Politecnico di Torino

Image Retrieval for Visual Geolocalization: Methods and Experiments

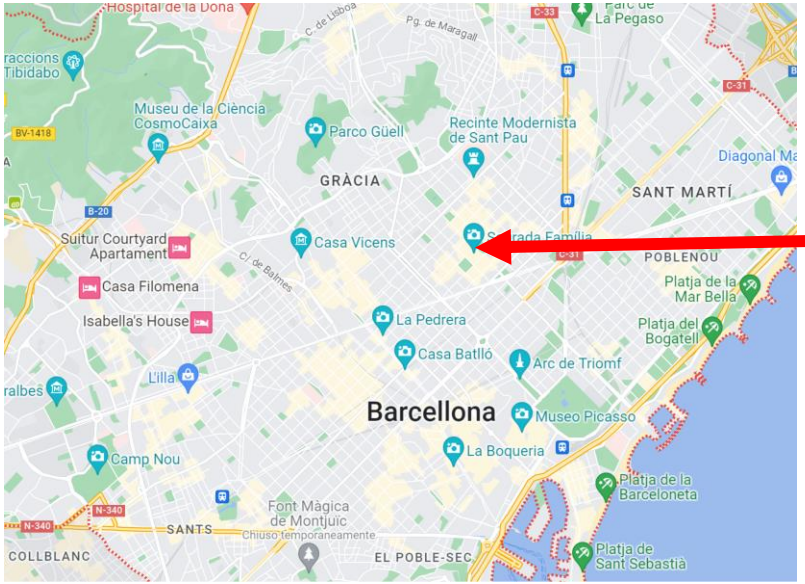
Curello Salvatore Junior s268066
Rogato Simone s289863
Cirigliano Antonio s275053

MLDL

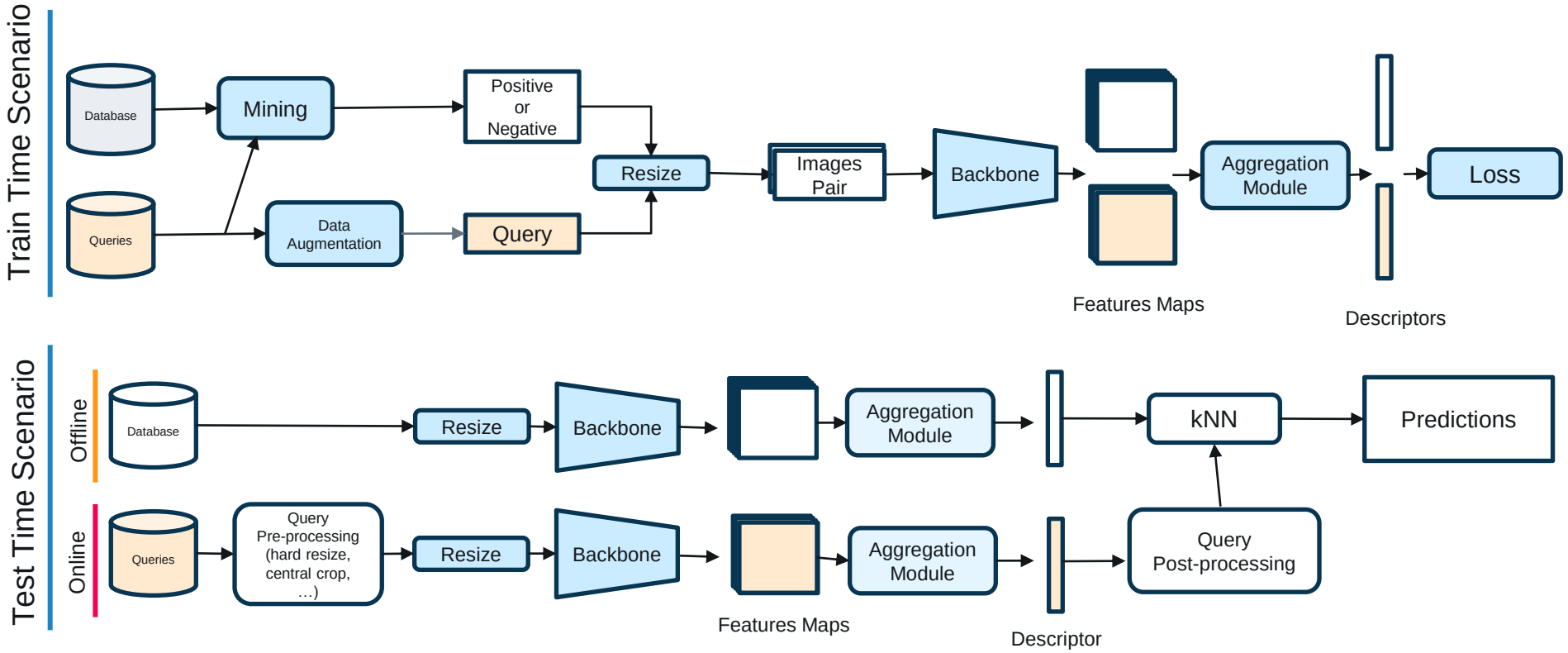
A.A. 2022/2023

Visual Place Recognition (VPR)

Visual Place Recognition (or Visual Geo-localization) is the task of finding the location where a given photo was taken.



VPR Pipeline



Deep Visual Geo-localization Benchmark, G. Berton, R. Mereu, G. Trivigno, C. Masone, G. Csurka, T. Sattler, B. Caputo

Datasets and Metric

- Datasets:
  - GSV-XS (Training)
  - SF-XS (Validation)
  - SF-XS and Tokyo-XS (Test)

Datasets and Metric

- Datasets:
  - GSV-XS (Training)
  - SF-XS (Validation)
  - SF-XS and Tokyo-XS (Test)

Example of queries in Tokyo-XS

Datasets and Metric

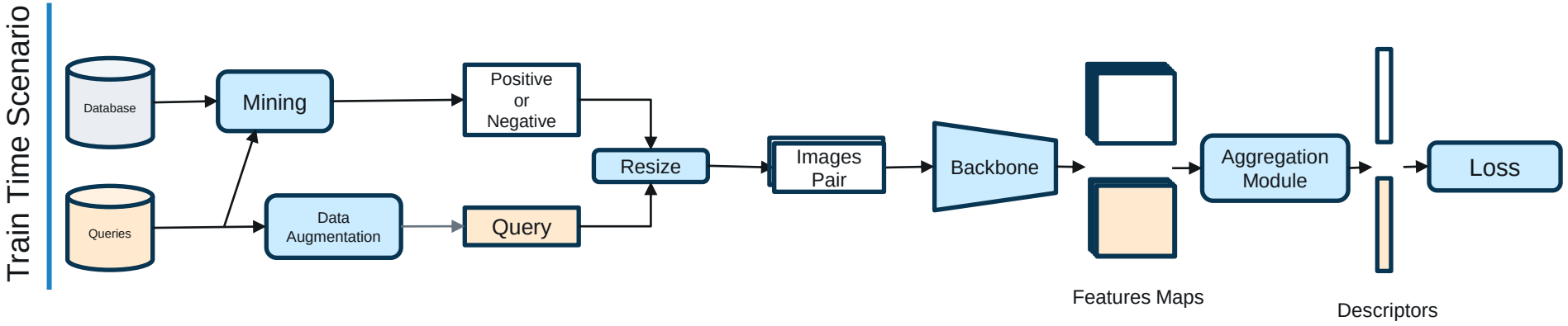
- **Datasets:**
  - **GSV-XS (Training)**
  - **SF-XS (Validation)**
  - **SF-XS and Tokyo-XS (Test)**

Example of queries in Tokyo-XS

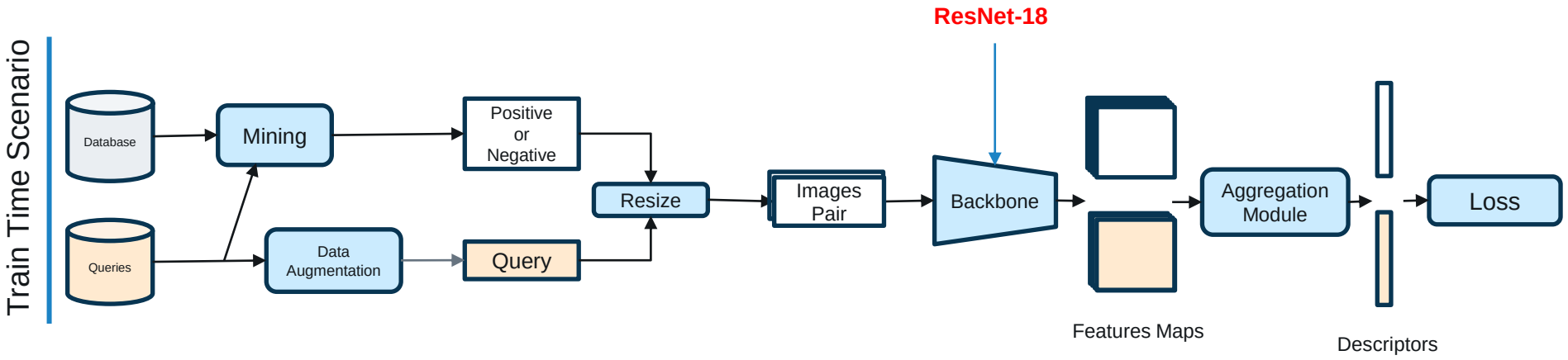
- **Metric:**

  Recall@N ⟶ percentage of queries with one of the first N predictions within 25 meters

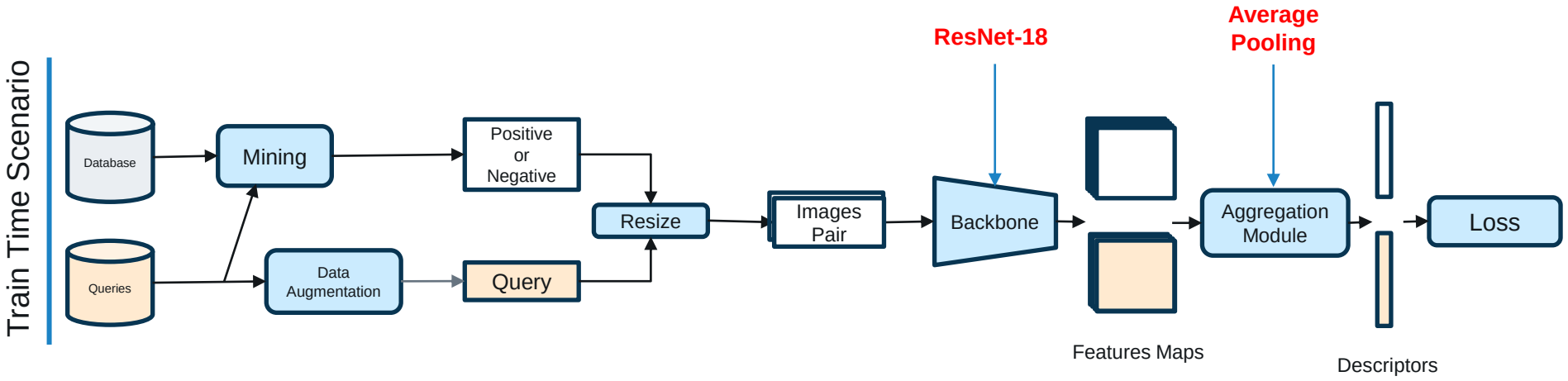
Baseline


Train Time Scenario
Database
Queries
Mining
Data Augmentation
Positive or Negative
Query
Resize
Images Pair
Backbone
Features Maps
Aggregation Module
Descriptors
Loss

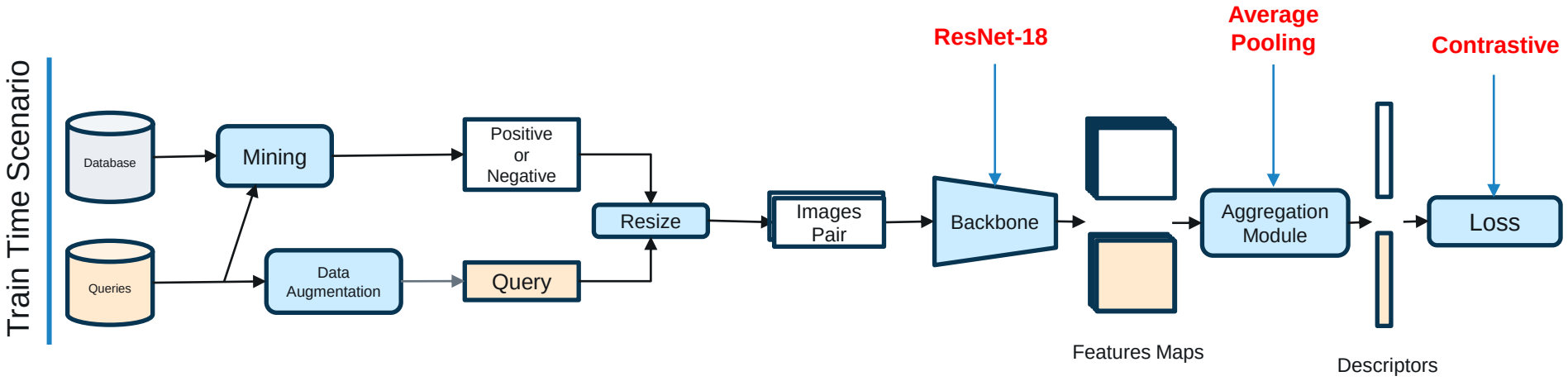
Baseline


Train Time Scenario
Database
Queries
Mining
Data Augmentation
Positive or Negative
Query
Resize
Images Pair
ResNet-18
Backbone
Features Maps
Aggregation Module
Descriptors
Loss

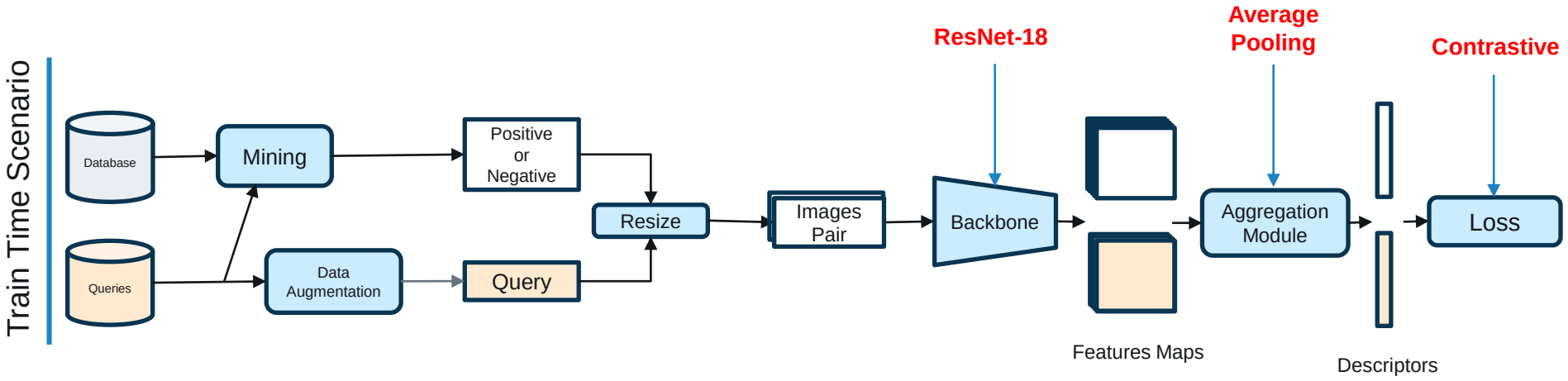
Baseline


Train Time Scenario
Database
Queries
Mining
Data Augmentation
Positive or Negative
Query
Resize
Images Pair
ResNet-18
Backbone
Features Maps
Average Pooling
Aggregation Module
Descriptors
Loss

Baseline

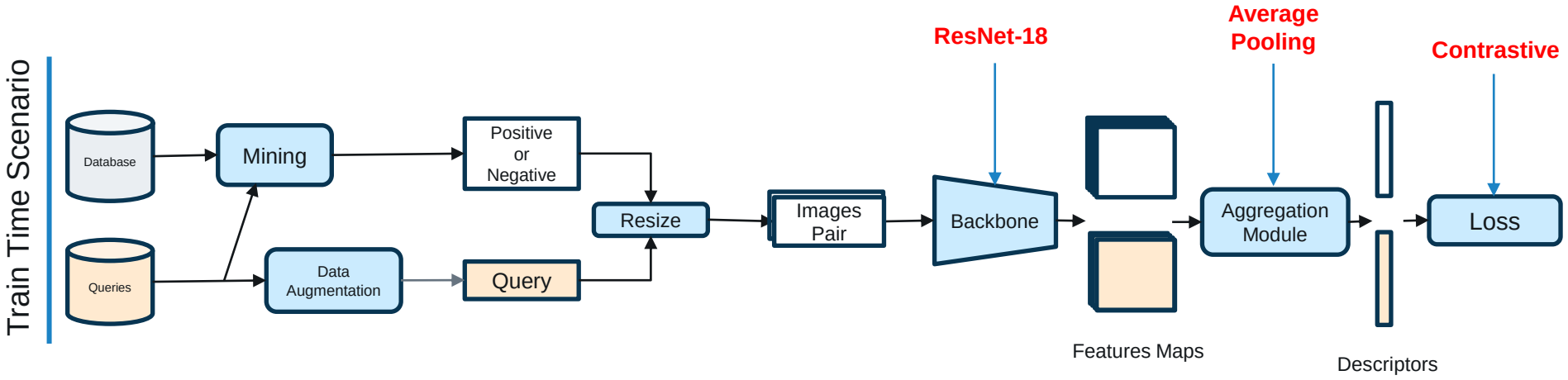

Train Time Scenario
Database
Queries
Mining
Data Augmentation
Positive or Negative
Query
Resize
Images Pair
ResNet-18
Backbone
Features Maps
Average Pooling
Aggregation Module
Descriptors
Contrastive
Loss

Baseline



- **Stochastic Gradient Descent (SGD) as optimizer.**

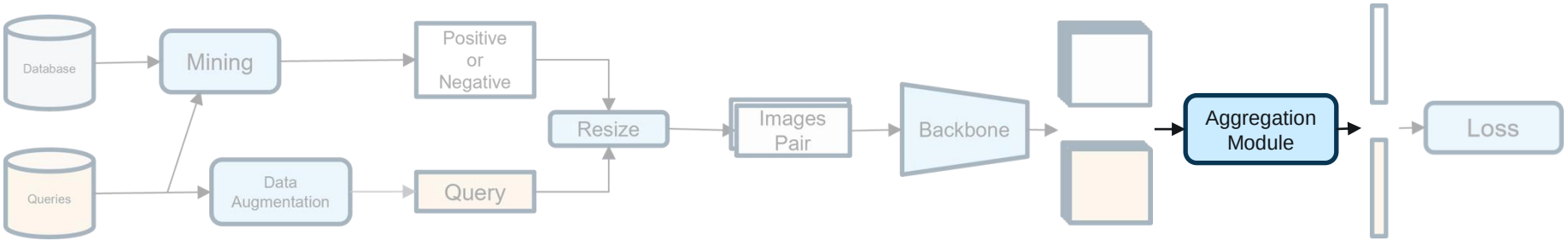
Baseline



- **Stochastic Gradient Descent (SGD) as optimizer.**
- **Number of epochs: 20**

	San Francisco		Tokyo	
	R@1	R@5	R@1	R@5
Baseline (Avg)	30.60	46.80	47.61	61.58

AGGREGATION MODULE



Aggregators

- **GeM (Generalized mean)**

$$\mathbf{f}^{(g)} = [\mathrm{f}_1^{(g)} \dots \mathrm{f}_k^{(g)} \dots \mathrm{f}_K^{(g)}]^\top, \quad \mathrm{f}_k^{(g)} = \left(\frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x^{p_k} \right)^{\frac{1}{p_k}}$$

Aggregators

- **GeM (Generalized mean)**

$$\mathbf{f}^{(g)} = [\mathrm{f}_1^{(g)} \dots \mathrm{f}_k^{(g)} \dots \mathrm{f}_K^{(g)}]^\top, \quad \mathrm{f}_k^{(g)} = \left(\frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x^{p_k} \right)^{\frac{1}{p_k}}$$

- **CosPlace**
  - L2 Normalization
  - GeM Pooling
  - Flattening
  - fc
  - L2 Normalization

Aggregators

- **GeM (Generalized mean)**

$$\mathbf{f}^{(g)} = [\mathrm{f}_1^{(g)} \dots \mathrm{f}_k^{(g)} \dots \mathrm{f}_K^{(g)}]^\top, \quad \mathrm{f}_k^{(g)} = \left(\frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x^{p_k} \right)^{\frac{1}{p_k}}$$

- **CosPlace**
  - L2 Normalization
  - GeM Pooling
  - Flattening
  - fc
  - L2 Normalization

- **Conv-AP**

$$\mathbf{z}_i = \mathtt{AAP}_{s_1 \times s_2}\big(\,\mathtt{Conv}_{1\times 1}\big(\,\mathbf{F}_i\,\big)\,\big)$$

Aggregators Results

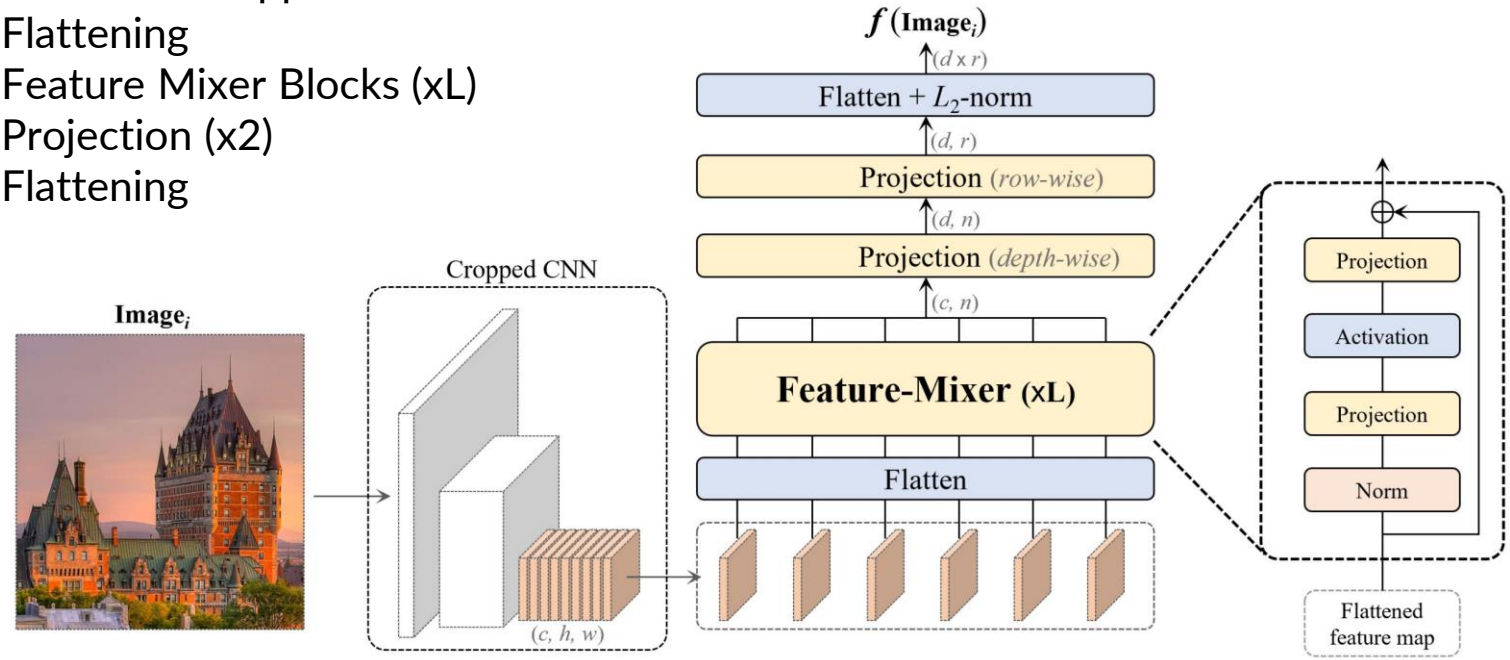
Aggregator	San Francisco		Tokyo	
	R@1	R@5	R@1	R@5
Baseline (Avg)	30.60	46.80	47.61	61.58
Gem	31.00	47.30	46.03	68.89
CosPlace	30.10	45.80	45.57	66.98
ConvAP	32.80	48.80	55.87	71.11

→ 20 epochs

All the experiments have been conducted setting the number of epochs to 10.

Aggregators

- **MixVPR**
  - Backbone cropped
  - Flattening
  - Feature Mixer Blocks (xL)
  - Projection (x2)
  - Flattening



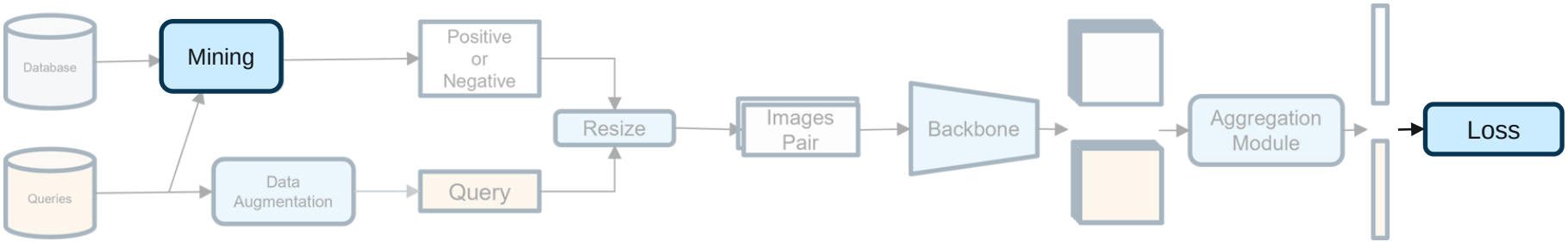
MixVPR: Feature Mixing for Visual Place Recognition, Amar Ali-bey, Brahim Chaib-draa, Philippe Giguère

MixVPR Results

Removed Layers	San Francisco		Tokyo	
	R@1	R@5	R@1	R@5
3, 4	30.80	43.50	51.11	67.62
4	**41.10**	**55.10**	**60.63**	**74.60**
None	33.00	48.30	55.87	73.65

All the subsequent experiments using MixVPR have been conducted cropping the backbone at the 4th residual block.

LOSSES and MINERS

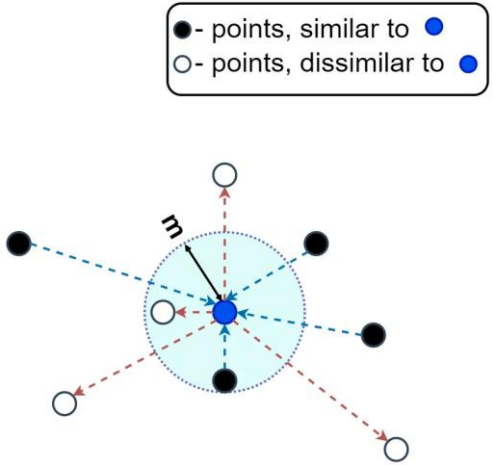


Metric learning losses

- **Contrastive loss**

$$L(W, Y, \vec{X}_1, \vec{X}_2) = (1 - Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2}\{max(0, m - D_W)\}^2$$

The aim is to minimize the distance between similar samples
and to maximize the distance between dissimilar samples over a margin m

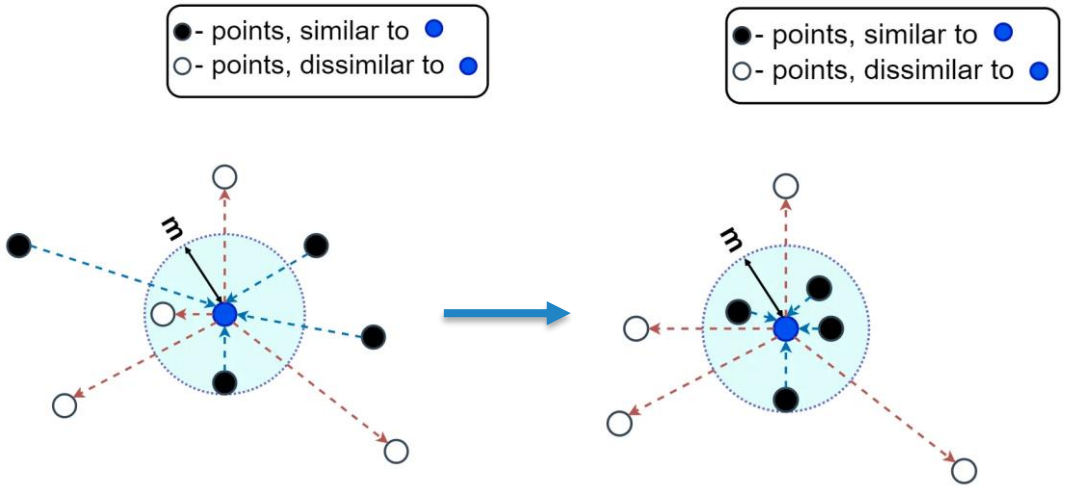


Metric learning losses

- **Contrastive loss**

$$L(W, Y, \vec{X}_1, \vec{X}_2) = (1 - Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2}\{max(0, m - D_W)\}^2$$

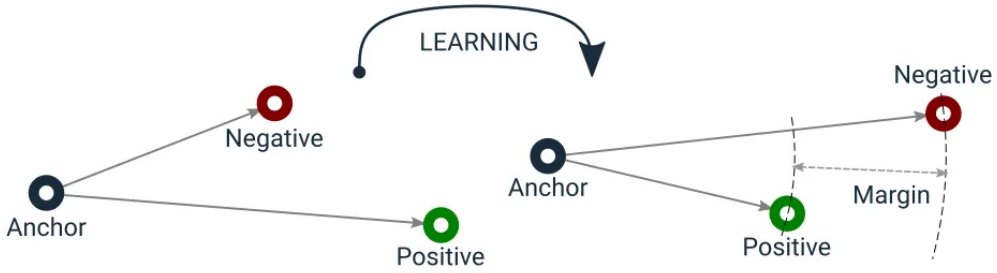
The aim is to minimize the distance between similar samples
and to maximize the distance between dissimilar samples over a margin m

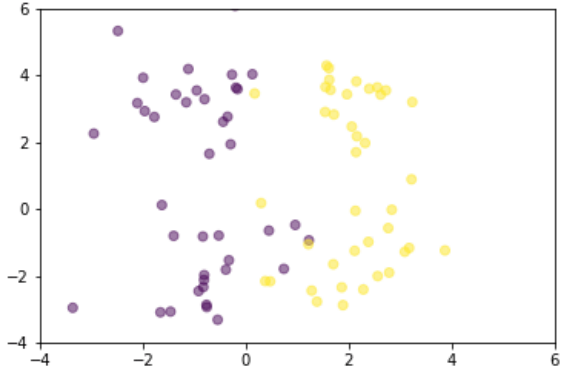


Metric learning losses

- **Triplet loss**

$$\ell(x_a, x_p, x_n) := max(\,0\,,\, m + d(x_a, x_p)\, -\, d(x_a, x_n)\,)$$

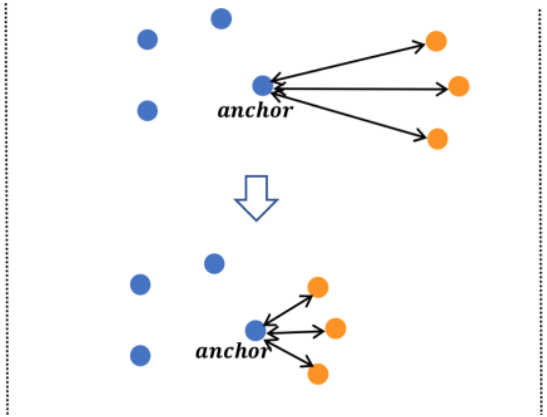




Metric learning losses

- **Multisimilarity loss**

Similarity-S

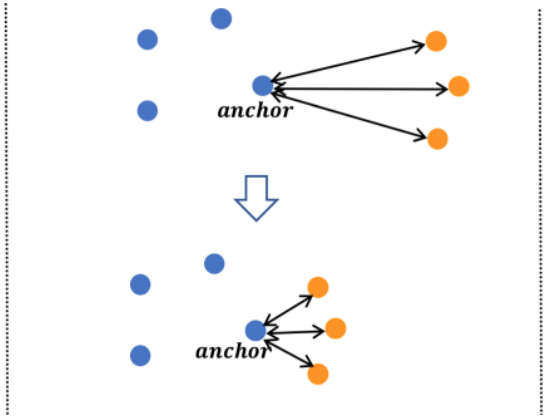


Weights of the negative pairs are increased

Metric learning losses

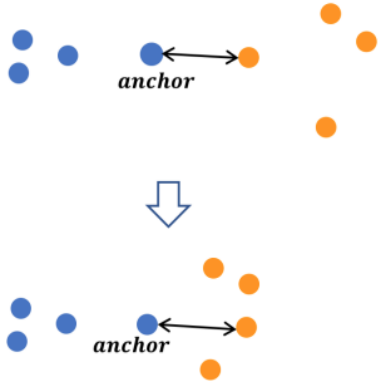
- **Multisimilarity loss**

Similarity-S



Weights of the negative pairs are increased

Similarity-N



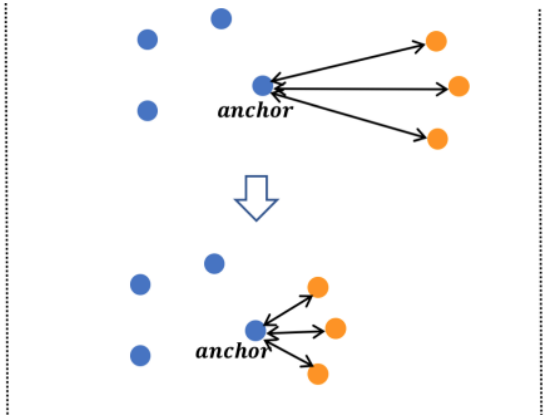
Relative weight of the reference negative pair is decreased

Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R. Scott. Multi-similarity loss with general pair weighting for deep metric learning, 2020

Metric learning losses

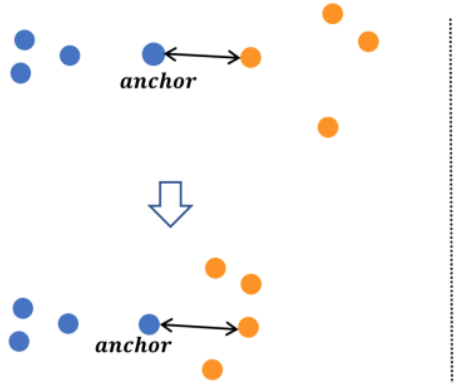
- **Multisimilarity loss**

Similarity-S



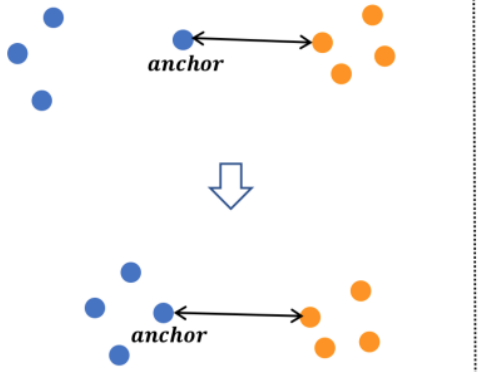
Weights of the negative pairs are increased

Similarity-N



Relative weight of the reference negative pair is decreased

Similarity-P

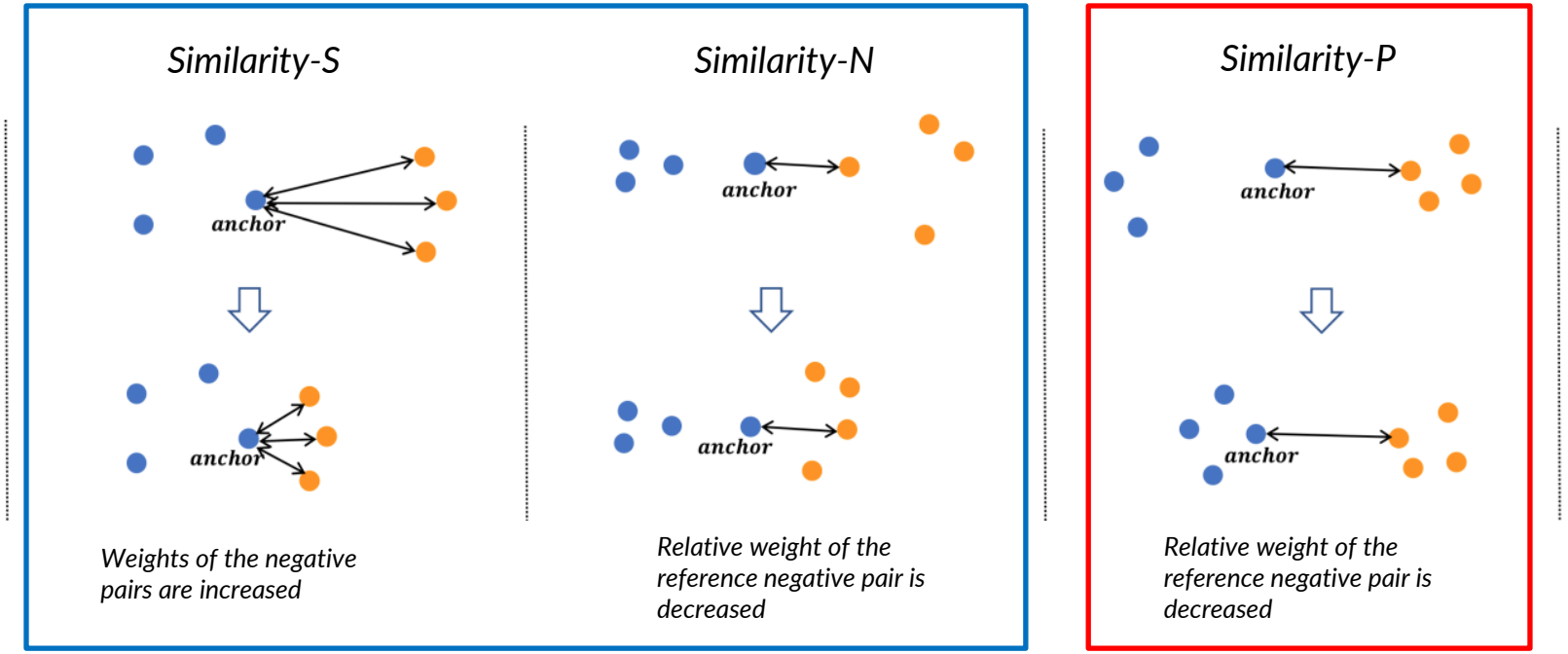


Relative weight of the reference negative pair is decreased

Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R. Scott. Multi-similarity loss with general pair weighting for deep metric learning, 2020

Metric learning losses

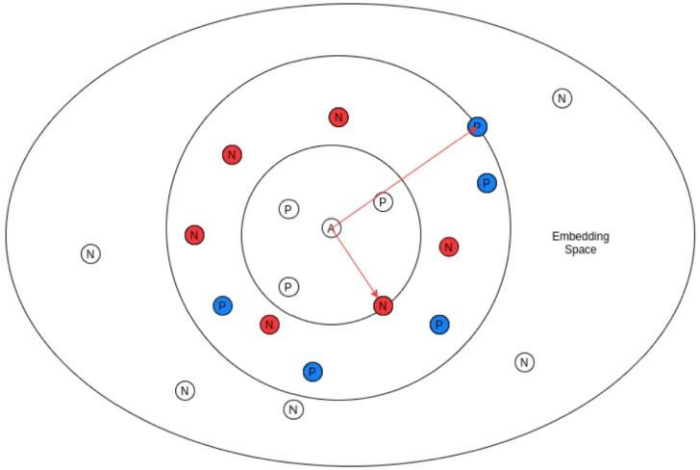
- **Multisimilarity loss**



Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R. Scott. Multi-similarity loss with general pair weighting for deep metric learning, 2020

Miners

- **Multisimilarity miner**

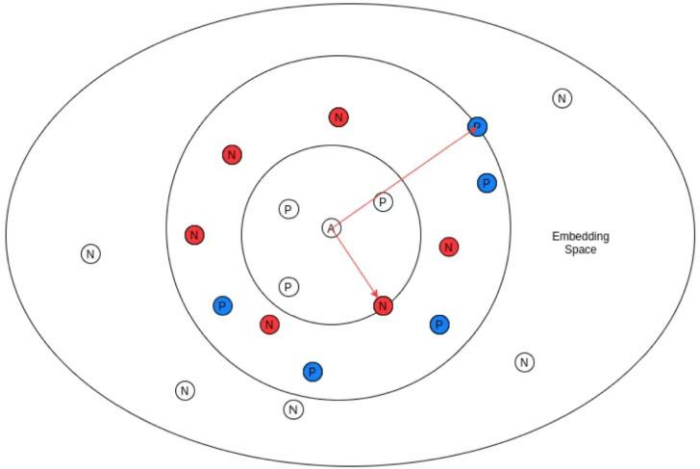


- Negative pairs whose similarity exceeds the similarity of the hardest positive by a margin.

- Positive pairs whose similarity is less then the similarity of the hardest negative by a margin.

Miners

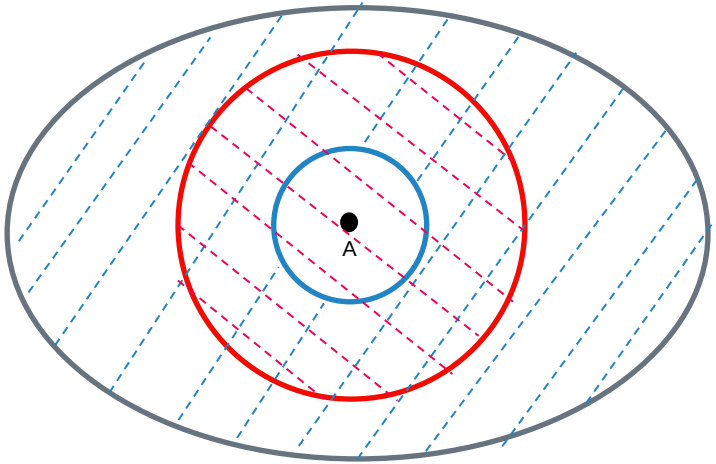
- **Multisimilarity miner**



- Negative pairs whose similarity exceeds the similarity of the hardest positive by a margin.

- Positive pairs whose similarity is less then the similarity of the hardest negative by a margin.

- **PairMargin miner**



- Negative pairs whose negative samples are within a certain margin (area in red).

- Positive pairs whose positive samples are over a certain margin (area in blu).

Losses&Miners comparison

Loss	Miner	San Francisco		Tokyo	
		R@1	R@5	R@1	R@5
Contrastive	No miner	32.80	48.80	55.87	71.11
Contrastive	Multisimilarity	33.20	49.00	49.52	69.21
Contrastive	PairMargin	32.20	48.70	52.06	69.24
Multisimilarity	No miner	28.00	42.80	46.35	68.89
Multisimilarity	Multisimilarity	29.60	44.50	48.57	65.39
Multisimilarity	PairMargin	28.70	43.40	47.93	66.98

Using ConvAP aggregator :

- Slight improvements only on San Francisco and R@1 on Tokyo

- Overall mediocre improvements

Losses&Miners comparison

Loss	Miner	San Francisco		Tokyo	
		R@1	R@5	R@1	R@5
Contrastive	No miner	32.80	48.80	55.87	71.11
Contrastive	Multisimilarity	33.20	49.00	49.52	69.21
Contrastive	PairMargin	32.20	48.70	52.06	69.24
Multisimilarity	No miner	28.00	42.80	46.35	68.89
Multisimilarity	Multisimilarity	29.60	44.50	48.57	65.39
Multisimilarity	PairMargin	28.70	43.40	47.93	66.98

Using ConvAP aggregator :

- Slight improvements only on San Francisco and R@1 on Tokyo

- Overall mediocre improvements

Using MixVPR aggregator :

- Slightest improvements only on Tokyo

- Overall no improvements

Loss	Miner	San Francisco		Tokyo	
		R@1	R@5	R@1	R@5
Contrastive	No miner	41.10	55.10	60.63	74.60
Contrastive	Multisimilarity	38.10	53.40	58.73	74.28
Contrastive	PairMargin	39.50	52.20	60.95	75.24
Multisimilarity	No miner	38.30	54.20	62.20	75.55
Multisimilarity	Multisimilarity	36.50	51.70	53.33	72.06
Multisimilarity	PairMargin	30.30	44.90	51.68	68.57

OPTIMIZER and SCHEDULER

Combinations of Optimizers, Learning rate and Weight decay

Loss	Miner	Optimizer	Learning Rate	Weight Decay	San Francisco		Tokyo	
					R@1	R@5	R@1	R@5
Contrastive	No miner	Adam	0.0001	0.001	36.80	49.90	52.38	68.89
Contrastive	No miner	AdamW	0.0001	0.001	43.30	56.30	60.63	73.65
Contrastive	PairMargin	AdamW	0.0001	0.001	42.80	55.50	60.95	76.51
Contrastive	PairMargin	SGD	0.0001	0.001	21.90	33.80	42.85	59.68
Multisimilarity	No miner	AdamW	0.0001	0.001	51.70	63.90	65.40	82.22
Multisimilarity	No miner	AdamW	0.0001	0.0001	46.60	63.10	64.62	81.27
Multisimilarity	No miner	AdamW	0.0001	0.01	50.70	64.60	66.98	79.68
Multisimilarity	No miner	AdamW	0.0001 (*10)	0.001	49.80	64.90	70.16	83.17
Multisimilarity	No miner	AdamW	0.0001 (*5)	0.001	**51.72**	**65.10**	**70.47**	**83.81**
Multisimilarity	No miner	RAdam	0.001	0	43.00	57.70	54.28	66.67
Multisimilarity	No miner	RAdam	0.0001	0	51.30	64.31	68.25	79.68
Multisimilarity	No miner	RAdam	0.0001 (*5)	0	49.30	65.70	68.89	82.54

All the experiments have been conducted using MixVPR aggregator.

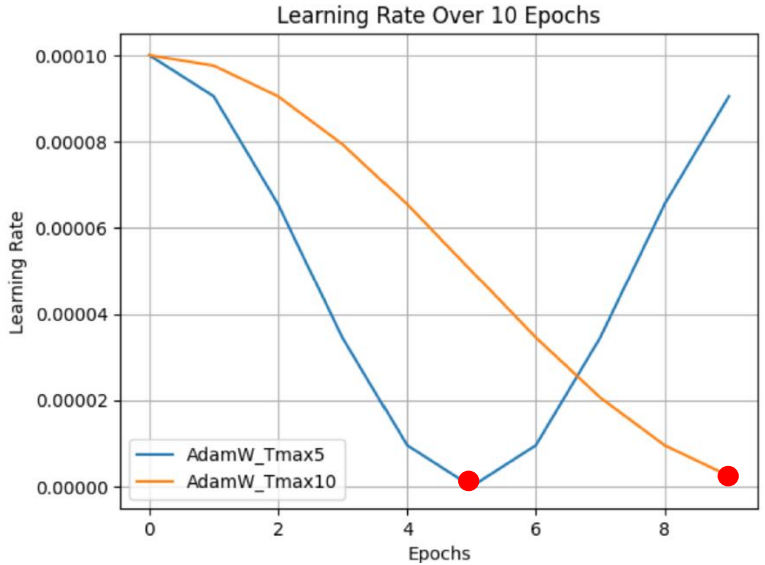
CosineAnnealingLR

Loss	Miner	Optimizer	Learning Rate	Weight Decay	San Francisco		Tokyo	
					R@1	R@5	R@1	R@5
Contrastive	No miner	Adam	0.0001	0.001	36.80	49.90	52.38	68.89
Contrastive	No miner	AdamW	0.0001	0.001	43.30	56.30	60.63	73.65
Contrastive	PairMargin	AdamW	0.0001	0.001	42.80	55.50	60.95	76.51
Contrastive	PairMargin	SGD	0.0001	0.001	21.90	33.80	42.85	59.68
Multisimilarity	No miner	AdamW	0.0001	0.001	51.70	63.90	65.40	82.22
Multisimilarity	No miner	AdamW	0.0001	0.0001	46.60	63.10	64.62	81.27
Multisimilarity	No miner	AdamW	0.0001	0.01	50.70	64.60	66.98	79.68
Multisimilarity	No miner	AdamW	0.0001 (*10)	0.001	49.80	64.90	70.16	83.17
Multisimilarity	No miner	AdamW	0.0001 (*5)	0.001	**51.72**	**65.10**	**70.47**	**83.81**
Multisimilarity	No miner	RAdam	0.001	0	43.00	57.70	54.28	66.67
Multisimilarity	No miner	RAdam	0.0001	0	51.30	64.31	68.25	79.68
Multisimilarity	No miner	RAdam	0.0001 (*5)	0	49.30	65.70	68.89	82.54

(*N) indicates that CosineAnnealingLR has been applied with T_max = N.

CosineAnnealingLR

$$\eta_t = \eta_{\min} + \frac{1}{2}\left(\eta_{\max} - \eta_{\min}\right)\left(1 + \cos\left(\frac{T_{\text{cur}}}{T_{\max}}\pi\right)\right)$$



- At epochs T_max the learnig rate is equal to zero.

The changing of the learnig rate over 10 epochs when we applied CosineAnnealingLR with AdamW optimizer.

Combinations of Optimizers, Learning rate and Weight decay

Loss	Miner	Optimizer	Learning Rate	Weight Decay	San Francisco		Tokyo	
					R@1	R@5	R@1	R@5
Contrastive	No miner	Adam	0.0001	0.001	36.80	49.90	52.38	68.89
Contrastive	No miner	AdamW	0.0001	0.001	43.30	56.30	60.63	73.65
Contrastive	PairMargin	AdamW	0.0001	0.001	42.80	55.50	60.95	76.51
Contrastive	PairMargin	SGD	0.0001	0.001	21.90	33.80	42.85	59.68
Multisimilarity	No miner	AdamW	0.0001	0.001	51.70	63.90	65.40	82.22
Multisimilarity	No miner	AdamW	0.0001	0.0001	46.60	63.10	64.62	81.27
Multisimilarity	No miner	AdamW	0.0001	0.01	50.70	64.60	66.98	79.68
Multisimilarity	No miner	AdamW	0.0001 (*10)	0.001	49.80	64.90	70.16	83.17
Multisimilarity	No miner	AdamW	0.0001 (*5)	0.001	**51.72**	**65.10**	**70.47**	**83.81**
Multisimilarity	No miner	RAdam	0.001	0	43.00	57.70	54.28	66.67
Multisimilarity	No miner	RAdam	0.0001	0	51.30	64.31	68.25	79.68
Multisimilarity	No miner	RAdam	0.0001 (*5)	0	49.30	65.70	68.89	82.54

10 epochs

This combination resulted in a remarkable increase of 21.05% and 22.86% in R@1 and 18.30% and 22.23% in R@5 on the SF-XS and Tokyo-XS dataset compared to the baseline (20 epochs).

QUALITATIVE RESULTS

Differences between Conv-AP and MixVPR

MixVPR aggregator is capable to give correct predictions even if the query is taken during the night while Conv-AP cannot.

Differences between Conv-AP and MixVPR

MixVPR aggregator is capable to give correct predictions even if the query is taken during the night while Conv-AP cannot.

The feature mixer blocks in MixVPR incorporates global relationships into each feature map.

Differences between Conv-AP and MixVPR

A query extracted from SF-XS which shows that in case of noise or weird weather in the image, MixVPR performs slightly better than Conv-AP .

For the future…

- **As future research could be interesting to repeat our experiments with a bigger training set.**

This could possibly increase the performance of CosPlace aggregator since it is not suited for training of small datasets.

For the future...

- **As future research could be interesting to repeat our experiments with a bigger training set.**

This could possibly increase the performance of CosPlace aggregator since it is not suited for training of small datasets.

- **It can be useful to analyze how the performances change by varying the distance threshold.**

A smaller value requires more precise matching.

Thank you for your attention