Deep Learning for Computer Vision

History

Vineeth N Balasubramanian

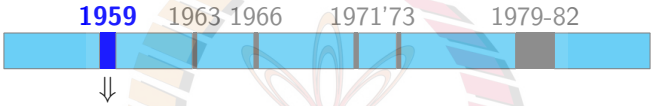
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Outline

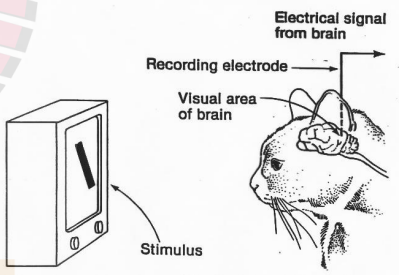
Disclaimer

- A history of the field as captured from multiple sources (including Szeliski's book and other sources credited on each page)
- A slightly biased history, as relevant to the topics we cover in this course. There is more to history in related topics (e.g. geometry-based vision, physics-based vision, image/video processing and compression, graphics, computational photography) not covered herein.
- A slight predisposition to work based on images, more than videos.

Early History[1]

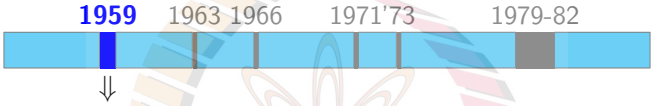


- David Hubel and Torsten Wiesel publish their work *"Receptive fields of single neurons in the cat's striate cortex"*
- Placed electrodes into primary visual cortex area of an anesthetized cat's brain
- Showed that simple and complex neurons exist, and that visual processing starts with simple structures such as oriented edges



[1] Credit: Rostyslav Demush, medium.com

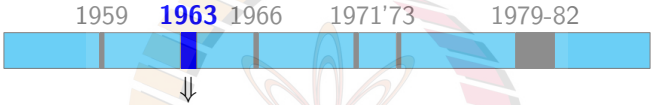
Early History[2]



- World's first digital image: Russell Kirsch and his colleagues develop an apparatus to transform images into number grids
- Image of Russell's infant son: grainy 5cm by 5cm photo, 30,976 pixels (176×176 array)
- Now stored in Portland Art Museum

[2] Credit: Rostyslav Demush, medium.com

Early History[3]



- Lawrence Roberts' PhD thesis: *"Machine Perception Of Three-Dimensional Solids"*
- Discussed extracting 3D information about solid objects from 2D photographs of line drawings
- Discussed issues such as camera transformations, perspective effects, and the rules and assumptions of depth perception

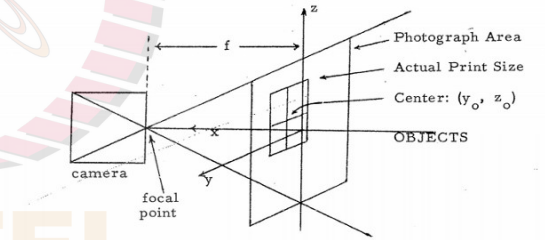
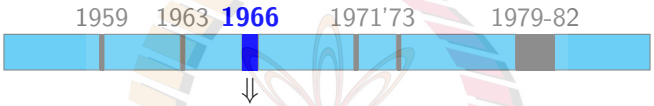


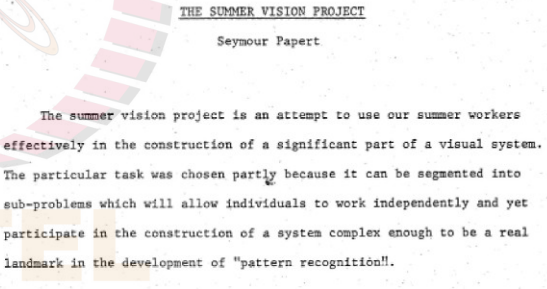
Figure 1: Camera Transformation

[3] Credit: Rostyslav Demush, medium.com

Early History[4]



- Seymour Papert (with Gerald Sussman) from MIT launched the *Summer Vision Project*
- Aimed to develop a platform to automatically segment background/foreground and extract non-overlapping objects from real-world images



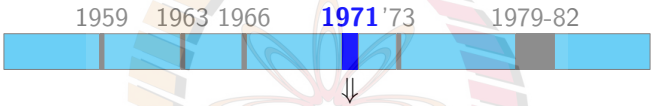
THE SUMMER VISION PROJECT

Seymour Papert

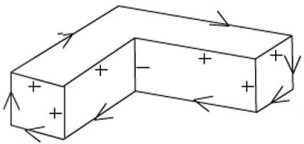
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

[4] Credit: Rostyslav Demush, medium.com

Early History[5]

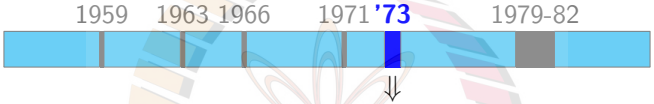


- Discern a shape in a line drawing by labeling lines as convex, concave, and occluded
  - David Huffman et al, *Impossible objects as nonsense sentences*, Machine Intelligence, 8:475-492, 1971
  - Max Clowes et al, *On seeing things*, Artificial Intelligence, 2:79-116, 1971

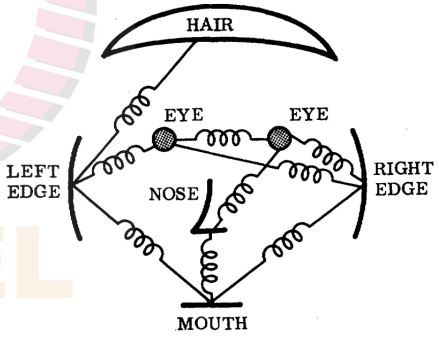


[5] Credit: Rostyslav Demush, medium.com

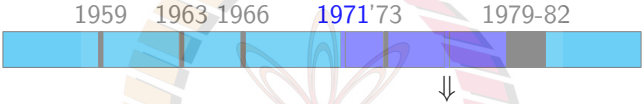
Early History



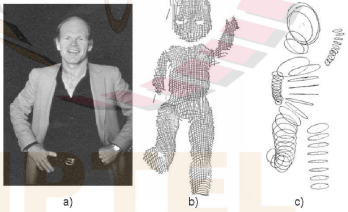
- *Pictorial Structures* model by Fischler and Elschlager
- Given a visual object's description, find the object in a photograph
- Part of the solution is specification of a descriptive scheme, and a metric on which to base the decision of "goodness" of matching or detection

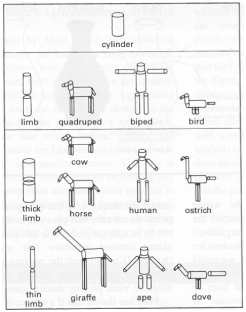


Early History

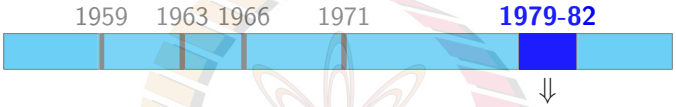


- Object recognition through shape understanding
  - Binford 1971, Generalized Cylinders
  - Marr and Nishihara 1978, Skeletons and Cylinders
- MIT's Artificial Intelligence Lab offers a "Machine Vision" course

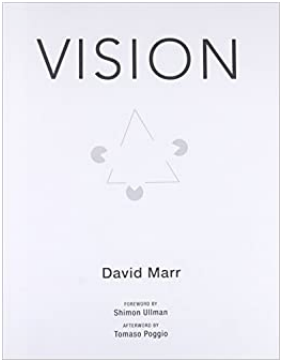




Early History[6]

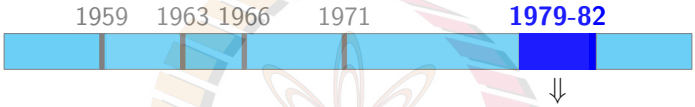


- David Marr, *Vision: A computational investigation into the human representation and processing of visual information*, 1982
- Established that vision is hierarchical
- Introduced a framework where low-level algorithms that detect edges, curves, corners, etc., are used to get high-level understanding of visual data



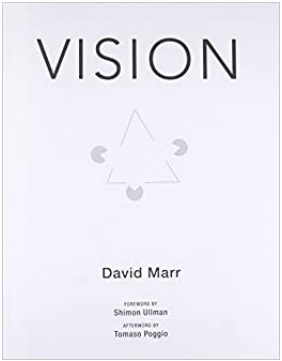
[6] Credit: Rostyslav Demush, medium.com

Early History[7]



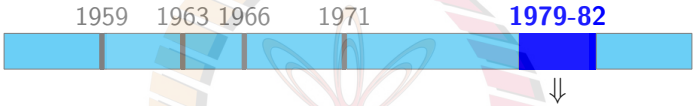
Marr's Representational Framework

- A primal sketch of an image, where edges, bars, boundaries etc., are represented
- A $2\frac{1}{2}$-D sketch representation where surfaces, information about depth, and discontinuities on an image are pieced together
- A 3D model that is hierarchically organized in terms of surface and volumetric primitives

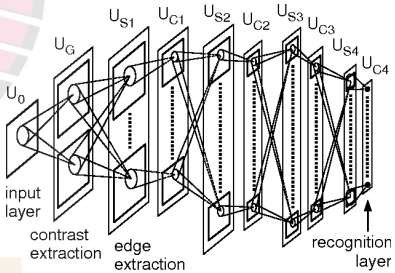


[7] Credit: Rostyslav Demush, medium.com

Early History[8]



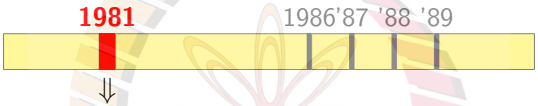
- Kunihiko Fukushima' *Neocognitron*, a self-organizing artificial network of simple and complex cells to recognize patterns, unaffected by position shifts
- The original ConvNet!
- Included convolutional layers with weight vectors (known as filters)



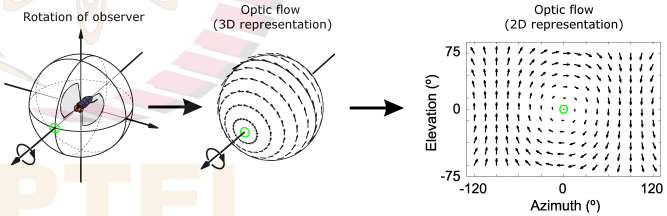
[8] Credit: Rostyslav Demush, medium.com

Outline

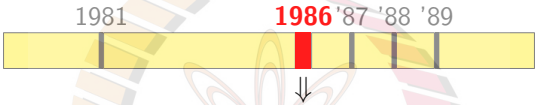
Towards Algorithms and Practice: Low-level Understanding



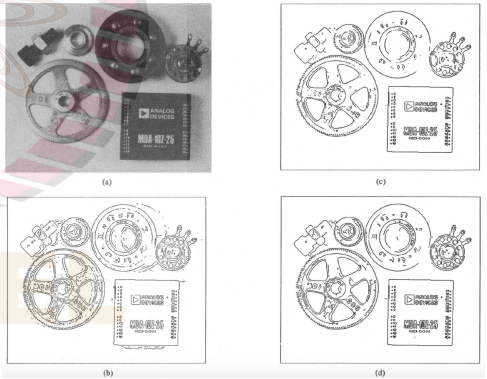
- **Optical Flow:** Horn and Schunck develop method to estimate the direction and speed of a moving object across two images
- Flow is formulated as a global energy functional which is minimized



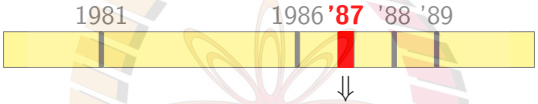
Towards Algorithms and Practice: Low-level Understanding



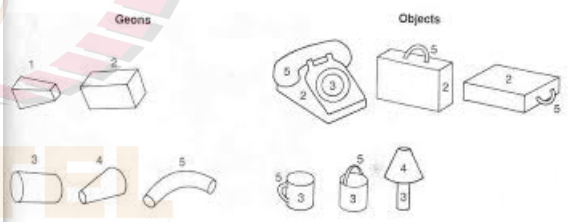
- **Canny Edge Detector:** Multi-stage edge detection operator, with a computational theory of edge detection
- Used calculus of variations to find the function that optimizes a given functional
- Well-defined method, simple to implement, became very popular for edge detection



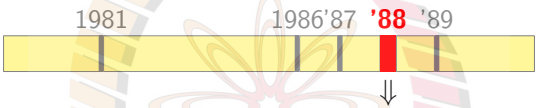
Towards Algorithms and Practice: Low-level Understanding



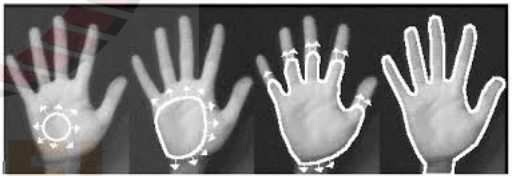
- **Recognition by Components Theory:** Proposed by Biederman
- Bottom-up process to explain object recognition
- Object's component parts: *geons*, based on basic 3-dimensional shapes (cylinders, cones, etc.) assembled to form the object



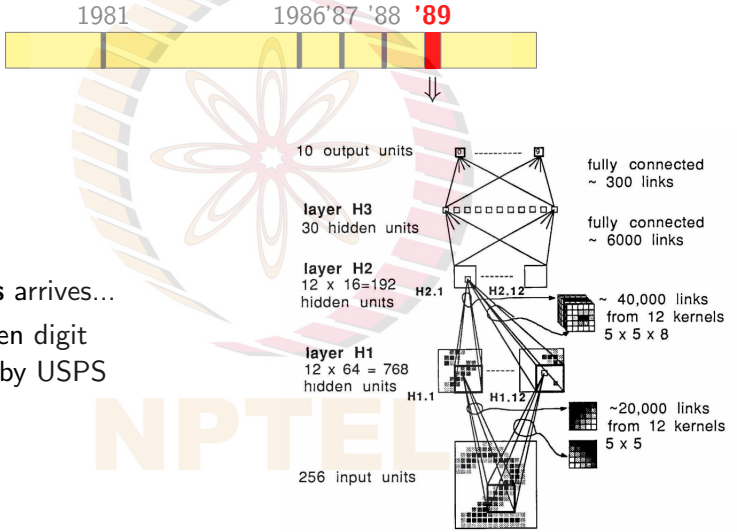
Towards Algorithms and Practice: Low-level Understanding



- **Snakes** or **active contour models** delineate an object outline from a possibly noisy 2D image
- Widely used in applications like object tracking, shape recognition, segmentation, edge detection and stereo matching

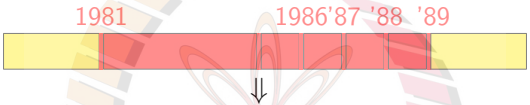


Towards Algorithms and Practice: Low-level Understanding



- **Backprop for CNNs** arrives...
- Applied to handwritten digit recognition provided by USPS

Towards Algorithms and Practice: Low-level Understanding



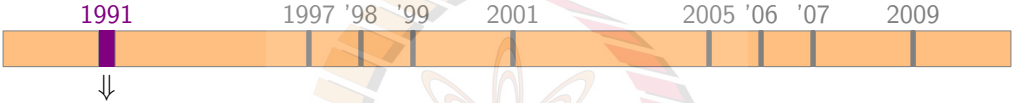
- Image Pyramids[9]
- Scale-space Processing[10]
- Wavelets[11]

- Shape-from-X[12]
- Variational Optimization methods[13]
- Markov Random Fields[14]

[9] Burt and Adelson, 1983
[10] Witkin, 1984
[11] Mallat, 1989
[12] Pentland, 1984; Blake et al, 1985
[13] Poggio et al, 1985
[14] Geman and Geman, 1985

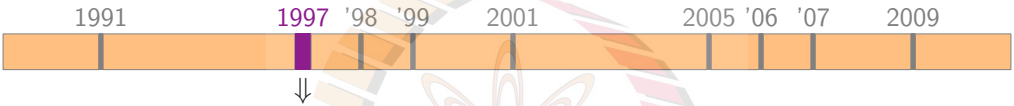
Outline

Towards Algorithms and Practice: Next-level Understanding



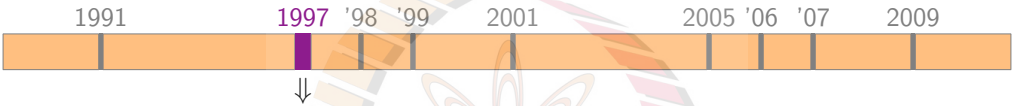
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)

Towards Algorithms and Practice: Next-level Understanding



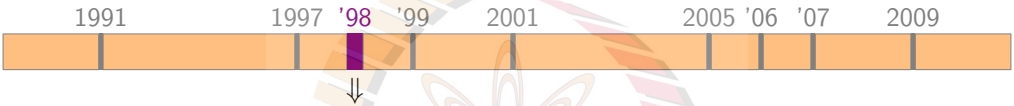
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)

Towards Algorithms and Practice: Next-level Understanding



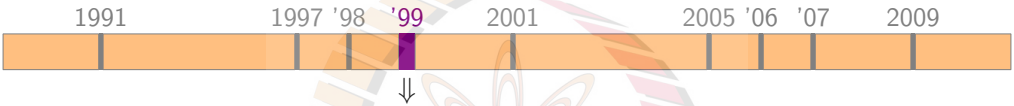
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)

Towards Algorithms and Practice: Next-level Understanding



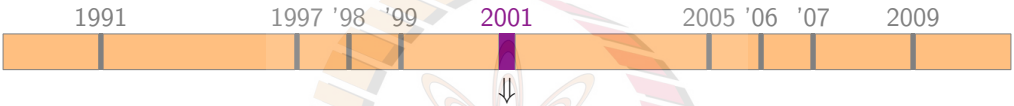
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)

Towards Algorithms and Practice: Next-level Understanding



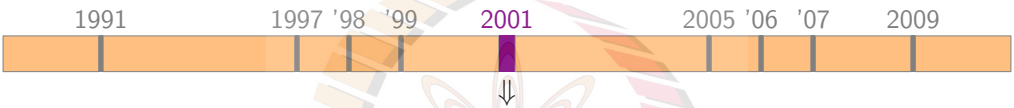
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)

Towards Algorithms and Practice: Next-level Understanding



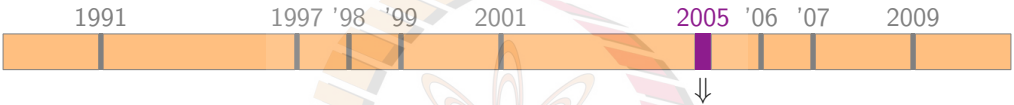
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)
- **Viola-Jones face detection** (Viola & Jones, 2001)

Towards Algorithms and Practice: Next-level Understanding



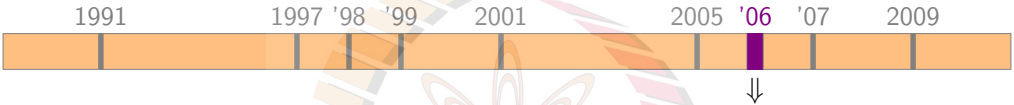
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)
- **Viola-Jones face detection** (Viola & Jones, 2001)
- **Conditional Random Fields** (Lafferty et al, 2001)

Towards Algorithms and Practice: Next-level Understanding



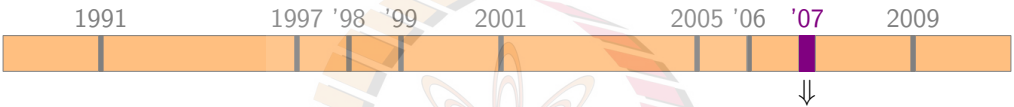
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)
- **Viola-Jones face detection** (Viola & Jones, 2001)
- **Conditional Random Fields** (Lafferty et al, 2001)
- **Pictorial structures** revisited (Felzenszwalb & Huttenlocher, 2005)

Towards Algorithms and Practice: Next-level Understanding



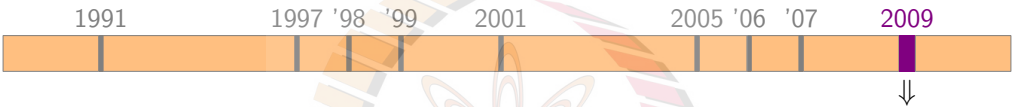
- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)
- **Viola-Jones face detection** (Viola & Jones, 2001)
- **Conditional Random Fields** (Lafferty et al, 2001)
- **Pictorial structures** revisited (Felzenszwalb & Huttenlocher, 2005)
- **PASCAL VOC** arrives; Scene/panorama/location recognition methods grow

Towards Algorithms and Practice: Next-level Understanding



- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)
- **Viola-Jones face detection** (Viola & Jones, 2001)
- **Conditional Random Fields** (Lafferty et al, 2001)
- **Pictorial structures** revisited (Felzenszwalb & Huttenlocher, 2005)
- **PASCAL VOC** arrives; Scene/panorama/location recognition methods grow
- **Constellation models** (Fergus, Perona & Zisserman, 2007)

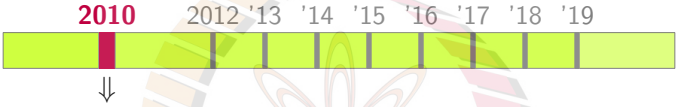
Towards Algorithms and Practice: Next-level Understanding



- **Eigenfaces for face recognition** (Turk & Pentland, 1991)
- **Computational theories of object recognition** (Edelman, 1997)
- **Perceptual grouping, Normalized cuts** (Shi & Malik, 1997)
- **Particle filters**, **Mean shift** for tracking (Liu & Chen, 1998)(Cheng, 1998)
- **SIFT** (Lowe, 1999) (Lowe, 2004)
- **Viola-Jones face detection** (Viola & Jones, 2001)
- **Conditional Random Fields** (Lafferty et al, 2001)
- **Pictorial structures** revisited (Felzenszwalb & Huttenlocher, 2005)
- **PASCAL VOC** arrives; Scene/panorama/location recognition methods grow
- **Constellation models** (Fergus, Perona & Zisserman, 2007)
- **Deformable part models** (Felzenszwalb et al, 2009)

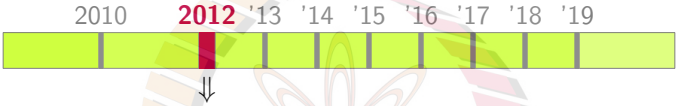
Outline

The Deep Learning Era

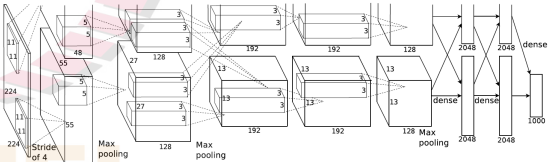


- ImageNet arrives

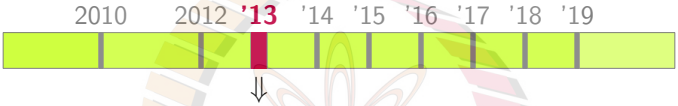
The Deep Learning Era



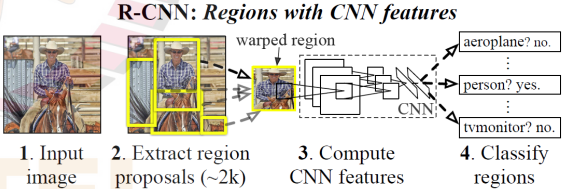
- ImageNet arrives
- AlexNet wins the ImageNet challenge



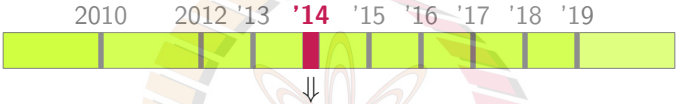
The Deep Learning Era



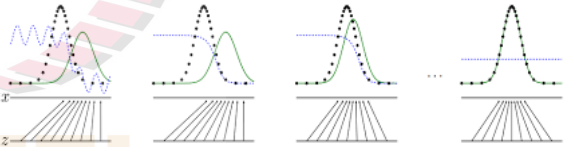
- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins



The Deep Learning Era

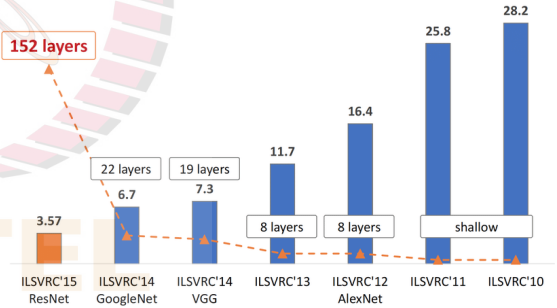


- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins
- InceptionNet, VGG models arrive; Human Pose Estimation CNNs; Deep generative models: GANs, VAEs



The Deep Learning Era

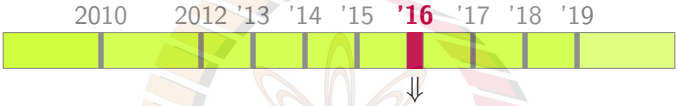
- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins
- InceptionNet, VGG models arrive; Human Pose Estimation CNNs; Deep generative models: GANs, VAEs
- ResNet arrives; CNNs match human performance on ImageNet



The Deep Learning Era

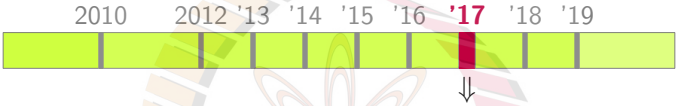
- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins
- InceptionNet, VGG models arrive; Human Pose Estimation CNNs; Deep generative models: GANs, VAEs
- ResNet arrives; CNNs match human performance on ImageNet
- FCN, SegNet and U-Net for semantic segmentation; COCO dataset arrives; VQA dataset arrives

The Deep Learning Era



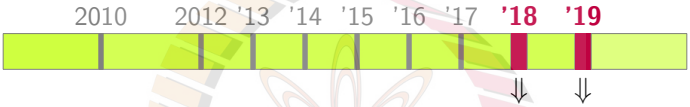
- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins
- InceptionNet, VGG models arrive; Human Pose Estimation CNNs; Deep generative models: GANs, VAEs
- ResNet arrives; CNNs match human performance on ImageNet
- FCN, SegNet and U-Net for semantic segmentation; COCO dataset arrives; VQA dataset arrives
- YOLO and SSD for object detection; Cityscapes dataset arrives, Visual Genome dataset arrives

The Deep Learning Era



- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins
- InceptionNet, VGG models arrive; Human Pose Estimation CNNs; Deep generative models: GANs, VAEs
- ResNet arrives; CNNs match human performance on ImageNet
- FCN, SegNet and U-Net for semantic segmentation; COCO dataset arrives; VQA dataset arrives
- YOLO and SSD for object detection; Cityscapes dataset arrives, Visual Genome dataset arrives
- Scene graph generation models

The Deep Learning Era



- ImageNet arrives
- AlexNet wins the ImageNet challenge
- A CNN, variant of ZFNet, wins ImageNet challenge; R-CNNs for object detection arrive; Understanding CNNs begins
- InceptionNet, VGG models arrive; Human Pose Estimation CNNs; Deep generative models: GANs, VAEs
- ResNet arrives; CNNs match human performance on ImageNet
- FCN, SegNet and U-Net for semantic segmentation; COCO dataset arrives; VQA dataset arrives
- YOLO and SSD for object detection; Cityscapes dataset arrives, Visual Genome dataset arrives
- Scene graph generation models
- VCR dataset, Panoptic segmentation
- ...

History of Applications[15]

- **1970s:** Optical Character Recognition (OCR)
- **1980s:** Machine vision, Smart cameras
- **1990s:** Machine vision in manufacturing environments, Biometrics, Medical imaging, Recording devices, Video surveillance
- **2000s:** More biometrics, Better medical imaging, Object/Face detection, Autonomous navigation, Google Goggles, Vision on social media
- **2010s:** Everywhere around us

[15] See https://www.phase1vision.com/resources/timeline for a longer historical timeline

Homework

Readings

- Chapter 1, Szeliski, *Computer Vision: Algorithms and Applications* (Sections 1.1 and 1.2, in particular)
- Other links provided on respective slides, especially
  - Rostyslav Demush, A Brief History of Computer Vision (and Convolutional Neural Networks)

Relevant References I

David Huffman. "Impossible objects as nonsense sentences | AMiner". In: *Machine Intelligence* 8 (1971), pp. 475–492.

M.A. Fischler and R.A. Elschlager. "The Representation and Matching of Pictorial Structures". In: *IEEE Transactions on Computers* C-22.1 (Jan. 1973), pp. 67–92.

Agin and Binford. "Computer Description of Curved Objects". In: *IEEE Transactions on Computers* C-25.4 (Apr. 1976), pp. 439–449.

D. Marr, H. K. Nishihara, and Sydney Brenner. "Representation and recognition of the spatial organization of three-dimensional shapes". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 200.1140 (Feb. 1978), pp. 269–294.

Kunihiko Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36.4 (Apr. 1980), pp. 193–202.

Berthold K. P. Horn and Brian G. Schunck. "Determining optical flow". In: *Artificial Intelligence* 17.1 (Aug. 1981), pp. 185–203.

Relevant References II

P. Burt and E. Adelson. "The Laplacian Pyramid as a Compact Image Code". In: *IEEE Transactions on Communications* 31.4 (Apr. 1983), pp. 532–540.

Stuart Geman and Donald Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6 (Nov. 1984), pp. 721–741.

A. Witkin. "Scale-space filtering: A new approach to multi-scale description". In: *ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 9. Mar. 1984, pp. 150–153.

Tomaso Poggio, Vincent Torre, and Christof Koch. "Computational vision and regularization theory". In: *Nature* 317.6035 (Sept. 1985), pp. 314–319.

John Canny. "A Computational Approach to Edge Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (Nov. 1986), pp. 679–698.

Irving Biederman. "Recognition-by-components: a theory of human image understanding". In: *Psychological Review* 94.2 (Apr. 1987), pp. 115–147.

Relevant References III

L. Sirovich and M. Kirby. "Low-dimensional procedure for the characterization of human faces". In: *JOSA A* 4.3 (Mar. 1987), pp. 519–524.

Michael Kass, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models". In: *International Journal of Computer Vision* 1.4 (Jan. 1988), pp. 321–331.

Y. LeCun et al. "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1.4 (Dec. 1989), pp. 541–551.

S.G. Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.7 (July 1989), pp. 674–693.

M.A. Turk and A.P. Pentland. "Face recognition using eigenfaces". In: *1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings*. June 1991, pp. 586–591.

Yizong Cheng. "Mean shift, mode seeking, and clustering". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.8 (Aug. 1995), pp. 790–799.

Relevant References IV

Pierre Del Moral. "Nonlinear filtering: Interacting particle resolution". In: *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* 325.6 (Sept. 1997), pp. 653–658.

Jianbo Shi and Jitendra Malik. "Normalized Cuts and Image Segmentation". In: *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*. CVPR '97. USA: IEEE Computer Society, June 1997, p. 731.

Jun S. Liu and Rong Chen. "Sequential Monte Carlo Methods for Dynamic Systems". In: *Journal of the American Statistical Association* 93.443 (Sept. 1998), pp. 1032–1044.

D.G. Lowe. "Object recognition from local scale-invariant features". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. Sept. 1999, 1150–1157 vol.2.

Jianbo Shi and J. Malik. "Normalized cuts and image segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8 (Aug. 2000), pp. 888–905.

Relevant References V

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data". In: *Proceedings of the Eighteenth International Conference on Machine Learning*. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., June 2001, pp. 282–289.

Kunihiko Fukushima. "Neocognitron for handwritten digit recognition". In: *Neurocomputing* 51 (Apr. 2003), pp. 161–180.

David G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". In: *International Journal of Computer Vision* 60.2 (Nov. 2004), pp. 91–110.

Paul Viola and Michael J. Jones. *Robust Real-Time Face Detection*. May 2004. URL: https://doi.org/10.1023/B:VISI.0000013087.49260.fb (visited on 04/08/2020).

Pedro F. Felzenszwalb and Daniel P. Huttenlocher. "Pictorial Structures for Object Recognition". In: *International Journal of Computer Vision* 61.1 (Jan. 2005), pp. 55–79.

Relevant References VI

R. Fergus, P. Perona, and A. Zisserman. "Weakly Supervised Scale-Invariant Learning of Models for Visual Recognition". In: *International Journal of Computer Vision* 71.3 (Mar. 2007), pp. 273–303.

Pedro F. Felzenszwalb et al. "Object Detection with Discriminatively Trained Part-Based Models". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.9 (Sept. 2010), pp. 1627–1645.

Richard Szeliski. *Computer Vision: Algorithms and Applications*. Texts in Computer Science. London: Springer-Verlag, 2011.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks". In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'12. Lake Tahoe, Nevada: Curran Associates Inc., Dec. 2012, pp. 1097–1105.

Weijun Wang and Ramakant Nevatia. "Robust Object Tracking Using Constellation Model with Superpixel". In: *Computer Vision – ACCV 2012*. Ed. by Kyoung Mu Lee et al. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 191–204.

Relevant References VII

Ross Girshick et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 580–587.

Ian J. Goodfellow et al. "Generative adversarial nets". In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'14. Montreal, Canada: MIT Press, Dec. 2014, pp. 2672–2680.

Diederik P. Kingma. "Auto-Encoding Variational Bayes.". In: (2014).

Alexander Toshev and Christian Szegedy. "DeepPose: Human Pose Estimation via Deep Neural Networks". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. June 2014, pp. 1653–1660.

Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv:1409.1556 [cs]* (Apr. 2015).

Relevant References VIII

Christian Szegedy et al. "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 1–9.

Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016, pp. 770–778.

Vidolab. *Computer Vision History: Milestones and Breakthroughs*. Apr. 2020. URL: https://computer-vision-ai.com/blog/computer-vision-history/ (visited on 04/06/2020).

Rostyslav Demush. *A Brief History of Computer Vision (and Convolutional Neural Networks) | Hacker Noon*. URL: https://hackernoon.com/a-brief-history-of-computer-vision-and-convolutional-neural-networks-8fe8aacc79f3 (visited on 04/06/2020).

History of Machine Vision. URL: https://www.phase1vision.com/resources/timeline (visited on 04/12/2020).

ImageNet. URL: http://www.image-net.org/ (visited on 04/12/2020).

Relevant References IX

The MIT Press. *Robot Vision | The MIT Press*. URL: https://mitpress.mit.edu/books/robot-vision (visited on 04/07/2020).

pubmeddev and Edelman S. *Computational theories of object recognition. - PubMed - NCBI*. URL: https://www.ncbi.nlm.nih.gov/pubmed/21223932 (visited on 04/08/2020).

The PASCAL Visual Object Classes Challenge 2006. URL: http://host.robots.ox.ac.uk/pascal/VOC/voc2006/index.html (visited on 04/08/2020).

Zbigniew Zdziarski. *The Beginnings of Computer Vision*. URL: https://zbigatron.com/the-early-history-of-computer-vision/ (visited on 04/06/2020).