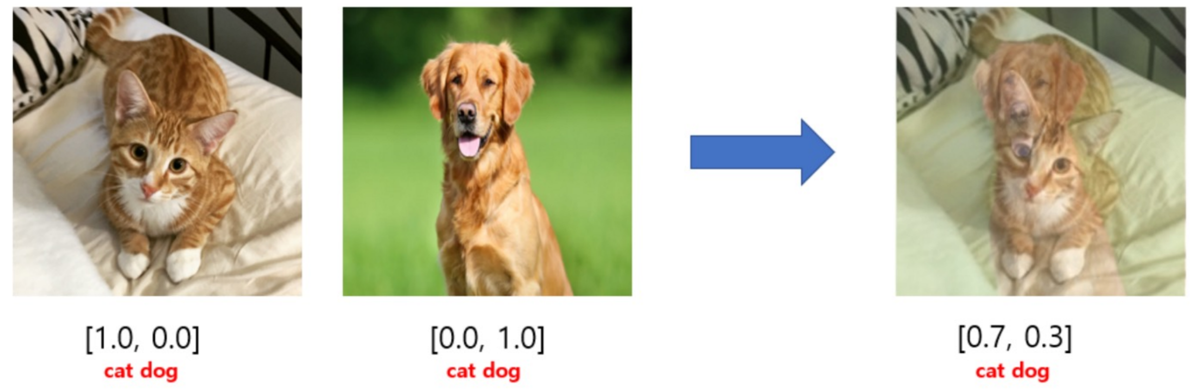
AlignMixup: Improving Representations By Interpolating Aligned Features

Shashanka Venkataramanan[1], Ewa Kijak[1], Laurent Amsaleg[1], and Yannis Avrithis[2]
[1]Inria, Univ Rennes, CNRS, IRISA; [2]Athena Research Center
shashanka.venkataramanan@inria.fr

Mixup improves generalization

- Data Augmentation technique that **interpolates** between pairs of examples (input/feature) and its labels.
- **Flattens** class representations, reduces overconfident incorrect predictions and **smoothens** decision boundaries.

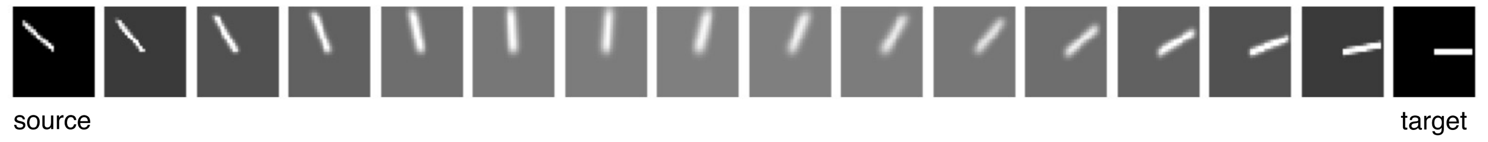


Existing Mixup Methods

- Recent works **combine multiple objects** in an image by cut-and-paste [3] or use salient regions [4, 5].
- This results in an **overlay** of one image onto another.
- These methods make **efficient use of training pixels**; interpolation can be better defined.

AlignMixup: Deformation

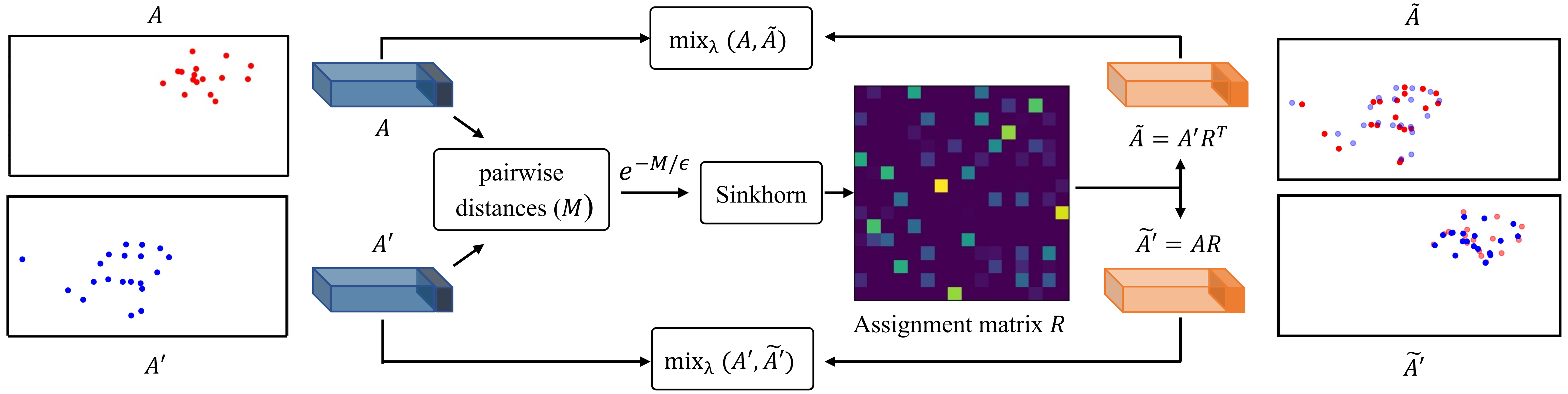
- **Deformation** a natural way of interpolating images e.g. one image may deform into another, in a continuous way.
- Traversing along the manifold of representations obtained from **deeper layers** captures **salient characteristics**.



AlignMixup:

- **Geometrically aligning** features feature in **soft correspondences**.
- Alignment is based on **optimal transport** and **Sinkhorn divergence**.

AlignMixup: Aligning feature tensors



Alignment:

- **Cost matrix** $M : r \times r$, where : $m_{ij} := ||a_i - a'_j||^2$. $r = h \times w$; $a_j, a'_j \in \mathbb{R}^c$ are columns of $\mathbf{A}, \mathbf{A}'$ representing a spatial position in the original image x, x'.
- **Transport plan**, $r \times r$ matrix $P \in U_r$, where $U_r := \{P \in \mathbb{R}_+^{r \times r} : P\mathbf{1} = P^\top \mathbf{1} = \mathbf{1}/r\}$
- **Optimization function** $P^* = \arg\min_{P \in U_r} \langle P, M \rangle - \epsilon H(P)$
- **Assignment Matrix** $R := rP^*$; $r_{ij} \in R$ expresses the probability with which column a_i of A corresponds to column a'_j of $\mathbf{A}'$.
- **Alignment** $\widetilde{\mathbf{A}} := \mathbf{A}'R^\top$ and $\widetilde{\mathbf{A}}' := \mathbf{A}R$ column $\widetilde{a}_i$ of $c \times r$ matrix $\widetilde{\mathbf{A}}$ is a convex combination of columns of A' that corresponds to the same column a_i of A.
- $\widetilde{\mathbf{A}}$ represents **A aligned to A′**, and $\widetilde{\mathbf{A}}'$ represents **A′ aligned to A**.

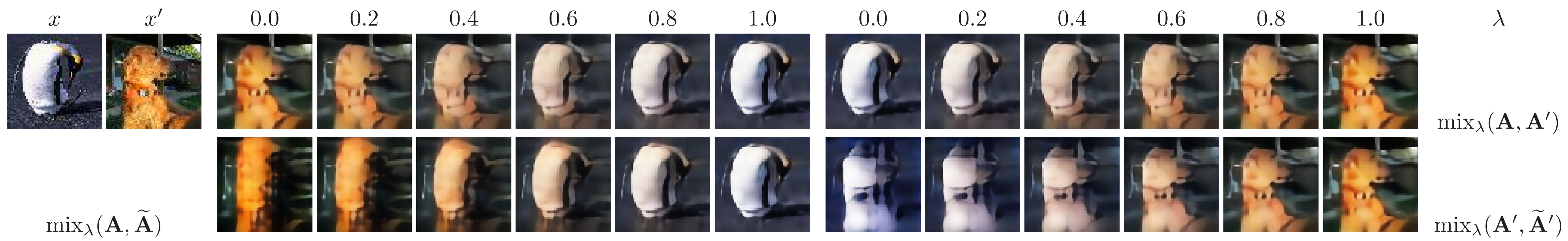
Interpolation:

- We interpolate between $\widetilde{\mathbf{A}}$ and $\mathbf{A}$, and between $\widetilde{\mathbf{A}}'$ and $\mathbf{A}'$ as

$$\text{mix}_\lambda(\mathbf{A}, \widetilde{\mathbf{A}}) = \lambda \mathbf{A} + (1 - \lambda)\widetilde{\mathbf{A}} \tag{1}$$

$$\text{mix}_\lambda(\mathbf{A}', \widetilde{\mathbf{A}}') = \lambda \mathbf{A}' + (1 - \lambda)\widetilde{\mathbf{A}}' \tag{2}$$

Visualization



- **Asymmetric** morphing, where one object continuously deforms into itself.
- Retain **geometry** or **pose** of the image and keep the coordinates and the **appearance** or **texture** of the other.

Experimental Results

Image Classification

Dataset Network	Cifar-10 R-18	Cifar-100 R-18	TI R-18	ImNet R-50
Baseline	5.19	23.24	43.40	23.68
Input [2]	4.03	20.21	43.48	22.58
Manifold [1]	2.95	19.80	40.76	22.50
PuzzleMix [4]	2.93	20.01	36.52	21.24
Co-Mixup [5]	2.89	19.81	35.85	–
StyleCutMix [9]	3.06	19.34	34.49	–
AlignMixup (ours)	2.95	18.29	33.13	20.68
AlignMixup/AE (ours)	**2.83**	**17.82**	**32.73**	**18.83**
Gain	**+0.06**	**+1.52**	**+1.76**	**+2.41**

Image classification top-1 error (%): lower is better. Gain: reduction of error. TI: TinyImagenet, ImNet: ImageNet; R: PreActResnet.

Weakly-supervised Object Localization

Metric	Top-1 Loc.		MaxBoxAcc-v2	
Network	VGG-GAP	ResNet-50	VGG-GAP	ResNet-50
Baseline CAM [8]	37.1	49.4	59.0	59.7
Input [2]	41.7	49.3	57.1	60.6
CutMix [3]	52.5	54.8	62.6	64.8
AlignMixup (ours)	**53.1**	**56.2**	**63.8**	**65.4**
Gain	**+0.6**	**+1.4**	**+1.2**	**+0.6**

Weakly-supervised object localization on CUB200-2011. Top-1 localization accuracy (%): higher is better. Gain: increase in accuracy.

References

[1] Verma et al. Manifold mixup: Better representations by interpolating hidden states *ICML*, 2019.
[2] Zhang et al. mixup: Beyond empirical risk minimization *ICLR*, 2018.
[3] Yun et al. Cutmix: Regularization strategy to train strong classifiers with localizable features *ICCV*, 2019.
[4] Kim et al. Puzzle mix: Exploiting saliency and local statistics for optimal mixup *ICML*, 2020.
[5] Kim et al. Co-Mixup: Saliency Guided Joint Mixup with Supermodular Diversity *ICLR*, 2021.
[6] Bengio et al. Better mixing via deep representations *ICML*, 2013.
[7] Choy et al. Universal Correspondence Network *NIPS*, 2016.
[8] Zhou et al. Learning deep features for discriminative localization *CVPR*, 2016.
[9] Hong et al. StyleMix: Separating Content and Style for Enhanced Data Augmentation *CVPR*, 2021.

Acknowledgement

This work was partially supported by the ANR-19-CE23-0028 MEERQAT and was performed using the HPC resources from GENCI-IDRIS Grant 2021 AD011012528.