2019 International Conference on Document Analysis and Recognition (ICDAR)

Identifying the Central Figure of a Scientific Paper

Sean T. Yang*, Po-Shen Lee*, Lia Kazakova†, Abhishek Joshi†, Bum Mook Oh†, Jevin D. West†, and Bill Howe†
*Department of Electrical and Computer Engineering, †Information School
University of Washington, Seattle, WA
{tyyang38, sephon, kazako, joshi1, bmo5, jevinw, billhowe}@uw.edu

***Abstract*—Publishers are increasingly using graphical abstracts to facilitate scientific search, especially across disciplinary boundaries. They are presented on various media, easily shared and information rich. However, very small amount of scientific publications are equipped with graphical abstracts. What can we do with the vast majority of papers with no selected graphical abstract? In this paper, we first hypothesize that scientific papers actually include a "central figure" that serve as a graphical abstract. These figures convey the key results and provide a visual identity for the paper. Using survey data collected from 6,263 authors regarding 8,353 papers over 15 years, we find that over 87% of papers are considered to contain a central figure, and that these central figures are primarily used to summarize important results, explain the key methods, or provide additional discussion. We then train a model to automatically recognize the central figure, achieving top-3 accuracy of 78% and exact match accuracy of 34%. We find that the primary boost in accuracy comes from figure captions that resemble the abstract. We make all our data and results publicly available at https://github.com/viziometrics/centraul_figure. Our goal is to automate central figure identification to improve search engine performance and to help scientists connect ideas across the literature.**

***Keywords*-Graphical Abstract; Central Figure; Machine Learning; Scientific Documents Analysis**

I. INTRODUCTION

The graphical abstract (GA), a visual summary of a scholarly article's main findings, is an emerging concept in scientific publishing. Elsevier, the largest publisher[1] of scholarly articles, requests that authors provide GAs and use them for online search results in facilitating the discovery process. With no specific guidance or requirements provided to authors, 68% and 65% of papers accepted in two of the top computer vision conferences (International Conference on Computer Vision (ICCV) and Conference on Computer Vision and Pattern Recognition (CVPR)) include "teaser figures," a form of GA. 350% increase of graphical abstracts use in social sciences from 2011 to 2015 is demonstrated by Yoon et al. [1]. The significant increase of the use of GA can be related to human's superior ability of perceiving visual materials. It is believed that the human's highly developed visual cortex [2] contributes to better perception of visual information than textual information [3]. As a result, visualizations play a significant role in scientific communication. With the abundance of scientific papers, GAs complement conventional text abstracts to help users quickly identify papers relating to their interests [1], [4].

Elsevier submission guidelines [2] describe a graphical abstract as a "single, concise, pictorial and visual summary of the main findings of the article" that should "allow readers to quickly gain an understanding of the main take-home message of the paper" and "encourage browsing, promote interdisciplinary scholarship, and help readers identify more quickly which papers are most relevant to their research interests" which could be a "concluding figure from the article or a figure that is specially designed for the purpose, which captures the content of the article for readers at a single glance.". Since not all publishing venues request a GA at the time of submission and not all authors elect to provide one, services that make use of graphical abstracts only apply to a small fraction of the scientific literature [1].

In this paper, we consider the automatic selection of a "central figure" (CF) that can be used as a graphical abstract to visually summarize the paper's objectives, results, or methods, afford fast assessment of relevance, and provide a basis for new search services. This framing assumes that these CFs actually exist. To test this hypothesis, we issued 488,590 survey invitations to authors of papers on PubMed Central, asking them to identify the CF of their own publications, or indicate if no CF exists (see Figure 1). We also asked authors to explain the information represented in the figure to understand what role it plays. Figure 1 shows the survey interface. We received responses from 6,263 distinct authors across 8,353 papers. Author respondents identified a central figure for 87.6% of the papers.

Next, we use the survey responses to train a model to predict the CF in a paper. Existing GAs and teaser images are unsuitable as training data due to selection biases toward particular domains (typically visually oriented fields such as computer vision, graphics, and visualization) and because many such figures are created specifically for the purpose. We use the term central figure (CF) in this paper to distinguish from GAs. In response to publisher request, authors create GAs at the time of submission. CFs are selected from existing figures after the paper has been published. A CF

[1]Elsevier is not the only publisher requiring GAs. Other large publishers are also requiring GAs, including Wiley-Blackwell.

[2]https://www.elsevier.com/authors/journal-authors/graphical-abstract

2379-2140/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDAR.2019.00173

may be suitable as a GA, and a GA may be identified as the central figure of a paper, but the two terms are not necessarily equivalent.

Using the results of our survey as training labels, we extract features from the figures relating to figure content, the surrounding text, and the overall paper layout. We use these features in two different models: a figure-level model that considers only one figure and its associated context at a time, and a paper-level model that considers the set of figures in a paper simultaneously. The paper-level model with features from figure content combined with the surrounding text and the overall paper layout produces the best CF identification performance. We achieve top-3 accuracy of 77.9% and exact match accuracy of 33.6% for identifying CFs with our features and model. The model outperforms heuristic baselines of selecting the first figure in the paper (25.8%), the last figure in the paper (26.9%), and uniform random selection (26.4%). We find that the section title in which the figure appears and the text similarity between the abstract, the caption, and the inline reference of the figure are predictive of the CF, suggesting that authors consider these concepts in the design of their papers.

We make the following contributions:

- We conduct a large-scale survey to determine the prevalence and nature of the "central figure" of a paper, with 6,263 distinct authors describing 8,353 papers.
- We combine features extracted from surrounding text, figure type, and overall paper layout and further propose image-level model and paper-level for automated identifying CF in scientific literature. The paper-level model with all features included achieves top-3 accuracy of 77.9% and exact match accuracy of 33.6%.
- We conduct ablation studies to measure the influence of individual features to provide information for authors and publishers in each features. The experimental results show that the similarity between image description, including captions and the inline reference paragraph of images, and abstract is significant in identifying central figures in scientific documents.

II. RELATED WORK

Yoon et al. [1] investigated the frequency of graphical abstracts and the type of graphical abstracts that are adopted in social science disciplines. Hullman [4] studied the design pattern of graphical abstracts. However, only a small collection of articles were examined in both studies (772 and 54 respectively) and both studies focused on analyzing existing GAs instead of creating tools to identify GAs.

Other studies have focused on automated tools to create a representation that summarizes scientific articles have also been considered. Strobelt et al. describe DocumentCards [5], a system to extract textual and visual content from a scientific literature and produce a high level representation. Their approach relies on simple rules to create the visual summary and can not be customized for different papers.

A number of studies have focused on the mining of scientific figures. Chart classification was studied by Shao et al. [6] and Lee et al. [7]. Recent studies have been focusing on extraction of quantitative data from scientific visualizations, including line charts [8], bar charts [9], and tables [10]. Researchers have also investigated the techniques to understand the semantic messages of the scientific figures. Kembhavi et al. [11] utilized a convolution neural network (CNN) to study the problem of diagram interpretation and reasoning. Elzer et al. [12] studied the intended messages in bar charts. Besides, several visualization-based search engines have been presented. DiagramFlyer [13], introduced by Chen et al., is a search engine for data-driven diagrams. VizioMetrics.org[14] and NOA[15] are both scientific figure search engines, yet they both work primarily by examining the captions around the figures rather than specific features in the images.

III. DATA

This study was conducted using scientific papers from PubMed Central (PMC), an archive of biomedical and life science literature.

IV. CENTRAL FIGURE SURVEY

To obtain the labeled data for CF, we launched a large-scale survey asking authors to identify CFs in their papers. We extracted email addresses from the XML files provided by PMC API and sent out 488,590 survey invitations.

Authors are asked to answer two questions for each paper:

- *Click on one of the images to select ONE figure that you could call the "graphical summary" of the paper, if one exists. A figure that summarizes the key aspects of the article for readers at a single glance.*
- *What does the figure you selected represent?*

For the first question, we used the descriptive term "graphical summary" rather than central figure to indicate our intention. Authors can select from among all the figures in the paper or select "No such figure." The latter option allows us to validate whether or not the CF is a recognizable concept in the current scientific literature. For the second question, authors may select from five options: "Results", "Discussion", "Model", "Methods", and "Other".

A. Survey Results

As of December 1st, 2018, we had collected data on 8,353 distinct papers, from 6,263 distinct authors. Some authors provided responses for more than one of their papers, and some papers generated responses from more than one of its authors. The publishing year distribution is shown in Figure 2. 74.0% of evaluated papers are published after 2010. Only 12.4% (1,036) of the evaluated paper were indicated not to have a figure that satisfies our definition of CF (890) or for

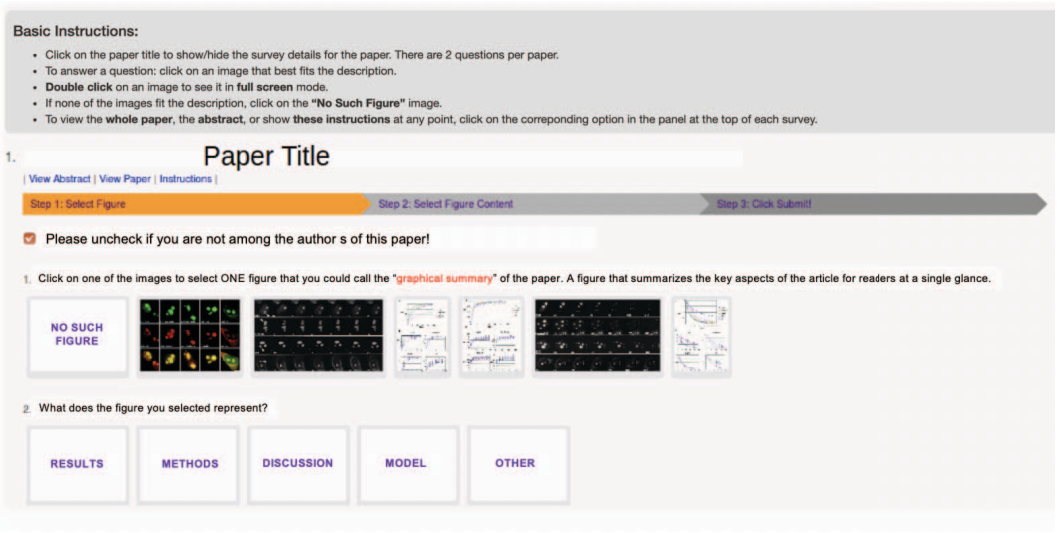


Figure 1: Snapshot of the survey. We asked authors of PubMed papers to identify the central figure of their own publications using this interface. Authors were asked to select a figure, if it exists, that summarizes the key aspects of the article, or choose "No such figure". We also asked authors to provide what kind of information the selected figure represents for the article from five options, which are "Results", "Discussion", "Model", "Methods", and "Other".

which multiple authors selected different figures (146). For the remaining 87.6% of the papers, the authors identified a single CF, suggesting acceptance of the concept.

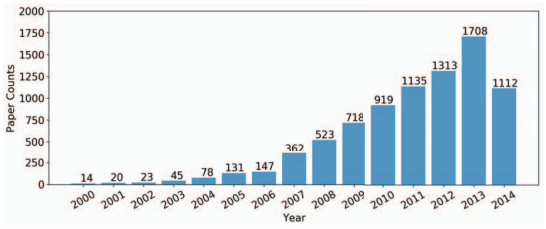
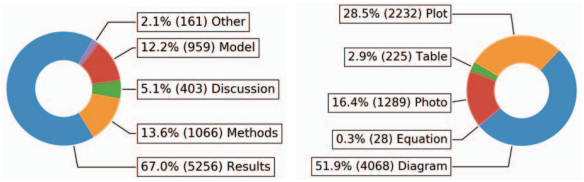


Figure 2: The publishing year distribution of evaluated papers. 74.0% of evaluated papers are published after 2010.

B. Analysis of Objectives of Central Figure

Figure 3(a) illustrates the purpose of the central figure. In 67.0% of the papers, the central figure represents results, corroborating Yoon et al. who found that graphical abstracts are most frequently used to present results [1]. This use of the central figure affords an interpretation that a paper is a delivery vehicle for one main result, which supports the idea toward a results-oriented publishing model, where the unit of publishing is a scientific workflow [16], [17] or a nano-publication [18]. Methods and model were the next most popular categories at 13.6% and 12.2%. Discussion is responsible for only 5.1% of central figures. In 2.1% of the papers the authors indicated the content as Other.



(a) (b)

Figure 3: (a) Author-indicated objective of the central figures. The survey results reveal that the most of the central figures are used to represent scientific results. (b) Pie chart of figure type distribution of central figures. 51.9% of central figures are diagrams.

C. Analysis of Figure Content of Central Figure

After collecting the survey results, the next step is to analyze the content within the figures. Using the class assignments compiled by Lee et al. [19] and classifier approach described by Lee et al. [7], we train a classifier to identify different figure types. The training dataset, including 1871 equations, 3347 photos, 2849 diagrams, 2193 tables and 4680 plots, is split into training set, validation set, and test set with 8:1:1 ratio. We finetune a pre-trained ResNet-50 [20] and obtain similar model performance reported by Lee et al. [7]. We label the central figure as one of five figure types. The totals are shown in Figure 3(b). 51.9% of the central figures in the evaluated papers are diagrams, which agrees with the findings of Yoon et al. [1] that most GAs are diagrams. We found that despite the fact that plots (graphs) and tables can both be used for presenting data,

plots are much more popular when it comes to presenting key information. This result agrees with Cleveland et al. [21], who showed that fractional graph area (FGA) increases as one moves from social to mathematical and then to natural science. This finding also agrees with results from Smith et al. [22] that suggest that technical fields of science tend to use more graph-oriented figures than table-oriented. The fact that equations are rarely found among central figures is consistent with findings by Fawcett and Higginson [23].

V. MODEL CENTRAL FIGURES

The next task is to train a model to select the CF from all figures in a paper. We consider three sources for identifying the CF: (1) the content of the image, (2) the text describing the image (and as it relates to the abstract), and (3) the location of the figure in the paper. We will elaborate on each source in the following subsections.

A. Image Content

The content of the image itself is not a good predictor of centrality, as we will show, since many figures in a paper look alike and our training set is limited. However, we find it useful to consider the broad type of the image as a feature. We classify each image into one of five categories, diagram, plot, table, equation, and photo, using the classifier developed by Lee et al. [19]. We label all the figures in the datasets by running the figure type classifier mentioned in Section IV-C. This categorical feature is encoded in a 5-d one hot vector to represent the visual content of the figure.

B. Text Features

Each figure is described in both a caption and in one or more inline references in the body of the paper. While both sources of text can be used as features alone, we also consider the similarity of these excerpts to the abstract as an indicator that the text serves as a summary of the overall paper. We will first explain the process of extracting surrounding text of a figure from the paper and then describe the similarity measures.

Text Extraction: We collect captions of the figures from PubMed. To extract inline references in the body of the paper, we use Science Parse [3] to parse the papers in pdf format provided by PubMed and obtain the full text in structural form. We then search the pattern that consist of words, including *Figure*, *Fig*, and *Table*, followed by a number using regular expression. We select the paragraph blocks contain the inline references in between two break line ($\backslash n$) characters. Finally, we match the index between the inline references and the captions of the figures.

Similarity Between Caption and Abstract: An abstract is a summary of the paper's results. High similarity between a figure's caption and the paper's abstract would therefore indicate that the figure plays a potential summarizing role as well. We experiment with three different similarity measures: (1) TF-IDF, (2) Elmo-avg and (3) Elmo-DynaMax.

- **TF-IDF + Cosine Similarity**: We preprocess the captions, the inline references and abstracts from training set by tokenizing the documents and removing the stop words. We pick the most frequent 1,024 words to construct TF-IDF weights and the weights are acquired from the preprocessed training set. The dimensionality is picked to match with the competitive similarity measures. For each image, we apply the weights to the concatenation of the caption and the inline reference. Every abstract is also embedded in the TF-IDF vector. We finally compute the cosine similarity between the two vectors. We will refer this similarity measure as TF-IDF for simplicity.
- **Elmo-avg** [24]: Elmo is one of the state-of-the-art contextualized word embedding models. The word representations are functions of the internal states of a bidirectional language model. Elmo has been trained in large-scale scientific documents from PubMed, making Elmo a natural candidate for our task. The contextualized word representation is obtained from the top layer of the pre-trained Elmo model, and we average the word representations to acquire the representation vector for both the image descriptions and abstract. The cosine similarity is computed between the averaged word vectors of image descriptions and abstract.
- **Elmo-DynaMax** [24] [25]: Zhelezniak proposed a similarity measure, DynaMax, that dynamically extracts max-pool features based on the sentence pair. This method outperforms current baselines on several tasks [25]. The DynaMax similarity is computed between the image descriptions and the paper's abstracts from Elmo word vectors.

C. Layout

We produced two numerical features and one categorical feature from image position: (1) normalized section index, (2) image order, and (3) section heading:

- **Normalized section index**: Normalized section index is used to represent the position of the image within the layout of the paper. For example, in a paper with sections "Introduction," "Methods," and "Results," the corresponding sequentially increasing section identifiers would be 0, 1, and 2. The normalized version of the identifiers would be its original value divided by the maximum identifier value.
- **Image Order**: The sequentially increasing numerical identifier for an image based on its order of occurrence in a paper.
- **Section Heading**: The survey shows 67% of the cases with central figures are used to represent results. To capture this feature, we constructed unigram representations of the section headings of papers in our dataset

[3]https://github.com/allenai/science-parse

for both the entire headings and their distinct words. We then transformed the top ten frequently occurring words in the section headings unigram model in to ten unique boolean classification features, each denoting "1" for whether the corresponding word occurred in a given section heading, and "0" otherwise.

VI. MODELS

In this section, we illustrate two different models to identify central figures.

A. Figure-level Model

This approach attempts to predict whether an individual figure is a central figure without considering the other figures in the paper. Let $X = \{x_i : i\}$ be the features of the images and each image corresponds to a label y_i, where $y_i \in \{-1, 1\}$. central figures are labeled as 1 and non-central figures are labeled as -1. We learn a mapping function $f : X \rightarrow Y$ using machine learning techniques, which include logistic regression, random forest, gradient boosting, support vector machine (SVM), and neural networks.

To pick the central figure from a paper $A = \{a_j : j\}$, we select the figure with highest probability predicted by each classifier f: $C_j = \arg\max_{x_i \in A_j}(P(f(x_i) = 1))$

B. Paper-level Model

This approach predicts the position of the central figure given all figures in a paper. For example, if a paper has 10 figures, we concatenate all 10 feature vectors, and then predict an integer 0..9 to indicate which figure is the central figure. Let $V = \{v_j : j\}$ represent a feature vector for each paper. v_j is a n x d vector where n is a parameter and d is the dimension of image feature. Since there are variable number of figures in different papers and basic machine learning models only take fixed dimension inputs, we introduce a hyperparameter n to serve as the threshold for the number of figures. We pad zero if the number of figure is smaller than n in a paper. For the case where the number of figure is larger than n, we select n figures whose captions are most similar to the abstract based on our TF-IDF model to fill v_j. The classifiers f will learn a mapping function $f : V \rightarrow I$, where $I \in \{0, 1, ..., n\}$ is the index of the central figure. We experiment on the ensemble and regression learning methods plus neural networks listed in previous sub section.

VII. EXPERIMENTS

In this section, we first define evaluation metrics on our task. Next, we explain the implementation details of our models. Baseline models are next introduced as comparisons. Finally, quantitative results of our image-based model and paper-based model are presented.

A. Evaluation Metrics

The image accuracy is applied to evaluate the image based model. The image accuracy is defined as:

$$imageACC = \frac{\text{True Positive + True Negative}}{\text{Total number of the images}} \quad (1)$$

We use two metrics, ACC and ACC@3, to evaluate the overall capability of selecting central figure from a paper.

$$ACC = \frac{N_c}{N_t} \quad (2)$$

where N_c is the number of the papers with correct central figure prediction and N_t is the total number of the papers.

$$ACC@3 = \frac{N_c@3}{N_t} \quad (3)$$

where N_c is the number of the papers with the correct central figure prediction, within the 3 figures with highest probability. N_t is the total number of papers.

B. Implementation Details

We remove the evaluated papers which do not have central figures and split the data into training, validation, and test set with 8:1:1 ratio. We run our experiments on the training and validation set. The final model is trained by the data from both training set and validation set and accuracy results reported below are conducted on test set.

Regression and ensemble models are trained using Scikit-learn and we use default values for hyperparameters. The neural network model include three fully connected layers with dimensions 100-100-n. Drop out layers with drop out rate 0.2 are inserted between the fully connected layers. All the models are trained with learning rate 0.01 and 0.01 decay for 100 epochs.

C. Baseline Models

We introduce three naive baseline models as comparisons.

- **Pick First**: First image is selected as prediction in this model. We pick first three images in the paper as top three guesses for ACC@3 evaluation metric.
- **Pick Last**: The last image is selected as prediction in this model. We pick last three images in the paper as top three guesses for ACC@3 evaluation metric.
- **Randomly Select**: We randomly select an image as prediction. Three images are randomly selected as the top 3 guesses for the ACC@3 evaluation metric.

Table I: Performance of baseline models.

Pick First		Pick Last		Randomly Select	
ACC	ACC@3	ACC	ACC@3	ACC	ACC@3
0.258	0.704	0.269	0.679	0.264	0.706

The performance of the baseline models is shown in Table I. There are 4.68 images in a paper on average in the dataset. Accuracy of 0.264 from randomly select model makes sense.

D. Image-level Model

Table II: Image accuracy (Equ. 1) of central figure classification from image-based models.

Logistic Regression	Random Forest	Gradient Boosting	SVM	Neural Networks
0.626	0.616	0.621	0.673	0.684

Table II shows the classification results from each classifier on identifying central figure. Overall, every classifier is able to achieve more than 60% accuracy on classifying between central figure and non-central figure on figure level. The accuracy of central figure prediction given paper is shown in Table III. Not surprisingly, this simple image-based model does not perform well on selecting the central figure from a list of figures. The model is not able to learn the structural relationships between figures from the same paper.

E. Paper-level Model

We run an experiment to determine hyperparamter n (the threshold for the number of figure to be accommodated for the input V). The experimental results are shown in Figure 4. The blue line, which corresponds to the y axis on the left, is the accuracy of the model and the red line shows the percentage of central figures that were left out because of our selection of n. The selection of n has insignificant influence to the model when n is larger than 6 and the maximum number of figure a paper has in our validation set is 12. Thus, we pick $n = 15$ for the rest of the experiments.

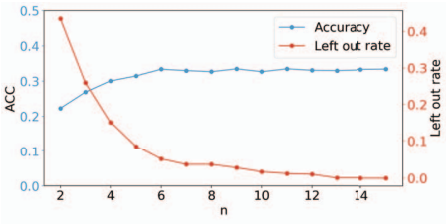


Figure 4: Experimental results on hyperparameter n. When n is larger than 6, selection of n does not affect the accuracy of the model.

The results for paper-level model with different feature combinations are shown in Table III. The logistic regression classifier performs the best among all the models, including neural networks. The poor performance from neural networks is likely due to insufficient data and low dimensional features. The text context has the most predictive power among the three sets of features, while the visual figure content has the least. Our interpretation of these results is that similarity between the figure caption and paper abstract not only provides the representation of the image but it also suggests the relationship to the paper. On the other hand, without any further information of the paper, figure type is irrelevant to determine central figure in this generation of the model. Also, surprisingly, the simple TF-IDF representation produces better performance than the Elmo word representations in more than half of the models. We speculate that the terms used in captions, and the contexts in which they are used are sufficiently specialized to allow the simpler representation to outperform pre-learned representations based on a larger corpus of text. Using max-pool followed by fuzzy Jaccard index to compute similarity between two documents has superior results over averaging the word vectors, which agrees with the findings of Zhelezniak et al. [25]. Considering the difficulty of the task and the variability of scientific figures, we see our results as a reasonable start for automatically identifying central figures.

We investigate the effectiveness of our feature selection. We replace the similarity between abstract, caption and inline reference with a text representation vector of the caption and inline reference. We also experiment with using image embedding extracted from pre-trained ResNet-50 [20] instead of categorical feature based on the figure content. Even though the model was trained on 1M natural images, we find that the embeddings that capture visual patterns and colors are sufficiently general to represent the combinations of edges and shapes that comprise artificial images as well. The results are presented in Table IV. The experiments demonstrate that using similarity between abstract, image caption, and inline reference boost the central figure identification performance and that the categorical label of image content is more beneficial than image embedding. Our interpretation is that the high similarity between a paper's abstract and a image's surrounding text does indicate the centralization of the figure and that the original text representations and image embeddings are too sparse and noisy for the model to learn an effective function.

VIII. DISCUSSION

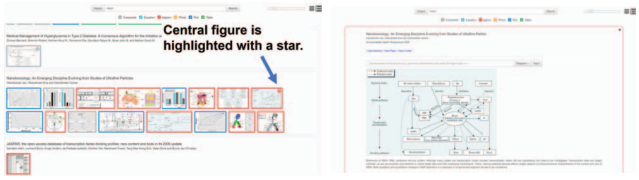
Scholarly communication is moving away from just a simple PDF. Individual insights, experiments, and conclusions can be communicated across different media and platforms. In this paper, we focus on the role that visual information plays in communicating the key results, models or concepts. The idea behind a central figure is that it provides an alternative access point to the content of the paper. In some papers, it can reveal the key results and conclusions better than the title, abstract, keywords or authors. Figure 5 shows two prototypes of how to introduce the central figure in an image-oriented scientific search interface, viziometrics.org. As shown in Figure 5(a), the central figure is highlighted with a star on the search interface. The entry page could feature the central figure along with textual abstracts as shown in 5(b). With these two additional features, users are able to quickly ascertain the overall concept of the article with the help of central figure at a single glance.

Table III: The results of paper-level model with different classifiers. Surprisingly, logistic regression outperforms random forest and gradient boosting. Textual content is the most useful feature on recognizing central figure, compared to visual content and the position feature.

			Logistic Regression		Random Forest		Gradient Boosting		SVM		Neural Network	
Text	Visual	Layout	ACC	ACC@3	ACC	ACC@3	ACC	ACC@3	ACC	ACC@3	ACC	ACC@3
Figure-level model												
TF-IDF	v	v	0.140	0.691	0.248	0.703	0.126	0.685	0.126	0.690	0.142	0.688
Paper-level model												
TF-IDF	-	-	0.302	0.764	0.318	0.724	0.314	0.760	0.314	0.756	0.311	0.718
-	v	-	0.289	0.730	0.278	0.693	0.286	0.731	0.319	0.748	0.282	0.724
-	-	v	0.296	0.757	0.284	0.723	0.284	0.741	0.299	0.746	0.284	0.756
TF-IDF	v	-	0.317	0.757	0.292	0.712	0.295	0.764	0.322	0.749	0.276	0.716
TF-IDF	-	v	0.323	0.782	0.277	0.690	0.335	0.765	0.273	0.708	0.280	0.750
-	v	v	0.312	0.742	0.267	0.703	0.302	0.720	0.300	0.724	0.293	0.739
Elmo-avg	v	v	0.329	0.771	0.262	0.670	0.314	0.771	0.299	0.729	0.299	0.738
Elmo-DynaMax	v	v	0.330	0.769	0.285	0.679	0.321	0.778	0.299	0.733	0.282	0.739
TF-IDF	v	v	**0.336**	**0.779**	0.267	0.701	0.314	0.760	0.302	0.727	0.306	0.745

Table IV: Experimental results on using text representation and image embedding. *Sim()* indicates the model uses similarity between paper's abstract, image caption, and inline reference computed by the text representation in the parenthesis. *Vec()* implies the model utilizes the representation vectors derived from the model in the parenthesis. *Label* represents the model includes the categorical label described in Section V-A as image content feature. We can observe that using similarity and the categorical label of image content produces better performance than using representations.

			Logistic Regression	
Text	Visual	Layout	ACC	ACC@3
Sim(TF-IDF)	Label	v	**0.336**	**0.779**
Using Image Embedding from Pre-trained ResNet-50				
Sim(TF-IDF)	Vec(ResNet)	v	0.323	0.761
Using Text Representation Vectors				
Vec(TF-IDF)	Label	v	0.310	0.722
Vec(Elmo-avg)	Label	v	0.300	0.723
Using Both Text Representation Vectors and Image Embedding				
Vec(TF-IDF)	Vec(ResNet)	v	0.288	0.741
Vec(Elmo-avg)	Vec(ResNet)	v	0.293	0.705



(a) (b)

Figure 5: Prototype interfaces on viziometrics.org allow individuals to search for images from scientific literature with the aid of "central figures". (a) Central figure is starred for easy recognition on searching interface. (b) Prototype of entry page for each article. The entry interface of each article could be led with the central figure along with textual abstract to help the users understand the articles quickly.

Publishing culture has changed dramatically over the last few decades due to the introduction of multiple open access platforms, such as arXiv and PubMed, as well as the significant increase of scientific publications. With more open access platforms available, the accessibility of innovative ideas pushes the advance of science and allows the community to share and communicate ideas in different formats. The presentation of new scientific ideas is no longer restricted to traditional document copies or digitized pdf formats. For example, we can easily find comprehensive ablation studies of the state-of-the-art deep learning models on GitHub. Google AI [4] hosts a blog to introduce and advertise their progress on innovative scientific findings and technologies. Plus, the overwhelming scale of scientific publications [26], [27] that are published every year. Several recent studies [1], [4] have explored new measures for the community to quickly grasp the main messages of the scientific documents. The scientific publishing enterprise has shifted to be more open to the public and less restrictive on format, and we believe the identification and extraction of central figures can play an important role in the evolution of the scientific communication. A central figure provides a visual summary of the key results, objectives, or methods of a paper. It is adaptable to varying media and platforms, easy to share, and information-rich. We can see each central figure as a visual "nanopublication" [18] and use it to reduce the redundancy of traditional publications. Every central figure is a module of condensed ideas and can be transmitted and shared easily. Therefore, central figures can contribute greatly in the evolution of scientific communication with quick idea transferring and flexible publishing platforms.

IX. CONCLUSION

Visualizations will play an increasingly important role in scholarly communication. The goal of this paper was to focus on visual objects that convey the central findings of a research paper. We collected more than eight thousand

[4]https://ai.googleblog.com/

labeled data for central figure identification from a large-scale survey. 87.6% of the evaluated papers included a central figure noted by the authors. This was evidence that central figures exist and they perform a function in scholarly communication. We extracted features from the figure content, surrounding text, and the overall paper layout as a way of training a figure-level model and a paper-level model. The results reveal that the paper-level model with all features produce the best performance overall in identifying central figures. We achieve top-3 accuracy of 77.9% and exact match accuracy of 34%. We also demonstrate that the caption, inline description, and layout shows higher importance than figure content in this task. Survey data and code are publicly available [5], and we hope the released data can attract the community to investigate this problem and further contribute to the scientific communication.

REFERENCES

[1] J. Yoon and E. Chung, "An investigation on graphical abstracts use in scholarly articles," *International Journal of Information Management*, no. 1, pp. 1371–1379, 2017.

[2] C. Ware, *Information visualization: perception for design*. Elsevier, 2012.

[3] D. L. Nelson, V. S. Reed, and J. R. Walling, "Pictorial superiority effect." *Journal of Experimental Psychology: Human Learning and Memory*, vol. 2, no. 5, p. 523, 1976.

[4] J. Hullman and B. Bach, "Picturing science: Design patterns in graphical abstracts," *DIAGRAMS*, 2018.

[5] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, and O. Deussen, "Document cards: A top trumps visualization for documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1145–1152, 2009.

[6] M. Shao and R. P. Futrelle, "Recognition and classification of figures in pdf documents," in *GREC*. Springer, 2005.

[7] P. Lee, T. S. Yang, J. West, and B. Howe, "Phyloparser: A hybrid algorithm for extractingphylogenies from dendrograms," in *ICDAR*, 2017.

[8] N. Siegel, Z. Horvitz, R. Levin, S. Divvala, and A. Farhadi, "Figureseer: Parsing result-figures in research papers," in *ECCV*. Springer, 2016, pp. 664–680.

[9] R. A. Al-Zaidy and C. L. Giles, "Automatic extraction of data from bar charts," in *K-CAP*. ACM, 2015, p. 30.

[10] J. Fang, P. Mitra, Z. Tang, and C. L. Giles, "Table header detection and classification." in *AAAI*, 2012, pp. 599–605.

[11] A. Kembhavi, M. Salvato, E. Kolve, M. Seo, H. Hajishirzi, and A. Farhadi, "A diagram is worth a dozen images," in *ECCV*. Springer, 2016, pp. 235–251.

[5]https://github.com/viziometrics/centraul_figure

[12] S. Elzer, S. Carberry, and I. Zukerman, "The automated understanding of simple bar charts," *Artificial Intelligence*, vol. 175, no. 2, pp. 526–555, 2011.

[13] Z. Chen, M. Cafarella, and E. Adar, "Diagramflyer: A search engine for data-driven diagrams," in *WWW*. ACM, 2015.

[14] P. Lee, J. West, and B. Howe, "Viziometrix: A platform for analyzing the visual information in big scholarly data," in *BigScholar*, 2016.

[15] J. Charbonnier, L. Sohmen, J. Rothman, B. Rohden, and C. Wartena, "Noa: A search engine for reusable scientific images beyond the life sciences," in *ECIR*. Springer, 2018, pp. 797–800.

[16] L. Bavoil, S. Callahan, P. Crossno, J. Freire, C. Scheidegger, C. Silva, and H. Vo, "Vistrails: Enabling interactive multiple-view visualizations," in *IEEE Visualization*, 2005.

[17] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover, M. R. Pocock, A. Wipat *et al.*, "Taverna: a tool for the composition and enactment of bioinformatics workflows," *Bioinformatics*, vol. 20, no. 17, pp. 3045–3054, 2004.

[18] B. Mons and J. Velterop, "Nano-publication in the e-science era," in *SWASD*, 2009.

[19] P. Lee, J. D. West, and B. Howe, "Viziometrics: Analyzing visual information in the scientific literature," *IEEE Transactions on Big Data*, 2017.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[21] W. S. Cleveland, "Graphs in scientific publications," *The American Statistician*, vol. 38, no. 4, pp. 261–269, 1984.

[22] L. D. Smith, L. A. Best, D. A. Stubbs, A. B. Archibald, and R. Roberson-Nay, "Constructing knowledge: The role of graphs and tables in hard and soft psychology." *American Psychologist*, vol. 57, no. 10, p. 749, 2002.

[23] T. W. Fawcett and A. D. Higginson, "Heavy use of equations impedes communication among biologists," *PNAS*, vol. 109, no. 29, pp. 11 735–11 739, 2012.

[24] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *NAACL*, 2018.

[25] V. Zhelezniak, A. Savkov, A. Shen, F. Moramarco, J. Flann, and N. Y. Hammerla, "Don't settle for average, go for the max: Fuzzy sets and max-pooled word vectors," in *ICLR*, 2019.

[26] P. Larsen and M. Von Ins, "The rate of growth in scientific publication and the decline in coverage provided by science citation index," *Scientometrics*, vol. 84, no. 3, 2010.

[27] L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, 2015.