Seminar Current Topics in Computer Vision and Machine Learning

Seminar Important Developments in Computer Vision and Machine Learning

Kickoff Meeting

18.10.2019

Prof. Dr. Bastian Leibe

RWTH Aachen University, Computer Vision Group
http://www.vision.rwth-aachen.de

Visual Computing Institute | RWTH AACHEN UNIVERSITY

Two Seminars

- Seminar “Current Topics in CV + ML”
  - Participants: Master students
- Seminar “Important Developments in CV + ML”
  - Participants: Bachelor students
- Organization
  - The seminars will be co-located
  - Joint event with seminar talks from both groups
  - Difference will be in the expectations we have of you
    - Master students: already familiar with CV/ML concepts
    - Bachelor students: often first encounter with CV/ML papers
  - In all cases, you should be familiar with the basics of ML

  ⇒ *If you haven’t already, take the ML lecture offered this semester!*

Organization

- Reports
  - English or German (depending on the supervisor)
  - ≥15 pages (but no more than 20)
    - Bibliography counts – TOC does not
    - Don’t use excessive white space or layout tricks to get more pages
  - LaTeX is mandatory

- Presentations
  - In English
  - 30-40 minutes
  - Block event at the end of the semester – 3 days of presentations
  - Slide Templates will be available on the webpage
  - Laptop can be provided for the presentation, if necessary

Schedule

- Deadlines
  - Hand in signed Declaration of Compliance (hardcopy – before outline)
  - Outline: Monday, Dec 2nd
  - Report: Monday, Jan 6th – graded version!
  - Slides: Monday, Jan 27th
  - Presentations: 3 days in the week of Feb 03-05 (block event)
  - Turn in corrected report at presentation day

Hints for Your Report – DOs

- Content
  - Read and understand your paper
  - Search for additional literature
  - Take part in a library tour (if you haven't already)
  - 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
- Form
  - Write a report in your own words
  - 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 if you exceed the deadline, up to failing the seminar!

- 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 if possible – 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 cite 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
  - State the name of the seminar in the subject
    (“Current Topics in CV+ML” / “Important Developments in CV+ML”)

Topic 1 – István

Unsupervised 3D Pose Estimation with Geometric Self-Supervision
Chen et al. (Amazon, Georgia Tech), CVPR 2019

Task: 2D human pose ➞ 3D human pose (“pose lifting”)

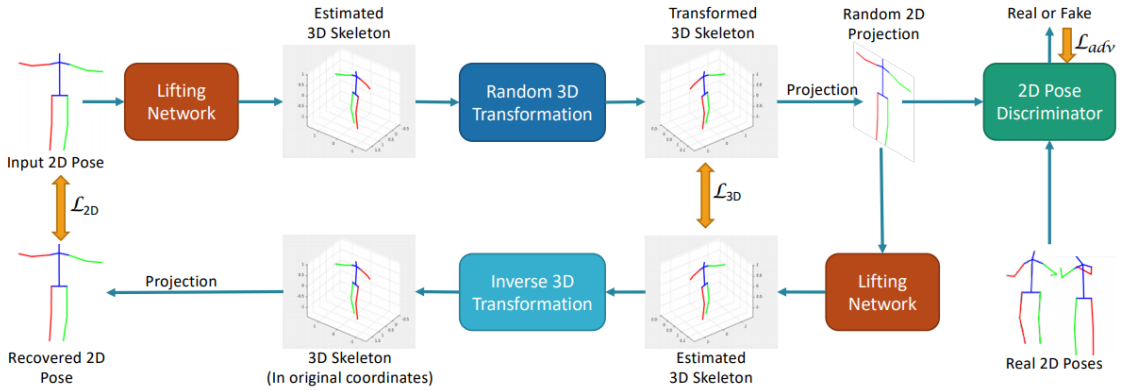
The general framework of decoupled 3D human pose estimation is

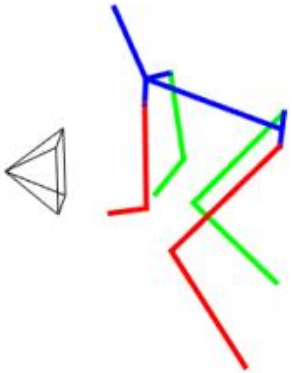
1) RGB image ➞ 2D pose (e.g. OpenPose)

2) 2D pose ➞ 3D pose (e.g. by regression)

However, labels are scarce for 3D, but widely available for 2D keypoints

Could we learn the 2D-to-3D “lifting” entirely from 2D data, never observing 3D annotations?





Topic 2 – István

Learnable Triangulation of Human Pose

Iskakov et al. (Samsung AI), ICCV 2019

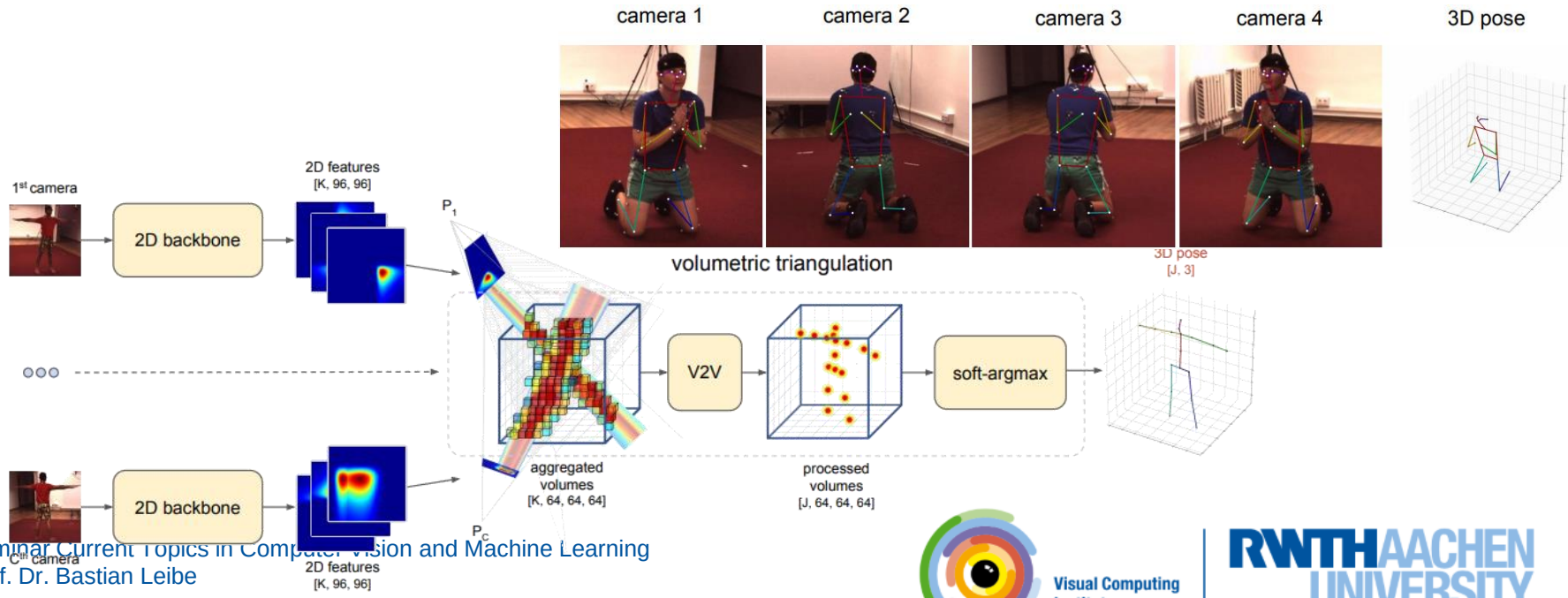
Task: Calibrated multi-view RGB ➞ 3D pose (“markerless motion capture”)

Baseline: Predict 2D keypoints in each view and *then* combine them by triangulation

This uses very limited info from each view (just points) and combines them purely by geometry

How could we *first* combine rich information from all views and *then* predict plausible 3D poses?

End-to-end learnable, so standard deep nets can be applied (e.g. ResNet)



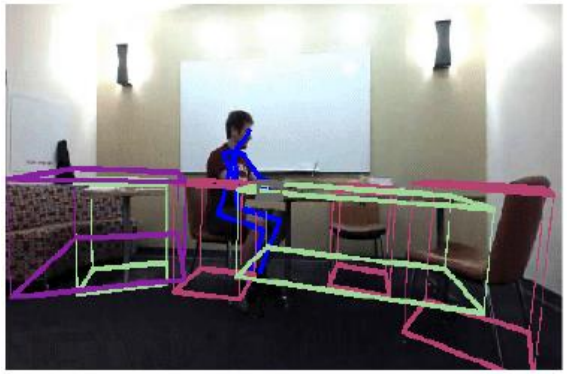
Topic 3 – István

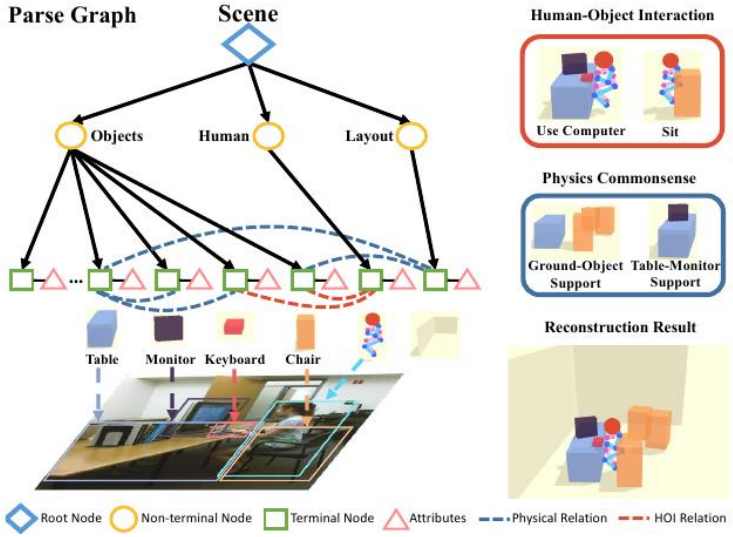
Holistic++ Scene Understanding: Single-view 3D Holistic Scene Parsing and Human Pose Estimation [...]
Chen et al. (UCLA), ICCV 2019

Task: RGB image ➞ 3D human poses + parsed 3D scene

Most pose estimation works consider people in isolation

How could we take into account **scene constraints** and **human-object interactions**?

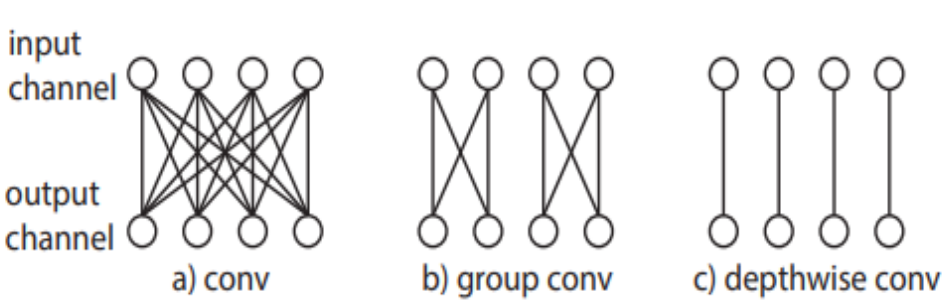


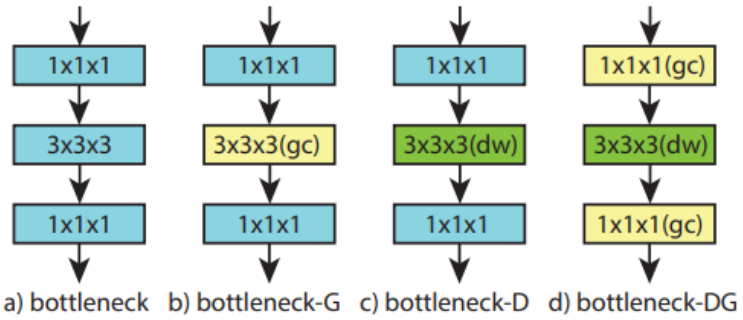


Video Classification with Channel-Separated Convolutional Networks

Du Tran, Heng Wang, Lorenzo Torresani, Matt Feiszli, ICCV'19 (Facebook AI)

- 3D Convolutions are computationally expensive.
- Group convolutions save computational cost in 2D. What are their effects in 3D convolutional networks?

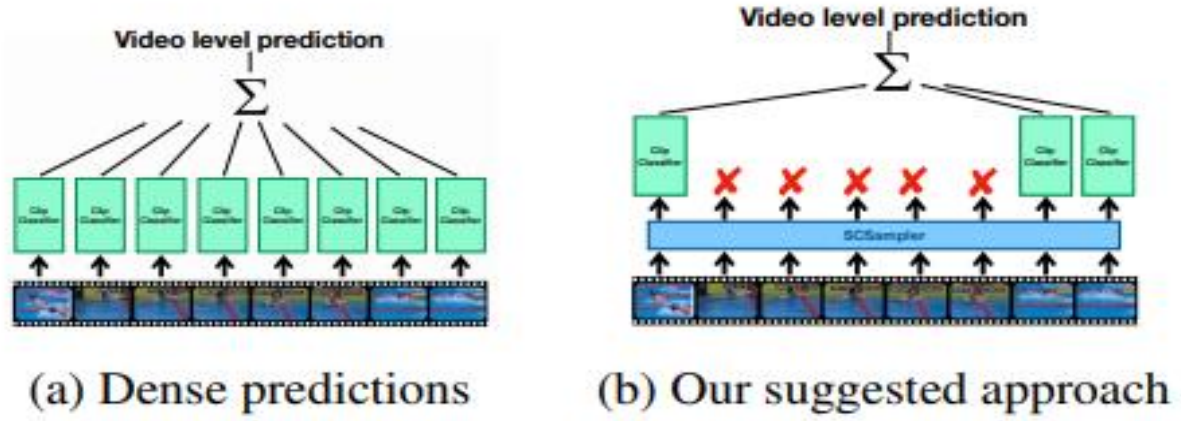




SCSampler: Sampling Clips from Video for Efficient Action Recognition

Bruno Korbar, Du Tran, Lorenzo Torresani, ICCV'19

- Processing large video clips are expensive, and often limited by GPU memory.
- Some of the frames within a video could be irrelevant for the task at hand.
- SCSampler learns to select salient clips from a large video.
- Uses a set of Video and Audio sampler.

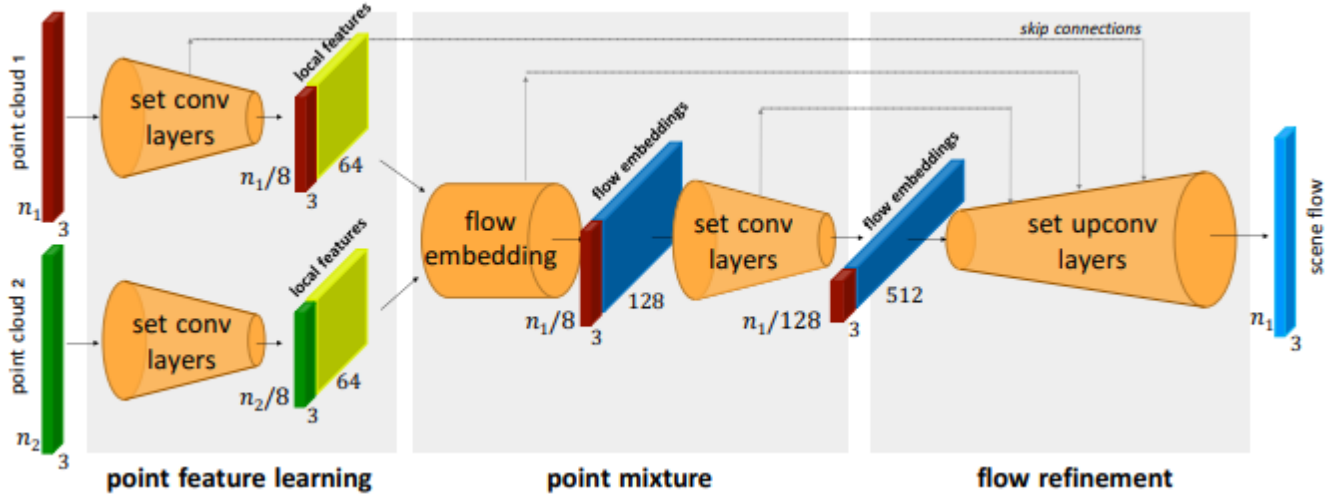


(a) Dense predictions (b) Our suggested approach

FlowNet3D: Learning Scene Flow in 3D Point Clouds

Xingyu Liu, Charles R Qi, Leonidas J. Guibas, CVPR'19 (Stanford University, Facebook AI Research)

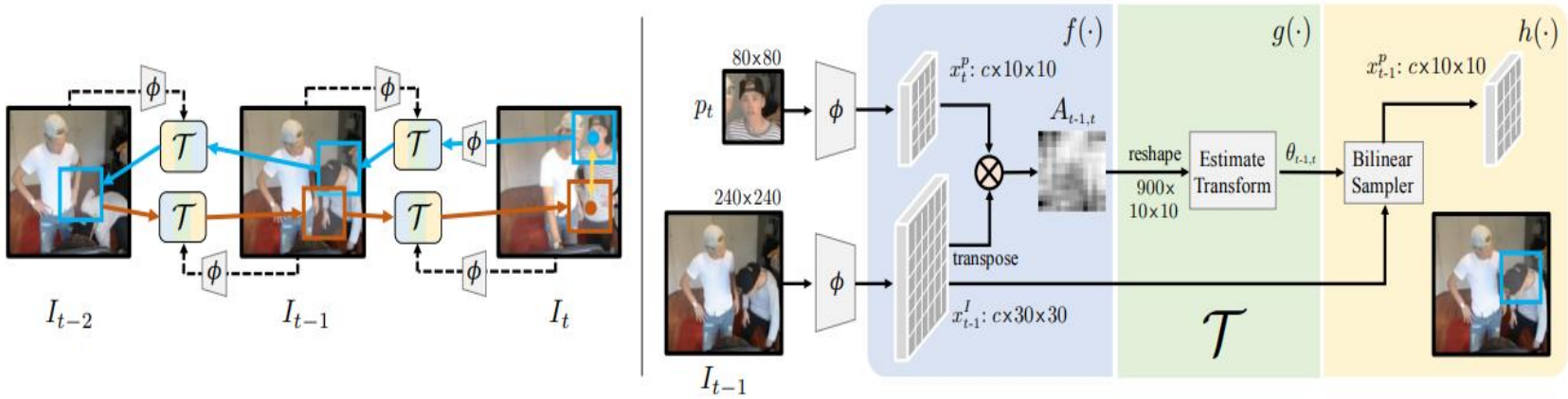
- End-to end learning of scene flow from point clouds.
- Uses 3 layers: set conv layers(PointNet++), flow embedding layer, and upsampling layers



Learning Correspondence from the Cycle-consistency of Time

Xiaolong Wang, Allan Jabri, Alexei A. Efros, CVPR'19

- Self-supervised learning of visual correspondences.
- Uses cycle consistency over time in a video as supervisory signal
- Applied for multiple tasks such as mask propagation, pose tracking, optical flow etc.



Topic 8 – Paul

DeepMOT: A Differentiable Framework for Training Multiple Object Trackers

Xu et al., ArXiv 2019

Task: Multi-Object Tracking (MOT)

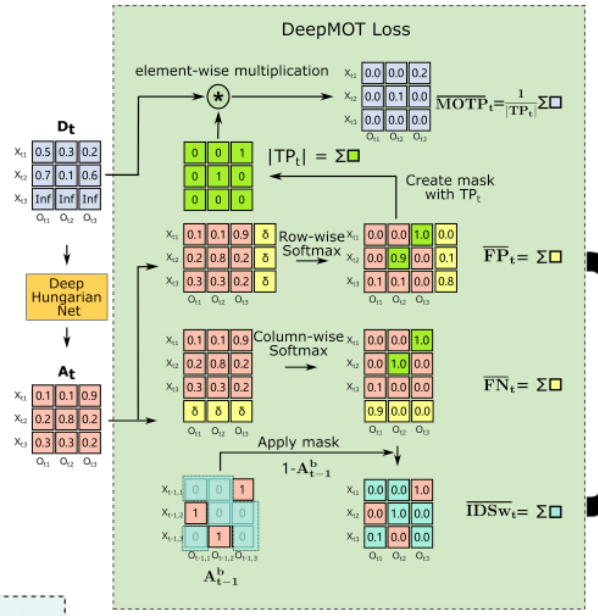
Evaluation criteria MOTA and MOTP non-differential

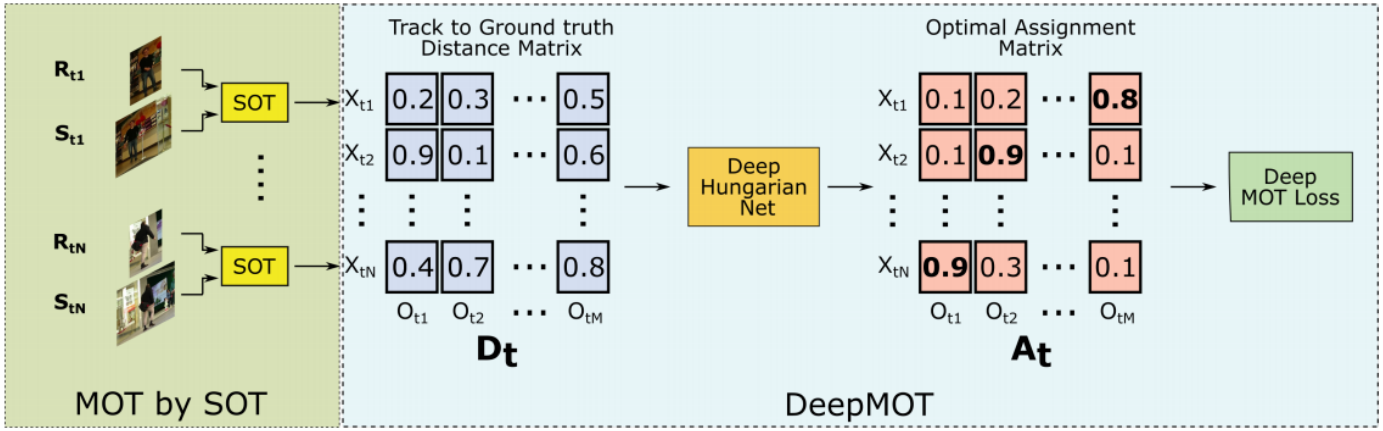
Use differentiable proxy to train end-to-end

Multi-Object Tracking by Single-Object Tracking + Matching

Replace Hungarian Algorithm by Deep Hungarian Net

Bidirectional RNNs

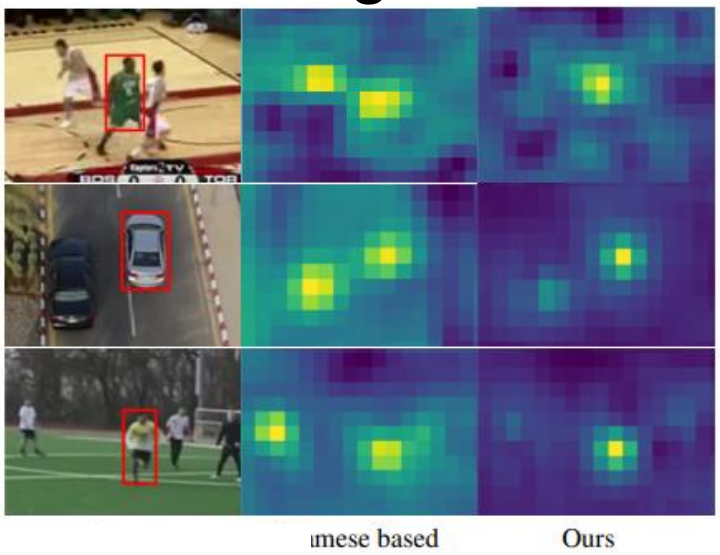




Topic 9 – Paul

Learning Discriminative Model Prediction for Tracking

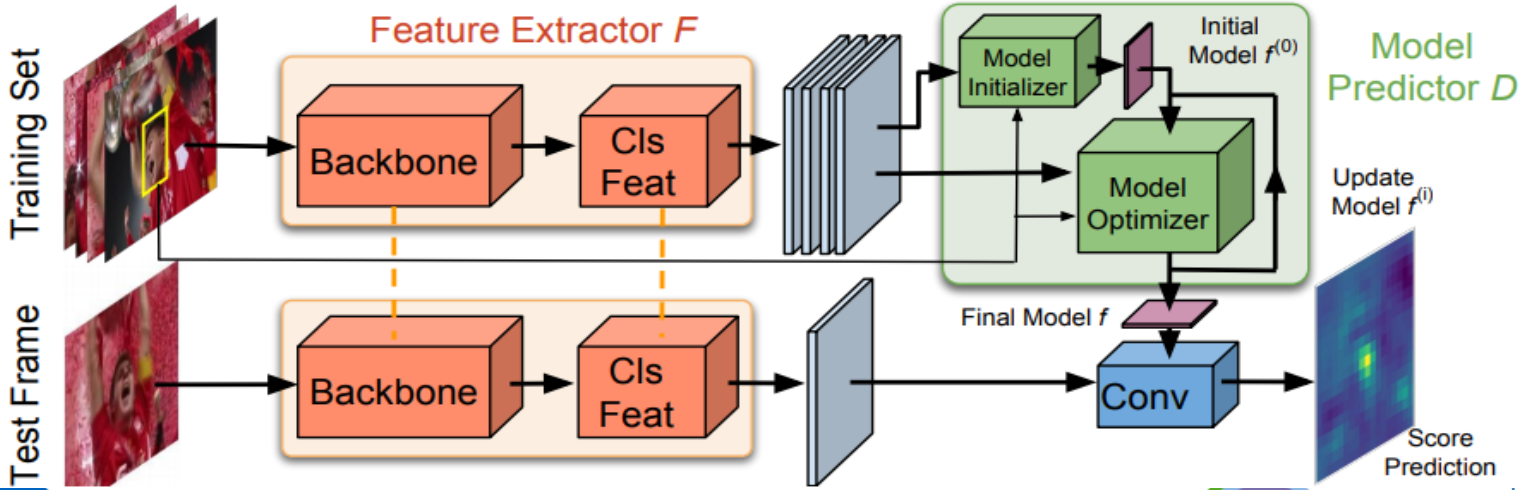
Bhat et al., ICCV 2019



Task: Single-Object Tracking

Current approaches extract template based on first-frame ground truth bounding box but neglect background

Meta-learning: Learn model predictor which at test time predicts model parameters for tracking



Tracking Holistic Object Representations (THOR)

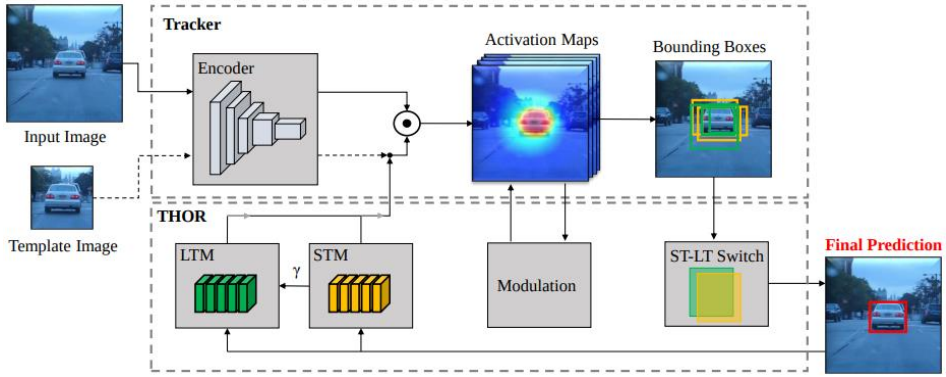
Sauer et al., BMVC 2019, Best Science Paper Award

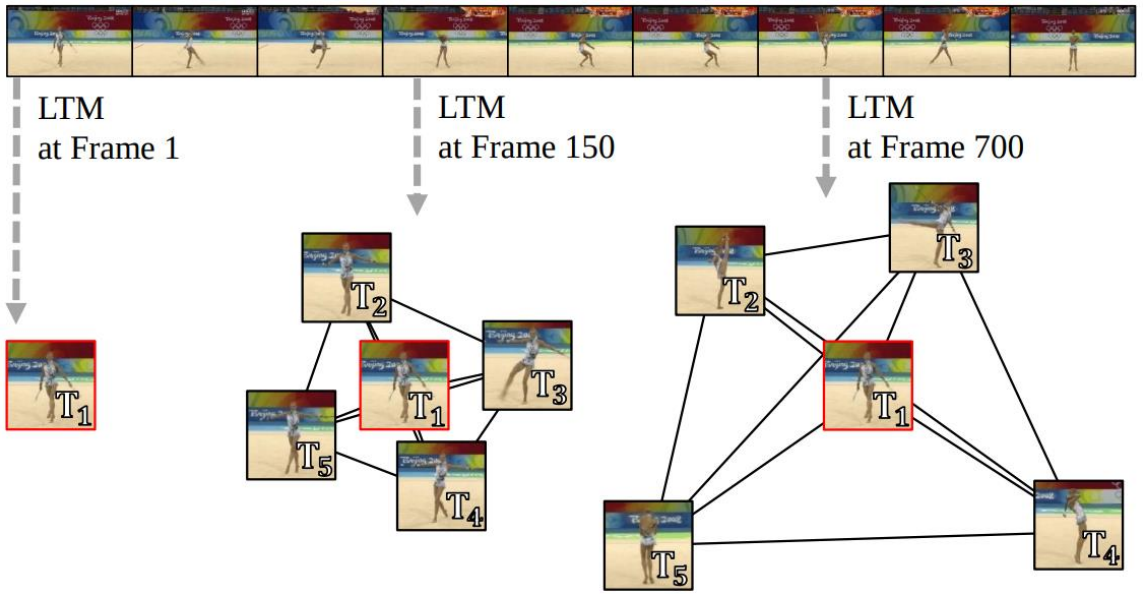
Task: Single-Object Tracking

Current approaches: use only first-frame ground truth box as template

THOR: use detected boxes as additional templates, subselect templates

Long-term module and short-term module





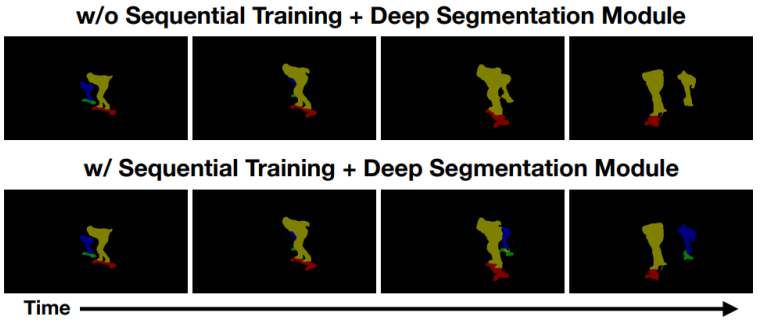
Going Deeper into Embedding Learning for Video Object Segmentation

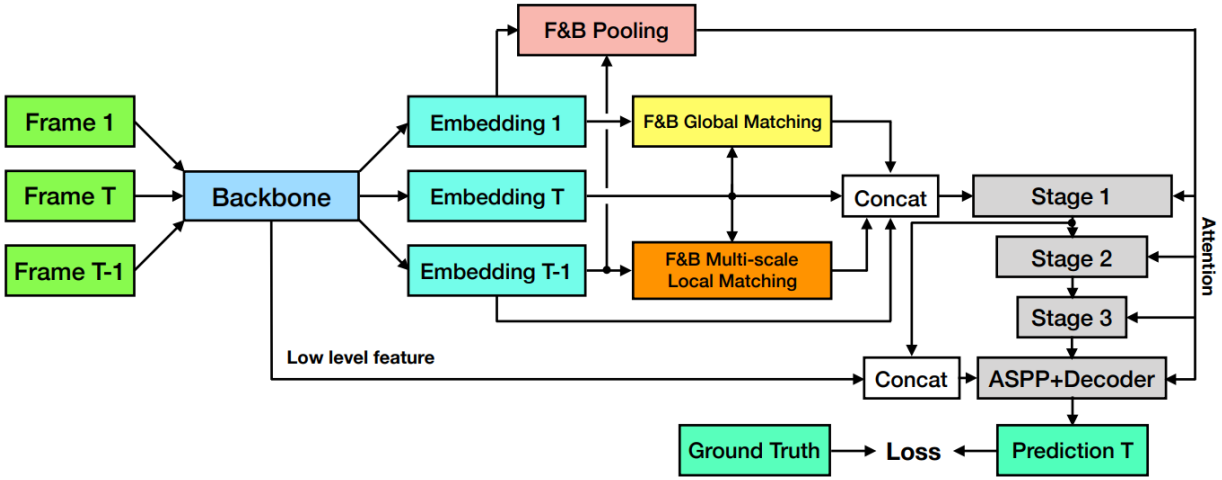
Yang et al., ICCV Workshop 2019

Task: Video Object Segmentation

Based on “Fast End-to-end Embedding Learning for Video Object Segmentation

Various Extensions for YouTube-VOS competition, won 3rd place





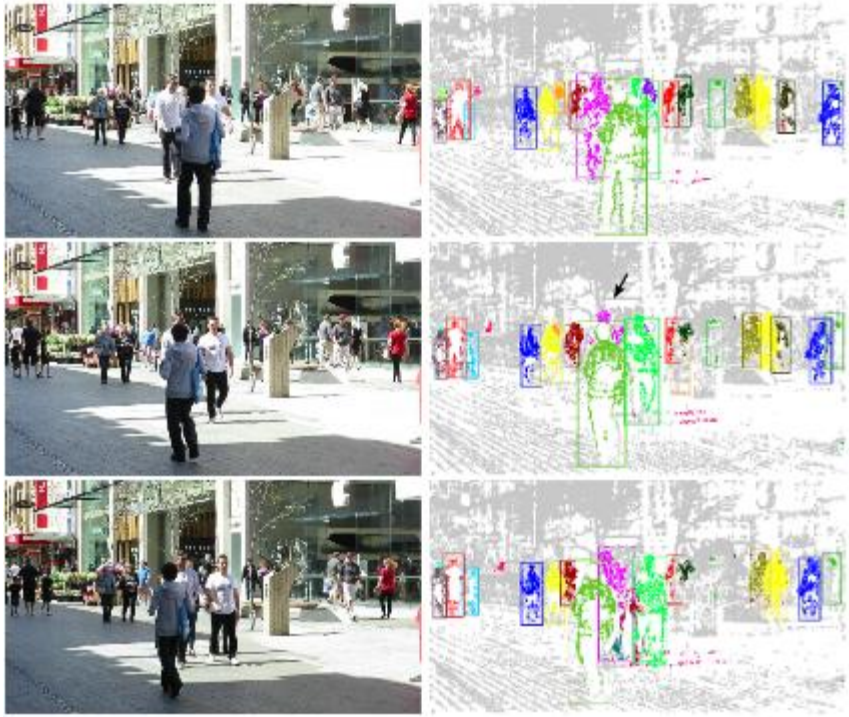
Approach	score	boost
FEELVOS (after adjusting hyper-parameters for YouTube-VOS)	75.1%	-
+ Foreground and Background Matching	76.2%	1.1%
+ Foreground and Background Attention	77.1%	0.9%
+ Balanced Random Crop	78.4%	1.3%
+ Deep Segmentation Module	79.5%	1.1%
+ Multi-scale Local Matching	80.2%	0.7%
+ Sequential Training	81.0%	0.8%
+ Multi-scale & Flip in Testing	82.4%	1.4%

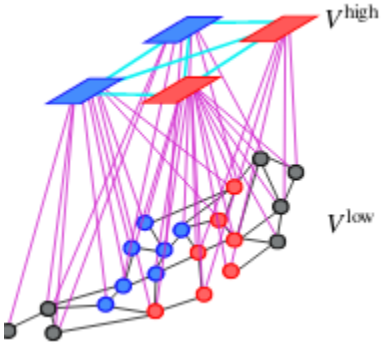
Table 1. The ablation study experiments on the validation set of YouTube-VOS 2019.

Motion Segmentation & Multiple Object Tracking by Correlation Co-Clustering

Margret Keuper, Siyu Tang, Bjoern Andres, Thomas Brox, Bernt Schiele, PAMI'18

Task: Tracking multiple objects in videos.
Novelty: Combining bottom-up and top-down approaches.
Jointly optimized using correlation co-clustering.
Evaluated on Multi-Object Tracking Challenge: dataset containing videos with several persons.



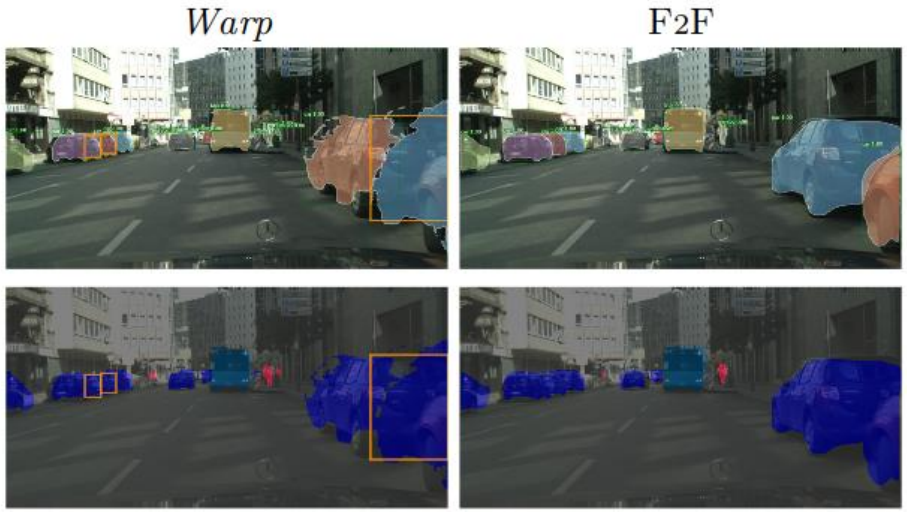


Topic 13 – Ali

Predicting Future Instance Segmentation by Forecasting Convolutional Features

Pauline Luc, Camille Couprie, Yann LeCun, Jakob Verbeek, ECCV'18

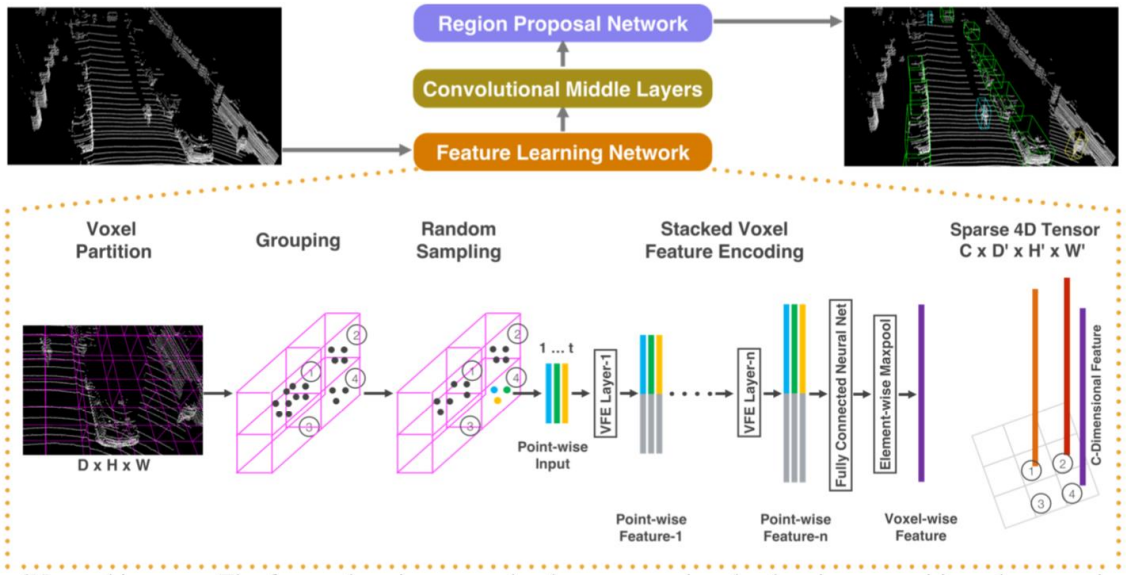
Task: Given a video sequence, predicting pixel masks for object instances in the future.
Novelty: Predict the feature maps for future image frames rather than directly predicting the pixel masks.
Autoregressive property: can feed the output of the network back as input to get predictions further in the future.

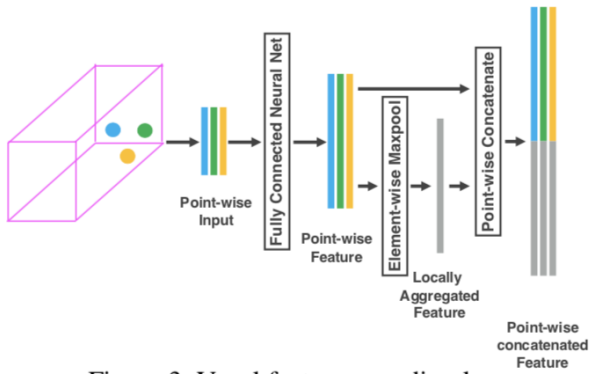


VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection

Yin Zhou and Oncel Tuzel, CVPR'18

- One of the pioneering approaches for 3d object detection on point clouds.
- Used as the backbone for many newer networks (cited by 283).

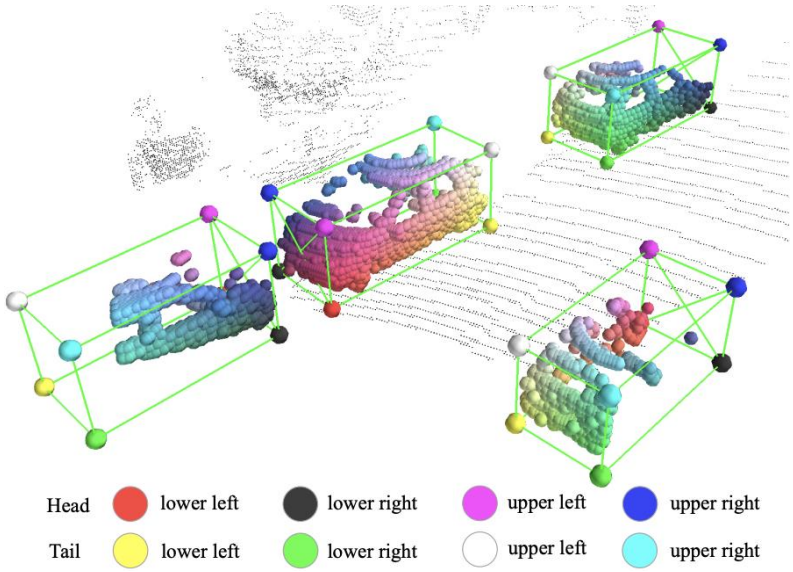


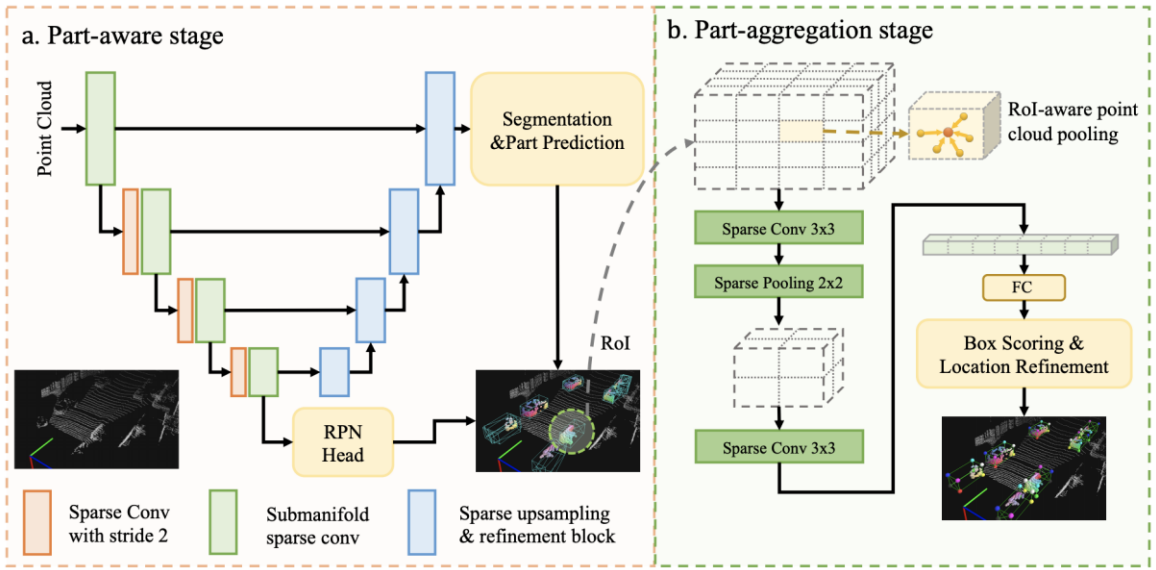


Part-A2 Net: 3D Part-Aware and Aggregation Neural Network for Object Detection from Point Cloud

Shaoshuai Shi, Zhe Wang, Xiaogang Wang, Hongsheng Li, arXiv'19

- 3d object detection on point clouds.
- State-of-the-art on KITTI 3d object detection.





Revealing Scenes by Inverting Structure from Motion Reconstructions

Francesco Pittaluga, Sanjeev J. Koppal, Sing Bing Kang, Sudipta N. Sinha, CVPR'19

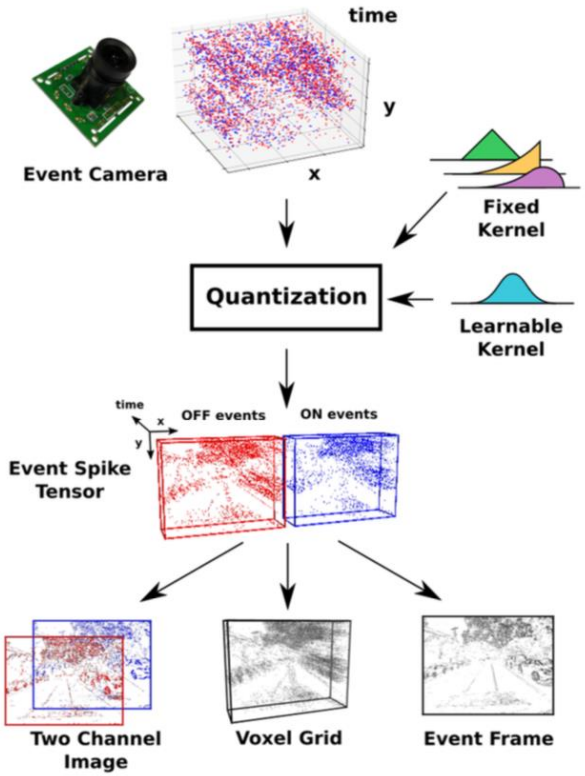
- Structure from motion (SfM) construct point clouds from images.
- Is it possible to invert the process, i.e. synthesis an image from point clouds (color and SIFT descriptor optional)?

(a) SfM point cloud (top view) (b) Projected 3D points (c) Synthesized Image (d) Original Image

End-to-End Learning of Representations for Asynchronous Event-Based Data

Daniel Gehrig, Antonio Loquercio, Konstantinos G. Derpanis, Davide Scaramuzza, ICCV'19

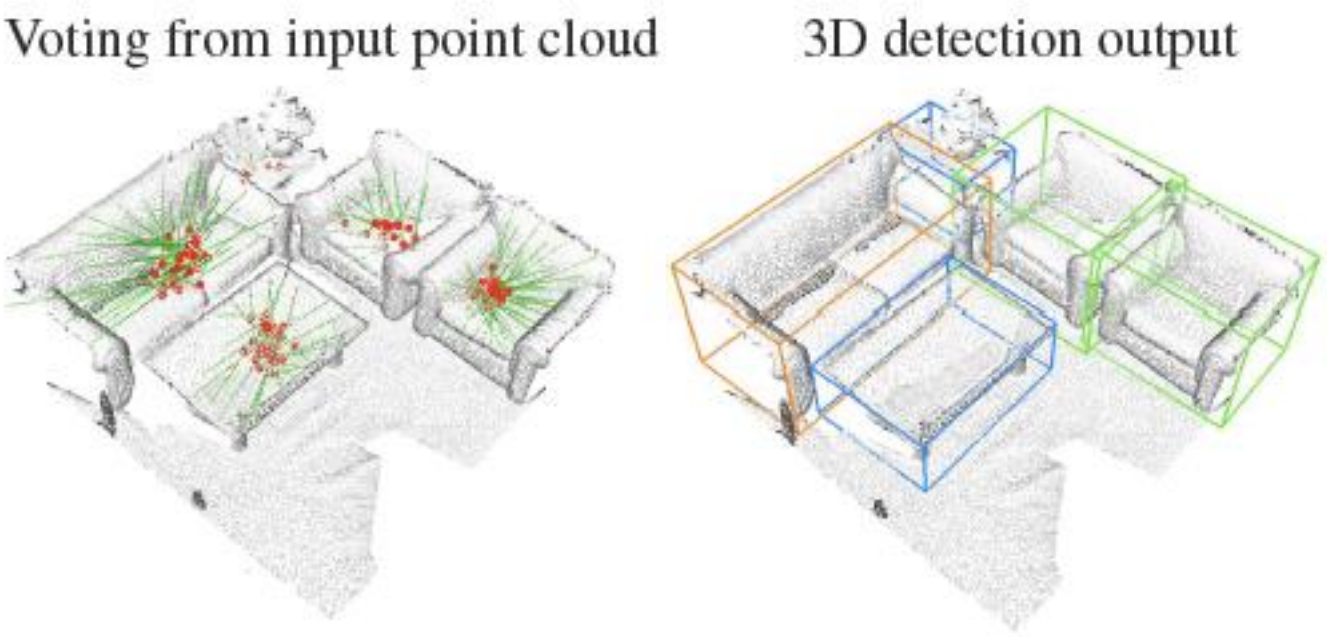
- Event cameras are bio-inspired vision sensors, which produce asynchronous event streams (intensity changes) instead of synchronous intensity measurement (images).
- How to convert such event streams into representations that can be processed by CNNs?



Topic 18 – Francis

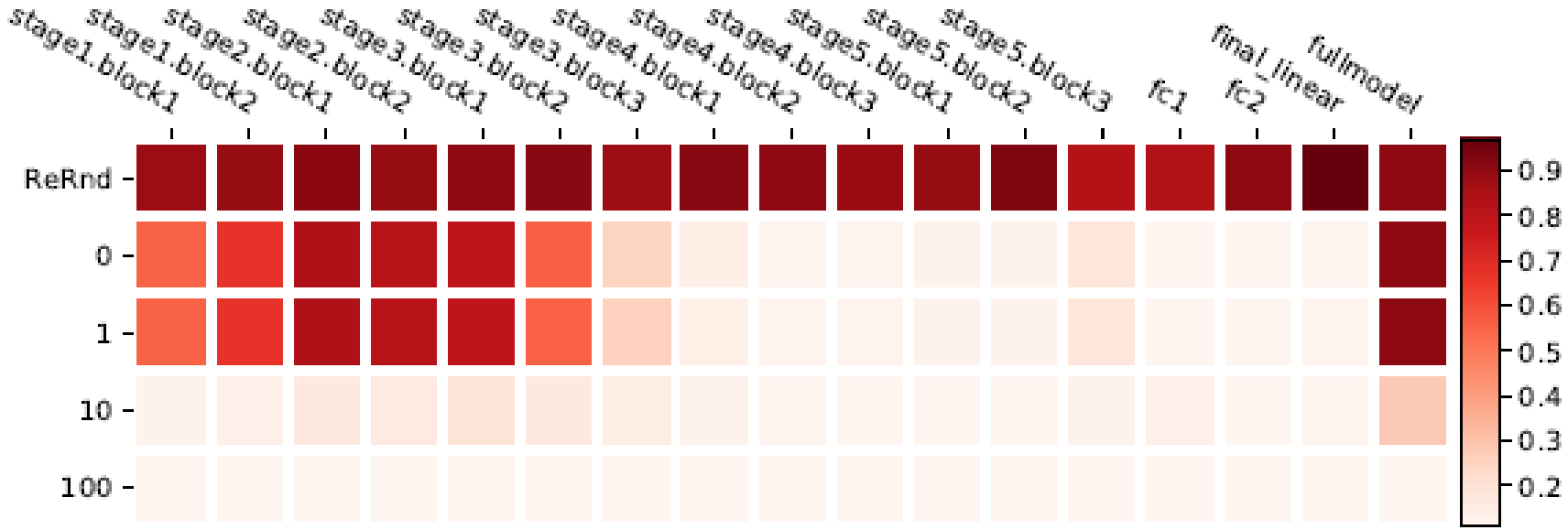
Deep Hough Voting for 3D Object Detection in Point Clouds

Qi, Litany, He, Guibas



Are All Layers Created Equal?

Zhang, Bengio, Singer

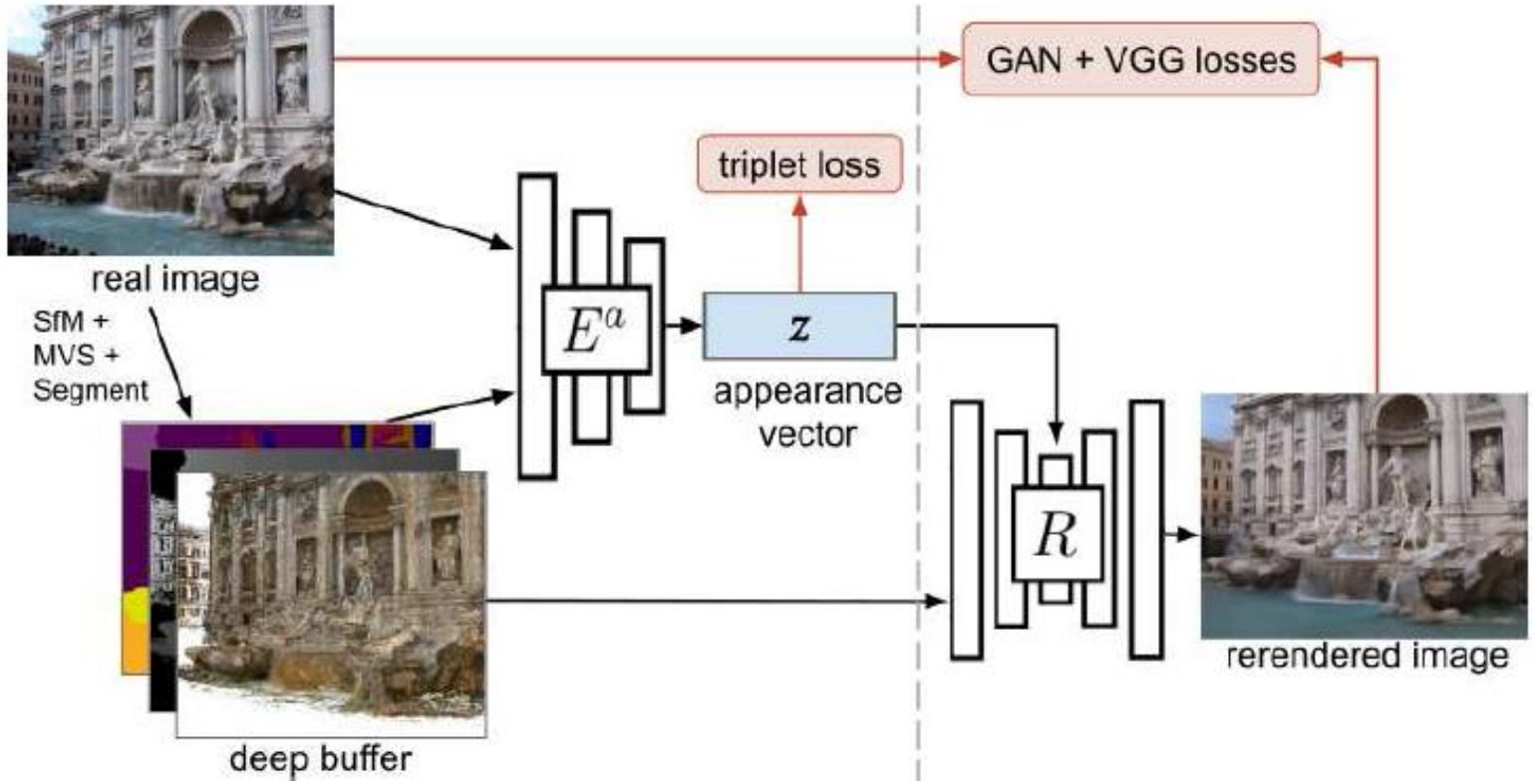


(b) VGG16 on CIFAR10

Topic 20 – Dora

Neural Rerendering in the Wild

Meshry et al., CVPR'19



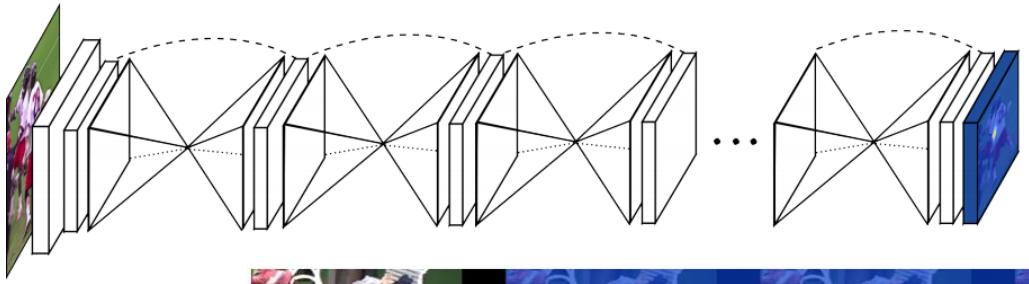
Topic 21 – István

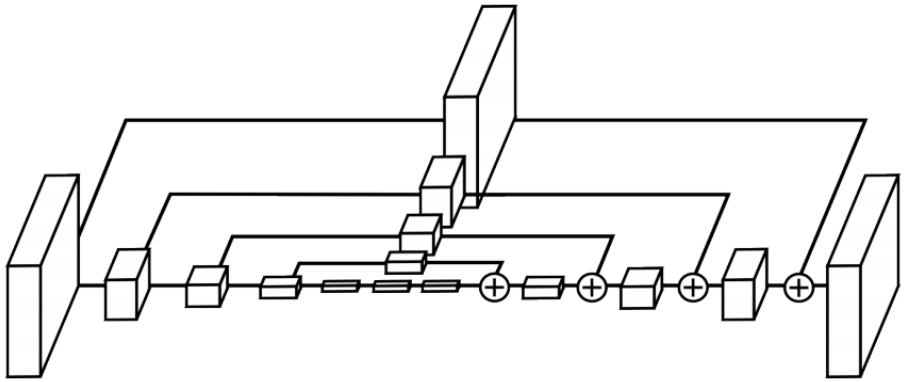
Stacked Hourglass Networks for Human Pose Estimation (Bachelor topic)
Newell et al. (UMichigan), ECCV 2016

Very influential fully-convolutional architecture for keypoint localization, over 1100 citations

Stacked encoder-decoder modules called “hourglasses”

Repeated bottom-up, top-down processing for long-range information aggregation and refinement with intermediate supervision





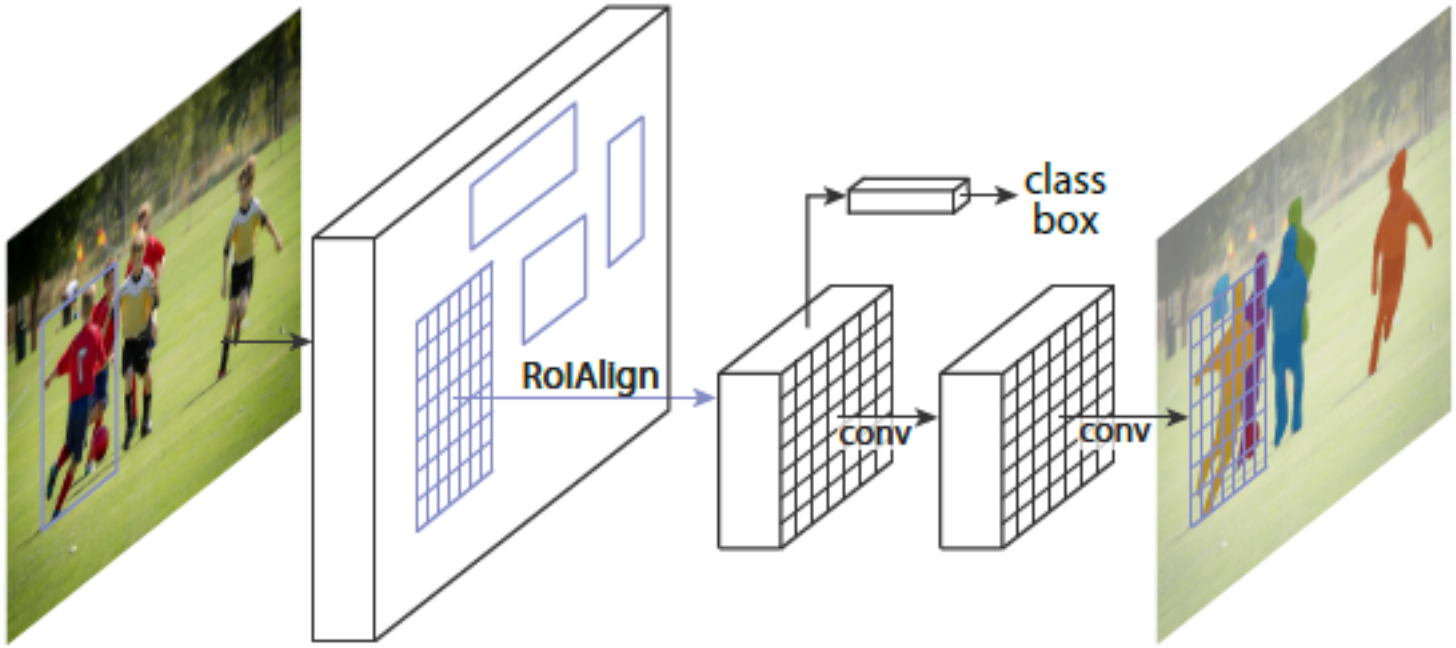
Topic 22 – Ali

Mask R-CNN

He, Gkioxari, Dollar, Girshick

The current state-of-the-art object detection approach

Hugely influential paper

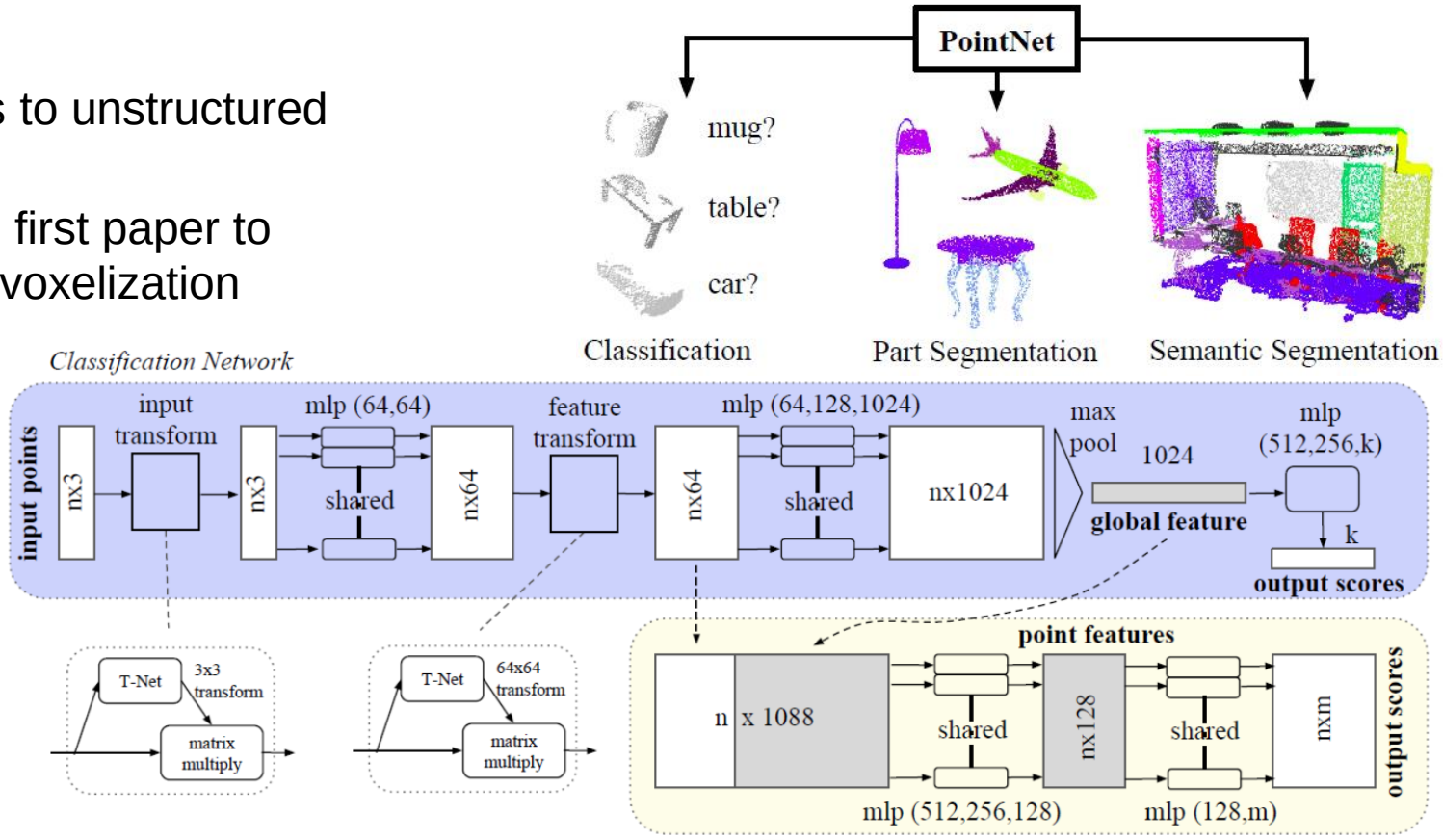


Topic 23 – Francis

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

C.R. Qi, H. Su, K. Mo, L. Guibas, CVPR'17

- Apply deep networks to unstructured point clouds
- Very influential work, first paper to address this without voxelization



Topic Assignment

- Pick three topics you might find interesting
  - No preference, just pick three
- Then we assign the topics
- We will quickly review the topics again...
  - Remember the numbers!