Learnable Uncertainty under Laplace Approximations (Supplementary Material)

Agustinus Kristiadi[*1] **Matthias Hein**[1] **Philipp Hennig**[1,2]

[1]University of Tübingen, Tübingen, Germany
[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany

APPENDIX A PROOFS

Proposition 1 (Properties). *Let $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ be a MAP-trained L-layer network under dataset $\mathcal{D}$, and let θ_{MAP} be the MAP estimate. Suppose $\widetilde{f} : \mathbb{R}^n \times \mathbb{R}^{\widetilde{d}} \to \mathbb{R}$ and $\widetilde{\theta}_{\text{MAP}} \in \mathbb{R}^{\widetilde{d}}$ are obtained via the previous construction, and $\widetilde{\mathcal{L}}$ is the resulting loss function under $\widetilde{f}$.*

(a) For an arbitrary input $x \in \mathbb{R}^n$, we have $\widetilde{f}(x; \widetilde{\theta}_{\text{MAP}}) = f(x; \theta_{\text{MAP}})$.

(b) The gradient of $\widetilde{\mathcal{L}}$ w.r.t. the additional weights in $\widetilde{W}^{(L)}$ is non-linear in $\widehat{\theta}$.

Proof. As the first order of business, for each layer $l = 1, \dots, L$ we denote the hidden units and pre-activations of $\widetilde{f}$ as $\widetilde{h}^{(l)}$ and $\widetilde{a}^{(l)}$, respectively.

We begin with (a). Let $x \in \mathbb{R}^n$ be arbitrary. We need to show that the output of $\widetilde{f}$, i.e. the last pre-activations $\widetilde{a}^{(L)}$, is equal to the last pre-activations $a^{(L)}$ of f. For the first layer, we have that

$$\begin{aligned}\widetilde{a}^{(1)} &= \widetilde{W}^{(1)} x + \widetilde{b}^{(1)} \\ &= \begin{pmatrix} W^{(1)} \\ \widehat{W}_1^{(1)} \end{pmatrix} x + \begin{pmatrix} b^{(1)} \\ \widehat{b}^{(1)} \end{pmatrix} \\ &= \begin{pmatrix} W^{(1)} x + b^{(1)} \\ \widehat{W}_1^{(1)} x + \widehat{b}^{(1)} \end{pmatrix} =: \begin{pmatrix} a^{(1)} \\ \widehat{a}^{(1)} \end{pmatrix} .\end{aligned} \tag{1}$$

For every layer $l = 1, \dots, L-1$, we denote the hidden units as the block vector

$$\widetilde{h}^{(l)} = \begin{pmatrix} \varphi(a^{(l)}) \\ \varphi(\widehat{a}^{(l)}) \end{pmatrix} = \begin{pmatrix} h^{(l)} \\ \widehat{h}^{(l)} \end{pmatrix} .$$

Now, for the intermediate layer $l = 2, \dots, L-1$, we observe that

$$\begin{aligned}\widetilde{a}^{(l)} &= \widetilde{W}^{(l)} \widetilde{h}^{(l-1)} + \widetilde{b}^{(l)} \\ &= \begin{pmatrix} W^{(l)} & 0 \\ \widehat{W}_1^{(l)} & \widehat{W}_2^{(l)} \end{pmatrix} \begin{pmatrix} h^{(l-1)} \\ \widehat{h}^{(l-1)} \end{pmatrix} + \begin{pmatrix} b^{(l)} \\ \widehat{b}^{(l)} \end{pmatrix} \\ &= \begin{pmatrix} W^{(l)} h^{(l-1)} + 0 + b^{(l)} \\ \widehat{W}_1^{(l)} h^{(l-1)} + \widehat{W}_2^{(l)} \widehat{h}^{(l-1)} + \widehat{b}^{(l)} \end{pmatrix} =: \begin{pmatrix} a^{(l)} \\ \widehat{a}^{(l)} \end{pmatrix} .\end{aligned} \tag{2}$$

Finally, for the last layer, we get

$$\begin{aligned}\widetilde{a}^{(L)} &= \widetilde{W}^{(L)} x + \widetilde{b}^{(L)} \\ &= \begin{pmatrix} W^{(L)} & 0 \end{pmatrix} \begin{pmatrix} h^{(L-1)} \\ \widehat{h}^{(L-1)} \end{pmatrix} + b^{(L)} \\ &= W^{(L)} h^{(L-1)} + 0 + b^{(L)} \\ &= a^{(L)} ,\end{aligned} \tag{3}$$

and thus we have the desired invariance.

For part (b), we denote the additional (zero) weights in $\widetilde{W}^{(L)}$ by $\widehat{W}^{(L)}$. It is clear from (3) that the gradient $\nabla_{\widehat{W}^{(L)}} a^{(L)}$ is given by $\widehat{h}^{(L-1)}$. Hence, by chain rule we have

$$\begin{aligned}\nabla_{\widehat{W}^{(L)}} \widetilde{\mathcal{L}} &= (\nabla_{a^{(L)}} \widetilde{\mathcal{L}}) \, (\nabla_{\widehat{W}^{(L)}} a^{(L)}) \\ &= (\nabla_{a^{(L)}} \widetilde{\mathcal{L}}) \, \widehat{h}^{(L-1)} .\end{aligned}$$

By observing (1) and (2), along the fact that the non-linearity φ is used in the forward pass, it is clear that $\widehat{h}^{(L-1)}$ is non-linear in $\widehat{\theta} = (\widehat{W}_1^{(1)}, \widehat{b}^{(1)}, \dots, \widehat{W}_1^{(L-1)}, \widehat{W}_2^{(L-1)}, \widehat{b}^{(L-1)})$ and therefore $\nabla_{\widehat{W}^{(L)}} \widetilde{\mathcal{L}}$ also is. □

Proposition 2 (Predictive Uncertainty). *Suppose $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is a real-valued network and $\widetilde{f}$ is as constructed above. Suppose further that diagonal Laplace-approximated posteriors $\mathcal{N}(\theta_{\text{MAP}}, \text{diag}(\sigma))$, $\mathcal{N}(\widetilde{\theta}_{\text{MAP}}, \text{diag}(\widetilde{\sigma}))$ are employed for f and $\widetilde{f}$, respectively. Under the linearization (4), for any input $x \in \mathbb{R}^n$, the variance over the output $\widetilde{f}(x; \widetilde{\theta})$ is at least that of $f(x; \theta)$.*

[*]Correspondence to: agustinus.kristiadi@uni-tuebingen.de

Algorithm 2 Adding LULA units.

Input:

L-layer net with a MAP estimate $\theta_{\text{MAP}} = (W^{(l)}_{\text{MAP}}, b^{(l)}_{\text{MAP}})_{l=1}^{L}$. Sequence of non-negative integers $(m_l)_{l=1}^{L}$.

1: **for** $l = 1, \dots, L-1$ **do**

2: $\quad \operatorname{vec} \widehat{W}_1^{(l)} \sim p(\operatorname{vec} \widehat{W}_1^{(l)})$ ▷ Draw from a prior

3: $\quad \operatorname{vec} \widehat{W}_2^{(l)} \sim p(\operatorname{vec} \widehat{W}_2^{(l)})$ ▷ Draw from a prior

4: $\quad \widehat{b}^{(l)} \sim p(\widehat{b}^{(l)})$ ▷ Draw from a prior

5: $\quad \widetilde{W}^{(l)}_{\text{MAP}} = \begin{pmatrix} W^{(l)}_{\text{MAP}} & 0 \\ \widehat{W}_1^{(l)} & \widehat{W}_2^{(l)} \end{pmatrix}$ ▷ $0 \in \mathbb{R}^{n_l \times m_{l-1}}$

6: $\quad \widetilde{b}^{(l)}_{\text{MAP}} := \begin{pmatrix} b^{(l)}_{\text{MAP}} \\ \widehat{b}^{(l)} \end{pmatrix}$

7: **end for**

8: $\widetilde{W}^{(L)}_{\text{MAP}} = (W^{(L)}_{\text{MAP}}, 0)$ ▷ $0 \in \mathbb{R}^{k \times m_{L-1}}$

9: $\widetilde{b}^{(L)}_{\text{MAP}} = b^{(L)}_{\text{MAP}}$

10: $\widetilde{\theta}_{\text{MAP}} = (\widetilde{W}^{(l)}_{\text{MAP}}, \widetilde{b}^{(l)}_{\text{MAP}})_{l=1}^{L}$

11: **return** $\widetilde{\theta}_{\text{MAP}}$

Proof. Let us denote the random variable taking values in the augmented parameter space by $\widetilde{\theta}$. W.l.o.g. we re-arrange $\widetilde{\theta}$ as $(\theta^\top, \widehat{\theta}^\top)^\top$ where $\widehat{\theta} \in \mathbb{R}^{\widetilde{d}-d}$ contains the weights corresponding to the the additional LULA units. If $g(x)$ is the gradient of the output $f(x;\theta)$ w.r.t. θ at θ_{MAP}, then the gradient of $\widetilde{f}(x;\widetilde{\theta})$ w.r.t. $\widetilde{\theta}$ at $\widetilde{\theta}_{\text{MAP}}$, say $\widetilde{g}(x)$, can be written as the concatenation $(g(x)^\top, \widehat{g}(x)^\top)^\top$ where $\widehat{g}(x)$ is the corresponding gradient w.r.t. $\widehat{\theta}$. Furthermore, $\operatorname{diag}(\widetilde{\sigma})$ has diagonal elements

$$\left(\sigma_{11}, \dots, \sigma_{dd}, \widehat{\sigma}_{11}, \dots, \widehat{\sigma}_{\widetilde{d}-d,\widetilde{d}-d}\right)^\top =: (\sigma^\top, \widehat{\sigma}^\top)^\top .$$

Let $x \in \mathbb{R}^n$ be an arbitrary input. Denoting the output variance of $\widetilde{f}(x;\widetilde{\theta})$ by $\widetilde{v}(x)$, we have

$$\begin{aligned} \widetilde{v}(x) &= \widetilde{g}(x)^\top \operatorname{diag}(\widetilde{\sigma}) \widetilde{g}(x) \\ &= \underbrace{g(x)^\top \operatorname{diag}(\sigma) g(x)}_{=v(x)} + \widehat{g}(x)^\top \operatorname{diag}(\widehat{\sigma}) \widehat{g}(x) \\ &\geq v(x)\,, \end{aligned}$$

since $\operatorname{diag}(\widehat{\sigma})$ is positive-definite by definition. □

APPENDIX B IMPLEMENTATION

We summarize the augmentation of a network with LULA units in Algorithm 2. Note that the priors of the free parameters $\widehat{W}^{(l)}$, $\widehat{b}^{(l)}$ (lines 2 and 3) can be chosen as independent Gaussians—this reflects the standard procedure for initializing NNs' parameters.

Table 4: UQ performances on UCI datasets. Values are the average (over all data points and ten training-prediction trials) predictive standard deviations, i.e. the standard deviation of the Gaussian (5). Lower is better for test data and vice-versa for outliers. By definition, MAP does not have (epistemic) uncertainty.

	Test set ↓			Outliers ↑		
Dataset	**DE**	**LA**	**LA-LULA**	**DE**	**LA**	**LA-LULA**
Housing	5.82	1.26	1.37	145.33	222.76	377.92
Concrete	8.11	10.44	16.89	964.63	30898.92	83241.42
Energy	4.40	1.05	1.08	126.11	1070.09	5163.53
Kin8nm	0.10	0.14	0.18	2.12	0.80	2.12
Power	19.85	2.85	3.20	12235.87	4148.98	221287.80
Wine	0.64	1.15	1.22	28.57	186.76	21383.17
Yacht	5.17	2.08	2.78	187.41	5105.69	13119.99

Table 5: Predictive performances on UCI regression datasets in terms of average test log-likelihood. The numbers reported are averages over ten training-prediction runs along with the corresponding standard deviations. The performances of LULA are similar to LA's. The differences between their exact values are likely due to MC-integration.

Dataset	**MAP**	**DE**	**LA**	**LA-LULA**
Housing	-2.794±0.012	-3.045±0.009	-3.506±0.055	-3.495±0.047
Concrete	-3.409±0.036	-3.951±0.062	-4.730±0.205	-4.365±0.094
Energy	-2.270±0.128	-2.673±0.015	-2.707±0.030	-2.698±0.014
Kin8nm	-0.923±0.000	1.086±0.022	-0.965±0.003	-0.969±0.003
Power	-3.154±0.002	-54.804±7.728	-3.273±0.015	-3.277±0.024
Wine	-1.190±0.014	-1.038±0.018	-1.624±0.075	-1.630±0.092
Yacht	-1.835±0.053	-3.272±0.079	-2.509±0.367	-2.663±0.276

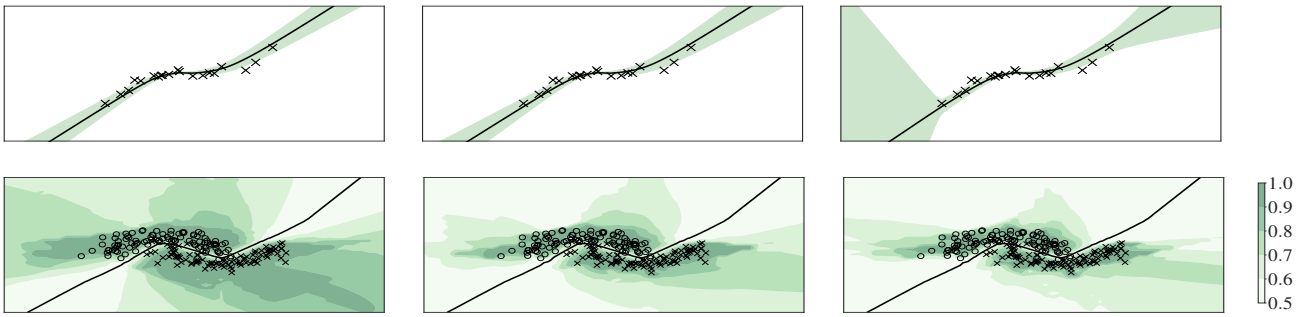
APPENDIX C ADDITIONAL RESULTS

C.1 TOY DATASET

To show the effectiveness of LULA training, we compare the standard Laplace, untrained LULA, and trained LULA in Fig. 6. As predicted by Proposition 2, untrained LULA increases predictive uncertainty estimates. However, this increase of uncertainty is not well-adapted to the data (b). Training $\widetilde{\theta}$ using (9) make it more calibrated to both inliers and outliers (c).

C.2 UCI REGRESSION

To validate the performance of LULA in regressions, we employ a subset of the UCI regression benchmark datasets. Following previous works, the network architecture used here is a single-hidden-layer ReLU network with 50 hidden units. The data are standardized to have zero mean and unit variance. We use 50 LULA units and optimize them for 40 epochs using OOD data sampled uniformly from



(a) Laplace (b) LULA-Untrained (c) LULA-Trained

Figure 6: The effect of LULA training.

$[-10, 10]^n$. For LA and LULA, each prediction is done via MC-integration with 100 samples. For the evaluation of each dataset, we use a 60-20-20 train-validation-test split. We repeat each train-test process 10 times and take the average.

In Table 4 we report the average predictive standard deviation for each dataset. Note that this metric is the direct generalization of the 1D uncertainty estimates in Fig. 1 to multi-dimension. The test outliers are sampled uniformly from $[-10, 10]^n$. Note that since the inlier data are centered around the origin and have unit variance, they lie approximately in a Euclidean ball with a radius of 2. Therefore, these outliers are far away from them. Thus, naturally, high uncertainty values over these outliers are desirable. Uncertainties over the test sets are generally low for all methods, although LULA has slightly higher uncertainties compared to the base LA. However, LULA yield much higher uncertainties over outliers across all datasets, significantly more than the baselines. Moreover, in Table 5, we show that LULA maintains the predictive performance of the base LA. Altogether, they imply that LULA can detect outliers better than other methods without costing the predictive performance.

C.3 IMAGE CLASSIFICATION

To complement Fig. 5, we present the ECE and Brier score results on CIFAR-10-C in Fig. 7. As observed in the main text, LULA consistently improves the base LA. Furthermore, LULA is competitive to the state-of-the-art DE, especially in higher severity levels.

We furthermore present the detailed results on OOD detection in terms of MMC, FPR95 (Tables 6 and 7), and additionally area under ROC (AUROC) and precision-recall (AUPRC) curves (Tables 8 and 9). We use standard datasets: EMNIST, KMNIST, FMNIST, and LSUN. Furthermore, we use the following artificial datasets:

- GRAYCIFAR10: obtained by converting CIFAR-10 test data into grayscale images.
- UNIFORMNOISE: obtained by uniformly sampling from the hypercube $[0, 1]^n$.
- SMOOTHEDNOISE: obtained by permuting, blurring, and contrast re-scaling the original test images [Hein et al., 2019].
- FMNIST3D: obtained by converting the grayscale FMNIST images into 3-channel images.

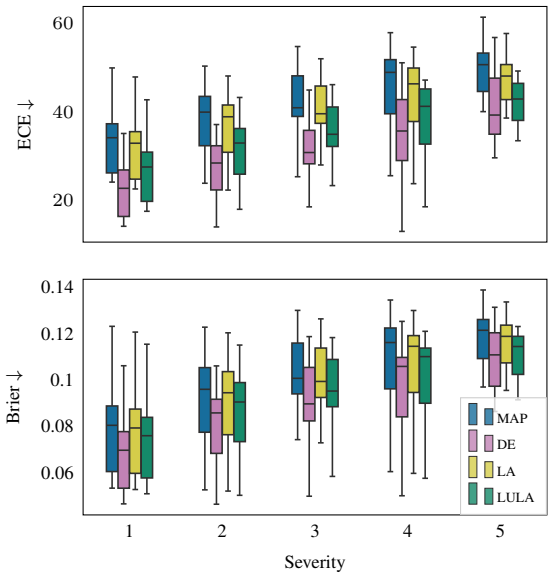


Figure 7: Summarized ECE and Brier score at each severity level of the CIFAR-10-C dataset.

We observe that LULA consistently improves the base LA. Especially, LULA makes the confidence estimates over OOD data lower without introducing underconfidence on in-distribution data.

Table 6: Detailed MMC results. Values are averages over five prediction runs.

Dataset	MAP	MAP-Temp	DE	DE-Temp	LA	LA-LULA	LLLA	LLLA-LULA	OE	OE-LULA
MNIST	99.8	99.8±0.0	99.7	99.8±0.0	99.7±0.0	98.3±0.0	99.3±0.0	99.2±0.0	99.4±0.0	71.5±0.4
EMNIST	84.6	86.2±0.0	82.8	87.4±0.0	84.1±0.0	56.6±0.5	73.7±0.2	67.4±0.3	81.2±0.0	31.7±0.2
KMNIST	71.3	73.8±0.0	67.8	76.1±0.0	70.5±0.0	34.7±0.4	56.4±0.3	45.6±0.5	66.5±0.0	23.7±0.1
FMNIST	76.7	79.0±0.0	69.6	80.1±0.0	75.7±0.0	37.6±0.9	57.3±0.3	50.3±1.2	32.6±0.0	22.4±0.1
GrayCIFAR10	68.2	71.0±0.0	55.4	66.7±0.0	66.9±0.0	32.4±0.6	46.2±0.2	42.5±0.7	10.2±0.0	22.1±0.2
UniformNoise	82.0	83.7±0.1	67.4	94.6±0.1	75.7±0.4	29.4±0.7	36.0±0.9	39.6±1.2	10.1±0.0	21.7±0.7
Noise	99.4	99.7±0.0	99.5	99.9±0.0	99.4±0.0	85.6±1.5	96.4±0.2	95.9±0.6	10.4±0.0	14.2±0.1
SVHN	98.5	97.1±0.0	98.1	97.5±0.0	98.5±0.0	97.5±0.0	91.8±0.5	95.9±0.1	98.4±0.0	98.4±0.0
CIFAR10	72.5	62.4±0.0	58.7	58.1±0.0	71.8±0.0	60.8±0.1	48.5±0.2	52.3±0.4	10.7±0.0	13.3±0.2
LSUN	73.7	63.9±0.0	59.0	59.6±0.0	73.0±0.0	61.5±0.2	48.2±0.3	52.5±0.5	10.3±0.0	12.8±0.3
CIFAR100	73.4	63.5±0.0	60.0	59.6±0.0	72.7±0.0	61.6±0.1	48.9±0.2	52.9±0.4	11.3±0.0	14.0±0.3
FMNIST3D	74.6	64.8±0.0	64.1	61.4±0.0	74.0±0.0	65.2±0.2	53.3±0.4	57.6±0.4	10.6±0.0	13.7±0.2
UniformNoise	79.1	70.8±0.1	54.6	63.8±0.2	77.8±0.2	62.5±0.6	43.9±0.5	51.6±0.3	10.0±0.0	12.4±0.3
Noise	64.2	55.1±0.2	53.3	51.7±0.2	63.5±0.2	53.6±0.1	41.3±0.2	45.8±0.4	55.3±0.1	54.3±0.1
CIFAR10	97.2	94.8±0.0	96.1	95.7±0.0	96.9±0.0	96.2±0.0	90.6±0.0	83.4±0.2	97.3±0.0	97.0±0.0
SVHN	70.6	57.2±0.0	57.2	52.6±0.0	67.7±0.1	63.2±0.3	42.1±0.5	35.0±0.5	56.1±0.0	53.3±0.1
LSUN	74.8	61.5±0.0	65.6	61.8±0.0	73.4±0.0	68.7±0.2	51.3±0.3	40.5±0.4	66.2±0.0	64.4±0.1
CIFAR100	78.7	67.1±0.0	71.2	68.3±0.0	77.3±0.0	73.4±0.1	56.6±0.1	46.6±0.2	78.1±0.0	76.6±0.0
FMNIST3D	68.8	53.7±0.0	60.7	54.7±0.0	66.5±0.1	61.2±0.1	40.4±0.4	32.9±0.3	61.4±0.0	59.2±0.0
UniformNoise	88.0	71.5±0.1	89.3	82.2±0.0	79.5±0.6	62.6±1.6	30.7±0.6	25.2±0.2	10.1±0.0	12.2±0.1
Noise	64.5	52.2±0.2	53.7	52.6±0.1	59.6±0.2	53.8±0.3	35.5±0.4	30.3±0.4	48.6±0.3	46.4±0.2
CIFAR100	85.7	76.8±0.0	81.5	80.6±0.0	80.4±0.0	72.6±0.1	75.7±0.1	63.8±0.2	86.5±0.0	81.2±0.1
SVHN	61.3	42.0±0.0	47.5	42.2±0.0	52.9±0.1	40.7±0.5	46.8±0.6	33.0±0.8	63.7±0.0	54.6±0.2
LSUN	64.9	47.8±0.0	51.7	49.3±0.0	56.0±0.2	46.1±0.1	49.1±0.4	37.5±0.8	58.4±0.0	50.8±0.3
CIFAR10	67.2	51.8±0.0	56.1	54.4±0.0	58.9±0.1	49.8±0.1	52.6±0.1	41.6±0.2	68.8±0.0	59.7±0.1
FMNIST3D	56.4	35.7±0.0	45.8	39.2±0.0	49.0±0.1	40.1±0.3	42.7±0.2	32.6±0.3	53.9±0.0	46.2±0.2
UniformNoise	68.3	56.5±0.1	29.5	43.7±0.1	45.3±0.5	33.0±0.8	36.5±0.9	24.7±0.8	1.7±0.0	1.7±0.0
Noise	68.7	55.3±0.2	50.5	50.2±0.3	58.1±0.2	36.3±1.0	51.2±1.0	29.0±1.4	64.5±0.2	53.8±0.4

Table 7: Detailed FPR95 results. Values are averages over five prediction runs.

Dataset	MAP	MAP-Temp	DE	DE-Temp	LA	LA-LULA	LLLA	LLLA-LULA	OE	OE-LULA
MNIST	-	-	-	-	-	-	-	-	-	-
EMNIST	23.9	24.0±0.0	22.3	22.4±0.0	23.9±0.0	23.6±0.2	24.0±0.2	23.5±0.1	27.5±0.0	23.5±0.6
KMNIST	2.4	2.4±0.0	1.8	2.3±0.0	2.4±0.0	0.8±0.0	1.8±0.2	1.0±0.1	5.1±0.0	3.6±0.4
FMNIST	2.4	2.4±0.0	1.1	1.8±0.0	2.3±0.0	0.8±0.0	1.5±0.1	0.9±0.1	0.2±0.0	1.8±0.2
GrayCIFAR10	0.1	0.0±0.0	0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.8±0.2
UniformNoise	1.1	1.0±0.0	0.0	0.2±0.0	0.3±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.9±0.5
Noise	0.1	0.1±0.0	0.2	0.2±0.1	0.1±0.0	9.4±3.0	7.6±2.2	1.4±0.3	0.0±0.0	0.2±0.2
SVHN	-	-	-	-	-	-	-	-	-	-
CIFAR10	24.0	23.2±0.0	11.3	14.1±0.0	23.8±0.1	20.7±0.2	23.7±1.8	19.6±0.4	0.0±0.0	0.0±0.0
LSUN	25.7	25.3±0.0	11.0	16.3±0.0	25.5±0.2	21.3±0.5	22.2±2.2	19.7±0.8	0.0±0.0	0.0±0.0
CIFAR100	25.5	24.8±0.0	13.3	16.8±0.0	25.3±0.1	21.9±0.2	24.3±1.6	20.5±0.4	0.2±0.0	0.1±0.0
FMNIST3D	29.7	28.9±0.0	22.5	22.5±0.0	29.8±0.1	29.4±0.3	33.0±1.6	29.2±0.3	0.0±0.0	0.0±0.0
UniformNoise	33.2	34.1±0.3	5.4	18.4±0.3	31.7±0.3	19.5±1.1	14.0±2.1	15.6±0.6	0.0±0.0	0.0±0.0
Noise	17.5	17.0±0.5	7.8	10.0±0.5	17.1±0.5	13.6±0.3	14.7±1.7	12.1±0.7	10.3±0.1	10.2±0.2
CIFAR10	-	-	-	-	-	-	-	-	-	-
SVHN	41.7	35.4±0.0	25.0	20.1±0.0	38.9±0.2	37.6±0.5	19.6±0.8	20.4±1.3	22.8±0.0	20.9±0.1
LSUN	50.7	45.7±0.0	45.3	39.3±0.0	50.9±0.2	48.9±0.5	41.9±0.2	37.4±1.1	38.3±0.0	38.4±0.2
CIFAR100	60.1	55.9±0.0	54.6	51.7±0.0	59.7±0.3	58.7±0.2	51.4±0.4	50.4±0.4	58.0±0.0	57.5±0.2
FMNIST3D	40.4	31.3±0.0	35.2	27.0±0.0	39.0±0.3	36.3±0.3	19.1±0.6	17.4±0.4	30.0±0.0	29.2±0.2
UniformNoise	89.0	81.6±0.5	99.9	99.3±0.1	73.9±1.7	31.3±5.9	0.1±0.1	0.7±0.3	0.0±0.0	0.0±0.0
Noise	36.6	31.8±0.5	25.7	31.6±0.2	28.9±0.4	24.3±0.7	9.9±0.7	11.0±1.0	15.4±0.2	14.0±0.3
CIFAR100	-	-	-	-	-	-	-	-	-	-
SVHN	73.8	67.9±0.0	62.1	58.2±0.0	73.3±0.3	68.8±0.6	72.4±0.9	67.5±0.9	75.9±0.0	74.1±0.3
LSUN	81.7	81.7±0.0	73.0	75.3±0.0	82.4±0.6	82.1±0.4	81.7±0.6	81.0±0.8	69.7±0.0	71.0±0.8
CIFAR10	83.0	81.5±0.0	77.2	78.2±0.0	82.9±0.2	82.8±0.3	82.3±0.2	82.7±0.2	82.4±0.0	81.5±0.2
FMNIST3D	70.2	59.5±0.0	64.3	58.8±0.0	70.6±0.2	69.1±0.6	69.1±0.3	67.8±0.9	63.1±0.0	62.9±0.4
UniformNoise	97.7	100.0±0.0	15.7	99.5±0.1	89.1±1.1	71.2±4.3	76.9±3.1	57.5±5.6	0.0±0.0	0.0±0.0
Noise	74.2	72.0±0.5	63.3	63.8±0.4	71.6±0.5	57.1±1.8	69.9±0.5	54.3±1.6	66.6±0.3	61.6±0.8

Table 8: Detailed AUROC results. Values are averages over five prediction runs.

Dataset	MAP	MAP-Temp	DE	DE-Temp	LA	LA-LULA	LLLA	LLLA-LULA	OE	OE-LULA
MNIST	-	-	-	-	-	-	-	-	-	-
EMNIST	89.5	89.5±0.0	89.8	89.6±0.0	89.5±0.0	90.6±0.2	89.6±0.1	90.3±0.0	92.9±0.0	93.2±0.1
KMNIST	98.9	98.9±0.0	99.1	98.9±0.0	98.9±0.0	99.5±0.0	99.3±0.0	99.5±0.0	98.6±0.0	98.9±0.1
FMNIST	98.8	98.8±0.0	99.2	99.0±0.0	98.9±0.0	99.3±0.0	99.3±0.0	99.3±0.1	99.7±0.0	99.2±0.0
GrayCIFAR10	99.7	99.6±0.0	99.8	99.8±0.0	99.7±0.0	99.6±0.0	99.8±0.0	99.7±0.0	100.0±0.0	99.3±0.0
UniformNoise	99.1	99.2±0.0	99.8	99.1±0.0	99.5±0.0	99.8±0.0	100.0±0.0	99.8±0.0	100.0±0.0	99.4±0.1
Noise	97.4	97.3±0.0	96.9	96.8±0.0	97.4±0.0	96.3±0.3	96.7±0.1	96.7±0.1	100.0±0.0	99.9±0.0
SVHN	-	-	-	-	-	-	-	-	-	-
CIFAR10	95.2	95.3±0.0	97.7	97.2±0.0	95.3±0.0	96.2±0.1	95.5±0.3	96.5±0.0	100.0±0.0	100.0±0.0
LSUN	94.9	94.9±0.0	97.9	96.9±0.0	94.9±0.0	96.0±0.1	95.8±0.3	96.5±0.1	100.0±0.0	100.0±0.0
CIFAR100	94.6	94.6±0.0	97.2	96.5±0.0	94.7±0.0	95.8±0.0	95.3±0.3	96.2±0.0	100.0±0.0	100.0±0.0
FMNIST3D	94.2	94.3±0.0	96.2	96.0±0.0	94.2±0.0	94.4±0.1	93.0±0.5	94.3±0.1	100.0±0.0	100.0±0.0
UniformNoise	93.8	93.4±0.1	98.5	96.5±0.0	94.1±0.1	96.6±0.2	97.4±0.2	97.3±0.1	100.0±0.0	100.0±0.0
Noise	96.6	96.6±0.1	98.3	97.9±0.1	96.6±0.1	97.4±0.0	97.2±0.3	97.7±0.1	97.9±0.1	97.9±0.1
CIFAR10	-	-	-	-	-	-	-	-	-	-
SVHN	94.6	95.3±0.0	96.6	97.1±0.0	94.9±0.0	95.0±0.1	96.9±0.1	96.6±0.2	97.0±0.0	97.2±0.0
LSUN	92.5	93.5±0.0	93.7	94.3±0.0	92.5±0.0	92.8±0.1	93.2±0.1	93.9±0.2	94.9±0.0	94.9±0.0
CIFAR100	90.0	90.6±0.0	91.1	91.6±0.0	90.1±0.0	90.1±0.0	90.2±0.1	90.0±0.1	90.1±0.0	90.2±0.0
FMNIST3D	94.7	95.8±0.0	95.3	96.3±0.0	94.9±0.0	95.3±0.0	97.0±0.1	97.2±0.1	95.9±0.0	96.0±0.0
UniformNoise	91.5	92.6±0.0	88.6	91.0±0.0	93.6±0.1	96.2±0.3	99.4±0.1	99.3±0.0	100.0±0.0	100.0±0.0
Noise	95.2	95.7±0.1	96.6	95.9±0.1	96.0±0.1	96.7±0.1	98.1±0.1	98.0±0.1	97.1±0.1	97.4±0.1
CIFAR100	-	-	-	-	-	-	-	-	-	-
SVHN	80.2	83.9±0.0	85.0	86.7±0.0	80.5±0.1	83.5±0.4	80.7±0.4	84.1±0.7	80.1±0.0	80.2±0.2
LSUN	78.1	80.1±0.0	82.5	82.7±0.0	78.5±0.2	79.1±0.1	79.4±0.4	79.8±0.9	83.7±0.0	83.2±0.2
CIFAR10	75.4	76.4±0.0	78.7	78.6±0.0	75.5±0.1	75.4±0.2	75.8±0.1	75.3±0.2	75.4±0.0	75.8±0.0
FMNIST3D	84.1	88.2±0.0	86.6	89.1±0.0	83.7±0.1	84.1±0.2	84.3±0.2	84.5±0.3	86.4±0.0	86.2±0.1
UniformNoise	78.7	75.5±0.1	96.8	88.5±0.0	88.1±0.4	91.2±0.6	90.4±0.6	93.0±0.6	100.0±0.0	100.0±0.0
Noise	69.3	71.5±0.2	80.9	78.1±0.2	74.3±0.3	86.2±0.9	75.5±0.9	87.0±1.4	75.1±0.2	78.5±0.4

Table 9: Detailed AUPRC results. Values are averages over five prediction runs.

Dataset	MAP	MAP-Temp	DE	DE-Temp	LA	LA-LULA	LLLA	LLLA-LULA	OE	OE-LULA
MNIST	-	-	-	-	-	-	-	-	-	-
EMNIST	67.3	67.2±0.0	67.0	66.5±0.0	67.3±0.2	69.7±0.7	67.9±0.4	69.1±0.3	84.5±0.0	81.0±0.3
KMNIST	97.9	97.9±0.0	98.4	98.0±0.0	98.0±0.0	99.5±0.0	99.2±0.1	99.4±0.0	98.6±0.0	99.0±0.0
FMNIST	98.3	98.4±0.0	98.9	98.5±0.0	98.4±0.0	99.3±0.0	99.1±0.0	99.3±0.1	99.7±0.0	99.3±0.0
GrayCIFAR10	99.7	99.7±0.0	99.9	99.8±0.0	99.7±0.0	99.7±0.0	99.8±0.0	99.7±0.0	100.0±0.0	99.4±0.0
UniformNoise	99.8	99.8±0.0	100.0	99.8±0.0	99.9±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.9±0.0
Noise	99.5	99.4±0.0	99.4	99.3±0.0	99.5±0.0	99.2±0.1	99.3±0.0	99.3±0.0	100.0±0.0	100.0±0.0
SVHN	-	-	-	-	-	-	-	-	-	-
CIFAR10	97.8	97.8±0.0	99.1	98.8±0.0	97.8±0.0	98.3±0.0	98.1±0.2	98.5±0.0	100.0±0.0	100.0±0.0
LSUN	99.9	99.9±0.0	100.0	100.0±0.0	99.9±0.0	99.9±0.0	99.9±0.0	100.0±0.0	100.0±0.0	100.0±0.0
CIFAR100	97.4	97.3±0.0	98.8	98.4±0.0	97.4±0.0	98.0±0.0	97.9±0.2	98.3±0.0	100.0±0.0	100.0±0.0
FMNIST3D	97.4	97.4±0.0	98.5	98.3±0.0	97.3±0.0	97.5±0.0	96.8±0.3	97.4±0.0	100.0±0.0	100.0±0.0
UniformNoise	99.4	99.3±0.0	99.9	99.7±0.0	99.4±0.0	99.7±0.0	99.8±0.0	99.8±0.0	100.0±0.0	100.0±0.0
Noise	99.7	99.7±0.0	99.8	99.8±0.0	99.7±0.0	99.8±0.0	99.8±0.0	99.8±0.0	99.8±0.0	99.8±0.0
CIFAR10	-	-	-	-	-	-	-	-	-	-
SVHN	91.5	92.2±0.0	94.5	95.0±0.0	91.9±0.0	92.0±0.1	94.3±0.2	93.9±0.3	94.3±0.0	94.6±0.0
LSUN	99.7	99.7±0.0	99.7	99.8±0.0	99.7±0.0	99.7±0.0	99.7±0.0	99.7±0.0	99.8±0.0	99.8±0.0
CIFAR100	90.3	90.6±0.0	91.2	91.6±0.0	90.3±0.0	90.3±0.0	89.6±0.1	89.1±0.1	90.3±0.0	90.3±0.0
FMNIST3D	95.3	96.1±0.0	95.7	96.6±0.0	95.5±0.0	95.7±0.0	97.0±0.1	97.2±0.1	96.1±0.0	96.2±0.0
UniformNoise	98.1	98.4±0.0	97.5	98.1±0.0	98.6±0.0	99.2±0.1	99.9±0.0	99.8±0.0	100.0±0.0	100.0±0.0
Noise	98.8	98.9±0.0	99.1	99.0±0.0	99.0±0.0	99.2±0.0	99.5±0.0	99.5±0.0	99.2±0.0	99.3±0.0
CIFAR100	-	-	-	-	-	-	-	-	-	-
SVHN	67.8	72.3±0.0	73.1	75.4±0.0	67.4±0.2	71.7±1.0	66.5±0.9	71.9±1.6	69.4±0.0	68.3±0.3
LSUN	99.0	99.1±0.0	99.2	99.2±0.0	99.0±0.0	99.0±0.0	99.0±0.0	99.0±0.1	99.3±0.0	99.2±0.0
CIFAR10	74.7	75.2±0.0	77.8	77.6±0.0	74.4±0.1	73.8±0.2	74.6±0.3	73.1±0.4	75.3±0.0	75.3±0.1
FMNIST3D	85.0	88.5±0.0	87.5	89.6±0.0	84.3±0.1	84.2±0.2	84.5±0.2	84.3±0.3	87.1±0.0	86.6±0.1
UniformNoise	94.7	94.0±0.0	99.3	97.4±0.0	97.2±0.1	98.0±0.2	97.8±0.2	98.4±0.1	100.0±0.0	100.0±0.0
Noise	90.2	91.0±0.1	94.2	92.8±0.1	92.2±0.1	96.2±0.3	92.5±0.4	96.4±0.5	92.1±0.1	93.3±0.2