Frederik Warburg
Javier Civera
Martin Jørgensen
Søren Hauberg

Bayesian Triplet Loss: Uncertainty Quantification in Image Retrieval

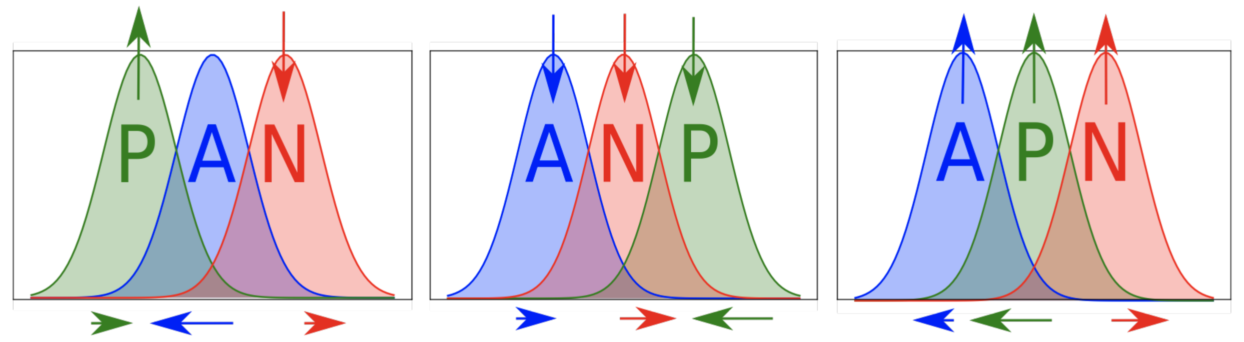
2021 ICCV OCTOBER 11-17 VIRTUAL

Overview

Uncertainty quantification in image retrieval is crucial for downstream decisions. In this work, we present the Bayesian Triplet Loss, which is the probabilistic equivalent of the triplet loss and produces state-of-the-art uncertainty quantification for image retrieval.

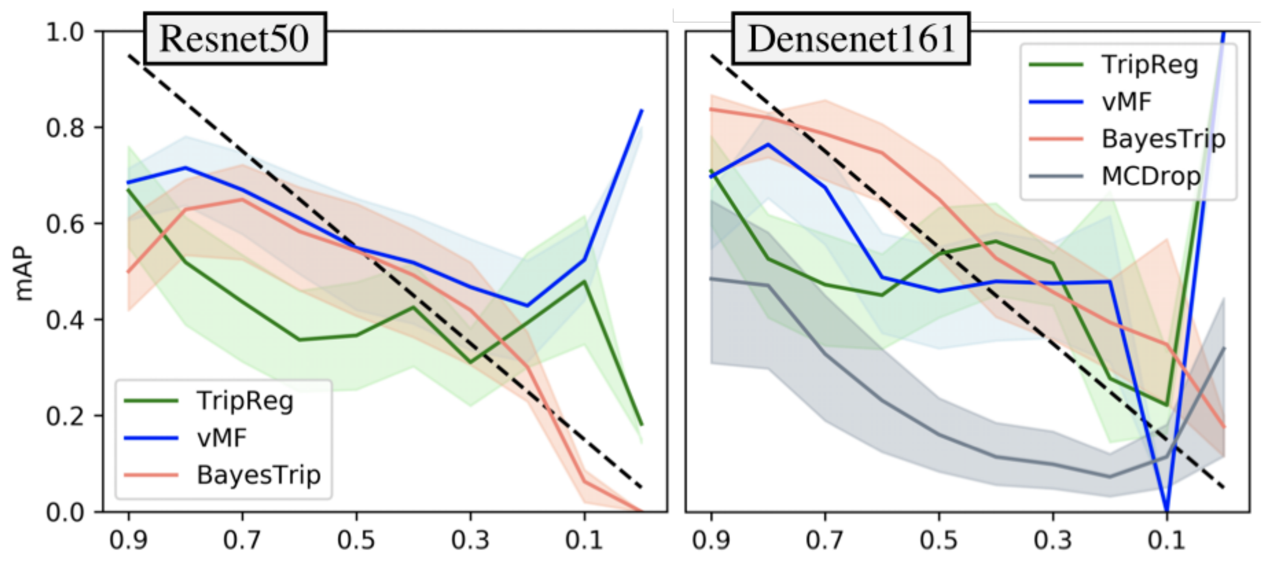
- Learned uncertainty estimates in latent space
- A likelihood function on the latent space that follows the intuition of the triplet loss
- Comparable retrieval performance to SOTA
- SOTA uncertainty quantification in image retrieval

Intuition



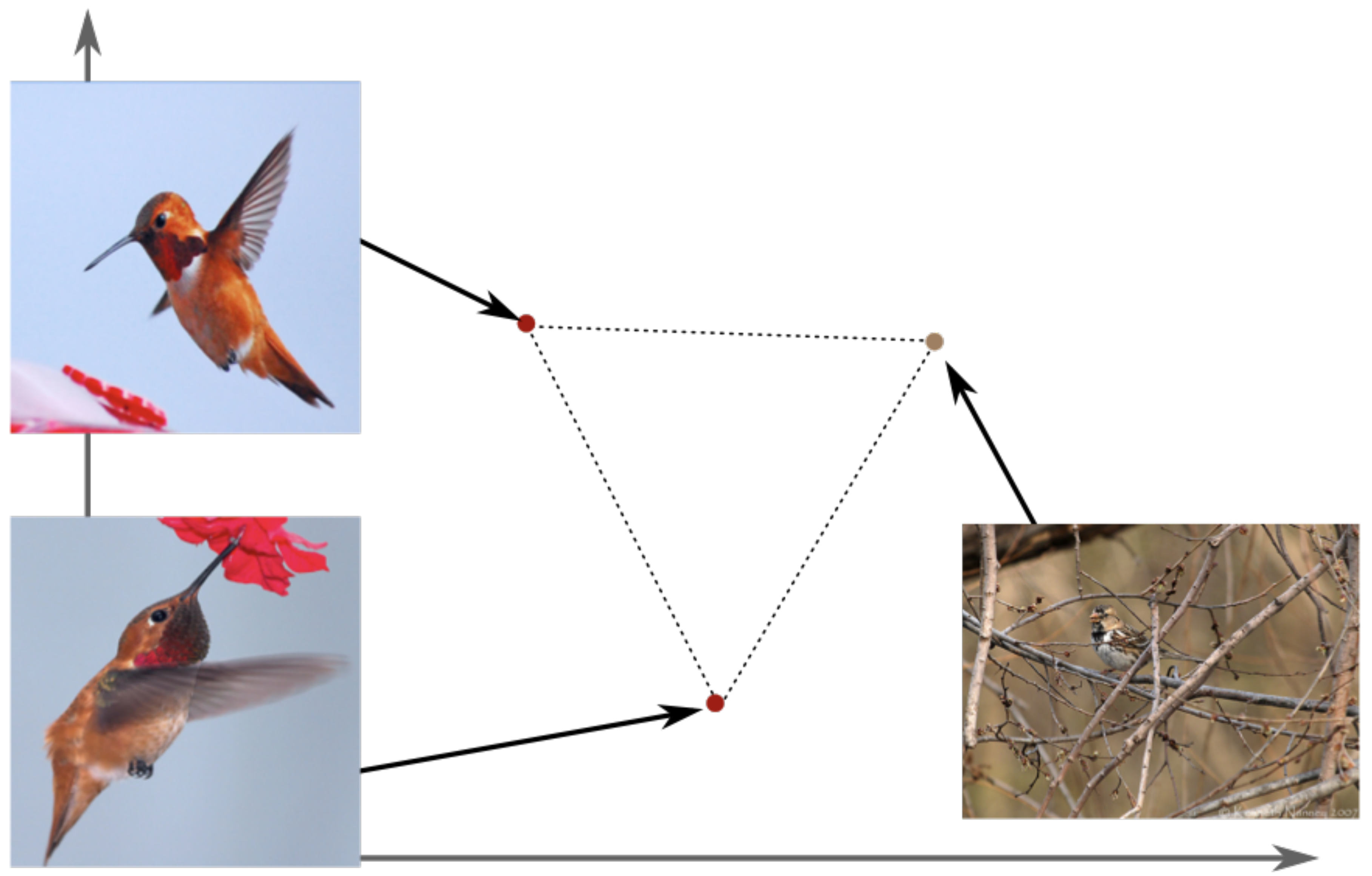
The arrows below the figures indicate the gradient direction and magnitude of the means, while the arrows above the distributions indicate the gradients of the variances (downwards indicate more spread, upwards means more peaky)

Calibration Plot

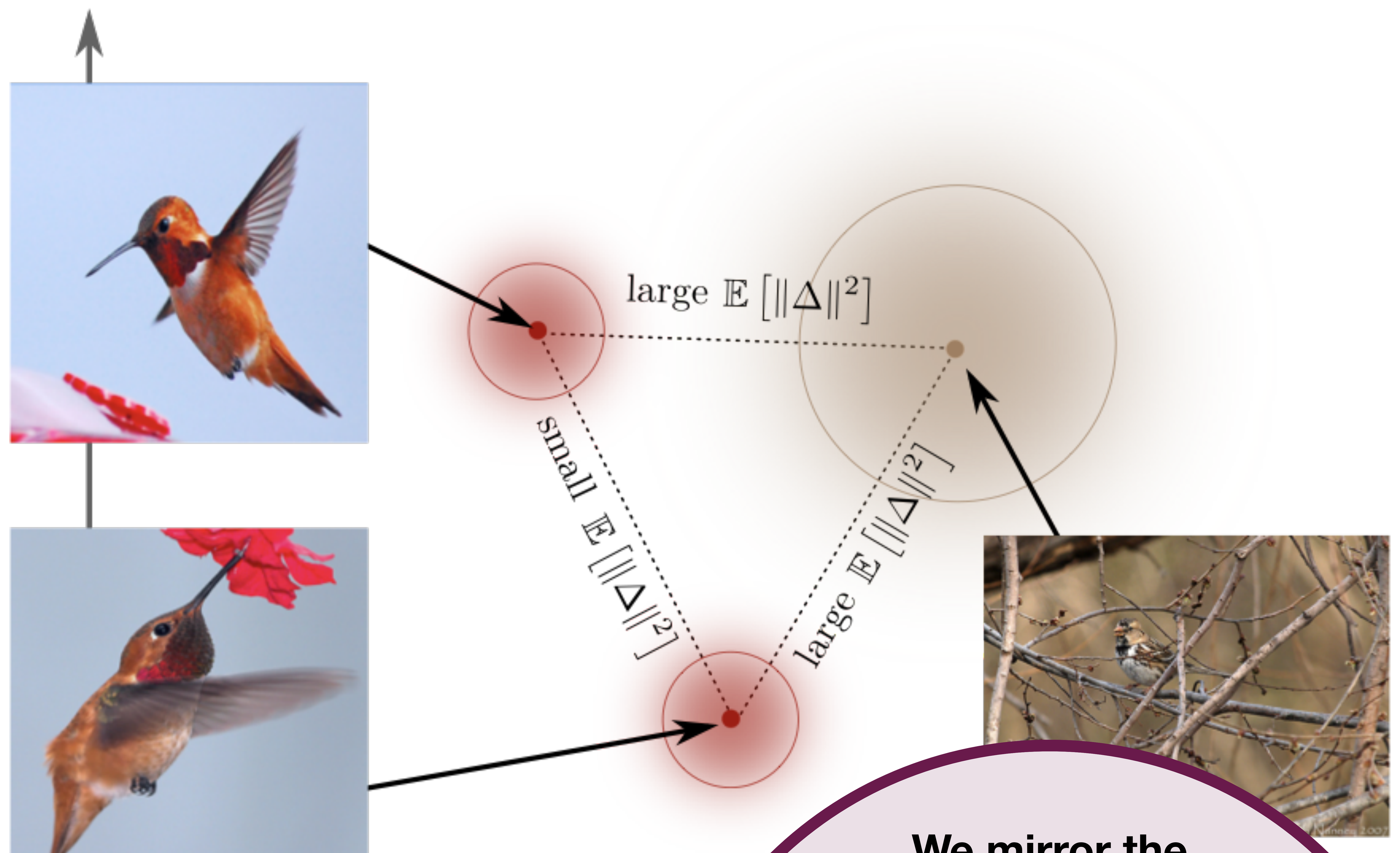


The solid line indicates mAP@5 and the shaded area covers from mAP@1 to mAP@10. Note how, for both backbones, the Gaussian distributed embeddings are better calibrated, especially for uncertain queries.

Triplet Loss



Bayesian Triplet Loss



We mirror the triplet loss, and learn stochastic features. We propose a likelihood that matches the triplet constraint and evaluates the probability of an anchor being closer to a positive than a negative.

$$P(\|a-p\|^2 + m < \|a-n\|^2)$$

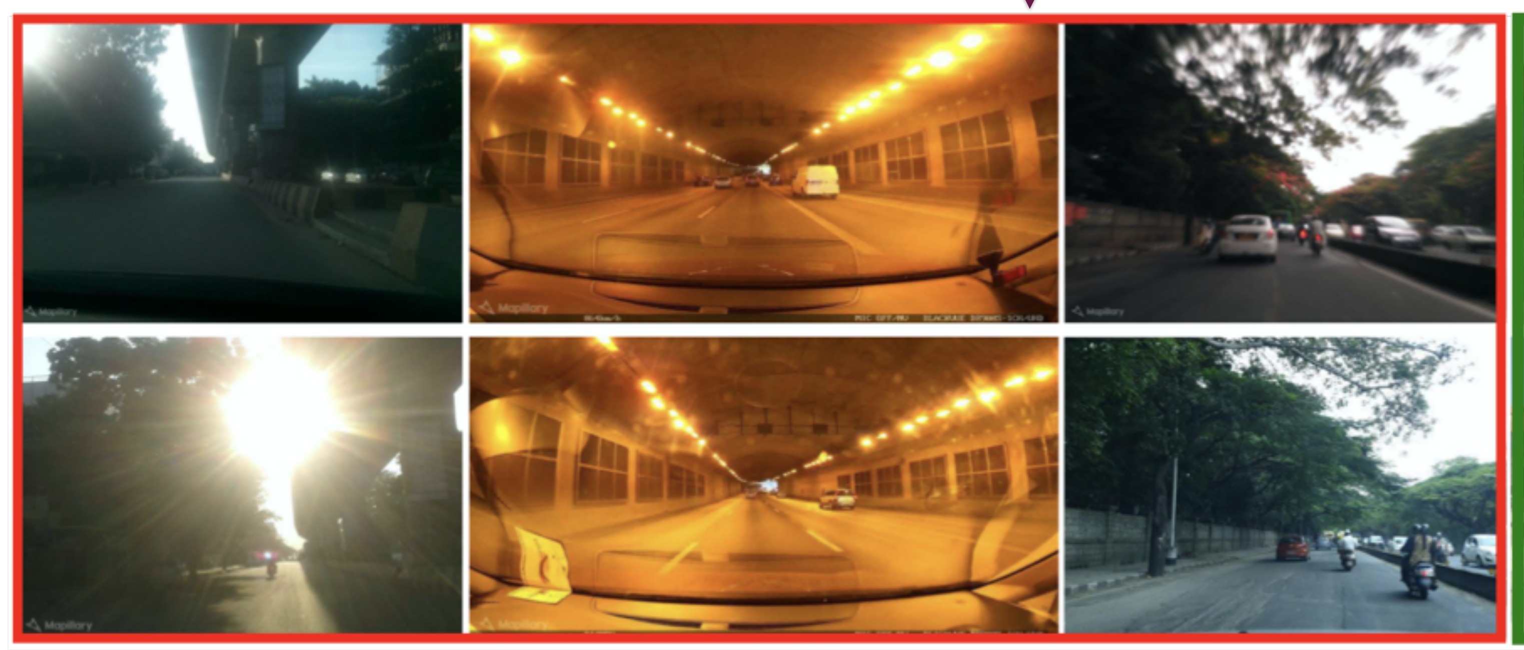
State of the art uncertainty quantification

Low Confidence

High Confidence

Our method provides SOTA uncertainty estimates, while matching the predictive performance of the triplet loss. It associates high uncertainty to hard examples, like birds that blend in with the background and are hardly discernible

We show on the large scale place recognition dataset, MSLS, that our model gives high uncertainty to challenging places w. harsh sunlight, blur and ambiguous tunnels.



Low Confidence

High Confidence

Likelihood Derivation

STEP 1: (define likelihood for all triplets)

$$\mathcal{L}(\Omega) = \prod_{X\in\Omega}\prod_{Y\in\Omega}\prod_{Z\in\Omega} P(I(X,Y,Z) \mid X,Y,Z)$$

STEP 2: (define multimodal model)

$$P(I(X,Y,Z) \mid X,Y,Z) = p^{\mathbb{1}\{I(X,Y,Z)=2\}}$$

STEP 3: (define p using triplet notation)

$$P(\|a-p\|^2 + m < \|a-n\|^2)$$

STEP 4: (assume independent dimensions)

$$P(\|a-p\|^2 - \|a-n\|^2 < -m) = P(\tau < -m),$$
$$\text{where } \tau = \sum_{d=1}^{D}(a_d-p_d)^2 - (a_d-n_d)^2,$$

STEP 5: (law of large numbers - surprisingly good approximation; see paper)

$$\lim_{D\to\infty} P\left(\frac{\tau-\mu}{\sigma} < -m\right) = \Phi(-m),$$

STEP 6: (find mean and variance)

$$\mathbb{E}[\tau] = \mu_p^2 + \sigma_p^2 - \mu_n^2 - \sigma_n^2 - 2\mu_a(\mu_p - \mu_n).$$

$$\begin{aligned}\frac{\mathrm{Var}(\tau)}{2} &= \sigma_p^2(\sigma_p^2 + 2\mu_p^2) + \sigma_n^2(\sigma_n^2 + 2\mu_n^2) - 4\sigma_a^2\mu_p\mu_n \\ &- 2\mu_a\left(\mu_a(\mu_p^2+\mu_n^2) - 2\mu_p\sigma_p^2 - 2\mu_n\sigma_n^2\right) \\ &+ 2(\sigma_a^2+\mu_a^2)\left((\sigma_p^2+\mu_p^2) + (\sigma_n^2+\mu_n^2)\right).\end{aligned}$$

STEP 7: (learn mean and variance for a, p, n)