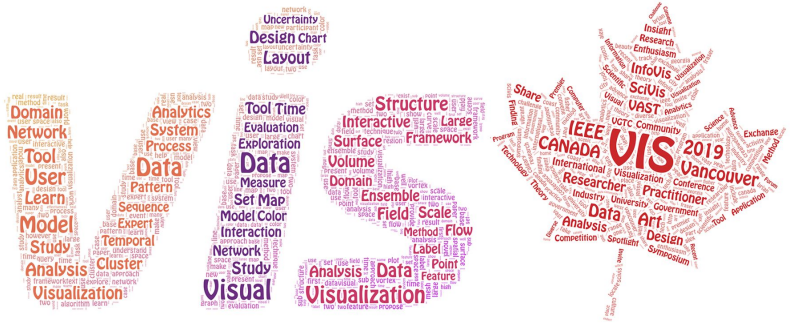
Creating Word Paintings Jointly Considering Semantics, Attention, and Aesthetics

ACM Transactions on Applied Perception, Vol. 19, No. 3, Article 13. Publication date: September 2022.

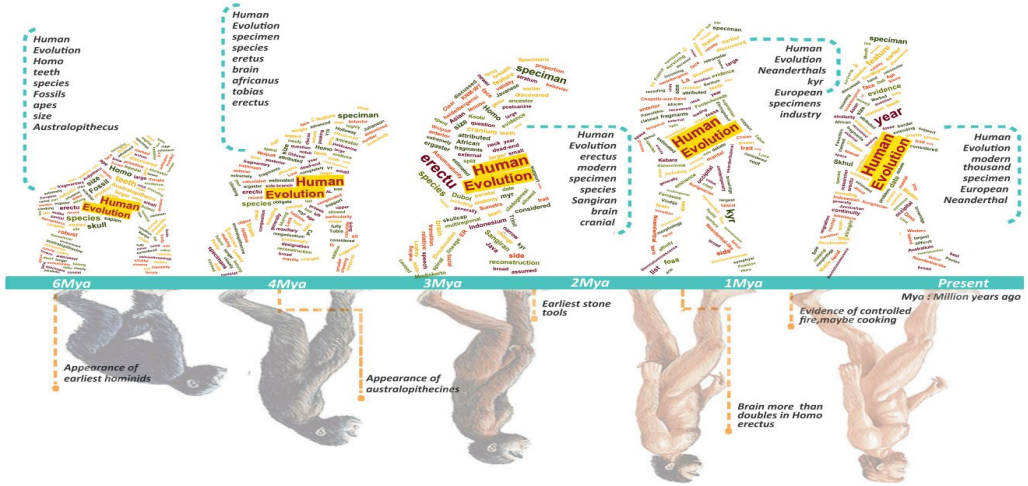
INTRODUCTION

- present a content-aware method for generating a word painting
- This artwork can well preserve visual features and content of a given image, such as color, texture, and semantics
- Traditional word painting typically is manually designed by professional artists with technical precision , which usually involves the following steps: (1) analyzing the **content, structure, and features** of a given image; (2) determining **where** and **what** the assigned word should be; and (3) **optimizing the layout** to create a visually appealing word painting.

RELATED WORK



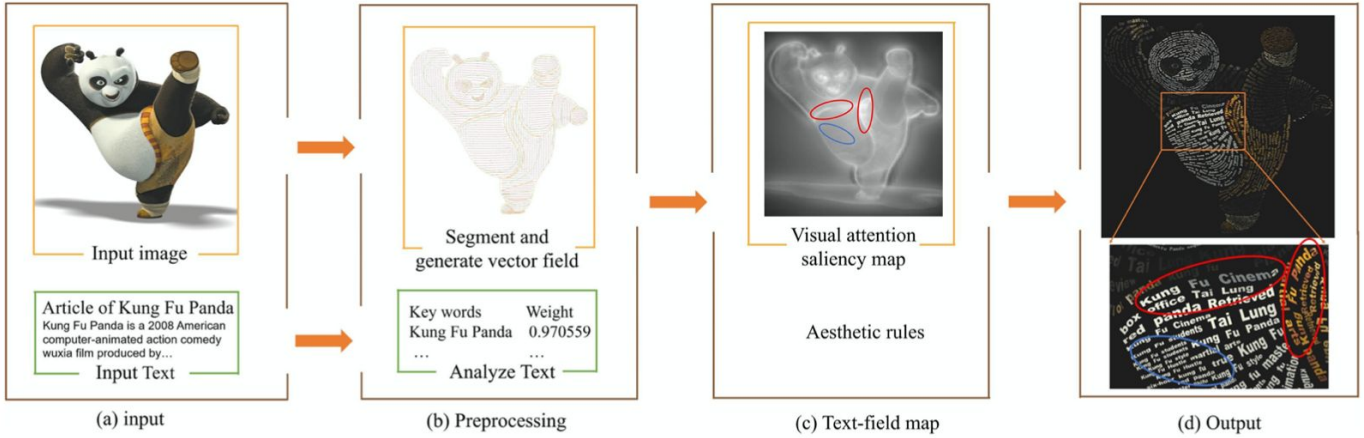
ShapeWordle(VIS 2019)



Morphable Word Clouds for Time-varying Text Data Visualization(TVCG 2015)

Digital micrography(TOG 2011)

METHOD

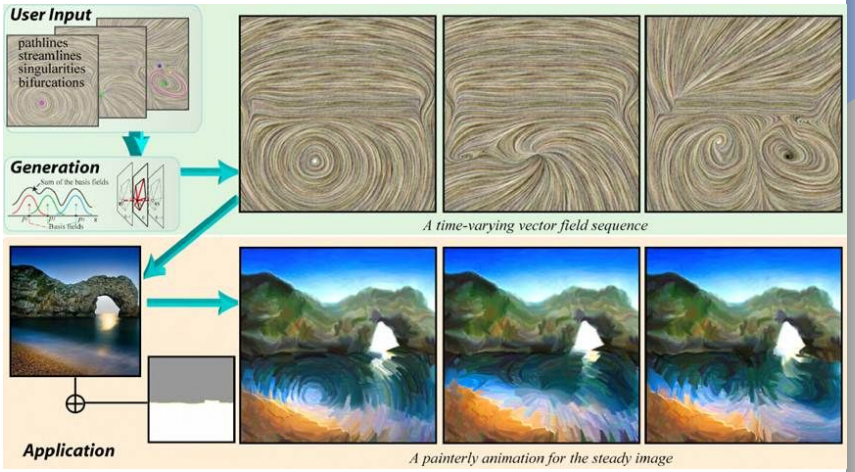


- adopt an **image and its content-related text** as the input
- **segment** the input image, approximate each region with a smooth vector field and extract weighted keywords from the text as graphic elements
- calculate a **saliency map** for the input image and jointly consider visual attention and **aesthetic rules** to optimize the textual layout
- obtain the final content-aware word painting

METHOD

- **Vector Fields Generation and Text Extraction**

Structured Forests for Fast Edge Detection(2013 ICCV)

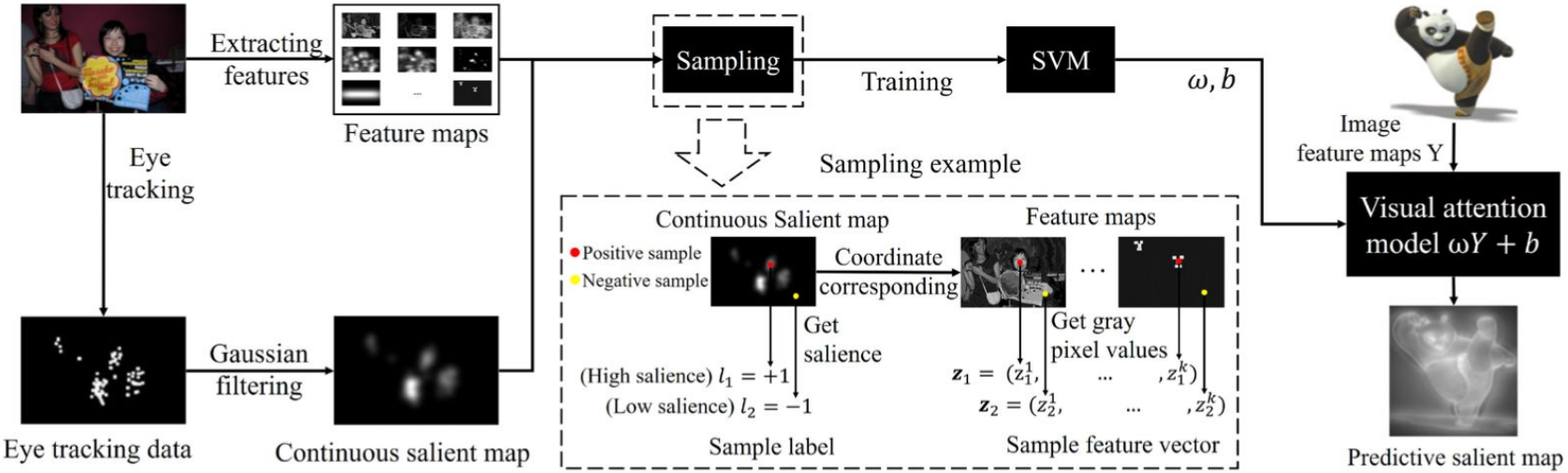


Design of 2D Time-Varying Vector Fields (2011 TVCG)

METHOD

- **Visual Attention Model**

Extract feature maps for each image, such as color, edges, intensity and human faces. Each feature map is a single channel gray map



METHOD

- **Energy-based Optimization Framework**
  - **Visual attention principle.** This principle ensures that the more theme-related keywords are arranged in the more attention-drawing positions with **a larger font size**, such that the viewers can quickly capture the semantic theme of word paintings at the first sight.
  - **Aesthetic rules.** Some rules, such as **visual balance**, **text non-overlap**, and **compactness**, show a stronger relationship to aesthetic response, so we follow these aesthetic rules to achieve a visually pleasing and attractive word painting.

$$\begin{aligned} \min E(x) &= \sum_{k=1}^{6} \alpha_k E_k(x), \\ E_k(x) &\in \{E_{va}(x), E_b(x), E_n(x), E_c(x), E_f(x), E_{vf}(x)\}, \\ s.t. x &\in X, \end{aligned} \tag{1}$$

METHOD

- **Energy-based Optimization Framework**
  - **Visual attention principle**

$$E_{va} = \sum_{i=1}^{n} (\omega_i^s - M(p_i))^2, \qquad (2)$$

METHOD

- **Energy-based Optimization Framework**
  - **Aesthetic Rules**
    - **Visual balance**

$$E_b = \frac{D(\sum_{i=1}^{n} \omega_i^v p_i, c)}{\sqrt{h^2 + w^2}}, \tag{3}$$

    - **Non-overlap**

$$S_p = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} overlap(i,j), \tag{4}$$

    - **Compactness**

$$E_c = \frac{|S_R - \sum_{i=1}^{n} S_i + S_p|}{S_R}, \tag{6}$$

    - **Font size**

$$E_f = \sum_{i=1}^{n} \left(f_i - k \cdot \omega_i^s\right)^2, \tag{7}$$

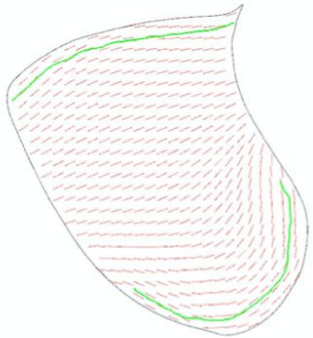
Recommendation system for automatic design of magazine covers(IUI 2013)

METHOD

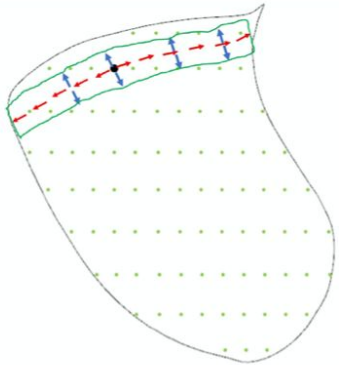
- **Energy-based Optimization Framework**
  - **Aesthetic Rules**
    - **Vector Field**

$$E_{vf} = \sum_{i=1}^{n} (d_i - \Theta(p_i))^2, \qquad (8)$$

(a)



(b)

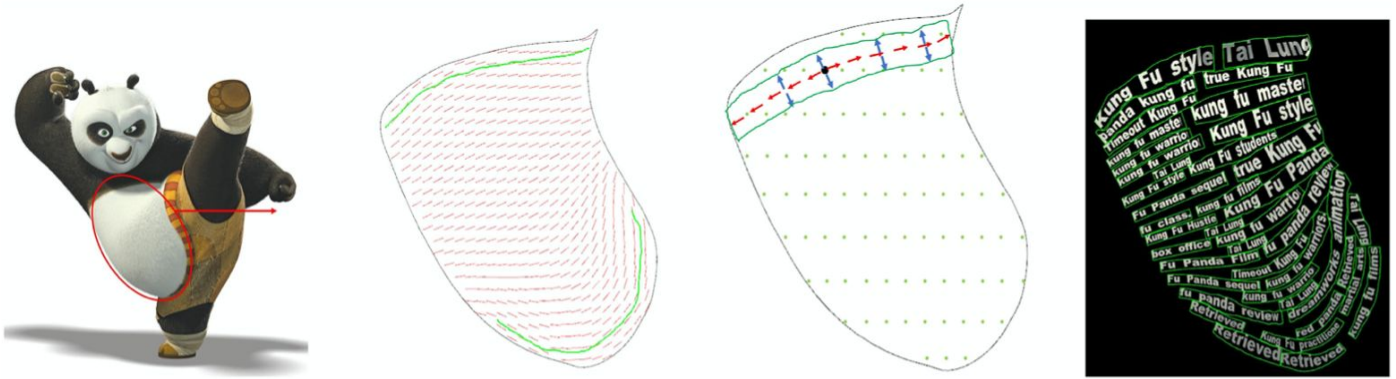


(c)

(d)

METHOD

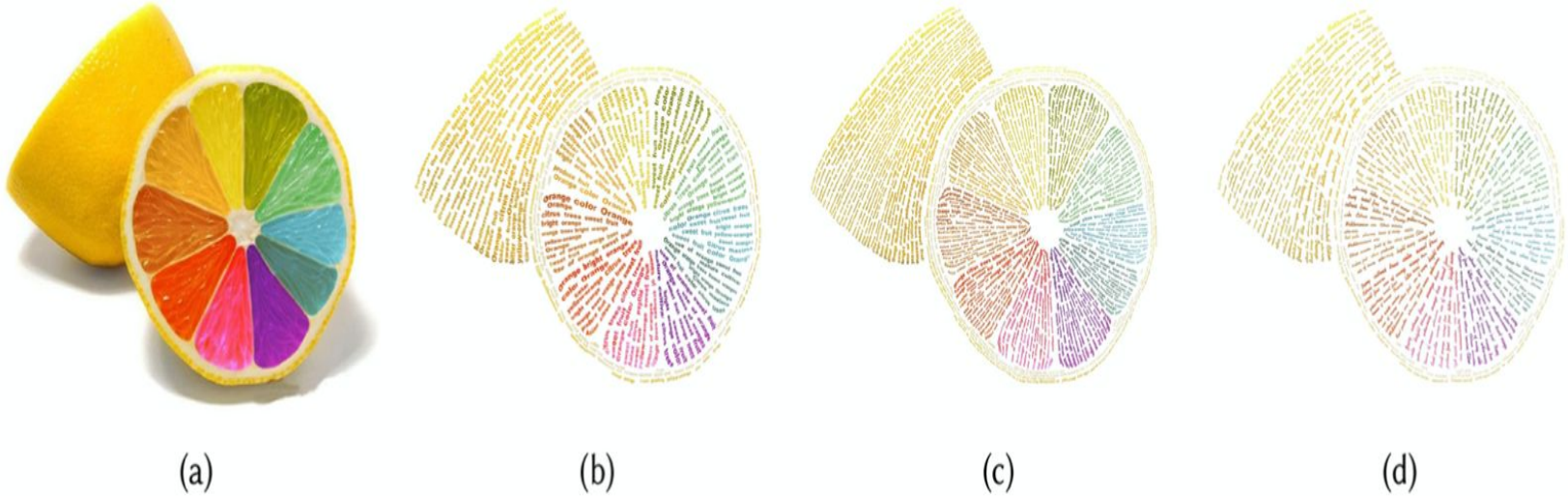
- **Textual Layout**
  - **Dimensionality Determination**
  - **Keywords-patches Mapping**



(a) (b) (c) (d)

RESULTS AND DISCUSSION

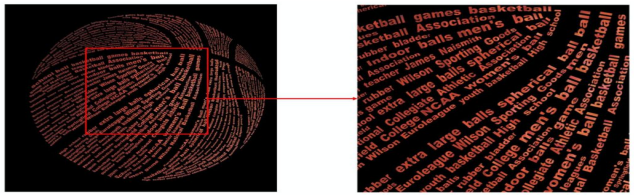
- System Interface
  - With the control panel, users can determine some parameters, such as user-defined keywords, typeface, font size, background color



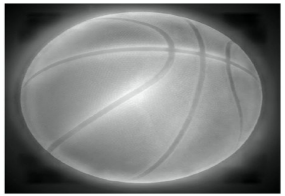
RESULTS AND DISCUSSION

- Effectiveness Evaluation
  - **Preserving image features**: Mark the degree of the generated result in preserving original image features, such as contour and color.
  - **Readability**: Measure the readability of keywords in the generated result, namely, whether the keywords can be recognized clearly and correctly.
  - **Orderliness**: Evaluate the orderliness, continuity, and fluency of the textual layout.
  - **Prominent theme**: Examine the degree of theme-prominence in the generated result.
  - **Overall attractiveness**: Evaluate the overall attractiveness of the generated result.

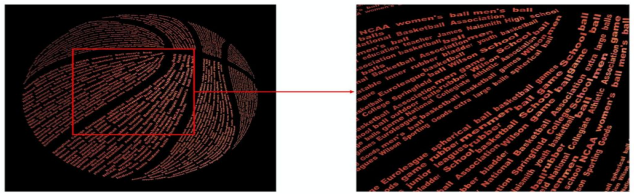
(a) The input image



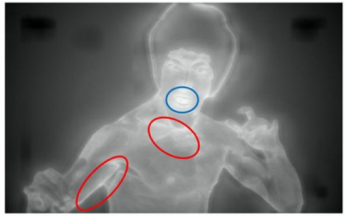
(b) With using the visual attention model

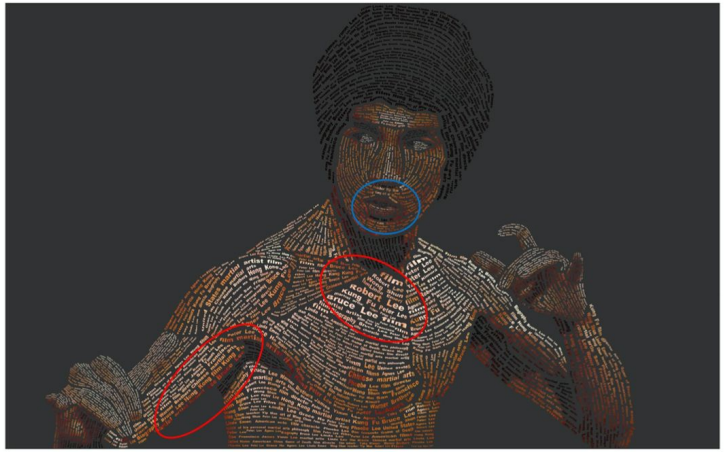


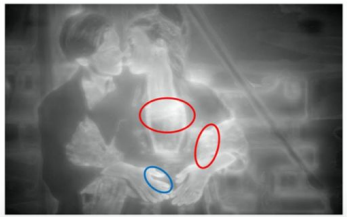
(c) The saliency map



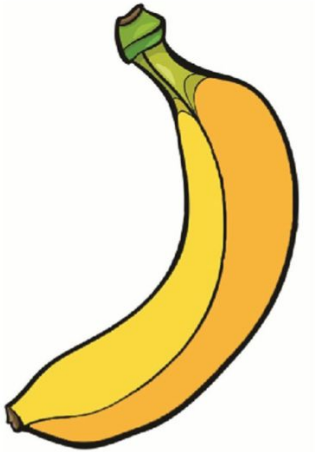
(d) Without using the visual attention model



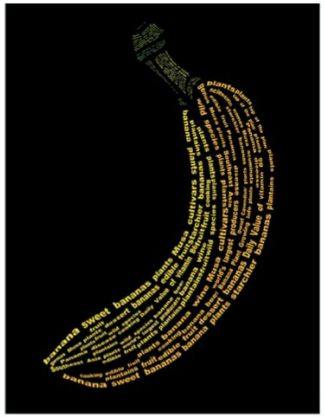

Robert Lee
Bruce Lee



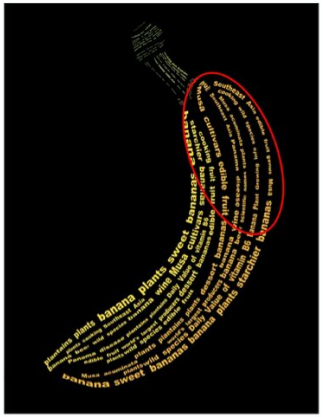

first film disaster film



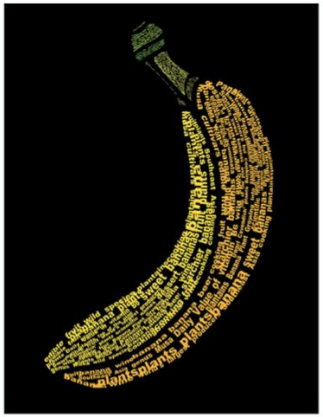
(a) The input image



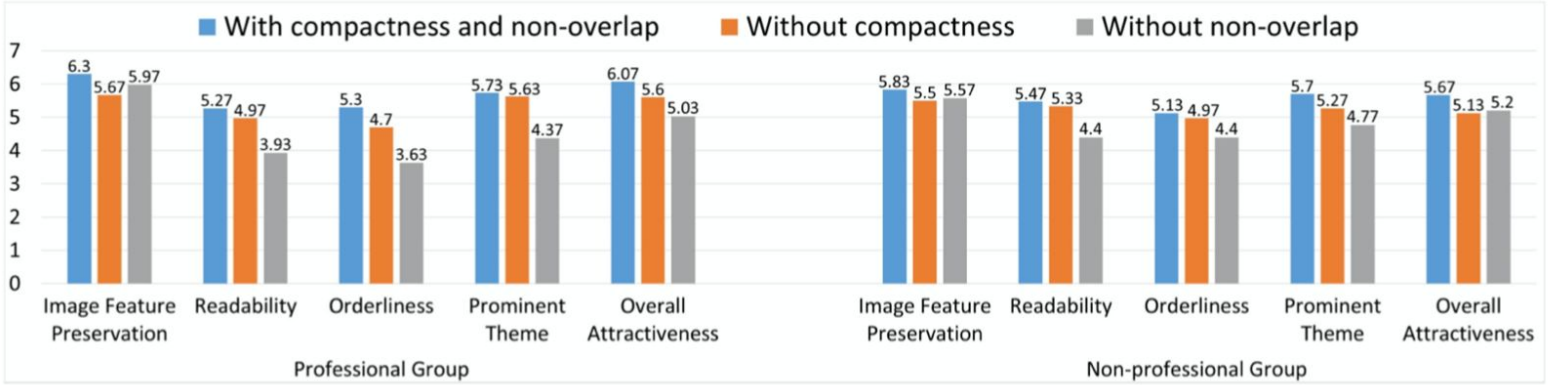
(b) With using both constraints



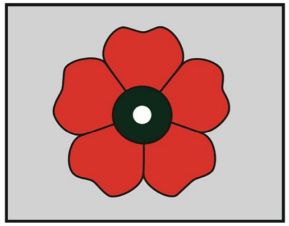
(c) Without using compactness constraint



(d) Without using non-overlap constraint



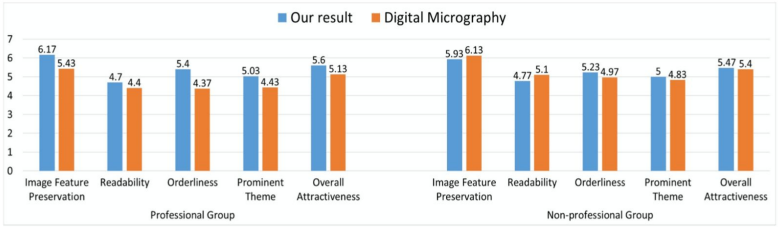
COMPARISONS BETWEEN DIFFERENT METHODS



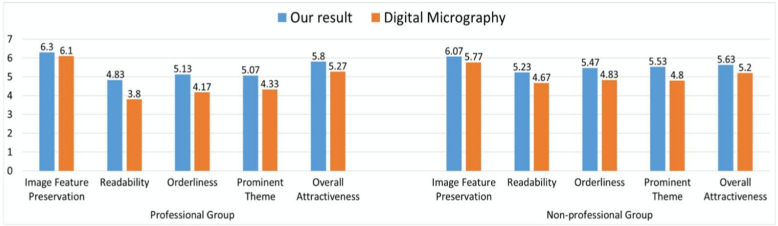
(a) The input images

(b) Digital Micrography

(c) Our method



(a)



(b)

CONCLUSION AND FUTURE WORKS

(1) To some extent, the visual effect of word paintings is **limited by the salient map** generation strategy, while automatic saliency map generation is not always reliable. However, how to generate a reliable salient map is not the focus of this article. Therefore, should there be a trustworthy robust visual attention model consistent with our design goals, the generated content-aware word paintings would be very visually pleasing.

(2) If the input image contains **too many details**, then it may lead to a more complex, **time-consuming task** when creating the word paintings.

(3) Sometimes, a conflict exists between **font size setting and image feature preservation**. For instance, when the most salient positions are located in the segmented regions with image details, accommodating enlarged key- words in these regions would impact the shape features of the input image.

In the future, we will try to construct a data set containing enough word paintings, which is not available at present, and adopt some data-driven methods, such as deep neural network, to generate a visually pleasing word painting.

(a) The input images

(a) The input images