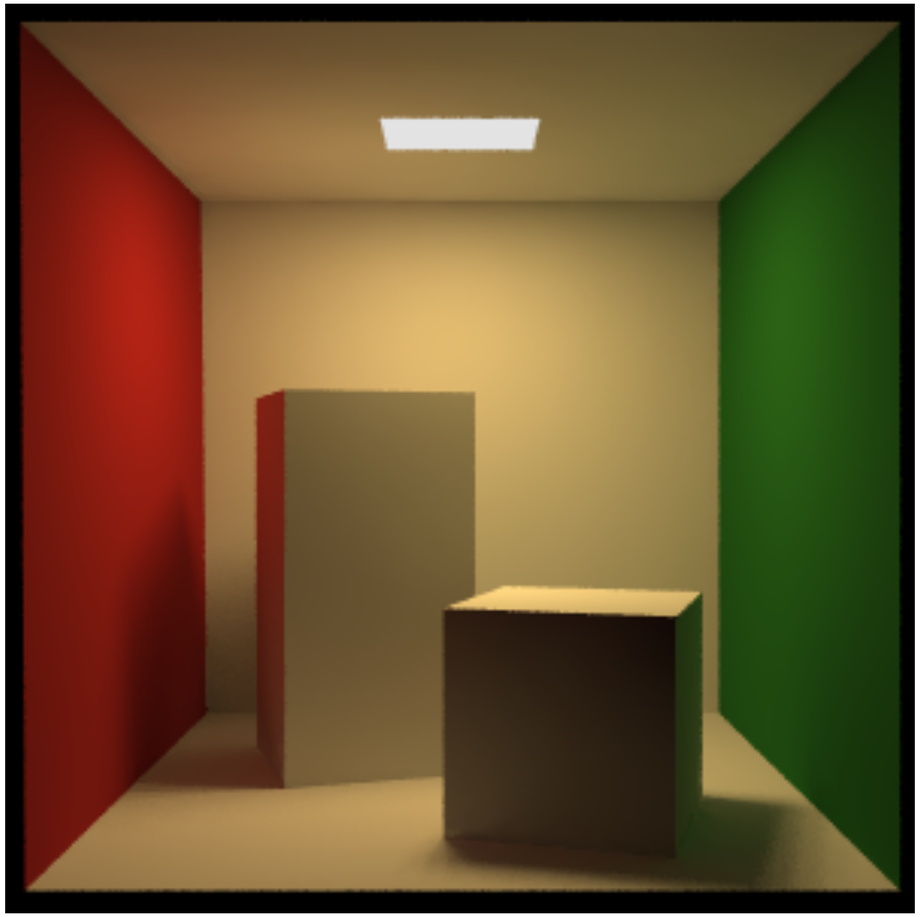
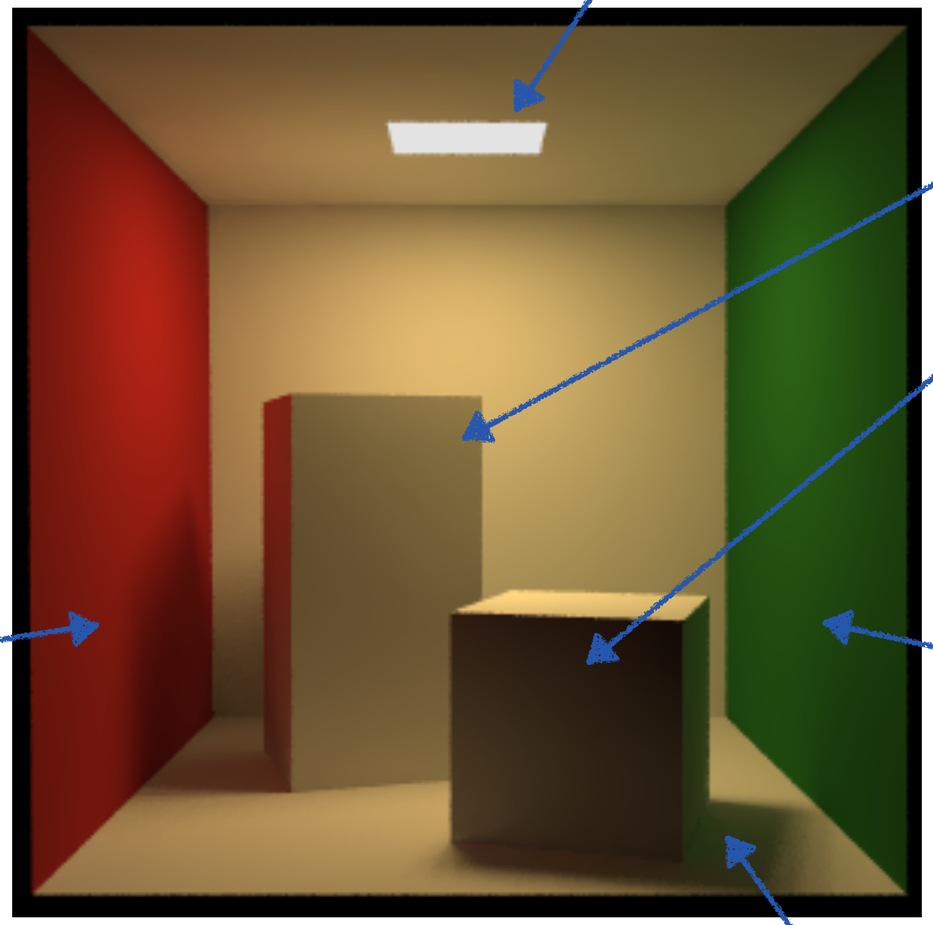
Image Representation with CNN

In comparison with the theories in Vision(1982) by David Marr

Weihao Qiu



Cornell box


Light
Further
Closer
Red Texture
Green Texture
Shadow

“Vision is the ***process*** of discovering from images ***what*** is present in the world, and ***where*** it is. Vision is therefore … **an information-processing task.**”

Vision(1982), David Marr

Vision - An Information-processing task

Three levels of the process

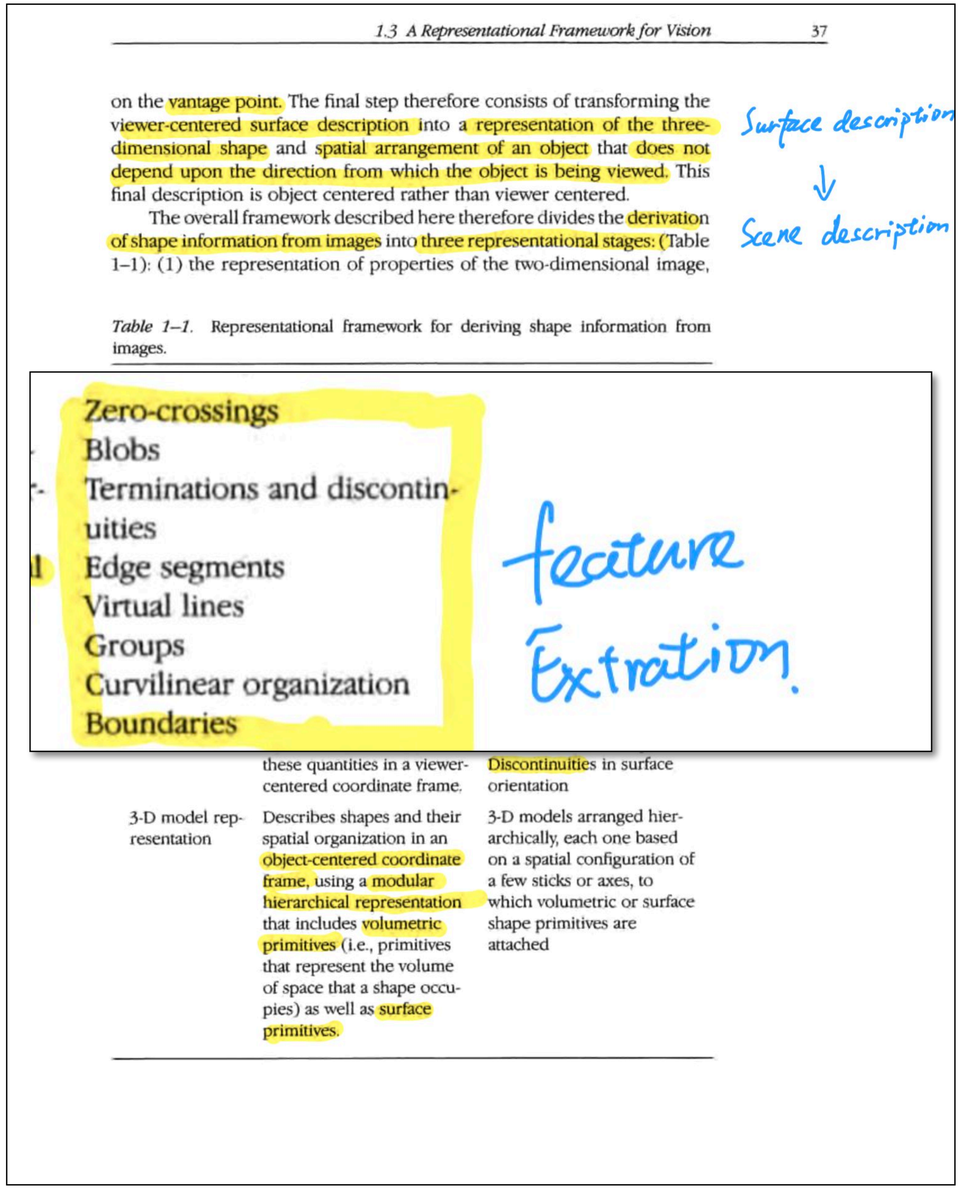
Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1–4. The three levels at which any machine carrying out an information-processing task must be understood.

Representational Framework

Analogies

- *Images* -> Pixel-value
- *Primal Sketch* -> Low-level features
- *2 1/2 -D sketch* -> High-level features
- *3D model representation* -> Representation



1.3 A Representational Framework for Vision 37

on the vantage point. The final step therefore consists of transforming the viewer-centered surface description into a representation of the three-dimensional shape and spatial arrangement of an object that does not depend upon the direction from which the object is being viewed. This final description is object centered rather than viewer centered.

The overall framework described here therefore divides the derivation of shape information from images into three representational stages: (Table 1–1): (1) the representation of properties of the two-dimensional image,

Surface description
↓
Scene description

Table 1–1. Representational framework for deriving shape information from images.

Zero-crossings
Blobs
Terminations and discontinuities
Edge segments
Virtual lines
Groups
Curvilinear organization
Boundaries

feature Extration.

	these quantities in a viewer-centered coordinate frame.	Discontinuities in surface orientation
3-D model representation	Describes shapes and their spatial organization in an object-centered coordinate frame, using a modular hierarchical representation that includes volumetric primitives (i.e., primitives that represent the volume of space that a shape occupies) as well as surface primitives.	3-D models arranged hierarchically, each one based on a spatial configuration of a few sticks or axes, to which volumetric or surface shape primitives are attached

Generating Image

Framework

Images → Representation → Image

"Apple"
"苹果"

Painting

Abstracting

Photo

Representation

Closer

Further

Light

Shadow

Red Texture

Green Texture

Geometry

Hand-coded Algorithms

3D rendering software

Rendered Image

Photo

Representation

Closer

Further

Light

Shadow

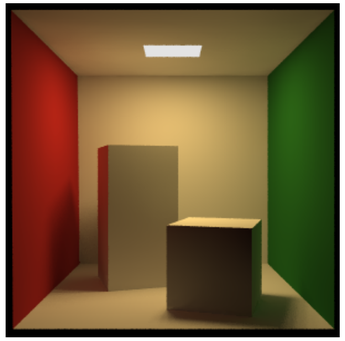
Red Texture

Green Texture

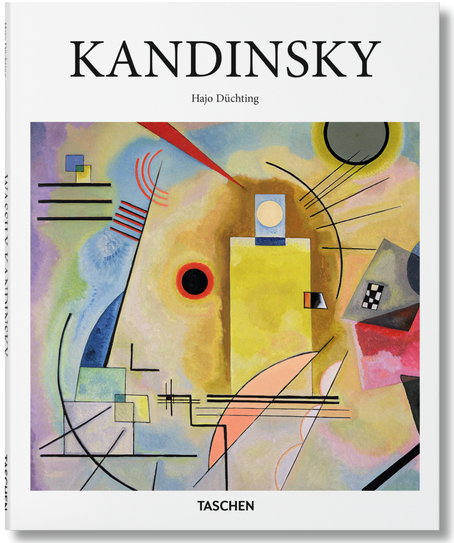
Geometry

Hand-coded Algorithms

3D rendering software



Rendered Image

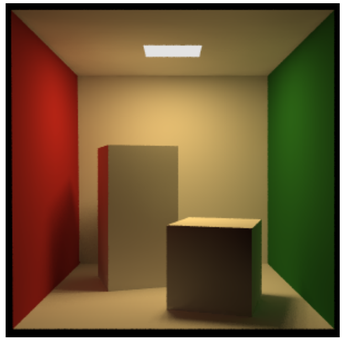


Photo

Representation

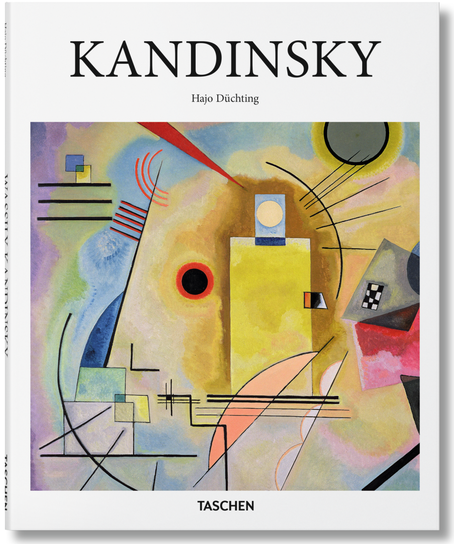
Hand-coded Algorithms

3D rendering software



Rendered Image

How to get the representation suitable for artistic purposes?

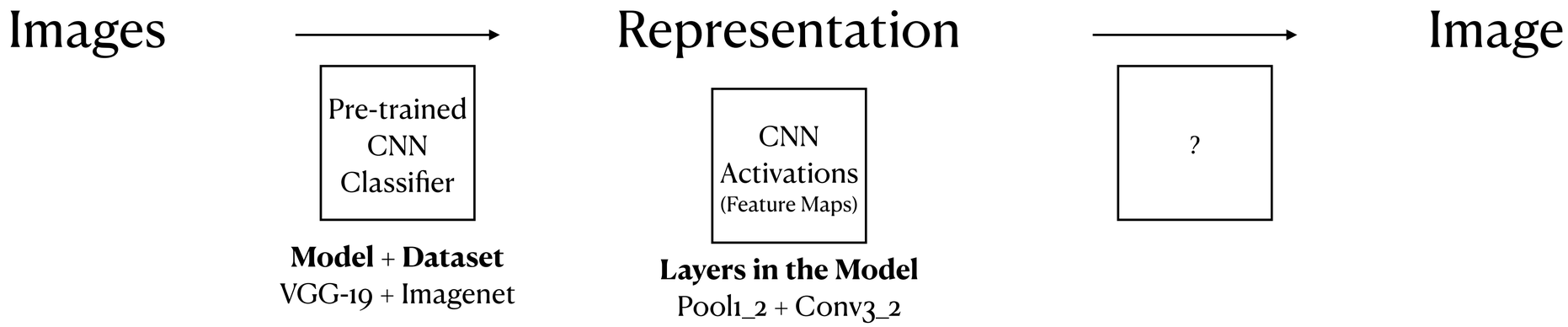


“Interestingly, the representations of CNNs trained to solve large-scale object recognition appear to be generally useful for visual information processing. They transfer to other datasets and tasks. It has become common practice in the computer vision community to use the **activations** of CNNs pre-trained on object recognition as the **ad-hoc feature representation** to solve **visual information processing tasks.**”

Texture and art with deep neural networks, Gatys et al.(2017)

Generating Image with CNN

Framework



Generating Image with CNN

Framework

Images → Pre-trained CNN Classifier

Model + **Dataset**
VGG-19 + Imagenet

→ Representation

CNN Activations (Feature Maps)

Layers in the Model
Pool1_2 + Conv3_2

→ Image

Hand-coded Algorithms

Strokes/Shapes
Visualizations

Omnia per Omnia, Sougwen Chung

Neural Nets

Model + **Loss function**
Deep dream
Texture Synthesis

Xavier Snelgrove

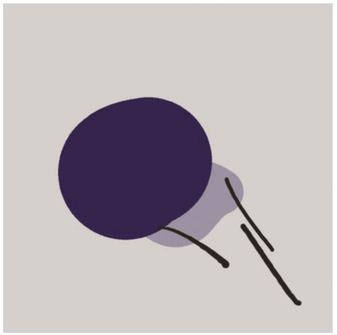
Generating Image with CNN

Framework

Images → Representation → Image

Pre-trained CNN Classifier

Model + **Dataset**
VGG-19 + Imagenet



Synthetic Abstraction, Tom White

CNN Activations (Feature Maps)

Layers in the Model
Pool1_2 + Conv3_2

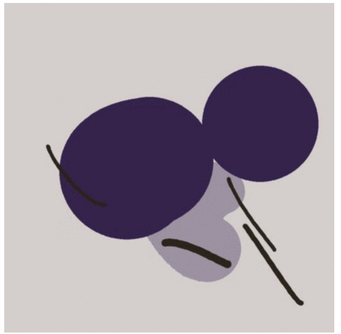
Score of the likeness to a Fan

Hand-coded Algorithms

Strokes/Shapes
Visualizations

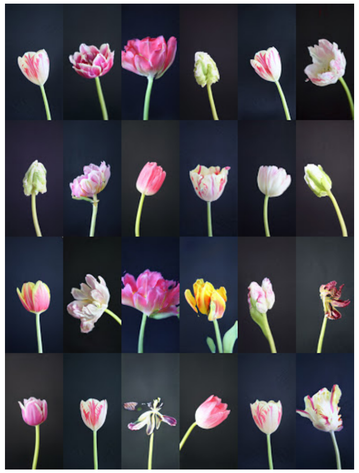
Use the customized drawing system the output images of only certain style

Try adding different strokes on top of the input image and pick the best one



Generating Image with CNN

Framework



Images → Pre-trained CNN Classifier → Representation → CNN Activations (Feature Maps) → Image

Model + Dataset

Layers in the Model

Age

Eyes

Traditional Model +
Customized/Synthetic Dataset

Interpretable Features
.....

Perspective

Mood

Thank you