Current Topics in Computer Vision and Machine Learning

Introductory Meeting

Prof. Dr. Bastian Leibe

Lehr- und Forschungsgebiet Informatik 8 (Visual Computing Institute)

RWTH Aachen University

http://www.vision.rwth-aachen.de

Organization

- **Reports**
  - English or German (depending on the supervisor)
  - 20 pages (+- 1 page) Bibliography counts – TOC does not
  - LaTeX is mandatory

- **Presentations**
  - In English
  - 45 minutes (+- 5 minutes)
  - Block at the semester end – 3 days of presentations
  - Slide Templates will be available on the webpage
  - Laptop can be provided for the presentation if necessary

Schedule

- Hand in Declaration of Compliance
  - as hardcopy – before outline
- Outline: Monday, November 21th
- Report: Monday, January 9th – graded version!
- Slides: Monday, February 6th
- Presentations: February – 20th-22th – 3 days
  - Turn in corrected report at presentation day

Hints for your report – DOs

- Read and understand your paper
- Write a report in your own words
- Search for additional literature
  - Take part in a library tour
  - Compare your paper to work of other authors
  - Explain the bigger picture
  - Describe something extra – content beyond the topic‘s original paper
- Discuss the advantages & disadvantages of the approach
- Make the reader understand the topic
  - Audience: Your fellow seminar participants
- Correctly cite all sources (also for all figures)

Hints for your report – DON‘Ts

- Do not simply copy or translate original text!
- Do not miss the deadlines
  - Penalty for every day you exceed any deadline
- We will check if you...
  - Have copied content / text from the paper or other sources
  - Have not correctly cited any material, etc.
  - If you do, you **immediately** fail the seminar

Reminder: How to cite

- General rule: For every piece of information it has to be clear if it is your own work or someone else‘s.
  - If your text contains “Our approach…”, “We propose...”, etc. you are doing it wrong...
- Direct Quote:
  - Smith et al. state that their “approach combines x and y in order to achieve z” [5].
  - You have to use direct quotes if you copy original text.
  - **Avoid** such direct quotes – and instead use your own words
- Indirect Quotes:
  - Smith et al. use an approach which combines x and y allowing to... [5].

Reminder: How to cite

- Mind credible sources
  - Papers published in journals or conference proceedings
    - Peer reviewed == reliable and good
  - arXiv.org
    - Depends!?
  - Wikipedia
    - Can be altered by anyone and it changes over time == not good
- Use the original sources
  - Instead of sources that only cites the original source
  - That requires to also look (and dig) for the original sources!
- Use BibTeX
  - Saves a lot of trouble
  - And good practice for your master thesis

Important Details – before we start...

- Declaration of Compliance
  - Read “Ethical Guidelines for the Authoring of Academic Work”
  - See seminar webpage for the document
  - Sign and hand in to me – as hardcopy – before outline deadline

- Send all submissions regarding the seminar to
  - seminar@vision.rwth-aachen.de

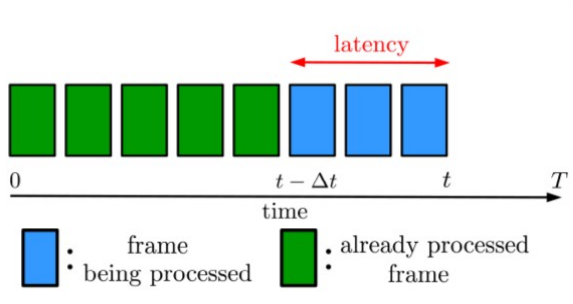
Topic 1 - Wolfgang

Improving Multi-frame Data Association with Sparse Representations for Robust Near-online Multi-object Tracking

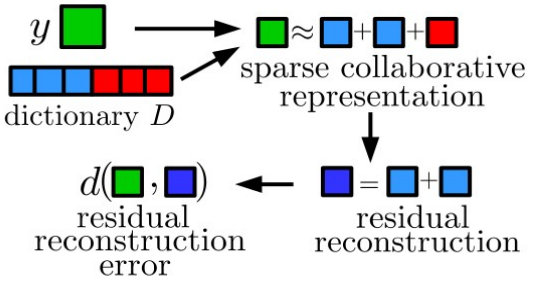
Fagot-Bouquet et al. (ECCV 2016)

- Near-online tracking-by-detection
- Solve data association using an energy formulation
- Focus on how to properly incorporate appearance

Multi-object Tracking by detection



Sliding window-based approach

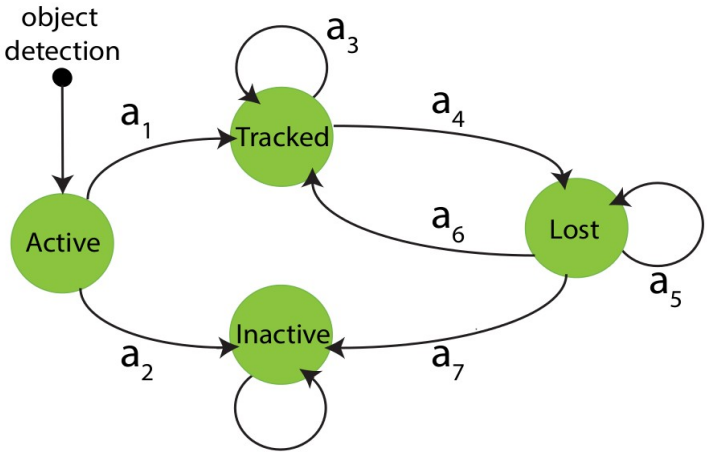


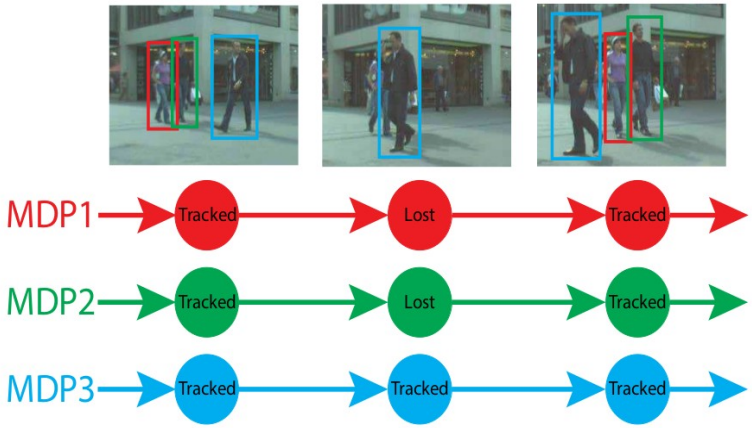
Classification with collaborative representations

Topic 2 - Wolfgang

Learning to Track: Online Multi-Object Tracking by Decision Making Xiang et al. (ICCV 2015)

- Model Tracking as a Markov Decision Process
- Explicitly handle Birth/Death/Occlusion/Reappearance
- Learn the Policies by Reinforcement Learning
  - Rather than designing them by hand

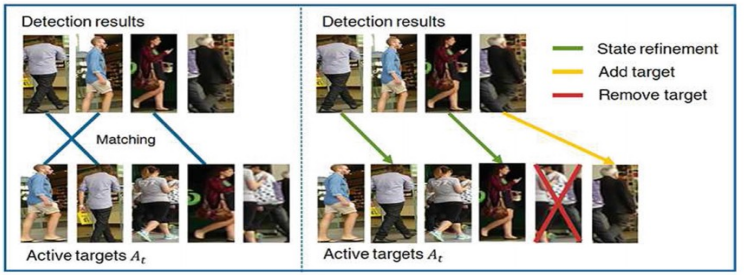


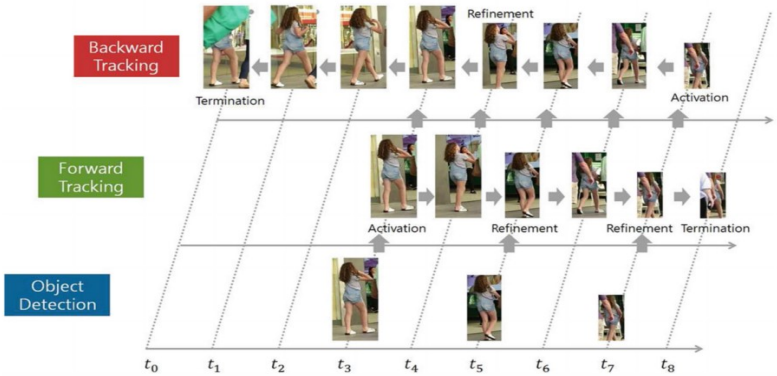


Topic 3 - Stefan

CDT: Cooperative Detection and Tracking for Tracing Multiple Objects in Video Sequences Kim et al. (ECCV 2016)

- Combine object detector and model-free tracker
- Manage tracks according to their visibility in a forward pass
- Refine and restore tracks in a backward pass (offline)



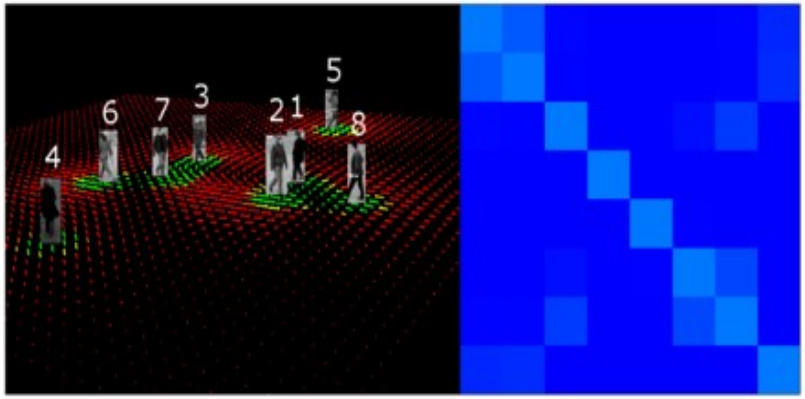


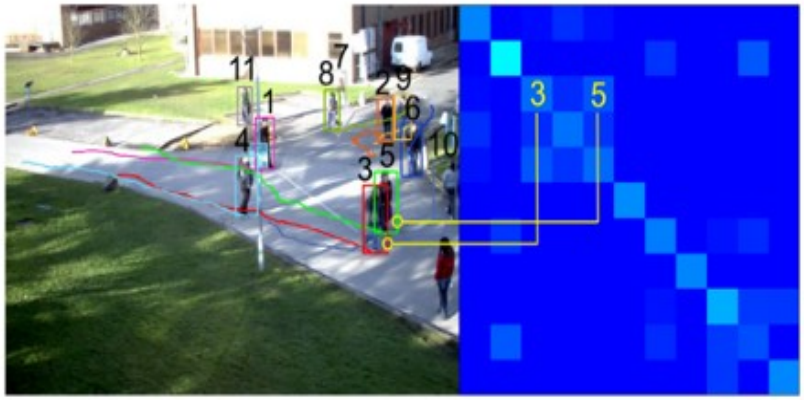
Topic 4 - Stefan

A Gaussian Process Based Multi-Person Interaction Model

Klinger et al. (ISPRS 2016)

- Take into account motion of all existing, interrelated targets
- Predictive model based on Gaussian Process Regression
- Update pedestrian positions with Dynamic Bayes Networks

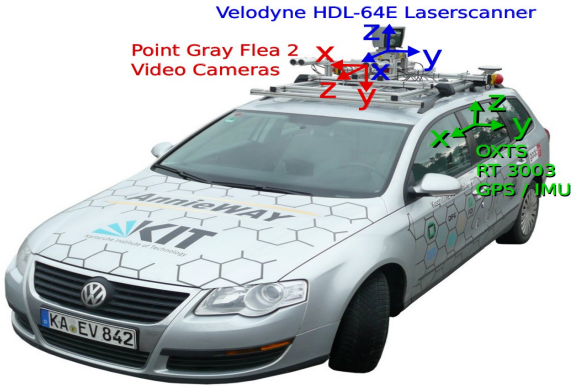


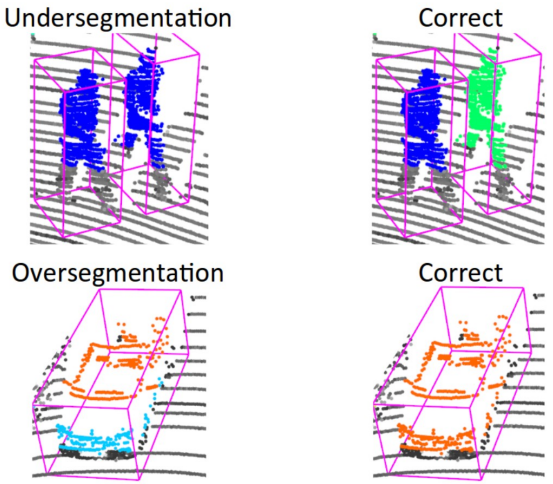


Topic 5 - Aljosa

A Probabilistic Framework for Real-time 3D Segmentation using Spatial, Temporal, and Semantic Cues Held et al. (RSS 2016)

- Scene understanding: robots must always be aware of their whereabouts!
- Task: segment 3D LiDAR scans into a collection of separate objects
- Probabilistic 3D segmentation method
- Utilize spatial, temporal and semantic information



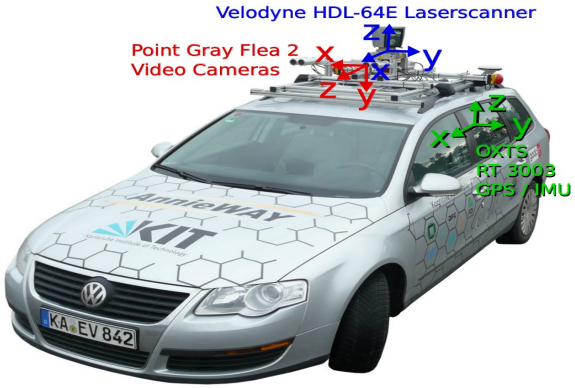


Topic 6 - Aljosa

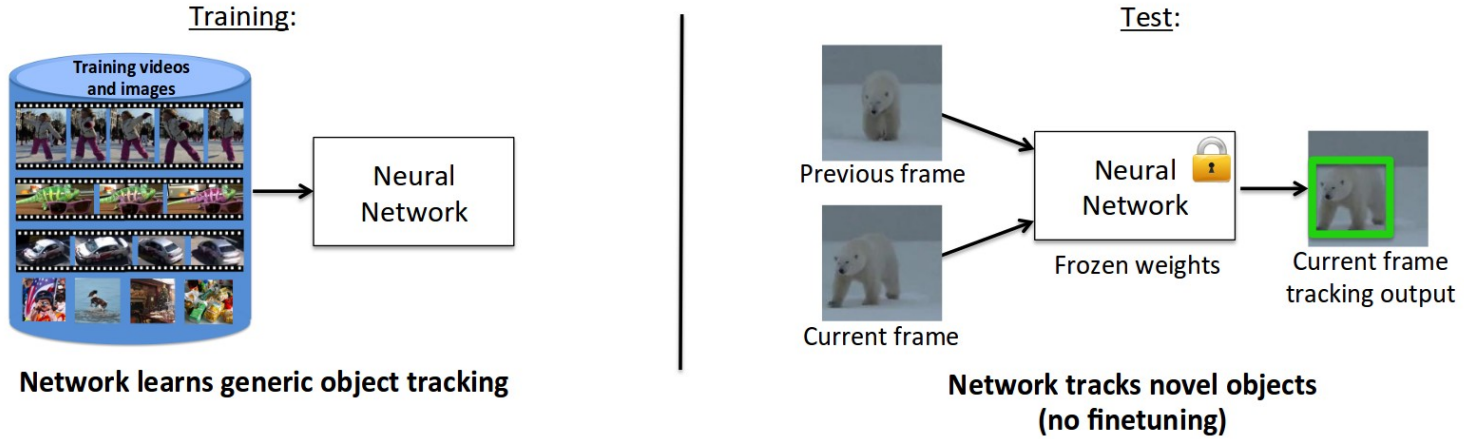
Learning to Track at 100 FPS with Deep Regression Networks

Held et al. (ECCV 2016)

- Single-target generic-object tracking
- Tracker learns generic relationship between object motion and appearance using CNNs
- State-of-the-art + fastest (runs at 100Hz)



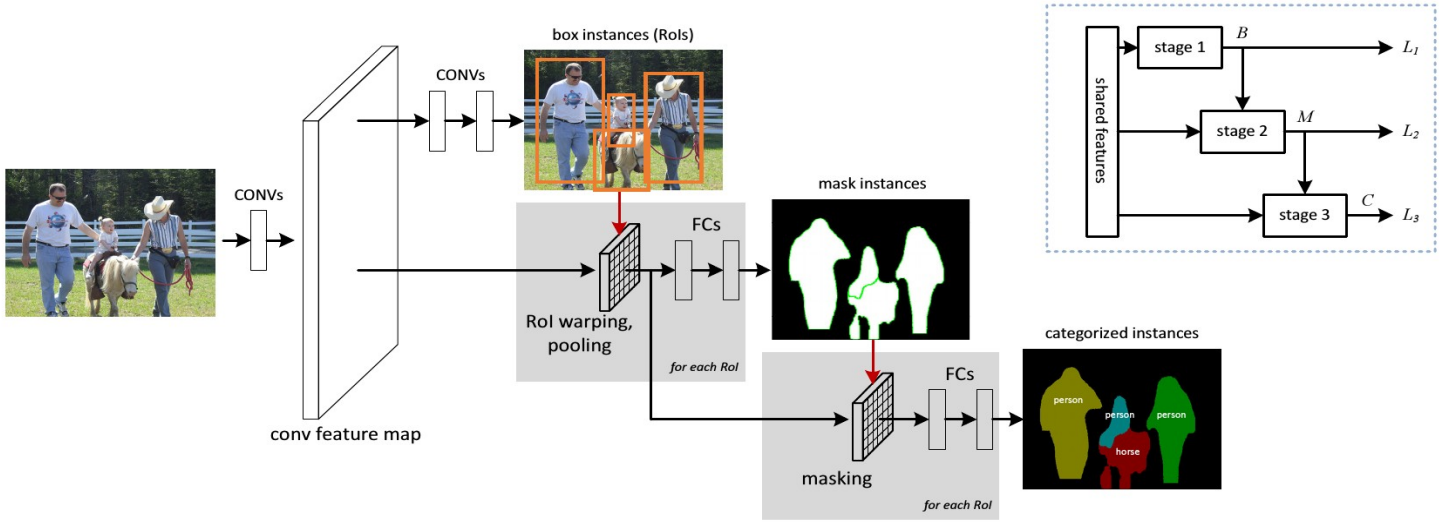
Code available, yay!



Topic 7 - Alex

Instance-aware Semantic Segmentation via Multi-task Network Cascades Dia et al. (CVPR 2016)

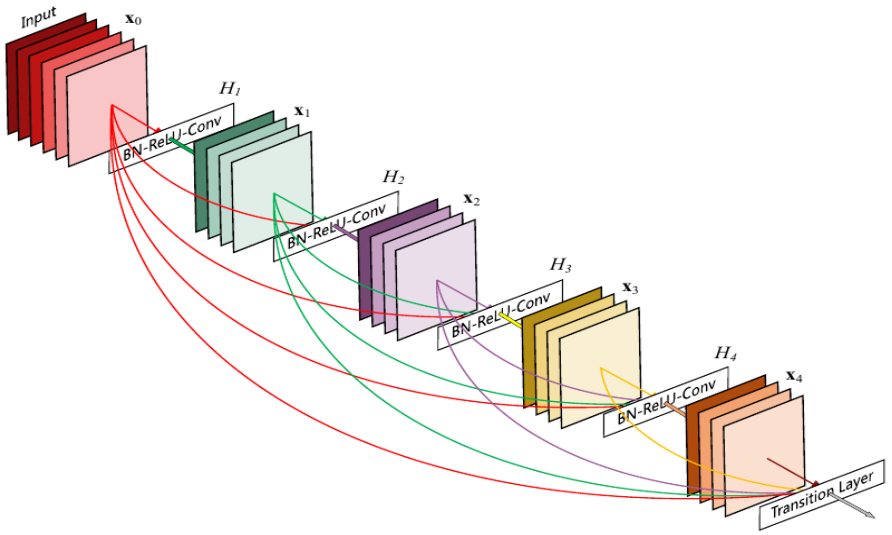
- A system combining CNN based object-detection, segmentation and classification.
- Goes beyond semantic segmentation by detecting object instances not just labels.



Topic 8 - Alex

Densely Connected Convolutional Networks Jhuang et al. (Arxiv 2016)

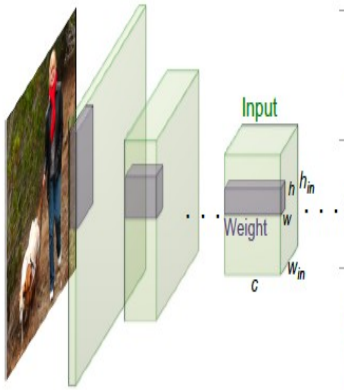
- ResNets, with many skip connections, have recently allowed us to train much deeper networks.
- Densely connected networks go one step further and improve the state-of-the art significantly.



Topic 9 - Umer

XOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks Rastegari et al (ECCV 2016)

- Deep Networks have been a great success for different computer vision problems.
- Need expensive GPU machines to run.
- Unsuitable for mobile devices or situation when less computational power is available.
- Find best efficient approximations e.g making all the weights binary.
- Provides 58x times speed ups and 32x times memory savings.



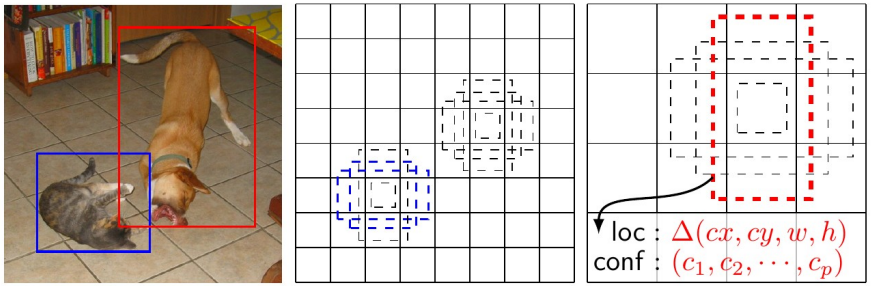
	Network Variations	Operations used in Convolution	Memory Saving (Inference)	Computation Saving (Inference)	Accuracy on ImageNet (AlexNet)
Standard Convolution	Real-Value Inputs 0.11 -0.21 ... -0.34 / -0.25 0.61 ... 0.52 — Real-Value Weights 0.12 -1.2 ... 0.41 / -0.2 0.5 ... 0.68	+ , − , x	1x	1x	%56.7
Binary Weight	Real-Value Inputs 0.11 -0.21 ... -0.34 / -0.25 0.61 ... 0.52 — Binary Weights 1 -1 ... 1 / -1 1 ... 1	+ , −	~32x	~2x	%56.8
BinaryWeight Binary Input (XNOR-Net)	Binary Inputs 1 -1 ... -1 / -1 1 ... 1 — Binary Weights 1 -1 ... 1 / -1 1 ... 1	XNOR , bitcount	~32x	~58x	%44.2

Topic 10 - Ishrat

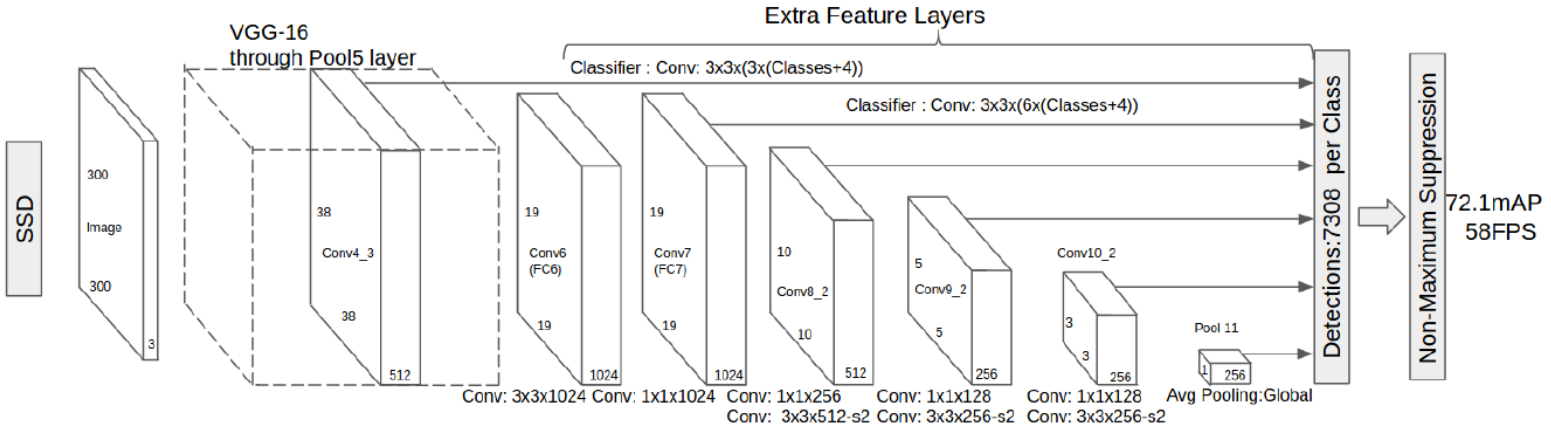
SSD: Single Shot MultiBox Detector

Liu et al. (ECCV 2016)

- CNN for object recognition and localization using separate filters for different aspect ratios and scales on multiple feature maps.

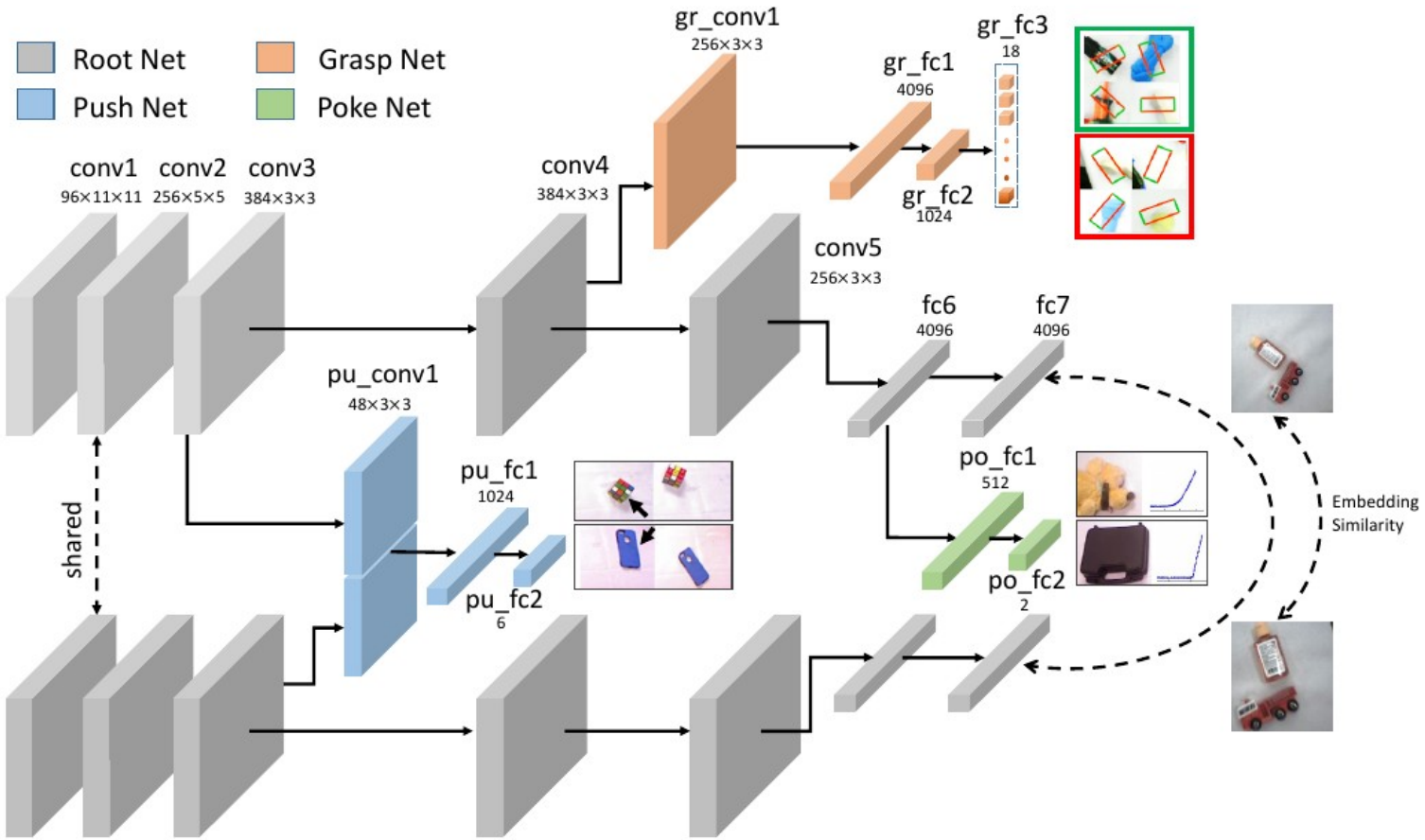


(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map



Topic 11 - Ishrat

The Curious Robot: Learning Visual Representations via Physical Interactions Pinto et al. (ECCV 2016)

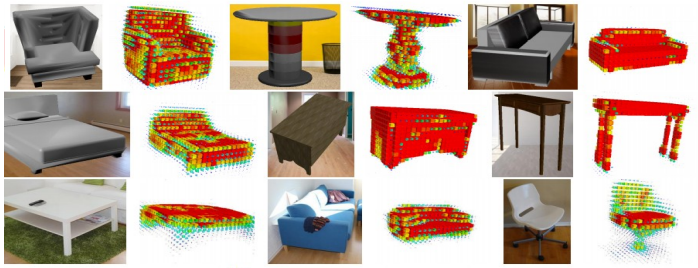


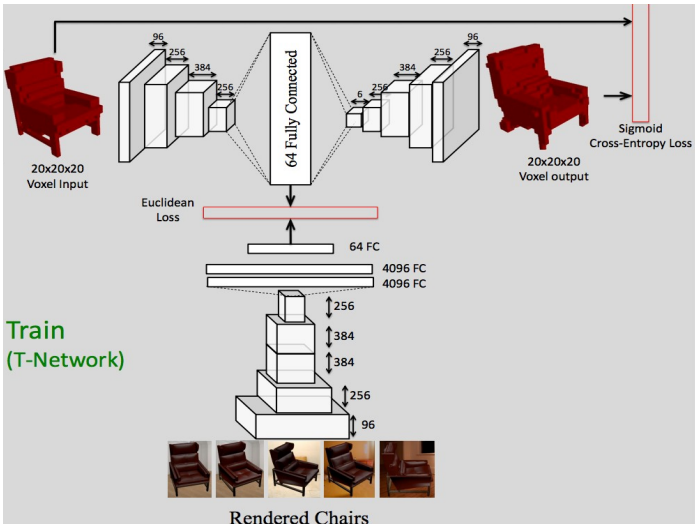
Topic 12 - Francis

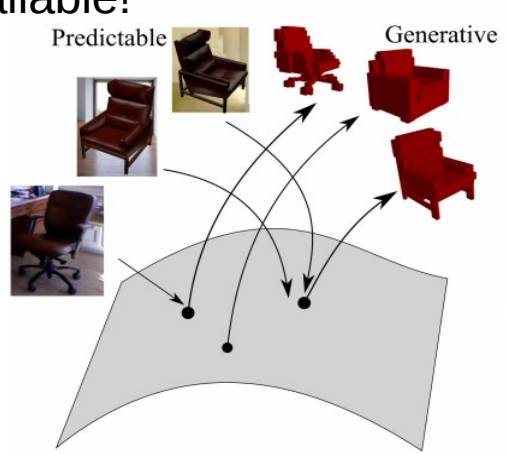
Learning a Predictable and Generative Vector Representation for Objects Girdhar et al. (ECCV 2016)

- What is a **good representation for objects?**
  1. Representable in 3D space
  2. Perceivable from 2D images
- Learn an **embedding space!**

  Combine autoencoder with ConvNet
- Code Available!





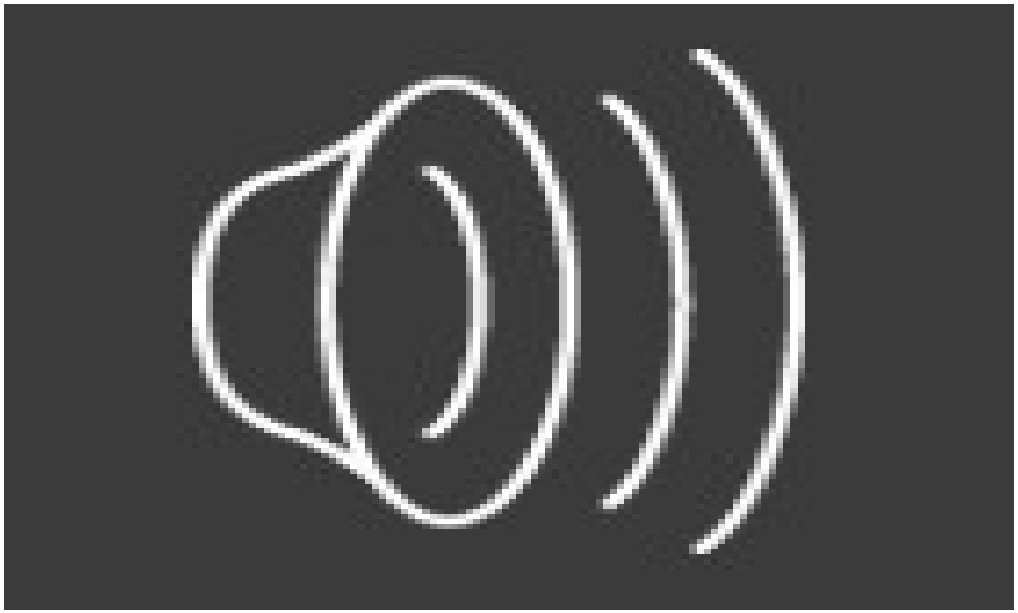


Topic 13 - Francis

Real-Time Monocular Segmentation and Pose Tracking of Multiple Objects Tjaden et al. (ECCV 2016)

Real Time system(50-100Hz) for Multiple object **tracking** and **segmentation**.

Novel approach for fast pose optimization.

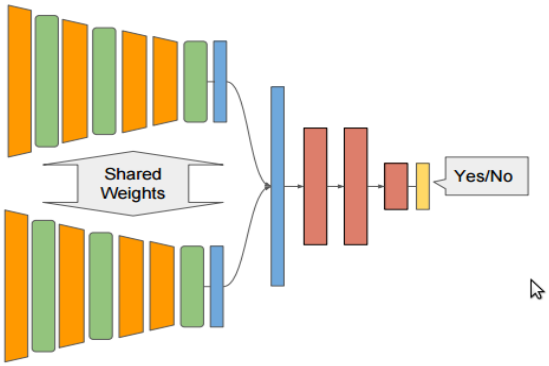


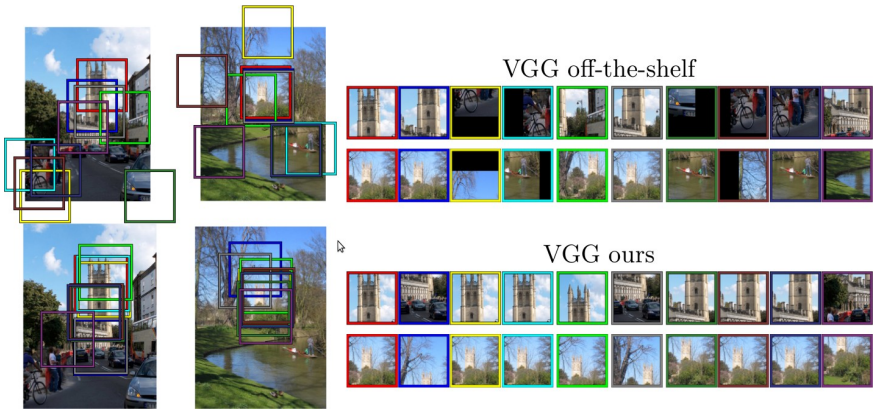
Topic 14 - Dora

CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples Radenovic et al. (ECCV 2016)

- Image Retrieval (returning a ranked list of relevant to a query image)
- Based usually on BOW + spatial verification or off-the-self CNNs
- Use of CNN siamese networks to learn features representation
- Hard positive and negative examples for training

Data and nets are available!!

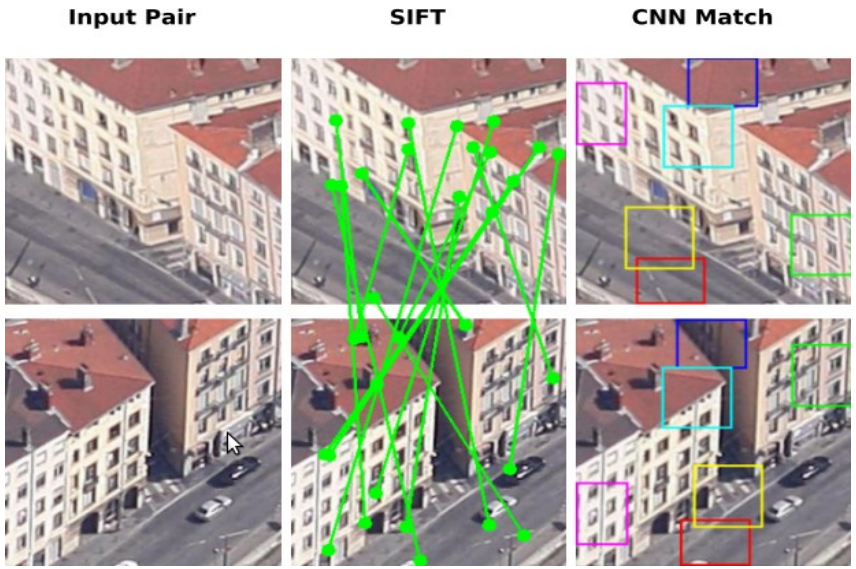


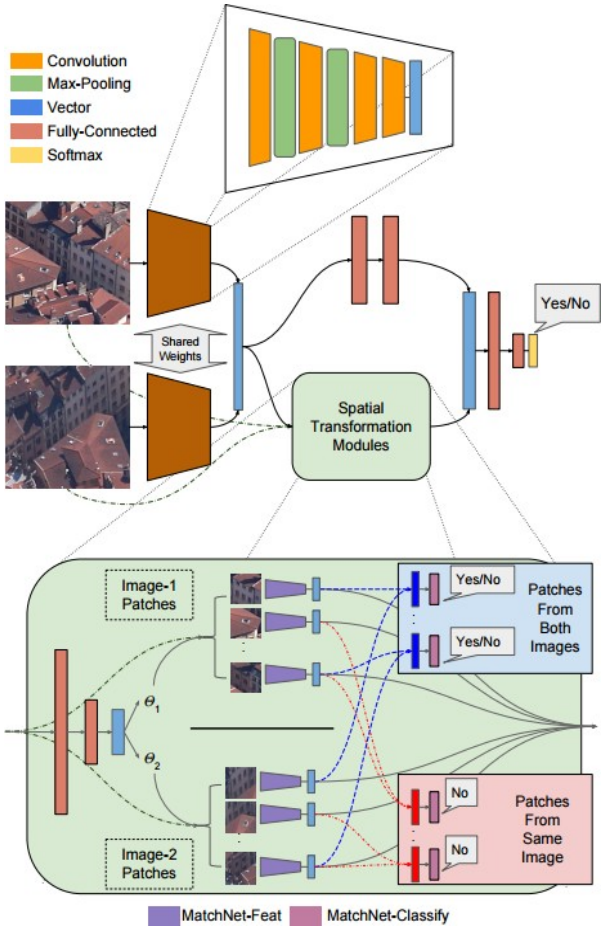


Topic 15 - Dora

Learning to Match Aerial Images with Deep Attentive Architectures Altwaijry et al. (CVPR 2016)

- Wide baseline image matching
- Siamese/Classification network
- Spatial Transformer networks as attention mechanism



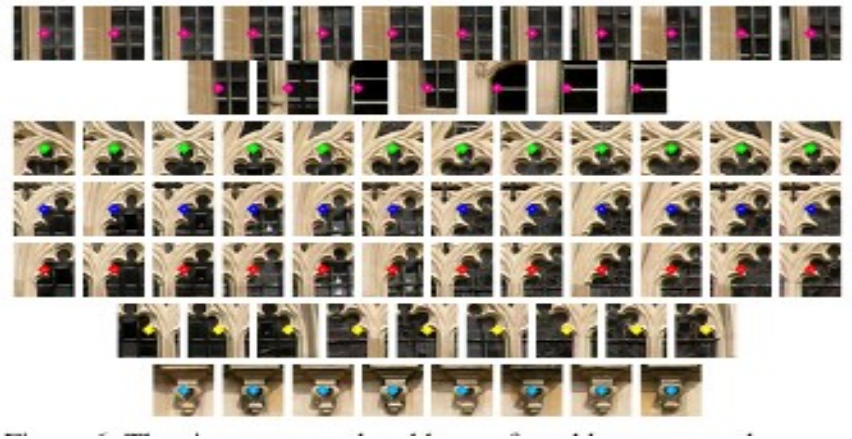


Topic 16 - Michael

Early Burst Detection for Memory-Efficient Image Retrieval

Shi et al., CVPR 2015

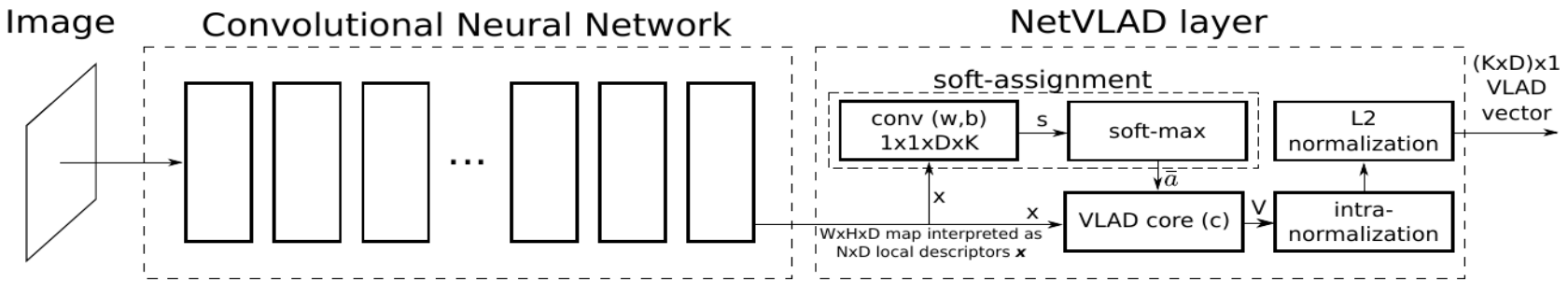
- Certain local features occur in high frequencies in images (bursts)
- idf normalization can only compensate a certain amount
- Early detection of bursts and aggregation into meta-features



Topic 17 - Michael

NetVLAD: CNN architecture for weakly supervised place recognition Arandjelović et al., CVPR 2016

- End-to-end trainable CNN for visual place recognition
- Use custom-built layer and propose a triplet ranking loss
- State-of-the-art results on retrieval and place recognition benchmarks
- Code available as open source



Assignment

- Pick three topics you might find interesting
  - No preference, just pick three
- Then we assign the topics
  - *Which algorithm can we use?*
- Let us quickly review the topics...
  - Remember the numbers!