Intermediate Domain Alignment and Morphology Analogy for Patent-Product Image Retrieval

Haifan Gong[1,2†], Xuanye Zhang[1,2†], Ruifei Zhang[1,2], Yun Su[3], Zhuo Li[1,2], Yuhao Du[1,2], Anningzhe Gao[2], Xiang Wan[1,2*], Haofeng Li[4,2*]

[1]The Chinese University of Hong Kong, Shenzhen,
[2]Shenzhen Research Institution of Big Data,
[3]University of Waterloo,
[4]School of Systems Science and Engineering, Sun Yat-sen University

Abstract

Recent advances in artificial intelligence have significantly impacted image retrieval tasks, yet Patent-Product Image Retrieval (PPIR) has received limited attention. PPIR, which retrieves patent images based on product images to identify potential infringements, presents unique challenges: (1) both product and patent images often contain numerous categories of artificial objects, but models pre-trained on standard datasets exhibit limited discriminative power to recognize some of those unseen objects; and (2) the significant domain gap between binary patent line drawings and colorful RGB product images further complicates similarity comparisons for product-patent pairs. To address these challenges, we formulate it as an open-set image retrieval task and introduce a comprehensive Patent-Product Image Retrieval Dataset (PPIRD) including a test set with 439 product-patent pairs, a retrieval pool of 727,921 patents, and an unlabeled pre-training set of 3,799,695 images. We further propose a novel Intermediate Domain Alignment and Morphology Analogy (IDAMA) strategy. IDAMA maps both image types to an intermediate sketch domain using edge detection to minimize the domain discrepancy, and employs a Morphology Analogy Filter to select discriminative patent images based on visual features via analogical reasoning. Extensive experiments on PPIRD demonstrate that IDAMA significantly outperforms baseline methods (+7.58 mAR) and offers valuable insights into domain mapping and representation learning for PPIR. (The PPIRD dataset is available at: https://loslorien.github.io/idama-project/)

1 Introduction

Patent systems play a critical role in protecting intellectual property and fostering innovation. The rapid advancement of artificial intelligence (AI) has revolutionized numerous domains, including patent retrieval [1, 2, 3, 4, 5]. Specifically, in the NLP area, significant progress has been made in both patent-patent retrieval [1, 2] and patent-product retrieval [3, 4], leveraging textual descriptions to identify similarities and potential infringements. In contrast, the CV community has primarily focused on patent-patent retrieval [5], leaving Patent-Product Image Retrieval (PPIR) unexplored. Considering the increasing complexity of e-commerce platforms (a product may have an incomplete, confusing name or descriptions, but its images are always clear) and the critical need to detect patent infringements efficiently, there is an urgent demand for research on PPIR.

PPIR faces two primary challenges: 1) product and patent images often contain numerous categories (more than 250,000) of artificial objects, such as mechanical parts or electronic devices, but most

*Corresponding Email: wanxiang@sribd.cn, lihf95@mail.sysu.edu.cn, † these authors contribute equally to this work.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

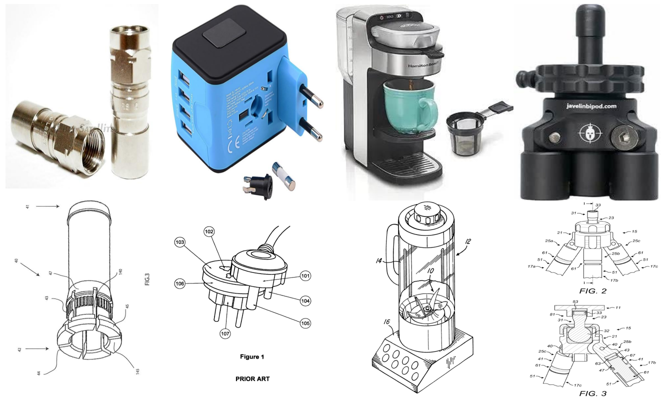

Figure 1: **Characteristics of Patent-Product Image Retrieval (PPIR) task:** (Upper: Product Images, Lower: Patent Images) 1) PPIR task has large-scale artificial categories, which are significantly different from those in general vision datasets; 2) Patent binary line-drawings images and Product colorful RGB images have huge domain gap, casting negative effect for PPIR tasks.

of these categories are distinct from those in general pre-training datasets like ImageNet [6]. 2) There is a significant domain gap between patent images (typically binary line drawings) and product images (typically colorful RGB images). This domain discrepancy further exacerbates the difficulty of accurate retrieval. (See Fig. 1 for details.)

Due to the unique characteristics of PPIR, existing general image retrieval methods [7] exhibit suboptimal performance for two reasons: 1) General image retrieval approaches often struggle to discriminate a large number of unseen artificial-category objects, which makes it challenging to grasp distinctive visual features of patents/products for accurate retrieval; 2) The domain gap causes distribution shift for specific categories, which further aggravate the retrieval performance, especially for some similar artificial objects.

We formulate PPIR as an open-set image retrieval task to simulate real-world scenarios. Moreover, to address the lack of suitable data for PPIR, we introduce the Patent-Product Image Retrieval Dataset (PPIRD), a large-scale dataset designed to facilitate research in this area. PPIRD comprises two components: (1) a testing set containing 439 product-patent pairs alongside a retrieval pool of 727,921 patent images and (2) an unlabeled pre-training set with 3,799,695 product/patent images. We offer highly detailed product descriptions for each product and the corresponding potential patent infringement pair to verify potential patent-patent infringement. (The extremely detailed information required and the difficulties of ascertaining potential patent-patent infringement both limit the scale of the infringed product-patent pair.)

Furthermore, we propose an extremely simple but effective Intermediate Domain Alignment and Morphology Analogy (IDAMA) strategy for PPIR. IDAMA contains an Intermediate Domain Mapping (IDM) strategy and a Morphology Analogy Filter (MAF) strategy: 1) IDM aligns binary line drawing patent images and colorful RGB product images by mapping them into an intermediate sketch domain using an edge detector. We provide a theoretical analysis to prove that this alignment effectively mitigates the domain discrepancy, enabling more accurate retrieval. 2) MAF leverages a cognitive principle of morphology analogy. An unknown object can be described by analogy to a known object—by using high classification confidence (regardless of the label) to select a discriminative image for artificial patent images for retrieval similarity comparison. By selecting discriminative images for unseen artificial patents, MAF can grasp distinctive visual features of the patents for PPIR.

Experiments on PPIRD demonstrate that IDAMA achieves significant improvements over baseline methods (+7.58 mAR), proving the effectiveness of IDM and MAF in addressing the unique challenges of PPIR. Furthermore, we systematically compare various domain alignment and representation learning methods on PPIRD, providing insights into their strengths and limitations in coping with unseen artificial objects in cross-domain scenarios. These findings offer valuable guidance for future research in PPIR and related areas.

In summary, this work makes three key contributions:

1. We formulate PPIR as an open-set image retrieval task and introduce PPIRD, a large-scale dataset for Patent-Product Image Retrieval;
2. We propose IDAMA, a novel strategy that combines intermediate domain alignment and morphology analogy to address challenges of PPIR; We further provide theoretical analysis to demonstrate the advances of intermediate domain alignment via compressive sensing.
3. We comprehensively evaluate domain alignment and representation learning methods, providing insights for efficient cross-domain learning.

2 Related Works

2.1 Image-based Patent/Sketch Datasets and Benchmarks

The availability of large-scale, well-structured datasets is essential for advancing AI research in patent analysis. [8] presented *PatentMatch*, a dataset specifically created for matching patent claims with prior art, facilitating studies in patent infringement detection and novelty assessment. For image-focused research, [5] developed *DeepPatent*, a large-scale benchmarking corpus aimed at patent drawing recognition and retrieval. [9] extended this work with *DeepPatent2*, focusing on technical drawing understanding. [10] introduced the Harvard USPTO Patent Dataset, a comprehensive corpus of patent applications designed to support diverse AI research tasks. However, the research on image-based patent retrieval mainly focuses on image retrieval between patents and solving patent infringement problems. Product infringement detection and corresponding patent-product image retrieval (PPIR) are under-explored. This paper introduces a large-scale patent-product image retrieval dataset to satisfy the requirements.

On the other hand, Im4Sketch [11] proposes a large-scale natural RGB/sketch image retrieval dataset together with a method to address the RGB-sketch retrieval challenges. However, the images of Im4Sketch dataset are collected from ImageNet-1K [6] and adopt the closed-set image retrieval formulation, which exists two main differences from PPIR task: 1) PPIR contains large-scale artificial categories that do not exist in ImageNet-1K. 2) The huge amounts of artificial categories and the continuously appearing products make PPIR not suitable to be formulated as a closed-set image retrieval task.

2.2 Patent Retrieval in NLP areas

The analysis and retrieval of patents have been pivotal research areas in the application of artificial intelligence (AI) across multiple domains, including classification and retrieval [12, 13]. Accurate patent classification is critical for efficiently organizing and accessing patent information. Several studies have explored AI-driven methods to automate patent classification, with notable contributions from [14, 15, 16, 17]. These works investigate using machine learning models to handle the complexity of patent data. For example, [18] examined the potential of AI in solving the patent classification problem, analyzing various machine learning models and their effectiveness in handling the diverse and intricate nature of patent data.

In patent retrieval, AI techniques have been widely used to enhance the accuracy and efficiency of retrieving relevant patents. Traditional text-based methods have seen significant advancements by applying deep learning models. [1] proposed an advanced patent prior art search method using deep learning, which relies on semantic embeddings to capture the contextual meaning of patent documents, offering improvements over keyword-based retrieval systems. Similarly, [19] integrated text embeddings with knowledge graph embeddings to further refine patent retrieval, highlighting the importance of external knowledge sources in improving search performance. More recently, [20] explored the use of large language models (LLMs) [21, 22] for patent retrieval, demonstrating their potential to improve retrieval accuracy. To satisfy the real-world requirements of product infringement detection, [23, 14, 15, 16, 17, 10] propose and explore patent-product retrieval. Patent-product retrieval is formulated as an unsupervised retrieval task to satisfy the continuously emerging new products.

In summary, the formulation of patent retrieval in NLP areas has developed from classification or supervised retrieval to open-set unsupervised retrieval using a retrieval pool. Considering that open-set unsupervised retrieval from a retrieval pool is closer to real-world scenarios, which can

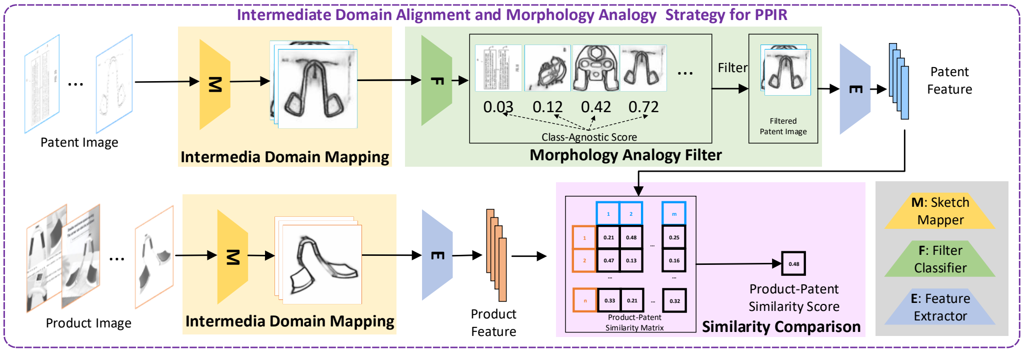

Figure 2: **Pipeline of IDAMA:** IDAMA consists of Intermediate Domain Mapping and Morphology Analogy Filter methods. The pipeline of IDAMA is as follows: 1) Using IDM to align product/patent images by mapping them into intermediate sketch domains; 2) Using MAF to filter discriminative mapped patent images; 3) Comparing product-patent similarity by obtaining product-patent similarity score from product-patent similarity matrix between filtered patent images and mapped product images for infringement detection.

address the issue of conducting infringement detection for continuously emerging products, the proposed PPIR tasks also adopt such formulations to satisfy real-world requirements.

2.3 Patent Retrieval with Images

Patent images often include technical drawings and diagrams, and image retrieval in this domain has emerged as an important area of study. [3] focused on improving the quality of convolutional neural network (CNN) [24] training datasets for patent image retrieval, emphasizing the importance of dataset quality in model performance. Building on this, [4] introduced methods using cross-entropy-based metric learning to enhance the accuracy of patent image retrieval systems. In another study, [2] proposed transformer-based deep metric learning for patent image retrieval, showcasing the effectiveness of transformer architectures in this domain. Some works achieve pure patent retrieval by transferring the binary line-draw patent images [25, 26, 27] into sketches to enhance performance.

However, the huge domain gap between patent binary line-drawing images and product colorful RGB images makes it extremely challenging for unsupervised open-set PPIR compared to patent retrieval tasks. Althought there are some works that aims to address the domain gap [28, 29, 30, 31, 32], especially in cross-domain image retrieval [33, 34]. However, these methods usually focus on close-set with options rather than the challenging open-world setting. Therefore, this paper proposes domain mapping methods to bridge the domain gaps between patent and product through an intermediate domain to alleviate the challenges.

3 Methodology

The Intermediate Domain Alignment and Morphology Analogy (IDAMA) strategy for PPIR consists of Intermediate Domain Mapping (IDM) (Sec. 3.1) and Morphology Analogy Filter (MAF) (Sec. 3.2). We also prove a brief for IDM's effectiveness (Sec. 3.3). The pipeline of IDAMA is demonstrated in Fig. 2: 1) Using IDM to align product/patent images by mapping them into intermediate sketch domains; 2) Using MAF to filter discriminative mapped patent images; 3) Comparing product-patent similarity by obtaining product-patent similarity score from product-patent similarity matrix between filtered patent images and mapped product images for infringement detection.

3.1 Intermediate Domain Mapping

The PPIR faces a primary challenge: the domain gap [35, 36, 37] between product and patent images. Specifically, product images are RGB images containing rich color, texture, and other information [38], while patent images are binary line drawings that abstractly represent products [39, 40]. Resolving the domain gaps between product and patent images thus becomes our primary task.

To address the challenge, we propose the Intermediate Domain Mapping (IDM) (See Fig. 2 for details) to align binary drawings, patent images, and colorful RGB product images by mapping both of them to the intermediate sketch domain.

Specifically, we denote a sketch mapper as $M_{\text{sketch}}(\cdot)$ [41], which can map both product images I_r and patent images I_p to the intermediate sketch domain. In the PPIR process, IDM for PPIR first maps raw product images I_r and patent images I_p to the intermediate sketch domain and obtains mapped sketch images I'_r and I'_p, which can be expressed as:

$$I'_r = M_{\text{sketch}}(I_r), \tag{1}$$

$$I'_p = M_{\text{sketch}}(I_p). \tag{2}$$

With the above-defined sketch-based intermediate domain mapping method, we can mitigate the domain gap between patents and products. Analysis in Fig. 4 further confirms the effectiveness of our process.

3.2 Morphology Analogy Filter

PPIRD needs to handle a large number of unseen man-made objects. Because general pretraining datasets contain limited man-made categories, extracting features with backbones pre-trained on these datasets is less discriminative for efficient product-patent retrieval.

To overcome the challenge, we observe how human beings grasp the distinct visual feature of an unseen object: human beings analogize the morphological feature of the unseen object to some familiar objects [42]. By doing this, even those who do not see the object can recognize the object through the morphology analogy. Inspired by this cognitive principle, we propose a Morphology Analogy Filter (MAF) for filtering patent images before mapping to the edge domain.

Specifically, given a Filter F (a classification network) and mapped patent images I_p, MAF first obtain the classification vector c_p by

$$c_p = F(I_p) \tag{3}$$

Then, MAF obtains discriminative patents I_{pd} image using the max classification scores regardless of the label itself (class-agnostic scores), which can be expressed as:

$$I_{pd} = argwhere(argmax(c_{p_j}) > \tau_{maf}) \tag{4}$$

where c_{p_j} is the classification score in c_p and τ_{maf} is the MAF threshold. If all the image classification scores are less than τ_{maf}, we preserve the image of max c_{p_j} for the patent.

We only conduct MAF for patent images as follows: The motivation of PPIR is to discover whether a product infringes any patent. Each image contains valuable visual features. It'd be better to follow the 'Better to have more than to miss out' philosophy for products. Because morphology similarity may indicate potential infringement for patent images, it'd be better to follow the rule 'Better to have less than to have something of poor quality' to select distance visual features of patents to avoid the negative influence. Therefore, we only conduct MAF for patent images.

Similarity Comparison: Then, a feature extractor $E(\cdot)$ are adopted to extract feature f_r and f_p from I_r and I_{pd} :

$$f'_r = E_{(\cdot)}(I_r) \tag{5}$$

$$f'_p = E_{(\cdot)}(I_{pd}) \tag{6}$$

After obtaining the intermediate sketch-domain feature f'_r and f'_p, PPIRD calculates the cosine similarity score for them to conduct PPIR:

$$s'_{r,p} = sim(f'_r, f'_p) \tag{7}$$

where $sim(\cdot)$ denotes the cosine similarity function.

For PPIR tasks, each patent p and product r may contain more than one image, which can be expressed as $r = \{I_{p1}, I_{p2}, ...\}$ and $p = \{I_{r1}, I_{r2}, ...\}$. IDM uses a maximum of image-level similarity scores (formulating product-patent similarity matrix) in the intermediate domain as product-patent similarity score $s_{p,r}$:

$$s_{p,r} = max(s'_{r1,p1}, s'_{r1,p2}, ..., s'_{r2,p2}, ...) \tag{8}$$

IDM for PPIR compares each product r with all the patents in the retrieval pools, denoted as $P = \{p_1, p_2, ...\}$, obtains their similarity score set, denoted as $S_r = \{s_{p1,r}, s_{p2,r}, ...\}$, and then sort S_r for patent-product retrieval.

3.3 Understanding Intermediate Domain Mapping via Compressive Sensing

Preliminaries. Let $I \in \mathbb{R}^n$ denote a vectorised image that can be decomposed as

$$I = s + \eta, \tag{9}$$

where s encodes the geometrical structure (edges, contours) while the term η contains texture, colour and background clutter. The mapper M used in intermediate domain mapping (IDM) has a linear front–end $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \ll n$, whose entries are i.i.d. $\mathcal{N}(0, 1/m)$ random variables. Such a matrix acts as a Johnson–Lindenstrauss embedding and, with overwhelming probability, satisfies the Restricted Isometry Property (RIP) required by compressed-sensing theory [43, 44, 45, 46, 47].

Model assumptions. 1) s is k-sparse in a known orthonormal basis Ψ; 2) the embedding dimension obeys $m \geq Ck\log(n/k)$ for a universal constant $C > 0$; 3) the backbone network E is L-Lipschitz, i.e. $\|E(x) - E(y)\|_2 \leq L\|x - y\|_2$ for all x, y (see, e.g., [48, 49]).

Definition 1 (Restricted Isometry Property [50])**.** *A matrix $\mathbf{A}$ satisfies $\mathrm{RIP}(2k, \delta)$ if, for every $2k$-sparse vector $x \in \mathbb{R}^n$,*

$$(1 - \delta)\,\|x\|_2^2 \;\leq\; \|\mathbf{A}x\|_2^2 \;\leq\; (1 + \delta)\,\|x\|_2^2. \tag{10}$$

Claim 2 (Distance preservation)**.** *Consider a patent–product pair (I_p, I_{pr}) whose structural components s_p, s_{pr} satisfy $\|s_p - s_{pr}\|_2 \leq \varepsilon$. If $\mathbf{A}$ fulfils $\mathrm{RIP}(2k, \delta)$, then*

$$\|\mathbf{A}s_p - \mathbf{A}s_{pr}\|_2 \;\leq\; (1 + \delta)\,\varepsilon. \tag{11}$$

Theorem 1 (IDM feature-space contraction)**.** *Under Assumptions 1–3 and Definition 1, the feature distance of the mapped images satisfies*

$$\|\tilde{z}_p - \tilde{z}_{pr}\|_2 \;\leq\; L\big((1 + \delta)\varepsilon + \tfrac{m}{n}\big(\|\eta_p\|_2 + \|\eta_{pr}\|_2\big)\big)\,, \tag{12}$$

where $\tilde{z}_p = E(\mathbf{A}I_p)$ and $\tilde{z}_{pr} = E(\mathbf{A}I_{pr})$.

Remark 3. *Because $m \ll n$, the nuisance term is suppressed by the factor m/n, whereas the structural term is almost isometrically preserved by RIP. Consequently, with high probability, $\|\tilde{z}_p - \tilde{z}_{pr}\|_2 \ll \|z_p - z_{pr}\|_2$, which explains the empirical robustness of Intermediate Domain Mapping.*

The proofs of Claim 2 and Theorem 1 are deferred to Appendix A.

4 Experiments

4.1 PPIRD: A Patent-Product Image Retrieval Dataset

Retrieving patents for infringing products is complex and time-consuming, heavily relying on patent experts to search massive patent databases to identify potential infringing patents [13]. This reliance makes it challenging to obtain a large-scale paired dataset of product-patent images to train models in a supervised manner. To address the challenge, we proposes PPIRD, consisting of PPIRD-testing and PPIRD-unlabeled, and formulates it as a retrieval task. To the best of knowledge, we provide the first benchmark for image-based patent-product retrieval at near million data level.

PPIRD-testing set: The PPIRD-testing set contains 439 product-patent pairs and a retrieval pool of 727,921 patent images. Experts annotate and validate the product-patent infringement pairs (Expert can refer the detailed descriptions of 439 products for more accurate validation). For the infringed patents, we use Google Patents to search 9,999 patents for each product (using their name as a keyword) and download their patent images to formulate a retrieval pool. We formulate PPIRD-testing as product-patent-level retrieval rather than image-level retrieval like DeepPatent [5] to simulate real-world scenarios better.

The formulation of PPIRD-testing is as follows: let the product dataset be D_r and the patent dataset be D_p. For a product x_i from the product dataset D_r, we aim to find its corresponding patent y_j in

Table 1: **Quatitative Results of IDAMA:** 1) IDAMA can bring significant performance enhancement compared with baseline methods, and both IDM and MAF can contribute to the improvement. 2) Comparison between DeepPatent [5] and MAF ('IDM+DeepPatent' and 'IDAMA') proves the intuitive idea of MAF can bring more performance enhancement even without extra pre-training. 3) Comparison between UCDIR [33] and UCDIR+IDM/IDAMA also demonstrates that our method can consistently boosts the performance. 4) Contrastive Learning ('IBOT') may be more suitable for PPIR in intermediate sketch domain.

Method	Backbone	Pre-train	mAR	R@100	R@500	R@1000	R@2000
Product-Patent	ResNet-18	Supervised	13.50	1.59	11.62	14.12	26.65
IDM	ResNet-18	Supervised	21.70	3.42	14.12	26.20	43.05
IDM+DeepPatent	ResNet-18	Supervised	21.98	3.19	13.90	25.74	45.10
IDAMA	ResNet-18	Supervised	22.84	4.10	15.03	26.42	45.79
Product-Patent	ResNet-50	Supervised	<u>18.05</u>	2.73	13.90	19.82	35.76
UCDIR	ResNet-50	UCDIR	18.64	2.87	14.03	20.17	35.81
UCDIR+IDM	ResNet-50	UCDIR	24.94	4.96	18.13	28.42	46.91
UCDIR+IDAMA	ResNet-50	UCDIR	25.36	5.31	18.47	28.97	47.28
IDM	ResNet-50	Supervised	25.06	5.24	18.45	28.93	47.61
IDM+DeepPatent	ResNet-50	Supervised	25.12	5.47	18.68	28.25	48.06
IDAMA	ResNet-50	Supervised	<u>25.63 (+7.58)</u>	5.47	19.13	29.84	48.06
IDAMA	Swin-B	Supervised	26.20	6.38	19.36	29.84	49.20
IDAMA	ViT-B	Supervised	26.43	6.83	20.05	30.30	48.52
IDAMA	Swin-L	Supervised	28.02	7.52	22.32	33.94	48.29
IDM	ViT-L	Supervised	26.99	6.83	22.55	31.44	47.15
IDAMA	ViT-L	Supervised	28.30	7.97	23.23	33.48	48.52
IDAMA	ViT-L	MAE [52]	23.35	5.24	17.31	28.02	42.82
IDAMA	ViT-L	IBOT [53]	**31.61**	**9.34**	**28.02**	**36.21**	**52.85**

the patent dataset D_p. We represent the labels as the correspondence between infringing products and patent images, forming a set $R = \{(x_i, y_{ij}) \mid i = 1, \dots, I = 439\}$. This work collected 439 product-patent image pairs to evaluate product-patent retrieval.

Evaluation metrics: Given the unsupervised formulation of PPIR tasks, we assess the retrieval performance by calculating the cosine similarity between patent and product image features. In PPIR tasks, one product or patent may contain more than one image, and choose to use the average similarity of all the product-image and patent-image pairs of the product or patent to denote the retrieval similarity in this pair. Moreover, PPIR contains 1-to-N situations, indicating that one product may infringe several patents, and we regard retrieving at least one patent as a successful retrieval.

We sort the patents in the retrieval pool according to their similarity scores and two metrics to systemically evaluate PPIR performance [51]: the mean Average Recall (mAR) and top-k recall at specific thresholds ($k = 100, 500, 1000, 2000$) (R@K) for each patent-product pair. mAR calculates the average recall rate of different top-k thresholds. AR considers the PPIR performance of different thresholds and can reflect the overall PPIR performance among other difficulties. Considering the 1-to-N matching situations, we use recall metric for evaluation. The R@K can demonstrate the model discrimination capability in specific scenarios.

4.2 Pre-training in Intermediate Domain

The capability of feature extractor $E_{(\cdot)}(\cdot)$ can seriously influence PPIR performance. Therefore, pre-training in the intermediate sketch-image domain may contribute to enhancing the representation and discrimination capability of the model, obtaining $E_{\text{sketch}}(\cdot)$ and improving PPIR performance. In this section, We compare the performance between supervised and unsupervised methods for feature learning in the intermediate domain. In this section, all the images, including images of both PPIRD-unlabeled $\mathcal{D}_{\text{unlabeled}}$ and ImageNet $\mathcal{D}_{\text{ImageNet}}$, are mapped to the intermediate domain using $M_{\text{sketch}}(\cdot)$ to extract sketch images for pre-training:

$$I'_{pre} = M_{sketch}(I_{pre}), \quad \forall I_{pre} \in \mathcal{D}_{unlabeled}, \mathcal{D}_{ImageNet}, \tag{13}$$

where I'_{pre} and I_{pre} denote the images of intermediate and raw domains, respectively.

Unsupervised Pre-training: Unsupervised pre-training has played a significant role in computer vision recently [54, 53, 55, 56, 57, 58, 52], enabling efficient visual representation encoders without

Table 2: **Quatitative Results of different domain mapping methods:** Compared with other mapping methods, IDM ('Product(Edge)-Patent(Edge)' is the more suitable mapping method for PPIRD and can bring more performance enhancement.

Method	mAR	R@100	R@500	R@1000	R@2000
Product-Patent	18.05	2.73	13.90	19.82	35.76
Product-Patent (Colorized by [27])	20.44	3.19	15.49	23.01	40.09
Product (Binary Line)-Patent	22.67	3.64	17.77	25.51	43.74
Product (Edge)-Patent	23.80	4.33	18.45	27.33	45.10
Product (Edge)-Patent (Edge)	**25.63**	**5.47**	**19.13**	**29.84**	**48.06**

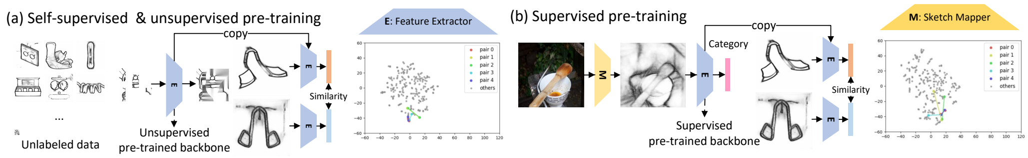

Figure 3: Comparison between self-/unsupervised pretraining and supervised pretraining strategies. **Subplot (a):** self-/unsupervised contrastive learning method 'IBOT' pre-trained on PPIRD-unlabeled, **Subplot (b):** Supervised pretraining on ImageNet1k-Edge. Matched product-patent image pairs are depicted using the same color. The visualization demonstrates that the unsupervised pretraining method effectively brings matched pairs closer in the feature space, enhancing their alignment.

the need for labeled training data. We utilize unsupervised pre-training methods of both Masked Image Modeling (MAE [52]) and Contrastive Learning (IBOT [53]) to obtain feature encoders for the intermediate domain by minimizing the unsupervised loss function $\mathcal{L}_{\text{unsup}}$:

$$\min_{E} \sum_{i=1}^{N} \mathcal{L}_{\text{unsup}} \left(E(I'_{pre}) \right), \tag{14}$$

where $E(\cdot)$ is the feature extractor and $\mathcal{L}_{\text{unsup}}$ is the unsupervised loss function specific to the pre-training method used. Images of PPIRD-unlabeled $D_{unlabeled}$ are used for unsupervised pre-training.

Supervised Pre-training: According to the conclusion of [59], supervised pre-training may be more suitable for learning representative and distinguishable features in the intermediate sketch-image domain due to the sparsity of this domain. Specifically, *Masked Image Modeling* or *Contrastive learning* may be challenging to reconstruct those large blank areas or discriminate blank areas from others, hindering the model's ability to learn representative and distinguishable feature representations. Considering this possibility, we conduct a simple yet effective supervised pre-training method by mapping the images in ImageNet and using the mapped well-annotated data for backbone pre-training. Based on these mapped sketch images, we train a classifier to obtain feature extractor $E(\cdot)$ using the original labels of these images. The training objective is to minimize the cross-entropy loss:

$$\mathcal{L} = -\sum_{i} y_i \log \hat{y}_i, \quad \text{with} \quad \hat{y}_i = \text{softmax}(F(I_{\text{sketch},i})), \tag{15}$$

where y_i is the ground truth label, and I'_{sketch} is the sketch images. The advantage of this approach is that our sketch feature extractor has an explicit learning objective and does not rely on negative samples in contrastive learning or masks in self-supervised learning. This endows feature extractor with excellent discriminative ability. More implemental details are provided in Sec. B.

4.3 Main Results

We compare IDAMA with raw ResNet [60] (denoted as 'Product-Patent') and the current SOTA patent retrieval method DeepPatent [5]. For a fair comparison, we add IDM before DeepPatent [5] since it is designed for patent image retrieval. Tab. 1 demonstrates the quantitative results on PPIRD. As is shown, IDAMA can bring significant performance enhancement compared with baseline methods, and both IDM and MAF can contribute to the improvement.

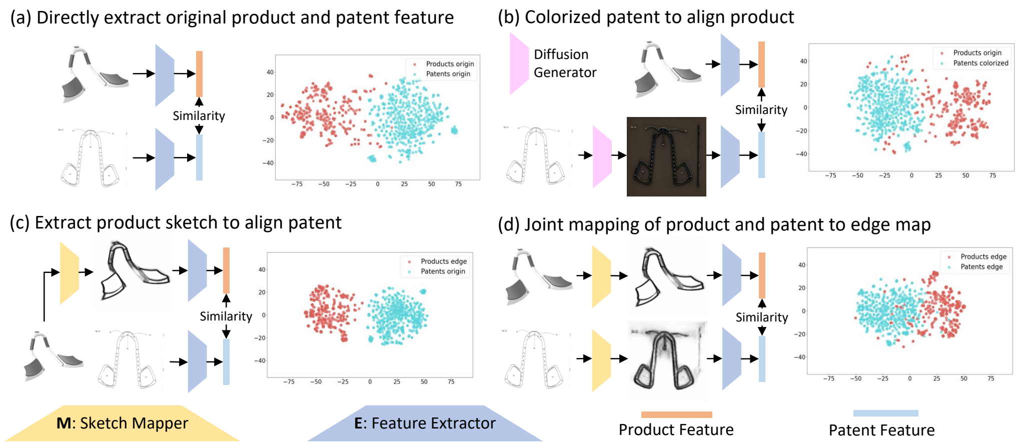

Figure 4: t-SNE results for different domain mapping methods: (a) *Product-Patent*: Directly extract original product and patent feature; (b) *Product-Patent (Colorized)*: Colorized patent to align product; (c) *Product (Binary Line)-Patent*: Extract product edge to align patent; (d) *IDM*: Extract both product images and patent images to sketch images. In each subplot, closer interleaving of the two colored points indicates a greater reduction in the domain gap, indicating better patent/product image alignment. IDM can best achieve the goal.

Comparison with DeepPatent: The comparison with DeepPatent [5] proves that PPIR (product-patent level retrieval) is quite different from Patent Image Retrieval (image leval retrieval) tasks. PPIR is more challenging because the product and the patent are not the same object. When comparing DeepPatent with MAF ('IDM+DeepPatent' and 'IDAMA' in Tab. 1), though the idea of MAF is more intuitive and even without requiring extra pre-training, it brings more performance enhancement, indicating the importance of integrating human beings' philosophy for challenging tasks.

Comparison of Different Pre-training methods: When comparing with different types of pre-training methods, the contrastive learning methods ('IBOT' in Tab. 1) may be the most suitable pre-training methods for PPIR tasks. We can notice that Masked Image Modeling methods ('MAE' in Tab. 1) perform worse than supervised learning, indicating these methods may not be suitable for such sketch domain lack of texture or color information. Fig. 3 reaches similar conclusion.

4.4 Comparison with Other Domain Mapping Approach

We compare IDM with some naive domain mapping approaches for PPIR:

Mapping colorful RGB product images into binary line drawings through Edge Detection (denoted as 'Product(Binary Line)-Patent'): We employ an Edge Detection Mapper $M_{\text{edge}}(\cdot)$ to extract edge from product images I_r, converting them into line drawings $I_{re} = M_{\text{edge}}(I_r)$. The advantage of this method is its computational efficiency and ease of implementation. However, the accuracy of the edge feature extraction algorithm is heavily dependent on it. Under varying lighting and texture conditions, the extracted edge maps from product images may exhibit discontinuities, affecting the performance of PPIR. Additionally, the feature extractor demonstrates in-robustness in extracting features from binary line-drawing images.

Converting patent images into RGB colorful images through diffusion (denoted as 'Product-Patent (Colorized)'): Recently, generative models [61, 62, 63] have gained increasing attention, with many models capable of efficiently generating RGB images based on line drawings [25, 26, 27]. We utilize a generative model $M_{\text{colorize}}(\cdot)$ [25] to transform patent images into pseudo-product images, $I_{pg} = M_{\text{colorize}}(I_p)$. The advantage of this method is that the generated images are in RGB format, allowing direct use of feature encoders pre-trained on natural images for feature extraction. However, challenges include the high computational cost of generative models and the lack of specific domain knowledge, which leads to the color distribution and texture of generated images may differ substantially from natural images.

We conduct extensive evaluations of the above three approaches using ResNet-50 [60], pre-trained on ImageNet-1K [6] (or ImageNet-1K-Edge), for feature extraction. According to the t-SNE [64] results in Fig. 4, IDM (denoted as 'Product (Edge)-Patent (Edge)', mapping patent images and product images into the intermediate sketch domain) is the best method to align patent/product images and alleviate the negative effects of domain gaps. Tab. 2 reaches a similar conclusion. We also conduct experiments by only mapping product images to directly sketch and compare their similarity with patent images. Experiment result 'Product (Edge)-Patent' in Tab. 2 proves the domain matching method is suboptimal for the tasks, indicating a domain gap between sketch and binary line drawings.

5 Conclusion

In summary, we formulate Patent-Product Image Retrieval (PPIR) as an open-set image retrieval task and propose a large-scale Patent-Product Image Retrieval Dataset (PPIRD) comprising (1) a testing set with 439 product-patent pairs (annotated and validated by experts with detailed descriptions) and a retrieval pool of 727,921 patents, and (2) an unlabeled pre-training set with 3,799,695 product/patent images. We propose a novel Intermediate Domain Alignment and Morphology Analogy (IDAMA) strategy tailored for PPIR. IDAMA contains an Intermediate Domain Mapping method to align binary line drawing patent images and colorful RGB product images by mapping them into an intermediate sketch domain using an edge detector to effectively mitigate the domain discrepancy and a Morphology Analogy Filter to select discriminative patent images containing distinctive visual feature of patents for efficient similarity comparison (inspired the cognitive principle—an unknown object can be described by analogy to a known object (patent image with high classification score regardless of label). Extensive experiments on PPIRD demonstrate that the intermediate domain is more suitable for aligning patent/product images and improving performance.

Acknowledgements

This work is supported in part by the Shenzhen Science and Technology Program (JCYJ20220818103001002), the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen, the Longgang District Special Funds for Science and Technology Innovation (LGKCSDPT2023002), the Project (No. 20232ABC03A25), the Scientific Research Program of Wuxi Health Commission (Project NO. Z202309).

References

[1] Dylan Myungchul Kang, Charles Cheolgi Lee, Suan Lee, and Wookey Lee. Patent prior art search using deep learning language model. In *Proceedings of the 24th Symposium on International Database Engineering & Applications*, pages 1–5, 2020.

[2] Kotaro Higuchi and Keiji Yanai. Patent image retrieval using transformer-based deep metric learning. *World Patent Information*, 74:102217, 2023.

[3] Alla Kravets, Nikita Lebedev, and Maxim Legenchenko. Patents images retrieval and convolutional neural network training dataset quality improvement. In *IV International research conference" Information technologies in Science, Management, Social sphere and Medicine"(ITSMSSM 2017)*, pages 287–293. Atlantis Press, 2017.

[4] Kotaro Higuchi, Yuma Honbu, and Keiji Yanai. Patent image retrieval using cross-entropy-based metric learning. In *Proc. of the 29th International Workshop on Frontiers of Computer Vision*, 2023.

[5] Michal Kucer, Diane Oyen, Juan Castorena, and Jian Wu. Deeppatent: Large scale patent drawing recognition and retrieval. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2309–2318, 2022.

[6] 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] Shiv Ram Dubey. A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5):2687–2704, 2021.

[8] Julian Risch, Nicolas Alder, Christoph Hewel, and Ralf Krestel. Patentmatch: a dataset for matching patent claims & prior art. *arXiv preprint arXiv:2012.13919*, 2020.

[9] Kehinde Ajayi, Xin Wei, Martin Gryder, Winston Shields, Jian Wu, Shawn M Jones, Michal Kucer, and Diane Oyen. Deeppatent2: A large-scale benchmarking corpus for technical drawing understanding. *Scientific Data*, 10(1):772, 2023.

[10] Mirac Suzgun, Luke Melas-Kyriazi, Suproteem Sarkar, Scott D Kominers, and Stuart Shieber. The harvard uspto patent dataset: A large-scale, well-structured, and multi-purpose corpus of patent applications. *Advances in Neural Information Processing Systems*, 36, 2024.

[11] Nikos Efthymiadis, Giorgos Tolias, and Ondřej Chum. Edge augmentation for large-scale sketch recognition without sketches. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 3595–3602. IEEE, 2022.

[12] Ralf Krestel, Renukswamy Chikkamath, Christoph Hewel, and Julian Risch. A survey on deep learning for patent analysis. *World Patent Information*, 65:102035, 2021.

[13] Homaira Huda Shomee, Zhu Wang, Sathya N Ravi, and Sourav Medya. A comprehensive survey on ai-based methods for patents. *arXiv preprint arXiv:2404.08668*, 2024.

[14] Yonghe Lu, Xin Xiong, Weiting Zhang, Jiaxin Liu, and Ruijie Zhao. Research on classification and similarity of patent citation based on deep learning. *Scientometrics*, 123:813–839, 2020.

[15] Selen Yücesoy Kahraman, Türkay Dereli, and Alptekin Durmuşoğlu. Forty years of automated patent classification. *International Journal of Information Technology & Decision Making*, pages 1–32, 2023.

[16] Ran Li, Wangke Yu, Qianliang Huang, and Yuying Liu. Patent text classification based on deep learning and vocabulary network. *International Journal of Advanced Computer Science and Applications*, 14(1), 2023.

[17] Hamid Bekamiri, Daniel S Hain, and Roman Jurowetzki. Patentsberta: A deep nlp based hybrid model for patent distance and classification using augmented sbert. *Technological Forecasting and Social Change*, 206:123536, 2024.

[18] Eleni Kamateri, Michail Salampasis, and Eduardo Perez-Molina. Will ai solve the patent classification problem? *World Patent Information*, 78:102294, 2024.

[19] L Siddharth, Guangtong Li, and Jianxi Luo. Enhancing patent retrieval using text and knowledge graph embeddings: a technical note. *Journal of Engineering Design*, 33(8-9):670–683, 2022.

[20] Hao-Cheng Lo, Jung-Mei Chu, Jieh Hsiang, and Chun-Chieh Cho. Large language model informed patent image retrieval. *arXiv preprint arXiv:2404.19360*, 2024.

[21] Chenhe Dong, Yinghui Li, Haifan Gong, Miaoxin Chen, Junxin Li, Ying Shen, and Min Yang. A survey of natural language generation. *ACM Computing Surveys*, 55(8):1–38, 2022.

[22] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[23] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[24] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[25] Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036*, 2024.

[26] Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, and Yi-Zhe Song. You'll never walk alone: A sketch and text duet for fine-grained image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16509–16519, 2024.

[27] Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, and Yi-Zhe Song. Text-to-image diffusion models are great sketch-photo matchmakers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16826–16837, 2024.

[28] Hyeonseob Nam, HyunJae Lee, Jongchan Park, Wonjun Yoon, and Donggeun Yoo. Reducing domain gap by reducing style bias. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8690–8699, 2021.

[29] Liuliu Xu, Haifan Gong, Yun Zhong, Fan Wang, Shouxin Wang, Lu Lu, Jinru Ding, Chen Zhao, Wenchao Tang, and Jie Xu. Real-time monitoring of manual acupuncture stimulation parameters based on domain adaptive 3d hand pose estimation. *Biomedical Signal Processing and Control*, 83:104681, 2023.

[30] Senmao Wang, Haifan Gong, Runmeng Cui, Boyao Wan, Yicheng Liu, Zhonglin Hu, Haiqing Yang, Jingyang Zhou, Bo Pan, Lin Lin, et al. Costal cartilage segmentation with topology guided deformable mamba: Method and benchmark. *arXiv preprint arXiv:2408.07444*, 2024.

[31] Haifan Gong, Yu Lu, Xiang Wan, and Haofeng Li. Domain generalized medical landmark detection via robust boundary-aware pre-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 3140–3148, 2025.

[32] Haifan Gong, Huixian Liu, Yitao Wang, Xiaoling Liu, Xiang Wan, Qiao Shi, and Haofeng Li. Fetal cerebellum landmark detection based on 3d mri: Method and benchmark. *IEEE Journal of Biomedical and Health Informatics*, 2025.

[33] Conghui Hu, Can Zhang, and Gim Hee Lee. Unsupervised feature representation learning for domain-generalized cross-domain image retrieval. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11016–11025, 2023.

[34] Bin Li, Ye Shi, Qian Yu, and Jingya Wang. Unsupervised cross-domain image retrieval via prototypical optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 3009–3017, 2024.

[35] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

[36] Zihang Xu, Haifan Gong, Xiang Wan, and Haofeng Li. Asc: Appearance and structure consistency for unsupervised domain adaptation in fetal brain mri segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 325–335. Springer, 2023.

[37] Pedro O Pinheiro. Unsupervised domain adaptation with similarity learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8004–8013, 2018.

[38] Yunhao Ge, Yao Xiao, Zhi Xu, Xingrui Wang, and Laurent Itti. Contributions of shape, texture, and color in visual recognition. In *European Conference on Computer Vision*, pages 369–386. Springer, 2022.

[39] Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, and Yi-Zhe Song. Picture that sketch: Photorealistic image generation from abstract sketches. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 6850–6861, 2023.

[40] Andrey Voynov, Kfir Aberman, and Daniel Cohen-Or. Sketch-guided text-to-image diffusion models. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023.

[41] Caixia Zhou, Yaping Huang, Mengyang Pu, Qingji Guan, Ruoxi Deng, and Haibin Ling. Muge: Multiple granularity edge detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25952–25962, 2024.

[42] Irving Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.

[43] David L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[44] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[45] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. In *Constructive Approximation*, volume 28, pages 253–263, 2008.

[46] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

[47] Yaodong Yu, Sam Buchanan, Druv Pai, Tianzhe Chu, Ziyang Wu, Shengbang Tong, Hao Bai, Yuexiang Zhai, Benjamin D Haeffele, and Yi Ma. White-box transformers via sparse rate reduction: Compression is all there is? *Journal of Machine Learning Research*, 25(300):1–128, 2024.

[48] William W Hager. Lipschitz continuity for constrained processes. *SIAM Journal on Control and Optimization*, 17(3):321–338, 1979.

[49] Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 2009.

[50] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[51] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. Aggregating local image descriptors into compact codes. In *European Conference on Computer Vision*, pages 776–789. Springer, 2010.

[52] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[53] Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Yujun Hou, Wenliang Jiang, Tianjun Gong, Yu Qiao, Thomas Y Hui, Ziwei Liu, et al. ibot: Image bert pre-training with online tokenizer. In *International Conference on Learning Representations*, 2022.

[54] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[55] 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.

[56] Haifan Gong, Guanqi Chen, Sishuo Liu, Yizhou Yu, and Guanbin Li. Cross-modal self-attention with multi-task pre-training for medical visual question answering. In *Proceedings of the 2021 international conference on multimedia retrieval*, pages 456–460, 2021.

[57] Yuxin Fang, Wen Wang, Binhui Xie, Quan Sun, Ledell Wu, Xinggang Wang, Tiejun Huang, Xinlong Wang, and Yue Cao. Eva: Exploring the limits of masked visual representation learning at scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19358–19369, 2023.

[58] Yuxin Fang, Quan Sun, Xinggang Wang, Tiejun Huang, Xinlong Wang, and Yue Cao. Eva-02: A visual representation for neon genesis. *Image and Vision Computing*, 149:105171, 2024.

[59] Wenxuan Li, Alan Yuille, and Zongwei Zhou. How well do supervised models transfer to 3d image segmentation. In *The Twelfth International Conference on Learning Representations*, volume 1, 2024.

[60] 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.

[61] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020.

[62] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023.

[63] Haifan Gong, Yitao Wang, Yihan Wang, Jiashun Xiao, Xiang Wan, and Haofeng Li. Diffuse-uda: Addressing unsupervised domain adaptation in medical image segmentation with appearance and structure aligned diffusion models. *arXiv preprint arXiv:2408.05985*, 2024.

[64] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne, 2008.

[65] 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. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[66] 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. In *ICLR*, 2021.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims exactly reflects the paper's contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are discussed in the last section of this paper.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We have provide the full set of assumptions and a complete proof.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We have provide all the information needed to reproduce the results in this paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The data are publicly available, while the code is available in the submission system.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All these details are available in the paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We have conducted the t-test in the comparison of our method and the previous methods.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have provided all these information in this paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This work follows the the NeurIPS code of ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We have discuss the broader impact in the discussion section.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We have cited all related paper in this work.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The new assets introduced in the paper are well documented.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: We use publicity available dataset.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This work uses the public available dataset as they already received the IRB approvals.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Proofs for Section 3.3

A.1 Proof of Claim 2

Let $x := s_p - s_{pr}$. By assumption, x is $2k$-sparse and $\|x\|_2 \le \varepsilon$. Because $\mathbf{A}$ satisfies $\mathrm{RIP}(2k, \delta)$,

$$\|\mathbf{A}x\|_2 \le \sqrt{1+\delta}\,\|x\|_2 \le (1+\delta)\,\varepsilon. \tag{16}$$

Since $\mathbf{A}$ is linear, $\mathbf{A}x = \mathbf{A}s_p - \mathbf{A}s_{pr}$, which concludes the proof. $\square$

A.2 Proof of Theorem 1

Decompose each image as $I = s + \eta$ and observe

$$\mathbf{A}I_p - \mathbf{A}I_{pr} = (\mathbf{A}s_p - \mathbf{A}s_{pr}) + (\mathbf{A}\eta_p - \mathbf{A}\eta_{pr}). \tag{17}$$

Applying the backbone E and using its L-Lipschitz continuity [48],

$$\begin{aligned} \|\tilde{z}_p - \tilde{z}_{pr}\|_2 &= \|E(\mathbf{A}I_p) - E(\mathbf{A}I_{pr})\|_2 \\ &\le L\,\|\mathbf{A}I_p - \mathbf{A}I_{pr}\|_2 \\ &\le L\big(\|\mathbf{A}s_p - \mathbf{A}s_{pr}\|_2 + \|\mathbf{A}\eta_p\|_2 + \|\mathbf{A}\eta_{pr}\|_2\big). \end{aligned} \tag{18}$$

The first term is bounded by Claim 2. For the nuisance terms we appeal to the concentration of the χ^2 distribution of Gaussian projections [46, Thm. 7.5]: for any fixed vector η and all $t \in (0,1)$,

$$\Pr\Big(\big|\|\mathbf{A}\eta\|_2^2 - \tfrac{m}{n}\|\eta\|_2^2\big| \ge t\tfrac{m}{n}\|\eta\|_2^2\Big) \;\le\; 2\exp\!\big(-ct^2 m\big), \tag{19}$$

where $c > 0$ is an absolute constant. Selecting $t = \frac{1}{2}$ and using $\sqrt{a \pm b} \le \sqrt{a} \pm b/(2\sqrt{a})$ gives, with probability at least $1 - 2e^{-cm/4}$,

$$\|\mathbf{A}\eta\|_2 \le \frac{m}{n}\|\eta\|_2. \tag{20}$$

Applying this bound to both η_p and η_{pr} and combining all inequalities establishes the asserted result. $\square$

B Implemental Details

The τ_{maf} is set to 0.4. We directly use the Supervised Pre-trained ResNet-50 on ImageNet-1K as the Morphology Analogy Filter. For all the supervised pre-training models (including ResNet [60], SwinTransformer [65], and ViT [66]), the pre-training dataset is ImageNet-1K-Edge (Edge is extracted using [25]) the optimizer is AdamW, the learning rate is set to 0.002, the beta of the optimizer is set to (0.9, 0.95), and the weight decay is set to 0.001. We train all the models for 100 epochs. For all the self-supervised/unsupervised pretraining, the pre-training dataset is PPIR-unlabeled-Edge (Edge is extracted using [25]), the batch size is 256, the optimizer is AdamW, the learning rate is 0.002. We pre-train all the models for 100 epochs. The batch size is set from 256 to 1024, with more details are available in the supplementary material.

C Appendix on Training Details

Table 3: Hyper-parameters for supervised training. Values separated by "/" correspond to the successive model sizes listed in the **Model** row.

Hyper-parameter	**ResNet**	**Swin**	**ViT**
Model	ResNet-18 / 50 / 101	Swin-T / S / B	ViT-B-16 / L-16
Dataset	ImageNet1K-edge	ImageNet1K-edge	ImageNet1K-edge
Batch size	4096 / 2048 / 1024	2048 / 1024 / 512	512 / 256
Optimizer	AdamW	AdamW	AdamW
Learning rate	0.002	0.002	0.002
β (AdamW)	(0.9, 0.95)	(0.9, 0.95)	(0.9, 0.95)
Epochs	100	100	100
Weight decay	0.001	0.001	0.001
Training time (h)	5.3 / 9.5 / 18.7	7.9 / 15.7 / 21.6	19.8 / 34.7

Table 4: Hyper-parameters for self-supervised learning.

Hyper-parameter	**iBOT**	**MAE**
Method	iBOT	MAE
Model	ViT-L-16	ViT-L-16
Dataset	PPIR-unlabeled	PPIR-unlabeled
Batch size	256	256
Optimizer	AdamW	AdamW
Learning rate	0.002	0.002
β (AdamW)	(0.9, 0.95)	(0.9, 0.95)
Epochs	100	100
Training time (h)	97.4	95.3

D Visualization Results

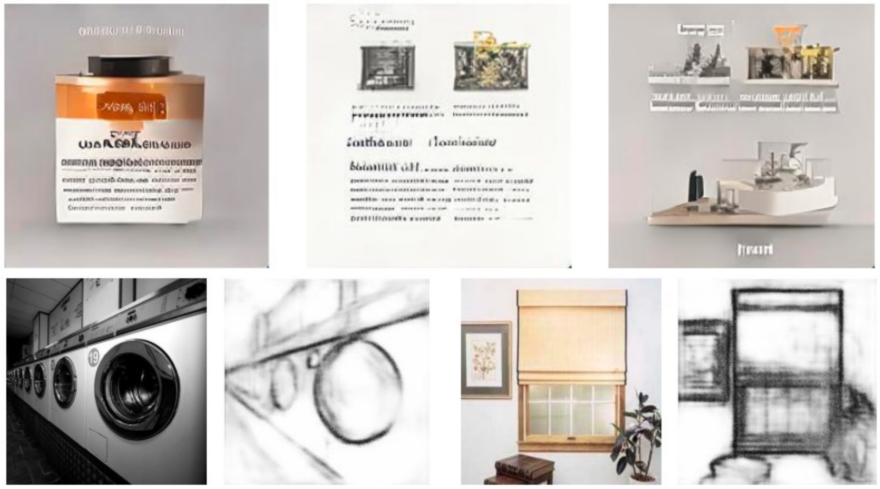

Figure 5: **Upper:** Diffusion generated colorful patent images
Lower: Sketches extracted from raw images

E Broader impact, Advantages, and Limitations

Broader Impact. Our work lowers the barrier for large–scale, automated monitoring of potential design infringements, thereby helping both innovators and regulators to protect intellectual property more efficiently and at earlier stages of the product life-cycle. The publicly released PPIRD dataset also creates a new benchmark that can stimulate research on cross-domain retrieval, representation learning under distribution shifts, and open-set recognition. Nevertheless, any technology that simplifies infringement search can also be misused: malicious actors might automate "design-around" strategies or target small businesses with aggressive litigation. We therefore encourage downstream users to combine PPIR systems with human expert review, publish detailed audit logs, and follow responsible-AI frameworks to mitigate unintended consequences.

Advantages. (1) **Comprehensive benchmark.** PPIRD is, to our knowledge, the largest and most realistic testbed for patent–product image retrieval, containing nearly one million patent drawings and millions of unlabeled images for pre-training. (2) **Effective domain bridging.** The proposed IDAMA pipeline explicitly maps both binary line drawings and RGB product photos into an intermediate sketch space, reducing domain discrepancy without requiring paired supervision. (3) **Open-set readiness.** By framing PPIR as an open-set problem and adding a morphology-analogy filter, our method retains high recall on novel object categories that are unseen during conventional pre-training. (4) **Plug-and-play.** IDAMA is architecture-agnostic; it can be integrated into existing CNN or vision-transformer backbones with minimal code changes and no task-specific annotations.

Limitations. The morphology-analogy filter relies on hand-crafted similarity thresholds; tuning them across industries or jurisdictions may require domain expertise. In addition, our dataset—while large—covers only 439 manually verified product–patent pairs, limiting fine-grained evaluation of failure modes such as functional equivalence or partial design overlap.