Patent Image Retrieval Using Cross-entropy-based Metric Learning

Kotaro Higuchi, Yuma Honbu, and Keiji Yanai

Department of Informatics, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585, Japan
{higuchi-k,honbu-y,yanai}@mm.inf.uec.ac.jp

Abstract. Intellectual property work covers a wide range of areas. In particular, prior art literature searching in the patent field requires finding documents that can be used to determine novelty and inventive steps from a vast amount of past literature. Concerning this search practice, research and development of a drawing search technology that directly searches drawings, and essential information about inventions, has long been desired. However, patent drawings are described as black-and-white abstract drawings, and their modal characteristics are very different from those of natural images, so they have yet to be explored. This study achieved higher accuracy than the previous one by introducing InfoNCE and ArcFace in the DeepPatent dataset instead of the conventional Triplet. In addition, we developed an application that enables users to search for patent drawings using any images. Our architecture can be applied to patent drawings and many other modal-like drawings, such as mechanical drawings, design patents, trademarks, diagrams, and sketches.

Keywords: Patent Image Retrieval · Metric Learning · Search Application

1 Introduction

The patent field has developed a combination of natural language processing that can easily interact with textual information and patent classification information. High-quality search has been a long-standing challenge in patent practice since patent search requires a person skilled in the technical field, both in making queries and in assigning classifications.

Since the advent of ResNet [6], research on image recognition of natural images has made dramatic progress. However, the development of patent drawing retrieval, which is described as an abstract drawing, has been challenging, and no de facto method or system has yet to emerge [1].

In this study, we develop a search application based on metric learning, which has rapidly developed in recent years, and demonstrate patent drawing retrieval. In particular, for metric learning, we employ cross-entropy-based methods such as InfoNCE [16] and ArcFace [3] instead of the conventional Triplet loss.

Briefly, our main contributions are to:

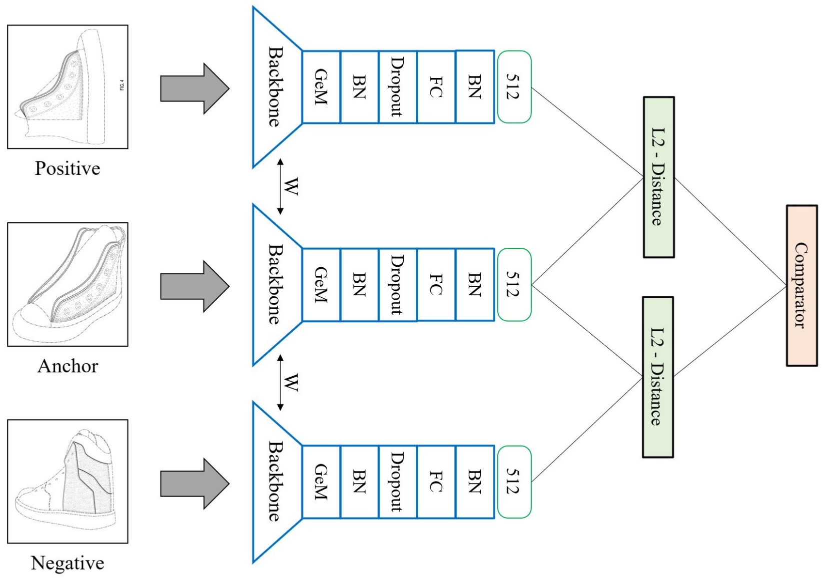

Fig. 1: The proposed architecture. Positive and Anchor are from USD0811075S1; Negative is from USD0811070S1. Both diagrams are cited from the drawings included in DeepPatent [11].

- achieve higher accuracy than previous papers by using cross-entropy-based metric learning methods on a dataset of patent drawings, which has been challenging to achieve in the past.
- develop an application that can search patent drawings using the proposed model (see Figure 1). There has yet to be a successful example of such an application in the past.

2 Related Works

Conventional Patent Searches and Issues Conventionally, patent searches have been performed using various search tools such as EspaceNet [4], making full use of text or patent classification. An example of an operation screen is shown in Figure 2.

However, conventional patent searches are based on the following two assumptions: (1) in the past, the annotator has assigned an appropriate classification to the patent, and (2) the searcher has learned appropriate terms (textual queries) that capture the essence of the patent.

In other words, both the person who assigns the classification and the person who performs the search need to be familiar with the technical field. In addition, the accuracy of patent classification is less than 100% due to the nature of cutting-edge technology being applied. Furthermore, the assignment of clas-

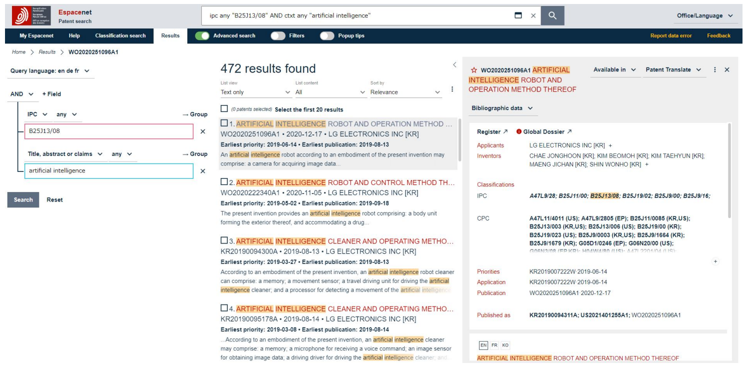

Fig. 2: EspaceNet. Users can perform text-based searches using the bibliography of the invention, the name of the invention, and the patent classification.

sifications and the creation of queries can sometimes be challenging, even for intellectual property specialists.

In conventional patent search, there is a need to search for shapes that are difficult to be expressed in words. However, there are many cases where such needs are challenging to be solved by text queries, and people concerned have to visually check several thousand to several tens of thousands of patent drawings.

Research in computer vision Because of the situation mentioned above, some research has been conducted on patent drawing retrieval based on deep learning techniques [9] [18]. However, the network in this study performed a primary task of estimating eight types of International Patent Classification (IPC, from A to H) and an auxiliary task of classifying nine types of drawings. Therefore, the image representation was strongly related to the existing patent classification, and the problem of omission of classification remained.

In addition, since the same IPC label is assigned to all drawings in the same application, there is an inherent issue of acquiring a sparse image representation. Kucer et al. [11] showed that patent image retrieval is possible by distance learning using ResNet50 [6] as the backbone, together with the DeepPatent dataset described below. However, there was still room for proof-of-concept for whether the model obtained in the paper could be developed as an actual application. Triplet depends on the sample selection in the batch; therefore, it is unpredictable whether a high accuracy can be obtained. For this reason, in this study, we experimented with a cross-entropy-based method using Triplet as a baseline.

Significance of drawings in patent practice Patent drawings are essential in practice [14], and in Japan, there are many cases in which patent drawings are the focus of infringement judgments. There are also precedents indicating that the composition (e.g. shape) can be read from the drawings [8]. Considering the above, the importance of drawing searches in patent practice is high.

3 Methods

We use the architecture shown in Figure 1 for a patent image retrieval system as a baseline for the DeepPatent dataset [11]. The Triplet network [2] [7] consists of three instances (with shared parameters) of the same forward propagation network. The system takes three samples from a batch, the network computes two distances between Anchor and Positive, and between Anchor and Negative. It calculates distances and adds a margin of m to update the model for using the three samples simultaneously.

Triplet method, however, has a drawback that it can handle one positive pair and one negative pair at the same time. This makes it difficult to use negative pairs effectively which exist much more than positive pairs in general. To resolve this problem, in this work, we use cross-entropy-based methods which can train many negative pairs at the same time. As cross-entropy-based methods, we employ InfoNCE [16] and ArcFace [3].

3.1 InfoNCE

InfoNCE [16] is one of the most popular methods used in self-supervised learning. Unlike Triplet, InfoNCE uses many Anchor-Negative pairs for each Anchor-Positive pair in sampling within a batch. The loss function L_i of infoNCE is shown below.

$$L_i = -\log \frac{e^{q\cdot k_+/\tau}}{e^{q\cdot k_+/\tau} + \sum_{i=0}^{K} e^{q\cdot k_i/\tau}}$$

3.2 ArcFace

ArcFace [3] achieves metric learning that increases the variance between classes by adding the following to the Softmax Cross-Entropy loss function used in the classification problem: 1) normalization of weights and features, and 2) a margin for the correct class. The loss function L_a of ArcFace is show below.

$$L_a = -\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1,j\neq y_i}^{N} e^{s\cos\theta_j}}$$

4 Experiments

4.1 DeepPatent Dataset

Kucer et al. [11] published a dataset containing more than 350,000 U.S. Design Patents in the public domain. The dataset is available from the project's Google Drive, and each drawing is assigned a publication number and a drawing number. Table 1 shows the details, which consist of 45,000 cases, divided into 70% as train, 15% as test, and 15% as validation.

The published DeepPatent dataset is not subject to copyright restrictions and is in the public domain, as stated by the U.S. Patent and Trademark Office (USPTO) [20].

Based on the above, we adopted the DeepPatent dataset as our benchmark for patent image retrieval.

Table 1: DeepPatent dataset

DeepPatent	figures	classes
Train	254,787	33,364
Test	38,834	6,927
Validation	44,815	5,888

4.2 Evaluation index

To evaluate the retrieval system, we used mAP score. mAP is the average precision APs from computing each query. Note that the AccuracyCalculator of the Pytorch Metric Learning [15] was used to implement the accuracy measurement.

4.3 Comparison to baseline

In the architecture shown in Figure 1, we compare baseline and metric learning with the proposed methods. Specifically, we made experiments with the baseline ResNet+Triplet, the EffNet [19]+Triplet, the EffNet+infoNCE [16], and the EffNet+ArcFace [3]. We used Hard Negative Mining [21] for Triplet training.

Table 2: mAP Score Comparison

Method	mAP
DeepPatent baseline [11]	0.379
EffNet + InfoNCE [16]	0.447
EffNet + ArcFace [3]	**0.622**

5 Results and Discussions

5.1 Comparison to baseline

Table 2 shows the results of the comparison between the baseline and the proposed method for the architecture shown in Figure 1.

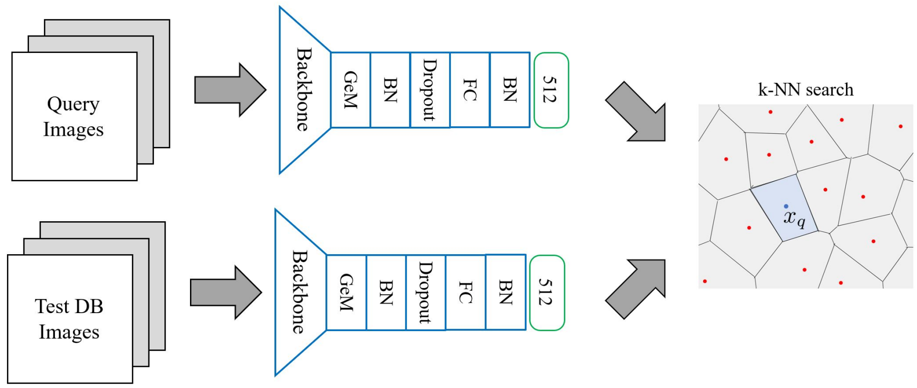

Fig. 3: Inference and retrieval architecture. The system performs indexing by Faiss.

The proposed method achieves a score of mAP=0.622. This score is higher than the conventional method because we select a more appropriate backbone and method for the patent drawing task (more complex than ResNet in this task). In addition, Triplet uses only the m parameter of margin for metric learning, which is easy to control, and the amount of VRAM memory usage is relatively small (within a few GB at batch size=512). Furthermore, the infoNCE can achieve a higher score because the Negative sample is larger than the Triplet. With the ArcFace method, we achieve a value that stands out above.

5.2 Search Application and Practitioners' Opinions

We implemented a patent drawing search application using the indexing method, Faiss [10], as shown in Figure 3. The index file size was about 80 MB, which was large enough to deploy on a server or in the cloud.

We implemented the system as a Web application using Streamlit on an on-premise Ubuntu server. Figure 4 shows the application screenshots.

Afterward, we received feedback from practitioners on the usability of the application. Overall, the feedback was favorable, and we have received positive comments from the users, who look forward to future research development. The following comments are parts of the feedback we received.

- The operation is more straightforward than expected. However, the accuracy still needs to be up to practical use.
- The search is easy because it is possible to search only by image query.

6 Conclusion

With the proposed method, we achieved an mAP score higher than that of previous papers and also realized a patent drawing retrieval application. For

(a) Shoes

(b) SmartPhone Case

Fig. 4: The screenshots of the developed patent image search system. Users can search for patent drawings by dragging and dropping any images. (a) Search results for a shoe. The search result is good and hits similar patent drawings of shoes. (b) Search results for a smartphone case. The search results are questionable, and future improvements are needed in search accuracy.

many years, numerous issues in the patent field have required the appearance of patent drawings, and our proposal will help solve these issues.

This research has revealed two significant possibilities: first, deep metric learning is possible on patent drawing datasets, and second, by combining a machine learning framework and trained models, a patent drawing application can be developed.

Future work Future work is to improve the mAP score and search accuracy. Where feedback from practitioners has revealed that the accuracy could be better at the level of practice, we recognize that there is a large room for growth in accuracy, as there are various other methods for backbone and metric learning. Examples include SwinTransformer [13] [12], and other losses [5] [17].

The above results show that patent image retrieval has many unexplored areas. A wide range of research is expected to be conducted in the future because it can be applied not only to patent drawings but also to similar modal drawings, sketches, trademarks, designs, utility models, mechanical drawings, and flowcharts.

References

1. Bhattarai, M., Oyen, D., Castorena, J., Yang, L., Wohlberg, B.: Diagram image retrieval using sketch-based deep learning and transfer learning. In: CVPR. pp. 174–175 (2020)
2. Cao, R., Zhang, Q., Zhu, J., Li, Q., Li, Q., Liu, B., Qiu, G.: Enhancing remote sensing image retrieval using a triplet deep metric learning network. International Journal of Remote Sensing **41**(2), 740–751 (2020)
3. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: CVPR. pp. 4685–4694 (2019)
4. European Patent Office: Espacenet (2023), `https://worldwide.espacenet.com/`
5. Gordo, A., Almazan, J., Revaud, J., Larlus, D.: End-to-end learning of deep visual representations for image retrieval. International Journal of Computer Vision **124**(2), 237–254 (2017)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
7. Hoffer, E., Ailon, N.: Deep metric learning using triplet network. In: Similarity-Based Pattern Recognition. pp. 84–92. Springer International Publishing (2015)
8. Intellectual Property High Court: Case number 2014(ke)10274 (2014), `https://www.ip.courts.go.jp/app/hanrei_jp/detail?id=4185`
9. Jiang, S., Luo, J., Pava, G., Hu, J., Magee, C.: A convolutional neural network-based patent image retrieval method for design ideation. In: IDETC-CIE. vol. 83983. American Society of Mechanical Engineers (2020)
10. Johnson, J., Douze, M., Jegou, H.: Billion-scale similarity search with gpus. Proc. of IEEE Transactions on Big Data **7**(03), 535–547 (2021)
11. Kucer, M., Oyen, D., Castorena, J., Wu, J.: Deeppatent: Large scale patent drawing recognition and retrieval. In: WACV. pp. 2309–2318 (January 2022)
12. Liu, Z., *et al.*: Swin transformer: Hierarchical vision transformer using shifted windows. ICCV pp. 9992–10002 (2021)
13. Liu, Z., *et al.*: Swin transformer v2: Scaling up capacity and resolution. In: CVPR. pp. 12009–12019 (2022)
14. Ministry of Justice: Japanese law translation, patent act, article70(2) (2023), `https://www.japaneselawtranslation.go.jp/en/laws/view/4097`
15. Musgrave, K., Belongie, S., Lim, S.: A metric learning reality check. In: ECCV. Lecture Notes in Computer Science, vol. 12370, pp. 681–699. Springer (2020)
16. Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
17. Revaud, J., Almazan, J., Rezende, R., de Souza, C.: Learning with average precision: Training image retrieval with a listwise loss. In: ICCV (2019)
18. Shalaby, W., Zadrozny, W.: Patent retrieval: A literature review. Knowledge and Information Systems **61**, 631–660 (2019)
19. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: ICML. pp. 6105–6114 (2019)
20. United States Patent and Trademark Office: Terms of use for uspto websites (2023), `https://www.uspto.gov/terms-use-uspto-websites`
21. Xuan, H., Stylianou, A., Liu, X., Pless, R.: Hard negative examples are hard, but useful. In: ECCV. pp. 126–142. Springer (2020)