

Optimization-Based Layout Synthesis for Multi-Model Medical Image Visualization

Chih-Yu Wang

National Yang Ming Chiao Tung University
wanglear6.ai12@nycu.edu.tw

Chi-Han Peng

National Yang Ming Chiao Tung University
pengchihan@nycu.edu.tw

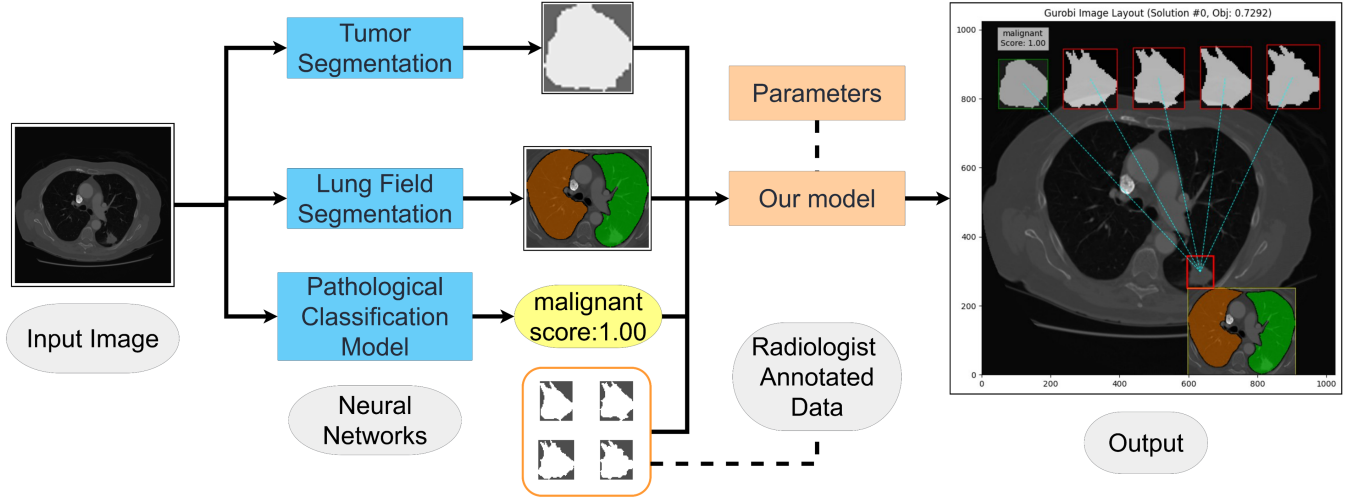


Figure 1: Workflow of our layout generation algorithm. First, given an input image, task-specific neural networks extract feature representations. Second, these features and radiologist-annotated data are converted into parameters for optimization. Finally, the resulting optimization problem is solved by the proposed MILP model with predefined parameters, and the solver generates the final optimized layout.

ABSTRACT

In computational clinical diagnosis, it is a common practice to examine multiple and possibly heterogeneous diagnostic results of a single image input. Existing layout approaches including trivial side-by-side / grid-based layouts or automatic 2D graphics layout methods can generate layouts with undesirable occlusions, loss of image aspect ratios, and inefficient uses of 2D space. Therefore, we propose a Mixed-Integer Linear Programming (MILP) optimization-based 2D layout method that is tailored for clinical diagnosis results. Tested on the LIDC-IDRI and QUBIQ medical image datasets, we show that our layout results are superior in terms of image overlaps, alignments, image aspect ratios, and 2D space utilization.

KEYWORDS

Data Visualization, Optimization-Based Algorithms, Mixed-Integer Linear Programming (MILP)

1 INTRODUCTION

Existing databases show that inter-observer variability exists in clinical diagnosis [1, 8]. Because different doctors may focus on different diagnostic features, clinical findings are often discussed among multiple doctors to reach a consensus. Similarly, many neural network-based models have been developed to assist in clinical diagnosis, some performing the same task and others performing different tasks such as lung segmentation and tumor detection [3, 5]. Therefore, a key challenge is how to present these multi-model outputs effectively in a 2D space, allowing doctors to reference outputs from different models at the same time.

Current layout generation research mainly relies on generative methods [6, 7, 9]. Although these approaches perform well on general design datasets such as Rico [2] and PubLayNet [10], their inherent randomness makes it difficult to meet the strict constraints and deterministic requirements of medical imaging reports.

To address this issue, we define layout synthesis as a Mixed-Integer Linear Programming (MILP) problem. Our method is specifically designed for clinical diagnostic needs and uses the Gurobi optimizer [4] to compute optimal layouts in real time.

A good medical image layout should satisfy several key requirements. First, the most fundamental quality of a medical layout is visual clarity. A good layout ensures that key diagnostic information, such as tumors and important anatomical structures, does

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CGW '26, July 09–10, 2026, Hsinchu City, Taiwan

© 2024 Copyright held by the owner/author(s).

not overlap. Second, related information should be placed in a connected and visually coherent manner. For example, a diagnostic text report should be placed close to its corresponding tumor image. Third, the layout should follow a clear order: it should first present the whole structure (e.g., full CT scans), then regional parts (e.g., lungs), and finally small regions (e.g., tumors). This organization provides a basis for assigning element weights in our method. Finally, the original aspect ratio of each image should be preserved so that doctors can correctly understand the size of tumors. An ideal medical image layout avoids unexpected image distortion or cropping, helping reduce the risk of incorrect understanding of the physical size of a lesion.

We summarize our contributions as follows:

- **Optimization-Based Layout synthesis Algorithm:** Unlike generative models that rely on random processes, we redefine the layout problem as a Mixed-Integer Linear Programming (MILP) problem to meet the requirements of medical image layout. We use a Gurobi optimizer to find the mathematically optimal layout.
- **Controlled Non-overlapping Constraints:** We introduce controlled non-overlapping constraints, which define a configurable and precise threshold for spatial overlap between diagnostic images. This mechanism balances information density and visual clarity.
- **Data-Independent Adaptability and Zero Retraining:** Our algorithm is data-independent and does not require retraining. It can adapt to different numbers of input images and heterogeneous outputs from different neural networks.
- **Empirical Validation on Clinical Datasets:** We conduct experiments on publicly available medical datasets for layout generation. Our method significantly outperforms state-of-the-art generative models in both layout accuracy and speed.

2 RELATED WORK

2.1 Medical Data

Large-scale public datasets are widely used for developing and evaluating medical imaging methods. The Cancer Imaging Archive (TCIA) provides a large collection of multimodal medical images (e.g., CT, MRI, and PET) in DICOM format, along with clinical annotations and related metadata. It hosts several widely used datasets, including the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), one of the most representative datasets for lung nodule analysis.

This study uses two datasets that exhibit inter-observer variability. The LIDC-IDRI dataset[1], hosted in TCIA, contains 1,018 lung CT scans for lung nodule analysis. Each scan was independently annotated by four radiologists who marked potential tumors and later reviewed each other’s annotations. This process results in multiple segmentation masks per case, where tumor boundaries often differ across radiologists. Therefore, LIDC-IDRI is widely used for studying inter-observer variability in tumor annotation.

We also use the Quantification of Uncertainties in Biomedical Image Quantification (QUBIQ) Challenge dataset[8]. QUBIQ consists of 2D slices from CT and MRI scans across multiple organs, including the prostate, brain, and kidney. It defines several segmentation tasks in which each structure is typically annotated three to seven

times by different doctors, resulting in multiple valid segmentation masks per case. The goal is to predict segmentations that match the distribution of these annotations, making QUBIQ a benchmark for studying inter-observer variability in multi-organ medical image segmentation.

These datasets show that medical images often involve multiple valid interpretations rather than a single ground truth, highlighting the inter-observer variability that motivates the need for effective visualization of multiple model outputs.

2.2 Layout Generation

Layout generation can be broadly categorized into early optimization-based methods and recent neural network-based generative models.

State-of-the-art layout generation methods primarily use generative models to learn from large-scale datasets. Recent architectures, such as LayoutDM and LayoutDiffusion, utilize discrete diffusion probabilistic models to synthesize layouts through iterative denoising. Early foundational work, particularly LayoutGAN++, adopts a generative adversarial network (GAN) framework and incorporates constraint-aware loss functions to improve spatial quality.

Although these generative methods have achieved strong performance on general design datasets, they are inherently random. Moreover, existing datasets are not specifically designed for medical image layout. However, in medical imaging applications, doctors require stable and reliable layout results.

In contrast, optimization-based methods define layout generation as a Mixed-Integer Linear Programming (MILP) problem, addressing the randomness of generative approaches. With recent advances in optimization solvers, high-quality solutions can be efficiently computed using tools such as Gurobi [4], enabling both stable and fast layout generation.

3 METHOD

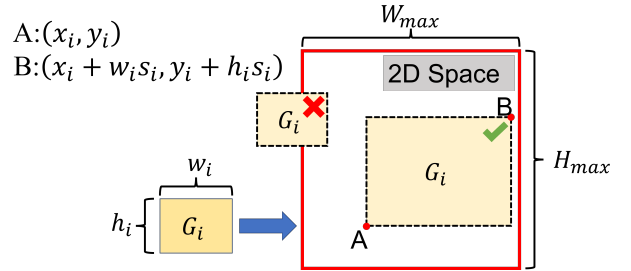


Figure 2: This figure shows a graph element G_i with its original size (w_i, h_i) and the bounded layout region defined by W_{max} and H_{max} . It displays cases of valid and invalid layouts caused by incorrect element placement that exceeds the layout boundary. In the 2D space, point A corresponds to the optimized bottom-left position of G_i , while $w_i s_i$ and $h_i s_i$ denote its scaled width and height after optimization.

As shown in Figure 1, our pipeline consists of two main stages. First, the input medical images are processed by neural networks designed for different tasks to extract relevant features. These features

are then fed into our proposed MILP-based algorithm to generate the final layout.

The MILP formulation includes four components: constants, variables, constraints, and objective functions.

3.1 Constants

To construct the Mixed-Integer Linear Programming (MILP) model, we first define a set of constant parameters derived from image metadata and predefined layout configurations, including images, text, and numerical information.

To represent these heterogeneous data types in a 2D space, we model them as graph elements. We denote the index set of graph elements as $\mathcal{N} = \{0, 1, \dots, N-1\}$, where N is the total number of elements. The set of elements to be placed in the 2D space is denoted as $G = \{G_i \mid i \in \mathcal{N}\}$. Each element G_i has a size (w_i, h_i) obtained from neural network outputs or metadata. The layout is bounded by a fixed region with maximum width $W_{\max}$ and height $H_{\max}$. We further define the lower and upper bounds for the scaling factors, denoted by $S_{\min}$ and $S_{\max}$. We define a minimum spatial gap D between non-overlapping elements to ensure visual clarity. Finally, each element G_i is associated with a visual priority weight ω_i that reflects its diagnostic importance in the report. To allow semantically related elements to be presented together, we define a group $\mathcal{G} \subseteq \mathcal{N}$, which contains elements with similar semantic features.

3.2 Variables

We solve the layout synthesis problem by optimizing the following variables to generate a high-quality layout.

Let $x = \{x_0, x_1, \dots, x_{N-1}\}$ and $y = \{y_0, y_1, \dots, y_{N-1}\}$ denote the sets of integer variables, where each x_i, y_i represents the bottom-left coordinates of element G_i in the 2D space.

In addition, let $s = \{s_0, s_1, \dots, s_{N-1}\}$ denote the set of continuous scaling factor variables, where each $s_i \in \mathbb{R}$ is bounded by $[S_{\min}, S_{\max}]$. This variable determines the scale of element G_i , preserving its original aspect ratio while improving space utilization within the layout.

Consequently, the optimized display width W'_i and height H'_i for each element are derived as $W'_i = s_i \cdot w_i$ and $H'_i = s_i \cdot h_i$.

3.2.1 Variables for overlap. To enable controllable overlapping, we introduce binary decision variables defined over pairs of elements. Specifically, for each pair (i, j) with $i < j$, we define directional binary variables $b_{i,j}^{\text{left}}, b_{i,j}^{\text{right}}, b_{i,j}^{\text{up}}, b_{i,j}^{\text{down}} \in \{0, 1\}$, which control the spatial relationship between graph elements G_i and G_j .

In addition, for numerical or text-based elements, we define $b_{\text{text}}^{\text{up}}, b_{\text{text}}^{\text{down}} \in \{0, 1\}$ to control the vertical relationship between the text element G_{text} and its associated graph element G_i . This ensures that the diagnostic label (benign or malignant) is positioned above or below the corresponding lesion.

3.3 Constraints

3.3.1 2D Space Boundary Constraints. To ensure all elements are fully contained within the 2D space, for all $i \in \mathcal{N}$, the following

constraints hold:

$$x_i \geq 0, \quad y_i \geq 0, \quad \forall i \in \mathcal{N} \quad (1)$$

$$x_i + W'_i \leq W_{\max}, \quad \forall i \in \mathcal{N} \quad (2)$$

$$y_i + H'_i \leq H_{\max}, \quad \forall i \in \mathcal{N} \quad (3)$$

3.3.2 Non-overlapping Constraints. The directional binary variables $b_{i,j}^{\text{left}}, b_{i,j}^{\text{right}}, b_{i,j}^{\text{up}},$ and $b_{i,j}^{\text{down}}$ are used to control the spatial relationship between graph elements G_i and G_j via the following constraints:

$$x_i + W'_i + D \leq x_j, \quad \text{if } b_{i,j}^{\text{left}} = 1 \quad (4)$$

$$x_j + W'_j + D \leq x_i, \quad \text{if } b_{i,j}^{\text{right}} = 1 \quad (5)$$

$$y_j + H'_j + D \leq y_i, \quad \text{if } b_{i,j}^{\text{up}} = 1 \quad (6)$$

$$y_i + H'_i + D \leq y_j, \quad \text{if } b_{i,j}^{\text{down}} = 1 \quad (7)$$

To ensure visual quality, we handle non-overlapping in two steps. First, we prevent elements with the same visual priority weight from overlapping, as shown in Eq. (8). Second, for elements with different visual priority weights, we limit the overlap depth to at most two layers, as shown in Eq. (9).

$$b_{i,j}^{\text{left}} + b_{i,j}^{\text{right}} + b_{i,j}^{\text{up}} + b_{i,j}^{\text{down}} \geq 1, \quad \text{if } \omega_i = \omega_j, \quad \forall i < j \quad (8)$$

$$\sum_{\text{dir} \in \{L, R, U, D\}} (b_{i,j}^{\text{dir}} + b_{i,k}^{\text{dir}} + b_{j,k}^{\text{dir}}) \geq 1, \quad \text{if } \omega_i \neq \omega_j, \omega_i \neq \omega_k, \omega_j \neq \omega_k, \quad \forall i < j < k \quad (9)$$

To prevent occlusion of critical information in selected elements G_i , we introduce an invisible element G_{crit} . This element is not rendered in the final layout and is used only to enforce spatial constraints.

The size of G_{crit} , denoted by $(w_{\text{crit}}, h_{\text{crit}})$, is determined from annotated critical region information. A fixed offset $C_{\text{offset}} = (\Delta x, \Delta y)$ is used to locate G_{crit} relative to G_i . The corresponding constraints are shown below:

$$s_{\text{crit}} = s_i \quad (10)$$

$$x_{\text{crit}} = x_i + \Delta x \cdot s_i \quad (11)$$

$$y_{\text{crit}} = y_i + \Delta y \cdot s_i \quad (12)$$

3.3.3 Alignment Constraints. To improve the spatial relationships between visual elements in the report layout, we define alignment constraints for grouped elements, diagnostic text, and tumor images.

For grouping elements, we enforce the following constraints on the subset $\mathcal{G} \subseteq \mathcal{N}$.

For horizontally aligned groups, all elements share the same vertical coordinate:

$$y_i = y_{i_0}, \quad \forall i \in \mathcal{G} \quad (13)$$

where i_0 is the first element in the group.

To maintain uniform horizontal spacing, we enforce the following constraint on three consecutive elements (a, b, c) in the group $\mathcal{G}$:

$$x_b - (x_a + w_a s_a) = x_c - (x_b + w_b s_b) \quad (14)$$

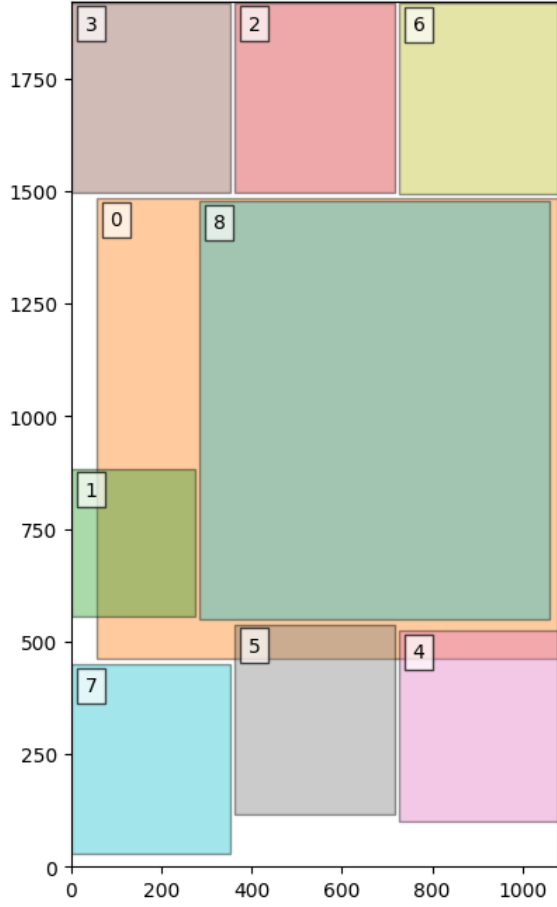


Figure 3: Critical Information Preservation mechanism. Element 8 represents the critical region of element 0, introduced to prevent overlap with its critical area. By using the constraint that elements with different weights are allowed to overlap up to two layers in Eq. 9, overlap on element 0 is prevented.

Similarly, for vertically aligned groups, the constraints as follows:

$$x_i = x_{i_0}, \quad \forall i \in \mathcal{G} \quad (15)$$

$$y_b - (y_a + h_a s_a) = y_c - (y_b + h_b s_b) \quad (16)$$

To align the features extracted from neural networks, such as diagnostic text and tumor images, we define alignment constraints between related visual elements. The diagnostic text and tumor image are constrained to share the same horizontal center. In addition, the diagnostic text is constrained to appear either above or below the tumor image. The corresponding binary variables satisfy a constraint to ensure that only one relative position is selected.

$$x_{\text{text}} + \frac{1}{2} w_{\text{text}} s_{\text{text}} = x_{\text{tumor}} + \frac{1}{2} w_{\text{tumor}} s_{\text{tumor}} \quad (17)$$

$$y_{\text{text}} = y_{\text{tumor}} + h_{\text{tumor}} s_{\text{tumor}} + D, \quad \text{if } b_{\text{text}}^{\text{up}} = 1 \quad (18)$$

$$y_{\text{tumor}} = y_{\text{text}} + h_{\text{text}} s_{\text{text}} + D, \quad \text{if } b_{\text{text}}^{\text{under}} = 1 \quad (19)$$

$$b_{\text{text}}^{\text{up}} + b_{\text{text}}^{\text{under}} = 1 \quad (20)$$

3.4 Objective Function

The final layout is obtained by maximizing an objective function composed of scale score, group centering score, and group scale variance score:

$$\max f_{\text{scale}} - f_{\text{center}} - f_{\text{var}} \quad (21)$$

where the components are defined as follows.

3.4.1 Scale Score. To ensure sufficient visibility, we define a normalized function that represents the scaling level of all elements:

$$f_{\text{scale}} = \frac{\sum_{i \in \mathcal{N}} \omega_i s_i}{S_{\text{max}} \cdot \sum_{j \in \mathcal{N}} \omega_j} \quad (22)$$

3.4.2 Group Centering Score. To improve the visual symmetry of the layout, we evaluate the overall centering quality by calculating the offset between the group center and the center of the 2D space:

$$f_{\text{center}} = \begin{cases} \left(\frac{\sum_{i \in \mathcal{G}} (x_i + \frac{1}{2} w_i s_i)}{|\mathcal{G}| \cdot \frac{W_{\text{max}}}{2}} - 1 \right)^2, & \text{if horizontal group} \\ \left(\frac{\sum_{i \in \mathcal{G}} (y_i + \frac{1}{2} h_i s_i)}{|\mathcal{G}| \cdot \frac{H_{\text{max}}}{2}} - 1 \right)^2, & \text{if vertical group} \end{cases} \quad (23)$$

3.4.3 Group Scale Variance Score. To avoid large scaling differences between elements in the same group, we evaluate this score by calculating the scale variance within the group:

$$f_{\text{var}} = \frac{\sum_{i \in \mathcal{G}} (s_i - \bar{s})^2}{(S_{\text{max}} - S_{\text{min}})^2}, \quad \text{where } \bar{s} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} s_i \quad (24)$$

4 RESULTS AND ANALYSIS

We executed all tasks on a computer with Intel i7-12700F CPU, 32 GB RAM, and an NVIDIA RTX 4090 GPU (24 GB VRAM).

4.1 Dataset

To evaluate the proposed layout synthesis method, we used the Brain Growth and Kidney tasks from the QUBIQ challenge and Lung CT tasks from the LIDC-IDRI dataset[1, 8].

The original medical images, doctor annotations, and features extracted from neural networks were used for layout synthesis, where the width and height of the corresponding bounding boxes were defined as the element dimensions (w_i, h_i) in the MILP algorithm.

4.2 Evaluation Metrics

We evaluate the generated layouts using both internal optimization metrics and external layout evaluation metrics. The internal metrics are directly derived from the objective function defined in Section 3.4, while the alignment and overlap metrics from previous work [7] are adopted for fair comparison.

The first three metrics (f_{scale} , f_{center} , and f_{var}) correspond to the optimization objectives and are used to analyze how well each

optimization goal is achieved. They are not intended to serve as independent perceptual quality metrics. Therefore, we additionally evaluate the generated layouts using $\mathcal{L}_{\text{align}}$, $\mathcal{L}_{\text{overlap}}$, and qualitative comparisons to provide complementary assessments of layout quality.

Unlike previous layout generation methods, we interpret $\mathcal{L}_{\text{overlap}}$ as a space utilization metric rather than a pure overlap penalty. Since our method allows controlled overlap in non-critical regions, a larger $\mathcal{L}_{\text{overlap}}$ indicates more efficient use of the available 2D space. However, this metric should be interpreted together with the proposed layout constraints. Excessive overlap can artificially increase $\mathcal{L}_{\text{overlap}}$ while violating layout constraints and reducing readability. Therefore, a larger $\mathcal{L}_{\text{overlap}}$ is desirable only for layouts that satisfy the proposed constraints.

4.3 Implementation Details

4.3.1 Our Method. We set the 2D space to $W_{\text{max}} = 1080$ and $H_{\text{max}} = 1920$, with a minimum padding of $D = 10$ across all experiments.

For the Lung CT task, we set the scaling range to $S_{\text{max}} = 10$, $S_{\text{min}} = 0.5$ and use a three-level weighting scheme (40-20-3) to balance CT images, segmentation masks, and expert annotations.

For the Brain Growth and Kidney tasks, we set the scaling range to $S_{\text{max}} = 4$, $S_{\text{min}} = 0.5$ and use a two-level weighting scheme (20-3) to balance primary images and annotations.

For the Brain Growth task, the group alignment constraint was disabled due to the large number of similar elements. This adjustment improves the overall layout readability and visual balance.

The optimization convergence is shown in Figure 4.

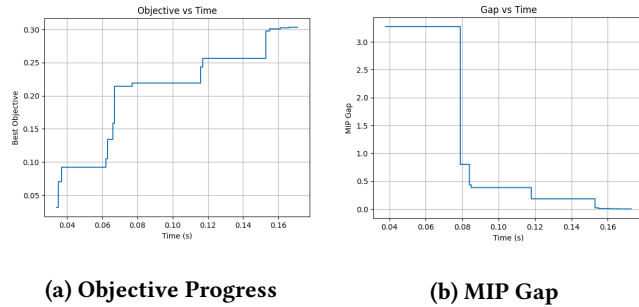


Figure 4: Optimization convergence of our method for Lung CT. The convergence behavior of the objective function (Eq. 21) is shown in (a). The optimization process terminates when the MIP gap reaches 0%, indicating that the optimal solution has been proven by the solver.

4.3.2 Baseline Layout Models. For all baseline methods, we use the same input configuration to ensure a fair comparison. Specifically, we set the 2D canvas size to $W_{\text{max}} = 1080$ and $H_{\text{max}} = 1920$, and provide each method with identical element attributes, including their widths w_i and heights h_i .

We evaluate a range of recent multimodal large language models (MLLMs), including GPT, Gemini, Grok, and Qwen. All models are provided with the same structured inputs and semantic descriptions, and are prompted to directly generate layout outputs.

We also experimented with directly providing visual inputs; however, it may lead to unstable outputs, as shown in Figure 5.

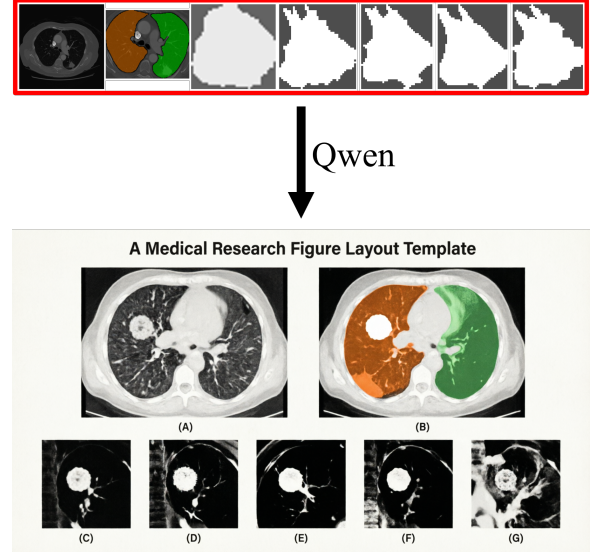


Figure 5: Unstable Layout Generation with Direct Visual Inputs. When visual inputs are directly provided to Qwen, we can see that the images in the generated layout may be different from the original images.

4.4 Quantitative Comparisons

We compare our method and neural network-based approaches under the evaluation metrics defined in Section 4.2. The quantitative results are summarized in Table 1–Table 3, while rule violation statistics are reported in Table 4.

Table 1: Performance comparison on Lung CT. Although GPT has higher f_{scale} , it produces layouts with occlusion errors. Gemini achieves competitive performance in terms of $\mathcal{L}_{\text{overlap}}$ and $\mathcal{L}_{\text{align}}$, while our method performs better on the remaining metrics, indicating superior overall space utilization and improved preservation of semantic relationships in the layout. Highlighting: **best**, **second-best**.

Model	$f_{\text{scale}} \uparrow$	$f_{\text{var}} \downarrow$	$f_{\text{center}} \downarrow$	Total Score $\uparrow$	$\mathcal{L}_{\text{overlap}} \uparrow$	$\mathcal{L}_{\text{align}} \downarrow$
Ours (Full Model)	0.3034	0.000037	0.000093	0.7678	0.6563	0.0427
Ours (Ablation)	1.0000	0.0000	0.0000	1.0000	3.9844	1.7026
LayoutGAN++	0.1468 ± 0.0581	0.0797 ± 0.0568	0.3413 ± 0.1351	0.5753 ± 0.0515	0.2564 ± 0.1719	0.0975 ± 0.0195
LayoutDiffusion	0.0892 ± 0.0266	0.0169 ± 0.0322	0.1951 ± 0.0847	0.6257 ± 0.0317	0.1478 ± 0.2442	0.0779 ± 0.0299
LayoutDM	0.1024	0.0008	0.2021 ± 0.1268	0.6332 ± 0.0423	0.1461	0.1083 ± 0.0377
Gemini	0.2180	0.0002	0.1783	0.6798	0.8810	0.0237
GPT	0.3325	0.0010	0.2500	0.6938	0.5348	0.0973

4.5 Qualitative Comparisons

We perform a qualitative visual comparison between our method and neural network-based approaches. Representative results for each task are shown in Figures 6–8.

From Figures 6–8, we observe that most neural network-based approaches fail to fully capture the semantic meaning of different

Table 2: Performance comparison on Brain Growth. Although LayoutDiffusion achieves higher $\mathcal{L}_{\text{overlap}}$, it produces distorted elements with incorrect overlaps. LayoutDM performs better in $\mathcal{L}_{\text{align}}$, but its layout is less efficient in space utilization.

Model	$f_{\text{scale}} \uparrow$	$f_{\text{var}} \downarrow$	$f_{\text{center}} \downarrow$	Total Score $\uparrow$	$\mathcal{L}_{\text{overlap}} \uparrow$	$\mathcal{L}_{\text{align}} \downarrow$
Ours (Full Model)	0.7135	0.0117	4.77×10^{-9}	0.9006	0.3253	0.0278
Ours (Ablation)	1.0000	0.0000	0.0000	1.0000	3.0931	0.0403
LayoutGAN++	0.3291 ± 0.0742	0.0701 ± 0.0497	0.2386 ± 0.0937	0.6735 ± 0.0071	0.0000	0.0929 ± 0.0108
LayoutDiffusion	0.4491 ± 0.2080	0.3500 ± 0.3129	0.4198 ± 0.2391	0.5597 ± 0.1273	1.2074 ± 1.3937	0.0980 ± 0.0676
LayoutDM	0.6921	0.0172	0.2827	0.7974	0.0889	0.0265
Gemini	0.6231	0.0000	0.5625	0.6869	0.8993	0.0709
GPT	0.6505	0.0000	0.3131	0.7791	0.5687	0.0540

Table 3: Performance comparison on Kidney. The proposed method is better than all baseline methods across all evaluation metrics.

Model	$f_{\text{scale}} \uparrow$	$f_{\text{var}} \downarrow$	$f_{\text{center}} \downarrow$	Total Score $\uparrow$	$\mathcal{L}_{\text{overlap}} \uparrow$	$\mathcal{L}_{\text{align}} \downarrow$
Ours (Full Model)	0.4740	0.000002	0.000002	0.8247	0.3573	0.0698
Ours (Ablation)	1.0000	0.000000	0.000000	1.0000	2.8064	0.4276
LayoutGAN++	0.2493 ± 0.1083	0.014440 ± 0.016148	0.247535 ± 0.007831	0.662436 ± 0.035400	0.0000	0.241998 ± 0.081081
LayoutDiffusion	0.093905 ± 0.026669	0.000863 ± 0.001127	0.214421 ± 0.168004	0.625907 ± 0.058141	0.151222 ± 0.284489	0.141612 ± 0.0794
LayoutDM	0.245424	0.000107	0.328483	0.134261 ± 0.044754	0.001077	0.156420 ± 0.021708
Gemini	0.401583	0.000006	0.250000	0.717192	0.0000	0.115236
GPT	0.328483	0.000327	0.246923	0.693744	0.038622	0.153304

Table 4: Rule violation quantification across different constraints. Each cell represents the violation rate for a specific constraint across all tasks. Lower is better. Note: LB = Layout Boundary; NIW = Non-overlapping for Identical Weights; MOL = Max Overlap Limit; CIP = Critical Information Protection; GA = Group Alignment

Model	LB	NIW	MOL	CIP	GA	Average
Ours (Full Model)	0.00	0.00	0.00	0.00	0.33	0.07
Ours (Ablation)	1.00	1.00	1.00	1.00	1.00	1.00
LayoutGAN++	0.33	0.00	0.00	0.13	0.37	0.17
LayoutDiffusion	0.07	0.15	0.45	0.30	0.48	0.29
LayoutDM	0.00	0.42	0.00	0.33	1.00	0.35
Gemini	0.00	0.00	0.00	0.33	0.67	0.20
GPT	0.00	0.33	0.33	0.67	1.00	0.47

layout elements. Compared with baseline methods, the proposed MILP-based algorithms achieves superior space utilization and information readability. Although Gemini and GPT can partially understand the semantic meaning of elements through prompts, they still cannot completely avoid the risk of information occlusion.

4.6 Ablation Study

To verify the necessity of each constraint, we perform an ablation study by removing constraints one at a time. We analyze their effects using the evaluation metrics defined in Section 4.2.

Quantitative results are reported in Table 5, and qualitative comparisons are shown in Figure 9.

From Table 5 and Figure 9, we analyze the effects of each constraint as follows:

- **Layout Boundary.** Removing this constraint significantly improves f_{scale} by enabling better utilization of the layout space. However, without proper boundary control, some

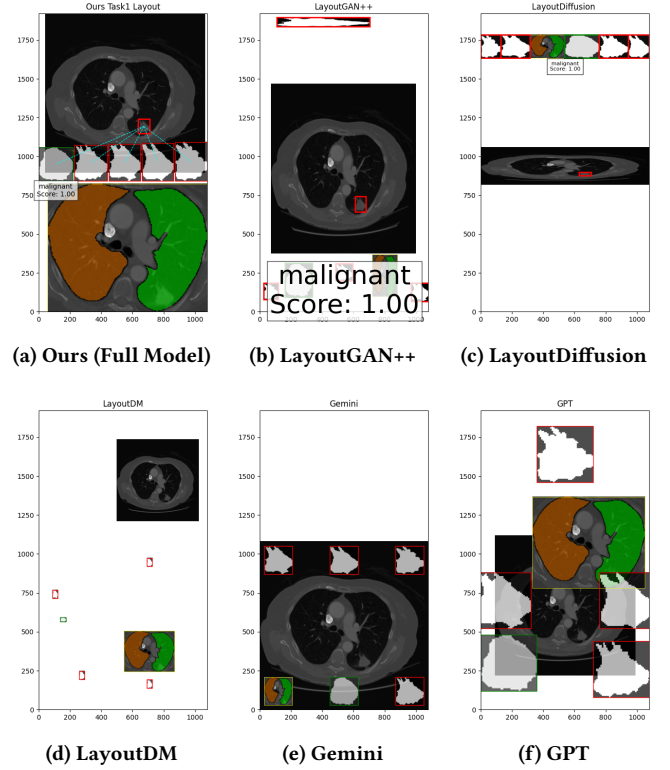


Figure 6: Qualitative comparison for Lung CT. (b) and (d) show randomly placed elements, resulting in poor readability. (c) preserves semantic relationships but leads to low space utilization due to small element sizes. (e) achieves a relatively structured layout but leaves the upper region of the 2D space largely unused. (f) improves element scaling but suffers from occlusion and weakened semantic relationships between elements.

elements may exceed the layout region, leading to a loss of important information.

- **Non-overlapping for Identical Weights, Max Overlap Limit, and Critical Region Protection.** Removing these constraints significantly increases $\mathcal{L}_{\text{overlap}}$, but overlaps between elements leads to poor information clarity.
- **Text Space Allocation.** Without this constraint, diagnostic text is randomly placed, making it difficult to associate it with the corresponding images.
- **Group Alignment.** Disabling this constraint causes semantically similar images to be spatially dispersed, resulting in reduced readability and weaker structural organization of the layout.

4.7 Non-uniform Scaling

We implement non-uniform scaling. The Total Score is 0.803648, f_{scale} is 0.411032, f_{var} is 0.000052, f_{center} is 0.000035, $\mathcal{L}_{\text{overlap}}$ is 0.853387, and $\mathcal{L}_{\text{align}}$ is 0.040715. Although this variant improves several quantitative metrics, as shown in Fig. 10, the distortion of aspect ratios

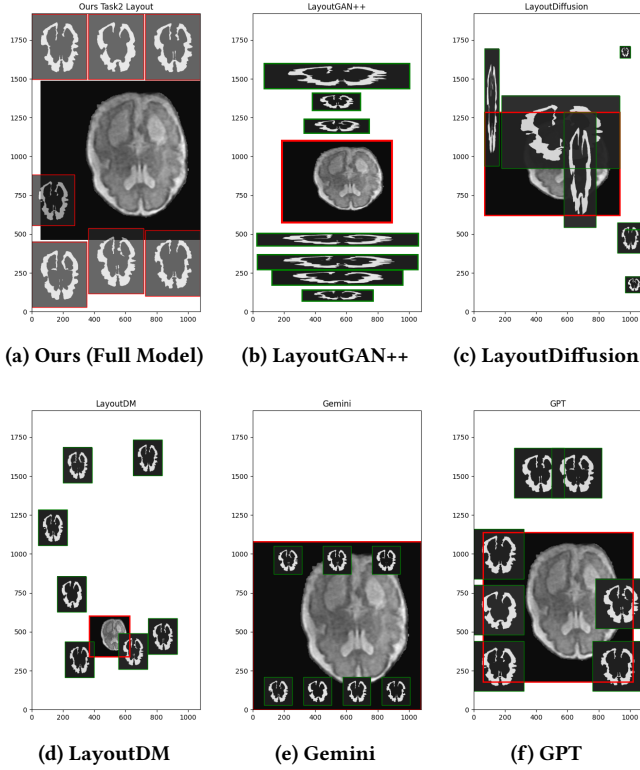


Figure 7: Qualitative comparison for Brain Growth. (b) achieves good center alignment among semantically related elements but suffers from incorrect aspect ratios. (c) shows randomly placed elements with distorted aspect ratios. (d) shows randomly placed elements. (e) achieves a relatively structured layout but leaves the upper region of the 2D space largely unused and introduces occlusion of important information. (f) also suffers from occlusion of critical information.

Table 5: Quantitative results of the ablation study. The first column indicates the specific constraint disabled in each configuration.

Disabled Constraint	$f_{scale} \uparrow$	$f_{var} \downarrow$	$f_{center} \downarrow$	$\mathcal{L}_{overlap} \uparrow$	$\mathcal{L}_{align} \downarrow$
Layout Boundary	0.8983	0.0005	0.0024	0.3772	0.6658
Non-overlapping for Identical Weights	0.3945	0.0000	0.0000	3.5553	0.0537
Max Overlap Limit	0.2777	0.0319	0.0001	1.7614	0.0380
Critical Region Protection	0.2557	0.0011	0.0021	0.9666	0.0557
Text Space Allocation	0.2239	0.0007	0.0004	0.9487	0.0487
Group Alignment	0.2361	0.0187	0.0000	0.9556	0.0451
None (Full Model Proposed)	0.2113	0.0039	0.0024	0.9457	0.0486

makes it difficult to perceive the original shape of lesions, leading to reduced readability.

4.8 Limitations

One limitation of our method is that enforcing group alignment constraints may become difficult when a semantic group contains a large number of elements. In such cases, to fit all elements into the

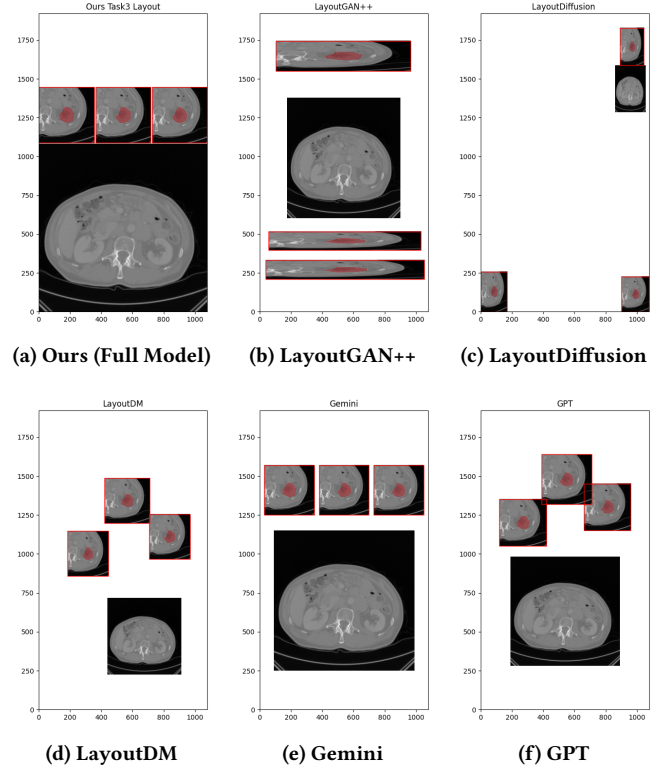


Figure 8: Qualitative comparison for Kidney. (b) achieves good center alignment among semantically related elements but suffers from incorrect aspect ratios. (c) shows randomly placed elements with distorted aspect ratios. (d) shows randomly placed elements. (e) achieves a relatively structured layout but could benefit from slightly larger element scaling. (f) shows misalignment among semantically related elements.

available 2D space, each element must be scaled down, which leads to reduced readability. As a result, when the number of elements within a semantic group is large, a trade-off must be made between preserving group alignment and maintaining readability. This may require either relaxing the alignment constraint or reducing the number of elements in the group.

Another limitation is that the proposed framework requires users to manually set element weights and layout boundaries. Although these parameters provide flexibility and allow the framework to handle different application domains without retraining, users may still need to redefine element weights and adjust the width and height of the 2D space according to task-specific requirements.

5 CONCLUSION

We show that layout synthesis can be defined as a Mixed-Integer Linear Programming (MILP) problem and efficiently solved using an optimization solver. This allows us to generate optimized layouts while satisfying both space utilization and visual clarity constraints. In addition, our approach achieves faster layout generation

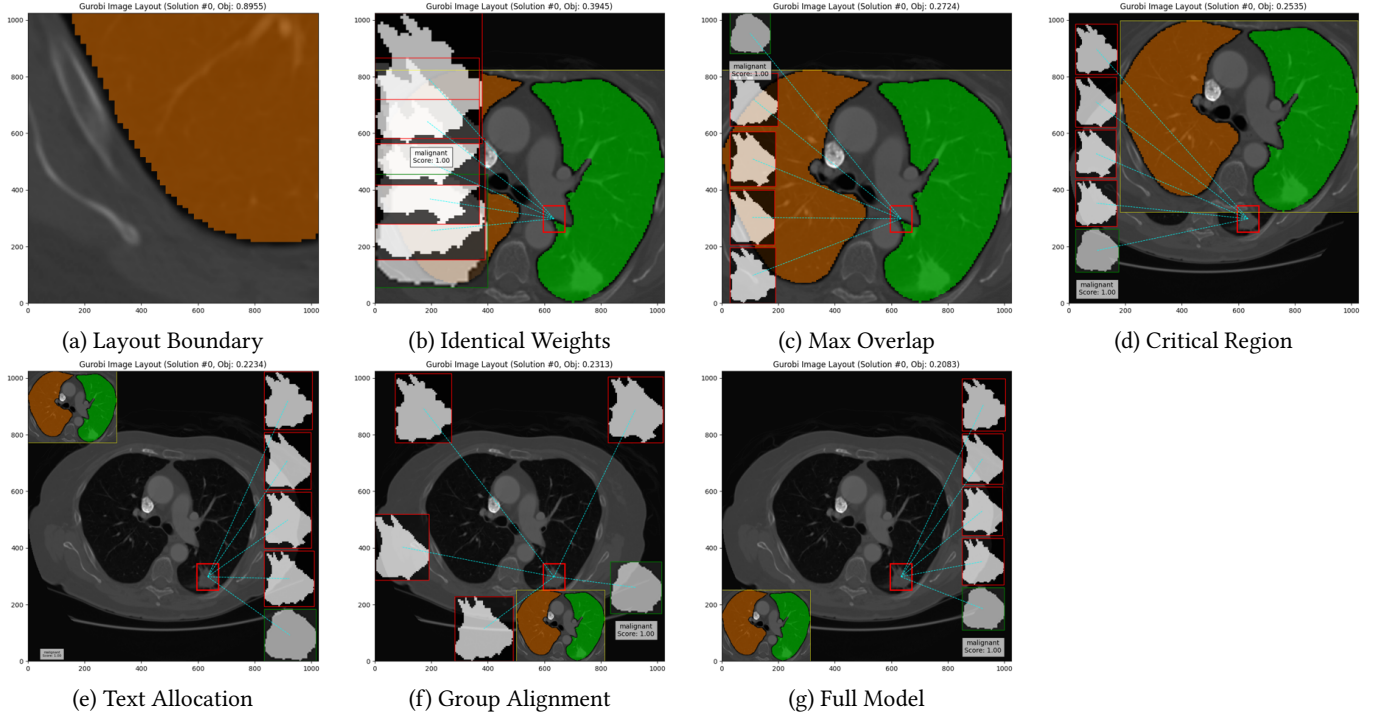


Figure 9: Qualitative ablation study of constraint removal. Each subfigure illustrates the effect of disabling one constraint in the MILP formulation.

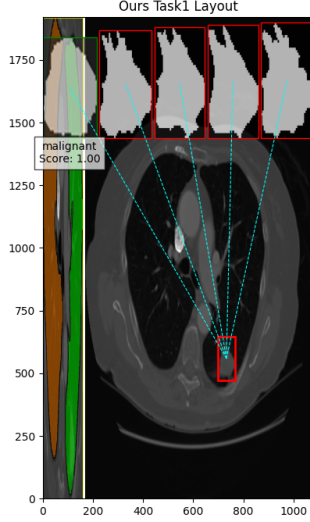


Figure 10: Non-uniform scaling of our method for Lung CT.

compared to existing methods, and does not require any training process.

REFERENCES

- [1] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric K Hoffman, et al. 2011. The lung image database consortium

- (LIDC) and infectious disease research institute (IDRI) node classification reference database: an evaluation of medical imaging data annotation. *Medical Physics* 38, 2 (2011), 915–931.
- [2] Biplob Deka, Zifeng Huang, Chad Franzen, Joshua Hibschan, Daniel Afargan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*. 845–854.
- [3] Vemund Fredriksen, Svein Ole M Seivle, Andre Pedersen, Thomas Langø, Gabriel Kiss, and Frank Lindseth. 2022. Teacher-student approach for lung tumor segmentation from mixed-supervised datasets. *Plos one* 17, 4 (2022), e0266147.
- [4] Gurobi Optimization, LLC. 2026. Gurobi Optimizer Reference Manual. <https://www.gurobi.com>
- [5] Johannes Hofmanninger, Forian Prayer, Jeanny Pan, Sebastian Röhrich, Helmut Prosch, and Georg Langs. 2020. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European radiology experimental* 4, 1 (2020), 50.
- [6] Naoto Inoue, Kotaro Kikuchi, Edgar Simo-Serra, Mayu Otani, and Kota Yamaguchi. 2023. Layoutdm: Discrete diffusion model for controllable layout generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10167–10176.
- [7] Kotaro Kikuchi, Edgar Simo-Serra, Mayu Otani, and Kota Yamaguchi. 2021. Constrained graphic layout generation via latent optimization. In *Proceedings of the 29th ACM International Conference on Multimedia*. 88–96.
- [8] Hongwei Bran Li, Fernando Navarro, Ivan Ezhov, Amirhossein Bayat, Dhritiman Das, Florian Kofler, Suprosanna Shit, Diana Waldmannstetter, Johannes C Paetzold, Xiaobin Hu, et al. 2024. QUBIQ: Uncertainty quantification for biomedical image segmentation challenge. *arXiv preprint arXiv:2405.18435* (2024).
- [9] Junyi Zhang, Jiaqi Guo, Shizhao Sun, Jian-Guang Lou, and Dongmei Zhang. 2023. Layoutdiffusion: Improving graphic layout generation by discrete diffusion probabilistic models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7226–7236.
- [10] Xu Zhong, Jianbin Tang, and Antonio Jimeno Yepes. 2019. Publaynet: largest dataset ever for document layout analysis. In *2019 International conference on document analysis and recognition (ICDAR)*. IEEE, 1015–1022.