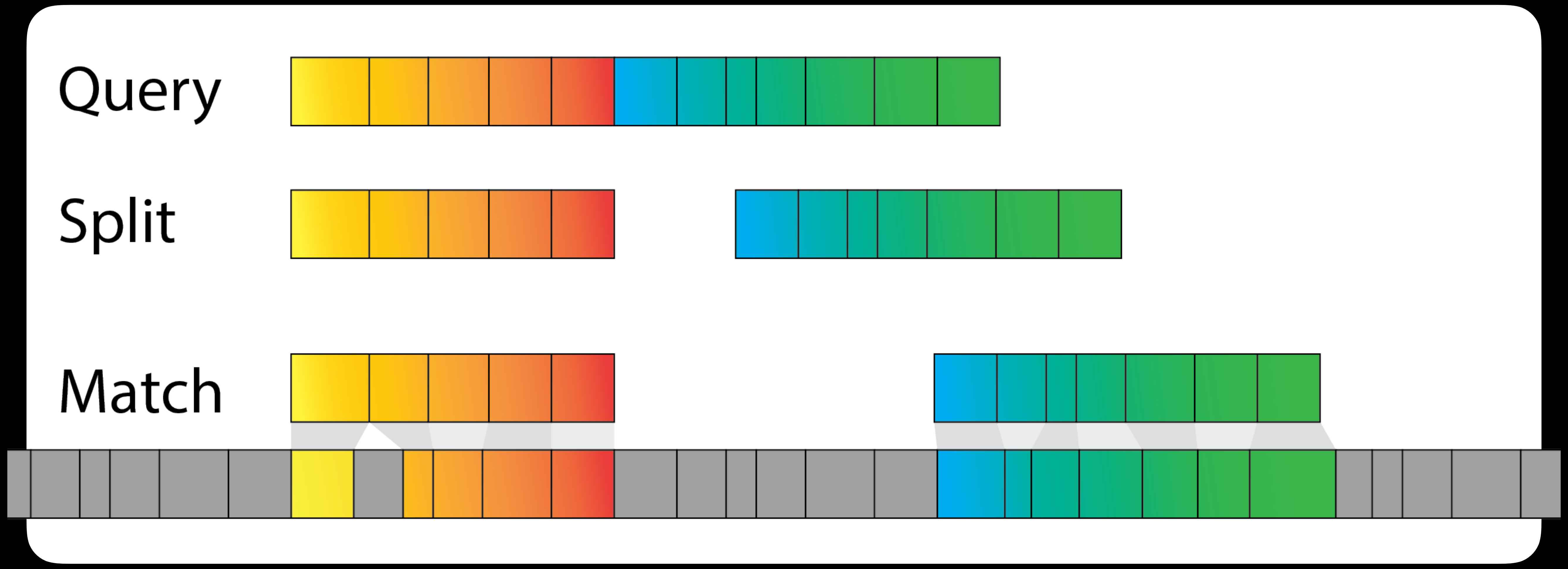
Viseme search

- Identical phonemes are likely to be visually similar
- Same for visemes (but less so)
- **Cannot** expect to find a good coherent viseme sequence for long edits

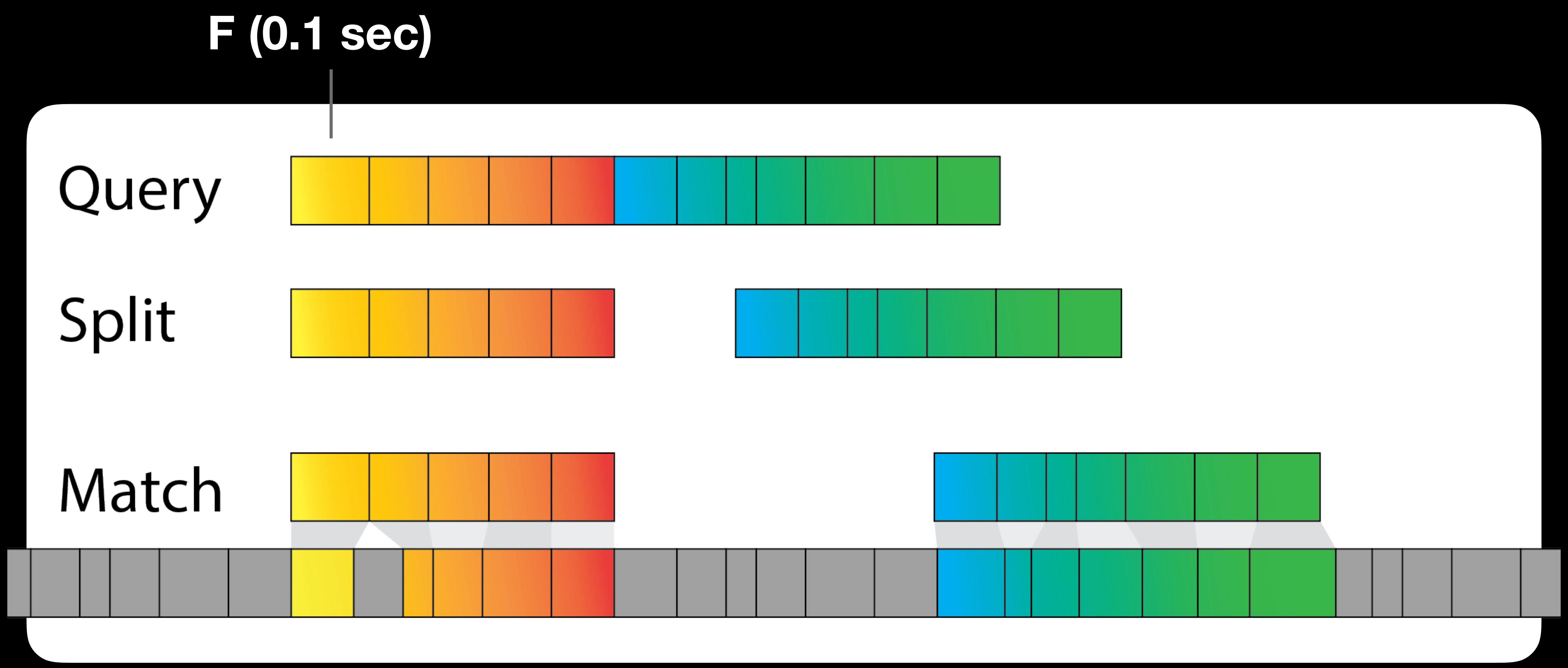
Viseme search

- Identical phonemes are likely to be visually similar
- Same for visemes (but less so)
- **Cannot** expect to find a good coherent viseme sequence for long edits
    - Instead, find several matching subsequences and combine

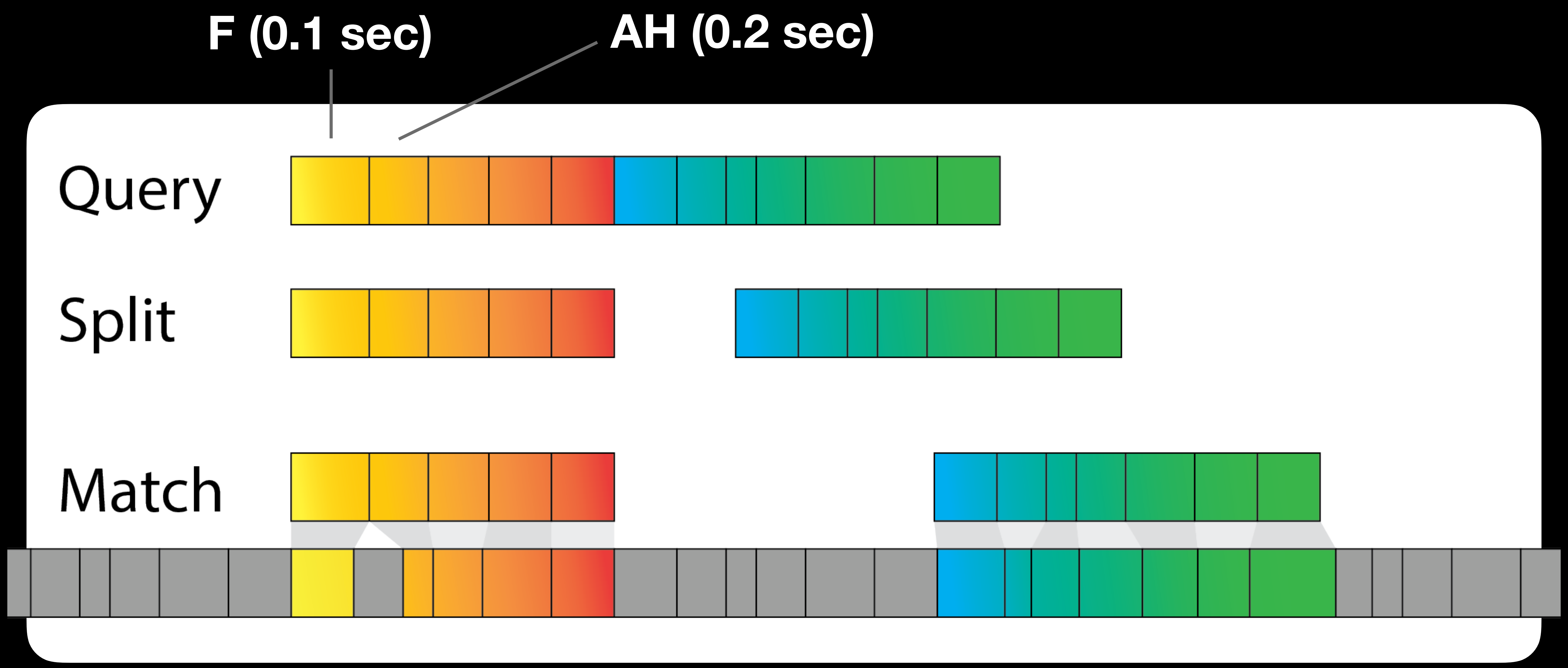
Viseme search



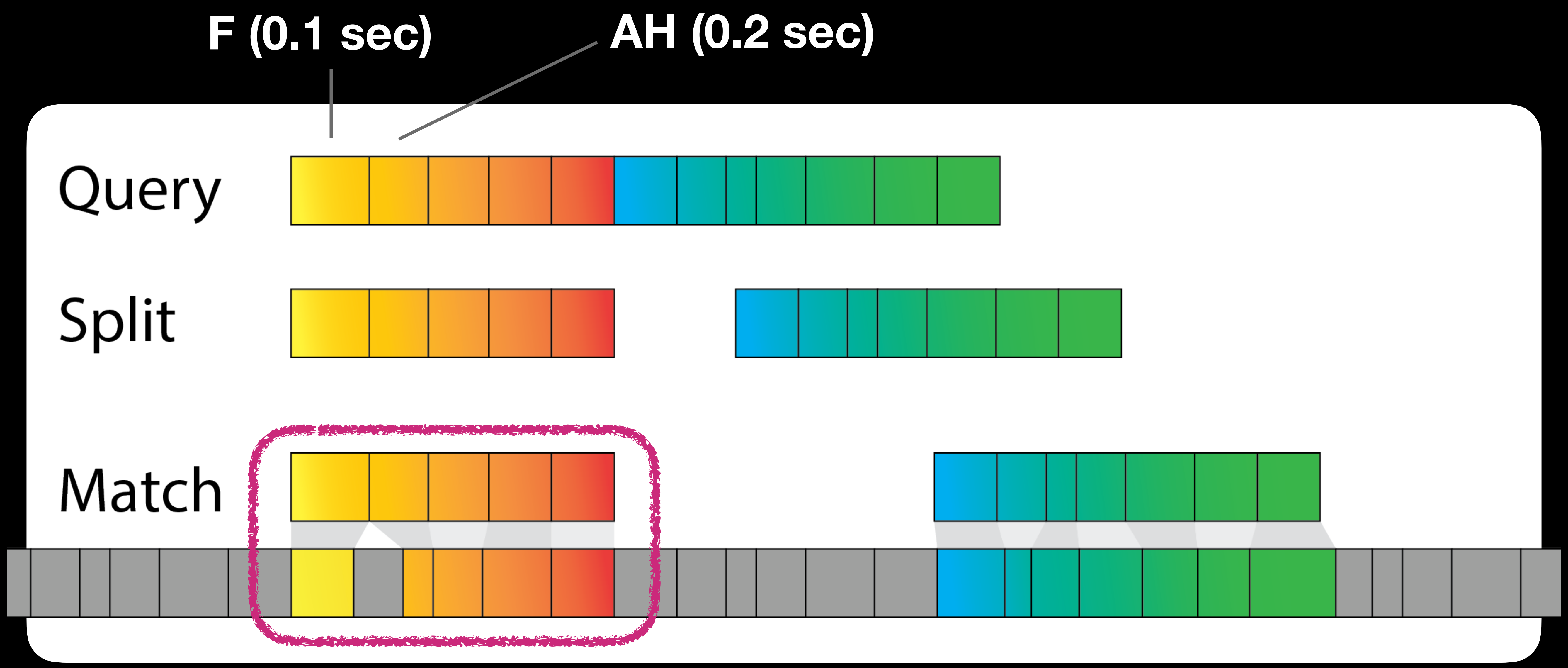
Viseme search



Viseme search



Viseme search



Viseme search

Viseme search

Modified Levenshtein edit distance

Viseme search

Modified Levenshtein edit distance

$$C_{insert} = 1$$

$$C_{delete} = 1$$

Viseme search

Modified Levenshtein edit distance

$$C_{insert} = 1$$

$$C_{delete} = 1$$

$$C_{swap} =$$

Viseme search

Modified Levenshtein edit distance

$$C_{insert} = 1$$

$$C_{delete} = 1$$

$$C_{swap} = C_{vis}(\text{🟨}, \text{🟧})$$

Viseme search

Modified Levenshtein edit distance

$$C_{insert} = 1$$

$$C_{delete} = 1$$

$$C_{swap} = C_{vis}(🟨, 🟧)(|🟨| + |🟧|)$$

Viseme search

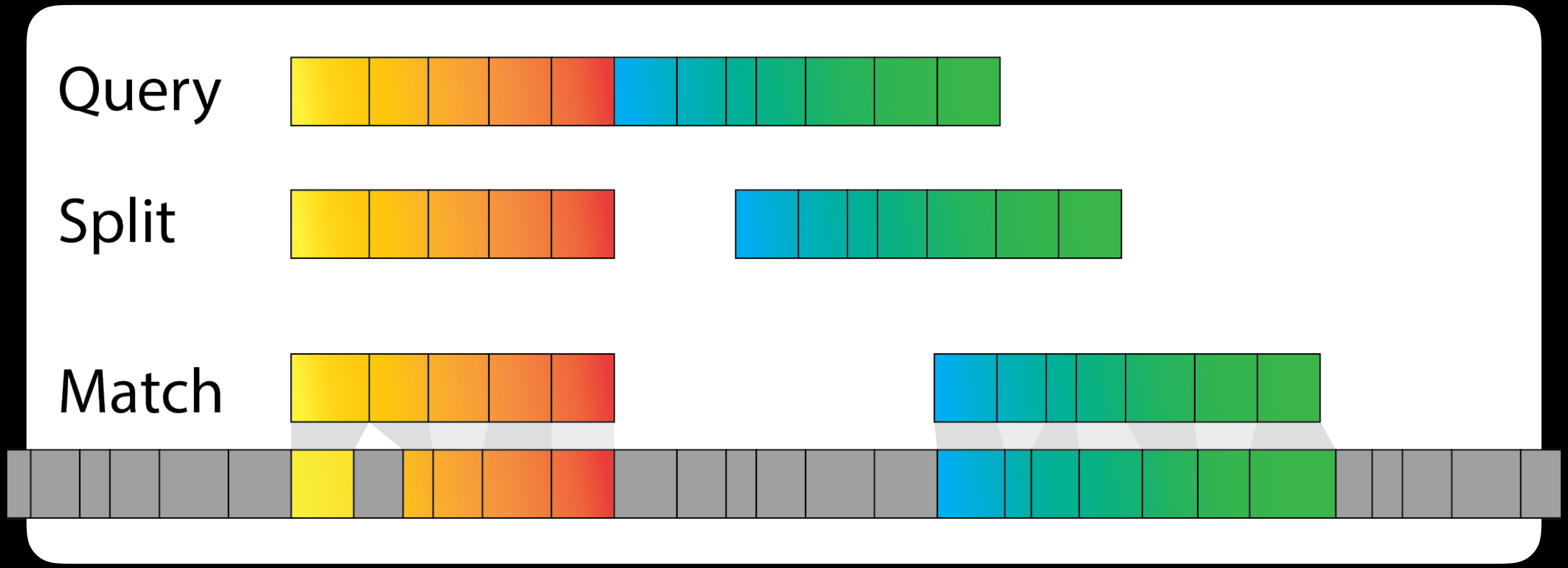
Modified Levenshtein edit distance

$$C_{insert} = 1$$

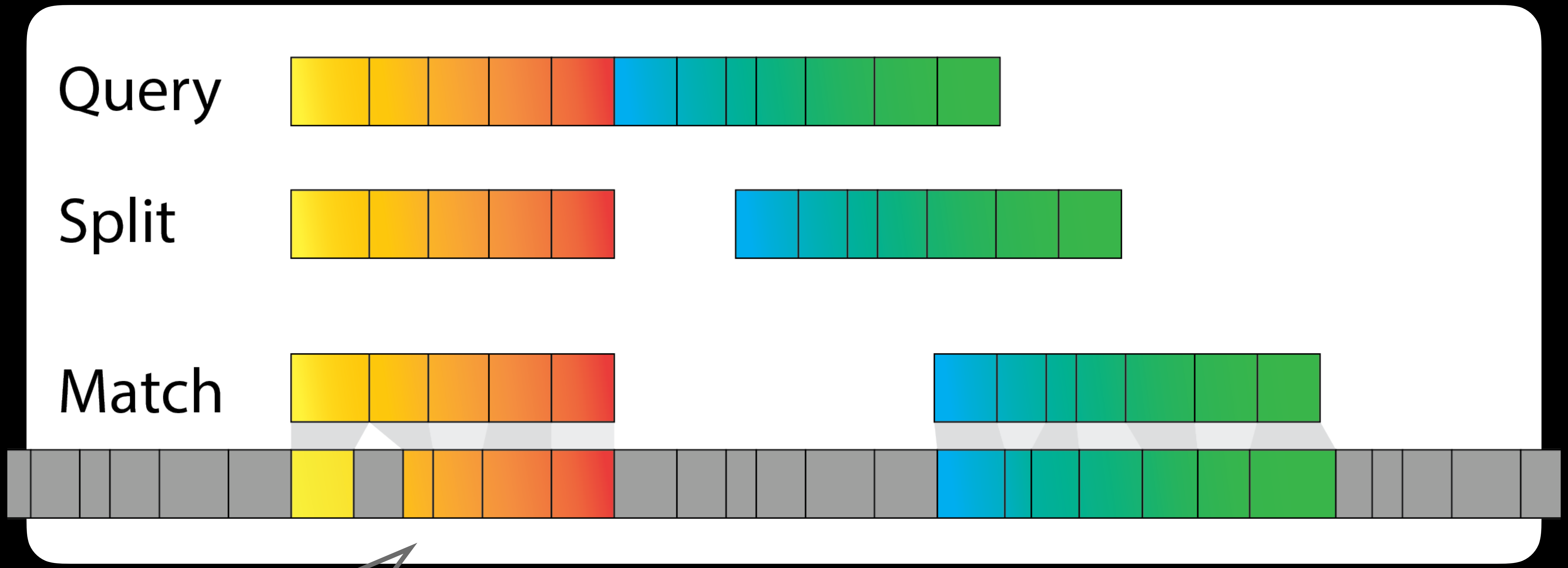
$$C_{delete} = 1$$

$$C_{swap} = C_{vis}(\text{🟨}, \text{🟧})(|\text{🟨}| + |\text{🟧}|) + \chi \Big|\, |\text{🟨}| - |\text{🟧}| \,\Big|$$

Viseme search

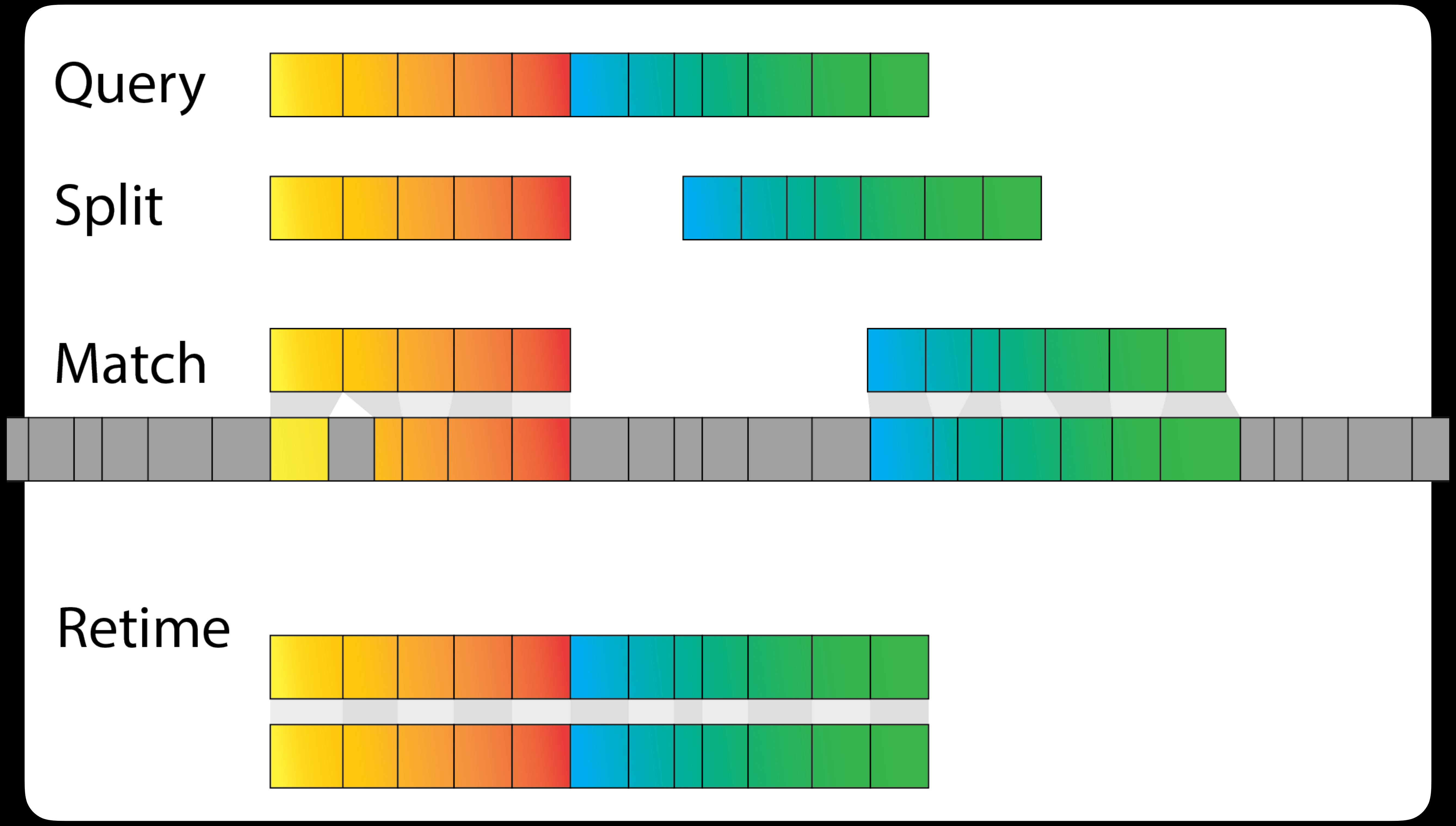


Viseme search



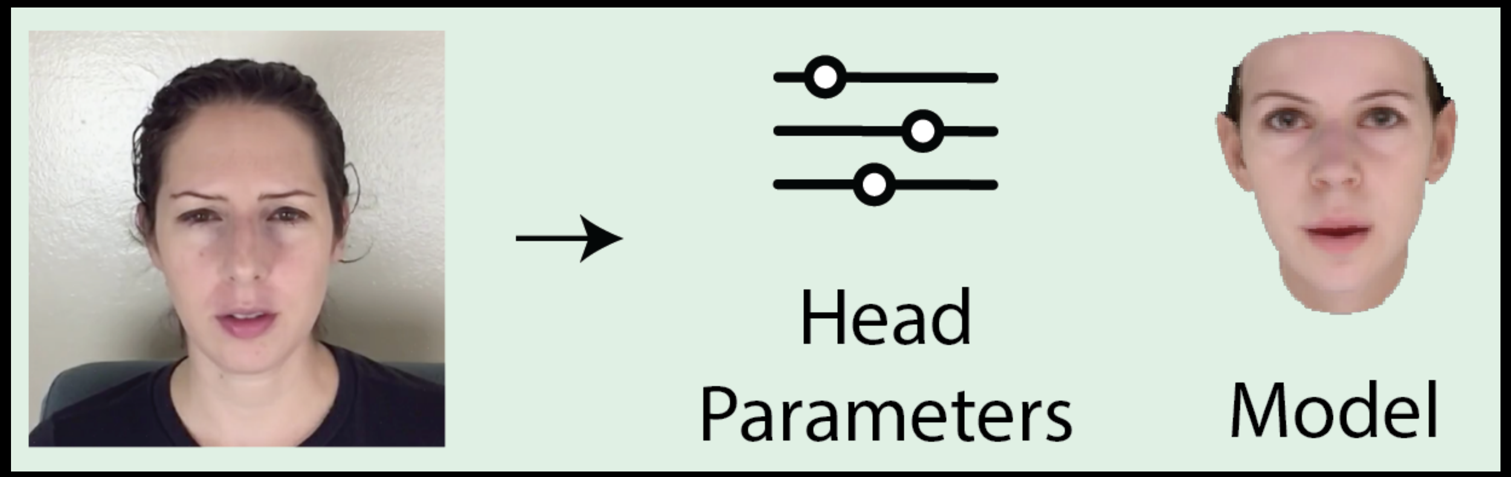
Match cost +
short length penalty

Viseme search



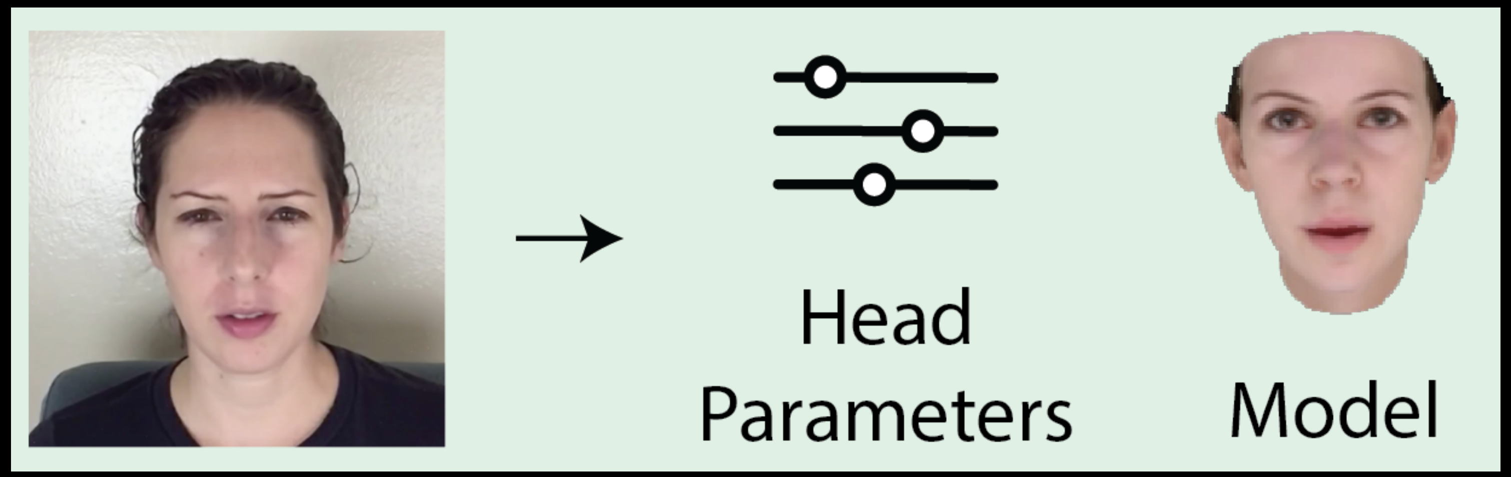
Parameter blending

Parameter blending

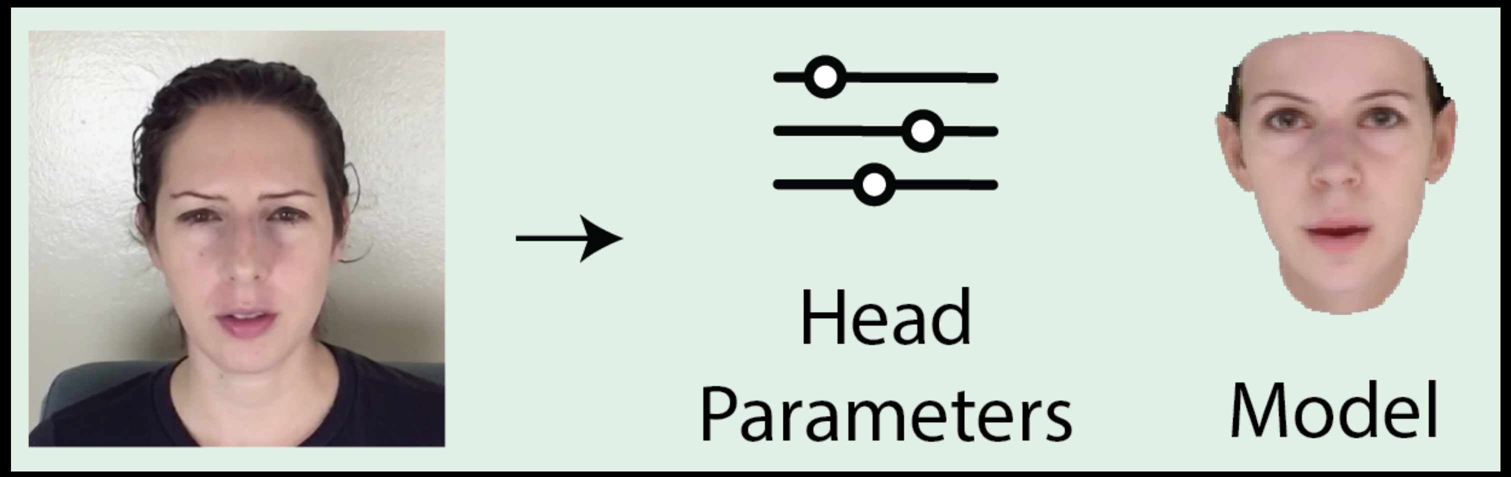


Parameter blending

- Geometry

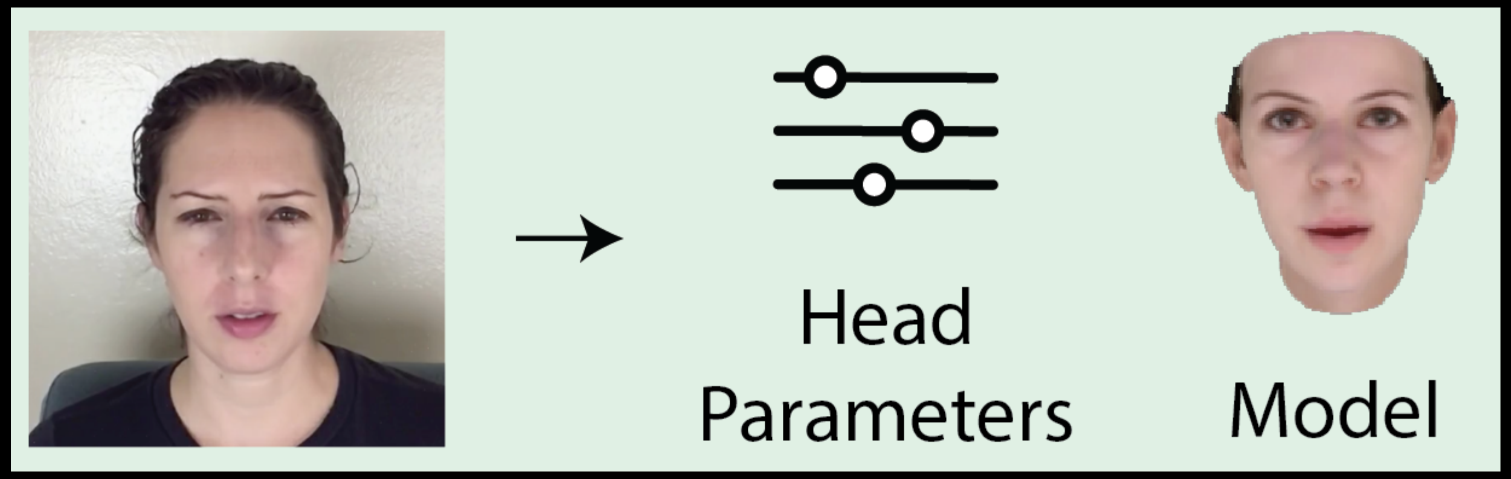


Parameter blending



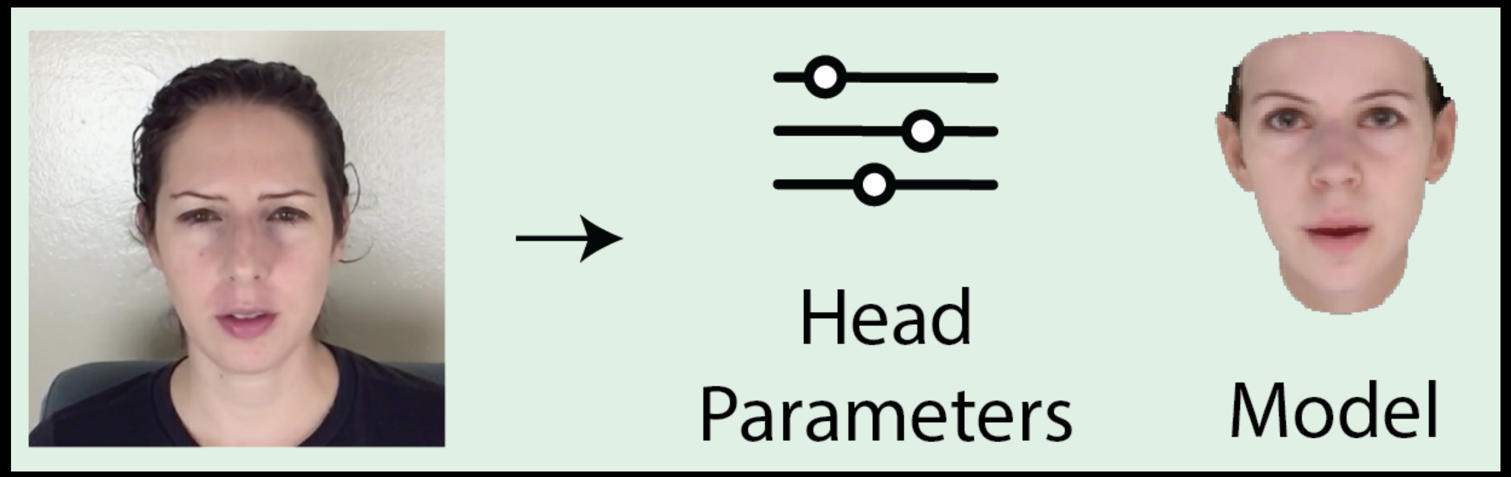
- Geometry
- Albedo

Parameter blending



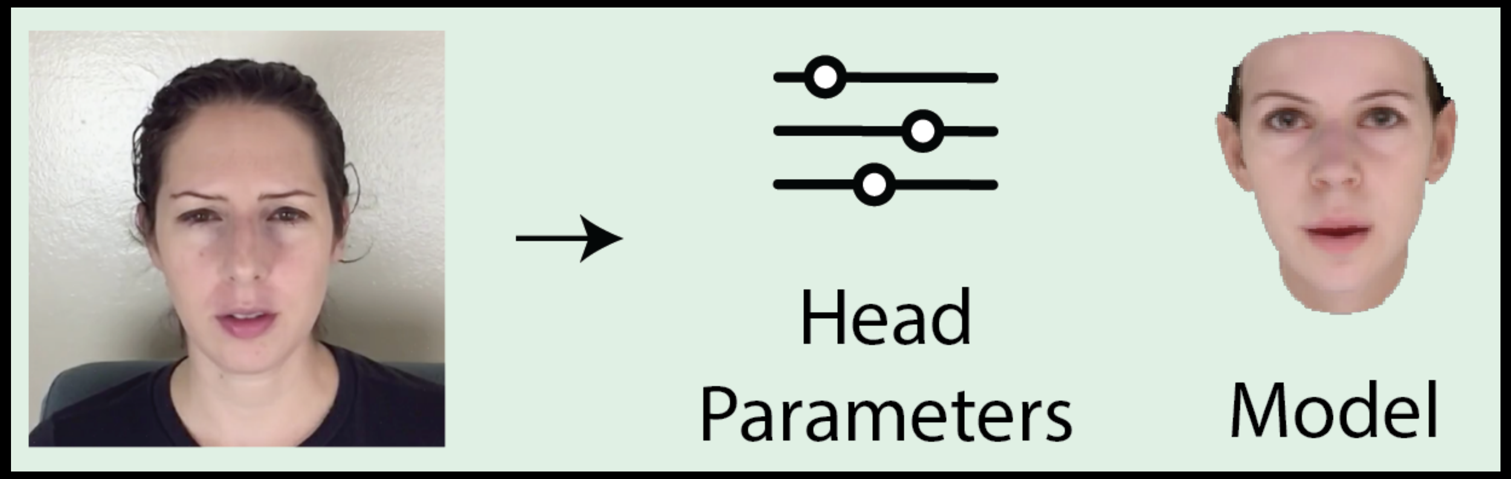
- Geometry
- Albedo
- Illumination

Parameter blending



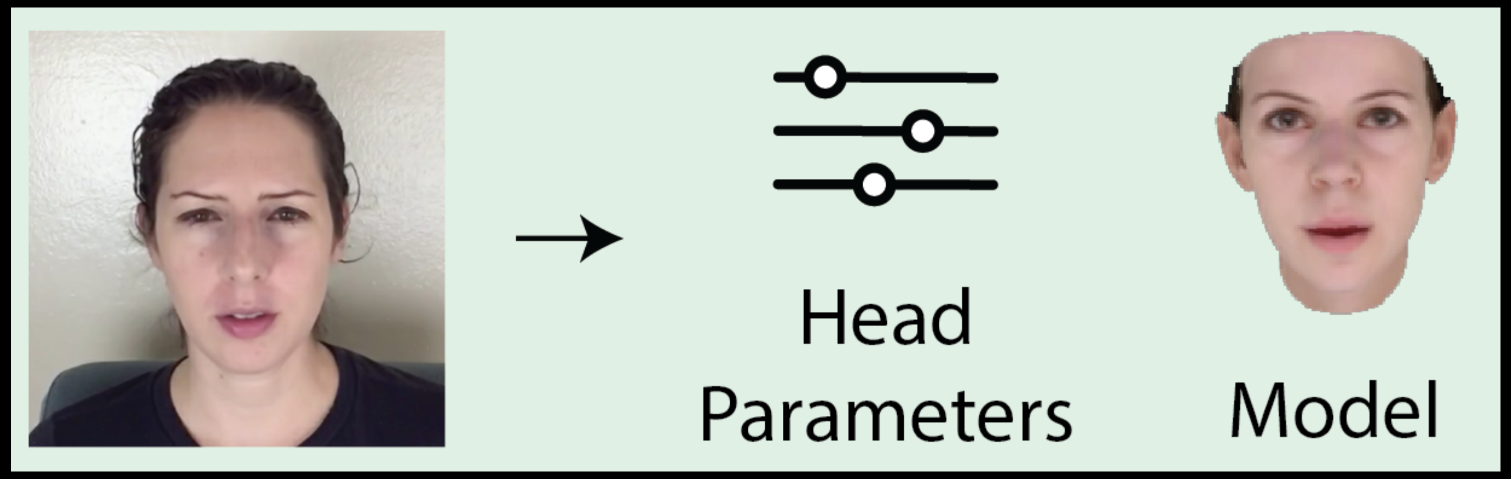
- Geometry
- Albedo
- Illumination
- Pose

Parameter blending



- Geometry
- Albedo
- Illumination
- Pose
- Expression

Parameter blending

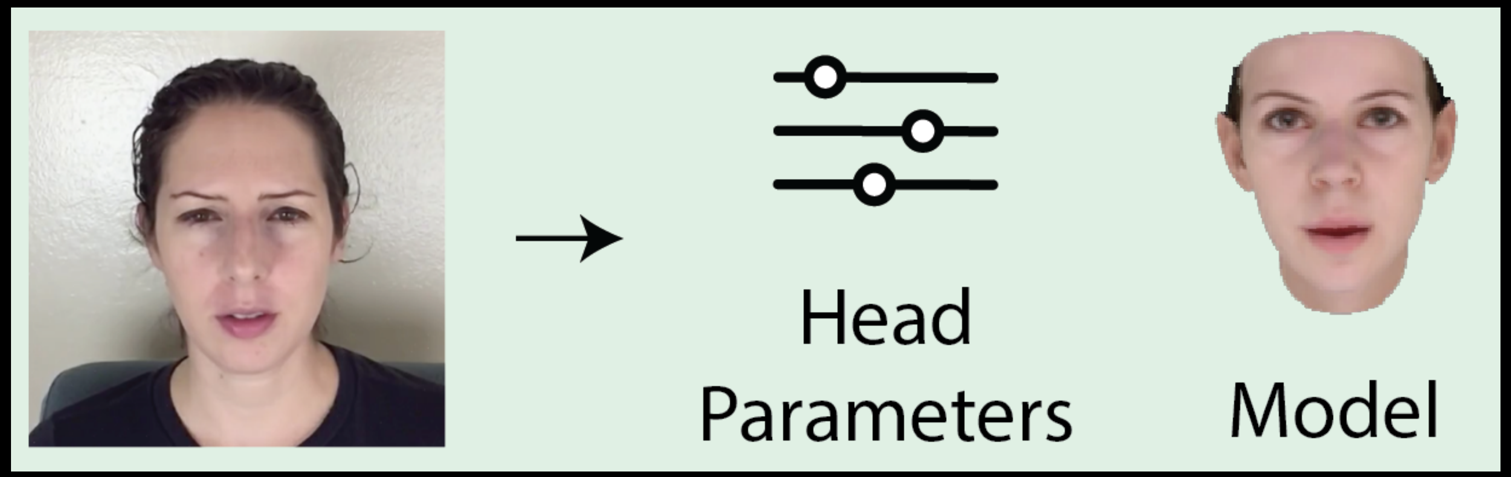


- Geometry
- Albedo

Constant

- Illumination
- Pose
- Expression

Parameter blending



- Geometry
- Albedo

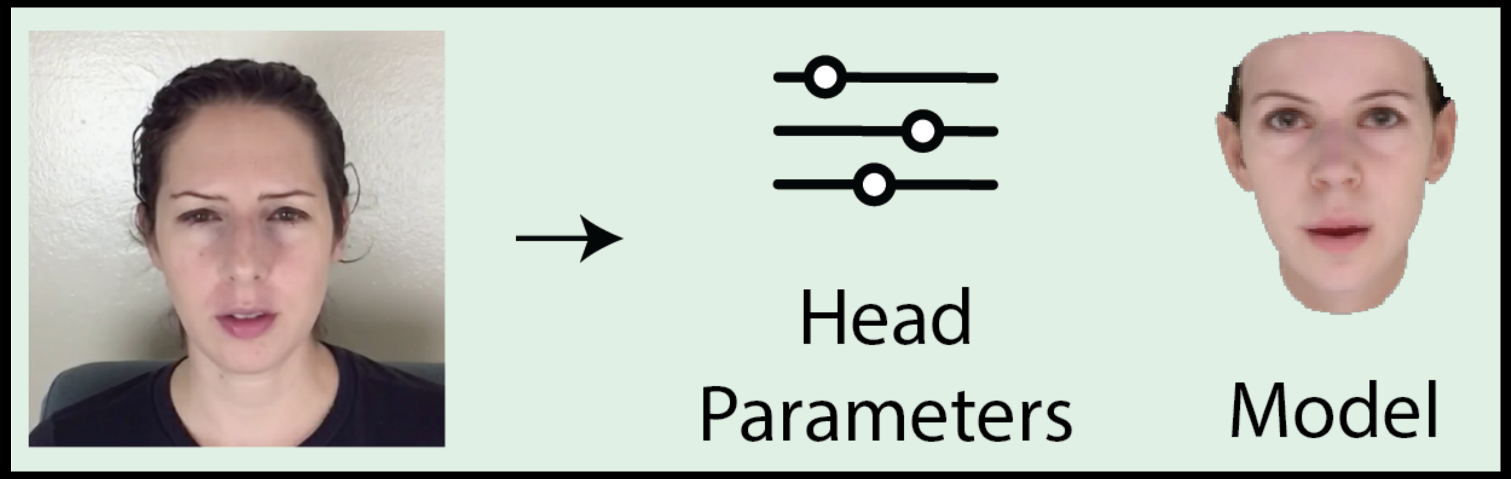
Constant

- Illumination

Linear interpolation in new region

- Pose
- Expression

Parameter blending



- Geometry
- Albedo

Constant

- Illumination

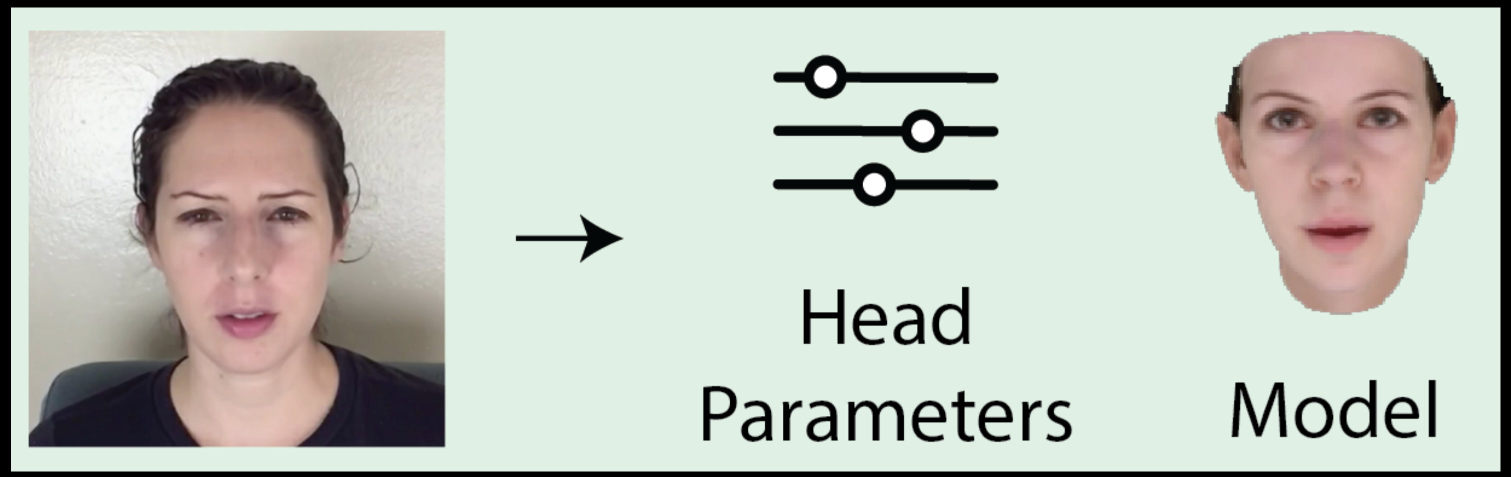
Linear interpolation in new region

- Pose

Later…

- Expression

Parameter blending



- Geometry
- Albedo

Constant

- Illumination

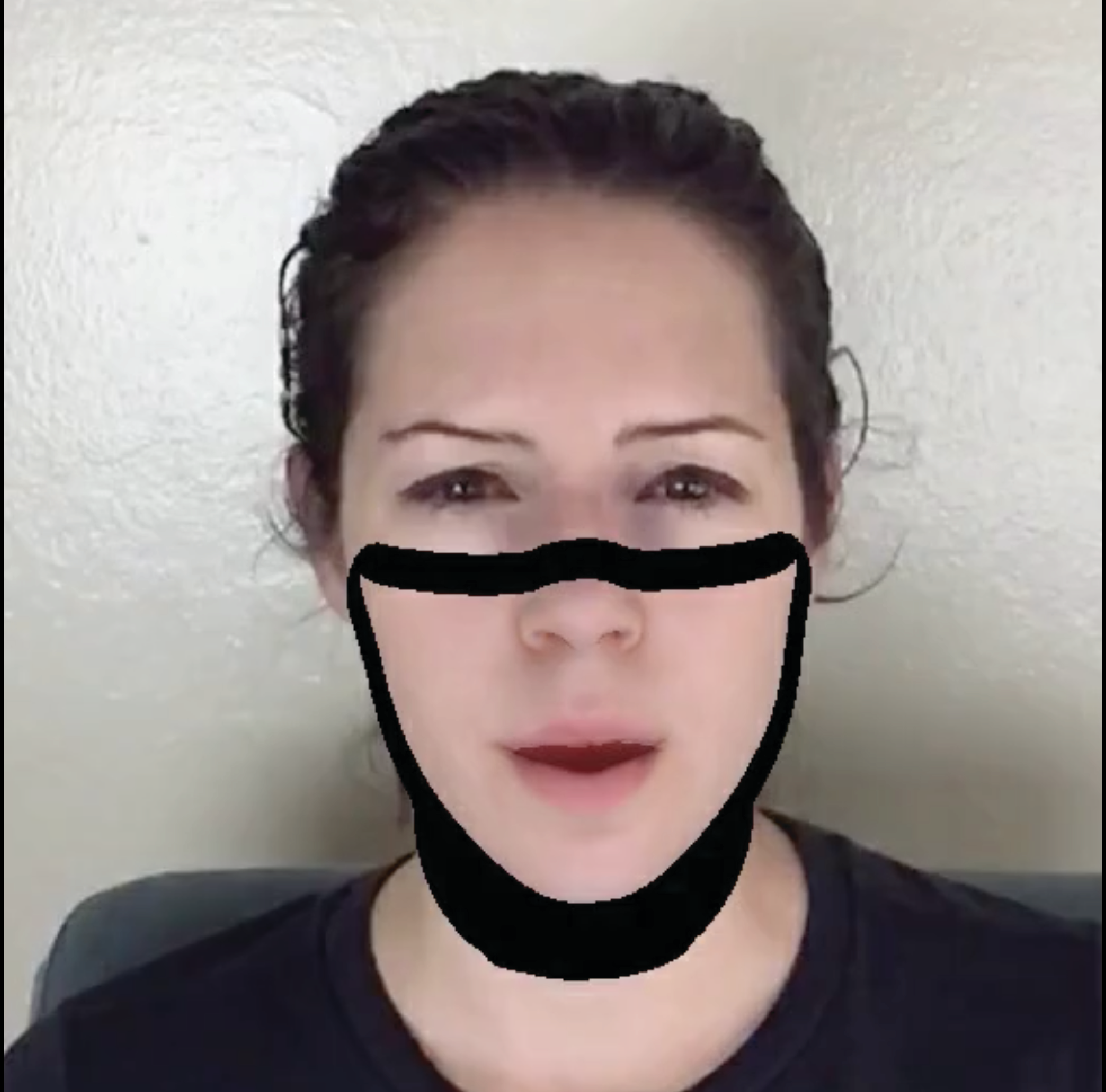
Linear interpolation in new region

- Pose

Later…

- Expression

Linear interpolation between snippets



Background retiming

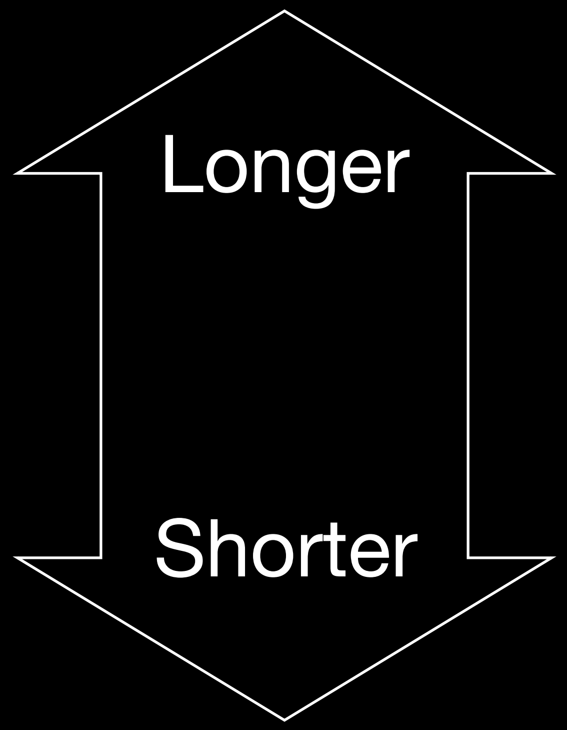
Background retiming

… The quick brown spider jumped …

… The quick brown fox jumped …

Background retiming

… The quick brown spider jumped …



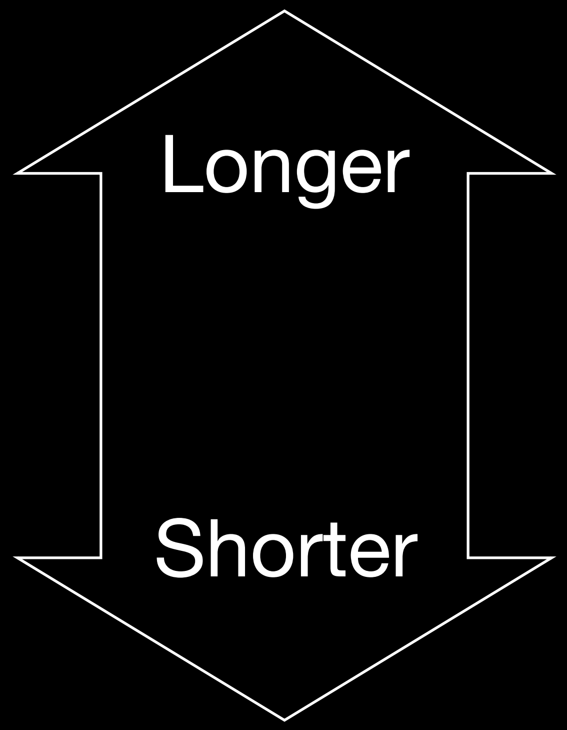
Longer

Shorter

… The quick brown fox jumped …

Background retiming

… The quick brown spider jumped …



Longer

Shorter

… The quick brown fox jumped …

We want localized edits. Everything else should stay the same

Background retiming

Background retiming

- Use longer sequence (even if edit is short)

Background retiming

- Use longer sequence (even if edit is short)
- Calculate number of frames to add / remove

Background retiming

- Use longer sequence (even if edit is short)
- Calculate number of frames to add / remove
  - Spread equally

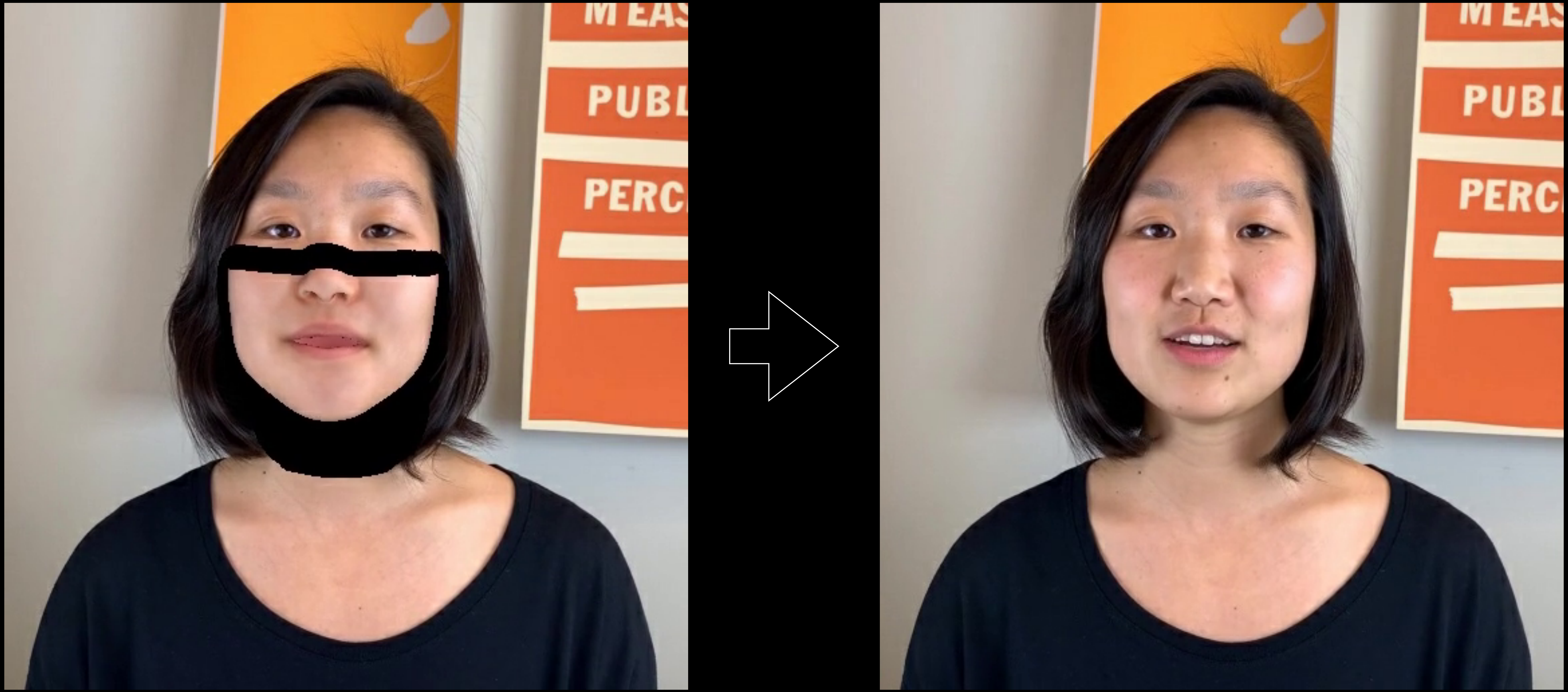
Background retiming

- Use longer sequence (even if edit is short)
- Calculate number of frames to add / remove
    - Spread equally
- Long enough —> no retiming artifacts

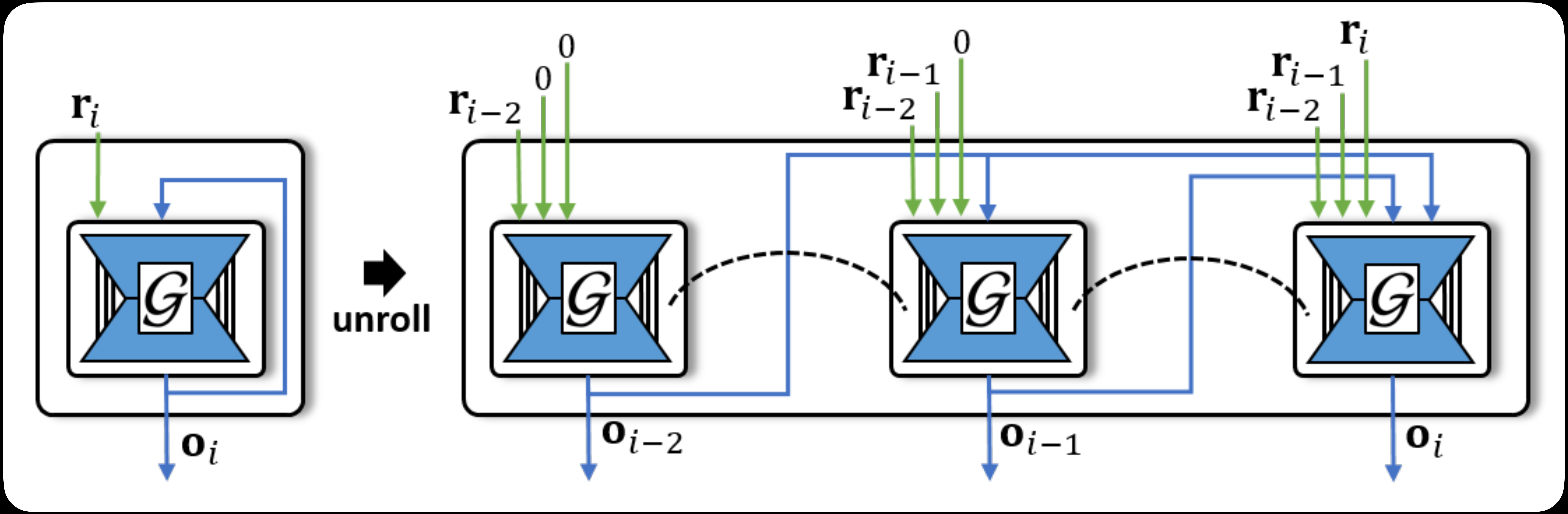
Background retiming

- Use longer sequence (even if edit is short)
- Calculate number of frames to add / remove
    - Spread equally
- Long enough —> no retiming artifacts
- Pose parameters taken from retimed background

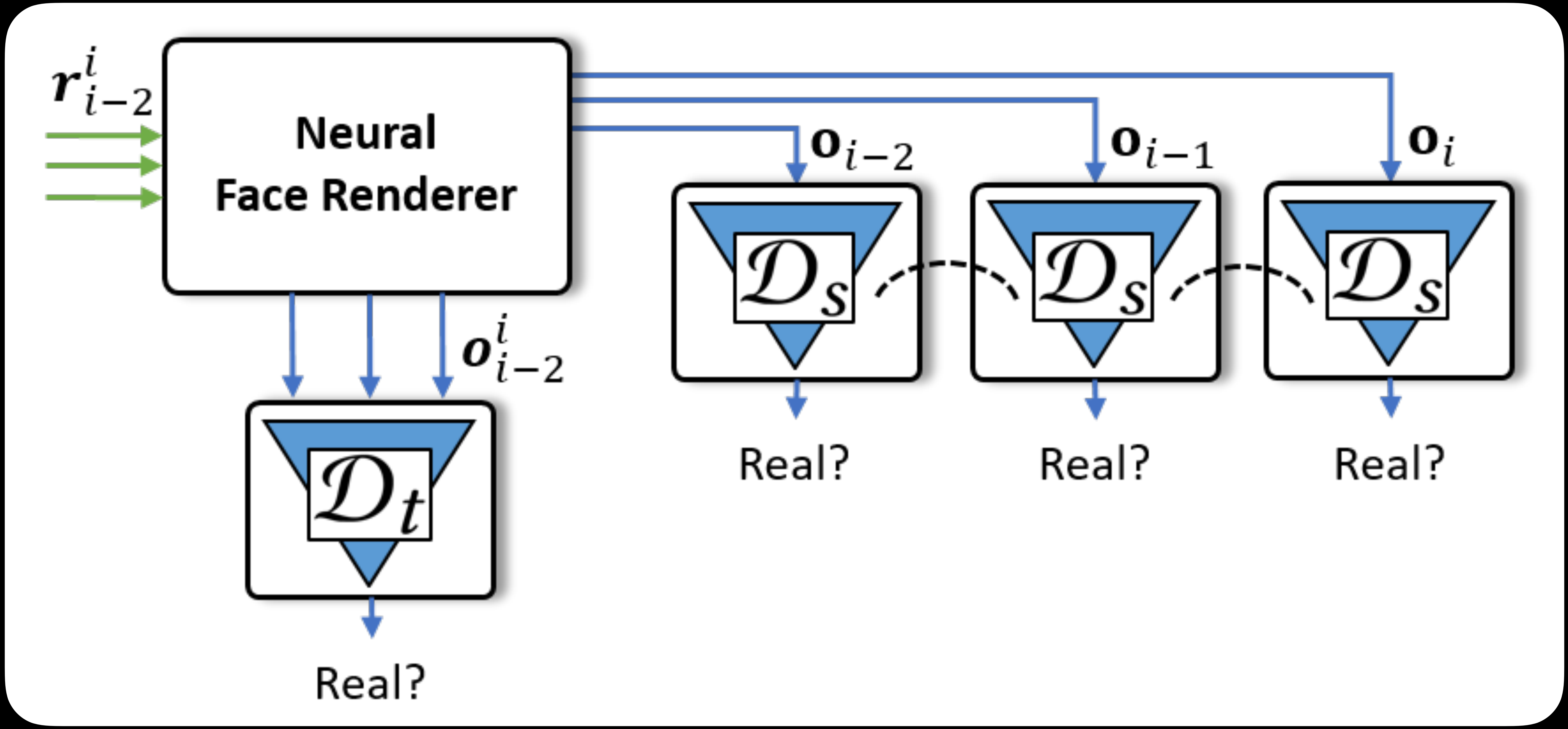
Neural Face Rendering



Neural Face Rendering

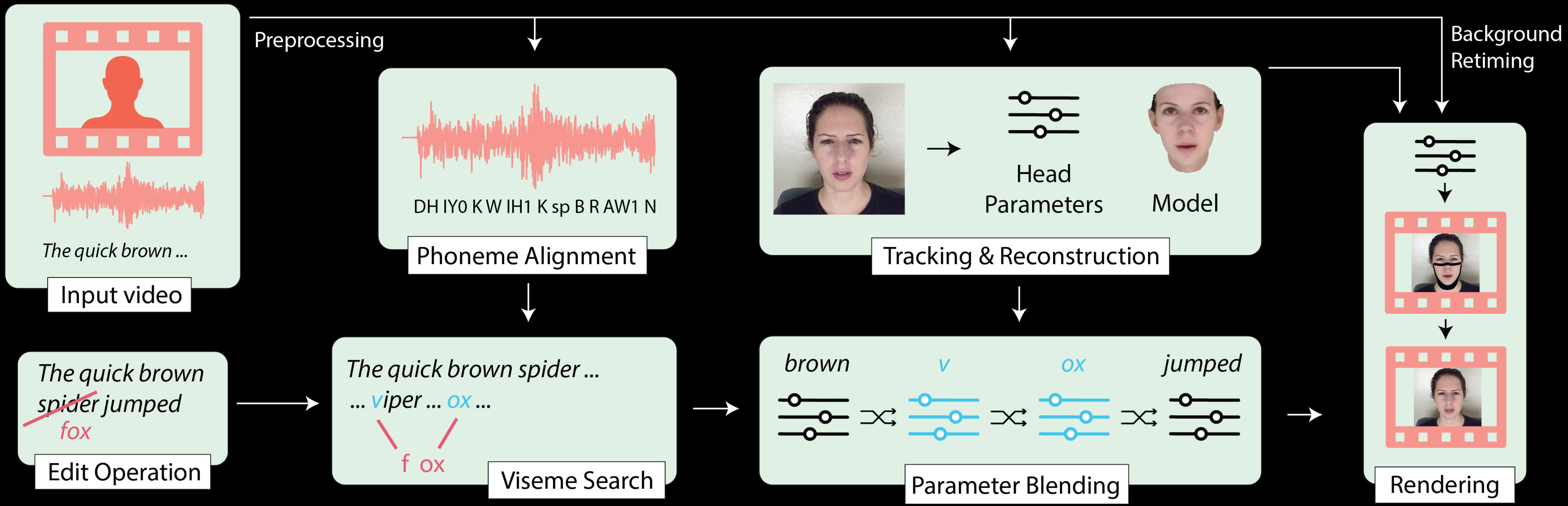


Neural Face Rendering



Neural Face Rendering

Loss = Photometric + Spatial adversarial + Temporal adversarial



The quick brown ...

Input video

Preprocessing

DH IY0 K W IH1 K sp B R AW1 N

Phoneme Alignment

Head Parameters

Model

Tracking & Reconstruction

Background Retiming

The quick brown ~~spider~~ jumped
fox

Edit Operation

The quick brown spider ...
... viper ... ox ...
f ox

Viseme Search

brown v ox jumped

Parameter Blending

Rendering

Results

Input

Input

Edit

Edit

MERA
VA
ON

Input

Input

Edit

Edit

Input

Input

Edit

Edit

Input

Input

Edit

Edit

Visual voice assistant

Visual voice assistant

MERA
VA
ON

Translation

Translation

Blend with input (temporal)

Blend with input
(temporal)

Blend with input (spatial)

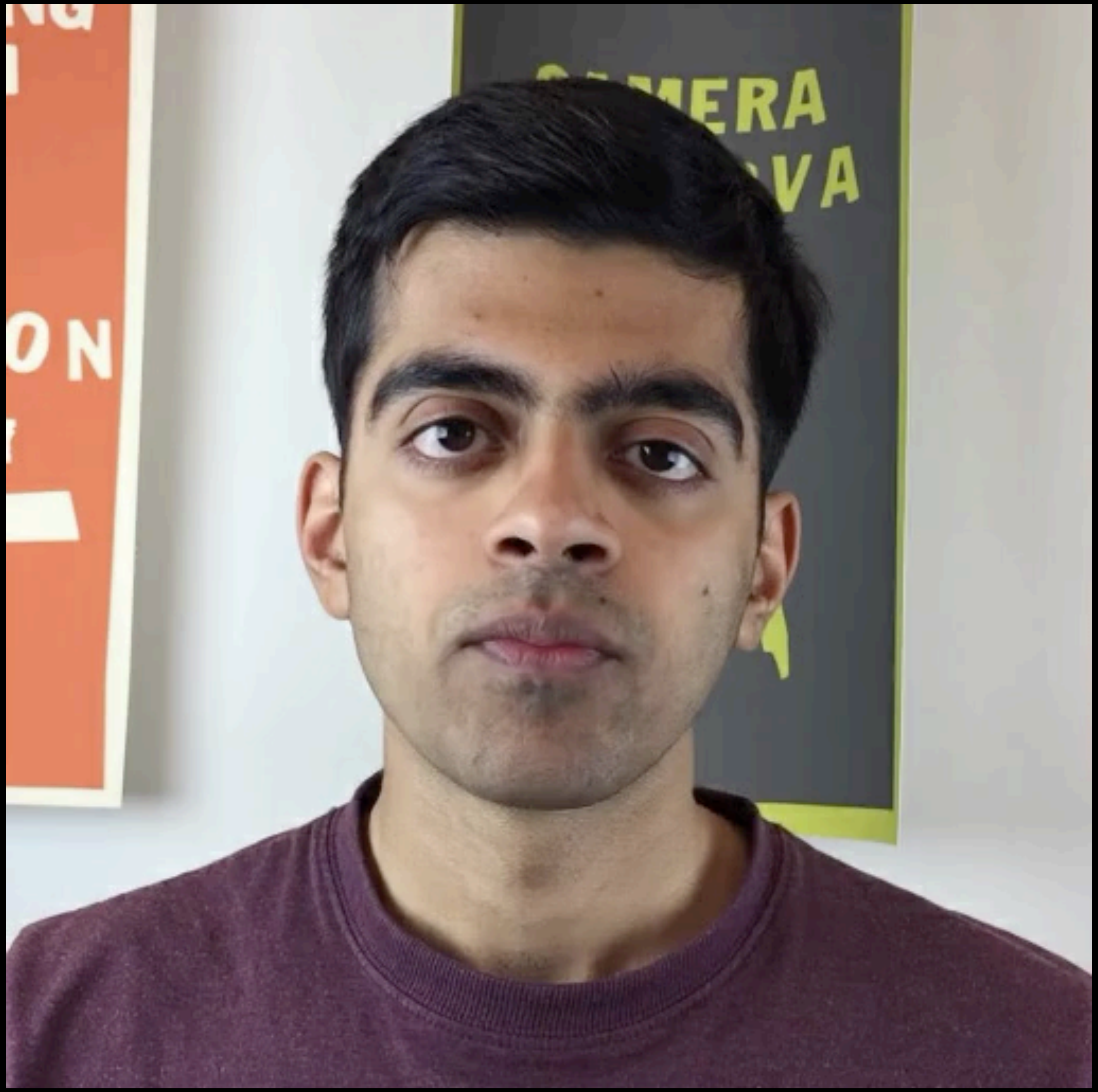
Blend with input (spatial)

Voice by VoCo
[Jin et al. ’17]

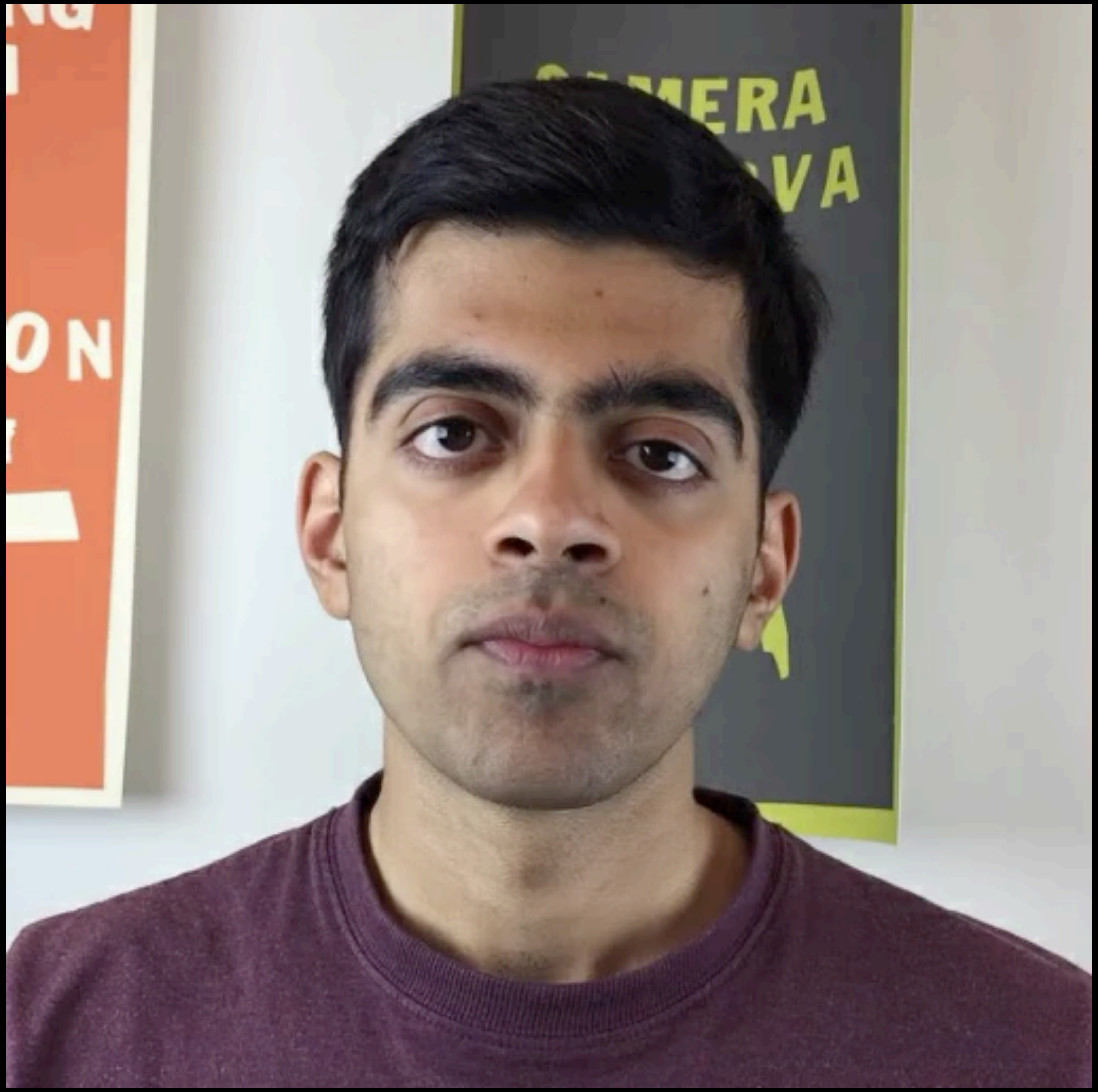
Voice by VoCo
[Jin et al. ’17]

Evaluation & comparisons

Parameter blending

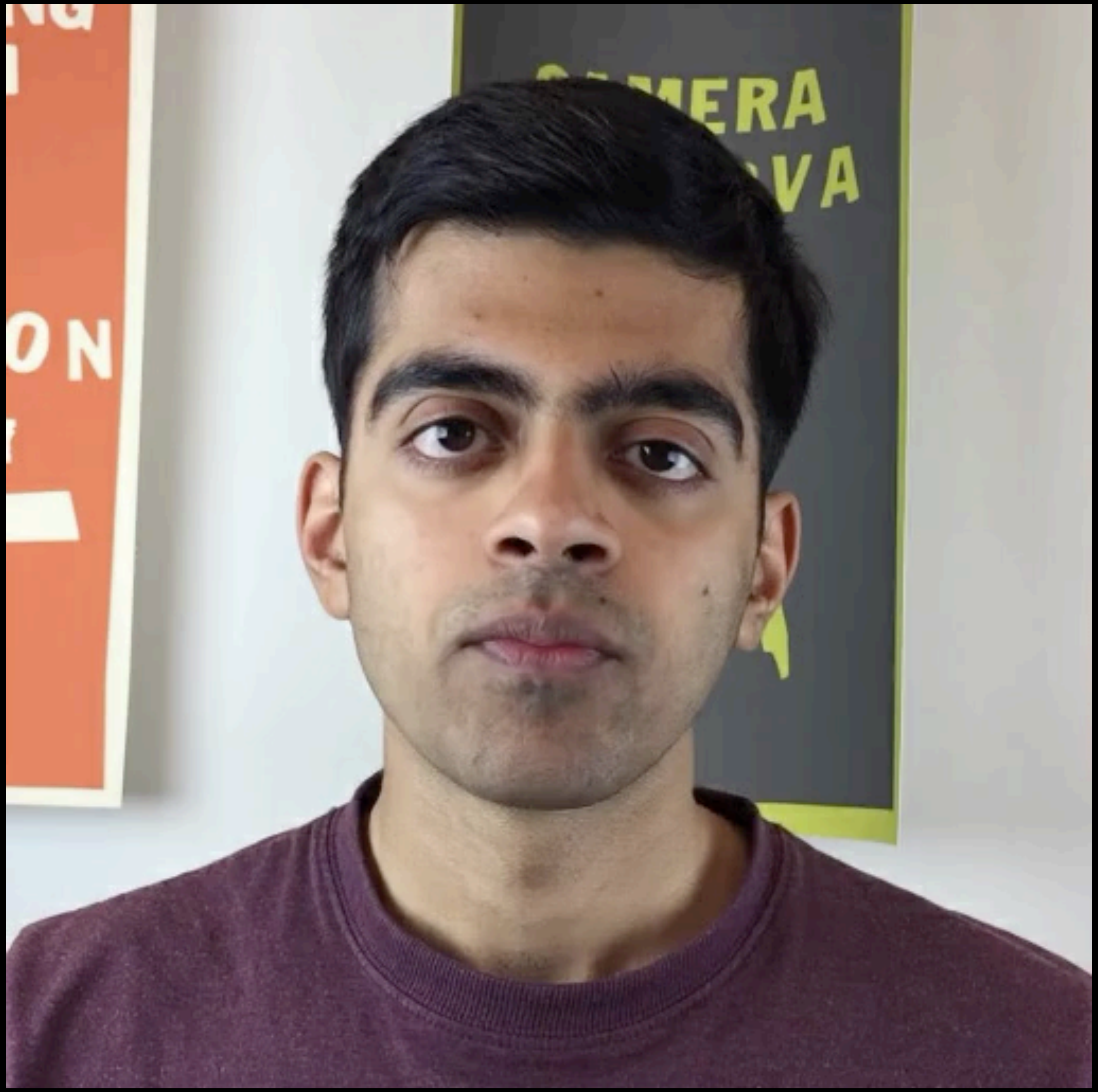


Without blending

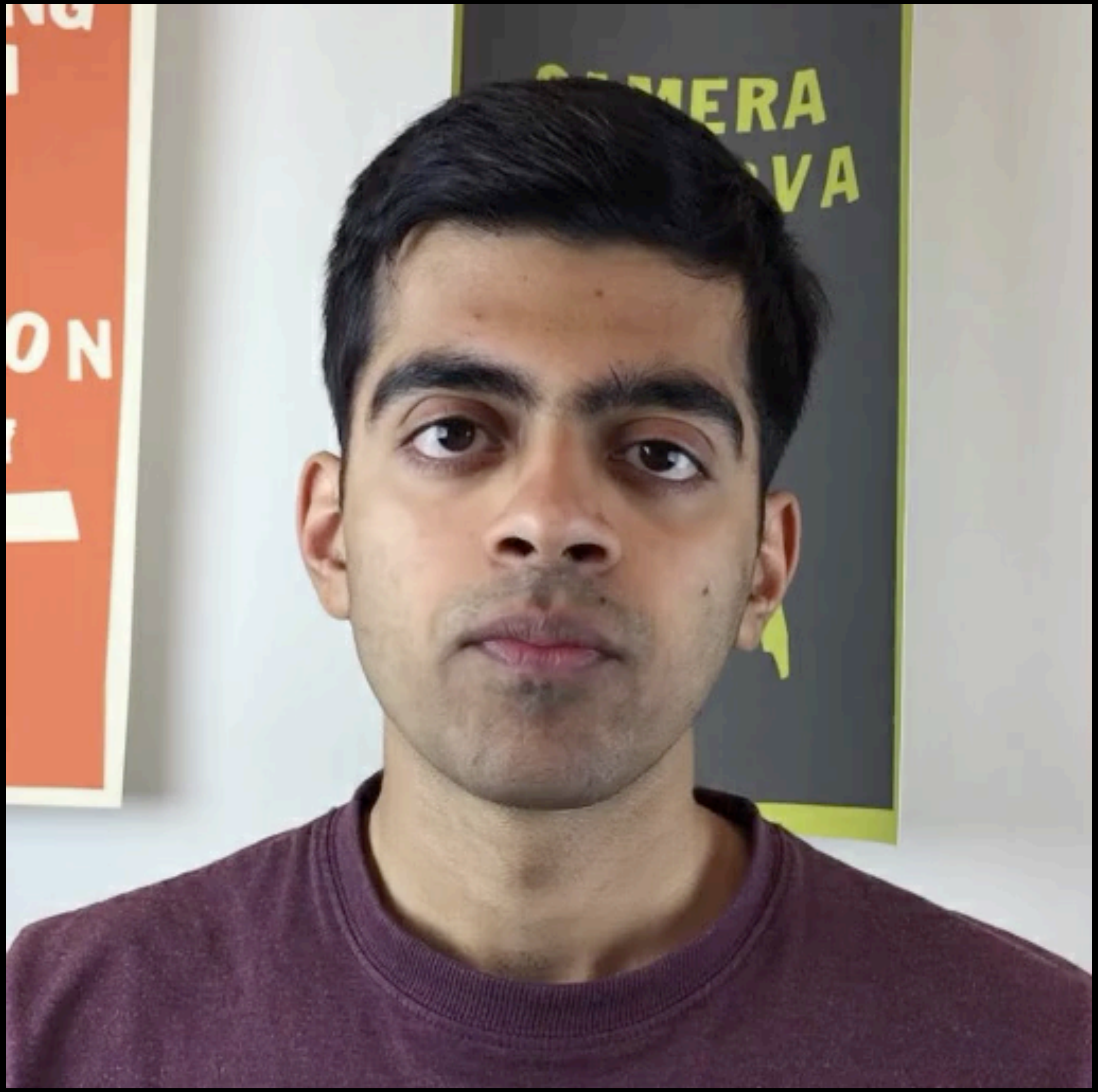


With blending

Parameter blending

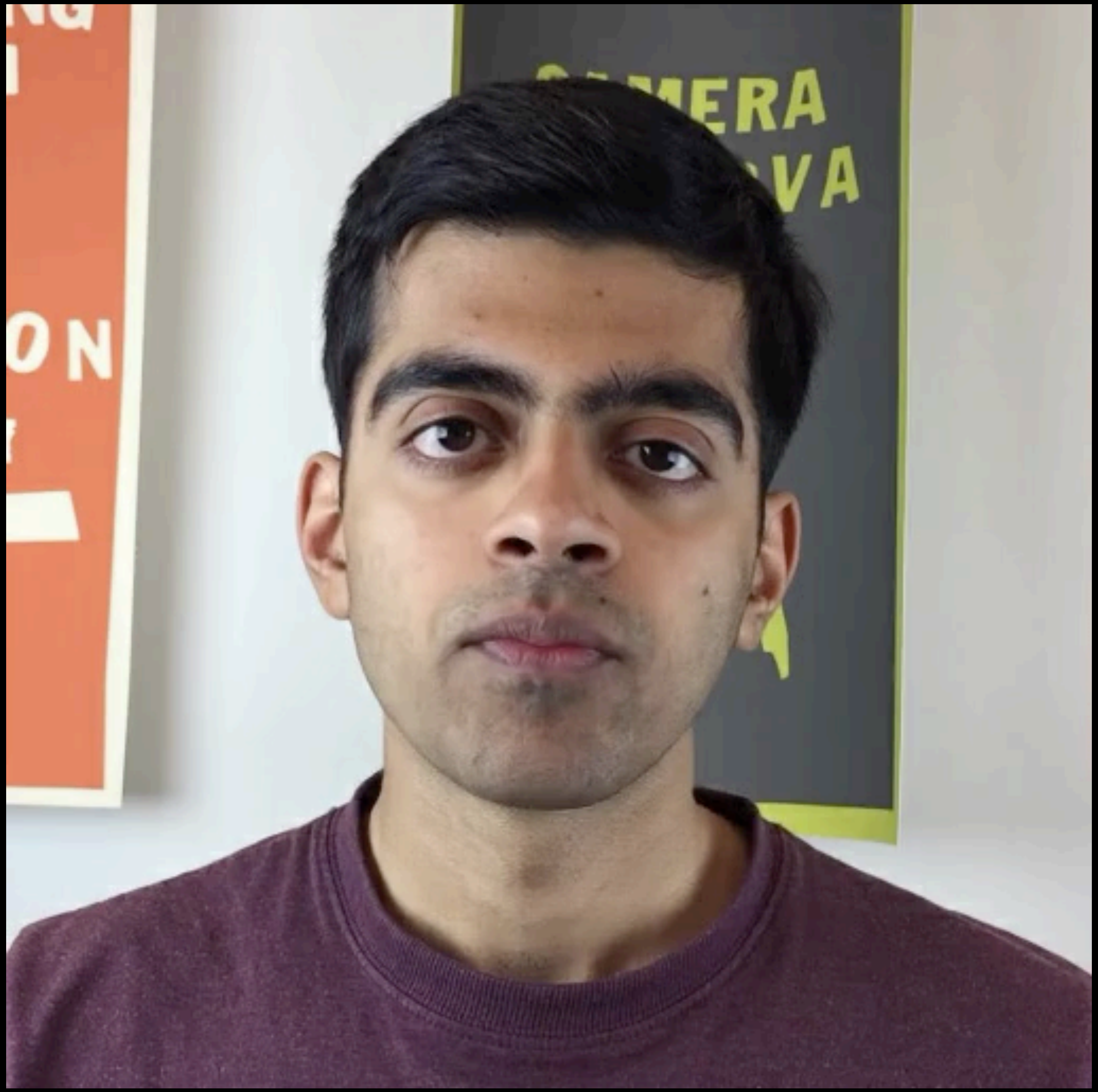


Without blending

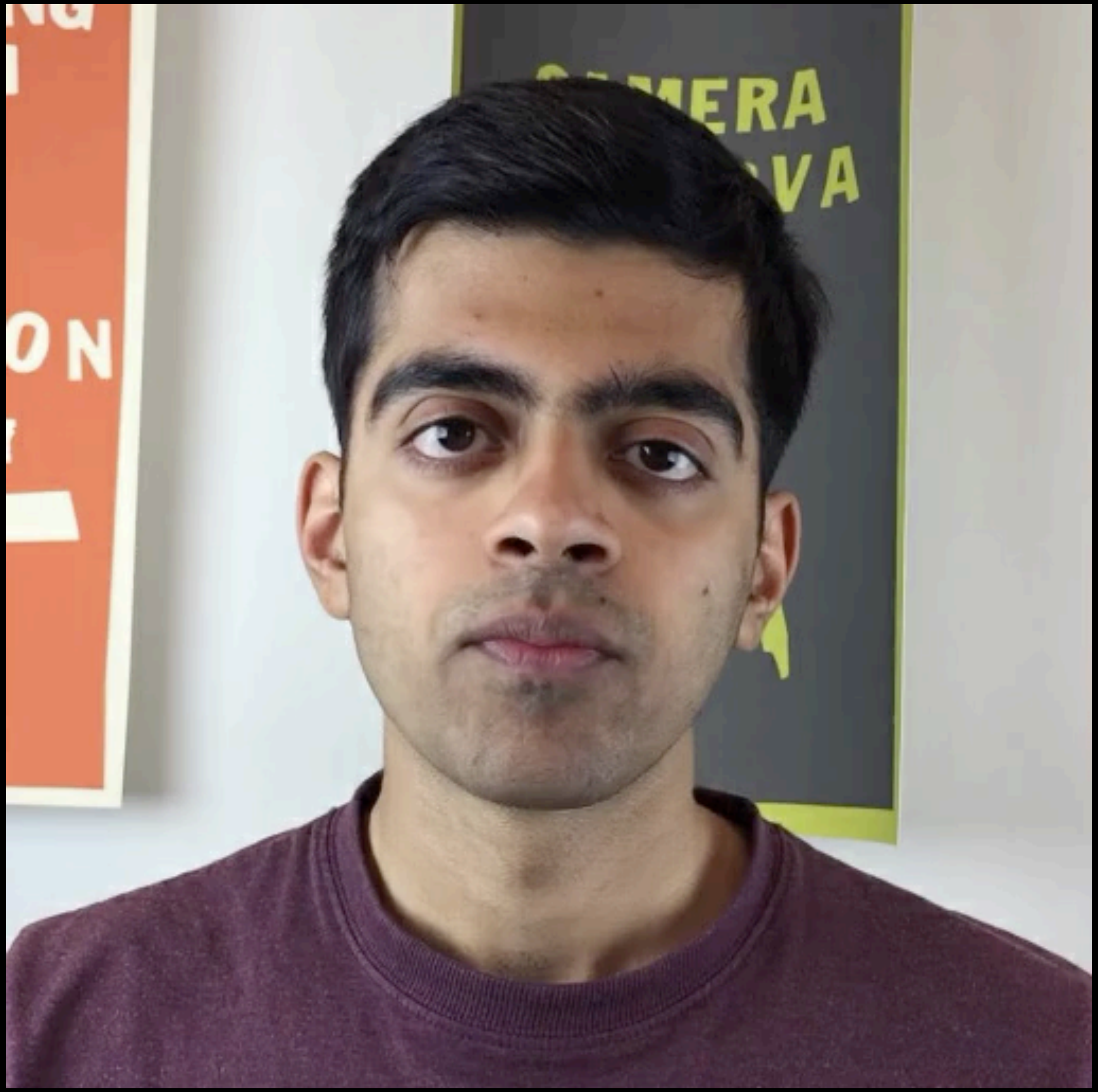


With blending

Parameter blending



Without blending



With blending

Parameter blending

Without blending

With blending

Parameter blending

Without blending

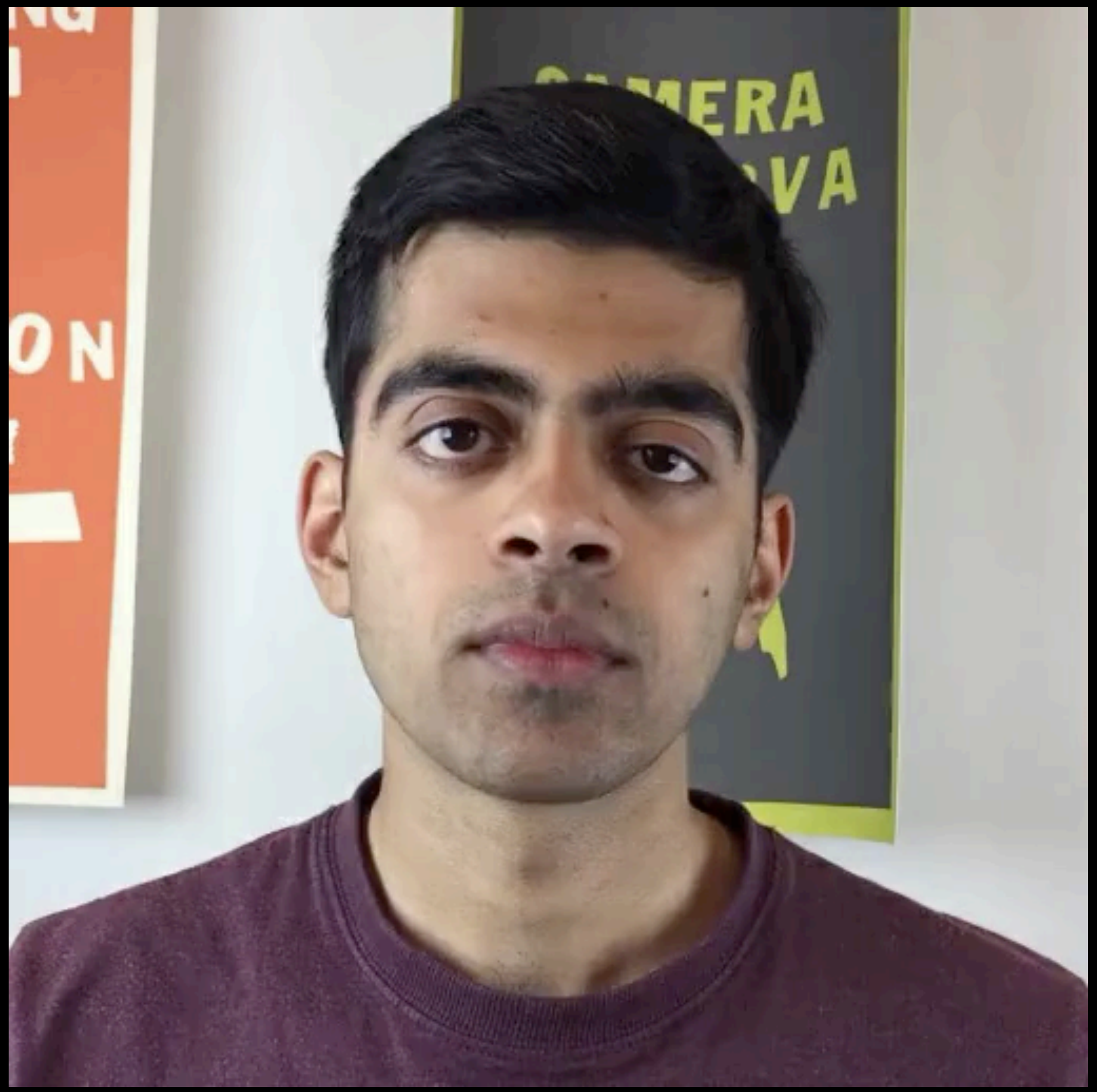
With blending

Parameter blending

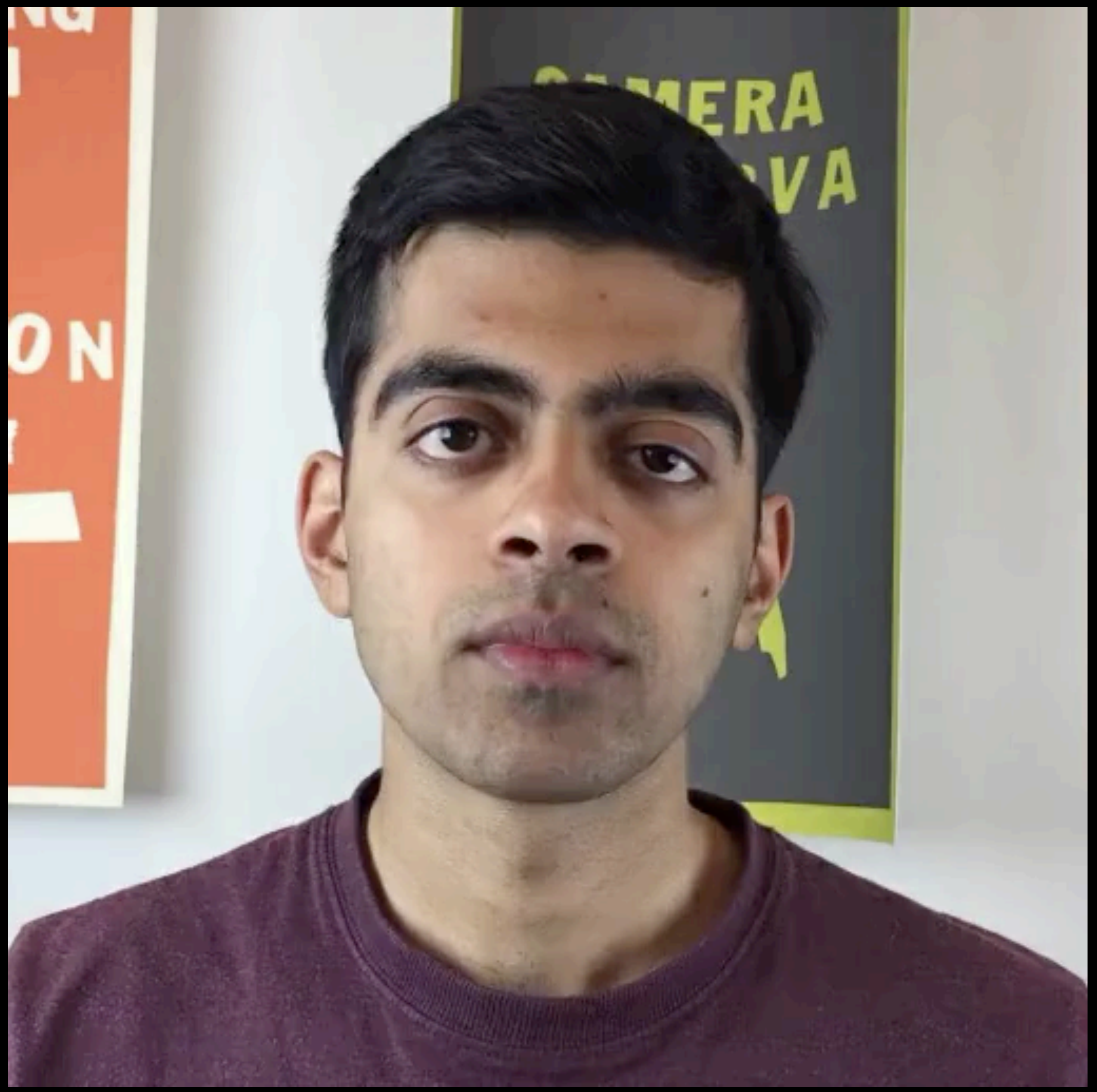
Without blending

With blending

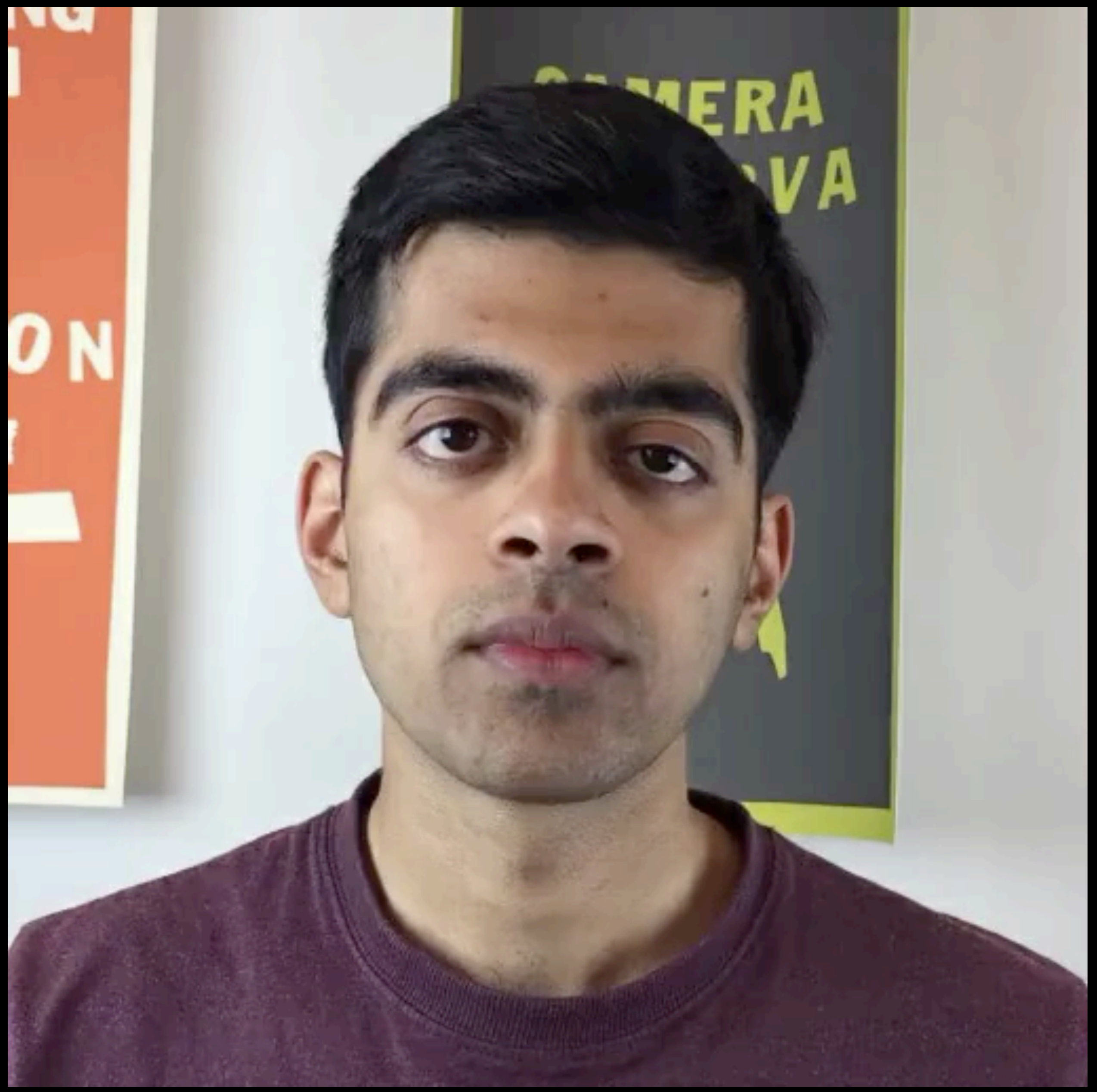
Dataset size



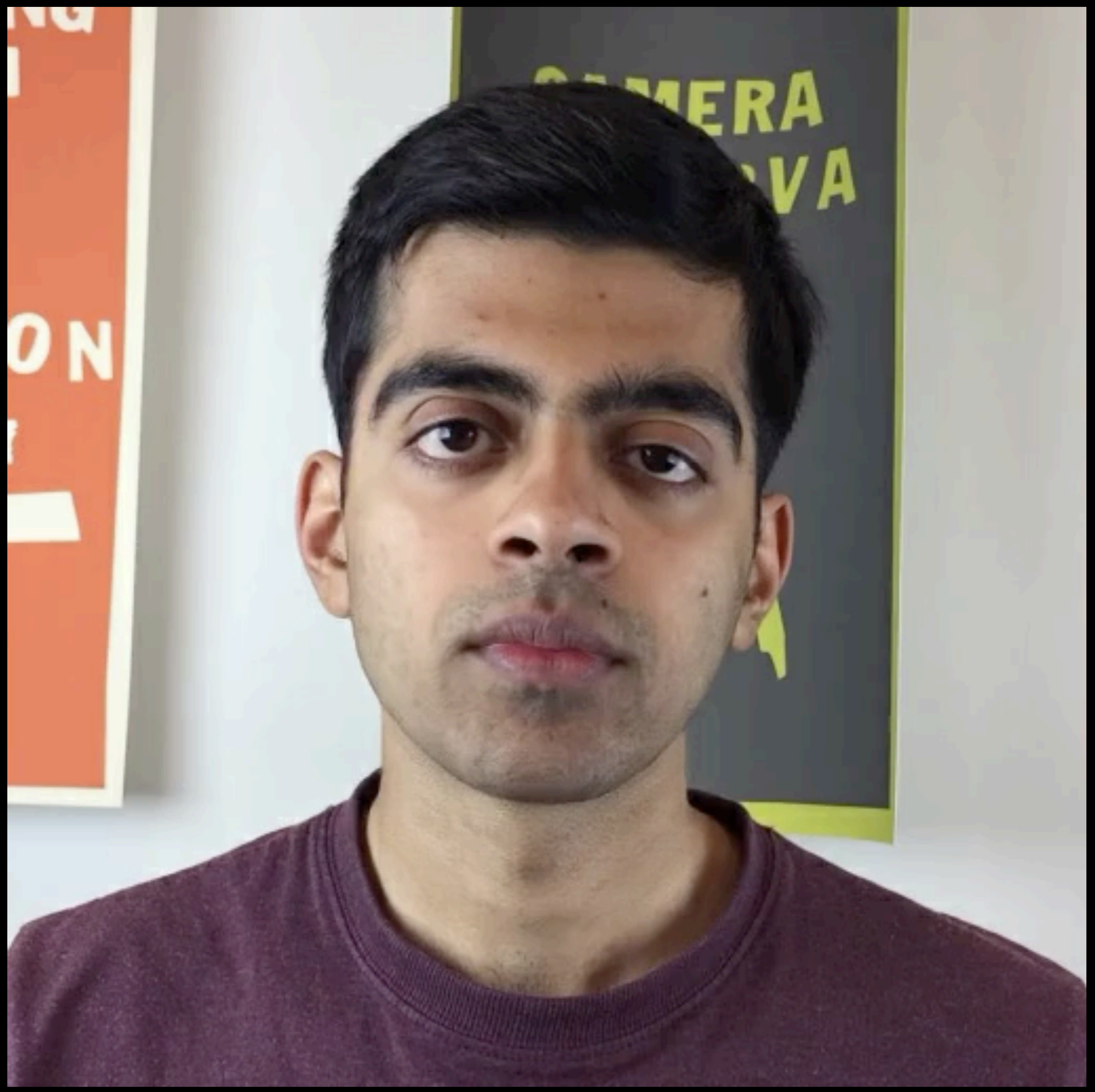
5% data



10% data

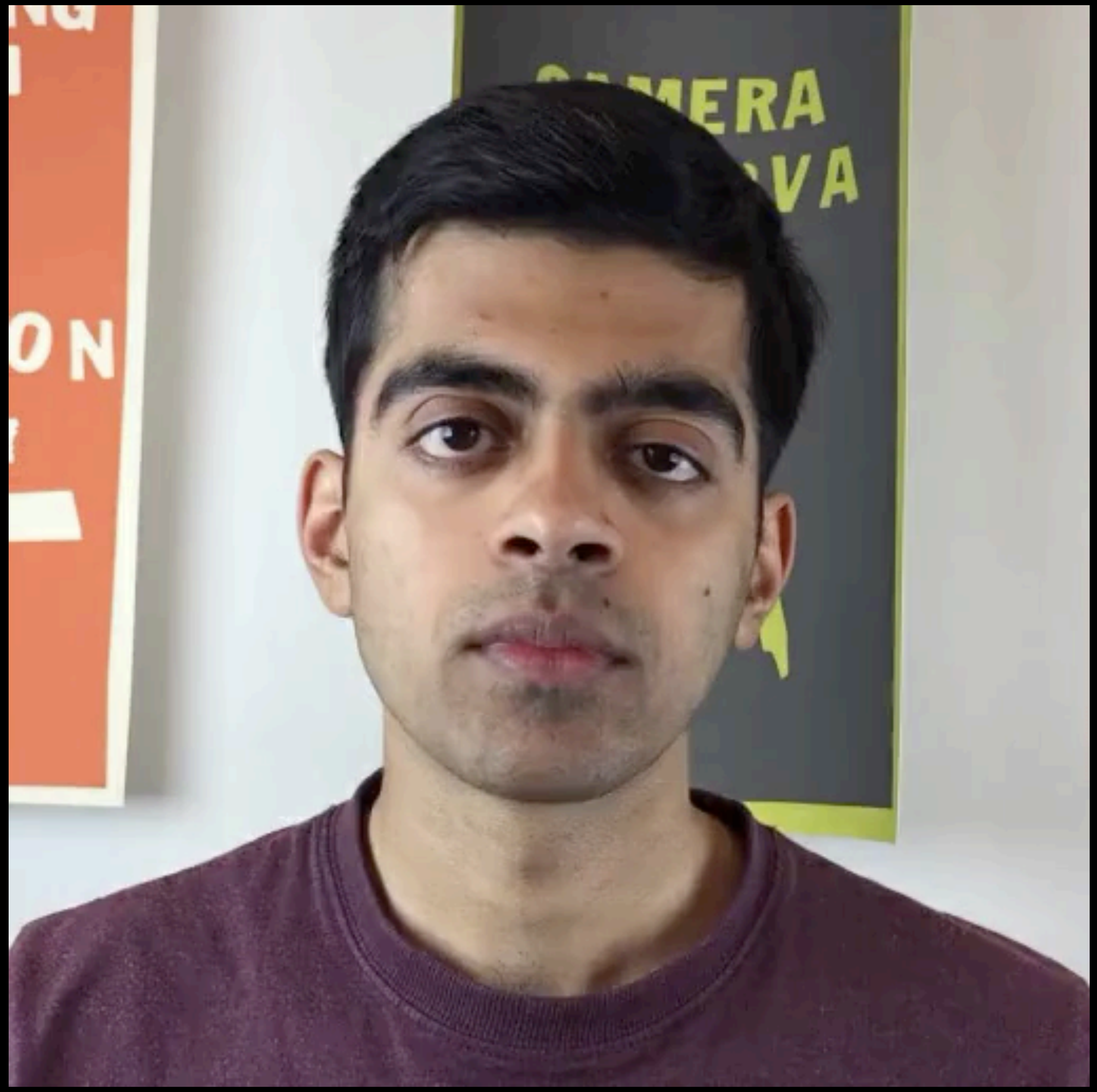


50% data

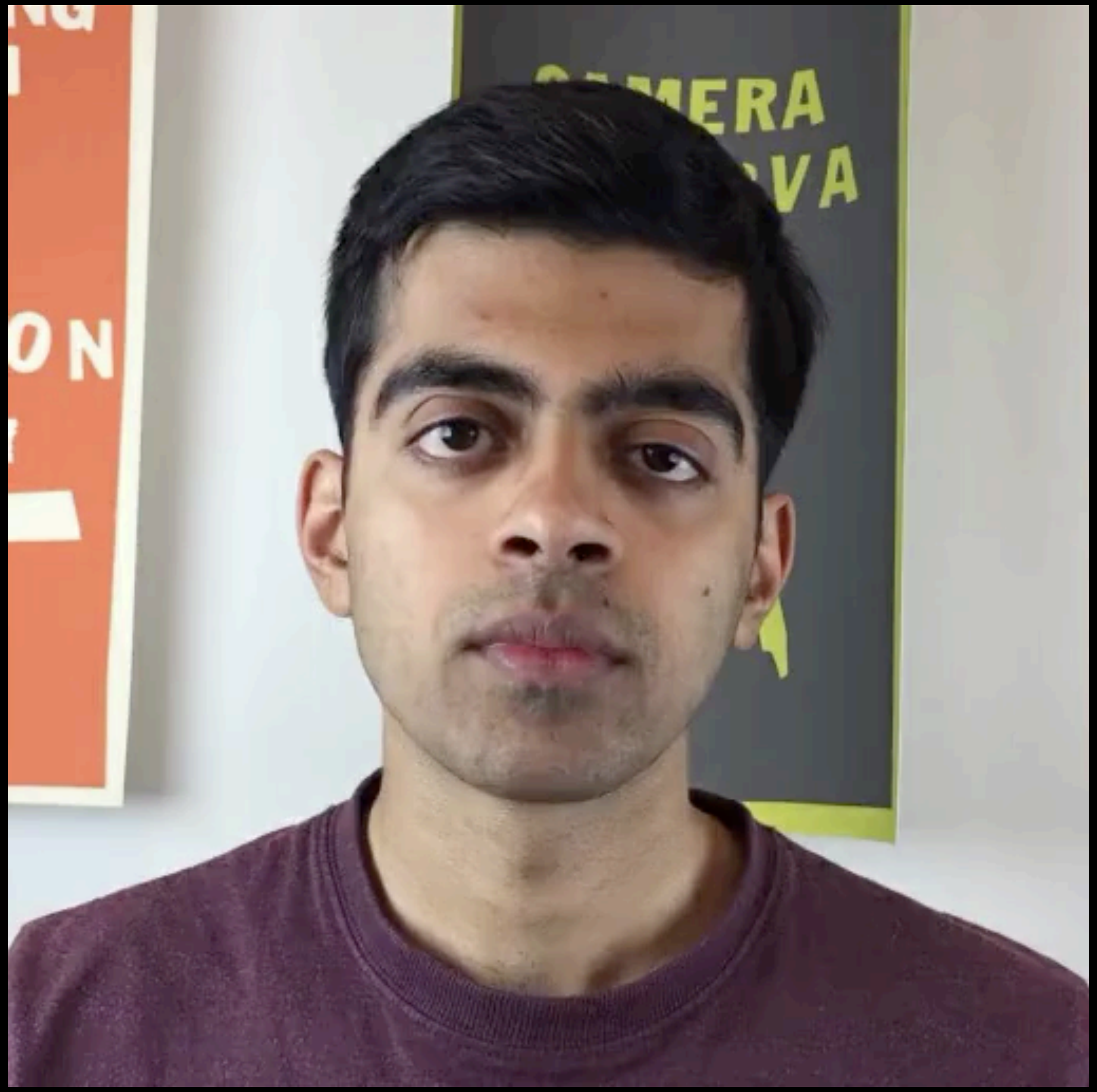


100% data

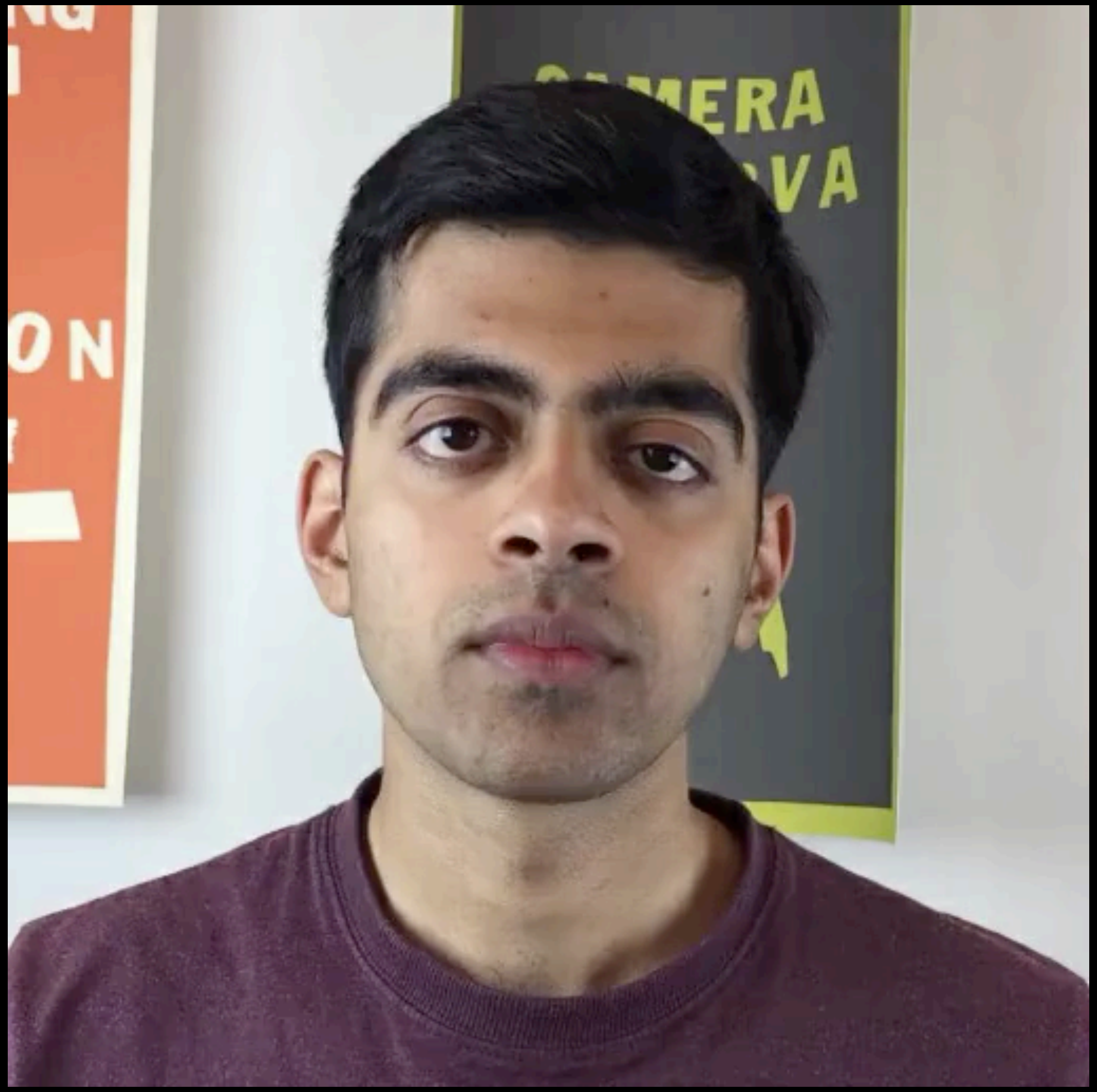
Dataset size



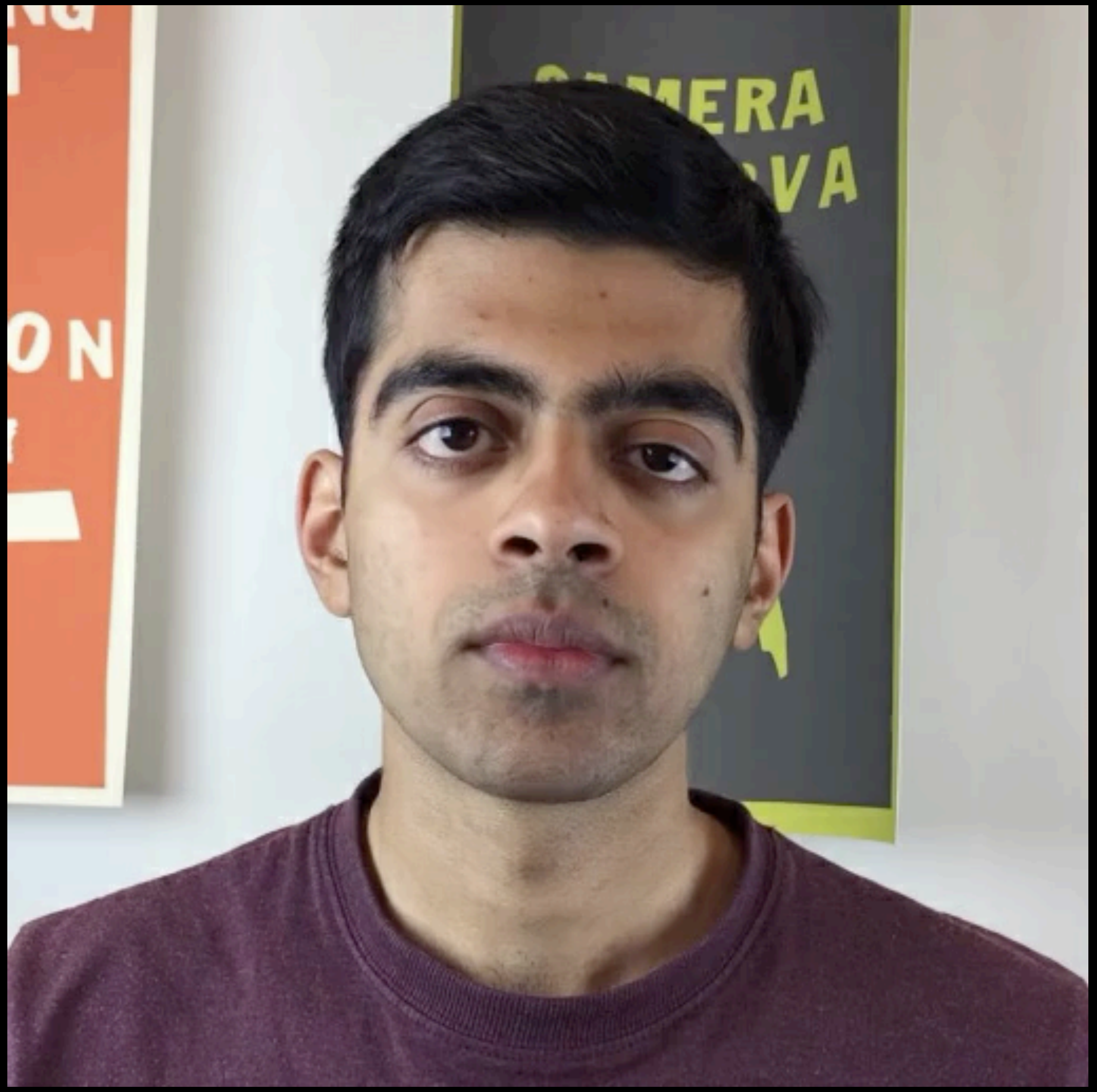
5% data



10% data

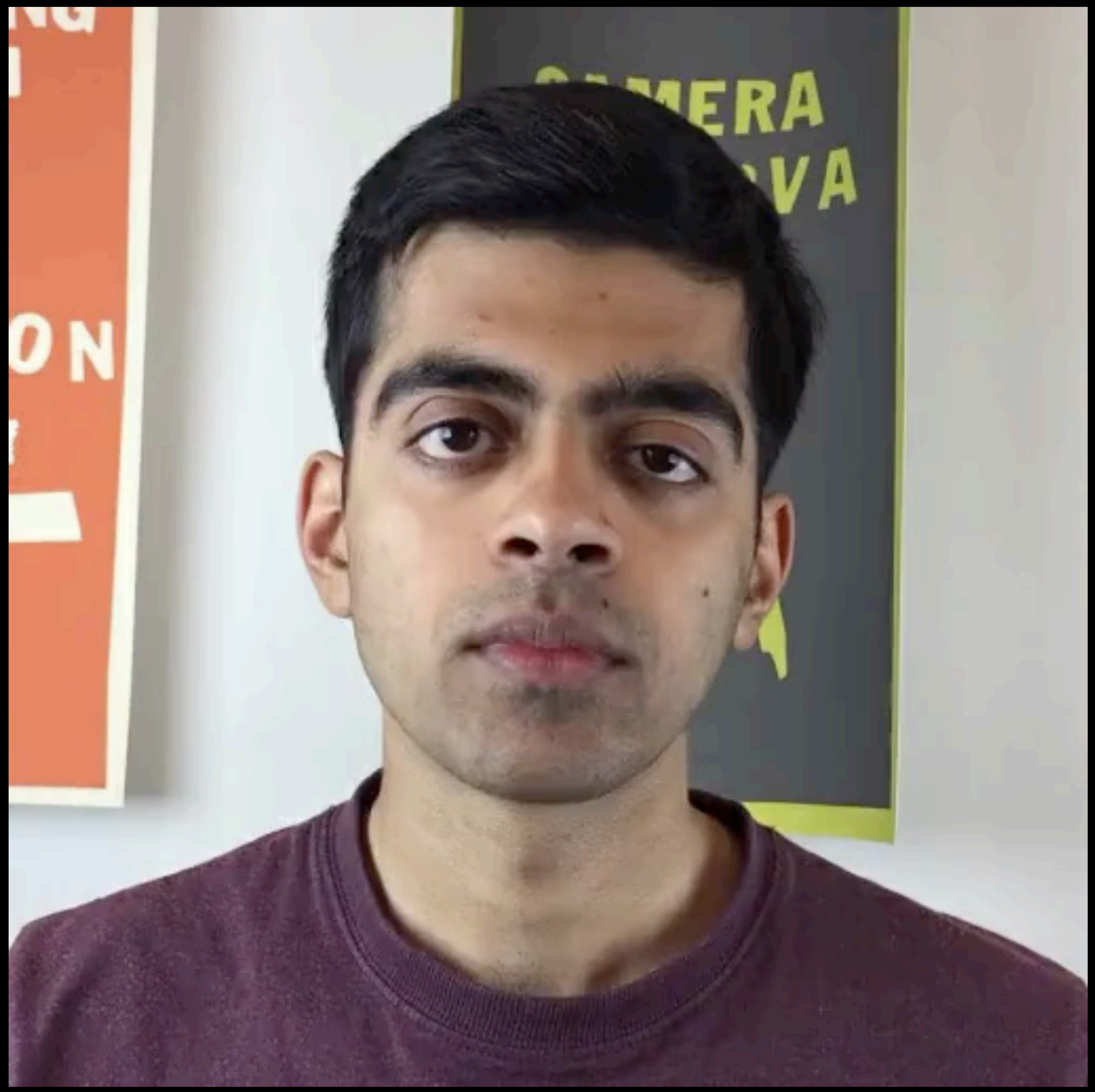


50% data

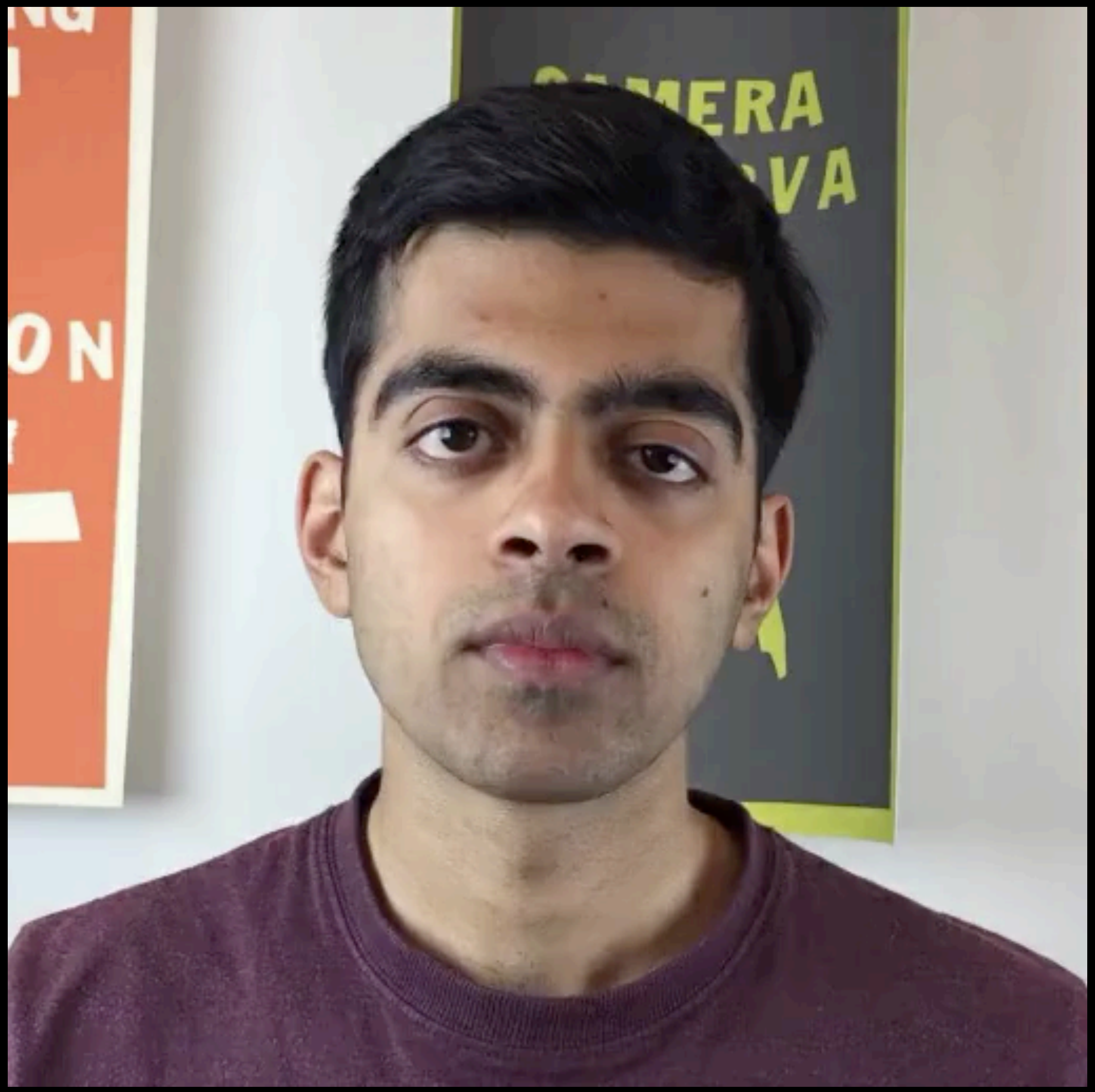


100% data

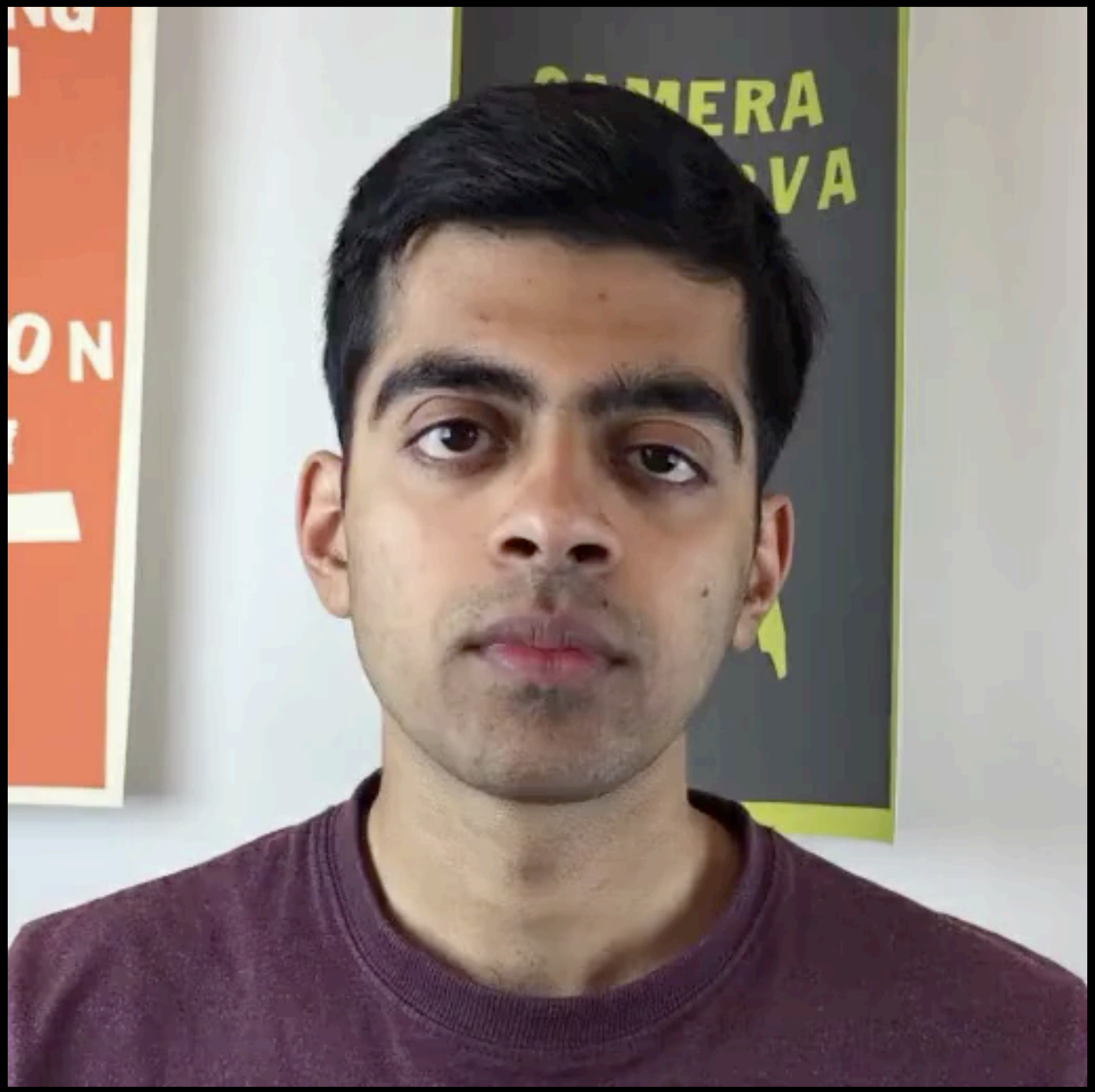
Dataset size



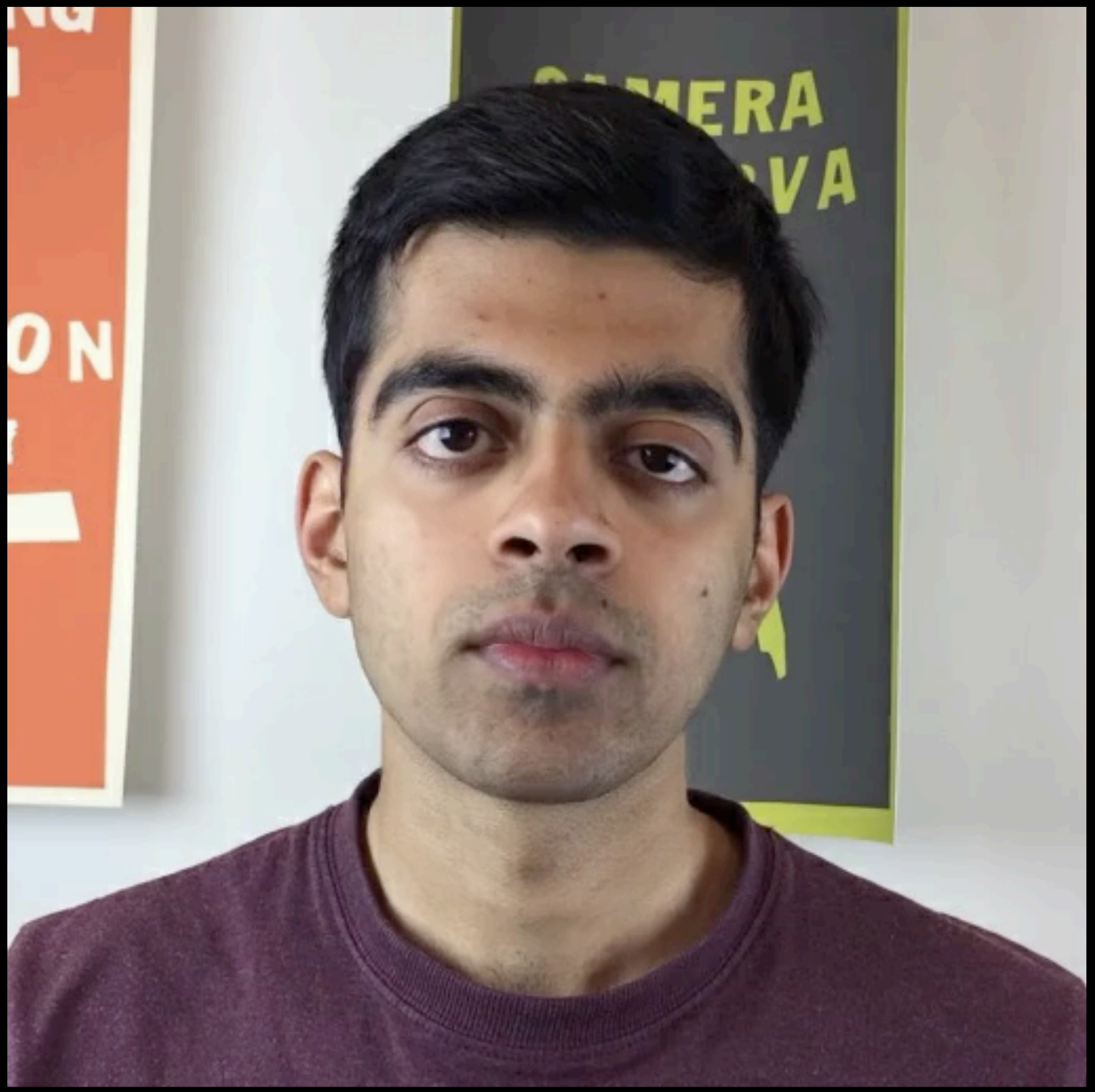
5% data



10% data

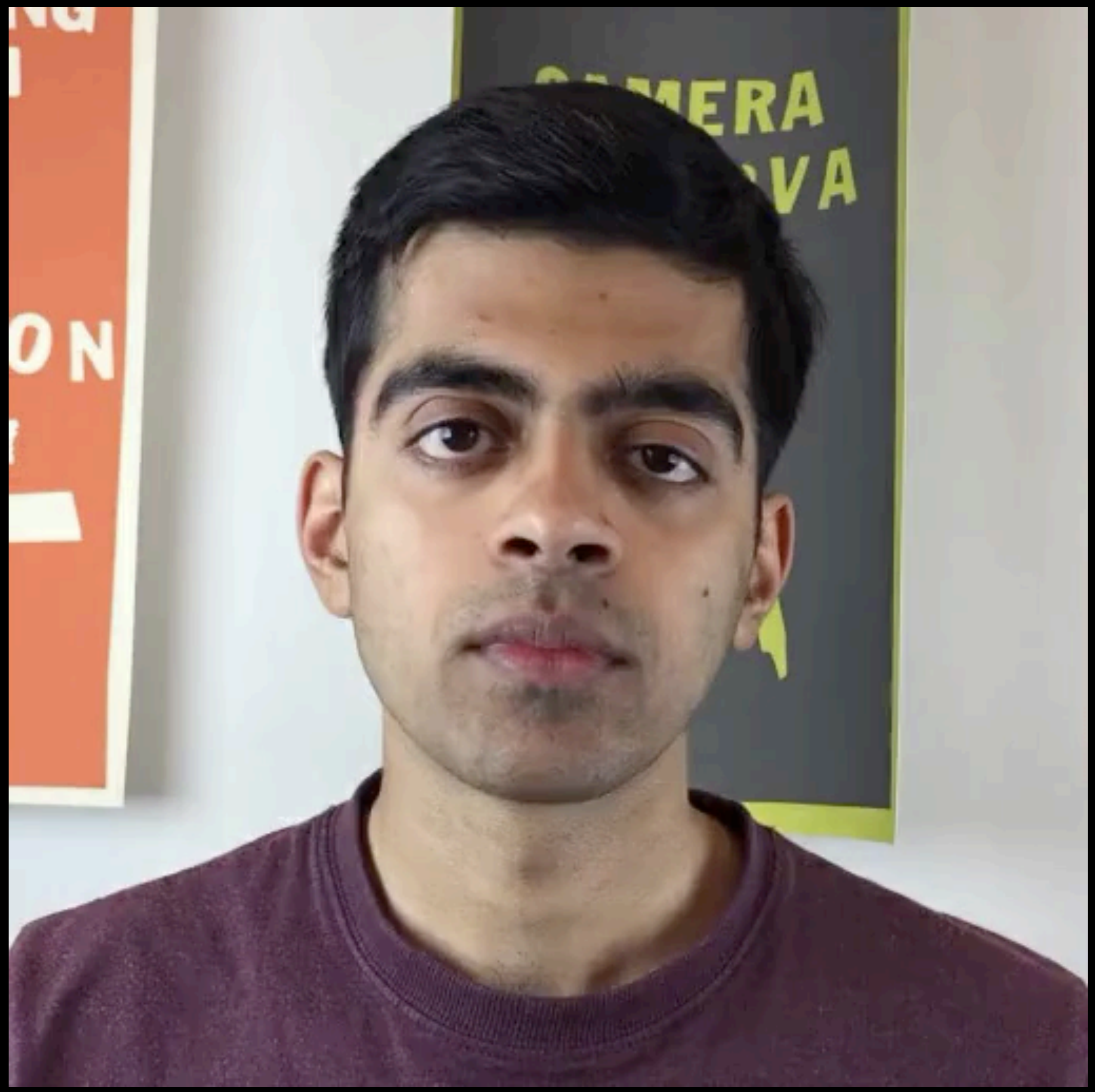


50% data

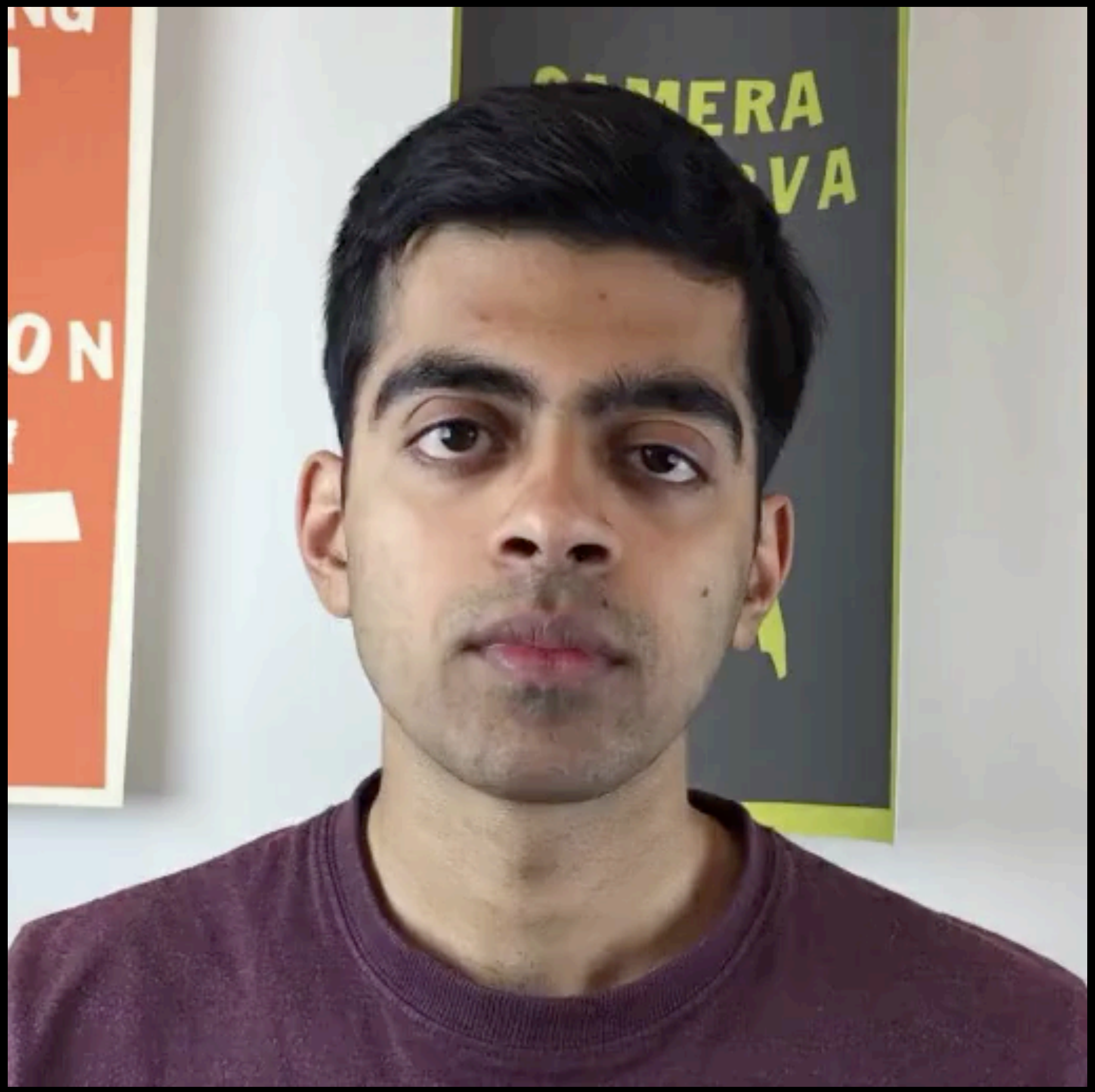


100% data

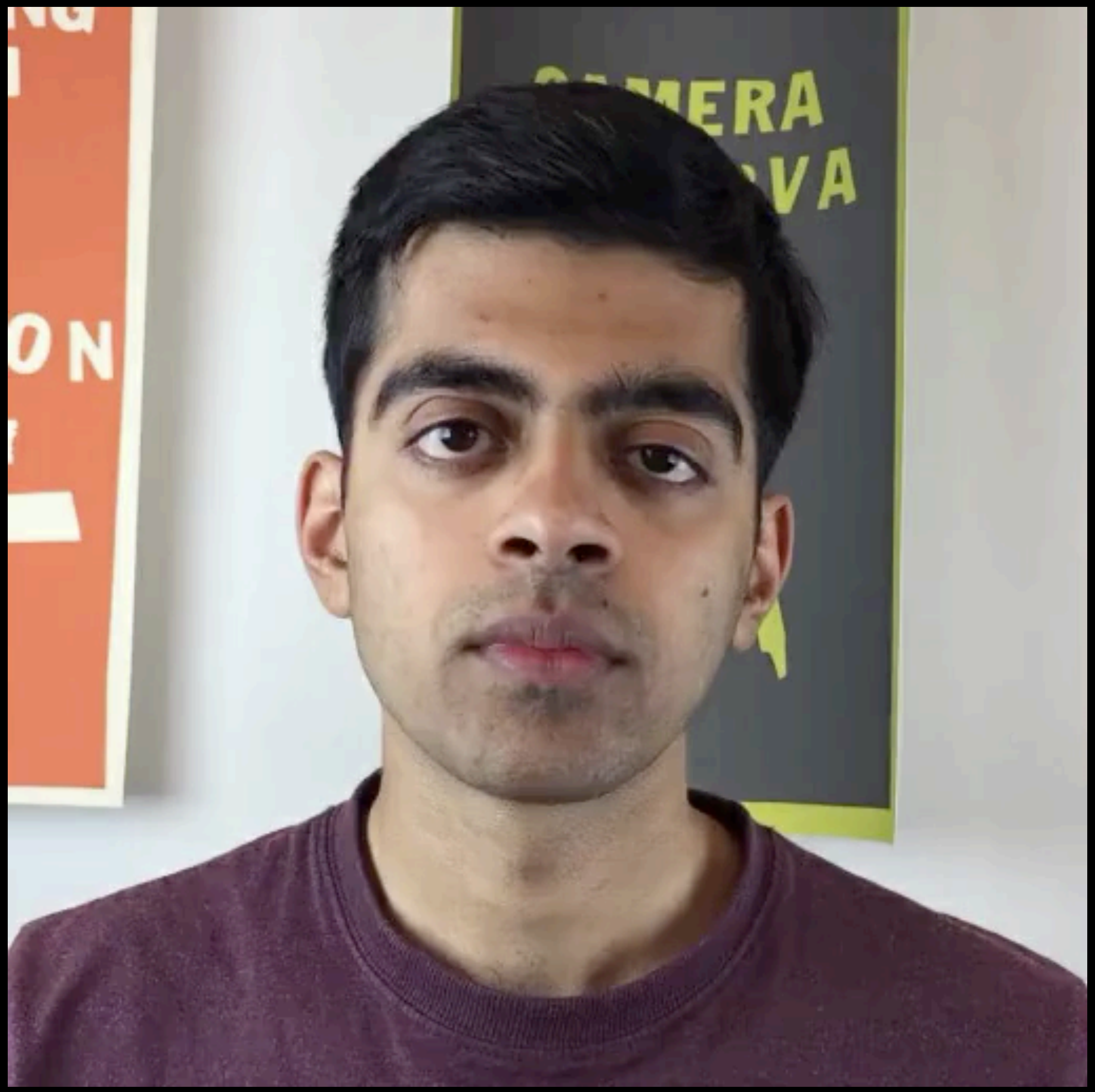
Dataset size



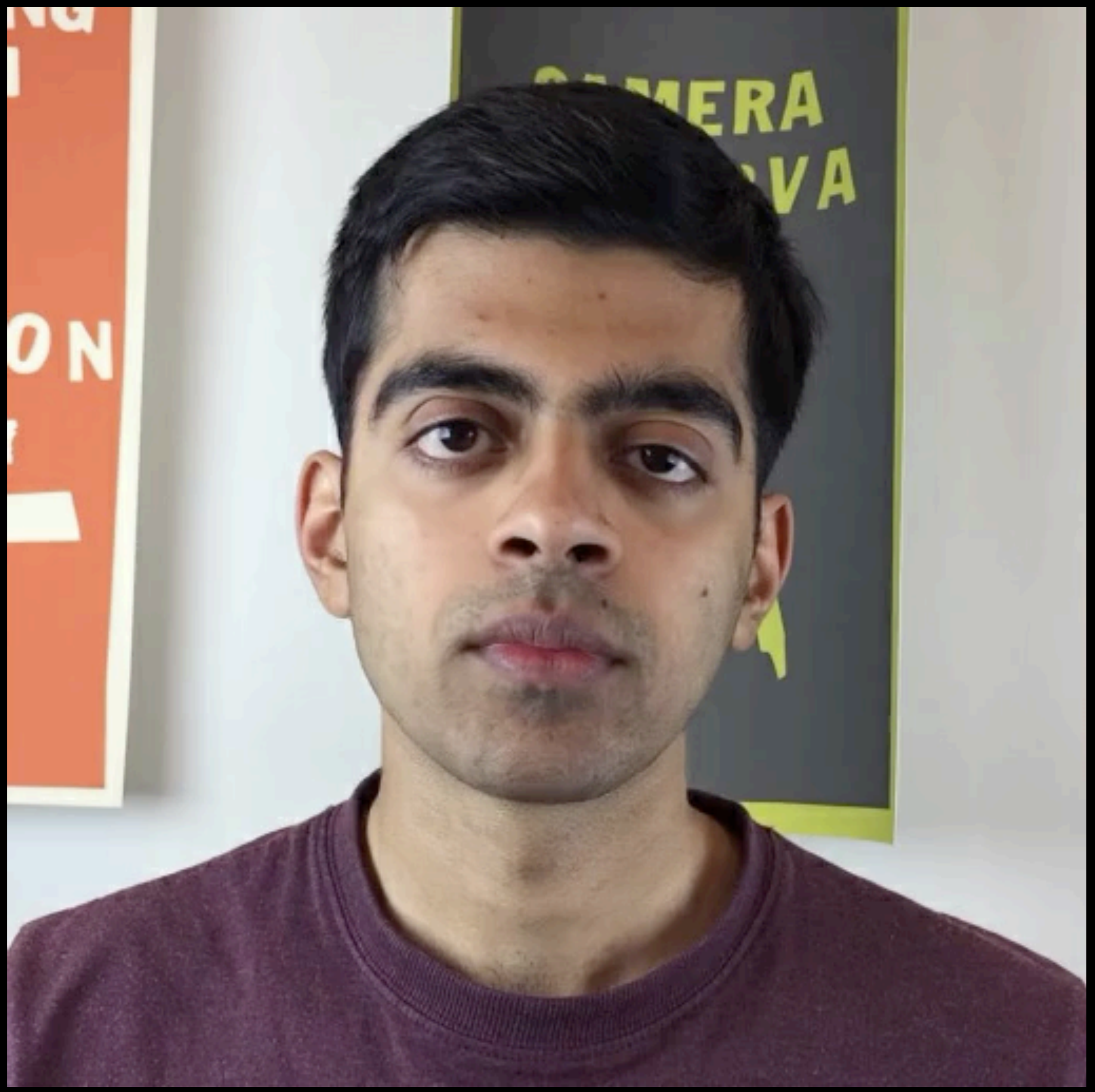
5% data



10% data

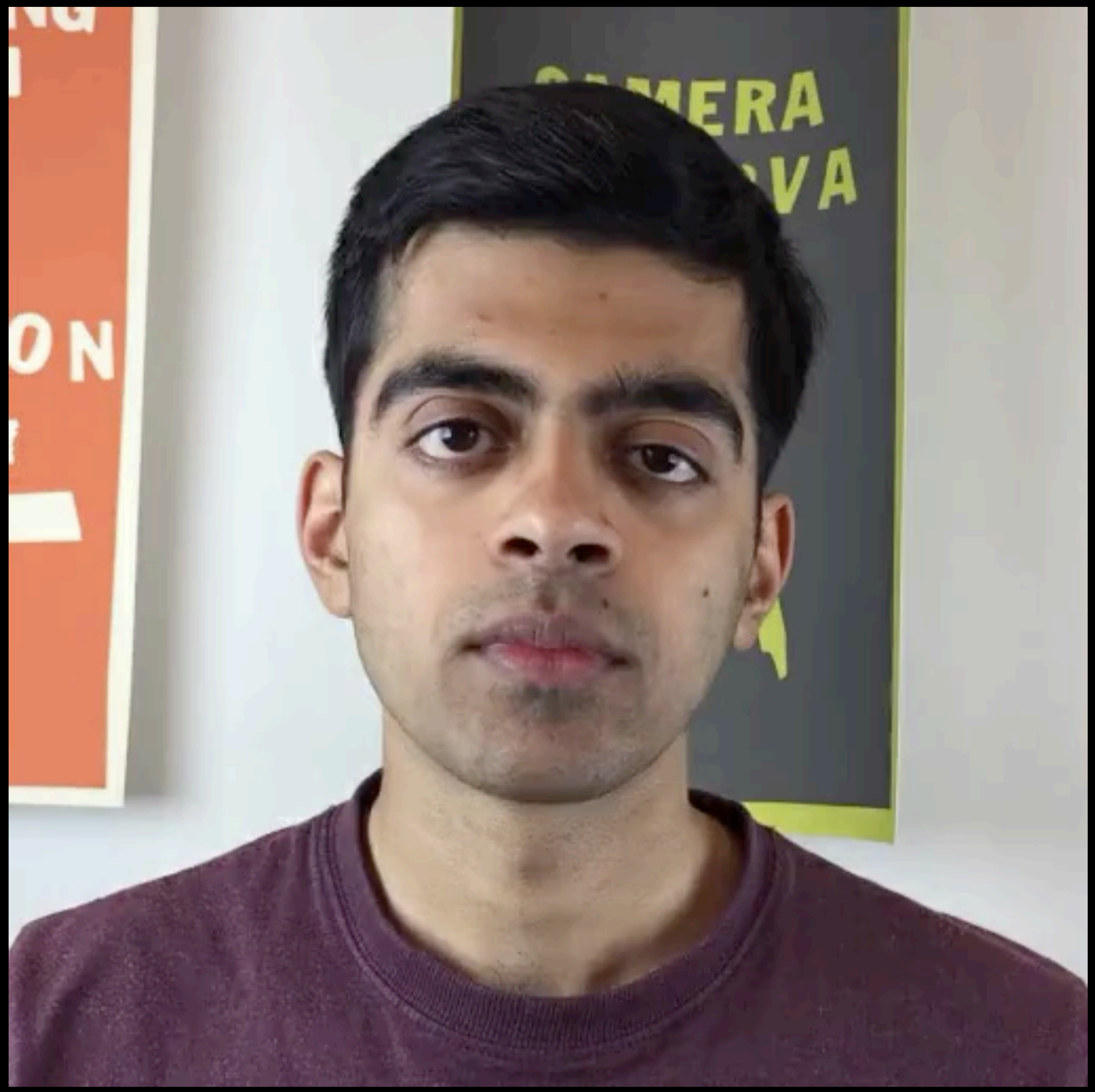


50% data

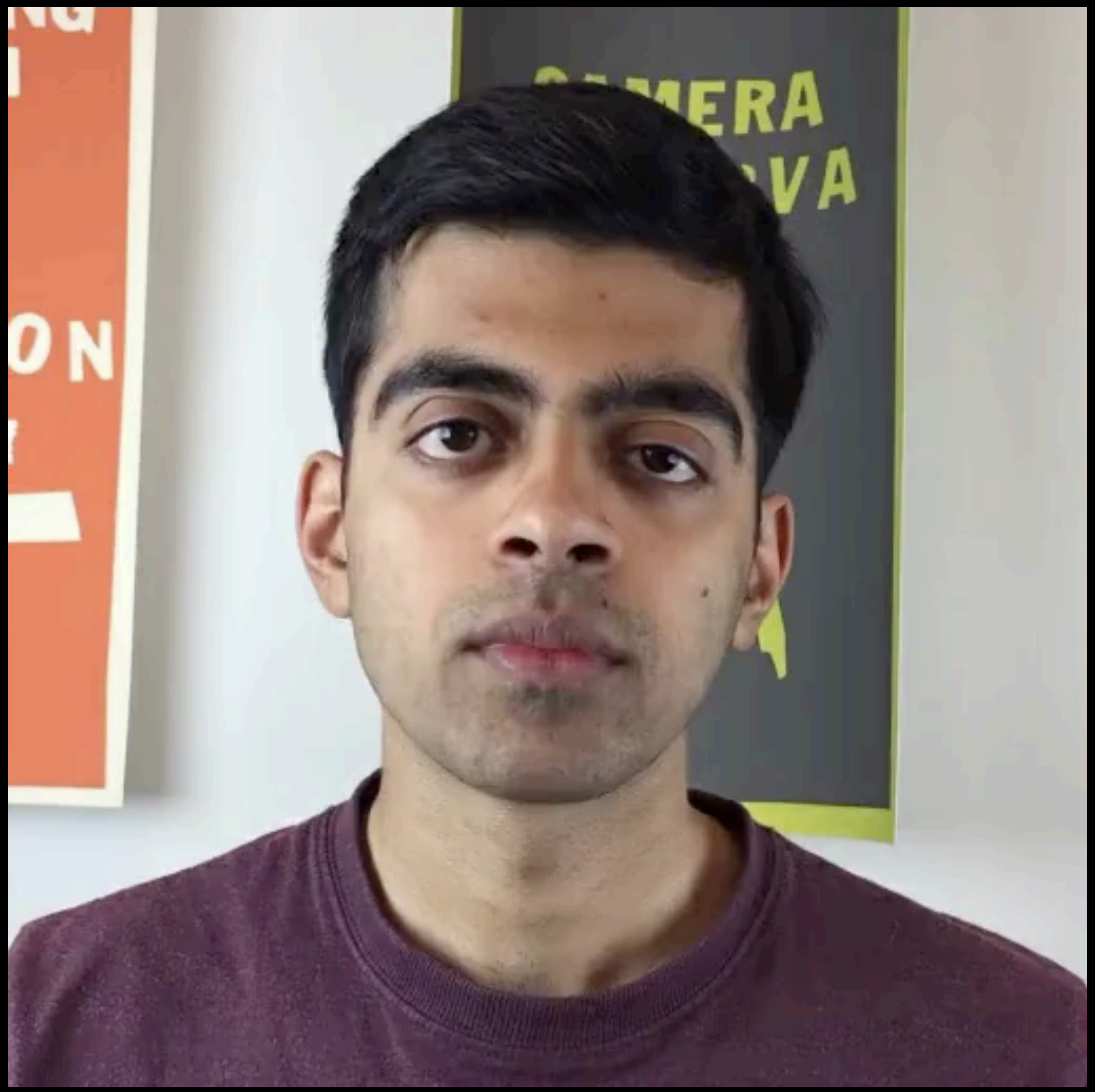


100% data

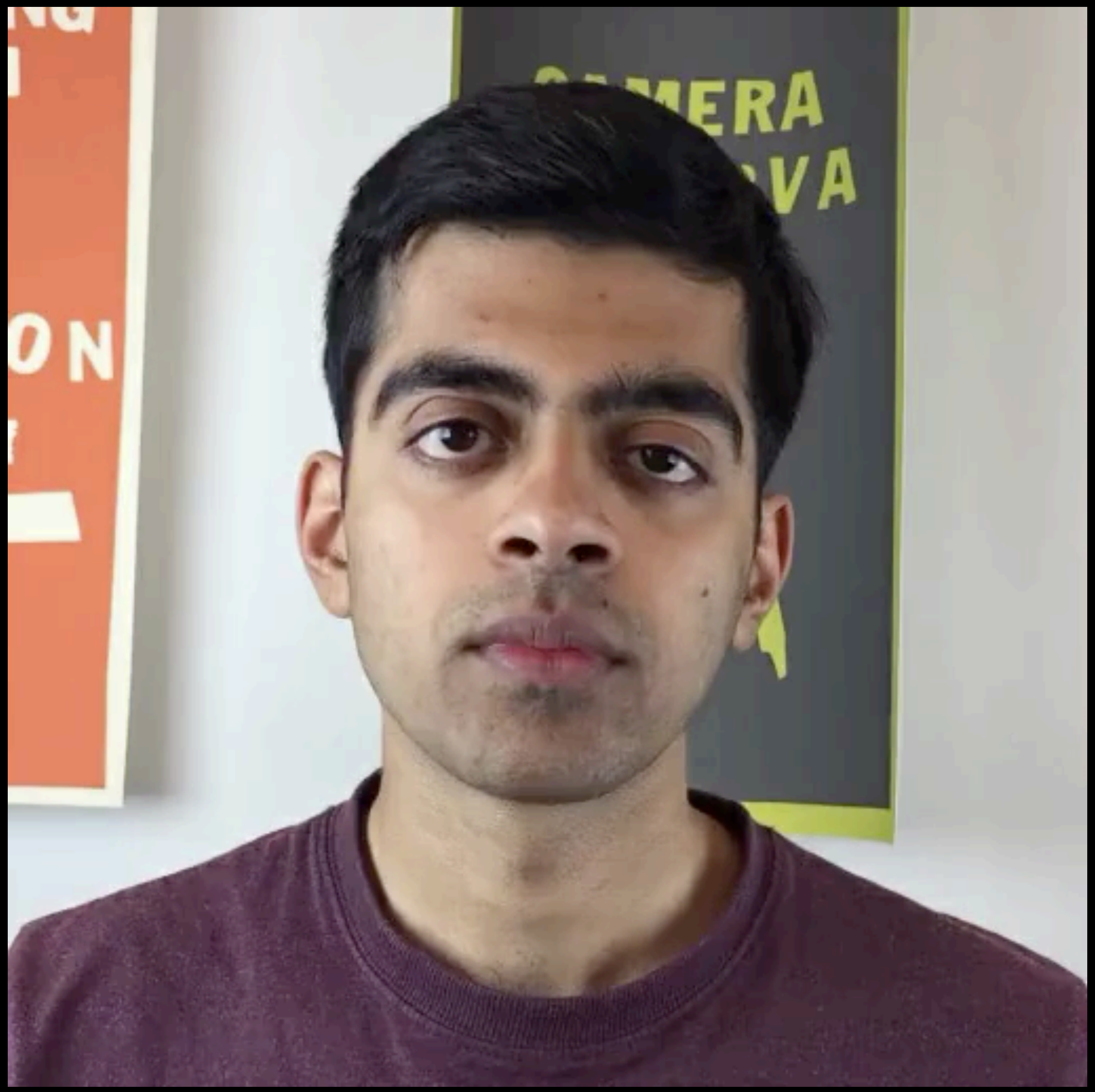
Dataset size



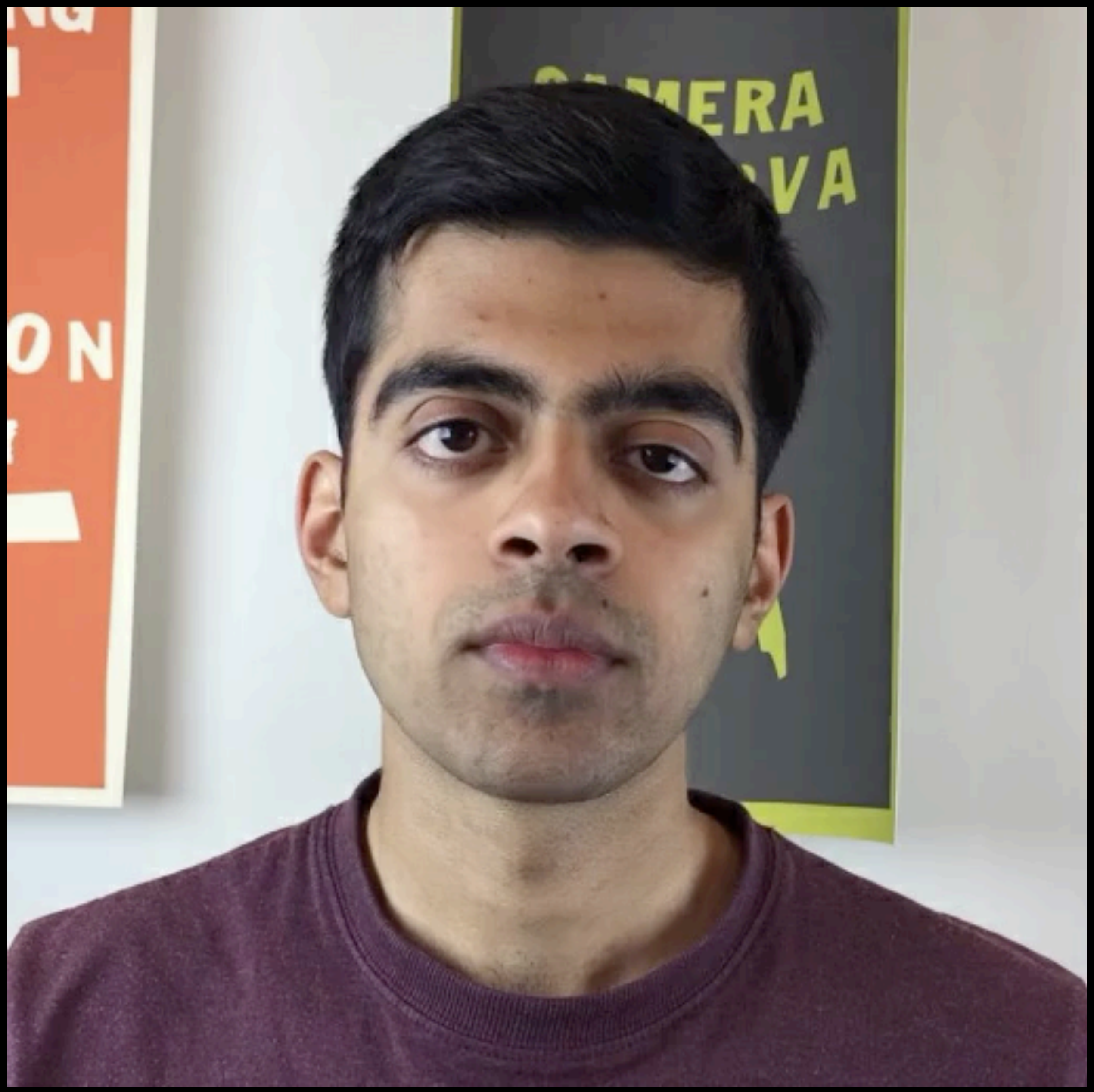
5% data



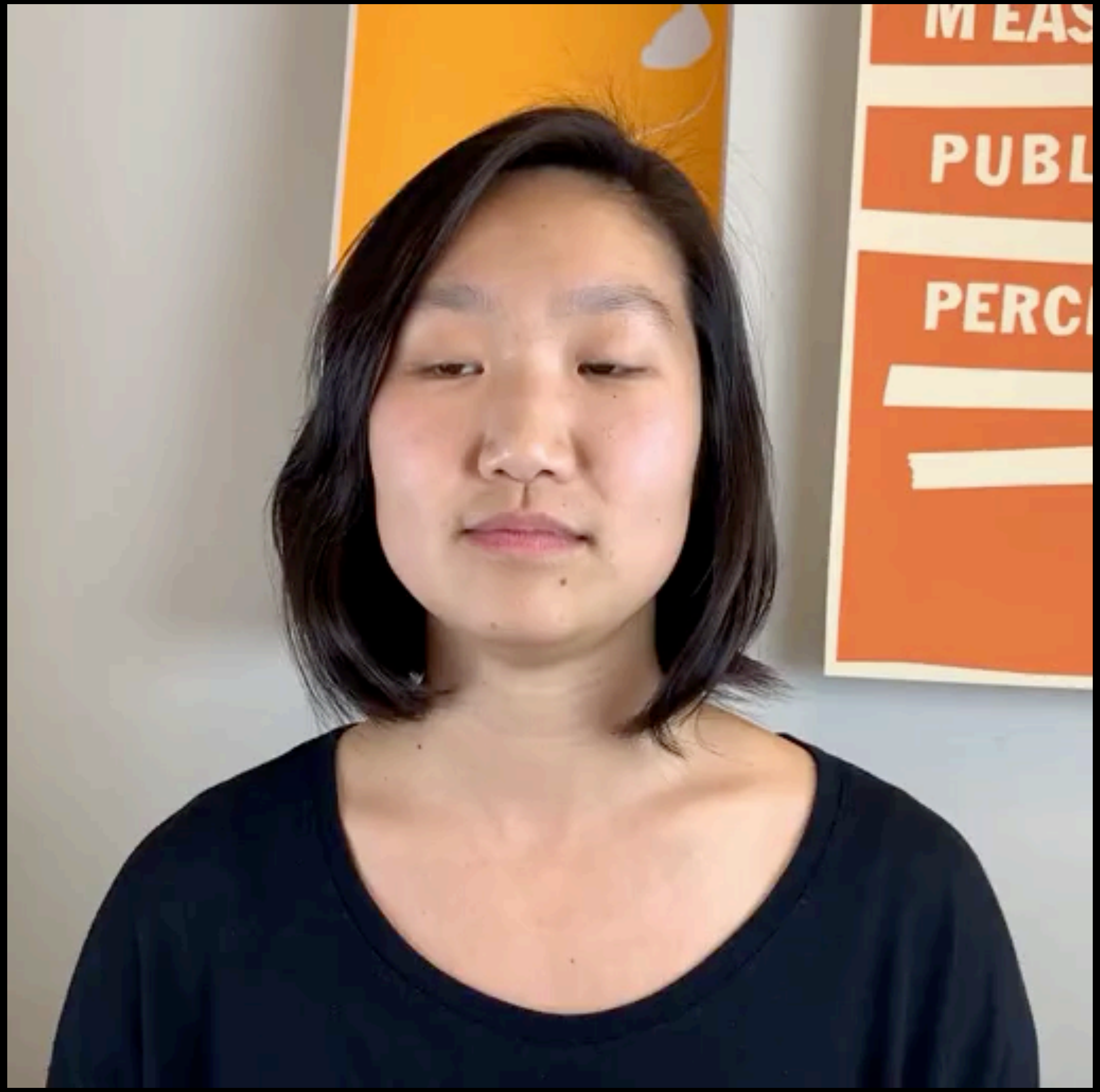
10% data



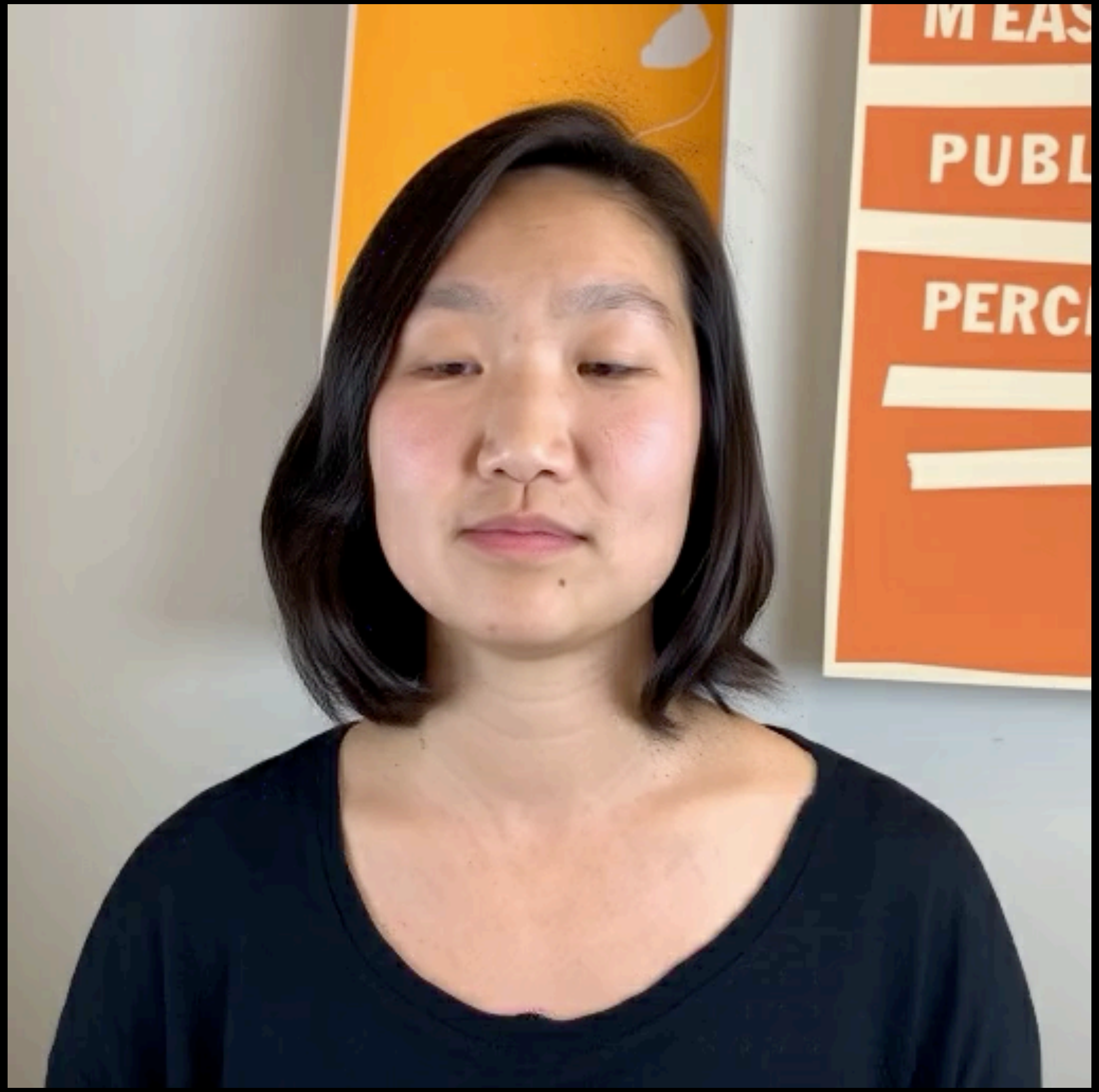
50% data



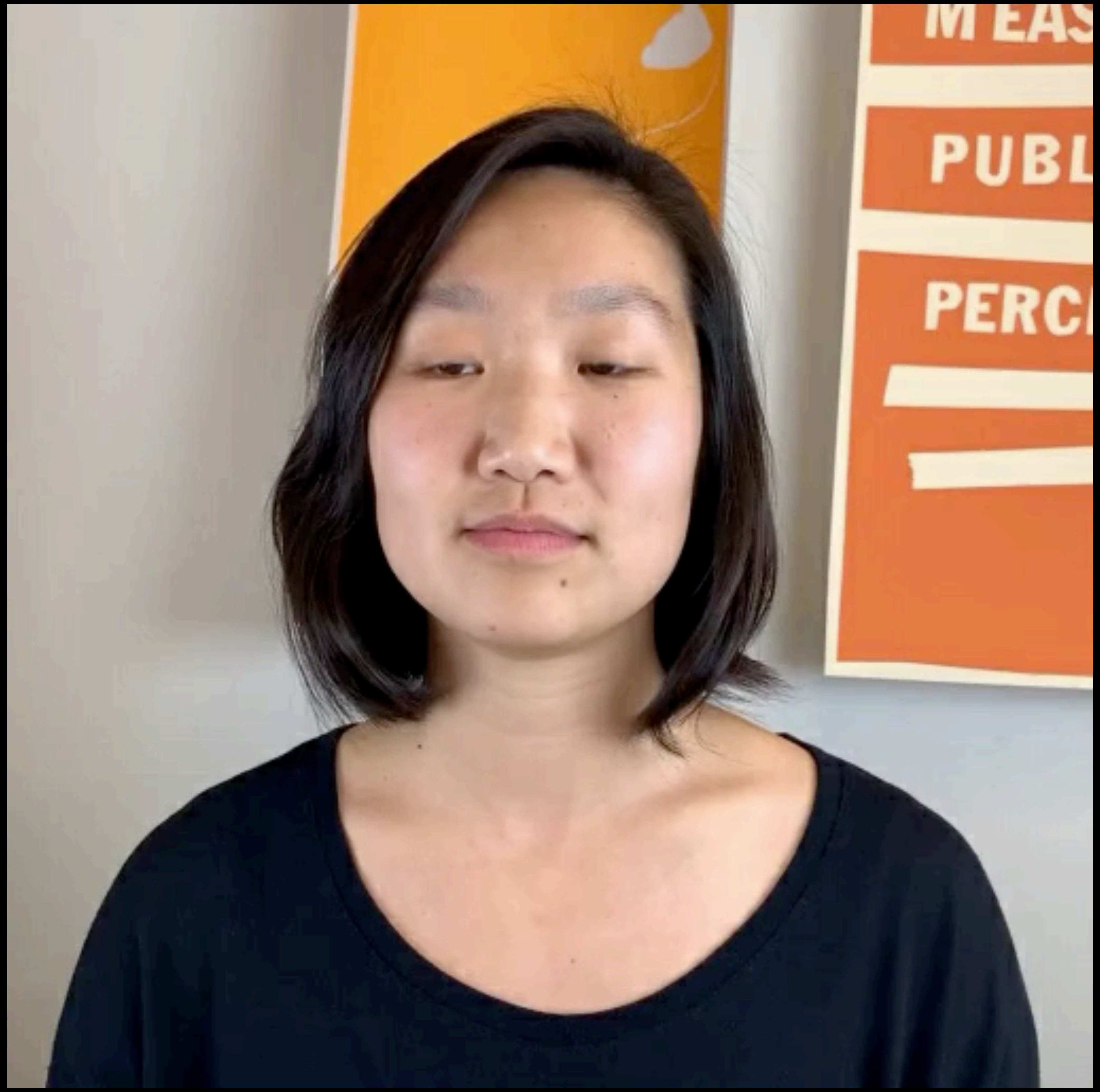
100% data



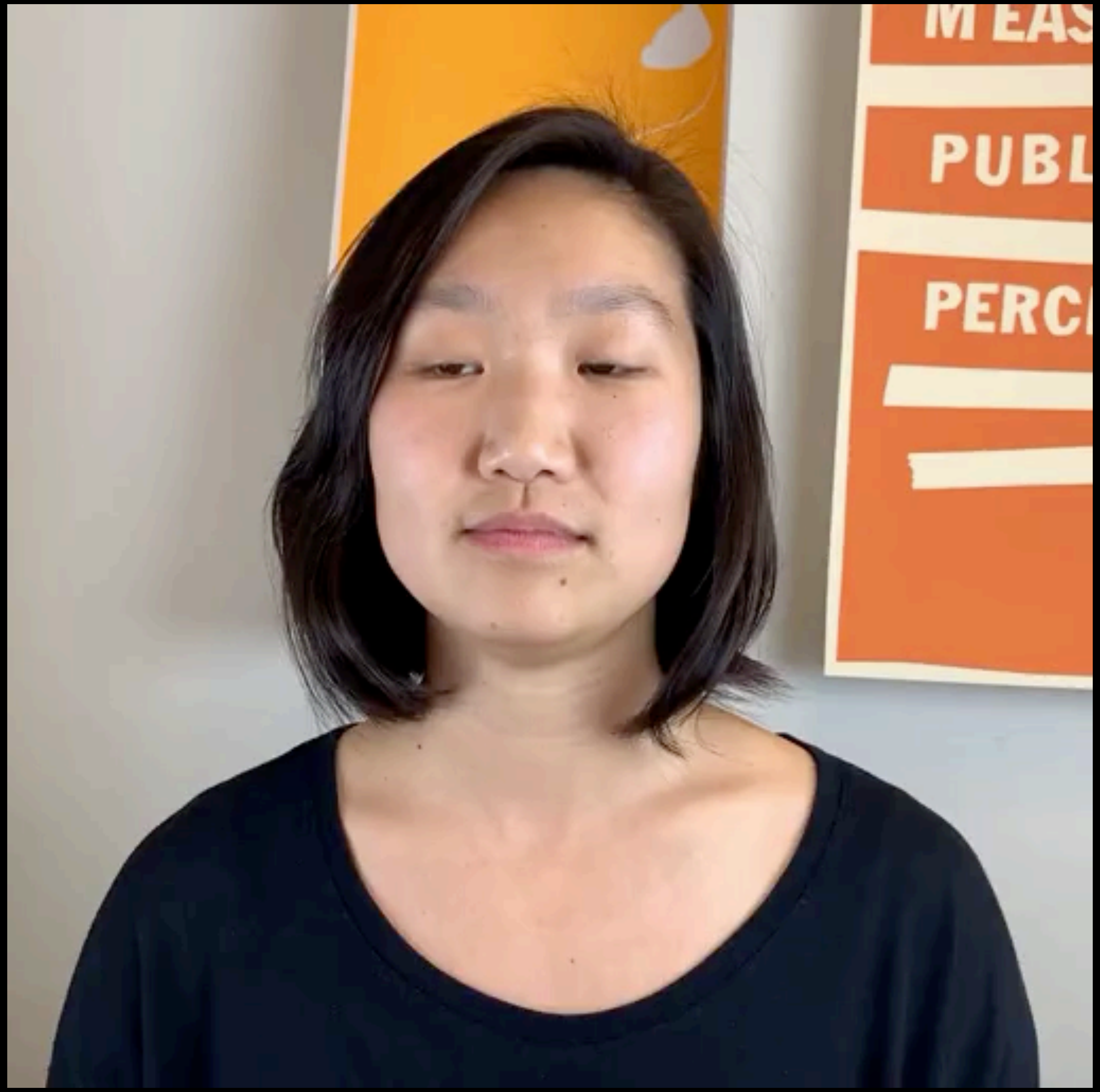
Ground truth



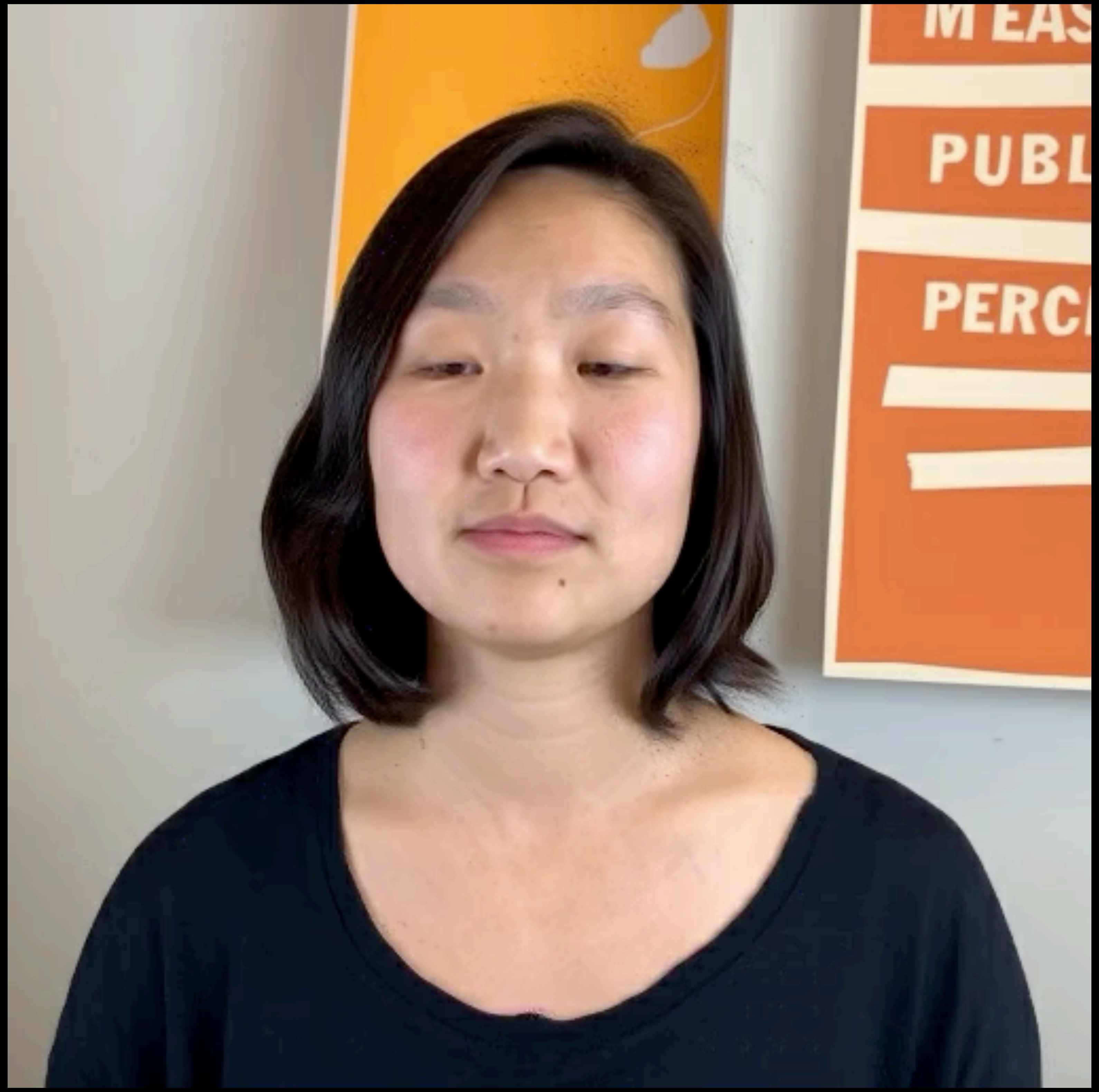
Deep Video Portraits
[Kim et al. ’18]



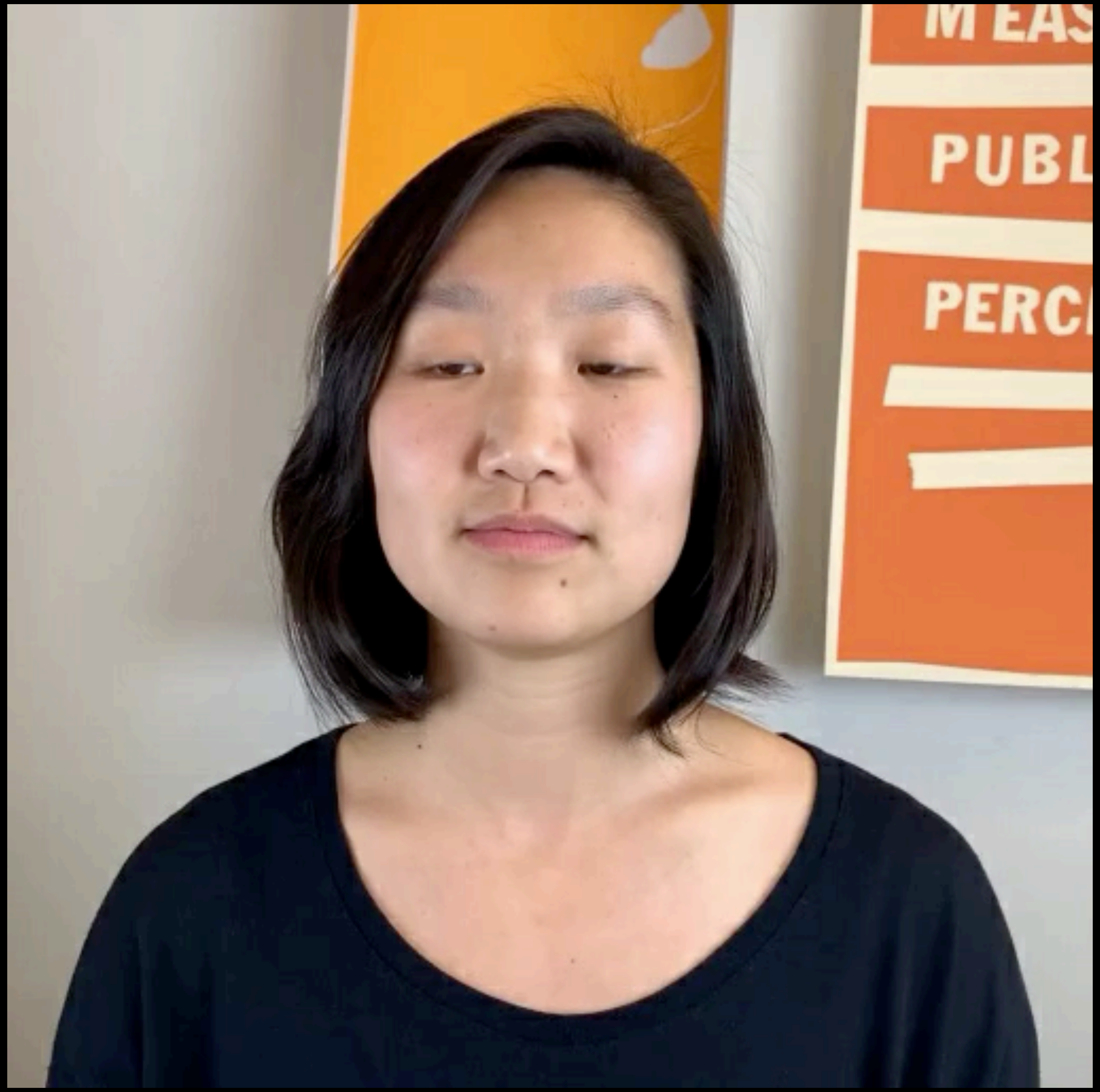
Ours



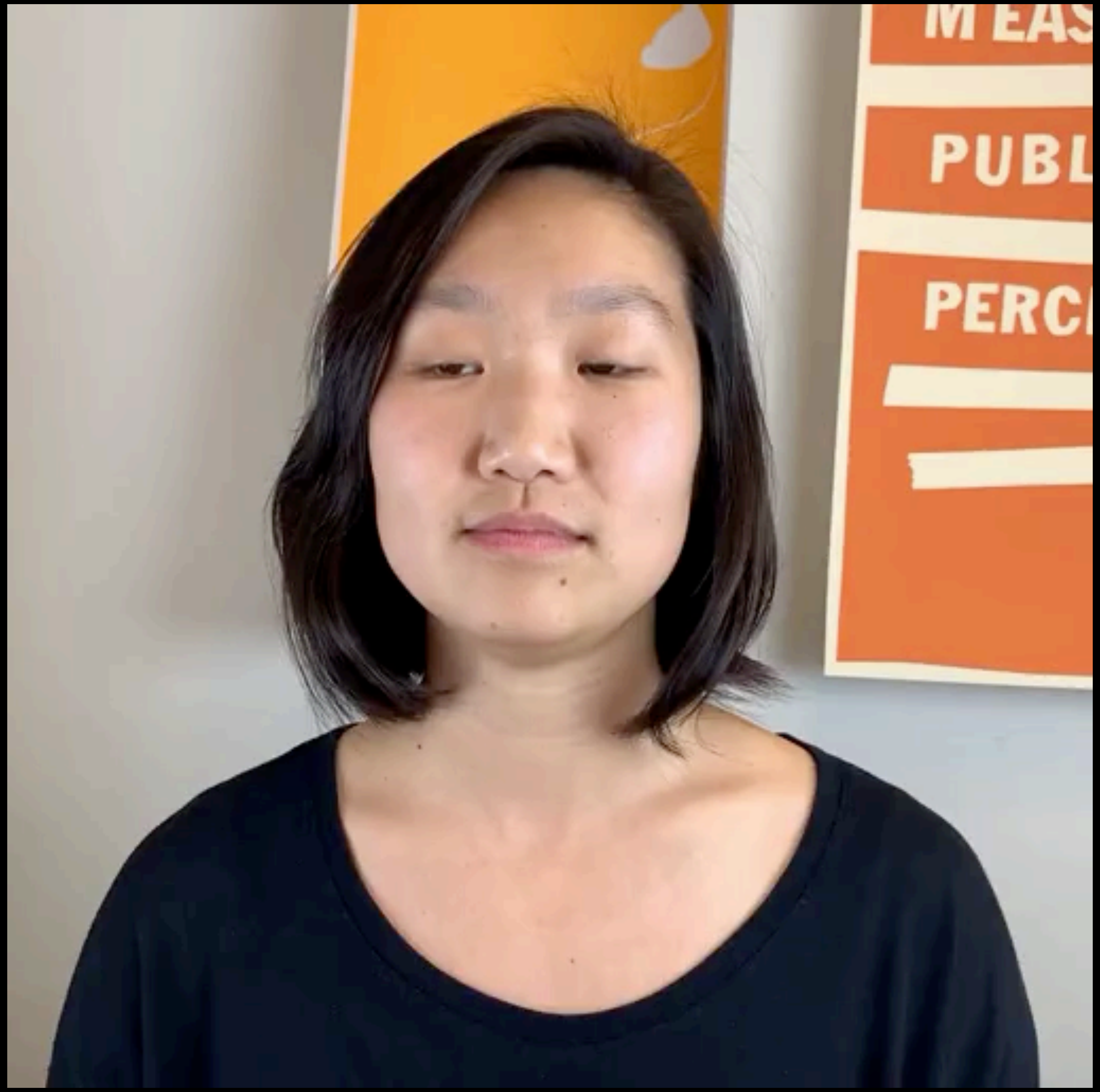
Ground truth



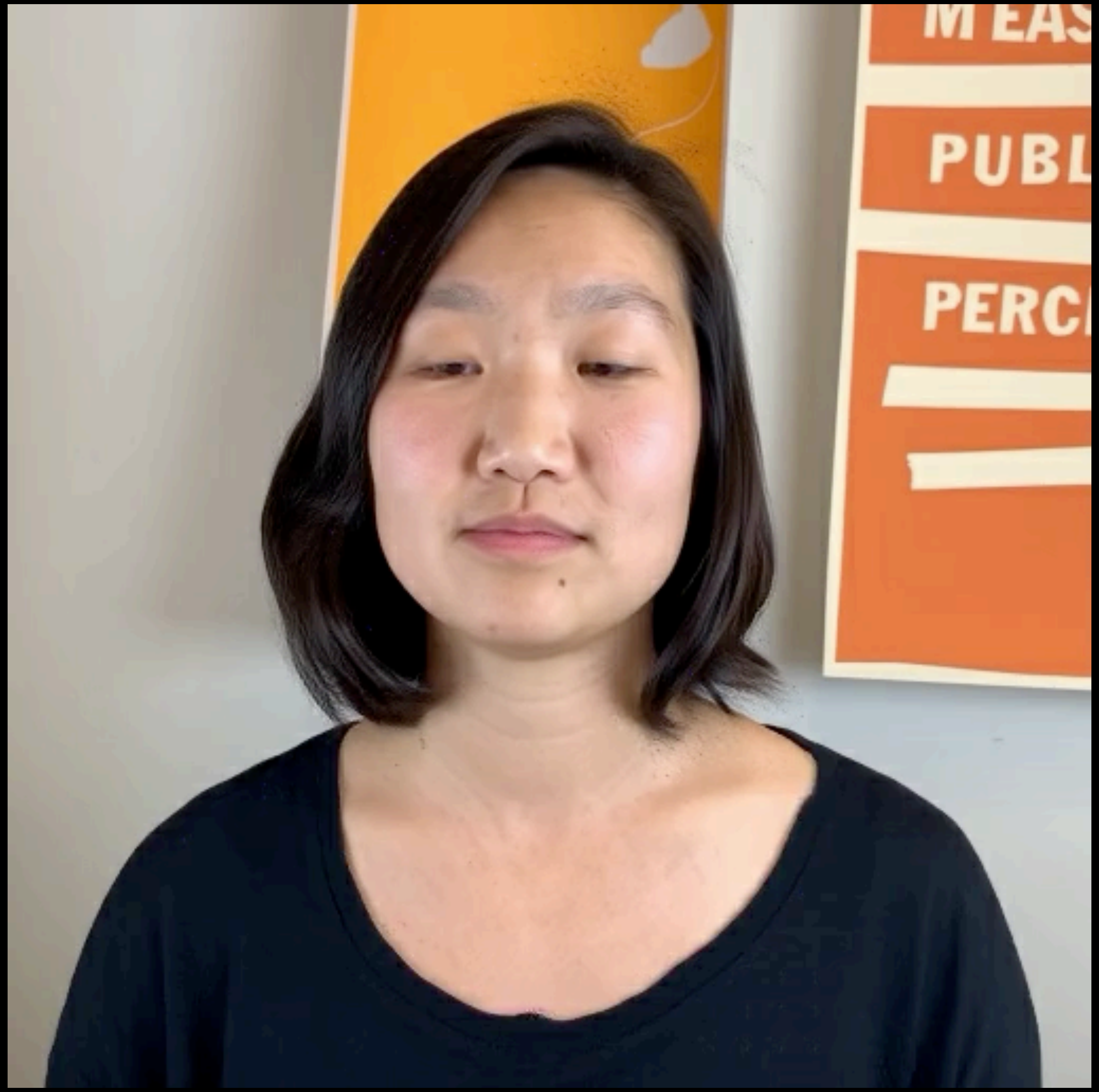
Deep Video Portraits
[Kim et al. ’18]



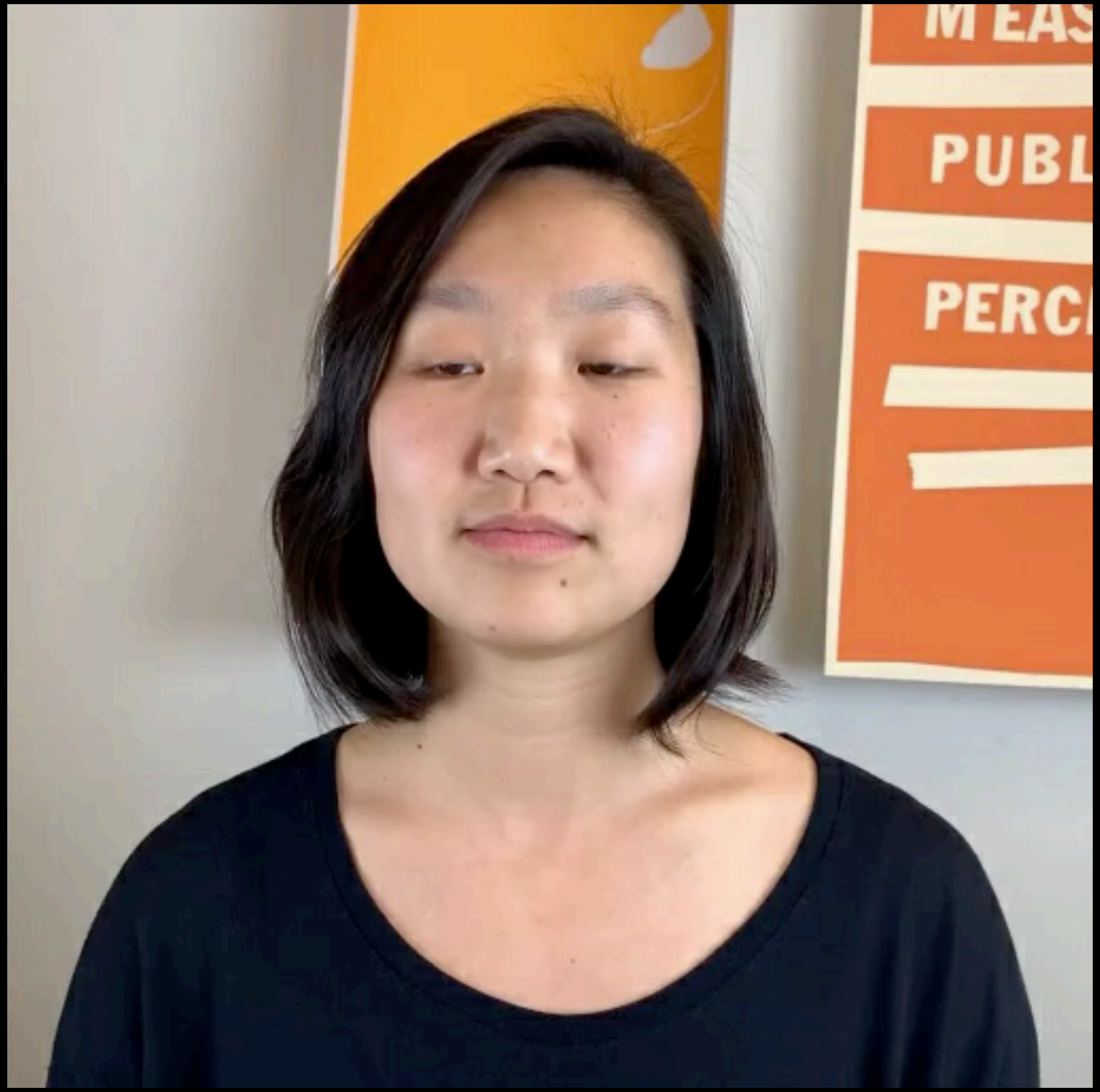
Ours



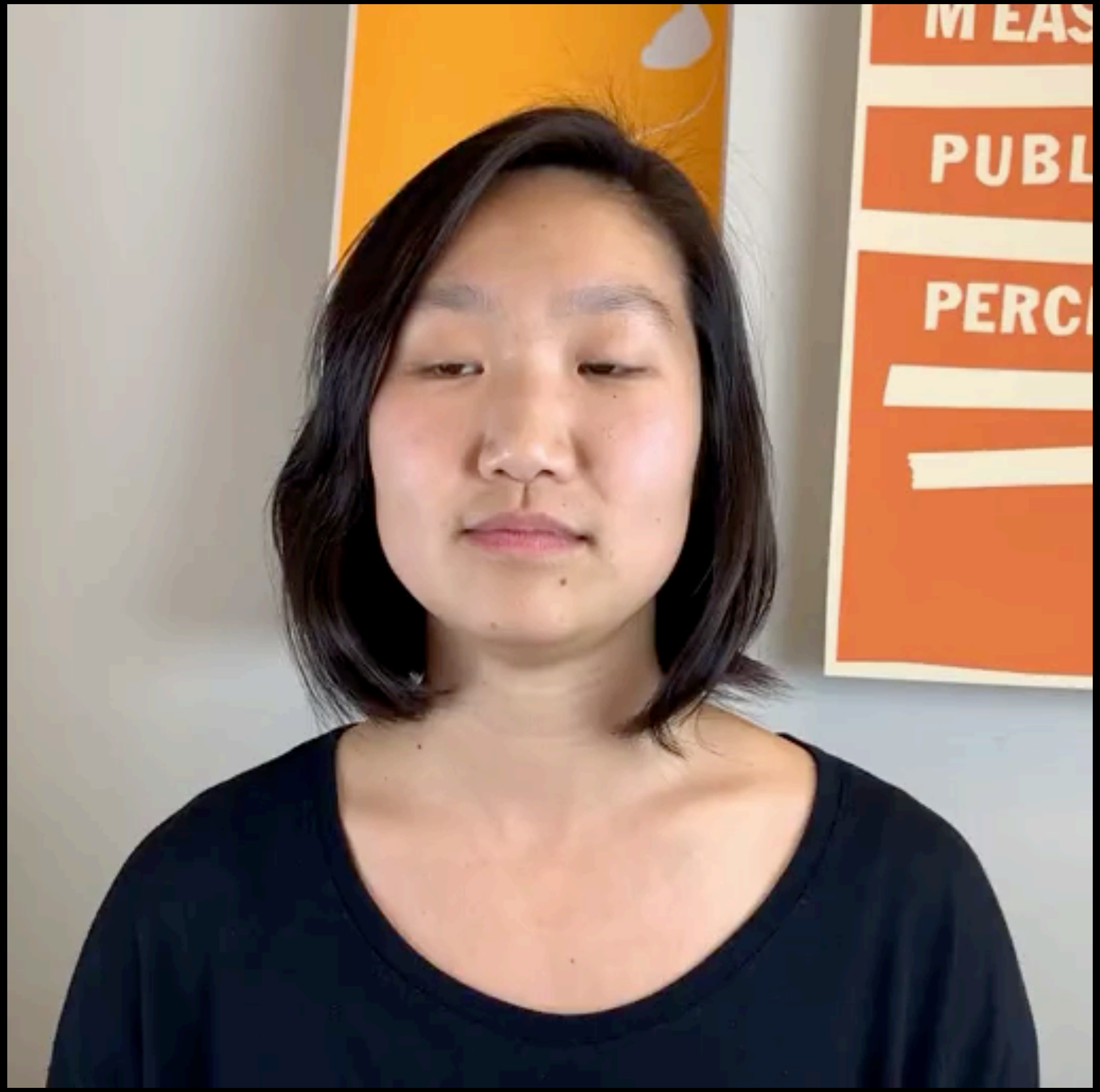
Ground truth



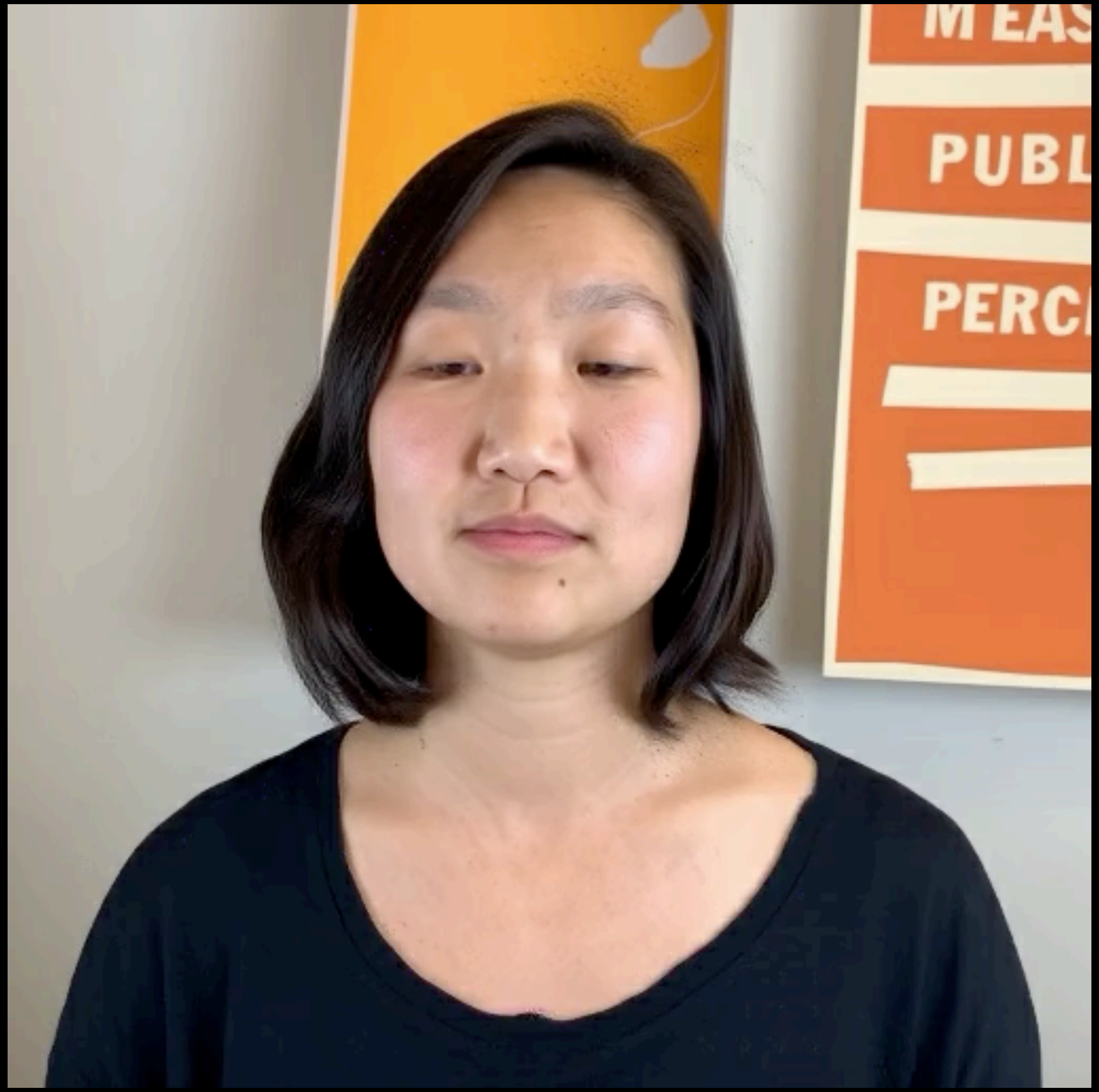
Deep Video Portraits
[Kim et al. ’18]



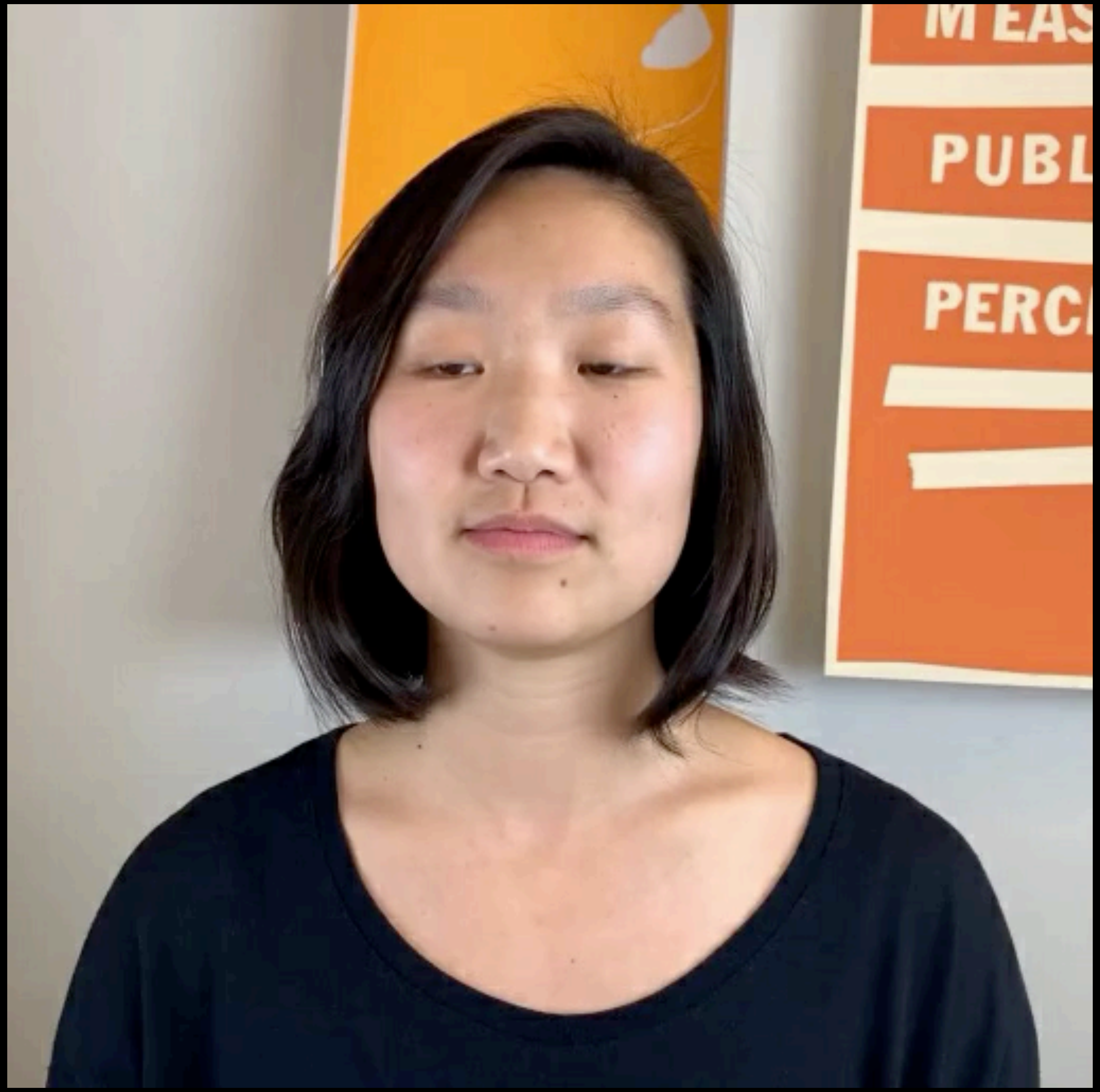
Ours



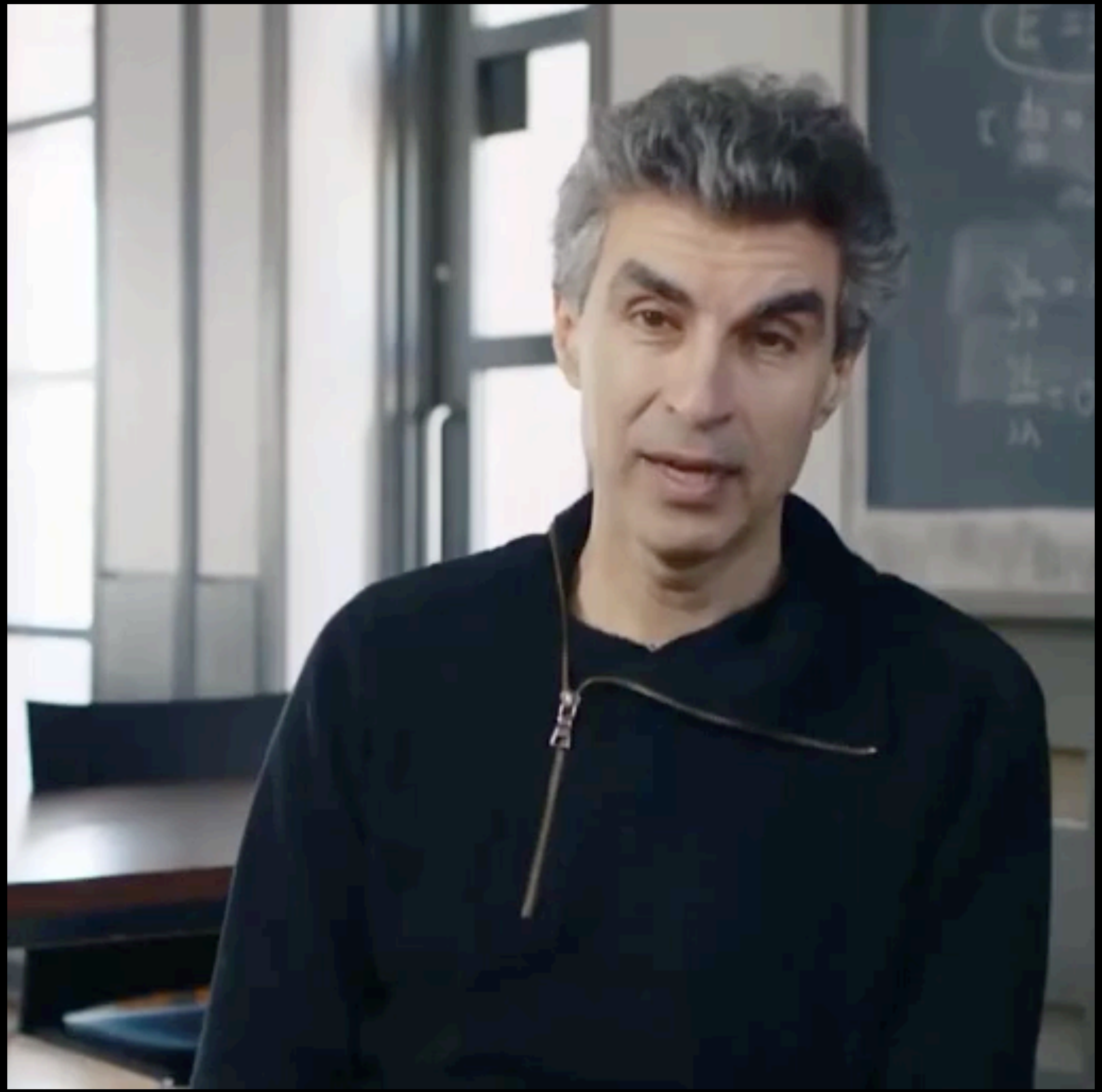
Ground truth



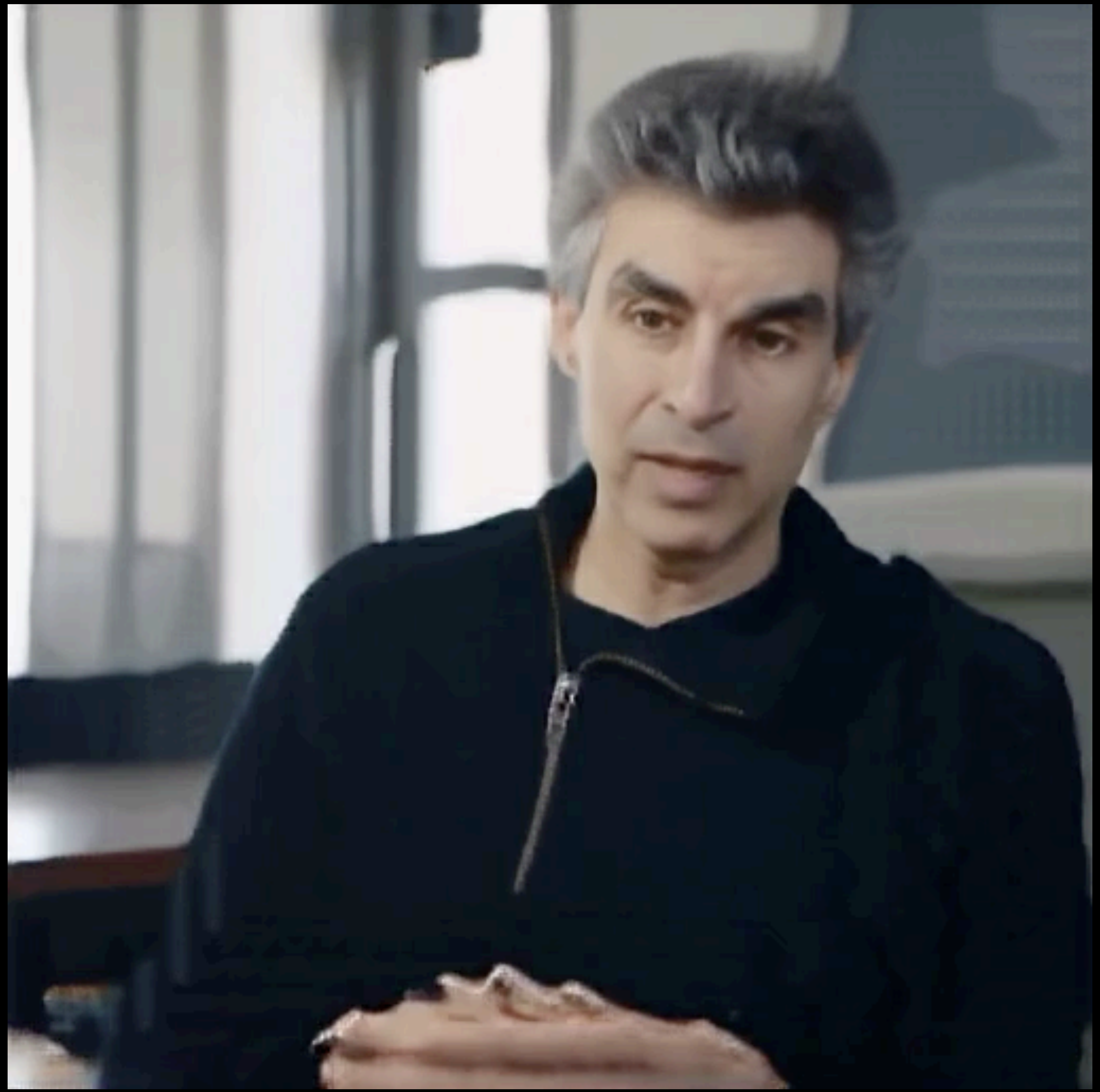
Deep Video Portraits
[Kim et al. ’18]



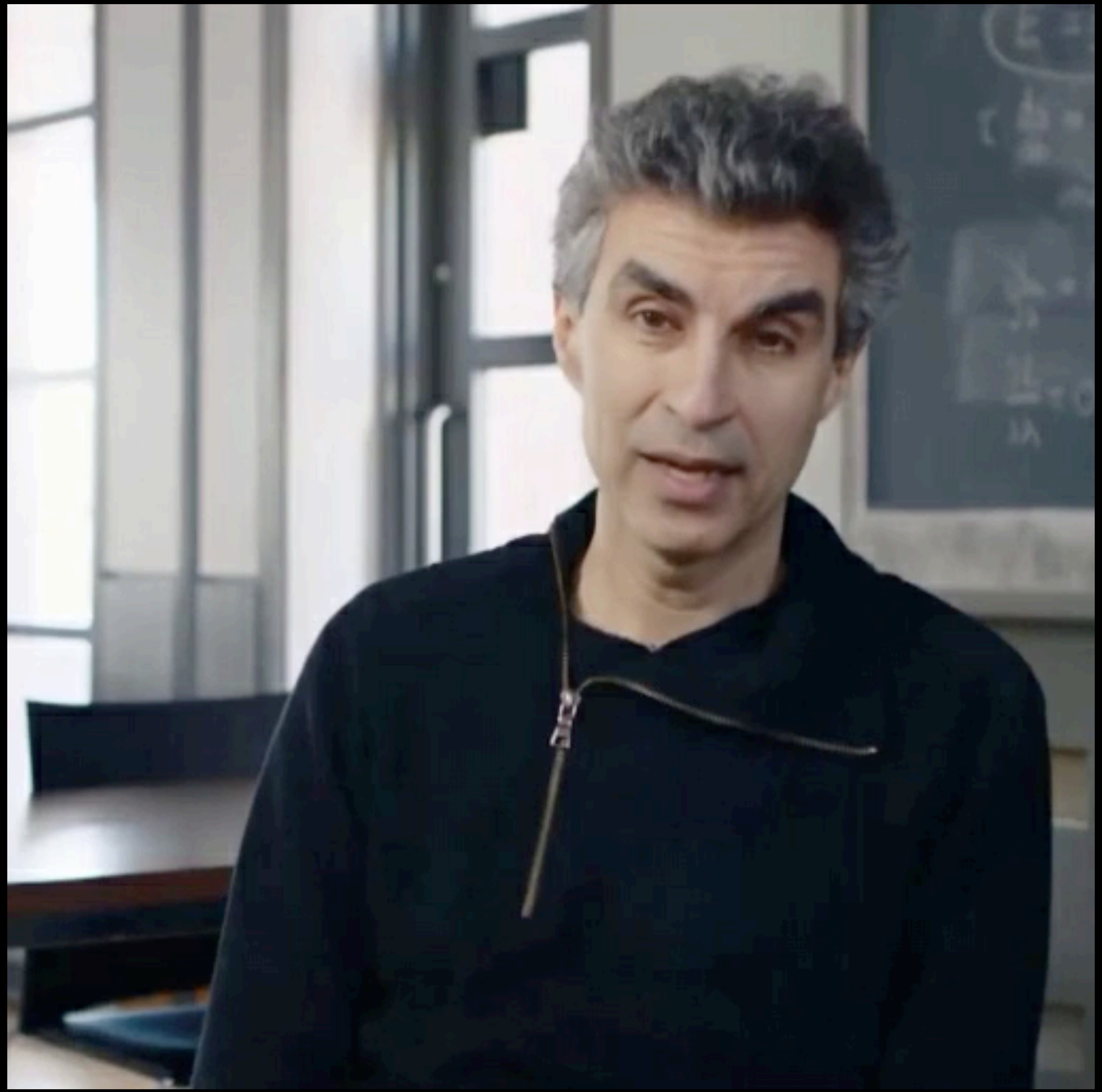
Ours



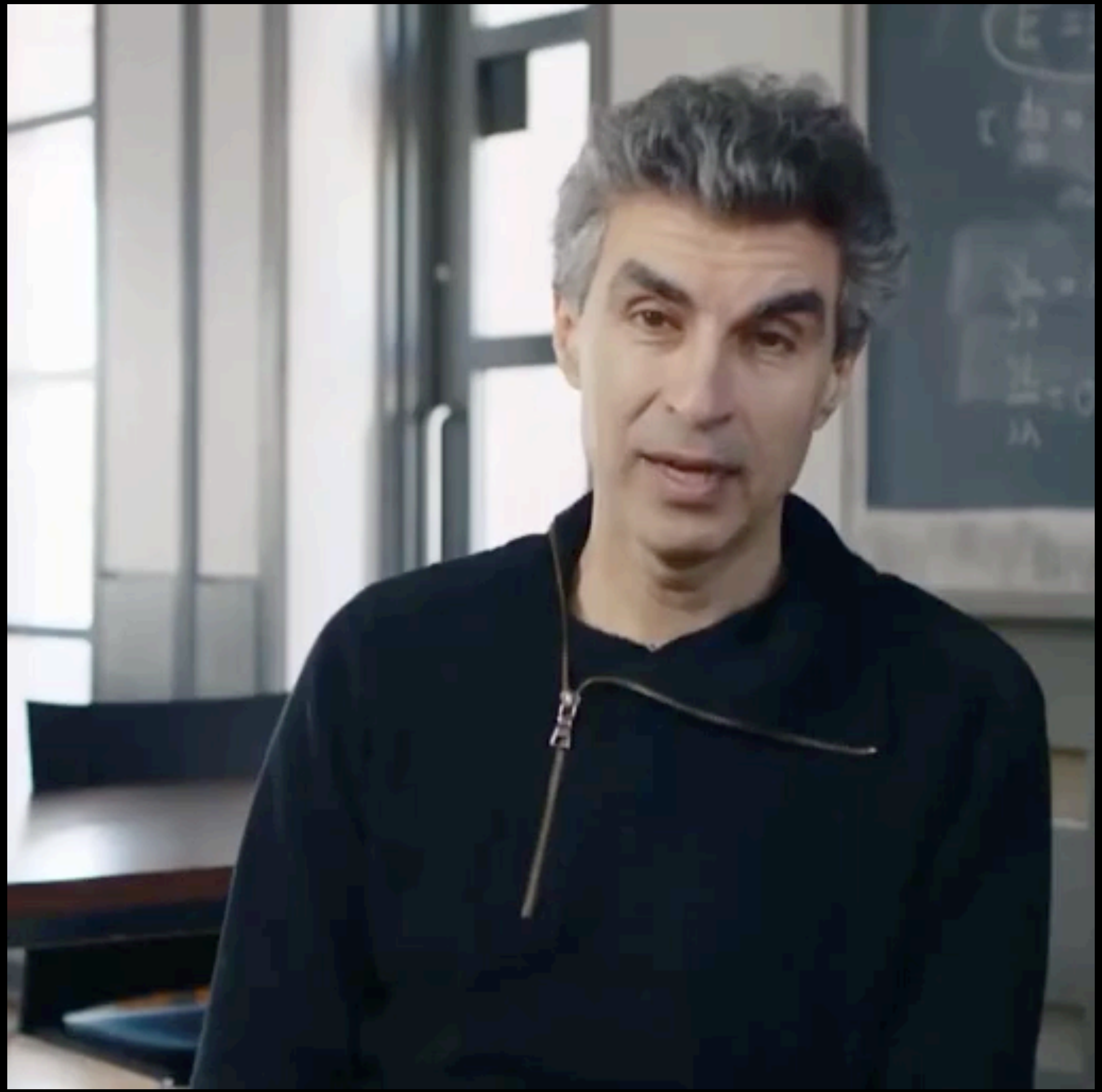
Ground truth



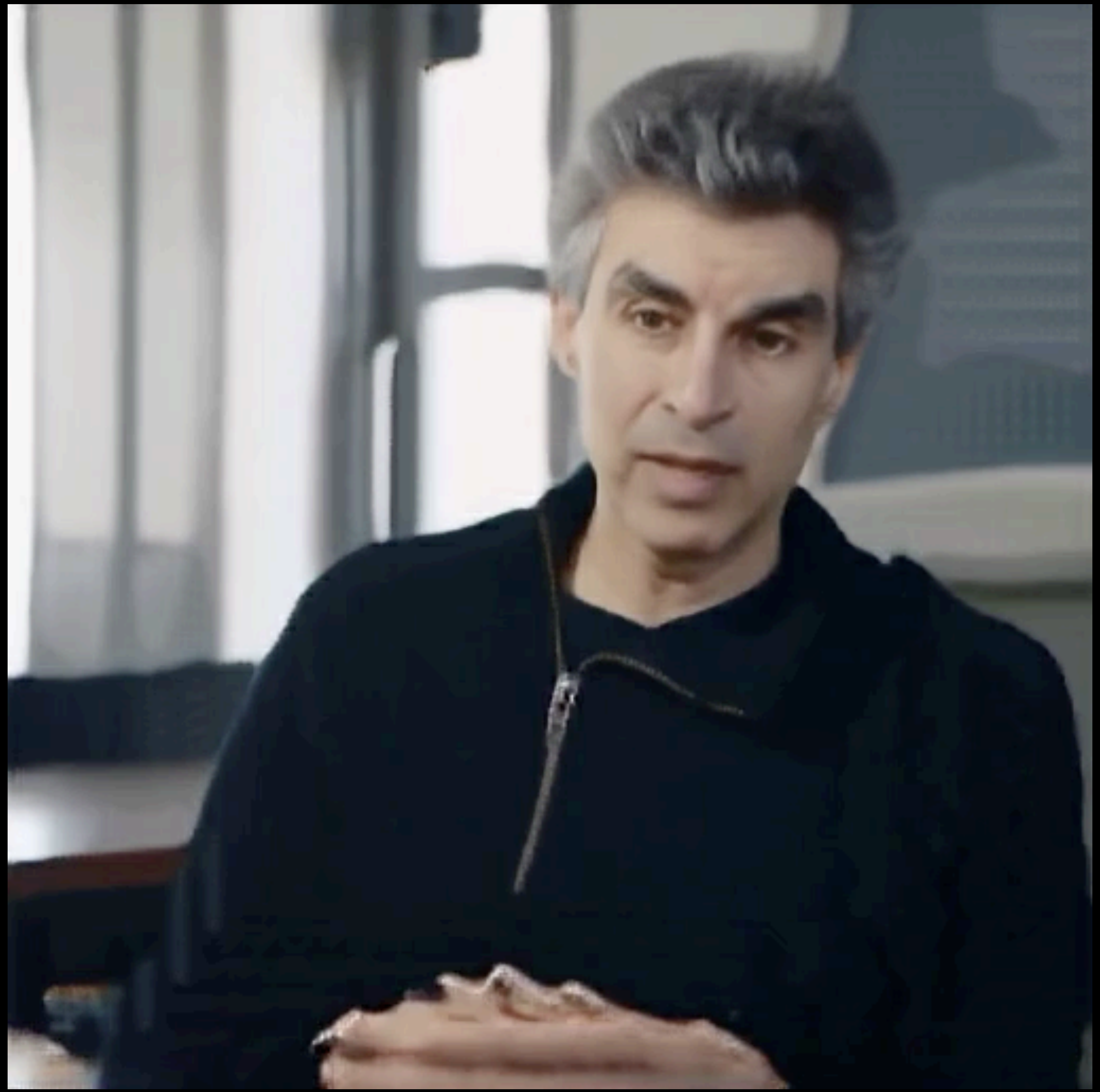
Deep Video Portraits
[Kim et al. ’18]



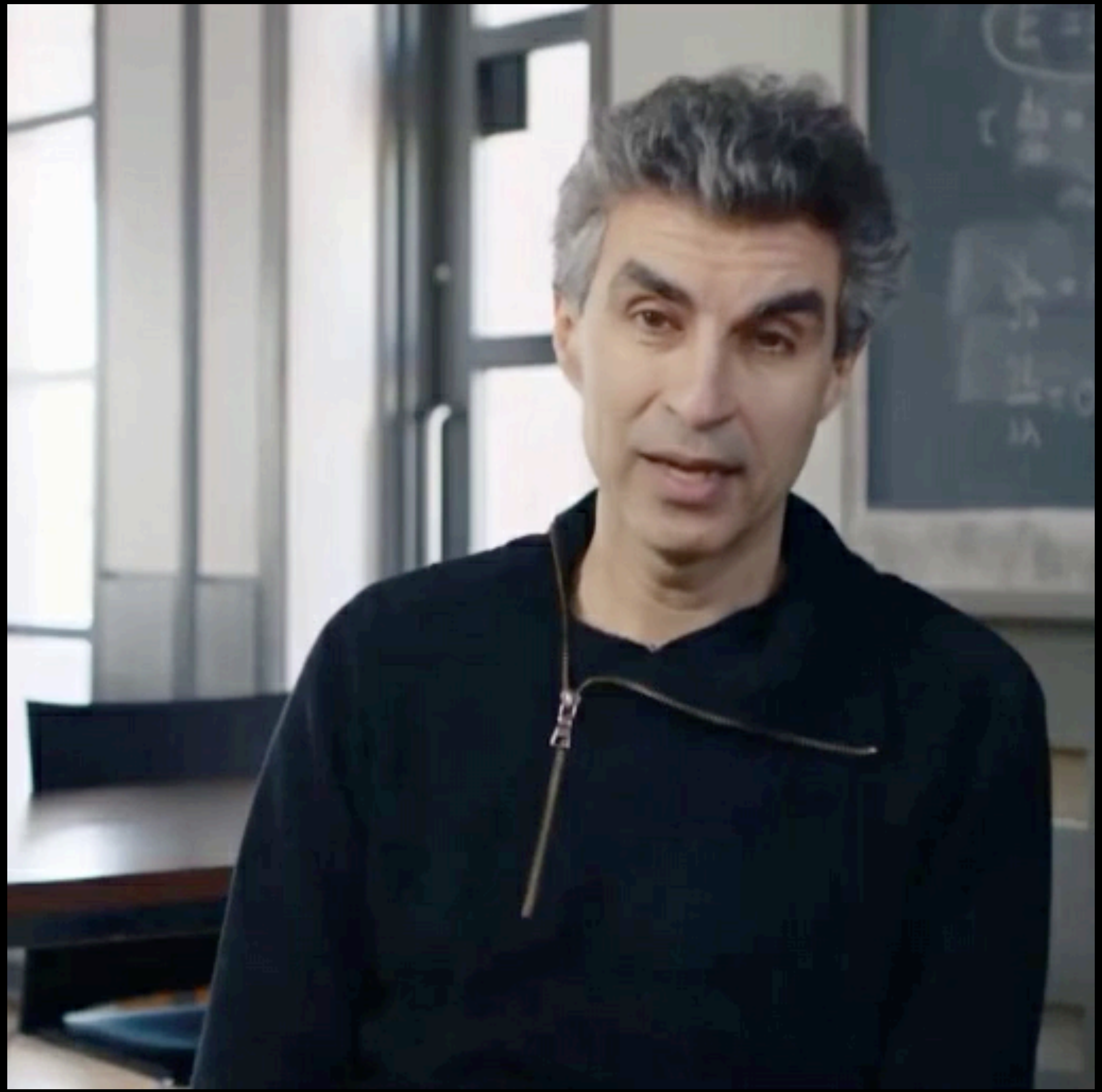
Ours



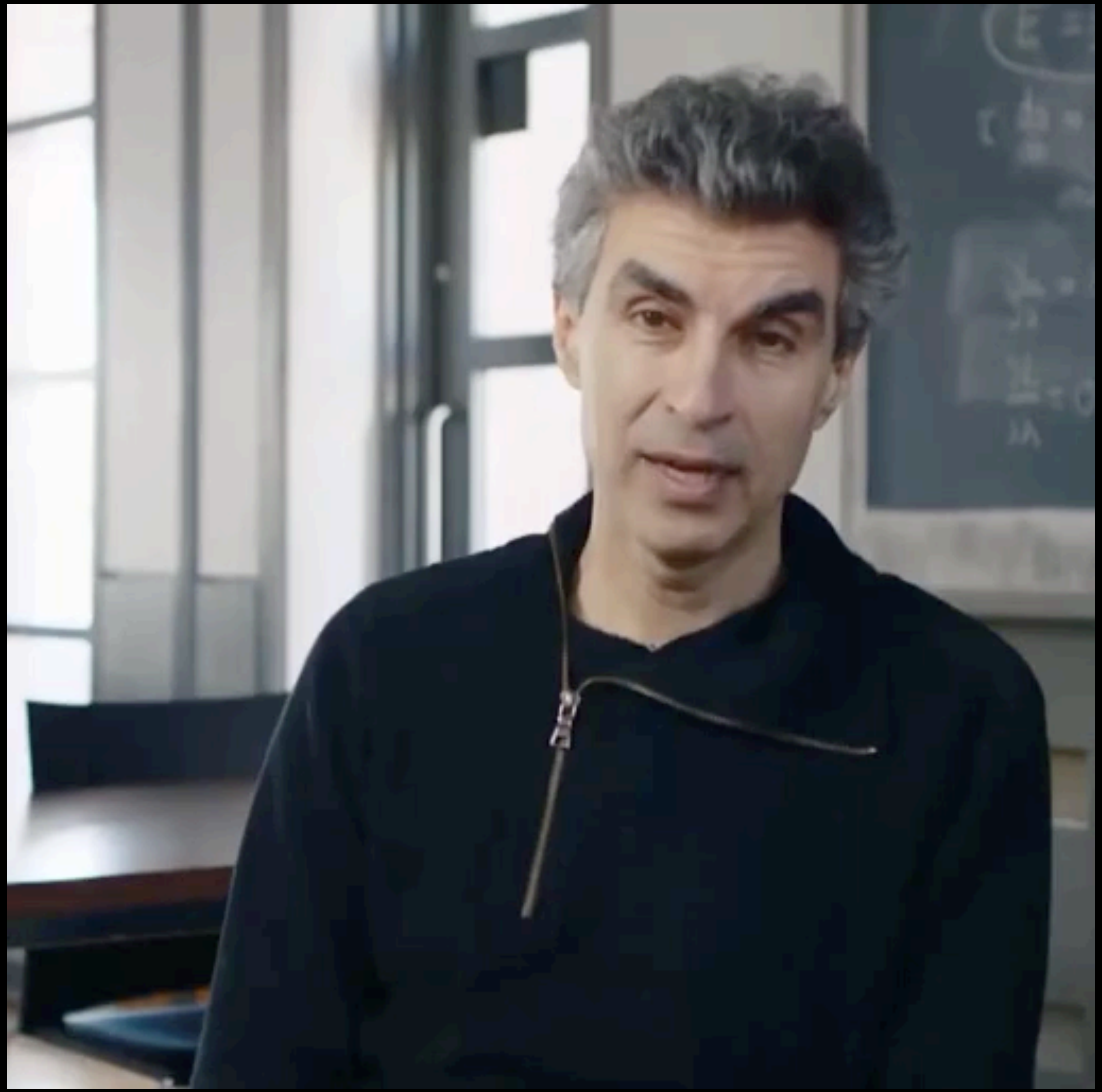
Ground truth



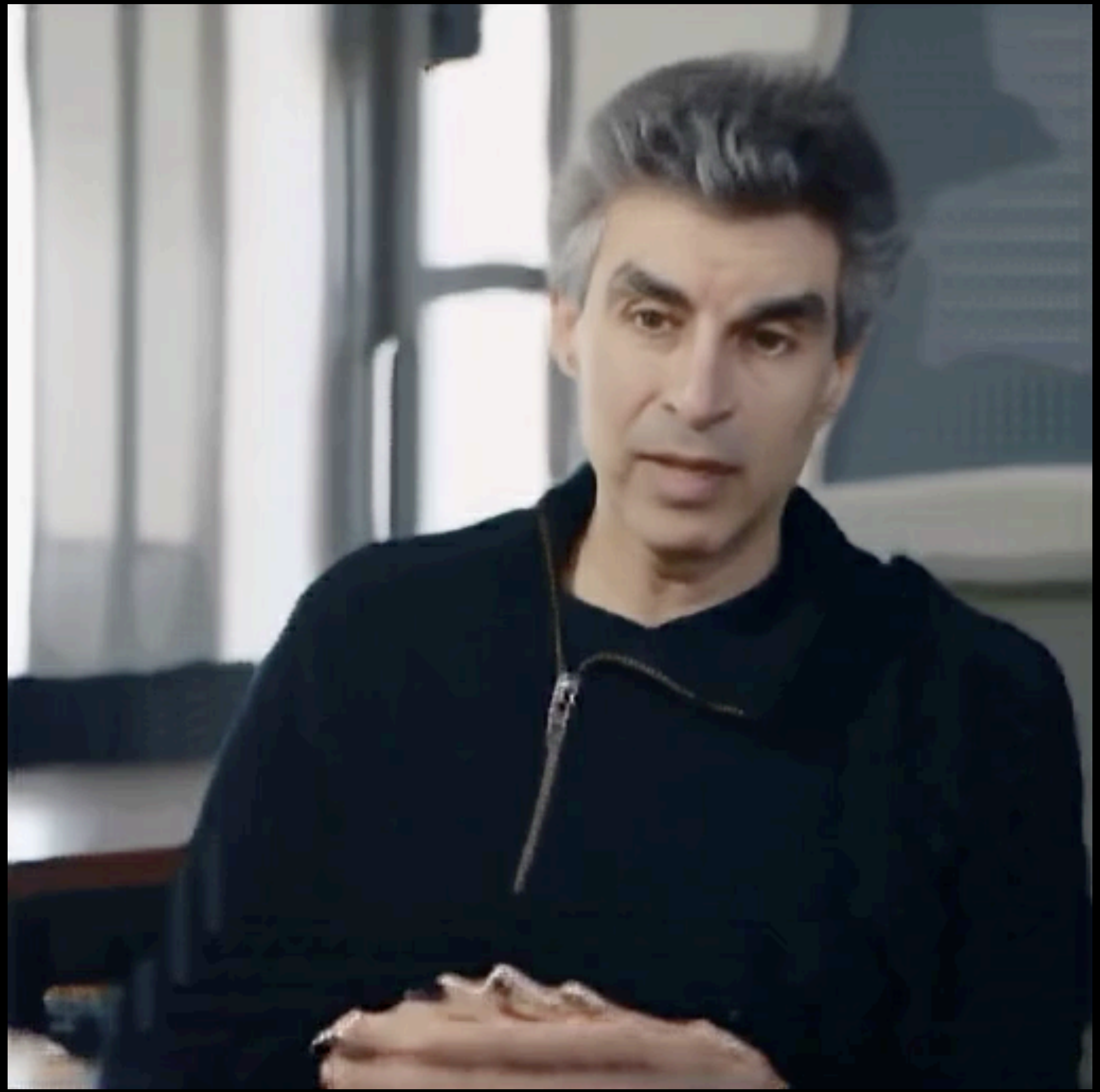
Deep Video Portraits
[Kim et al. ’18]



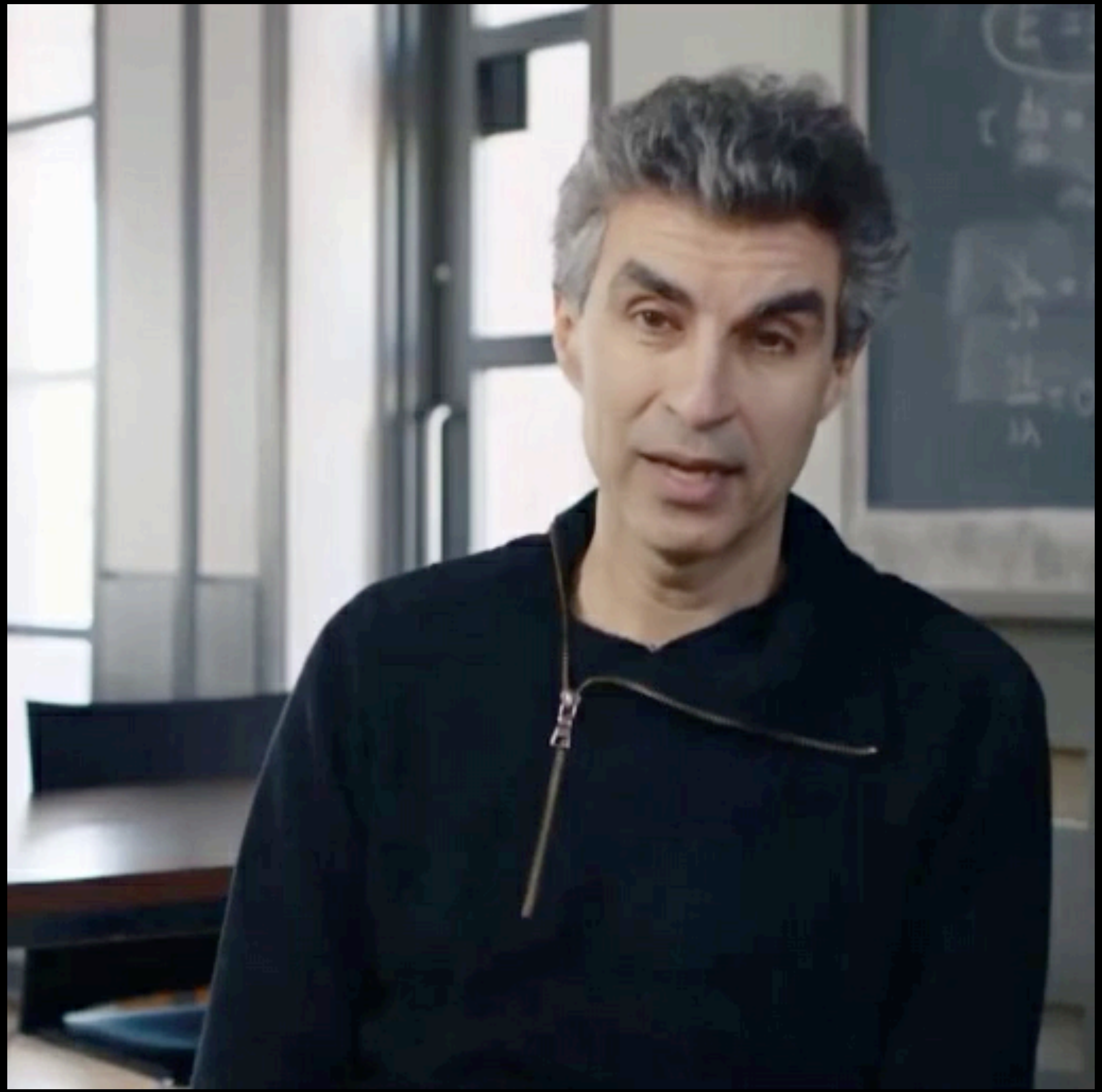
Ours



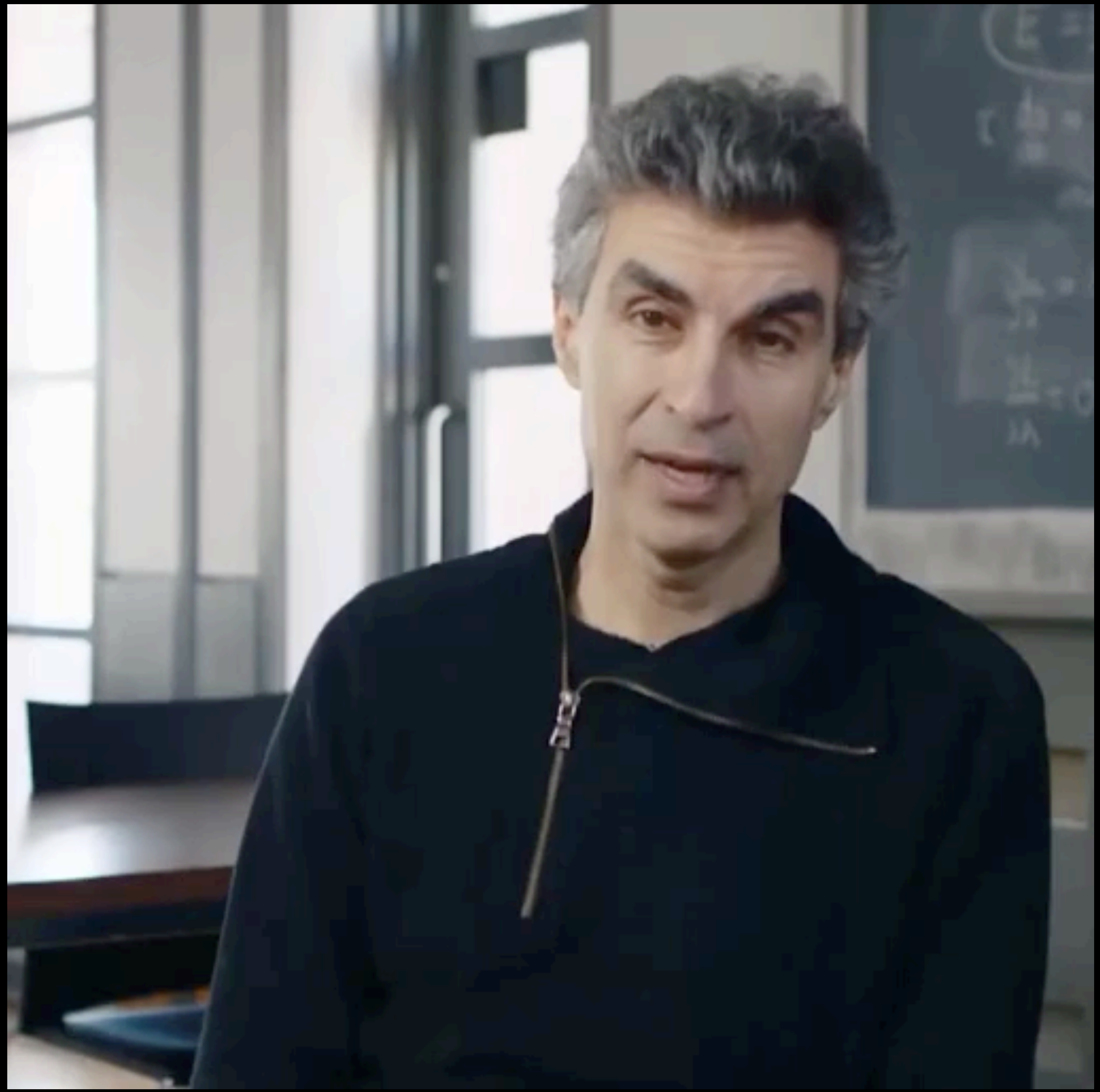
Ground truth



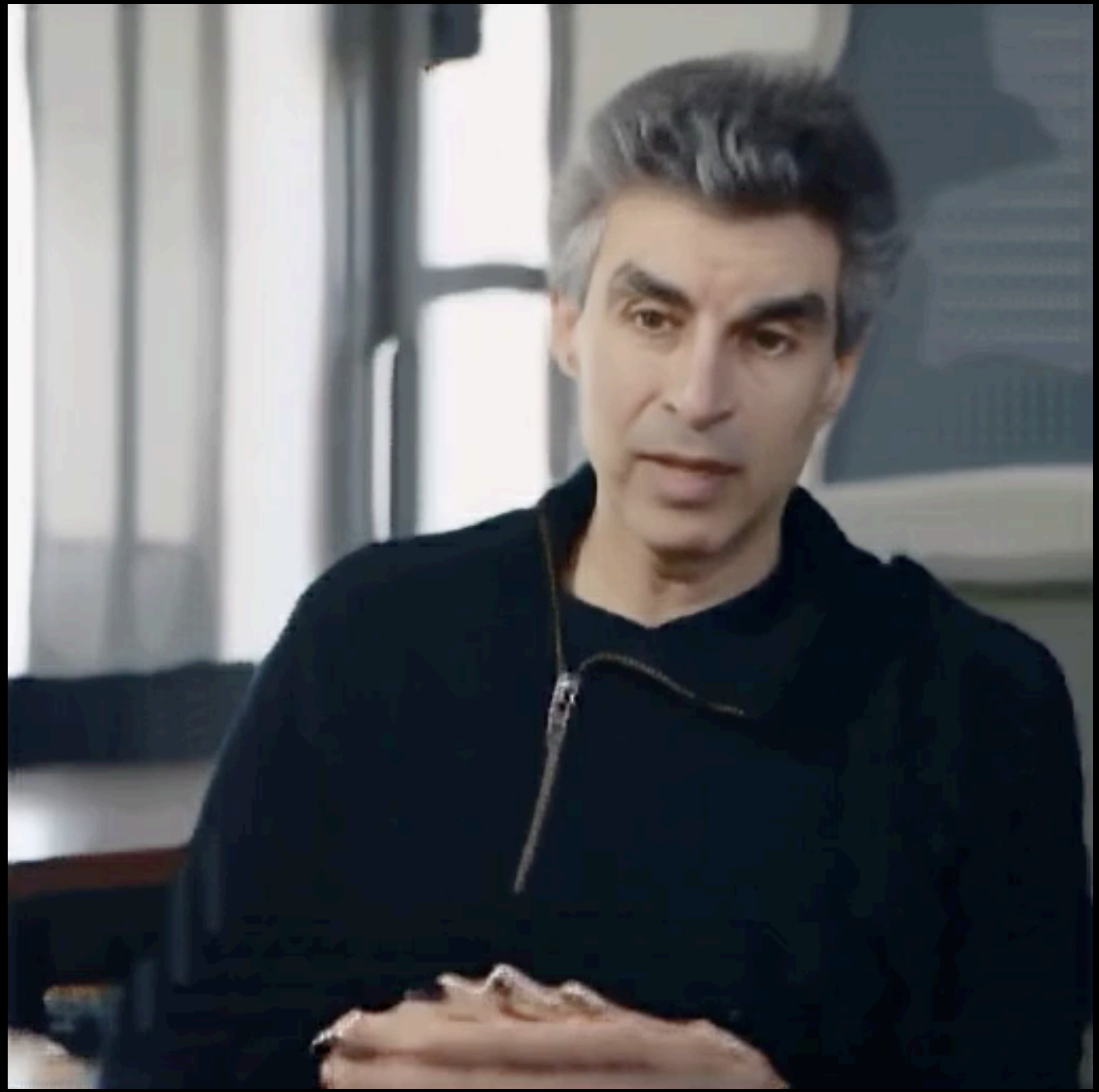
Deep Video Portraits
[Kim et al. ’18]



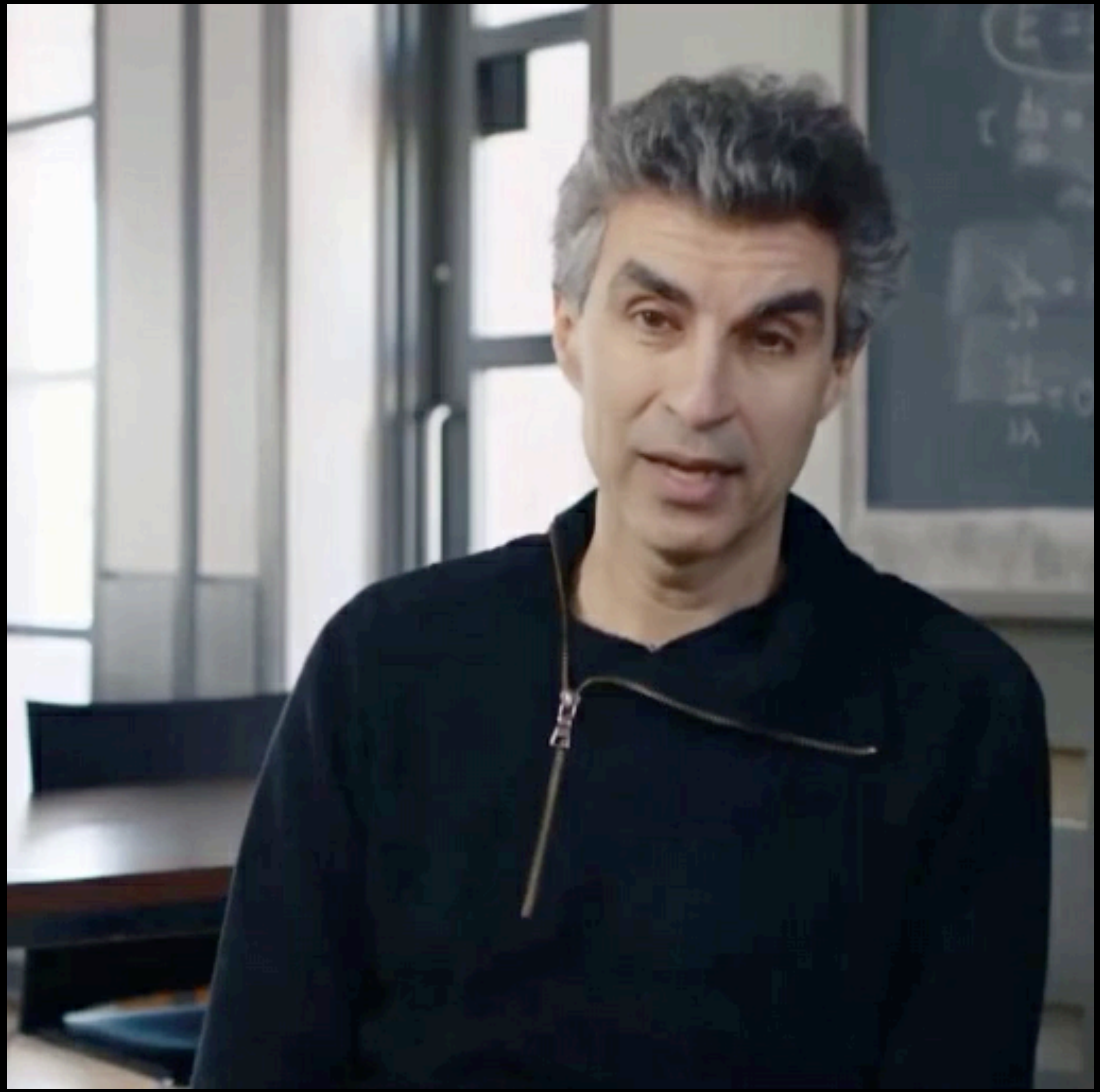
Ours



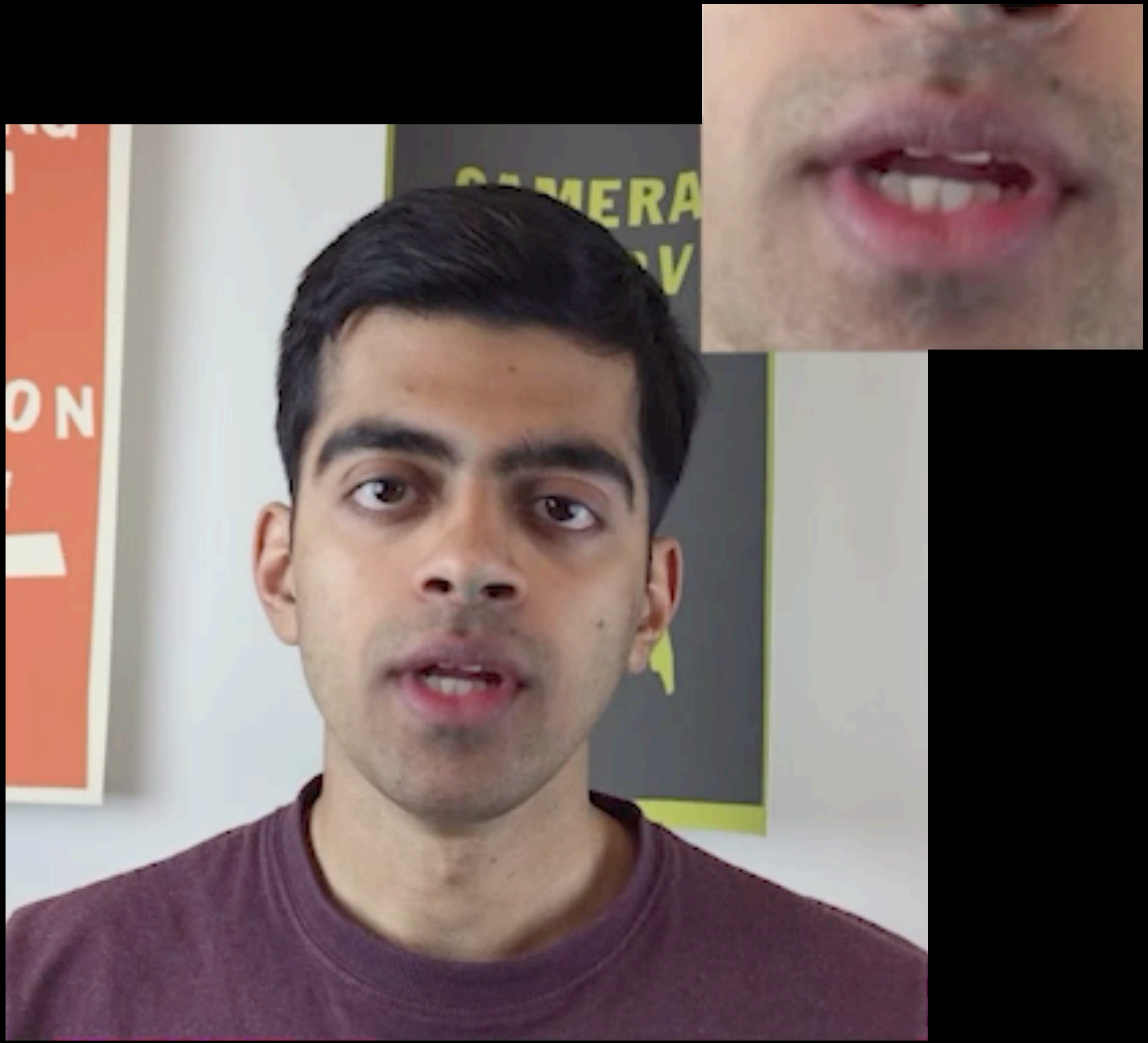
Ground truth



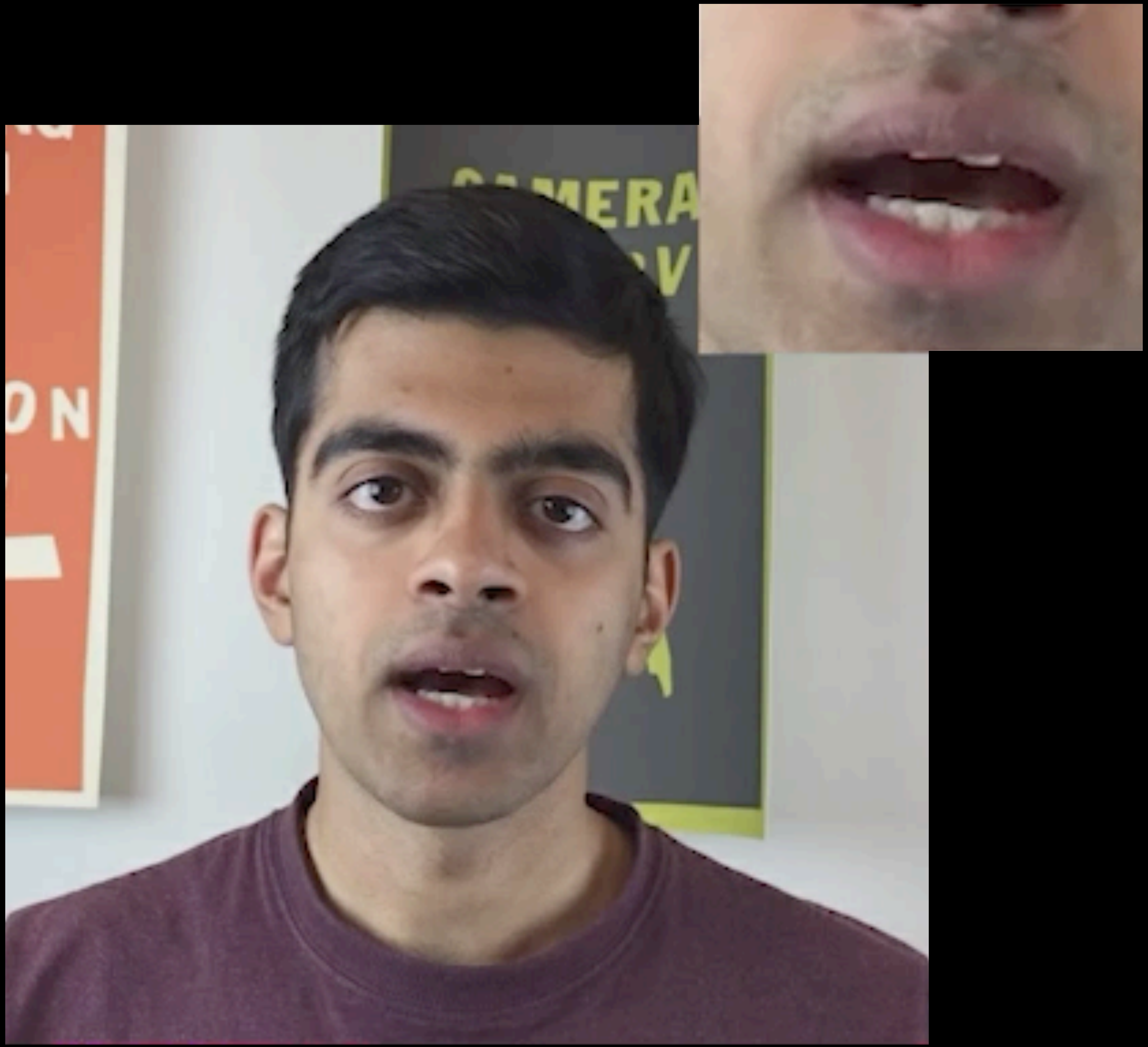
Deep Video Portraits
[Kim et al. ’18]



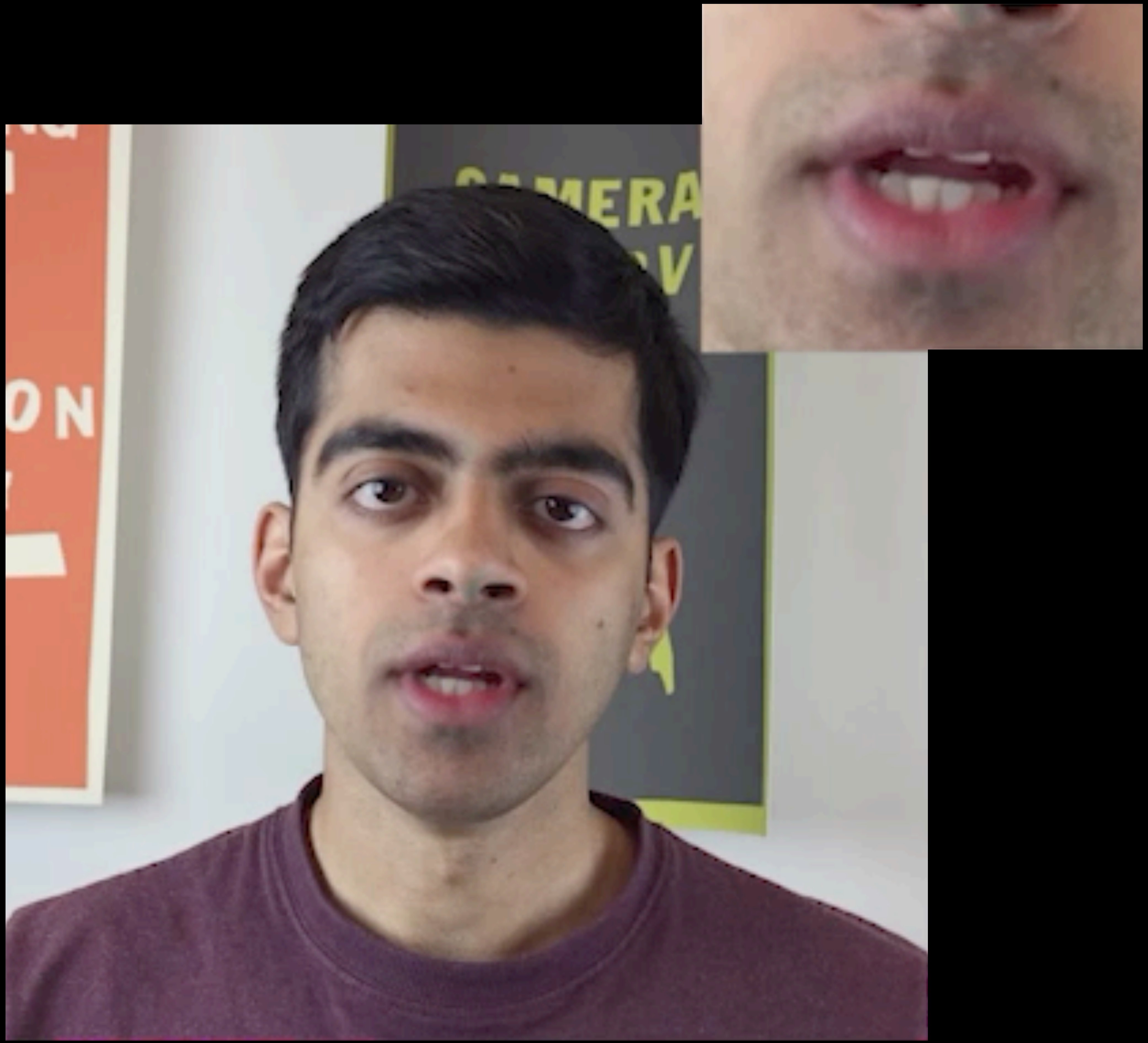
Ours



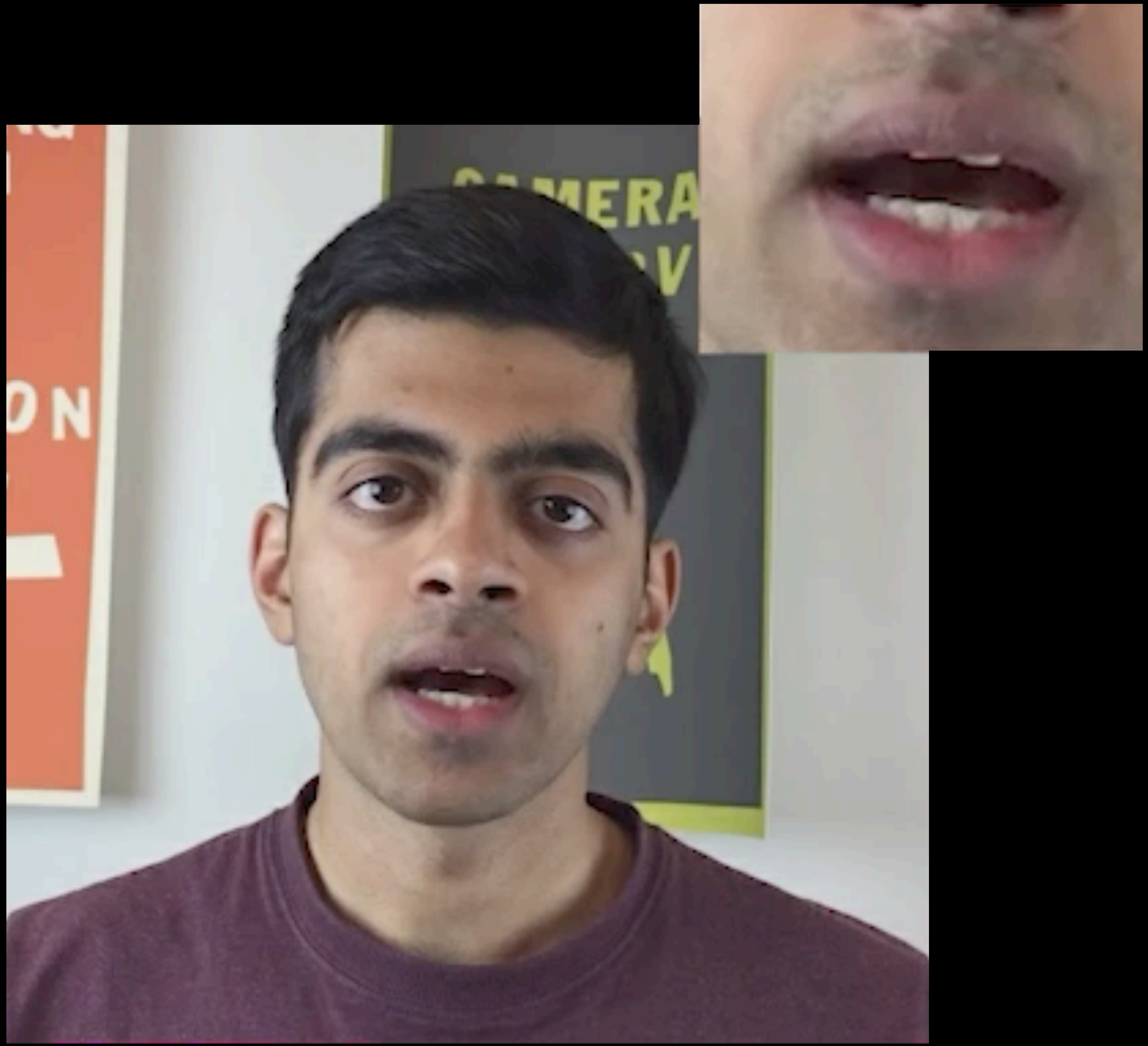
Face2Face
[Thies et al. 18]



Ours



Face2Face
[Thies et al. 18]



Ours

Morph Cut
[Berthouzoz et al. 12]

Ours

“It’s one of the approaches ~~to~~ to machine learning”

Morph Cut
[Berthouzoz et al. 12]

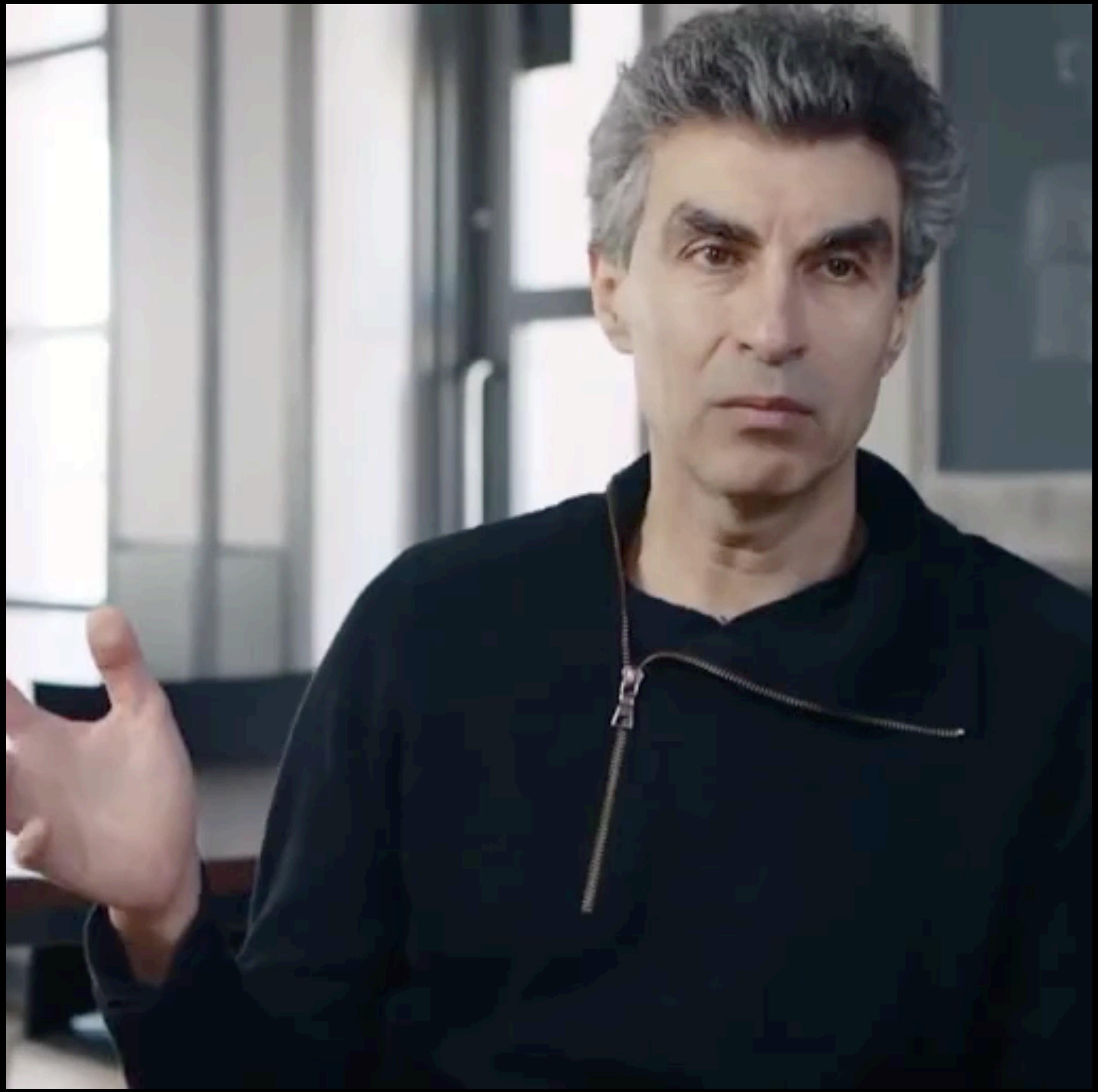
Ours

“It’s one of the approaches ~~to~~ to machine learning”

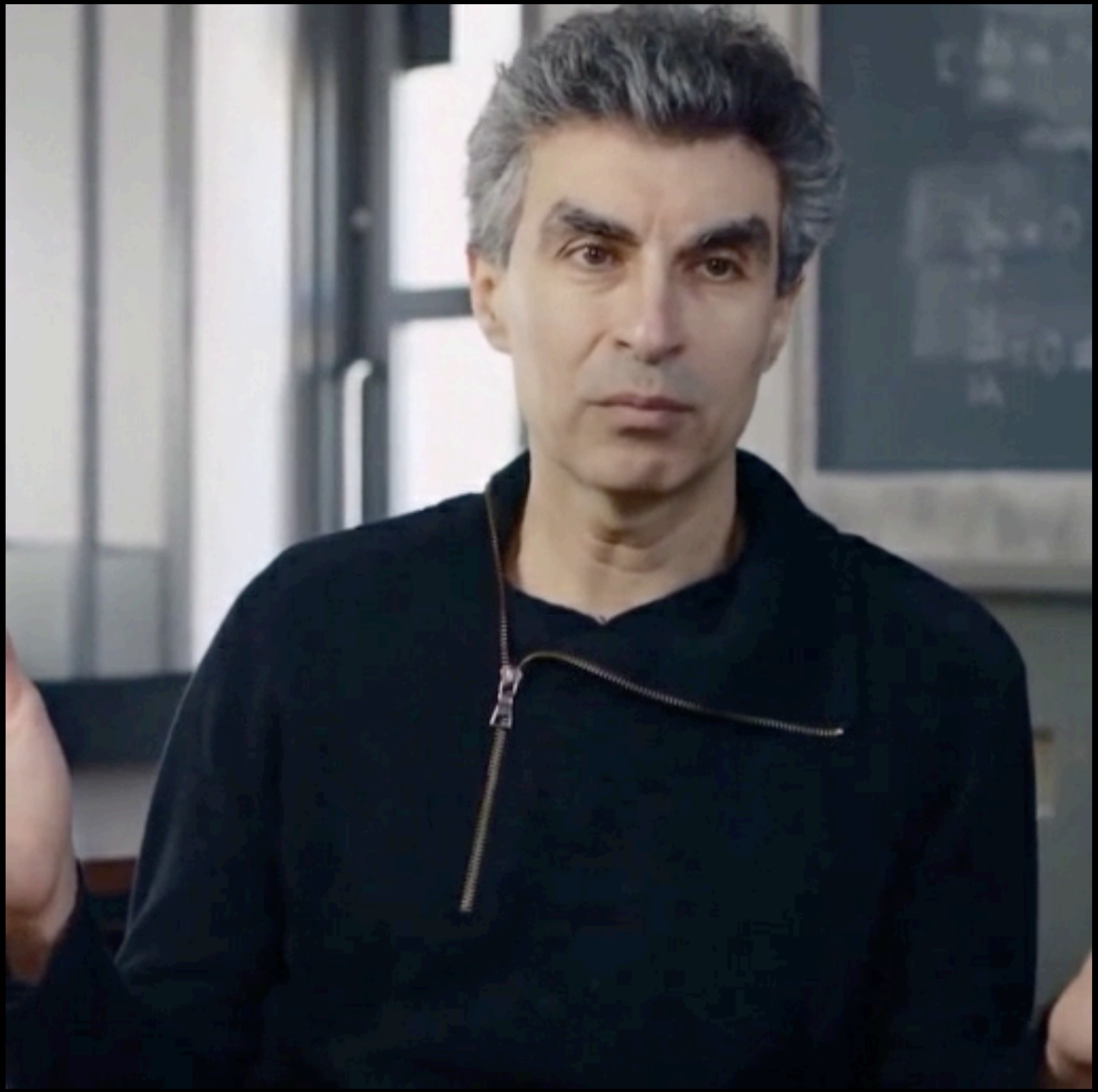
Morph Cut
[Berthouzoz et al. 12]

Ours

“It’s one of the approaches ~~to~~ to machine learning”

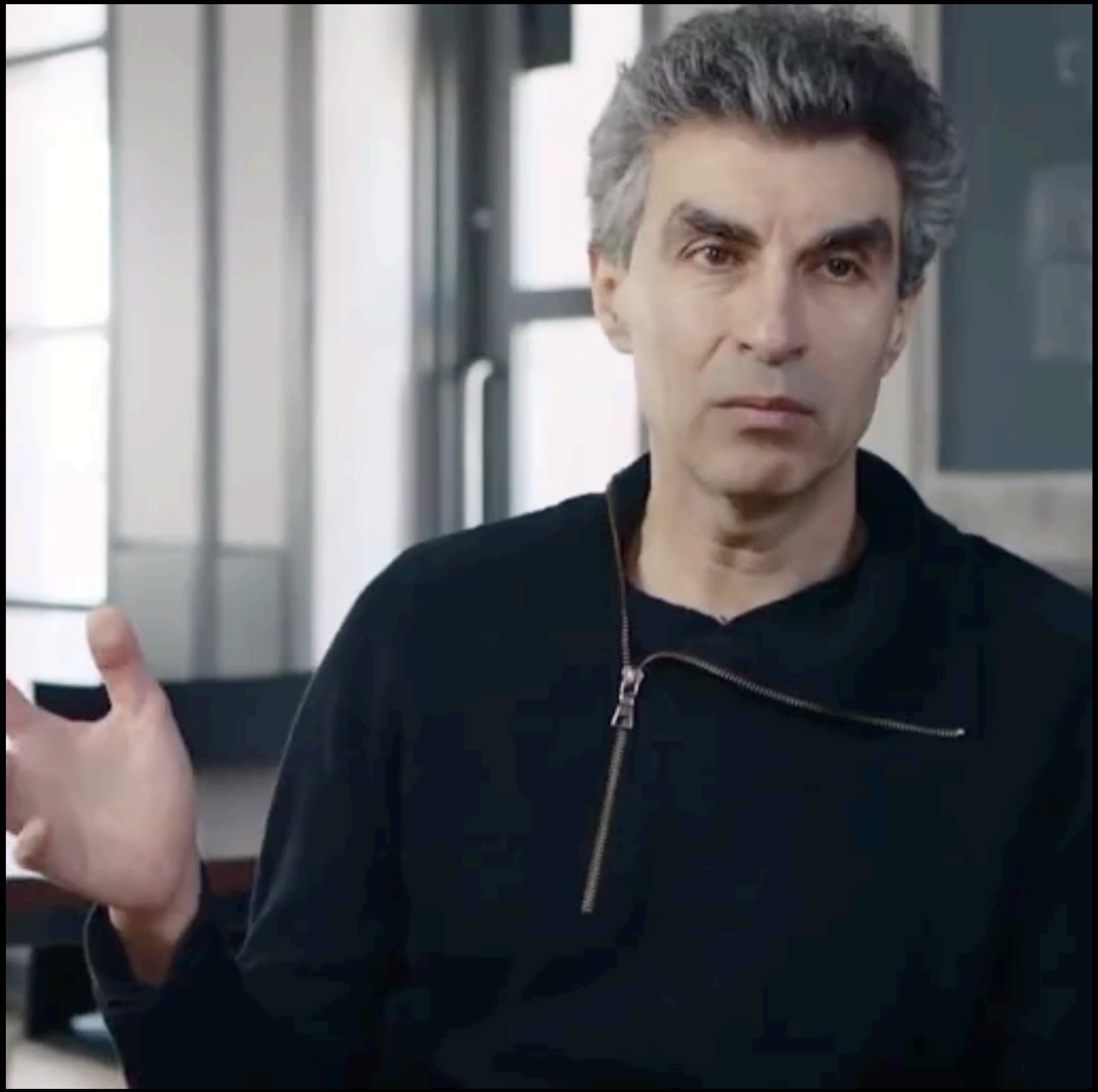


Morph Cut
[Berthouzoz et al. 12]

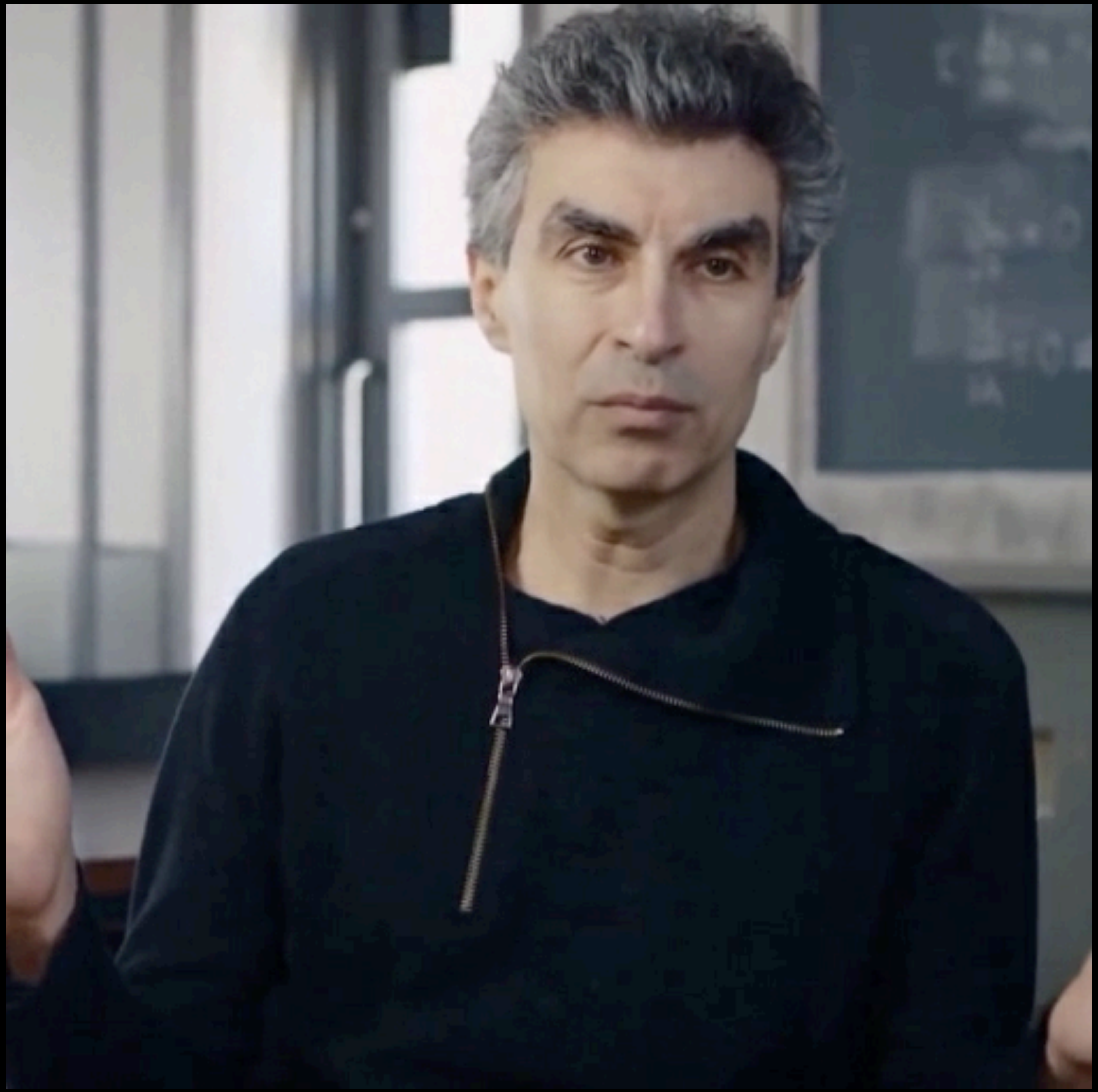


Ours

“… learning from examples ~~and~~ and scientists …”

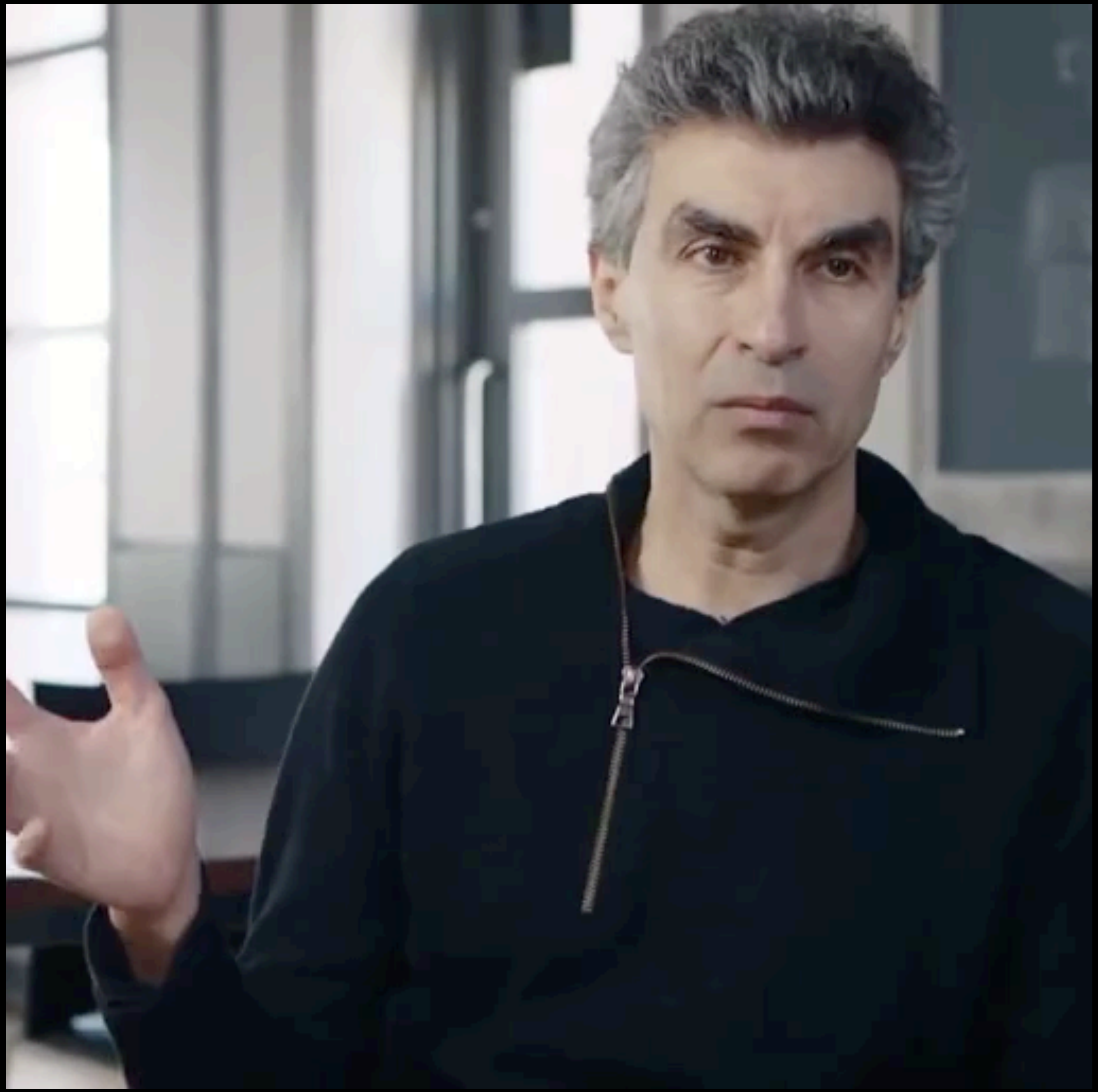


Morph Cut
[Berthouzoz et al. 12]

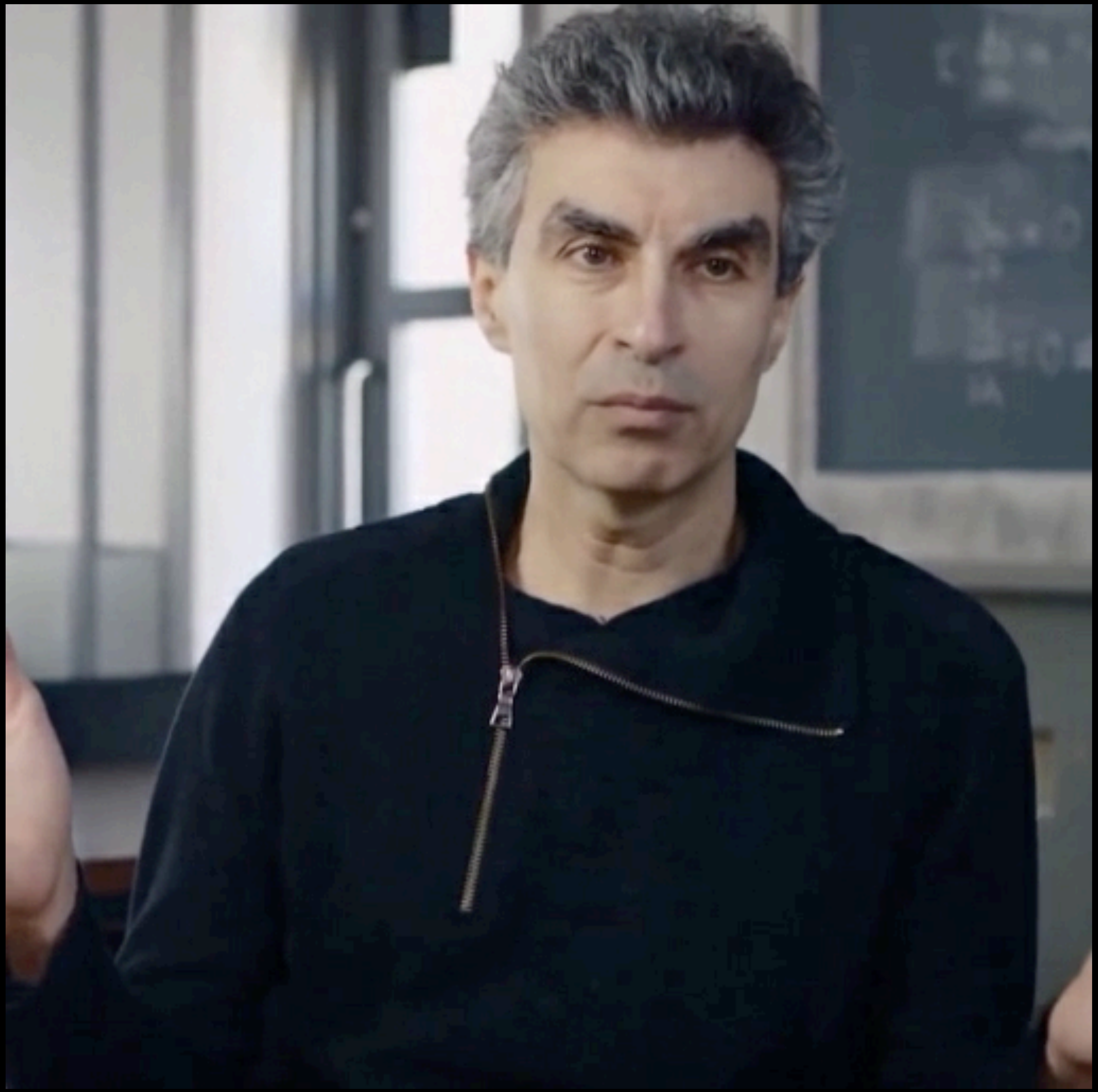


Ours

“… learning from examples ~~and~~ and scientists …”



Morph Cut
[Berthouzoz et al. 12]

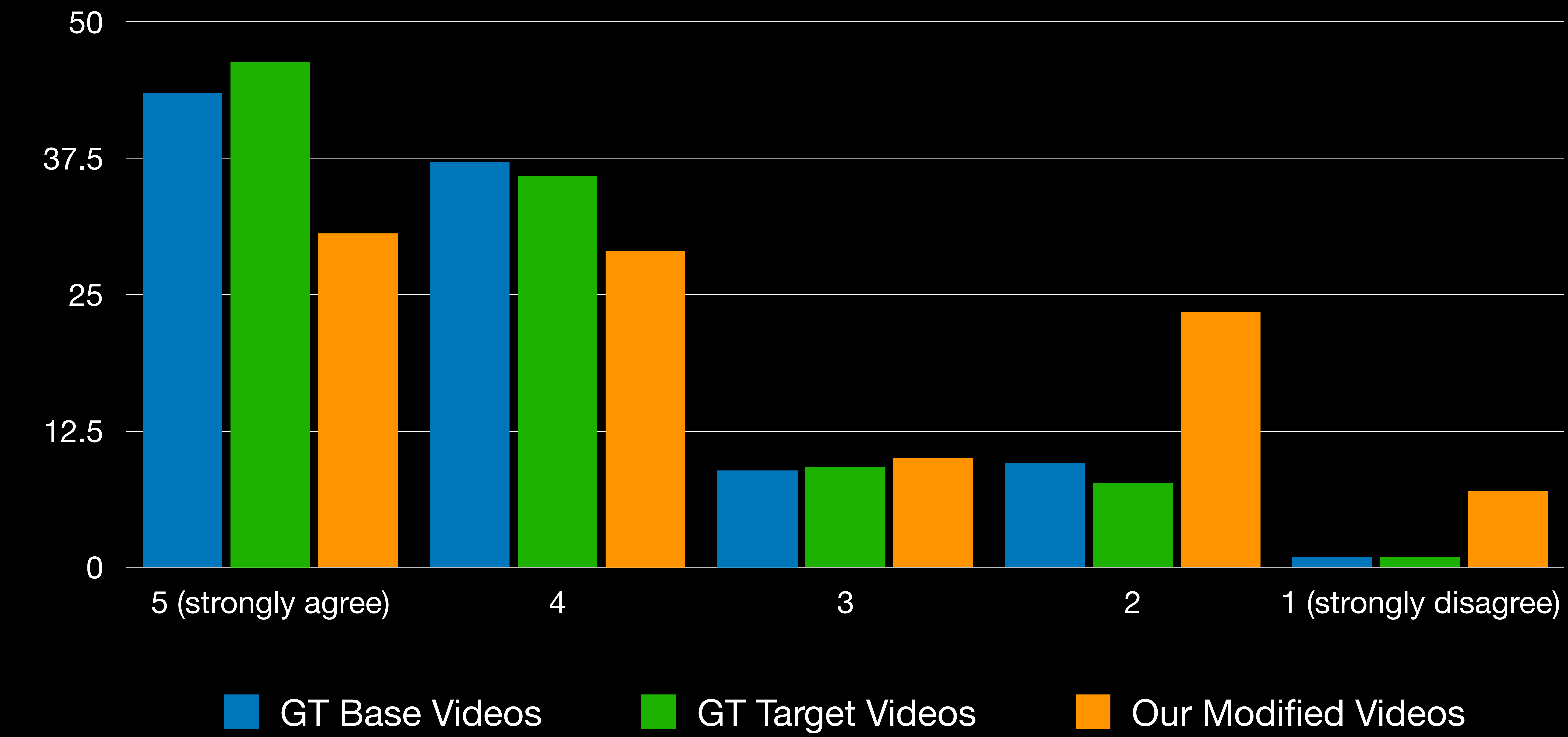


Ours

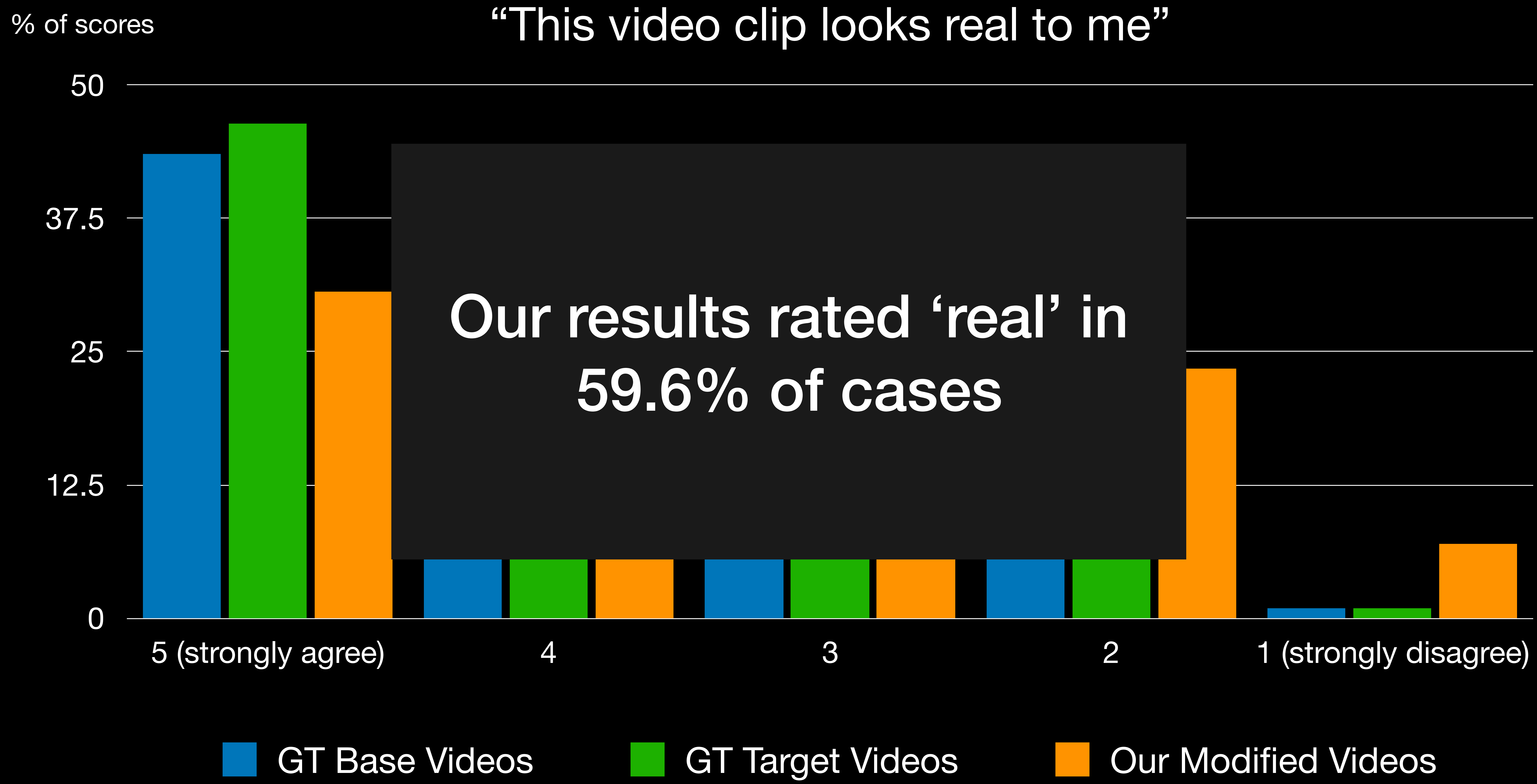
"… learning from examples ~~and~~ and scientists …"

User study

“This video clip looks real to me”



“This video clip looks real to me”



Our results rated ‘real’ in 59.6% of cases

Limitations & future work

Limitations & future work

- Moods and facial expressions
  - We might blend incompatible sequences
  - Can we control these when synthesizing?

Limitations & future work

- Moods and facial expressions
    - We might blend incompatible sequences
    - Can we control these when synthesizing?
- Viseme search is slow
    - Can speed up with some relaxations

Limitations & future work

- Moods and facial expressions
    - We might blend incompatible sequences
    - Can we control these when synthesizing?
- Viseme search is slow
    - Can speed up with some relaxations
- Interactivity
    - Algorithm speedups
    - Editing UI

Constrained Local Models

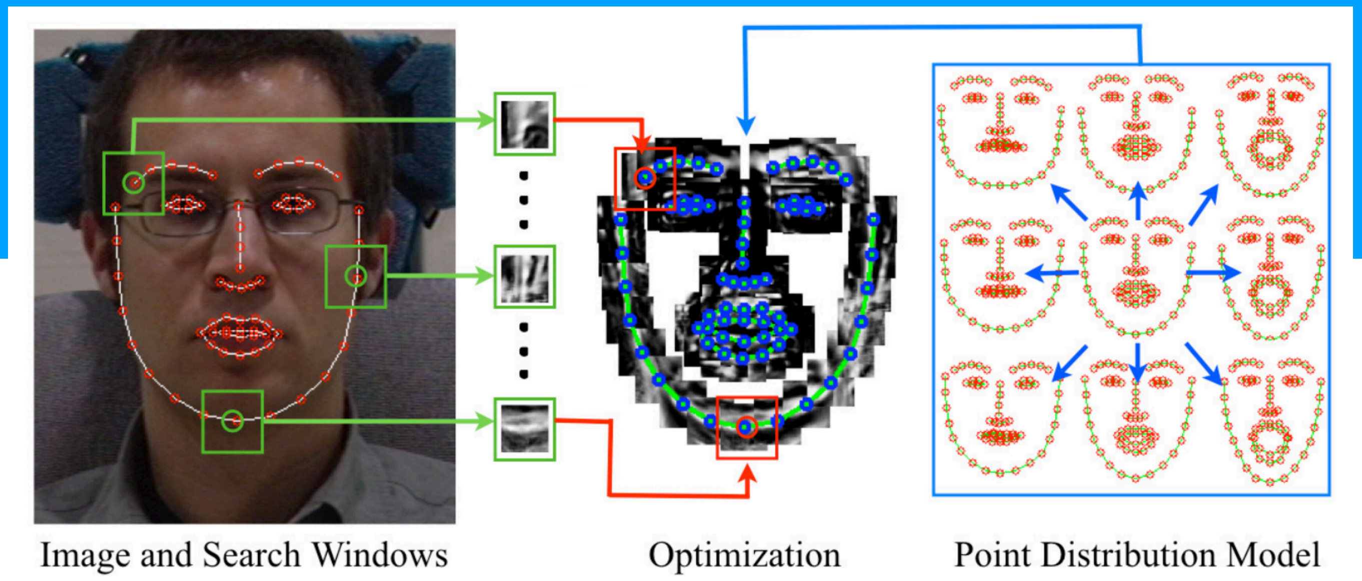
Improving Portraits

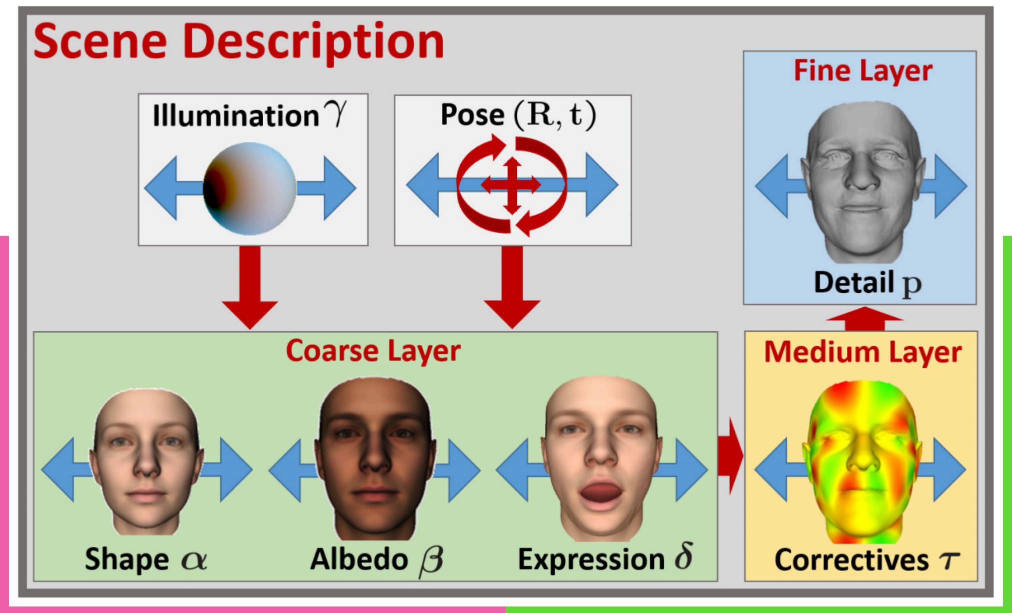
Editing Video

Constrained Local Models

Improving Portraits

Editing Video





Constrained Local Models

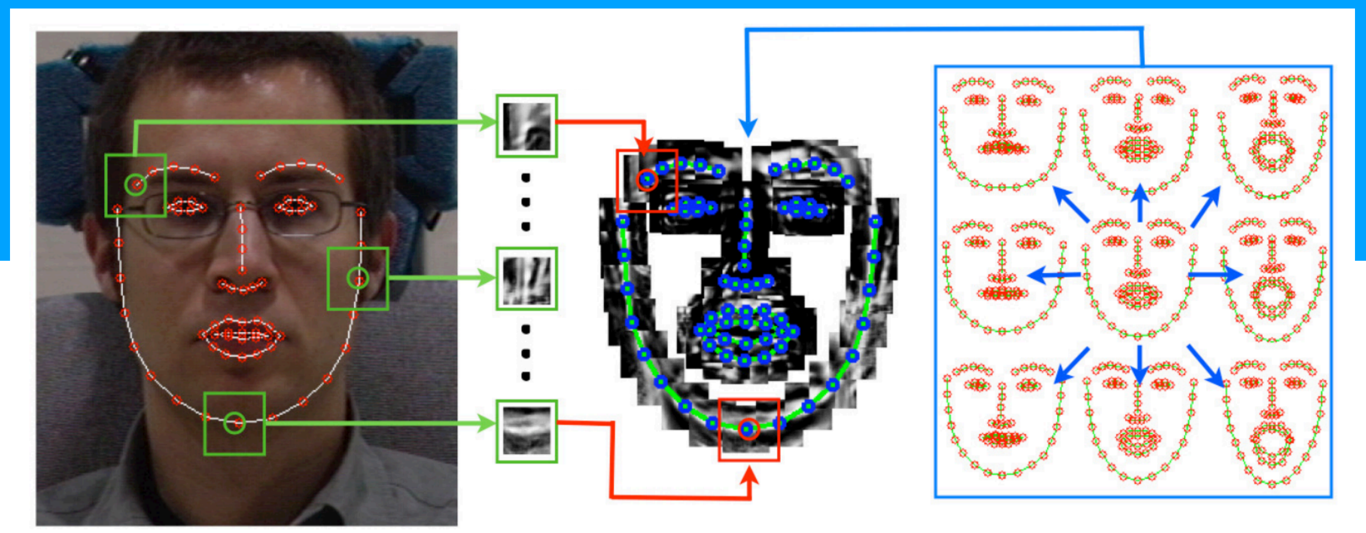


Image and Search Windows

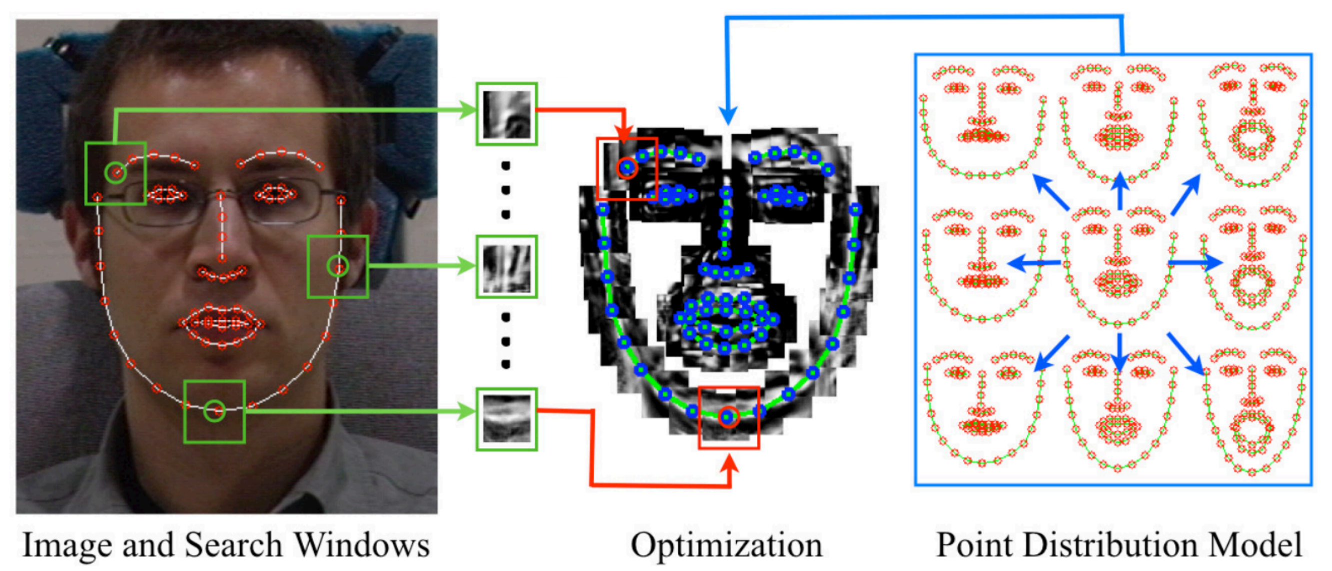
Optimization

Point Distribution Model

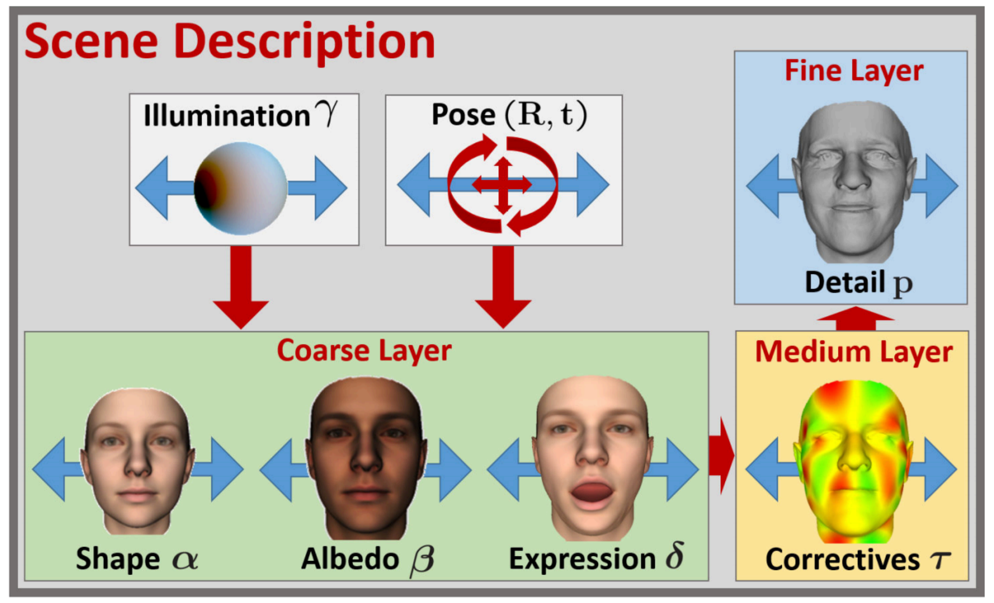
Improving Portraits

Editing Video

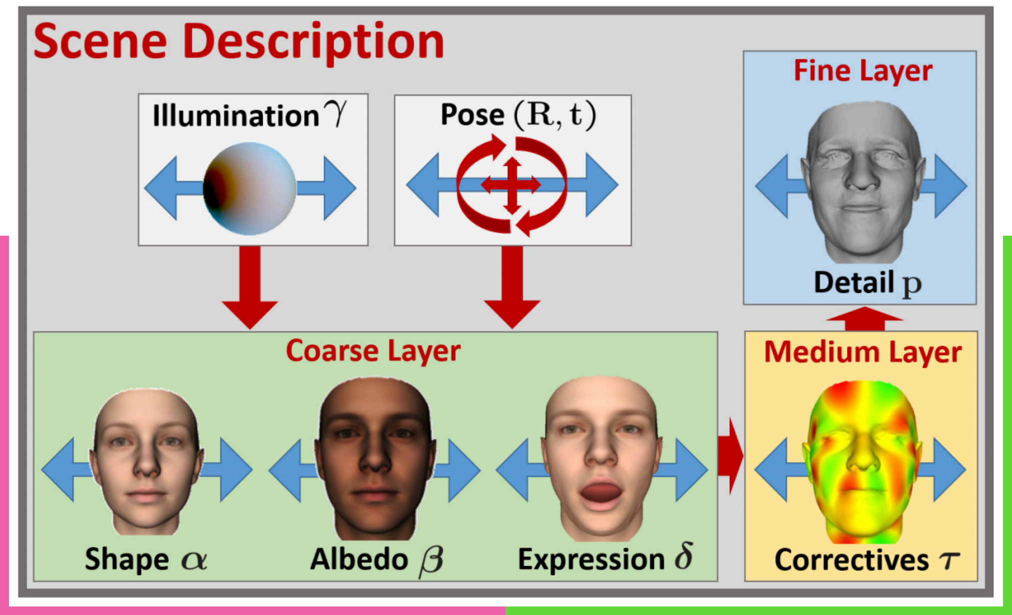
Constrained Local Models



Improving Portraits



Editing Video



Constrained Local Models

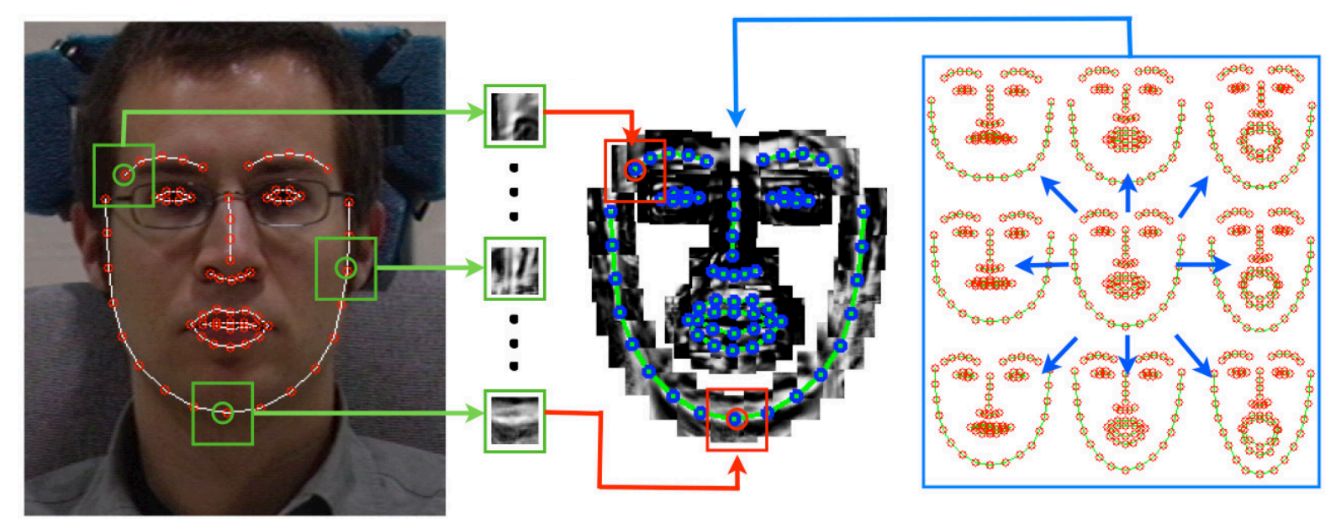


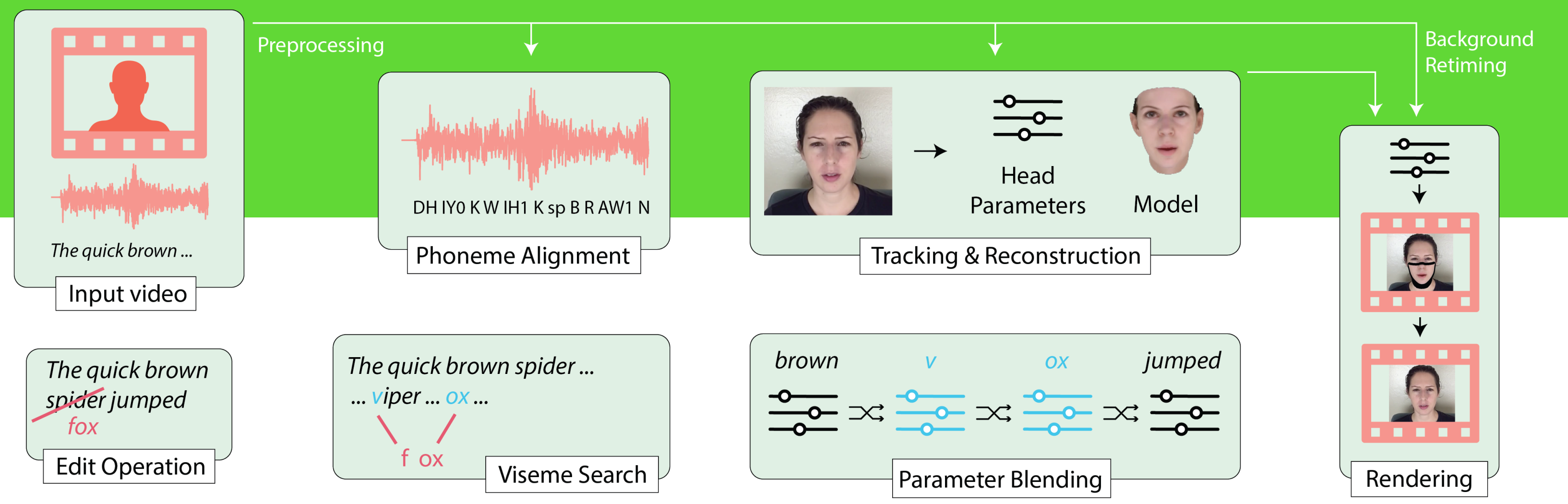
Image and Search Windows

Optimization

Point Distribution Model

Improving Portraits

Editing Video


Preprocessing
Background Retiming
The quick brown ...
Input video
DH IY0 K W IH1 K sp B R AW1 N
Phoneme Alignment
Head Parameters
Model
Tracking & Reconstruction
Rendering
The quick brown ~~spider~~ jumped
fox
Edit Operation
The quick brown spider ...
... viper ... ox ...
f ox
Viseme Search
brown
v
ox
jumped
Parameter Blending