LCD^3: Layout-Conditioned Diffusion for Dataset Distillation in Object Detection

Yue Cao[1] Mohsen Zardadi[2] Yu Hu[1] Yanshuo Fan[1] Jozsef Hamari[2] Zheng Liu[1] Jianyang Gu[3]

[1]*The University of British Columbia* [2]*TerraSense Analytics* [3]*The Ohio State University*

NeurIPS 2026 · Poster #3857

Object Detection Datasets Contain Redundant Visual Information



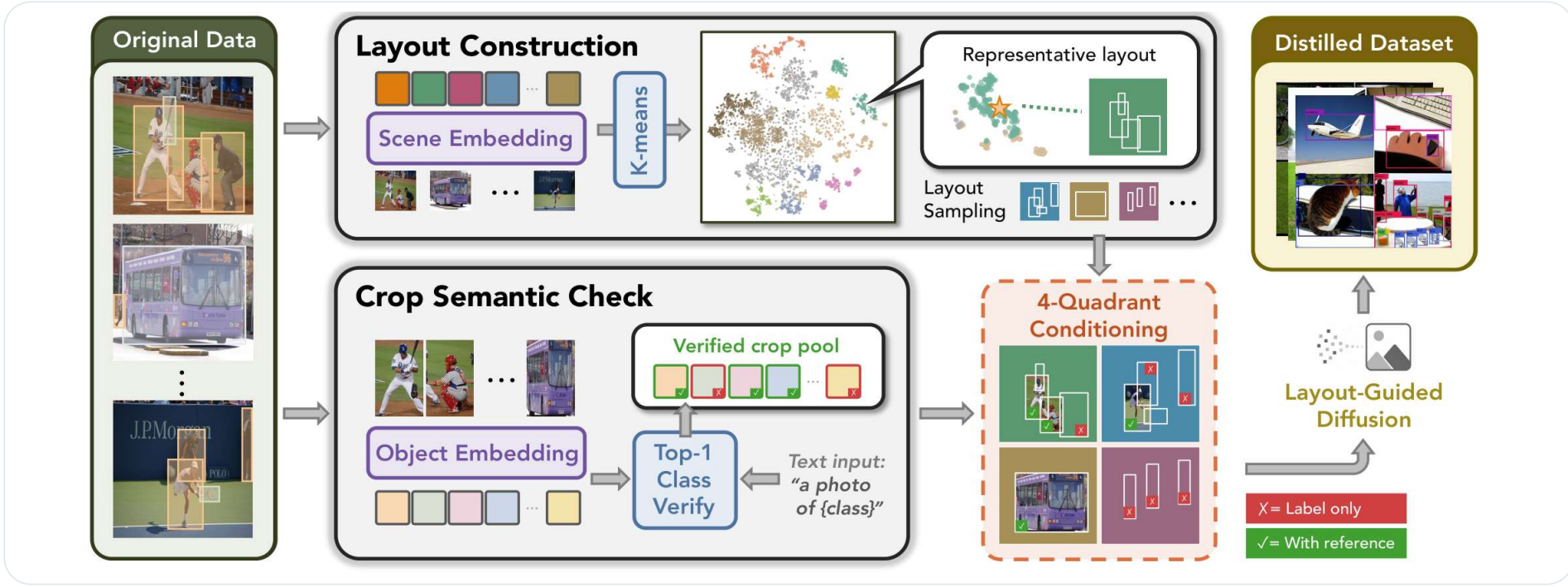
The problem

Similar objects and scenes repeat within and across images.

The insight

Layouts define detection targets; their appearances can vary.

LCD³ preserves layouts while diversifying appearances



01 Scene-diverse layouts

Combine layouts from different scene clusters.

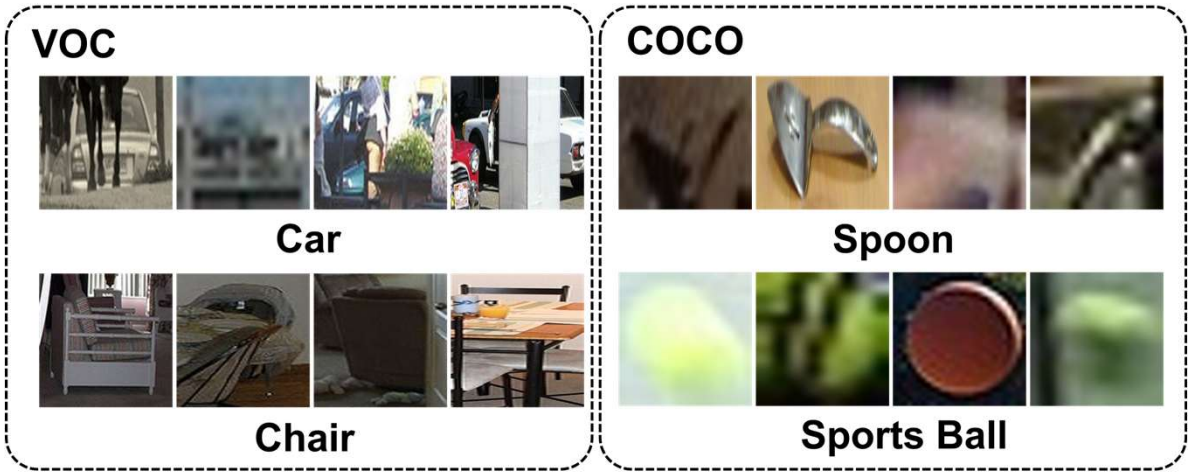
02 Crop verification

Keep reliable crops as appearance references.

03 Guided diffusion

Render new appearances using layouts and verified crops.

Semantic verification filters unreliable appearance references



What is rejected?

Occluded, ambiguous or low-detail object crops

What is kept?

Class labels and bounding boxes

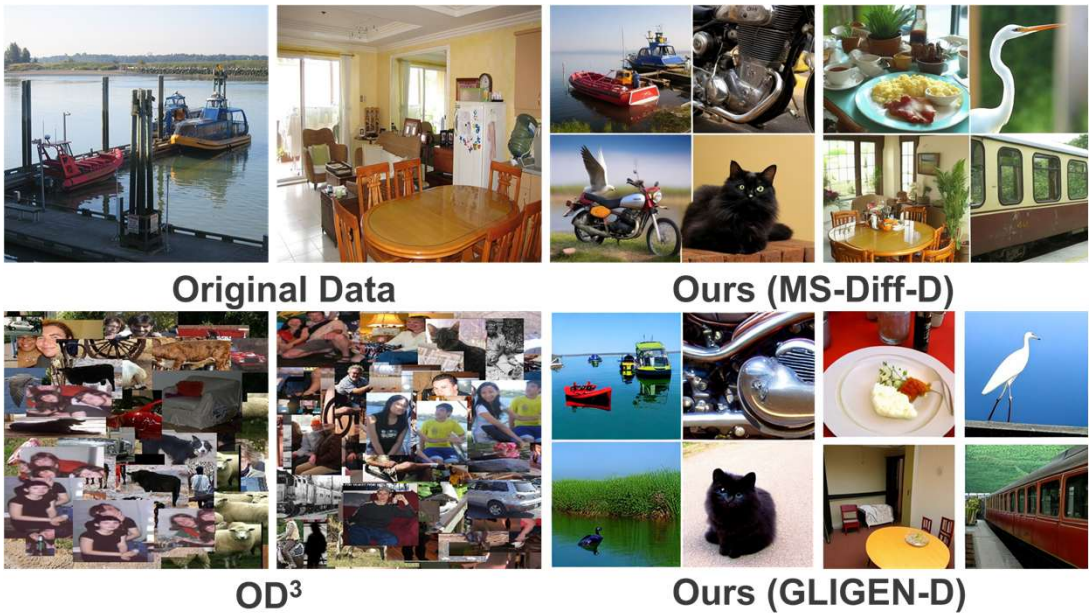
Rejected crops are not used as visual references; their objects are still generated from layout and label conditions.

LCD³ improves detection performance on MS COCO

Label	Method	IPD = 0.25% mAP%	IPD = 0.25% mAP_{50}%	IPD = 0.25% mAP_{75}%	IPD = 0.5% mAP%	IPD = 0.5% mAP_{50}%	IPD = 0.5% mAP_{75}%	IPD = 1.0% mAP%	IPD = 1.0% mAP_{50}%	IPD = 1.0% mAP_{75}%
Hard	Random	$0.5_{\pm 0.1}$	$1.7_{\pm 0.3}$	–	$3.7_{\pm 0.2}$	$10.1_{\pm 0.3}$	–	$7.2_{\pm 0.8}$	$17.3_{\pm 0.9}$	–
	Uniform	$0.8_{\pm 0.2}$	$2.4_{\pm 0.5}$	–	$3.4_{\pm 0.4}$	$9.5_{\pm 0.6}$	–	$7.4_{\pm 0.5}$	$17.6_{\pm 0.5}$	–
	K-Center	$0.4_{\pm 0.1}$	$1.5_{\pm 0.2}$	–	$3.2_{\pm 0.5}$	$9.3_{\pm 0.5}$	–	$6.1_{\pm 0.3}$	$15.4_{\pm 0.6}$	–
	Herding	$0.5_{\pm 0.1}$	$1.8_{\pm 0.4}$	–	$3.5_{\pm 0.3}$	$9.7_{\pm 0.3}$	–	$6.7_{\pm 0.4}$	$16.3_{\pm 0.7}$	–
	DCOD	–	–	–	$3.7_{\pm 0.3}$	$8.4_{\pm 0.2}$	–	$6.1_{\pm 0.4}$	$13.5_{\pm 0.3}$	–
	UniDD	$4.5_{\pm 0.3}$	$10.3_{\pm 0.4}$	–	$7.1_{\pm 0.4}$	$16.9_{\pm 0.3}$	–	$10.8_{\pm 0.4}$	$22.5_{\pm 0.5}$	–
	OD^3	$5.5_{\pm 0.3}$	$12.9_{\pm 0.5}$	$3.9_{\pm 0.2}$	$8.8_{\pm 0.1}$	$19.3_{\pm 0.3}$	$7.0_{\pm 0.3}$	$11.2_{\pm 0.3}$	$23.4_{\pm 0.3}$	$9.6_{\pm 0.3}$
	LCD^3 (GLIGEN-C)	$6.4_{\pm 0.5}$	$13.9_{\pm 0.9}$	$5.2_{\pm 0.4}$	$9.2_{\pm 0.4}$	$18.3_{\pm 0.7}$	$8.3_{\pm 0.4}$	$11.2_{\pm 0.5}$	$20.2_{\pm 0.7}$	$11.1_{\pm 0.5}$
	LCD^3 (GLIGEN-D)	$6.6_{\pm 0.7}$	$14.3_{\pm 1.5}$	$5.5_{\pm 0.6}$	$9.5_{\pm 0.3}$	$19.0_{\pm 0.6}$	$8.6_{\pm 0.4}$	$10.8_{\pm 0.3}$	$20.4_{\pm 0.6}$	$10.2_{\pm 0.4}$
	LCD^3 (MS-Diff-C)	$6.8_{\pm 0.3}$	$15.8_{\pm 0.6}$	$4.9_{\pm 0.3}$	$9.5_{\pm 0.2}$	$20.0_{\pm 0.2}$	$7.8_{\pm 0.3}$	$12.5_{\pm 0.1}$	$\mathbf{24.8}_{\pm 0.3}$	$\mathbf{11.5}_{\pm 0.2}$
	LCD^3 (MS-Diff-D)	$\mathbf{8.7}_{\pm 0.3}$	$\mathbf{16.8}_{\pm 0.6}$	$\mathbf{6.8}_{\pm 0.3}$	$\mathbf{10.4}_{\pm 0.3}$	$\mathbf{20.6}_{\pm 0.4}$	$\mathbf{9.6}_{\pm 0.4}$	$\mathbf{12.6}_{\pm 0.1}$	$24.7_{\pm 0.1}$	$11.3_{\pm 0.2}$
Soft	OD^3	$11.9_{\pm 0.3}$	$23.4_{\pm 0.8}$	$10.8_{\pm 0.3}$	$17.8_{\pm 0.2}$	$33.2_{\pm 0.2}$	$17.1_{\pm 0.1}$	$21.5_{\pm 0.3}$	$38.9_{\pm 0.3}$	$21.3_{\pm 0.3}$
	LCD^3 (GLIGEN-C)	$14.2_{\pm 0.3}$	$26.2_{\pm 0.4}$	$14.0_{\pm 0.3}$	$18.6_{\pm 0.2}$	$32.2_{\pm 0.4}$	$19.0_{\pm 0.2}$	$23.6_{\pm 0.1}$	$38.0_{\pm 0.2}$	$25.0_{\pm 0.2}$
	LCD^3 (GLIGEN-D)	$14.0_{\pm 0.2}$	$25.7_{\pm 0.4}$	$13.7_{\pm 0.2}$	$18.8_{\pm 0.2}$	$32.8_{\pm 0.5}$	$19.1_{\pm 0.3}$	$23.7_{\pm 0.1}$	$38.1_{\pm 0.1}$	$\mathbf{25.2}_{\pm 0.1}$
	LCD^3 (MS-Diff-C)	$14.7_{\pm 0.1}$	$27.5_{\pm 0.2}$	$14.2_{\pm 0.1}$	$19.3_{\pm 0.3}$	$34.1_{\pm 0.4}$	$19.6_{\pm 0.1}$	$23.6_{\pm 0.2}$	$39.7_{\pm 0.2}$	$24.8_{\pm 0.2}$
	LCD^3 (MS-Diff-D)	$\mathbf{16.3}_{\pm 0.2}$	$\mathbf{29.1}_{\pm 0.3}$	$\mathbf{16.5}_{\pm 0.4}$	$\mathbf{21.1}_{\pm 0.3}$	$\mathbf{35.9}_{\pm 0.4}$	$\mathbf{22.1}_{\pm 0.3}$	$\mathbf{23.8}_{\pm 0.2}$	$\mathbf{40.1}_{\pm 0.3}$	$24.9_{\pm 0.1}$
Whole Dataset (Hard)		38.4	59.0	42.0	38.4	59.0	42.0	38.4	59.0	42.0

LCD³ improves mAP_{50} by +5.7 points over OD³ at 0.25% IPD

LCD³ generates more diverse images with fewer objects



Objects / image

86.7 → **12.0**

OD³ LCD³

Object similarity ↓

0.2279 → **0.2164**

Fewer objects, lower visual redundancy.

Takeaway: LCD³ preserves detection layouts, filters unreliable references, and generates more diverse training images.