インタラクティブ機械学習の教示プロセスでの生成AIの利用可能性の検討

山本 恒輔[1]　周 中一[1]　ゼファン シュラーメク[1]　矢谷 浩司[1]

概要：
機械学習は大量の学習データがあれば難しい課題を効率的に解決できるが，モデル作成には深い理解が必要であり，一般の人々が利用するにはハードルが高い．インタラクティブ機械学習（IMT）は機械学習のモデル作成に対話的要素を組み込み，ユーザの理解を促進することを目的としており，すでにウェブカメラなどを使って簡単にデータを提供できるシステムが公開されている．しかし，これらの IMT ツールではユーザが提供する学習データと実際に学習させたい概念との乖離が生じることがあり，本研究ではその乖離を生成 AI によるデータの拡張によって解決することを目指す．本稿では，生成 AI を組み込んだ IMT システムの開発と実験を行い，IMT の教示プロセスにおける生成 AI の利用可能性を検討した．

Examining the Feasibility of Integration of Generative AI into the Teaching Process in Interactive Machine Teaching

Kosuke YAMAMOTO[1]　Zhongyi ZHOU[1]　Zefan SRAMEK[1]　Koji YATANI[1]

1. はじめに

機械学習は，既存のアルゴリズムでは効率的に解くことが難しい課題を十分な学習データがあれば簡潔に解決できるという特徴を持っており，昨今の技術発展もあり急速に人々に使われ始めている．しかし，機械学習技術が社会に普及し多くの人々が利用し始めた一方で，学習モデルを作成するためには機械学習技術への深い理解が必要であり，誰もがモデルを自由に作成し活用できるという状況にはなっていない．このような状況において，過去にプログラミング分野において関数型言語やバージョン管理システムの研究が進んだように，最近では機械学習の分野においてもモデルの精度向上だけではなく生産性向上やユーザの機械学習モデルの理解を促進するような分野の研究が進んでいる [1].

こうした研究分野の一つにインタラクティブ機械学習（IMT: Interactive Machine Teaching）がある．これは機械学習のモデル作成に対話的要素を組み込むことで，簡単にモデルを作成することを可能にしたり，ユーザの技術への理解を促したりすることを目的としている．IMT ツールは機械学習の非専門家が使用することを想定して作成されており，特にユーザがウェブカメラなどを利用して簡単にデータを提供できるようにしているものを Vision-based IMT（V-IMT）と呼ぶ．V-IMT の代表例としては Teachable Machine [2] や Lobe.ai [3] などがあり，これらのツールではウェブカメラに数種類の物体をかざして写真を撮りラベリングをするだけで，物体の分類モデルを簡単に作成することが出来る．

これらの V-IMT ツールを用いるとユーザは簡単に分類モデルを作成することができる一方で，ユーザが学習に利用する学習データは一般的にはユーザがモデル作成時に所有しているものに限られることが多く，ユーザが本当に学習させたい概念と提供できるデータ内容に乖離が起こることがある．例えば，ユーザがリンゴと洋ナシを分類するモデルを作成する場合，ユーザは赤いリンゴと緑色の洋ナシを学習データとして分類モデルを学習させる可能性がある．ユーザは提供したデータで学習モデルがリンゴと洋ナシを十分判別できると考えるかもしれないが，その学習モデルは実は色だけで判別を行っており，青リンゴを入力す

[1] 東京大学 Interactive Intelligent Systems Laboratory

ると洋ナシと判別してしまうというケースが考えられる．そこで本研究では，ユーザが V-IMT ツールに入力した画像データをもとに，生成 AI を用いてユーザが提示した物体の類似画像を生成して学習データの拡張を行う．そのために生成 AI を組み込んだ IMT システムを開発し実験することで，生成 AI の利用可能性を検討する．

2. 関連研究

2.1 インタラクティブ機械学習

インタラクティブ機械学習（IMT: Interactive Machine Teaching）は，ML 分野と HCI 分野で異なる使われ方をしている用語である．ML 分野では主に，人間がモデルの学習ループに介入すること (Human in the loop) で人間の知識を取り込み，モデルの学習精度を上昇させることを目的としており，人間とシステムのどちらが主体性を持つかに応じて，AL(Active learning) や IML（Interactive machine learning）といった用語が使い分けられている [4]．一方 HCI 分野では，ML 分野同様人間が学習ループに介入するものの，ユーザがモデルを作成することや，モデルの作成を通して知識を得ることなどが主目的とされており，近年の機械学習技術の発展に伴い，機械学習を利用する人が増えたものの機械学習モデルを作成できる人は少ないという状況を解決するための手法としての意味も持つ．なお本稿では，Ramos ら [5] が定めた HCI 分野での IMT の定義を利用することとする[*1]．

IMT ツールは一般的に，機械学習分野に詳しくない非専門家が特別な知識や技術を持たずとも機械学習モデルを作成できることを目的に設計されている [6]．例えば，Zho ら [7] は，機械学習における特徴量選択に着目し，特徴量選択とモデルの評価をインタラクティブに支援するツール FeatureExplorer を作成した．ユーザは FeatureExplorer を使うことで交換可能な特徴量やそれぞれの特徴量の重要性などを理解可能であり，自身のドメイン知識を適用することで，モデルを効率的に改良・診断することができる．また，Kulesza ら [8] は，機械学習の分類タスクに置いて必要不可欠なラベリングを支援するために，構造化ラベリングという一貫したラベリングを支援する手法を提案している．機械学習のラベリングは簡単に思えるが，ラベル対象の変化や抽象的概念の変化によって一貫性がなくなることで，過去のラベリング内容がモデルの学習にとって不利に働くことがある．そこで Kuleasza らは，yes, could be, no の 3 種類に分類すると同時に，その理由などでグループを作成することで，ラベリングの判断基準をいつでも再確認できるようにし，ラベリングの一貫性を向上させている．

[*1] IMT is “an IML (interactive machine learning) process in which the human-in-the-loop takes the role of a teacher, and their goal is to create a machine-learned model”

2.2 インタラクティブ機械学習の設計

既存の研究のいくつかは，IMT ツールを用いて ML 初心者を対象にした定性的な研究を行い，IMT ツールにおけるユーザの要求や設計のガイドラインを提示している [9,10]．Yang ら [11] の研究では，ML の初心者はモデルの精度だけを唯一の指標と認識しており，ML モデルはより簡単かつロバストなものを目指すべきだと提案している．Fiebrink ら [12] の研究では，交差検証やモデルの直接評価によって，ユーザの行動に対するフィードバックを行い，機械学習の経験や知識を持たないユーザでもモデルの改善を促せることを示した．ユーザはツールからのフィードバックをもとに，作成が困難だと判明した当初の目的を変化させてモデルを作成したり，モデルの予想外の動きを柔軟に取り入れたモデルを作成したりすることもあった．

IMT ツールの研究は画像を用いた分類問題を解くものが多く，Carney ら [2] の Teachable Machine や Microsoft が発表した Lobe.ai [3] などが有名である．Teachable Machine では，ユーザが簡単に画像分類を行うモデルを作成するための GUI を提案しており，ツールはラベリングされたデータ入力パネル，学習用パネル，プレビューパネルの 3 つから構成されている．Lobe.ai でも同様に分類モデルの作成は 3 つのステップで構成されており，ラベリングとデータ入力・学習・利用の 3 つである．Lobe.ai では学習時に失敗したデータ例を提示する機能があり，モデル改善のために追加する必要があるデータがわかりやすいという特徴を持つ．

2.3 学習データの情報量を増加させるツール

ユーザが V-IMT ツールを使用する際，ツールがデモプレイ等で出力する推定精度は，必ずしも学習モデルがユーザの示した画像や意図した特徴を理解していることに直結しないという点に注意する必要がある．Zhou と Yatani [13] の研究では，V-IMT ツールの評価フェーズにおいて Saliency Map という，画像内の注目度が高い部分を示す仕組みを導入することで，学習モデルが間違った特徴をもとに学習していることの検知を可能にしている．こうした学習データに含まれるユーザが意図しない特徴での学習を防ぐためには，入力画像データのアノテーションを提供する必要がある．画像データのどの部分が学習データとしての特徴を持つかを示すためにはセグメンテーションマスクを利用することが一般的であり，セグメンテーションマスクは画像内の物体の輪郭をなぞったり，CV の技術を使うことでクリックやドラッグ等で作成が可能である [14,15]．一方で，これらの技術は IMT ツールに最適化されたものではなく，学習データの入力フェーズに置いてインタラクティブ性を著しく損なうものであった．そこで Zhou ら [16] は，ユーザがカメラに向かってデモを行う際

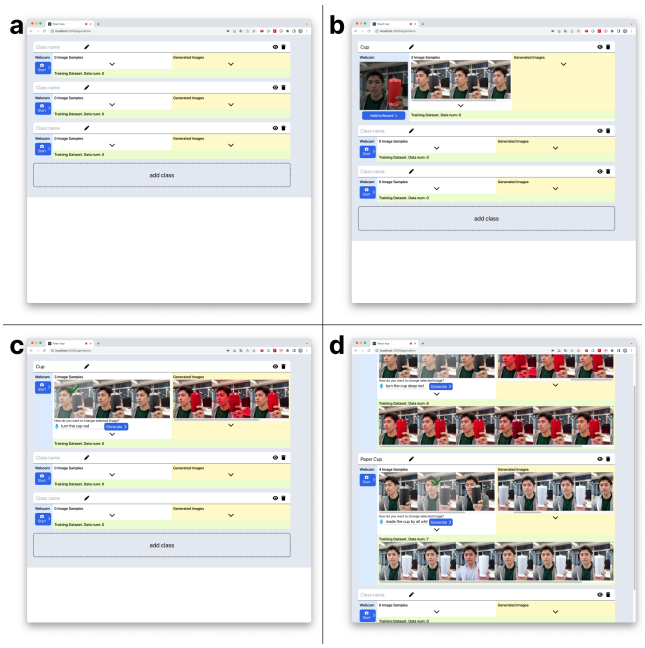

図 1　作成したシステムの UI.

に自然と行うハンドジェスチャーを利用して，画像内の物体のセグメントを自動で行うアルゴリズムを開発した．

Zhou らのハンドジェスチャーによる自動セグメンテーション機能を備えた V-IMT ツール LookHere は，画像内の対象物体のセグメントには成功した一方で，学習データがユーザの所有しているものに偏るという課題は残っており，ユーザが意図した概念を学習モデルに正確に伝えられているとはいえない．

2.4　生成 AI を用いた学習データの拡張

Stable Diffusion [17] や DALL-E [18] など，画像生成を行う大規模拡散モデルは急激に進化しており，実際の写真と見間違うような写真の生成が可能になってきている．そこで Azizi ら [19] は，画像生成モデルで生成したデータが実際に画像分類タスクの改善に役立つかどうかを研究した．その結果，大規模拡散モデルを fine tuning して作成した画像分類モデルは分類精度で SoTA を記録したことに加え，ImageNet の学習データに拡散モデルで生成したデータを加えることで，分類精度が大幅に改善されることを示した．

これらのことから，機械学習モデルを作成する際に画像生成 AI を用いて学習データを拡張すると，モデルの判別精度が改善されることが期待でき，本稿では実際に生成 AI を組みこんだ IMT の作成を目指す．

3.　生成 AI を組み込んだ IMT システム

3.1　UI と作業手順

提案する V-IMT ツールの UI を図 1 に示す．基本的な UI は Teachable Machine [2] を参考にしている．この UI では，分類するクラスの数だけ写真群を作成するための領域があり，ラベル名の設定（領域上部），画像の入力（青背景），入力画像の保存（白背景），類似画像の生成結果（黄色背景），データセットとして利用する画像（緑背景）という役割をそれぞれ持つ．

このツールの作業手順は以下のようになる．(1) まずユーザは分類対象のラベル名を設定し，各ラベルにおいて分類対象の物体をウェブカメラに提示することで類似画像の生成に用いるデータを入力する（図 1（a），(b)）．(2) 入力されたデータをもとに生成 AI を用いて画像の生成をインタラクティブに行い，データセットとして利用するに値する画像を選択する．（図 1（c））．(3) 分類対象のクラスの数だけ（1）-（2）の操作を繰り返して学習に用いるデータセットを作成する（図 1（d））．

ユーザは手順（2）において生成 AI を用いて類似画像の生成を行うが，その際の入力としてはプロンプトと画像変化率の二つを用いる．プロンプトとは Stable Diffusion などの画像生成モデルにおいて，画像をどのように変化させるかを指定するための文章であり英語で入力を行う．画像変化率とは，元の画像からどれだけ画像を変化させるかを示す値であり，画像生成分野で CFG（classifier-free guidance）と呼ばれる値を調整している．プロンプトや画像変化率を設定した後，画像生成のボタンを押すことでサーバーで画像の生成が始まり，数秒後画像生成が完了するとブラウザ側で生成結果を確認できる．画像は一度に 4 枚生成される．

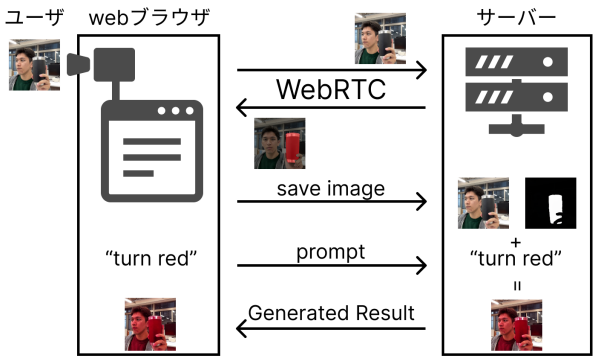

図 2　作成したシステムのアーキテクチャ

3.2　システムアーキテクチャ

提案するシステムの通信や画像のやり取りについて図 2 に示した．ユーザがウェブカメラを通じて入力した画像データは，ブラウザとサーバー間で接続された WebRTC のパイプラインを通して一度サーバーへ送られ，ハンドジェスチャーによるセグメンテーションがなされた図と合成された元画像データがブラウザで表示される．ユーザはブラウザ上で自分の入力に対してセグメンテーションマス

クがどの部分に生成されているかを常に確認でき，Recordボタンを押すと，その瞬間の入力画像と対応するマスク画像がサーバーで保存される．画像生成時には，ユーザは選択した画像生成に用いる画像，画像の編集内容であるプロンプト，画像変化率の値をサーバーに送信する．サーバーでは受け取った情報を元に画像を生成し，生成が完了したらブラウザへ生成結果を送り返す．

4. ユーザ実験・結果

4.1 実験概要

東京大学の学生 2 人を対象に，IMT へ組み込む生成 AI の利用可能性を明らかにするため，3 章で作成したシステムを用いてユーザ実験を行った．IMT への習熟度の差が結果に影響することを避けるため，IMT への知識がないことを条件に，参加者を募集した．ユーザ実験では，はじめに Teachable Machine の説明動画を視聴してもらい，IMT の概念を理解してもらったあと，タスク内容を説明し，実際にシステムを利用してもらった．実験の目的は，通常のバナナのみが手元にある状態で，普通のバナナ，腐ったバナナ，熟していないバナナの 3 種類を判別できる画像分類モデルを作成するのに必要なデータセットの作成とした．

4.2 実験結果・考察

本節では，ユーザ実験を通して得られたインタビュー結果と，実験中のユーザ操作に関する考察を記す．

画像生成の困難さ

実験参加者全員がシステムの利用において最も困難な点に上げたことが，この画像生成の困難さであった．”画像を生成するにあたって，ユーザにとって未知の概念であるプロンプトと画像変化率という 2 つの変数を最適化する必要があり手間取った”という意見や，”目的の画像が生成出来なかった場合，次にどのようなことをしたら良いか分からない”という意見が得られた．

学習に必要な画像枚数

2 人の実験参加者は，約 20 分程度の所要時間でシステムを用いてデータセットを作成したが，そのデータセットに含まれる画像枚数は共に 10 枚以下であった．インタビューを通しても，”学習に必要な画像枚数が不明であり，自分の満足のいく画像を数枚作成できた段階で十分だと考えた”という意見が得られた．

これに関しては，本システムの画像生成における困難さが大きく影響していると考えられる．ウェブカメラを用いて多くの画像を一度に手軽に入力出来る既存の IMT と異なり，本システムでは目的の画像を 1 枚生成するだけ 5-10 分程度の時間を必要とし，必要に駆られなければユーザは画像を大量に作成しないと考えられる．

画像生成の所要時間

実験終了後に，画像生成の所要時間に関して尋ねた所，2 人の実験参加者は共に”画像生成にかかる時間が長い”と回答した．一方で，画像生成の高速化に言及することは無かった．その理由としては，”画像生成は本来時間がかかるもので，高速化出来るという視点がなかった”ことや，”実験を行っていくうちに，生成のスピードに慣れてしまった”という回答が得られた．

UI デザインについて

実験を通して，”画面内に操作可能な項目がたくさんありすぎるせいで操作に戸惑った”．”次に行うべき操作をもっと明示して欲しい”．という声が聞かれた．これに関しては，本システムの作り込みの不十分さも一因として考えらるが，IMT と生成 AI というユーザにとって不慣れな概念を 2 つ同時に提案していることが理由として挙げられ，ユーザにとってわかりやすい誘導が必要である．

5. 考察

この章では，4 章で示したシステムを実際に利用することで判明した課題において，提案システムのインタラクティブ性に特に大きな影響を及ぼすと予想される，以下の 2 つの課題に対して議論する．

5.1 画像生成の困難さ

1 つ目の課題として，システムに入力した画像をもとに，生成 AI を用いて理想的な画像を生成することが難しいというものがある．その要因としては，画像生成に利用するモデルの特徴と，生成内容を示すプロンプト，画像の変化率を調整するパラメータなど，ユーザに取って不慣れな概念かつ生成される画像への影響が大きい項目があることがあげられる．

まず生成 AI で利用するモデルについて，本稿では，Inpaint[*2] と Instructpix2pix[*3] の 2 種類のモデルを利用した．

Inpaint 用のモデルでは，ユーザが入力画像に対してマスクを設定可能であり，画像の 1 部だけを編集しやすいという特徴がある (図 3)．一方で，画像生成時には，マスク領域の元の画像を参考に生成を行うわけではなく，プロンプトを元に新たに画像を生成するため，ユーザが入力とした画像からかけ離れた画像が生成されやすい．また，マスク領域を設定した場合はマスク外領域の変更が出来ないため，原理上元の画像からサイズが変わるような物体の生成が出来ない (図 3(c))．

Instructpix2pix は，画像全体の編集を可能とすることを目的に作成されたモデルであり，入力画像全体を元画像を保って変化させるような画像生成が得意である．しかし

[*2] https://huggingface.co/stabilityai/stable-diffusion-2-Inpainting
[*3] https://huggingface.co/timbrooks/instruct-pix2pix

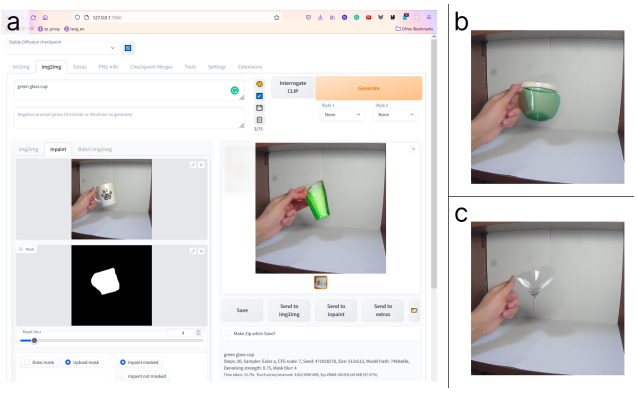

図 3　マスクを用いた Inpaint 機能での画像生成結果．(a) では変更前の画像とマスク，生成結果を示しており正しく画像が生成されている．(c) 一方で，マスク領域を超えるような物体の生成は難しいことがわかる．

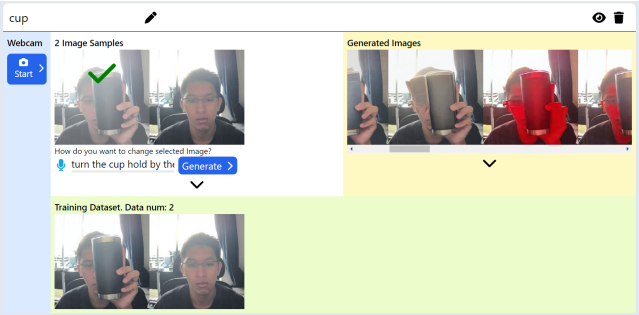

図 4　Instructpix2pix を用いた画像生成結果例．手に持ったコップだけでなく，人間の顔まで赤く変化してしまっている．

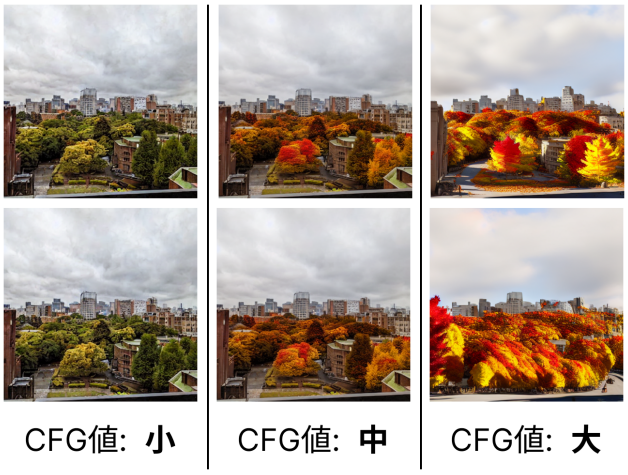

図 5　画像変化率（CFG 値）を変えた時の画像生成例．プロンプトとしては，”sunny day, fallen leaves, autumn”を設定した．値が大きくなるほど，画像が大きく変化していることがわかる．

Instructpix2pix では，マスクを利用した画像の再生成ができず，画像内の編集したい物体が，プロンプトによって明確に定まる場合は画像生成が比較的上手くいくが，プロンプトでの指定が不明瞭な場合，画像全体の色が変化する事例が多数確認された (図 4)．

プロンプトに関しては，生成を何度か繰り返すことで，

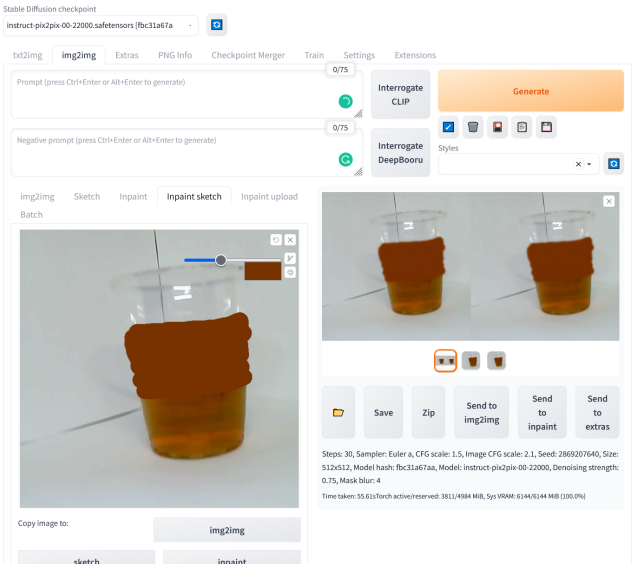

図 6　Inpaint sketch を利用した画像生成例．コップ内の麦茶の量を増やすにあたっては，色と量を sketch で指定することで適切な画像が生成できる．

内容の修正が比較的簡単に可能だったが，画像変化率を調整するパラメータの最適化には課題が残る．本稿で画像変化率と呼んでいるものは，生成モデル上では CFG と呼ばれるパラメータを意味しており，これは Classifier-Free Guidance の略である．

Inpaint モデルでは，prompt CFG と Image CFG の 2 つが，Instructpix2pix では prompt CFG だけが存在する．prompt CFG は，どれだけプロンプトの指示内容を反映するかの指標となり，概ね大きい程に画像が大きく変化する．Image CFG は入力画像忠実度と言い換えることかでき，大きいほどに入力画像に忠実，つまり画像変化量が少なくなる．そのため，ユーザが生成した画像において，変化量が不足している場合は prompt CFG を増やし，Image CFG を下げるという戦略を取ることができる．本稿ではこれ以降簡便のため，これら 2 つのパラメータの役割に着目し，画像変化率という変数名で話を進める．

実験で判明した大きな課題として，画像の変化率は，生成画像の変化が少ない場合，一概に変化率を上げればいい訳では無い，ということがある．画像の変化率を上げるということは，入力画像に大して多くの変化を許容することになり，特にマスクを設定していない場合，想定外の領域の変化を引き起こしかねない．しかし，画像生成時にマスクを利用した場合でも問題が発生することが分かっている．マスクを設定すると，設定しない場合に比べて，相対的に画像の変化量は大きくなり，その分，画像の変化率をあげる必要がある．ただ，画像の変化率をあげると，マスク領域内で入力画像を参照して欲しい部分まで変化するというジレンマが発生する．

このジレンマは，Inpaint Sketch という機能を使うことで

ある程度解消できることが分かっている．Inpaint Sketch では，ユーザが，画像の変化させたい領域をマスクすると同時に，どんな色に変化させるかを指定することができる．画像の編集内容をユーザが直接指定可能になるため，画像の変化率は比較的小さく収めることが可能になるが，Inpaint Sketch で必要となるマスクは入力画像と 1 対 1 対応するものであり，最終的にデータセットを作るために複数の入力画像から画像を生成する場合，同じ数だけマスクが必要となる．

5.2 画像生成にかかる時間

二つ目の課題として，画像生成に必要な時間の短縮があげられる．現状のプロトタイプでは，ユーザが提示した画像から類似した画像を生成するのに，通常のマシンだと 20 秒程度，GPU を比較的搭載したクラウドサーバーを用いても 5 秒程度が必要となる．これは，Stable Diffusion の画像生成方式が，ノイズが混ざった画像から少しづつノイズを減らしていくというステップ（イテレーション）を繰り返してより綺麗な画像を生成する方法であることに起因する．

生成画像はイテレーションを繰り返すほどにクオリティが向上していくが，Salimans ら [20] は生成時間を短縮する手法の研究をしており，それによると，モデルを改良することで比較的イテレーションが浅い段階でも高クオリティの画像を出力できることを示しており，実際に本稿でもイテレーションを短縮して画像を生成している．これは，本システムでは機械学習モデル用のデータセットを作成するという目的上，人間がある程度視認できる画質であれば問題ないという考えに基づいている．

一方で，GAN を用いた画像生成手法を採用することも考えられる．Kang ら [21] は Text-to-Image が可能な GAN を研究しており，0.13 s での画像生成が可能である．これは，GAN の画像生成手法が Stable Diffusion などで採用されている拡散モデルと異なることが理由である．

6. 結論

本稿では，既存の IMT における課題である，ユーザのメンタルモデル実データの乖離を書いけるするために，生成 AI を組み込んだ IMT を実際に開発した．開発したシステムを用いてユーザ実験を行った結果，画像の生成や生成速度に関する課題が存在し，特に画像生成においては，生成 AI に不慣れなユーザが自由に画像を生成するためには，生成に利用するモデルの選定やプロンプト，画像変化率のパラメータの最適化など多くの考慮すべき要素があることが明らかになった．また本稿では異なる画像生成モデルや画像生成手法を提示し，これらの課題の解決方針を示した．

謝辞

本研究を進めるにあたり，アドバイスをくださったすべての研究室メンバー，ユーザ実験に参加してくださった実験参加者に感謝申し上げます．

参考文献

[1] Simard, P. Y., Amershi, S., Chickering, D. M., Pelton, A. E., Ghorashi, S., Meek, C., Ramos, G., Suh, J., Verwey, J., Wang, M. et al.: Machine teaching: A new paradigm for building machine learning systems, *arXiv preprint arXiv:1707.06742* (2017).

[2] Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A. and Chen, A.: Teachable machine: Approachable Web-based tool for exploring machine learning classification, *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pp. 1–8 (2020).

[3] : Lobe — Machine Learning Made Easy, `https://www.lobe.ai/`.

[4] Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J. and Fernández-Leal, Á.: Human-in-the-loop machine learning: A state of the art, *Artificial Intelligence Review*, pp. 1–50 (2022).

[5] Ramos, G., Meek, C., Simard, P., Suh, J. and Ghorashi, S.: Interactive machine teaching: a human-centered approach to building machine-learned models, *Human–Computer Interaction*, Vol. 35, No. 5-6, pp. 413–451 (2020).

[6] Dudley, J. J. and Kristensson, P. O.: A review of user interface design for interactive machine learning, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 8, No. 2, pp. 1–37 (2018).

[7] Zhao, J., Karimzadeh, M., Masjedi, A., Wang, T., Zhang, X., Crawford, M. M. and Ebert, D. S.: FeatureExplorer: Interactive feature selection and exploration of regression models for hyperspectral images, *2019 IEEE Visualization Conference (VIS)*, IEEE, pp. 161–165 (2019).

[8] Kulesza, T., Amershi, S., Caruana, R., Fisher, D. and Charles, D.: Structured labeling for facilitating concept evolution in machine learning, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3075–3084 (2014).

[9] Sanchez, T., Caramiaux, B., Françoise, J., Bevilacqua, F. and Mackay, W. E.: How do people train a machine? Strategies and (Mis) Understandings, *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. CSCW1, pp. 1–26 (2021).

[10] Gil, Y., Honaker, J., Gupta, S., Ma, Y., D'Orazio, V., Garijo, D., Gadewar, S., Yang, Q. and Jahanshad, N.: Towards human-guided machine learning, *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 614–624 (2019).

[11] Yang, Q., Suh, J., Chen, N.-C. and Ramos, G.: Grounding interactive machine learning tool design in how non-experts actually build models, *Proceedings of the 2018 designing interactive systems conference*, pp. 573–584 (2018).

[12] Fiebrink, R., Cook, P. R. and Trueman, D.: Human model evaluation in interactive supervised learning, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 147–156 (2011).

[13] Zhou, Z. and Yatani, K.: Enhancing Model Assessment

in Vision-based Interactive Machine Teaching through Real-time Saliency Map Visualization, *The Adjunct Publication of the 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 112–114 (2021).

[14] Sofiiuk, K., Petrov, I., Barinova, O. and Konushin, A.: f-brs: Rethinking backpropagating refinement for interactive segmentation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8623–8632 (2020).

[15] Chang, C.-M., Lee, C.-H. and Igarashi, T.: Spatial labeling: leveraging spatial layout for improving label quality in non-expert image annotation, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12 (2021).

[16] Zhou, Z. and Yatani, K.: Gesture-aware Interactive Machine Teaching with In-situ Object Annotations, *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–14 (2022).

[17] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-resolution image synthesis with latent diffusion models, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695 (2022).

[18] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M. and Sutskever, I.: Zero-shot text-to-image generation, *International Conference on Machine Learning*, PMLR, pp. 8821–8831 (2021).

[19] Azizi, S., Kornblith, S., Saharia, C., Norouzi, M. and Fleet, D. J.: Synthetic Data from Diffusion Models Improves ImageNet Classification, *arXiv preprint arXiv:2304.08466* (2023).

[20] Salimans, T. and Ho, J.: Progressive distillation for fast sampling of diffusion models, *arXiv preprint arXiv:2202.00512* (2022).

[21] Kang, M., Zhu, J.-Y., Zhang, R., Park, J., Shechtman, E., Paris, S. and Park, T.: Scaling up gans for text-to-image synthesis, *arXiv preprint arXiv:2303.05511* (2023).