

Peer Review

Review of: "Grade Inflation in Generative Models"

Attila Fazekas¹

1. University of Debrecen, Hungary

A very interesting paper that deals with the fact that quality scores used to characterize generative models often give better results in two-dimensional cases than they should. This is the so-called "grade inflation problem". The limitation of the article, which the authors also describe, is that they focus exclusively on the 2D case, and it would be a difficult task to qualify this in a subjective human experiment if we have higher, especially more than 3-dimensional samples.

I would like to draw attention to two important things:

1. The statement that the quality score gives a better result than it should means that we have the perfect quality score, which the researchers consider objectively better. If there were such a thing, then why would this paper be interesting? This *circulus vitiosus* should be resolved by refining the text.
2. The concept of fit based on human experiments can raise many questions. The participants in the experiment cannot be considered a representative sample of humans. It would be worth doing more such experiments and also looking at the variance of human observers' responses.

I think that in the case of generative models, the only expectation is that they generate data that is consistent in certain parameters. If there are no constraints on the geometric properties in the feature space of the samples, then why is it a problem that they are not met?

Declarations

Potential competing interests: No potential competing interests to declare.