Iris Groen – Day 1 Lecture 1

07.07

Part 1

→ Hubel & Wiesel (cats :))

→ orientation tuning

→ retinotopic organisation

* Kobatake & Tanaka (1994)
* Desimone et al (1984)
* Grill-Spector & Weiner (2014)

fMRI ← where

MEG, EEG ← when

! Alexia part appears when you learn to read and it's always in the same place of the brain.

→ hierarchical theory of object recognition

Part 2

How human vision influenced computer vision?

→ Gabor filters

→ computer vision meets deep learning

→ solve invariant object recognition

→ ImageNet & AlexNet

* Krizhevsky, Suts... (2012)
* Fukushima (1980) Neocognition
* LeNet, LeCun
→ poloclub github cnn explainer
* Lindsay (2020) computational neural networks
* Zeiler & Fergus (2014)
* Gucly & van Gerven (2015)

LeCun et al, Nature (2015)

O'reilly (2020) (?)

diCarlo & Cox (2007)

* The Woreds I See Fei-Fei Li

Part 3

→ Do CNN networks predict (features) brain activity?

→ HCNN layers (?)

→ Grouping similarity

→ Networks seem to mimic human brain

- Lindsay (2020)
- Yamins et al. (2014)
- Kriegeskorte (2014)
- Kriegeskorte & Mur (2012)
- King & Groen et al (2018)
- Doerig et al. (2022)
- Groen, Silson & Baker (2017)

→ Brain-Score Algonauts

→ Neuroconnectionist research cycle

→ object recognition & scene processing

Part 4 How computer vision & human vision (still) different?

→ retinal image vs cortical map.

→ transformers are better than neural networks

→ Neuroscience & computer vision work together, they feed eachother

→ global statistics extraction, feedback/recurrent connections, peripheral sampling, ecological training, etc. ? ! ?

- * Da Costa et. al. (2021)
- * Cheung, Wells ...
- * Muller, Scholte & Groen (2024)
- * Geirhos et. al. (2018)
- * Geirhos et. al. (2019)
- * Dapello, Marques et. al Simulating a Primary ...
- * An image worth 16x16 words
- * Tuli, Dasgupta, ... Are convolutional NNs more like com. vis.?
- * Conwell et. al (2024)

Andrew Tracks - Day 1 Lecture 2

"no learning"?

* Classical tracking
  → motion detection & tracking
  → individual vs. multiple targets

→ pixel-level motion
→ optic flow
→ background models

} detect movement not tracking ∴

* if we can model the motion, we can predict objects' future locations.
  → HCI, robotics, surveillance, medicine...

flow-field ↔ quiver plot

→ motion difference & background subtraction

two basic approaches in motion detection

motion detection

- multiple object tracking -

* Markov Chain Monte Carlo tracking
* Metropolis-Hastings' algorithm
  → chain of predictions

→ Khan Z. et.al. (2004) An MCMC...

* social extensions
  → behaviours
  → sharing motion information

- tracking -

* uncertainty
* we want to be able to PREDICT where our target will be, and UPDATE our guess with a measurement.
* Kalman Filter
* Particle filters
  → real tracking is often multimodal

→ Isard & Blake (1998) Condensation

* appearance model(?)
* contour tracking
* mixed-state condensation
  → transition probability
* behavioural recognition !
* the curse of dimensionality

} still relevant in the deep-learning era

Michael Wray - Day 2 Lecture 1

08.07

Egocentric Vision - Making Sense of the First-Person Perspective

* Steve Mann, WearCam, WearComp
  applications in Personal Imaging
* Wearable devices
* Not limited to humans
  ↳ dogs, robots, shopping cards
* Blue Sky Outcomes
  ↳ robotics to understand surroundings
  ↳ personal assistants
    - remind you what you forget
    - help you fix things
    - remind you a recipe's next step etc.
  ↳ healthcare for assisting people
  ↳ ethic & privacy is an issue.!
  ↳ Hololens, Google Glass, etc.?

Ego4D dataset
GTEA, BEOID, GTEA Gaze, UT Ego,
ADL, EGTEA Gaze +, EPIC-Kitchens-55
Charades-Ego, EPIC-Kitchens-100
↳ combining every dataset
Ego4D

Past — **Present** — **Future**

- Episodic memory
- Hand & object
- Audio visual diarization
- Social interaction
- Forecasting

HD-EPIC

Digital Twin of kitchen videos
- narration, fine grained annot
  audio annotation, object detection

- object-fixture assignment

- Blind-Language: Llama 3.2, Gemini Pro
- Video-Language: VideoLlama 2, LongVA, Gemini Pro

Ego-Centric Tasks

- video understanding tasks
  ↳ action recognition, VQA
- captioning, retrieval, grounding
  audio localisation, activity & object recognition

→ EgoVis workshop

HierVL: learning Hierarchical Video-Language Embeddings

- EgoVLP - previous study, not good (?)
- Assembly101 dataset (2022)
  ↳ assembly Hands
  ↳ SVEgoNet : single view ego net

LEGO: Learning Egocentric Action Frame Generation via Visual Instruction Tuning

---

Privacy & Ethics

- Egocentric data is highly personal
- capturing should be done carefully
- biases can be added easily.
- indoors only, other people are not involved & recorders review their videos.
- anonymisation / de-identification

Prof. Victor Sanchez - Warwick

Day 2 Lecture 2

CLIP: Contrastive Language-Image Pretraining for Video Analytics

* Deepfakes, IAPR head of.
* SIPlab,
- multimedia forensics & biometrics
  - ↳ platform provenance

Project: multicamera Trajectory Forecasting tracking via person re-identification

Styles, Guha, Sanchez (2021) Pattern Analysis...
Nooker, Sanchez (2025) WACV

---

* CV + DL

which? → when? → where?

trajectory tensor

→ Enc Dec → which, when, where? CNN-GRU

mapping features to outputs

feature extraction

historical data as input

→ used tensors instead of vectors to convey the uncertainty

* FC, CNN, GRU, tCNN
* binary cross-entropy loss

Tiny Faces project

MTCNN + IFS

Deyra, Sanchez, Li (2018)

Deyra, Shen, Bahadur ... (2025)

(Automatic Face & Gesture Recogn.)

IEEE

Tiny Faces IFS detector

↳ Fisher Vector

* IFS to FVs → reduces the # of operations

Krishnashtha & Guha (2018)

(ICIP)

MEVA dataset

multiview extended video with activities

Qi, Tan, Yao, ... (2023) – Yolo5face

Yu, Huang ... (2024) – YoloFacev2

tinyfaces is still a challage to be solved.!

———— o ————

Recall: CLIP is proposed to be used in downstream tasks

→ WebImageText (WIT)

CLIP for segmentation

Lüddecke, Ecker (2022) → CVF

DACE: Distance-Aware Cross-Entropy Loss

HuggingFace: spaces/yiming-M/CLIP-EBC (?)

Crowd Counting

→ rely on annotated 2D coordinates of individuals' head centres in images.

Ma, Sonche2, Guha (2022) → ICIP

Ma, Sonche2, Guha (2025) → ICME (CLIP-EBC)

↳ Reframe crowd counting as a classification task

⇓

they introduced this idea to CLIP.

* zero-shot learning: can we facilitate?

↳ natural language supervision

* contrastive learning

Radford, Kim, Hallacy ... learning transferable visual models from natual language supervision

* symmetric crossentropy loss

Attention is all you need

Encoders of CLIP => ResNet50: attention pooling

Vision Transformer

Dr Andrew Gilbert - UV Surrey

Day 2 Lecture 3

Impact & Application

Surveillance
Healthcare
Autonomous Vehicles
Entertainment

From Paragraphs to Pixels: LLMs in Video Understanding

@andrewjohngilbert.github.io

* Video understanding
  - ↳ long range temporal dependencies
  - ↳ ambiguous / overlapping events
  - ⋮
  - lots of challenges in this area.

YouTube: WhoDunnIt? It is easy to miss stuff you're not looking for.

* Video Understanding with Large Language Models: A Survey, Yunlong Tang et. al.

* Unsupervised learning of Visual Representations Using Videos, ICCV 2015
* Shuffle Learn ECCV (2016)
* Self-Supervised Video Rep CVPR 2017
* Arrow of Time, CVPR 2018 (?)
* ECCV 2018 - Tracking Emerges by Colorizing Videos
* VideoMAE, NeurIPS 2022
* MOFO:, NeurIPS 2022 workshop
* CVPR 2023 - learning Video ..
* Vid2Seq CVPR 2023
* VAST NeurIPS 2023, Sihan Chen
* FILS: Self-Supervised Video Feature Prediction...
* Fine-Grained audible video description
* SWINBert (?)
* AutoAD: CUF 2023
- → ? Donte AD ?

gap between pixels & nl reasoning

AI Storytelling

Art in Storytelling.

* Visual Agents in Software
  - ↳ unreal engine
  - or Game Engines nowtey

} Visual Interface Agent.

* MAGMA: Multimodal Agentic Foundation

  Yang et al. 2025

* UI-TARS, 2025

Audio-Visual Fusion (?)

* Ramazonaw... "OWL..." CUF 2023
* Egocentric audio-visual object localization, CUF 2023
* Self-supervised moving vehicle...
* Epic-Fusion: Audio visual temporal binding for egocentric action recognition

→ Dataset: Youtube8M THUMOS:

Local (?)
Learnable Gating
Cross Attention

* Multi-Resolution Audio-Visual Feature Fusion...
* DEL: Dense Event Localisation...? 2025

→ Action Quality Assessment ↔ related to this. ECCV 2022 with temporal parsing transformer

* learning to Score Olympic Events

Oisin Mac Aodha - U. of Edinburgh

Day 2 lecture 6

Representation learning -

@ sthalles.github.io/a-few-words-on-representation-learning

* bad vs good features

* Once trained, the representation can be transferred to other tasks.

- Wei Koh et. al. Concept Bottleneck Models, ICML 2020

* success depends on data representation

↳ what is good representation?

- compact (minimal)
- explanatory (sufficient)
- disentangled (independent factors)
- interpretable
- make subsequent problem easier

Eli Cole, Caltech (slide credits)

* Transfer learning (ImageNet)

- Bengio et al. Representation learning: A Review and Perspectives

credit: Justin Johnson eecs498

Self-Supervised learning

"Most of human and animal learning ..." ~Yann lecun (On true AI)

- How to learn features from the unlabelled data?

* Papers With Code: Self Supervised Image Classification

- Effective methods for learning representations.
- Gidaris et al. Unsupervised Repr. learning... ICLR 201
- Weng & Kim, Self-Supervised learning, NeurIPS 2021

Self-prediction

SimCLR (2018)

- Chen et. al. A Simple framework for contrastive learning ...
  - Github: google-research/simclr

Multi-modal Contrastive learning

- Radford et. al. learning Transferable Visual models from NL Supervision ICML 2021

* zero-shot classification

* there are limitations

↳ can require large batch size

↳ relies on "good" negative selection

↳ space of plausible augmentations can be dataset specific and must be defined in advance

* masked language models

↳ BERT masked language models

* vision transformers (ViT)

- Dosovitskiy et. al. An Image is worth 16x16 Words. ICLR
- He et al. Masked Autoencoders Are Scalable Vision learners CVPR 2022

- Zhang et al. 2016, 2017
- Doersch et al. 2015
- Noroozi & Favaro, 2016?
- Noroozi et. al. 2017

Neill DF Campbell - UCL

Day 3 lecture 1

09.07

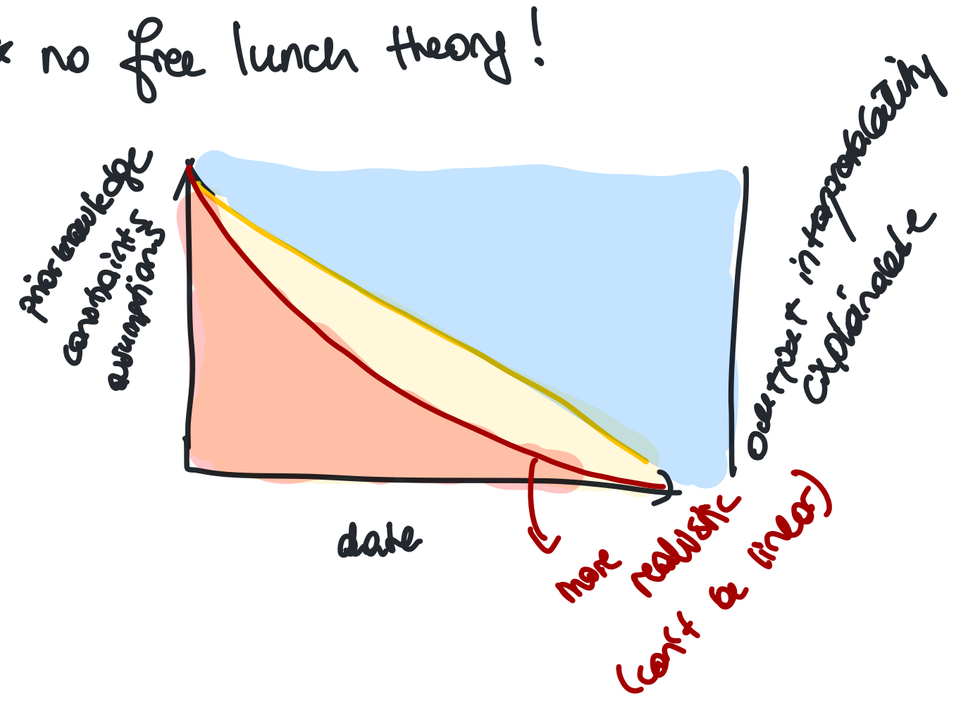
Uncertainty & Evaluation in Computer Vision

- Poggi ... On the uncertainty of self-supervised...

* why do we care about uncertainty?=
  - ↳ ambiguity in task, in our models
  - ↳ downstream decision making
    - safe - robust - transparent
  - ↳ improved performance
    - data efficiency
    - self supervision
  - ↳ evaluation & model selection!

* there are danger if we avoid probabilistic approch

* no free lunch theory!



Laplace: "the theory probability is the common sense reduced to calculus." (?)

* statistical significance

→ Youtube: Crimes against data.

* Human36M dataset

* ablation studies

- Simple and Scalable predictive uncertainty estimation using deep ensembles
- on uncertainty of self-supervised monocular depth estimation
- learning structured gaussians to upp. deep ensembles.

Recent advances

* non-contrastive Siamese networks

* BYOL, SimSiam

* DINO

- Emerging Properties in self-Supervised Vision Transformers, ICCV 2021

* DINOv2 builds on DINO & iBOT

- DinoV2 : learning Robust Visual Features...

Limits of SSL

* most SSL pretrained with ImageNet

- Cole et al. When Does Contrastive Visual Representation learning work ? CVPR'22

* dataset size matters
  - ↳ amount of unlabelled data for pretraining
  - ↳ amount of labelled data for supervised learning

* iNAT
  - Van Horn et al Benchmarking... (?)
  - Zong, Aodha... Self Supervised ...

Vision @ UoE

---

* Monty Hall problem.

* Bayes' rule

→ How can we make it better if we don't know where it fails...

→ Understanding Deep Learning - Simon Pierce

* Bias-variance trade-off

Paul Henderson - Uni. of Glasgow

Day 3 lecture 2

* GAN, VAE, DDM
* no semantic manipulation
  ↳ so structured generative models
* generative model: probability dist. we can sample from

→ give latent variables Semantic meanings, by forcing the decoder

- AIR (Eslami NIPS'16) - SPACE (Lin ICLR'20)
- SPAIR (Crawford AAAI'?)
- Anciukevicius: ICML 2020 OOL workshop

- NeRF

meshes

- SoftRas (Liu 2019)
- NMR (Kato 2018)
- DIRT (Henderson (2018)

} differentiable rendering

- Mildenhall et. al, ECCV 2020
- GRAF, Schwartz et. al. NeurIPS 2020
- Henderson, Tsiminaki, Lampert - CVPR (2020)
- RenderDiffusion, CVPR 2023
- Denoising Diffusion via Image-Based Rendering, ICLR 2024
- PixelNeRF
- Sampling 3D Gaussian Scenes in Seconds with latent Diffusion models · Henderson et al. 2024

Elliot Crawley - Uni. of Edinburgh

Day 3 lecture 3

Efficient models for computer vision

- arxiv.org/abs/1711.02613
- arxiv.org/abs/2211.10438v7

*methods
  ↳ pruning → structured
                → unstructured
  ↳ quantisation
  ↳ knowledge distillation
  ↳ neural architecture search

*channel pruning → heuristically

- arxiv.org/abs/1802.03494

* FP32 to INT8
* symmetric quantisation
* asymmetric quantisation
* quantising weights
  activations (post-training)

* logits, softmax, log loss
* distillation
  ↳ teacher & student

- arxiv.org/abs/1312.6184

- arxiv.org/abs/1612.03928
- ieee: 9446880
- arxiv.org/abs/1805.04770
- arxiv.org/abs/2203.05469
- AlexNet (2012)
- VGGNet (2014)
- ResNets (2015)
- arxiv.org/abs/1512.03385
- Inception V3 (2016)
- NASNet (2018)
- arxiv.org/abs/1808.05377

- Zoph & Le (2017)
- arxiv.org/pdf/1707.07012

* CIFAR-10

- arxiv.org/pdf/1802.01548.pdf
- arxiv.org/pdf/1807.11626.pdf
- arxiv.org/pdf/1808.05377.pdf

* one-shot NAS
* zero-shot NAS

*DARTS

- arxiv.org/pdf/1806.09055.pdf

* NASWOT

- arxiv.org/pdf/2006.04647.pdf
- arxiv.org/pdf/2301.11300.pdf
- arxiv.org/pdf/2010.11929v2.pdf

- GPViT (ICLR 2023)

*specialist architectures

Mike Mangan – Uni. of Sheffield

* SLAM Primer
* SLAM/AI
* robotics industry.

Localisation is solved ...?

↳ insufficiently accuracy
↳ subject to interference
↳ requires infrastructure
↳ not a general solution

---

* Simultaneous Localisation & Mapping (SLAM)

  • Boilg, Tim & Hugh "Simultaneous localization and mapping" Part 1 & 2 IEEE robotics & automation (2006)

* Using vision to create world models (maps)

  • NeRF-SLAM: Real-time Dense Monocular SLAM with Neural Radiance Fields

* indirect SLAM

  • ORB SLAM3 → totally open source

* Where SLAM deployed today?

  ↳ AR/VR

• Youtube: Optosan 1 minute introduction

  reverse-engineering biological brains, creating general purpose neuromorphic software

* SoTA autonomy → create human algorithms
* Natural autonomy → how humans solve autonomy
                   → extract nature algorithms
                   → how nature solve auto.

* the struggle is to build adaptable machines

* Natural Intelligence

  ↳ identify valuable natural behaviours in nature
  ↳ separate algorithms from neural implementation
  ↳ map algorithms to standard compute h/w

* neuromorphic software for machines.

* intelligence isn't artificial, it's natural.

Armin Mustafa – Uni. of Surrey

4D Machine Perception of Complex Scenes

* Single & multi-view reconstructions

* generative ai
* video understanding

* 4D Perception

Computer vision | NLP
Generative AI | machine learning

* Personalised media

  ↳ AI for me project

* 4D Vision

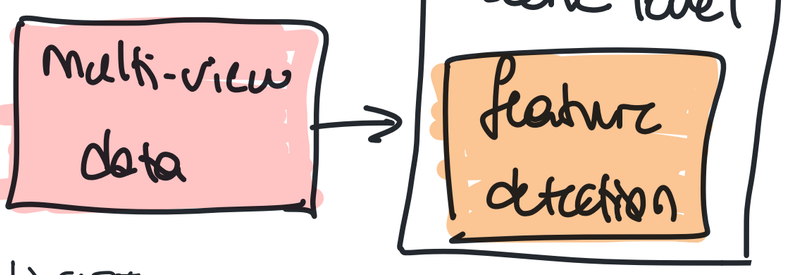
3D from monocular videos

Uparametric reconstruction

- depend on parametric 3D models
- estimate model parameters
- ??

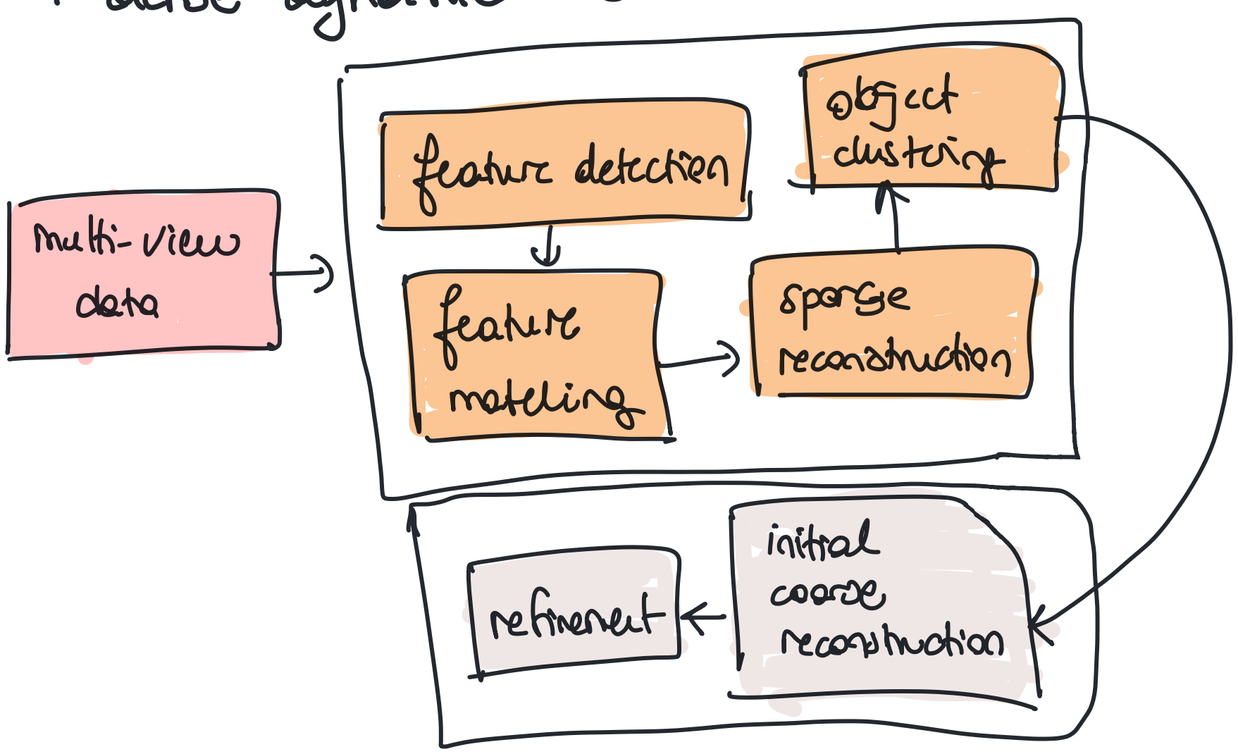
(photos)

* general scene reconstruction



↳ SIFT
↳ SFD
} algorithms

* dense dynamic 3D reconstruction



* machine learning based 3D reconstruction

* dense dynamic 3D reconstruction

↳ training data S2P2 dataset (?)

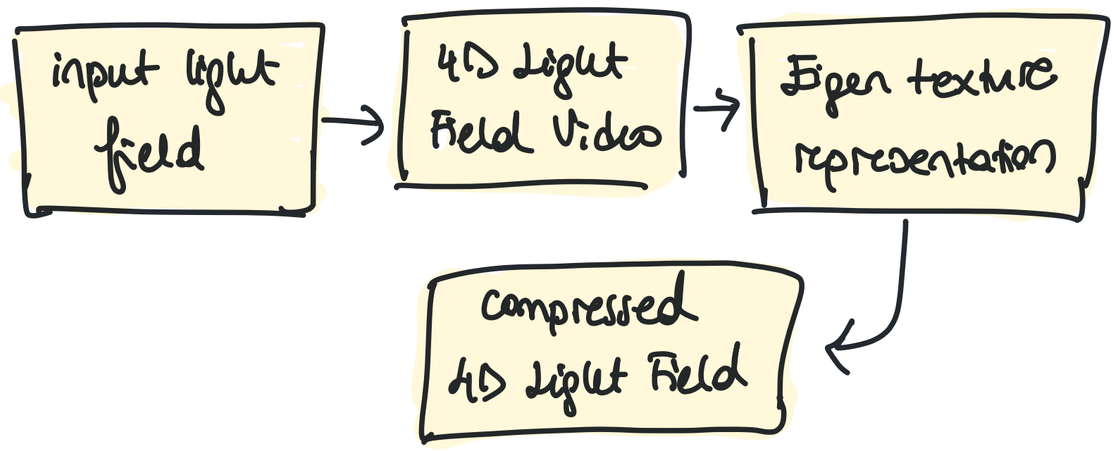
* 3D Video to 4D Models
(photo)

- video: 4D light-field

* Kinch & the Double World : movie :)

* epipolar plane information (EPI)

* light-field Video



Why 4D Scene understanding?

* estimate semantics, reconstruction and motion simultaneously

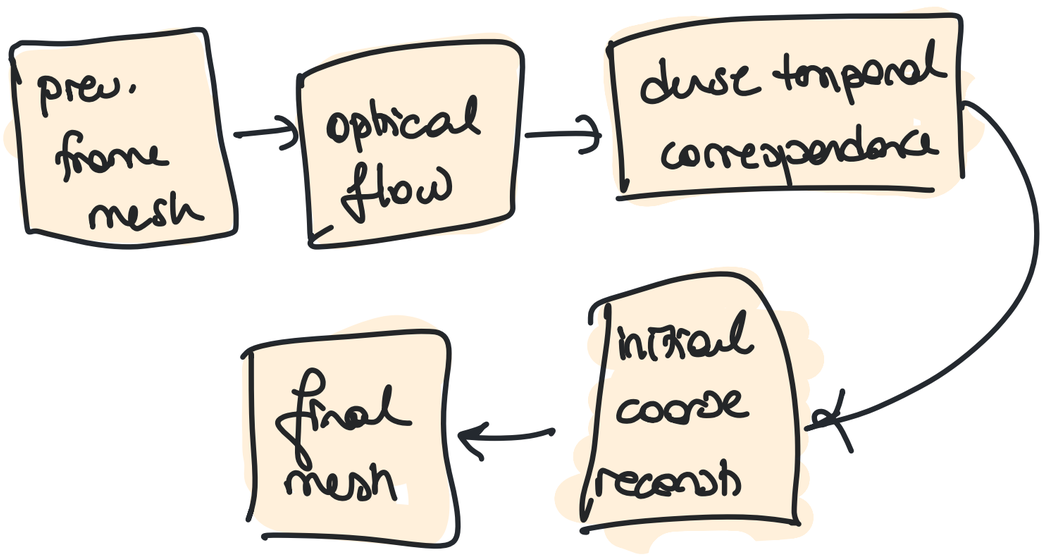
*sparse temporal tracks

↳ introduce temporal coherence

↳ appearance, motion & semantic inf.

↳ 3D human pose - part of sparse temporal tracks.

* temporal coherence



* geodesic star convexity (GSC) in optimisation

⇒ semantics + segmentation + depth inform

! could be applied to scene reconstruction methods

VR, video creation, ...

⇒ single-image 3D human reconstruction

⇒ 3D Virtual Human Dataset

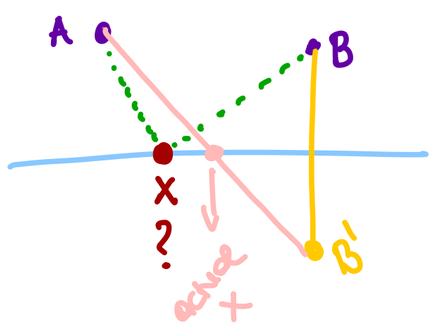
Tolga Birdal - ICL Day 4 lecture 3

Shape Shifters: Geometry, Topology and Learning For and Beyond Computer Vision

* Representation has been known to be notorious problem in 3D vision

* 14th century Parisian manuscript of euclid's elements

* Heron's solution



$Z = \{x, y\}$ (x: inputs, y: labels) $S \supset Z$ → data; S → training set

$\hat{R}(f) = \frac{1}{n} \sum_{i=0}^{n} \ell(f(x_i), y_i)$ → loss function

$R(f) = \mathbb{E}[\ell(f(x_i), y_i)]$ → integral over life

$|\hat{R}(f) - R(f)|$ → generalisation error

* invariance & equivariance
  @ maurice-weiler.gitlab.io

* Fourier transforms
  * Laplacian → Pierre Laplace

→ what is not representable with graphs?

@robertpepperell.com
"induce an indeterminate mental state"

→ micropenesis, Gestalt

! the whole ≠ the sum of its parts

* canto 33 · Dante

→ cognition is synergistic

Binod Bhattarai - Uni. of Aberdeen

Day 4 lecture 4

Towards Trustworthy AI in medical Imaging

* ~80 of all healthcare data comes from medical images

* Modality 1: X-Ray (Radiography)
* Modality 2: Computed Tomography (CT)
* Modality 3: Magnetic Resonance Imaging (MRI)
* Modality 4: Ultrasound Imaging.
  (proven to be very safe??)

⇒ endoscopic imaging.

The Human Factor and Motivation of AI

→ most abnormalities are missed by human

→ 1% increase in detection in abnormalities
3% decrease in cancer risk

* AI as an Assistive Tool in Medical Imaging.

- 3D U-Net (Cicek et. al. 2016)
- SegFormer-3D, Perera et al. 2024
- Explainable Vision Transformers and Radiomics for COVID-19 Detection in Chest X-Rays

* Vision Language models
  CLIP, BLIP, Flamingo, LLaVA, GIT, MedCLIP, BioViL

* Medical Report generation
  - Zhou et. al. ICCV 2021

* Question answering
  - Lee et. al. ICLR 2024

Challenges and Future Direction for Trustworthy AI

* European High Commision Expert Group on AI (2019), trustworthy ai:
  - ↳ reliability / robustness
  - ↳ privacy
  - ↳ fairness
  - ↳ explainability

* Federated learning for privacy protection
  - Fed Average (mcmahon et.al. AISTATS 2017)
  - FedProx (Sahu et al. MLSys 2020)
  - Federated learning for Optical Coherence Tomography (Anpain et. al. 2024)
  - Li et. al. MLSys 2020
  - Karimireddy et. al. PMLR 2020
  - Poudel 2024 (?)
  - Poudel 2025 (?)
  - Li 2025 (?)

* multimodal FL with missing modalities
  - Multimodal FL by Augmentation by Retrieval in Clients (Poudel et. al., MICCAI 2024)
  - Feature Imputation Network (FIN) (Poudel et.al MIUA 2025)

* What makes medical AI unreliable?
  - heterogeneous data
  - long-tail problem
  - ? ? ?

- Feature-Based OOD detection using NCDD score in GI (Pokhrel et. al. MIUA 2025)

↳ nearest enter ? deficit ?? !

- NERO: Explainable Out-of-Distribution Detection with Neuron-level Relevance (Chhetri et. al., MICCAI 2025)
- Layer-wise relavence propagation: an overview (2018)
- From Attribution Maps to Human-Understandable explanations through Concept Relevance Propagation, 2023 Nature machine Intelligence
- Benchmark for Hallucination Detection in VLM for GI Imagining. (Pokhrel et al. MICCAI 2025)

Pascal Mettes – Uni of Amsterdam

Day 5 lecture 1

Hyperbolic & Hyperspherical Learning

→ House of cards in deep learning

→ why are we so euclidean?

- euclidean is not
  - ↳ hierarchical
  - ↳ compact
- the world is not always euclidean.

• Bochman et. al. ICML 2020

↳ we need a hierarchical geometry to represent the learning.

* hyperbolic geometry?
  - ↳ euclid's 5 postulates
  - ↳ the rise of non-euclidean geometry
  - ↳ the poincaré ball model
    - escher drawings
* hyperbolic ... embeddings?
  - ↳ poincare embeddings
    - Nickel and Kiela, NeurIPS 2017
  - ↳ hyperbolic entailment cone
    - ganea et al. ICML 2018
    - van Spengler & Mettes, 2025 (under subm.)
* hyperbolic computer vision?
  - ↳ hierarchical learning for action recog.
    - Long et al. CVPR 2020

! hyperbolic grammars (?)

↳ hierarchies need to be given, not learned

↳ hyperbolic image embeddings
- Khrulkov et al. CVPR 2020

↳ Hyperbolic zero-shot learning
- Liu et al. CVPR 2020

* Hyperbolic image segmentation
  - Ghadimi et. al. CVPR 2022

* Hyperbolic vision-language models
  - Radford et al. ICML 2021

"all vision-language models should be hyperbolic"

⇒ vision-language models explicitly assume that image and text are equal.

But one sentence does not uniquely define an image!

Texts are more general than images. (Language is more general than vision)

- Alper et. al. (2024)

↳ "foundation VLMs exhibit zero-shot hierarchical understanding."

* CLIP: embed images and text in a Euclidean space

* MERU: embed images and text in a hyperbolic space.
  - Desai et al. ICML 2023
  - Ibrahimi et al. TMLR 2024

* Hyperbolic CLIP better for:
  - spatial awareness
  - ambiguous tasks
  - distribution shifts
- Pal et al. ICLR 2025

* Hyperbolic Safety-Aware Vision-Language
  - Poppi, Kasarla et al CVPR 2025

* Using hyperbolic geometry to remove samples
  - Kim et al. ECCV 2024

* Hyperspherical Deep learning
  - Zhao et al (2018)
  - Cohen et al. (2018)

* spherical losses dominate deep learning
  - Aggarwal et al. (2001)
  - Hinneburg et al. (2000)

* Decoupled Networks
  - Liu et al. CVPR 2018

* maximum class separation
  - Kasarla et al. NeurIPS 2022
  - ↳ problem is known as Tammes problem

Potential of Hyperbolic learning

→ Hierarchical learning

→ robust learning

→ multi-modal learning

→ brain-like networks

Challenges of Hyperbolic learning

→ fully hyperbolic learning

→ computational challenges

→ open source community

→ learning at scale

Oya Celiktutan - King's College London

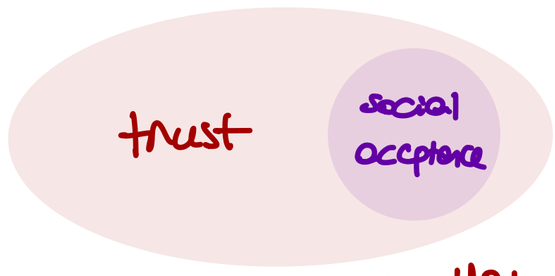
Day 5 lecture 2

Multimodal Human Behaviour Understanding and Generation for Human-Robot Interaction

* Classical Vision vs. Robot Vision

* Three application areas (Raychoudhury et al., 2023):
  - → internal state est.
  - → external environment est. (nav. & manip.)
  - → human-robot interaction (social or physical)

- Google DeepMind, Gemini Robotics video
- Silvia Rossi : Socio-Physical Human-Robot Inter.
- Hong-Long Cao et al. [comic image]



trustworthy of HRI systems

* socially-aware navigation

* robocentric indoor crowd analysis
  - ↳ f-formation theory
  - Zheng & Chen, 2018
  - Hamilton et al. 2017
  - } Group detection with link Prediction

- Advantage Actor-Critic (A2C) learning
  - Do et. al. 2020

social interaction reward function

- Navigation Turing Test (Devlin et. al. 2021)
- PSI Questionnaire (Barchard et al. 2018)

* Understanding Human Gaze Behaviour
  - Recasens et. al 2015
  - Fan et al. 2018

* Learning to Generate Co-Speech Gestures
  - rule-based
  - data-driven (Yoon et al. ICRA 2016)

* diffusion based models
  - too slow
  - switch to Vector Quatised Variational AutoEncoder (VQVAE)

* Real-time motion generation from text (image: Habitat 3.0)

* GENEA Challenge 2023

* Artificial Empathy *

* MAPTRAITS: Real-time personality Prediction

* UDIVA: Personality Recognition Track ICCV '21 Challenge

* PE-HRI - TEMPORAL Dataset (Nasir et al. J. Social Robotics 2021)

* Utility-Fairness Framework
  - Spurious correlation