Learning Compatible Representations

Alberto Del Bimbo, Niccolò Biondi, Simone Ricci and Federico Pernici
Dipartimento Ingegneria dell'Informazione, Università degli Studi di Firenze,
Via Santa Marta 3, 50139, Firenze, Italy

EXTENDED ABSTRACT

Retrieval systems are designed to identify similar data from a gallery set by comparing representations generated from queries with those stored in the gallery. The performance of these systems heavily relies on the quality of the learned representations, which has been greatly enhanced by advancements in deep neural networks. Traditional methods operate under the assumption of static scenarios: after training a network, feature vectors are extracted and stored for all gallery data. The search process then involves matching the query features with those stored in the gallery. However, the assumption of static scenarios does not hold in real-world settings. Here, novel data constantly emerge after deployment, requiring periodic updates to neural networks through additional training data. Moreover, network updating is also required by the emerging of new architectures, or revised training methodologies. Additionally, it is becoming a common practice to rely on large models pre-trained by third parties accessed by APIs. Such models often undergo frequent updates driven by evolving training strategies, architectural advancements, or access to higher-quality datasets.

While developers typically focus on enhancing performance, they often overlook how these updates can drastically alter the behavior of applications or services built on these models due to changes in feature representations. In such cases, it becomes essential to regenerate feature vectors for the entire gallery using the updated model due to inconsistencies between the old and new representations. This reprocessing is not only computationally expensive but may be impractical in real-world scenarios with galleries containing billions of data or impossible if the original images are inaccessible for reprocessing due to privacy restrictions. Recent research has therefore focused on developing methods for learning *backward-compatible* representations, addressing the challenge of aligning representations across models.

The concept of *backward compatibility* was first introduced in (Y. Shen, et al., 2020) and further explored in (Q. Meng, et al., 2021; T. Pan, et al., 2023; K. Chen, et al., 2019; J. Hu, et al., 2019; C.-Y. Wang, et al., 2020), and (F. Pernici, et al., 2019), among several others. In (Y. Shen, et al., 2020), the authors proposed the *Backward Compatible Training* method, which achieves compatibility by regularizing the training of the new model using the old classifier as a reference. A defining feature of learning compatible representations is the imposition of constraints on the semantic distances within the feature space. They established a general criterion for evaluating compatibility: a new, compatible representation model must perform at least as well as its predecessor in clustering images from the same class while effectively separating images from different classes. A new representation model φ_{new} is therefore *compatible* with an old representation model φ_{old} if:

$$\text{dist}\,(\varphi_{new}(x_u), \varphi_{old}(x_v)) \leq \text{dist}\,(\varphi_{old}(x_u), \varphi_{old}(x_v)) \quad \forall (u,v) \in \{(u,v) \mid y_u = y_v\} \quad \text{and}$$

$$\text{dist}\,(\varphi_{new}(x_u), \varphi_{old}(x_v)) \geq \text{dist}\,(\varphi_{old}(x_u), \varphi_{old}(x_v)) \quad \forall (u,v) \in \{(u,v) \mid y_u \neq y_v\},$$

where x_u and x_v are two input samples, y_u and y_v are their classes and dist $(\cdot,\cdot)$ is a distance in feature space. As it imposes constraints on all pairs of samples, for practical application this criterion is simplified into the *Empirical Compatibility Criterion:*

$$M\,(\varphi^{Q}_{new}, \varphi^{G}_{old}) > M\,(\varphi^{Q}_{old}, \varphi^{G}_{old})$$

where M is a metric used to evaluate the performance based on dist $(\cdot,\cdot)$.

In real-world applications, multi-step upgrades are often necessary, meaning that different representation models must be learned sequentially over time through multiple stages of updates. In such case, the *Empirical Compatibility Criterion* can be immediately extended to its multistep version as:

$$M\,(\varphi_{t'}{}^{Q}, \varphi_{t}{}^{G}) > M\,(\varphi_{t}{}^{Q}, \varphi_{t}{}^{G}) \;\forall\, t' > t \quad \text{with } t' \in \{2,3,...,T\} \text{ and } t \in \{1,2,...,T-1\},$$

The approach introduced in (Y. Shen, et al., 2020) was further developed by other researchers, notably (Q. Meng, et al., 2021) and (T. Pan, et al., 2023). In (Q. Meng, et al., 2021), additional regularization loss

Del Bimbo, A., Biondi, N., Ricci, S. and Pernici, F.
Learning Compatible Representations.
DOI: 10.5220/0013453300003905
In *Proceedings of the 14th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2025)*, pages 9-10
ISBN: 978-989-758-730-6; ISSN: 2184-4313

functions were introduced to align the new representation with the old one. While (T. Pan, et al., 2023) proposed using an adversarial learning discriminator to distinguish whether an embedding originates from the new or old model, ensuring that features from different models remain indistinguishable. However, in all these methods, while aligning with the previously learned model, the new backward-compatible model often struggles to match the performance of a newly independently trained model. In (K. Chen, et al., 2019; J. Hu, et al., 2019; and C.-Y. Wang, et al., 2020), the issue of feature compatibility was addressed following a different approach, by learning a mapping between two representation models to enable direct comparison between new and old feature vectors. However, this approach is clearly unsuited for sequential multi-model learning or large gallery sets.

In contrast, we propose a different approach that requires a suitable architectural change to the network. In (E. Hoffer, et al., 2018), it was demonstrated that a deep network with a fixed classification layer (i.e., one that is not trainable) initialized with random weights can be nearly as effective as a trainable classifier, offering significant savings in computational and memory resources. In fixed classifiers, the functional complexity of the classifier is entirely shifted to the internal layers of the neural network. Since the classifier prototypes' parameters are fixed, the feature vector directions must align with the directions of the class prototypes. We therefore argued that a compatibility training procedure should be defined by directly exploiting the stationarity of the representation as provided by fixed classifiers. In CoReS (N. Biondi, et al., 2024), we use a special class of fixed classifiers where the classifier weights are fixed to values taken from the coordinate vertices of regular polytopes. Regular polytopes extend the concept of regular polygons into higher dimensions and embody the idea of partitioning the available space into nearly equiangular regions. Among the various polytopes in a multi-dimensional feature space with 5 or more dimensions, we selected the *d*-Simplex because it is the only polytope that positions class prototypes equidistant from one another and maximally separated in the representation space, preserving their spatial configuration as new classes are added (F. Pernici, et al., 2022). The K classifier prototypes are computed as $W = [\, e_1, e_2, \ldots, e_{K-1}, (1 - \mathrm{sqrt}(K-1) / K-1) \sum_{i=1}^{k-1} e_i \,]$ where e_i is the standard basis in R_d, with $i \in \{1,2,\ldots,K-1\}$. Training is performed using the standard cross-entropy loss.

With our approach, there is no need to learn mappings between representations or to perform pairwise training with previously learned models. To ensure compatibility across model upgrades, we learn the features of new classes in designated regions of the representation space at each upgrade, while the features from previous upgrades remain fixed due to the stationarity property. Regions for the future classes are established at the start of training by setting the number of classifier outputs higher than the number of initial classes and leaving unassigned regions for future upgrades. We can demonstrate that the stationary representation provided by fixed classifiers ensures compatibility by satisfying the two inequality constraints.

REFERENCES

Y. Shen, Y. Xiong, W. Xia, and S. Soatto, “Towards backward-compatible representation learning,” in *Proc. IEEE/CVF CVPR.*, 2020, pp. 6368–6377.

Q. Meng, C. Zhang, X. Xu, and F. Zhou, “Learning compatible embeddings,” in *Proc. IEEE/CVF CVPR.*, 2021, pp. 9939–9948.

T. Pan, F. Xu, X. Yang, S. He, C. Jiang, Q. Guo, F. Qian, X. Zhang, Y. Cheng, L.Yang, et al. “Boundary-aware backward-compatible representation via adversarial learning in image retrieval” In Proceedings of the *IEEE/CVF CVPR*, pp. 15201–15210, 2023.

K. Chen, Y. Wu, H. Qin, D. Liang, X. Liu, and J. Yan, “R3 adversarial network for cross model face recognition,” in *Proc. IEEE/CVF CVPR.*, 2019, pp. 9868–9876.

J. Hu, R. Ji, H. Liu, S. Zhang, C. Deng, and Q. Tian, “Towards visual feature translation,” in *Proc. IEEE/CVF CVPR.*, 2019, pp. 3004–3013.

C.-Y. Wang, Y.-L. Chang, S.-T. Yang, D. Chen, and S.-H. Lai, “Unified representation learning for cross model compatibility,” in *Proc. 31st* BMVA, 2020.

F. Pernici, M. Bruni, C. Baecchi, and A. Del Bimbo, “Maximally compact and separated features with regular polytope networks,” in *Proc. IEEE/CVF CVPR. Workshops*, 2019, pp. 46–53.

E. Hoffer, I. Hubara, and D. Soudry, “Fix your classifier: The marginal value of training the last weight layer,” in *Proc. ICLR*, 2018.

N. Biondi, F. Pernici, M. Bruni, and A. Del Bimbo “CoReS: Compatible Representations via Stationarity” submitted to IEEE TPAMI, under revision, 2024.

F. Pernici, M. Bruni, C. Baecchi, and A. D. Bimbo, “Regular polytope networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4373–4387, Sep. 2022.