**Dataset Source:**

IAPR TC11 is the International Association for Pattern Recognition (IAPR) Technical Committee Number 11. IAPR TC11 is concerned with the theory and applications of Reading Systems. We seek to study and develop systems that recognize character content and structure in handwritten and typeset documents, images, and video.

<http://www.iapr-tc11.org/mediawiki/index.php?title=IAPR-TC11:Reading_Systems>

**Dataset Source Link:**

<http://www.iapr-tc11.org/mediawiki/index.php?title=The_Street_View_House_Numbers_(SVHN)_Dataset>

Dataset Description:

SVHN is a real-world image dataset for developing machine learning and object recognition algorithms with minimal requirement on data preprocessing and formatting. It can be seen as similar in flavor to MNIST (e.g., the images are of small cropped digits), but incorporates an order of magnitude more labeled data (over 600,000-digit images) and comes from a significantly harder, unsolved, real world problem (recognizing digits and numbers in natural scene images). SVHN is obtained from house numbers in Google Street View images.

* 10 classes, 1 for each digit. Digit '1' has label 1, '9' has label 9 and '0' has label 10.
* 73257 digits for training, 26032 digits for testing, and 531131 additional, somewhat less difficult samples, to use as extra training data