A Review: Abstract Visual Reasoning with Tangram Shapes

The paper "Abstract Visual Reasoning with Tangram Forms" proposes utilizing tangrams to study abstract visual reasoning. Unfortunately, the research fails to clearly demonstrate how the current problem setting would be applicable to real-world circumstances, limiting its significance.

The Kilogram dataset was created by the authors and comprises 1,016 tangram arrangements and 13,404 annotations. The dataset is unique and the first of its kind, however, the annotation variability is only moderate. The authors construct reference games for an annotated text-image pair at random by picking extra k-1 pictures from the data while adhering to numerous constraints. They also take into account a variety of input variations, including entire descriptions, descriptions with part information, tangrams without part segmentation, and tangrams with part segmentation. However, the authors may well have linked the descriptions of the parts with the colour of the segmentations to check if linking the two truly helps.

The model's evaluation is deficient in various respects. The authors did not present adequate prior work or baselines against which to assess their model, and the use of solely CLIP and VILT to provide evaluation outcomes is not a good fit for the problem setting. The evaluation fails to clearly demonstrate what the research topic sought to solve (i.e visual reasoning) and the ablation study is insufficiently detailed. The research could not explain why ViLT with one encoder outperformed CLIP with two encoders. Because the problem setting is artificial, contrived and similar to training on a random supervised learning dataset, the paper's contribution to visual reasoning is limited. The authors' evaluation metric is also inadequate and does not give a full assessment of the model's performance.

Overall, I believe the article lacks innovation and makes minimal contributions to the subject of abstract visual reasoning. The paper contains various shortcomings and fails to effectively answer the research question it sought to answer. While the Kilogram dataset is unique, the research lacks evidence to support the assertion that tangrams are a valuable tool for comprehending complexity in visual thinking. The authors should have supplied a more varied dataset, a clearer explanation of how this aids in real-world image reasoning through experimentation, and a more complete assessment that addresses the problem setting's constraints.