chist-era

CHIST-ERA Project Periodic Report

XAIface
Measuring and Improving Explainability for AI-based Face Recognition

Periodic report n° 1
25 April 2022

This document must be filled in by the project coordinator with the help of the project partners and must be uploaded online in the dedicated web portal at the end of each period (typically every year after project start). The Joint Secretariat ensures distribution to the concerned research funding organisations. The project coordinator is responsible for sending a copy of the report to the project partners.

The information provided should cover the whole duration since project start (information from a previous period should be kept in for the next period if still relevant; the report for the final period thus also constitutes the project final report covering the whole project duration).

You are also encouraged to take advantage of this reporting to update your project factsheet on the CHIST-ERA website as well as associate to your project the scientific publication in open access: https://www.chistera.eu/toolbox

1. Progress Report

1.1. Project objectives and activities implemented

(Indicative length: 2 pages per period)

Describe the work performed during each period and assess it with respect to the initial work plan. Clearly indicate who performed each part of the work and which parts are done in cooperation, describing the nature of the cooperation. Mention any difficulty encountered and the solutions implemented.

If applicable, indicate the work planned during the rest of the project, relating it to the initial work plan and the work already performed. Mention any open issue (e.g.: technical deadlock, service provider default, failure to meet deadlines, budget control), the solutions envisaged, and any foreseen need for a contractual project content revision or schedule extension.

Work already performed:

- EURECOM: Project management: organisation and coordination of meetings & minutes; coordination of the collaborative tools (Mailing list, Google Drive, Slack); creation of document templates and XAIface project logo; literature review of Face Detection (FD) and Face Recognition (FR) methods and SotA explainability techniques; Project website xaiface.eurecom.fr design and implementation (M3); Quality Control and Risk Register document (M6); Dissemination Plan document (M6); Contributor to D2.1: Progress report on influencing factors and models (M10) with sensor-related and social (artificial beautification & drug abuse) factors; Contributor to D2.2: Evaluation metrics and protocols (M10) with databases description; D4.1: Explainability Protocol and Methods (document structure, 2 new approach descriptions and contribution to different

explainability techniques); Presentation of XAIface and participation to the annual CHIST-ERA Projects Seminar 2021 & 2022; Paper submitted and accepted to CVPRw 2022;

- JRS: project management; generic literature review and study (IBM's XAI toolbox); contributor DMP; contributor D2.2 (evaluation protocols); contributor D4.1 (DL-methods for face recognition and new approach description); contributor database selection process; contributor reference pipeline definition and selection (e.g. definition of selection criteria + approaches); joint GIT-Hub contributions and assessment (x86 compiling issues Retina-Face MXNET + PyTorch); provision reference (baseline) results for detection, verification and identification (IARPA Janus Benchmark - IJB-B and -C - reference protocol implementation for identification and 1:1 single face verification) and comparison with other partners (e.g. IT); IJB-B,C database understanding (protocol, API); contributions to GIT-documentation; bilateral seminar regarding technical, actual state (basics) and legal aspects of face-recognition (spec. Austria);

- UNIVIE: D3.1 Data Management Plan (DMP) based on the contributions of each partner; D3.2: Face Image Dataset, analysis of current legal obligations for research institutions concerning data collection from freely available sources and their use in research projects with a special focus on international data protection law; privileges and exceptions within the GDPR regarding scientific research; evaluation of administrative & court decisions; criteria for database selection; D4.1: Description of legal and ethical aspects consisting of distinction between explainability, interpretability and explainability; accumulation of current legal obligations and discussion of available guidelines and secondary literature; assessment of legally defined visualisation techniques; explanation of modalities of transparency according to the GDPR; preview of future legal obligations under the AI-Act on intrepretability; analysis of the article on human oversight (design requirements) according to the proposal of the AI-Act; analysis of the ethics guidelines by the HLEG on AI;

- IT: Study and written survey on the influencing factors that impact the recognition performance of AI-based facial recognition systems, in general, and those based on deep learning (DL)-based facial recognition systems, in particular. The work and survey also reviewed existing publications that study the impact of the various influencing factors in face recognition performance, and try to model the nature and strength of their effect when compared to each other, as well as any interference between them; this survey was contributed to D2.1. Definition of a biometrics-compliant face recognition verification protocol, including appropriate assessment metrics; the result was contributed to D2.2. Design and implementation of a XAIFace GitHub project with the selected face recognition pipelines, notably RetinaFace for face detection and ArcFace and MagFace for face recognition; this will be central for D5.1;

- EPFL: Investigation of the influencing factors from extrinsic environment and processing operations that are common in real world; proposal of a generic assessment framework to evaluate the impact of various influencing factors to different recognition systems or detectors; submission of two papers about the proposed assessment framework to conferences; investigation of different evaluation metrics for face verification task; understanding and implementation the ArcFace pipeline and code; investigation of the verification and identification protocols for IJB-B/C and LFW dataset and implement with ArcFace pipeline; literature review of explainable AI methods and proposal of a plan to develop XFace technique; coordination and contribution to D2.2 (evaluation protocol and metrics); contribution to D4.1 (explainable face

recognition); contribution to D2.1 (influencing factors); review of the state-of-the-art face recognition methods and contribution to face recognition pipeline selection; review of face recognition databases and contribution to database selection.

Work in progress:

- EURECOM: Face image dataset (M12); Explainability protocol and methods (M12); Explanatory video. Contributions to all deliverables;
- JRS: Contributions to outstanding deliverables, project reports and project video(s);
- UNIVIE: First version of legal guidelines for AI-based face recognition (D3.3 v1) and ethical guidelines for AI-based face recognition (D3.4 v1) presentation and discussion of the results within the consortium in both WP 3 and 4;
- IT: The impact of image coding is one of those influencing factors, as nowadays compression is used whenever acquiring, storing, or transmitting an image, including face images. This technology can impact AI-based face recognition, as using decoded images might lead to recognition performance degradation. It is therefore important to include an interpretable reasoning of how image compression technology can impact the decision made by AI-based face recognition systems. This work assesses and analyses the impact of image coding/compression on AI-based face recognition systems' final decisions, notably considering both conventional and AI-based image coding/compression solutions on AI-based facial recognition, thus providing a contribution to understand and explain AI-based face recognition system behaviour. This work will be contributed to the second version of D2.1;
- EPFL: Contribution to deliverables, especially D2.2; implementation EPFL proposed assessment framework on the reference ArcFace/MagFace face recognition pipeline and producing results; investigation of the evaluation metrics for face identification task; investigation and implementation of face verification and identification protocols for other datasets; investigation of the RISE-based explainable face recognition technique.

Work delayed:

- EPFL SNSF funded part started with 3 months delay when compared to the start of the project for the other project participants. Consequently, deliverable D2.2 "Progress report on evaluation metrics and protocols" (responsible EPFL) is shifted by 3 months;
- D6.4 Explanatory video has been postponed (responsible EURECOM) due to the global health situation that did allow the consortium to have a meeting in attendance to record the video. The video is currently in preparation and its parts are recorded separately by each partner locally and will be merged later by EURECOM with the help of a professional filmmaker;
- Data Management Plan (responsible UNIVIE): A first version (v1) of the DMP has been established based on the inputs of all project partners. This working document will be updated according to the project progress and to the specification of the data processing activities.

1.2. Transnational collaboration

Describe the added value and synergies in the collaboration, any obstacles to the transnational collaboration, and the proposed solution (if necessary).

-

1.3. Significant events and results

(Indicative length: 2-4 pages)

Describe the main achievements of the project. For example:

- *New ideas, new knowledge, new interpretative models of complex phenomena;*
- *Realization of new scientific instrumentation and/or advanced devices;*
- *Implementation of new advanced scientific methodologies;*
- *Realization of prototypes;*
- *Proposal of new technologies;*
- *Contribution to innovation in the production of goods and services;*
- *Development of innovative software;*
- *Economic impact and results exploitation.*

For each achievement, provide a description with factual and, if relevant, quantitative information.

For significant results you would like to publicise using the communication channels of CHIST-ERA, please feel free to forward the information to CHIST-ERA Joint Secretariat using the Toolbox dedicated to the funded projects: https://www.chistera.eu/toolbox

FR pipeline selection and setup of GIT-project page; experiments regarding baseline performance; Database selection Technical webinar: PhD and Postdoc presentations

1.4. Technology readiness level (TRL)

Describe the global positioning of the project (from 'idea to application', or from 'lab to market'). Refer to Technology Readiness Levels (see definition here) at the beginning and at the end of the project.

TRL 2 – technology concept formulated

1.5. *Consortium meetings*

Provide the cumulative list of consortium meetings from project start.

Meetings				
N°	**Date**	**Location**	**Attending partners**	**Purpose**
1	12 May 2021	virtual	All partners	Kick-off meeting
2	29 October 2021	Virtual	All partners	To present project progress. To present tools set up by EURECOM for collaboration. To present the deliverables already submitted (e.g. project website) and to be submitted soon. To plan work for the upcoming months.
3	15 November 2021	Virtual	All partners	Technical webinar to present the planned work to be carried out in XAIface by each partner and discuss possible collaborations.
4	6 December 2021	Virtual	EURECOM, IT, JRS, EPFL	XAIface technical meeting. This meeting aims to select the face recognition (FR) pipeline to be used as a baseline in XAIface.
5	10 January 2022	Virtual	All partners	XAIface technical meeting. This meeting aims to select the image databases to be used in XAIface.
6	3 March 2022	Virtual	All partners	XAIface consortium meeting – Deliverables. This meeting aims to discuss upcoming deliverables: current status and plan towards submission.
7	9 March 2022	Virtual	EURECOM, IT, JRS, EPFL	XAIface technical meeting - Deliverable 2.2. This meeting aims to discuss deliverable 2.2 “Progress report on evaluation metrics and protocols”: current status and plan towards submission.
8	22 March 2022	Virtual	EURECOM, IT, JRS, EPFL, UNIVIE	XAIface technical meeting – GitHub project structure.
9	25 March 2022	Virtual	EURECOM, IT, JRS, EPFL, UNIVIE	XAIface technical meeting - Deliverable 4.1. This meeting aims to discuss deliverable 4.1 “Explainability protocol and methods”: current status and plan towards submission.

1.6. *Deliverables*

Provide the cumulative list of deliverables from project start.

Deliverables					
N°	**Title**	**Nature**	**Delivery date (month)**		**Partner in charge**
			Contractual	**Actual**	
6.1	Project website		July 2021	July 2021	EURECOM
1.2	Quality control and risk register		October 2021	October 2021	EURECOM
6.2	Dissemination plan		October 2021	October 2021	EURECOM
6.4	Explanatory videos		October 2021	In progress	EURECOM
3.1	Data Management Plan v1		December 2021	April 2022	UNIVIE
2.1	Progress report on influencing factors and models		February 2022	March 2022	EURECOM

2.2	Progress report on evaluation metrics and protocols		February 2022	June 2022	EPFL
1.1	Global annual report on project progress and activities		April 2022	April 2022	EURECOM
3.2	Face image dataset		April 2022	April 2022	EURECOM
4.1	Explainability protocol and methods		April 2022	April 2022	EURECOM

1.7. Free comments

Compliance with project objectives, interaction between the partners, issues, questions to CHIST-ERA...

To request a project modification, please use the dedicated form on the Toolbox: https://www.chistera.eu/toolbox

-

2. Dissemination of results, exploitation, impact

2.1. Scientific publications (conferences/workshops, book chapters, etc.)

Indicate the publications resulting from the project. Mention only those that result directly from the project (after it started, and which mention the support of CHIST-ERA and the project reference). Indicate whether they correspond to single or multi-partner communications (multi-partner means involving several project partners). Indicate if they are available in Open Access and linked to the respective underlying data. Provide the corresponding Digital Object Identifiers (DOI).

Distinguish the different categories of publications (journals/conference proceedings, technical reports, etc.). Use the usual citation standards for the field reference. If the publication is accessible on line, indicate the URL.

Please harmonise the bibliography and use only one font.

	Scientific publications					
Reference (list of authors, journal/conference proceedings/other, pages, year of publication, ...)	**Multi-project partners of same country (Yes/No)**	**Multi-project partners of different countries (Yes/No)**	**Open Access (Yes/No)**	**DOI**	**URL**	**DOI(s) of underlying data**
T. Chakraborty, U. Trehan, K. Mallat, J.-L. Dugelay: "Generalizing Adversarial Explanations with Grad-CAM", CVPRw 2022. June 19-23, Louisiana.	**No**	**No**	**Yes**	-	**https://arxiv.org/abs/2204.05427**	
Lu, Y., Luo, R. and Ebrahimi, T., 2022. A Novel Framework for Assessment of Learning-based Detectors in Realistic Conditions with Application to Deepfake Detection. arXiv preprint arXiv:2203.11797. (Submitted to ICIP 2022)	**No**	**No**	**Yes**	-	**https://arxiv.org/abs/2203.11797**	
Lu, Y. and Ebrahimi, T., 2022. A New Approach to Improve Learning-based Deepfake Detection in Realistic Conditions. arXiv preprint arXiv:2203.11807. (Submitted to EUSIPCO 2022)	**No**	**No**	**Yes**	-	**https://arxiv.org/abs/2203.11807**	

URL of Data Management Plan (optional):

2.2 Exploitation plan

Outline an exploitation plan of your most significant exploitable results including:

- *Who will exploit the result output (project participant/if someone else then who and how will they be informed);*
- *Use type (commercial/other use);*
- *Intellectual property rights arrangements if relevant;*
- *Target end user;*
- *Roadmap and goals during and after the project's lifetime (plan of actions to be taken to achieve exploitation);*
- *Timeframe.*

Dissemination and exploitation actions are aimed to maximize the impact of the research. The public disclosure of the results by any appropriate means as well as utilisation of results in further research activities other than those covered by the action are the essential elements of good research practice.

The report D6.2 "Dissemination report" has been released in its first version in M6. The report will be updated at M24 and D6.3 "Dissemination and **exploitation** plan" will be released at M30 and updated at M36.

2.3 Exploitation overview (software, products, spin-offs, etc.)

Use the table below to outline your current progress in the exploitation plan (see previous section): achievements so far and next steps. Fill in the goals foreseen in your plan for every year of your project and 3 subsequent years after the end of your project (column 1) and actual exploited results up to date (column 2).

Period	Planned goals	Actual exploited results
Year 1	Selection of FR pipeline and face image dataset.	Github project providing baseline implementation of selected FR recognition anchors and list of datasets to be used for training, validation, and testing. This is currently used by the consortium but will be useful for other researchers later to reproduce the project results.
Year 2	Module-level explainability	-
Year 3 *(if applicable)*	End-to-end explainability	-
Project end + 1 year		n/a
Project end + 2 year		n/a
Project end + 3 year		n/a

Describe project spin-off effects, for example:

- *Software and any other prototype;*
- *Standardization actions;*
- *National and international patents, licences, and other elements of intellectual property;*
- *Launching of product or service, new project, contract, etc.;*
- *Development of a new partnership;*
- *Creation of a platform available to a community;*
- *Company creation, spin-off companies, fund-raising.*

-

2.4 Other dissemination of results

Mention any communication actions, including the project website creation and management and the target audience.

Project website: https://xaiface.eurecom.fr/ Explanatory video (in progress)

chist-era

3. Resources and Funding

3.1. Project level (from project start)

Budget used				
N°	**Partner**	**Person.months**	**Total costs**	**Percentage of requested budget**
1	EURECOM [end of March 2022]	17	86000	31 %
2	JRS [end of March 2022]	2.57	30936,25	28 %
3	UNIVIE [until April 2022]	8	36669,05	33,5 %
4	IT	8	19295,41	21,5%
5	EPFL [reported till end of April 2022]	8	34600	26%

Comments on expenses

IT: The hired post-doc started only in September 2021 (not in May 2021 when the project started) and thus there is a delay in the use of these human resources.

EPFL: The activity financed by SNSF started in August 2021 (instead of May 2021 for the project) and a doctoral student has been hired since August 2021, who is full time on the project.