Reconstruct and Match: Out-of-Distribution Robustness via Topological Homogeneity
(Supplementary Materials)

Chaoqi Chen
Shenzhen University
Shenzhen, China
cqchen1994@gmail.com

Luyao Tang
Xiamen University
Xiamen, China
lytang@stu.xmu.edu.cn

Hui Huang[*]
Shenzhen University
Shenzhen, China
hhzhiyan@gmail.com

A   Additional Results

A.1   Results on transferability

To assess the contributions of different components of REMA in terms of transferability, we calculate the Pearson correlation coefficient (ρ) between TETOT/Entropy and the ground truth at each corruption level. A higher negative correlation implies higher transferability of the model, which is trained on the source CIFAR10 and tested on the target CIFAR10-C.

TETOT [1] is a novel and effective metric that estimates the distribution difference between the source and unknown domains by quantifying the distance between them using optimal transport. TETOT consists of two distance measures: the dissimilarity between the features of the source and target domains, and the difference between the probability distributions in the predicted label space of the two distributions. TETOT has been validated on multiple cross-domain datasets for its excellent performance in representing model transferability. Entropy (computed as $\frac{1}{n} \sum_{x \in \mathcal{D}_T} H(h(g(x)))$, where $H(\hat{y}) = -\sum_{c \in \mathcal{Y}} p(\hat{y}_c) \log(p(\hat{y}_c))$) is a popular indicator that solely relies on the predicted probabilities of the target domain data and is widely used to evaluate the degree of domain shift.

The transferability coefficients for corruption levels 1-5 are shown in Figure 1 to Figure 5. Two notable trends are observed: (1) Compared to ERM, REMA (Full) generally exhibits higher transferability in both TETOT and Entropy evaluation systems. (2) The relationship between Entropy and Accuracy is not entirely linear, and there are more outliers for Entropy compared to TETOT, indicating that it is a less robust indicator.

A.2   Results on visualization

We further visualize the feature distributions under fog, frost, and pixelate corruption. As shown in Figure 6, Figure 7 and Figure 8, SSR and HORR have specific effects on each type of corruption. They not only increase the inter-class distance to avoid confusion between semantically similar samples but also reduce the inter-class distance, resulting in tighter clusters and enhanced discriminative power of the model. The ability of REMA to improve model generalization is unaffected by the type and severity of OOD, indicating the importance and necessity of higher-order semantic dependencies and abstracting dense pixels into slots.

---

[*]Corresponding author.

B   Additional Discussion

B.1   Limitations

The primary limitation of our paper lies in its compatibility with tasks other than object recognition. For scenes lacking clear foreground-background distinctions, such as natural landscape images, sparse modeling based on slots may struggle to accurately separate the scenes into several main components. For such data with specific styles or attributes, it may be necessary to collect relevant annotations to train robust models.

B.2   Broader Impacts

We summarize two potential broader impacts of our work as follows.

Advancing Robust Deep Learning in Dynamic Environments: By addressing the challenge of out-of-distribution (OOD) generalization in deep learning models, this research contributes to the advancement of robust artificial intelligence systems. The proposed Reconstruct and Match (REMA) framework not only enhances model performance in non-stationary environments but also lays the groundwork for future developments in transfer learning and domain adaptation, fostering the deployment of more reliable AI systems across various real-world applications.

Fostering Understanding of Human-like Object Recognition: Through the inspiration drawn from human object recognition processes, this study offers insights into the fundamental principles underlying visual cognition. By emulating the way humans decompose objects into major components and identify their topological relations, REMA opens avenues for interdisciplinary collaboration between cognitive science and artificial intelligence, potentially leading to novel approaches in both fields and promoting a deeper understanding of human perception and learning.

B.3   Future Works

We provide three potential directions for future research.

Refinement of Topological Representation: Future studies could delve deeper into refining the topological representations of objects within the REMA framework. This could involve exploring alternative methods for selective slot-based reconstruction or investigating the integration of additional information sources to enhance the fidelity of object component identification. By refining the representation of object topology, researchers can further improve the capabilities of deep learning models in capturing the intrinsic structure of complex objects.

Extension to Multimodal Learning: Another promising direction for future work is the extension of REMA to multimodal learning scenarios. Integrating information from multiple modalities such as images, text, and audio could enrich the understanding of object semantics and enhance the robustness of models across diverse data distributions. Investigating how REMA can effectively leverage multimodal information to improve object recognition performance would be an intriguing avenue for research.

Application to Real-World Domains: Beyond standard benchmark datasets, future research could focus on applying REMA to real-world object recognition tasks in domains such as robotics, autonomous driving, and augmented reality. Evaluating the effectiveness of REMA in practical scenarios with complex environmental conditions and diverse object categories would provide valuable insights into its scalability and generalization capabilities. Additionally, exploring potential adaptations of REMA for specific application domains could lead to the development of tailored solutions with enhanced performance and applicability.

References

[1] Akshay Mehra, Yunbei Zhang, and Jihun Hamm. Test-time assessment of a model's performance on unseen domains via optimal transport. *arXiv preprint arXiv:2405.01451*, 2024.
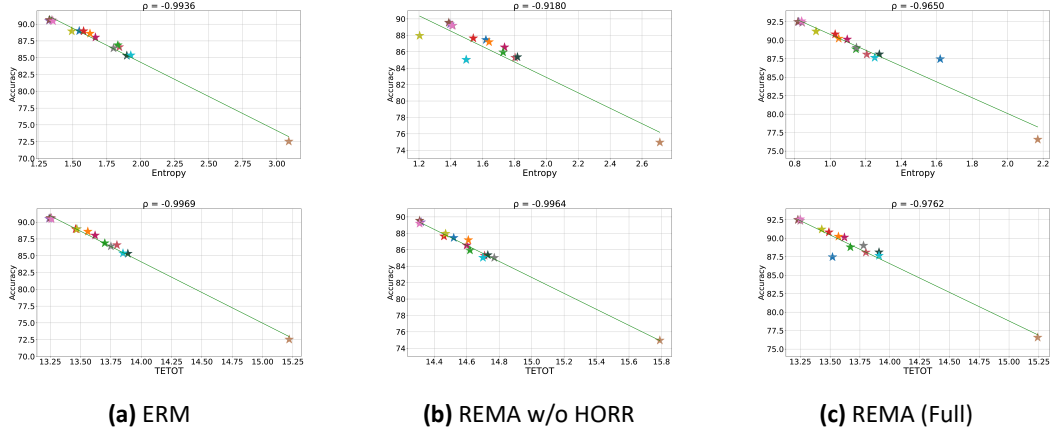
Figure 1: The transferability of TETOT and Entropy at level-1 corruption of the CIFAR10-C dataset.
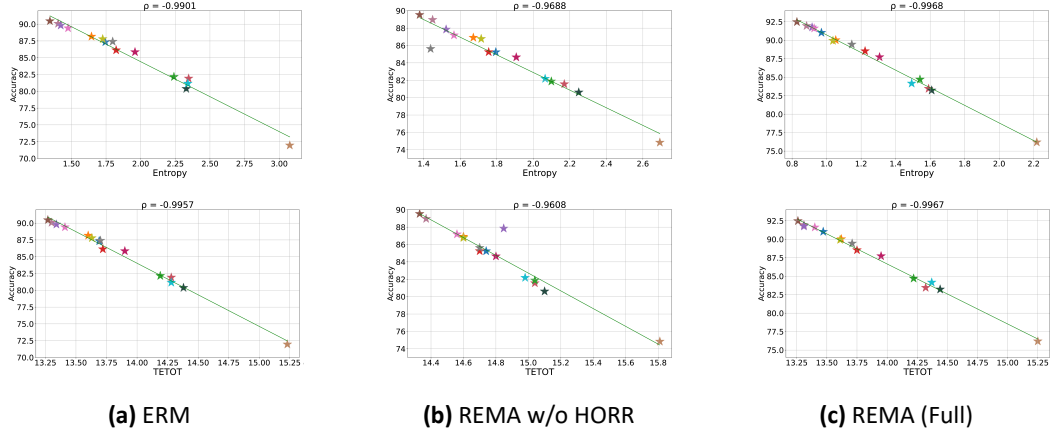

Figure 2: The transferability of TETOT and Entropy at level-2 corruption of the CIFAR10-C dataset.
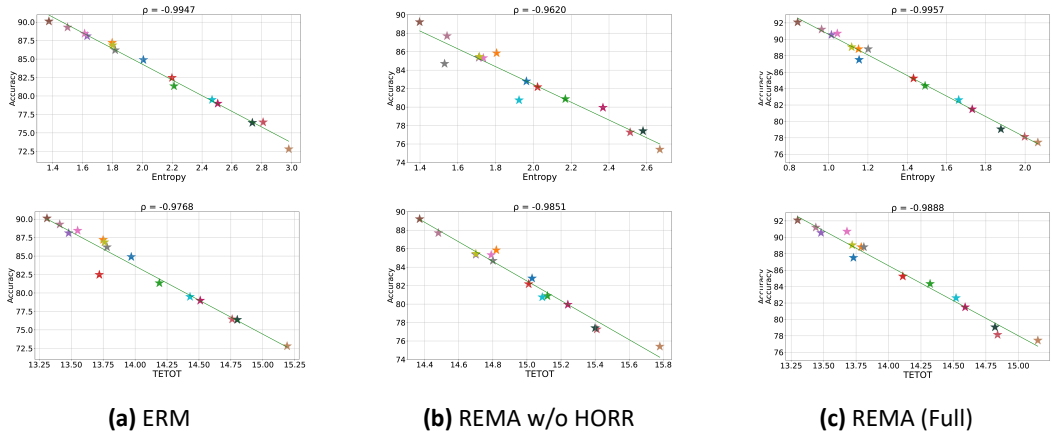

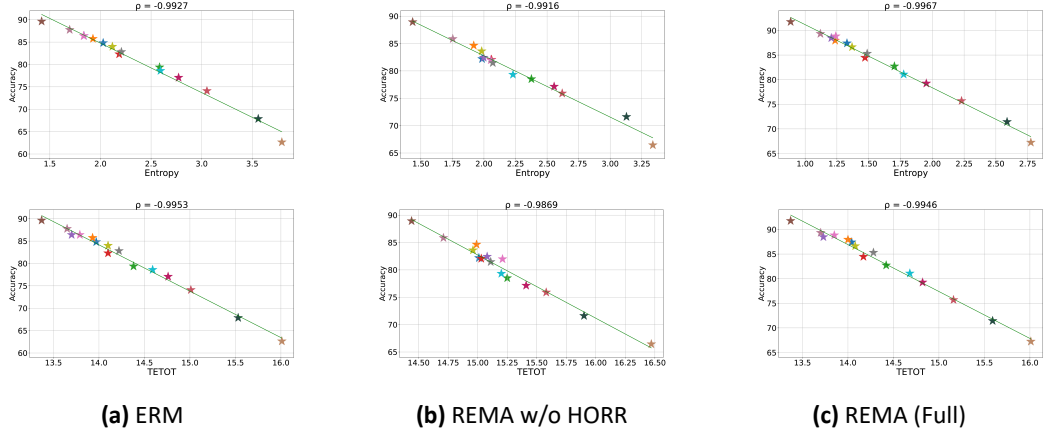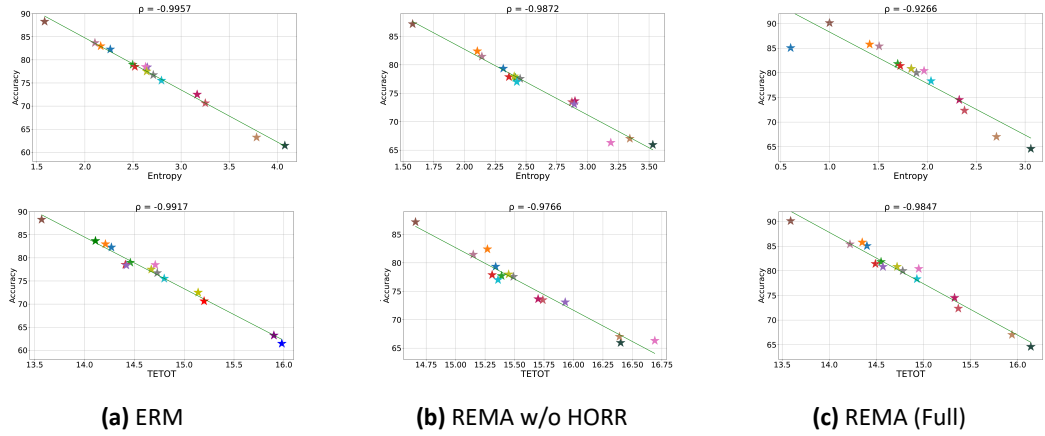Figure 3: The transferability of TETOT and Entropy at level-3 corruption of the CIFAR10-C dataset.

(a) ERM   **(b)** REMA w/o HORR   **(c)** REMA (Full)

Figure 4: The transferability of TETOT and Entropy at level-4 corruption of the CIFAR10-C dataset.

(a) ERM   **(b)** REMA w/o HORR   **(c)** REMA (Full)

Figure 5: The transferability of TETOT and Entropy at level-5 corruption of the CIFAR10-C dataset.

(a) ERM   **(b)** REMA w/o SSR   **(c)** REMA w/o HORR   **(d)** REMA (Full)
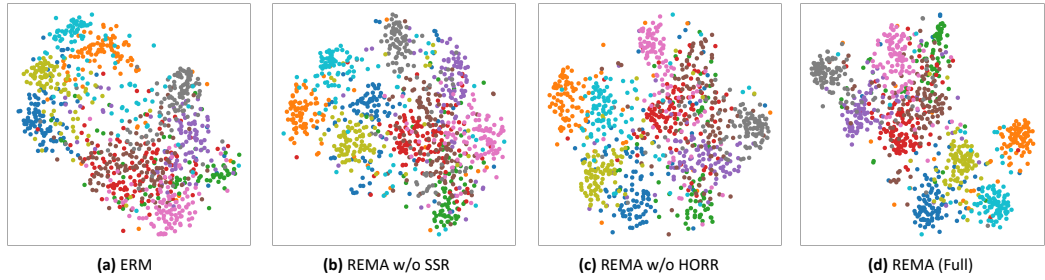
Figure 6: Feature visualization of different methods using t-SNE at level-5 fog corruption of the CIFAR10-C dataset.
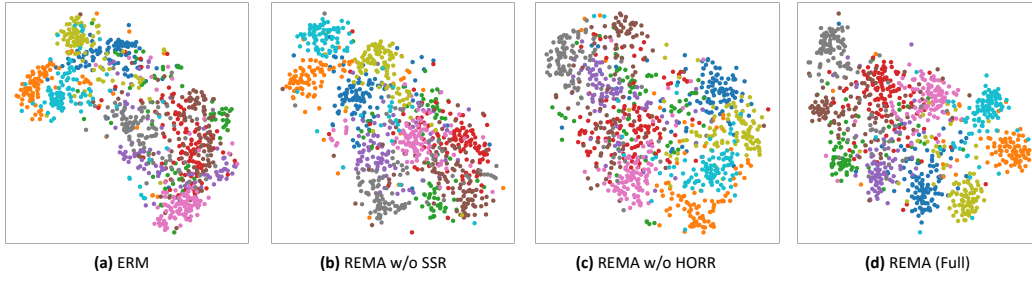
Figure 7: Feature visualization of different methods using t-SNE at level-5 frost corruption of the CIFAR10-C dataset.
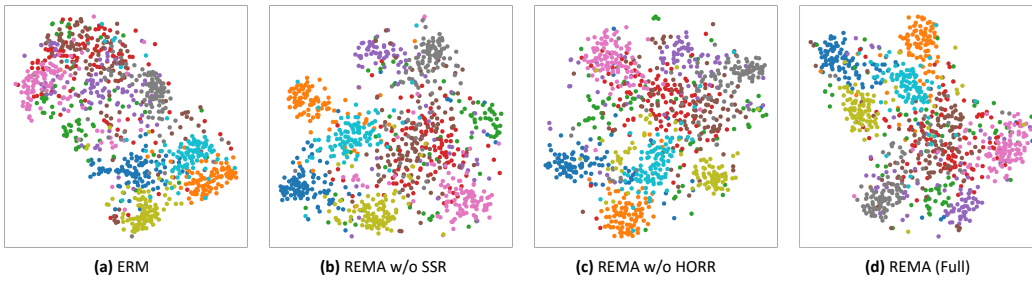

Figure 8: Feature visualization of different methods using t-SNE at level-5 pixelate corruption of the CIFAR10-C dataset.