# Review

The problem statement of generating images at pixel-level is well described. The main idea is to decreasing image sequence length using encoding methods from NLP, such as BPE and its variations. Comparison with discrete feature based regression is presented.

According to report structure, following points can be improved. Proposed method can be put into separate section "Proposed Method". Link to Figure 2 is missing in Section 3.1. Figure 2 needs axes labelling. Experiment protocol and presented results seems reasonable, however, they can be expanded in future. `README.md` may include installation section.

The problem is interesting and usage of BPE encoding seems reasonable when using GPT model. However, there are other encoding schemes like Huffman coding, Discrete Cosine Transform (it is used in JPEG), etc. Using image encoding technic may also give vocabulary to be used in GPT. Or you can compress image with image compression method and then apply BPE. Moreover, converting RGB value (R, G, B) to 3-char value can be done differently, for example, you can reduce size of (R, G, B) vocabulary of size 256*256*256 to smaller size using clustering. It is important as quality of input image may affect quality of output.

One more idea is to replace GPT with Transformer which attention block has linear complexity regarding number of input tokens/sequence length. For example, efficient Transformers can be found in [1] and GPT compression in [2].

---

[1] Ziyaden, A., Yelenov, A., & Pak, A. (2021). Long-context Transformers: A survey. 2021 5th Scientific School Dynamics of Complex Networks and their Applications (DCNA), 215-218.
[2]Edalati, A., Tahaei, M.S., Rashid, A., Nia, V., Clark, J.J., & Rezagholizadeh, M. (2022). Kronecker Decomposition for GPT Compression. ACL.