Diagram Understanding in Geometry Questions

Min Joon Seo¹, Hanna Hajishirzi¹, Ali Farhadi¹, Oren Etzioni²
Sample Geometry Question
In the diagram at the right, circle O has a radius of 5, and CE = 2. Diameter AC is perpendicular to chord BD. What is the length of BD?
In the diagram at the right, circle O has a radius of 5, and CE = 2. Diameter AC is perpendicular to chord BD. What is the length of BD?
In the diagram at the right, circle O has a radius of 5, and CE = 2. Diameter AC is perpendicular to chord BD. What is the length of BD?
Diagram Understanding

1. Discovering locations of visual elements.
2. Discovering their geometric properties.
3. Aligning them with the text.

In the diagram at the right, circle O has a radius of 5, and CE = 2. Diameter AC is perpendicular to chord BD. What is the length of BD?
In the diagram at the right, circle O has a radius of 5, and CE = 2. **Diameter AC** is perpendicular to chord BD. What is the length of BD?
Standard Vision Techniques

- Usually uses **Hough Transform**:
  - **Detection**: Finds hundreds of lines and circles, each with some confidence level.
  - **Filtering**: Removes those with low confidence (thresholding); removes similar lines and circles (non-maximum suppression).

**Problem**: Parameter-sensitive (5 params)
Performance of Pure Hough

1 Extra Line

1 Missing Line

*Same parameters for both figures
Use both **diagram** and **text**
Intuition

- Start with unfiltered lines and circles (primitives, $L$).

- **Goal**: Find a subset of primitives ($\hat{L}$) that best represents the diagram, using information from *both text and diagram*.
Evaluation Function

How do we know if $\hat{L}$ represents diagram “well”?

$$F(\hat{L})$$

- Pixel **coverage** of primitives $P$
- Visual **coherence** between primitives $C$
- **Alignment** with textual information $S$
Optimal subset of primitives $\hat{L}$ should explain most of the non-white pixels in the diagram.

Maximize \( P(D, \hat{L}) = \frac{|D_\hat{L}|}{|D|} = \frac{\text{# pixels covered by the set } L}{\text{# all pixels in the diagram}} \)
Detected visual elements are visually coherent if they agree on corners.

$$C(H, \hat{L}) = \frac{|H_{\hat{L}}|}{|H|} = \frac{\text{# corners covered by } L}{\text{# all Corners in the diagram}}$$
Maximizes alignment between primitives and textual mentions

- Mention triangle ABC → three lines AB, AC, BC
- Aligned mention: corresponding primitives are detected

Alignment = \frac{\# \text{ aligned mentions in } L}{\# \text{ all mentions}}

- Also penalize redundancy

Text alignment

\[ S(T, \hat{L}) = \frac{|T \hat{L}|}{|T|} - r_{\hat{L}} \]
\[
\hat{L}^* = \arg\max_{\hat{L} \in 2^L} F(\hat{L}, D, H, T) = \\
P(D, \hat{L}) \quad \text{Coverage} \\
+ C(H, \hat{L}) \quad \text{Visual Coherence} \\
+ S(T, \hat{L}) \quad \text{Text alignment}
\]

**Problem:** Optimization requires \(2^{|L|}\) operations

**Solution:** \(F\) is submodular
Suboptimal Efficient Algorithm

Greedy Algorithm

- $\geq 1 - \frac{1}{e}$ of maxima, by submodularity of $F$
- $O(n)$, where $n$ is the number of primitives

$$l^{k+1} = \arg\max_{l \in L \setminus \hat{L}^k} F(\hat{L}^k \cup \{l\}) - F(\hat{L}^k)$$
Experiments

Dataset (100 questions, 482 alignments)

- Questions compiled from four websites for high school geometry (RegentsPrepCenter, EdHelper, SATMath, SATPractice)

- Manually recorded ground truth for visual primitives and textual alignment

- Dataset can be downloaded at: cs.washington.edu/research/ai/geometry
Experiments: Detecting Primitives

G-Aligner

Best Hough, tuned on entire dataset
Experiments: Alignment Accuracy

Alignment Accuracy (%)

- Hough
- G-Aligner
Custom Problem
Upload your own problem by specifying image URL and question text.

OR

Examples
Click an image below.

Diagram Understanding in Geometry Questions
To Appear in AAAI '14
View PDF

©2014 University of Washington
Conclusion & Future Work

- Diagram understanding
  - Overproduce lines and circles via Hough
  - Find a “good” subset
  - Objective function uses text and diagram

- Textual understanding
  - Knowledge representation of both text and diagram
  - Inference engine to solve questions
Thank you!

*Diagram Understanding in Geometry Questions*

For more information, please visit:

[cs.washington.edu/research/ai/geometry](http://cs.washington.edu/research/ai/geometry)
Hough Transform

http://www.cs.utah.edu/~vpegorar/courses/cs7966/