Code Generation From Visual Representations

Tommaso Calò

PhD Candidate:

1. Introduction

Translating web UI designs into code is labor and time-intensive due to the frequent adjustments between designers and developers. While deep learning offers to automatically translate visual desing to code, many existing methods rely on basic datasets and primarily use Recurrent Neural Networks, which limits their applicability in real world scenarios.

2. Objectives

We explore the introduction of a novel dataset containing real web screenshots and their associated code along with a technique to reduce the collected code to the only element affecting the visual appearance of the related websites. Furthermore we introduce and investigate the performance of a multimodal Transformer architecture tailored to translate visual web design into compilable code.

3. Method

For dataset collection, the webpage's HTML is downloaded and its CSS URLs extracted.
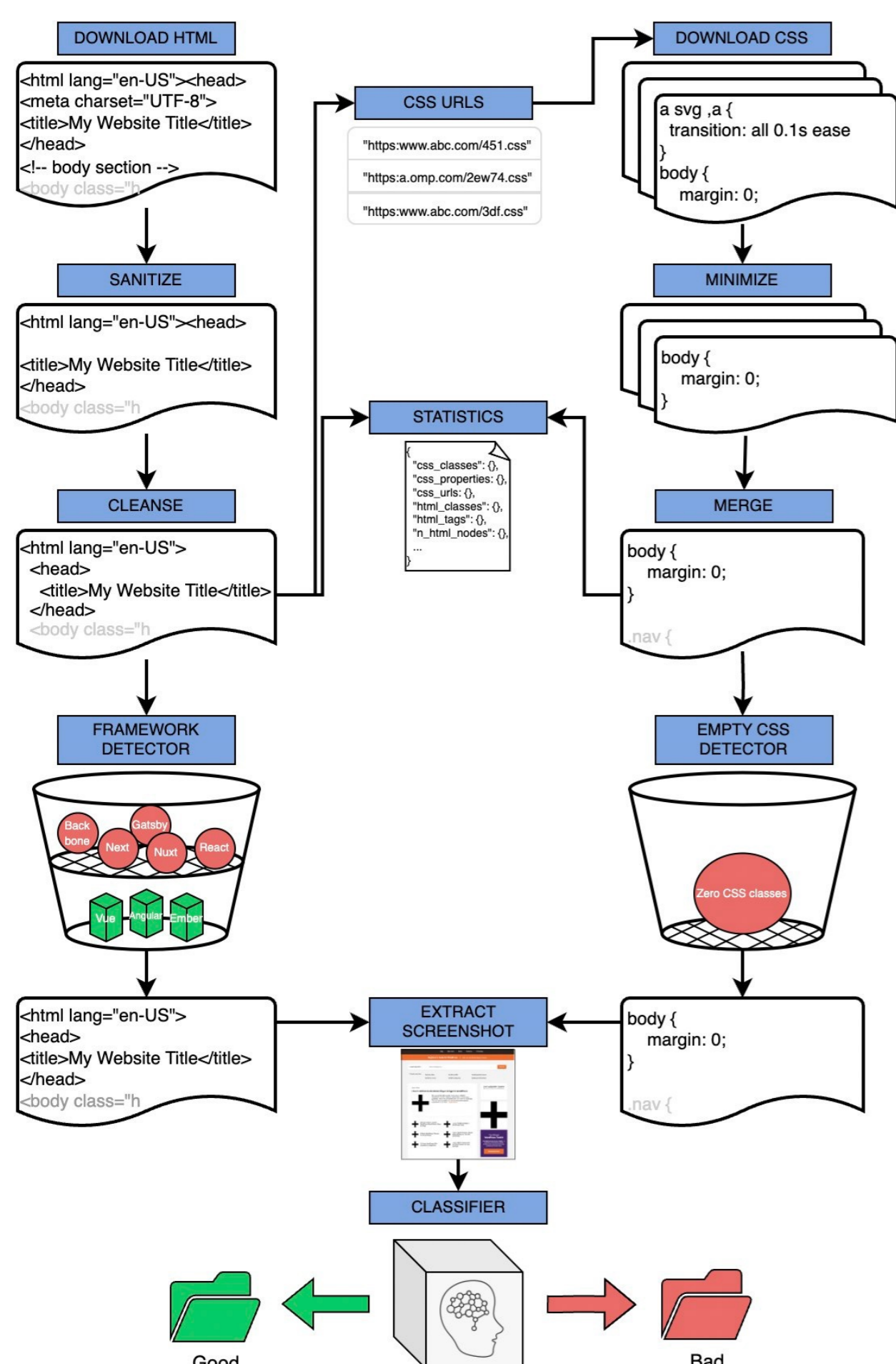
The HTML is sanitized, and the CSS is minimized. Frameworks are identified, and unused CSS discarded. A webpage screenshot is taken, and along with the refined code, it's classified to assess the translation quality.

Next, we introduce a multimodal Transformer architecture to translate visual web designs into code. It simultaneously processes textual and visual inputs. Text is tokenized and combined with position data, while visuals are segmented into patches with positional embeddings. Through attention mechanisms, it aligns text and visuals, eventually decoding them into a predictive code sequence.
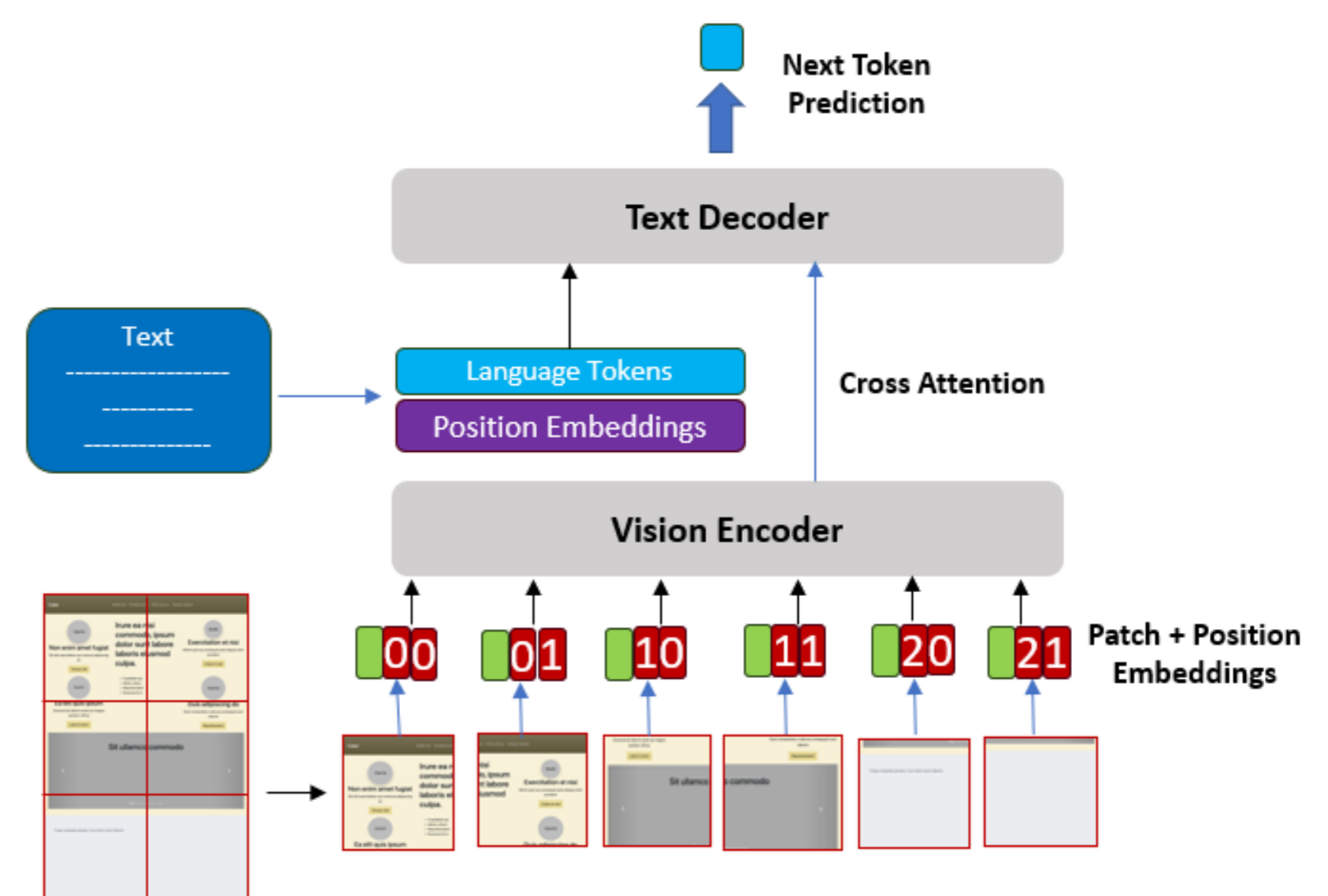
4. Results

We conducted preliminary tests of our architecture on both the pix2code and ui2code datasets. The results indicated that our approach outperformed the existing RNN-based architectures in translating visual web designs into code, while it is less performant on our proposed real world web design dataset.

5. Conclusions

Our study showcases the efficacy of a multimodal Transformer architecture in the domain of web design automatic translation and introduces a real-world design/code dataset. As the next step, it will be crucial to focus on extensive real-world evaluations, to adapt our method to diverse and complex web designs and procedures encountered in practical scenarios.

[1] Beltramelli, Tony. "pix2code: Generating code from a graphical user interface screenshot." *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems.* 2018

C. Chen, T. Su, G. Meng, Z. Xing and Y. Liu, "From UI Design Image to GUI Skeleton: A Neural Machine Translator to Bootstrap Mobile GUI Implementation," *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE),*