The CREENDER Tool for Creating Multimodal Datasets of Images and Comments

Alessio Palmero Aprosio
Fondazione Bruno Kessler
Trento, Italy
aprosio@fbk.eu

Stefano Menini
Fondazione Bruno Kessler
Trento, Italy
menini@fbk.eu

Sara Tonelli
Fondazione Bruno Kessler
Trento, Italy
satonelli@fbk.eu

Abstract

English. While text-only datasets are widely produced and used for research purposes, limitations set by image-based social media platforms like Instagram make it difficult for researchers to experiment with multimodal data. We therefore developed CREENDER, an annotation tool to create multimodal datasets with images associated with semantic tags and comments, which we make freely available under Apache 2.0 license. The software has been extensively tested with school classes, allowing us to improve the tool and add useful features not planned in the first development phase.[1]

Italiano. *Mentre i dataset testuali sono ampiamenti creati e usati per scopi di ricerca, le limitazioni imposte dai social media basati sulle immagini (come Instagram) rendono difficile per i ricercatori sperimentare con dati multimodali. Abbiamo quindi sviluppato CREENDER, un tool di annotazione per la creazione di dataset multimodali in cui immagini vengono associate a etichette semantiche e commenti, e che abbiamo reso disponibile gratuitamente con la licenza Apache 2.0. Il software è stato testato in un laboratorio con alcune classi scolastiche, permettendoci di ottimizzare alcune procedure e di aggiungere feature non previste nella prima release.*

1 Introduction

In the last years, the NLP community has started to focus on the challenges of combining vision and language technologies, proposing approaches towards multimodal data processing (Belz et al., 2016; Belz et al., 2017). This has led to an increasing need of multimodal datasets with high-quality information to be used for training and evaluating the developed systems. While several datasets have been created by downloading and often adding textual annotation to real online data (see for example the Flickr dataset[2]), this poses privacy and copyright issues, since downloading and using pictures posted online without the author's consent is often forbidden by social network privacy policies. Instagram terms of use, for example, explicitly forbid collecting information in an automated way without express permission from the platform.[3]

In order to address this issue, we present CREENDER, a novel annotation tool to create multimodal datasets of images and comments. With this tool it is possible to simulate a scenario where different users access the platform and are displayed different pictures, having the possibility to leave a comment and associate a semantic tag to the image. The same pictures can be shown to different users, allowing a comparison of their comments and online behaviour.

CREENDER can be used in contexts where simulated scenarios are the only solution to collect datasets of interest. One typical example, which we detail in Section 4, is the analysis of the online behaviour of teenagers and young adults, a task that poses relevant privacy issues since underage users are targeted. Giving the possibility to comment images in an Instagram-like setting without giving any personal information to register is indeed of paramount importance, and can be easily achieved with the tool presented in this paper.

[1] "Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)."

[2] https://yahooresearch.tumblr.com/post/89783581601/one-hundred-million-creative-commons-flickr-images

[3] See, for example, https://help.instagram.com/581066165581870.

Given its flexibility, CREENDER can however be used for any task where images need to be tagged and/or commented, and multiple annotations of the same image should be preferably collected.

2 Related Work

Several tools have been developed to annotate images with different types of information. Most of them are designed to be run only on a desktop computer and are meant to select parts of the picture to assign a semantic tag or a description, so that the resulting corpora can be used to train or evaluate image recognition or captioning software. In this scenario, users often need to be trained to use the annotation tool, which requires some time that is usually not available in specific settings like schools (Russell et al., 2008). Other tools for image annotation or captioning are web-based, like CREENDER, but the software is not available for download and must be used as a service. This paradigm can lead to privacy issues, as the data are not stored locally or on an owned server (Chapman et al., 2012). This could be problematic when the pictures to be annotated are copyright-protected or when users involved in the data collection do not want/cannot create an account with personal information. Finally, some software is not distributed open source, and could suddenly become unavailable or not usable when not maintained any more (Halaschek-Wiener et al., 2005; Hughes et al., 2018).

Regarding the datasets, Mogadala et al. (2019) focus on prominent tasks that integrate language and vision by discussing their problem formulations, methods, existing datasets, and evaluation measures, comparing the results obtained with different state-of-the-art methods. Ethical and legal issues on the use of pictures and texts taken from social networks are also relevant, as discussed in (Lyons, 2020; Prabhu and Birhane, 2020; Fiesler and Proferes, 2018). Our tool has been developed to address specifically also this kind of issues, preserving the privacy of users and avoiding the collection of real data.

3 Annotation Tool

The CREENDER tool can be accessed both via browser and mobile phone, so that users can use it even if no computer connected to Internet is available. The web interface is multi-language, since English, French and Italian are already included, while other language files can be added as needed. The interface language can be assigned at user level, meaning that the interface for users on the same instance can be configured in different languages.

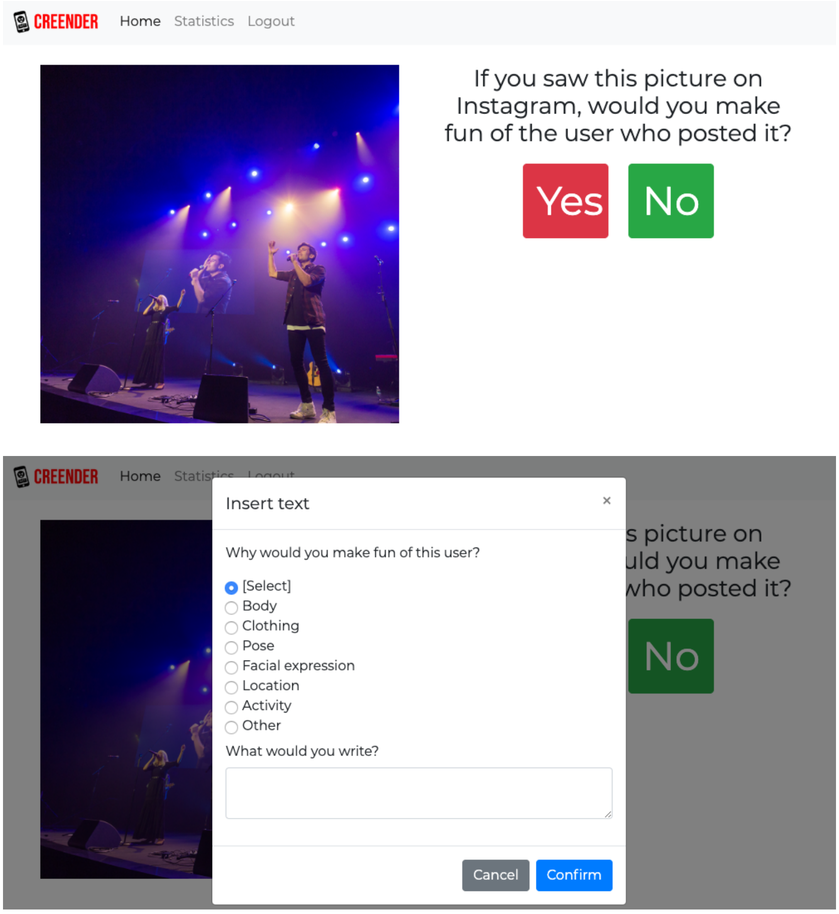
Once the tool is installed on a server, a super user is created, who can access the administration interface where the projects are managed with the password chosen during installation (see Figure 2).

For each project, on the configuration side, a set of photos (or a set of external links to images on the web) needs to be given to the tool. Then, one can set the number of users and the number of annotations that are required for each photo. Finally, the system assigns the photos to the users and creates the login information for them. Social login is also supported (only Google for now), so that there is no need to spread users and password: the administrator chooses a five-digit code and gives it to every annotator, that can then log in using the code and his/her social account.

Given a picture, the system can be set to perform three actions in sequence or in isolation, as needed by the task: *i)* the picture can be skipped by the user, so that no annotation is stored and the next one is displayed; *ii)* the user can insert free text associated to the image. This can be used to write a caption, comment the picture, list the contained objects, etc. *iii)* one or more pre-defined categories can be assigned to the picture. Categories can range from specific ones related to the portrayed objects (e.g. male, female, animals, etc.) to more abstract ones, like for example the emotions provoked by looking at the picture.

In the configuration screen, the administrator can edit the prompted questions and the possible answers, so that the tool can be used for a variety of different tasks.

Using the administration web interface, it is also possible to monitor the task with information about the number of annotations that each user has performed. This enables to check whether some users experience difficulties in the annotation, or if some annotators are anomalously fast (for example by skipping too many images). Once the annotation session is closed, the administrator can download the resulting corpus containing the images and the associated information. The export is available in three formats: SQL database, CSV, and JSON.



If you saw this picture on Instagram, would you make fun of the user who posted it?

Insert text

Why would you make fun of this user?

- ● [Select]
- ○ Body
- ○ Clothing
- ○ Pose
- ○ Facial expression
- ○ Location
- ○ Activity
- ○ Other

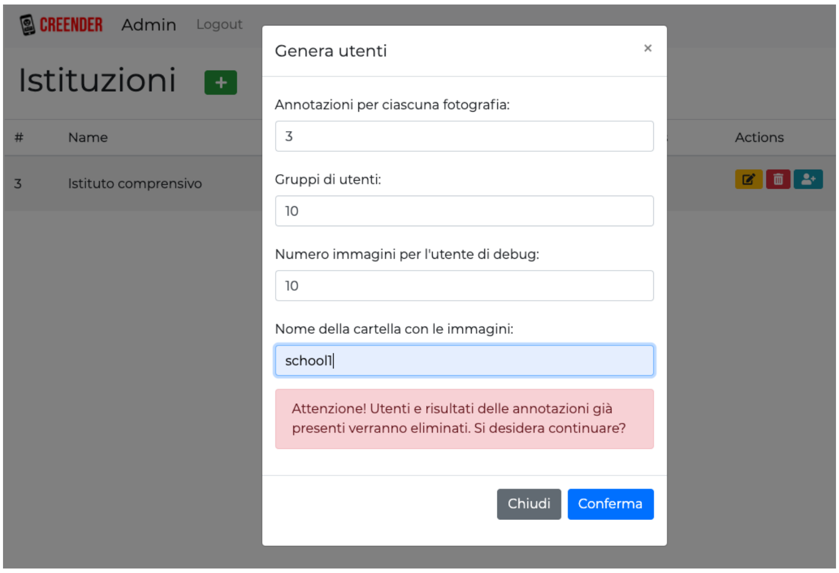
What would you write?

Figure 1: CREENDER interface configured for the collection of potentially offensive comments

4 Use Case: Creation of Offensive Posts

The CREENDER tool was used to collect abusive comments associated to images, simulating a setting like Instagram in which pictures and text together build an interaction which may become offensive. The data collection was carried out in several classes of Italian teenagers aged between 15 and 18, in the framework of a collaboration with schools aimed at increasing awareness on social media and cyberbullying phenomena (Menini et al., 2019). The data collection was embedded in a larger process that required two to three meetings with each class, one per week, involving every time two social scientists, two computational linguists and at least two teachers. During these meetings several activities were carried out with students, including simulating a WhatsApp conversation around a given plot as described in (Sprugnoli et al., 2018), commenting on existing social media posts, and annotating images as described in this paper.

Overall, 95 students were involved in the annotation. The sessions were organised so that different school classes annotated the same set of images, in order to collect multiple annotations on the same pictures. The pictures were retrieved from online sources and then manually checked by the researchers involved in the study to remove pornographic content. In the preparatory phase, the filtered pictures were uploaded in the CREENDER image folder. Then, a login and password were created for each student to be involved in the data collection and printed on paper, so that they could be given to each student before an annotation session without the possibility to associate login information with the students' identity. CREENDER was configured to first take a random picture from the image folder, and display it to the user with a prompt asking "*If you saw this picture on Instagram, would you make fun of the user who posted it?*". If the user selects "*No*", then the system picks another image randomly and the same question is asked. If the user clicks on "*Yes*", a second screen opens where the user is asked to specify the reason why the image would trigger



Genera utenti

Annotazioni per ciascuna fotografia: 3

Gruppi di utenti: 10

Numero immagini per l'utente di debug: 10

Nome della cartella con le immagini: school1

Attenzione! Utenti e risultati delle annotazioni già presenti verranno eliminati. Si desidera continuare?

Figure 2: The administration interface to define the number of users and the images per user

such reaction by selecting one of the following categories: "*Body*", "*Clothing*", "*Pose*", "*Facial expression*", "*Location*", "*Activity*" and "*Other*". Two screenshots of the interface are displayed in Figure 1. The user should also write the textual comment s/he would post below the picture. After that, the next picture is displayed, and so on. A screenshot of the tool configured for this specific task is displayed in Figure 1.

At the end of the activities with schools, all collected data were exported. The final corpus includes almost 17,912 images, 1,018 of which have at least one associated comment, as well as a trigger category (e.g. *facial expression*, *pose*) and the category of the subject/s (*female*, *male*, *mixed* or *nobody*). The number of annotations for each picture may vary between 1 to 4. A more detailed description of the corpus is reported in (Menini et al., 2021).

The use of CREENDER allowed a seamless and very fast data collection, without the need to send images to each student, to exchange or merge files and to install specific applications. On the other hand, the data collection with students, who used the online platform in classes while researchers were physically present and could check the flow of the interaction, was useful to improve the tool. Some bug fixes and small improvements were indeed implemented after the first sessions. For example, a small delay (2 seconds) was added after the image is displayed to the user and before the *Yes/No* buttons appear, so that users are more likely to look at the picture before deciding to skip it or not.

5 Release

The software is distributed as an open source package[4] and is released under the Apache license (version 2.0). The API (backend) is written in php and relies on a MySQL database. The web interface (frontend) is developed using the HTML/CSS/JS paradigm using the modern Bootstrap and VueJS frameworks.

The interface is responsive, so that one can use it from any device that can open web pages (desktop computers, smartphones, tablets).

6 Conclusions

In this work we present a methodology and a tool, CREENDER, to create multimodal datasets. In this framework, participants in online annotation sessions can write comments to images, assign pre-defined categories or simply skipping an image. The tool is freely available with an interface in three languages, and allows setting up easily annotation sessions with multiple users.

CREENDER has been extensively tested during activities with schools around the topic of cyberbullying, involving 95 Italian high-school students. The tool is particularly suitable for this kind of settings, where privacy issues are of paramount importance and the involvement of un-

[4] `https://github.com/dhfbk/creender`

derage people requires that personal information is not shared.

In the future, we plan to continue the annotation of images related to cyberbullying, creating and comparing subsets of pictures related to different topics (e.g. religious symbols, political parties, football teams). From an implementation point of view, we will extend the analytics panel, adding for example scripts for computing inter-annotator agreement.

Acknowledgments

Part of this work has been funded by the KID ACTIONS REC-AG project (n. 101005518) on "Kick-off preventIng and responDing to children and AdolesCenT cyberbullyIng through innovative mOnitoring and educatioNal technologieS". In addition, the authors want to thank all the students and teachers who participated in the experimentation.

References

Anya Belz, Erkut Erdem, Krystian Mikolajczyk, and Katerina Pastra, editors. 2016. *Proceedings of the 5th Workshop on Vision and Language*, Berlin, Germany, August. Association for Computational Linguistics.

Anya Belz, Erkut Erdem, Katerina Pastra, and Krystian Mikolajczyk, editors. 2017. *Proceedings of the Sixth Workshop on Vision and Language*, Valencia, Spain, April. Association for Computational Linguistics.

Brian E Chapman, Mona Wong, Claudiu Farcas, and Patrick Reynolds. 2012. Annio: a web-based tool for annotating medical images with ontologies. In *2012 IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology*, pages 147–147. IEEE.

Casey Fiesler and Nicholas Proferes. 2018. "participant" perceptions of twitter research ethics. *Social Media + Society*, 4(1):2056305118763366.

Christian Halaschek-Wiener, Jennifer Golbeck, Andrew Schain, Michael Grove, Bijan Parsia, and Jim Hendler. 2005. Photostuff-an image annotation tool for the semantic web. In *Proceedings of the 4th international semantic web conference*, pages 6–10. Citeseer.

Alex J Hughes, Joseph D Mornin, Sujoy K Biswas, Lauren E Beck, David P Bauer, Arjun Raj, Simone Bianco, and Zev J Gartner. 2018. Quanti.us: a tool for rapid, flexible, crowd-based annotation of images. *Nature methods*, 15(8):587–590.

Michael J Lyons. 2020. Excavating" excavating ai": The elephant in the gallery. *arXiv preprint arXiv:2009.01215*.

Stefano Menini, Giovanni Moretti, Michele Corazza, Elena Cabrio, Sara Tonelli, and Serena Villata. 2019. A system to monitor cyberbullying based on message classification and social network analysis. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 105–110.

Stefano Menini, Alessio Palmero Aprosio, and Sara Tonelli. 2021. A multimodal dataset of images and text to study abusive language. In *7th Italian Conference on Computational Linguistics, CLiC-it 2020*.

Aditya Mogadala, Marimuthu Kalimuthu, and Dietrich Klakow. 2019. Trends in integration of vision and language research: A survey of tasks, datasets, and methods. *arXiv preprint arXiv:1907.09358*.

Vinay Uday Prabhu and Abeba Birhane. 2020. Large image datasets: A pyrrhic win for computer vision?

Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3):157–173.

Rachele Sprugnoli, Stefano Menini, Sara Tonelli, Filippo Oncini, and Enrico Piras. 2018. Creating a WhatsApp Dataset to Study Pre-teen Cyberbullying. In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, pages 51–59. Association for Computational Linguistics.