画風変換 LoRA の内部パラメータによる 変換特徴を考慮したモデルの埋め込み表現の獲得

金田　悠路† 大江　優真† ファムフーロン†† 加藤　誠†††,†††† 大島　裕明††
藤田　澄男††††† 荘司　慶行†

† 静岡大学大学院　〒 432 – 8011 静岡県 浜松市 中央区 城北 3-5-1
†† 兵庫県立大学　〒 651 – 2197 兵庫県 神戸市 西区 学園西町 8-2-1
††† 筑波大学　〒 305 – 8550 茨城県 つくば市 天王台 1-1-1
††††† LINE ヤフー株式会社　〒 102 – 8282 東京都 千代田区 紀尾井町 1-3
†††† 国立情報学研究所　〒 101 – 8430 東京都 千代田区 一ツ橋
E-mail: †{kanada.yuro.21,oe.yuma.21}@shizuoka.ac.jp, ††af23a009@guh.u-hyogo.ac.jp,
†††mpkato@slis.tsukuba.ac.jp, ††††ohshima@ai.u-hyogo.ac.jp, †††††sufujita@lycorp.co.jp,
††††††shojiy@inf.shizuoka.ac.jp

あらまし　本研究では，画風変換 LoRA の内部パラメータから，変換特徴を反映した LoRA モデルの埋め込み表現を獲得する手法を提案する．複数の層から構成される LoRA モデルを適切にベクトル化するために，層ごとに内部パラメータを抽出し，flat 化と次元圧縮を施すことで，ベクトル系列として表現した．このベクトル系列を入力として，重みを共有した三つの Transformer Encoder と MLP 層からなる Triplet Network で，それぞれの LoRA で変換された画像同士がどれだけ画像的に類似しているかを推定する距離学習を行った．学習の妥当性に関する自動評価，埋め込み表現と人間の類似性判断との一致度の検証，および検索タスクにおけるランキング性能の評価を通じて，提案手法が人間の判断と整合した埋め込みを獲得し，安定した類似 LoRA 検索を実現できることを示した．

キーワード　Low-Rank Adaptation，Model Embedding，Model Retrieval，Metric Learning, Weight Parameter

1　はじめに

近年の画像生成 AI を取り巻く環境の進歩は凄まじく，個人や企業などが，様々な目的で独自に追加学習した画像生成モデルを使うようになってきている．中でも，画風や特徴を覚え込ませた Low-Rank Adaptation（LoRA）[1] モデルは特に多く使われるようになってきており，トレーニングした LoRA モデルを共有されるサイトも登場してきている．実際に，HaggigFace[1] や Civitai[2]では，数十万件を超える LoRA モデルが，実際に共有されている．また，一般的な企業でも，社内の業務のために多数の LoRA を学習し，保有し再利用することが一般的になりつつある．

このような状況下で，大量の学習済みモデルを管理し，活用可能にする技術の重要性が高まってきている．一つの解決策として，事前学習済みモデルのベクトル空間上への埋め込み（Embedding）が注目を集めている．モデル埋め込みでは，モデルをベクトルで表現することで，意味的類似性に基づいてモデルを検索・推薦可能にしている．この際，多くの手法では，モデルのメタデータや出力例をもとにモデルをベクトル化してきている．

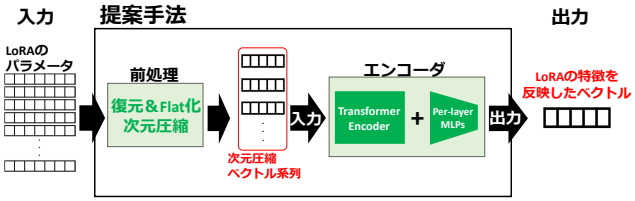

図 1　提案手法の推論時の概要図．LoRA のパラメータを入力すると，その LoRA の変換特徴を反映したベクトルを出力する．

一方で，こうしたメタデータや入出力を用いた埋め込みは，そういったデータのない単体のモデルを埋め込みできないという欠点を持つ．たとえば実際の業務では，メタデータや入出力例のない LoRA の実ファイルだけが残されていた場合もありうる．既存手法でモデルの埋め込みが行えない例として，

- 新たに構築された多数の学習済みモデルに対して，出力例を生成するコストが高い，または
- 古いモデルが残されているが，当時のメタデータやサンプル出力が失われている

などの場面が考えられる．実際に多くの追加学習済みモデルを管理する際に，モデル本体以外の付帯情報に依存したベクトル化手法では，十分な網羅性と柔軟性を確保できない場合がある．

1：「Hugging Face – The AI community building the future.」：https://huggingface.co/
2：「Civitai: The Home of Open-Source Generative AI」：https://civitai.com/

そこで本研究では，画像生成 AI モデルの持つ内部パラメータそのものから，モデルの変換特徴を考慮した埋め込み表現を獲得する手法を提案する．図 1 に，提案手法の推論時の概要図を示す．今回の手法では，画像変換用の LoRA を単体で与えると，それを LoRA の特徴を反映したベクトルに変換するモデルを学習する．

そして，

（1） LoRA の内部パラメータを flat 化，

（2） PCA による次元圧縮，

（3） Transformer Encoder [2] をベースとする Triplet Network による距離学習

を行うことで，LoRA の特徴を反映した 1 本の密ベクトル（すなわち，埋め込み表現）を出力する．こうすることで，LoRA の内部に存在するパラメータ（重み行列）の構造を直接利用して，モデルの変換特徴を反映した埋め込み表現を獲得できる．

提案手法は，出力例やメタデータが存在しない LoRA モデルに対しても，内部パラメータそのものから意味的な LoRA モデルベクトルを獲得可能にする．こうしたモデルの重みそのものからの埋め込みは，今後さらに増大が見込まれる LoRA の管理や検索を自動化するうえで重要である．本研究の主要な貢献は，推論時に出力例やメタデータに依存せず，LoRA の内部パラメータのみから，変換特徴を反映した埋め込み表現を得る手法を提案し，実際に有効性を実証した点にある．本研究の成果は，大規模モデル群に対する実用的な検索や推薦を行う上で，重要な基盤技術であると考えられる．本論文は，WebDB Forum で発表した研究 [3] を発展させ，MLP 層の構成および学習データの設計に変更を加え，手法の拡張と性能検証を行ったものである．

2 関 連 研 究

本研究は，LoRA の内部パラメータに基づいた埋め込み表現の学習手法を提案する．そのため本研究は，LoRA の基本的な仕組みとその応用，モデル埋め込み（model embedding）に関する先行研究と密接に関連する．

2.1 Low-Rank Adaptation

LoRA（Low-Rank Adaptation）は，大規模モデルの重み更新を低ランク行列として分離することで，効率的なファインチューニングを実現する手法である．具体的には，元の重み W_0 に対して低ランク行列 $B \in \mathbb{R}^{d\times r}$，$A \in \mathbb{R}^{r\times d}$ を用い，$W = W_0 + BA$ として更新を表現する．

LoRA の重要な特徴は，モデル全体の挙動変化が，レイヤごとの低ランク更新行列として明示的に分離されている点にある．すなわち，各 LoRA は，「どのレイヤで」「どの方向に」モデルの表現を変換しているかを，パラメータ空間上に構造化された形で保持している．この性質は，LoRA の内部パラメータ列そのものが，モデルの変換特性を直接表す表現になっていることを意味する．

近年，LoRA を対象とした検索・推薦・マージ，生成に関する研究が活発化している．例えば，LoraRetriever は入力プロンプトに基づいて複数の LoRA を動的に検索・組み合わせる手法を提案している [4]．また，SemLA は CLIP 空間上の画像・テキスト特徴を用いて LoRA の選択的マージを行う [5]．さらに，K-LoRA [6] や IterIS [7]，LoRACLR [8] など，再学習を伴わない統合手法も提案されている．また，条件付きノイズやタスク記述から LoRA モデル自体を生成する研究も提案されている [9, 10]．

これらの研究は，主に出力画像や外部埋め込み空間に基づいて LoRA の選択や統合を行う点に特徴がある．一方で，LoRA が本質的にレイヤ単位の重み更新として構造化された表現であることを踏まえると，その内部パラメータ列を直接埋め込み空間へ写像し，モデル間の類似性を比較するという視点は，LoRA 特有の構造を活かした自然なアプローチであると考えられる．本研究は，この観点に基づき，LoRA の内部パラメータから変換特性を反映した埋め込み表現を学習することを目的とする．

2.2 Model Embedding

LoRA の検索や推薦には，モデルをベクトル空間に写像する埋め込み手法が必要である．既存の手法は，出力例やメタデータに基づくものと，パラメータ自体に基づくものに大別される．

現状，LoRA 埋め込みの獲得には，出力例やメタデータに基づく手法が主流である．例えば，LoraRetriever では入力プロンプトの平均ベクトルを，SemLA では生成画像の CLIP 特徴量を，LoRA 埋め込みの代替表現として利用している．また，ModelSpider [11] や LogME [12]，およびモデル差分やログ解析に基づく手法 [13–15] では，モデルの出力ログに基づいて LoRA の埋め込み表現を学習している．

一方で，LoRA の内部パラメータそのものに基づく埋め込み表現の獲得手法も近年注目されている．例えば，低ランク行列 A および B を flat 化し，PCA によって圧縮することで LoRA 埋め込みを獲得する手法 [16, 17] や，タスク性能の予測を目的とした回帰モデルを通じて LoRA 埋め込みを学習する手法 [18] が提案されている．

しかし，既存のパラメータベース手法は，レイヤ構造やレイヤ間の関係といった LoRA が本来持つ構造的情報を明示的に考慮していない．これに対し本研究は，LoRA の内部パラメータをレイヤ順序を保った系列データとして扱い，Transformer Encoder による系列学習を通じて，LoRA の構造的な変換特性を反映した重みベースの埋め込み表現を学習する．

3 LoRA の内部パラメータによる変換特徴を考慮したモデルの埋め込み表現の獲得

本節では，画風変換 LoRA の内部パラメータに基づき，変換特徴を反映した埋め込み表現を獲得するためのフレームワークについて述べる．図 2 に，本手法の学習時の処理過程を示す．はじめに，Stable Diffusion [19] などの生成モデルに適用される画風変換 LoRA に対し，内部パラメータを flat 化，PCA で次元圧縮することで，計算可能な次元数へと変換する．その

後，Transformer Encoder を基盤とした triplet 距離学習により LoRA の変換特徴を反映した埋め込み表現を獲得する．

3.1 LoRA の内部パラメータの flat 化

はじめに，与えられた LoRA モデルを展開し，複数のベクトルの列に変換する．この際，LoRA の構造的特徴を保持しつつ，各 LoRA を共通の形式へ変換するために，構造を考慮して 1 次元のベクトルに変換（すなわち，flat 化）する．

画風変換 LoRA の内部パラメータは，複数の圧縮行列と復元行列の 2 種類の低ランク行列として表現される．今回実験で使用した Stable Diffusion-v1-5[3]では，264 個のレイヤそれぞれに対して，圧縮行列と復元行列が 1 つずつ与えられている．つまり，今回の場合，LoRA は 528 個の行列として表される．これらの行列 1 つ 1 つを flat 化し，ベクトルに変換する．

一般的に共有されている LoRA は圧縮度合い（すなわち，ランク）が異なる場合がある．そのため，単純に各行列を flat 化した場合，ランクの異なる LoRA 間ではベクトルの次元数が一致しなくなる．そこで，本研究では圧縮行列と復元行列の積を取り，ベースモデルと同形式に復元してから flat 化を行った．こうすることで，ランクに依らず，共通の長さのベクトルが得られる．

復元後のパラメータは，264 個のレイヤごとに分割し，各レイヤのパラメータを個別に flat 化して 1 次元ベクトルとする．LoRA のレイヤ構造を保ったまま flat 化することで，各レイヤが持つ構造的特徴やパラメータ特徴を保持しつつ，次元圧縮やその後の埋め込み表現学習へと接続可能な形式へ変換する．

3.2 Incremental PCA による次元圧縮

次に，flat 化されただけのベクトルは冗長で学習に向かないので，次元圧縮を行う．

レイヤごとの flat 化処理によって，LoRA の内部パラメータを，レイヤ構造を保持した高次元のベクトル列として表現できた．この際，各ベクトルはレイヤによって数十万から数千万次元になる場合もある．このような長大で長さの異なるベクトルを機械学習の入力として扱うのは，非効率的である．

そこで本研究では，各レイヤを表すそれぞれのベクトルを次元圧縮し，意味を保ちながら各ベクトルを統一された低次元ベクトルに変換する．今回の実装では，次元圧縮手法として IPCA を用いた．IPCA は，バッチ単位で主成分を逐次更新する仕組みを持ち，従来の PCA に比べてメモリ効率に優れ，大規模なベクトル集合に対しても実用的に適用可能である．

この際，flat 化した後でも特に次元数が大きい一部のレイヤは，分散処理による次元圧縮を行った．これは，巨大なベクトルを正しく次元圧縮するためには，データ量と計算機資源を大量に必要とするためである．具体的には，ベクトルを複数のチャンクに分割し，各チャンクごとに独立に次元圧縮を行った後，それらを結合した．

こうしたレイヤ単位での次元圧縮によって，LoRA の構造的情報を保ったまま全体のベクトル表現をコンパクトに再構成しなおすことができた．こうすることで，LoRA モデルを複数の低次元の密ベクトルとして表せるので，一般的な Transformer による埋め込み学習が可能になった．

3.3 Transformer Encoder を基盤とする Triplet Network による距離学習

こうして得られた LoRA を表すベクトル列をもとに，実際にモデルを埋め込み表現として表す．この際，元のベクトル列から，似たモデルのベクトルどうしは近くに，似ていないモデルどうしは遠くになるよう，距離学習を用いてエンコードする．そのために，Transformer Encoder を用いて，Triplet Network による学習を行う．

本研究のアイデアの中核は，レイヤ単位で次元圧縮された LoRA のベクトル列に対して，意味的な類似性に基づく埋め込み表現を学習する点にある．人間がモデルの類似性を評価する際には，絶対的な尺度ではなく，複数の対象を比較する相対的な判断に基づいて類似性を捉えることが多い．たとえば，あるモデルを基準としたときに，別の二つのモデルのうちどちらがより近いかを比較するような判断は,「どの程度似ているか」を絶対値で評価するよりも自然である．このような人間の比較的・関係的な判断様式に近い学習枠組みが，表現学習において有効であることが近年の研究により示されている [20–22]．そこで本手法では，Transformer Encoder を基盤とする Triplet Network を構築した．Triplet 損失に基づく距離学習によって，人間の類似性判断に近い形で LoRA 間の相対的な変換特徴の違いを学習することを目的とする．

Triplet Network を構成するエンコーダは，Transformer Encoder とその後段の per-layer MLPs から構成される．Transformer Encoder は token embedding 層を持たず，学習可能な position embedding 層と Multi-Head Attention 層，Feedforward Network 層からなる構造を採用する．各 LoRA は 264 個のレイヤからなる圧縮済みベクトル列として入力され，position embedding 層によって各レイヤの位置情報が付与される．これにより，LoRA の重み情報を保持したまま，レイヤの順序に関する構造的情報（どのレイヤが何番目の位置にあるレイヤか）を明示的に導入できる．

Multi-Head Attention 層は，トークン間の相互関係を考慮した表現へと変換を行う．ここでいう相互関係とは，Text Encoder における層の連鎖や，UNet における attention ブロックと feed-forward ブロックの配置といった，LoRA 内部のレイヤ構造を指す．これらの関係性を自己注意機構によって捉えることで，系列全体から高次の変換特徴を抽出する．最終的に，Transformer Encoder は $T \times D$ のトークンベクトル列（T: トークン数，D: 埋め込み次元）を出力する．

Transformer Encoder の出力は，per-layer MLPs に入力される．近年，Transformer 出力を単純な平均プーリングによって集約することの限界が指摘されており，トークンごとの重要度を学習的に推定するプーリング機構の有効性が報告されてい

3：「stable-diffusion-v1-5/stable-diffusion-v1-5」: https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5

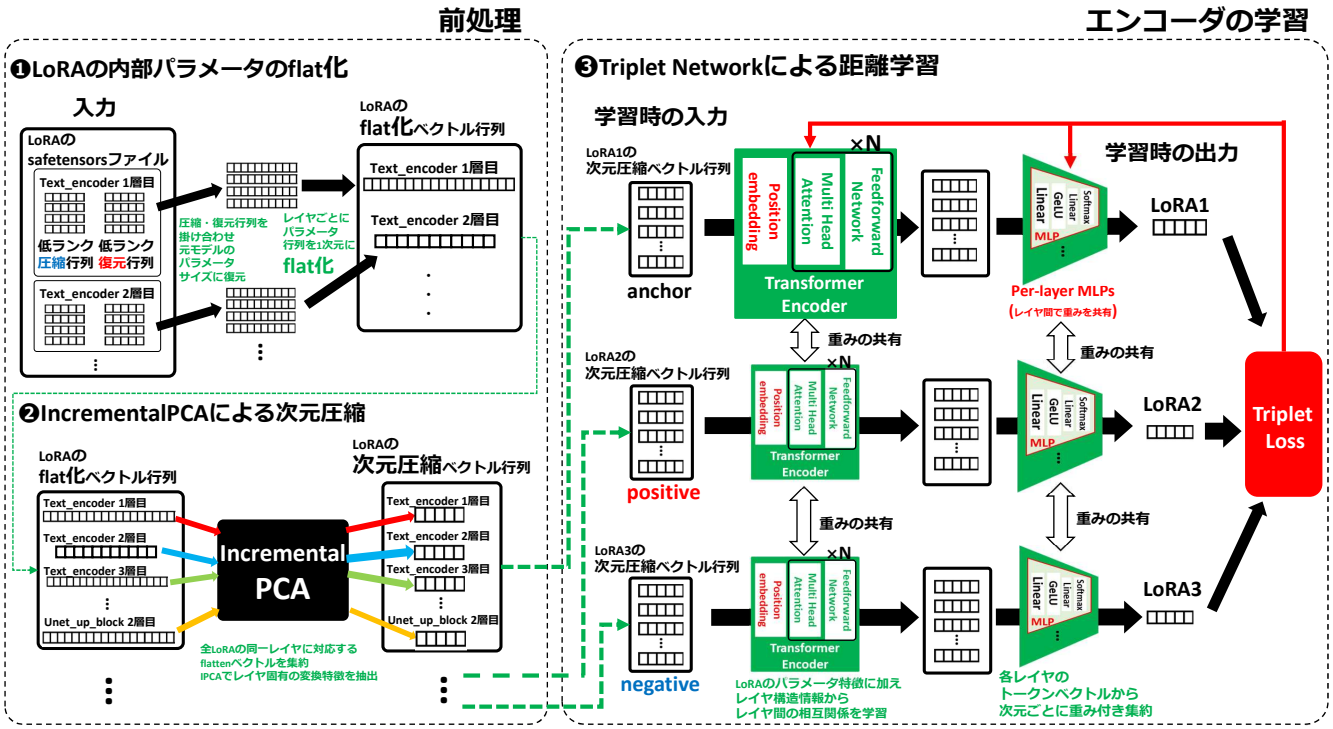

図 2　提案手法の学習時の流れ．入力の各 LoRA モデルの内部パラメータから 264 レイヤ分のもとの行列を復元・flat 化し，レイヤ単位で IPCA を適用することで，圧縮ベクトル列として LoRA を表現する．得られた表現から triplet を構成し，Transformer Encoder と per-layer MLPs を通じて LoRA 全体を 1 ベクトルに集約し，triplet 損失により構造的特徴に沿った埋め込みを学習する．

る [23–25]．本研究で用いる MLP も，Transformer 出力を適応的に統合するプーリング機構として位置づけられる．per-layer MLPs は，各レイヤのトークンに対して同一の MLP を適用する構成であり，この MLP はレイヤ間およびモデル間で重みを共有する．MLP は GELU [26] を挟んだ 2 層の全結合構造と softmax 層からなり，各トークンの D 次元ベクトルを同一次元の重みスコアベクトルへと変換し，softmax 層によってこれらの係数は正規化され，トークンごとの重みに変換される．得られた重みを各トークンベクトルに要素ごとに乗算し，すべてのトークンにわたって平均を取ることで，D 次元の固定長ベクトルを得る．

学習には，triplet 損失を用いる．これは，アンカー例・正例・負例のラベルがついた三つ組の埋め込み表現に対して，埋め込み空間上でアンカー例と正例の距離が負例よりも小さくなるように訓練を行うものである．この三つ組による比較構造は，「ある対象に対してどちらがより近いか」を判断する人間の類似性評価の形式と対応している．この学習により，ある LoRA をアンカーとしたときの相対的な類似関係を反映した埋め込み表現を獲得することが可能となる．

このとき，MLP も triplet 損失によって Transformer Encoder と同時に学習されるため，MLP が出力するトークンベクトルに対応した各次元の重みは，アンカー例と正例を近づけ，負例と区別するという triplet 識別目的を最小化するように最適化される．したがって，これらの重みは，LoRA モデル間の識別において各トークンベクトルの各次元がどれほど寄与したかを反映した量と解釈できる．この性質により，本手法は各レイヤの寄与を考慮した重み付き集約を実現し，単純な平均プーリングよりも識別的な埋め込み表現を得ることが可能となる．このように，Transformer Encoder と per-layer MLPs を組み合わせた Triplet Network による距離学習によって，LoRA 内部の構造的特徴およびレイヤ間関係を捉え，LoRA モデル間の相対的な類似関係を反映した埋め込み表現を構築している．

4　評　　価

本節では，提案手法の有効性，実用性を検証するため，

（1）　学習が成功したかを計測するための自動評価，

（2）　得られた埋め込み表現が人間の類似性判断とどの程度一致するかを判定するための被験者評価，および，

（3）　実用性を評価するための類似 LoRA 検索タスクにおけるランキング性能の被験者評価

について，それぞれ実験した．

4.1　データセット

実験のために，Stable Diffusion 用の画風変換 LoRA を収集し，変換画像の類似度から triplet を作成し，学習用データセットを構築した．使用した LoRA の safetensors ファイルと付随するメタデータは，Civitai から収集した．

はじめに，1,000 件の LoRA を，学習，検証，評価用に分割

し，それぞれ用のデータセットとした．こうすることで，PCA の次元圧縮モデルの学習，Transformer の距離学習において，訓練時と評価時で全く異なる LoRA を使うようにした．無作為に収集した 1,000 件の LoRA のうち，うまく変換の行えなかったものや，倫理的に問題があると判断されたものは取り除いた．実際には，549 件を学習用，100 件を検証用，150 件を評価用 LoRA として用いた．

こうして分割されたデータセットの中から LoRA を組み合わせて triplet を作成した．Triplet の学習では，1 件のデータはアンカー例，正例，負例の 3 つ組からなる．ここで，「実際にそれらの LoRA で画像を変換した際に，変換後の画像が似た画像になるか」を基準に，正例と負例をそれぞれ設定した．Triplet を作成するために，それぞれの LoRA で，10 枚のサンプル画像を画風変換した．変換後の画像を，画像エンコーダである DINOv2 [27] でベクトル化し，総当たりで LoRA 間の変換画像の類似度を計算した．こうして得られた LoRA どうしの変換類似度について，類似度 0.6 以上のペアをアンカー例と正例ペアとし，類似度 0.5 以下のペアをアンカー例と負例のペアとした．同じアンカー例を共有するペアを結合し，アンカー例，正例，負例からなる triplet として扱った．

この際，データセットの質を高めるために，ベースモデルの最適化や，人手でのクレンジングを行った．変換もととなるサンプル画像は，被写体や画像の種類が多様になるよう，Stable Diffusion XL で生成した．サンプル画像の変換に，LoRA 適用に特化された Stable Diffusion のモデルである AnyLoRA[4]を用いた．このモデルは，Stable Diffusion 1.5 用に学習された LoRA を適用すると，もとの Stable Diffusion 1.5 よりも強く LoRA の特徴を反映させ，出力結果も奇麗になりやすいモデルである．

サンプル画像をそれぞれの LoRA で画風変換した際に，Seed 値によっては，うまく変換が実施されなかったり，変換結果が乱れる場合があった．この場合，画風とは別に，変換が成功したかどうかによって類似度計算が行われる危険性がある．そのため，人手ですべての変換結果を目視で評価し，うまく変換されていない画像を取り除いた．具体的には，変換結果が意味をなさない（1 色の画像や，模様のような変換が行われた）場合や，変換もとからほとんど変更がない場合には，Seed 値を変えて再変換した．

こうして作成された triplet の中から，学習の効率化のために，semi-hard negative 例を含む triplet を抽出した．Triplet Network の学習時に，学習に寄与しない簡単すぎるデータや，学習を不安定にしうる難しすぎるデータを減らし，境界例のデータを使って学習を進める手法が知られている [20, 28, 29]．そこで本研究では，学習時の LoRA の組み合わせとして，1 件のアンカー例と正例のペアに対し，複数の難易度の異なる負例を，異なる比率で組み合わせて triplet を構築した．具体的には，類似度が 0.4 ～ 0.5 の semi hard な組み合わせが 15，0.3 ～ 0.4 の easy な組み合わせが 5，0.0 ～ 0.3 の very easy な組み合わせが 1 の割合になるよう，データセットを構築した．この構成比は，予備実験を通じて，経験的に設定したものである．

本研究で構築した LoRA 類似度評価データセットは，再現性確保のため GitHub リポジトリ [5]上で公開している．

4.2 比 較 手 法

本研究では，以下の 5 手法を比較対象としてアブレーションテストを実施した：

- **提案手法**：LoRA の内部パラメータを flat 化した後，次元圧縮を施し，レイヤ系列として Transformer Encoder に入力し，per-layer MLPs による集約をもとに Triplet 距離学習を行う構成．
- **MLP なし**：Transformer Encoder の出力に対して per-layer MLPs による重み付き集約を行わず，単純平均によって最終ベクトルを得る構成．
- **位置 E なし**：提案手法から絶対位置エンコーディングを除去し，レイヤ間の順序関係を考慮せずに学習したモデル．
- **MLP・位置 E なし**：提案手法から位置エンコーディングおよび per-layer MLPs の両方を除去し，triplet 距離学習のみを行った構成．
- **ベースライン**：提案手法から triplet 距離学習を除去し，flat 化および次元圧縮後のベクトルを単純平均して扱う手法．

これらの手法を比較することで，距離学習の有無が性能に与える影響，ならびに位置エンコーディングおよび per-layer MLPs 層の導入によって LoRA モデルの変換特徴をどの程度考慮できるかを明らかにする．

4.3 実　　装

実装には Python を用い，PyTorch[6]を用いて Transformer Encoder および per-layer MLPs などの各モジュールを構築した．学習時の主なハイパーパラメータとして，バッチサイズを 128，学習率を 1×10^{-4}，Triplet Loss のマージンを 0.3 に設定した．

提案モデルに関するパラメータとして，Transformer への入力次元（IPCA 圧縮後のトークン次元）を 256，Feedforward Network 層の次元数を 512，MLP の隠れ層の次元数を 128，最終出力次元数を 256 とし，学習エポック数を 15 に設定した．エポック数については，第 4.1 節で構築した検証用データに対する Triplet Loss が最小となったエポックを基準として決定している．

提案手法の実装に用いたコードは，GitHub リポジトリ [7]に

4：「Lykon/AnyLoRA」：
https://huggingface.co/Lykon/AnyLoRA

5：「LoRA 類似度評価データセット」：
https://github.com/shoji-lab/LoRA_triplet_dataset_Kanada

6：「PyTorch」：
https://pytorch.org/

7：「LoRA 埋め込み学習モデル」：
https://github.com/YuroKanada/Learning-Embedding-Representations-of-LoRA-Models-from-Adapter-Weights

表 1　評価データに対する各手法の Triplet Loss および Triplet Accuracy.（***は提案手法の Triplet Accuracy に対して McNemar [30] 検定により $p < 0.001$，を示す.）

手法名	Triplet Loss	Triplet Accuracy
提案手法	0.218	0.731
MLP なし	0.221	***0.720
位置 E なし	0.272	***0.631
MLP・位置 E なし	0.299	***0.594
ベースライン	0.322	***0.505

て公開している.

4.4　実験 1：学習妥当性に関する自動評価

提案手法が適切に学習されているかを検証するため，各手法に対して Triplet Loss および Triplet Accuracy に基づく推論性能の定量評価を行った．Triplet Loss は学習時に用いた損失関数に準拠した誤差指標であり，Triplet Accuracy は各 triplet において，アンカーとポジティブ間の類似度がネガティブよりも高い割合を評価する指標である．特に後者は，距離ではなく類似度に基づく評価指標であり，人間の類似性判断により近い観点から性能を評価できる.

第 4.1 節で構築した評価データに対する，各手法の Triplet Loss および Triplet Accuracy を表 1 に示す．提案手法は，すべての比較手法の中で，最も低い Triplet Loss および最も高い Triplet Accuracy を示した．一方で，triplet 距離学習を行わないベースラインは，他の手法と比較して最も低い性能となった.

4.5　実験 2：人間の相対的類似性判断と埋め込み表現との整合性評価

本実験では，各手法によって得られた埋め込み表現が，人間による相対的な類似性判断とどの程度整合しているかを検証することを目的とする．ここでは，「類似した LoRA は類似した変換特徴を有する.」という仮定に基づき，LoRA によって生成された出力画像を介して埋め込み表現の有効性を評価する.

評価対象とする 150 件の LoRA モデルのそれぞれをアンカーとして 10 回ずつ使用し，アンカー LoRA1 件と候補 LoRA2 件からなる計 1,500 件の triplet を構成した．1 件の triplet について，アンカー例および候補 LoRA により生成された変換画像を用いて評価を行った．1 件の triplet に対し被験者 5 名に，アンカー LoRA で変換した画像と比較してより類似していると判断される変換画像を選択させ，多数決により正解ラベルを付与した．被験者には図 3 のように，LoRA の内部パラメータやメタデータを提示せず，元画像とアンカー LoRA による変換結果を参照した上で，候補①および候補②のうちどちらが同様の変換過程を経て生成された画像であるかを判断させた．本実験の設定は，著者らの既発表研究 [3] および予備実験の結果に基づき設計したものであり，変換特徴の比較を促すため，元画像からアンカー LoRA による変換結果までの過程を明示的に提示する実験画面を構成した．得られた正解 triplet データに対して，各手法の Triplet Loss および Triplet Accuracy を算出するこ

図 3　変換特徴を比較させる実験画面．元画像とベース LoRA によって変換された画像の変換過程に注目させ，候補 1，候補 2 の LoRA で変換された画像を見て，似たような変換がされたのはどちらかを選択させた.

表 2　人間の相対的類似性判断と埋め込み表現との整合性評価の結果．被験者が出力画像に基づいて付与した正解ラベルから構築した triplet に対する推論結果を示す.（**は提案手法に対して McNemar 検定により $p < 0.01$，*** は $p < 0.001$ を示す）

手法名	Triplet Loss	Triplet Accuracy
提案手法	0.170	0.780
MLP なし	0.182	0.772
位置 E なし	0.211	***0.712
MLP・位置 E なし	0.195	**0.741
ベースライン	0.256	***0.621

とで，学習された埋め込み表現が人間の直感的な類似性判断とどの程度整合しているかを定量的に評価した.

人間の類似性判断と埋め込み表現の整合性に関する実験結果を表 2 に示す．Triplet Loss および Triplet Accuracy のいずれにおいても，提案手法はすべての比較手法と比べて最も高い性能を示した．そして，triplet 距離学習を行わないベースラインは，4.4 節と同様に他の手法と比較して最も低い性能となった．なお，被験者 5 名による回答の完全一致率は 0.722，平均ペア一致率は 0.867 であった.

4.6　実験 3：視覚的類似性に基づく検索性能評価

本実験では，各手法から得られた LoRA 埋め込み表現を用い

表 3　検索タスクに基づく各手法のランキング性能評価結果．30 件のクエリ LoRA に対する検索結果を，人手評価に基づいて構築した正解ランキングと比較し，全クエリの平均 Recall@10 および NDCG@10 により評価した．(* は提案手法に対して Wilcoxon の符号付き順位検定 [32] により $p < 0.05$，** は $p < 0.01$ を示す)

手法名	Recall@10(± std)	NDCG@10(± std)
提案手法	0.420(± 0.106)	0.513(± 0.117)
MLP なし	**0.337(± 0.140)	**0.411(± 0.122)
位置 E なし	**0.310(± 0.161)	*0.401(± 0.165)
MLP・位置 E なし	*0.323(± 0.168)	**0.406(± 0.167)
ベースライン	*0.353(± 0.131)	*0.437(± 0.119)

た LoRA モデル検索の実用的性能を評価するため，検索タスクに基づくランキング評価を行った．本タスクは第 4.5 節と同様に，「LoRA の埋め込み空間上の距離は，対応する変換画像の視覚的類似性を反映する」という仮定に立脚している．

評価対象の LoRA モデルの中からランダムに選択した 30 件のクエリ LoRA に対し，各手法の埋め込み空間上で，評価対象の全 LoRA とのコサイン類似度を総当たりで算出し，上位 10 件を類似 LoRA として検索した．得られた検索結果に対しては，各クエリ LoRA をアンカー例として固定し，検索結果として得られた LoRA モデルの中から候補 LoRA を 2 件ずつ選択することで triplet を構成した．候補 LoRA の選択にあたっては，検索結果として得られた LoRA モデル間のすべての組合せを対象として triplet を構成した．本実験では，9 名の被験者に第 4.5 節と同様に LoRA の内部情報を提示せず，出力画像のみに基づいてクエリ LoRA による変換と類似した変換特徴がみられる LoRA を選択させた．被験者による相対的類似性判断に基づく triplet ラベルを用いて，Copeland 法 [31] により全順序化を行い，正解ランキングを構築した．この正解ランキングに対する各手法の検索結果を，Recall@10 および NDCG@10 により評価した．

なお，事前実験として，クエリと候補を一対一で比較する絶対評価形式を試みたが，人間にとって LoRA モデル間の類似性について「どれくらい似ているか」と評価することは困難であることが確認された．そのため，本研究では，triplet 形式による相対評価を用いることで，疑似的なランキング評価を実現している．本実験は，変換後の画像を介した人手評価に基づく正解ランキングを用いることで，埋め込み表現が検索タスクにおいて有効かつ安定的に機能しているかを間接的に検証することを目的とする．

検索タスクに基づくランキング評価の結果を表 3 に示す．各手法のランキング性能を比較すると，提案手法は Recall@10 および NDCG@10 の両指標において最も高いスコアを示した．また，各指標の標準偏差に注目すると，提案手法は他手法と比較して分散が小さいことが確認された．

実際に検索された結果のケーススタディを図 4 に示す．

図 4　クエリ LoRA に類似する LoRA 上位 3 件の人間の回答と比較手法に関するケーススタディ．一部画像については著作権の問題でモザイクを付与．

5　考　　察

本節では，学習妥当性に関する自動評価，視覚的判断との整合性評価，検索性能評価の結果に基づき，提案手法の有効性と課題を考察する．

まず，学習妥当性に関する自動評価について考察する．ベースラインが他のすべての手法と比較して Triplet Loss および Triplet Accuracy の両指標で大きく劣る結果となったことから，出力画像に基づく距離学習を導入することで，LoRA の内部パラメータからモデル間の類似関係を反映した埋め込み表現を学習可能であることが示された．さらに，絶対位置エンコーディングと per-layer MLPs による重み付き集約を組み合わせた構成が，LoRA 埋め込み表現の性能向上に寄与することが確認された．per-layer MLPs 単体では効果が限定的であったが，位置エンコーディングと併用することで，各レイヤの構造的な順序情報を考慮した重み付けが可能となり，LoRA のレイヤ構造に基づく重要度の差異を反映した集約が実現されたと考えられる．

次に視覚的判断との整合性評価である．提案手法は，Triplet Loss が最も低く，Triplet Accuracy が最も高い値を示しており，人間の視覚的判断に近いモデル間類似関係を学習できていることが示された．一方で，per-layer MLPs による集約を行わない構成との差が有意でなかったことから，per-layer MLPs による重み付き集約は，人間には明示的に判断しにくい細粒度な変換差異を捉えている可能性を示唆している．さらに，位置エンコーディングを導入した構成では人間の判断との整合性が向上していることから，LoRA のレイヤ構造に由来する特徴が，最終的な出力画像の視覚的印象にも反映されている可能性が示唆される．

最後に視覚的類似性に基づく検索性能評価について考察する．まず，提案手法が最も高精度であり，Recall@10 については，ベースラインと比較して 0.07 の改善が見られ，相対的には約 20%の性能向上に相当する．また，NDCG@10 においても 0.51 と最も高い値を示しており，検索結果の上位順位においても，人間の評価と整合した並びが得られていることが分かる．さらに，提案手法は Recall@10 および NDCG@10 の両指標において標準偏差が最も小さく，クエリに依存せず安定した検索性能を示している．これは，LoRA モデル検索のように正

解が一意に定まらない実用タスクにおいて，特に重要な性質であると考えられる．実際に，図 4 に示す検索結果を確認すると，衣服の画風変換を行うクエリ LoRA に対して，提案手法は人間の判断に近い結果を返しており，同一の衣服変換コンセプトを持つ LoRA を上位に検索できている．一方で，他の手法では，キャラクター LoRA や無関係なコンセプト LoRA が上位に現れており，クエリ LoRA の変換意図と乖離した結果となっていることが分かる．

加えて，提案手法の実用性を確認するため，未知 LoRA モデルに対する埋め込み表現の算出時間を測定した．未知 LoRA100 個を入力とした場合，重みの復元と flat 化，次元圧縮，および学習済み提案モデルによる推論までに要した時間は，合計で 8 分 36 秒であった．なお，次元圧縮については，追加の LoRA に対する変換処理に要する時間のみを理論値として算出している．この結果は，一度次元圧縮器および埋め込みモデルを構築してしまえば，未知 LoRA に対しても現実的な時間でベクトル化が可能であることを示しており，大規模な LoRA モデル群を対象とした検索・管理タスクにおいても，本手法が実用的に適用可能であることを示唆している．

6 まとめと今後の課題

本研究では，LoRA モデルの内部パラメータに基づき，Transformer ベースの Triplet Network による距離学習を通じて，推論時に出力例やメタデータを必要としない埋め込み表現の学習を実現した．

学習妥当性に関する自動評価では，出力画像に基づく距離学習を導入することで，LoRA の内部パラメータからモデル間の類似関係を適切に獲得できることを確認した．さらに，人間の相対的類似性判断との整合性評価により，提案手法が視覚的判断と整合したモデル間類似関係を捉えられることを示した．加えて，検索タスクに基づくランキング評価において，提案手法は比較手法に対して平均性能および安定性の両面で最も高い性能を示し，LoRA モデル検索における実用性を確認した．

今後の課題として，まず学習データのジャンル分布の偏りが挙げられる．本研究では 549 件の多様な画風変換 LoRA を用いて学習を行ったが，収集された LoRA モデルのジャンルバランスは必ずしも均衡ではなく，より多様で均等なデータ構成により性能向上の余地がある．また，本研究は Stable Diffusion v1.5 をベースモデルとして検証したが，他の生成モデルへの適用可能性は未検討である．提案手法はベースモデルに応じたハイパーパラメータ設定により他モデルにも適用可能な設計であるため，今後は SDXL など他のベースモデルに対する有効性を検証する必要がある．

生成モデル検索のように正解が一意に定まらず主観的判断が関与するタスクにおいては，平均的な性能の向上に加えて，性能の安定性も重要な評価軸となる．本研究の結果は，未知のモデルに対しても内部構造情報を考慮した埋め込み表現がその両立に寄与することを示唆しており，生成モデルの管理・検索・再利用といった実用的応用に向けた有効なアプローチであると考えられる．

謝　　辞

本研究の一部は JSPS 科研費 25K03229，25K03228，および 24K03228 の助成を受けて実施されました．ここに記して謝意を表します．

文　　献

[1] Edward J Hu, et al. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[2] Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Neural Information Processing Systems*, 2017.

[3] 金田悠路, 大江優真, ファムフーロン, 加藤誠, 大島裕明, 藤田澄男, 荘司慶行. 画風変換 lora の内部パラメータによるモデルの埋め込み表現の獲得. 2025.

[4] Ziyu Zhao, et al. Loraretriever: Input-aware lora retrieval and composition for mixed tasks in the wild. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 4447–4462. Association for Computational Linguistics, August 2024.

[5] Reza Qorbani, et al. Semantic library adaptation: Lora retrieval and fusion for open-vocabulary semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9804–9815, June 2025.

[6] Ziheng Ouyang, et al. K-lora: Unlocking training-free fusion of any subject and style loras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13041–13050, June 2025.

[7] Hongxu Chen, et al. Iteris: Iterative inference-solving alignment for lora merging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4829–4838, June 2025.

[8] Enis Simsar, et al. Loraclr: Contrastive adaptation for customization of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13189–13198, June 2025.

[9] Xiaolong Jin, et al. Conditional lora parameter generation, 2024.

[10] Bedionita Soro, et al. Diffusion-based neural network weights generation. In *The Thirteenth International Conference on Learning Representations*, 2025.

[11] Yi-Kai Zhang, et al. Model spider: learning to rank pre-trained models efficiently. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2023.

[12] Kaichao You, et al. Logme: Practical assessment of pre-trained models for transfer learning. In *International Conference on Machine Learning*, 2021.

[13] Yuanchun Li, et al. Modeldiff: testing-based dnn similarity comparison for model reuse detection. In *Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis*, pp. 139–151. Association for Computing Machinery, 2021.

[14] Vu Thi Ngoc Anh, Yoshiyuki Shoji, Yuma Oe, Huu-Long Pham, and Hiroaki Ohshima. Image-generation ai model retrieval by contrastive learning-based style distance calculation. In *Proceedings of the 31st International Conference on Multimedia Modeling (MMM 2025)*, p. 101–114, 2025.

[15] Huu-Long Pham, Ryota Mibayashi, Takehiro Yamamoto, Makoto P. Kato, Yusuke Yamamoto, Yoshiyuki Shoji,

and Hiroaki Ohshima. Pre-trained bert model retrieval: Inference-based no-learning approach using k-nearest neighbour algorithm. *IEICE Transactions on Information and Systems*, Vol. E108.D, No. 8, pp. 872–882, 2025.

[16] Chenxi Liu, et al. A lora is worth a thousand pictures, 2024.

[17] Amil Dravid, et al. Interpreting the weight space of customized diffusion models. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2025.

[18] Theo Putterman, et al. Learning on loras: Gl-equivariant processing of low-rank weight spaces for large finetuned models. In *Workshop on Neural Network Weights as a New Data Modality*, 2025.

[19] Robin Rombach, et al. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

[20] Florian Schroff, et al. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[21] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019.

[22] Martin Hebart, Charles Zheng, Francisco Pereira, and Chris Baker. Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, Vol. 4, pp. 1–13, 11 2020.

[23] Jiacheng Chen, Hexiang Hu, Hao Wu, Yuning Jiang, and Changhu Wang. Learning the best pooling strategy for visual semantic embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15789–15798, June 2021.

[24] Chen Huang, Walter Talbott, Navdeep Jaitly, and Joshua M Susskind. Efficient representation learning via adaptive context pooling. In *Proceedings of the 39th International Conference on Machine Learning*, Vol. 162 of *Proceedings of Machine Learning Research*, pp. 9346–9355. PMLR, 17–23 Jul 2022.

[25] Piotr Nawrot, Jan Chorowski, Adrian Lancucki, and Edoardo Maria Ponti. Efficient transformers with dynamic token pooling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, p. 6403–6417. Association for Computational Linguistics, 2023.

[26] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv: Learning*, 2016.

[27] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russ Howes, Po-Yao (Bernie) Huang, Shang-Wen Li, Ishan Misra, Michael G. Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision. *ArXiv*, Vol. abs/2304.07193, , 2023.

[28] Edgar Simo-Serra, et al. Discriminative learning of deep convolutional feature point descriptors. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[29] Yao Zhai, et al. In defense of the classification loss for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[30] Quinn Mcnemar. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, Vol. 12, pp. 153–157, 1947.

[31] A. Copeland. A reasonable social welfare function. *Seminar on Applications of Mathematics to Social Sciences*, 1951.

[32] Frank. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, Vol. 1, pp. 196–202, 1945.