

Title	イラスト推薦と可変グリッドを用いた模写練習支援システムの提案
Author(s)	金山, 春香
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18288
Rights	
Description	Supervisor: 宮田 一乗, 先端科学技術研究科, 修士(知識科学)

修士論文

イラスト推薦と可変グリッドを用いた模写練習支援システムの提案

金山 春香

主指導教員 宮田 一乗

北陸先端科学技術大学院大学
先端科学技術研究科
(知識科学)

令和5年3月

Abstract

In recent years, the proliferation of digital devices has led to a growing number of self-taught digital illustrators. However, there are limited resources available to support beginners in their digital illustration practice. Observation is a crucial aspect of reproduction, which is often used in illustration practice. While many existing copying support systems provide guidance by displaying the results of observation, few support the observation process itself.

To address this issue, we propose a framework for reproduction practice support that incorporates a variable grid system for copying and the recommendation of reference illustrations based on rough sketches. Our system takes a rough sketch drawn by the user as input, and presents similar illustrations with similar feature values as candidates from the dataset. The user can then select their preferred illustration and reproduce it using a reproduction interface that allows them to adjust the fineness of the grid. The results of our evaluation experiments and a questionnaire survey confirm the effectiveness of the system in supporting the practice of reproduction.

目次

第1章	はじめに	1
1.1	研究背景	1
1.2	研究目的	5
1.3	論文構成	6
第2章	関連研究	7
2.1	イラスト生成・制作支援に用いられる深層学習	7
2.1.1	GAN	7
2.1.2	拡散モデル	8
2.2	イラスト制作支援	9
2.3	模写支援	11
2.4	本研究の位置づけ	12
第3章	提案手法	13
3.1	概要	13
3.2	データセット	14
3.3	画像推薦システム	18
3.3.1	特徴量を用いた類似画像検索	18
3.3.2	特徴量の比較(予備実験)	20
3.4	ラフスケッチの洗練化	21
3.5	インターフェース	24
第4章	ユーザスタディ	26
4.1	評価実験	26
4.2	System Usability Scale	30
4.3	NASA-TLX	31
4.4	実験環境	32
第5章	結果	33
5.1	事前アンケート調査の結果	34
5.2	アンケート調査の結果	34
5.2.1	模写後のアンケート結果(支援あり)	43
5.3	アンケート結果のまとめ	45

5.4	System Usability Scale	46
5.5	NASA-TLX	48
第6章	おわりに	49
6.1	まとめ	49
6.2	今後の課題	49

目次

1.1	ペイントソフト例 (CLIP STUDIO PAINT)	2
1.2	イラスト制作過程	2
1.3	模写とトレースの違い	3
1.4	自画像を描くときに用いられるグリッド付き鏡	4
1.5	デッサンとクロッキーの違い	5
2.1	GAN の概要図 [1]	7
2.2	DDPM の概要図 [2]	9
2.3	dualFace の概要図 [3]	9
2.4	dualFace のインターフェース [3]	10
2.5	shadowDraw の概要図 [3]	10
2.6	自動レイヤ分けシステムの流れ図 [4]	11
2.7	制作過程を編集した例 [5]	11
2.8	The drawing assistant の概要図 [6]	12
3.1	システム概要図	14
3.2	顔範囲抽出の例 [7]	15
3.3	線画抽出・2値化例 [8][9]	16
3.4	データセット例 (上: 顔イラスト, 下: 対応する線画画像)	17
3.5	SIFT 特徴量の計算方法 [10]	19
3.6	HOG 特徴量の計算方法 [10]	20
3.7	予備実験の結果例	21
3.8	DeepPS の概要図 [11]	22
3.9	洗練度 l による画像の洗練 [11]	22
3.10	スケッチ洗練例 ($l = 1$)	23
3.11	提案インターフェース	24
3.12	グリッド追加例	25
4.1	実験の流れ	29
5.1	実験の様子	33
5.2	制作例	34
5.3	制作時間の比較	35

5.4	アンケート結果 (設問1～設問3)	36
5.5	アンケート結果 (設問10,12,14)	43
5.6	SUS結果	47
5.7	NASA-TLX結果 (MD: 知的・知覚的要求, PD: 身体的要求, TD: タイムプレッシャー, OP: 作業成績, EF: 努力, FR: フラストレー ション, WWL: 全体的な評価)	48

表 目 次

4.1	事前アンケート内容	26
4.2	模写後のアンケート内容	27
4.3	模写後のアンケート内容(支援あり)	28
4.4	SUS アンケート内容	30
4.5	実験環境	32
5.1	設問 4: 気をつけたこと(支援なし)	37
5.2	設問 5: 難しかったこと(支援なし)	37
5.3	設問 7: どこに違和感があるか(支援なし)	38
5.4	設問 8: 修正するとしたらどこをどのようにするか(支援なし)	38
5.5	設問 9: 感想(支援なし)	39
5.6	設問 4: 気をつけたこと(支援あり)	40
5.7	設問 5: 難しかったこと(支援あり)	40
5.8	設問 7: どこに違和感があるか(支援あり)	41
5.9	設問 8: 修正するとしたらどこをどのようにするか(支援あり)	41
5.10	設問 9: 感想(支援あり)	42
5.11	設問 11: 設問 10 の理由	44
5.12	設問 13: 設問 12 の理由	44
5.13	設問 15: 設問 14 の理由	45

第1章 はじめに

本章では、本研究の背景及び目的について述べ、最後に本論文の構成を記述する。

1.1 研究背景

近年，SNS の発展により，イラスト制作についての情報が得やすくなっている．特に，pixiv¹を初めとするイラスト SNS や，大手 SNS である twitter²や YouTube³においてイラスト投稿や練習法の共有が盛んに行われている．また，Apple 社の iPad や Microsoft 社の Surface を初めとするタブレット型 PC や，Wacom 社の Wacom Cintiq や Wacom Intuos を初めとするペンタブレットなど，ペン型デバイスで操作することのできる製品の普及が進んでいる．さらに，ペイントソフトの普及が進んでいる．ペイントソフトとは，コンピュータ上でイラストを描くソフトウェアである．例を図 1.1 に示す．現実における紙やキャンバスの上での描画とは違い，コンピュータ上の仮想キャンバスを使うため，自由に修正が可能である．そして，絵具の種類に制限がなく，ソフトの種類によってはフィルタや素材 (パターン，3D モデル等) が利用できることが特徴である．以前から，Adobe 社の Photoshop やセルシス社の CLIP STUDIO PAINT といったペイントソフトが販売されていたが，近年では購入方法に月額サブスクリプションサービスが追加され，一定期間ではあるがより安価に利用できるようになっている．また，MediBang 社の MediBang Paint のような無料かつ高性能なペイントソフトもリリースされている．このように，デジタル環境下でのイラスト制作に必要な機材が身近になっており，デジタルイラスト制作を行う人が増えている．ペンタブレットメーカーである wacom 社が 2020 年に行った調査 [12] では，イラスト制作において，ユーザの 85 % がデジタル画材 (PC，タブレット，イラスト制作ソフト等) を取り入れており，66 % が制作の半分以上においてデジタル画材を使用していることがわかった．

イラストを描く工程は，一般的には以下のような手順を踏むことが多い．例を図 1.2 に示す．

¹<https://www.pixiv.net/>

²<https://twitter.com/>

³<https://www.youtube.com/>

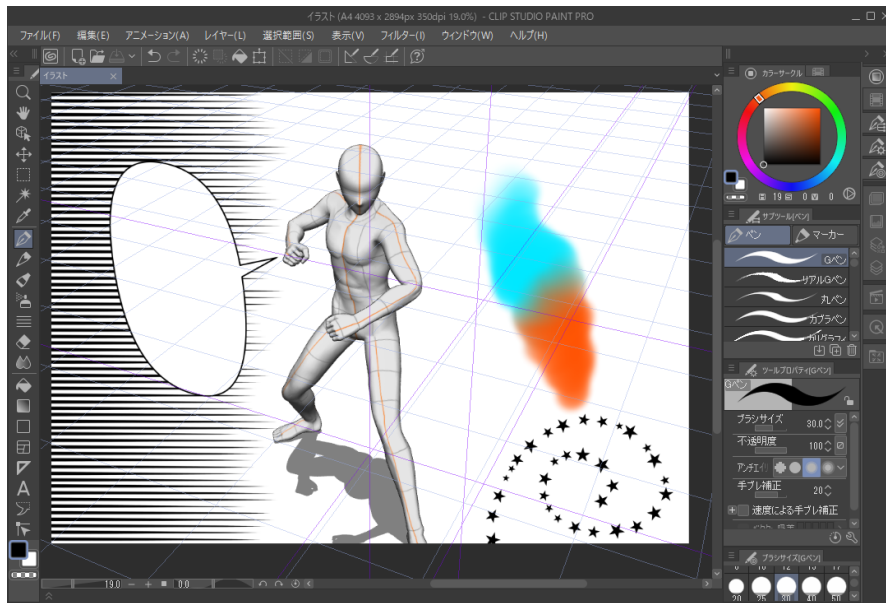


図 1.1: ペイントソフト例 (CLIP STUDIO PAINT)



図 1.2: イラスト制作過程

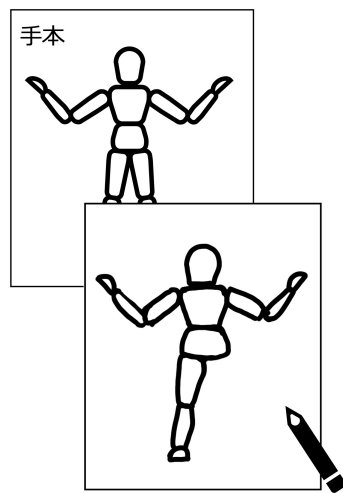
1. 構図を考え、アタリ (大まかな位置決め作業) を描く
2. アタリに沿ってラフを描く
3. ラフを線画にする (清書する)
4. 配色を考える・彩色
5. 仕上げ

この過程の中で、特に自分の思い通りに描くための実践的な技術が必要となるのは3以降である。構図や対象物についての情報、色彩などの知識も必要であるが、よりクオリティの高いイラストを制作するためには、実際に描く練習が不可欠である。また、デジタル環境においては、ペン型デバイスを用いるため、紙に鉛筆やペンを用いて描くことと感覚が異なり、その環境に慣れる必要がある。

イラストを上達するための練習として、主に3つの練習法がある。

1つ目は模写である。プロのイラストレータの作品を観察し再現することで、作風や意図を理解することが目的である。単なるトレースでは観察が伴わないので練習としては適さない。模写とトレースの違いを図1.3に示す。

模写 …観察して描く



トレース …手本をなぞって描く



図 1.3: 模写とトレースの違い

グリッドとよばれる画面を水平線と垂直線で分断した格子状の線を手本と自らのキャンバスに追加することで補助を行うこともある。このグリッドは、自画像を描くときなどにも使用される (図 1.4)。2つ目はデッサンである。デッサンとは、明暗や線の強弱を使い分けることによって、彩色を行わずに対象の形態を平面に描画する技法である。とりわけ人物イラストにおいては、体の部位



図 1.4: 自画像を描くときに用いられるグリッド付き鏡

の比率、骨格、筋肉など、人体として正確に表現するため、練習として人体デッサンがよく行われている。3つ目はクロッキーである。クロッキーとは対象を素早く描画することである。短時間で簡潔に描く練習をすることで、少ない情報量で対象を表現する練習となる。デッサンとクロッキーはどちらも3次元の対象物を観察して描画を行うが、その性質上、作品として大きく異なる。例を図 1.5⁴に示す。これらの練習法は、全て何らかの参考となる対象がある。しかし、創作を行う場合は自らのイメージを描くため、現実において参考とし、比較できるものはない。そのため、イラスト制作の初心者は、創作において自ら描いたイラストを正確に評価できず、上手く試行錯誤ができない可能性がある。また、デッサンやクロッキーは輪郭線のない立体を2次元のキャンバスに起こすため、線を選ぶのが初心者にとって難しい。そのため、初心者が初めて行うのに最適なイラストの練習法は線がはっきりと描かれているイラストの模写を行うことであると考えられる。

人物イラストの創作において参考に用いられるのは、ポーズ集や3Dモデルなどが挙げられる。しかし、それらは自らのイメージを実現するヒントとなりえるが、活用するためにはある程度の知識や経験が必要になるため、初心者にとってはハードルが高いと考えられる。また、キャラクターイラストにおいては、顔が人物の印象の多くを決定してしまう [13] ため、特に顔の作画のクオリティやバランスが重要となる。そのため、初心者はまずキャラクター顔を練習することで、イラストの上達を実感できやすく、モチベーションの維持につながると考えられる。

このような課題が存在する中、キャラクターのイラスト制作に関する研究がいくつか報告されている。例えば、元となるキャラクターの顔イラストと生成したい表情のランドマーク (顔のパーツや輪郭を決める特徴点) からその表情をしたキャラクターイラストを生成する手法 [14] や StyleGAN を用いて全身イラスト

⁴<https://oekaki-zukan.com/articles/4008>

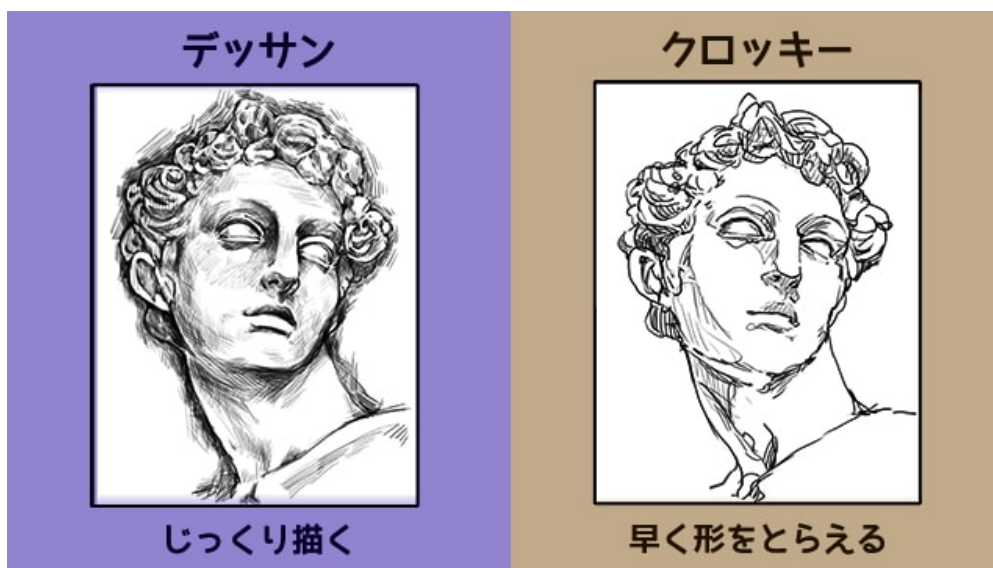


図 1.5: デッサンとクロッキーの違い

を生成する手法 [15] がある。しかし、これらの手法はインタフェースが用意されておらず、初心者にとって使いやすいとはいえない。インタラクティブに画像の生成を行う手法も報告されている [11]。この手法では、スケッチベース画像編集ネットワークを用いて、手描きスケッチから実世界に基づいた画像の生成や編集をすることができる。このように、イラスト生成を行う研究は行われているが、イラスト練習を支援する研究は少ない。そこで本研究では、イラスト練習に焦点を当て、初心者を対象としたイラスト練習の支援を行う。

デジタル環境で模写を行う場合であっても、参考イラストを見つけることとそのイラストと模写で用いるキャンバスにグリッドを描画する必要がある。そのためには、参考イラストをイラストサイト等で探したり、ペイントソフトの理解が必要であることから、初心者にとってはハードルが高いと考えられる。そこで本研究では、デジタル環境における初心者向け模写システムを提案する。模写に必要な参考イラスト探しから完成イラストの比較まで支援することで、初心者にとって使いやすいシステム構築を目指す。

1.2 研究目的

本研究では、模写過程においてグリッドを可変にすることで効果的な模写支援を可能にするという仮説を立てた。これに対し、ラフスケッチを入力とした参考イラスト推薦システムと、模写過程における可変グリッドシステムを組み合わせた模写練習支援のフレームワークを提案し、仮説の検証を行った。本研究では、イラストの中でも、キャラクターの顔イラストを対象として模写過程の支援を行う。本システムにより、ユーザイラスト練習の補助や、練習の意欲向

上の助けになることを目指す.

1.3 論文構成

本論文は5章で構成する. 第2章では関連研究を紹介し, 本研究の位置づけを示す. 第3章では提案手法について説明する. 第4章ではユーザスタディについて説明し, 第5章ではその結果を述べる. 第6章では本研究を総括する.

第2章 関連研究

本章では、イラスト生成・制作支援に用いられる深層学習、イラスト制作支援、模写支援に関する関連研究を紹介し、最後に本研究の位置づけを述べる。

2.1 イラスト生成・制作支援に用いられる深層学習

近年、深層学習を用いた画像生成が発表されている。本節では、特にイラスト生成として活用されているGANと拡散モデルについて説明する

2.1.1 GAN

Generative Adversarial Network(GAN)とは、日本語で敵対的生成ネットワークとよばれる教師なし学習による生成モデルである。オリジナルのGAN [1]が提案された2014年以降、GANに関する多数の研究が実施されている。GANはgenerator(生成器)とdiscriminator(識別器)の2つのネットワークで構成されていることが特徴である。生成器 G は、ランダムなノイズ z を入力とし、偽のデータ $G(z)$ を生成する。識別機 D は、入力が真のデータ X か偽のデータ $G(z)$ かを判定する。この2つのネットワークを同時に競わせながら学習することで、生成器 G が真のデータに近い偽のデータを生成することと、識別機 D が生成器 G が生成した偽のデータを見分けることを目指している。[1]での損失関数

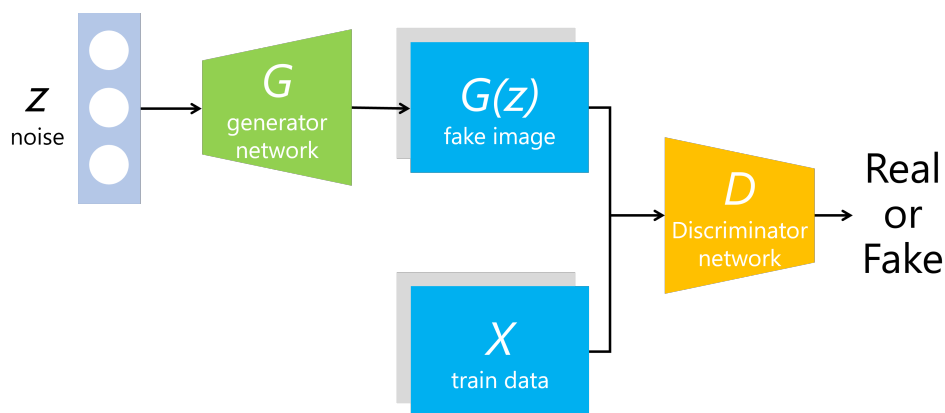


図 2.1: GAN の概要図 [1]

は以下の式 (2.1) のように表せる。損失関数とは、機械学習モデルが算出した予測と、実際の正解のズレを計算するための関数であり、この損失関数の値を最適化することでモデルの精度を向上させている。

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{data}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (2.1)$$

式 (2.1) では、識別機 D が正しく判別する確率 $D(\mathbf{x})$ を最大化し、識別機 D が生成器 G で生成されたデータを偽だと判断する確率 $(1 - D(G(\mathbf{z})))$ の最小化を目指している。

オリジナルの GAN では、学習が不安定であることや判定基準が不透明であることが課題となっており、その後様々な研究によって改善・発展が遂げられている。その一つに条件付き GAN (CGAN: Conditional GAN) [16] がある。オリジナルの GAN はノイズだけを入力するが、CGAN では条件ベクトルを与えることで、条件に対応した生成を可能とした。これを応用したものに、pix2pix [17] がある。pix2pix は、CGAN の条件ベクトルとして画像を用いる手法である。

2.1.2 拡散モデル

近年、テキストから画像を生成する画像生成 AI が発表されている。特に、2022 年に発表された Stability AI による Stable Diffusion¹ はオープンソースで提供され、無料で使用できるため、大きな話題となった。更に、Stable Diffusion を発展させ、キャラクターに特化したイラスト生成を可能にした NovelAI Diffusion² が発表された。これらの画像生成 AI は、プロンプトと呼ばれるテキストの記述に従って高品質の画像を生成することができる。この他にも様々な画像生成 AI サービスがリリースされているが、その多くで拡散モデルが使用されている。

拡散モデルとは、非平衡統計物理学から着想を得た深層学習モデルの一つで、画像からノイズへの変換の逆変換を学習することによって画像生成を行う。現在、普及している拡散モデルの多くは、ノイズ除去拡散確率モデル (DDPM; denoising diffusion probabilistic models) [2] に基づいている。DDPM では、図 2.2 に示すように Forward Process として元画像 \mathbf{x}_0 にガウスノイズを付加するマルコフ過程 $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ を T 回適用することでノイズ画像 \mathbf{x}_T の生成を行う。次に、Forward Process の逆、Reverse Process ではノイズを推定し、ノイズ成分を除去するマルコフ過程 $p_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})$ をニューラルネットワークで学習することで、ノイズ画像からの画像生成を可能にする。損失関数は、データを示す確率 $p_\theta(\mathbf{x}_0)$ を最大化するために、対数尤度 $\log p_\theta(\mathbf{x}_0)$ を最大化することを目指す。DDPM の問題点としては、何度もノイズを推論・除去して画像を生成するため、計算量が多く処理に時間がかかってしまうことが挙げられる。

¹<https://stablediffusionweb.com/>

²<https://novelai.net/>

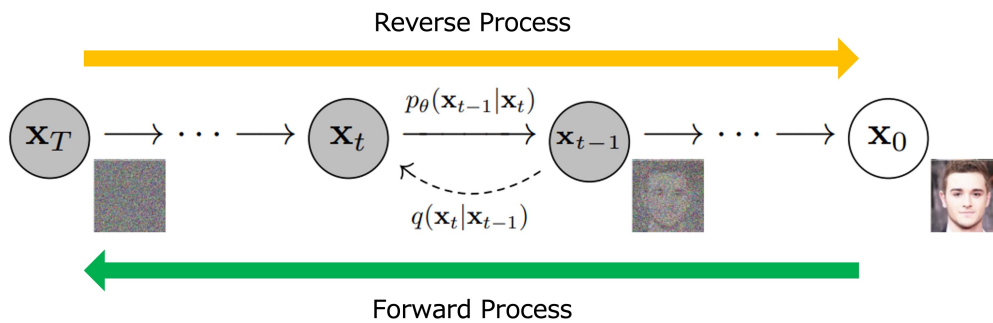


図 2.2: DDPM の概要図 [2]

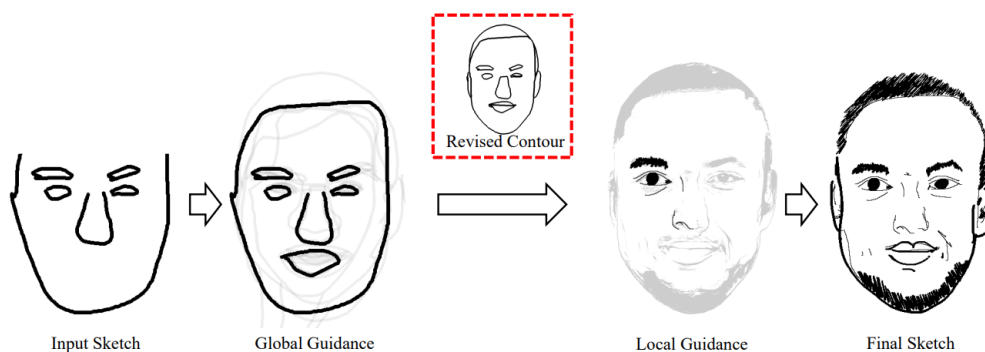


図 2.3: dualFace の概要図 [3]

Stable Diffusion では、潜在拡散モデル (LDM; latent diffusion model) [18] が元になっている。LDM では、画像を低次元の潜在表現へと変換し、その潜在空間上で学習・推論を行うことで、DDPM に比べて計算量を抑えることを可能とした。

2.2 イラスト制作支援

イラスト制作支援の研究事例として、大域的支援と局所的支援に分けてリアル調の似顔絵の描画支援を行う dualFace [3] がある。この手法では、似顔絵を書く際に、まず大まかな輪郭を描き、その後に細かな顔のパーツを描く過程から着想を得ている。大域的支援では、ユーザが描いた線からデータベースの画像を検索し、背景にガイダンス候補として表示する。局所的支援では、ユーザが描いた線から深層生成モデルを用いて詳細な描画ガイダンスを生成し表示する。

また、同じく描画ガイダンスシステムとして ShadowDraw [19] がある。この研究では、ユーザの描いている対象の関連画像を大規模データセットから検索することで複数見つけ出す。それらの画像を合成したものを、ユーザの描いて

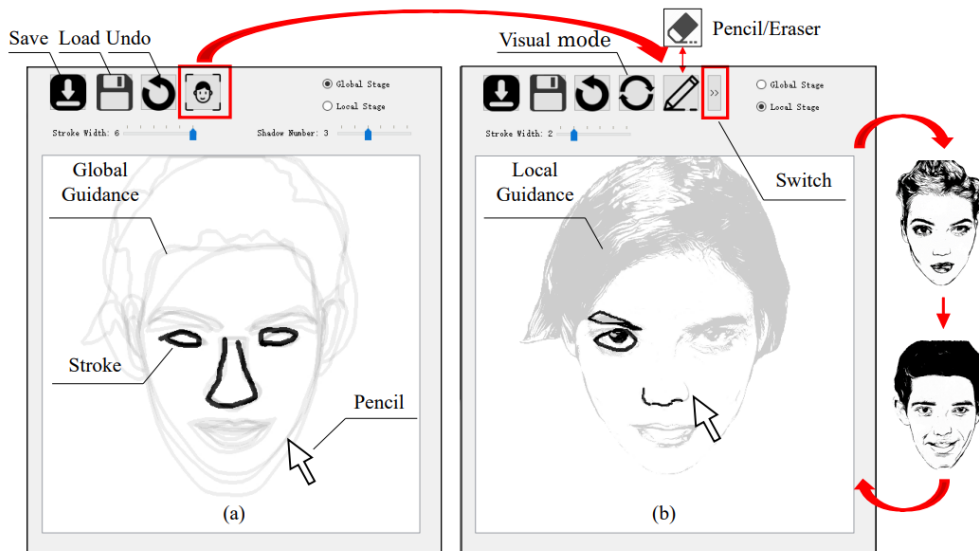


図 2.4: dualFace のインターフェース [3]

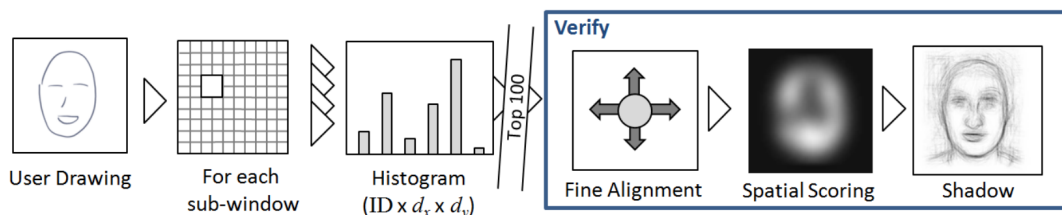


図 2.5: shadowDraw の概要図 [3]

いるキャンバスに影としてガイドを表示する。ユーザの描画にリアルタイムで適応し、影を更新することができる。画像検索はヒストグラムを用いているが、ハッシュ化を行うことで局所的と大域的な類似性を高速に求めることが可能となった。

デジタルイラスト制作における自動レイヤ分けシステム [4]がある。デジタルイラストでは、レイヤとよばれる層に分けてイラストを描くことができる。線画と着色、背景とキャラクターなどを別のレイヤに分けて描くことで、描画や修正をより簡単に行うことができる。レイヤ分けは、ペイントソフトの塗りつぶしや選択範囲ツールを用いて自動で行うこともあるが、描写が細かい箇所は失敗することも多く、手動で修正する必要がある。この研究では、pix2pix [17]を用いて、キャラクターのパーツ毎に学習を行うことで、パーツを認識し、レイヤ分け作業の自動化を行っている。

また、アーティストの作品制作過程をモデル化した画像生成・編集フレームワーク [5]が提案されている。CGAN [16]をベースにして制作過程を段階ごとに学習することで、作品の制作支援を行っている。また、既存の作品の制作過

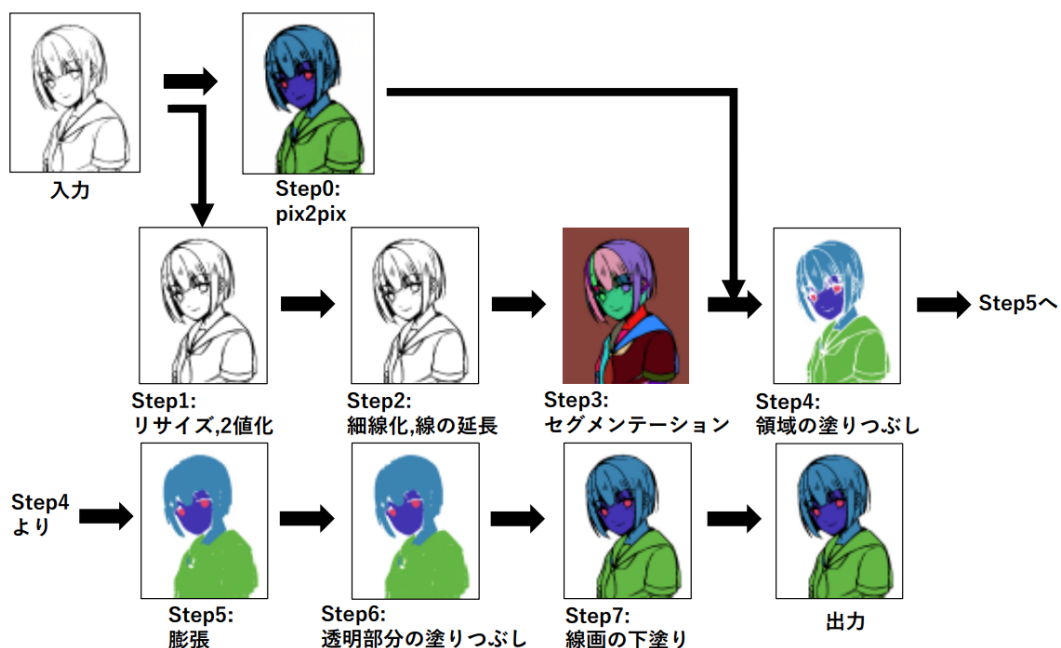


図 2.6: 自動レイヤ分けシステムの流れ図 [4]

程を推測することができ、制作の途中過程を編集することで、自然かつ元とは異なる作品を生成することができる。

2.3 模写支援

実人間の模写を通して、模写を行う上で重要な観察を支援する研究がある [20]. この研究では、イラスト描画の初心者の問題として挙げられる絵画対象の観察をおろそかにするという点を解決するために、自身の絵の修正点への気づきを促すシステムを提案した。提案システムでは、観察したことを言語化することによって意識的に観察を促すことや、部分的に遮蔽することでより深い観察を

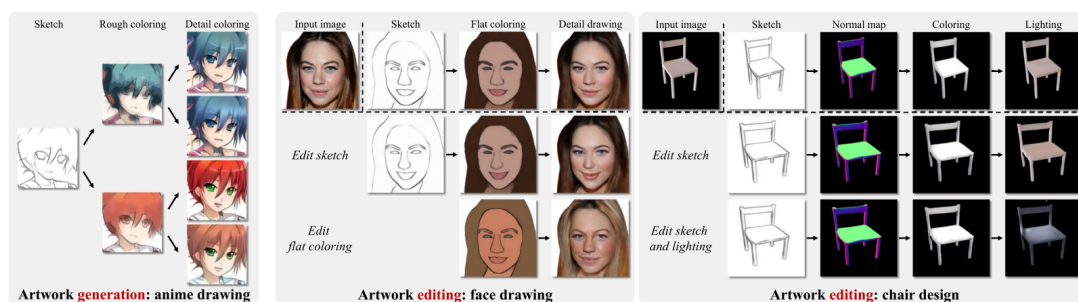


図 2.7: 制作過程を編集した例 [5]

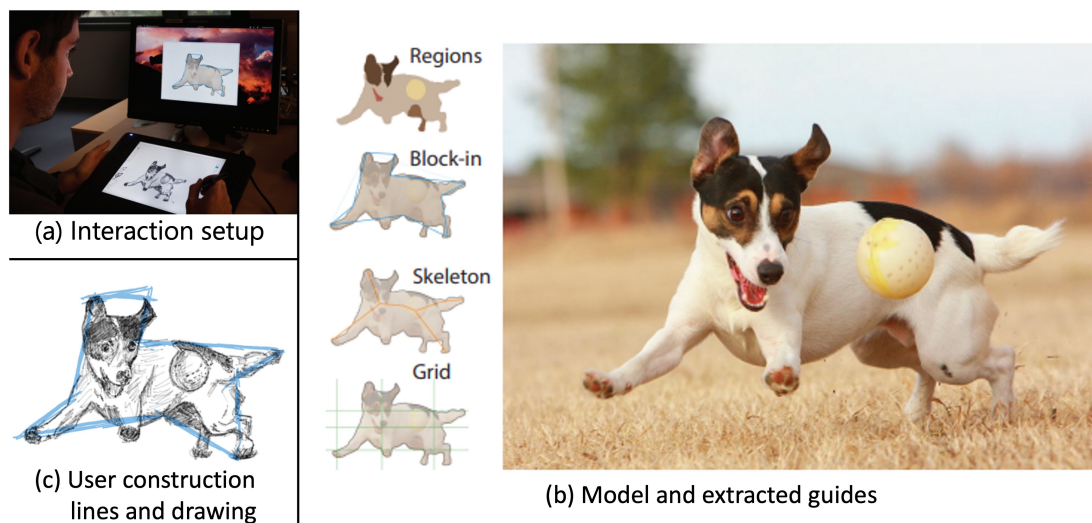


図 2.8: The drawing assistant の概要図 [6]

させるシステムを実装している。

また、模写の支援として、様々なガイド (図 2.8(b) 左側) を用いて描画を支援するシステム The drawing assistant [6] がある。このシステムで用いられているインタフェースでは、既存研究とは異なり、ガイドとして模写対象自体を用いることはせず、模写対象から抽出した幾何学的特徴からガイドを作成する。

2.4 本研究の位置づけ

既存のイラスト制作支援はユーザのイラスト練習を目的とはしておらず、模写支援では模写のツールとしてよく利用されているグリッドに焦点を当てた支援は少ない。本研究はキャラクタイラストを対象にして、模写で用いるグリッドを可変にすることが、観察に焦点を当てた効果的な模写支援を可能にするという仮説に基づく。そして、ラフスケッチを入力とした参考イラストの推薦と、模写における可変グリッドシステムを組み合わせた模写練習支援のフレームワークを提案し、検証した。

第3章 提案手法

3.1 概要

提案システムでは、ユーザによるスケッチを入力として、それを元としたキャラクターイラストを推薦する。また、スケッチの作成・編集と生成したイラストを表示するインターフェースを作成する。イラスト画像の推薦においてはラフスケッチから洗練化 [11] を行い、入力とする。入力画像の特徴量を抽出し、データセット内の画像と比較し、類似度が高い上位5枚のイラストを推薦する。その中からユーザが手本にしたいイラストを選択し、模写過程に移る。模写過程では、グリッドを変化させることができ、手本イラストと比較することができる。

文献 [11] では、洗練化においてデータセットとして実世界の人の顔画像である CelebA-HQ データセット [21] を用いているが、本手法ではキャラクターイラストから作成したデータセットでネットワークを学習させる。システム概要図を図 3.1 に示す。

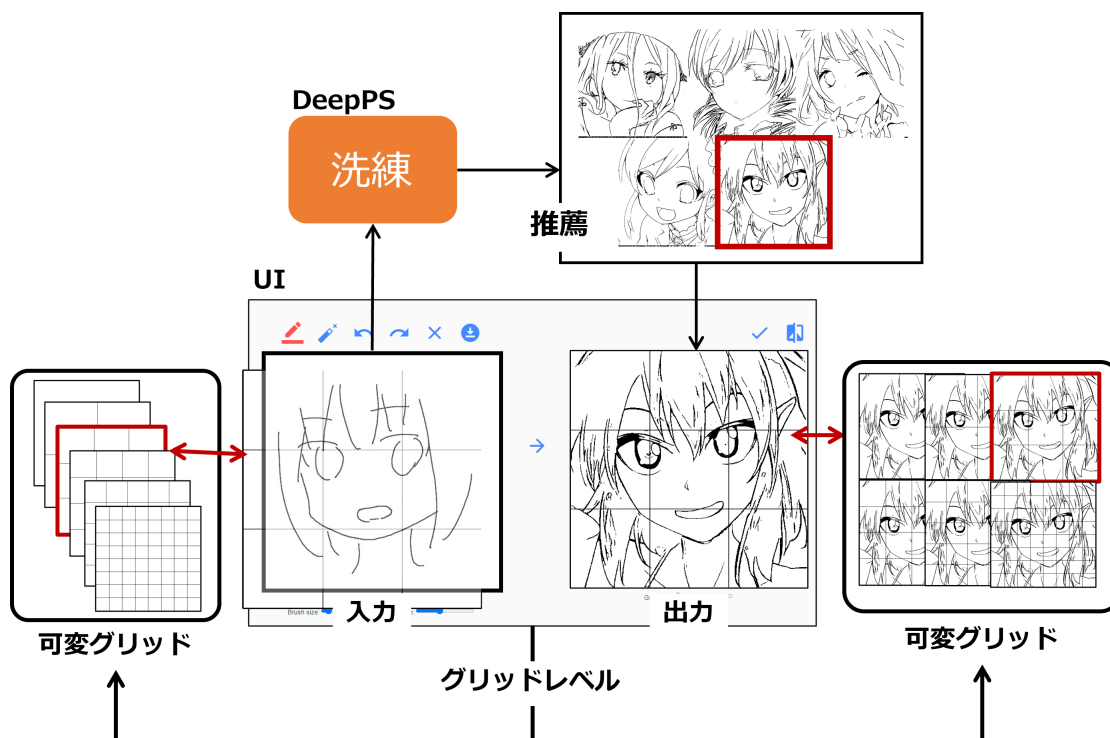


図 3.1: システム概要図

3.2 データセット

本研究において、データセット作成のために、イラスト画像データセットとして Face of Pixiv Top Daily Illustration 2018 ~ 2020¹²³と Tagged Anime Illustrations⁴を使用している。

データセット作成のために、イラスト画像内のキャラクターの顔範囲画像の抽出、線画抽出、2値化処理を行う。この処理は、PythonとOpenCVを用いて行う。

イラスト画像内のキャラクターの顔範囲画像の抽出は、nagadomiが作成したアニメ顔検出用のライブラリ `lbpcascade_animeface[7]` を使用する。本実験ではサイズが 512×512 px の顔イラスト画像を使用するが、データセットから顔範囲画像の抽出を行うと、 512×512 px 以上のサイズの顔範囲画像はごく少数となる。そのため、サイズが 300×300 px 以上の顔範囲画像を対象とし、OpenCV

¹<https://www.kaggle.com/datasets/stevenevan99/face-of-pixiv-top-daily-illustration-2018>

²<https://www.kaggle.com/datasets/stevenevan99/face-of-pixiv-top-daily-illustration-2019>

³<https://www.kaggle.com/datasets/stevenevan99/face-of-pixiv-top-daily-illustration-2020>

⁴<https://www.kaggle.com/datasets/mylesoneill/tagged-anime-illustrations>

の関数 `cv2.resize` を用いて 512×512 px に拡大する. `cv2.resize` は画像を拡大・縮小するための関数であり, デフォルトでは線形補間法を用いている. そして, `waifu2x-ncnn-vulkan`[22] を用いて画像拡大において発生したノイズ除去を行う. `waifu2x-ncnn-vulkan` は畳み込みニューラルネットワークによる二次元画像のための超解像システムである `waifu2x`⁵を, 高性能なニューラルネットワークの推論フレームワークである `ncnn`⁶を用いて実装したものである. これらの処理の結果, データセットとして使用する顔イラスト画像は 4,847 枚である. 顔範囲抽出の例を図 3.2 に示す⁷.



図 3.2: 顔範囲抽出の例 [7]

顔範囲の抽出後, Mukosame が作成した `Anime2Sketch`[8][9] を用いて線画抽出を行う. その後 2 値化処理を行う. この処理では, 画像ごとに最適なしきい値を求めるため, 判断分析法 (Discriminant Analysis Method; 大津の手法 [23]) を用いた.

⁵<https://waifu2x.udp.jp/index.ja.html>

⁶<https://github.com/Tencent/ncnn>

⁷<https://www.pixiv.net/en/artworks/85705876>

顔範囲抽出

線画抽出

2値化

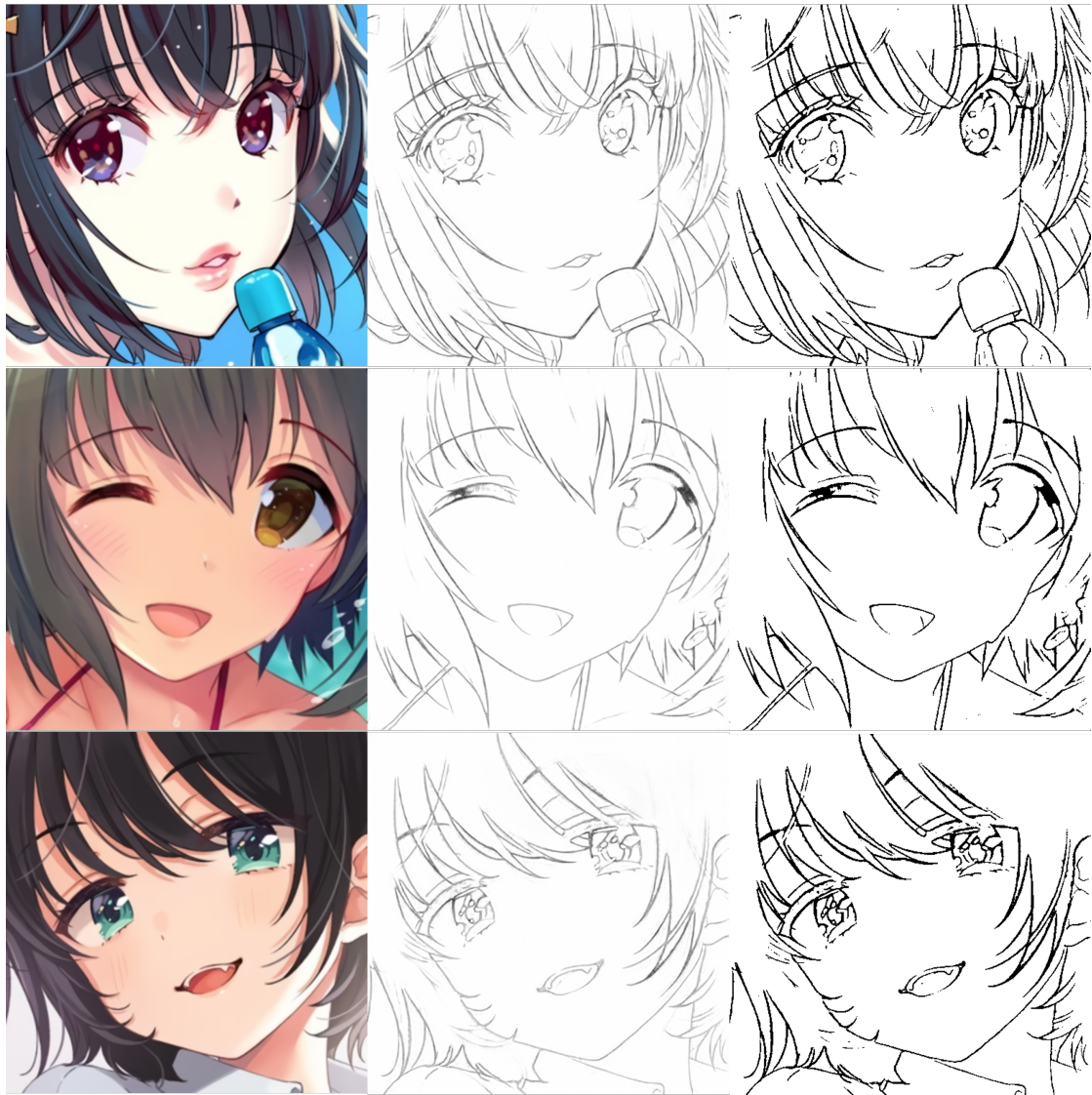


図 3.3: 線画抽出・2値化例 [8][9]

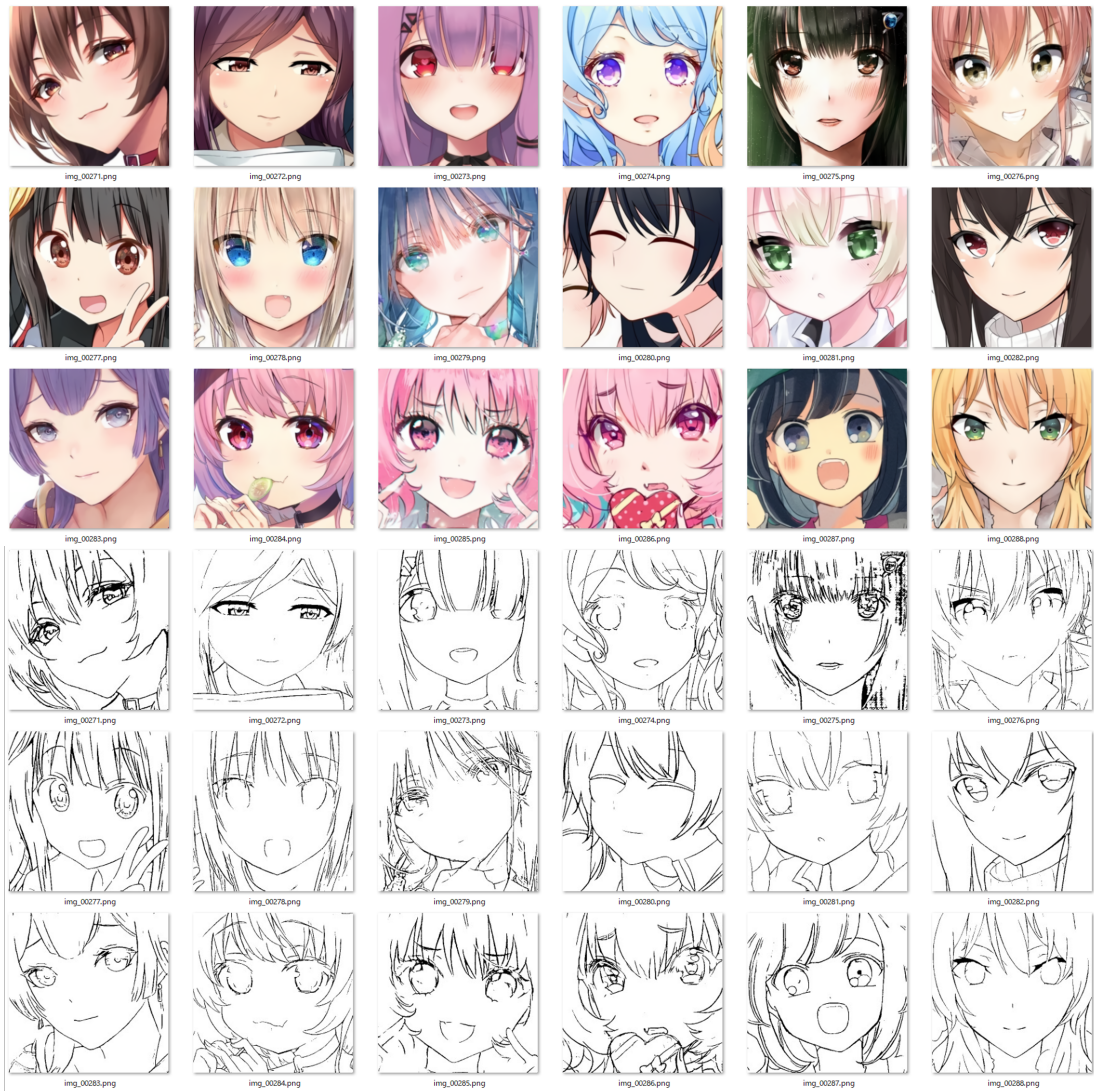


図 3.4: データセット例 (上: 顔イラスト, 下: 対応する線画画像)

3.3 画像推薦システム

本研究では、模写における手本イラストの推薦を行う。推薦については、ユーザのラフスケッチを入力として、データセットから類似画像を検索し、類似度の高い上位5枚を推薦する。本節では、類似画像検索手法の検討と、ユーザのラフスケッチの処理について示す。

3.3.1 特徴量を用いた類似画像検索

現在、類似画像検索で用いられる代表的な手法には、CNNを使った特徴量抽出や、Siamese Network[24]などがあげられる。これらの手法は色情報を用いて学習を行っていることが多い。しかし、モノクロ画像である手本イラストの推薦には、色情報を必要としない手法のほうが適していると考えられる。色情報を用いずに画像の類似度を求める手法として、ヒストグラムベースの方法や、特徴量抽出、テンプレートマッチングなどがあげられる。ヒストグラムベースの方法は画像内の輝度値をヒストグラムで表す手法であり、輝度の分布を使用する。特徴量抽出は画素中に存在する重要な点である特徴量のうち、周辺の濃淡の変化が多い特徴点の周辺領域の輝度や画素の微分値から得られる特徴ベクトルを抽出する方法である。テンプレートマッチングはテンプレート画像に一致する部分を見つける手法である。本研究ではラフスケッチからデータセット内の画像検索を行うため、ラフスケッチの大まかな特徴を捉えられるかが重要である。そのため、ヒストグラムベースとテンプレートマッチングは適さないと考えられる。したがって、本研究では類似画像検索の手法として特徴量抽出を用いる。

画像の特徴量には様々なものがあるが、代表的なものとしてSIFTとHOGがある。

SIFT

SIFT(Scale-Invariant Feature Transform)[25]は、画像中の特徴点周辺における勾配情報を表すベクトルを特徴量とする。SIFTの処理は、特徴点の検出と特徴量の記述に分けて行われる。図3.5にSIFT特徴量の計算の流れを示す。

特徴点の検出では、スケールの異なる平滑化画像の差分を取ることで、特徴点と思われる点を決定する。具体的には、各スケール間においてDoG (Difference of Gaussian)を用いて画像から特徴点を検出する。DoGを用いることで、物体と背景の境目などの画像の変化が起こる部分を算出し、この変化が極大となる点を特徴点とする。次に、検出された特徴点に対して、周囲の画素値を用いて特徴量を計算する。この特徴量は、特徴点の周囲を格子とした画像の勾配ヒストグラムや、画像の変化がどの範囲に及んでいるかのスケール情報を含

む。図 3.5 中央右の画像における円の半径がスケール情報を表す。特徴点周辺の画像勾配から特徴点の方向を求めるため、画像の回転があっても特徴形状が変化しない限り同じ特徴点を求めることができる。

各特徴点の持つ特徴量の次元は、勾配ヒストグラムの格子数と勾配の方向数で決まる。図 3.5 では、格子数が 4×4 、勾配の方向数が 8 であるため、各特徴点の持つ特徴量の次元は $(4 \times 4) \times 8 = 128$ 次元となる。

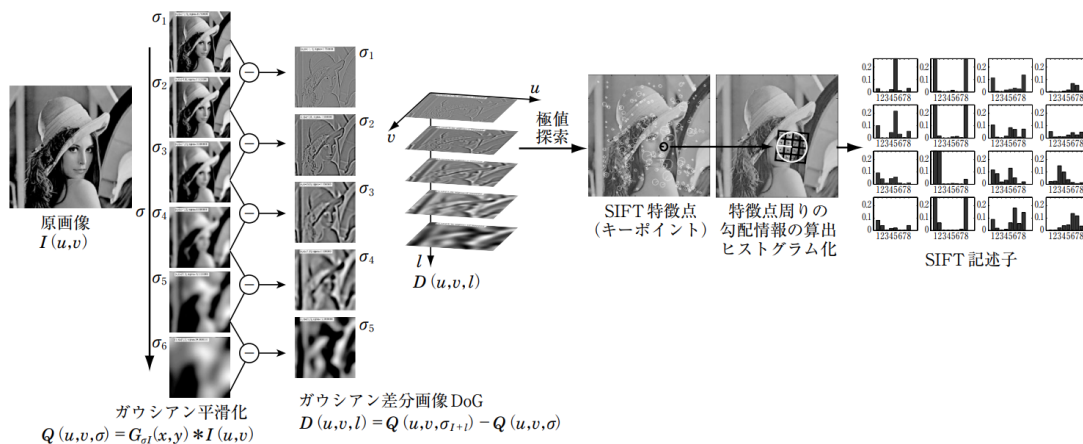


図 3.5: SIFT 特徴量の計算方法 [10]

HOG

HOG(Histogram of Oriented Gradients)[26] は、画像の輝度勾配をヒストグラム化して記述した特徴量である。図 3.6 に HOG 特徴量の計算の流れを示す。

HOG では、(1) 輝度勾配の計算、(2) 局所ヒストグラム化、(3) 局所ブロックによる正規化の 3 つの処理を行う。

はじめに、画像を小さな領域(セル)に分割する。これにより、画像の局所的な特徴の抽出を行う。次に、画像内の輝度勾配方向の計算を行い、セル内部の勾配方向をヒストグラム化する。セルの近傍セル領域をブロックとし、ブロック内でヒストグラムの正規化を行うことで、あるセルに対する特徴量を得る。

セルの持つ特徴量の次元は、ヒストグラム化において指定する勾配方向数と、ブロックの大きさに依存する。例えば、図 3.6 の例では勾配方向が 9、ブロックを 3×3 としているので 1 つのセルが持つ特徴量の次元は $9 \times (3 \times 3) = 81$ 次元となる。画像全体の大きさを 256×256 とすると、セルの個数は 16×16 個となるため、全体では $(16 - 2)^2 \times 81 = 15876$ 次元の特徴量ベクトルを得る(近傍がない画像端セルは考慮しない)。

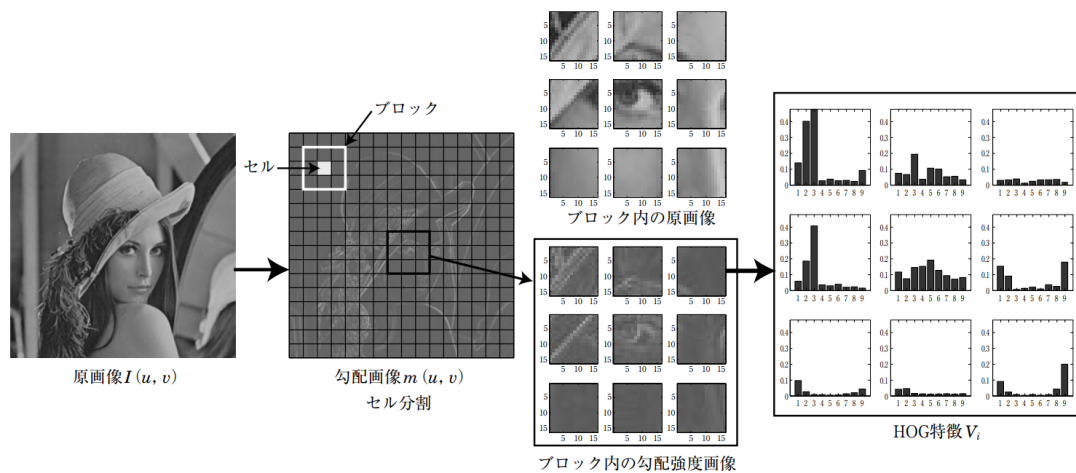


図 3.6: HOG 特徴量の計算方法 [10]

本研究では，この2つ特徴量のどちらを使用するかを決定するための予備実験を行う。

3.3.2 特徴量の比較 (予備実験)

予備実験として，HOG 特徴量と SIFT 特徴量を利用した類似画像検索を行い，両者の比較を行う。実験の内容は，データベース内の画像の特徴量を保存するインデックスの作成時間の比較，インデックスを利用したラフスケッチを入力とした類似画像検索の質と検索時間の比較である。類似画像検索の質は5段階評価を行う。どちらの特徴量に対しても，特徴量のマッチングは OpenCV の `BFMatcher.knnMatch()` 関数を用いて行った。 `BFMatcher.knnMatch()` は，特徴量同士を距離で比較し，最も距離の近い k 個の結果を返す関数である。本予備実験は `BFMatcher.knnMatch` の引数を $k = 2$ とし，類似画像検索は類似度が高い5枚のイラストを示すことで評価を行った。本予備実験の被験者は5人で，全員20代男性であった。

結果

データベース内の画像の特徴量を保存するインデックスの作成時間は，HOG 特徴量が3分48秒，SIFT 特徴量が2分56秒となった。また，平均検索時間に関しては，HOG 特徴量が0.36秒，SIFT 特徴量が17.26秒となった。類似画像検索の質を5段階評価で行ったところ，平均値はHOG 特徴量が4，SIFT 特徴量が2.2となった。検索結果例を図3.7に表す。

平均検索時間は HOG 特徴量のほうが非常に早く、主観的ではあるが画像検索の質も HOG 特徴量のほうが SIFT 特徴量に対し 1.8 も差をつけるほど高かったため、提案システムでは HOG 特徴量を使用することにする。

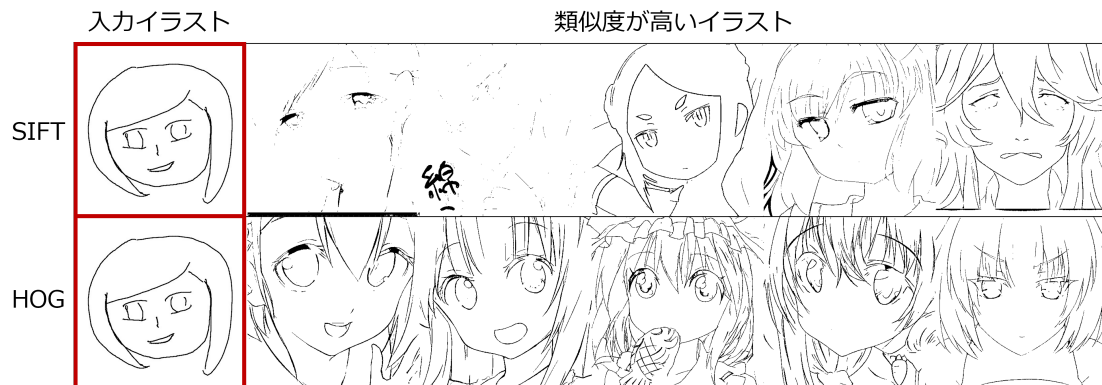


図 3.7: 予備実験の結果例

3.4 ラフスケッチの洗練化

類似画像検索においてユーザが描いたラフスケッチを入力とするが、ラフスケッチはデータセット内のイラストとは違い詳細度は低く、未完成である。そのため、類似画像検索に使用するためには詳細度を上げることでより良い精度が得られると考えられる。本研究では、ラフスケッチの詳細度を上げるために DeepPS[11] を参考としたスケッチ洗練ネットワークを利用する。

DeepPS では、線画の膨張 (Dilation) を行い、膨張レベル l ごとに膨張した線画から元線画を出力するような学習を pix2pix[17] を用いて行っている。この過程は図 3.8 の G にあたる。また G では、洗練レベル l に応じて畳み込み層の出力を調整している。洗練度 l による画像の洗練結果が図 3.9 である。また図 3.8 の F は pix2pix で線画をカラー画像に出力するように学習し、 D は GAN である。

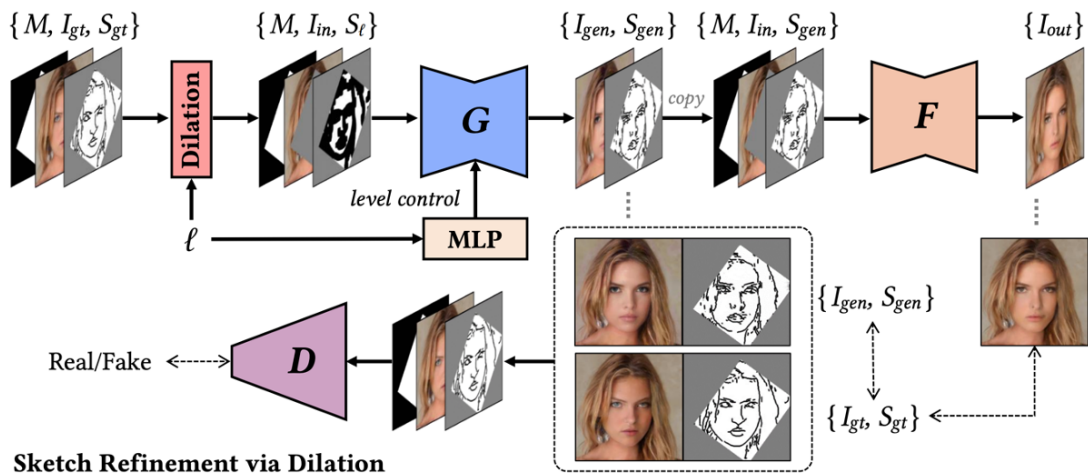


図 3.8: DeepPS の概要図 [11]

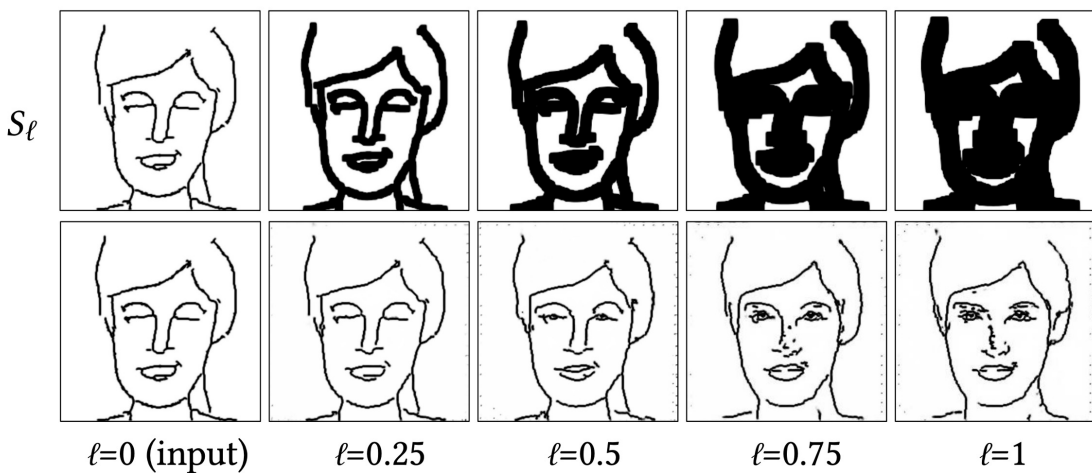


図 3.9: 洗練度 ℓ による画像の洗練 [11]

DeepPS では実際の写真である CelebA-HQ[21] を使って学習を行っているが、本研究はイラストを対象にしているため 3.2 節で作成したデータセットで学習を行った。本研究におけるデータセットを適用させた洗練例を図 3.10 に示す。図 3.10 では、上部の粗いイラストの詳細度が上がっているが、下部の比較的細かいイラストの場合線が統合することで減っていることが確認できる。本研究におけるデータセットは DeepPS で用いている CelebA-HQ データセット [21] (30,000 枚、サイズ 1024×1024) に比べて小規模なため、精度が落ちてしまうと考えられる。また、実際の顔写真ではなくキャラクタイラストを扱っているため、それを考慮した処理を行う必要があると考えられる。

入力

洗練後



図 3.10: スケッチ洗練例 ($l = 1$)

3.5 インターフェース

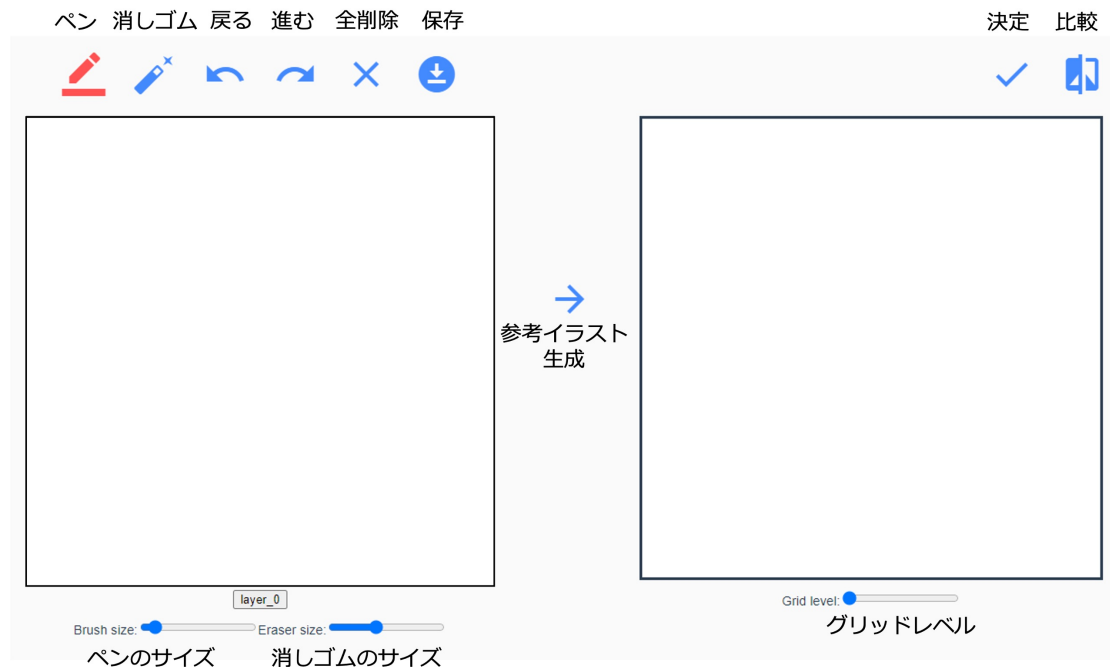


図 3.11: 提案インターフェース

図3.11 に示すインターフェースの開発を行った. 本インターフェースはJavaScriptの2Dキャンバスライブラリであるkonva.jsとJavaScriptのフレームワークであるVue.js, Pythonを用いて実装した. 本インターフェースでは, スケッチの作成・グリッド操作を行うために, 以下の機能を有している.

- ペンと消しゴム
- ペンと消しゴムのサイズ変更
- 操作の取り消しとやり直し
- キャンバスの削除
- スケッチの保存
- 参考イラスト生成
- 参考イラスト・グリッド決定
- 比較
- スケッチキャンバスと参考イラストを表示するウィンドウ

グリッドレベルはグリッドなし、 5×5 、 4×4 、 3×3 、 2×2 、を設定していたが、もう一段階細かなグリッドが必要と考えたため、 5×5 よりグリッド1つあたりの面積が半分以下の 8×8 を加えて6段階の設定を行った。例を図3.12に示す。

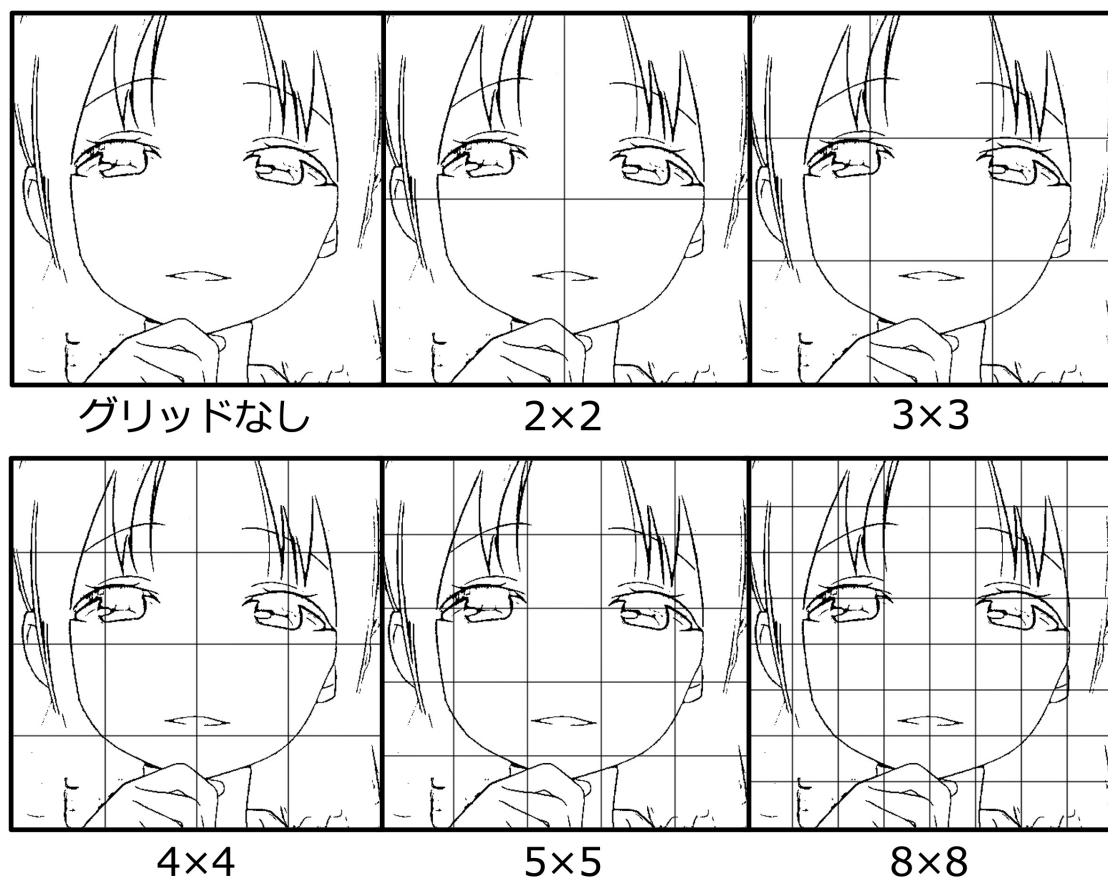


図 3.12: グリッド追加例

第4章 ユーザスタディ

4.1 評価実験

提案システムを評価するためのユーザ実験を行った。本実験では、20代から30代の男女計10人(男8,女2)に対して本システムを利用してもらった。提案システムによる支援ありと支援無しでアンケート調査を行い、最後にSUS(System Usability Scale)とNASA-TLXによる評価を行った。SUSとNASA-TLXの詳細はそれぞれセクション4.2, 4.3内で説明する。

被験者は図4.1のようなパターン①とパターン②に半分に分けられ、それぞれの流れに沿って実験を行った。実験の流れを図4.1に示す。

実験手順としては、まず本研究の趣旨や提案システムの概要説明を行い、被験者のイラスト経験などを調査する事前アンケートを行った。次に、提案インタフェースの説明を行い、5分程度の練習を行った。提案インタフェースの説明では、実際に著者がインタフェースを利用した際の動画を流しながら説明を行った。練習では、提案インタフェースに慣れるために、実験の一連の流れを練習した。

事前アンケートでは、表4.1に示す内容の調査を行った。

表 4.1: 事前アンケート内容

- | |
|--------------------------|
| 1 イラストをどのくらいの頻度で描きますか |
| 2 模写をしたことはありますか |
| 3 (前設問であると答えた場合) 何回ありますか |
| 4 グリッドを用いて模写をしたことがありますか |

その後、インタフェースを利用してシステムによる支援あり・支援無しで模写の制作を行った。その際、制作過程を動画として収録し、制作時間と用いたグリッドを記録した。使用するグリッドは被験者が選択したものを基準とした。支援ありの場合はグリッドを変化させることができるが、必ずしも変化させる必要はないことを説明した。

模写の制作を行った後、表 4.2 に示すアンケート調査を行った。

表 4.2: 模写後のアンケート内容

- | |
|---------------------------------|
| 1 絵の満足度 (5 段階評価) |
| 2 絵の完成度 (5 段階評価) |
| 3 上手く描けたと思うか (5 段階評価) |
| 4 気をつけたこと |
| 5 難しかったこと |
| 6 作成した模写に違和感があるか (ある・無し) |
| 7 (前設問で違和感があると答えた場合) どこに違和感があるか |
| 8 修正するとしたらどこをどのようにするか |
| 9 感想 |

加えて、支援ありで模写の製作を行う場合は、表 4.3 の内容も調査した。

表 4.3: 模写後のアンケート内容 (支援あり)

10 グリッドが可変なのは模写で有用であったか (5段階評価)
11 前設問の理由
12 画像推薦システムは模写で有用であったか (5段階評価)
13 前設問の理由
14 またこのシステムを使って模写をしたいと思うか (5段階評価)
15 前設問の理由

支援あり・無しで模写制作を行いアンケートを回答した後, SUS と NASA-TLX による評価を行った.

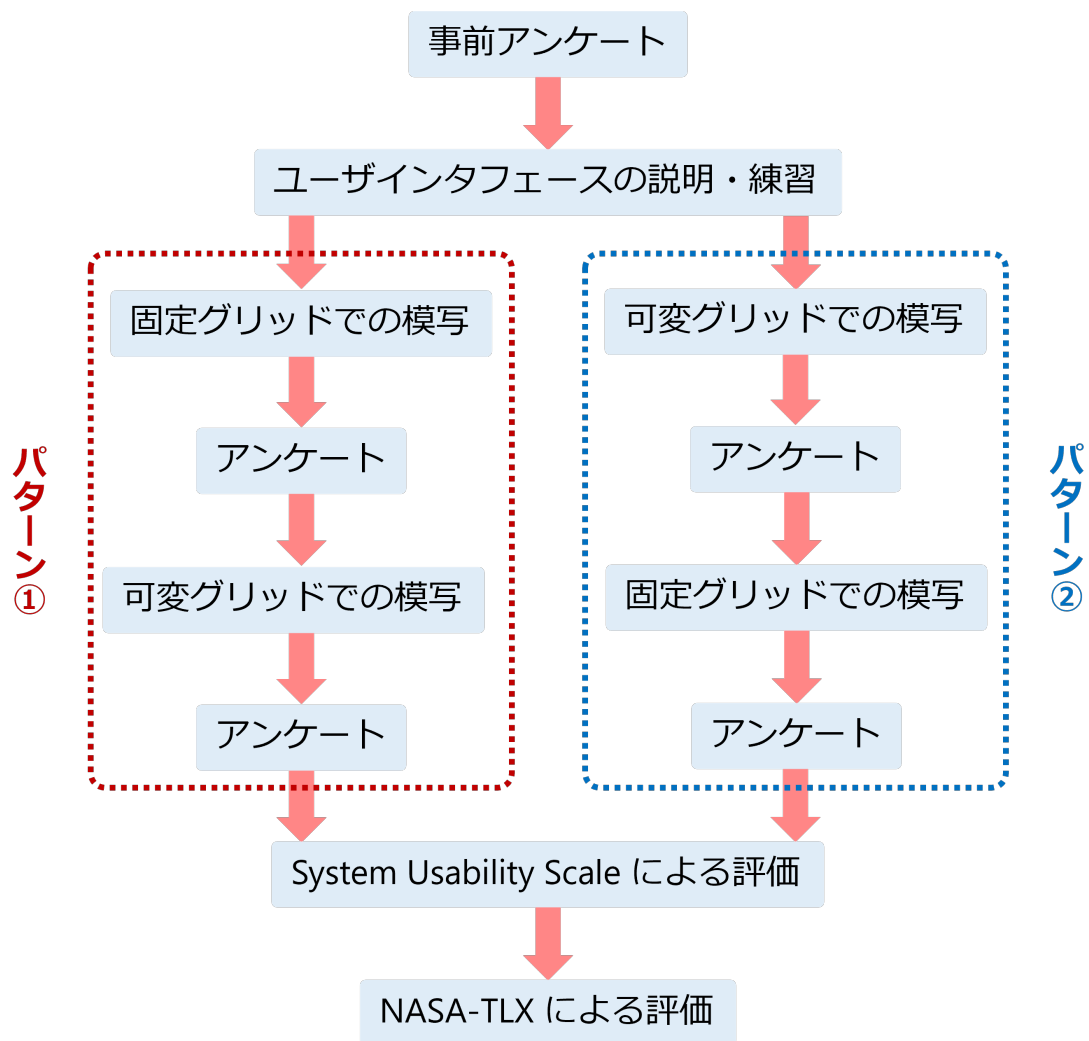


図 4.1: 実験の流れ

4.2 System Usability Scale

System Usability Scale(SUS)[27]は、ユーザビリティに関する主観的な評価の全体像を示すための評価手法であり、有効さ、効率、満足度を評価することができる。SUSは、10問の質問に対し、5段階のリッカート尺度で評価を行う。本実験で使用したアンケート項目を表4.4に示す。

表 4.4: SUS アンケート内容

- 1 (デジタル環境で模写を行う場合) このシステムをしばしば使用したいと思う。
- 2 このシステムを使用するには説明が必要となるほど複雑であると感じた。
- 3 このシステムは使いやすいと感じた。
- 4 このシステムを利用するには、専門家・技術者のサポートが必要だと思う。
- 5 このシステムの様々な機能には統一感があると感じた。
- 6 このシステムには一貫性が無いところが多いと感じた。
- 7 ほとんどの人がこのシステムはすぐ使いこなせるようになると思う。
- 8 このシステムは操作しづらいと感じた。
- 9 このシステムを自信を持って使える。
- 10 このシステムを使いこなすには事前にたくさんのことを学ぶ必要があると思う。

SUSでは、上記で示すように奇数番号ではポジティブな質問、偶数番号ではネガティブな質問を設けている。このように性質の異なる問いを交互に配置することで、回答者が質問に対して熟考することなく安易に回答することを防いでいる。そのため、スコア計算は質問の性質を考慮して行う必要がある。したがって、各質問の評価に対して後処理を行う。SUSの評価計算は、各質問への評価を x とすると、ポジティブな質問である奇数番号の質問では $x-1$ 、ネガティブな質問である偶数番号の質問では $5-x$ として計算する。全体の評価は0から100のスケールで表すため、後処理を行った評価を合計し、2.5を掛けることで算出する。

4.3 NASA-TLX

NASA-TLX[28](以下, TLX)はあるタスクに対するメンタルワークロード(精神的作業負荷・負担)を評価する手法である. 6種類の評価尺度を持ち, 3つのグループに分けることができる. オリジナル [28] は英語であるが, 本実験では日本語版 [29] を採用した. 以下は日本語で評価尺度の定義を表したものである.

1. 課されたタスクによる要求

知的・知覚的要求 (小さい/大きい)

知的・知覚的活動(考える, 決める, 計算する, 見るなど)の必要度

身体的要求 (小さい/大きい)

身体的活動(押す, 引く, 回す, 制御する, 動き回るなど)の必要度

タイムプレッシャー (弱い/強い)

タスクの頻度やペースのために感じる時間切迫感の度合い

2. 求められたタスクに対する反応

作業成績 (良い/悪い)

タスクで課された目標に対する達成度や満足度

3. 実行したタスクに対する自己評価

努力 (少ない/多い)

タスクを達成・維持するために必要な努力度

フラストレーション (低い/高い)

タスク中に感じた不安やストレス度

回答者は, これら6つの評価尺度に対して, 21段階のアナログスケールにより評価を行う. この評価は0~100のスコアに置き換えられて記録される. TLXは, このスコアをそのまま評価として用いるのではなく, 尺度ごとのスコアに加え, 各尺度の一対比較によって決まる重み係数を掛けることで, 各評価尺度のスコアを算出する. この一対比較で重要なのは, 比較される2つの評価尺度に関してどちらのスコアが大きいのではなく, どちらが重要な関わりを持つかを比較することである. この点を回答者に対して十分に理解させた上で回答させる必要がある. 各評価尺度のスコアを v_i , 重み係数を w_i とすると, 全体的なワークロード評価 (WWL: Weighted Workload) は次式で算出される.

$$WWL = \frac{\sum_{i=1}^6 (w_i \times v_i)}{\sum_{i=1}^6 w_i} \quad (4.1)$$

4.4 実験環境

実験環境を表 4.5 に示す.

表 4.5: 実験環境

OS	Windows 10 Education
RAM	32GB
CPU	Intel® Core™ i7-10700
GPU	NVIDIA GeForce RTX 2070 SUPER
液晶タブレット	Wacom Cintiq 22 (21.5 型)

第5章 結果

本章では，提案システムを利用したユーザスタディの結果より，本研究の有効性を示す。



図 5.1: 実験の様子

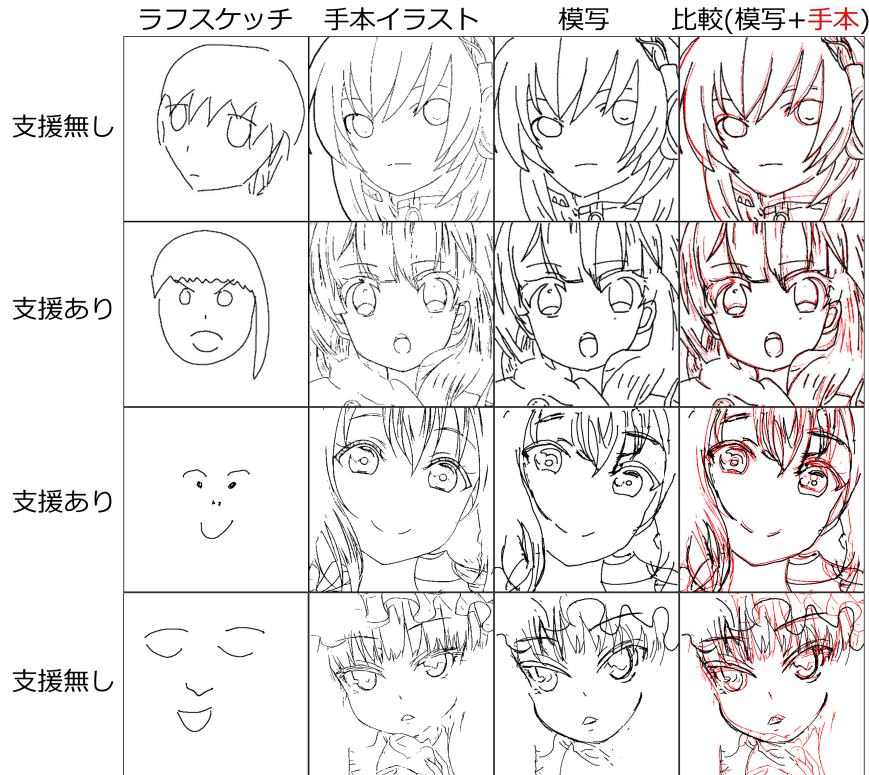


図 5.2: 制作例

5.1 事前アンケート調査の結果

被験者 10 人のうち、イラストを描いたことがあるのは 8 人で、そのうち 2 人が月に数回以上描いていた。また、模写の経験については 8 人が経験があり、そのうち 2 人は 10 回程度、1 人は頻繁に行うと答えていた。グリッドを使った模写の経験がある被験者は 1 人であった。

5.2 アンケート調査の結果

制作時間の平均は支援なしで 14 分 42 秒、支援ありで 14 分 34 秒となった。比較を図 5.3 に示す。

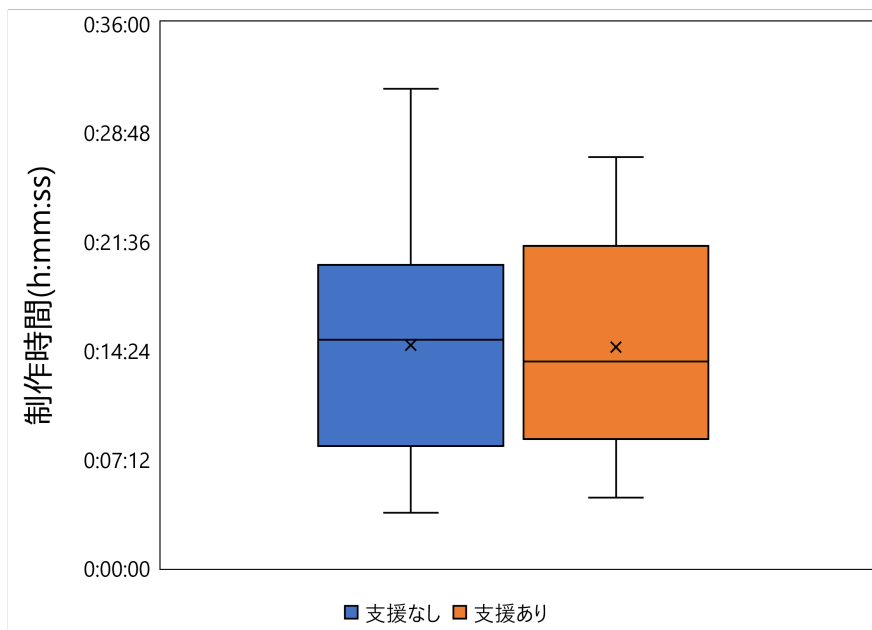


図 5.3: 制作時間の比較

図5.3から読み取れるように、制作時間の最大値は支援なしのほうが大きかったものの、支援ありと支援無しの制作時間に違いが見られるか検定を行った。サンプル数が少ないため、ノンパラメトリック検定の1つであるマン・ホイットニーのU検定を用いた。その結果、制作時間に有意差は確認されなかった ($U = 47, p > 0.10$)。支援なしの制作時間の最大値が大きかった理由は、被験者が指定した参考イラストが他で選ばれた参考イラストより線の描写が多く、時間がかかってしまったためだと考えられる。

次に、表4.2で示したアンケート項目の比較を行う。5段階評価である設問1～設問3の結果を図5.4で示す。

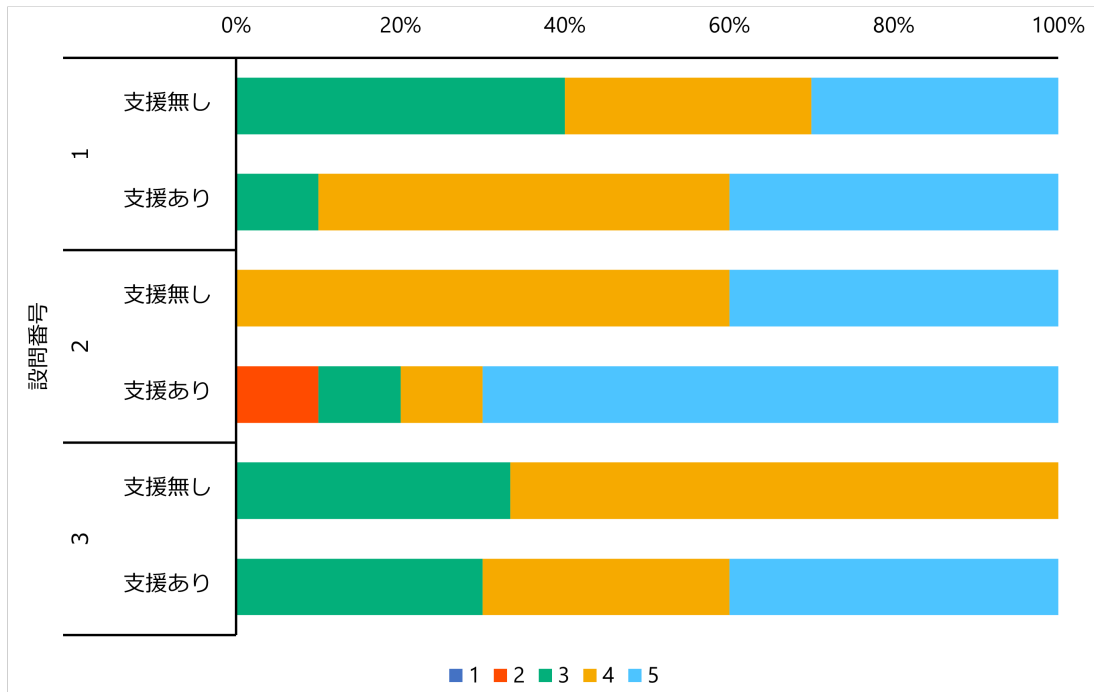


図 5.4: アンケート結果 (設問 1~設問 3)

また、設問 1,2,3 において支援なしの評価の平均値はそれぞれ 3.9, 4.4, 3.9, 支援ありの評価の平均値はそれぞれ 4.3, 4.4, 4.1 であった。

設問 1: 絵の満足度と設問 3: 上手く描けたと思うかは、支援ありの方が評価が高かった。一方、設問 2: 絵の完成度は同じ値となっている。つまり、平均ではどの被験者も支援ありと支援なしで同程度の完成度で制作したということを示している。制作時間の比較と同じく、マン・ホイットニーの U 検定を用いた結果、設問 1~3 に有意差は確認されなかった (設問 1: $U = 36.5, p > 0.05$, 設問 2: $U = 41, p > 0.05$, 設問 3: $U = 42.5, p > 0.05$)。

次に、記述で回答を行った設問 4~9 のコメントを以下に示す。

アンケート結果: 支援なしの場合

表 5.1: 設問 4: 気をつけたこと (支援なし)

見本において、グリッドと見本絵との交点をなるべくずれないように描いた.
顎の角度
グリッド線とのバランスと細かな影の表現
描いている線が他の線と上手く組み合わさるように気をつけた.
目の位置に注意した
模写なので線の位置を極力合わせるように気をつけた
グリッドを見ながらどの線がどこにあるかをしっかり意識しようと思った.
目と輪郭の位置とグリッドの位置
線とグリッドの位置関係を意識すること。
顔の大まかなバランス

表 5.2: 設問 5: 難しかったこと (支援なし)

交点を意識すると、ぱっと見の絵の見た目が歪というか、無骨に見えたので、その後の修正で妥協点を見つけるのが難しかった。
髪と顎の角度
グリッドが固定だったので細かな調整が難しかった
グリッド線から離れた線を書き始めるのが難しかった.
目の細かい曲線
マス目が粗いので、位置がどこか探ることに時間をかけた
髪の毛を描くのが難しかった
顔から下を描くときに複雑でややこしかった
滑らかな線を描くのが難しかった。

表 5.3: 設問 7: どこに違和感があるか (支援なし)

顎と髪 特に顎の角度
目や首のバランスがおかしく感じる.
目がアンバランスだった点
顔の形
目の位置と顔の輪郭。

表 5.4: 設問 8: 修正するとしたらどこをどのようにするか (支援なし)

描けていない線を付け足したい。
顎
顔や細かな毛の位置や角度のバランスを整えたい
目や首のバランスを何回か書き直して修正する.
目をバランスよく描く
目の位置と髪の位置
目、鼻、口のバランス。
顔以外をもっと描く
右目の位置を上調査、顔の線をちょっと控えめに小顔方向へ調整するかと。
満足しています。

表 5.5: 設問9: 感想 (支援なし)

自分の中ではとても上手な漫画のような絵が描けた！満足です。
面白い ずっと書きたいです
グリッドが調整できない分難しかったです、 2回目目もあって上達が見られたのが嬉しかったです。
グリッド線がもう少し細かくできると書きやすくなると感じた。
2回目の方が慣れて上手くかけましたが、 グリッドを複数使い分けて描く方が描きやすいと感じました。
グリッドを粗くすることでかなり描きづらくなることを実感しました
純粹に模写が難しかった。
自分には模写の才能があるかもしれない
グリッドの存在はとても助かる。
グリッドが大きいと、顔の大きなバランスを意識できて、 グリッドが小さいとパーツの細かい形がうまくかかるので、一長一短だと感じた

アンケート結果: 支援ありの場合

表 5.6: 設問 4: 気をつけたこと (支援あり)

点と点をつなげて、なるべく正しい位置に線を書くことを心がけた。
グリッドの箇所を間違えないようにした。
目と眉毛と髪
可愛くさせようと思った、元が可愛かったから。
前髪の位置
模写なので、似たように描くことをとにかく意識した。線の位置など
絵のバランスを崩さないように、ある程度は妥協して見本を真似た。
グリッドを参考にしながら線を描いた
線とグリッドの位置関係と線の傾きとか
グリッドを基準として正しい位置に正しいパーツを配置すること

表 5.7: 設問 5: 難しかったこと (支援あり)

グリッドと絵の交点を繋げて描いていたので、丸みを帯びた線を描き直すのが難しかった。
髪の毛の模写が難しかった
髪に分目
目がキラキラしてたので、そのキラキラがめちゃくちゃ難しかった。
線の太さの使い分け
時々このマス目なのか見失うことがあった
目を描くことが、細かい部分や丸があり、むずかしかった。
線の太さを扱うのに慣れていなかったなので、ほとんど同じ太さで描いてしまった
線を出るだけ滑らかに描くことに気を向けば角度とかに気を配る余裕がなくなる。
全体的な形どりが難しかった

表 5.8: 設問 7: どこに違和感があるか (支援あり)

髪の毛の表現が足りない.
ない
目
目の位置がズレている
顔の輪郭のバランスがうまく描けない
顔の輪郭が大きい

表 5.9: 設問 8: 修正するとしたらどこをどのようにするか (支援あり)

顎をもっとシュッとさせる。口をもう少し閉じさせる。
髪の毛をもうちょっと細かく修正したい。もっと細かいグリッドを所望する。
髪の毛の分目と目の細かく部分
目の形をもっと綺麗にしたい。
塗りつぶすところを塗りつぶす
前髪の位置を下げたり、目の細かい線を描く。細かい線が描きづらかったのでもう少しペンの太さが細いものもできると嬉しい
目の場所を正確に描くといいと思った
グリッドの調整機能を存分に活用する
グリッドを大きくして、顔の輪郭を書いたらうまくいくかもしれないと思った

表 5.10: 設問9: 感想 (支援あり)

描いている時はこれであっているか不安だったが、絵をぱっと見ると、自分がこんなに上手な絵を描けているとは思わなかったのが嬉しかった。
最小のグリッド数が割りとちょうどよいサイズだったが、もう少し細かくなるとより良く模写できるかも。
面白い 長い時間書きますけどつまらないの気がないです
グリッドがあると描きやすく感じつつも、グリッドの線が乗っているせいでその部分にちゃんと線が描かれているのかわかりにくくなっていたかと思った。
自分には模写の才能があるかもしれない
グリッドがあるものは時々見るのですが、グリッドが変わることで線の位置が違うことがよくわかり、かなり描きやすかったです。
普段の自分では考えられないくらいちゃんと絵になっていた。
元の絵を意識したいときは細かいグリッドが選べて、少し自分らしさを出したい場合は粗いグリッドを選べる。 全体的には元の絵に囚われることもなく自分が満足できる絵を出力できた。
久しぶりに絵を描いて楽しかった。グリッドを使って模写したのは初めてだったが、描きやすいと感じた。

5.2.1 模写後のアンケート結果 (支援あり)

支援ありの場合のみに行ったアンケート結果を示す。5段階評価である設問10,12,14の結果を図5.5で示す。

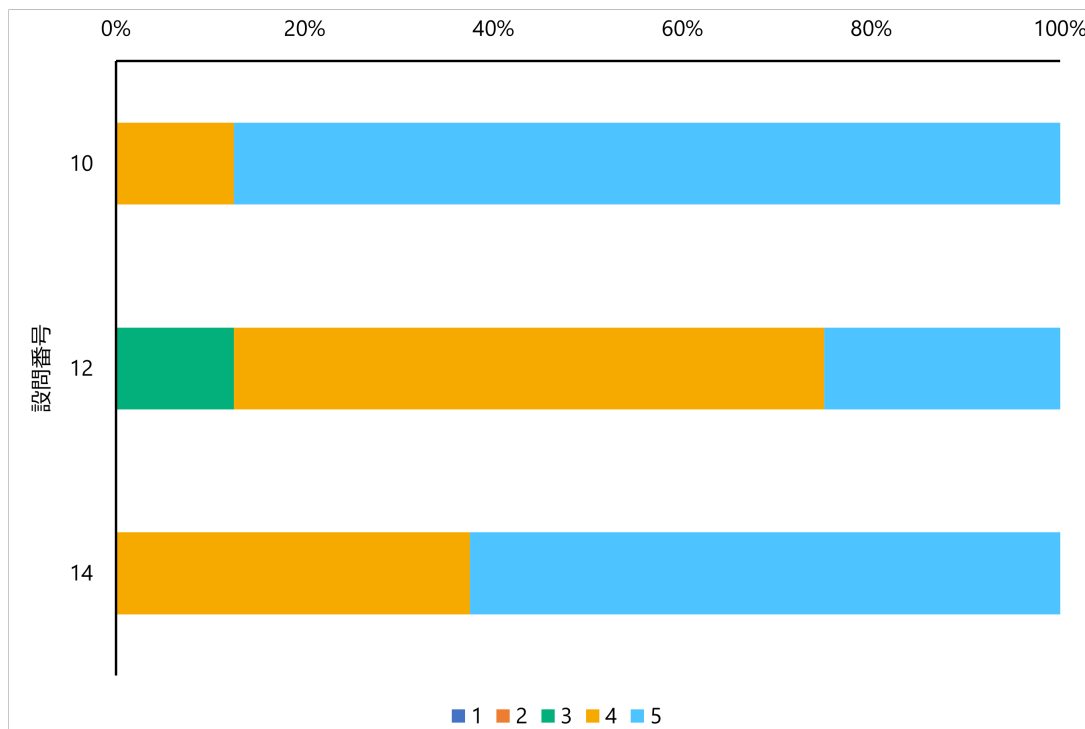


図 5.5: アンケート結果 (設問 10,12,14)

また、設問 10,12,14 において評価の平均値はそれぞれ 4.9, 4.1, 4.6 であった。次に、記述回答である設問 11,13,15 のコメントを表 5.11~5.13 に示す。

表 5.11: 設問 11: 設問 10 の理由

「この線どこ？」みたいなときに細かいグリッドだとどこにあるのかがわかりやすくなったから。
グリッドが大きすぎる場合に細かく変化することで調整しやすかったため
局所の線の配置がわからないとき多いのですごく助かります。
はじめに大まかに位置を書き、あとで詳細を書きたかったから。
顔の全体のバランスを整えるためには大きなグリッドを、細かいパーツ(目とか)のバランスを整えるためには小さいグリッドを利用するといいいことを教えてもらってから意識してみたら、とてもやりやすかったため。
大きめのグリッドで輪郭を大まかにとって、細かいグリッドで目などのパーツを書くことで書きやすくなったため
大まかに模写した後の微調整が容易に行えるため
大まかに書きたいところと細かく書きたいところを自由に設定できるから

表 5.12: 設問 13: 設問 12 の理由

描きたい女の子や可愛い女の子が出てきたから。
選択肢から自分の描きたいイラストを探すことができた
ぱっと書きたい絵を思い浮かばないので、推奨イラストは参考になれます。
推薦してもらうことで、たくさんある中から選ぶ手間が省けたから。
欲しい顔イラストの特徴が反映できていたりできていなかったりしていたため。
ある程度思い描いていた女性が出てきたため
簡単な絵から参考になるイラストが生成されたため
書きたい絵が大体出てきてよかった。

表 5.13: 設問 15: 設問 14 の理由

模写をするときには必要だなあと思ったから。
模写の楽しさおよび、イラスト力向上に役立つと感じた。
絵の練習に有用だと思います。
いつもより上手に漫画家のように絵が描けたから。
グリッドの存在を知らなかったので、グリッドがとても便利だと感じたため。
模写が好きなのわけではないが、模写をする必要があるなら便利なので使いたい
絵を描くことに慣れていなくてもある程度描けるため
グリッドを自分で作るのはめんどくさいし、一度作ってしまったグリッドは 変えられず、「もっとグリッド細かく/粗くすればよかった～」とならないため。

5.3 アンケート結果のまとめ

アンケート評価の結果、模写過程においてグリッドを可変にすることは設問 10 の平均値が 4.8 であることや、設問 11 のコメントから高評価であることが確かめられた。また、顔イラストの模写において特に目や顎、髪が難しいことがわかった。目は細かく描写されていることが多く、加えてイラストの印象を大きく左右するため、特に時間をかける被験者が多いことが実験を観察していると分かった。髪や顎は逆に描写が粗いため、周囲の線の位置などから空間を把握して描写する必要がある。描写が粗い箇所に関しては、特に細かなグリッドを使用しないと初心者にとっては模写が難しいと考えられる。また、初めから細かなグリッドを利用するとグリッド同士の対応関係を間違えるといった意見もあった。グリッドが粗いと大域的な意識ができ、グリッドが細かいと局所的に意識を向けることができるという意見を踏まえると、グリッドを粗いものから細かいものにすることによって、イラスト描画の基本の流れである大域的に描くこと(ラフ)から局所的な箇所を仕上げる(線画)を自然に身に付けられるのではないかと考えられる。また、模写過程の初めの方でグリッドを細かくして、そのまま最後まで描き切る被験者も多かったため、グリッドの細かさはユーザの描画具合を読み取ることで、システム内で自動的に行う必要を感じた。

グリッドを用いた模写自体に関しては、支援あり・無しのどちらに関しても設問 3 の平均値が 4 であることや、設問 9 のコメントに上手なイラストが描けたと意見があることから、上手く描けたと評価されている。一方で、グリッドとイラストが重なっている箇所が書きづらいとの意見もあった。これに対しては、グリッドの色をイラストと同じ黒色から変化させることで対処できると考えられる。

また、インタフェースの問題として細い線が描きづらい(ペンのサイズが太い)ことや線の太さの扱いに困るという意見もあった。これに関しては、ペンのサイズを調整するスライダとペンのサイズの対応関係がよくわからなかったのではないかと考えられる。スライダを調整する際に、調整したペンのサイズを視覚的に把握できるようにすることで解決できると考えられる。

5.4 System Usability Scale

SUSに関する結果を図5.6に示す。SUSの総和平均の結果は、最小値が67.5、最大値は90、平均は78.25であった。[30]では、SUSスコアと形容詞評価(Adjective ratings; Best Imaginable, Excellent, Good, OK, Poor, Worst Imaginableの6段階の評価)の間に有意な正の相関が確認されている。[30]によると、本システムのSUSスコアである78.25は 71.4 ± 11.6 の区分に含まれており、形容詞評価で「Good」に該当し、提案インタフェースが良好な評価を持つことが確認できた。

SUSの各設問結果は、全体の平均と比較して、設問1、設問6、設問7、設問10が高評価、設問2と設問5が低評価となった。設問1、設問6、設問7、設問10はそれぞれシステムの使用意欲、一貫性、理解の簡単さ、事前知識の不必要さを質問している。これらが高評価であることは、提案システムが使用しやすく、かつ使用したいという評価を得ていると考えられる。設問2はシステムの複雑さ、設問5はシステムの統一感を表している。これらが低評価であることは、ユーザの操作する機能が多いことが考えられる。特に、グリッドの操作はスライダを動かすだけでなく決定ボタンを押す必要があるため、煩わしかったのではないかと考える。また、決定ボタンは推薦された参考イラストの決定とグリッド操作の決定のどちらの操作にも用いるため、複数の機能を持つことでまとまりがなく、統一感が無いと判断されたのではないかと考える。

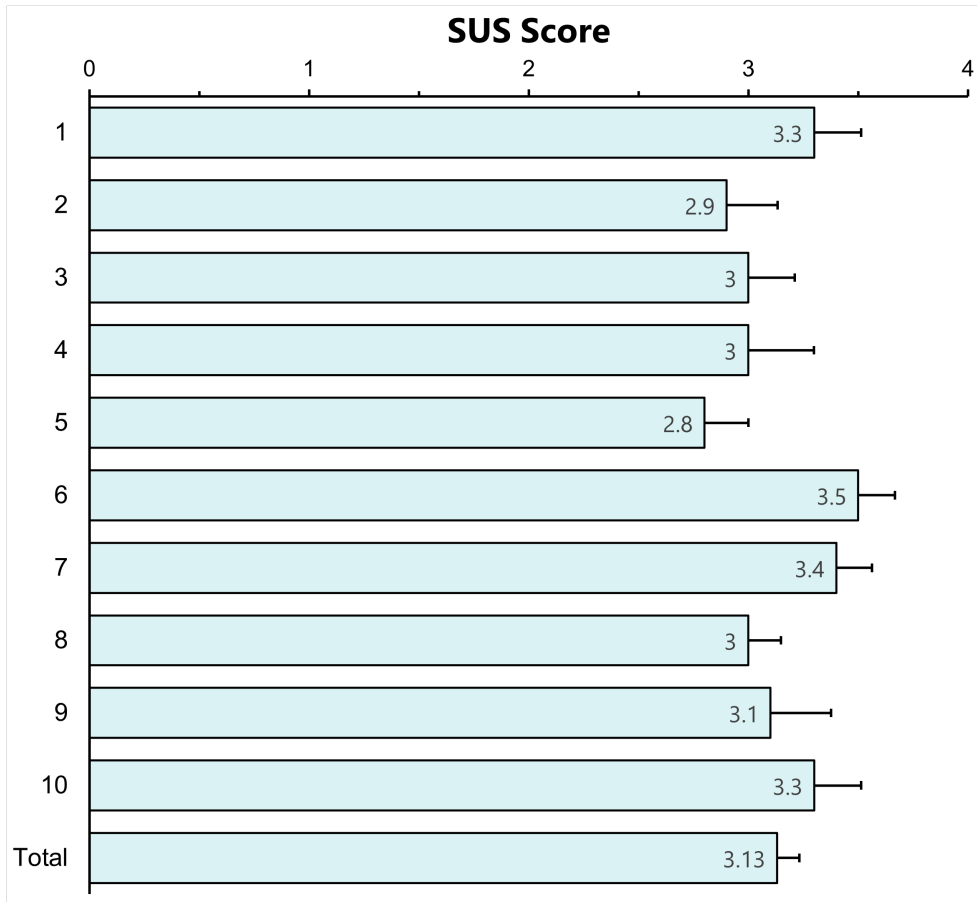


図 5.6: SUS 結果

5.5 NASA-TLX

TLXに関する結果を図5.7に示す。アナログスケールと重みの積に対して、特に評価の大きいものは知的・知覚的要求と努力、特に評価の小さいものはタイムプレッシャーとなった。

ユーザスタディは、提案システムを用いて模写を制作することであり、参考イラストの観察と完成までの努力が必要不可欠である。そのため、知的・知覚的要求と努力の評価が大きくなることは自明である。タイムプレッシャーについては、特に制限時間を設けなかったため評価が小さくなったと考えられる。

また、重みにおいては作業成績についても大きく評価されていた。つまり、提案システムを使用する際には達成度や満足度が重要であるが、それによる負荷はあまり発生しないことを示唆していると考えられる。

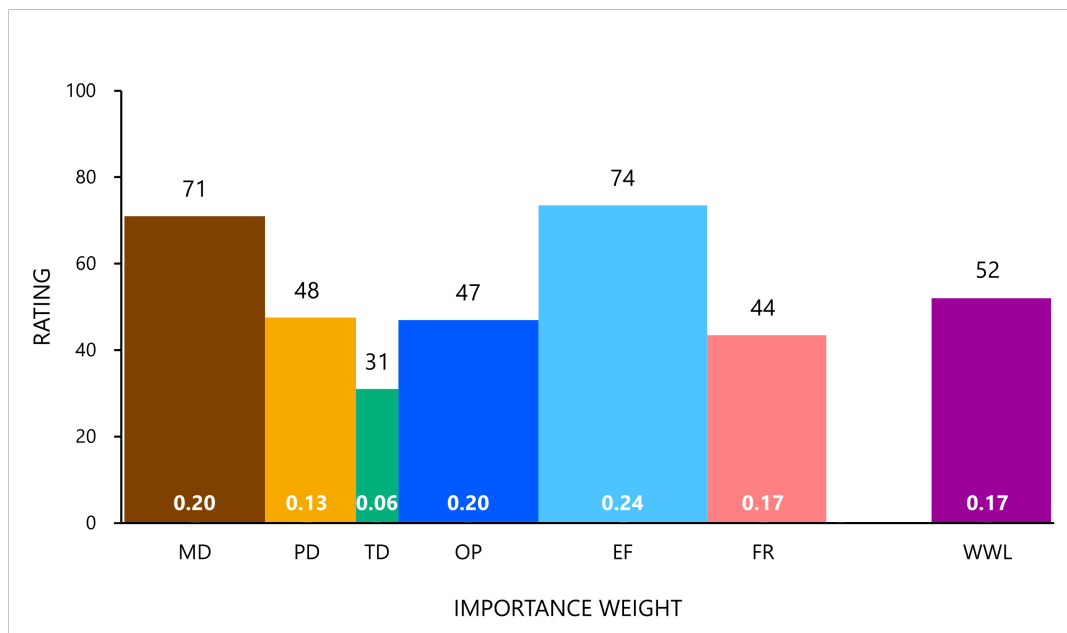


図 5.7: NASA-TLX 結果 (MD: 知的・知覚的要求, PD: 身体的要求, TD: タイムプレッシャー, OP: 作業成績, EF: 努力, FR: フラストレーション, WWL: 全体的な評価)

第6章 おわりに

本章では、本研究のまとめと今後の課題について述べる。

6.1 まとめ

本研究では、模写で用いるグリッドを可変にすることで、観察に焦点を当てた効果的な模写支援を可能にするという仮説に基づいて、ラフスケッチを入力とした参考イラストの推薦と、模写における可変グリッドシステムを組み合わせた模写練習支援のフレームワークを提案した。本システムでは、ユーザの描いたラフスケッチを入力とし、データセットから特徴量の近い類似イラストを候補として提示する。ユーザはそこから好ましいイラストを選び、グリッドの細かさを変化できる模写インターフェースにより模写を行う。ユーザスタディを行った結果、グリッドを可変にすることで効果的な模写を可能にすることが確かめられた。また、インターフェースの評価を行ない、提案インターフェースの有用性を確認できた。

6.2 今後の課題

本研究で提案したインターフェースは、ユーザが全て手動で操作する必要がある。ユーザスタディからも、ユーザの負担が低くはないことが示されている。そのため、ユーザが行う動作をある程度自動化することを考えている。例えば、グリッドをユーザ側のキャンバスと参考イラストで同じタイミングで変化するように実装を行うことや、ユーザの描画具合を読み取ることで、適切なグリッドを設定することなどが考えられる。また、カラーイラストからデータセットを作成したことを活かし、モノクロだけでなくカラーイラストの模写にも適応させていきたい。

謝辞

本研究を遂行するにあたり、多くの方々にご支援頂けたことに深く感謝申し上げます。

主指導教員としてご指導とご鞭撻、温かい激励を賜りました宮田一乗教授に心より感謝申し上げます。研究だけでなく、生活や進路に至るまで多面的な角度から多くの意見を頂戴し、暖かく見守っていただきました。また、先生方との議論を通じて、研究課題の目的や意義などを客観的に捉えることができ、研究テーマの方針を固めることが出来ました。感謝の念に堪えません。ありがとうございました。

また、謝浩然講師におかれましては、日頃から研究の進捗を気にかけてくださり、研究テーマの方向性など様々な意見をくださいました。研究に苦戦していた時、テーマを変えてみてはどうかと助言くださったあの日は忘れられません。ここまで来れました。心より感謝申し上げます。

副テーマでご指導いただいた日高昇平准教授に感謝申し上げます。初めての国際会議の発表や小学生に向けたワークショップなど、貴重な経験ができました。特に国際会議では、不得意な英語で論文を執筆・発表にあたり色々ご助力をいただきました。心より感謝申し上げます。

最後に、研究活動や娯楽、食事といったプライベートを共に過ごした宮田研究室内外の同期、先輩、後輩の皆様、多くの友人達に心より感謝申し上げます。様々な知識や技術、娯楽の共有は大学院で生活する上でとても助けになりました。また、私にはない価値観や考え方に触れられたことは、この先にも生きてくると思います。本当にありがとうございました。

参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc., 2014.
- [2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, Vol. 33, pp. 6840–6851. Curran Associates, Inc., 2020.
- [3] Zhengyu Huang, Yichen Peng, Tomohiro Hibino, et al. Dualface: Two-stage drawing guidance for freehand portrait sketching. *Computational Visual Media*, Vol. 8, pp. 63–77, 03 2022.
- [4] 渡邊優, 阿倍博信. pix2pix を用いたデジタルイラスト制作における自動レイヤ分けシステム. 情報処理学会 研究報告コンシューマ・デバイス&システム (CDS) , Vol. 2020-CDS-27, No. 27, p. 1–8, Jan 2020.
- [5] Hung-Yu Tseng, Matthew Fisher, Jingwan Lu, Yijun Li, Vladimir Kim, and Ming-Hsuan Yang. Modeling artistic workflows for image generation and editing. In *Computer Vision – ECCV 2020*, pp. 158–174. Springer International Publishing, 2020.
- [6] Emmanuel Iarussi, Adrien Bousseau, and Theophanis Tsandilas. The drawing assistant: Automated drawing guidance and feedback from photographs. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, p. 183–192. Association for Computing Machinery, 2013.
- [7] lbpcascade_animeface: A face detector for anime/manga using OpenCV. https://github.com/nagadomi/lbpcascade_animeface.
- [8] X. Xiang, D. Liu, X. Yang, et al. Anime2sketch: A sketch extractor for anime arts with deep networks. <https://github.com/Mukosame/Anime2Sketch>, 2021.

- [9] X. Xiang, D. Liu, X. Yang, et al. Adversarial open domain adaptation for sketch-to-photo synthesis. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 944–954. IEEE Computer Society, jan 2022.
- [10] 庄野逸. 局所画像特徴量～ sift, hog を題材に～. *映像情報メディア学会誌*, Vol. 67, No. 3, pp. 256–258, 2013.
- [11] Shuai Yang, Zhangyang Wang, Jiaying Liu, and Zongming Guo. Deep plastic surgery: Robust and controllable image editing with Human-Drawn sketches. January 2020.
- [12] wacom. 画材・文具に対する意識調査.
<https://tablet.wacom.co.jp/article/regarding-art-supplies-and-stationery-awareness-survey>, 2021.
- [13] 中島 楓華, 山西 良典, 巽 優人, 他. アニメキャラの顔パーツの位置バランスとキャラクタ属性の関係性に関する基礎検討. 第6回コミック工学研究会予稿集, pp. 24–28, 2021.
- [14] 大芝 潤気, 岩田 基, 黄瀬 浩一. C2GAN を用いたランドマークに基づくアニメキャラクタの顔画像自動生成手法. 第3回コミック工学研究会予稿集, pp. 57–62, 2020.
- [15] 小笠原 渚, 清 雄一, 田原 康之, 大須賀 昭彦. アニメキャラクター風イラストの全身画像生成手法の提案. *電子情報通信学会技術研究報告*, Vol. 120, No. 379, pp. 30–35, 2021.
- [16] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.
- [18] Robin Rombach, Andreas Blattmann, Dominik Lorenz, 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.
- [19] Yong Jae Lee, C Lawrence Zitnick, and Michael F Cohen. ShadowDraw: Real-time user guidance for freehand drawing. *ACM Transactions on Graphics (TOG)*, Vol. 30, No. 4, pp. 1–10, 2011.

- [20] 菅野一平, 中村聡史. 個人のイラスト制作における観察に対する支援手法の検討. 第 2022-DCC-30 巻, pp. 1–8, 2022.
- [21] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018.
- [22] waifu2x-ncnn-vulkan: waifu2x converter ncnn version, runs fast on intel / amd / nvidia / apple-silicon GPU with vulkan. <https://github.com/nihui/waifu2x-ncnn-vulkan>.
- [23] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp. 62–66, 1979.
- [24] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a "siamese" time delay neural network. NIPS'93. Morgan Kaufmann Publishers Inc., 1993.
- [25] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [26] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, pp. 886–893 vol. 1, 2005.
- [27] John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, Vol. 189, No. 194, pp. 4–7, 1996.
- [28] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human mental workload*, Vol. 1, No. 3, pp. 139–183, 1988.
- [29] 芳賀繁, 水上直樹. 日本語版 nasa-tlx によるメンタルワークロード測定各種室内実験課題の困難度に対するワークロード得点の感度. *人間工学*, Vol. 32, No. 2, pp. 71–79, 1996.
- [30] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, Vol. 4, No. 3, p. 114–123, may 2009.