In this supplementary material, we show more analysis of our proposed LecSD dataset, annotation details, synonym based paraphrase sentence generation.

1. Dataset analysis

To analyze the distribution of topics in the dataset, we extracted trigram keywords from the text data of each slide image using KeyBERT [3] and obtained 252K keywords from the dataset. Figure 1 shows the sunburst of common keywords, and the data structure topics such as ‘tree’, ‘list’, ‘heap’, ‘queue’, and ‘stack’ are fairly distributed in the dataset. The inner-circle keywords occur at least 10K times, and the outer-circle keywords occur more than 50 times in the entire dataset.

To analyze the collected slide images, we use two approaches such as a) an Optical Character Recognizer (OCR) with text properties and b) layout-based image segmentation. In the first approach, we utilize off-the-shelf Google lens OCR engine¹ to extract the text information from the slide images. Further, we categorize the text in the slide based on font size. Figure 4(a) shows the distribution of words with various font sizes in the dataset. The font size is the height of a word, and the number of pixels represents the height. As shown in the figure, we divide the histogram into four regions as follows– S: small, M: medium, H: high, and L: large font size with a height of (0, 15], (15, 30], (30, 45], and > 45 pixels respectively and create the word cloud images using the words in each region. Word clouds provide a simple and effective means to communicate the most frequent words of the documents visually. We further analyzed our dataset – LecSD using image properties such as the number of color channels in the image. In the analysis, first, we sorted out the slide images based on its RGB channels and obtained only 6261 grayscale images out of 54K slide images. Generally, grayscale slide images have a less complex layout design than color slide images. Next, we studied the layout components and types of figures used in the slide images. Figure 2 shows that the ‘Enumeration’ region has a higher frequency, and the ‘Footnote’ has the lowest frequency. Figure 3 shows the most frequently occurring document figure is the ‘flowchart’.

Figure 4(b) shows the word cloud of the four regions. The word cloud of regions M, H, and L have common words, and the word cloud image of the S region has different words related to the copyright and publication details. Figure 4(c) shows the typical example of four regions of text appearing in a slide image. The L, H, and M regions are ‘title’, ‘section’, and ‘paragraph/list’ items. Hence, it shares common words on the topic of Data structures. However, the small font size text belongs to represent ‘affiliation’, ‘date’, ‘slide number’, ‘press titles’, ‘footnote’ and ‘website’. Hence, the words in the S region’s word cloud differ from other regions.

We extract the text information from the slide and combine the text from similar logical regions and plot its word cloud [4] as shown in Figure 5. By analyzing Figure 5,
the layout segmentation has higher accuracy in segmenting most logical regions except ‘Date’, and ‘Slide number’. The word cloud of ‘Affiliation’, ‘Pres title’, ‘Website’, and ‘Footnote’ have high similarities, indicating a high confusion in segmenting these regions. We also plot the word cloud of all texts in the slide, and it has a high similarity to the word cloud of the ‘Enumeration’ text.

2. Slide Image summary

We used a modified version of VIA annotation software [2] and a screenshot of the web-based annotation tool is shown in Figure 10. Table 2 and 3 shows sample slide images and its manual summary, ChatGPT [1] generated paraphrase, Automatic summary, Synonym based paraphrased sentences.

In the synonym based paraphrase generation, first we identifies the slide topic \( T \) and semantic regions \( C \). We replace each semantic region with its corresponding synonyms. Some of the sample synonyms are shown in Table 1. We also randomly change structure of the sentence. The half of the paraphrased sentence with structure \( T \) explain using \( C \) and half with \( C \) is used to explain \( T \).

3. Figure bounding box annotation

We manually annotate figure bounding boxes in the slide image. A screenshot of the web-based annotation tool is shown in Figure 11.

4. Figure sketches drawing

Our annotation team drew the annotated figures on a specially designed A4 paper. A screenshot of the web-based annotation tool is shown in Figure 7.

References

Figure 3. Distribution of figure logical labels in LecSD dataset.

Figure 4. Visualize various font size text in the slide images (a) histogram of words with various font sizes and divide the histogram into four regions, S: small, M: medium, H: high, and L: large font size. (b) The word cloud images of words from S, M, H, and L regions. (c) A typical example of text with various font sizes in a slide image (best view in color)

<table>
<thead>
<tr>
<th>block diagram</th>
<th>Diagram</th>
<th>enumeration</th>
<th>equation</th>
<th>Line graph</th>
<th>paragraph</th>
<th>explained using</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chart of relationships</td>
<td>Sketch</td>
<td>List items</td>
<td>Formula</td>
<td>Line chart</td>
<td>Block of text</td>
<td>Depicts using</td>
</tr>
<tr>
<td>Conceptual diagram</td>
<td>Artwork</td>
<td>Itemized points</td>
<td>Algebraic expression</td>
<td>Trendline</td>
<td>Section</td>
<td>elucidated by</td>
</tr>
<tr>
<td>Schematic diagram</td>
<td>Illustration</td>
<td>Key details</td>
<td>Numeric expression</td>
<td>Graph</td>
<td>text Segment</td>
<td>Used for describing</td>
</tr>
<tr>
<td>Structural diagram</td>
<td>Rendering</td>
<td>bullet points</td>
<td>Mathematical formula</td>
<td>Chart</td>
<td>Passage</td>
<td>Employed to elucidate</td>
</tr>
</tbody>
</table>

Table 1. A sample synonyms of commonly occurring words such as 'block diagram', 'sketch', 'enumeration', 'line graph', 'paragraph', and 'used to explain' in the summary of slide image
Figure 5. The word cloud image of the text extracts from the various logical regions in the slide dataset. The odd rows indicate the logical region, and the figure below is its word cloud image.

Figure 6. Figure bounding box annotation using a web-based annotation tool.
Manual Summary: A block diagram depicts the recursive calls to the sum methods.

ChatGpt: A block diagram illustrates the repetitive invocations of the sum function.

Manual Summary: Recursive calls to the sum method explains with a block diagram

Synonym based paraphrase: Abstract code Used for describing Recursive calls to the sum method.

Manual Summary: The definition of program explains using enumeration.

ChatGpt: The program’s definition elucidates the utilization of enumeration.

Automatic Summary: What is Program explains using Enumeration

Synonym based paraphrase: Itemized points Employed to elucidate the What is Program
Three algorithms for computing $1 + 2 + \ldots + n$ for an integer $n > 0$

Selection Sort: C++ Implementation

```cpp
void selectionSort(DataType a[], int n )
{
    for( int last = n - 1; last > 1; last-- )
    {
        // set imax to the index of the largest element in a[0..last]
        int imax = 0;
        for( int i = 1; i <= last; i++ )
            if( a[i] > a[imax] ) imax = i;
        // swap a[imax] with a[last]
        DataType temp = a[last];
        a[last] = a[imax];
        a[imax] = temp;
        // invariant: a[0..a[n-1]] are the largest elements in sorted order
    }
}
```

Measuring Algorithm Efficiency explains using a table and Paragraph

Selection Sort: C++ Implementation explains using a Pseudocode

Synonym based paraphrase: Measuring Algorithm Efficiency Describes with tables, and text Segment.

Table 3. Showing sample slide images and its manual summary, ChatGPT generated paraphrase, Automatic summary, Synonym based paraphrased sentences
Figure 7. The web-based annotation tool shows the figure and its cell id to sketch the image.

Figure 8. The left side shows the A4 paper canvas with 15 table cells with unique cell id. The right side shows the A4 paper canvas after drawing the figure image.
Figure 9. Sample slide image figures and their corresponding drawn sketch of our dataset. The figure columns show the figure annotated from the slide images and the right-side Sketch columns shows its hand-drawn sketches.
Search algorithm #2: Binary Search

- Basic idea: On each step, look at the middle element of the remaining list to eliminate half of it.

Figure 10. The slide image summary writing using a web-based annotation tool.
Search algorithm #2: Binary Search

- Basic idea: On each step, look at the middle element of the remaining list to eliminate half of it.

Figure 11. Figure bounding box annotation using a web-based annotation tool.