Metric Compatible Training for Online Backfilling in Large-Scale Retrieval

Seongukk Seo [1] **Mustafa Gokhan Uzunbas** [2] **Bohyung Han** [1 3]
Sara Cao [2] **Joena Zhang** [2] **Taipeng Tian** [2] **Ser-Nam Lim** [2]

Abstract

In large-scale retrieval systems, model upgrades require re-extracting all gallery embeddings from upgraded models. However, it inevitably spends a prohibitively large amount of computational cost and even entails the downtime of the service. To alleviate this bottleneck, backward-compatible learning is proposed to learn feature space of new model while being compatible with those of old model. Although it sidesteps this challenge by tackling query-side representations, this leads to suboptimal solutions in principle because gallery embeddings cannot benefit from model upgrades. We address this dilemma by introducing an online backfilling algorithm, which enables us to achieve a progressive performance improvement during the backfilling process without sacrificing the full performance of the new model after the completion of backfilling. To this end, we first show that a simple distance rank merge is a reasonable option for online backfilling. Then, we incorporate a reverse transformation module and metric-compatible contrastive learning, resulting in desirable merge results during backfilling with no extra overhead.

1. Introduction

Image retrieval models (Sun et al., 2014; Li et al., 2014; Wang et al., 2018a) have achieved remarkable performance by adopting deep neural networks for representing images. Yet, all models need to be upgraded at times to take advantage of improvements in training datasets, network architectures, and training techniques. This unavoidably leads to the need for re-extracting the features from millions or even billions of gallery images using the upgraded new model. This process, called *backfilling* or *re-indexing*, needs to be completed before the retrieval system can benefit from the new model, which may take months in practice.

To sidestep this bottleneck, several backfilling-free approaches based on backward-compatible learning (Shen et al., 2020; Meng et al., 2021; Su et al., 2022; Wan et al., 2022; Duggal et al., 2021) have been proposed. They learn a new model while ensuring that its feature space is still compatible with the old one, thus avoiding the need for updating old gallery features. Although these approaches have achieved substantial performance gains without backfilling, they achieve feature compatibility at the expense of feature discriminability and their performance is suboptimal. We argue that backward-compatible learning is not a fundamental solution and backfilling is still essential to accomplish state-of-the-art performance without performance sacrifices.

To resolve this compatibility-discriminability dilemma, we relax the backfill-free constraint and propose a novel online backfilling algorithm equipped with three technical components. We posit that an online backfilling technique needs to satisfy three essential conditions: 1) immediate deployment after the completion of model upgrade, 2) progressive and non-trivial performance gains in the middle of backfilling, and 3) no degradation of final performance compared to offline backfilling. In this paper, we proposed a fundamental distance rank merge framework and metric compatible training approach to satisfy this constraints. The main contributions of our work are summarized as follows.

- We propose an online backfilling approach, a fundamental solution for model upgrades in image retrieval systems, based on distance rank merge to overcome the compatibility-discriminability dilemma in existing compatible learning methods.
- We incorporate a reverse query transform module to make it compatible with both the old and new galleries while computing the feature extraction of query only once in the middle of the backfilling process.
- We adopt a metric-compatible learning technique to make the merge process robust by calibrating distances in the feature embedding spaces of old and new models.

[1]ECE, Seoul National University [2]Meta AI [3]IPAI, Seoul National University. Correspondence to: Ser-Nam Lim <ser-nam@meta.com>.

In *ICML Workshop on Localized Learning (LLW)*, Honolulu, Hawaii, USA. 2023.

2. Image Retrieval by Rank Merge

2.1. Overview

Our goal is to develop a fundamental solution to overcome the compatibility-discriminability trade-off via online backfilling in compatible model upgrade. We aim to remove inherent limitations of backfill-free backward-compatible learning—the inability to benefit from upgraded representations of gallery images extracted from new models—while avoiding system downtime and outdated service during offline backfilling process until the completion of backfilling.

Image retrieval systems with online backfilling should satisfy the following three conditions:

1. The system can be deployed as soon as the model upgrade is completed.
2. The performance should monotonically increase without negative flips[1] as backfill progresses.
3. The final performance should be better than or equivalent to the algorithm relying on offline backfilling.

We present a distance rank merge approach for image retrieval, which enables online backfilling in arbitrary model upgrade scenarios. This method maintains two separate retrieval pipelines corresponding to the old and new models and merges the retrieval results from the two models based on distances from a query embedding. This allows us to run the retrieval system without a warm-up period and achieve surprisingly good results during the backfilling process. Note that the old and new models are not required to be compatible at this moment but we will make them so to further improve performance in the subsequent sections.

2.2. Formulation

Let $\mathbf{q} \in \mathbf{Q}$ be a query image and $\mathbf{G} = \{\mathbf{g}_1, ..., \mathbf{g}_N\}$ be a gallery composed of N images. An embedding network $\phi(\cdot)$ projects an image onto a learned feature space. To retrieve the closest gallery image given a query, we solve $\arg\min_{\mathbf{g}\in\mathbf{G}} \text{dist}\left(\phi(\mathbf{q}), \phi(\mathbf{g})\right)$, where $\text{dist}(\cdot,\cdot)$ is a distance metric. Following (Shen et al., 2020), the retrieval accuracy is given by

$$\mathcal{M}(\phi(\mathbf{Q}), \phi(\mathbf{G})), \tag{1}$$

where $\mathcal{M}(\cdot,\cdot)$ is an evaluation metric such as mAP or CMC.

Backward compatibility Denote the old and new embedding networks by $\phi^{\text{old}}(\cdot)$ and $\phi^{\text{new}}(\cdot)$ respectively. If $\phi^{\text{new}}(\cdot)$ is backward compatible with $\phi^{\text{old}}(\cdot)$, then we can perform search on a set of old gallery embeddings using a new query embedding, *i.e.*, $\arg\min_{\mathbf{g}\in\mathbf{G}} \text{dist}(\phi^{\text{new}}(\mathbf{q}), \phi^{\text{old}}(\mathbf{g}))$. As stated in (Shen et al., 2020), the backward compatibility is achieved when the following criterion is satisfied:

$$\mathcal{M}(\phi^{\text{new}}(\mathbf{Q}), \phi^{\text{old}}(\mathbf{G})) > \mathcal{M}(\phi^{\text{old}}(\mathbf{Q}), \phi^{\text{old}}(\mathbf{G})). \tag{2}$$

From now, we refer to a pair of embedding networks for query and gallery as a retrieval system, *e.g.*, $\{\phi^{(\cdot)}, \phi^{(\cdot)}\}$.

Rank merge Assume that the first M out of a total of N images are backfilled, *i.e.*, $\mathbf{G}^{\text{new}} = \{\mathbf{g}_1, ..., \mathbf{g}_M\}$ and $\mathbf{G}^{\text{old}} = \{\mathbf{g}_{M+1}, ..., \mathbf{g}_N\}$. Note that the total number of stored gallery embeddings is fixed to N during the backfilling process, *i.e.*, $\mathbf{G}^{\text{old}} = \mathbf{G} - \mathbf{G}^{\text{new}}$. Then, we first conduct image retrieval using the individual retrieval systems, $\{\phi^{\text{old}}, \phi^{\text{old}}\}$ and $\{\phi^{\text{new}}, \phi^{\text{new}}\}$, independently as

$$\mathbf{g}_m = \underset{\mathbf{g}_i\in\mathbf{G}^{\text{old}}}{\arg\min}\ \text{dist}\left(\phi^{\text{old}}(\mathbf{q}), \phi^{\text{old}}(\mathbf{g}_i)\right), \tag{3}$$

$$\mathbf{g}_n = \underset{\mathbf{g}_j\in\mathbf{G}^{\text{new}}}{\arg\min}\ \text{dist}\left(\phi^{\text{new}}(\mathbf{q}), \phi^{\text{new}}(\mathbf{g}_j)\right). \tag{4}$$

For each query image $\mathbf{q}$, we 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

$$\mathcal{M}_t := \mathcal{M}(\{\phi^{\text{old}}(\mathbf{Q}), \phi^{\text{new}}(\mathbf{Q})\}, \{\phi^{\text{old}}(\mathbf{G}_t^{\text{old}}), \phi^{\text{new}}(\mathbf{G}_t^{\text{new}})\}), \tag{5}$$

where $t \in [0, 1]$ indicates the rate of backfilling completion, *i.e.*, $|\mathbf{G}_t^{\text{new}}| = t|\mathbf{G}|$ and $|\mathbf{G}_t^{\text{old}}| = (1-t)|\mathbf{G}|$. The criteria discussed in Section 2.1 are formally defined as

$$\mathcal{M}_0 \geq \mathcal{M}(\phi^{\text{old}}(\mathbf{Q}), \phi^{\text{old}}(\mathbf{G})), \tag{6}$$

$$\mathcal{M}_1 \geq \mathcal{M}(\phi^{\text{new}}(\mathbf{Q}), \phi^{\text{new}}(\mathbf{G})), \tag{7}$$

$$\mathcal{M}_{t_1} \geq \mathcal{M}_{t_2} \ \text{ if } \ t_1 \geq t_2. \tag{8}$$

Comprehensive evaluation To measure both backfilling cost and model performance comprehensively during online backfilling, we utilize the following metrics that calculate the area under mAP or CMC curves as

$$\text{AUC}_{\text{mAP}} = \int_0^1 \text{mAP}_t dt \ \text{ and } \ \text{AUC}_{\text{CMC}} = \int_0^1 \text{CMC}_t dt.$$

3. Metric Compatible Training

3.1. Reverse Query Transform

One may argue that the proposed approach is computationally expensive at inference time because we need to conduct feature extraction twice per query for both the old and new models. To reduce the computational cost incurred by computing query embeddings twice for two separate systems at

[1]The "negative flip" refers to performance degradation caused by incorrect retrievals in the new model, which were correct in the old model.

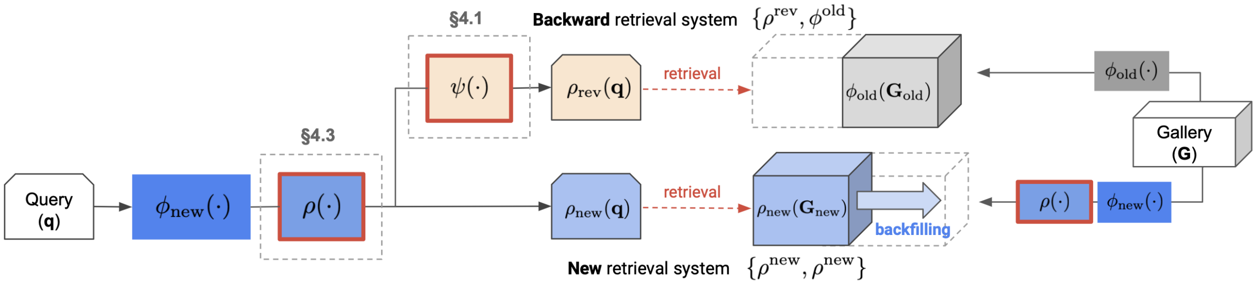

Figure 1: Image retrieval with our final rank merge framework including Section 3.1-3.3. Backward retrieval system consists of reversely transformed new query and old gallery, $\{\rho^{\text{rev}}, \phi^{\text{old}}\}$. The final image retrieval results are given by merging the outputs from $\{\rho^{\text{rev}}, \phi^{\text{old}}\}$ and $\{\rho^{\text{new}}, \rho^{\text{new}}\}$.

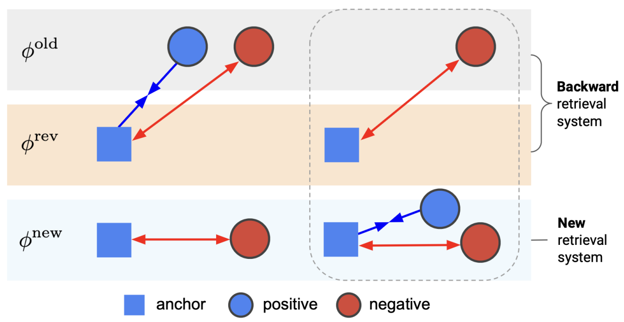

Figure 2: Illustration of metric compatible learning loss with two systems. Two boxes with dotted lines corresponds to two terms in equation 11. For each retrieval system, the distances between positive pairs are learned to be both smaller than those of negative pairs in the two systems.

the inference stage, we compute the embedding using the new model and transform it to the version compatible with the old model through the reverse query transform module. To establish such a mechanism, we fix the parameters of the old and new models $\{\phi^{\text{old}}, \phi^{\text{new}}\}$ after training them independently, and train a lightweight network, $\psi(\cdot)$, which transforms the embedding in the new model to the one in the old model. For each training example $\mathbf{x}$, our objective is minimizing the following loss:

$$\mathcal{L}_{\text{RQT}}(\mathbf{x}) := \text{dist}\left(\psi\left(\phi^{\text{new}}(\mathbf{x})\right), \phi^{\text{old}}(\mathbf{x})\right), \tag{9}$$

where $\text{dist}(\cdot,\cdot)$ is a distance metric such as ℓ_2 or cosine distances. Because we only update the parameters in $\psi(\cdot)$, not the ones in $\phi^{\text{new}}(\cdot)$ or $\phi^{\text{old}}(\cdot)$, we can still access the representations given by the new model at no cost even after the optimization of $\psi(\cdot)$. Note that this reverse query transform module differs from FCT (Ramanujan et al., 2022) mainly in terms of transformation direction and requirement of side information. FCT performs a transformation from the old representation to the new, while the opposite is true for our proposed approach. Since the embedding quality of a new model is highly likely to be better than that of an old one, our reverse transformation module performs well even without additional side information and, consequently, is more practical and efficient.

Integration into baseline retrieval system Figure 1 illustrates the distance rank merge process together with the proposed reverse transformation module. The whole procedure consists of two retrieval systems defined by a pair of query and gallery representations, backward retrieval system $\{\phi^{\text{rev}}, \phi^{\text{old}}\}$ and new retrieval system $\{\phi^{\text{new}}, \phi^{\text{new}}\}$, where $\phi^{\text{rev}} := \psi(\phi^{\text{new}})$. Note that we obtain both the new and compatible query embeddings, $\phi^{\text{new}}(\mathbf{q})$ and $\phi^{\text{rev}}(\mathbf{q}) = \psi(\phi^{\text{new}}(\mathbf{q}))$, using a shared feature extractor, $\phi^{\text{new}}(\cdot)$.

The entire image retrieval pipeline consists of two parts: 1) feature extraction of a query image and 2) search for the nearest image in a gallery from the query. Compared to the image retrieval based on a single model, the computational cost of the proposed model with rank merge requires negligible additional cost which corresponds to feature transformation $\psi(\cdot)$ in the first part. Note that the number of total gallery embeddings is fixed, *i.e.*, $|\mathbf{G}^{\text{new}}| + |\mathbf{G}^{\text{old}}| = |\mathbf{G}|$, so the cost of the second part is almost the same in both cases.

3.2. Metric Compatible Contrastive Learning

While the rank merge with the basic reverse transformation works well, the objective in equation 9 cares about the positive pairs ϕ^{old} and ϕ^{rev} with no consideration of negative ones, which sometimes lead to misranked position. To handle this issue, we employ a supervised contrastive learning loss (Oord et al., 2018; Khosla et al., 2020) to consider both positive and negative pairs as

$$\mathcal{L}_{\text{CL}}(\mathbf{x}_i, y_i) = -\log \frac{\sum_{y_k = y_i} s_{ik}^{\text{old}}}{\sum_{y_k = y_i} s_{ik}^{\text{old}} + \sum_{y_k \neq y_i} s_{ik}^{\text{old}}}, \tag{10}$$

where $s_{ij}^{\text{old}} = \exp\left(-\text{dist}\left(\phi^{\text{rev}}(\mathbf{x}_i), \phi^{\text{old}}(\mathbf{x}_j)\right)\right)$ and y_i denotes the class membership of the i^{th} sample. For more robust contrastive training, we perform hard example mining for both the positive and negative pairs[2]. Such a contrastive learning approach facilitates distance calibration and improves feature discrimination because it promotes

[2]For each anchor, we select the half of the examples in each of positive and negative labels based on the distances from anchors.

Table 1: Comparison with existing compatible learning methods on four standard benchmarks in homogeneous model upgrades. *Gain* denotes relative gain that each method achieves from old model in terms of AUC_{mAP}, compared to the gain of new model. Note that $\text{RM}_{\text{naïve}}$ indicates the basic version of distance rank merge described in Sec. 2.2 and that *Old* and *New* denote embedding models of gallery images.

	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%
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%**

separation of the positive and negative examples.

Now, although the distances within the backward retrieval system become more comparable, they are still not properly calibrated in terms of the distances in the new retrieval system. Considering distances in both retrieval systems jointly when we train the reverse transformation module, we can obtain more comparable distances and consequently achieve more reliable rank merge results. From this perspective, we propose a metric compatible contrastive learning loss as

$$\mathcal{L}_{\text{MCL}}(\mathbf{x}_i, y_i) = \tag{11}$$
$$-\log \frac{\sum_{y_k=y_i} s_{ik}^{\text{old}}}{\sum_{y_k=y_i} s_{ik}^{\text{old}} + \sum_{y_k\neq y_i} s_{ik}^{\text{old}} + \sum_{y_k\neq y_i} s_{ik}^{\text{new}}}$$
$$-\log \frac{\sum_{y_k=y_i} s_{ik}^{\text{new}}}{\sum_{y_k=y_i} s_{ik}^{\text{new}} + \sum_{y_k\neq y_i} s_{ik}^{\text{new}} + \sum_{y_k\neq y_i} s_{ik}^{\text{old}}},$$

where $s_{ij}^{\text{new}} = \exp(-\text{dist}(\phi^{\text{new}}(\mathbf{x}_i), \phi^{\text{new}}(\mathbf{x}_j)))$ and $s_{ij}^{\text{old}} = \exp(-\text{dist}(\phi^{\text{rev}}(\mathbf{x}_i), \phi^{\text{old}}(\mathbf{x}_j)))$. Figure 2 depicts the concept of the loss function. The positive pairs from the backward retrieval system $\{\phi^{\text{rev}}, \phi^{\text{old}}\}$ are trained to locate closer to the anchor than not only the negative pairs from the same system but also the ones from the new system $\{\phi^{\text{new}}, \phi^{\text{new}}\}$, and vice versa. We finally replace equation 9 with equation 11 for training the reverse transformation module. Compared to equation 10, additional heterogeneous negative terms in the denominator of equation 11 play a role as a regularizer to make the distances from one model directly comparable to those from other one, which is desirable for our rank merge strategy.

3.3. Learnable New Feature Embedding

Until now, we do not jointly train the reverse transformation module $\psi(\cdot)$ and the new feature extraction module $\phi^{\text{new}}(\cdot)$. This hampers the distance compatibility between the backward and new retrieval systems because the backward retrieval system $\{\phi^{\text{rev}}, \phi^{\text{old}}\}$ is the only part to be optimized while the new system $\{\phi^{\text{new}}, \phi^{\text{new}}\}$ is fixed. To provide more flexibility, we add another transformation module $\rho(\cdot)$ on top of the new model where $\rho^{\text{new}} = \rho(\phi^{\text{new}})$ and $\rho^{\text{rev}} = \psi(\rho(\phi^{\text{new}}))$. In this setting, we use ρ^{new} as the final new model instead of ϕ^{new}, and our rank merge process employs $\{\rho^{\text{rev}}, \phi^{\text{old}}\}$ and $\{\rho^{\text{new}}, \rho^{\text{new}}\}$ eventually. This strategy helps to achieve a better compatibility by allowing both systems to be trainable. The final loss function to train the reverse transformation module has the identical form to $\mathcal{L}_{\text{MCL}}$ in equation 11 except for the definitions of s_{ij}^{new} and s_{ij}^{old}, which are given by

$$s_{ij}^{\text{new}} = \exp\left(-\text{dist}\left(\rho^{\text{new}}(\mathbf{x}_i), \rho^{\text{new}}(\mathbf{x}_j)\right)\right) \tag{12}$$
$$s_{ij}^{\text{old}} = \exp\left(-\text{dist}\left(\rho^{\text{rev}}(\mathbf{x}_i), \phi^{\text{old}}(\mathbf{x}_j)\right)\right). \tag{13}$$

This extension still does not constrain ρ^{new} to be feature compatible with ϕ^{old}, thus preserving its feature discriminability while further improving distance compatibility.

4. Experiments

Table 1 and Figure A3 compare the proposed framework, referred to as RM (Rank Merge), with existing compatible learning approaches, including BCT (Shen et al., 2020), FCT (Ramanujan et al., 2022), and BiCT (Su et al., 2022). As shown in the table, RM consistently outperforms all the existing compatible training methods by remarkably significant margins in all datasets. BCT learns backward compatible feature representations, which is backfill-free, but its performance gain is not impressive. FCT achieves meaningful performance improvement by transforming old gallery features, but most of the gains come from side-information.[3] Also, such side-information is not useful for re-identification dataset, Market-1501, mainly because the model for the side-information is trained for image classification using ImageNet, which shows its limited generalizability. On the other hand, although BiCT takes advantage of online backfilling with less backfilling cost, it suffers from degraded final performance and negative flips in the middle of backfilling. Note that $\text{RM}_{\text{naïve}}$, our naïve rank merging between old and new models, is already competitive to other approaches.

[3] If side-information is not available, the performance gain of FCT drops from 62% to 28% on ImageNet.

References

Bai, Y., Jiao, J., Wu, S., Lou, Y., Liu, J., Feng, X., and Duan, L.-Y. Dual-tuning: Joint prototype transfer and structure regularization for compatible feature learning. *arXiv preprint arXiv:2108.02959*, 2021.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In *ICLR*, 2020.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

Duggal, R., Zhou, H., Yang, S., Xiong, Y., Xia, W., Tu, Z., and Soatto, S. Compatibility-aware heterogeneous visual search. In *CVPR*, 2021.

Gordo, A., Almazán, J., Revaud, J., and Larlus, D. Deep image retrieval: Learning global representations for image search. In *ECCV*, 2016.

He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. In *CVPR*, 2016.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. Supervised contrastive learning. *NeurIPS*, 2020.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Li, W., Zhao, R., Xiao, T., and Wang, X. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014.

Li, Y., Yosinski, J., Clune, J., Lipson, H., and Hopcroft, J. Convergent learning: Do different neural networks learn the same representations? *arXiv preprint arXiv:1511.07543*, 2015.

Loshchilov, I. and Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

Meng, Q., Zhang, C., Xu, X., and Zhou, F. Learning compatible embeddings. In *ICCV*, 2021.

Oord, A. v. d., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Pan, X., Luo, H., Chen, W., Wang, F., Li, H., Jiang, W., Zhang, J., Gu, J., and Li, P. Dynamic gradient reactivation for backward compatible person re-identification. *arXiv preprint arXiv:2207.05658*, 2022.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

Ramanujan, V., Vasu, P. K. A., Farhadi, A., Tuzel, O., and Pouransari, H. Forward compatible training for large-scale embedding retrieval systems. In *CVPR*, 2022.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3): 211–252, 2015.

Shen, Y., Xiong, Y., Xia, W., and Soatto, S. Towards backward-compatible representation learning. In *CVPR*, 2020.

Su, S., Zhang, B., Ge, Y., Xu, X., Wang, Y., Yuan, C., and Shan, Y. Privacy-preserving model upgrades with bidirectional compatible training in image retrieval. *arXiv preprint arXiv:2204.13919*, 2022.

Sun, Y., Chen, Y., Wang, X., and Tang, X. Deep learning face representation by joint identification-verification. *NIPS*, 2014.

Wan, T. S., Chen, J.-C., Wu, T.-Y., and Chen, C.-S. Continual learning for visual search with backward consistent feature embedding. In *CVPR*, 2022.

Wang, F., Chen, L., Li, C., Huang, S., Chen, Y., Qian, C., and Loy, C. C. The devil of face recognition is in the noise. In *ECCV*, 2018a.

Wang, L., Hu, L., Gu, J., Hu, Z., Wu, Y., He, K., and Hopcroft, J. Towards understanding learning representations: To what extent do different neural networks learn the same representation. *NeurIPS*, 2018b.

Weyand, T., Araujo, A., Cao, B., and Sim, J. Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval. In *CVPR*, 2020.

Zhang, B., Ge, Y., Shen, Y., Li, Y., Yuan, C., Xu, X., Wang, Y., and Shan, Y. Hot-refresh model upgrades with regression-free compatible training in image retrieval. In *ICLR*, 2021.

Zhang, B., Ge, Y., Shen, Y., Su, S., Yuan, C., Xu, X., Wang, Y., and Shan, Y. Towards universal backward-compatible representation learning. *arXiv preprint arXiv:2203.01583*, 2022.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. Scalable person re-identification: A benchmark. In *ICCV*, 2015.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. Learning deep features for scene recognition using places database. *NIPS*, 2014.