GPS: A Probabilistic Distributional Similarity with Gumbel Priors for Set-to-Set Matching

Ziming Zhang, Fangzhou Lin*, Haotian Liu*, Jose Morales, Haichong Zhang, Kazunori Yamada, Vijaya B Kolachalama, and Venkatesh Saligrama*

ICLR 2025

Code is available at http://github.com/Zhang-VISLab/ICLR2025-GPS

Introduction

Problems Statements:

- Many computer vision tasks can be recognized as a set-to-set matching problem.
- Learning effective set representations requires training feature extractors with a loss that directly optimizes inter-set similarity.
- Our approach can improve model robustness while reducing computational overhead.

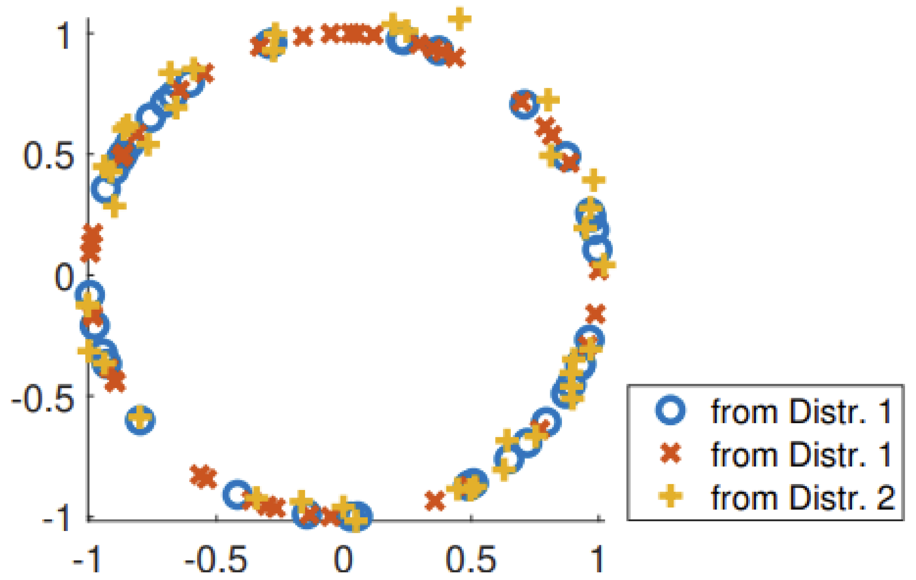

Fig1. Three sets randomly sampled from a circular distribution and a Gaussian

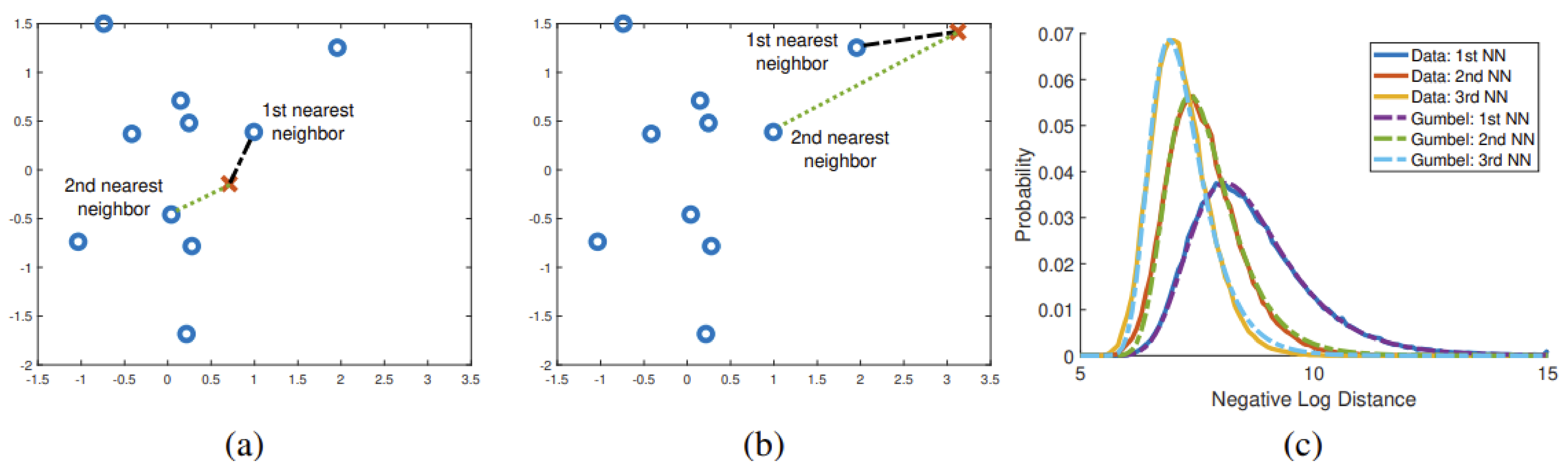

Fig2. Illustration of (a-b) samples from the same/different distributions, and (c) data fitting with Gumbel prior distributions for 1st, 2nd, 3rd smallest distance distributions between two 2D point sets.

Motivation:

- We model minimum-distance distributions between point sets using Gumbel distributions.
- We borrow NLP's distributional similarity concept to enhance matching.
- We demonstrate better performance while keeping linear complexity like Chamfer Distance.

The probability density function (PDF) of a **Gumbel distribution**:

$$p(x) = \frac{1}{\sigma}\exp\left(-(y + \exp(-y))\right), \quad y = \frac{x - \mu}{\sigma}$$

Methodology

Distributional Signatures :

$$\mathcal{D}(\mathcal{X}_1, \mathcal{X}_2) = \left\{ d_{min}^{(k)}(x_{1,i}) = |x_{1,i} - x_{2,i_k}|, d_{min}^{(k)}(x_{2,j}) = |x_{2,j} - x_{1,j_k}| \mid \forall k \in [K], \forall i, \forall j \right\}$$

Probabilistic Modeling :

$$p(\mathcal{P}_1 = \mathcal{P}_2 \mid \mathcal{X}_1, \mathcal{X}_2) = \sum_{q \in \mathcal{Q}} \sum_{d_{min} \in \mathcal{D}} p(\mathcal{P}_1 = \mathcal{P}_2, q, d_{min} \mid \mathcal{X}_1, \mathcal{X}_2)$$

$$= \sum_{q \in \mathcal{Q}} \sum_{d_{min} \in \mathcal{D}} p(q) p(\mathcal{P}_1 = \mathcal{P}_2 \mid q) p(d_{min} \mid q, \mathcal{X}_1, \mathcal{X}_2)$$

Probabilistic Modeling with Mixture of Models based on KNN:

$$p(\mathcal{P}_1 = \mathcal{P}_2 \mid \mathcal{X}_1, \mathcal{X}_2) \propto \sum_{k,m} \left[\sum_i p\left(d_{min}^{(k)}(x_{1,i}); \alpha_{k,m}, \beta_{k,m}\right) + \sum_j p\left(d_{min}^{(k)}(x_{2,j}); \alpha_{k,m}, \beta_{k,m}\right) \right]$$

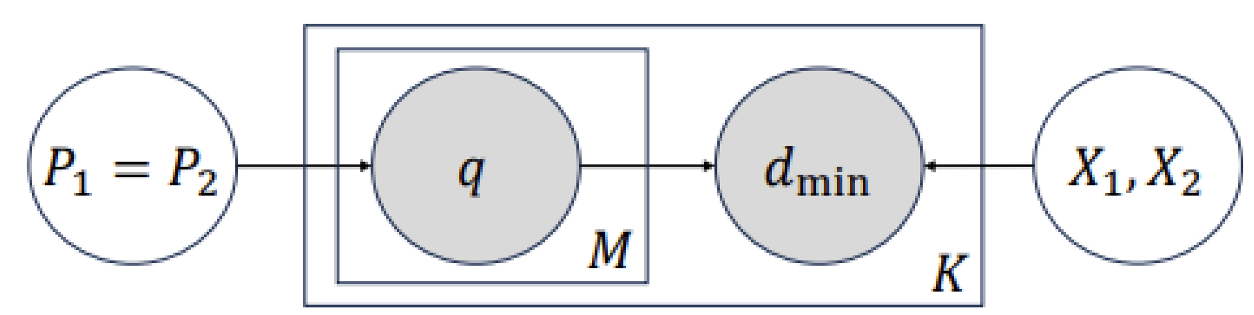

Fig 3. Graphical model for computing the conditional probability

Pseudocode for GPS :

```
def Gumbel_Fit(dis,a,b):

    min_dis = a * dis ** b
    t = exp(log(min_dis))
    sim = mean(-t * exp(-t))
    return sim

def Set_Similarity(X1,X2,a1,a2,b1,b2):

    D = distance_matrix(X1,X2)
    sim0 = Gumbel_Fit(D.min(0),a1,b1)
    sim1 = Gumbel_Fit(D.min(1),a2,b2)
    return sim0 + sim1
```

Performance

Few-shot Image Classification

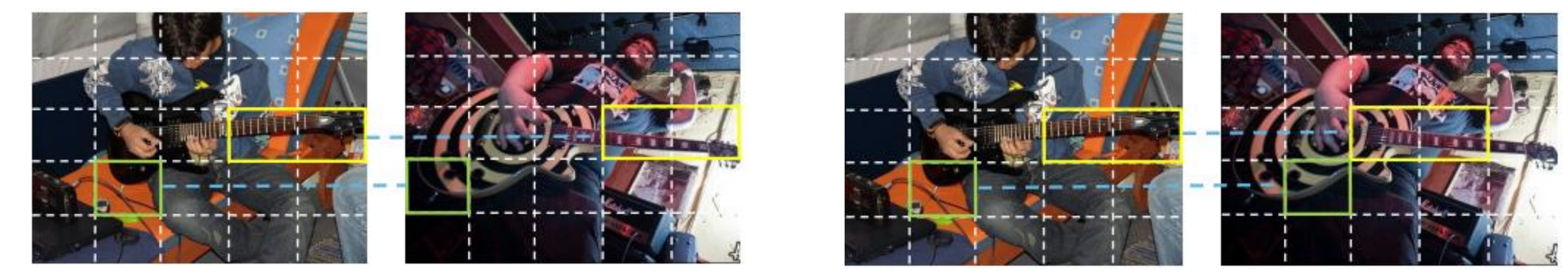

Fig 4. Visual matching results by (left) our GPS and (right) DeepEMD.

Method	*mini*ImageNet		*tiered*ImageNet	
	1-shot	5-shot	1-shot	5-shot
DeepEMD	63.36±0.75	79.15±0.66	70.48±0.78	83.89±0.67
CD	63.40±0.46	79.54±0.39	70.23±0.64	84.01±0.31
PWD	63.92±0.77	78.77±0.37	70.69±0.92	83.88±0.34
SWD	63.15±0.76	78.46±0.41	69.72±0.93	83.02±0.33
GSWD	63.66±0.72	78.92±0.47	70.25±0.86	83.62±0.31
ASWD	63.16±0.75	78.87±0.45	69.30±0.91	83.71±0.38
HyperCD	63.63±0.65	79.78±0.73	70.58±0.81	84.27±0.48
InfoCD	64.01±0.32	80.87±0.64	70.97±0.59	84.54±0.36
Ours: GPS	**66.27±0.37**	**81.19±0.47**	**73.16±0.43**	**85.52±0.48**

Tab.1 Results of 5-way (%) on miniImageNet and tieredImageNet datasets.

Point Cloud Completion

Fig 5. Row-1: Inputs of incomplete point clouds. Row-2: Outputs of Seedformer with CD. Row-3: Outputs of Seedformer with GPS. Row-4: Ground truth.