History of object and scene representations

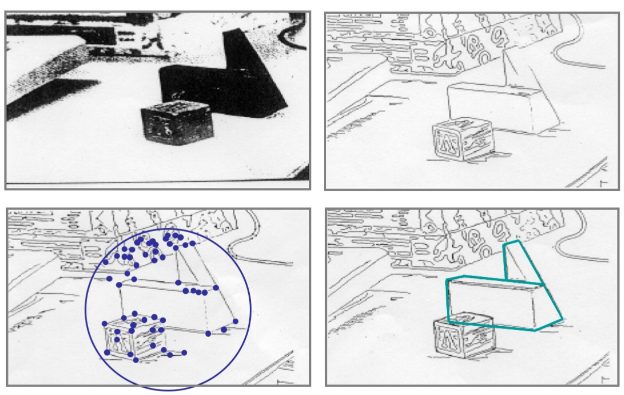
https://en.wikipedia.org/wiki/Landscape_with_the_Fall_of_Icarus

Outline

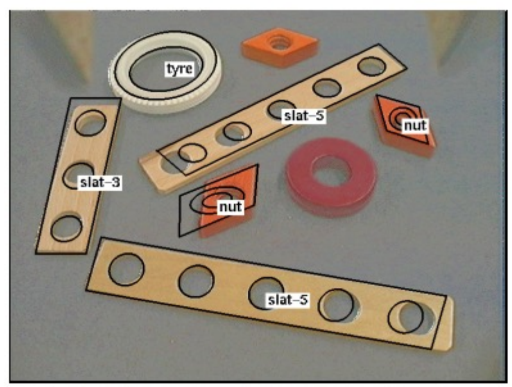
- Object representations
  - 3D shape
  - 3D primitives
  - 2D appearance-based models
  - 2D part-based models (deformable templates)
  - CNNs
- Scene representations
  - Structured representations
  - Appearance-based representations
  - Bottom-up and top-down perceptual organization
- Trends

3D object representations

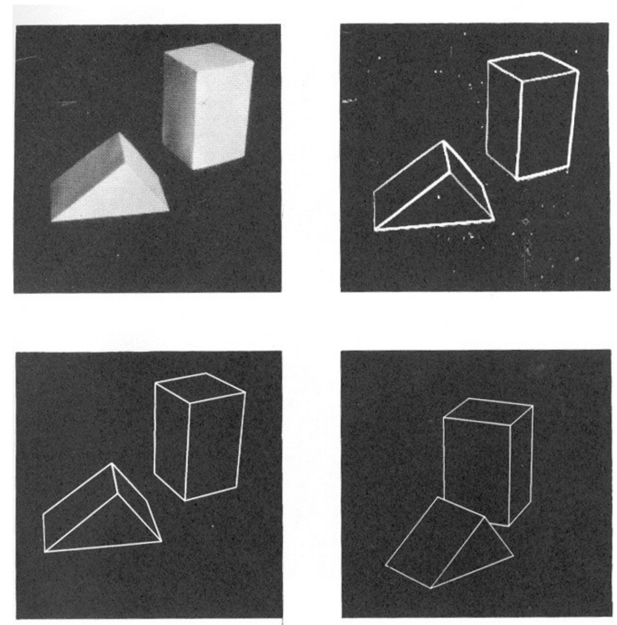
- Represent an object using its 3D model
- Recognition by *alignment* or *geometric invariants*



Alignment: Huttenlocher & Ullman (1987)

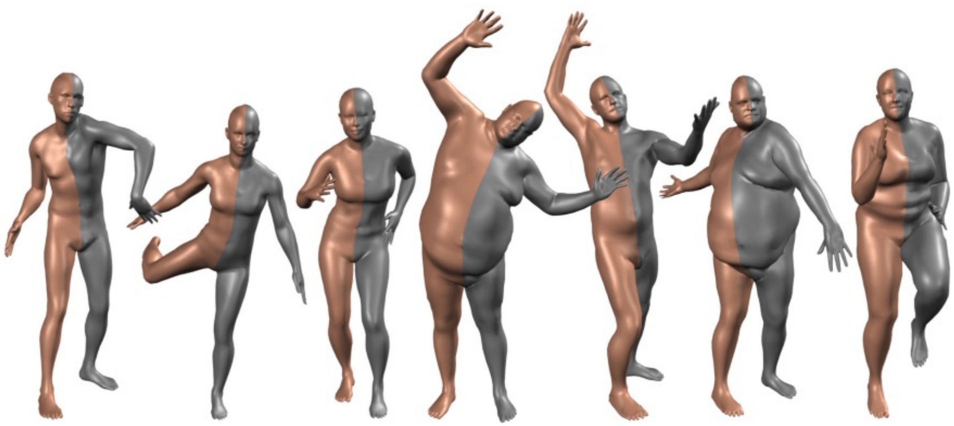


Invariants: Graf (2000)

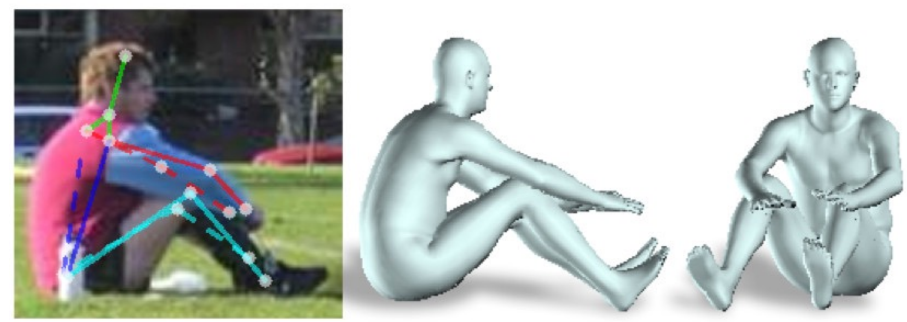


Roberts (1963)

Today: Category-specific 3D models

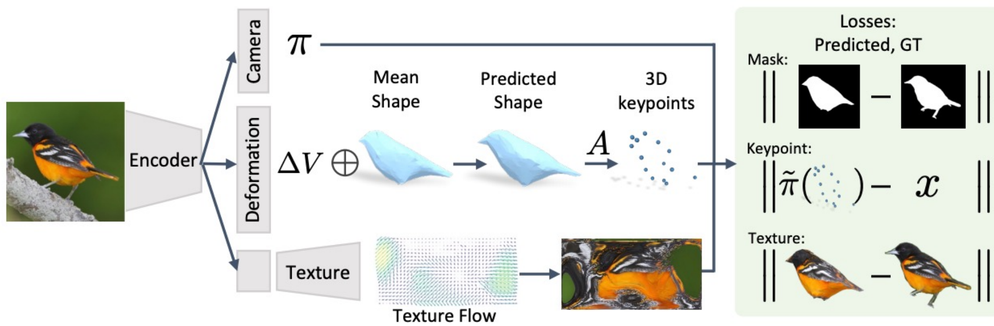


M. Loper et al. SMPL: A Skinned Multi-Person Linear Model. SIGGRAPH Asia, 2015



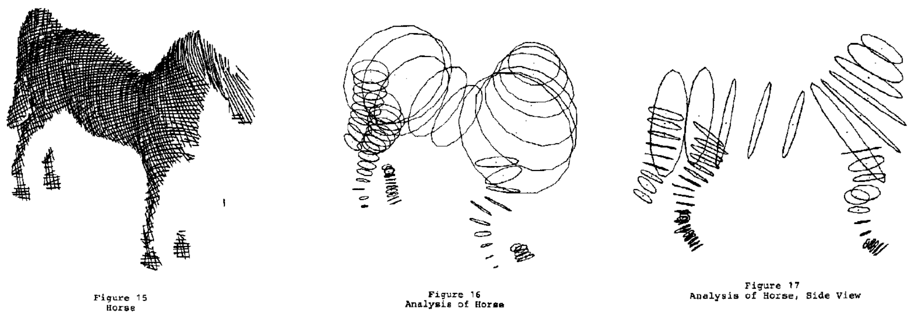
F. Bogo et al., Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image, ECCV 2016

Today: Generic category-level 3D models

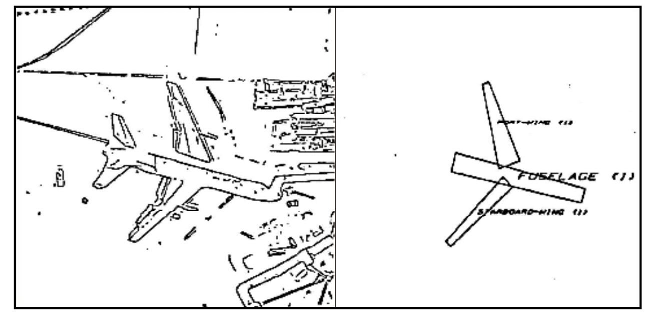


A. Kanazawa et al. [Learning Category-Specific Mesh Reconstruction from Image Collections](). ECCV 2018

3D shape primitives: Generalized cylinders

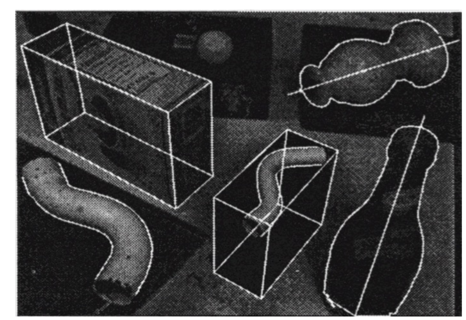


Binford (1971), Agin & Binford (1973)



Brooks (1981)

Zerroug & Nevatia (1994)



Zisserman et al. (1995)

Marr’s 3D object representation

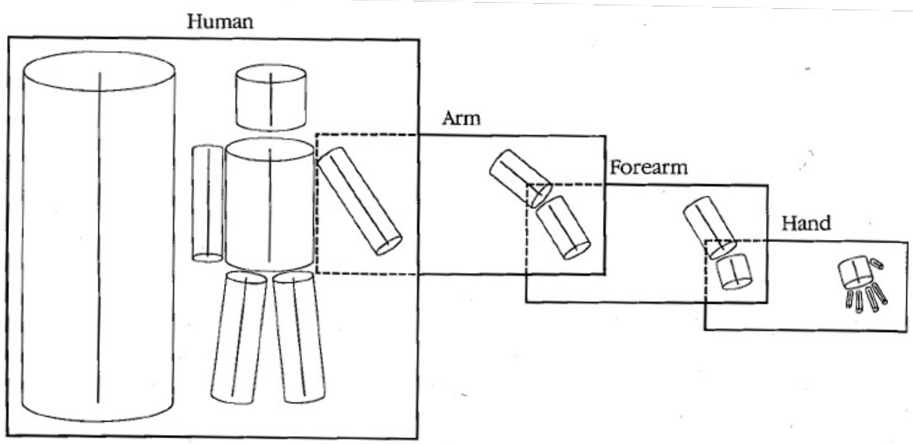
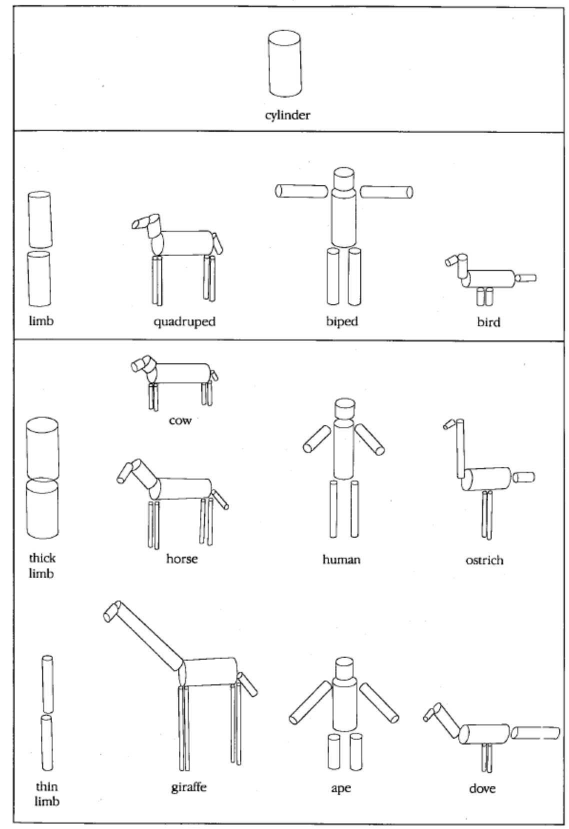


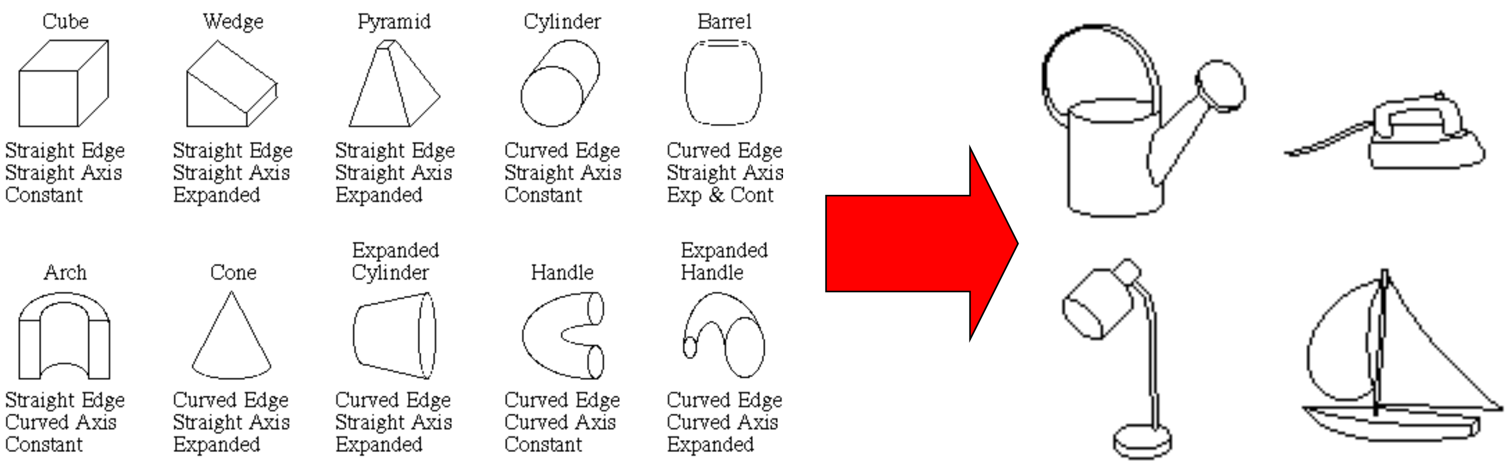
Figure 5–3. This diagram illustrates the organization of shape information in a 3-D model description. Each box corresponds to a 3-D model, with its model axis on the left side of the box and the arrangement of its component axes on the right. In addition, some component axes have 3-D models associated with them, as indicated by the way the boxes overlap. The relative arrangement of each model’s component axes, however, is shown improperly, since it should be in an object-centered system rather than the viewer-centered projection used here (a more correct 3-D model is given by the table shown in Figure 5–5c). The important characteristics of this type of organization are: (1) Each 3-D model is a self-contained unit of shape information and has a limited complexity; (2) information appears in shape contexts appropriate for recognition (the disposition of a finger is most stable when specified relative to the hand that contains it); and (3) the representation can be manipulated flexibly. This approach limits the representation’s scope, however, since it is only useful for shapes that have well-defined 3-D model decompositions. (Reprinted by permission from D. Marr and H. K. Nishihara, “Representation and recognition of the spatial organization of three-dimensional shapes,” *Proc. R. Soc. Lond. B 200*, 269–294.)



Marr & Nishihara (1978)

Psychological theory: Recognition by components

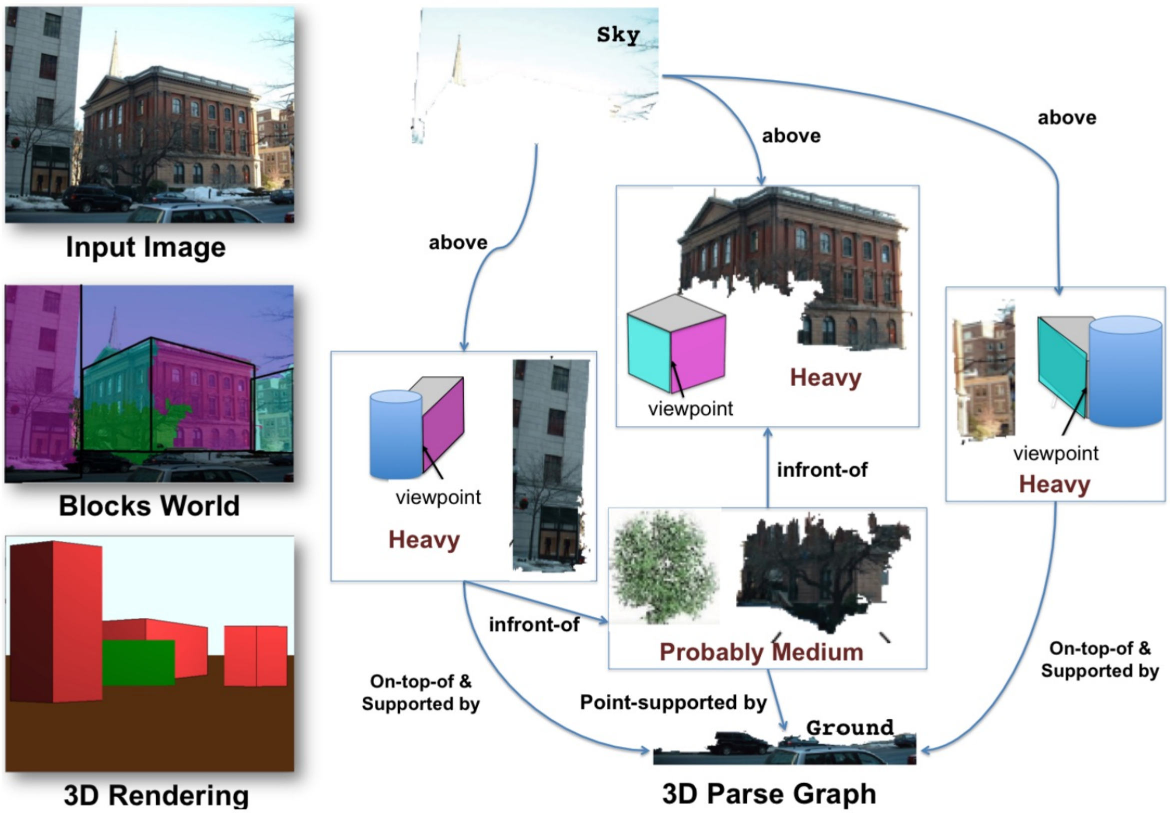
Primitives (*geons*) Objects



[Biederman (1987)]

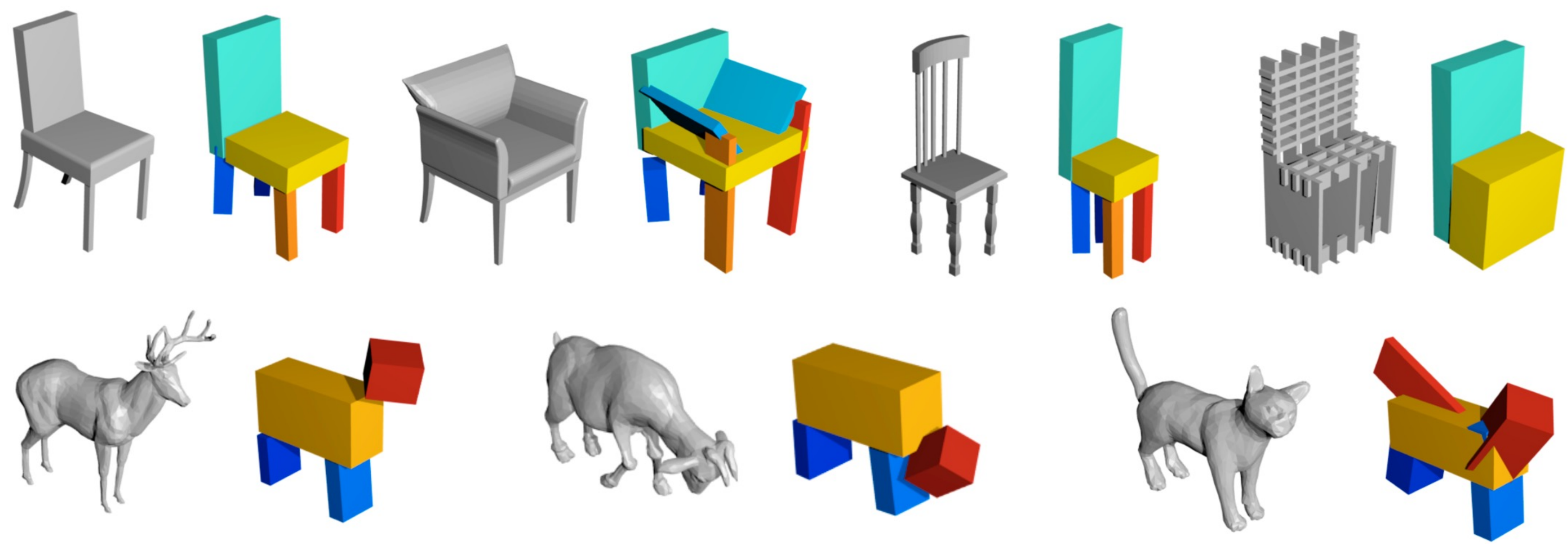
[http://en.wikipedia.org/wiki/Recognition_by_Components_Theory]

Today: Revival of 3D primitives



A. Gupta et al. [Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics.](#) ECCV 2010

Today: Revival of 3D primitives



"Here we do not wish to reprise the classic debates on the value of volumetric primitives – while they were oversold in the 70s and 80s, they suffer from complete neglect now, and we hope that this demonstration of feasibility of learning how to assemble an object from volumetric primitives will reignite interest."

S. Tulsiani et al. [Learning Shape Abstractions by Assembling Volumetric Primitives]. CVPR 2017

Today: Revival of 3D primitives

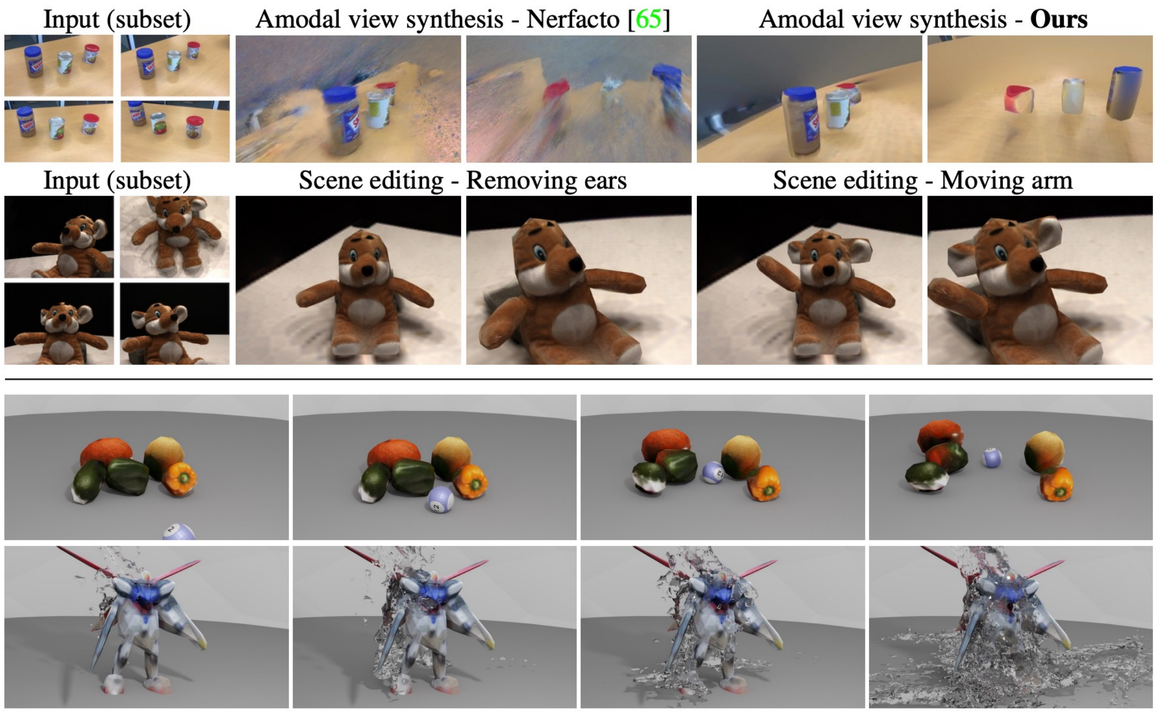


Figure 5: **Applications.** Amodal completion (1st row), scene editing (2nd) and physical simulations (bottom).

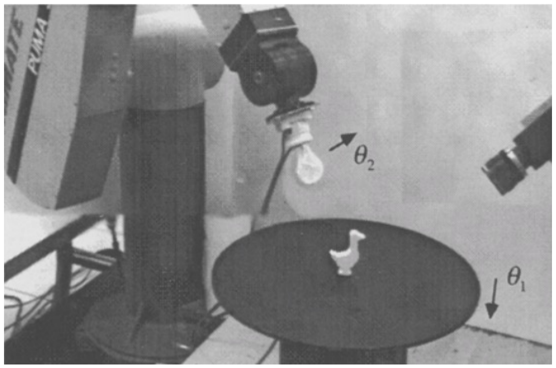
W. Liu et al. Marching-Primitives: Shape Abstraction from Signed Distance Function. CVPR 2023

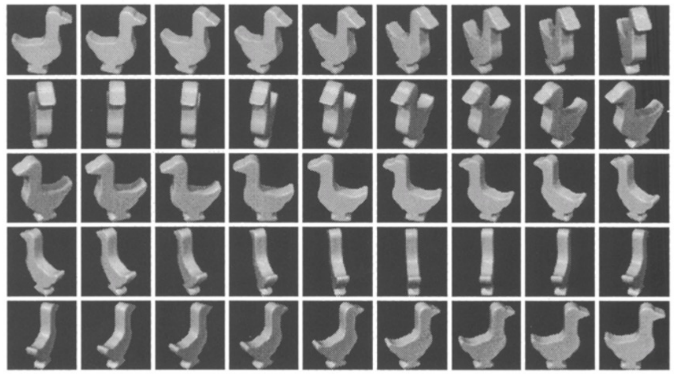
T. Monnier et al. Differentiable Blocks World: Qualitative 3D Decomposition by Rendering Primitives. arXiv 2023

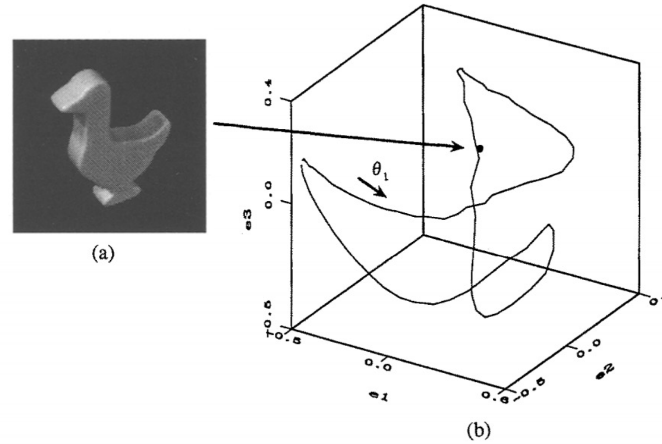
Outline

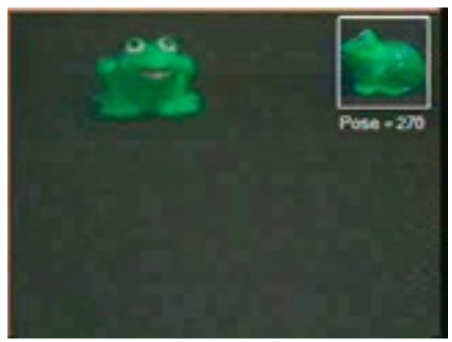
- Object representations
  - 3D shape
  - 3D primitives
  - 2D appearance-based models
  - 2D part-based models (deformable templates)
  - CNNs

Global appearance representations









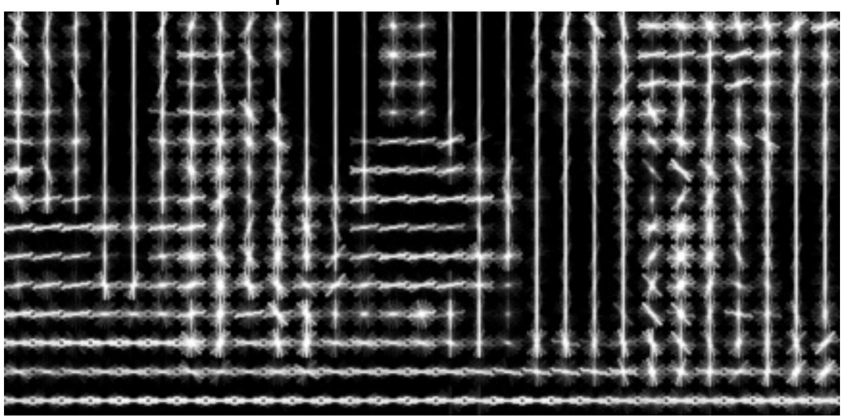
Recognition demo movie

H. Murase and S. Nayar, Visual learning and recognition of 3-d objects from appearance, IJCV 1995

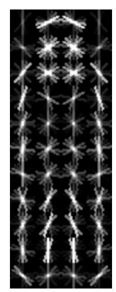
J. Mundy et al., An Experimental Comparison of Appearance and Geometric Model Based Recognition, 1996

Global appearance representation: HOG template

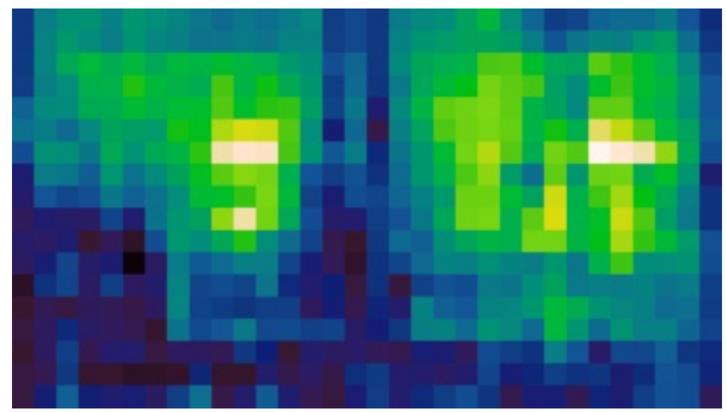
HOG feature map



Template



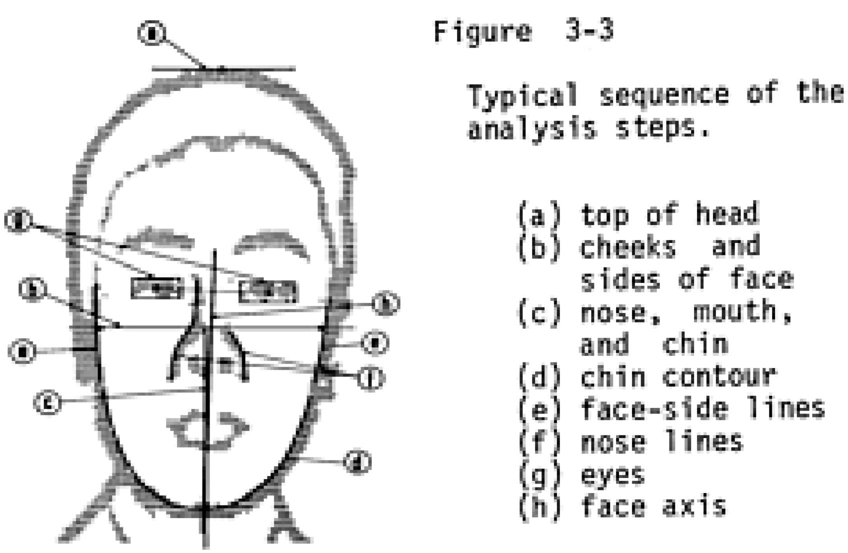
Detector response map



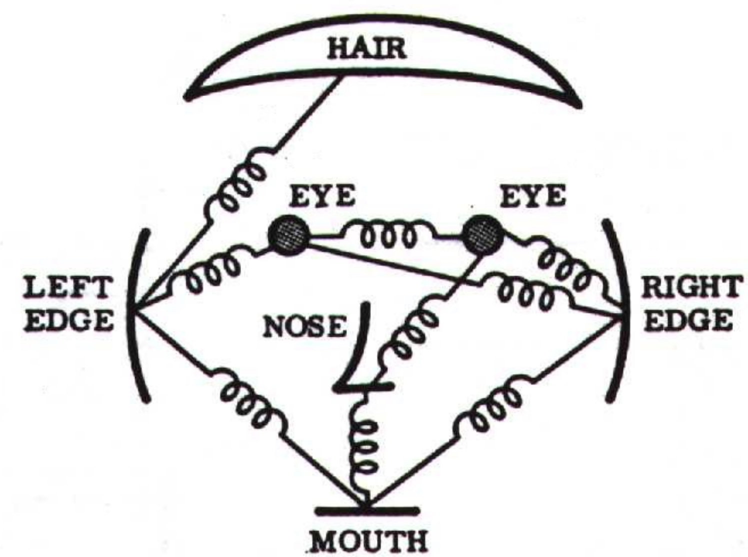
* =

N. Dalal and B. Triggs, [Histograms of Oriented Gradients for Human Detection](), CVPR 2005

2D part-based representations:
Local appearance + deformable shape



[Kanade](#) (1973)



[Fischler and Elschlager](#) (1973)

A procedural part-based recognition system

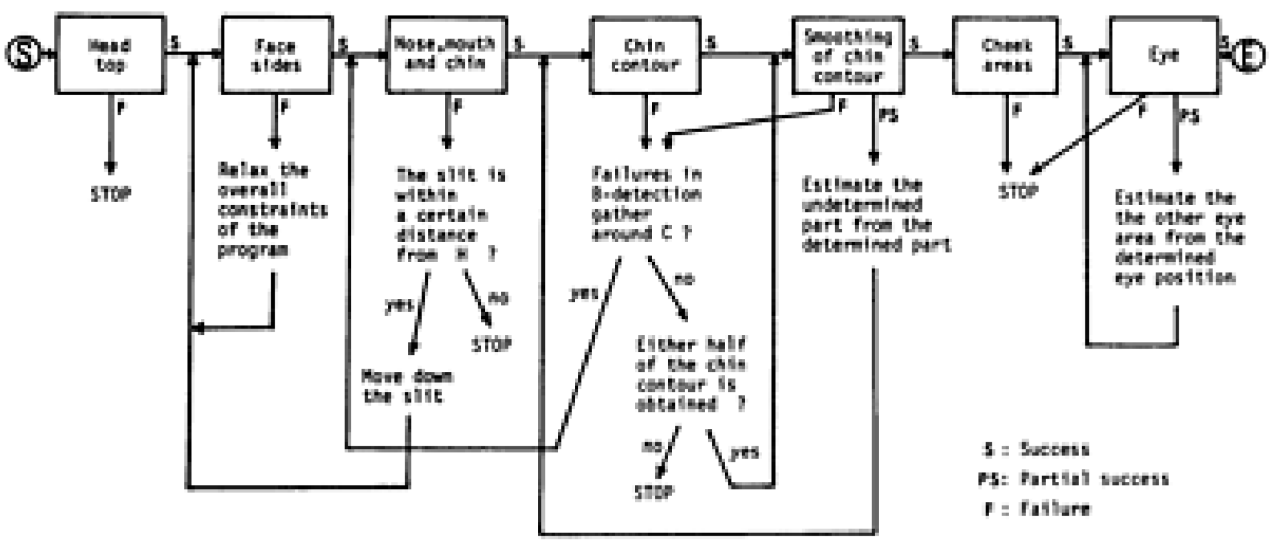


Figure 3-12 General flow of analysis program.

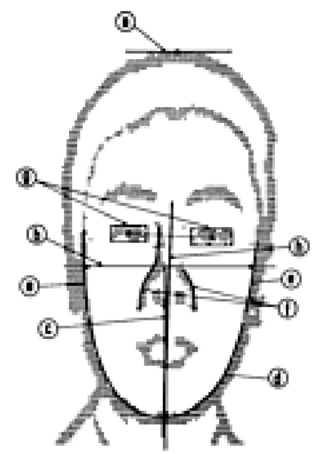


Figure 3-3

Typical sequence of the analysis steps.

(a) top of head
(b) cheeks and sides of face
(c) nose, mouth, and chin
(d) chin contour
(e) face-side lines
(f) nose lines
(g) eyes
(h) face axis

T. Kanade. Picture Processing System by Computer Complex and Recognition of Human Faces. Ph.D. dissertation, 1973

A procedural part-based recognition system

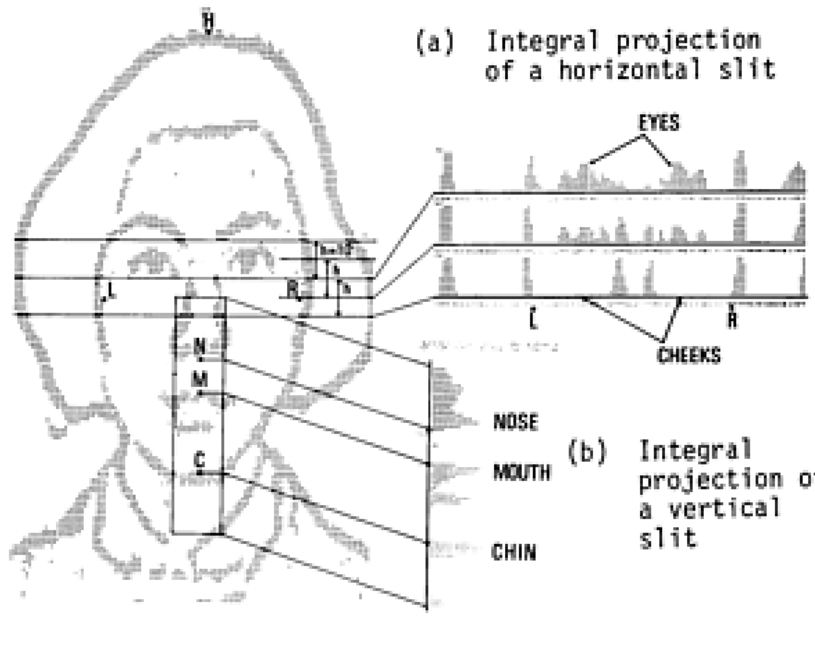


Figure 3-14 Detection of the face sides, nose, mouth, and chin by the application of slits.

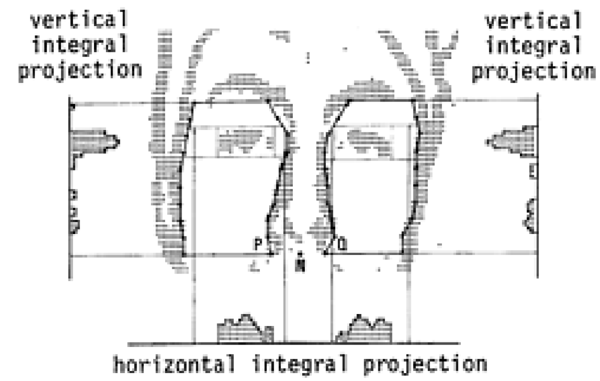


Figure 3-18 Cheek areas and rectangles which contain the eye.

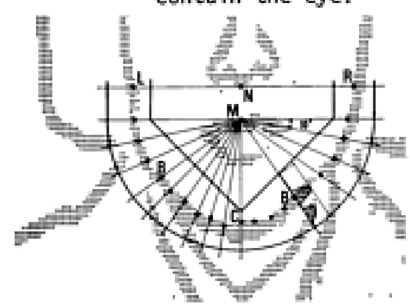


Figure 3-17

Extraction of chin contour. The search area is established and a slit is placed along each radial line.

T. Kanade. [Picture Processing System by Computer Complex and Recognition of Human Faces](). Ph.D. dissertation, 1973

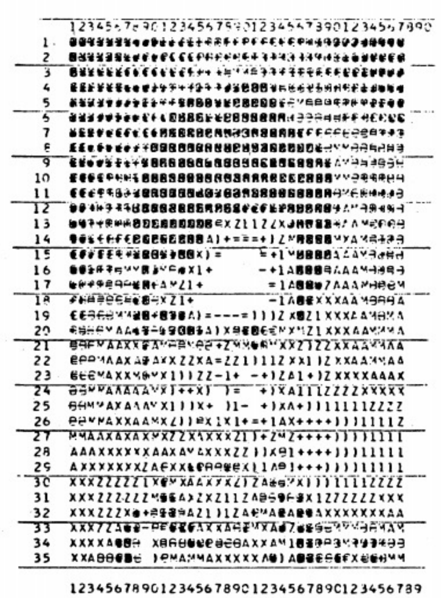
A procedural part-based recognition system

category of faces	number of faces	correct results	error or un-recovered failure	step in which the error or unrecovered failure occurred				
				face sides	nose mouth chin	chin contour	eyes	nose width
full face with no glasses or beard	670	608	62	5	14	17	19	7
full face with glasses	77	2	75	4	4	2	65	
face with turn or tilt	79	63	16	4	3	3	5	1
face with beard, and others	27		27		12	4		11

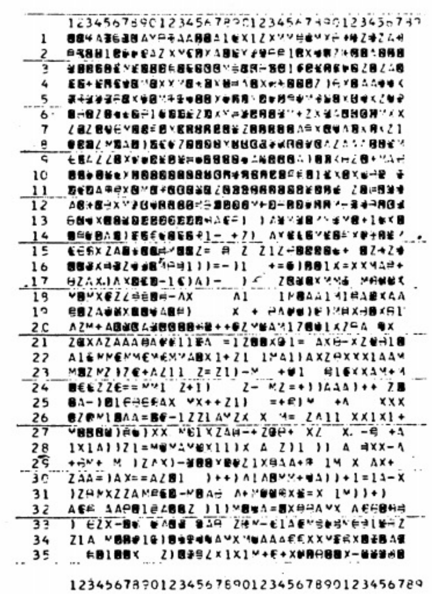
Table 3-1 Summary of results of analysis

T. Kanade. Picture Processing System by Computer Complex and Recognition of Human Faces. Ph.D. dissertation, 1973

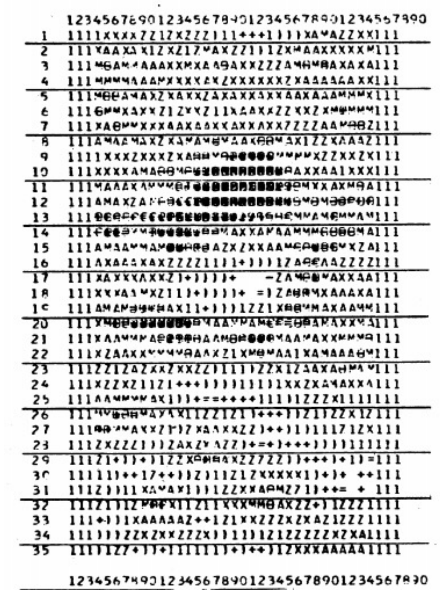
Part-based recognition by optimization



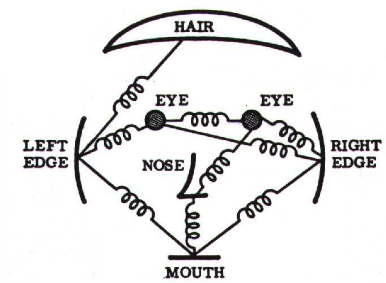
Original picture.



Noisy picture (sensed scene) as used in experiment.



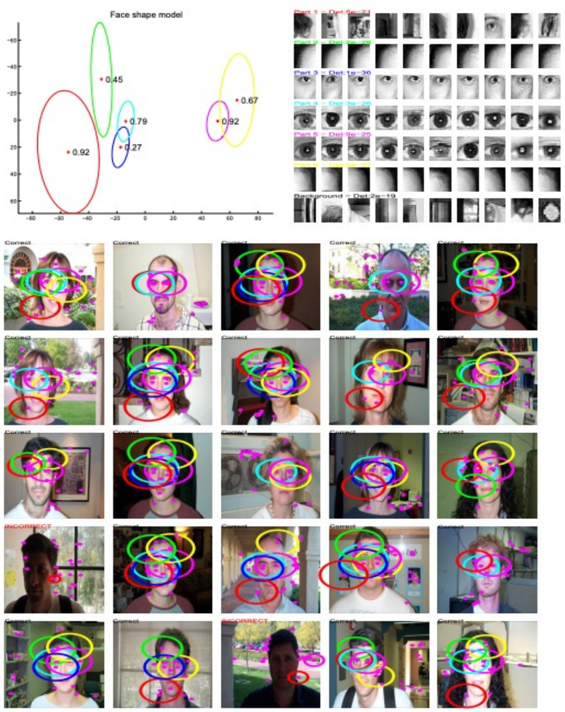
L(EV)A for hair. (Density at a point is proportional to probability that hair is present at that location.)

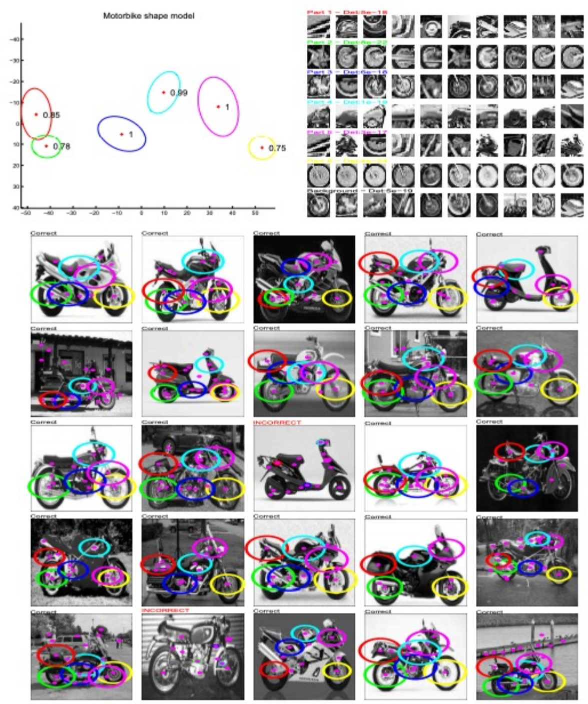


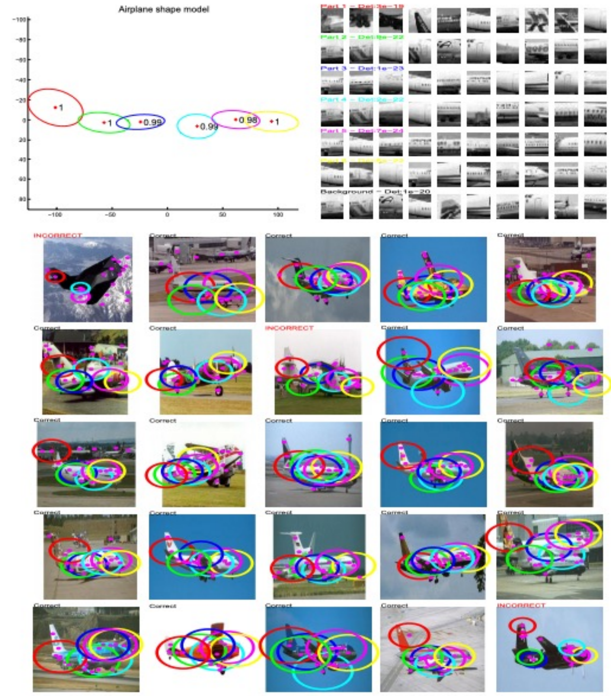
HAIR WAS LOCATED AT (11, 21)
L/EDGE WAS LOCATED AT (25, 11)
R/EDGE WAS LOCATED AT (25, 24)
L/EYE WAS LOCATED AT (21, 15)
R/EYE WAS LOCATED AT (21, 21)
NOSE WAS LOCATED AT (26, 18)
MOUTH WAS LOCATED AT (29, 17)

M. Fischler and R. Elschlager, [The representation and matching of pictorial structures](), IEEE Trans. on Computers, 1973

Constellation models

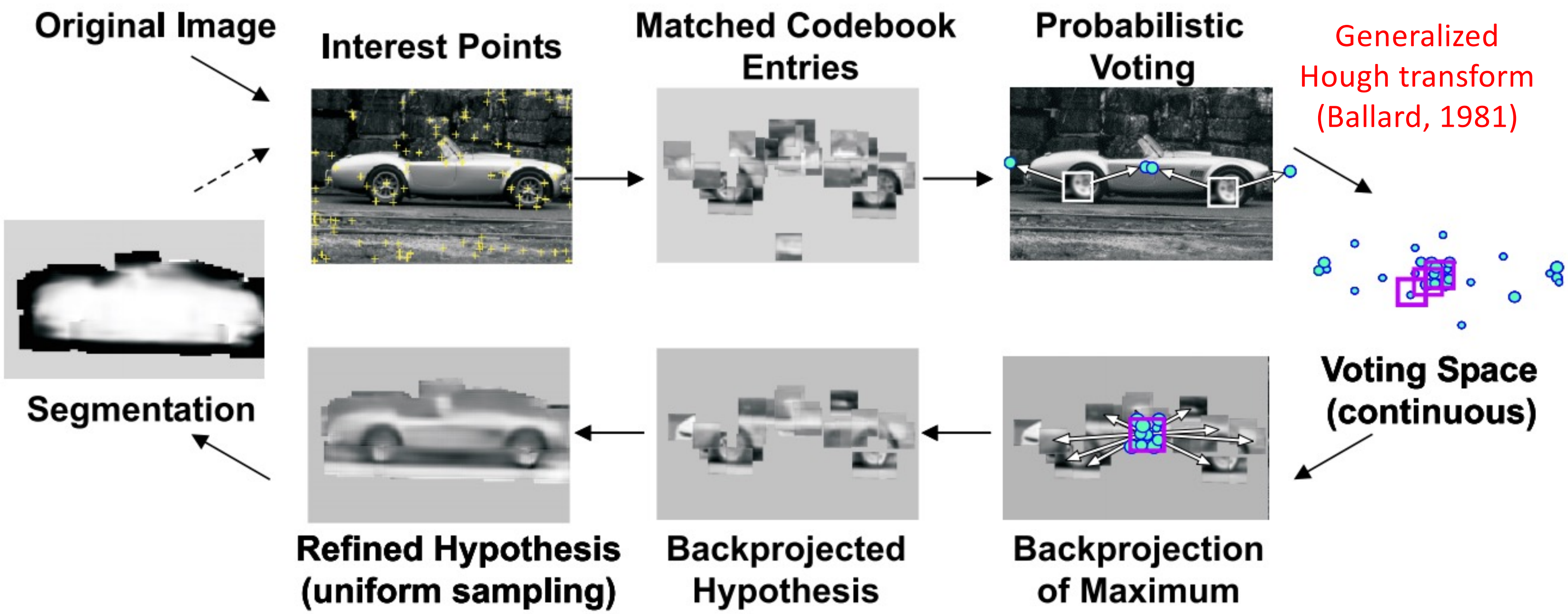






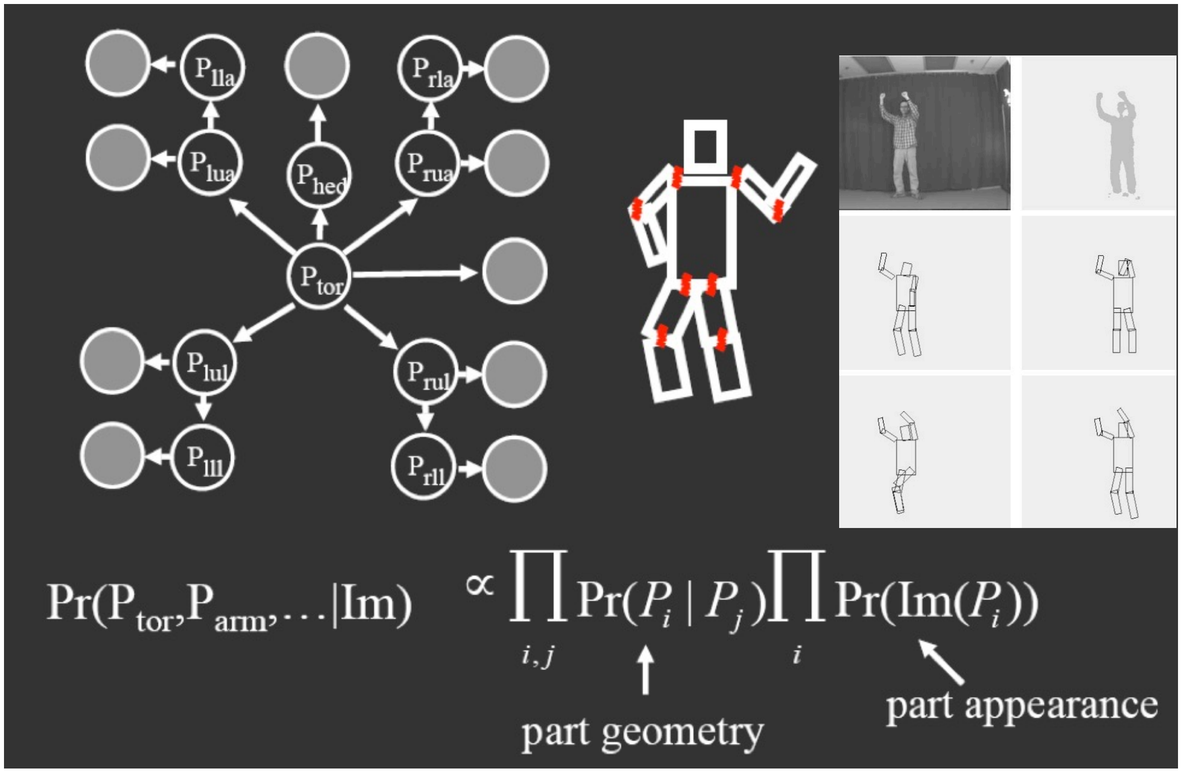
Burl, Weber, & Perona (1998); Weber, Welling & Perona (2000); Fergus, Perona & Zisserman (2003)

Implicit shape models



B. Leibe, A. Leonardis, and B. Schiele, [Combined Object Categorization and Segmentation with an Implicit Shape Model](), ECCV Workshop on Statistical Learning in Computer Vision, 2004

Pictorial structures revived



P. Felzenszwalb and D. Huttenlocher, Efficient matching of pictorial structures, CVPR 2000

P. Felzenszwalb and D. Huttenlocher, Pictorial structures for object recognition, IJCV 2005

Last time: Object and scene representations

- Object representations
  - 3D shape
  - 3D primitives
  - 2D appearance-based models
  - 2D part-based models (deformable templates)
  - CNNs
- Scene representations
  - Structured representations
  - Appearance-based representations
  - Bottom-up and top-down perceptual organization
- Trends

Finding people used to be *really hard!*

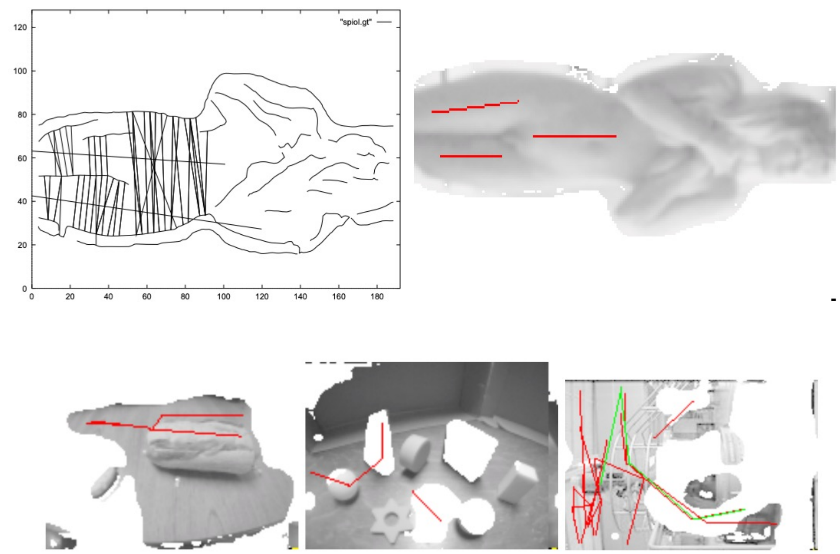
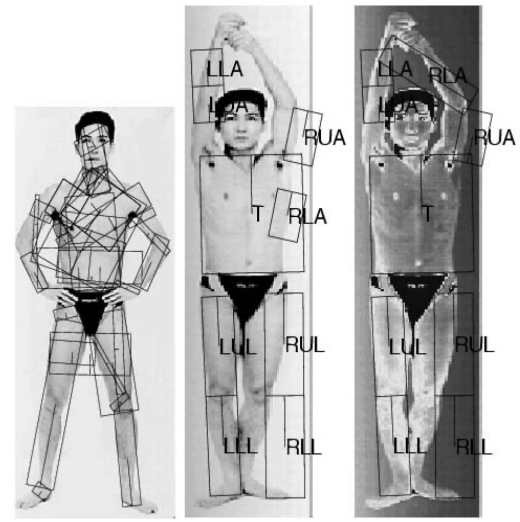


Fig. 6. Typical control images wrongly classified as containing naked people. These images contain people or skin-colored material (animal skin, wood, bread, off-white walls) and structures which the geometric grouper mistakes for spines or girdles. The grouper is frequently confused by groups of parallel edges, as in the industrial image.

[Fleck, Forsyth & Bregler (1996)]



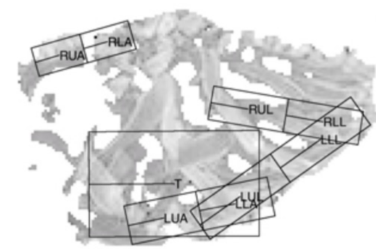
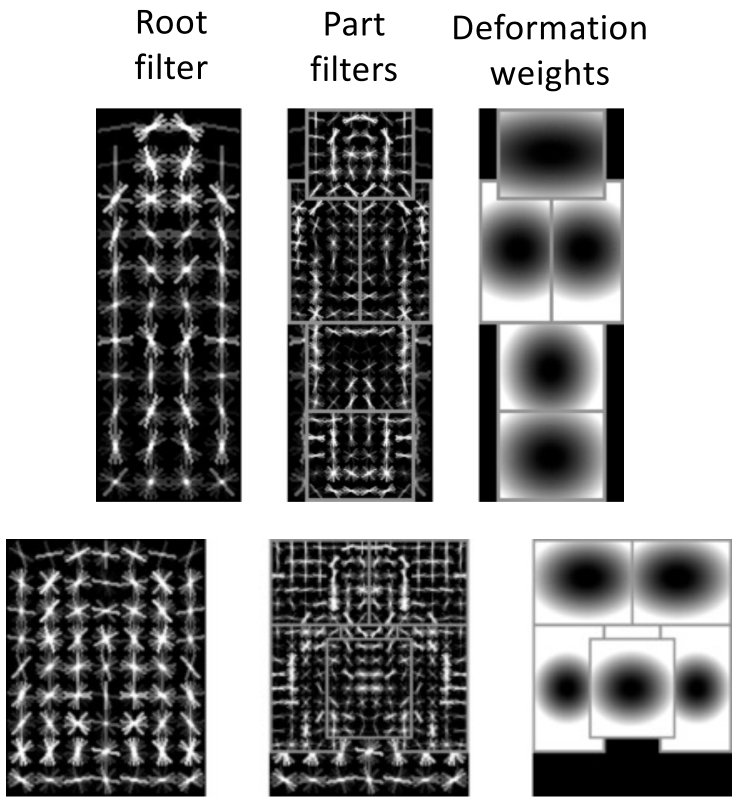


Figure 17. A negative image for which a human assembly was found. The assembly indeed looks like a configuration of a person. A better segment finder would not produce these segments and thus a person would not be detected. The white regions in the image are the pixels that have been masked out because they could not belong to a person due to their color.

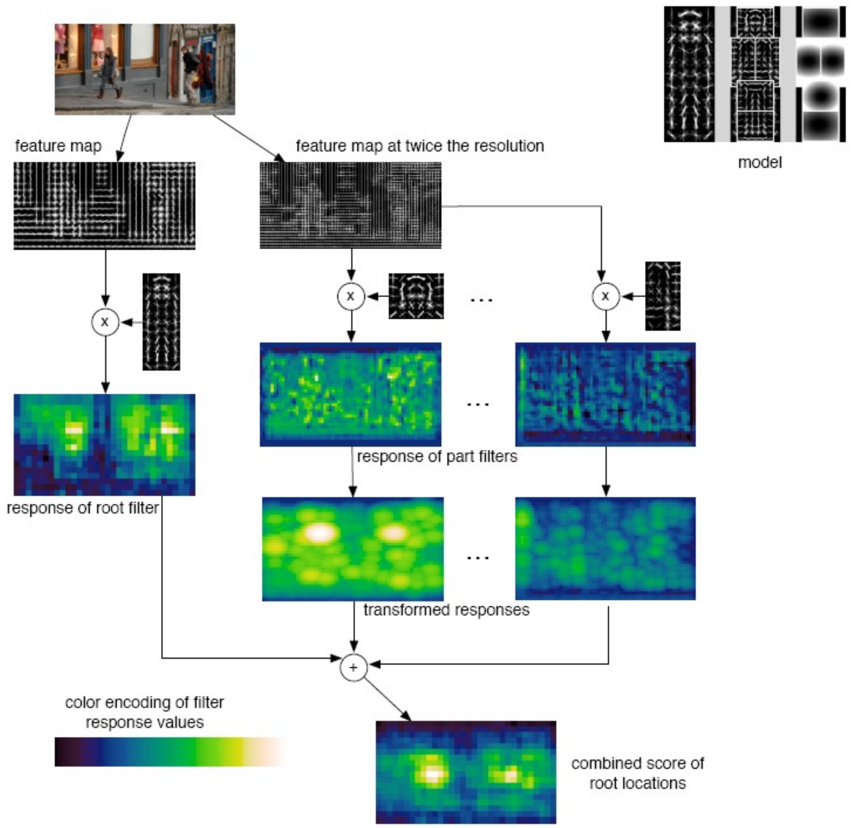
[Ioffe & Forsyth (2001)]

Discriminative deformable part models



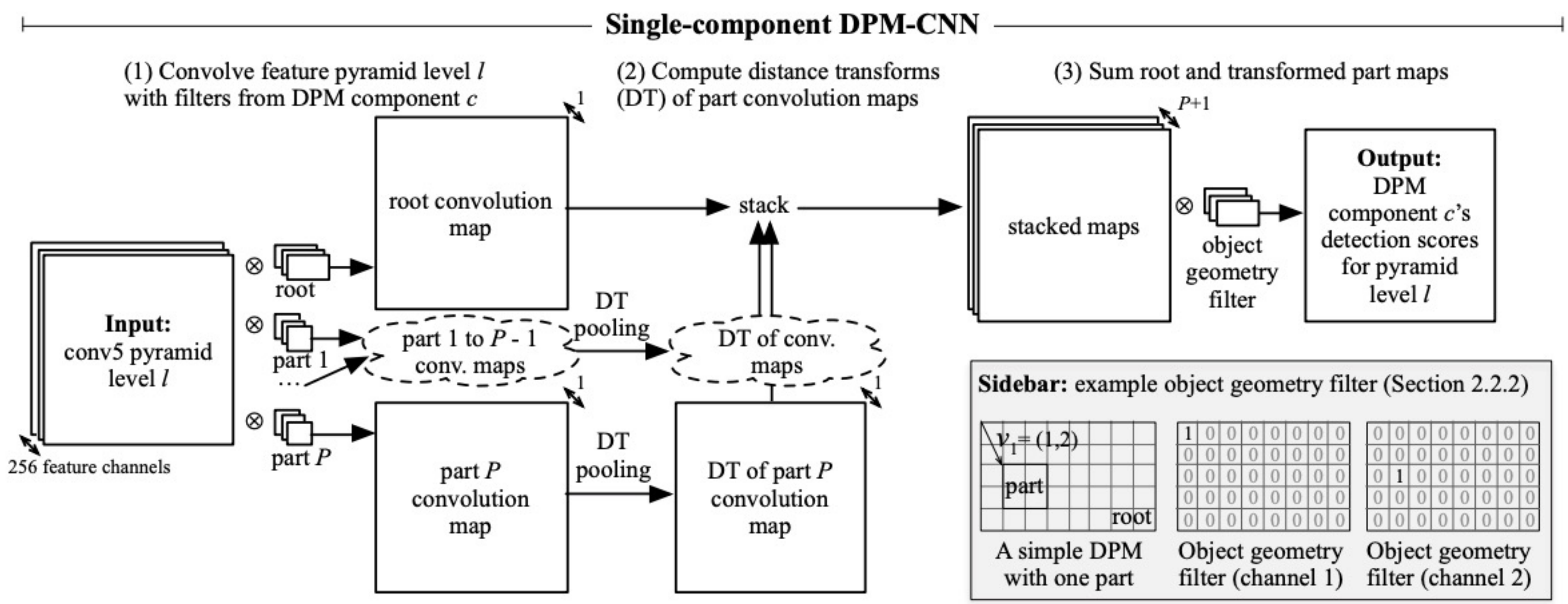
P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, [Object Detection with Discriminatively Trained Part-Based Models](), PAMI 2009

Discriminative deformable part models



P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, [Object Detection with Discriminatively Trained Part-Based Models](), PAMI 2009

Deformable part models as CNNs



R. Girshick, F. Iandola, T. Darrell, and J. Malik, Deformable Part Models are Convolutional Neural Networks, CVPR 2015

CNNs as deformable part models

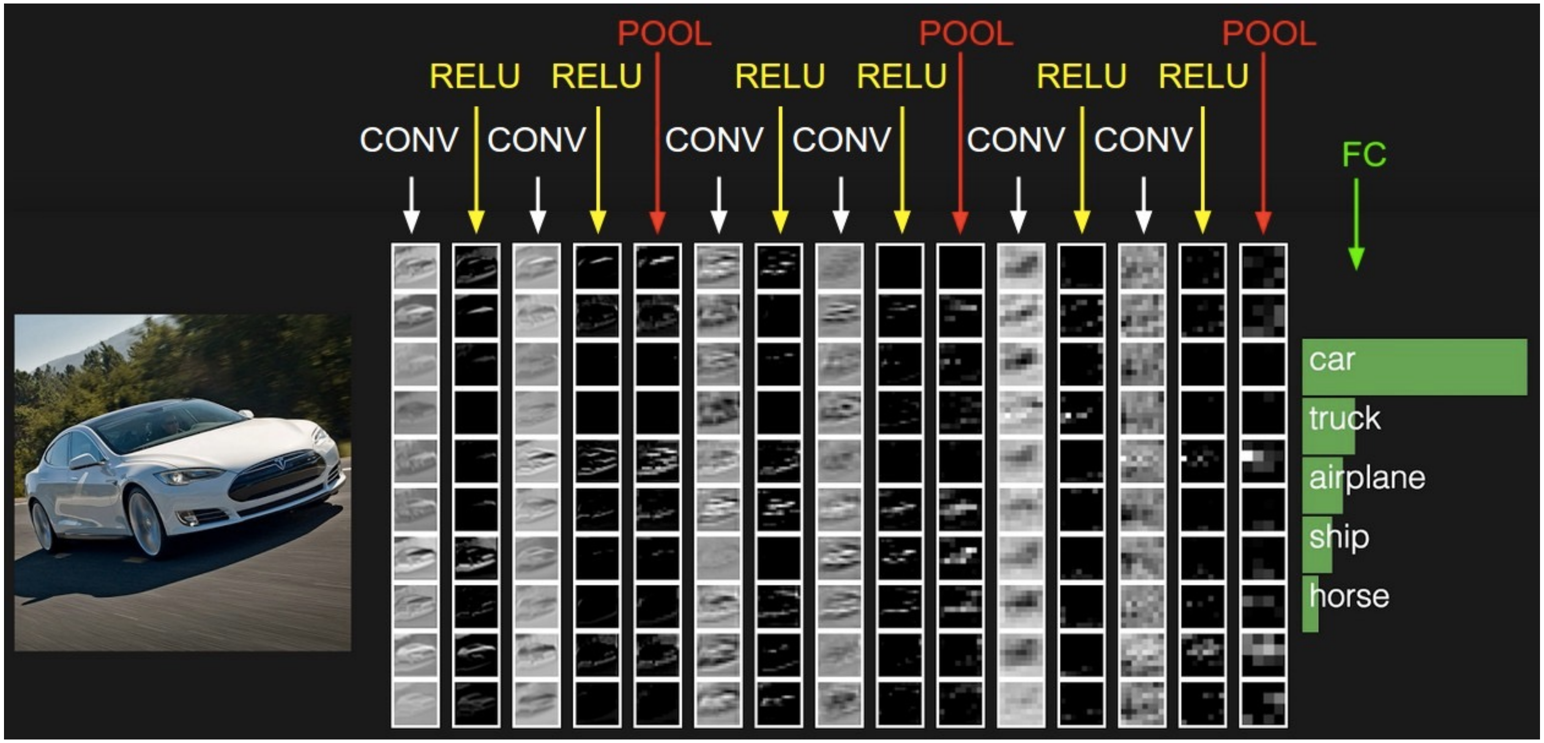
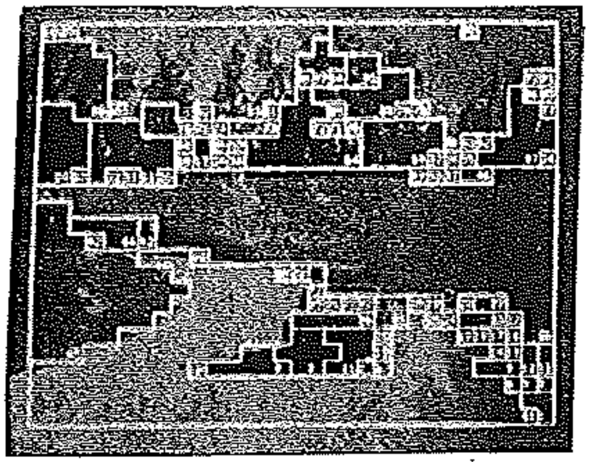


Image source: A. Karpathy

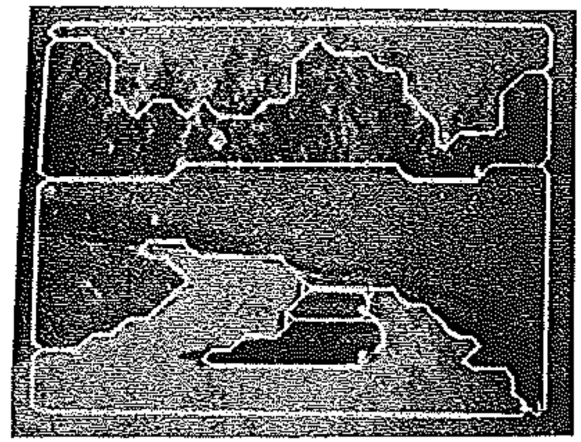
Outline

- Object representations
  - 3D shape
  - 3D primitives
  - 2D appearance-based models
  - 2D part-based models (deformable templates)
  - CNNs
- Scene representations
  - Structured representations
  - Appearance-based representations
  - Bottom-up and top-down perceptual organization

Structured scene representations



(B-2) Output of the non-semantic weakest boundary melted first region grower.



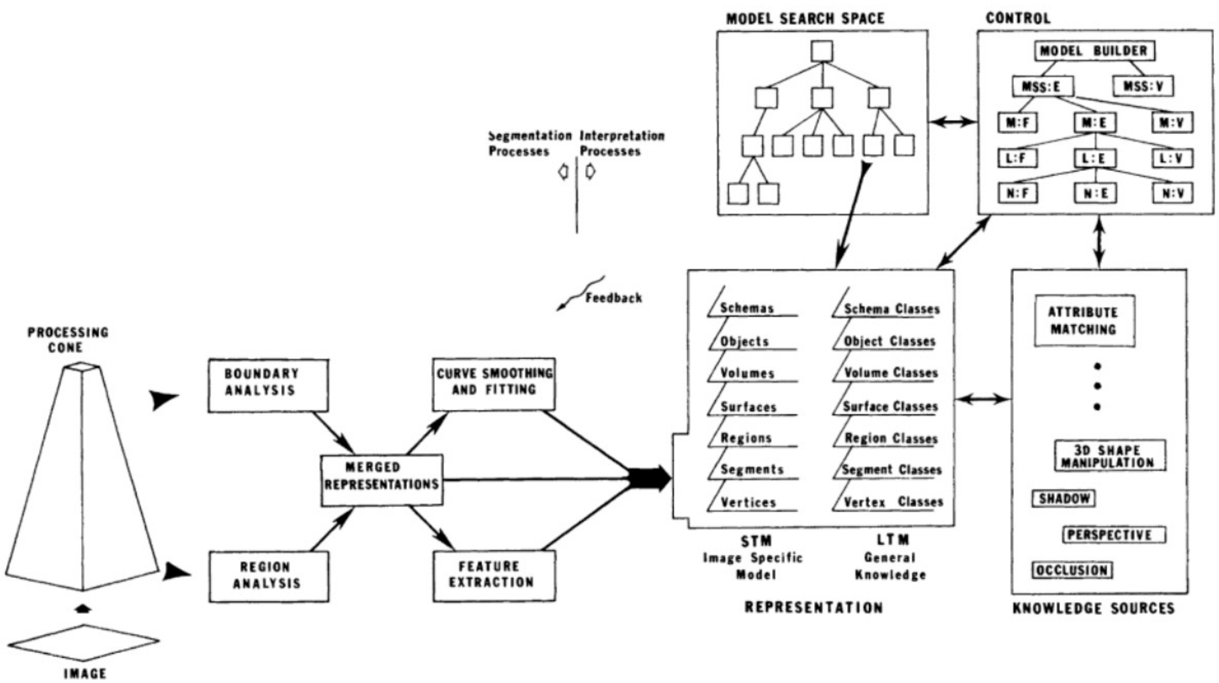
(B-5) Grouping regions by their assigned meaning, all regions considered mergable.

Slide credit: A. Efros

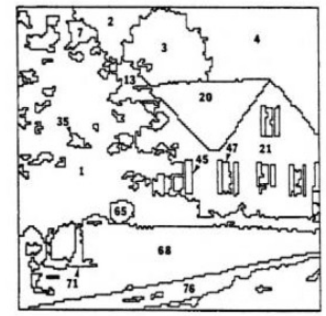
- Approach has everything: color, super-pixels, bottom-up segmentation, top-down parsing, inter- and intra-region reasoning, Bayesian formulation!

Y. Yakimovsky and J. Feldman, [A semantics-based decision theory region analyzer](), IJCAI 1973

Structured scene representations



(a)



(b)

Summary of Identification Accuracy

	All Regions: All 5 Objects	All Regions: 4 Objects After Collaps-ing Bush-Tree	Regions in Which at Least One 5×4 Window Will Fit: All 5 Objects	Regions in Which at Least One 5×4 Window Will Fit: 4 Objects	Large Regions (area ≥ 65 pixels): 5 Objects	Large Regions (area ≥ 65 pixels): 4 Objects
Total Number of Regions	209		83		45	
Number of Target Regions	99		50		25	
Number of Non-Target Regions	110		33		20	
Number of Target Regions Correctly Identified	76	91	40	45	19	23
Number of Target Regions Incorrectly Identified	23	8	10	5	6	2
% Target Regions Correct	76.7	91.9	80.	90.	76.	92.

A. Hanson and E. Riseman, VISIONS: A computer system for interpreting scenes, Computer Vision Systems, 1978

Structured scene representations

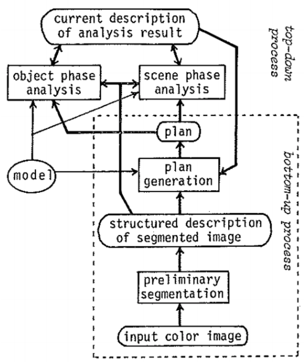


Figure 1. Flow of data in the analysis process.

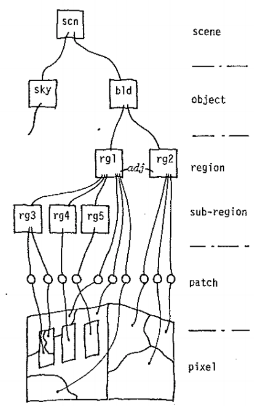


Figure 3. Structure of description built by the system.

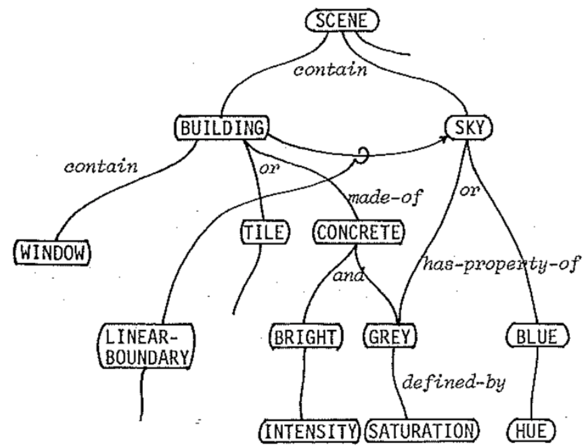


Figure 2. Semantic network for knowledge representation.

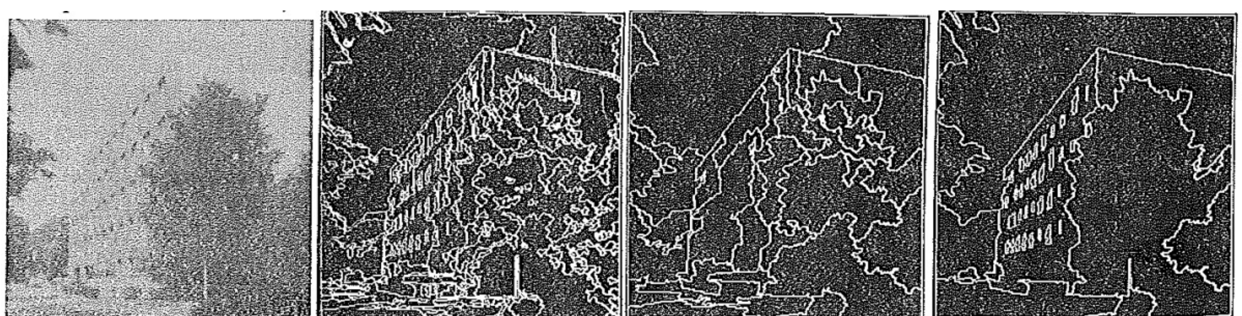
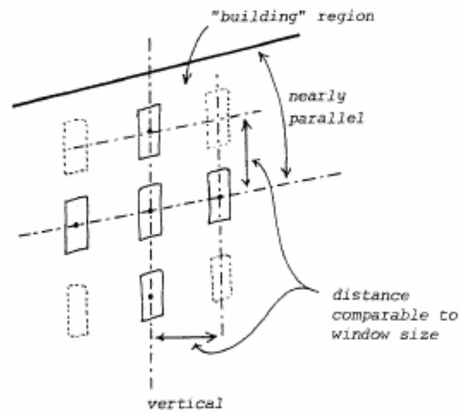


Figure 5-a. Digitized color scene. 5-b. Result of preliminary segmentation. 5-c. Plan image. 5-d. Result of semantic segmentation.

Y. Ohta, Takeo Kanade and T. Sakai, [An Analysis System for Scenes Containing objects with Substructures,](#) Proc. of the Fourth International Joint Conference on Pattern Recognition, pp. 752-754, January, 1978

What went wrong?



(a) "windows" and "building"

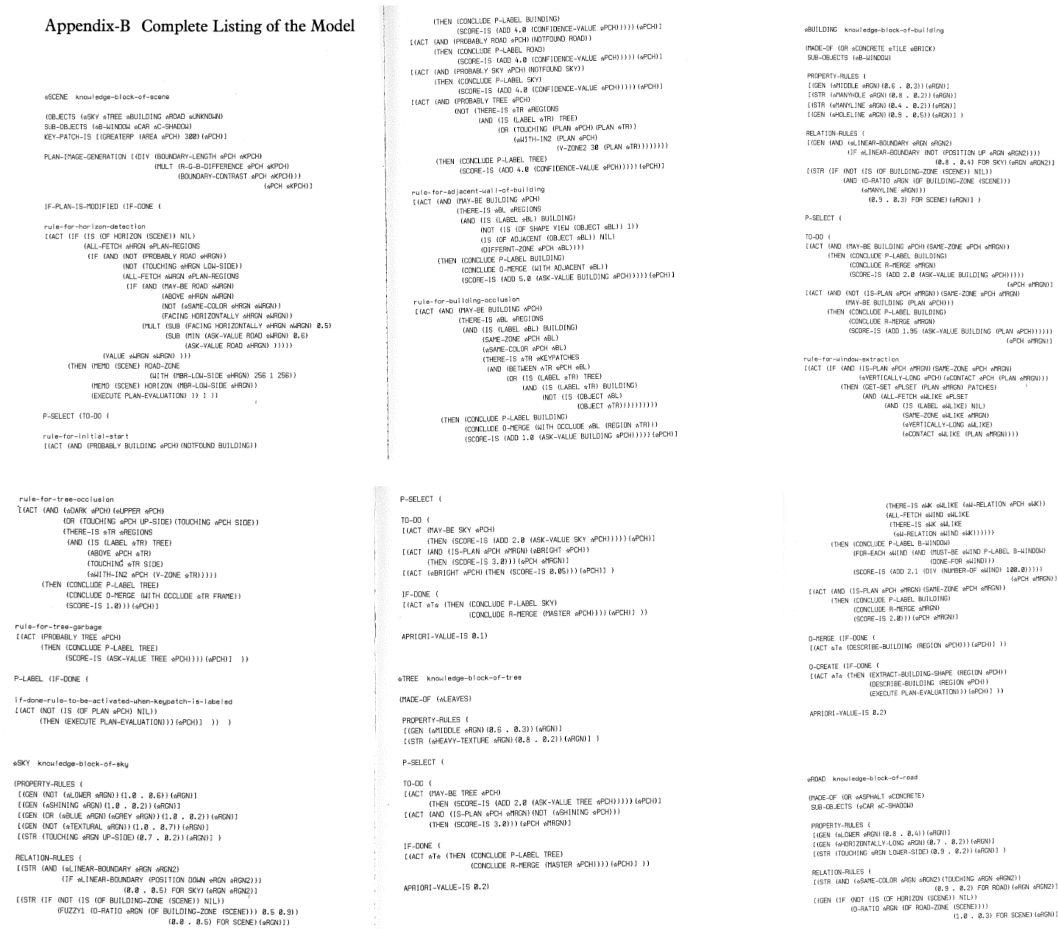
```
[(ACT (IF (AND (IS-PLAN *PCH *MRGN)                        ........... (1)
               (*VERTICALLY-LONG *PCH))
          (THEN (GET-SET *PLSET (PLAN *MRGN) PATCHES)       ....... (2)
                (AND (ALL-FETCH *WLIKE *PLSET               .......... (3)
                          (AND (IS (LABEL *WLIKE) NIL)
                               (*VERTICALLY-LONG *WLIKE)))
                     (ALL-FETCH *WIND *WLIKE                ........... (4)
                          (THERE-IS *WK *WLIKE
                           (*W-RELATION *WIND *WK))))))
      (THEN (CONCLUDE P-LABEL B-WINDOW)
            (FOR-EACH *WIND (AND (MUST-BE *WIND P-LABEL B-WINDOW)
                                 (DONE-FOR *WIND)))
            (SCORE-IS (ADD 2.1 (DIV (NUMBER-OF *WIND) 100.0)))))
                                                     (*PCH *MRGN)]
```

(b) listing of the to-do rule for "windows" detection

Slide credit:
A. Efros

Ohta & Kanade (1978)

What went wrong?



Appendix-B Complete Listing of the Model

```
*SCENE  knowledge-block-of-scene

(OBJECTS (*SKY *TREE *BUILDING *ROAD *UNKNOWN)
SUB-OBJECTS (*B-WINDOW *CAR *C-SHADOW)
KEY-PATCH-IS [(GREATERP (AREA *PCH) 300)(*PCH)]

PLAN-IMAGE-GENERATION [(DIV (BOUNDARY-LENGTH *PCH *KPCH)
                            (MULT (R-G-B-DIFFERENCE *PCH *KPCH)
                                  (BOUNDARY-CONTRAST *PCH *KPCH)))
                                                    (*PCH *KPCH)]

IF-PLAN-IS-MODIFIED (IF-DONE (

rule-for-horizon-detection
[(ACT (IF (IS (OF HORIZON (SCENE)) NIL)
          (ALL-FETCH *HRGN *PLAN-REGIONS
           (IF (AND (NOT (PROBABLY ROAD *HRGN))
                    (NOT (TOUCHING *HRGN LOW-SIDE))
                    (ALL-FETCH *WRGN *PLAN-REGIONS
                     (IF (AND (MAY-BE ROAD *WRGN)
                              (ABOVE *HRGN *WRGN)
                              (NOT (*SAME-COLOR *HRGN *WRGN))
                              (FACING HORIZONTALLY *HRGN *WRGN))
                         (MULT (SUB (FACING HORIZONTALLY *HRGN *WRGN) 0.5)
                               (SUB (MIN (ASK-VALUE ROAD *WRGN) 0.6)
                                    (ASK-VALUE ROAD *HRGN) )))))
               (VALUE *WRGN *WRGN) )))
      (THEN (MEMO (SCENE) ROAD-ZONE
                           (WITH (MBR-LOW-SIDE *HRGN) 256 1 256))
            (MEMO (SCENE) HORIZON (MBR-LOW-SIDE *HRGN))
            (EXECUTE PLAN-EVALUATION) )) ] ))

P-SELECT (TO-DO (

rule-for-initial-start
[(ACT (AND (PROBABLY BUILDING *PCH)(NOTFOUND BUILDING))
```

```
      (THEN (CONCLUDE P-LABEL BUINDING)
            (SCORE-IS (ADD 4.0 (CONFIDENCE-VALUE *PCH)))))(*PCH)]
[(ACT (AND (PROBABLY ROAD *PCH)(NOTFOUND ROAD))
      (THEN (CONCLUDE P-LABEL ROAD)
            (SCORE-IS (ADD 4.0 (CONFIDENCE-VALUE *PCH)))))(*PCH)]
[(ACT (AND (PROBABLY SKY *PCH)(NOTFOUND SKY))
      (THEN (CONCLUDE P-LABEL SKY)
            (SCORE-IS (ADD 4.0 (CONFIDENCE-VALUE *PCH)))))(*PCH)]
[(ACT (AND (PROBABLY TREE *PCH)
           (NOT (THERE-IS *TR *REGIONS
                 (AND (IS (LABEL *TR) TREE)
                      (OR (TOUCHING (PLAN *PCH)(PLAN *TR))
                          (*WITH-IN2 (PLAN *PCH)
                                     (V-ZONE2 30 (PLAN *TR)))))))
      (THEN (CONCLUDE P-LABEL TREE)
            (SCORE-IS (ADD 4.0 (CONFIDENCE-VALUE *PCH)))))(*PCH)]

rule-for-adjacent-wall-of-building
[(ACT (AND (MAY-BE BUILDING *PCH)
           (THERE-IS *BL *REGIONS
            (AND (IS (LABEL *BL) BUILDING)
                 (NOT (IS (OF SHAPE VIEW (OBJECT *BL)) 1))
                 (IS (OF ADJACENT (OBJECT *BL)) NIL)
                 (DIFFERNT-ZONE *PCH *BL))))
      (THEN (CONCLUDE P-LABEL BUILDING)
            (CONCLUDE O-MERGE (WITH ADJACENT *BL))
            (SCORE-IS (ADD 5.0 (ASK-VALUE BUILDING *PCH)))))(*PCH)]

rule-for-building-occlusion
[(ACT (AND (MAY-BE BUILDING *PCH)
           (THERE-IS *BL *REGIONS
            (AND (IS (LABEL *BL) BUILDING)
                 (SAME-ZONE *PCH *BL)
                 (*SAME-COLOR *PCH *BL)
                 (THERE-IS *TR *KEYPATCHES
                  (AND (BETWEEN *TR *PCH *BL)
                       (OR (IS (LABEL *TR) TREE)
                           (AND (IS (LABEL *TR) BUILDING)
                                (NOT (IS (OBJECT *BL)
                                         (OBJECT *TR))))))))))
      (THEN (CONCLUDE P-LABEL BUILDING)
            (CONCLUDE O-MERGE (WITH OCCLUDE *BL (REGION *TR)))
            (SCORE-IS (ADD 1.0 (ASK-VALUE BUILDING *PCH)))))(*PCH)]
```

```
rule-for-tree-occlusion
[(ACT (AND (*DARK *PCH)(*UPPER *PCH)
           (OR (TOUCHING *PCH UP-SIDE)(TOUCHING *PCH SIDE))
           (THERE-IS *TR *REGIONS
            (AND (IS (LABEL *TR) TREE)
                 (ABOVE *PCH *TR)
                 (TOUCHING *TR SIDE)
                 (*WITH-IN2 *PCH (V-ZONE *TR)))))
      (THEN (CONCLUDE P-LABEL TREE)
            (CONCLUDE O-MERGE (WITH OCCLUDE *TR FRAME))
            (SCORE-IS 1.0)))(*PCH)]

rule-for-tree-garbage
[(ACT (PROBABLY TREE *PCH)
      (THEN (CONCLUDE P-LABEL TREE)
            (SCORE-IS (ASK-VALUE TREE *PCH))))(*PCH)]  ))

P-LABEL (IF-DONE (

if-done-rule-to-be-activated-when-keypatch-is-labeled
[(ACT (NOT (IS (OF PLAN *PCH) NIL))
      (THEN (EXECUTE PLAN-EVALUATION)))(*PCH)]  ))  )


*SKY  knowledge-block-of-sky

(PROPERTY-RULES (
 [(GEN (NOT (*LOWER *RGN))(1.0 . 0.6))(*RGN)]
 [(GEN (*SHINING *RGN)(1.0 . 0.2))(*RGN)]
 [(GEN (OR (*BLUE *RGN)(*GREY *RGN))(1.0 . 0.2))(*RGN)]
 [(GEN (NOT (*TEXTURAL *RGN))(1.0 . 0.7))(*RGN)]
 [(STR (TOUCHING *RGN UP-SIDE)(0.7 . 0.2))(*RGN)] )

RELATION-RULES (
[(STR (AND (*LINEAR-BOUNDARY *RGN *RGN2)
           (IF *LINEAR-BOUNDARY (POSITION DOWN *RGN *RGN2)))
                        (0.0 . 0.5) FOR SKY)(*RGN *RGN2)]
[(STR (IF (NOT (IS (OF BUILDING-ZONE (SCENE)) NIL))
          (FUZZY1 (O-RATIO *RGN (OF BUILDING-ZONE (SCENE))) 0.5 0.9))
                                   (0.0 . 0.5) FOR SCENE)(*RGN)])
```

```
P-SELECT (

TO-DO (
[(ACT (MAY-BE SKY *PCH)
      (THEN (SCORE-IS (ADD 2.0 (ASK-VALUE SKY *PCH)))))(*PCH)]
[(ACT (AND (IS-PLAN *PCH *MRGN)(*BRIGHT *PCH))
      (THEN (SCORE-IS 3.0)))(*PCH *MRGN)]
[(ACT (*BRIGHT *PCH)(THEN (SCORE-IS 0.05)))(*PCH)] )

IF-DONE (
[(ACT *T* (THEN (CONCLUDE P-LABEL SKY)
                (CONCLUDE R-MERGE (MASTER *PCH))))(*PCH)] ))

APRIORI-VALUE-IS 0.1)


*TREE  knowledge-block-of-tree

(MADE-OF (*LEAVES)

 PROPERTY-RULES (
 [(GEN (*MIDDLE *RGN)(0.6 . 0.3))(*RGN)]
 [(STR (*HEAVY-TEXTURE *RGN)(0.8 . 0.2))(*RGN)] )

 P-SELECT (

 TO-DO (
 [(ACT (MAY-BE TREE *PCH)
       (THEN (SCORE-IS (ADD 2.0 (ASK-VALUE TREE *PCH)))))(*PCH)]
 [(ACT (AND (IS-PLAN *PCH *MRGN)(NOT (*SHINING *PCH)))
       (THEN (SCORE-IS 3.0)))(*PCH *MRGN)]

 IF-DONE (
 [(ACT *T* (THEN (CONCLUDE P-LABEL TREE)
                 (CONCLUDE R-MERGE (MASTER *PCH))))(*PCH)] ))

 APRIORI-VALUE-IS 0.2)
```

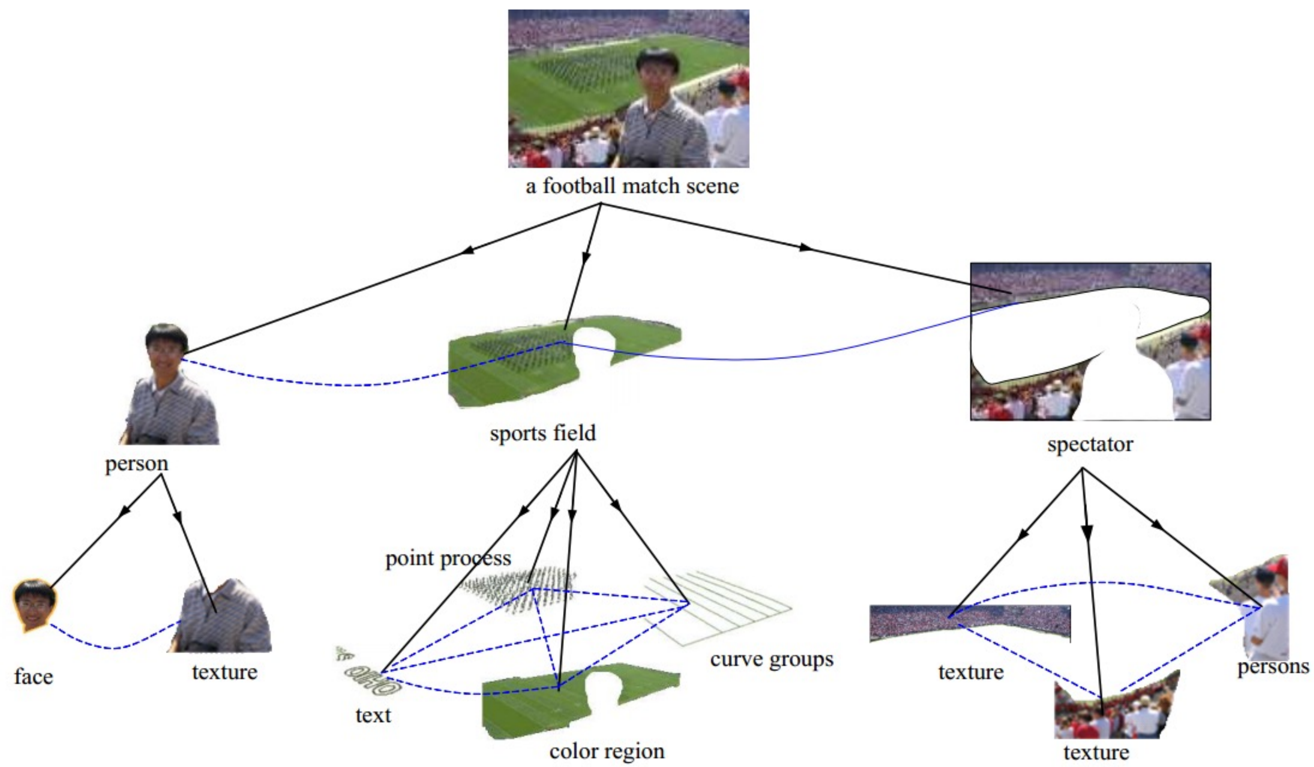
```
*BUILDING  knowledge-block-of-building

(MADE-OF (OR *CONCRETE *TILE *BRICK)
 SUB-OBJECTS (*B-WINDOW)

 PROPERTY-RULES (
 [(GEN (*MIDDLE *RGN)(0.6 . 0.3))(*RGN)]
 [(STR (*MANYHOLE *RGN)(0.8 . 0.2))(*RGN)]
 [(STR (*MANYLINE *RGN)(0.4 . 0.2))(*RGN)]
 [(GEN (*HOLELINE *RGN)(0.9 . 0.5))(*RGN)] )

 RELATION-RULES (
 [(GEN (AND (*LINEAR-BOUNDARY *RGN *RGN2)
            (IF *LINEAR-BOUNDARY (NOT (POSITION UP *RGN *RGN2))))
                               (0.8 . 0.4) FOR SKY)(*RGN *RGN2)]
 [(STR (IF (NOT (IS (OF BUILDING-ZONE (SCENE)) NIL))
           (AND (O-RATIO *RGN (OF BUILDING-ZONE (SCENE)))
                (*MANYLINE *RGN)))
                 (0.9 . 0.3) FOR SCENE)(*RGN)] )

P-SELECT (

TO-DO (
[(ACT (AND (MAY-BE BUILDING *PCH)(SAME-ZONE *PCH *MRGN))
      (THEN (CONCLUDE P-LABEL BUILDING)
            (CONCLUDE R-MERGE *MRGN)
            (SCORE-IS (ADD 2.0 (ASK-VALUE BUILDING *PCH)))))
                                                  (*PCH *MRGN)]
[(ACT (AND (NOT (IS-PLAN *PCH *MRGN))(SAME-ZONE *PCH *MRGN)
           (MAY-BE BUILDING (PLAN *PCH)))
      (THEN (CONCLUDE P-LABEL BUILDING)
            (CONCLUDE R-MERGE *MRGN)
            (SCORE-IS (ADD 1.95 (ASK-VALUE BUILDING (PLAN *PCH))))))
                                                  (*PCH *MRGN)]

rule-for-window-extraction
[(ACT (IF (AND (IS-PLAN *PCH *MRGN)(SAME-ZONE *PCH *MRGN)
               (*VERTICALLY-LONG *PCH)(*CONTACT *PCH (PLAN *MRGN)))
          (THEN (GET-SET *PLSET (PLAN *MRGN) PATCHES)
                (AND (ALL-FETCH *WLIKE *PLSET
                      (AND (IS (LABEL *WLIKE) NIL)
                           (SAME-ZONE *WLIKE *MRGN)
                           (*VERTICALLY-LONG *WLIKE)
                           (*CONTACT *WLIKE (PLAN *MRGN))))
```

```
                     (THERE-IS *WK *WLIKE (*W-RELATION *PCH *WK))
                     (ALL-FETCH *WIND *WLIKE
                      (THERE-IS *WK *WLIKE
                       (*W-RELATION *WIND *WK))))))
      (THEN (CONCLUDE P-LABEL B-WINDOW)
            (FOR-EACH *WIND (AND (MUST-BE *WIND P-LABEL B-WINDOW)
                                 (DONE-FOR *WIND)))
            (SCORE-IS (ADD 2.1 (DIV (NUMBER-OF *WIND) 100.0)))))
                                                  (*PCH *MRGN)]
[(ACT (AND (IS-PLAN *PCH *MRGN)(SAME-ZONE *PCH *MRGN))
      (THEN (CONCLUDE P-LABEL BUILDING)
            (CONCLUDE R-MERGE *MRGN)
            (SCORE-IS 2.0)))(*PCH *MRGN)]

O-MERGE (IF-DONE (
[(ACT *T* (DESCRIBE-BUILDING (REGION *PCH)))(*PCH)] ))

O-CREATE (IF-DONE (
[(ACT *T* (THEN (EXTRACT-BUILDING-SHAPE (REGION *PCH))
                (DESCRIBE-BUILDING (REGION *PCH))
                (EXECUTE PLAN-EVALUATION)))(*PCH)] ))

APRIORI-VALUE-IS 0.2)


*ROAD  knowledge-block-of-road

(MADE-OF (OR *ASPHALT *CONCRETE)
 SUB-OBJECTS (*CAR *C-SHADOW)

 PROPERTY-RULES (
 [(GEN (*LOWER *RGN)(0.8 . 0.4))(*RGN)]
 [(GEN (*HORIZONTALLY-LONG *RGN)(0.7 . 0.2))(*RGN)]
 [(STR (TOUCHING *RGN LOWER-SIDE)(0.9 . 0.2))(*RGN)] )

 RELATION-RULES (
 [(STR (AND (*SAME-COLOR *RGN *RGN2)(TOUCHING *RGN *RGN2))
                              (0.9 . 0.2) FOR ROAD)(*RGN *RGN2)]
 [(GEN (IF (NOT (IS (OF HORIZON (SCENE)) NIL))
           (O-RATIO *RGN (OF ROAD-ZONE (SCENE))))
                                   (1.0 . 0.3) FOR SCENE)(*RGN)]
```

Slide credit:
A. Efros

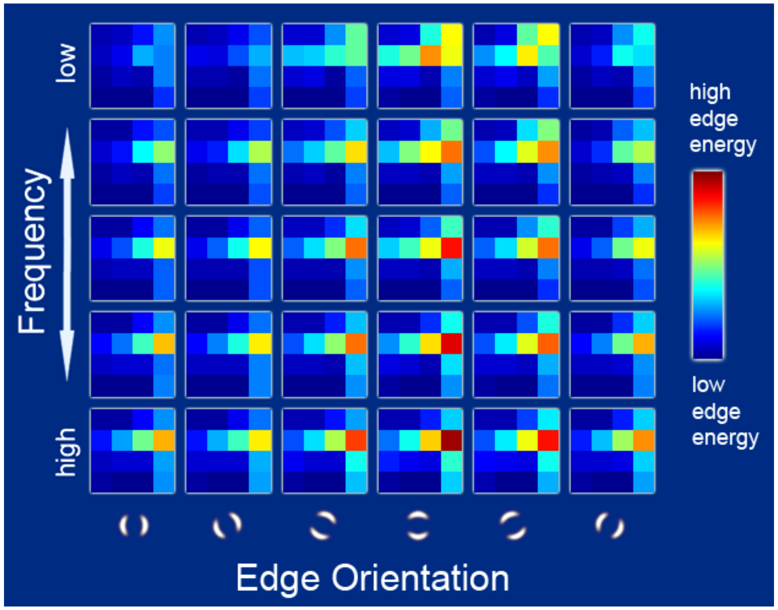
Ohta & Kanade (1978)

Structured scene representations revisited



Z. Tu et al. [Image Parsing: Unifying Segmentation, Detection, and Object Recognition](), ICCV 2003, IJCV 2005

Appearance-based scene representation: GIST



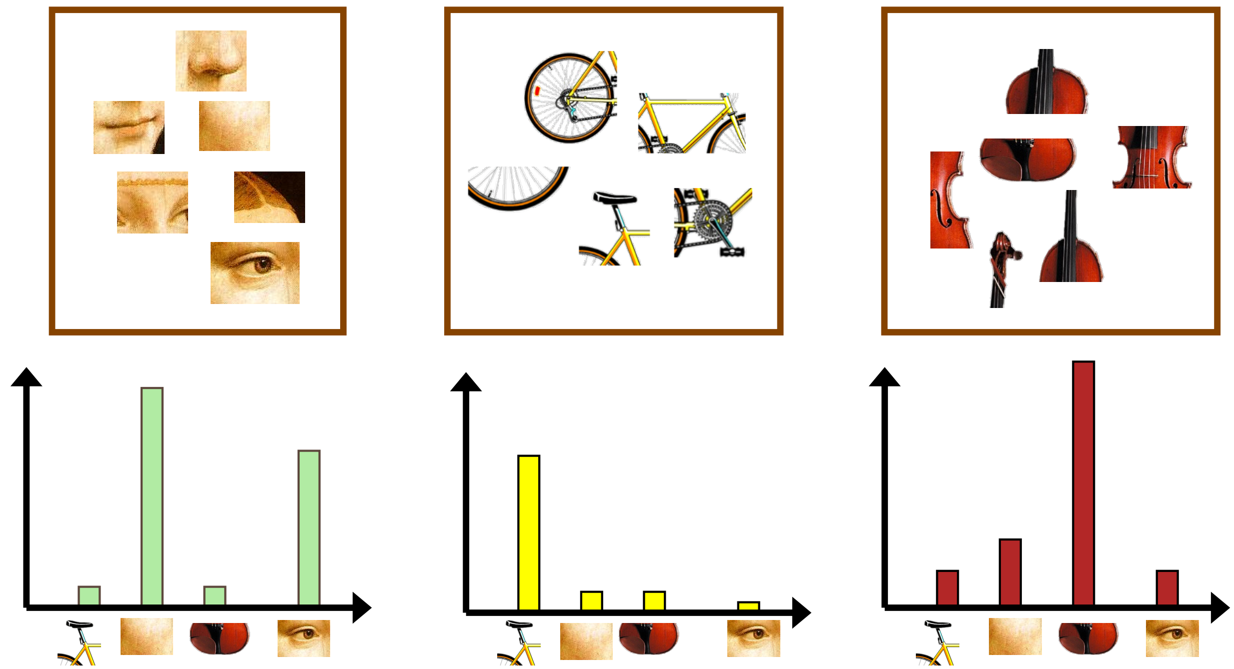
A. Oliva and A. Torralba. [Modeling the shape of the scene: A holistic representation of the spatial envelope](). IJCV 2001

Appearance-based scene representation: GIST

- Matching of scenes based on GIST works surprisingly well – *given a large enough dataset*

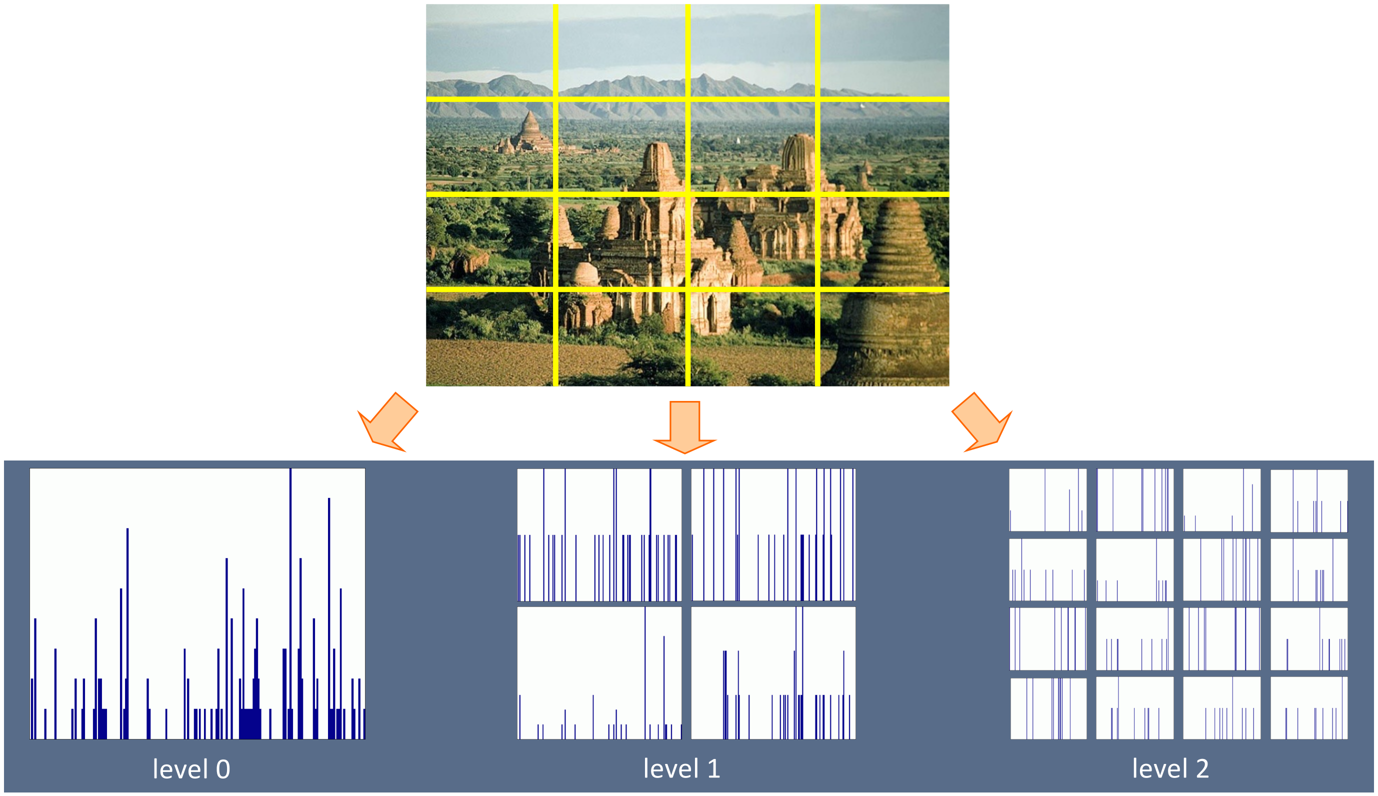
J. Hays and A. Efros. [Scene Completion Using Millions of Photographs](). SIGGRAPH 2007

Appearance-based scene representation: Bag of features



Csurka et al. (2004), Willamowski et al. (2005), Grauman & Darrell (2005), Sivic et al. (2003, 2005)

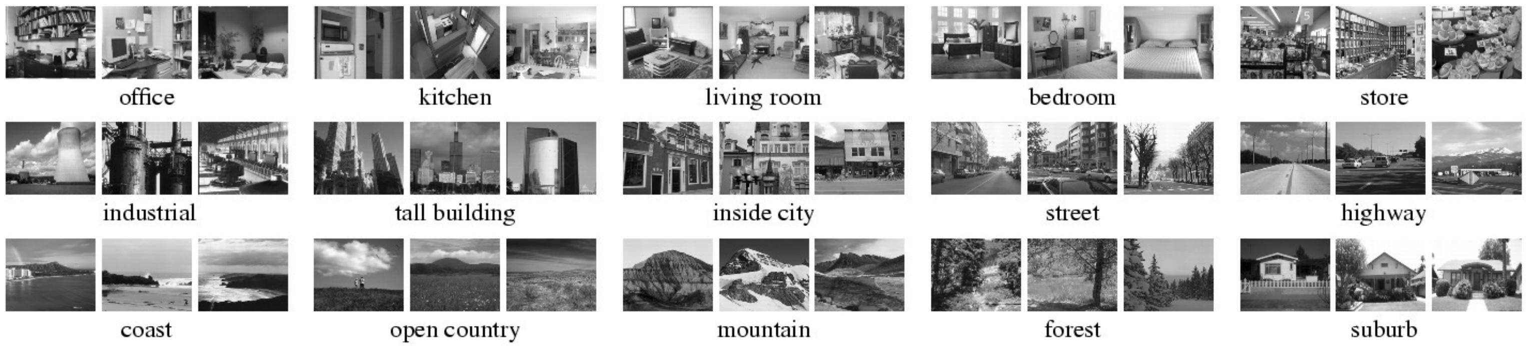
Spatial pyramids



Lazebnik, Schmid & Ponce (2006)

Spatial pyramids

15-category scene dataset

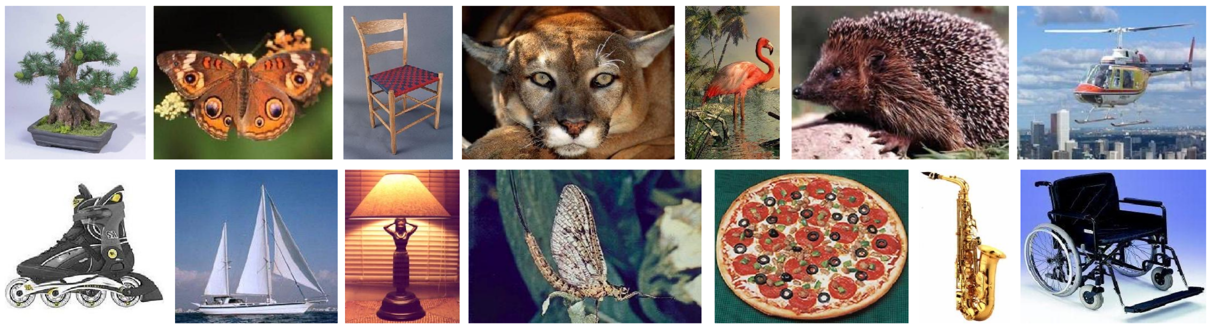


Multi-class classification results
(100 training images per class)

	Weak features (vocabulary size: 16)		Strong features (vocabulary size: 200)	
Level	Single-level	Pyramid	Single-level	Pyramid
$0\ (1 \times 1)$	45.3 ± 0.5		72.2 ± 0.6	
$1\ (2 \times 2)$	53.6 ± 0.3	56.2 ± 0.6	77.9 ± 0.6	79.0 ± 0.5
$2\ (4 \times 4)$	61.7 ± 0.6	64.7 ± 0.7	79.4 ± 0.3	$\mathbf{81.1} \pm 0.3$
$3\ (8 \times 8)$	63.3 ± 0.8	$\mathbf{66.8} \pm 0.6$	77.2 ± 0.4	80.7 ± 0.3

Spatial pyramids

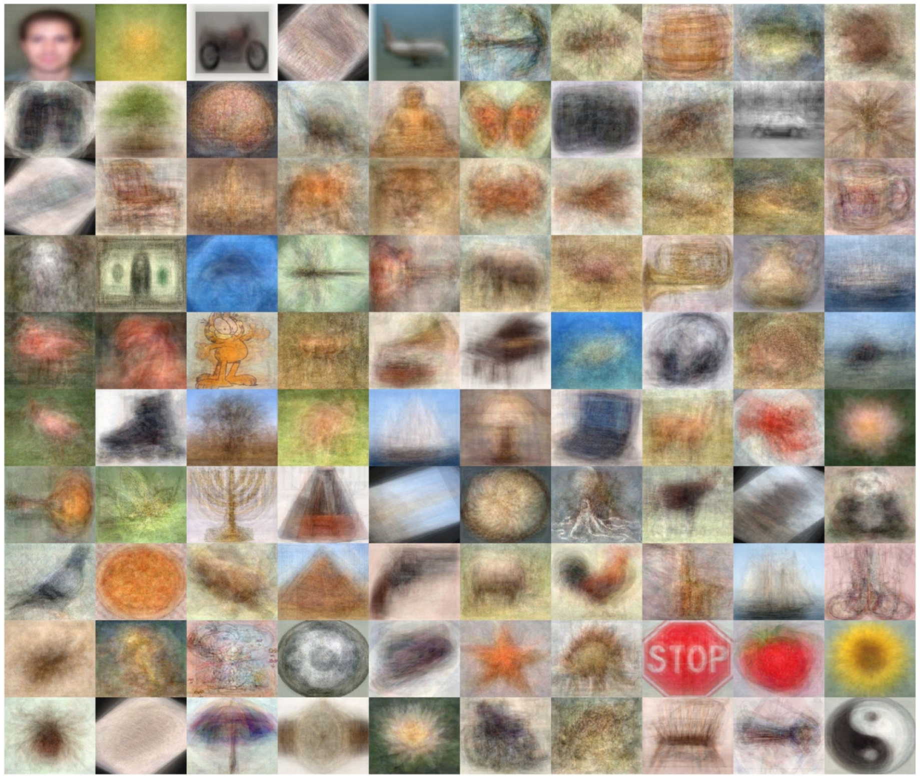
Multi-class classification results (30 training images per class)



Multi-class classification results (30 training images per class)

	Weak features (16)		Strong features (200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0	15.5 ±0.9		41.2 ±1.2	
1	31.4 ±1.2	32.8 ±1.3	55.9 ±0.9	57.0 ±0.8
2	47.2 ±1.1	49.3 ±1.4	63.6 ±0.9	**64.6** ±0.8
3	52.2 ±0.8	**54.0** ±1.1	60.3 ±0.9	64.6 ±0.7

Caltech-101 dataset



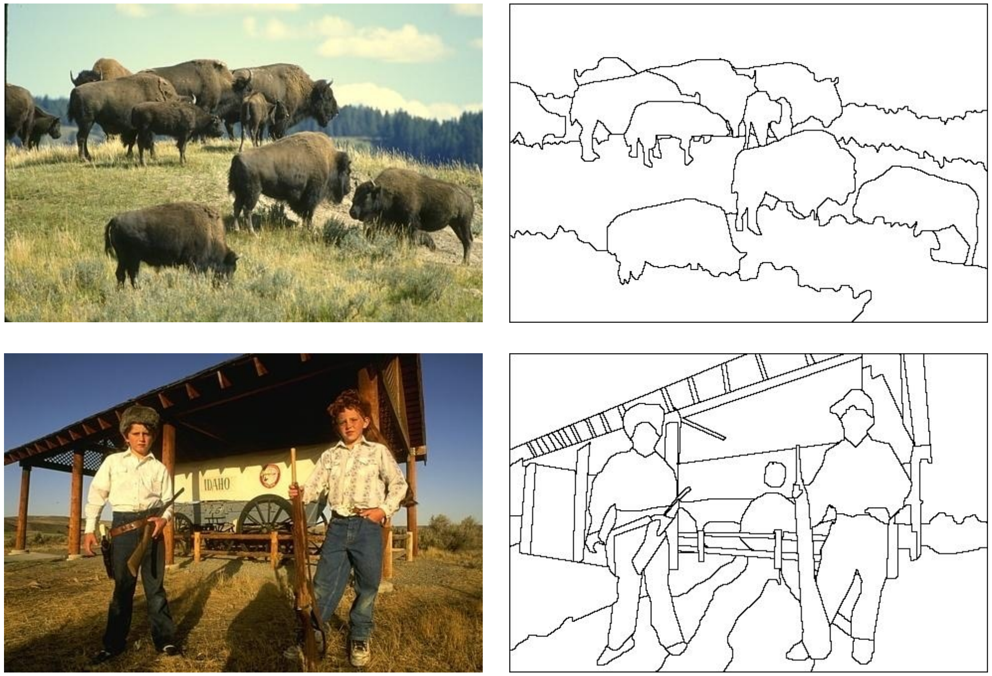
L. Fei-Fei et al. [Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories](), CVPR 2004 Workshop on Generative-Model Based Vision

http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Outline

- Object representations
  - 3D shape
  - 3D primitives
  - 2D appearance-based models
  - 2D part-based models (deformable templates)
  - CNNs
- Scene representations
  - Structured representations
  - Appearance-based representations
  - Bottom-up and top-down perceptual organization

Bottom-up perceptual organization

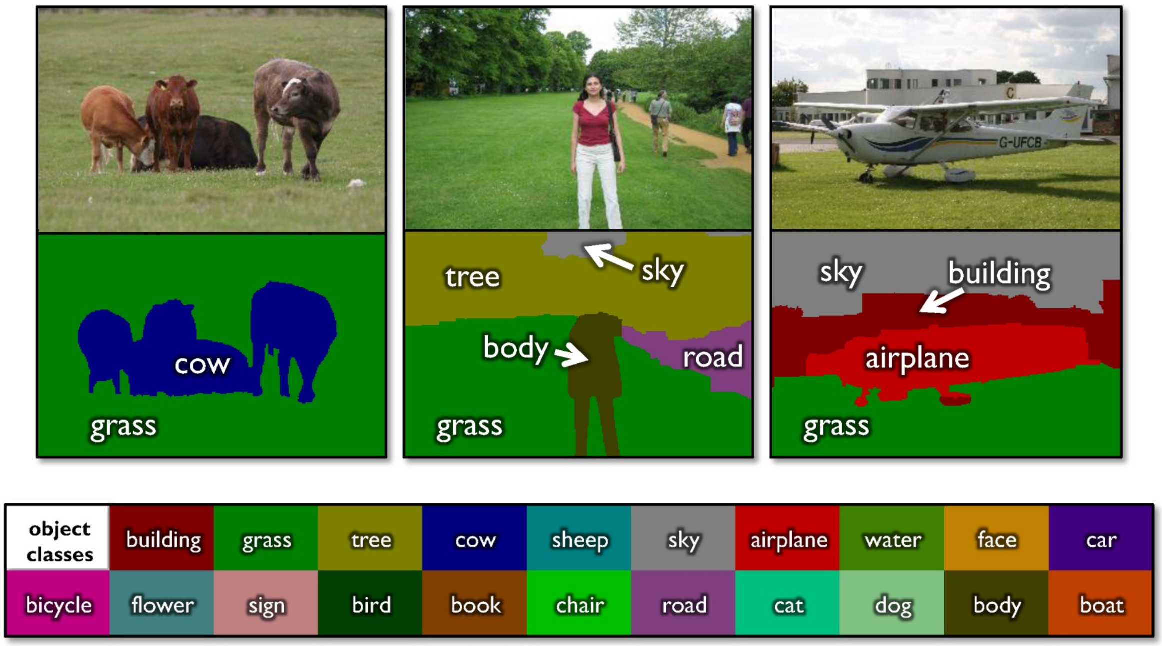


J. Malik et al. Contour and Texture Analysis for Image Segmentation. IJCV 2001

D. Martin et al. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. ICCV 2001

Top-down perceptual organization: Semantic segmentation

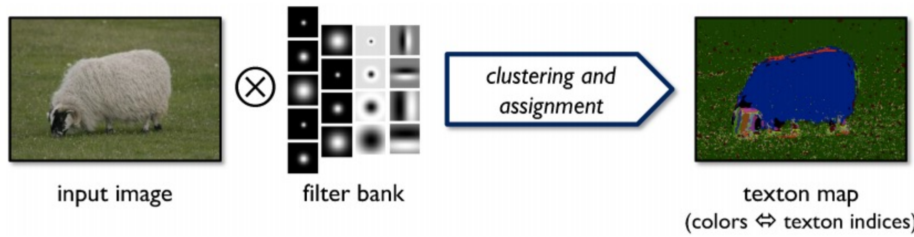
MSRC Dataset (2006)



J. Shotton et al. [TextonBoost: Joint Appearance, Shape And Context Modeling For Multi-class Object Recognition And Segmentation](). ECCV 2006

Top-down perceptual organization: Semantic segmentation

$$\log P(\overset{\text{Label field}}{\mathbf{c}}\,|\,\overset{\text{Image}}{\mathbf{x}}, \overset{\text{Model parameters}}{\boldsymbol{\theta}}) =$$



$$\boxed{\sum_i \overbrace{\psi_i(c_i, \mathbf{x}; \boldsymbol{\theta}_\psi)}^{\text{texture-layout}} + \overbrace{\pi(c_i, x_i; \boldsymbol{\theta}_\pi)}^{\text{color}} + \overbrace{\lambda(c_i, i; \boldsymbol{\theta}_\lambda)}^{\text{location}}}$$

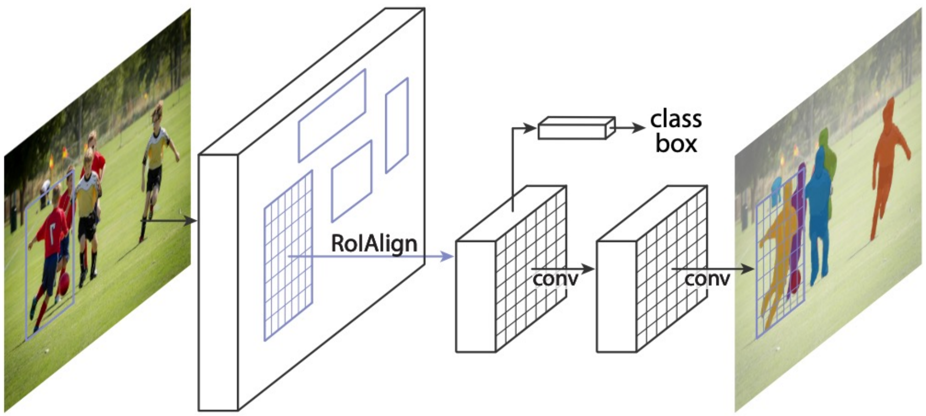
Local data term

$$+ \boxed{\sum_{(i,j)\in\mathcal{E}} \overbrace{\phi(c_i, c_j, \mathbf{g}_{ij}(\mathbf{x}); \boldsymbol{\theta}_\phi)}^{\text{edge}}}$$

Smoothing term

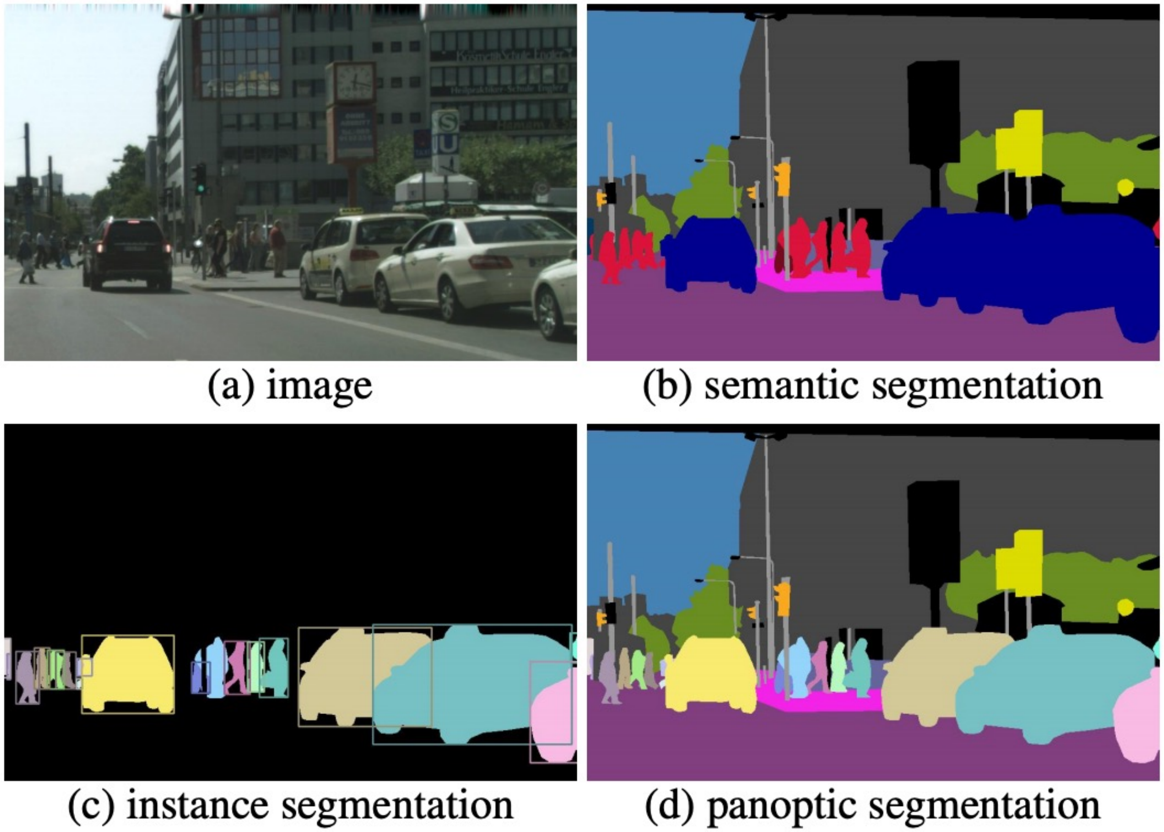
J. Shotton et al. [TextonBoost: Joint Appearance, Shape And Context Modeling For Multi-class Object Recognition And Segmentation](). ECCV 2006

Semantic segmentation today: Mask R-CNN



K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN, ICCV 2017 (Best Paper Award)

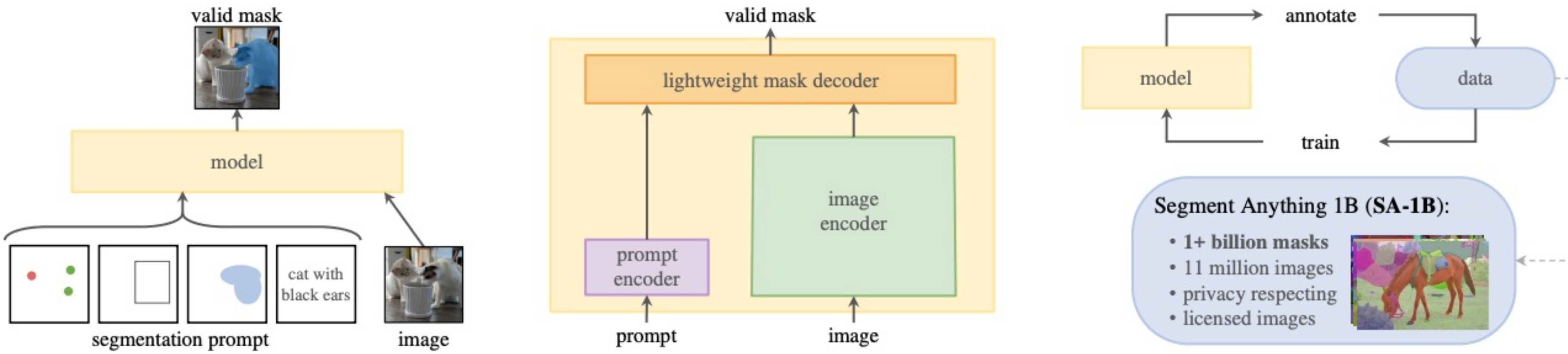
Panoptic segmentation



(a) image (b) semantic segmentation (c) instance segmentation (d) panoptic segmentation

A. Kirillov et al. Panoptic segmentation. CVPR 2019

Segment Anything Model



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

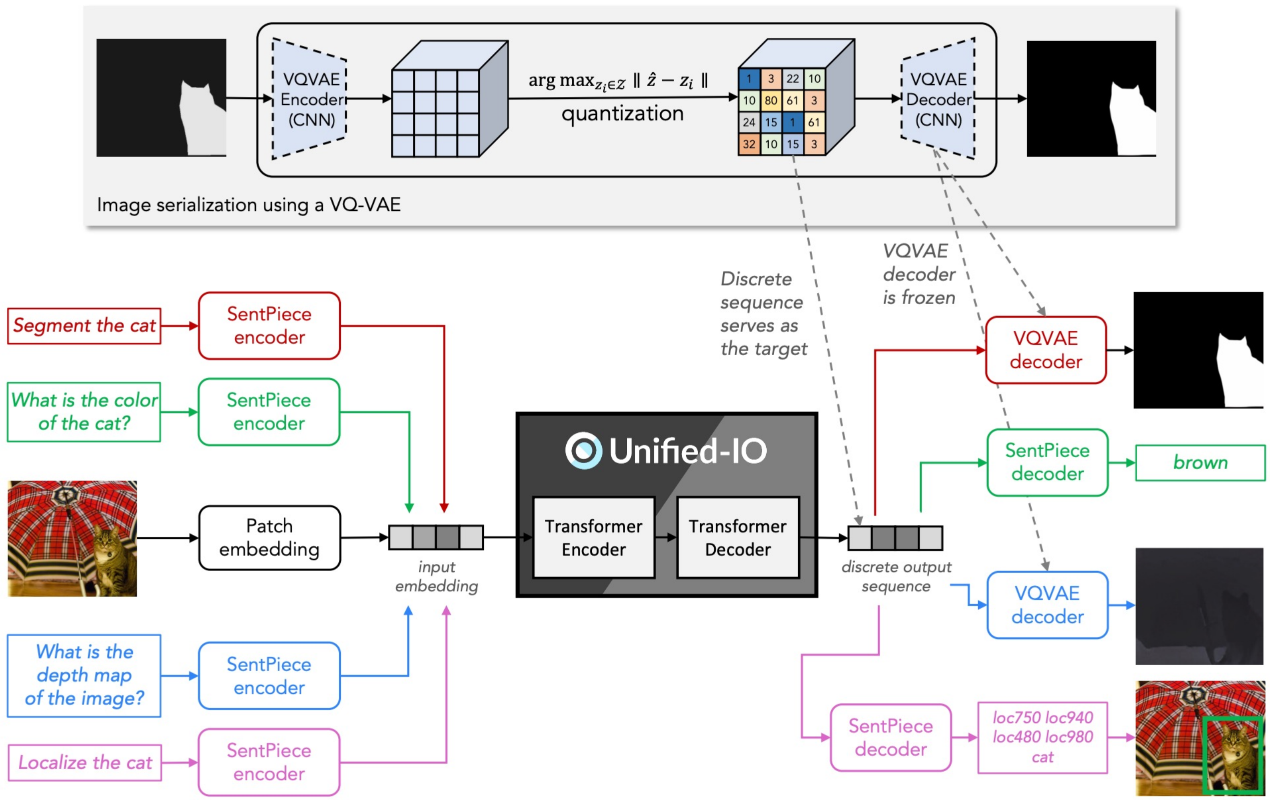
Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data* engine for collecting SA-1B, our dataset of over 1 billion masks.

A. Kirillov et al. [Segment Anything](). arXiv 2023
https://segment-anything.com/

Outline

- Object representations
    - 3D shape
    - 3D primitives
    - 2D appearance-based models
    - 2D part-based models (deformable templates)
    - CNNs
- Scene representations
    - Structured representations
    - Appearance-based representations
    - Bottom-up and top-down perceptual organization
- Trends

Unified representations for many tasks



J. Lu et al. [Unified-IO: A Unified Model for Vision, Language, and Multi-Modal Tasks.](#) ICLR 2023

Unified representations for many tasks

T. Gupta and A. Kembhavi. [Visual Programming: Compositional visual reasoning without training](). CVPR 2023

Generative models

Diffusion models

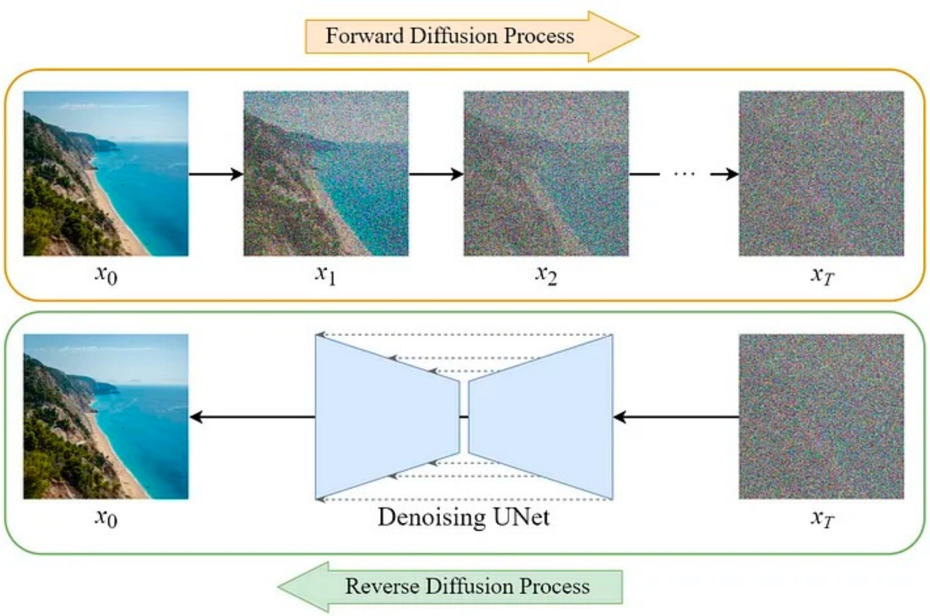
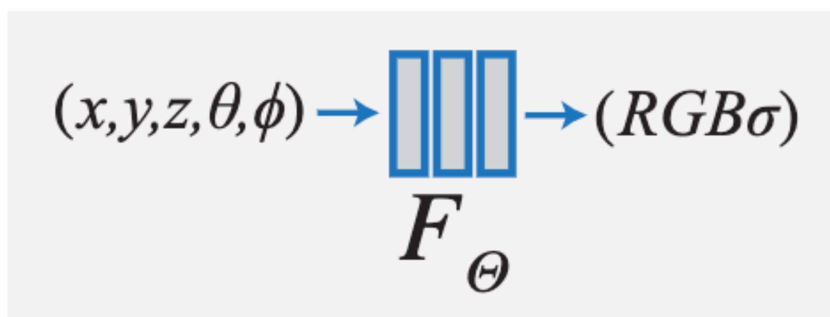


Figure source

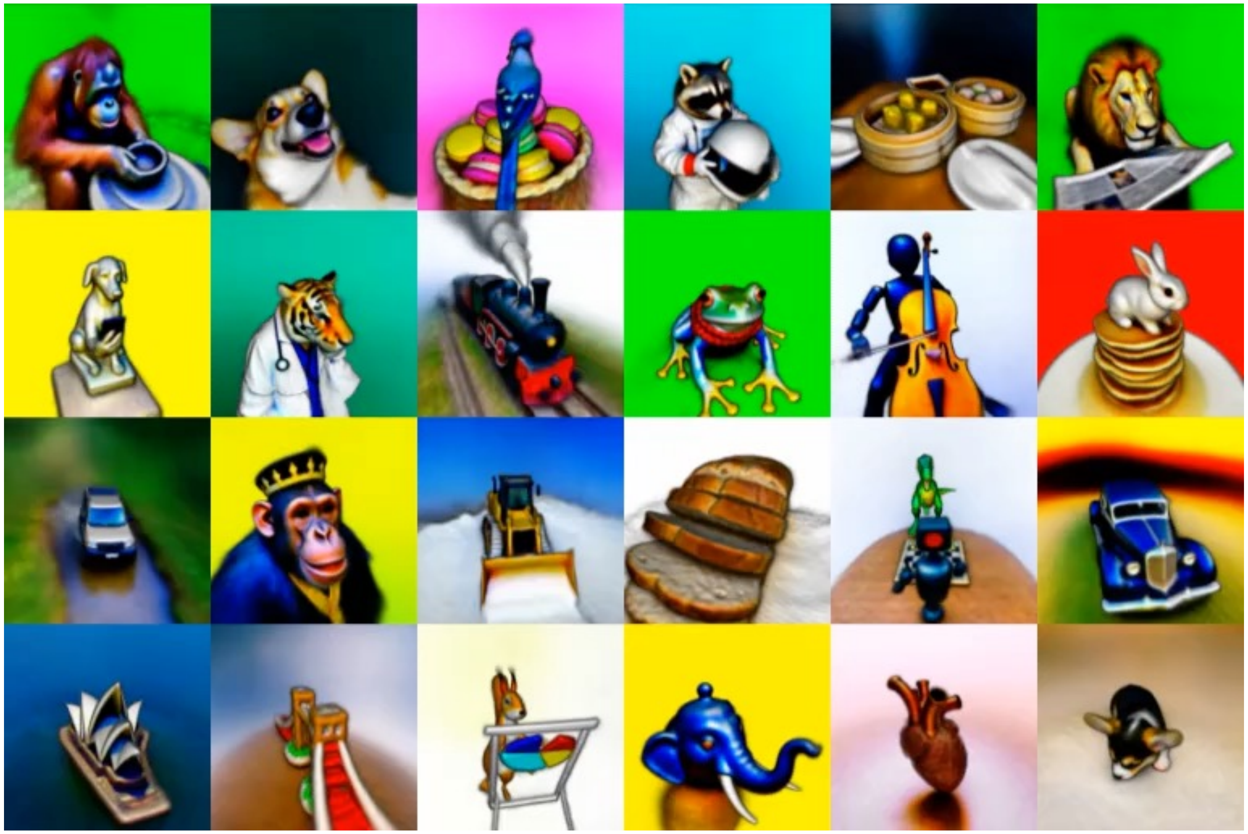
C. Saharia et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. NeurIPS 2022

Neural 3D representations: NERFs



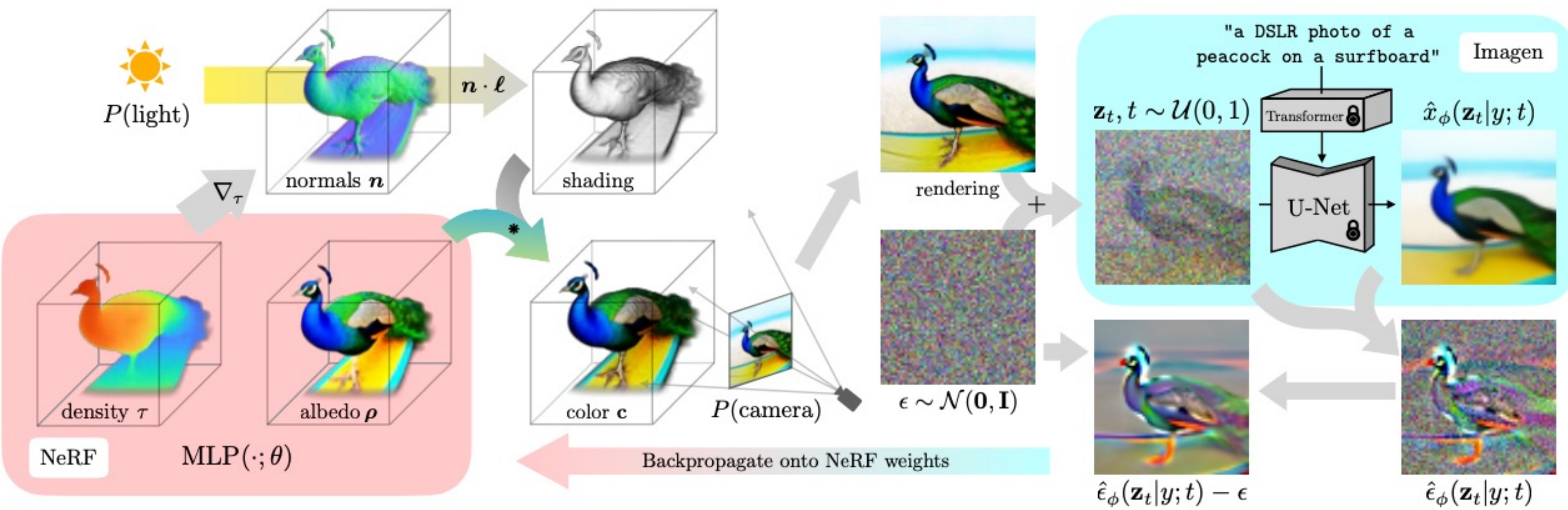
B. Mildenhall et al., [Representing Scenes as Neural Radiance Fields for View Synthesis](), ECCV 2020

Connecting 2D to 3D: DreamFusion



B. Poole et al. DreamFusion: Text-to-3D using 2D Diffusion. arXiv 2022

Connecting 2D to 3D: DreamFusion



B. Poole et al. DreamFusion: Text-to-3D using 2D Diffusion. arXiv 2022