Supplementary Materials for VISAtlas

In this document, we provide supplementary materials for our submission, including the use cases of our system on two more visualization collections, PlotQA[1] and VisImages[2], and the cross validation result on VisImages.

1. APPLICATIONS TO MORE COLLECTIONS

We present applications of our system for comparative analysis of two more visualization collections, including PlotQA[1] and VisImages[2].

- *PlotQA*. PlotQA[1] is a large synthetic collection built to train model to answer questions based on visualization images. It contains more than 30,000 images in its test set. Figure S1 presents the analysis of PlotQA in our system. As shown in the *Embedding Overview* (Figure S1 (a)), despite the significantly larger size, this synthetic collection still suffers from limited diversity and richness of visualization, just like Data2Vis. Even worse, PlotQA is concentrated at only three types: *Bar*,*Line* and *Point*. Through filtering, zooming and visual query (Figure S1 (b), (c)), we can explore the images and find in the *Selection* panel and *Visualization Gallery* that the synthetic charts still have simplistic styles, even though the images look slightly more diverse than that in Data2Vis. For machine learning and visualization researchers who want to develop a more general question answering model for visualization images, a more diverse collection is needed, which should include more types and more complex real-world visualization images. For designers, this collection also does not provide enough diversity for design reference and inspiration.

- *VisImages*. VisImages[2] is a collection sourced from 1,397 papers in IEEE InfoVis and VAST, which contains 12,267 images, in which 10,055 images are annotated to contain visualizations. The images are collected and annotated by experts according to a comprehensive taxonomy. The analysis of this collection is presented in Figure S2. We can see that this real-world collection contains a more balanced distribution compared to synthetic collections, which is consistent with our findings from the VIS30K collection. In fact, as shown in Figure S2 (a), VisImages looks even more balanced than VIS30K because it does not have a single dominant type as the Diagram in VIS30K. By further exploring the images with our histogram + lasso filtering (Figure S2 (b)) and visual query interactions, we can find that the diagrams in VisImages are mostly flow charts and some illustrations. However, we cannot find many scientific visualizations as we do in VIS30K. This is because VisImages only crawls data from InfoVis and VAST.

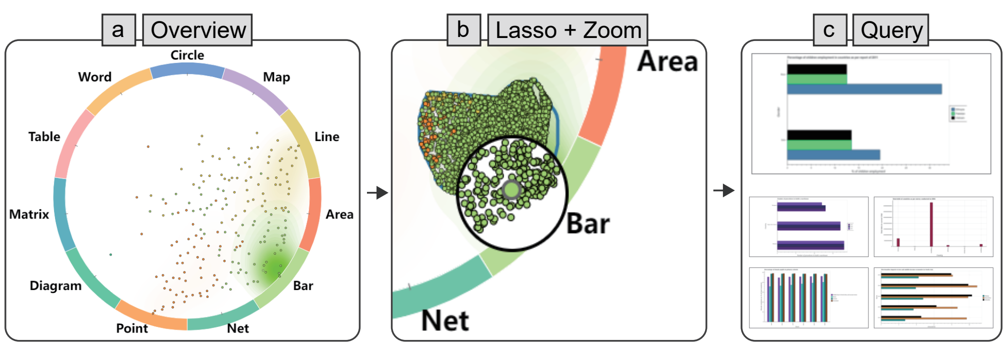

Fig. S1. The overview and exploration of PlotQA collection shows that this synthetic collection is also unbalanced in type distribution.

2. CROSS VALIDATION OF MODEL ON VISIMAGES

For more evaluation results, we provide here cross validation of models trained on our dataset and VisImages[2] respectively. We followed the practice in VisImages and provide cross validation

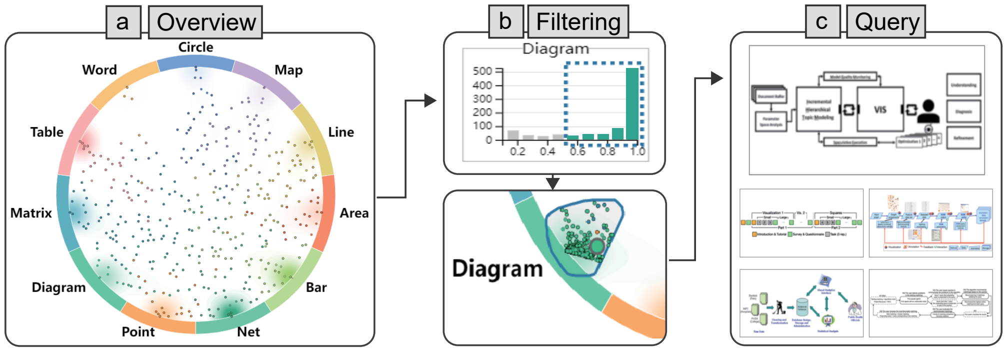

Fig. S2. The overview and exploration of VisImages collection shows that this real-world collection is balanced and contains diverse visualizations.

accuracy on the VisImages collection. First, as VisImages' taxonomy has more classes, we select the 11 common major classes from our taxonomy and VisImages. Then we tested the models trained on our training set and VisImages' training set. Similar to the procedure in VisImages, the two models are tested on both our testing set and VisImages' testing set and we provide Top-1 and Top-3 accuracies of our model.

Similar to the findings in VisImages[2], the results here show that models trained on one collection will experience a drop of performance when tested on another collection which has a different taxonomy or a different data generation process. We can also see that the VisImages collection is generally harder to classify, even for the model trained on VisImages' own training set. This reflects the existing challenges of classifying complex real-world visualizations. For example, node-link tree and treemap are variants of the same major type, but they are visually more different than other visualizations of different types.

Table S1. Test accuracy of ResNet50 models trained on our collection and the VisImages collection

Training Set	*Ours*		*VisImages*	
Testing Set	*Ours*	*VisImages*	*Ours*	*VisImages*
Top-1 Accuracy	**96.35%**	70.08%	68.22%	**78.35%**
Top-3 Accuracy	**98.57%**	81.05%	85.87%	**92.64%**

REFERENCES

1. N. Methani, P. Ganguly, M. M. Khapra, and P. Kumar, "Plotqa: Reasoning over scientific plots," in *Proc. WACV,* (2020), pp. 1527–1536.
2. D. Deng, Y. Wu, X. Shu, J. Wu, S. Fu, W. Cui, and Y. Wu, "Visimages: A fine-grained expert-annotated visualization dataset," IEEE Trans. Vis. Comput. Graph. pp. 1–1 (2022).