Alejandro **Pardo**

PHD CANDIDATE · COMPUTER VISION

✉ alejandro.pardo@kaust.edu.sa | ⌂ pardoalejo.github.io | PardoAlejo | luis-alejandro-pardo-gonzalez

Professional Experience

Adobe Research *San Jose, California*

RESEARCH INTERN *June 2023- September 2023*

- Research Internship in vision and language models for video editing tasks. Supervised by Fabian Caba Heilbron.

Embodied AI Labs at Intel Research *Munich, Germany*

RESEARCH INTERN *August 2022- November 2022*

- Research Internship working on Vision and Language models for general segmentation tasks. Supervised by Matthias Mueller.

Publication and Research Experience

MatchDiffusion: Training-free Generation of Match-Cuts

UNDER REVIEW - 2024

- **Publication:** **Alejandro Pardo**, Fabio Pizzati, Tong Zhang, Alexander Pondaven, Philip Torr, Juan Camilo Perez, & Bernard Ghanem. (2024). "Generative Timelines for Instructed Visual Assembly." Under Review.
- **Description:** We introduce a training-free method for generating match-cuts using text-to-video diffusion models. By leveraging the denoising process, our approach creates visually coherent video pairs with shared structure but distinct semantics, enabling the creation of seamless and impactful transitions.

Generative Timelines for Instructed Visual Assembly

NEURIPS WORKSHOP ON VIDEO-LANGUAGE MODELS - 2024

- **Publication:** **Alejandro Pardo**, Jui-Hsien Wang, Josef Sivic, Bryan Russell, Bernard Ghanem, & Fabian Caba Heilbron. (2024). "Generative Timelines for Instructed Visual Assembly." At NeurIPS 2024 Workshop.
- **Description:** We introduce the Assembler, a generative model trained to perform instructed visual assembly tasks, enabling users to edit video timelines via natural language. We develop a large multimodal language model to process visual content, compactly represent timelines, and interpret editing instructions. Our automatic dataset generation method allows efficient training, and we validate our model on two novel datasets, where it outperforms recent baselines like GPT-4o in real-world assembly tasks.

Towards Automated Movie Trailer Generation

CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR) - 2024

- **Publication:** Dawit Mureja Argaw, Mattia Soldan, **Alejandro Pardo**, Chen Zhao, Fabian Caba Heilbron, Joon Son Chung, & Bernard Ghanem. (2024). "Towards Automated Movie Trailer Generation." On CVPR, 2024.
- **Description:** Automatic Trailer generation system that generates plausible trailers from a full movie. Our approach is inspired by machine translation, and approaches trailer generation as a sequence-to-sequence task, significantly outperforming existing methods across various metrics.

MovieCuts: A New Dataset and Benchmark for Cut Type Recognition

EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV) - 2022

- **Publication:** **Alejandro Pardo**, Fabian Caba Heilbron, Juan León Alcázar, Ali Thabet, & Bernard Ghanem. (2021). "MovieCuts: A New Dataset and Benchmark for Cut-Type Recognition." On ECCV, 2022.
- **Description:** Understanding movies and their structural patterns is a crucial task in decoding the craft of video editing. We construct a large-scale dataset called MovieCuts, which contains more than 170K video clips labeled among ten cut types.

MAD: A Dataset for Language Grounding in Videos from Movie Audio Descriptions

CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR) - 2022

- **Publication:** Mattia Soldan, **Alejandro Pardo**, Juan León Alcázar, Fabian Caba Heilbron, Chen Zhao, Silvio Ginacola & Bernard Ghanem. (2021). "MAD: A Scalable Dataset for Language Grounding in Videos from Movie Audio Descriptions." On CVPR, 2022.
- **Description:** A novel benchmark that departs from the paradigm of augmenting existing video datasets with text annotations and focuses on crawling and aligning available audio descriptions of mainstream movies. MAD's collection strategy enables a novel and more challenging version of video-language grounding, where short temporal moments (typically seconds long) must be accurately grounded in diverse long-form videos that can last up to three hours.

Learning to Cut by Watching Movies

INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV) - 2021

- **Publication:** **Alejandro Pardo**, Fabian Caba Heilbron, Juan León Alcázar, Ali Thabet, & Bernard Ghanem. (2021). "Learning to Cut by Watching Movies." On ICCV, 2021.
- **Description:** We propose a new method and pipeline to create video editing cut recommendations. Our method utilizes the information of already edited content to learn patterns between plausible and non-plausible cuts via contrastive learning.

RefineLoc: Iterative Refinement for Weakly-Supervised Action Localization

WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV) - 2021.

- **Publication:** Alejandro Pardo, Humam Alwassel, Fabian Caba Heilbron, Ali Thabet, & Bernard Ghanem. (2021). RefineLoc: Iterative Refinement for Weakly-Supervised Action Localization. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 3319-3328).
- **Description:** RefineLoc is a weakly-supervised temporal action localization method. RefineLoc uses an iterative refinement approach by estimating and training on snippet-level pseudo ground truth at every iteration. Additionally, our iterative refinement process significantly improves the performance of two state-of-the-art methods, setting a new state-of-the-art on THUMOS14.

Academic Experience

ICVSS - Computer Vision in the Age of Large Language Models   *Sicily, Italy*

ATTENDEE   *2024*

- Attendance to the prestigious International Computer Vision Summer School (IEEE PAMI Mark Everingham Prize - 2017). Focused on Computer Vision in the Age of Large Models.

ELLIS Winter School on Foundation Models   *Amsterdam, Netherlands*

ATTENDEE   *2024*

- Attendance, First Winter School on Foundation Models. Organized by the ELLIS Unit at University of Amsterdam. Focused on the construction, application, and analysis of current foundation models within Computer Vision and Natural Language Processing domains.

AI for Creative Video Editing and Understanding (CVEU) Workshop   *ICCV-2021*

CO-ORGANIZER AND PR CHAIR

- Co-organizer of the first CVEU Workshop at ICVV2021. I worked as Web and Public Relations Chair in the main role while supporting the other tasks like the call for papers, industry speakers invitations, among others..

LatinX in AI Workshop   *CVPR-2021*

CO-ORGANIZER AND WEB CHAIR

- Co-organizer and web chair of the first LatinX in AI workshop at CVPR.

Universidad de los Andes   *Bogota, Colombia*

TEACHING ASSITANT   *2014-2015*

- Teaching Assistant for the course Science, Technology and Gender, by Professor Alba Avila.

Universidad de los Andes   *Bogota, Colombia*

TEACHING ASSITANT   *2013*

- Teaching Assistant for the course Digitial Electronics, by Professor Antonio Garcia Rozo.

Presentations

Perceiving Systems Lab   *Max Planck Institue for Intelligent Systems*

INVITED SPEAKER   *September. 2022*

- I gave a talk about Computer Vision and automated Video Editing at the Perceiving Systems Lab, lead by Michael Black.

Education

King Abdullah University of Science and Technology - KAUST   *Thuwal, Saudi Arabia - 23955-6900*

PH.D. IN ELECTRICAL ENGINEERING   *2019 - Now*

Universidad de los Andes   *Bogotá, Colombia - 111711*

M.S. IN BIOMEDICAL ENGINEERING   *2017 - 2018*

Honors & Awards

2022	**Awarded,** Outstanding Reviewer Award	*ECCV 2022*
2021	**First Place,** Best Paper Award	*LatinX in AI @ CVPR*
2020	**Awarded,** Outstanding Reviewer Award	*CVPR 2020*
2020	**Awarded,** Outstanding Reviewer Award	*BMVC 2020*
2019	**Awarded,** KAUST Fellowship for PhD Studies	