Supplementary Materials

Anonymous CVPR submission

Paper ID 13612

Dataset	Point Cloud Sequence	Frames
MVUB	*Andrew*	318
	David	216
	Phil	245
	Ricardo	216
	Sarah	207
8iVFBv2	*Longdress*	300
	Loot	300
	Redandblack	300
	Soldier	300

Table 1. The frame count of each point cloud sequence

A. More Details on Network Architecture

We have introduced the network architecture of SparsePCAC combined with proposed method in the main text. Next, we will present the network architecture obtained by integrating the proposed method with two other baseline models.

A.1. TSC-PCAC with Our Method

The structure of TSC-PCAC [4] is shown in Figure 1, which includes transformer and sparse convolutional module (TSCM) that can better extract local and global information. The structure of TSCM is shown in Figure 4. $\hat{\boldsymbol{y}}$ is divided into C groups according to channels in the TSCM-based channel context module shown in Figure 2, with parallel encoding and decoding within each group and serial encoding and decoding between groups, and using TSCM to extract contextual information between channels. Therefore, we also divide the generated $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ into C groups based on channels, and then use TSCM-based channel context module to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\pi}$. We set $C = 8$ in the experiment.

A.2. SPAC with Our Method

As shown in Figure 3, SPAC [6] uses Fast Fourier Transform to layer the point cloud and trains a compression network similar to SparsePCAC for each layer of the point cloud. Figure 5 shows the details of each module, where offset attention [3] is used to capture both local and global features. The hyperprior module utilizes the hyperprior obtained from the $\mathcal{P}4$ to assist in the encoding and decoding of $\mathcal{P}1$, $\mathcal{P}2$, $\mathcal{P}3$, and $\mathcal{P}4$. Each entropy model has an independent context module. The hyperprior module and context module in Figure 3 are the same as those in SparsePCAC, so we directly refer to the modifications made to SparsePCAC to modify the entropy models of $\mathcal{P}1$, $\mathcal{P}2$, $\mathcal{P}3$, and $\mathcal{P}4$ respectively.

B. Dataset

Figure 6 shows some examples of the training dataset generated using the method in Section 6.1 based on ShapeNet [2] and COCO [5].

MVUB [7] and 8iVFBv2 [1] comprise 9 point cloud sequences. Table 1 presents the frame count of each point cloud sequence, while Figure 7 illustrates the visual results of the first frame of each point cloud sequence.

C. Visual Quality

In Figure 8, we compare the visual quality of reconstruction parts with complex textures.

References

[1] 8i Labs. 8i voxelized full bodies version 2 – a voxelized point cloud dataset. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document, 2017. 1

[2] Angel X. Chang, Thomas Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report Tech Report, Stanford University, 2015. 1, 4

[3] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R. Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7(2):187–199, 2021. 1

[4] Zixi Guo, Yun Zhang, Linwei Zhu, Hanli Wang, and Gangyi Jiang. Tsc-pcac: Voxel transformer and sparse convolution-based point cloud attribute compression for 3d broadcasting. *IEEE Transactions on Broadcasting*, 2024. 1

[5] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312, 2014. 1, 4

[6] Xiaolong Mao, Hui Yuan, Tian Guo, Shiqi Jiang, Raouf Hamzaoui, and Sam Kwong. Spac: Sampling-based progressive attribute compression for dense point clouds. *arXiv preprint arXiv:2409.10293*, 2024. 1

[7] Microsoft. Voxelized Upper Bodies – A Voxelized Point Cloud Dataset. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document, 2016. 1

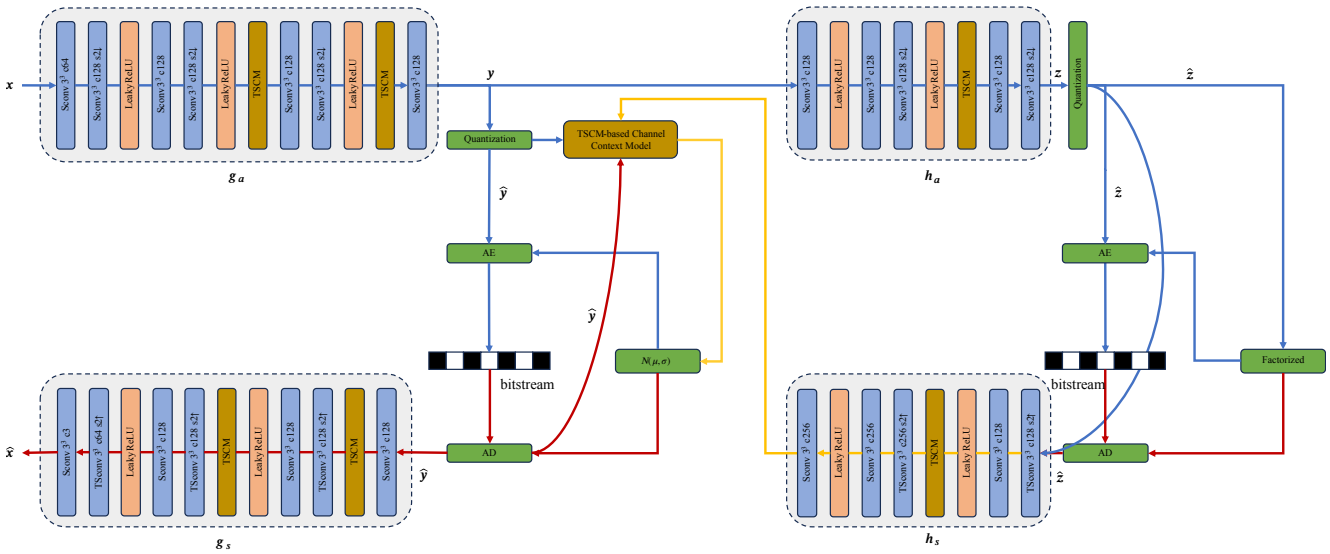


Figure 1. The framework of TSC-PCAC.

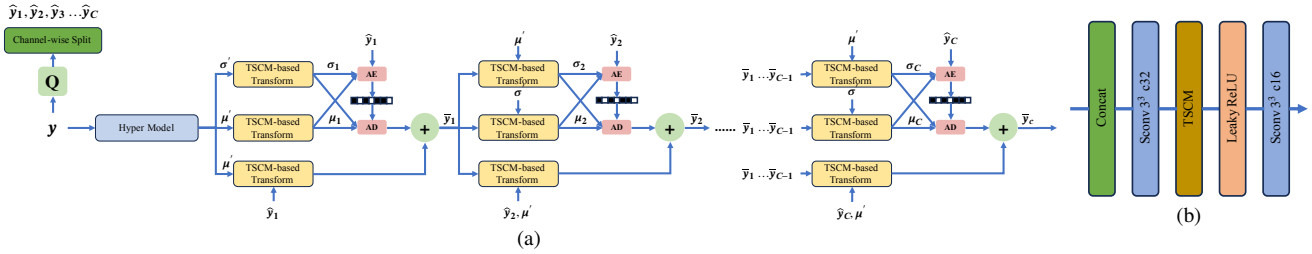


Figure 2. Context module of TSC-PCAC. (a) TSCM-based channel context module. (b) TSCM-based transform.

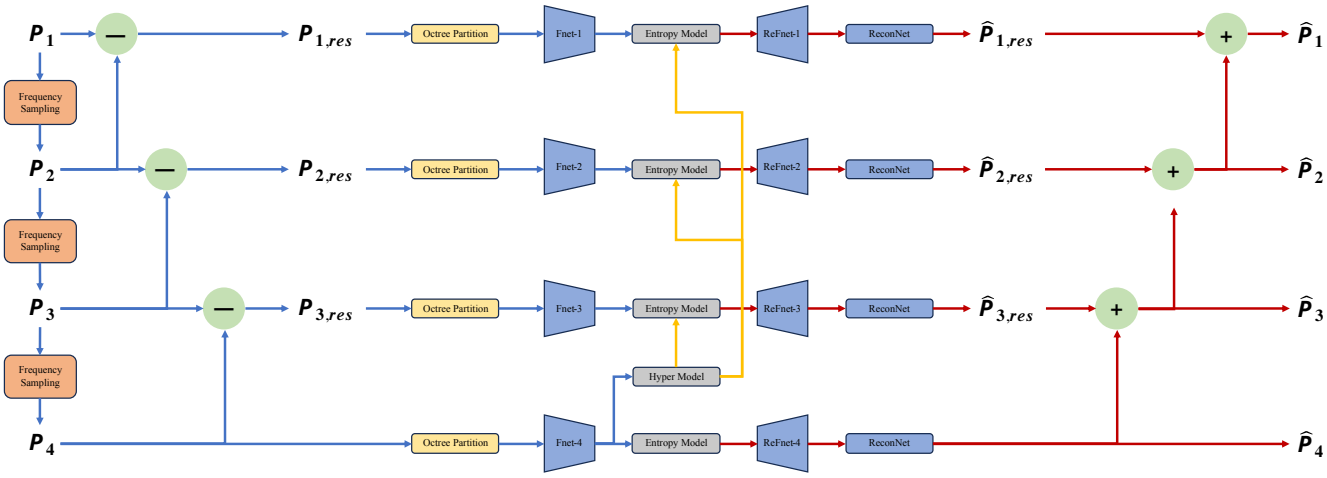


Figure 3. The framework of SPAC.

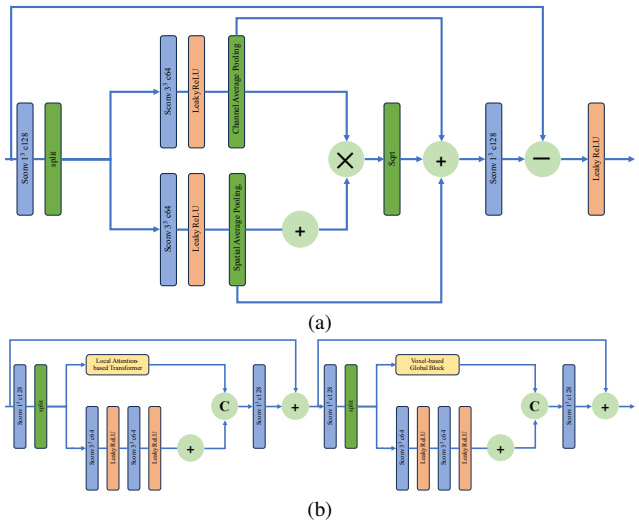


Figure 4. The structure of TSCM. (a) Voxel-based global block. (b) TSCM.

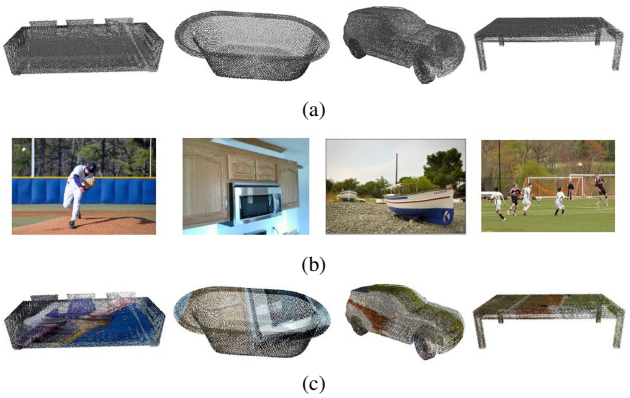


Figure 6. Training dataset. (a) ShapeNet [2]. (b) COCO [5]. (c) Generated point clouds with attributes for training.

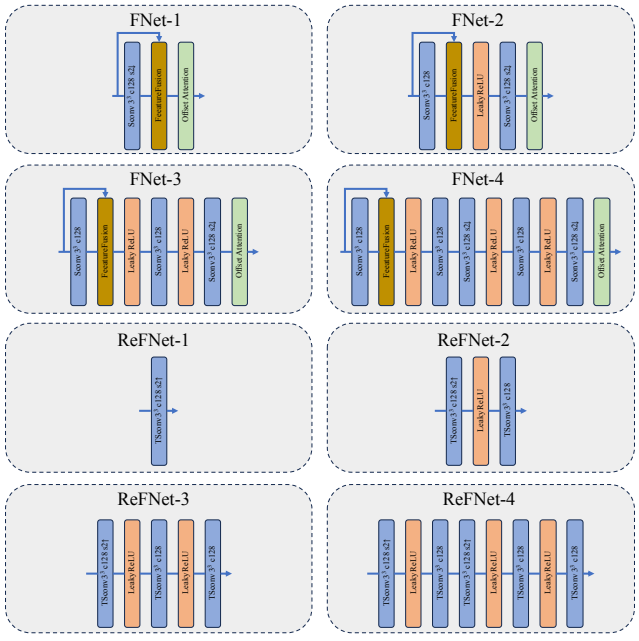


Figure 5. Details of SPAC modules.

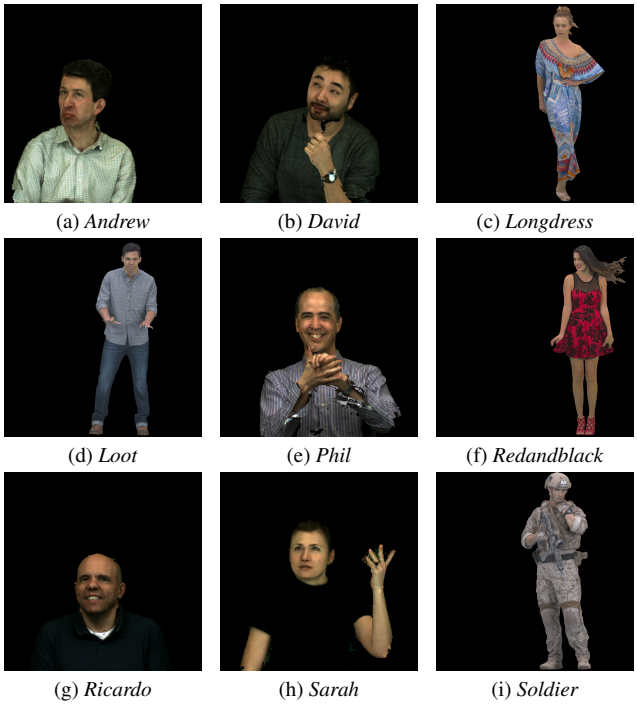
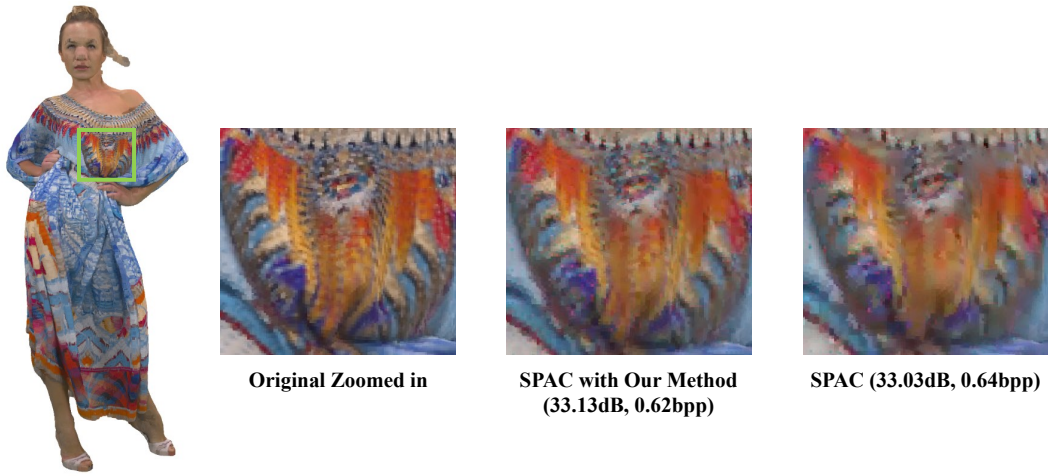


Figure 7. The first frame of 9 point cloud sequences in testing dataset.

CVPR
#13612

(a) *Soldier*



(b) *Longdress*

(c) *Redandblack*

Figure 8. Visual quality comparison.