Combining Statistical Depth and Fermat Distance for Uncertainty Quantification

Hai-Vy Nguyen[1,2,3], Fabrice Gamboa[2], Reda Chhaibi[2], Sixin Zhang[3], Serge Gratton[3], Thierry Giaccone[1]

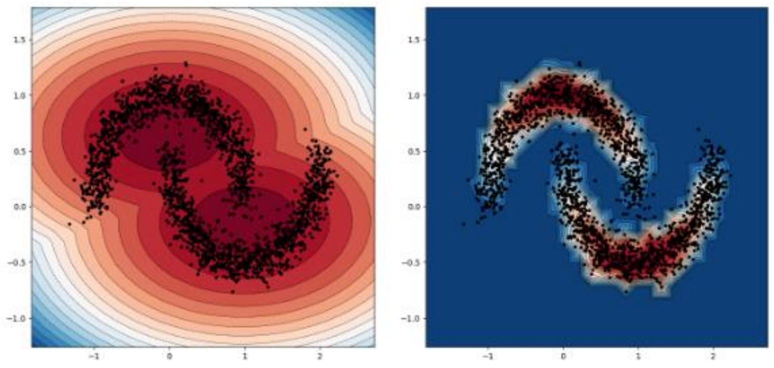
[1] Ampere Software Technology

[2]Institut de mathématiques de Toulouse

[3]Institut de Recherche en Informatique de Toulouse

High-level Idea and General Scheme

- Measure “how central” a point is with respect to (w.r.t.) the training set taking into account density and geometry of the data in the feature space.



(a) Gaussian fitting. (b) Our method.

Figure 1. Motivation example.

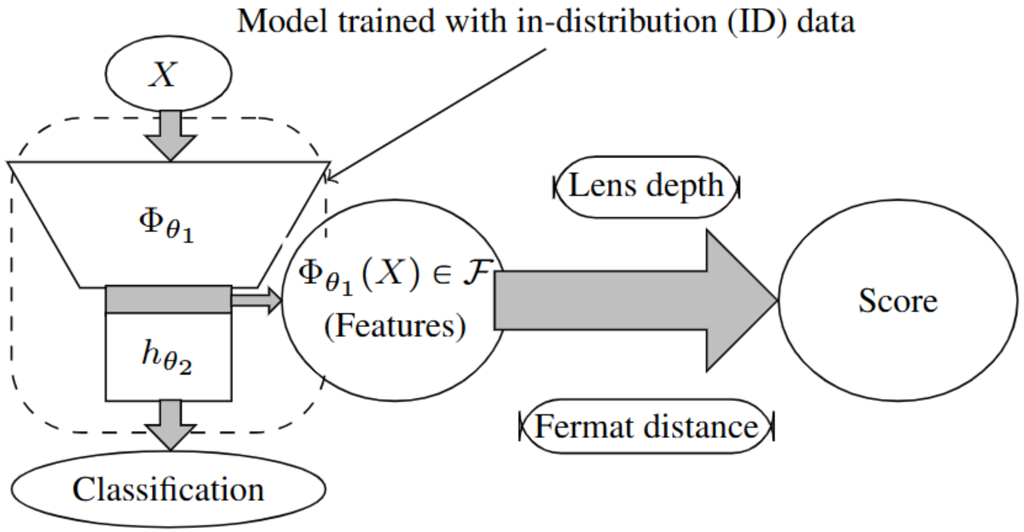
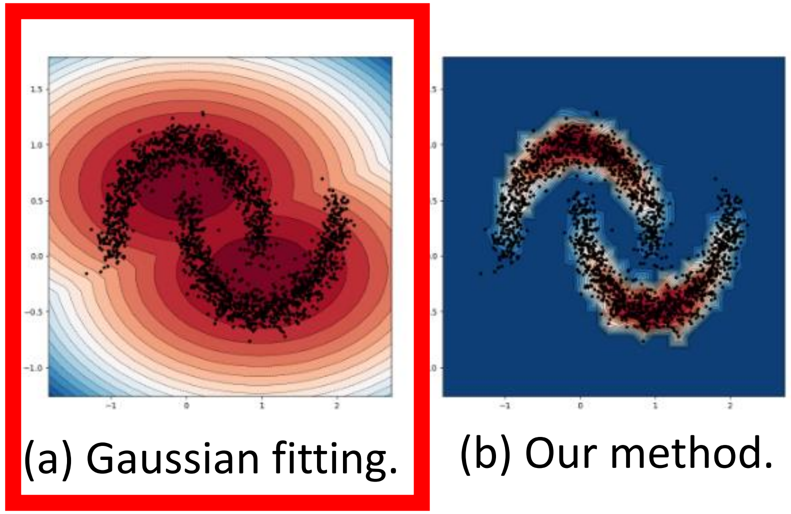


Figure 2. General scheme.

High-level Idea and General Scheme

- Measure “how central” a point is with respect to (w.r.t.) the training set taking into account density and geometry of the data in the feature space.
- Standard methods consist in assuming prior distribution such as Gaussian.



(a) Gaussian fitting. (b) Our method.

Figure 1. Motivation example.

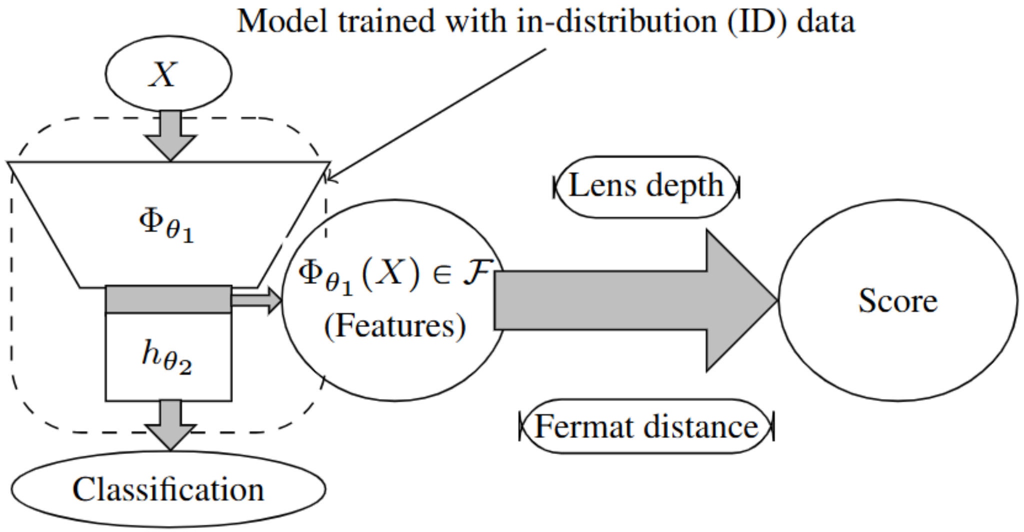
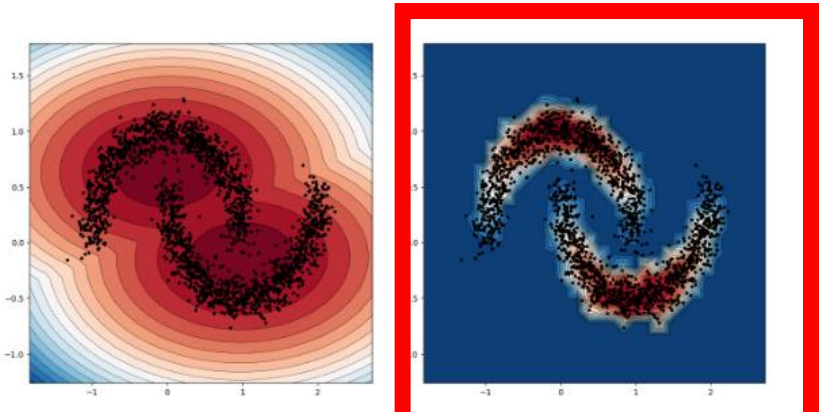


Figure 2. General scheme.

High-level Idea and General Scheme

- Measure "how central" a point is with respect to (w.r.t.) the training set taking into account density and geometry of the data in the feature space.
- Standard methods consist in assuming prior distribution such as Gaussian.
- Our method has no distributional assumption, by combining Lens Depth with FD.



(a) Gaussian fitting. (b) Our method.

Figure 1. Motivation example.

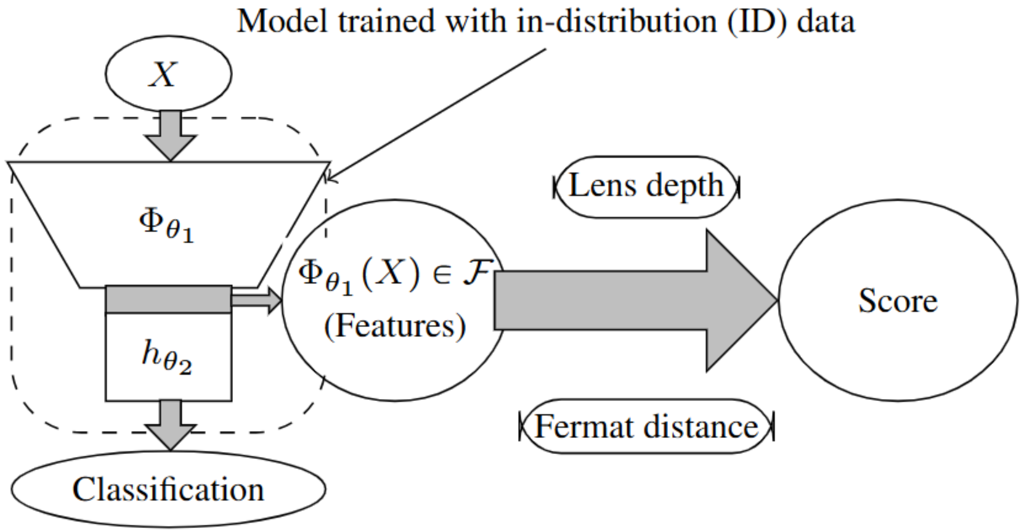
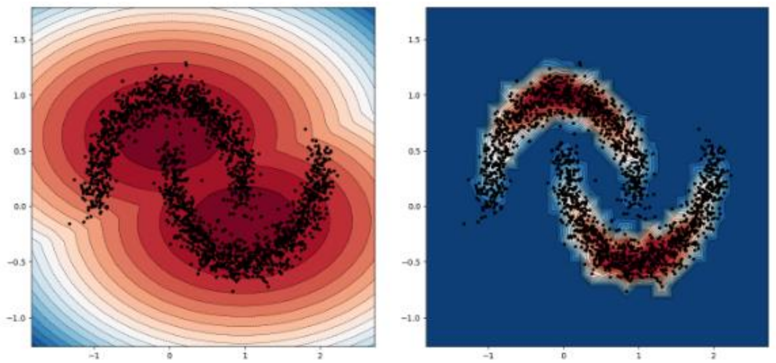


Figure 2. General scheme.

High-level Idea and General Scheme

- Measure "how central" a point is with respect to (w.r.t.) the training set taking into account density and geometry of the data in the feature space.
- Standard methods consist in assuming prior distribution such as Gaussian.
- Our method has no distributional assumption, by combining Lens Depth with FD.
- The larger the depth of an example, the more typical this point is (relative to the training set), the easier it is for the model to classify.
- This gives us an out-of-domain uncertainty score.



(a) Gaussian fitting. (b) Our method.

Figure 1. Motivation example.

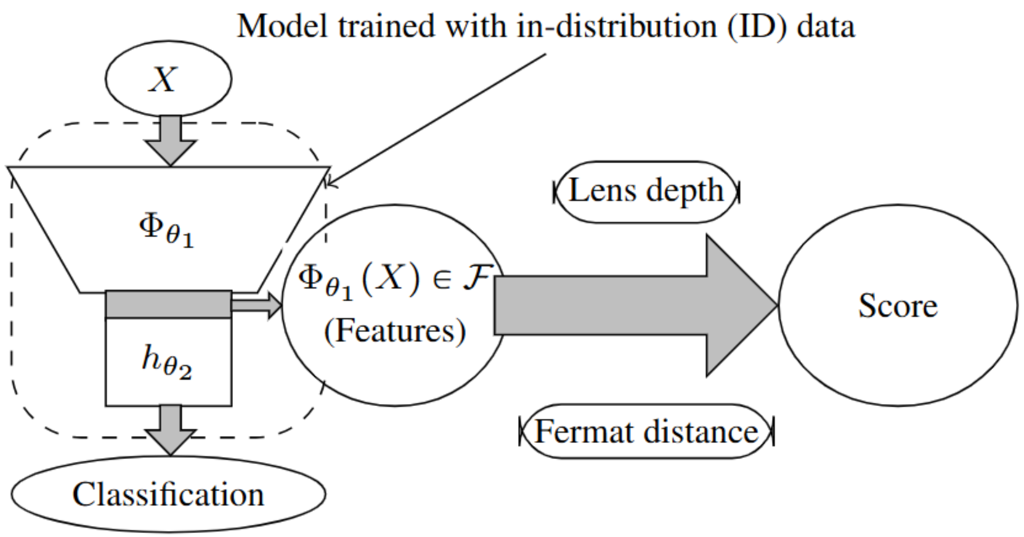
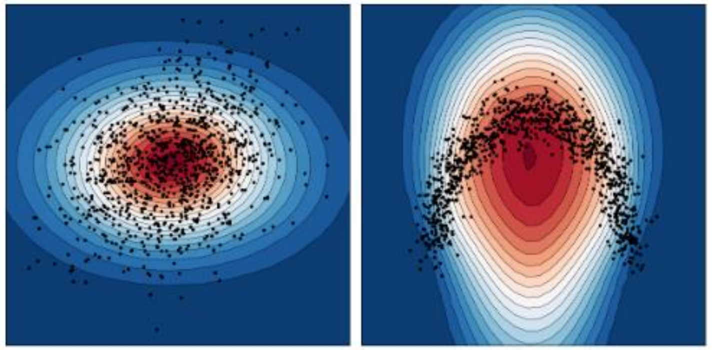


Figure 2. General scheme.

Lens Depth

- LD measures the centrality of a point w.r.t. to a cluster.
- The greater the depth, the more central the point is to the distribution.
- Empirical LD is defined as

$$\widehat{LD}_n(x) := \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} 1_{A(X_{i_1}, X_{i_2})}(x) \ .$$



(a) Gaussian (b) Moon

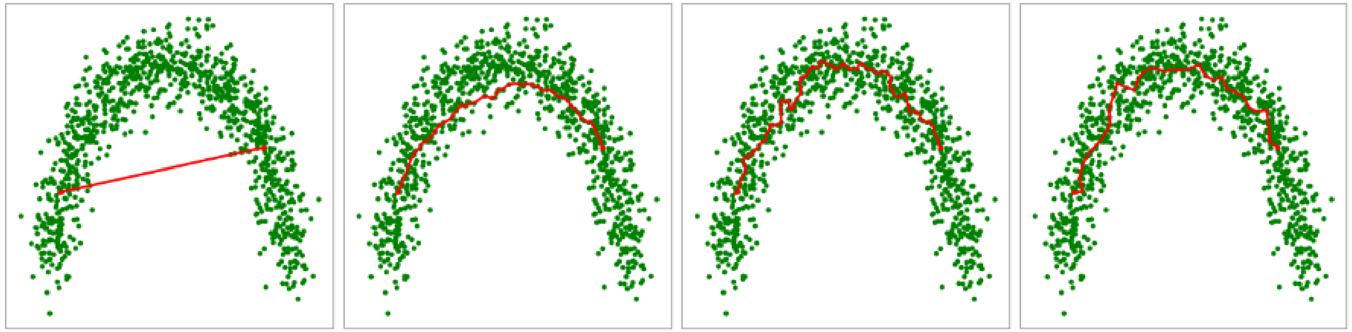
Figure 3. Lens Depth using Euclidean distance.

- $A(X_1, X_2) = \{x: \max_{i=1,2} D(x, X_i) \le D(X_1, X_2)\}$ ⇒ The choice of the distance D is important !
- A naive choice is the Euclidean one ⇒ Unsatisfying results (Fig. 3) .
- Need some distance that can adapt to the distribution.

Sample Fermat Distance (SFD)

- Need some distance that can adapt to the distribution ⇒ Fermat Distance
- Q: a non-empty, locally finite, subset of $\mathbb{R}^d$. $|x|$ denotes Euclidean norm of x, $q_Q(x) \epsilon Q$ is the particle closest to x in Euclidean distance. For α ≥ 1, and x, $y \epsilon \mathbb{R}^d$, SFD is defined as

$$D_{Q,\alpha}(x,y) := \min\left\{\sum_{j=1}^{k-1} |q_{j+1}-q_j|^\alpha : (q_1, \dots, q_k) \epsilon Q^k \text{ with } q_1 = q_Q(x), q_k = q_Q(y), k \geq 1\right\}. \quad (3)$$



(a) $\alpha = 1$ (b) $\alpha = 1.2$ (c) $\alpha = 3$ (d) $\alpha = 7$

Figure 4. Sample Fermat path

Artifact from Sample Fermat Distance (SFD)

- $\widehat{LD}$ using SFD.
- The shape of datasets is much better captured.
- However, some zones have constant LD value (represented by the same color).
- We can show that $\widehat{LD}_n$ is constant over the Voronoï cells associated to Q.

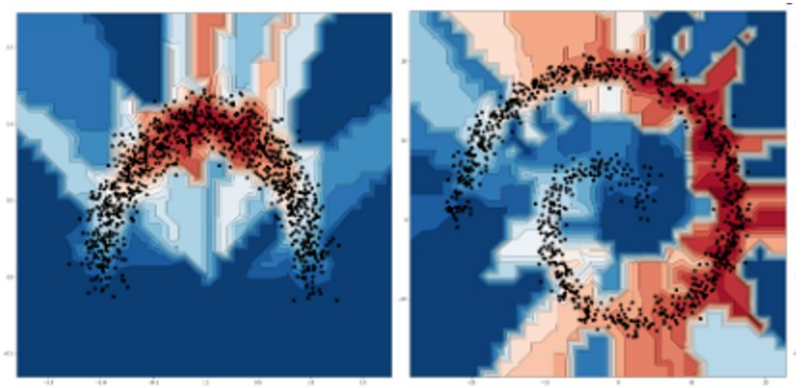
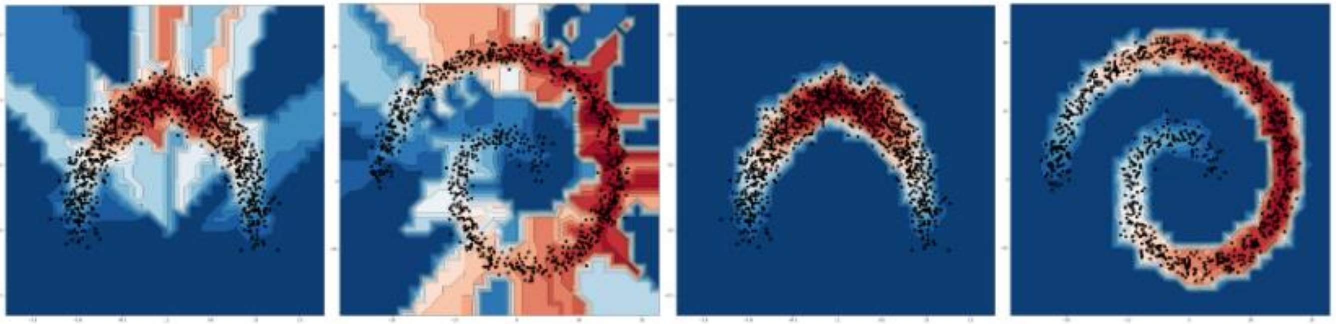


Figure 5. Lens Depth using SFD.

Modified Sample Fermat Distance

To avoid this undesirable artifact, we proposed modified SFD defined as

$$\boxed{D_{Q,\alpha}^{modif}(x,y) := \min_{q \in Q}\{|x-q|^{\alpha} + D_{Q,\alpha}(q,y)\}.}$$



(a) Before (b) Before (c) After (d) After

SFD Modified SFD

Figure 6. Lens Depth using SFD and Modified SFD.

Two-moon Dataset

Different methods for uncertainty estimation are applied on a same neural net trained to classify 2 classes in moon-shape. Uncertainty estimations are computed based solely on the feature space of the model.

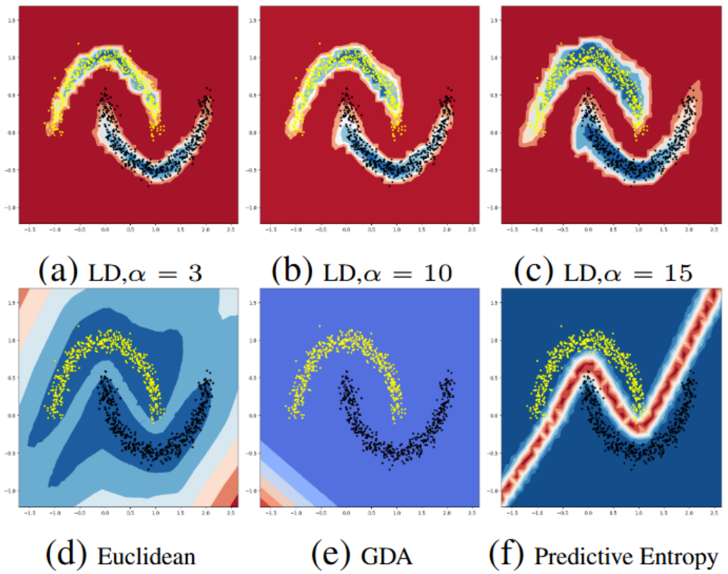


Figure 7. Different methods for uncertainty estimation. Red represents high uncertainty.

Experiments with different neural nets and data sets

Table 1. Results on CIFAR10 with Tiny-ImageNet, CIFAR100 and SVHN as OOD sets.
SN: SpectralNormalisation, JP: Jacobian Penalty.

Method	Model	Penalty	AUROC Tiny-ImageNet	AUROC CIFAR100	AUROC SVHN
LD (ours)	ResNet18	No	**0.965** $\pm$0.003	**0.892** $\pm$0.002	**0.936** $\pm$0.006
GDA	ResNet18	No	0.945 $\pm$0.005	0.864 $\pm$0.003	0.914 $\pm$0.014
LD (ours)	ResNet18	SN	0.927 $\pm$0.003	**0.900** $\pm$0.001	**0.950** $\pm$0.008
DDU (GDA + SN)	ResNet18	SN	**0.937** $\pm$0.009	0.872 $\pm$0.005	0.947 $\pm$0.015
DUQ	ResNet18	JP	—	—	0.927 $\pm$0.013
LD (ours)	Wide-ResNet-28-10	SN	**0.926** $\pm$0.002	0.906 $\pm$0.001	0.939 $\pm$0.007
DDU (GDA + SN)	Wide-ResNet-28-10	SN	0.9107 $\pm$0.005	**0.913** $\pm$0.004	**0.979** $\pm$0.002
DUQ	Wide-ResNet-28-10	JP	0.868 $\pm$0.001	0.859 $\pm$0.003	0.937 $\pm$0.006
SNGP	Wide-ResNet-28-10	SN	0.899 $\pm$0.002	0.911 $\pm$0.002	0.940 $\pm$0.001
Energy-based	Wide-ResNet-28-10	No	0.881 $\pm$0.006	0.889 $\pm$0.007	0.945 $\pm$0.005
5-Ensemble	Wide-ResNet-28-10	No	0.9006 $\pm$0.003	0.921 $\pm$0.002	0.977 $\pm$0.003

Table 2. AUROC score with CIFAR100 as ID data and Tiny-Imaget as OOD data.

Method	AUROC
LD (ours)	0.8310 $\pm$0.0013
Softmax Entropy	0.8153 $\pm$0.0005
Energy-based	0.8133 $\pm$0.0006
SNGP	0.7885 $\pm$0.0004
DDU	0.8313 $\pm$0.0006
5-Ensemble	0.8295 $\pm$0.0009

Thank you !!!