

PosterGen: Aesthetic-Aware Multi-Modal Paper-to-Poster Generation via Multi-Agent LLMs

Supplementary Material

Table of Contents

A Extended Related Work	1
B Implementation Details	1
B.1. Agent Specifications	1
B.2. Experimental Environment	1
B.3. Configuration Parameters	2
C Prompts	2
C.1. Baseline Prompt	2
C.2. PosterGen Prompts	2
C.3. VLM-as-Judge Evaluation Prompt Template	2
D Human Evaluation Details	2
E Additional Qualitative Results	2
F. Limitations	4
G Broader Impacts	4

A. Extended Related Work

Slide Generation. A similar task to poster generation is the automatic generation of presentation slides from documents [11, 17, 19, 20, 27, 39, 41, 63, 67]. Some works develop agents for general purposes, such as efficient slide editing [19] or narrated presentation videos [39]. Other methods [67] focus on holistically improving the content, design, and coherence of the slides. Among these, several works specifically aim to generate slides for academic presentations. Early approaches like PPSGen [17], SlidesGen [41], and SlideSpawn [20] utilized summarization and information extraction techniques to generate draft slides. More recent approaches utilize end-to-end systems [11] or persona-aware models [27] to generate more tailored slides. However, academic poster design is far more challenging than slide generation, as slides can distribute content across multiple pages, and work together with a presenter’s oral explanation to convey the full message of the paper, while an academic poster needs to contain all necessary information from a paper onto a single page, and ought to be visually appealing to attract the attention and help initiate a dialogue [10] with the authors.

Table 7. Agent specifications in PosterGen.

Agent/Module	Type	Model	Role
Parser	Hybrid	GPT 4.1	Structure section content
↳Marker	Tool	N/A	Extract content from PDF
Curator	LLM	GPT 4.1	Create spatial content plan and storyboard
Color	LLM	GPT 4.1	Generate color palette
Layout	Hybrid	N/A	Control precise element coords (CSS-style)
↳Balancer	LLM	GPT 4.1	Optimize column utilization
Font	LLM	GPT 4.1	Apply typography and keyword highlighting
Renderer	Proced.	N/A	Generate final PPTX and PNG files

Table 8. Key hyperparameters for PosterGen Implementations. The default LLM/VLM model is `gpt-4.1-2025-04-14`. The “Content Constraints” are controlled via instructions within the agent prompts. The symbol ‘#’ denotes “number of.”

Category	Parameter	Value
LLM/VLM Configuration	Model	GPT-4.1
	↳ Alternatives	GPT-4o
		GPT-4.1 mini Claude-Sonnet-4
	Temperature	0.7
Content Constraints	# Sections	[5, 8]
	# Visual Assets	[4, 6]
	Max Words per Section	1000
Text Height Estimation	Height Precision (ϵ)	0.001 inches
	Newline Offset Ratio	1.0

B. Implementation Details

B.1. Agent Specifications

Table 7 details the type and role of each agent in our pipeline. Most agents (CuratorAgent, ColorAgent, FontAgent) operate via a single LLM call, while ParserAgent and LayoutAgent adopt a hybrid design that runs mostly procedural logic and only calls the LLM for specific subtasks. Renderer is entirely procedural and requires no LLM call.

B.2. Experimental Environment

The framework is compatible with Windows, Linux and macOS operating systems, and is implemented in Python 3.11. While a GPU is not strictly required, it is strongly recommended to accelerate document parsing and optical character recognition (OCR) tasks handled by the `marker-pdf` tool. The framework relies on several key libraries; the most critical dependencies for reproducing our work include:

- `python-pptx` = 1.0.2 for PPT generation.
- `langchain` = 0.3.25 and `langgraph` = 0.4.8 for building the multi-agent workflow.

- `marker-pdf` = 1.8.2 for PDF-to-Markdown conversion and parsing. Notably, we slightly modify the official `marker-pdf` source code to extract tables as images rather than Markdown text.
- `Pillow` = 10.4.0 for image manipulation.

B.3. Configuration Parameters

To ensure the reproducibility of our implementation and experiments, we list the most critical parameters that directly influence the behavior of the agents and the quality of generated posters in Table 8. Aesthetic parameters related to layout, color and typography (font sizes, margins, color values) are defined in the configuration files within our source code but omitted here for clarity.

C. Prompts

C.1. Baseline Prompt

We present the prompt of GPT-4o Image Generation via ChatGPT web interface (as illustrated in Figure 8), which is alongside the input paper file.

C.2. PosterGen Prompts

We present the detailed prompt design [64] used in our PosterGen multi-agent workflow as follows.

ParserAgent. This includes: (1) Title and Authors Extraction (Figure 9); (2) Narrative (ABT) Extraction (Figure 10); (3) Visual Asset Classification (Figure 11); and (4) Structured Section Extraction (Figure 12).

CuratorAgent. CuratorAgent generates an effective storyboard through strategic content planning and by applying visual height constraints. Due to space limitations, we split the prompt into three parts: (1) Input and Design Patterns (Figure 13), (2) visual asset selection and content organization (Figure 14), and (3) output format (Figure 15). We also omit less important parts and replace them with ellipsis mark (...). The full prompt is available in the source code.

Layout Balancer. This is a sub-agent of LayoutAgent designed to improve column utilization and prevent overflows. Its prompt is detailed in Figure 16 and Figure 17.

ColorAgent. We present only the prompt for extracting the theme color from an affiliation logo using a VLM (as shown in Figure 18); the fallback method, which uses a key visual asset, is omitted for clarity.

FontAgent. FontAgent calls LLM once to extract and classify different keywords. The detailed prompt is illustrated in Figure 19.

C.3. VLM-as-Judge Evaluation Prompt Template

We present the prompt template used for our VLM-as-Judge evaluation, as shown in Figure 20. This standardized template is applied to every evaluation focus area and uses a 5-point scale. To counteract the tendency of Vision-Language

Models (VLMs) to provide overly generous scores, our prompt design incorporates a targeted few-shot example strategy. For high scores (4 and 5), we provide positive examples of desired qualities, while for low-to-mid scores (1, 2, and 3), we provide negative examples of common flaws. This approach is designed to calibrate the VLM’s judgment and yield more accurate, evidence-based scoring.

D. Human Evaluation Details

We invite 30 participants to take part in our user study using the Qualtrics platform. To gain objective evaluation results, we survey each user for information regarding their field of study and experience in making academic posters, as shown in Figure 22. This verifies that our group of human evaluators possesses adequate research background to evaluate academic posters properly.

To alleviate evaluator fatigue and cognitive load, we did not ask each participant to assess all of our benchmark. Instead, each participant only needs to evaluate aesthetic scores for a random subset of 10 posters. Each set consists of three variations of posters: generated by our approach PosterGen, the baseline PosterAgent [32], and designed by authors of the paper. A rigorous randomization procedure is adopted to reduce possible bias: (i) each of the 10 sets is randomly ordered for evaluation, and (ii) for each set, the order of the three posters is also randomized. This makes sure that the evaluation is completely blind to all of our approaches for its validity to be maximized. An example set of questions is shown in Figure 21.

In addition, we did not include the GPT-4o-Image approach in our human study. As discussed in the qualitative analysis in Section 5.5 of the main paper, diffusion-based generation via GPT-4o-Image constantly suffers from gibberish text, distorted text blocks, and the hallucination of visual assets. It is thus pointless to consider it for comparison with the multi-agent approach and human designs.

E. Additional Qualitative Results

In this section, we provide 15 additional representative qualitative results drawn from our 30-paper benchmark, as shown in Figures 24 to 38. All the 30 papers were selected from top-tier AI conferences, e.g., NeurIPS, ICML, ICLR, CVPR and ECCV in the last four years, with the distribution shown in Figure 23. Our selection criteria require that both the full paper and a corresponding high-quality human-made poster are publicly available.

As illustrated in Figures 24 to 38, these supplementary qualitative results further validate the observations presented within the main paper, regarding the limitations of baseline approaches for aesthetic design. The end-to-end GPT-4o-Image method is prone to basic layout constraints, manifesting serious boundary problems like too

GPT-4o Image Generation

You are an expert specializing in designing automated academic poster.

Primary Task: Analyze the provided research paper and autonomously design and generate a complete, professional academic poster.

Key Guidelines:

- **Fixed Dimensions:** The final poster layout must be exactly `width` pixels wide and `height` pixels high.
- **Content Fidelity:** All content (text, figures, tables) must be extracted or summarized exclusively from the source paper. Do not invent or infer information.
- **Visual Design:** The poster must be visually appealing. Apply a clean and professional theme consistently across all elements.
- **Layout Design:** The layout must be well-balanced, effectively utilizing the available space. Actively avoid element overflow and large, underutilized blank areas.

Required Structure & Content:

- **Header:** Must include the full paper title and a complete list of all authors.
- **Body:** The main content area must be organized into distinct sections. These sections should feature a well-balanced and carefully arranged composition of:
 - Concise text summaries.
 - high-quality figures from paper pdf.
 - Key data tables from paper pdf.

Output Format: The final output must be a single PNG image file with dimensions of `width` × `height` pixels.

Figure 8. Prompt for GPT-4o Image Generation.

ParserAgent (1): Title and Authors Extraction

You are an expert academic paper parser. Your task is to extract the title and authors from the provided academic paper text.

Please extract:

1. **Title:** The main title of the paper
2. **Authors:** All author names using initials (no affiliations, emails, or other metadata)

Strict Formatting Requirements:

- **Title:** Use proper title case where each word has only the first letter capitalized, EXCEPT for established acronyms, technical terms, or proper nouns that are conventionally written in all uppercase letters (such as abbreviations for organizations, technologies, or methodologies). Keep such terms in their original case. Example: “A Study of Machine Learning Methods” but preserve acronyms like “Using LLM for Data Analysis” or “CNN Architecture Design”.
- **Authors:** Use initials for authors’ names. Convert full names to initials format, preserving middle initials when present. Examples: “Kevin W. Jones” → ”K.W. Jones”, “Yann LeCun” → ”Y. LeCun”, ”Mary Smith Johnson” → ”M.S. Johnson”. Separate multiple authors with “, ”. Remove all affiliations, emails, institutions, departments, addresses, and other metadata.

Input text: `{{ markdown document }}`

Required JSON structure:

```
1 {
2   "title": "Title With Proper Case Formatting",
3   "authors": "F. Author, S. Author, T. Author"
4 }
```

Figure 9. Prompt for ParserAgent to extract the title and authors from the source paper.

much whitespace and cut-off vertical content, thus verifying that it is unable to have robust canvas control. Though PosterAgent [32] enhances content fidelity over diffusion-

based methods, it consistently shows aesthetic inadequacies like improper alignment of elements, poor whitespace use, and a monotonous visual aesthetic that lacks typographi-

ParserAgent (2): Narrative (ABT) Extraction

Extract ABT narrative structure optimized for poster presentation:

Input: {{ Academic paper markdown text }}

Output: JSON with poster-ready ABT structure

Guidelines:

- Each section (and/but/therefore) should be 1-2 concise sentences
- Focus on visual impact and poster audience understanding
- Emphasize key contributions and results
- Avoid technical jargon where possible

Required JSON structure:

```
1 {
2   "and": "Current knowledge and established facts (background context)",
3   "but": "Specific problem, gap, or challenge identified",
4   "therefore": "Your solution, contribution, and key findings",
5   "poster_hook": "One compelling sentence that grabs attention",
6   "key_impact": "Why this research matters (practical implications)"
7 }
```

Paper content: markdowndocument

Figure 10. Prompt for ParserAgent to extract ABT-structured narratives.

cal hierarchy to create visual flow. In contrast, PosterGen consistently generates high-quality artistic posters showcasing pleasant colors, effective visual hierarchy generation, and natural reading paths. Therefore, it demonstrates that PosterGen is capable of including effective aesthetic concepts within its agent workflow.

F. Limitations

Although **PosterGen** demonstrates significant improvement over existing methods for generating aesthetic-aware posters, we identify several limitations that present directions for future work:

- Although the marker-pdf tool is very effective for most use cases for ParserAgent, occasionally it yields errors for extractions. Such artifacts may include lack of text information (for example, title or author names) or possibly for visual components at low resolutions. Areas for improvement may include development of a specialized document parser module designed for academic-style pdf files containing intricate formats.
- CuratorAgent is now limited to using only the visual elements it finds in the source paper. It is not capable of creating generative visual elements based on paper content only. One promising direction to take this agent further is to empower its functionality to generate additional diagrams, such as creating a flowchart to intuitively summarize a text-heavy Introduction.
- PosterGen is dependent on a three-column grid layout as its current paradigm. This acts as a solid and safe choice for landscape-oriented posters, but does come with its limitations in design flexibility. Going further, the ex-

ploration of LayoutAgent should include diverse ratios for various portrait-oriented positions, as well as dynamic and diverse layout designs.

G. Broader Impacts

In this paper, we are the first to explore the integration of design and aesthetic principles into a multi-agent framework for academic poster generation, and propose a novel workflow **PosterGen**. By mirroring the specialized workflow of professional designers, **PosterGen** achieves content fidelity that rivals the state-of-the-art while significantly enhancing the final poster’s visual design and aesthetic quality. Though it may not fully eliminate the need for human fine-tuning, its core contribution lies in systematically embedding design principles into multi-agent design, a step often overlooked even in manual creation.

We hope our work will advance this emerging yet meaningful field of automated scientific communication, and firmly believe that PosterGen can substantially relieve researchers of the time and effort required for poster creation, allowing them to focus more on the one-on-one scholarly dialogue that poster sessions are meant to facilitate.

ParserAgent (3): Classify Visual Assets

Classify visual assets by column-aware poster placement and research role:

Available visuals with captions: visuals list

Classify each visual into exactly one category based on column-specific poster design:

Column-Aware Categories:

1. **key visual:** Most important method visual representing core research innovation (max 1, middle column)
2. **problem illustration:** Visuals showing research problem, challenges, or motivation (left column introduction)
3. **method workflow:** Method architecture, system diagrams, algorithmic workflows (middle column method)
4. **main results:** Primary experimental results, performance tables, key findings (right column)
5. **comparative results:** Baseline comparisons, ablation studies, validation charts (right column)
6. **supporting:** Background concepts, supplementary analysis, minor details (flexible placement)

Classification Guidelines:

- Problem Context: Figures showing “what’s wrong” or “why this matters” → problem illustration
- Method Core: Most important technical diagram → key visual
- Method Details: Architecture/workflow diagrams → method workflow
- Primary Evidence: Main performance results → main results
- Validation Evidence: Comparisons with baselines → comparative results
- Background/Supplementary: Minor or supporting content → supporting

Consider:

- Visual content and research narrative role
- Optimal column placement for logical flow
- Visual impact and audience comprehension

Required JSON output:

```
1 {
2   "key_visual": "visual_id or null",
3   "problem_illustration": ["visual_id1", ...],
4   "method_workflow": ["visual_id1", ...],
5   "main_results": ["visual_id1", ...],
6   "comparative_results": ["visual_id1", ...],
7   "supporting": ["visual_id1", ...]
8 }
```

Ensure every visual id appears exactly once across all categories.

Figure 11. Prompt for ParserAgent to classify extracted visual assets.

ParserAgent (4): Structured Section Extraction

Extract structured sections from academic paper text for poster creation.

Paper Text: {{ raw text }}

Task: Extract all major sections from the paper and organize them with their content. Focus on sections that would be relevant for an academic poster.

Section Extraction Guidelines:

1. Identify Major Sections:

- Introduction/Background
- Related Work (if substantial)
- Methodology/Approach
- Experiments/Results
- Discussion/Analysis

2. Content Processing:

- Extract the main content for each section
- Keep section content under **1000** words
- Preserve key technical details, formulas, and findings
- Maintain important bullet points and lists
- Remove excessive citations and references

3. Section Classification:

- **foundation:** Introduction, background, motivation, problem statement
- **method:** Methodology, approach, algorithm, system design
- **evaluation:** Experiments, results, analysis, validation

Required JSON structure:

```
1 {
2   "paper_sections": [
3     {
4       "section_name": "Introduction",
5       "section_type": "foundation",
6       "content": "Main content of the section (max 1000 words)",
7       "key_points": [ ... ],
8       "importance": "high|medium|low",
9       "contains_figures": ["figure_1", "figure_2"],
10      "contains_tables": ["table_1"]
11    }
12  ],
13  "paper_structure": {
14    "total_sections": 5,
15    "foundation_sections": 2,
16    "method_sections": 2,
17    "evaluation_sections": 1
18  }
19 }
```

Critical Requirements:

- Extract ALL major sections (don't skip any)
- Keep each section under 1000 words
- Preserve technical accuracy
- Identify which figures/tables belong to each section
- Classify section importance for poster layout

Generate structured sections that provide comprehensive paper coverage for poster creation.

Figure 12. Prompt for ParserAgent to extract structured sections.

CuratorAgent: Spatial Story Board Generation (Part 1 of 3)

You are an Expert Academic Poster Designer specializing in visual-dense poster layouts with strategic spatial organization.

Mission: Transform research papers into spatially-organized poster sections that maximize visual asset utilization while following human design patterns. Prioritize visual impact over text density.

Input:

- Paper Structured Sections: {{ structured sections }}
- Enhanced ABT Narrative: {{ narrative content }}
- Classified Visuals: {{ classified visuals }}
- Available Images: {{ available images }}
- Available Tables: {{ available tables }}
- Visual Heights Information: {{ visual heights info }}
- Available Height Per Column: {{ available height per column }}

Human Poster Design Patterns:

Based on analysis of successful academic posters:

1. **Left Column Strategy - Foundation & Context:**
 - Introduction/Background/Motivation (priority placement)
 - Problem definition and challenges
 - Related work and background context
 - Method overview or workflow diagrams
2. **Middle Column Strategy - Core Technical Content:**
 - Primary methodology (highest priority content)
 - Technical details and algorithms
 - Theoretical analysis and key innovations
 - System architecture diagrams
3. **Right Column Strategy - Experiments & Results:**
 - Experimental results (tables and performance charts)
 - Key findings and validation data
 - Performance comparisons and analysis

Figure 13. Part 1 of CuratorAgent prompt, focusing on the inputs, high-level instructions and human design patterns.

Oversized Visual Exclusion:

- **Exclusion Rule:** Any visual with height percentage > 50% in visual heights info **MUST BE EXCLUDED** from poster
- **Reasoning:** Even with 80% shrinking, these visuals would still exceed 40% column height
- **Smart Substitution:** ...
- **FALLBACK RULE:** If only ONE oversized visual (> 50%) is selected, allow it to proceed. For multiple oversized visuals, only select the one with **SMALLEST** height percentage.

Visual Asset Strategic Selection Process

1. Key Visual Mandatory Placement:

- Identify the “key visual” from classified visuals. This is the **MOST** important visual
- Place key visual in middle column, top priority section
- This anchors the entire poster layout around the core research contribution

2. Column-Based Visual Distribution:

• **Column 1 (Left) - Foundation & Context:**

- **MINIMUM:** 1 visual asset required
- **Purpose:** Express core research problem or contradiction visually
- **Selection Priority:** Choose visuals that illustrate problem context, background concepts, or prior work limitations
- **Maximum:** 2 visual assets

• **Column 2 (Middle) - Methodology:**

- **MANDATORY:** Contains key visual from classified visuals
- **Additional:** May include 1 supporting method diagram
- **Maximum:** 2 visual assets

• **Column 3 (Right) - Results & Impact:**

- **STRICT MAXIMUM:** 2 visual assets **ONLY**
- **Selection Criteria:** Choose the 2 most critical visuals that directly validate main claims
- **Priority Order:** ...

3. Visual Distribution Enforcement: ...

4. Column Space Optimization Strategy: ...

Core Task: Create 5-8 poster sections with **BOTH** content organization **AND** strategic spatial placement to achieve perfect space utilization across all three columns. **DO NOT** create any conclusion, takeaway, future work, or impact sections. Focus **ONLY** on problem, method, and results/experiments.

Content Organization Guidelines:

1. Section Requirements:

- **Section titles:** Maximum 4 words (e.g., “Our Method”, “Key Results”)
- **Text content:** 2-3 concise entries using different rich hierarchical formatting (see examples below) based on section contents
- **Visual integration:** Each visual assigned to exactly **ONE** section
- **Complete content:** No ellipsis (...), write full bullet points

2. Rich Text Formatting Options:

• **A) Nested Bullet Structure:**

- 1 “* Primary concept or finding”,
- 2 “ - Supporting detail or sub-point”,
- 3 “ - Additional supporting evidence”

- Other formats like **Bold Headers** and **Ordered Lists** are also available.

Figure 14. Part 2 of CuratorAgent prompt, specifying the detailed rules for visual asset selection, content organization, and other planning requirements.

CuratorAgent: Spatial Story Board Generation (Part 3 of 3)

Output Format:

```
1 {
2   "spatial_content_plan": {
3     "poster_strategy": {
4       "narrative_flow": "How the story progresses across columns",
5       "space_utilization_approach": "Strategy for filling all three columns",
6       "column_balance_rationale": "Why content is distributed this way"
7     },
8     "sections": [
9       {
10        "section_id": "unique_identifier",
11        "section_title": "Max 4 Words",
12        "column_assignment": "left|middle|right",
13        "vertical_priority": "top|middle|bottom",
14        "importance_level": 1,
15        "content_type": "foundation|method|results",
16        "expected_content_density": "high|medium|low",
17        "text_content": [
18          "* **Key Innovation:** Core contribution with bold emphasis",
19          "  - Supporting technical detail",
20          "* **Impact:** Quantifiable result or benefit"
21        ],
22        "visual_assets": [
23          {
24            "visual_id": "figure_1",
25            "visual_purpose": "How this supports the section",
26            "placement_rationale": "Why this visual belongs in this spatial location"
27          }
28        ],
29        "spatial_rationale": "Why this section belongs in this column/position"
30      }
31    ]
32  },
33  "column_distribution": {
34    "left_column": {
35      "focus": "Foundation and context",
36      "assigned_sections": ["section_id_1", "section_id_2"],
37      "content_strategy": "Build problem understanding and motivation"
38    },
39    "middle_column": {
40      "focus": "Core methodology",
41      "assigned_sections": ["section_id_3", "section_id_4"],
42      "content_strategy": "Present technical innovation and approach"
43    },
44    "right_column": {
45      "focus": "Results and impact",
46      "assigned_sections": ["section_id_5", "section_id_6"],
47      "content_strategy": "Demonstrate effectiveness and validations"
48    }
49  }
50 }
```

Figure 15. Part 3 of CuratorAgent prompt, defining the exact JSON output format and data structure required from the agent.

Balancer Agent: Balance Column Space Utilization (Part 1 of 2)

You are an expert academic poster layout optimization specialist. Your goal is to achieve optimal three-column space utilization through conservative within-column content adjustments only.

Current Column Status:

- **Column 1 (Left):** {left utilization} utilization - {left status}
- **Column 2 (Middle):** {middle utilization} utilization - {middle status}
- **Column 3 (Right):** {right utilization} utilization - {right status}
- **Available Height per Column:** {available height} inches

Target Utilization: 85-95% for each column

Core Optimization Principle: Prioritize content reduction over content expansion. Better to have 80% utilization than risk overflow beyond available space.

Column Content Rules:

1. Left Column: Foundation & Context
 - **Purpose:** Introduction, background, prior work, problem setup, supporting context
 - **Content Types:** Motivation, challenges, related work, problem definitions, supporting materials
 - **Reading Role:** Sets up the research problem and provides necessary background
2. Middle Column: Core Methodology
 - **Purpose:** Method details, algorithms, implementation, technical innovation
 - **Content Types:** Core methods, algorithms, technical approach, key innovations
 - **Reading Role:** Presents the technical contribution and methodology
 - **CRITICAL:** Contains key visual (importance level=1). NEVER remove method sections
3. Right Column: Results & Impact
 - **Purpose:** Experiments, evaluation, findings, conclusions, future work
 - **Content Types:** Experimental results, performance analysis, conclusions, future directions
 - **Reading Role:** Demonstrates validation and impact of the proposed method

Figure 16. Part 1 of the Balancer sub-agent prompt, outlining its role, the current column status, and the fundamental content rules for each column.

Balancer Agent: Balance Column Space Utilization (Part 2 of 2)

Within-Column Optimization Strategies:

1. Strategy A: Conservative Text Content Adjustment (for 80-100% utilization)
 - **When to use:** Column utilization is close to optimal range (80-100%)
 - **Actions allowed:**
 - MINIMAL text expansion: Add only 1-2 short phrases to underutilized columns (75-85%)
 - Aggressive text reduction: Significantly shorten content in overflow columns (>95%)
 - CONSERVATIVE APPROACH: Prefer slight underutilization over any risk of overflow
 - **Text Length Limits:**
 - Maximum per bullet: 25 words (count carefully)
 - Maximum sub-bullets: 2 per main bullet
 - Expansion limit: Add maximum 10-15 words total per section
 - Reduction target: Remove 30-50% of content from overflow sections
2. Strategy B: Section Management (for <80% or >100% utilization)
 - **When to use:** Column has severe underutilization (<80%) or overflow (>100%)
 - **Actions allowed:**
 - Add sections from structured sections: Use additional content from paper sections that fit the column's purpose
 - Remove less important sections: Remove sections with importance level = 3 or lower importance
 - **Section Removal Priority:**
 - NEVER remove: Method sections with key visual (importance level = 1)
 - NEVER remove: Core experimental results or main findings
 - Remove first: Supporting context, minor experiments, supplementary details (importance level = 3)
 - Remove second: Secondary analysis, additional background (importance level = 2)

Strict Constraints:

1. **NO CROSS-COLUMN MOVES:** Never change column assignment for any existing section
2. **PRESERVE READING FLOW:** Maintain left→middle→right logical progression
3. **SECTION ID PRESERVATION:** Never change section id, section title, visual assets, or other identifying fields
4. **IMPORTANCE RESPECT:** Never remove critical sections (importance level=1 or core results)
5. **TARGET UTILIZATION:** Achieve 85-95% utilization for each column

Input: {{structured sections}}, {current story board}, {column analysis}}

Output Format:

Output the complete optimized story board JSON. Each section's 'text content' must be an array of complete strings only:

```
1 "text_content": [  
2   "**Point Title:** Complete description text here",  
3   " - Supporting detail in complete sentences",  
4   "**Another Point:** Full explanation without truncation"  
5 ]
```

Preserve all original structure and field names. Only modify content within string values.

Figure 17. Part 2 of the Balancer sub-agent prompt, detailing the specific optimization strategies, strict constraints, and the required input/output format.

ColorAgent: Theme Color Extraction

Extract a sophisticated theme color from an affiliation logo that will work well as a poster accent color.

Core Task: Analyze the provided affiliation logo and identify the most prominent, meaningful color that can serve as a poster theme color. This color should be:

- Representative of the organization's visual identity
- Suitable for poster design applications (text highlights, accents)
- Professional and readable when used on white backgrounds
- Harmonious for academic poster contexts

Color Extraction Guidelines:

1. Primary Color Identification:

- Look for the main brand color of the organization
- Ignore pure white, black, and very light grays (background/outline colors)
- Focus on colored elements that define the logo's visual identity
- Consider text colors, graphic elements, symbols, and emblematic elements

2. Color Suitability Assessment:

- Too Bright: If the main color is very bright/saturated (e.g., neon yellow #FFFF00), generate a more subdued version
- Appropriate Saturation: Aim for colors that are vibrant but professional
- Readability: Ensure the color provides sufficient contrast on white backgrounds for text

3. Color Adjustment Rules:

- If original color is too bright (lightness > 85% or saturation > 90%), reduce brightness by 15-25%
- If original color is too dark (lightness > 25%), lighten slightly for better visibility
- Maintain the color's hue character while optimizing for poster applications

Output Requirements: Return ONLY a JSON object with the following structure:

```
1 {
2   "extracted_color": "#1E3A8A",
3   "color_name": "Professional Navy Blue",
4   "adjustment_made": "reduced_brightness | lightened | none",
5   "original_color": "#0000FF",
6   "suitability_score": 8.5,
7   "reasoning": "Extracted the primary blue from the university emblem. Reduced brightness
8     from bright blue to professional navy to ensure readability and sophisticated
9     appearance on white backgrounds.",
10  "usage_notes": "Excellent for text highlights, section headers, and accent elements.
11    Provides strong contrast while maintaining professional appearance."
```

Scoring Criteria (1-10 scale):

- **Contrast/Readability:** How well it works on white background
- **Professional Appearance:** Appropriate for academic/research contexts
- **Brand Representation:** How well it represents the organization
- **Poster Suitability:** Effectiveness for highlights and accents

Figure 18. Prompt for ColorAgent to extract theme color from affiliation logo.

FontAgent: Keyword Extraction

Analyze poster content and identify keywords for strategic visual highlighting using three distinct formatting styles.

Input Data:

- Enhanced Narrative: enhanced narrative
- Curated Content: curated content

Core Task: For each section, identify keywords and assign them to specific highlighting styles based on their semantic importance and role in the research narrative.

Highlighting Style Categories:

1. BOLD + CONTRAST COLOR:

- **Purpose:** Core method/methodology names that represent the paper's unique contribution
- **Criteria:** Novel algorithms, architectures, or techniques introduced by this work; the main methodological innovation that defines the paper; must be unique to this research (not generic terms)
- **Limit:** Maximum 2 per section, prefer 1 if it captures the main contribution

2. BOLD:

- **Purpose:** Important quantitative results and core technical terms within each section
- **Criteria:** Performance metrics and numerical results (e.g., "**95% accuracy**", "**5.2x speedup**"); key technical concepts central to understanding the section; architecture names, dataset names, established method names; word-level emphasis, not entire phrases
- **Limit:** Maximum 3 per section

3. ITALIC:

- **Purpose:** Defining terms, single-word emphasis, and foreign terminology
- **Criteria:** Technical terms being defined or introduced for the first time; single-word emphasis (e.g., "This was the *only* experiment"); foreign words, Latin terms, or specialized vocabulary; word-level application only, never entire sentences
- **Limit:** Maximum 2 per section

Output Format:

```
1 {
2   "section_keywords": {
3     "motivation": {
4       "bold_contrast": ["DP-CutMixSL"],
5       "bold": ["Vision Transformers", "privacy leakage"],
6       "italic": ["federated"]
7     },
8     "method": {
9       "bold_contrast": ["CutMix", "differential privacy"],
10      "bold": ["95% accuracy", "ResNet-50"],
11      "italic": ["only"]
12    },
13    "results": {
14      "bold_contrast": ["TransformerNet"],
15      "bold": ["top-1 accuracy", "5.2x speedup", "CIFAR-10"],
16      "italic": ["a priori"]
17    }
18  },
19  "formatting_summary": {
20    "total_bold_contrast": 4,
21    "total_bold": 7,
22    "total_italic": 3,
23    ...
24  }
25 }
```

Return a JSON object with the exact schema above that maximizes research impact through strategic visual emphasis.

Figure 19. Prompt for FontAgent to extract different types of keywords.

VLM-as-Judge Evaluation Template

You are an expert academic poster design critic. Your evaluation must be strict, detailed, and evidence-based. For each dimension, assess the poster against the specific examples provided in the 5-point scale. **High scores require adherence to professional design principles (positive examples). Low scores should be assigned when common design failures (negative examples) are present.** A poster that is merely functional but exhibits poor design choices must be scored significantly lower on design metrics than one that is both functional and visually excellent. Please carefully examine the provided poster image and evaluate it across the following metrics. Provide a detailed explanation and score on a 5-point scale.

FOCUS AREA: < Focus Area >

Metric: < Metric Description >

- **Score 5 (Excellent):**
 - **Descriptor:** ...
 - **Positive Examples:** ...
- **Score 4 (Good):**
 - **Descriptor:** ...
 - **Positive Examples:** ...
- **Score 3 (Acceptable):**
 - **Descriptor:** ...
 - **Negative Examples:** ...
- **Score 2 (Poor):**
 - **Descriptor:** ...
 - **Negative Examples:** ...
- **Score 1 (Failed):**
 - **Descriptor:** ...
 - **Negative Examples:** ...

Output Format: Please provide your evaluation as a JSON array with exactly 1 object for the metric:

```
1 [
2   {
3     "metric": "Core Graphic Principles",
4     "explanation": "Detailed analysis of the application of repetition, alignment, contrast,
5       and proximity principles...",
6     "score": 4
7   }
8 ]
```

Evaluation Instructions: ...

Please evaluate the poster step by step according to these criteria.

Figure 20. Prompt Template for VLM-as-Judge Evaluation.

	(A)	(B)	(C)
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
1. Layout Quality. Rate the overall layout. Consider: Does it use a clear grid? Is white space balanced? Is content distributed evenly? Is there a clear visual anchor?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
2. Color Quality. Rate the overall color system. Consider: Does it use a professional palette? Does color create a clear hierarchy? Is the text easy to read (high contrast)? Do colors feel purposeful?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
3. Typography Quality. Rate the overall typography. Consider: Are the fonts professional and readable? Is emphasis (bold/italic) used well? Is the text large and clear enough?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
4. Technical Precision. Rate the poster's polish. Consider: Are elements precisely aligned? Is it free from errors like text overflow, element overlap, or blurry images?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Figure 21. A sample set of questions from Qualtrics survey for human evaluation.

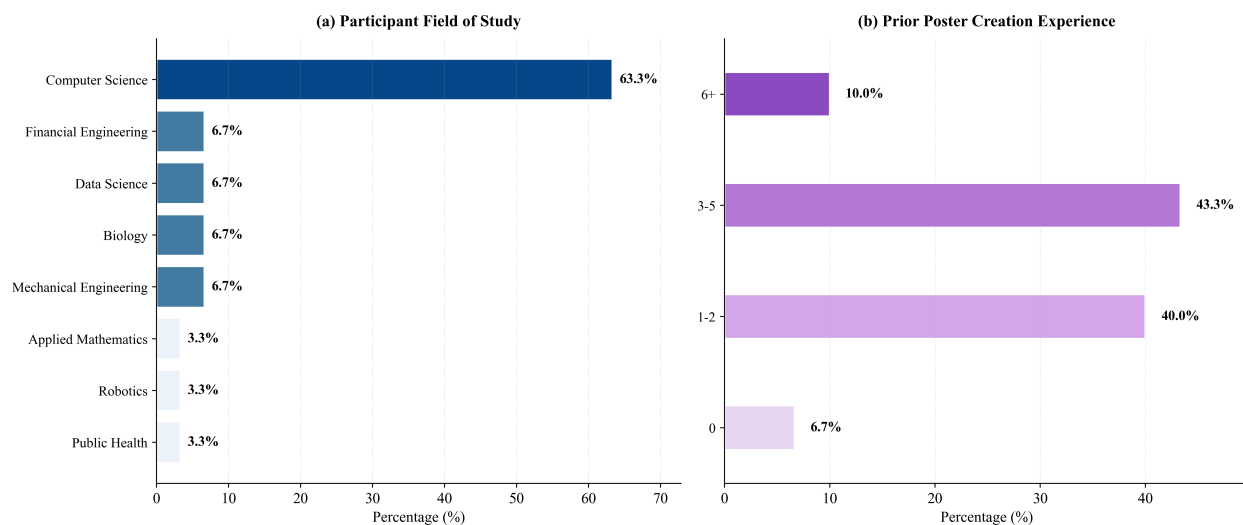


Figure 22. Participant demographics regarding field of study and prior poster creation experience.

Deep Compression of Pre-trained Transformer Models:
 Wang, Chi-Chun Liu, Swagath Venkataramani, Sanchai Sen, Chia-Yu Chh
 Kaoutar El Maghraoui, Vijayalakshmi Srinivasan, Leland Chen
 [wang.cdu@us.ibm.com, us.ibm.cro@us.ibm.cu]

Abstract
 Recent pre-training Transformer models have achieved success in NLP, vision, and speech. Transformer models are competitive in accuracy and speed. However, they require three arguments: operations across three model applications and large memory footprint.

4-bit Quantization
 FP32 Data Flow INT4 Pruned

50% Fine-Group Sparsity

Figure 1: 4-bit+50% sparse quantization for transformer models

Quantization and Sparsity

- All weights and activations in transformer models quantized to 4-bit
- Intermediate 4-bit difference address zero-preserving quantizers
- Paired-group quantizers for QKV mix
- Sparsity
- 50% sparsity is enforced using fine-tuning
- Searching Dropout is crucial regularizer

Conclusion
 All weights and activations in transformer models quantized to 4-bit

(a) Generated by GPT-4o-Image

Deep Compression of Pre-trained Transformer Models

Abstract
 Transformer models excel in NLP, vision, and speech. However, they require three arguments: operations across three model applications and large memory footprint.

Introduction
 Transformer models excel in NLP, vision, and speech. However, they require three arguments: operations across three model applications and large memory footprint.

Quantization & Sparsity
 All weights and activations in transformer models quantized to 4-bit. Intermediate 4-bit difference address zero-preserving quantizers. Paired-group quantizers for QKV mix. Sparsity: 50% sparsity is enforced using fine-tuning. Searching Dropout is crucial regularizer.

Hardware Implementation
 Evaluation conducted on a systolic array-based accelerator. Achieves significant memory capacity savings. Performance improvements through structured weight sparsity.

Conclusion
 We introduced a novel methodology for compressing transformer models. Careful selection of quantizers, data formats, and initialization strategies. Achieves significant compression while maintaining accuracy. The foundation for efficient deployment in various scenarios.

(b) Generated by PosterAgent

Deep Compression of Pre-trained Transformer Models

Pre-trained Transformer Models (aka. Foundation Models)

Quantization

Quantization and Sparsity-aware Fine Tuning

Performance

Acknowledgements

(c) Designed by humans

Deep Compression of Pre-trained Transformer Models

Why Compress Transformers

Our Approach

Key Results

Prior Solutions

Performance Comparison

(d) Generated by PosterGen (Ours)

Figure 27. Qualitative results of Wang et al. [49].

Punctuation-level Attack: Single-shot and Single Punctuation Attack Can Fool Text Models

Wenqiang Wang, Chongyang Du, Tao Wang, Kaihao Zhang, Kai Li, Jun Luo, Wenhao Luo, Lin Ma, Wei Liu, Xiaochun Cao

ABSTRACT
 Adversarial attack has attracted increasing attention in various fields, including natural language processing. This paper focuses on foot-text models. The anarchy-level attack achieves promising foot-text models against SOTA text models on typical text methods. Foot-text position punctuation and replacement priors and replacement promoting and replacement aware, and prepadding punctuation. Unmanageable misanthropic method's size area in natural language processing. Unmanageable misanthropic method's size area in natural language processing. Unmanageable misanthropic method's size area in natural language processing.

INTRODUCTION
 Adversarial examples cause a devastating impact in various domains, but being imperceptible in text. This paper focuses on foot-text models. The anarchy-level attack achieves promising foot-text models against SOTA text models on typical text methods. Foot-text position punctuation and replacement priors and replacement promoting and replacement aware, and prepadding punctuation. Unmanageable misanthropic method's size area in natural language processing. Unmanageable misanthropic method's size area in natural language processing. Unmanageable misanthropic method's size area in natural language processing.

(a) Generated by GPT-4o-Image

Punctuation-level Attack: Single-shot and Single Punctuation Attack Can Fool Text Models

Abstract
 Adversarial attacks are rising in NLP. This paper focuses on foot-text models. The anarchy-level attack achieves promising foot-text models against SOTA text models on typical text methods. Foot-text position punctuation and replacement priors and replacement promoting and replacement aware, and prepadding punctuation. Unmanageable misanthropic method's size area in natural language processing. Unmanageable misanthropic method's size area in natural language processing. Unmanageable misanthropic method's size area in natural language processing.

Related Works

Conclusion

(b) Generated by PosterAgent

Punctuation-level Attack: Single-shot and Single Punctuation Attack Can Fool Text Models

Wenqiang Wang, Chongyang Du, Tao Wang, Kaihao Zhang, Kai Li, Jun Luo, Wenhao Luo, Lin Ma, Wei Liu, Xiaochun Cao

Introduction

Embedding Method

Experimental Results

TPEP Training Algorithm

TPEP Searching Algorithm

(c) Designed by humans

Punctuation-level Attack: Single-shot and Single Punctuation Attack Can Fool Text Models

Research Challenge

Our Approach

Key Results

Prior Approaches

Attack Modes

Comparative Analysis

(d) Generated by PosterGen (Ours)

Figure 28. Qualitative results of Du et al. [8].

Abstract
DC Gaussian: Improving 3D Gaussian Splatting for Reflective Dash Cam Videos. DC-Gaussian achieves better results than the state-of-the-art method, showing superior performance in handling reflections and occlusions.

Introduction
Neural Radiance Field (NeRF) has achieved the state-of-the-art in novel view synthesis. However, existing methods struggle to handle reflections and occlusions in dash cam videos. DC-Gaussian addresses these challenges by introducing a geometry-guided Gaussian enhancement module and an illumination-aware obstruction modeling module.

Method
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.
Observation 1: The obstruction is like an image shared by all the frames in a dash cam video.
Observation 2: The strength of reflections is conditioned on cars' positions.

Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(a) Generated by GPT-4o-Image

DC-Gaussian: Improving 3D Gaussian Splatting for Reflective Dash Cam Videos

Abstract
DC-Gaussian generates novel views from dash cam videos, addressing reflections and occlusions in current methods. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

Method
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.

Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(b) Generated by PosterAgent

DC-Gaussian: Improving 3D Gaussian Splatting for Reflective Dash Cam Videos

Overview
We present DC-Gaussian, a new method for generating novel views from in-vehicle dash cam videos.

Method
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.

Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(c) Designed by humans

DC-Gaussian: Improving 3D Gaussian Splatting for Reflective Dash Cam Videos

Obstruction Challenge
Dash Cam Videos: Present events, not world. Obstructions: Present events, not world. Obstructions: Present events, not world.

Obstruction Types
Common Obstructions: Mobile phone, traffic light, etc. Obstructions: Present events, not world.

DC-Gaussian Method
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.

Main Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(d) Generated by PosterGen (Ours)

Figure 29. Qualitative results of Wang et al. [48].

Federated Online Adaptation for Deep Stereo

Abstract
Depth sensing plays a key role in several applications in fields of computer vision, robotics, and more. The use of stereo images (SI) for the purpose has been one of the most studied topics for decades, consisting of matching corresponding pixels and estimating horizontal disparity between corresponding pixels and, consequently, their depth through challenging environments, such as a semantic segmentation task.

Introduction
Depth sensing plays a key role in several applications in fields of computer vision, robotics, and more. The use of stereo images (SI) for the purpose has been one of the most studied topics for decades, consisting of matching corresponding pixels and, consequently, their depth through challenging environments, such as a semantic segmentation task.

(a) Generated by GPT-4o-Image

Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery

Abstract
Unsupervised landmark discovery is challenging. We propose a pose-guided self-training framework for unsupervised landmark discovery. It achieves superior performance in landmark discovery and is more robust to varying lighting conditions.

Method
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.

Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(b) Generated by PosterAgent

Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery

Background
Unsupervised Landmark Discovery (ULD) is a challenging problem. We propose a pose-guided self-training framework for unsupervised landmark discovery. It achieves superior performance in landmark discovery and is more robust to varying lighting conditions.

Method
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.

Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(c) Designed by humans

Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery

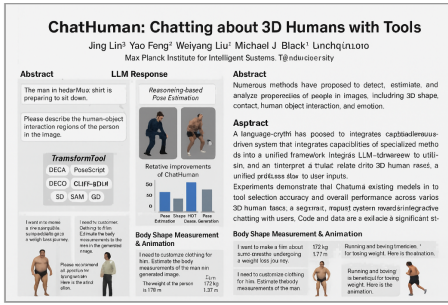
Problem & Motivation
Unsupervised Landmark Discovery (ULD) is a challenging problem. We propose a pose-guided self-training framework for unsupervised landmark discovery. It achieves superior performance in landmark discovery and is more robust to varying lighting conditions.

Our Approach
(1) Adaptive Image Decomposition: Decomposing images with learnable capacity map $J(u, v, z) = (1 - \phi)I_r(u, v, z) + \phi + I_o(u, v, z)$.
(2) Illumination-Aware Obstruction Modeling (IOM): Modeling obstruction with Global-aware Multi-resolution Hash Encoding and Local Frequency Modulation.

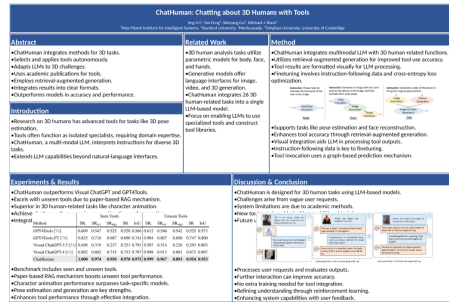
Key Results
DC-Gaussian outperforms the state-of-the-art method in handling reflections and occlusions. It achieves superior performance in novel view synthesis and is more robust to varying lighting conditions.

(d) Generated by PosterGen (Ours)

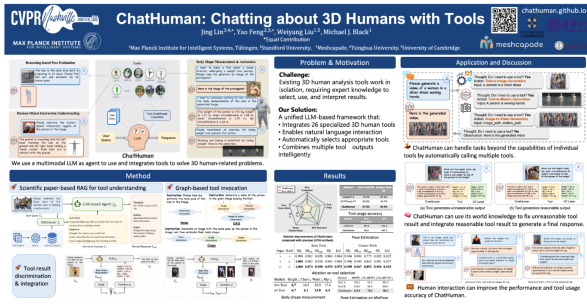
Figure 30. Qualitative results of Tourani et al. [46].



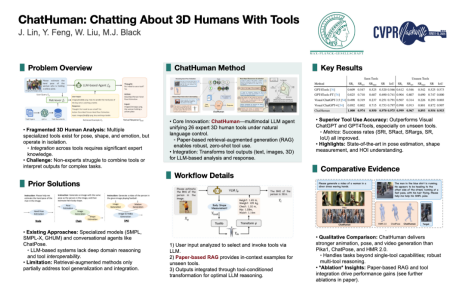
(a) Generated by GPT-4o-Image



(b) Generated by PosterAgent

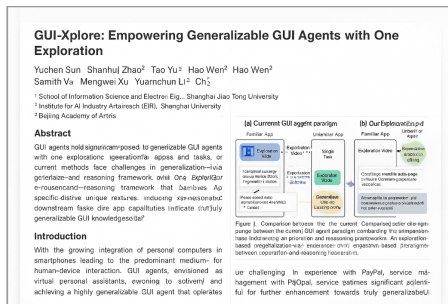


(c) Designed by humans

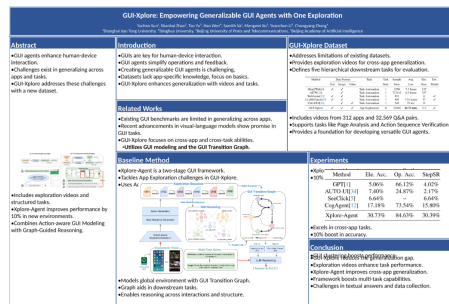


(d) Generated by PosterGen (Ours)

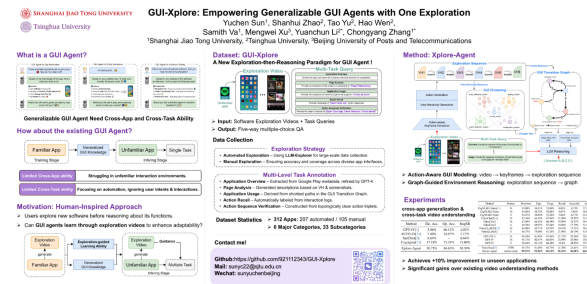
Figure 33. Qualitative results of Lin et al. [24].



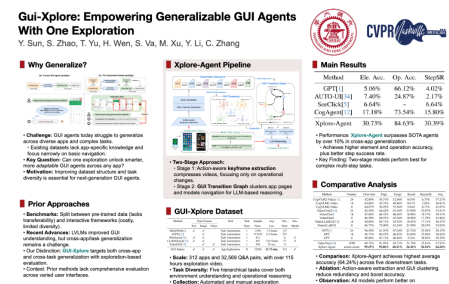
(a) Generated by GPT-4o-Image



(b) Generated by PosterAgent

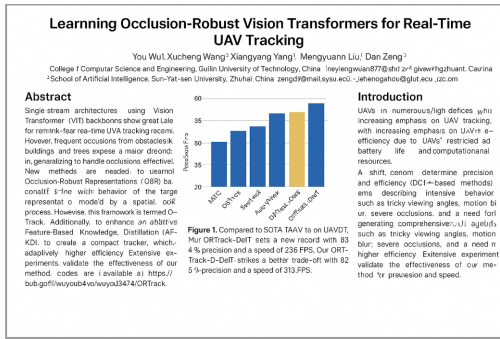


(c) Designed by humans

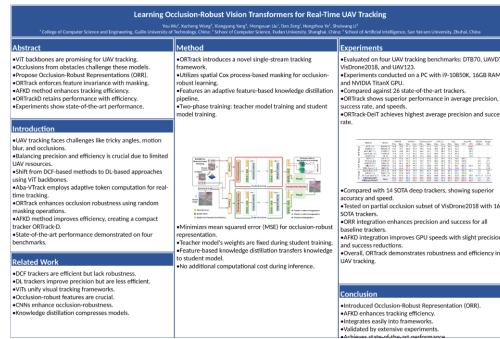


(d) Generated by PosterGen (Ours)

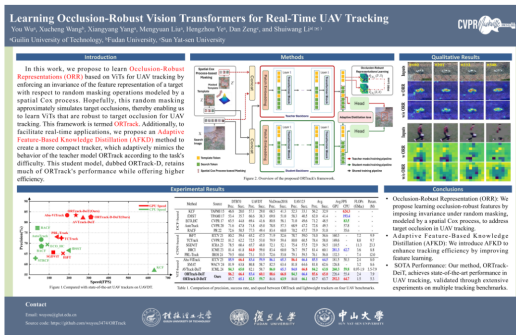
Figure 34. Qualitative results of Sun et al. [45].



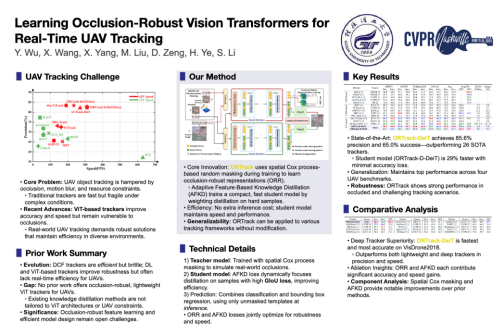
(a) Generated by GPT-4o-Image



(b) Generated by PosterAgent

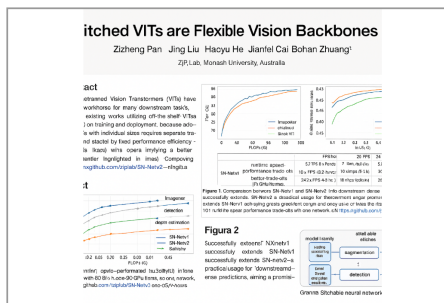


(c) Designed by humans

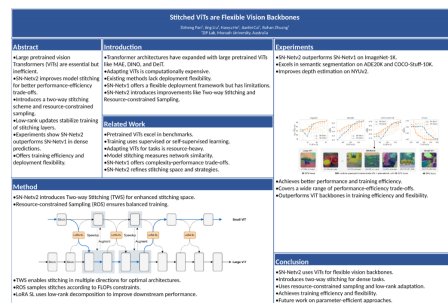


(d) Generated by PosterGen (Ours)

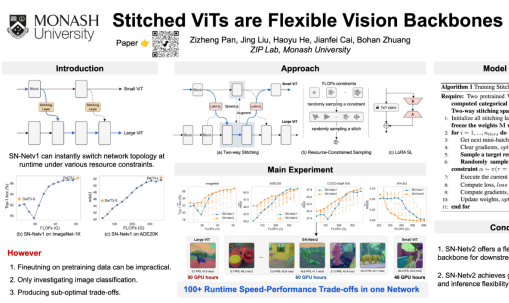
Figure 35. Qualitative results of Wu et al. [53].



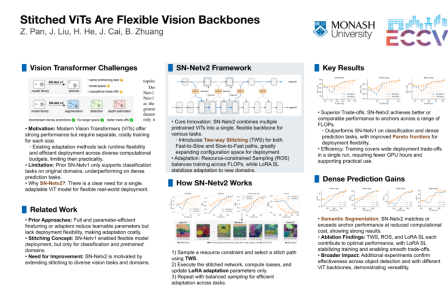
(a) Generated by GPT-4o-Image



(b) Generated by PosterAgent



(c) Designed by humans



(d) Generated by PosterGen (Ours)

Figure 36. Qualitative results of Pan et al. [31].

