

Computer Vision with Videos

(Introduction)

2019. 1. 14.
Jihun **Kim**, Hanyang Univ.

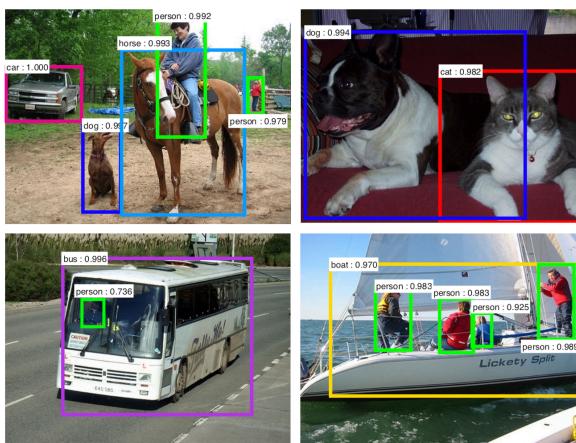
Recap: Computer Vision with Images

Visual Classification



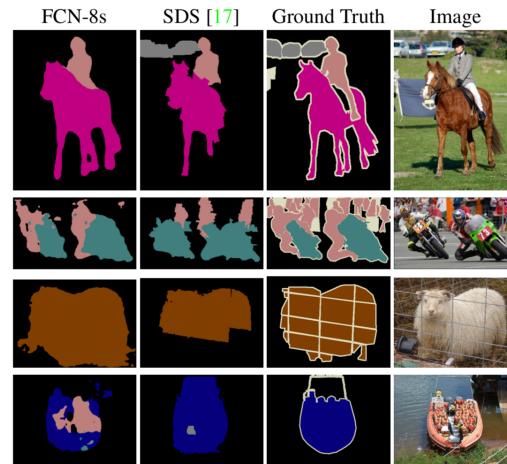
(Alex et al. Imagenet classification with deep convolutional neural networks. NIPS 2012.)

Object Detection



(Ren et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NIPS 2015.)

Semantic Segmentation



(Long et al. Fully convolutional networks for semantic segmentation. CVPR 2015.)

Instance Segmentation



(He et al. Mask R-CNN. ICCV 2017.)

AlexNet, 2012
SqueezeNet, 2016

R-CNN, 2013
Fast/Faster R-CNN, 2015
YOLO, 2015

FCN, 2015

Mask R-CNN, 2017

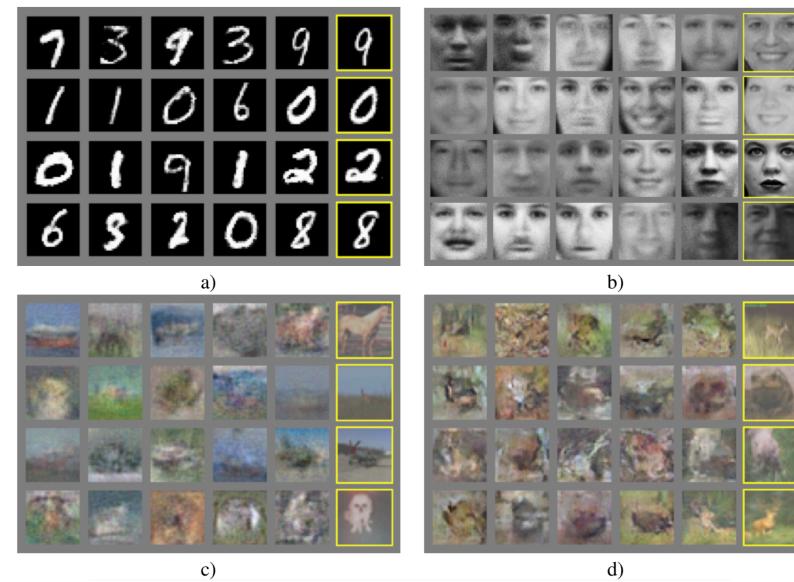
Recap: Computer Vision with Images

Image Captioning



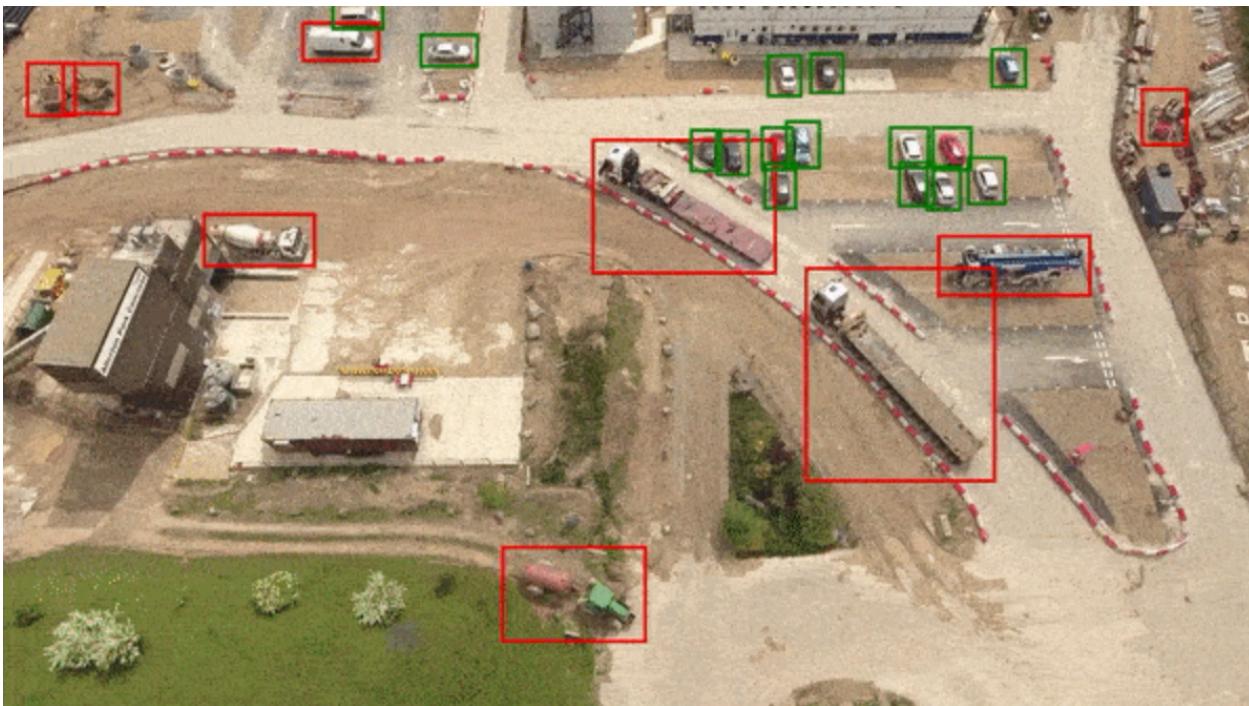
(Oriol, et al. Show and tell: A neural image caption generator. CVPR 2015.)

Generative Models



(Oriol, et al. Show and tell: A neural image caption generator. CVPR 2015.)

Going further



Beach



Golf



Train Station



Baby

<https://medium.com/nanoneets/how-we-flew-a-drone-to-monitor-construction-projects-in-africa-using-deep-learning-b792f5c9c471>

<https://www.theverge.com/2016/9/12/12886698/machine-learning-video-image-prediction-mit>

Simple approach: classical methods

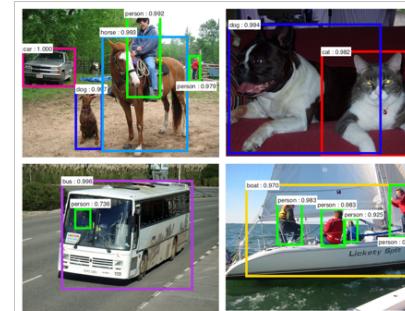
Problem?

- Not using previous frame(s)
- Large computational cost
 - Can't track specific objects accurately

Visual Classification

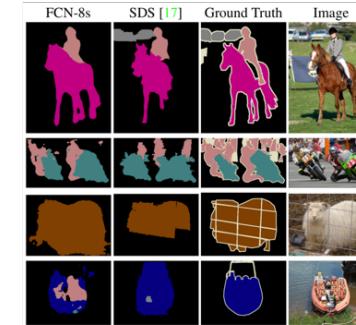


Object Detection



(Ren et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NIPS 2015.)

Semantic Segmentation



(Long et al. Fully convolutional networks for semantic segmentation. CVPR 2015.)

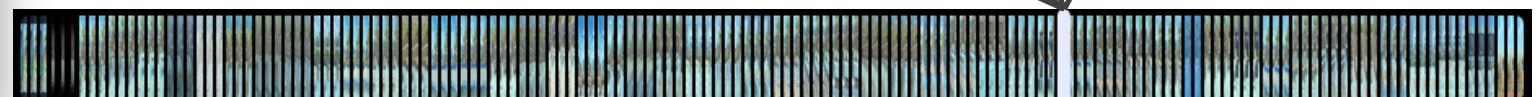
Instance Segmentation



(He et al. Mask R-CNN. ICCV 2017.)



Sequence of image (frames)



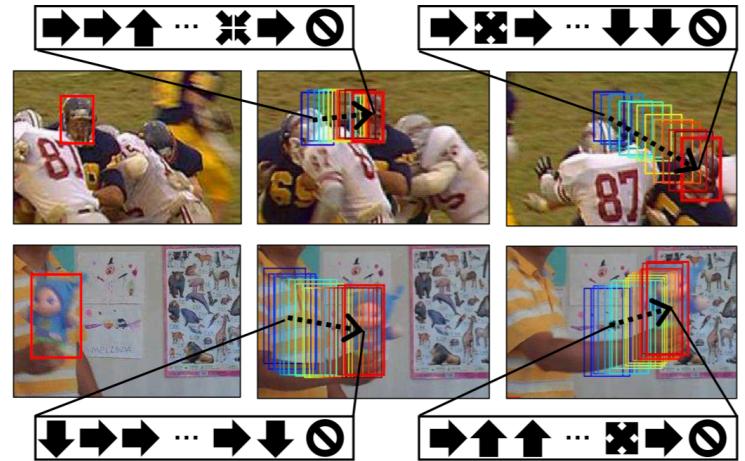
Simple approach: classical methods

Problem?

- Not using previous frame(s)
- Large computational cost
 - Can't track specific objects accurately

The concept of the proposed visual tracking controlled by sequential actions

Yun et al. Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning. CVPR 2017.



Object 1

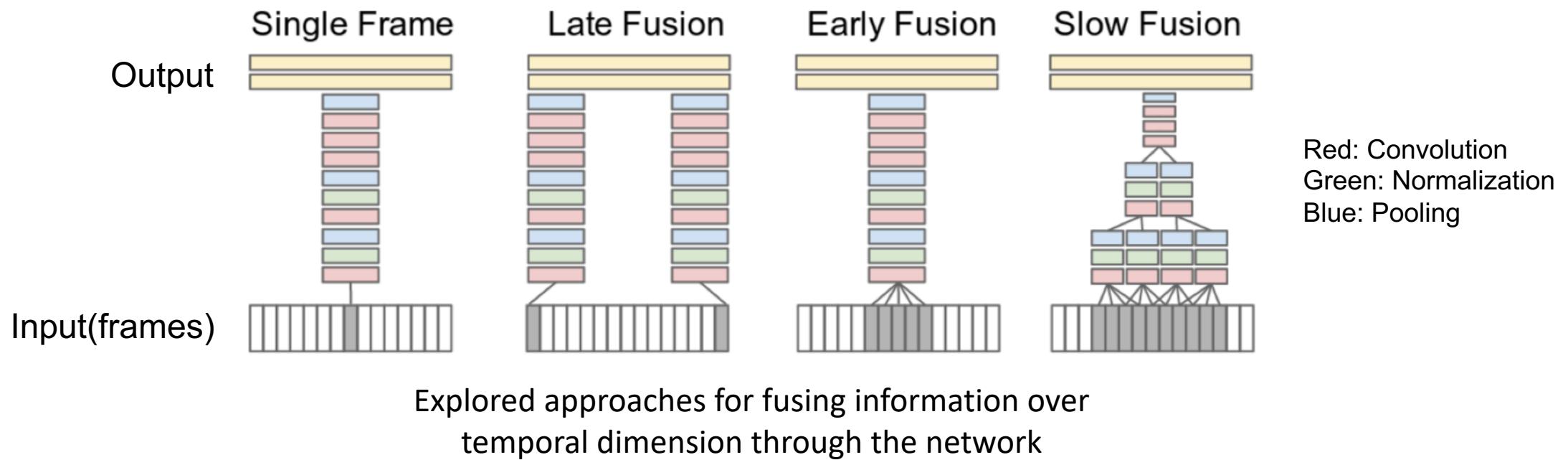
Object 2

Object 3

Video Classification

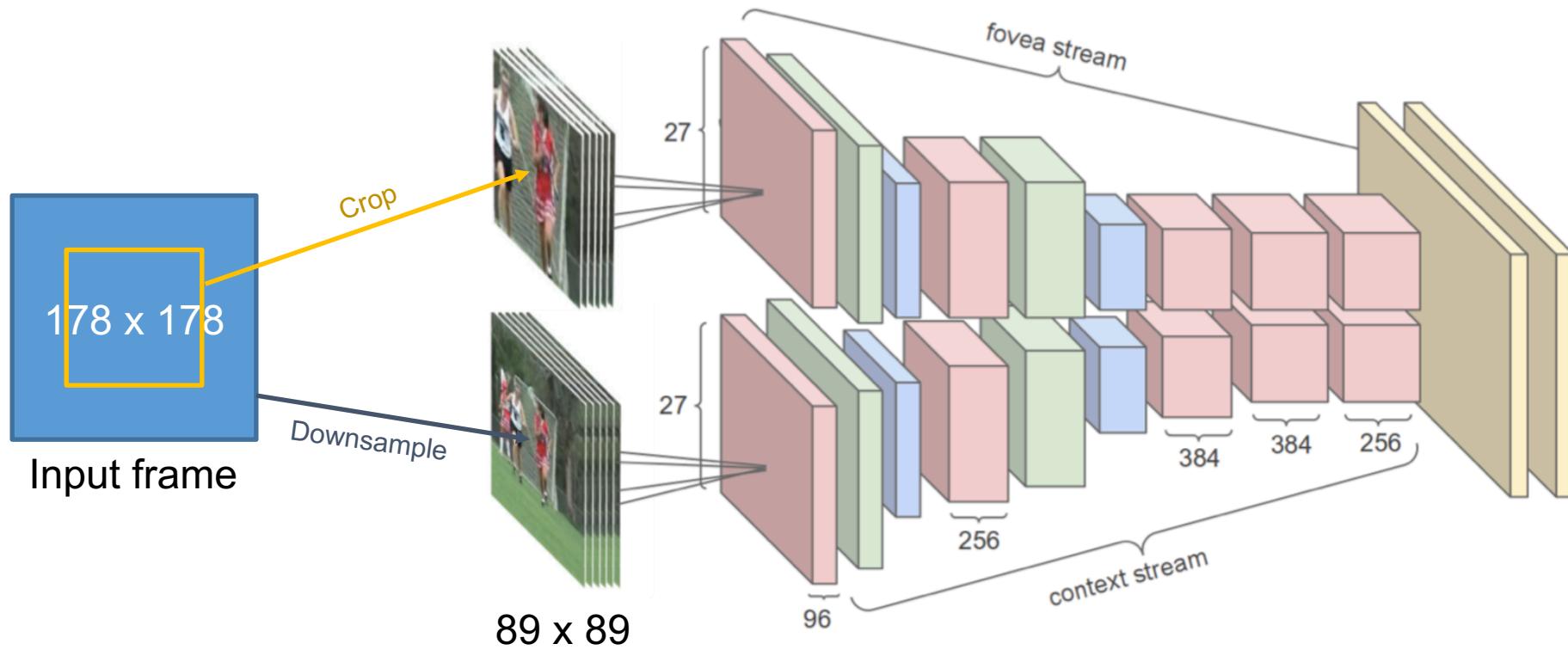
Karpathy et al. Large-scale Video Classification with
Convolutional Neural Networks. CVPR 2014.

Time Information Fusion



Karpathy et al. Large-scale Video Classification with Convolutional Neural Networks. CVPR 2014.

Multiresolution CNNs



Input frames are fed into two separate streams of processing.

Results



track cycling
cycling
track cycling
road bicycle racing
marathon
ultramarathon



ultramarathon
ultramarathon
half marathon
running
marathon
inline speed skating



heptathlon
heptathlon
decathlon
hurdles
pentathlon
sprint (running)



bikejoring
mushing
bikejoring
harness racing
skijoring
carting



longboarding
longboarding
aggressive inline skating
freestyle scootering
freeboard (skateboard)
sandboarding



ultimate (sport)
ultimate (sport)
hurling
flag football
association football
rugby sevens



demolition derby
demolition derby
monster truck
mud bogging
motocross
grand prix motorcycle racing



telemark skiing
snowboarding
telemark skiing
nordic skiing
ski touring
skijoring



whitewater kayaking
whitewater kayaking
rafting
kayaking
canoeing
adventure racing



arena football
indoor american football
arena football
canadian football
american football
women's lacrosse



reining
barrel racing
rodeo
reining
cowboy action shooting
bull riding



eight-ball
nine-ball
blackball (pool)
trick shot
eight-ball
straight pool

Results

Model	Clip Hit@1	Video Hit@1	Video Hit@5
Feature Histograms + Neural Net	-	55.3	-
Single-Frame	41.1	59.3	77.7
Single-Frame + Multires	42.4	60.0	78.5
Single-Frame Fovea Only	30.0	49.9	72.8
Single-Frame Context Only	38.1	56.0	77.2
Early Fusion	38.9	57.7	76.8
Late Fusion	40.7	59.3	78.7
Slow Fusion	41.9	60.9	80.2
CNN Average (Single+Early+Late+Slow)	41.4	63.9	82.4

Karpathy et al. Large-scale Video Classification with Convolutional Neural Networks. CVPR 2014.