Deep Transformation-Invariant Clustering

Tom Monnier[1] Thibault Groueix[1] Mathieu Aubry[1]

http://imagine.enpc.fr/~monniert/DTIClustering/

Pytorch code

Motivation

Goal → efficiently cluster images, even in the wild

Challenge → distances not invariant to image transformations

spatial transformation morpho color & illumination

Previous work

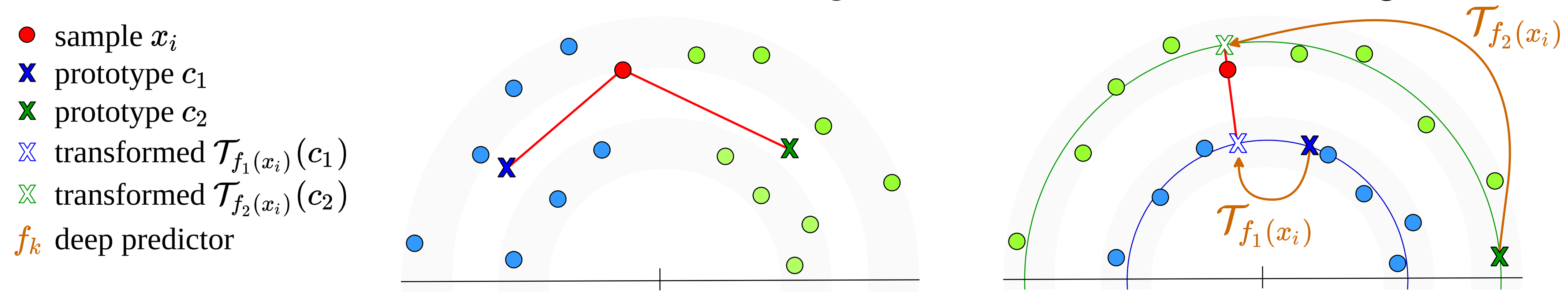
1. Clustering in a **feature space**
   state-of-the-art struggle with real images hard to interpret
2. Align images in **pixel space** before clustering them
   highly interpretable difficult optimization simple alignments

Contributions

1. Joint learning of clustering and deep alignment in pixel space
2. Approach with state-of-the-art and interpretable results

Method

Standard clustering **DTI clustering**



Clustering loss

Standard $\mathcal{L}(c_{1:K}) = \sum_{i=1}^{N} l(x_i, \{c_1, \dots, c_K\})$

Transformation-Invariant (TI)

$$\mathcal{L}_{\text{TI}}(c_{1:K}) = \sum_{i=1}^{N} \min_{\beta_{1:K}} l(x_i, \{\mathcal{T}_{\beta_1}(c_1), \dots, \mathcal{T}_{\beta_K}(c_K)\})$$

Deep Transformation-Invariant (DTI)

$$\mathcal{L}_{\text{DTI}}(c_{1:K}, f_{1:K}) = \sum_{i=1}^{N} l(x_i, \mathcal{T}_{f_1(x_i)}(c_1), \dots, \mathcal{T}_{f_K(x_i)}(c_K)\})$$

Applications

K-means $\mathcal{L}_{\text{K-means}}(c_{1:K}) = \sum_{i=1}^{N} \min_k \|x_i - c_k\|^2$

$$\mathcal{L}_{\text{DTI K-means}}(c_{1:K}, f_{1:K}) = \sum_{i=1}^{N} \min_k \|x_i - \mathcal{T}_{f_k(x_i)}(c_k)\|^2$$

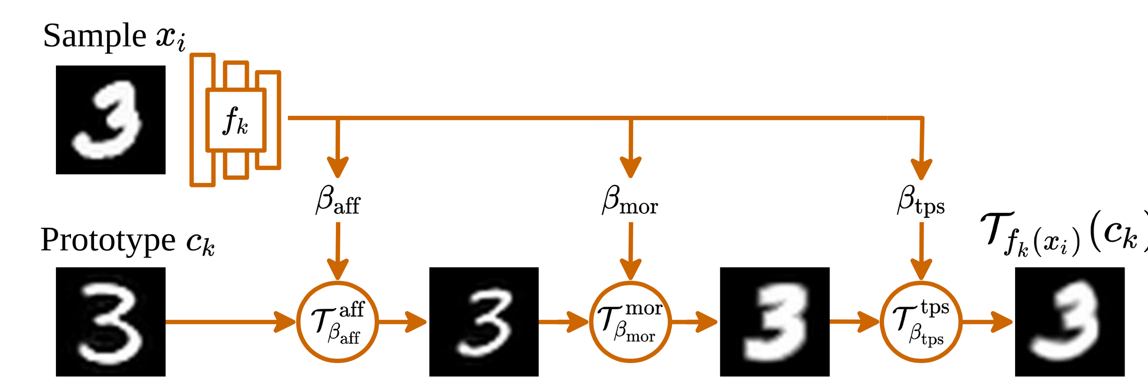
GMM $\mathcal{L}_{\text{GMM}}(\mu_{1:K}, \Sigma_{1:K}, \pi_{1:K}) = -\sum_{i=1}^{N} \log\Big(\sum_{k=1}^{K} \pi_k G(x_i\,; \mu_k, \Sigma_k)\Big)$

$$\mathcal{L}_{\text{DTI GMM}}(\mu_{1:K}, \Sigma_{1:K}, \pi_{1:K}, f_{1:K}) = -\sum_{i=1}^{N} \log\Big(\sum_{k=1}^{K} \pi_k G(x_i\,; \mathcal{T}_{f_k(x_i)}(\mu_k), \mathcal{T}^*_{f_k(x_i)}(\Sigma_k))\Big)$$

Transformation sequence $\mathcal{T}_{\beta_M}^{M} \circ \dots \circ \mathcal{T}_{\beta_1}^{1}$

Transformation modules

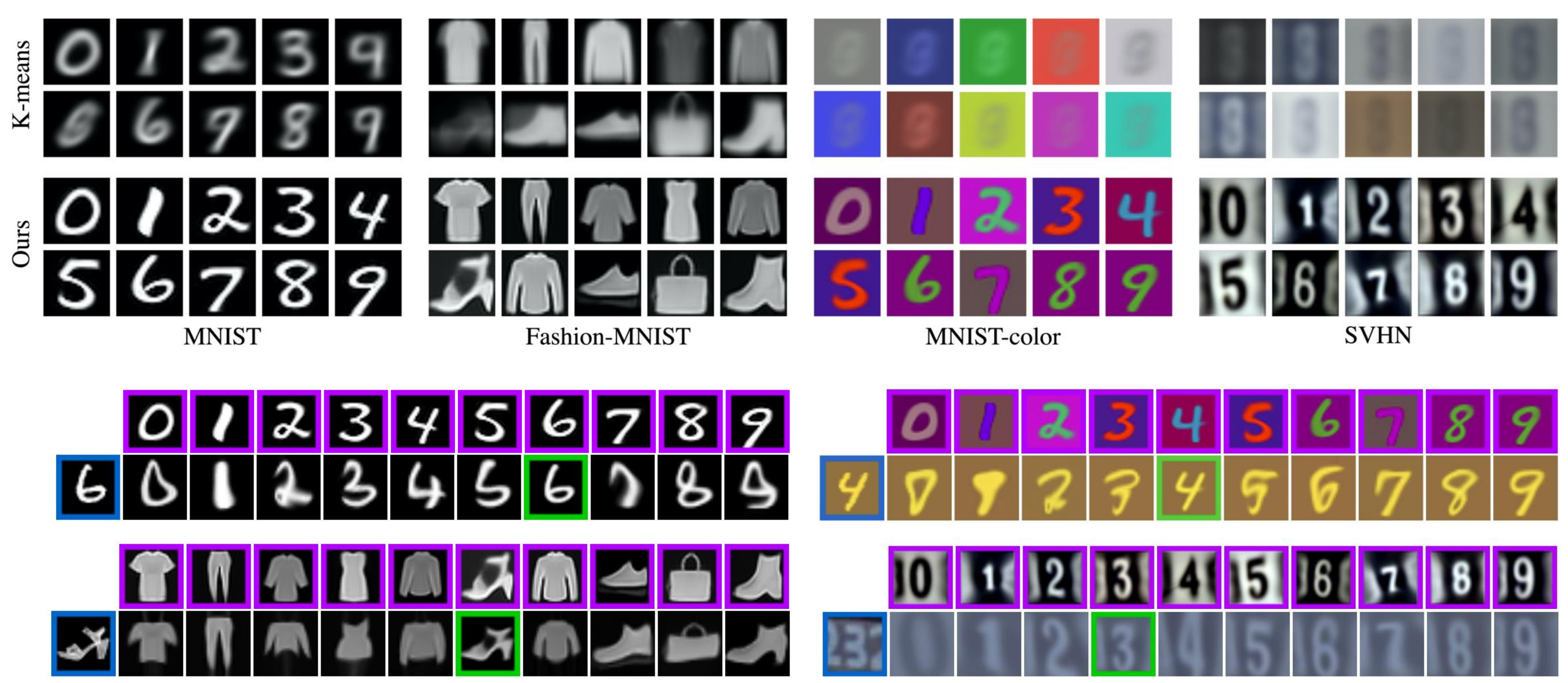
- spatial transformers [1] → affine $\mathcal{T}_\beta^{\text{aff}}$, projective $\mathcal{T}_\beta^{\text{proj}}$, thin plate spline $\mathcal{T}_\beta^{\text{tps}}$
- color transformation $\mathcal{T}_\beta^{\text{col}}$
- morphological transformation $\mathcal{T}_\beta^{\text{mor}}$ (new)



Key elements for training
→ curriculum learning + cluster reassignment

Results

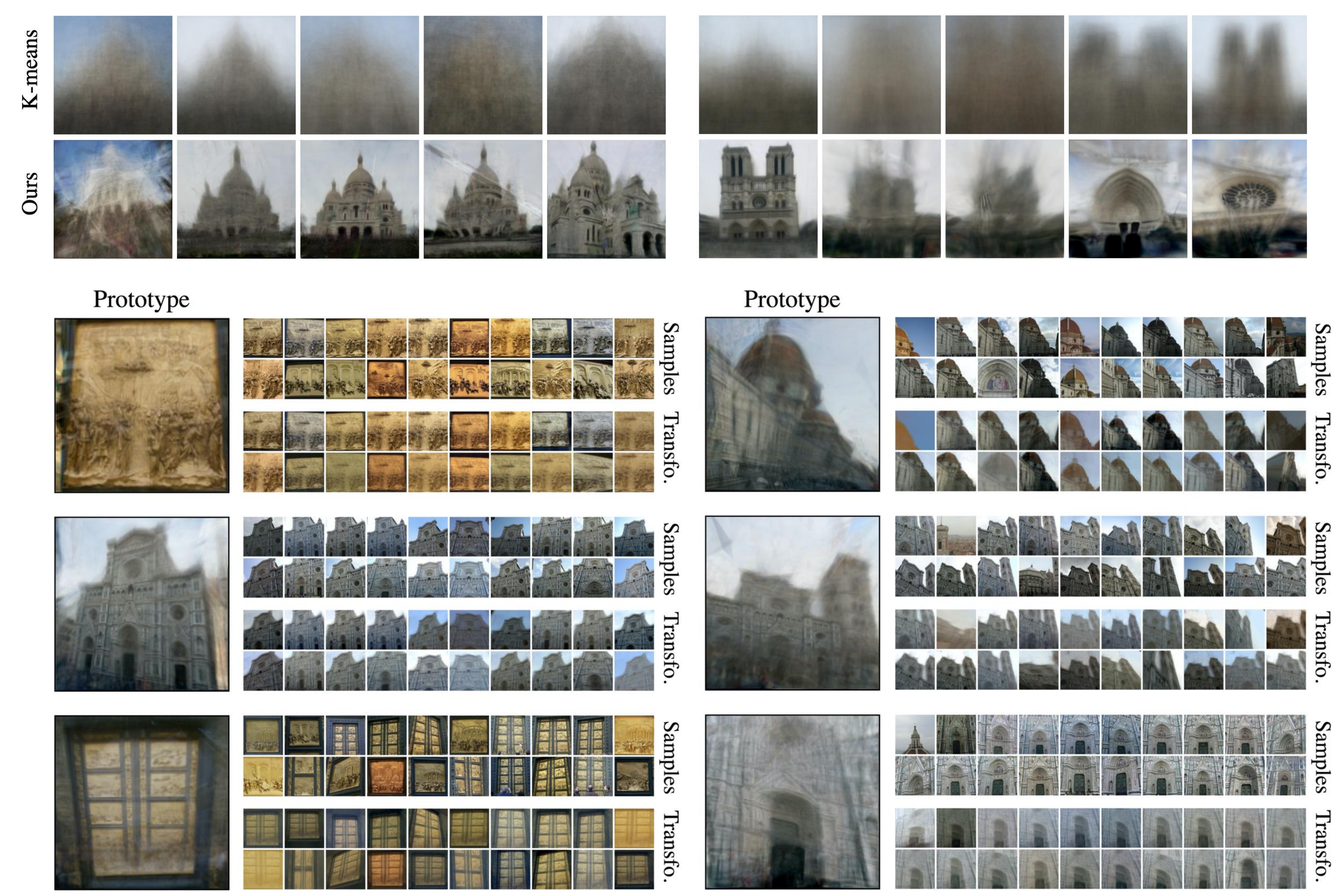
Standard image clustering benchmarks



prototype / query sample / closest aligned prototype

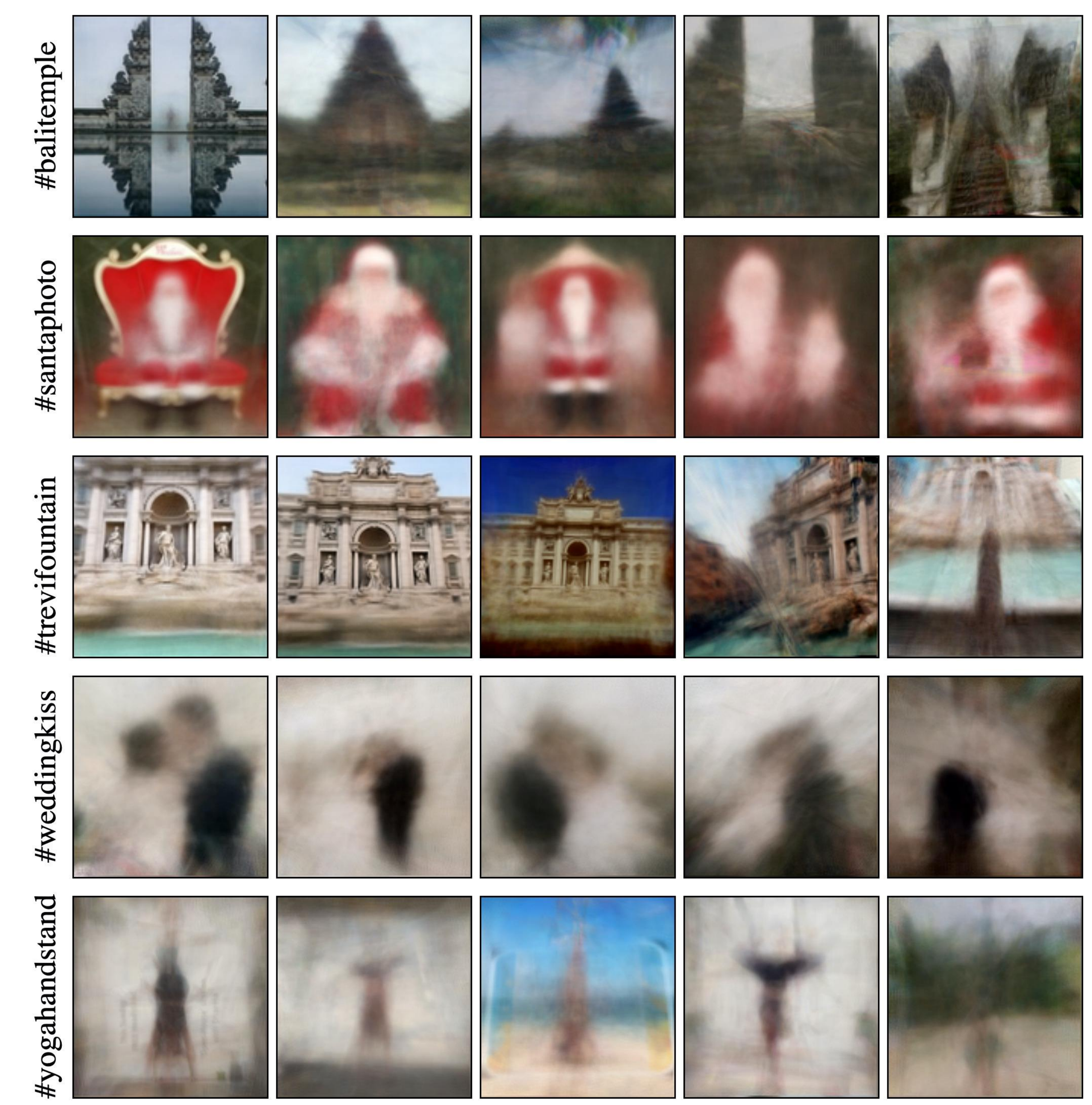
Method	Runs	Eval	MNIST ACC	MNIST NMI	MNIST-test ACC	MNIST-test NMI	USPS ACC	USPS NMI	F-MNIST ACC	F-MNIST NMI	SVHN ACC	MNIST-color ACC
Clustering on a learned feature												
DEPICT [2]	5	avg	96.5	91.7	96.3	91.5	**96.4**	**92.7**	39.2	39.2	-	-
DSCDAN [3]	10	avg	**97.8**	**94.1**	**98.0**	**94.6**	86.9	85.7	**66.2**	**64.5**	-	-
Clustering on a learned feature with data augmentation and/or ad hoc data representation												
IMSAT [4]	12	avg	98.4$^\triangledown$	-	-	-	-	-	-	-	**57.3**$^{\triangledown\dagger}$	10.6
IIC [5]	5	avg	98.4$^\triangledown$	-	-	-	-	-	-	-	-	10.6
	5	minLoss	**99.2**$^\triangledown$	-	-	-	-	-	-	-	-	10.6
Clustering on pixel values												
K-means	10	avg	54.8	50.2	55.9	51.2	65.3	61.2	54.1	51.4	12.2	10.5
DTI K-means	10	avg	**97.3**	**94.0**	96.6	94.6	86.4	88.2	**61.2**	63.7	44.5*	96.7
	10	minLoss	97.2	93.8	**98.0**	**95.3**	**89.8**	**89.5**	57.4	**64.1**	**62.6***	**96.8**

Filtered web images (MegaDepth [6])



[1] Spatial Transformer Networks, Jaderberg et al. 2015
[2] Deep Clustering via Joint Convolutional Autoencoder Embedding and Relative Entropy Minimization, Dizaji et al. 2017
[3] Deep Spectral Clustering Using Dual Autoencoder Network, Yang et al. 2019
[4] Learning Discrete Representations via Information Maximizing Self-Augmented Training, Hu et al. 2017
[5] Invariant Information Clustering for Unsupervised Image Classification and Segmentation, Ji et al. 2019
[6] MegaDepth: Learning Single-View Depth Prediction from Internet Photos, Li and Snavely 2018

Raw Instagram hashtags



[1] LIGM, Ecole des ponts, Univ Gustave Eiffel, CNRS, France