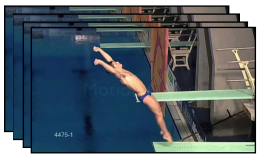
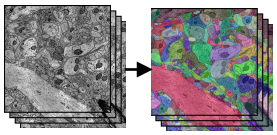
Parallel Separable 3D Convolution for Video and Volumetric Data Understanding

Felix Gonda, Donglai Wei, Toufiq Parag, Hanspeter Pfister

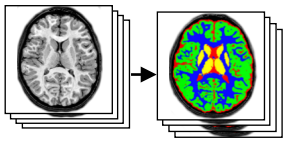
Motivation



Action Rec. in Videos



Neuron Segmentation



MRI Segmentation

3D convolutions capture spatial-temporal signal in video and volumetric data.

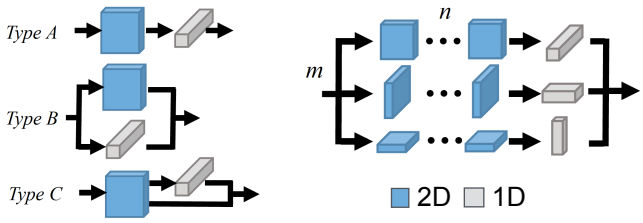
But are expensive:

- Computation speed
- Memory constraints

We use tensor factorization to improve speed and reduce model size.

Our Approach

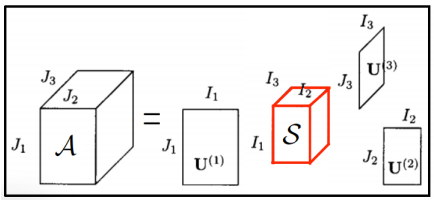
Replace 3D convolution with computational graph of orthogonal pairs of 2D and 1D



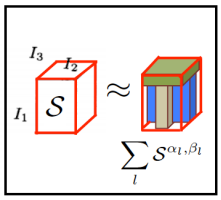
Method	# Streams	# 2D Conv.	Type
P3D(+)	1	1	ABC
R(2+1)D(#)	1	1	A
P_mSC_n(ours)	m	n	A

(+) Pseudo 3D (#) Facebook

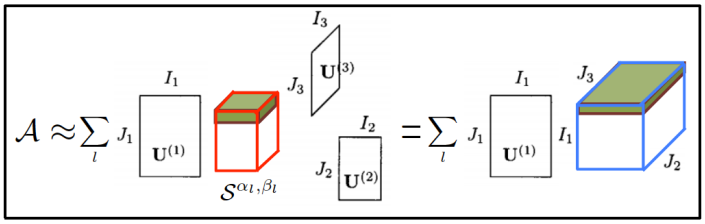
Parallel Streams



Decompose A into orthogonal matrices U(k)

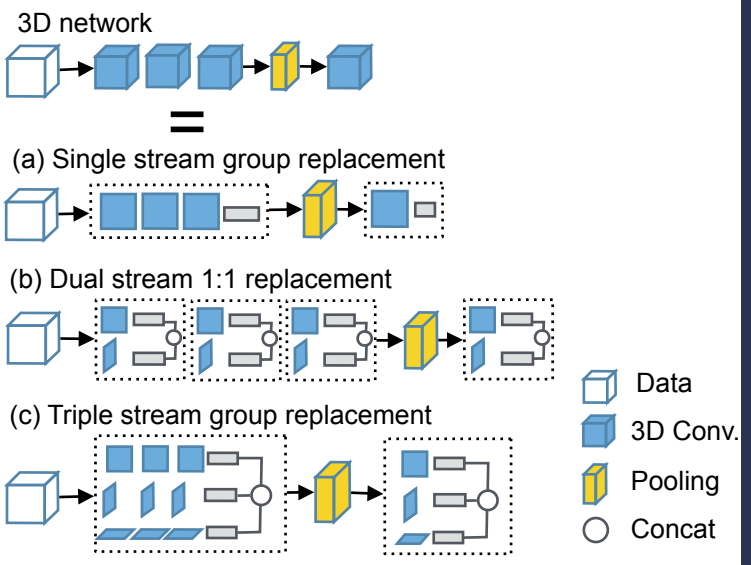


Sparse approximation



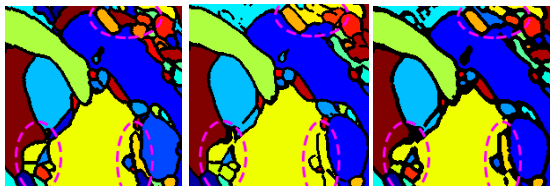
Re-group into (2+1)D (different 2D non-zero planes)

Approximation Examples



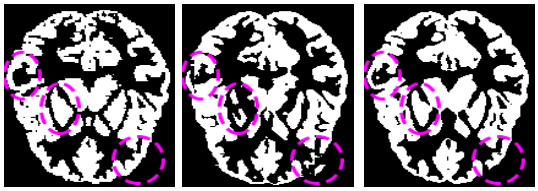
Results

Neuron Segmentation (U-Net on FIBSEM)



Type	VI	# Param
3D	0.10,0.48	21M
P_1SC_1(ours)	0.11,0.24	11M
P_2SC_2(ours)	**0.07,0.23**	12M
P_2SC_3(ours)	0.08,0.28	10M

MRI Segmentation (DenseNet on IBSR)



Type	WM	GM	CSF	#Param
3D	91.3	91.6	94.7	1.6M
P_1SC_1(ours)	95.1	94.1	93.2	4.7M
P_2SC_2(ours)	**95.7**	**96.1**	96.3	2.5M
P_2SC_3(ours)	95.2	**96.1**	**97.6**	**1.4M**

Action Recognition (ResNet on UCF101)

Type	Acc.	#Param
P3D(+)	88.6	261M
3D	85.4	64M
P_1SC_1 (#)	**93.6**	39M
P_1SC_1(ours)	89.7	39M
P_2SC_2(ours)	**92.3**	49M
P_2SC_3(ours)	91.4	39M