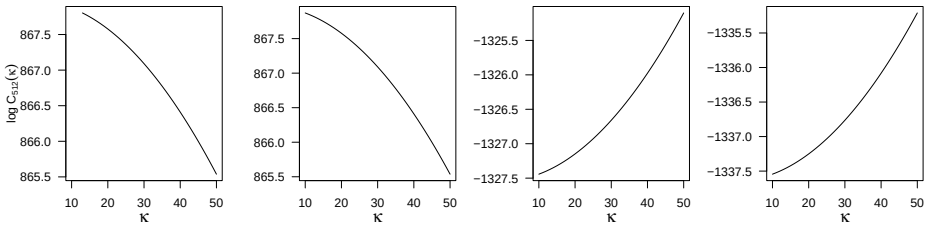
Supplementary Material: A Non-isotropic Probabilistic Take on Proxy-based Deep Metric Learning

Michael Kirchhof[1,*], Karsten Roth[1,*], Zeynep Akata[1], and Enkelejda Kasneci[1]

[1]University of Tübingen, Germany. (*) equal contribution

A Approximation of the von Mises-Fisher Distribution's Normalizing Constant



(a) Exact values (b) Our approx. (c) Approx. of [3] (d) Approx. of [9]

Fig. 8: Comparison of approximations and exact values of the logarithmized normalization constant of the vMF distribution $\log C_M(\kappa)$ for $M = 512$ dimensions.

As we aim to resolve sample-specific ambiguities captured by κ_z, we need to calculate the logarithmic normalizing constant of the vMF distribution:

$$\log C_M(\kappa) = \log \frac{\kappa^{M/2-1}}{(2\pi)^{M/2} I_{M/2-1}(\kappa)}, \tag{11}$$

where I_d is the modified Bessel function of first kind at order d and M is the dimensionality of the embedding space. However, I_d is expensive to compute and impossible to backpropagate through in high dimensions since it has no closed form. Hence, it is commonly approximated in the literature. [3] and [9] for example utilize approximations from lower and upper bounds which are shown in Figure 8c and 8d for $M = 512$. However, if we calculate $\log C_M$ from the exact Bessel functions implemented in `R 4.1.1`'s base package [7], we see in Figure 8a that $\log C_M$ is monotonically decreasing, because I_d is monotonically increasing with κ [5, Section 10.37].

To account for this issue, we thus choose to derive an approximation by directly fitting a quadratic model to the exact Bessel function for $M \in \{128, 512\}$ with $\kappa \in$

$\{10, \dots, 50\}$. The resulting approximations are

$$\log C_{128}(\kappa) \approx 127 - 0.01909 \cdot \kappa - 0.003355 \cdot \kappa^2 \text{ and} \tag{12}$$

$$\log C_{512}(\kappa) \approx 868 - 0.0002662 \cdot \kappa - 0.0009685 \cdot \kappa^2. \tag{13}$$

The mean squared error of these approximations to the ground truth values is smaller than 0.1%, which is visually confirmed in Figure 8b. During experimentation, we found that the model is insensitive to perturbations in the precise coefficients. Also, we found that a linear model is too simple and an exponential model imposed very high gradients and inverts the behaviour of the metrics when κ is high. Hence, we decided for the quadratic approximation as the simplest yet well extrapolating function. As a reference for future work, we note that [2] recently gave an additional approximation implemented in PyTorch.

B Derivation of the Non-isotropic von Mises-Fisher Distribution

The nivMF can be motivated by a transformed vMF distribution, which we assume to be parametrized by $\mu \in \mathcal{S}^{M-1}$ and $K = \text{diag}(\kappa) \in \mathbb{R}_{>0}^{(M \times M)}$, $\kappa \in \mathbb{R}_{>0}^{M}$. Transforming our parameters into $\tilde{\mu} = \frac{K\mu}{\|K\mu\|}$ and $\tilde{\kappa} = \|K\mu\|$, we can define an ordinary vMF distribution $\tilde{X} \sim \text{vMF}(\tilde{\mu}, \tilde{\kappa})$ with density

$$f_{\tilde{X}}(\tilde{x}) = C_M(\tilde{\kappa}) \, \exp\left(\tilde{\kappa} \tilde{x}^\top \tilde{\mu}\right) \ . \tag{14}$$

For ease of notation, we do not include the subscript p to denote specific proxies. Now, we substitute $\tilde{x} := g(x) = \frac{Kx}{\|Kx\|}$. Note that g is bijective as a function $g : \mathcal{S}^{M-1} \to \mathcal{S}^{M-1}$, but non-bijective when seen as a function $g : \mathbb{R}^M \to \mathbb{R}^M$, since it would lose a degree of freedom due to normalization. We will still treat it as the latter and ignore the non-bijectivity, such that the following should be seen as motivation and not proof, and comment on the implications further below. We now seek the density of $X = g^{-1}(\tilde{X})$. The change-of-variable theorem gives

$$f_X(x) = f_{\tilde{X}}(\tilde{x}) |\det \frac{\partial g(x)}{\partial x}| \, . \tag{15}$$

By Equation 130 given in [6] and the chain rule, we obtain

$$\frac{\partial g(x)}{\partial x} = \left(\frac{1}{\|Kx\|} I_m - \frac{K^\top x x^\top K}{\|Kx\|^3} \right) K^\top \tag{16}$$

$$= \left(\frac{1}{\tilde{\kappa}} I_M - \frac{(\tilde{\kappa}\tilde{\mu})(\tilde{\kappa}\tilde{\mu})^\top}{\tilde{\kappa}^3} \right) K^\top \tag{17}$$

$$= \frac{1}{\tilde{\kappa}} \left(I_M - \tilde{\mu}\tilde{\mu}^\top \right) K^\top \ . \tag{18}$$

Since the first part of this matrix is a projection on the orthogonal complement of $\tilde{\mu}$, the matrix has rank $M - 1$ and the determinant becomes zero. This is a consequence of

the broken bijectivity assumption from above. However, we can see that Equation 18 essentially projects K on the tangential plane of $\tilde{\mu}$. By taking its determinant, we measure the volume of the remaining $(M-1)$-dimensional concentration sphere. Performing a singular value decomposition on Equation 18 reveals that μ is the eigenvector with eigenvalue 0. So, if we substract the contribution of μ to the volume of K, which is $\|K\mu\| = \tilde{\kappa}$, we obtain

$$D(K) = \frac{\prod_{m=1}^{M} \kappa_m}{\tilde{\kappa}} \,. \tag{19}$$

When we plug this heuristic into Equation 15, we arrive at the nivMF density:

$$f_X(x) = C_M(\tilde{\kappa}) \, \exp\left(\tilde{\kappa}\tilde{x}^\top \tilde{\mu}\right) D(K) \tag{20}$$

$$= C_M(\|K\mu\|) \, D(K) \, \exp\left(\|K\mu\| \left(\frac{Kx}{\|Kx\|}\right)^\top \frac{K\mu}{\|K\mu\|}\right) \tag{21}$$

$$= C_M(\|K\mu\|) \, D(K) \, \exp\left(\|K\mu\| \, s(Kx, K\mu)\right) \,. \tag{22}$$

We stress that $D(K)$ is a heuristic choice, such that the proposed nivMF density strictly speaking yields only a measure and not necessarily a probability measure. An analytical solution is promising material for future work. It may also enable the density of the nivMF to become a true expansion of the vMF density, i.e., $D(K)$ may vanish when $K = \kappa I_M$ for $\kappa > 0$, which is currently not the case. In empirical tests, dropping $D(K)$ lead to a considerably severed performance.

C Further distribution-to-distribution Metrics

We can define further distribution-to-distribution metrics beyond $d_{\text{EL-nivMF}}$. One starting point are probability product kernels (PPK) [1]. They are a family of metrics to compare two distributions ρ and ζ by the product of their densities:

$$\text{PPK}_\gamma(\rho, \zeta) = \int_{\mathcal{E}} \rho(a)^\gamma \zeta(a)^\gamma da, \text{ with } \gamma > 0. \tag{23}$$

Since the loss in Equation 2 takes the exponential of the distance metrics, we take their logarithms here to retain the PPK as actual score in nominator and denominator. In particular, if we assume a vMF distribution for both ρ and ζ

$$d_{\text{B-vMF}}(\rho, \zeta) := -\log(\text{PPK}_{0.5}(\rho, \zeta)) \tag{24}$$

gives the Bhattacharyya distance and

$$d_{\text{EL-vMF}}(\rho, \zeta) := -\log(\text{PPK}_{1}(\rho, \zeta)) \tag{25}$$

gives the expected likelihood distance, also known as mutual likelihood score [10]. Their analytical solutions are provided in Supp. D.

The previous metrics are symmetric in ρ and ζ. To capture the inherent asymmetry between samples and proxies, we also study the Kullback-Leibler divergence $d_{\text{KL-vMF}}(\rho, \zeta) := \text{KL}(\zeta||\rho)$. Its analytical solution if both ρ and ζ are vMF densities is given in Supp. E.

D Analytical Solutions of Bhattacharyya and Expected Likelihood Distance

Let ζ and ρ be densities of two vMF-distributed random variables with parameters $\nu_z = \kappa_z \mu_z$ and $\nu_p = \kappa_p \mu_p$, respectively.

Bhattacharyya distance. Since the vMF is a member of the exponential family, [1] gives us that

$$\text{PPK}_{0.5}(\rho, \zeta) = \exp(K(\nu_z/2 + \nu_p/2) - K(\nu_z)/2 - K(\nu_p)/2), \text{ with} \tag{26}$$

$$K(\nu) = -\log C_M(\|\nu\|)\,. \tag{27}$$

Thus,

$$d_{\text{B-vMF}}(\rho, \zeta) = -\log(\text{PPK}_{0.5}(\rho, \zeta)) \tag{28}$$

$$= \log C_M(\|\nu_z + \nu_p\|/2) - \log C_M(\nu_z)/2 - \log C_M(\nu_p)/2\,. \tag{29}$$

Expected likelihood distance. We can extend

$$\text{PPK}_1(\rho, \zeta) = \int_{\mathcal{E}} \zeta(\tilde{z})\rho(\tilde{z})d\tilde{z} \tag{30}$$

$$= C_M(\kappa_z) \cdot C_M(\kappa_p) \int_{\mathcal{E}} \exp((\kappa_z \mu_z + \kappa_p \mu_p)^\top \tilde{z})d\tilde{z} \tag{31}$$

$$= \frac{C_M(\kappa_z) \cdot C_M(\kappa_p)}{C_M(\|\nu_0\|)} \int_{\mathcal{E}} C_M(\|\nu_0\|) \exp(\nu_0^\top \tilde{z})d\tilde{z}, \text{ with} \tag{32}$$

$$\nu_0 := \kappa_z \mu_z + \kappa_p \mu_p, \tag{33}$$

such that the latter is again the density of a vMF distributed random variable, whose integral over the embedding space is 1. Then,

$$d_{\text{EL-vMF}}(\rho, \zeta) = -\log(\text{PPK}_1(\rho, \zeta)) \tag{34}$$

$$= \log C_M(\|\nu_z + \nu_p\|) - \log C_M(\nu_z) - \log C_M(\nu_p)\,. \tag{35}$$

Note that both $d_{\text{EL-vMF}}$ and $d_{\text{B-vMF}}$ depend on $\|\nu_z + \nu_p\|$ which implicitely respects the cosine similarity between μ_z and μ_p, but also processes κ_z and κ_p.

E Analytical Solution of KL-Divergence

Let ζ and ρ be densities of two vMF-distributed random variables with parameters μ_z, κ_z and μ_p, κ_p, respectively. Then

$$KL(\zeta||\rho) = \int_{\mathcal{E}} \zeta(\tilde{z}) \log \frac{\zeta(\tilde{z})}{\rho(\tilde{z})} d\tilde{z} \tag{36}$$

$$= \int_{\mathcal{E}} \log C_M(\kappa_z) - \log C_M(\kappa_p) + (\kappa_z \mu_z^\top - \kappa_p \mu_p^\top)\tilde{z} d\zeta(\tilde{z}) \tag{37}$$

$$= \log C_M(\kappa_z) - \log C_M(\kappa_p) + (\kappa_z \mu_z^\top - \kappa_p \mu_p^\top) \int_{\mathcal{E}} \tilde{z} d\zeta(\tilde{z}) \tag{38}$$

$$= \log C_M(\kappa_z) - \log C_M(\kappa_p) + (\kappa_z \mu_z^\top - \kappa_p \mu_p^\top)\mu_z \tag{39}$$

F Gradients of d_{L2} and d_{Cos}

We are interested in differentiating the loss $\mathcal{L}_{\text{NCA++}}$ from Equation 1 in §3.2 by the cosine similarity between the image z and a proxy of interest p. Let p^* denote the ground-truth proxy of z and $\frac{\delta}{\delta s} := \frac{\delta}{\delta s(\mu_p, \mu_z)}$. Then,

$$\frac{\delta}{\delta s}\mathcal{L}_{\text{NCA++}} = \begin{cases} \frac{\delta}{\delta s} d(\rho^*, \zeta)/t + \frac{\delta}{\delta s} \log(\sum_{c=1}^{C} \exp(-d(\rho_c, \zeta)/t)) & , \text{if } p = p^* \\ \frac{\delta}{\delta s} \log(\sum_{c=1}^{C} \exp(-d(\rho_c, \zeta)/t)) & , \text{else} \end{cases} \tag{40}$$

and by the chain rule we get

$$\frac{\delta}{\delta s} \log\left(\sum_{c=1}^{C} \exp(-d(\rho_c, \zeta)/t)\right) = -\frac{\exp(-d(\rho, \zeta)/t)}{\sum_{c=1}^{C} \exp(-d(\rho_c, \zeta)/t)} \frac{\delta}{\delta s} d(\rho_c, \zeta)/t \ . \tag{41}$$

Let's consider the $\mathcal{L}^{\text{Cos}}_{\text{NCA++}}$ loss, i.e., $d(\rho, \zeta) = -s(\mu_p, \mu_z)$. We can plug $\frac{\delta}{\delta s} d(\rho, \zeta) = -1$ into Equations 40 and 41 and obtain:

$$\frac{\delta}{\delta s}\mathcal{L}^{\text{Cos}}_{\text{NCA++}} = \begin{cases} \frac{1}{t}\left(-1 + \frac{\exp(-d(\rho,\zeta)/t)}{\sum_{c=1}^{C} \exp(-d(\rho_c,\zeta)/t)}\right) & , \text{if } p = p^* \\ \frac{1}{t} \frac{\exp(-d(\rho,\zeta)/t)}{\sum_{c=1}^{C} \exp(-d(\rho_c,\zeta)/t)} & , \text{else} \end{cases} \tag{42}$$

$$= \begin{cases} \frac{1}{t}\left(-1 + \frac{\exp(s(\mu_p,\mu_z)/t)}{\sum_{c=1}^{C} \exp(s(\mu_{p_c},\mu_z)/t)}\right) & , \text{if } p = p^* \\ \frac{1}{t} \frac{\exp(s(\mu_p,\mu_z)/t)}{\sum_{c=1}^{C} \exp(s(\mu_{p_c},\mu_z)/t)} & , \text{else} \end{cases} \ . \tag{43}$$

Now, consider $\mathcal{L}^{\text{L2}}_{\text{NCA++}}$, i.e., $d(\rho, \zeta) = \|\nu_p - \nu_z\|^2 = \kappa_p^2 + \kappa_z^2 - 2\kappa_p\kappa_z s(\mu_p, \mu_z)$, following from the law of cosines. Here, $\frac{\delta}{\delta s} d(\nu_p, \nu_z) = -2\kappa_p\kappa_z$, which we can again plug into Equations 40 and 41 and obtain:

$$\frac{\delta}{\delta s}\mathcal{L}^{\text{L2}}_{\text{NCA++}} = \begin{cases} -\frac{2\kappa_p\kappa_z}{t} + \frac{2\kappa_p\kappa_z}{t} \frac{\exp(-d(\rho,\zeta)/t)}{\sum_{c=1}^{C} \exp(-d(\rho_c,\zeta)/t)} & , \text{if } p = p^* \\ \frac{2\kappa_p\kappa_z}{t} \frac{\exp(-d(\rho,\zeta)/t)}{\sum_{c=1}^{C} \exp(-d(\rho_c,\zeta)/t)} & , \text{else} \end{cases} \tag{44}$$

$$= \begin{cases} -\frac{2\kappa_p\kappa_z}{t} + \frac{2\kappa_p\kappa_z}{t} \frac{\exp((\kappa_p^2 + 2\kappa_p\kappa_z s(\mu_p,\mu_z))/t)}{\sum_{c=1}^{C} \exp((\kappa_{p_c}^2 + 2\kappa_p\kappa_z s(\mu_{p_c},\mu_z))/t)} & , \text{if } p = p^* \\ \frac{2\kappa_p\kappa_z}{t} \frac{\exp((\kappa_p^2 + 2\kappa_p\kappa_z s(\mu_p,\mu_z))/t)}{\sum_{c=1}^{C} \exp((\kappa_{p_c}^2 + 2\kappa_{p_c}\kappa_z s(\mu_{p_c},\mu_z))/t)} & , \text{else} \end{cases} \ . \tag{45}$$

G Summary of Loss Calculation

Algorithm 1 sketches how **EL-nivMF** is implemented practically. As discussed, the parameters of the proxies are learnable parameters, whereas the vMF distributions of points are predicted by an encoder. Thus, the module in Algorithm 1 can be plugged on-top of an encoder and trained jointly. Since test-time retrieval only requires access to the image-embeddings, the module can be discarded after training.

Algorithm 1: Module to compute **EL-nivMF** loss

Function `initialize` (C: *num proxies,* M: *dimensions,* N: *num samples*) **:**
  $\mu_\rho \leftarrow$ learnable tensor $\in [C, M]$
  $\kappa_\rho \leftarrow$ learnable tensor $\in [C, M]$
  $t \leftarrow$ learnable parameter $\in [1]$
  Save C, M, N
Function `loss` (z: *image embedding* $\in [1, M]$, c^*: *ground-truth proxy index*) **:**
  `samples` $\leftarrow$ empty matrix $\in [N, D]$
  for $n = 1, \dots, N$ **do**
    `samples`$[n, :] \sim \text{vMF}\left(\mu = \frac{z}{\|z\|}, \kappa = \|z\|\right)$
  end
  `sim_to_proxy` $\leftarrow$ empty vector $\in [C]$
  for $c = 1, \dots, C$ **do**
    `logls` $\leftarrow$ empty vector $\in [N]$
    for $n = 1, \dots, N$ **do**
      `logls`$[n]$
      $= \log(\texttt{nivmf\_likelihood}(z, \mu = \mu_\rho[c, :], K = \text{diag}(\kappa_\rho[c, :]))$
    end
    `sim_to_proxy`$[c] \leftarrow$ `logsumexp`(`logls`$/t$)
  end
  `logloss` $\leftarrow -$`sim_to_proxy`$[c^*] +$ `logsumexp`(`sim_to_proxy`)
  return *logloss*

H Experimental Details

As already noted in §3.3, we generally utilize $N \approx 10$ for our Monte-Carlo estimation of the PPK kernel (Eq. 5), but switch to $N = 5$ for hyperparameter searches and $N = 20$ for our ablation experiments, as within this range, we found performance to be similar.

I Experimental Details Ablation Study

To reduce any influences of covariates, we seek to keep experimental settings in the ablation study in §4.3 constant across all benchmarked metrics. Hence, we fixed all hyperparameters as in the previous experiment, and tuned the following hyperparameters for each approach on validation data:

$$t \in \{1, 1/32, 1/256\} \tag{46}$$

$$\kappa_p \in \{10, 50, 200\} \text{ (for } \textbf{ni-vMF}\text{, this is for each dimension)}\,. \tag{47}$$

Across all metrics, we used the dimensionality $M = 512$, a batchsize of 106, and 150 epochs on CARS and 50 on CUB. To reduce the initialization noise, we initiated each hyperparameter-tuning experiment 3 times with random seeds, then calculated the median of the maximum $R@1$ performance on the validation set, and ran the best hyperparameter settings with 5 seeds.

J L_2 Distance as Retrieval Metric

Table 4: R@1 of the same trained models from Figure 5, but using the euclidean instead of the cosine distance for retrieval.

Method	CUB	CARS
d_{L2}	61.89 ± 0.36	76.61 ± 0.17
d_{Cos}	62.01 ± 0.35	76.94 ± 0.49
d_{nivMF}	63.74 ± 0.18	78.62 ± 0.41
$d_{\text{B-vMF}}$	62.29 ± 0.34	79.69 ± 0.15
$d_{\text{EL-vMF}}$	62.49 ± 0.56	80.17 ± 0.24
$d_{\text{KL-vMF}}$	61.68 ± 0.36	76.65 ± 0.20
$d_{\text{EL-nivMF}}$	63.69 ± 0.56	76.37 ± 5.32

K Qualitative Impact on Image Norms

To understand in more detail the difference in learned and assigned image norms produced when training with $d_{\text{EL-nivMF}}$, we compare the distribution of image norms between those belonging to originally correctly and incorrectly classified samples (initial separation done using a standard baseline DML model operating on d_{cos}) for CUB & CARS, respectively. Results are shown in Fig. 9, which reveal that correct classifications on average have higher norms while miss-classifications are more often attributed to lower norms. This aligns well with the underlying motivation assigning low norms to ambiguous images (compare to e.g. Sec.4.4).

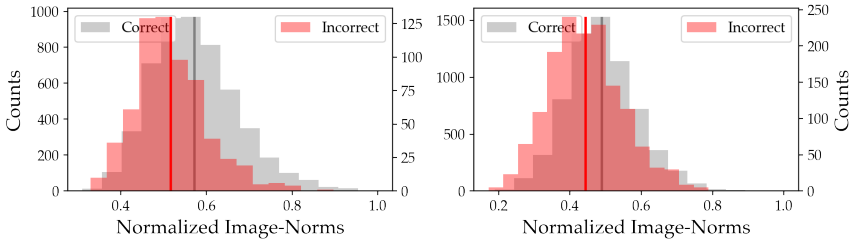


Fig. 9: Norms of prev. correct/incorrect pred. on CUB/CARS.

L Non-isotropic Proxies Encourage Diverse Representations

Finally, we qualitatively investigate the metric representation spanned by metric learners trained using $d_{\text{EL-nivMF}}$. To do so, we follow both [8] and look at the feature diversity, as well as evaluating the cluster diversity to see whether encouraging unique class-proxy distributions helps in learning a more diverse class-specific encoding. For the former, we follow [8] and evaluate the uniformity of the sorted spectral value distribution of all training image embeddings to measure the number of significant directions of variances in feature space. The latter is simply computed as the variance (i.e. diversity) of intraclass distances for each class-cluster. For both cases, we specifically care about relative changes compared to models trained without probabilistic treatment (i.e. using $d_{\cos}$) as well as changes going from an isotropic ($d_{\text{EL-vmf}}$) to a non-isotropic setup ($d_{\text{EL-nivMF}}$). Results are summarized in Tab. 5, showcasing a consistent improvement

Dataset	Metric	$d_{\cos} \rightarrow d_{\text{EL-vMF}}$	$d_{\cos} \rightarrow d_{\text{EL-nivMF}}$
CARS	Cluster-Div.↑	$+24\%$	$+31\%$
	Feat.-Div. ↑	$+13\%$	$+14\%$
CUB	Cluster-Div. ↑	$+11\%$	$+25\%$
	Feat.-Div. ↑	$+6\%$	$+8\%$

Table 5: Metrics on how **EL-nivMF** structures the embeddings.

in both feature and cluster diversity when incorporating both a probabilistic treatment and a non-isotropic encoding of proxy distributions. This provides further heuristic evidence linking the usage of $d_{\text{EL-nivMF}}$ to a better capture of the semantic class variability as well as an improved incorporation of a more diverse feature set, shown to facilitate generalisation [8,4].

M Further Qualitative Embedding Norm Studies

Fig. 10: CARS train images with lowest (left) to highest (right) embedding norms on a $M = 512$ dimensional ResNet-50 backend.

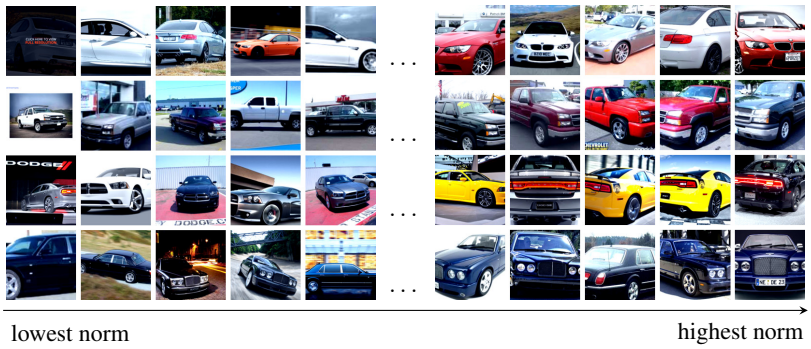


Fig. 11: Images for four randomly chosen classes (rows) of the CARS training set, ordered by their norm from lowest (left) to highest (right). Obtained from the $d_{\text{EL-vMF}}$ model on a ResNet-50, where the norms of image embeddings range from 70.58 to 140.09 whereas the proxy norms are between 45.95 to 79.98.

References

1. Jebara, T., Kondor, R.: Bhattacharyya and expected likelihood kernels. In: Learning Theory and Kernel Machines (2003) 3, 4
2. Kim, M.: On PyTorch implementation of density estimators for von Mises-Fisher and its mixture. arXiv preprint arXiv:2102.05340 (2021) 2
3. Kumar, S., Tsvetkov, Y.: Von Mises-Fisher loss for training sequence to sequence models with continuous outputs. In: International Conference on Learning Representations (ICLR) (2019) 1
4. Milbich, T., Roth, K., Bharadhwaj, H., Sinha, S., Bengio, Y., Ommer, B., Cohen, J.P.: Diva: Diverse visual feature aggregation for deep metric learning. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Proceedings of the European Conference on Computer Vision (ECCV) (2020) 8
5. Olver, F.W.J., Daalhuis, A.B.O., Lozier, D.W., Schneider, B.I., Boisvert, R.F., Clark, C.W., Miller, B.R., Saunders, B.V., Cohl, H.S., M. A. McClain, e.: NIST Digital library of mathematical functions, https://dlmf.nist.gov/10.37 1
6. Petersen, K.B., Pedersen, M.S., et al.: The matrix cookbook. Technical University of Denmark **7**(15) (2008) 2
7. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2021) 1
8. Roth, K., Milbich, T., Sinha, S., Gupta, P., Ommer, B., Cohen, J.P.: Revisiting training strategies and generalization performance in deep metric learning. In: Proceedings of the 37th International Conference on Machine Learning (ICML) (2020) 8
9. Scott, T.R., Gallagher, A.C., Mozer, M.C.: von Mises-Fisher loss: An exploration of embedding geometries for supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021) 1
10. Shi, Y., Jain, A.K.: Probabilistic face embeddings. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019) 3