Efficiently Crowdsourcing Visual Importance with Punch-Hole Annotation

Minsuk Chang[*1], Soohyun Lee[†2], Aeri Cho[† 2], Hyeon Jeon[† 2], Seokhyeon Park[† 2],
Cindy Xiong Bearfield[* 1], Jinwook Seo[† 2]

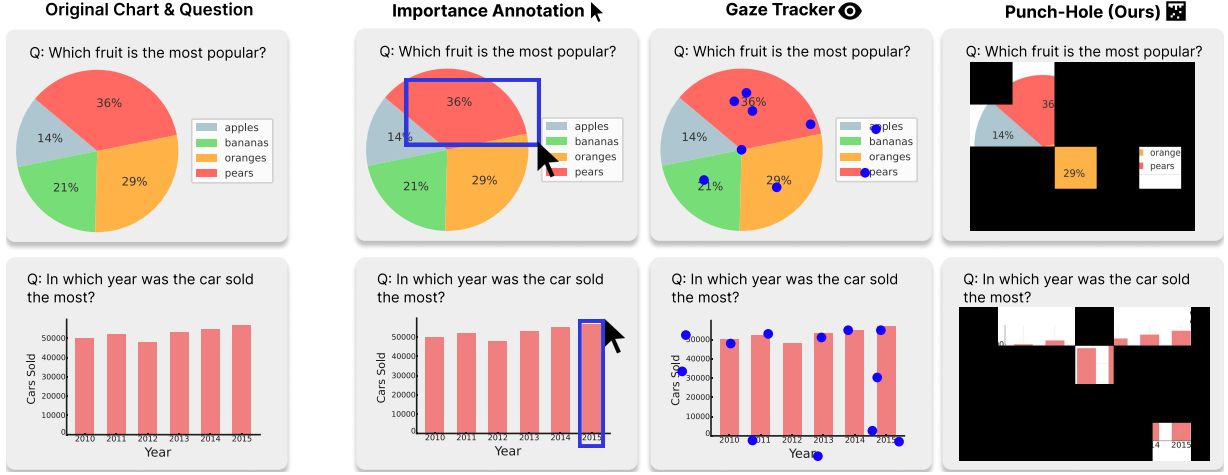[1]Georgia Institute of Technology   [2]Seoul National University

Figure 1: Comparison of three methods used to extract important areas in charts. **Importance Annotation** [4] (left): Drawing a box or polygon with a mouse. **Gaze Tracker** [5] (middle): Capture the points where the eye stays long. **Punch-Hole Labeling** (right): Extract important patches for proper question answering. The example annotation for Importance Annotation and Punch-Hole labeling is retrieved from the pilot study.

ABSTRACT

We introduce a novel crowdsourcing method for identifying important areas in graphical images through punch-hole labeling. Traditional methods, such as gaze trackers and mouse-based annotations, which generate continuous data, can be impractical in crowdsourcing scenarios. They require many participants, and the outcome data can be noisy. In contrast, our method first segments the graphical image with a grid and drops a portion of the patches (punch holes). Then, we iteratively ask the labeler to validate each annotation with holes, narrowing down the annotation only having the most important area. This approach aims to reduce annotation noise in crowdsourcing by standardizing the annotations while enhancing labeling efficiency and reliability. Preliminary findings from fundamental charts demonstrate that punch-hole labeling can effectively pinpoint critical regions. This also highlights its potential for broader application in visualization research, particularly in studying large-scale users' graphical perception. Our future work aims to enhance the algorithm to achieve faster labeling speed and prove its utility through large-scale experiments.

Index Terms: Crowdsourcing, visual annotation, importance labeling, punch-hole annotation, graphical image analysis.

1 INTRODUCTION

Extracting saliency in graphical images has gained importance for analyzing cognitive activities, such as perceptual bias [1] and sub-

---

[*]e-mail: {minsuk, cxiong}@gatech.edu
[†]e-mail: {shlee, archo, hj, shpark, jseo}@hcil.snu.ac.kr

jective importance [3]. Various researchers use crowdsourcing to retrieve and utilize such human-centric data on a large scale [1, 5].

However, crowd workers often annotate images directly (e.g., bounding boxes, polygons) or indirectly (e.g., gaze capture). While intuitive, this method introduces two main problems. First, labelers may overemphasize certain regions, neglecting other important areas. For example, TurkEyes [4] shows that the extracted importance in a resume is concentrated near the title, overshadowing other sections. Second, free-form annotations (e.g., colored pixels, gaze points) are highly variable, requiring data processing or combining responses from multiple participants. ScannerDeeply [5] addressed this by implementing active noise removal in gaze data to counter low accuracy of webcam-based gaze trackers, adding to processing efforts and raising labeling costs.

As a breakthrough, we suggest a novel grid-based annotation to effectively crowdsource the importance or saliency of the graphical image. Punch-hole labeling reduces the continuous annotation task into multiple binary questions, fostering consensus among participants and reducing labeling costs. The process involves dividing the image into small grids and hiding one patch at a time (similar to punching holes) to identify less important areas. If users can answer the given question based on the shown patches, it indirectly indicates that the punched holes lack significance. This approach aims to prevent omitting critical but overlooked regions, minimize unnecessary discrepancies among annotations, and simplify the annotation task for greater time efficiency.

Our preliminary study with two fundamental charts in Figure 1 suggests that Punch-hole labeling may achieve faster labeling speed and produce a more reliable importance map than traditional approaches. Our next goal is to enhance the punching algorithm to increase labeling speed and deploy a large-scale experiment in the crowdsourcing platforms.

2 PUNCH-HOLE LABELING

Punch-Hole annotation is based on the idea that **simplified questions and responses** in a crowdsourcing environment can improve both the speed and reliability of labeling tasks [2]. Compared to the original annotation task of "*Annotate the important area related to this question.*", it is reduced to the binary question of "*Can you answer the question based on these patches?*". Our punch-hole algorithm and its advantages follow this principle.

2.1 Algorithm

We begin by dividing an image into square patches using a pre-defined grid size and displaying only a subset of these patches to crowd workers. This approach involves sequentially hiding each patch, like punching holes in a piece of paper, and asking whether the remaining visible portions of the image are sufficient to answer the given question. In Figure 1, the remaining areas are essential for answering the questions, compared to unnecessary black patches. For instance, the pie chart's top-right corner or the bar chart's middle, which appears black, is unimportant for retrieving the answer.

This process is iterative, with holes being punched until all visible patches are crucial for the given question. Subsequently, the area's resolution can be enhanced by reducing the grid size and repeating the process, theoretically achieving pixel-level granularity. Labeling time and patch size present an inverse tradeoff that deployers can adjust based on their budget or quality requirements.

2.2 Advantages

The potential advantages of punch-hole labeling can be expressed with three key terms:

- **Show-And-Verify** strategy prevents underestimation of specific areas. Punch-hole labeling controls the information presented to users, ensuring no sections are overlooked. Previous methods using active responses like gaze points and polygons [5, 4] allowed too much freedom, leading to incomplete coverage. For example, users might skim the legend and focus solely on the data, neglecting the legend's importance. Our approach includes all crucial patches in the final annotation, enhancing quality.

- **Discrete Responses** reduce the number of required participants. Previous methods like gaze trackers [5] and mouse-based interfaces [4] face issues with merging diverse responses, requiring many participants to suppress noise [4]. Punch-hole annotations, with lower granularity and standardized format, separate controversial or subjective areas from consensus ones. This method achieves the desired granularity after iterations with smaller patch sizes, reducing the impact of outliers and noise.

- **Task Simplicity** increases accessibility. Punch-hole labeling simplifies and standardizes the task, requiring only two buttons (yes/no). Traditional tools require facial position calibration [5] or precise clicking and brushing [4], relying heavily on motor skills. This can limit accessibility for elderly users, significant contributors to crowdsourcing. Poor devices can also produce noisy annotations, requiring iterations for accuracy [5]. The punch-hole algorithm addresses this by focusing on the cognitive task, reducing complexity, and enhancing accessibility.

3 PRELIMINARY STUDY

As a preliminary study, we tested the effectiveness and efficiency of our punch-hole annotation with a pilot test with two labelers. We analyzed whether our approach could effectively find the important area in the image while reaching a faster labeling speed.

3.1 Procedure

We started our study by generating the fundamental graphical image for the experiment. Leveraging ChatGPT, we created one pie chart with fruit sales and one bar chart with car sales. Trailing questions were also generated based on the charts' contents. We then tested our algorithm on a simple web-based labeling tool and compared its results with the box annotation from importance annotation [4]. Users were allowed to annotate more than one box for a fair comparison. The time spent to label each annotation was recorded for later analysis.

3.2 Results

- **Extracted Annotations.** An example of punch-hole annotation is shown in Figure 1. Punch-hole annotations include the core area needed to answer the question correctly. In importance annotation, users' marked areas varied greatly; one marked the whole bar, another only the top. Users also often miss or over-mark legends as important.

- **Time Analysis.** The average labeling speed was 1.32 seconds per annotation candidate and 30.36 seconds per chart. Punch-hole labeling takes a similar time to the 30 seconds reported for importance annotation [4]. By dynamically adjusting the punch-hole shape and refining which patches to hide, we aim to create more precise areas in less time.

3.3 Discussion & Future Work

We have confirmed that punch-hole labeling can extract relative importance from the image within a reasonable time. Our findings also indicate that traditional annotation techniques often miss essential areas and produce highly variable results. Consequently, we aim to advance our punch hole algorithm by: 1) minimizing the number of punch holes to reduce the number of micro-tasks for crowdsourcing workers, and 2) dynamically determining the optimal punch-hole size for each image.

We also observed that the order in which we punch holes may influence the final output. Randomizing the order of punching and analyzing how the final areas are extracted will be another future direction. To achieve this, we plan to experiment with this approach in a crowdsourcing platform to increase the reliability of the results and conduct a large-scale analysis of the extracted areas, ultimately forming a dataset. We expect that our work will offer researchers focused on human cognition and visual perception the opportunity to crowdsource large-scale annotations, even with limited budgets.

REFERENCES

[1] C. X. Bearfield, L. van Weelden, A. Waytz, and S. Franconeri. Same data, diverging perspectives: The power of visualizations to elicit competing interpretations. *IEEE Transactions on Visualization and Computer Graphics*, 30(6):2995–3007, 2024. 1

[2] R. M. Borromeo, T. Laurent, and M. Toyama. The influence of crowd type and task complexity on crowdsourced work quality. In *Proceedings of the 20th International Database Engineering & Applications Symposium*, pp. 70–76, 2016. 2

[3] Z. Bylinskii, N. W. Kim, P. O. Donovan, S. Alsheikh, S. Madan, H. Pfister, F. Durand, B. Russell, and A. Hertzmann. Learning visual importance for graphic designs and data visualizations. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology*, 2017. 1

[4] A. Newman, B. McNamara, C. Fosco, Y. B. Zhang, P. Sukhum, M. Tancik, N. W. Kim, and Z. Bylinskii. Turkeyes: A web-based toolbox for crowdsourcing attention data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. 1, 2

[5] S. Shin, S. Chung, S. Hong, and N. Elmqvist. A scanner deeply: Predicting gaze heatmaps on visualizations using crowdsourced eye movement data. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):396–406, 2022. 1, 2