

Performance Analysis of Data Augmentation in Machine Learning

Recycling materials image recognition

Alin Ivan

Source code: <https://labcode.mdx.ac.uk/alinivan/ml-augmentation>

Setup

- Objective
- Approach

Evaluate the performance of data augmentation in recyclable materials image recognition

Recyclable materials: carton and plastic

Initial dataset: 100 images

Classes: 2 (carton and plastic)

Approach:

- Augmentation (noise, color and mixed)
- Training (real, augm and various hybrid)
- Testing (real, augm and various hybrid)

Tools: Python 3, Anaconda, OpenCV, Scikit-Learn, TensorFlow

Command line menu: load, augment data, train, test

Experiments

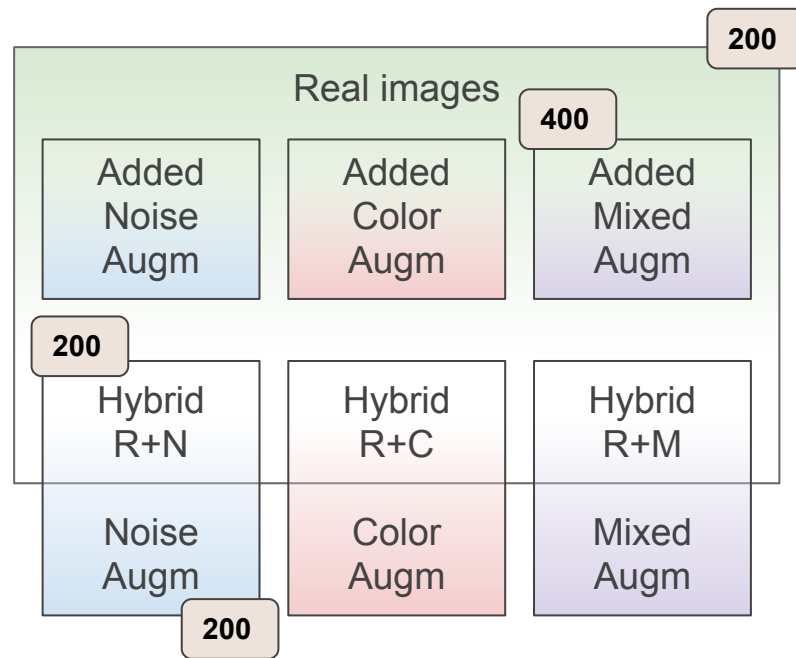
- Real, Augmented and Hybrid Data
- 2 ML methodologies
- 10 different data sets
- 20 different ML runs
- 10 trials each
- 200 repetitions

Machine Learning Methodologies:

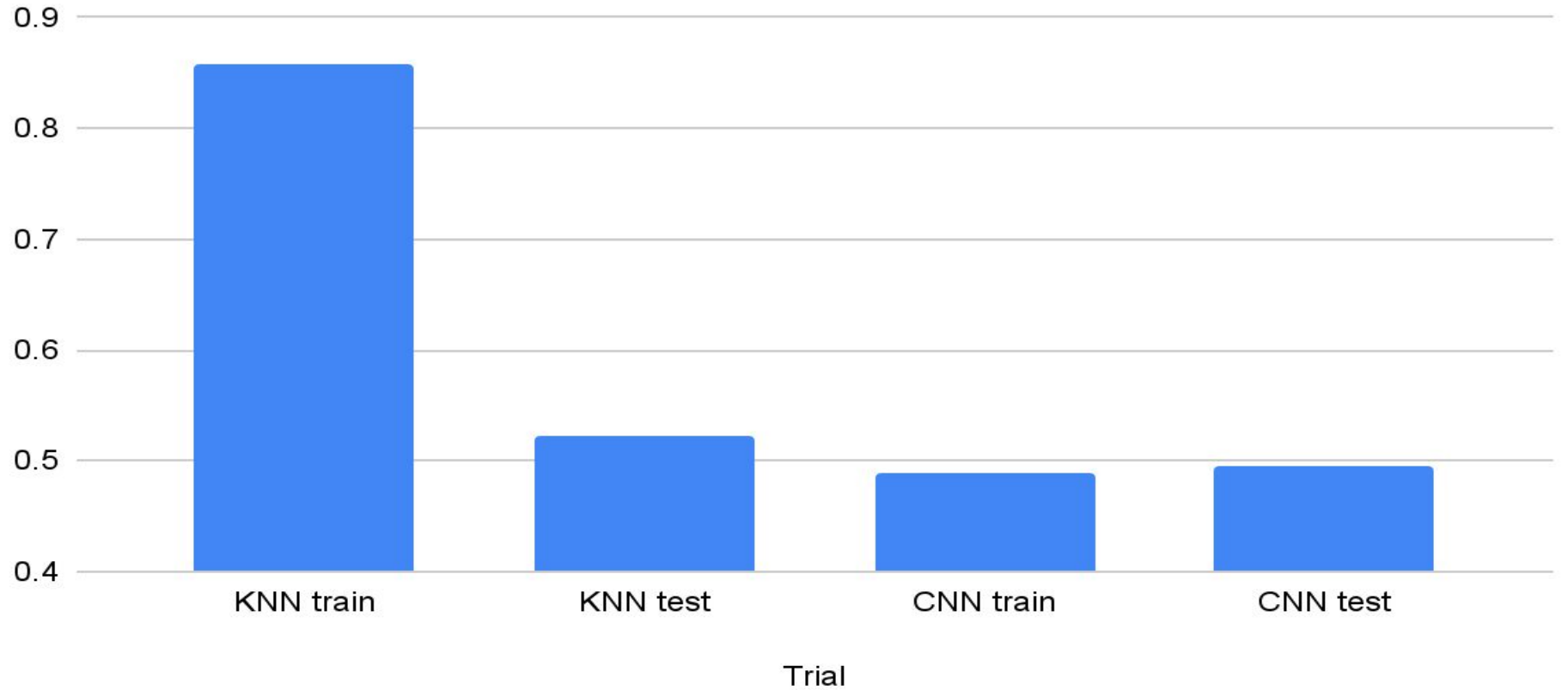
- K-Nearest Neighbors (3 neighbors)
- Convolutional Neural Networks (3 convolutional and 2 pooling layers)

10 different data sets **Train:Test Ratio 80:20**

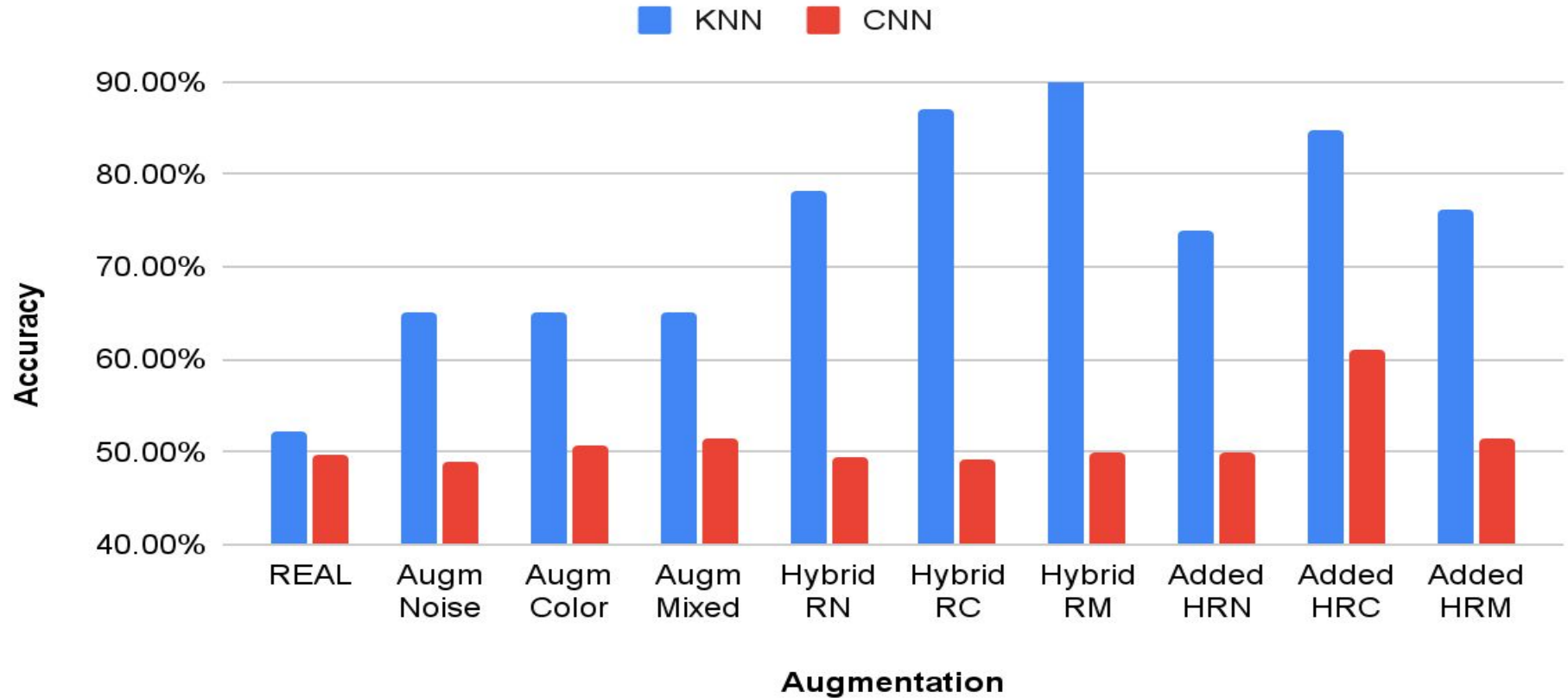
7 X 100 img/class & 3 x 200 img/class



Cross-Validation



ML with various augmentation



ML with various augmentation: grouped

