Information Block Detection in Infographic Based on Spatial Proximity and Structural Similarity (Student Abstract)

Jie Lin,[1,2] Xin Wu,[1,2] Jianwei Lu,[1,2] Yi Cai[1,2*]

[1] School of Software Engineering, South China University of Technology, Guangzhou, China
[2] Key Laboratory of Big Data and Intelligent Robot (South China University of Technology), Ministry of Education
{se_jielin, sexinwu, jianweilu}@mail.scut.edu.cn, ycai@scut.edu.cn

Abstract

The infographic is a type of visualization chart used to display information. Existing infographic understanding works utilize spatial proximity to group elements into information blocks. However, these works ignore structural features such as background color and boundary, which results in poor performance towards complex infographics. We propose a Spatial and Structural Feature Extraction model to group elements based on spatial proximity and structural similarity. We introduce a new dataset for information block detection. Experiments show that our model can effectively identify the information blocks in the infographic.

Introduction

The infographic is a type of visualization chart to display information, data, knowledge, etc. Recently, with the emerging demands for automated infographic authoring, several works have been proposed on infographic understanding (Lu et al. 2020) and generating (Chen et al. 2019). As the Gestalt Principles of visual perception (Desolneux, Moisan, and Morel 2004; De Koning et al. 2009) suggested, by grouping elements into blocks, people can speed up the perceiving and authoring process of the infographic. Following the intuition, attempt (Lu et al. 2020) has been made to automatically divide elements into **information blocks** based on spatial proximity, i.e., the closer two elements are, the more likely they belong to the same information block. However, it's common in infographics that two neighboring elements belong to different information blocks, as shown in Figure 1. The model (Lu et al. 2020) based only on spatial proximity may incorrectly group these elements into one block. Structural features (such as background color and boundary) can explicitly indicate the block relationship between elements, thereby avoiding the above grouping error. It's necessary to consider spatial proximity together with the structural features of infographics when detecting information blocks.

Inspired by **Spatial Proximity** and **Structural Similarity** of the Gestalt organization principle in psychology (De Koning et al. 2009), we propose the Spatial and Structural Feature Extraction (SS-FE) model to determine whether two infographic elements are in the same information block. Our model is factorized into a spatial feature extractor and a structural feature extractor. The spatial feature extractor obtains spatial proximity utilizing the relative spatial relationships (neighboring and disjointing) of highlighted element bounding boxes. The structural feature extractor constructs multi-level structural features with a feature pyramid network (FPN). We adaptively fuse spatial and structural features to produce the information block prediction. Besides, we introduce one dataset and conduct experiments on it to validate the effectiveness of our model.

Figure 1: An example of information block detection. People can easily divide the infographic elements into 7 groups (left) according to the background and boundaries of different blocks. By contrast, the model (Lu et al. 2020) incorrectly groups the number 6 and the icon phone together (right) as these elements are the closest, disregarding the boundary splitting them.

Proposed Methodology

Task Definition

The information block detection task is defined as follows: predicting the set of information blocks $S = \{s_1, \ldots, s_k\}$, where $s_k = \{e_{k1}, \ldots, e_{kt}\}, e_{ki} \in E$, according to the infographic image P and elements E. We consider this problem as *paired elements block relationship detection*, i.e., given element pairs in the infographic $T_{pair} = \{c_1, \cdots, c_n\}, c = (e_i, e_j), e_i, e_j \in E, i \neq j$, determine whether each paired elements (e_i, e_j) belong to the same information block.

Spatial and Structural Feature Extraction Model

We propose Spatial and Structural Feature Extraction (SS-FE) model to handle the paired elements block relationship detection problem. As depicted in Figure 2, our model includes two modules: spatial feature extractor and structural feature extractor. These two modules extract the spatial

*Corresponding author: Yi Cai, ycai@scut.edu.cn

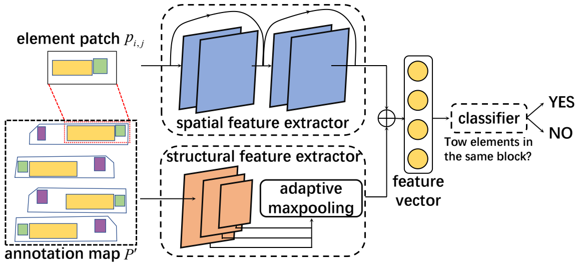

Figure 2: Architecture of the SS-FE model.

proximity and structural similarity of paired elements.

Given the infographic image P and its element set $E = \{e_1, \cdots, e_n\}$, we conduct edge detection on P to highlight the boundaries across elements and possible blocks. We then highlight each element box of E with specific color according to its type to obtain annotation map P'. Given paired elements (e_i, e_j), we crop a patch $p_{i,j}$ from P' to contain them. We take P' and $p_{i,j}$ as the inputs of our model.

Spatial Feature Extractor We use the ResNet model as spatial feature extractor to extract the spatial proximity features $f_p = \mathrm{ResNet}(p_{i,j})$ from element patch $p_{i,j}$. The resent model determines the spatial proximity by extracting the relative positional features (such as overlapping, neighboring, and disjointing) of the two highlighted element boxes.

Structural Feature Extractor Inspired by (Chen et al. 2019), we use feature pyramid network (FPN) as structural feature Extractor. FPN model takes the annotation map P' as input, and obtains structural similarity f_I by extracting the multi-level features of the detected edges in P'. We use adaptive max pooling to retain the most significant structural features while unifying the dimension of the multi-level features, and concatenate the vectors after max pooling:

$$f_I = \left[\sum_{u=1}^{n} \mathrm{maxpool}_{\mathrm{adaptive}}\left(f_u\right) \right], \tag{1}$$

where $f_u = \mathrm{FPN_u}\left(P'\right)$ denotes the feature of the uth layer in top-down part of FPN model, n represents the number of feature pyramid layers, $[\cdot]$ is the concatenation operation.

We integrate the spatial and structural features and use multi-layer perceptron to obtain the information block relationship possibility of paired elements (e_i, e_j): $t_{i,j} = \mathrm{MLP}\left(\mathrm{attention}\left(f_I, f_p, f_p\right)\right)$. Given the outputs of our model $T_{block} = \{(e_i, e_j, t_{i,j})\}$, we choose a specific category of elements $\{e_s\}$ as seeds, and group non-seed element e_v into the same block as e_s according to $t_{s,v}$. The grouping result of T_{block} is the information block prediction S'.

Empirical Experiments and Results

Dataset We annotate infographics in the dataset proposed by Lu et al. 2020. For each infographic element, we label its bounding box, category (*icon, title, text, mark*) and which information block it belongs to. We obtain 1,417 infographics, which are further divided into 149,511 element pairs with ground truth information block relationships.

Model	Precision		Recall		F1 score		block_acc
	same	non	same	non	same	non	
baseline	-	-	-	-	-	-	0.4643
SS-FE \wo st	0.78	0.96	0.80	0.96	0.79	0.96	0.5062
SS-FE	0.85	0.96	0.80	0.97	0.82	0.97	**0.5614**

Table 1: Information block detection results. We conduct paired elements block relationship detection (*same* for in the same block and *non* for the counterpart) on *SS-FE* and its variation, *SS-FE \wo st* (without structural features). The *baseline* (Lu et al. 2020) predicts the information blocks directly without detecting relationships of paired elements.

Evaluation Metric We propose *Jaccard-based* metric $block\_acc$ to evaluate the performance of our model:

$$block\_acc = \frac{1}{n} \sum_{i}^{n} J\left(S_i, S_i'\right), \tag{2}$$

where $J(S_i, S_i')$ is the jaccard similarity of the ground truth S_i and the predicted one S_i' towards ith infographic.

Analysis The performance is illustrated in Table 1. Our model achieves the state-of-the-art performance (0.5614) on $block\_acc$. The $block\_acc$ has dropped about 5% after removing the structural features, indicating that the structural similarity is beneficial to information block detection.

Conclusion

We propose the SS-FE model to effectively detect information blocks with spatial proximity and structural similarity of paired infographic elements. We propose a dataset and an evaluation metric for information block detection.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities, SCUT (No.D2182480), the National Key Research and Development Program of China, the Science and Technology Programs of Guangzhou (No.201802010027, 201902010046), National Natural Science Foundation of China (62076100).

References

Chen, Z.; Wang, Y.; Wang, Q.; Wang, Y.; and Qu, H. 2019. Towards Automated Infographic Design: Deep Learning-based Auto-Extraction of Extensible Timeline. *IEEE TVCG* 26(1): 917–926.

De Koning, B. B.; Tabbers, H. K.; Rikers, R. M.; and Paas, F. 2009. Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review* 21(2): 113–140.

Desolneux, A.; Moisan, L.; and Morel, J.-M. 2004. Gestalt theory and computer vision. In *Seeing, Thinking and Knowing*, 71–101. Springer.

Lu, M.; Wang, C.; Lanir, J.; Zhao, N.; Pfister, H.; Cohen-Or, D.; and Huang, H. 2020. Exploring Visual Information Flows in Infographics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12.