Noname manuscript No.
(will be inserted by the editor)

GTCreator: a Flexible Annotation Tool for Image-based Datasets

Jorge Bernal[1], · Aymeric Histace[2] · Marc Masana[1] · Quentin Angermann[2] · Cristina Sánchez-Montes[3] · Cristina Rodríguez de Miguel[3] · Maroua Hammami[2] · Ana García-Rodríguez[3] · Henry Córdova[3] · Olivier Romain[2] · Gloria Fernández-Esparrach[3] · Xavier Dray[2,4] · F. Javier Sánchez[1]

Received: date / Accepted: date

Abstract Purpose: Methodology evaluation for decision support systems for health is a time consuming-task. To assess performance of polyp detection methods in colonoscopy videos, clinicians have to deal with the annotation of thousands of images. Current existing tools could be improved in terms of flexibility and ease of use. **Methods:** We introduce GTCreator, a flexible annotation tool for providing image and text annotations to image-based datasets. It keeps the main basic functionalities of other similar tools while extending other capabilities such as allowing multiple annotators to work simultaneously on the same task or enhanced dataset browsing and easy annotation transfer aiming to speed up annotation processes in large datasets. **Results:** The comparison with other similar tools shows that GTCreator allows to obtain fast and precise annotation of image datasets, being the only one which offers full annotation editing and browsing capabilites. **Conclusions:** Our proposed annotation tool has been proven to be efficient for large image dataset annotation, as well as showing potential of use in other stages of method evaluation such as experimental setup or results analysis.

Keywords Annotation tool · Validation Framework · Benchmark · Colonoscopy · Evaluation

[1] Computer Vision Center and Universitat Autonòma de Barcelona, Spain
{jorge.bernal,marc.masana,javier.sanchez}@cvc.uab.cat
[2] ETIS lab, ENSEA, University of Cergy-Pontoise, CNRS, Cergy, France
{quentin.angermann,maroua.hammami,olivier.romain,aymeric.histace}@ensea.fr
[3] Endoscopy Unit, ICMDiM, Hospital Clnic, IDIBAPS, CIBEREHD, University of Barcelona, Spain
{mgfernan,crsanchez,crodrigm,hcordova}@clinic.cat, anagrod4@gmail.com
[4] St. Antoine Hospital, APHP, Paris, France
xavier.dray@aphp.fr

1 Introduction

The development and validation of decision support systems (DSS) for health has seen an increase in interest in recent years due to the improvement in computing capabilities and the evolution of imaging technology. These systems provide objective information to clinicians to help them in the different stages of a clinical procedure, from intervention planning (automatic location of cancer nodules in a CT scan [1]), to assistance during in-vivo explorations (renal lesion detection in ultrasound images [2]), or in decision making processes (resection of colon polyps with automatic histology prediction [3]).

The scope of our research is the development and validation of intelligent systems for colonoscopy. These systems couple the expertise of clinicians and technicians to develop efficient solutions that can help to tackle some of the most demanding clinical needs. For the case of colonoscopy, polyp detection and in-vivo histology prediction [4] are being identified as the key areas in which a computational system could be of great help.

Several computational methods have been proposed to tackle both polyp detection (the interested reader can find a comparison between several detection methodologies in [5]) and in-vivo histology prediction [6]. Despite the large number of approaches, none of them is used in clinical facilities. The two main reasons for it being low practical feasibility due to not meeting real-time constraints, and the requirement of excessive computational resources. Furthermore, the lack of uniform and public frameworks hinder the performance assessment of these systems before their use by clinicians.

With respect to the latter, several attempts have been made to develop public validation frameworks, especially in the context of challenges in medical imaging conferences such as MICCAI or ISBI. However, there is still a lack in terms of quality and quantity with respect to clinically generated medical image annotations. For the case of colonoscopy video analysis ground truth definition is a highly time consuming task, requiring clinicians to precisely annotate thousands of images.

Several tools have been proposed to assist in ground truth generation which can be divided into two groups: general and domain-specific. General tools include LabelMe [7] or VGG Image Annotator (VIA) [8], while tools such as Ratsnake [9] or Arthemis [10] were specifically designed for medical imaging domains. None of the tools from the latter group were designed along their final users, which results in powerful applications that, unfortunately, are not practically usable due to being complex to use.

In conclusion, the contributions of this article are two-fold:

- GTCreator, a novel flexible tool for providing image and text-based annotations of image datasets,
- a comparison study between image annotation tools.

GTCreator has been designed to ease the task of Ground Truth (GT) annotation and management. It allows a great flexibility with respect to the number and type of annotations to be provided for each image, making it suitable to use for any image dataset annotation task (see Section 2.1.2).

To show the benefits of our tool, we present a study in which we compare qualitatively and quantitatively six different annotation tools. In the proposed use case of annotation of colonoscopy videos and still images by experts, our tool keeps the main functionalities provided by other tools as well as incorporating new ones such as annotation transfer or full annotation editing capabilities (see Section 3).

The rest of the paper is structured as follows: Section 2 details the features and usage of GTCreator. We present the results of the comparison study against other similar tools in Section 3. Conclusions are presented in Section 4.

2 GTCreator: A Flexible Image Annotation Tool

2.1 Overview of GTCreator

GTCreator is a tool for creating and managing annotations of image databases. The result of an image annotation is a set of metadata (image and text-based) which will be associated to that image. The main innovation of GTCreator is its flexibility in comparison to other existing tools, as it allows to freely determine the number and type of metadata to be associated to each image during the annotation task. This makes our tool suitable for its use in any image annotation task.

We show in Fig. 1 our proposed pipeline for image dataset annotation. GTCreator consists of 3 stages after dataset collection: 1) definition file configuration for ground truth definition, 2) image annotation using graphical user interface (GUI) and 3) ground truth data exportation.

2.1.1 Definition file configuration

The definition file is the key element of GTCreator and consists of three different parts: the header containing global information of the annotation task, the list of metadata to be associated to each image, and the list of dataset images to be annotated.

The header contains: a) name of the annotation set to be generated, b) description of the annotation task and c) the relative paths in which both dataset images and annotations will be stored. This allows to univocally identify and locate the annotation task and the data. To cover the range of possible annotation tasks, GTCreator admits three types of metadata:

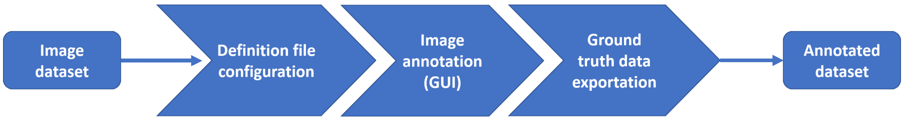


Fig. 1 Proposed pipeline for image dataset annotation using GTCreator.

- images/masks, to represent the objects in the image. This annotation type admits two different options: annotation through freehand drawing or by using predetermined shapes.
- text, to provide formatted information associated to the image. This category subtypes' include bool (for binary metadata), string, or list of values.
- semantic labels, to classify each of the predefined regions in the image.

To make annotations by multiple users easier and more coherent, each metadata can have an associated description (visual and/or text) to provide an example of how the annotation process is expected to be done. Finally, the definition file includes the list of dataset images to be annotated, and image names can be added manually or automatically through the GTCreator GUI.

2.1.2 Image annotation through Graphical User Interface

The GUI of our tool has been designed to allow an easy access to all functionalities. The GUI is divided into six different areas, as shown in Fig. 2:

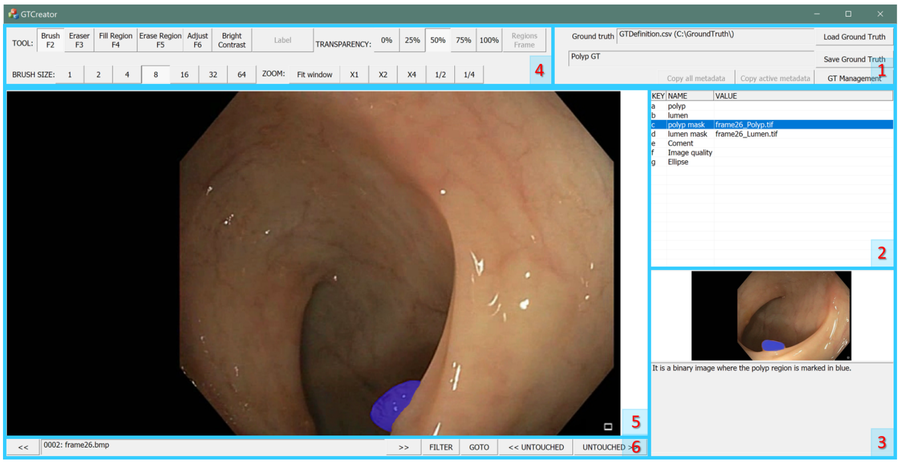
1. Definition file and database management operations, which include the possibility of adding images to the annotation task.
2. Metadata editing area, in which clinicians can modify the value of the different metadata associated to each image.
3. Annotation assistance area, which might include a representative image and additional text to guide the annotation task.
4. Image annotation tools, including the possibility to change image scale change, mask transparency, brush size as well as additional capabilities to improve freehand contour definition and to change local image contrast to enhance image visibility.
5. Image annotation area.
6. Database browsing tools, including metadata-based filtering and shortcuts to ease navigation through non-annotated images.

An annotation session starts by loading the definition file associated to the task. This also allows to resume an already started annotation session. To ease the annotation of video datasets, we allow annotation transfer between consecutive images. Finally, considering that a large annotation task might be divided among different annotators, annotation merging from different observers is also supported. All those functionalities are accessible through the described areas of the GUI. [1].

2.1.3 Ground truth data exportation

Once the desired dataset has been annotated, it is important that it can easily be used widely by both clinical and technical users. As it will be discussed in Section 3, the format used to export the annotation data should be simple and easy-to-use. Therefore, all information related to the annotation task is

[1] a demo version of GTCreator is available at `https://tinyurl.com/GTCreator`



Ground truth: GTDefinition.csv (C:\GroundTruth\)

Polyp GT

KEY	NAME	VALUE
a	polyp	
b	lumen	
c	polyp mask	frame26_Polyp.tif
d	lumen mask	frame26_Lumen.tif
e	Coment	
f	Image quality	
g	Ellipse	

It is a binary image where the polyp region is marked in blue.

0002: frame26.bmp

Fig. 2 Distribution of the different areas of GTCreator Graphical User Interface as described in Section 2.1.2.

organized as a table and stored in a CSV file, which is compatible with most commonly used software environments (Matlab, Python or SPSS). This allows the use of the annotations for method validation purposes in many frameworks without further interaction with the data. In the CSV file, rows correspond to the different images in the dataset and the columns to the different metadata included in the definition file. Image-based annotations are stored as binary masks in TIFF files to ensure efficiency in data storage.

The way our tool is designed allows modification of the definition file during the annotation process while keeping intact the information already stored. In this case, the user should only complete the new metadata information for the already labeled images.

2.2 Use-case scenarios

GTCreator has been mainly designed to support annotation tasks but it can also be used in other stages of the development and validation of an image processing method, as shown in Fig. 3. More precisely, we foresee the following scenarios in which our tool could play a key role:

- Provide image and text-based annotations of image datasets (**GT Creation & Annotation**). The interface of our tool allows to resume an already started annotation process as well as incorporating new metadata on the fly without having to start the annotation process from scratch. For the case of large datasets, our tool incoroporates annotation merging capabilities in case annotation task was split into different annotators.
- Annotation review (**GT Revision**). Our tool incorporates metadata types that fosters annotation review by other users. In an example case, a novice clinician marks mark an image for later inspection by an expert by using a

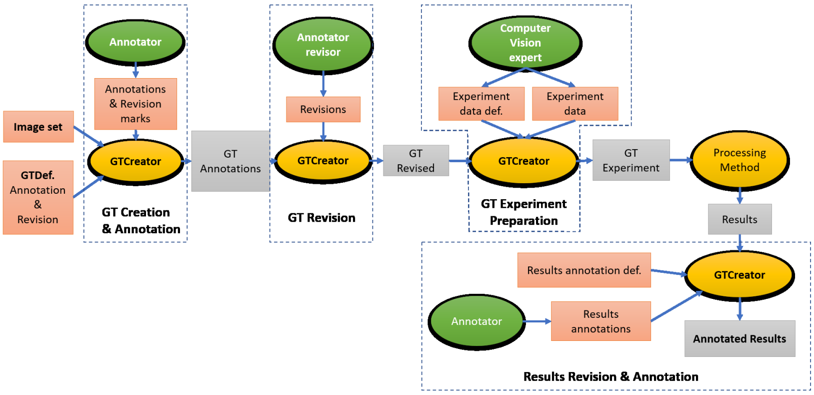


Fig. 3 Graphical scheme showing all the potential interactions of GTCreator through a complete validation of an image processing method.

boolean metadata type. The expert can easily navigate through the images marked for inspection by using the filtering capabilities of our tool.

- The CSV file produced as a result of dataset annotation can be used to perform experimental setups for methodology validation. For instance, specific text-based metadata (set by the user or imported from an external source) can be used to split a dataset into training, validation and test sets (**GT Experiment Preparation**).
- Our tool can also be used to assist in method validation (**Result Revision and Annotation**) by using the semantic label metadata type. In this case, the user can assign any of the predefined labels (set in the definition file) to each of the regions of interest provided by a given image processing method. The user can also browse through numerical or textual output values using the filtering functions once these values are imported into the corresponding CSV file.

3 Comparison study of annotation tools

The development of manual image annotation tools has already been covered in the literature. We can divide existing tools into two groups: domain-specific (such as Arthemis [10] or iPad [11]) or general (such as LabelMe [7] or VGG Image Annotator (VIA) [8]). We present in this section a comparative study between GTCreator and other similar tools.

The inclusion criteria for the comparison has been the free availability of the tool, ease of installation and use by means of the potential final users. Five tools were finally selected: RatSnake [9], LabelMe [7], VGG Image Annotator (VIA) [8], Video Image Annotation Tool (VIAT) [12], and ImageJ [13].

RatSnake is based on the use of a snake model to enable semiautomatic segmentation, being model customization capabilities key to succeeding in providing accurate object masks. LabelMe is a web-based tool allowing individual region annotation (which needs no installation), being its usability compromised by requiring a network connection. VIA is another web-based tool which stands out for its simplicity in use, including the possibility of adding formated text annotations. VIAT offers great flexibility in the type of image annotations to be created, ranging from polygons to freehand annotations, supporting semiautomatic segmentation capabilities. Finally, ImageJ offers a great variability in output formats and incorporates a full image processing suite to support the annotation process.

3.1 Qualitative comparison

Table 1 presents a qualitative comparison of the mentioned annotation tools. We describe next some of the most relevant differences among those tools.

A desirable feature of an annotation tool is to allow users a fast and easy navigation experience through the images to be annotated. With respect to this, we can clearly divide the tools into two groups: those which allow the creation of image collections (LabelMe, VIA and GTCreator) and those which require the user to manually open images and saving results. It is clear that, for the case of large dataset annotation, the use of tools from the second group is less practical since a large amount of the annotation time would be devoted to tasks outside actual image annotation.

Regarding the type of annotations that can be generated, VIA and GTCreator are the only tools allowing the inclusion of formatted text metadata. LabelMe allows to incorporate semantic labels to the objects marked by the annotator and add unformatted comments for each annotation. Finally VIAT allows to include a general unformatted comment for each image to be annotated. Despite other tools offering the possibility of including text annotations, GTCreator is the only one allowing browsing the dataset using filters defined according to text metadata values.

All tools allow image annotation using pre-determinated shapes but only three of them (VIAT, ImageJ and GTCreator) allow the user to draw freely the contour of the object to be annotated. The use of pre-determinated shapes allows a faster dataset annotation at the cost of losing annotation precision, which we consider is key when validating segmentation methods. Our tool is the only one supporting complete pixel-wise editing of image annotations, whereas other approaches simply allow slight polygon points displacement.

Some tools (RatSnake, VIA and GTCreator) allow annotation transfer among images, which is specially useful when providing image annotations in video datasets in which variability between consecutive frames tends to be low. GTCreator is the only tool allowing text metadata transfer between images.

The majority of tools accept all common image input formats with the exception of LabelMe and VIAT with respect to TIFF images, which needs of image format conversion. On ground truth data exportation, only GTCreator and RatSnake directly provide as a result the binary ground truth masks. The rest of tools require further file editing operation, which are not always immediate for the final user. Both GTCreator and VIA allow users to also

Feature	RatSnake	LabelMe	VIA	VIAT	ImageJ	GTCreator
Annotation types	Image masks	Image masks, semantic labels	Image masks, formatted text	Image masks, unformatted text	Image masks	Image mask, formatted text
Mask annotation	Polygon	Polygon	Polygon, predetermined shapes	Polygon, predetermined shapes, freehand	Polygon, predetermined shapes, freehand	Polygon, predetermined shapes, freehand
Mask editing	Annotation transfer	None	Annotation transfer	None	Annotation transfer	Annotation transfer, pixelwise editing
Dataset browsing	Single image	Collection	Collection	Single image	Single image	Collection
Input format	BMP, JPEG, PNG, TIFF	BMP, JPEG, PNG	BMP, JPEG, PNG, TIFF	BMP, JPEG, PNG	BMP, JPEG, PNG, TIFF	BMP, JPEG, PNG, TIFF
Output format	Binary masks	XML file	CSV file	XML file	Text file	Binary masks, CSV file
Extra features	Semantic ontology	Semi-automatic segmentation	None	MPEG-7 descriptors	Image processing suite	Filtering-based browsing, annotation merging and reviewing

Table 1 Comparison of mean and standard deviation of the annotation precision with respect to the annotation tool.

obtain CSV files with the full annotation results (text metadata values and mask annotation image names).

Finally, some of the tools include additional capabilities aiming to provide an extra value to the annotation process. For instance, RatSnake allows the creation of a semantic ontology for the annotations and VIAT provides directly the output of MPEG-7 descriptors for the target image. With respect to image processing capabilities, LabelMe incorporates semi-automatic segmentation capabilities whereas ImageJ includes a full image processing suite. Furthermore, the use of some of these processing capabilities require excessive user training which might prevent less experienced users from actually taking profit from them. GTCreator incorporates some basic image processing capabilities such as the possibility of adjusting free hand-made contours to actual image contours using watershed segmentation with markers or the possibility to change image contrast locally.

3.2 Quantitative comparison

3.2.1 Experimental Setup

To explore practical differences in image annotation, we performed an experiment in which 6 different experts were asked to annotate 6 different sets of images, extracted from the datasets of GIANA challenge[2]. Each set consists of 13 different images divided into two subsets: ten consecutive standard definition frames from a colonoscopy video, and three high definition images. All images contain a polyp and each set was chosen to present a similar difficulty level. The annotation task consists of generating a mask covering the polyp. Fig. 4 shows a set of images used in the experiment.

The annotation experiment was performed as follows: each annotator would use a different set of images for each of the tools included in the comparison.

[2] available at `https://giana.grand-challenge.org/`

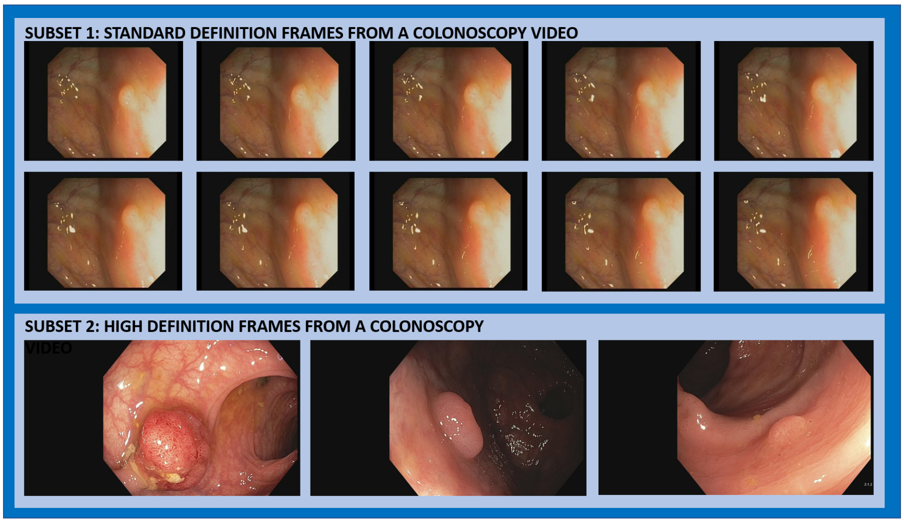

SUBSET 1: STANDARD DEFINITION FRAMES FROM A COLONOSCOPY VIDEO
SUBSET 2: HIGH DEFINITION FRAMES FROM A COLONOSCOPY VIDEO

Fig. 4 Content example of a set of images used in the comparison study.

Those sets were shared across all annotators and they were randomly assigned to each tool so that no set and tool would be repeated for two different annotators. Each of the tools was presented to the annotators before the start of the experiment and, during the experiment, support was given if necessary to explain where the boundaries of the polyps were.

Quantiative comparison was based on the measurement of annotation time and precision. For annotation time, a separate measurement was done to account for the time spent in the interaction with the system, like dataset browsing. All annotators performed the experiment under the same equipment. An expert clinician, not participating in the annotation experiment, was asked to provide pixel-wise annotations of the different images using common image editing software. The different annotations generated in the experiment were compared to this clinical ground truth using common segmentation metrics (DICE coefficient, Jaccard index).

3.2.2 Experimental results

Annotation time results with respect to the annotation tool can be found in Table 2. We can observe a big difference in mean total annotation time between the group of tools that include the use of image collections (LabelMe, VIA and GTCreator) with respect to the ones that do not. This latter group of tools (RatSnake, VIAT and ImageJ) all present a mean interaction time per image higher than 28 seconds. As results show, some of these tools would benefit greatly if this interaction time was removed. As an example, mean total annotation time for ImageJ would be reduced to its half.

GTCreator appears as the tool with a lowest mean total annotation time, followed by VIA. It is worth to mention that this difference is higher if we only consider video sequence subsets. Both GTCreator and VIA benefit from

	RatSnake	LabelMe	VIA	VIAT	ImageJ	GTCreator
Total annotation time (mean ± standard deviation)						
All	13′18″ ± 125″	6′3″ ± 88″	4′54″ ± 54″	12′9″ ± 96″	10′11″ ± 81″	**4’52”** ± 56″
VidFr	9′38″ ± 120′	4′12″ ± 75″	3′10″ ± 51″	8′30″ ± 88″	7′35″ ± 68″	**2’59”** ± 54″
HDFr	3′38″ ± 19″	1′52″ ± 27″	1′44″ ± 33″	3′48″ ± 30″	2′40″ ± 22″	**1’43”** ± 14″
Interaction time per image (mean ± standard deviation)						
	27″ ± 2.4	**0”** ± 0″	**0”** ± 0*s*	29″ ± 4.1″	28″ ± 3.4″	**0”** ± 0*s*
Actual image annotation time (minutes and seconds)						
All	7′31″ ± 105″	6′3″ ± 88″	4′54″ ± 54″	5′51″ ± 75″	**4’13”** ± 76″	4′52″ ± 56″
VidFr	5′12″ ± 107″	4′12″ ± 75″	3′10″ ± 51″	3′40″ ± 70″	3′01″ ± 66″	**2’59”** ± 54″
HDFr	2′18″ ± 13″	1′52″ ± 27″	1′44″ ± 33″	2′13″ ± 21″	**1’17”** ± 23″	1′43″ ± 14″

Table 2 Annotation time with respect to the annotation tool. VidFr stands for video sequence frames, HDFr for still HD images.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
All	8′14″ ± 199″	**10’16”** ± 280″	7′50″ ± 218″	7′36″ ± 223″	8′41″ ± 267″	8′50″ ± 217″
VidFr	5′27″ ± 158″	**7’54”** ± 216″	5′32″ ± 173″	5′10″ ± 174″	5′49″ ± 203″	6′10″ ± 165″
HDFr	2′37″ ± 56″	2′27″ ± 68″	2′18″ ± 50″	2′25″ ± 53″	**2’51”** ± 67″	2′39″ ± 56″

Table 3 Mean and standard deviation of the annotation time with respect to the set to be annotated. VidFr stands for video sequence frames, HDFr for still HD images. Bold results point to highest values of annotation time which we associate to higher difficulty

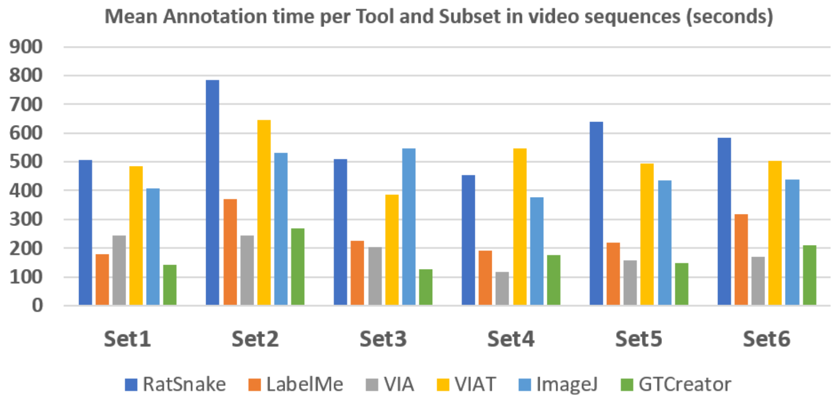


Fig. 5 Mean annotation time with respect to image subset and annotation tool

including annotation transfer capabilities between consecutive frames, which points into this feature to be key for any given annotation tool.

In order to better understand the reason behind differences between tools, we present results respect to sets instead of annotators in Table 3. We observe that set 2 of the video sequences appears to be the most difficult one as it has the largest mean annotation time. The small differences between mean annotation time among still frame subsets indicate a similar difficulty level.

Finally we show in Fig. 5 a break down of the mean annotation time per tool and video sequence subset. The graph shows how GTCreator is the tool with smaller mean annotation time in 3 out of 6 subsets, followed by VIA.

Annotation precision results with respect to the annotation tool can be found in Table 4. We observe that GTCreator is the tool which achieves the highest score in both DICE and Jaccard metrics, showing the robustness in annotation quality when generated by our tool. Differences between GTCreator and the rest of the tools increase if we only consider video sequence subsets.

	RatSnake	LabelMe	VIA	VIAT	ImageJ	GTCreator
DICE score (mean ± standard deviation)						
All	0.886 ± 0.040	0.860 ± 0.080	0.876 ± 0.059	0.848 ± 0.068	0.852 ± 0.082	**0.908**±0.032
VidFr	0.870 ± 0.058	0.851 ± 0.103	0.862 ± 0.076	0.824 ± 0.089	0.832 ± 0.099	**0.899**±0.037
HDFr	0.935 ± 0.026	0.917 ± 0.022	0.924 ± 0.015	0.927 ± 0.038	0.920 ± 0.031	**0.937**±0.029
Jaccard Index (mean ± standard deviation)						
All	0.938 ± 0.025	0.925 ± 0.051	0.932 ± 0.035	0.915 ± 0.042	0.917 ± 0.052	**0.951**±0.017
VidFr	0.929 ± 0.035	0.916 ± 0.064	0.924 ± 0.046	0.901 ± 0.056	0.905 ± 0.063	**0.946**±0.002
HDFr	0.965 ± 0.009	0.956 ± 0.013	0.960 ± 0.008	0.960 ± 0.024	0.958 ± 0.017	**0.967**±0.010

Table 4 Mean and standard deviation of the annotation precision with respect to the annotation tool. VidFr stands for video sequence frames, HDFr for still HD images.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
DICE score (mean ± standard deviation)						
All	0.921 ± 0.011	0.928 ± 0.012	0.871 ± 0.027	0.887 ± 0.026	0.869 ± 0.014	**0.761**±0.058
VidFr	0.915 ± 0.014	0.936 ± 0.008	0.812 ± 0.032	0.869 ± 0.033	0.847 ± 0.019	**0.720**±0.071
HDFr	0.939 ± 0.014	**0.900** ± 0.023	0.937 ± 0.007	0.942 ± 0.011	0.940 ± 0.007	0.907 ± 0.026
Jaccard Index (mean ± standard deviation)						
All	0.959 ± 0.006	0.962 ± 0.007	0.934 ± 0.014	0.939 ± 0.015	0.929 ± 0.008	**0.861**±0.038
VidFr	0.955 ± 0.007	0.967 ± 0.005	0.919 ± 0.019	0.929 ± 0.018	0.917 ± 0.011	**0.839**±0.004
HDFr	0.969 ± 0.006	**0.945** ± 0.015	0.967 ± 0.004	0.970 ± 0.006	0.969 ± 0.004	0.948 ± 0.013

Table 5 Mean and standard deviation of the annotation precision with respect to the set to be annotated. VidFr stands for video sequence frames, HDFr for still HD images. Bold results point to lowest values for each of the metrics which we associate to higher difficulty.

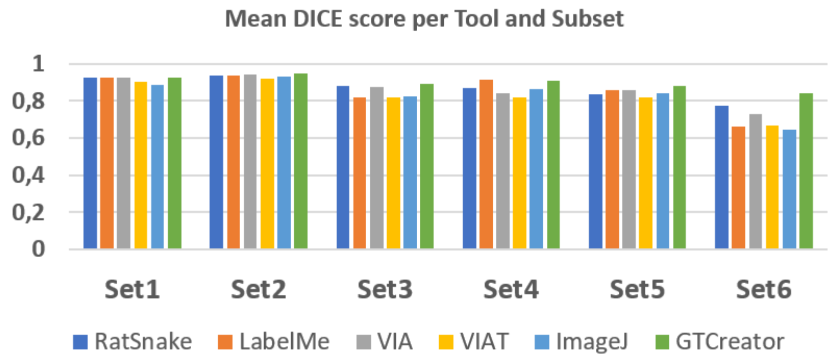


Fig. 6 Mean DICE score with respect to image subset and annotation tool

We associate these differences to the intensive use of the annotators of pixel-wise mask editing capabilities offered by GTCreator, especially for the case of small polyps. In this case, a precise delimitation of the polyp contour can be more difficult to achieve by polygon approximation offered by the other tools. Differences among tools are smaller when only still frames are considered.

As for annotation time, we study in Table 5 the dependence of annotation precision with respect to the specific subset to be annotated. We observe how video sequence of subset 6 appears to be the most difficult whereas difficulty appears to be balanced between still frame subsets.

Finally we show in Fig. 6 a break down of the mean DICE score per tool and video sequence subset. The graph shows how GTCreator is the tool with highest mean DICE score in all subsets, being this difference remarkable for the case of the mentioned subset 6.

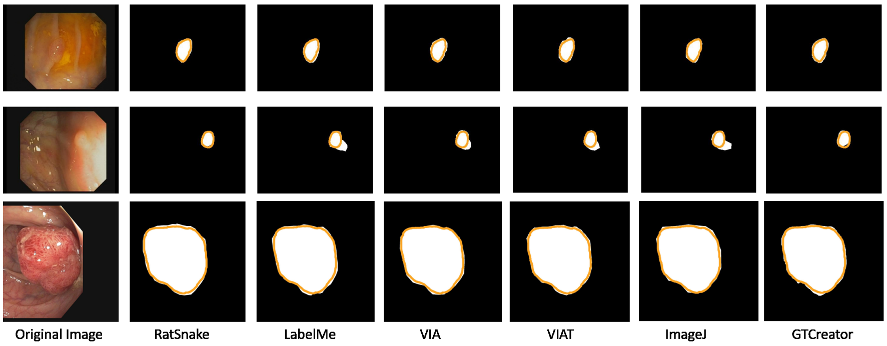


Fig. 7 Comparison between annotations performed with the different tools and the ground truth provided by an external expert. In annotation images, pixels in white represent those marked by the annotators whereas the orange contour represents the ground truth for the particular image.

We present in Fig. 7 a graphical comparison of the annotations generated with the different proposed tools. We have chosen three different representative examples: first row shows annotation results of an image belonging to a subset in which all annotation tools lead to achieve good annotations. Second rows shows an example of an image from subset 6 in which all annotators found difficulties when delimiting the area occupied by the polyp. Finally the last row shows an example of HD image annotation.

From these examples we can observe that indeed there are no big differences in annotation quality when the polyp appears clearly defined in the image. With respect to the example image from subset 6. In this case we can observe clear discrepancies between annotators and the external expert which provided the ground truth, which suggest that both having several annotators for a same image and annotation review and editing capabilities are key to generate high quality annotations of image datasets. Finally, HD image annotation task allows us to observe better the limitation of some tools when providing high quality pixe-wise annotations. We can observe steep contour direction changes in those tools which allow less flexibility in mask generation using pre-determined shapes such as VIAT or ImageJ; in these cases, the final mask is obtained by joining the different points marked by the user requiring a large amount of user interaction by marking several consecutive points to obtain a smoother final contour.

4 Conclusion and perspectives

Evaluating decision support systems for health is a challenging task, mainly due to the lack of public annotated datasets. Video annotation by clinicians is a high time-consuming task as precise masks have to be provided for thousand of images. We have proposed an annotation tool designed to ease annotation by clinicians while keeping the functionalities of other existing tools. Flexibility, efficient data management and browsing capabilities allow our tool to be used during the main stages of method evaluation for any image domain.

We have performed a comparison study to rank our tool among similar ones with the conclusion of GTCreator being the tool which offers the best compromise between annotation time and precision. Easy image browsing and the inclusion of image editing capabilities play a key role in the generation of fast and precise annotations.

As future work, we plan to extend GTCreator by incorporating image processing tools to support annotation tasks as well as video annotation capabilities. We plan to use GTCreator to propose new benchmarks in colonoscopy image analysis such as polyp segmentation in high definition images or to extend the current one on automatic histology prediction.

Funding

This work has been funded by Spanish Government through iVENDIS (DPI2015-65286-R), DeepMTL (TIN2016-79717-R) and HISINVIA(PI17/00894) projects, Catalan government through SGR-2017-1669 , SGR-2017-653 and CERCA programme, Région Île de France through SATT funding ”iPolyp” (Project 184). A. Histace and J. Bernal acknowledge the Institute of Advanced Studies from UCP (Invited Prof. Position grant) as well as Initiative Paris Seine through which the position was obtained in the context of ”iPolyp”. M. Masana acknowledges 2017FIB-00218 grant of Generalitat de Catalunya. We also acknowledge the generous GPU support from NVIDIA.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

References

1. El-Regaily, S.A., Salem, M.A., Abdel Aziz, M.H., Roushdy, M.I.: Survey of computer aided detection systems for lung cancer in computed tomography. Current Medical Imaging Reviews **14**(1) (2018) 3–18
2. Gui, L., Yang, X.: Automatic renal lesion segmentation in ultrasound images based on saliency features, improved lbp, and an edge indicator under level set framework. Medical physics **45**(1) (2018) 223–235
3. Sánchez, F.J., Bernal, J., Sánchez-Montes, C., de Miguel, C.R., Fernández-Esparrach, G.: Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. Machine Vision and Applications **28**(8) (Nov 2017) 917–936
4. Mori, Y., Kudo, S.e., Berzin, T.M., Misawa, M., Takeda, K.: Computer-aided diagnosis for colonoscopy. Endoscopy **49**(08) (2017) 813–819
5. Bernal, J., Tajkbaksh, N., Snchez, F.J., Matuszewski, B.J., Chen, H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., Pogorelov, K., Choi, S., Debard, Q., Maier-Hein, L., Speidel, S., Stoyanov, D., Brandao, P., Crdova, H., Snchez-Montes, C., Gurudu, S.R., Fernndez-Esparrach, G., Dray, X., Liang, J., Histace, A.: Comparative validation of polyp detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision challenge. IEEE Transactions on Medical Imaging **36**(6) (June 2017) 1231–1249
6. Chen, P.J., Lin, M.C., Lai, M.J., Lin, J.C., Lu, H.H.S., Tseng, V.S.: Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology **154**(3) (2018) 568–575
7. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. International journal of computer vision **77**(1) (2008) 157–173
8. Dutta, A., Gupta, A., Zissermann, A.: VGG image annotator (VIA). http://www.robots.ox.ac.uk/ vgg/software/via/ (2016) Accessed: January 21st 2018.
9. Iakovidis, D., Goudas, T., Smailis, C., Maglogiannis, I.: Ratsnake: a versatile image annotation tool with application to computer-aided diagnosis. The Scientific World Journal **2014** (2014)
10. Liu, D., Cao, Y., Kim, K.H., Stanek, S., Doungratanaex-Chai, B., Lin, K., Tavanapong, W., Wong, J., Oh, J., De Groen, P.C.: Arthemis: Annotation software in an integrated capturing and analysis system for colonoscopy. Computer methods and programs in biomedicine **88**(2) (2007) 152–163
11. Rubin, D.L., Rodriguez, C., Shah, P., Beaulieu, C.: ipad: Semantic annotation and markup of radiological images. In: AMIA annual symposium proceedings. Volume 2008., American Medical Informatics Association (2008) 626
12. Several authors.: Video image annotation tool. https://sourceforge.net/projects/via-tool/ (2013) Accessed: January 22nd 2018.
13. Abràmoff, M.D., Magalhães, P.J., Ram, S.J.: Image processing with imagej. Biophotonics international **11**(7) (2004) 36–42