TDP1

Case study: CNNs for video analysis

Tim Dunn

Singapore 2019

Slide 1

TDP1 Timothy Dunn, Ph.D., 7/10/2019

Examples we will cover today

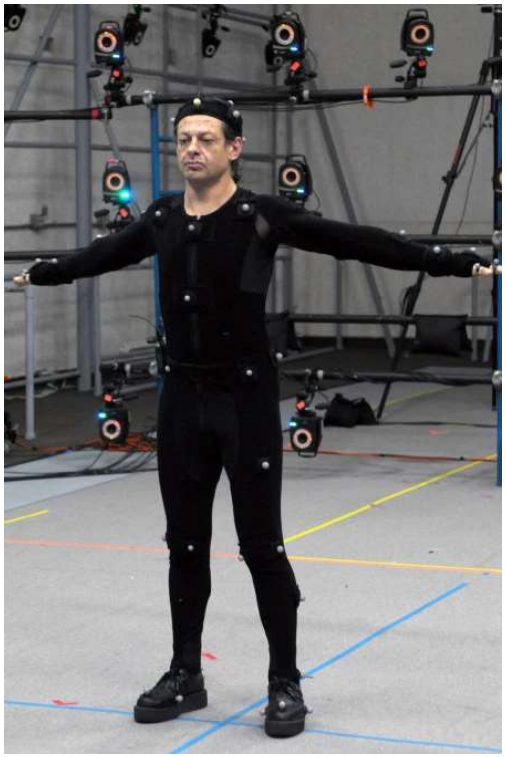
- 2D Human keypoint detection
- 2D Animal keypoint detection
- 3D Human & Animal keypoint detection
- Action recognition
- The Sound of Pixels

2D Human keypoint detection

(or *markerless motion capture*)

2D Human keypoint detection

(or *markerless motion capture*)



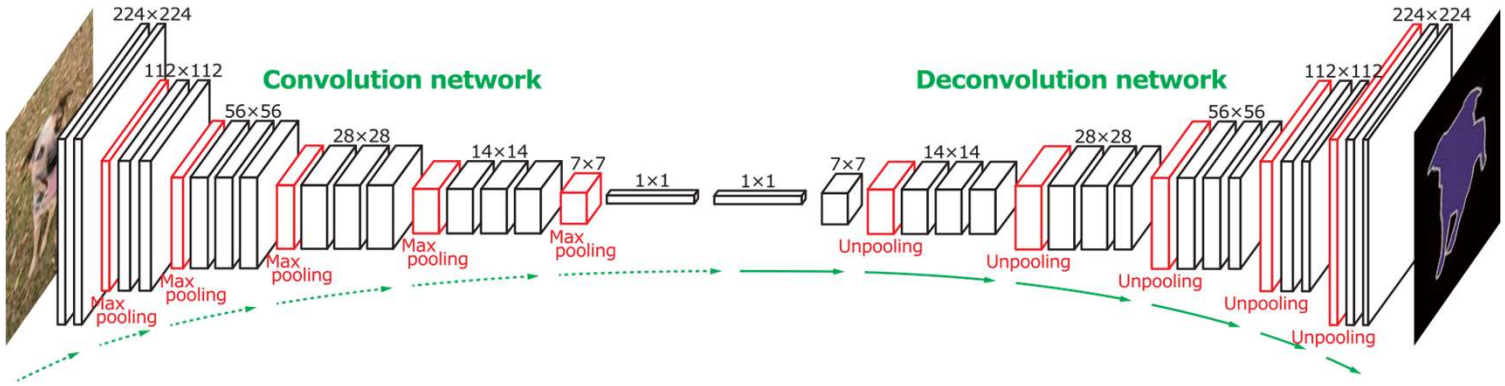
- Motion capture technology measures body movement at multiple positions across the body
- These measurements are typically used by the entertainment industry (video games, movies)
- It requires a special suit (un-natural)
- For 3D, it requires a large camera array for tracking points on the suit (expensive, non-portable).

2D Human keypoint detection

(or *markerless motion capture*)

- Markerless motion capture seeks to provide the same readout of movement but without the special suit and with a normal RGB camera(s)
- As in marker-based motion capture, these techniques track a set of user-defined “keypoints” at specific locations on the body
- Deep learning has revolutionized this technology

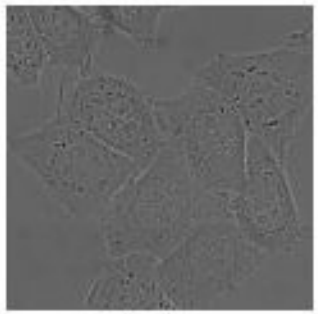
2D Human keypoint detection

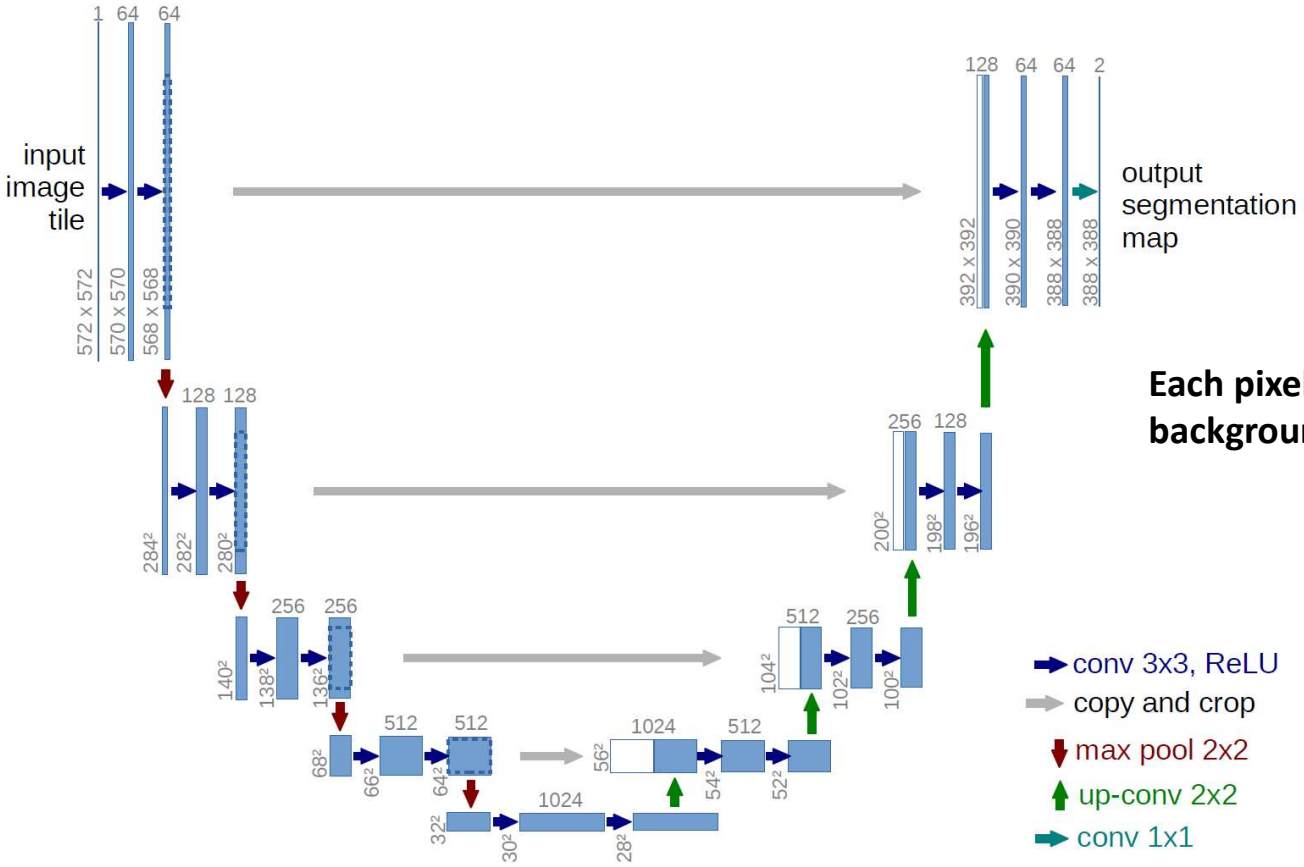


- Like most image segmentation networks, 2D keypoint detection networks are **fully convolutional**.
- This allows for input images of different sizes
- This allows for image classification at the level of individual pixels (knee? foot? hand?)

2D Human keypoint detection

The U-Net

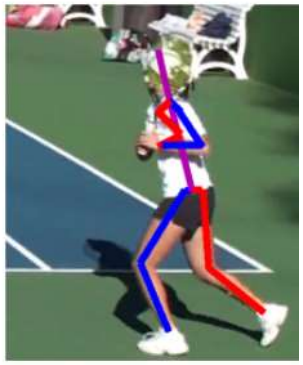


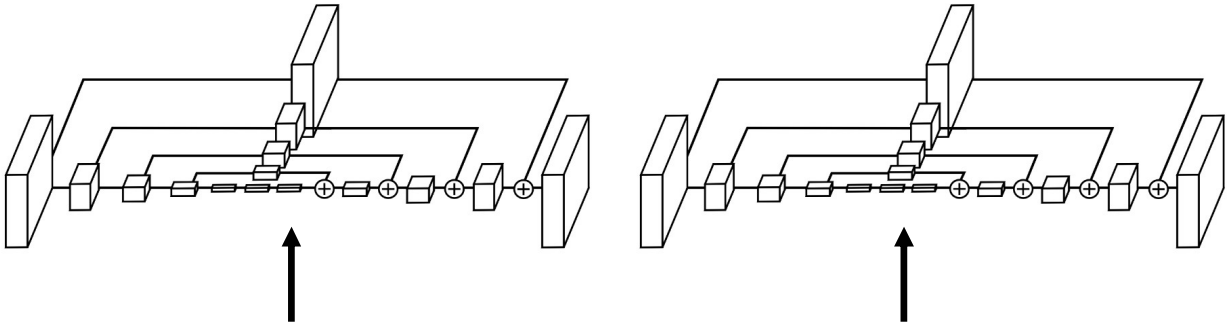


Each pixel is classified as cell (1) or background (0)

2D Human keypoint detection

Stacked Hourglass





U-Net

Ground truth (hand-labeled)
for display only

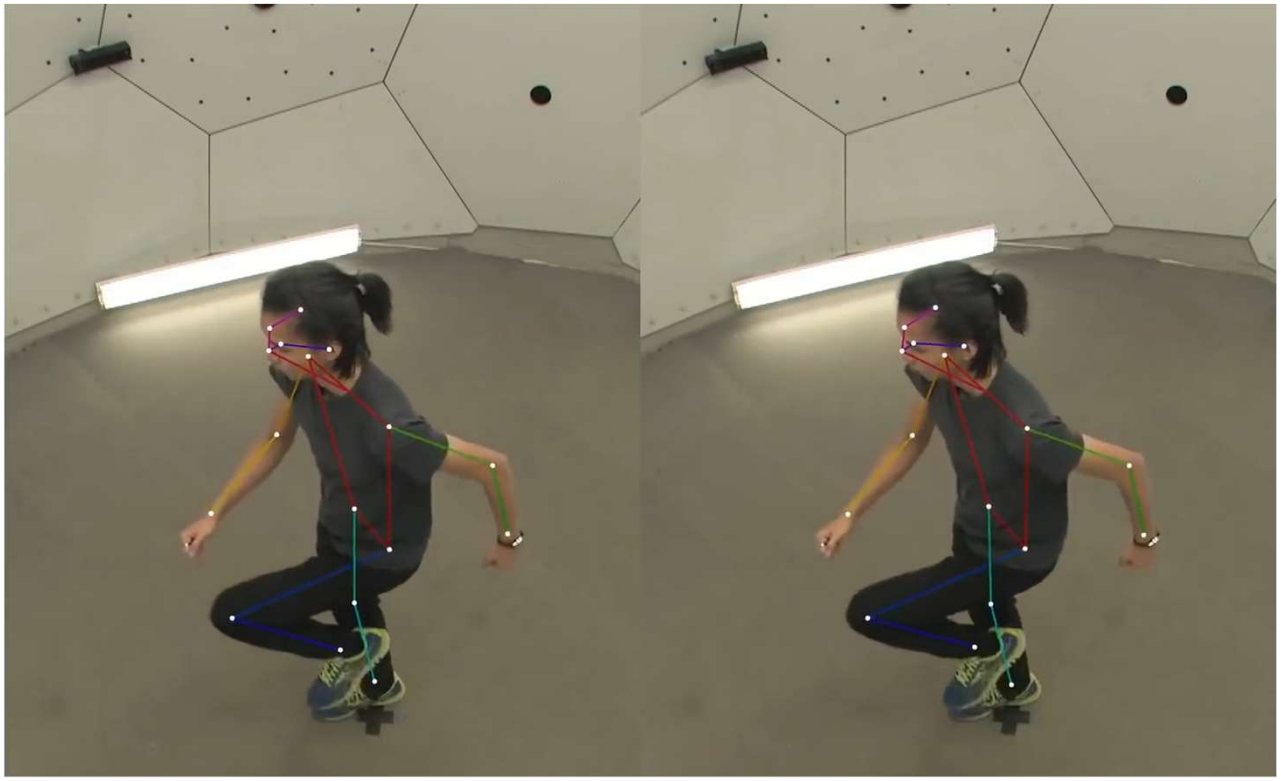
Each pixel is classified as
shoulder, elbow, knee, etc.

Newell et al. (2016)

2D Human keypoint detection

Stacked Hourglass

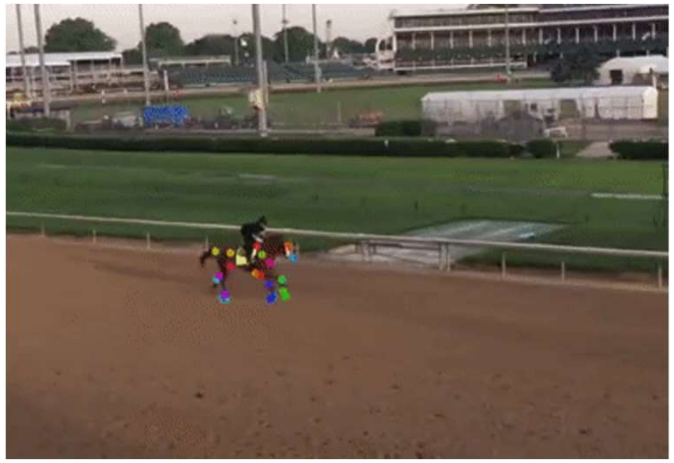
2D Human keypoint detection



2D Animal keypoint detection

2D Animal keypoint detection

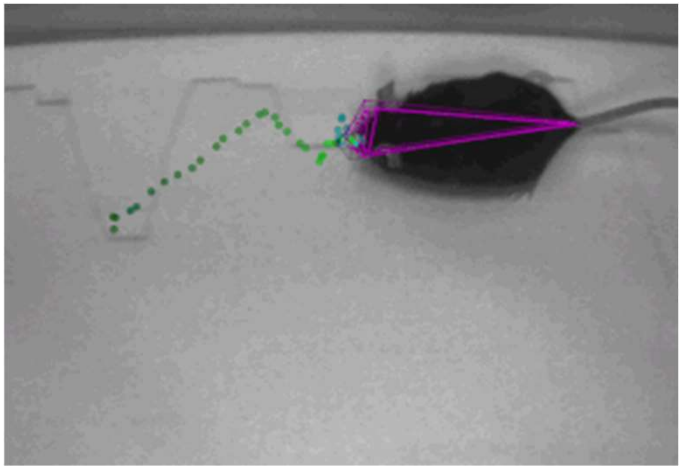
DeepLabCut

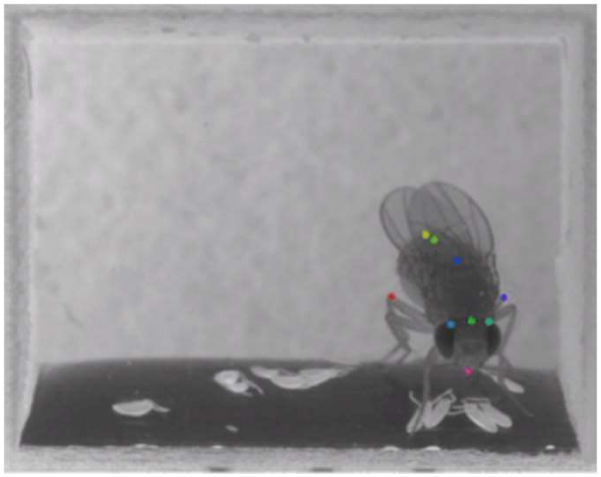


Mathis et al. (2018)

2D Animal keypoint detection

DeepLabCut





2D Animal keypoint detection

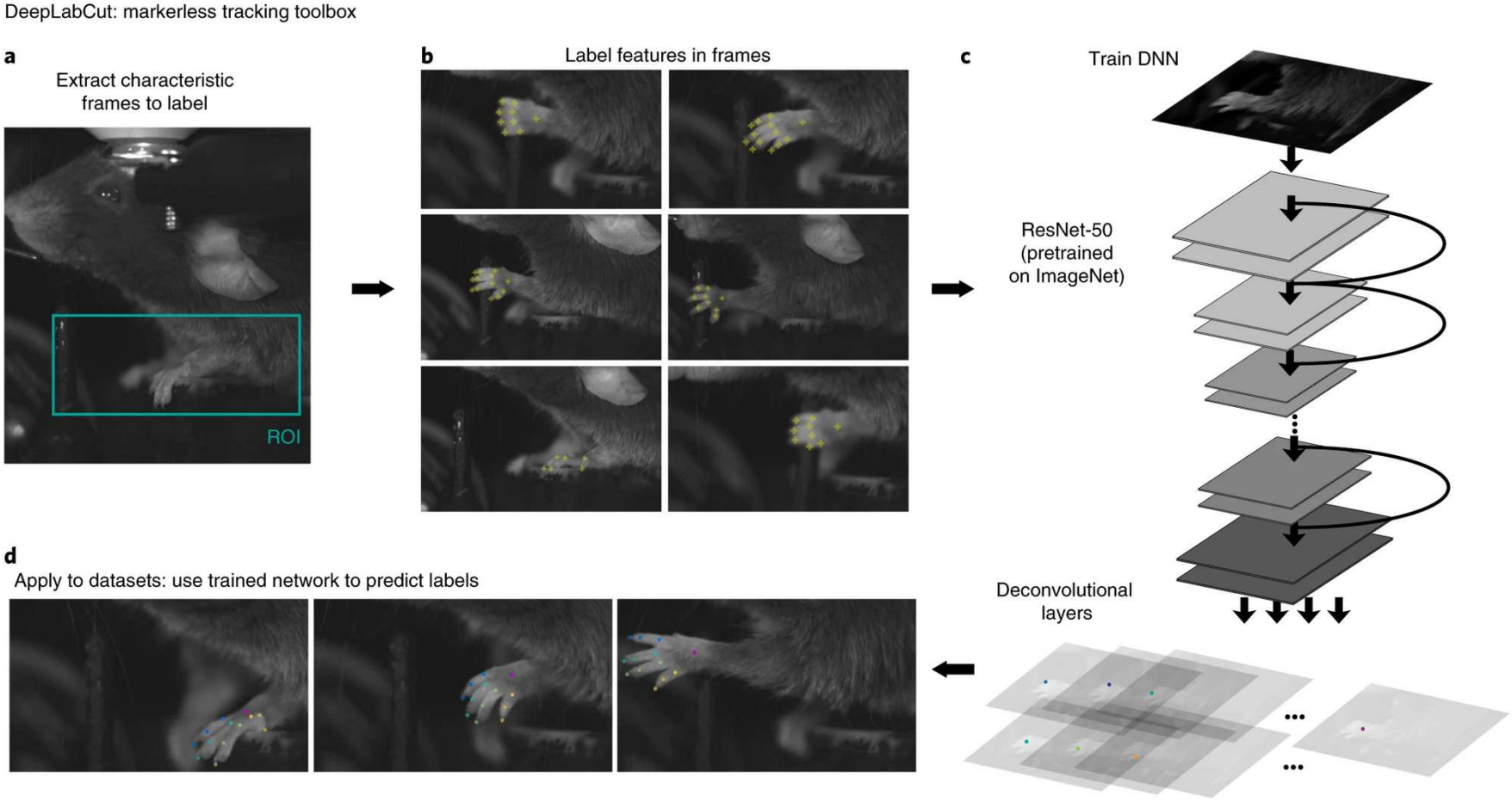
DeepLabCut

- The stacked hourglass network used a public database of over 45K labeled ground truth images.

- No such database exists for lab animals, and labeling is costly and time consuming. DeepLabCut uses **transfer learning** to train a network with <500 labeled images.

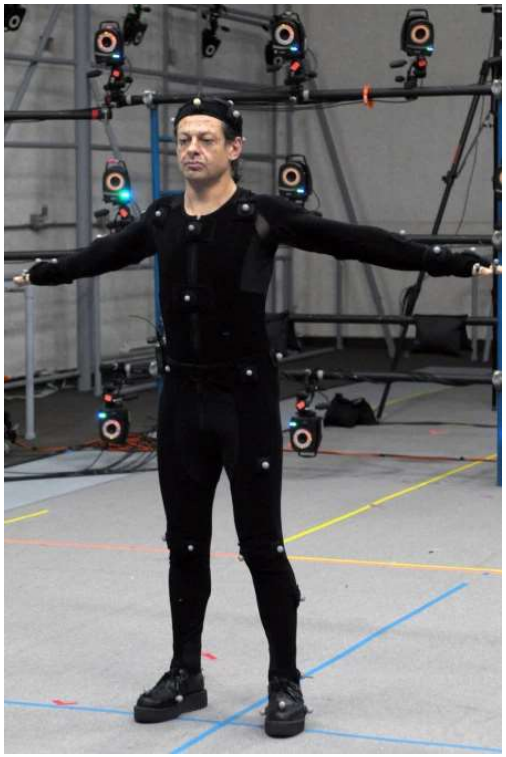
2D Animal keypoint detection

DeepLabCut



3D Human and Animal keypoint detection

3D Human and Animal keypoint detection



- Motion capture technology measures ***3-Dimensional*** body movement at multiple positions across the body
- We cannot completely describe movement without this third dimension.
- Motion capture uses a large camera array for tracking points on the suit *so that the 3D position of each point can be triangulated from 2D*.

3D Human and Animal keypoint detection

Utility in biology and medicine

- Precise quantification of body movements will allow us to understand the brain's control of movement

- Precise quantification of body movements will allow us to build a better understanding of devastating movement disorders (e.g. Parkinson's disease) and their treatments

3D Human and Animal keypoint detection

Utility in biology and medicine

Parkinson's disease causes a range of different, devastating movement defects

- Tremor

Shuffling gait

3D Human and Animal keypoint detection

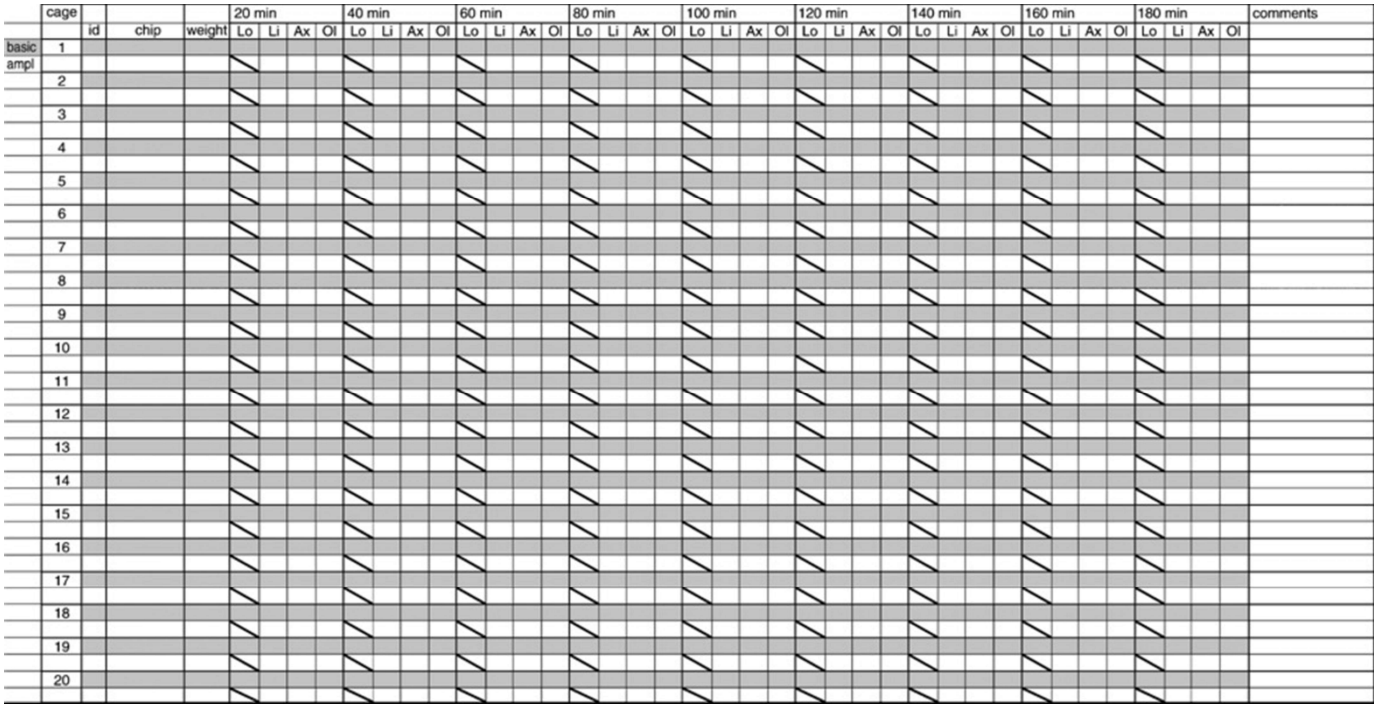
Utility in biology and medicine

- L-DOPA induced dyskinesia

3D Human and Animal keypoint detection

Utility in biology and medicine

Laborious manual annotation / what are we missing?



	cage				20 min				40 min				60 min				80 min				100 min				120 min				140 min				160 min				180 min				comments
		id	chip	weight	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	Lo	Li	Ax	Ol	
basic	1																																								
ampl																																								
	2																																							
	3																																							
	4																																							
	5																																							
	6																																							
	7																																							
	8																																							
	9																																							
	10																																							
	11																																							
	12																																							
	13																																							
	14																																							
	15																																							
	16																																							
	17																																							
	18																																							
	19																																							
	20																																							

Cenci & Lundblad (2007)

3D Human and Animal keypoint detection

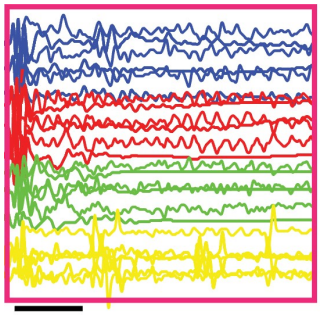
Utility in biology and medicine

Screening of the future?



knee velocity

foot velocity



250 ms

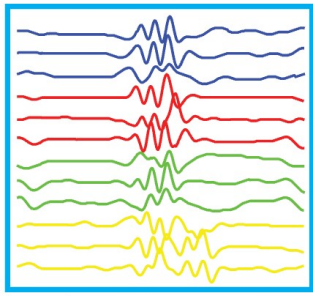
Treatment I

PD Type I



knee velocity

foot velocity



200 ms

PD Type II

Treatment II

3D Human and Animal keypoint detection

DANNCE: 3-D Aligned Neural Network for Computational Ethology

Goal: Predicted the 3D coordinates of body keypoints from normal RGB videos taken from multiple viewpoints

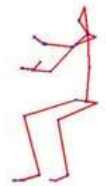
Dunn, Marshall* et al. (Submitted)*

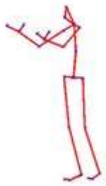
3D Human and Animal keypoint detection

3D training data - Humans

S9 sitingdown Fr 146

S11 waiting Fr 1629

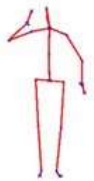




Ground truth 3D skeleton (from motion capture)

S9 greeting Fr 146

S11 walkdog Fr 207

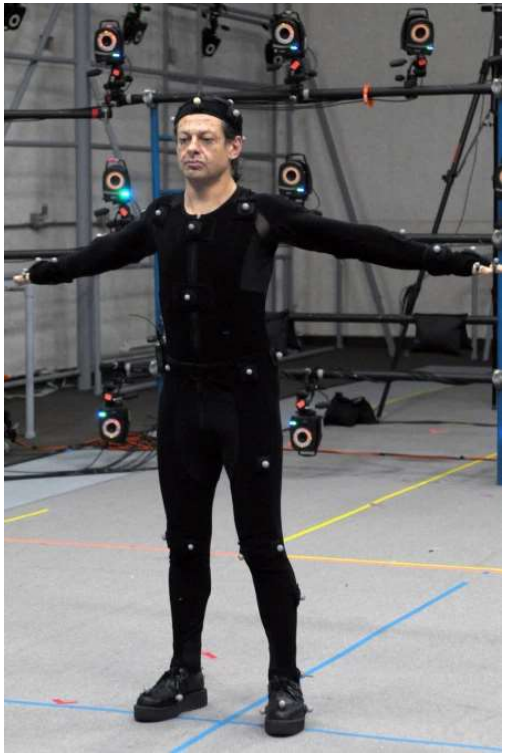


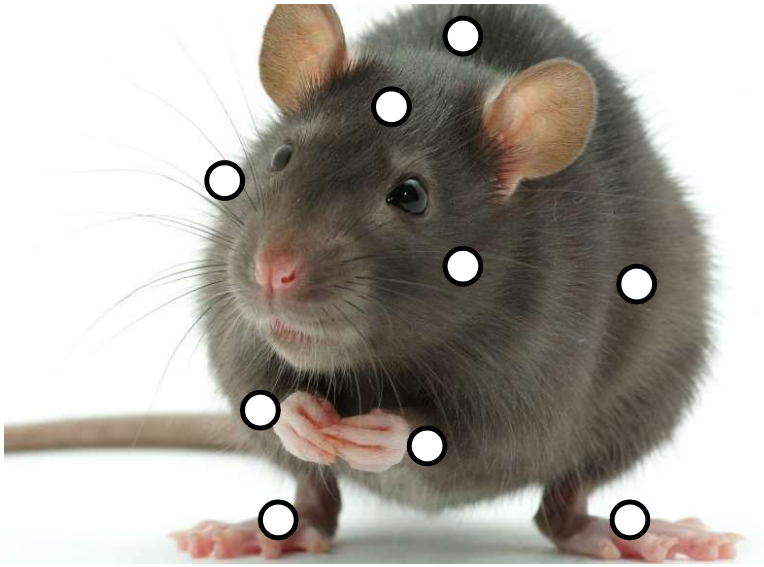
Ground truth 3D skeleton (from motion capture)

dannce

3D Human and Animal keypoint detection

3D training data - Rodents

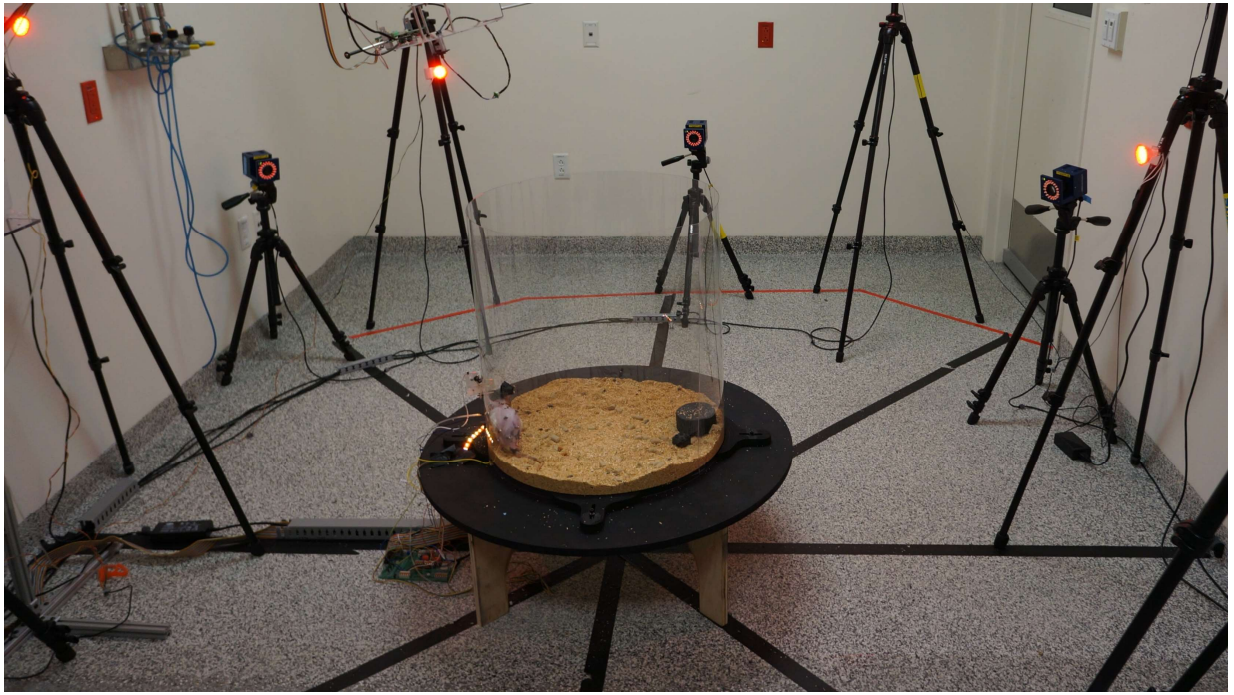




dannce

3D Human and Animal keypoint detection

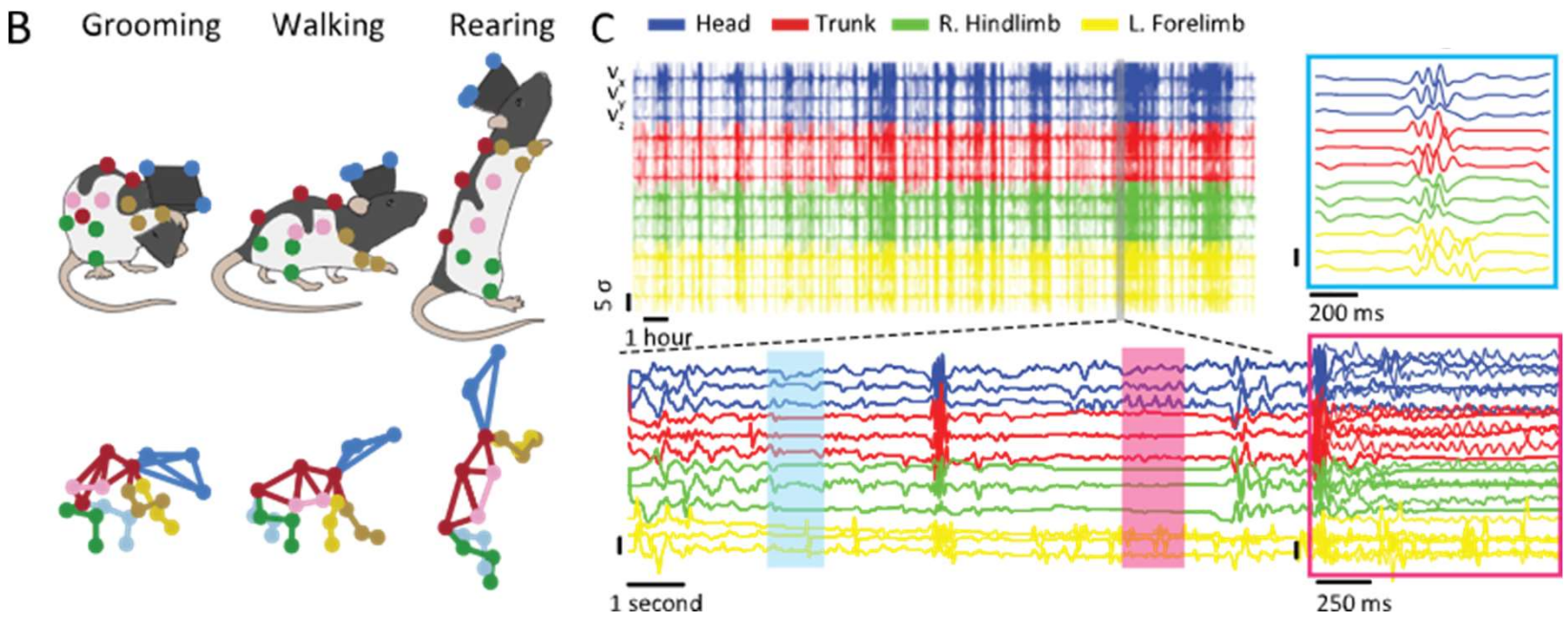
3D training data - Rodents



dannce

3D Human and Animal keypoint detection

3D training data - Rodents

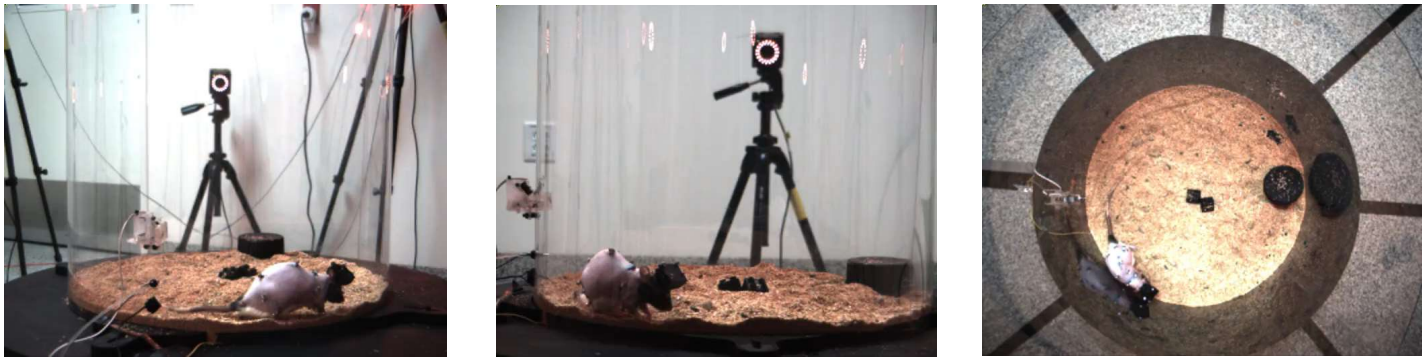


dannce

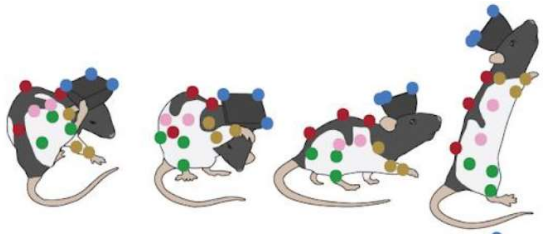
3D Human and Animal keypoint detection

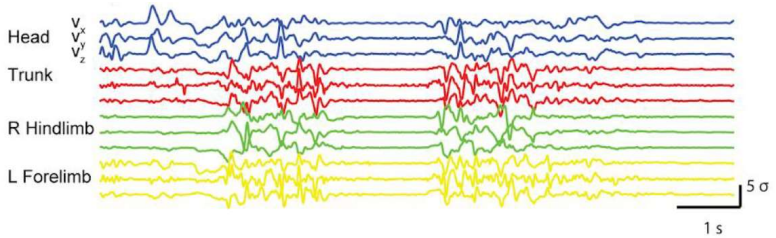
3D training data - Rodents

Camera 1 Camera 2 Camera 3



CNN

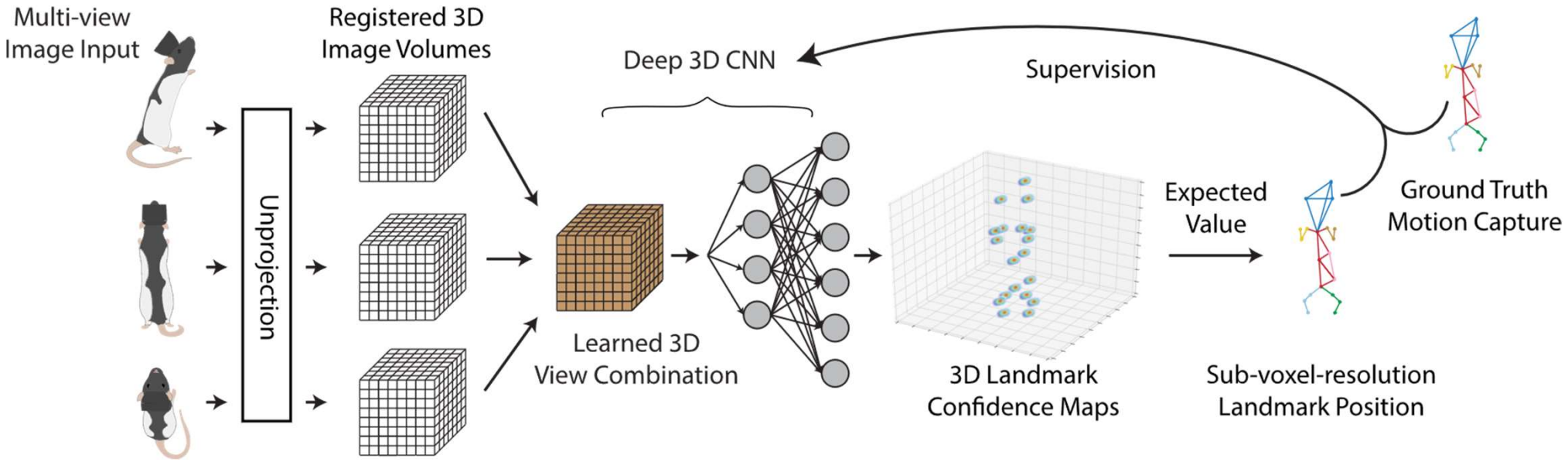




dannce

3D Human and Animal keypoint detection

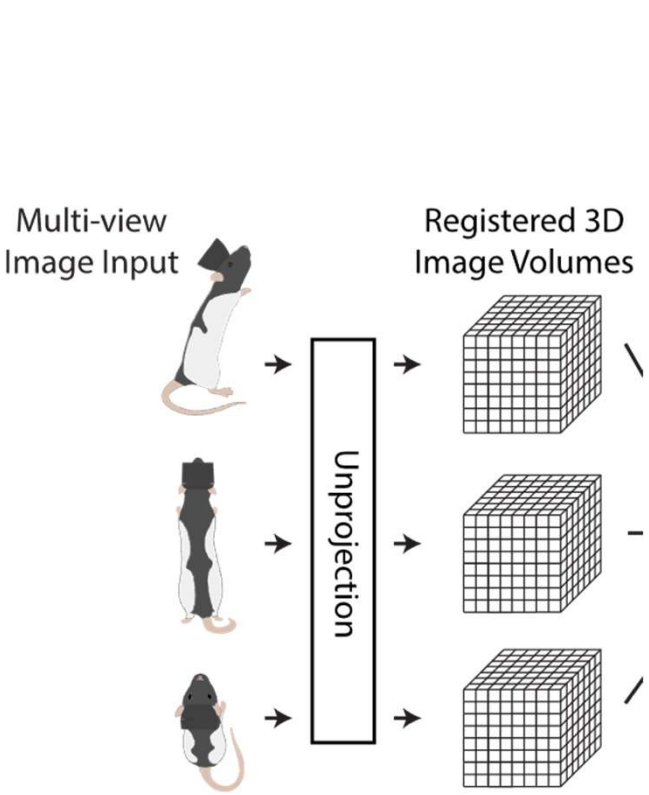
Full pipeline

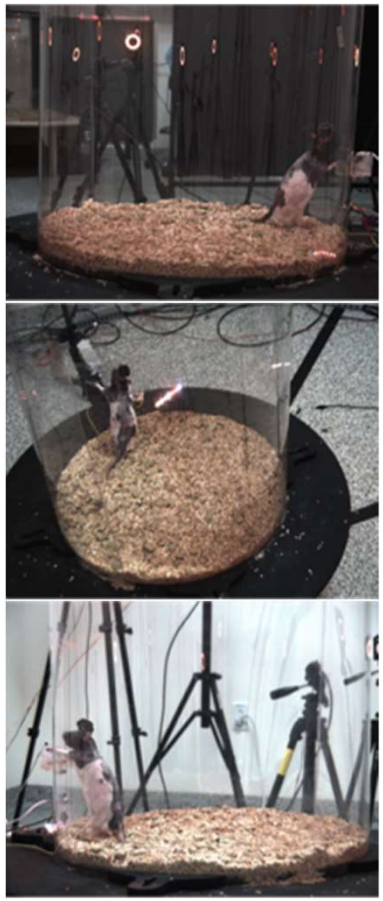


dannce

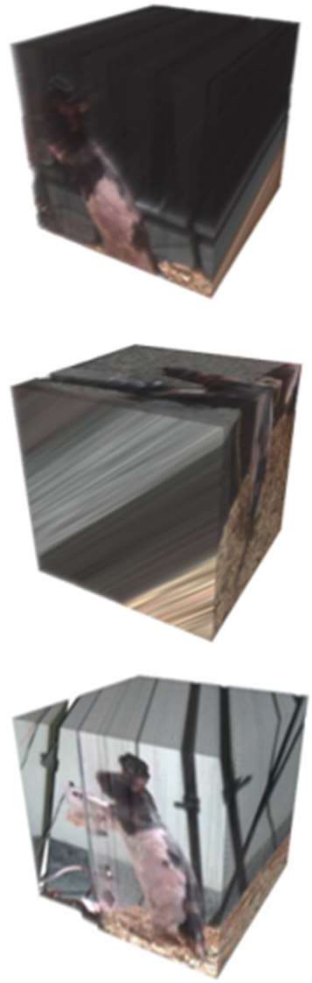
3D Human and Animal keypoint detection

Turning 2D into 3D



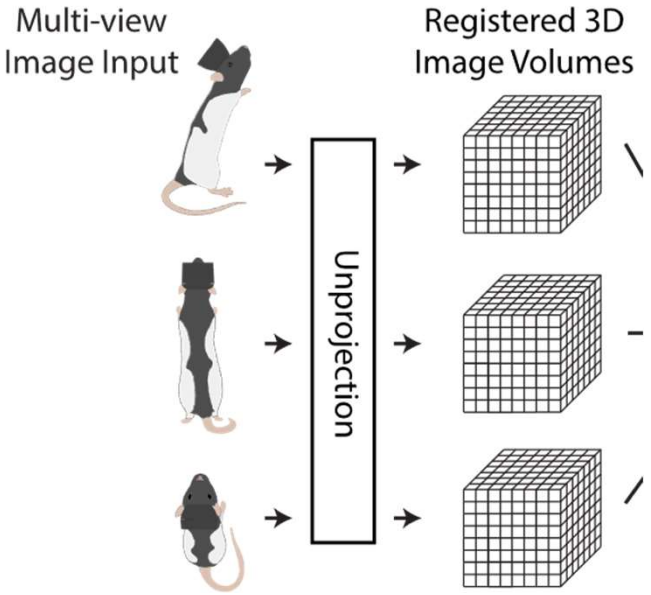


$$p_{2D} = K[R|t]p_{3D}$$



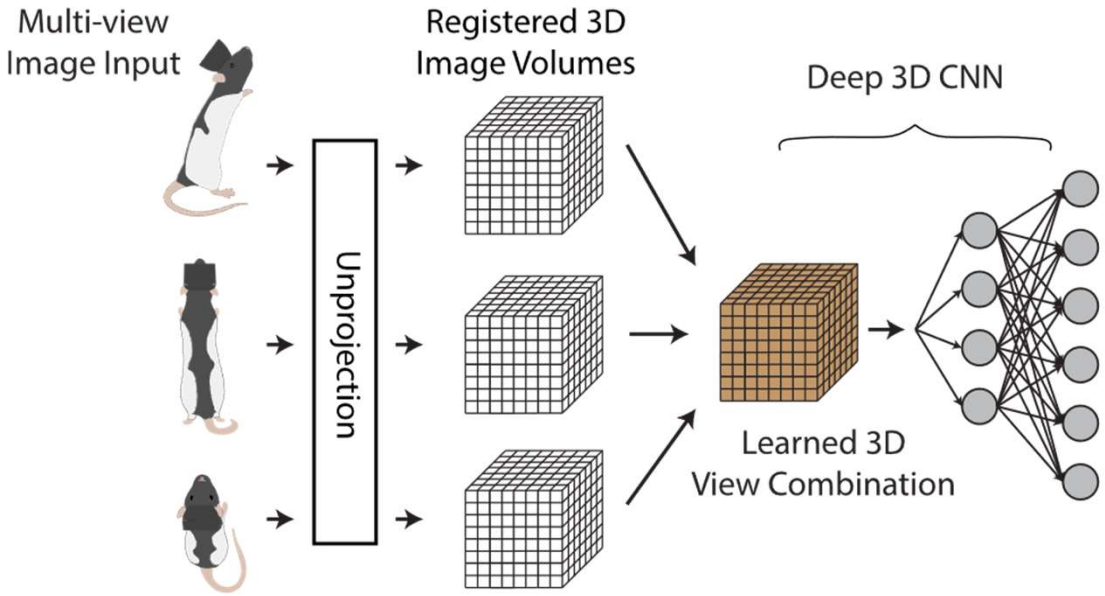
3D Human and Animal keypoint detection

Turning 2D into 3D

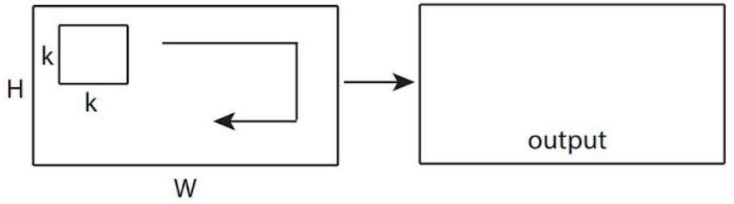


3D Human and Animal keypoint detection

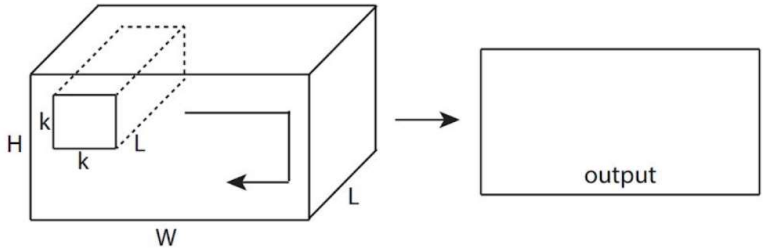
Output: 3D probability distributions for each keypoint



2D Convolution

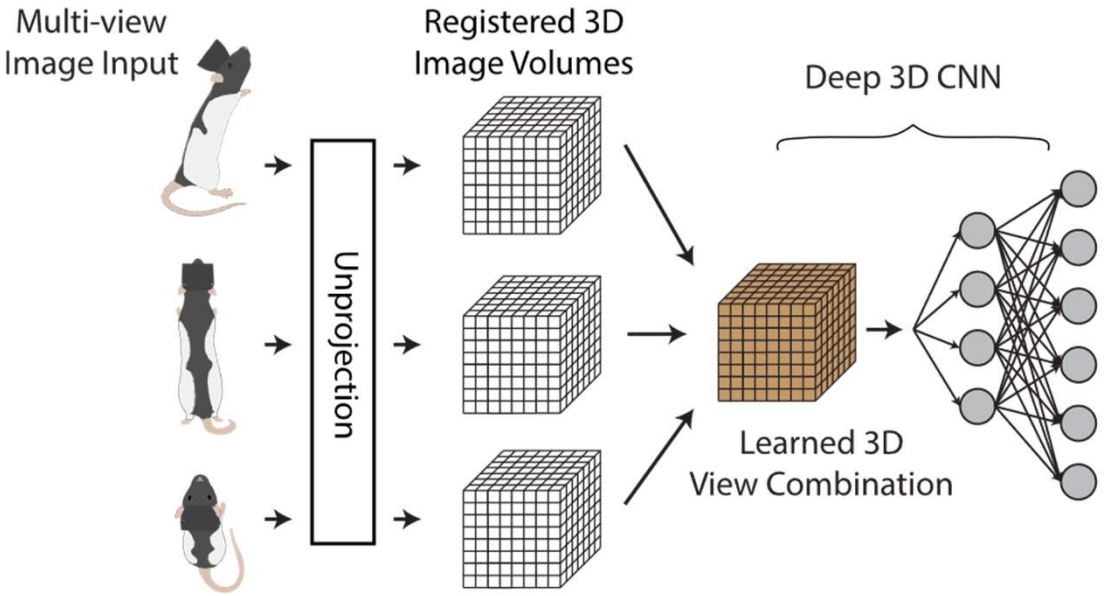


2D Convolution (volumetric input)

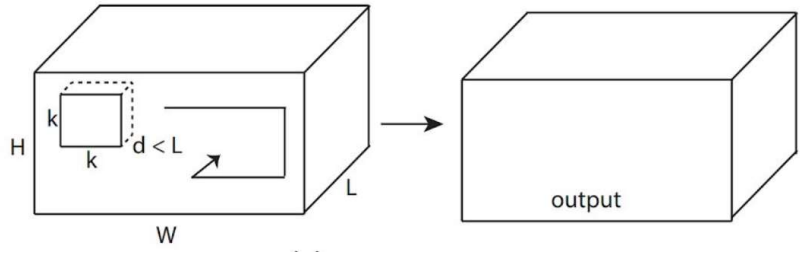


3D Human and Animal keypoint detection

Output: 3D probability distributions for each keypoint

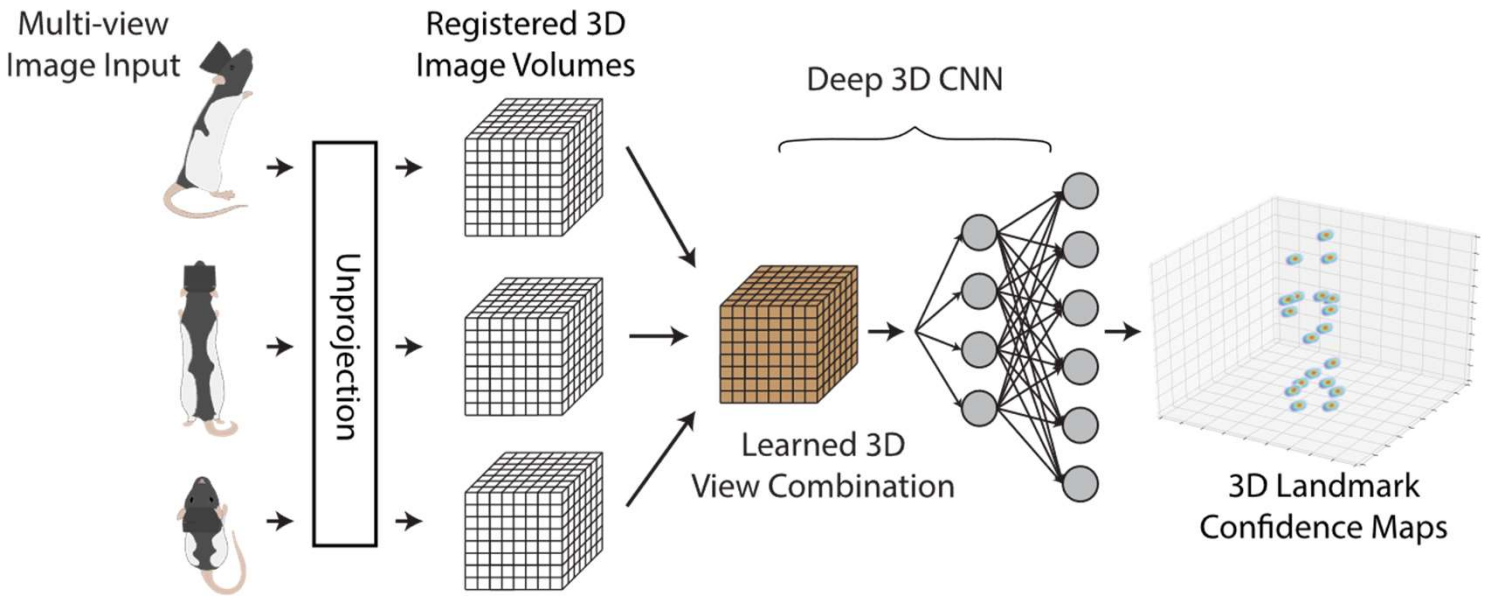


3D Convolution



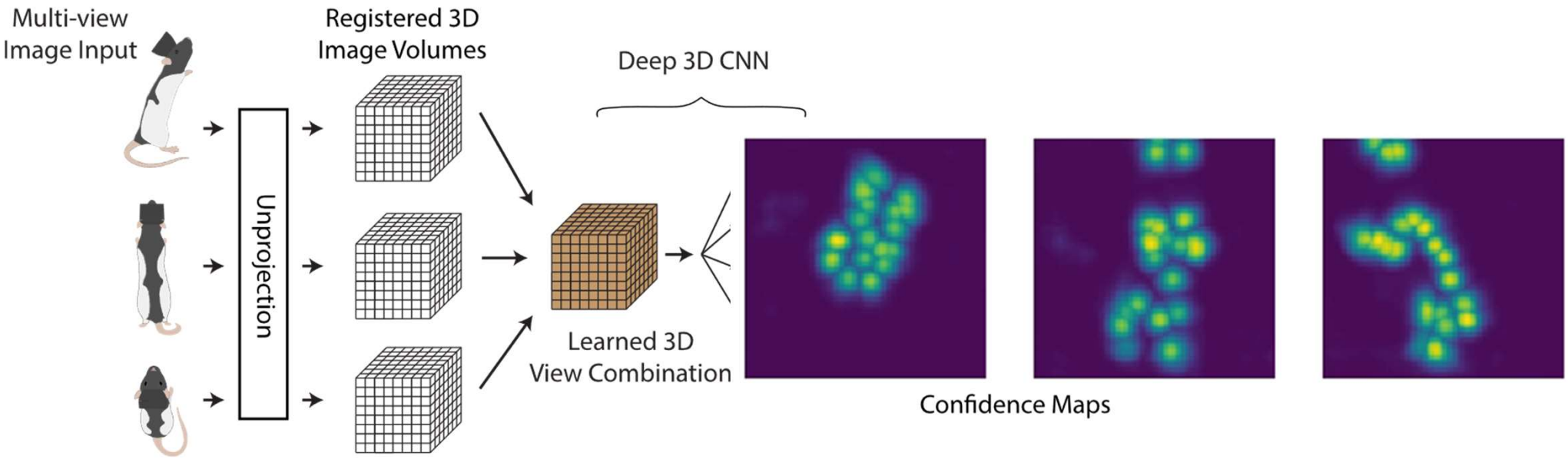
3D Human and Animal keypoint detection

Output: 3D probability distributions for each keypoint



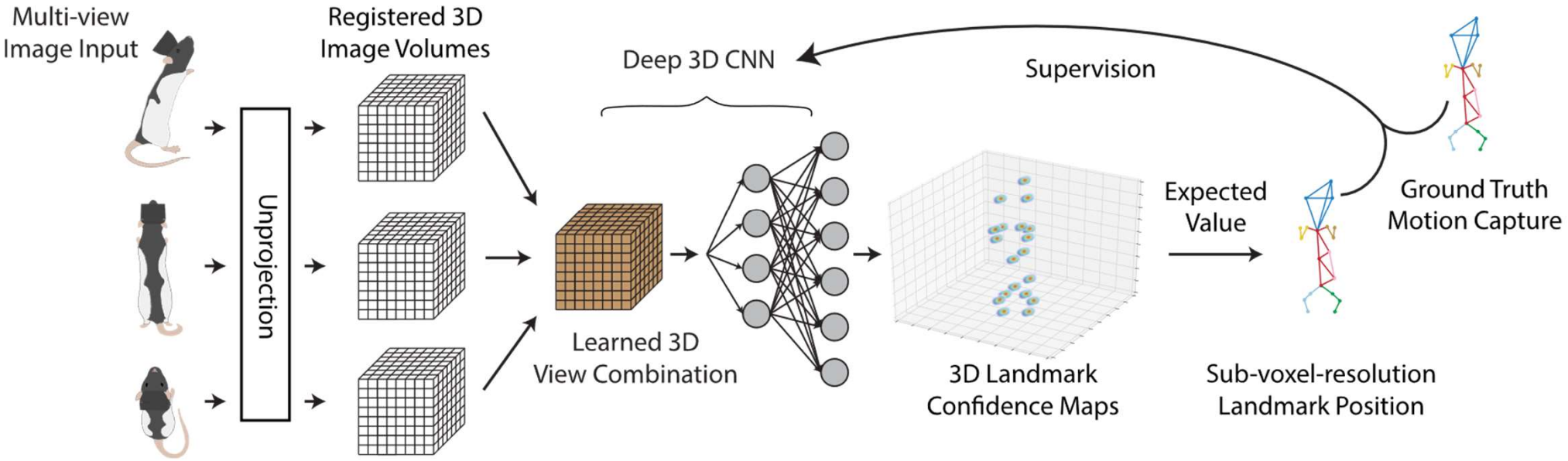
3D Human and Animal keypoint detection

Output: 3D probability distributions for each keypoint



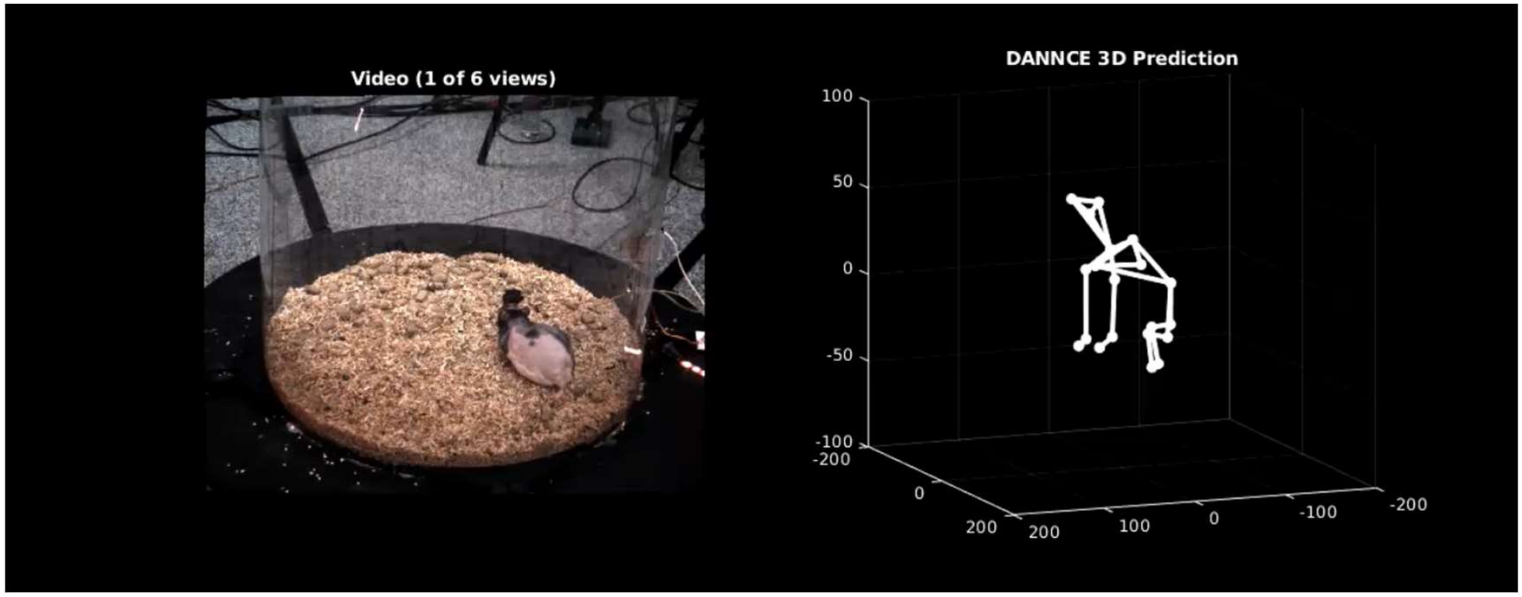
3D Human and Animal keypoint detection

Output: spatial average of 3D probability distributions for each keypoint



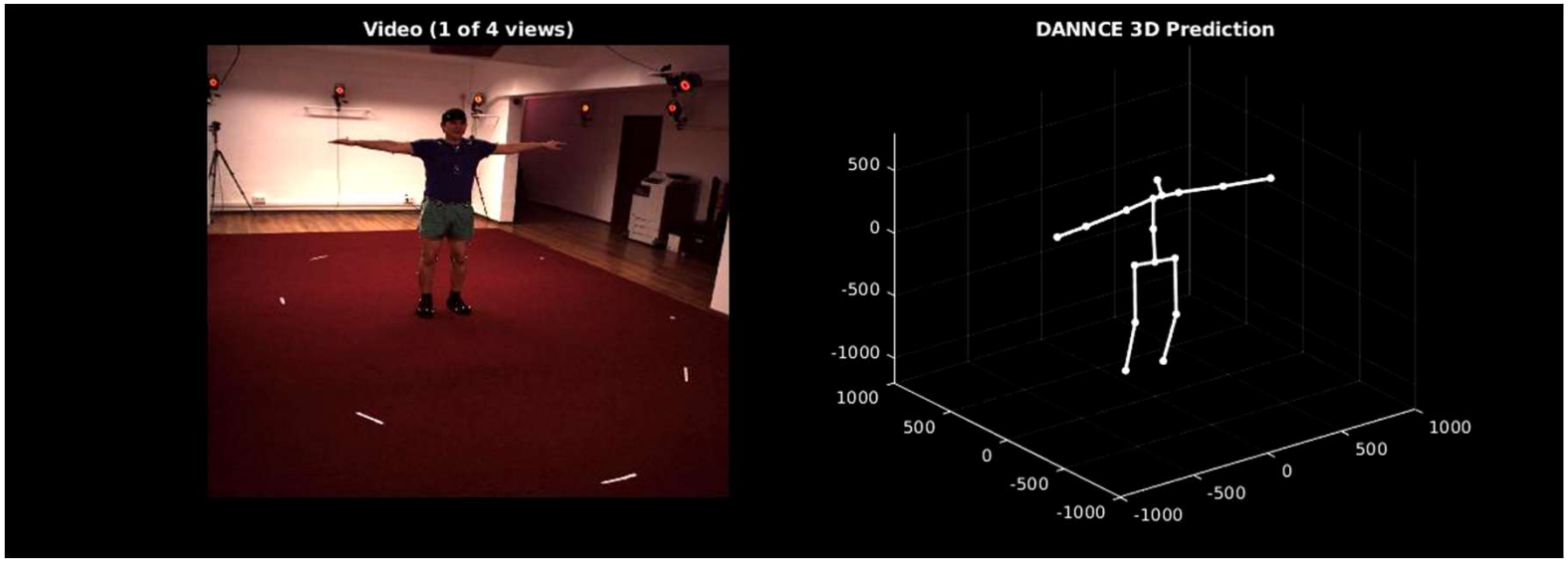
3D Human and Animal keypoint detection

Performance on rodents



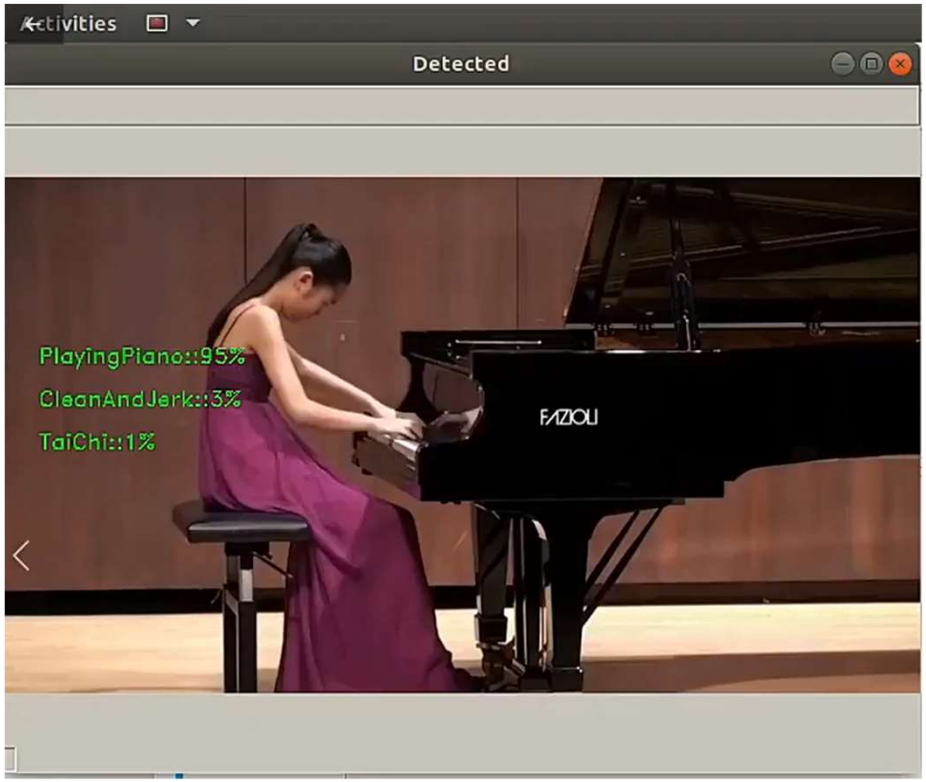
3D Human and Animal keypoint detection

Performance on humans



Action Recognition

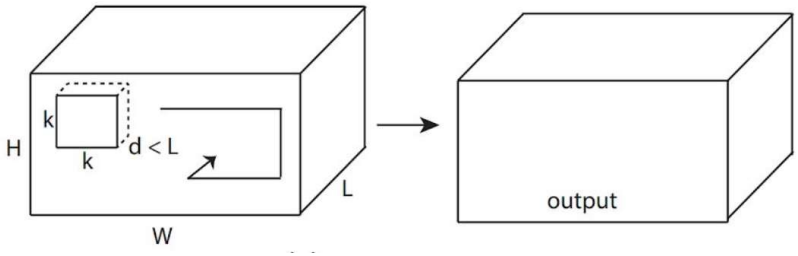
Action Recognition



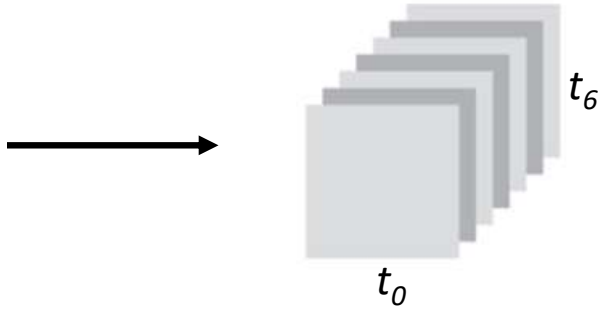
Action Recognition

3D CNN for action recognition

Recall the 3D Convolution: previously, the third dimension was space

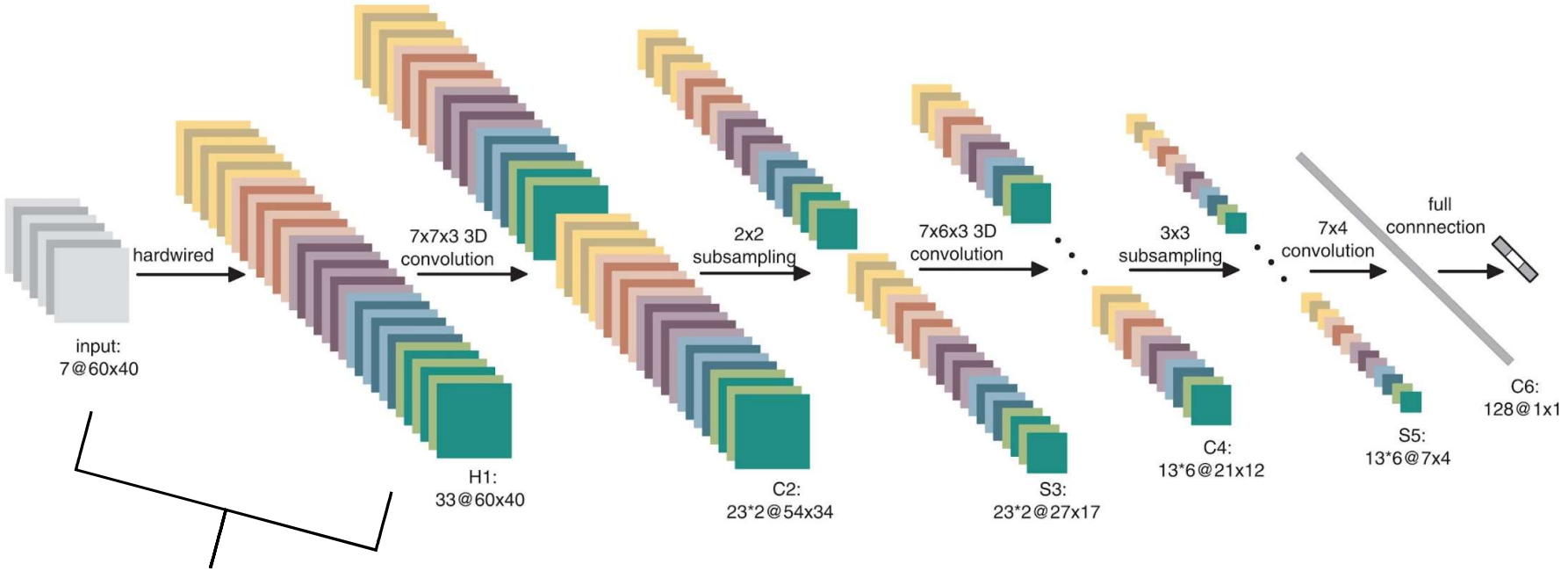


But we can also create a 3D volume using *time* as the third dimension



Action Recognition

3D CNN for action recognition

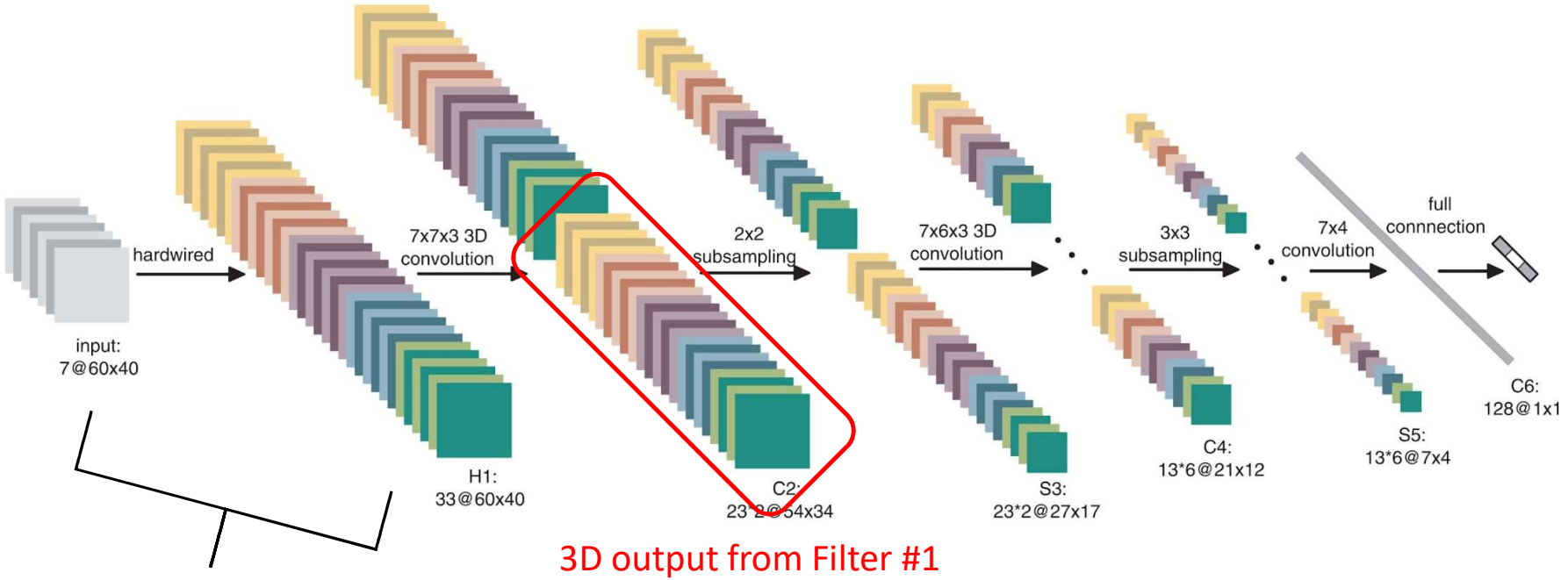


Preprocessing: add spatial gradients and “optic flow” for each frame

Ji et al. (2013)

Action Recognition

3D CNN for action recognition



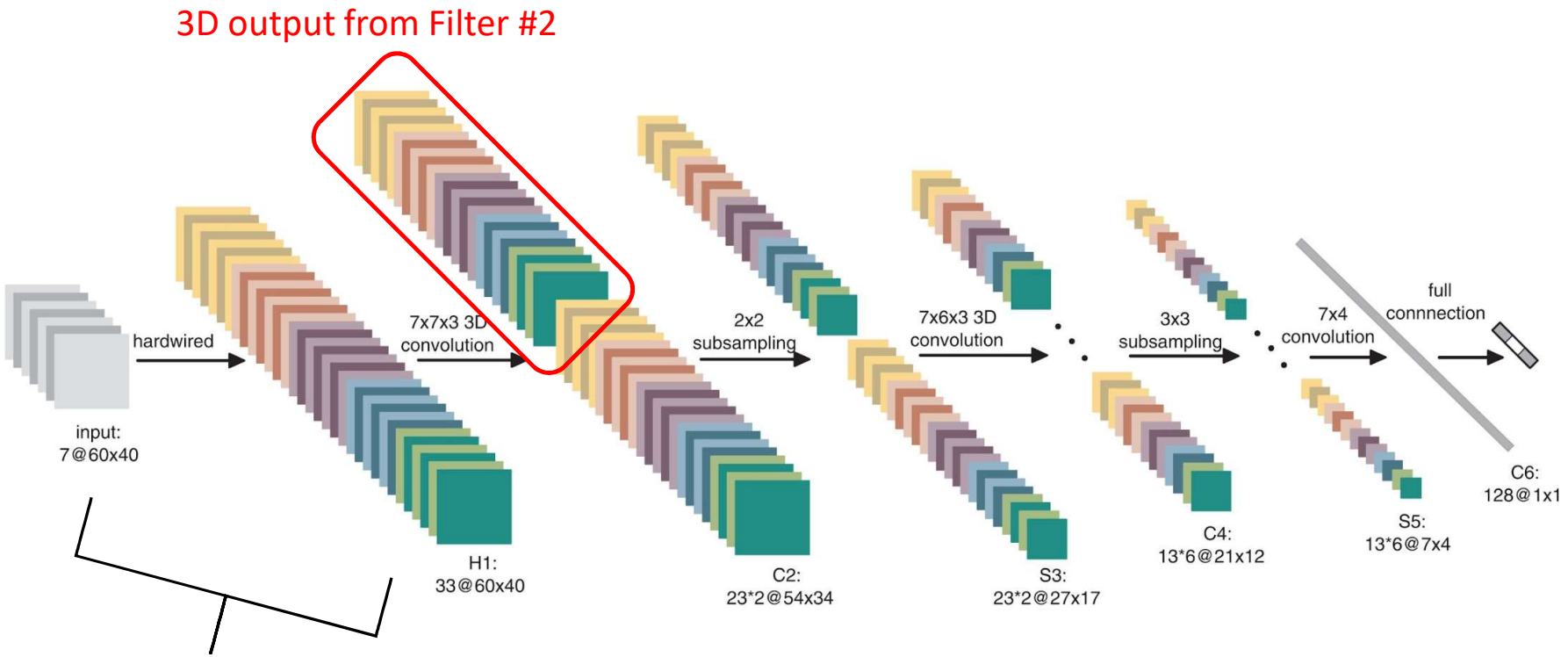
3D output from Filter #1

Preprocessing: add spatial gradients and “optic flow” for each frame

Ji et al. (2013)

Action Recognition

3D CNN for action recognition



Preprocessing: add spatial gradients and “optic flow” for each frame

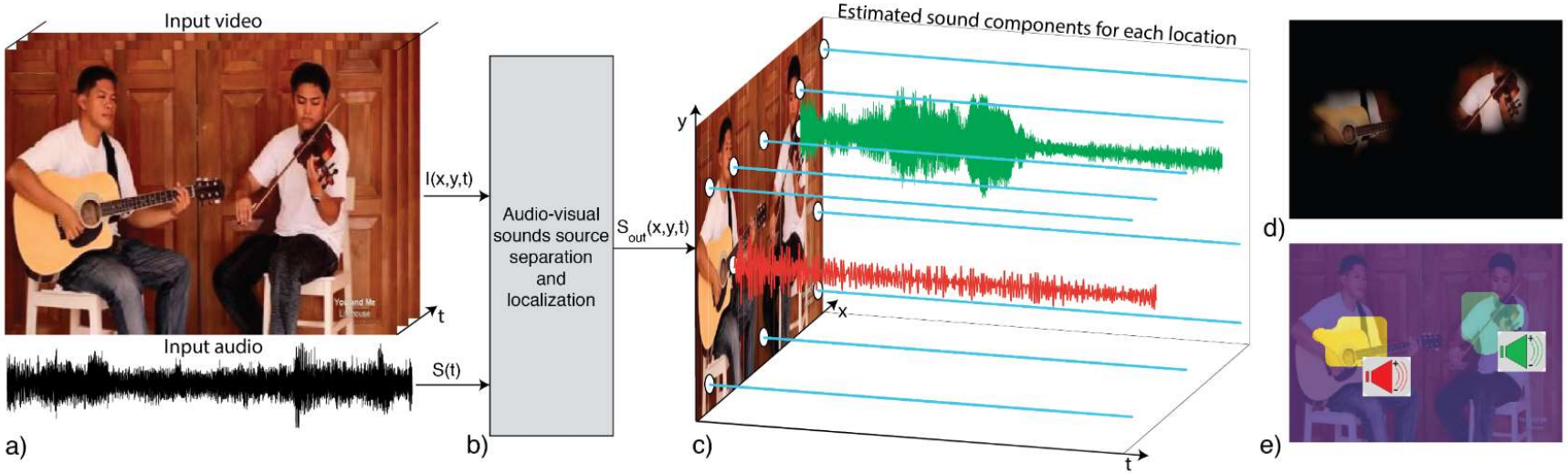
Ji et al. (2013)

The Sound of Pixels

The Sound of Pixels

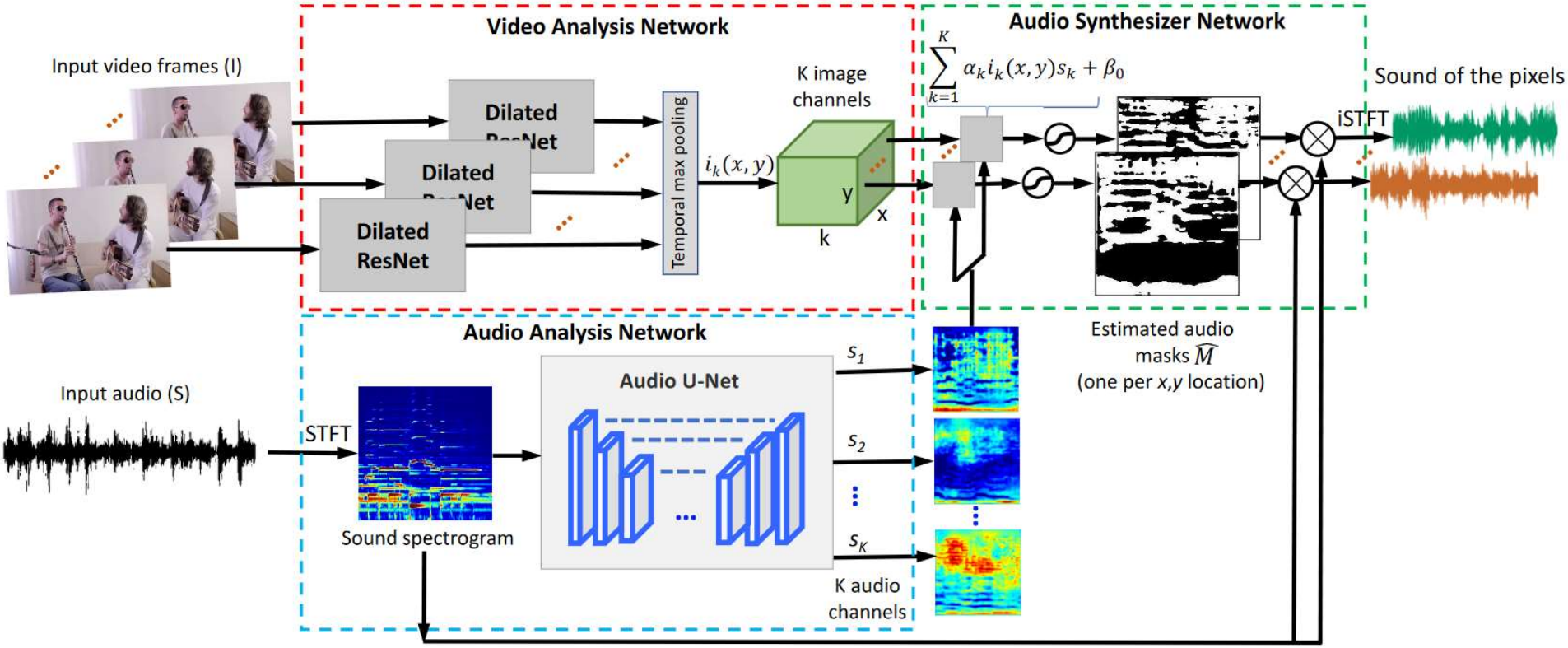
Zhao et al. (2018)

The Sound of Pixels



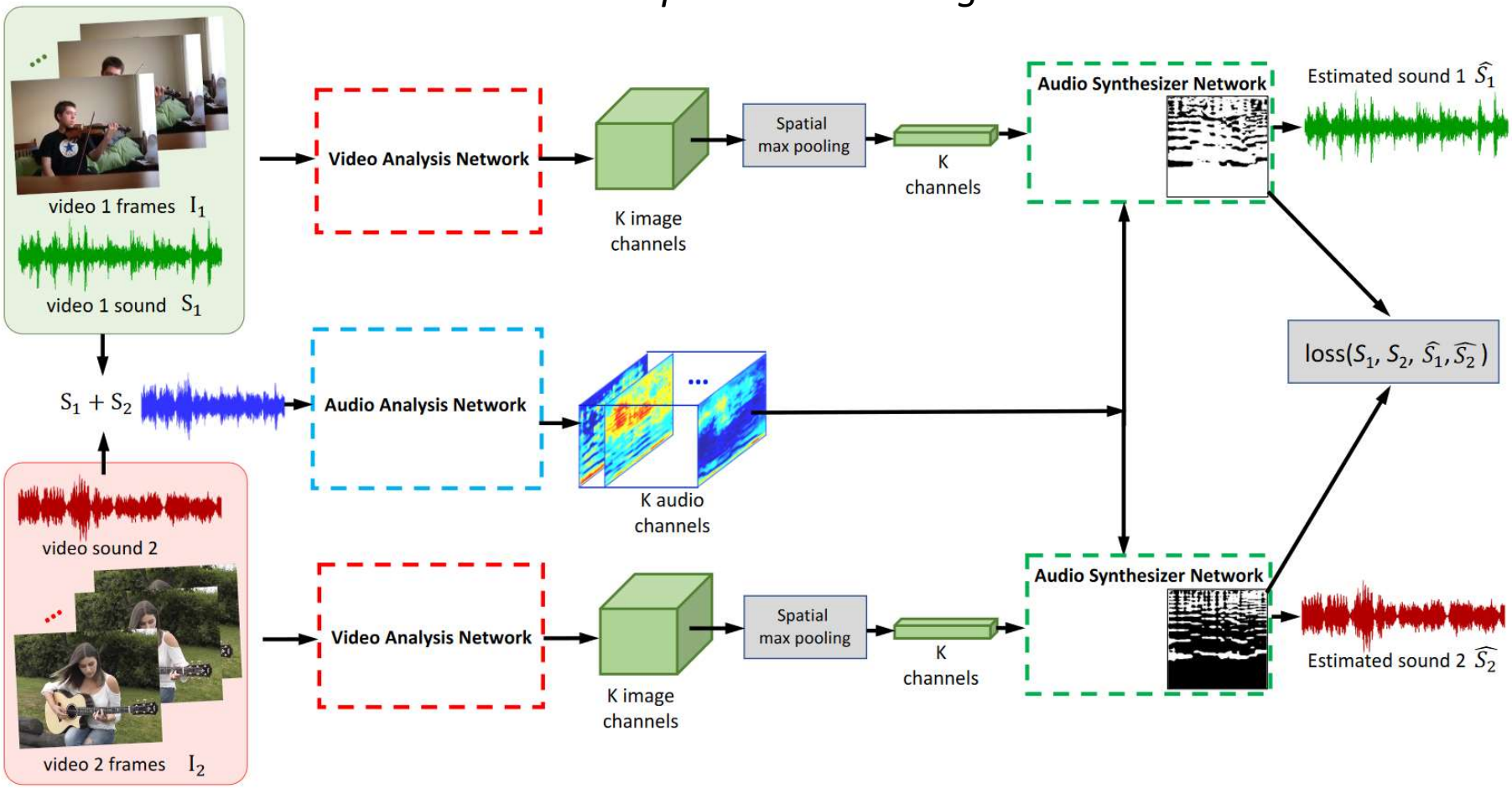
Zhao et al. (2018)

The Sound of Pixels



The Sound of Pixels

Unsupervised Learning



The Sound of Pixels

Online Demo