# Drawing Boundaries: Evaluating a Heuristic Baseline for Panel Detection in Jack Kirby Comics

John A. Walsh[1] [0000−0002−1824−210X] and Faye Jelenski[2]

[1] Indiana University, Bloomington IN 47408, USA jawalsh@iu.edu
https://johnwalsh.name
[2] Indiana University, Bloomington IN 47408, USA fjelensk@iu.edu

**Abstract.** This paper forms part of a larger diachronic study of Jack Kirby's page design, with particular attention to how his use of panel structure and page layout changes across major phases of his comics work. Such analysis requires panel layout to be represented as data, yet large-scale study remains limited by the difficulty of identifying panels reliably across diverse page designs. Rather than proposing a new state-of-the-art panel detector, this paper evaluates an intentionally simple and interpretable computer-vision baseline based on preprocessing, edge detection, contour analysis, and bounding-box filtering. Using a manually annotated set of 100 Kirby pages with 472 gold-standard panels, the pipeline produced 302 predicted boxes: 299 true positives, 3 false positives, and 173 false negatives. Results show very high precision (0.9901) but limited recall (0.6335), with an F1 score of 0.7726. The detector is therefore not sufficient for fully automated layout analysis. Its value lies instead in its use as a diagnostic and analytical tool: by showing which panels can be recovered through simple contour-based methods and which are missed, the baseline helps identify the visual conditions under which Kirby's pages conform to, interrupt, or depart from standard bounded-panel conventions.

**Keywords:** computer vision · comics studies · panel detection · page layout · Jack Kirby · image processing · computational comics analysis

## 1 Introduction

Panel layout is a fundamental organizing feature of comics. The size, shape, placement, and sequencing of panels help structure narrative rhythm, visual emphasis, page-level composition, and the reader's movement through the page. Yet large-scale empirical analysis of panel layout remains difficult because identifying panels across hundreds or thousands of pages usually requires slow, labor-intensive manual annotation.

This difficulty is especially significant for studies that seek to examine change over time or variation across artists. The present study forms part of a larger project on panel usage and page layout in Jack Kirby's Silver and Bronze Age work, with future comparisons to artists such as Jim Steranko, Frank Brunner, Barry Windsor-Smith, and Paul Gulacy, whose comics are often associated with more experimental approaches to page structure. Before such diachronic and comparative questions can be addressed at scale, however, a basic methodological question must be considered: can panel structures be detected with enough reliability to support computational analysis?

This paper does not present a general-purpose solution to comics panel detection. Instead, it establishes and evaluates a classical computer-vision baseline in order to identify which aspects of comics page layout are tractable through contour-based methods and which remain resistant to automation. The method is intentionally simple and interpretable. It uses image preprocessing, edge detection, contour analysis, and bounding-box filtering to identify candidate panels in comic-book pages drawn by Kirby, and its output is evaluated against a manually annotated set of pages.

The method evaluated here should be understood in relation to a larger body of work on comics panel and frame extraction. Prior approaches have used recursive division-line detection, connected-component and blob analysis, contour and morphology-based segmentation, line-segment combination, edge-segment analysis, probabilistic graphical models, and more recent object-detection or segmentation models [18, 16, 13, 10, 21, 20, 17, 1]. These methods address many of the same problems identified in this paper, including overlapping panels, interrupted gutters, missing borders, irregular layouts, and the limitations of treating panels only as rectangles or four-edge polygons. The present study does not claim to improve on that literature algorithmically. Its detector is intentionally closer to earlier heuristic and contour-based approaches than to more advanced line-segment, graphical-model, or learning-based systems. This choice reflects the paper's main purpose: to establish a transparent, reproducible lower-bound baseline for a historically specific Kirby corpus and to use its errors diagnostically as part of a broader computational comics-studies project. A full comparison with existing implementations and public benchmark datasets remains necessary future work before stronger claims can be made about the relative difficulty of Kirby pages or the general performance of the method.

The purpose of this baseline is not only to measure performance, but also to understand failure. Pages with intact rectangular borders may be relatively amenable to contour-based detection, while pages with broken borders, open gutters, inset structures, irregular compositions, caption boxes, or visual elements crossing panel boundaries pose more difficult problems. These failures are not merely technical obstacles; they point to formal features of comics page design that matter for future computational analysis.

The paper asks four questions: how accurately can a heuristic computer-vision pipeline identify panels in a sample of Kirby pages; what kinds of page layouts does the method handle well; what visual or formal features cause the method to fail; and how can this baseline inform future computational comics analysis, including supervised machine-learning approaches? By foregrounding both the strengths and limitations of a transparent baseline method, this study provides an initial empirical assessment of a panel-detection workflow and clarifies the methodological challenges that must be addressed before large-scale analysis of comics layout can support claims about artistic development, historical change, or stylistic difference across creators.

## 2 Why Panel Detection Matters

Panels are among the basic units through which comics organize narrative, direct visual attention, guide sequential reading, and contribute to the mood and aesthetic qualities of the page. Comics theorists have long emphasized the importance of this spatial organization, Eisner's and McCloud's accounts of panels, sequence, timing, and gutters,[5, 11] Peeters's typology of page layout,[15] Groensteen's account of the page's "spatio-topia,"[6] and Cohn's work on comics' "visual language"[3] and readers' navigation of panel order across the page.[2]

Page layout matters not only because it arranges panels in space, but because it contributes to pacing, rhythm, emphasis, and narrative hierarchy. A regular grid can establish a steady visual tempo; a splash page can suspend sequential progression in favor of spectacle; inset panels can compress or subordinate narrative moments; and irregular, borderless, or overlapping panels can complicate the reader's path through the page. Layout is therefore part of the page's formal logic rather than a neutral container for narrative content.

For comics scholars interested in large-scale formal analysis, this creates a practical problem. Questions about page composition require panel structure to be represented as data: the number of panels per page, the relative size of panels, the distribution of grids and non-grid layouts, and the frequency of splash pages, inset panels, or irregular compositions. Pederson and Cohn's diachronic analysis of American superhero comics demonstrates the value of treating page layout as a measurable formal feature, but such work depends on the ability to identify and encode panels consistently across many pages. Their study describes page layouts as "one of the most overt features of comics' structure" and uses a corpus of 40 comics published from 1940 through 2014.[14]

Manual annotation remains the most reliable way to identify panels in difficult cases, especially when pages include broken borders, images that cross gutters, implied boundaries, or visually and structurally experimental layouts. But manual annotation is time-consuming and difficult to scale. Automatic panel detection, or semi-automated detection followed by human correction, offers one possible solution. In computer-vision research on comics and manga, panel detection is often treated as an enabling task for subsequent analysis. Ngo Ho, Burie, and Ogier describe panel extraction as "essential" because panels contain key narrative moments, support the deduction of reading order from layout, and can be further analyzed for indexing and retrieval.[12] Yanagisawa and Watanabe similarly detect panels before assigning speech balloons and character faces to panel-level structures.[22]

The present study takes this gap between reliable manual annotation and scalable computational analysis as its starting point. It treats panel detection as a necessary baseline task for future computational comics analysis, while recognizing that a simple computer-vision pipeline cannot capture all of the formal complexity of comics pages. This tension is especially important for a project focused on Kirby, whose pages often combine legible panel organization with dynamic figures, captions, machinery, collages, and graphic effects that interrupt, obscure, or otherwise complicate panel boundaries.

## 3 Corpus and Annotation

This study evaluates the panel-detection pipeline on a manually annotated set of 100 comic-book pages drawn by Jack Kirby. The evaluation set is drawn from a larger project that uses computational methods to study Kirby's page design across major phases of his career, especially his 1960s Marvel work and his 1970s DC work. Kirby is a useful test case because his work is historically central to American superhero comics and formally dynamic, including regular grids, splash pages, irregular row structures, dense action sequences, and varied relationships between text and image.

The broader project focuses on sustained bodies of Kirby's work rather than isolated examples: for Marvel, especially *Fantastic Four* and *Thor*; for DC, the interconnected Fourth World titles—*Superman's Pal Jimmy Olsen, New Gods, Forever People*, and *Mister Miracle*—as well as *Kamandi*. The present paper does not yet attempt that full diachronic comparison. Instead, it uses a smaller manually annotated evaluation set as a first step toward that larger goal.

The evaluation set contains 100 pages[3] and 472 manually annotated panel boxes. These annotations serve as the gold standard against which the detector's output is evaluated. The set was annotated using the Computer Vision Annotation Tool (CVAT), an open-source image and video annotation platform commonly used to produce ground-truth data for computer-vision tasks.[4] The accompanying repository, *drawing-boundaries-panel-detection*, includes the corpus manifest, annotation guidelines, CVAT-to-CSV conversion workflow, and gold-standard annotation CSV used in the evaluation.[19]

Annotations were created as rectangular bounding boxes around each interpreted panel unit, using a single annotation label, *panel*. This approach reflects both the capabilities of the baseline detector and the goals of the initial evaluation. The method tested here is designed to identify rectangular panel regions, and the 100 pages annotated for this study do not include clear examples of non-rectangular panels. The annotated sample therefore does not test how the detector would perform on panels requiring polygonal or segmentation-based annotation. This limits direct comparison with datasets that include non-

[3] The 100 pages in the evaluation set consist of ten pages each from *Fantastic Four* #6 (September 1962), *Journey into Mystery* #124 (January 1966), *Fantastic Four* #53 (August 1966), *Thor* #171 (December 1969), *Superman's Pal Jimmy Olsen* #141 (September 1971), *New Gods* #5 (October–November 1971), *Forever People* #5 (October–November 1971), *Kamandi* #1 (October–November 1972), *Mister Miracle* #12 (January–February 1973), and *Kamandi* #36 (December 1975).

rectangular, open, or free-shape panel definitions. Together, these annotations provide a consistent basis for evaluating a first-pass rectangular panel detector, while reserving explicitly non-rectangular panel detection for future work.

The annotation policy treats splash pages as single panels when the whole page functions as one image unit. Captions, title areas, logos, and other textual features are not annotated as panels by themselves unless they are clearly part of a larger panel region. When a caption appears within, above, or adjacent to bordered artwork and functions with that artwork as a single narrative or visual unit, it is included within the panel's bounding box rather than marked separately. Ambiguous cases are resolved according to narrative and visual function: the annotation asks whether the region operates as a distinct panel in the page's sequence, not simply whether it is enclosed by a visible border.

This annotation policy is necessarily interpretive. Comics pages often blur the boundaries among panels, captions, image fields, and decorative or paratextual space. A human annotator can often identify a panel through conventions of reading order, visual grouping, and narrative segmentation, even when the border itself is incomplete or interrupted. A detector based on edges and contours, by contrast, is more likely to depend on visible, continuous, high-contrast boundaries. The annotated set is exploratory rather than comprehensive. It supports baseline evaluation and error analysis, not a general account of Kirby's work or of comics page design as a whole. The sample is best understood as a controlled starting point: large enough to reveal recurring patterns in the detector's performance, but limited enough that the results should not be generalized too broadly. Its purpose is to establish what this simple heuristic method can and cannot detect reliably, and to identify layouts that would require more advanced methods in future work.

## 4 Detection Method

This study uses an intentionally simple, interpretable panel detector as a classical computer-vision baseline. The detector is not a trained model and does not learn panel features from annotated examples. Instead, it applies a fixed sequence of image-processing operations to each page image to identify candidate panel regions. This design makes the detector's behavior relatively easy to inspect: when it succeeds, the result can usually be explained in terms of visible borders, gutters, and rectangular page structure; when it fails, the failure can often be traced to broken boundaries, merged artwork, inset text boxes, or other features of comics layout that do not correspond neatly to closed contours.

The pipeline consists of four main stages: image preprocessing, edge detection, contour identification, and heuristic filtering of candidate bounding boxes. For each input page, the script first reads the image and converts it from color to grayscale. Because the goal is to detect structural boundaries rather than analyze color, grayscale conversion reduces the image to intensity values while preserving the visual contrasts most relevant to borders and gutters. The grayscale image is then smoothed with a bilateral filter. This step reduces small variations and noise while preserving relatively sharp edges, which is useful for comics pages where inked panel borders, caption boxes, balloons, and figure outlines may all produce strong visual boundaries.

After preprocessing, the detector applies Canny edge detection to identify high-contrast edges in the page image. The resulting edge map is then slightly thickened through dilation and processed with a morphological closing operation. These operations help connect nearby edge fragments into more continuous shapes. This is important because scanned comic-book pages often contain imperfect lines, textured printing, uneven reproduction, or small breaks in borders. At the same time, this step also illustrates one of the limitations of the method: strengthening edges can help recover panel outlines, but it can also emphasize non-panel structures such as caption boxes, speech balloons, sound effects, or interior artwork.

The script then identifies contours in the processed edge image. A contour is a connected boundary-like shape found in the edge map. Since many contours on a comics page do not correspond to panels, the detector treats each contour only as a candidate. For each contour, the script calculates its area, perimeter, and approximate polygonal shape. Very small contours are discarded, as are contours that are nearly the size of the entire page. The detector also excludes shapes that are too irregular for the intended baseline: candidate contours must approximate to a convex polygon with between four and eight vertices. This means that the method is especially suited to panels that appear as closed rectangular or near-rectangular regions. It is not designed to detect highly irregular, open, implied, or non-convex panel boundaries.

For contours that pass these tests, the detector extracts a rectangular bounding box. Each predicted panel is encoded as four numeric values, $(x_1, y_1, x_2, y_2)$, corresponding to the upper-left and lower-right corners of the rectangle. Additional heuristic filters are then applied to reduce obvious false positives. Boxes below a minimum width or height are discarded. A detected contour corresponding to the page border is removed if it occupies most of the image. Smaller boxes contained inside larger boxes are also removed when they appear likely to be internal elements rather than distinct panels. Finally, a non-maximum suppression step removes near-duplicate boxes that overlap heavily with one another, keeping the larger candidate. The remaining boxes are sorted into approximate reading order by grouping boxes into rows and then ordering them from left to right within each row.

The output of the detector is a set of predicted panel boxes for each page image. The script writes these detections in a CSV format compatible with the manually annotated data exported from CVAT: image filename followed by four numeric values encoding the bounding box, $(x_1, y_1, x_2, y_2)$. This shared format allows predicted boxes to be compared directly with the manually annotated "gold" boxes used for evaluation. In the evaluation stage, predicted boxes are matched against manual annotations using overlap between bounding boxes, allowing the system to report true positives, false positives, false negatives, precision, recall, F1 score, and page-level panel-count accuracy. The accompany-


**Fig. 1.** Successful panel detection on a page from *Fantastic Four #6* (September 1962) with clear borders and closed panel regions. The detector performs best when panel boundaries form strong, continuous contours that can be converted into rectangular bounding boxes.[8]

ing repository, *drawing-boundaries-panel-detection*, includes the detection script, evaluation script, prediction CSV, per-page results, and aggregate metrics used for the results reported here.[19]

This baseline is deliberately limited. It assumes that many panels can be detected as visible, bounded regions in the image. That assumption works best for pages with clear black borders, strong gutters, and relatively regular panel geometry. It works less well when panel boundaries are implied rather than drawn, when figures or captions interrupt the gutter, when artwork bleeds across panel divisions, or when page design integrates panels into a larger visual composition. These limitations are not incidental to the experiment; they are central to its purpose. By testing a transparent contour-based method against manually annotated Kirby pages, the study identifies which aspects of panel layout can be recovered through simple image processing and which aspects are likely to require more sophisticated modeling, richer page understanding, or continued human interpretation.

## 5 Evaluation Metrics and Results

### 5.1 Evaluation Metrics

The panel-detection pipeline was evaluated by comparing its predicted panel boxes with a manually annotated ground-truth set. Each page in the evaluation set was annotated by hand, and the detector's output was then compared

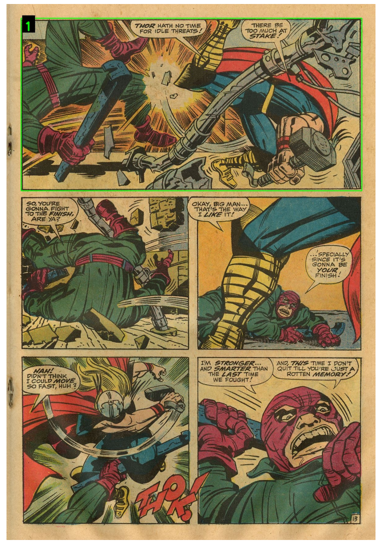


**Fig. 2.** Example failure case involving broken or visually integrated panel boundaries. The contour-based method may miss panels when gutters are interrupted, borders are incomplete, or the page design does not produce closed rectangular regions. In this example from *Thor* #171, the four lower panels are not detected. Note the speech balloon and the visual sound effect ("THOK!") that break the panel borders and traverse the gutters.[9]

against those annotations to determine how many predicted boxes corresponded to actual panels and how many annotated panels were missed. The goal of this evaluation is not only to produce a single accuracy score, but to understand the detector's behavior: whether it tends to over-detect panels, miss panels, or identify some kinds of layouts more successfully than others.

The evaluation uses four main measures: precision, recall, F1 score, and page-level panel-count accuracy. Together, these metrics distinguish between different kinds of success and failure. This distinction is important because a detector can perform well by one measure while still being inadequate for some research purposes. For example, a method might rarely identify false panels, but still miss a large number of real panels. In that case, it would have high precision but limited recall.

**Precision** measures how many of the detector's predicted boxes are correct. In this study, a predicted box is counted as correct when its overlap with a manually annotated panel box meets an intersection-over-union threshold of 0.5. A high precision score means that the detector rarely mistakes non-panel regions for panels.

**Recall** measures how many of the manually annotated panels the detector successfully finds. A lower recall score means that the detector misses many real panels. Recall is especially important for layout analysis because missing panels can distort page-level interpretations. If a page contains six panels but the detector finds only four, the resulting data may misrepresent the density, rhythm, or structure of the page even if all four detected boxes are correct.

**F1 score** combines precision and recall into a single measure. As the harmonic mean of the two values, it rewards methods that perform reasonably well on both measures rather than extremely well on only one. In this study, F1 is useful as a summary statistic, but it should not be treated as a substitute for interpreting precision and recall separately. The gap between those two values is itself one of the most important findings of the experiment.

**Page-level panel-count accuracy** measures the proportion of pages on which the detector predicts the correct number of panels. This measure is stricter than box-level precision because it asks whether the detector captures the overall panel structure of the page. It is especially relevant for comics layout analysis, where many research questions depend on characterizing the composition of whole pages rather than merely identifying isolated panels. A detector that finds several correct panels on a page may still fail to capture the page's layout if it misses one or more panels.

The current evaluation set consists of 100 manually annotated pages containing 472 annotated panel boxes. The detector predicted 302 boxes. Of these, 299 were true positives and 3 were false positives. The detector missed 173 annotated panels. These results produce a precision score of 0.9901, a recall score of 0.6335, an F1 score of 0.7726, and a page-level panel-count accuracy of 0.2700.

**Table 1.** Panel-detection evaluation results

| Metric | Result |
|---|---:|
| Pages evaluated | 100 |
| Gold boxes | 472 |
| Predicted boxes | 302 |
| True positives | 299 |
| False positives | 3 |
| False negatives | 173 |
| Precision | 0.9901 |
| Recall | 0.6335 |
| F1 | 0.7726 |
| Page-count accuracy | 0.2700 |

Table 1 summarizes the main evaluation results, highlighting the contrast between the detector's very high precision and its more limited recall and page-level panel-count accuracy.

### 5.2 Results: High Precision, Limited Recall

The evaluation results show that the detector behaves less like an over-eager classifier than like a conservative extractor. When it identifies a panel, it is usually correct; however, it fails to identify many panels that a human annotator recognizes. This pattern is evident in the sharp gap between precision and recall: the detector produces very few false positives, but it misses a substantial number of annotated panels.

This imbalance is central to interpreting the experiment. The pipeline is very good at avoiding false detections. It rarely mistakes non-panel regions for panels. This is an important strength, especially for an exploratory method based on classical computer-vision operations rather than supervised learning. The detector's output is generally trustworthy when considered box by box: a predicted panel is very likely to correspond to an actual panel. In this respect, the method succeeds as a conservative baseline.

The main weakness is omission. The detector found 299 annotated panels but failed to find 173 others. These missed panels substantially limit the usefulness of the current method for fully automated layout analysis. Many comics studies questions require an accurate account of the entire page: how many panels it contains, how panels are arranged, whether the page uses a regular grid or an irregular structure, and how page density changes across issues, titles, or periods. For these purposes, high precision alone is not enough. A detector that misses more than one third of the annotated panels cannot yet provide a reliable basis for large-scale analysis without further correction or refinement.

The page-level panel-count accuracy reinforces this point. The detector predicted the correct number of panels on only 27 percent of pages. This does not mean that the detector failed completely on the remaining pages; in many cases, it identified several correct panels but missed one or more others. However, from the perspective of layout analysis, a partially correct page can still be analytically misleading. A page-level accuracy of 0.2700 is low, and it indicates that the current detector is not reliable enough for fully automated claims about page density, grid structure, or historical change in Kirby's layouts. The method's value is therefore not as a production detector but as a conservative diagnostic baseline and possible pre-annotation aid.

The results therefore suggest at least two different uses for the current pipeline. First, it can serve as a baseline for future computational work. Because the method is simple and interpretable, its successes and failures can be inspected directly. This makes it useful for identifying which visual features of comics pages are well suited to contour-based detection and which require more sophisticated approaches. Second, the detector may be useful in a semi-automated annotation workflow. Since its false-positive rate is very low, predicted boxes could provide a starting point for human correction: an annotator would primarily need to add missing panels rather than delete many incorrect ones.

The pattern of errors is also analytically useful. The lower recall score points toward layout features that complicate computational treatment of comics pages. Panels with intact rectangular borders are more likely to be detected successfully. By contrast, panels are more likely to be missed when boundaries are broken, gutters are interrupted, or borders are implied rather than drawn. Floating captions, open panels, and visually dense action sequences may also interfere with contour-based detection. These are not merely technical edge cases; they are common features of comics page design and are often central to the aesthetic and narrative effects of the page.

For that reason, the detector's failures should not be understood simply as errors to be eliminated. They also reveal where the formal properties of comics resist straightforward computational segmentation. The pipeline performs best when a panel behaves like a visually discrete rectangular object. It performs less well when panelhood depends on conventions of reading, implied boundaries, or contextual interpretation. This distinction is important for future work because it suggests that panel detection is not only an image-processing problem. It is also a problem of modeling comics-specific visual conventions.

## 6 Error Analysis

The detector's errors are most informative when read in relation to the visual assumptions built into the baseline method. Because the pipeline is based on edge detection, contour identification, and rectangular bounding-box filtering, it performs best when panels appear as visually discrete, closed, high-contrast regions. Its failures occur most often when a panel is legible to a human reader but not reducible to a continuous rectangular contour.

The successful cases are therefore relatively easy to characterize. Pages with strong black panel borders, regular gutters, and conventional grid structures are well suited to the detector's assumptions. This makes the method especially relevant to a study of Kirby and other mainstream American superhero comics of the Silver and Bronze Age, where many pages use clearly bordered panels and grid-based or grid-adjacent layouts. Pederson and Cohn's corpus study of American superhero comics identifies grid-type layouts as a major historical feature of the form, even as such layouts decrease over time and non-grid features become more prominent.[14] The successful example in Fig. 1 illustrates this condition.

Most errors, however, are false negatives. The detector missed 173 of the 472 annotated panels while producing only 3 false positives. Missed panels typically arise when the panel boundary is broken, incomplete, interrupted, or visually integrated into the surrounding composition. Borders may be absent on one or more sides; figures, captions, balloons, or sound effects may cross gutters; or the page may imply a panel division through layout and reading order rather than through a fully enclosed frame. In such cases, the human annotation can rely on narrative segmentation and comics-reading conventions, while the detector depends on the presence of a recoverable contour.

This distinction matters because many missed cases are not merely noisy versions of standard rectangular panels. They are features of comics form. A floating caption may belong narratively and visually to a panel even when no single contour encloses both the caption and the image area, as Fig. 3 illustrates. A figure or sound effect may break the gutter in a way that strengthens the dynamism of the page while weakening the detector's ability to segment the layout. These cases show that the formal expressiveness of comics often depends on departures from the visual conditions that make automated detection easiest. In this sense, a missed panel can sometimes become analytically useful: it marks not only a failure of the detector, but also a moment when Kirby's page design departs from strict bounded-panel conventions and asks readers to infer panel structure through context, composition, or reading order.


**Fig. 3.** Floating caption and incomplete panel boundary from *Kamandi* #1 (October–November 1972). In cases like this, the human annotation treats the caption and bordered artwork as a single panel unit, while a contour-based detector may fail because no continuous border encloses the complete artwork-and-caption region.[7]

False positives are comparatively rare, but they also reveal the method's assumptions. Because the detector searches for closed rectangular or near-rectangular regions, it can mistake caption boxes, interior rectangular image regions, decorative frames, or architectural shapes for panels when they satisfy the same geometric filters. The low number of false positives suggests that the current filtering steps remove many such candidates, but the remaining cases indicate a general problem for panel detection: some non-panel elements share the same visual features that make panels computationally detectable.

For this reason, the error analysis reframes the detector's limitations as part of the paper's contribution. The baseline identifies a meaningful division between panels that can be detected as closed rectangular image regions and panels that require a richer understanding of comics layout, narrative segmentation, and page design.

## 7 Discussion: What the Baseline Shows

This experiment shows that a simple classical computer-vision pipeline can serve as a useful baseline for computational comics analysis, even though it is not sufficient as a finished panel detector. Its value lies not only in its scores but in the fact that its successes and failures can be connected directly to visible page features: strong borders, regular gutters, closed contours, interrupted frames, open panels, and visually integrated compositions.

The most important result is the contrast between high precision and limited recall. The detector is conservative, which makes it potentially useful for semi-automated annotation workflows: it could pre-populate likely panel boxes on pages with clear borders, leaving human annotators to add missed panels and correct ambiguous cases. At the same time, the low recall and low page-level count accuracy show that the method cannot yet support fully automated analysis of page layout. A high-precision detector may still produce misleading layout data if it systematically misses panels with open, broken, implied, or interrupted boundaries.

The results also suggest that panel detection is not merely a technical problem. Some panels are graphic objects with explicit borders, but others are interpretive units produced by page composition, reading order, captions, gutters, and the relation among adjacent image areas. A method that relies only on closed contours will therefore capture only part of what human readers and annotators recognize as panel structure.

## 8 Future Work: From Heuristic Baseline to Supervised Detection

Because this study evaluates only a single heuristic detector, its recall and page-level accuracy cannot yet be separated from the simplicity of the method itself. The results should therefore be read as a lower-bound baseline, to be tested against existing panel-detection systems, public datasets, and supervised object-detection approaches.

Future work will export the CVAT annotations in standard formats such as COCO or YOLO so that established detection models can be trained or fine-tuned on the annotated Kirby pages. The goal is not to replace interpretive annotation, but to measure whether supervised methods improve recall on layouts where the heuristic baseline fails: broken borders, interrupted gutters, open or implied boundaries, floating captions, and panel divisions signaled by composition rather than drawn frames.

Further experiments should compare annotation schemes—including bounding boxes, polygons, segmentation masks, reading-order markup, and separate labels for captions, balloons, gutters, and other page elements—to determine which forms of markup best support the study of panel usage and page layout across Kirby's work and in comparison with other artists.

## 9 Conclusion

This paper does not present a completed general-purpose detector for comics panels. Instead, it presents a classical computer-vision baseline, evaluates it against 100 manually annotated Kirby pages, and uses its successes and failures to clarify the computational problem while beginning to interrogate Kirby's adherence to, and departures from, conventional panel and page-layout structures. The detector produced very high precision but limited recall, showing that it rarely identifies non-panels as panels but frequently misses panels that human annotators recognize. This pattern makes the method useful as a conservative baseline, but not sufficient for fully automated page-layout analysis.

The results show that clearly bounded rectangular panels can often be detected with a high degree of reliability. At the same time, the missed panels show that comics page layouts frequently depend on ambiguous, interrupted, or implied boundaries that resist simple contour-based detection. Floating captions, open borders, interrupted gutters, figures or sound effects crossing panel divisions, and compositions that imply panel divisions without fully enclosing them all complicate the assumption that a panel can be identified as a closed rectangular region. These are not merely technical anomalies; they are part of the formal expressiveness of comics pages and therefore part of what computational methods must learn to represent.

For that reason, the baseline's limitations should be understood as analytically useful rather than simply as poor performance. By showing where a transparent heuristic detector succeeds and where it breaks down, the study identifies the visual and interpretive conditions that existing and future methods must address on this corpus. It also establishes a measurable point of comparison for supervised object-detection models, hybrid annotation workflows, and more expressive forms of panel and page markup. The larger argument is not that panel detection has been solved, but that Kirby's layouts can be made more tractable for computational study through careful baselining, explicit annotation policy, and error analysis.

*Data and code availability.* Code and derived evaluation materials for this study are available in the repository *drawing-boundaries-panel-detection*. The repository includes the detection script, evaluation script, CVAT-to-CSV conversion stylesheet, corpus manifest, gold-standard annotation CSV, prediction CSV, per-page results, aggregate metrics, annotation guidelines, and data-format documentation.[19] Copyrighted comic-book page images and image overlays derived from those pages are not redistributed. Readers with lawful access to the relevant images may rerun the detector using the filenames listed in the corpus manifest.

## References

1. Arai, K., Herman, T.: Method for automatic e-comic scene frame extraction for reading comic on mobile devices. In: 2010 Seventh International Conference on Information Technology: New Generations. pp. 370–375 (2010). https://doi.org/10.1109/ITNG.2010.22
2. Cohn, N.: Navigating comics: An empirical and theoretical approach to strategies of reading comic page layouts. Frontiers in Psychology **4**, 186 (2013). https://doi.org/10.3389/fpsyg.2013.00186

3. Cohn, N.: The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images. Bloomsbury Academic, London and New York (2013)
4. CVAT.ai Corporation: Computer vision annotation tool (CVAT). https://github.com/cvat-ai/cvat (2026), version 2.62.1. Software used for manual image annotation; installed locally via Docker.
5. Eisner, W.: Comics & sequential art. Poorhouse Press Distributed by Eclipse Books, Tamarac, Fla. Guerneville, Calif. (1985)
6. Groensteen, T.: The System of Comics. U of Mississippi P, Jackson (2007)
7. Kirby, J.: Kamandi, the Last Boy on Earth #1. DC Comics (October–November 1972), comic book
8. Lee, S., Kirby, J.: Fantastic Four #6. Marvel Comics (September 1962), comic book
9. Lee, S., Kirby, J.: Thor #171. Marvel Comics (December 1969), comic book
10. Li, L., Wang, Y., Suen, C.Y., Tang, Z., Liu, D.: A tree conditional random field model for panel detection in comic images. Pattern Recognition **48**(7), 2129–2140 (Jul 2015). https://doi.org/10.1016/j.patcog.2015.01.011
11. McCloud, S.: Understanding comics : The invisible art. Tundra Pub, Northampton, MA (1993)
12. Ngo Ho, A.K., Burie, J.C., Ogier, J.M.: Panel and speech balloon extraction from comic books. In: Proceedings of the 10th IAPR International Workshop on Document Analysis Systems. pp. 424–428. IEEE (2012)
13. Nguyen Nhu, V., Rigaud, C., Burie, J.C.: What do we expect from comic panel extraction? In: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). vol. 1, pp. 44–49 (2019). https://doi.org/10.1109/ICDARW.2019.00013
14. Pederson, K., Cohn, N.: The changing pages of comics: Page layouts across eight decades of american superhero comics. Studies in Comics **7**(1), 7–28 (2016)
15. Peeters, B.: Four conceptions of the page. ImageText **3**(3) (2021)
16. Sharma, R., Kukreja, V.: Image segmentation, classification and recognition methods for comics: A decade systematic literature review. Engineering Applications of Artificial Intelligence **131**, 107715 (May 2024). https://doi.org/10.1016/j.engappai.2023.107715
17. Stommel, M., Merhej, L.I., Müller, M.G.: Segmentation-Free Detection of Comic Panels, p. 633–640. Springer Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-33564-8_76
18. Vivoli, E., Souibgui, M.A., Barsky, A., LLabrés, A., Bertini, M., Karatzas, D.: One missing piece in vision and language: A survey on comics understanding (2025), https://arxiv.org/abs/2409.09502
19. Walsh, J.A., Jelenski, F.: drawing-boundaries-panel-detection. https://github.com/jawalsh/drawing-boundaries-panel-detection (2026), code and derived evaluation materials
20. Wang, Y., Zhou, Y., Liu, D., Tang, Z.: Comic storyboard extraction via edge segment analysis. Multimed Tools Appl **75**(5), 2637–2654 (May 2015). https://doi.org/10.1007/s11042-015-2680-8
21. Wang, Y., Zhou, Y., Tang, Z.: Comic frame extraction via line segments combination. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR). pp. 856–860. IEEE (2015)
22. Yanagisawa, H., Watanabe, H.: Recognition of panel structure in comic images using faster r-cnn. In: Proceedings of the 4th IIAE International Conference on Intelligent Systems and Image Processing. Da Nang, Vietnam (2017)