XDGAN

Multi-Modal, 3D-Shape Generation in 2D Space

Hassan Abu Alhaija[1]
habualhaija@nvidia.com

Alara Dirik[2]
alara@pch-innovations.com

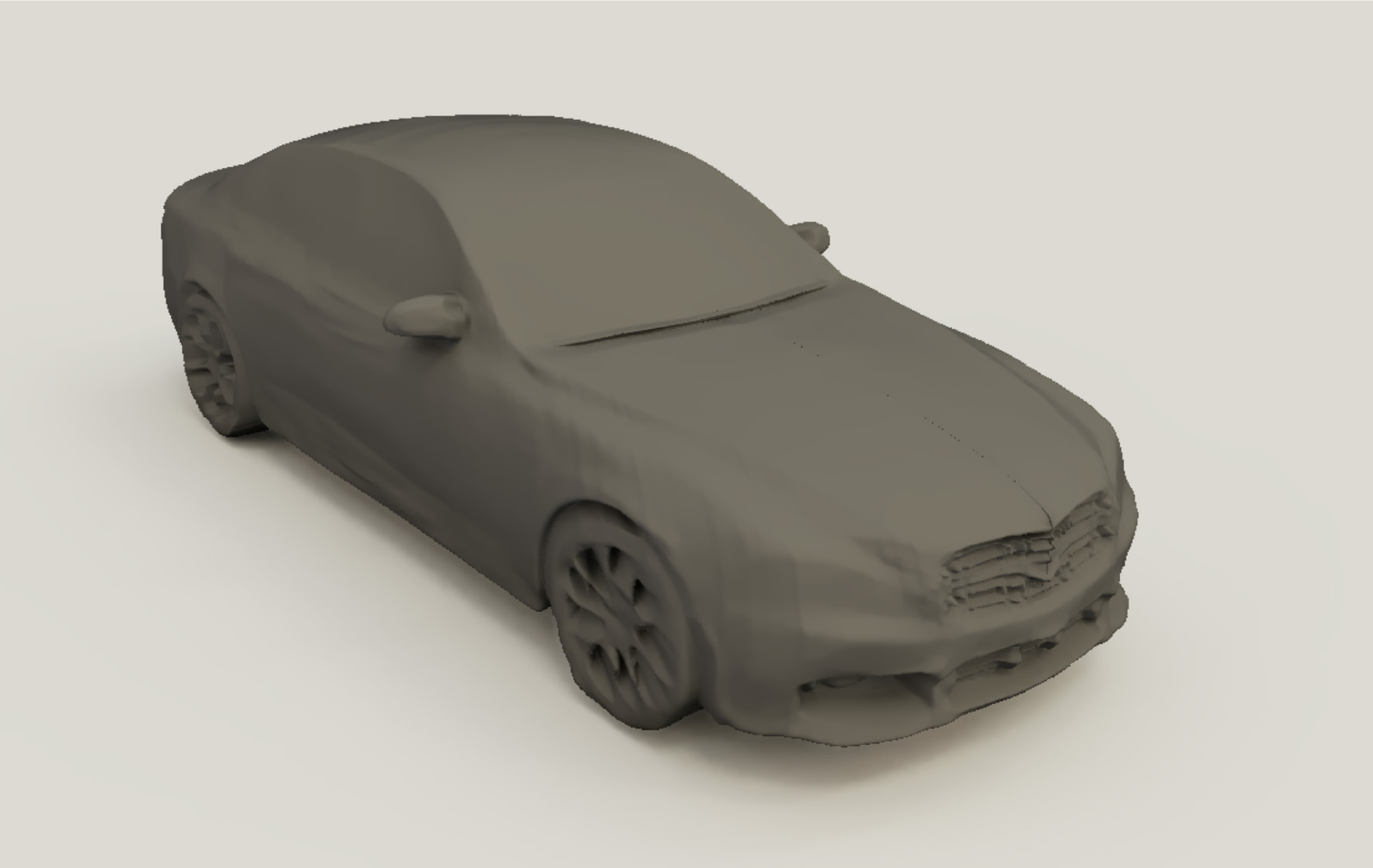
André Knörig[2]
andre@pch-innovations.com

Sanja Fidler[1,3]
sfidler@nvidia.com

Maria Shugrina[1]
mshugrina@nvidia.com

[1] NVIDIA
[2] PCH Innovations GmbH
[3] University of Toronto

nv-tlabs.github.io/XDGAN



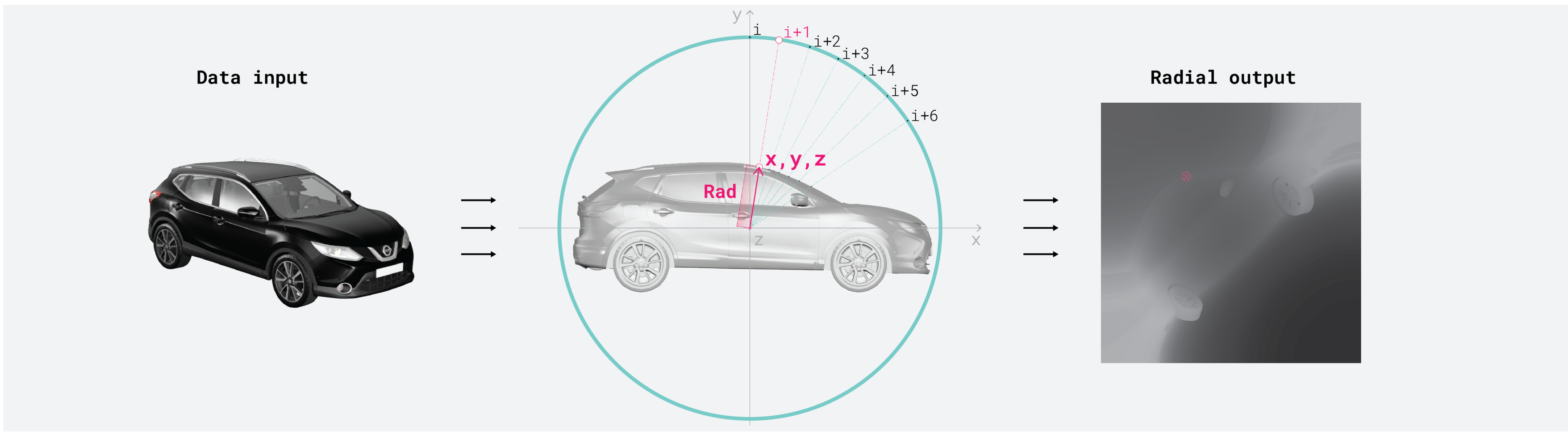
Generative models for 2D images have recently seen tremendous progress in quality, resolution and speed as a result of the efficiency of 2D convolutional architectures. However, it is difficult to extend this progress into the 3D domain since most current 3D representations rely on custom network components. Our paper addresses a central question:

Is it possible to directly leverage 2D image generative models to generate 3D shapes?

To answer this, we propose XDGAN, an effective and fast method for applying 2D image GAN architectures to the generation of 3D object geometry combined with additional surface attributes, like color, textures and normals. Specifically, we propose a novel method to convert 3D shapes into compact 1-channel geometry images and leverage StyleGAN3 and image-to-image translation networks to generate 3D objects in 2D space.

	Representations				
	Voxels	Points	Implicit	Mesh	**Ours**
Real-time generation	✓	✓	✗	✓	✓
Real-time rendering	✓	✓	✗	✓	✓
High-quality surface	✗	✗	✓	✗	✓
Texture	✗	✗	✗	✓	✓
Variable topology	✓	✗	✓	✓ | ✗	✗

Comparison of 3D generative methods by the representation used. While there have been remarkable advances in 3D shape generation in the recent years, deployment of 3D generative models poses challenges due to generation quality and speed, as well as compatibility of the output format with downstream tasks.



Generation of Rad geometry images from 3D meshes using spherical projection. We represent the 3D location of a surface point as a single value measuring the distance (which we call "Rad") from the origin along the projection ray, effectively reducing the degrees of freedom of the representation from 3 channels per pixel to only 1.

Generation Samples

XDGAN allows generation of high-resolution textured 3D meshes, and supports projection of 3D models into the latent space, where generation of multiple textures and semantic editing are possible.

Original *Projection (full)* *Projection (w/o texture)* *Projection (w/o normals)*

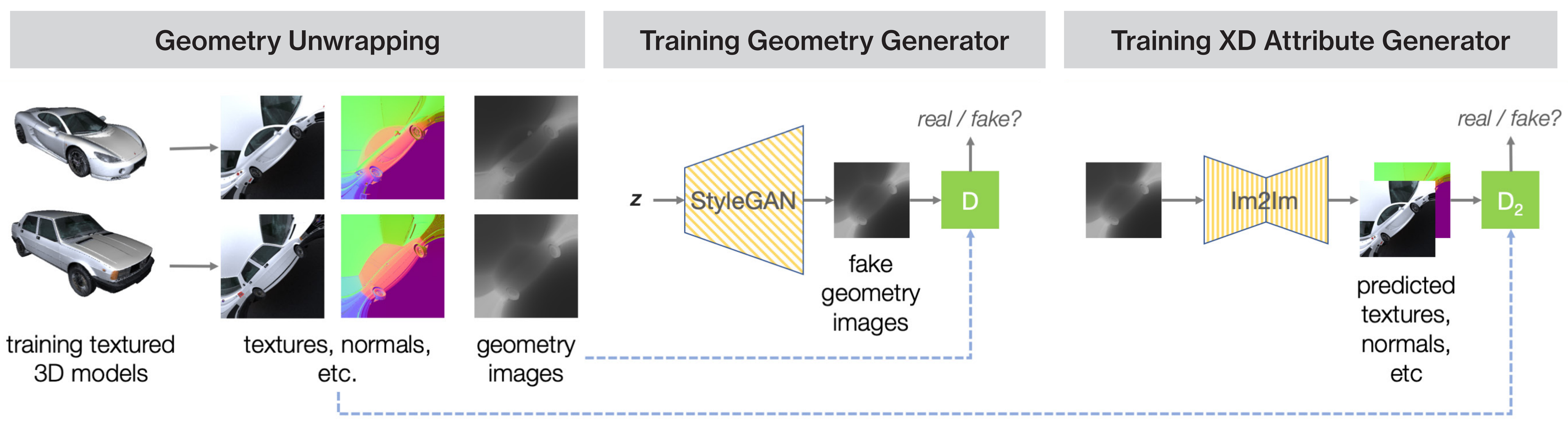
Model Architecture

To train our model **(a)**, we first convert a training dataset of textured 3D meshes into 2D geometry images with corresponding textures, normals or any other surface attributes.

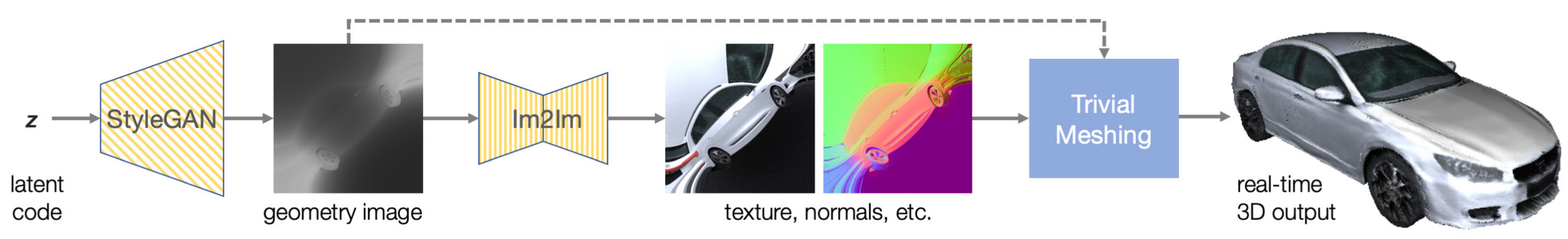
Next, we train a GAN model on geometry images to generate geometry, and an image-to-image translation network to generate X-channel attribute images for an input geometry image.

At test-time, feed-forward evaluation of these networks followed by a trivial meshing step produces textured 3D output meshes in real-time.

Training Pipeline



Generation Pipeline



PCH INNOVATIONS | NVIDIA | UNIVERSITY OF TORONTO