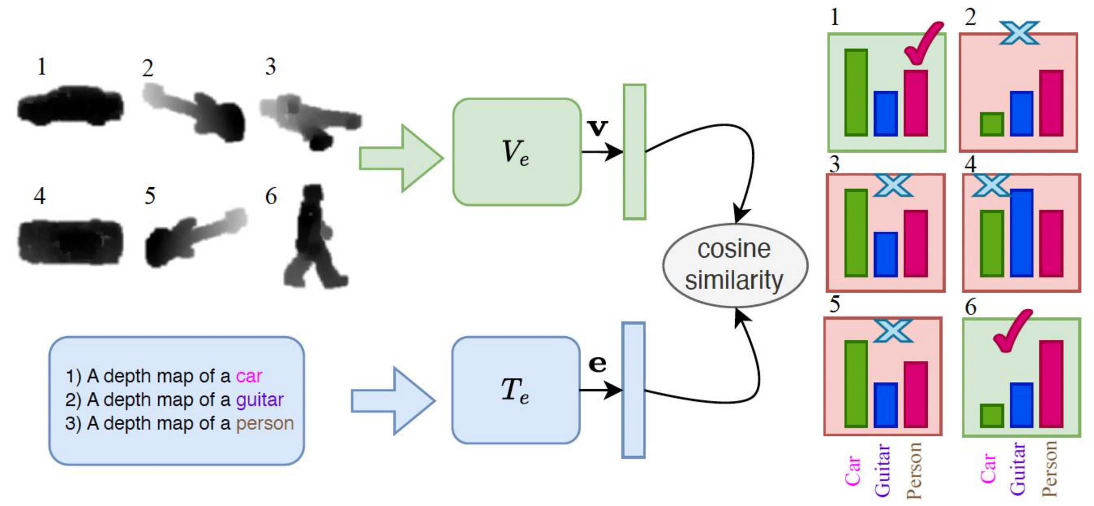
Canonical Shape Projection is All You Need for 3D Few-shot Class Incremental Learning

Ali Cheraghian, Zeeshan Hayder, Sameera Ramasinghe,
Shafin Rahman, Javad Jafaryahya, Lars Petersson, Mehrtash Harandi

Australian National University · MONASH University · North South University · CSIRO · DATA 61 · amazon

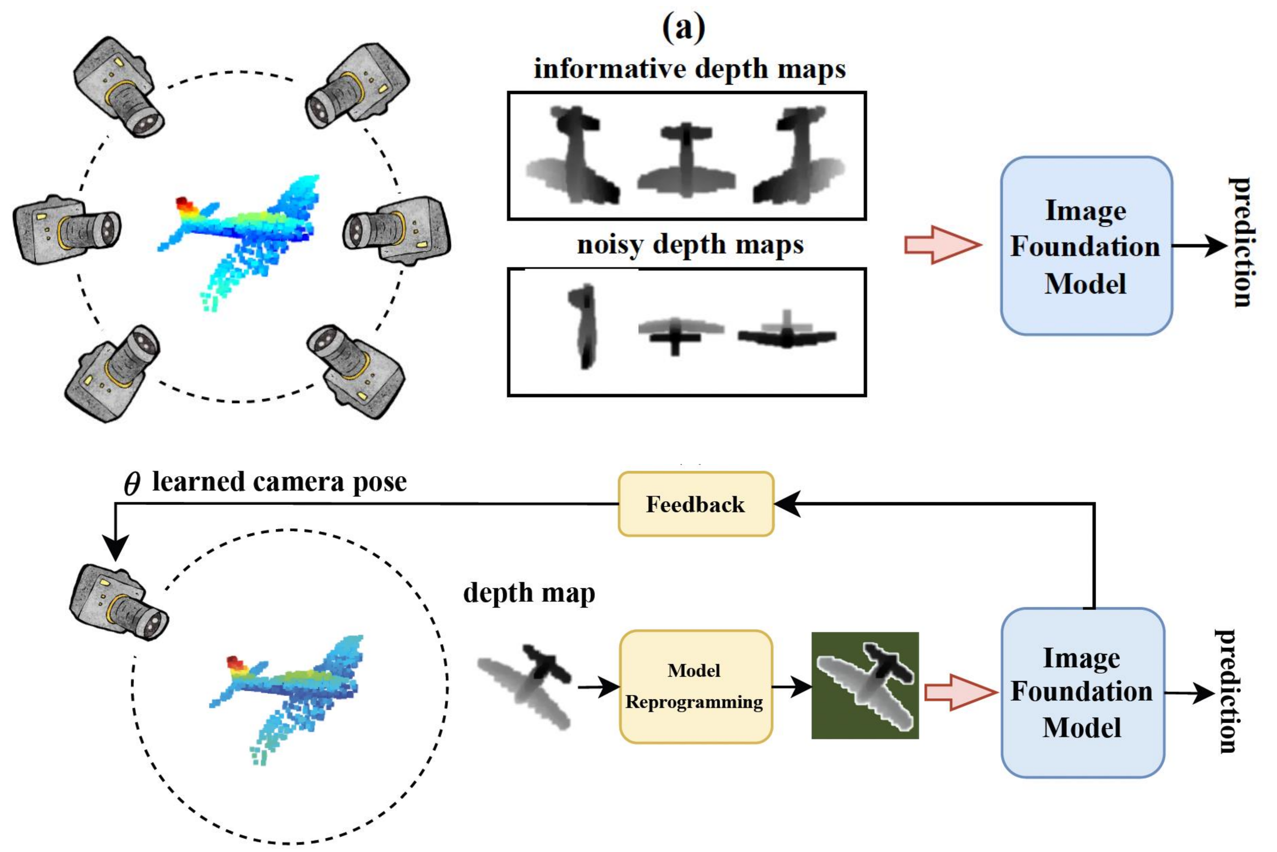
ECCV 2024

Motivation



❑ **Image foundation model for 3D FSCIL:** The existing image foundation models, such as CLIP, are not well-suited for 3D tasks like Few-Shot Class-Incremental Learning (FSCIL).

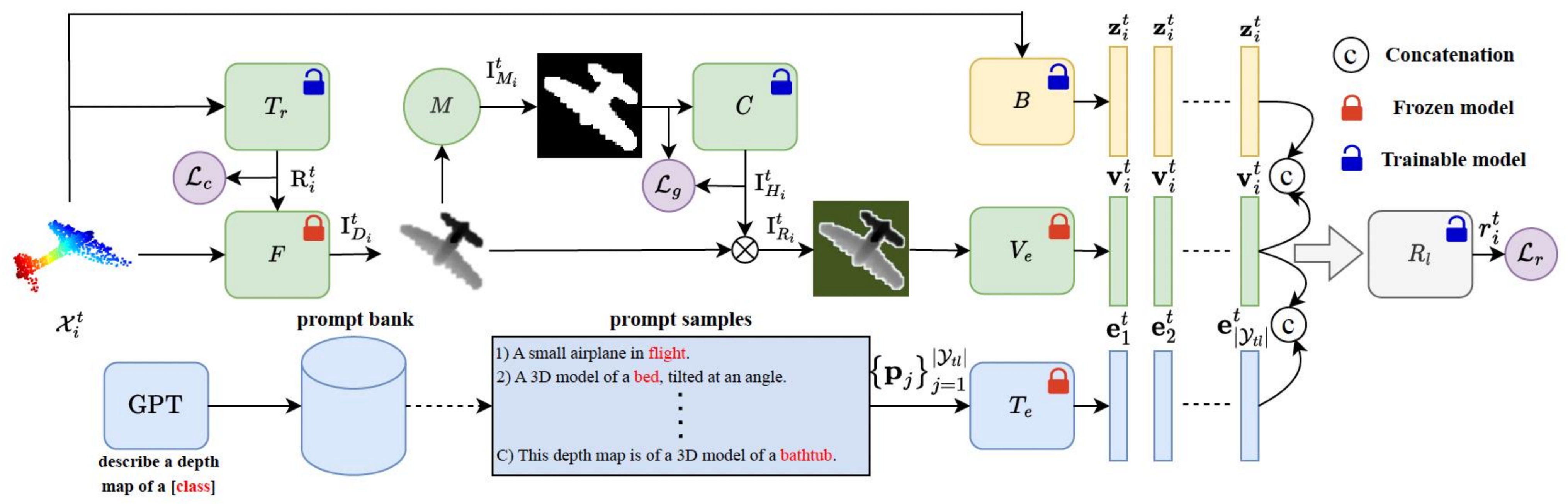
Solution



❑ We propose a novel approach, C3PR, for 3D FSCIL.

- Learning the preferred object-based camera pose for 3D to 2D projections;
- A novel model reprogramming approach for 2D CLIP encoders targeting 3D point cloud data;
- Prompt engineering based on a GPT model for novel few-shot classes to mitigate overfitting.

Our Proposed Method



- **3D to depth map projection**: C3PR learns the optimal camera pose for 3D-to-depth projection using $\mathcal{L}_c$.

$$\mathcal{L}_c = \frac{1}{n}\sum_{i=1}^{n}\left(\left\|\mathbf{R}_i^T\mathbf{R}_i - \mathbf{I}\right\|_{\mathcal{F}}^2 + \left|\det(\mathbf{R}_i) - 1\right|\right)$$

- **Model reprogramming**: This module is used to further adapt the projected depth for the CLIP using $\mathcal{L}_g$.

$$\mathcal{L}_g = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \mathbf{I}_{M_i}}{\partial x} - \frac{\partial \mathbf{I}_{H_i}}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{I}_{M_i}}{\partial y} - \frac{\partial \mathbf{I}_{H_i}}{\partial y}\right)^2$$

- **Prompt engineering:** To mitigate overfitting, we used prompt engineering with GPT to generate additional samples for few-shot novel classes.

- **Vision-language alignment:** The language-vision domains are aligned using a relation model and the corresponding loss $\mathcal{L}_g$.

$$\mathcal{L}_r = -\frac{1}{|\mathcal{Y}_{tl}|\,|S|}\sum_{k\in\mathcal{Y}_{tl}}\sum_{\mathcal{X}_i\in S}\left(\mathbf{1}(y_i^t == k)\log(r_{ik}) + \left(1 - \mathbf{1}(y_i^t == k)\right)\log(1 - r_{ik})\right)$$

- **Training:** The entire pipeline is trained with,

$$\mathcal{L}_t = \lambda_1\mathcal{L}_c + \lambda_2\mathcal{L}_g + \lambda_3\mathcal{L}_r$$

Experiments

Summary of FSCIL results for within-dataset experiments.

	ModelNet					
Method	20	25	30	35	40	Δ↓
FT	89.8	9.7	4.3	3.3	3.0	96.7
Joint	89.8	88.2	87.0	83.5	80.5	10.4
LwF [22]	89.8	36.0	9.1	3.6	3.1	96.0
IL2M [2]	89.8	65.5	58.4	52.3	53.6	40.3
ScaIL [3]	89.8	66.8	64.5	58.7	56.5	37.1
EEIL [5]	89.8	75.4	67.2	60.1	55.6	38.1
FACT [50]	90.4	81.3	77.1	73.5	65.0	28.1
Sem-aware [9]	91.3	82.2	74.3	70.0	64.7	29.1
Microshape [11]	**93.6**	**83.1**	**78.2**	**75.8**	67.1	28.3
C3PR (ours)	91.6	82.3	75.8	72.2	**70.9**	**22.5**

	CO3D						
Method	25	30	35	40	45	50	Δ↓
FT	76.7	11.2	3.6	3.2	1.8	0.8	99.0
Joint	76.7	69.4	64.8	62.7	60.7	59.8	22.0
LwF [22]	76.7	14.7	4.7	3.5	2.3	1.0	98.7
IL2M [2]	76.7	31.5	27.7	18.1	27.1	21.9	71.4
ScaIL [3]	76.7	39.5	34.1	24.1	30.1	27.5	64.1
EEIL [5]	76.7	61.4	52.4	42.8	39.5	32.8	57.2
FACT [50]	77.9	67.1	59.7	54.8	50.2	46.7	40.0
Sem-aware [9]	76.8	66.9	59.2	53.6	49.1	42.9	44.1
Microshape [11]	78.5	67.3	60.1	56.1	51.4	47.2	39.9
C3PR (ours)	**81.5**	**69.4**	**66.5**	**63.0**	**54.2**	**53.8**	**34.0**

	ShapeNet							
Method	25	30	35	40	45	50	55	Δ↓
FT	87.0	25.7	6.8	1.3	0.9	0.6	0.4	99.5
Joint	87.0	85.2	84.3	83.0	82.5	82.2	81.3	6.6
LwF [22]	87.0	60.8	33.5	15.9	3.8	3.1	1.8	97.9
IL2M [2]	87.0	58.6	45.7	40.7	50.1	49.4	49.3	43.3
ScaIL [3]	87.0	56.6	51.8	44.3	50.3	46.3	45.4	47.8
EEIL [5]	87.0	77.7	73.2	69.3	66.4	65.9	65.8	22.4
FACT [50]	87.5	75.3	71.4	69.9	67.5	65.7	62.5	28.6
Sem-aware [9]	87.2	74.9	68.1	69.0	68.1	66.9	63.8	26.8
Microshape [11]	87.6	**83.2**	**81.5**	**79.0**	76.8	73.5	72.6	17.1
C3PR (ours)	**88.0**	81.6	77.8	76.7	**76.9**	**76.2**	**74.7**	**15.1**

Summary of FSCIL results for cross-dataset experiments.

	ShapeNet → CO3D											
Method	39	44	49	54	59	64	69	74	79	84	89	Δ↓
FT	81.0	20.2	2.3	1.7	0.8	1.0	1.0	1.3	0.9	0.5	1.6	98.0
Joint	81.0	79.5	78.3	75.2	75.1	74.8	72.3	71.3	70.0	68.8	67.3	16.9
LwF [22]	81.0	57.4	19.3	2.3	1.0	0.9	0.8	1.3	1.1	0.8	1.9	97.7
IL2M [2]	81.0	45.6	36.8	35.1	31.8	33.3	34.0	31.5	30.6	32.3	30.0	63.0
ScaIL [3]	81.0	50.1	45.7	39.1	39.0	37.9	38.0	36.0	33.7	33.0	35.2	56.5
EEIL [5]	81.0	75.2	69.3	63.2	60.5	57.9	53.0	51.9	51.3	47.8	47.6	41.2
FACT [50]	81.4	76.0	70.3	68.1	65.8	63.5	63.0	60.1	58.2	57.5	55.9	31.3
Sem-aware [9]	80.6	69.5	66.5	62.9	63.2	63.0	61.2	58.3	58.1	57.2	55.2	31.6
Microshape [11]	82.6	77.9	73.9	72.7	67.7	66.2	65.4	63.4	60.6	58.1	57.1	30.9
C3PR (ours)	**83.6**	**80.0**	**77.8**	**75.4**	**72.8**	**72.3**	**70.3**	**67.9**	**64.9**	**64.1**	**63.2**	**24.4**

	ModelNet → ScanObjectNN				
Method	26	30	34	37	Δ↓
FT	88.4	6.4	6.0	1.9	97.9
Joint	88.4	79.7	74.0	71.2	19.5
LwF [22]	88.4	35.8	5.8	2.5	97.2
IL2M [2]	88.4	58.2	52.9	52.0	41.2
ScaIL [3]	88.4	56.5	55.9	52.9	40.2
EEIL [5]	88.4	70.2	61.0	56.8	35.7
FACT [50]	89.1	72.5	68.3	63.5	28.7
Sem-aware [9]	88.5	73.9	67.7	64.2	27.5
Microshape [11]	**89.3**	73.2	68.4	65.1	27.1
C3PR (ours)	88.3	**75.7**	**70.6**	**67.8**	**23.2**

	ShapeNet → ScanObjectNN				
Method	44	49	54	59	Δ↓
FT	81.4	38.7	4.0	0.9	98.9
Joint	81.4	82.5	79.8	78.7	3.3
LwF [22]	81.4	47.9	14.0	5.9	92.8
IL2M [2]	81.4	53.2	43.9	45.8	43.7
ScaIL [3]	81.4	49.0	46.7	40.0	50.9
EEIL [5]	81.4	74.5	69.8	63.4	22.1
FACT [50]	82.3	74.6	69.9	66.8	18.8
Sem-aware [9]	81.3	70.6	65.2	62.9	22.6
Microshape [11]	82.5	74.8	71.2	67.1	18.7
C3PR (ours)	**84.5**	**77.8**	**75.5**	**71.9**	**14.9**

The impact of all our contributions to the proposed architecture.