The 2021 ImageCLEF Benchmark Multimedia Retrieval in Medical, Nature, Internet and Social Media Applications

Bogdan Ionescu[1], Henning Müller[2], Renaud Péteri[3], Asma Ben Abacha[4], Dina Demner-Fushman[4], Sadid A. Hasan[12], Obioma Pelka[5], Christoph M. Friedrich[5], Alba G. Seco de Herrera[7], Janadhip Jacutprakart[7], Vassili Kovalev[6], Serge Kozlovski[6], Jon Chamberlain[7], Adrian Clark[7], Antonio Campello[9], Hassan Moustahfid[8], Thomas Oliver[8], Abigail Schulz[8], Paul Brie[10], Raul Berari[10], Dimitri Fichou[10], Andrei Tauteanu[10], Mihai Dogariu[1], Liviu Daniel Stefan[1], Mihai Gabriel Constantin[1], Jérôme Deshayes[11], and Adrian Popescu[11]

[1] University Politehnica of Bucharest, Romania bogdan.ionescu@upb.ro
[2] University of Applied Sciences Western Switzerland (HES-SO), Switzerland
[3] University of La Rochelle, France
[4] National Library of Medicine, USA
[5] University of Applied Sciences and Arts Dortmund, Germany
[6] Belarussian Academy of Sciences, Belarus
[7] University of Essex, UK
[8] NOAA/ US IOOS, USA
[9] Wellcome Trust, UK
[10] teleportHQ, Romania
[11] CEA LIST, France
[12] CVS Health, USA

Abstract. This paper presents the ideas for the 2021 ImageCLEF lab that will be organized as part of the Conference and Labs of the Evaluation Forum — CLEF Labs 2021 in Bucharest, Romania. ImageCLEF is an ongoing evaluation initiative (running since 2003) that promotes the evaluation of technologies for annotation, indexing and retrieval of visual data with the aim of providing information access to large collections of images in various usage scenarios and domains. In 2021, the 19th edition of ImageCLEF will organize four main tasks: (i) a *Medical* task addressing visual question answering, a concept annotation and a tuberculosis classification task, (ii) a *Coral* task addressing the annotation and localisation of substrates in coral reef images, (iii) a *DrawnUI* task addressing the creation of websites from either a drawing or a screenshot by detecting the different elements present on the design, and a new (iv) *Aware* task addressing the prediction of real-life consequences of online photo sharing. The strong participation in 2020, despite the COVID pandemic, with over 115 research groups registering and 40 submitting over 295 runs for the tasks shows an important interest in this benchmarking campaign. We expect the new tasks to attract at least as many researchers for 2021.

Keywords: User awareness · medical image classification · medical image understanding· coral image annotation and classification · recognition of hand drawn website UIs · ImageCLEF benchmarking · annotated data

1 Introduction

The ImageCLEF evaluation campaign was started as part of the CLEF (Cross Language Evaluation Forum) in 2003 [6, 7]. It has been held every year since then and delivered many results in the analysis and retrieval of images [15, 12]. Medical tasks started in 2004 and have in some years been the majority of the tasks in ImageCLEF [10, 11]. The objectives of ImageCLEF have always been the multilingual or language-independent analysis of visual content. A focus has often been on multimodal data sets, so combining images with structured information, free text or other information that helps in the decision making, usually based on real user needs [14].

Since 2018, ImageCLEF uses the crowdAI (now migrated to AIcrowd[13]) platform to distribute the data and receive the submitted results. The system allows having an online leader board and gives the possibility to keep data sets accessible beyond competition, including a continuous submission of runs and addition to the leader board.

Over the years, ImageCLEF and also CLEF have shown a strong scholarly impact that was captured in [18, 19]. This underlines the importance of evaluation campaigns for disseminating best scientific practices. In the 2020 ImageCLEF campaign [11], despite the outbreak of the COVID-19 pandemic and lock-down during the benchmark, 115 teams registered, 40 teams completed the tasks and submitted over 295 runs. Although the number of registrations was lower than in 2019, the rate of the participants actually submitting runs increased by over 8 percentage points.

In the following, we introduce the four tasks that are planned for 2021[14], namely: ImageCLEFmedical, ImageCLEFcoral, ImageCLEFdrawnUI and the new ImageCLEFaware. Figure 1 captures with a few images the specificity of the tasks.

2 ImageCLEFmedical

The *concept detection task* concentrates on developing systems that are capable of predicting Unified Medical Language System (UMLS®) Concept Unique Identifiers (CUIs) on a given text. In 2021 the task will include a larger data set compared to 2020 [16]. The distributed corpus will be an extension of the Radiology Objects in Context (ROCO) [17] data set, that originates from image-caption pairs extracted from the PubMed Central Open Access Subset. The development data includes radiology images grouped into 7 sub-classes denoting the imaging

[13] `https://www.aicrowd.com/`
[14] `http://clef2021.clef-initiative.eu/`

acquisition technique, with a corresponding set of concepts. In 2021, the data set will be manually curated to reduce the data variability, something that participants had asked for in previous editions. The automatically predicted concepts can be further adopted as first steps towards the *Medical Visual Question Answering (VQA-Med)* task.

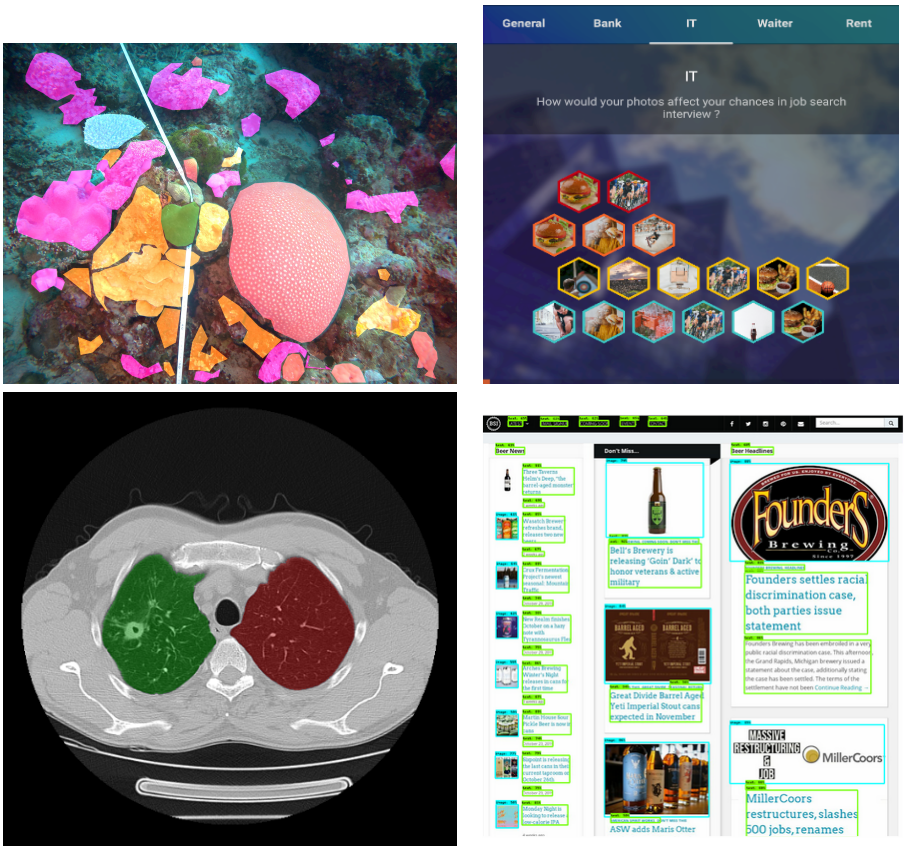

Fig. 1. Sample images from (left to right, top to bottom): ImageCLEFmedical with *(a slice of a chest CT with tuberculosis)*, ImageCLEFcoral with *(an example of an annotated coral reef image)*, ImageCLEFdrawnUI with recognition of UI elements from website screenshots, and ImageCLEFaware with an example of user photos and predicted influence when searching for an job in IT.

The Medical Visual Questions Answering task (VQA med [2]) will focus in 2021 on chest x-rays as they are the most commonly performed radiology exams. In this context, participants will be encouraged to use available resources in addition to the provided data in order to build robust VQA models dedicated to

chest x-rays. An additional objective will be to combine the concept detection task with the VQA task by using a common subset of radiology images.

The ImageCLEF tuberculosis task [13] will focus in 2021 on a larger data set than in 2020 and also on a larger number of concepts to extract for structured report generation. As in previous editions the task will use 3D Computed Tomography (CT) data of the chest. Lung masks will be supplied [8]and as in 2020 the report generation will be on a lung basis, so separate for left and right lung.

3 ImageCLEFcoral

The increasing use of structure-from-motion photogrammetry for modelling large-scale environments from action cameras attached to drones has driven the next-generation of visualisation techniques that can be used in augmented and virtual reality headsets. The main goal of the ImageCLEFcoral task since its first edition is to addresses this particular issue for monitoring coral reef structure and composition, in support of their conservation. In this third edition, it will follow a similar format as in the previous edition [3, 4] containing the same two subtasks with some modifications. The two tasks will be: *coral reef image annotation and localisation* and *coral reef image pixel-wise parsing*.

In the *coral reef image annotation and localisation* subtask, the participants are asked to annotate substrates in coral reef images using bounding boxes while in the *coral reef image pixel-wise parsing* subtask, participants need to submit a series of boundary image coordinates which form a single polygon around each identified substrate (see Figure 1). In both tasks, the participants will also identify the substrate type annotated in each coral reef image. The performance of the submitted algorithms will be evaluated using the PASCAL VOC style metric of intersection over union (IoU) and the mean of pixel-wise accuracy per class.

Previous editions of ImageCLEFcoral in 2019 and 2020 have shown improvements in task performance and promising results on cross-learning between images from different geographical regions. In 2021, the task will continue to explore how cross-learning can improve performance by offering supplemental data sets the participants may wish to use, as well as increased training data for the task itself. In addition, the training and test data will form the complete set of images required to form a 3D reconstruction of the environment. This allows the participants to explore novel probabilistic computer vision techniques based around image overlap and transposition of data points.

4 ImageCLEFdrawnUI

The increasing importance of User Interfaces (UIs) for companies highlights the need for novel ways of creating them. Currently, this activity can be slow and error prone due to the constant communication between the specialists involved in this field, e.g., designers and developers. The use of machine learning and automation could speed up this process and ease access to the digital space

for companies who could not afford it with today's tools. A first step to build a bridge between developers and designers is to infer the intent from a hand drawn UI (wireframe) or from a web screenshot. This is done by detecting atomic UI elements, such as images, paragraphs, containers or buttons.

Inspired by recent progress of machine learning usage for UI creation [1, 5], the previous edition of drawnUI challenged the participants to perform object detection on hand-drawn representations of websites (wireframes). The participant submissions offered promising results [9] and encouraged the further extension of the task at hand.

In this edition, two tasks are proposed to the participants, both requiring them to detect rectangular bounding boxes corresponding to the UI elements from the images. The first task, *wireframes annotation*, is a continuation of the previous edition, where 1,000 more wireframes are added to the existing 3,000 images of the data set. These new images will contain a bigger proportion of the rare classes to tackle the long tail problem found in the previous edition. For the second task we present the new challenge of *screenshot annotation*, where 10,000 screenshots of real websites were compiled into a data set by utilizing an in-house parser. Due to the nature of the web, the data set is noisy, e.g., some of the annotations correspond to invisible elements, while other elements have missing annotations. The training set will be provided without cleaning and will contain 8,000 images. The remaining images will be cleaned manually and split into validation and test subsets.

The performance of the algorithms will be evaluated using the standard mean Average Precision over IoU 0.50 and recall over IoU 0.50.

5 ImageCLEFaware

Images constitute a large part of the content shared on social networks. Their disclosure is often related to a particular context and users are often unaware of the fact that, depending on their privacy status, images can be accessible to third parties and be used for purposes which were initially unforeseen. For instance, it is common practice for employers to search information about their future employees online. Another example of usage is that of automatic credit scoring based on online data. Most existing approaches which propose feedback about shared data focus on inferring user characteristics and their practical utility is rather limited. We hypothesize that user feedback would be more efficient if conveyed through the real-life effects of data sharing. The objective of the task is to automatically score user photographic profiles in a series of situations with strong impact on her/his life.

This is the first edition of the task. A data set of 500 user profiles with 100 photos per profile will be created and annotated with an "appeal" score for a series of real-life situations via crowdsourcing. The averaged "appeal" score will be used to create a ground truth composed of ranked users in each modeled situation. The set will be split into train/validation/test and participants will be provided with the train and validation parts along with the associated rankings.

They will be required to provide the automatic ranking for the test subset. The objective of the task will be to produce an automatic ranking which is as closely correlated as possible to the manual ranking. Correlation will be measured using a classical measure such as the Pearson correlation coefficient.

Task-related resources will be provided by organizers to encourage participation of different communities. These resources include: (i) a labeled visual dataset which includes concepts relevant for raising users' awareness about sharing practices, (ii) visual concept ratings for each of the modeled situations, (iii) automatically extracted predictions for the photos with compose user profiles.

6 Conclusions

In this paper, we presented an overview of the upcoming ImageCLEF 2021 campaign. ImageCLEF has organized many tasks in a variety of domains over the past 18 years, from general stock photography, medical and biodiversity data to multimodal lifelogging. The focus has always been on language independent or multi-lingual approaches and most often on multimodal data analysis. 2021 has a set of interesting tasks that are expected to again draw a large number of participants. As in 2020, the focus for 2021 has been on the diversity of applications and on creating clean data sets to provide a solid basis for the evaluations.

Acknowledgement

Part of this work is supported under the H2020 AI4Media "A European Excellence Centre for Media, Society and Democracy" project, contract #951911.

References

1. Beltramelli, T.: pix2code : Generating Code from a Graphical User Interface Screenshot. Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems pp. 1–9 (2018)
2. Ben Abacha, A., Datla, V.V., Hasan, S.A., Demner-Fushman, D., Müller, H.: Overview of the vqa-med task at imageclef 2020: Visual question answering and generation in the medical domain. In: CLEF 2020 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece (September 22-25 2020)
3. Chamberlain, J., Campello, A., Wright, J.P., Clift, L.G., Clark, A., García Seco de Herrera, A.: Overview of ImageCLEFcoral 2019 task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, vol. 2380. CEUR-WS.org <http://ceur-ws.org>, Lugano, Switzerland (2019)
4. Chamberlain, J., Campello, A., Wright, J.P., Clift, L.G., Clark, A., García Seco de Herrera, A.: Overview of the ImageCLEFcoral 2020 task: Automated coral reef image annotation. In: CLEF2020 Working Notes. CEUR Workshop Proceedings, vol. 1166. CEUR-WS.org <http://ceur-ws.org>, Thessaloniki, Greece (September 22-25 2020)

5. Chen, C., Su, T., Meng, G., Xing, Z., Liu, Y.: From UI Design Image to GUI Skeleton : A Neural Machine Translator to Bootstrap Mobile GUI Implementation. International Conference on Software Engineering **6** (2018)
6. Clough, P., Müller, H., Sanderson, M.: The CLEF 2004 cross–language image retrieval track. In: Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.) Multilingual Information Access for Text, Speech and Images: Result of the fifth CLEF evaluation campaign. Lecture Notes in Computer Science (LNCS), vol. 3491, pp. 597–613. Springer, Bath, UK (2005)
7. Clough, P., Sanderson, M.: The CLEF 2003 cross language image retrieval task. In: Proceedings of the Cross Language Evaluation Forum (CLEF 2003) (2004)
8. Dicente Cid, Y., Jiménez del Toro, O.A., Depeursinge, A., Müller, H.: Efficient and fully automatic segmentation of the lungs in ct volumes. In: Goksel, O., Jiménez del Toro, O.A., Foncubierta-Rodríguez, A., Müller, H. (eds.) Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI. pp. 31–35. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (May 2015)
9. Fichou, D., Berari, R., Brie, P., Dogariu, M., Stefan, L., Constantin, M., Ionescu, B.: Overview of imageclefdrawnui 2020: the detection and recognition of hand drawn website uis task. In: CLEF2020 Working Notes, CEUR Workshop Proceedings. CEUR-WS. org, Thessaloniki (2020)
10. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), vol. 11438. LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 9-12 2019)
11. Ionescu, B., Müller, H., Péteri, R., Dang-Nguyen, D.T., Zhou, L., Piras, L., Riegler, M., Halvorsen, P., Tran, M.T., Lux, M., Gurrin, C., Chamberlain, J., Clark, A., Campello, A., de Herrera, A.G.S., Ben Abacha, A., Datla, V., Hasan, S.A., Liu, J., Demner-Fushman, D., Pelka, O., Friedrich, C.M., Cid, Y.D., Kozlovski, S., Liauchuk, V., Kovalev, V., Berari, R., Brie, P., Fichou, D., Dogariu, M., Stefan, L.D., Constantin, M.G.: ImageCLEF 2020: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020), LNCS Lecture Notes in Computer Science, Springer, Thessaloniki, Greece (2020)
12. Kalpathy-Cramer, J., García Seco de Herrera, A., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems: Overview of the medical image retrieval task at ImageCLEF 2004–2014. Computerized Medical Imaging and Graphics **39**(0), 55 – 61 (2015)
13. Kozlovski, S., Liauchuk, V., Dicente Cid, Y., Tarasau, A., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2020 - automatic CT-based report generation. In: CLEF2020 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Thessaloniki, Greece (September 22-25 2020)
14. Markonis, D., Holzer, M., Dungs, S., Vargas, A., Langs, G., Kriewel, S., Müller, H.: A survey on visual information search behavior and requirements of radiologists. Methods of Information in Medicine **51**(6), 539–548 (2012)

15. Müller, H., Clough, P., Deselaers, T., Caputo, B. (eds.): ImageCLEF – Experimental Evaluation in Visual Information Retrieval, The Springer International Series On Information Retrieval, vol. 32. Springer, Berlin Heidelberg (2010)
16. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2020 concept prediction task: Medical image understanding. In: CLEF2020 Working Notes. CEUR Workshop Proceedings, vol. 1166. CEUR-WS.org, Thessaloniki, Greece (September 22-25 2020)
17. Pelka, O., Koitka, S., Rückert, J., Nensa, F., Friedrich, C.M.: Radiology Objects in COntext (ROCO): a multimodal image dataset. In: Proceedings of the Third International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis (LABELS 2018), Held in Conjunction with MICCAI 2018. vol. 11043, pp. 180–189. LNCS Lecture Notes in Computer Science, Springer, Granada, Spain (September 16 2018)
18. Tsikrika, T., García Seco de Herrera, A., Müller, H.: Assessing the scholarly impact of ImageCLEF. In: CLEF 2011. pp. 95–106. Springer Lecture Notes in Computer Science (LNCS) (sep 2011)
19. Tsikrika, T., Larsen, B., Müller, H., Endrullis, S., Rahm, E.: The scholarly impact of CLEF (2000–2009). In: Information Access Evaluation. Multilinguality, Multimodality, and Visualization, pp. 1–12. Springer (2013)