Neural Rendering
Rendering

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

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

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

Prof. Leal-Taixé and Prof. Niessner
Photo-realistic Image Synthesis

The Rendering Equation [Kajiya 86]

\[ L_o(x, \omega_o, \lambda, t) = L_e(x, \omega_o, \lambda, t) + \int_{\Omega} f_r(x, \omega_i, \omega_o, \lambda, t) L_i(x, \omega_i, \lambda, t) (\omega_i \cdot n) \, d\omega_i \]
Need 3D Content for Rendering

Geometry

Textures

Material & Lighting

Prof. Leal-Taixé and Prof. Niessner
3D Digitization

Computer Graphics

Computer Vision

Prof. Leal-Taixé and Prof. Niessner
Traditional Graphics vs Deep Learning

3D Model + Textures + Shading -> Synthetic Image

Generative Adversarial Networks

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

\[
J^{(G)} = -J^{(D)}
\]

Prof. Leal-Taixé and Prof. Niessner
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

• 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
Deep Voxels

Simplified overview for novel view synthesis

Source View $R, t$

Lifting Layer $2D ightarrow 3D$

3D U-Net

Projection Layer $3D ightarrow 2D$

Target View $R, t$

Renderin $g$

2D U-Net

Output

Prof. Leal-Taixé and Prof. Niessner

[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

Prof. Leal-Taixé and Prof. Niessner
Deep Voxels

Best Baseline: Pix2Pix [Isola et al. 2017]
Deep Voxels

Pix2Pix [Isola et al. 2017]  DeepVoxels (Ours)

Prof. Leal-Taixé and Prof. Niessner
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
### 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)

Prof. Leal-Taixé and Prof. Niessner
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 Sizmann (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
- “True 3D”
- High quality
- Generalizes

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
- "True 3D"
- High quality
- Generalizes

**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

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
- Requires good SFM

**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

**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)
Scene Representation Networks
Sitzmann et al., Neurips 2019

Scene Representation
ReLU MLP

Renderer
Generalized (learned) sphere-tracing

Generalization
Hypernetwork

$\mathbb{R}^3 \rightarrow \mathbb{R}^n$
Scene Representation Networks
Sitzmann et al., Neurips 2019

Full 3D Reconstruction from single image!
NERF: Neural Radiance Fields
Mildenhall et al., arXiv 2020

Scene Representation
ReLU MLP + Positional Encoding
View Direction

Renderer
Volumetric, stratified sampling

Generalization
None.

$\mathbb{R}^6 \rightarrow \mathbb{R}^3$
NERF: Neural Radiance Fields
Mildenhall et al., arXiv 2020

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

Scene Representation

Multi-Plane Images
Voxelgrids
Image-based
Point Clouds
Implicit Function

Renderer

(Alpha) compositing
Volumetric Ray-based
Rasterization
Splatting
Sphere-Tracing Volumetric

Pros

Fast rendering
“True 3D”
High quality
High quality
High quality
High quality
May generalize!

Cons

Only 2.5D Size
No reconstruction priors
Memory $O(n^3)$
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
Neural Textures: Features on 3D Mesh

Prof. Leal-Taixé and Prof. Niessner
Neural Textures: Features on 3D Mesh

3D Geometry

Neural Texture

Rendering
3D → 2D

View $R, t$

UV-Map

Sampled Texture

Renderer

Output Image

Prof. Leal-Taixé and Prof. Niessner

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

Deferred Renderer

Handcrafted "Feature Maps"

Prof. Leal-Taixé and Prof. Niessner

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

Prof. Leal-Taixé and Prof. Niessner

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

3D Geometry

UV-Map

Renderer

Output Image

Neural Texture

Sampled Texture

Prof. Leal-Taixé and Prof. Niessner

Siggraph’19 [Thies et al.]: Neural Textures
Neural Textures: Features on 3D Mesh

Novel View Synthesis

Scene Editing

Animation Synthesis

Prof. Leal-Taixé and Prof. Niessner

Siggraph’19 [Thies et al.]: Neural Textures
Novel View-Point Synthesis

Input UV-Map

Ours

Siggraph’19 [Thies et al.]: Neural Textures
Novel View-Point Synthesis

Ground Truth

Ours

Prof. Leal-Taixé and Prof. Niessner

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

Input Sequence

Geometry Editing

Prof. Leal-Taixé and Prof. Niessner
Scene Editing

Prof. Leal-Taixé and Prof. Niessner

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

Prof. Leal-Taixé and Prof. Niessner

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

Prof. Leal-Taixé and Prof. Niessner

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

Prof. Leal-Taixe and Prof. Niessner

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

Animation Synthesis

Source Actor

Target UV-Map

Target Background

Output

Prof. Leal-Taixé and Prof. Niessner

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

We use Siri as synonym for a digital assistant.
How does it work?

Pipeline Overview
Neural Voice Puppetry

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

We use a time window (n=16)
Neural Voice Puppetry

Audio Input → DeepSpeech RNN → Extracted Feature → Audio Expressions

Audio2ExpressionNet: generalized

Expression Basis: specialized

Intermediate 3D Model → Neural Renderer → Video Output

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Audio Input → DeepSpeech RNN → Extracted Feature → Audio Expressions → Expression Basis → Intermediate 3D Model → Neural Renderer → Video Output

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Neural Voice Puppetry

Person-specific Blendshape Expression Model

Audio Input → DeepSpeech RNN → Extracted Feature → Audio Expressions → Expression Basis → Intermediate 3D Model → Video Output

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Audio2Expression Training

Audio Input → DeepSpeech RNN → Extracted Feature → Audio Expressions → Expression Basis → Intermediate 3D Model → Neural Renderer → Video Output

- Generalized
- Specialized

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

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

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Flame Model

Basel Model

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Image description:
- Audio Input
- DeepSpeech RNN
- Extracted Feature
- Audio Expressions
- Expression Basis
- Intermediate 3D Model
- Neural Renderer
- Video Output

Text:
Neural Voice Puppetry

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry

Rasterized 3D Model

Neural Texture Rendering Network

Neural Rendering

Compositing Network

Output Image

Target Video Frame

Eroded Image

Prof. Leal-Taixé and Prof. Niessner
Neural Voice Puppetry
Big Open Challenges
Big Open Challenges

Photo-realistic Reconstruction
Big Open Challenges: How much can AI do?

Using a Bounding Box as Proxy

Input UV-Map → Sampled Texture → Ours → Ground Truth

Prof. Leal-Taixé and Prof. Niessner

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


1MPI Informatics 2Stanford University 3Technical University of Munich 4Facebook Reality Labs 5Adobe Research 6Google 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:
Comparisons

Model-based Methods
Neural Voice Puppetry

Comparisons

2D-based Methods
Neural Voice Puppetry

More learnable weights (4x4 kernels)
Less memory consumption because of down convs

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)

~6 hours training time

~50 hours training time

Prof. Leal-Taixé and Prof. Niessner
Facial Reenactment

Dense

Sparse
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 LI, University of Hong Kong, Max Planck Institute for Informatics
WEIPENG XU, Max Planck Institute for Informatics
MICHAEL ZOLL-HOFER, Stanford University, Max Planck Institute for Informatics
HYEONGWOO KIM, FLOHAN BEINARD, and MARC HABERMAANN, Max Planck Institute for Informatics
WENPING WANG, University of Hong Kong
CHRISTIAN THEOBALD, Max Planck Institute for Informatics

Textured Neural Avatars

Aliaksandra Shysheya,1,2,3,4 Egor Zakharov,1,2,3 Karim Iskakov,1,2,3,4
Egor Barkov,1,2,3,4 Karim Iskakov,1,2,3,4
Igor Pasechnik1,2,3,4,5
Dmitry Ulyanov1,2,3,4,5

1 Samsung AI Center, Moscow
2 Skolkovo Institute of Science and Technology, Moscow
3 National Research University Higher School of Economics, Moscow
4 Russian Quantum Center, Moscow
5 Skolkovo Institute of Science and Technology, Moscow

Everybody Dance Now

Caroline Chan\textsuperscript{1} Shiry Ginosar\textsuperscript{1} Tinghui Zhou\textsuperscript{1} Alexei A. Efros

UC Berkeley

G(x_1), \ldots, G(x_t)
Open Challenges

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