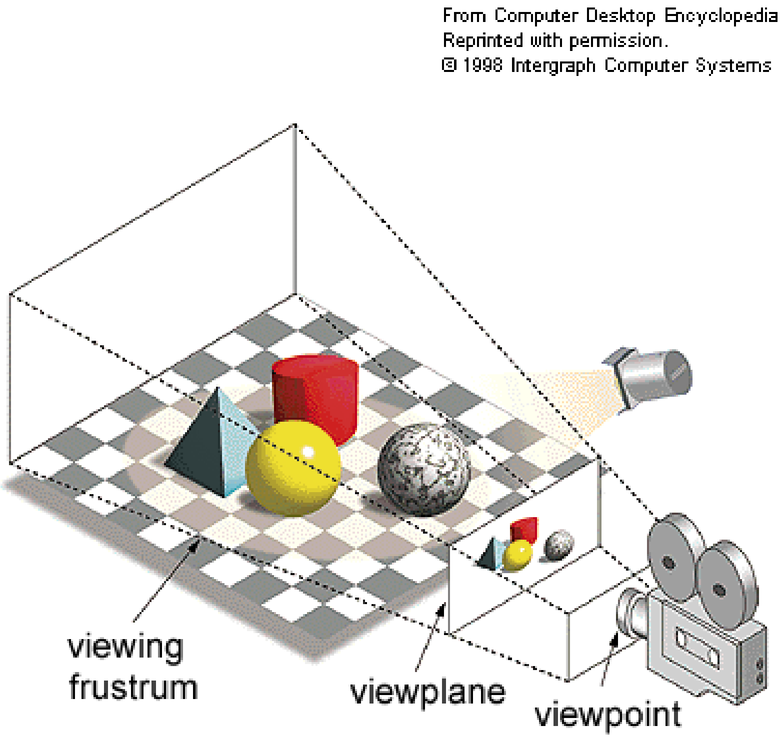
Neural Rendering

Rendering

3D Scene:
- Material
- Lighting
- Geometry
  (incl. animation)



Camera Def.
- Intrinsics
- Often:
  - focal length
  - principal point)

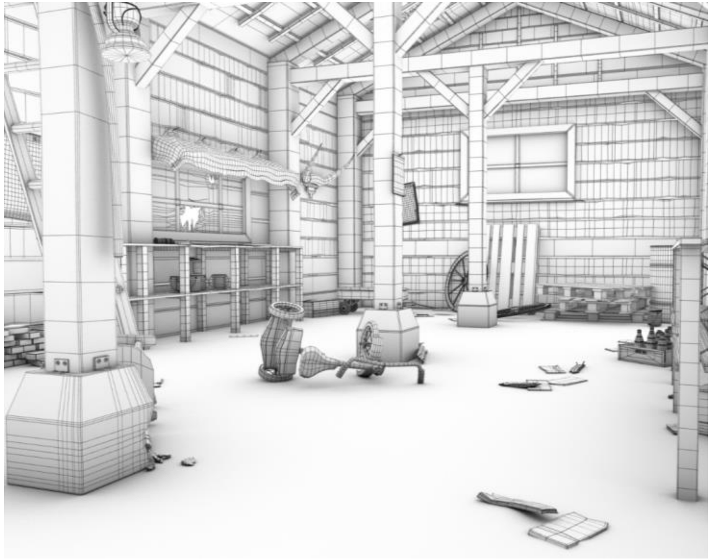
Camera View Point
- Extrinsics
- 6 DoF (3rot, 3trans)

Photo-realistic Image Synthesis

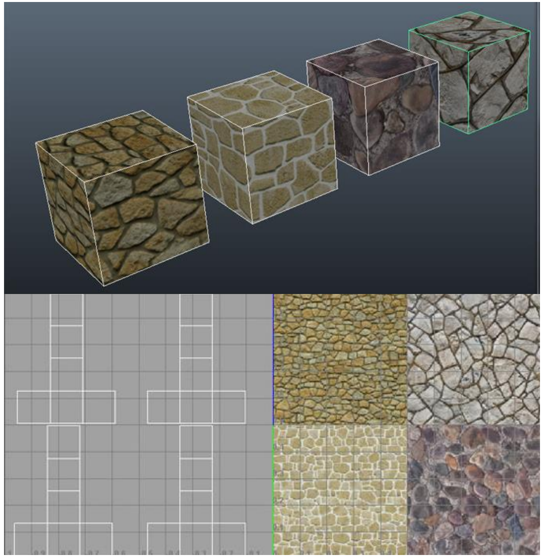
The Rendering Equation [Kajiya 86]

$$L_{\mathrm{o}}(\mathbf{x},\,\omega_{\mathrm{o}},\,\lambda,\,t) \;=\; L_e(\mathbf{x},\,\omega_{\mathrm{o}},\,\lambda,\,t) \;+\; \int_\Omega f_r(\mathbf{x},\,\omega_{\mathrm{i}},\,\omega_{\mathrm{o}},\,\lambda,\,t)\, L_{\mathrm{i}}(\mathbf{x},\,\omega_{\mathrm{i}},\,\lambda,\,t)\,(\omega_{\mathrm{i}}\,\cdot\,\mathbf{n})\,\mathrm{d}\,\omega_{\mathrm{i}}$$

Need 3D Content for Rendering



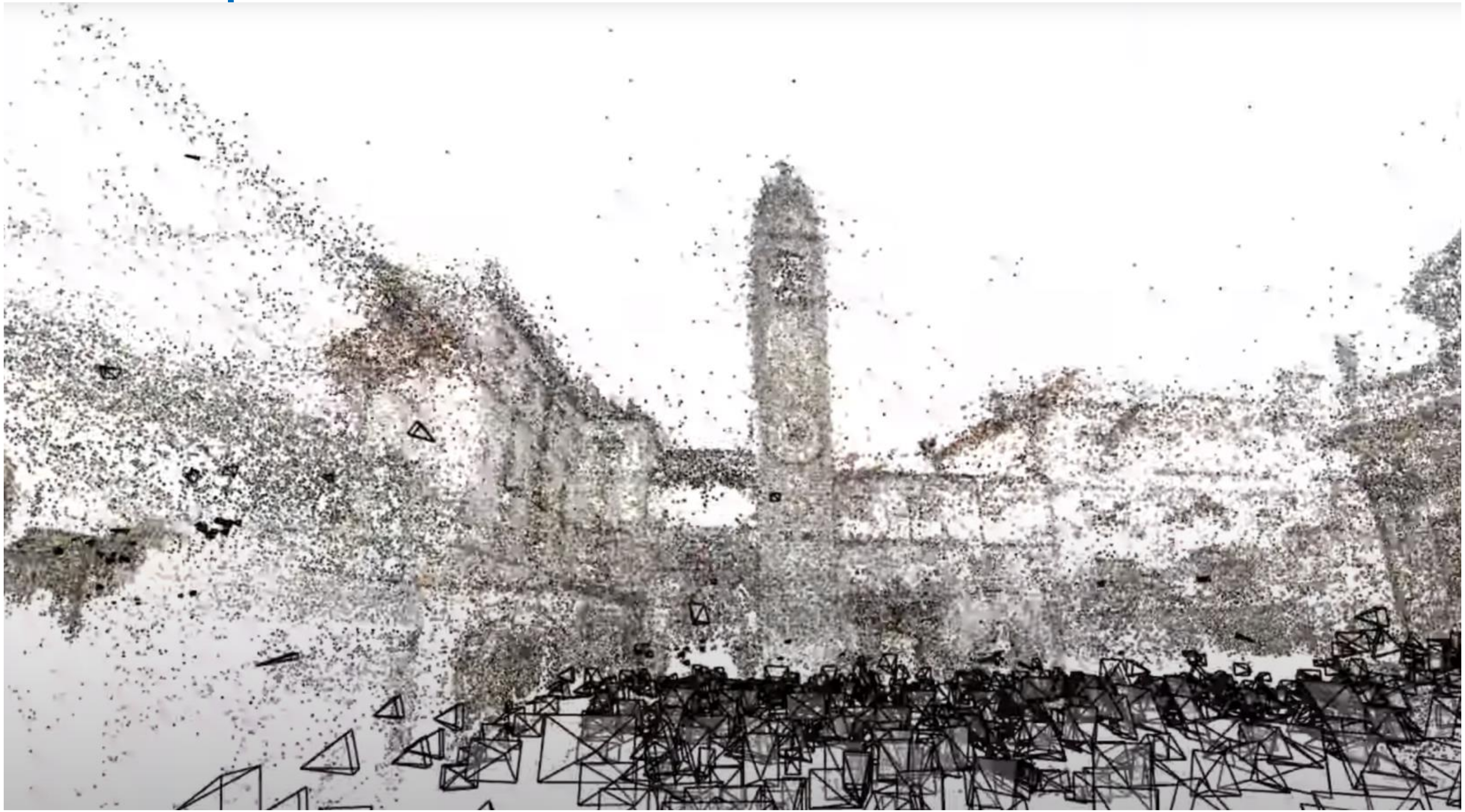
Geometry



Textures

Material & Lighting

Computer Vision for Reconstruction



ICCV’09 [Agarwal et al.]: Building Rome in a Day

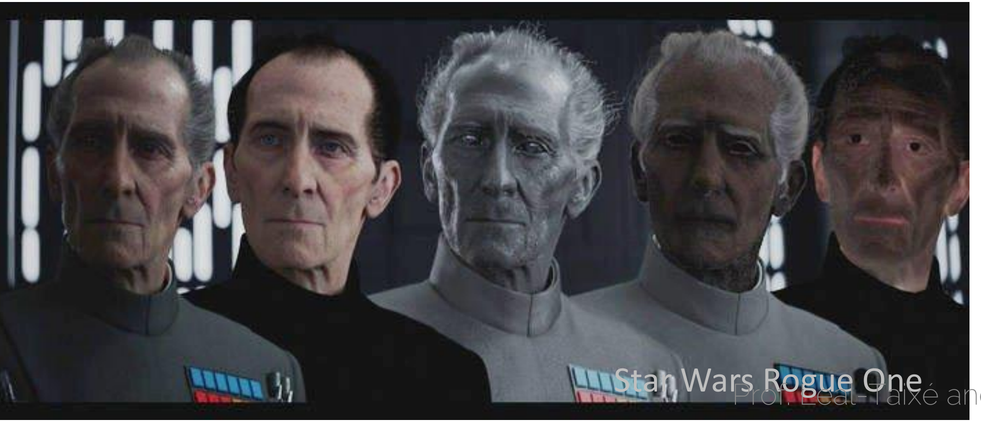
3D Digitization

Computer Graphics

Computer Vision

Traditional Graphics vs Deep Learning

3D Model + Textures + Shading -> Synthetic Image



Generative Adversarial Networks

Discriminator loss

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}} \log D(\boldsymbol{x}) - \frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log\left(1 - D\left(G(\boldsymbol{z})\right)\right)$$

Generator loss

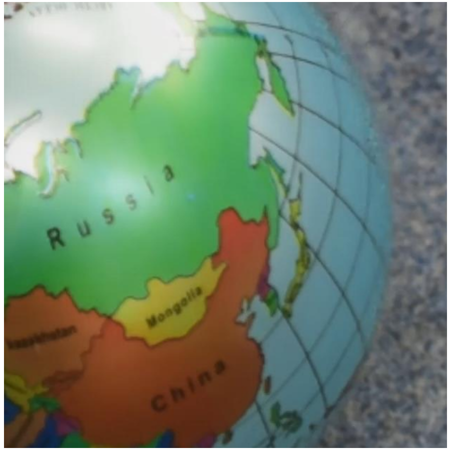
$$J^{(G)} = -J^{(D)}$$

Idea of Neural Rendering

Novel View point synthesis:

6 DoF Camera
Pose / View Point

Neural Network
-> Encodes entire scene description, lighting, materials, etc.



Neural Rendering with Pix2Pix

Ground truth for training

- Pose + Target Image (e.g., observed from real world)
- Constrain with re-rendering loss

Testing

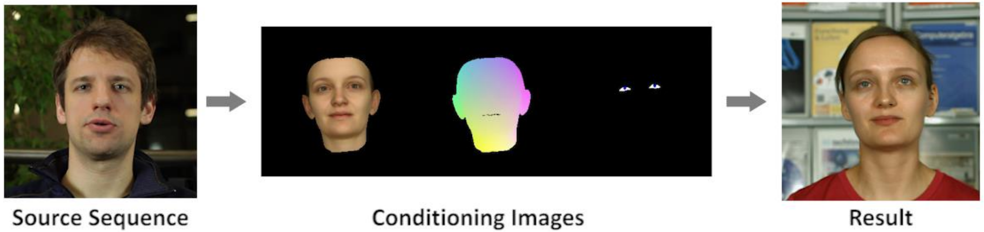
- Given unseen pose, generate image

Neural Rendering with Pix2Pix

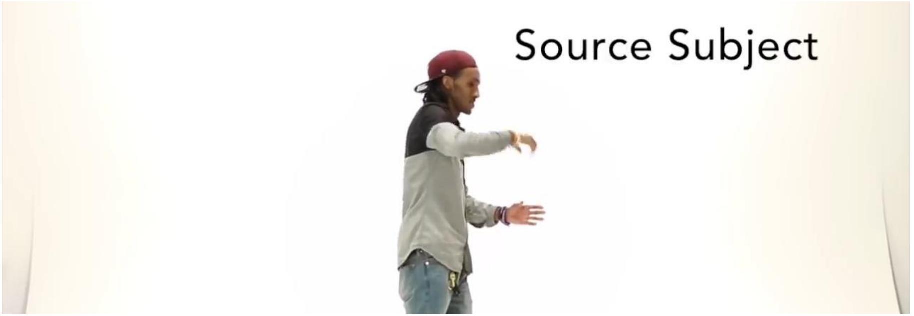
Pix2Pix [Isola et al. 2017]

Other Neural Rendering

- Conditioned on Faces (Deep Video Portraits)



- Conditioned on Human Skeleton (Everybody Dance Now)



Neural Rendering with Pix2Pix

Pix2Pix [Isola et al. 2017]

Deep Voxels

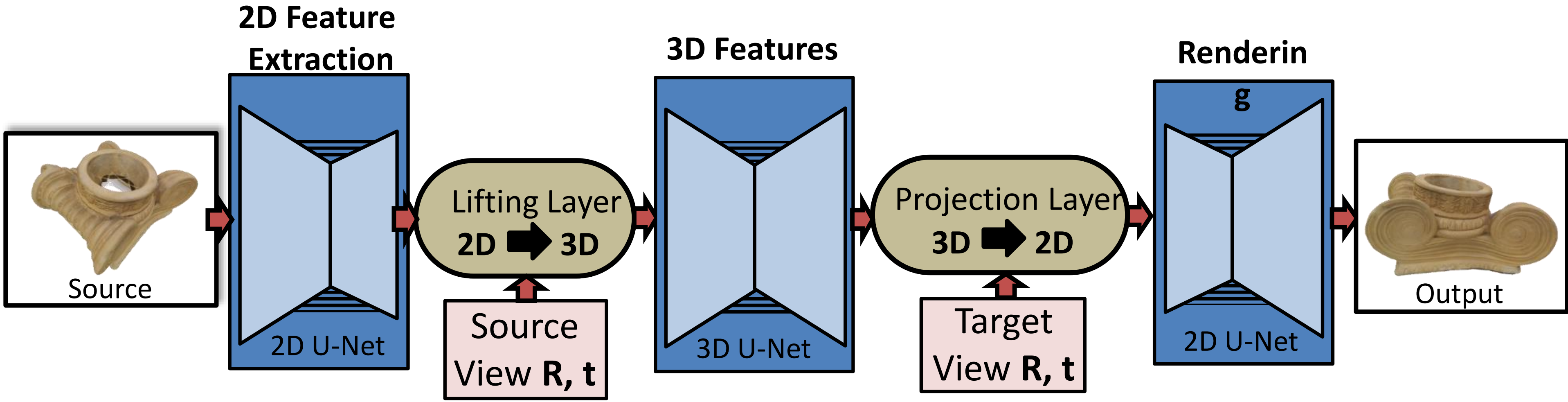
[Sitzmann et al. CVPR'19] Deep Voxels

Deep Voxels

- Main idea for video generation:
  - Why learn 3D operations with 2D Convs !?!?
  - We know how 3D transformations work
    - E.g., 6 DoF rigid pose [R | t]
  - Incorporate these into the architectures
    - Need to be differentiable!
  - Example application: novel view point synthesis
    - Given rigid pose, generate image for that view

[Sitzmann et al. CVPR'19] Deep Voxels

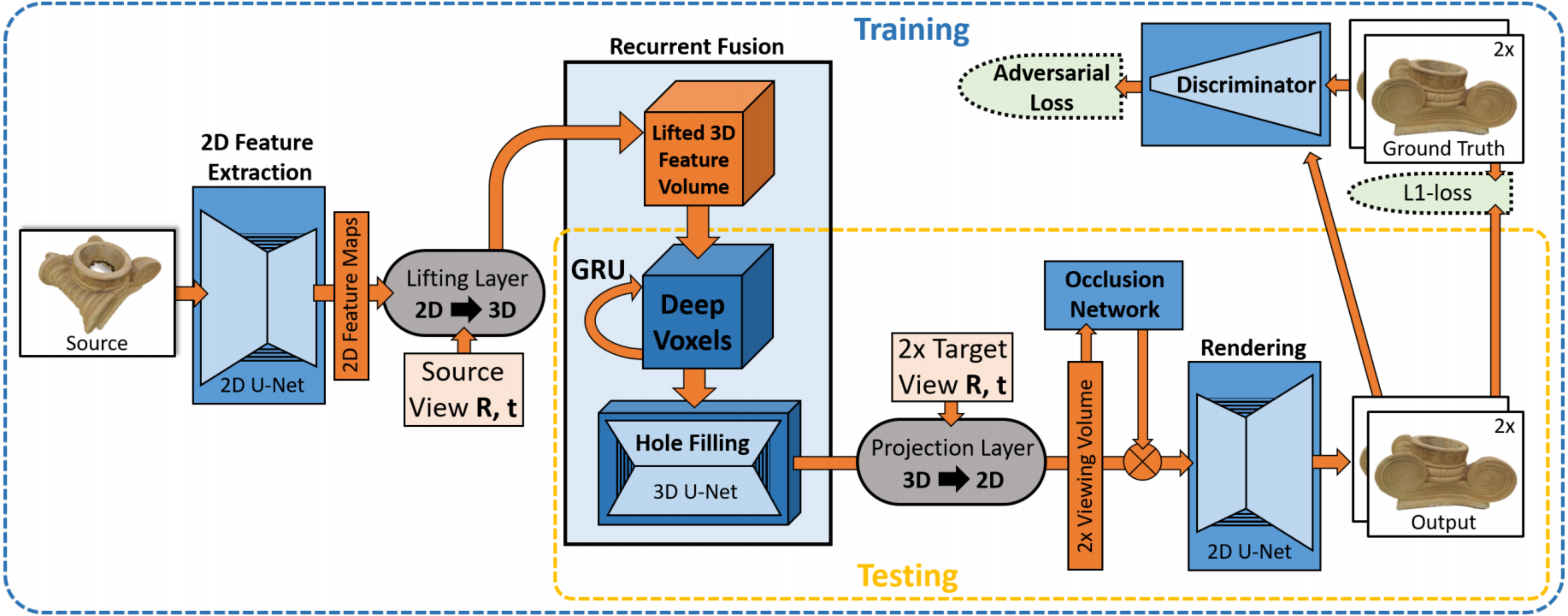
Deep Voxels



Simplified overview for novel view synthesis

[Sitzmann et al. CVPR'19] Deep Voxels

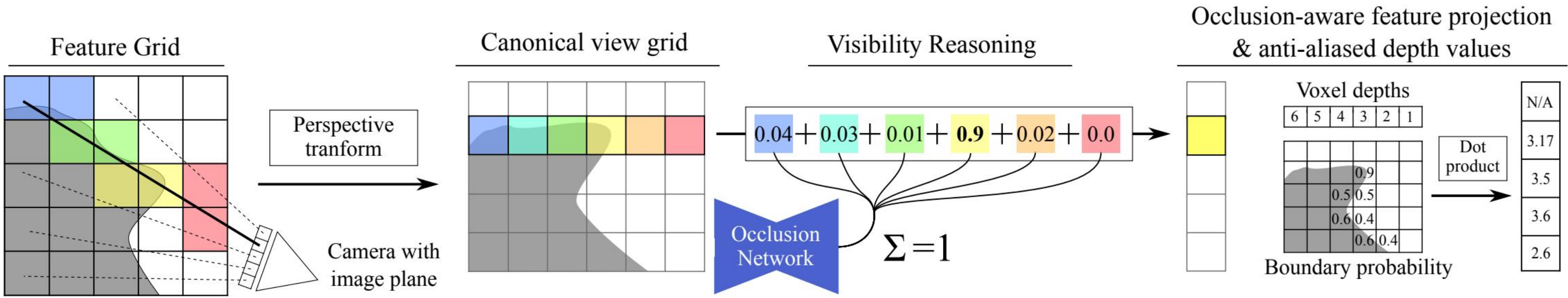
Deep Voxels



[Sitzmann et al. CVPR'19] Deep Voxels

Deep Voxels

Occlusion Network:



Issue: we don't know the depth for the target!
-> Per-pixel softmax along the ray
-> Network learns the depth

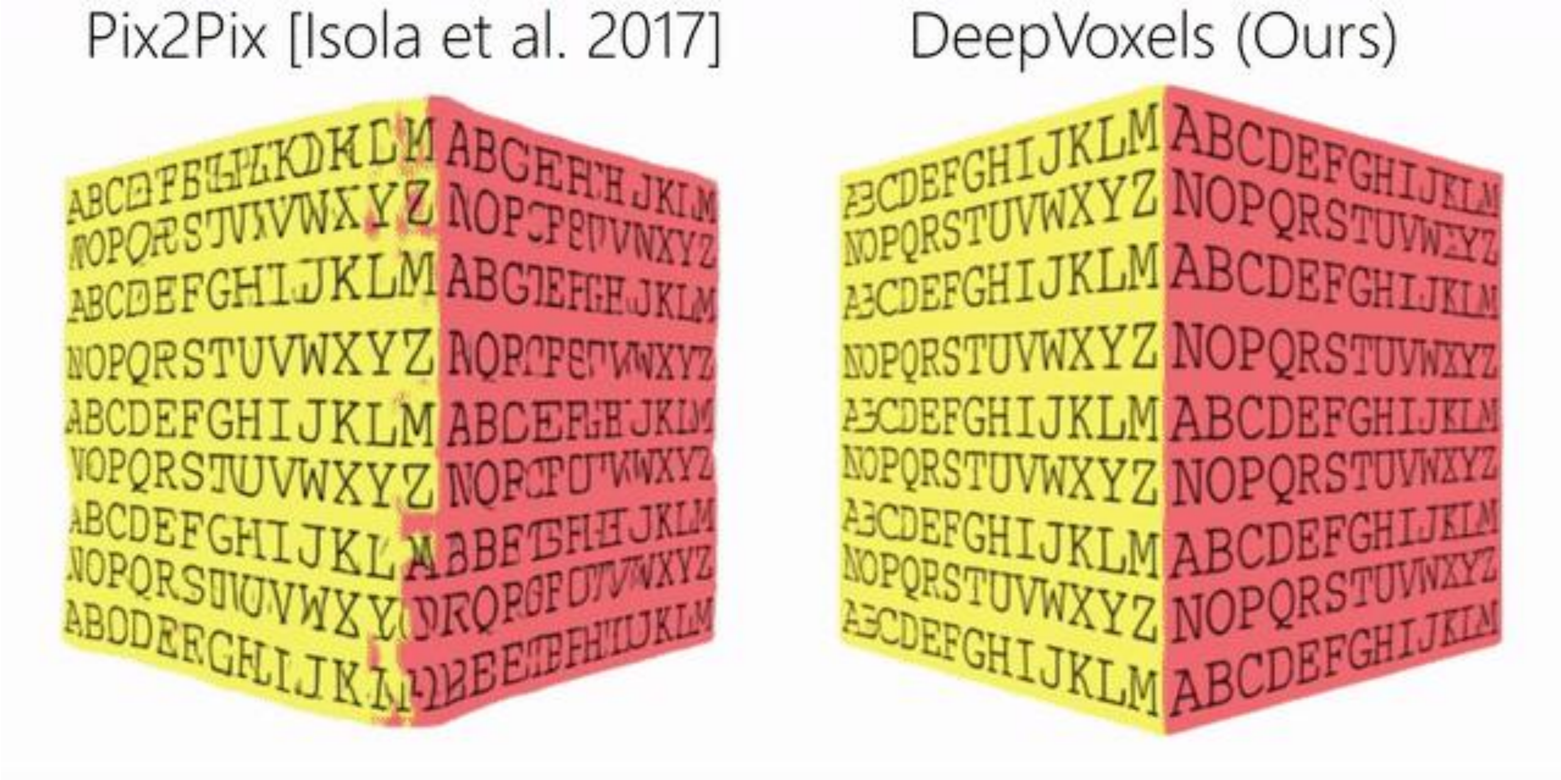
[Sitzmann et al. CVPR'19] Deep Voxels

Deep Voxels

DeepVoxels

Best Baseline: Pix2Pix [Isola et al. 2017]

Deep Voxels



[Sitzmann et al. '18] Deep Voxels

Deep Voxels: Insights

- Lifting from 2D to 3D works great
  - No need to take specific care for temp. coherency!

- All 3D operations are differentiable

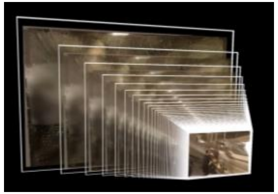
- Currently, only for novel view-point synthesis
  - I.e., cGAN for new pose in a given scene

- But: limited resolution due to dense 3D voxel grid

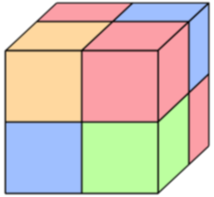
[Sitzmann et al. '18] Deep Voxels

Importing 3D structure from CG

Scene Representation

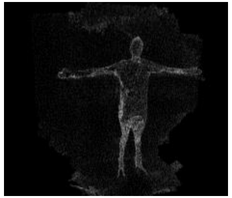


Multi-Plane Images

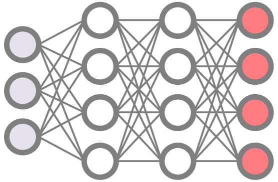


Voxelgrids

Image-based



Point Clouds



Implicit Function

Renderer

(Alpha) compositing

Volumetric Ray-based

Rasterization

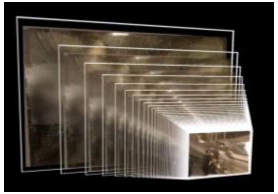
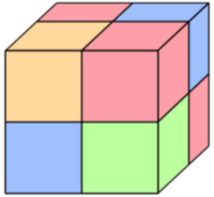
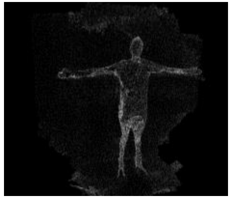
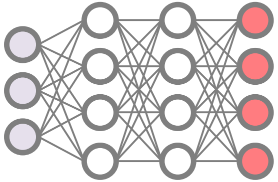
Splatting

Sphere-Traced Volumetric

Scene Representation Differentiable Renderer

Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

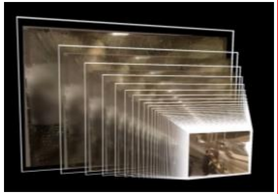
Importing 3D structure from CG

Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros					
Cons					

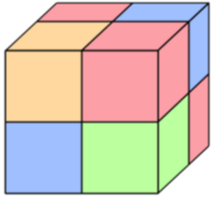
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG

Scene Representation

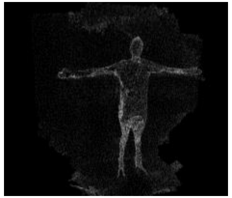


Multi-Plane Images

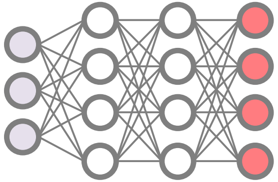


Voxelgrids

Image-based



Point Clouds



Implicit Function

Renderer

(Alpha) compositing

Volumetric Ray-based

Rasterization

Splatting

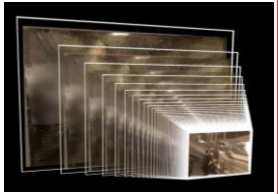
Sphere-Traced Volumetric

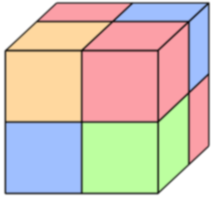
Pros

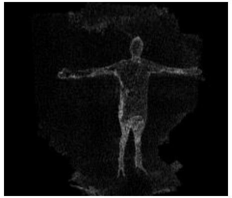
Cons

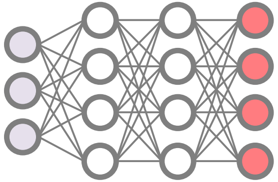
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG





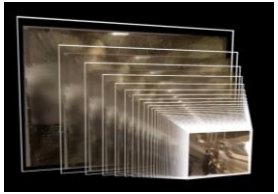
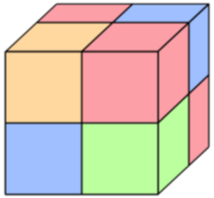
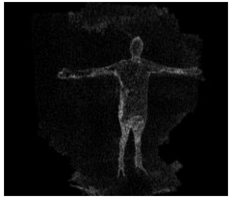
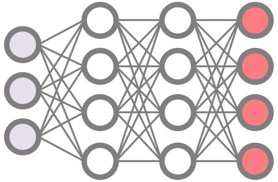




Scene Representation	Multi-Plane Images	Voxelgrids	Image-based	Point Clouds	Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes				
Cons	Only 2.5D Size				

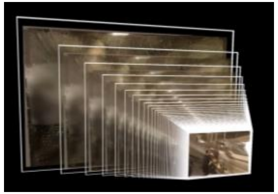
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

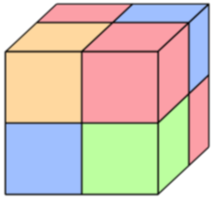
Importing 3D structure from CG

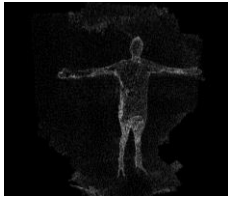
					
Scene Representation	**Multi-Plane Images**	**Voxelgrids**	**Image-based**	**Point Clouds**	**Implicit Function**
Renderer	**(Alpha) compositing**	**Volumetric Ray-based**	**Rasterization**	**Splatting**	**Sphere-Traced Volumetric**
Pros	**Fast rendering High quality Generalizes**				
Cons	**Only 2.5D Size**				

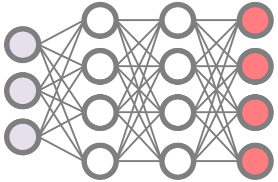
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG





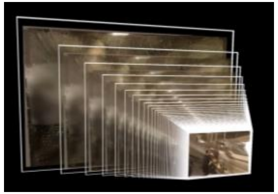
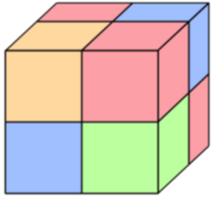
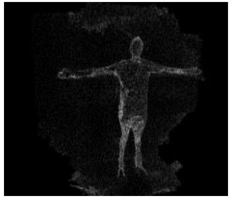
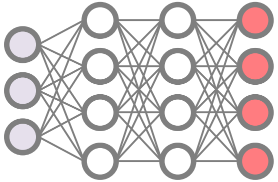




Scene Representation	Multi-Plane Images	Voxelgrids	Image-based	Point Clouds	Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality			
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$			

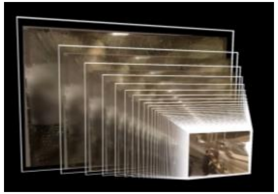
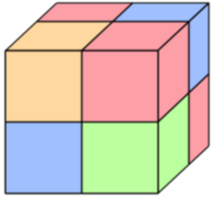
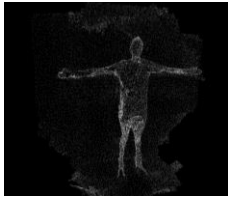
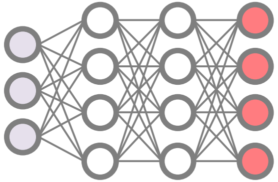
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG

Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality			
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$			

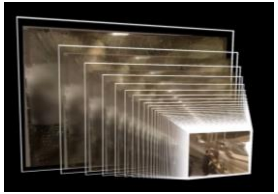
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

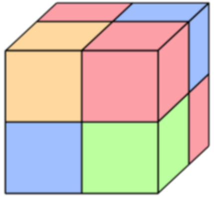
Importing 3D structure from CG

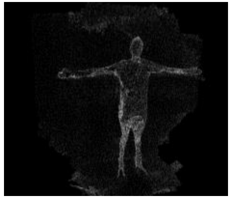
Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality	High quality		
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$	Requires good SFM No compact representation		

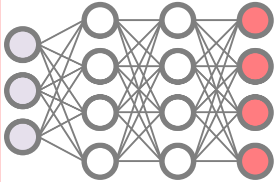
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG





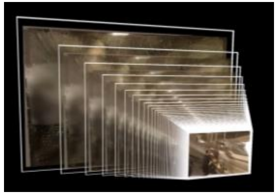
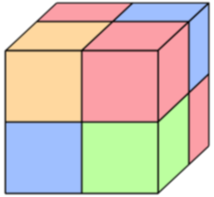


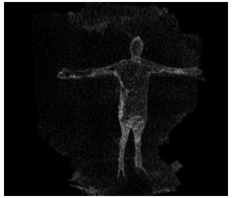
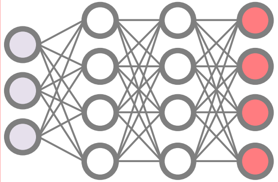


Scene Representation	Multi-Plane Images	Voxelgrids	Image-based	Point Clouds	Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality	High quality		
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$	Requires good SFM No compact representation		

Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG

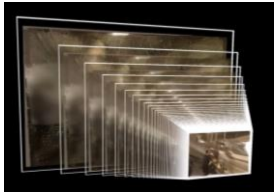



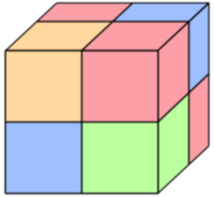




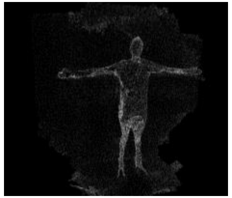
Scene Representation	Multi-Plane Images	Voxelgrids	Image-based	Point Clouds	Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality	High quality	High quality	
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$	Requires good SFM No compact representation	Requires good SFM	

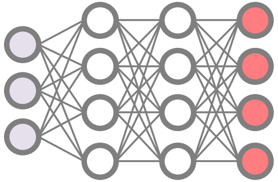
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Importing 3D structure from CG







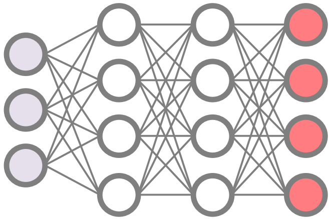


Scene Representation	Multi-Plane Images	Voxelgrids	Image-based	Point Clouds	Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality	High quality	High quality	
Cons	Only 2.5D Size	No reconstruction priors Memory O(n^3)	Requires good SFM No compact representation	Requires good SFM	

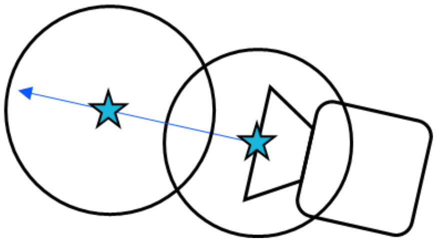
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

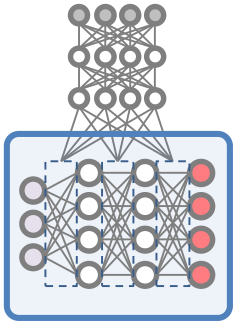
Scene Representation Networks
Sitzmann et al., Neurips 2019

Scene Representation	Renderer	Generalization
ReLU MLP	Generalized (learned) sphere-tracing	Hypernetwork



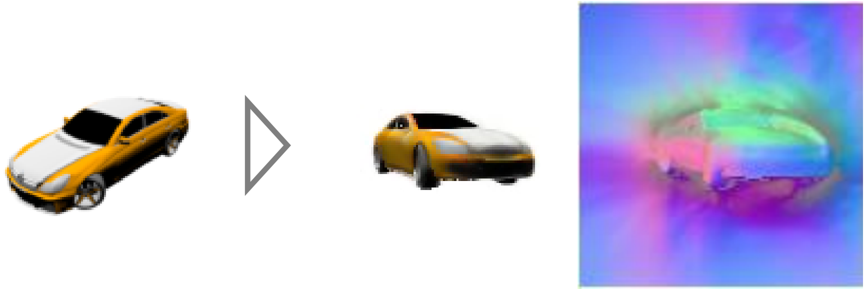
$\mathbb{R}^3 \rightarrow \mathbb{R}^n$





Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Scene Representation Networks
Sitzmann et al., Neurips 2019



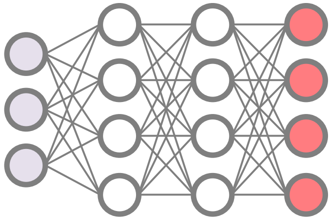
Full 3D Reconstruction from single image!

Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

NERF: Neural Radiance Fields Mildenhall et al., arXiv 2020

Scene Representa tion

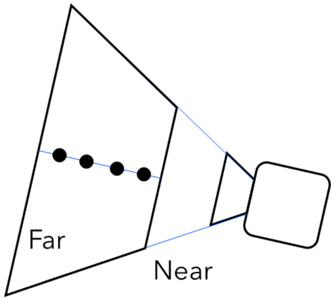
ReLU MLP +
Positional Encoding
View Direction



$\mathbb{R}^6 \rightarrow \mathbb{R}^3$

Renderer

Volumetric,
stratified sampling



Generalizati on

None.

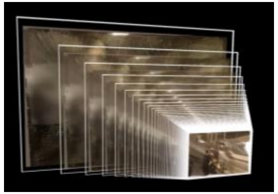
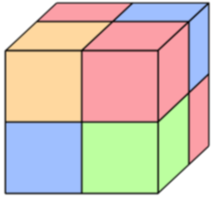
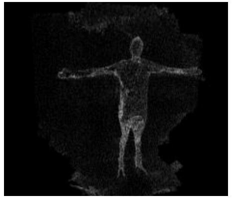
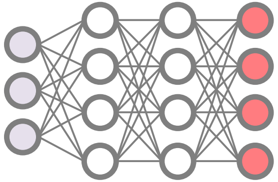
Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

NERF: Neural Radiance Fields
Mildenhall et al., arXiv 2020

Photorealistic, including view-dependence!
(~100 images)

Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Requirements

Scene Representa tion	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Tracing Volumetric
Pros	Fast rendering High quality Generalizes	“True 3D” High quality	High quality	High quality	High quality May generalize!
Cons	Only 2.5D Size	No reconstruction priors Memory O(n³)	Requires good SFM No compact representation	Requires good SFM	Expensive rendering, training

Slides: Vincent Sitzmann (Eurographics State-of-the-art on Neural Rendering)

Neural Textures: Features on 3D Mesh

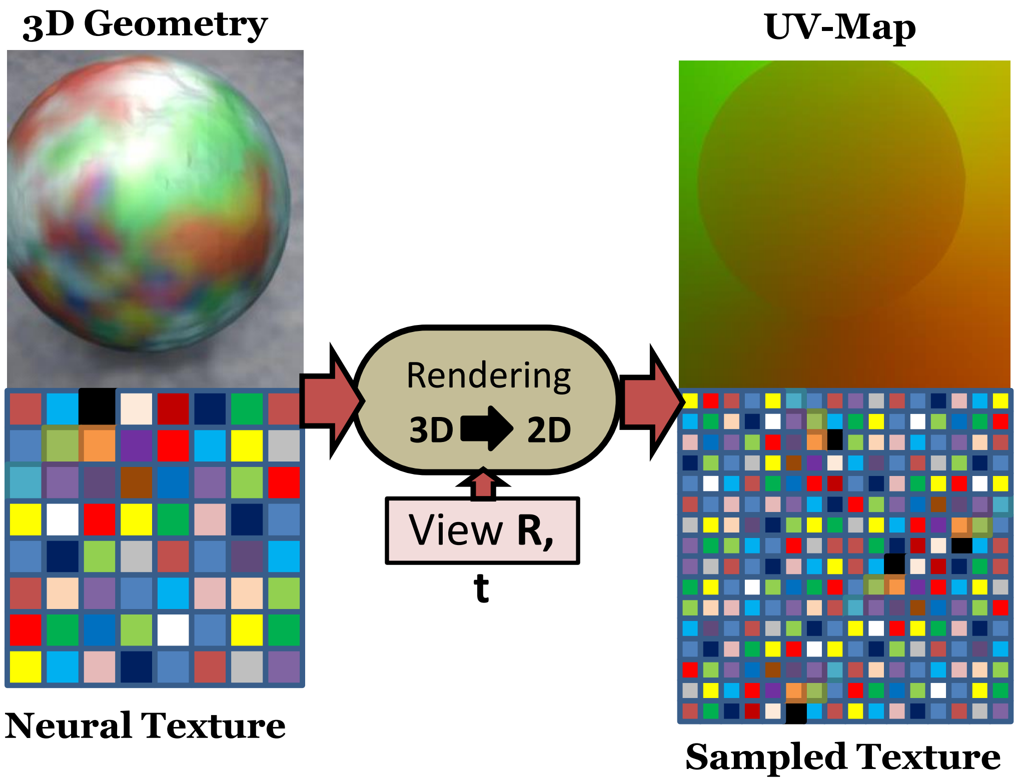
Neural Textures: Features on 3D Mesh

3D Geometry

Neural Texture

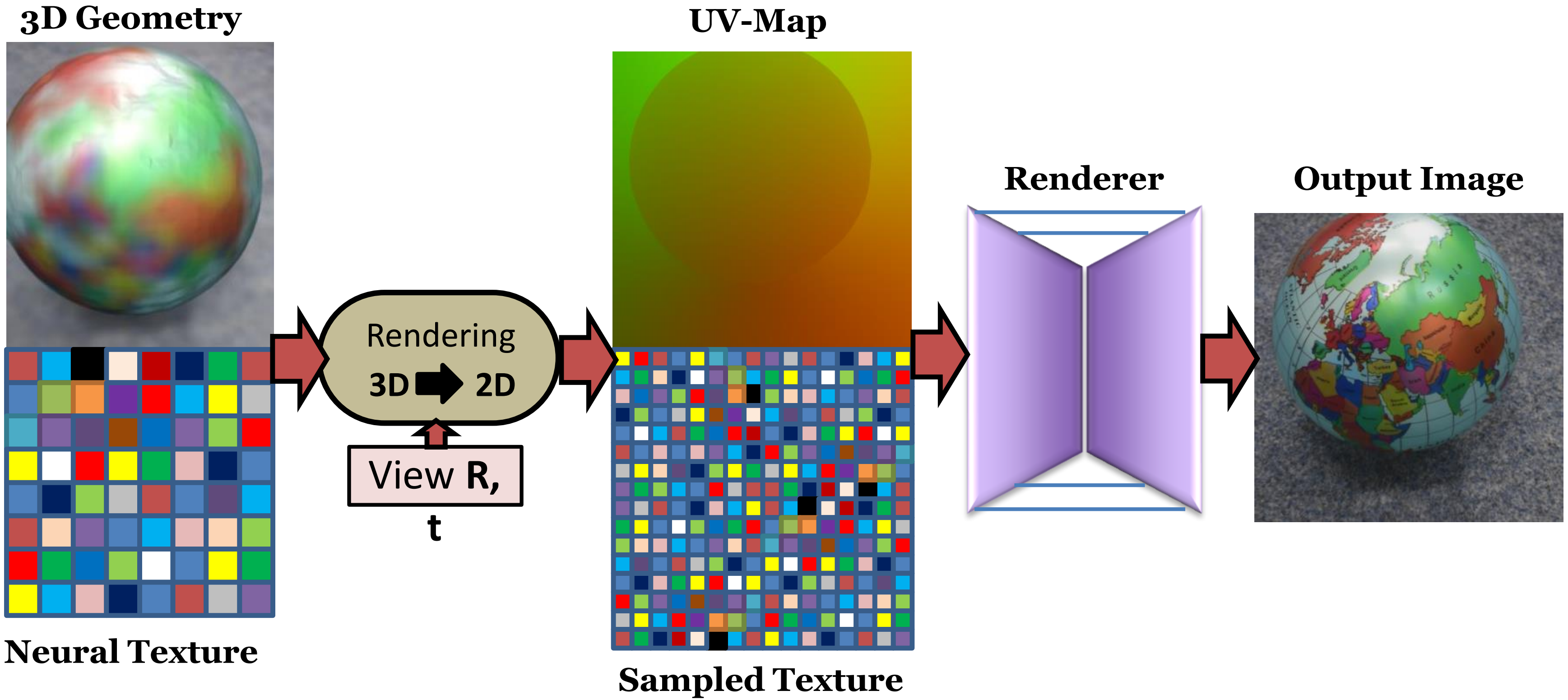
Siggraph'19 [Thies et al.]: Neural Textures

Neural Textures: Features on 3D Mesh



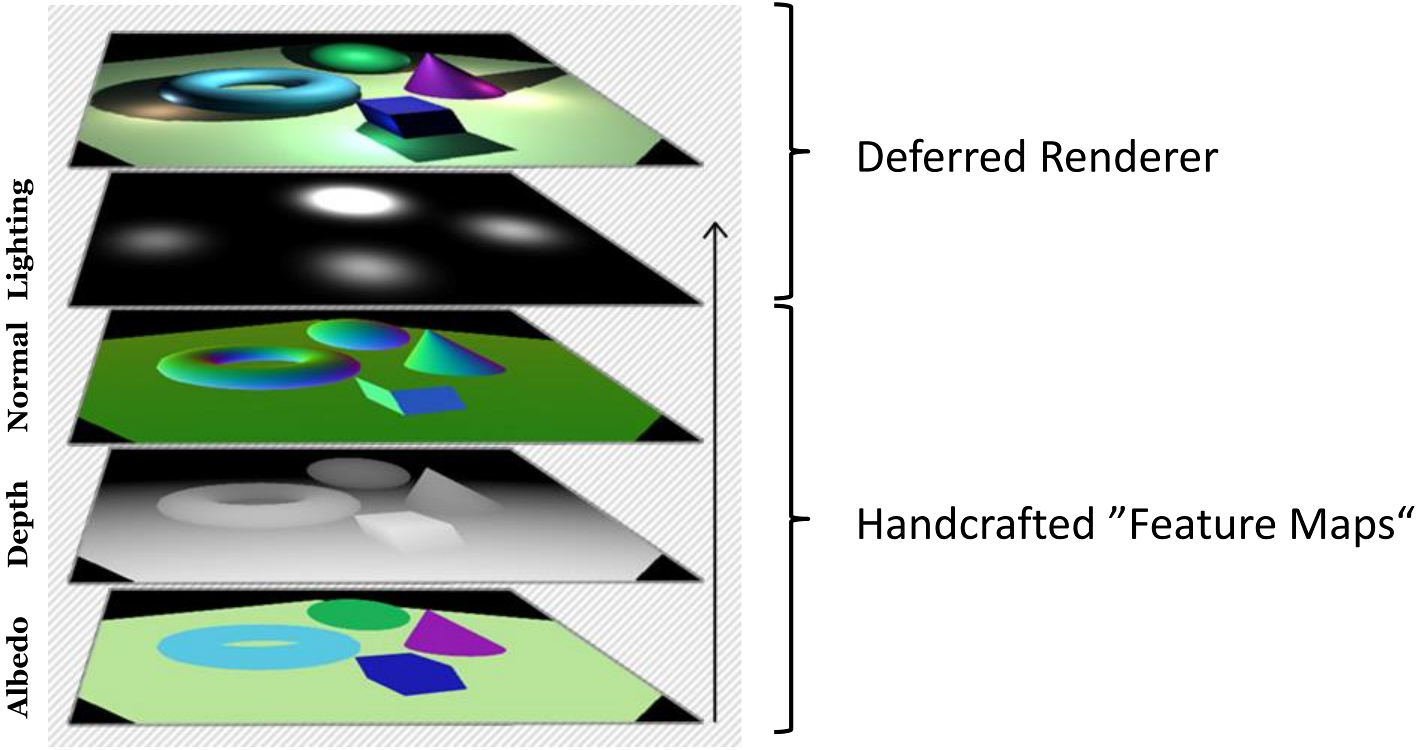
Siggraph'19 [Thies et al.]: Neural Textures

Neural Textures: Features on 3D Mesh



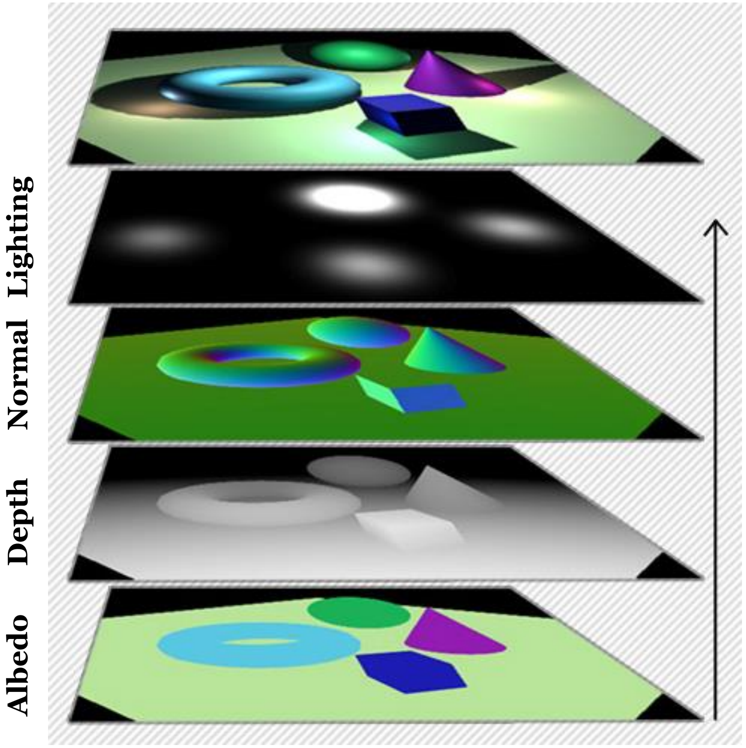
Siggraph'19 [Thies et al.]: Neural Textures

Deferred Neural Rendering



Siggraph’19 [Thies et al.]: Neural Textures

Deferred Neural Rendering

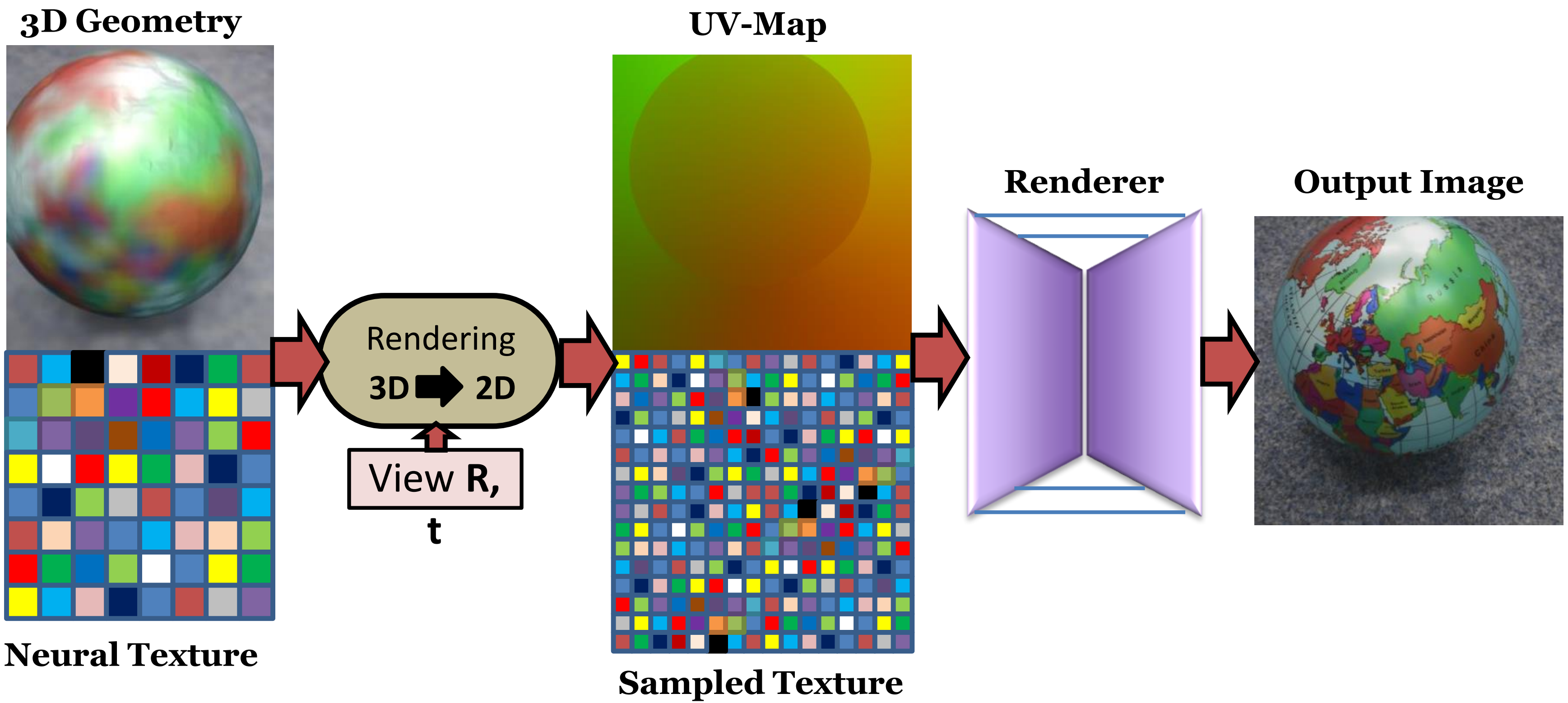


Neural
Deferred Renderer

Learned
~~Handcrafted~~ ”Feature Maps“

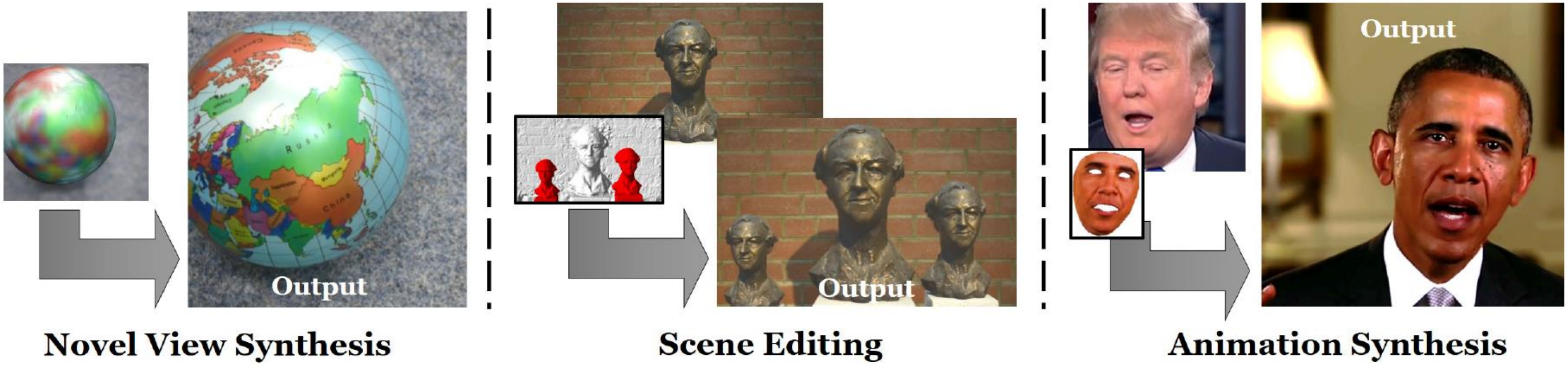
Siggraph’19 [Thies et al.]: Neural Textures

Deferred Neural Rendering



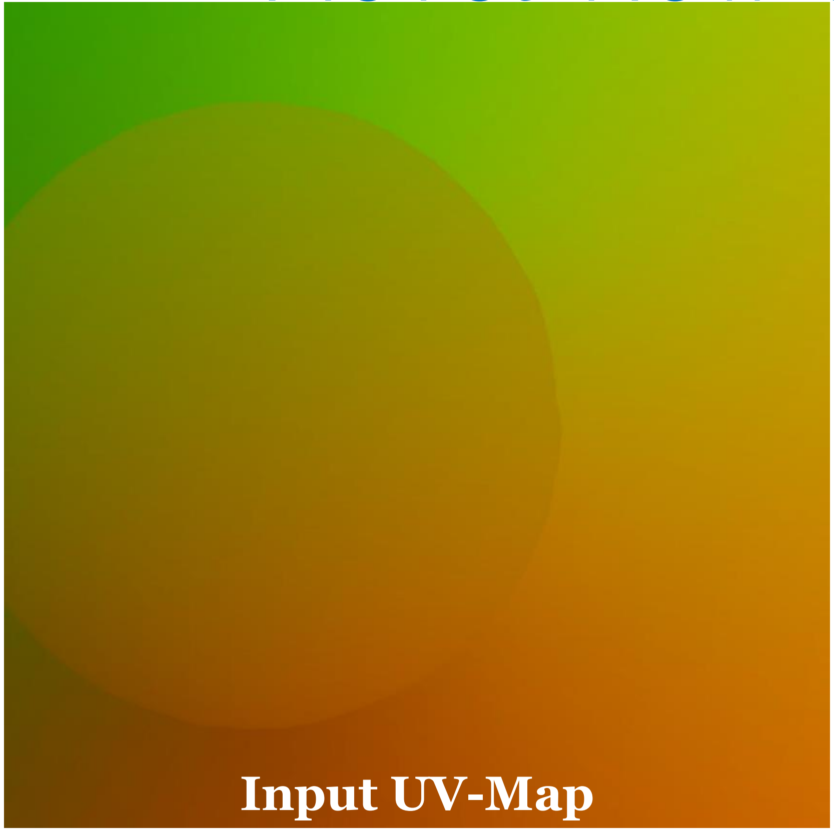
Siggraph'19 [Thies et al.]: Neural Textures

Neural Textures: Features on 3D Mesh



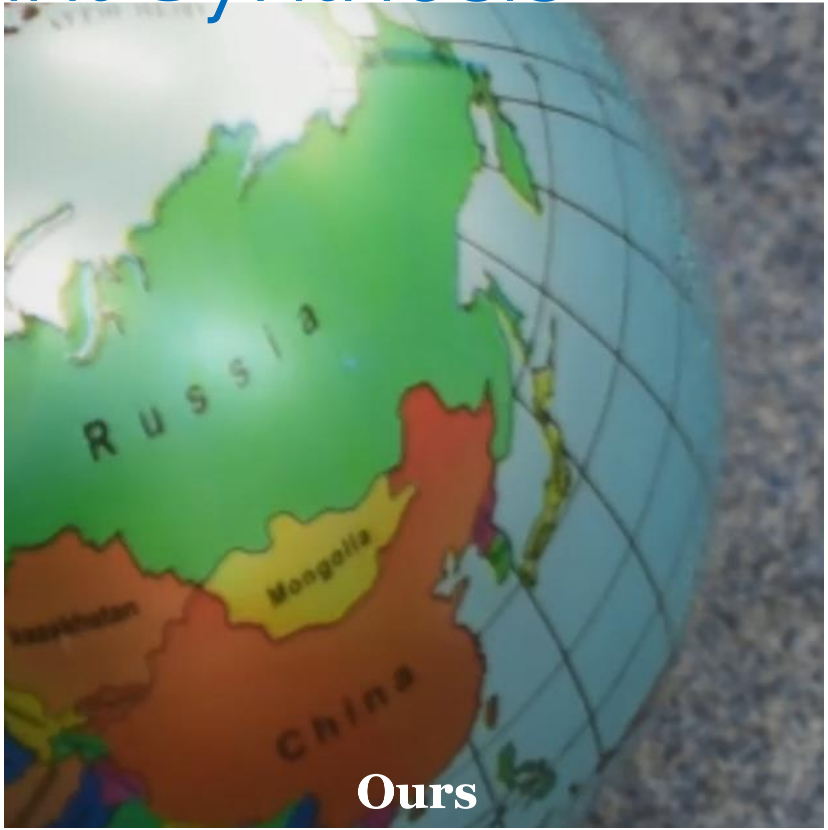
Siggraph'19 [Thies et al.]: Neural Textures

Novel View-Point Synthesis



Siggraph'19 [Thies et al.]: Neural Textures

Novel View-Point Synthesis



Siggraph'19 [Thies et al.]: Neural Textures

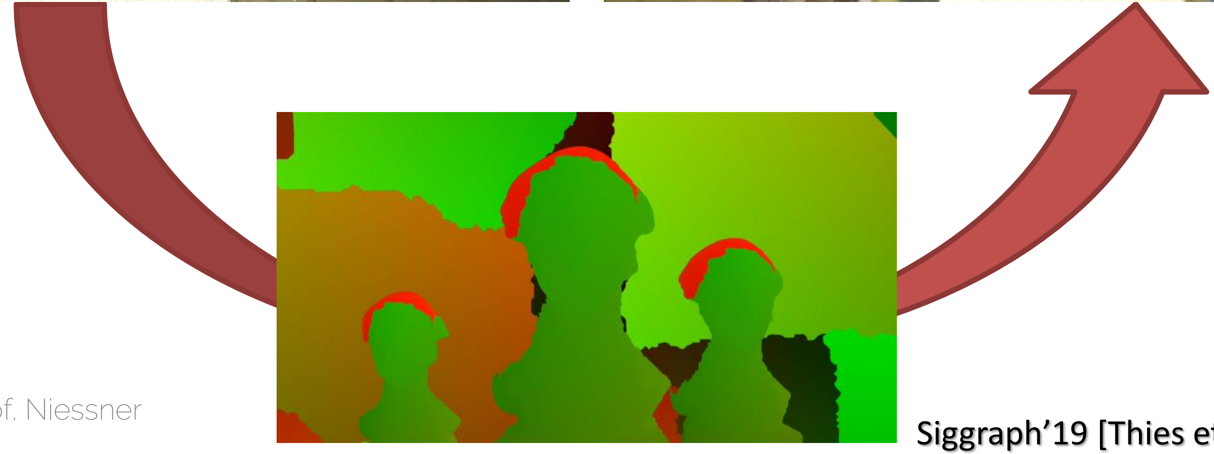
Scene Editing

Input Sequence

Geometry Editing

Siggraph'19 [Thies et al.]: Neural Textures

Scene Editing

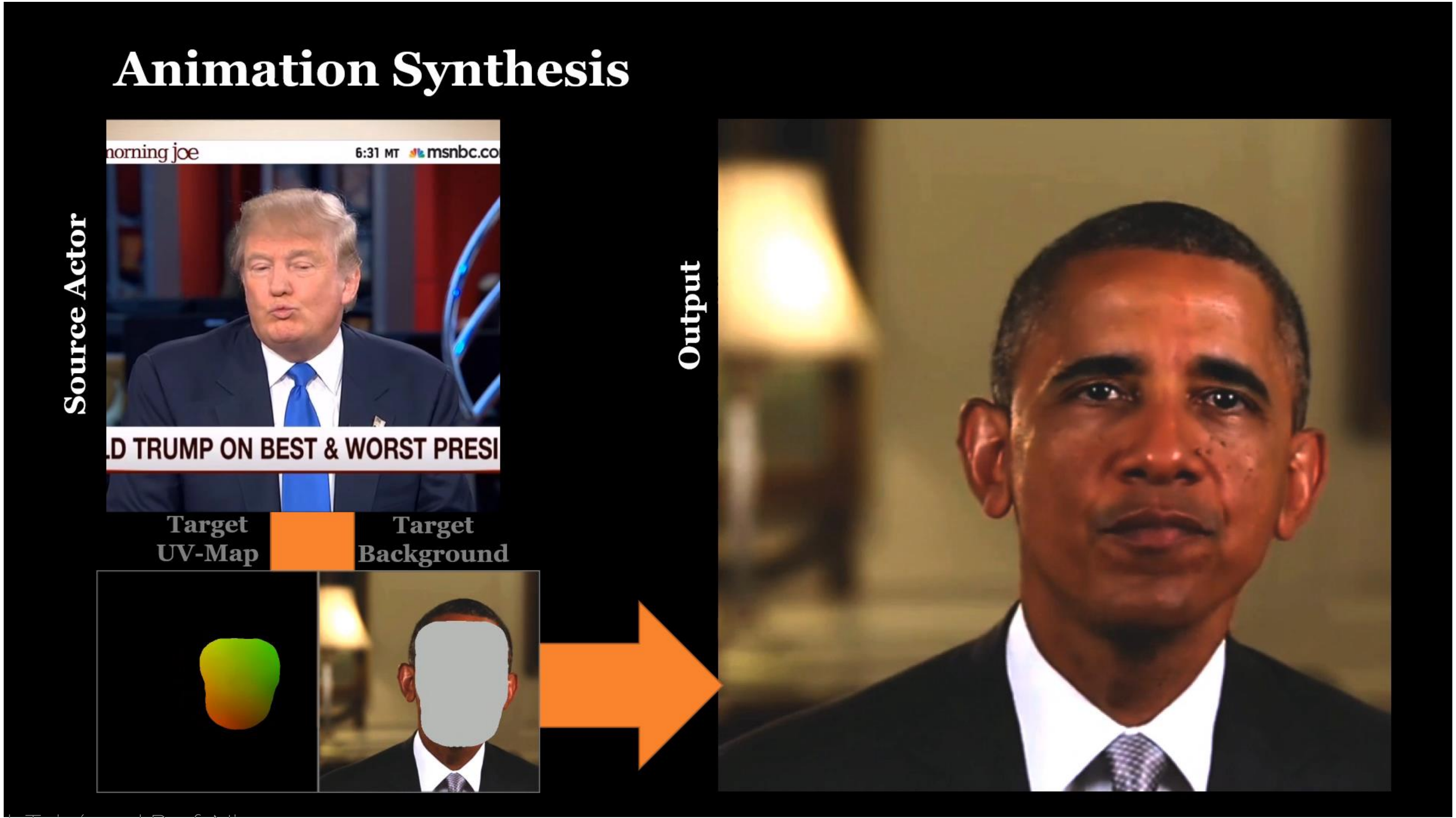


Siggraph’19 [Thies et al.]: Neural Textures

Scene Editing

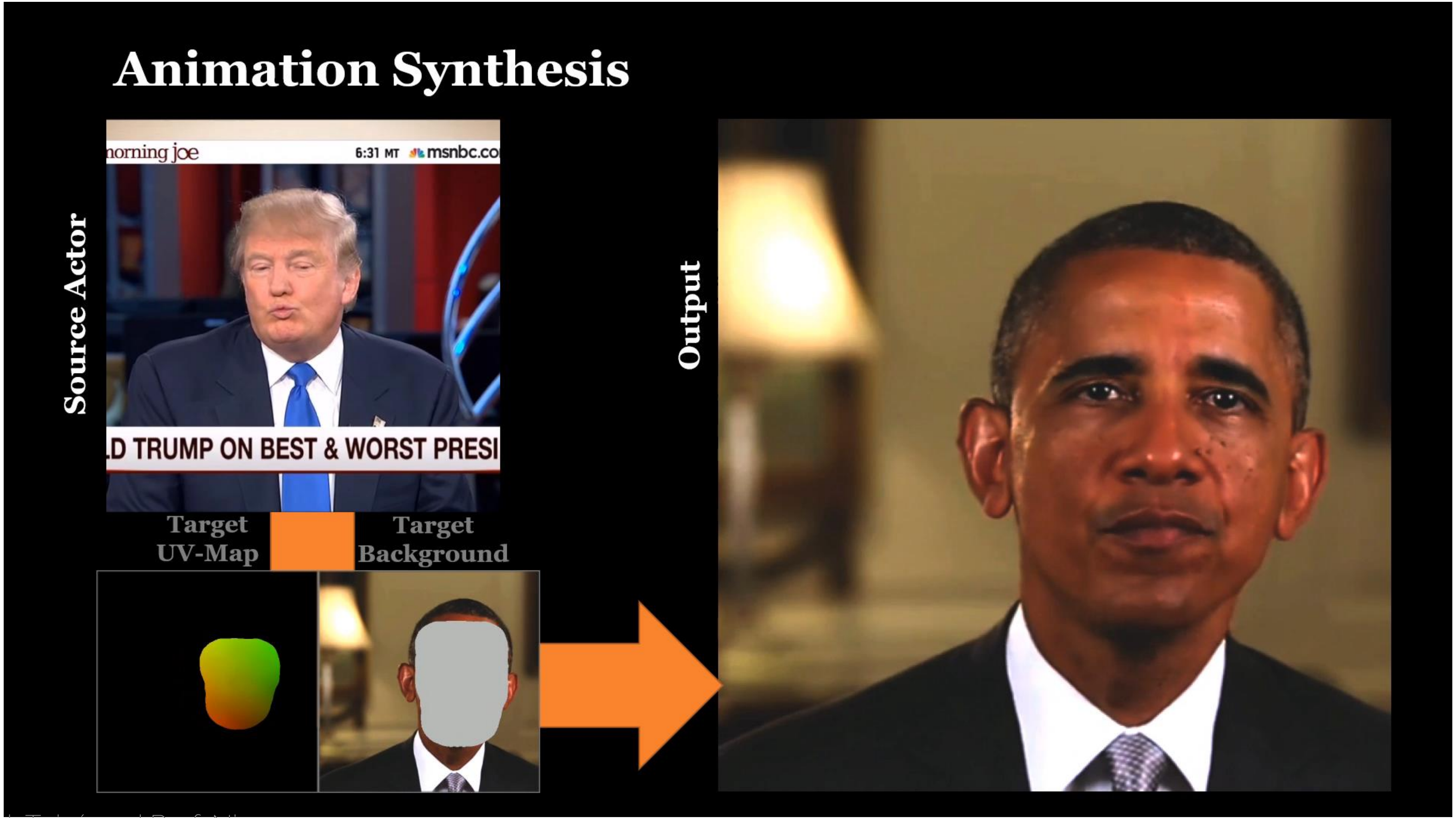
Siggraph'19 [Thies et al.]: Neural Textures

Facial Animation



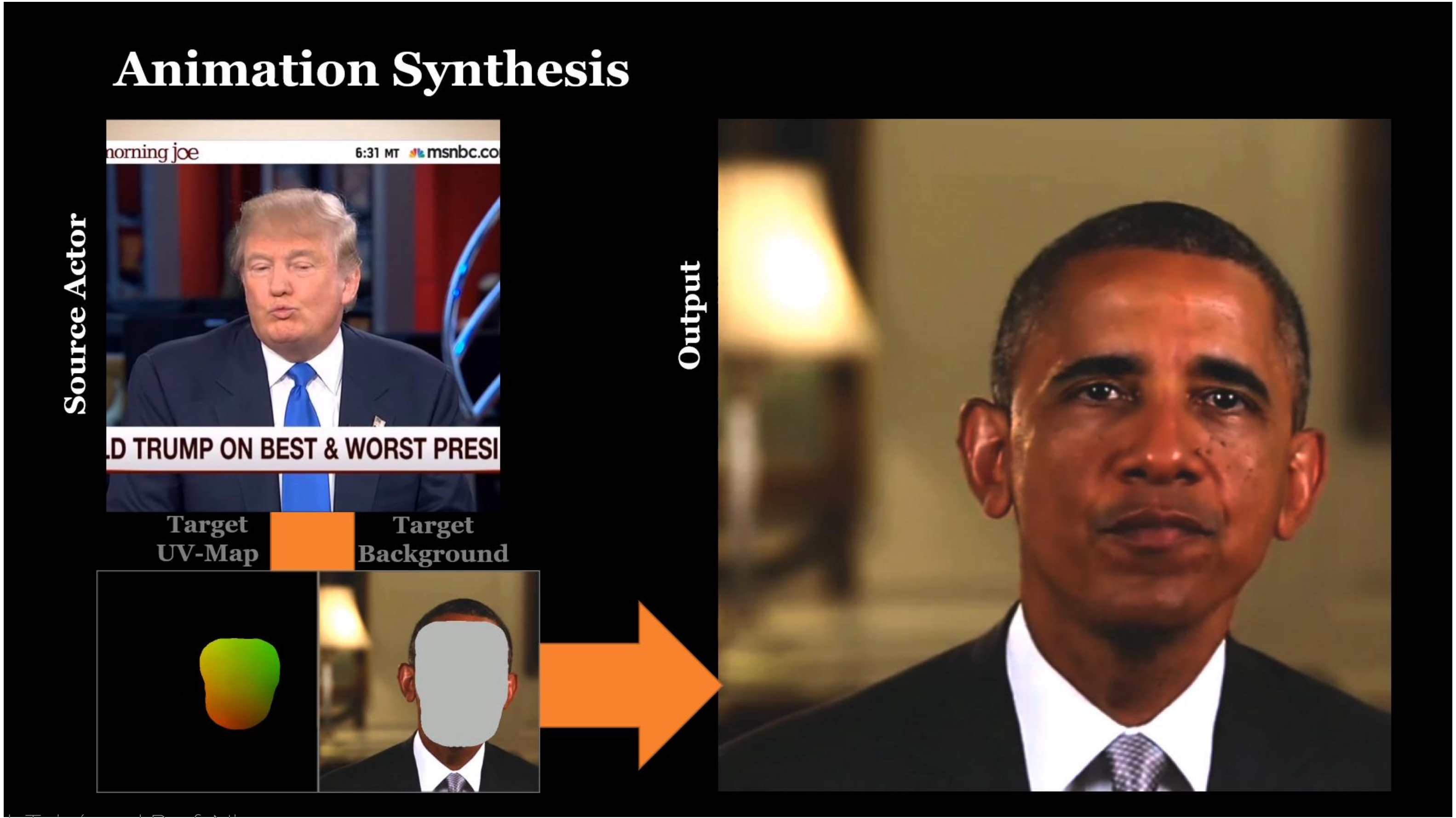
Siggraph'19 [Thies et al.]: Neural Textures

Facial Animation



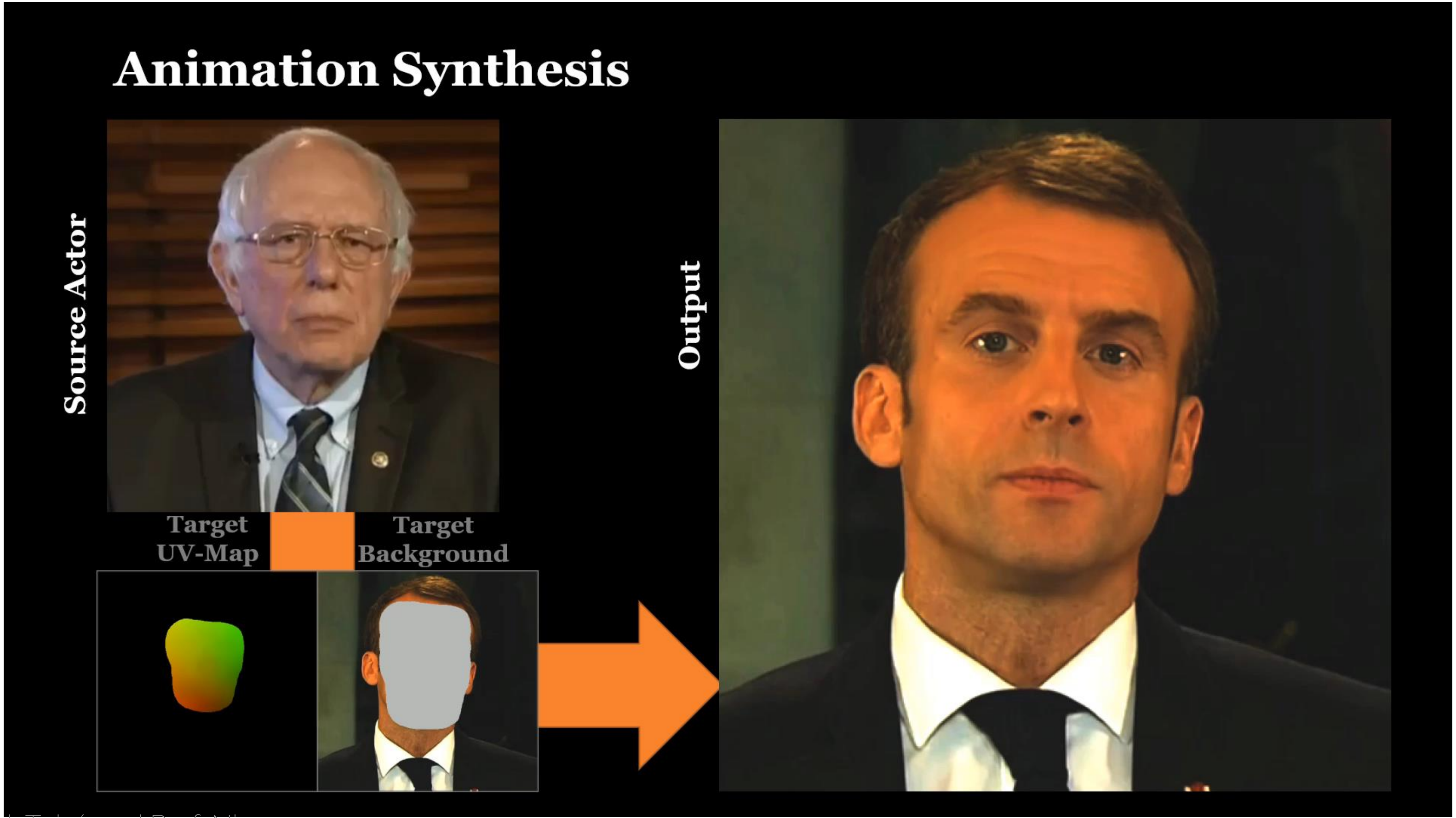
Siggraph'19 [Thies et al.]: Neural Textures

Facial Animation



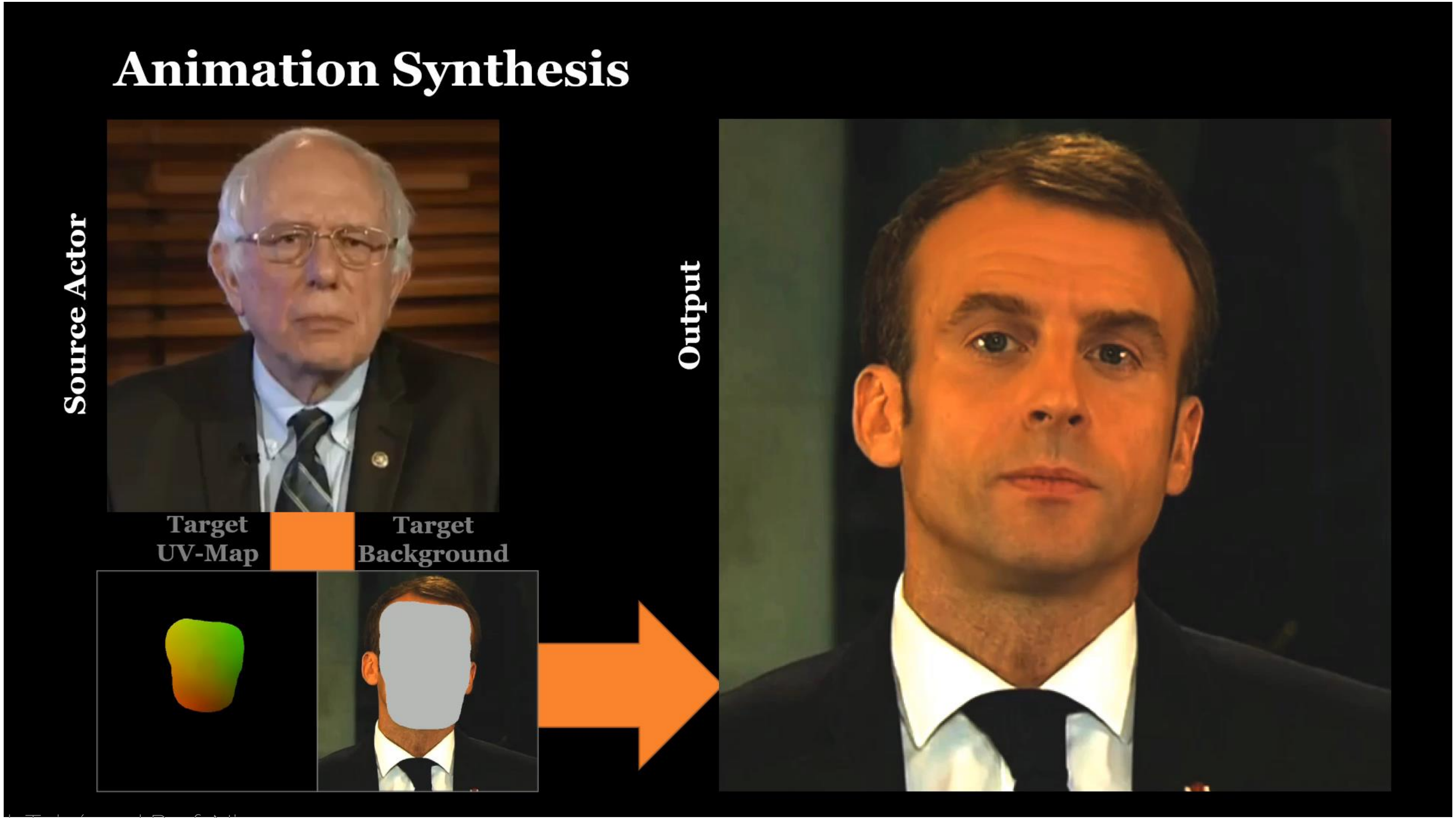
Siggraph'19 [Thies et al.]: Neural Textures

Facial Animation



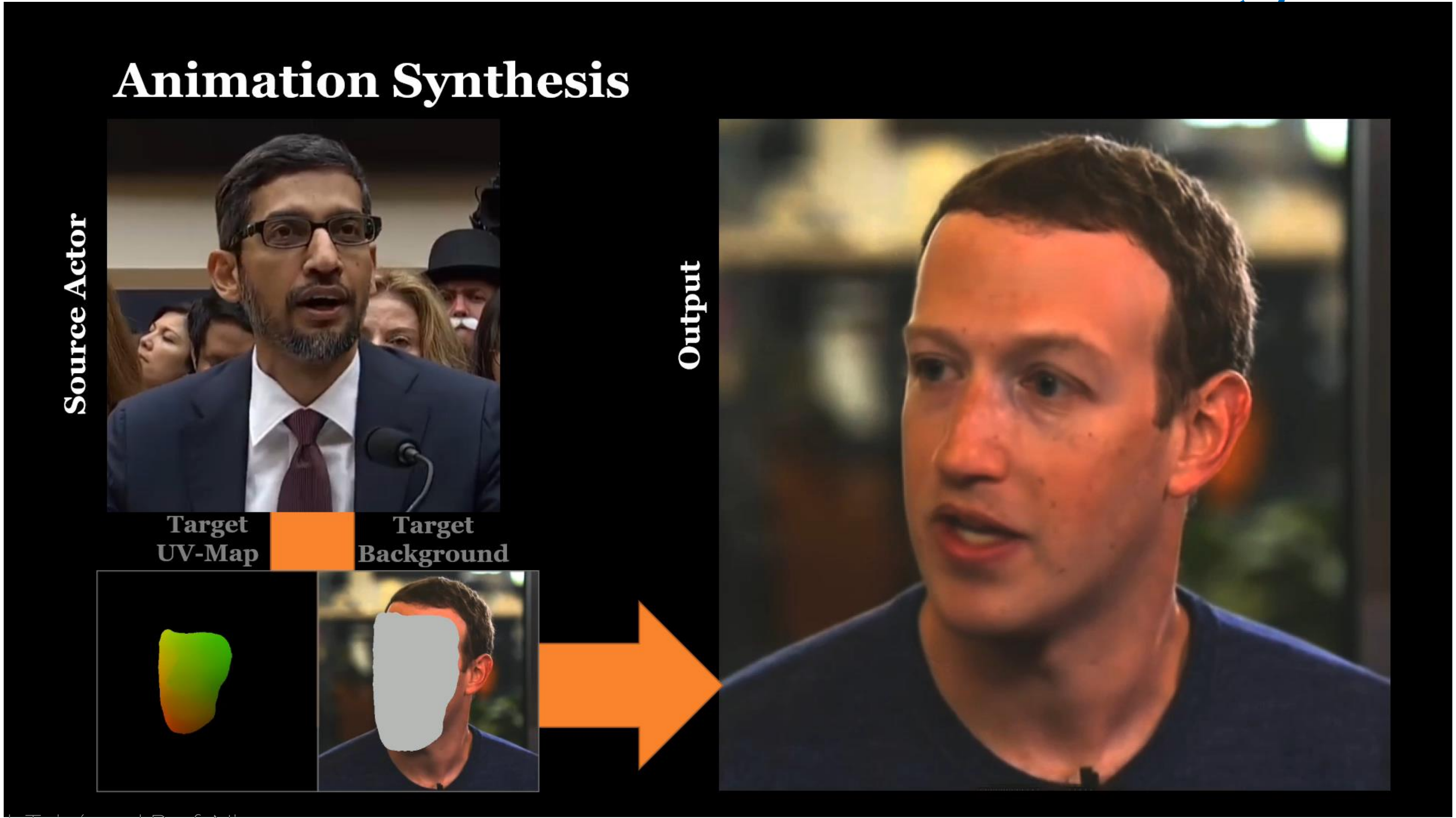
Siggraph'19 [Thies et al.]: Neural Textures

Facial Animation



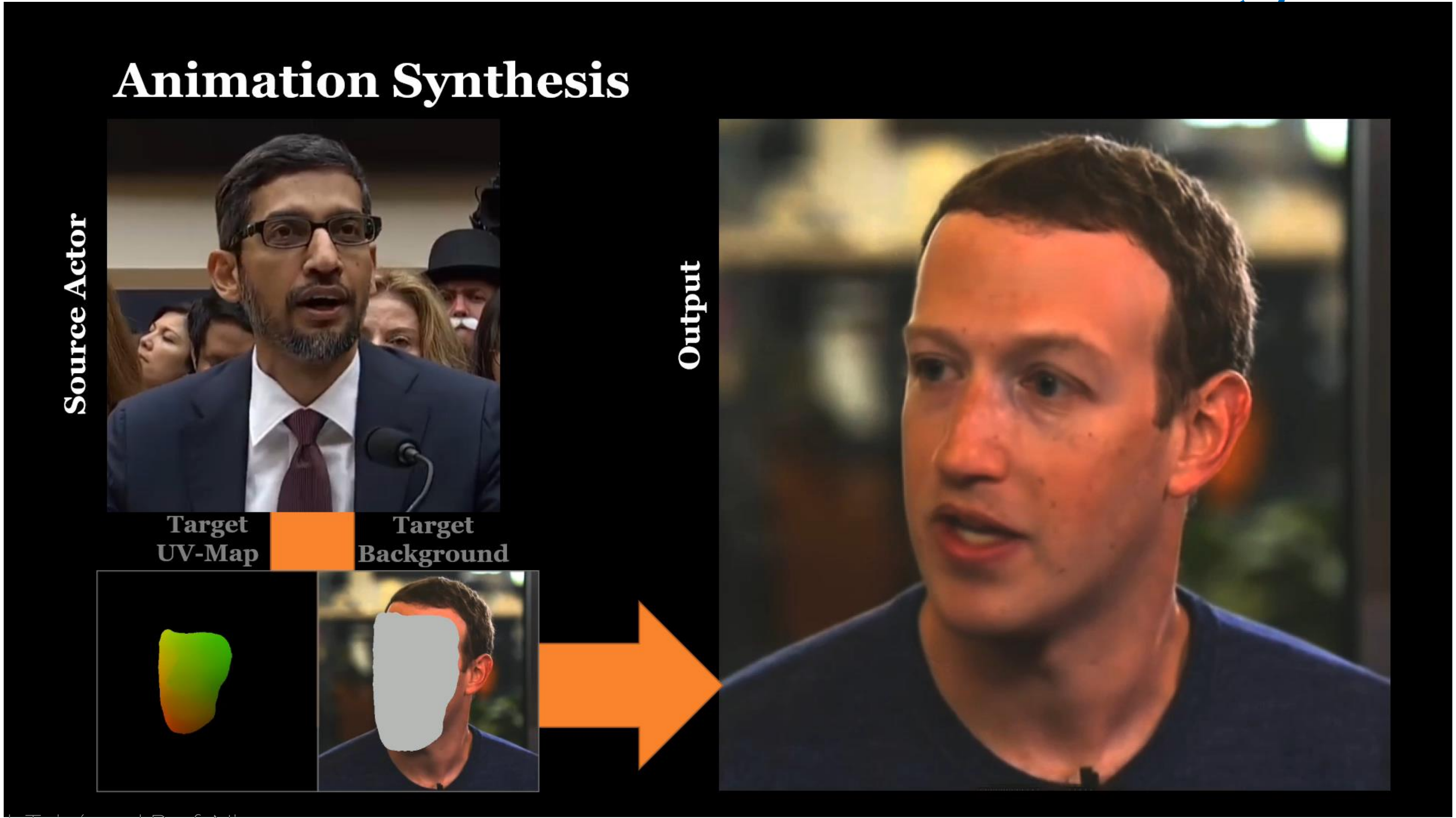
Siggraph'19 [Thies et al.]: Neural Textures

Deferred Neural Rendering



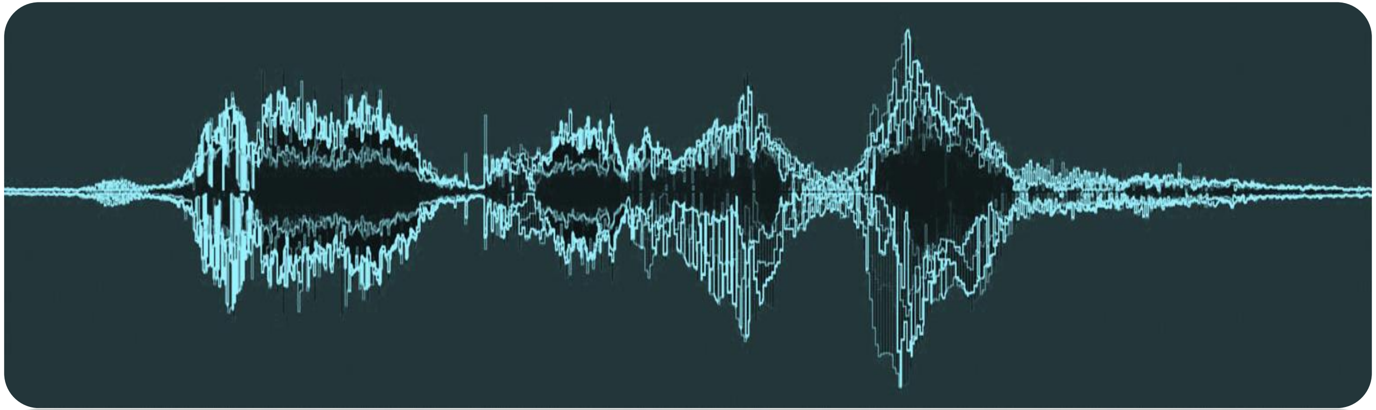
Siggraph'19 [Thies et al.]: Neural Textures

Deferred Neural Rendering

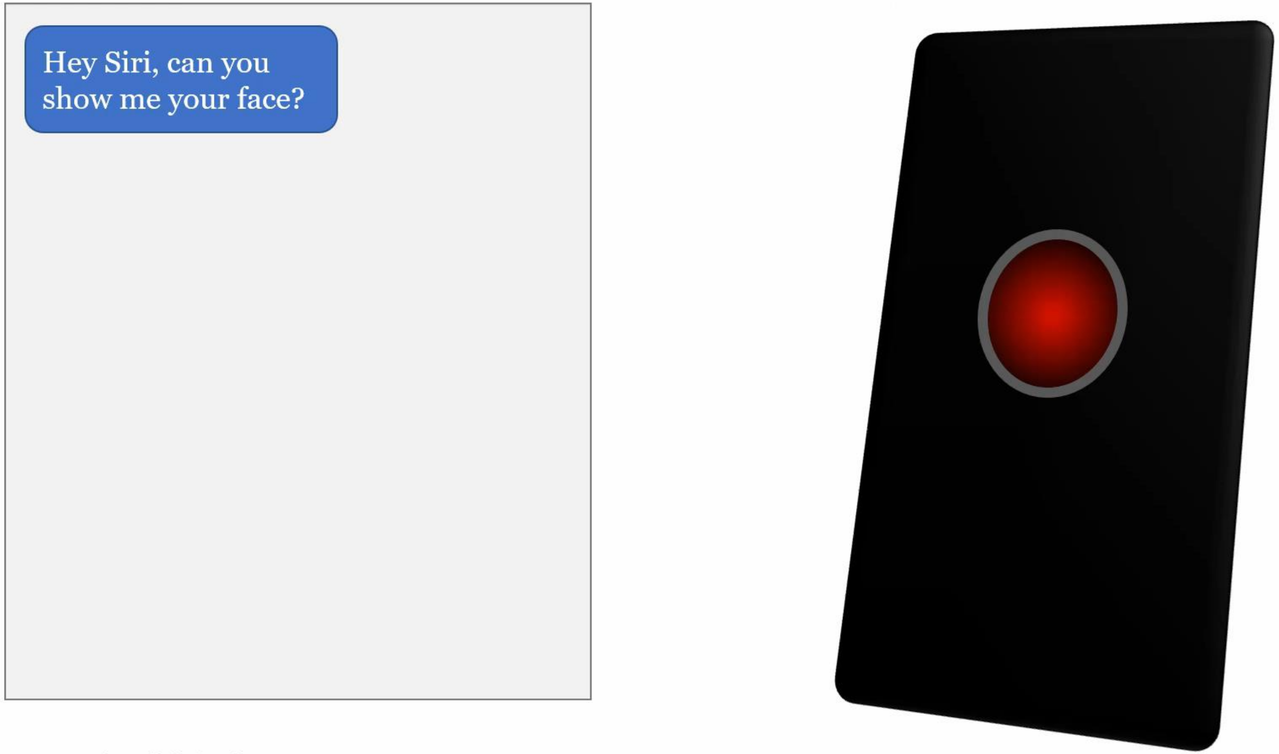


Siggraph'19 [Thies et al.]: Neural Textures

Neural Voice Puppetry



Neural Voice Puppetry



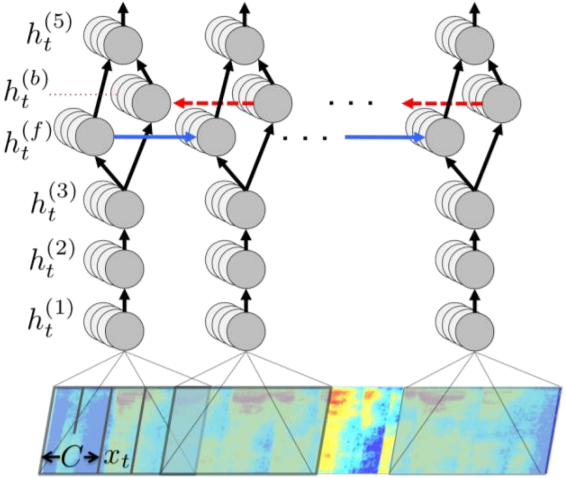
We use Siri as synonym for a digital assistant.

Neural Voice Puppetry

How does it work?

Pipeline Overview

Neural Voice Puppetry

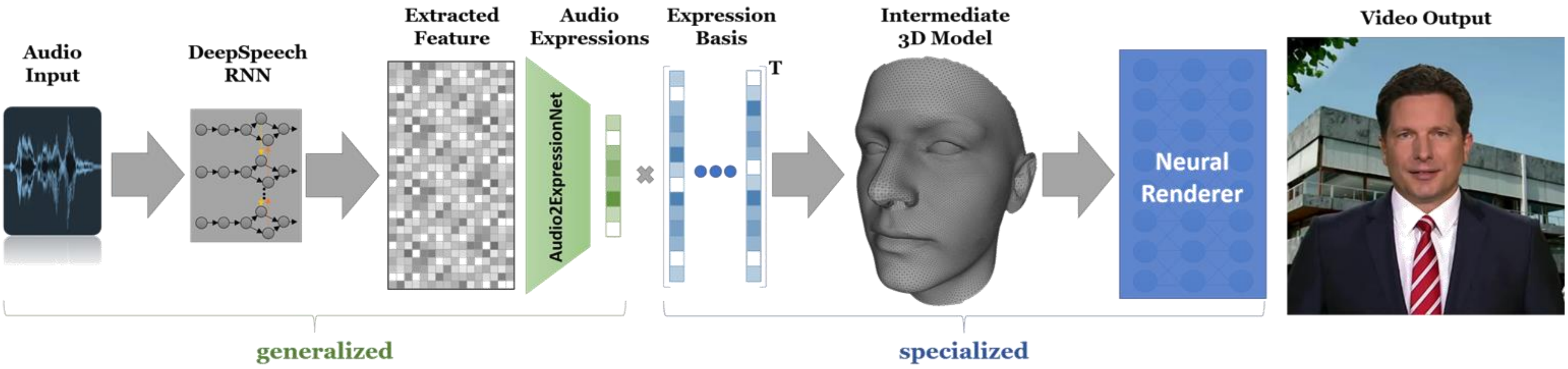


[Hannun et al.] DeepSpeech RNN

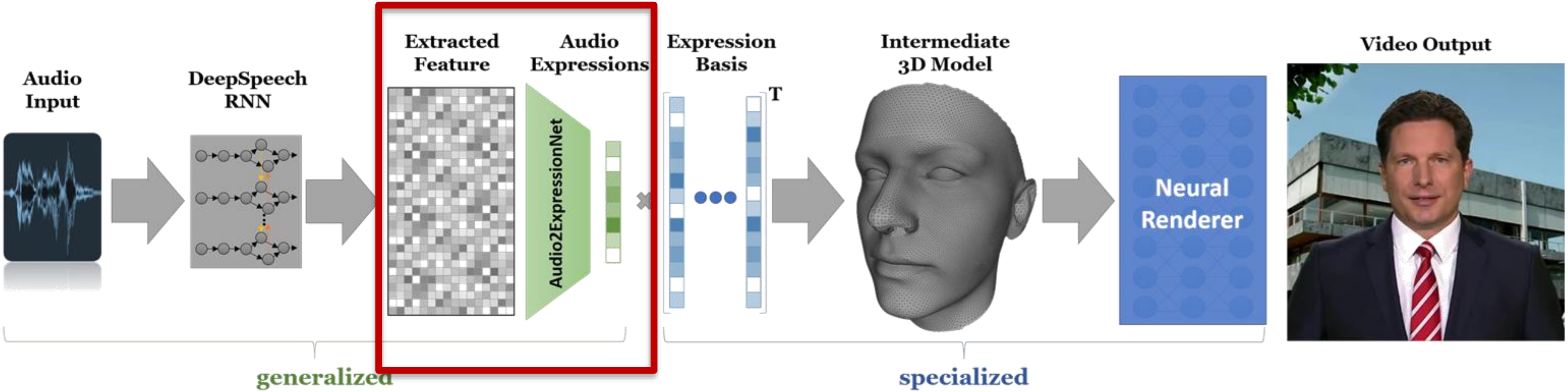
Output of the RNN of DeepSpeech:
- Logits of alphabet (|alphabet|=29)

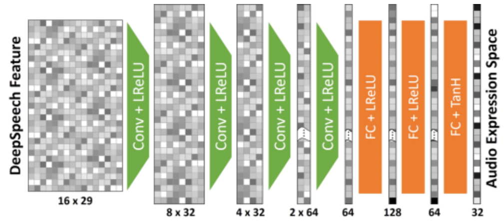
We use a time window (n=16)

Neural Voice Puppetry

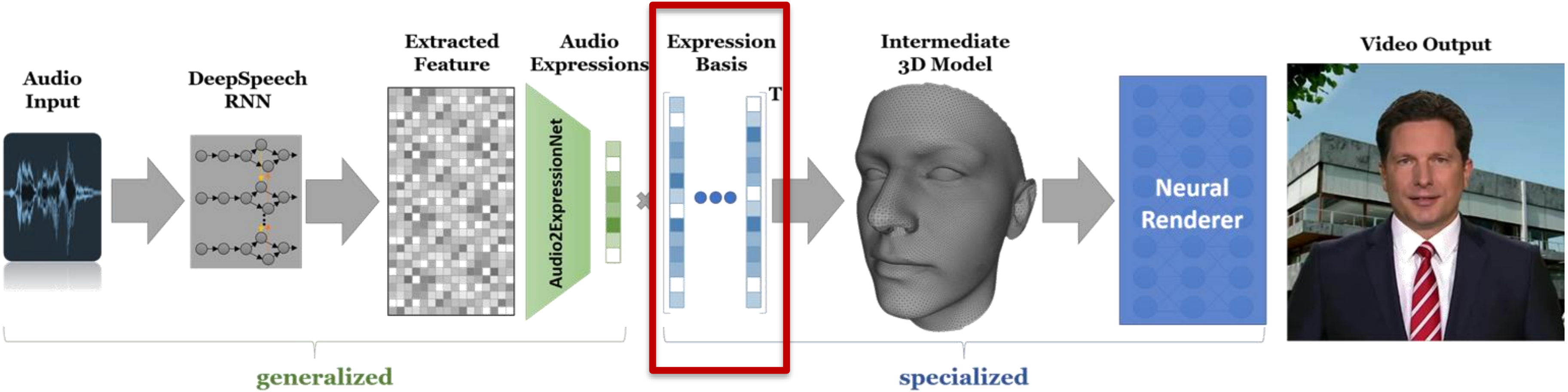


Neural Voice Puppetry





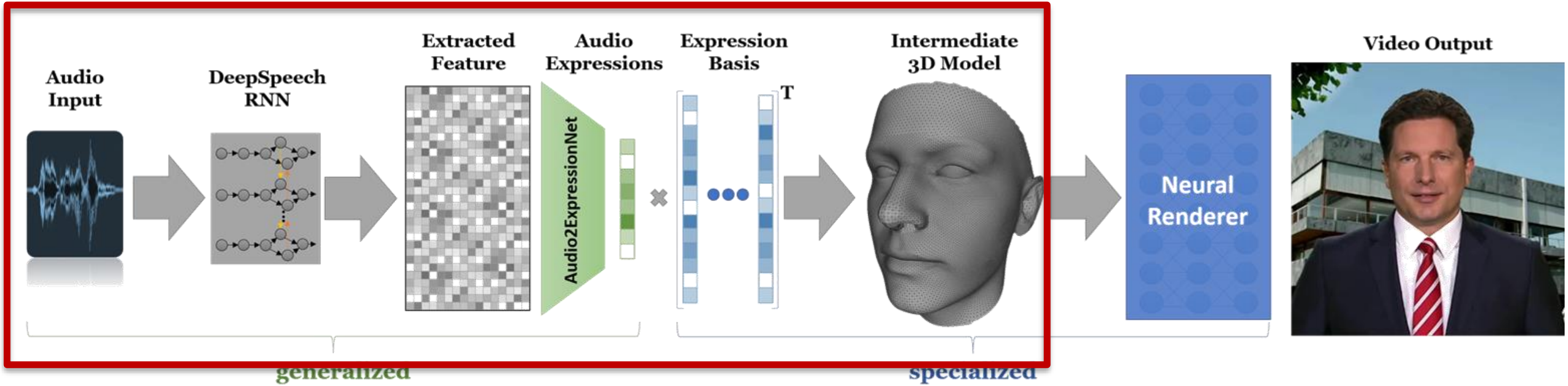
Neural Voice Puppetry



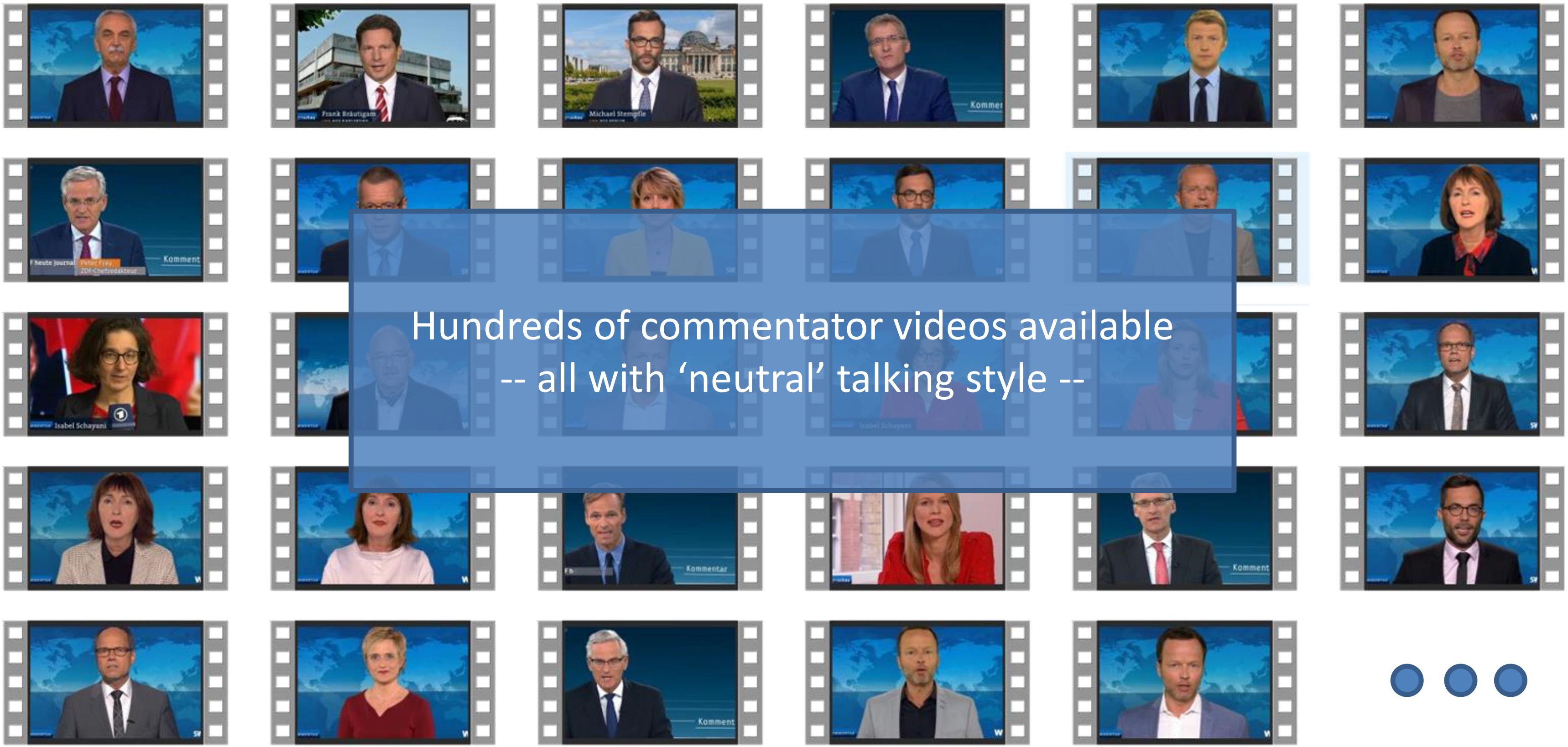
Person-specific
Blendshape Expression Model

Neural Voice Puppetry

Audio2Expression Training

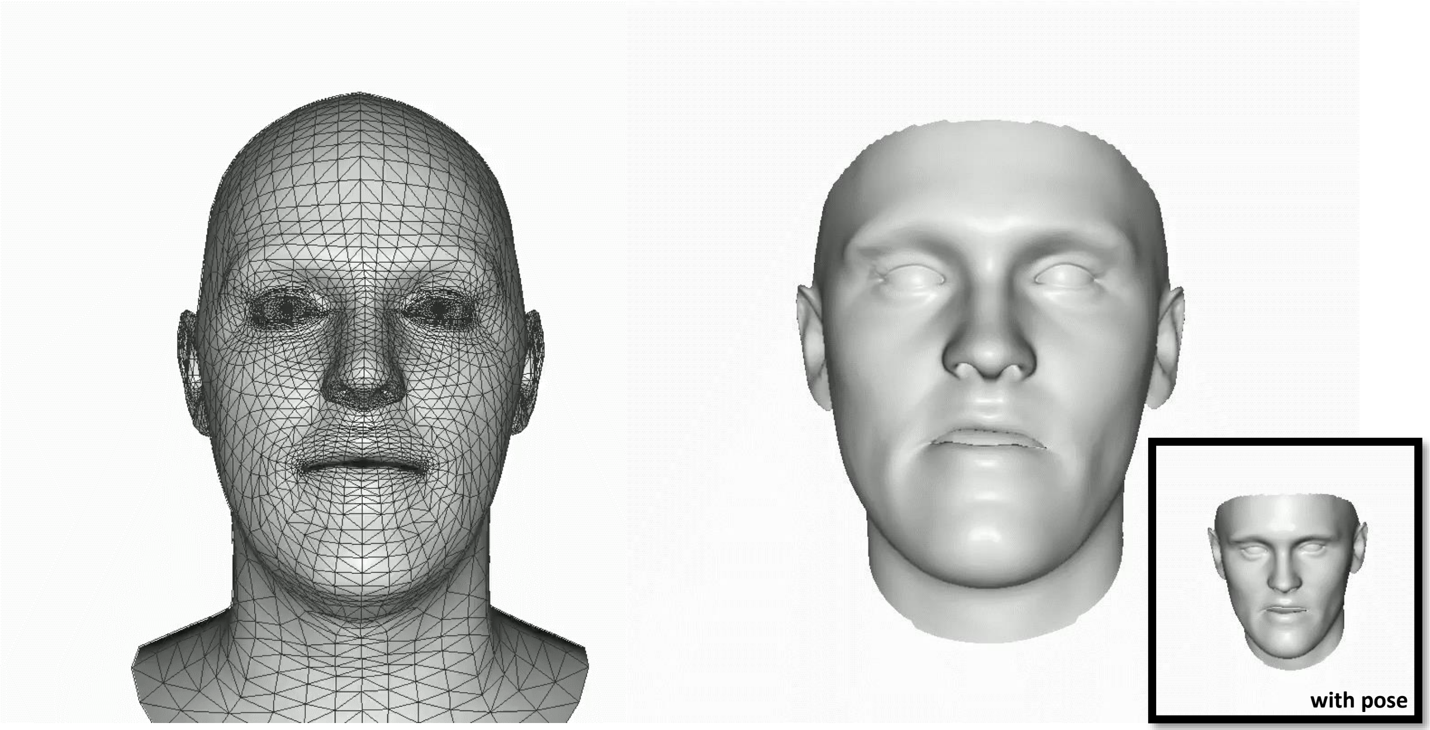


Neural Voice Puppetry



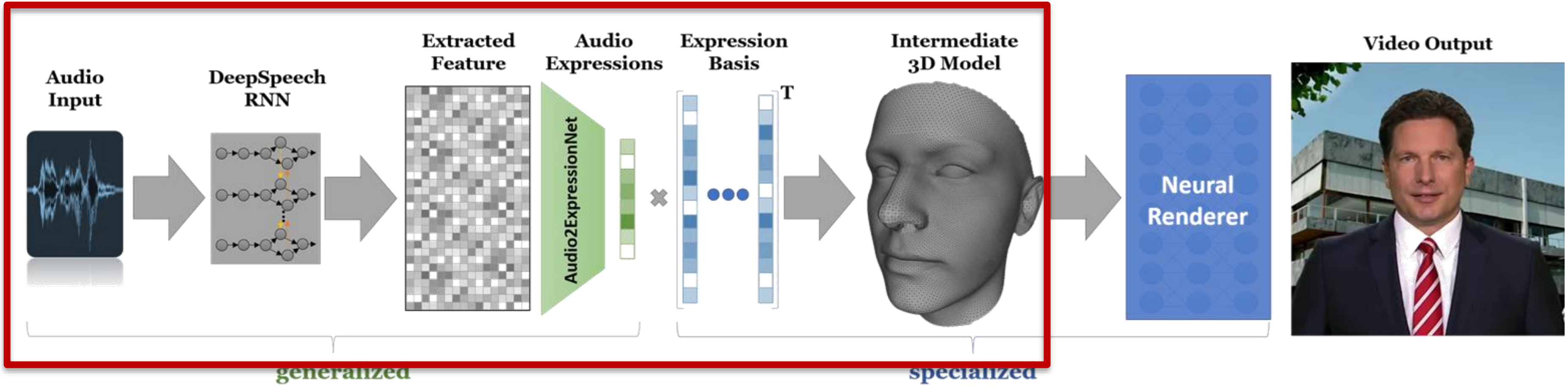
Hundreds of commentator videos available
-- all with ‘neutral’ talking style --

Neural Voice Puppetry

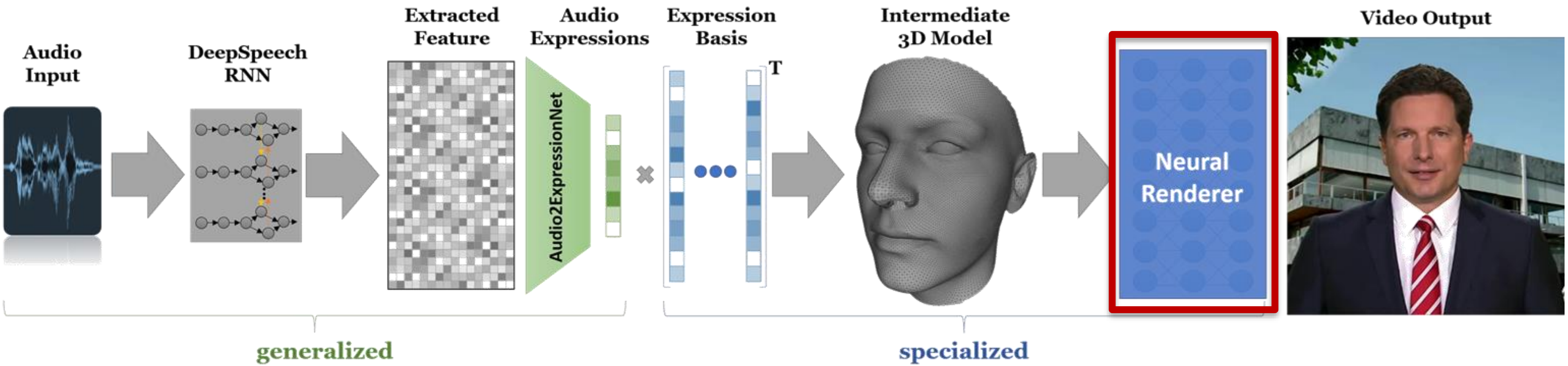


Flame Model **Basel Model**

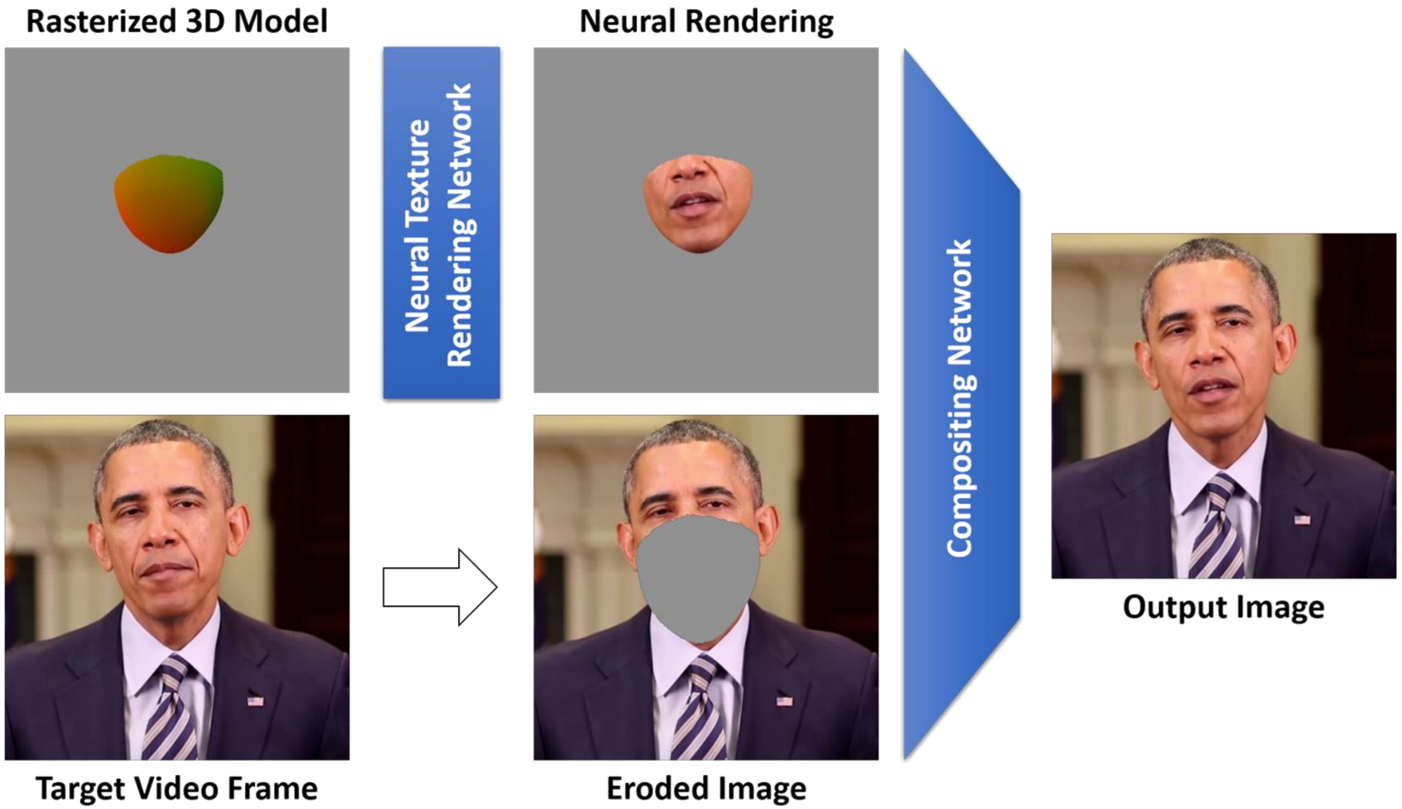
Neural Voice Puppetry



Neural Voice Puppetry



Neural Voice Puppetry



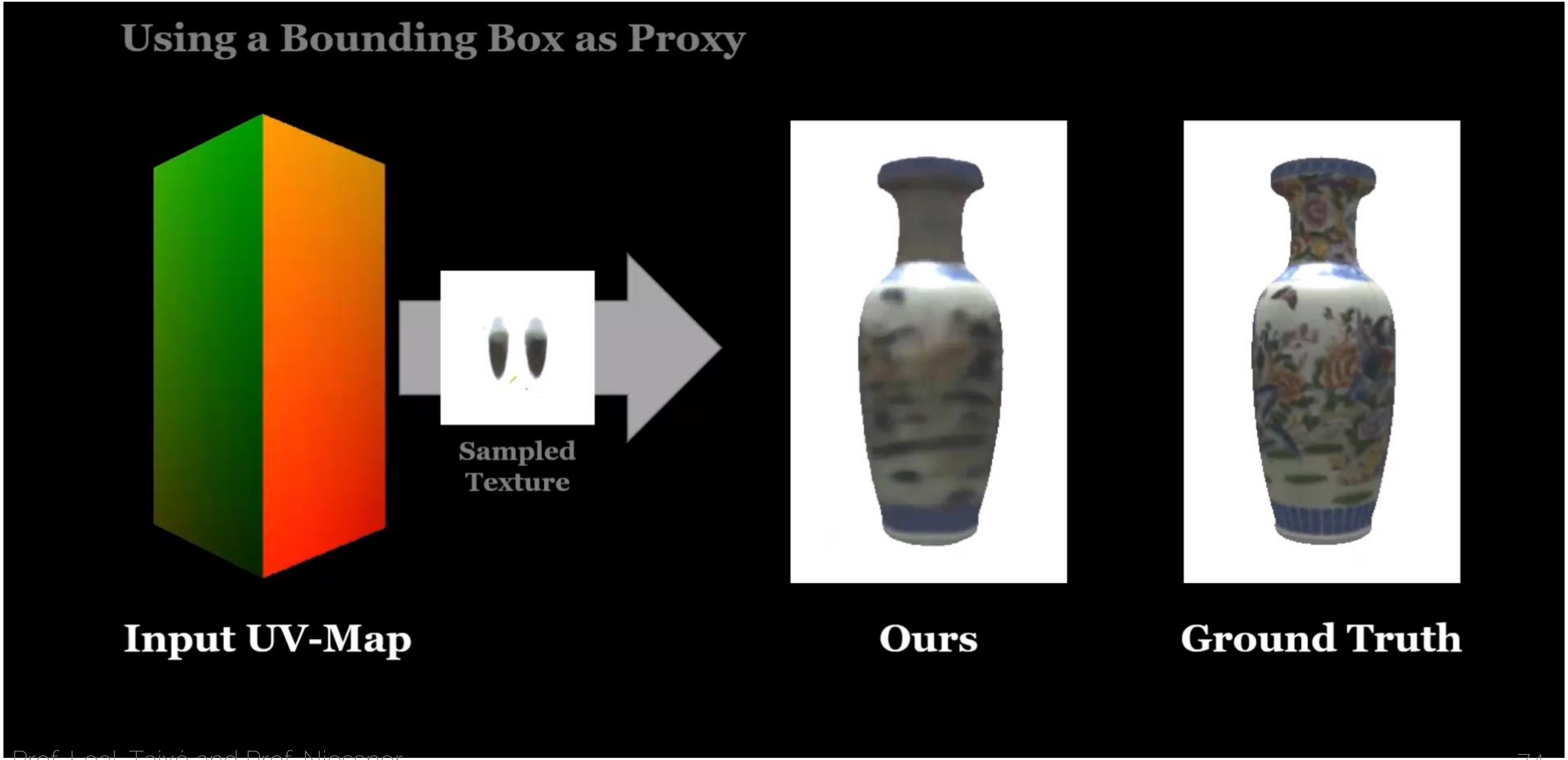
Neural Voice Puppetry

Big Open Challenges

Big Open Challenges

Photo-realistic Reconstruction

Big Open Challenges: How much can AI do?

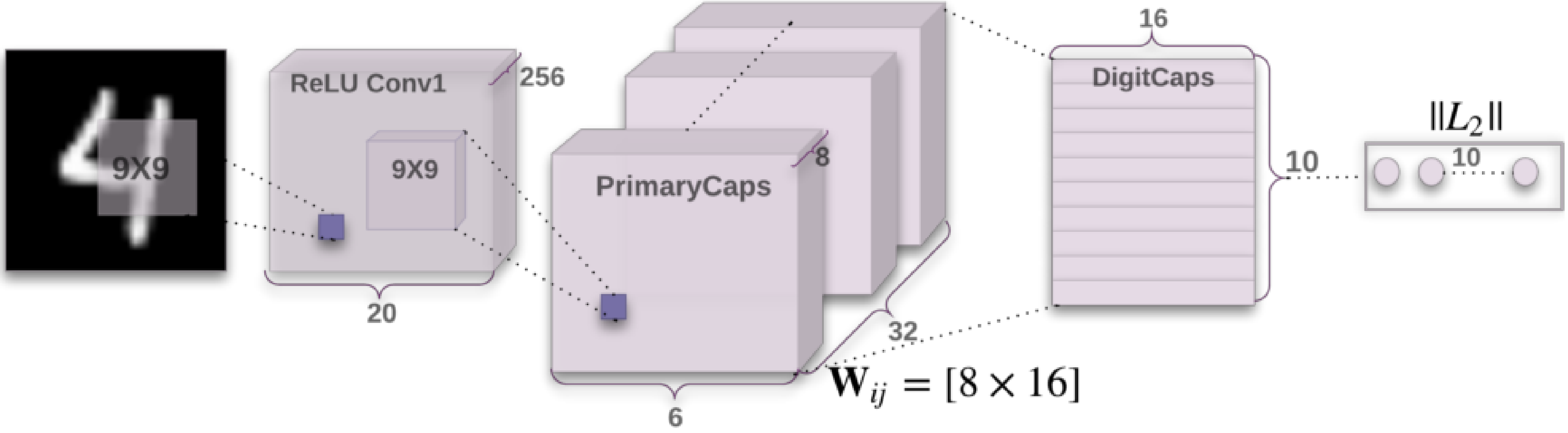


Siggraph'19 [Thies et al.]: Neural Textures

Big Open Challenges: 3D in Networks

Why learn 3D operations, such as transformations?

-> *differentiate known operators*



Capsule networks are motivated by *inverse graphics* [Sabour et al. 17]

State of the Art on Neural Rendering

A. Tewari[1⋆] O. Fried[2⋆] J. Thies[3⋆] V. Sitzmann[2⋆] S. Lombardi[4] K. Sunkavalli[5] R. Martin-Brualla[6] T. Simon[4] J. Saragih[4] M. Nießner[3] R. Pandey[6] S. Fanello[6] G. Wetzstein[2] J.-Y. Zhu[5] C. Theobalt[1] M. Agrawala[2] E. Shechtman[5] D. B Goldman[6] M. Zollhöfer[4]

[1]MPI Informatics [2]Stanford University [3]Technical University of Munich [4]Facebook Reality Labs [5]Adobe Research [6]Google Inc [⋆]Equal contribution.

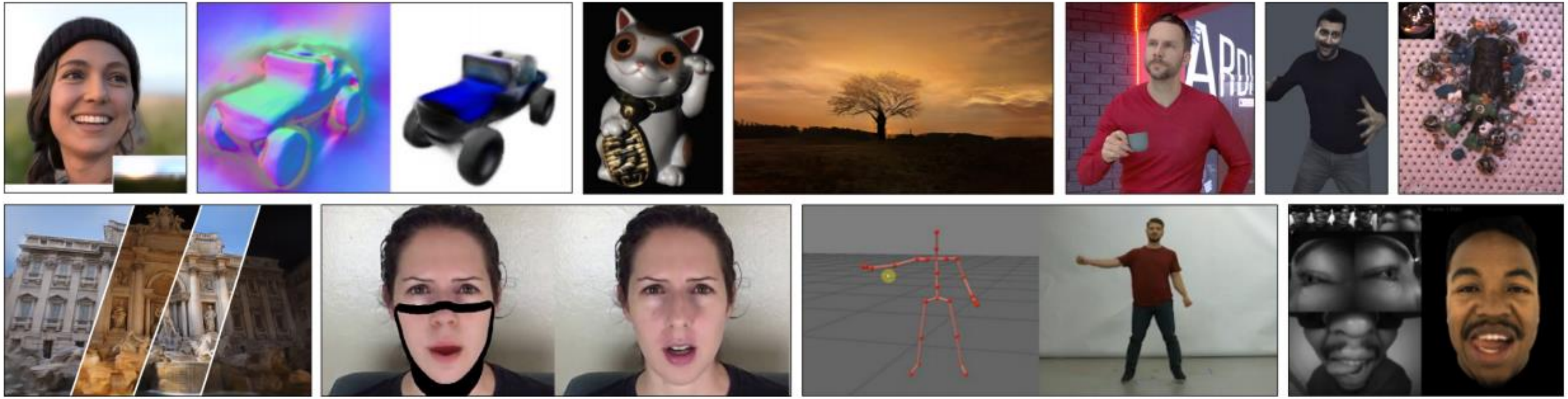


Figure 1: *Neural renderings of a large variety of scenes. See Section 6 for more details on the various methods. Images from [SBT* 19, SZW19, XBS* 19, KHM17, GLD* 19, MBPY* 18, XSHR18, MGK* 19, FTZ* 19, LXZ* 19, WSS* 19].*

See you next week ☺

Some Extra Slides:

Neural Voice Puppetry

Comparisons

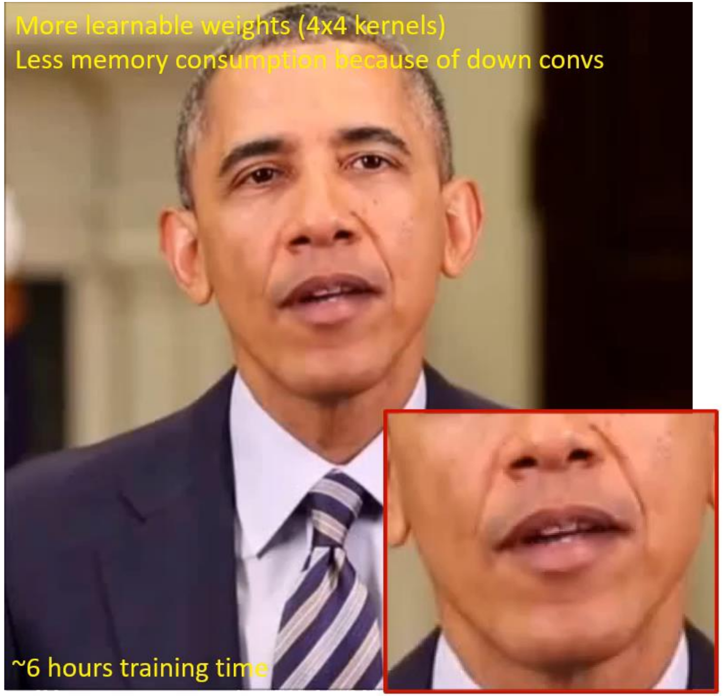
Model-based Methods

Neural Voice Puppetry

Comparisons

2D-based Methods

Neural Voice Puppetry



Strided Convolutions

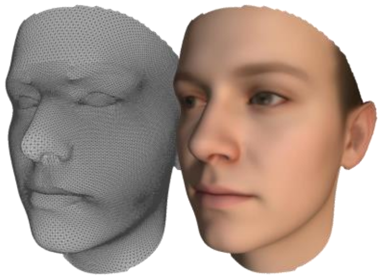
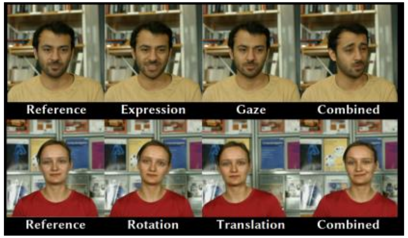
(classical U-Net, 5 down & up convs, kernel size 4)

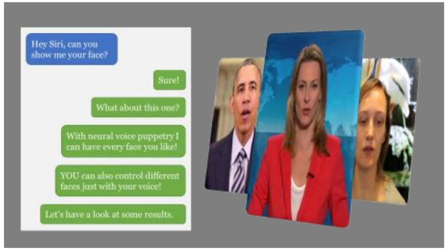
Dilated Convolutions

(U-Net, dilated instead of strided convs increasing dilation per layer, kernel size 3)

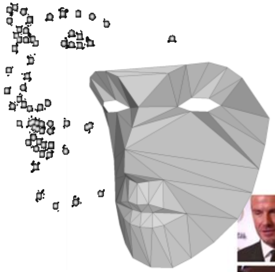
Facial Reenactment

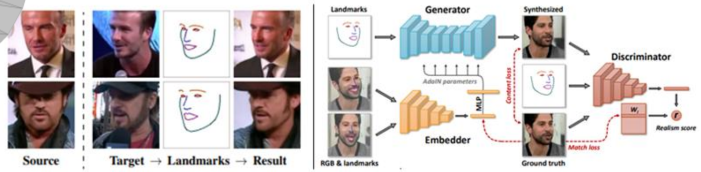
Dense



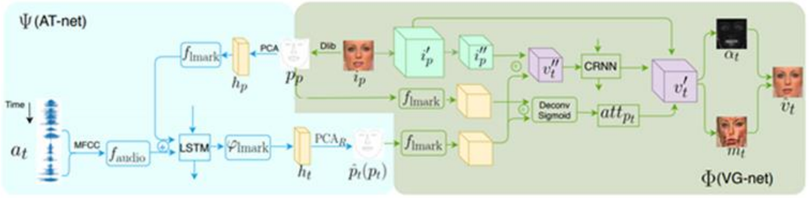
Sparse





Few-Shot Adversarial Learning of Realistic Neural Talking Head Models

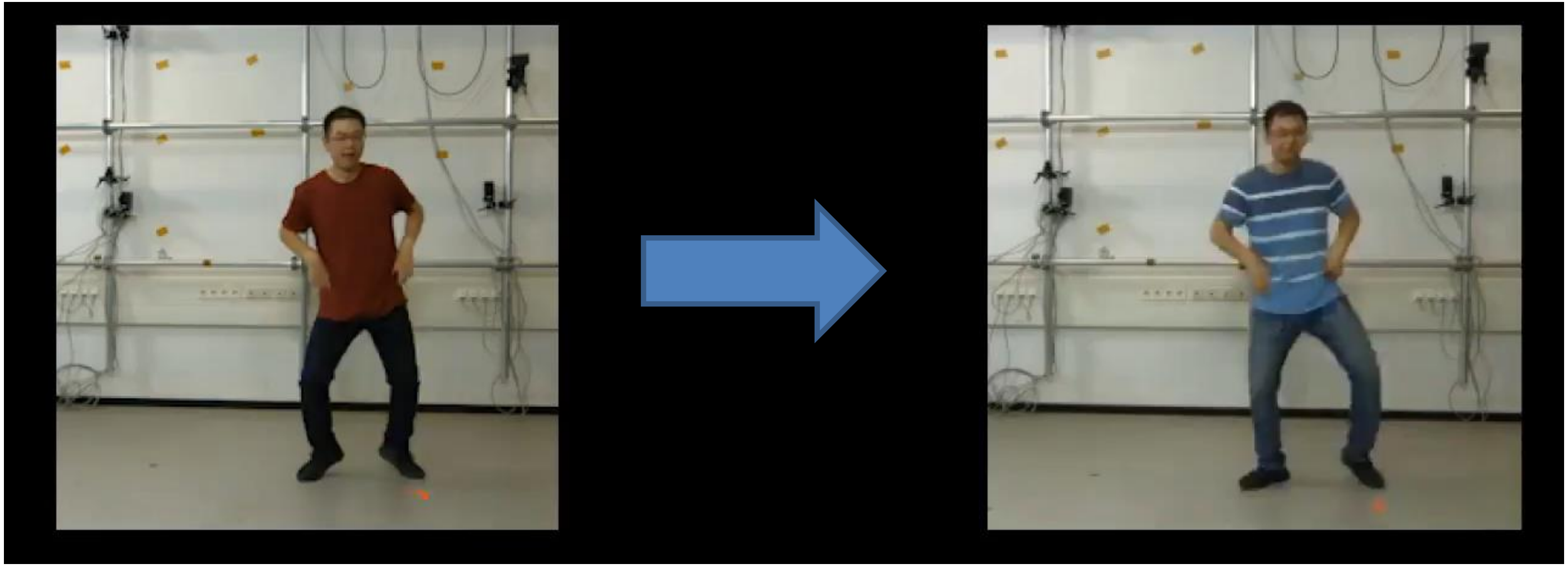
Egor Zakharov[1,2] Aliaksandra Shysheya[1,2] Egor Burkov[1,2] Victor Lempitsky[1,2]
[1]Samsung AI Center, Moscow [2]Skolkovo Institute of Science and Technology



Hierarchical Cross-Modal Talking Face Generation with Dynamic Pixel-Wise Loss

Lele Chen Ross K. Maddox Zhiyao Duan Chenliang Xu
University of Rochester, USA

Neural Rendering and Reenactment of Human Actor Videos

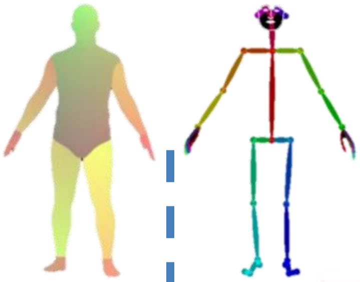


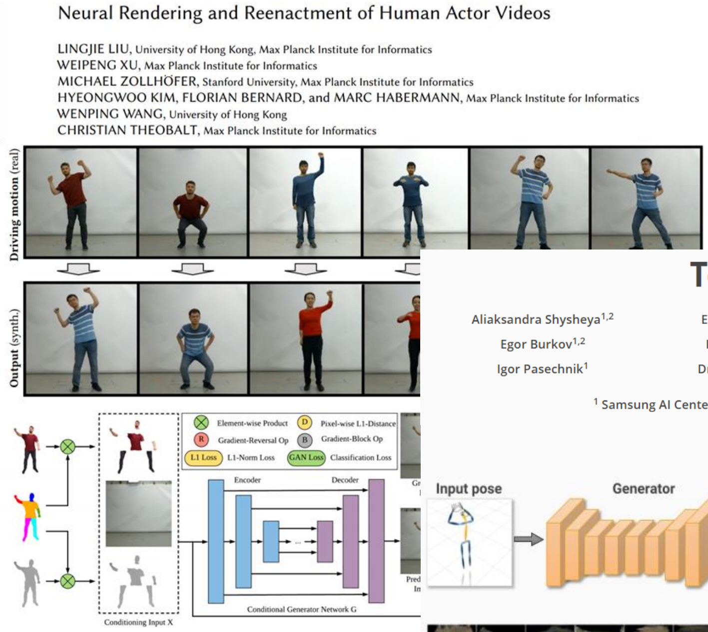
[Liu et al. 19] Neural Rendering and Reenactment of Human Actor Videos

Body Reenactment

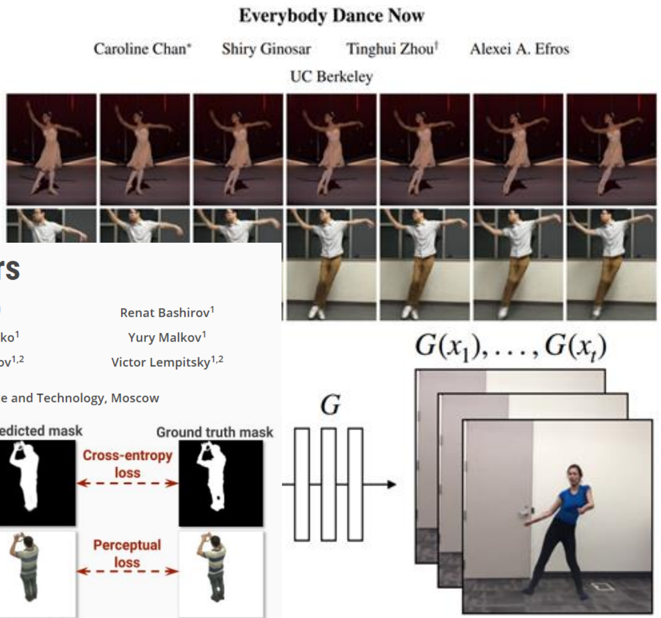
Dense

Sparse



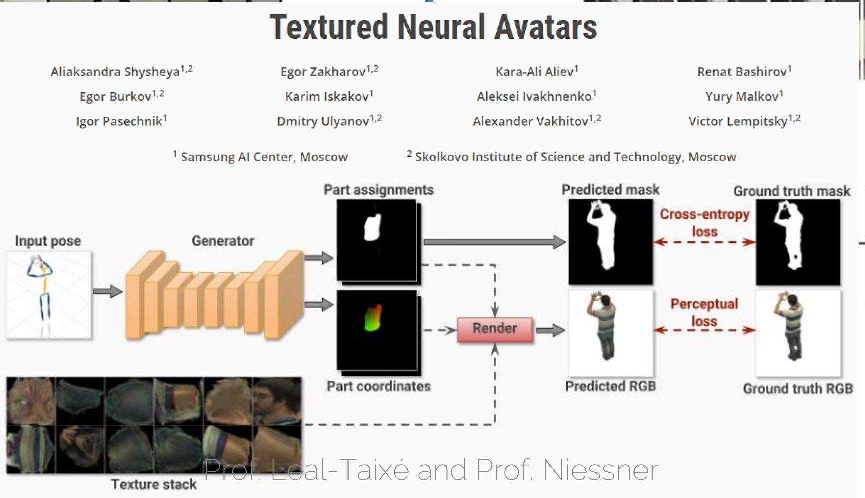

Neural Rendering and Reenactment of Human Actor Videos

LINGJIE LIU, University of Hong Kong, Max Planck Institute for Informatics
WEIPENG XU, Max Planck Institute for Informatics
MICHAEL ZOLLHÖFER, Stanford University, Max Planck Institute for Informatics
HYEONGWOO KIM, FLORIAN BERNARD, and MARC HABERMANN, Max Planck Institute for Informatics
WENPING WANG, University of Hong Kong
CHRISTIAN THEOBALT, Max Planck Institute for Informatics


Everybody Dance Now

Caroline Chan* Shiry Ginosar Tinghui Zhou† Alexei A. Efros

UC Berkeley


Textured Neural Avatars

Aliaksandra Shysheya[1,2] Egor Zakharov[1,2] Kara-Ali Aliev[1] Renat Bashirov[1]
Egor Burkov[1,2] Karim Iskakov[1] Aleksei Ivakhnenko[1] Yury Malkov[1]
Igor Pasechnik[1] Dmitry Ulyanov[1,2] Alexander Vakhitov[1,2] Victor Lempitsky[1,2]

[1] Samsung AI Center, Moscow [2] Skolkovo Institute of Science and Technology, Moscow

Open Challenges

- Motion Capturing
- Person-specific Motions/Expressions
- Temporal Stability
- Image Quality