

# The Name of the Title Is Hope

Ben Trovato\*  
G.K.M. Tobin\*  
trovato@corporation.com  
webmaster@marysville-ohio.com  
Institute for Clarity in Documentation  
Dublin, Ohio, USA

Lars Thørväld  
The Thørväld Group  
Hekla, Iceland  
larst@affiliation.org

Valerie Béranger  
Inria Paris-Rocquencourt  
Rocquencourt, France

Aparna Patel  
Rajiv Gandhi University  
Doimukh, Arunachal Pradesh, India

Huifen Chan  
Tsinghua University  
Haidian Qu, Beijing Shi, China

Charles Palmer  
Palmer Research Laboratories  
San Antonio, Texas, USA  
cpalmer@prl.com

John Smith  
The Thørväld Group  
Hekla, Iceland  
jsmith@affiliation.org

Julius P. Kumquat  
The Kumquat Consortium  
New York, USA  
jpkumquat@consortium.net



Figure 1: Seattle Mariners at Spring Training, 2010.

## Abstract

A clear and well-documented  $\text{\LaTeX}$  document is presented as an article formatted for publication by ACM in a conference proceedings or journal publication. Based on the “acmart” document class, this article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work.

## CCS Concepts

• **Do Not Use This Code** → **Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate

\*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-XXXX-X/2018/06  
<https://doi.org/XXXXXXX.XXXXXXX>

the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

## Keywords

Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

### ACM Reference Format:

Ben Trovato, G.K.M. Tobin, Lars Thørväld, Valerie Béranger, Aparna Patel, Huifen Chan, Charles Palmer, John Smith, and Julius P. Kumquat. 2018. The Name of the Title Is Hope. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 7 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 Introduction

ACM’s consolidated article template, introduced in 2017, provides a consistent  $\text{\LaTeX}$  style for use across ACM publications, and incorporates accessibility and metadata-extraction functionality necessary for future Digital Library endeavors. Numerous ACM and SIG-specific  $\text{\LaTeX}$  templates have been examined, and their unique features incorporated into this single new template.

If you are new to publishing with ACM, this document is a valuable guide to the process of preparing your work for publication.

If you have published with ACM before, this document provides insight and instruction into more recent changes to the article template.

The “acmart” document class can be used to prepare articles for any ACM publication — conference or journal, and for any stage of publication, from review to final “camera-ready” copy, to the author’s own version, with very few changes to the source.

## 2 Template Overview

As noted in the introduction, the “acmart” document class can be used to prepare many different kinds of documentation — a double-anonymous initial submission of a full-length technical paper, a two-page SIGGRAPH Emerging Technologies abstract, a “camera-ready” journal article, a SIGCHI Extended Abstract, and more — all by selecting the appropriate *template style* and *template parameters*.

This document will explain the major features of the document class. For further information, the *L<sup>A</sup>T<sub>E</sub>X User’s Guide* is available from <https://www.acm.org/publications/proceedings-template>.

### 2.1 Template Styles

The primary parameter given to the “acmart” document class is the *template style* which corresponds to the kind of publication or SIG publishing the work. This parameter is enclosed in square brackets and is a part of the `documentclass` command:

```
\documentclass[STYLE]{acmart}
```

Journals use one of three template styles. All but three ACM journals use the `acmsmall` template style:

- `acmsmall`: The default journal template style.
- `acmlarge`: Used by JOCCH and TAP.
- `acmtog`: Used by TOG.

The majority of conference proceedings documentation will use the `acmconf` template style.

- `sigconf`: The default proceedings template style.
- `sigchi`: Used for SIGCHI conference articles.
- `sigplan`: Used for SIGPLAN conference articles.

### 2.2 Template Parameters

In addition to specifying the *template style* to be used in formatting your work, there are a number of *template parameters* which modify some part of the applied template style. A complete list of these parameters can be found in the *L<sup>A</sup>T<sub>E</sub>X User’s Guide*.

Frequently-used parameters, or combinations of parameters, include:

- `anonymous, review`: Suitable for a “double-anonymous” conference submission. Anonymizes the work and includes line numbers. Use with the `\acmSubmissionID` command to print the submission’s unique ID on each page of the work.
- `authorversion`: Produces a version of the work suitable for posting by the author.
- `screen`: Produces colored hyperlinks.

This document uses the following string as the first command in the source file:

```
\documentclass[sigconf,authordraft]{acmart}
```

## 3 Modifications

Modifying the template — including but not limited to: adjusting margins, typeface sizes, line spacing, paragraph and list definitions, and the use of the `\vspace` command to manually adjust the vertical spacing between elements of your work — is not allowed.

**Your document will be returned to you for revision if modifications are discovered.**

## 4 Typefaces

The “acmart” document class requires the use of the “Libertine” typeface family. Your T<sub>E</sub>X installation should include this set of packages. Please do not substitute other typefaces. The “lmodern” and “ltimes” packages should not be used, as they will override the built-in typeface families.

## 5 Title Information

The title of your work should use capital letters appropriately — <https://capitalizemytitle.com/> has useful rules for capitalization. Use the `title` command to define the title of your work. If your work has a subtitle, define it with the `subtitle` command. Do not insert line breaks in your title.

If your title is lengthy, you must define a short version to be used in the page headers, to prevent overlapping text. The `title` command has a “short title” parameter:

```
\title[short title]{full title}
```

## 6 Authors and Affiliations

Each author must be defined separately for accurate metadata identification. As an exception, multiple authors may share one affiliation. Authors’ names should not be abbreviated; use full first names wherever possible. Include authors’ e-mail addresses whenever possible.

Grouping authors’ names or e-mail addresses, or providing an “e-mail alias,” as shown below, is not acceptable:

```
\author{Brooke Aster, David Mehldau}
\email{dave,judy,steve@university.edu}
\email{firstname.lastname@phillips.org}
```

The `authornote` and `authornotemark` commands allow a note to apply to multiple authors — for example, if the first two authors of an article contributed equally to the work.

If your author list is lengthy, you must define a shortened version of the list of authors to be used in the page headers, to prevent overlapping text. The following command should be placed just after the last `\author{}` definition:

```
\renewcommand{\shortauthors}{McCartney, et al.}
```

Omitting this command will force the use of a concatenated list of all of the authors’ names, which may result in overlapping text in the page headers.

The article template’s documentation, available at <https://www.acm.org/publications/proceedings-template>, has a complete explanation of these commands and tips for their effective use.

Note that authors’ addresses are mandatory for journal articles.

## 7 Rights Information

Authors of any work published by ACM will need to complete a rights form. Depending on the kind of work, and the rights management choice made by the author, this may be copyright transfer, permission, license, or an OA (open access) agreement.

Regardless of the rights management choice, the author will receive a copy of the completed rights form once it has been submitted. This form contains  $\LaTeX$  commands that must be copied into the source document. When the document source is compiled, these commands and their parameters add formatted text to several areas of the final document:

- the “ACM Reference Format” text on the first page.
- the “rights management” text on the first page.
- the conference information in the page header(s).

Rights information is unique to the work; if you are preparing several works for an event, make sure to use the correct set of commands with each of the works.

The ACM Reference Format text is required for all articles over one page in length, and is optional for one-page articles (abstracts).

## 8 CCS Concepts and User-Defined Keywords

Two elements of the “acmart” document class provide powerful taxonomic tools for you to help readers find your work in an online search.

The ACM Computing Classification System — <https://www.acm.org/publications/class-2012> — is a set of classifiers and concepts that describe the computing discipline. Authors can select entries from this classification system, via <https://dl.acm.org/ccs/ccs.cfm>, and generate the commands to be included in the  $\LaTeX$  source.

User-defined keywords are a comma-separated list of words and phrases of the authors’ choosing, providing a more flexible way of describing the research being presented.

CCS concepts and user-defined keywords are required for all articles over two pages in length, and are optional for one- and two-page articles (or abstracts).

## 9 Sectioning Commands

Your work should use standard  $\LaTeX$  sectioning commands: `\section`, `\subsection`, `\subsubsection`, `\paragraph`, and `\subparagraph`. The sectioning levels up to `\subsubsection` should be numbered; do not remove the numbering from the commands.

Simulating a sectioning command by setting the first word or words of a paragraph in boldface or italicized text is **not allowed**.

Below are examples of sectioning commands.

### 9.1 Subsection

This is a subsection.

9.1.1 *Subsubsection*. This is a subsubsection.

*Paragraph*. This is a paragraph.

Subparagraph This is a subparagraph.

## 10 Tables

The “acmart” document class includes the “booktabs” package — <https://ctan.org/pkg/booktabs> — for preparing high-quality tables.

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
$\pi$	1 in 5	Common in math
\$	4 in 5	Used in business
$\Psi_1^2$	1 in 40,000	Unexplained usage

Table captions are placed *above* the table.

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper “floating” placement of tables, use the environment **table** to enclose the table’s contents and the table caption. The contents of the table itself must go in the **tabular** environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules. Again, detailed instructions on **tabular** material are found in the  *$\LaTeX$  User’s Guide*.

Immediately following this sentence is the point at which Table 1 is included in the input file; compare the placement of the table here with the table in the printed output of this document.

To set a wider table, which takes up the whole width of the page’s live area, use the environment **table\*** to enclose the table’s contents and the table caption. As with a single-column table, this wide table will “float” to a location deemed more desirable. Immediately following this sentence is the point at which Table 2 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed output of this document.

Always use `midrule` to separate table header rows from data rows, and use it only for this purpose. This enables assistive technologies to recognise table headers and support their users in navigating tables more easily.

## 11 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

### 11.1 Inline (In-text) Equations

A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$ . . . $`. You can use any of the symbols and structures, from  $\alpha$  to  $\omega$ , available in  $\LaTeX$  [10]; this section will simply show a few examples of in-text equations in context. Notice how this equation:  $\lim_{n \rightarrow \infty} x = 0$ , set here in in-line math style, looks slightly different when set in display style. (See next section).

### 11.2 Display Equations

A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in  $\LaTeX$ ; this section will just give a couple of examples of display equations in context. First, consider the



Table 2: Some Typical Commands

Command	A Number	Comments
<code>\author</code>	100	Author
<code>\table</code>	300	For tables
<code>\table*</code>	400	For wider tables

equation, shown as an inline equation above:

$$\lim_{n \rightarrow \infty} x = 0 \tag{1}$$

Notice how it is formatted somewhat differently in the `display-math` environment. Now, we'll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty} x_i = \int_0^{\pi+2} f \tag{2}$$

just to demonstrate  $\LaTeX$ 's able handling of numbering.

12 Figures

The “figure” environment should be used for figures. One or more images can be placed within a figure. If your figure contains third-party material, you must clearly identify it as such, as shown in the example below.



Figure 2: 1907 Franklin Model D roadster. Photograph by Harris & Ewing, Inc. [Public domain], via Wikimedia Commons. (<https://goo.gl/VLCRBB>).

Your figures should contain a caption which describes the figure to the reader.

Figure captions are placed *below* the figure.

Every figure should also have a figure description unless it is purely decorative. These descriptions convey what’s in the image

to someone who cannot see it. They are also used by search engine crawlers for indexing images, and when images cannot be loaded.

A figure description must be unformatted plain text less than 2000 characters long (including spaces). **Figure descriptions should not repeat the figure caption – their purpose is to capture important information that is not already provided in the caption or the main text of the paper.** For figures that convey important and complex new information, a short text description may not be adequate. More complex alternative descriptions can be placed in an appendix and referenced in a short figure description. For example, provide a data table capturing the information in a bar chart, or a structured list representing a graph. For additional information regarding how best to write figure descriptions and why doing this is so important, please see <https://www.acm.org/publications/taps/describing-figures/>.

12.1 The “Teaser Figure”

A “teaser figure” is an image, or set of images in one figure, that are placed after all author and affiliation information, and before the body of the article, spanning the page. If you wish to have such a figure in your article, place the command immediately before the `\maketitle` command:

```
\begin{teaserfigure}
  \includegraphics[width=\textwidth]{sampleteaser}
  \caption{figure caption}
  \Description{figure description}
\end{teaserfigure}
```

13 Citations and Bibliographies

The use of Bib $\TeX$  for the preparation and formatting of one’s references is strongly recommended. Authors’ names should be complete — use full first names (“Donald E. Knuth”) not initials (“D. E. Knuth”) — and the salient identifying features of a reference should be included: title, year, volume, number, pages, article DOI, etc.

The bibliography is included in your source document with these two commands, placed just before the `\end{document}` command:

```
\bibliographystyle{ACM-Reference-Format}
\bibliography{bibfile}
```

where “bibfile” is the name, without the “.bib” suffix, of the Bib $\TeX$  file.

Citations and references are numbered by default. A small number of ACM publications have citations and references formatted in the “author year” style; for these exceptions, please include this command in the **preamble** (before the command “`\begin{document}`”) of your  $\LaTeX$  source:

```
\citestyle{acmauthoryear}
```

Some examples. A paginated journal article [? ], an enumerated journal article [? ], a reference to an entire issue [? ], a monograph (whole book) [? ], a monograph/whole book in a series (see 2a in spec. document) [? ], a divisible-book such as an anthology or compilation [? ] followed by the same example, however we only output the series if the volume number is given [? ] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [? ], a chapter in a divisible book in a series [? ], a multi-volume work as book [? ], a couple of articles in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [6? ], a proceedings article with all possible elements [? ], an example of an enumerated proceedings article [? ], an informally published work [? ], a couple of preprints [3, 4], a doctoral dissertation [5], a master's thesis: [? ], an online document / world wide web resource [1, 14, 18], a video game (Case 1) [13] and (Case 2) [12] and [11] and (Case 3) a patent [17], work accepted for publication [16], 'YYYYb'-test for prolific author [? ] and [? ]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [9]. Boris / Barbara Beeton: multi-volume works as books [8] and [7]. A couple of citations with DOIs: [9? ]. Online citations: [18–20]. Artifacts: [15] and [2].

## 14 Acknowledgments

Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgment section, which is placed just before the reference section in your document.

This section has a special environment:

```
\begin{acks}
...
\end{acks}
```

so that the information contained therein can be more easily collected during the article metadata extraction phase, and to ensure consistency in the spelling of the section heading.

Authors should not prepare this section as a numbered or un-numbered \section; please use the “acks” environment.

## 15 Appendices

If your work needs an appendix, add it before the “\end{document}” command at the conclusion of your source document.

Start the appendix with the “appendix” command:

```
\appendix
```

and note that in the appendix, sections are lettered, not numbered. This document has two appendices, demonstrating the section and subsection identification method.

## 16 Multi-language papers

Papers may be written in languages other than English or include titles, subtitles, keywords and abstracts in different languages (as a rule, a paper in a language other than English should include an English title and an English abstract). Use language=. . . for every language used in the paper. The last language indicated is the main language of the paper. For example, a French paper with

additional titles and abstracts in English and German may start with the following command

```
\documentclass[sigconf, language=english, language=german,
language=french]{acmart}
```

The title, subtitle, keywords and abstract will be typeset in the main language of the paper. The commands \translatedXXX, XXX begin title, subtitle and keywords, can be used to set these elements in the other languages. The environment translatedabstract is used to set the translation of the abstract. These commands and environment have a mandatory first argument: the language of the second argument. See sample-sigconf-i13n.tex file for examples of their usage.

## 17 SIGCHI Extended Abstracts

The “sigchi-a” template style (available only in L<sup>A</sup>T<sub>E</sub>X and not in Word) produces a landscape-orientation formatted article, with a wide left margin. Three environments are available for use with the “sigchi-a” template style, and produce formatted output in the margin:

**sidebar:** Place formatted text in the margin.

**marginfigure:** Place a figure in the margin.

**marginable:** Place a table in the margin.

## Acknowledgments

To Robert, for the bagels and explaining CMYK and color spaces.

## References

- [1] Rafal Ablamowicz and Bertfried Fauser. 2007. *CLIFFORD: a Maple 11 Package for Clifford Algebra Computations, version 11*. Retrieved February 28, 2008 from <http://math.tntech.edu/rafal/cliff11/index.html>
- [2] Sam Anzaroot and Andrew McCallum. 2013. *UMass Citation Field Extraction Dataset*. Retrieved May 27, 2019 from <http://www.iesl.cs.umass.edu/data/data-umasscitationfield>
- [3] Sam Anzaroot, Alexandre Passos, David Belanger, and Andrew McCallum. 2014. Learning Soft Linear Constraints with Application to Citation Field Extraction. arXiv:1403.1349
- [4] Lutz Bornmann, K. Brad Wray, and Robin Haunschild. 2019. Citation concept analysis (CCA)—A new form of citation analysis revealing the usefulness of concepts for other researchers illustrated by two exemplary case studies including classic books by Thomas S. Kuhn and Karl R. Popper. arXiv:1905.12410 [cs.DL]
- [5] Kenneth L. Clarkson. 1985. *Algorithms for Closest-Point Problems (Computational Geometry)*. Ph. D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171.
- [6] Torben Hagerup, Kurt Mehlhorn, and J. Ian Munro. 1993. Maintaining Discrete Probability Distributions Optimally. In *Proceedings of the 20th International Colloquium on Automata, Languages and Programming (Lecture Notes in Computer Science, Vol. 700)*. Springer-Verlag, Berlin, 253–264.
- [7] Lars Hörmander. 1985. *The analysis of linear partial differential operators. III*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. viii+525 pages. Pseudodifferential operators.
- [8] Lars Hörmander. 1985. *The analysis of linear partial differential operators. IV*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. vii+352 pages. Fourier integral operators.
- [9] Markus Kirschmer and John Voight. 2010. Algorithmic Enumeration of Ideal Classes for Quaternion Orders. *SIAM J. Comput.* 39, 5 (Jan. 2010), 1714–1747. doi:10.1137/080734467
- [10] Leslie Lamport. 1986. *L<sup>A</sup>T<sub>E</sub>X: A Document Preparation System*. Addison-Wesley, Reading, MA.
- [11] Newton Lee. 2005. Interview with Bill Kinder: January 13, 2005. Video. *Comput. Entertain.* 3, 1, Article 4 (Jan.-March 2005). doi:10.1145/1057270.1057278
- [12] Dave Novak. 2003. Solder man. Video. In *ACM SIGGRAPH 2003 Video Review on Animation theater Program: Part I - Vol. 145 (July 27–27, 2003)*. ACM Press, New York, NY, 4. doi:99.9999/woot07-S422 <http://video.google.com/videoplay?docid=6528042696351994555>

- [13] Barack Obama. 2008. A more perfect union. Video. Retrieved March 21, 2008 from <http://video.google.com/videoplay?docid=6528042696351994555>
- [14] Poker-Edge.Com. 2006. Stats and Analysis. Retrieved June 7, 2006 from <http://www.poker-edge.com/stats.php>
- [15] R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- [16] Bernard Rous. 2008. The Enabling of Digital Libraries. *Digital Libraries* 12, 3, Article 5 (July 2008). To appear.
- [17] Joseph Scientist. 2009. The fountain of youth. Patent No. 12345, Filed July 1st., 2008, Issued Aug. 9th., 2009.
- [18] Harry Thornburg. 2001. *Introduction to Bayesian Statistics*. Retrieved March 2, 2005 from <http://ccrma.stanford.edu/~jos/bayes/bayes.html>
- [19] TUG 2017. *Institutional members of the T<sub>E</sub>X Users Group*. Retrieved May 27, 2017 from <http://www.tug.org/instmem.html>
- [20] Boris Veytsman. 2017. *acmart—Class for typesetting publications of ACM*. Retrieved May 27, 2017 from <http://www.ctan.org/pkg/acmart>

## A Dataset construction

The temperature parameter of all MLLM models used in data construction is 0.5.

### A.1 The process of step1

We initially utilized the Gemini-2.5-Pro model for initial annotations. Each input comprised the original image and its four sub-images, which were obtained by cropping the original image into a 2×2 non-overlapping grid and upsampling them by a factor of two using Lanczos interpolation. These five images were concatenated into a single visual input sequence and fed into the model for annotation. The model was then with prompt:

"Imagine you need to use this image to solve a specific problem or complete a specific task(e.g., geographical positioning, temporal inference, route planning, urban management, disaster detection, future prediction, event tracing, traffic safety, sports analysis, human intention speculation). Note that you should not be limited to the given scenarios; if you have other scenarios more suitable for the given image, you can design them yourself. Please first clarify the specific context or task you set, then carry out detail mining and logical exploration: in the context you set, please carefully examine the image, paying special attention to those unobtrusive details that contain rich information. Your goal is: based on the specific context you set and the special details you find, put forward a key question. This question should: 1.Go beyond surface description: It should not be a simple object recognition ('What is this?'), counting ('How many are there?'), or color question ('hat color is it?'). 2. Be supported by details: It needs to be answered by combining (visually small) detailed visual elements in the image. 3.Inspire in-depth thinking: Answering the question often requires further inference. 4.Be closely tied to the set context: It is directly related to the task you initially envisioned to solve or complete. 5.Focus on the content itself: Avoid discussing image quality, shooting techniques or angles, and avoid diverging towards art, literature, etc. Please first explain the context/task you set, then put forward the question you designed that requires in-depth reasoning combined with visual details to solve. Finally, give the answer and a description of the visual clues supporting the answer and the reasoning. Note that some images have more than two clues, and you should find all clues that can support the answer. The result output format is:

```
{
  "Image": "<image_id>",
```

```
  "Scenario": "<generated context or task>",
  "Question": "<question>",
  "Answer": "<answer>",
  "Visual Clue1": "<description of clue 1>",
  "Reasoning": "<logical inference based on clue 1>",
  ...
}
```

### A.2 The process of step2

At first, we try to use clip model to measure the relevance between image and scenario or question, but the experiment result show that, mostly instances' score is lower than 4. The reason we analyses is the scenario or question generated by MLLM is different with a image caption, and final we choose to use another MLLM to judge.

In this step, the masked image is not set by randomly masking each pixel in the image. Instead, the entire image is divided into 30×30=900 small squares, and 450 of these squares are randomly selected for masking. The detail prompt used for MLLM as follows:

You are an impartial evaluator. Your job is only to judge the relevance between the masked image and situation. Score the relevance from 0 to 10. Only output one number.

### A.3 The process of step3

In this step the cross-verification is performed as follow. At first, we input image with the crossponeding question to other MLLM. Let the Answer1, Answer2 are output in this step, initial answer are from MLLM in step1. If the instance satisfy:

if  $\text{similarity}(\text{answer1}, \text{answer2}) < 0.7$  or  $\text{similarity}(\text{answer1}, \text{answer3}) < 0.7$  or  $\text{similarity}(\text{answer2}, \text{answer3}) < 0.7$ :

this instance will be removed because the question may not has strong relation with image.

## B Dataset

The distribution of bbox annotated for clues is shown as follows:

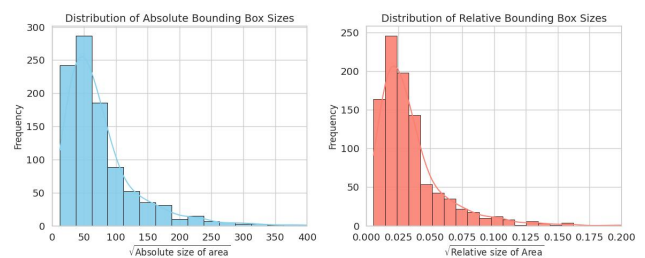


Figure 3: The distribution of annotated bbox

In the left sub image in Fig.3, its show the distribution of absolute size for bbox. More than 50% of the boxes have a scale smaller than 50×50. As can be seen from the right subgraph, boxes corresponding to clues with a relative size (area) of less than 0.05

### B.1 The clue type of six tasks

We have broadly categorized the types of clues for each task in VER-Bench, as listed below:

Geospatial Reasoning Clue Type: Landmarks, regional brands, written language, flags, vehicle license plates, phone number/country codes, boat registration identifiers, official organization insignia.

Temporal Reasoning Clue Types: Dates on posters, clocks, seasonal indicators, specific events, social context, holiday/festival indicators, product date labels

Situational Adaptation Reasoning Clue Types: Activity signage, posters, various functional devices, brand logos, landmark buildings, crowd behavior, temporary structures

Intent Inference Clue Types: Slogans, signs, public announcements, specific equipment/devices, product packaging, advertisements, architectural features, spatial layout

System State Understanding Clue Types: Notice boards, environmental cues, pedestrian/vehicle flow, state of equipment

Symbolic Analysis Clue Types: Symbols, slogans, text, sculptures, artwork, characteristic architecture, posters.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009