The 40th Annual AAAI Conference on Artificial Intelligence

ShapBPT: Image Feature Attributions using Data-Aware Binary Partition Trees

Muhammad Rashid [1], Elvio G. Amparore [1], Enrico Ferrari [2], Damiano Verda [2]

1 UNIVERSITÀ DI TORINO https://www.unito.it/

2 RuleX https://www.rulex.ai/

EU HORIZON-KDT-JU-2023-2-RIA

Chips JU EUROPEAN PARTNERSHIP

DistriMuSe https://distrimuse.eu/

We introduce **ShapBPT**, a novel **explainable computer vision** (XCV) method that assigns **Shapley** values over a multiscale **B**inary **P**artition **T**ree. By using image-adaptive hierarchical regions, ShapBPT produces semantically meaningful attributions while reducing computational cost. Experiments show improved explanation quality and stronger human preference compared to axis-aligned Shapley methods.

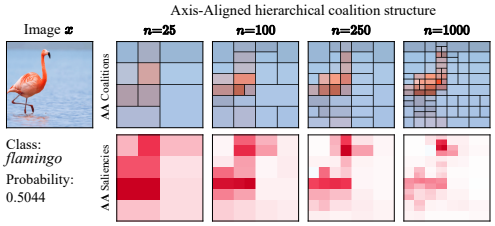
Shap: Shapley image feature attributions

The recursive Shapley/Owen formulation for explainable computer vision exploits a hierarchical partition of the image to decompose feature attributions across scales. By applying the Owen value recursively on an arbitrary hierarchical coalition structure (HCS), Shapley values are efficiently propagated from coarse regions to finer subregions while preserving axiomatic guarantees. However, the result strongly depends on the chosen HCS. Popular choices of HCS are **Axis-Aligned** grids (**AA**), which lack data-awareness.

$$\Omega_i(Q,T) = \begin{cases} \frac{1}{2}\Omega_i(Q \cup T_2, T_1) + \frac{1}{2}\Omega_i(Q, T_1) & \\ \quad \text{if } T\downarrow = \{T_1, T_2\}, i \in T_1 & \\ \frac{1}{|T|}\Delta_T(Q) \quad \text{if } T \text{ is indivisible} & \end{cases}$$

Recursive Owen formula.

$$\Delta_i(S) = \nu(S \cup \{i\}) - \nu(S)$$

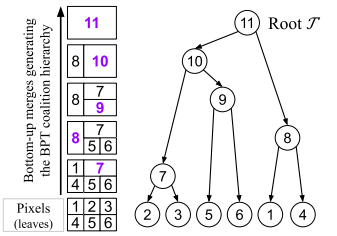


- Relies on predefined or weakly structured regions, often leading to fragmented or semantically incoherent explanations.
- Slow convergence, requiring many iterations to localize relevant regions.
- Region discovery is agnostic to image content, which limits shape fidelity.
- Higher computational cost.

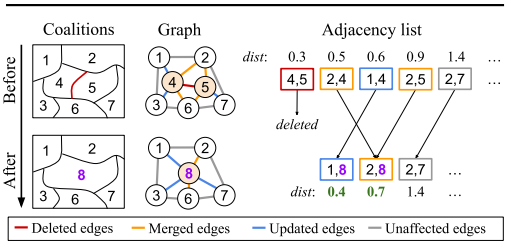
BPT: Binary Partition Tree

The Binary Partition Tree (BPT) defines a hierarchical coalition structure in which image regions form nested, non-overlapping groups across multiple scales.

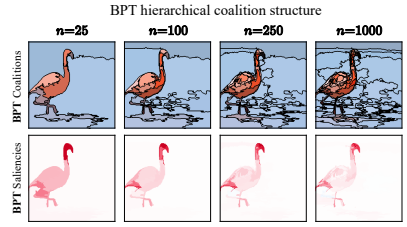
This organization naturally encodes feature dependencies and shared contributions within regions, aligning well with the assumptions of cooperative game theory.



Such a hierarchy integrates well with the Owen formula. As a result, Shapley attributions can be computed efficiently and consistently across scales while preserving fairness and additivity guarantees.



ShapBPT: Shapley coefficients over BPT coalitions



- Content-aware region decomposition via the Binary Partition Tree enables precise and meaningful shape identification.
- Adaptive, multiscale splitting that naturally aligns explanations with image structures. Fast convergence toward salient regions.
- Robust and consistent explanations across diverse datasets, architectures, and tasks.

The ShapBPT approach consists in computing Shapley feature attributions over a data-aware hierarchical partition of the image defined by a Binary Partition Tree, rather than over fixed pixel grids.

Its rationale is that image classifiers rely on structured, multiscale visual patterns, so explanations should follow the intrinsic morphology of the image instead of arbitrary geometric splits.

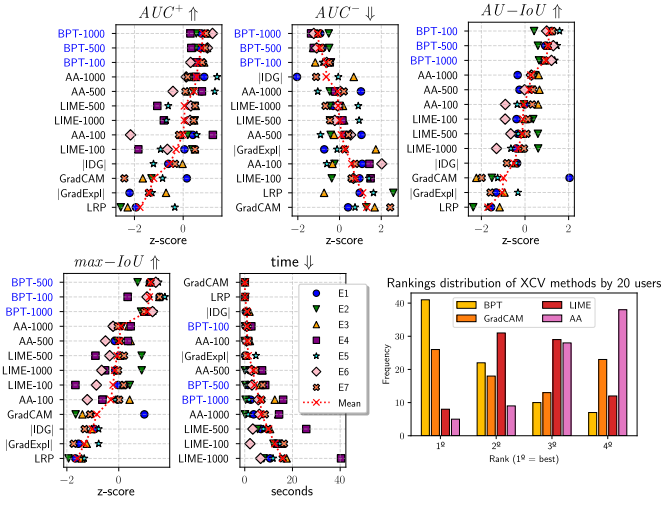
ShapBPT integrates the BPT hierarchy with the Owen formula, recursively propagating Shapley values from coarse, semantically meaningful regions down to finer details in an adaptive manner.

This strategy prioritizes relevant regions early in the explanation process, significantly reducing the number of recursive evaluations needed to localize objects.

As a result, ShapBPT achieves faster convergence, better alignment with object boundaries, and more stable and human-preferred explanations across models and datasets

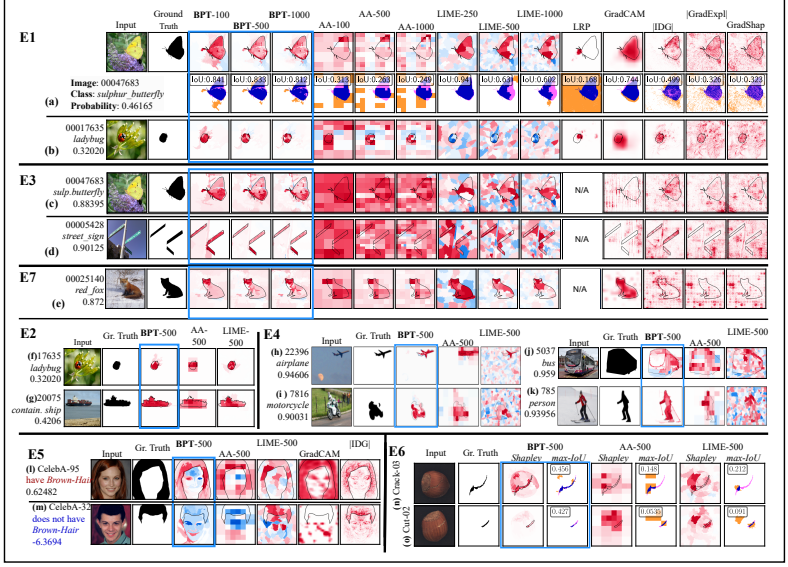
Results on multiple scores and experiments confirm the effectiveness of ShapBPT in producing accurate and faithful explanations

Across response-based and ground-truth-based metrics, ShapBPT consistently outperforms data-agnostic Shapley methods under equal evaluation budgets. indings are further supported by human preference studies, indicating improved interpretability and practical usefulness.



Broad applicability to various computer vision tasks, including image classification, object detection, anomaly localization, and attribute recognition

ShapBPT is model-agnostic and remains effective across CNNs, Vision Transformers, and hybrid architectures. Its data-aware hierarchical design ensures consistent performance across diverse datasets and imaging conditions.



Muhammad Rashid[1] muhammad.rashid@unito.it muhammad.rashid@rulex.ai

Prof. Elvio G. Amparore[1] elviogilberto.amparore@unito.it

Enrico Ferrari[2] enrico.ferrari@rulex.ai

Damiano Verda[2] damiano.verda@rulex.ai

Links

Code: https://github.com/amparore/shap_bpt
Tests: https://github.com/rashidrao-pk/shap_bpt_tests
Tech. Appendix: https://zenodo.org/records/17570695
Python Package: pip install shap-bpt
PyPi: https://pypi.org/project/shap-bpt

ShapBPT code

Test