

Research Overview

CV Group

Feb 3rd, 2022

Prof. Dr. Simone Frintrop

Computer Vision, Department of Informatics,
University of Hamburg, Germany

Uni Hamburg

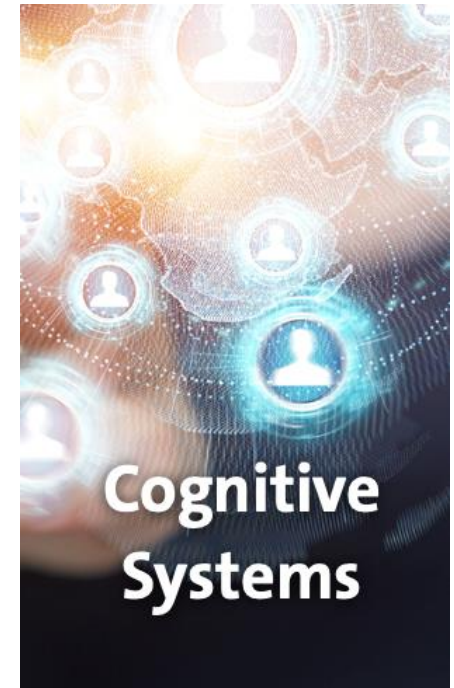
Department of Informatics

- University of Hamburg: Excellence University
- Informatics: 24 Research groups, 200+ researchers & staff in total
- About 2400 students in 10 degree programs + 110 in PhD program



Department of Informatics

- [Image Processing \(BV\)](#)
- [Computer Vision \(CV\)](#)
- [Ethics in Information Technology \(EIT\)](#)
- [Human-Computer Interaction \(HCI\)](#)
- [Language Technology \(LT\)](#)
- [Natural Language Systems \(NATS\)](#)
- [Semantic Systems \(SEMS\)](#)
- [Signal Processing \(SP\)](#)
- [Technical Aspects of Multimodal Systems \(TAMS\)](#)
- [Knowledge Technology \(WTM\)](#)



Computer Vision Group



Simone Frintrop



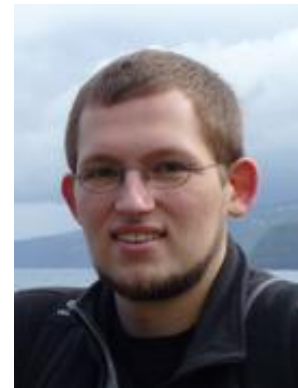
Mikko Lauri



Ehsan Yaghoubi



Noha Sarhan



Christian Wilms



Tim Rolff

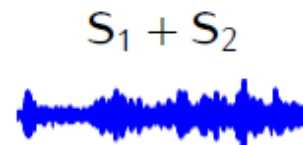
Research in CV Group



Human Behavior Analysis



Saliency/Attention



Cross-modal data processing

Gaze Prediction

Personality Recognition

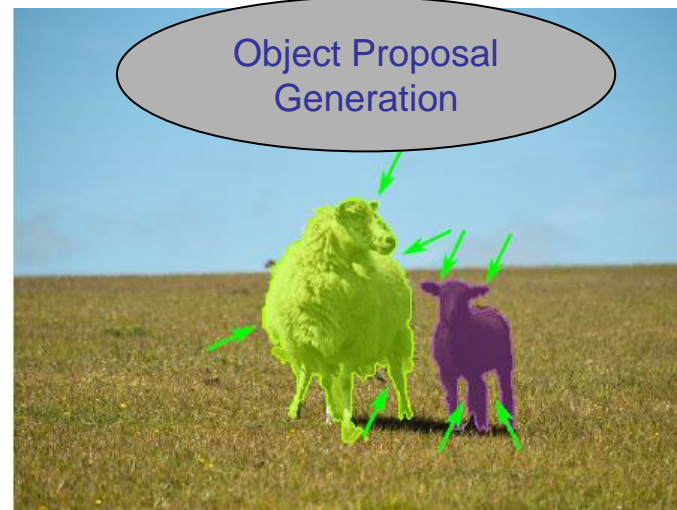
Attentive Sign Language Recognition



Active Perception



Object Pose Estimation



Object Proposal Generation

Human Behavior Analysis

Attentive Sign
Language
Recognition



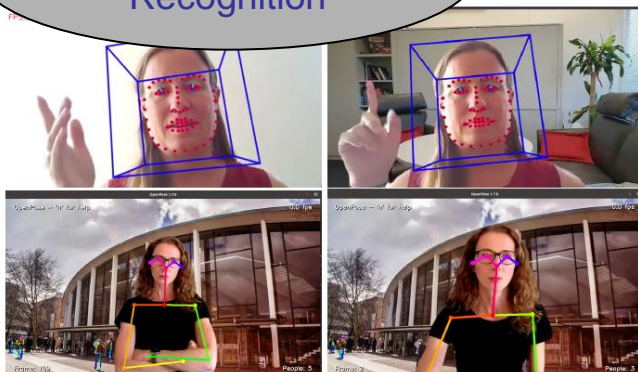
Noha Sarhan

Gaze Prediction



Tim Rolff

Personality
Recognition



Human Attribute
Recognition



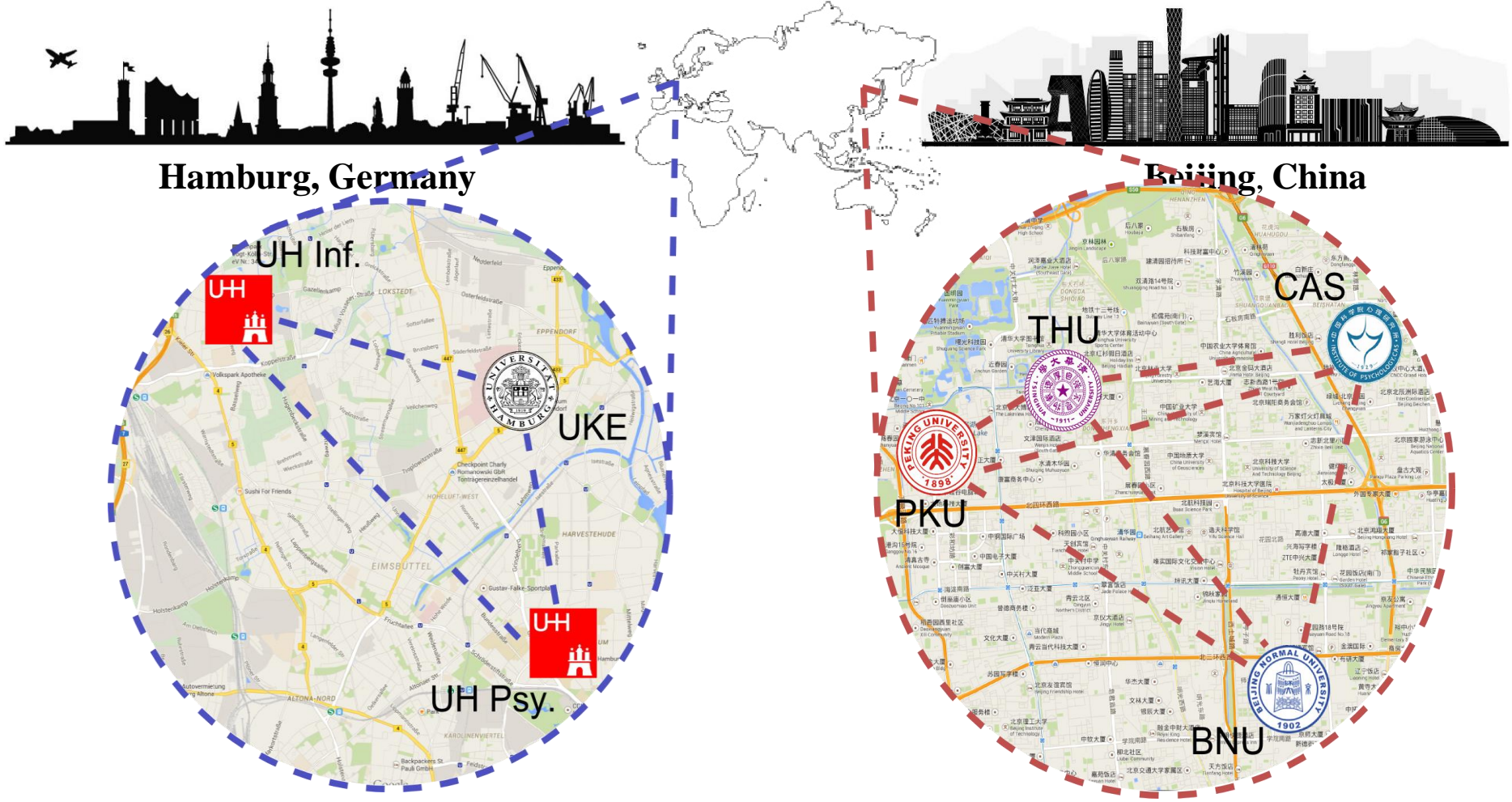
Ehsan Yaghoubi



SFB TRR 169

Crossmodal Learning: Adaptivity, Prediction and Interaction

Cross-regional Collaborating Institutions



Project Leaders (16 in Germany + 17 in China)



Learning as a Means to Cope with the Real World

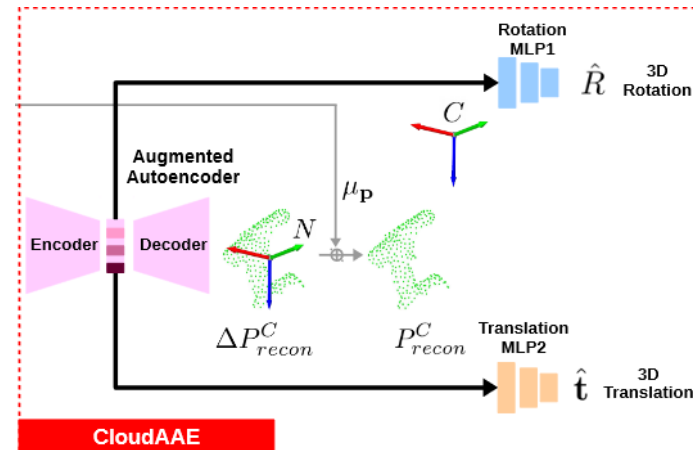
Learning is becoming increasingly central for **interrelated studies of intelligent systems**, including neuroscience, cognitive science and robotics.

Learning enables

- **Integration** of local information with established, global knowledge
- **Interaction** with a changing world
- **Cooperation** with other adaptive learning systems
- **Consolidation** of knowledge into internal, predictive models
- **Combination** of top-down and bottom-up modelling



Object Pose Estimation



Ge Gao

- Ge Gao: **Learning 6D Object Pose from Point Clouds**, PhD thesis, 2021, [\[PDF\]](#)
- Ge Gao, Mikko Lauri, Xiaolin Hu, Jianwei Zhang, Simone Frintrop: **CloudAAE: Learning 6D Object Pose Regression with On-line Data Synthesis on Point Clouds**, Proceeding of International Conference on Robotics and Automation (ICRA), 2021, [\[PDF\]](#), [\[Code\]](#)
- Ge Gao, Mikko Lauri, Yulong Wang, Xiaolin Hu, Jianwei Zhang and Simone Frintrop: **6D Object Pose Regression via Supervised Learning on Point Clouds**, Proceeding of International Conference on Robotics and Automation (ICRA) 2020, [\[PDF\]](#), [\[Code\]](#)
- Ge Gao, Mikko Lauri, Jianwei Zhang, Simone Frintrop: **Occlusion Resistant Object Rotation Regression from Point Cloud Segments**, Proceeding of the ECCV workshop on Recovering 6D Object Pose, 2018, [\[PDF\]](#), [\[arXiv\]](#)

Ahoi.digital project

Audio-visual processing: with Timo Gerkmann & Julius Richter

ICPR 2020: Sound source separation

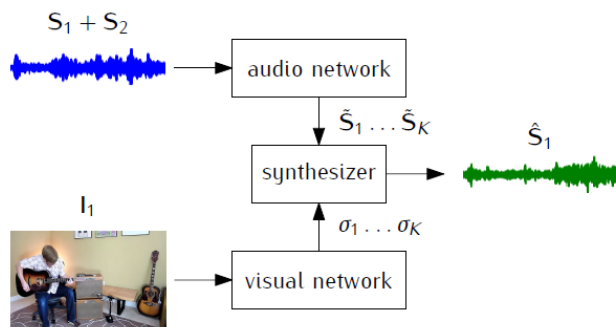
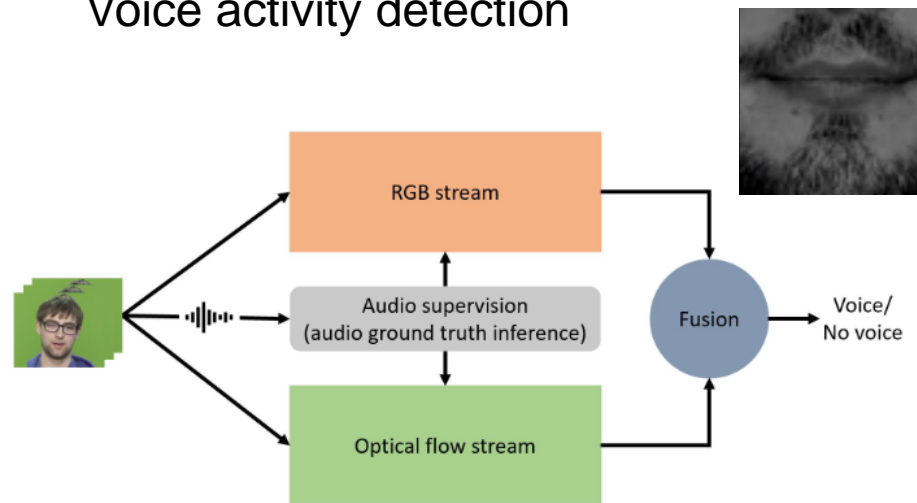


Fig. 1. Our audio-visual sound source separation framework at test time. The audio network takes an audio mixture, e.g., the sum of two spectrograms S_1 and S_2 , and outputs a list of K spectrograms $\hat{S}_1 \dots \hat{S}_K$. The visual network takes a single video frame I_1 as an object prior according to the sound source S_1 and outputs a discrete probability distribution $p(\text{type} = i) = \sigma_i$. The synthesizer generates the separated sound source estimate \hat{S}_1 according to its inputs.

Quan Nguyen, Julius Richter, Mikko Lauri, Timo Gerkmann, Simone Frintrop: **Improving mix-and-separate training in audio-visual sound source separation with an object prior**, ICPR 2020

ICVS 2021: Voice activity detection

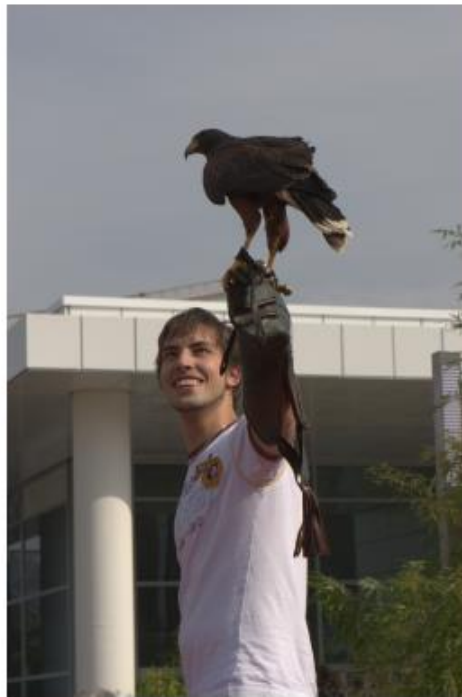


Danu Caus, Guillaume Carbajal, Timo Gerkmann, Simone Frintrop: **See the silence: improving visual-only voice activity detection by optical flow and RGB fusion**, International Conference on Computer Vision Systems (ICVS) 2021

Saliency-based Image Enhancement

With Adobe Research, Hamburg:

- Image enhancement with saliency



MIT-Adobe FiveK Dataset:

<https://data.csail.mit.edu/graphics/fivek/>

[Soroka 2018]

Adobe Research

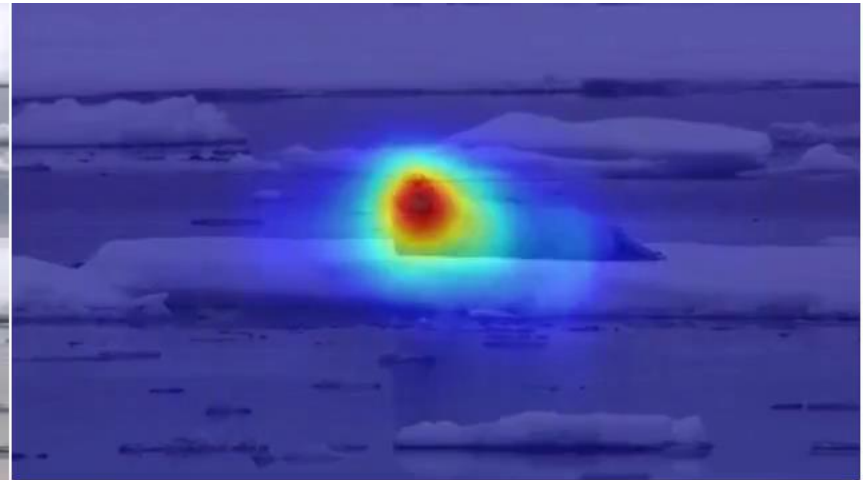
- At Adobe, a saliency system is used to enable smart cropping of videos:



- <https://www.adobe.com/de/marketing/experience-manager-assets/smart-crop.html>
- <https://www.youtube.com/watch?v=MlImphmT5dy8>

Adobe Research

- Deep-learning-based video saliency model to enable focusing on motion (coop with Adobe):



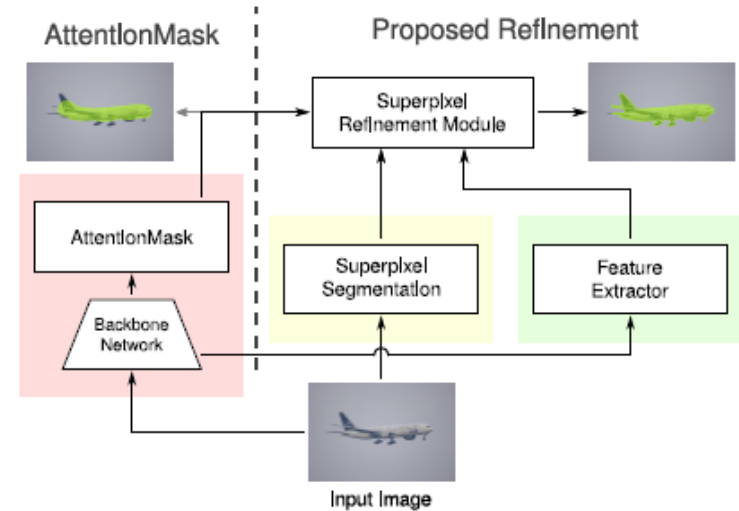
A Novel Dynamic Saliency Network and Its Application to Automatic Reframing,
Ozan Özdemir, Master's thesis, University of Hamburg, 2020



AttentionMask

AttentionMask [Wilms/Frintrop 2018]

- Scale-based attention focuses processing on promising parts of the image
- use attention to sample windows sparsely
- more resources for small objects
- Superpixel refinement enables precise boundaries



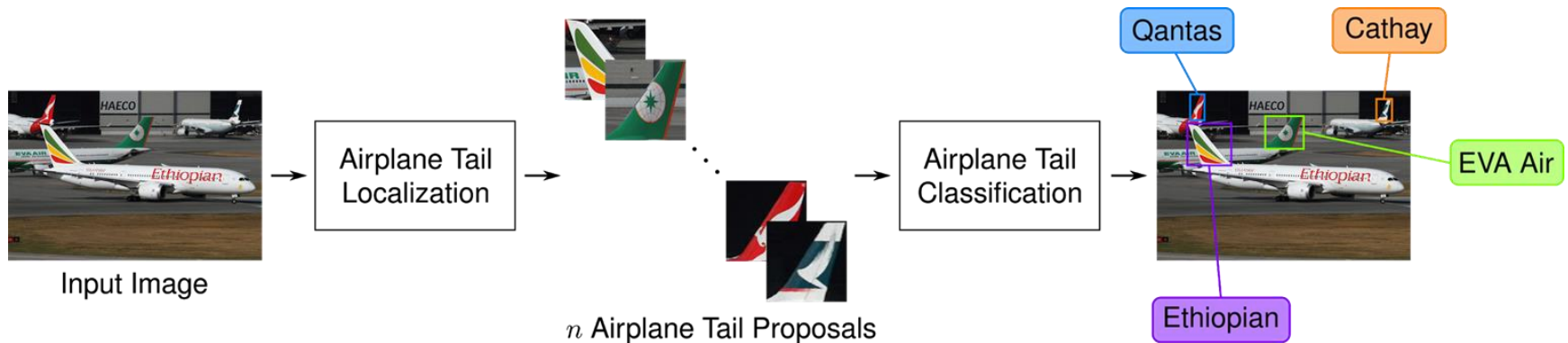
Christian Wilms, Simone Frintrop: **DeepFH Segmentations for Superpixel-based Object Proposal Refinement**, Image and Vision Computing (IMAVIS) 2021

Christian Wilms, Simone Frintrop: **AttentionMask: Attentive, Efficient Object Proposal Generation Focusing on Small Objects**, Asian Conference on Computer Vision (ACCV), 2018



Airline Logo Detection

Task: Localization and classification of airline logos (with ZeroG)



Automatic localization based
on AttentionMask

Simplified VGG-style
architecture as classifier


Research in CV Group



Human Behavior Analysis



Saliency/Attention

$S_1 + S_2$




Cross-modal data processing

Gaze Prediction

Personality Recognition

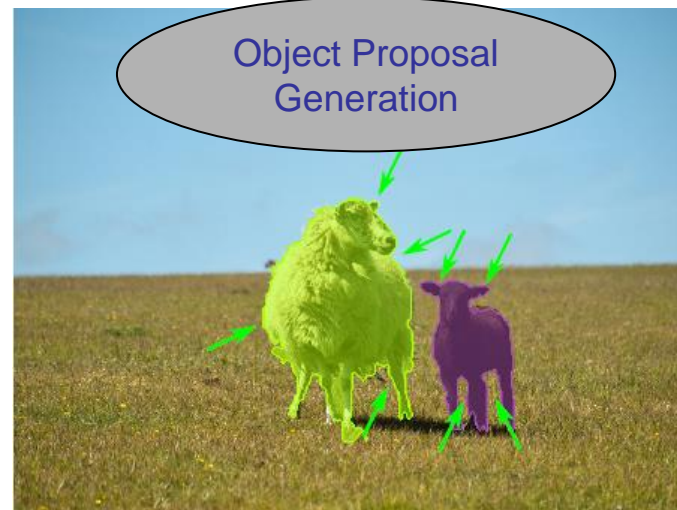
Attentive Sign Language Recognition



Active Perception



Object Pose Estimation



Object Proposal Generation