This writing summarizes the paper titled "From Synthetic Images Towards Detectors of Real Objects: A Case Study on Road Sign Detection", published in ICAISC, 2019, by Przemyslaw Klesk. In this work, the author contends that collecting images and labelling them for use in training a detection algorithm can be taxing, and thus synthetic generation of such images with imposed labels could seem tempting. The paper sets out to determine the usefulness of such an approach, using road sign detection as an example.

To use synthetically created images for machine learning violates the principle of having random sampling of data, which could lead to higher error rates when testing a classifier on real life data while it was trained only on synthetic data. The author thus foresees high error rates in testing the performance of the classifier against real data.

In the study, the author limits the scope of the road signs to 42 warning signs, all of which consist of yellow triangles with red borders. The training sample images are created by randomly laying the icon set of the road signs over a set of backgrounds which consist of realistic representations like road views and landscapes, and unrealistic ones like in-doors, nature, etc. The images may have additional transformations such as rotations, brightness changes, noise, blurring, or sharpening. The probability of no transformation occurring, that is the images remaining plain, is 1/8, with a similar probability for other combinations of transformations taking place on any single image.

The created images are then cropped as fragments representing positive (contain road sign) and negative images. Next, the images go through a feature extraction phrase using two methods, Haar-like Features (HFs) and Fourier Moments (FMs). HFs could be used for detecting patterns based on edges, which would be well-suited for the contrast offered by road sign images against most backgrounds. FMs, on the other hand, may perform better for smoother patterns. Hence, the author takes it that HFs would do better in this study. Both methods can be used to extract features in constant time if aided with the math of integral images, he adds.

After feature extraction, the features are passed to the classifier, an ensemble learning algorithm, RealBoost with shallow decision trees. Ensemble algorithms like RealBoost or Real AdaBoost attempt to create a strong classifier from a combination of weak classifiers such as shallow decision trees. The algorithm creates consecutive models where each model tries to rectify the errors of the previous model, until a preset parameter is reached.

Doing some experimentations, erroneous false positives were detected such that smaller windows were detected lying inside actual targets, which meant in some cases, small areas within the triangle road signs were detected, but not the signs as a whole. To fix this issue, the author extends his training data, adding samples he calls hard negatives. Hard negatives are randomly cropped windows from the vicinity of the actual targets but not completely centered on them.

In the experimental setup, the classifier trained on different combinations of the synthetic training data types is tested against detection on real images. The different data types consist of samples with plain or transformed synthetic images, the presence or absence of the hard negative samples, whether HFs or FMs were used for feature selection, and the number of features passed from the total HFs or FMs to the classifier. The author observes that HFs-based detectors perform better than ones based on FMs. The observation is extended to add that detectors trained on transformed images in general surpassed ones trained on plain images. The best results, the author adds, are achieved when using 3 948 features from among the 10 125 available HFs, yielding 91.7% sensitivity with only 5% false alarms per image.

In conclusion, the author calls the experiment a success, and notes that despite the violation of the principle of random sampling of the data, synthetic images can be used to train classifiers for real-life data. He adds that HFs, as expected, are better suited than FMs to distinguish patterns with sharp edges. He closes with the remark that transformations on the synthetic images can increase performance of the classifiers, despite being simple to perform.