

Back Home: A Computer Vision Solution to Seashell Identification for Ecological Restoration

Alexander Valverde*
FIFCO

Luis Solano
FIFCO

André Montoya
FIFCO

Abstract

Illegal souvenir collection strips an estimated five tonnes of seashells from Costa Rica’s beaches each year, yet once these specimens are seized their coastal origin—Pacific or Caribbean—cannot be verified easily because the lack of information, preventing their return when are confiscated by the local authorities. To solve this issue, we introduce BackHome19K, the first large-scale image dataset annotated with coast-level labels, and propose a lightweight pipeline that infers provenance in real time on a mobile-grade CPU. A similar PaDiM-inspired anomaly filter pre-screens uploads, increasing robustness to user-generated noise. On a held-out test set the classifier attains 86.3 % balanced accuracy, while the filter rejects 93 % of 180 out-of-domain objects with zero false negatives. Deployed as a web application, the system has already processed 70 000 shells for wildlife officers in under three seconds per image, enabling confiscated specimens to be safely repatriated to their native ecosystems. The dataset is available at [FIFCO/BackHome19K](#).

1. Introduction

Seashells are a keystone element of coastal ecosystems. Beyond providing habitat and shelter, their biological structures regulate sediment chemistry, buffer pH, and contribute to nutrient cycling. Recent materials-science studies [6, 22] have further shown that molluscan shells possess exceptional hardness, fracture toughness, corrosion resistance, and bio activity. Preserving this natural capital is therefore critical for biodiversity, shoreline stability, and sustainable innovation.

Despite their importance, seashell populations along Costa Rica’s beaches have sharply declined owing to unregulated souvenir collection by tourists. Customs officers at Juan Santamaría International Airport routinely confiscate these shells, yet repatriation is impossible because the spec-



Figure 1. Web interface displaying the final output of our pipeline: for each uploaded seashell image the system returns its predicted coastal provenance with a confidence score

imens’ provenance—Pacific or Caribbean coast—cannot be determined post-hoc. Returning them to the wrong habitat risks spreading invasive parasites, disrupting local gene pools, and skewing ecological surveys [24, 30]. Consequently, thousands of shells remain in storage each year instead of being restored to their native ecosystems [1, 41, 42]

This conservation bottleneck presents a classic computer vision challenge: fine-grained classification where visually similar species from different ecosystems must be distinguished based on subtle morphological differences [3, 17]. Traditional taxonomic identification focuses on species-level classification, but conservation applications require ecosystem-level provenance determination that demands specialized datasets and architectures optimized for geographic rather than purely biological distinctions.

The challenge extends beyond traditional fine-grained classification because Pacific and Caribbean shells often share nearly identical morphological features—similar color palettes, growth ring patterns, and overall silhouettes—yet originate from fundamentally different marine ecosystems [8, 14]. This requires our model to learn extremely subtle distinguishing characteristics: minute texture variations, slight hue shifts, and microscopic geometric deviations that reflect different environmental conditions, water chemistry, and ecological pressures.

*Project developed while the main author was working at FIFCO

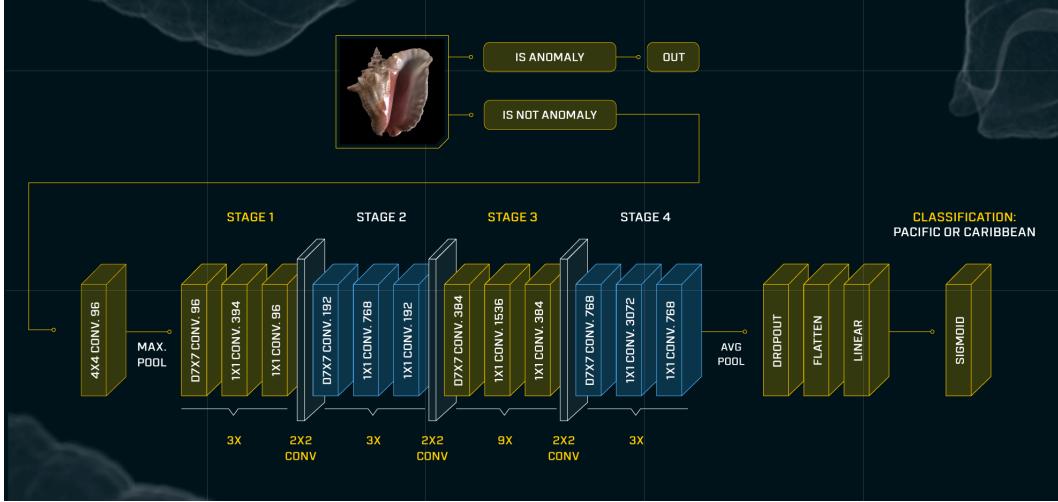


Figure 2. Overview of the proposed two-stage pipeline for seashell origin classification. The system first applies anomaly detection to identify and filter out abnormal specimens, then employs a classification network to predict the geographic origin of authentic seashells. The integrated approach ensures robust performance by eliminating outliers before classification.

We tackle this logistical and conservation bottleneck through an image-based classification model. Specifically, we formulate the task as a binary recognition problem—Pacific versus Caribbean—and introduce a large-scale dataset comprising $\sim 19\,000$ high-resolution photographs containing 516 mollusk species presented in each location. Building on this resource, we design a lightweight convolutional neural network (CNN) tailored to the fine-grained visual cues that distinguish shells from the two coasts.

Our contributions are threefold:

- 1. Dataset.** We release **BackHome19K**, the first annotated image corpus of Costa Rican seashells with coast-level labels, captured *in situ* under natural lighting conditions across 516 species.
- 2. Model.** We propose a compact CNN architecture based on ConvNeXt-Tiny that balances classification accuracy with the low-latency constraints of field deployment.
- 3. System.** We integrate the model into a production-ready mobile and web application that incorporates anomaly detection for robust real-world performance, enabling wildlife authorities to triage confiscated shells in real time and streamline the path from seizure to ecosystem restoration.

2. Related Work

Our research builds upon two primary areas: ecological classification models and image embeddings for real-time anomaly detection systems. The former involves techniques for identifying species or ecosystems based on visual or environmental data, while the latter focuses on methods for detecting outliers or non-conforming inputs to ensure reli-

ability and robustness in real-world deployments.

2.1. Architectures

Convolutional Neural Networks (CNNs) have been central to image classification since the breakthrough by [23]. Efficient variants like ResNet [15], MobileNetV2 [38], and DenseNet [16] improved performance using residual connections, depthwise separable convolutions, and dense connectivity.

ConvNeXt [28] modernized CNNs by integrating transformer-inspired elements (large kernels, LayerNorm, GELU), narrowing the performance gap with transformers.

Vision Transformers (ViTs) [11], and later Swin [27] and DeiT [44], replaced convolutions with attention-based mechanisms, achieving state-of-the-art results by modeling global patch relationships. This shift from CNNs to transformer-based models continues to improve recognition tasks across domains.

2.2. Ecology Classification Models

Initial research framed marine shell recognition as a fine-grained classification task. [46] introduced the DDI dataset and proposed Shuffle-Xception, enhancing accuracy on cluttered seabed scenes [47]. [49] compiled a large benchmark (7,894 species, 59,000 images), evaluating traditional descriptors with classical classifiers.

To address class imbalance and visual similarity, [48] proposed FLNet, a CNN with filter pruning and a hybrid loss tailored to skewed distributions. More recently, [12] applied transformers for fish recognition, leveraging transfer learning for high accuracy.

Despite progress, most approaches emphasize taxon-

omy and overlook ecological context—such as geographic provenance—which is crucial for conservation and real-world deployment.

2.3. Image Embeddings and Anomaly Detection

Deep learning enables encoding images as high-dimensional embeddings that capture semantic content, supporting tasks like retrieval, classification, and anomaly detection. These embeddings cluster semantically similar inputs in feature space, offering more meaningful comparisons than raw pixels. Self-supervised methods have largely superseded supervised CNNs by learning from unlabeled data.

SimCLR [5] leveraged contrastive learning to encode image semantics via augmented views. CLIP [33] extended this to multimodal embeddings, enabling zero-shot transfer across vision-language tasks. DINO [4] showed that vision transformers can learn structural and semantic features without labels.

For scalable retrieval, embeddings are indexed using approximate nearest neighbor (ANN) algorithms like FAISS, enabling efficient search in large datasets [34].

Embeddings also underpin anomaly detection—key to identifying out-of-distribution (OOD) inputs. Similar techniques are used in NLP to detect hallucinations in LLMs [39, 40, 43]. Visual anomaly detection methods fall into three categories: (i) reconstruction-based (autoencoders, VAEs [2, 37]), (ii) probabilistic modeling (PaDiM [9], CS-Flow [36]), and (iii) confidence-based methods like ODIN [25].

While not central to our system, anomaly detection modules improve robustness against unexpected user-generated inputs in field conditions.

3. Method

This work makes three key contributions. First, we curate and release the first large-scale seashell image corpus, encompassing photographs from seashell species collected from both Costa Rica’s Pacific and Caribbean coasts—an essential resource for future marine-biodiversity studies. Second, we trained a lightweight ConvNeXt-Tiny-based classifier that distinguishes between the two ecosystems even for visually similar species from the same family that inhabit different shores. Third, we wrap the model in a production-ready web service that delivers predictions in under three seconds per image and incorporates an integrated anomaly detector, ensuring corrupted or out-of-distribution inputs are flagged before they reach the classifier.

3.1. Dataset Construction

We introduce **BackHome19K**, the first comprehensive seashell dataset specifically designed for ecosystem-level

classification. Our dataset addresses a critical gap in existing marine datasets [31, 49], which focus on taxonomic classification rather than bio geographic origin inference.



Figure 3. Representative examples of bivalves and gastropods collected from Caribbean coasts.

Because Pacific and Caribbean shells often share near-identical color palettes, growth rings, and overall silhouettes, a small dataset would leave the model unable to tease apart these subtle cues. We therefore assembled a broad and balanced collection spanning 516 species based on a list of seashells provided by the Biology School from the Universidad de Costa Rica.

Our dataset encompasses a diverse collection of marine mollusks from Pacific and Caribbean waters, representing major taxonomic groups across multiple families. Among the gastropods, we include specimens from families such as Acmaeidae (limpets), Buccinidae (whelks), Bullidae (bubble shells), Calliostomatidae (top shells), Cassidae (helmet shells), Cypraeidae (cowries), and Fasciolariidae (spindle shells), among numerous others. The bivalve collection features representatives from Arcidae (ark shells), Ostreidae (oysters), Pectinidae (scallops), Tellinidae (tellins), Semeliidae (semele clams), and Veneridae (venus clams), providing comprehensive coverage of both filter-feeding and burrowing species across varied marine habitats.



Figure 4. Representative examples of bivalves and gastropods collected from Pacific coasts.

We systematically cataloged the species endemic to Pacific and Caribbean coasts through consultation with marine biology experts and cross-referencing multiple authoritative databases [7, 19, 20, 32]. Species selection criteria included: (1) confirmed presence in target ecosystems,

Table 1. Dataset composition and characteristics for Pacific and Caribbean divided into Gastropods and Bivalves

| Metric | Pacific | Caribbean |
|---------------------|---------|-----------|
| Total Species | 237 | 279 |
| Total Images | 9,505 | 9,553 |
| Gastropod Species | 130 | 149 |
| Bivalve Species | 107 | 130 |
| Avg. Images/Species | 40.1 | 34.2 |

(2) sufficient morphological distinctiveness for visual classification, and (3) availability of high-quality reference imagery.

The dataset was assembled by three researchers under the guidance of Dr. Yolanda Camacho García, who provided the species list. Using her taxonomic expertise, the lead researcher defined a labeling protocol to ensure consistency within classes and clear separation between them. Images were chosen based on key morphological traits—texture, color, shape, and distinctive external features—and organized into class-specific folders grouped by ecosystem (Pacific vs. Caribbean). In Figure 5, we then computed each family’s mean embedding and applied t-SNE to visualize clusters of families with similar color, shape, texture, and size.

Over 10 months, we assembled 19,051 images. Each of them underwent rigorous quality control including: taxonomic verification, removal of synthetic/composite images and standardization of lighting conditions To ensure morphological diversity, we collected 30–40 pictures per species showing different growth stages, orientations, and preservation states.

The dataset was divided into Pacific and Caribbean subsets for the specific purpose of training the model introduced in the next section. However, the images are also organized by family, genus, and species, providing broader opportunities for future research. This structure enables studies focused on inter-species similarities and supports classification tasks based on taxonomic attributes rather than geographic origin.

3.2. Model Architecture

We adopt ConvNext-Tiny [28] as our classification backbone, initialized with ImageNet-1K pre-trained weights. Our architectural choice addresses two critical deployment constraints: computational efficiency for real-time inference and model size limitations for mobile deployment scenarios.

ConvNext modernizes traditional CNN design by incorporating key innovations from Vision Transformers while maintaining computational efficiency. The architecture employs large 7×7 convolution kernels in early stages, enabling effective global feature capture without the quadratic complexity of self-attention mechanisms. This design proves particularly advantageous for fine-grained classification tasks where subtle morphological differences must be detected across varying spatial scales. We deliberately exclude Vision Transformers from consideration due to model size, ViT-Base requires 86M parameters *vs.* ConvNext-Tiny’s 28M, creating storage constraints for field deployment; second, self-attention scales quadratically with input resolution, limiting real-time performance on resource-constrained devices.

This architecture already encodes a rich hierarchy of visual cues, from global shape and size in its early layers to increasingly fine color bands and micro-textures in its final block. By freezing the stem and first three stages we lock in the generic edge, contour, and chromatic primitives that distinguish broad shell geometries, while selectively unfreezing only the last block to relearn the millimeter-scale ridges, growth rings, and pigment speckles that separate near-identical families from Costa Rica’s Pacific and Caribbean coasts.

3.3. Anomaly Detection

Real-world deployment necessitates filtering non-seashell inputs that could degrade classification performance. Therefore, our approach focuses on using a vector representation of each training image and storing these vectors in a vector database. This enables efficient comparison against images provided by volunteers, resulting in a more accurate classification based on their similarities. This method ensures that only images containing the seashell as the primary element of interest are selected, reducing the impact of extraneous noise that could adversely affect the classification. Such an approach is particularly beneficial for a web application utilized by a wide range of users, many of whom may lack professional or scientific photography skills.

Our anomaly-detection module adopts the embedding logic of PaDiM [9] while using a more lightweight backbone. Concretely, each image is forwarded through SqueezeNet 1.0 [18]; in line with PaDiM we take the global-average-pooled activations of the final convolutional layer as the feature vector, yielding a 1 000-dimensional descriptor.

The choice of SqueezeNet is driven by its compactness and representational strength. With only ≈ 1.2 M parameters—roughly $20 \times$ fewer than canonical backbones such as ResNet-50—it fits comfortably on resource-constrained devices. In addition, its Fire modules (1×1 “squeeze” followed by $1 \times 1 / 3 \times 3$ “expand” filters) and delayed down-sampling keep high-resolution feature maps deep into the network, enabling it to capture the fine local textures that reveal subtle, pixel-level anomalies.

Having obtained a 1 000-dimensional embedding e_q for

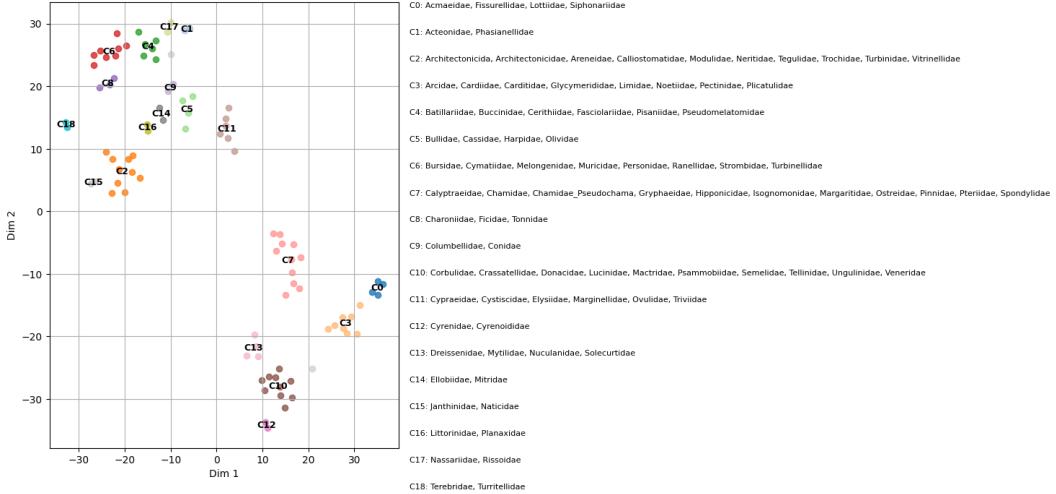


Figure 5. Seashell family clusters based on the mean feature vectors for each species using embedding representations from all the images in the dataset using t-SNE [45] and DBSCAN [13]



Figure 6. Representative specimens from our dataset showing morphological diversity across families. Top row (left to right): Pectinidae (scallop with characteristic radial ribs), Cardiidae (cockle with prominent radial sculpture), Cardiidae (heart cockle with distinctive coloration), Veneridae (venus clam with concentric growth lines). Middle row: Fasciolariidae (tulip shell with elongated spire), Cypraeidae (cowrie with glossy, oval shell), Turbinidae (turbo shell with nacreous interior visible). Bottom row: Mitridae (mitre shell with geometric pattern), Turritellidae (tower shell with high spire), Tellinidae (tellin with smooth, elongated form), Buccinidae (whelk with pointed spire), Fasciolariidae (spindle shell with fusiform shape).

each query image, the system then measures its affinity to the dataset by computing the mean cosine similarity to its k nearest neighbors:

$$S = \frac{1}{k} \sum_{i=1}^k \frac{\langle \mathbf{e}_q, \mathbf{e}_i \rangle}{\|\mathbf{e}_q\| \|\mathbf{e}_i\|}.$$

If $S < \lambda$, where λ is a predefined threshold, the image is classified as an anomaly:

$$\text{Classification} = \begin{cases} \text{Valid Image} & \text{if } S \geq \lambda \\ \text{Anomaly} & \text{if } S < \lambda \end{cases}$$

This method ensures that only valid seashell images are processed by the classification system, improving overall accuracy and reliability. The threshold λ was determined empirically by analyzing the distribution of similarity scores between known seashell images in our dataset.



Figure 7. Feature extraction and embedding generation from input images using the SqueezeNet architecture

4. Experiments

In the following section we indicate the procedure for the model training and hyperparameter selection.

4.1. Dataset Setting

The dataset was divided into three subsets for training, validation, and testing, with 70% allocated to the training set for model training, 15% to the validation set for hyperparameter tuning and performance evaluation during training, and 15% to the test set for evaluating the final model performance on unseen data. This split ensured a balanced representation of families across all subsets, supporting a rigorous and reliable evaluation of the model’s performance. The decision to retain almost 2,900 pictures for testing was made because of the need to ensure that the model would perform correctly on completely new seashells, as it was required to be used in a real-time framework with completely new seashells, ensuring that our dataset distribution could cover those undiscovered seashells. In addition, all images were resized to 224x224 pixels to maintain consistent input dimensions with the model architecture.

To improve the diversity of the training data and help the model perform better in real-world conditions, we applied several data augmentation techniques during training. These included: (1) geometric transformations, such as random rotations (up to $\pm 45^\circ$), horizontal and vertical flips, and zooming in or out (scaling between $0.8 \times$ and $1.2 \times$) to reflect different angles at which seashells might be photographed; (2) photometric changes, including variations in brightness (up to $\pm 20\%$), contrast (up to $\pm 15\%$), and saturation (up to $\pm 10\%$), to simulate differences in lighting; (3) spatial adjustments, such as random cropping (while keeping the aspect ratio) and slight elastic distortions, mimicking natural variations in how seashells may appear. These augmentations made the training set more realistic and better suited for field deployment, especially in cases where the images are taken by non-experts under uncontrolled conditions.

4.2. Implementation Details and Hyperparameter Analysis

All experiments were implemented in PyTorch and conducted on a single NVIDIA A100 GPU. We performed extensive hyperparameter optimization across seven distinct

Table 2. Top-1 test accuracy (mean \pm std across ten experiments) of four backbone CNNs. ConvNeXt-Tiny outperforms the next-best DenseNet121 by roughly 6 percentage points.

| Architecture | Test Accuracy (%) |
|---------------|-----------------------------------|
| ResNet50 | 78.3 ± 0.4 |
| DenseNet121 | 80.2 ± 0.3 |
| MobileNetV2 | 79.3 ± 0.5 |
| ConvNext-Tiny | 86.28 ± 0.5 |

configurations, systematically evaluating the impact of unfrozen layer counts, training duration, and learning rate scheduling on model performance.

We evaluated three optimizers (AdamW [29], Adam [21], and Stochastic Gradient Descent (SGD) [35]) with initial learning rates spanning 1×10^{-4} to 1×10^{-2} , comparing two scheduling strategies: cosine annealing and ReduceLROnPlateau. Our final configuration employs SGD with an initial learning rate of 0.001 and cosine annealing, which demonstrated superior convergence characteristics.

The architectural decisions were guided by empirical findings:

- Unfreezing the final 30 layers of ConvNext-Tiny provided optimal feature adaptation without overfitting
- Strategically placed dropout layers mitigated overfitting on underrepresented seashell families
- The chosen configuration particularly benefits fine-grained classification, where subtle inter-class differences demand careful feature extraction

This systematic exploration establishes a robust baseline for seashell classification while providing insights into transfer learning optimization for fine-grained visual tasks.

5. Results

5.1. Classification Performance

As shown in table 3, the model achieves balanced performance across both ecosystems, with slightly higher accuracy for Caribbean specimens (87.10% vs. 85.43%). This difference may reflect the larger number of Caribbean species in our training set (279 vs. 237 Pacific species).

ConvNeXt-Tiny’s superior performance stems from its modern architectural design, which effectively preserves high-frequency visual details crucial for fine-grained classification tasks. The network’s large 7×7 convolution kernels in early stages enable comprehensive feature capture, while its hierarchical structure maintains discriminative micro-textures and subtle morphological variations that distinguish shells from different ecosystems.

In contrast, traditional architectures like ResNet50 [15] and MobileNetV2 [38] struggle to preserve these fine-grained details due to their aggressive down-sampling

Table 3. Classification performance on BackHome19K test set (2,858 images) showing balanced accuracy across Pacific and Caribbean ecosystems

| Ecosystem | Accuracy (%) | Precision (%) | Recall (%) | F1 (%) |
|----------------|--------------|---------------|--------------|--------------|
| Pacific | 85.43 | 88.70 | 85.43 | 87.00 |
| Caribbean | 87.10 | 83.45 | 87.10 | 85.29 |
| Overall | 86.28 | 86.23 | 86.13 | 86.17 |

strategies and smaller receptive fields. This limitation becomes particularly problematic when classifying morphologically similar seashells from the same taxonomic family but different geographic origins, where the distinguishing features may be limited to minute surface textures.

The results demonstrate that the model performs consistently across both ecosystems, achieving a balanced accuracy of 85.43% on Pacific samples and 87.10% on Caribbean samples. This indicates that the system correctly classifies nearly 9 out of 10 seashells into their respective native ecosystems, highlighting its robustness and generalization capabilities across diverse marine environments.

5.2. Anomaly Detection Performance

Our anomaly detection system successfully filters non-seashell inputs while preserving legitimate specimens. Testing on 200 images across 20 object categories (10 images each, except 40 seashell images), the system achieved 100% recall on seashell images while correctly identifying 90.5% of non-seashell objects as anomalies using the empirically determined threshold $\lambda = 0.955$ and $k = 5$.

To probe the robustness of our similarity-based filter, we curated an out-of-domain corpus from different standard vision benchmarks such as COCO [26], ImageNet-1k [10], and Places365 [50]. For each of 18 everyday object categories, we sampled ten high-quality photos, yielding 180 non-seashell test images. We also added forty previously unseen in-domain seashell pictures for a recall check.



Figure 8. Anomaly detection interface displaying rejection of non-seashell objects using similarity-based filtering with threshold $\lambda = 0.955$, achieving 93% precision on out-of-domain images

A fixed threshold of 0.955 rejected 168 of the 180 off-

domain images (93 % true-positive rate). Perfect rejection was achieved for visually distinctive classes such as cats, cars, dogs, trucks, and backgrounds (10 / 10 each). More heterogeneous textures—notably reptiles, people, and frogs—accounted for the remaining misses, yet still exceeded a 60 % rejection rate. Crucially, all forty seashell controls scored above the threshold, producing zero false negatives and confirming that the detector preserves recall on in-domain data while aggressively filtering other objects.

6. Web Application

To make the classification model truly accessible, we released a production-ready two tier web service that lets users drag-and-drop shell photos and receive provenance predictions in ≤ 3 s. On the client side, a React application served via FastAPI “render” service, which exposes a receive-files endpoint: incoming images undergo Base64 encoding before being bundled with a JWT and asynchronously proxied, via HTTP/2 and Uvicorn’s event loop, to a second FastAPI service on a predict endpoint. That service forwards the payload to a custom prediction endpoint on Google Cloud Platform. The predict service then streams the JSON response received from Vertex AI back to the frontend, which dynamically updates the UI.

The entire stack, React assets, FastAPI render and predict containers, is Dockerized and deployed to Cloud Run, leveraging scale to zero, sub second cold starts, and horizontal auto scaling. This architecture ensures sub 3 seconds end to end response times under heavy concurrency while minimizing idle costs.

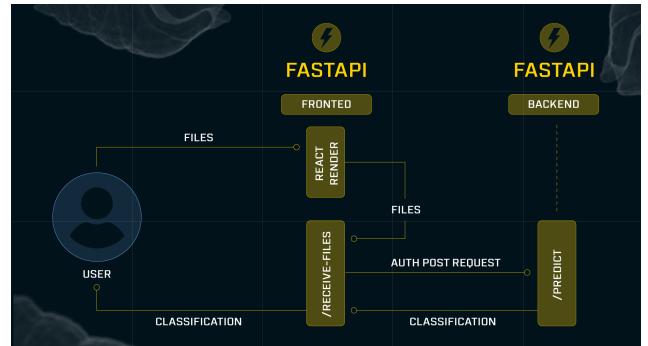


Figure 9. End-to-end technical pipeline for real-time shell classification system with sub-3-second response times via distributed FastAPI services and Cloud Run deployment

Cold-start latency on Cloud Run is ≈ 1.2 s, while warm inference adds a median 1.6 s, yielding a 95th-percentile end-to-end time of < 3 s from upload to prediction display.

During a three-day public launch, the app served ~ 200 unique users who collectively classified $\approx 36\,000$ seashells averaging 6.7 images s^{-1} at peak traffic—without a single unhandled error.

7. Ablations

To evaluate the performance of the ConvNext-Tiny architecture and our decision to unfreeze a set of layers given the efficiency and considering the benefits of initial ones, we ran a set of experiments using the ConvNext-Tiny architecture introduced above with a set of same hyperparameters to evaluate the changes. We trained these models with SGD [35] and with a learning rate of 1×10^{-2} and just differing the learning rate epoch scheduler and the number of epochs.

Table 4. Ablation Study: Model Performance with Different Configurations to observe unfrozen layers importance

| Unfrozen | Epochs | Scheduling | Accuracy |
|-----------|------------|------------|--------------|
| 0 | 50 | 25 | 83.24 |
| 0 | 50 | 25 | 84.11 |
| 10 | 100 | 50 | 80.15 |
| 10 | 150 | 75 | 84.56 |
| 10 | 250 | 125 | 84.16 |
| 13 | 150 | 75 | 84.16 |
| 30 | 100 | 68 | 86.28 |

Based on Table 4, we observe that unfreezing 30 layers yields the best performance with 86.28% accuracy. Notably, this configuration required only 100 training epochs compared to the 13-layer unfrozen experiment, which trained for 150 epochs yet achieved 2.12% lower accuracy (84.16%). The accuracy is presented on the test dataset.

The results demonstrate a clear relationship between the number of unfrozen layers and model performance. Experiments with 10 unfrozen layers consistently underperformed compared to the 30-layer configuration, with accuracies ranging from 80.15% to 84.56%. While the baseline frozen models (0 unfrozen layers) achieved reasonable performance (83.24%–84.11%), they were still outperformed by the optimal unfrozen configuration.

It is important to note that the varying number of epochs across experiments was determined based on validation accuracy convergence patterns observed during training using an early stop function. Each configuration was trained until the validation accuracy plateaued, indicating that additional epochs would incur unnecessary computational costs without performance gains.

8. Limitations

While the classification system demonstrated strong performance, several limitations were observed. The model occasionally failed under suboptimal lighting conditions or unconventional viewing angles/distances, significantly affecting extracted visual features and reducing accuracy. Additionally, some Caribbean seashells lack sufficient training images in research databases, creating classification difficulties.

Although the anomaly detection mechanism filtered many irrelevant inputs, it was not infallible. Non-seashell objects with similar textures (coral fragments, rocks, marine debris) were sometimes misclassified as seashells, indicating the need for better fine-grained visual distinction.



Figure 10. Example of seashells that were misclassified into Caribbean due to challenging conditions such as viewpoint, depth, lighting, and lack of visible texture or detail. From left to right: *Pectinidae*, *Arcidae*, *Cypraea* (x2), and a small *Cardiidae* shell.

The current dataset includes 516 species based on available scientific knowledge, though this initial version may not cover all existing species. Future updates will address these gaps as new discoveries emerge.

9. Conclusion

This research demonstrates that modern computer vision can effectively support conservation efforts through practical, deployable systems. Our BackHome19K dataset and classification pipeline achieve 86.3% accuracy in determining seashell ecosystem provenance, enabling the automated return of confiscated specimens to their native habitats.

This work establishes a framework for ecosystem-level classification that could extend to other marine organisms and geographic regions. The integration of anomaly detection with fine-grained classification proves essential for real-world deployment, where user-generated content varies significantly in quality and relevance.

The public release of BackHome19K will enable comparative studies and methodological improvements across the computer vision for ecology community.

Acknowledgments

We would like to thank FIFCO and the advertising agency Joystick for the funding and support provided throughout the duration of this project. We are especially grateful to Yolanda Camacho, PhD for her invaluable work and dedication in compiling the seashell lists. We also express our appreciation to the *Sistema Nacional de Áreas de Conservación* (SINAC), the *Universidad de Costa Rica* (UCR), and AERIS for their collaboration and assistance throughout this process. Finally, we extend our thanks to Vivian Mayorga from Joystick for her support in designing the web application and visual elements of this paper.

References

- [1] A.M. Costa Rica. Four tons of seashells seized from tourists in costa rica. *A.M. Costa Rica*, 2023. Online news article. 1
- [2] Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. Technical report, SNU Data Mining Center, 2015. Technical Report. 3
- [3] George Ashline, Joanna Ellis-Monaghan, Zsuzsanna Kadas, and Declan McCabe. Modeling seashell morphology. *UMAP/ILAP Modules*, 2009. 1
- [4] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 3
- [5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 3
- [6] Meiqi Cheng, Man Liu, Lirong Chang, Qing Liu, Chunxiao Wang, Le Hu, Ziyue Zhang, Wanying Ding, Li Chen, Sihan Guo, Zhi Qi, Panpan Pan, and Jingdi Chen. Overview of structure, function and integrated utilization of marine shell. *Science of The Total Environment*, 870:161950, 2023. 1
- [7] ConchyliNet. Seashell database, n.d. Accessed: 2024-12-26. 3
- [8] Francesco De Comité. Modelling seashells shapes and pigmentation patterns: Experiments with 3d printing. 2017. 1
- [9] Thomas Defard, Alex Setkov, Angeline Loesch, and Romaric Audiger. PaDiM: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition (ICPR)*, 2021. 3, 4
- [10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 7
- [11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2
- [12] Ebru Ergün. Swinfishnet: A swin transformer-based approach for automatic fish species classification using transfer learning. *PLOS ONE*, 20(5):1–27, 2025. 2
- [13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, page 226–231. AAAI Press, 1996. 5
- [14] Juan Gefaell, Juan Galindo, and Emilio Rolán-Alvarez. Shell color polymorphism in marine gastropods. *Evolutionary Applications*, 16(2):202–222, 2022. 1
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 2, 6
- [16] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017. 2
- [17] Xianpeng Huang, Zhenlin Hao, Junxia Mao, Luo Wang, Xubo Wang, and Ying Tian. Morphometric analysis of two shell color strains of the bay scallop *Argopecten irradians*. *Fishes*, 9(7), 2024. 1
- [18] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size, 2016. 4
- [19] iNaturalist. inaturalist: Connect with nature, n.d. Accessed: 2024-12-26. 3
- [20] Conchology Inc. Welcome to conchology, inc., n.d. Accessed: 2024-12-26. 3
- [21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2015. 6
- [22] Kevin M. Kocot, Felipe Aguilera, Carmel McDougall, Daniel J. Jackson, and Bernard M. Degnan. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Frontiers in Zoology*, 13:23, 2016. 1
- [23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. 2
- [24] Siqi Lai, Ling Shi, Yida Han, Ying Tian, and Zhenlin Hao. Adaptation of shell morphology to different tidal zones—insights into phenotypic plasticity of *Littorina brevicula*. *Frontiers in Ecology and Evolution*, Volume 12 - 2024, 2025. 1
- [25] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks, 2020. 3
- [26] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015. 7
- [27] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030*, 2021. 2
- [28] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11976–11986, 2022. 2, 4
- [29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2019. 6
- [30] Olivia Martínez-Ruiz, Rodrigo Riera, Víctor M. Tuset, and Joana Vasconcelos. Coastal exposure and artificialization: Drivers of shell shape variation in intertidal limpets. *Estuarine, Coastal and Shelf Science*, 322:109344, 2025. 1
- [31] Matthewwacs. South florida atlantic sea shell types, n.d. Accessed: 2024-12-26. 3

- [32] Florida Institute of Technology. Marine & environmental systems: Mollusk shell collection, n.d. Accessed: 2024-12-26. 3
- [33] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 3
- [34] MD Shaikh Rahman, Feiroz Humayara, Syed Maudud E Rabbi, and Muhammad Mahbubur Rashid. Efficient medical image retrieval using densenet and faiss for birads classification, 2024. 3
- [35] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951. 6, 8
- [36] Lukas Rudolph, Bastian Wandt, and Bodo Rosenhahn. Fully convolutional cross-scale-flow for image-based anomaly detection and localization. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022. 3
- [37] Mayu Sakurada and Takashi Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the 2nd Workshop on Machine Learning for Sensory Data Analysis (MLSDA), in conjunction with the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 4:4–4:11. ACM, 2014. 3
- [38] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018. 2, 6
- [39] Weihang Su, Yichen Tang, Qingyao Ai, Changyue Wang, Zhijing Wu, Yujia Zhou, and Yiqun Liu. Mitigating entity-level hallucination in large language models. *arXiv preprint arXiv:2407.09417*, 2024. 3
- [40] Weihang Su, Changyue Wang, Qingyao Ai, Yiran Hu, Zhijing Wu, Yujia Zhou, and Yiqun Liu. Unsupervised real-time hallucination detection based on the internal states of large language models. *arXiv preprint arXiv:2403.06448*, 2024. 3
- [41] The Costa Rican Times. Leave the shells, take memories: The hidden cost of your costa rican souvenir. *The Costa Rican Times*, 2023. Online news article. 1
- [42] The Tico Times. Shells must remain on costa rican beaches. *The Tico Times*, 2023. Online news article. 1
- [43] S. M. Towhidul Islam Tonmoy, S. M. Mehedi Zaman, Vinija Jain, Anku Rani, Vipula Rawte, Aman Chadha, and Amitava Das. A comprehensive survey of hallucination mitigation techniques in large language models. *arXiv preprint arXiv:2401.01313*, 2024. 3
- [44] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. *arXiv preprint arXiv:2012.12877*, 2021. 2
- [45] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008. 5
- [46] Bing Xue, Baoxiang Huang, Ge Chen, Haitao Li, and Weibo Wei. Deep-sea debris identification using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:8909–8915, 2021. 2
- [47] H. Xue, Y. Li, and X. Wang. Deep-sea debris identification with shuffle-xception. In *Proc. OCEANS*, 2021. 2
- [48] J. Yue, P. Luo, and S. Liu. Flnet: Filter pruning and repair for imbalanced shellfish recognition. *Pattern Recognition*, 142:109–119, 2023. 2
- [49] Qi Zhang, Jianhang Zhou, Jing He, Xiaodong Cun, Shaoning Zeng, and Bob Zhang. A shell dataset for shell features extraction and recognition. *Scientific Data*, 6(226), 2019. 2, 3
- [50] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 7