Alibaba Group

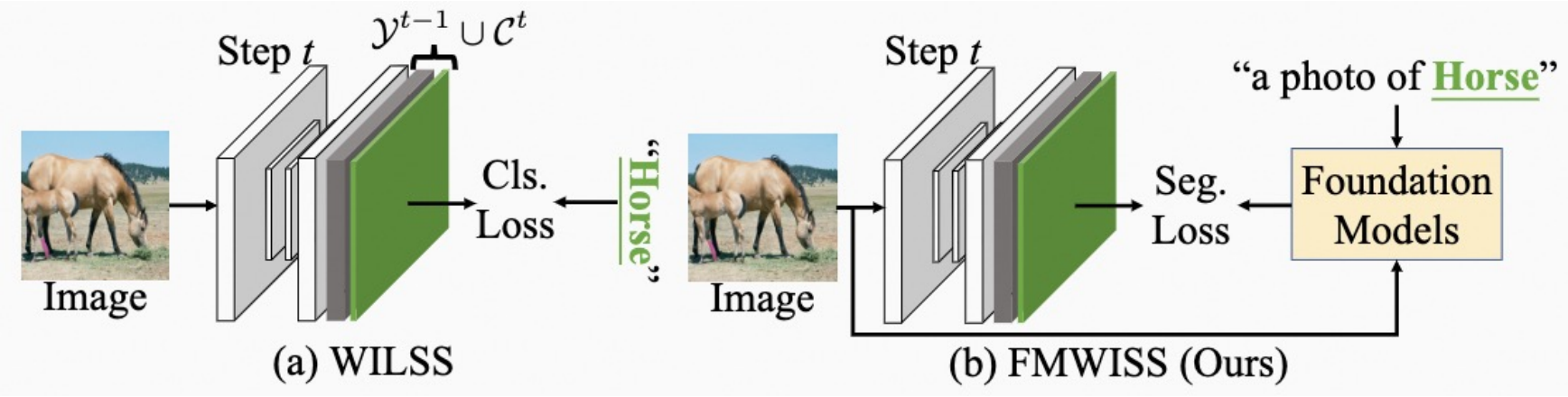
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Foundation Model Drives Weakly Incremental Learning for Semantic Segmentation

Chaohui Yu, Qiang Zhou, Jingliang Li, Jianlong Yuan, Zhibin Wang, Fan Wang

1. Background

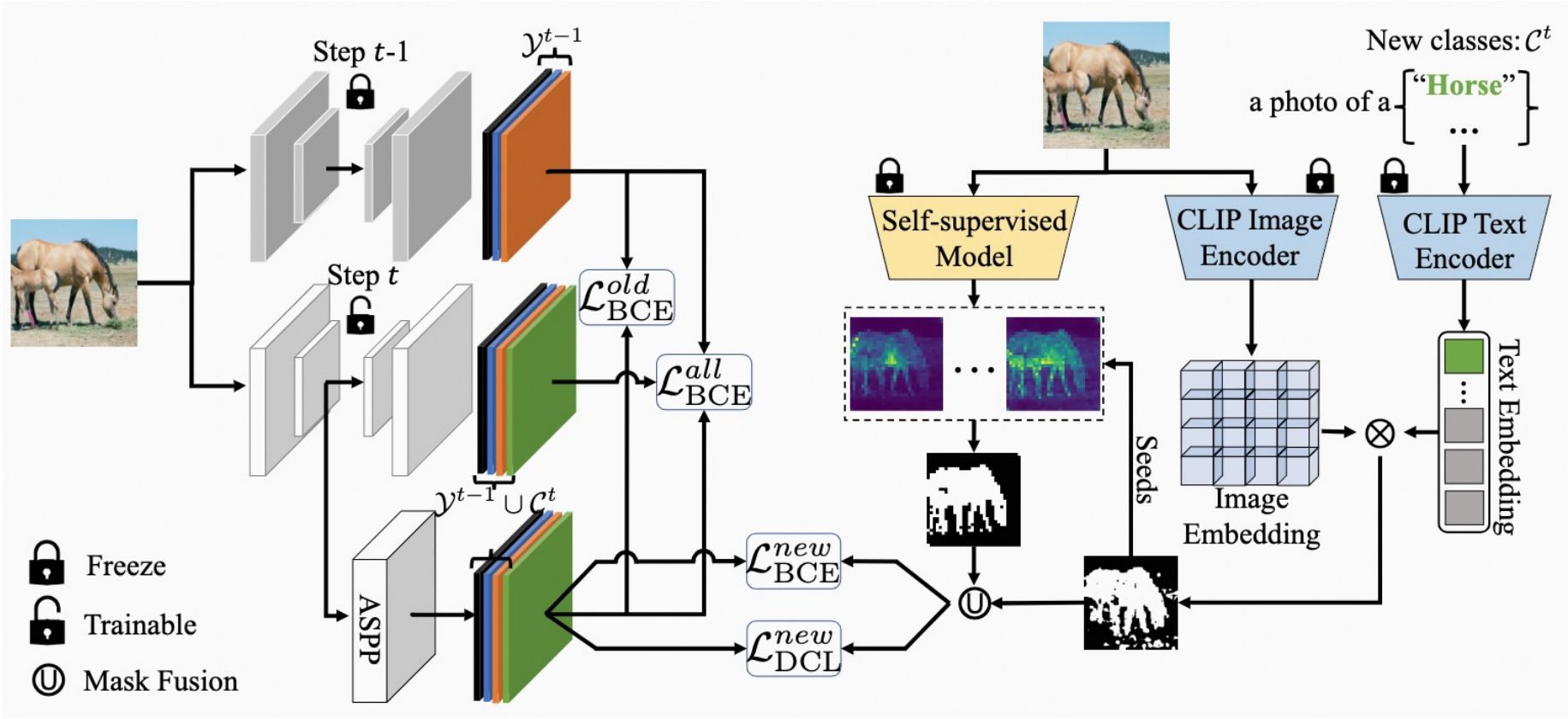
- Modern incremental learning for semantic segmentation (ILSS) methods usually learn new categories based on dense annotations,
- Pixel-by-pixel labeling is costly and time-consuming;
- Weakly incremental learning for semantic segmentation (WILSS) is a novel and attractive task, which aims at learning to segment new classes from cheap and widely available image-level labels.
- Image-level labels can not provide details to locate each segment, which limits the performance of WILSS
- We propose **FMWISS** to improve and effectively utilize the supervision of new classes given image-level labels.



Major difference of pipeline between previous WILSS work and FMWISS

2. Our Method

- Pre-training Based Co-segmentation
- Pseudo Label Optimization
- Memory-based Copy-Paste Augmentation



The proposed FMWISS framework

2. Our Method

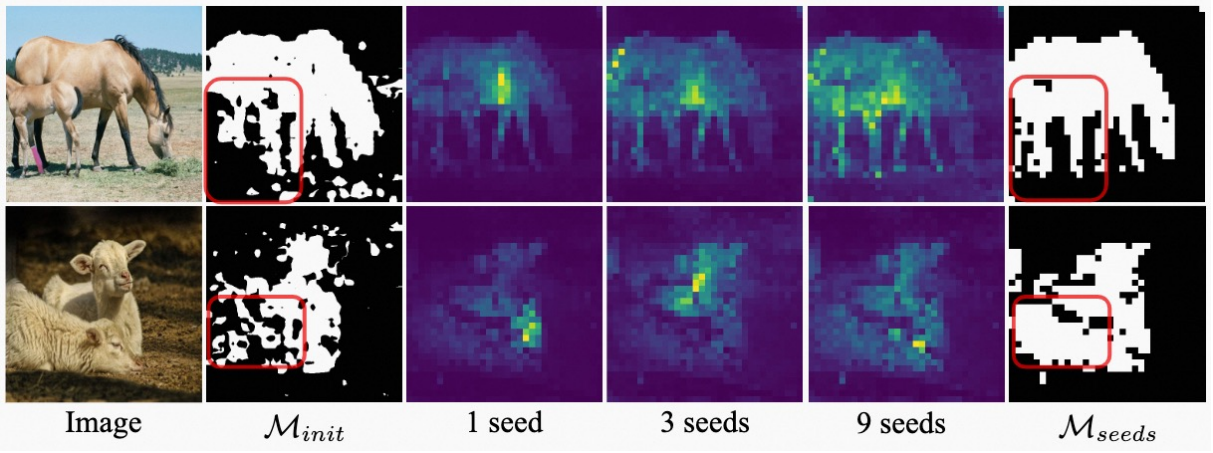
- Pre-training Based Co-segmentation
  - Initial mask generation:

$$\mathcal{M}_{init} = \overline{F} \cdot \overline{T}^{\top}, \ \mathcal{M}_{init} \in \mathbb{R}^{h \times w \times \mathcal{C}^t},$$

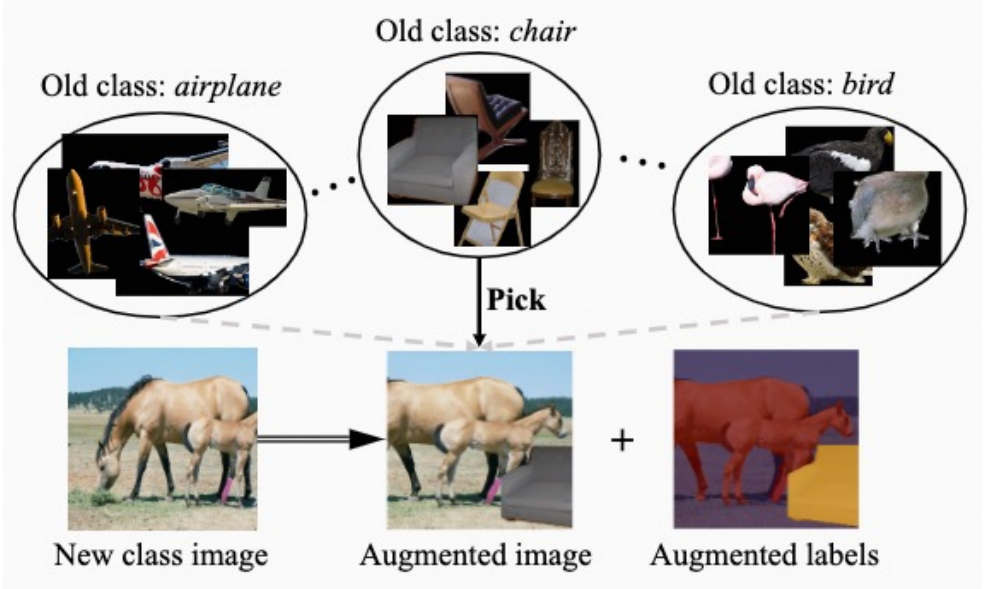
  - Refine mask via seeds guidance:

$$\mathcal{M}^c_{seeds} = [\frac{1}{N}\sum_{p=1}^{N}\frac{1}{n}\sum^{n} S(x)]_{\texttt{binary}},$$

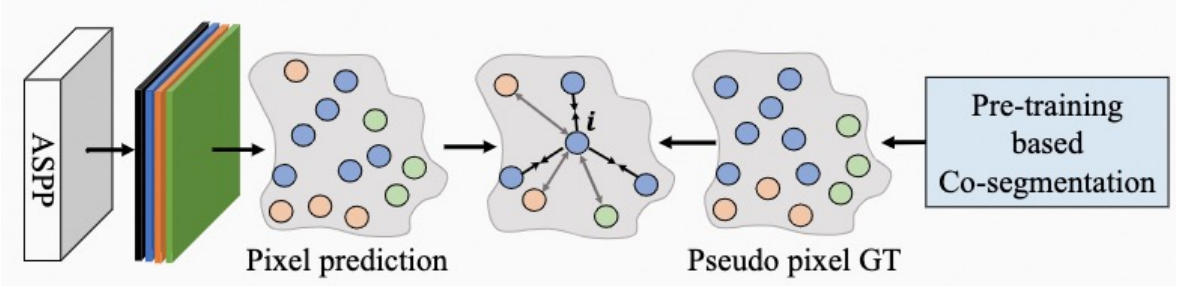
$$\mathcal{M}_{refine} = \mathcal{M}_{init} \cup \mathcal{M}_{seeds}, \ \mathcal{M}_{refine} \in \mathbb{R}^{h \times w \times \mathcal{C}^t},$$



- Memory-based Copy-Paste Augmentation



- Pseudo Label Optimization



Dense Contrastive Learning

3. Experimental Results

Method	Sup	Disjoint			Overlap		
		1-10	11-20	All	1-10	11-20	All
Joint*	P	76.6	74.0	75.4	76.6	74.0	75.4
FT*	P	7.7	60.8	33.0	7.8	58.9	32.1
LWF* [34]	P	63.1	61.1	62.2	70.7	63.4	67.2
LWF-MC* [45]	P	52.4	42.5	47.7	53.9	43.0	48.7
ILT* [38]	P	67.7	61.3	64.7	70.3	61.9	66.3
CIL* [29]	P	37.4	60.6	48.8	38.4	60.0	48.7
MIB* [9]	P	66.9	57.5	62.4	70.4	63.7	67.2
PLOP [18]	P	63.7	60.2	63.4	69.6	62.2	67.1
SDR* [39]	P	67.5	57.9	62.9	70.5	63.9	67.4
RECALL* [36]	P	64.1	56.9	61.9	66.0	58.8	63.7
CAM†	I	65.4	41.3	54.5	70.8	44.2	58.5
SEAM† [48]	I	65.1	53.5	60.6	67.5	55.4	62.7
SS† [3]	I	60.7	25.7	45.0	69.6	32.8	52.5
EPS† [32]	I	64.2	54.1	60.6	69.0	57.0	64.3
WILSON† [8]	I	64.5	54.3	60.8	70.4	57.1	65.0
FMWISS (Ours)	I	**68.5 (+4.0)**	**58.2 (+3.9)**	**64.6 (+3.8)**	**73.8 (+3.4)**	**62.3 (+5.2)**	**69.1 (+4.1)**

Table 1:Results on the 10-10 setting of Pascal VOC

Method	Sup	Disjoint			Overlap		
		1-15	16-20	All	1-15	16-20	All
Joint*	P	75.5	73.5	75.4	75.5	73.5	75.4
FT*	P	8.4	33.5	14.4	12.5	36.9	18.3
LWF* [34]	P	39.7	33.3	38.2	67.0	41.8	61.0
LWF-MC* [45]	P	41.5	25.4	37.6	59.8	22.6	51.0
ILT* [38]	P	31.5	25.1	30.0	69.0	46.4	63.6
CIL* [29]	P	42.6	35.0	40.8	14.9	37.3	20.2
MIB* [9]	P	71.8	43.3	64.7	75.5	49.4	69.0
PLOP [18]	P	71.0	42.8	64.3	75.7	51.7	70.1
SDR* [39]	P	73.5	47.3	67.2	75.4	52.6	69.9
RECALL* [36]	P	69.2	52.9	66.3	67.7	54.3	65.6
CAM†	I	69.3	26.1	59.4	69.9	25.6	59.7
SEAM† [48]	I	71.0	33.1	62.7	68.3	31.8	60.4
SS† [3]	I	71.6	26.0	61.5	72.2	27.5	62.1
EPS† [32]	I	72.4	38.5	65.2	69.4	34.5	62.1
WILSON† [8]	I	73.6	43.8	67.3	74.2	41.7	67.2
FMWISS (Ours)	I	**75.9 (+2.3)**	**50.8 (+7.0)**	**70.7 (+3.4)**	**78.4 (+4.2)**	**54.5 (+12.8)**	**73.3 (+6.1)**

Table 2: Results on the 15-5 setting of Pascal VOC

Method	Sup	COCO			VOC
		1-60	61-80	All	61-80
FT†	P	1.9	41.7	12.7	75.0
LWF† [34]	P	36.7	49.0	40.3	73.6
ILT† [38]	P	37.0	43.9	39.3	68.7
MIB† [9]	P	34.9	47.8	38.7	73.2
PLOP† [18]	P	35.1	39.4	36.8	64.7
CAM†	I	30.7	20.3	28.1	39.1
SEAM† [48]	I	31.2	28.2	30.5	48.0
SS† [3]	I	35.1	36.9	35.5	52.4
EPS† [32]	I	34.9	38.4	35.8	55.3
WILSON† [8]	I	39.8	41.0	40.6	55.7
FMWISS (Ours)	I	**39.9 (+0.1)**	**44.7 (+3.7)**	**41.6 (+1.0)**	**63.6 (+7.9)**

Table 3: Results on COCO-to-VOC

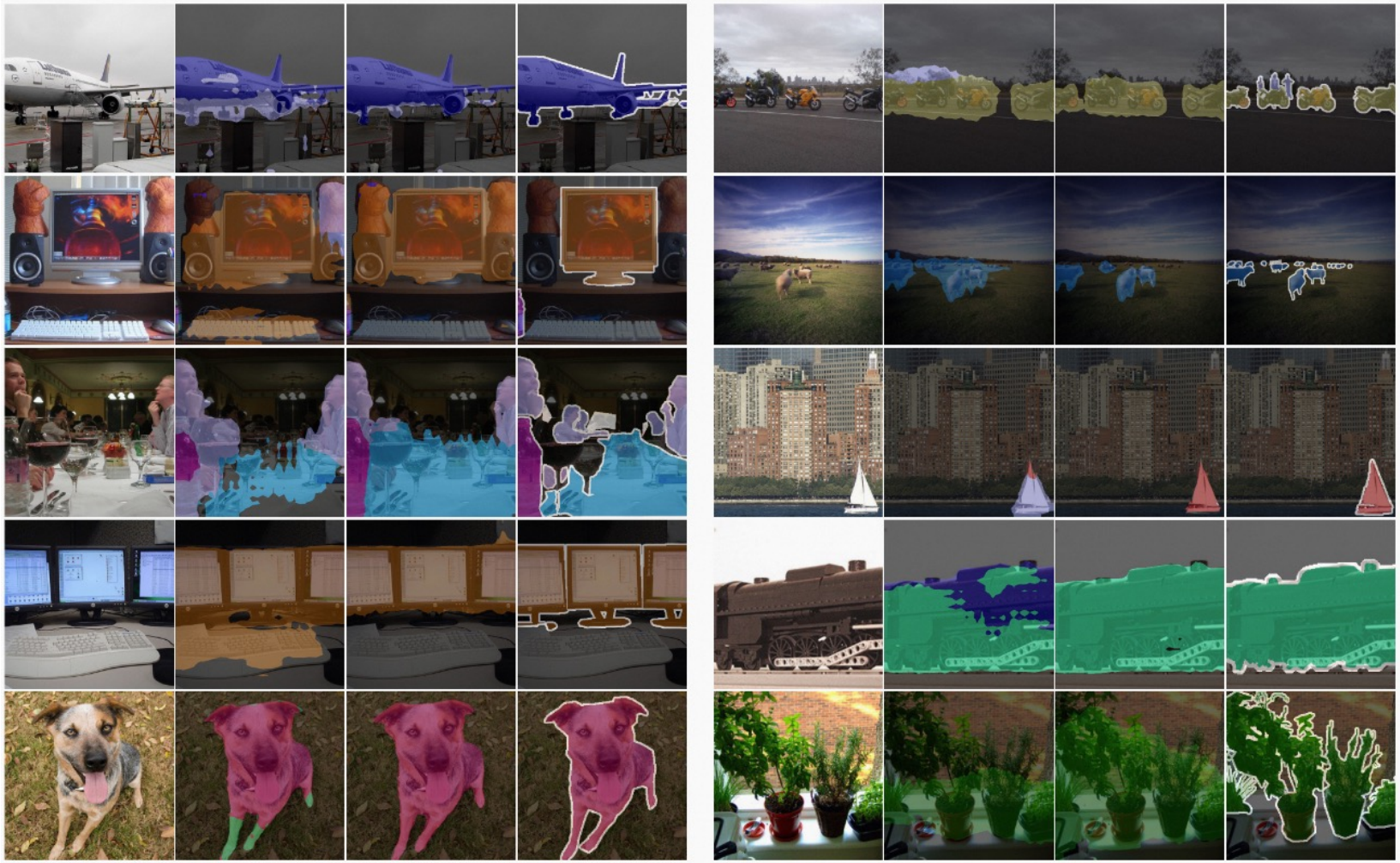
Method	Sup	Disjoint			Overlap		
		1-10	11-20	All	1-10	11-20	All
WILSON [8]	I	64.5	54.3	60.8	70.4	57.1	65.0
WILSON [8]	P	69.5	56.4	64.2	73.6	57.6	66.7
FMWISS (Ours)	I	68.5	58.2	**64.6**	73.8	62.3	**69.1**

Table 4: Comparison trained with dense annotations

Method	Train Data	Disjoint			Overlap		
		1-10	11-20	All	1-10	11-20	All
WILSON [8]	100%	64.5	54.3	60.8	70.4	57.1	65.0
FMWISS (Ours)	100%	68.5	58.2	64.6	73.8	62.3	69.1
	50%	66.7	56.0	62.7	72.1	60.5	67.4
	30%	68.5	51.5	61.5	75.7	55.7	66.8

Table 5: Performance comparison with fewer training data

3. Experimental Results



More comparison on the 10-10 VOC setting. From left to right: image, WILSON, FMWISS (ours) and the ground-truth.

Thanks !