EUROGRAPHICS 2022

Machine Learning for Graphics: A Brief Overview

Deepak Vembar

Research Scientist, Intel Labs

email: deepak.s.vembar@intel.com

intel®

Agenda

- Overview of machine learning in graphics (10 mins)
- ML in content generation pipelines (12 mins)
- ML to augment rendering (8 mins)
- Challenges and opportunities (10 mins)

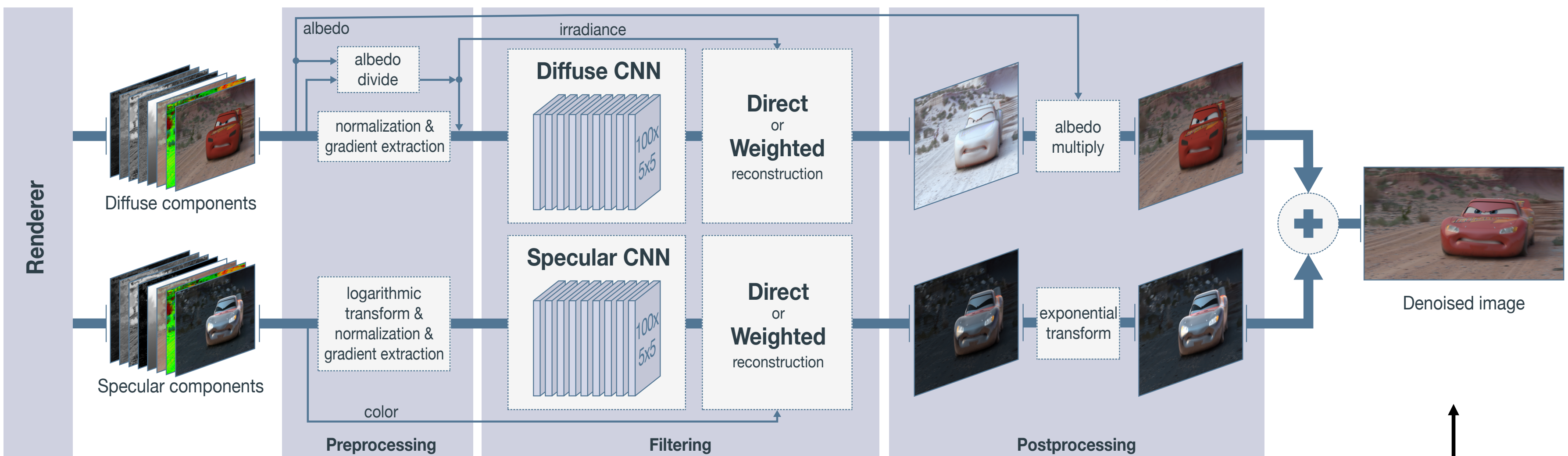
Increased use of ML in computer graphics

- Asset curation, real-time and offline rendering
  - Across the entire production pipeline – games, VFx, interactive rendering
- Improved quality and/or performance, reduced power
  - Authoring time, final frame rendering, better quality at same power
- Improved tools and learnings
  - Hardware and system support – CPUs, GPUs, TPUs, ASICs

Challenges – Datasets, models, generalization, deployment

Iterative ML training workflow

Image Credits: S. Bako ©Disney/ Pixar



GFX buffers

Dataset generation

Neural network

Post processing

Result

ML for content generation

Image credit: T. Aila

Progressive GAN

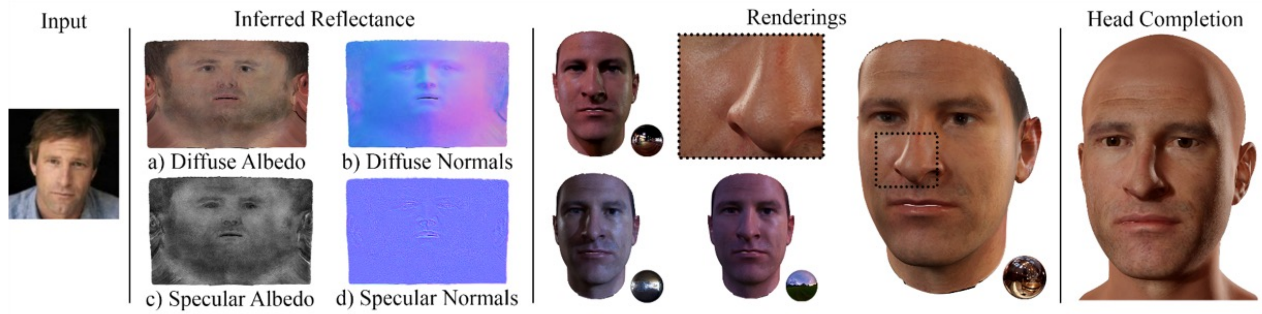


Image credit: A. Lattas

AvatarMe

Image credit: T Komura

Animation

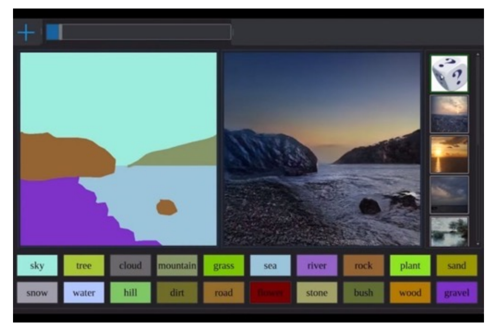


Image credit: T.S. Park

GauGAN

ML integrated with rendering

Image Denoising

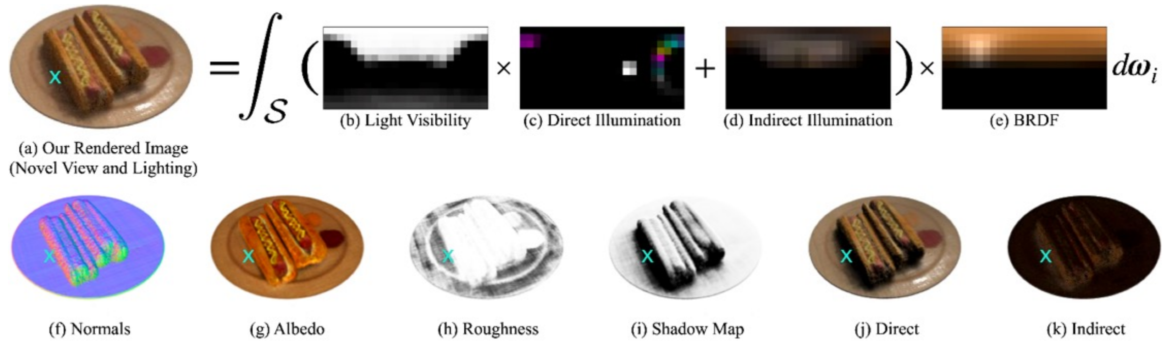


Image credits: P. Srinivasan

Scene relighting

©Nvidia

Neural Scene representation and shading

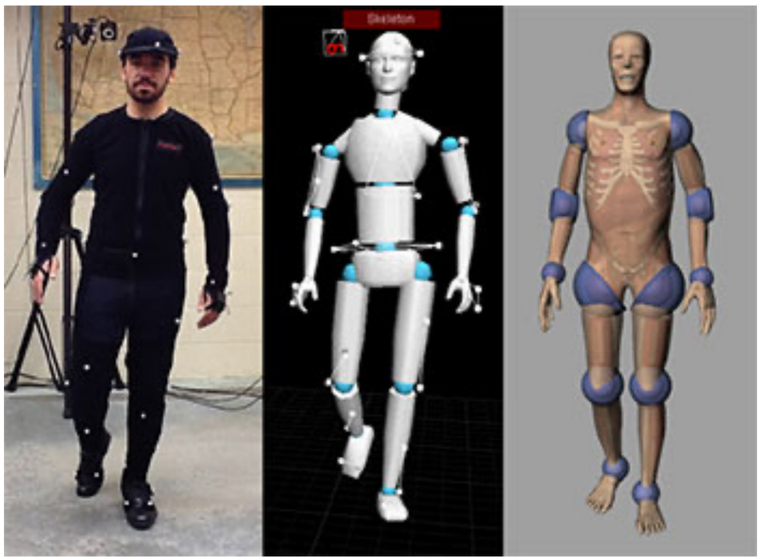
©Nvidia

DL Super Sampling

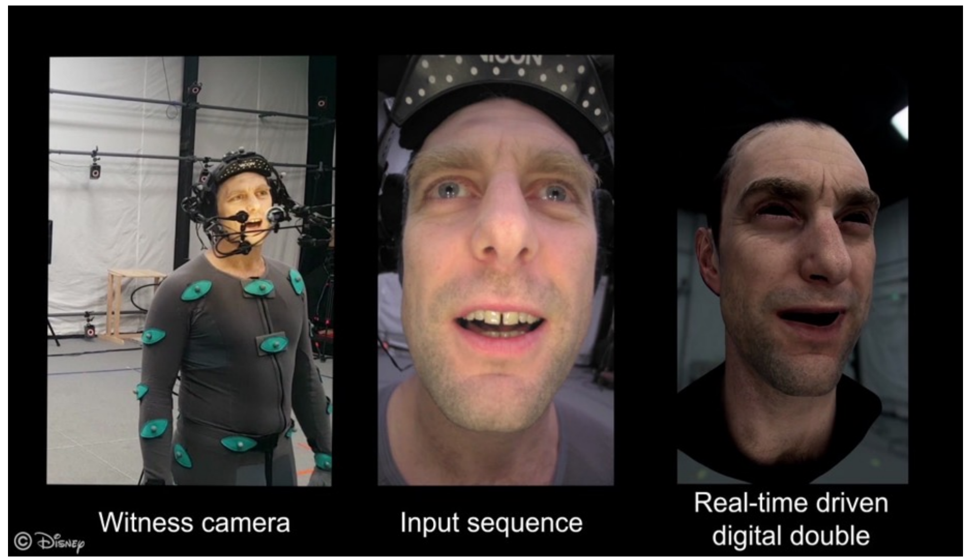
ML for content generation

Neural Animation, Codec avatars, Photorealistic backgrounds

Avatar authoring is time consuming



Motion capture



Facial animation capture ©Disney

Phase-functioned neural network (PFNN)

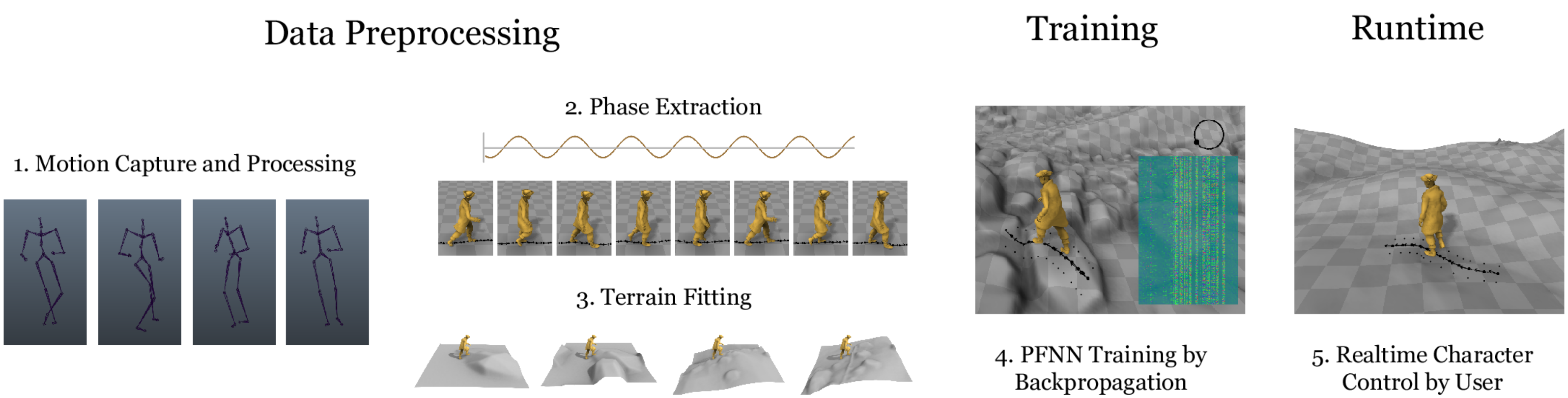


Image credits D. Holden ©ACM

Using mocap data for character animation in real-time games

PFNN – Network topology

- Relatively simple network
  - Additional cyclic function
- Prior frame, user input and scene geometry into consideration
- Outputs next step/ motion
- Fast performance (ms)
  - Integrated into games

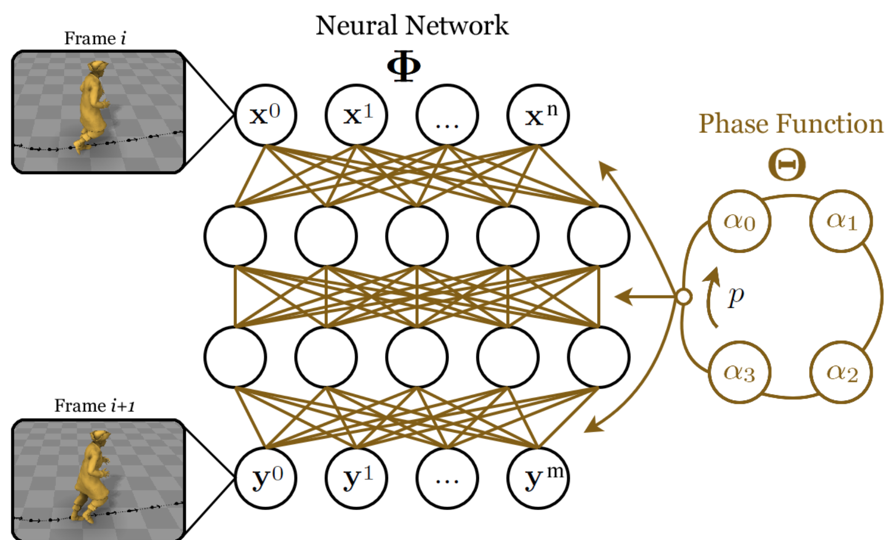


Image credits D. Holden ©ACM

Face rendering for virtual reality

- Facial animation is important for VR experiences
  - Improved presence
- Hard to convey with an HMD
  - Augmentation with extra sensors
- Fast transmission to support distributed participants
  - Social interaction in multi-user scenarios

Facebook – Codec Avatars using deep VAEs

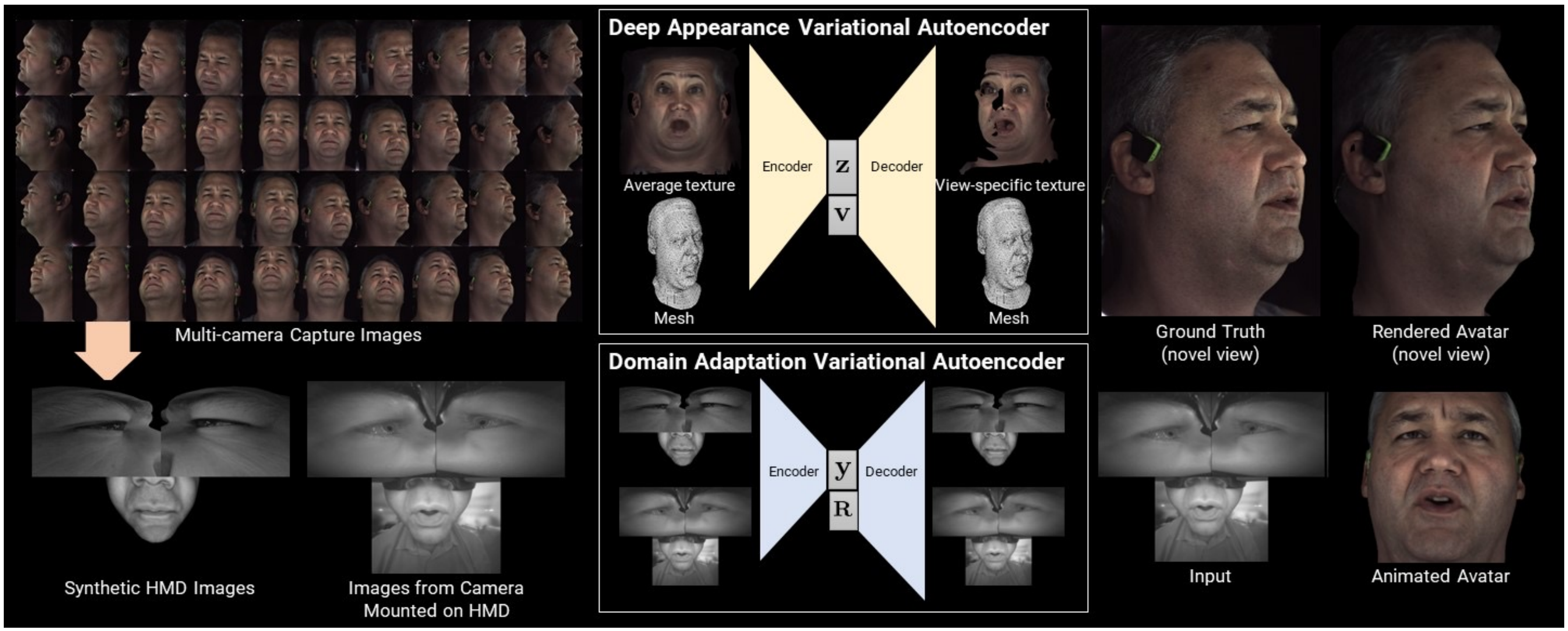


Image credits S. Lombardi ©ACM

Deep Appearance Models to render avatars

Cameras in HMD with multi-view capture dataset

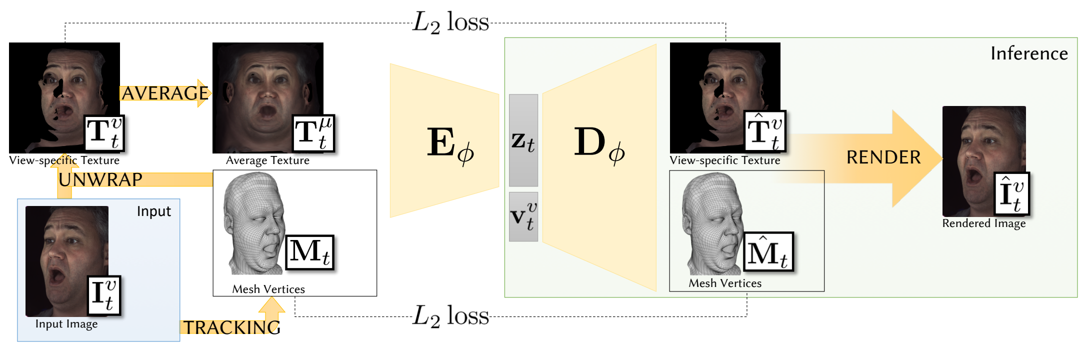
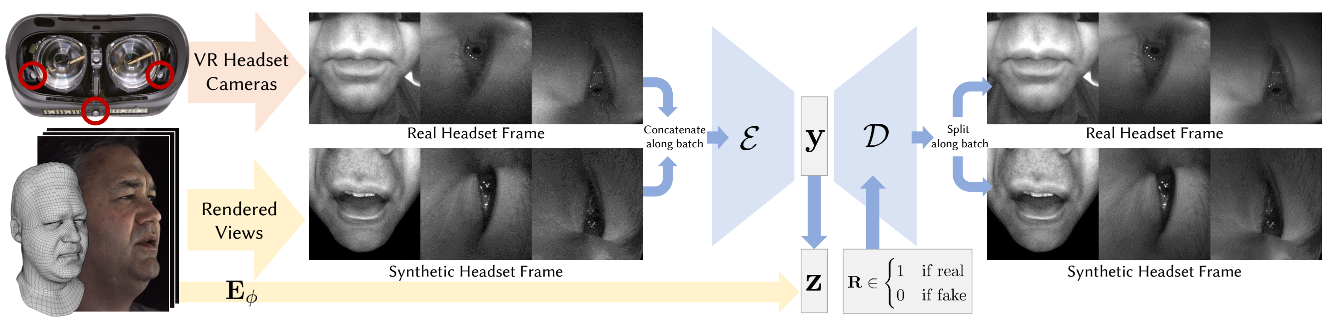


Image credits S. Lombardi ©ACM



Improved telepresence experience

Image credits S. Lombardi

Recent work with relightable face models (SIGGRAPH 2021)

Fast content generation from semantic maps

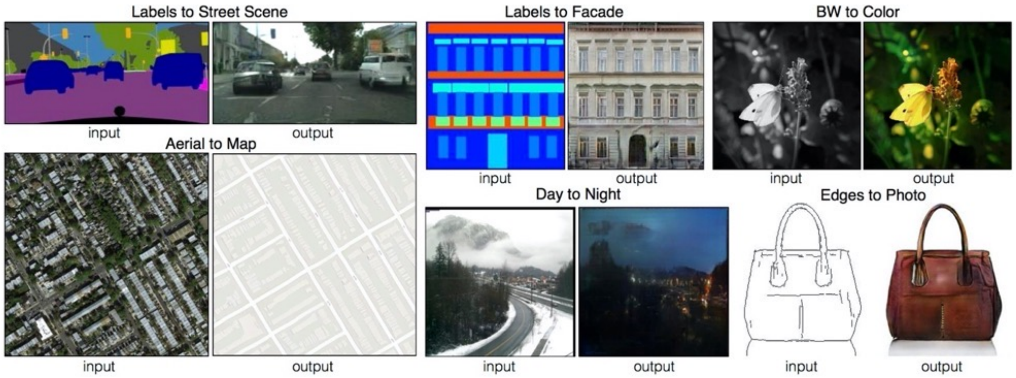


Image credits P. Isola ©IEEE

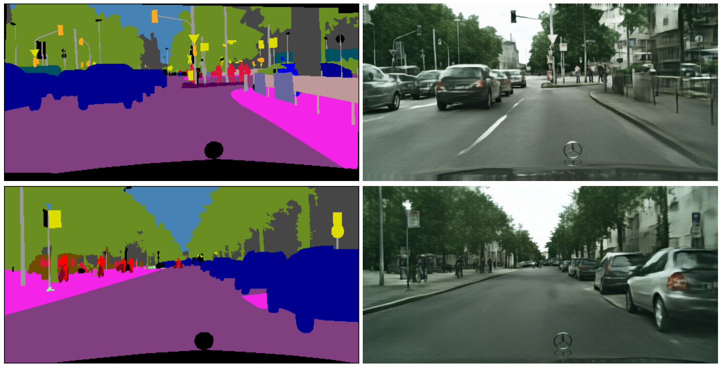


Image credits Q. Chen© ICCV

Image to Image translation

Synthesizing images

NeRF – Novel view synthesis

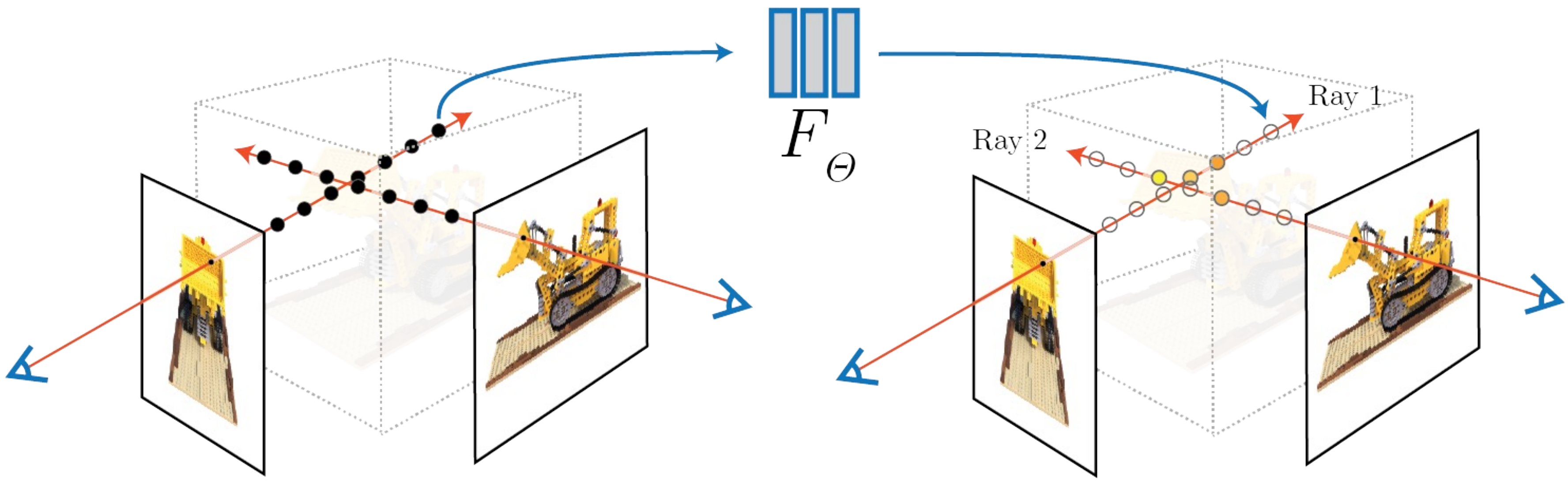
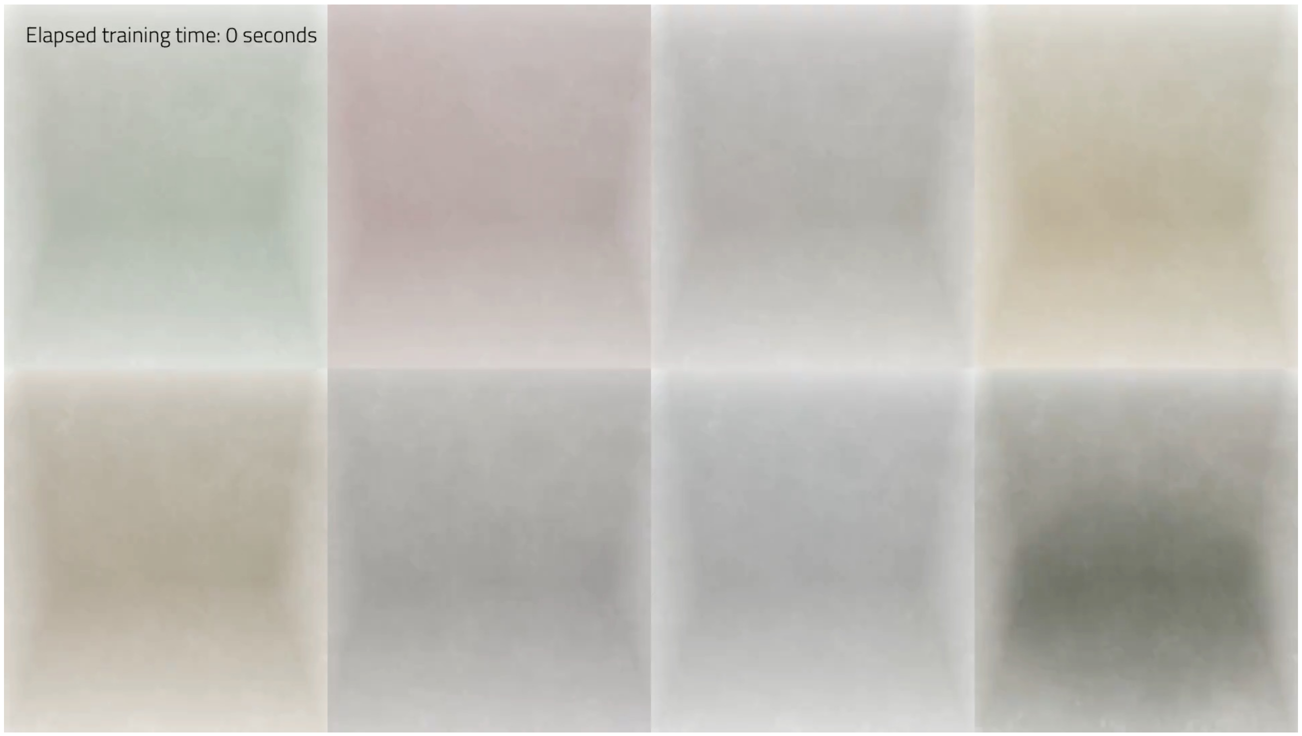


Image credits B. Mildenhall

Easing real-world content capture

Instant NGP - ~real-time training



ML for Rendering

Post processing, super sampling, denoising

Deep learning for post-processing effects

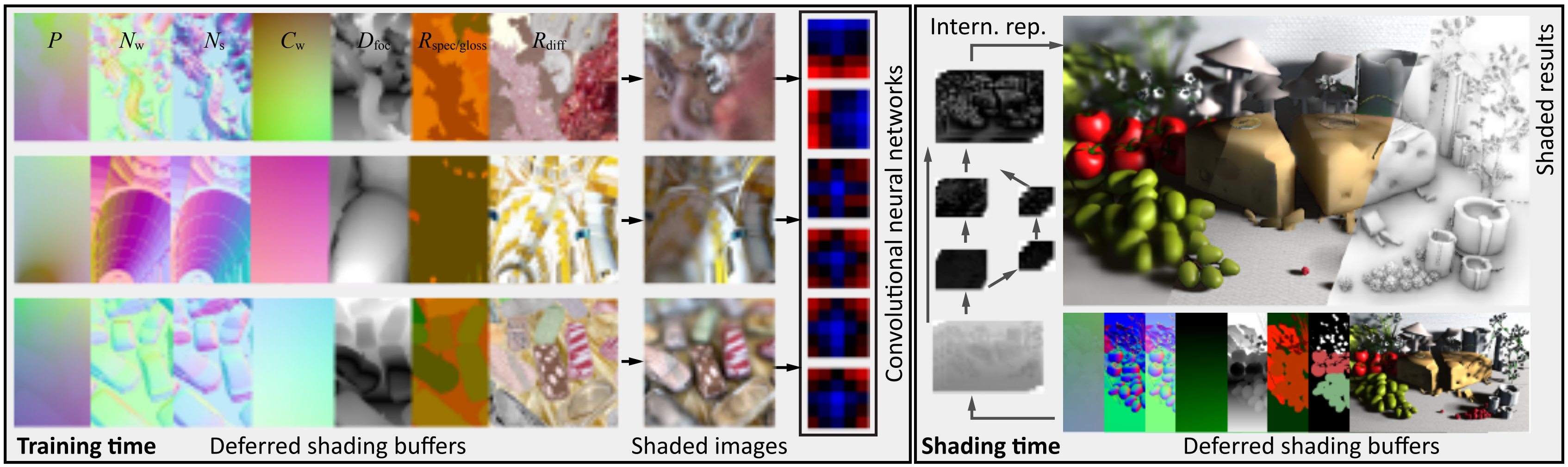


Image credit: O. Nalbach ©Eurographics

Deep Shading- Synthesizing screen space effects using CNNs

Deep Shading network architecture

- U-shaped CNN
- Input buffers depend on post processing effect desired
  - Usually Normals, albedo, motion vectors
- Combined effects using same network
- Fast inference performance

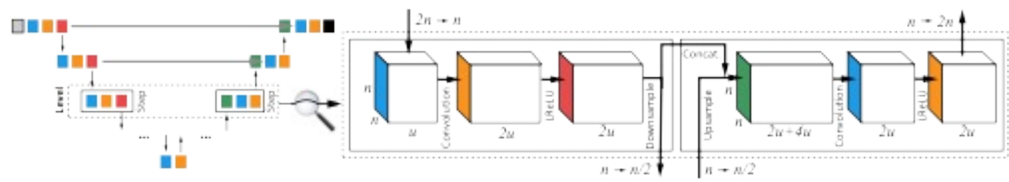
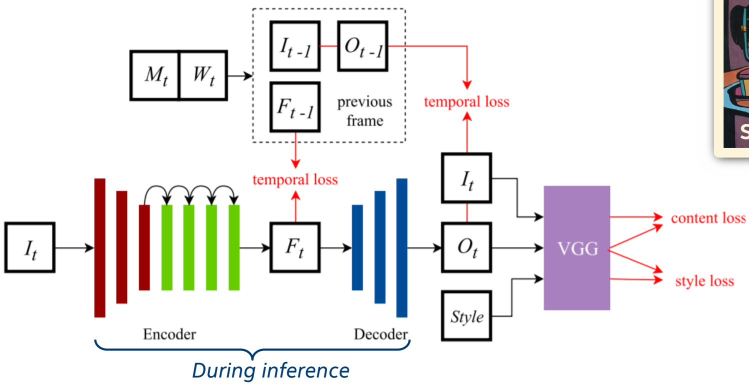
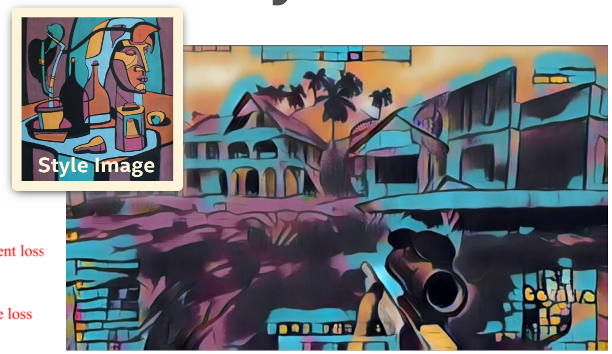


Image credit: O. Nalbach ©Eurographics

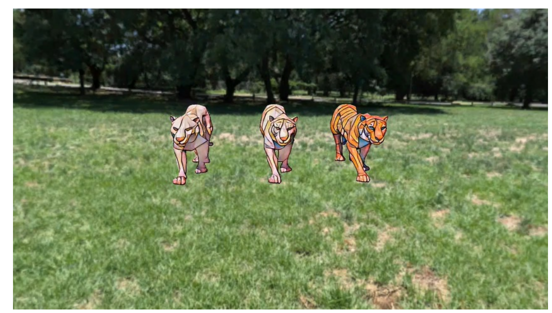
Real-time Segmented Style Transfer





Style Transferred Video

Video content



3D rendered content

Goal: Real-time, temporal consistent, high resolution, per object

- A Feedforward Network design using VGG for perpetual loss
- Use exact pixel segmentation for synthesized content

Challenges in rendering high resolution games

- Interactive gaming at high resolutions/ high fps
  - 4K gaming @60fps
- Bottlenecks in texture sizes, model detail
  - Many millions of polygons, multi GB textures
- Hybrid rendering
  - Global illumination, ray traced reflections, post processing effects

Traditional rendering methods may not suffice

Deep learning super sampling (DLSS)

Image credit ©Nvidia

DLSS 2.0 – Auto encoder with motion vectors

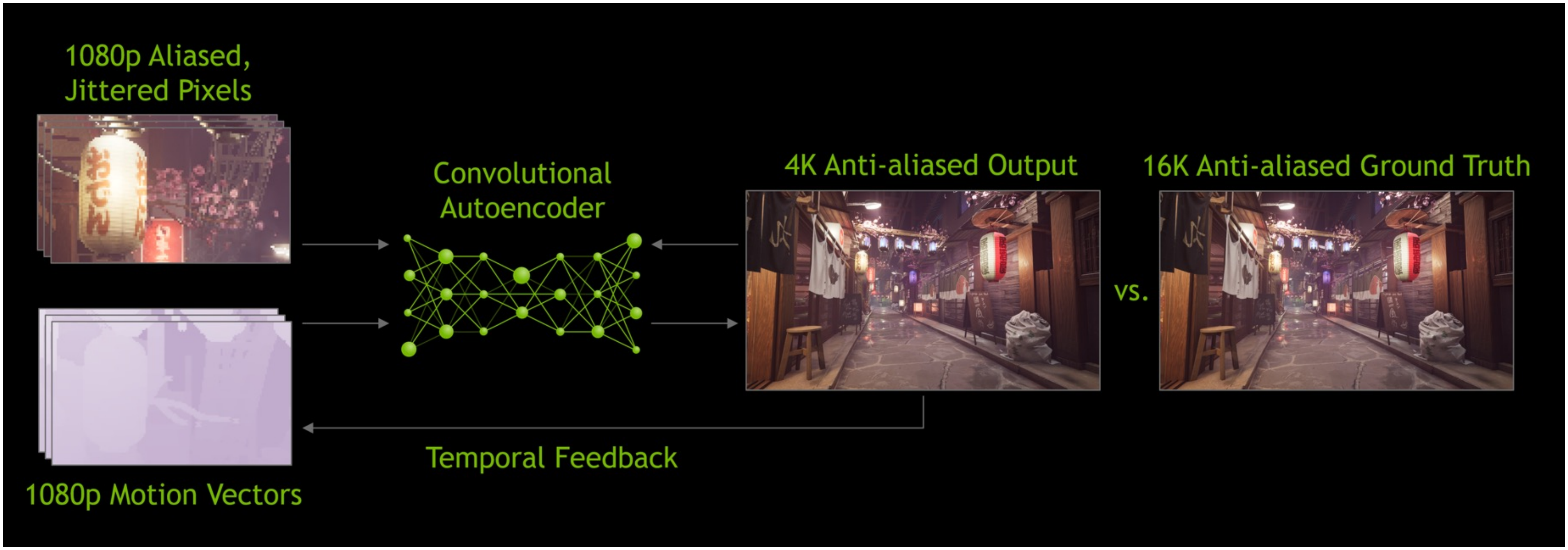


Image credit ©Nvidia

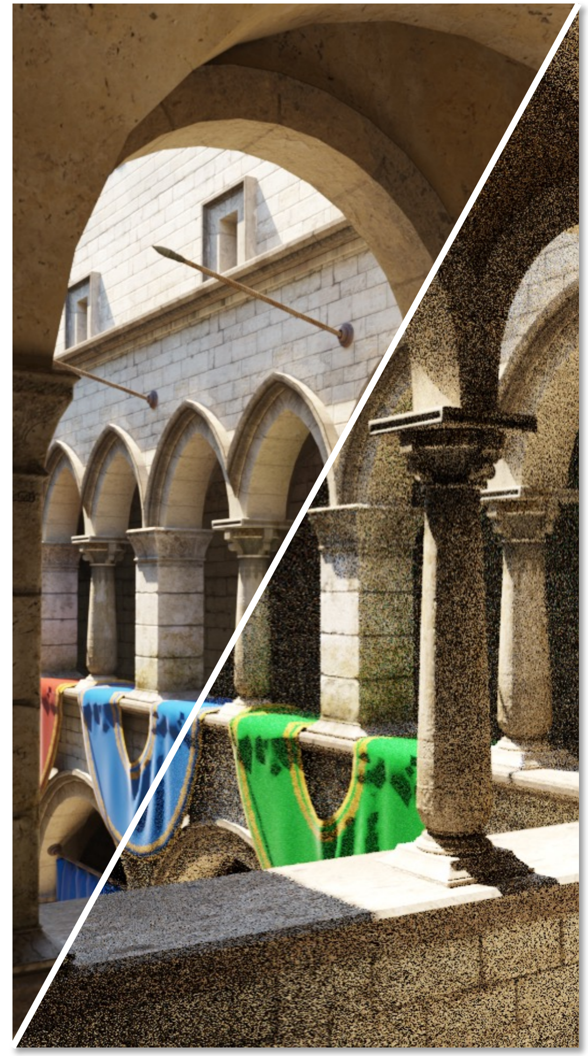
Improved performance in games

- Render low resolution image
  - Image upscale using DL
- Async compute using Tensorcores
  - Significant performance improvement v/s rendering at higher resolution
- Wide adoption
  - Unity/ Unreal

Image credit ©Nvidia

Intel® Open Image Denoise

- Denoising library for ray traced images
  - Final frames and baked lightmaps
- High-quality ML-based denoising filters
- Suitable for interactive and offline rendering
- Simple C/C++ API
- Easy integration into rendering applications
- Open Source under Apache* 2.0 license
  - www.openimagedenoise.org



Scene courtesy of Frank Meinl, downloaded from Morgan McGuire's Computer Graphics Archive.

*Other names and brands may be claimed as the property of others.


Ground truth:
32K spp
RETAIL
blender
The Junk Shop by Alex Treviño. Original Concept by Anaïs Maamar.


Ground truth:
32K spp
Low sample count:
16 spp
RETAIL
3RD PIECE
-50%
blender
The Junk Shop by Alex Treviño. Original Concept by Anaïs Maamar.


Ground truth:
32K spp
Low sample count:
16 spp (denoised)
RETAIL
blender
The Junk Shop by Alex Treviño. Original Concept by Anaïs Maamar.
intel
OPEN IMAGE
DENOISE

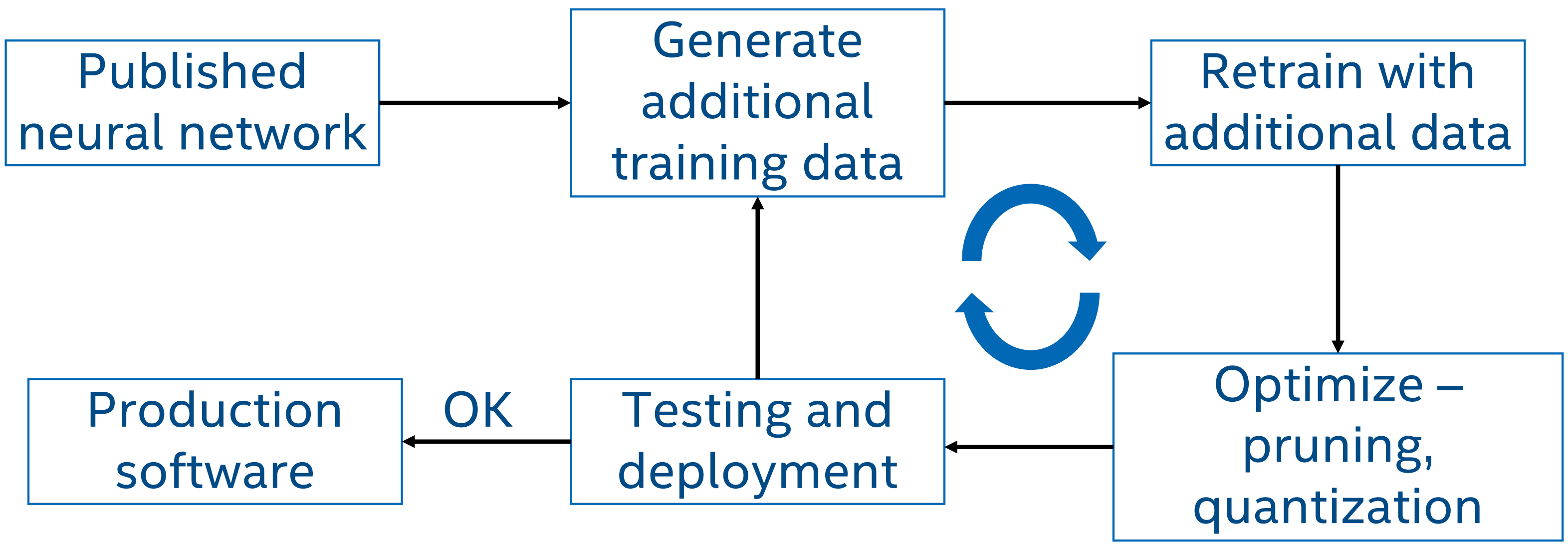
More recent papers show promise

- Photogrammetry and novel view synthesis
- Vfx usages
  - Relighting, appearance capture
- Ray tracing and path tracing
  - Importance sampling, adaptive sampling with denoising
- Improving photorealism in games

Challenges

Datasets, neural networks, deployment

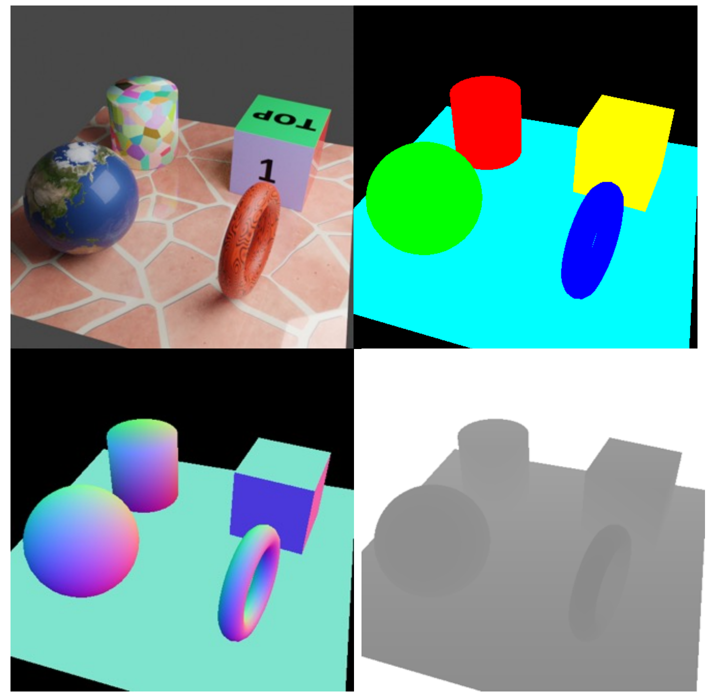
Testing and deploying a published ML model



Iterative process and requires a lot of additional steps

Dataset curation and augmentation

- Common buffers used – normal, albedo, color, position, depth, specular, motion vectors
- Most data can be directly obtained from renderers
- Input resolution
- Rendering time for dataset



Some considerations for dataset generation

- Size of dataset to be collected – small v/s large
  - Training time v/s quality
- Licensing of datasets – open v/s closed
- Generalizability across different scenes
  - Rendering time implications
- Compressed v/s uncompressed data
  - Memory costs and training time
- Data format – color space, dynamic range

Neural network architecture and optimizations

- Takes some effort to deploy published work
  - Understand performance targets and deployment system
- Common network optimizations
  - Pruning, quantization, sparsity
- Use tools such as TensorRT, OpenVINO for auto optimization
  - Most take an ONNX file as input
- Considerations for extending from images to videos
  - Minimize flicker, include temporal loss terms

Deployment considerations

- Training generally uses Pytorch/ Tensorflow
  - Impractical for deployment in real-time usages
- Hardware compatibility and driver support
  - Fallback software path may not be as performant as hardware supported path
- Standards and APIs evolving to support ML
  - DirectML with DirectX, ONNX as model interchange format
- Third party and ecosystem support
  - E.g: Unity Barracuda for inference deployments

Exciting time to be in graphics

- Increased use of ML in graphics
- Potential to improve quality, reduce rendering times and democratize content generation costs
- Improved hardware and systems support,

But..

- Challenges – datasets, networks, deployments

We have just scratched the surface

intel®

References

- Kernel-Predicting Convolutional Networks for Denoising Monte Carlo Renderings, S. Bako*, T. Vogels*, B. McWilliams, M. Meyer, J. Novák, A. Harvill, P. Sen, T. DeRose, and F. Rousselle, "Kernel-Predicting Convolutional Networks for Denoising Monte Carlo Renderings," ACM Transactions on Graphics, Vol. 36, No. 4, Article 97, July 2017, (Proceedings of ACM SIGGRAPH 2017)
- ProgressiveGAN - https://arxiv.org/abs/1710.10196
- AvatarMe: Realistically Renderable 3D Facial Reconstruction "In-the-Wild", Lattas, Alexandros and Moschoglou, Stylianos and Gecer, Baris and Ploumpis, Stylianos and Triantafyllou, Vasileios and Ghosh, Abhijeet and Zafeiriou, Stefanos, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020, Animation - https://dl.acm.org/doi/abs/10.1145/3386569.3392450
- Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu., "Semantic Image Synthesis with Spatially-Adaptive Normalization", in CVPR, 2019.
- Denoising with kernel prediction and asymmetric loss functions, Thijs Vogels, Fabrice Rousselle, Brian McWilliams, Gerhard Röthlin, Alex Harvill, David Adler, Mark Meyer, Jan Novák, ACM Transactions on Graphics, Volume 37, Issue 4, August 2018 Article No.: 124, pp 1–15, https://doi.org/10.1145/3197517.3201388
- Srinivasan, Pratul & Deng, Boyang & Zhang, Xiuming & Tancik, Matthew & Mildenhall, Ben & Barron, Jonathan. (2020). NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis.
- Compositional Neural Scene Representations for Shading Inference, Jonathan Granskog, Fabrice Rousselle, Marios Papas, Jan Novák, Transaction on Graphics (Proceedings of SIGGRAPH 2020), vol. 39, no. 4
- DLSS 2.0, retrieved June 2021, https://www.nvidia.com/en-us/geforce/news/nvidia-dlss-2-0-a-big-leap-in-ai-rendering/
- S. McDonagh, M. Klaudiny, D. Bradley, T. Beeler, I. Matthews and K. Mitchell, "Synthetic Prior Design for Real-Time Face Tracking," 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 639-648, doi: 10.1109/3DV.2016.72.
- Phase-functioned neural networks for character control : ACM Transactions on GraphicsVolume 36Issue 4July 2017 Article No.: 42pp 1–13, https://doi.org/10.1145/3072959.3073663
- Deep appearance models for face rendering, Stephen Lombardi , Jason Saragih, Tomas Simon, Yaser Sheikh, ACM Transactions on GraphicsVolume 37Issue 4August 2018 Article No.: 68pp 1–13https://doi.org/10.1145/3197517.3201401
- Image-to-Image Translation with Conditional Adversarial Networks, Isola, Phillip and Zhu, Jun-Yan and Zhou, Tinghui and Efros, Alexei A, CVPR, 2017
- Photographic Image Synthesis with Cascaded Refinement Networks, Qifeng Chen and Vladlen Koltun, International Conference on Computer Vision (ICCV), 2017
- Deep Shading: Convolutional Neural Networks for Screen Space Shading, O Nalbach, E Arabadzhiyska, D Mehta, HP Seidel, T Ritschel, Computer Graphics Forum 36 (4), 65-78
- Thomas, M. M. and Forbes, A. G., "Deep Illumination: Approximating Dynamic Global Illumination with Generative Adversarial Network", arXiv e-prints, 2017