

文化財におけるデータサイエンス ——ディープラーニングによる龍門寺仏像の線図生成——

劉 学 平
李 博 月
李 為

目 次

1. 問題提起
2. 先行研究の概観
3. 研究設計および研究方法
4. 仏像の線形生成アルゴリズム
5. 社会的意義と今後の課題

1. 問題提起

中国の長い歴史の中に多くの文化財が残されているが、文化財の遺留は重要な考古価値を持っている。考古学における遺跡、造像、碑文、器物などの文化財は異なる形状と紋様の特徴を持っており、文字を使って正確で客観的に記述することは困難であり、画像を使って文化財を記述することは客観的にできるが、文化財の大きさと抽象的な特徴を記述することは困難である。文化財線図は文化財の形状、構造と紋様を直感的に表すことができるが、調査報告書の重要な構成部分であり、文化財の復元と学術研究を行う重要な材料でもある。他方、仏像線図の作成は石窟の考古作業に不可欠な構成要素であり、考古作業の全過程を跨っている。

文化財に対する保護力を強化することは、文化財の環境要因に対する統制を強化し、文化財の安定した状態を維持することである。しかし、地球温暖化による気候変動や地質作用、たとえば、龍門石窟、雲岡石窟、大足石刻などの石窟寺の洞窟環境の悪化、周辺岩石の崩壊、洞窟の漏出などのリスク¹⁾に直面しており、仏像も彫刻風化などの病害に直面しており、これらの悪化は不可逆的であり、環境要因を抑えることは悪化の過程をある程度遅らせるしかない。一方、文化財が保留している情報をできるだけ早く抽出し、調査報告書の形で残すことは文化財線図として調査報告書の重要な構成要素である。

文化財線図の作成には高い専門性があり、製図者に投影製図の知識だけでなく、製図者にも文化財に関する専門知識が求められる。そのため、現在、文化財線図を作成するには主に手作業であり、

1) 盧繼文, 李彬「雲岡石窟 21-30 窟及 5 窟附窟差異性風化因素初步研究」『石窟寺研究』2020 (10), 245-246.

高い人的・時間的コストがかかる。近年、ディープラーニングに代表される人工知能技術は急速に発展し、畳み込みニューラルネットワーク（CNN）と敵対的生成ネットワーク（GAN）はコンピューター視覚分野で広く応用されている。そのため、ディープラーニングを文化財線図作成に応用すれば、文化財線図を作成する時間を大幅に短縮し、作業の効率を高め、文化財線図の自動生成を実現するには客観的な可能性と必要性がある。本研究は文化財のデジタル化保護問題に対して、ディープラーニングを利用して文化財線図の生成を実現し、伝統的な応用分野の制限を打ち破ることに位置づけられる。

本研究はディープラーニングを中心に、エッジ抽出タスクの下ディープネットワークと画像翻訳タスクの下ディープネットワークに対して研究とテストを行うことにより、高品質と高精度を持つ仏像線図生成アルゴリズムを考案した。仏像線図データセットを作成して仏像線図生成アルゴリズムをトレーニングし、ハイパーパラメータチューニングを行うことで、生成品質と正確率において現在の既存アルゴリズムより優れたことを明らかにした。また、仏像線図データセットの数が少なく、入手が難しいなどの問題に対しては、まず、エッジ抽出アルゴリズムに自然画像を用いてトレーニングした後、文化財画像を入力して粗サンプル拡張データセットを作成し、粗サンプルを用いて改善されたネットワーク構造を事前にトレーニングし、仏像線図生成アルゴリズムが限られたデータセットの下でよりロバスト性と表現力を高めた。更に仏像画像を入力してクライアントから仏像線図を出力する機能を実現し、後に生成した線図を修正・修正できるようにするネットワーク・クライアントを開発したことで、修正した線図を真値としてデータセットの拡張を図ることができる。このように本研究は仏像線図生成システム全体の最適化と向上を達成し、この分野の空白をある程度埋めることを目指している。

2. 先行研究の概観

文化財の正投影画像から高品質な文化財線図を生成するタスクを実現するために、まず文化財線図の作成方法とコンピューターによる文化財線図生成の先行研究について概観しておこう。このタスクは、従来の画像処理方法で実現することも、近年急速に進展しているディープラーニングに関するアルゴリズムによって実現することもできる。これらの方法のうち、ディープラーニングに基づくエッジ検出アルゴリズムと画像翻訳アルゴリズムは文化財線図生成の重要な手段である。

（1）文化財線図の作成における問題点

文化財線図は絵画の対象によって分けられ、野外考古絵画と器物絵画に分けることができる。現在、中国では国の基準において、強制的な基準にしても推奨的な基準にしても文化財線図に対する規定がない。業界基準では、野外考古製図に関する基準²⁾があるが、文化財線図を定義することができる。

2) <http://hbba.sacinfo.org.cn/> (WW/T 0035-2012 田野考古製図 2012 [EB/OL]. [2021-08-16])

文化財線図は点、線、面の形と、線の太さ、虚实、疎密の処理によって文化財対象の成層、大きさ、構造などの情報を表現する。また、北京大学考古学部の教材本³⁾では文化財線図の作成についてより詳しく紹介している。文化財線図の伝統的な作成方法には、メッシュ法とベースライン法がある。伝統的な方法は作成周期が長く、文化財を運搬すると二次的な破壊が生じやすいため、現在、広く採用されている方法として、コンピューターで文化財の正投影画像を用いて線図を作成することである。作図の過程で作図者には文化財の専門知識が求められ、投影作図法を身につけてエッジ情報を利用して文化財を忠実に描く必要がある。

(2) 従来の画像処理方法に見る問題点

従来のエッジ検出オペレータは、強度、色、テクスチャなどの急激な変化を発見することで、画像中のエッジ情報を検出する。従来のエッジ検出オペレータは、フィルタによって画像の特徴を抽出する。フィルタ内のパラメータは固定であり、特定の規則で画像の特徴を抽出することができる。たとえば、Sobel は勾配情報を検出してエッジ検出を行い、水平方向検出オペレータと垂直方向検出オペレータに分けられる。しかし、Sobel 演算子は物体と背景をうまく区別できず、複雑なシーンで抽出された線が乱雑である。Scharr は Sobel オペレータの中心的な要素よりも重みが大きく、いくつかの検出精度を高めた。Laplacian 演算子は二次演算子であり、回転不変性があるため、各方向の異なるエッジに対して一つの演算子を使えば検出できるが、Laplacian 演算子はノイズに敏感であるため、平滑演算子⁴⁾と組み合わせて使う必要がある。Canny 演算子⁵⁾は 1986 年に提案され、エッジ検出に広く応用されている。Canny 演算子は 2 つの勾配の閾値を手動で設定し、閾値の中間のエッジ情報のみを残す必要があるが、乱雑な背景と物体をうまく区別できないという問題点がある。

Kang⁶⁾などの研究者たちはストリームに基づくガウス差分フィルタ (FDog) 方法を考案して原稿の抽出を実現したが、He⁷⁾などの研究者は階層ブロックを検出して原稿を抽出することを考案した。これらの従来の画像処理方法は特徴として比較的簡単な画像に対してエッジ情報を良好に抽出することができるが、比較的複雑な画像、特にノイズや損傷が多い文化財の画像に対しては、背景が乱雑で線が一貫していないという問題が残る。

(3) 畳み込みニューラルネットワークによるエッジ抽出アルゴリズムの問題点

畳み込みニューラルネットワーク (CNN) は畳み込み層、活性化層、プーリング層などの基本構造からなり、2012 年に AlexNet が ILSVRC 大規模画像認識チャレンジ優勝を獲得してから、畳み込み

3) 馬鴻藻『田野考古絵図』北京大学出版社, 2010:10-14.

4) 孫増国, 韓崇昭「基于 Laplacian 算子の圖像增強」『計算機応用研究』2007, 24 (001):222-223.

5) JOHN CANNY. A Computational Approach to Edge Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986 (6):679-687.

6) Kang H, Lee S, Chui C K. Coherent Line Drawing [C]. In Proceedings of the 5th International Symposium on Nonphotorealistic Animation and Rendering, 2007, 43-50.

7) He J N, Wang S, Zhang Y, et al. A Computational Fresco Sketch Generation Framework [C]. IEEE International Conference on Multimedia and Expo, 2013, 1-6.

ニューラルネットワークは大きく発展した。畳み込みニューラルネットワークは自動的にデータから特徴を学習し、結果を同種の未知データに汎化することができ、現在、画像分類、画像検索、意味分割など多くの分野で非常に高い効果を得ている。しかし現在、畳み込みニューラルネットワークを用いた文化財線図に関する研究は少なく、公開されている文化財線図データベースもトレーニングに利用できることは皆無に等しい。エッジ検出はコンピューター視覚領域での重要なタスクであり、多くの研究者がそれを研究しており、エッジ検出は文化財線図抽出タスクと強い相関があることがすでに明らかにされている。

2015年、Xie⁸⁾らは全体ネストエッジ検出ネットワーク（HED）がエンドツーエンドのトレーニングと予測を提案した。HEDはVGG16をバックボーンネットワークとして使用し、VGG16のそれぞれの畳み込み層の出力は1つの出力層に接続され、5つの測定出力層の画像を異なる重みで融合して出力する。HEDはマルチスケールとマルチレベルの思想をエッジ検出に取り入れ、高い効果が得られた。しかし、同じ真値を使用しても、各サイド出力レイヤにすべてのエッジ特性を反映させることには限界がある。

同様にマルチスケールの発想をVGG16バックボーンネットワークに融合し、WangらはDeep Crisp Boundaries（CED）⁹⁾を提案し、逆細分化パスを追加することで、異なる解像度の特徴マップを一つずつ加算し、逆畳み込みの代わりにサブピクセル畳み込みを用いてアップサンプリングを実現した。HEDとCEDは共にディープラーニングの手法を用いてエッジ検出を実現する先駆的な研究であり、本研究の文化財線図の生成にとって大いに示唆してくれた。

他方、HeらはHED中の異なるスケールの特徴図を一つの真値だけで監視する問題を最適化したと考え、エッジ検出を感知する双方向カスケードネットワーク（BDCN）¹⁰⁾を提案した。畳み込み層が深くなるにつれて、畳み込み核の感受野が大きくなることを考慮して、浅い層の情報は細部に敏感で、深い層の情報は輪郭に敏感であるため、各層の出力に対して適切な監視データを持つことが望ましい。しかし、真値データを異なるスケールで手作業になると、分解することは現実的ではなく、双方向カスケードネットワークは異なるスケールで近似された真値を自動的に生成することができる。双方向のカスケードネットワークを設計したのは、真値を分解する際の勾配消失問題を解決するためである。下記の通り、最終的な損失関数は、両側の損失と融合画像の損失の和として表現できる。

8) Xie S N, Tu Z W. Holistically-Nested Edge Detection [J]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015:1395-1403.

9) Wang Y P, Zhao X, Huang K Q. Deep Crisp Boundaries [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:3892-3900.

10) He J Z, Zhang S L, Yang M, Shan Y H, Huang T J. Bi-Directional Cascade Network for Perceptual Edge Detection [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019:3828-3837.

$$\mathcal{L} = w_{side} \cdot \mathcal{L}_{side} + w_{fuse} \cdot \mathcal{L}_{fuse}(P, Y)$$

$$\mathcal{L}_{side} = \sum_{s=1}^S \mathcal{L}(P_s^{s2d}, Y_s^{s2d}) + \mathcal{L}(P_s^{d2s}, Y_s^{d2s})$$

BDCN は多階層マルチスケール抽出エッジに新しいネットワーク構造を提供し、異なる階層の情報をよりよく保持することができるが、そのネットワーク構造が複雑で、ピクセルベースの損失のみを使用しているため、線が完全に揃っていない線図のトレーニングには難しい。

これらの畳み込みニューラルネットワークに基づく方法では、より複雑な線を学習することができるが、勾配伝播の過程でグローバル情報に注目すると、生成される線が太くなり、詳細情報が失われる。

(4) 敵対的生成ネットワークに基づく画像翻訳アルゴリズムの問題点

文化財の正投影画像から文化財線図を生成する過程は画像翻訳の仕事だとも考えることもできる。線図を生成する過程は単なるマルチスケールのエッジ抽出ではなく、エッジ情報に基づく画像翻訳であり、エッジ情報には存在しない繊細な特徴を体現する必要がある。近年、GAN ネットワークは画像翻訳タスクで大きな成果をあげている。

敵対的生成ネットワーク (GAN) は Ian J. Goodfellow¹¹⁾ らによって 2014 年に提唱され、一つの生成ネットワーク (Generator) と識別ネットワーク (Discriminator) から構成されている。生成ネットワークの目標はサンプルのデータ分布を学習して識別ネットワークを欺くことである。識別ネットワークは、サンプルが生成ネットワークによって生成された確率或いは真の確率であると判断する。GAN の価値関数は次のように表すことができる。

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$$

ここで、 x はサンプルデータからのサンプリング、 z はランダムノイズからのサンプリングであり、生成ネットワークの目標は価値関数 V の最小化と最大化である。理論的には、任意の識別ネットワークと生成ネットワークに対して、大域的な最適解を見つけることができる。GAN ネットワークは直接データ計算損失を用いて生成ネットワークを更新するのではなく、識別ネットワークの勾配を用いるため、非常に鋭い、あるいは退化した分布を表すことができる。これは、AI に画像作成・編集の能力を持たせ、以後の研究に継承されている。

Mehdi¹²⁾ らは GAN ネットワークに加えて、条件制約を追加し、監視トレーニングに利用できる GAN ベースの条件敵対的生成ネットワーク CGAN を提案した。生成ネットワークと識別ネットワー

11) I Goodfellow, J Pouget-Abadie, M Mirza, etc. Generative Adversarial Nets [J]. Conference and Workshop on Neural Information Processing Systems (NIPS), 2014.

12) Mirza, S Osindero. Conditional Generative Adversarial Nets 2014 [EB/OL]. [2021-08-16]. arXiv:1411.1784.

クにそれぞれ同じ条件入力 y を加えて生成ネットワークの出力を拘束する。生成ネットワークと識別ネットワークはすべて全連結層と活性化関数からなるため、生成画像の品質と解像度に大きな不足が生ずる。とはいえ、CGANの提案は、そのネットワーク構造を共通の戦略として他のGANネットワークに埋め込み、監視トレーニングを実現することができる。

2016年、ディープコンボリューション敵対的生成ネットワーク(DCGAN)¹³⁾はGAN(Generative Adversarial Networks)を元に改善が加えられたアルゴリズムで、より安定した枠組みを提案した。初期のGANネットワークはMINSTのような単純なデータセットでしか画像生成を行っていなかったが、DCGANは生成ネットワークと識別ネットワークの構造をそれぞれステップコンボリューション層とデコンボリューション層に変更し、全接続層の隠れ層を除去した。これにより、ネットワーク構造がより深くなり、画像生成タスクでの効果が大幅に向上した。Radfordらはまた、DCGANの無監視特徴学習能力について研究した結果、識別ネットワークは大型画像データセットで良い階層構造特徴を学習できることを考案した。

2017年、Phillip¹⁴⁾らはCGANの画像翻訳タスクにおける汎用的な解決策を研究し、ペアデータ監視下での入力マップから出力マップへのマッピングだけでなく、そのマッピングをトレーニングするための損失関数も学習できる画像翻訳のための条件生成式ネットワークPix2Pixを提案した。これにより、Pix2Pixは、手作業で損失関数を設計することなく、画像の彩色、線画と写真の変換など、さまざまな画像翻訳タスクに適用できる。生成ネットワークでは、U-net構造の自動エンコード-デコーダを使用して、生成ネットワークが入力と出力の間に大量の低レベル情報を共有できるようにする。識別ネットワークはpatchGAN識別ネットワークを用いて、入力画像を $N \times N$ の領域に分割し、識別ネットワークの出力はすべての領域推定構造の出力に変わった。Pix2Pixは、特に高度に構造化された画像生成において、GANネットワークが画像翻訳に大きな可能性を持っていることを明らかにした。

しかし、Pix2Pixの生成ネットワークを直接レイヤに追加してより大きな解像度の画像を生成させると、トレーニングプロセスが不安定になり、生成される画像の品質もよくなかった。Wang¹⁵⁾らはこれに基づいてPix2PixHDネットワーク構造を提案し、その構造はcoarse-to-fine生成ネットワークを含める。Coarse-to-fine生成ネットワークの低解像度生成ネットワークG1は大域的な連続性を学習し、高解像度生成ネットワークG2は局所的な細かい特徴を学習するため、生成される画像は局所的な特徴と大域的な特徴のリアリティを両立させる。識別ネットワークには、マルチスケールの識別ネットワークを用い、3つのスケールで判別を行う。また、Pix2PixHDネットワークが意味分

13) A Radford, L Metz, S Chintala. Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks [J]. International Conference on Learning Representations (ICLR), 2016.

14) P Isola, Zhu J Y, Zhou T H, Alexi A. Efros. Image-to-Image Translation with Conditional Adversarial Networks [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:1125-1134.

15) Wang T C, Liu M Y, Zhu J Y, etc. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018:8798-8807.

割図に基づく画像編集に用いられることによって、GAN の画像翻訳分野への応用を広げた。

画像翻訳タスクにはペアデータのトレーニングが必要になるが、ペアデータの取得にはコストがかかる。この問題について、Zhu らは非ペア画像翻訳のための循環一貫性対抗ネットワーク (CycleGAN)¹⁶⁾ を提案し、異なるドメイン間の画像マッピングを学習し、ペアデータセットがなくてもスタイル遷移、物体変形、季節変換、画像増強などのタスクを達成できるようにした。そのネットワーク構造は二つのドメイン X, Y の間に G と F の二つの生成ネットワークを通じて双方向のマッピング関係を確立し、二つの識別ネットワーク D_x, D_y はそれぞれ $\{x\}, \{f(y), g(x)\}$ を判別する。ネットワーク構造には二つの敵対ネットワークが含まれており、敵対損失は二つの敵対損失の和として表現できる。

$$\begin{aligned}\mathcal{L}_{GANloss} &= \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) = \\ &E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log (1 - D_Y(G(x)))] \\ &E_{x \sim p_{data}(x)}[\log D_X(x)] + E_{y \sim p_{data}(y)}[\log (1 - D_X(F(y)))]\end{aligned}$$

2つのドメインの誤ったクロスマッピングを避けるために、循環整合性損失も加えた。ある画像 x がある生成ネットワーク G を経て別のドメイン Y に変換された後、別の生成ネットワーク D を経て元のドメイン X に変換された後、2枚の画像は近似的になるはずである。損失関数は次のように表現される。

$$\mathcal{L}_{cycloss} = E_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1]$$

価値関数は次のとおりである。

$$G^*, F^* = \min_{G, F} \max_{D_X, D_Y} [\mathcal{L}_{GANloss} + \lambda \mathcal{L}_{cycloss}(G, F)]$$

一方、これらの研究は物体の変形に弱く、同じタスクでの生成効果は監視のいる Pix2Pix より劣っているが、CycleGAN の提案は監視のいない学習の応用を広げ、監視のいない画像翻訳に新たな解決策を提供した。TraVeLGAN¹⁷⁾ は、CycleGAN の循環的な一貫性の損失の代わりに、3番目のネットワーク学習の高度な特徴を設計することで、トレーニングの難易度を下げた。

16) Zhu J Y, T Park, P Isola, Alexi A.Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:2223-2232.

17) Amodio M, Krishnaswamy S. Tra Vel GAN: Image-to-Image Translation by Transformation Vector Learning [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019:8983-8992.

以上、GAN ネットワークには画像翻訳タスクに関する基本的な枠組みについて概観したが、画像生成の文化財線図タスクにとって大いに期待できる。

また、より細分化されたタスク領域に対して、Ran Yi¹⁸⁾ らは画像から芸術肖像画を生成するためのインターネット APDrawingGAN を提案した。肖像画の特徴のいくつかは仏像線図と似ている。①いずれも高度に抽象化されており、まばらではあるが連続した線の要素を含んでいる。②強い意味制約がある。③手作業によって描かれた真値の線は、元の図の特徴と正確に一致していない。④画像の中の低レベルの特徴に関係のない線が存在する。これらの特徴に対して、ネットワーク構造に階層的な GAN 構造を採用し、局所的な特徴と全体的な特徴を両立させた。生成ネットワークは1つのグローバル生成ネットワークと五感に対応する6つのローカル生成ネットワークで構成され、最小プール化され融合される。とりわけ、損失関数の設計では、真値の線と元の図中の特徴が正確に一致しないという問題を解決するために、線図を生成する線と真値の線とをピクセル単位で一致させる必要はなく、2つの線の間のピクセル単位の距離の和が最小になるようにするだけの距離変換損失を設計した。これに対して、APDrawingGAN は、階層的に特徴を抽出するアイデアを GAN ネットワークに取り入れ、肖像画に対して距離変換損失と局所損失を設計し、線図生成に重要な啓発的意義を持っている。

また、CoCosnet¹⁹⁾ や Pixel2Style2Pixel²⁰⁾ などのネットワーク構造は注意力メカニズムを取り入れ、より高品質でリアルな画像翻訳を模索した。BicycleGAN²¹⁾ ターゲット翻訳ドメインの隠れた空間コードを生成ネットワークに統合し、モデルの多様性の問題点を解決する。文化財線図の生成に関連する画像翻訳アルゴリズムを以下の図1にまとめる。

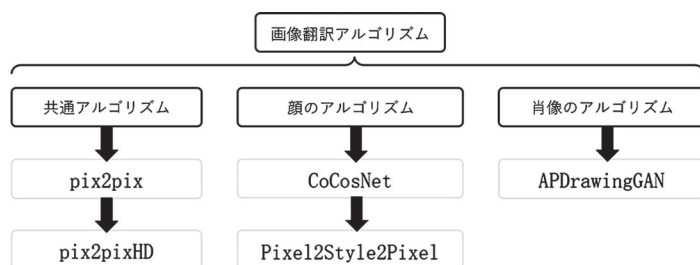


図1 文化財線図生成に関する画像翻訳アルゴリズム

18) Yi R, Liu Y j, Lai Y K, Pual L.Rosin. APDrawingGAN: Generating Artistic Portrait Drawings from Face Photos with Hierarchical GANs [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019:10743-10752.

19) Zhang P, Zhang B, Chen D, Yuan L. Cross-domain Correspondence Learning for Exemplar-based Image Translation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

20) E Richardson, Y Alaluf, O Patashnik, Y Nitzan. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation [C]. Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

21) Zhu J Y, Zhang R, Pathak D, et al. Toward Multimodal Image-to-Image Translation [C]. Advances in Neural Information Processing Systems, 2017, 465-476.

(5) これまでの研究に見る問題点

従来の画像処理方法では、画像勾配情報に基づいてエッジ抽出を行っていたが、文化財画像中の勾配がはっきりしない領域ではエッジ情報を捉えることができず、勾配変化が激しすぎる文化財の損傷領域ではノイズや非エッジ情報が多く抽出されることが多い。本研究は現在のいくつかの伝統的な方法について仏像の画像を用いて実験を行ったが、結果は理想的ではなかった（図2の参照）。

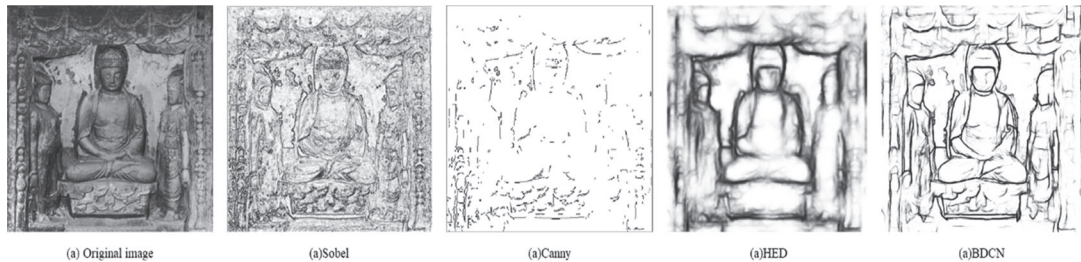


図2 従来の画像処理方法とエッジ抽出アルゴリズムの効果の比較

畳み込みニューラルネットワークに基づくエッジ抽出アルゴリズムは線の特徴を自動的に学習することができるが、ネットワークが深層になるにつれて元の図の情報が失われ、線が太くなるという問題が生ずる。本研究でも、既存のエッジ検出アルゴリズムで実験を行った結果、従来の画像処理方法に比べて改善を行ったが、まだ不十分である。敵対的生成ネットワークに基づく画像翻訳アルゴリズムは、画像の品質が高く、スタイルがリアルである。しかし、仏像線図生成専用の枠組みはまだ、ネットワークには線図生成に対する事前情報が欠落している。また、生成式はネットワークトレーニングに敵対し、生成された線図の形状とエッジに欠陥があり、新しいネットワーク構造をさらに改善し、設計する必要がある。さらに仏像線図データセットの入手が困難で、どのように少ないサンプルでトレーニングを行うかも考慮すべき問題である。

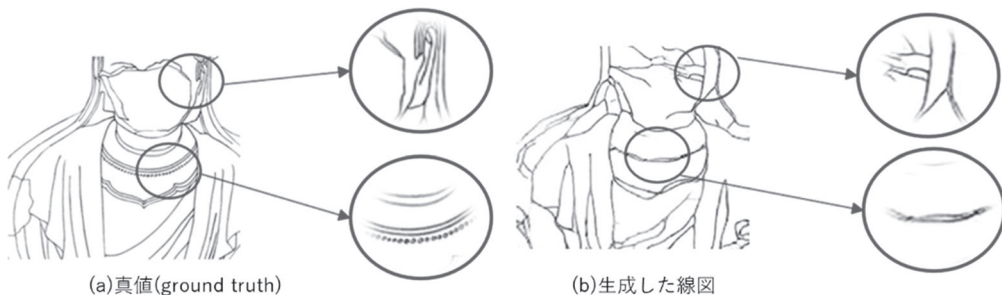


図3 画像翻訳アルゴリズムによって生成した線図

2. 研究設計および研究方法

以上の先行研究の問題点を踏まえて、本研究は既存のアルゴリズムについて比較実験を行い、仏像線図の生成品質に影響する要素について検討した。その中にはデータセットの裁断スケールが生成品質に与える影響、異なるネットワーク構造が生成品質に与える影響、損失関数の設計が生成品質に与える影響が含まれており、本研究は後述の仏像線図生成アルゴリズムの設計に理論的根拠と実験検証を提供することを可能にした。

仏像線図生成アルゴリズム設計において、既存のアルゴリズムは文化財線図生成タスクに特化して設計されたものではなく、生成品質に影響する要素を明確にした後、既存のアルゴリズムを再設計する必要がある。したがって、現在のアルゴリズムと特徴を比較分析し、線図を選択して代替アルゴリズムを生成し、実験検証により最終的に最も性能の良い改善後のネットワークを選択する。アルゴリズム設計も本研究のポイントの一つであり、アルゴリズムの性能の良さは後述のクライアントシステムの妥当性に直接関係する（図4の参照）。

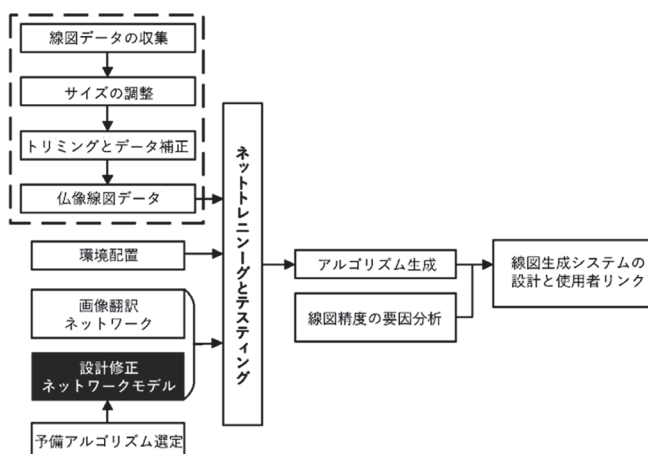


図4 全体的な技術ロードマップ

クライアントシステムの構築と最適化における線図データセットの数が少なく、入手が困難であるという問題に対して、アルゴリズムの実用化に焦点を当てながら、本稿で説明しているのは、ネットワーク・クライアント実装を開発してアルゴリズムのカプセル化を行い、使用者が遠隔地から仏像線図の生成を実現できるように、生成した線図の修正を支援し、修正した線図を真値としてデータセットを拡張できるようにした。

データセットの作成においては、データセットの品質と数はディープラーニングアルゴリズムの性能とトレーニングの安定性に直接関係するため、文化財線図の原図は解像度が大きいため、切り出してからネットワークに入力してトレーニングする必要がある。更に線図を生成する際には深層

ネットワーク出力をつなぎ合わせる必要があるため、スケール問題は非常に重要であり、大解像度の原図に基づくトレーニングデータ、検証セット、テストセットの区分を行い、マルチスケールを繰り返し切り出し方式でデータセットを作成する必要がある。

ネットワークトレーニングプログラムの構築では、まず真値とオリジナルイメージを使ってエッジ抽出ネットワークでトレーニングを行い、その後ネットワークを微調整して粗線図を得る。新しい線図よりマルチレベルを持つ仏像線図生成アルゴリズムを、生成された粗線図を用いて事前にトレーニングし、最終的に仏像線図データセットを用いて正式なトレーニングを行う。このトレーニング方法は、限られたデータセットでアルゴリズムのロバスト性と表現力を向上させるとともに、ネットワーク深化レーニングの不安定性の問題をある程度解決した（図5の参照）。

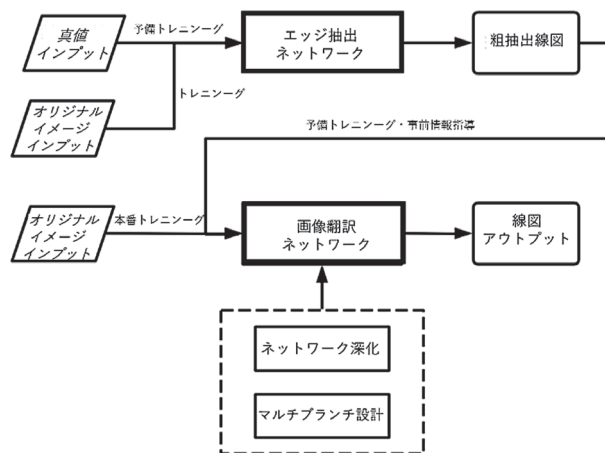


図5 ネットワークトレーニングのシナリオ

クライアントシステムの構築では、クライアントシステムには対話型インタフェース、線図生成アルゴリズムモジュール、フィードバックモジュール、線図データベースが含まれている。使用者はインタラクティブインタフェースにオリジナルイメージを入力し、アルゴリズムモジュールで計算した後、仏像線図をインタラクティブインタフェースに提示する。それに基づいてフィードバックモジュールで線図を修正し、修正した線図とそれに対応するオリジナルイメージをデータベースに保存し、データセットとしてアルゴリズムを再トレーニングする。クライアントシステムにより、アルゴリズムの自己最適化反復を実現することができる（図6の参照）。

(2) 単一スケールデータセットの作成

まず、解像度の大きい正投影画像とそれに対応する線図（オリジナルイメージ）の切り出し処理を行い、トレーニングに関係のないマーキングや文字情報を除去する。次に PS ソフトで原画を細かく調整し、線図中の線と正投影画像中の意味情報を揃える。そして、原図の解像度を調整して、図中の造像がほぼ同じスケールを持つようにする。次に、位置合わせイメージマップと線図を同期してトリミングし、トリミング後の両者の高さが揃っていることを確認してから、 256×256 のスケールにズームする。作成した単一スケールのデータセットは画像 443 枚で、トレーニングセット、テストセット、検証セットを 7:1.5:1.5 の比率で分割した（表 1 の参照）。

表 1 単一スケールデータセット数

データセット	画像数（枚）
トレーニング数	331
テスト数	56
検証数	56
合計	443

(3) マルチスケールデータセットの作成

仏像の正投影画像とその対応線図を解析することにより、大きな目標境界とスケールの小さい細部から構成される特徴を発見した。造像と紋織の意味情報の幅が普遍的に小さい、内容が豊富で、細部が複雑である。したがって、3つのスケールを設計して長方形の領域で原画をトリミングすることで、より多くの意味情報を保持することができる（表 2 の参照）。具体的な処理手順は次のとおりである。

- ① オリジナルイメージをほぼ同じスケールに調整する。
- ② トレーニングセット、テストセット、検証セットの区分を原図全体の単位で行う。
- ③ 3つのスケールをデザインして原図をカットし最大スケールの辺の長さが原図の最短辺を覆う。
- ④ オリジナルイメージを x, y 両方向にカットする。
- ⑤ 繰り返しトリミングの倍数を設定する。
- ⑥ 回転、反転などの方法を用いてデータの拡張を繰り返す。

表 2 単一スケールデータセット数

データセット	画像数（枚）
トレーニング数	1356
テスト数	240
検証数	240
合計	2016

続いて、従来の画像翻訳アルゴリズムのネットワーク構造（図8の参照）の改善を行った。

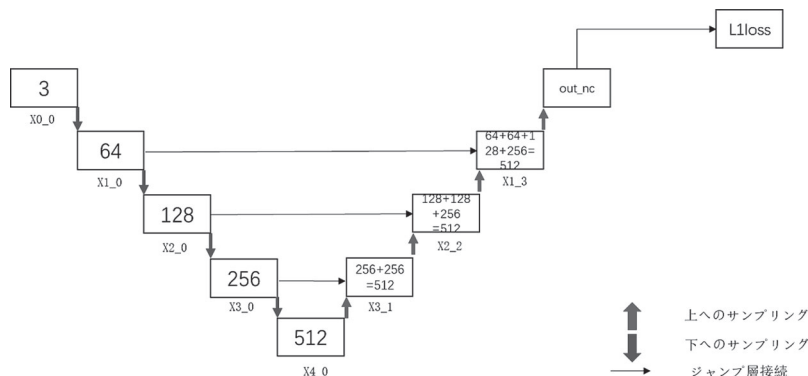


図8 U-net 構造

pix2pix の元のネットワーク構造の生成ネットワークはジャンプ層構造を持つ U-net を採用して、その中の浅い層構造は浅い層構造が画像の簡単な特徴，すなわち境界，色をつかむことができる。深層構造は大きな感覚野で抽象的な情報を捉えることができるため，深層，浅層の情報はジャンプ層構造によってチャンネル次元のつなぎ合わせを行う。仏像の正投影画像の低レベルの情報は非常に豊富で，密結合と深層監視を追加して U-net 構造を改良してから，4 種類の異なる構造を設計して比較した。

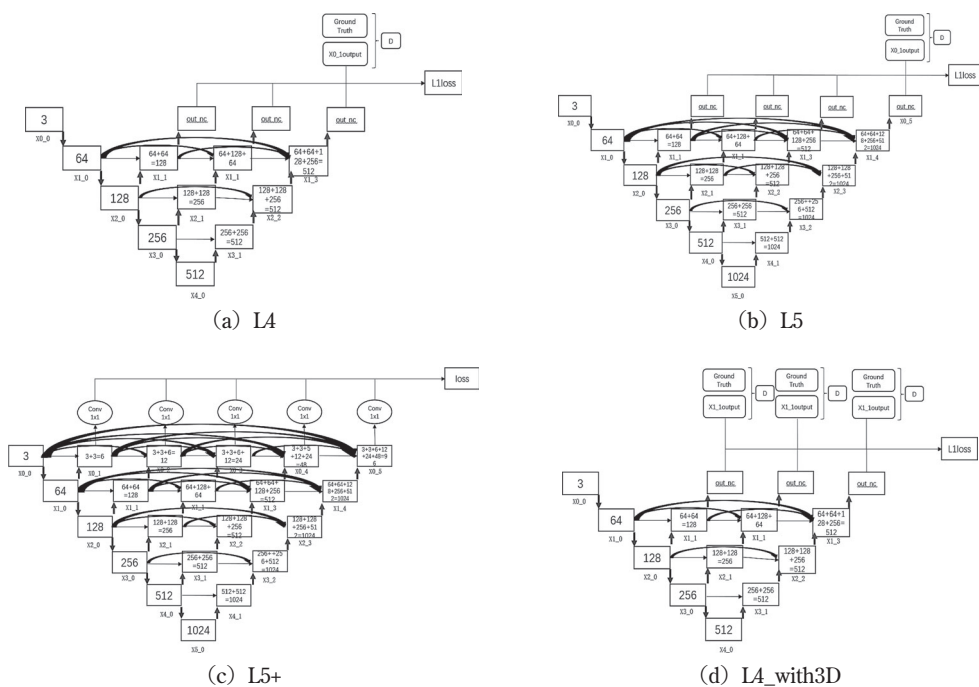


図9 四つのネットワーク構造の改善

そのうち、L4 ネットワークは U-net に稠密接続を追加し、中間層に深層監視を加えた。L5 は L4 と同じ構造で、層数が一段深くなる。L5+ は L5 に基づいて出力層に 1x1 の畳み込み層を追加し、L4_with3D は L4 に基づいて中間出力層ごとに識別ネットワークを追加して考案した（図 9 の参照）。その結果、改善された U-net 構造トレーニング曲線を比較すれば一目瞭然である（図 10 の参照）。それによってネットワーク出力結果（図 11 の参照）を確認すると、直感的に L4 の生成した仏像線図が非常に綺麗になっていることが分かる。

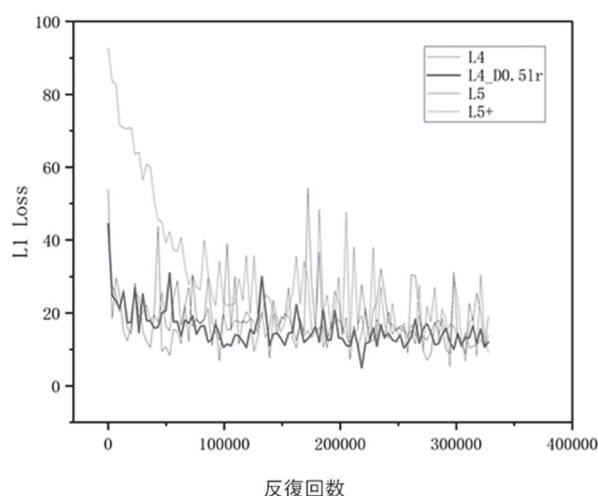


図 10 改善された U-net 構造トレーニング曲線の比較

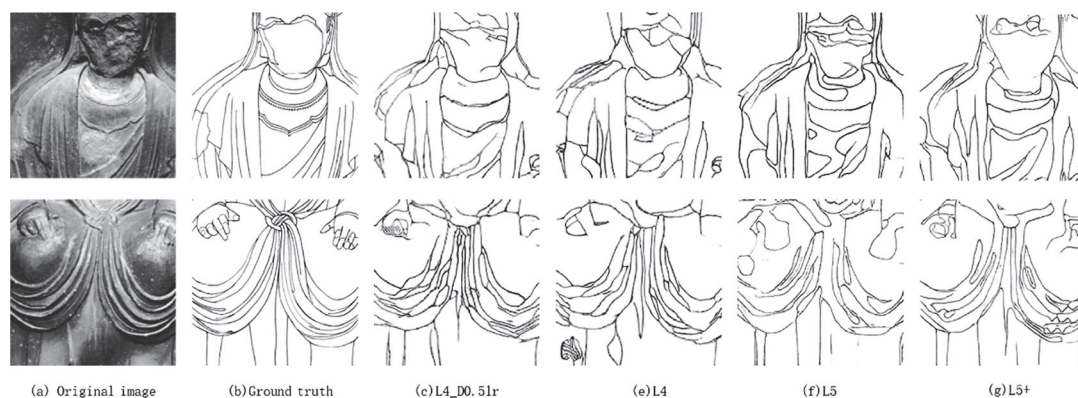


図 11 ネットワーク出力結果の比較

予想したようにトレーニング結果から見ても、L4 ネットワークの収束速度が最も速く、生成効果が最も高い。L5 効果が低い原因は、層数が深くなるにつれて、密連結とホップ連結が過剰な浅い層の情報を出力層に重ねてしまうため、トレーニングが深い層になりにくくなり、ネットワークが浅

い層の情報をより多く学習してしまうからである。L4_with3D ネットワークはトレーニング中に収束できない。なぜなら、現在の生成ネットワークの性能はまだ向上させる必要があり、3つの識別ネットワークの判別能力は生成ネットワークの生成能力よりはるかに大きいからである。L4 ネットワークにおける識別ネットワークの学習率を下げ調整すると、トレーニングプロセスがより安定していることが分かる（表3の参照）。

表3 異なる構造改善のためのトレーニングハイパーパラメータ

ネットワーク	step1	step2	max iter	optimizer	G_lr	D_lr	batch size
L4	600	400	331000	Adam	0.0002	0.0002	1
L4_D0.5lr	600	400	331000	Adam	0.0002	0.0001	1
L5	600	400	331000	Adam	0.0002	0.0002	1
L5+	600	400	331000	Adam	0.0002	0.0002	1
L4_with3D	600	400	331000	Adam	0.0002	0.0002	1

(4) マルチスケールデータセットによる pix2pixHD のトレーニング

現在の設備条件に適応するためには、pix2pixHD ネットワークを微調整する必要がある。出力解像度を 1024×2048 から 512×1024 に調整し、マルチスケールデータセットでトレーニングする²²⁾。トレーニングのハイパーパラメータとトレーニング曲線を表4と図12に示している。これによって実験した結果のアウトプットは図13と図14で示しているように、完璧とは言えないが、現段階では比較的良好な効果が得られた。

なお、今回の実験に使用したハードシステムとプログラミングには、CPU (Inter i7-11700K), GPU (NVIDIA RTX3080Ti), OS (WIN10), コンパイラ (Pycharm Community), プログラミング言語 (Python 3.8.5), ディープラーニングフレームワーク (Pytorch1.4 torchvision 0.9.1) が含まれている。中にはPytorchはオープンソースのpython言語を使ったディープラーニングライブラリで、Facebookによって開発されたものである。

22) 劉学平・李博月「敵対的生成ネットワークに基づく考古線図の自動生成方法」は2021年に特許として中華人民共和国特許局に特許出願し受理された (CN20211082115.7)。

表4 トレーニングのハイパーパラメータ

ハイパーパラメータ名称	ハイパーパラメータ数
トレーニング数	1536
テスト数	240
トレーニング時間	約 68h
Epoch	200
Batchsize	1
学習率(lr)	0.0002(100epoch 後減衰)
オプティマイザ	Adam
α	0.5
β	0.999

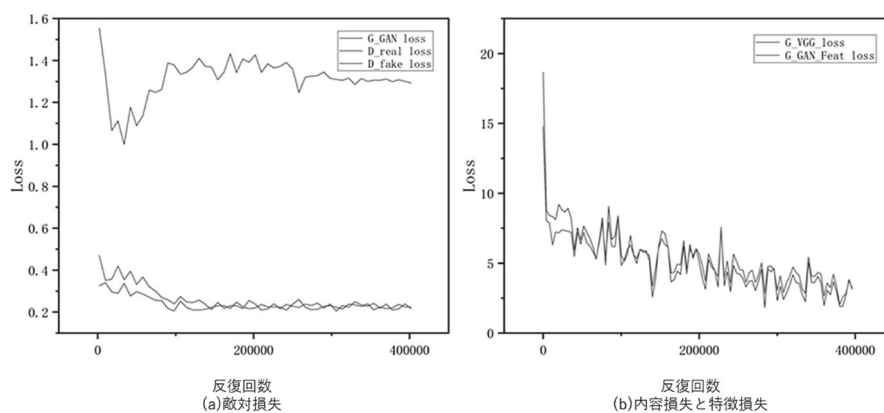


図12 pix2pixHD におけるトレーニング曲線

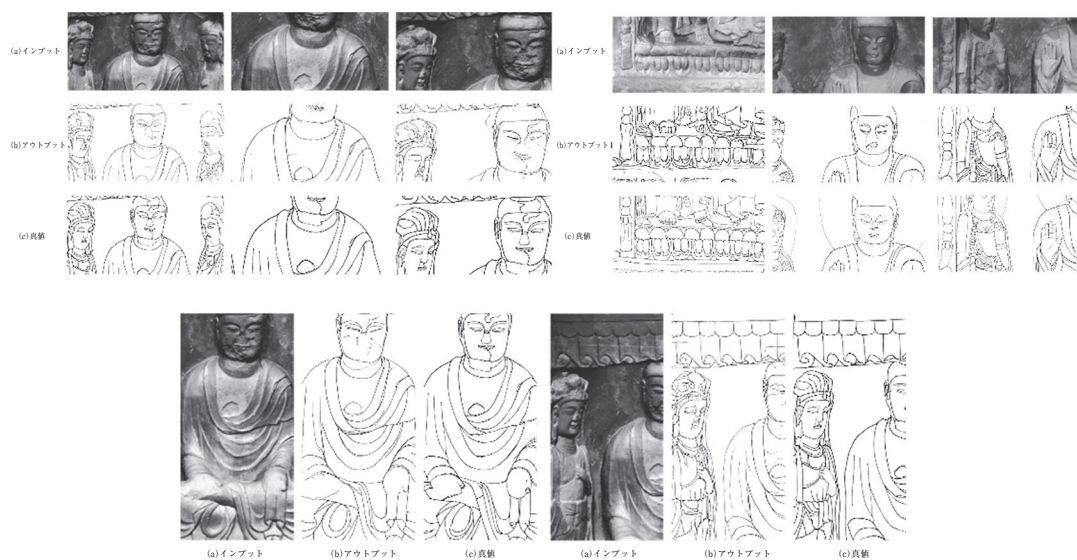


図13 部分の実験結果

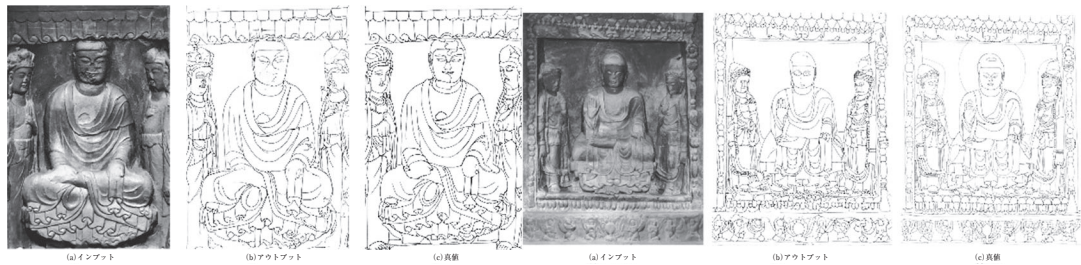


図14 全体の実験結果

5. 社会的意義と今後の課題

近代科学は仮説に基づくモデリングであり、思考は前にあり、回答は後にある。つまり、データがないところは予測できるが、要素が見落とされがちなのが特徴である。それに対して、データサイエンスは機械学習に基づくモデリングであり、答えは前にあり、思考は後になる。つまり、データのないところは予測できないが、要素を漏れなく考えやすいのが特徴である。したがって、本研究はデータサイエンスの範疇に属する。この前提から文化財を対象に、本研究は仏像の線図生成アルゴリズムを設計し、ディープラーニングと文化財線図作成のアイデアを組み合わせ、仏像線図の高品質かつ高精度な自動生成を実現し、文化財線図分野におけるディープラーニングの空白を埋めた。人手で作成した人的・物的・時間的コストを節約し、手作業で線図を作成する過程で文化財を運搬する二次的破壊を回避し、文化財保護のデジタル化の水準を高め、文化財保護への貢献に資する本研究の初歩的な試みを明らかにした。

他方、現在利用可能な公開されている仏像線図データセットはまだ存在していないため、本研究では仏像線図データセットを作成し、ディープラーニングアルゴリズムの異なるスケールでのデータセットの表現能力を模索してきた。仏像画像に含まれている意味情報は非常に豊富で、マルチスケールでトリミングして作成したデータセットは深層ネットワークの仏像線図の作成をよりよくトレーニングすることを期待できる。とりわけ、仏像線図データセットの数が少なく、入手が困難な問題に対して、まず真値とオリジナルイメージを用いてエッジ抽出ネットワークにおいてトレーニングを行い、後にネットワークを微調整して一次的な真値を得るディープネットワークトレーニングという方法を考案した。マルチスケールを持つ文化財線図生成アルゴリズムを、粗線図の真値を用いて事前にトレーニングし、最終的に仏像線図データセットを用いて本番トレーニングを行う。このトレーニング方法は、限られたデータセットでアルゴリズムのロバスト性と表現力を向上させるとともに、深層ネットワークトレーニングの不安定性の問題点もある程度解決する。今後の課題として安定性をさらに高める必要がある。

仏像線図生成ネットワーク・クライアントを開発し、使用者がリモートで仏像線図生成を実現で

きるようにする。また、生成された線図の修正を支援し、修正された線図を真値としてデータセットを拡張できるようにする。ネットワーク・クライアントは仏像線図生成アルゴリズムの工学的応用を促進する一方で、より多くの仏像線図データを収集することができ、さらなる研究の基礎を築き、本研究の目的である文化財保護に応用できるように期待できる。

参考文献

1. JOHN CANNY. A Computational Approach to Edge Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986 (6):679-687.
2. 龍門文物保管所, 北京大学考古系編『龍門石窟』平凡社, 1987.12-1988.8.
3. 曾布川寛編『龍門石窟石刻集成』京都大學人文科學研究所附屬東洋學文獻センター, 2000.
4. 佐藤安之助文庫, 拓殖大学図書館編『「龍門石窟造像記」拓本分類目録』拓殖大学図書館, 2000.
5. Kang H, Lee S, Chui C K. Coherent Line Drawing [C]. In Proceedings of the 5th International Symposium on Nonphotorealistic Animation and Rendering, 2007, 43-50.
6. 孙增国, 韩崇昭「基于 Laplacian 算子的图像增强」『计算机应用研究』2007, 24 (001):222-223.
7. 馬鴻藻『考古器物绘图』北京大学出版社, 2008:1-4.
8. 馬鴻藻『田野考古绘图』北京大学出版社, 2010:10-14.
9. He J N, Wang S, Zhang Y, et al. A Computational Fresco Sketch Generation Framework [C]. IEEE International Conference on Multimedia and Expo, 2013, 1-6.
10. I Goodfellow, J Pouget-Abadie, M Mirza, etc. Generative Adversarial Nets [J]. Conference and Workshop on Neural Information Processing Systems (NIPS), 2014.
11. Mirza, S Osindero. Conditional Generative Adversarial Nets 2014 [EB/OL]. [2021-08-16]. arXiv:1411.1784.
12. Xie S N, Tu Z W. Holistically-Nested Edge Detection [J]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015:1395-1403.
13. A Radford, L Metz, S Chintala. Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks [J]. International Conference on Learning Representations (ICLR), 2016.
14. P Isola, Zhu J Y, Zhou T H, Alexi A. Efros. Image-to-Image Translation with Conditional Adversarial Networks [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:1125-1134.
15. Zhu J Y, T Park, P Isola, Alexi A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:2223-2232.
16. Zhu J Y, Zhang R, Pathak D, et al. Toward Multimodal Image-to-Image Translation [C]. Advances in Neural Information Processing Systems, 2017, 465-476.
17. Wang Y P, Zhao X, Huang K Q. Deep Crisp Boundaries [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:3892-3900.

18. Wang T C, Liu M Y, Zhu J Y, etc. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs[J].
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018:8798-8807.
19. He J Z, Zhang S L, Yang M, Shan Y H, Huang T J. Bi-Directional Cascade Network for Perceptual Edge Detection [J].
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019:3828-3837.
20. Amodio M, Krishnaswamy S. TraVelGAN: Image-to-Image Translation by Transformation Vector Learning [C].
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019:8983-8992.
21. Yi R, Liu Y j, Lai Y K, Pual L. Rosin. APDrawingGAN: Generating Artistic Portrait Drawings from Face Photos with Hierarchical GANs [J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019:10743-10752.
22. Zhang P, Zhang B, Chen D, Yuan L. Cross-domain Correspondence Learning for Exemplar-based Image Translation [C].
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
23. 盧繼文, 李彬「雲岡石窟 21-30 窟及 5 窟附窟差異性風化因素初步研究」『石窟寺研究』2020 (10).245-246.
24. <http://hbba.sacinfo.org.cn/> 〈行业标准信息服务平台. WW/T 0035-2012 田野考古制图 2012 [EB/OL]. [2021-08-16]〉
25. E Richardson, Y Alaluf, O Patashnik, Y Nitzan. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation [C]. Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

Data Science in Cultural Relics: Generation of line diagram of Longmen Temple Buddha statue based on deep learning

Xueping LIU

Boyue LEE

Wei LEE

ABSTRACT

Modern science is based on hypothesis modeling, with thinking ahead and answers behind. It is characterized by the fact that there is no data to predict, but the elements are often ignored. On the contrary, data science is a modeling based on machine learning, with the answer in front and thinking behind. It is characterized by unpredictability where there is no data, but it is easy to consider the elements without omission. Therefore, this research belongs to the category of data science. On this premise, taking cultural relics as the object, this study designed the generation algorithm of cultural relics line graph, combined with the creativity of deep learning and cultural relics line graph making, realized the high-quality and high-precision automatic generation of cultural relics line graph, and filled the gap of deep learning in the field of cultural relics line graph. According to the preliminary attempt of this study, that is, to save the manpower, material resources and time cost of manual drawing, avoid the secondary damage of carrying cultural property in the process of manual drawing, improve the digitization level of cultural relics protection, and contribute to cultural relics protection.

