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

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

- \( S_\Omega \) = retrieval set
- \( S_p \) = positive retrieval set
- \( R \) = ranking of instance in retrieval set
- \( R^+ \) = 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.

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

\(S_\Omega = \) retrieval set \quad \(S_p = \) positive retrieval set \quad \(R = \) ranking of instance in retrieval set \quad \(R^+ = \) ranking of instance in positive retrieval set
Image Retrieval Training

- Embedding must be trained for good ranking.
- Achieved using loss functions.
Ranking Surrogate loss (e.g. Triplet Loss)

\[ L_{\text{triplet}} \propto \max(S_2 - S_1 + \alpha, 0) \]

\( S_2 = \) positive instance relevance score
\( S_1 = \) negative instance relevance score

Differentiable: \(\checkmark\) Optimises Ranking metric: \(\times\)
Smoothing the Average Precision Loss

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

<table>
<thead>
<tr>
<th>Score</th>
<th>Rank</th>
<th>Similarity Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>0.74</td>
<td>2</td>
<td>0.74</td>
</tr>
<tr>
<td>0.39</td>
<td>3</td>
<td>0.41</td>
</tr>
<tr>
<td>0.41</td>
<td>4</td>
<td>0.39</td>
</tr>
<tr>
<td>0.90</td>
<td>5</td>
<td>0.24</td>
</tr>
</tbody>
</table>

\[ H(x) = \begin{cases} 
0 & \text{if } x < 0 \\
1 & \text{if } x \geq 0 
\end{cases} \]
Smoothing the Average Precision Loss

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

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

Ranks: 1

Similarity scores: 0.9, 0.74, 0.41, 0.39, 0.24

Query

0.24
0.74
0.39
0.41
0.90

-0.50
0
-0.35
-0.33
0.16

0
1
0
0
1

sum

2
Smooth-AP

**Average Precision Loss**

\[ L_{AP} = (1 - AP_q) \]

**Smooth-AP Loss**

\[ L_{Smooth-AP} \propto G(x) = \frac{1}{1 + e^{-\frac{x}{\tau}}} \]

**Ranking Surrogate loss (e.g. Triplet Loss)**

\[ L_{triplet} \propto \max(S_2 - S_1 + \alpha, 0) \]

Differentiable?\quad\quad Optimises Ranking metric?

\[ \times \quad \checkmark \]

\[ \times \quad \checkmark \]

\[ \checkmark \quad \times \]

\[ S_2 = \text{positive instance relevance score} \quad \quad S_1 = \text{negative instance relevance score} \]
Smooth-AP

Average precision

$$AP_q = \frac{1}{|S_P|} \sum_{i \in S_P} \frac{1 + \sum_{j \in S_p, j \neq i} 1\{D_{ij} > 0\}}{1 + \sum_{j \in S_p, j \neq i} 1\{D_{ij} > 0\} + \sum_{j \in S_N} 1\{D_{ij} > 0\}}$$

Smooth average precision

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

$$AP_q \approx \frac{1}{|S_P|} \sum_{i \in S_P} \frac{1 + \sum_{j \in S_p} G(D_{ij}; \tau)}{1 + \sum_{j \in S_p} G(D_{ij}; \tau) + \sum_{j \in S_N} G(D_{ij}; \tau)}$$

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

Average precision

\[
\begin{align*}
AP_q &= \frac{1}{|S_P|} \sum_{i \in S_P} \frac{1 + \sum_{j \in S_P, j \neq i} 1\{D_{ij} > 0\}}{1 + \sum_{j \in S_P, j \neq i} 1\{D_{ij} > 0\} + \sum_{j \in S_N} 1\{D_{ij} > 0\}} \\
\end{align*}
\]

Smooth average precision

\[
G(x; \tau) = \frac{1}{1 + e^{-\frac{x}{\tau}}}
\]

\[
\begin{align*}
AP_q &\approx \frac{1}{|S_P|} \sum_{i \in S_P} \frac{1 + \sum_{j \in S_P} G(D_{ij}; \tau) + \sum_{j \in S_N} G(D_{ij}; \tau)}{1 + \sum_{j \in S_P} G(D_{ij}; \tau) + \sum_{j \in S_N} G(D_{ij}; \tau)} \\
\end{align*}
\]

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

Change in loss = 0.5

Change in loss = 0.2
Smooth-AP
Smooth-AP

\[ S_2 = \text{positive instance relevance score} \quad S_1 = \text{negative instance relevance score} \]
Smooth-AP

Query

Ranks:
1
2
3
4
5
6
7

Similarity scores:
0.9
0.74
0.68
0.63
0.41
0.39
0.37
0.24

Option 1: +0.02

Option 2: +0.13

$S_2$ = positive instance relevance score

$S_1$ = negative instance relevance score

$\Delta L_{\text{Smooth-AP}}$

$\Delta L_{\text{Triplet}}$
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** $\rightarrow$ larger reduction in loss

- Triplet favours the large score change $\rightarrow$ larger reduction in loss

$S_2 = \text{positive instance relevance score}$

$S_1 = \text{negative instance relevance score}$
Smooth-AP

Query

Baseline Network (AP = 0.09)

+ Smooth-AP (AP = 0.58)

Precision

Recall

rank: 1 2 3 18 19 26 27

rank: 1 2 3 189 412 1041 1253
Smooth-AP

Smooth average precision

\[ g(x; \tau) = \frac{1}{1 + e^{-\frac{x}{\tau}}} \]

\[ AP_q \approx \frac{1}{|S_p|} \sum_{i \in S_p} \frac{1 + \sum_{j \in S_p} g(D_{ij}; \tau)}{1 + \sum_{j \in S_p} g(D_{ij}; \tau) + \sum_{j \in S_N} g(D_{ij}; \tau)} \]

\[ L_{AP} = \frac{1}{m} \sum_{k=1}^{m} (1 - AP_k) \]
Effect of Sigmoid Temperature

\[ AP_e = |AP_{\text{pred}} - AP| \]
Effect of Sigmoid Temperature

- $H(x)$
- $G(x, 0.01)$
- $G(x, 0.1)$
- $G(x, 1)$
- $G(x) = \begin{cases} x & \text{if } -2 < x < 2 \\ H(x) & \text{otherwise} \end{cases}$
- $G(x) = e^x$

(a) $\frac{dH(x)}{dx}$
(b) $\frac{dG(x)}{dx}$
(c) $\frac{dG(x)}{dx}$
(d) $\frac{dG(x)}{dx}$
(e) $\frac{dG(x)}{dx}$
(f) $\frac{dG(x)}{dx}$
Effect of Sigmoid Temperature

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

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>VF2</th>
<th>IJB-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.824</td>
<td>0.726</td>
</tr>
<tr>
<td><strong>0.01</strong></td>
<td><strong>0.844</strong></td>
<td><strong>0.736</strong></td>
</tr>
<tr>
<td>0.001</td>
<td>0.839</td>
<td>0.733</td>
</tr>
</tbody>
</table>

$|P| = 4, B = 128$
Effect of Batch Size

Table 5: Ablation study over different parameters: temperature $\tau$, 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    | $|B|$ | mAP    |
|----------------|--------|------|--------|
|                | VF2    | IJB-C| VF2    | IJB-C |
| 4              | 0.844  | 0.736| 64     | 0.824  |
| 8              | 0.833  | 0.734| 128    | 0.844  |
| 16             | 0.824  | 0.726| 256    | 0.853  |
| $\tau = 0.01$, $B = 128$ |        |      | $\tau = 0.01$, $|\mathcal{P}| = 4$ |        |
## Results - INaturalist

### Table of Recall@K for INaturalist

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

* Recent AP-approximating approaches
# Results - VehicleID and Stanford Products

<table>
<thead>
<tr>
<th>Recall@K</th>
<th>VehicleID</th>
<th>Stanford Online Products</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small</td>
<td>Medium</td>
</tr>
<tr>
<td>Divide (CVPR ’19)</td>
<td>87.7</td>
<td>92.9</td>
</tr>
<tr>
<td>MIC (ICCV ’19)</td>
<td>86.9</td>
<td>93.4</td>
</tr>
<tr>
<td>FastAP (CVPR ’19)</td>
<td>91.9</td>
<td>96.8</td>
</tr>
<tr>
<td>Cont. w/M (CVPR ’20)</td>
<td>94.7</td>
<td>96.8</td>
</tr>
<tr>
<td>Smooth-AP</td>
<td><strong>94.9</strong></td>
<td><strong>97.6</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recall@K</th>
<th>Stanford Online Products</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Margin (CVPR ’17)</td>
<td>72.7</td>
</tr>
<tr>
<td>Divide (CVPR ’19)</td>
<td>75.9</td>
</tr>
<tr>
<td>FastAP (CVPR ’19)</td>
<td>76.4</td>
</tr>
<tr>
<td>MIC (ICCV ’19)</td>
<td>77.2</td>
</tr>
<tr>
<td>Blackbox AP (CVPR ’20)</td>
<td>78.6</td>
</tr>
<tr>
<td>Cont. w/M (CVPR ’20)</td>
<td><strong>80.6</strong></td>
</tr>
<tr>
<td>Smooth-AP BS=224</td>
<td>79.2</td>
</tr>
<tr>
<td>Smooth-AP BS=384</td>
<td>80.1</td>
</tr>
</tbody>
</table>

Smooth-AP uses hard negative mining

* Recent AP-approximating approaches
Results - Face Retrieval

baseline network  (AP = 0.32)

+ Smooth-AP  (AP = 0.87)

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