# MaRI: Material Retrieval Integration across Domains

## Supplementary Material

This supplementary material provides additional details and insights into the MaRI framework. Section 1 outlines the implementation details, including training configurations and the use of pre-trained DINOv2 [8] backbones. Section 2 presents an analysis of how different backbone architectures impact material retrieval performance on both trained and unseen datasets. Practical applications of MaRI, such as assigning materials to 3D models for design workflows, are demonstrated in Section 3, showcasing the system's versatility and ease of use. In addition, Section 4 reports a user study, and Section 5 offers further discussion on limitations, and potential future improvements.

## 1. Implementation Details

We use DINOv2 as the backbone for both the image and material encoders, initialized with pre-trained weights. The training process consists of two main stages: first, the model is fine-tuned on the synthetic dataset for 1 epoch using the Adam optimizer with a learning rate of $1 \times 10^{-4}$. This initial phase helps establish a robust baseline. Next, the model undergoes fine-tuning on the real-world dataset for 25 epochs at a reduced learning rate of $1 \times 10^{-5}$ to enhance generalization by capturing more intricate real-world features. To facilitate effective alignment in the shared feature space, the temperature parameter $\tau$ in the contrastive loss is set to 0.07. The dataset is divided into 90% for training and 10% for validation, and a batch size of 256 is employed. The entire process is conducted on four NVIDIA A100 GPUs (80GB each), completing within 3 hours.

## 2. Effect of Backbone Variations

The performance of different backbone architectures for material retrieval is summarized in Table 1. DINOv2 demonstrates the strongest overall performance, achieving top-1 instance accuracy (T1I) of 26.0% and top-5 instance accuracy (T5I) of 90.0% on the Trained dataset, along with 54.0% and 89.0% on the Unseen dataset, respectively.

Table 1. Backbone comparison for material retrieval on Trained and Unseen datasets. Best values are highlighted in **blue**.

| Backbone | Trained | | | | Unseen | |
|---|---|---|---|---|---|---|
| | T1I | T5I | T1C | T3IoU | T1I | T5I |
| ResNet50 [4] | 7.5% | 28.0% | 64.5% | 0.58 | 30.0% | 60.5% |
| ViT [1] | 15.0% | 33.5% | 54.5% | 0.67 | 21.0% | 65.0% |
| Swin Transformer [5] | 14.0% | 41.5% | 77.0% | 0.68 | 38.5% | 77.0% |
| ConvNeXt [6] | 16.5% | 42.0% | 70.0% | 0.67 | 35.0% | 73.5% |
| EfficientNet [11] | 8.5% | 21.5% | 52.0% | 0.55 | 22.0% | 54.0% |
| CLIP [9] | 12.0% | 44.5% | 61.0% | 0.68 | 29.5% | 72.5% |
| DINOv2 [8] | **26.0%** | **90.0%** | **81.5%** | **0.77** | **54.0%** | **89.0%** |

Table 1 highlights DINOv2's capability to effectively capture complex material characteristics and generalize across diverse data distributions. Other architectures show moderate performance, reflecting limited generalization capacity. These findings emphasize the critical role of backbone selection in enhancing material retrieval tasks, with DINOv2 emerging as the most effective backbone for bridging the synthetic-to-real domain gap.

## 3. Applications

Having established the effectiveness of MaRI in retrieving materials accurately, we now highlight a practical application that leverage MaRI's capabilities in real-world 3D design workflows. MaRI empowers users to effortlessly assign desired materials to different parts of a 3D model by providing a streamlined paradigm for material retrieval and application. For instance, a user working with a 3D chair model can specify preferred materials—such as leather for the seat, metal for the legs, and wood for the armrests—through simple input reference images. MaRI efficiently retrieves matching physically based rendering (PBR) materials and applies them to the respective parts. The workflow involves the following steps:

1. **Object Segmentation:** The 3D model is segmented into distinct components (e.g., seat, legs, and armrests) based on its structural design or user annotations.
2. **Material Retrieval:** Users provide reference images for their desired material patterns, such as leather textures or wood grains. MaRI leverages its robust library to retrieve accurate and visually consistent matches.
3. **Material Application:** The retrieved materials are mapped and applied to the segmented parts of the model, enabling users to visualize their design with photorealistic fidelity.

This process, exemplified through the chair model in Figure 2, highlights MaRI's ability to offer users an intuitive and efficient workflow for realizing their creative goals. MaRI transforms the traditionally labor-intensive process, allowing users to quickly locate and apply desired materials, resulting in outcomes that closely match their creative intent.

## 4. User Study

We conducted a user study with 20 participants (40% designers, 30% researchers, and 30% general users). Each participant provided three material images as queries, and the top-5 retrievals from MaRI and baseline methods were rated on relevance, realism, and perceptual consistency using a 5-point Likert scale. Table 2 shows that MaRI achieved

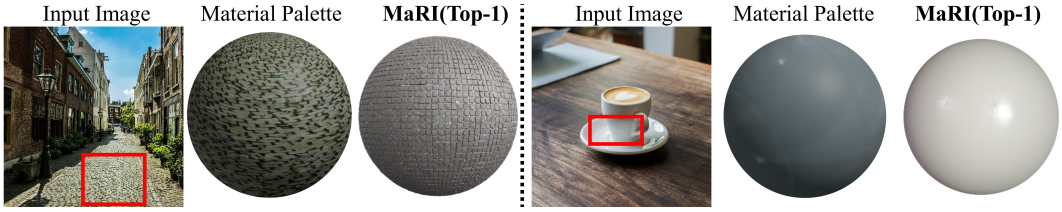


Figure 1. Comparison between Material Palette and MaRI.

Table 2. User study average scores for top-k retrievals. Best values are highlighted in **blue**.

| Method | Top-1 | Top-3 | Top-5 |
|---|---|---|---|
| MaPa [13] | 2.60 | 3.05 | 3.80 |
| Make-it-Real [2] | 3.25 | 3.50 | 3.90 |
| MaRI | **4.15** | **4.50** | **4.85** |

consistently higher scores across all top-k retrieval results. These results indicate that our method retrieves materials that are more relevant, realistic, and perceptually consistent compared to the baselines.

## 5. Discussion

We further analyze the behavior of our retrieval system. As shown in Figure 7, the last row presents a case where a query for sponge material returns several sand-like materials. This outcome is due to the limited number of sponge samples available in our gallery, which causes the system to favor visually similar textures. The case highlights a potential limitation of our approach compared to direct generative methods like Material Palette [7], which can synthesize a broader variety of material appearances. However, our retrieval-based approach is faster than diffusion-based methods. While Material Palette generates novel material images using diffusion, our method retrieves assets from an existing database. This distinction is crucial, as it ensures that our results are not only of high quality but also readily support PBR rendering. Figure 1 shows a comparison case.

In our work, we adopt material spheres as the shape representation, which is a widely used choice among content creators. MaRI effectively captures essential material properties and supports robust retrieval performance. We believe that exploring alternative representations, such as blob [12] or Havran's [3] shapes, may further enhance perceptual alignment [10] by capturing subtle material nuances. This presents a promising direction for future research.

## References

[1] 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, 2021. 1

[2] Ye Fang, Zeyi Sun, Tong Wu, Jiaqi Wang, Ziwei Liu, Gordon Wetzstein, and Dahua Lin. Make-it-real: Unleashing large multimodal model for painting 3d objects with realistic materials, 2024. 2

[3] V. Havran, J. Filip, and K. Myszkowski. Perceptually motivated brdf comparison using single image. *Comput. Graph. Forum*, 35(4):1–12, 2016. 2

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 1

[5] 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, 2021. 1

[6] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s, 2022. 1

[7] Ivan Lopes, Fabio Pizzati, and Raoul de Charette. Material palette: Extraction of materials from a single image, 2023. 2

[8] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2024. 1

[9] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. 1

[10] Ana Serrano, Bin Chen, Chao Wang, Michal Piovarči, Hans-Peter Seidel, Piotr Didyk, and Karol Myszkowski. The effect of shape and illumination on material perception: model and applications. *ACM Trans. Graph.*, 40(4), 2021. 2

[11] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. 1

[12] Peter Vangorp, Jurgen Laurijssen, and Philip Dutré. The influence of shape on the perception of material reflectance. In *ACM SIGGRAPH 2007 Papers*, page 77–es, New York, NY, USA, 2007. Association for Computing Machinery. 2

[13] Shangzhan Zhang, Sida Peng, Tao Xu, Yuanbo Yang, Tianrun Chen, Nan Xue, Yujun Shen, Hujun Bao, Ruizhen Hu, and Xiaowei Zhou. Mapa: Text-driven photorealistic material painting for 3d shapes, 2024. 2


Figure 2. Material assignment and rendering for a 3D chair model.


Figure 3. Material assignment and rendering for a 3D plant model.


Figure 4. Material assignment and rendering for a 3D desk model.


Figure 5. Material assignment and rendering for a 3D bucket model.

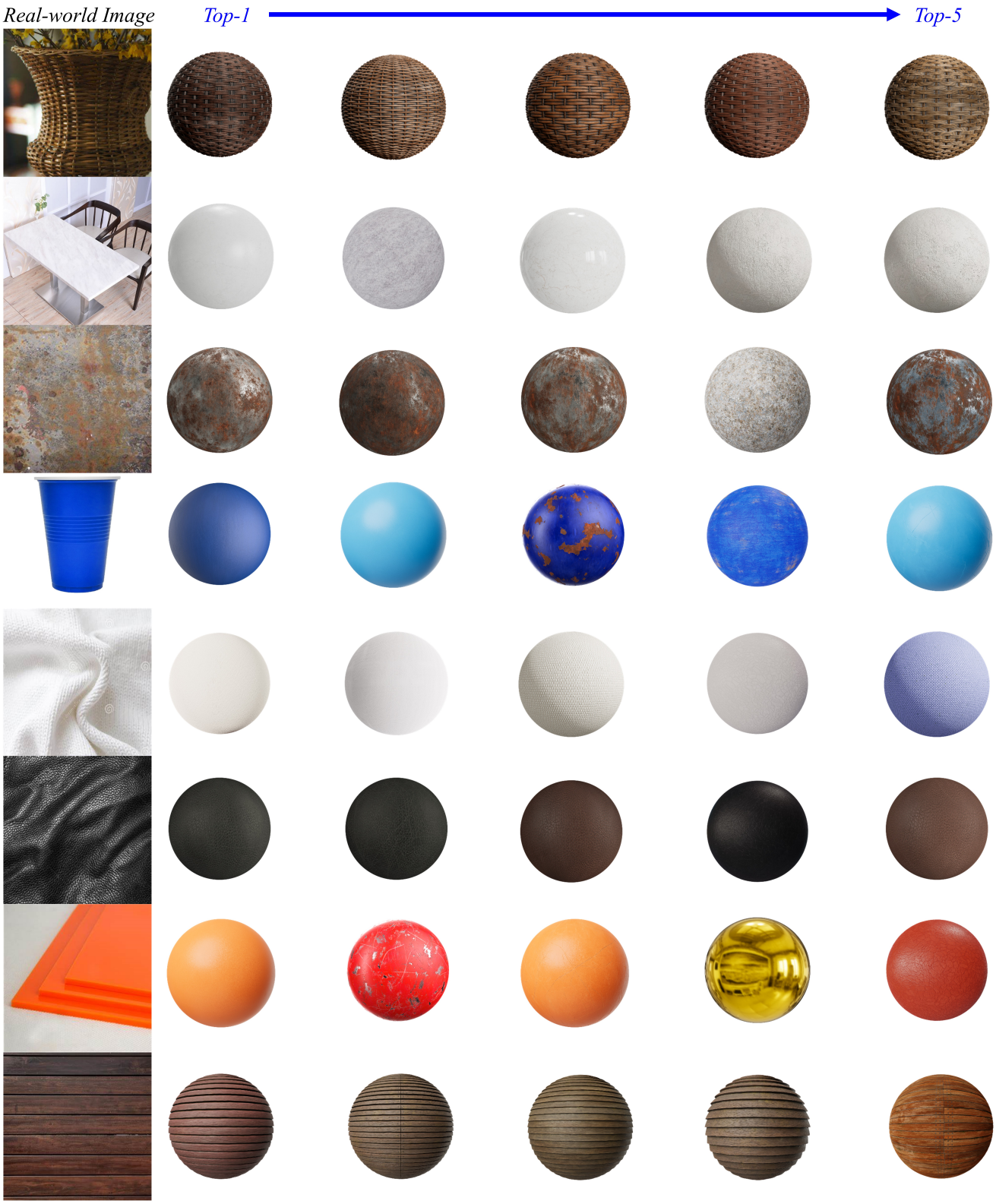


Figure 6. Top-5 material retrieval results for real-world images.


Figure 7. Top-5 material retrieval results for real-world images.