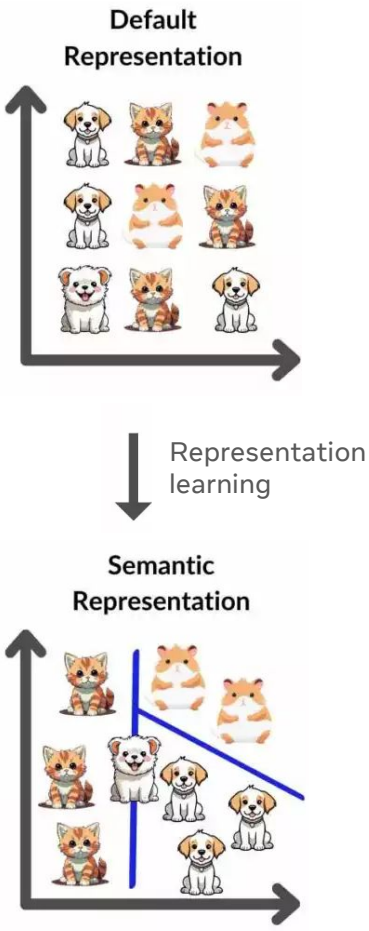


Robust image retrieval with deep learning.

Scientific seminar @ Orasis 2025 – Elias Ramzi
June 13, 2025

le cnam
cedric EA4629

COEXYA
Connect skills, create more
valeo.ai

Summary.

Robust image retrieval:

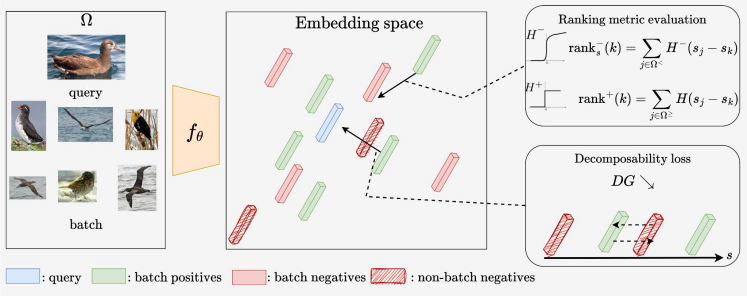
- 🗺️ ROADMAP: optimization of rank losses.
- 😸 HAPPIER: Hierarchical Image Retrieval for Robust Ranking.

Few shot robustness in the era of foundation models:

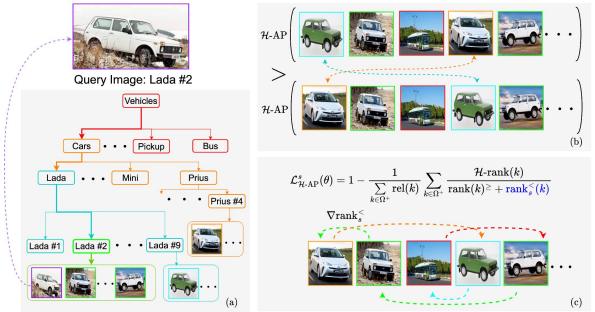
- 🔥HEAT: Post-hoc out-of-distribution detection.
- 🐎 GalLoP: Few shot adaptation of vision-language models.

Summary and contributions.

🗺️ ROADMAP: Optimization of Ranking Losses for Image retrieval.



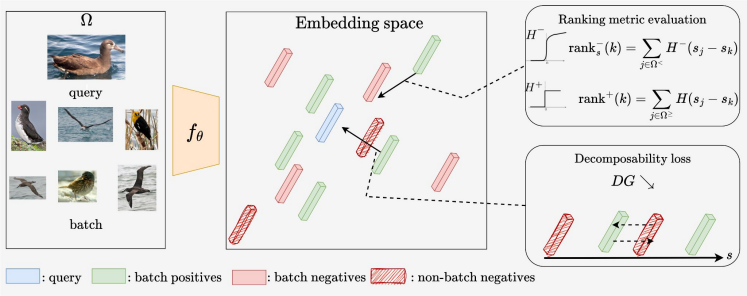
😺 HAPPIER: Hierarchical Image Retrieval for Robust Ranking.



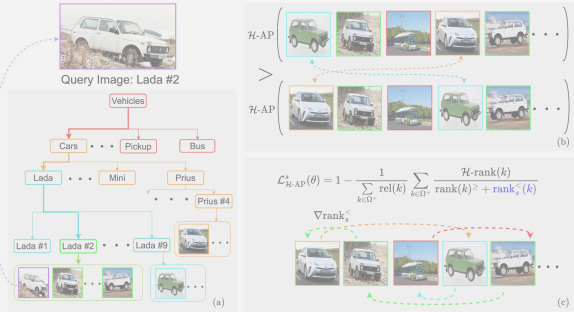
- Ramzi, Elias, et al. “Robust and Decomposable Average Precision for Image Retrieval.” *NeurIPS*, 2021.
- Ramzi, Elias, et al. “Hierarchical Average Precision Training for Pertinent Image Retrieval.” *ECCV*, 2022.
- Ramzi, Elias, et al. “Optimization of Rank Losses for Image Retrieval.” *TPAMI*, 2025.

Summary and contributions.

🗺️ ROADMAP: Optimization of Ranking Losses for Image retrieval.



😸 HAPPIER: Hierarchical Image Retrieval for Robust Ranking.



📕 Ramzi, Elias, et al. “Robust and Decomposable Average Precision for Image Retrieval.” *NeurIPS*, 2021.
📕 Ramzi, Elias, et al. “Optimization of Rank Losses for Image Retrieval.” *TPAMI*, 2025.

Image retrieval.

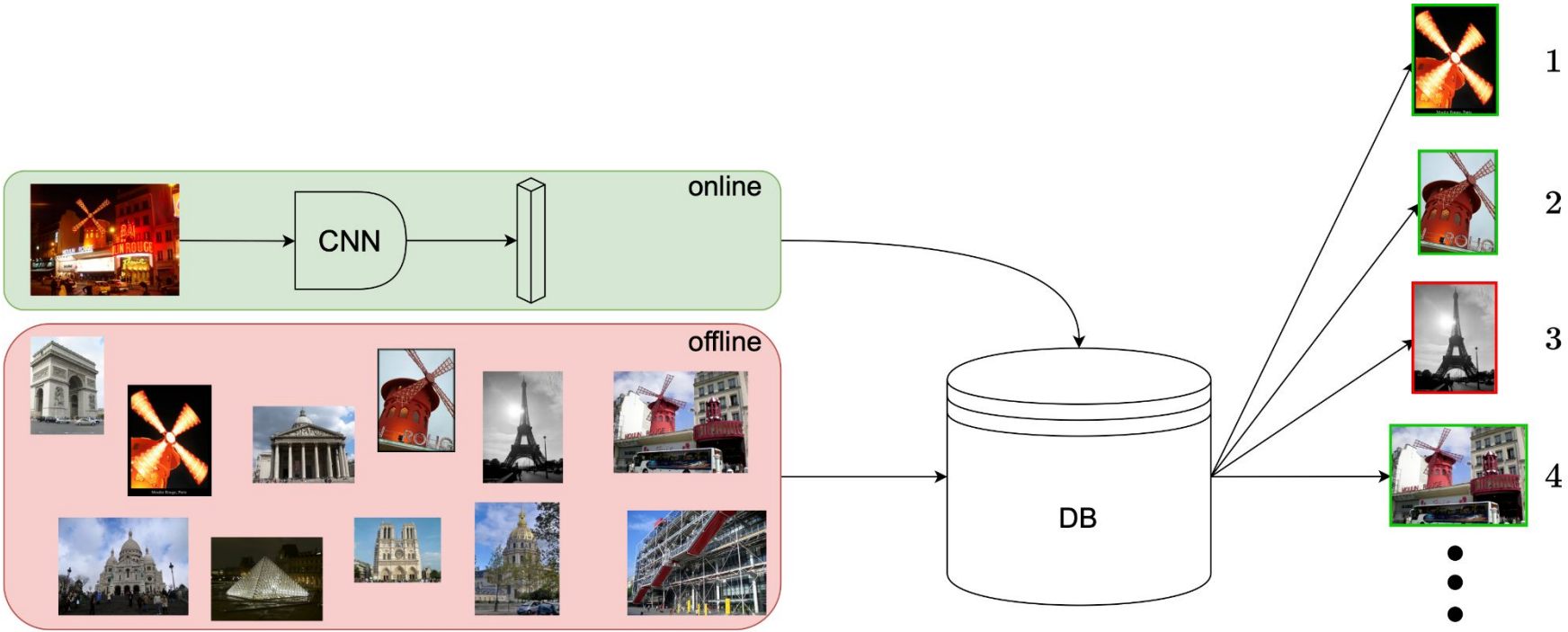
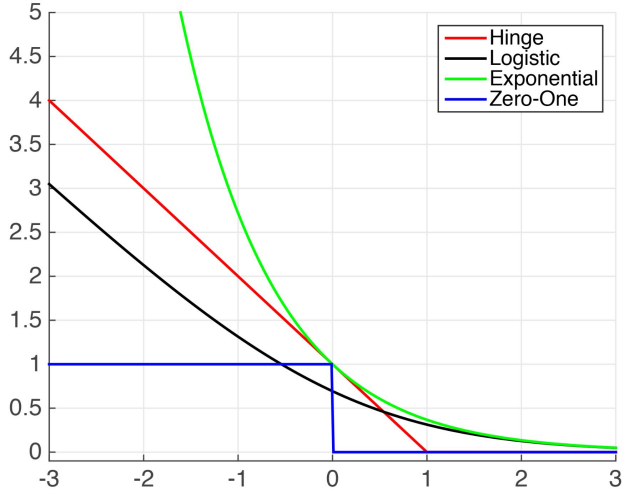
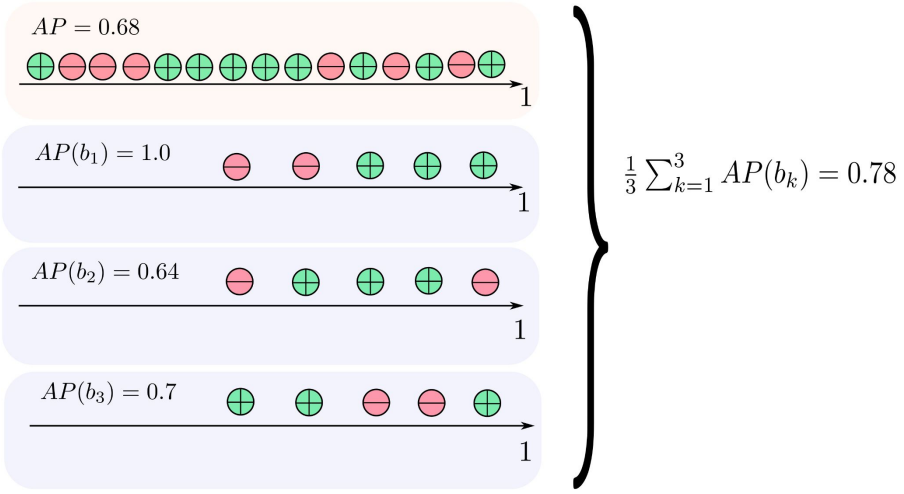


Image retrieval → retrieve similar images to the query image.

Optimizing rank losses.



Evaluation metrics are not differentiable → can not be used directly for SGD.



Evaluation metrics are non-decomposable → shortcomings for mini-batch training.

Ranking in stochastic gradient optimization.

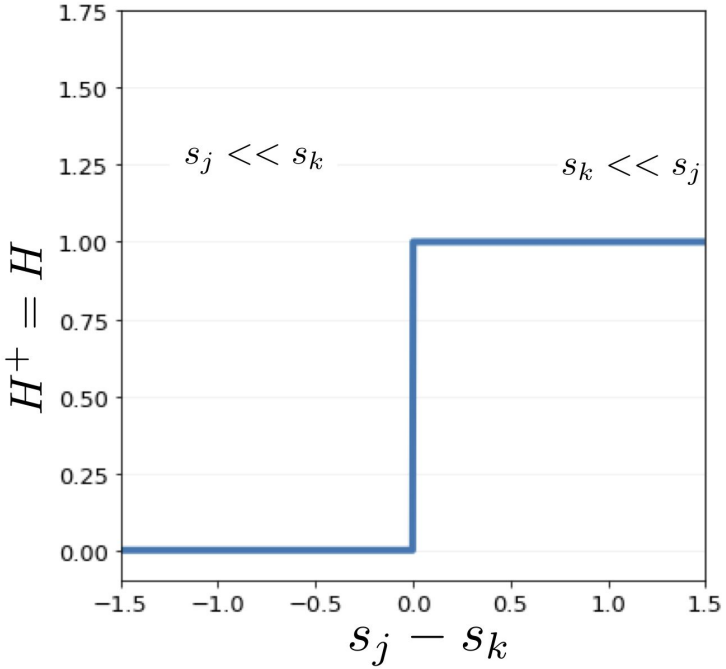
Evaluation metrics in image retrieval → rank-based:

$$\mathrm{AP} = \frac{1}{|\Omega^+|} \sum_{k \in \Omega^+} \frac{\mathrm{rank}^+(k)}{\mathrm{rank}(k)}$$

$$\mathrm{Recall@}n = \frac{\#\text{ number of positive before n}}{|\Omega^+|} = \frac{1}{\min(|\Omega^+|, n)} \sum_{k \in \Omega^+} H\left(n - \mathrm{rank}(k)\right)$$

Definition of the rank with Heaviside functions:

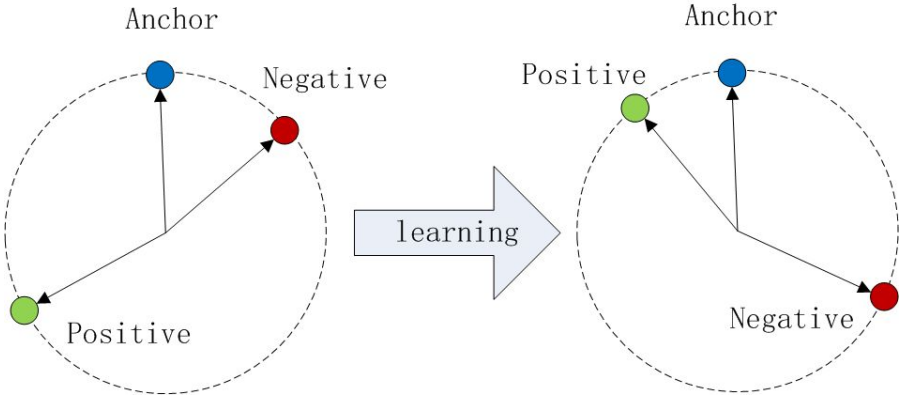
$$\mathrm{rank}(k) = \#\text{ number of similarities greater than } s_k = 1 + \sum_{j \in \Omega} H(s_j - s_k)$$



s_k : cosine similarity between image k and the query.

Surrogate losses: approximation of the metric.

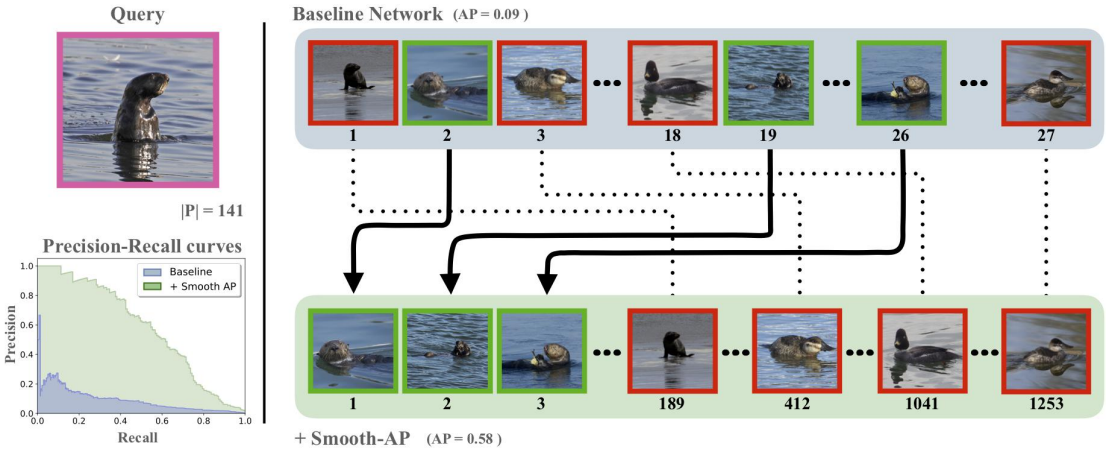
- Coarse upper-bound of the metrics.
- Not well-aligned with evaluation metrics: supports bottom vs. top of the ranking



Surrogate losses:

- Triplet loss – Wu, Chao-Yuan, et al. *ICCV*, 2017.
- NSM: Zhai, Andrew, et al. *BMVC*, 2018.
- Multi-similarity loss – Wang, Xun, et al. *CVPR*, 2019.

Surrogate losses: approximation of the rank.



- No strong theoretical guarantees: not an upper bound.
- Ill-behaved gradient flow.

Long line of work:

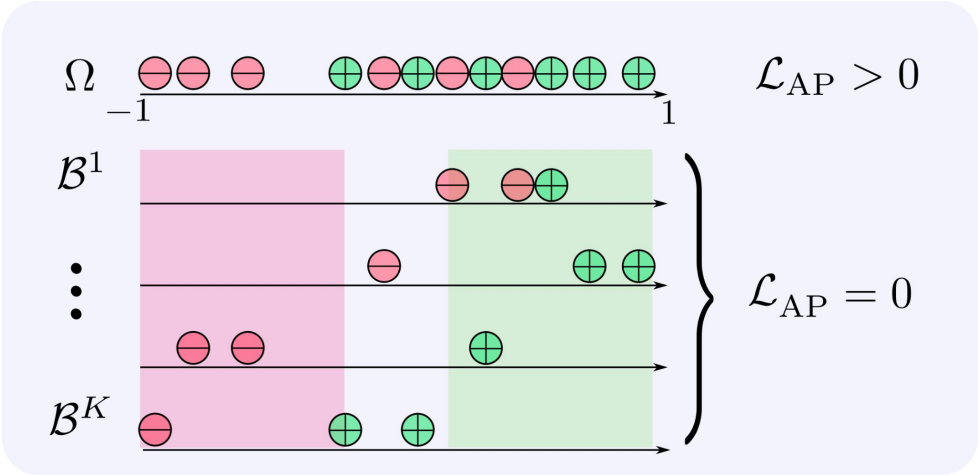
- Learned Ranking – SoDeep: Engilberge, Martin, et al. *CVPR*, 2019.
- Histogram binning – SoftBin: Revaud, Jerome, et al. *ICCV*, 2019.
- Blackbox optimization of AP: Rolínek, Michal, et al. *CVPR*, 2020.
- Smooth approximations of AP: Brown, Andrew, et al. *ECCV*, 2020.

Non-decomposability.

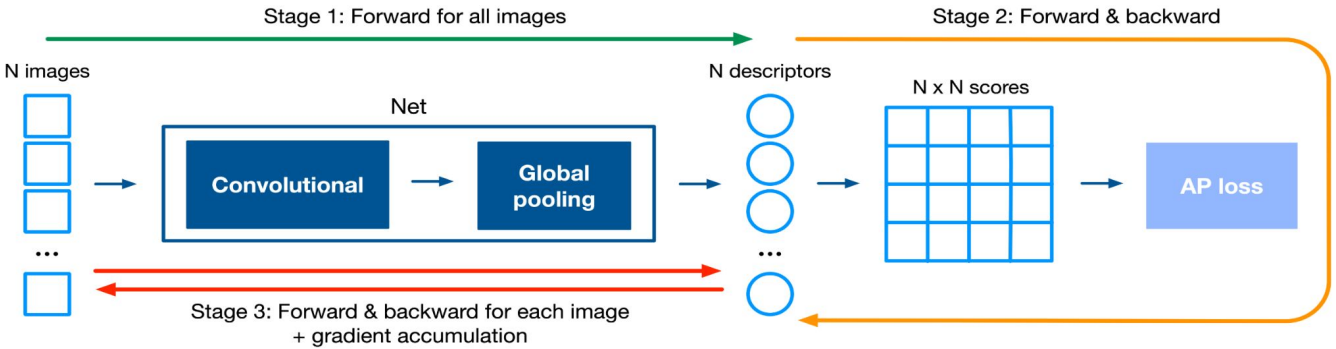
Ranking-based metrics are not decomposable:

$$\mathcal{M}(\Omega) \neq \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \mathcal{M}(b)$$

- Loss on batches are biased.
- Training can stop before global loss is zero.



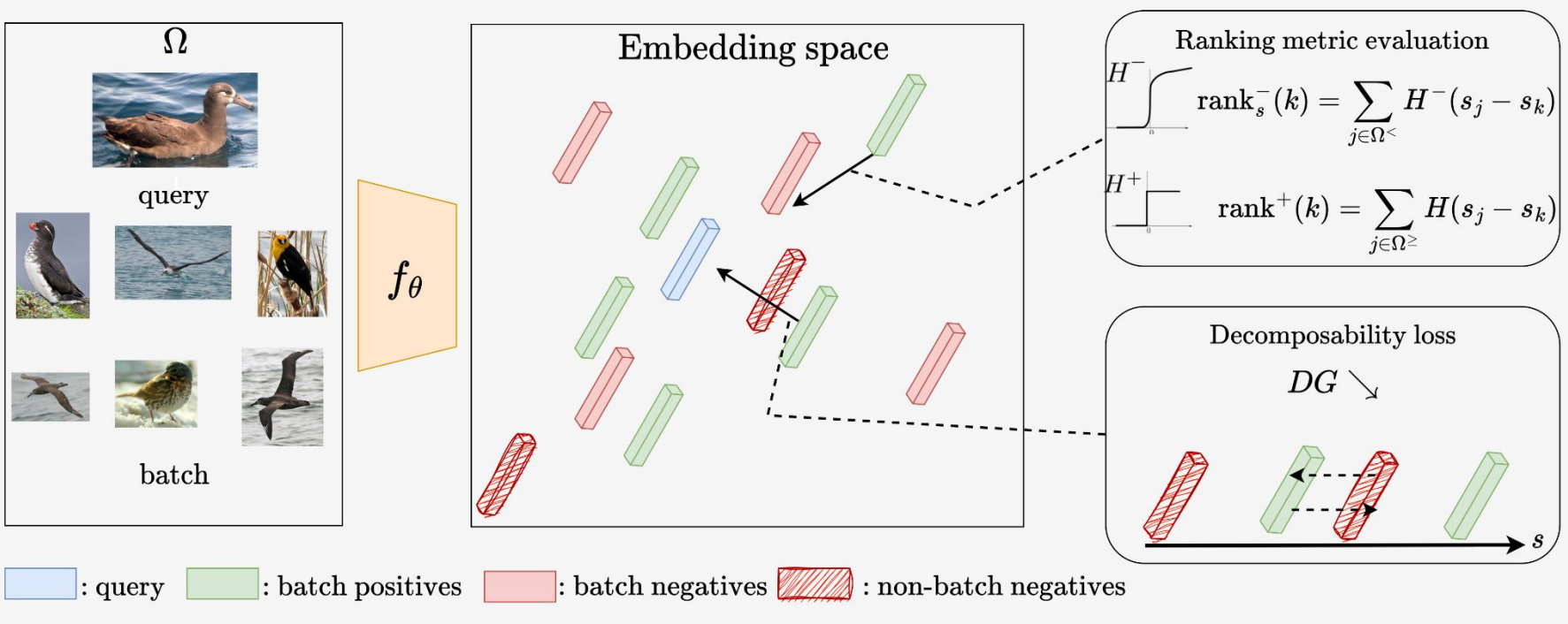
Addressing non-decomposability.



- ➔ Fewer works: brute force approaches
  - Hard negative mining.
  - Doubling the number of forward passes.
  - Storing the dataset.
- ➔ Overhead in training time.

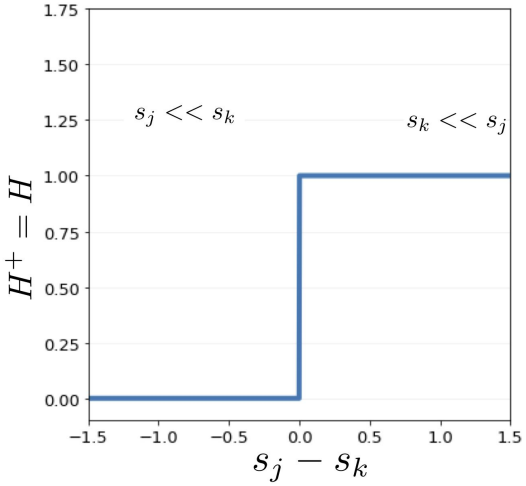
Negative mining – Wu, Chao-Yuan, et al. *ICCV*, 2017.
Large batches – Revaud, Jerome, et al. *ICCV*, 2019.
cross batch memory – Wang, Xun, et al. *CVPR*, 2020.

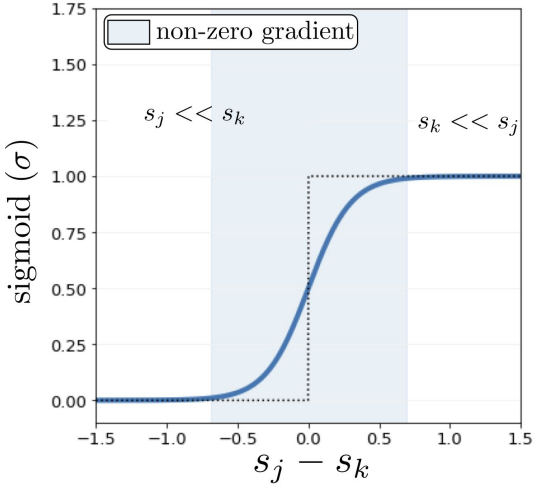
Robust and decomposable rank losses.

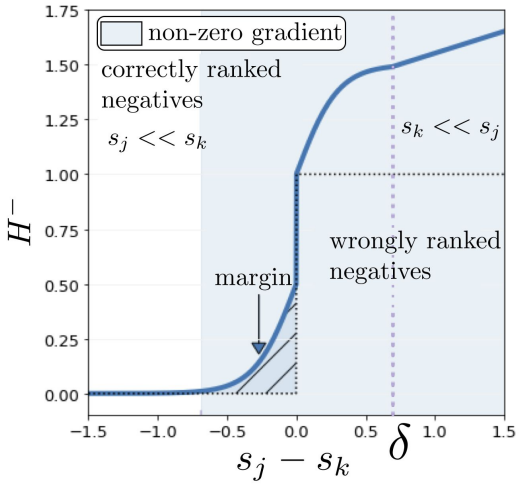


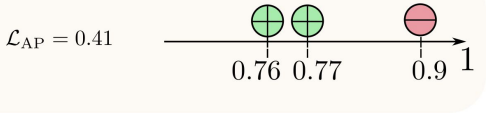
SupRank: smooth approximation of the rank.

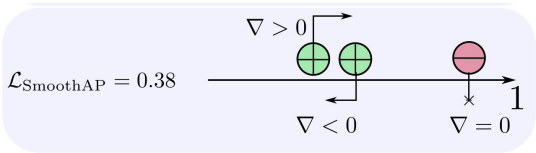
ROADMAP

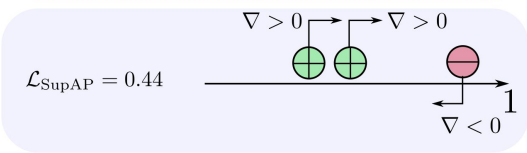




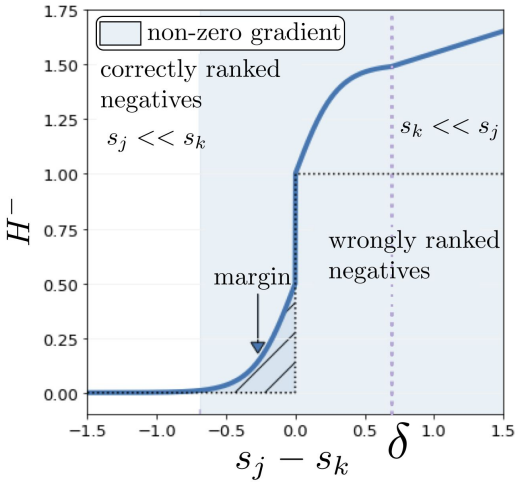


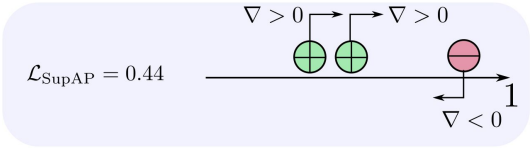






SupRank: smooth approximation of the rank.





$$\text{AP} = \frac{1}{|\Omega^+|} \sum_{k \in \Omega^+} \frac{\text{rank}^+(k)}{\text{rank}^+(k) + \text{rank}^-(k)}$$

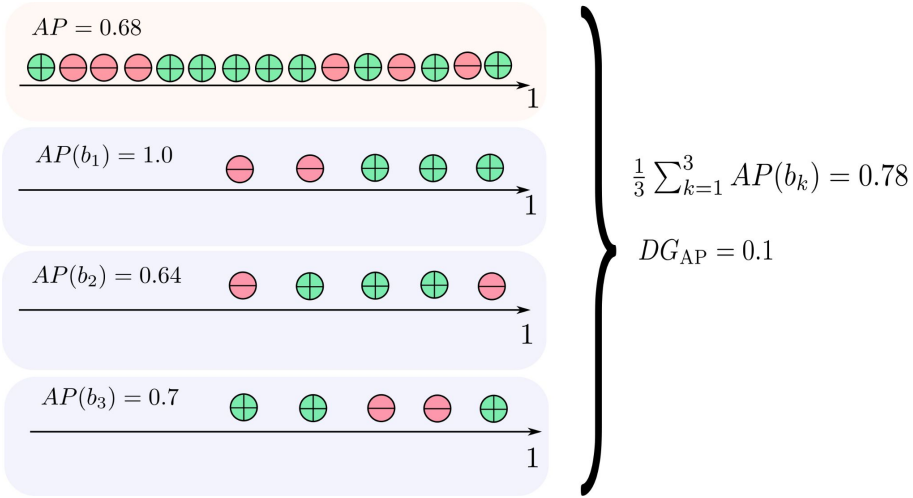
$$\text{rank}_s^-(k) = \sum_{j \in \Omega_k^-} H^-(s_j - s_k)$$

- Optimizing rank⁻ → smooth approximation + upper bound.
- Not optimizing rank⁺ → well-behaved gradients.

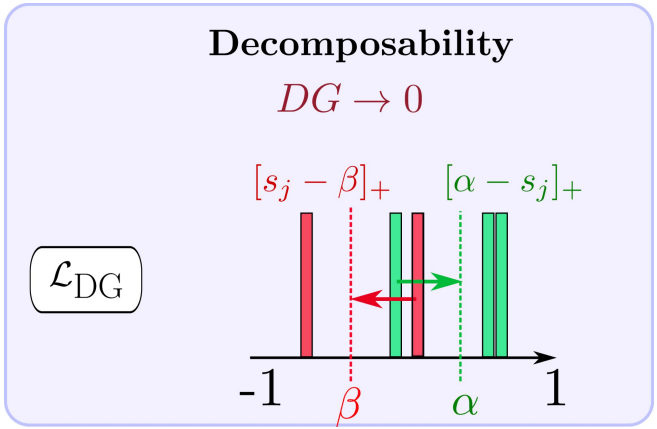
The decomposability gap.

$$\mathrm{DG}(\Omega) = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \mathcal{M}(b) - \mathcal{M}(\Omega)$$

→ Difference between the average ranking-based metrics on batch and its value on the whole dataset.



Decomposable rank losses.



$$\mathcal{L}_{DG}(\boldsymbol{\theta}) = \frac{1}{|\Omega^+|} \sum_{\boldsymbol{x_j} \in \Omega^+} [\alpha - \boldsymbol{s}_j]_+ + \frac{1}{|\Omega^-|} \sum_{\boldsymbol{x_j} \in \Omega^-} [\boldsymbol{s}_j - \beta]_+$$

- ➔ Calibrates scores across batches:
  - positive scores greater than α.
  - negative scores lower than β.
- ➔ Explicitly optimize an upper-bound on the decomposability gap.

Instantiation to image retrieval.

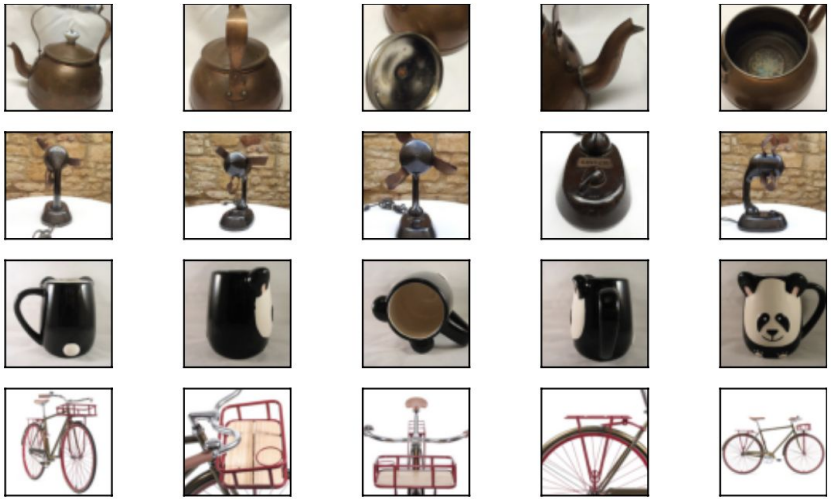
Framework applicable with ranking-based loss: *e.g.* AP, Recall, NDCG.

Application to Average Precision:

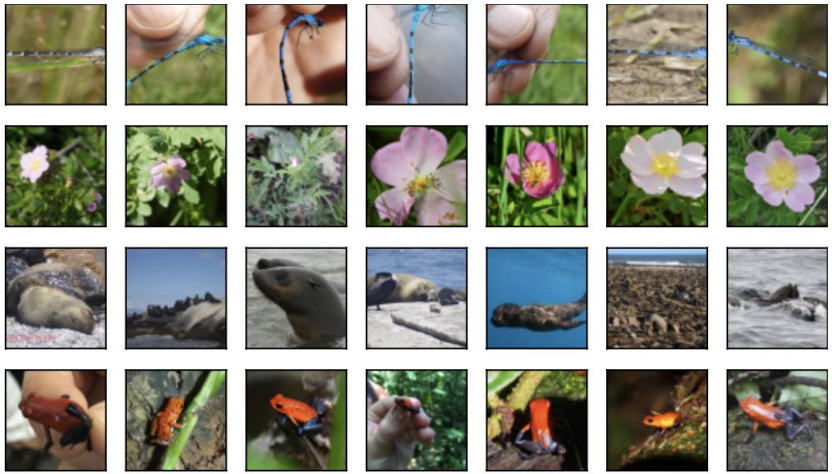
$$\mathcal{L}_{\text{SupAP}} = 1 - \frac{1}{|\Omega^+|} \sum_{k \in \Omega^+} \frac{\text{rank}^+(k)}{\text{rank}^+(k) + \text{rank}_s^-(k)}$$

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

Experimental validation.



SOP: retail products.

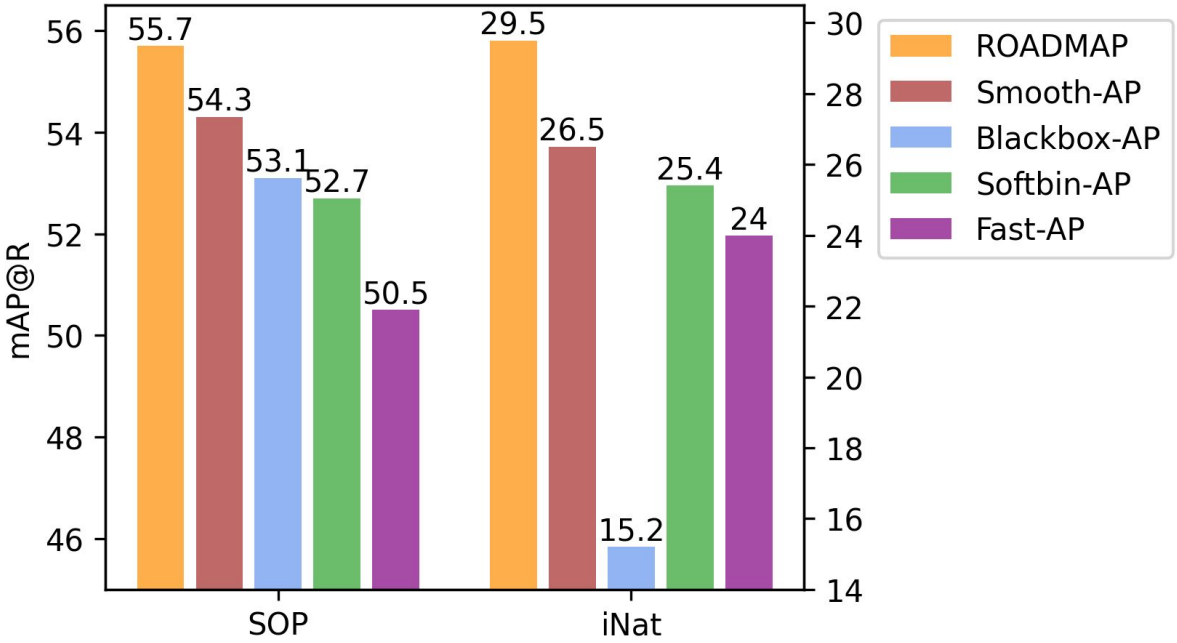


iNaturalist: wildlife images.

Song, Hyun Oh, et al. “Deep Metric Learning via Lifted Structured Feature Embedding.” *CVPR*, 2016.

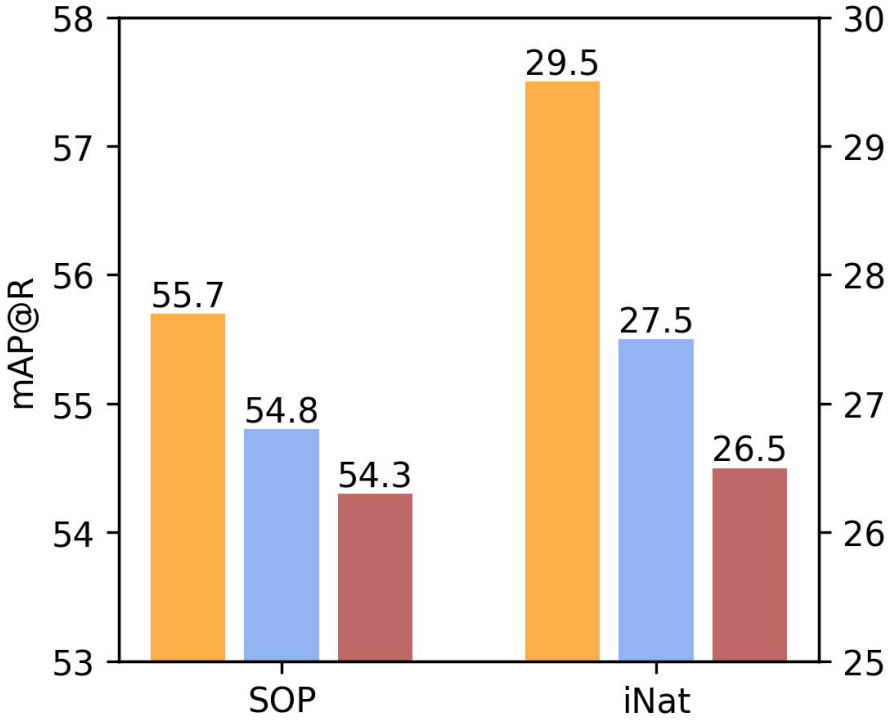
Van Horn, Grant, et al. “The iNaturalist Species Classification and Detection Dataset.” *CVPR*, 2018.

Comparison to AP methods.



➔ **Fair comparison** under the same setting (backbone, batch size, data, optimization…)
➔ Significant gains when **changing the loss**.

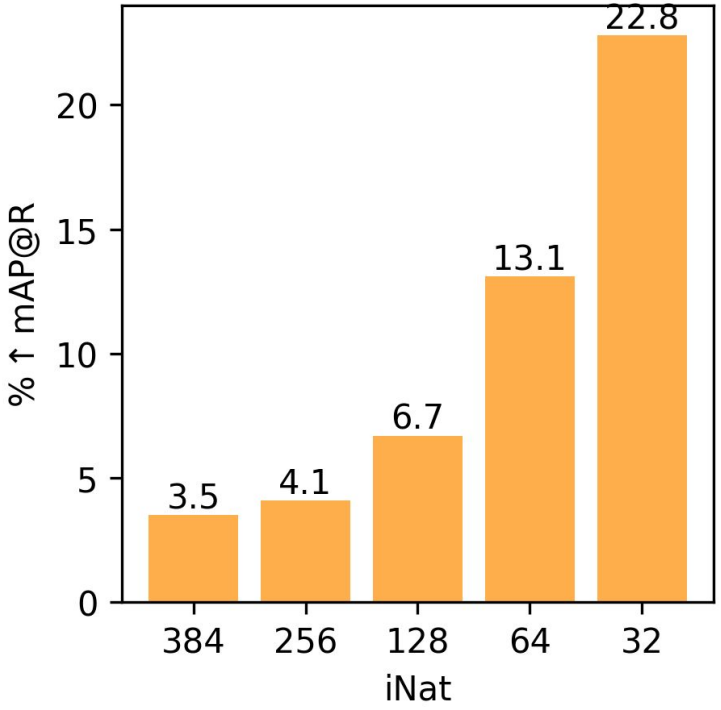
Ablation studies.



→ Significant gain with SupRank.

→ Further boost with decomposability loss.

Impact of the decomposability loss.



→ Biggest **relative** increase on small batches.

→ ROADMAP works with **small batches**.

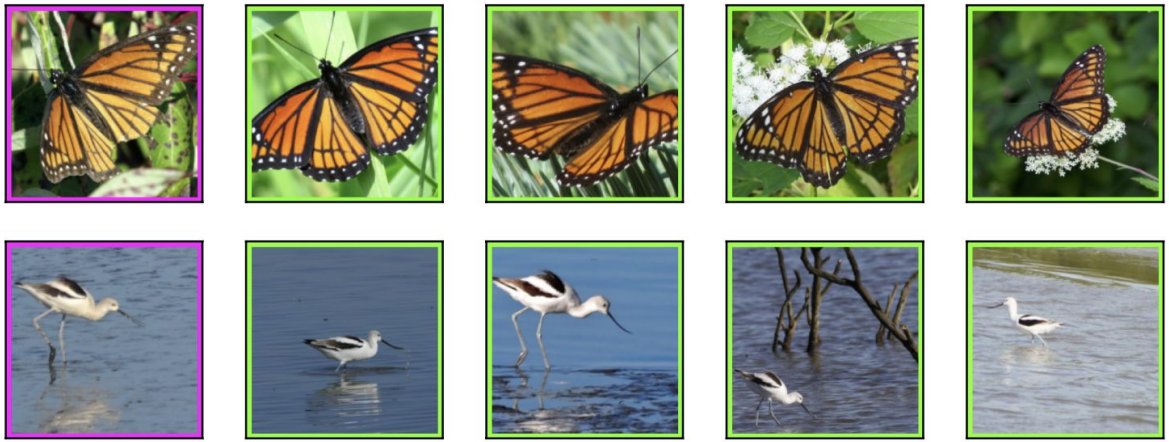
Conclusion on ROADMAP.

Pros:

- ➔ **Smooth** and **decomposable** surrogate for rank losses.
- ➔ Consistent and **significant improvement**s.

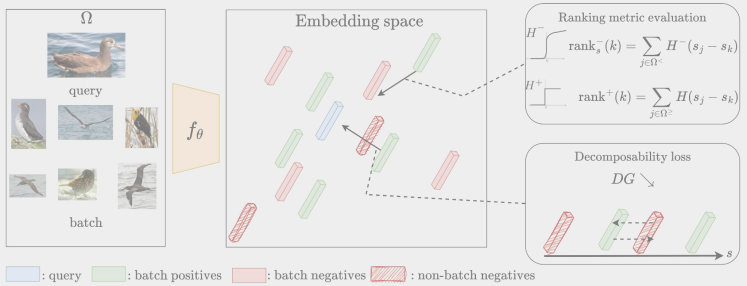
Limitation:

- ➔ Only defined for binary labels.

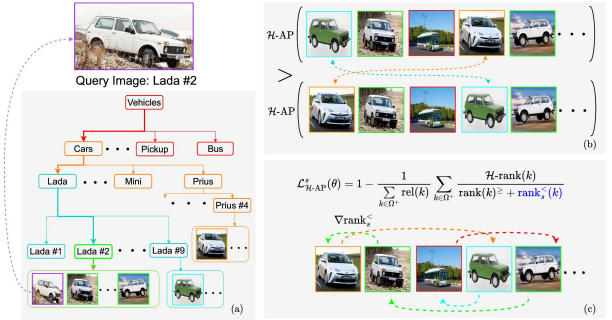


Summary and contributions.

🗺️ ROADMAP: Optimization of Ranking Losses for Image retrieval.



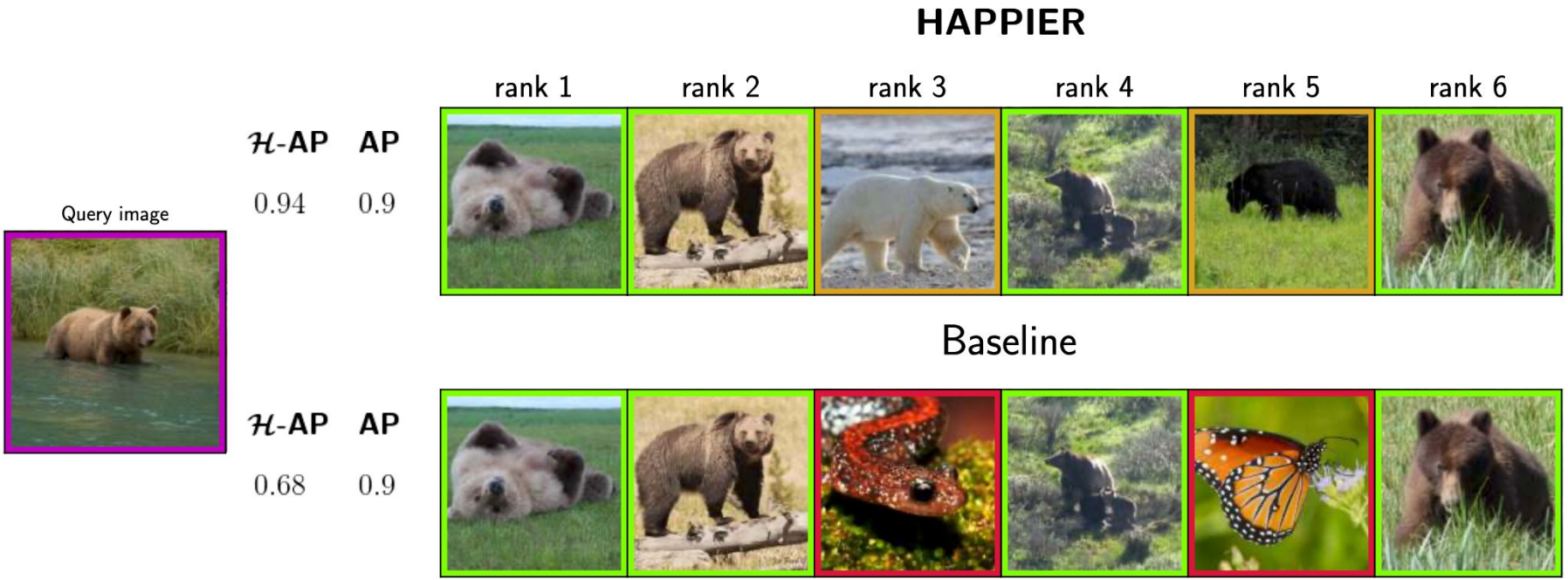
😺 HAPPIER: Hierarchical Image Retrieval for Robust Ranking.



📕 Ramzi, Elias, et al. “Hierarchical Average Precision Training for Pertinent Image Retrieval.” *ECCV* ,2022.
📕 Ramzi, Elias, et al. “Optimization of Rank Losses for Image Retrieval.” *TPAMI,* 2025.

Hierarchical Image Retrieval for Robust Ranking.

HAPPIER

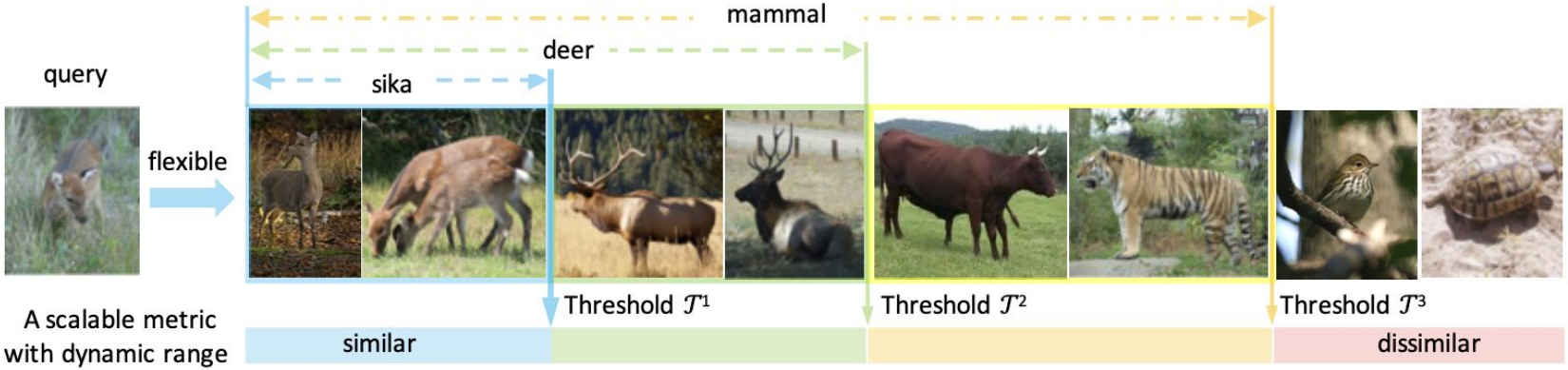


- Image retrieval is binary → do not take into account mistake severity.
- Extend AP to graded setting to take importance of errors into account.

Hierarchical learning.

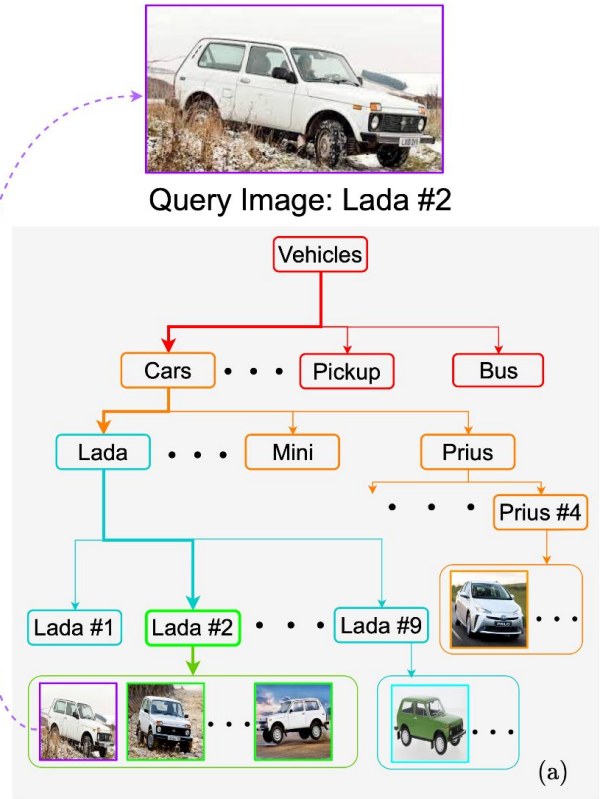
- Hierarchy is a good proxy for human perception of mistake severity.
- Several public datasets include hierarchical annotations.

Hierarchical image retrieval in the literature.

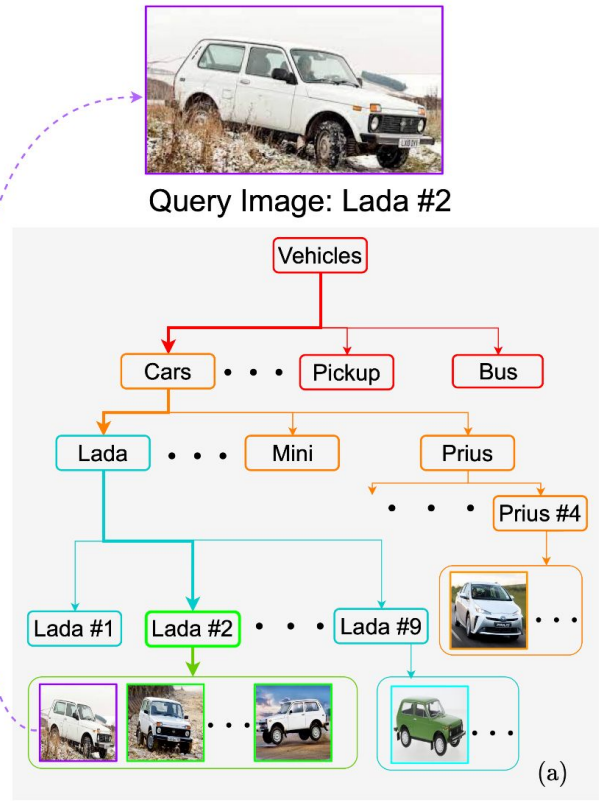


- ➔ Introduction of the “Dynamic Metric learning” datasets.
- ➔ CSL = triplet + proxy loss for hierarchical setting.
  - ◆ Limitation: **not aligned** with evaluation metrics.

📕 Sun, Yifan, et al. "Dynamic metric learning: Towards a scalable metric space to accommodate multiple semantic scales." *CVPR*, 2021.


Query Image: Lada #2
Vehicles
Cars
Pickup
Bus
Lada
Mini
Prius
Prius #4
Lada #1
Lada #2
Lada #9
(a)

Relevance function: graded similarities.



- Leverage a hierarchical tree to design a **graded similarity**, or "relevance" between categories.
- **Decreasing function of the distance** in the hierarchical tree.

$$\mathrm{rel}(k) = \frac{l/L}{|\Omega^{(l)}|}$$

- l: level of the closest ancestor in the tree.
- L total number of levels.

Hierarchical rank.

$$\mathcal{H}\text{-rank}(k) = \text{rel}(k) + \sum_{j \in \Omega^+} \min(\text{rel}(k), \text{rel}(j)) \cdot H(s_j - s_k)$$

Query Image:
Lada #2

$$\mathcal{H}\text{-rank}(1) = 1/3 = 1/3$$
$$\mathcal{H}\text{-rank}(2) = 1 + (1/3) = 4/3$$
$$\mathcal{H}\text{-rank}(4) = 2/3 + (1/3 + 2/3) = 5/3$$
$$\mathcal{H}\text{-rank}(5) = 1 + (1/3 + 1 + 2/3) = 3$$

Lada #2
rel := 1

Lada #9
rel := 2/3

Prius #4
rel := 1/3

Bus
rel := 0

- Errors in ranking → **weighted by relevance**.
- Correct $\mathcal{H}$-rank → decreasing order of relevance.

Hierarchical average precision.

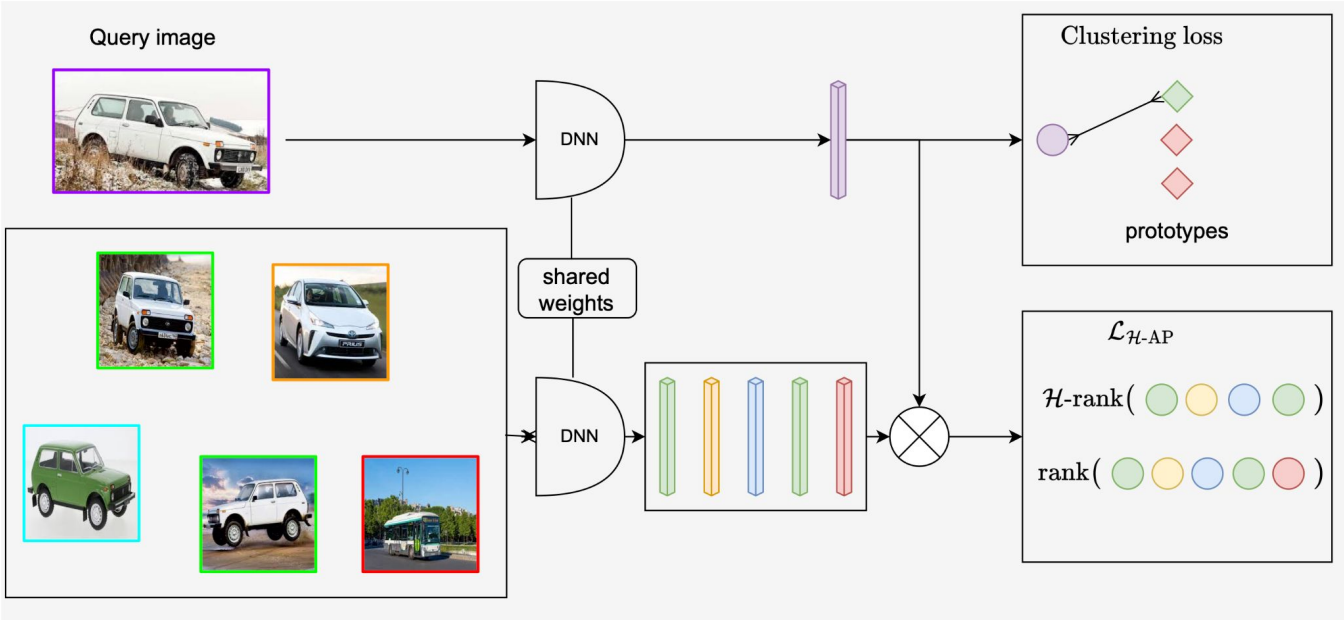
HAPPIER

$$\mathcal{H}\text{-AP} = \frac{1}{\sum_{k \in \Omega^+} \text{rel}(k)} \sum_{k \in \Omega^+} \frac{\mathcal{H}\text{-rank}(k)}{\text{rank}(k)}$$

$\mathcal{H}$-AP properties:

- Consistent generalization of AP.
- Keeps desirable properties of AP.
  - Penalize wrong rankings.
  - Emphasis for the top of the ranking.
  - Relevant in imbalanced settings.
- Is flexible wrt. the relevance.

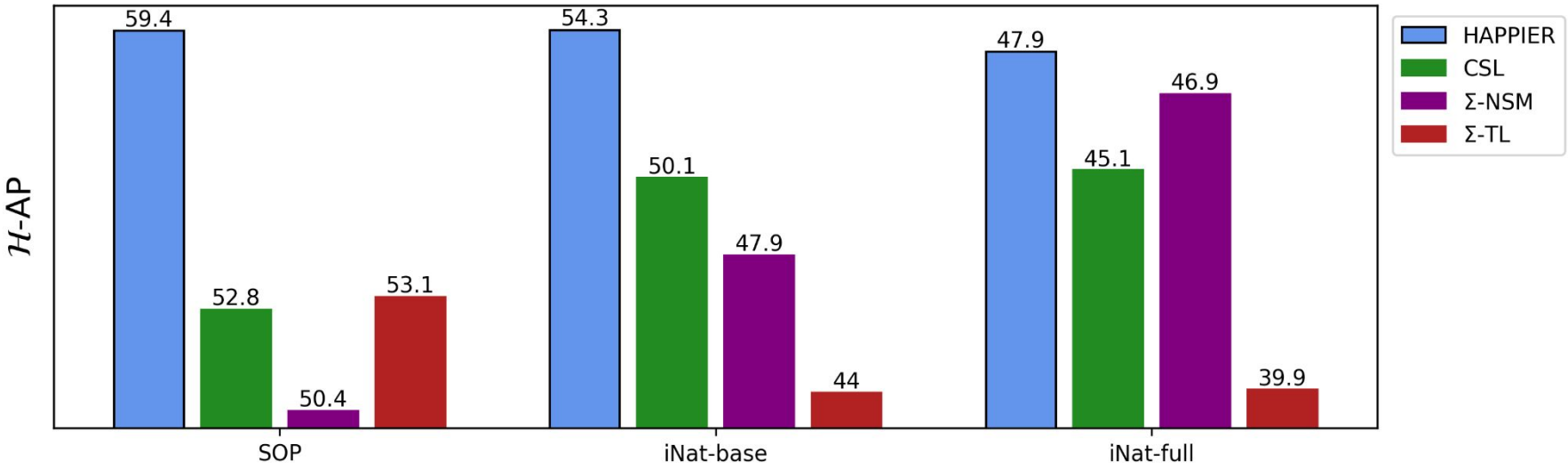
Optimizing $\mathcal{H}$-AP.



$$\mathcal{L}_{\text{Sup-}\mathcal{H}\text{-AP}}(\theta) = 1 - \frac{1}{\sum_{k \in \Omega^+} \text{rel}(k)} \sum_{k \in \Omega^+} \frac{\mathcal{H}\text{-rank}(k)}{\text{rank}^+(k) + \text{rank}_s^-(k)}$$

$$\mathcal{L}_{\text{HAPPIER}}(\theta) = (1 - \lambda) \cdot \mathcal{L}_{\text{Sup-}\mathcal{H}\text{-AP}}(\theta) + \lambda \cdot \mathcal{L}^*_{\text{DG}}(\theta)$$

Comparison to hierarchical methods.

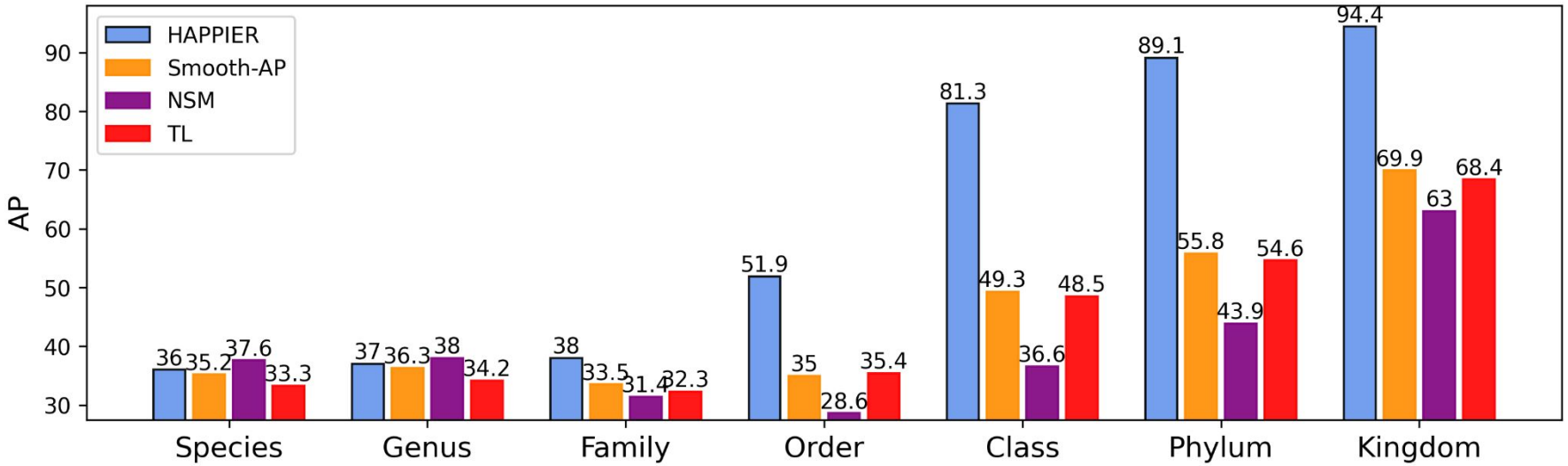


➔ HAPPIER outperforms the state-of-the-art hierarchical loss CSL on fine-grained and hierarchical retrieval.

TL: Wu, Chao-Yuan, et al. *ICCV*, 2017.
NSM: Zhai, Andrew, et al. *BMVC*, 2018.
CSL: Sun, Yifan, et al. *CVPR*, 2021.

Comparison to fine-grained methods.

HAPPIER



➔ On par for fine-grained retrieval (“Species”).

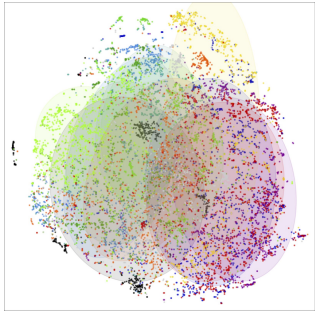
➔ **Large gains** on other hierarchical levels from “Family”.

TL: Wu, Chao-Yuan, et al. ICCV, 2017.

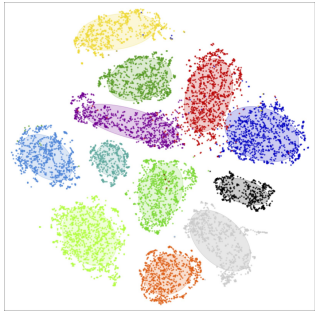
NSM: Zhai, Andrew, et al. BMVC, 2018.

Qualitative results.

HAPPIER

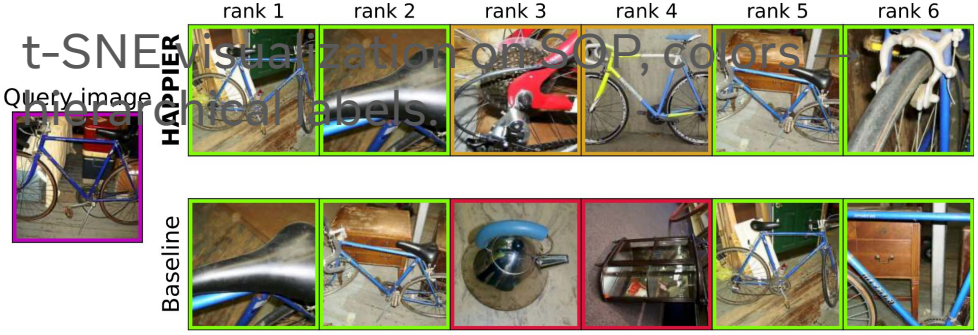


Baseline



HAPPIER

t-SNE visualization of SOP, colors: hierarchical labels.



HAPPIER:

➔ High quality of clusters.

➔ Better mistakes, even in failure cases.

$\mathcal{H}$-GLDv2: a hierarchical landmark dataset.

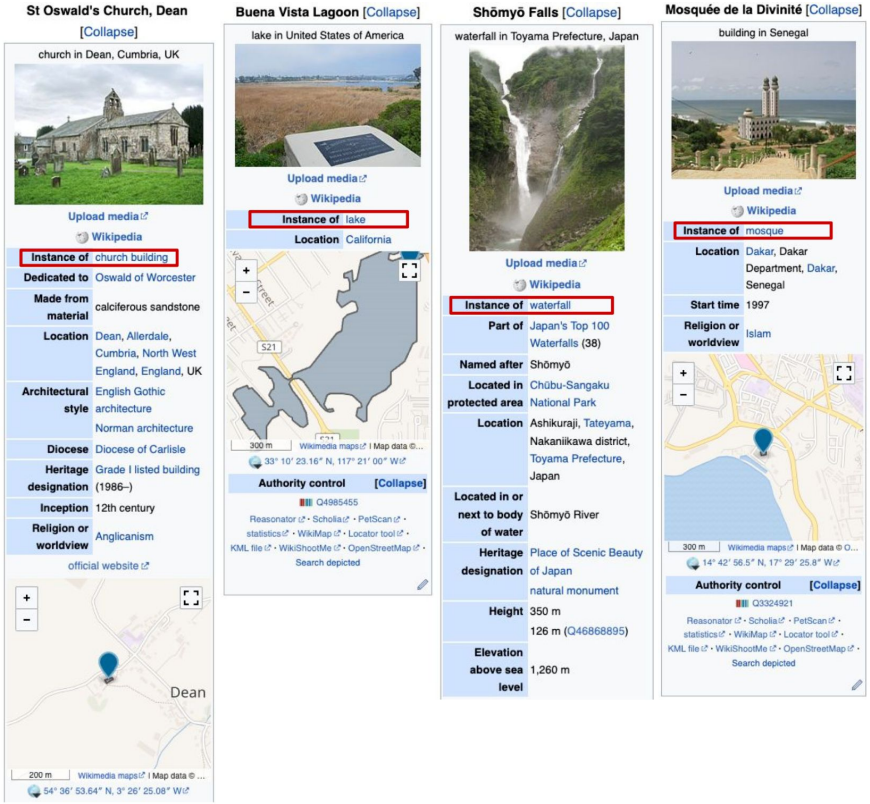
HAPPIER

GLDv2 → **large scale** landmarks retrieval dataset.

No hierarchical annotations → how difficult is it to create hierarchical annotations?

📕 Weyand, Tobias, et al. "Google landmarks dataset v2 - a large-scale benchmark for instance-level recognition and retrieval." *CVPR,* 2020.

Scraping Wikimedia Commons.

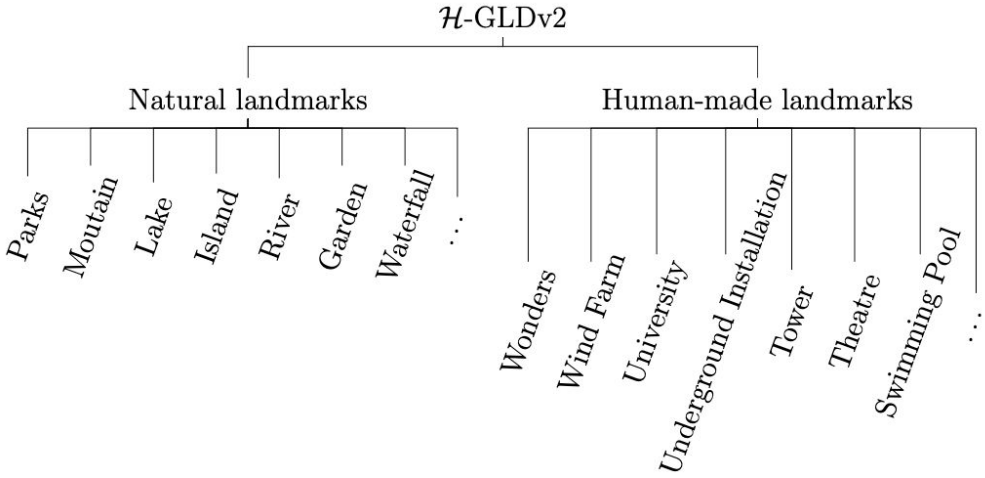


Wikimedia Commons → the largest open database of landmarks.

Scraped labels include:

- Church building.
- Church building (1172–1954)
- Cathedral.
- Castle.
- Corsican nature reserve.
- New Zealand great walks.
- Waterfall.
- Arch-gravity dam.
- Canal.
- Association football venue.
- Astronomical observatory.
- Village.

Final super categories.



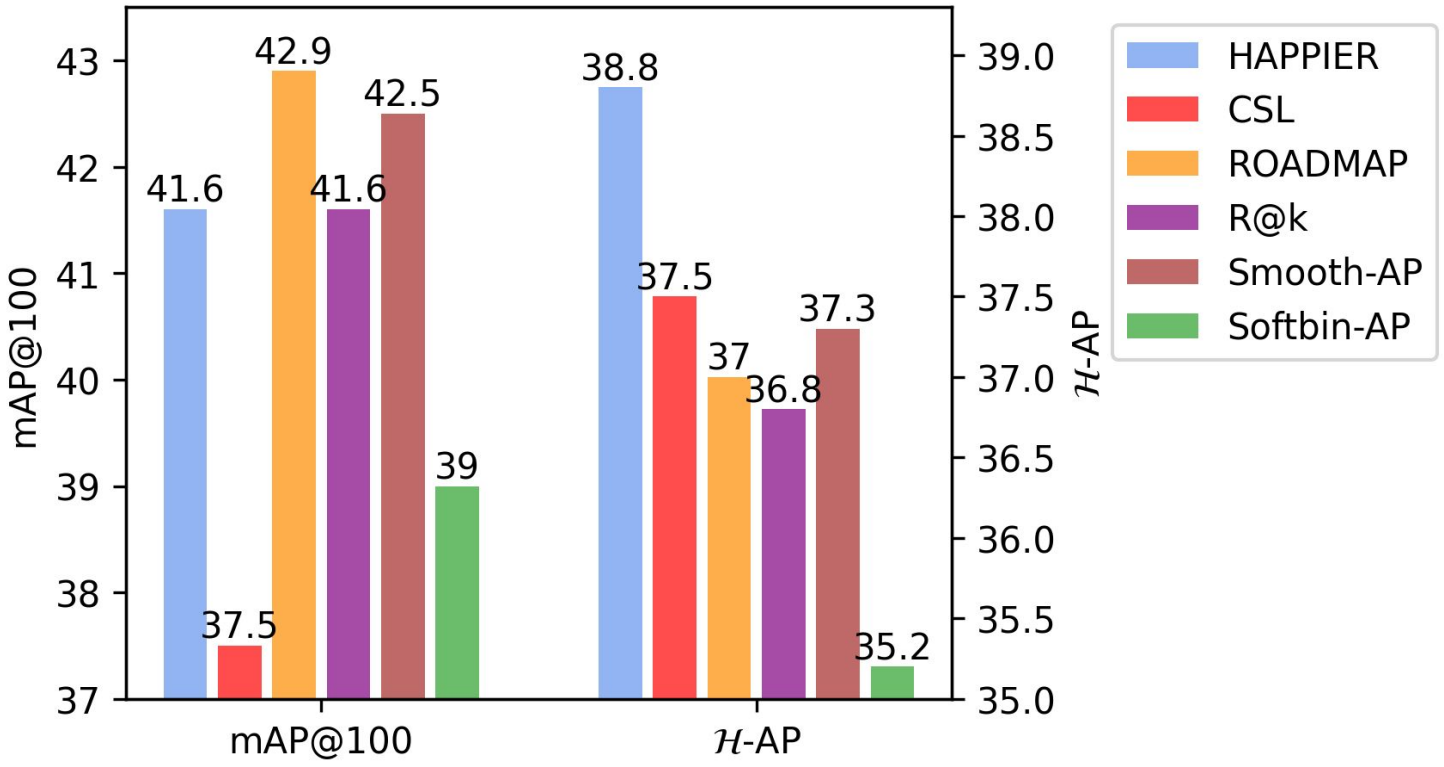

Bridge.


Castle.

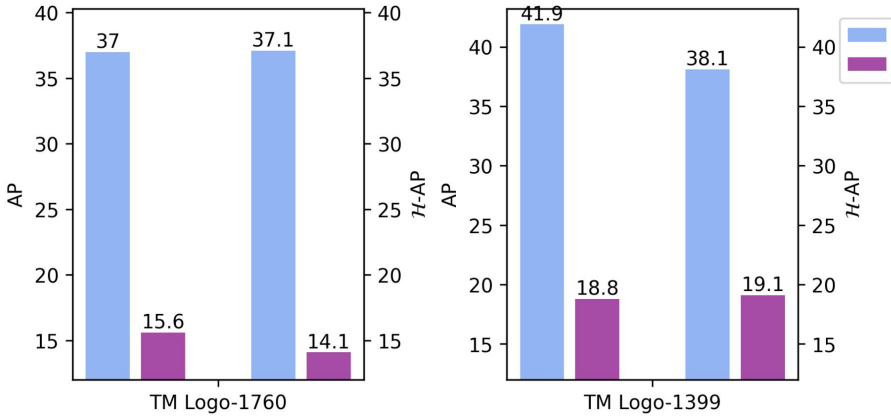

Waterfall.


Volcano.

$\mathcal{H}$-GLDv2 results.



Results on Coexya data.



➔ HAPPIER can be extended to **multi-label setting**.

➔ Optimizing HAPPIER works best on both **fine-grained** and **hierarchical metrics**.

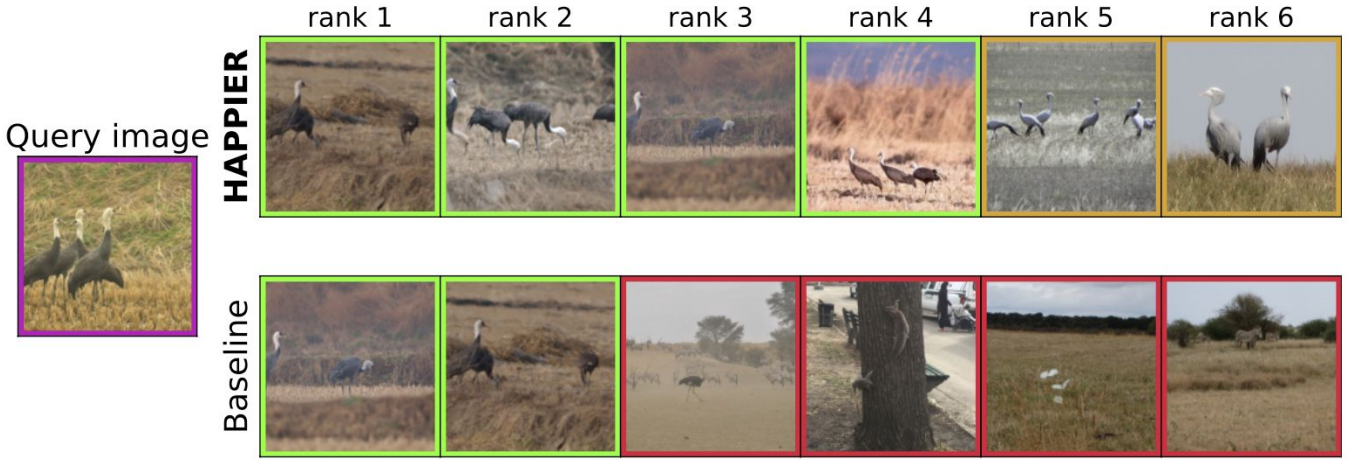
Qualitative results on Coexya data.

- HAPPIER → retrieves images with **both labels**.

Conclusion on HAPPIER.

HAPPIER

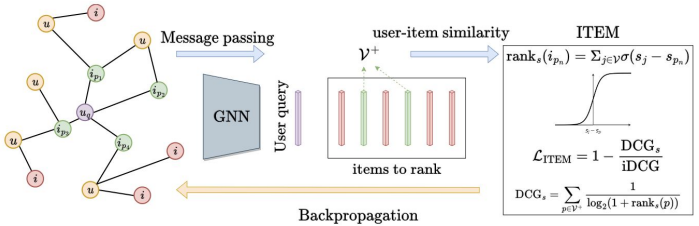
- Optimizing hierarchical metrics:
  - **Better robustness** wrt. to mistake severity.
  - Performances on par on fine-grained metrics.
- Hierarchical annotations → **not too costly** + boost models' robustness.
- HAPPIER → production in Acsepto



Short term extensions.

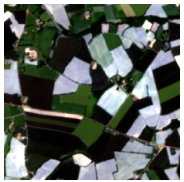
🗺️ ROADMAP:

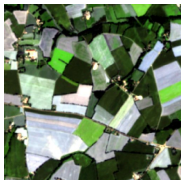
- Optimize other metrics: Recall, NDCG.
- Use in GNN recommender systems.



😸 HAPPIER:

- Use in multi-label setting + business driven insights.
- Aerial image time series ranking.


2019-06-01


2019-08-30

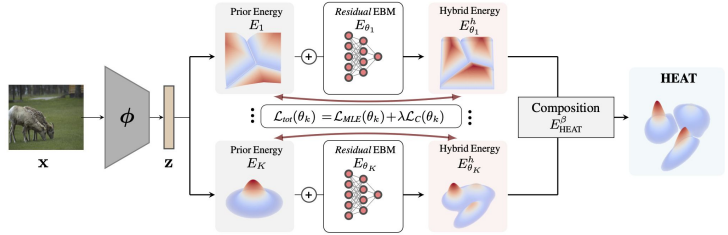

2019-09-19

Few shot robustness in the era of foundation models.

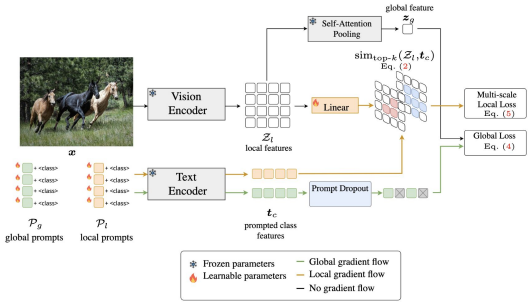
le cnam
cedric EA4629
COEXYA
Connect skills, create more
valeo.ai

Summary and contributions.

🔥HEAT:
Post-hoc out-of-distribution detection.



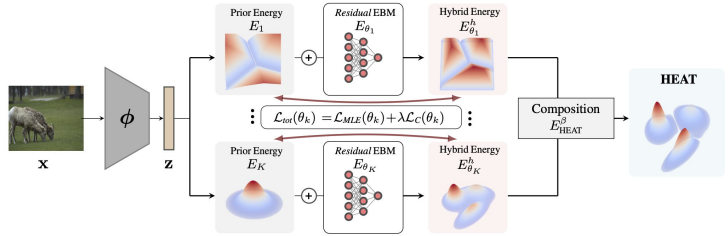
🐎 GalLoP:
Few shot adaptation of vision-language models.



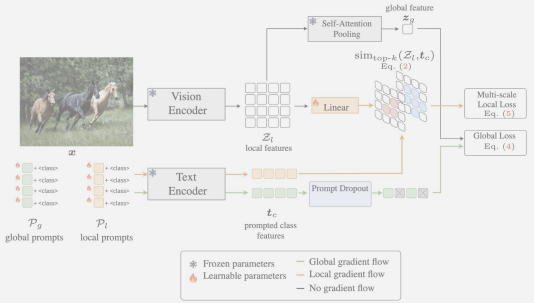
📕 Lafon, Marc, et al. “Hybrid Energy Based Model in the Feature Space for Out-of-Distribution Detection.” *ICML*, 2023.
📕 Lafon, Marc, et al. “GalLoP: Learning global and local prompts for vision-language models”, *ECCV*, 2024

Summary and contributions.

🔥HEAT:
Post-hoc out-of-distribution detection.



🐎 GalLoP:
Few shot adaptation of vision-language models.



📕 Lafon, Marc, et al. “Hybrid Energy Based Model in the Feature Space for Out-of-Distribution Detection.” *ICML*, 2023.

Out-of-distribution detection.

AI confuses a horse-drawn carriage (OOD) for a truck.

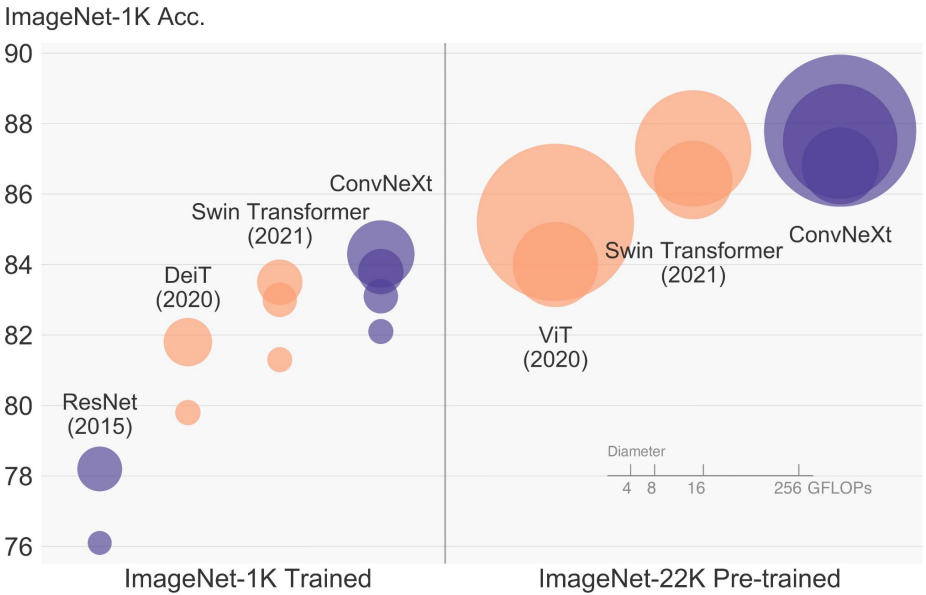
- Detection of in-distribution (ID) = similar to train dataset, vs. out-of-distribution (OOD) = dissimilar to train dataset.
- Important task for real world deployment.

OOD → binary problem:

$$G_\lambda(x) = \begin{cases} \text{ID}, & \text{if } E(x) \leq \lambda_{95\%} \\ \text{OOD}, & \text{if } E(x) > \lambda_{95\%} \end{cases}$$

Post-hoc out-of-distribution detection.

HEAT

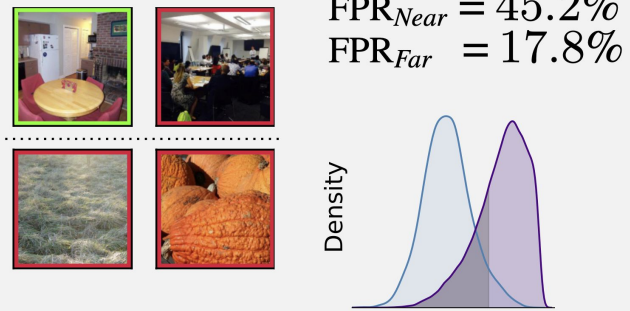


- Enjoy strong predictive performances from off-the-shelf neural networks.
- Use any backbones → ConvNets, transformers.
- Limited compute overhead at inference time.

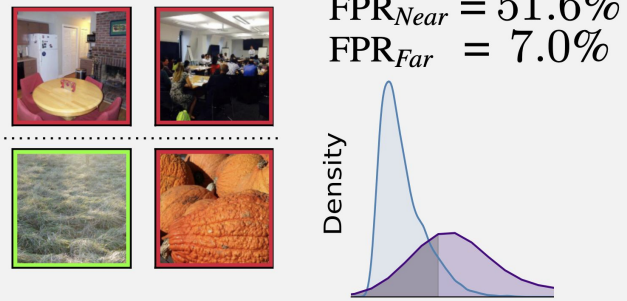
Prior out-of-distribution scorers.

HEAT

Prior K - (*EL*)



Prior 1 - (*GMM*)



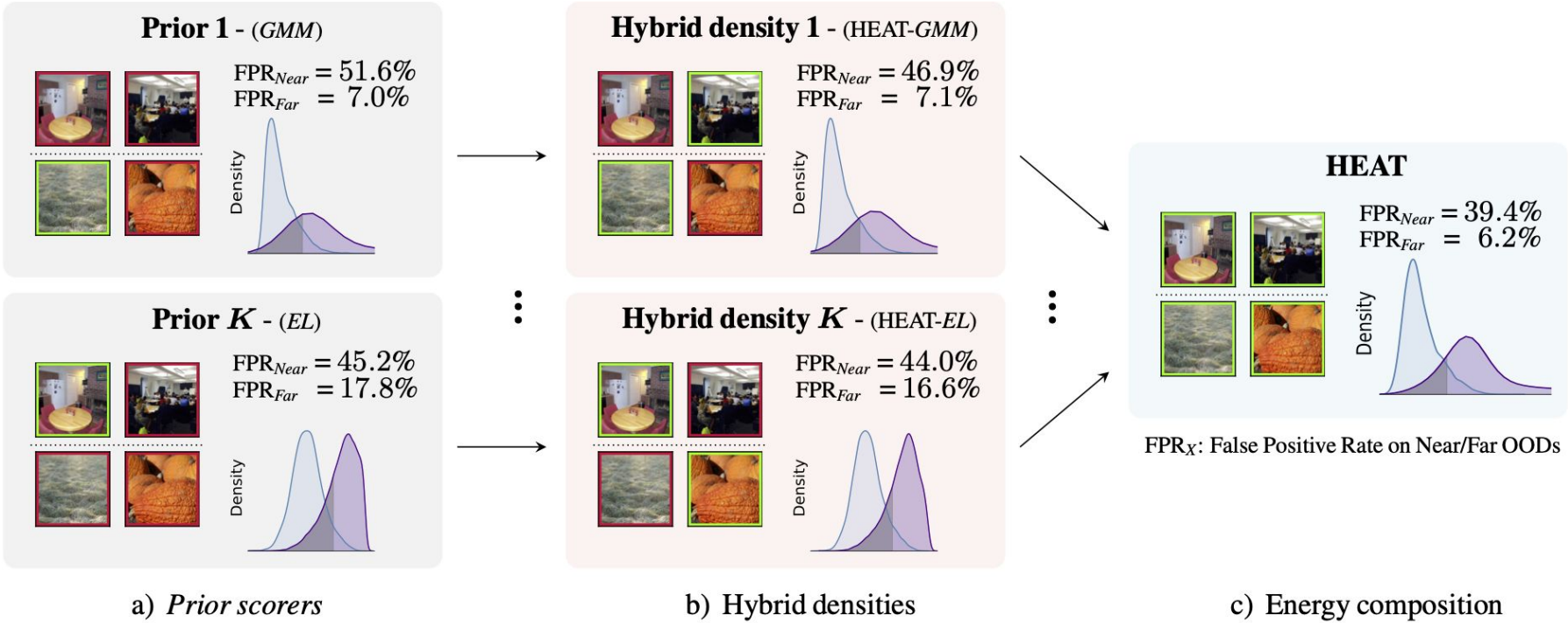
Classification based methods:
→ Better at distinguishing near-OOD.
→ No control for far-OOD.

Parametric methods:
→ Strong guarantees for far-OOD.
→ Lower performances on near-OOD.

- Energy logits (EL): Liu, Weitang, et al. NeurIPS, 2020.
- DICE: Sun, Yiyou, et al. ECCV, 2022.

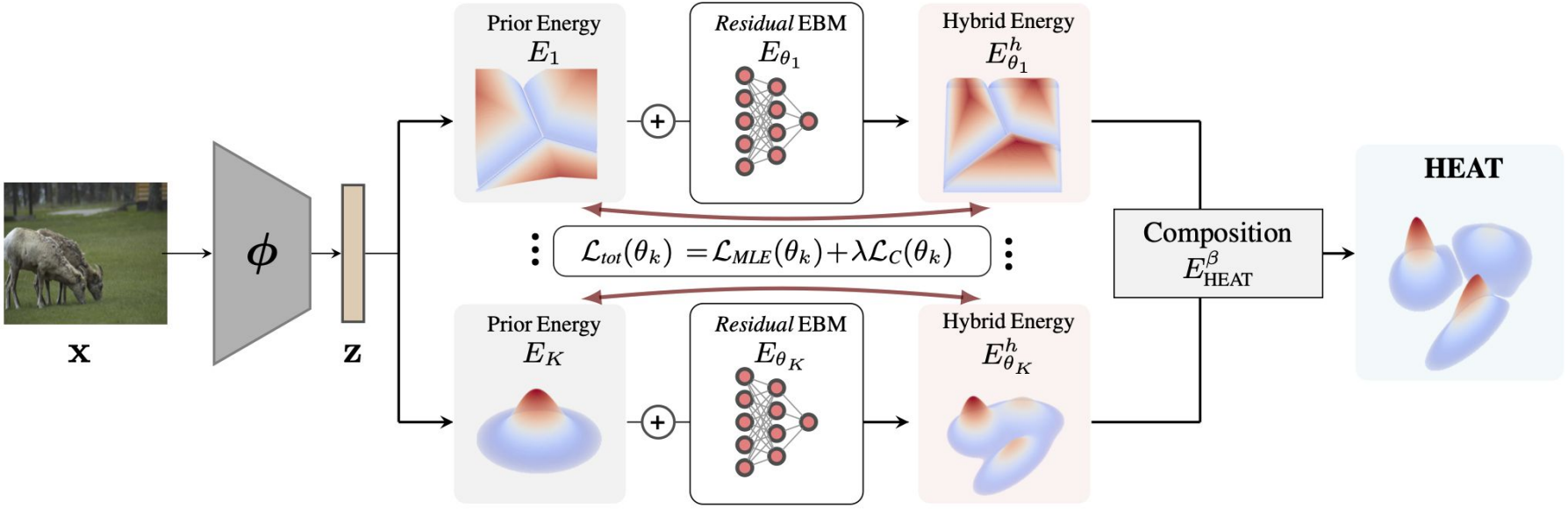
- SSD (GMM): Sehwag, Vikash, et al. ICLR, 2021.
- KNN: Sun, Yiyou, et al. ICML, 2022.

Motivations.



HEAT for out-of-distribution detection.

HEAT



- EBM → data driven residual learning.
- Composition → leverage ≠ modeling biases.

Hybrid energy-based density estimation.

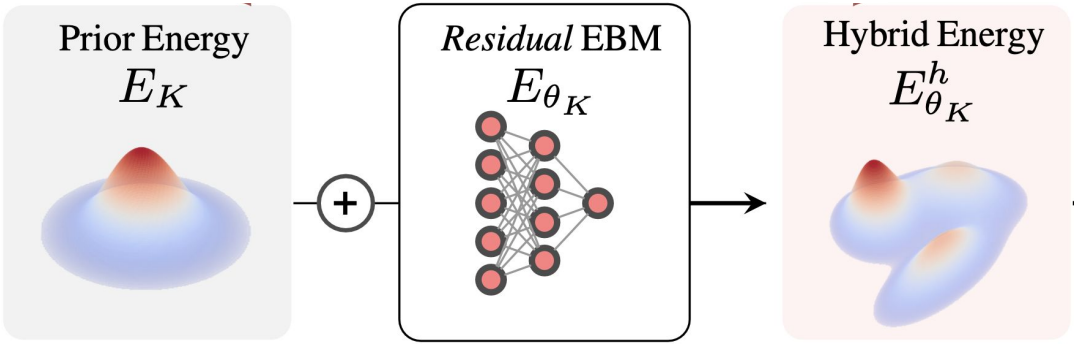
HEAT

Hybrid density:

$$p^h_{\theta_k}(z) = \frac{1}{Z(\theta_k)} p^r_{\theta_k}(z) q_k(z)$$

Hybrid energy:

$$E^h_{\theta_k}(z) = E_{q_k}(z) + E_{\theta_k}(z)$$



EBM – LeCun, Yann, et al. *Predicting structured data* 1.0 (2006).
Energy correction – Deng, Yuntian, et al. *ICLR 2020*.

Composition of refined prior density estimators.

HEAT

Class prior:

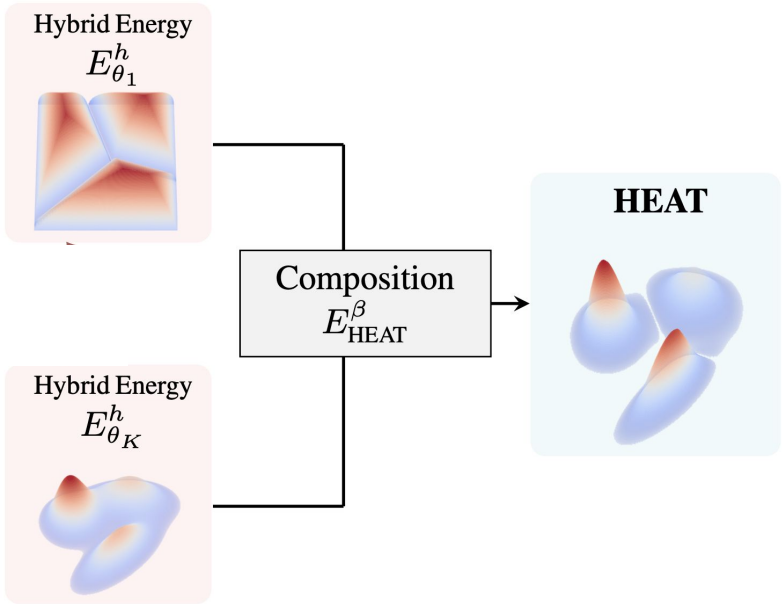
$$E^h_{\theta_l}(z) = -\log \sum_c e^{f(z)[c]} + E^r_{\theta_l}(z)$$

Feature & style prior:

$$E^h_{\theta_g}(z) = -\log \sum_c e^{-\frac{1}{2}(z-\mu_c)^T \Sigma^{-1}(z-\mu_c)} + E^r_{\theta_g}(z)$$

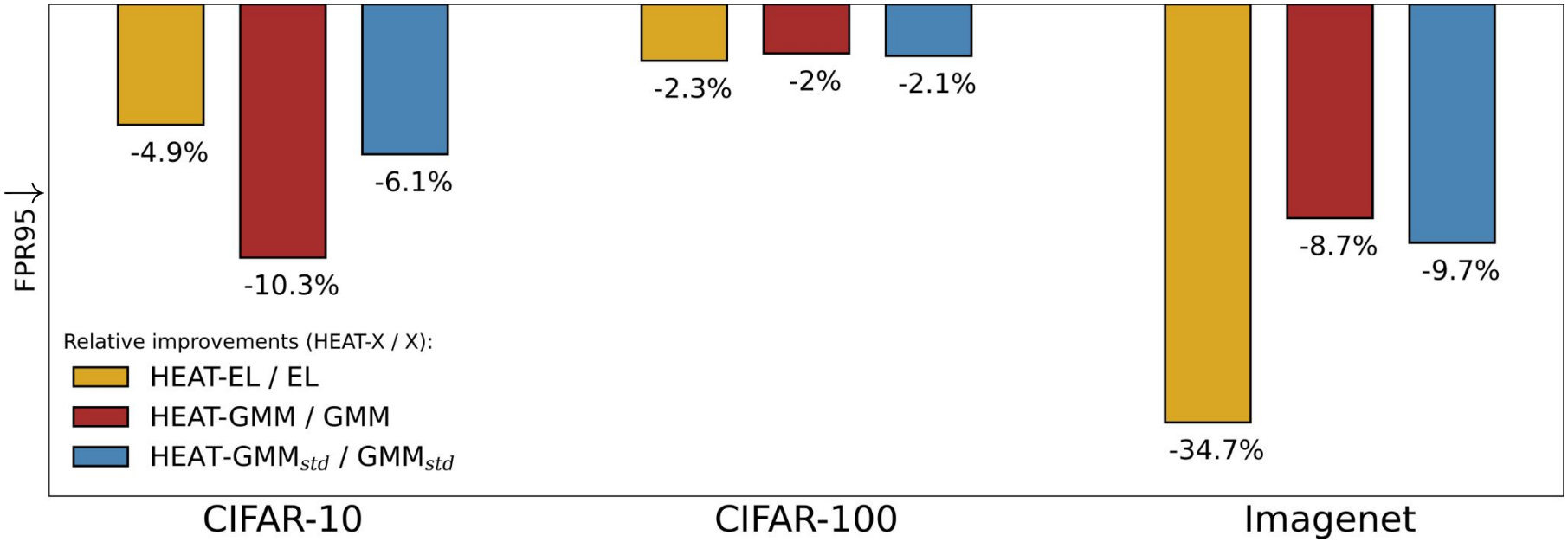
Energy function composition:

$$E^\beta_{\text{HEAT}} = \frac{1}{\beta} \log \sum_{k=1}^{K} e^{\beta E^h_{\theta_k}}$$



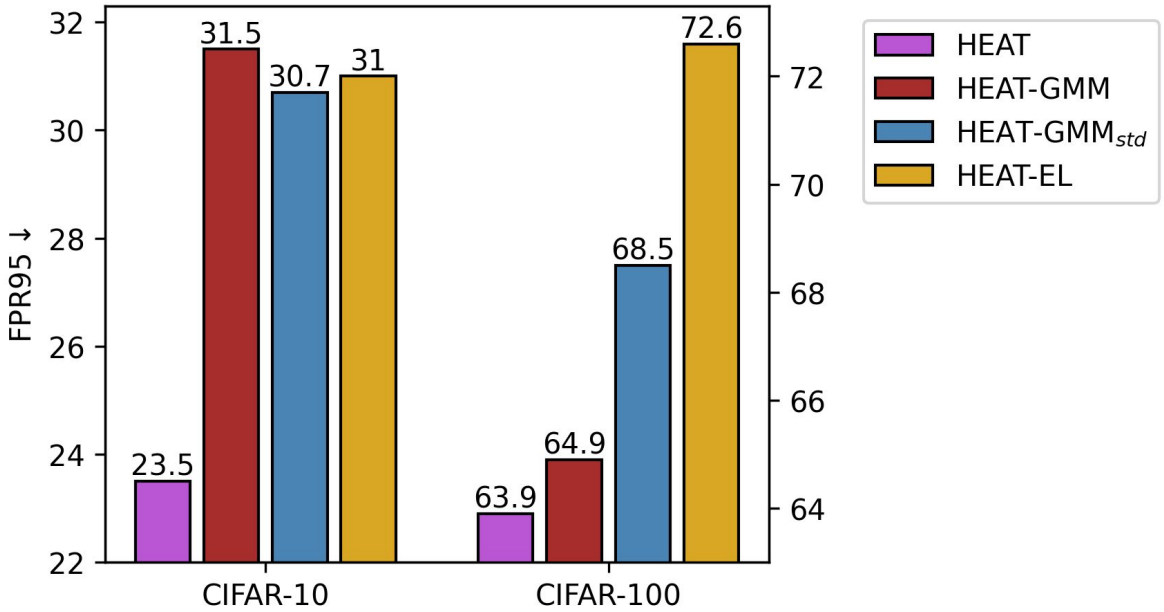
EL – Liu, Weitang, et al. NeurIPS, 2020.
GMM – Lee, Kimin, et al. NeurIPS, 2018.
Gram matrix – Sastry, Chandramouli Shama, et al. ICML, 2020.

Residual learning.



- Residual learning → better performances

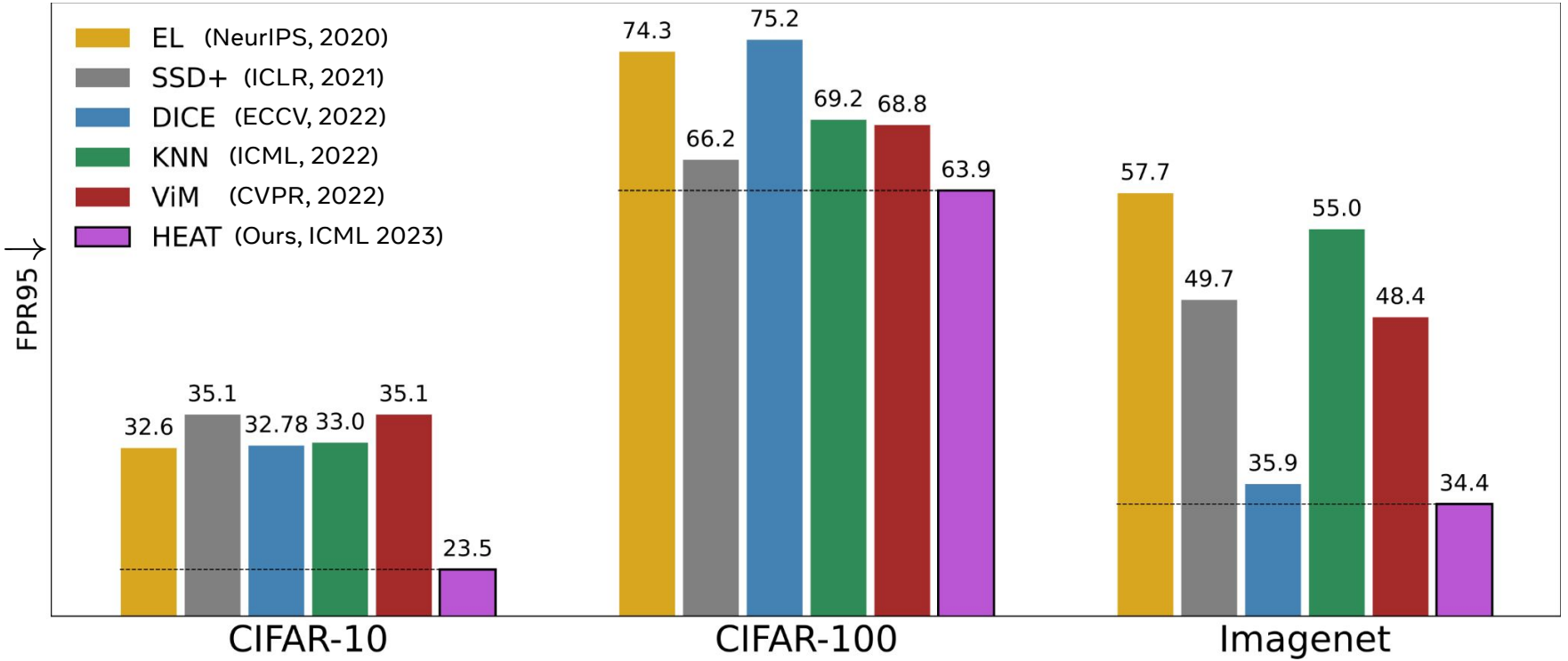
Composition.



- Results better than individual scorers.
- Composition → performance boost.

Comparison to state-of-the-art methods.

HEAT

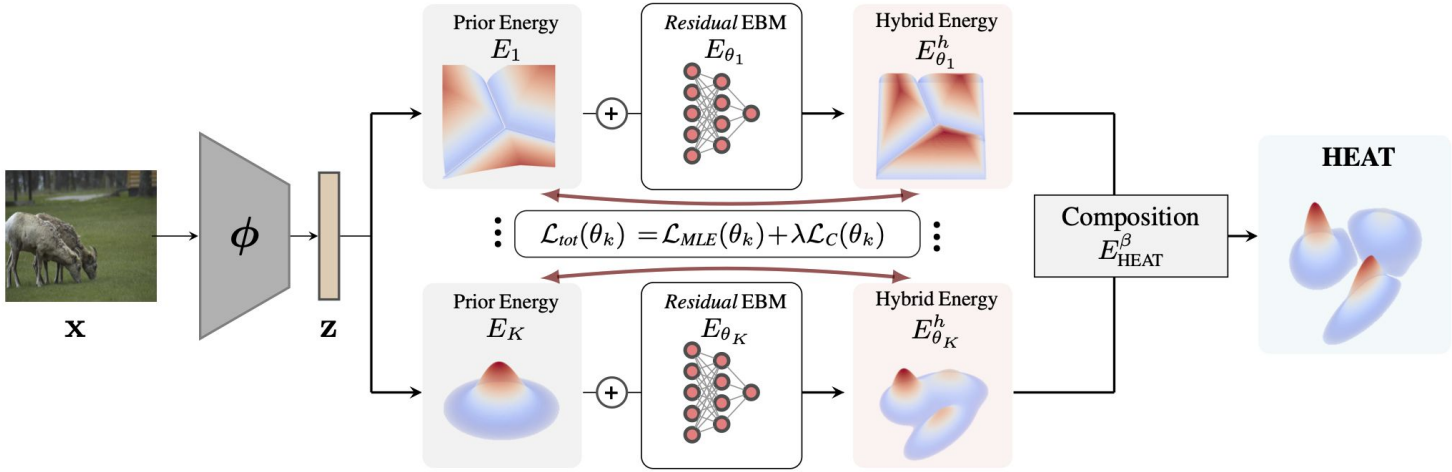


- HEAT → competitive performances on 3 datasets

Conclusion on HEAT.

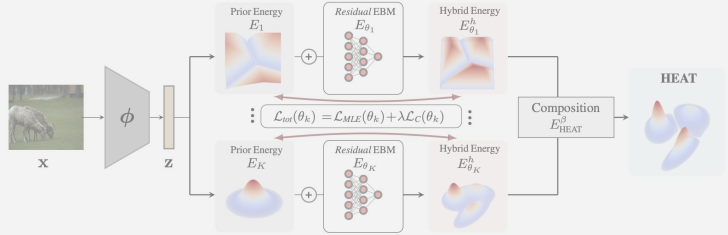
HEAT

- Post-hoc OOD detection → no assumption on models.
- EBM models boost prior OOD scorers.
- Composing refined OOD scorers → boost performances.

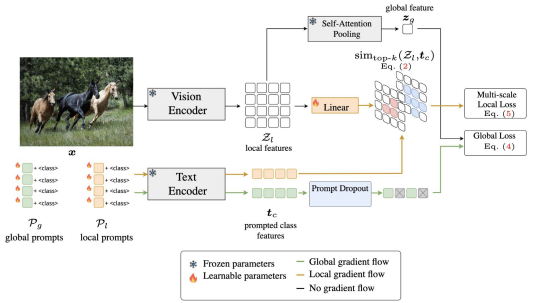


Summary and contributions.

🔥HEAT:
Post-hoc out-of-distribution detection.

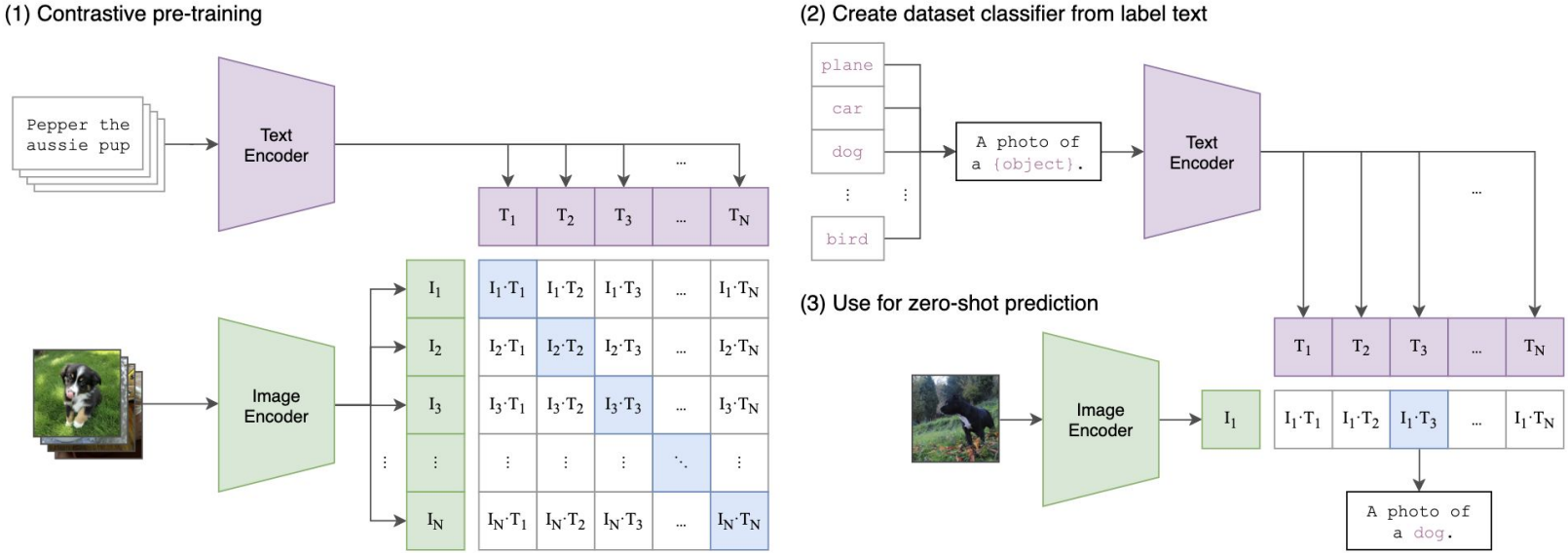


🐎 GalLoP:
Few shot adaptation of vision-language models.



📕 Lafon, Marc, et al. “GalLoP: Learning global and local prompts for vision-language models”, *ECCV*, 2024

Vision-language models.

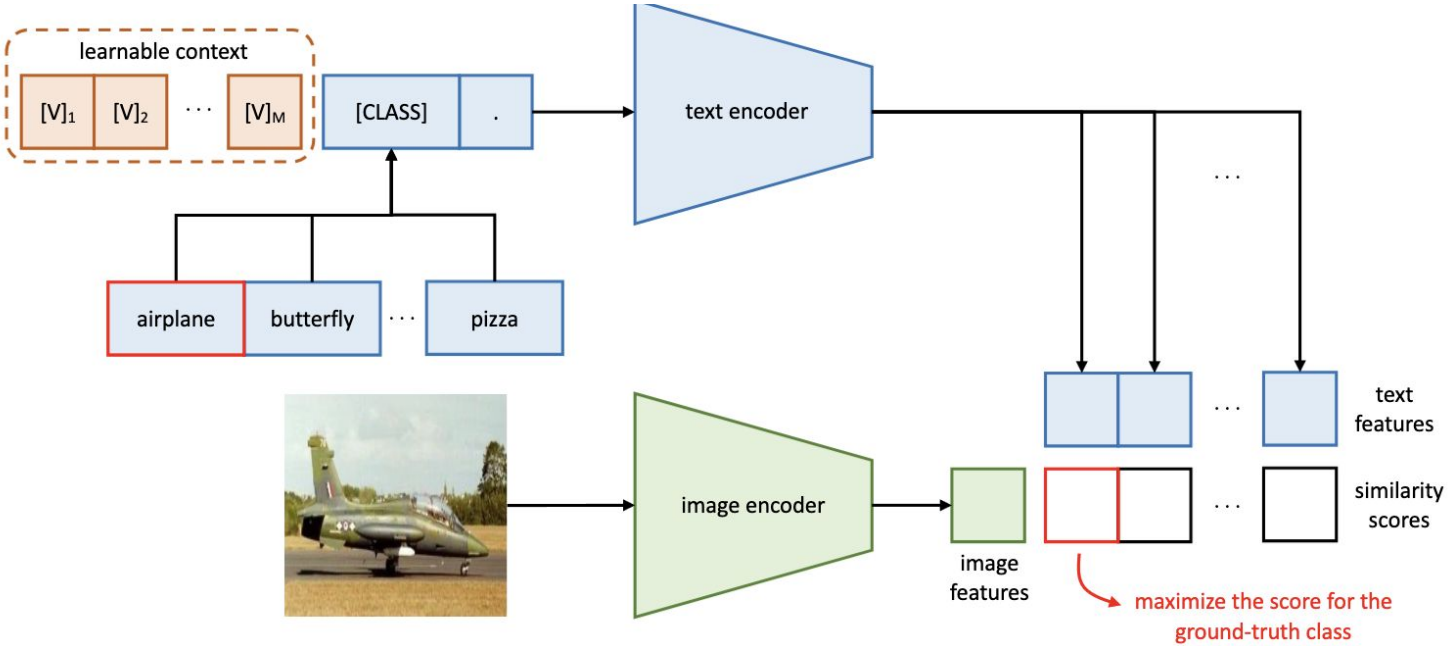


CLIP:

- Vision-language pre-training.
- Zero-shot classification on downstream datasets.

CLIP – Radford, Alec, et al. ICML, 2021.
ALIGN – Jia, Chao, et al. ICML, 2021.

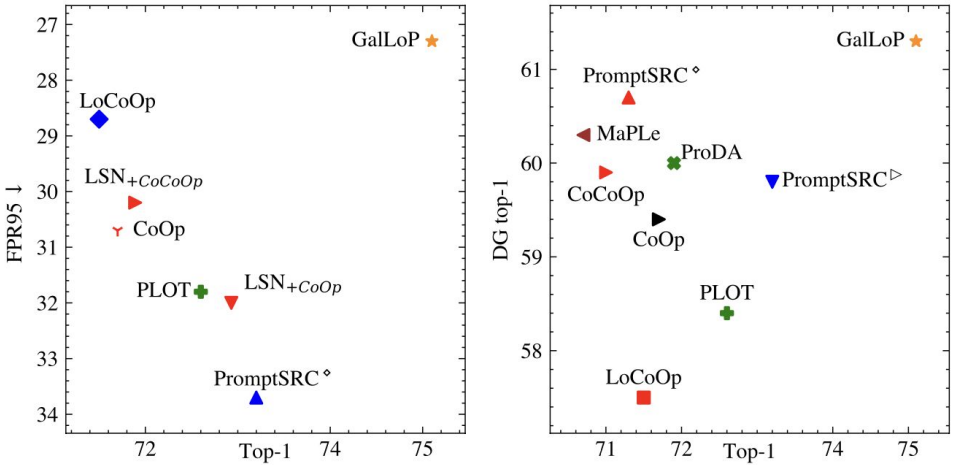
Few shot adaptation with prompt learning.



- Few shot adaptation of VLMs to downstream dataset.
- No need for “prompt engineering”.

CoOp – Zhou, Kaiyang, et al. IJCV, 2022.

Motivations.

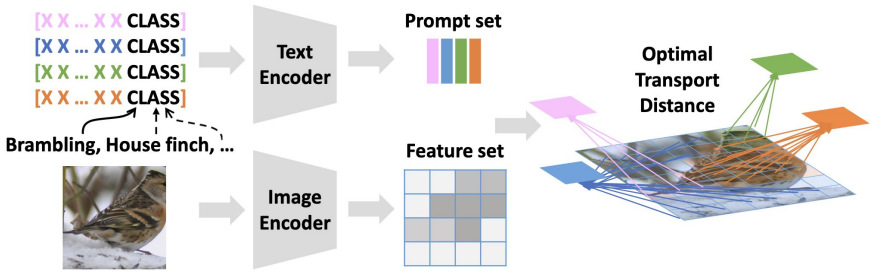


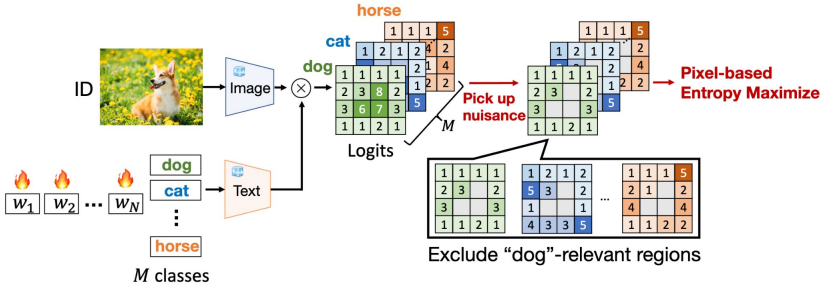
(a) OOD detection (b) Domain Generalization

- Prompt learning trade off top-1 vs. robustness.
- Ensembling and local features → robustness.

LoCoOp – Miyai, Atsuyuki, et al. NeurIPS, 2023.
PromptSRC – Khattak, Muhammad Uzair, et al. ICCV, 2023.

Prompt learning with local features.





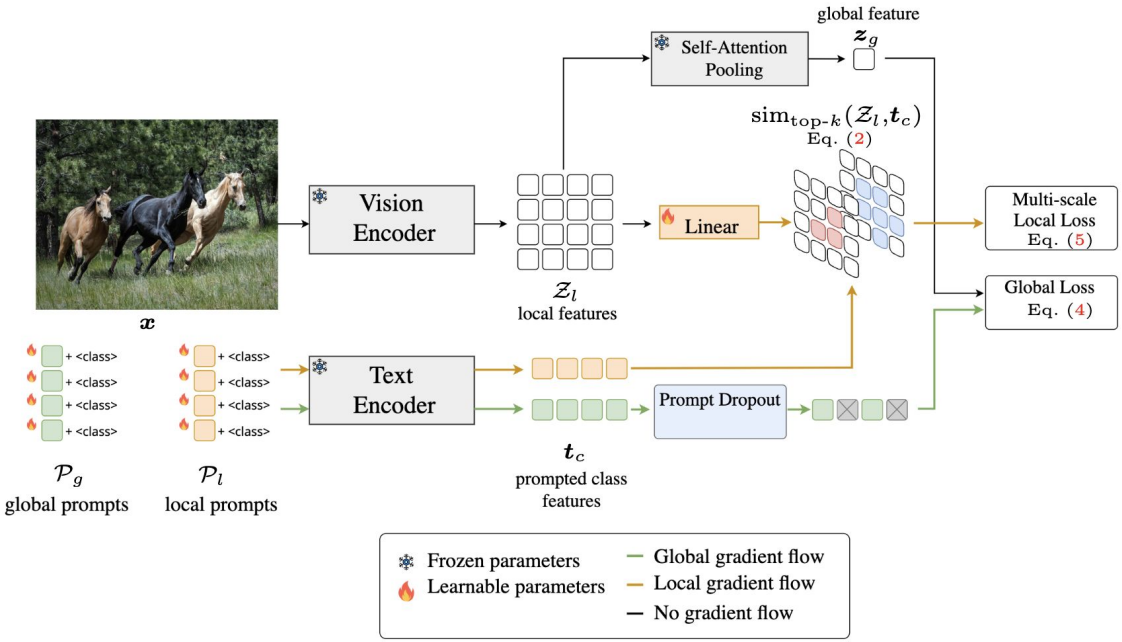
- Optimal transport for assignment.
- All local features → including irrelevant features.

📕 PLOT – Chen, Guangyi, et al. "Plot: Prompt learning with optimal transport for vision-language models." ICLR, 2023.

- Local features for regularization.
- Lower top-1 accuracy

📕 LoCoOp – Miyai, Atsuyuki, et al. "Locoop: Few-shot out-of-distribution detection via prompt learning." NeurIPS, 2023.

GalLoP.



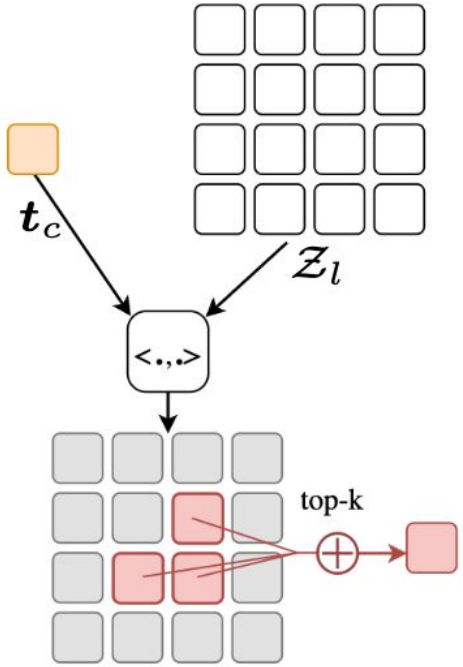
- Multiple global prompts w/ “prompt dropout”.
- Local alignment w/ sparsity and linear projection.
- Multiple local prompts w/ multiscale loss.

Prompt learning on local features.

GalLoP

$$\mathrm{sim}_{\text{top-}k}(\mathcal{Z}_l, \boldsymbol{t}_c) \coloneqq \frac{1}{k}\sum_{i=1}^{L} \mathbb{1}_{\text{top-}k}(i) \cdot \langle \boldsymbol{z}_i^l, \boldsymbol{t}_c \rangle$$

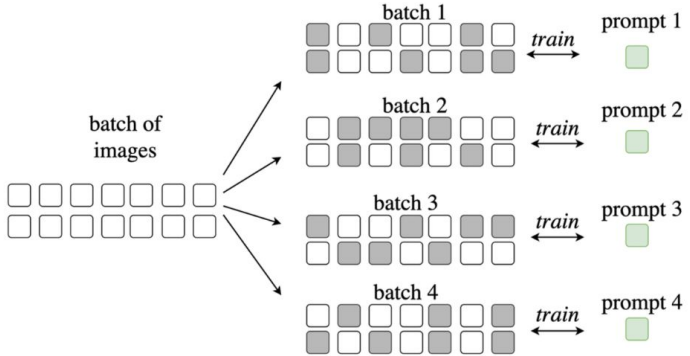
$$\text{where} \qquad \mathbb{1}_{\text{top-}k}(i) = \begin{cases} 1 & \text{if} \quad \mathrm{rank}_i(\langle \boldsymbol{z}_i^l, \boldsymbol{t}_c \rangle) \leq k, \\ 0 & \text{otherwise.} \end{cases}$$



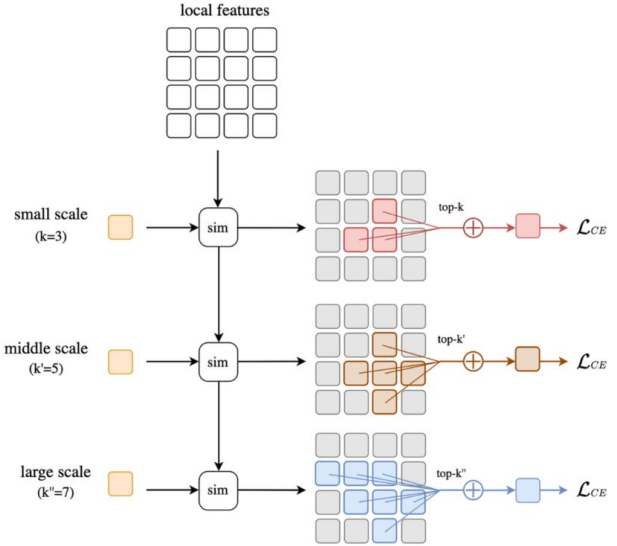
Exploit local features for prompt learning:

- Sparse local similarity.
- Learnable projection layer.

Diversity: prompt dropout & multiscale loss.



(a) Prompt dropout

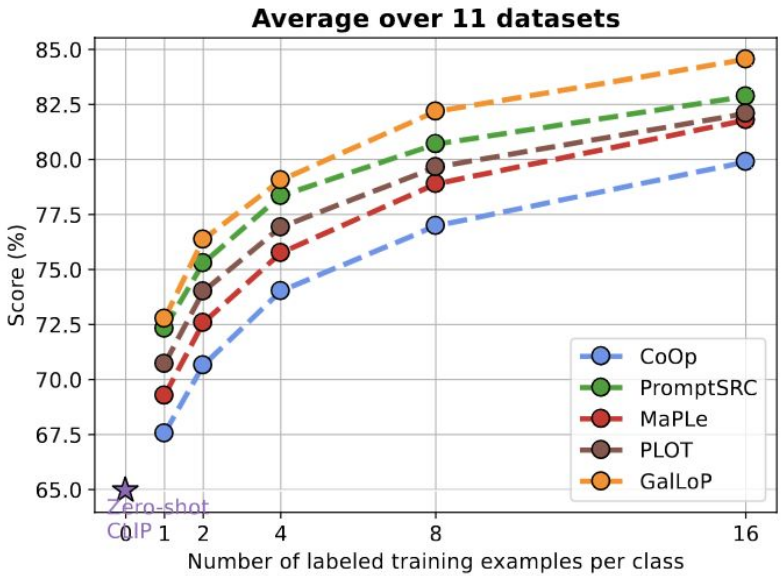


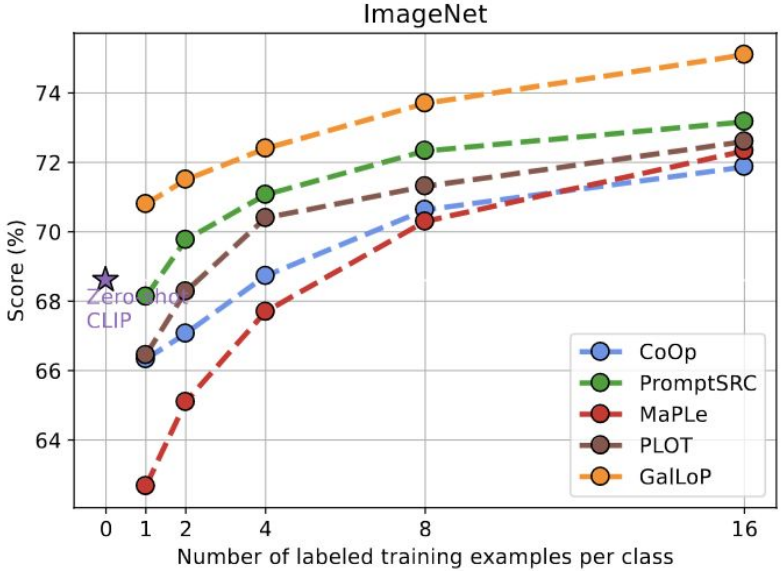
(b) Multiscale loss

- Diversity by enforcing different gradient signal.

Few shots results.

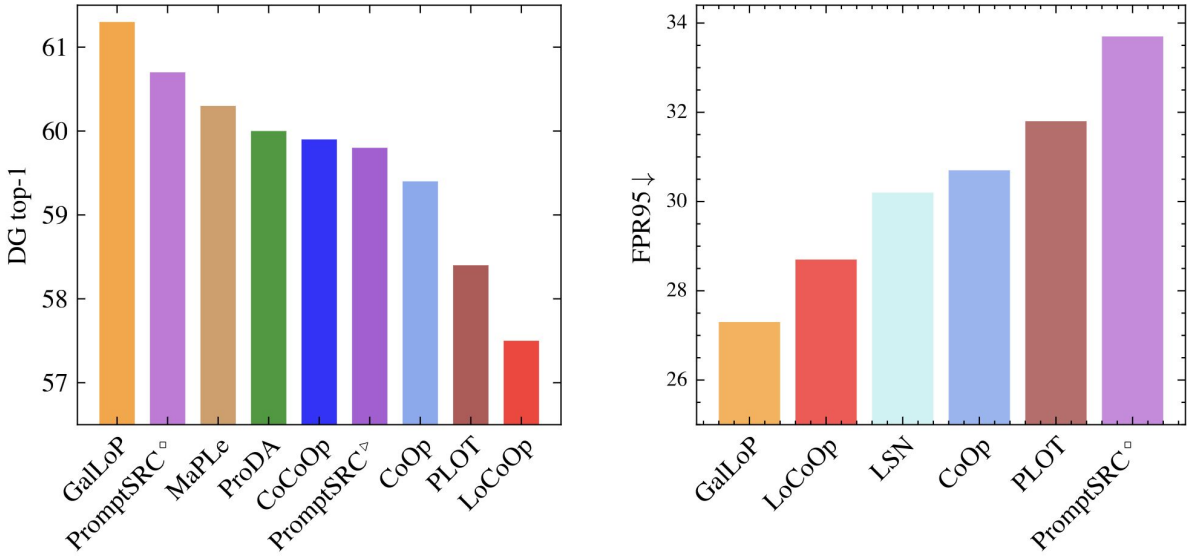
GalLoP





- Strong performances in **low shots** settings.
- Outperforms state-of-the-art prompt learning methods.

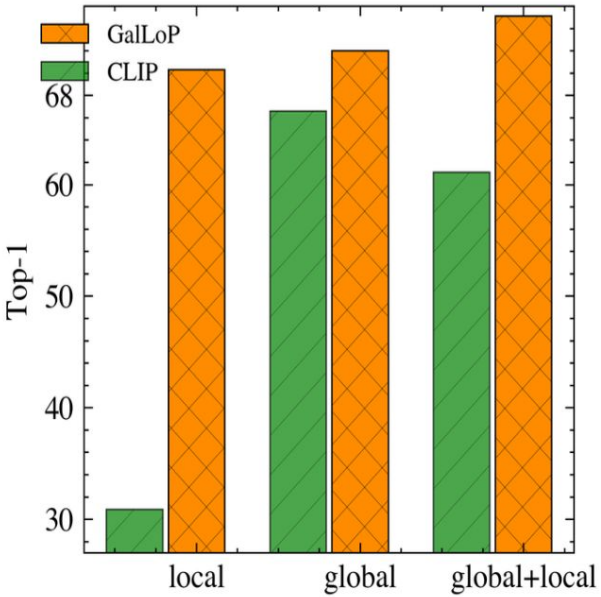
Robustness results.



(a) Domain generalization. **(b)** OOD detection.

- Strong domain generalization results (top-1).
- Outperforms dedicated OOD detection methods.

Combining global and local features



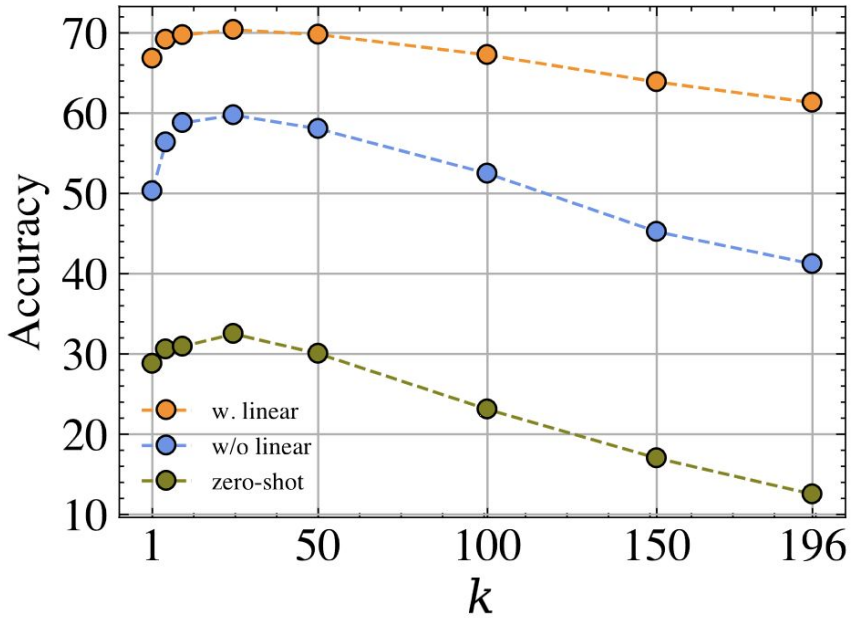
(c) Global-local

- Vanilla local features → no improvement.
- **Complementarity** of global / local features in GalLop.

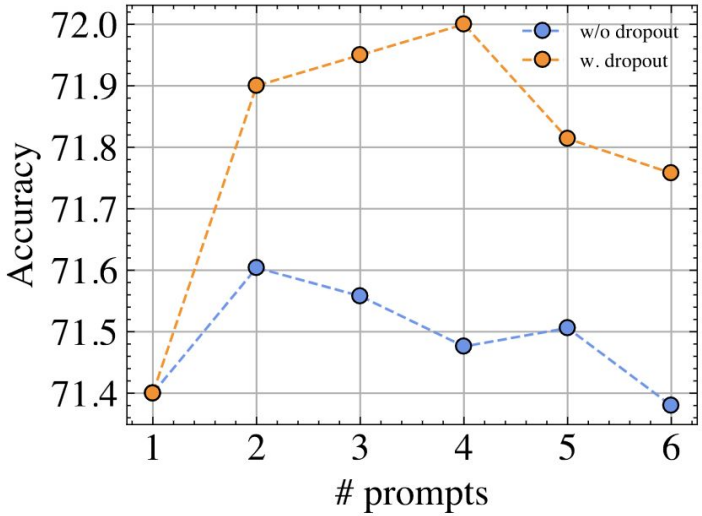
The need for sparsity.

GalLoP

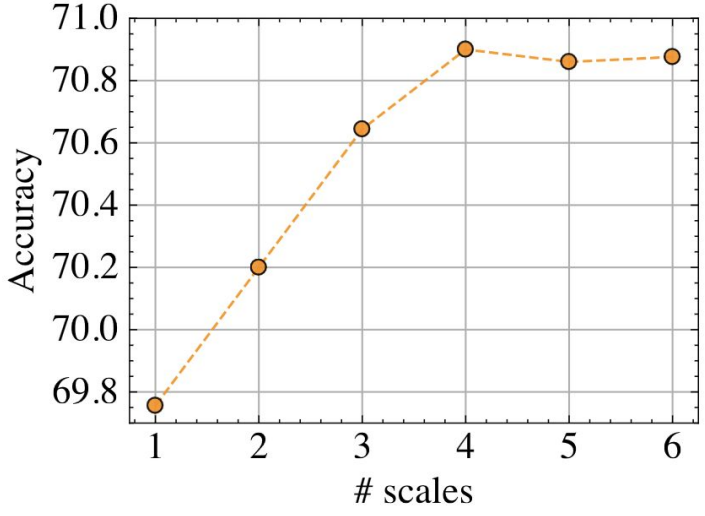
- Prompt learning → boosts performances.
- Sparsity → large gains in three regimes.
- Linear projection → better alignment.



Learning multiple prompts.



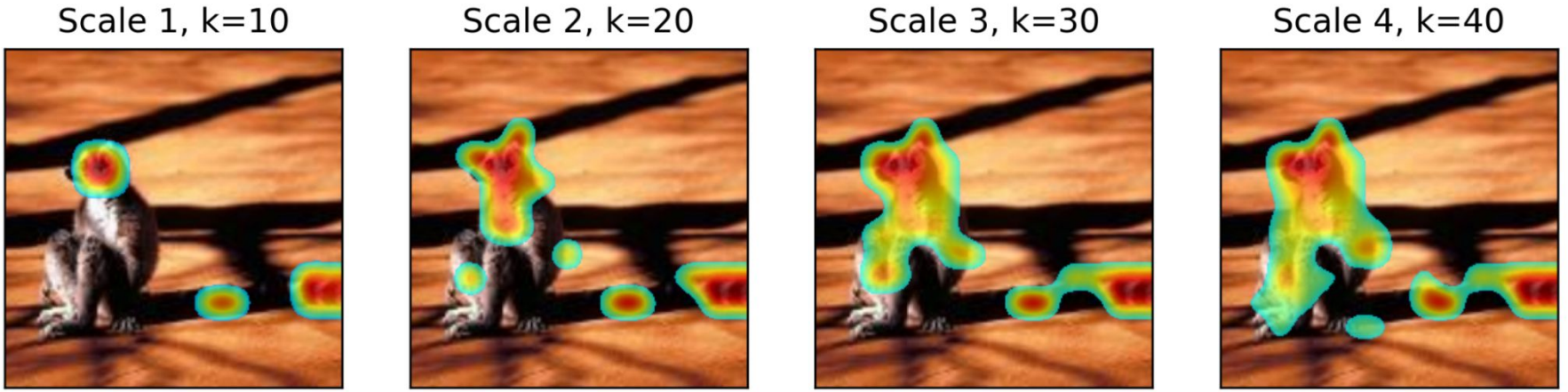
(a) Impact of prompt dropout when learning multiple global prompts.



(b) Impact of multiple scales when learning local prompts, with $k_1 = 10$, $\Delta_k = 10$.

- Prompt dropout → gains when using multiple global prompts.
- Multiscale loss → boosts local performances.

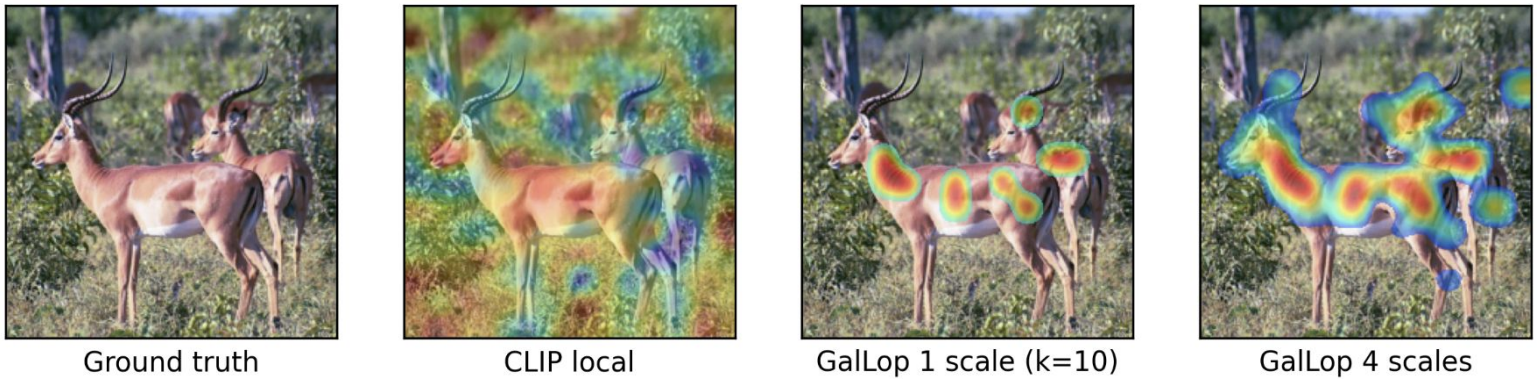
Qualitative results.



- Localization of objects.
- Few shot and weakly supervised segmentation.

Conclusion on GalLoP.

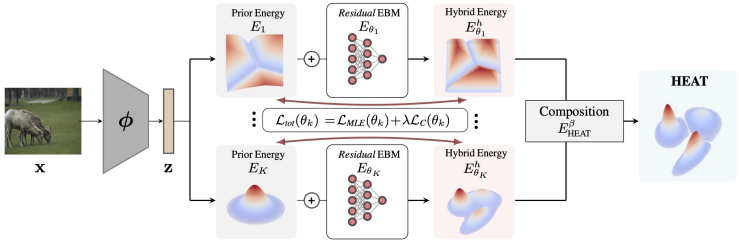
- Exploit local Features: sparsity & learnable projection.
- Global + local prompt learning → strong top-1 + robustness.



Short term extensions.

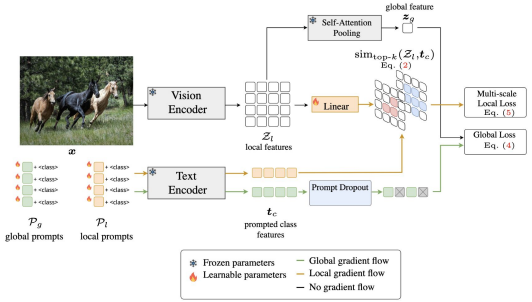
🔥HEAT:

- Extend other scorers, e.g. kNN.
- Apply HEAT to dense prediction.
- Extension to other modalities, e.g. NLP, audio.



🐎GalLoP:

- Integrate an adaptive sparsity ratio.
- Learn the linear projection on a bigger vision-language dataset.

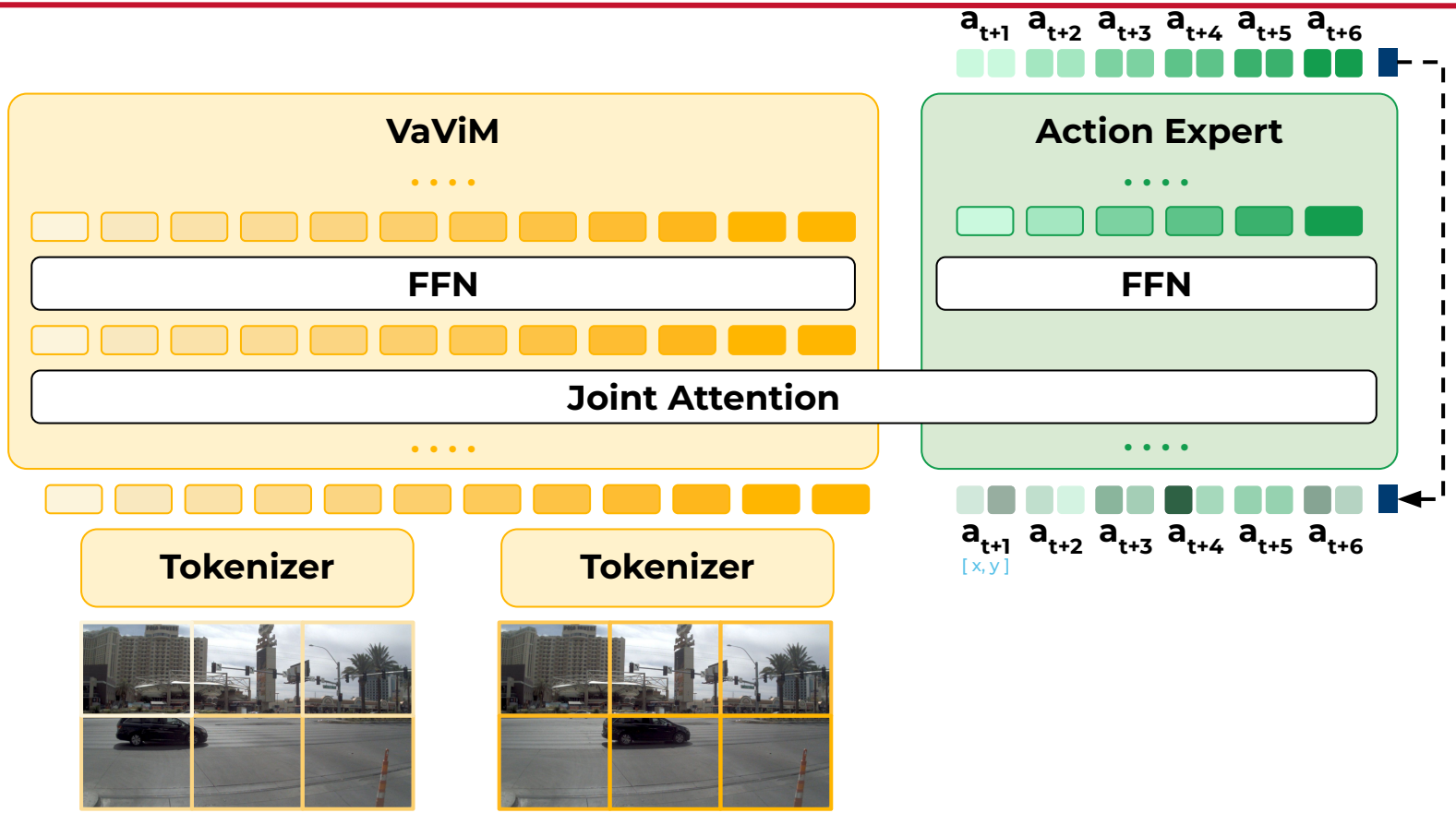


Current research interests.

le cnam
cedric EA4629

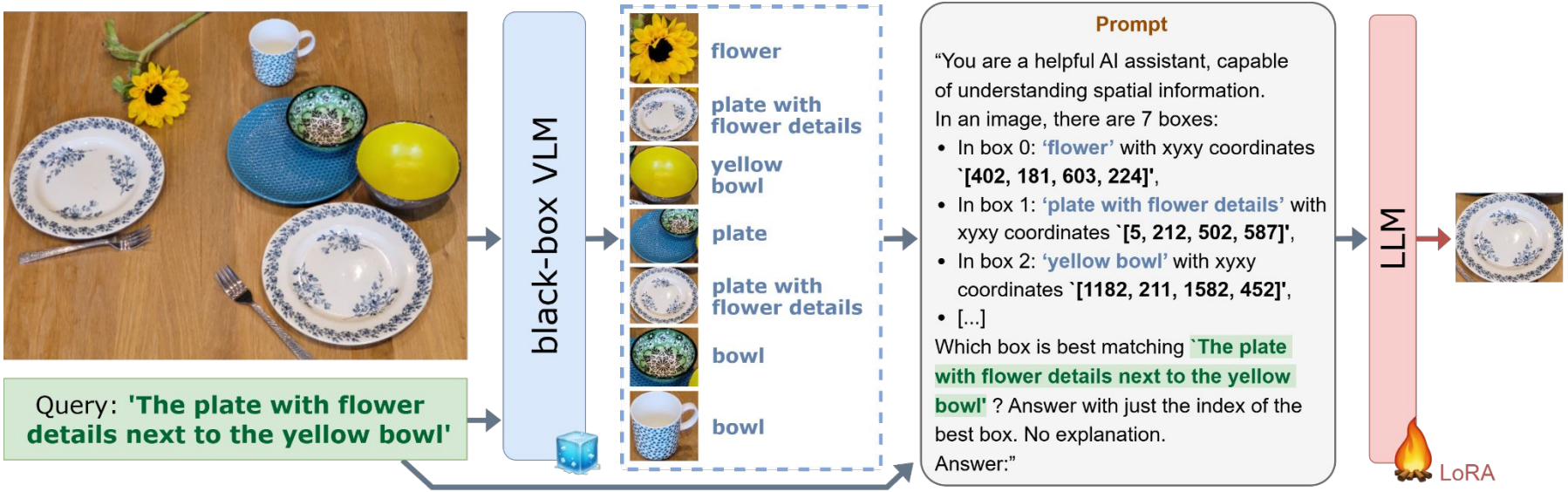
COEXYA
Connect skills, create more
valeo.ai

World model for autonomous driving



📕 Bartoccioni, Florent, et al. “VaViM and VaVAM: Autonomous Driving through Video Generative Modeling.” *ArXiv*, 2025.

Vision langue models



📕 Cardiel, Amaia, et al. "LLM-wrapper: Black-Box Semantic-Aware Adaptation of Vision-Language Models for Referring Expression Comprehension." ICLR, 2025.

Thank you
for your attention.

Elias Ramzi, et al. “Robust and Decomposable Average Precision for Image Retrieval.” *NeurIPS*, 2021 & *RFIAP*, 2022.
https://github.com/elias-ramzi/ROADMAP

EliasRamzi, et al. “Hierarchical Average Precision Training for Pertinent Image Retrieval.” *ECCV* ,2022.
https://github.com/elias-ramzi/HAPPIER

Marc Lafon, et al. “Hybrid Energy Based Model in the Feature Space for Out-of-Distribution Detection.” *ICML,* 2023.
https://github.com/MarcLafon/heatood

Elias Ramzi, et al. “Optimization of Rank Losses for Image Retrieval.” *TPAMI,* 2025.
https://github.com/cvdfoundation/google-landmark

Marc Lafon, et al. “GalLoP: Learning global and local prompts for vision-language models”, *ECCV* , 2024
https://github.com/MarcLafon/gallop

le cnam
cedric EA4629
COEXYA
Connect skills, create more
valeo.ai