ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CVL Computer Vision Lab

Image Communication and Understanding (Current Trends)

Danda Pani Paudel
Computer Vision Lab @ ETH Zurich

Computer Vision and Image Understanding

Prof. Dr. Luc Van Gool - https://www.vision.ee.ethz.ch

QUALITY
Image enhancement and super-resolution

TRACE
Toyota Research on Automated Cars in Europe

TRACKING
Benchmark fast trajectory annotation with path supervision

Input: Path supervision

VARCITY
Semantic and dynamic city modelling-building a city from images

RECOGNITION
Detecting and segmenting objects in images and videos

MOBILE BASED 3D
Mobile + Cloud generated 3D for Buildings

Deep Image Enhancement

- Make pictures taken by a cheap smartphone camera look as if taken by a DSLR!
- Transform 3D sequences into photorealistic movies

Bad quality images   Enhanced
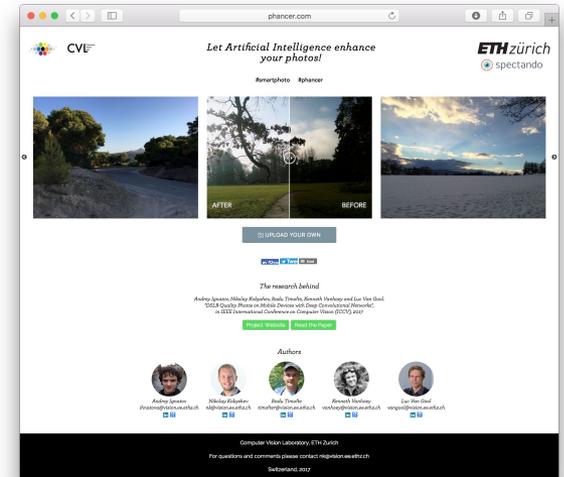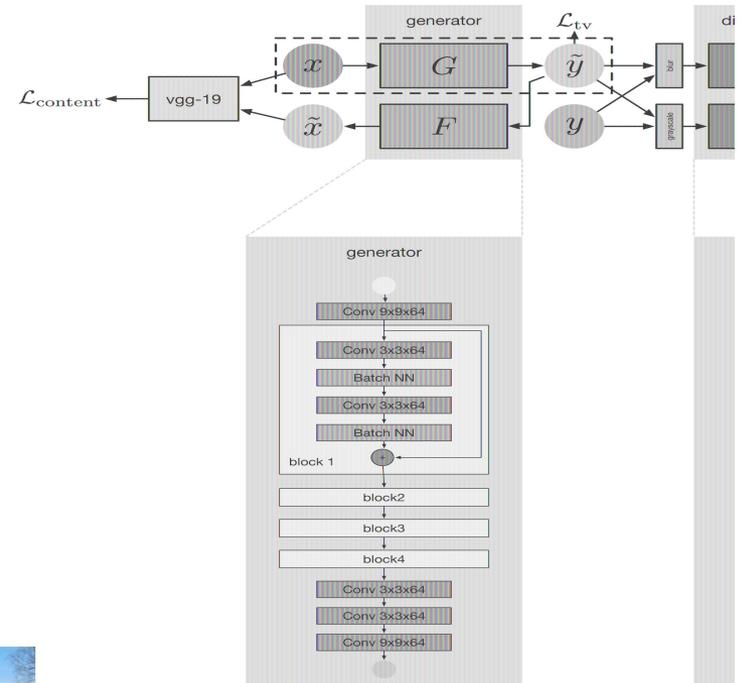
Unsupervised Image Super-Resolution

- Learn the downsampling operation with unpaired images
- Train the super-resolution network on the generated paired data
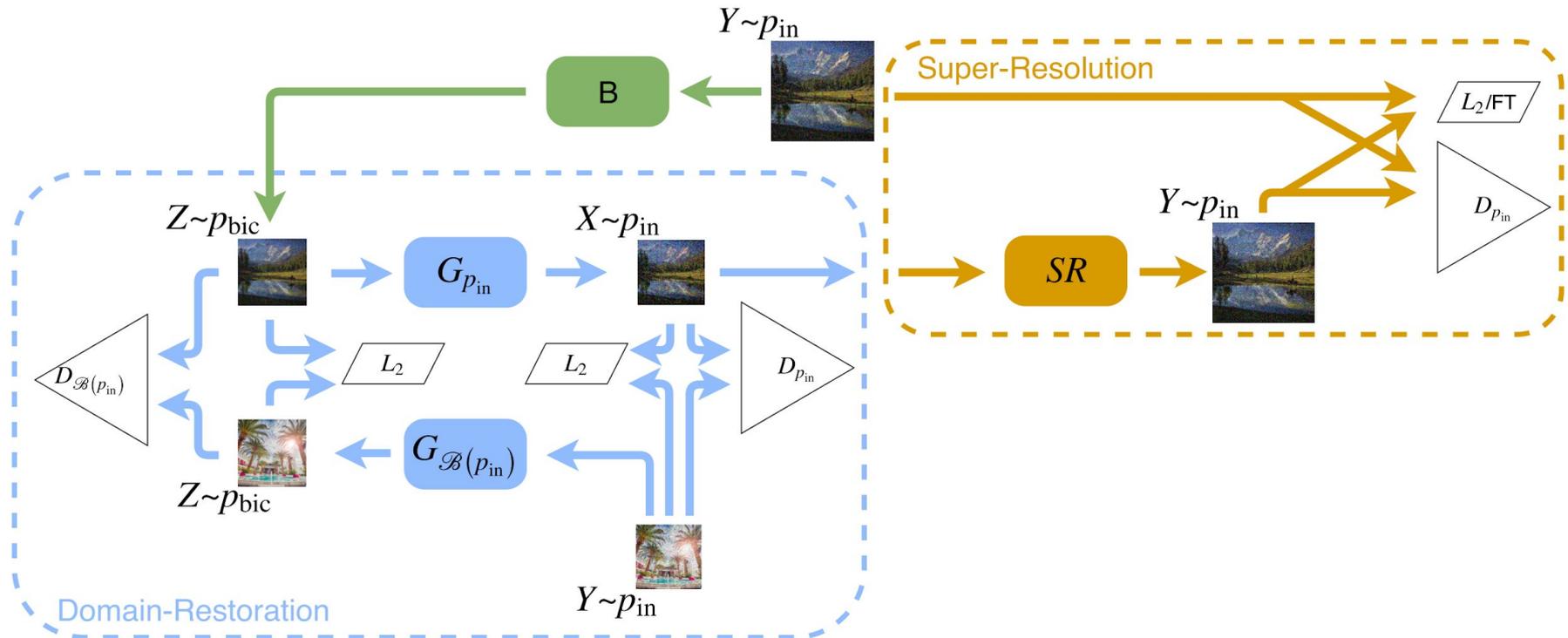
Compression

- Learned image compression for better quality
- Practical lossless compression

Original   Ours   1567 Bytes [B]

JP2K   3138B   *+100%*   BPG   3573B   *+120%*

Ours 1567 Bytes   **BPG** 3573 Bytes   *128% larger*   JPEG 13959B   *+790%*   WebP 9437B   *+502%*

x

E

w

q

$\hat{w}$

G

$\hat{x}$

D

[Agustsson et al.]
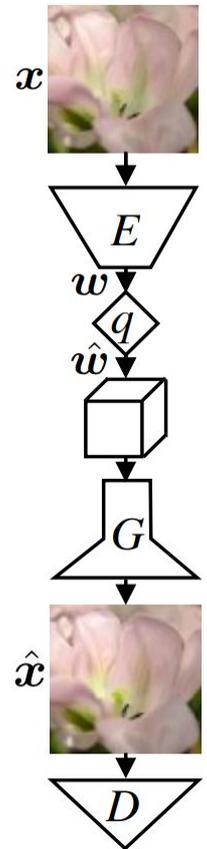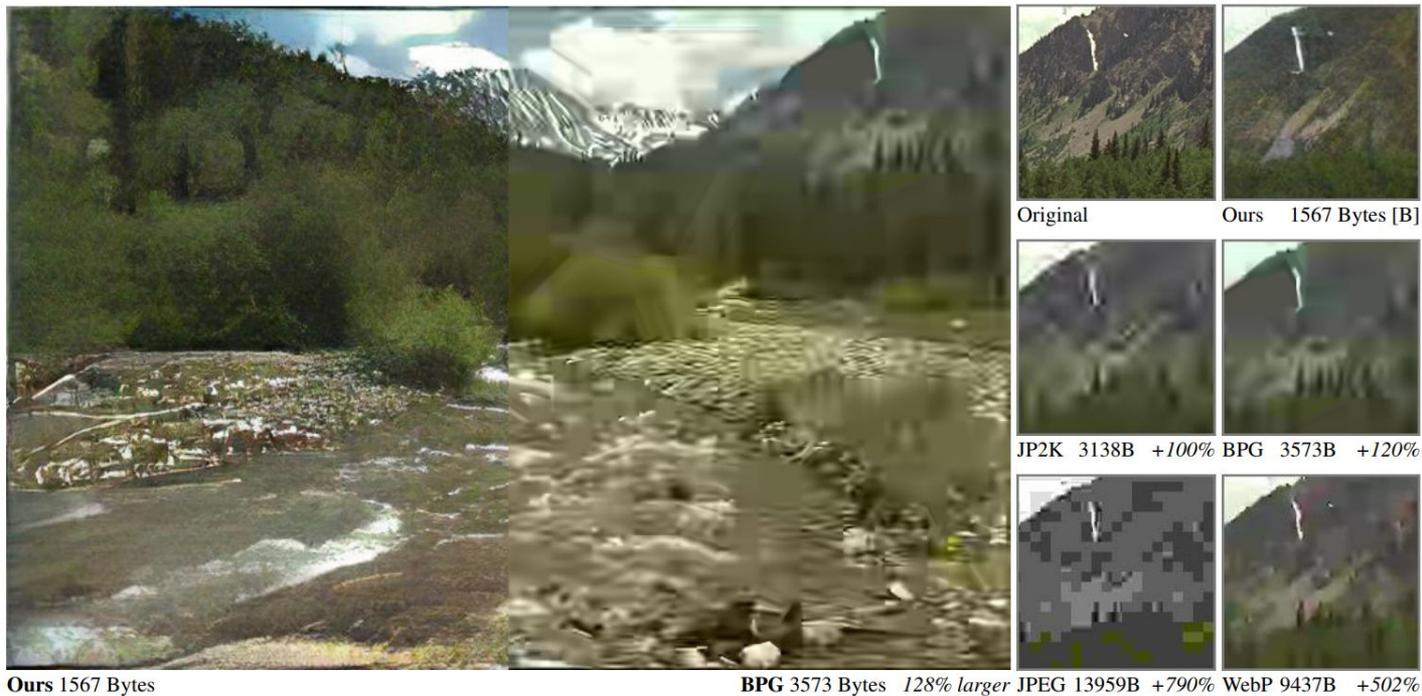
Unsupervised Video Generation

Progressive Video Generation

Sliced Wasserstein GAN loss
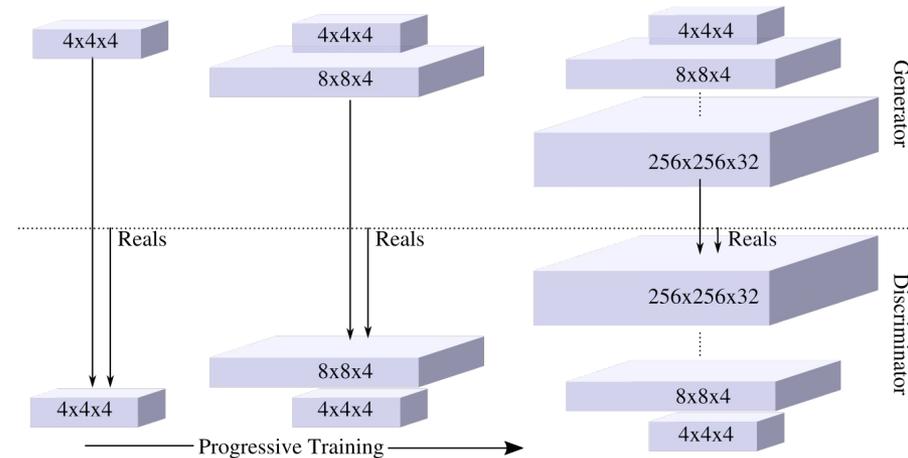
Progressive growing of 3D GANs

- [CVPR'19], Sliced Wasserstein Generative Models, Jiqing Wu*, Zhiwu Huang*, Dinesh Acharya, Wen Li, Danda Pani Paudel, Janine Thoma, Luc Van Gool (*equal contributed)

TRACE Autonomous Driving: Robust Perception

clear weather — snow — fog — rain — nighttime
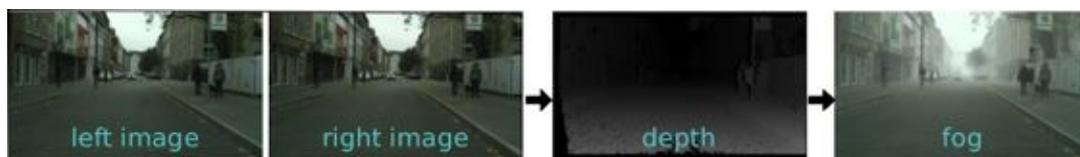
a) Fog Simulation

b) Training with Synthetic Fog

c) Foggy Scene Understanding
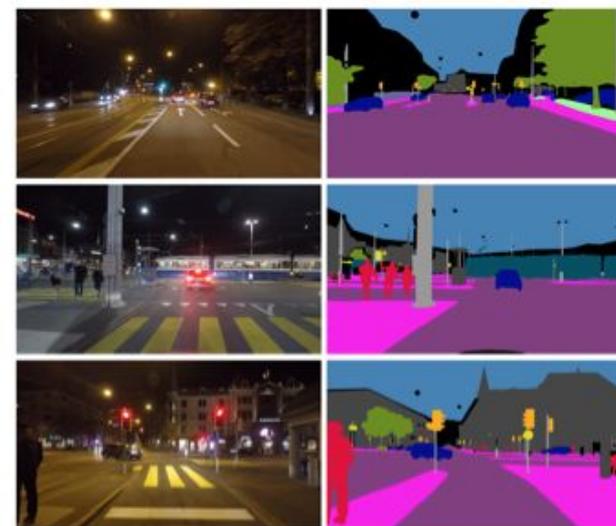
(a) Image — (b) Semantic GT

- Develop robust algorithms for all (adverse) weather and illumination conditions
- [IJCV18, ECCV18, ITSC18, IJCV19, arXiv19], Christos Sakaridis, Dengxin Dai, Luc Van Gool
- CVPR'19 workshop "Vision for All Seasons: Bad Weather and Nighttime"

TRACE Autonomous Driving: Learning to Drive (VP)
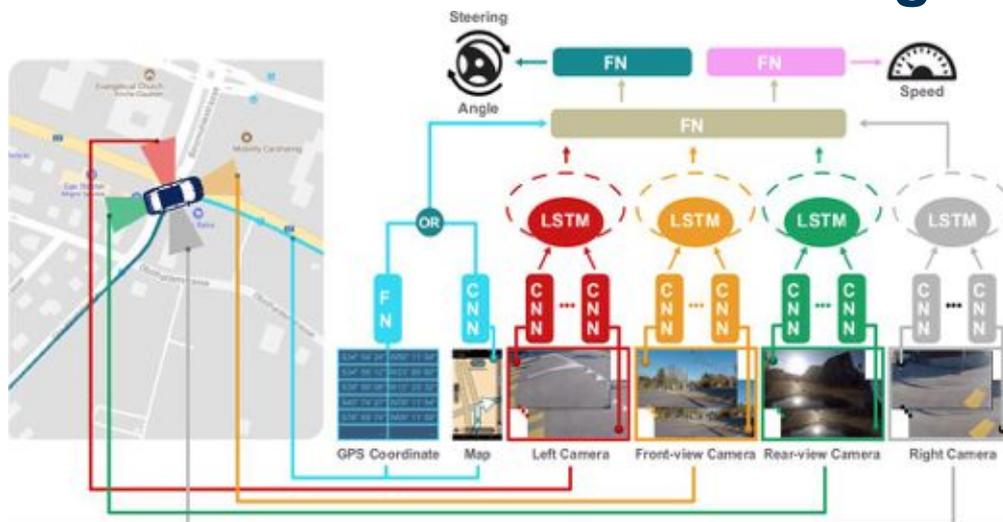
1) **Camera Rig Systems**
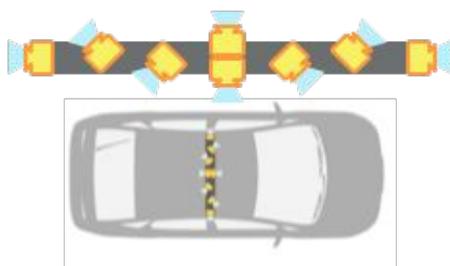
2) **Navigation Systems**

a) TomTom   b) HERE   c) Raw GPS
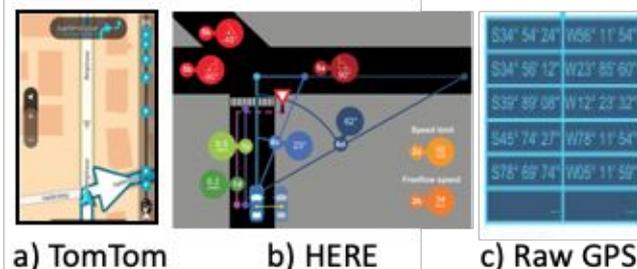
1. **End-to-End Learning**: from perception to control directly
2. **Accurate**: Surround-view Cams + Maps
3. **Comfortable**: reduce motion sickness (jerk)
4. **Human-like**: render human-like driving style

- Accurate, Comfortable and Human-like Driving with Cameras, Sensors and Navigational Maps
- [ECCV18, IV18,, arXiv19], Simon Hecker, Dengxin Dai, Luc Van Gool
- ICCV'19 workshop "Autonomous Driving: Challenges and Trends"

TRACE Autonomous Driving: Talk to your car (VP)
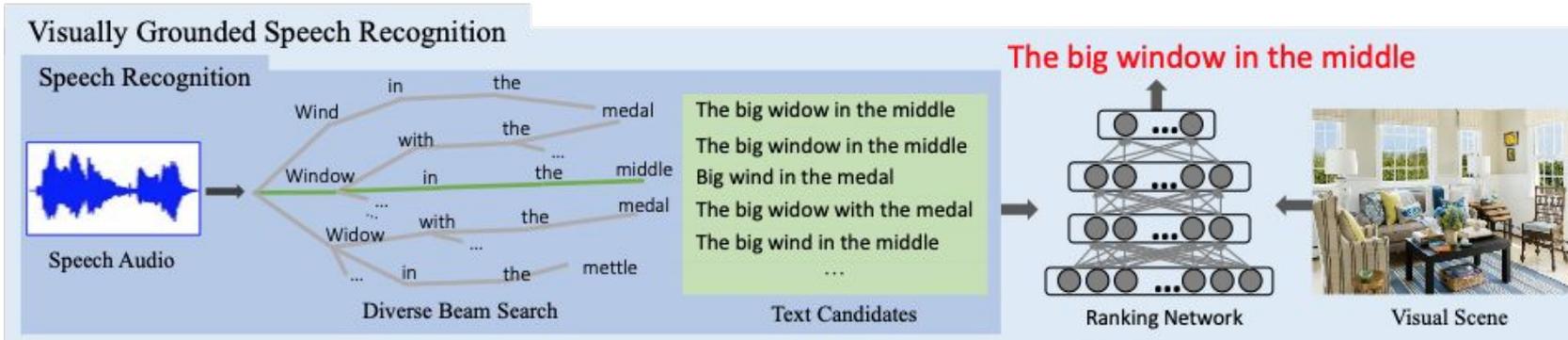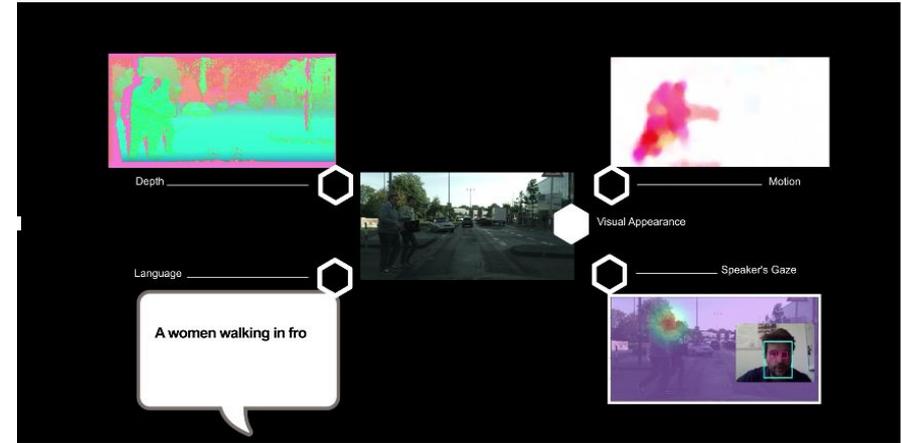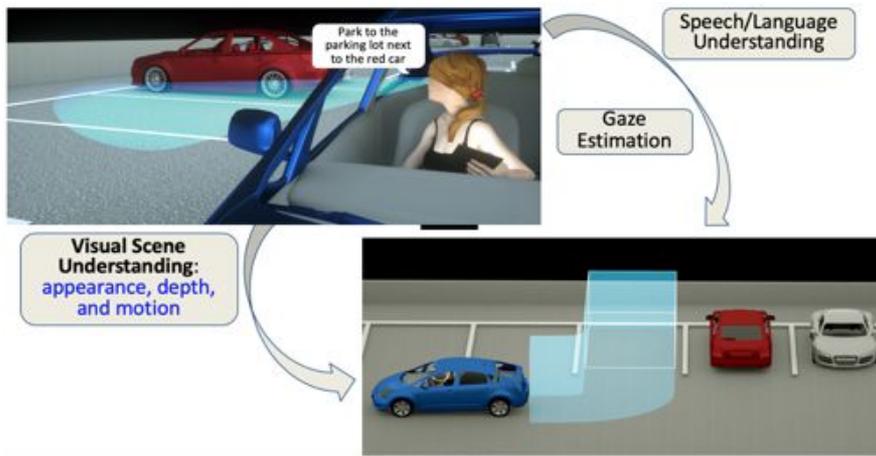
- Human-Robot Communication via Spoken Language
  - Speech understanding, natural language modeling, and computer vision
- [CVPR18, WACV18] Arun Balajee Vasudevan, Dengxin Dai, Luc Van Gool

Visual Object Tracking

- **Only** initial target box is given

- Need to learn and update a model (e.g. a classifier) of the target appearance **online**

- Need to estimate accurate target bounding box

ATOM: Accurate Tracking by Overlap Maximization

- Accurate bounding box estimation:
  - We train a network that regresses the bounding box overlap (IoU) between target and estimate
  - Prediction is conditioned on reference target appearance
  - Refined by maximizing the overlap during tracking

[M. Danelljan, G. Bhat, F. Khan, M. Felsberg. CVPR 2018]
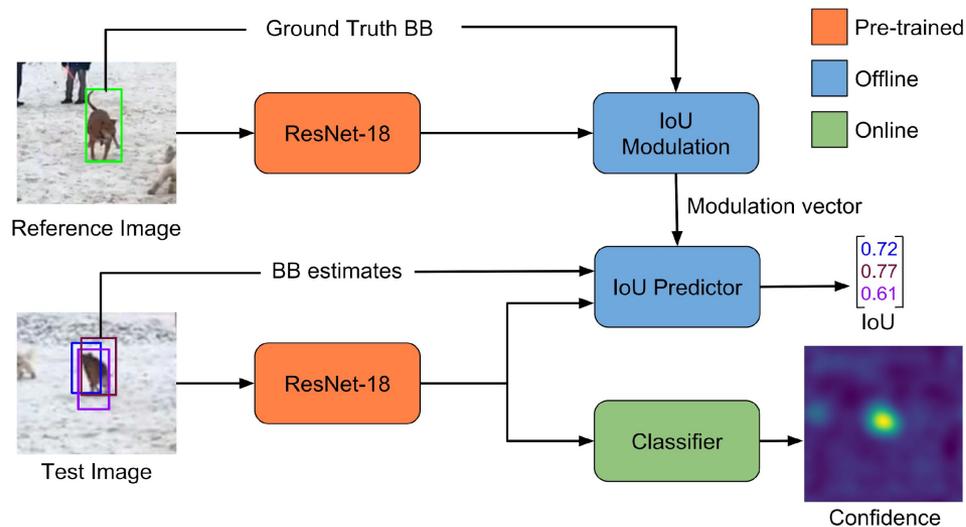
Two views of the same city

Semantic city modeling   &   Capture the dynamic city buzz

the static city model provides the
ideal backdrop to show dynamic events

a living city rather than a ghost town,
with traffic included
+ use consumer video for change detection

| Buildings | Landmarks | Traffic Flow | Events | AR nouveau |

3D All The Way

1. Dense 3D using
Structure-from-Motion
(CMP MVS)

3. Facade splitting

5. Textured 3D model
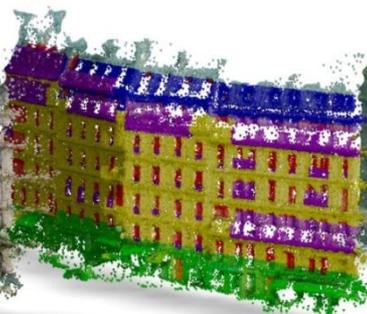
2. Semantic
segmentation
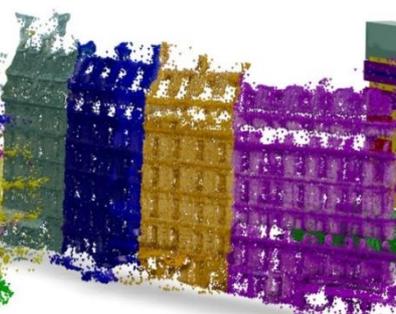
4. Architectural rules
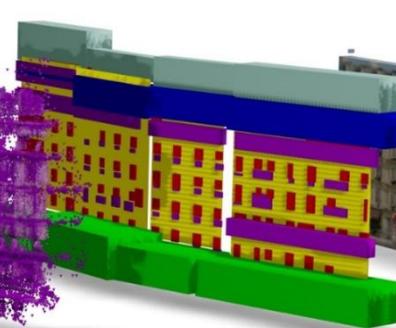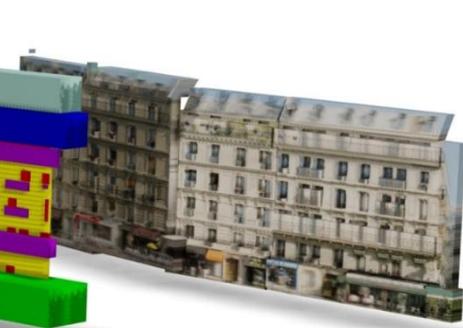
❑ 3D All The Way – 8 mins for whole street
Semantic Segmentation of Urban Scenes From Start to End in 3D (CVPR 2015)
Martinovic, Knopp, Riemenschneider, Van Gool

Dynamic Modelling
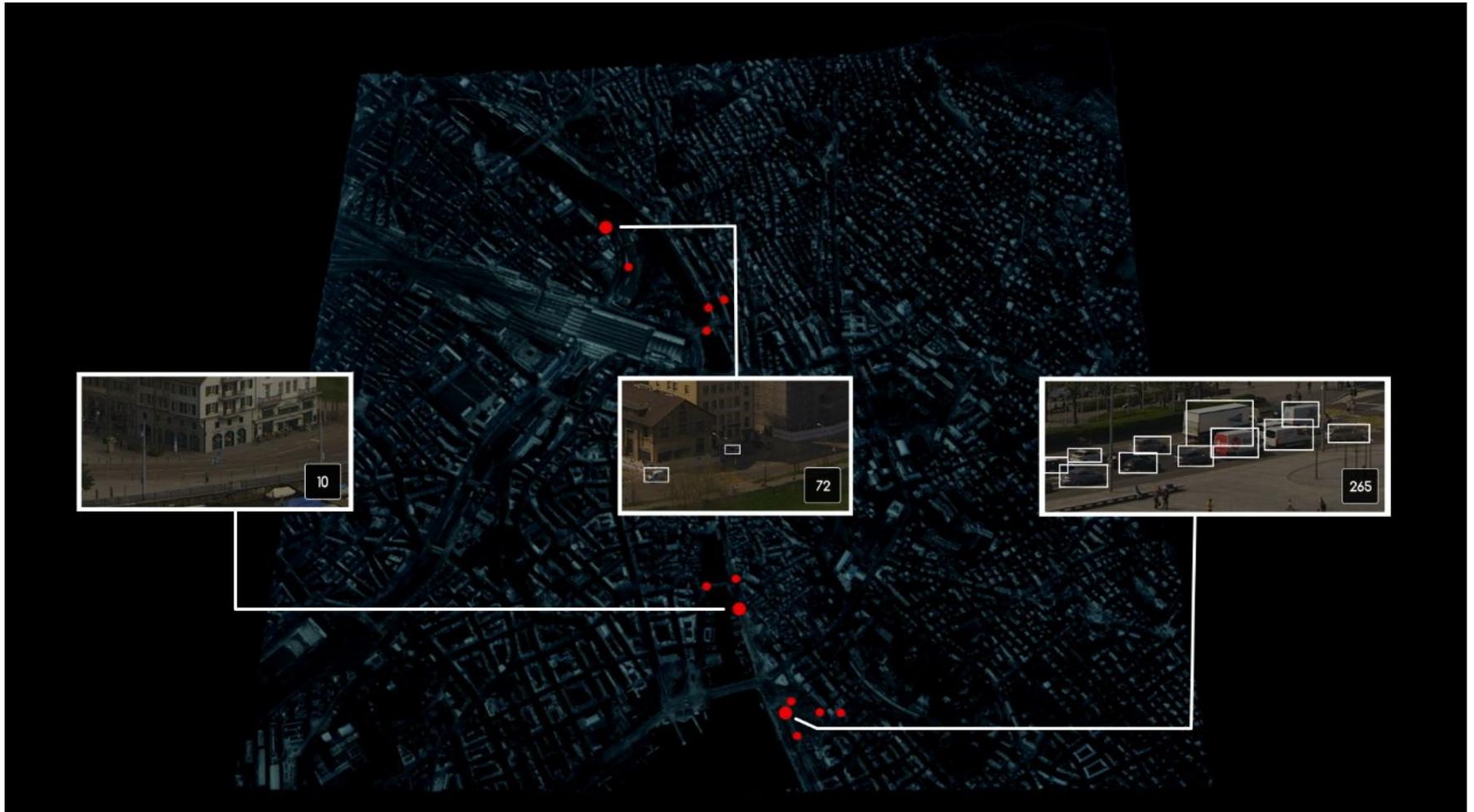
Traffic Modelling

Avatars
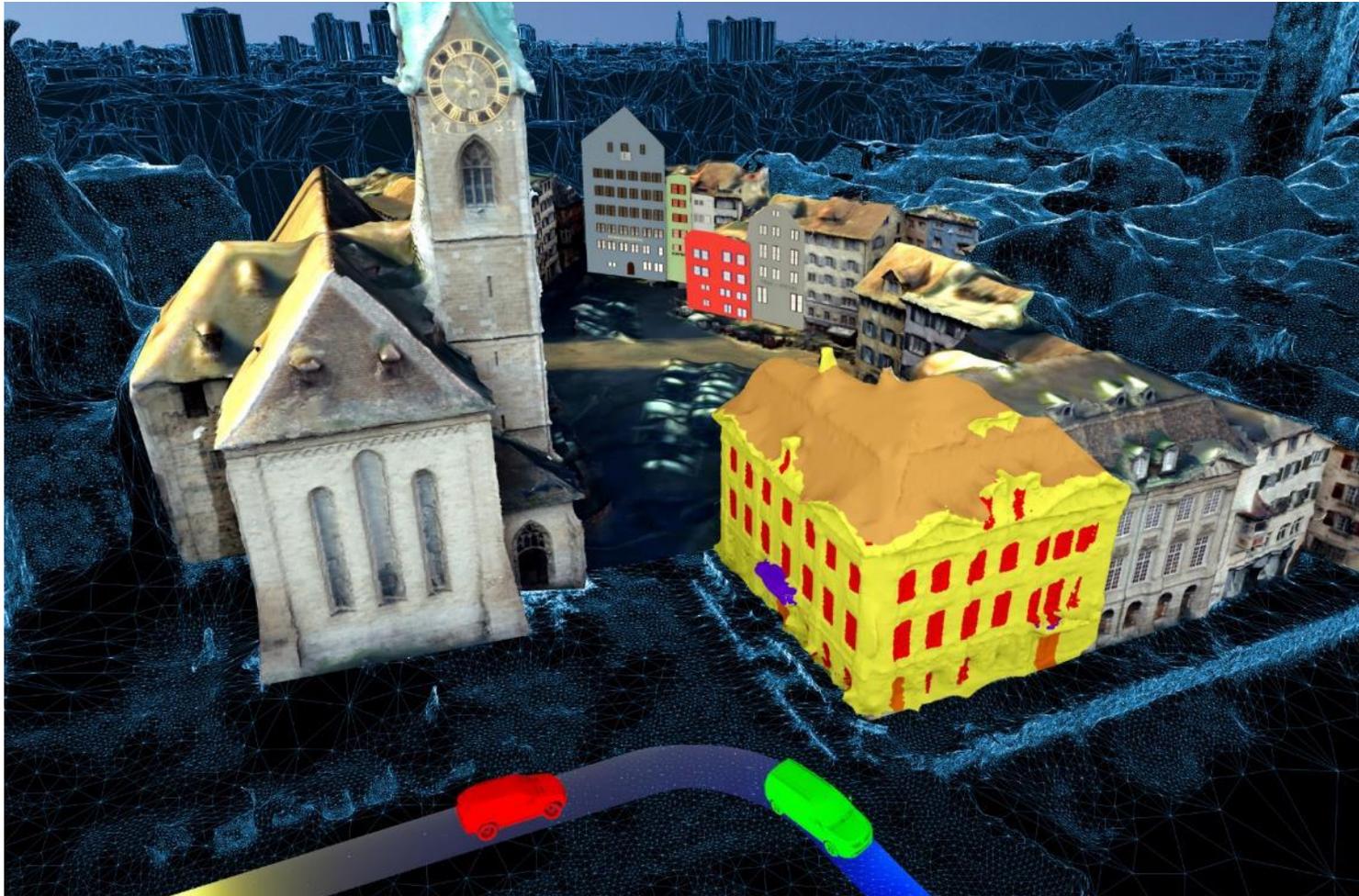
Visual Recognition by Learning from Web Data

Learn deep models with limited human supervision:

- Deep neural networks are data-hungry, and annotating large scale training data is time consuming and expensive
- Abundant weakly labeled data can be retrieved from Internet

queries → **tiger airplane, ...** → Image Retrieval → **Large Scale Dataset** → CNN

WebVision database 2.0:
- 5,000 categories
- 16M images
- 290K validation images
- 290K test images

Large scale image recognition task by learning from web data
- Workshops at CVPR 2017, 2018, 2019

WebVision Database: Visual Learning and Understanding from Web Data. arXiv: 1708.02862, 2017.

http://www.vision.ee.ethz.ch/webvision/workshop.html

Facial Expression Recognition

Facial expressions captured

Image-based covariance pooling

Original	Correctly Predicted	Incorrectly Predicted	Predicted
Angry			Neutral, Neutral, Neutral,
Disgust			Sad, Sad, Surprise,
Fear			Happy, Happy, Neutral,
Happy			Sad, Neutral, Neutral,
Neutral			Angry, Happy, Happy,
Sad			Neutral, Angry, Happy,
Surprise			Happy, Happy, Happy,
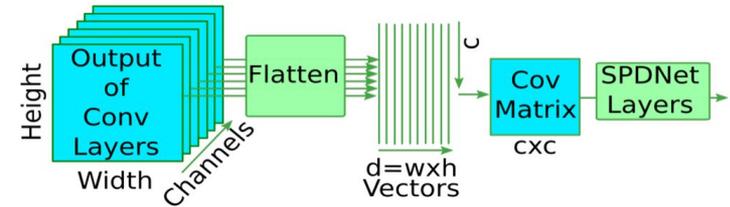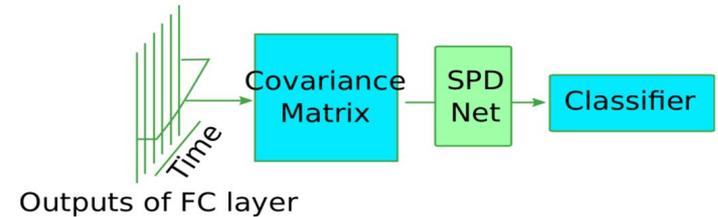
Video-based covariance pooling

- [AAAI'17/18], SPD and Grassmann Manifold Networks, Zhiwu Huang, Jiqing Wu, Luc Van Gool
- [CVPRW'18] Covariance Pooling for Facial Expression Recognition, Dinesh Acharya, Zhiwu Huang, Danda P. Paudel, Luc V. Gool

Human Action Recognition
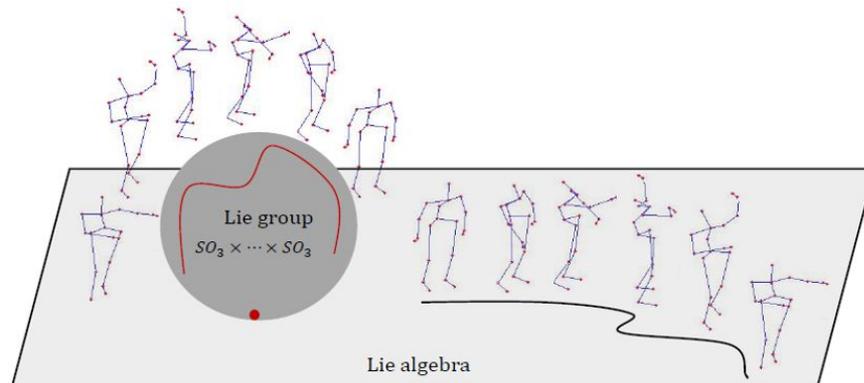
Predict Actions with 3D Skeletons

Lie group representation for skeletons
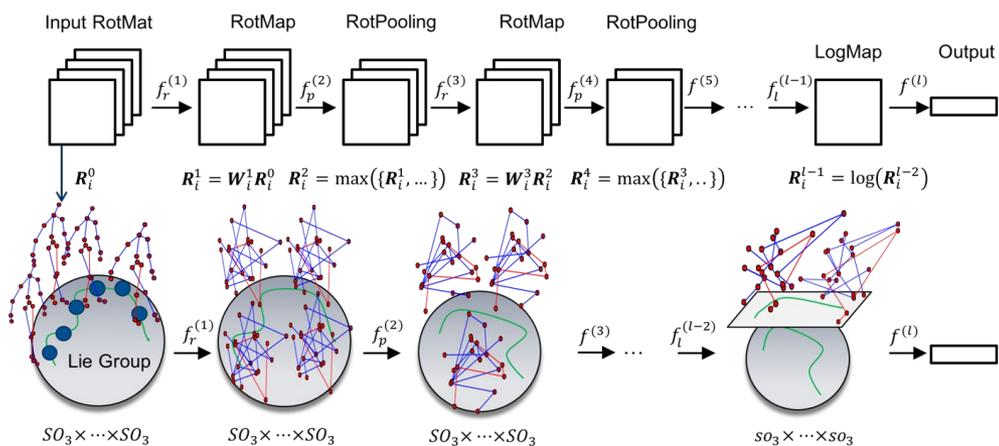
Deep Learning for compact Lie group representations

- [CVPR'17], Deep Learning on Lie Gropus for Skeleton-based Action Recognition, Zhiwu Huang, Chengde Wan, Thomas Probst, Luc Van Gool

Semantic 3D Model of a Building

- 3D Reconstruction using phone + cloud
- Usages: Building renovation, AR/VR
- Mapping contain semantics as well as geometry- e.g. window and its shape

- **Input:** Images from multiple users
- **Goal:** 3D CAD model of Buildings
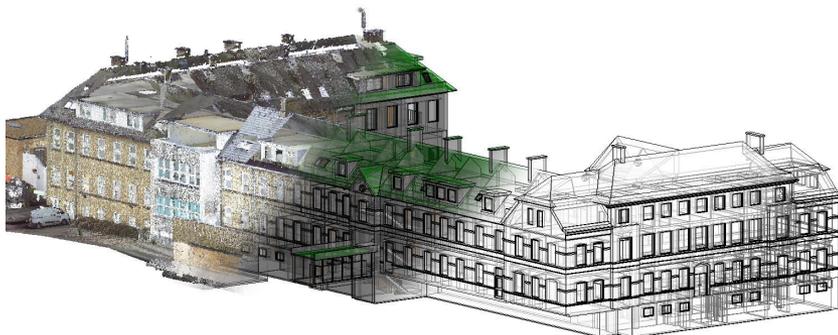
Mobile App
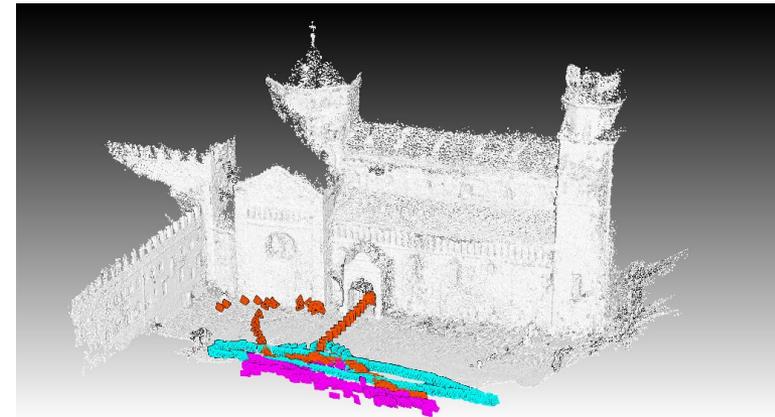
Collaborative 3D

[Locher et al.]

CAD model

Stereo Vision in Robot-Assisted Retinal Surgery
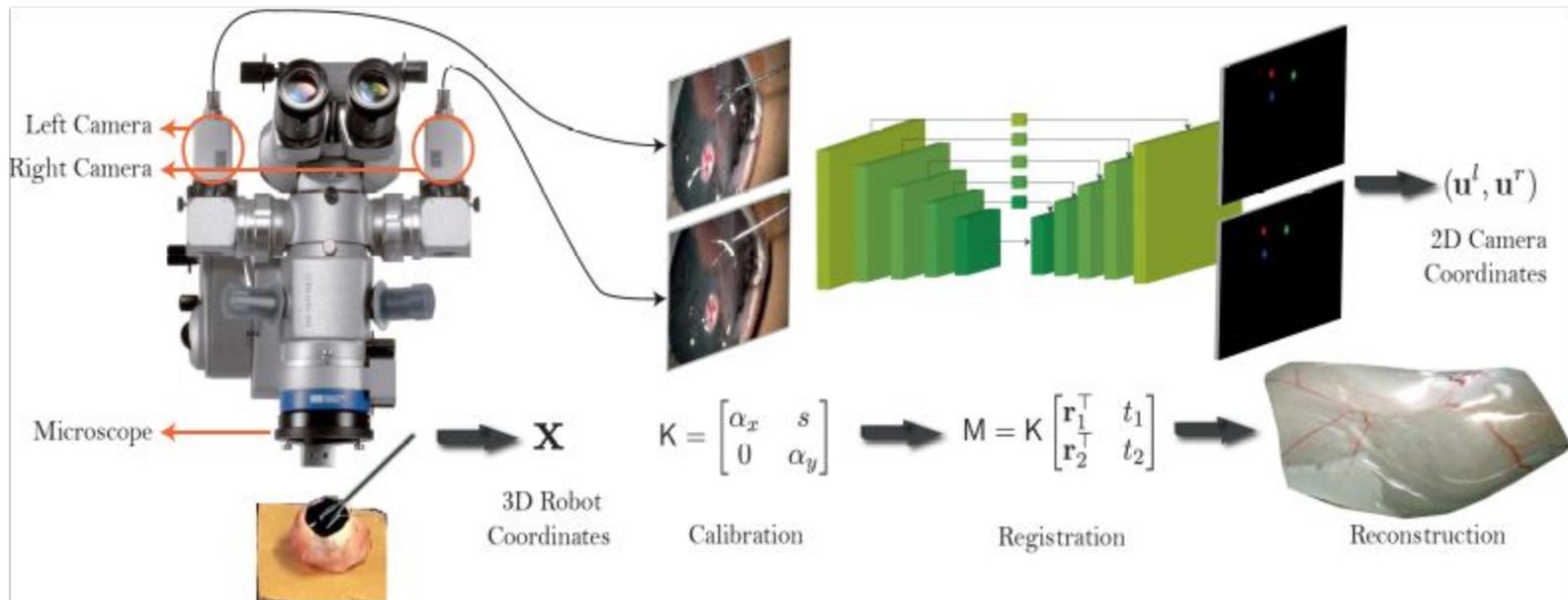
- Calibrate camera using tool landmarks and robot kinematics
- Real-time landmark detection (GPU)
- Triangulation of tool to continuously update hand-eye calibration

[Probst et al.]

Parking Spot Modelling—Static and Dynamic cameras

Dynamic + Static cameras

Deep Fashion Parsing

- Locate Fashion items
  - on images
- Learning
  - for semantic instances
- Product recognition
  - for business
- Search and Shop
  - by image

Coat	Bag
Pants	Scarf

Thank you for your attention!