

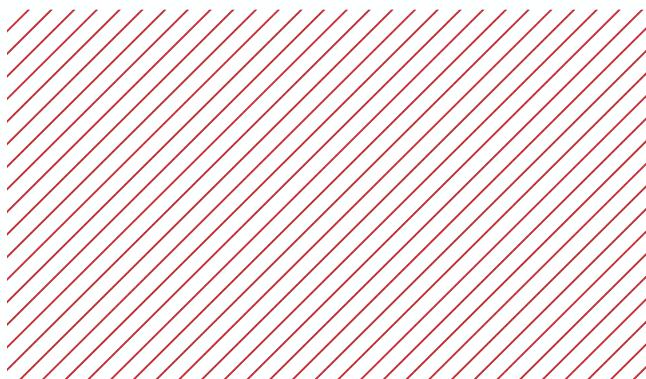
академия
больших
данных



Глубокое нейронные сети для обработки видео

Андрей Бояров

Ведущий программист-исследователь в командах Машинного обучения почты и
Машинного зрения Mail.ru, кандидат физ.-мат. наук





План лекции

- Зачем обрабатывать и анализировать видео
- Отличия анализа видео с помощью DL от анализа изображений
- Основные задачи анализа видео
- Методы для решения этих задач
- Нахождение хайлайтов

Роль видео

55%
of people watch
videos online every
day



3.7 Billion
daily views for
video at facebook



500 Million
hours of videos
watched daily in
Youtube



30%
video ad spend
increased 30% from
2015 to 2016



2.6 X
people spend 2.6x
more time on
pages w/ video
than w/o



1200%
video generates
1200% more
shares than text and
image





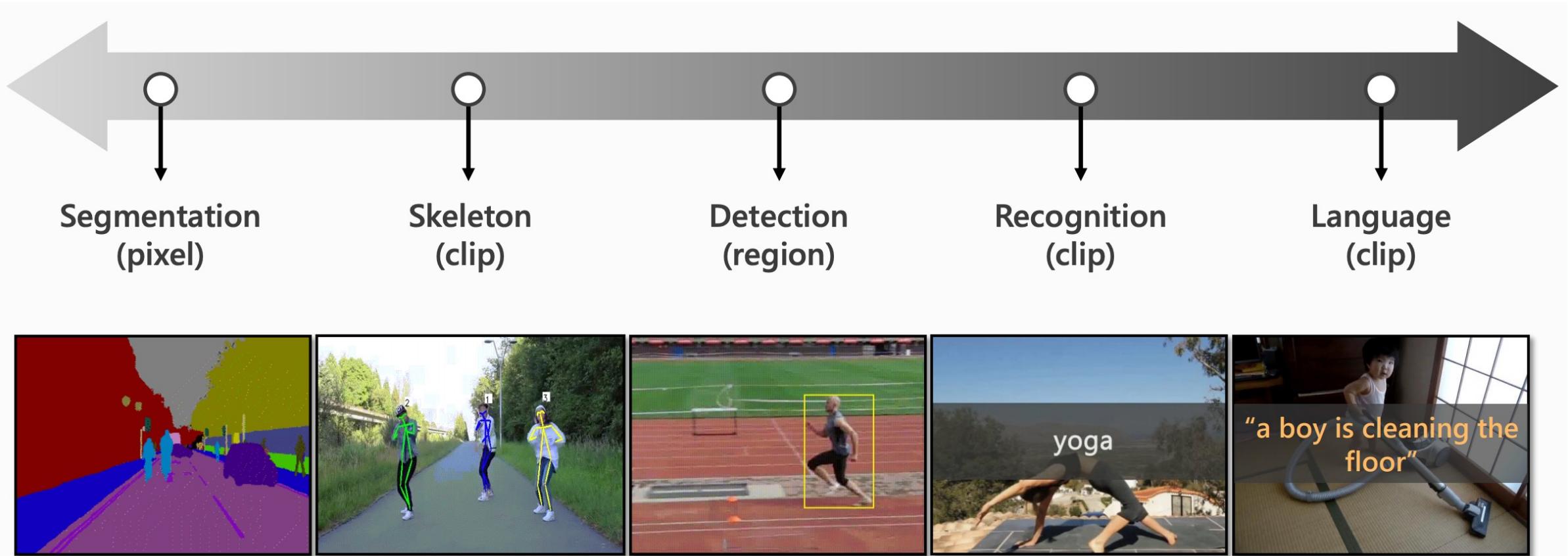
Сложности обработки видео

- Видео делится на большое количество кадров (фреймов)
 - Больше времени на обработку
 - Больше места для хранения
- Важно учитывать временные (temporal) связи между фреймами
- У видео есть звук

«Понимание» видео



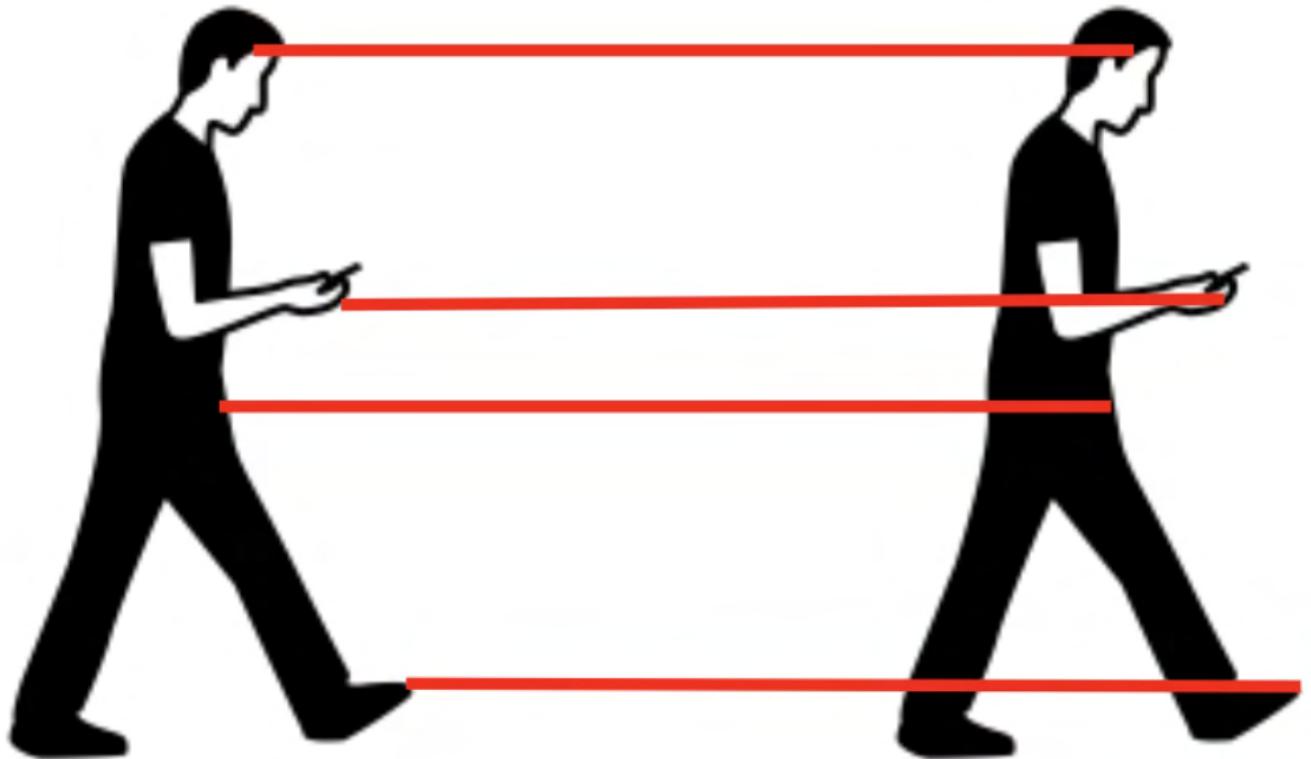
«Понимание» видео: основные задачи



Object tracking

Optical flow

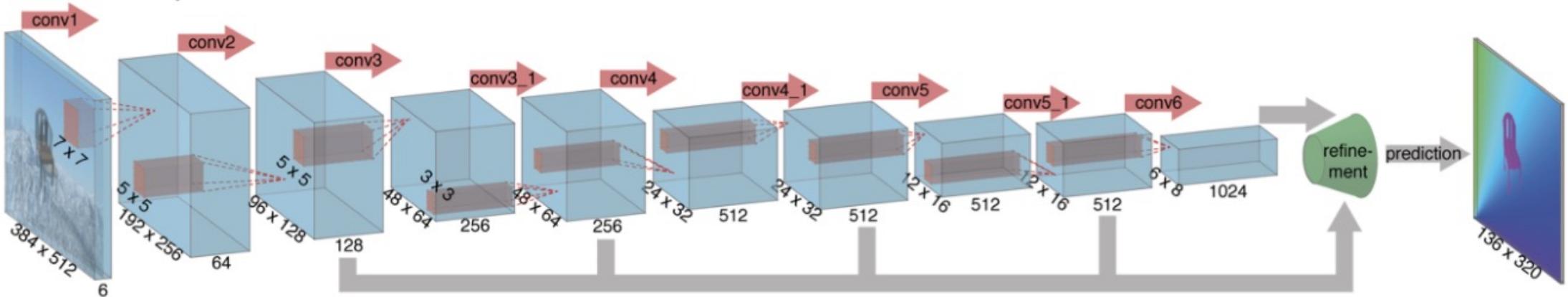
Вычисление сдвига между
пикселями между
двумя фреймами



Frame 0 — Frame 1

FlowNet

FlowNetSimple



Принимает на вход два RGB изображения $\rightarrow 3 + 3$ каналов

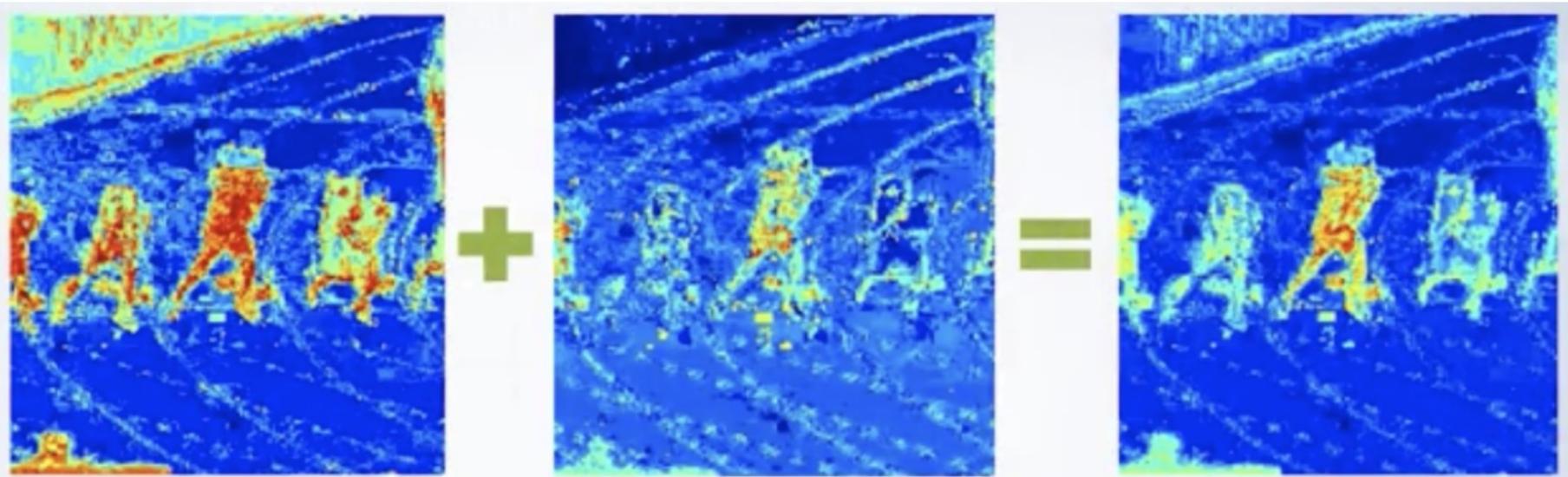
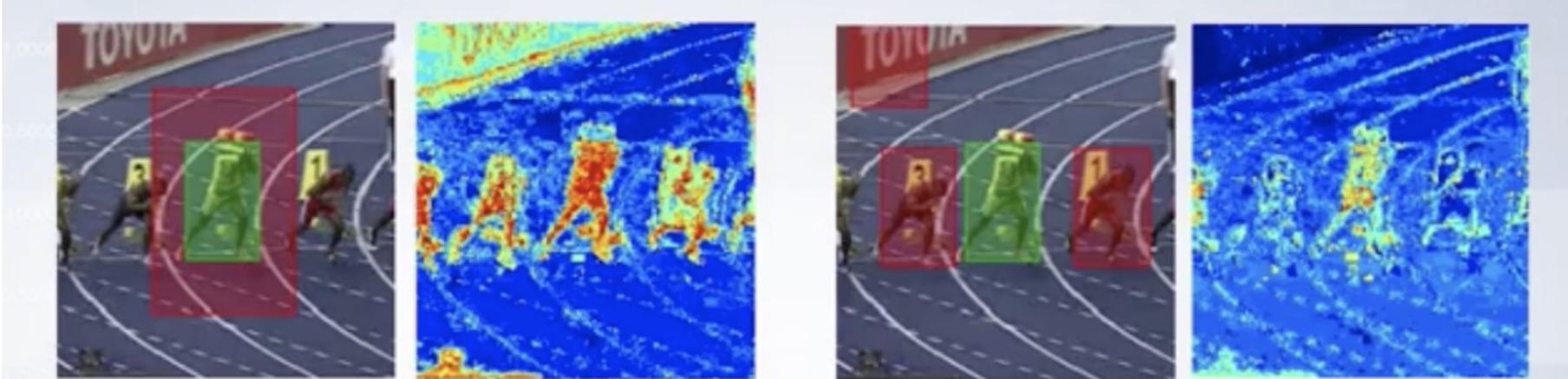
FlowNet: Learning Optical Flow with Convolutional Networks, [Fischer, 2015]



Visual object tracking

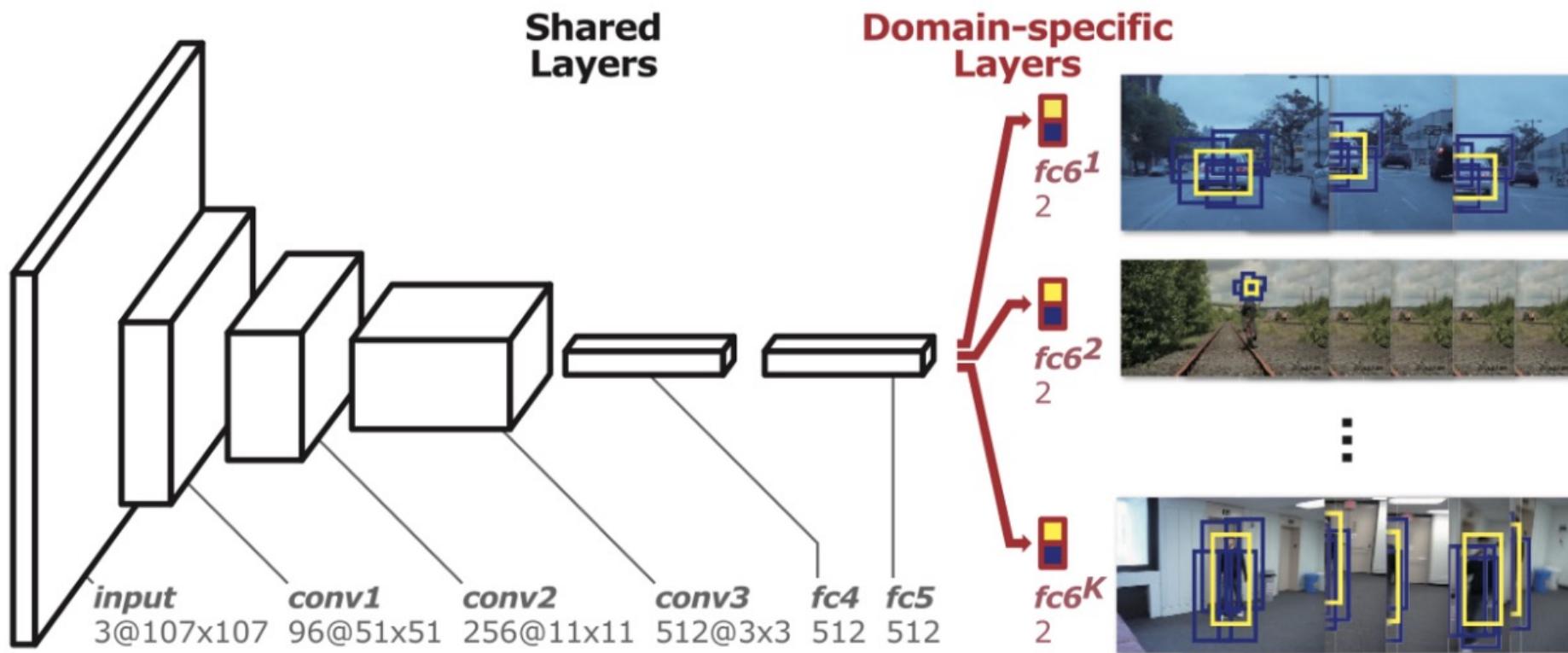
1. Детектируем объект для трекинга
2. Вычисляем его гистограмму цветов
3. Вычисляем цвет фона вокруг объекта
4. Удаляем цвет объекта из всего изображения
5. Получаем основанный на цвете трекер

Visual object tracking



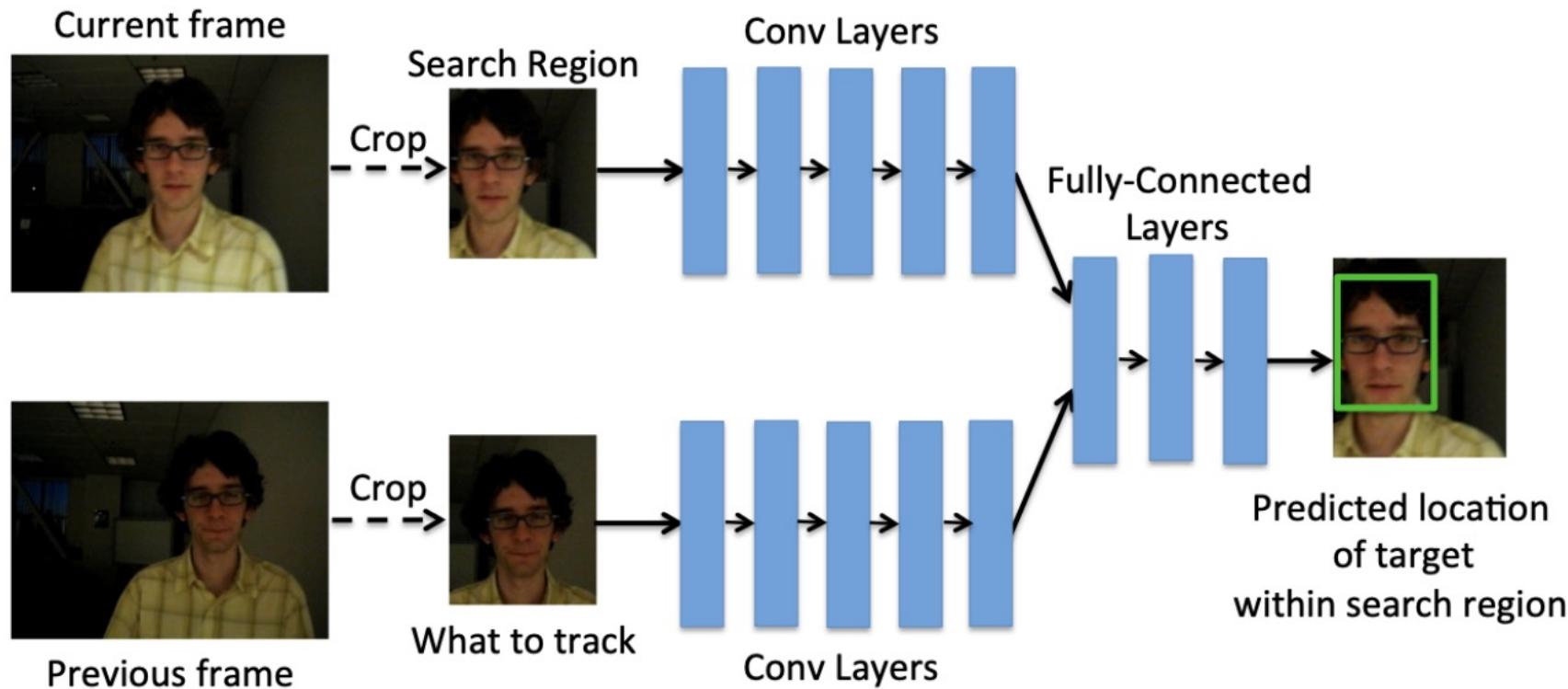
Multi-Domain Net

Детектирует объекты и фон



Learning Multi-Domain Convolutional Neural Networks for Visual Tracking [Nam, 2015]

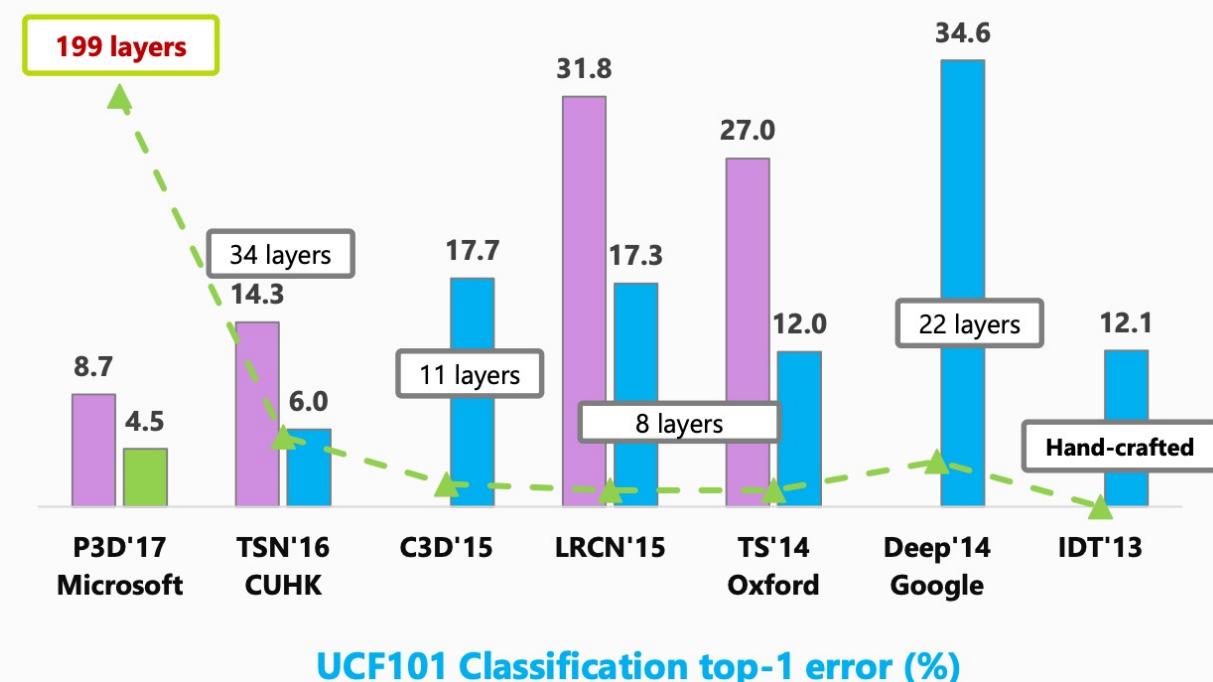
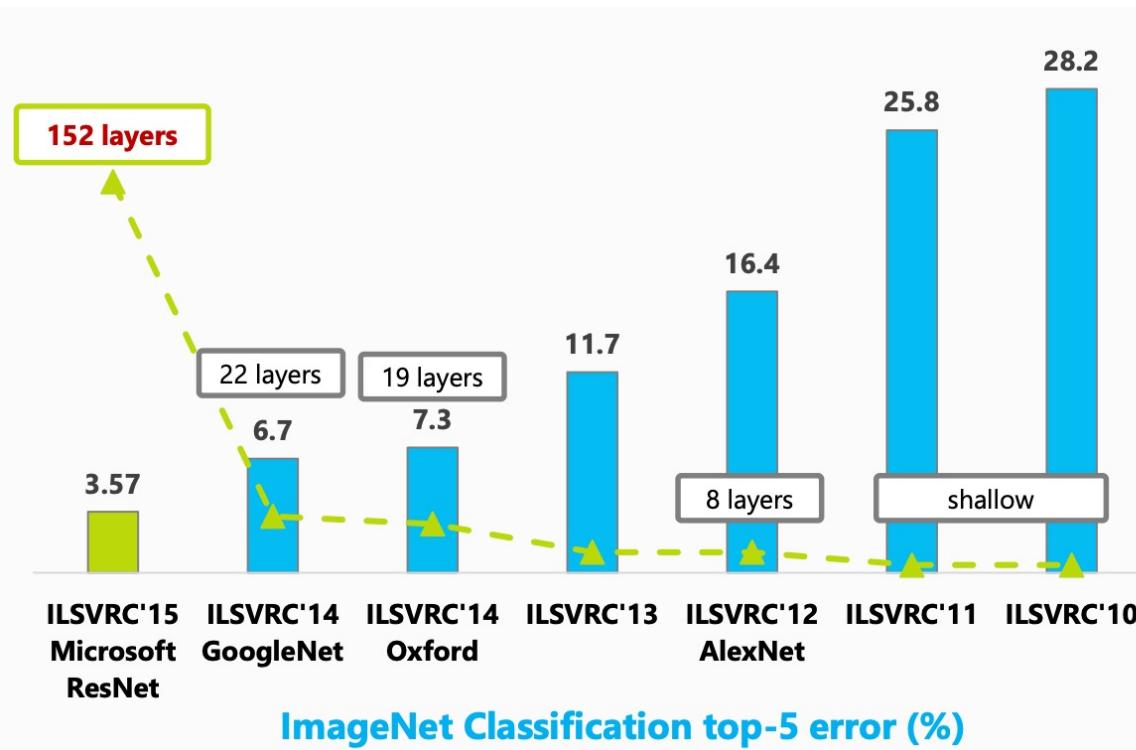
GOTURN (Generic Object Tracking Using Regression Networks)



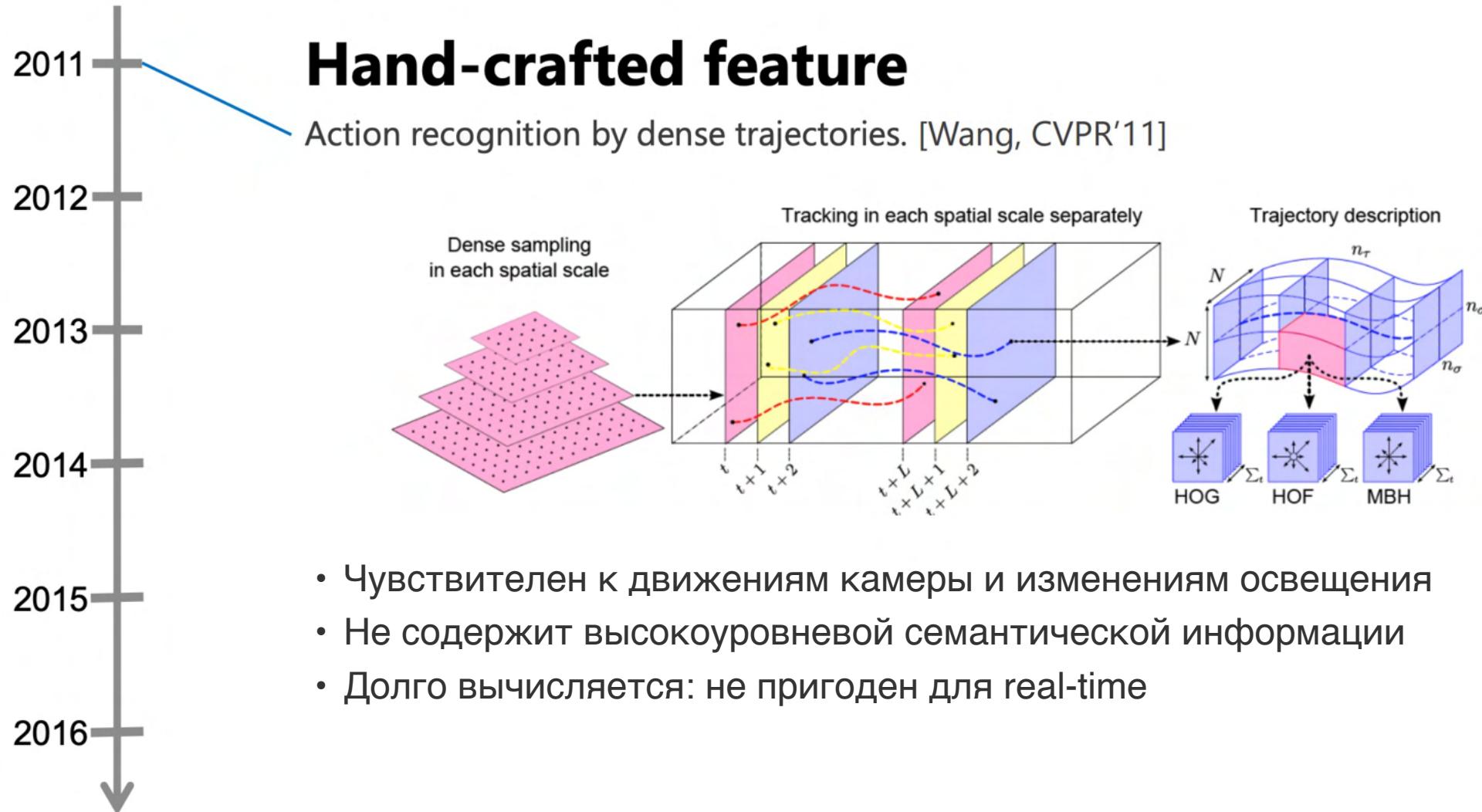
Learning to Track at 100 FPS with Deep Regression Networks, [Held, 2016]

Video representation
learning

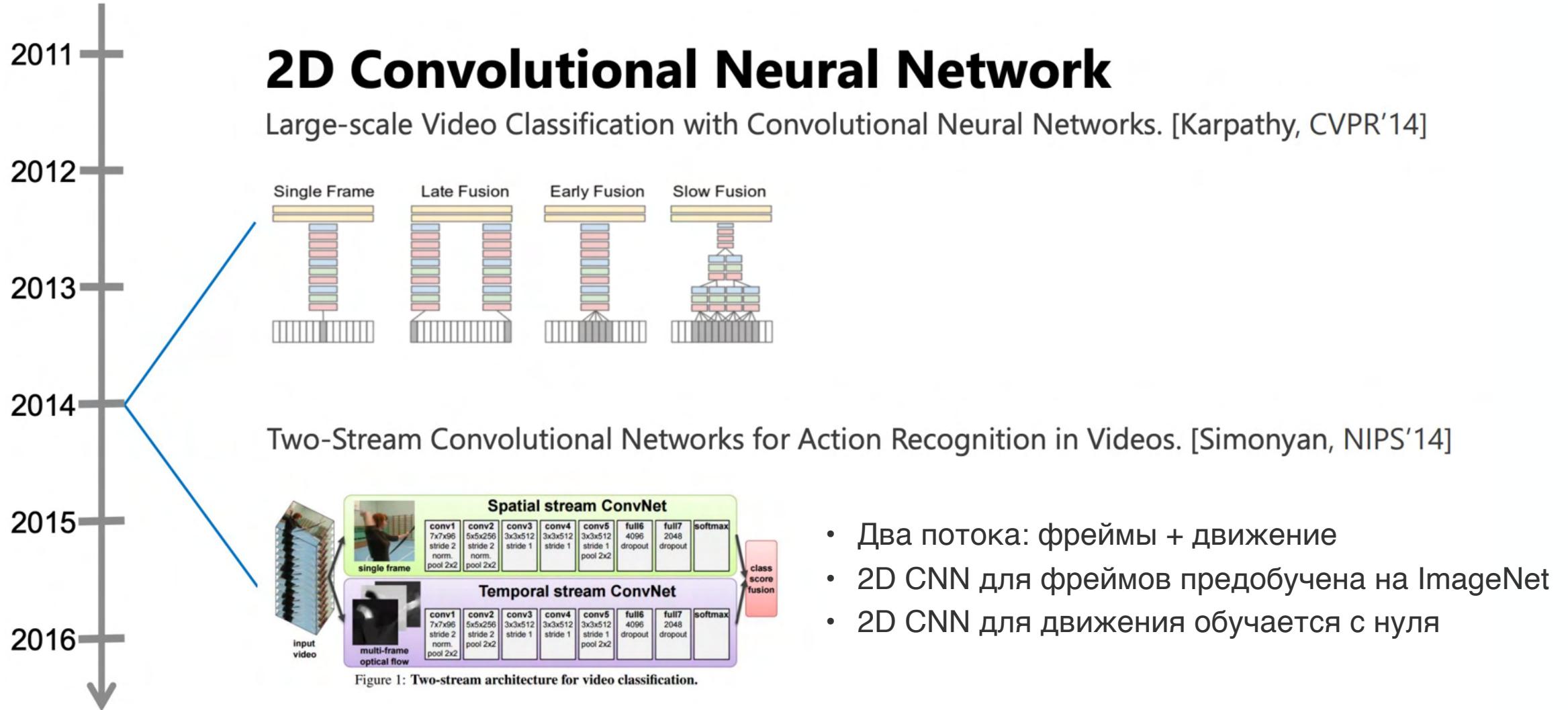
ImageNet для видео



Video representation learning



Video representation learning

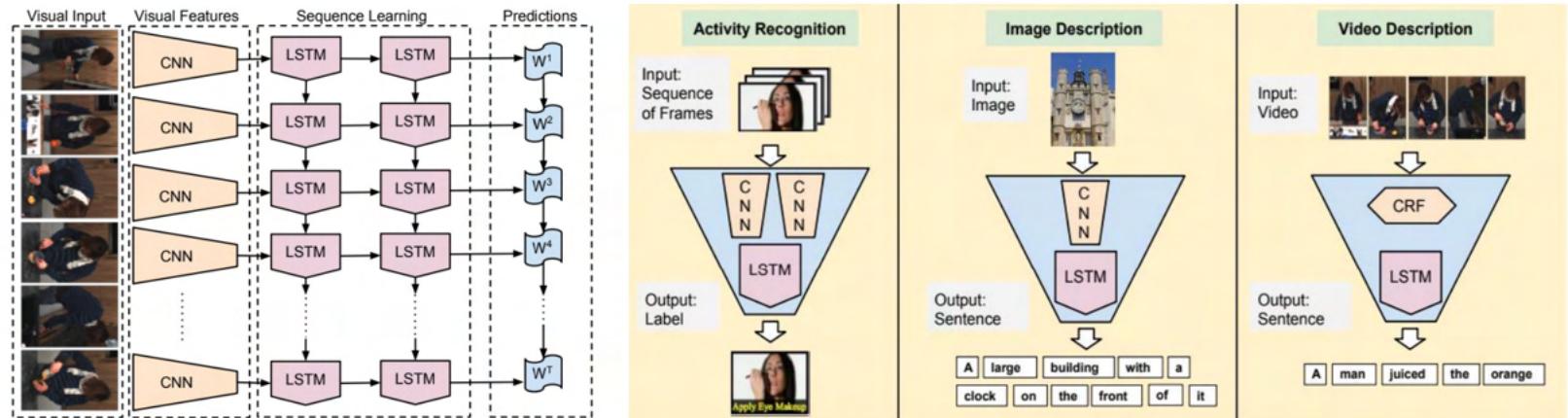


Video representation learning



2D CNN + LSTM (LRCN)

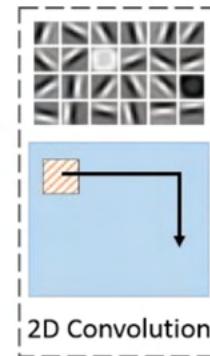
Long-term Recurrent Convolutional Networks for Visual Recognition and Description [Donahue, CVPR'15]



- Смесь CNN и RNN
- Применяется для распознавания действий и описания видео

Video representation learning: Conv 3D

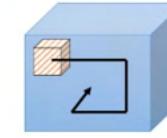
ResNet:
[MSRA, CVPR'16]



3D CNN:
[FAIR & NYU, ICCV'15]



video



3D ConvNet



Video representation learning: Conv 3D

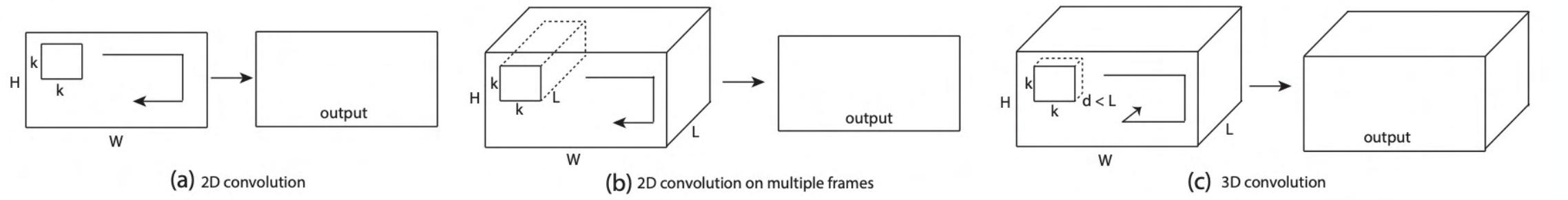
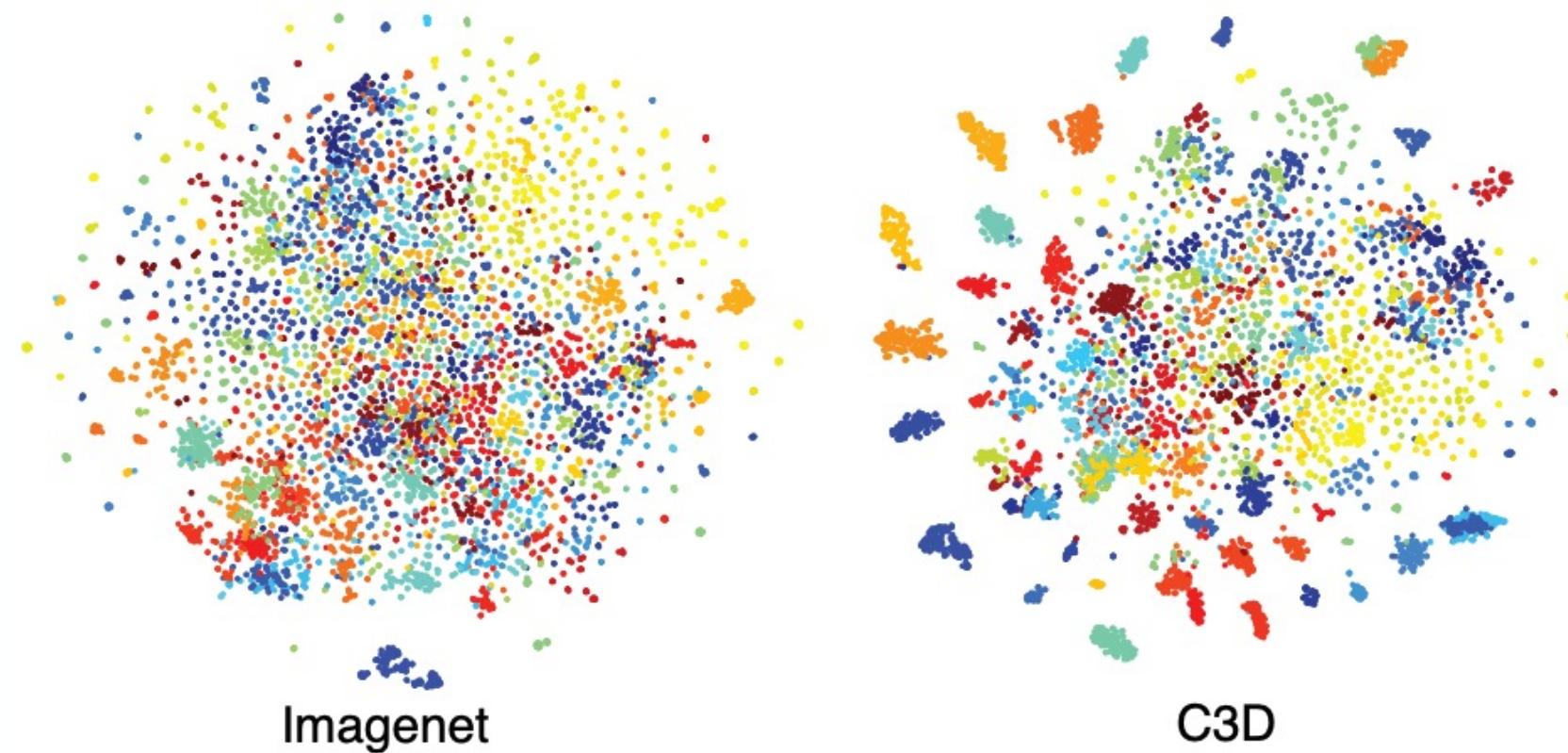
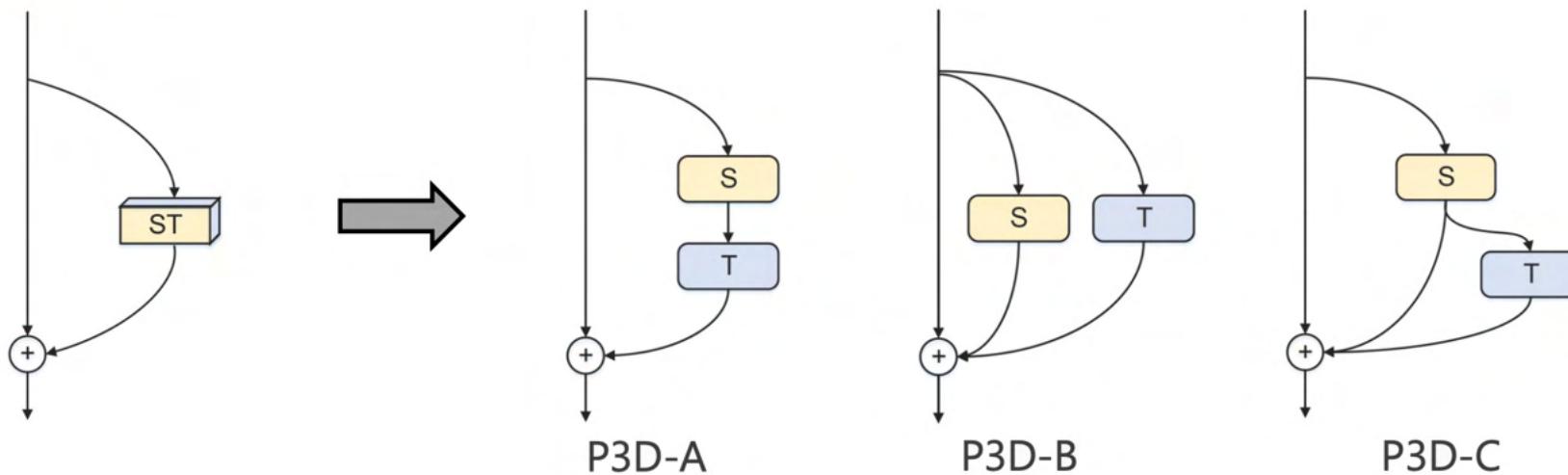
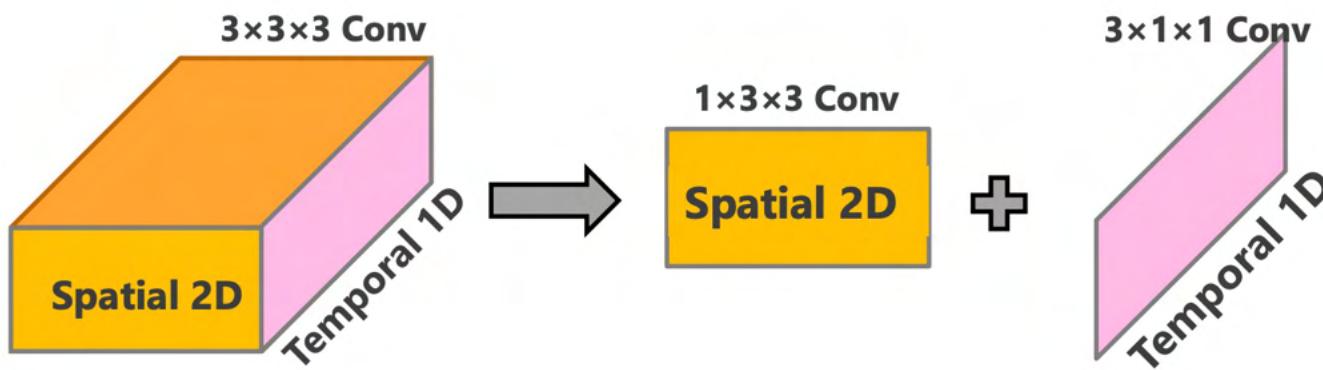


Figure 3. C3D architecture. C3D net has 8 convolution, 5 max-pooling, and 2 fully connected layers, followed by a softmax output layer. All 3D convolution kernels are $3 \times 3 \times 3$ with stride 1 in both spatial and temporal dimensions. Number of filters are denoted in each box. The 3D pooling layers are denoted from pool1 to pool15. All pooling kernels are $2 \times 2 \times 2$, except for pool1 is $1 \times 2 \times 2$. Each fully connected layer has 4096 output units.

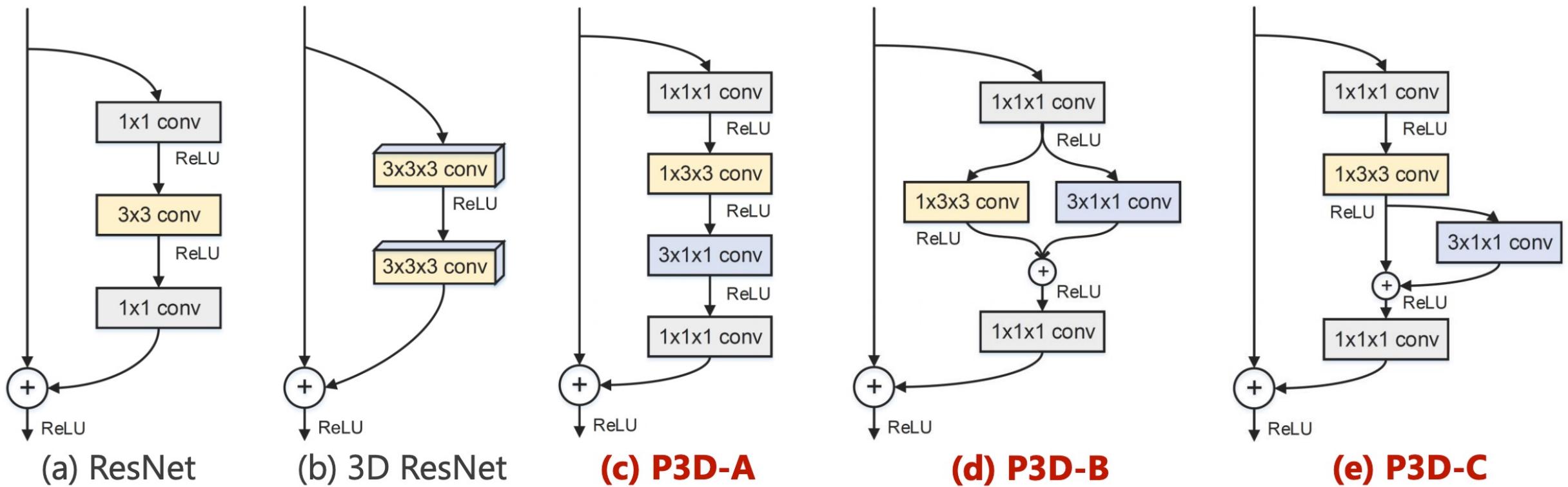
Video representation learning: Conv 3D



Video representation learning: Pseudo 3D ResNet

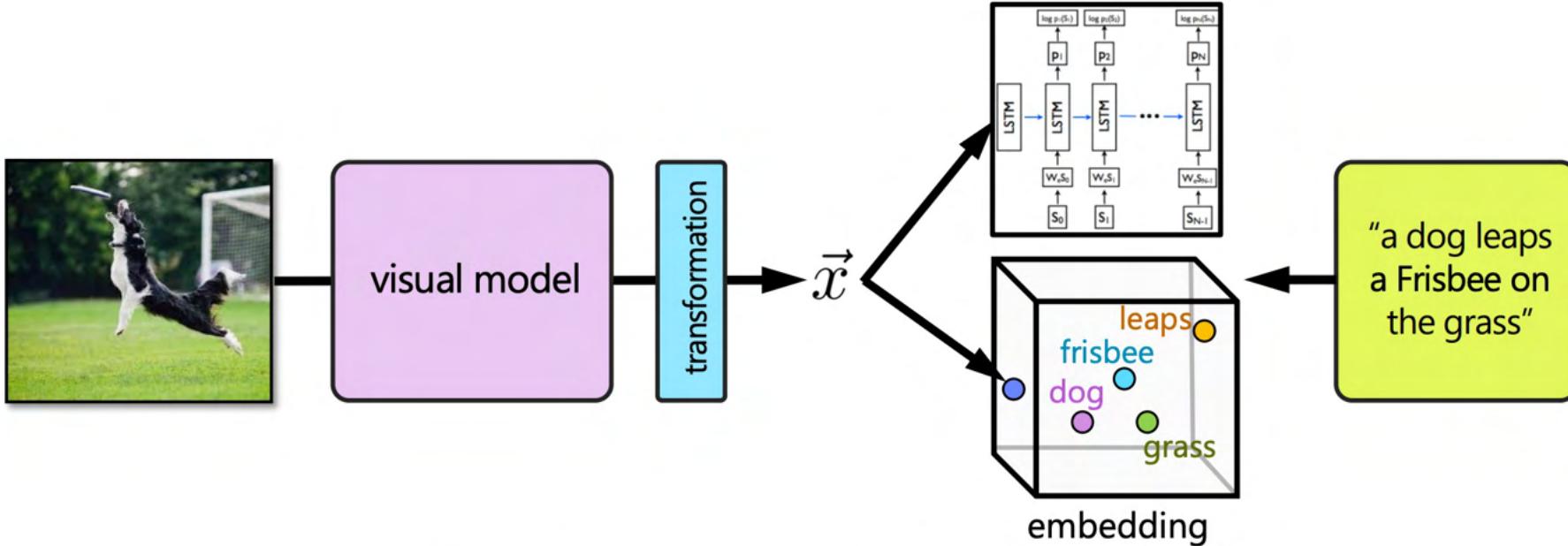


Video representation learning: Pseudo 3D ResNet



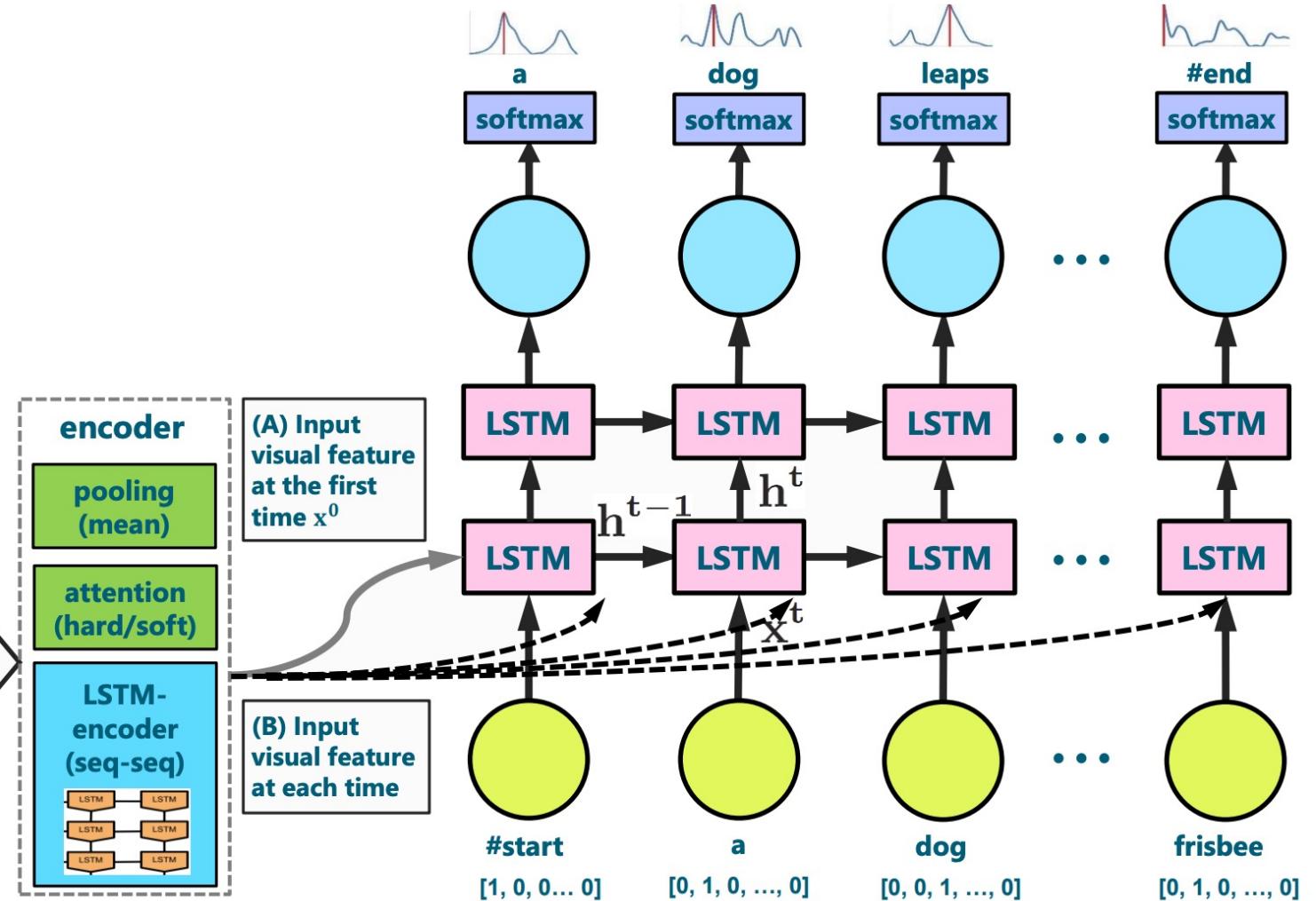
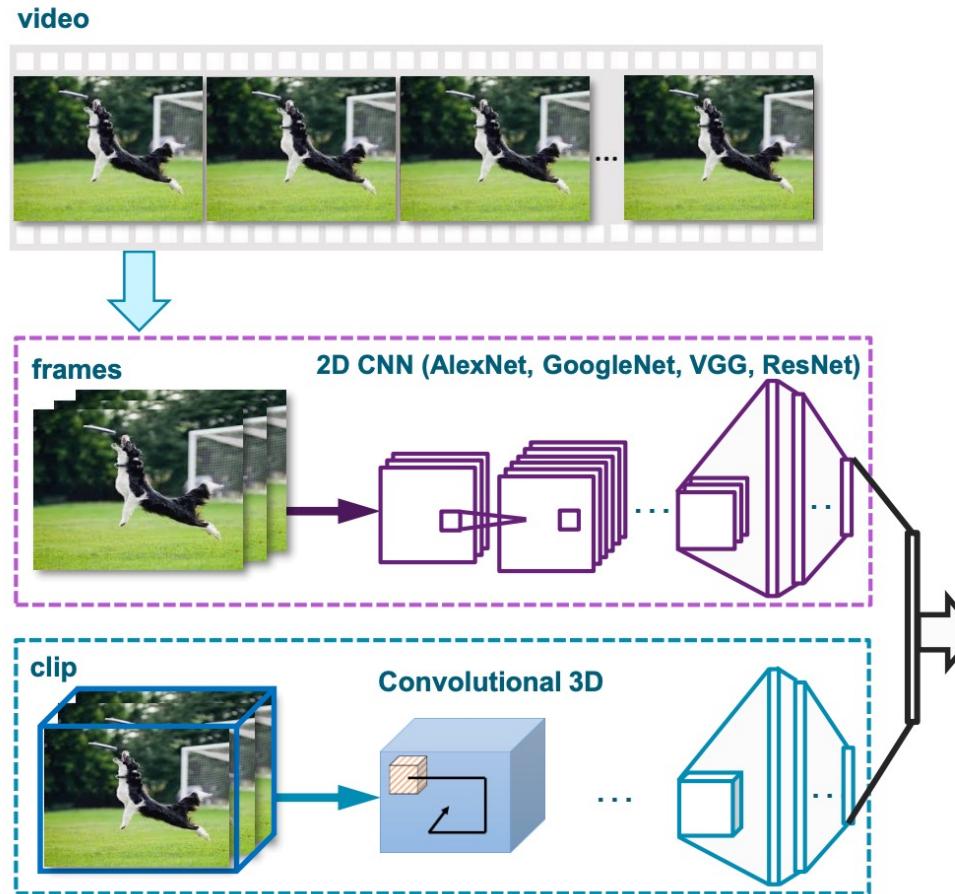
Video captioning

Video captioning: idea

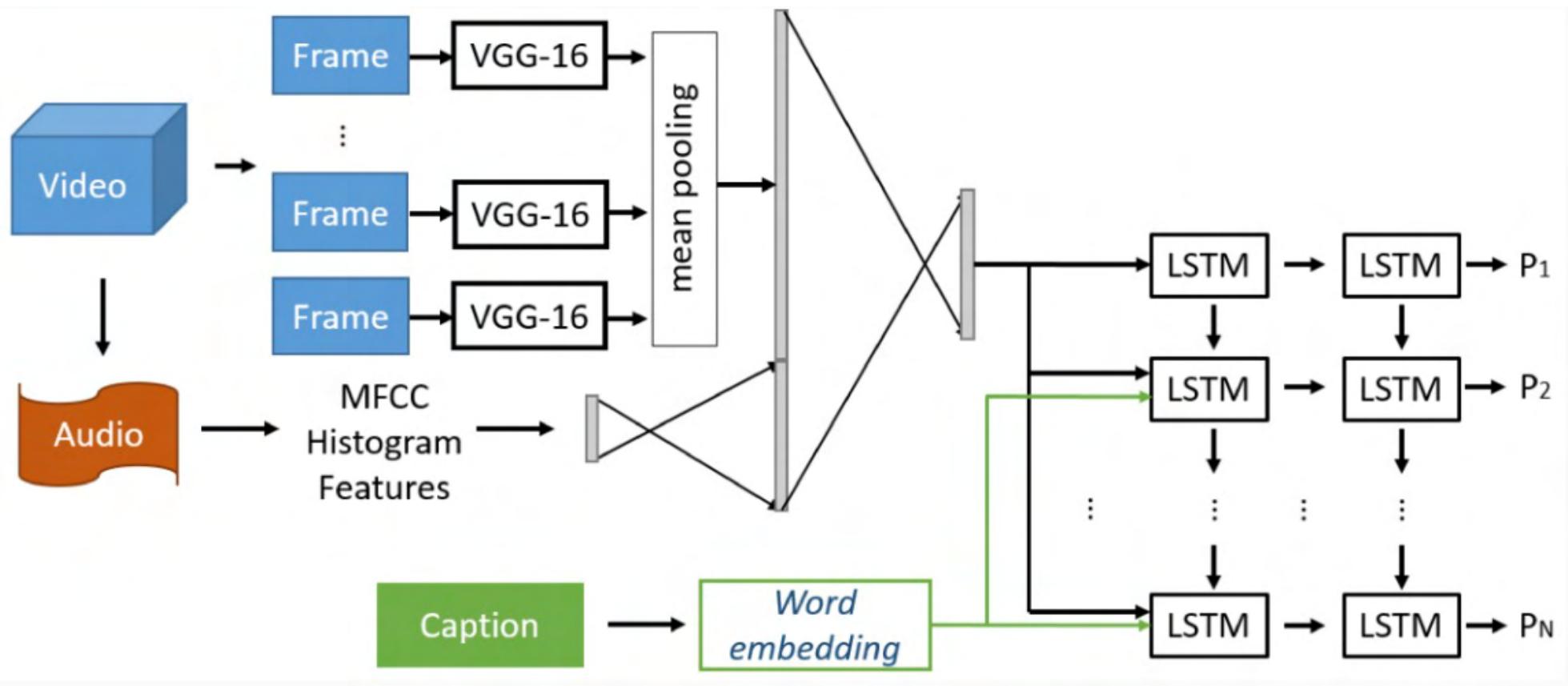


- Transforming an image/clip to a vector in visual space
 - CRF, CNN, Semantic Vector, CNN+Attention
- Transforming description to a vector in semantic space
 - Collection of words (BoW), sequence of words (RNN)
- Creating an embedding space
 - Language template (FGM, ME), RNNs (Encoder-Decoder), LSTM
- Methodologies
 - Search-based
 - Language template-based
 - Sequence learning-based
 - Generation: learning-decoder
 - Translation: encoder-decoder

Video captioning

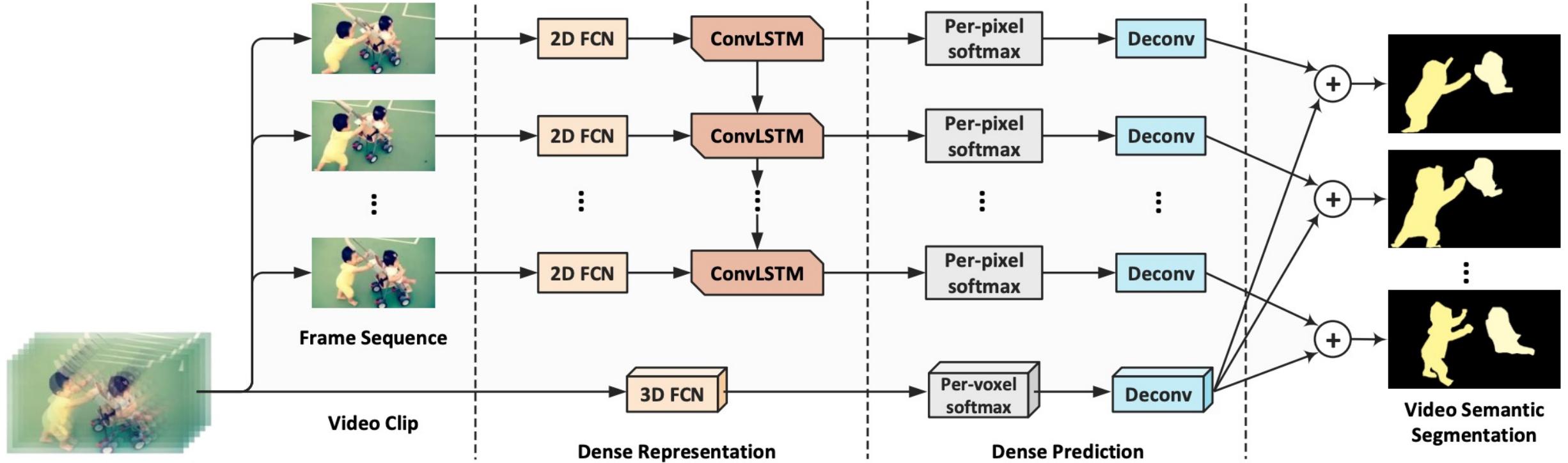


Video captioning



Video segmentation

Video segmentation



Highlights detection

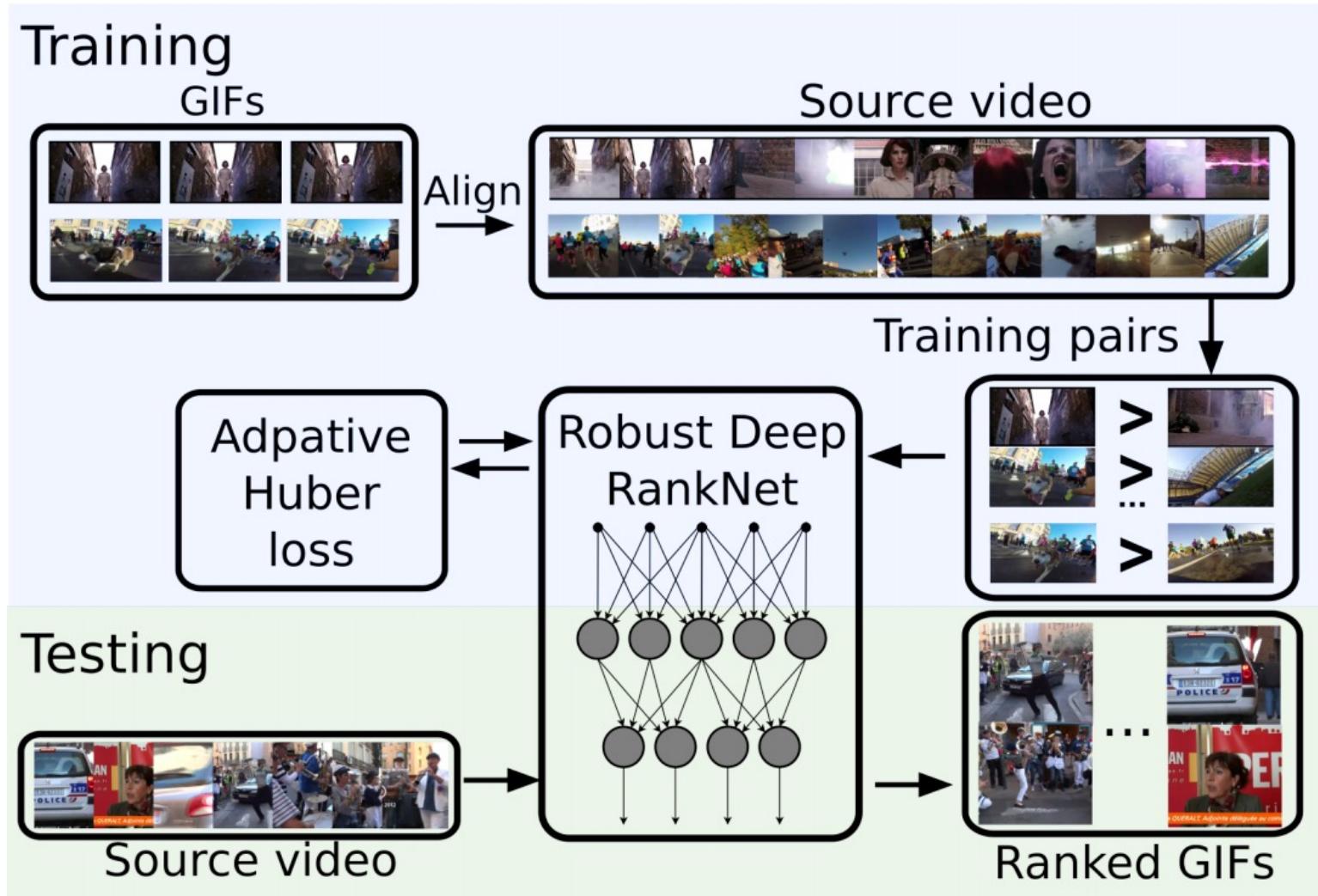


Video highlights

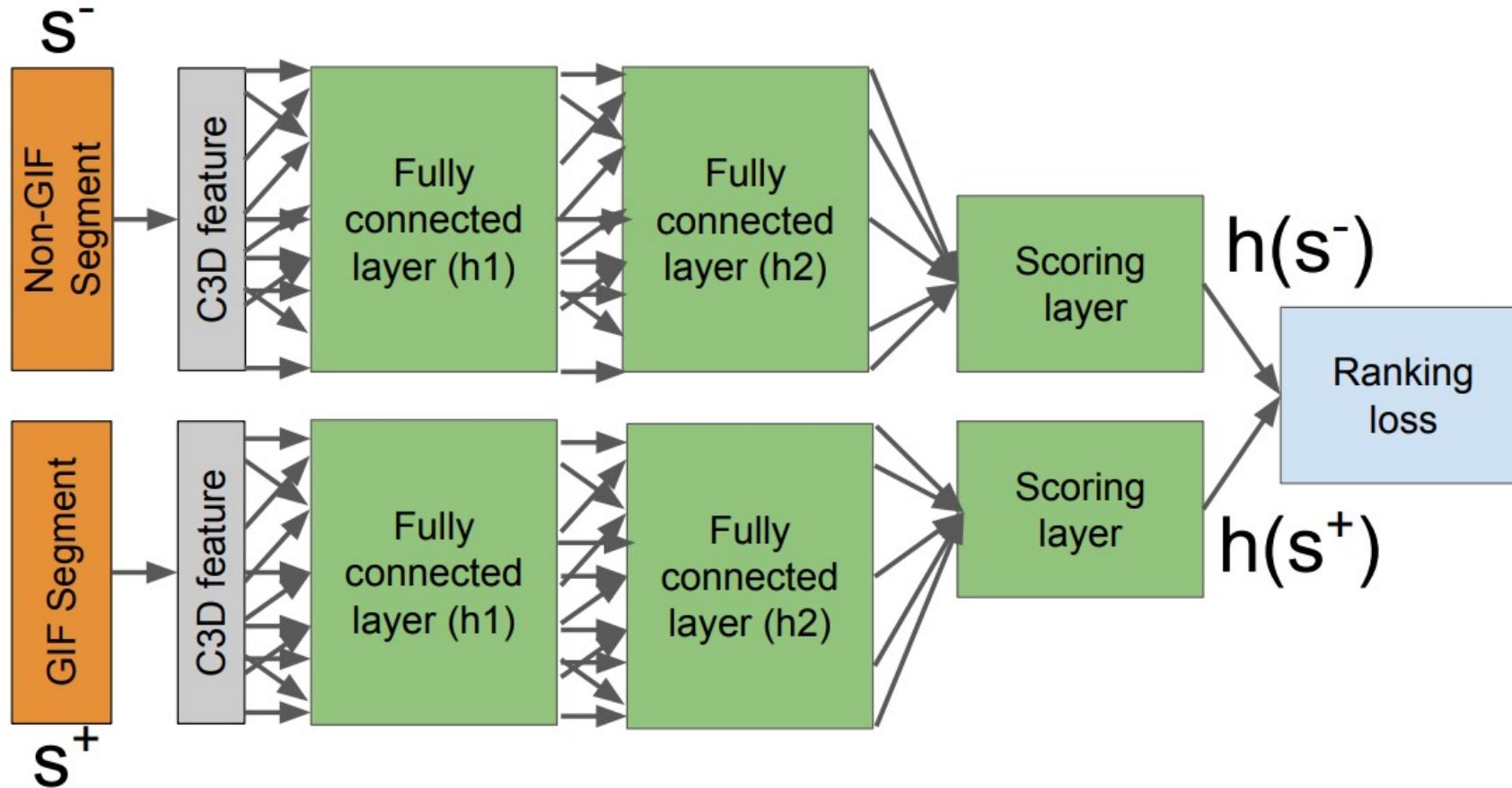
- Хайлайт — это важная, интересная часть видео
- Может использоваться для выделения основной информации о сюжете видео
- Разделим видео на сцены (segments) и будем ранжировать их (чем выше значение ранга, тем выше вероятность попадания сцены в хайлайт)
- Обучаемся на датасете Video2GIF
 - 120K видео из 80K видео

Video2GIF: Automatic Generation of Animated GIFs from Video, [Gygli, 2016]

Video highlights detection pipeline



Video highlights detection net



Loss function

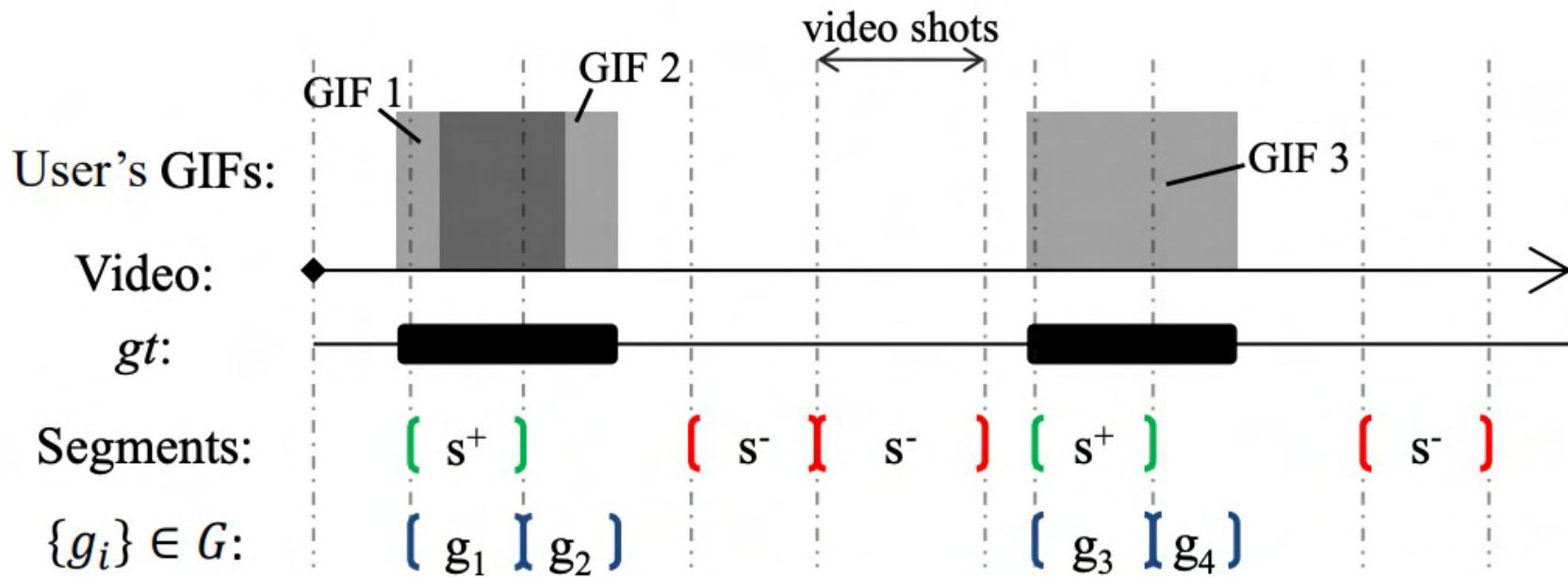
$$h(s^+) > h(s^-), \quad \forall (s^+, s^-) \in S.$$

$$l_p(s^+, s^-) = \max(0, 1 - h(s^+) + h(s^-))^p$$

$$l_{\text{Huber}}(s^+, s^-) = \begin{cases} \frac{1}{2}l_2(s^+, s^-), & \text{if } u \leq \delta \\ \delta l_1(s^+, s^-) - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases}$$

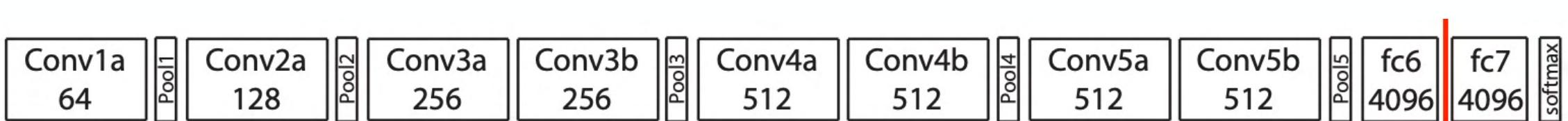
$$L(\mathcal{D}, \mathbf{W}) = \sum_{S_i \in \mathcal{D}} \sum_{(s^+, s^-) \in S_i} l_{\text{Huber}}(s^+, s^-) + \lambda \|\mathbf{W}\|_F^2$$

Positive / negative segments



Video segments embeddings

- Используем Conv 3D
- Sports-1M Dataset
 - 1 133 158 видео, 487 классов



YouTube preview

YouTube

Search

SIGN IN

Home

Trending

Subscriptions

Library

History

Sign in to like videos, comment, and subscribe.

BEST OF YOUTUBE

Music

Sports

Gaming

News

Live

Fashion

360° Video

Browse channels

Trending

Bad Boys For Life Trailer 2

Will Smith

886K views • 12 hours ago

HIGHLIGHTS | Canelo vs. Sergey Kovalev

DAZN USA

5.1M views • 2 days ago

Know Your Bro with Chris and Scott Evans

The Tonight Show Starring...

1.5M views • 21 hours ago

Cowboys vs. Giants Week 9 Highlights | NFL 2019

NFL

1.9M views • 21 hours ago

Recommended

Using Basic Geometry to play Poly Bridge

RTGame

4.9M views • 1 year ago

The History of Europe: Every Year

Cottreau

12M views • 1 year ago

Diablo 4 - Official Announcement Cinematic...

IGN

3.1M views • 4 days ago

How does a puma purr: 10 minutes of relaxation with...

I_am_puma

709K views • 2 weeks ago

Call of Duty: Modern Warfare

Viewing A Solar Eclipse From

NBA Highlights

Stephen Curry

What highlights would you like to see from Stephen Curry?

Top Plays

Last Game

Playoffs

Enjoy the highlights!



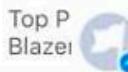
Watch This!

Stephen Curry beats the buzzer vs. the Thunder 05/30/2016

Watch

Watc

Top P
Blazei



LeBron James

What highlights would you like to see from LeBron James?

Top Plays

Last Game

Playoffs

Enjoy the highlights!



Watch This!

LeBron James rises for the jam! 05/27/2016

Watch



Watc

LeBor
hand g

Do you want to receive LeBron James' top plays after each game?

IBM Wimbledon Highlights

Wimbledon AI Highlights With Watson™

IBM

All Events | All Rounds | All Statistics | Day 3 Wed 3 July

Type Player Name COVERAGE 13169 Points 48 Hours HIGHLIGHTS 14 Produced 0 Published TOP 5 MAIN EVENTS HIGHLIGHTS 0 Published 0 Secs Published 900 Secs Remaining

Set Excitement Threshold: 0.0 0.3 1.00

Most Excitement | Most Recent

0.90 • Wednesday, 3 July 2019, 11:42
Stan Wawrinka vs Reilly Opelka
Set 1: 30-40 : Break Point; Wawrinka loses the point with a forehand forced error.

0.89 • Wednesday, 3 July 2019, 11:11
Su-Wei Hsieh vs Kirsten Flipkens
Set 1: 0-40 : Break Point; Hsieh wins the point with a backhand winner.

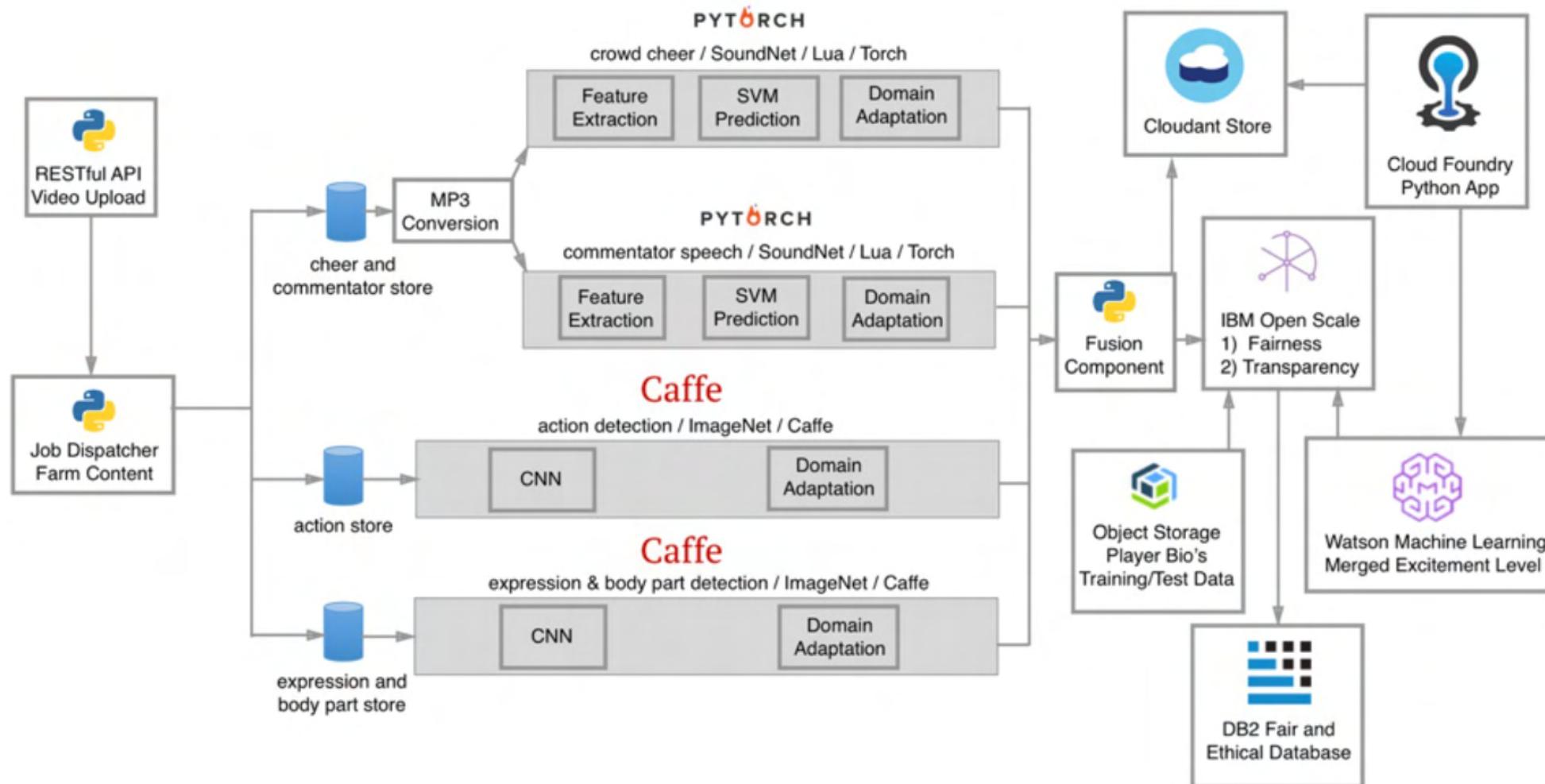
0.88 • Wednesday, 3 July 2019, 12:16
Elina Svitolina vs Margarita Gasparyan
Set 2 : 30-40 : Break Point; Svitolina loses the point with a forehand forced error.

0.86 • Wednesday, 3 July 2019, 13:06
Jeremy Chardy vs David Goffin
Set 1: 15-40 : Break Point; Goffin wins the point with a forehand winner.

Stan Wawrinka vs Reilly Opelka
Set 1: 30-40 : Break Point; Wawrinka loses the point with a forehand forced error.

0.65 Crowd Cheering
0.70 Match Analysis
0.03 Player Gestures
0.90 Overall Excitement

IBM Wimbledon Highlights





Заключение

- Анализ видео на порядок сложнее анализа изображений
 - Больше данных
 - Дольше время обработки
- Важно учитывать порядок фреймов по времени
- Есть звук (отдельная тема)
- Используется смесь из нескольких моделей
 - Conv 3D
 - RNN