

Прикладные задачи компьютерного зрения

Распознавание позы и действий человека

О чём пойдет речь?

- Распознавание позы человека
- Распознавание действий
- Отслеживание направления взгляда

Распознавание действия



(a) Cross-over



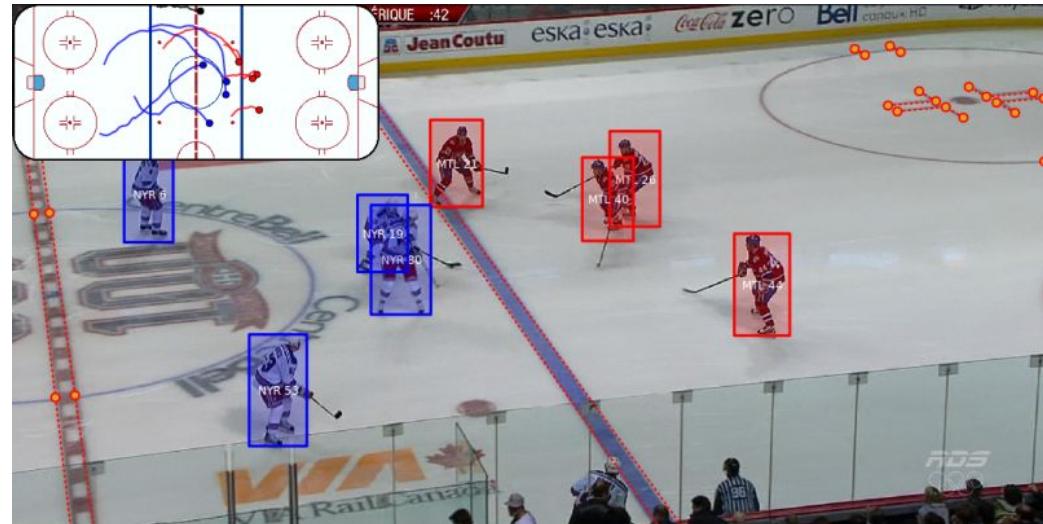
(b) Straight skating



(c) Pre-shot



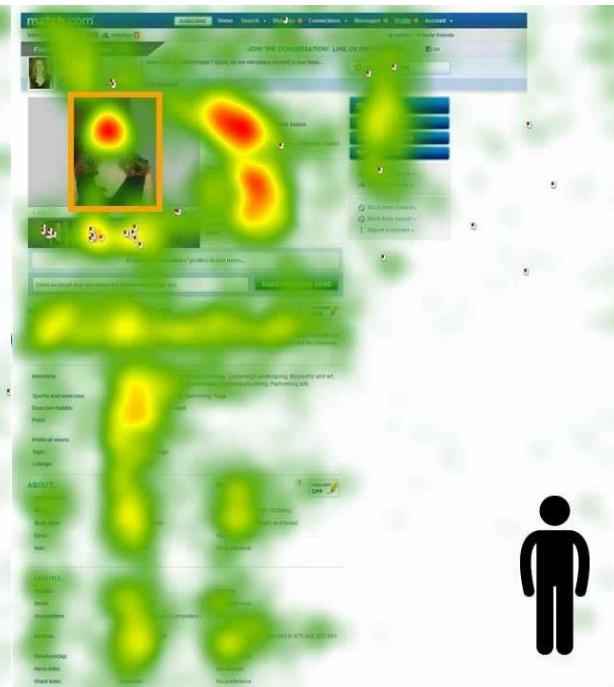
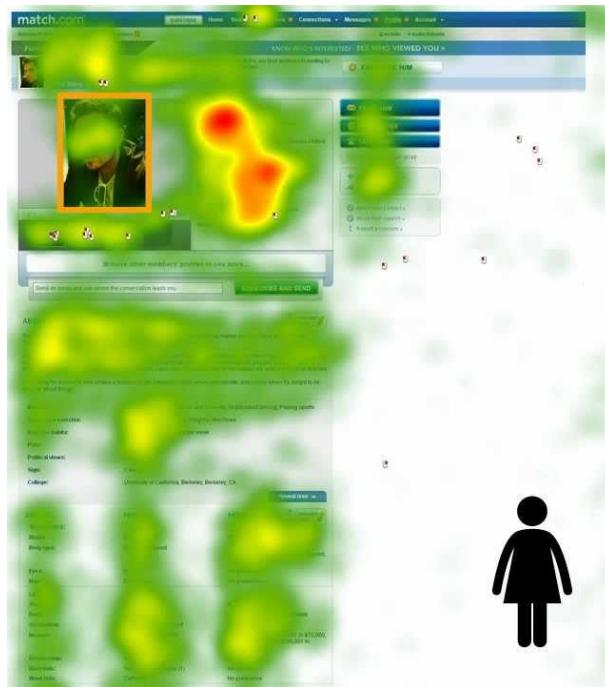
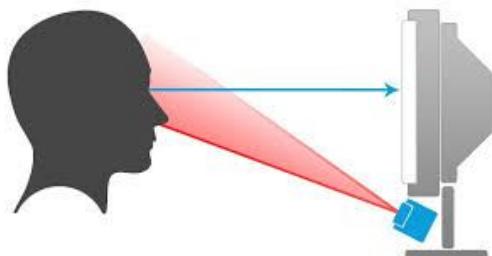
(d) Post-shot



Отслеживание внимания водителя



Отслеживание взгляда (eye tracking)

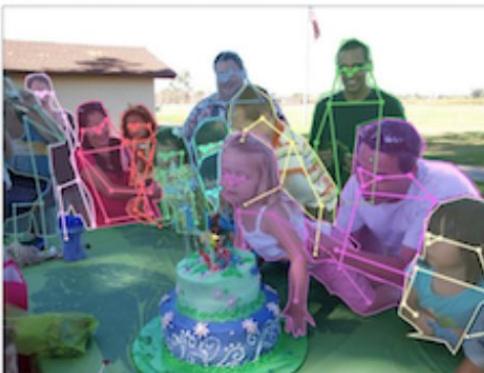


Распознавание эмоций



Детектор ключевых точек

COCO 2018 Keypoint Detection Task



<http://cocodataset.org/#keypoints-2018>

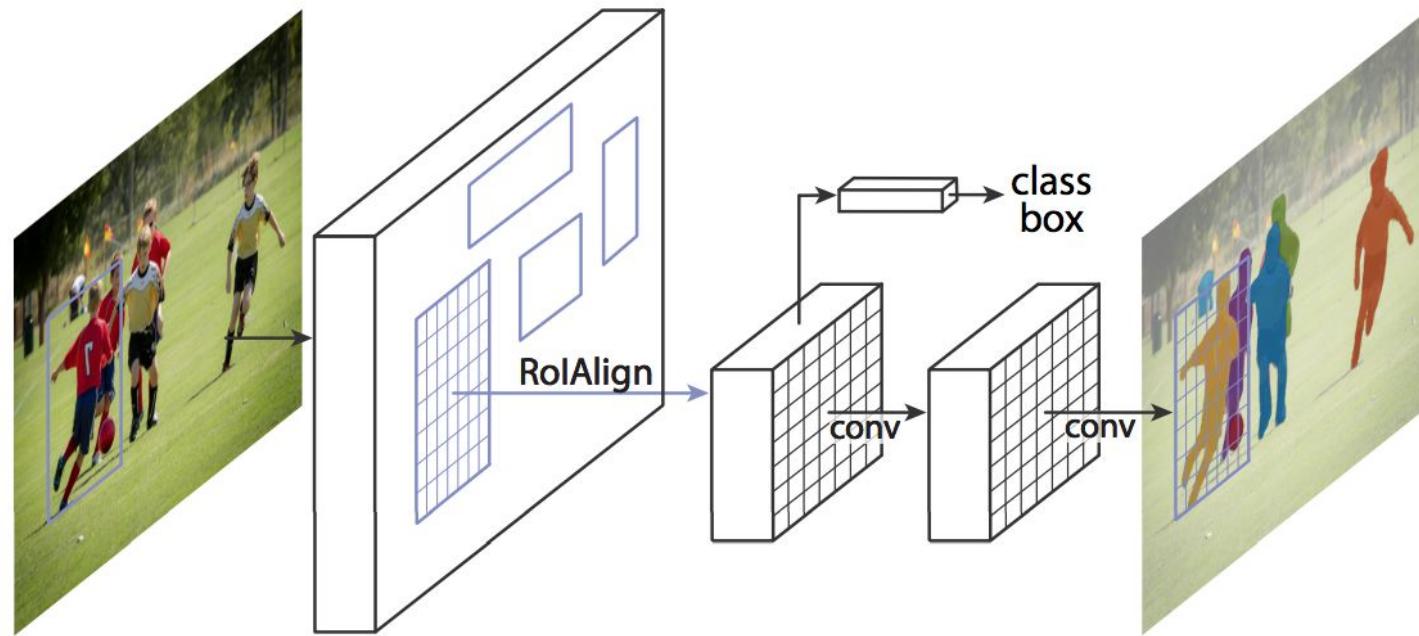
COCO 2018 Keypoint Detection Task

- Детекция объекта на изображении и нахождение ключевых точек
- Положение объекта не известно во время применения модели
- Качество оценивается аналогично качеству детектора
- Вместо IoU используется Object Keypoint Similarity (OKS)

$$\text{OKS} = \frac{\sum_i [\exp(-d_i^2/2s^2k_i^2)\delta(v_i>0)]}{\sum_i [\delta(v_i>0)]}$$

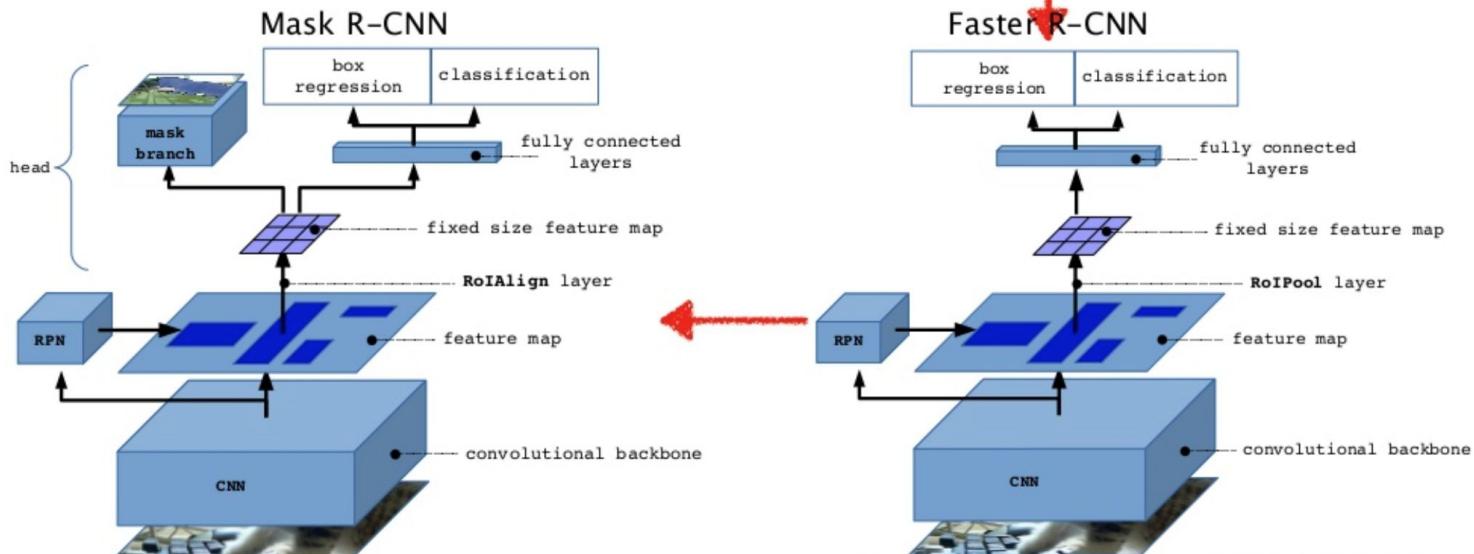
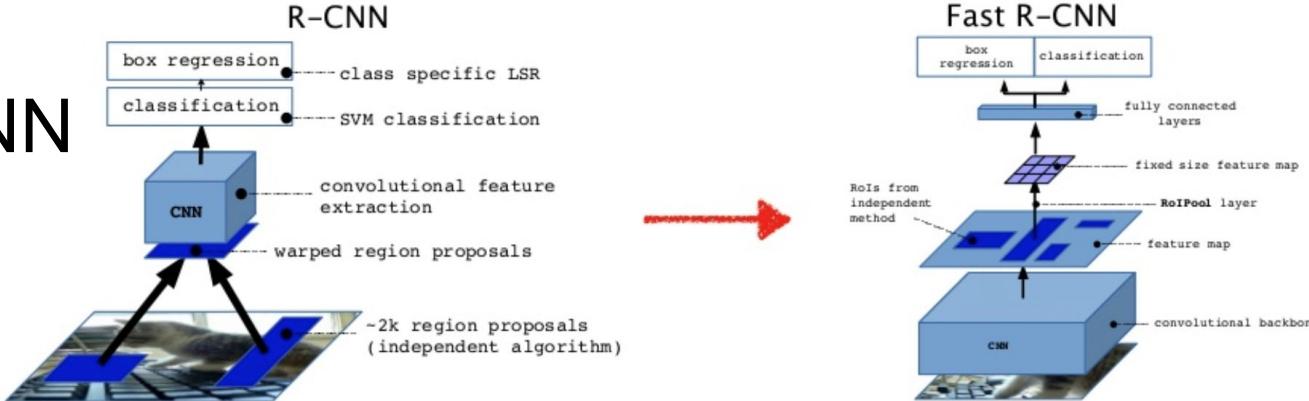
- d_i - расстояние между i -ой точкой из разметки и предсказанной
- s - площадь объекта
- v_i - флаг наличия (видимости) точки в разметке {0,1,2}
- k_i - нормирующая константа

Mask R-CNN

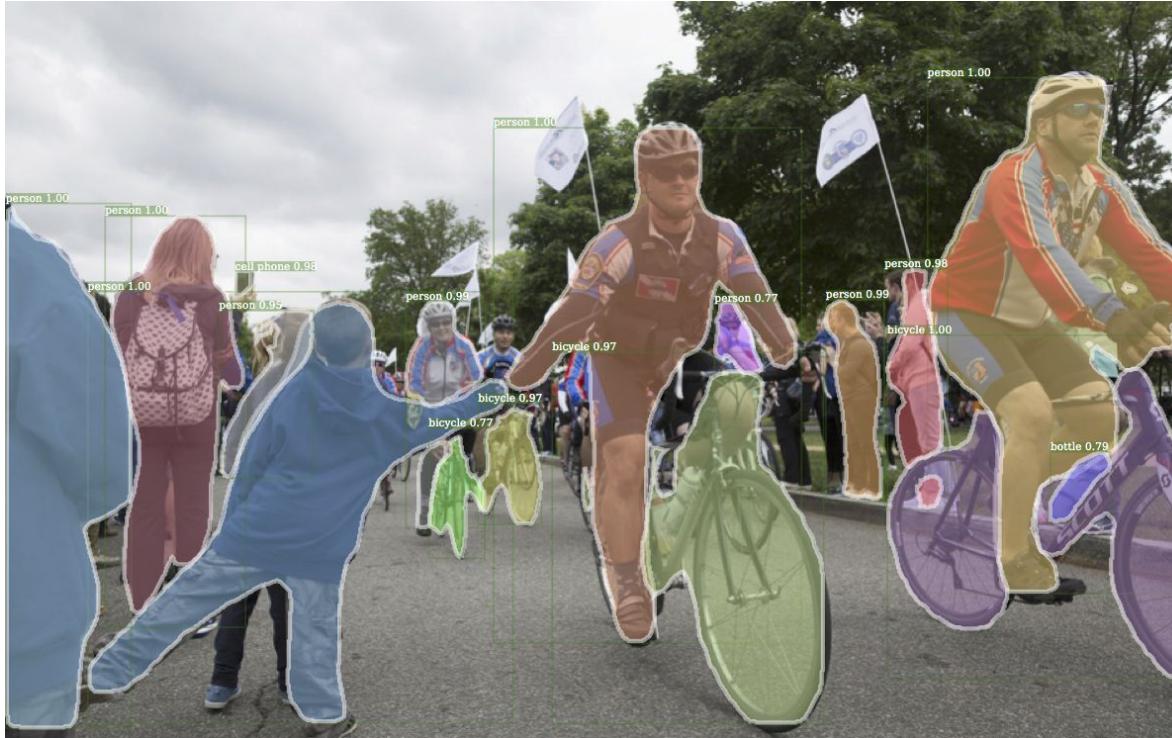


<https://arxiv.org/pdf/1703.06870.pdf>

Mask R-CNN

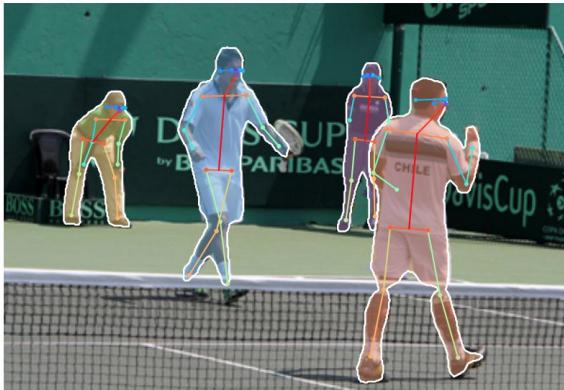


Mask R-CNN



<https://github.com/facebookresearch/maskrcnn-benchmark>

Mask R-CNN (KeyPoints)



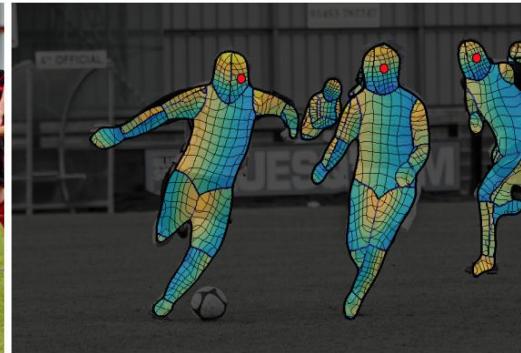
Распознавание позы человека

DensePose: Dense Human Pose Estimation In The Wild

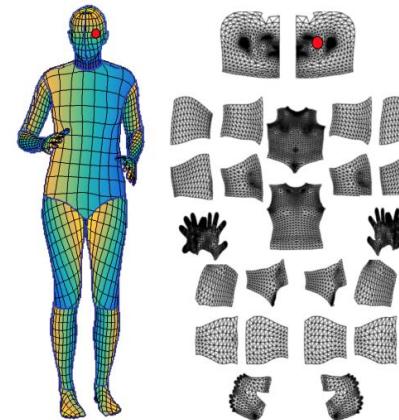


<http://densepose.org>

DensePose



DensePose-RCNN Results



DensePose COCO Dataset

DensePose Dataset

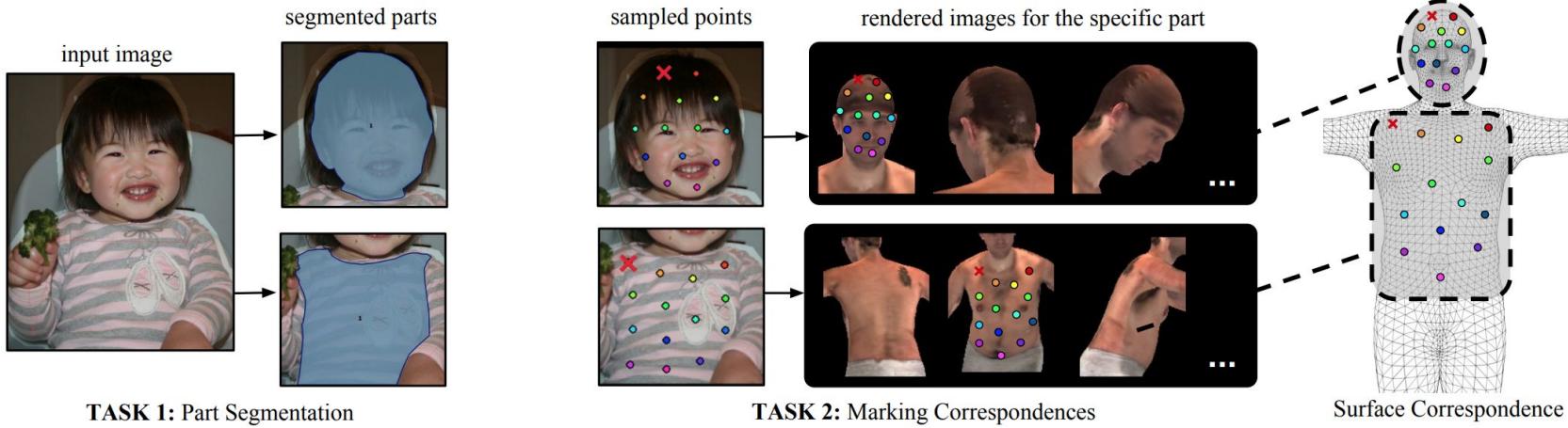
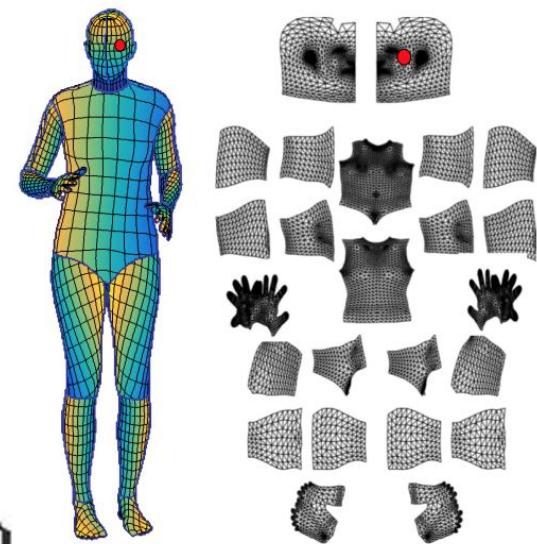
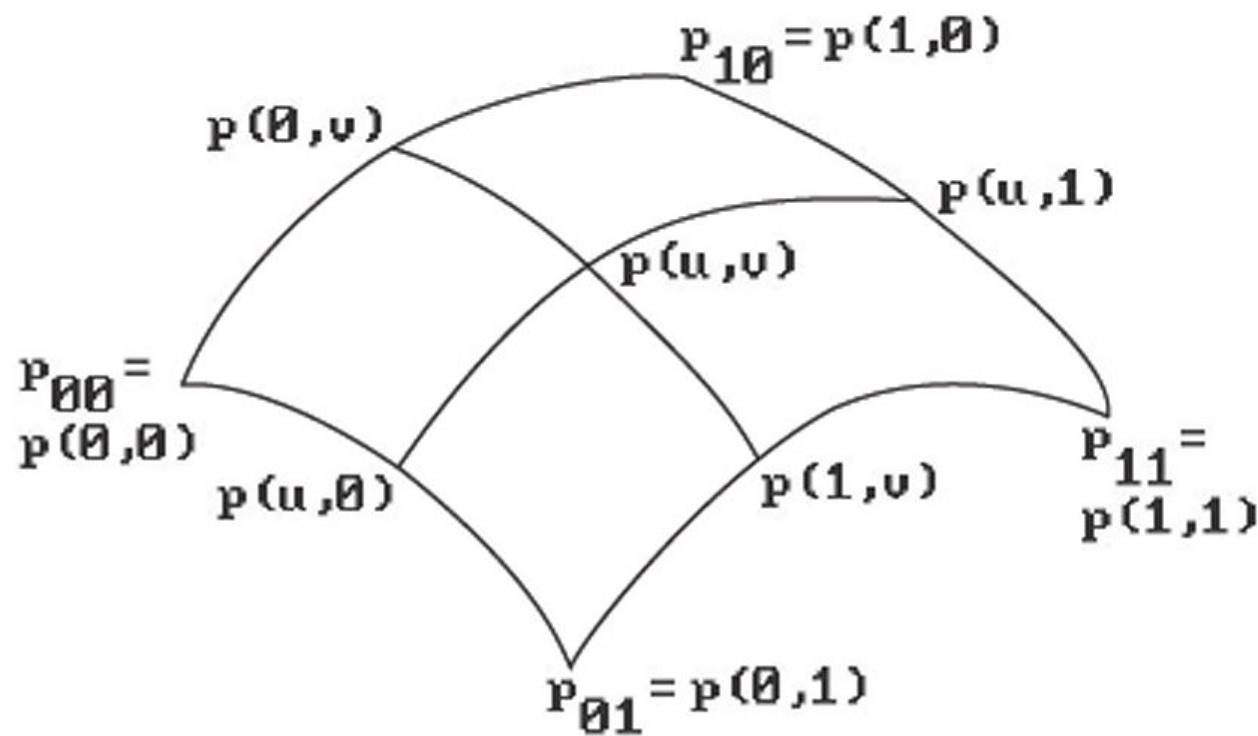


Figure 2: We annotate dense correspondence between images and a 3D surface model by asking the annotators to segment the image into semantic regions and to then localize the corresponding surface point for each of the sampled points on any of the rendered part images. The red cross indicates the currently annotated point. The surface coordinates of the rendered views localize the collected 2D points on the 3D model.

UV coordinates



DensePose

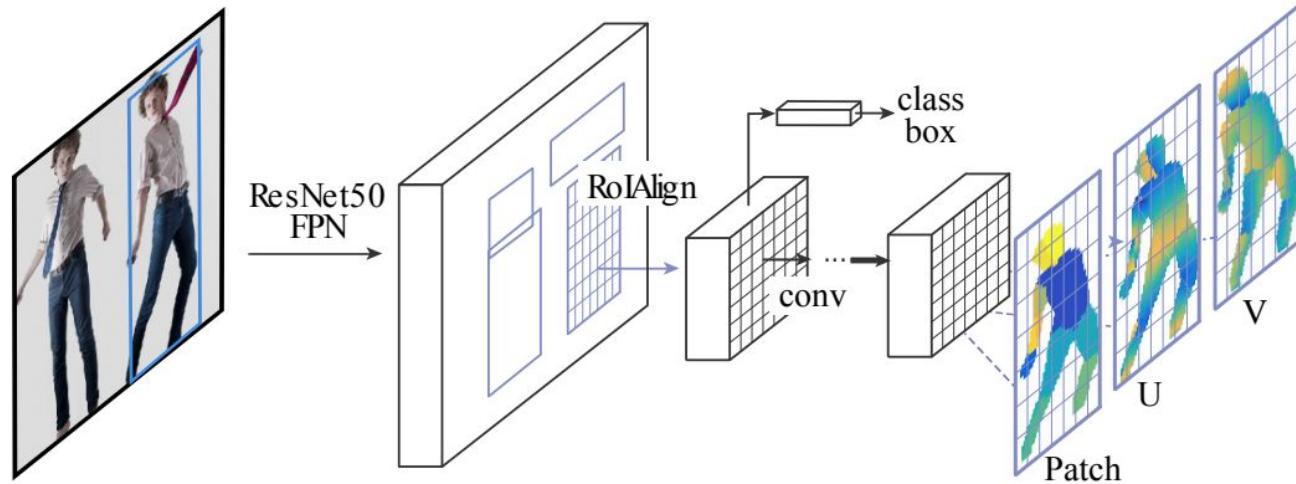


Figure 7: DensePose-RCNN architecture: we use a cascade of region proposal generation and feature pooling, followed by a fully-convolutional network that densely predicts discrete part labels and continuous surface coordinates.

DensePose

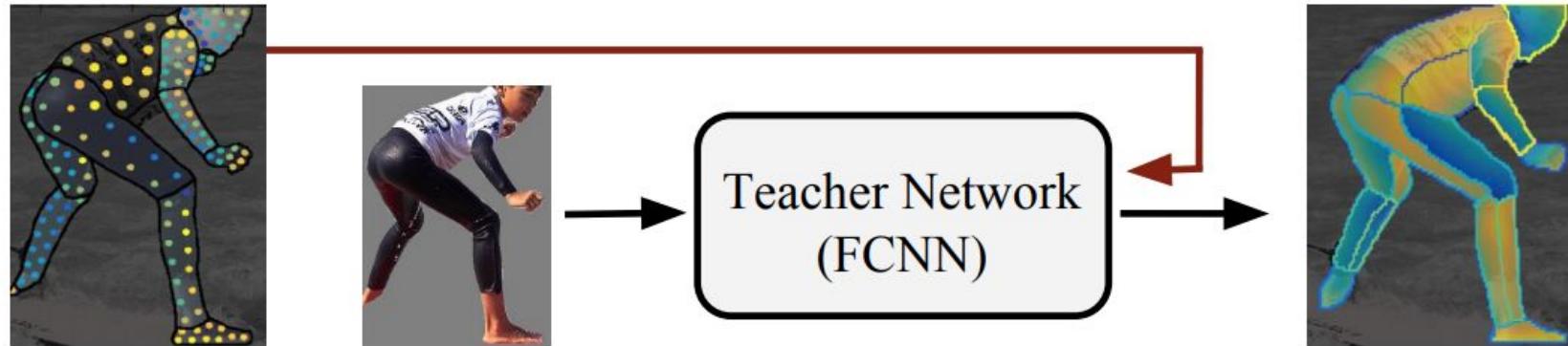
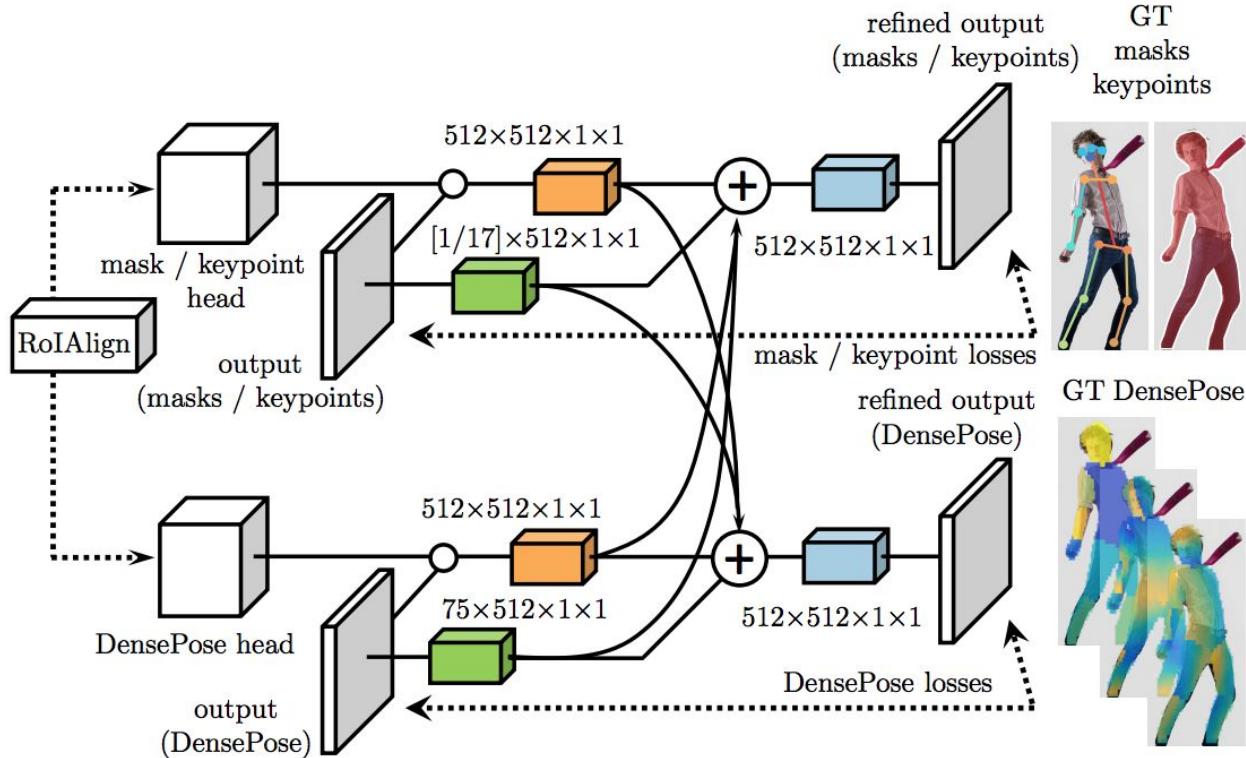


Figure 9: We first train a ‘teacher network’ with our sparse, manually-collected supervision signal, and then use the network to ‘inpaint’ a dense supervision signal used to train our region-based system.

DensePose

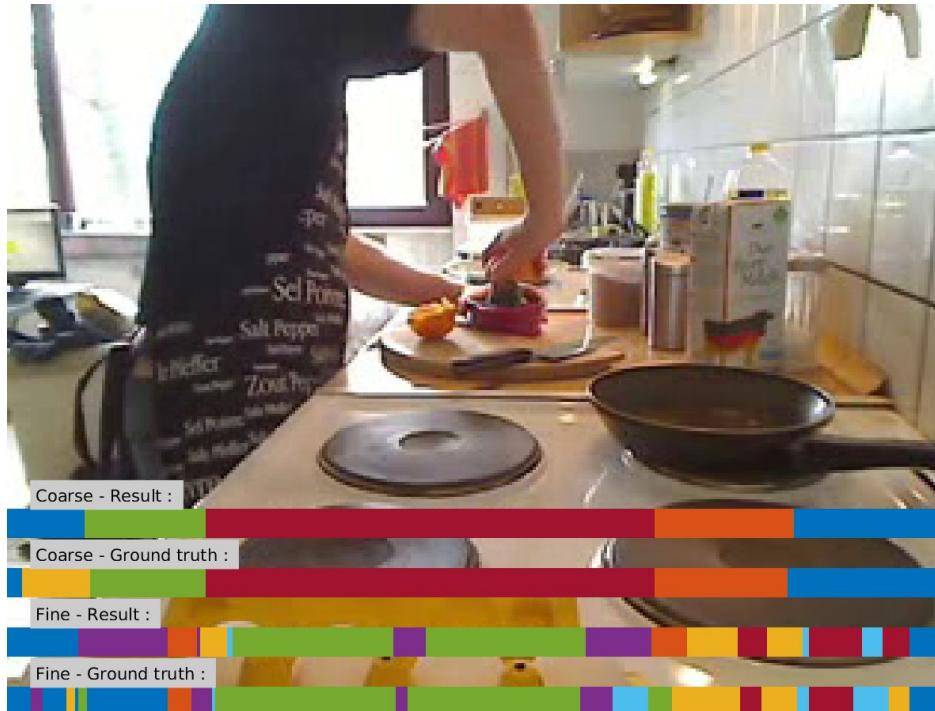


DensePose - Текстуры



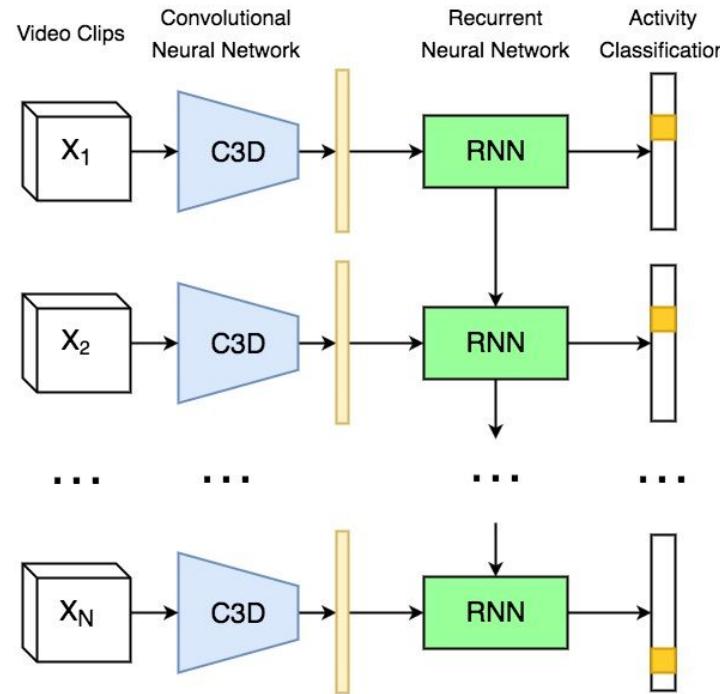
Распознавание действий

The Breakfast Actions Dataset



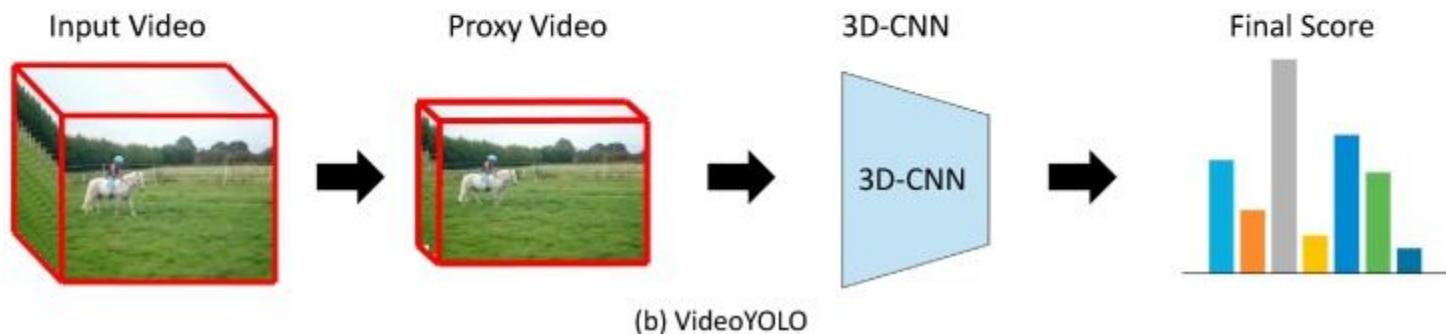
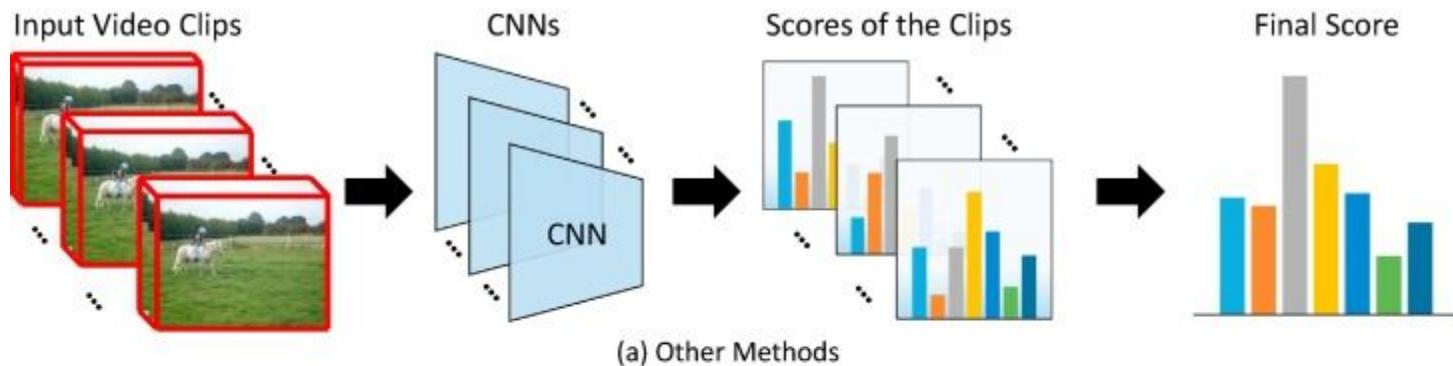
<http://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset/>

Рекурентные сети



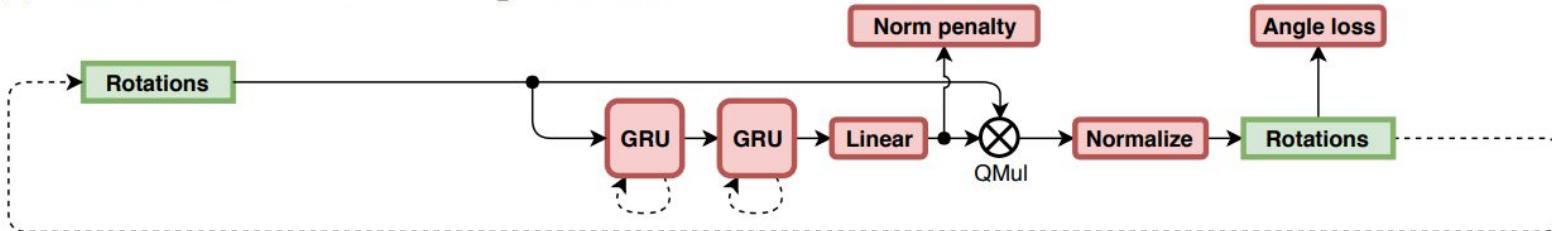
Temporal Activity Detection in Untrimmed Videos with Recurrent Neural Networks

3D свертки



Анализ последовательности поз

(a) Architecture for short-term prediction



(b) Architecture for long-term generation

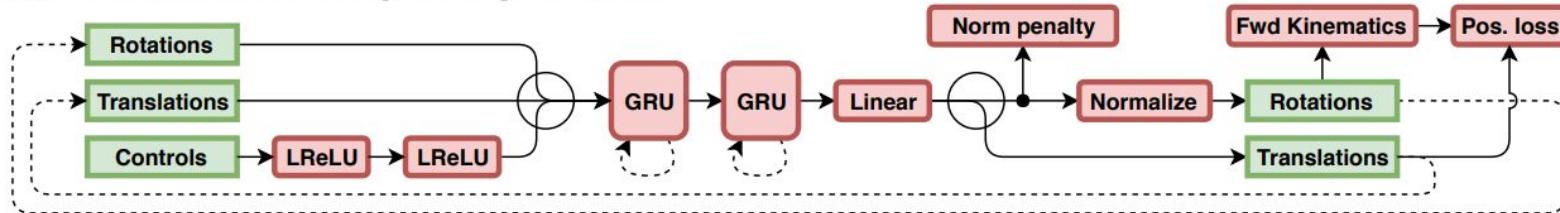


Figure 1: Architectures. “QMul” stands for quaternion multiplication: if included, it forces the model to output velocities; if bypassed, the model emits absolute rotations.

[QuaterNet: A Quaternion-based Recurrent Model for Human Motion](#)

Предсказание последовательности поз

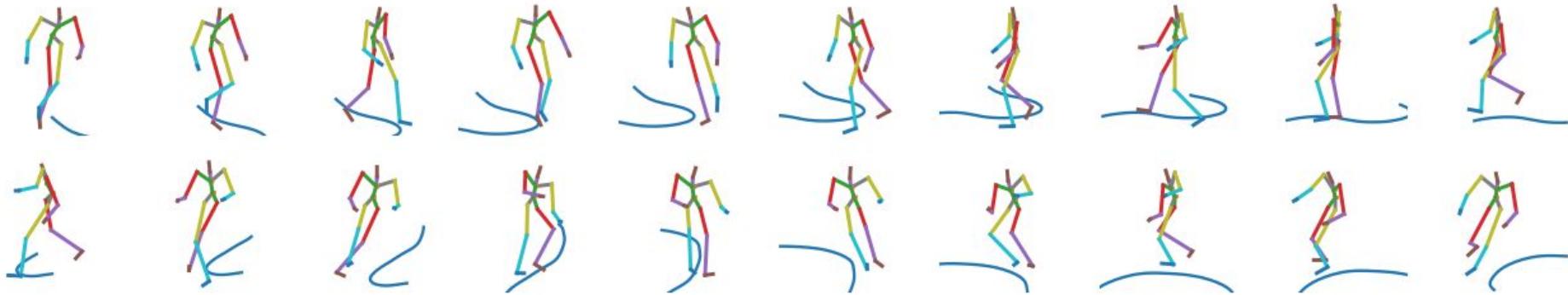
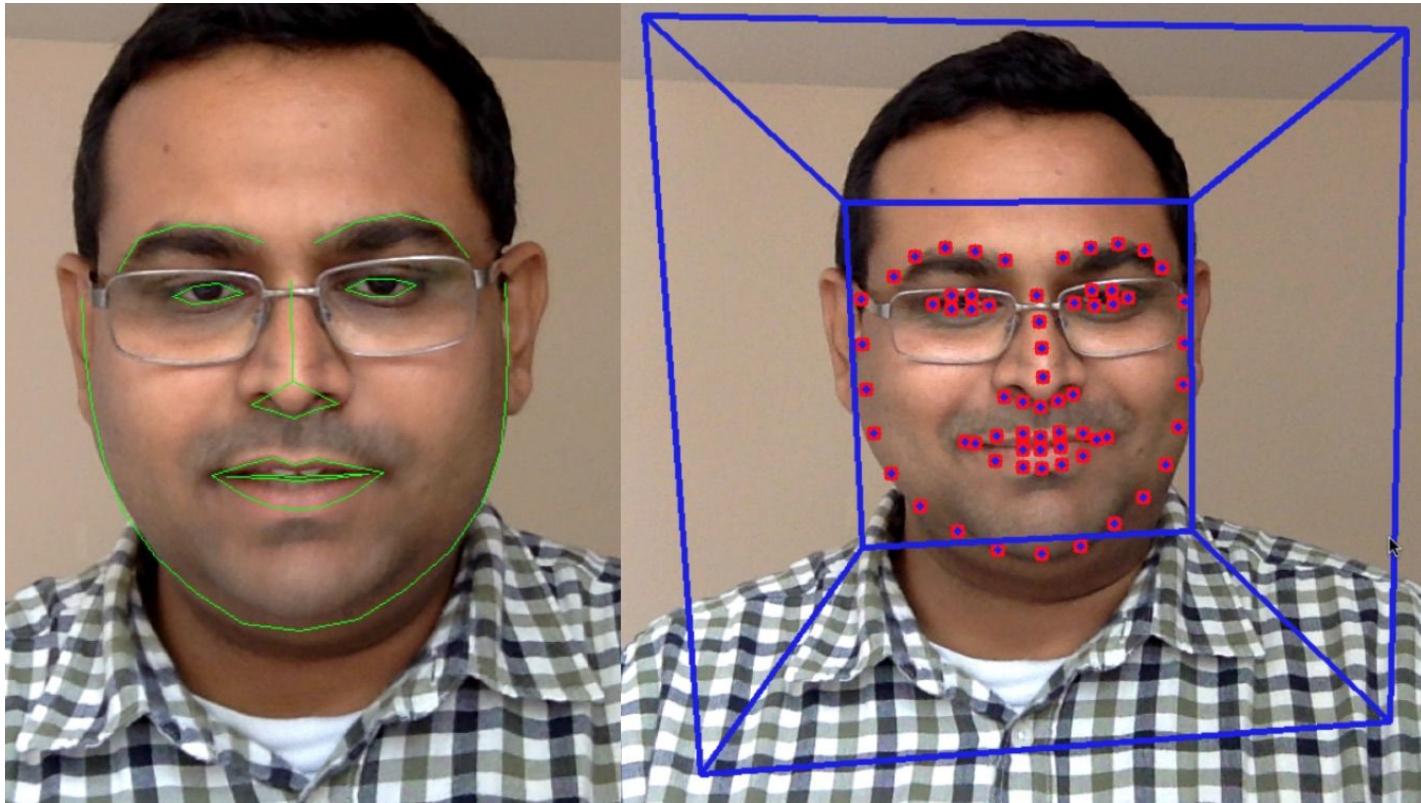


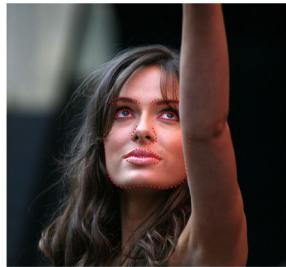
Figure 6: Example of locomotion generation. **Above:** walking. **Below:** running.

Ключевые точки лица направление взгляда

Ключевые точки лица направление взгляда

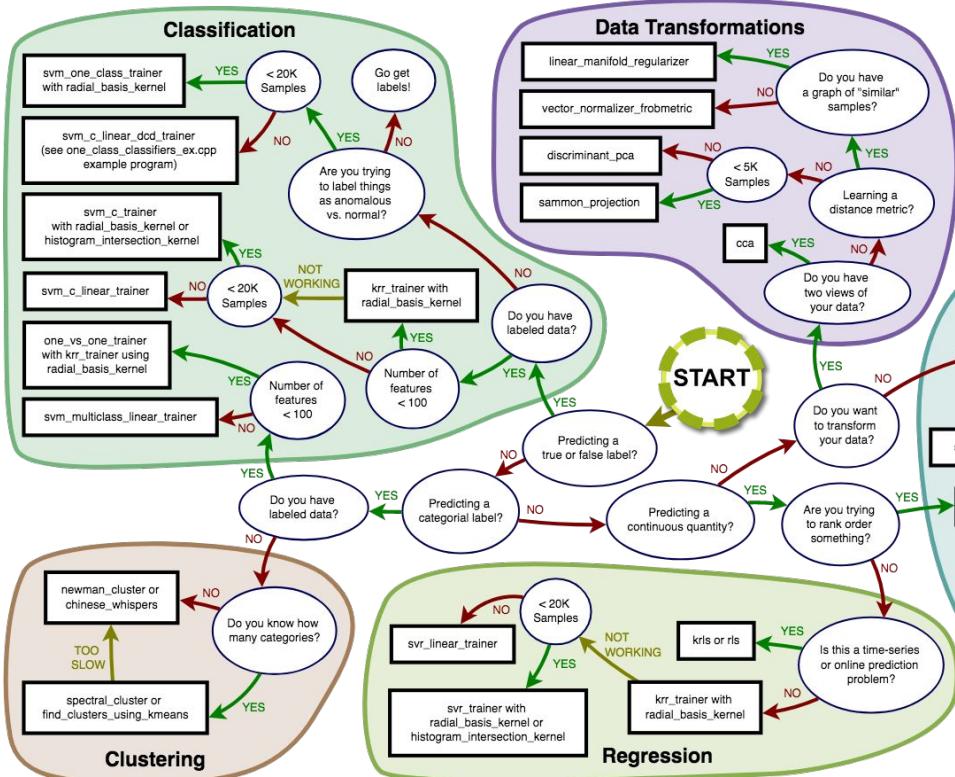


Helen Facial Feature Dataset



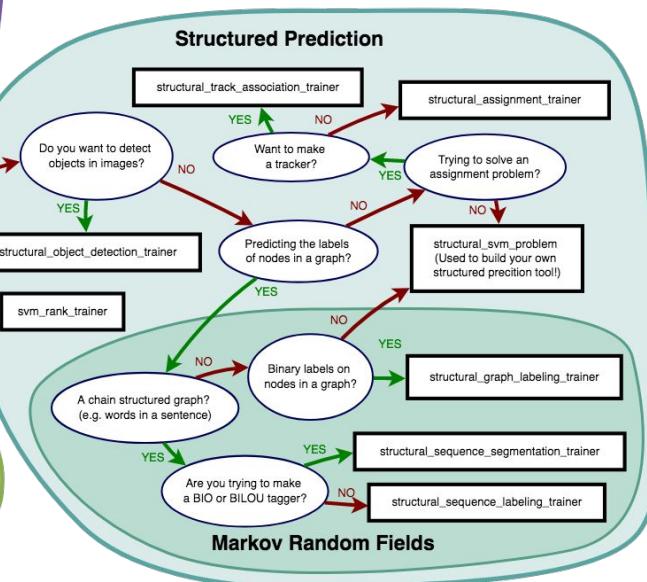
<http://www.ifp.illinois.edu/~vuongle2/helen/>

Dlib

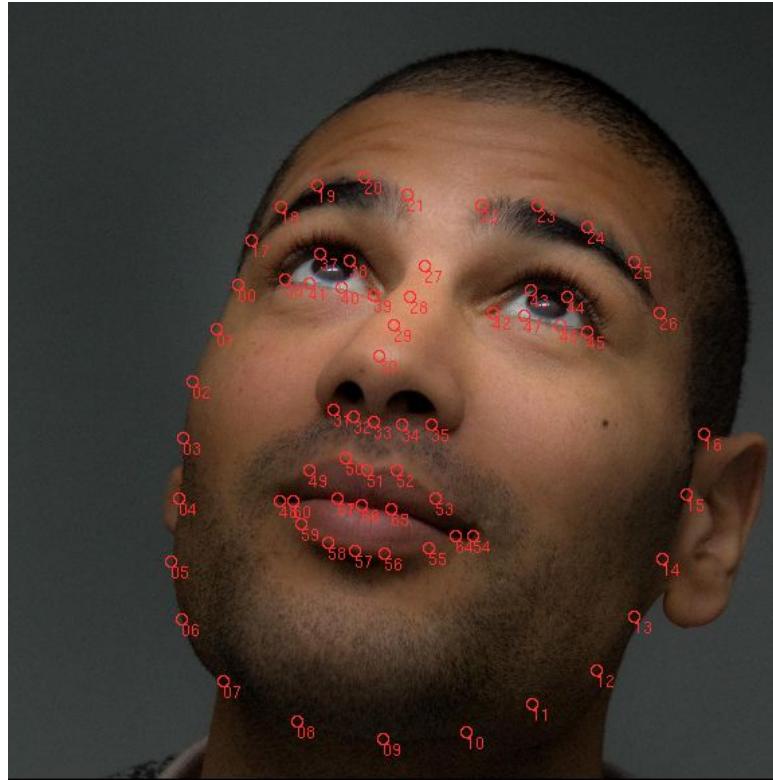


dlib C++ Library

Machine Learning Guide

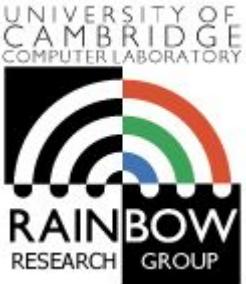


Dlib Face Landmarking



<http://blog.dlib.net/2014/08/real-time-face-pose-estimation.html>

OpenFace: facial behavior analysis toolkit



<https://github.com/TadasBaltrusaitis/OpenFace>

Eye Tracking for Everyone

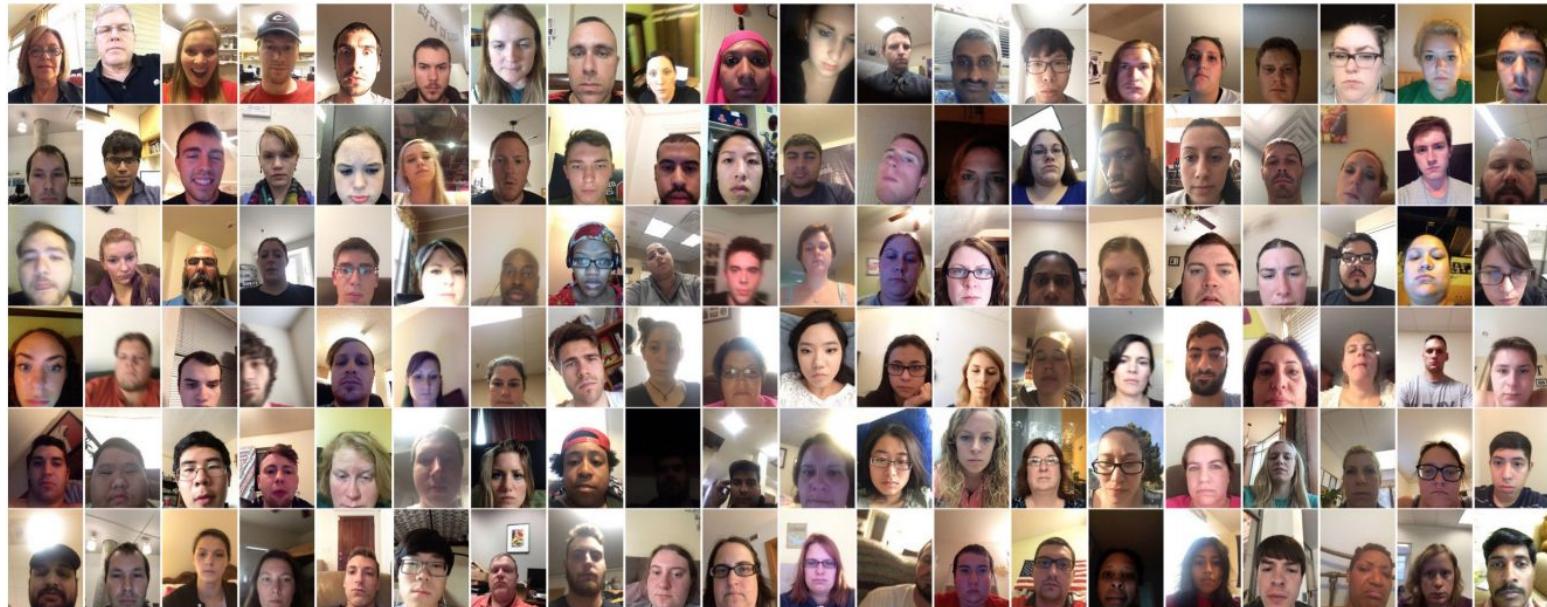
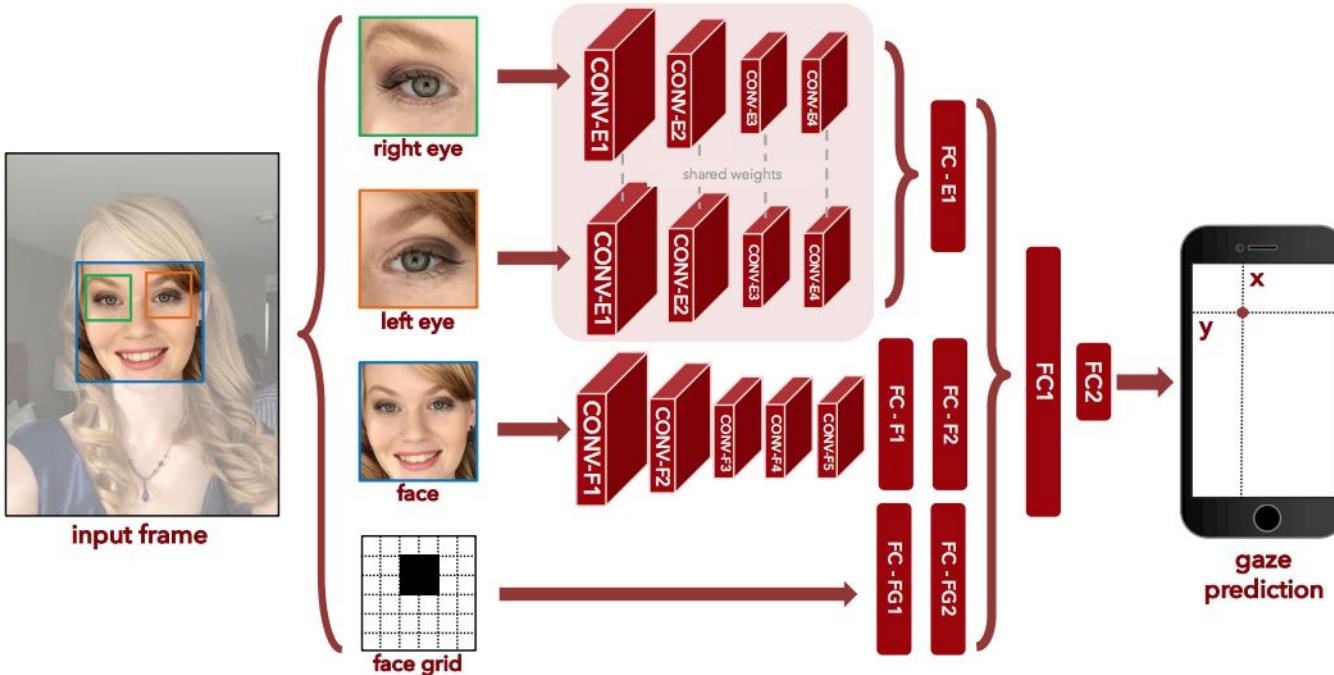


Figure 3: Sample frames from our GazeCapture dataset. Note the significant variation in illumination, head pose, appearance, and background. This variation allows us to learn robust models that generalize well to novel faces.

http://gazecapture.csail.mit.edu/cvpr2016_gazecapture.pdf

Eye Tracking for Everyone



<https://github.com/CSAILVision/GazeCapture>

Датасеты

- [COCO 2018 Keypoint Detection Task](#)
- [DensePose-COCO dataset](#)
- [Fashionista Clothing Parsing](#)
- [CAESAR: The most comprehensive source for body measurement data](#)
- [CMU Graphics Lab Motion Capture Database](#)
- [CMU Multi-Modal Activity Database](#)
- [GazeCapture Dataset](#)
- [State Farm Distracted Driver Detection](#)
- [Challenges in Representation Learning: Facial Expression Recognition Challenge](#)
- [Video Dataset Overview](#)

Дополнительные материалы

- [MPG: Perceptual User Interfaces](#)
- [Eye Tracking for Everyone](#)
- [Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution](#)
- [Detectron: Object detection algorithms](#)
- [Detecting and Recognizing Human-Object Interactions](#)
- [One Millisecond Face Alignment with an Ensemble of Regression Trees](#)