Using GAN, we generate inconspicuous adversarial perturbations to traffic sign images and feed the perturbed images to a traffic sign classifier in an attempt to cause targeted misclassification and a reduction in the accuracy of the classifier.

- **German Traffic Sign (GTS) Dataset**
  - More than 50,000 images in 43 unbalanced classes
  - Image sizes vary between 15x15 to 250x250 pixels (RGB). All data were converted to 32x32 RGB images.
  - 67% training, 8% validation, and 25% testing datasets with similar relative class size

By developing an optimized GAN (with perturbation threshold of 0.05, 10:1 adversarial perturbation weighted-loss ratio, 16/32/64 filters in the generator, and a 1:1 generator/discriminator learning ratio), we generated inconspicuous adversarial perturbations to traffic sign images and fed the perturbed images to a TSR classifier, reducing the classifier test accuracy from 92.1% to 17.4% and resulting in a significantly high misclassification percentage of 82.9%.

To mitigate the effect of these attacks, autonomous vehicle software should include perturbed images in the training set for their traffic sign recognition classifier (future work suggestion). Another solution is defensive distillation.