GAN-based Photo Video Synthesis

Summary of 3D Convolutional Neural Networks for Human Action Recognition
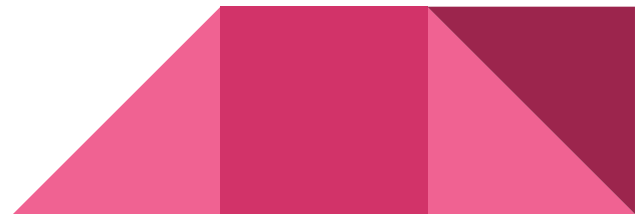Lei Zhang
CS 297

Limitation of 2D CNNs

- Handle only 2D input
- Maps to the spatial dimensions only
- Cannot handle the motion information

3D Convolutional Neural Networks

- Capture both spatial and temporal dimensions
- Extract multiple features from contiguous frames
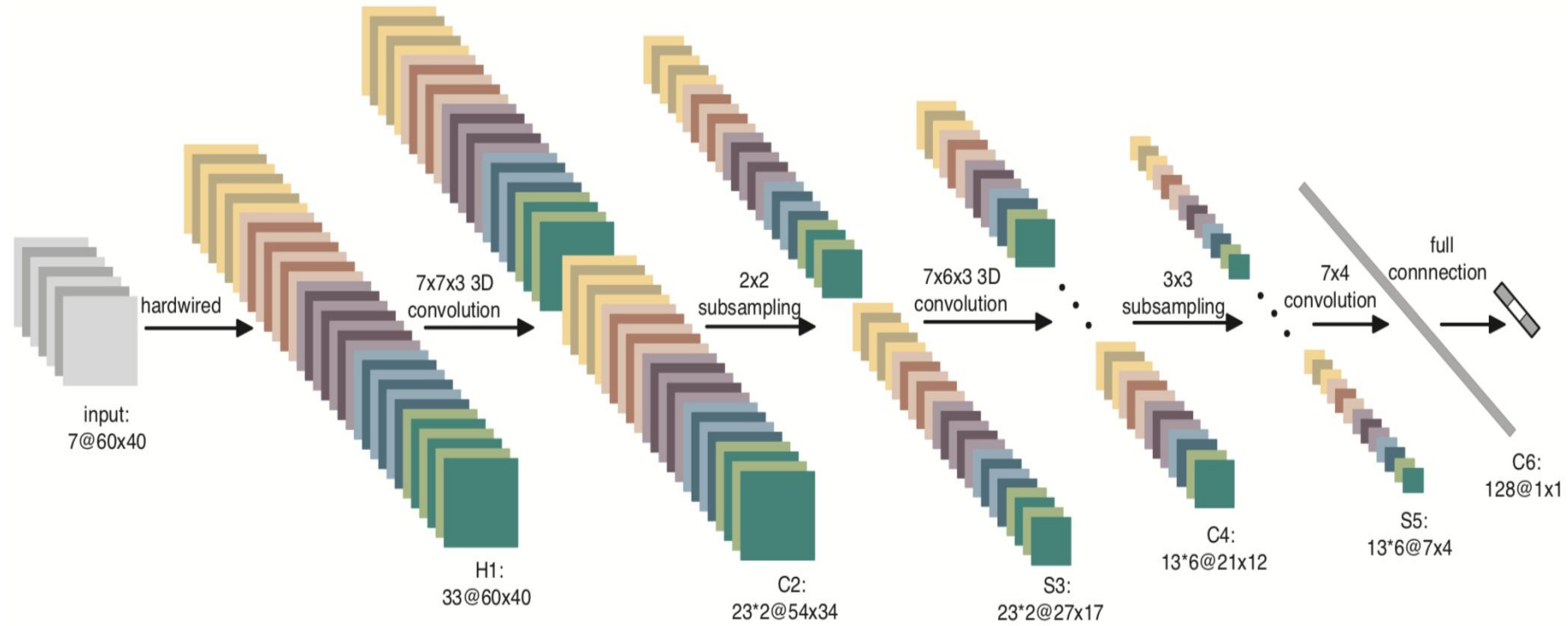- Cannot handle the motion information

3D Convolutional Cube

3D CNN Architecture for Human Action Recognition

Conclusion

- 3D CNN model performs better

METHOD	FPR	MEASURE	CELLTOEAR	OBJECTPUT	POINTING	AVERAGE
3D CNN	0.1%	PRECISION	**0.6433**	**0.6748**	0.8230	**0.7137**
		RECALL	**0.0282**	**0.0256**	0.0152	**0.0230**
		AUC($\times 10^3$)	**0.0173**	**0.0139**	0.0075	**0.0129**
	1%	PRECISION	**0.4091**	**0.5154**	0.7470	**0.5572**
		RECALL	**0.1109**	**0.1356**	0.0931	**0.1132**
		AUC($\times 10^3$)	**0.6759**	**0.7916**	0.5581	**0.6752**
2D CNN	0.1%	PRECISION	0.3842	0.5865	**0.8547**	0.6085
		RECALL	0.0097	0.0176	**0.0192**	0.0155
		AUC($\times 10^3$)	0.0057	0.0109	**0.0110**	0.0092
	1%	PRECISION	0.3032	0.3937	0.7446	0.4805
		RECALL	0.0505	0.0974	**0.1020**	0.0833
		AUC($\times 10^3$)	0.2725	0.5589	**0.6218**	0.4844
SPM$_{GRAY}^{CUBE}$	0.1%	PRECISION	0.3576	0.6051	0.8541	0.6056
		RECALL	0.0088	0.0192	0.0191	0.0157
		AUC($\times 10^3$)	0.0044	0.0108	**0.0110**	0.0087
	1%	PRECISION	0.2607	0.4332	0.7511	0.4817
		RECALL	0.0558	0.0961	0.0988	0.0836
		AUC($\times 10^3$)	0.3127	0.5523	0.5915	0.4855
SPM$_{MEHI}^{CUBE}$	0.1%	PRECISION	0.4848	0.5692	0.8268	0.6269
		RECALL	0.0149	0.0166	0.0156	0.0157
		AUC($\times 10^3$)	0.0071	0.0087	0.0084	0.0081
	1%	PRECISION	0.3552	0.3961	**0.7546**	0.5020
		RECALL	0.0872	0.0825	0.1006	0.0901
		AUC($\times 10^3$)	0.4955	0.4629	0.5712	0.5099

REFERENCE

Ji, Shuiwang, et al. "3D convolutional neural networks for human action recognition." IEEE transactions on pattern analysis and machine intelligence 35.1 (2012): 221-231.