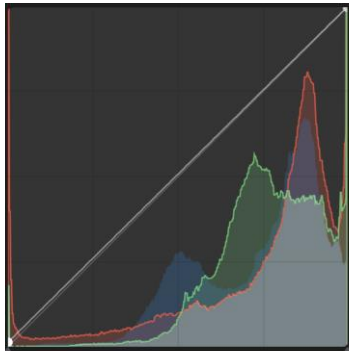
METRIC LEARNING WITH ADAPTIVE DENSITY DISCRIMINATION

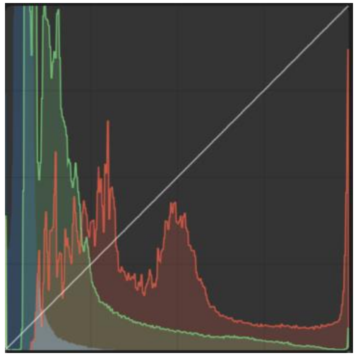
[OREN RIPPEL, MANOHAR PALURI, PIOTR DOLLAR, LUBOMIR BOURDEV 2016]

PRESENTATION BY PHILIPP REISER

INTRODUCTION

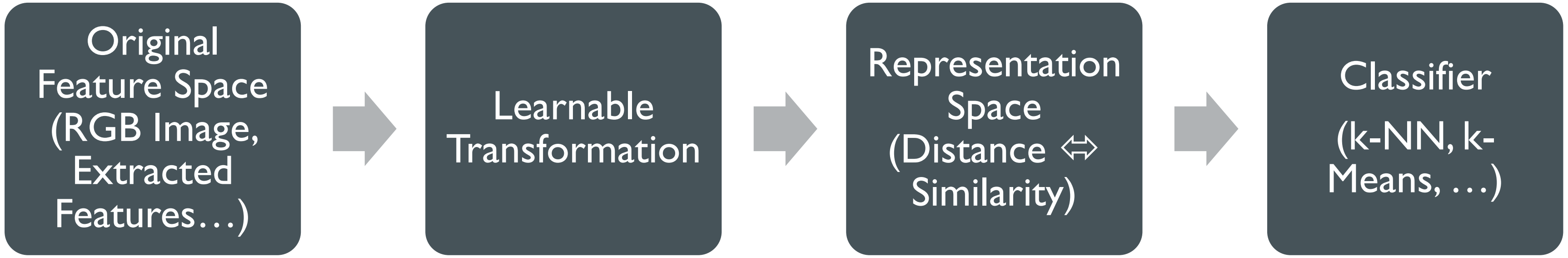
- Color similarity with color histograms
- Distance measure: L1, L2, …
- Problems:
  - Not robust
  - Works only with toy datasets





INTRODUCTION

- Metric Learning:

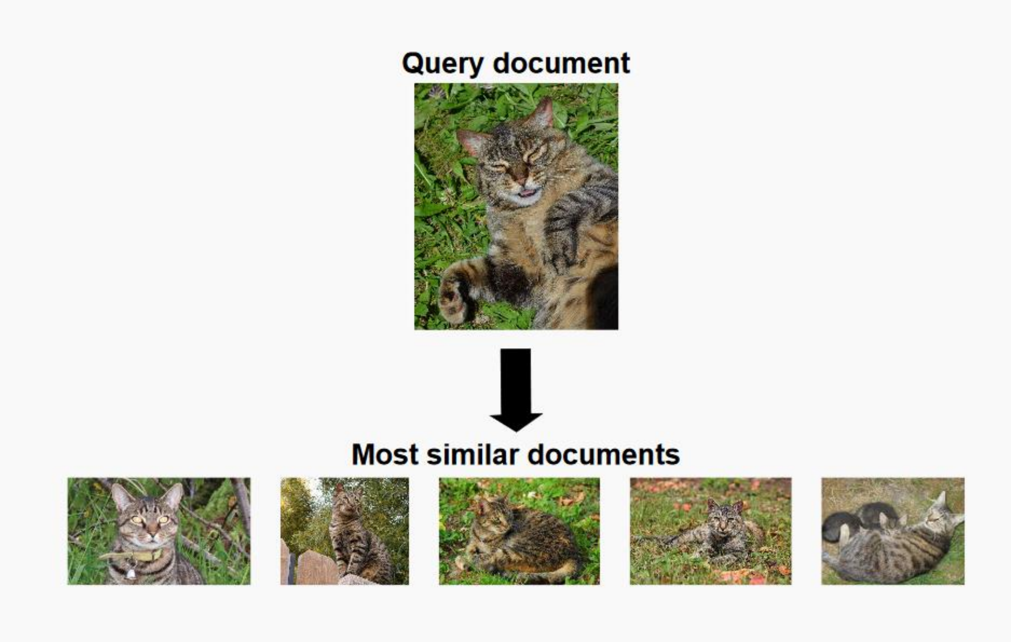


INTRODUCTION

- Recipe for metric learning:
  - Choose parametric distance/similarity function $D_M(x, x')$
  - Collect sets of data pairs/triplets with similarity information:
    - $S = \{(x_i, x_j)$: x_i and x_j are similar$\}$
    - $D = \{(x_i, x_j)$: x_i and x_j are dissimilar$\}$
    - $R = \{(x_i, x_j, x_R)$: x_i is more similar to x_j than to $x_R\}$
  - Estimate best parameters:
    - $\widehat{M} = arg \min_{M}[L(M, S, D, R) + l\, reg(M)]$

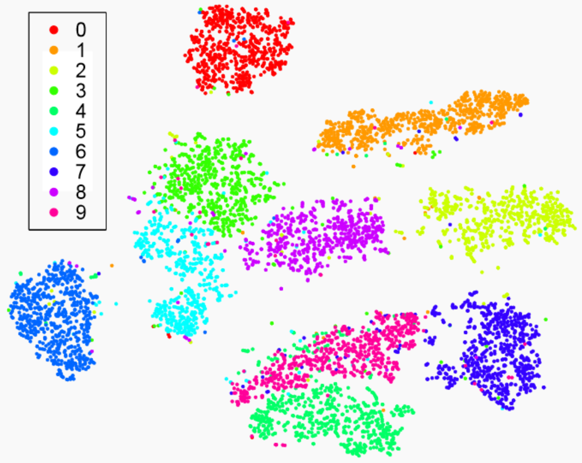
MOTIVATION

- Zero-shot learning
- Information retrieval



MOTIVATION

- Visualization of high-dimensional data
- Graceful scaling to instances with millions of classes
- Finetuning looses information and destroys intra- and inter-class variation



(image taken from [van der Maaten and Hinton, 2008])

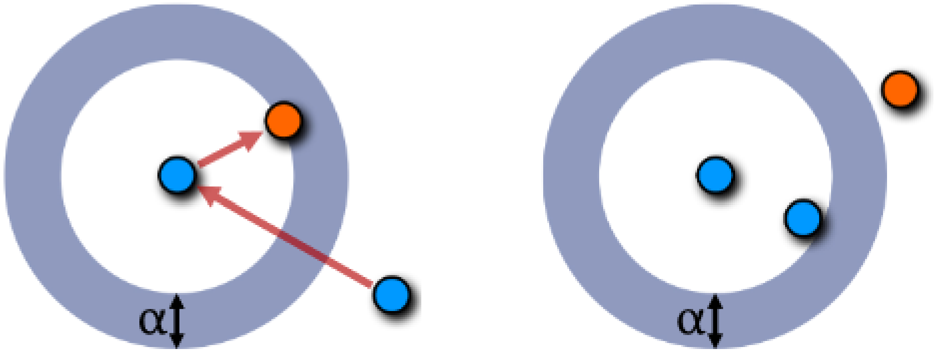
CHALLENGES

- **Issue #1** (predefined target neighborhood structure):
  - Semantic similarity ⇔ cluster classes together
  - Local similarity: target neighbors in same class (*in prior*)
  - Problem: similarity is defined in original input space
  - → Adaptive definition of similarity

CHALLENGES

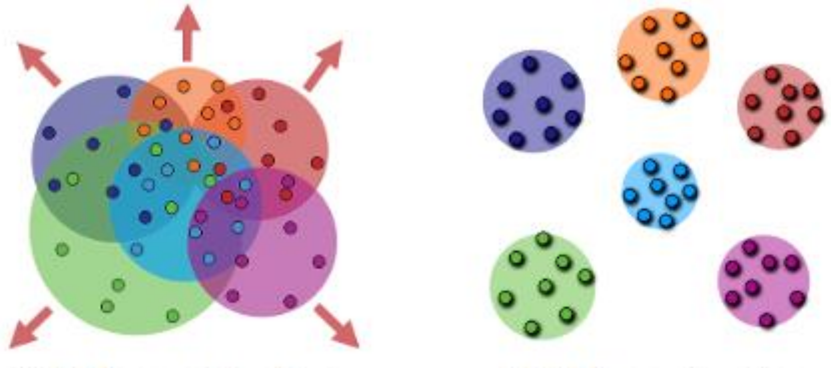
- **Issue #2** (objective formulation):
  - Popular class of DML: Triplet Loss

$$\mathscr{L}_{\text{triplet}}(\boldsymbol{\Theta}) = \frac{1}{M}\sum_{m=1}^{M}\left\{\left\|\mathbf{r}_m - \mathbf{r}_m^-\right\|_2^2 - \left\|\mathbf{r}_m - \mathbf{r}_m^+\right\|_2^2 + \alpha\right\}_+$$



(a) Triplet: before. (b) Triplet: after.

- Alternative: inform algorithm about distribution of different classes → Magnet Loss



(c) Magnet: before. (d) Magnet: after.

MODEL

- Parameterized map to representation space

$$\mathbf{r}_n = \mathbf{f}(\mathbf{x}_n; \boldsymbol{\Theta})$$

- Cluster dataset (K-Means)

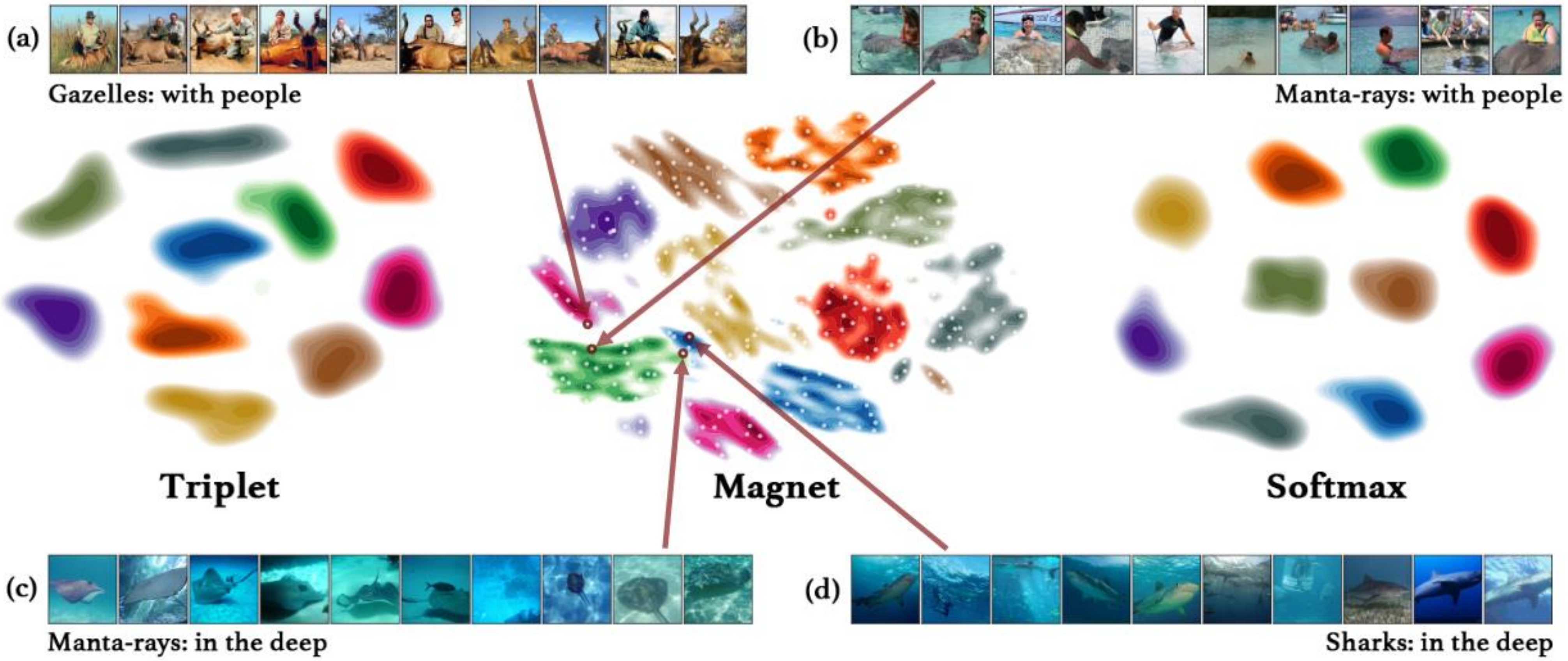
$$\mathcal{I}_1^c, \ldots, \mathcal{I}_K^c = \arg \min_{I_1^c, \ldots, I_K^c} \sum_{k=1}^{K} \sum_{\mathbf{r} \in I_k^c} \|\mathbf{r} - \boldsymbol{\mu}_k^c\|_2^2 ,$$

$$\boldsymbol{\mu}_k^c = \frac{1}{|I_k^c|} \sum_{\mathbf{r} \in I_k^c} \mathbf{r} .$$

- Objective function:

$$\mathscr{L}(\boldsymbol{\Theta}) = \frac{1}{N} \sum_{n=1}^{N} \left\{ -\log \frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{r}_n - \boldsymbol{\mu}(\mathbf{r}_n)\|_2^2 - \alpha}}{\sum_{c \neq C(\mathbf{r}_n)} \sum_{k=1}^{K} e^{-\frac{1}{2\sigma^2}\|\mathbf{r}_n - \boldsymbol{\mu}_k^c\|_2^2}} \right\}_+$$

MAGNET LOSS



(a)
Gazelles: with people
(b)
Manta-rays: with people
Triplet
Magnet
Softmax
(c)
Manta-rays: in the deep
(d)
Sharks: in the deep

TRAINING

- **Component #1**:
- Sample initial cluster
- Get nearest clusters
- Compute Loss
- Backpropagate

- **Component #2**:
- Stop training
- Compute forward pass
- Index representations by clusters

- → Representations Space is changed during training

EXPERIMENTS

- Fine grained classification
  - Datasets with similar classes
  - Comparison of Magnet Loss, Triplet Loss and Softmax
  - kNN for Triplet Loss evaluation
  - kNC for Magnet Loss evaluation

RESULTS

Approach	Error
Angelova & Long	51.7%
Gavves et al.	49.9%
Xie et al.	43.0%
Gavves et al.	43.0%
Qian et al.	30.9%
Softmax	26.6%
Triplet	35.8%
Magnet	**24.9%**

(a) Stanford Dogs.

Approach	Error
Angelova & Zhu	23.3%
Angelova & Long	19.6%
Murray & Perronnin	15.4%
Sharif Razavian et al.	13.2%
Qian et al.	11.6%
Softmax	11.2%
Triplet	17.0%
Magnet	**8.6%**

(b) Oxford 102 Flowers.

Approach	Error
Angelova & Zhu	49.2%
Parkhi et al.	46.0%
Angelova & Long	44.6%
Murray & Perronnin	43.2%
Qian et al.	19.6%
Softmax	11.3%
Triplet	13.5%
Magnet	**10.6%**

(c) Oxford-IIIT Pet.

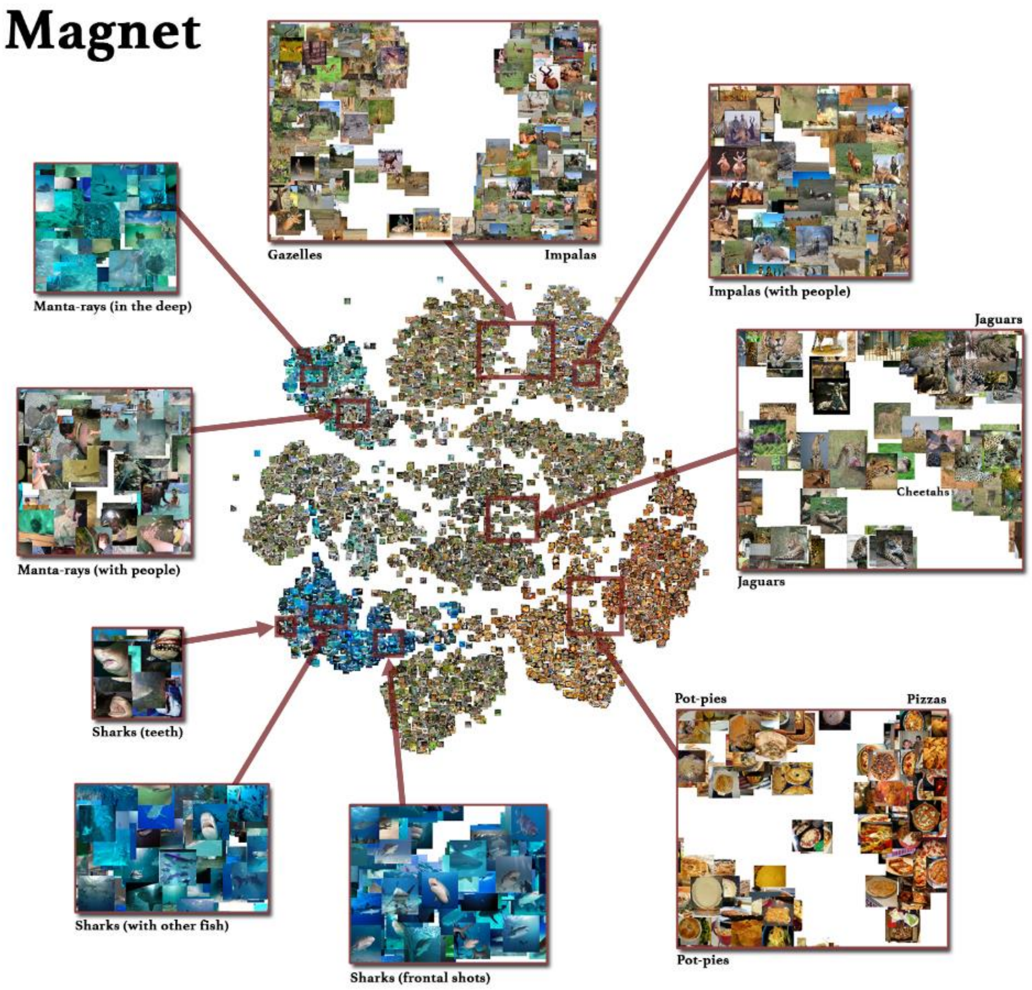
EXPERIMENTS

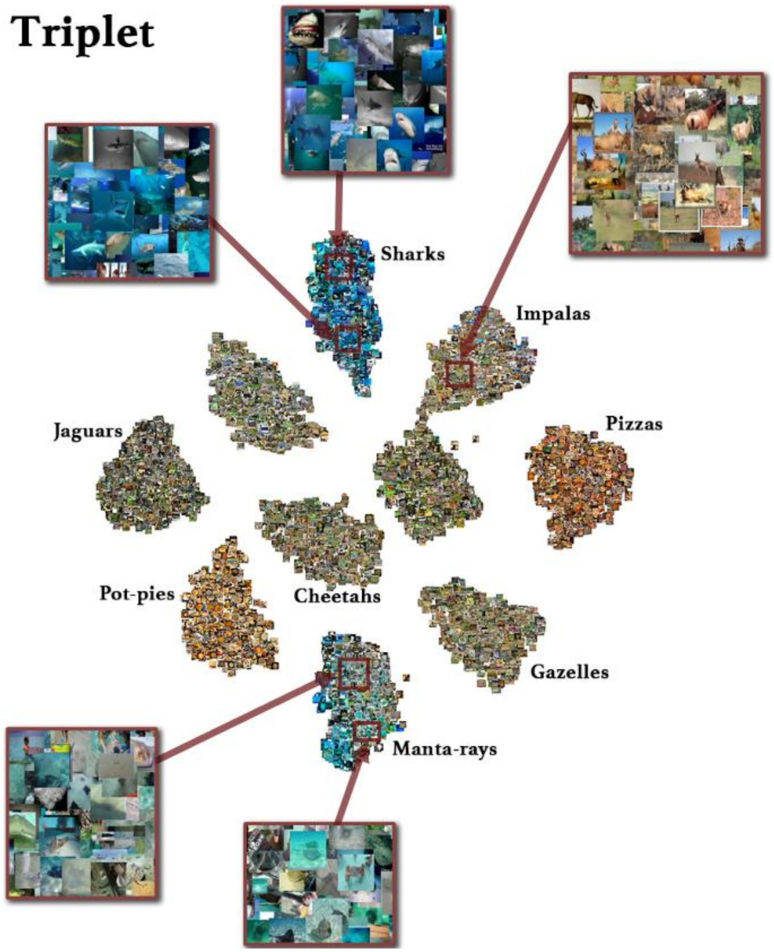
- Hierarchy recovery
  - Collapse finer grained classes to one class for training
  - Test performance in discovering intra-class variances

Approach	Error@1	Error@5
Softmax	30.9%	15.0%
Triplet	44.6%	23.4%
Magnet	**28.6%**	**7.8%**

(f) Hierarchy recovery on ImageNet Attributes.

RESULTS


Magnet
Gazelles
Impalas
Impalas (with people)
Manta-rays (in the deep)
Manta-rays (with people)
Jaguars
Cheetahs
Jaguars
Sharks (teeth)
Sharks (with other fish)
Sharks (frontal shots)
Pot-pies
Pizzas
Pot-pies


Triplet
Sharks
Impalas
Jaguars
Pizzas
Pot-pies
Cheetahs
Gazelles
Manta-rays

SOURCES

- Oren Rippel, Manohar Paluri, Piotr Dollar, and Lubomir Bourdev. Metric learning with adaptive density discrimination. International Conference on Learning Representations, 2016
- Weinberger, Kilian Q. and Saul, Lawrence K. Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res., 10:207–244, June 2009. ISSN 1532-4435.
- http://www.is.informatik.uni-duisburg.de/courses/ir_ss05/folien/colorsearch-ho.pdf
- http://researchers.lille.inria.fr/abellet/talks/metric_learning_tutorial_ECML_PKDD.pdf
- Image sources:
- https://en.wikipedia.org/wiki/File:Sunset_2007-1.jpg
- https://www.atlantisbahamas.com/thingstodo/waterpark/beaches
- http://researchers.lille.inria.fr/abellet/talks/metric_learning_tutorial_ECML_PKDD.pdf
- "Metric Learning with Adaptive Density Discrimination" by Oren Rippel, Manohar Paluri, Piotr Dollar, Lubomir Bourdev in 2016, p.3, 4, 8, 11, 12