Image Retrieval From Scientifi Publications: Text and Image Content Processing to Separate Multipanel Figures

Emilia Apostolova*, Daekeun You*, Zhiyun Xue*, Sameer Antani, Dina Demner-Fushman and George R. Thoma

Lister Hill National Center for Biomedical Communications, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894 USA. E-mail: emilia.aposto@gmail.com, {youd, xuez, santani, ddemner, gthoma}@mail.nih.gov

Images contained in scientifi publications are widely considered useful for educational and research purposes, and their accurate indexing is critical for efficient and effective retrieval. Such image retrieval is complicated by the fact that figure in the scientifi literature often combine multiple individual subfigure (panels). Multipanel figure are in fact the predominant pattern in certain types of scientifi publications.

The goal of this work is to automatically segment multipanel figures— necessary step for automatic semantic indexing and in the development of image retrieval systems targeting the scientifi literature. We have developed a method that uses the image content as well as the associated figur caption to: (1) automatically detect panel boundaries; (2) detect panel labels in the images and convert them to text; and (3) detect the labels and textual descriptions of each panel within the captions. Our approach combines the output of image-content and text-based processing steps to split the multipanel figure into individual subfigure and assign to each subfigur its corresponding section of the caption. The developed system achieved precision of 81% and recall of 73% on the task of automatic segmentation of multipanel figures

Background

The amount of information in digital image form is ever-increasing because of technological advances and various socio-economic factors. This growth is particularly manifested in the scientifi and medical domains. In the clinical domain, for example, Aucar, Fernandez, and Wagner-Mann (2007) report a trend of an increasing use of medical images. They examined medical images associated with trauma patients over a period of 4 years and observed that the number of radiographic studies increased by 82% during this time. Images are also abundantly used in scientifi publications, particularly in the biomedical literature. The mean number of images per article in the leading biological journals ranges from 6.5 (Yu, 2006) to 31 (Cooper et al., 2004).

With the proliferation of digital images comes the need to organize and easily retrieve image data. Easy access to the scientifi journal-article components such as tables and figure greatly enhances the search experience of researchers and educators (Sandusky and Tenopir, 2008; Divoli Wooldridge, & Hearst, 2010). Image retrieval techniques are therefore an active research field Datta, Joshi, Li, and Wang (2008) observe that the image retrieval fiel has grown tremendously since 2000 both in terms of researchers involved and papers published. The authors of the study searched for publications containing the phrase "Image Retrieval" for each year from 1995 to 2005. The results show a roughly exponential growth in interest in image retrieval and closely related topics during that period.

The interest in image retrieval and semantic image indexing is also manifested by the Image Retrieval Track of the Cross Language Evaluation Forum (ImageCLEF[1]) established in 2003. The goal of ImageCLEF is to create an evaluation platform and to further research on cross language image retrieval. The forum attracts a large number of participants. For example, in 2010 a record number of 112 research groups registered for the four subtasks of the 2010 ImageCLEF (Müller et al., 2010).

Within the general fiel of image retrieval, the retrieval of images from the scientifi literature has prompted avid interest. Images retrieved from scientifi literature are a useful

*These authors contributed equally to this work.

Received: April 3, 2012; revised August 21, 2012; accepted August 21, 2012

 DOI: 10.1002/asi.22810

[1]http://www.imageclef.org/

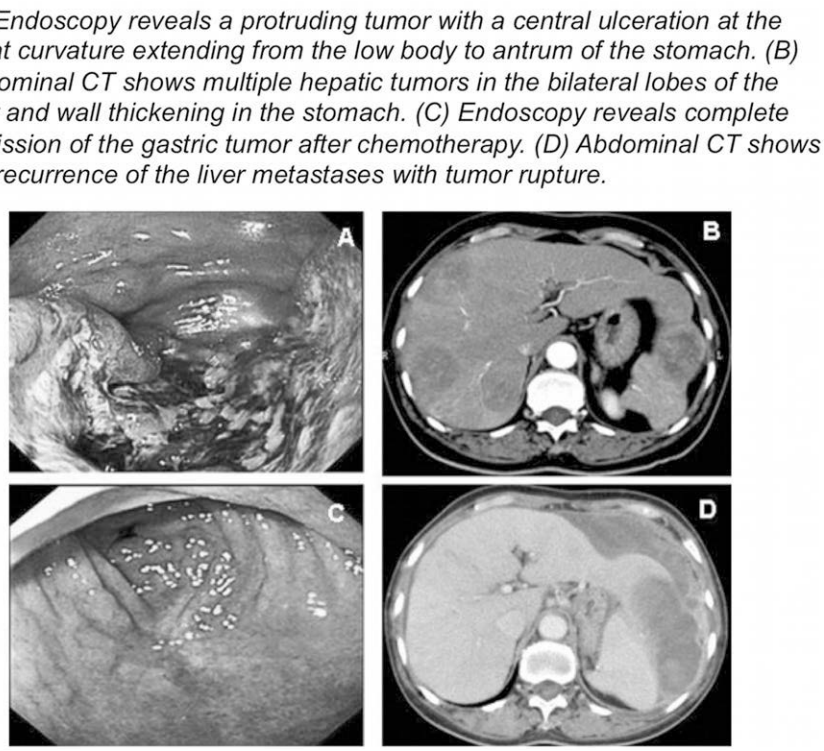


FIG. 1. A sample multipanel figur consisting of four subfigure with panel labels in the upper right corners. The figur caption consists of a correspondingly labeled list of subcaptions.

educational and research tool and their accurate semantic indexing is of significan research interest. Such indexing of images in the biomedical literature will help to address the heterogeneous requirements and searching methods of the intended users: patients and their families looking for explanations; students seeking additional information for their studies; and clinicians who need images for a variety of retrieval tasks. For example, the tasks that prompt clinicians to search for images include: patient education ("showing a patient what I mean with a picture"); comparison or confirmatio of a diagnosis; educational and scientifi presentations; and self-education (Kalpathy-Cramer, 2011). The searching methods could range from submitting a sample image to using sophisticated filter and Boolean operators made possible by the meta-annotation of the biomedical bibliographic citations provided by the NLM[2] indexers.

Images in scientifi publications have some unique characteristics that distinguish the image retrieval task from the task of retrieval of general purpose images. One such distinction is the presence of detailed and reliable text descriptions of images in scientifi publications (figur captions and the text within the article that refers to the figures i.e., "mentions"). This text is often used to provide reliable semantic annotation of the image for indexing and retrieval (Xu, McCusker, & Krauthammer, 2008; You, Antani, Demner-Fushman, Rahman, Govindaraju, & Thoma, 2010; Simpson, Demner-Fushman, & Thoma, 2010). Scientifi publications are also characterized by the abundance of figure consisting of multiple individual panels (subfigures) Multipanel figure are very useful in illustrating complex phenomena and providing comparisons. For example, medical finding are often depicted by multiple panels presenting various image slices, imaging modalities, or comparison images. Figure 1 shows a multipanel figur and its caption.

Multipanel figure are in fact the predominant pattern in certain types of scientifi publications. For example, 53% of 2,422 images randomly selected from the 2011 ImageCLEF medical retrieval track[3] data set (comprised of articles published in 3,277 biomedical journals) were multipanel figure similar to the example in Figure 1.

Although multipanel figure are an accepted and useful tool in journal publications, they do pose a challenge for image retrieval systems. Even though multiple panels combined in a single figur are related in the context of the publication, they may represent distinct entities for semantic image indexing and retrieval, in addition to presenting problems for image content indexing. Thus an image retrieval system needs to separate and distinguish between multiple images present in a single figure

[2]US National Library of Medicine, National Institutes of Health

[3]http://www.imageclef.org/2011/medical

DOI: 10.1002/asi

Label: A

Description: *Endoscopy reveals a protruding tumor with a central ulceration at the great curvature extending from the low body to antrum of the stomach.*

Image:

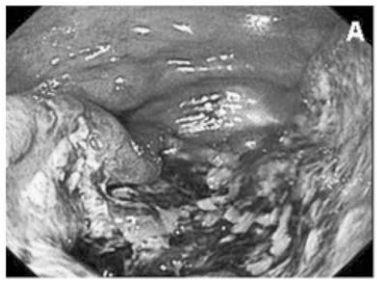


FIG. 2. A sample single-panel output of the panel segmentation system. The four-panel figur in Figure 1 was split into four output entities, each consisting of: (1) a panel label; (2) the segment of the image containing the corresponding label and delimited by the panel boundaries, and (3) the corresponding description extracted from the figur caption.

Purpose

This work presents a method and a system for automatic segmentation of multipanel figure typically present in scientifi publications. The procedure involves segmenting both the figur caption and the actual image content (i.e., findin panel boundaries). Such segmentation will facilitate the accurate automatic semantic indexing and retrieval of images from scientifi publications.

Given a figur and its associated caption, our panel segmentation system determines if the figur consists of multiple panels and, if so, separates the panels and segments the caption. Figure 2 shows a single panel from the output of the system applied to the multipanel figur shown in Figure 1 that consists of four subfigures Each panel in Figure 1 has an associated panel label (*A*, *B*, *C*, or *D*) that is independently detected both in the figur caption (text-based processing) and in the image (image-based processing). In addition, the caption is segmented into text snippets applicable to specifi panels. The resulting image panels and the associated caption segments serve as input to our multimodal biomedical information retrieval system (Demner-Fushman, Antani, Simpson, & Thoma, 2012).

Related Work

The individual image and text processing methods that we apply for panel and caption segmentation and image label recognition are fairly well known. The contributions of our work are: (1) the identificatio of a novel problem; (2) the algorithm that combines the suggestions of the basic text and image processing methods in a way that improves the overall system performance; and (3) the evaluation of feasibility of the proposed automated multipanel figur segmentation solution. Below, we describe some representative work in related areas of text and image processing and reference the relevant review literature for additional details.

Image Processing

In general, image segmentation is a very vast area of research and methods need to be carefully selected or developed based on the type of image and desired goal at hand. In our approach, several fundamental image processing methods were used to segment multipanel figures and are described in the following sections. As such, these methods are fairly generalizable and applicable in a wide variety of image processing applications. We refer the reader to standard image processing texts such as (Gonzalez & Woods, 2008), (Sonka, Hlavac, & Boyle, 2007), and (Russ, 1994) for a broad but useful description of segmentation methods.

Image Text OCR

Once the panels are segmented from the figure any graphical overlays (panel labels and other markup) need to be extracted and recognized using Optical Character Recognition (OCR: the fiel of recognizing image pixels that are in fact characters). As with image segmentation, there are many methods that have been developed over the decades for this problem. The OCR methods are reviewed in Plamondon and Srihari (2000).

Text Processing

Extraction of labels from the caption text is a specifi and relatively simple instance of the information extraction research. For a review of the latest developments in information extraction in the biomedical domain see Simpson and Demner-Fushman (2012). Label extraction could also be viewed as a form of "understanding" the figur captions. The levels of understanding range from extracting the structure of the caption (i.e., label extraction) to identifying the regions of interest shown in the image and described in the caption and the relations between them. Similarly to our

DOI: 10.1002/asi

task, Cohen, Wang, & Murphy (2003) focused on extracting image labels and then classifie the labels into three classes according to their linguistic function: (1) as indicators of a bulleted list; (2) as proper nouns, for example, in "..a procedure used in (A);" and (3) as references interspersed with the text. Note that our task is limited to extracting only the indicators of bulleted lists. For that task, Cohen et al. found the rule-based methods (that are similar to our rules described in the section Method) to have high precision (98.0%) but only moderate recall (74.5%). Despite the similarity of the task and methods, the results of our caption segmentation module cannot be directly compared, because we are interested in findin and classifying only the subcaption labels.

Method

Our multipanel figur segmentation procedure involves fi e distinct submodules: two text-based and two image-based processing modules, and a module that combines the outputs of the previous processing steps. The fi e submodules are described below.

1. Text label extraction: the goal of this module is to identify panel labels present in the figur caption.
2. Panel subcaption extraction: the goal of this module is to identify the individual panel descriptions within the figur caption.
3. Image panel segmentation: the module uses image pixel data to identify panel boundaries.
4. Image label extraction: the module uses image pixel data to identify labels present in the individual panels.
5. Panel splitting: the module combines the outputs of the text label extraction, image panel segmentation, and image label extraction modules to split and name individual subfigures

Methods used for each of the individual system subtasks are described in detail next.

Text Label Extraction

The goal of the text label extraction module is to identify references to panel labels in the associated figur caption. For example, given the caption snippet shown below, the task of the text label extraction module is to identify the panel labels "***A***" and "***B.***"

*(**A**) Endoscopy reveals a protruding tumor with a central ulceration at the great curvature extending from the low body to antrum of the stomach. (**B**) Abdominal CT shows . . . . . .*

In addition to detecting all references of panel labels within the caption, the module also expands label sequences and ranges. For example, the module has the task of expanding labels such as "(*a, b, c-f*)" to their full label list: "*a,b,c,d,e,f."*

The text label detection task lends itself well to a rule-based approach. Even though a number of label inconsistencies and ambiguities were observed, the rule-based approach produced satisfactory results (see section Results).

In our approach, label candidates are identifie through common label patterns and delimiters. These patterns and delimiters are encoded as regular expressions. For example, one of the identifie patterns matches a single alphanumeric character (followed by an optional digit) surrounded by parenthesis or followed by a colon (e.g., "(*a1*)", "*A:*", "*1:*"). Several similar patterns (regular expressions) were created and used to identify sets of label candidates.

In the next step, label expansion rules are applied to each label candidate identifie in the previous step. For example, label ranges such as "(*a-c*)" are expanded to their full label set *"a,b, c*." Similarly, label sequences (labels separated by commas or conjunction) are normalized, for example, the candidate "*a,b, and c*" is normalized to the label set: "*a,b,c.*"

Lastly, the module applies a set of filter to eliminate false positive candidates. For examples, labels that are out of numeric or alphabetic range of a sequence (e.g., "*a, b,* ***p****, c*," "*1, 2,* ***10***") or label candidates surrounded by mathematical or statistical notations are removed from the fina label list (e.g., "$\pm$ ***1***," "$\boldsymbol{p} \leq$ ").

Panel Subcaption Extraction

The goal of the panel subcaption extraction module is to correctly identify portions of the caption text pertaining to a particular panel. For example, text relevant to panel labeled "***A***" in the figur caption below is shown in bold: the description of panel "***A***" consists of the firs sentence (referring to both panels "***A***" and "***B***") and the second sentence (describing panel "***A***" only).

Radiographs performed after closed reduction. *(A)* ***Anteroposterior view showing incongruity of the elbow joint.*** *(B) Lateral view. A bone fragment is clearly identifie into the joint*.

In our approach, extraction of text relevant to a particular panel also relies on a set of hand-crafted rules. First, the module classifie the label sets (extracted by the text label extraction module) into labels preceding the panel descriptions (e.g., ***A:*** . . .) or following the description (e.g., ... **(*A*)**). Then, the module applies rules for identifying the scope of each panel description. An example of a panel scope detection rule is shown below:

The scope of a subcaption that follows the panel text label is the caption text from the current label until the next label.

For example, the detected scope of the description of panel "A" is shown in bold: *". . . . A:* ***Anteroposterior view showing incongruity of the elbow joint.*** *B:* . . . . . ".

Image Panel Segmentation

The goal of the image panel segmentation module is to fin the boundaries of individual panels in a multipanel

DOI: 10.1002/asi

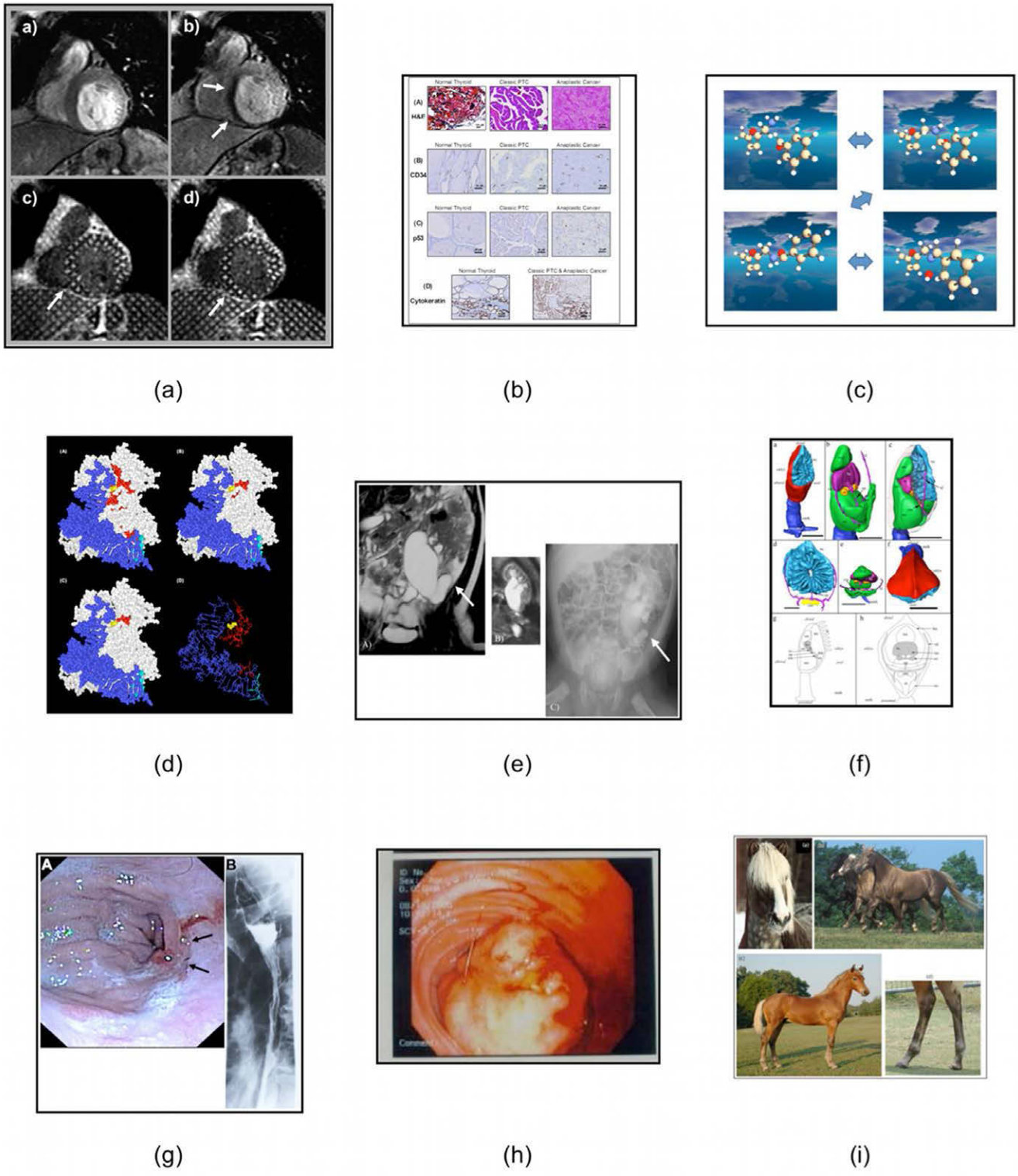


FIG. 3. Sample single- and multipanel figure in the data set. Visual characteristics (color, layout, overlay markup, etc.) for each are discussed in the article text with respect to their effect on the image processing methods. [Color figur can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

figur and split out the individual panels. Figure 3 shows examples of various single and multipanel figure present in the data set.

As illustrated by the examples shown in Figure 3, the major challenge in the image panel segmentation task is the large variety across the data set. For example, the color of the figur background, the layout and size of individual panels, and the image resolutions could vary significantl across figures In some cases, there is no clear panel boundary, or the width of the panel boundary is very small (only a

DOI: 10.1002/asi

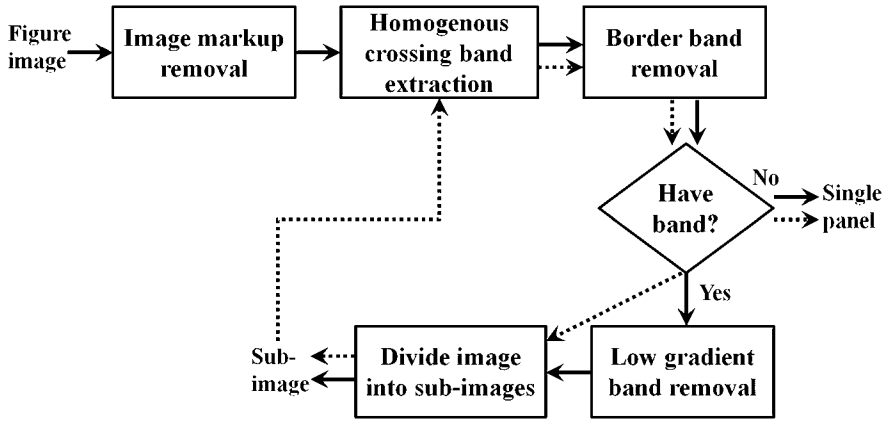


FIG. 4. Process diagram describing the image panel segmentation procedure.

few pixels). In addition, panel labels, text overlays or visual markers (such as arrows) placed in the vicinity of the panel boundaries could interfere with the panel segmentation procedure. To address these challenges, we developed an approach based on the observation that in the majority of multipanel figures individual panels are separated by homogenous horizontal or vertical crossing regions (bands) of uniform color (Cheng, Antani, Stanley, Demner-Fushman, & Thoma, 2011). The systematic errors of this algorithm and the proposed solutions will be discussed in section Results.

Figure 4 summarizes the image panel segmentation procedure. Our algorithm performs two iterations (shown by solid and dashed arrows in Figure 4) of fi e major steps: (1) image overlay/markup removal; (2) homogenous crossing band extraction; (3) border band (homogenous band that is located on the boundary of the panel) identification (4) low gradient band (a band that does not have a sharp boundary line) removal; and (5) image division based on crossing bands. The second iteration is needed for the irregular grid layout. For example, extraction of homogenous bands that cross the entire image will divide Figure 3(f) into three subimages, each of which is also a multipanel figur that needs to be split further. Each of these subimages is divided into subfigure by the homogenous bands that cross the entire subimage and are similar to the homogenous bands extracted in the firs iteration of the algorithm. We will firs describe each of the fi e major steps individually and then explain the fl w.

Image markup removal. The removal of the image markup (such as text) in the areas surrounding potential panels facilitates extraction of the homogenous crossing bands. As shown in Figure 3(b), markup may hinder detection of the crossing areas that separate panels. The markup outside of the panels is typically contained in small isolated regions. We replace these small regions with the surrounding background color as follows:

a. Get the bounding box of each connected component (CC: a blob of black or white pixels) in the binary edge map obtained using the Sobel filte (Gonzalez & Woods, 2008);
b. Remove bounding boxes enclosed by the larger bounding boxes and then merge the overlapping bounding boxes;
c. For each small-area bounding box (with the area less than 10% of the area of the largest bounding box in the image), replace the intensity of all the pixels in the bounding box with the average intensity of the pixels located on the lines that enclose the bounding box.

Homogenous crossing band extraction. The goal of this step is to extract the homogenous bands that cross the entire image horizontally or vertically. The method computes the variance and mean of the pixel intensity on each horizontal and vertical line. Most often, the homogenous bands have high intensity (figure on white paper). Therefore, we try to extract bright homogenous bands first That is, the lines with intensity variance under an empirically established threshold (15) and intensity mean above the threshold (200) are identifie and merged into a rectangle band if the distance between those lines is small (less than 5% of the image width or height). If no bands are extracted in the firs step, only the intensity variance of each line is considered. That is, we identify and merge the low variance lines.

Border band identification The goal of this step is to determine which of the homogenous crossing bands obtained in Step 2 are located close to the image border. For example, the gray image border in Figure 3(a) and the white image border in Figure 3(c) need to be removed for panel extraction. Similarly, in Figure 3(g) the subimage containing panel A extracted in the firs iteration (solid arrow path) includes the white space below the panel A. The white space is a border band for the subimage. Likewise, the white space in Figure 3(e) is also identifie as a border band.

Low gradient band removal. The goal of this step is to filte out the crossing bands that do not represent panel

DOI: 10.1002/asi

boundaries. Using the binary edge map obtained in the markup removal step, the procedure examines the longest edge of each of the crossing bands. The band is removed if the longest edge is too short compared to the corresponding image dimension, or the ratio between the length of the longest edge and the corresponding image dimension is too small.

Image division based on crossing bands. In this fina step, the images are either classifie as single-panel images or divided into subimages using the coordinates of the extracted horizontal and vertical crossing bands. An input image is classifie as a single-panel figure if no homogeneous crossing bands are identifie in the second step or remain after the fourth step. Otherwise, the algorithm outputs the number of individual panels and their coordinates obtained using the locations of the homogeneous crossing bands in the image.

Image Label Extraction

The goal of the image label extraction module is to detect panel labels superimposed on each individual panel of multipanel figures Several image processing and optical character recognition (OCR) techniques are used to segment panel label connected components (CCs) and recognize them. The module output consists of the recognized panel labels (e.g., *A, B, a, b*, etc.) together with their location within the entire multipanel figure The image label extraction algorithm performs three steps that are described below: (1) image preprocessing; (2) OCR; and (3) panel label detection.

Image preprocessing (binarization). In the preprocessing step an input image is binarized (each pixel is stored as either black or white) to extract character CCs. We observed that panel labels in the data set are usually black or white and hence a binarization-based method is sufficien to segment overlay characters. Two empirically established fi ed threshold values (50 and 200) are used to extract black and white characters, respectively. A threshold of 128 and an adaptive thresholding method[4] are applied for characters colored other than black or white (e.g., intensities between 50 and 200). Figure 5(b) shows the binarization result of the input image shown in Figure 5(a). The result was obtained by firs thresholding the input at 200 and then taking the negative of the binarized image to segment black CCs of white panel labels.

Image text recognition. The goal of this step is character recognition. We tested publicly available OCR tools and determined that the standard OCR tools are not well-suited for our task. We therefore developed an alphanumeric OCR engine based on contour features and neural network (NN) theory (You, Antani, Demner-Fushman, Govindaraju, & Thoma, 2011). The average recognition rate of this approach (measured on a test set consisting of more than 66,700 character samples extracted from biomedical images) is close to 99%. Each black CC identifie in the previous step is processed by the OCR engine that outputs a recognition result (character label) and a score. Figure 5(b) shows the OCR results next to the corresponding CCs.

Panel label detection. As shown in Figure 5(b), the OCR results include true panel labels, as well as multiple false positive characters. A method for detecting true panel labels (i.e., *A~F*) in the OCR output is necessary. We apply the Markov Random Field (MRF) modeling approach (Li, 2009) for this detection task based on the following characteristics of the panel labels:

Alignment: panel labels are aligned horizontally or vertically. For example, in Figure 5(a), panel labels *A, B*, and *C* are aligned horizontally, while *A* and *D* are aligned vertically. They are marked by dashed and solid narrow rectangles, respectively.
Order: panel labels are ordered alphabetically from left to right and/or top to bottom.
Size: the sizes of CCs of panel labels are very close.

The characteristics and relationships among panel labels are modeled using MRF to classify each OCR-ed CC as a true panel label or noise (You et al., 2011). Characters that satisfy the characteristics and relationships compose a candidate label set and several candidate sets are obtained as a result of the MRF modeling. Figures 5(c) shows two candidate sets. Both consist of characters that satisfy the characteristics. It is difficul to determine the true candidate set only from the MRF results. Characters in the false positive set shown in the dotted box in Figure 5(c) are also apparently good candidates for true panel labels (e.g., for a three-panel figure) Other results such as text labels and image panel boundaries help selecting the true label set.

Panel Splitting

The goal of the panel splitting task is to combine the results of the (1) text label extraction; (2) image panel segmentation; and (3) image label extraction to split out and name the individual subfigures We observed that the three results agree and successfully split the figure only for about 30% of the multipanel figures Here "agree" means that all panel labels and borders are accurately extracted by the three detection algorithms. For example, for the image shown in Figure 5(a) the three results would be in agreement only if the following three conditions are met: (1) the text label extraction module find all panel labels *A*, *B*, *C*, *D*, *E*, and *F*; (2) the image panel detection module delineates all boundaries of the six panels; and (3) all image labels (*A~F*) are correctly recognized and located. The three results frequently disagree and hence need to be combined and adjusted for successful panel splitting. To that end, we firs combine text labels and image labels, and then match the panel labels with the extracted subpanels.

[4]http://www.xdp.it/cximage.htm

DOI: 10.1002/asi

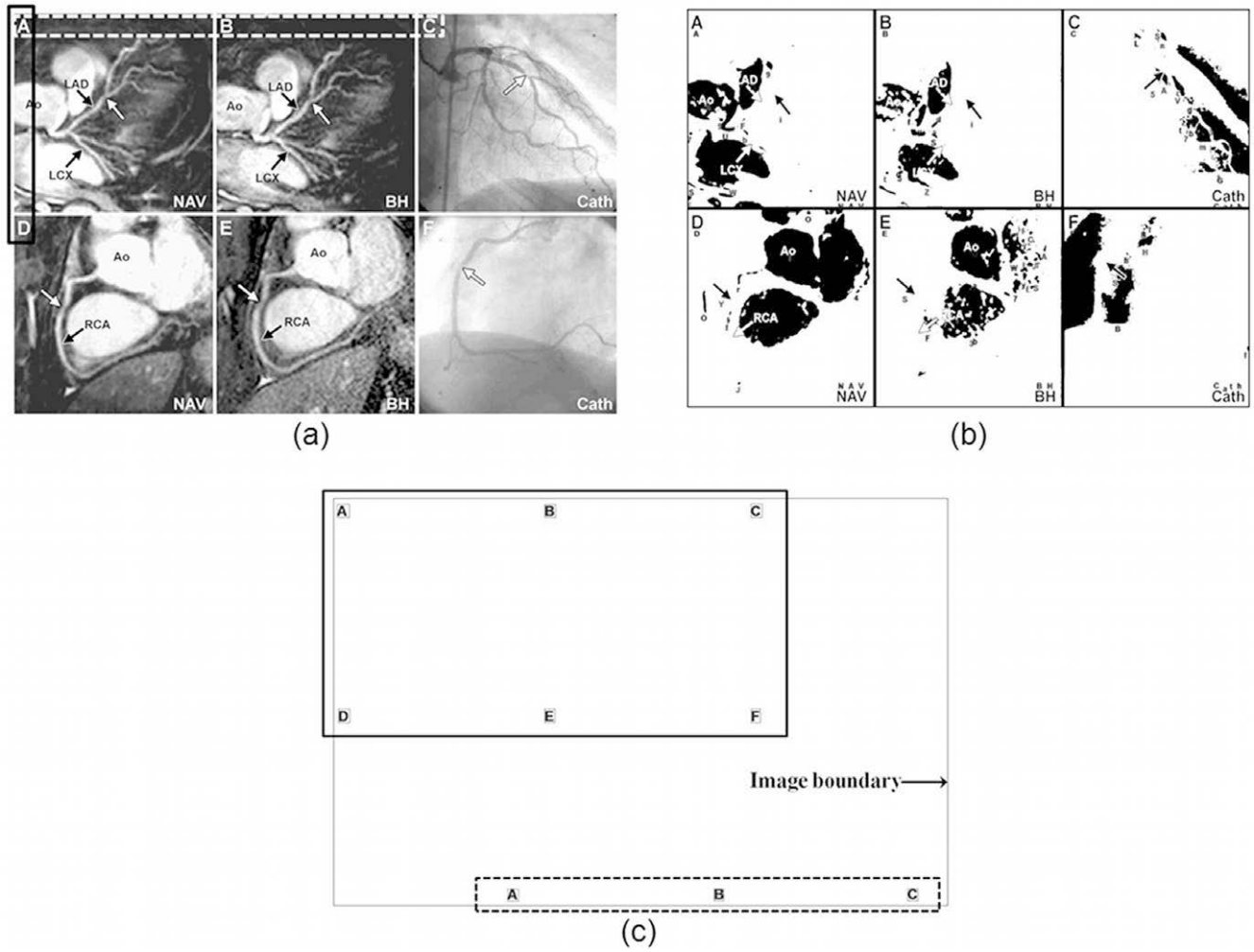


FIG. 5. Output from various image processing steps in the image panel label detection process. (a) Input multipanel figur showing vertical and horizontal alignment of panel labels. (b) Image binarization result for image in subfigur (a) and OCR-ed letter labels. (c) Output of the panel label detection step showing a true panel label set (marked by a box with solid line) and a false positive (marked by a box with dashed line).

Combining text and image labels. The text label extraction and image label extraction produce the same type of information: character panel labels that can be combined into a single-panel label set. Text labels could be useful for filte - ing out false positive candidate sets obtained by the image label extraction module. Because either or both results could be imperfect, we select panel labels that are found in both results. We consider such panel labels highly probable. Figure 6(a) shows an example in which both module results have some missing labels. The text label extraction module detected labels *a*, *b*, *c*, *d*, *e*, *f*, and *g* (white overlaid letters in panel a) but missed labels *h* and *i*. The image label extraction module, on the other hand, detected labels *a*, *b*, *d*, *e*, and *g* (white overlaid letters below each panel label) but missed four labels (*c*, *f*, *h*, and *i*). Hence only labels *a*, *b*, *d*, *e*, and *g* are considered panel labels. This AND selection operation may cause labels accurately detected by one module (but not by both) to be excluded from the fina output (e.g., labels *c* and *f*). However, it can successfully eliminate noisy labels erroneously detected by either module (e.g., labels *f*, *m*, and *v* in the text label extraction result from Figure 6(b)).

Combining panel labels with image panels. Panel labels (combined text and image labels) are now available to be assigned to the extracted panels (subfigures) Each panel should be named by its corresponding panel label found within or near the panel. Panels that are not properly split by the image panel segmentation module and hence form a super panel could be split and named successfully using the corresponding panel labels. This label-based splitting is possible because of the regularities in the label layout and assignment: (1) labels are usually located at the top or bottom corners (and sometimes at the center) of each panel, and (2) panel labels and subfigure are arranged from left to right and/or top to bottom. Both Figures 6(a) and 6(b) present left to right and top to bottom order of the subfigure and labels are placed at the top-left corners. Sometimes labels are placed outside of their panels (e.g., labels *F* and *J*

DOI: 10.1002/asi

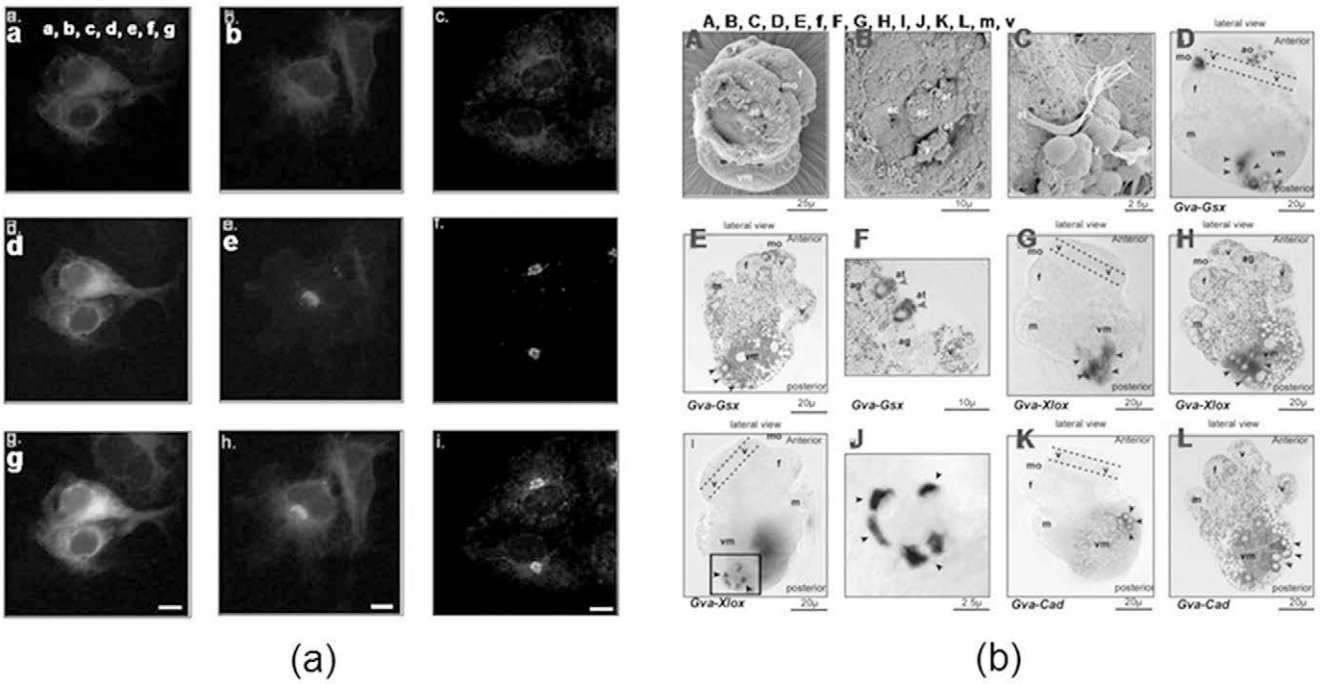


FIG. 6. Example images with missing or noisy panel label detection. Such errors can affect the process for combining text and image labels. (a) Labels missed by both text and image processing. Text processing extracted labels *a* through *g* (shown in white in top left subfigure and image processing extracted labels *a*, *b*, *d*, *e*, and *g* (shown in white below the panel labels). (b) Extraneous labels identifie by text processing (shown above the image) were removed because of accurate image processing.

in Figure 6(b)). However, the location of the labels is at the top-left corner, similarly to the rest of the labels. The fina panel splitting method is implemented based on the above two panel label patterns and consists of three steps.

Step 1: Splitting panels containing multiple labels. The image panel segmentation module sometimes fails to detect panel boundaries and split out the subfigures Figure 7(a) shows a multipanel figur where the algorithm failed because of the absence of clear panel boundaries. The panel labels, however, were successfully detected in terms of characters and their locations. In this case, the super panel can be split based on the panel labels and their arrangement pattern. Panel labels are assigned from left to right and from top to bottom, and they are located at the bottom-left corner. The space margin between a label (e.g., *A* or *C*) and the left panel border can be easily computed, and the margin can be placed to the left of each label (*B* and *D*) to determine their left panel border. Similarly, the bottom margin can be computed using labels *C* or *D* and the bottom border of the super panel. The margin offset is then used to determine the bottom border of panels A and B. The newly inferred borders may not be as accurate as the visually observed borders between the panels, however, they are fairly acceptable in this case and other similar cases in which all panel labels are located uniformly within each subfigure Figure 7(b) shows such inferred segmentation results.

Step 2: Matching single labels within or outside of their panel borders. After completion of Step 1, it can be assumed that each panel has one label or none (depending on the label extraction results). Panel labels may be located in or out of detected panel borders (Figure 8(a)). For certain subfigures however, there may be no detected labels (e.g., *A*, *D*, *E*, *G*, and *H* in Figure 8(b)). In this step, panels with labels detected within or outside of their borders are split and named. The algorithm firs examines panels with a single label in them and then identifie the location pattern of the labels in the panels. The location pattern is used to search for labels that are located outside of the corresponding panels. For the detection results in Figure 8(a), for example, we split and named panels B, D, and F first Then a label location pattern, that is, top-left corner, was detected in these panels and used to match labels *A*, *C*, and *E* to their panels. Panels F and J and their labels in Figure 6(b) were also successfully matched because their labels are near the top-left corner of the panels.

Step 3: Assign labels to panels with no available label. Panels with available labels are successfully split and named in the firs two steps. In Step 3, panels without available labels are processed based on their panel arrangement pattern. The pattern is determined using the panels successfully split in the two prior steps, and a missing label is assigned based on the pattern and neighboring panel

DOI: 10.1002/asi

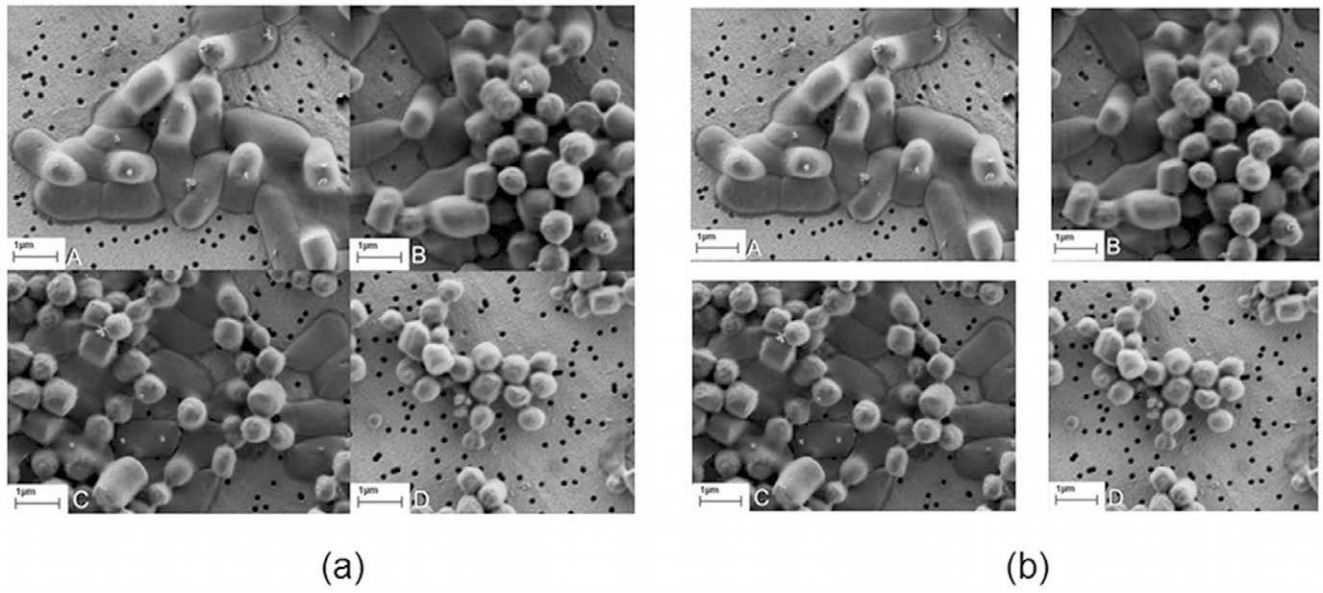


FIG. 7. In spite of one algorithm failing, if two of three algorithms agree a successful panel splitting may still be achieved. Figure shows illustrations of successful splitting from imperfect results. Image in (a) with indistinct panel boundaries, but with successful text and image panel label recognition, results of a successful panel splitting shown in (b).

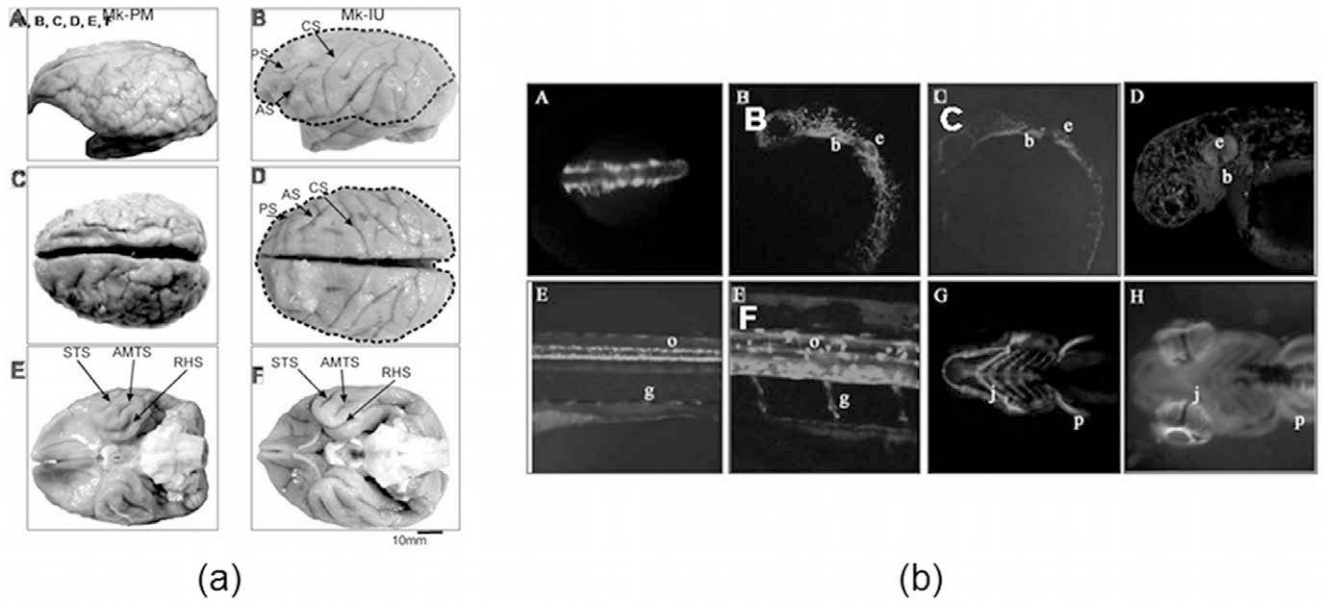


FIG. 8. (a) Example showing correct panel segmentation in spite of unusual panel label location affecting image-based recognition. (b) In spite of incomplete image processing-based panel label recognition, panel layout clues enable correct panel segmentation.

labels. For example, in Figure 8(b), the three labels *B*, *C*, and *F* are sufficien to detect a left to right order of panels in a row and a top to bottom order between the two rows. Then the missing labels *A*, *D*, *E*, *G*, and *H* can be easily assigned to their corresponding panels. If the extracted labels are not sufficien for detecting a pattern, the default arrangement pattern (left to right and/or top to bottom) and the default labels (upper case letters) are applied. Figure 9 summarizes the multistep panel segmentation approach.

Data Set

The data set used for multipanel figur analysis and evaluation of the extraction methods consists of 2,348

DOI: 10.1002/asi

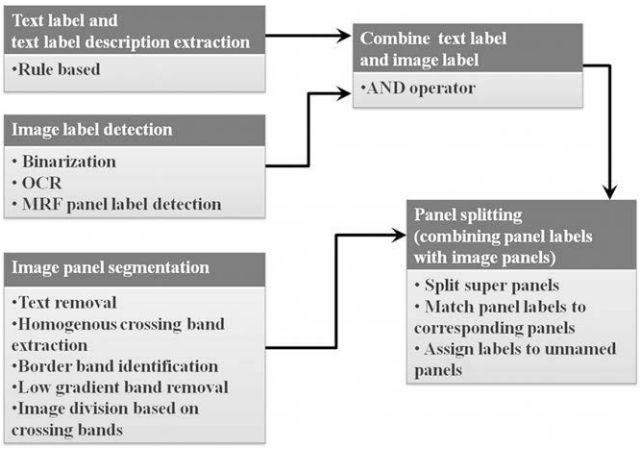


FIG. 9. Process diagram showing contribution of each step to the multipanel figur segmentation algorithm.

images and their captions extracted from scientifi publications in the biomedical domain and the life sciences. This is a subset of the data provided by the medical retrieval track of ImageCLEF2011[5] containing 231,000 images taken from the Open Access subset of PubMed Central.[6]

Evaluation data sets (a total of 400 images and their associated captions) were created for each of the tasks described in section Method: the two text-based, two image-based tasks, and the fina panel splitting task. Two independent annotators (a biomedical informatition and a medical student) annotated both the image and the associated figur caption following a set of predefine guidelines. The annotation procedure was based on the Delphi communication method (Linstone & Turoff, 1976). A procedure for modifi cation and refinemen of the annotation guidelines was established and followed. Annotator disagreement was resolved via an adjudication procedure, in which the two annotators were joined by a physician trained in medical informatics. Most disagreements were because of the differences in the boundaries annotations. For example, one of the annotators would accidentally include punctuation adjacent to a label while annotating labels in the captions. Such disagreements were easily reconciled. An annotated reference set was created from the consensus of both annotators.

A custom Web-based annotation tool was developed for the text-based annotation tasks-text panel label and panel description annotations as shown in Figure 10. The LabelMe (Russell, Torralba, Murphy, & Freeman, 2008) Web-based annotation tool was utilized for the image-based annotation tasks: subfigur segmentation and panel label annotation. Figure 11 illustrates the image-based annotation process.

The annotators achieved high interannotator agreement demonstrating the successful guidelines and annotation procedures. Cohen's kappa of 0.80 was measured on the label annotation task, and of 0.78 on the panel description annotation task.

Results

The text and image-based modules were independently evaluated using the annotated reference set. Table 1 shows the results for the text-based modules: text label extraction and panel description extraction. Results vary from an F1-score of 72.7% to 74.7% on the task of text-based panel label detection using exact and inexact boundaries, respectively. The results of the panel description extraction vary more widely with an F1-score of 65.4% using exact boundaries and 83.6% using inexact boundaries.

Table 2 shows the results of the image panel segmentation module.

To evaluate the image panel segmentation, we consider each panel individually and use the precision and recall measures with nonexact panel boundaries. An extracted panel is considered "true positive" if it satisfie the following two criteria:

1) The overlapping area between the extracted panel and the matching reference set panel is larger than 75% of the area of the reference set panel;
2) The overlapping area between the extracted panel and a reference set panel adjacent to the matching reference set panel is less than 5% of the area of the adjacent panel.

The 400 test images contain 1,764 reference set panels. The image panel segmentation module extracted 1,482 panels, of which 1,276 were correct, leading to precision of 86.1% and recall of 72.3%. Under-segmentation accounts for the majority of failures in the images that do not contain homogeneous crossing bands, whereas over-segmentation occurs in the images in which homogenous crossing bands do not delimit subfigures We plan to use the image label information to reduce segmentation errors.

For evaluation of the image label extraction, we firs ran the algorithm on the test set and then compared the extracted panel labels with the reference annotation.

Recall and precision of 70.6% and 97.3%, respectively (shown in Table 3) were achieved on the detected label sets retained after filterin based on the text extraction information. In addition, we evaluated the number of images in which more than 50% of panel labels were successfully detected. As shown in section Method, it is not always necessary to detect all panel labels for successful fina panel extraction and labeling. Our algorithm detected more than 50% of panel labels in 85.3% of the test images (341 images out of 400); however, it detected no labels in 7.0% (28 images) of the test set. Error analysis revealed that the main causes of undetected labels are 1) low image resolution/quality and 2) irregular alignment of panel labels.

[5]http://www.imageclef.org/2011/medical
[6]http://www.ncbi.nlm.nih.gov/pmc/

DOI: 10.1002/asi

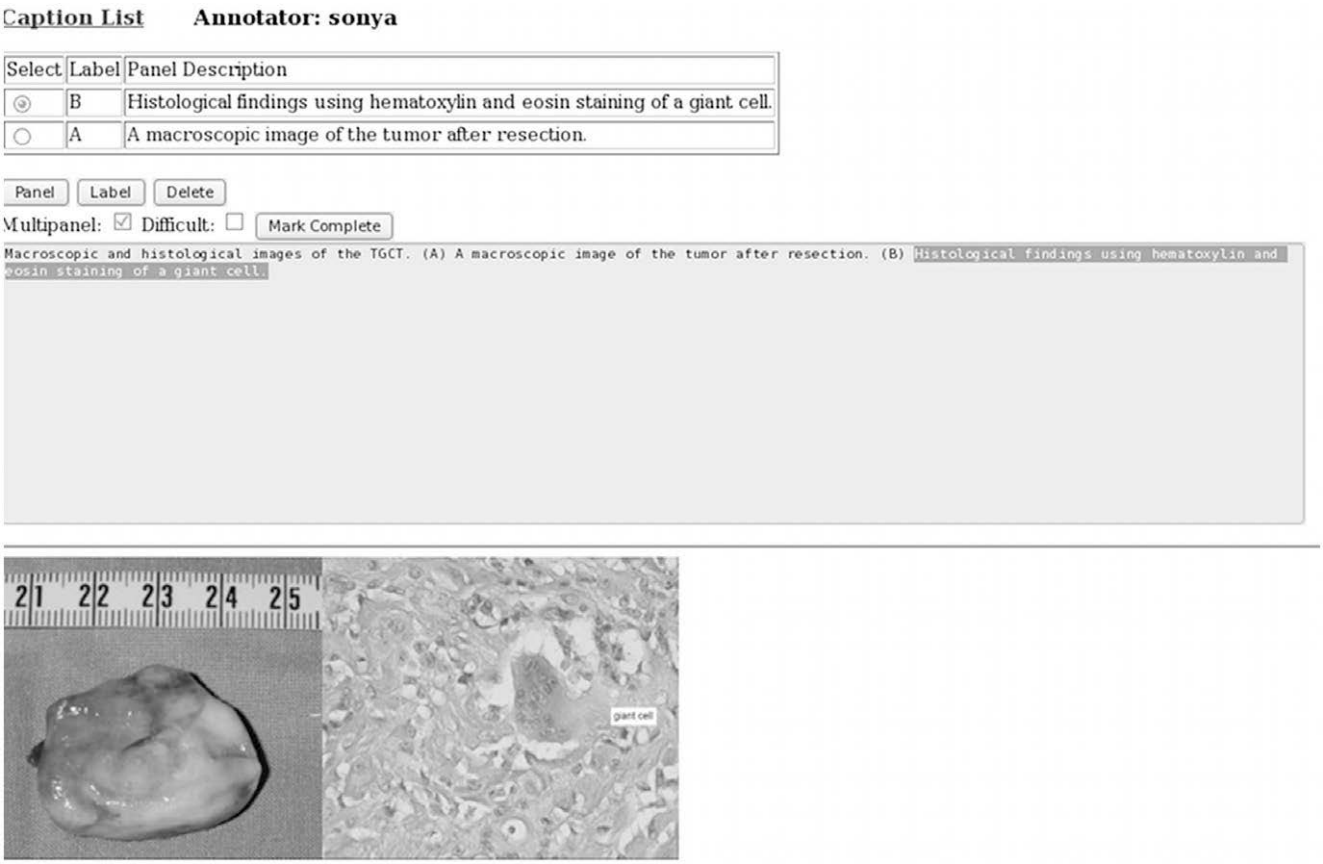


FIG. 10. A screenshot showing the custom Web-based annotation tool for text-based panel label and description annotation. In the interface, annotators are shown an image and its caption and are asked to annotate in the caption the panel labels and their associated descriptions.

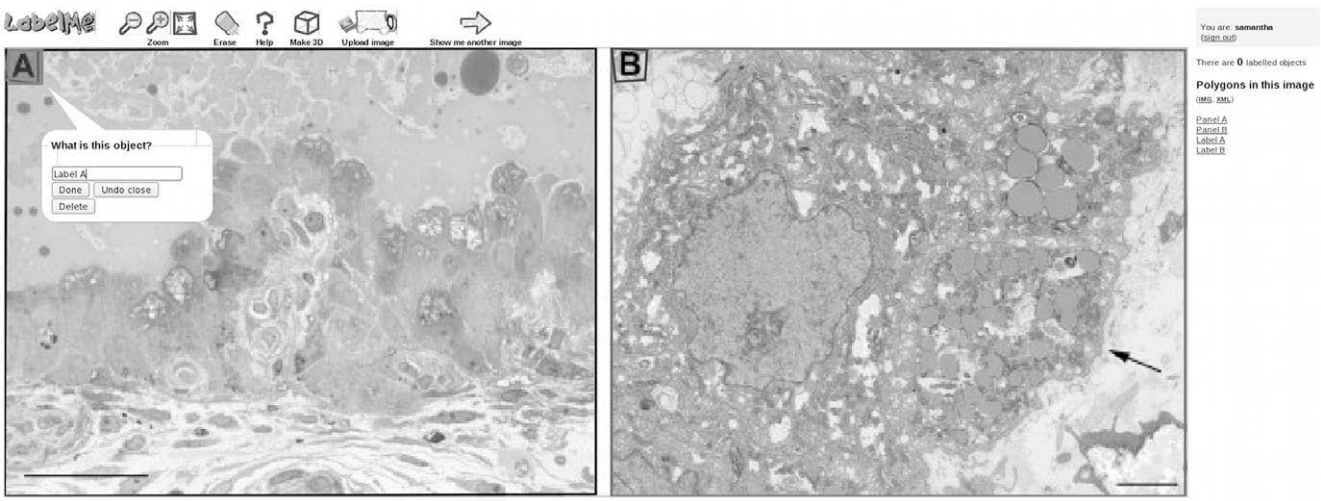


FIG. 11. A screenshot showing the image-based annotation process. The LabelMe tool (Russel et al., 2008) was used for the task. Human annotators were asked to delineate (as polygons) subfigure and their associated panel labels.

Overall Evaluation of the Panel Splitting Algorithm

In addition to evaluating the performance of individual system modules, we evaluated the overall performance of the developed system. The combined panel splitting algorithm was evaluated in realistic and simulated oracle conditions. First, all actual results from the three extraction modules (text label extraction, image panel segmentation, and image label extraction) were used as input to the panel splitting algorithm, and the results (extracted panels and

DOI: 10.1002/asi

TABLE 1. Caption segmentation results—panel label and panel description extraction. Panel description extraction metrics were computed using the reference set panel labels.

	Inexact boundary match			Exact boundary match		
	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)
Text panel label	79.03	70.75	74.66	76.92	68.84	72.65
Text panel description	81.06	86.29	83.59	63.35	67.47	65.35

TABLE 2. Results of the image panel segmentation.

	Annotated	Extracted	Correctly extracted	Precision (%)	Recall (%)	F1-score (%)
Number of panels	1764	1482	1276	86.10	72.34	78.62

TABLE 3. Results of the image label extraction.

Total reference set labels	Total detected labels	Total matched labels	Precision (%)	Recall (%)
1877	1363	1326	97.29	70.64

TABLE 4. Evaluation results of the combined panel splitting algorithm. The *Default* column reports the actual results, The *RS_* prefi es in the column headers denote tests with corresponding actual result(s) replaced with reference set annotations. *OCR*, *Text*, and *Image* after the *RS_* prefi es denote replaced results of the image label extraction, text label extraction, and image panel segmentation, respectively. Up to two actual results are replaced.

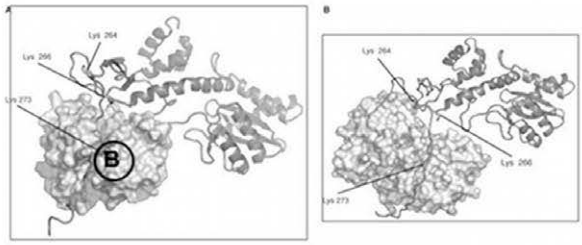
	Default	RS_OCR	RS_Text	RS_Image	RS_OCR_Text	RS_OCR_Image	RS_Text_Image
Precision (%)	**80.92**	84.97	87.07	83.30	91.34	88.28	89.48
Recall (%)	**73.39**	78.65	82.34	83.98	91.29	89.88	90.00

their labels) were compared to the reference set. Second, one or two of the intermediate results (e.g., text label extraction) were replaced with the reference annotations. The results of these evaluations are summarized in Table 4. Precision and recall were computed for every test run.

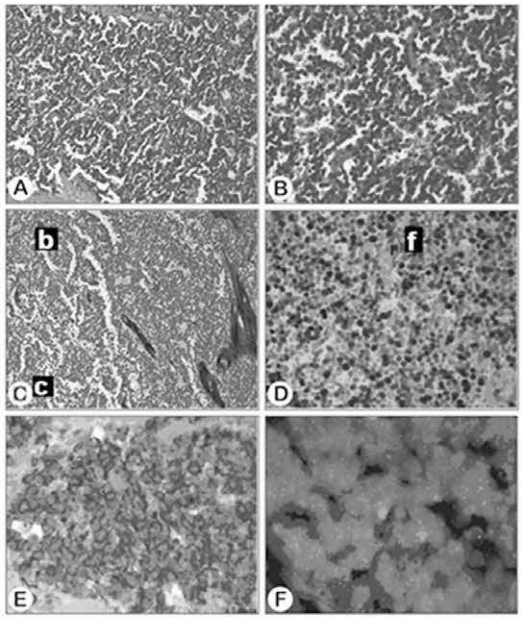
The realistic test run based on the actual output of each intermediate module (the *Default* column in Table 4) achieved about 80.9% and 73.4% precision and recall, respectively. We identifie three main causes of errors in the panel splitting algorithm: (1) image label extraction failed to detect panel labels in their true location (OCR error), (2) lower case labels were extracted from text, but the image labels were upper cases (case mismatch), and (3) image panel segmentation algorithm failed. Figure 12(a) shows an example of case (1). The image labels are too small to be successfully recognized and a wrong label *B* (in black circle) was detected in the panel A. As a result, panel A was named *B*, and then panel B was named *C*, which is the next label to *B* in alphabetic order. Label *C* was assigned by default because the panel B had no associated label detected within or outside of it. Figure 12(b) shows an example of case (2). The actual image labels are upper case, but the lower case labels were extracted from the caption, and this led to selection of a wrong candidate set (incorrect lower case labels in panels C and D shown in black background boxes). Figure 12(c) shows an example in which the image panel segmentation algorithm failed to detect an entire region of panel b and e (case 3), and hence they were not counted as successful or successes because the extracted region is smaller than 75.0% of the reference set panel.

Figures 12(d) and (e) show additional failure examples. Figure 12(d) shows an image in which no text labels were detected and the split panels were labeled using the default naming convention. The label order meets the rule; however, the labels are all lower case, not upper case. Keeping upper case as default label characters, however, achieves higher performance than a lower case default. Another test run (not shown in Table 4) with lower case as default achieved 70.2% and 63.7% precision and recall, respectively. These results are approximately 10.0% worse than the results for the

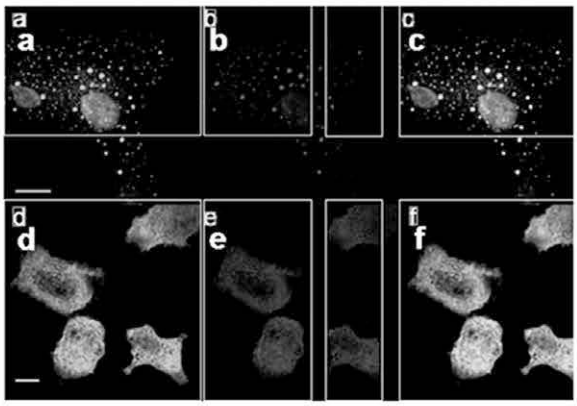
DOI: 10.1002/asi



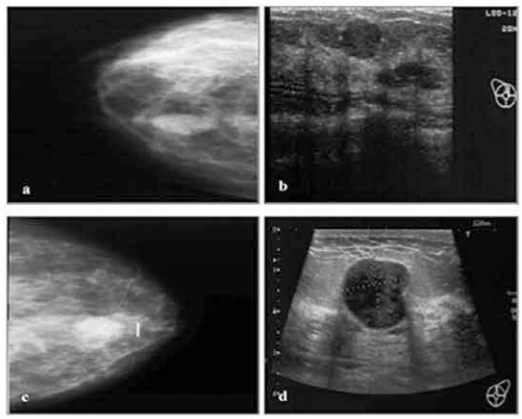
(a) OCR error



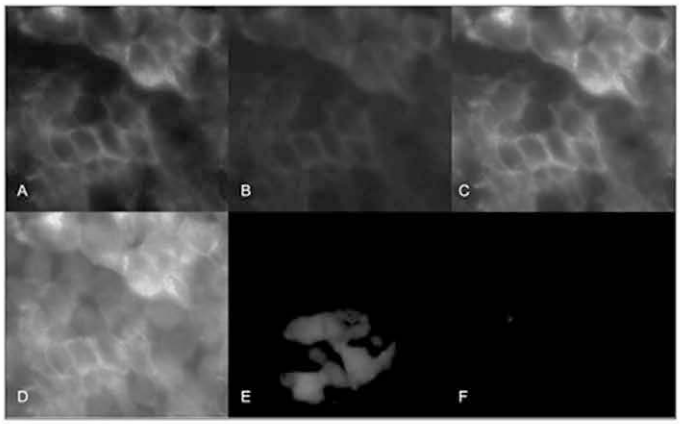
(b) Text label detection error



(c) Image panel detection error



(d) An error case due to default naming convention



(e) Both label and panel detection failed

FIG. 12. Examples of failure cases.

DOI: 10.1002/asi

upper case default (compare with *Default* in Table 4). If no labels or panels were detected, an image could not be split at all, as shown in Figure 12(e).

Test runs replacing one or two actual extraction results with reference annotations achieved better performance as shown in Table 4. Replacing image panel segmentation results increased recall by approximately 10.0%. The failure cases shown in Figure 12(c) and (e) were successfully split and named with correct panel borders. Such cases mainly contributed to the increase in recall. Using both the reference set text and image labels achieved the highest performance. Images similar to Figure 12(e) that was classifie as a single-panel in the actual evaluation were successfully split based on the location of panel labels. Other failure cases except (c) shown in Figure 12 were corrected as well. Our oracle test results indicate that improving each individual module is important to achieve better performance. It is also noticeable that improvement of the text and image panel label detection algorithms will provide for a better overall system. Accurate panel label detection could give the extracted panels (regardless of the panel segmentation errors) a better chance to be correctly split and named.

Conclusion

The scientifi literature presents a vast and mostly untapped source of image data. On the one hand, images found in publications are abundant, and on the other, they are typically accompanied by meaningful textual descriptions that lend themselves to accurate automatic semantic indexing. A significan obstacle in the image indexing process is the predominant presence of multipanel figures Multiple figure are often collated into a single figur described by a single figur caption. Segmenting multipanel figure into individual subfigure is a necessary preprocessing task for systems targeting scientifi image data.

We have developed a system capable of automatically segmenting multipanel figure and captions into individual subfigure and their associated textual descriptions. We have combined text extraction modules with image content-based processing modules. Two text-based processing modules firs extract the panel labels and panel descriptions from figur captions. Subsequently, an image panel segmentation module detects individual panels and another image-content processing module extracts panel labels. Although each individual text and image processing module performs satisfactorily, the cumulative errors might result in an unsatisfactory overall system performance. To avoid aggregating individual processing errors, we combined the results of individual modules in a way that improves the overall system performance, rendering results superior to each of the individually evaluated system modules. Although the algorithm for combining the results is developed specificall for images in the biomedical literature, it should be generalizable to any multipanel figure accompanied by captions and containing identical labels in both the images and the captions. The panel splitting module that combines the labels extracted in the text and image processing steps achieves precision of 80.9% and recall of 73.4% on the overall task. These results indicate that the automatic segmentation of multipanel figure is a feasible task that could considerably improve image retrieval and indexing systems targeting the scientifi literature.

Acknowledgments

This work was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health. We would like to thank the ImageCLEF organizers and the Radiological Society of North America (RSNA), publisher of Radiology and RadioGraphics, for making the database available for the experiments under the ImageCLEF medical image retrieval task.

References

Aucar, J. A., Fernandez, L., & Wagner-Mann, C. (2007). If a picture is worth a thousand words, what is a trauma computerized tomography panel worth? The American Journal of Surgery, 194(6), 734–740.

Cheng, B., Antani, S., Stanley, R.J., Demner-Fushman, D., & Thoma, G.R. (2011). Automatic segmentation of subfigur image panels for multimodal biomedical document retrieval. Proceedings of SPIE Electronic Imaging Science and Technology, Document Retrieval and Recognition XVIII. San Francisco, CA. January 2011; 7874: 78740Z.

Cohen, W.W., Wang, R., & Murphy, R. (2003). Understanding captions in biomedical publications. Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2003), pp. 499–504.

Cooper, M.S., Sommers-Herivel, G., Poage, C.T., McCarthy, M.B., Crawford, B.D., & Phillips, C. (2004). The zebrafis DVD exchange project: A bioinformatics initiative. Methods in Cell Biology, 77, 439–457.

Datta, R., Joshi, D., Li, J., & Wang, J.Z. (2008). Image retrieval: Ideas, influences and trends of the new age. ACM Computing Surveys (CSUR), 40(2), 5.

Demner-Fushman, D., Antani, S., Simpson, M., & Thoma, G.R. (2012). Design and development of a multimodal biomedical information retrieval system. Journal of Computing Science and Engineering, 6(2), 168–177.

Divoli, A., Wooldridge, M.A., Hearst, M.A. (2010). Full text and figur display improves bioscience literature search. PLoS ONE 5(4): e9619.

Gonzalez, R., & Woods, R. (2008). Digital Image Processing, 3rd edition. Upper Saddle River, NJ: Prentice Hall Inc.

Kalpathy-Cramer, J. (2011). Secondary use of medical images: User perspectives on image retrieval. Annual Symposium of the American Medical Information Association (AMIA 2011). Panel presentation. Washigton, DC.

Li, S.Z. (2009). Markov Random Field modeling in image analysis. New York: Springer.

Linstone, H. A., & Turoff, M. (1976). The Delphi method: Techniques and applications, 18(3). Boston, MA: Addison-Wesley.

Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Kahn, C.E., & Hersh, W. (2010). Overview of the CLEF 2010 medical image retrieval track, Working Notes of CLEF.

Plamondon, R., & Srihari, S.N. (2000). Online and off-line handwriting recognition: A comprehensive survey. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(1), 63–84.

Russ, J.C. (1994). The Image Processing Handbook. 2nd edition. Boca Raton, FL: CRC Press, Inc.

DOI: 10.1002/asi

Russell, B.C., Torralba, A., Murphy, K.P., & Freeman, W.T. (2008). LabelMe: A database and Web-based tool for image annotation. International Journal of Computer Vision, 77(1), 157–-173.

Sandusky, R.J., & Tenopir, C. (2008). Finding and using journal article components: Impacts of disaggregation on teaching and research practice. Journal of the American Society for Information Science & Technology, 59(6): 970–982.

Simpson, M.S., & Demner-Fushman, D. (2012). Biomedical text mining: A survey of recent progress. Mining Text Data. Berlin/Heidelberg, Germany, 2012: 465–517.

Simpson, M., Demner-Fushman, D., & Thoma, G.R. (2010). Evaluating the importance of image-related text for ad-hoc and case-based biomedical article retrieval. Proceedings of the 2010 Annual Symposium of the American Medical Information Association (AMIA 2010) 2010: 752. Washington, DC, USA.

Sonka, M., Hlavac, V., & Boyle, R. (2007). Image processing, analysis, and machine vision. 3rd Edition. CL Engineering, Stamford, CT, USA.

Xu, S., McCusker, J., & Krauthammer, M. (2008). Yale Image Finder (YIF): A new search engine for retrieving biomedical images. 24(17), 1968–1970. Oxford, England: Oxford University Press.

You, D., Antani, S., Demner-Fushman, D., Govindaraju, V., & Thoma, G. R. (2011). Detecting figure-pane labels in medical journal articles using MRF. Proceedings of 11th International Conference on Document Analysis and Recognition (ICDAR 2011) (pp. 967–971). Beijing, China.

You, D., Antani, S., Demner-Fushman, D., Rahman, M., Govindaraju, V., & Thoma, G.R. (2010). Biomedical article retrieval using multimodal features and image annotations in region-based CBIR. Document Recognition and Retrieval XVII. Proceedings of the SPIE. San Jose, CA, USA.

Yu, H. (2006). Towards answering biological questions with experimental evidence: Automatically identifying text that summarize image content in full-text articles. AMIA Annual Symposium Proceedings, 2006, 834. American Medical Informatics Association. AMIA 2006, Washington DC, USA.

DOI: 10.1002/asi