Supplementary materials for EFOMS: Effective, Efficient, and Environmentally Friendly Out-of-Model-Scope Detection Methodology

ANONYMOUS AUTHORS, Institution, Anonymous

ACM Reference Format:
Anonymous authors. 2018. Supplementary materials for EFOMS: Effective, Efficient, and Environmentally Friendly Out-of-Model-Scope Detection Methodology. *Proc. ACM Meas. Anal. Comput. Syst.* 37, 4, Article 111 (August 2018), 12 pages. https://doi.org/XXXXXXX.XXXXXXX

In this document, we provide supplementary materials for our submission entitled "*EFOMS: Effective, Efficient, and Environmentally Friendly Out-of-Model-Scope Detection Methodology*". First, we illustrate the image examples for CIFAR-10 and Tiny-ImageNet datasets considered in our evaluation. Then, we present additional empirical results that justify selecting $tc_{OMS} = 1$ or $tc_{OMS} = 2$ as the out-of-model-scope detection thresholds. Finally, we provide an additional theoretical explanation for the Sobol indices.

1 Image examples and image classes for the CIFAR-10 dataset

The image classes of the CIFAR-10 dataset are displayed in Tbl. 1, and some image examples for illustrating the CIFAR-10 and its OMS datasets are displayed in Fig. 1. The correspondence for the images in each column exists only for the first three rows.

Table 1. CIFAR-10 Image Classes

Label	Description	Label	Description	Label	Description
0	Airplane	1	Automobile	2	Bird
3	Cat	4	Deer	5	Dog
6	Frog	7	Horse	8	Ship
9	Truck				

2 Image examples for the Tiny-Imagenet dataset

A few image examples for illustrating the Tiny-Imagenet and the Tiny-Imagenet-C are shown in Fig. 2, and there is no correspondence between the images displayed in each column.

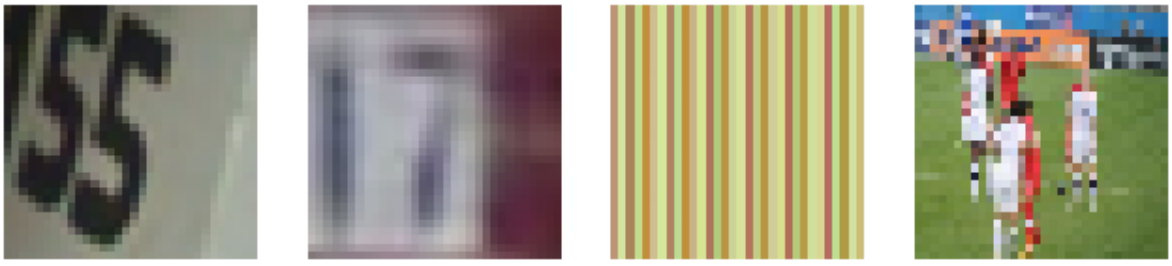
Author's Contact Information: Anonymous authors, Institution, Anonymous, Anonymous.

ACM 2476-1249/2018/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

(a) CIFAR-10 image examples

(b) CIFAR-10-C image examples, the displayed transformations are motion blur s3, fog s4, frost s2 and gaussian noise s5

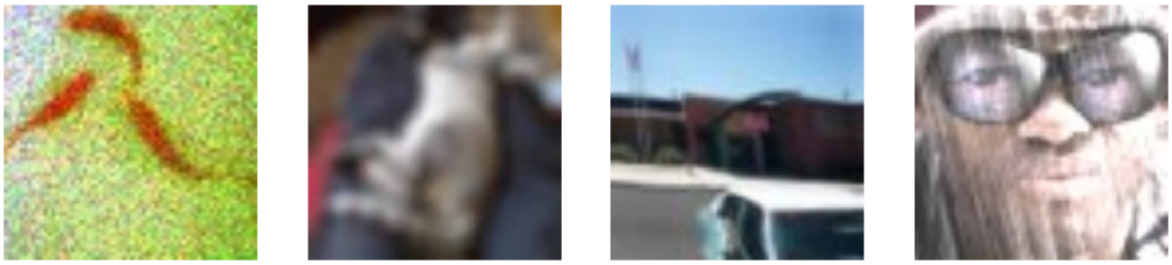
(c) Adversarial attack examples generated on ResNet34, the displayed attack order is FGSM, DeepFool, Carlini&Wagner and JSMA



(d) Novelty image examples, they are respectively from SVHN, SVHN, DTD, and Places365.

Fig. 1. Image examples of CIFAR-10 and its OMS datasets

(a) Tiny-Imagenet image examples



(b) Tiny-Imagenet-C image examples, the displayed transformations are shot noise s3, defocus blur s4, elastic transform s2 and snow s5

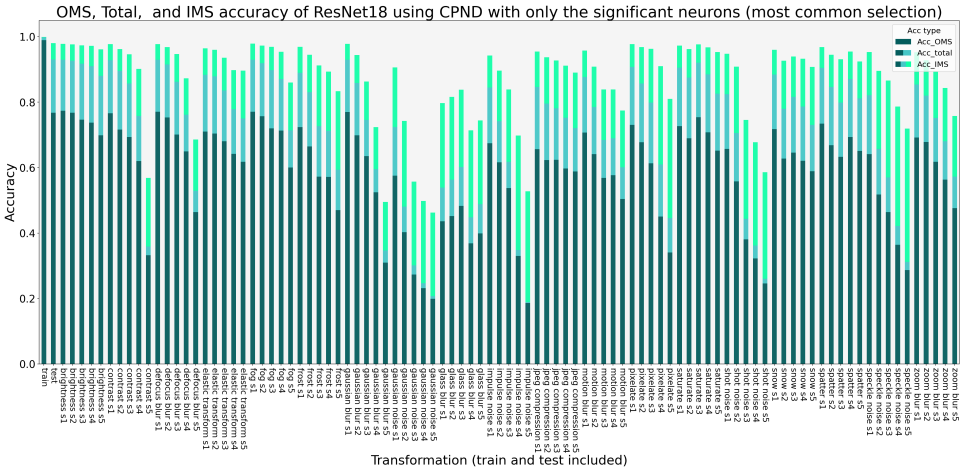
Fig. 2. Image examples of Tiny-Imagenet and Tiny-Imagenet-C

3 Additional accuracy results for the CIFAR-10 dataset using CPND with $tc_{OMS} = 1$

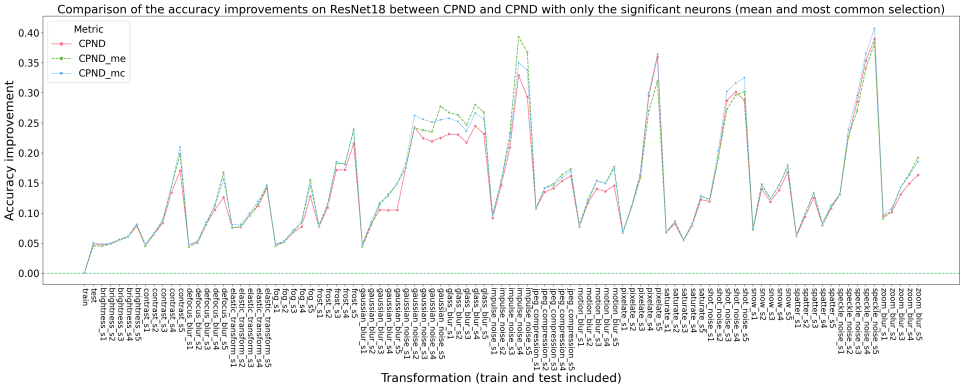
In Fig. 3, 4 and 5, We display the additional accuracy-related results on the distribution shift dataset of CIFAR10 (i.e., CIFAR-10-C) using CPND with $tc_{OMS} = 1$. These supplementary results help to justify the choice of tc_{OMS} described in the paper. More precisely, if we take the ResNet50 as the example, the results shown in 5b justify that we can still observe a significant accuracy improvement for various distribution-shift transformations. The performance in the best cases is even better compared to the one obtained with $tc_{OMS} = 2$ (i.e., 40% compared to 30% observed in Fig. 7). Hence, we can justify that $tc_{OMS} = 1$ is a good choice for the threshold coefficient and more robust than $tc_{OMS} = 2$ since it applies more strict criteria for being In-Model-Scope (IMS).

4 Additional accuracy results for the CIFAR-10 dataset using CPND and deep-KNN with $tc_{OMS} = 2$

In Fig. 6-9, We display the additional accuracy-related results on the distribution shift dataset of CIFAR10 (i.e., CIFAR-10-C) using CPND and deep-KNN with $tc_{OMS} = 2$. These results demonstrate that the application of EFOMS methodology in CPND and deep-KNN still maintains the OMS detection performance for various ResNet architectures on the distribution-shift dataset of CIFAR-10. For example, in Fig. 6a, the Acc_{IMS} is constantly higher than Acc_{total} for the presented transformations, and the accuracy improvements in Fig. 6b are all above 0, and have remarkable performance on some specific transformations (e.g., $CPND_{me}$ and $CPND_{mc}$ provide an accuracy improvement around 30%, which is even better than the original method $CPND$). The results in Fig. 6a and 6b justify that the application of the EFOMS methodology maintains the OMS detection effectiveness of the original method for ResNet18.



(a) OMS, Total, and IMS Accuracy using $CPND_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet18

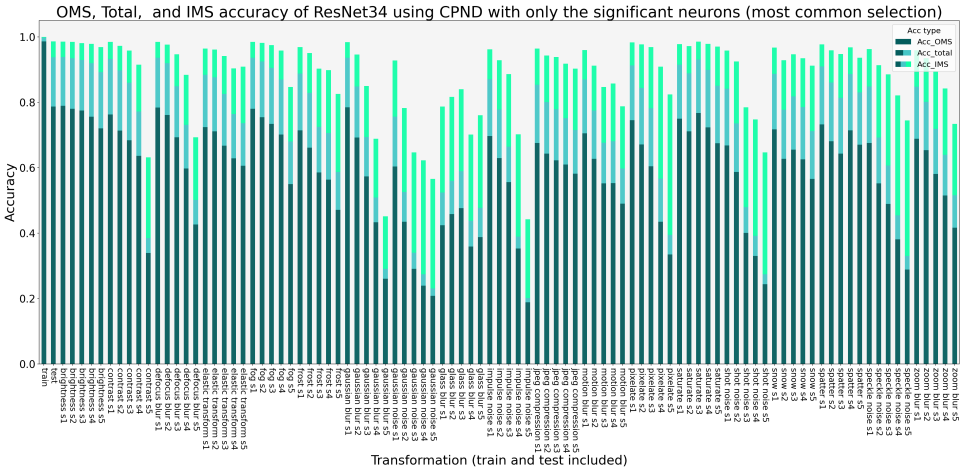


(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for CPND (i.e., $CPND$, $CPND_{me}$ and $CPND_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet18

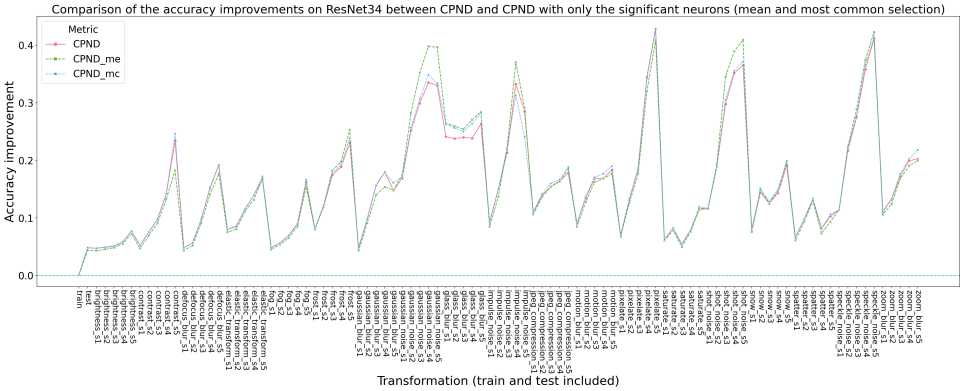
Fig. 3. ResNet18 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for CPND with $tc_{OMS} = 1$

5 Additional accuracy results for the Swin-b-1024 using deep-KNN with $tc_{OMS} = 2$

In Fig. 10, We display the complete accuracy improvement results on the Tiny-Imagenet-C for Swin-b-1024 using deep-KNN with $tc_{OMS} = 2$. The results provide more detailed information for the effectiveness of EFOMS on Swin-b-1024 mentioned in RQ3. More precisely, the accuracy improvements obtained with $deep - KNN_{me}$ for different severities of the contrast transformation or the glass blur are very similar to the ones of the original method $deep - KNN$, and the performance of $deep - KNN_{me}$ on other transformations is almost the same as $deep - KNN$.



(a) OMS, Total, and IMS Accuracy using $CPND_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet34

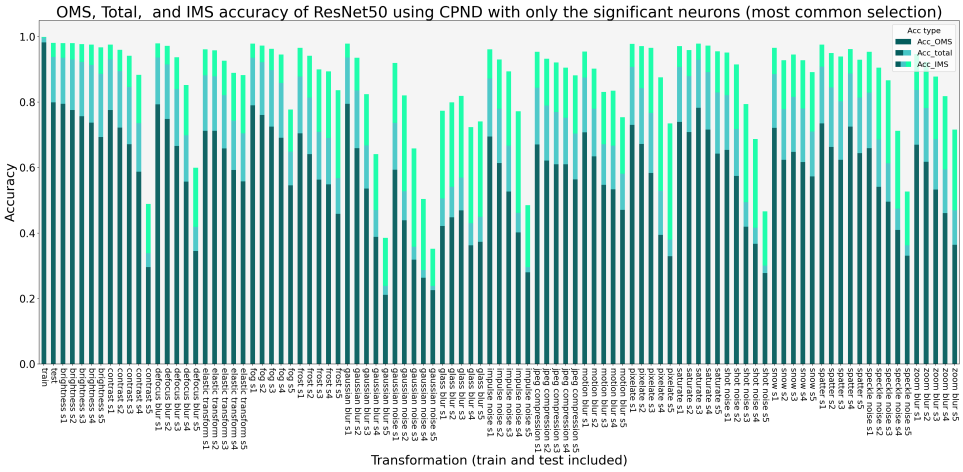


(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for CPND (i.e., $CPND$, $CPND_{me}$ and $CPND_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet34

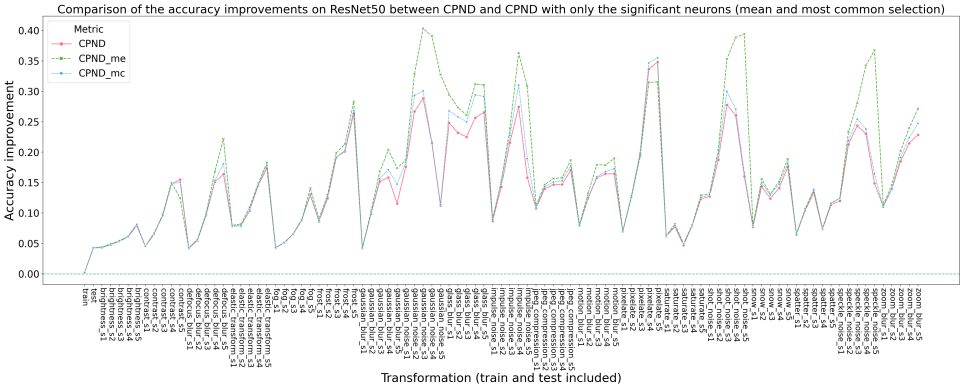
Fig. 4. ResNet34 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for CPND with $tc_{OMS} = 1$

6 More theoretical details about the Sobol indices

In the paper, we mentioned that the Hoeffding decomposition holds if the input variables are independent. The decomposition can still remain true in general cases if the input variables are not independent but with some assumptions on the joint distribution [1]. This point can help justify the success of our experiments in identifying the significant neurons.

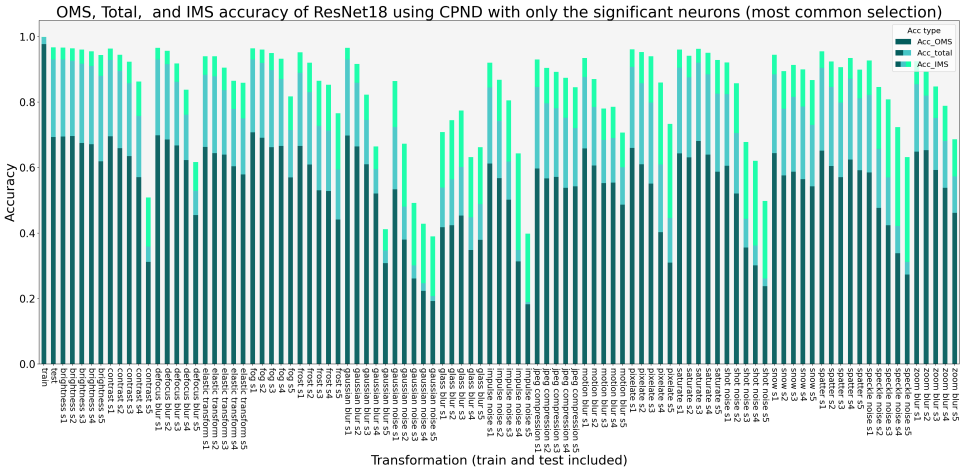


(a) OMS, Total, and IMS Accuracy using $CPND_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet50

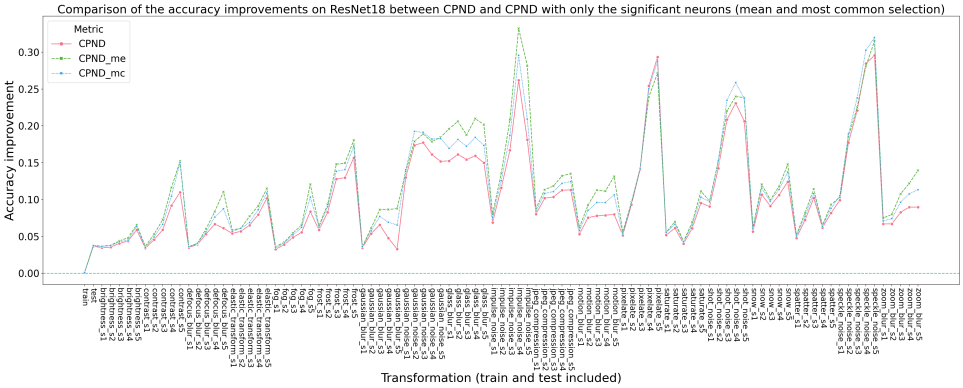


(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for CPND (i.e., $CPND$, $CPND_{me}$ and $CPND_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet50

Fig. 5. ResNet50 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for CPND with $tc_{OMS} = 1$

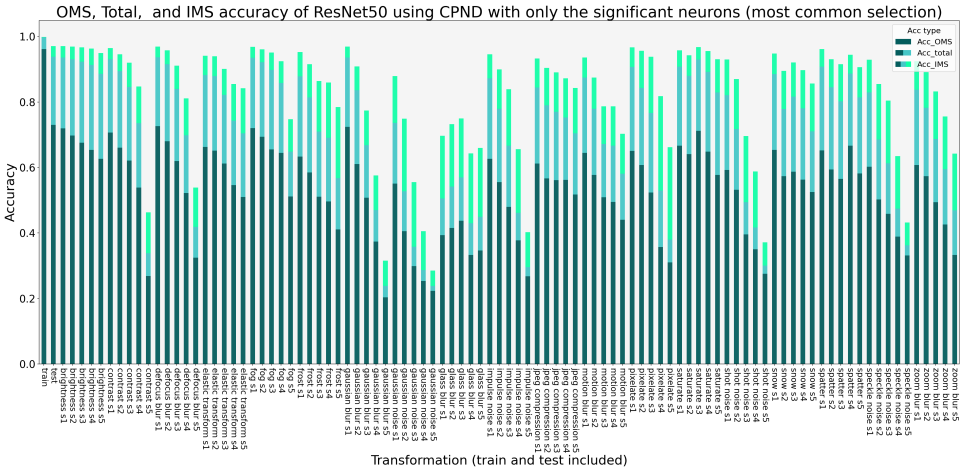


(a) OMS, Total, and IMS Accuracy using $CPND_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet18

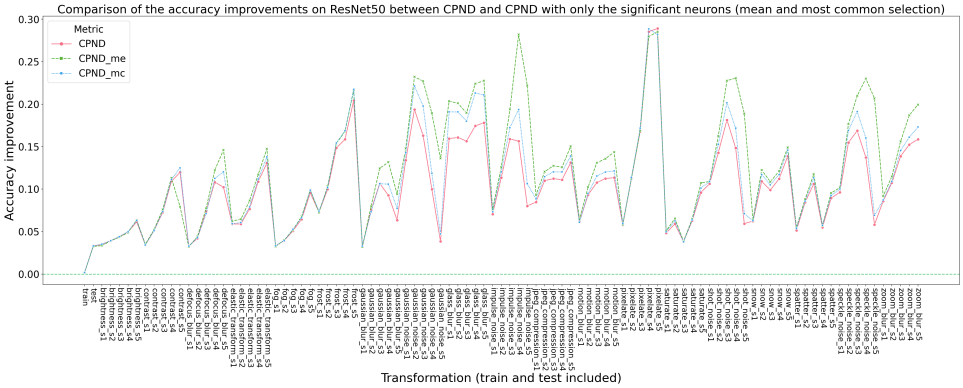


(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for CPND (i.e., $CPND$, $CPND_{me}$ and $CPND_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet18

Fig. 6. ResNet18 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for CPND with $tc_{OMS} = 2$

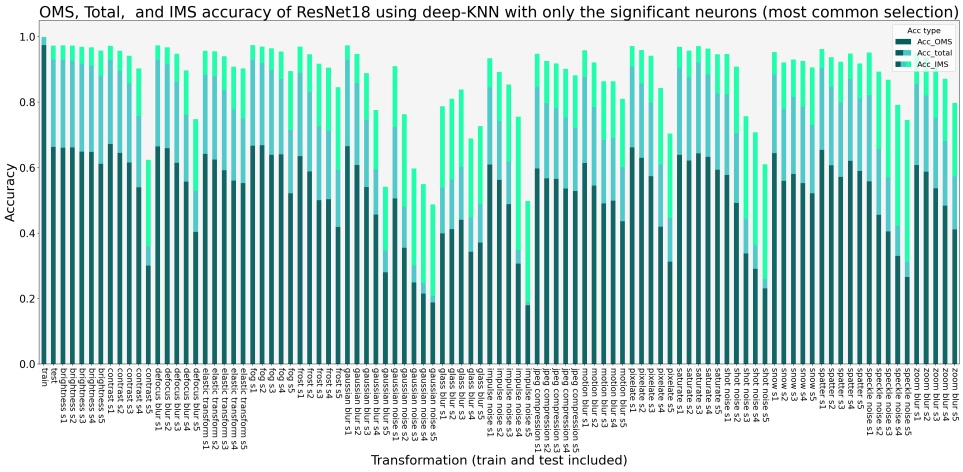


(a) OMS, Total, and IMS Accuracy using $CPND_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet50

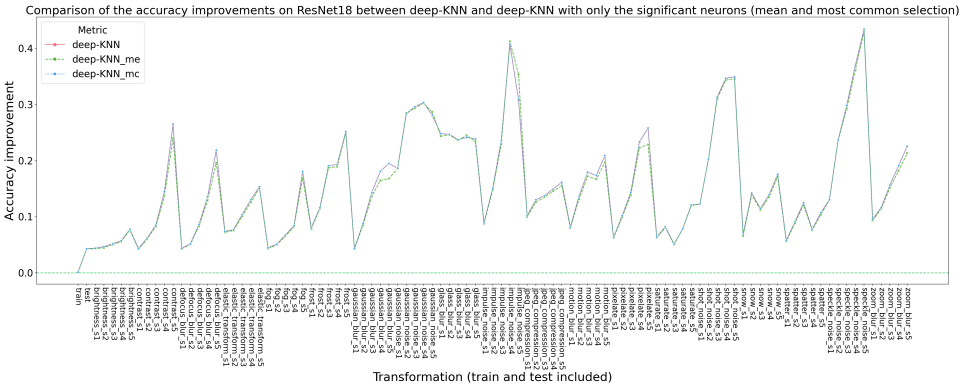


(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for CPND (i.e., $CPND$, $CPND_{me}$ and $CPND_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet50

Fig. 7. ResNet50 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for CPND with $tc_{OMS} = 2$

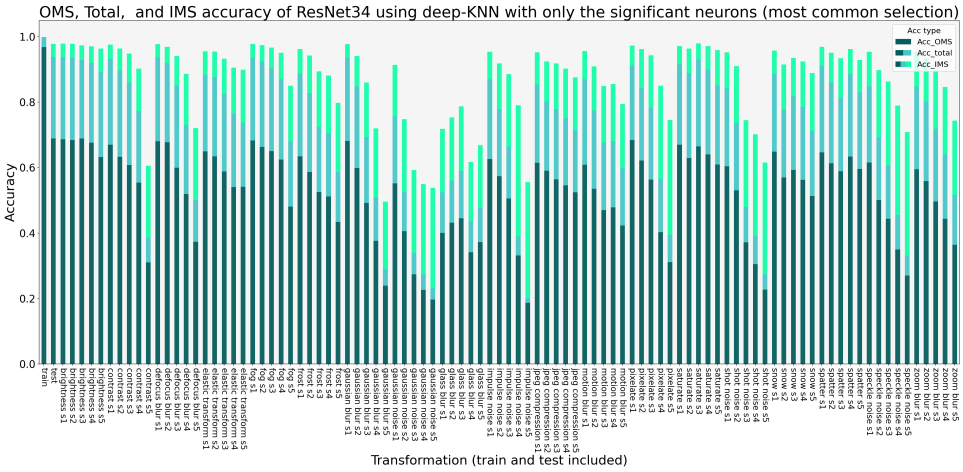


(a) OMS, Total, and IMS Accuracy using $deep - KNN_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet18

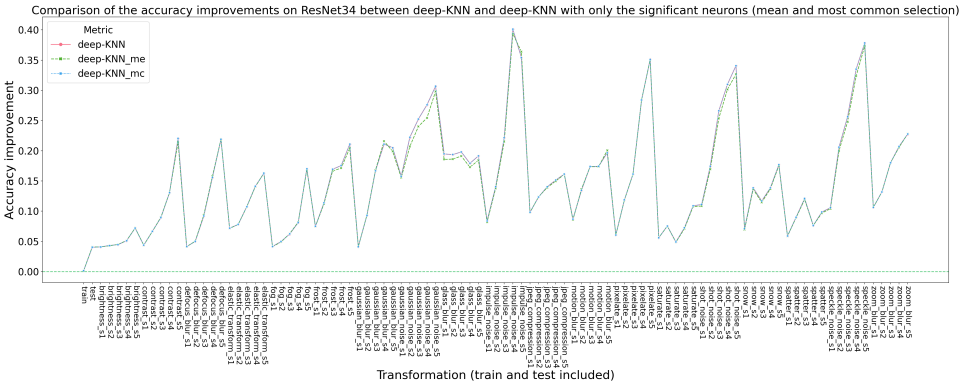


(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for deep-KNN (i.e., $deep - KNN$, $deep - KNN_{me}$ and $deep - KNN_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet18

Fig. 8. ResNet18 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for deep-KNN with $tc_{OMS} = 2$



(a) OMS, Total, and IMS Accuracy using $deep - KNN_{mc}$ on the CIFAR-10 and CIFAR-10-C datasets for ResNet34



(b) Comparison of accuracy improvements Acc_{imp} between all the tested methods for deep-KNN (i.e., $deep - KNN$, $deep - KNN_{me}$ and $deep - KNN_{mc}$) on the CIFAR-10 and CIFAR-10-C datasets for ResNet34

Fig. 9. ResNet34 Accuracy results on the CIFAR-10 and CIFAR-10-C datasets for deep-KNN with $tc_{OMS} = 2$

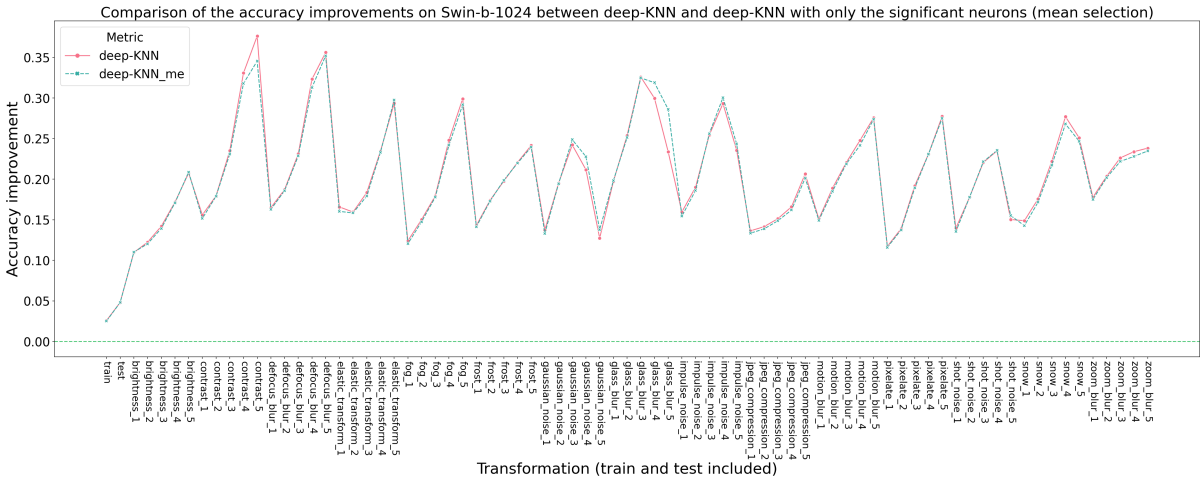


Fig. 10. Comparison of accuracy improvements Acc_{imp} between the tested methods of deep-KNN (i.e., $deep - KNN$ and $deep - KNN_{me}$) on the Tiny-Imagenet and Tiny-Imagenet-C datasets for Swin-b-1024

References

[1] Gaëlle Chastaing, Fabrice Gamboa, and Clémentine Prieur. 2012. Generalized Hoeffding-Sobol Decomposition for Dependent Variables -Application to Sensitivity Analysis. arXiv:1112.1788 [math.ST]