Qiang YT, Fu YW, Yu X *et al.* Learning to generate posters of scientific papers by probabilistic graphical models. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 34(1): 155–169 Jan. 2019. DOI 10.1007/s11390-019-1904-1

Learning to Generate Posters of Scientific Papers by Probabilistic Graphical Models

Yu-Ting Qiang[1], Yan-Wei Fu[2], *Member, ACM, IEEE*, Xiao Yu[1], Yan-Wen Guo[1,*], *Member, IEEE*
Zhi-Hua Zhou[1], *Fellow, CCF, ACM, IEEE*, and Leonid Sigal[3], *Member, ACM, IEEE*

[1] *National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210046, China*
[2] *School of Data Science, Fudan University, Shanghai 200433, China*
[3] *Disney Research Pittsburgh, Pittsburgh 15241, U.S.A.*

E-mail: qiangyuting.new@gmail.com; yanweifu@fudan.edu.cn; mg1533099@smail.nju.edu.cn
ywguo.nju@gmail.com; zhouzh@nju.edu.cn; lsigal@disneyresearch.com

Received January 14, 2018; revised November 12, 2018.

Abstract Researchers often summarize their work in the form of scientific posters. Posters provide a coherent and efficient way to convey core ideas expressed in scientific papers. Generating a good scientific poster, however, is a complex and time-consuming cognitive task, since such posters need to be readable, informative, and visually aesthetic. In this paper, for the first time, we study the challenging problem of learning to generate posters from scientific papers. To this end, a data-driven framework, which utilizes graphical models, is proposed. Specifically, given content to display, the key elements of a good poster, including attributes of each panel and arrangements of graphical elements, are learned and inferred from data. During the inference stage, the maximum a posterior (MAP) estimation framework is employed to incorporate some design principles. In order to bridge the gap between panel attributes and the composition within each panel, we also propose a recursive page splitting algorithm to generate the panel layout for a poster. To learn and validate our model, we collect and release a new benchmark dataset, called NJU-Fudan Paper-Poster dataset, which consists of scientific papers and corresponding posters with exhaustively labelled panels and attributes. Qualitative and quantitative results indicate the effectiveness of our approach.

Keywords graphical design, layout automation, probabilistic graphical model

1 Introduction

The emergence of a large number of scientific papers in various academic fields and venues (conferences and journals) is noteworthy. For example, ArXiv, a premiere on-line scientific repository, reports upload rate of over 9 000 papers and reports a month in 2016[①]. It is time-consuming to read and digest all of these papers for researchers, particularly those interested in holistically assessing the state-of-the-art, or understanding just core scientific ideas explored in the last year. Converting a scientific paper into a poster provides an important way to efficiently and coherently convey core ideas and findings of the original paper.

To achieve this goal, it is therefore essential to keep the posters readable, informative and visually aesthetic. It is challenging, however, to design a high-quality scientific poster which meets all of the above design principles, particularly for novice researchers who may not be proficient at design tasks or familiar with design tools (e.g., Adobe Illustrator). In general, poster design is a complicated and timeconsuming task; both understanding of the paper content and experience in design work are required.

Regular Paper

A preliminary version of the paper was published in the Proceedings of AAAI 2016.

This work was supported by the Natural Science Foundation of Jiangsu Province of China under Grant No. BK20150016, the National Natural Science Foundation of China under Grant Nos. 61772257 and 61672279, and the Fundamental Research Funds for the Central Universities of China under Grant No. 020214380042.

*Corresponding Author

① https://arxiv.org/year/start/16, Dec. 2018.

Automatic tools for scientific poster generation would help researchers by providing them with an easier way to effectively share their research. Further, given a large amount of scientific papers on ArXiv and other on-line repositories, such tools may also provide a way for other researchers to consume the content more easily. Instead of browsing raw papers, they may be able to browse automatically generated poster previews (potentially constructed with their specific preferences in mind).

Page layout generation[1−3] has been popular in recent years with the goal of generating graphical design layout, such as photo collage[4], furniture object arrangements[5,6], comics panel layouts[7] and so on. These studies pay more attention on visual aesthetics than informativeness and readability. On the other hand, there are also lots of studies that investigate presentation layout automation[8−10], which aim at document generation. These studies often focus on microtypography problems such as line breaking, margins inference and so on. In addition, some studies utilize templates as input to their layout algorithms[11].

In general, in order to generate a scientific poster in accordance with, and representative of, the original paper, many problems need to be solved.

1) *Content Extraction.* Both important textual and graphical contents need to be extracted from the original paper.

2) *Panel Layout.* The extracted content from each section should fit each panel; besides, the shape and position of each panel should be optimized for readability and design appeal.

3) *Graphical Elements* (*Figure and Table) Arrangement.* Within each panel, textual content can typically be sequentially presented, but for graphical elements, their size and placement should be carefully considered. Due to these challenges, to our knowledge, no automatic tool for scientific poster generation exists.

In this paper, we propose a data-driven method for automatic scientific poster generation (given a corresponding paper). Content extraction and layout generation are two key components in this process. For content extraction, we use TextRank[12] to extract textual content, and provide an interface for extraction of graphical content (e.g., figures, tables). Our approach focuses primarily on poster layout generation and we address this problem in three steps. First, we propose a probabilistic graphical model to infer panel attributes. Second, we introduce a tree structure to represent panel layout, based on which we further design a recursive algorithm to generate new layouts. Third, in order to synthesize layout within each panel, we train another probabilistic graphical model to infer the attributes of graphical elements.

To the best of our knowledge, this paper presents the first method for scientific poster generation from the original academic papers. A preliminary version of this work appeared as a conference paper[13]. This paper extends the previous version in the following perspectives.

1) *Enlarged Dataset.* We have enlarged and released our dataset② to the community as a new benchmark dataset for evaluating the problem of scientific poster generation.

2) *Improved Methodology.* We improve our method in several ways. First, we propose a novel loss function to evaluate the panel arrangement, which helps our algorithm to find better panel layouts. Second, we refine the probabilistic graphical model framework for element composition within each panel, and this refinement takes some design principles into consideration and makes our approach more effective.

3) *Additional Experiments.* We provide more detailed performance analysis and extensive experiments to show the effectiveness of the new method.

The remainder of this paper is organized as follows. The related work is briefly introduced in Section 2. In Section 3, we describe our dataset and preprocessing work in detail. In Section 4 and Section 5, we present a high-level overview and key components of our method respectively. Experiments and evaluation are discussed in Section 6. Section 7 concludes the paper.

2 Related Work

In this section, we review three heavily studied topics of layout generation, i.e., general graphical design (Subsection 2.1), comic layout generation (Subsection 2.2) and presentation layout automation (Subsection 2.3), and the differences between these topics and our task of scientific poster generation.

2.1 General Graphical Design

Graphical design has been studied extensively in computer graphics community. This involves several related, yet different topics. Geigel and Loui[4] made use of genetic algorithm[14,15] for photo album layout,

②https://drive.google.com/open?id=1N5AL3VSezYcXDjqgv4NjLLe7VEZ73NSF, Nov. 2018.

which addresses the placement of each photo in an album. Yu *et al.*[5] automatically synthesized furniture objects arrangements using simulated annealing algorithm. In contrast, Merrell *et al.*[6] applied some simple design guidelines to solve a similar problem. Other graphical design problems such as interface design[16], circuit board layout[17], and graph layout[18] have also been studied. These studies often present an optimization framework along with some design guidelines to synthesize and evaluate plausible layouts.

Nevertheless, these studies are concerned more about graphical elements (e.g., photo, furniture), and they take visual aesthetics as the highest priority. In contrast, for scientific poster generation, textual content, original paper structure, and the order of contents need to be considered to ensure the readability of a scientific poster.

2.2 Comic Layout Generation

Due to the popularity of comics, many related research topics, such as manga retargeting[19], comic episodes generation[20] and manga-like rendering[21], have drawn considerable research attention in computer graphics community. Particularly, several techniques have been studied to facilitate layout generation. For example, Arai and Herman[22] and Pang *et al.*[23] studied how to automatically extract each panel from e-comics and display e-comics on different devices. In order to convert conversational videos to comics, Jing *et al.*[24] made use of a rule-based optimization scheme for layout generation. Cao *et al.*[25] presented a generative probabilistic framework to arrange input artworks into a manga page, and then used optimization techniques to refine it. Furthermore, Cao *et al.*[7] took text balloons and picture subjects into consideration for manga layout generation to guide readers' attention. However, in our poster generation, one has to consider both texts and graphical elements composition within each panel, which has not been discussed previously.

Our panel layout generation method is partially inspired by the recent work on manga layout[25]. We use a binary tree to represent the panel layout. By contrast, Cao *et al.*[25] trained a Dirichlet distribution to sample a split configuration, and different Dirichlet distributions for different kinds of instance have to be trained as a result. Instead, we propose a recursive algorithm to search for the best split configuration along a binary tree. Similar to our panel layout splitting strategy, previous studies on 2D packing[26] and floorplanning[27] also try to split a chip/floor using vertical or horizontal lines, and have been applied to draw tag-cloud[28]. However, 2D packing problems aim at minimizing the space waste, which is totally different with the goal of our panel layout problem. On the other hand, floorplanning tries to fill the floor using rectangular items, but it ignores the order of different items, which would affect the readability when applied to scientific poster generation.

2.3 Presentation Layout Automation

The emergence of data and information that we need to present, challenges our ability to present them manually; thus, automated layout of presentations is becoming increasingly important[8]. For automated document formatting, early work, such as [9, 10], focuses largely on line breaking, paragraph arrangement and some other micro-typography problems. A common way to solve these problems is modeling it as a constrained optimization problem[29]. More recent studies pay attention to presentation document layout. Jacobs *et al.*[30] presented a grid-based dynamic programming method to select a page layout template. Damera-Venkata *et al.*[11] made use of Probabilistic Document Model (PDM) to facilitate document layout. By contrast, we focus on both macro-typography problems (e.g., panel layout) and microtypograph (e.g., graphical elements size decision) in this paper. Additionally, rather than use simple design guidelines as previous work[9,10], we learn our layout generating model from the annotated training datasets.

Another piece of related work is called single page graphical design[3], which made use of an energy-based model derived from design principles for graphic design layout. However, they regard texts as a rectangle block rather than text flow, which is inappropriate for scientific poster generation. Harrington *et al.*[31] described a measure of document aesthetics, and an aesthetics driven layout engine was proposed in [32]. However, these approaches do not put constraints on the ordering of content, which is clearly important for scientific poster generation.

3 NJU-Fudan Paper-Poster Dataset

In this paper, we propose a new research topic of learning to generate posters of scientific papers. According to our observation, a typical scientific poster usually follows some general design principles. The whole poster is often divided into several distinct panels and each panel usually includes several bullet points and sentences that explain the corresponding bullet point. Each bullet point often corresponds to a subsection or a paragraph in the paper. Important figures and tables in each paper section would also be included in the corresponding poster panel. Fig.1 shows such an example of human designed poster[33]. This type of scientific poster is readable, informative and visually aesthetic since it considers both the structure and key

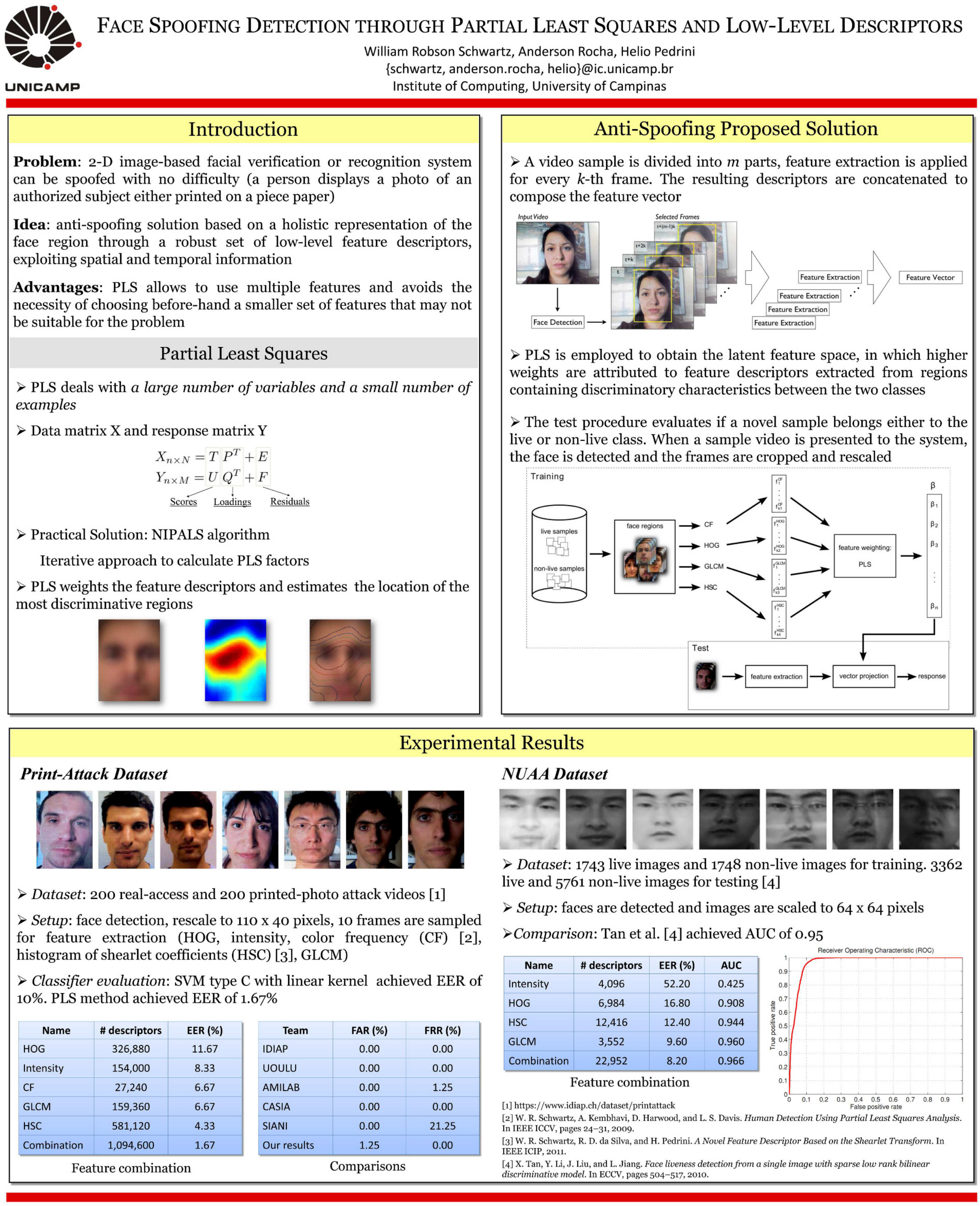


Fig.1. Example of human designed poster.

messages conveyed by the original paper, which makes it easy for readers to understand.

To further study the tasks of poster generation for scientific papers, we introduce an NJU-Fudan Paper-Poster dataset which contains pairs of scientific posters and their corresponding papers. A total of 85 computer science research paper-poster pairs were collected from an online website.

We further annotate the meta information for each paper-poster to facilitate the research of this topic. For each poster, we label both layout attributes (e.g., panel position, figure size) and content attributes (e.g., text length in each panel). In the corresponding paper, layout related information (e.g., figure size in original paper) is also manually labelled. We also provide annotation tool which can enable the annotation and labeling of further data.

4 Method Overview

Overview. To generate a readable, informative and aesthetic poster, we simulate the rule-of-thumb on how researchers design posters in practice. We generate the panel layout for a scientific poster first, and then arrange the textual and the graphical elements within each panel. As shown in Fig.2, the framework overall has four steps, namely, content extraction, panel attributes inference, panel layout generation, and composition within each panel.

Problem Formulation. We formally introduce the problem of learning to generate posters of scientific papers before developing our contributions to each section. We have a set of posters M and their corresponding scientific papers. Each poster $m \in M$ includes a set of panels P_m, and each panel $p \in P_m$ has a set of graphical elements (figures and tables) G_p. Each panel p is characterized by six attributes:

- text length within a panel (l_p);
- text ratio (t_p), text length within a panel relative to text length of the whole poster, $t_p = l_p / \sum_{q \in P_m} l_q$;
- number of graphical elements within a panel (n_p);
- graphical elements ratio (g_p), the size of graphi-

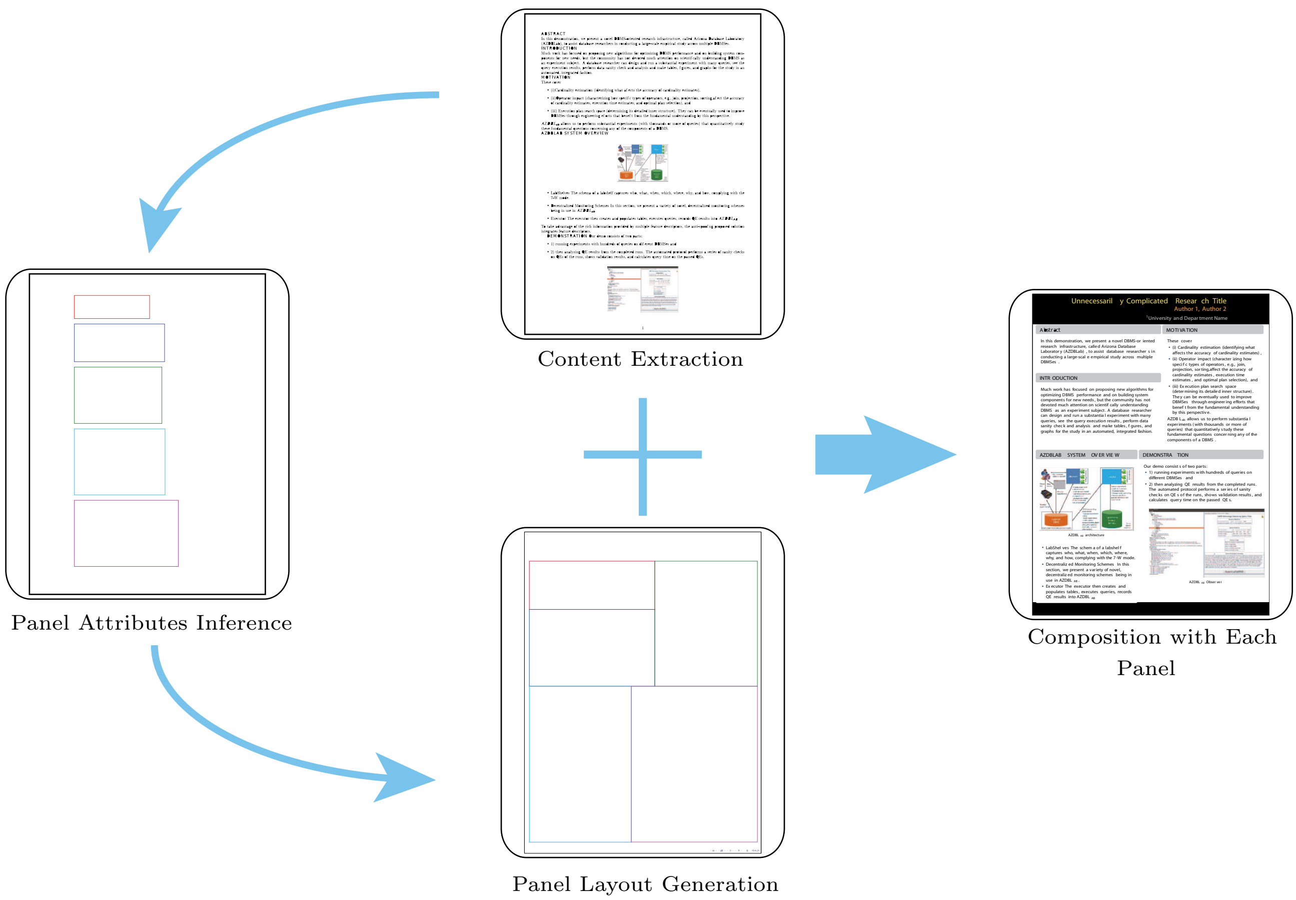


Fig.2. Overview of our proposed approach.

cal elements within a panel relative to the total size of graphical elements in the poster (note that there is a little difference between g_p and t_p; here instead of predicting the fixed figure size in poster, we directly use the corresponding figure from the original paper);

• panel size (s_p) and aspect ratio (r_p), $s_p = w_p \times h_p$ and $r_p = w_p/h_p$, where w_p and h_p denote the width and the height of a panel with respect to the poster.

Each graphical element $g \in G_p$ has four attributes:

• graphical element size (s_g) and aspect ratio (r_g), $s_g = w_g \times h_g$ and $r_g = w_g/h_g$, where w_g and h_g denote the width and the height of a graphical element relative to the whole paper respectively;

• horizontal position (h_g), inspired by the way how latex beamer makes poster, we arrange that panel content sequentially from top to bottom; hence only relative horizontal position needs to be considered, which is defined by a discrete variable $h_g \in \{left, center, right\}$;

• graphical element size in poster (u_g), the ratio of the width of the graphical element to the width of the panel it belongs to.

To learn how to generate a poster, our goal is to determine the above attributes for each panel $p \in P_m$ and each graphical element $g \in G_p$, as well as the arrangement of the panels.

Intuitively, a trivial solution is to use a learning model (e.g., support vector regression (SVR)) to learn how to regress these attributes, including s_p, r_p, u_g, and h_g, while regarding attributes which can be known according to corresponding scientific paper (i.e., t_p, g_p, l_p, r_g, and s_g) as features. However, such a solution takes these features as a whole, and thereby lacks an insight mechanism for exploring the relationships between specific attributes (e.g., s_p and g_p). It may fail to meet the requirements of readability, informativeness, and aesthetics. We thus propose a Bayesian network to characterize the relationships among these attributes, where the Bayesian network is trained on the paper-poster dataset we collected. Then according to the Bayesian network we trained, we can infer the layout attributes by using likelihood-weighted sampling method.

5 Our Methodology

In this section, we will further explain each step of our framework as illustrated in Fig.2. Particularly, 1) in Subsection 5.1, we extract from the paper the text content and the graphical content. The textual content can be summarized by the textual summary algorithms; and the graphical content (figures and tables) usually occupies a rectangular area of the poster, and would be extracted by user interactions. 2) Inference of the key attributes for initial panel (such as panel size s_p and aspect ratio r_p) is then conducted by learning a probabilistic graphical model from the training data in Subsection 5.2. 3) Furthermore, Subsection 5.3 synthesizes panel layout by developing our recursive algorithm to further update these key attributes and generate an informative and aesthetic panel layout. 4) Finally, we compose these panels by utilizing our graphical algorithm to further synthesize the visual properties of each panel (such as the size and the position of graphical elements) in Subsection 5.4.

5.1 Content Extraction

Content extraction, which includes both textual content extraction and graphical content extraction, is the first step in our proposed scientific poster generation system.

For textual content, we employ the state-of-the-art textual summary algorithm to summarize the content of each section. In particular, we use TextRank[12].

For graphical content, our algorithm will parse the key meta data of the layout (i.e., width and height) of each figure and table. To better select the most important figures/tables, we add user interaction here to rank the importance of the tables and figures.

5.2 Panel Attributes Inference

In our proposed approach, we assume that each section of the original scientific paper should be represented by one rectangular panel, which should not only be of an appropriate size to contain the textual and graphical content of each corresponding section, but also be in a reasonable shape (aspect ratio) to maximize visual aesthetic appearance.

To enable such a goal, we learn a Bayesian network to infer the initial size and aspect ratio for each panel. As each panel is composed of both textual description and graphical elements, we assume that panel size (s_p) and aspect ratio (r_p) are conditionally dependent on text ratio t_p, the number of graphical elements n_p, and graphical element ratio g_p. Therefore, we define the joint probability of a set of panels P as,

$$Pr(P|T, N, G) = \prod_{p \in P} Pr(s_p|t_p, n_p, g_p)Pr(r_p|t_p, n_p, g_p),$$

where $T = \{t_p|p \in P\}$, $N = \{n_p|p \in P\}$, and $G = \{g_p|p \in P\}$ denote attributes sets. $Pr(s_p|t_p, n_p, g_p)$

and $Pr(r_p|t_p, n_p, g_p)$ are conditional probability distributions (CPDs) of s_p and r_p respectively given t_p, n_p and g_p. We further model them as two conditional linear Gaussian distributions:

$$Pr(s_p|t_p, n_p, g_p) = N(s_p; \boldsymbol{w}_s(t_p, n_p, g_p, 1)^{\mathrm{T}}, \sigma_s),$$
$$Pr(r_p|t_p, n_p, g_p) = N(r_p; \boldsymbol{w}_r(t_p, n_p, g_p, 1)^{\mathrm{T}}, \sigma_r),$$

where t_p and g_p are defined by the content extraction step demonstrated in Fig.2; $\boldsymbol{w}_s$ and $\boldsymbol{w}_r$ are parameters that leverage the influence of various factors; σ_s and σ_r are the variances. The parameters ($\boldsymbol{w}_s$, $\boldsymbol{w}_r$, σ_s and σ_r) are estimated using maximum likelihood (ML) estimator.

Note that in order to learn from limited data, this step actually employs two assumptions: 1) s_p and r_p are conditionally independent; 2) the attribute sets for panels are independent.

We need the panels to be neither too small or large in size (s_p), nor too distorted in aspect ratio (r_p), to ensure a readable, informative and aesthetic poster. The two assumptions introduced here are sufficient for this task since the attribute values estimated in this step are just good initial values for each panel. We use the next two steps to further relax these assumptions and discuss the relationship between s_p and r_p, as well as the relationship among different panels.

To ease exposition, we denote the set of panels as $P = \{(s_{p_1}, r_{p_2}), (s_{p_2}, r_{p_2}), \cdots, (s_{p_k}, r_{p_k})\}$, where s_{p_i} and r_{p_i} are the size and the aspect ratio of the i-th panel p_i, respectively, and $|P| = k$.

5.3 Panel Layout Generation

One conventional way to design posters is to simply arrange them in two or three columns style. This scheme, although simple, makes posters designed in this way look similar. Inspired by manga layout generation[25], we propose a more vivid panel layout generation method. Specifically, we arrange the panels with a binary tree structure to help represent the panel layout. The panel layout generation is then formulated as a process of recursively splitting of a page, as illustrated in Fig.3. The first split is vertical with the split ratio (0.5, 0.5). The poster is further divided into three panels in the left, and two panels in the right. This makes the whole page as two equal columns. For the left column, we resort to a horizontal split with the split ratio (0.33, 0.67). The larger one is further horizontally divided into two panels with the split ratio (0.5, 0.5). We only split the right column once, with the split ratio (0.5, 0.5).

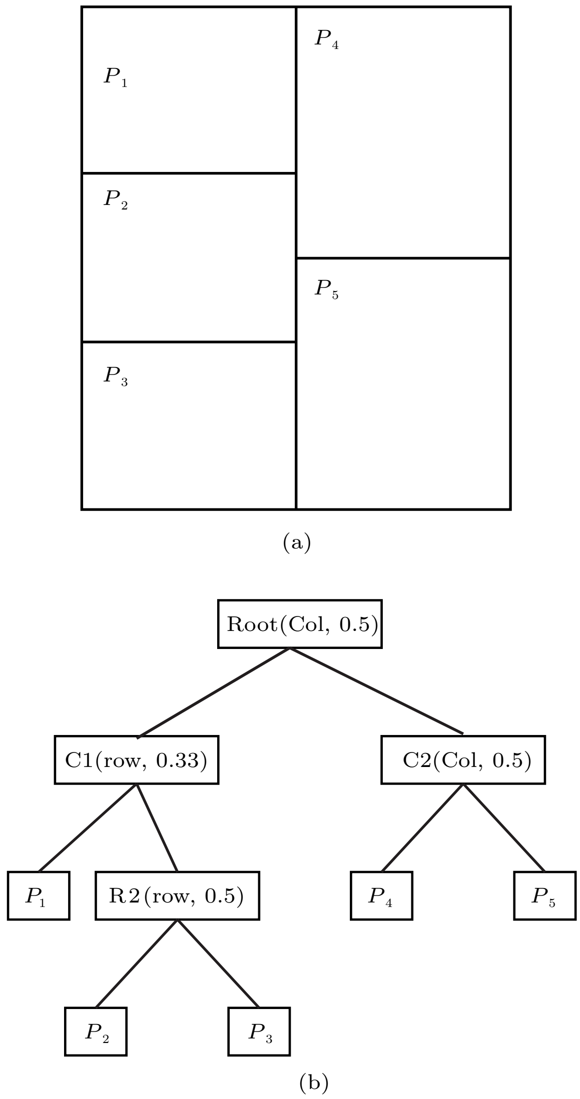


Fig.3. Example of panel layout and the corresponding tree structure. (a) Panel layout. (b) Tree structure.

Conveying information is the most important goal for a scientific poster; thus we attempt to maintain the relative size for each panel during panel layout generation. This motivates the following loss for the panel shape variation,

$$l_{\mathrm{var}}(p_i) = |r_{p_i} - r'_{p_i}|, \qquad (1)$$

where r'_{p_i} is the aspect ratio of a panel after optimization.

On the other hand, we also evaluate the aesthetic for the split configuration. In our approach, the split configuration is composed of several splits. Each split divides a set of panels into two panels, and the split ratio is decided by the ratio of the total area of the two parts of panels. Since balance is an important guideline for design work[3], we evaluate the aesthetic for the panel layout configuration based on the symmetry of each partition. In particular, if a panel set is divided by a split ζ_i as $p_1, p_2, \cdots, p_k$ and $p_{k+1}, p_{k+2}, \cdots, p_m$, then the aesthetic loss for this split is defined as follows:

$$l_{\text{aes}}(\zeta_i) = \alpha|\sum_{i=1}^{k} s_{p_i} / \sum_{i=1}^{m} s_{p_i} - 0.5|. \tag{2}$$

The loss for panel shape variation ((1)) and split configuration ((2)) leads to a combined loss for the panel layout arrangement:

$$Loss(P, P', Z) = \sum_{i=1}^{k} l_{\text{var}}(p_i) + \sum_{\zeta \in Z} l_{\text{aes}}(\zeta), \tag{3}$$

where P' is the panel set after optimization and Z is the set of splitting steps.

In each splitting step, the combinatorial choices for splitting positions can be recursively computed and compared with respect to the loss function ((3)) above and we choose the panel attributes with the lowest loss. The whole algorithm is summarized in Algorithm 1.

Algorithm 1 . Panel Layout Generation

Require:
panels which we learned from graphical model: $L = \{(s_{p_1}, r_{p_1}), (s_{p_2}, r_{p_2}), \cdots, (s_{p_k}, r_{p_k})\}$;
rectangular page area x, y, w, h
Ensure: none

if $k == 1$ **then**
  Adjust panels[0] to adapt to the whole rectangular page area, return the aesthetic loss: $|r_{p_0} - w/h|$;
else
  for each $i \in [1, k-1]$ **do**
    $t = \sum_{j=1}^{i} s_{p_j} / \sum_{j=1}^{n} s_{p_j}$;
    $Loss_1 = PanelArrangement((s_{p_1}, r_{p_1}), \cdots, (s_{p_i}, r_{p_i}), x, y, w, h \times t)$;
    $Loss_2 = PanelArrangement((s_{p_{i+1}}, r_{p_{i+1}}), \cdots, (s_{p_k}, r_{p_k}), x, y + h \times t, w, h \times (1-t))$;
    if $Loss > Loss_1 + Loss_2 + \alpha|t - 0.5|$ **then**
      $Loss = Loss_1 + Loss_2 + \alpha|t - 0.5|$;
      Record this arrangement;
    end if
    $Loss_1 = PanelArrangement((s_{p_1}, r_{p_1}), \cdots, (s_{p_i}, r_{p_i}), x, y, w \times t, h)$;
    $Loss_2 = PanelArrangement((s_{p_{i+1}}, r_{p_{i+1}}), \cdots, (s_{p_k}, r_{p_k}), x + w \times t, y, w \times (1-t), h)$;
    if $Loss > Loss_1 + Loss_2 + \alpha|t - 0.5|$ **then**
      $Loss = Loss_1 + Loss_2 + \alpha|t - 0.5|$;
      Record this arrangement;
    end if
  end for
end if
return loss and arrangement

5.4 Composition Within a Panel

Having inferred the layout of panels, we turn our attention to the composition of raw contents within each panel. Generally, each panel in a scientific poster is composed of textual and graphical content. Considering the readability of a scientific poster, each panel can be filled by these contents sequentially. However, for aesthetic consideration, the horizontal position and the size of each graphical element need to be specified carefully. Therefore, we pose automated panel composition as an inference problem in a Bayesian network that incorporates some design constraints.

Designing the composition for each panel is complicated, because both panel attributes and raw contents need to be considered. We aim at designing a Bayesian network to characterize how these variables interact with each other. Given the placement of each graphical element, textual contents can be filled into the panel sequentially; therefore, the composition of a panel can be defined by the horizontal position (h_g) and the size (s_g) of each graphical element. In our approach, the layout within each panel is composed by first sampling random variable h_g representing the choice of horizontal position (*left, right, center*), and then sampling variable s_g representing the size of a graphical element.

In our Bayesian network, horizontal position (h_g) of a graphical element relies on both the shape (r_p) of the panel which the element belongs to and attributes (r_g, s_g) of the element itself. For example, a portrait figure is more likely to be presented in the left or right of a landscape panel. To describe such relationship, the horizontal position h_g of a graphical element g in panel p is sampled from a soft-max function,

$$Pr(h_g = i | r_p, r_g, s_g) = \frac{e^{\boldsymbol{w}_{hi} \cdot (r_p, r_g, s_g, 1)^{\mathrm{T}}}}{\sum_{j=1}^{H} e^{\boldsymbol{w}_{hj} \cdot (r_p, r_g, s_g, 1)^{\mathrm{T}}}}, \tag{4}$$

where $H = 3$ is the cardinality of the value set of h_g, and $\boldsymbol{w}_{hi}$ is the i-th row of $\boldsymbol{w}_h$.

The size of a graphical element (u_g) has to meet two requirements: 1) it needs to be appropriate to fill the panel; 2) it also needs to harmonize with the occupation of the graphical element in the original paper. To this end, in our model, the size of each graphical element (u_g) is governed by both the panel attributes (l_p, s_p) and each element's own properties (s_g, h_g). We may sample the size of each graphical element from the conditional linear Gaussian distribution,

$$\begin{aligned} & Pr(u_g | s_p, l_p, s_g, h_g) \\ = \ & N(u_g | \boldsymbol{w}_u \cdot (s_p, l_p, s_g, h_g, 1)^{\mathrm{T}}, \boldsymbol{\sigma}_u), \end{aligned} \tag{5}$$

where $\boldsymbol{w}_u$ is the parameter to balance the influence of different factors, and $\boldsymbol{\sigma}_u$ represents variance.

For a set of graphical elements G which belongs to the same panel p, the probability of sampling process described above is simply the product of the probabilities of all design choices made during the sampling process, and it can be represented by the following distribution,

$$\begin{aligned} &Pr(h_G, u_G|s_p, r_p, l_p, s_G, r_G) \\ = &\prod_{g\in G} Pr(h_g|r_p, r_g, s_g)Pr(u_g|s_p, l_p, s_g, h_g), \end{aligned} \quad (6)$$

where h_G and u_G denote the assignments of horizontal position and the size for all graphical elements in G, respectively; s_G and r_G represent the input attributes of G.

Learning. The goal of the learning stage in this step is to estimate the parameters in our Bayesian network from training data, and this can be done by maximizing the complete-data log likelihood since all the random variables in our model are observed. For conditional linear Gaussian distribution ((5)), with some algebraic manipulation we can compute the optimal ML estimate of $\boldsymbol{w}_u$ and $\boldsymbol{\sigma}_u$ in a closed form:

$$\begin{aligned} \boldsymbol{w}_u^* &= (\sum_i^n x^{(i)}x^{(i)\mathrm{T}})^{-1}(\sum_i^n u_g^i x^i), \\ \boldsymbol{\sigma}_u^* &= \frac{1}{n}\sum_{i=1}^n (u_g^{(i)} - \boldsymbol{w}_u^{*\mathrm{T}} x^{(i)})^2, \end{aligned} \quad (7)$$

where $x^{(i)} = (s_p, l_p, s_g, h_g, 1)^{(i)}$ denotes the training data. For soft-max function ((4)), while there is no known closedform ML solution, we can resort to an iterative optimization algorithm — iteratively reweighted least squares (IRLS) algorithm.

The Bayesian network described above models the relationship between different variables explicitly. However, it is also desirable to consider the relationship between panel size and content occupation. In a human designed poster, contents usually fill each panel up exactly, which makes the poster seem clean and informative. Therefore, we incorporate the design principles with our Bayesian network, and our goal is to find solution to this function:

$$\begin{aligned} h_G^*, u_G^* &= \underset{h_G, u_G}{\operatorname{argmax}} f(h_G, u_G|s_p, r_p, l_p, s_G, r_G) \\ &= \lambda_1 \log Pr(h_G, u_G|s_p, r_p, l_p, s_G, r_G) + \\ &\quad \lambda_2 \log N(w_p \times h_p|\beta t_p + \sum_{g\in p} s_g, \rho). \end{aligned} \quad (8)$$

In (8), the first term is defined in (6). It is a likelihood that determines how well the solution fits our Bayesian network. The second term measures how well the contents fit the panel size, and it assigns high probability if the contents fill the panel precisely and low probability for deviations from the ideal.

Since the exact MAP inference is not tractable in our model, we perform approximate inference by using likelihood-weighted sampling method[34].

6 Experimental Results

6.1 Experimental Setup

NJU-Fudan Paper-Poster Dataset. Our dataset includes 85 well-designed pairs of scientific papers and their corresponding posters, which are selected from 600 publicly available pairs we have collected. These papers are all about computer science topics, and their posters have relatively similar design styles. We further annotate panel attributes, such as panel width, panel height and so on. The annotated meta data is saved into an XML file.

Implementation Details. The input content to our scientific poster generation approach is also specified in an XML file. This file specifies the structure and contents of a scientific paper, including chapters, sections, paragraphs, and graphical elements. The other attributes such as caption and key words are also saved in the corresponding content block. Note that the equation and formulas are taken as normal texts since they can be written in latex format. For graphical elements, we only save the width and the height in the XML file. In our experiment, sections and subsections correspond to panels and bullets respectively. We get textual content from XML file and use TextRank to get summarization. In order to give different importance of different sections, we can set different extraction ratio for each of them. This will result in important sections generating more content and hence occupying bigger panels. For simplicity, this paper uses equal important weights for all sections. The Bayesian Network Toolbox (BNT)[34] is used for key parameters estimation and sampling. For graphical element attributes inference, we generate 1 000 samples by the likelihood weighted sampling method[35] for (8). With the inferred metadata, the final poster is generated in latex Beamerposter format with Lankton theme.

Competitors and Evaluation Metrics. We compare several baselines on different sections of our model to evaluate the methods of attributes inference. Particularly, we compare ridge regression, regression tree, support vector regression (SVR) with linear kernel and RBF kernel respectively. And for graphical elements position (h_g) inference, we regard it as a classification problem, and then compare the performance of our method with that of the nearest neighbors classification (KNN), decision tree, support vector classification (SVC) with linear and RBF kernel. We employ the corresponding values for the original (human) designed posters as the ground-truth. We split the dataset into 80 pairs for training and validation, and the rest (5 pairs) for testing.

Comparison with Human Designed Posters. We then evaluate how well our approach facilitates scientific poster generation, compared with novice designers and the original poster (which is designed by the author). We invite three second-year Ph.D. students, who are not familiar with our project, to hand design posters for the test set. These three students work in computer vision and machine learning and have not yet published any papers on these topics; hence they are novices to research. Given the test set papers, we ask the students to work together and design a poster for each paper.

Running Time. Our framework is very efficient in terms of running cost. Our experiments are done on a PC with an Intel® Xeon® 3.6 GHz CPU and 11.6 GB RAM. Table 1 shows the average time we need for each step. The total running time is significantly less than the time experienced designers require to design a good poster, and it is also less than the time spent to generate the posters made by three novices.

Table 1. Running Time of Each Step

Step	Stage	Average Time (s)
Text extraction		9.236 2
Panel attributes inference	Learning stage	0.330 0
	Inferring stage	0.004 0
Panel layout generation		0.001 0
Composition within panel	Learning stage	0.570 0
	Inferring stage	0.913 0

6.2 Quantitative Evaluation

Effectiveness of Attribute Inferences. To validate the effectiveness of this step, our model is compared against several state-of-the-art regression methods, including ridge regression, regression tree, linear support vector regression (SVR), and RBF-SVR.

The results are shown in Table 2. We use the panel attributes of original posters as the ground-truth and root-mean-square error (RMSE) is computed for the inferred size and aspect ratio of each panel. Specifically, we use the design of original poster as the ground-truth and the RMSE is computed as,

$$RMSE = \sqrt{\sum_{i=1}^{n}(s_p - s'_p)/n}, \quad (9)$$

where s_p represents the panel size of original panel, s'_p represents the panel size inferred by learning model, and n indicates the total number of panels of all the posters. In (9), we use s_p as an example; the RMSE for r_p and u_g can be calculated in the same way.

Table 2. Performance of Attributes Inference

Method	Panel Size (s_p)	Panel Aspect Ratio (r_p)	Graphical Element Position (h_g)
Our method	0.071 0	0.695	0.014 4
Ridge regression	0.075 0	0.696	0.289 0
Regression tree	0.009 0	0.819	0.287 0
Linear-SVR	0.073 0	0.702	0.361 0
RBF-SVR	0.120 0	0.737	1.041 0

Note: Here we only consider the relative size of each panel which is normalized into [0, 1]. The lower the value, the better the performance.

To infer the panel size (s_p) and the aspect ratio (r_p), we use the text ratio (t_p) and the graphical elements ratio (g_p) as features. Compared with all the other methods, the RMSE of our method is only 0.71 and 0.695 respectively, which is lower than all the other methods. This shows that our algorithm can better estimate the panel attributes than the other methods, due to our probabilistic graphical formulation that effectively models the correlations and dependence among variables.

For graphical elements size (u_g) and horizontal position (h_g), we use s_p, r_p, l_p, s_g, r_g as features and our model is compared against all the other methods. RMSE and accuracy are used to evaluate the performance of each method on u_g and h_g, respectively. The accuracy is computed as

$$Accuracy = \sum_{i=1}^{n} I(h_g, h'_g)/n,$$

$$I(h_g, h'_g) = \begin{cases} 1, \text{ if } & h_g = h'_g, \\ 0, \text{ otherwise}, \end{cases}$$

where h_p represents the horizontal position in the original panel, and h'_p represents the horizontal position inferred by the learning model. As shown in Table 2 and Table 3, our results beat all those other methods since design constraints are introduced in the inference stage by (8).

Table 3. Accuracy of Horizontal Position Prediction

Method	Graphical Element Size (u_g) (%)
Our method	88.9
KNN	66.7
Decision tree	66.7
Linear-SVC	72.2
RBF-SVC	72.2

Note: The higher the value, the better the performance.

6.3 Qualitative User Study Evaluation

User Study. User study is employed to compare our results with original posters and posters made by novices. We invite 10 researchers (who are experts on the evaluated topic) to evaluate these results on readability, informativeness, and aesthetics. Each researcher is sequentially shown the three results generated (in randomized order) and asked to score the results from 0 to 10, where 0, 5 and 10 indicate the lowest, middle and highest scores of corresponding metrics respectively. The final results are averaged across subjects. Note that since our method mainly considers the layout of a poster, we provide novice designers and our method with contents as same as the original poster. We argue that this is a more objective way to evaluate our method because both texts extracted by TextRank and novice designers may not be so good as the text in original poster which is summarized by the authors of the paper, and different contents would affect poster layout evaluation.

In Table 4, on readability and informativeness, our result is comparable to the original poster, and it is significantly better than posters made by novices. This validates the effectiveness of our method. On the one hand, the inferred panel attributes and the generated panel layout will save most valuable and important information. Besides, the composition within each panel inferred by our method would give proper emphasis on figures and tables, which may be overlooked by novice designers. In contrast, our method is lower than the original posters on aesthetics metric (yet, still higher than those from novices). This is reasonable because aesthetics is a relatively subjective metric and it generally needs to involve lots of human interactions. Human designers can adjust the poster layout via lots of latex commands again and again. In general, it is an open problem to generate more aesthetic posters from papers.

Table 4. User Study of Different Posters Generated

Method	Readability	Informativeness	Aesthetics	Average
Our method	7.32	7.08	6.70	7.03
Posters by novices	6.82	6.80	6.58	6.73
Original posters	7.36	7.10	7.44	7.30

Qualitative Evaluation of Three Methods. We qualitatively compare our results (Fig.4(b) and Fig.4(e)) with the posters from novices (Fig.4(a) and Fig.4(d)) and the original posters (Fig.4(c) and Fig.4(f)). All of them are for the same paper and with same contents.

It is interesting to show that when compared with the panel layout of original poster, our panel layout looks more similar to the original one than the one by novices. This is due to that, firstly, the Paper-Poster dataset has a relatively similar graphical design with high quality, and secondly our split and panel layout algorithm works well to simulate the way how people design posters. In Figs.4(a)–4(c), we can see that in order to arrange contents in the poster aesthetically, the order of each panel is rearranged in the poster from the novice designer (Fig.4(a)), and this would affect the readability of a poster. Figs.4(d)–4(f) show that, compared with novice designers, our method also achieve good performance on attributes inference for graphical elements. The size of graphical elements inferred by our method seems similar to that of the original poster. In contrast, the poster designed by novices in Fig.4(d) loses emphasis on figures in order to keep the content fit each panel.

6.4 Qualitative Evaluation by Design Principles

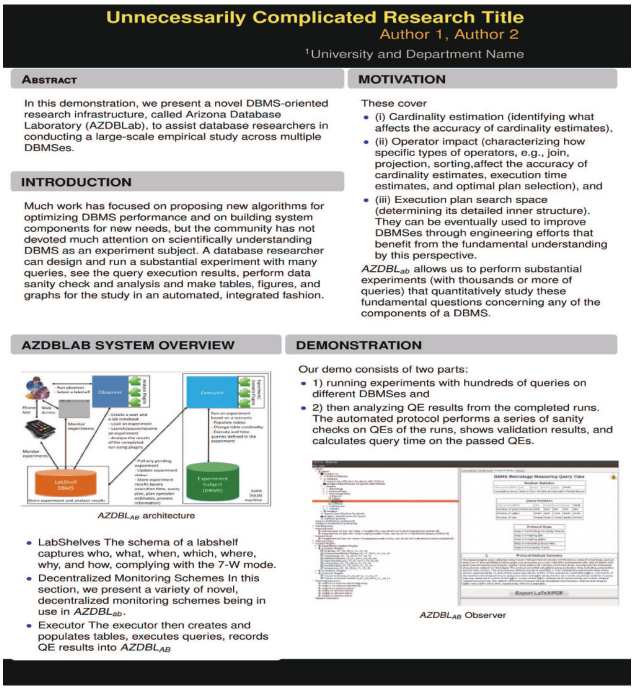
We further qualitatively evaluate our results (Fig.5) by the general graphical design principles[3], i.e., flow, alignment, and overlap and boundaries.

Flow. It is essential for a scientific poster to present information in a clear read-order, i.e., readability. People always read a scientific poster from left to right and from top to bottom. Since Algorithm 1 recursively splits the page of poster into left and right, or top and bottom, the panel layout we generate ensures that the read-order matches the section order of original paper. Within each panel, our algorithm also sequentially organizes contents which also follow the section order of the original paper and this improves the readability.

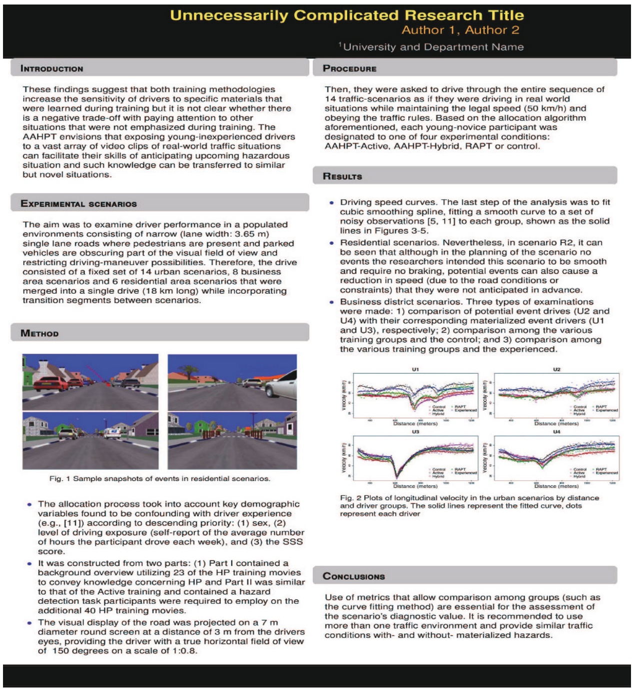
Alignment. Compared with the complex alignment constraint in [3], our formulation is much simpler and

Fig.4. Results generated by different ways. (a)(d) Designed by novices. (b)(e) Our result. (c)(f) Original posters[36,37].

uses an enumeration variable to indicate the horizontal position of graphical elements h_g. This simplification does not spoil our results which still have reasonable alignment as illustrated in Fig.5 and quantitatively evaluated by three metrics in Table 4.



(a)



(b)

Fig.5. Example of our results. (a) Our result 1. (b) Our result 2.

Overlap and Boundaries. Overlapped panels will make the poster less readable and less aesthetic. To avoid this, our approach 1) recursively splits the page for panel layout; 2) sequentially arranges panels; 3) incorporates a design constraint into our Bayesian network ((8)) to penalize the cases of overlapping between graphical elements and panel boundaries. As a result, our algorithm can achieve reasonable results without significant overlapping and/or crossing boundaries. Similar to the manually created posters (Fig.4(c)), our result (e.g., Fig.4(b)) does not have significantly overlapped panels and/or boundaries.

7 Conclusions

Automatic tools for scientific poster generation are important for poster designers. Designers can save a lot of time with these kinds of tools. Design is a hard work, especially for scientific posters, which require careful consideration of both utility and aesthetics. Abstract principles about scientific poster design cannot help designers directly. In contrast, we proposed an approach to learning design patterns from existing examples, and this approach can be used as an assistant tool for scientific poster generation to aid the designers.

As the future work, our framework can be also applicable to directly learn the general design patterns such as the web-page design and single-page graphical design, given the corresponding layout styles. Currently, we do not consider font types of posters which will be addressed in future.

References

[1] Jahanian A, Liu J, Tretter D R, Lin Q, Damera-Venkata N, O'Brien-Strain E, Lee S, Fan J, Allebach J P. Automatic design of magazine covers. In *Proc. IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics*, January 2012, Article ID. 83020.

[2] Hunter A, Slatter D, Greig D. Web-based magazine design for self publishers. In *Proc. IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics*, January 2011, Article ID. 787902.

[3] O'Donovan P, Agarwala A, Hertzmann A. Learning layouts for single-page graphic designs. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20(8): 1200-1213.

[4] Geigel J, Loui A. Using genetic algorithms for album page layouts. *IEEE Multimedia*, 2003, 10(4): 16-27.

[5] Yu L F, Yeung S K, Tang C K, Terzopoulos D, Chan T F, Osher S J. Make it home: Automatic optimization of furniture arrangement. *ACM Transactions on Graphics*, 2011, 30(4): Article No. 86.

[6] Merrell P, Schkufza E, Li Z, Agrawala M, Koltun V. Interactive furniture layout using interior design guidelines. *ACM Transactions on Graphics*, 2011, 30(4): Article No. 87.

[7] Cao Y, Lau R W, Chan A B. Look over here: Attention-directing composition of manga elements. *ACM Transactions on Graphics*, 2014, 33(4): Article No. 94.

[8] Hurst N, Li W, Marriott K. Review of automatic document formatting. In *Proc. the 9th ACM Symposium on Document Engineering*, September 2009, pp.99-108.

[9] Knuth D E, Plass M F. Breaking paragraphs into lines. *Software: Practice and Experience*, 1981, 11(11): 1119-1184.

[10] Peels A J H, Janssen N J M, Nawijn W. Document architecture and text formatting. *ACM Transactions on Information Systems*, 1985, 3(4): 347-369.

[11] Damera-Venkata N, Bento J, O'Brien-Strain E. Probabilistic document model for automated document composition. In *Proc. the 11th ACM Symposium on Document Engineering*, September 2011, pp.3-12.

[12] Mihalcea R, Tarau P. TextRank: Bringing order into text. In *Proc. the 2004 Conference on Empirical Methods in Natural Language Processing*, July 2004, pp.404-411.

[13] Qiang Y T, Fu Y W, Zhou Y W, Zhou Z H, Sigal L. Learning to generate posters of scientific papers. In *Proc. the 30th AAAI Conference on Artificial Intelligence*, February 2016, pp.51-57.

[14] Holland J H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. A Bradford Book, 1992.

[15] Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning (1st edition). Addison Wesley, 1989.

[16] Gajos K, Weld D S. Preference elicitation for interface optimization. In *Proc. the 18th Annual ACM Symposium on User Interface Software and Technology*, October 2005, pp.173-182.

[17] Sarrafzadeh M, Lee D T. Algorithmic Aspects of VLSI Layout. World Scientific Pub. Co. Inc., 1993.

[18] Battista G D, Eades P, Tamassia R, Tollis I G. Graph Drawing: Algorithms for the Visualization of Graphs (1st edition). Pearson, 1998.

[19] Matsui Y, Yamasaki T, Aizawa K. Interactive manga retargeting. In *Proc. ACM SIGGRAPH 2011 Posters*, August 2011, Article No. 35.

[20] Hoashi K, Ono C, Ishii D, Watanabe H. Automatic preview generation of comic episodes for digitized comic search. In *Proc. the 19th ACM International Conference on Multimedia*, November 2011, pp.1489-1492.

[21] Qu Y, Pang W M, Wong T T, Heng P N. Richness-preserving manga screening. *ACM Transactions on Graphics*, 2008, 27(5): Article No. 155.

[22] Arai K, Herman T. Method for automatic e-comic scene frame extraction for reading comic on mobile devices. In *Proc. the 7th International Conference on Information Technology: New Generations*, April 2010, pp.370-375.

[23] Pang X, Cao Y, Lau R W H, Chan A B. A robust panel extraction method for manga. In *Proc. the 22nd ACM International Conference on Multimedia*, November 2014, pp.1125-1128.

[24] Jing G, Hu Y, Guo Y, Yu Y, Wang W. Content-aware video2comics with manga-style layout. *IEEE Transactions on Multimedia*, 2015, 17(12): 2122-2133.

[25] Cao Y, Chan A B, Lau R W H. Automatic stylistic manga layout. *ACM Transactions on Graphics*, 2012, 31(6): Article No. 141.

[26] Lodi A, Martello S, Monaci M. Two-dimensional packing problems: A survey. *European Journal of Operational Research*, 2002, 141(2): 241-252.

[27] Stockmeyer L. Optimal orientations of cells in slicing floorplan designs. *Information and Control*, 1983, 57(2/3): 91-101.

[28] Kaser O, Lemire D. Tag-cloud drawing: Algorithms for cloud visualization. arXiv:0703109, 2007. https://arxiv.org/abs/cs/0703109, August 2018.

[29] Lok S, Feiner S. A survey of automated layout techniques for information presentations. In *Proc. the 1st International Symposium on Smart Graphics*, March 2001, pp.61-68.

[30] Jacobs C, Li W, Schrier E, Bargeron D, Salesin D. Adaptive grid-based document layout. *ACM Transactions on Graphics*, 2003, 22(3): 838-847.

[31] Harrington S J, Naveda J F, Jones R P, Roetling P, Thakkar N. Aesthetic measures for automated document layout. In *Proc. the 2004 ACM Symposium on Document Engineering*, October 2004, pp.109-111.

[32] Purvis L, Harrington S, O'Sullivan B, Freuder E C. Creating personalized documents: An optimization approach. In *Proc. the 2003 ACM Symposium on Document Engineering*, Nov. 2003, pp.68-77.

[33] Pinto A, Pedrini H, Schwartz W R, Rocha A. Face spoofing detection through visual codebooks of spectral temporal cubes. *IEEE Transactions on Image Processing*, 2015, 24(12): 4726-4740.

[34] Murphy K. The bayes net toolbox for Matlab. Technical Report, Computing Science and Statistics, 2001, http://people.cs.ubc.ca/∼/Papers/bnt.pdf, Nov. 2018.

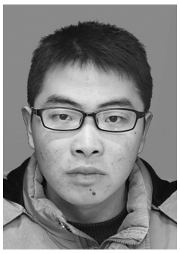
[35] Fung R, Chang K C. Weighing and integrating evidence for stochastic simulation in Bayesian networks. *Machine Intelligence and Pattern Recognition*, 1990, 10: 209-220.

[36] Zhao Y, Zhu S C. Image parsing with stochastic scene grammar. In *Proc. the 25th International Conference on Neural Information Processing Systems*, December 2011, pp.73-81.

[37] Chen Z, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R. Leveraging multi-domain prior knowledge in topic models. In *Proc. the 23rd International Joint Conference on Artificial Intelligence*, August 2013, pp.2071-2077.

Yu-Ting Qiang is a Ph.D. student at the National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing. She received her B.S. degree in software engineering from Jilin University, Changchun, in 2013. Her research interests include computer vision and machine learning.

Yan-Wei Fu received his Ph.D. degree in electric engineering and computer science in Queen Mary University of London, in 2016. He is currently an assistant professor (actually the rank of associate professor) at the School of Data Science, Fudan University, Shanghai. His research includes attribute learning for image/video understanding, life-long learning, open set recognition, open-vocabulary object detection/retrieval, and some multimedia applications such as multi-view video summarization and image segmentation.



Xiao Yu is a Master student of Department of Computer Science and Technology, Nanjing University, Nanjing. He received his B.S. degree in digital media technology from Nanjing University of Posts and Telecommunications, Nanjing, in 2015. His research interests include image processing and computer vision.

Yan-Wen Guo is currently a full professor in the Department of Computer Science and Technology, Nanjing University, Nanjing. He is also affiliated with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing. He received his Ph.D. degree from College of Computer Science and Technology, Zhejiang University, Zhejiang, in 2006. He worked as a visiting professor in the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2006, 2008, and 2009 and a visiting professor in Department of Computer Science, The University of Hong Kong, in 2008, 2012, and 2013, and the University of Illinois at Urbana-Champaign (UIUC) in 2013. His research interests mainly include graphics, image and video analysis and processing, and computer vision.

Zhi-Hua Zhou received his B.Sc., M.Sc. and Ph.D. degrees in computer science from Nanjing University, Nanjing, in 1996, 1998 and 2000, respectively, all with the highest honor. He joined the Department of Computer Science and Technology of Nanjing University as an assistant professor in 2001, and at present he is a professor and the head of the Department of Computer Science and Technology, Standing Deputy Director of the National Key Laboratory for Novel Software Technology, and the founding director of LAMDA (the Institute of Machine Learning and Data Mining) at Nanjing University, Nanjing. His research interests mainly include artificial intelligence, machine learning, data mining, pattern recognition, evolutionary computation and multimedia retrieval, among which machine learning and data mining are his core research area.

Leonid Sigal is an associate professor in the Department of Computer Science at the University of British Columbia, Kelowna. Prior to this, from 2009 to 2017, he was a senior research scientist at Disney Research, Pittsburgh, and an adjunct faculty of Carnegie Mellon University, Pittsburgh. Earlier he was a postdoctoral fellow in the Department of Computer Science at University of Toronto, Toronto. He completed his Ph.D. degree at Brown University (2008), Providence, received his B.Sc. degree in computer science and mathematics from Boston University (1999), Boston, his M.A. degree from Boston University (1999), Boston, and his M.S. degree from Brown University (2003), Providence. From 1999 to 2001, he worked as a senior vision engineer at Cognex Corporation, where he developed industrial vision applications for pattern analysis and verification. Leonid's research interests mainly lie in the areas of computer vision, machine learning, and computer graphics, but also borderline fields of psychology and humanoid robotics.