Smooth-AP: Smoothing the Path Towards Large-Scale Image Retrieval

Andrew Brown, Weidi Xie, Vicky Kalogeiton, Andrew Zisserman
European Conference on Computer Vision (ECCV) 2020

Presented by: Yash Patel
Visual Recognition Group, Czech Technical University in Prague

Image Retrieval

Retrieve images of the same class as the query from an image collection

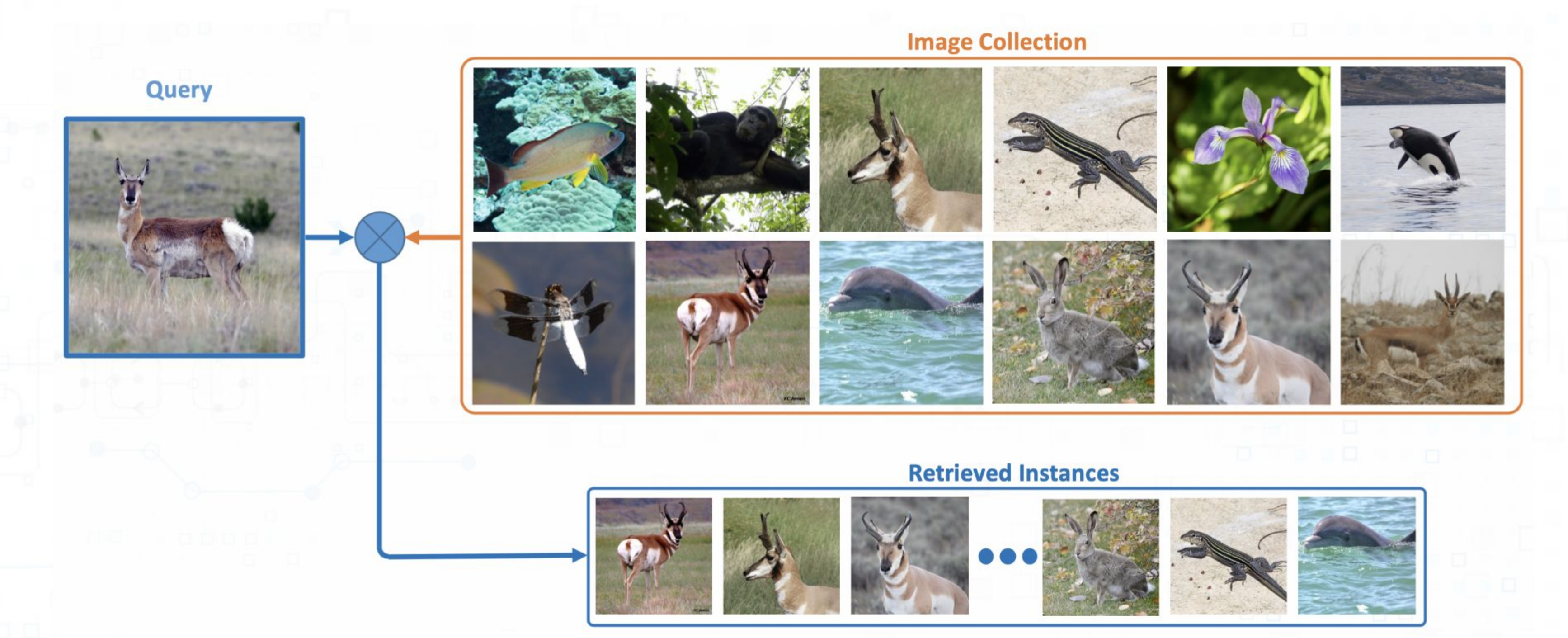


Image Retrieval

Classification:

(closed set)

Class? (animal species)

Image Retrieval

Classification:
(closed set)

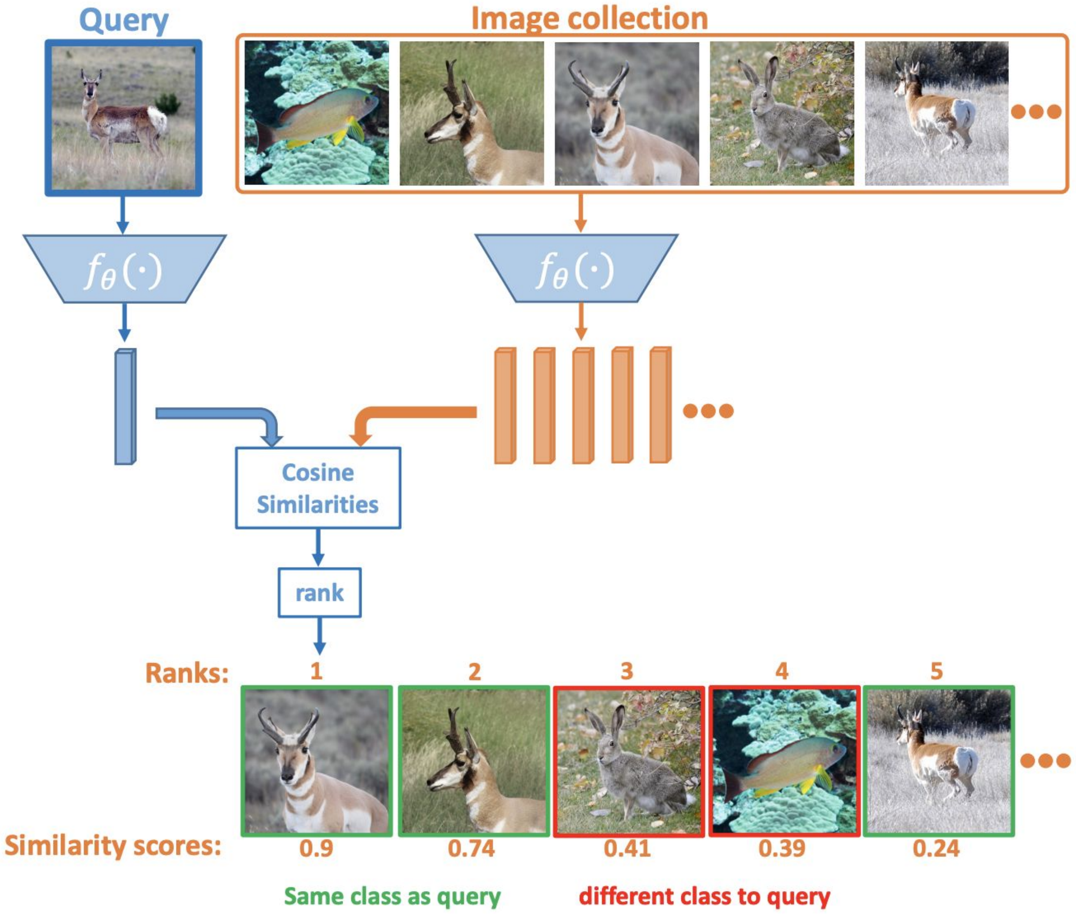
→ Class? (animal species)

Retrieval:
(open set)

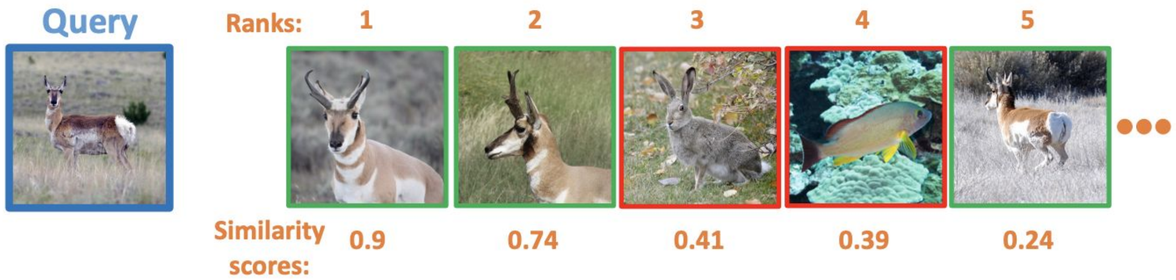
→ Retrieve all instances of the same class from retrieval set

Image Retrieval Inference

- Extract embeddings from query and image collection.
- Compute similarity scores.
- Rank according to relevance to the query.



Mean Average Precision

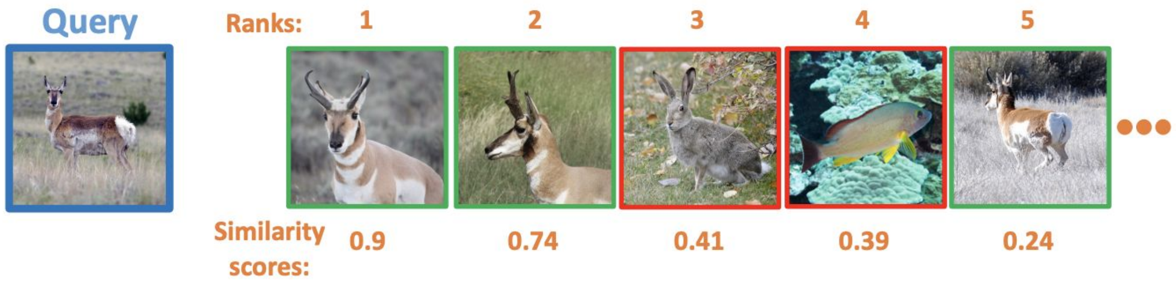


$$AP_q = \frac{1}{|S_p|} \sum_{i \in S_p} \frac{\mathcal{R}^+(i, S_p)}{\mathcal{R}(i, S_\Omega)} \quad \rightarrow \quad = \frac{1}{3}\left(\frac{1}{1} + \frac{2}{2} + \frac{3}{5}\right) \approx 0.87$$

S_Ω = retrieval set S_P = positive retrieval set $\mathcal{R}$ = ranking of instance in retrieval set $\mathcal{R}^+$= ranking of instance in positive retrieval set

Mean Average Precision

- Average Precision used to benchmark retrieval systems.
- **Non-differentiable ranking**, so cannot train end-to-end directly.
- Goal – Optimise a smoothed version of the Average Precision Metric.

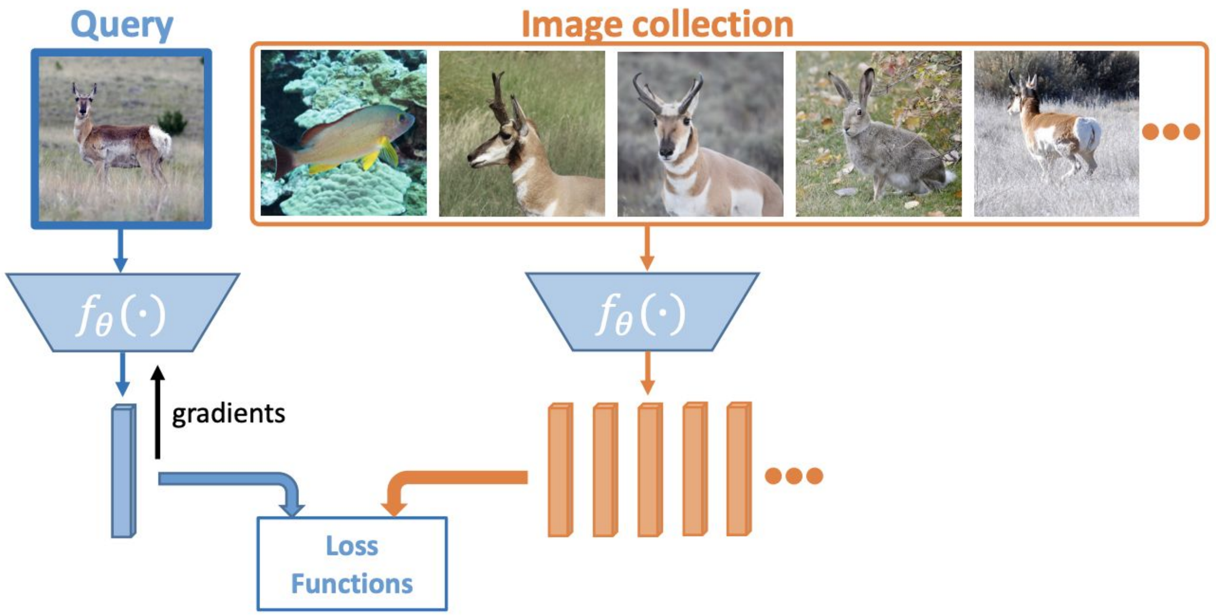


$$AP_q = \frac{1}{|S_p|} \sum_{i \in S_p} \frac{\mathcal{R}^+(i, S_p)}{\mathcal{R}(i, S_\Omega)} \rightarrow = \frac{1}{3}\left(\frac{1}{1} + \frac{2}{2} + \frac{3}{5}\right) \approx 0.87$$

S_Ω = retrieval set S_P = positive retrieval set $\mathcal{R}$ = ranking of instance in retrieval set $\mathcal{R}^+$ = ranking of instance in positive retrieval set

Image Retrieval Training

- Embedding must be trained for good ranking.
- Achieved using loss functions.

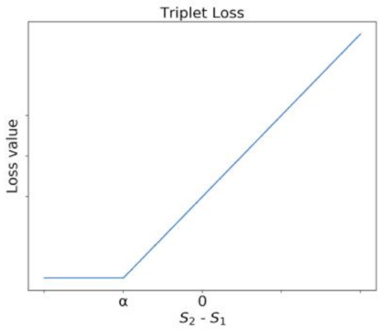


Triplet Loss

Differentiable?

Optimises Ranking metric?

Ranking Surrogate loss
(e.g. Triplet Loss)



$$\mathcal{L}_{triplet} \propto \max(S_2 - S_1 + \alpha, 0)$$

✓

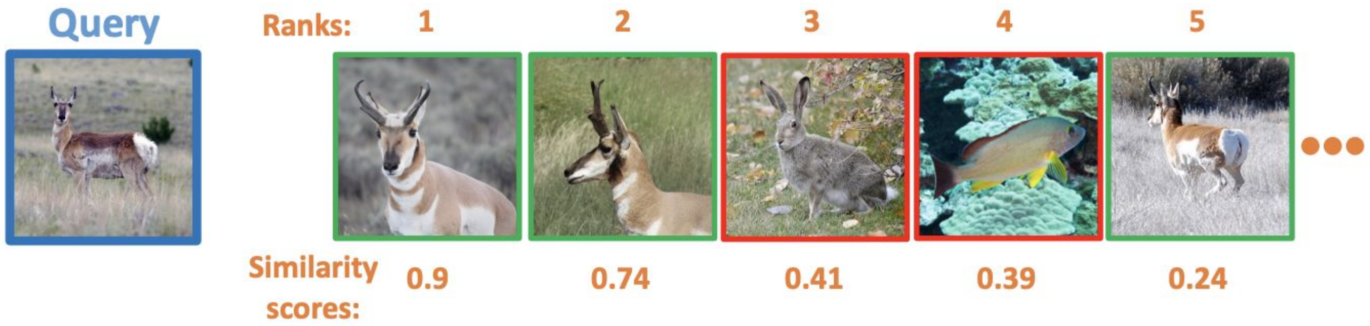
✗

S_2 = positive instance relevance score

S_1 = negative instance relevance score

Smoothing the Average Precision Loss

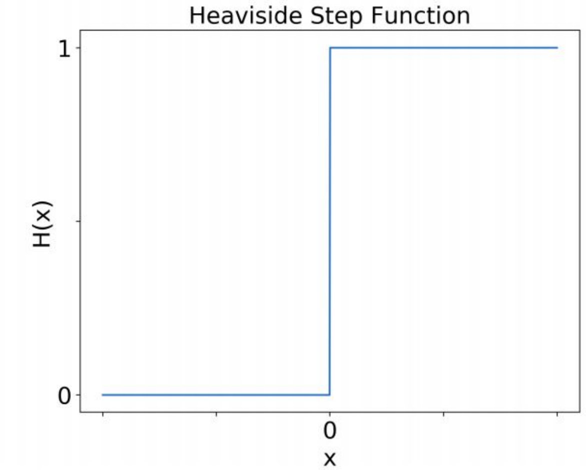
- Non-differentiable ranking.
- Find the rank of the first number.



0.24	0	1
0.74	0.50	1
0.39	0.15	1
0.41	0.17	1
0.90	0.56	1

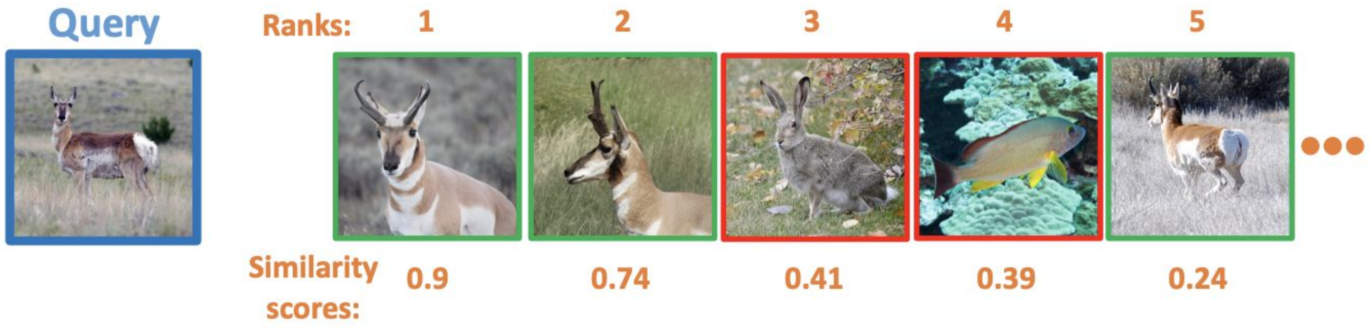
$\xrightarrow{-0.24}$ $\xrightarrow{H(x)}$ $\xrightarrow{\text{sum}}$ 5

$$H(x) = \begin{cases} 0 & if\ x < 0 \\ 1 & if\ x \geq 0 \end{cases}$$



Smoothing the Average Precision Loss

- Non-differentiable ranking.
- Find the rank of the first number.



0.24
0.74
0.39
0.41
0.90

$\xrightarrow{-0.74}$

-0.50
0
-0.35
-0.33
0.16

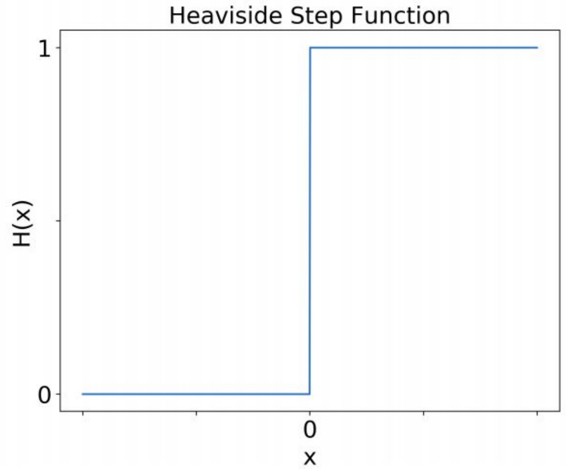
$\xrightarrow{H(x)}$

0
1
0
0
1

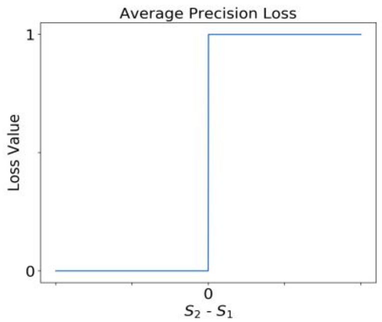
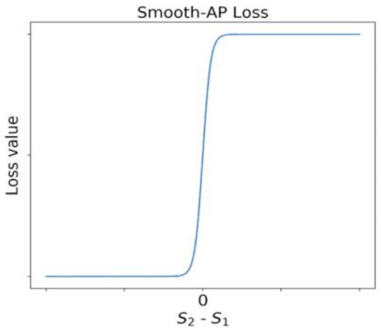
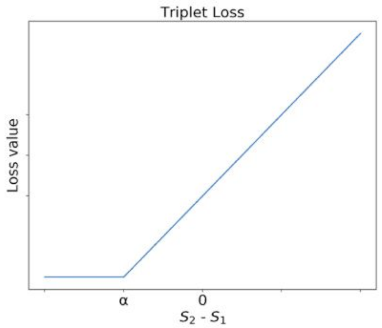
$\xrightarrow{\text{sum}}$

2

$$H(x) = \begin{cases} 0 & if\ x < 0 \\ 1 & if\ x \geq 0 \end{cases}$$



Smooth-AP

			Differentiable?	Optimises Ranking metric?
Average Precision Loss $\mathcal{L}_{AP} = (1 - AP_q)$		$\mathcal{L}_{AP} \propto H(x) = \begin{cases} 0 & if\ x < 0 \\ 1 & if\ x \geq 0 \end{cases}$	✗	✓
Smooth-AP Loss		$\mathcal{L}_{Smooth-AP} \propto G(x) = \dfrac{1}{1 + e^{\frac{-x}{\tau}}}$	✓	✓
Ranking Surrogate loss (e.g. Triplet Loss)		$\mathcal{L}_{triplet} \propto \max(S_2 - S_1 + \alpha, 0)$	✓	✗

S_2 = positive instance relevance score S_1 = negative instance relevance score

Smooth-AP

Average precision

$$AP_q = \frac{1}{|\mathcal{S}_P|} \sum_{i \in \mathcal{S}_P} \frac{1 + \sum_{j \in \mathcal{S}_p, j \neq i} \mathbb{1}\{D_{ij} > 0\}}{1 + \sum_{j \in \mathcal{S}_P, j \neq i} \mathbb{1}\{D_{ij} > 0\} + \sum_{j \in \mathcal{S}_N} \mathbb{1}\{D_{ij} > 0\}}$$

Smooth average precision

$$\mathcal{G}(x; \tau) = \frac{1}{1 + e^{\frac{-x}{\tau}}}.$$

$$AP_q \approx \frac{1}{|\mathcal{S}_P|} \sum_{i \in \mathcal{S}_P} \frac{1 + \sum_{j \in \mathcal{S}_P} \mathcal{G}(D_{ij}; \tau)}{1 + \sum_{j \in \mathcal{S}_P} \mathcal{G}(D_{ij}; \tau) + \sum_{j \in \mathcal{S}_N} \mathcal{G}(D_{ij}; \tau)}$$

$$\mathcal{L}_{AP} = \frac{1}{m} \sum_{k=1}^{m} (1 - AP_k)$$

Smooth-AP

Smooth average precision

$$\mathcal{G}(x;\tau) = \frac{1}{1 + e^{\frac{-x}{\tau}}}.$$

Average precision

$$AP_q = \frac{1}{|\mathcal{S}_P|} \sum_{i \in \mathcal{S}_P} \frac{1 + \sum_{j \in \mathcal{S}_p, j \neq i} \mathbb{1}\{D_{ij} > 0\}}{1 + \sum_{j \in \mathcal{S}_P, j \neq i} \mathbb{1}\{D_{ij} > 0\} + \sum_{j \in \mathcal{S}_N} \mathbb{1}\{D_{ij} > 0\}}$$

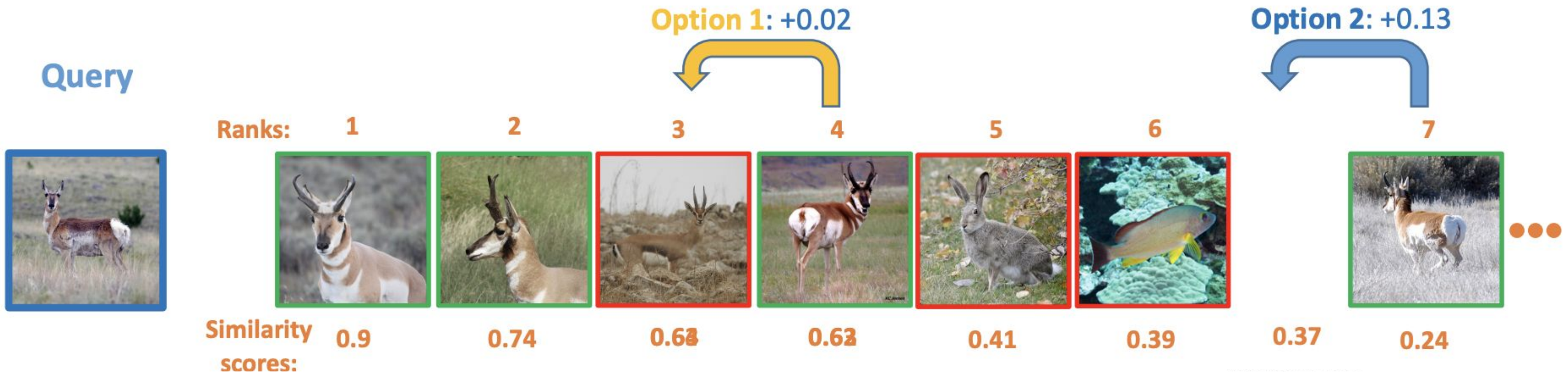
$$AP_q \approx \frac{1}{|\mathcal{S}_P|} \sum_{i \in \mathcal{S}_P} \frac{1 + \sum_{j \in \mathcal{S}_P} \mathcal{G}(D_{ij};\tau)}{1 + \sum_{j \in \mathcal{S}_P} \mathcal{G}(D_{ij};\tau) + \sum_{j \in \mathcal{S}_N} \mathcal{G}(D_{ij};\tau)}$$

$$\mathcal{L}_{AP} = \frac{1}{m} \sum_{k=1}^{m} (1 - AP_k)$$

Change in loss = 0.5

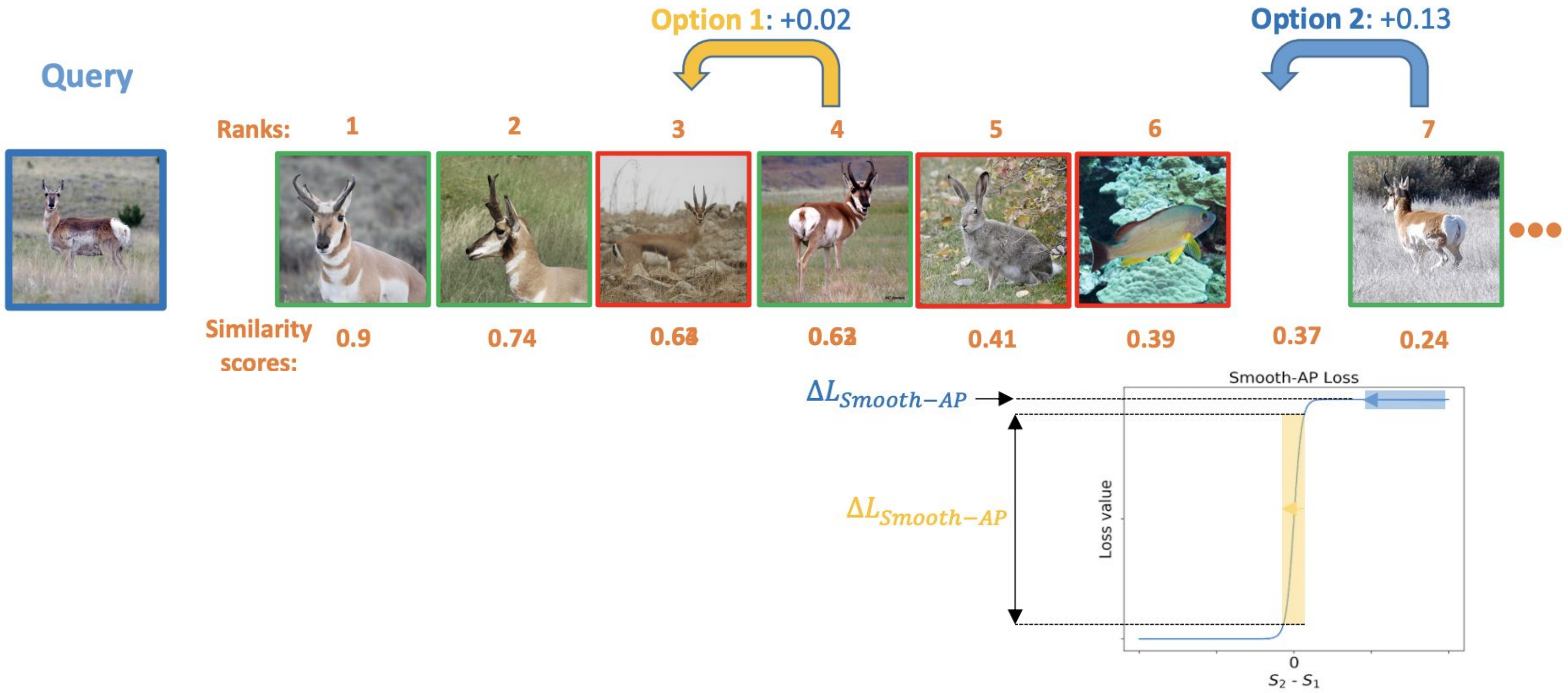
Change in loss = 0.2

Smooth-AP



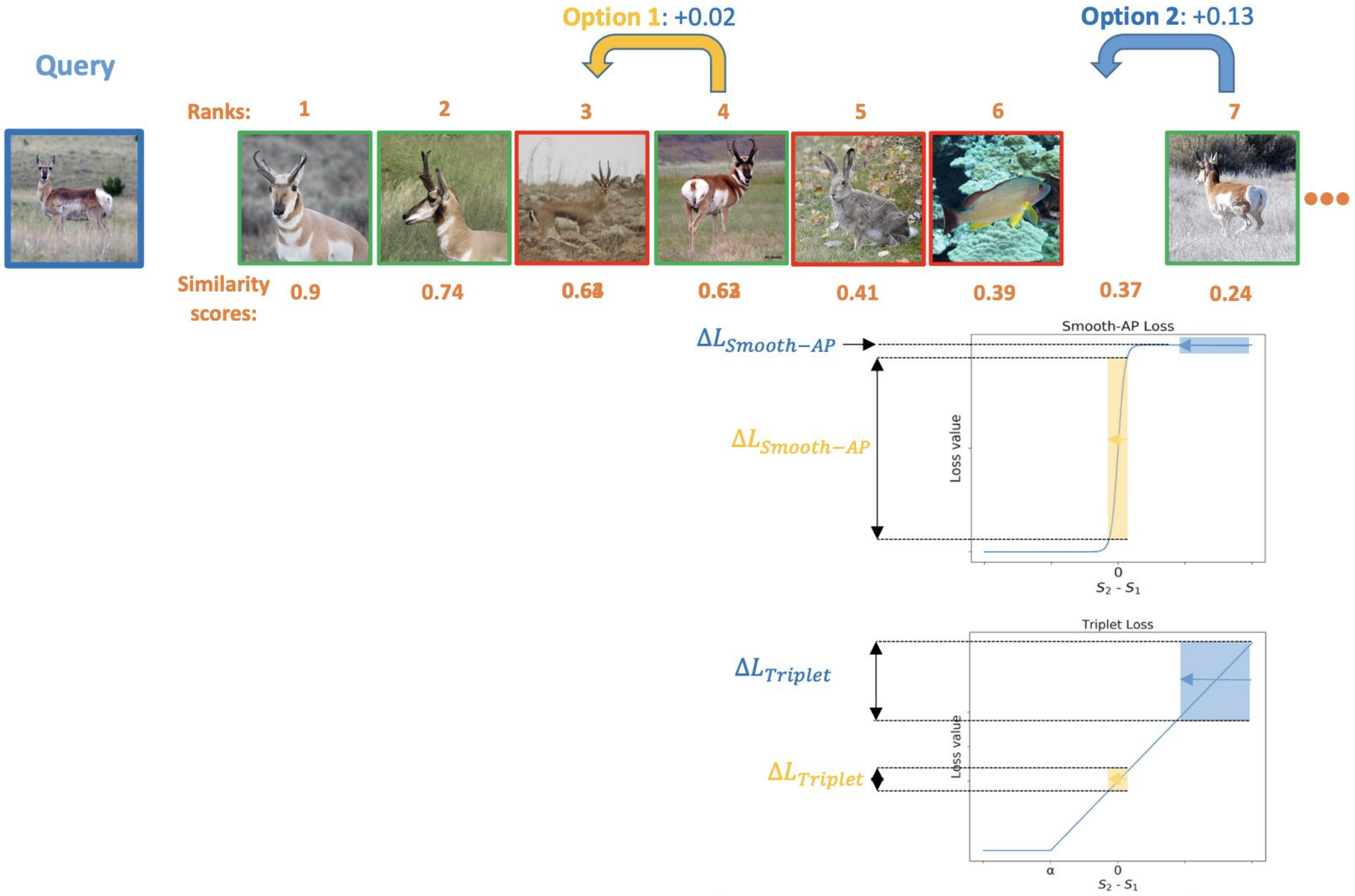
Option 1: +0.02
Option 2: +0.13
Query
Ranks:
1
2
3
4
5
6
7
Similarity scores:
0.9
0.74
0.64
0.63
0.41
0.39
0.37
0.24

Smooth-AP



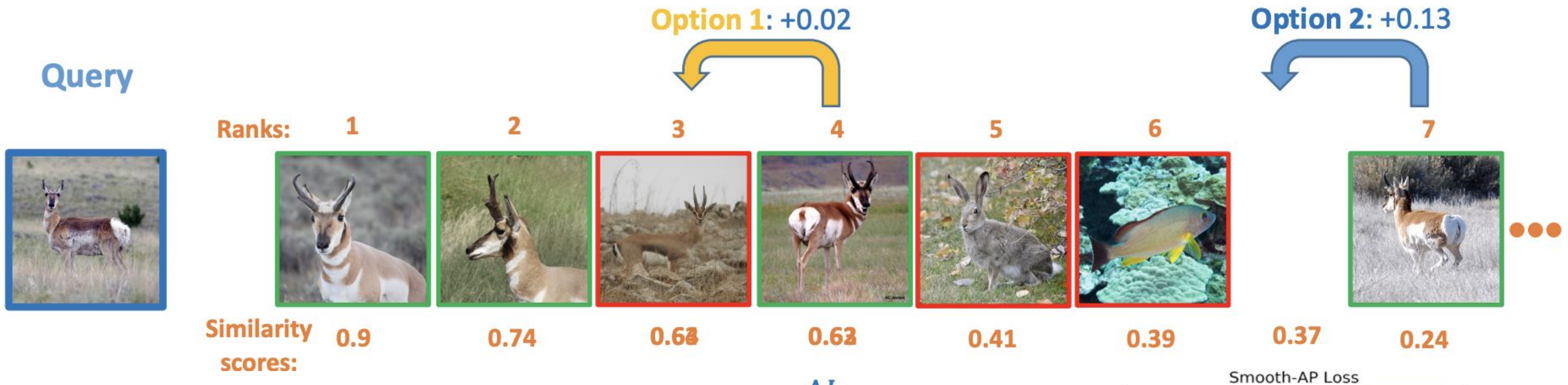
S_2 = positive instance relevance score S_1 = negative instance relevance score

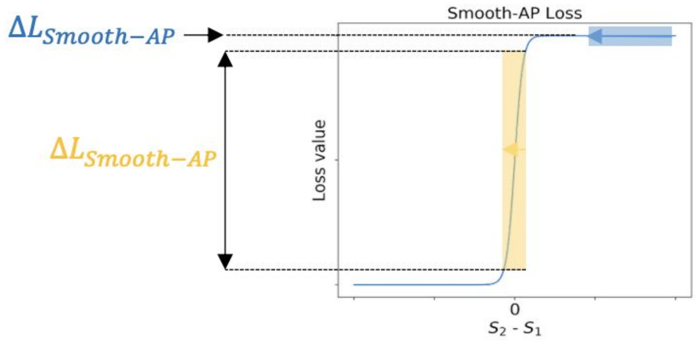
Smooth-AP

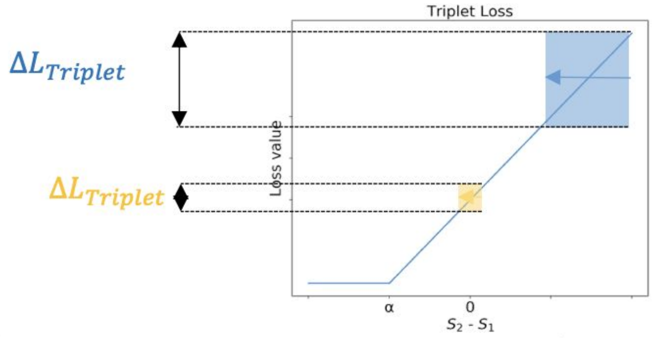


S_2 = positive instance relevance score S_1 = negative instance relevance score

Smooth-AP



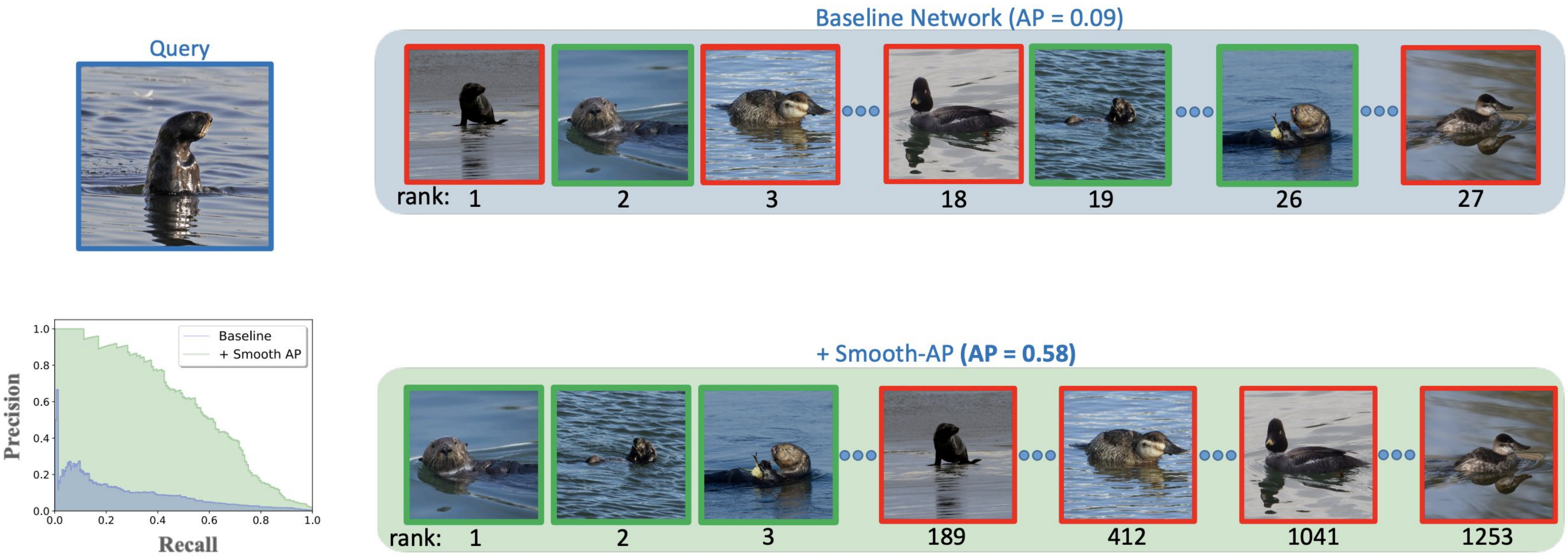




- Smooth-AP optimises a ranking metric
  - **Option 1**: small score change (+0.02), rank change →$\Delta AP > 0$
  - **Option 2**: large score change (+0.13), no rank change →$\Delta AP = 0$
- **Smooth-AP favours the rank change** → larger reduction in loss
- Triplet favours the large score change → larger reduction in loss

S_2 = positive instance relevance score S_1 = negative instance relevance score

Smooth-AP



Query
Baseline Network (AP = 0.09)
rank: 1
2
3
18
19
26
27
+ Smooth-AP (AP = 0.58)
rank: 1
2
3
189
412
1041
1253
Baseline
+ Smooth AP
Precision
Recall

Smooth-AP

Smooth average precision

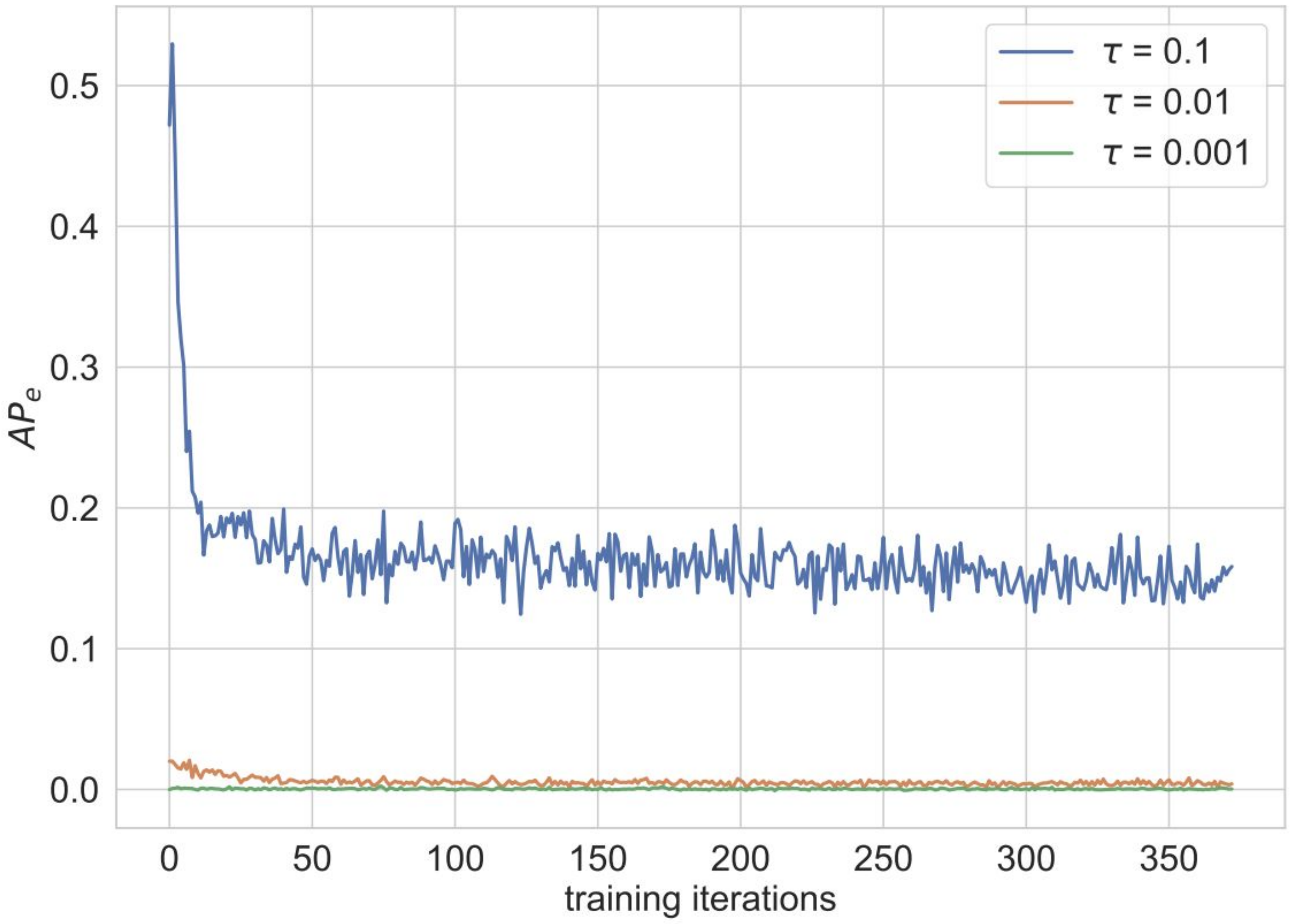
$$\mathcal{G}(x;\tau) = \frac{1}{1+e^{\frac{-x}{\tau}}}.$$

$$AP_q \approx \frac{1}{|\mathcal{S}_P|} \sum_{i \in \mathcal{S}_P} \frac{1 + \sum_{j \in \mathcal{S}_P} \mathcal{G}(D_{ij};\tau)}{1 + \sum_{j \in \mathcal{S}_P} \mathcal{G}(D_{ij};\tau) + \sum_{j \in \mathcal{S}_N} \mathcal{G}(D_{ij};\tau)}$$

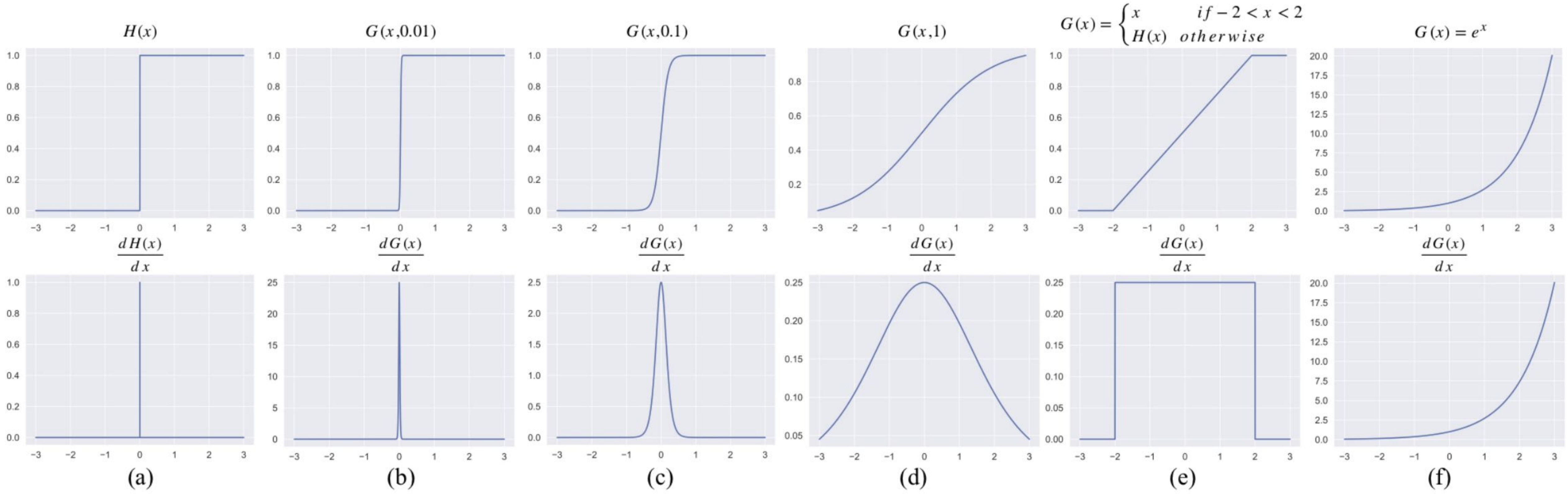
$$\mathcal{L}_{AP} = \frac{1}{m} \sum_{k=1}^{m} (1 - AP_k)$$

Effect of Sigmoid Temperature

$$AP_e = |AP_{pred} - AP|$$



Effect of Sigmoid Temperature



Effect of Sigmoid Temperature

Table 5: **Ablation study** over different parameters: temperature τ, size of positive set during minibatch sampling $|\mathcal{P}|$, and batch size B. Performance is benchmarked on VGGFace2-Test and IJB-C.

τ	mAP	
	VF2	IJB-C
0.1	0.824	0.726
0.01	**0.844**	**0.736**
0.001	0.839	0.733

$|\mathcal{P}| = 4$, $B = 128$

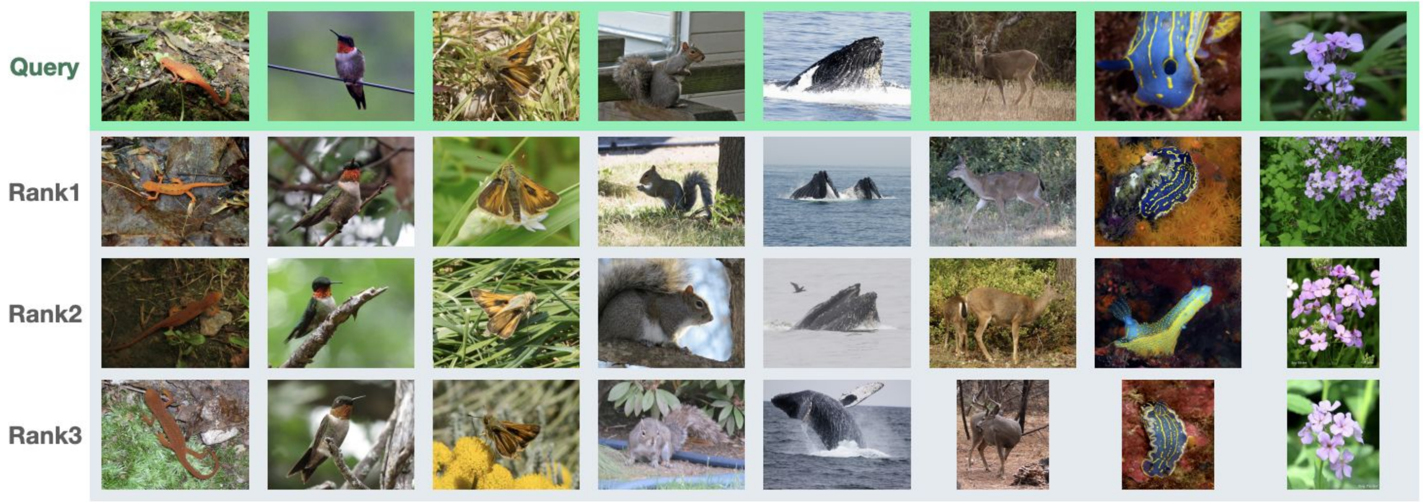
Effect of Batch Size

Table 5: **Ablation study** over different parameters: temperature τ, size of positive set during minibatch sampling $|\mathcal{P}|$, and batch size B. Performance is benchmarked on VGGFace2-Test and IJB-C.

$\|\mathcal{P}\|$	mAP	
	VF2	IJB-C
4	**0.844**	**0.736**
8	0.833	0.734
16	0.824	0.726
$\tau = 0.01$, $B = 128$		

$\|\mathcal{B}\|$	mAP	
	VF2	IJB-C
64	0.824	0.726
128	0.844	0.736
256.	**0.853**	**0.754**
$\tau = 0.01$, $\|\mathcal{P}\| = 4$		

Results - INaturalist



	INaturalist			
Recall@K	1	4	16	32
Triplet Semi-Hard (NeurIPS '06)	58.1	75.5	86.8	90.7
Proxy NCA (CVPR '17)	61.6	77.4	87.0	90.6
* FastAP (CVPR '19)	60.6	77.0	87.2	90.6
* Blackbox AP (CVPR '20)	62.9	79.0	88.9	92.1
Smooth-AP BS=224	65.9	80.9	89.8	92.7
Smooth-AP BS=384	**67.2**	**81.8**	**90.3**	**93.1**

* Recent AP-approximating approaches

Results - VehicleID and Stanford Products

	VehicleID					
	Small		Medium		Large	
Recall@K	1	5	1	5	1	5
Divide (CVPR '19)	87.7	92.9	85.7	90.4	82.9	90.2
MIC (ICCV '19)	86.9	93.4	-	-	82.0	91.0
* FastAP (CVPR '19)	91.9	96.8	90.6	95.9	87.5	95.1
* Cont. w/M (CVPR '20)	94.7	96.8	**93.7**	95.8	**93.0**	95.8
Smooth-AP	**94.9**	**97.6**	93.3	**96.4**	91.9	**96.2**

	Stanford Online Products			
Recall@K	1	10	100	1000
Margin (CVPR '17)	72.7	86.2	93.8	98.0
Divide (CVPR '19)	75.9	88.4	94.9	98.1
* FastAP (CVPR '19)	76.4	89.0	95.1	98.2
MIC (ICCV '19)	77.2	89.4	95.6	-
* Blackbox AP (CVPR '20)	78.6	90.5	96.0	98.7
* Cont. w/M (CVPR '20)	**80.6**	**91.6**	96.2	98.7
Smooth-AP BS=224	79.2	91.0	96.5	98.9
Smooth-AP BS=384	80.1	91.5	**96.6**	**99.0**

Smooth-AP uses hard negative mining

* Recent AP-approximating approaches

Results - Face Retrieval

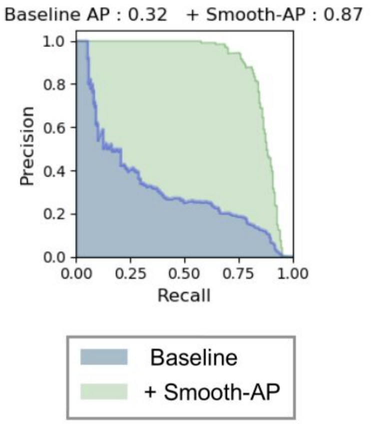
baseline network (AP = 0.32)

+ Smooth-AP **(AP = 0.87)**

Query Same class as query different class to query

VGGFace2 (ICAFGR '18)	VGGFace2 Test	IJB-C
Softmax	0.828	0.726
+Pairwise	0.828	0.728
+Triplet	0.845	0.740
+Smooth-AP	**0.850**	**0.754**
ArcFace (CVPR '19)	**VGGFace2 Test**	**IJB-C**
ArcFace	0.858	0.772
+Pairwise	0.861	0.775
+Triplet	0.880	0.787
+Smooth-AP	**0.902**	**0.803**

Table 2: Smooth-AP boosts mAP scores for strong face verification baselines on VGGFace2 test set and IJB-C datasets



Thank you!

@yash0307

patelyas@fel.cvut.cz

Homepage: https://yash0307.github.io