Institute of Cartography
and Geoinformation

From Vector Features to Stylized Maps – Exploration of Stable Diffusion Applied to Maps

Claudio Affolter
Master’s Thesis, Spring Semester 2024
Supervisors: Sidi Wu, Dr. Yizi Chen, Prof. Dr. Lorenz Hurni

1 Introduction

In recent years, diffusion models, a class of generative AI models, have set the state-of-the-art in image generation due to their ability to create visually stunning images from text prompts (i.e., user-given descriptions). However, when it comes to complex layouts, such as maps, text prompts fall short, as they offer limited control over spatial composition (i.e., the placement of features within a map tile) and semantic layout (i.e., the assignment of features to specific classes). Additionally, the prompt sequence length in text-to-image generators is typically restricted to a few hundred characters, which is insufficient to accurately describe even a single map tile.

2 Goals

1. Exploration of how image diffusion models, particularly Stable Diffusion (SD) **[1]**, can be controlled using text prompts in combination with vector data. With input vector data as conditioning controls, the trained model should be capable of generating map tiles in predefined styles specified by a text prompt (see Figure 1).
2. Development of a web-based application for map tile generation, allowing users to upload their own vector data to control the spatial and semantic layout of the output tile.

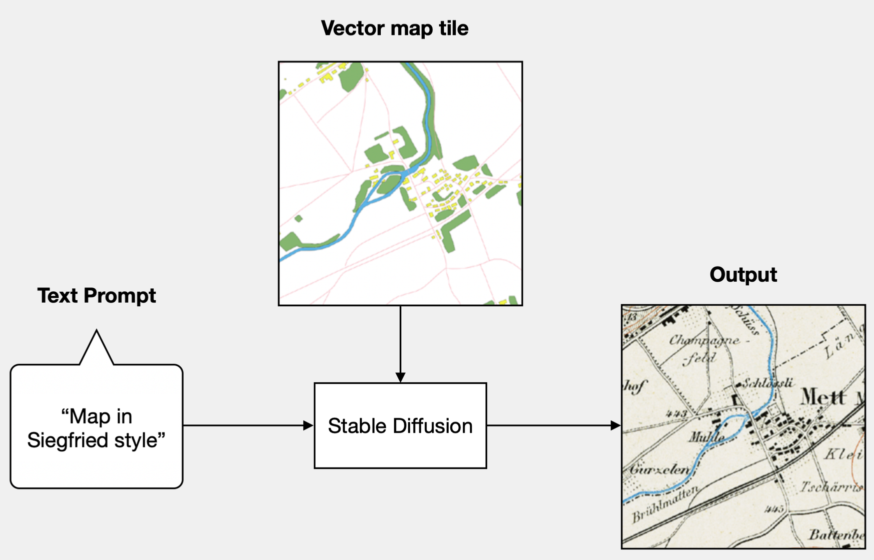

Figure 1: Overall idea.

3 Method Overview

- Collecting *Swisstopo*, *Old National* and *Siegfried* map sheets, along with the associated vector data. As there is no corresponding historical vector data for the Old National Map and Siegfried Map, modern Swisstopo data had to be used instead.
- Symbolizing the vector data in QGIS to ensure the best possible alignment with the corresponding raster data.
- Processing the raster data to mask the map labels, as leaving them in the training data could cause the model to learn to generate fake labels. For this, the annotation vector layers and/or the optical character recognition model keras-ocr **[2]** were employed.
- Utilizing the created datasets to train the neural network structure ControlNet **[3]** (see Figure 2) to control SD using vector map tiles.

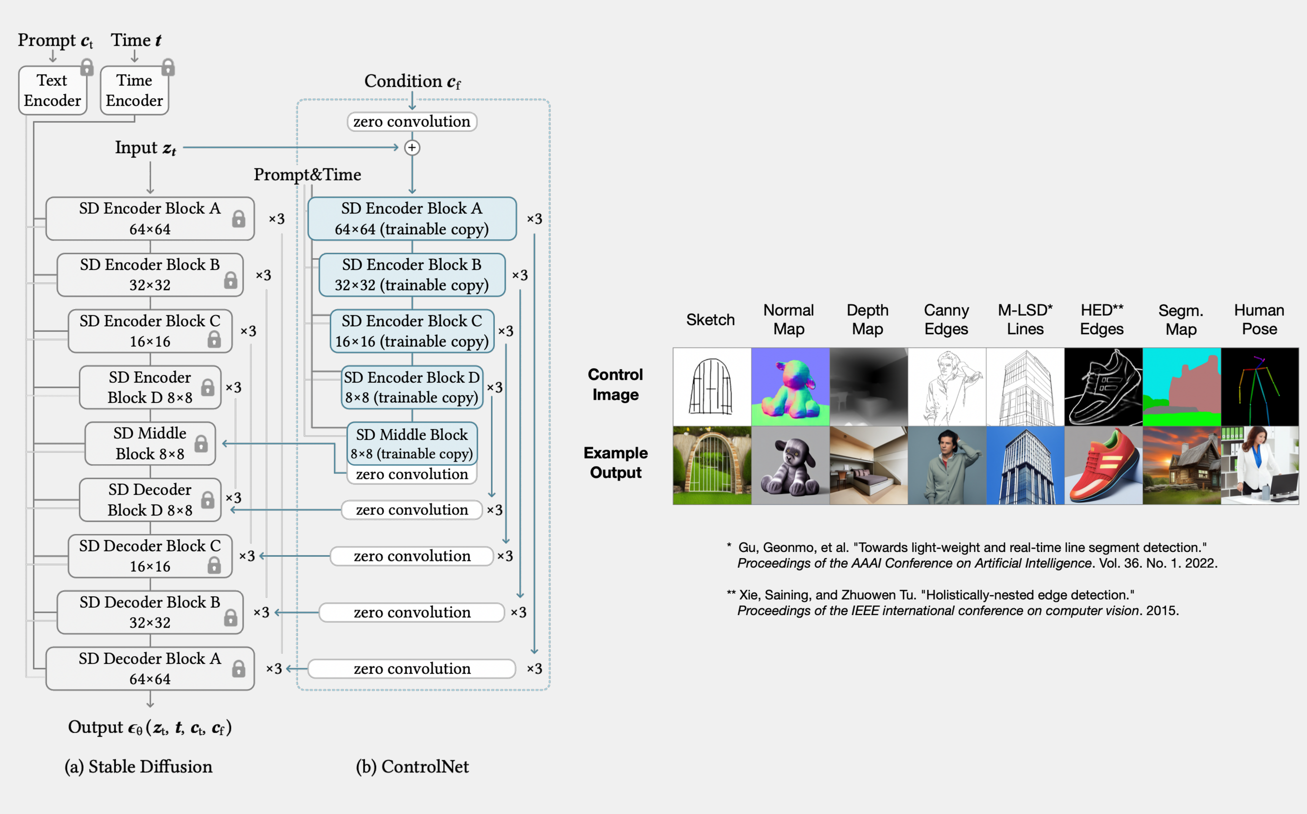

Figure 2: Left: U-Net architecture of SD connected with ControlNet. Right (top row): Control images supported by ControlNet. Right (bottom row): Example images generated by SD while controlled by ControlNet.

4 Results and Discussion

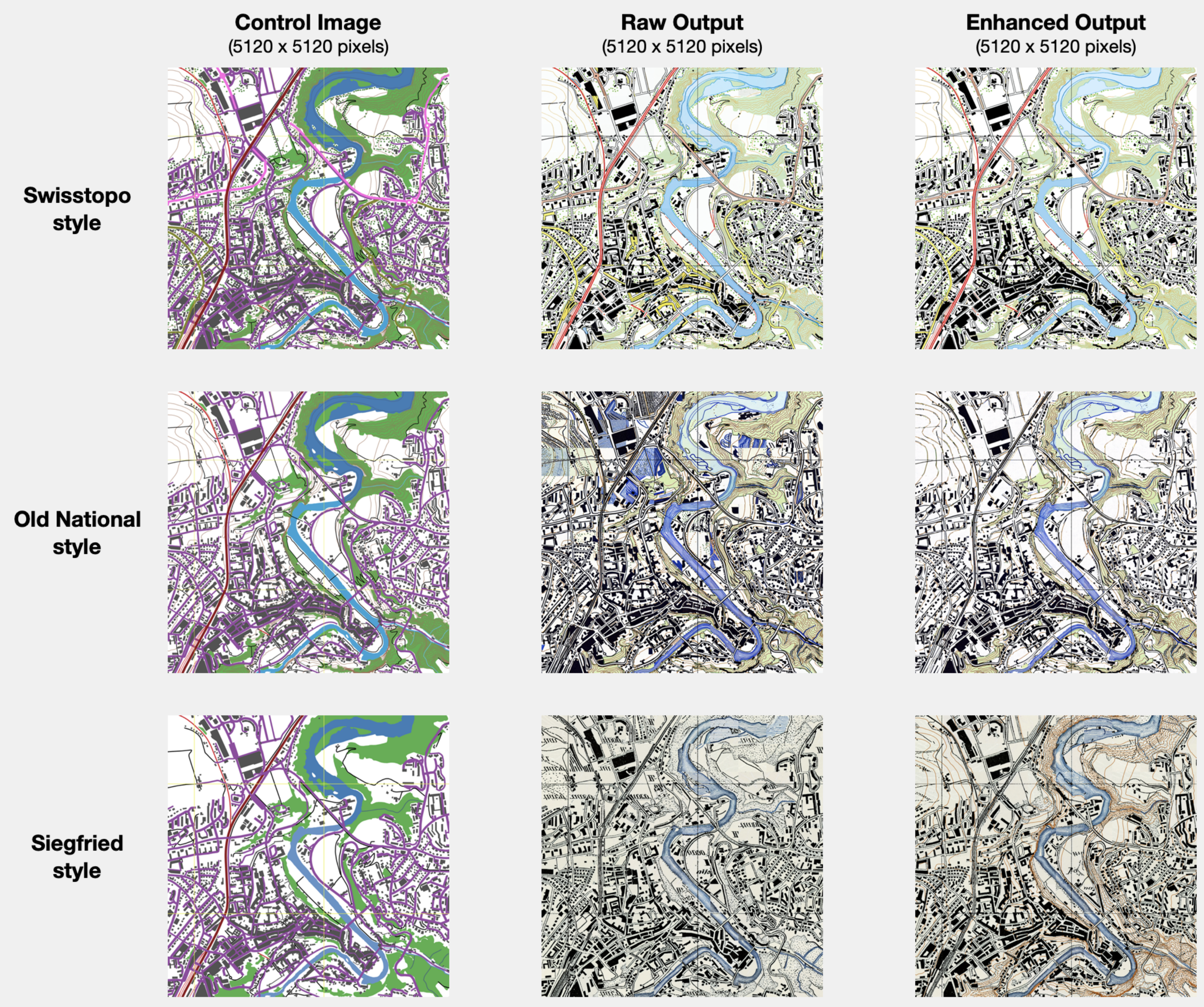

Figure 3: Qualitative performance overview of the three trained specialized ControlNet models (one for each map style) on the test set. Left column: Control images. Center column: Corresponding generated output maps. Right column: Output after post-processing / augmentation. All outputs were created by stitching together 100 generated map tiles.

Model	MSE	MSE_{post}
Swisstopo style	5647	4670
Old National style	17459*	15488*

Model	MIoU	Mask MSE	MIoU_{aug}	Mask MSE_{aug}
Siegfried style	0.25	7444	0.33	5307

Table 1 (left) lists the MSE relative to the ground-truth maps for *Swisstopo* and *Old National* styles, both before and after post-processing. The two MSE values marked with an asterisk (*) could not be calculated using true ground-truth data (as it does not exist); instead, the original Old National Map raster data was used. Table 2 (right) presents the quantitative performance overview for the *Siegfried* style. For evaluation, since no ground truth exists, a segmentation model calculating the MIoU between segmented generated map tiles and corresponding control images was employed, along with *Mask MSE* that calculates the extent of fake labels in a generated tile. These metrics were calculated both before and after augmentation.

Furthermore, a web application employing a combined ControlNet model capable of generating map tiles in all three styles has been developed. It allows users to upload, draw, or edit their own vector map tiles to control the spatial and semantic layout.

Depending on the degree of misalignment between the vector and raster data in the training set, the models produce map tiles that closely resemble the intended layouts and styles to varying extents. Simple post-processing and augmentation techniques further improved the overall output quality by correcting colors and removing fake labels and undesired randomly generated structures.

5 Conclusion

ControlNet has been successfully employed to control the image diffusion model Stable Diffusion using vector maps for map tile generation in three distinct styles. The resulting map tiles, after simple post-processing and augmentation steps, are of usable quality and could be applied to tasks such as change simulation and inpainting. Moreover, this work could serve as a foundation for future research and development in the field of cartography, paving the way for even more advanced methods of accurate map tile generation.

References

[1] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[2] Fausto Morales. keras-ocr. 2019. Retrieved July 27, 2024, from https://keras-ocr.readthedocs.io.

[3] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.