

# 深層距離学習の特許図面検索への適用

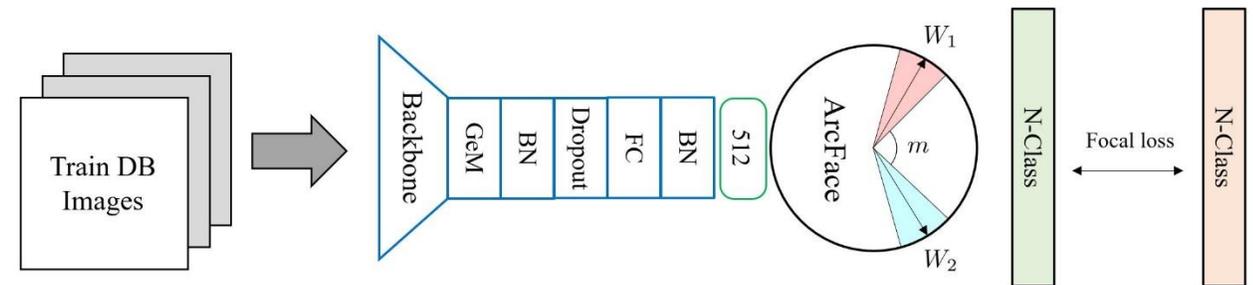
樋口幸太郎 柳井啓司

電気通信大学

PRMU 2023.03

# Abstract

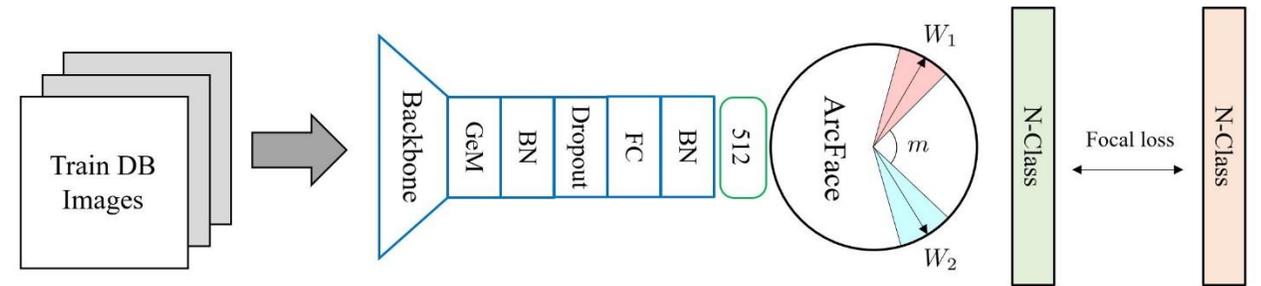
- 特許検索
  - 膨大な先行技術文献から、新規性・進歩性に係る文献をヒットさせる必要
- 特許図面は白黒の抽象的な図
  - 特許図面検索は、長年の課題
- 従来のTripletの距離学習手法
  - 最新のInfoNCE, ArcFaceを適用
- DeepPatent datasetにおいてSOTA達成



提案アーキテクチャ(後述)

# 1. Introduction

- 本論文の主な貢献は次の通り.  
特許図面データセットにおいて,
  1. 既存の距離学習手法をベンチマークし,  
有効性を評価する.
  2. 従来性能を上回るアーキテクチャ,  
及び, SOTA手法を提案する.



提案アーキテクチャ(後述)

# 2. Related work

## 従来の特許検索と課題

- Patent Public Searchのような様々な検索ツール

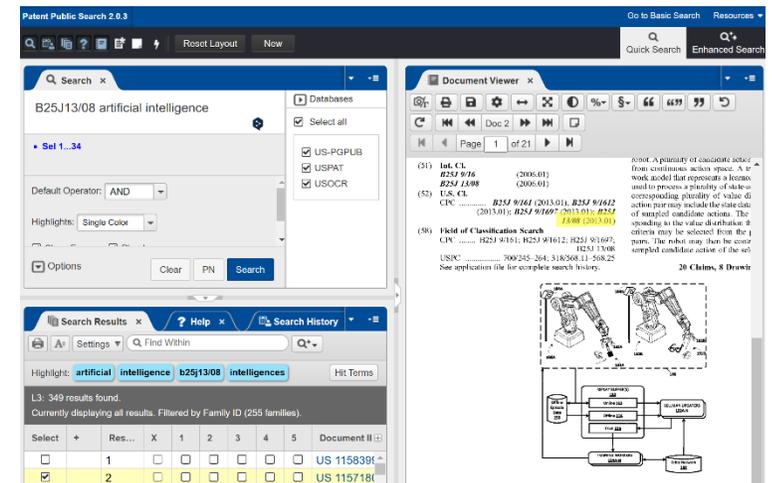
- 従来の検索手法は次の2つを前提とする:

1. アノテータが適切な分類を付与
2. 検索者が適切なテキストクエリを選択

- 特許分類の付与精度 < 100%

- 言葉で言い表すのが難しい形状

- 知財実務家でさえ、テキストクエリの作成は容易ではない
- サーチの際は、1万枚以上の図面を目視確認する必要あり



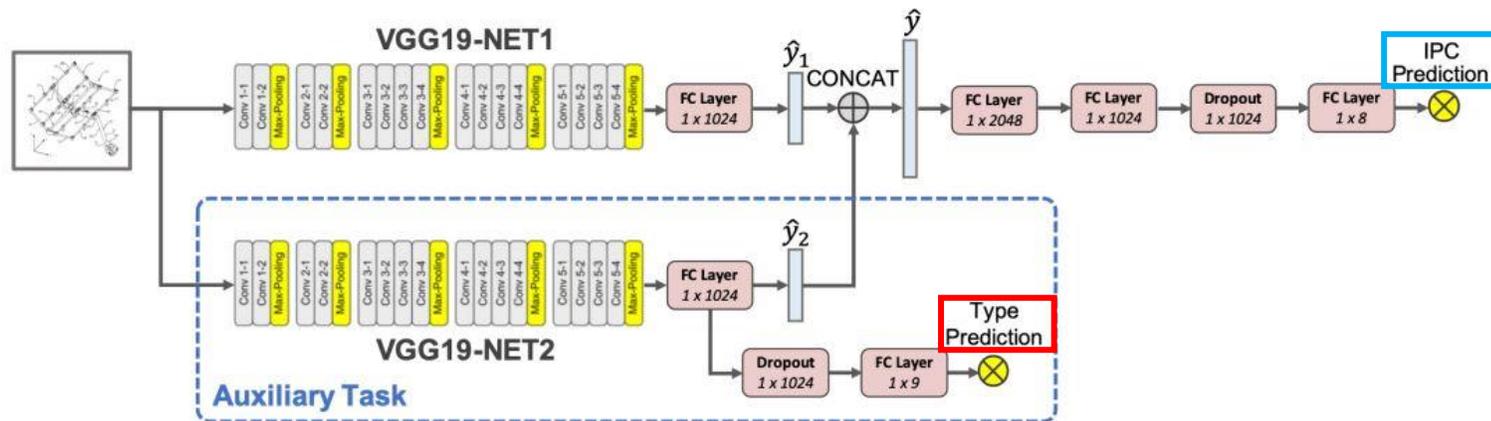
Patent Public Search([US](#))  
テキストに基づく検索が可能

# 2. Related work

## コンピュータビジョンにおける研究

- Dual VGG Network for patent image retrieval[14]
  - 主タスク: predict 8 types of International Patent Classification(IPC)
  - 副タスク: classify 9 types of drawings
  - Dataset: CLEP-IP 2011 Project, 38k images

- × 獲得する画像表現が既存の特許分類に強く関連する
- × 同じIPCラベルが, 同じ特許内の全ての図面に付与される



Dual VGG Network[14]

Drawing types
Abstract
Flowchart
Graph
Chemistry
Chart
DNA
Math
Program
Symbol

副タスクの図面種別[14]



# 2. Related work

## 特許実務における，図面の重要性

- 特許図面は実務において非常に重要な役割を担う
- 特許法第70条第2項: 特許発明の技術的範囲

“…願書に添付した明細書の記載及び図面を考慮して、特許請求の範囲に記載された用語の意義を解釈するものとする” ([条文摘記](#))

- 判例: 構成 (例: 形状) を図面から認定可能

“引用発明においては、[第4図](#)及び[第6図](#)において円弧として示されているギャッシュの底面領域部分だけではなく、ギャッシュの底面全域の形状が、同じ円弧面(略)によって形成されていることを刊行物1の記載から認定することができる。”  
(平成26(行ケ)10274 [知財高裁](#))

- 画像検索の実現は特許実務において重要な意味を持つ



DeepPatent sample[1]

# 3. Method

- **Baseline**

- Triplet[17]は, 同一の順伝播ネットワークの3つのインスタンスからなる(右下図).

- Triplet は1つの **positive** pair と1つの **negative** pair を扱う.

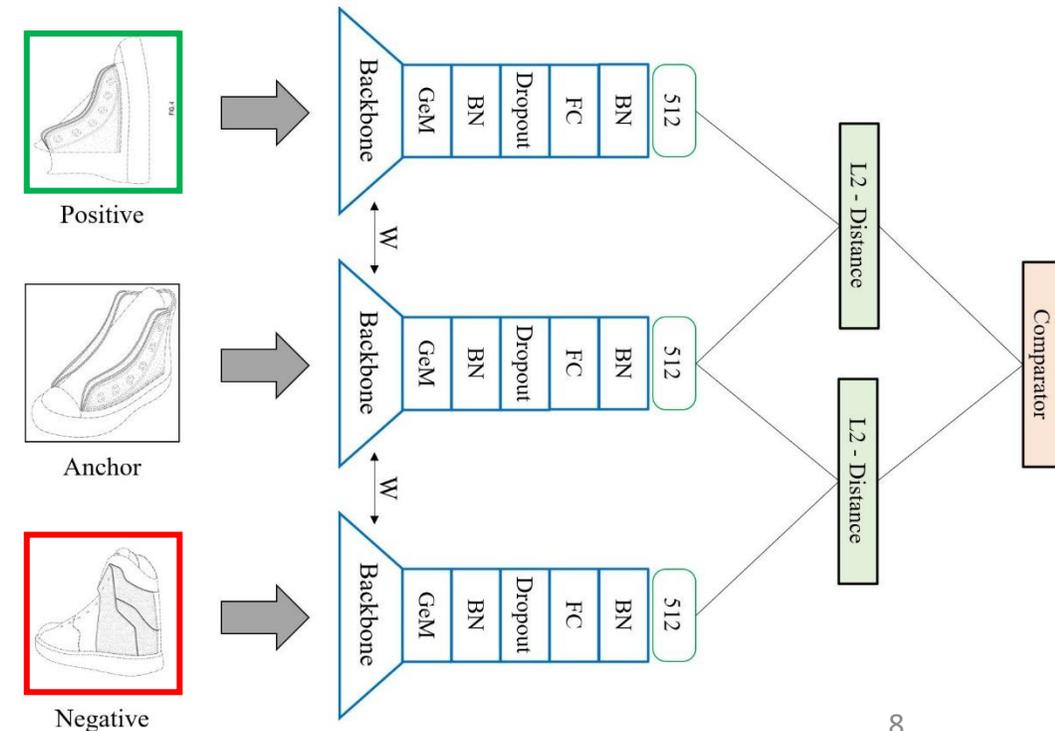
- × 一般に, **positive** pairs より多く存在する, **negative** pairs を有効に利用することが困難

- 当該課題に対して, 交差エントロピに基づく手法を適用

  - InfoNCE と ArcFace

$$L_t(I_q, I^+, I^-) = \frac{1}{2} \max(0, m + \|q - d^+\|^2 - \|q - d^-\|^2)$$

Anchor-**Positive**, Anchor-**Negative**, 2つのL2距離からなる0以上の距離をTriplet損失とする.



Baseline architecture(Triplet)

# 3. Method

- InfoNCE[18]
- 自己教師あり学習から発展
- Tripletと異なり, InfoNCE は多くの Anchor-Negative Pairs を利用

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

- ArcFace[16]
- 交差エントロピー損失に2つの要素を追加
  - 重み及び特徴量の正規化
  - マージン  $m$  を正解クラスに加える

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

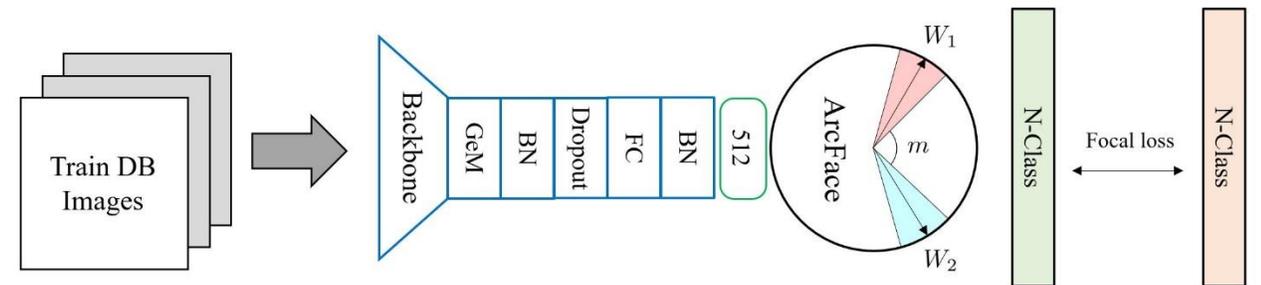
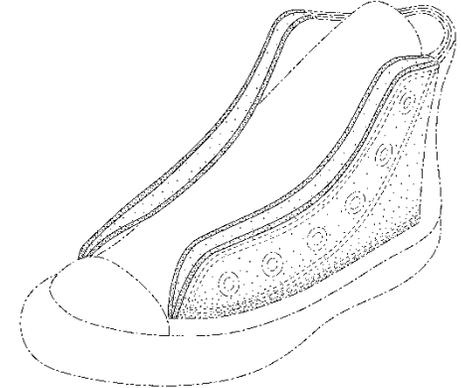


Fig. 1: 提案アーキテクチャ.  
ArcFaceに基づく学習を行う。

# 3. Method

- DeepPatent dataset
- Kucer *et al.* が米国意匠のデータセットを公開[1]
  - プロジェクトの [GitHub](#) から利用可能
  - 各特許は公開番号と図面番号を有する
- Public domain[20]
  - the U.S. Patent and Trademark Office(USPTO)による記述
  - DeepPatent dataset をベンチマークとして選択



引用: DeepPatent dataset  
公開番号: [USD811075S1](#)  
図面番号: 1

DeepPatent dataset 詳細

	画像枚数(図面数)	クラス数(出願件数)
Train	254,787	33,364
Test	38,834	6,927
Validation	44,815	5,888

# 4. Experiments

## • 実験設定

- Train/Test/Validation分割はDeepPatent[1]と同一
- 学習はPyTorch Metric Learningを用いて実装
- NVIDIA RTX A6000(48GB of VRAM) と Intel Xeon

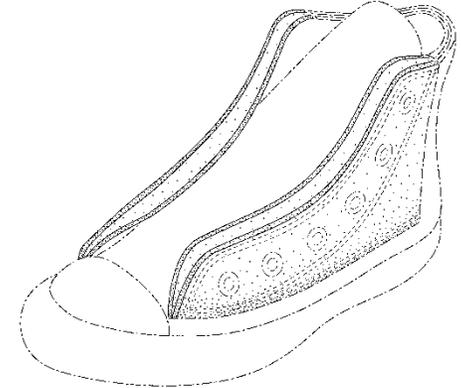
## • 前処理

- OpenCVを用いて, 図面外周の白色部分を削除
- バッチサイズ256

## • 評価指標

### • mAP score

- 各クエリのAverage precisionの平均結果
- Pytorch Metric Learning を用いて実装



引用: DeepPatent dataset  
公開番号: [USD811075S1](#)  
図面番号: 1

DeepPatent dataset 詳細

	画像枚数(図面数)	クラス数(出願件数)
Train	254,787	33,364
Test	38,834	6,927
Validation	44,815	5,888

# 5. Comparison to SOTA

## SOTAとの比較

- 我々の手法は mAP = **0.856** を達成
- DeepPatent baselineは, Res50+Triplet
- DIRA Challengeは, 手法非公開の1位スコア
  - mAP@10のため, 参考値\*
  - ECCV2022併催のコンペ. 評価サーバは公開終了
- Ours
  - 従来(baseline)を大幅に上回る mAPスコアを達成
  - SwinV2+ArcFace+データ拡張

## 最先端手法との比較

Method	mAP
RI-LBP[1]	0.069
DeepPatent (Res50+Triplet)[1]	0.379
DIRA Challenge Winner	0.849
Ours (SwinV2+ArcFace)	<b>0.856</b>

➤ Backbone, 距離学習手法, データ拡張の適切な組合せで, 特許図面検索の精度が向上すると判明

# 6. Ablation studies: 距離学習の比較

## 距離学習手法の比較

- Tripletは, baselineと非常に近い値
  - EffNet+Triplet と ResNet50+Triplet を比較
  - アーキテクチャ及び自己の実装を確認
- InfoNCE
  - より高いmAPスコア及びVRAM使用率
  - Tripletより negative samples を多く用いる点から妥当
- ArcFace
  - Baselineよりも突出して高い値 mAP=0.622 と判明
  - 顔認識だけでなく, 特許図面にも高い識別性

距離学習手法の比較

Method	mAP
Res50 + Triplet[1]	0.379
EffNet + Triplet[17]	0.384
EffNet + InfoNCE[18]	0.447
EffNet + ArcFace[16]	<b>0.622</b>

Backboneは軽量性及び精度を考慮し,  
EfficientNet[23]で統一  
画像サイズは256 × 256で統一

# 6. Ablation studies: Backboneの比較

## Backboneの比較

- EfficientNet
  - パラメータが少なく学習が高速にも関わらず, ViTを上回る
- ViT
  - Transformer手法が特許図面で動作すると判明
- Swin
  - 従来手法を上回るmAPスコアを達成
  - V2は更に顕著な結果

Backboneの比較

Backbone	mAP	#param.
EffNet-B0[23]	0.622	4M
ViT-B/16[24]	0.614	86M
Swin-B[25]	0.676	87M
SwinV2-B[26]	<b>0.767</b>	87M

学習はArcFaceで統一

画像サイズは256 × 256で統一

# 6. Ablation studies: 画像サイズの比較

## 画像サイズの比較

- 256 × 256
  - DeepPatentは, 1000 × 1000を超える大きい画像が多い
  - Cropにより必要な情報が失われる
- 384 × 384
  - 非常に高い精度
  - データセット規模にも依存するが, 実用に耐え得る精度
  - GPUの制約から今回の実験環境の上限

画像サイズの比較

Image size	Backbone	mAP
224 × 224	SwinV2-S	0.638
256 × 256	SwinV2-B	0.767
384 × 384	SwinV2-B	<b>0.831</b>

学習は SwinV2+ArcFace で統一  
原則, Backboneは, SwinV2-B  
例外として, 224 × 224は, SwinV2-S

# 6. Ablation studies: データ拡張の比較

## データ拡張の比較

- SimCLR-V2[27]及びMoCo-V3[38]を参考
  - 共通して弱めのGaussianBlurを加える
- Crop方法
  - 画像サイズの影響が大きいと判明したため実験
- RandomCropにより, 精度向上すると判明
  - 特許図面の意味合いを大きく変化せず,  
多様なバリエーションを学習出来た点が要因と推察

## 画像サイズの比較

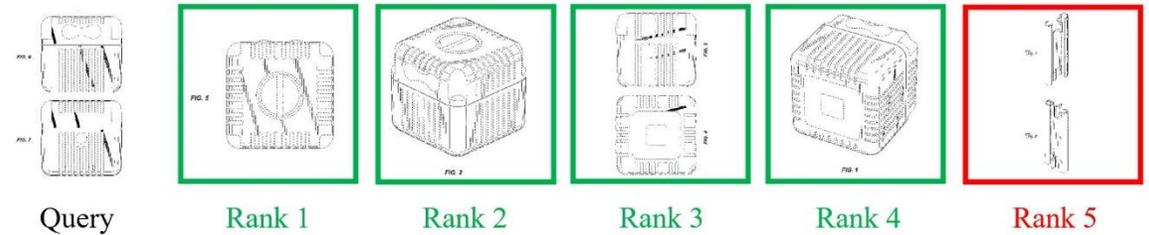
Method	mAP
CenterCrop	0.831
RandomCrop	<b>0.856</b>

学習は SwinV2+ArcFace で統一  
画像サイズは384 × 384で統一

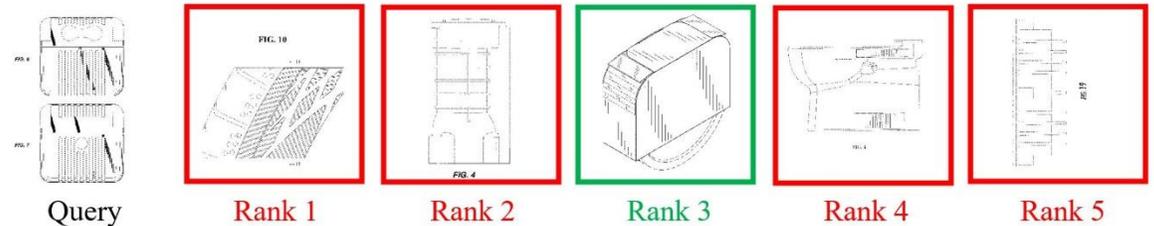
# 7. Application

## 検索結果の定性的評価

- 検索結果の例示
  - 一番左側がQuery画像
  - Rank順に1位から5位まで表示
- 提案手法(a)
  - 正解数が多く, 豊かな画像表現を獲得
- 提案手法(b)
  - 正解数が少ない



(a) 提案手法による, 類似した特許図面を含む検索結果の例.



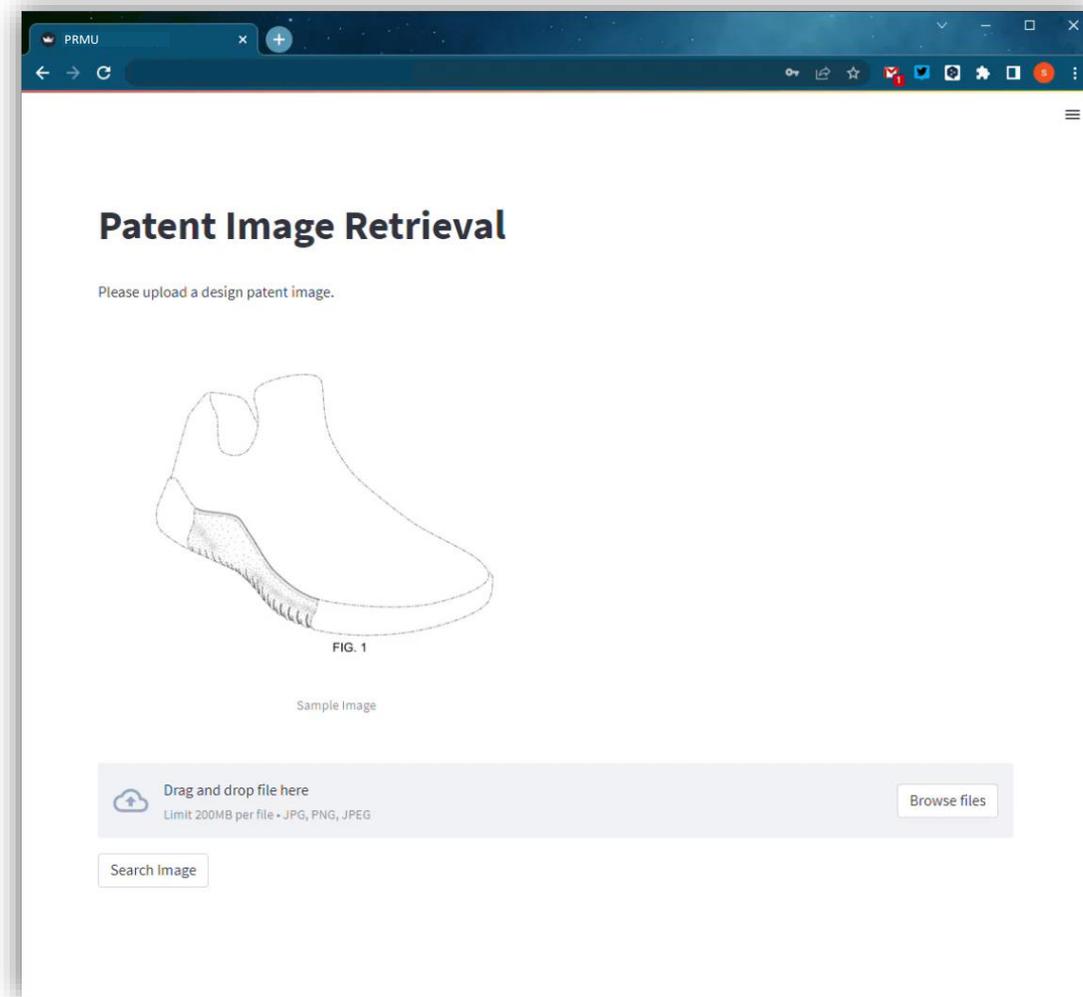
(b) 従来手法 (Triplet) による, 同じ Query の検索結果の例.

図 5: 検索結果の定性的評価. 提案手法で, 高品質の検索を実現. 緑色が正解で, 赤色が不正解を示す.

# 7. Application

## Webアプリ及び実務家のご意見

- 検索・推論アーキテクチャを実装
  - FAISS はベクトルの類似検索ライブラリ
  - Index file のサイズは約80MB
  - サーバにデプロイ可能
- 実務家からのご意見
  - 予想よりも動作が軽い. 精度も悪くない.
  - 画像Queryのみで検索可能のため, 簡便.
  - 日本の特許図面への適用にも関心がある.



特許検索アプリの画面

# 8. Conclusion

---

- TransformerとArcFaceを組合せた最新のアーキテクチャを適用し, 従来性能を上回った
  - mAPスコアでSOTA達成
- 特許凶面検索アプリケーションの実現
- 今後の課題
  - 更に大規模データサイズ(~数億)に適用した際, 学習収束し, 実務に耐える性能を獲得可能か

# References

- [1] M. Kucer, D. Oyen, J. Castorena, and J. Wu. Deep patent: Large scale patent drawing recognition and retrieval. In WACV, pp. 2309–2318, January 2022.
- [2] Ministry of Economy Trade and Industry. Release of japan patent office annual report, 2022. [https://www.meti.go.jp/english/press/2022/0727\\_003.html](https://www.meti.go.jp/english/press/2022/0727_003.html).
- [3] V. Stefanos, P. Symeon, M. Anastasia, S. Panagiotis, P. Emanuelle, and K. Ioannis. Towards content-based patent image retrieval: A framework perspective. *World Patent Information*, Vol. 32, No. 2, pp. 94–106, 2010.
- [4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv:1512.03385, 2015.
- [5] Y. Maehara, A. Kuku, and Y. Osabe. Macro analysis of decarbonization-related patent technologies by patent domainspecific bert. *World Patent Information*, Vol. 69, p. 102112, 2022.
- [6] W. Shimoda and K. Yanai. Learning food image similarity for food image retrieval. In *BigMM*, pp. 165–168, 2017.
- [7] F. Radenovic, A. Iscen, G. Tolias, Y. Avrithis, and O. Chum. Revisiting oxford and paris: Large-scale image retrieval benchmarking. In *CVPR*, 2018.
- [8] F. Radenovic, G. Tolias, and O. Chum. Fine-tuning cnn image retrieval with no human annotation. In *PAMI*, Vol. 41, pp. 1655–1668, 2019.
- [9] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural networks for one-shot image recognition. In *Proc. of ICML Workshop in deep learning*, 2015.
- [10] E. Hoffer and N. Ailon. Deep metric learning using triplet network. In *Similarity-Based Pattern Recognition*, pp. 84–92, 2015.
- [11] A. Gordo, J. Almazan, J. Revaud, and D. Larlus. End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, Vol. 124, No. 2, pp. 237–254, 2017.
- [12] J. Revaud, J. Almazan, R.S. Rezende, and C.R. de Souza. Learning with average precision: Training image retrieval with a listwise loss. In *ICCV*, 2019.
- [13] United States Patent and Trademark Office. Patent public search, 2023. <https://ppubs.uspto.gov/pubwebapp/static/pages/landing.html>.
- [14] S. Jiang, J. Luo, G. Pava, J. Hu, and C. Magee. A convolutional neural network-based patent image retrieval method for design ideation. In *IDETC-CIE*, Vol. 83983, 2020.
- [15] M. Bhattarai, D. Oyen, J. Castorena, L. Yang, and B. Wohlberg. Diagram image retrieval using sketch-based deep learning and transfer learning. In *CVPR*, pp. 174–175, 2020.
- [16] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, pp. 4685–4694, 2019.
- [17] R. Cao, Q. Zhang, J. Zhu, Q. Li, Q. Li, B. Liu, and G. Qiu. Enhancing remote sensing image retrieval using a triplet deep metric learning network. *International Journal of Remote Sensing*, Vol. 41, No. 2, pp. 740–751, 2020.
- [18] A. Oord, Y. Li, and O. Vinyals. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.
- [19] F. Piroi, M. Lupu, A. Hanbury, and V. Zenz. CLEF-IP 2011: Retrieval in the intellectual property domain. In *Conference and Labs of the Evaluation Forum*, 2011.
- [20] United States Patent and Trademark Office. Terms of use for uspto websites, 2023. <https://www.uspto.gov/terms-use-uspto-websites>.
- [21] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. In *ICCV*, pp. 2999–3007, 2017.
- [22] K. Musgrave, S. Belongie, and S. Lim. A metric learning reality check. In *ECCV*, pp. 681–699, 2020.
- [23] M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*, pp. 6105–6114, 2019.
- [24] A. Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.
- [25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *ICCV*, pp. 9992–10002, 2021.
- [26] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *CVPR*, pp. 12009–12019, 2022.
- [27] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton. Big self-supervised models are strong semi-supervised learners. *NeurIPS*, Vol. 33, pp. 22243–22255, 2020.
- [28] X. Chen, S. Xie, and K. He. An empirical study of training self-supervised vision transformers. In *CVPR*, pp. 9640–9649, 2021.
- [29] J. Johnson, M. Douze, and H. Jegou. Billion-scale similarity search with gpus. *BigMM*, Vol. 7, No. 03, pp. 535–547, 2021.