**AUTNTdataset**

Aliah University Text Non-Text (AUTNT) dataset contains component level multi-script text non-text images. There are three-fold utilities of this dataset: (i) component level image classification, (ii) script identification, and (iii) character recognition. Ground truth data are also prepared and included in this repository. Component images are chosen from diverse sources and conditions to ensure unconstrained working environment and applicability in practical scenario.

**Dataset Summary**

Total number of component level images: 10771.

Text component images: 7890.

Non-text component images: 2881.

Dataset is divided into 5:1 ratio for training and test set. In the training set, 6314 text and 2305 non-text components along with ground truths are present. Among all text components of the training set, 3264 components are taken from complex document images and the rest 3050 components are taken from natural indoor/outdoor sources. Dataset contains multi-script text components viz. *Latin*, *Devanagari* and *Bengali.* Components are scale invariant and multi oriented in nature. Component images are having near-homogenous as well as complex background with uneven illumination effects

**Contributors**

* Tauseef Khan, PhD Scholar, Department of Computer Science and Engineering, Aliah University, IIA/27 New Town, Kolkata 700160, India
* Dr. Ayatullah Faruk Mollah, Assistant Professor, Department of Computer Science and Engineering, Aliah University, IIA/27 New Town, Kolkata 700160, India

**Special Acknowledgment**

* Munsi Md Iftabudin, student of PGMCA, Department of Computer Science and Engineering, Aliah University, IIA/27 New Town, Kolkata 700160, India.
* Fatema Rahman, student of B.tech 4th year, Department of Computer Science and Engineering, Aliah University, IIA/27 New Town, Kolkata 700160, India.
* Sk. Shamim, student of PGMCA, Department of Computer Science and Engineering, Aliah University, IIA/27 New Town, Kolkata 700160, India.

Contact: [iilab.cse@gmail.com](mailto:iilab.cse@gmail.com)