Exploring Spherical Autoencoder for Spherical Video Content Processing

Jin Zhou, Na Li, Shuochao Yao, Yao Liu, Songqing Chen
George Mason University, Rutgers University

DOI:10.1145/3503161.3548364

Why spherical autoencoder?

- 360-degree Video with 3 DoF (Degree of Freedom)
- Widely use in many fields and research areas
- Traditional Video processing:
  - Project frames to 2D content
  - Process with traditional schemes, MPEG
- Spherical autoencoder:
  - Project frames onto 3D mesh content
  - Process with convolution neural network

DOI:10.1145/3503161.3548364

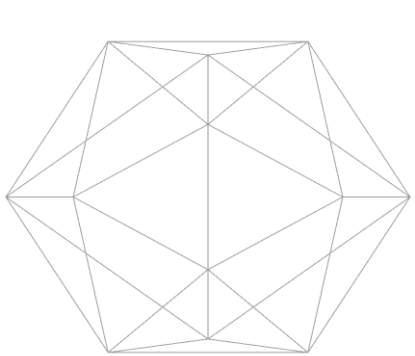
Previous work

Spherical CNN[1]:

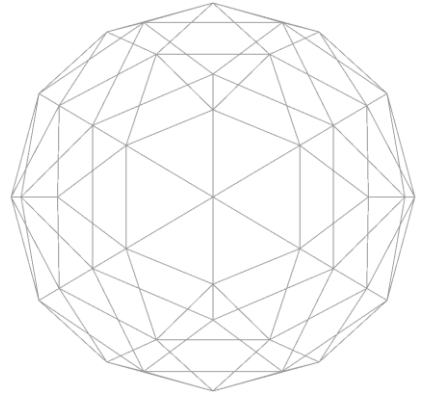
- Lower numbers of network parameters
- Better performance than others

Icosahedral Mesh[2]:

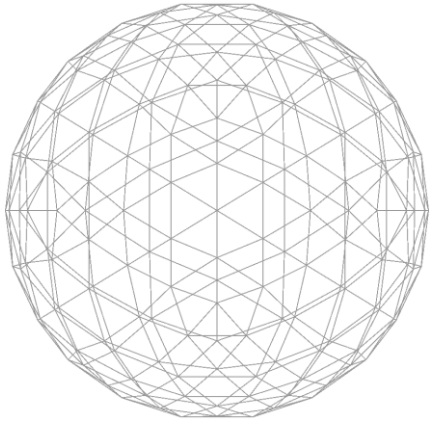
- Discretize use refined icosahedral mesh
- Start with a regular icosahedron
- Divided each faces into 4 smaller faces
- Level-9 mesh contains 2621442 vertices



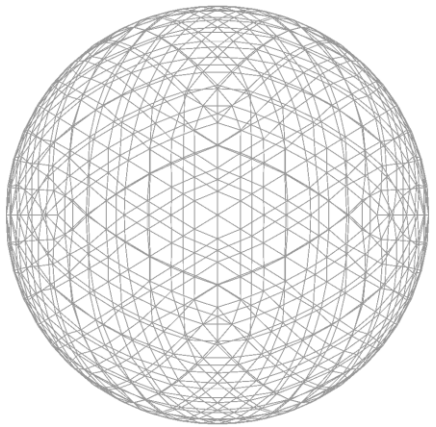
Level-0 mesh



Level-1 mesh



Level-2 mesh

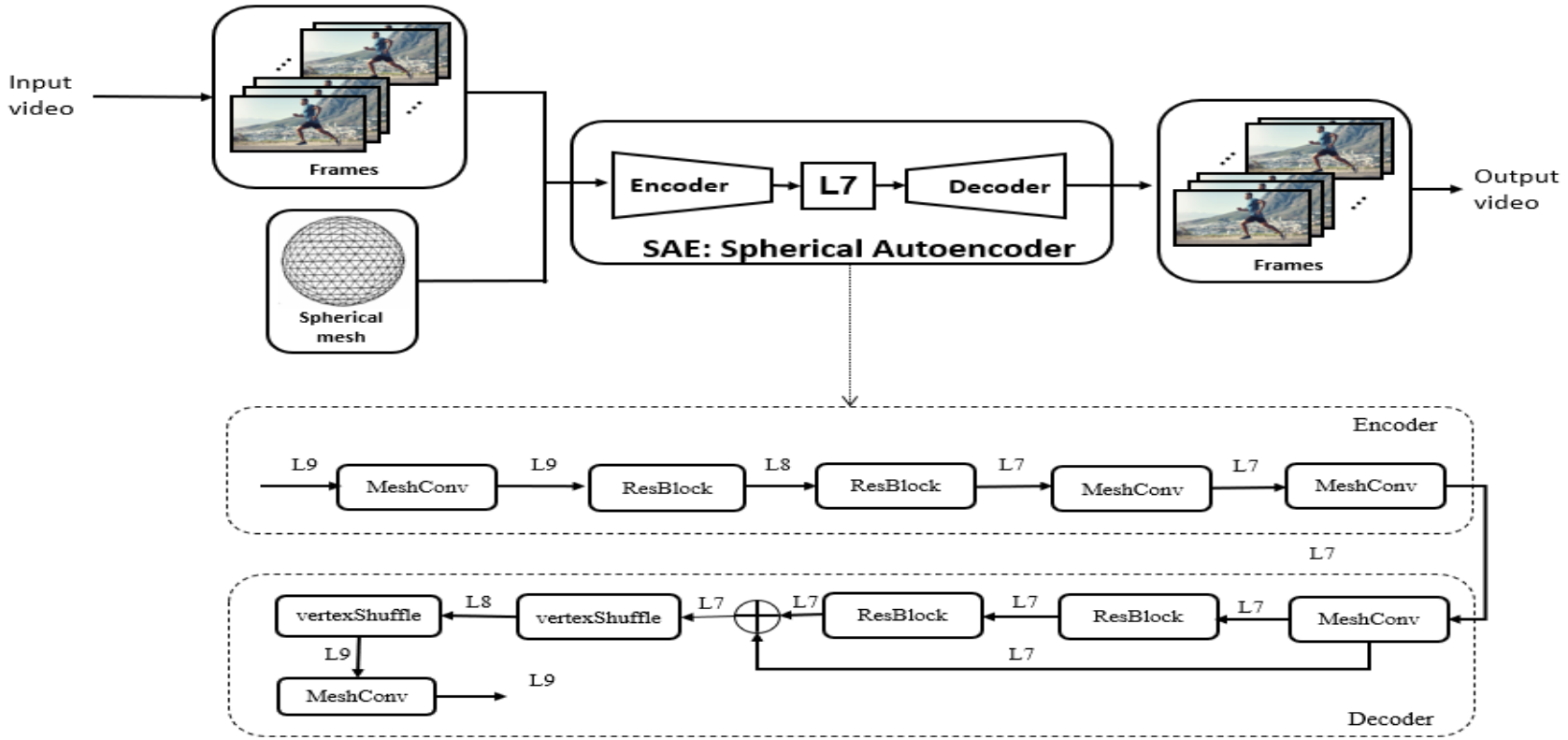


Level-3 mesh

[1] Chiyu" Max" Jiang, et al. "Spherical CNNs on Unstructured Grids." ICLR (Poster). 2019.
[2] Baumgardner, et al. "Icosahedral discretization of the two-sphere." *SIAM Journal on Numerical Analysis* 22.6 (1985)

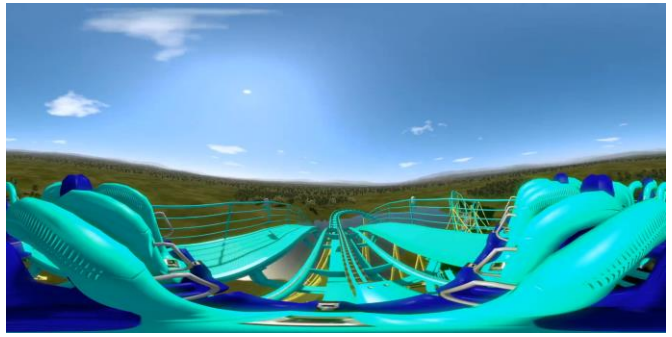
DOI:10.1145/3503161.3548364

Spherical Autoencoder (SAE)

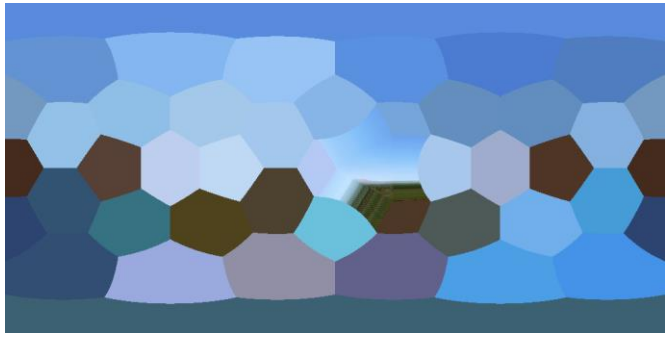


DOI:10.1145/3503161.3548364

Optimization - Partial Spherical Autoencoder



(a) SAE sample frame



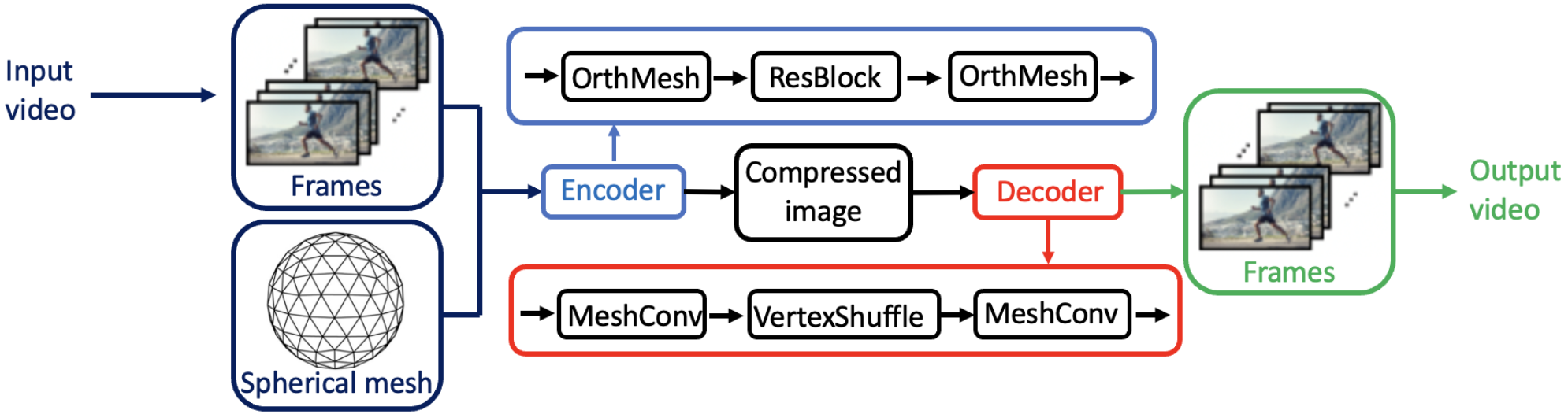
(b) p-SAE sample frame

- Process whole frame each time
- Directly compress it to lower level and use vertexShuffle to recover it.

- Use field-of-view prediction
- Only process 1/80 face each time
- Rotate 79 time to process the whole frame

DOI:10.1145/3503161.3548364

Optimization - Compressive Sensing



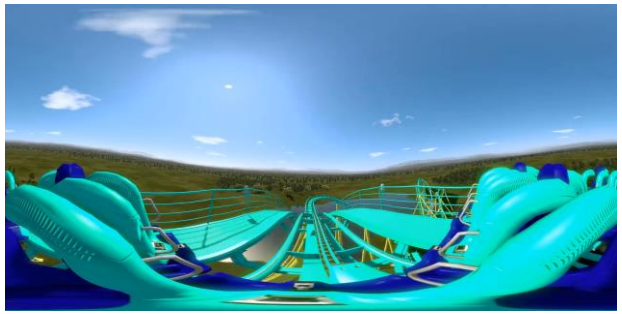
- Orthogonal Regularization < Restricted Isometry Property
- Spectral Normalization < Lipschitz continuity

DOI:10.1145/3503161.3548364

Experiments

- Dataset[1]
  - Highly dynamic, Dynamic, Low motions, Static
  - Different Resolutions：960p, 1080p, 2k, 4k
- Baseline model:
  - 2d Convolutional autoencoder
- Evaluation methods :
  - PSNR, model size, VI-VMAF

(a) classroom



(b) roller-coaster

(c) Indoor

(c) football

[1]These videos are from the “The Psychology of 360-Video” repository, available at https://github.com/vhilab/psych-360

DOI:10.1145/3503161.3548364

Result – PSNR

Model	CAE	SAE				p-SAE			
Model Compression Ratio	none	none	26.01%	41.54%	58.30%	none	36.03%	45.97%	55.11%
Indoor	24.7384	39.2887	41.2183	39.0652	38.3835	37.8229	37.7869	36.4909	35.9083
City	17.1524	39.7354	38.9868	38.5078	37.6546	33.4587	33.6031	32.7565	31.8372
Roller-Coaster	17.8414	34.1936	32.8375	32.0299	31.5185	32.5046	32.1746	31.5830	30.3547
Football	20.6538	36.2050	36.3093	36.2165	36.1787	34.4323	33.9375	33.7339	33.2091
Model	co-CAE	c-SAE				c-p-SAE			
Model Compression Ratio	37.21%	none	26.01%	41.54%	58.30%	none	36.03%	45.97%	55.11%
Indoor	20.1124	40.5602	41.3595	39.7724	38.7182	38.6803	38.5517	37.6399	37.5349
City	15.0681	40.3514	39.4804	39.1991	38.9915	34.5360	36.4247	33.6042	33.0823
Roller-Coaster	13.5168	35.5432	33.8656	33.4480	32.6312	33.2020	32.2679	32.1994	31.0052
Football	14.8910	35.5992	36.8822	36.4333	36.0842	36.6811	36.8463	36.4243	36.1195

- CAE – traditional 2D convolutional autoencoder
- co-CAE – CAE with compressive sensing
- SAE – Spherical autoencoder
- p-SAE – partial mesh in SAE
- c-SAE – Spherical autoencoder with compressive sensing
- c-p-SAE – p-SAE with compressive sensing

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

GEORGE MASON UNIVERSITY

DOI:10.1145/3503161.3548364

Result – Model Size

Model	CAE	SAE				p-SAE			
Model Compression Ratio	none	none	26.01%	41.54%	58.30%	none	36.03%	45.97%	55.11%
Indoor	34.6512	73.4501	73.2218	72.6442	72.6351	70.4519	70.1362	69.8044	69.2314
City	34.0079	72.7804	72.6419	71.9773	71.0025	70.7577	70.5893	69.7488	69.6832
Roller-Coaster	28.5093	68.1255	68.1238	67.9836	67.4821	66.2091	65.9343	65.4290	65.1028
Football	29.8409	68.8549	68.6027	68.4981	68.0195	67.5121	67.4982	67.0034	66.2105
Model	co-CAE	c-SAE				c-p-SAE			
Model Compression Ratio	37.21%	none	26.01%	41.54%	58.30%	none	36.03%	45.97%	55.11%
Indoor	32.8001	75.8872	75.4571	74.7633	73.8041	72.6344	72.1874	71.0801	70.3342
City	31.5367	75.0090	74.6623	74.1342	73.3103	71.9523	71.6345	70.7638	70.1020
Roller-Coaster	25.0956	71.8376	71.0186	71.0123	70.8327	69.7491	68.4566	68.1483	67.3970
Football	26.7573	72.9037	71.9642	71.6107	70.9362	69.9907	69.6016	69.2106	68.6433

Model	CAE	SAE				p-SAE			
Model Compression Ratio	none	none	26.01%	41.54%	58.30%	none	36.03%	45.97%	55.11%
Model Size (KB)	8600	268	195	154	110	250	180	152	126
GPU Usage	4702	13261	13053	12371	11647	2973	2619	2578	2477
Model	co-CAE	c-SAE				c-p-SAE			
Model Compression Ratio	37.21%	none	26.01%	41.54%	58.30%	none	36.03%	45.97%	55.11%
Model Size (KB)	5400	330	207	174	135	321	202	187	131
GPU Usage	3514	15479	15291	14695	14027	1692	1447	1397	1375

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

GEORGE MASON UNIVERSITY

DOI:10.1145/3503161.3548364

Conclusion

- Spherical Autoencoder maintains the high quality of compression by projecting 360-degree content onto spherical mesh
- Partial Mesh Autoencoder achieves the low computing power and small size of transition data
- Compressive sensing efficiently and effectively guarantees the compression ratio and video quality

DOI:10.1145/3503161.3548364

Thank you!

DOI:10.1145/3503161.3548364