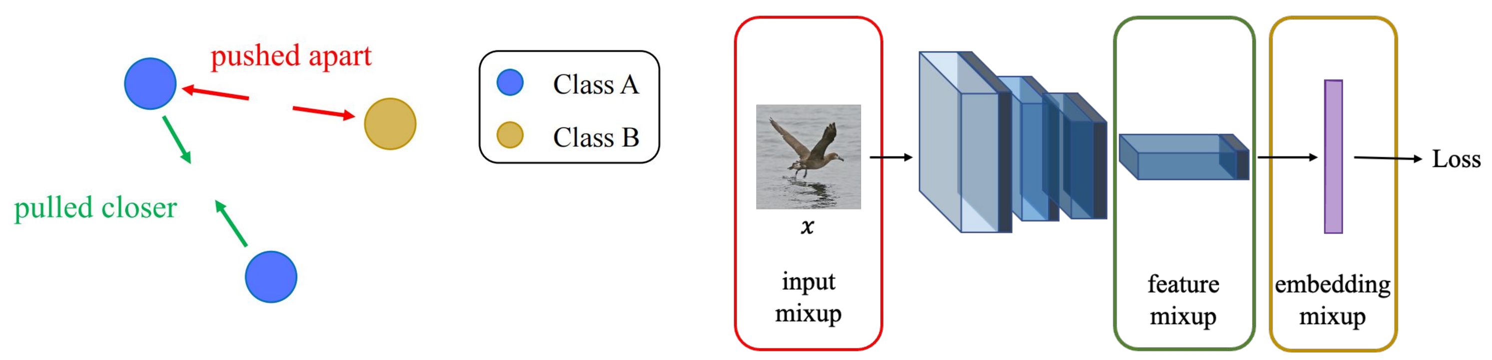
It Takes Two to Tango: Mixup for Deep Metric Learning

Shashanka Venkataramanan, Bill Psomas, Ewa Kijak, Laurent Amsaleg, Konstantinos Karantzalos, Yannis Avrithis

Deep Metric Learning & Mixup

- **Goal** - Learning a discriminative representation that generalizes to unseen classes.
- **How?** - Intra-class embeddings are pulled closer and inter-class embeddings are pushed apart.
- **Motivation** - Classes during training and inference are different, interpolation-based data augmentation e.g. mixup plays significant role.



Left: Deep Metric Learning has binary labels (*positive*/*negative*).
Right: Mixup [1, 2] interpolates between examples (*input*, *feature or embedding*) and has non-binary mixed labels.

Generic Loss Formulation

Additive losses e.g., Contrastive [3] and non-additive losses e.g., Multi-similarity [4] involve:

- A sum over positives $P(a)$ and a sum over negatives $N(a)$.
- A decreasing function ρ^+ of similarity $s(a,p)$ for $p \in P(a)$ and an increasing function ρ^- of similarity $s(a,n)$ for $n \in N(a)$.

Non-additive losses also involve non-linear functions σ^+ and σ^-.

$$\ell(a;\theta) := \sigma^+\left(\sum_{p\in P(a)} \rho^+(s(a,p))\right) + \sigma^-\left(\sum_{n\in N(a)} \rho^-(s(a,n))\right)$$

Positives $P(a)$ and negatives $N(a)$ of anchor a have the same or different class label as the anchor.
A binary class label $y \in \{0,1\}$ for each example in $P(a) \cup N(a)$ is defined: $y=1$ for positives, $y=0$ for negatives.

$$\ell(a;\theta) := \tau\left(\sigma^+\left(\sum_{(x,y)\in U(a)} y\rho^+(s(a,x))\right) + \sigma^-\left(\sum_{(x,y)\in U(a)} (1-y)\rho^-(s(a,x))\right)\right)$$

y is binary, only one of the two contributions is non-zero.

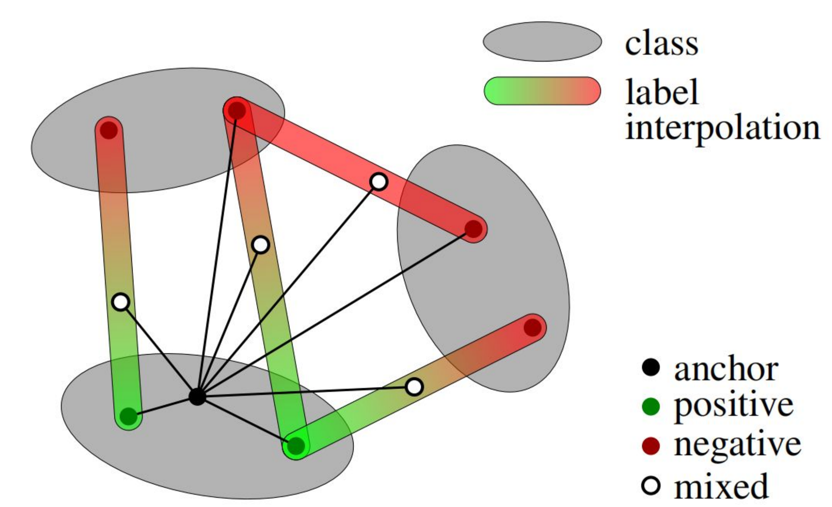
Interpolating Labels

Given $M(a)$, which is the possible choices of mixing pairs (*positive-positive*, *positive-negative*, *negative-negative*), the labeled mixed embedding is:

$$V(a) = \{f_\lambda(x,x'), \mathrm{mix}_\lambda(y,y') : (x,y),(x',y') \in M(a)\}$$

$$\tilde{\ell}(a;\theta) := \tau\left(\sigma^+\left(\sum_{(v,y)\in V(a)} y\rho^+(s(a,v))\right) + \sigma^-\left(\sum_{(v,y)\in V(a)} (1-y)\rho^-(s(a,v))\right)\right)$$

y ∈ [0,1], both contributions are non-zero.



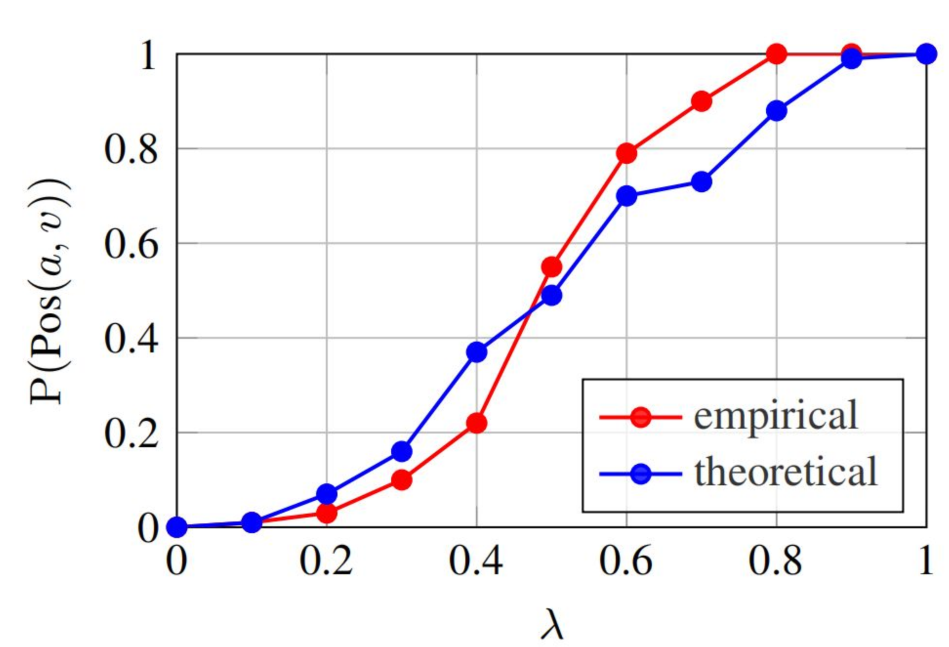
Metrix (=Metrix Mix) allows an anchor to interact with *positive* (same class), *negative* (different class) and *interpolated examples*, which also have interpolated *labels*.

Analysis: Mixed Embeddings and Positivity

- $Pos(a,v)$: a mixed embedding v behaves as "positive" for anchor a.
- "Positivity" is equivalent to $\partial\tilde{\ell}(a;\theta)/\partial s(a,v) \leq 0$.
- Under *positive-negative* mixing, i.e. $M(a) \subset U^+(a) \times U^-(a)$, the probability of $Pos(a,v)$ as a function of λ is:

$$\mathrm{P}(\mathrm{Pos}(a,v)) = F_\lambda\left(\frac{1}{\beta+\gamma}\ln\left(\frac{\lambda}{1-\lambda}\right) + m\right)$$

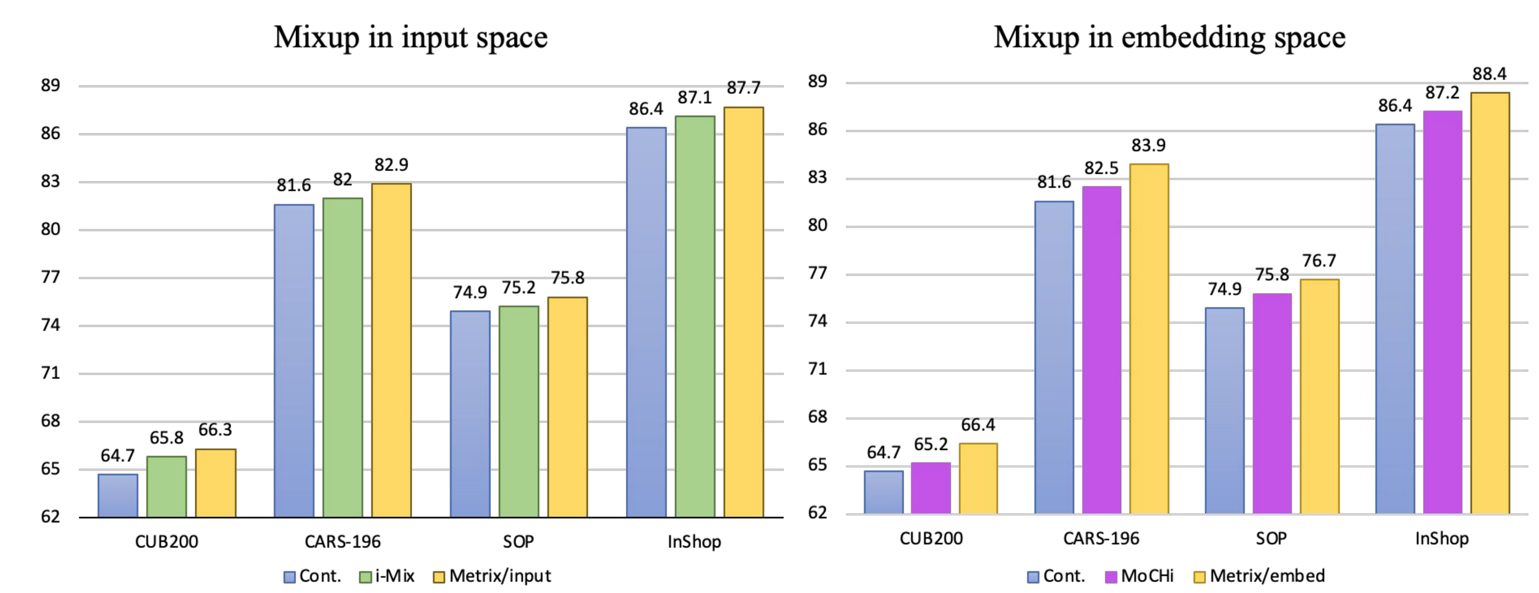
- We measure this function both empirically and theoretically:



Improving Losses with Metrix

Method	CUB200 [5] R@1	R@2	R@4	CARS196 [6] R@1	R@2	R@4	SOP [7] R@1	R@10	R@100	IN-SHOP [8] R@1	R@10	R@20
MS	67.8	77.8	85.6	87.8	92.7	95.3	76.9	89.8	95.9	90.1	97.6	98.4
MS +Metrix	**71.4**	**80.6**	**86.8**	**89.6**	**94.2**	**96.0**	**81.0**	**92.0**	**97.2**	**92.2**	**98.5**	**98.6**
PA [9]	69.5	79.3.	87.0	87.6	92.3	95.5	79.1	90.8	96.2	90.0	97.4	98.2
PA +Metrix	**71.0**	**81.8**	**88.2**	**89.1**	**93.6**	**96.7**	**81.3**	**91.7**	**96.9**	**91.9**	**98.2**	**98.8**

Comparison with other Mixing Methods

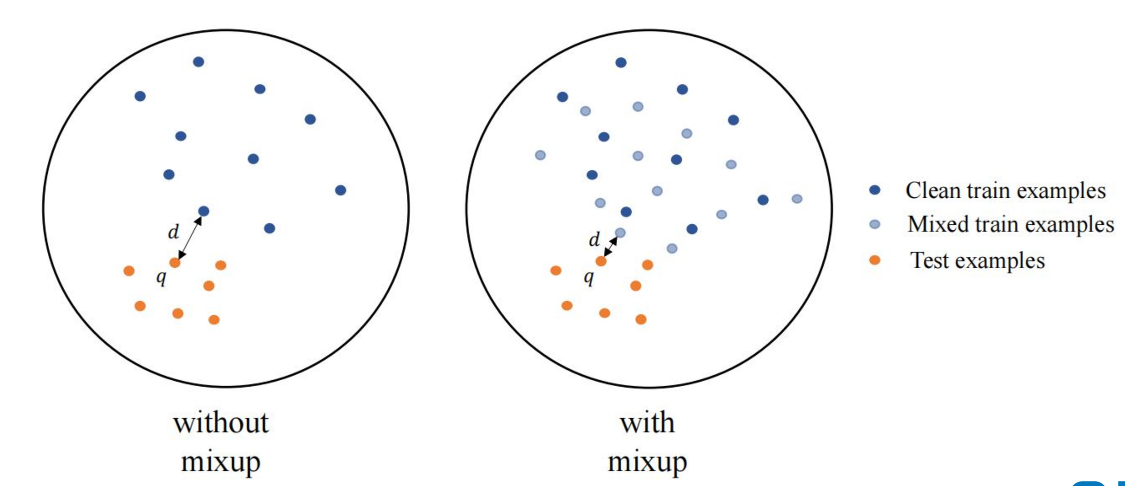


How Does Metrix Improve Representations?

- *Utilization* of the training set X by the test set Q as:

$$u(Q,X) = \frac{1}{|Q|}\sum_{q\in Q}\min_{x\in X}\|f(q)-f(x)\|^2$$

- Low utilization indicates that there are examples in the training set that are similar to test examples.



References

[1] Zhang et al., mixup: Beyond Empirical Risk Minimization. In: ICLR, 2018.
[2] Verma et al., Manifold Mixup: Better Representations by Interpolating Hidden States. In: ICML, 2019.
[3] Hadsell et al., Dimensionality reduction by learning an invariant mapping. In: CVPR, 2006.
[4] Wang et al., Multi-similarity loss with general pair weighting for deep metric learning. In: CVPR, 2019.
[5] Wah et al., The Caltech-UCSD Birds-200-2011 Dataset. In: CNS-TR-2011-001, 2011.
[6] Krause et al., 3d object representations for fine-grained categorization. In: ICCVW, 2013.
[7] Song et al., Deep metric learning via lifted structure feature embedding. In: CVPR, 2016.
[8] Liu et al., Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: CVPR, 2016.
[9] Kim et al., Proxy Anchor Loss for Deep Metric Learning. In: CVPR, 2020.
[10] Lee et al., I-mix: A domain-agnostic strategy for contrastive representation learning. In: ICLR, 2021.
[11] Kalantidis et al., Hard Negative Mixing for Contrastive Learning. In: NeurIPS, 2020.