Robust and Decomposable Average Precision for Image Retrieval

Elias Ramzi[1,2], Nicolas Thome[1], Clément Rambour[1], Nicolas Audebert[1], Xavier Bitot[2]

[1]CEDRIC, Conservatoire National des Arts et Métiers, Paris, France
{elias.ramzi,nicolas.thome,clement.rambour,nicolas.audebert}@cnam.fr

[2]Coexya, Paris, France
xavier.bitot@coexya.eu

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

le cnam

COEXYA
Connect skills, create more

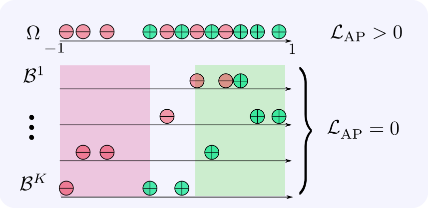
Average Precision for Image Retrieval

Average Precision (AP): most popular metric for ranking evaluation in several tasks (Recommendation systems, Object detection, Image Retrieval, NLP).

$$\mathcal{L}_{\text{AP}} = 1 - \text{AP} = 1 - \frac{1}{3}\left(\frac{1}{1} + \frac{2}{4} + \frac{3}{8}\right) = 37.5\%$$

Two issues when optimizing $\mathcal{L}_{\text{AP}}$ with SGD:

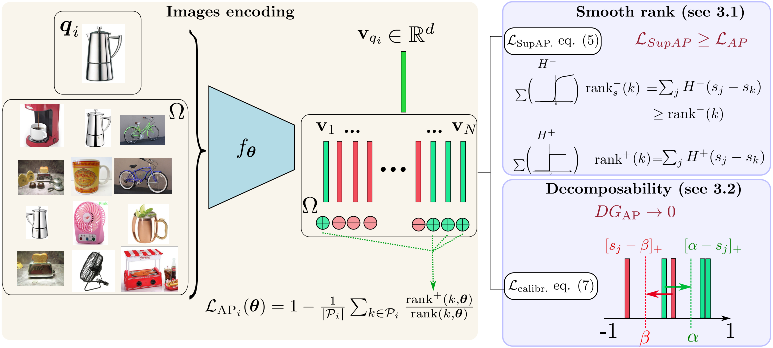
- **Non-differentiability**
- **Non-decomposability**



ROADMAP

ROBust And DecoMposable Average Precision (**ROADMAP**) for Image Retrieval:

- $\mathcal{L}_{\text{SupAP}}$: a new **smooth surrogate loss** for $\mathcal{L}_{\text{AP}}$
- $\mathcal{L}_{\text{calibr.}}$: a training objective to **reduce the non-decomposability** of AP



$$\mathcal{L}_{\text{ROADMAP}}(\boldsymbol{\theta}) = (1 - \lambda) \cdot \mathcal{L}_{\text{SupAP}}(\boldsymbol{\theta}) + \lambda \cdot \mathcal{L}_{\text{calibr.}}(\boldsymbol{\theta})$$

SupAP: direct optimization of AP

SupAP:

- **smooth approximation of the rank**.

$$\mathcal{L}_{AP} = 1 - \frac{1}{|\mathcal{P}_i|} \sum_{k \in \mathcal{P}_i} \frac{\text{rank}^+(k)}{\text{rank}(k)}$$

$\Rightarrow$ different than coarse upper bound on $\mathcal{L}_{\text{AP}}$: Structured SVM (SIGIR 2007), Blackbox (ICLR 2020)

- Re-writting the rank with the **Heaviside** step function

$$\text{rank}(k) = 1 + \sum_{j \in \Omega} H(s_j - s_k) = 1 + \underbrace{\sum_{j \in \mathcal{P}_i} H(s_j - s_k)}_{\text{rank}^+} + \underbrace{\sum_{j \in \mathcal{N}_i} H(s_j - s_k)}_{\text{rank}^-}$$

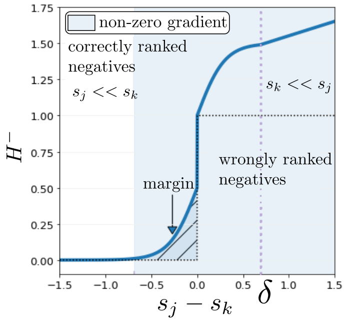
$\Rightarrow$ Other state of the art rank approximations do not provide upper bounds on $\mathcal{L}_{\text{AP}}$: FastAP (CVPR'19), SmoothAP (ECCV'20)

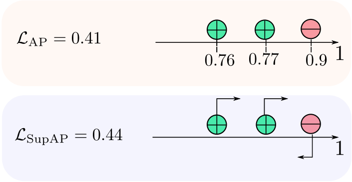
SupAP: direct optimization of AP

$$\mathcal{L}_{\text{SupAP}} = 1 - \frac{1}{|\mathcal{P}_i|} \sum_{k \in \mathcal{P}_i} \frac{\text{rank}^+(k)}{\text{rank}^+(k) + \text{rank}_s^-(k)}$$

$rank_s^-$: **Smooth upper bound** on $rank^-$ to ensure that:

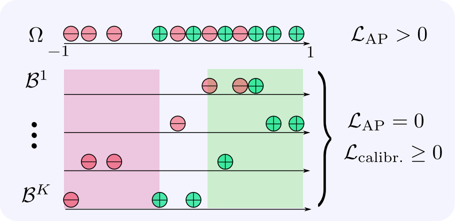
- SupAP is an **upper bound** on $\mathcal{L}_{\text{AP}}$.
- SupAP has **correct gradient flow**.





Decomposability Gap:
Approximation error between the AP per batch and the global AP.

$$DG_{\text{AP}}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{b=1}^{K} \text{AP}_i^b(\boldsymbol{\theta}) - \text{AP}_i(\boldsymbol{\theta})$$



Training objective to **reduce the Decomposability Gap**:

$$\mathcal{L}_{\text{calibr.}}(\boldsymbol{\theta}) = \frac{1}{M}\sum_{i=1}^{M} \underbrace{\frac{1}{|\mathcal{P}_i|}\sum_{\mathbf{x_j}\in\mathcal{P}_i}[\alpha - s_j]_+}_{\mathcal{L}^+_{\text{calibr.}}} + \underbrace{\frac{1}{|\mathcal{N}_i|}\sum_{\mathbf{x_j}\in\mathcal{N}_i}[s_j - \beta]_+}_{\mathcal{L}^-_{\text{calibr.}}}$$

$\mathcal{L}_{\text{calibr.}}$: reducing DG_{AP}

$\mathcal{L}_{\text{calibr.}}$ compares the scores in each batch with a **fixed threshold**.

Upper bound on DG_{AP} without $\mathcal{L}_{\text{calibr.}}$:

$$0 \leq DG_{\mathbf{AP}} \leq 1 - \frac{1}{\sum_{b=1}^{K} |\mathcal{P}_i^b|} \left(\sum_{b=1}^{K} \sum_{j=1}^{B} \frac{j + |\mathcal{P}_i^1| + \dots + |\mathcal{P}_i^{b-1}|}{j + |\mathcal{P}_i^1| + \dots + |\mathcal{P}_i^{b-1}| + |\mathcal{N}_i^1| + \dots + |\mathcal{N}_i^{b-1}|} \right)$$

Refined and **tighter upper bound** on DG_{AP} with $\mathcal{L}_{\text{calibr.}}$:

$$0 \leq DG_{\mathbf{AP}} \leq 1 - \frac{1}{\sum_{b=1}^{K} |\mathcal{P}_i^b|} \left(\sum_{b=1}^{K} \left[\sum_{j=1}^{G_b^+} \frac{j + G_1^+ + \dots + G_{b-1}^+}{j + G_1^+ + \dots + G_{b-1}^+ + E_1^- + \dots E_{b-1}^-} + \sum_{j=1}^{E_b^+} \frac{j + G_b^+ + |\mathcal{P}_i^1| + \dots + |\mathcal{P}_i^{b-1}|}{j + G_b^+ + |\mathcal{P}_i^1| + \dots + |\mathcal{P}_i^{b-1}| + |\mathcal{N}_i^1| + \dots + |\mathcal{N}_i^{b-1}|} \right] \right)$$

Comparison direct optimization AP methods under the **same settings** (backbone, data augmentation, batch size, optimizer . . .)

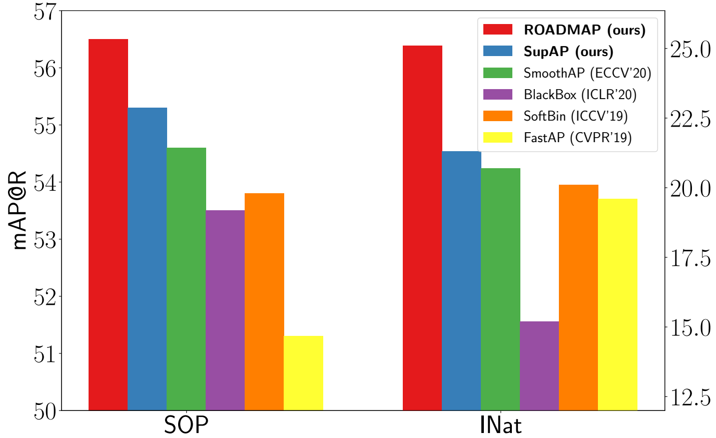
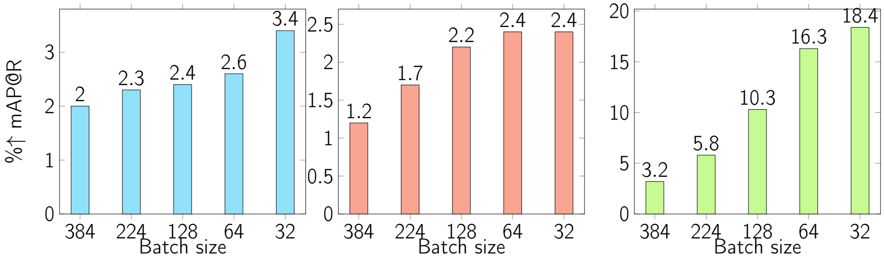


Figure: Models are ResNet-50 with a batch size of 64.

ROADMAP impact of decomposability ($\mathcal{L}_{\text{calibr.}}$)



(a) CUB (b) SOP (c) INaturalist

- **Good results** even with **small batches**!
- **Simple and effective** method, **no overhead** $\neq$ memory methods (FastAP, XBM)

	SOP	INat
Method	mAP@R	mAP@R
XBM (CVPR'20)	54.9	18.5
ROADMAP (ours)	**56.5**	**25.1**

Thank you for your attention !

ROADMAP a method for AP optimization:

- $\mathcal{L}_{\text{SupAP}}$: a new smooth surrogate loss for $\mathcal{L}_{\text{AP}}$;
- $\mathcal{L}_{\text{calibr.}}$: a training objective to reduce the decomposability gap.

Elias Ramzi

elias.ramzi@cnam.fr

https://github.com/elias-ramzi/ROADMAP