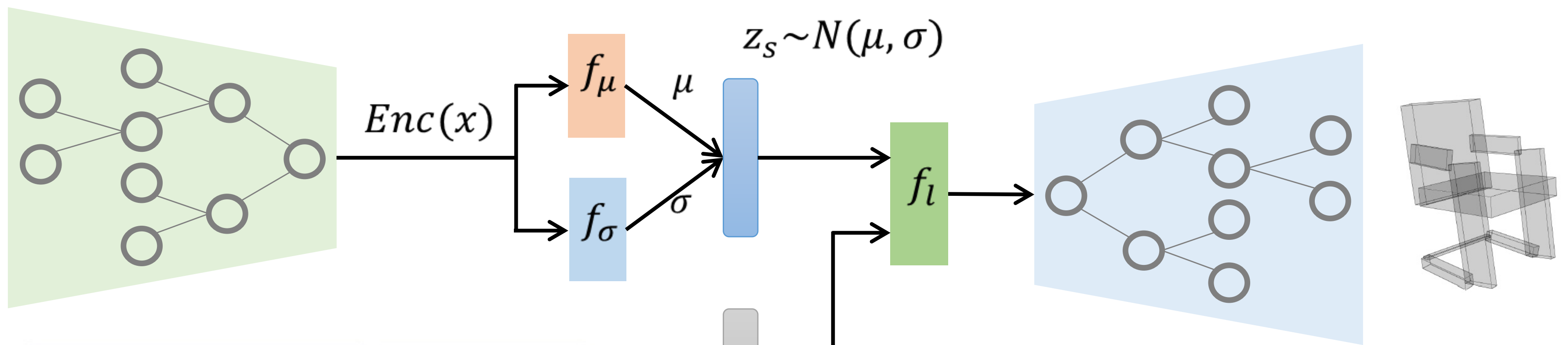
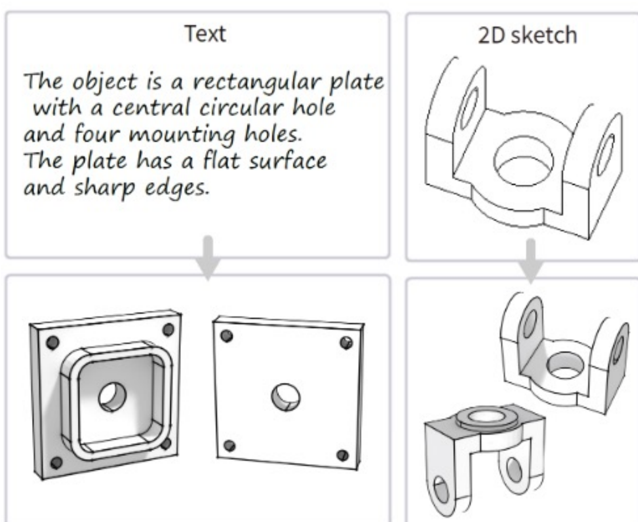
The New Computer Graphics in the Age of AI and Robotics

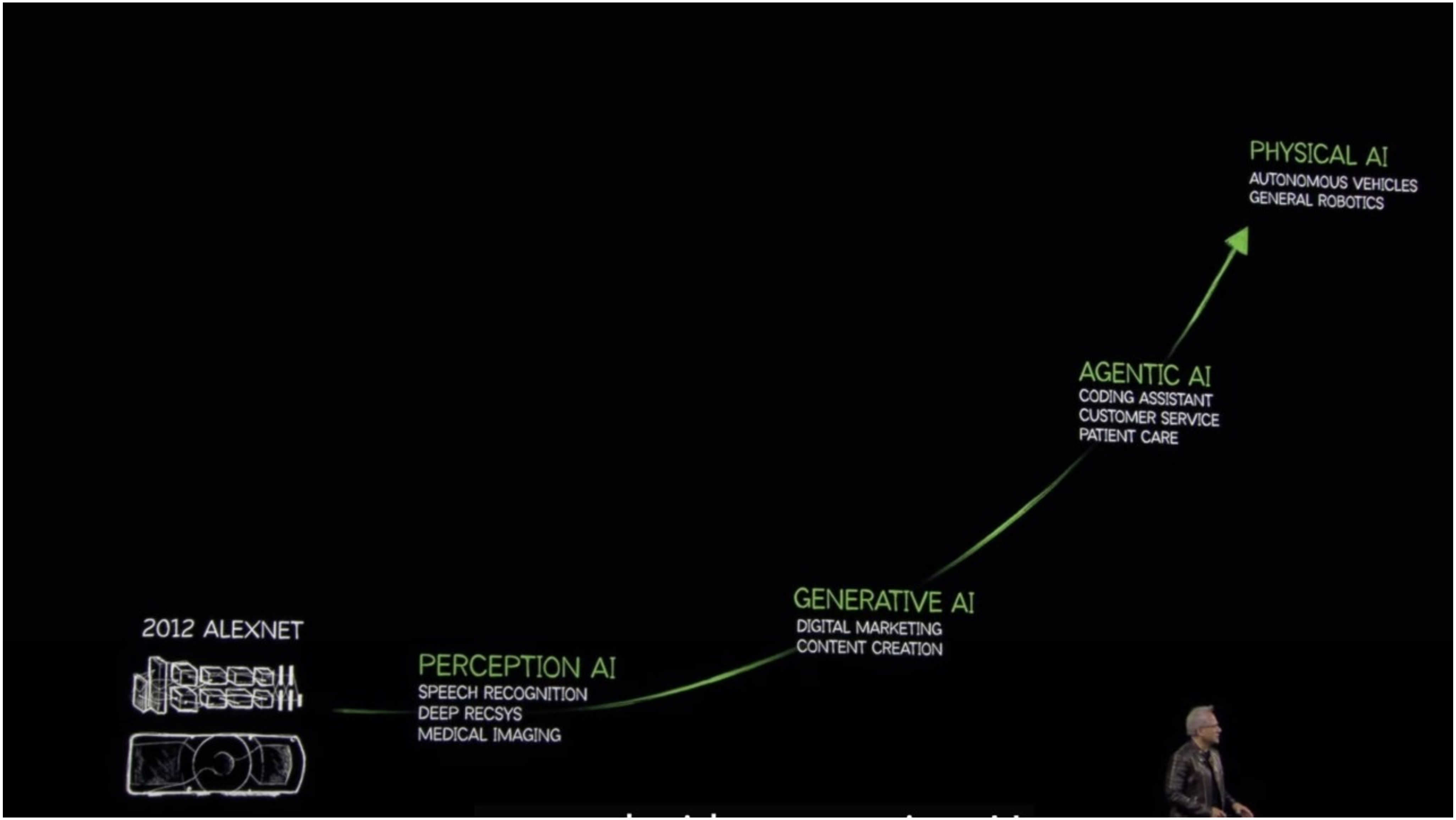




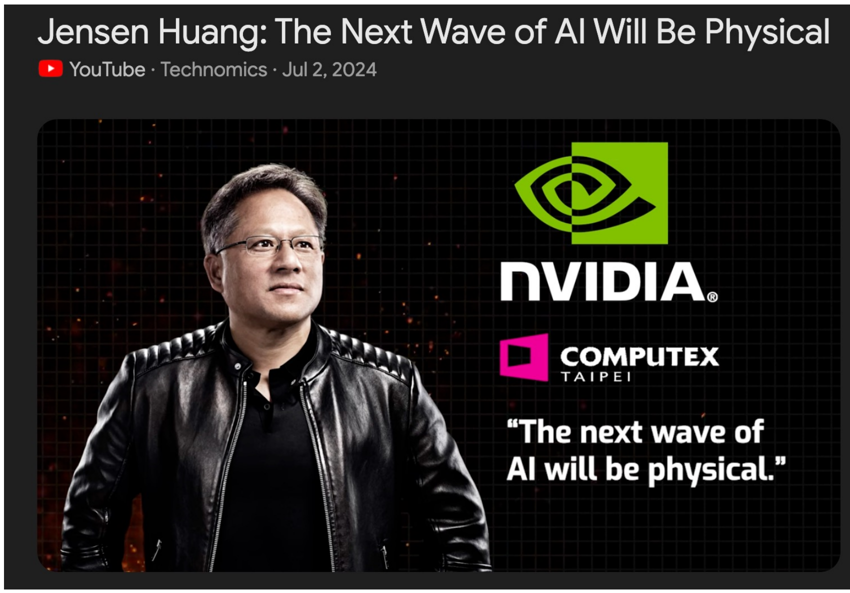
CMPT 464/764 Lecture 0

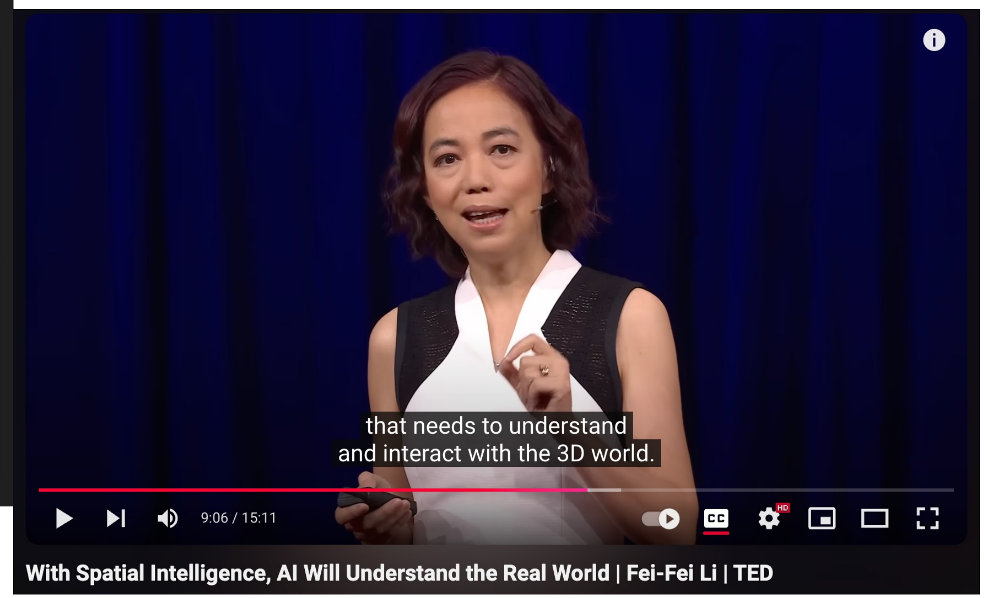
Why robotics?

Nvidia GPU Tech Conf (GTC) 2025



Next wave of AI: spatial and physical

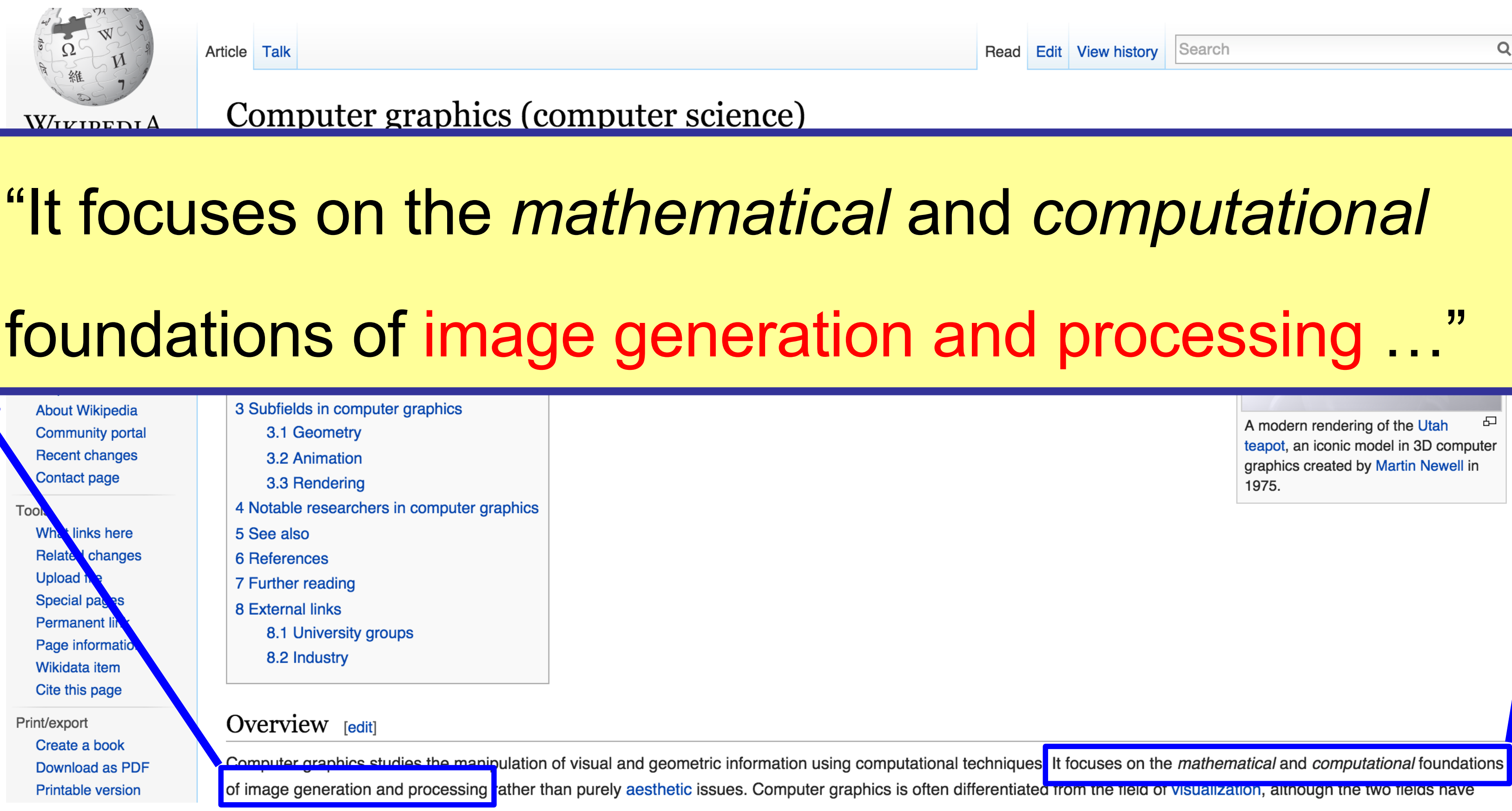




- 🙁 Contemporary Large Foundational Models (e.g., GPT-4o) still far from having spatial intelligence

What is computer graphics (CG)?

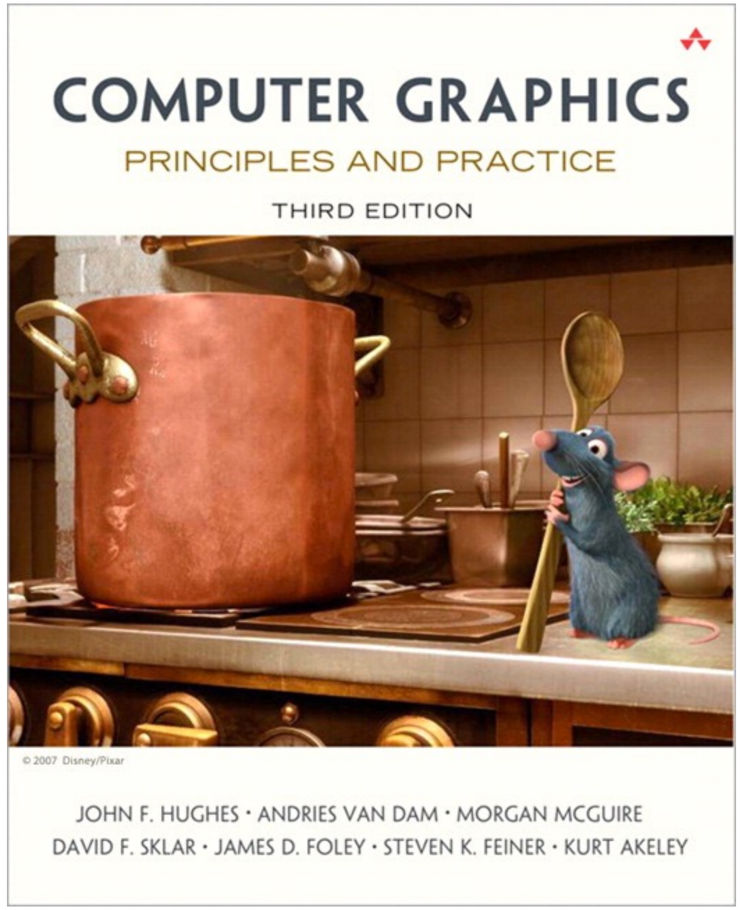
- Wikipedia: computer graphics (computer science)



"It focuses on the *mathematical* and *computational* foundations of image generation and processing …"

What do the experts say?

Perhaps the most classic computer graphics textbook

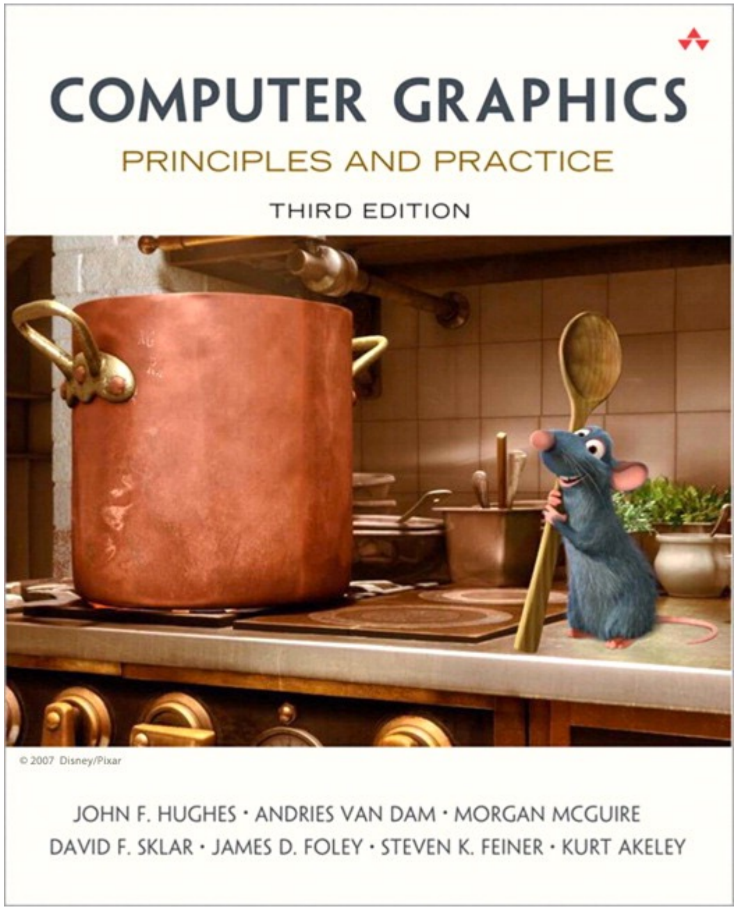


Third edition @ 2014

What do the experts say?

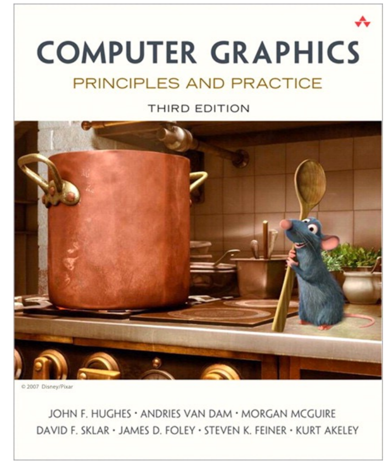
- Hughes, van Dam, et al.:
  - “Computer graphics is the science and art of communicating visually via a computer’s display and its interaction devices.”

------ *page 1.*



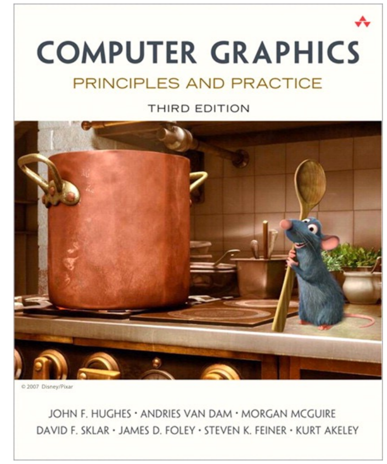
Third edition @ 2014

Classical CG



- Hughes, van Dam, et al.:
  - "Taking a model of the objects in a scene and a model of the light emitted into the scene and producing a representation of a particular view of the scene."

Classical CG



- Hughes, van Dam, et al.:
  - "Taking a model of the objects in a scene and a model of the light emitted into the scene and producing a representation of a particular view of the scene."
  - "A glorified multiplication: multiplying incoming light by reflectivity of objects … for all light reaching the camera"

------ *page 2, "A narrow definition".*

List of CG topics from CMPT 361

- The graphics (vertex & pixel) pipeline
- Transformation, viewing, projection, clipping & visibility
- Light, color, local & global illumination
- Sampling and reconstruction: Fourier transform; aliasing
- Image representation, manipulation, and texture mapping
- Curves, surfaces, meshes, and other geometry reps

All about model representation and rendering

Classical CG

- Explicit scene description is given
- Key problem #1: how to best represent geometry, texture, and lighting for the given scene

Classical CG

- Explicit scene description is given
- Key problem #1: how to best represent geometry, texture, and lighting for the given scene
- Key problem #2: how to render the scene with
  - Efficiency

A forward problem:

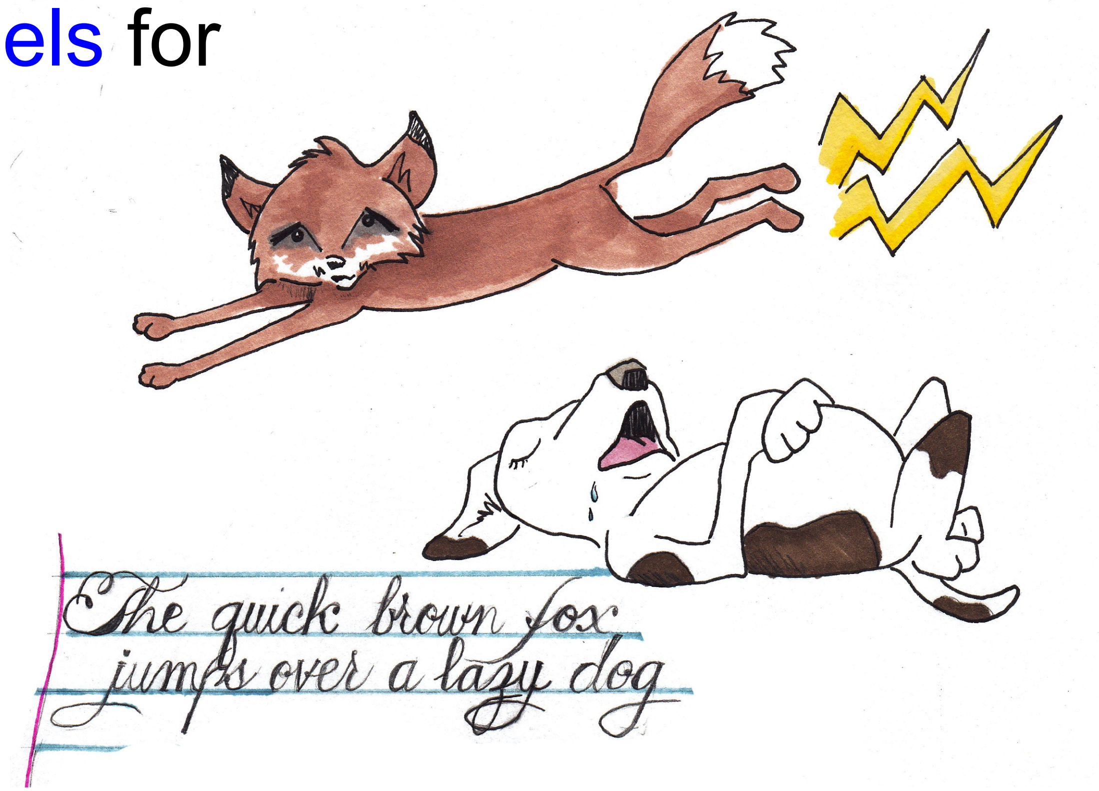
Explicit model description ➡ rendered image

The “forward” problem

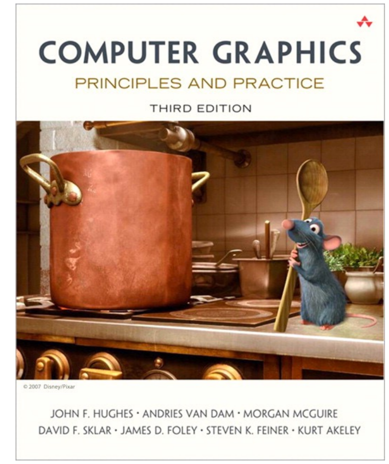
- “The quick brown fox jumps over a lazy dog.”
- Need explicit models for
  - A brown fox
  - A dog
  - Quick jump
  - Sleeping dog …

The "forward" problem

- "The quick brown fox jumps over a lazy dog."
- Need explicit models for
  - A brown fox
  - A dog
  - Quick jump
  - Sleeping dog …

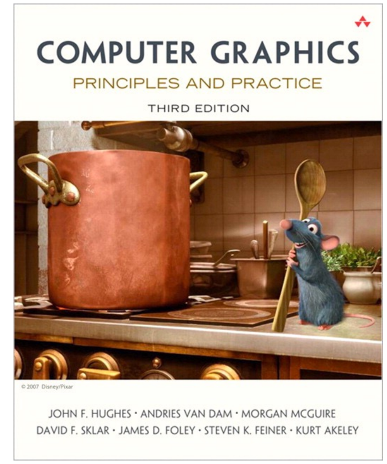


What about computer vision?



- Lower level:
  - Analysis: given one view of a scene, determine the illumination and the scene's content, which a graphics system could use to produce the scene

What about computer vision?



- Lower level:
  - Analysis: given one view of a scene, determine the illumination and the scene's content, which a graphics system could use to produce the scene
- Higher level: infer an understanding of what are

An inverse problem:
From a rendered image to a model description

The “inverse” problem

- Ask Claude: Please describe this image.

The "inverse" problem

- Ask Claude: Please describe this image.
  - There is a fox
  - There is a dog
  - Fox jumps over dog
  - Fox is quick
  - Dog is lazy …

Graphics vs. vision

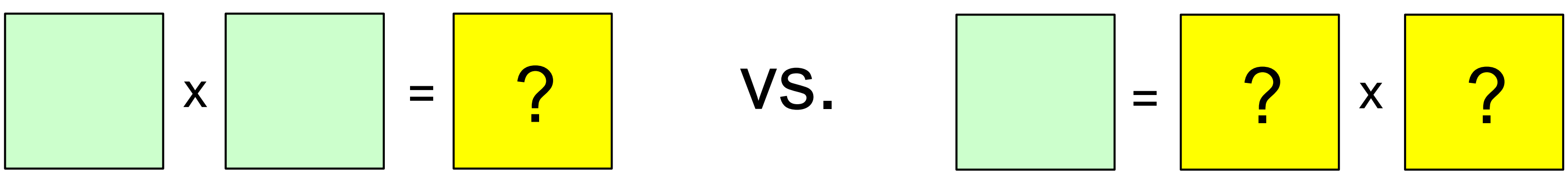
- Graphics is about synthesis
  - Classical graphics is about image synthesis

Graphics vs. vision

- Graphics is about synthesis
  - Classical graphics is about image synthesis
- Vision is about image analysis

Graphics vs. vision – classically

- Graphics is about synthesis
  - Classical graphics is about image synthesis
- Vision is about image analysis
- In classical setting, they were opposite problems
  - Forward vs. inverse problems: which is harder? ☺



Difference in DATA – classically

- Before jumping to “the new graphics”, what would be one big difference between “data in classic CG” and “data in classic CV”?

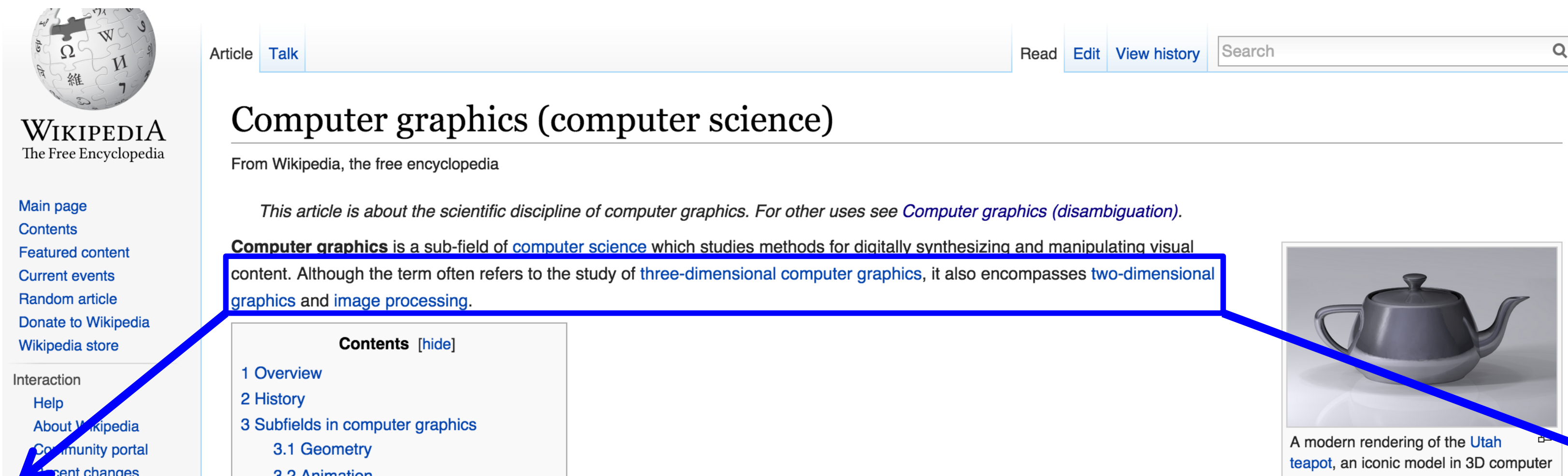
2D image data for computer vision

vs.

3D shape/scene data for computer graphics

Difference in DATA – classically

- Wikipedia: computer graphics (computer science)



"Although the term (computer graphics) often refers to the study of three-dimensional computer graphics, it also encompasses two-dimensional graphics and image processing …"

3D Vision 2026, Vancouver, 03/24-27

General Chairs

Manolis Savva
Simon Fraser University

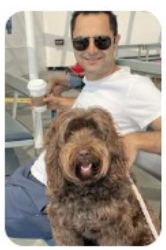
Christian Theobalt
MPII / Saarland

Angjoo Kanazawa
UC Berkeley

Program Chairs

Contact: 3dv-2026- pcs [at] googlegroups [dot] com

Lingjie Liu
University of Pennsylvania



Arash(Ali) Mahdavi-Amiri
Simon Fraser University

Daniel Ritchie
Brown University

Young Min Kim
Seoul National University

Keynote Speakers

Jitendra Malik

Angela Dai

Christian Rupprecht

Alec Jacobson

Regardless: challenges with 3D data

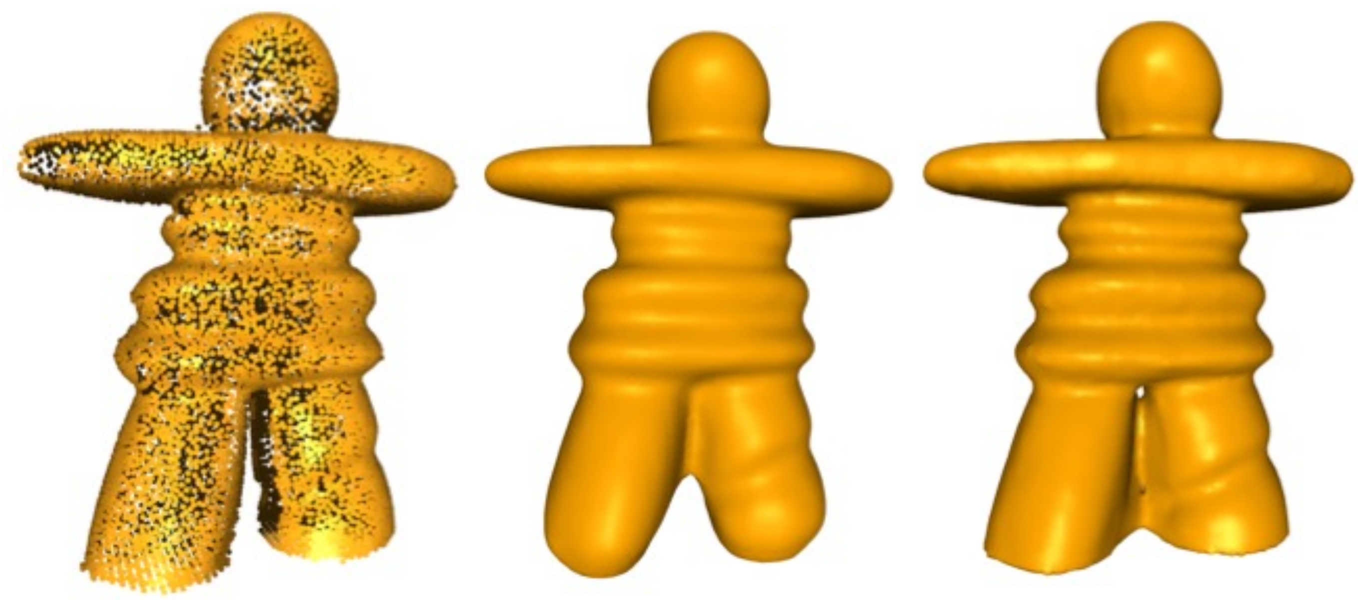
- Acquisition of 3D models is hard
- 3D modeling is hard
- Spatial reasoning and computation in 3D is hard
- Interaction in 3D is hard

3D challenge: acquisition

- Harder to acquire 3D chairs than chair images

3D challenge: classical acquisition

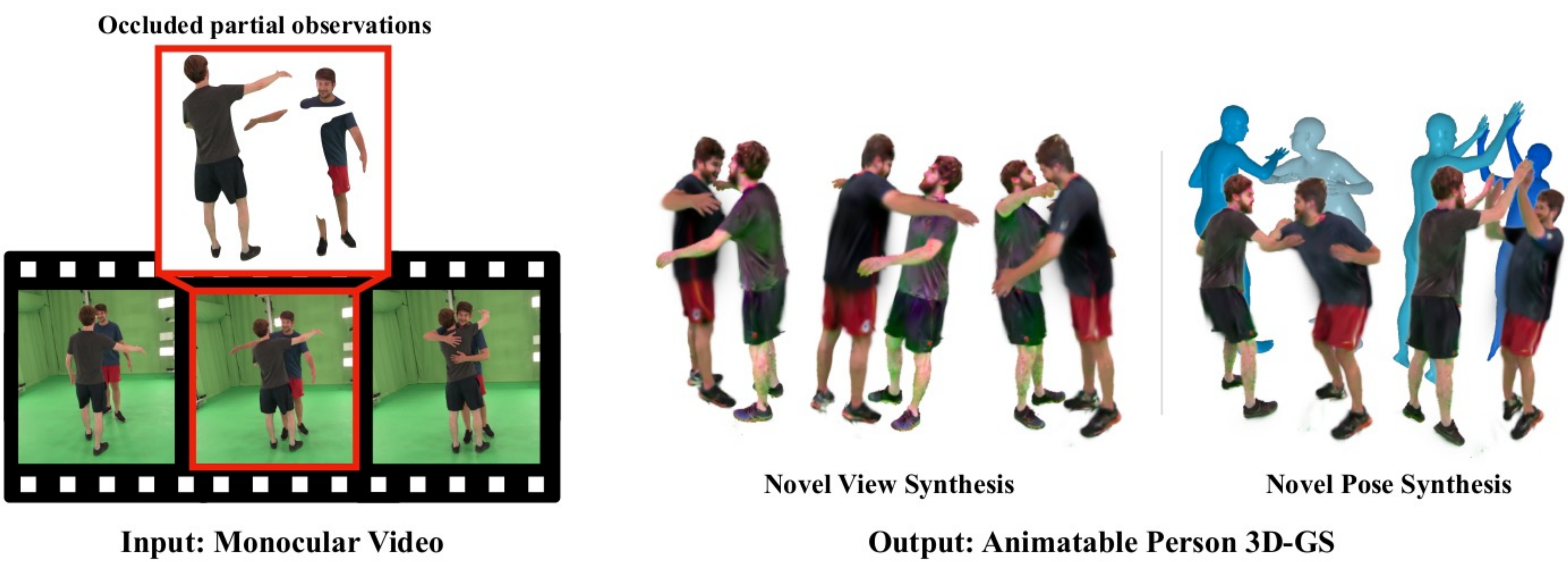
- Reconstruction or modeling from 3D point clouds
- Missing data



Laser scan Poisson MPU

3D challenge: classical acquisition

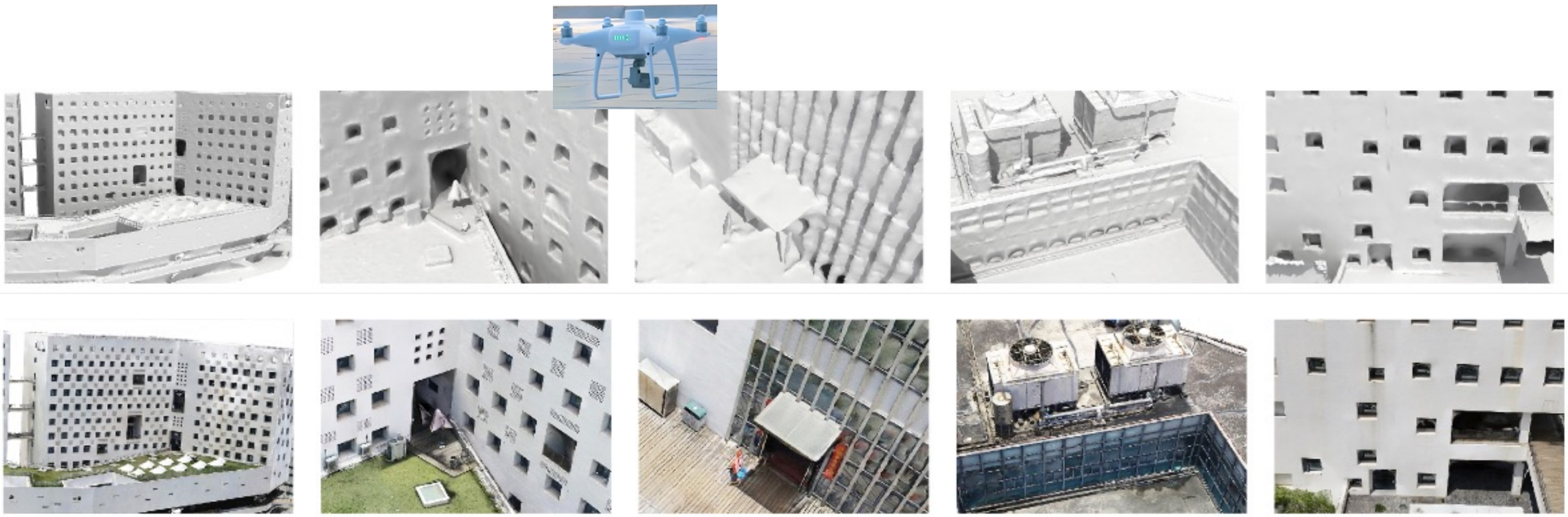
- Reconstruction or modeling from 3D point clouds
- Missing data + dynamic data



[Lee et al. *CVPR* 2024]

3D challenge: classical acquisition

- Reconstruction or modeling from 3D point clouds
- Missing data + dynamic data + large scales



[Liu et al. *SIGGRAPH Asia* 2022]

3D challenge: acquisition in 2025

- Image-to-3D, text-to-3D, or text-to-image-to-3D

Text:
"A bunny on pancakes"

Single-view image

NeRF [Mildenhall et al. ECCV 2020] via multi-view

Magic3D [Lin et al. CVPR 2023]

3D challenge: acquisition in 2025

- Image-to-3D, text-to-3D, or text-to-image-to-3D

Text:
“A potted cactus plant”

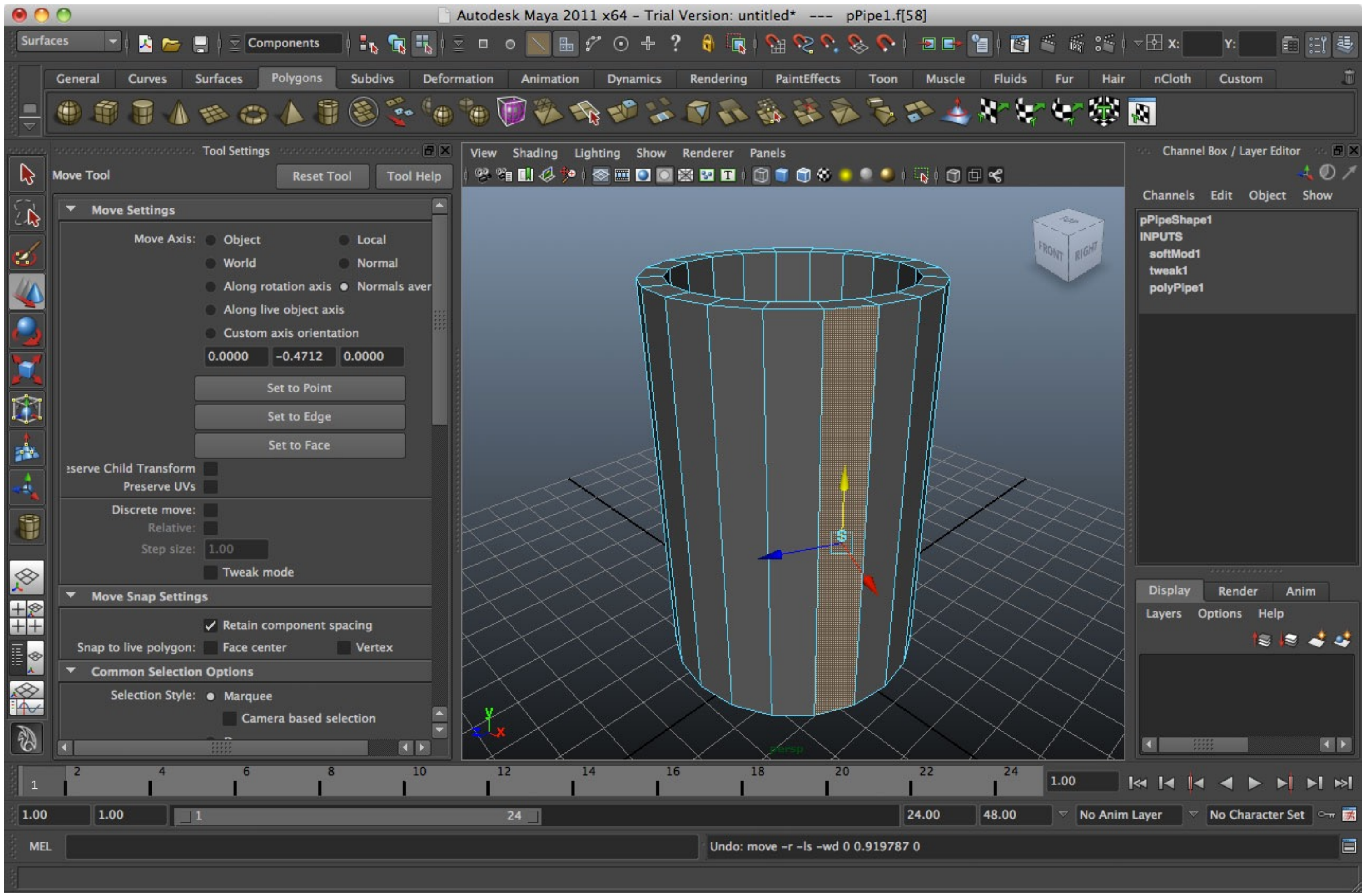
Single-view image

3D Gaussians via multi-view
[Kerbl et al. SIGGRAPH 2023]

DreamGaussian
[Tang et al. ICLR 2024]

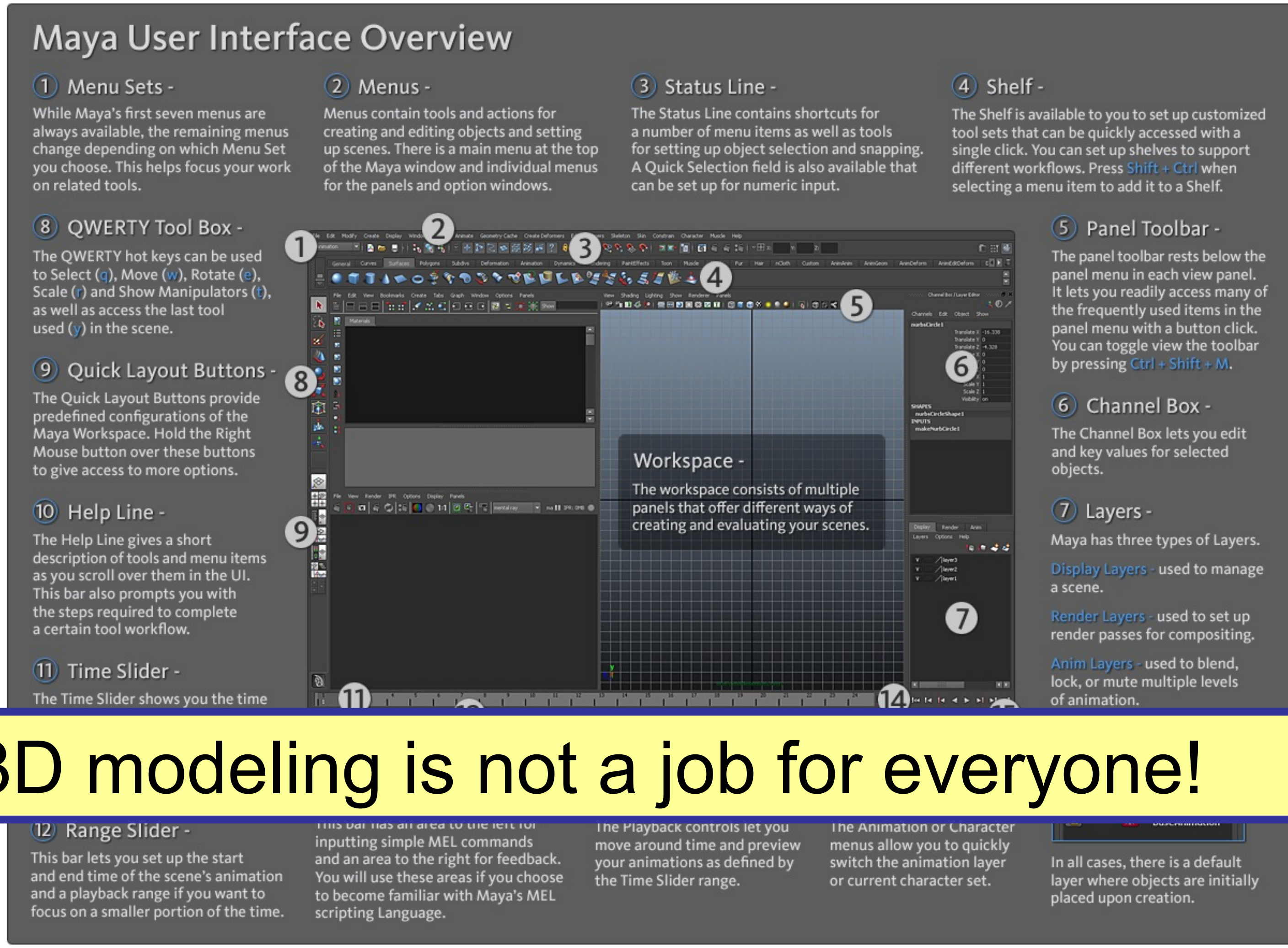
3D challenge: modeling

Autodesk Maya



3D challenge: modeling

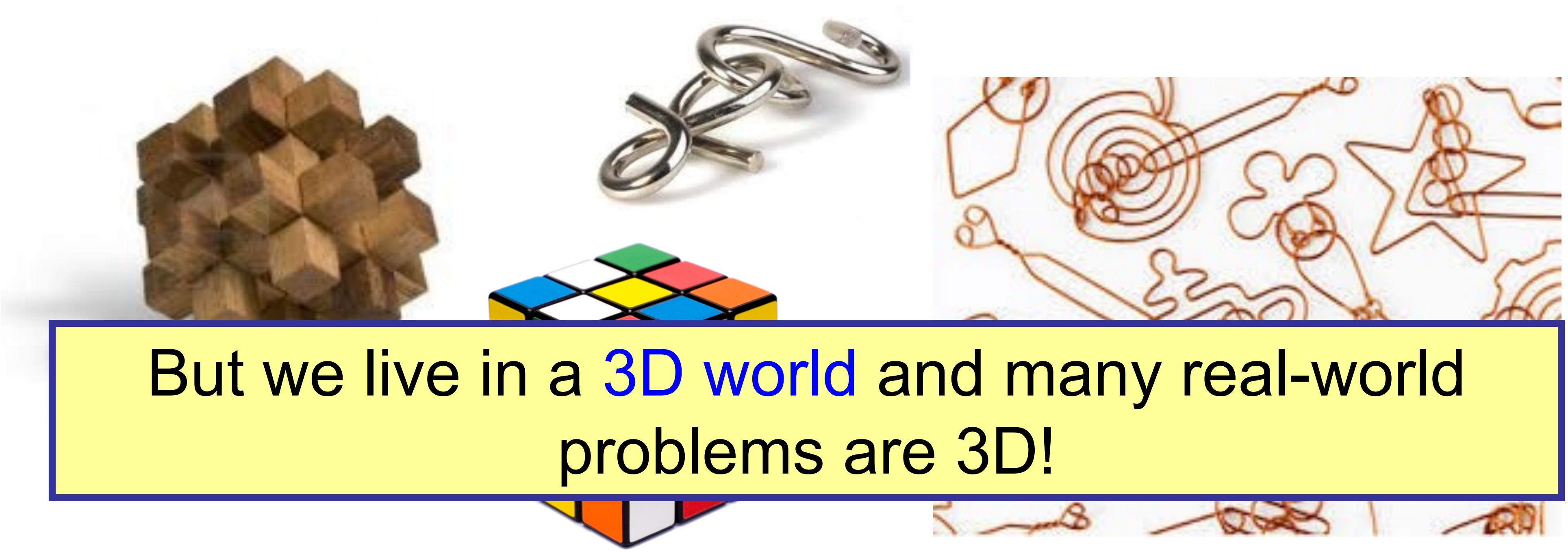
Autodesk Maya



3D modeling is not a job for everyone!

3D reasoning is hard

- Humans very good at pattern recognition (vision)
- But not so good at 3D reasoning or manipulation



But we live in a 3D world and many real-world problems are 3D!

Many real-world problems are 3D

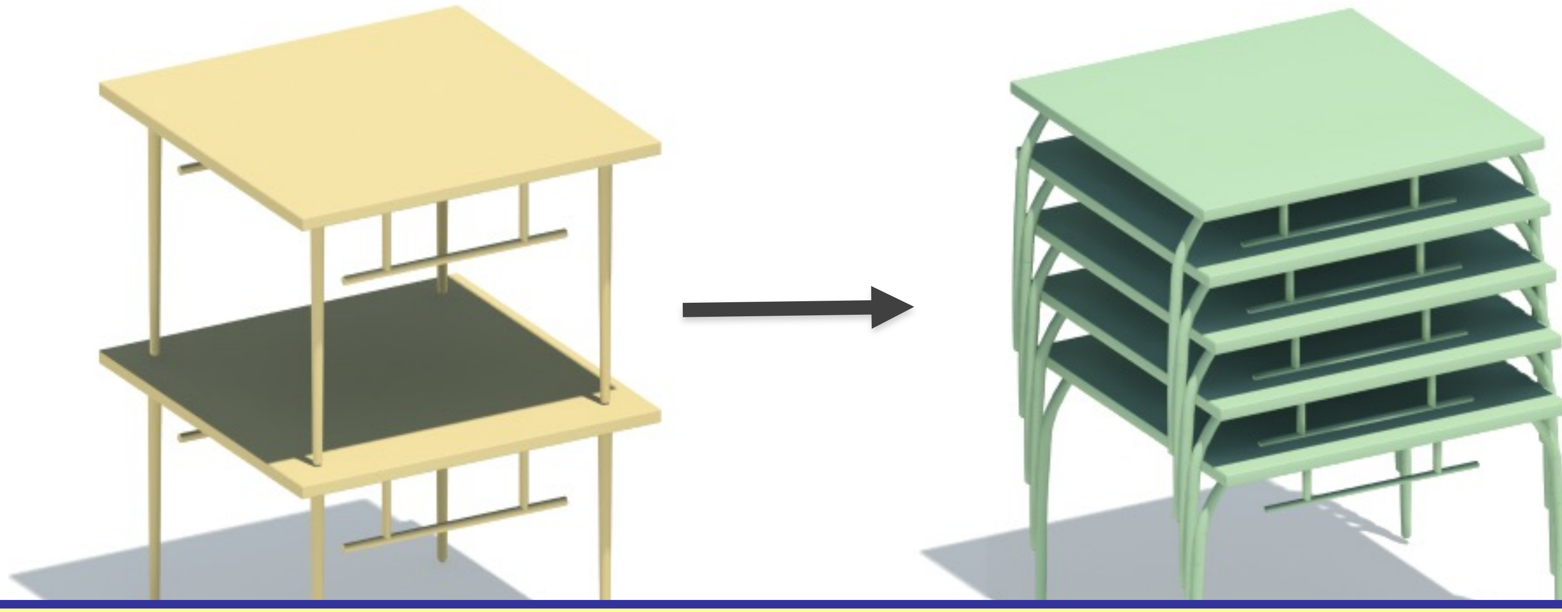
How to easily generate many 3D indoor scenes that are realistic and diverse, e.g., for AR/VR?

Many real-world problems are 3D

- How to subtly make the table stackable?

Stackabilization

- How to subtly make the table stackable?



Requires precise measurement and transform of 3D objects: difficult for human users to model

Foldabilization

- How to subtly make the chair foldable?

Like solving a puzzle:
acute 3D spatial reasoning
skills are needed

Computationally hard with
very large search space

[Li et al. *Siggraph* 2015]

Foldabilization

EMBRY-RIDDLE Aeronautical University
SCHOLARLY COMMONS

International Journal of Aviation, Aeronautics, and Aerospace

Volume 8 | Issue 1 Article 1

2021

Folding Methodology for Flexible Aircraft Interiors

Aditya Venkatesh
Ryerson University, Toronto, Canada, avenkatesh@ryerson.ca
Fengfeng (Jeff) Xi
Ryerson University, Toronto, Canada, fengxi@ryerson.ca
Joon Chung
Ryerson University, Toronto, Canada, j3chung@ryerson.ca

concepts.

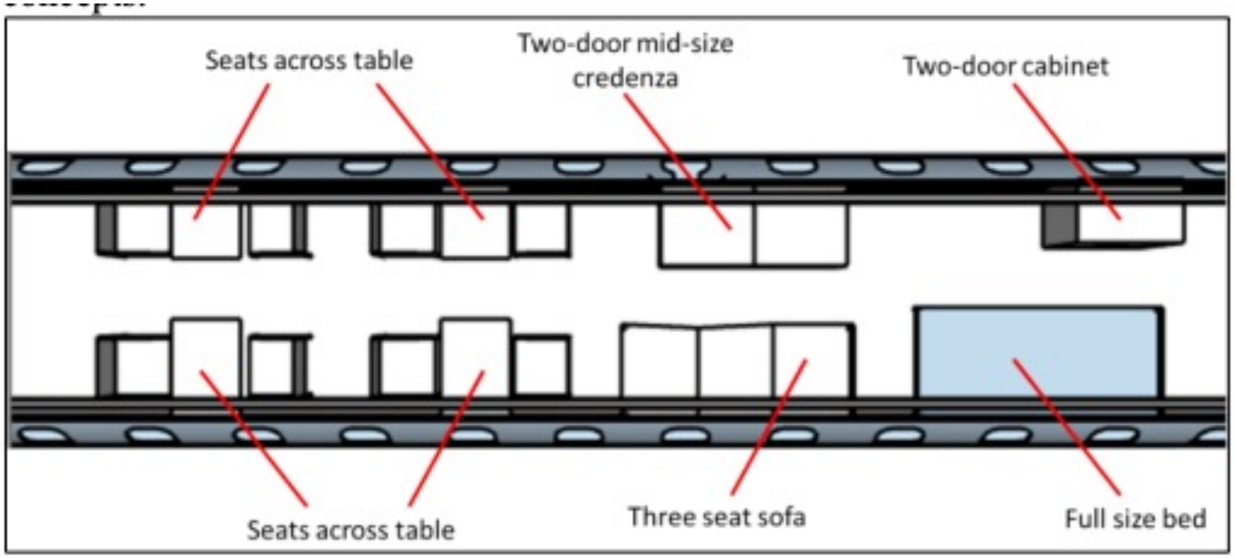


Figure 3. Partially sectioned top view with extended furniture pieces.

Figures 4 and Figure 5 illustrate the Open-on-Demand concept along the starboard and port sides of the partially sectioned cabin with zero-thickness furniture pieces. The area highlighted in yellow refers to the floor space the furniture pieces occupy in their extended forms.

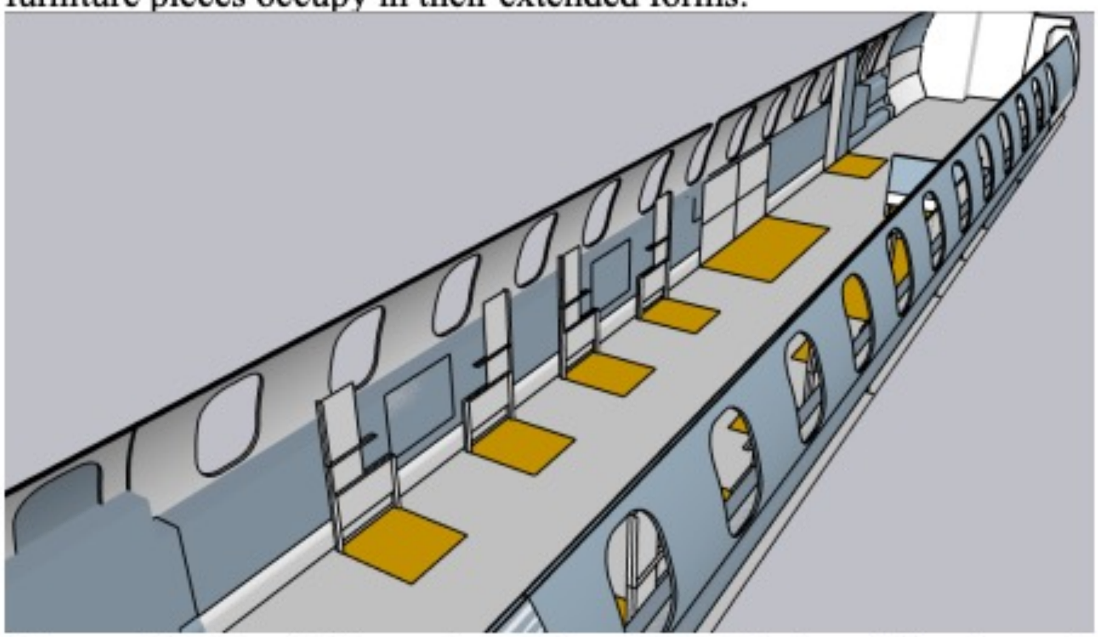


Figure 4. Starboard side partial section view with stowed furniture pieces.

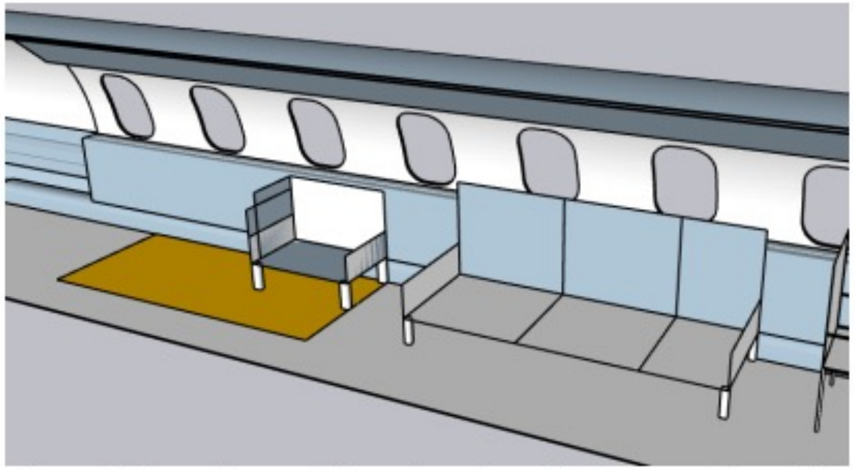


Figure 9. Port side partial section view of a bed reconfigured into a seat.

The furniture folding methodology can also be quantitively assessed to analyze advantages. A zero-thickness seat frame with measurements reflecting the lounge seat on a business jet was recreated using SketchUp 2020. Figure 10 illustrates an extended seat and stowed seat, following the furniture folding methodology, with a length of 27.25", width of 31.33" and height of 44.23" and new a length of 2", width of 31.33" and height of 44.23", respectively.

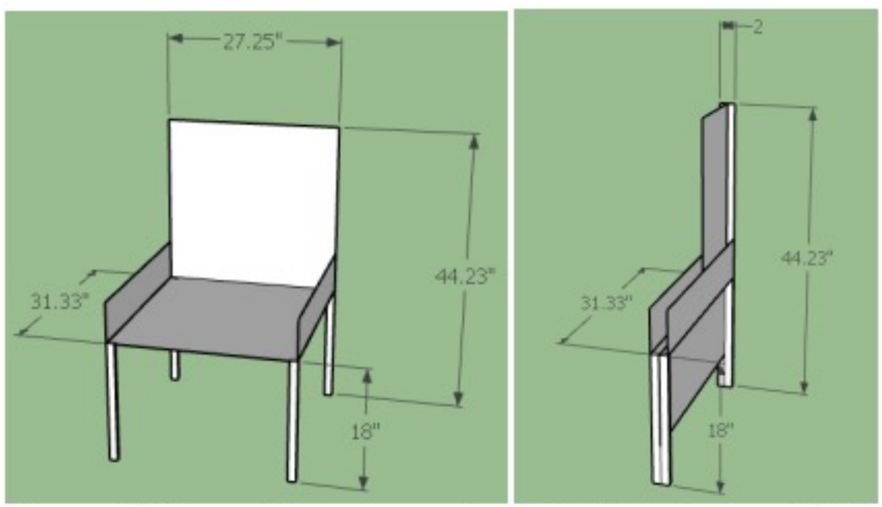
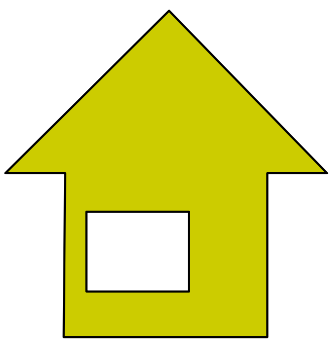
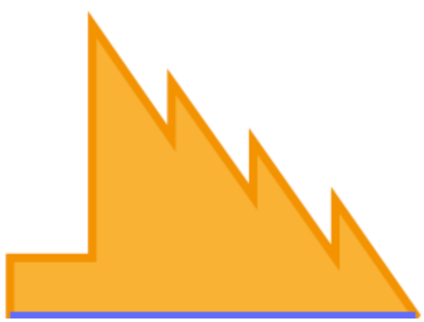


Figure 10. Recreation of an extended(left) and stowed(right) zero-thickness business jet lounge seat frame.

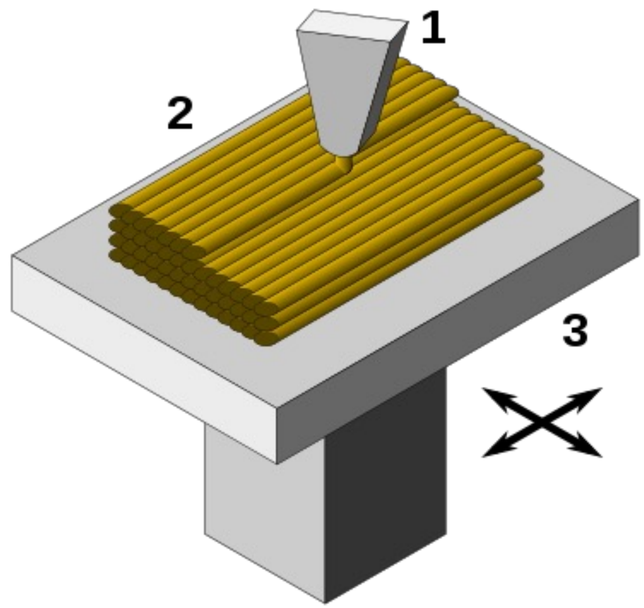
A cool decomposition for 3D printing

- How to decompose into few terrain-like parts?





Terrain-like part



Zero material waste for layer-based 3D printing

Pyramidal decomposition

- How to decompose into few terrain-like parts?

[Hu et al., *Siggraph Asia* 2014]

Pyramidal decomposition

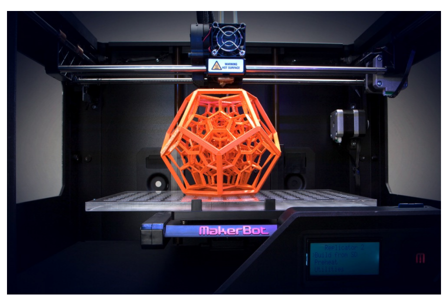
- How to decompose into few terrain-like parts?

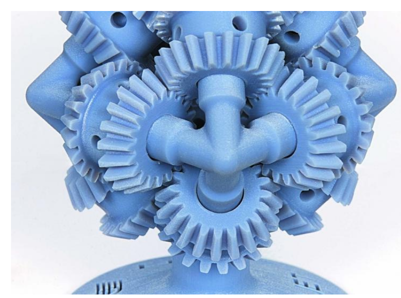
3D problem is provably NP-hard

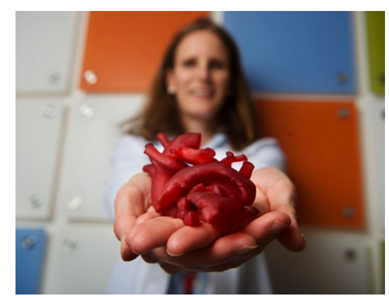
Computer graphics is responsible for addressing the various 3D data challenges

3D printing may be a blessing

- Graphics likes 3D to be wanted & used everywhere
- The internet has not made 3D data ubiquitous as promised: remember VRML around 15 years ago?
- 3D printing just might!







New graphics

- Keep doing synthesis, but focus on modeling
- Synthesis of all visual contents, not just images

explicit model description → synthesized content

rendering

New graphics: no explicit model

- Keep doing synthesis, but focus on modeling
- Synthesis of all visual contents, not just images

~~explicit~~ model description → synthesized content

rendering

Model description is only abstract (e.g., texts or a sketch), hard to quantify (functional or creative), or unknown entirely (input = set of examples)

New graphics: novel content

- Synthesis and manipulation of ~~images~~

$\boxed{\text{novel}}$ + $\boxed{\text{visual contents}}$

Implicit or abstract inputs, examples, … → novel content especially 3D, e.g., geometric modeling

New kinds of inputs

A rough sketch

New kinds of inputs

A rough sketch

One or more images

New kinds of inputs

A rough sketch

One or more images

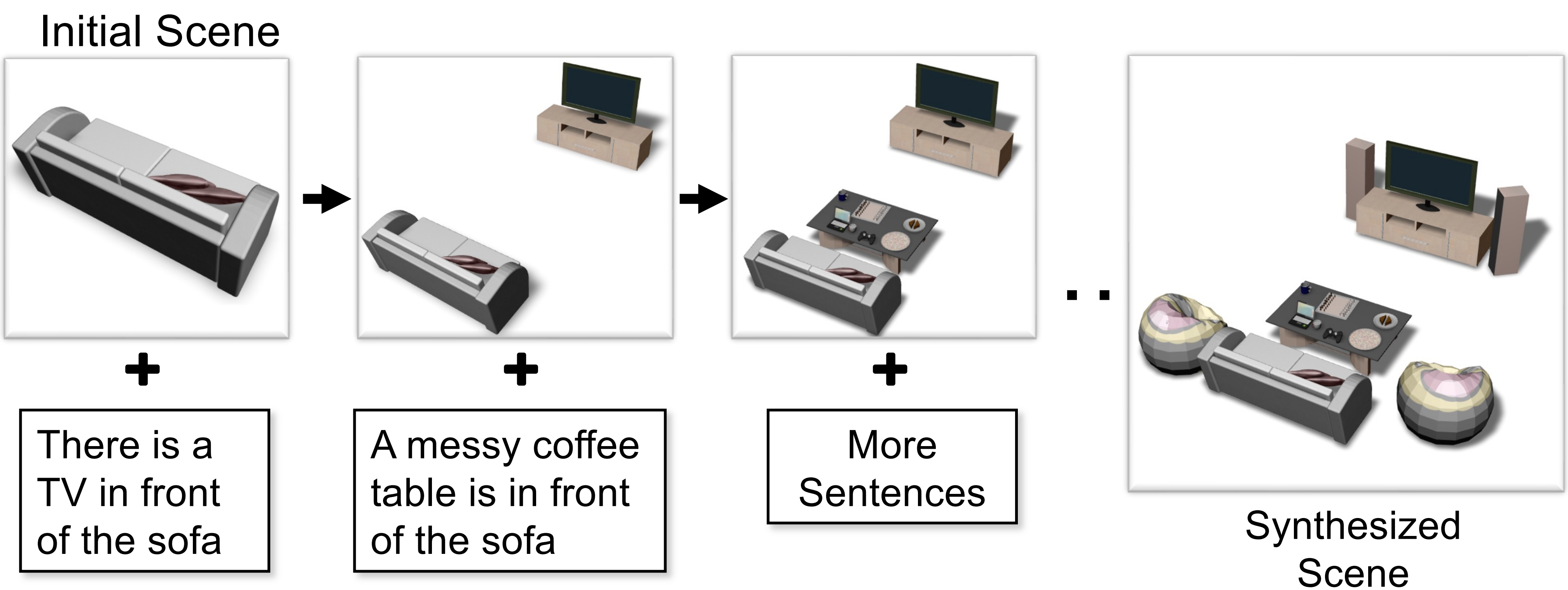
Just some examples

Text-to-3D

"A baby bunny sitting on top of a stack of pancakes"

3D model generated from text [Zhu et al. 2023]

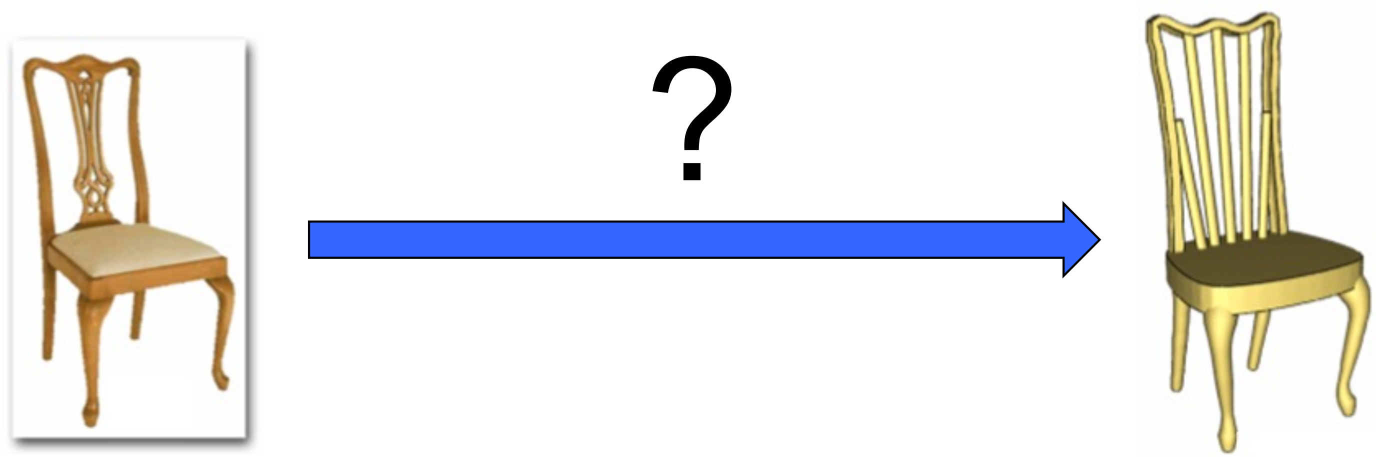
Text-driven 3D scene synthesis



[Li et al. *SIGGRAPH* 2018]

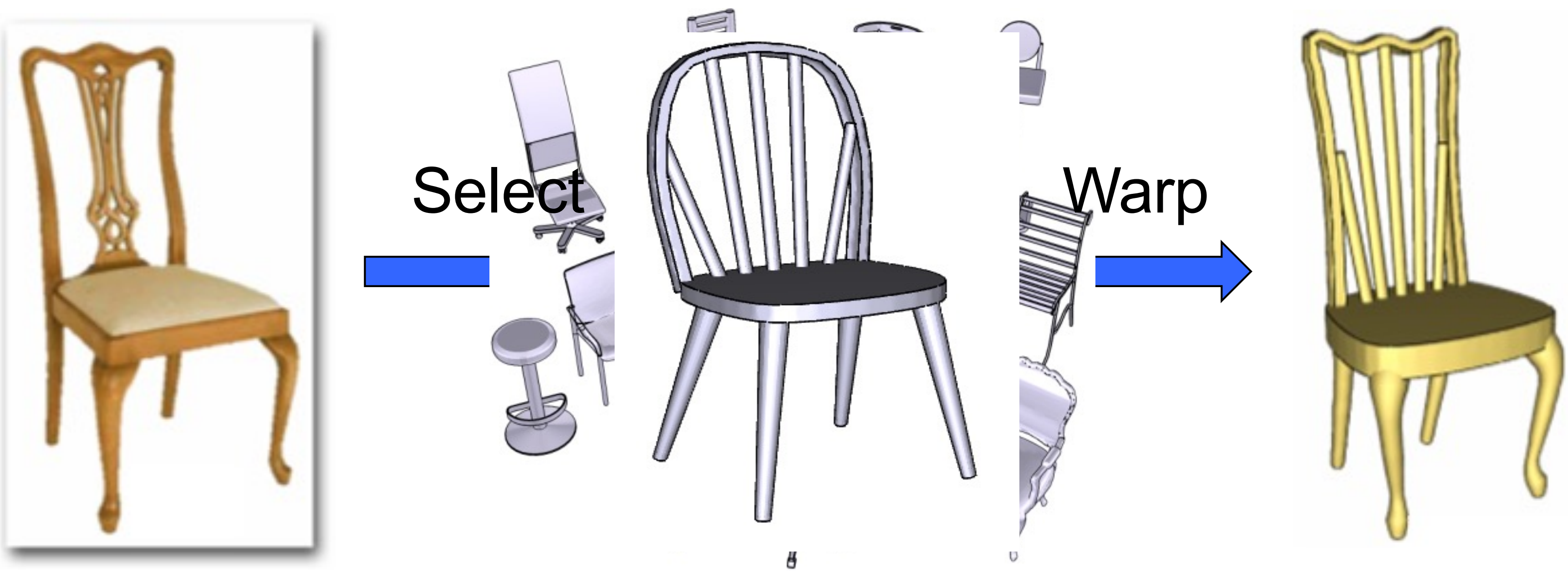
Photo-inspired modeling

- 3D model from a single photograph



Has to be data-driven

- Abstract inputs: ill-posed synthesis problem
- Needs extra knowledge, e.g., pre-existing dataset



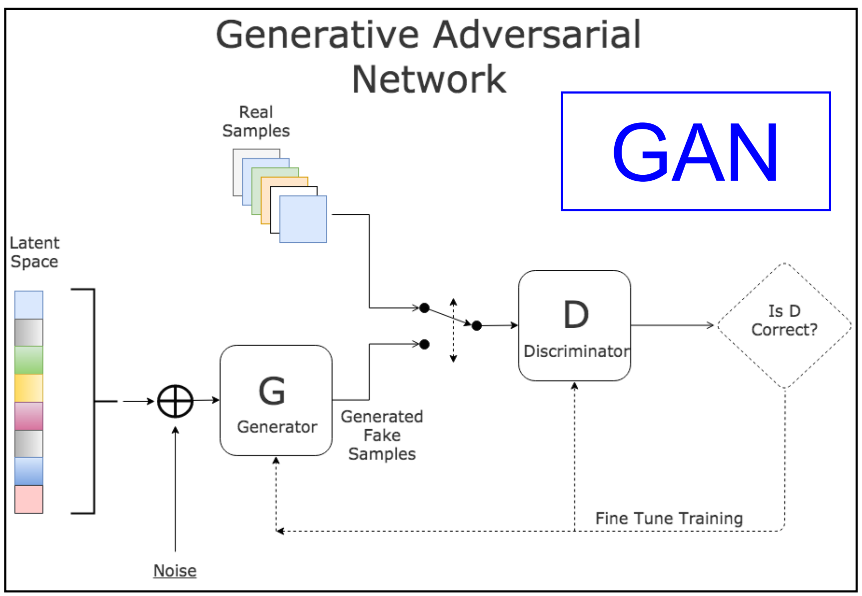
[Xu et al. *SIGGRAPH* 2011]

Example-based modeling

- Key: understand the set of examples by learning
- To infer commonality and diversity in the set

Inverse modeling

- Learn a generative model, e.g., from examples
- Apply the model forwardly, maybe with a random input, to synthesize novel contents



In the new graphics

We generate/synthesize objects/scenes that are

- Plausible: “What makes a chair a chair?”
- With the right style: “Essence of Gothic style?”
- Functional: “Shape vs. functional similarities?”
- Ergonomic: “How to quantify human comfort?”
- Creative: “What is a model of human creativity?”
- …

To generate/synthesize

2D/3D objects or scenes that are

- Plausible, stylistically compatible, aesthetically pleasing, functional, ergonomic, or creative, etc.

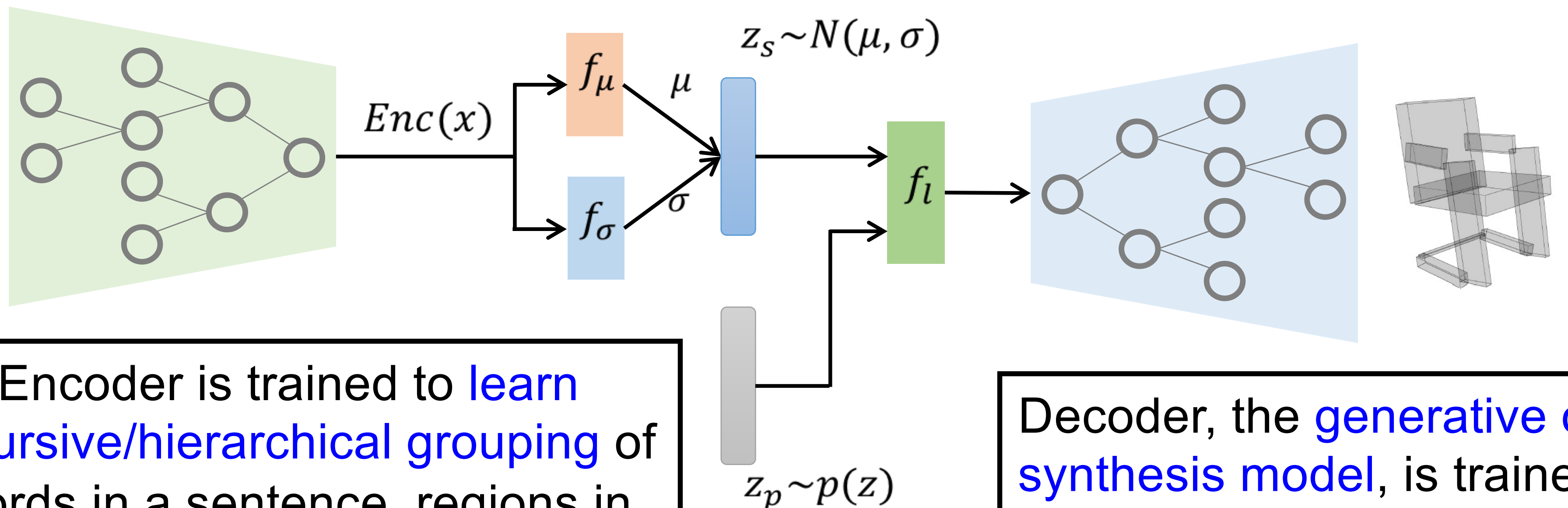
- We must first learn plausibility, style, function, etc.
- To generate from texts/sketches, we must learn the right mapping/regression model

Workflow of new graphics research

- Analysis to acquire understanding of
  - grouping/clustering patterns, object/scene composition, human activities, styles, functionality, creativity, etc.
- Synthesis of images, shapes, scenes via
  - interactive modeling, genetic algorithm, statistic models, deep regression and generative neural networks, etc.

Example: generative structural reps

- Generative autoencoders via recursive neural nets



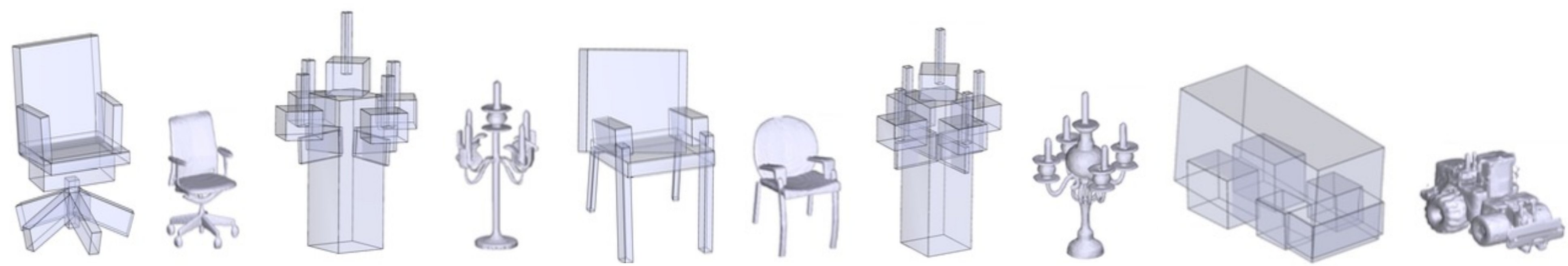
Encoder is trained to learn recursive/hierarchical grouping of words in a sentence, regions in an image, parts in a 3D shape, or objects in an indoor scenes

Decoder, the generative or synthesis model, is trained to convert a random code into a hierarchy

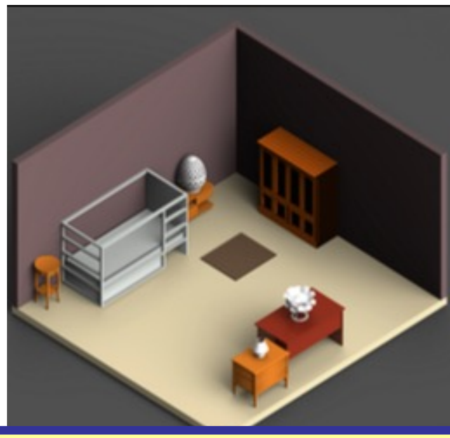
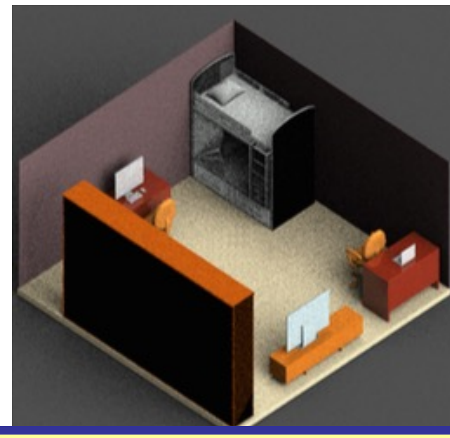
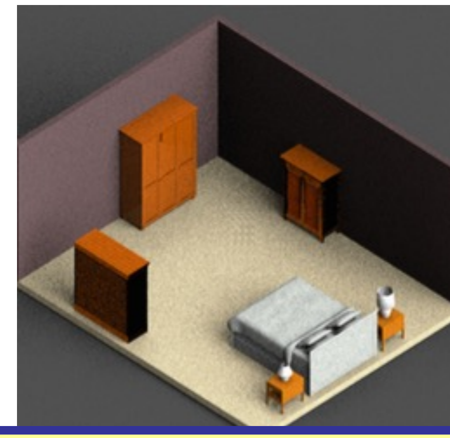
GRASS: Generative Recursive Autoencoders for Shape Structures [SIG 2017]

Example: generative structural reps

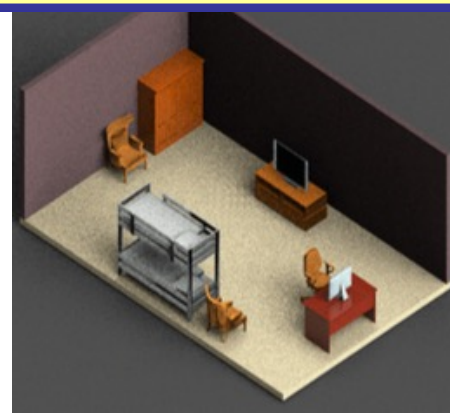
- Combine autoencoder with GAN: VAE-GAN
- Structure-aware; coarse-to-fine; clean parts; high res



Example: scenes from random codes

Takes less than one second to generate a 3D scene!


GRAINS: Generative Recursive Autoencoders for INdoor Scenes [2018]

New graphics: not forward problem

- Inverse analyses and learning generative models
- Keying on shape/scene understanding
- Only with a good understanding of a shape/scene category (“bicycle” or “kitchen”) can one recreate!

Important problems in new graphics

- Modeling from abstract description, e.g., texts
- Modeling from few examples
- Inverse procedural modeling
- Learning generative neural networks

Knowledge, learning, and data play the key roles!

Our new view of graphics

- Wikipedia was already catching up:
  - Something I hid: Computer graphics = methods for digitally synthesizing and manipulating visual content
- From image production to all (3D) visual content

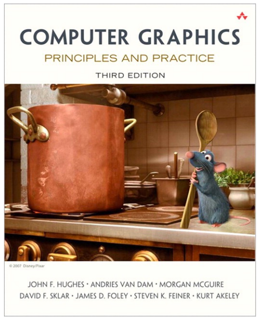
Novelty

of the synthesized content is the BIG challenge!

Current graphics and computer vision

- Hughes, van Dam, et al.:
  - "Much of current research in graphics is in methods for creating geometric models, methods for representing surface reflectance, the animation of scenes …, and in recent years, an increasing integration of techniques from computer vision."

------ *page 2 of*



Computer graphics vs. vision again

- Shape understanding and inverse modeling are very much “vision-like” research problems
- So, graphics is “catching up” ☺
- But could we do more?

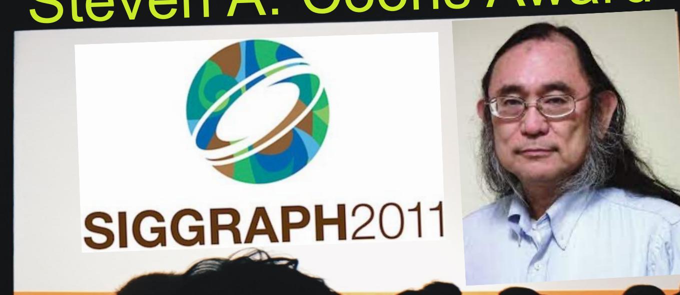

Steven A. Coons Award
SIGGRAPH2011

Jim Kajiya: What human capabilities does each CS discipline try to enhance/replace?

Artificial intelligence:
human intelligence

Computer vision:
pattern recognition

Steven A. Coons Award



What about computer graphics?

Human imagination!

Graphics and imagination

- Think of the various VFX we see in films and games
- Think of the VR/AR/MR CG has helped create

Computer graphics allows our imagination to be realized into (virtual) reality!

It allows a mental concept to turn into a digital representation, and now fabricated!

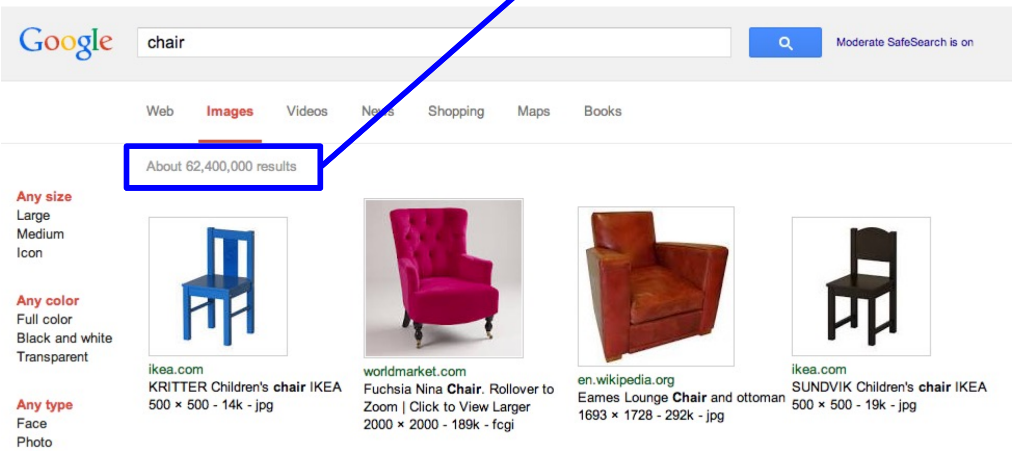
Still a long way to go

- Only scratching the surface in the new graphics
- Smart ideas: data-driven, data reuse, co-analysis, supervised learning, active learning, etc.
- Future of modeling in computer graphics

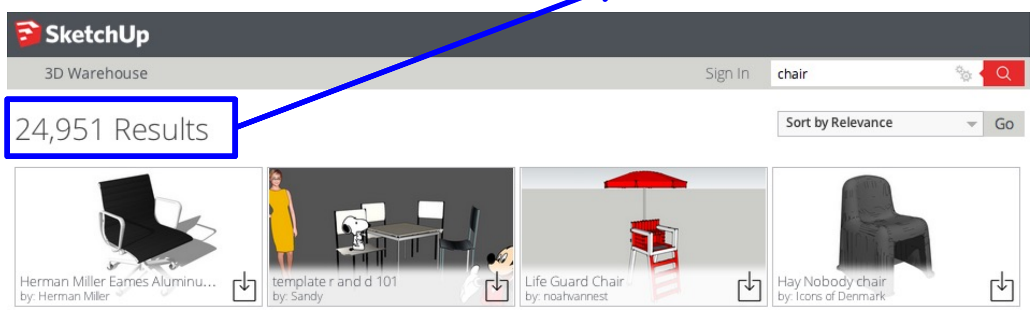
Data + knowledge + learning

A new 3D data challenge

Google image search for chair: 64,000,000 results

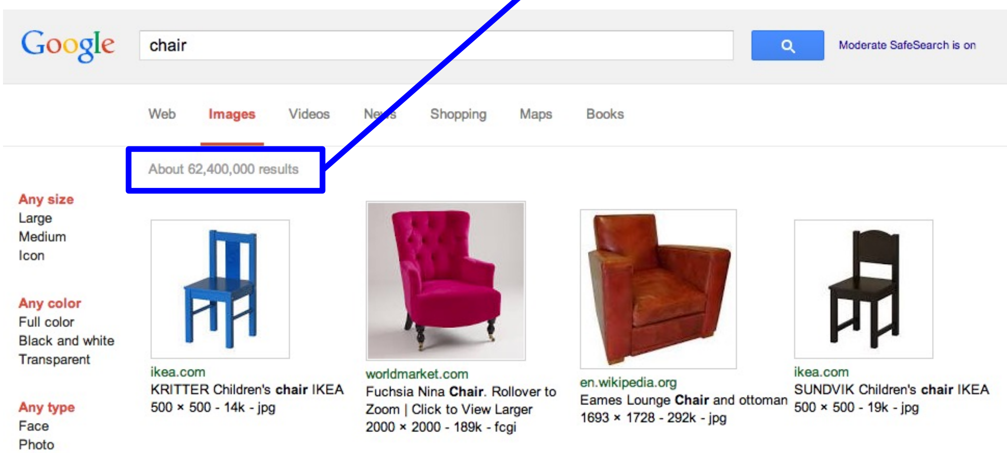


3D Warehouse: 24,951 results

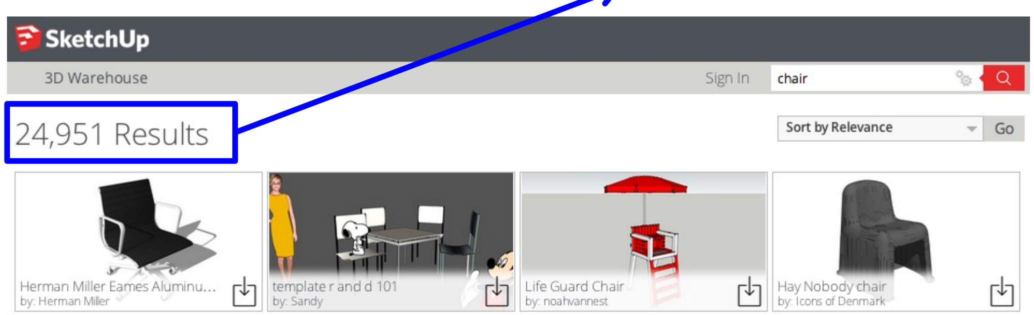


No “Big 3D Data” yet

Google image search for chair: 64,000,000 results



3D Warehouse: 24,951 results



- Bicycles

  29,900,000 vs. 1,225

- Strollers

  5,070,000 vs. 36

Problems with “small 3D data”?

- Lack of knowledge for learning-based 3D analyses
- Lack of examples for example-driven 3D syntheses
- Small data is *the* detriment to

Data + knowledge + learning

To fix the "small 3D data" problem

- Need to synthesize more and more 3D models!
- Not only volume, but variation, variety, and novelty!

Novelty of synthesized content may enhance knowledge!
Computer graphics is responsible for producing such Big 3D Data

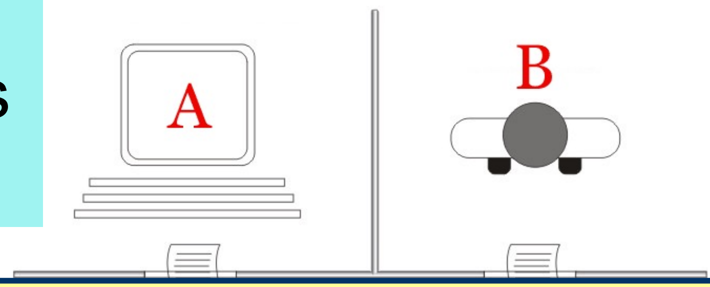
What is an AI?

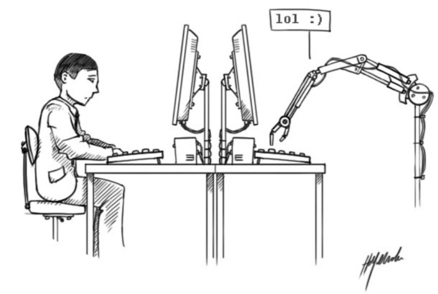
What about machine becoming human?

When is a machine becoming human?

- Well-known Turing test: indistinguishability between human and machine in natural conversation

Multiple human judges vote who is human/machine

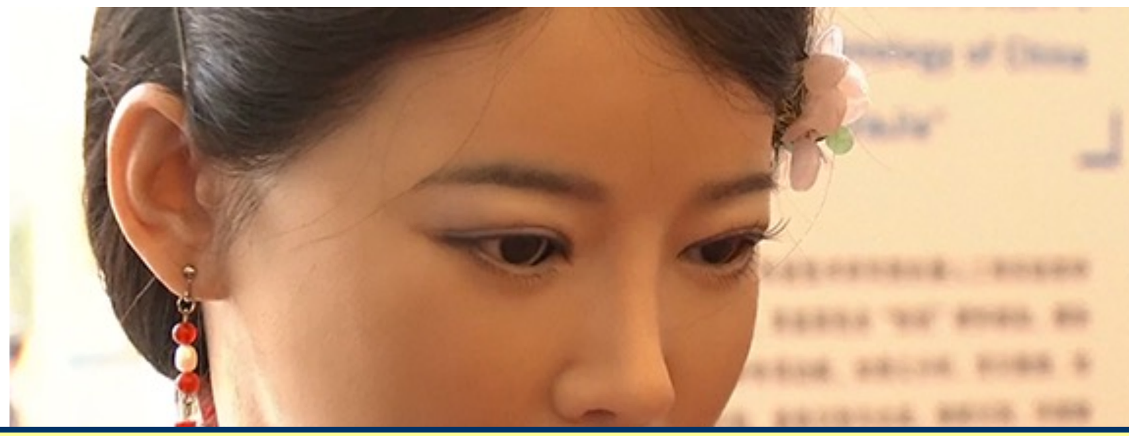




Turing (1950’s) predicted the test would be passed around year 2000. An easy version of the test was passed in 2014 by *Eugene*, a ChatBot.

When is a machine becoming human?

- Total Turing test: machines linguistically and physically indistinguishable from a human



Hans Moravec (1999): Total Turing Test to be passed by the year 2040

Is this Turing test too easy?

- Ada Lovelace (1815-1852)
  - Pre-dates Turing (1912 – 1954)
  - Worked on world's first computer
  - AKA: the first computer programmer

Computers can't create anything (Humans can!). Creation requires originality. But computers originate nothing; they only do that what we order them, via programs, to do.
- Ada Lovelace

A harder test: Lovelace Test

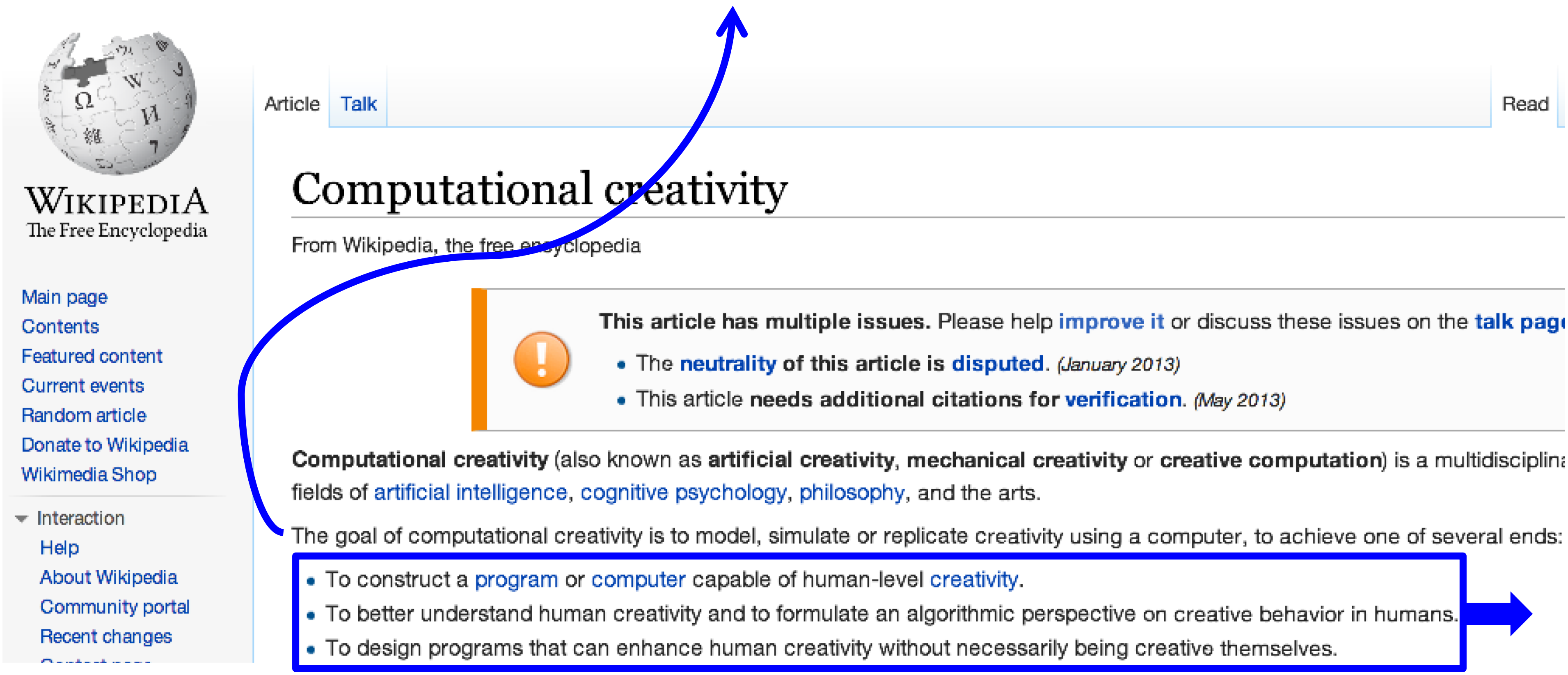
- Test on machines' ability to create an artifact
  - e.g., a story, poem, painting, or a 3D shape
- Test or judging criterion can vary

Key message: What really separates humans from machines is not the ability to make conversation, but the ability to create!

Creation and synthesis of visual content is the goal of computer graphics!

Computational creativity

- Goal: model or simulate creativity using a computer



Goals of computational creativity

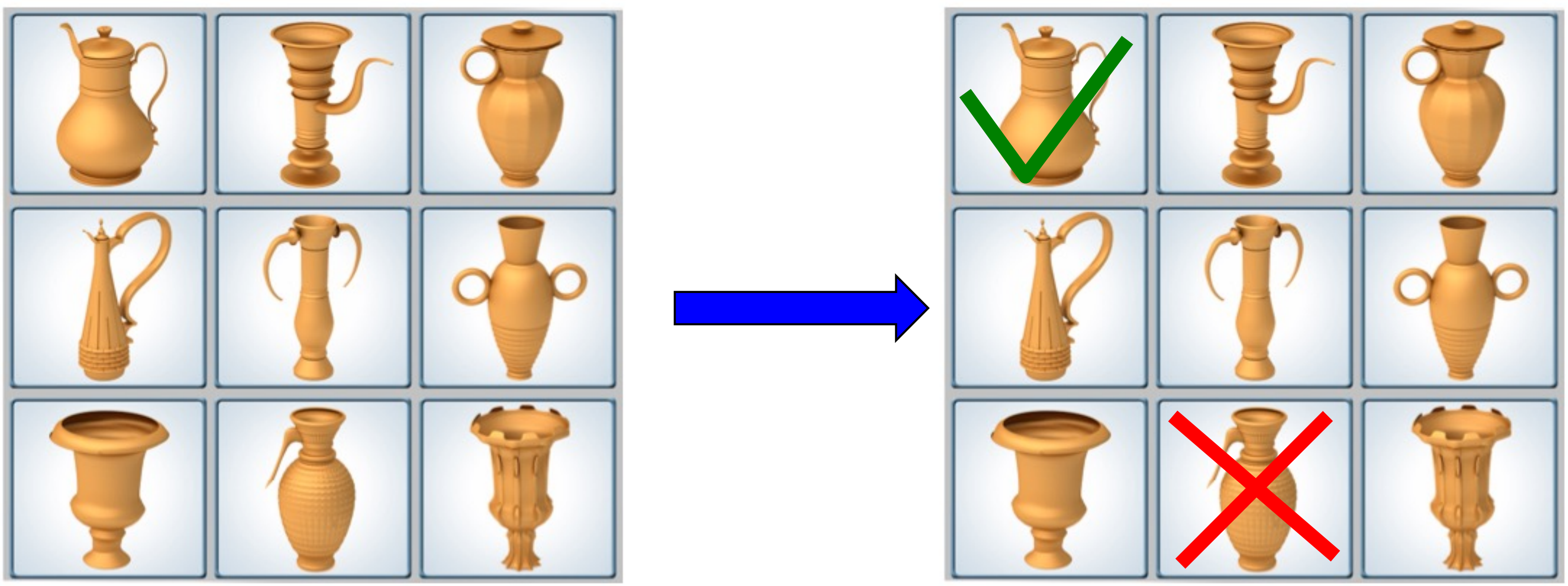
*** Construct a program capable of human creativity

** Understand creativity and formulate an algorithmic perspective on creative behavior in humans

* Design a program which may enhance human creativity without the program being creative itself

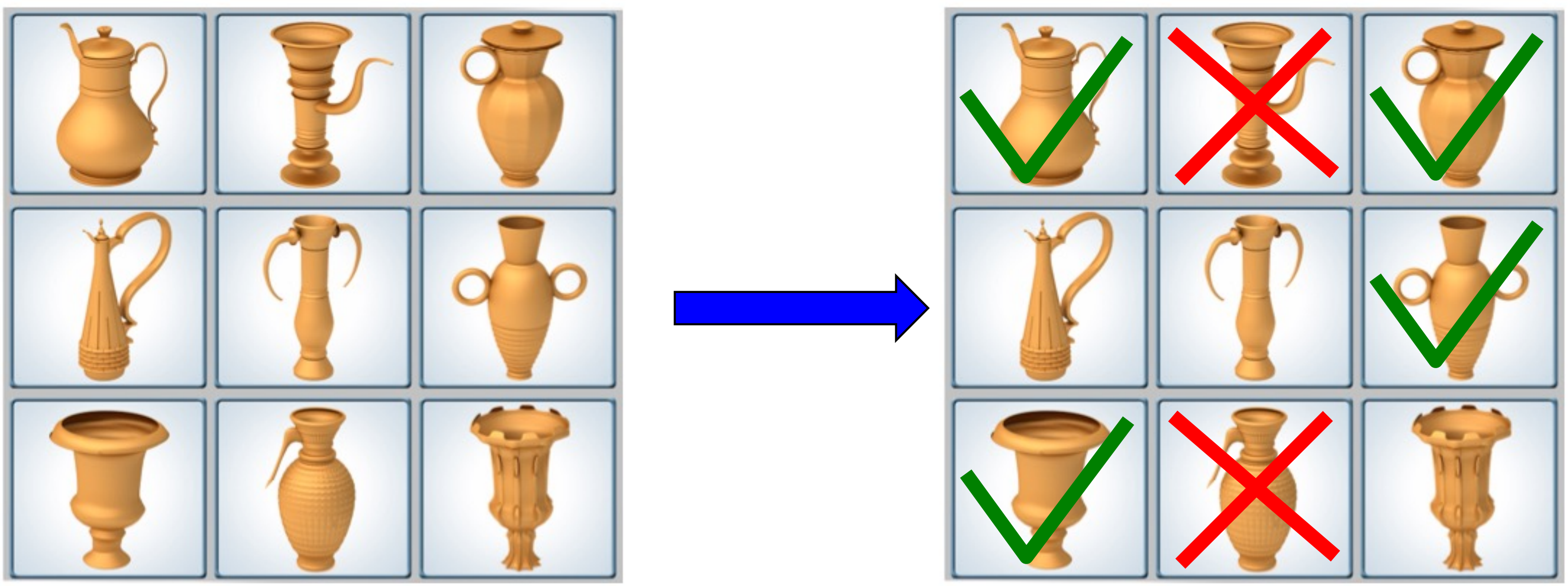
Creative 3D modeling

- Creativity: machines stochastically generate models
- Has to be controlled



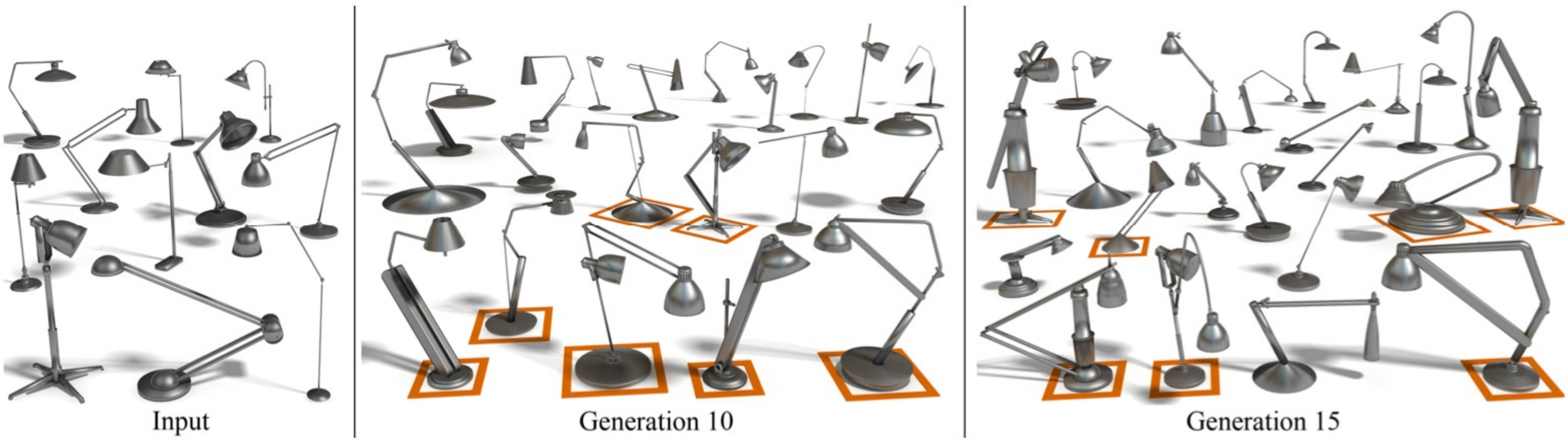
Creative 3D modeling

- Creativity: machines stochastically generate models
- Control by humans operating on a “design gallery”



Creative 3D modeling via evolution

- Evolve an entire set (initial population) to obtain generations of fit and diverse new creations



[Xu et al. *SIGGRAPH* 2012]

Creative 3D modeling

- Allows creative generation of novel 3D contents

Genetic algorithm + Interactive modeling

Very far from an intelligent machine that is creative

Machine to generate creative logos?

What can we do now?

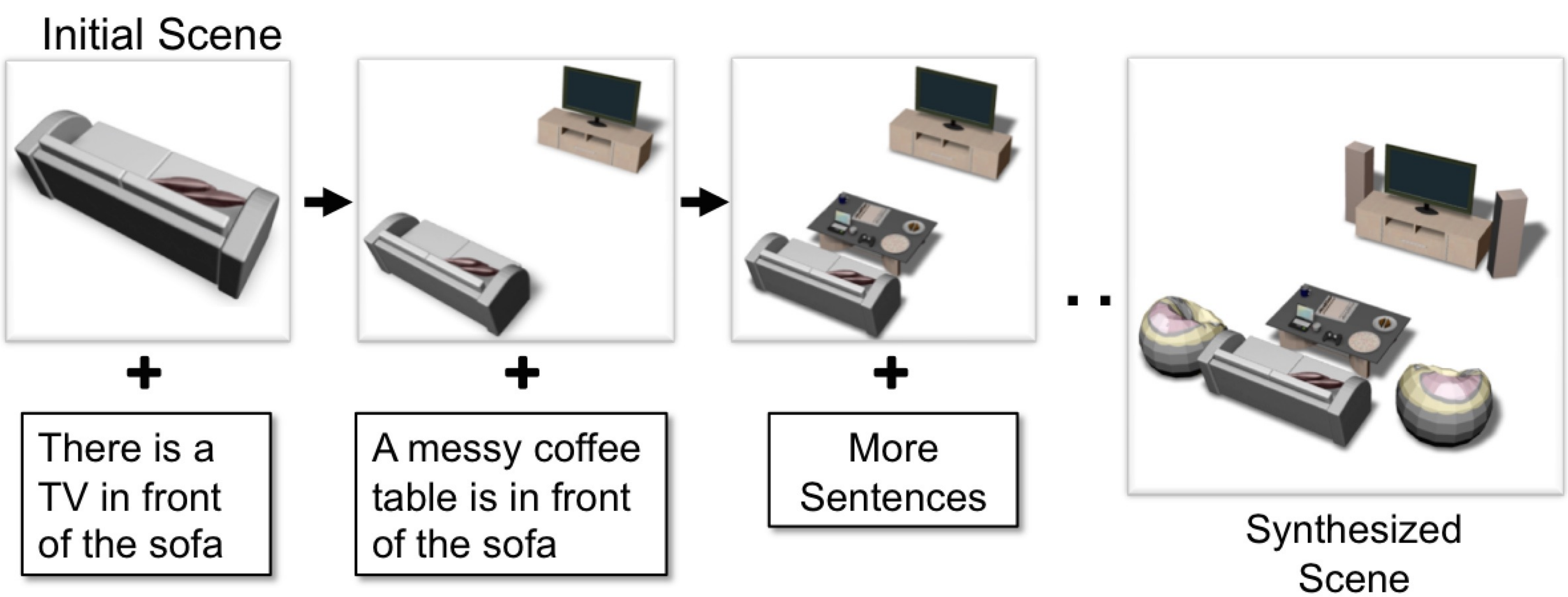
[Tanveer et al. *ICCV* 2023]

DS-Fusion: Artistic Typography via Discriminated and Stylized Diffusion
Paper ID 1938

Graphics in the age of AI
Graphics is accessible for

Graphics is responsible for

- Addressing the various 3D data challenges: acquisition, modeling, interaction, etc.

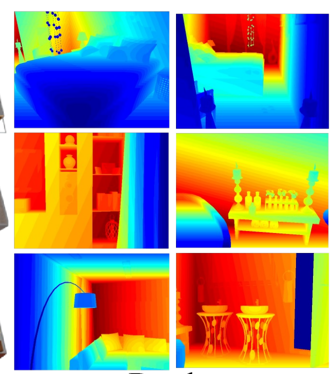


Graphics is responsible for

- Producing BIG 3D data

ScanNET

SUNCG dataset



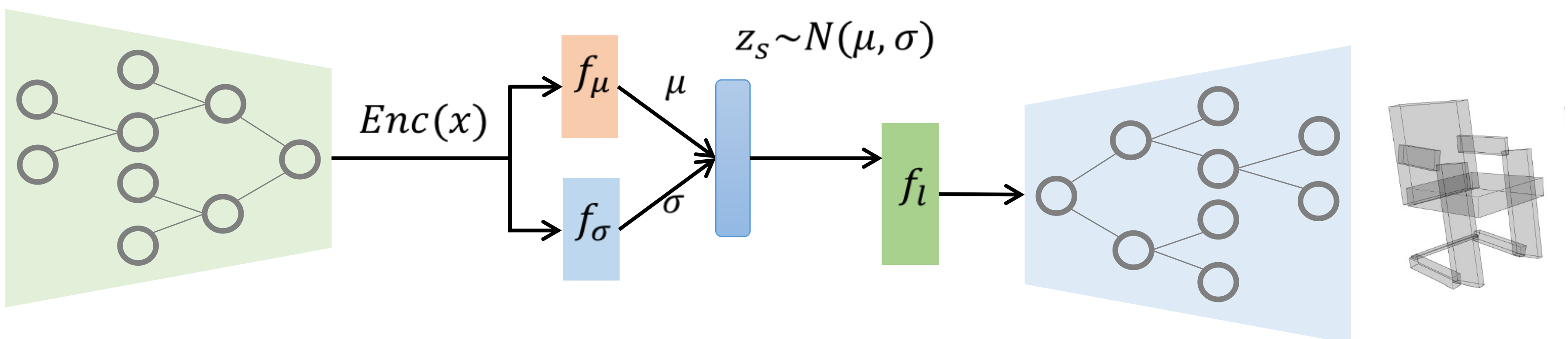
Depth

Physical based rendering

Volumetric ground truth

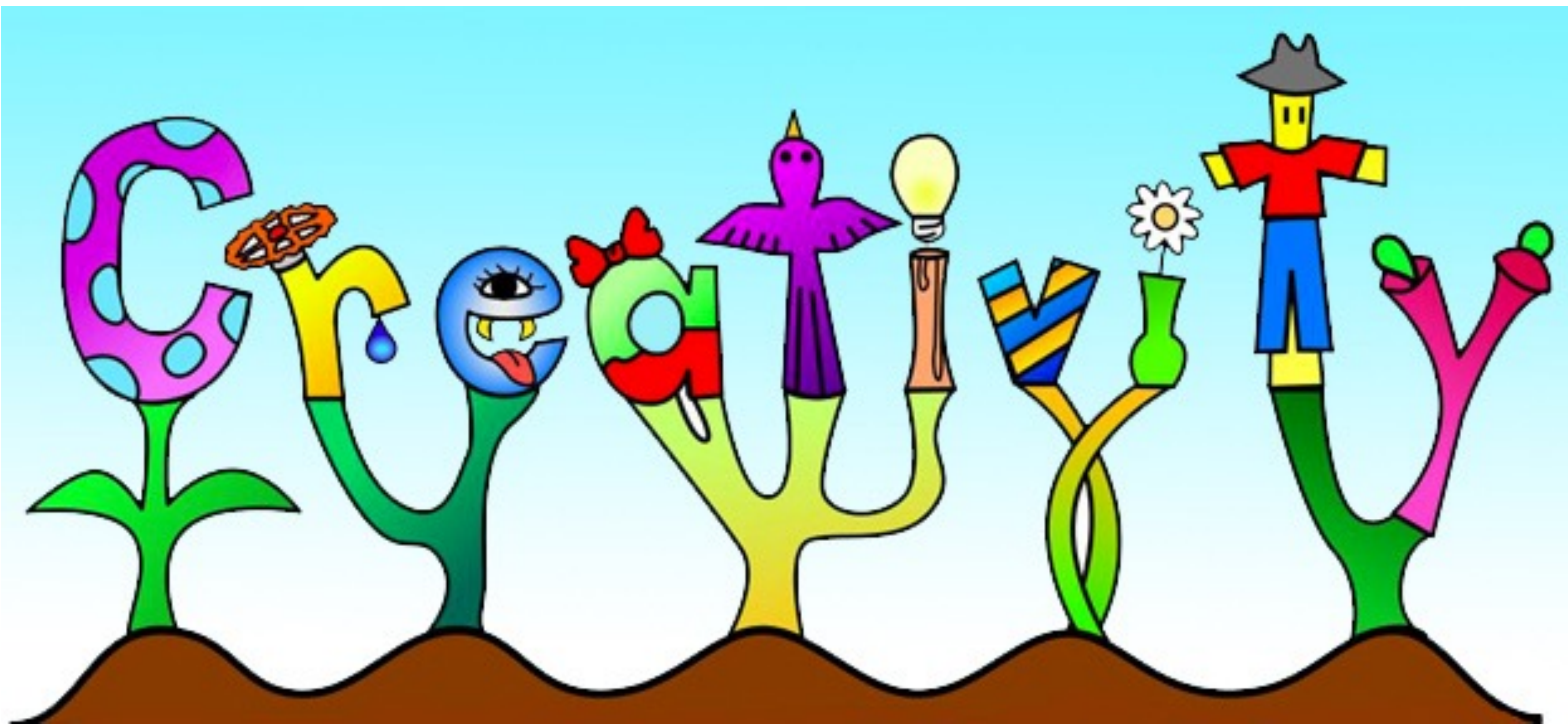
Graphics is responsible for

- Training machines and neural networks capable of generating (novel) 3D content from abstract, implicit descriptions, sets of examples, etc.



Graphics is responsible for

- Realizing and enhancing human imagination



New graphics: synthesis challenges

- No explicit model description
- Synthesize novel 3D content
- Synthesize Big 3D Data
  — 4 V's: Volume + Variation + Variety + NoVelty
- Synthesize creative 3D content = 3D generative AI

We are only scratching the surface!

Graphics in the age of robotics

Foundational model for robotics (RFM)

[Submitted on 13 Dec 2023]

Foundation Models in Robotics: Applications, Challenges, and the Future

Roya Firoozi, Johnathan Tucker, Stephen Tian, Anirudha Majumdar, Jiankai Sun, Weiyu Liu, Yuke Zhu, Shuran Song, Ashish Kapoor, Karol Hausman, Brian Ichter, Danny Driess, Jiajun Wu, Cewu Lu, Mac Schwager

We survey applications of pretrained foundation models in robotics. Traditional deep learning models in robotics are trained on small datasets tailored for specific tasks, which limits their adaptability across diverse applications. In contrast, foundation models pretrained on internet-scale data appear to have superior generalization capabilities, and in some instances display an emergent ability to find zero-shot solutions to problems that are not present in the training data. Foundation models may hold the potential to enhance various components of the robot autonomy stack, from perception to decision-making and control. For example, large language models can generate code or provide common sense reasoning, while vision-language models enable open-vocabulary visual recognition. However, significant open research challenges remain, particularly around the scarcity of robot-relevant training data, safety guarantees and uncertainty quantification, and real-time execution. In this survey, we study recent papers that have used or built foundation models to solve robotics problems. We explore how foundation models contribute to improving robot capabilities in the domains of perception, decision-making, and control. We discuss the challenges hindering the adoption of foundation models in robot autonomy and provide opportunities and potential

… significant research challenges remain, particularly around the scarcity of robot-relevant data, safety guarantees and uncertainty quantification, and real-time performance.

3D datasets for RFMs

- 😀 High-quality: e.g., Amazon-Berkeley Objects (ABO)

3D datasets for RFMs

- 😀 High-quality
- 😀 Affordance: robots know what objects can do, how they are used

3D AffordanceNet [Deng et al. CVPR 2021]: 23K shapes

3D datasets for RFMs

- 😀 High-quality
- 😀 Affordance
- 😀 Structured
- 😀 Text-grounded

Active learning [Yu et al. ICCV 2023]

Zero-shot, open-vocabulary, arbitrary referring segmentation in image, then back-project to 3D [Wang et al. 2025]

3D datasets for RFMs

- 😀 High-quality
- 😀 Affordance
- 😀 Structured
- 😀 Text-grounded
- 😀 Motion-enabled

Robots need to recognize and interact with object parts to expect proper responses (motions) and fulfill tasks. But such RFMs are still far away!

~2,300 models from PartNet-Mobility [Xiang et al. CVPR 2020]