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

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

Problem:

Patent-Product Image Retrieval (PPIR) is underexplored despite advances in AI for image retrieval.

Goal:

Retrieve patent images given product photos to flag potential infringements.

Challenges:

Many artificial-object categories; pretrained models struggle with unseen objects.
Large domain gap: binary patent line drawings vs. RGB product photos.

Our setup and data (PPIRD):

- Open-set image retrieval to mirror real-world conditions.
- Test: 439 product–patent pairs.
- Retrieval pool: 727,921 patent images.
- Unlabeled pre-training: 3,799,695 product/patent images.
- Detailed product descriptions to aid infringement verification.

Fig. 1 Characteristics of Patent-Product Image Retrieval (PPIR) task

Method

IDAMA (Intermediate Domain Alignment and Morphology Analogy)

Intermediate Domain Mapping (IDM): map both image types to a sketch domain via edge detection to reduce domain discrepancy.

Morphology Analogy Filter (MAF): select discriminative patent images using high-confidence visual analogies, boosting retrieval on unseen artificial objects.

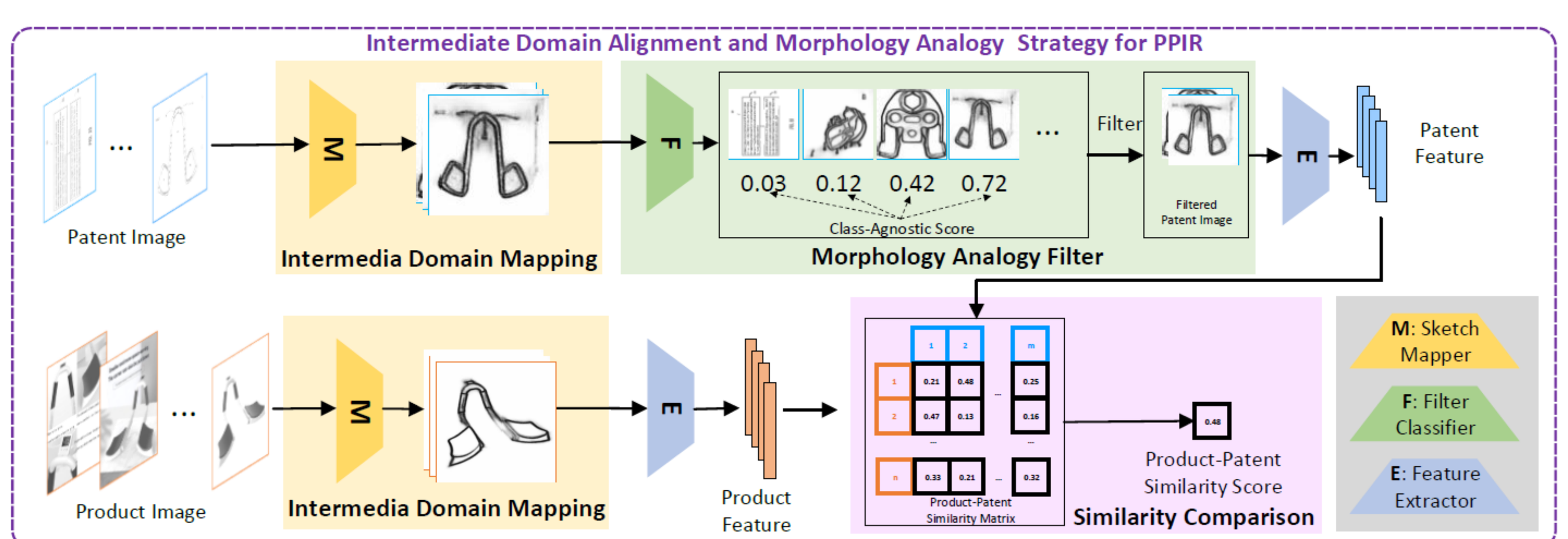

Fig.2 Pipeline of IDAMA: IDAMA consists of Intermediate Domain Mapping and Morphology Analogy Filter methods. The pipeline of IDAMA is as follows: 1) Using IDM to align product/patent images by mapping them into intermediate sketch domains; 2) Using MAF to filter discriminative mapped patent images; 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.

Theory Analysis 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.*

Results

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	18.05	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	25.63 (+7.58)	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**

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**

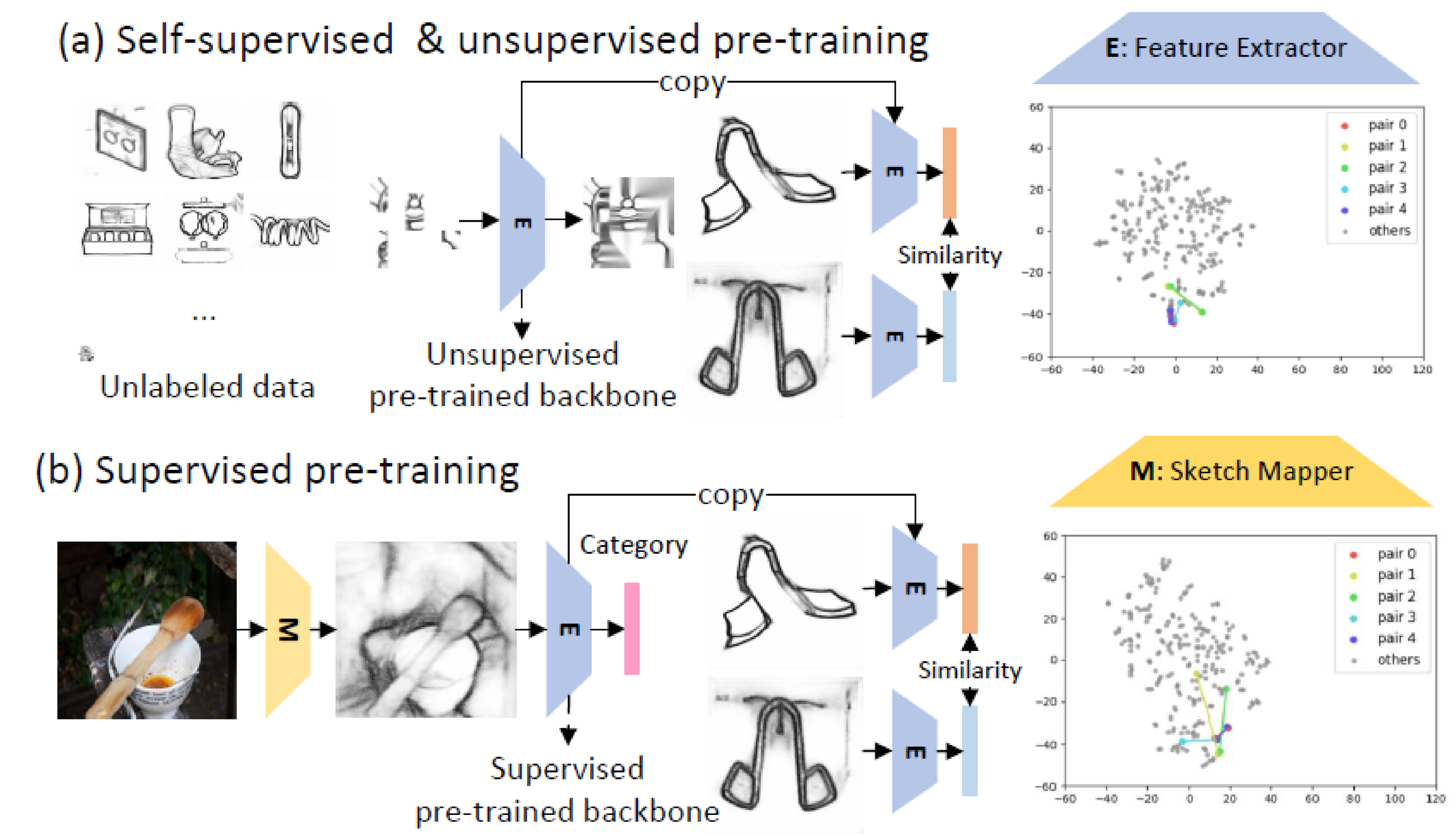

Fig.3 Comparison between self-/unsupervised pretraining and supervised pretraining strategies. Subplot (a): self-/unsupervised contrastive learning method `IBOT' pre-trained on PPIRD-unlabeled, Subplot (b): Supervised pretraining on ImageNet1k-Edge. Matched product-patent image pairs are depicted using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

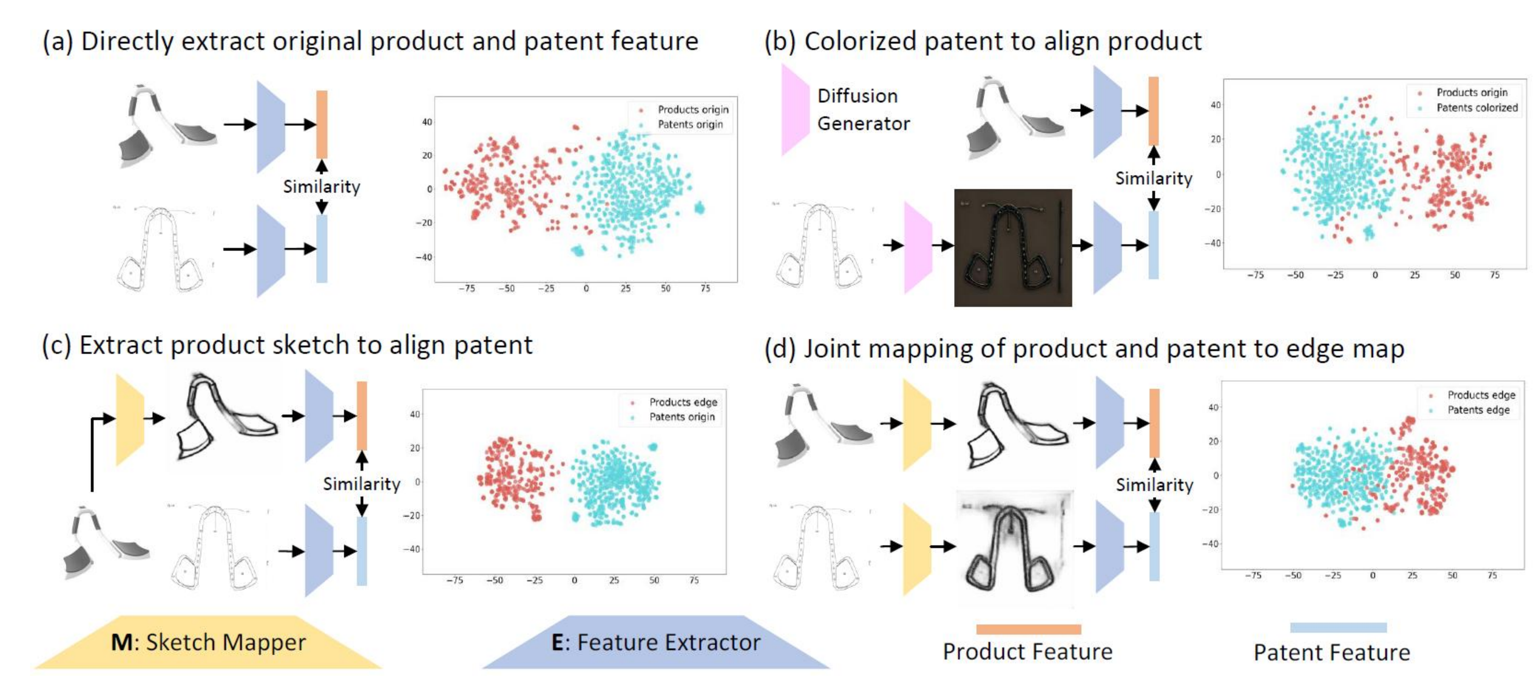

Fig. 4 t-SNE results for different domain mapping methods: (a) Product-Patent: Directly extract original product and patent feature; (b) Product-Patent (Colorized): Colorized patent to align product; (c) Product (Binary Line)-Patent: Extract product edge to align patent; (d) IDM: Extract both product 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.