Horospherical Learning with Smart Prototypes

Supplementary Material

Paul Berg[1]
paul.berg@irisa.fr
Bjoern Michele[1,2]
bjoern.michele@valeo.ai
Minh-Tan Pham[1]
minh-tan.pham@irisa.fr
Laetitia Chapel[1,3]
laetitia.chapel@irisa.fr
Nicolas Courty[1]
nicolas.courty@irisa.fr

[1] IRISA
Université de Bretagne Sud
Vannes, France
[2] Valeo AI
Paris, France
[3] Institut Agro
Rennes, France

Supplementary Materials - Horospherical Learning with Smart Prototypes

A Impact of $\phi(d)$ regularization

In order to evaluate the impact of the regularization introduced in Equation 6 we perform an experimental study of varying the value of the $\phi(d)$ parameter. Since we use a constant regularization defined as $\phi(d) = \lambda \times d$, we vary the value of the λ parameter in the range $[0,2]$.

B Impact of the bias a

The bias a in Equation 6 acts on the radius of each horosphere. We experiment with disabling the bias parameter in order to evaluate its impact on the final performance of the model. As can be seen in Table 1, the bias provides in 2 dimensions an increase in performance but seems to be of less importance for higher dimensions.

C Radius of an Horosphere

Given an horosphere parameterized by an ideal prototype $\mathrm{p} \in \mathbb{S}^{d-1}$ and a bias $a \in \mathbb{R}$. We can compute the radius of the said horosphere. Remember that horospheres in the Poincaré ball

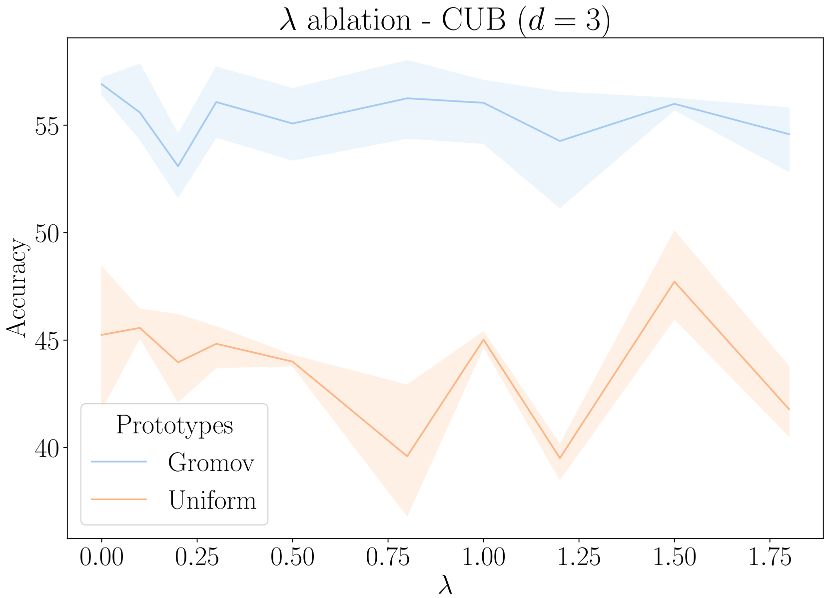

Figure 1: Evolution of the accuracy when moving the λ regularisation parameter. We observe that varying the value of the parameter has little impact on the resulting performance.

Table 1: Performance when enabling or disabling biases during training of horospherical classifiers on the CUB dataset.

Method	Bias	Dimensions 2	3	4	50
Horospherical - Smart		34.4±6.2	56.4±0.8	56.0±0.4	**57.5**±**0.2**
Horospherical - Smart	✓	**38.7**±**4.3**	56.4±1.6	**56.3**±**0.8**	56.6±0.6

model are hyperspheres tangent to the boundary of the ball. To compute the radius dependent on a, we will find the two points of the horosphere which are located on the Poincaré ball radius, one of this point is p, and we refer to the other one as x. By taking p as our first base vector, x has a single non-null dimension x_0.

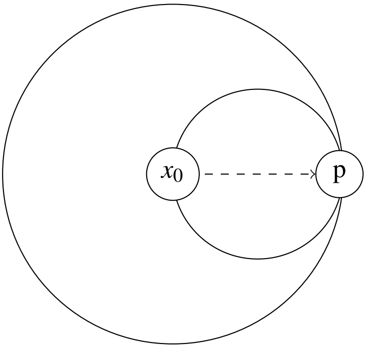

Figure 2: The position of x along the vector p is x_0.

$$
\begin{aligned}
-B_{\mathrm{p}}(x) + a &= 0 && (1)\\
\log\left(\frac{\|\mathrm{p}-x\|^2}{1-\|x\|^2}\right) - a &= 0 && (2)\\
\log\left(\exp(-a)\frac{\|\mathrm{p}-x\|^2}{1-\|x\|^2}\right) &= 0 && (3)\\
\exp(-a)\frac{\|\mathrm{p}-x\|^2}{1-\|x\|^2} &= 1 && (4)\\
\|\mathrm{p}-x\|^2 &= \frac{1-\|x\|^2}{\exp(-a)} && (5)\\
(1-x_0)^2 &= \frac{1-x_0^2}{\exp(-a)} && (6)\\
1-2x_0+x_0^2 &= \frac{1-x_0^2}{\exp(-a)} && (7)\\
1-2x_0+x_0^2 &= (1-x_0^2)\exp(a) && (8)\\
1-\exp(a)-2x_0+(1+\exp(a))x_0^2 &= 0. && (9)
\end{aligned}
$$

We find the roots for this polynomial:

$$
\begin{aligned}
\Delta &= (-2)^2 - 4(1-\exp(a))(1+\exp(a)) && (10)\\
&= 4-4(1-\exp(2a)) && (11)\\
&= 4\exp(2a). && (12)
\end{aligned}
$$

$$
\begin{aligned}
x_0 &= \frac{2 \pm \sqrt{\Delta}}{2(1+\exp(a))} && (13)\\
&= \frac{2 \pm \sqrt{4\exp(2a)}}{2(1+\exp(a))} && (14)\\
&= \frac{2 \pm 2\sqrt{\exp(2a)}}{2(1+\exp(a))} && (15)\\
&= \frac{2(1 \pm \exp(a))}{2(1+\exp(a))} && (16)\\
&= \frac{1 \pm \exp(a)}{1+\exp(a)} && (17)\\
&= \begin{cases} 1. & \text{if } x = \text{p}. \\ -\tanh(a/2) & \text{otherwise.} \end{cases} && (18)
\end{aligned}
$$

Therefore, the radius of an horosphere with a bias term a is:

$$
r(a) = \frac{1+\tanh(a/2)}{2}. \tag{19}
$$

D Hierarchies

In this section, we include the hierarchies used for positioning prototypes in our experiments on different datasets. For large hierarchies, we do not include the label of nodes for readability purpose. The nodes are coloured following the topological sort of the tree.

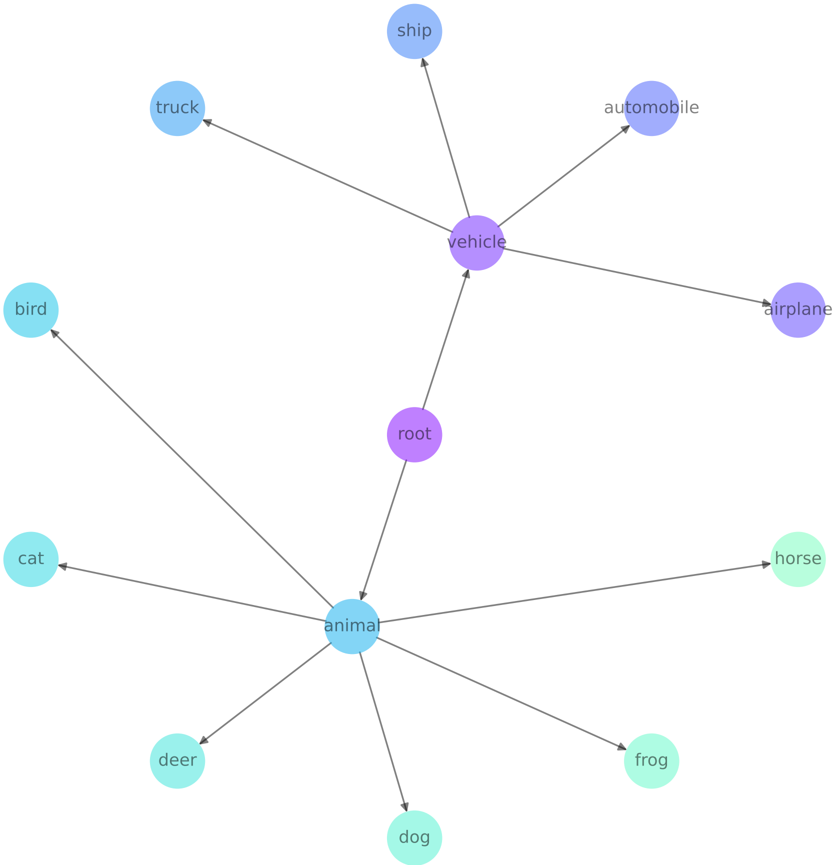

Figure 3: CIFAR10 Hierarchy.

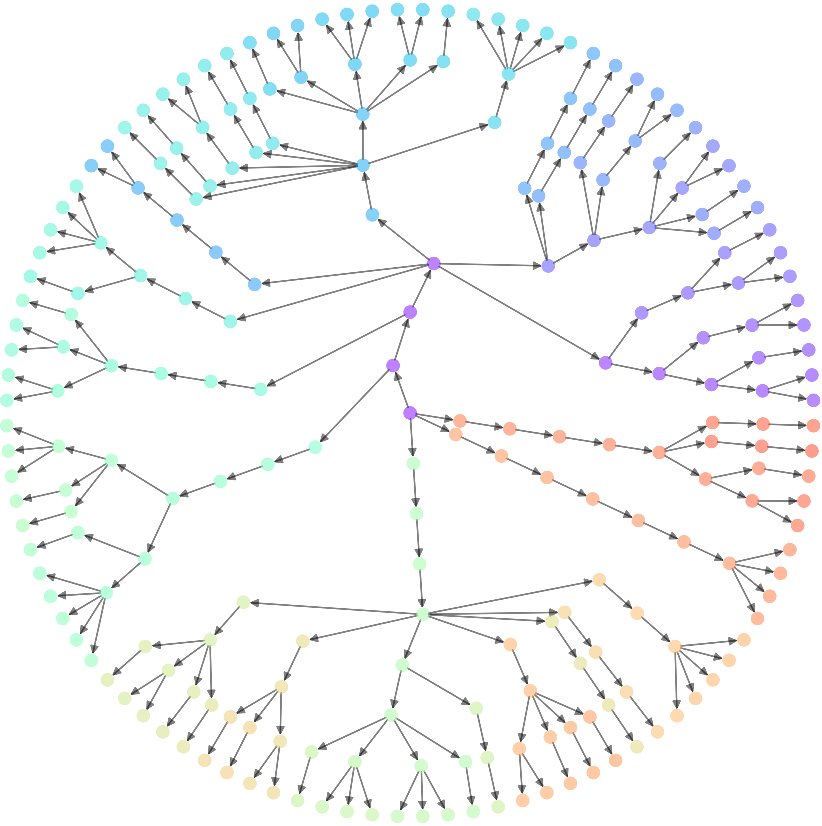

Figure 4: CIFAR100 Hierarchy.

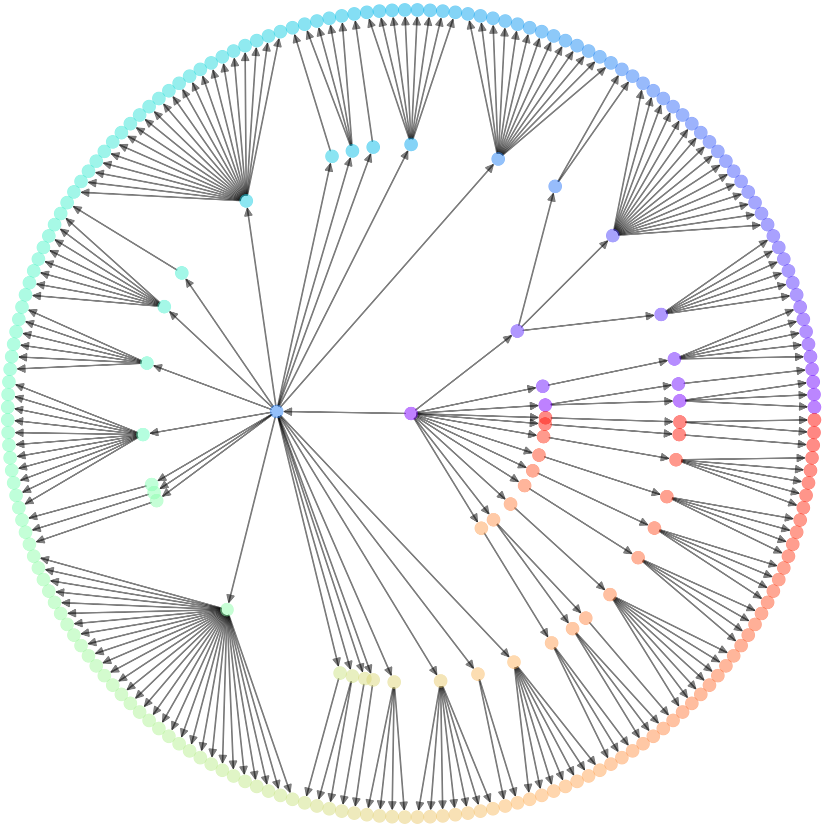

Figure 5: CUB200 Hierarchy.

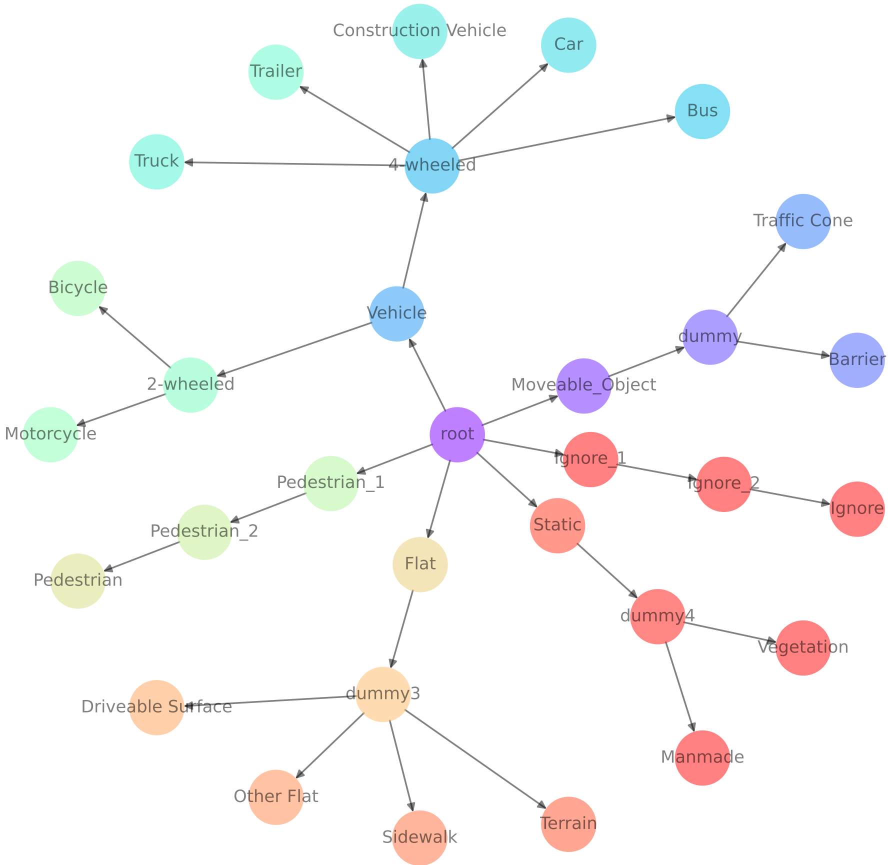

Figure 6: NuScenes Hierarchy.

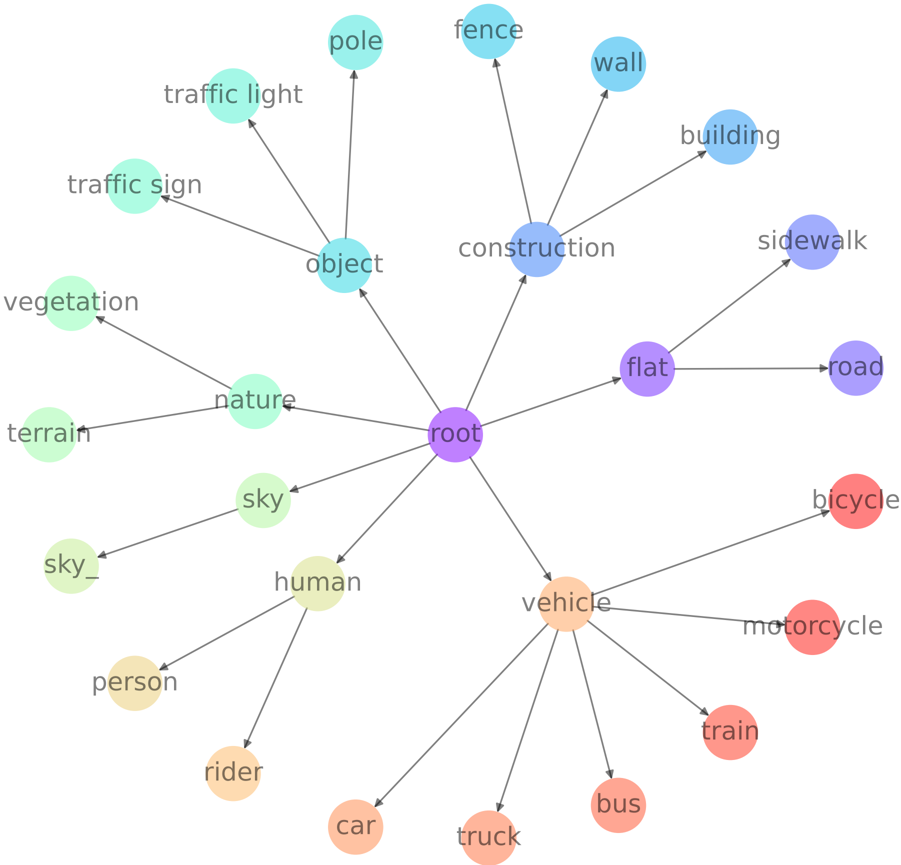

Figure 7: Cityscapes Hierarchy.