4F12 Course summary
Q&A/Discussion

We will start at **10:05 am** (to allow everyone to join from previous in-person lectures/kitesurfing/juicing etc.)

If you are reading this before 10:05 am, you may wish to apply the computer vision researcher algorithm:

If 5 mins spare: Read a classic text on computer vision/ML to get a sense of how things have evolved. *Recommendation:* Chapter 3 of the PDP book (1986)

If 3 mins spare: The latest MLPerf results are out (1st December 2021)! Check to see who has built the best accelerator....

If 1 min spare: Stay hydrated.

Focus of the session

- Give a high-level summary of the course
- Give suggestions for how to prepare for exams

Instructors: Ignas Budvytis and Samuel Albanie

Course summary

Introduction: Lecture 1

- Computer vision: what is it, why study it and how?
- Vision as an information processing task?
- Geometrical and statistical frameworks for vision?

Image Structure: Lectures 2 - 5

- Image intensities and structure
- 2D convolution with gaussians for low-pass filtering
- Edge detection, the aperture problem and corner detection
- Image pyramids, blob detection with band-pass filtering
- The SIFT feature descriptor for matching
- Characterising textures.

Projection: Lectures 6 - 9

- Orthographic projection
- Planar perspective projection, vanishing points and lines
- Homogeneous coordinates and the projection matrix
- Camera calibration, recovery of world position
- Weak perspective and the affine camera

Stereo vision and Multiple View Geometry: Lectures 10 - 11

- Recovery of depth by triangulation
- Epipolar geometry and the essential matrix
- The correspondence problem
- Structure from motion
- 3D shape examples from multiple view stereo and photometric stereo

Deep Learning for Computer Vision: Lectures 12 - 14

- Basic architectures for deep learning in computer vision
- Detection, classification and semantic segmentation
- Recognition, feature embedding and metric learning

Everything above the red dashed line is examinable

Everything below the red dashed line is not examinable

Guest Lecture from Matthew Jonson

- Fourier Feature Networks and Neural Volume RenderingPage

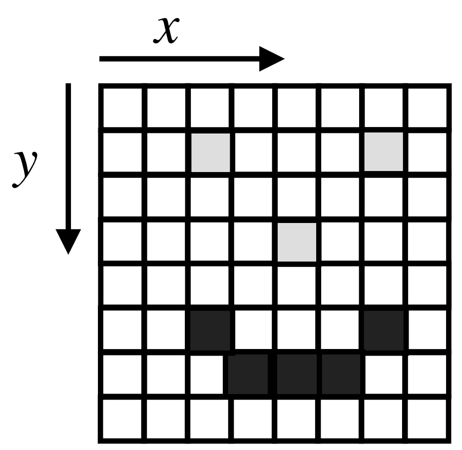
Deep Learning for Computer Vision: Lectures 15 - 16

- Transformers, scaling laws for computer vision, neural architecture search
- Self-supervised learning and pseudo-labelling

Image Structure 1
Image intensities, Edge detection, Convolution

Summary

- How to represent images as matrices
- Nuisance factors in pixel intensity data
- Data reduction in computer vision and Marr's hierarchy
- Image structures: featureless regions, edges and corners
- Edge detection in 1D (and how to do it quickly)
- Edge detection in 2D (and how to do it quickly)
- Implementation details (truncated summations; convolution)



Images as matrices

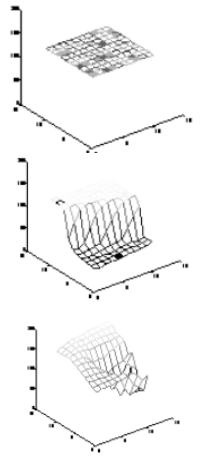
Geometric — Photometric

Nuisance factors

Raw Image data
↓
Generic salient features
↓
Interpretation

O(100+ Kbits/sec)

Data reduction

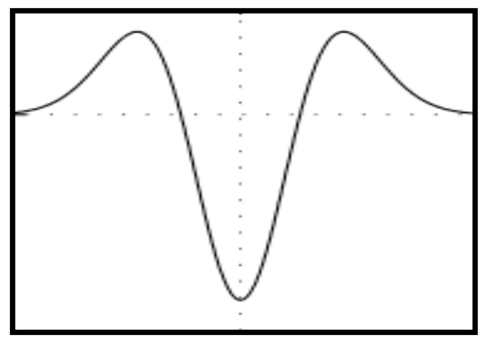


0D
Featureless region

1D
Edge

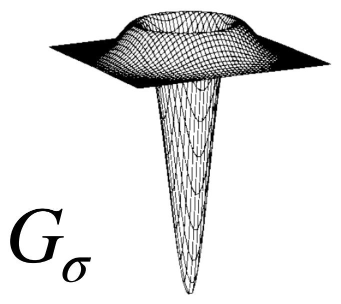
2D
Corner

Image structures



1D Edge Detection

$\nabla^2 G_\sigma$



2D Edge Detection

$$S(x, y) = \sum_{u=-n}^{n} \sum_{v=-n}^{n} G_\sigma(u, v) I(x-u, y-v)$$

Discrete convolution via
truncated summation

$$\frac{\partial S}{\partial x} = \frac{S(x+1, y) - S(x-1, y)}{2}$$

Differentiation via
convolution (Taylor)

Example questions

Differentiation and 1D edge detection: Q6 from 4F12 Image Structures Examples paper

Smoothing, Image Gradients and Discrete Convolutions: Past Paper 4F12 2019: Q1, parts (a) and (b)

Key idea

What is image structure, and how to find it?

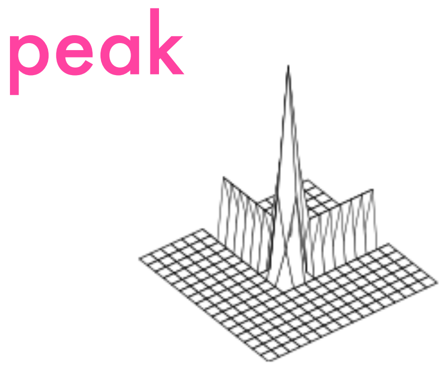
Image Structure 2

The Aperture Problem, Cross-correlation, Corner Detection and Scale

Summary

- The Aperture Problem
- Cross-correlation, and corners
- Corner Detection - Harris
- Scale invariance
- Blobs and band pass filtering
- Selecting characteristic scale
- Scale space for scale invariance
- Computational tricks for scale space

Aperture problem



Cross-correlation and corners

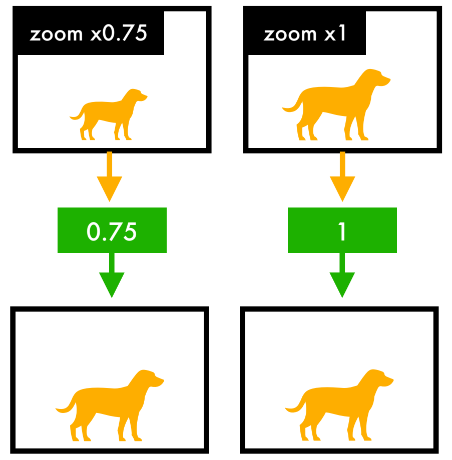
Arbitrary corner angles

Reduce noise

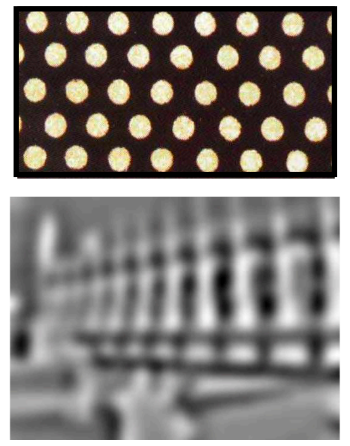
Eigenvalues -> structure

The det/trace trick

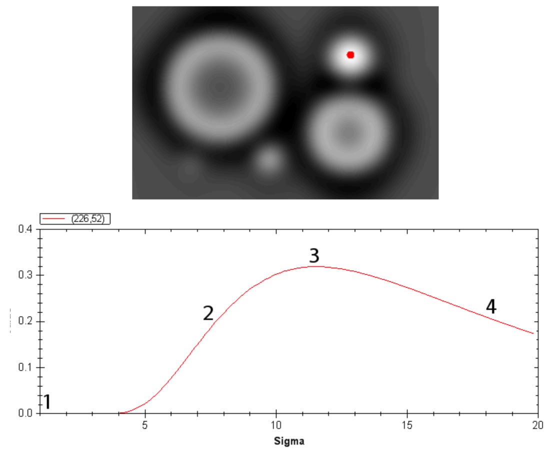
Harris Corners



Scale Invariance



Blobs & filtering



Selecting scale

$S(x, y, \sigma) = G(x, y, \sigma) \circledast I(x, y)$

Scale space for scale invariance

Sparse sampling

Image Pyramids

Incremental blurs

Difference of Gaussians

Computational tricks

Example questions

Harris corner detection: Past Paper 4F12 2018: Q1 (b)

Blobs and Image Pyramids: Past Paper 4F12 2016: Q1 (a) and (b)

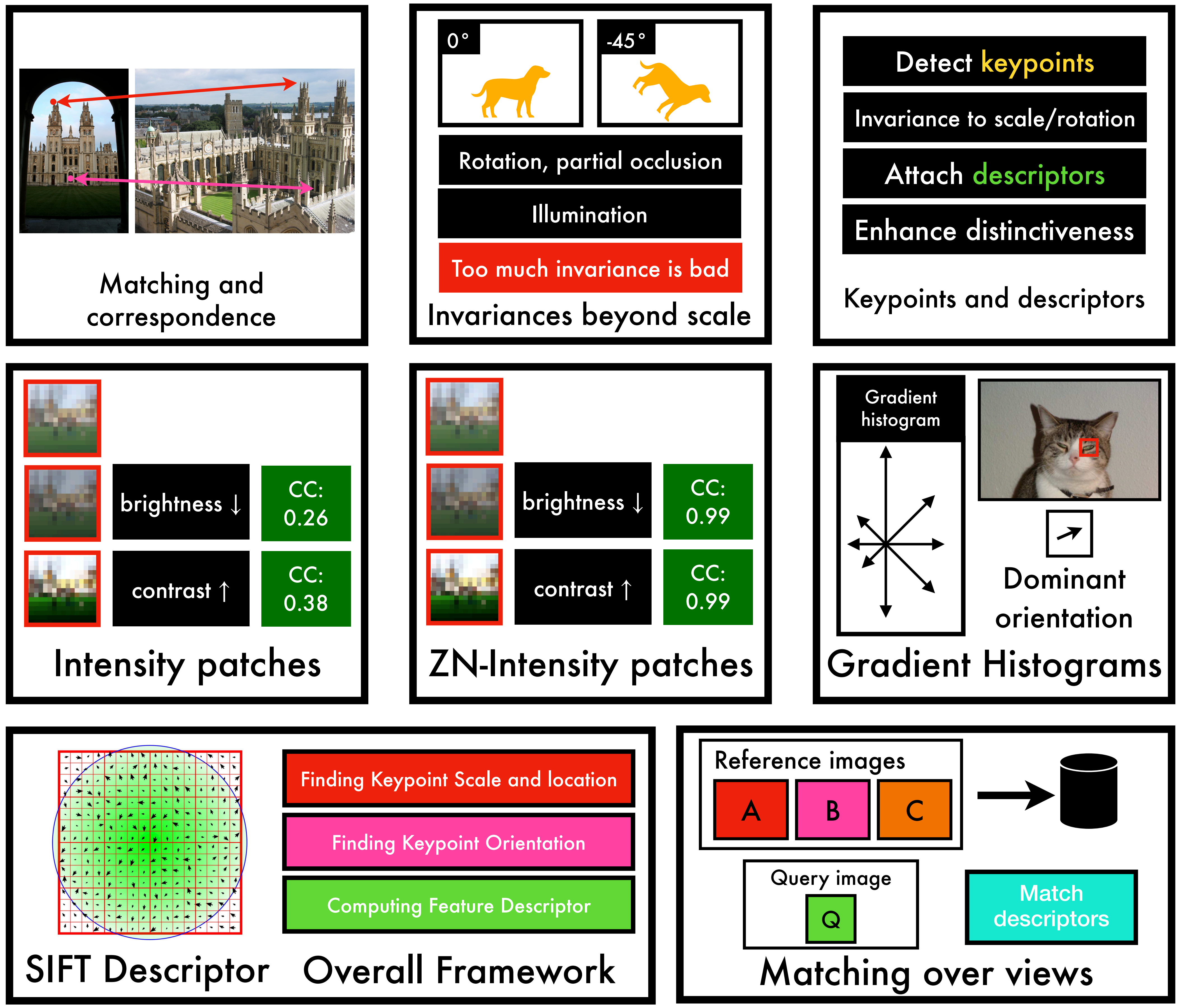
Key idea

Efficient, scale-invariant feature detection

Image Structure 3
Matching and correspondence, keypoints & descriptors, SIFT

Summary

- Matching and correspondence
- Invariance beyond scale: 3D viewpoint, partial occlusion, illumination
- Strategy: use keypoints *and* descriptors
- Descriptors: intensity patches
- Descriptor: zero-normalised intensity patches
- Histogram of Oriented Gradients & Dominant Orientations
- SIFT Descriptor
- The Keypoint and Descriptor Framework
- Matching over multiple views



Example questions

Normalised descriptors Past Paper 4F12 2015: Q1 (c)

SIFT descriptor/Matching: Past Paper 4F12 2021: Q1 (b)

Key idea

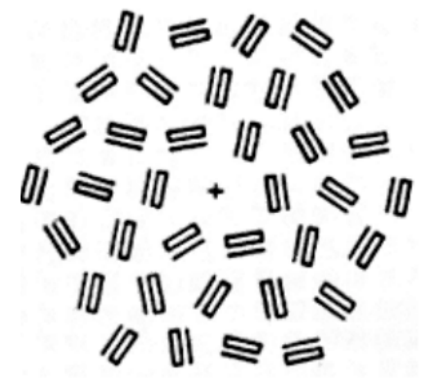
Matching over multiple views

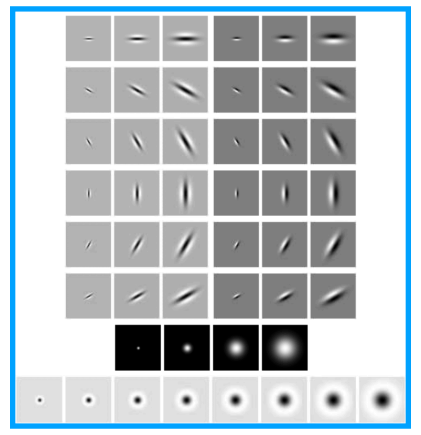
Image Structure 4
Image textures, textons and filter banks

Summary

- Image textures
- Historical context: preattentive vision
- Textons revisited
- Link to deep learning: "hand-crafted" feature extractors vs "data-driven" feature extractors

Image Textures


Studied in preattentive vision


Textons - prototypes from clustering filter bank responses

Key idea

Filter banks can be used to characterise texture

Deep Learning 4

Strategies for Neural Network Design, Scaling Phenomena, Transformers

Summary

- Strategies for Neural Network Design
- Scaling phenomena
- Transformers
- Vision Transformers
- Transformer Explosion
- Energy Consumption

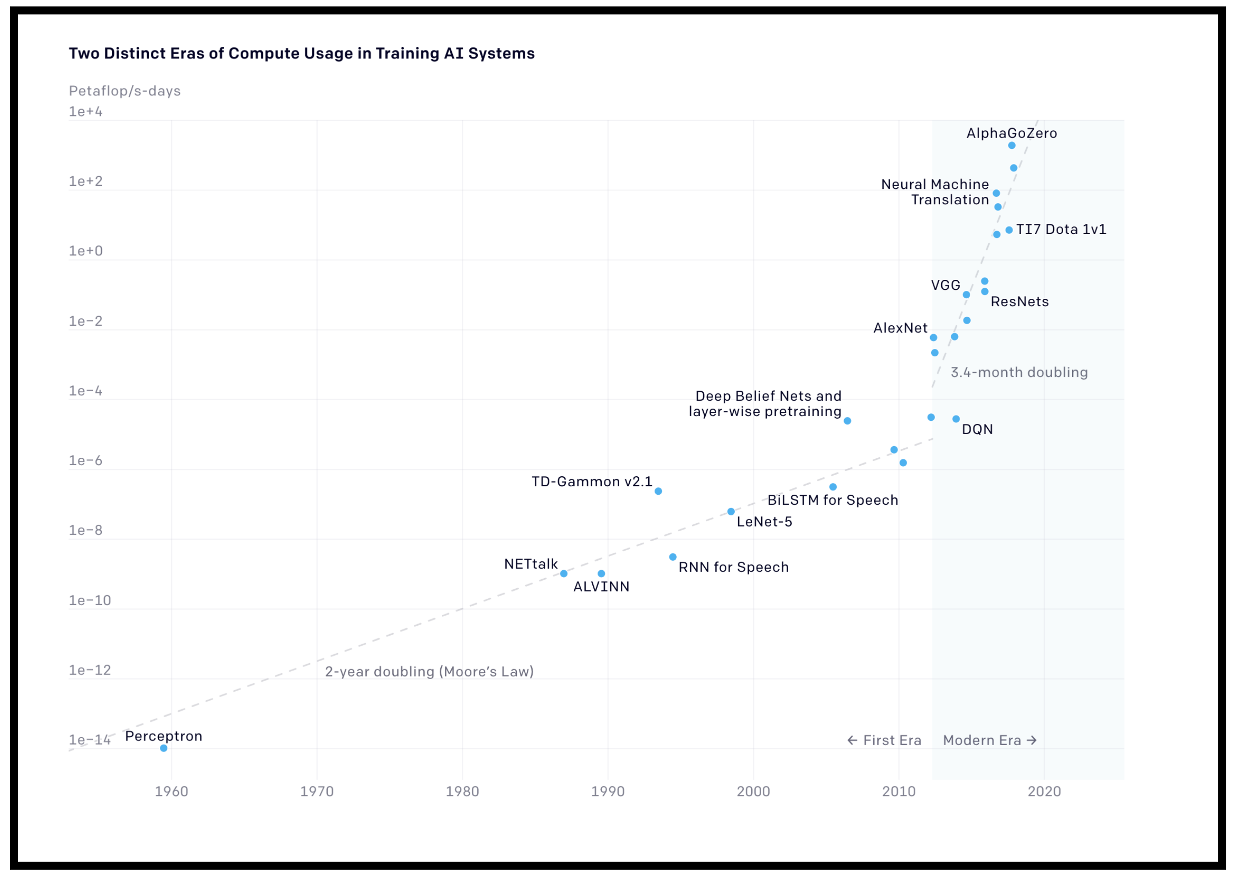
Hand designed

Random Wiring

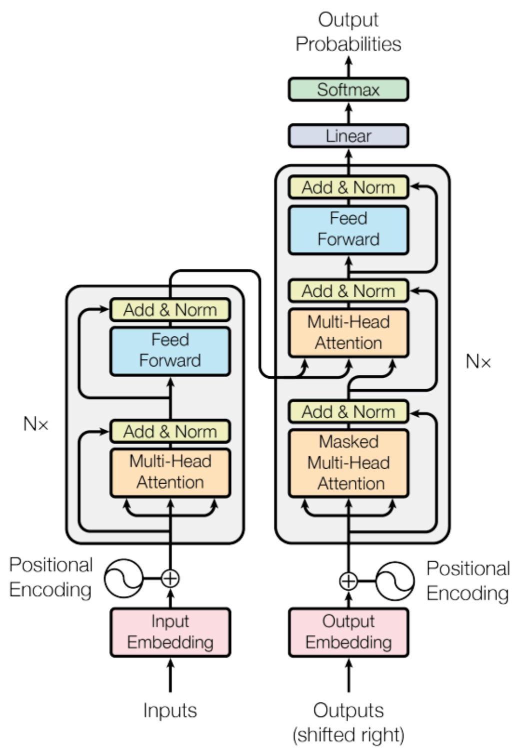
Evolution

Architecture Search

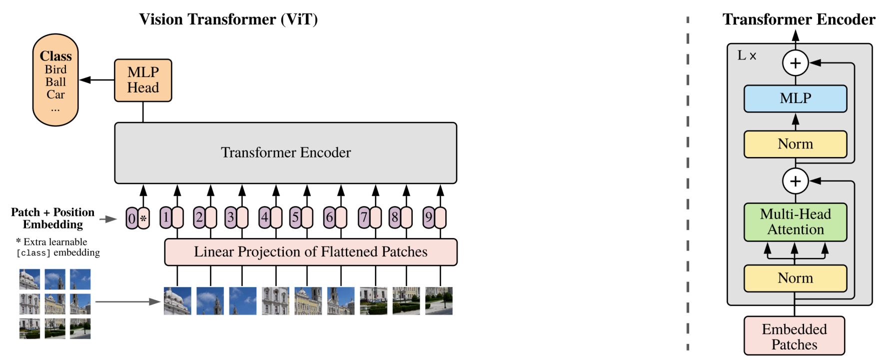
Design Strategies



Scaling Phenomena: Models are getting bigger



Transformers



Vision Transformers

Key takeaway

Architecture design grows increasingly important as more methods employ Deep Learning

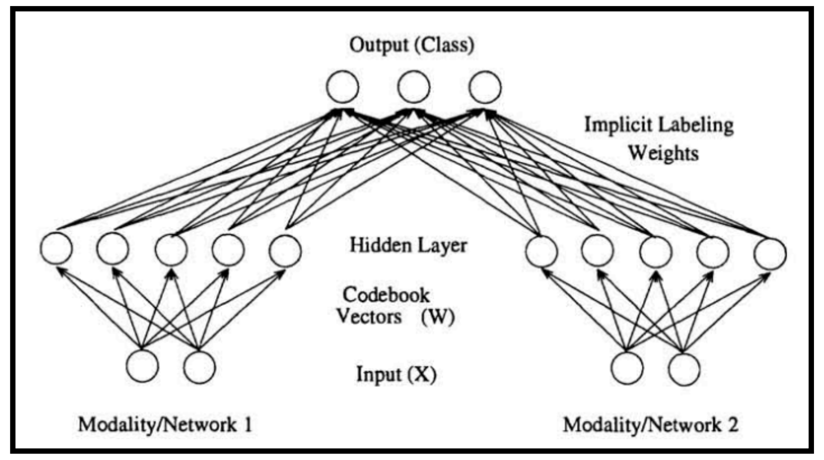
Deep Learning 5
Self-supervised learning and Pseudo-Labelling

Summary

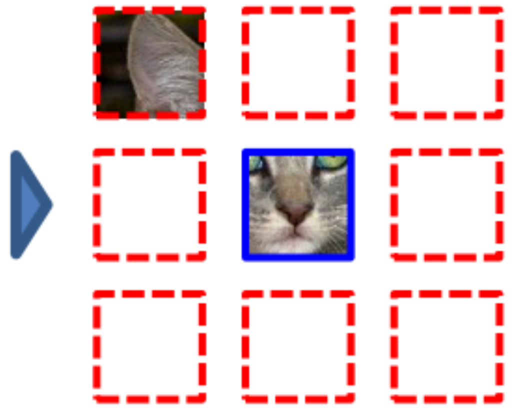
- Motivation
- Self-supervised learning
- Pretext tasks
- Beyond Image Representations
- Pseudo-labelling

Incremental | Social | Physical
Exploratory | Language-based | Multi-modal

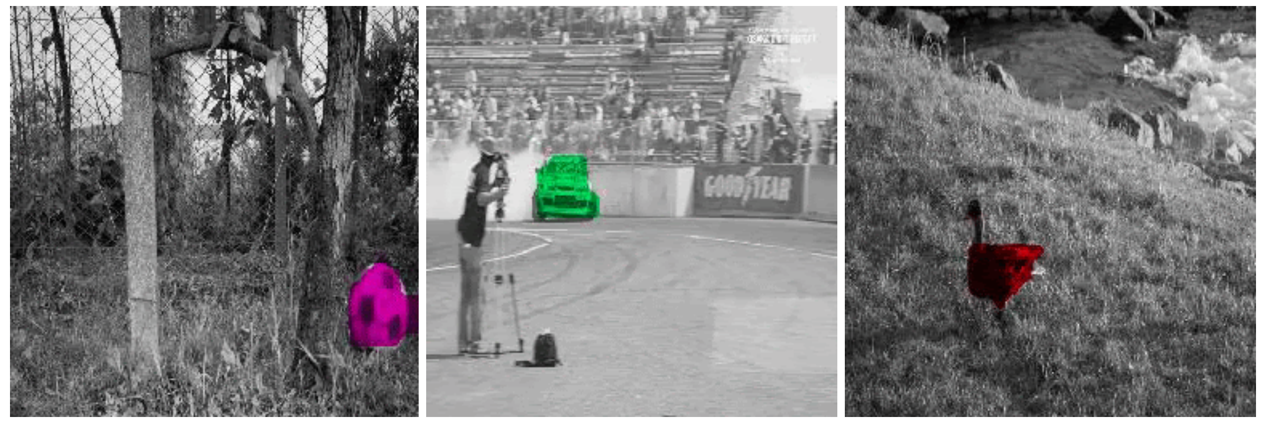
Motivation: What can we learn from babies?



Self-supervised: create your supervision



Pretext tasks



Beyond Image Representations

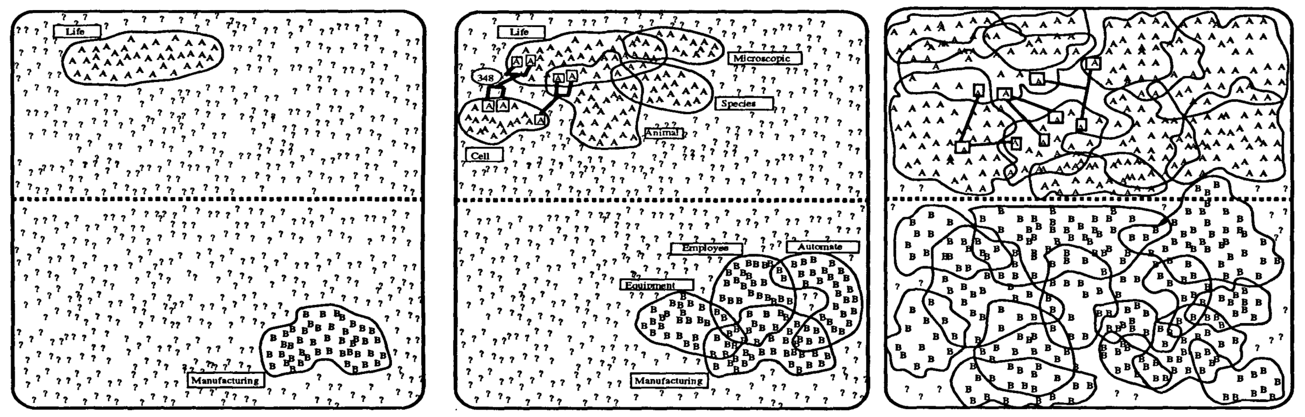


Figure 1: Sample Initial State
Figure 2: Sample Intermediate State
Figure 3: Sample Final State

Pseudo-labelling

Key takeaway

Not everything can be solved with supervised learning. Self-supervised/semi-supervised learning are promising ways forwards.