Remove as much noise as possible

1. Crop image to squared matrix with dimension fewer than initial keeping whole useful image in there. Modified stochastic hill climbing. Decision function – sum of nonzero points in frame. By visual inspection frame size is assigned to be 30x30. Algorithm starts by randomly initialized top left coordinates xi, yi. Frame moves in direction where decision function has max value. Step size is 1. When max is reached, new coordinates are stored in sorted list. Algorithm repeats until number of top scores reaches N=10. Coordinates that correspond to max of decision function is considered to be top left of new image with size 30x30. All 10000 images from training and test set are processed.
2. Use outlier detection algorithms to remove more noise. Contamination fraction is 0.3 by inspection. IsolationForest, LocalOutlierFactor, EllipticEnvelope and OneClassSVM from scikit-learn are used for outlier’s detection. Decision is based on results from all 4 algorithms. Training and test arrays are processed and stores in new files.
3. Convert from RGB to BW format. This is required to simplify our data.