AI Graphics

Machine Learning for Media Applications

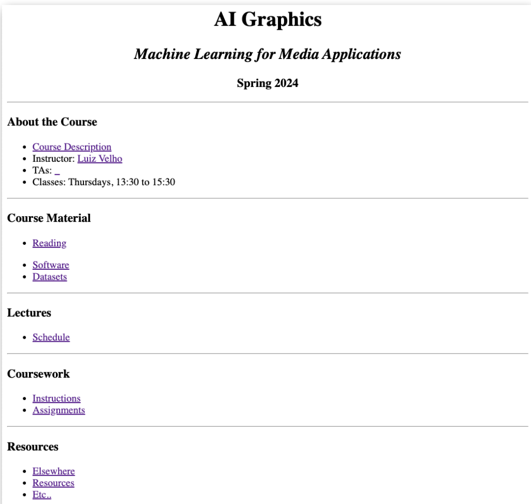
Luiz Velho
IMPA

Overview

- Course Overview
- 3D Graphics and AI
- Basic Concepts for Modeling and Rendering
- Machine Learning Fundamentals
- 3D Deep Learning

The Course

Course Page



AI Graphics

Machine Learning for Media Applications

Spring 2024

About the Course

- Course Description
- Instructor: Luiz Velho
- TAs: _
- Classes: Thursdays, 13:30 to 15:30

Course Material

- Reading
- Software
- Datasets

Lectures

- Schedule

Coursework

- Instructions
- Assignments

Resources

- Elsewhere
- Resources
- Etc..

https://lvelho.github.io/ai-graphics-2024/

Course Affairs

Administrivia

- Classes
- Reading
- Projects
- Grading

3D Graphics & AI

Scope

3D

What we will study

- 3D Graphics
- Synthetic Images

Areas of Application

- Scientific Visualization
  - Medicine, Biology, Mathematics
- CAD / CAM
  - Engineering, Architecture, Design
- Entertainment
  - Television, Film, Games

Other Areas

- Vision

- 3D Animation
- Virtual Reality
- Multimedia

2D

What we will not cover

- 2D Graphics

- Human-Computer Interaction

Areas of Applications

- Electronic Publishing
  - Illustration
  - Paint Systems
  - Page Layout
- Image Processing
  - Video Effects
  - Pattern Recognition
- Interface Design
  - Window Systems

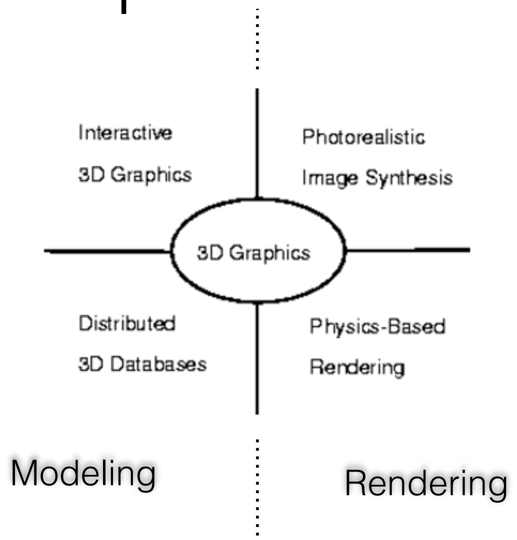
Subject

3D Graphics Systems / Media / Artificial Intelligence

- Theory

- Practice

3D Graphics Standards

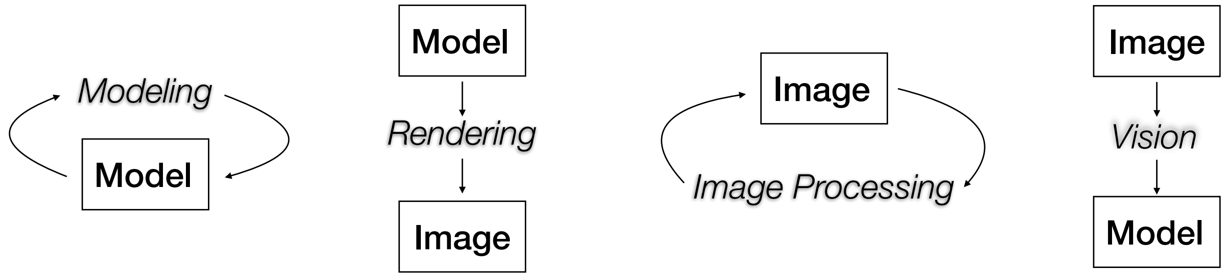


TRENDS: Integration (VR / AR / Games)

Evolution of the Area

Traditional *Graphics*

- Four Separated Areas



Modern *Graphics*

- Integrated Areas

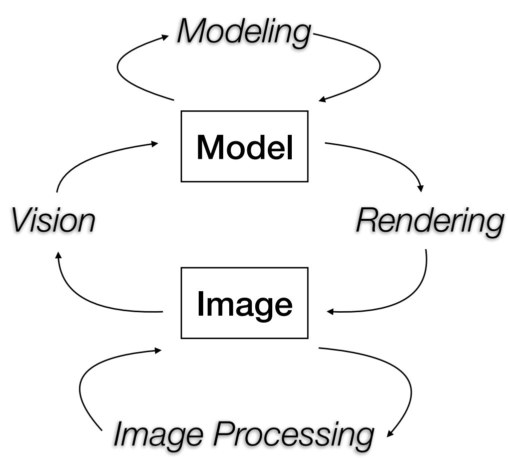
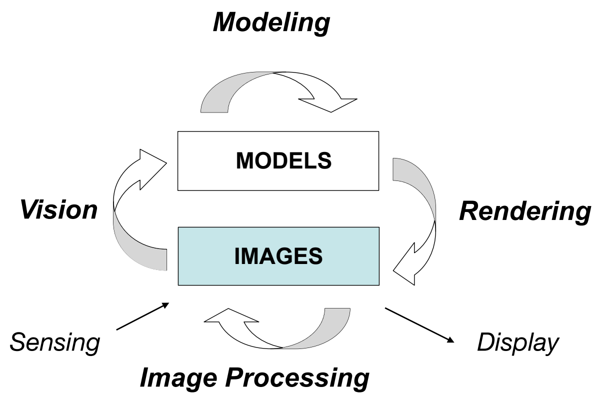


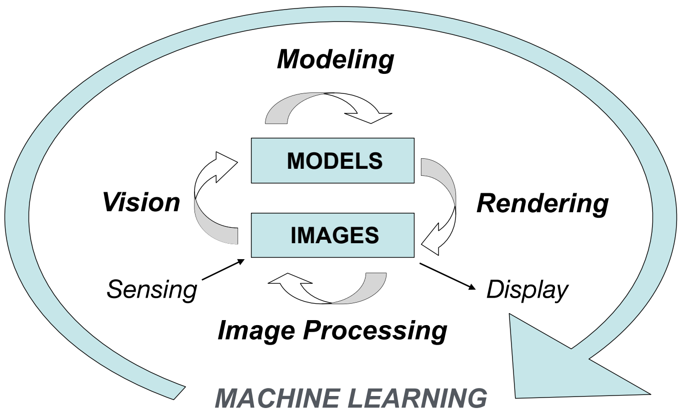
Image-Based *Graphics*

- From Appearance to Algorithms



A.I. *Graphics*

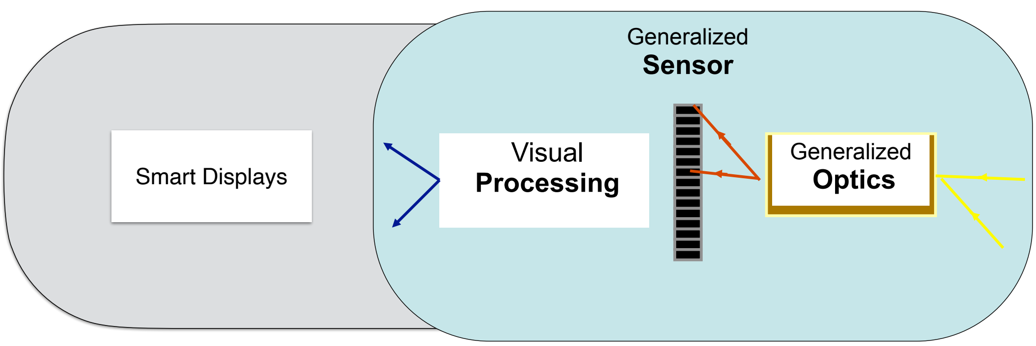
- From Data to Networks

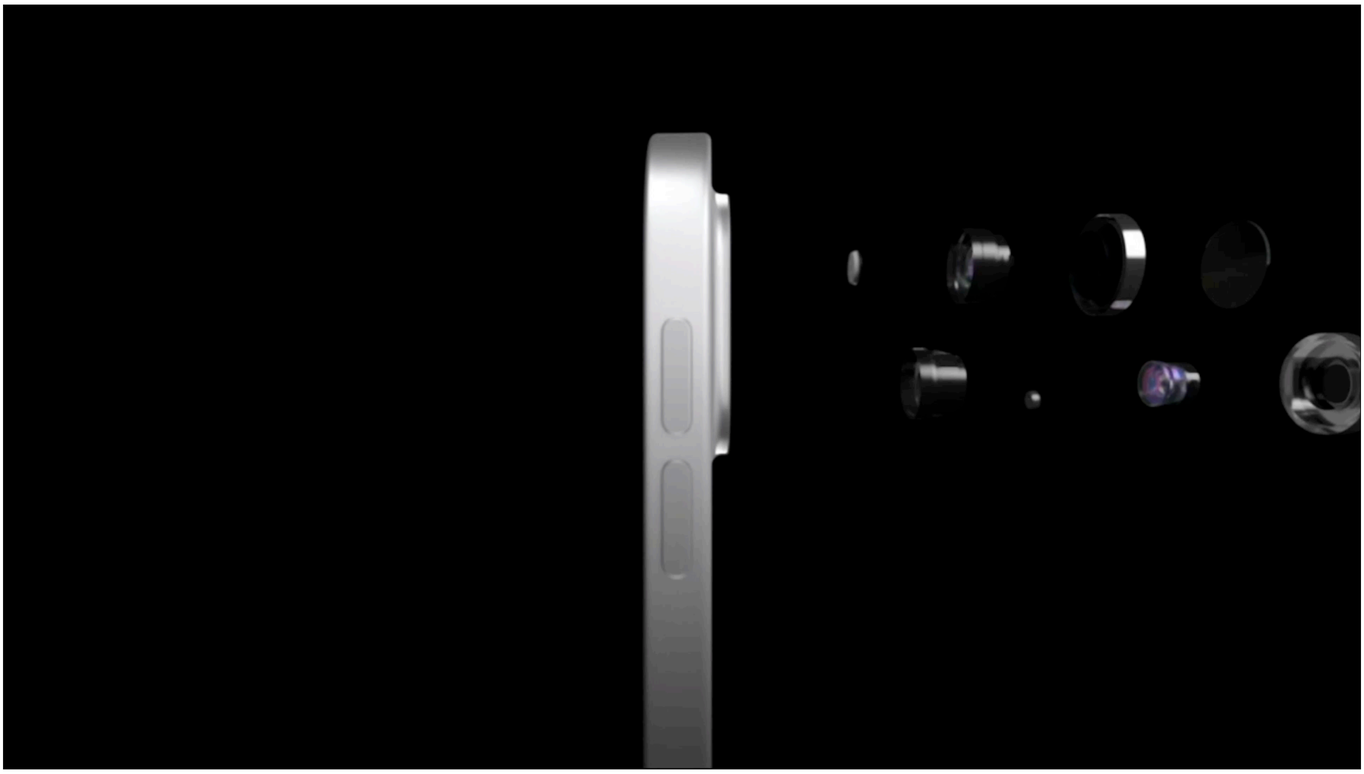


Media Technology

Computational Imaging

- Intelligent Acquisition / Display Devices





Case Study

GANCraft: Rendering Realistic Worlds

GANCRAFT

Unsupervised 3D Neural Rendering of Minecraft Worlds

Zekun Hao*+, Arun Mallya*, Serge Belongie+, Ming-Yu Liu*
NVIDIA*, Cornell University+

Pause for a Break

"At least this supports my theory that we live in a computer simulation."

Case Study

Physics-Based Realistic Avatars

(for the Metaverse)

Basic Concepts for 3D Graphics Systems

Concepts

- Paradigm of the 4 Universes
- Media Objects

4 Universes Paradigm

- Levels of Abstraction

Physical Universe

↕

Mathematical Universe

↕

Representation Universe

↕

Implementation Universe

(Gomes and Velho, 1995)

Example: Numbers

Scalar Quantities	length, area
Real Numbers	7.598
Floating Point Representation	$7598 * 10^{-3}$
IEEE Standard	mantissa exponent

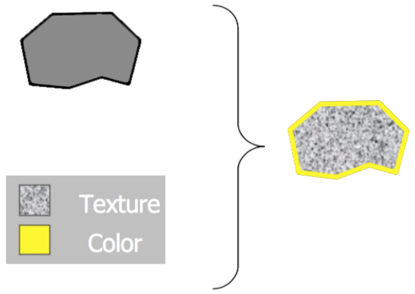
Issues

- Study each Level
  - Objects in level
- Relation between Levels
  - How to Map (e.g. Correspondences)
  - Equivalences and Losses

Media Objects

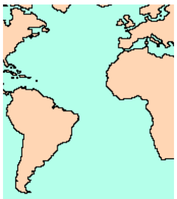
$$o = (U, f) \qquad\qquad f : U \subset \mathbb{R}^m \to \mathbb{R}^n$$

- Support ***U***
  - Shape / Geometry
- Properties ***f***
  - Attributes / Appearance



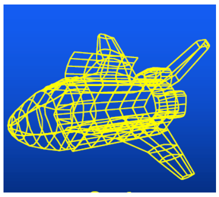
Unifying Concept

Examples

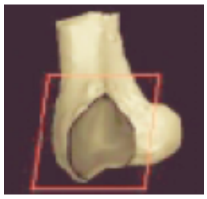


drawings

images



surfaces



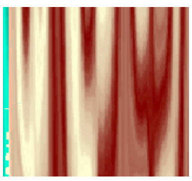
volumes

time

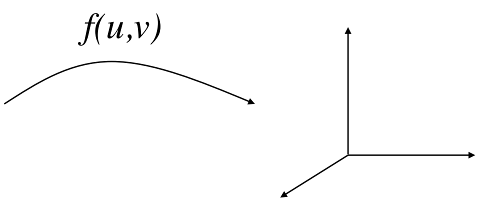
- General Encompassing Definition (Gomes and Velho, 1996)

Defining Attributes

- Ex: Image
  - Simple Shape: $U = [0,1] \times [0,1]$
  - Complex Attributes: (color, density, etc) $f(u, v) = [r, g, b, a, \ldots]$



U



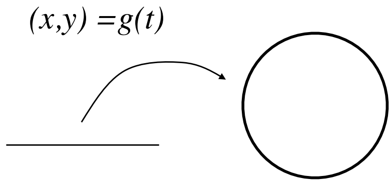
C

Defining Shape

- Ex: Curves
  (unit circle)
  - Parametric Description

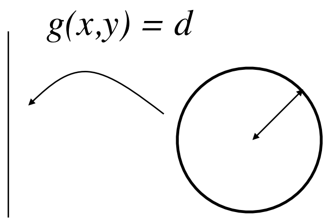
    $g(\theta) = (\cos\theta, \sin\theta)$

    $\theta \in [0, 2\pi]$



  - Implicit Description

    $g^{-1}(0) = \{(x, y) \,|\, x^2 + y^2 - 1 = 0\}$



Putting it All Together

- Ex: 3d Object
  (Diana statue)
  - Shape:
    - Scanned 3D Surface
    - Solid Object
  - Properties:
    - Material Marble

(Wann Jensen et al.)

Questions

- How to Define:
  - Function (model)
  - Support (domain)
  - Attribute (range)

Mathematical Models

- Deterministic
  - Function (one object)
- Probabilistic
  - Stochastic Process (class of objects)
- Procedural
  - Generators + Operators (algebra / expression)

Computation and Representation

- How to Compute
- Conversion between Models / Representations

✳ Hybrid Models

Media Objects Operators

- Spaces of Media Objects

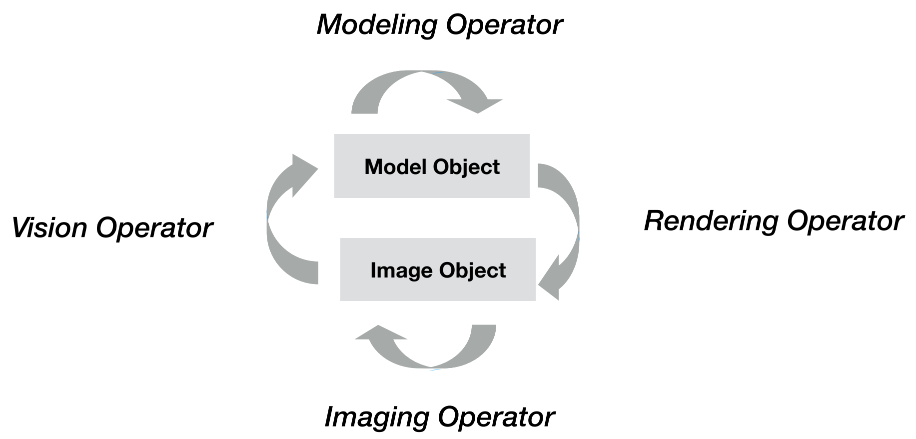
$$o \in \mathcal{O}$$

- Media Objects Operators

$$T : \mathcal{O}_1 \mapsto \mathcal{O}_2$$

- Intra $(i = j)$ / Inter Levels $(i \neq j)$
- Representation

Landscape of Operations



Problems

Direct Problems

- Given T and x, find y

$$y = Tx$$

- Ex: Visualization
  - x is the scene (geometry, lighting, camera)
  - T is the rendering operator
  - y is the rendered image

Inverse Problems

- Given T and y, find x such that $T(x) = y$

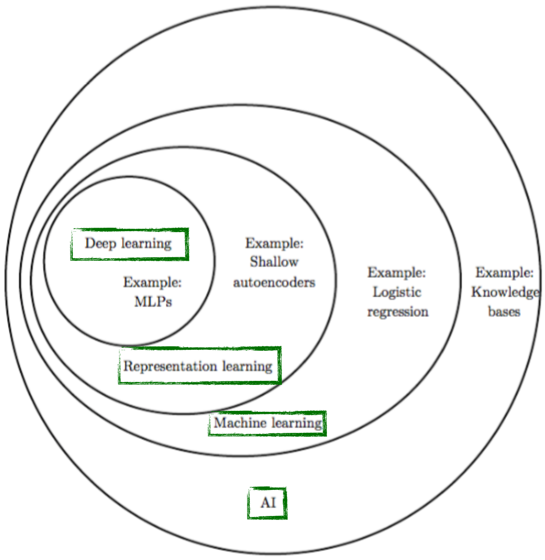
  Ex: Object Recognition
  - y is a captured image
  - T is an acquisition system
  - x is a template

- Given x and y, find T

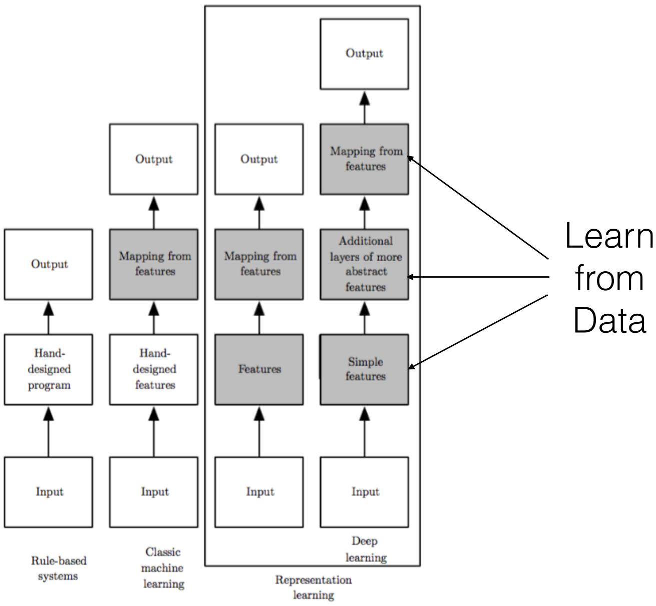
  Ex: Camera Calibration
  - x is a calibration pattern
  - y is a transformed pattern
  - T is a projective transformation

Machine Learning Fundamentals

Contextualization



Historical Evolution



Representation Learning

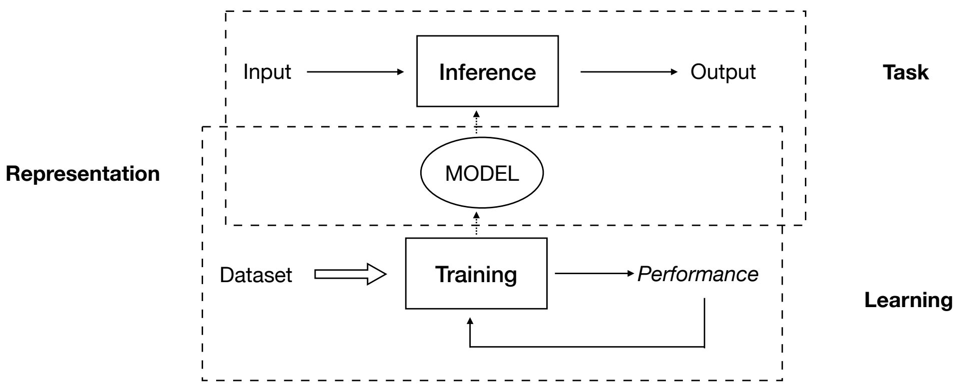
- Set of Methods
  - Allows a Machine
  - to Automatically Discover Representation (Model)
  - for a Task (Analysis / Synthesis)

Learning Algorithms

Machine Learns from Data

- Task (based on a model)
  - Analysis / Synthesis
- Performance (measure of success)
  - Training / Inference
- Experience (from a dataset)
  - Supervised / Unsupervised / Reinforcement

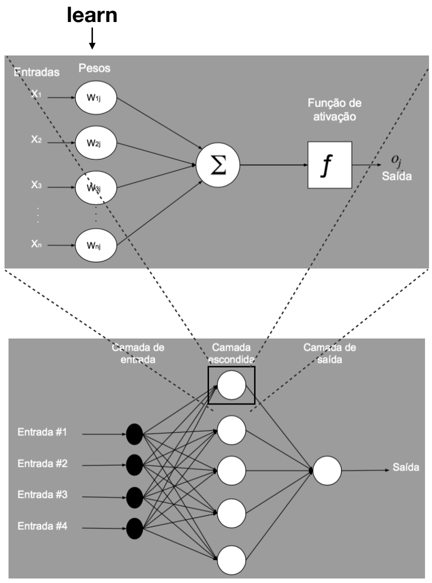
Conceptual Scenario



Deep Learning Methods

- Representation Learning
  - Neural Network Model
- Multiple Levels of Representation
  - Features
- Composition
  - Simple, but Non-Linear Operators

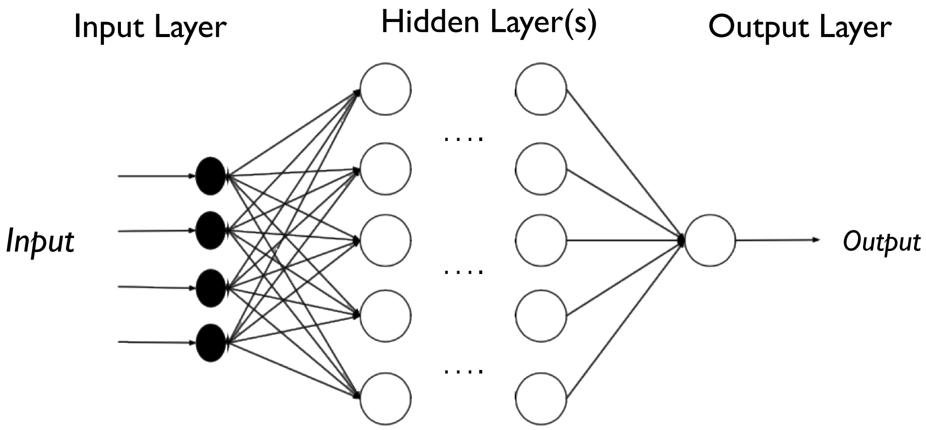
Mathematical Framework



- Perceptron (operator)
- Neural Nets (composition)

Neural Network

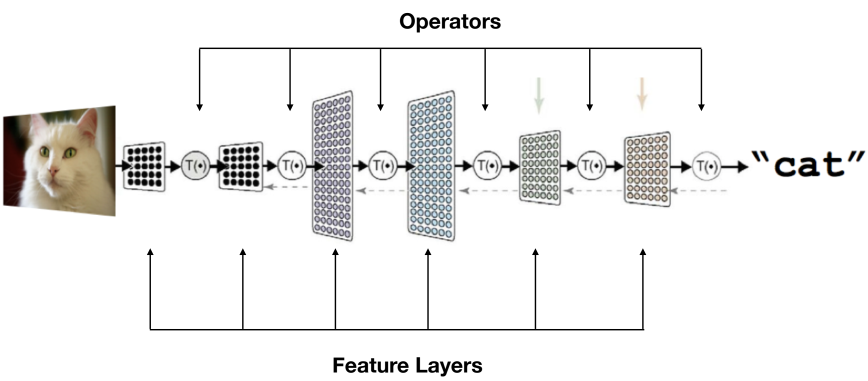
- Multiple Interconnected Layers



OBS: Computational Graph

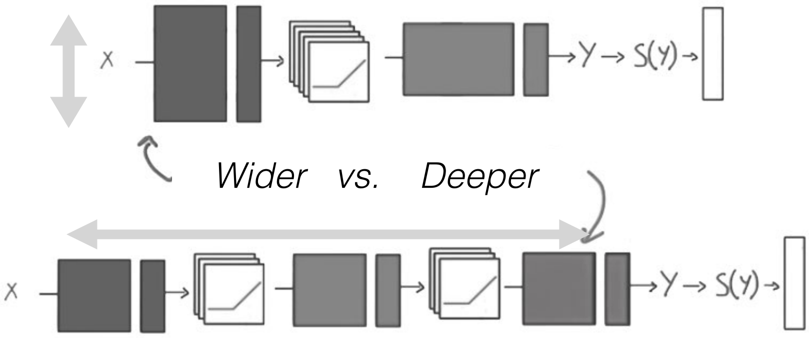
Deep Neural Networks

- Multi Layer Architecture



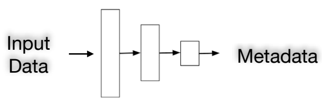
Network Size

- Number of Nodes / Number of Layers

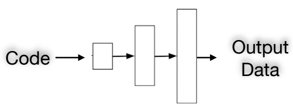


Types of Neural Networks

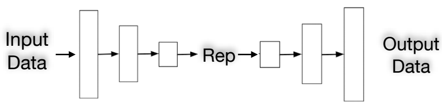
- Analysis



- Synthesis



- Hybrid

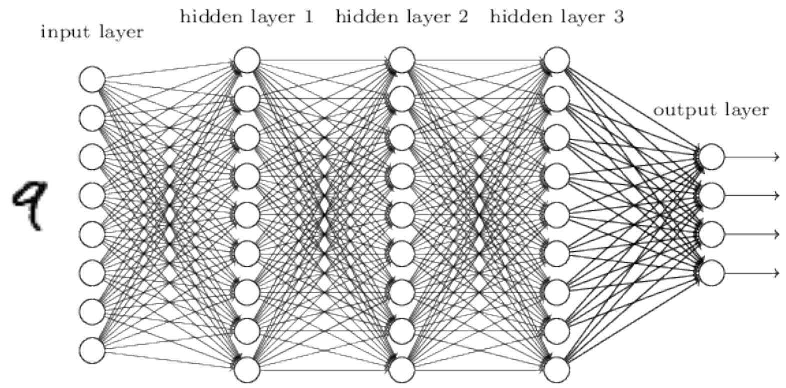


Network Architectures

- Fully Connected
- Convolutional
- Auto-Encoder
- Generative / Adversarial
- Recurrent
- etc…

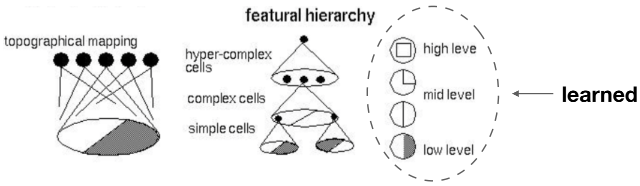
Fully Connected MLP

- Learn General Functions

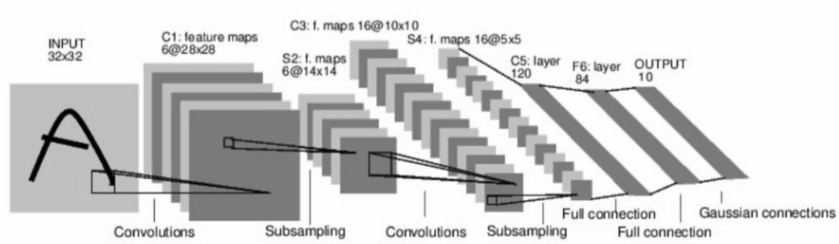


Convolutional Networks

- Convolution Operators

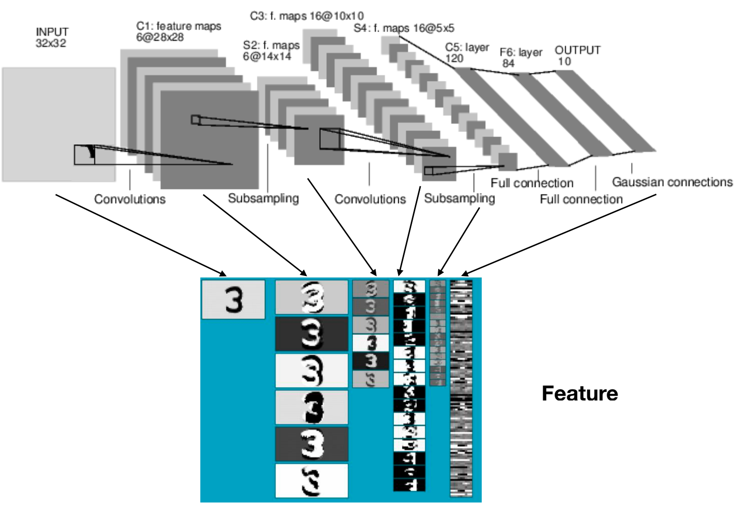


- Deep Convolutional Neural Networks



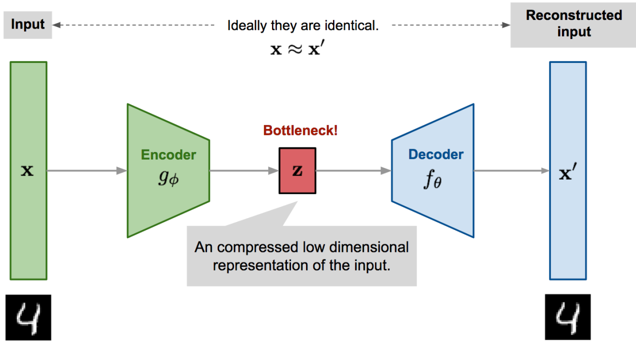
CNN in Action

- Learn Image Features



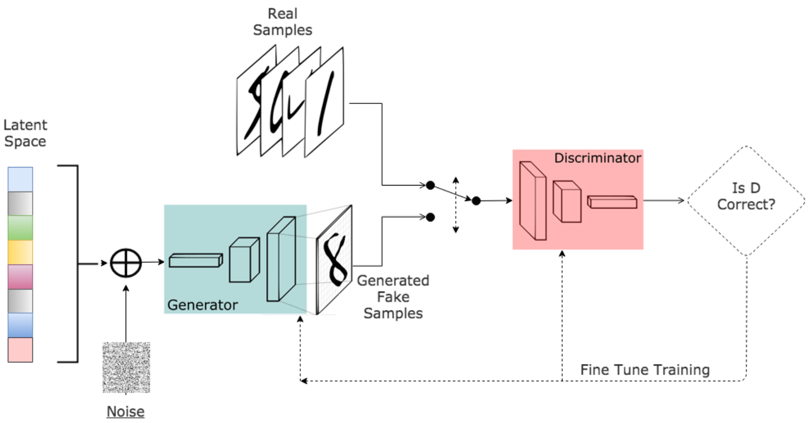
Variational Auto-Encoders

- Learn <u>Representations</u>



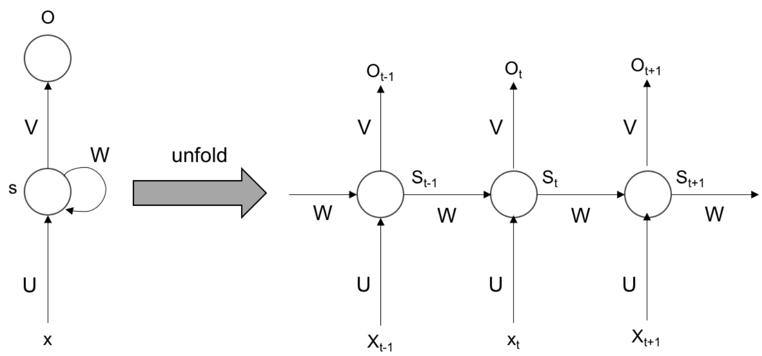
Generative Networks

- Learn to <u>Reproduce from a Distribution</u>



Recurrent Networks

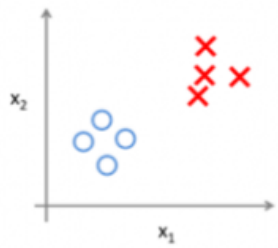
- Learn a <u>Time Dependent Process</u>



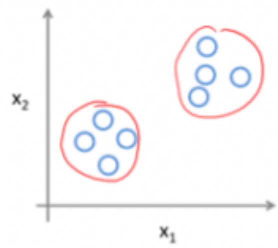
Learning

Learning Methods

Supervised **Unsupervised**



Labeled Data



Unlabeled Data

OBS: Semi-Supervised

Comparison

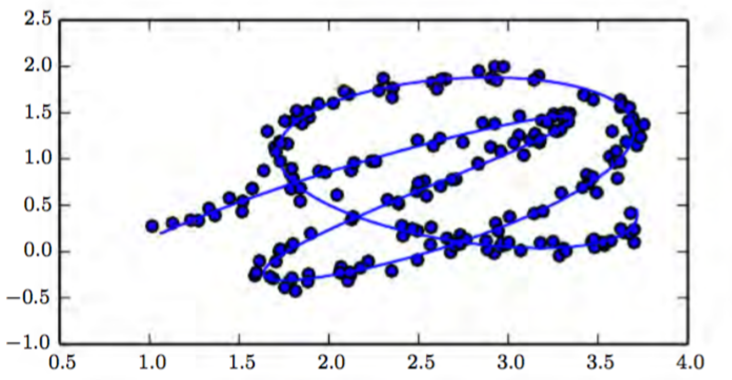
Approaches to learning features

- Supervised Learning
  - End-to-end learning of deep architectures (e.g., deep neural networks) with back-propagation
  - Works well when the amounts of labels is large
  - Structure of the model is important (e.g. convolutional structure)

- Unsupervised Learning
  - Learn statistical structure or dependencies of the data from unlabeled data
  - Layer-wise training
  - Useful when the amount of labels is not large

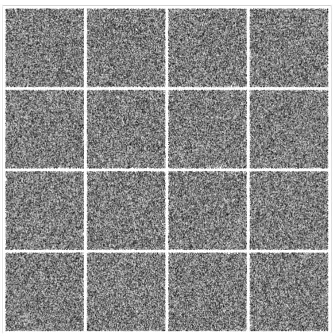
Manifold Learning

- Embedded Model Subspace

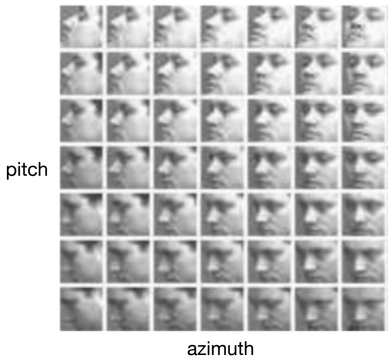


Model Parametrization

- Discover and Disentangle Manifold Coordinates

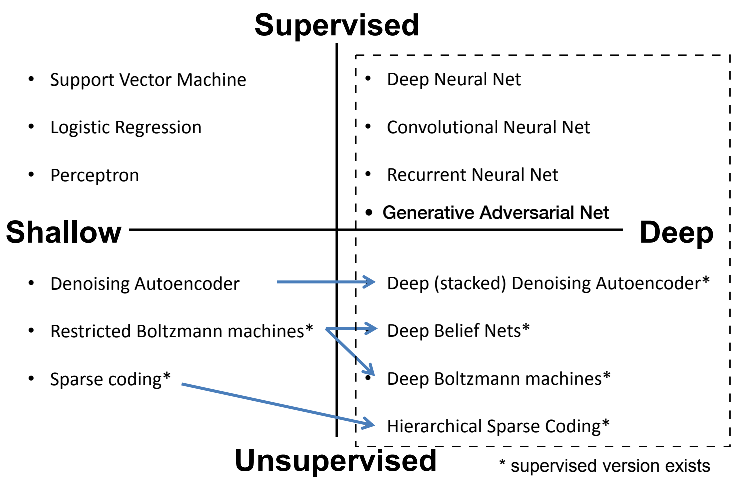


Random Images



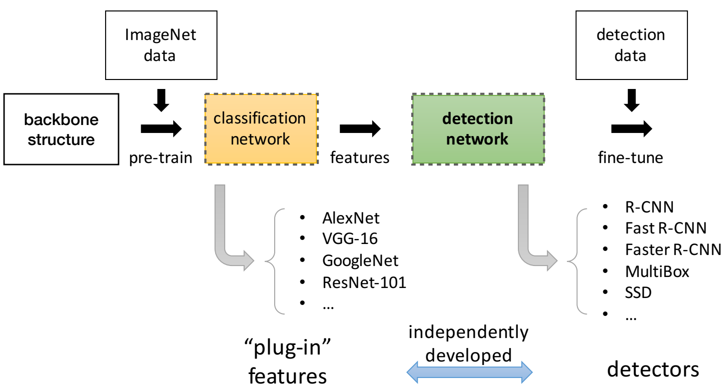
Multiview Face Dataset

Taxonomy



Combining Networks

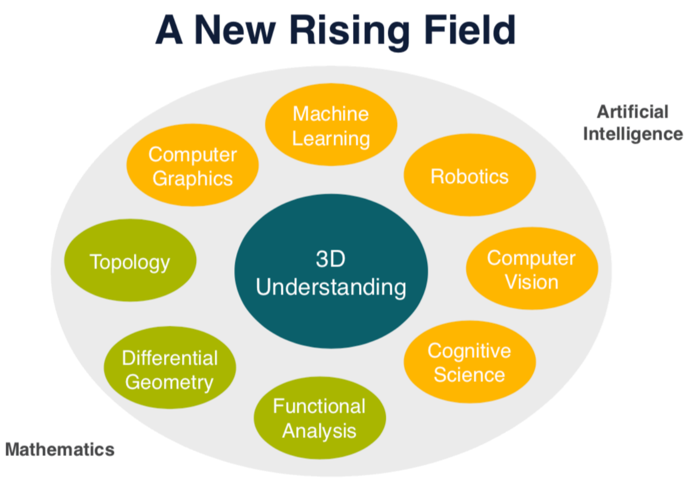
- Example: Classification / Detection



3D Deep Learning

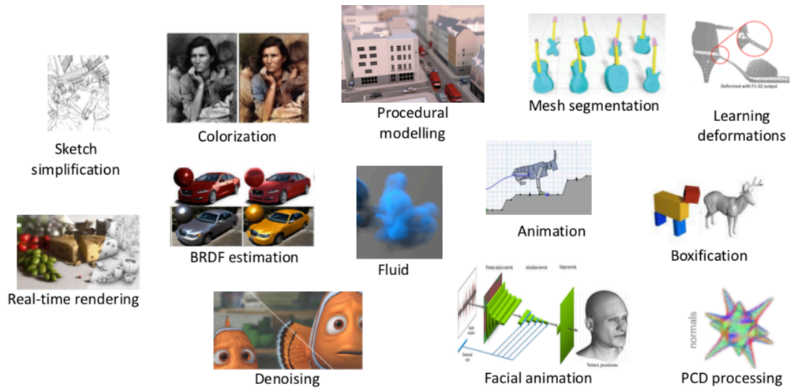
Slides from: (Mitra et. al, "CreativeAI: Deep Learning for Computer Graphics", 2019)

Deep Learning for 3D Data



(Mitra et. al, 2019)

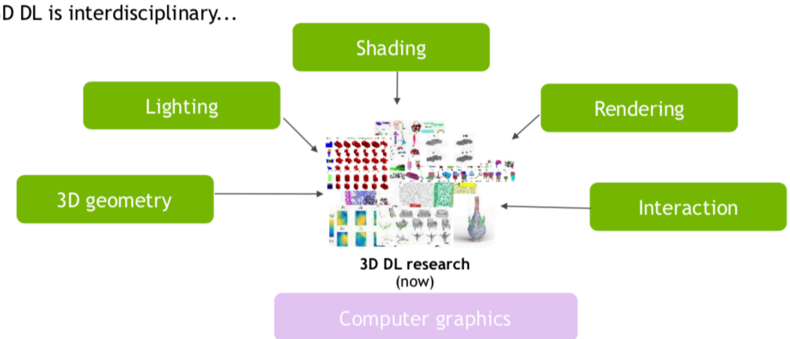
Applications



(Mitra et. al, 2019)

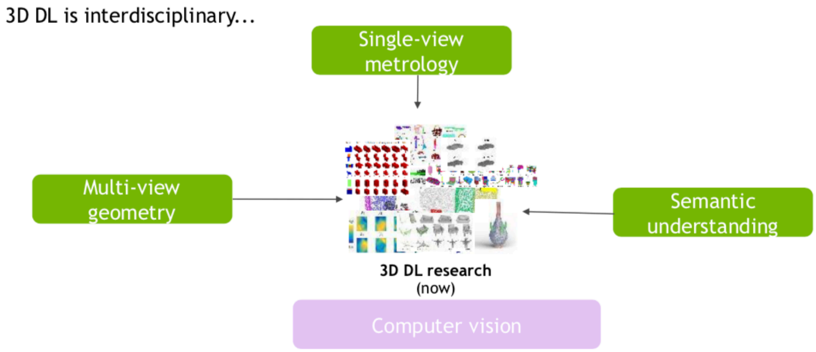
Computer Graphics

3D DL is interdisciplinary…



(Mitra et. al, 2019)

Computer Vision



(Mitra et. al, 2019)

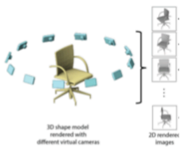
Synergistic Opportunities

What is Special about CG?

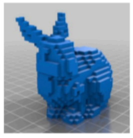
1. **Regular data structure** and easy to parallelize (e.g., image translation)
2. Many sources of input data — **model building** (e.g., images, scanners, motion capture)
3. Many sources of **synthetic data** — can serve as supervision data (e.g., rendering, animation)
4. Many problems in **generative models** and need for **user-control**

(Mitra et. al, 2019)

Representation Challenge



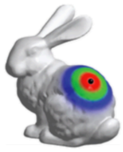
Multi-view



Volumetric

Part Assembly

Point Cloud



Mesh (Graph CNN)

$$F(x) = 0$$

Implicit Shape

(Mitra et. al, 2019)

Representations in CG

- Images (e.g., pixel grid)
- Volume (e.g., voxel grid)
- Meshes (e.g., vertices/edges/faces)
- Point clouds (e.g., collection of points)
- Animation (e.g., skeletal positions over time; cloth dynamics over time)

(Mitra et. al, 2019)

Problems in GC

- Feature detection (image features, point features) $\mathbb{R}^{m \times m} \rightarrow \mathbb{Z}$
- Denoising, Smoothing, etc. $\mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$
- Embedding, Metric learning $\mathbb{R}^{m \times m, m \times m} \rightarrow \mathbb{R}^{d}$

analysis

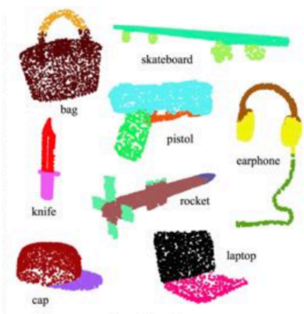
- Rendering $\mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$
- Animation $\mathbb{R}^{3m \times t} \rightarrow \mathbb{R}^{3m}$
- Physical simulation $\mathbb{R}^{3m \times t} \rightarrow \mathbb{R}^{3m}$
- Generative models $\mathbb{R}^{d} \rightarrow \mathbb{R}^{m \times m}$

synthesis

(Mitra et. al, 2019)

3D Geometry Analysis

Classification



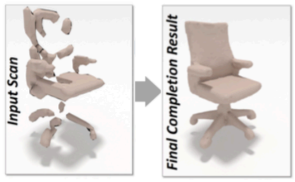
Parsing (object/scene)

Correspondence

(Mitra et. al, 2019)

3D Synthesis

Monocular
3D reconstruction



Shape completion

Shape modeling

(Mitra et. al, 2019)

3D Deep Learning

Goal: Learn a Parametric Function

$$f_\theta : \mathbb{X} \longrightarrow \mathbb{Y}$$

θ: function parameters, these are learned

$\mathbb{X}$: source domain

$\mathbb{Y}$: target domain

Examples:

Image Classification: $f_\theta : \mathbb{R}^{w \times h \times c} \longrightarrow \{0, 1, \ldots, k-1\}$

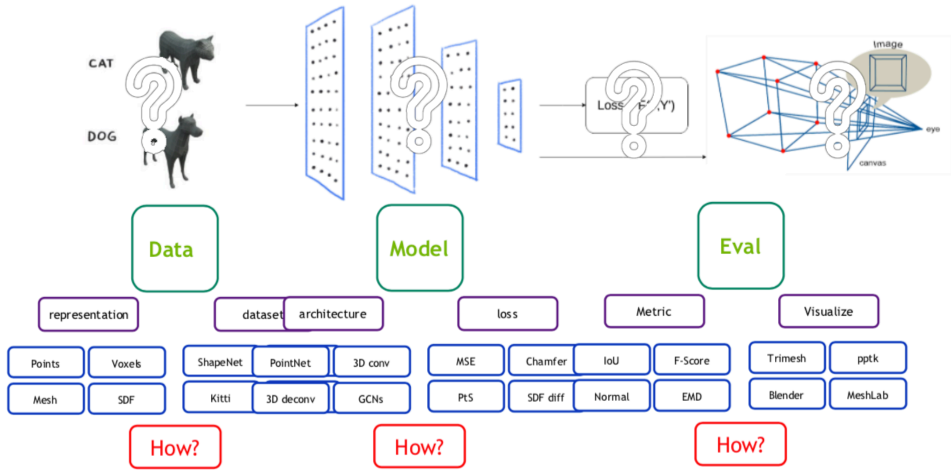
$w \times h \times c$: image dimensions k: class count

Image Synthesis: $f_\theta : \mathbb{R}^{n} \longrightarrow \mathbb{R}^{w \times h \times c}$

n : latent variable count $w \times h \times c$: image dimensions

(Mitra et. al, 2019)

3D DL Research



(Mitra et. al, 2019)

Main Challenges

1. **Representation:** How is the data organised and structured?
2. **Training data:** Is it synthetic or real, or mixed?
3. **User control:** End-to-end or in small steps?
4. **Loss functions:** Hand-crafted or learned from data?

(Mitra et. al, 2019)

Data is the New Currency

- Synthetic data
  - Generative model + photo-realistic rendering
  - Object geometry + physical simulation
  - Object geometry + synthetic materials + realistic simulations

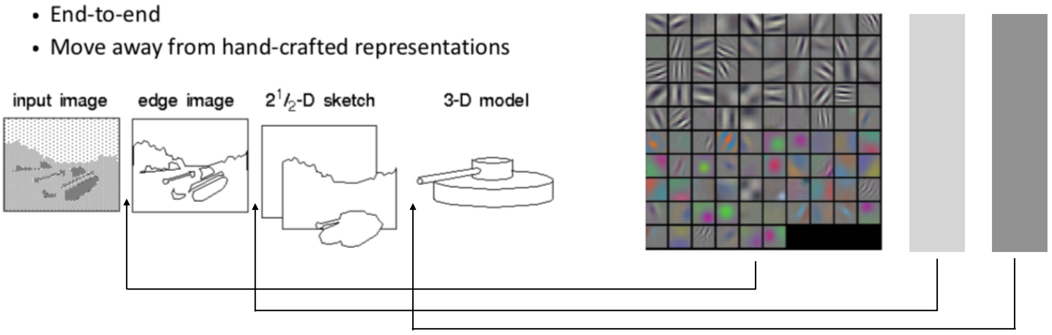
- Real data
  - Collected from images, scans, mocap sessions
  - Collected using specialized equipments (e.g., light-field, pressure gloves)

(Mitra et. al, 2019)

Learning Features

End-to-end: Learned Features

- *Before*
  - Handcrafted feature extraction, e.g., edges or corners (hand-crafted)
  - Mostly with linear models (PCA)
- *Now*
  - End-to-end
  - Move away from hand-crafted representations



(Mitra et. al, 2019)

Metrics

End-to-end: Learned Loss

- *Before*
  - Evaluation came after
  - It was a bit optional
    - You might still have a good algorithm without a good way of quantifying it
    - Evaluation helped publishing
- *Now*
  - It is essential and build-in
  - If the loss is not good, the result is not good
  - (Extensive) Evaluation happens automatically

- While still much is left to do, this makes graphics much more reproducible

(Mitra et. al, 2019)

A.I. Graphics