

Beyond In-Domain Scenarios: Robust Density-Aware Calibration

Christian Tomani* ^{1 2},

Futa Waseda* ³,

Yuesong Shen^{1 2},

Daniel Cremers^{1 2}

* Equal contribution

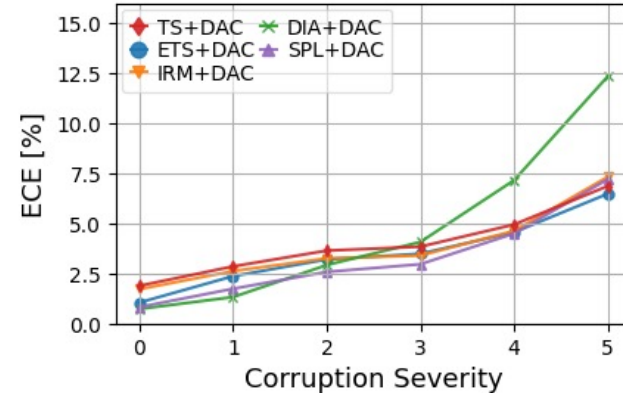
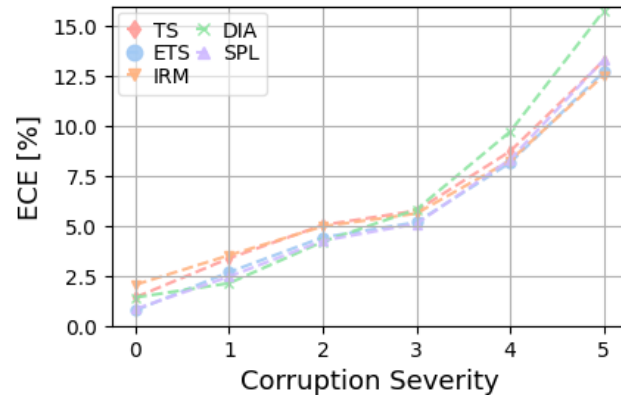
¹ Technical University of Munich

² Munich Center for Machine Learning

³ The University of Tokyo

Correspondence to: Christian Tomani <christian.tomani@tum.de>, Futa Waseda <futa-waseda@g.ecc.u-tokyo.ac.jp>

Handling out-of-distribution (OOD) and corrupted data is a necessity for machine learning applications



SOTA calibration methods:

- **Cannot** provide accurate uncertainty scores under **domain-shift** and **OOD scenarios**, although well-calibrated in-domain

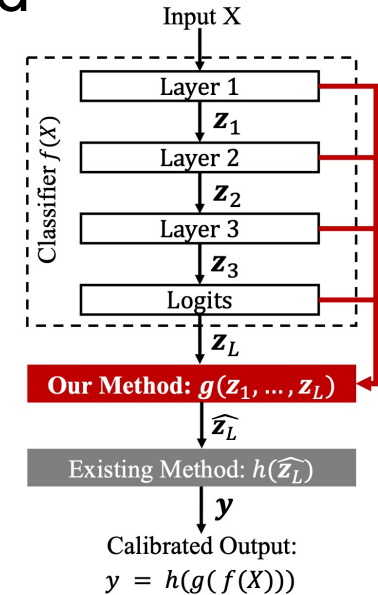


Our approach + SOTA calibration methods:

- **Improve** calibration performance under **domain-shift** and **OOD scenarios**.

DAC is a Density-Aware Calibration method

- **DAC can be combined with any existing post-hoc method:**
 - More robust and reliable uncertainty estimates with DAC.
- **DAC utilizes hidden layers:**
 - Unlike existing post-hoc methods, DAC utilizes the hidden layers of classifiers as a source of uncertainty-related information.
- **DAC is a post-hoc method:**
 - DAC does not require any training of the underlying classifier.

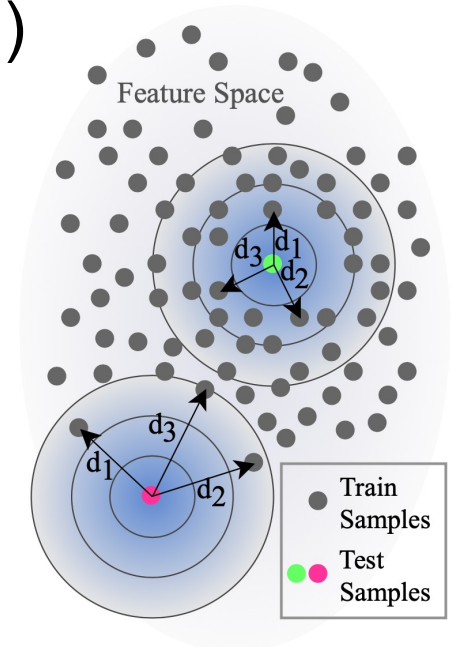


➤ DAC is **fully backbone agnostic** and can directly be added to **any** existing post-hoc calibration pipeline.

- ✓ **Any DNN-based classifier**
- ✓ **Any post-calibration pipeline to combine**

DAC is based on k-nearest-neighbors (KNN)

- **DAC utilizes non-parametric density estimation through k-nearest-neighbor (KNN)** based on **feature embeddings** extracted from the classifier.
- High-density regions => tend to have higher prediction confidence.
- Low-density regions => lack sufficient nearby training samples, leading to unreliable predictions.
- **By leveraging information from feature embeddings** across the entire classifier, DAC is capable of identifying corrupted or OOD samples.



● High Density → Low Uncertainty
 ● Low Density → High Uncertainty

➤ DAC is **density aware** and **domain agnostic** due to KNN.

DAC uses linear regression to aggregate density information from different layers

Density Proxy is used for scaling:

$$\hat{Q}(\mathbf{x}, w) = \sigma_{SM}(\mathbf{z}_L / S(\mathbf{x}, w))$$

Calibrated output distribution Softmax (Logits / Temperature)

where

$$S(\mathbf{x}, w) = \sum_{l=1}^L w_l \boxed{s_l} + w_0$$

Temperature Linear regression

density proxy

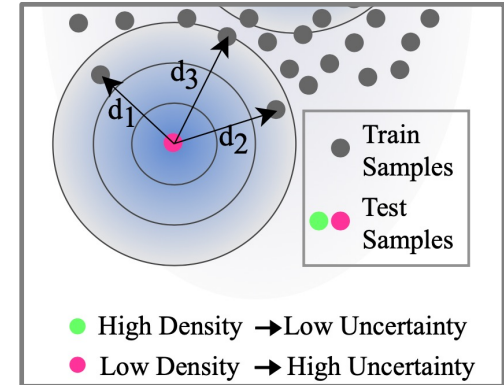
Mean squared error (MSE) loss:

$$L_w = \sum_{c=1}^C (I_c - \sigma_{SM}(\mathbf{z}_L / S(\mathbf{x}, w))^{(c)})^2$$

Loss for linear regression $I_c \dots$ Binary indicator for correct class
 $C \dots$ Number of classes

Accuracy preserving &
Density-aware scaling of logits

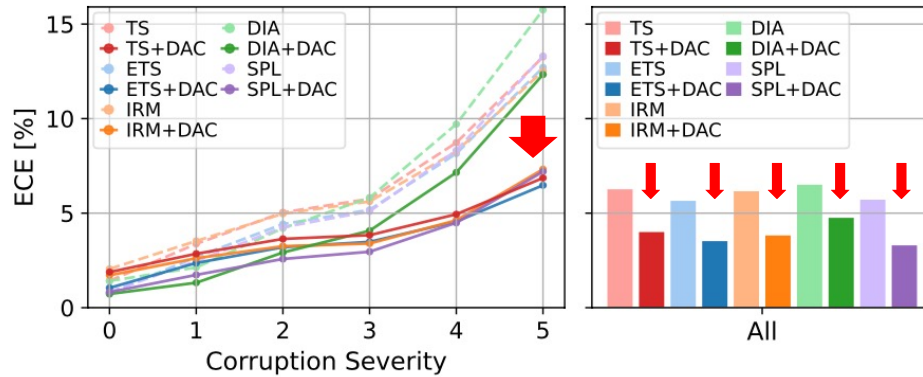
Feature space:



$S = d_{(k)} \dots$ Distance between a test sample to nearby training samples in feature space via KNN.
 The larger $d_{(k)}$, the lower the density $\Rightarrow$ higher uncertainty.

Optimized with a validation set.

Calibration performance can be improved for various state-of-the-art post-hoc methods when using DAC.



ImageNet-BiT-M

Delta of ECE [%] when DAC is added

	IND	SEV. 5	ALL	
TS	+0.47	-6.40	-2.26	36%
ETS	+0.28	-6.23	-2.13	38%
IRM	-0.34	-5.20	-2.34	38%
DIA	-0.69	-3.42	-1.75	27%
SPL	-0.01	-6.11	-2.41	42%

D.) IMAGENET BiT-M

- ✓ Boosts domain-shift calibration performance
- ✓ Maintains in-domain calibration performance

➤ DAC **boosts** the robustness of calibration performance in **domain-shift scenarios**, while **maintaining in-domain** predictive uncertainty estimates

Conclusion

- DAC can be combined with **existing post-hoc methods** and **boosts** calibration performance in **domain-shift scenarios**, while maintaining the in-domain calibration performance.
- DAC is a **density aware**, as well as **domain and backbone agnostic** method that can be applied to a wide range of classifiers, including Transformers.
- **Hidden layers** in classifiers **carry valuable information** for accurately predicting uncertainty estimates.