社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

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†

E-mail: †higuchi-k@mm.inf.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし　知的財産に関する業務は多岐に渡る．特に，特許分野の先行技術文献調査では，過去の膨大な文献の中から，新規性・進歩性に係る判断材料となる文献をヒットさせる必要がある．調査業務の中で，発明の重要な情報である図面を直接検索する図面検索技術の研究開発が，長年望まれていた．しかし，特許図面は白黒の抽象的な図として記述される性質があり，自然画像と大きくモーダルの異なる図面であることから，研究開発は困難を極めていた．
そこで，本研究では，深層距離学習を特許図面検索に適用することで，ネットワークに良質な画像表現を獲得させ，特許図面データセットにおける図面検索が可能であることを示し，DeepPatent [1] dataset 上で従来手法を大幅に上回る性能を達成した．本研究の手法は，特許図面だけでなく，モーダルの類似する，商標，意匠，ダイアグラム，スケッチ等に応用が可能と考えられる．

キーワード　Deep Metric Learning, Patent Image Retrieval, Swin Transformer

1. はじめに

特許の先行技術文献調査は，あらゆる知財ユーザにおいて，競合調査及び重複投資の防止の観点で，重要である．一方，特許出願は，我が国で現在約 28.9 万件 [2] 出願されており，上述のユーザは，膨大な文献の中から，新規性・進歩性に係る判断材料となる文献を，発見する必要がある．特に，特許分野はテキスト情報及び特許分類情報による検索技術が発達しており，近年の自然言語処理の発達に伴い，特許分野の研究開発は飛躍的な進化を遂げた．しかし，特許図面に関しては，未開拓の領域が残されている．

そのような状況の中で，発明の重要な情報の一つである，図面を直接検索する図面検索技術の研究開発が，長年望まれていた．例えば，日本の類似画像検索システムは，いわゆる DeepLearning 技術の登場前に構築されたものが多く，従来の画像特徴量を抽出することで，類似画像を検索する手法が採用されていた [3]．しかし，ResNet [4] の登場以降，DeepLearning に基づく画像認識を利用した特許図面の検索技術の開発は困難を極め，未だ特許分野においてデファクトとなる手法・システムは登場していない．

理由は，特許図面は白黒の抽象的な図として記述される性質があり，自然画像と大きくモーダルの異なる図面であるため，従来手法の単純適用では，所望の性能到達が困難な点が挙げられる [1]．加えて，特許分野は，図面の認識及び理解に知財専門家 (弁理士，発明者等) の知見が必要である場合が多く，需要に対して供給 (専門家の人数) が少ないことに加え，データセットの新規構築は高コストであり，現実的ではなかった．

本論文では，上述の状況を踏まえ，近年急速な発展を遂げた，深層距離学習 (Deep Metric Learning) を特許図面に適用することで，従来困難であった，特許図面検索が可能であることを示す．

本論文の主な貢献は次の通りである．

- 特許図面データセットにおいて，既存の DeepLearning 手法を含む手法をベンチマークし，その有効性を評価する．
- 特許図面データセットにおいて，従来の検索性能を上回るアーキテクチャ及び SOTA 手法を提案する．

2. 関連研究

2.1 機械学習の特許文献への適用

特許分野は，テキスト情報及び特許分類情報の活用が発達してきた．まず，テキスト検索では，特許出願の明細書等で利用される文言を検索に利用し，ユーザは様々な類語・シソーラス・技術用語を駆使して先行技術文献調査を実行する．広範な検索が行える反面，検索クエリの作成には専門的な知見が不可欠である．また，特許分類を用いる検索は，特許出願に付与される分類データを利用して，調査を行う手法であり，IPC, CPC 等が存在する．IPC とは，International Patent Classification の略であり，1975 年に国際特許分類に関するストラスブール協定に基づいて作成された，世界共通の特許分類である．また，CPC は，Cooperative Patent Classification の略であり，2013 年に欧州特許庁と米国特許庁が特許分類を共同で利用する分類である．特許の先行技術文献調査では，上記のテキストと特許分類を掛け算することで，サーチが行われてきた．

上述の背景から，特許分野は，テキスト情報及び特許分類情報と相互作用しやすい，自然言語処理技術との発達が進んだ．例えば，前原ら [5] は，公開特許公報の記載を用いて，BERT モデルに FC 層を追加して finetune することで，特許文献の CPC における自動分類を実現した．上記のように，特許分野は NLP により飛躍的な進化を遂げた．一方，CV の分野では，同様の改善はまだ登場していない．

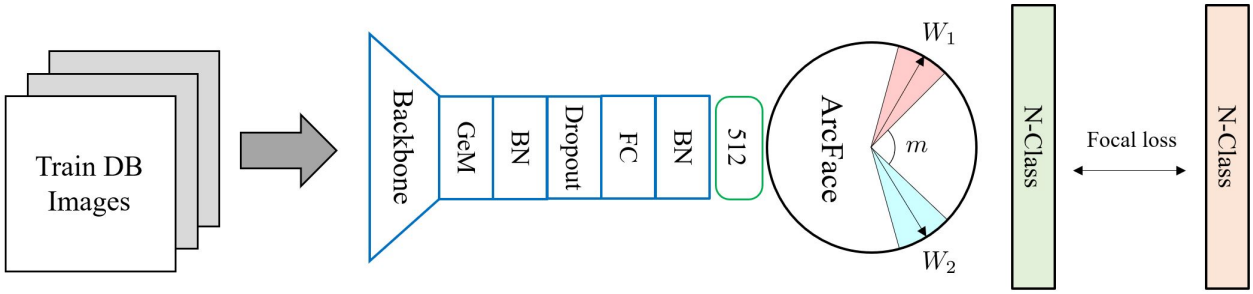


図 1: 提案するアーキテクチャ．ArcFace に基づく深層距離学習を行う．

他方で，CNN 特徴量を用いた画像検索技術は，近年発達を遂げた．例えば，下田ら [6] は，CNN ベースの画像検索を，食事画像に適用し，本技術の応用可能性を開拓した．当該論文では，特徴量の類似性を学習する CNN を種々用いることで，食品画像検索における有効性を検証している．

2.2 Image Retrieval

画像検索に CNN 等のネットワークを用いる手法について，様々な研究がなされてきた．例えば，Radenović ら [7] は，画像検索データセットのアノテーションを再検討し，従来の局所特徴量に基づく手法と，CNN に基づく手法とを比較した．当該比較により，CNN に基づく大域記述子を用いる手法で，性能向上が図れると判明した．

また，別の研究で同 Radenović [8] らは，GeM Pooling を提案した．GeM Pooling は，Generalized Mean Pooling の略であり，Avg Pooling, Max Pooling を一般化し，パラメータ p_k を訓練可能とした手法である．GeM Pooling の数式を以下に示す．

$$\mathbf{f}^{(g)} = [\mathrm{f}_1^{(g)} ... \mathrm{f}_k^{(g)} ... \mathrm{f}_K^{(g)}]^\top, \quad \mathrm{f}_k^{(g)} = \left(\frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x^{p_k} \right)^{\frac{1}{p_k}}$$

当該手法は，画像検索における様々なネットワークの pooling 層として，近年の画像検索タスクに広く用いられている．

上述の研究を踏まえ，近年，CNN ベースの画像検索において，Siamese [9] や，Triplet [10] のアーキテクチャを採用した上で，通常は微分（学習）不可能な，評価指標（mAP 等）を最適化する手法が，最先端の性能を示してきた [11] [12].

2.3 従来の特許検索及び課題

従来，Patent Public Search [13] 等の様々な検索ツールを用いて，テキストあるいは特許分類を駆使し，特許検索が実行されてきた．操作画面の一例を図 2 に示す．

しかし，既存の特許検索は，1. 検索対象の文献に適切な分類が付与されている，かつ，2. 発明の本質を捉える適切な技術用語（クエリ）を習得している（技術分野に熟達している），2 点を前提としている．特許分類の付与精度は 100% でなく，かつ分類が異なる特許にも，類似する特許は存在する．また，テキストクエリの作成は，発明が最先端技術を包含する特性から，知財専門家でも困難な場合が多い．

加えて，特許検索には，言葉で表現しにくい形状を検索するニーズが存在する．しかし，その検索には，検索者は数千枚〜数万枚の図面を目視確認する必要がある．

上述の状況を鑑みて，CNN に基づく特許図面検索の研究もなされている [14]．当該研究のネットワークは，特許図面から得られる embedding で，特許分類 IPC(大分類 A から H まで)8 種類を推定する主タスクと，図面の 9 種分類を行う補助タスクを実行する．しかし，得られる画像表現は，既存の特許分類と強く関連するため，分類の付与漏れの課題が残存していた．また，同一出願の全ての図面に，同じ IPC ラベルが付与されるため，疎な画像表現を獲得する課題も内在していた．

特許図面検索に取り組んだものとして，Bhattarai [15] らの研究も挙げられる．当該研究は，VAE で，画像表現を獲得した．この研究は，図面サーチの有効性が高い (靴の) 分野に限定している点で，一定の効果が見込まれる．しかし，業務ニーズが高いのは，技術分野を限定せず，類似図面を発見することである．

そこで，Kucer ら [1] は，後述する DeepPatent dataset と共に，ResNet50 で Triplet による距離学習を行うことで，特許図面検索が可能であると示した．ただ，mAP スコアが実務にあと一歩届かず，従来の特許検索には，課題が残されたままであった．

2.4 提案する特許図面検索

上記を勘案して，最先端の画像検索システムとして次のようなアーキテクチャを提案する．

- CNN や Transformer を Backbone として，良質な画像表現を獲得すべく深層距離学習を実施．
- 学習済 NW を用いて，Query 画像から特徴量を抽出し，Database と照合して類似度を演算．
- 類似度に基づいて，最も類似した画像をユーザに提示．

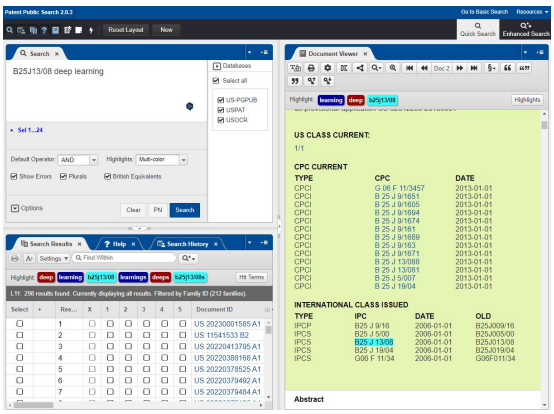


図 2: Patent Public Search [13].

3. 手　　法

3.1 Deep Metric Learning

Deep Metric Learning(DML) における画像検索は，DCNN 埋込みによる画像表現が用いられる．正規化処理した画像を，クラス内距離が小さく，クラス間距離が大きいマップ上に埋込む学習を行う．DML は，大きく 2 つの系統に分かれる．1 つ目は，Triplet loss [10] のように，マイニングしたサンプルからネットワークを介して埋込みを抽出し，当該埋込同士の距離を学習するものである．2 つ目は，ネットワークを介した埋込みに対して，分類器のように，独立したクラスを分類できる，多クラス分類器を学習するものである [16]．以下，具体的に説明する．

3.1.1 Triplet

深層距離学習の代表的な手法の 1 つとして，Triplet [17] が挙げられる．Triplet ネットワークは，同じ順伝播ネットワークの 3 つのインスタンス（パラメータは共有）から構成される．バッチから 3 つのサンプルを取り出し，ネットワークは，Anchor と Positive，Anchor と Negative，2 つの距離を計算し，マージン m を加えることで，誤差逆伝播が 3 つのサンプルに関して同時にモデルを更新することを可能にする．

L_t を損失関数，m をマージン変数とすると，Triplet loss の式は次の通り．

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

ここで，(I_q, I^+, I^-) は画像，(q, d^+, d^-) は query 及び embedding の triplet(三つ組) を示す．一般に，$(anchor, positive^+, negative^-)$ の組で説明される．

3.1.2 infoNCE

深層距離学習に加えて，自己教師あり学習でも用いられる代表的な手法の 1 つに，infoNCE [18] が挙げられる．infoNCE は，Triplet と同様に，バッチ内のサンプリングにおいて，Anchor と Positive の 1 つのペアに対して，Anchor と Negative のペアを多数用いるものである．そのため，バッチサイズが大きい程，良質なバッチサンプリングが可能であり，精度の高い学習が可能となる．infoNCE の損失関数 L_i を次に示す．

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

3.1.3 ArcFace

ArcFace [16] は，クラス分類問題において用いられる Cross Entropy の損失関数に対して，1. 重み・特徴量の正規化，2. 正解クラスにマージンを設ける工夫を加えることで，クラス間の分散を大きくする距離学習を実現した．具体的には，クラス内距離を小さくし，クラス間距離を大きくすべく，Cross Entropy 損失を式変形し，角度 margin ペナルティ m を導入している．当該 margin の導入は，正規化した超球上の距離ペナルティ margin と等価であることから，ArcFace と名付けられた．ArcFace の損失関数 L_a は以下の通り．

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

3.2 Patent Image Dataset

特許図面について，2011 年の CLEF-IP プロジェクト [19] は，タスク別のデータセットを提案した．検索タスクは 211 件の特許出願から構成される．また，分類タスクは，約 3.8 万件の案件が含まれ，クラス種別は，フローチャート等を含む 9 種類に限られていた．当時最大級の特許データセットであったが，現代の深層学習に必要なデータ量には届かない課題があった．

3.2.1 DeepPatent Dataset

Kucer ら [1] は，米国意匠 (Design Patent) を 35 万件以上含む，データセットを公開した．詳細を表 1 に示す．2018 年から 2019 年の前半にかけた，約 4.5 万件から構成され，70% を train，15% を test，15% を validation として分割されている．本データセットは，当該プロジェクトの Google Drive から入手可能であり，図面ごとに公開番号及び図面番号が付されている．

なお，公開された DeepPatent データセットは，当該論文及び米国特許商標庁 (USPTO) において，著作権制限の対象ではない，パブリックドメインである旨が記載されている [1] [20]．

以上を踏まえ，本研究では，ベンチマークに用いる特許図面検索のデータセットとして，DeepPatent dataset を採用した．

表 1: DeepPatent dataset 詳細

DeepPatent	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

3.3 距離学習手法の検討

知財実務に求められる検索性能は高く，かつ，特許図面検索タスクの難度は高い．背景は 2 点あり，1. 特許図面は自然画像とモーダルが大きく異なる点，かつ，2. データセットの構成が案件毎に複数の視点を含む大規模の図面群からなる点である．上記を勘案して，本研究では，DeepPatent データセットにおいて，最新の画像検索コンペ等で顕著な成果を残している，ArcFace [16] を中心に，有効性を評価する実験を行った．

本研究の訓練アーキテクチャを図 1 に示す．画像バッチを，バックボーンを含む画像エンコーダに通して，次元圧縮を行い，ArcFace で距離学習させることで，N-Class 分類を行う（表 1 から，N=33364）．損失関数には Focal loss [21] を用いる．

4. 実　　験

4.1 実 験 設 定

データセットには上述の DeepPatent dataset を用いた．また，Train/Test/Validation の分割も DeepPatent 論文 [1] と同一である．全てのネットワークは，Pytorch Metric Learning [22] を使用して実装した．また，Backbone には timm [1]に含まれる EfficientNet-B0, ViT-B, Swin-B, SwinV2-B を使用する．学習は，NVIDIA RTX A6000 (48GB of VRAM) と Intel Xeon CPU の組合せで行う．前処理は，白黒画像であることを考慮し，入力チャンネル

（注1）：https://github.com/rwightman/pytorch-image-models

数 1 の GrayScale，Resize 及び CenterCrop を行う．また，平均及び標準偏差を 0.5 で正規化する．さらに，DeepPatent dataset は，比較的画像サイズが大きく，余白（図面外周の白色部分）が大きいため，当該余白は OpenCV によりカットしている．

また，バッチサイズ 256 で学習を行った．最適化には Adam を利用し，初期学習率は $1e-4$ に設定した．

4.2 評価指標

検索システムの評価には，mAP を用いた．mAP は平均精度 AP を各クエリ q に対して計算し，平均を計算したものである．

なお，精度計測の実装は，Pytorch Metric Learning 標準の AccuracyCalculator を用いた．

5. 最先端手法との比較

DeepPatent dataset において，本提案の手法と，最先端手法と比較した結果を次の表 2 に示す．最先端手法として，DeepPatent 論文 [1] と比較した．Backbone に SwinV2 を用いて，ArcFace で距離学習し，後述のデータ拡張を実施した結果，最先端スコアと比較して大幅に (0.477 mAP) 上回った．

なお, 参考値として, ECCV2022 併催の DIRA Workshop Image Retrieval Challenge [2]における，1 位スコアを記載した．当該スコアは，評価サーバが公開終了かつ，mAP@10 であるため，参考値として報告する．

以上から，特許図面タスクにおいて，Backbone，距離学習手法，前処理の適切な組合せにより，特許図面検索の精度が向上し，本提案の手法が SOTA と判明した．

表 2: 最先端手法との比較

Method	mAP
DeepPatent baseline [1]	0.379
DIRA Challenge [2]	0.849
ours	**0.856**

6. アブレーション実験

実験は，上述の 4. 1 節で用いた実験設定を共通して用いた．また，下記のネットワークは，原則，画像サイズ 224×224，バッチサイズ 512 で 20epoch 学習させることで，実験を行った．

6.1 学習手法の比較

図 1 のアーキテクチャにおいて，距離学習の 3 種類の手法で比較した結果を表 3 に示す．

DeepPatent Baseline において，ResNet50 を Backbone として，Triplet により学習したところ, mAP=0.379 と報告されている [1]. ただし，本実験の Backbone は，モデルの軽量性と精度を備えた，EfficientNet [23] で統一する．

本実験で実装した EfficientNet と Triplet との組合せは，mAP が 0.384 と判明した．当該値（mAP=0.384）は，DeepPatent baseline とほぼ同じ値 (mAP=0.379) であることから，深層距離学習のアーキテクチャ及び自己の実装について確認することが出来た．次に，infoNCE は，Triplet に対して少し精度が高い結果となった．これは，infoNCE が，Anchor と Negative のペアを多数用いる点から妥当であると考えられる．なお，VRAM のメモリ使用量も Triplet に対して高かった．最後に，ArcFace は，baseline よりも非常に高い mAP スコアが得られた．ArcFace は，顔認識だけでなく，特許図面においても高い識別性が得られることが判明した．

本結果を踏まえ，以降は，ArcFace を中心に実験を行った．

表 3: 深層距離学習の手法別比較

Method	mAP
DeepPatent baseline [1]	0.379
Triplet [17]	0.384
InfoNCE [18]	0.447
ArcFace [16]	**0.622**

6.2 Backbone の比較

図 1 のアーキテクチャにおいて，Backbone 別の mAP スコアを比較した．具体的には，EfficientNet [23] と，Vision Transformer [24], そして，最新の Swin Transformer [25] の 3 種類である．距離学習の手法は，ArcFace で統一している．

なお，ViT-B/16 及び Swin-B については，GeM Pooling を省略したアーキテクチャで実験を行った．その他の構造は，head 部分を含め統一している．

まず，EffNet はパラメータ数が少なく学習が高速であるにも関わらず，ViT をやや上回る性能を達成した．EffNet は，VRAM メモリ使用量も少なく，各種タスクの利用実績がある点を理解出来た．また，Transformer 手法の一つである，ViT も，特許図面検索で問題なく動作することが確認出来た．さらに，Swin について実験し，mAP スコア 0.676 と，従来の Baseline を大幅に超える性能を獲得することが出来た．加えて，最新の SwinV2 [26] について実験を行ったところ，他のモデルと比較して，顕著な結果を得ることが出来た．

表 4: Backbone の比較

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

6.3 画像サイズの比較

図 1 のアーキテクチャにおいて，Backbone に Swin-B を採用し，Crop する画像サイズを変更して比較した結果を表 5 に示す．

まず，224×224 の画像については，所望の精度が得られた．画像サイズを大きくするにつれて，高い性能が得られている．これは，DeepPatent dataset に含まれる画像が，1000×1000 を超える大きなサイズのものが多く，画像を小さく Crop することで，検索に必要な情報が失われてしまうからと考えられる．

（注2）：https://sites.google.com/view/eccv-dira/home

特に，384 × 384 の画像は，非常に高い精度であり，データセットのサイズにも依存するが，十分実用に耐え得るものと考えられる．ただし，今回の実験環境では，384 × 384 の画像サイズが，GPU の制約から実験可能な上限であった．

表 5: 画像サイズの比較

Image size	mAP
224 × 224	0.676
256 × 256	0.742
384 × 384	**0.831**

6.4 データ拡張の比較

図 1 のアーキテクチャにおいて，Backbone に SwinV2-B を採用し，データ拡張を加えた結果を表 6 に示す．

また，当該実験において，SimCLR-V2 [27] 及び MoCo-V3 [28] を参考に，共通して弱めの GaussianBlur を加えている．

上記の実験から，画像サイズの影響が大きいと判明したため，Crop 方法に絞って実験を行った．いずれも画像サイズは上限の 384 × 384 に固定している．RandomCrop により精度が向上した理由は，特許図面の意味合いを大きく変化せず，epoch が進むごとに多様なバリエーションを学習出来た点が，精度向上に繋がったと考えられる．

表 6: データ拡張の比較

Augmentation	mAP
CenterCrop	0.831
RandomCrop	**0.856**

6.5 定性的評価

上述の実験から，特許図面における検索精度の有効性を定量的に確認することが出来た．これを踏まえ，最も性能の高いモデルと，従来のモデルとを比較する趣旨で，定性的評価を行う．また，学習したモデルが，局所解に陥っていないか確認する．具体的には，当該モデルの検索結果の一部を例示する．

6.5.1 推論・検索アーキテクチャ

上記の図 3 のアーキテクチャを用いて，特許図面の検索を行った．Faiss [29] は，埋込みベクトルに対する，効率的な類似検索が可能なライブラリであり，本実験で indexing の実装に利用した．また，train/test/validation 分割は DeepPatent [1] と同一のため，訓練データと，推論・検索に用いたデータは分離されている．さらに，Query と DB に同一の画像は含まれておらず，検索結果には一定の汎用性があると考えられる．

6.5.2 検索結果の定性的評価

図 5 に，検索結果の例示を行った．図 5 の一番左側が Query で，検索結果を Rank 順に 1 位から 5 位まで表示している．提案アーキテクチャは，タスクに対して適切な Backbone 及び距離学習を実施したため，正解数が多く，豊かな画像表現を獲得したことがわかる．

また，検索システムにおける距離学習では，学習曲線または

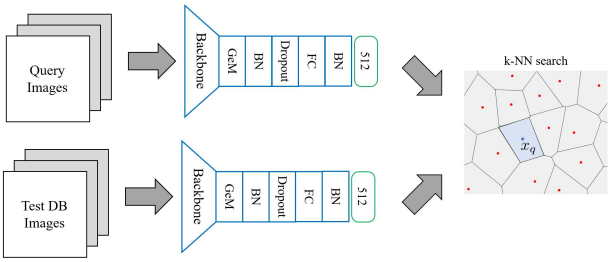


図 3: 推論・検索アーキテクチャ．Faiss による indexing を実施．

損失値等から，学習成功のように見え，かつ，高 mAP スコアの場合でも，実際の検索結果はどの Query 画像でも同じ，局所解に陥ってしまう場合が存在する．今回の定性的評価からは，そのような失敗事例には該当しないと確認することが出来た．

ただし，提案手法には，図 4 のように，数は少ないが検索の失敗例も存在する．現時点では，DB サイズ不足の可能性は否定出来ないものの，単純な形状（例えば，丸型）の場合において，検索の品質が低下するのではないか，と考えられる．ただし，従来論文 [1] においても，単純な形状で類似の失敗例が例示されており，共通の課題を内在すると認識している．

6.5.3 Web アプリ及び実務家のご意見

Faiss の indexFlatL2 による演算を行った結果，index ファイルの容量は約 80MB となった．当該ファイルの容量は，embedding の次元数（今回の実験は 512）と関連しており，今回はサーバあるいはクラウドにデプロイが十分可能なサイズであると判明した．そこで，上記の図 3 のアーキテクチャをユーザが手軽に利用可能とするため，Web アプリ化を行った上で，実務家に使用感のフィードバックを頂いた．なお，実装には Streamlit と，オンプレミスの Ubuntu サーバを用いた．

頂いたご意見は以下の通りである．全体として好評であり，今後の研究の発展を期待する温かい声を頂いた．

- 動作が予想よりも軽い．精度も悪くない．
- 画像 Query のみで検索可能のため，サーチが簡便．
- 日本の特許図面への適用にも関心がある．

7. おわりに

本提案の手法により，mAP の定量的評価において SOTA，かつ，定性的評価において高品質の画像検索を実現できた．長年，特許分野には図面検索の登場が望まれる課題があったが，本提案により，当該課題を解決可能であると確信する．本提案で，新たに 2 点が実証された．1 つ目は，特許図面データセットにおいて，既存の DeepLearning 手法が一定程度有効である点．2 つ目は，Transformer と距離学習を組合せた最新のアーキテク

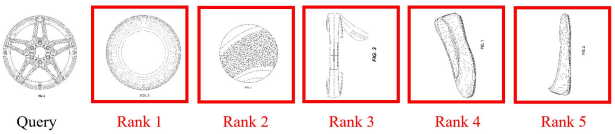
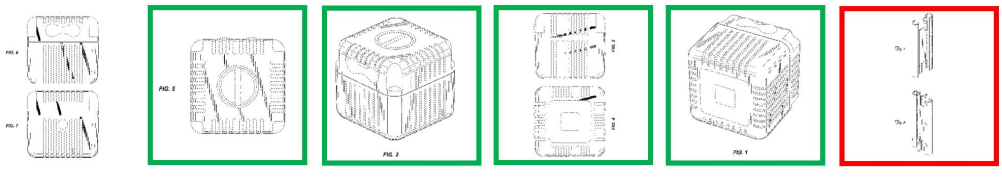
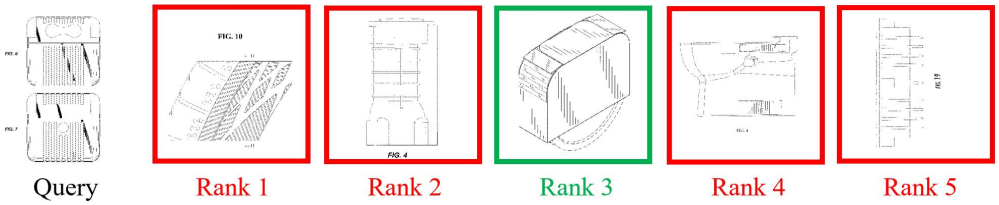


図 4: 提案手法の検索失敗例．



(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 画像の検索結果の例.

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

チャを特許図面データセットに適用することで，従来の検索性能を大幅に上回った点である．今後の課題としては，本提案の手法を更に大規模データセットサイズ（数千万〜数億）に適用した際に，学習を収束させ，実務に耐え得る性能を獲得出来るかという点が挙げられる．以上から，特許図面検索は未開拓の領域が多くあり，特許図面だけでなく，モーダルの類似する，スケッチ，商標，意匠，実用新案，機械製図，フローチャート等にも応用可能であり，今後も幅広い研究が行われることが期待される.

文　　献

[1] M. Kucer, D. Oyen, J. Castorena, and J. Wu. Deeppatent: 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.

[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 domain-specific 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. Radenović, A. Iscen, G. Tolias, Y. Avrithis, and O. Chum. Revisiting oxford and paris: Large-scale image retrieval benchmarking. In *CVPR*, 2018.

[8] F. Radenović, 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.