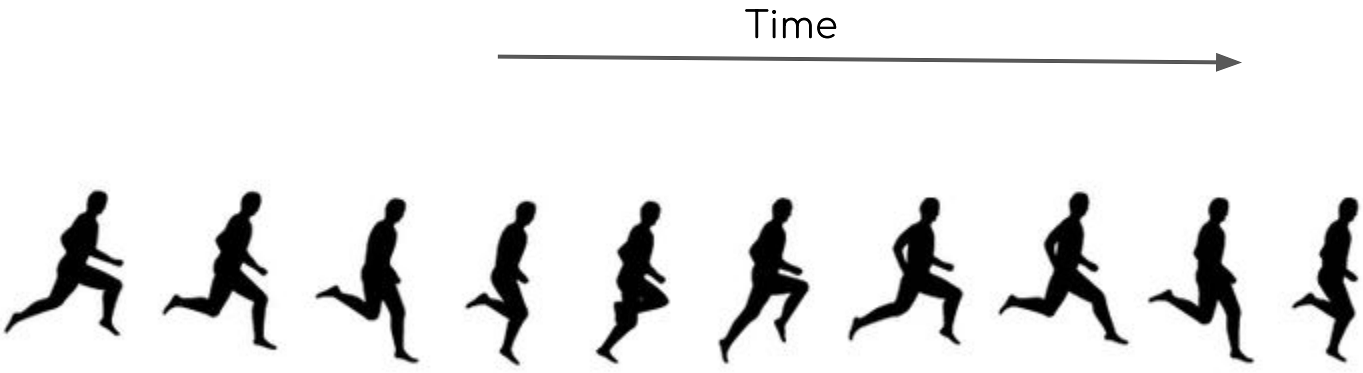
Video Classification

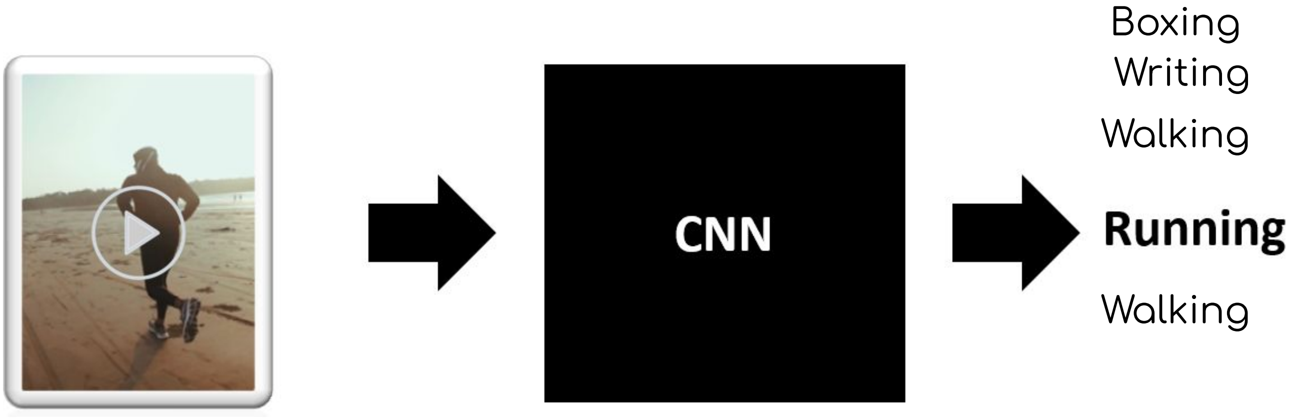
Dr. Konda Reddy Mopuri
Deep Learning for Computer Vision (DL4CV)
IIT Guwahati
Aug-Dec 2022

Video

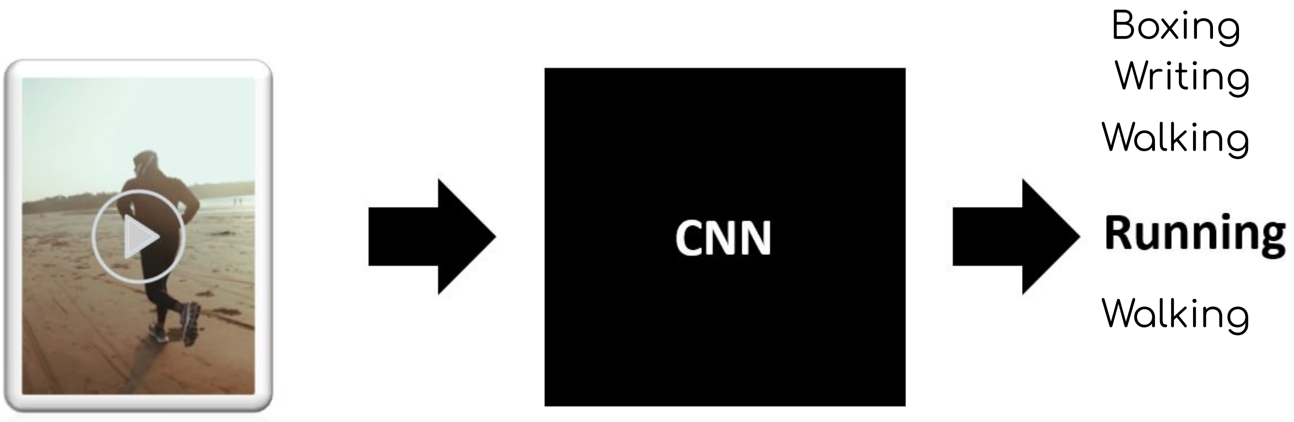


Sequence of frames
4D tensor
T X 3 X H X W

Video Classification



Video Classification



Recognizes the **actions (verbs)** as opposed to the **objects (nouns)** in Image Classification

Challenge

- Videos are BIG!
- Uncompressed video (30fps, 24-bit)
  - Standard Definition (640 X 480) =
  - High Definition (1920 X 1080) =

Challenge

- Videos are BIG!
- Uncompressed video (30fps, 24-bit)
  - Standard Definition (640 X 480) ~1.5GB/minute
  - High Definition (1920 X 1080) ~ 10GB/minute

Solution

- Short clips (3-5 seconds)
- Lower resolution
- E.g. 3.2s video (5 fps, 112 X 112) → 16 frames → 588KB

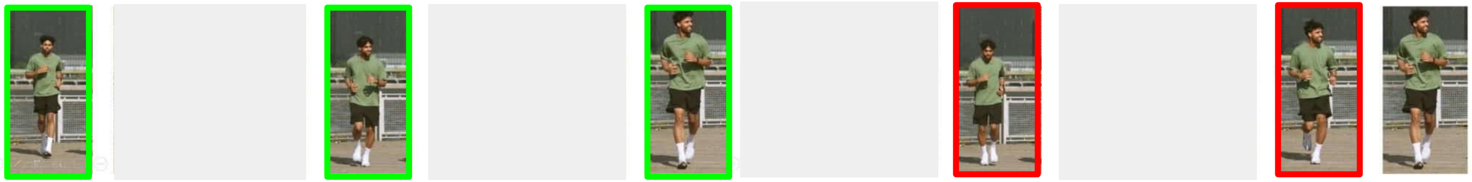
Training on video clips

Raw: long, high fps

Training on video clips

Raw: long, high fps

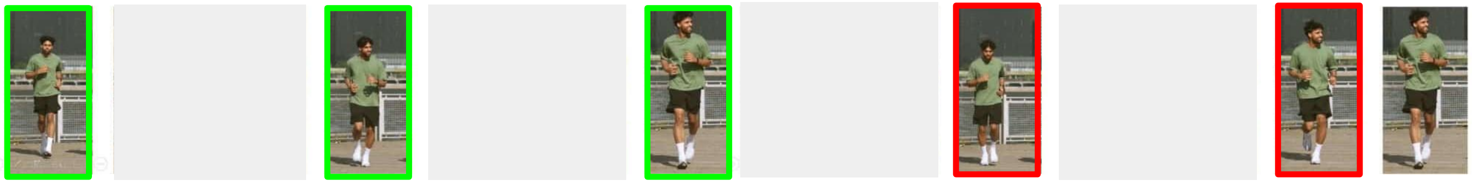
Train: short clips, low fps



Training on video clips

Raw: long, high fps

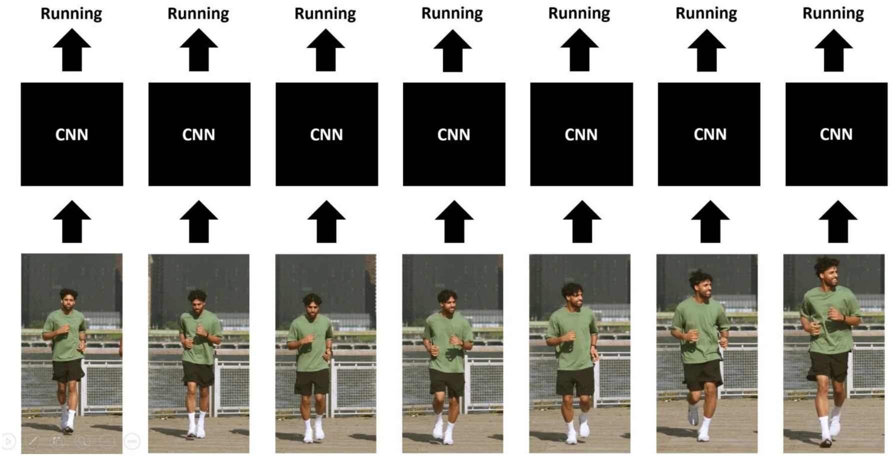
Train: short clips, low fps



Test: run on multiple clips, fuse the predictions

Classification from single frame

- Train 2D CNN to classify video frames
- Test: average the predictions on all the frames of the video



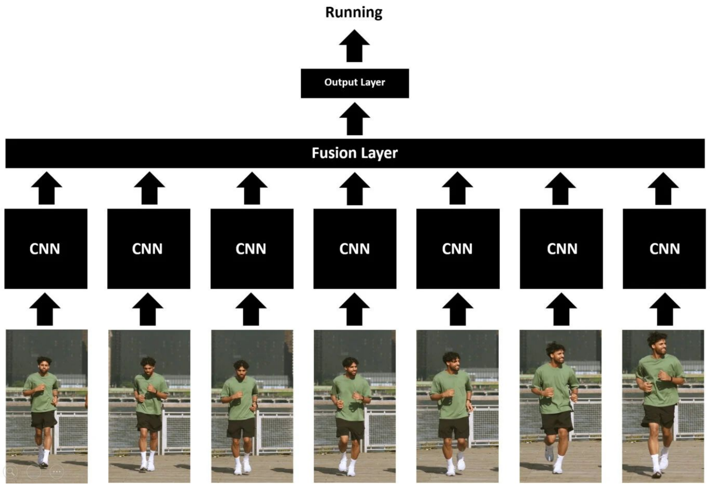
Very strong baseline for video classification!

Classification with Late Fusion

- Extract semantic (high-level appearance) features from each frame
- Combine it from all the frames

T X D X H' X W'

T X 3 X H X W



Flatten/GAP, etc.

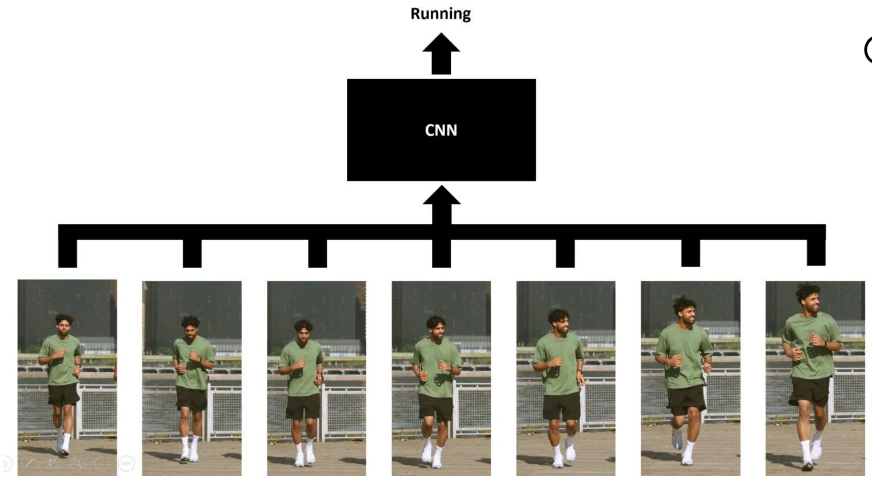
Hard to perceive the low-level motion across the frames!

Classification with Early Fusion

- Compare the frames very early in the network (1st Conv)
- Then operate a 2D CNN

3T X H X W

T X 3 X H X W

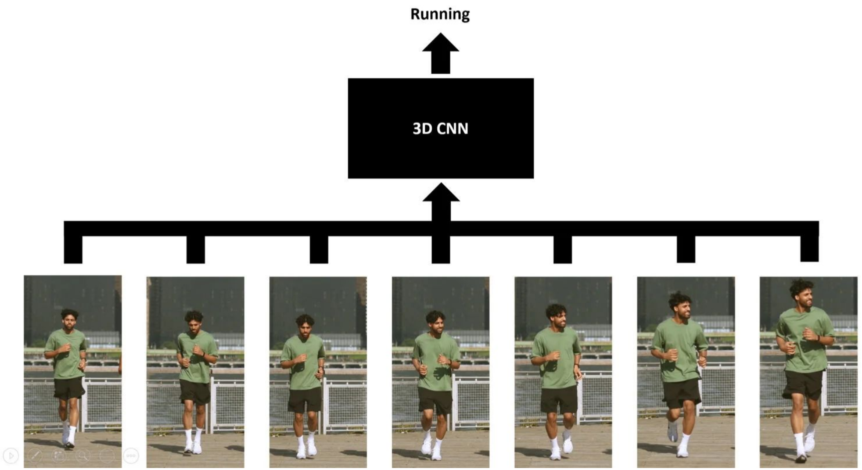


Conv1 consumes all the frames
i/p: 3T X H X W
o/p: D X H' X W'

Only one layer of Temporal Processing!

Classification with 3D CNN

- Use the 3D versions of Convolution and Pooling
- Slowly fuse the temporal information over the layers



3 X T X H X W

Layers are 4D tensors
D X T X H X W

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)		

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3

Late fusion

Example frame and architectures

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)		

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)		

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP		

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)		

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)		

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)		

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP		

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)		

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3
Pool3D (4X4X4)		

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3
Pool3D (4X4X4)	12X5X16X16	6X6X6

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3
Pool3D (4X4X4)	12X5X16X16	6X6X6
Conv3D (3X3X3, 12->24)		

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3
Pool3D (4X4X4)	12X5X16X16	6X6X6
Conv3D (3X3X3, 12->24)	24X5X16X16	14X14X14

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3
Pool3D (4X4X4)	12X5X16X16	6X6X6
Conv3D (3X3X3, 12->24)	24X5X16X16	14X14X14
GAP		

3D CNN

Example video and processing

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3->12)	12X20X64X64	1X3X3
Pool2D (4X4)	12X20X16X16	1X6X6
Conv2D (3X3, 12->24)	24X20X16X16	1X14X14
GAP	24X1X1X1	20X64X64

Late fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv2D (3X3, 3*20->12)	12X64X64	20X3X3
Pool2D (4X4)	12X16X16	20X6X6
Conv2D (3X3, 12->24)	24X16X16	20X14X14
GAP	24X1X1X1	20X64X64

Early fusion

layer	Size (CXTXHW)	Receptive field (TXHXW)
i/p	3X20X64X64	
Conv3D (3X3X3, 3->12)	12X20X64X64	3X3X3
Pool3D (4X4X4)	12X5X16X16	6X6X6
Conv3D (3X3X3, 12->24)	24X5X16X16	14X14X14
GAP	24X1X1X1	20X64X64

3D CNN

Example video and processing

- Spatial: builds slowly
- Temporal: all at one, in the end

Late fusion

- Spatial: builds slowly
- Temporal: all at one, in the beginning

Early fusion

- Spatial: builds slowly
- Temporal: builds slowly
- Slow Fusion

3D CNN

Early Fusion vs. 3D CNN

- Input
  - C_{in} X T X H X W
  - (3D Grid; at each location Cin dimensional feature)
- Weights
  - C_{out} X C_{in} X T X 3 X 3
- Output
  - C_{out} X H X W
  - 2D Grid; C_{out} dimensional features at each location

- Input
  - C_{in} X T X H X W
  - (3D Grid; at each location Cin dimensional feature)
- Weights
  - C_{out} X C_{in} X 3 X 3 X 3
- Output
  - C_{out} X T X H X W
  - 3D grid with Cout dimensional feature at each location

No temporal shift-invariance; separate filters need to be learnt based on the location!

Datasets

- Sports 1M
  - 1M YouTube videos, 487 classes
  - Fine-grained sports
- [HMDB-51]
  - 6849 clips, 51 actions
  - Collected from hollywood movies

Actions	101
Clips	13320
Groups per Action	25
Clips per Group	4-7
Mean Clip Length	7.21 sec
Total Duration	1600 mins
Min Clip Length	1.06 sec
Max Clip Length	71.04 sec
Frame Rate	25 fps
Resolution	320×240
Audio	Yes (51 actions)

Table 1. Summary of Characteristics of UCF101

Early vs. late vs. slow fusion comparison (2014)

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

Table 1: Results on the 200,000 videos of the Sports-1M test set. Hit@k values indicate the fraction of test samples that contained at least one of the ground truth labels in the top k predictions.

C3D: the VGG of video classification (ICCV 2015)

- 3D CNN that uses all 3X3X3 convolutions and 2X2X2 pooling (except the first)
- Popular as the go-to pretrained net for videos

Layer	Size
Input	3 x 16 x 112 x 112
Conv1 (3x3x3)	64 x 16 x 112 x 112
Pool1 (1x2x2)	64 x 16 x 56 x 56
Conv2 (3x3x3)	128 x 16 x 56 x 56
Pool2 (2x2x2)	128 x 8 x 28 x 28
Conv3a (3x3x3)	256 x 8 x 28 x 28
Conv3b (3x3x3)	256 x 8 x 28 x 28
Pool3 (2x2x2)	256 x 4 x 14 x 14
Conv4a (3x3x3)	512 x 4 x 14 x 14
Conv4b (3x3x3)	512 x 4 x 14 x 14
Pool4 (2x2x2)	512 x 2 x 7 x 7
Conv5a (3x3x3)	512 x 2 x 7 x 7
Conv5b (3x3x3)	512 x 2 x 7 x 7
Pool5	512 x 1 x 3 x 3
FC6	4096
FC7	4096
FC8	C

C3D: the VGG of video classification (ICCV 2015)

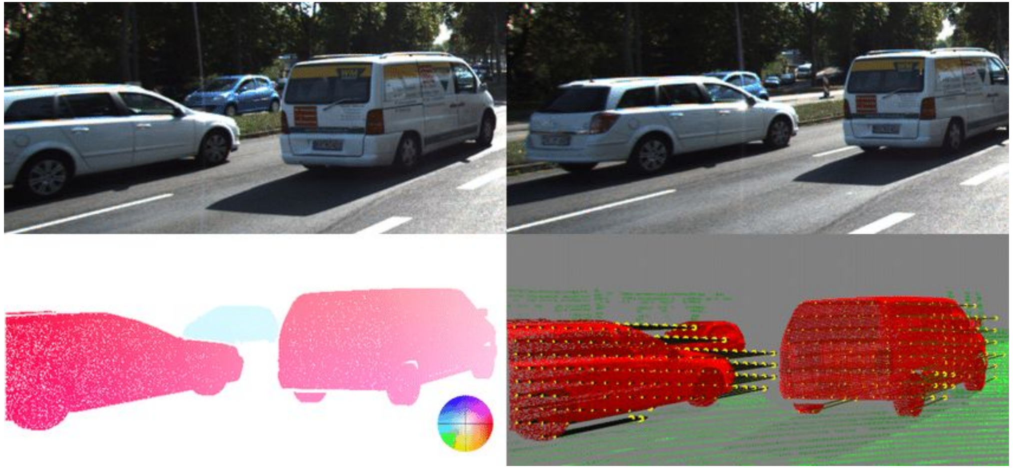
- 80.2 → 84.4
- However, 3X3X3 convolutions are very expensive (C3D ~40GFlops)

Layer	Size
Input	3 x 16 x 112 x 112
Conv1 (3x3x3)	64 x 16 x 112 x 112
Pool1 (1x2x2)	64 x 16 x 56 x 56
Conv2 (3x3x3)	128 x 16 x 56 x 56
Pool2 (2x2x2)	128 x 8 x 28 x 28
Conv3a (3x3x3)	256 x 8 x 28 x 28
Conv3b (3x3x3)	256 x 8 x 28 x 28
Pool3 (2x2x2)	256 x 4 x 14 x 14
Conv4a (3x3x3)	512 x 4 x 14 x 14
Conv4b (3x3x3)	512 x 4 x 14 x 14
Pool4 (2x2x2)	512 x 2 x 7 x 7
Conv5a (3x3x3)	512 x 2 x 7 x 7
Conv5b (3x3x3)	512 x 2 x 7 x 7
Pool5	512 x 1 x 3 x 3
FC6	4096
FC7	4096
FC8	C

Classifying actions based on motion

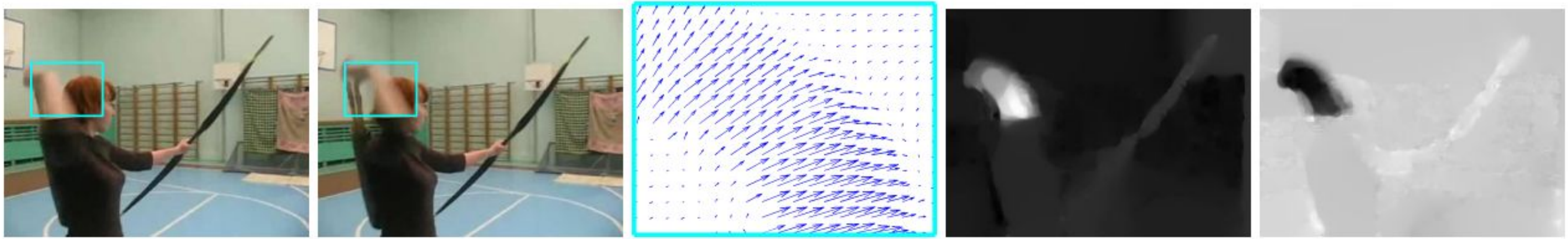
- Humans are good at recognizing from motion

How to represent motion: Optical Flow



Gives the displacement (dx, dy) of each pixel between two images

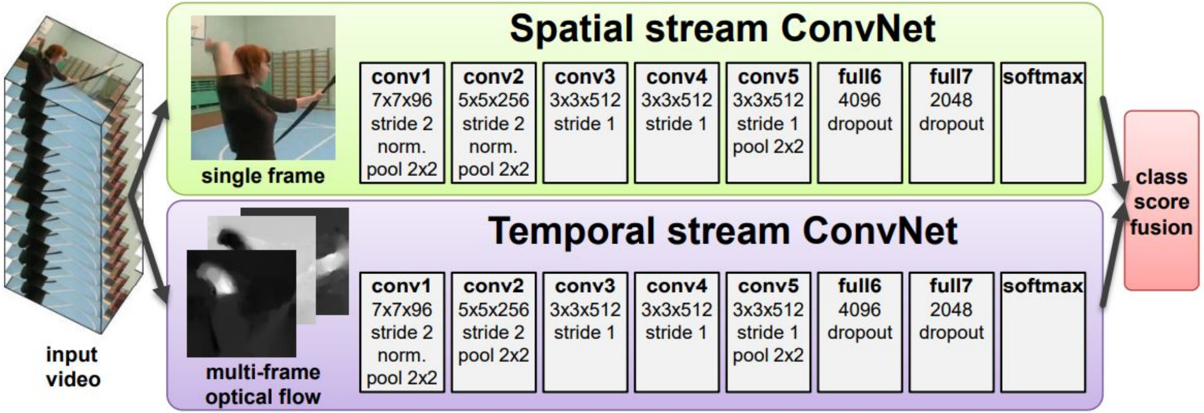
How to represent motion: Optical Flow



Gives the displacement (dx, dy) of each pixel between two images

Two stream networks

- Exploiting the motion information properly



Two stream networks

- Results on UCF-101, HMDB-51

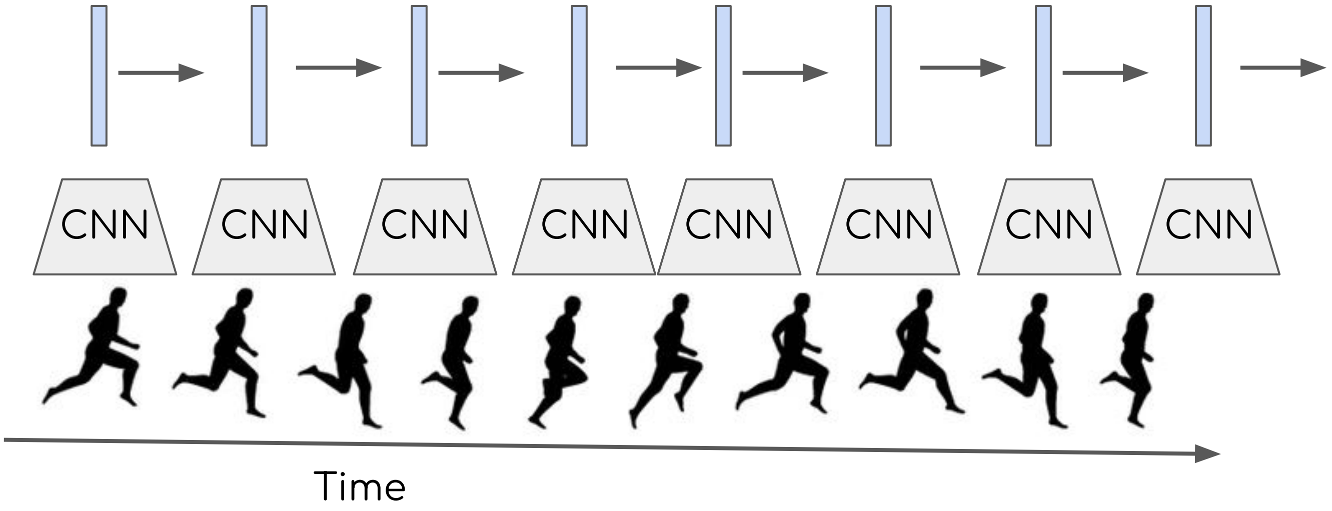
Method	UCF-101	HMDB-51
Improved dense trajectories (IDT) [26, 27]	85.9%	57.2%
IDT with higher-dimensional encodings [20]	**87.9%**	61.1%
IDT with stacked Fisher encoding [21] (based on Deep Fisher Net [23])	-	**66.8%**
Spatio-temporal HMAX network [11, 16]	-	22.8%
"Slow fusion" spatio-temporal ConvNet [14]	65.4%	-
Spatial stream ConvNet	73.0%	40.5%
Temporal stream ConvNet	83.7%	54.6%
Two-stream model (fusion by averaging)	86.9%	58.0%
Two-stream model (fusion by SVM)	**88.0%**	**59.4%**

Can we model the long-term temporal structure?

- Analyze the local features (extracted by a CNN) with an RNN

Can we model the long-term temporal structure?

- Analyze the local features (extracted by a CNN) with an RNN



Next: Generative Models