Geometry Diagram Understanding

Trideep Rath
Computer Science
Arizona State Univeristy
Email:trath@asu.edu

Arindam Mitra
Computer Science
Arizona State Univeristy
Email:amitra7@asu.edu

1. Introduction

Solving geometry questions is a long desired goal in AI. A geometry question generally consists of a diagram and a text describing the diagram followed by a question. To solve a geometry problem one needs to understand both the text and the image. In this project we will be working on diagram understanding. Fig 1 shows an example of a geometry problem.

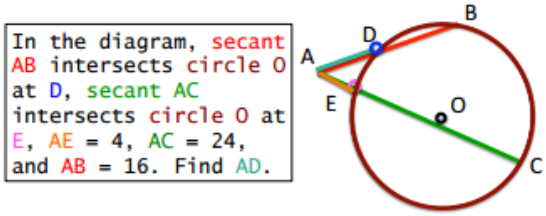

Figure 1. Diagram understanding: identifying visual elements in the diagram and aligning them with their textual mentions. Visual elements and their corresponding textual mentions are color coded. This Figure is best viewed in color.

The goal of the diagram understanding is to detect the visual elements from the diagram and to align those with the mentions in the texts as shown in Fig 1. The set of visual elements includes lines, circle, triangle, chords, arcs etc. A textual mention is a word or phrase that corresponds to a visual element, such as "circle O", "secant AB".

2. Related Work and Our Approach

In the work of [1] the goal is achieved in multiple steps. First the basic visual elements such as circle segments and lines are extracted using Hough transform, then post-processing is done to extract the endpoints and finally corner detection method is used to glue the parts. For the textual mention extraction they use keyword search and finally align the extracted mentions with the visual elements using a set of constraints. For example to align the textual mention of "circle o" with it visual element, it looks for a circle for which all the points on its perimeter is equidistant with the point 0. As we can see the methods has several steps. Our first goal is to formulate the problem in such a way that it is end to end trainable using a deep learning architecture. We would further experiment with various neural networks used in vision to compare with their performance. A high level break down of the overall task and a tentative completion date is shown in Table 1. We will further divide this tasks among ourselves as we progress further along our goal.

Task	Owner	Completion Date
Understanding annotations in geometry problems	Arindam, Trideep	1^{st}March
Reformulating the problem for end-to-end training	Arindam Trideep	14^{th} March
Implementation	Arindam, Trideep	28^{th} March
Evaluation	Arindam, Trideep	20^{th} April

TABLE 1. A TENTATIVE BREAKDOWN OF OUR TASKS AND THE PROJECTED COMPLETION DATE. SINCE IT IS OUR FIRST WORK ON VISION PROBLEM WE WANT TO SPEND SOME TIME UNDERSTANDING HOW THE ANNOTATIONS ARE PROVIDED IN THE TRAINING DATASET. SINCE THERE IS NO DOCUMENTATION OF THAT WE WANT TO SPEND SOME TIME UNDERSTANDING THAT PART FIRST. WE HAVE EMAILED THE FIRST AUTHOR OF [1] REGARDING THIS AND WE HAVE STARTED COMMUNICATION IN THIS REGARD.

3. Dataset

The dataset was released as the part of the work in [1]. The dataset consists of plane geometry questions each having textual description in english and accompanied by a diagram. The diagrams are annotated with correct visual elements as well as alignments between textual mentions and visual elements.Questions are taken from (RegentsPrep-Center; EdHelper; SATMath; SATPractise) websites for high school geometry. The dataset has 100 questions with 482 ground truth labels.

4. Conclusion

In many scenarios one needs to ground the image to a problem specific ontology for understanding the information it contains. Understanding diagrams in physics, geometry, geography and bio-science are all examples of this. By attempting to solve geometry problems we are moving one step towards it. Beside to understand an image one often needs to work with both the image and the accompanying

text. Currently several challenges are proposed such as Visual Question Answering [3], CLEVER [2] that test the skill in vision and language together. The task of solving geometry problems is also related to this. Also, there are several objectives in computer vision such as discovering spacial relationship between objects (concentric circles, pericircle, parallel lines), comparing size of objects (which circle refers to the mention 'smaller circle'?) and finding contour of complex regions which still needs to be triumphed and solving geometry problems can lead us towards that.

References

[1] Seo, M. J., Hajishirzi, H., Farhadi, A., & Etzioni, O. (2014, July). Diagram Understanding in Geometry Questions. In AAAI (pp. 2831-2838).

[2] Johnson, J., Hariharan, B., van der Maaten, L., Fei-Fei, L., Zitnick, C. L., & Girshick, R. (2016). CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning. arXiv preprint arXiv:1612.06890.

[3] Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C., & Parikh, D. (2015). Vqa: Visual question answering. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2425-2433).