

令和 5 年 6 月 22 日現在

機関番号：32660

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K12115

研究課題名(和文) 技能伝承・ノウハウ共有のための人間参加型機械学習による映像組織化

研究課題名(英文) Organizing Videos by Human-in-the-Loop Machine Learning for Skill Transfer and Know-How Sharing

研究代表者

谷口 行信 (Taniguchi, Yukinobu)

東京理科大学・工学部情報工学科・教授

研究者番号：70759422

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究の目的は、未編集の作業映像を、画像認識により自動的に組織化(意味的なシーンに分割し、タグを付与)し、技能伝承・ノウハウ共有を促進することにある。画像認識モデル開発に伴うアノテーション負担を低減するため、次の二つのアプローチを検討した：(1) 弱教師あり学習・転移学習に基づく映像組織化：ユーザが付与する少量・不完全な注釈データを活用する、(2) 人間参加型機械学習に基づく映像組織化：画像認識の結果をユーザに能動的に提示し、ユーザフィードバックを得ながら画像認識モデルを改良する。

研究成果の学術的意義や社会的意義

製造業・看護など様々な業界において、熟練者から初心者への技能伝承、組織内でのノウハウ共有が重要な課題となっている。技能伝承・ノウハウは、企業やコミュニティに閉じたドメイン固有の概念を扱うため、学習データの整備、汎用的な画像認識モデルの構築が難しいことが問題となっている。本研究は、この問題点の解決に向けたもので、技能伝承・ノウハウ共有の促進に資するものである。なお、ここで開発した手法は、学習データ整備の困難を軽減するもので、当初想定した分野以外への展開も可能であることを示した。

研究成果の概要(英文)：This research aims to promote skill transfer and know-how sharing by automatically organizing unedited work videos --- dividing them into semantic scenes and tagging them --- using image recognition. To reduce the annotation burden associated with the development of image recognition models, we investigated the following two approaches: (1) weakly supervised and transfer learning approach, which trains image recognition models by taking advantage of the small amount of incomplete annotation data provided by the user, and (2) a human-in-the-loop machine learning approach, which improves image recognition models by presenting the image recognition results to the user and obtaining the user feedback.

研究分野：知覚情報処理

キーワード：人間参加型機械学習 Human-in-the-loop 弱教師あり学習 技能伝承 画像認識

1. 研究開始当初の背景

製造業・看護・介護・営業など様々な業界において、熟練者から初心者への技能伝承、組織内でのノウハウ共有が重要な経営課題とされてきた。言語化が難しい技能やノウハウを共有する手段として、画像・映像の有用性が様々な事例で示されている。画像・映像によるノウハウ共有の環境は整備されつつあるものの、映像編集に手間がかかるため、素人にとっては心理的なハードルが高く、技能伝承・ノウハウ共有が十分に進んでいない現状がある。未編集の状態映像を共有しても、ユーザが求めるノウハウ・コツを含むシーンを検索できず、検索してもどこがポイントか分からない。したがって、映像を組織内で共有する前に最低限の編集作業(映像のシーンにコメントや注釈を付与したり、不要なシーンを削除したりする作業)は必須である。このような背景から、手間をかけずに作業映像を自動編集・組織化する技術が求められている。

2. 研究の目的

本研究の目的は、素人が撮影した作業映像を画像認識により自動的に「組織化」する(具体的には、映像を意味的なシーンに分割し、タグを付与する)ことで、技能伝承・ノウハウ共有を促進することにある。

画像認識の精度は、深層学習の導入によって飛躍的に向上しているものの、画像認識モデルを学習するには、大量の学習データが必要となり、その作成に多大な時間とコストがかかることが問題となる。技能伝承・ノウハウは、企業やコミュニティに閉じたドメイン固有の概念を扱うため、汎用的な画像認識モデルの構築が難しい。例えば、看護教育の分野で利用するには、看護技術に特化した概念(医療器具、動作など)を認識するように画像認識モデルを学習する必要があるが、学習データ収集が困難である。

3. 研究の方法

本研究の特徴は、企業やコミュニティに特化した高精度な画像認識モデルを効率的に構築するために、人間と機械が協調しながら、システムを賢くしていく人間参加型(Human-in-the-loop)機械学習、及び、弱教師あり深層学習のアプローチを採用する点にある。

図1に示すワークフローの実現を目指して研究を進めた。まず、(a) 動画投稿者が、未編集の作業映像をノウハウ共有システムへアップロードする、(b) 映像組織化エンジンは、作業映像の内容を解析し意味的なシーンに分割し、タグを付与する、(c) ユーザがコメントや注釈を付与しシーン分割・タグの修正を行う(ユーザアノテーションとよぶ)、(d) 映像組織化エンジンが、ユーザに認識結果を能動的に提示し、認識結果の正誤を回答してもらう(ユーザフィードバックとよぶ)、(e) ユーザアノテーション、及び、ユーザフィードバックに基づいて弱教師あり学習により画像認識モデルを改良する。

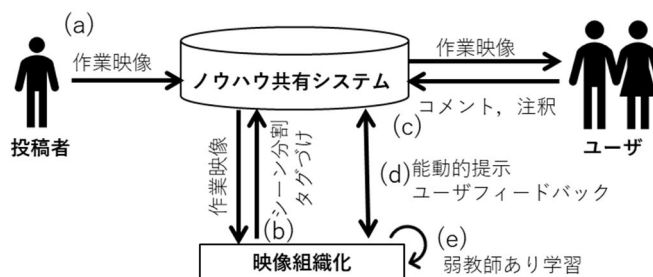


図1 映像活用型ノウハウ共有システムのワークフロー

本研究では、このワークフローの実現に向けて、以下の二つのアプローチを検討した：

- ・弱教師あり学習・転移学習に基づく映像組織化(図1(e))：ユーザが付与する少量・不完全な注釈データ(ユーザアノテーション)を用いて、ドメインに特化した画像認識モデルを構築する。
- ・人間参加型機械学習に基づく映像組織化(図1(d))：画像認識の結果をユーザに能動的に提示し、ユーザフィードバックを得ながら画像認識モデルを改良していく。

当初は、製造業における製品組み立て作業などを想定して公開データセット(料理、家具組立など)を用いた検討を進めたが、他分野(漫画制作、酪農業)への展開を意識した追加検証も実施した。

4. 研究成果

本節では、弱教師あり学習・転移学習アプローチに基づく研究成果(1)、(2)と、人間参加型機械学習アプローチに基づく研究成果(3)について述べる。

(1) 映像と言語の類似度学習による作業映像のシーン分割

作業映像に付随する作業手順書などの言語情報を、弱教師データとして利用した作業映像のシーン分割手法を提案した。例えば、料理動画は、図2に示すように、複数の工程(シーン)

からなり、工程を説明する言語情報（レシピ、キャプションとよぶ）が付随している。ここで、シーンとキャプションの間の時間的対応（工程の開始・終了時刻の情報）は未知である。我々は、シーンが時間的に重複しないことを仮定し、シーン分割を「キャプションとフレームの1対多マッチング」と捉えることでシーン分割を実現した。図3に提案手法の概要を示す。提案手法は学習と推論の二つのフェーズからなる。学習フェーズでは、キャプション付きの作業動画データセットを用いて、フレームとキャプションの間の類似度を学習する。推論フェーズでは、作業動画とキャプションを入力し、フレームとキャプションの間の類似度行列を算出し、動的計画法を用いたフレームとキャプションの1対多マッチング(DP マッチング)により作業動画をシーンに分割する。料理動画データセット YouCook2[Zhou2018]を用いた検証により、提案法の精度は従来法（画像特徴のみを用いた階層型クラスタリング法、等分割法）を上回ることを示した。

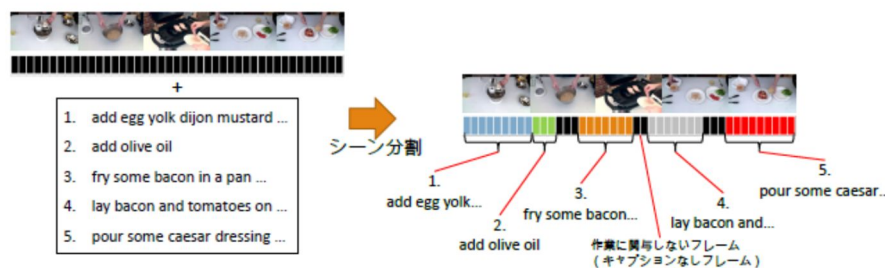


図2 料理動画におけるシーン分割の例。料理動画とレシピを入力し、動画内フレームとレシピ文のマッチングによりシーン分割を実現する。

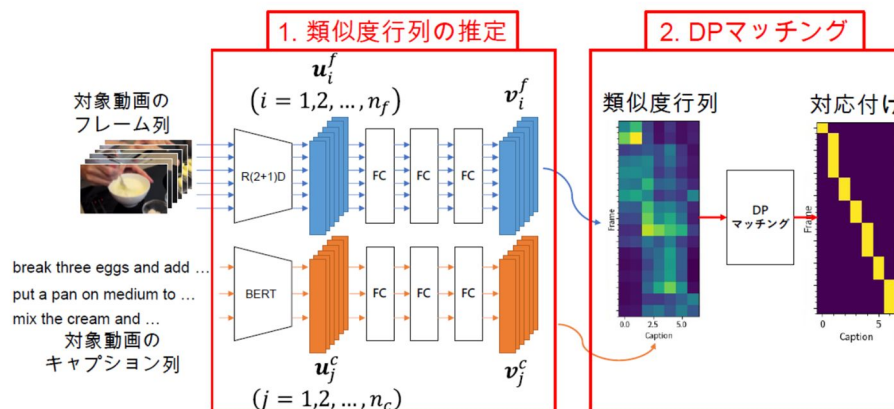


図3 映像と言語の類似度学習による作業映像のシーン分割

(2) 弱教師付き学習による画像への自動タグ付け

弱教師付き学習に基づく映像構造化アプローチを以下のタスクに適用し、その有効性を検証した。

Human-Object Interaction 検出(HOI 検出) :

HOI 検出は画像内に存在する人と物の位置を矩形として検出し、同時に人と物の関係性を推定するタスクである。人と作業対象物との関係性を理解することは、技能伝承のタグ付けにも必要である。従来は、教師データとして、人と物の外接矩形と、それらの関係性（動作ラベル）を与える必要があった。本研究では、弱教師付き HOI 検出のタスクに始めて取り組み、図4に示すように、画像全体に対して付与した動作ラベル（例えば、“eat obj”）のみを用いて HOI 検出器を学習し、画像から人・物の位置と関係性を推論する手法を提案した。弱教師付き物体検出手法である Proposal Cluster Learning (PCL) [Tang2018]を拡張した手法である。既存データセットで学習済みの人物検出器によって検出した人矩形を与えることで精度が向上することを、V-COCO データセット[Gupta 2015]を用いた実験により示した。

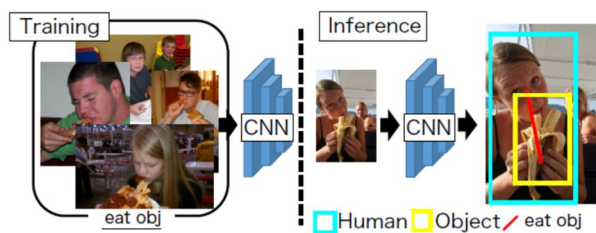


図4 弱教師付き Human-Object Interaction 検出

漫画の半自動彩色 :

日本の漫画は伝統的にモノクロ画像として制作されているが、海外ではカラー漫画の人気が高い。モノクロ漫画からカラー漫画を制作する彩色工程は、イラストレータの高い技術に頼らざ

るをえず、多大な人手と時間がかかっている。イラストレータの負担を軽減するため、本研究では、ベタ塗り画像をヒントとして与えることで、モノクロ画像（スクリーン・トーン画像）から高品質な彩色画像を生成する半自動彩色手法を提案した（図 5）。ベタ塗り画像は色指定が必要になるものの、陰影まで精緻に表現する彩色画像に比べれば短時間で作成できる。具体的には、モノクロ画像とベタ塗り画像のペアを入力し、彩色画像を出力する U-Net を敵対的生成ネットワーク(GAN)で学習した。Manga109 データセット[Manga109]を用いた評価により、従来の全自動彩色手法よりも高い画像品質(Peak signal-to-noise ratio: PSNR)が得られることを示した。また、プロのイラストレータ 4 名を被験者として、ベタ塗り画像と彩色画像の制作に要する時間を計測したところ、ベタ塗り画像を利用する提案法により約 53%の時間削減効果が得られることが示された。

ベタ塗り画像の制作時間の更なる削減を目的として、(1) 彩色画像からベタ塗り画像を自動生成するネットワークを学習するアプローチ、(2) ユーザの入力したスクリーン・トーン（カラーヒント）から Watershed 法を用いてインタラクティブにベタ塗り画像を生成するシステムを試作した。(1)は、大規模な彩色画像データセットから、疑似的に彩色画像・ベタ塗り画像・モノクロ画像の三つ組を生成することで、彩色ネットワークを学習する半教師あり学習手法である。(2)は、ユーザとのインタラクションを通じて学習データを生成する取り組みの一つである。



図 5 ベタ塗り画像を用いた漫画の半自動彩色（©えびふらい）

乳牛のマウンティング行動検知：

酪農家の高齢化や後継者不足、酪農業の大規模化に伴い、乳牛個体管理の効率化が重要な課題となっている。乳牛の発情兆候を見落とさず発見するために、牛舎天井に取り付けた監視カメラ映像から乳牛のマウンティング行動（図 6）を自動検出する手法を開発した。具体的には、通常行動をしている乳牛個体画像と牛舎背景画像を合成することで疑似マウンティング画像を作成し、事前学習した上で、少量の実マウンティング画像でファインチューニングを行う。実際の牛舎で撮影した映像を用いて、非マウンティング画像 38,219 枚、実マウンティング画像 124 枚からなるデータセットを作成し、疑似画像を用いた事前学習の有効性を確認した。



図 6 乳牛のマウンティング行動

(3) 画像認識結果の能動的提示・ユーザフィードバックの弱教師あり学習による映像組織化

効率的な技術・技能伝承を支援する一つ的手段として、熟練者と非熟練者の作業映像を見比べながら学ぶ方法がある。ただし、単純に二つの動画を並べて先頭から再生すると、作業者ごとに作業速度が異なるため作業シーンが徐々にずれてしまい、作業工程を同時に比較することが難しくなる。この映像比較を容易にするため、同じ作業工程を自動的に対応付ける時系列アライメント手法を開発した（図 7）。ここでは作業工程の順序が同一であること（順序一貫性）を仮定している。

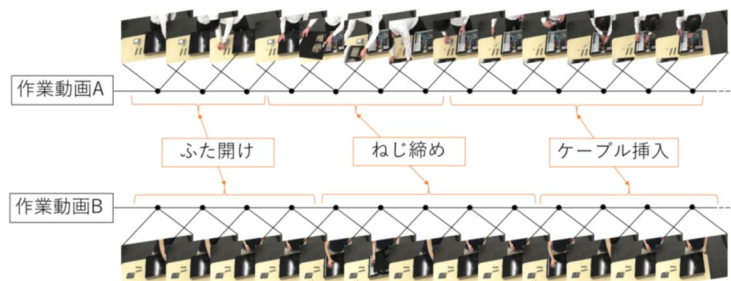


図 7 作業映像の時系列アライメント。同じ作業工程の開始時刻を対応付けることにより、作業映像を比較しやすくする。

提案手法は、図 8 に示すように、三つのステップからなる：(a) 特徴表現学習、(b) 特徴抽出、(c) 時系列アライメント。(a)特徴表現学習ステップでは、正解ラベル付きの作業映像データセット（作業工程ラベルと時間情報を付与した作業映像データセット）を入力し、画像特徴抽出器を学習する。(b)特徴抽出ステップは、比較対象となる二つの作業映像を入力し、(a)で学習した特徴抽出器を用いて、二つの特徴ベクトル列を抽出する。具体的には、動画を 1 秒間隔で

分割し、それぞれの区間から 512 次元の画像特徴ベクトルを抽出した。(c)時系列アライメントステップは、(b)で抽出した二つの特徴ベクトル列から距離行列を算出し、DP マッチングを用いて距離の総和が最小となる最短経路を求めることにより、フレームの対応付けを行う。

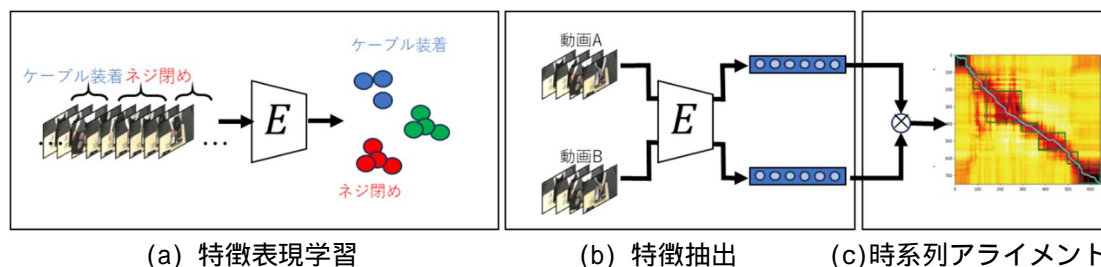


図 8 作業映像の時系列アライメント処理の概要

一つの問題は(a)特徴表現学習のための教師データの作成に時間と手間がかかることである。提案した人間参加型アプローチの処理の流れを、図 9 に示す。時系列アライメント結果に基づいて映像クリップペアを抽出し、オペレータに提示し正誤判定(一対比較)した結果を学習データセットに追加することで、学習データを拡充していく。オペレータに提示する映像クリップ対をシーン境界付近から多く選択する戦略を採用することで、正誤判定の回数(アノテーション作業時間)を短縮できることを明らかにした。被験者 12 名に、通常のアノテーション作業(作業シーンの開始終了時刻を付与する作業)と提案法(6 秒の映像クリップを並列再生し同一シーンか判定する作業)を実施してもらい、作業時間を測定したところ、提案法により約 6 割の作業時間を削減できる見込みを得た。

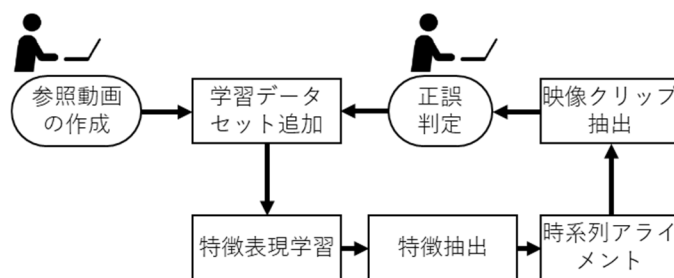


図 9 人間参加型(Human-in-the-loop)機械学習アプローチの手順。

さらなる作業時間の削減のためには、映像クリップを選択する際に作業シーン境界を含まないようにする必要がある。この機能を実現するために、手の動きに着目したシーン境界検出手法を提案した。机上作業においては、作業対象物や道具に手を伸ばすといった手の動きが、シーン境界で現れることが多い点に着目した。

応用範囲拡大に向けた今後の課題として、時系列アライメントにおける作業順序一貫性の仮定の緩和、より効率的な映像クリップ抽出、ユーザフィードバック取得方法の開発がある。

参考文献

- [Zhou, AAI2018] L. Zhou, et al., "Towards automatic learning of procedures from web instructional videos," AAI, 2018.
- [Tang2018] P. Tang et al., "PCL: Proposal cluster learning for weakly supervised object detection," TPAMI, 2018.
- [Gupta2015] S. Gupta and J. Malik, "Visual semantic role labeling," arXiv:1505.04474, 2015.
- [Manga109] Y. Matsui, et al., "Sketch-based manga retrieval using Manga109 dataset," Multimedia Tools and Applications, 2017.

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件／うち国際共著 0件／うちオープンアクセス 3件）

1. 著者名 Y. Okuda, Y. Yamamoto, K. Nakamura and Y. Taniguchi	4. 巻 -
2. 論文標題 Detecting Mounting Behaviors of Dairy Cows by Pre-Training with Pseudo Images	5. 発行年 2023年
3. 雑誌名 Proceedings the 29th International Workshop on Frontiers of Computer Vision (IW-FCV 2023)	6. 最初と最後の頁 05-5
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Shimizu Yugo, Furuta Ryosuke, Ouyang DeLong, Taniguchi Yukinobu, Hinami Ryota, Ishiwatari Shonosuke	4. 巻 -
2. 論文標題 Painting Style-Aware Manga Colorization Based On Generative Adversarial Networks	5. 発行年 2021年
3. 雑誌名 Proceedings of 2021 IEEE International Conference on Image Processing (ICIP 2021)	6. 最初と最後の頁 1739-1743
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICIP42928.2021.9506254	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Ouyang DeLong, Furuta Ryosuke, Shimizu Yugo, Taniguchi Yukinobu, Hinami Ryota, Ishiwatari Shonosuke	4. 巻 -
2. 論文標題 Interactive Manga Colorization with Fast Flat Coloring	5. 発行年 2021年
3. 雑誌名 ACM SIGGRAPH Asia 2021 Posters	6. 最初と最後の頁 1-2
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3476124.3488628	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Masaki Sugimoto, Ryosuke Furuta, and Yukinobu Taniguchi	4. 巻 5
2. 論文標題 Weakly-supervised Human-object Interaction Detection	5. 発行年 2021年
3. 雑誌名 Proceedings of the 16th International Joint Conferences on Computer Vision, Imaging and Computer Graphics Theory and Applications	6. 最初と最後の頁 293-300
掲載論文のDOI（デジタルオブジェクト識別子） 10.5220/0010196802930300	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計10件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Y. Okuda, Y. Yamamoto, K. Nakamura and Y. Taniguchi
2. 発表標題 Detecting Mounting Behaviors of Dairy Cows by Pre-Training with Pseudo Images
3. 学会等名 The 29th International Workshop on Frontiers of Computer Vision (IW-FCV 2023) (国際学会)
4. 発表年 2023年

1. 発表者名 高木 利菜, 山本 洋太, 古田 諒佑, 谷口 行信
2. 発表標題 手検出を用いた作業映像の教師なしアクション境界検出
3. 学会等名 画像電子学会第303回研究会
4. 発表年 2023年

1. 発表者名 奥田 悠太, 山本 洋太, 中村 和晃, 谷口 行信
2. 発表標題 疑似マウンティング画像生成による乳牛の発情兆候検知
3. 学会等名 第50回画像電子学会年次大会
4. 発表年 2022年

1. 発表者名 清水友悟, 古田諒佑, 欧陽徳龍, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 ベタ塗り画像を利用した半自動漫画彩色手法
3. 学会等名 映像表現・芸術科学フォーラム 2022
4. 発表年 2022年

1. 発表者名 欧陽徳龍, 古田諒佑, 清水友悟, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 Watershed法を用いたインタラクティブ漫画彩色システム
3. 学会等名 映像表現・芸術科学フォーラム 2022
4. 発表年 2022年

1. 発表者名 林純平, 山本洋太, 谷口行信
2. 発表標題 映像と文章の類似度学習による作業映像のシーン分割
3. 学会等名 画像電子学会第297回研究会
4. 発表年 2021年

1. 発表者名 Masaki Sugimoto, Ryosuke Furuta, and Yukinobu Taniguchi
2. 発表標題 Weakly-supervised Human-object Interaction Detection
3. 学会等名 the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (国際学会)
4. 発表年 2021年

1. 発表者名 林 純平, 古田諒佑, 谷口行信
2. 発表標題 作業工程の言語情報を利用した作業映像のシーン分割
3. 学会等名 電子情報通信学会2021年総合大会, 学生ポスターセッション
4. 発表年 2021年

1. 発表者名 杉本将来, 古田諒佑, 谷口行信
2. 発表標題 弱教師付き Human-Object Interaction 検出
3. 学会等名 映像情報メディア学会 創立70周年記念大会
4. 発表年 2020年

1. 発表者名 杉本将来, 古田諒佑, 谷口行信
2. 発表標題 弱教師付きHuman Object Interaction検出
3. 学会等名 第23回画像の認識・理解シンポジウム (MIRU2020)
4. 発表年 2020年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 画像着色方法及び装置、着色画像生成モデルの生成装置、並びにプログラム	発明者 清水友悟, 古田諒佑, 欧陽徳龍, 谷口行信, 日並遼太, 石	権利者 Mantra株式会社, 東京理科大学
産業財産権の種類、番号 特許、特開2023-011381	出願年 2021年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

谷口研究室ホームページ https://www.rs.tus.ac.jp/vml/
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	古田 諒佑 (Furuta Ryosuke) (20843535)	東京大学・生産技術研究所・助教 (12601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関