THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

電子情報通信学会 論文誌 D 情報・システム

VOL. J107-D NO. 4
APRIL 2024

情報・システムソサイエティ

一般社団法人 電子情報通信学会
THE INFORMATION AND SYSTEMS SOCIETY
THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

画風に基づく作品検索に向けた生成モデルと距離学習に基づく深層クラスタリング手法

朴　炳宣†　　松下　光範†

A Deep Clustering Method with Generation Model and Metric Learning for Content-Based Retrieval Focused on Art Style

Byeongseon PARK† and Mitsunori MATSUSHITA†

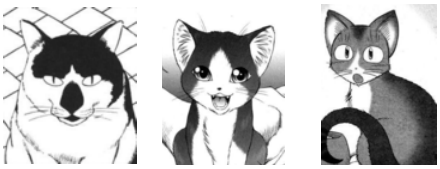
あらまし　本論文では，コミックやイラストのように絵画によって表現されるコンテンツにおける画風に基づいた作品検索システムの実現に向け，深層距離学習を用いたクラスタリングモデルを提案する．既存手法では，人によって設計された特定の要素による特徴量や，固定されたクラスによって最適化された特徴量によってモデルが学習されるため，未知データに対する頑健性が懸念される．そこで，提案手法では，Variational Autoencoder の構造によって画像を再構成するために最適化される潜在空間を Triplet loss によって共同最適化する．未知データを用いた定量評価の結果，提案手法は NMI スコアにおいて 51.71% を達成しており，従来手法よりも 10.61% 向上していることが確認された．更に，定性的評価として，各モデルによって生成された特徴量をもとに任意の画像に対する類似検索を行った結果，提案手法は従来手法よりもクエリ画像と画風における類似度の高いサンプルをより多く提示することができることを確認した．

キーワード　画風，検索，クラスタリング，Metric learning, Triplet loss, Variational Autoencoder

1. まえがき

近年，電子書籍やウェブプラットホームを用いたコンテンツ提供の定着化により，膨大な数のコミックが日々生まれている．しかし，こういった状況により，ユーザが膨大なコンテンツの中から，自身の趣向に適した特徴 (e.g., かっこいいタッチのコミックを読みたい) をもつ作品を探すことはより困難となっている．

作品を構成する特徴の中で，ユーザの趣向に大きく関わる要素として，「画風」が挙げられる．画風とは，絵画に表れた作者の特色であり，自然や光景をそのまま写実する場合を除いて，各作者と作品には必ず固有のスタイルを有するとされる [1]．例えば，図 1 で示すように，同じ被写体であっても作者によって表現方法 (e.g., 線の太さ，彩色方法，目や顔の輪郭などの描写方法) が大きく異なるように，イラスト作品において画風は作品や作者を特徴づける重要な要素となる．



© Shinpei Ito　© Ebifurai　© Kaasan

図 1　画風による描写の違いの一例

各作品の画風が作者の固有の特徴に影響を受けることから，これまで作者に基づく作品検索に向け，画風を構成する特定の要素を用いて特徴量を設計する手法 [2]〜[4] や，深層ニューラルネットワークによって表現された特徴量から作者を分類する手法 [5], [6] が提案されている．しかし，これらの手法では，人によって設計された特徴量の柔軟性や学習時に活用する作者のデータセットの網羅性がボトルネックになり，未知データへの対応が困難となる．

そこで本論文では，画風に基づく作品検索に向けて，深層学習に基づくクラスタリングモデルを提案する．

† 関西大学大学院総合情報学研究科，高槻市
Graduate School of Informatics, Kansai University, 2–1–1 Ryozenji, Takatsuki-shi, 569–1095 Japan
DOI:10.14923/transinfj.2023PDP0042

本手法では，生成モデルである Variational Autoencoder (VAE) [7] によって得られる高度な潜在特徴量について，特徴量同士の距離に基づき学習を行う距離学習 (metric learning) の一種である Triplet loss [8] によって共同最適化を行うことで，未知の作者やドメインに頑健なクラスタリングモデルを構築する．

クラスタリングモデルの評価指標である normalized mutual information (NMI) [9] を用いた定量評価の結果，提案手法は 51.71% を達成しており，従来手法 [6] よりも 10.61% 向上していることが確認された．更に，特定のイラストをクエリーとして類似検索を行う場面において，既存手法より直感に近い結果を得られることを確認した．

2. 関 連 研 究

2.1 画風に基づく検索に関する研究

作者の画風を構成する要素の抽出とその要素に基づく作者の分類に関する研究として，Chu [2] らや安田ら [3]，福田 [4] の研究が挙げられる．安田ら [3] は，コミックの 1 ページあたりのコマの数やキャラクターの数といった各要素を作者の特徴量として定量化し，統計モデルによって作者を分類する手法を提案した．また，福田ら [4] は作者の画風を特徴づける要素として人物の顔に着目し，目や口の大きさや位置などの要素を特徴量として用いて作者を分類する手法を提案した．しかし，これらの手法では，特徴量の品質は人間による設計や各要素の抽出するための付加タスクの精度に依存している．

一方で，Kim [5] と増子ら [6] は，人によって設計された特徴量ではなく，深層ニューラルネットワークによって画像そのものから得られる特徴量のみを用いて作者を分類する手法を提案した．Kim [5] は，コミックのページやコマを入力とし畳み込みニューラルネットワークから得られる特徴量を用いた作者分類手法を提案している．また，増子ら [6] は，距離学習の中でも Softmax loss [10] に基づく手法である ArcFace [11] によって特徴量を最適化することにより，高い分類精度を達成した．しかし，これらの手法では，モデルが学習データ内の既知の作者クラスの確率分布を再現するように最適化されるため，未知クラスへの対応性が懸念される．

2.2 Deep clustering model

特定の成分や既定のクラスを用いずデータを分類またはクラスタリングする手法として，VAE や GAN [12] などの生成モデルによって得られる特徴量を用いる手法や，距離学習を用いた特徴量の最適化を用いる手法が挙げられる．

まず，生成モデルを用いたクラスタリング手法として，Dilokthanakul ら [13] らや Lim ら [14]，Chang ら [15] の研究が挙げられる．Dilokthanakul ら [13] や Lim ら [14] は，VAE は入力データを制約された潜在変数に圧縮した後，再度潜在変数を用いて入力データを復元する生成過程から教師データを用いずともクラスタリングに有効な潜在空間が得られる特徴をもつ点に着目し，VAE における潜在変数を混合ガウスモデル (gaussian mixture model, GMM) [16] を用いて最適化を行うことによってクラスタリング精度を向上させた．また，Chang ら [15] は，VAE による再構成損失に加え，GAN における Discriminator を用いた損失関数を導入することによってクラスタリング精度を高めている．これらの手法によって特定のクラスを教師データとして用いずクラスタリングモデルを構築できる一方，色や形といった明示的な特徴をもとにモデル自らデータを分類することによって学習が行われるため，明確な指示を用いずに図 1 の画像群のような同じ視覚的特徴をもつ被写体の中から画風の違いを見出すことは困難であると考えられる．

また，距離学習を用いた分類手法として，Zeng ら [17] や Yang ら [18]，Sain ら [19] の研究が挙げられる．距離学習は，モデルによる特徴量について既定のクラスの確率分布を再現するように学習する一般的な分類モデルの学習とは異なり，各データの特徴量の距離がクラス間の関係性 (e.g., 同一クラスであるか否か) に基づいて学習されることにより，未知のデータにも頑健な分類精度を得られる特徴を持つ [20]．Zeng らは，様々な方向から撮影された人物を識別する人物再同定 (person re-identification) タスクについて，距離学習の一種である Triplet loss [8] を用いることでモデルの識別精度を向上させることに成功している．また，Yang ら [18] は，画像に含まれる感情という抽象的な要素に対する分類タスクについて，Triplet loss を用いて感情ラベル間の階層関係に基づく特徴量の最適化を行うことで，様々なデータセットにおいて汎化性能の高いモデルを構築している．また Sain ら [19] は，スケッチに基づく画像検索 (Sketch-based image retrieval) における，同一の被写体について異なる形で描かれたスケッチの分類精度が低下する問題について，VAE モデルの学習時に Triplet loss を組み合わせることでスケッチ

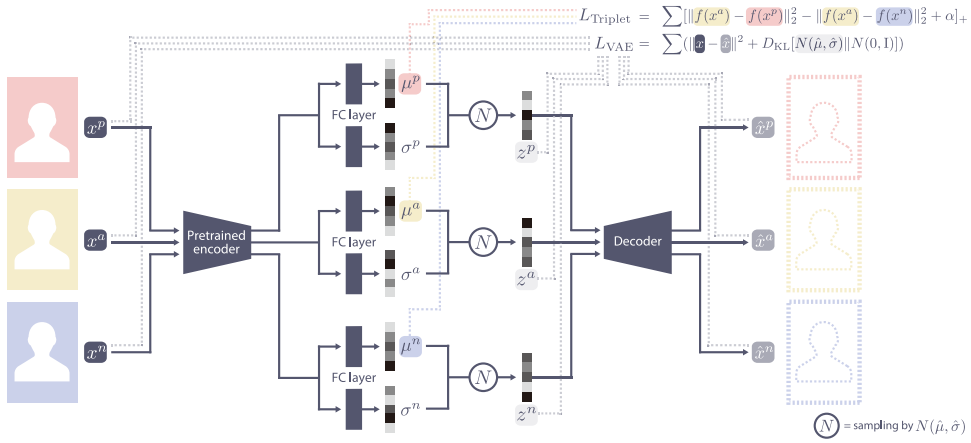


図 2　提案手法の概要図

と元画像間の良質なマルチモーダル特徴量空間を獲得し，検索精度を向上させた．しかし Sain らの手法では，Triplet loss を用いる際，スケッチと元画像の特徴量間の距離を近づけるよう設計することで，異なる画風で描かれたスケッチ間の多様性 (diversity) に対する頑健性を向上している反面，スケッチ間の画風の特徴を区別可能な (discriminative) 状態にできるかについては必ずしも保証されていない．更に，VAE モデルの学習時の入力データと潜在変数によって復元されたデータ間の再構成損失 (reconstruction loss) は，学習初期に望ましくない極小値に陥りやすいため学習が不安定になることが多いとされる点 [21], [22] や，Triplet loss では triplet の選択にモデルのパフォーマンスが大きく依存する点 [11] といった懸念も残っている．

2.3　本研究の位置付け

本研究では，画風に基づく作品検索に向け，生成モデルと距離学習に基づく深層クラスタリングモデルを提案する．具体的には，VAE の生成過程によって獲得できる高度な潜在空間をクラスタリングに活用することにより，人間によって設計された特徴量や既定のラベルセットに制約をうけないクラスタリングモデルを構築する．また，VAE の潜在空間について，各画像の作者クラスにおける関係性 (i.e., 同一作者であるか否か) に基づき Triplet loss を用いた距離学習によって共同最適化することにより，図 1 のように同様の視覚的特徴をもつ被写体間でも画風に基づくクラスタリングが可能になることを目指す．最後に，VAE や Triplet loss を用いる場合，各手法の特性により学習が不安定になる問題について，Jaing ら [23] や Seo ら [24] が事前学習によって良質な初期パラメーターを得ることでモデルの学習を安定化した事例を踏まえ，教師なし学習に基づく事前学習の導入による学習の安定化を目指す．

3.　提 案 手 法

提案手法の概要図を図 2 に示す．提案モデルでは，VAE モデルの一般的な設計に基づき，入力データ x について潜在変数 z を得るためのエンコーダーと，潜在変数 z に基づき入力データ x を再構成するためのデコーダーによって構成される．このとき，潜在変数 z は，平均パラメーター μ と分散パラメーター σ によって構成される正規分布 $N(\mu, \sigma)$ に従ってサンプリングされる．モデルのパラメーターはニューラルネットワークによって構成され，以下の損失関数 L_{VAE} を最小化するように学習される．

$$L_{\text{VAE}} = \sum_{i}^{N_d}(\|x_i - \hat{x}_i\|^2 + D_{\text{KL}}[N(\hat{\mu}_i, \hat{\sigma}_i)\|N(0, \text{I})]) \tag{1}$$

ここで，N_d はデータ数を，$N(0, \text{I})$ は標準多変量ガウス分布，D_{KL} は Kullback-Leibler (KL) 情報量を表す．

VAE に基づくクラスタリング手法では，潜在変数 z や平均パラメーター μ をクラスタリングのための特徴量として活用する [14], [15]．しかし，潜在変数 z はガウス分布によって制限され，生成プロセス中にクラスター間のスムーズな遷移のために異なるクラスター

間のギャップが排除される傾向があることから，異なるクラスター間で深刻な重複が発生する可能性がある [14], [23], [25]．そこで，提案手法では Chang ら [15] の手法に倣い，平均パラメーター μ をクラスタリング特徴量として活用する．

また，提案手法では，L_{VAE} によって最適化される潜在変数の平均パラメーター μ について，Triplet loss による最適化を加える．Triplet loss では，入力データ x について d 次元を保つユークリッド空間 (Euclidean space) に埋め込む Embedding 表現 $f(x) \in \mathbb{R}^d$ について，ある画像 x^a (anchor) に対して同じ被写体を含む全ての画像 x^p (positive) がより近く，それ以外の画像 x^n (negative) について遠くなるような表現を得ることを目的とする [8]．以下の式は，前述した問題を定式化したものである．

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \tag{2}$$

$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in T \tag{3}$$

ここで，α は positive pair (i.e., x^a と x^p) と negative pair (i.e., x^a と x^n) 間のマージン (margin) を表すパラメーターである．また，T はデータセットに存在する全ての triplet の N_t 個の組み合わせを表す．

提案手法では，triplet を表現するための画像間の関係性について，被写体の一致ではなく，作者の一致によって表現することで，Embedding 表現 $f(x)$ が作者の画風間の距離を再現できるように学習されることを目指す．つまり，特定の作者によって描かれた画像 x_i^a について，同じ作者によって描かれた全ての画像 x_i^p の距離を近づけ，それ以外の画像 x_i^n について距離を遠ざける Embedding 表現 $f(x)$ を得ることを目的とする．

上記の問題を最小化する損失関数 L_{Triplet} は，

$$\sum_i^{N_t} [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha]_+ \tag{4}$$

として定式化される [8]．提案手法では，L_{Triplet} によって最適化する Embedding 表現 $f(x)$ を $\mu(x)$ から得られるベクトルとして扱い，モデルを L_{VAE} と L_{Triplet} によって共同最適化する．

一方，増子ら [6] が作者のクラスタリングに活用した ArcFace [11] では，入力データ x の Embedding 表現 $f(x) \in \mathbb{R}^d$ と n 種類の正解クラスのベクトル $W \in \mathbb{R}^{d \times n}$ の中心位置とのコサイン類似度 $cos(\theta)$ を用いて以下の式のように最適化を行う．

$$-\frac{1}{N_d} \sum_i^{N_d} \log\left(\frac{e^{s(\cos(\theta_{y_i} + \alpha))}}{e^{s(\cos(\theta_{y_i} + \alpha))} + \sum_{j \neq y_i} e^{s(\cos(\theta_j))}}\right) \tag{5}$$

ここで，y_i は i 番目のサンプルの正解ラベル y を，s は $cos(\theta)$ のスケーリングのためのパラメータを示す．このように ArcFace は，正解クラスの確率分布を活用することにより，距離学習手法の中でも既知クラスに対する高い分類性能を発揮する一方，未知クラスへの頑健性が懸念される特徴をもつ．

3.1 Triplet selection

Triplet loss による学習を行う際，式 (3) によって得られる triplet の全ての組み合わせ T にはモデルの最適化に寄与しないサンプルも多く含まれることから，モデルの最適化に有益な triplet を適度に選択することが望ましいとされる [8]．この問題に対して Schroff ら [8] は，学習データの中から既定の数のクラスをもつ画像を mini batch としてサンプリングし，mini batch 中にあるサンプルから得られる positive pair について，Semi-hard negative と呼ばれる下記の式の条件を満たす negative サンプルを加えることにより，有益な tirplet のみを選択することによって学習を安定化及び効率化させている．

$$\|f(x_i^a) - f(x_i^p)\|_2^2 < \|f(x_i^a) - f(x_i^n)\|_2^2 \tag{6}$$

そこで，提案手法では Schroff ら [8] に倣い，各 mini batch ごとに一定の数の作者クラスを含むようにサンプリングし，式 (6) の条件を満たす triplet のみを用いて L_{Triplet} を求める．

3.2 事前学習

VAE モデルの学習の際，式 (1) におけるモデルによって再構成された特徴量 $\hat{x}$ に対する再構成損失を求める項である $(\|x_i - \hat{x}_i\|)^2$ は，学習初期に望ましくない極小値に陥りやすいため学習が不安定になることが多いとされている [21], [22]．また，Triplet loss を用いる場合，適切な triplet の選択にモデルのパフォーマンスが大きく依存することから，データセットによってモデルの学習が不安定になることがあるとされている [11]．このような問題について，Jaing ら [23] や Seo ら [24] は，事前学習モデルを導入し良質なモデルの初期パラメーター得ることによってモデルの学習を安定化した．そこで，提案手法では，自己教師学習に基づく事前学習手法である SimCLR [26] によって事前学

習したエンコーダーを用いる．

SimCLR は，Triplet loss と同様，任意の画像 x^a について x^p と x^n をサンプリングし，positive pair については特徴量の類似度を高く，negative pair については特徴量の類似度が低くなる特徴量を得るように学習を行う．しかし，SimCLR において x^p は x^a から色や形状について変換を加えた生成画像を用いており，x^n はそれ以外の画像全てとしている．つまり，SimCLR では各画像の positive pair と negative pair を作成する際，特定のラベルを必要としないため，画像のみを用いた自己教師学習を実現しており，SimCLR によって学習されたエンコーダーを用いることで様々なタスクに関するパフォーマンスが向上しているとされている．

増子ら [6] の手法でも事前学習モデルが活用されており，事前学習に用いるデータセットとして ImageNet [27] を活用している．しかし，ImageNet に含まれる画像のほとんどは写真であり，Fine-tuning を行う際に用いるイラスト画像のデータセットとはドメイン特性が大きく異なる．そこで，提案手法では，事前学習を行う際に用いるデータセットとして Danbooru 2021 データセット [28] を用いることで，モデルのパフォーマンスの向上を図る．

4. 評　　価

4.1 データセット

表 1 は提案手法を評価するためのデータセットの設定を示す．まず，SimCLR による事前学習のために，ImageNet ILSVRC-2012 データセット [27] の学習セットとして分類されている約 120 万枚の画像と，Danbooru 2021 [28] に含まれる約 500 万枚の画像の中からランダムにサンプリングされたイラスト画像 100 万枚をそれぞれ活用する．

また，Fine-tuning を行う際の学習データとして，Danbooru 2021 データセットのうち事前学習のための画像を取り除いた 400 万枚の画像の作者の中から，事前学習の学習データに含まれていない，かつ 150 枚以上の画像が登録されている作者を 1,000 名をランダムで選定し，各作者のイラスト画像を収集したサブセットを学習データとして活用する．この際，作者によって画像数が偏りすぎないように，各作者ごとの最大画像数を 300 としている．また，評価データとしては，Fine-tuning の際の学習データに含まれる作者 1,000 名のイラスト画像を各作者ごとに 50 枚ずつ別途サンプリングしたサブセットを既知 (seen) 作者に関する評価セットとし，事前学習及び Fine-tuning の際の学習データに含まれない作者 100 名について各作者ごとに 50 枚ずつ収集したサブセットを未知 (unseen) 作者に関する評価セットとして活用する．

表 1　学習及び評価に用いるデータセット

用途	データセット	クラス数	画像数
事前学習	ImageNet [27]	-	1,281,167
	Danbooru [28]	-	1,000,000
学習	Danbooru	1,000	157,330
評価 (Seen)	Danbooru	1,000	50,000
評価 (Unseen)	Danbooru	100	5,000
評価 (Face)	Manga109 [29], [30]	92	157,152
評価 (Body)	Manga109	92	118,715
評価 (Frame)	Manga109	92	103,900
評価 (Cover)	Manga109	91	107

更に，イラスト画像のクラスタリング性能を評価するデータセットとして Manga109 を用いる [29], [30]. Manga109 は，92 名の漫画家によって 1970 年代から 2010 年代に公開された日本のコミック 109 冊で構成されており，日本コミック特有の表現方式によって描かれている点，描かれた時代が大きく離れる点，表紙を除くほとんどの画像は線画によって描かれているといった点において Danbooru 2021 と異なる特性をもつ．本論文では，Manga109 に含まれるアノテーション情報を活用して，各コミックに描かれた人物の顔 (face) や全身 (body)，コマ全体 (frame)，そして表紙 (cover) の画像を抽出し，評価データとして活用する．ここで，表紙画像については一部の作品において画像が含まれなかったり単色のみで描かれているケースがあり，データセットから排除している点について注意されたい．

4.2 モデル詳細

提案手法を評価するために，本論文では，事前学習の有無やデータセット，損失関数の設定などで構成された各条件の組み合わせに基づき，表 2 に列挙されているモデルを構築し調査を行った．まず，距離学習を用いたイラスト作者のクラスタリングにおけるベースラインシステムとして，増子ら [6] と同様に ImageNet によって事前学習を行い，ArcFace を損失関数として活用 (i.e., 表 2 の L_{Arc}) したモデルを用いる．また，VAE モデルを用いた教師なしクラスタリングにおけるベースラインシステムとして，Dilokthanakul ら [13] と同様に既存の VAE モデルの構造に加え GMM を活用して潜在変数をクラスタリングモデルとして最適化する損失関数 (i.e., 表 2 の L_{GMVAE}) によって学習したモデ

表 2　Danbooru 2021 データセットを用いた定量評価結果

Model	Pretraining domain	Loss	Seen			Unseen		
			Accuracy	ARI	NMI	Accuracy	ARI	NMI
(a)	-	L_{Arc}	6.42%	1.08%	47.02%	14.82%	5.13%	35.60%
(b) [6]	ImageNet	L_{Arc}	8.45%	1.93%	48.93%	19.70%	8.17%	41.10%
(c)	Danbooru	L_{Arc}	9.80%	2.60%	50.30%	20.46%	8.93%	43.52%
(d)	-	L_{Triplet}	4.73%	0.45%	45.28%	10.42%	2.57%	29.98%
(e)	ImageNet	L_{Triplet}	11.69%	3.87%	51.41%	26.68%	13.90%	46.45%
(f)	Danbooru	L_{Triplet}	14.61%	5.57%	53.51%	30.22%	16.65%	50.02%
(g)	-	L_{VAE}	4.21%	0.46%	39.06%	7.06%	1.01%	19.93%
(h)	ImageNet	L_{VAE}	4.26%	0.47%	40.98%	7.10%	1.06%	21.22%
(i)	Danbooru	L_{VAE}	4.27%	0.47%	41.28%	7.10%	0.92%	21.65%
(j) [13]	-	L_{GMVAE}	4.42%	0.50%	41.85%	7.44%	1.01%	21.54%
(k)	ImageNet	L_{GMVAE}	4.37%	0.46%	42.76%	7.08%	1.01%	23.11%
(l)	Danbooru	L_{GMVAE}	4.48%	0.54%	42.86%	7.88%	1.20%	23.64%
(m)	-	$L_{\mathrm{Arc}} + L_{\mathrm{VAE}}$	6.69%	1.24%	47.07%	14.60%	5.11%	35.08%
(n)	ImageNet	$L_{\mathrm{Arc}} + L_{\mathrm{VAE}}$	8.45%	1.95%	48.68%	20.04%	8.83%	40.86%
(o)	Danbooru	$L_{\mathrm{Arc}} + L_{\mathrm{VAE}}$	10.53%	3.02%	50.60%	24.66%	12.28%	45.80%
(p)	-	$L_{\mathrm{Triplet}} + L_{\mathrm{VAE}}$	4.00%	0.17%	44.02%	7.62%	1.03%	25.74%
(q)	ImageNet	$L_{\mathrm{Triplet}} + L_{\mathrm{VAE}}$	13.06%	4.55%	52.12%	26.64%	14.52%	47.06%
(r) **(i.e., ours)**	Danbooru	$L_{\mathrm{Triplet}} + L_{\mathrm{VAE}}$	**15.44%**	**6.06%**	**53.94%**	**32.70%**	**18.71%**	**51.71%**

ルを用いる．これらのベースラインシステムを含め，評価に用いる全 18 個のモデルは，最適化に用いる損失関数の設定によって下記の六つのグループに大別される．

(1) L_{Arc} **を用いる** - (a), (b), (c)
(2) L_{Triplet} **を用いる** - (d), (e), (f)
(3) L_{VAE} **を用いる** - (g), (h), (i)
(4) L_{GMVAE} **を用いる** - (j), (k), (l)
(5) L_{Arc} **と** L_{VAE} **を用いる** - (m), (n), (o)
(6) L_{Triplet} **と** L_{VAE} **を用いる** - (p), (q), (r)

更に，事前学習に関する設定による下記の三つのサブセットグループが存在する．

(1) 事前学習を行わない - (a), (d), (g), (j), (m), (p)
(2) ImageNet を用いる - (b), (e), (h), (k), (n), (q)
(3) Danbooru を用いる - (c), (f), (i), (l), (o), (r)

4.2.1　事前学習

事前学習には，18 層の畳み込みニューラルネットワークによって構成された ResNet (i.e., ResNet18) [31] をエンコーダーとしてもち，128 次元の潜在特徴量投影する 2 層によって構成された Multilayer perceptron (MLP) [32] をもつ SimCLR を用いる．学習の際は，学習率を 2.4 とし Layerwise adaptive rate scaling (LARS) [33] によって最適化され，バッチサイズは 2,048 とし 100 エポック間学習を行っている．この際，最初の 10 エポックは Warm-up として，最初は少ない学習率を与え徐々に目標学習率まで向上させており，10 エポック後は 10^{-6} の減少率を用いて徐々に学習率を低下させるように調整している [26]．また，[26] と同様，学習時の入力画像の解像度は 224×224 とし，各画像には Gaussian noise や Gaussian blur を含む 9 種類の Data augmentation を行っている．なお，Data augmentation に関する詳細は [26] を参照されたい．

4.2.2　Fine-tuning

Fine-tuning には，事前学習が行われた SimCLR のエンコーダーに対して，潜在変数 z を構成する平均パラメーター μ と分散パラメーター σ を生成するための全結合層と，z をもとに入力画像 x を再構成するためのデコーダー結合したモデルを用いる．潜在変数 z の次元数は 256 とし，従って平均パラメーター μ と分散パラメーター σ の出力次元数も同様に 256 となる．デコーダーはエンコーダーと同様 ResNet 18 の構造をもっており，各層の順序を逆順にすることによって潜在変数 z を入力とし再構成画像 $\hat{x}$ を出力する．

また，VAE モデルを活用しないモデル (i.e., モデル (a) から (f)) は，エンコーダーによって得られる特徴量を Embedding 表現に変換する全結合層のみをもち，この全結合層によって得られた Embedding についてそれぞれの損失関数に基づいて学習が行われる．この際，Embedding の次元数 d は上記の VAE モデルを用いたモデルの潜在変数 z と同様 256 としている．更に，事前学習を活用しないモデル (i.e., モデル (a), (d), (g), (j), (m), (p)) は初期状態の ResNet 18 のエンコーダーをも

ち，それぞれの損失関数によって学習される．

全てのモデルは，学習率 10^{-4} とし，Adam [34] によって最適化される．この際，Adam のパラメータ $(\beta_1, \beta_2, \varepsilon)$ はそれぞれ $(0.9, 0.999, 10^{-8})$ としている．また，学習は 40 エポック間行われ，**3.1** で先述した手法によって tirplet を選定するために，1 バッチあたり 20 クラスの作者についてそれぞれ 20 枚の画像をサンプリングするため，バッチサイズは 400 となる．このサンプリングは，L_{Triplet} を活用しないモデルにも同じく適用される．更に，L_{Triplet} と L_{Arc} におけるマージン α は 1.0 とし，L_{Arc} における特徴量のスケール s は増子ら [6] と同様 10 としている．

4.2.3 評価指標

各モデルのクラスタリング性能を評価するために，本論文では unsupervised clustering accuracy [35]，average rand index (ARI) [36]，normalized mutual information (NMI) [9] を用いる．ここで，unsupervised clustering accuracy はクラスタの正解ラベルとクラスタリング結果について Kuhn-Munkres アルゴリズム [37] によって対応づけされた結果に基づいて精度を求める指標である．また，ARI はクラスタの正解ラベルと推定されたクラスタリング結果の相関度を評価する指標である．最後に NMI はクラスタ間の正規化された相互情報量を用いた指標である．各指標は 0 から 1 の間を取り(注1)，1 に近ければ近いほど生成されたクラスタが正解データに近いことを表す．

4.3 Danbooru 2021 を用いた評価

4.3.1 定量評価

表 2 は **4.1** で先述した Danbooru 2021 データセットを用いた評価結果を示す．まず，提案手法に基づくモデル (r) は，ベースラインシステムであるモデル (b) と (j) について，全ての評価指標において優位であることが確認された．特に，距離学習を用いたベースライン手法に基づいて学習されたモデル (b) は，既知の作者 (i.e., 表 1 の Seen テストセット) に対する評価において NMI スコアが 48.93% を達成したものの，未知の作者 (i.e., 表 1 の Unseen テストセット) でのスコアは 41.10% に留まり 7.83% 低下している．その反面，提案手法に基づくモデル (r) は，既知作者について 53.94% を，未知作者については 51.71% を達成しており，低下率は 2.23% に留まっている．また，モデル (b) と同様 ImageNet によって事前学習を行い，損失関数として L_{Triplet} のみを用いたモデル (e) の場合でも，未知作者に対する NMI スコアの低下率が 4.96% となり，同様の傾向が伺える．一方，Accuracy や ARI などのスコアにおいては，Seen での精度よりも Unseen での精度が高くなっており，これは各指標が正解データと推定データ間の総ペア数に大きく影響を受けることから，Seen サブセットと Unseen サブセットのクラス数及びサンプル数が大きく異なることが起因していると考えられる．しかし，各サブセットごとに Accuracy 及び ARI スコアを用いて各モデルを比較すると，ベースラインと本手法との優劣 (i.e., モデル (b) vs. (r)) が変わることはなかった．これらの結果から，画風に基づくクラスタリングにおいて，Triplet loss は ArcFace より有効であることが示唆される．

また，VAE モデルを用いた教師なしクラスタリングにおけるベースライン手法に基づいて学習されたモデル (j) は，VAE モデルのみを活用したモデル (g) よりも高い性能を示すものの，ArcFace や Triplet loss などの距離学習に基づいて学習されたモデル (i.e., モデル (a) から (f)) に比べ劣化する傾向が見られた．しかし，VAE モデルに加え距離学習による共同最適化を用いているモデル (i.e., モデル (m) から (r)) は，それぞれ同じ損失関数を用いているモデル同士を比較した場合，明確な性能向上が見られた (i.e., モデル (a)-(c) vs. モデル (m)-(o), モデル (e) & (f) vs. モデル (q) & (r))．これらの結果により，距離学習による特徴量の最適化に加え，VAE モデルの生成過程によって最適化される潜在空間も，画風に基づくクラスタリングにおいて有効であることが確認された．ただし，モデル (d) とモデル (p) を比較した場合，他の条件とは異なり性能劣化が伺える．これは，**3.2** で述べた特徴により，事前学習を用いないことで初期の学習が不安定になったことが原因であると考えられる．

ImageNet データセットや Danbooru データセットを用いた事前学習の有効性は，事前学習を用いていないモデル (i.e., モデル (a), (d), (g), (j), (m), (p)) とそれ以外のモデルを比較することによって明確に確認できる．更に，Danbooru データセットを事前学習に用いたモデル (i.e., モデル (c), (f), (i), (l), (o), (r)) は，ImageNet データセットを事前学習に用いたモデル (i.e., モデル (b), (e), (h), (k), (n), (q)) に比べ未知作者に対する大幅な性能向上が見られ，特にモデル (q) と (r) 間では未知作者に対して 4.65% の大幅な NMI スコア差があったことが確認された．これは，同一ドメインに該当する他の

(注1)：ARI は期待値より不一致度が高い場合 0 以下の値を取り得る．

表 3　t-SNE を用いた各モデルごとの画風特徴量空間の可視化

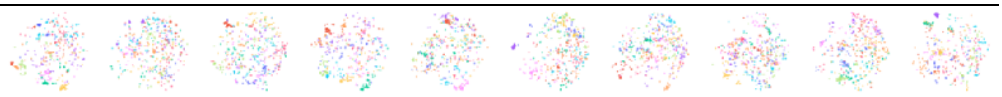
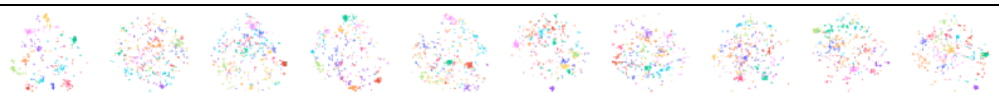
Model	Embeddings
(j)	
(b)	
(r)	

表 4　Manga 109 データセットを用いた定量評価結果.

	Face			Body			Frame			Cover		
Model	**Accuracy**	**ARI**	**NMI**	**Accuracy**	**ARI**	**NMI**	**Accuracy**	**ARI**	**NMI**	**Accuracy**	**ARI**	**NMI**
(b) [6]	4.92%	1.12%	9.98%	5.85%	1.51%	11.41%	6.33%	1.58%	12.21%	74.77%	−0.31%	95.15%
(j) [13]	4.38%	0.63%	5.54%	2.85%	0.28%	3.05%	2.86%	0.30%	2.93%	74.77%	−0.36%	94.88%
(f)	7.29%	2.14%	14.24%	8.66%	2.57%	15.77%	9.00%	2.56%	14.75%	79.44%	15.51%	95.51%
(r) **(i.e., ours)**	**7.56%**	**2.28%**	**14.89%**	**9.21%**	**2.79%**	**15.99%**	**9.21%**	**2.63%**	**15.14%**	**80.37%**	**15.97%**	**95.70%**

データによって事前学習された特徴量が大きく貢献していると考えられる．一方で，ImageNet データセットを事前学習に用いたモデルもまた，事前学習を用いないモデルから大きな性能向上が見られることから，今後 Vincent ら [38] の研究のように，異なるドメインが混合されたデータセットを用いて事前学習に行うことによって汎化性能が向上することも期待される．

4.3.2　t-SNE による可視化

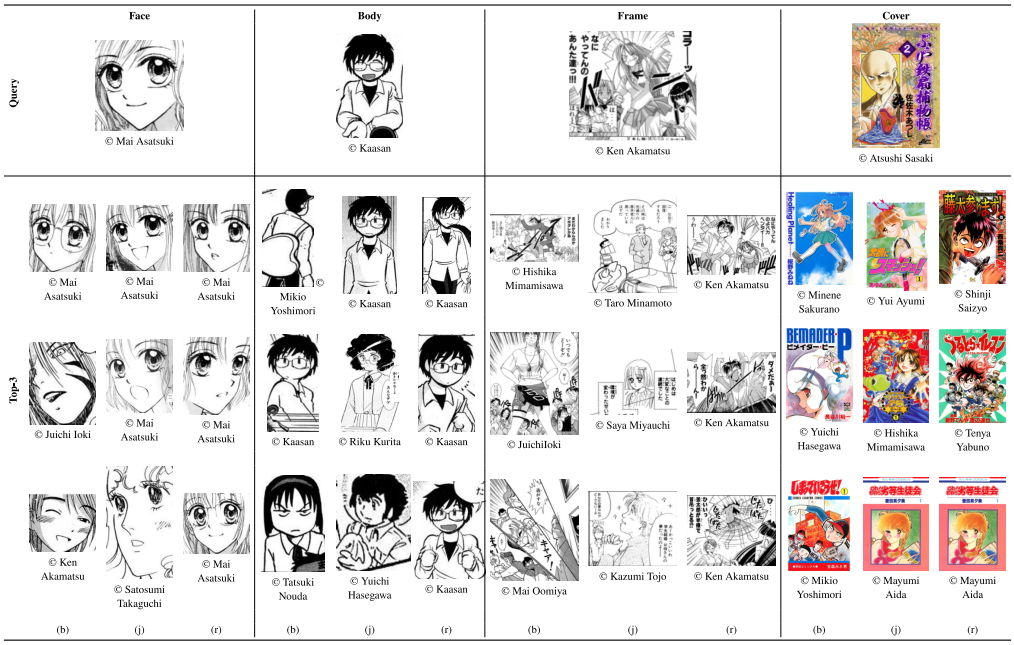
表 3 は，Danbooru データセットの Unseen サブセットに対して，データセット内の全クラス (i.e., 100 作者) をランダムに 10 個のグループで分割し，各グループごとのサンプルについてベースライン手法に基づいて学習されたモデル (b), (j) と，提案手法によって学習されたモデル (r) によって得られる画風特徴量を，t-SNE [39] によって 2 次元に圧縮し可視化した状態を表す．表 3 を確認すると，モデル (j), (b), (r) の順番で各クラスタごとのサンプルがより収束しており，定量評価結果に準じていることが確認できる．

4.4　Manga 109 を用いた評価

4.4.1　定 量 評 価

表 4 は，ベースライン手法に基づいて学習されたモデル (b) と (j)，**4.3.1** において最も高い性能を示したモデル (f) と (r) に対して，Manga 109 データセットを用いた評価結果を示す．この際，各モデルは **4.2** で述べた設定に基づき Danbooru データセットを用いて学習されており，別途の Fine-tuning は施されていないことに注意されたい．評価の結果，各サブセットの全ての評価指標において，提案手法に基づいて学習されたモデル (r) が最も高い性能を示した．まず，学習データとなる Danbooru データセットには稀にしか登場しない，人物の顔のみが線画で描かれた画像によって構成された Face サブセットについて，モデル (r) はモデル (b) と (j) に比べ NMI スコアにおいてそれぞれ 4.91% と 9.39% の性能向上を示している．また，キャラクターの全身が描かれているといった点で Face サブセットよりも Danbooru データセットに含まれる画像との共通点をもつ Body 及び Frame サブセットでは，モデル (b) と (j) の各精度が Face サブセットでの評価に比べ改善しているものの，モデル (f) や (r) を上回ることはなかった．最後に，彩色が行われている点や，コマやテキストといった漫画特有の表現が含まれない点において最も Danbooru データセットに類似した特性をもつ Cover サブセットでは，全てのモデルが各評価指標において他のサブセットよりも高い精度を示していることを確認できる．しかし，モデル (b) と (j) の ARI スコアは負の値を示しており，これは ARI による評価において正解データと各モデルの推定データ間の類似度が期待値よりも低いことを意味するため，各モデルのクラスタリング結果が顕著に不安定であったことを表すと考えられる．これらの結果から，提案手法

表 5　Manga 109 データセットの Nearest neighbor retrieval test



は **4.3.1** で述べた特徴により，各ベースライン手法に比べ，未知ドメインでも有効に活用できる頑健性をもつことが示唆された．

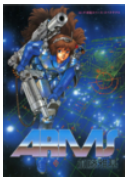
4.4.2　Nearest neighbor search

画風は線の描き方や色彩方法などの複数の要素が複雑に組み合わさっているため，言語化が難しい特徴をもつ．そこで，表 5 は，画風に基づく作品検索の際，特定の画像をもとに類似した画風をもつ画像を検索する場面において各モデルによって得られる結果を表したものである．具体的には，モデル (b), (j), (r) について Manga 109 データセットの各サブセットに含まれる各画像ごとに各モデルによって得られる画風特徴量間のユークリッド距離を求めることで最近傍探索 (nearest neighbor search) を行っており，クエリとなった画像の特徴量ともっとも距離の近かった上位三つのサンプルを提示している．表 5 を確認すると，Face サブセットについてモデル (b) の結果は，与えられたクエリ画像に対して人物の目の大きさや描き方が大きく異なるサンプルを含めている反面，モデル (r) は近似している画風によって人物の顔が描かれたサンプルを提示していることが確認できる．また，Frame サブセットについて，モデル (b) と (j) は人物の描写や吹き出しの配置，コマ内の構図などの各要素から考慮しても共通点が少ないサンプルが上位の結果になっている反面，モデル (r) は同じ作者の画像の中でもより多くの共通点をもつ画像が上位に現れていることを確認できる．更に，Cover サブセットにおいても，モデル (r) は鋭い目や彫り深い顔つきの人物描写といった共通点をもつ画像を上位に提示していることを確認できる．これらの結果から，**4.4.1** で示された性能の差に比例し，提案手法は既存手法により Manga 109 に含まれる未知データについてより正確にクラスタリングしていると考えられる．

4.4.3　t-SNE による可視化

図 3 は，Manga 109 データセットの Cover サブセットに含まれる各作品の表紙画像データについて，**4.3.2** と同様にモデル (r) による画風特徴量を t-SNE によって 2 次元に圧縮し可視化したものである．また，図 3 に含まれるグループ (a) から (d) は，可視化内容をより詳細に把握するために一定の特徴をもつ一部のサンプルやその周辺を拡大したものである．図 3 を確認すると，各表紙画像は画風の特徴に基づいてグルーピングされており，その中には各作品のジャンルの特徴にも紐づいている傾向が確認された．具体的には，グループ (a) と (b) は写実的な人物描写に合わせ細かい陰影によって彩色された共通点をもつが，人物の目の書き

表 6　Manga 109 データセットの Art style morphing

	Query A	Intermediate images				Query B
(1)	 © Masaki Kato	 © Yuzuru Shimazaki	 © Riku Kurita	 © Minene Sakurano	 © Ani Kuzuhara	 © Taro Minamoto
(2)	 © Kei Kazuna	 © Tatsuki Nouda	 © Hiroyuki Kanno	 © Hiroyuki Kanno	 © Shoei Ishida	 © Machiko Satonaka
(3)	 © Hotaru Unno	 © Machiko Satonaka	 © Shoei Ishida	 © Tadashi Sato	 © Tadashi Sato	 © Yoko Sanri

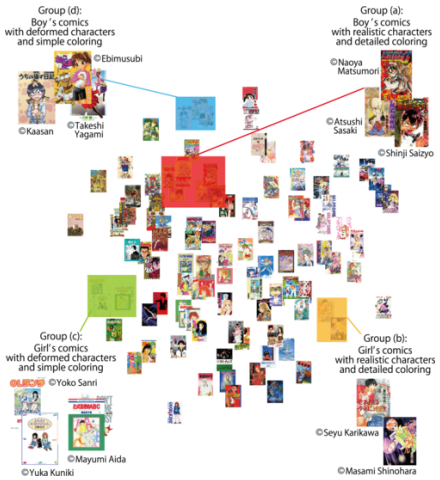


図 3　Manga 109 データセットの t-SNE による可視化

方や頭身の度合いといった細部において違いが存在する．これらの違いは，各グループの作品のジャンルである少年漫画や少女漫画における固有の特徴がそれぞれ反映されている．この傾向は，グループ (c) と (d) にも同様に確認できる．これらの結果から，電子書籍やウェブプラットホームに提供される表紙画像について，提案手法を活用することにより，画風に基づく作品検索が実現できることが期待される．

4.4.4　Art style morphing

Saito ら [40] は，イラスト画像に含まれる内容に基づいて検索を行う際，従来のキーワードベースの検索では対応できない点が多いことから，任意の二つの画像をクエリ A と B として与え，モデルによる意味特徴量に基づきクエリ A から B に遷移する際に参照される画像によって内容に基づく柔軟な検索を行う Semantic morphing を提案している．本論文では，Saito らに倣い，最近傍探索から拡張された画風に基づく作品検索手段として，Art style morphing を試みる．

具体的には，Manga 109 データセットの Cover サブセットについて，モデル (r) によって得られる画像特徴量について，各特徴量間のユークリッド距離を求め，互いにもっとも距離の近い 5 個のサンプルを一つのノードとした距離グラフを作成する．表 6 は，上記の距離グラフをもとに，与えられるクエリ A と B における最短パスを探索し，クエリ A から B へ遷移するパスに含まれる画像を提示したものである．

まず，表 6 のサンプル (1) では，色彩が豊かな表紙画像とモノクロの表紙画像をそれぞれクエリとして与えており，遷移パスに含まれる各画像の彩度が徐々に増減していることが確認できる．また，表 6 のサンプル (2) では，デフォルメ化された人物が描かれている表紙画像と写実的な描写で描かれた表紙画像について，遷移パス上で頭身の低い画像から頭身の高い画像に徐々に変化している様子が確認できる．これらの結果により，提案手法を用いた Art style morphing によって，異なる画風をもつ任意の表紙画像から，両方の画風における特徴が混合された特徴をもつ画像を検索したい場面 (e.g., サンプル (2) のクエリ A と B の間くらいの頭身をもつキャラクターが描かれた作品を探したい) に有効に活用できることが期待できる．

5. むすび

本論文では，画風に基づく作品検索に向け，距離学習と生成モデルを組み合わせた深層クラスタリングモデルを提案した．定量評価では，提案手法は未知データやドメインにおいて既存手法よりも強い頑健性を示した．また，提案手法によって獲得できる画風特徴量を t-SNE によって可視化した結果，人物の描写方法や彩色方法に関する共通点をもつクラスタが得られることが確認された．更に，最近傍探索や Art style morphing などの方法によって，任意の画像をクエリとした画風に基づく類似検索を行う場面においても，提案手法は既存手法よりもクエリに類似した特徴をもつ画像を提示することができた．今後は，同じ作者によって描かれているものの，作者の意図や技術によって異なった画風で描かれるような状況に対応できる方法として，半教師あり学習を用いたクラスタリングモデルなどについて検討していきたい．

謝辞　本研究は JSPS 科研費 22K12338 の助成を受けた．記して謝意を表す．

文　献

[1] B. Lang, The Concept of Style, Cornell University Press, 1987.
[2] W.-T. Chu and Y.-C. Chao, “Line-based drawing style description for manga classification,” ACM Int. Conf. Multimedia, pp.781–784, 2014. DOI:10.1145/2647868.2654962
[3] 安田幸生，佐藤雅明，村井　純，“漫画における著者の個性の定量化による漫画家推定，” 第 2 回コミック工学研究会，pp.46–50，2019.
[4] 福田光範，“漫画キャラクターの顔と畳み込みニューラルネットワークに基づく漫画の作者推定，” 第 6 回コミック工学研究会，pp.50–54，2021.
[5] Y.-M. Kim, “Feature visualization in comic artist classification using deep neural networks,” J. Big Data, vol.6, no.56, 2019. DOI: 10.1186/s40537-019-0222-3
[6] 増子達哉，松澤智史，“イラスト作家検索に向けた深層相関特徴の再考，” 第 36 回人工知能学会全国大会，2022. DOI: 10.11517/pjsai.JSAI2022.0-3P3GS202
[7] D.P. Kingma and M. Welling, “Auto-Encoding variational bayes,” Proc. Int. Conf. Learning Representations, 2014.
[8] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” Proc. Conf. Computer Vision and Pattern Recognition, pp.815–823, 2015. DOI: 10.1109/CVPR.2015.7298682
[9] A.F. McDaid, D. Greene, and N. Hurley, “Normalized Mutual Information to evaluate overlapping community finding algorithms,” arXiv preprint, 2011. DOI:10.48550/arXiv.1110.2515
[10] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-margin softmax loss for convolutional neural networks,” Proc. 33th Int. Conf. Machine Learning, vol.48, pp.507–516, 2016.
[11] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive angular margin loss for deep face recognition,” Proc. Conf. Computer Vision and Pattern Recognition, pp.4685–4694, 2019. DOI: 10.1109/CVPR.2019.00482
[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” Proc. Advances in Neural Information Processing Systems, vol.27, 2014.
[13] N. Dilokthanakul, P.A.M. Mediano, M. Garnelo, M.C.H. Lee, H. Salimbeni, K. Arulkumaran, and M. Shanahan, “Deep unsupervised clustering with gaussian mixture variational autoencoders,” arXiv preprint, 2017. DOI:10.48550/arXiv.1611.02648
[14] K.-L. Lim, X. Jiang, and C. Yi, “Deep clustering with variational autoencoder,” IEEE Signal Processing Lett., vol.27, pp.231–235, 2020. DOI:10.1109/LSP.2020.2965328
[15] S. Chang, “Deep clustering with fusion autoencoder,” arXiv preprint, 2022. DOI:10.48550/arXiv.2201.04727
[16] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[17] K. Zeng, M. Ning, Y. Wang, and Y. Guo, “Hierarchical clustering with hard-batch Triplet loss for person re-identification,” Proc. Conf. Computer Vision and Pattern Recognition, pp.13657–13665, 2020. DOI:10.1109/CVPR42600.2020.01367
[18] J. Yang, D. She, Y.-K. Lai, and M.-H. Yang, “Retrieving and classifying affective images via deep metric learning,” Proc. Conf. Artificial Intelligence, vol.32, no.1, 2018. DOI:10.1609/aaai.v32i1.11275
[19] A. Sain, A.K. Bhunia, Y. Yang, T. Xiang, and Y.-Z. Song, “StyleMeUp: Towards style-agnostic sketch-based image retrieval,” arXiv preprint, 2021. DOI:10.48550/arXiv.2103.15706
[20] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” Proc. European Conf. Computer Vision, vol.9911, pp.499–515, 2016.
[21] C.K. Sønderby, T. Raiko, L. Maaløe, S.r.K. Sønderby, and O. Winther, “Ladder variational autoencoders,” Proc. Advances in Neural Information Processing Systems, vol.29, pp.3738–3746, 2016.
[22] D.P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever,

and M. Welling, “Improved Variational Inference with Inverse Autoregressive Flow,” Advances in Neural Information Processing Systems, vol.29, pp.4743–4751, 2016.

[23] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, “Variational deep embedding: An unsupervised and generative approach to clustering,” Proc. Int. Joint Conf. Artificial Intelligence, pp.1965–1972, 2017. DOI:10.24963/IJCAI.2017/273

[24] S. Seo, D. Kim, Y. Ahn, and K.-H. Lee, “Active learning on pre-trained language model with task-independent Triplet loss,” Proc. Conf. Artificial Intelligence, vol.36, no.10, pp.11276–11284, 2022. DOI:10.1609/aaai.v36i10.21378

[25] V. Prasad, D. Das, and B. Bhowmick, “Variational clustering: Leveraging variational autoencoders for image clustering,” Proc. Int. Joint Conf. Neural Networks, pp.1–10, 2020. DOI:10.1109/IJCNN48605.2020.9207523

[26] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” Proc. Int. Conf. Machine Learning, vol.119, pp.1597–1607, 2020.

[27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, “ImageNet large scale visual recognition challenge,” Int. J. Computer Vision, vol.115, no.3, pp.211–252, 2015. DOI:10.1007/s11263-015-0816-y

[28] Anonymous, Danbooru community, and G. Branwen, “Danbooru2021: A large-scale crowdsourced and tagged Anime illustration dataset,” 2022. https://gwern.net/danbooru2021

[29] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, “Sketch-based Manga Retrieval using Manga109 Dataset,” Multimedia Tools and Applications, vol.76, no.20, pp.21811–21838, 2017. DOI:10.1007/s11042-016-4020-z

[30] K. Aizawa, A. Fujimoto, A. Otsubo, T. Ogawa, Y. Matsui, K. Tsubota, and H. Ikuta, “Building a Manga dataset “Manga109” with annotations for multimedia applications,” IEEE MultiMedia, vol.27, no.2, pp.8–18, 2020. DOI:10.1109/mmul.2020.2987895

[31] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. Conf. Computer Vision and Pattern Recognition, pp.770–778, 2016. DOI:10.1109/CVPR.2016.90

[32] S. Haykin, Neural networks: a comprehensive foundation, Prentice Hall PTR, 1994.

[33] Y. You, I. Gitman, and B. Ginsburg, “Large batch training of convolutional networks,” arXiv preprint, 2017. DOI:10.48550/arXiv.1708.03888

[34] D.P. Kingma and J. Ba, “Adam: A Method for stochastic optimization,” Proc. Int. Conf. Learning Representations, 2015.

[35] J. Xie, R. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” Proc. 33th Int. Conf. Machine Learning, vol.48, pp.478–487, 2016.

[36] L. Hubert and P. Arabie, “Comparing partitions,” J. Classification, vol.2, no.1, pp.193–218, 1985. DOI:10.1007/BF01908075

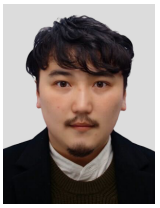
[37] H.W. Kuhn, “The Hungarian method for the assignment problem,” Naval Research Logistics, vol.2, no.1, pp.83–97, 1995. DOI:10.1007/978-3-540-68279-0_2

[38] G. Vincent, A. Yepremyan, J. Chen, and E. Goh, “Mixed-domain training improves multi-mission terrain segmentation,” Proc. European Conf. Computer Vision Workshop, pp.96–111, Cham, 2023. DOI:10.1007/978-3-031-25056-9_7

[39] L.v.d. Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Machine Learning Research, vol.9, no.86, pp.2579–2605, 2008.

[40] M. Saito and Y. Matusi, “Illustration2Vec: A semantic vector representation of illustrations,” Proc. SIGGRAPH Asia, vol.5, pp.1–4, 2015. DOI:10.1145/2820903.2820907

（2023 年 5 月 29 日受付，10 月 4 日再受付，
12 月 21 日早期公開）



朴　炳宣　（学生員）

2019 関西大学大学大学院総合情報学研究科知識情報学専攻博士課程前期課程了．同年関西大学院大学同学科博士課程後期課程に入学．現在，コミック工学と機械学習に関する研究に従事．情報処理学会，人工知能学会各会員．

松下　光範　（正員）

1995 大阪大学大学院基礎工学研究科物理系専攻制御工学分野博士前期課程了．同年日本電信電話株式会社入社．2008 関西大学総合情報学部准教授．2010 同教授．自然言語理解，インタラクションデザインに関する研究に従事．博士 (工学)．2003 情報処理学会論文賞，2013 年 LavalVirtual Award ほか各賞受賞．情報処理学会，人工知能学会，芸術科学会，ACM 各会員．