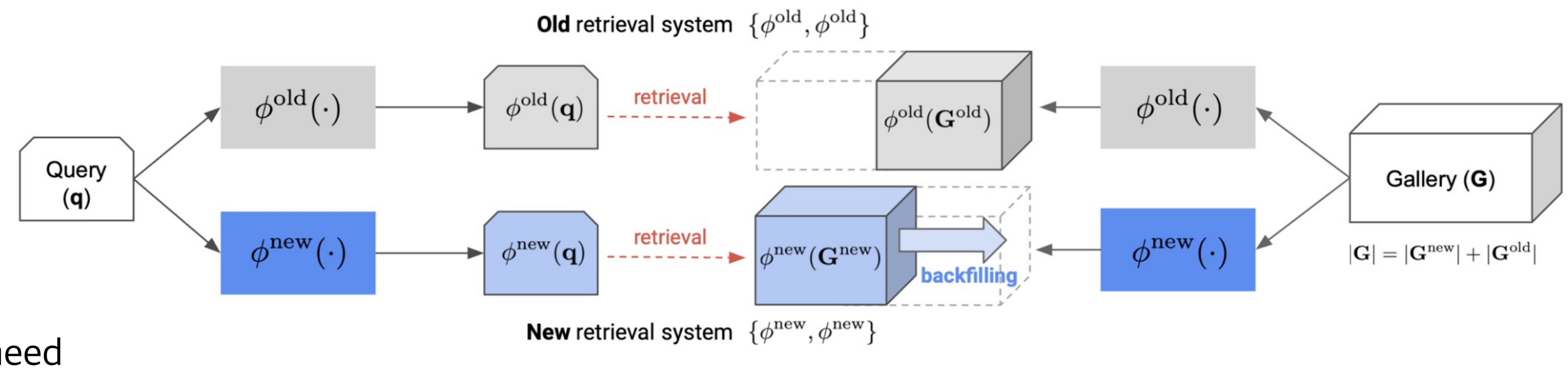
Online Backfilling with No Regret for Large-Scale Image Retrieval

Seonguk Seo[1] Gokhan Uzunbas[2] Bohyung Han[1] Sara Cao[2] Joena Zhang[2] Taipeng Tian[2] Ser-Nam Lim[2]

1 Computer Vision Lab, Seoul National University 2 Meta



Introduction

- **[Motivation]** Backward-compatible training avoids the need for updating all gallery images during model upgrade, but it achieves feature compatibility at the expense of feature discriminability, resulting in sub-optimal performance.
- **[Contribution]** To resolve this compatibility-discriminability dilemma, we relax the backfill-free constraint and propose a novel online backfilling algorithm to alleviate the bottleneck.

Preliminaries

- **Image retrieval** evaluates the retrieval accuracy as

$$\mathcal{M}(\phi(\mathbf{Q}), \phi(\mathbf{G}))$$

- **Backward-compatible training** aim to learn a new model while its feature space being compatible to those of old model. Backward compatibility is achieved when

$$\mathcal{M}(\phi^{\text{new}}(\mathbf{Q}), \phi^{\text{old}}(\mathbf{G})) > \mathcal{M}(\phi^{\text{old}}(\mathbf{Q}), \phi^{\text{old}}(\mathbf{G}))$$

Rank Merge

- Assume that the first M out of a total of N images are backfilled. We can first conduct image retrieval using the individual retrieval systems independently as

$$\mathbf{g}_m = \underset{\mathbf{g}_i \in \mathbf{G}^{\text{old}}}{\arg\min}\ \text{dist}\left(\phi^{\text{old}}(\mathbf{q}), \phi^{\text{old}}(\mathbf{g}_i)\right), \quad \mathbf{g}_n = \underset{\mathbf{g}_j \in \mathbf{G}^{\text{new}}}{\arg\min}\ \text{dist}\left(\phi^{\text{new}}(\mathbf{q}), \phi^{\text{new}}(\mathbf{g}_j)\right)$$

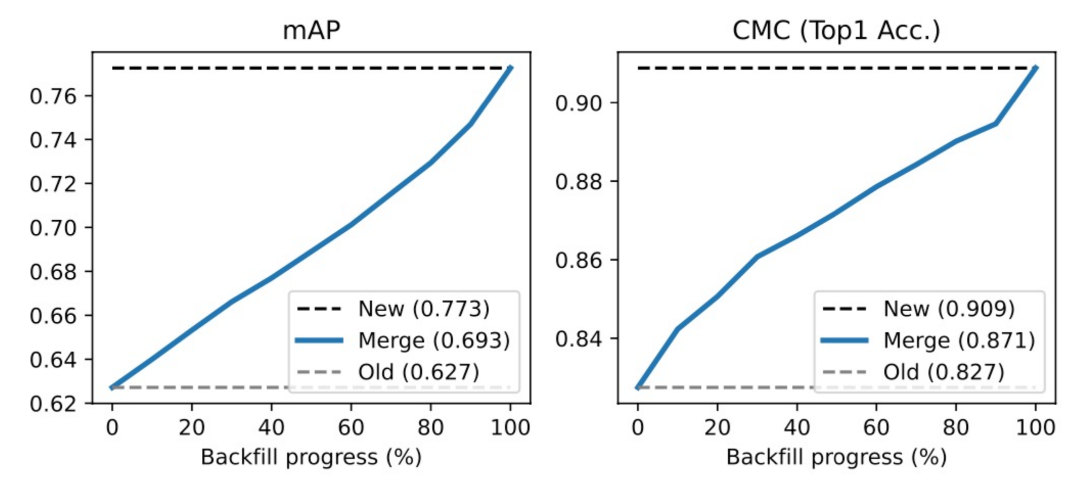
and finally select $\mathbf{g}_m$ if $\text{dist}(\phi^{\text{old}}(\mathbf{q}), \phi^{\text{old}}(\mathbf{g}_m)) < \text{dist}(\phi^{\text{new}}(\mathbf{q}), \phi^{\text{new}}(\mathbf{g}_n))$ and $\mathbf{g}_n$ otherwise.

- The retrieval performance after rank merge during backfilling is given by

$$\mathcal{M}_t := \mathcal{M}(\{\phi^{\text{old}}(\mathbf{Q}), \phi^{\text{new}}(\mathbf{Q})\}, \{\phi^{\text{old}}(\mathbf{G}_t^{\text{old}}), \phi^{\text{new}}(\mathbf{G}_t^{\text{new}})\}),$$

and should satisfy:

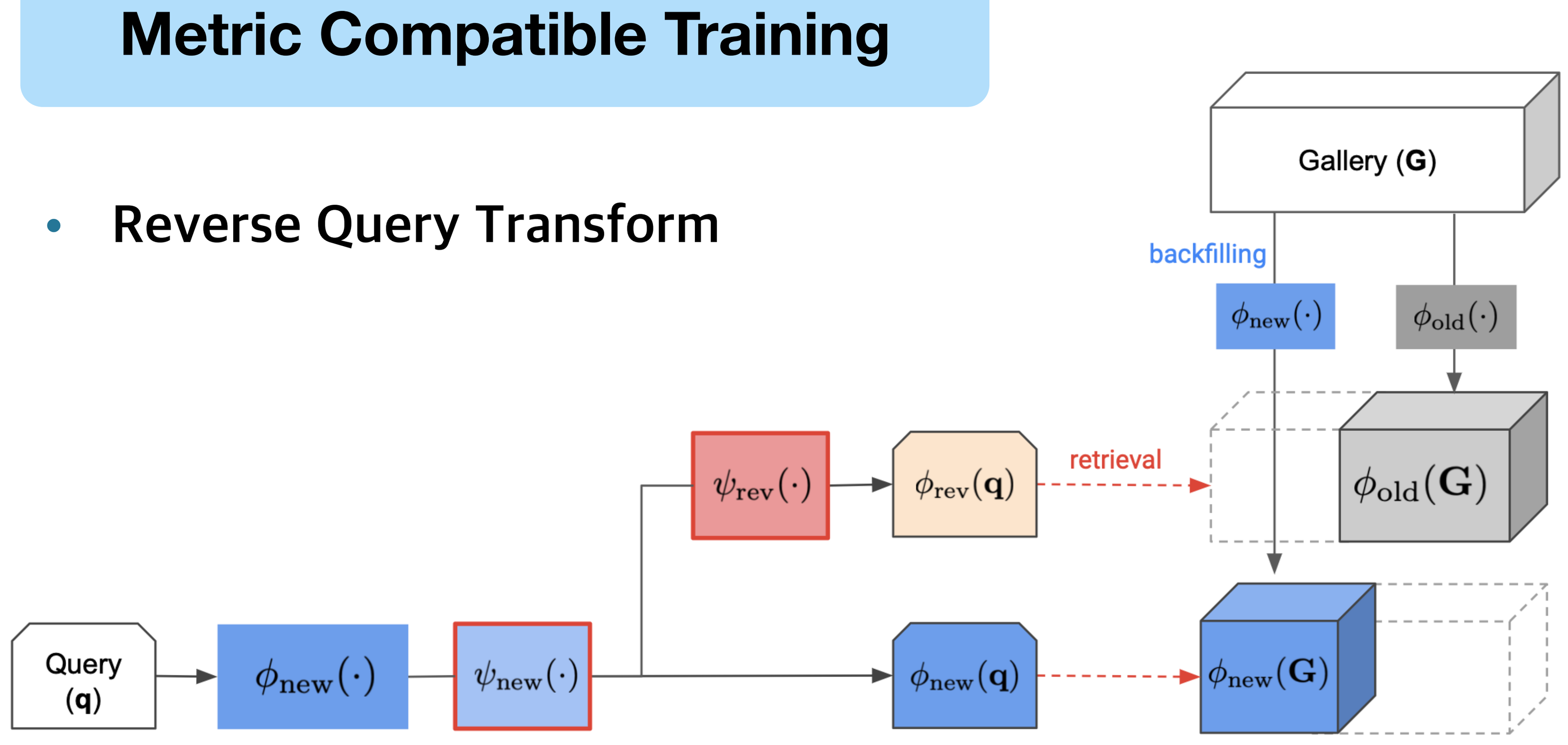
$$\mathcal{M}_0 \geq \mathcal{M}(\phi^{\text{old}}(\mathbf{Q}), \phi^{\text{old}}(\mathbf{G})),$$
$$\mathcal{M}_1 \geq \mathcal{M}(\phi^{\text{new}}(\mathbf{Q}), \phi^{\text{new}}(\mathbf{G})),$$
$$\mathcal{M}_{t_1} \geq \mathcal{M}_{t_2} \ \text{ if } t_1 \geq t_2.$$



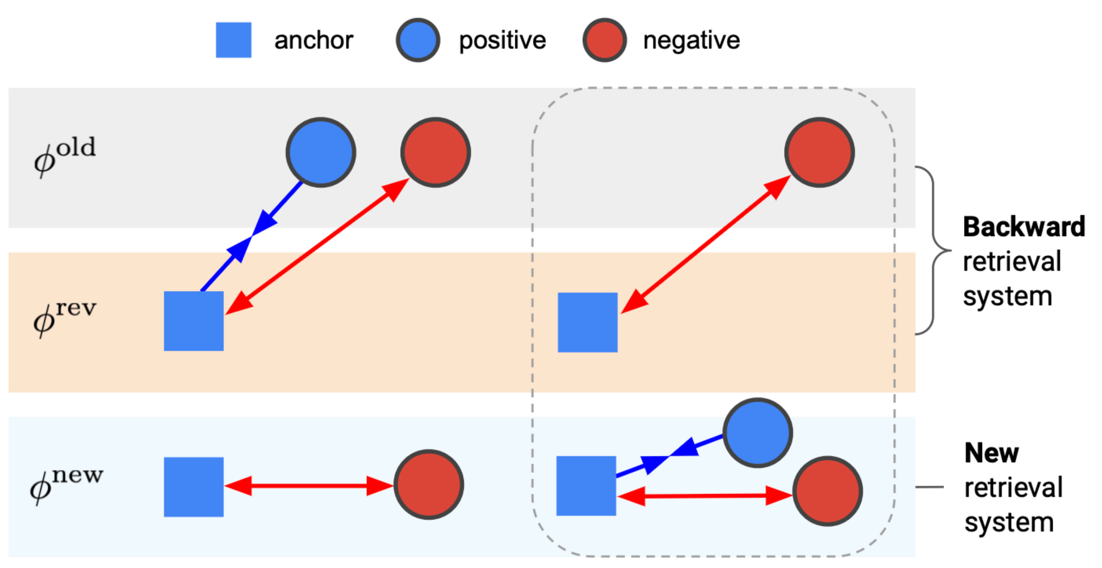
Backward compatible training leads to suboptimal solutions in principle.
Online backfilling can address the compatibility-discriminability dilemma effectively.

Metric Compatible Training

- **Reverse Query Transform**



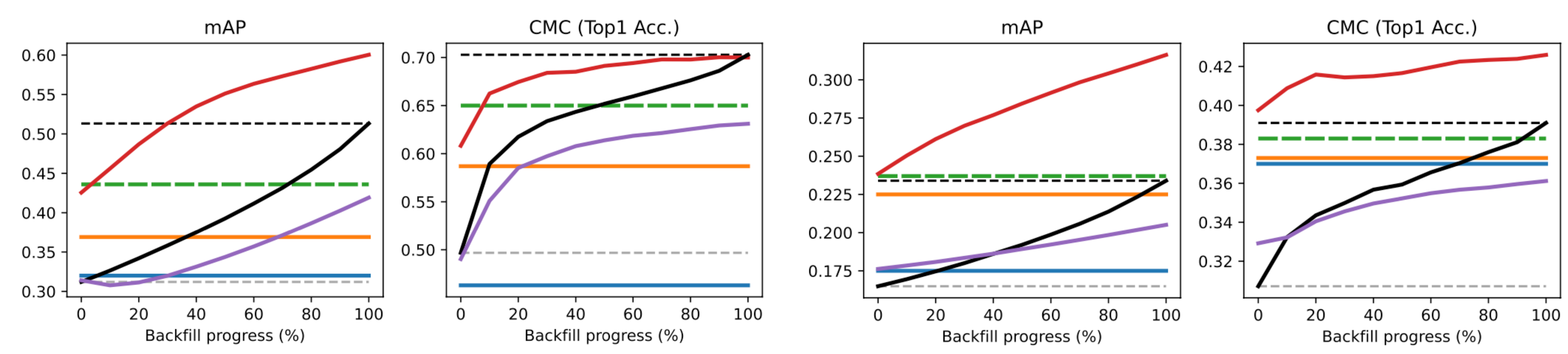
- **Metric Compatible Contrastive Learning**



$$\mathcal{L}_{\text{MCL}}(\mathbf{x}_i, y_i) = -\log \frac{\sum_{y_k=y_i} s_{ik}^{\text{old}}}{\sum_{y_k=y_i} s_{ik}^{\text{old}} + \sum_{y_k\neq y_i} s_{ik}^{\text{old}} + \sum_{y_k\neq y_i} s_{ik}^{\text{new}}} - \log \frac{\sum_{y_k=y_i} s_{ik}^{\text{new}}}{\sum_{y_k=y_i} s_{ik}^{\text{new}} + \sum_{y_k\neq y_i} s_{ik}^{\text{new}} + \sum_{y_k\neq y_i} s_{ik}^{\text{old}}},$$

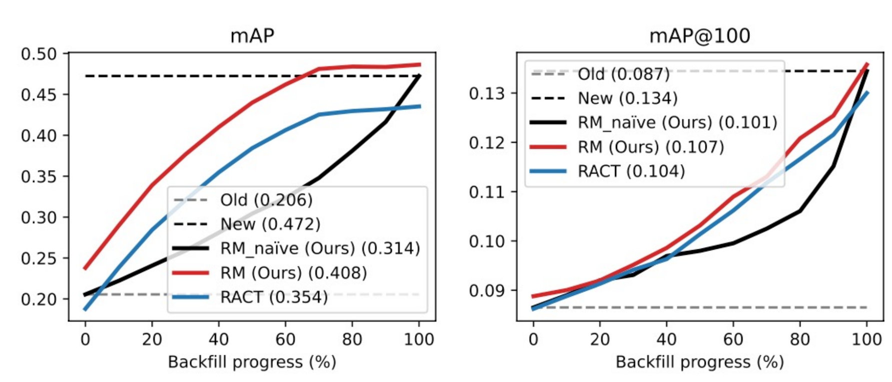
$$s_{ij}^{\text{new}} = \exp\left(-\text{dist}\left(\rho^{\text{new}}(\mathbf{x}_i), \rho^{\text{new}}(\mathbf{x}_j)\right)\right)$$
$$s_{ij}^{\text{old}} = \exp\left(-\text{dist}\left(\rho^{\text{rev}}(\mathbf{x}_i), \phi^{\text{old}}(\mathbf{x}_j)\right)\right)$$

Experiments



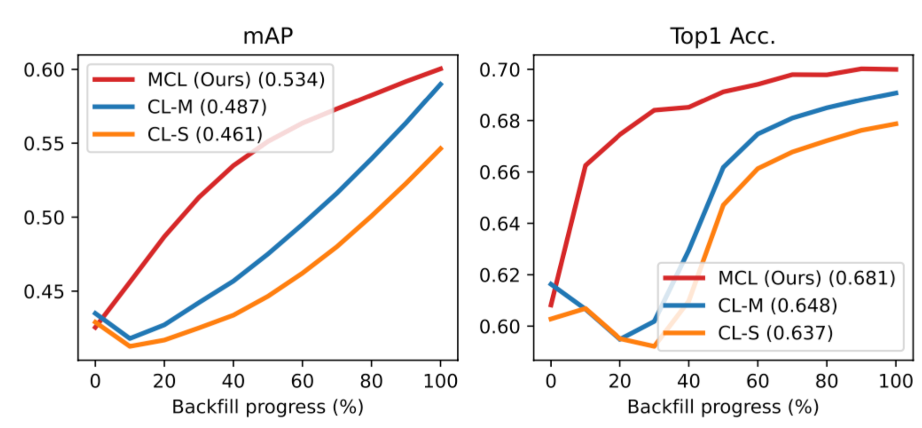
	ImageNet-1K			CIFAR-100			Places-365			Market-1501		
	AUC_{mAP}	AUC_{CMC}	Gain	AUC_{mAP}	AUC_{CMC}	Gain	AUC_{mAP}	AUC_{CMC}	Gain	AUC_{mAP}	AUC_{CMC}	Gain
Old	31.2	49.7	0%	21.6	34.3	0%	16.5	30.7	0%	62.7	82.7	0%
New	51.3	70.3	100%	47.4	62.6	100%	23.4	39.1	100%	77.3	90.9	100%
$\text{RM}_{\text{naïve}}$ (Ours)	40.0	63.9	44%	30.8	49.1	36%	19.5	35.8	43%	69.2	87.0	45%
BCT	32.0	46.3	4%	26.4	43.5	19%	17.5	37.0	14%	66.6	84.3	27%
FCT	36.9	58.7	28%	27.1	49.4	21%	22.5	37.3	87%	66.4	84.2	25%
FCT (w/ side-info)	43.6	65.0	62%	37.0	53.9	60%	23.7	38.3	104%	66.4	84.4	25%
BiCT	35.1	59.7	19%	29.0	48.3	29%	19.0	34.9	36%	65.0	82.4	16%
RM (Ours)	**53.4**	**68.1**	**110%**	**41.4**	**60.7**	**78%**	**28.2**	**41.7**	**170%**	**70.7**	**87.6**	**55%**

- Open-class setting



(a) CIFAR-100 (b) GLDv2-test

- Ablation study (MCL)



$$\mathcal{L}_{\text{CL-M}}(\mathbf{x}_i, y_i) = -\log \frac{\sum_{y_k=y_i} s_{ik}^{\text{old}}}{\sum_{y_k=y_i} s_{ik}^{\text{old}} + \sum_{y_k\neq y_i} s_{ik}^{\text{old}}} - \log \frac{\sum_{y_k=y_i} s_{ik}^{\text{new}}}{\sum_{y_k=y_i} s_{ik}^{\text{new}} + \sum_{y_k\neq y_i} s_{ik}^{\text{new}}}$$

Selected References

[BCT] Shen et al., "Towards backward-compatible representation learning", CVPR (2020)

[FCT] Ramanujan et al., "Forward compatible training for large-scale embedding retrieval systems", CVPR (2022)