Traffic sign data augmentation using Stable Diffusion conditioned with ControlNet

Tomáš Železný*

1 Introduction

With the recent trend towards autonomous vehicles, traffic sign recognition has become an important automotive task. Like other machine learning tasks, it requires large amounts of data. The datasets are often created using dashcam or similar videos and then manually annotated. As traffic signs are specific to each country, the datasets are expensive. Therefore, there's a need to augment existing data to increase its diversity and size without incurring the high costs of manual annotation.

Our goal is to apply strong augmentations to create a diverse dataset that remains realistic. We use the Stable Diffusion image generation model[1] (Rombach et al. (2022)) and condition it using ControlNet (Zhang (2023)) - a model specifically designed to condition Stable Diffusion using another image. We train ControlNet to use Canny's edge detection map as an input to condition Stable Diffusion. This allows us to generate new realistic images as an extension to the existing dataset.

2 Dataset

The dataset we use was created in 2013 as part of the TAČR project with the aim of passportization in the Czech Republic. The dataset was later processed and annotated, resulting in a subset of 267 classes corresponding to Czech traffic sign IDs in 2013. The dataset statistics show that some traffic signs are rarer than others on Czech roads. Some signs even occur only once in the dataset. This fact encourages us for the future work (see Conclusion).

Figure 1: Left: Source image. Middle: Canny edge detector map. Right: Output image from the Stable Diffusion conditioned with the Canny map

* Student of the doctoral degree program Cybernetics, field of study Computer Vision,
e-mail: zeleznyt@kky.zcu.cz
[1]Specifically, we use v2.1 model: `https://github.com/CompVis/stable-diffusion`

3 Experiments

First, we detect the edges in the source image using the Canny edge detector. Then it is used to condition the Stable Diffusion using the ControlNet. We prompt the Stable Diffusion model with "A traffic sign `[id]`", where `[id]` is the name of the traffic sign (e.g. 'Speed limit').

Figure 2: Examples of the output of Stable Diffusion (SD) conditioned with ControlNet. Top: ControlNet trained with frozen SD. Bottom: ControlNet trained together with SD.

4 Conclusion

We conducted preliminary experiments using Stable Diffusion to generate synthetic traffic sign data. Our results show that freezing Stable Diffusion during the training of ControlNet results in the generation of traffic signs with incorrect colors. We believe that this behavior is caused by the nature of the Stable Diffusion training data, which include traffic signs from different countries with different color schemes. In our follow-up experiments, we plan to train a classification network on our dataset and utilize Stable Diffusion to augment the data.

For future work, we aim to use ControlNet-conditioned Stable Diffusion for one-shot classification. This approach is motivated by the lack of certain traffic signs in our dataset, which prevents effective classifier training. Therefore, we intend to use reference images of traffic signs scraped from e.g. Wikipedia and apply augmentations with Stable Diffusion to create a diverse set of synthetic data.

Acknowledgement

The work has been supported by the grant of the University of West Bohemia, project No. SGS-2022-017. Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

References

Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conf

Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. erence on computer vision and pattern recognition. 2022.