様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １３ 日現在

機関番号：１４４０１
研究種目：若手研究
研究期間：2020～2022
課題番号：２０Ｋ１９８２２

研究課題名（和文）Understanding Concrete and Abstract Representations in Art

研究課題名（英文）Understanding Concrete and Abstract Representations in Art

研究代表者
ＧＡＲＣＩＡ・ＤＯＣＡＭＰＯ　ＮＯＡ（Garcia Docampo, Noa）

大阪大学・データビリティフロンティア機構・特任助教（常勤）

研究者番号：８０８７０００５

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：本プロジェクトでは、コンピュータビジョン（CV）と人工知能（AI）の技術を用いて、アートにおける具体的表現と抽象的表現を理解するための研究を行いました。このプロジェクトの成果は3つある：
1）美術品に関する質問に答えるためのデータセットを作成した。2）絵画から説明文を生成するモデルを開発した。3）写真から作品を特定するためのデータセットを作成した。
私たちの研究は、トップレベルの学会で発表され、いくつかの招待講演を通じて研究コミュニティに提供されました。

研究成果の学術的意義や社会的意義
With this research, we aimed to make art more accessible to people by developing artifical intelligence models that facilitate the understanding of art. For example, by generating automatic explanations of paintings, museum visitors can understand the intrinsic meaning of each artwork easily.

研究成果の概要（英文）：In this project, we have conducted research to understand concret and abstract representation in art using computer vision (CV) and artificial intelligence (AI) techniques. The achievements of the project are three fold:
1) We created a dataset for answering art-related questions about fine-art paintings. 2) We developed a model to generate descriptions from paintings. 3) We created a dataset for identifying artworks given photo.
Our work was published at top conferences and delivered to the research community through several invited talks.

研究分野：computer vision

キーワード：computer vision for art　vision and language　description generation　image search

１．研究開始当初の背景

Art is essential for the conservation of cultures and their history. It exists in every community and every culture in the world since time began, and it reflects a society's beliefs, values, and accomplishments. In today's world, art brings us the unique opportunity to look at the past and understand the complex historical processes that lead us to the present. Nowadays, with the large-scale digitization of artistic representations from collections all over the world, computer vision (CV) and machine learning have become important tools in the conservation and dissemination of cultural heritage. Some of the most promising work on this direction involves the automatic analysis of paintings.

Due to the latest advancements in deep learning, the study of paintings in CV has already achieved outstanding results in the recognition of low-level attributes (e.g., brushwork, illumination, palette etc.) as well as medium-level attributes (e.g., style, author, depicted objects etc.). However, it still remains as a challenging unsolved problem when it comes to high-level attributes (e.g., motivation, influences, ideas etc.), as current datasets and models are only focused on the concrete concepts that appear in an image. Thus, questions like "what inspired the author to create this painting?" cannot be addressed with the current techniques. As an attempt to study abstract representations, semantic art understanding [1] propose to find a painting inside a dataset that best matches a description. However, similar to medium-level attribute recognition methods, the proposed models only identify the keywords in the description that have a direct correspondence with the concrete objects represented in the image, without paying attention to the abstract concepts in the artwork.

２．研究の目的

The purpose of this research is to study abstract high-level representations in art. To that end, we propose a new framework to evaluate both concrete and abstract understanding based on VQA. In VQA, models have to answer questions about images, which are usually related to their content. By adopting the VQA framework in the art domain, we will be able to evaluate different levels of understanding by using different categories of questions: 1) questions about low-level attributes (e.g., illumination, colors, brushwork); 2) questions about medium-level attributes (e.g., objects, style, author); and 3) questions about abstract high-level attributes (e.g., influences, motivations, abstract ideas). The research objectives (RO) of this proposal are:

- RO1: Design a VQA framework with multiple categories of questions for art understanding evaluation.
- RO2: The implementation of architectures to extract different level of attributes from paintings.
- RO3: The generalization of the proposed architectures to different visual domains.

３．研究の方法

The methodology of the project has been developed into several parts:

- **Visual question answering dataset**: we created the first dataset for visual question answering in art, which we called AQUA (RO1). The dataset was created using automatic natural language processing and computer vision techniques from our previous existing dataset in [1]. The dataset contains diverse types of questions about paintings, such as questions about the visual content, objects, and attributes that are depicted, as well as knowledge-based questions about the context and history of the painting.
- **Visual question answering model for art:** we designed a computer vision and machine learning-based model to address visual question answering in art, which we called VIKING (RO2). The model consists of three parts: a module selector, a visual-based question answering, and a knowledge-based question answering. For a given question and painting, the module selector decides whether the question is visual-based (e.g., about the objects in the image) or knowledge-based (e.g., about the history of the painting). Then, the visual-based question answering and the knowledge-based

question answering modules are in charge of finding the correct answer for the question by using either the image (visual-based question answering) or external knowledge bases (knowledge-based question answering).

- **Visual question answering for general images:** We developed visual questions answering models for the general image domain and studied how to transfer knowledge between different domains (RO3).
- **Artwork recognition** (AR): we created a dataset and benchmark for recognizing artworks in museum collections, which we called The MET (RO3). Artwork recognition requires the extraction and representation of low-level features, which need to be able to capture the specific details of each piece of art in particular. The dataset was created by using the open-source collection of The MET museum in New York (US) and manually matching museum visitor photos with their original artworks. Along with the dataset, we introduced a benchmark and evaluated the latest state-of-the-art algorithms on image recognition.
- **Image captioning for paintings** (IC): we developed an algorithm to automatically generate descriptions and explanations from fine-art paintings (RO3). The algorithm was able to generate three types of sentences based on the three main topics in art (content, form, and context), and thus, generating longer and more diverse explanations than standard image captioning methods.

４．研究成果

The results of this research are summarized below.

Datasets: We generated several datasets that can be accessed by the research community and contribute to advance the field of automatic art analysis.

- AQUA dataset: a dataset for visual question answering. URL: https://github.com/noagarcia/ArtVQA
- The MET dataset: a dataset for artwork recognition. URL: http://cmp.felk.cvut.cz/met/
- KnowIT VQA dataset for video question answering and knowledge transfer. URL: https://knowit-vqa.github.io/
- Art descriptions annotations: data for studying different types of description in art. URL: https://sites.google.com/view/art-description-generation

Code: We published the code of our models so that other researchers can implement and reproduce our results, which leads to more transparency in science and faster progress.

- VIKING model for visual question answering. URL: https://sites.google.com/view/art-description-generation
- Knowledge-based Video Question Answering model. URL: https://github.com/noagarcia/knowit-rock
- Benchmarks for artwork recognition. URL: http://cmp.felk.cvut.cz/met/
- Image captioning for paintings model. URL: https://github.com/JosephPai/Art-Description

International conferences presentations:

- Yankun Wu, Yuta Nakashima**, Noa Garcia**. Not Only Generative Art: Stable Diffusion for Content-Style Disentanglement in Art Analysis. In: ACM International Conference on Multimedia Retrieval (ICMR), to appear, Thessaloniki, Greece, 2023.
- Zechen Bai, Yuta Nakashima, **Noa Garcia**. Explain Me the Painting: Multi-Topic Knowledgeable Art Description Generation. In: International Conference on Computer Vision (ICCV), pp. 5422 - 5432, Online 2021.
- Nikolaos-Antonios Ypsilantis, **Noa Garcia**, Guangxing Han, Sarah Ibrahimi, Nanne Van Noord, Giorgos Tolias. The MET Dataset: Instance-level Recognition for Artworks. In: Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks), pp. 1-12, Online 2021.

- Tianran Wu, **Noa Garcia**, Mayu Otani, Chenhui Chu, Yuta Nakashima, Haruo Takemura. Transferring Domain-Agnostic Knowledge in Video Question Answering. In: British Machine Vision Conference (BMVC), pp. 1-13, Online 2021.
- Cheikh Brahim El Vaigh, **Noa Garcia**, Benjamin Renoust, Chenhui Chu, Yuta Nakashima, Hajime Nagahara. GCNBoost: Artwork Classification by Label Propagation through a Knowledge Graph. In: ACM International Conference on Multimedia Retrieval (ICMR), pp. 92-100, Online 2021.
- **Noa Garcia**, Yuta Nakashima. Knowledge-Based Video Question Answering with Unsupervised Scene Descriptions. In: European Conference on Computer Vision (ECCV), pp 581-598, Online 2020.

Journal papers:

- Zekun Yang, **Noa Garcia**, Chenhui Chu, Mayu Otani, Yuta Nakashima, Haruo Takemura. A Comparative Study of Language Transformers for Video Question Answering. In: Neurocomputing (Elsevier), vol. 445, pp. 121-133, March 2021.
- **Noa Garcia**, Benjamin Renoust, Yuta Nakashima. ContextNet: Representation and Exploration for Painting Classification and Retrieval in Context. In: International Journal of Multimedia Information Retrieval (Springer), pp. 17-30, December 2019.

International workshops with Proceedings:

- Yusuke Hirota, **Noa Garcia**, Mayu Otani, Chenhui Chu, Yuta Nakashima, Ittetsu Taniguchi, Takao Onoye. Visual Question Answering with Textual Representations. In: International Conference on Computer Vision (ICCV) Workshop Proceedings - Closing the Loop Between Vision and Language, 2021.
- Nikolai Huckle, **Noa Garcia**, Yuta Nakashima. Demographic Influences on Contemporary Art with Unsupervised Style Embeddings. In: European Conference on Computer Vision (ECCV) Workshop Proceedings - 5th Workshop on Computer Vision for Art Analysis, 2020.
- **Noa Garcia**, Chentao Ye, Zihua Liu, Qingtao Hu, Mayu Otani, Chenhui Chu, Yuta Nakashima, Teruko Mitamura. A Dataset and Baselines for Visual Question Answering on Art. In: European Conference on Computer Vision (ECCV) Workshop Proceedings - 5th Workshop on Computer Vision for Art Analysis, 2020.

Invited talks:

2022/08 *Through the latent space. A conversation on Dall-E and art (history)*
Panel discussion at Digital Art History Summer School 2022 (Online)

2022/07 *Automatic interpretability in art*
Thousand Words of Art Tutorial. Conference on Digital Humanities 2022 (Online)

2022/06 *Content, Form, and Context. What we talk about when we talk about art?*
5th Workshop on Computer Vision for Fashion, Art, and Design, CVPR 2022 (US)

2021/10 *AI and Art*
The First SCAI-IDS Workshop for Future Collaboration on AI (Online)

2021/03 *Understanding fine-art paintings through visual and language representations*
CAI+CAI Workshop, NLP Conference 2021 (Online)

References:

[1] Noa Garcia and George Vogiatzis. How to read paintings: semantic art understanding with multi-modal retrieval. In Proceedings of the European Conference on Computer Vision, pp. 676-691, (2018).

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　0件）

１．著者名 Zekun Yang, Noa Garcia, Chenhui Chu, Mayu Otani, Yuta Nakashima, Haruo Takemura	４．巻 445
２．論文標題 A comparative study of language transformers for video question answering	５．発行年 2021年
３．雑誌名 Neurocomputing	６．最初と最後の頁 121-133
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.neucom.2021.02.092	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 該当する

〔学会発表〕　計13件（うち招待講演　5件／うち国際学会　11件）

１．発表者名 Zechen Bai, Yuta Nakashima, Noa Garcia
２．発表標題 Explain Me the Painting: Multi-Topic Knowledgeable Art Description Generation
３．学会等名 IEEE/CVF International Conference on Computer Vision 2021（国際学会）
４．発表年 2021年

１．発表者名 Nikolaos-Antonios Ypsilantis, Noa Garcia, Guangxing Han, Sarah Ibrahimi, Nanne Van Noord, Giorgos Tolias
２．発表標題 Instance-level Recognition for Artworks: The MET Dataset
３．学会等名 Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track 2021（国際学会）
４．発表年 2021年

１．発表者名 Tianran Wu, Noa Garcia, Mayu Otani, Chenhui Chu, Yuta Nakashima, Haruo Takemura
２．発表標題 Transferring Domain-Agnostic Knowledge in Video Question Answering
３．学会等名 British Machine Vision Conference 2021（国際学会）
４．発表年 2021年

１．発表者名
Cheikh Brahim El Vaigh, Noa Garcia, Benjamin Renoust, Chenhui Chu, Yuta Nakashima, Hajime Nagahara

２．発表標題
GCNBoost: Artwork Classification by Label Propagation through a Knowledge Graph

３．学会等名
ACM International Conference in Multimedia Retrieval 2021（国際学会）

４．発表年
2021年

１．発表者名
Yusuke Hirota, Noa Garcia, Mayu Otani, Chenhui Chu, Yuta Nakashima, Ittetsu Taniguchi, Takao Onoye

２．発表標題
Visual Question Answering with Textual Representations

３．学会等名
Workshop on Closing the Loop Between Vision and Language, IEEE/CVF International Conference on Computer Vision 2021（国際学会）

４．発表年
2021年

１．発表者名
Noa Garcia

２．発表標題
Understanding Fine-Art Paintings through Visual and Language Representations

３．学会等名
CAI+CAI Workshop, Natural Language Processing 2021（招待講演）

４．発表年
2021年

１．発表者名
Noa Garcia, Yuta Nakashima

２．発表標題
Knowledge-Based Video Question Answering with Unsupervised Scene Descriptions

３．学会等名
European Conference on Computer Vision（国際学会）

４．発表年
2020年

１．発表者名
Noa Garcia, Chentao Ye, Zihua Liu, Qingtao Hu, Mayu Otani, Chenhui Chu, Yuta Nakashima and Teruko Mitamura

２．発表標題
A Dataset and Baselines for Visual Question Answering on Art

３．学会等名
VISART workshop at European Conference on Computer Vision 2020（国際学会）

４．発表年
2020年

１．発表者名
Nikolai Huckle, Noa Garcia and Yuta Nakashima

２．発表標題
Demographic Influences on Contemporary Art with Unsupervised Style Embeddings

３．学会等名
VISART workshop at European Conference on Computer Vision 2020（国際学会）

４．発表年
2020年

１．発表者名
Noa Garcia

２．発表標題
Through the latent space. A conversation on Dall-E and art (history)

３．学会等名
Panel discussion at Digital Art History Summer School 2022（招待講演）（国際学会）

４．発表年
2022年

１．発表者名
Noa Garcia

２．発表標題
Automatic interpretability in art

３．学会等名
Thousand Words of Art Tutorial. Conference on Digital Humanities 2022（招待講演）（国際学会）

４．発表年
2022年

１．発表者名
Noa Garcia

２．発表標題
Content, Form, and Context. What we talk about when we talk about art?

３．学会等名
5th Workshop on Computer Vision for Fashion, Art, and Design, CVPR 2022（招待講演）（国際学会）

４．発表年
2022年

１．発表者名
Noa Garcia

２．発表標題
AI and Art

３．学会等名
The First SCAI-IDS Workshop for Future Collaboration on AI（招待講演）

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

Art Description Generation
https://sites.google.com/view/art-description-generation
The Met Dataset
http://cmp.felk.cvut.cz/met/
Visual Question Answering on Art
https://github.com/noagarcia/ArtVQA
contempArt
https://contempart.org/

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

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