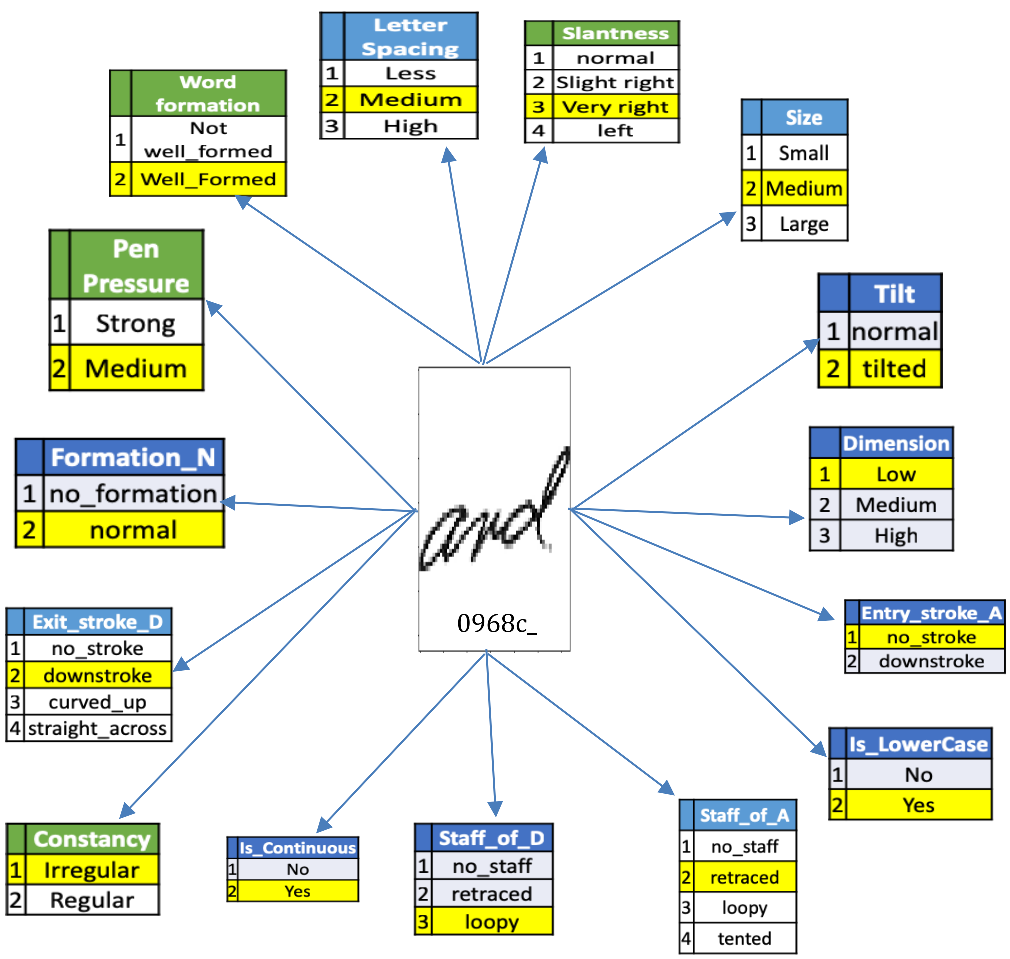
Writer Verification-Explainable AI

Abstract

Given the "and" dataset, using the handcrafted features the task is to perform bayesian inference ,deep learning inference to determine if the word "and" written by two writers is similar or dissimilar.

1 Dataset Annotation

Our dataset consists of AND images where each image is associated with 15 features. We handcraft the features for each image . The values for each feature depends on the way the word "and " is written. Also, each feature has a specific number of classes associated with it. Some of the features include pen pressure, is continous, word formation, letter spacing, slantness, dimension, staff of a and d, constancy etc.



2 Bayesian Inference

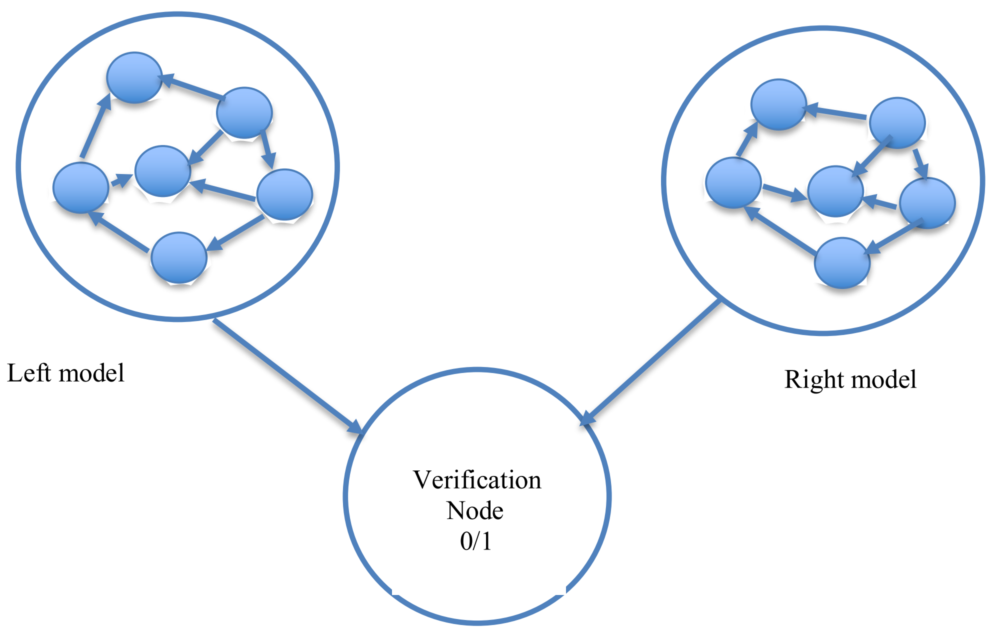
We build mutiple bayesian models for the task of verifying if two “and” images are similar or dissimilar. The created models are compared based on accuracy, time to train, time to infer and number of edges.

Since we compare the feature sets of two “and” images to compute their similarity , we created two bayesian models one for the left image and one for the right image. The first model is created by doing a hill climb search on the 15 feature set using k2 score to estimate the best edges needed to construct a network.

The first model is then replicated again to create another bayesian model . We differentiate the features used in both the models by creating two files with different suffixes. Hillclimb search on both models result in the same set of features.

We use a verification node as a means of verifying the similarity between two images. This node takes the value of 0 or 1. The value 0 implies that two images are dissimilar and value 1 implies that two images are similar.

A final bayesian model is created with all the edges of the left and right model .Two more edged are added to these models. One edge is from any one feature of the left model to the verification node. The same feature from the right model is connected to the verification node.



2.1. Calculating accuracy:

Once the bayesian model is created we train the bayesian model using the training images. We then infer the label values from the created model using infer.map query function.

It is then compared with label values we already have in the validation dataset. Based on the correct number of predicted values the accuracy is calculated.

2.1.1 Results:

The seen dataset produced the maximum accuracy among all datasets. The time to train to train and infer was also comparatively less since it had lesser number of images than unseen and shuffled dataset.
The accuracy of the seen dataset: **69.89**
The accuracy of unseen dataset: **65.32**
The accuracy of shuffled dataset: **65.32**

The seen dataset seemes to produce the best accuracy in comparison to the unseen and shuffled dataset.

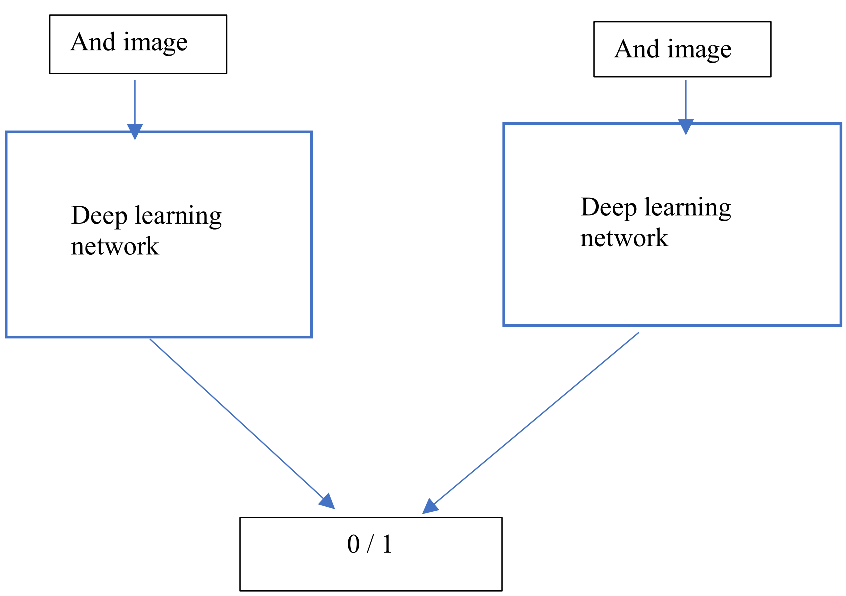
3 Deep learning Inference:

A siemese network is build for the task of verification. A model is built where convolution and max pooling are performed alternatively. A flattened model is created. The final model takes the "and" images as inputs with output being the model where the left and the right images are concatenated and batch normalization performed.

3.1 Siamese network:

Siamese networks are a special type of neural networks. It consists of two identical neural networks, each taking one of the two input images. Instead of a learning to classify the inputs , the network learns to differentiate between two inputs. By which it learns the similarity between them.

A Siamese network consists of two identical neural networks, each taking one of the two input images. There are two sister networks, which are identical neural networks, with the exact same weights. Each image in the image pair is fed to one of these networks.



3.1.1 Data creation:

For each dataset we obtain left and the right images along with the target data for training and validation.

3.1.2 Training and accuracy:

We train the siamese network with left and right image from the training dataset.

Then we evaluate the model with the left and the right images from the validation dataset.

Result:

Seen dataset:

Training accuracy for seen dataset:0.90

Training loss for seen dataset: 0.10

Validation accuracy for seen dataset:0.73

Validation loss for seen dataset: 0.39

Unseen dataset:

Training accuracy for unseen dataset:0.8345

Training loss for unseen dataset: 0.32

Validation accuracy for unseen dataset:0.6945

Validation loss for unseen dataset: 0.45

Shuffled dataset :

Training accuracy for shuffled dataset:0.8020

Training loss for shuffled dataset: 0.310

Validation accuracy for shuffled dataset:0.6545

Validation loss for shuffled dataset: 0.412

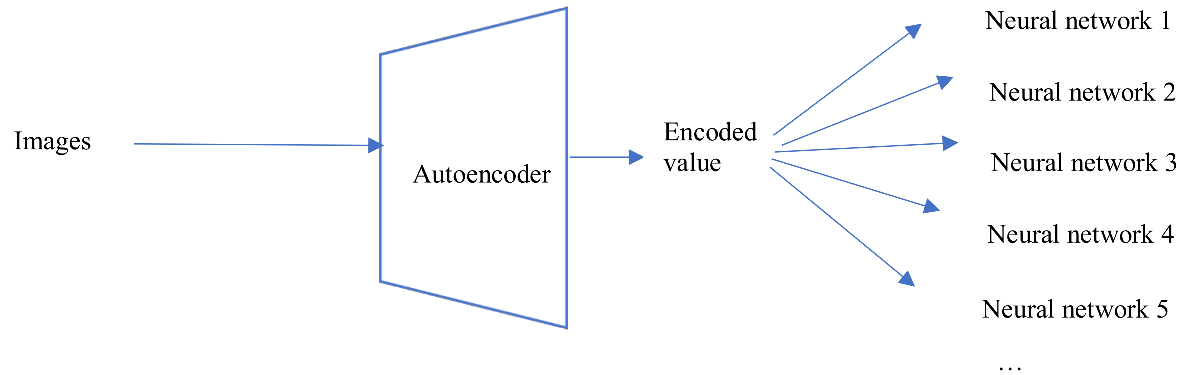
4 Explainable AI

We build a multitask learning model to learn the mapping between the "and" images and the handcrafted features. We use autoencoder to first pass the images to get the encoded value which is then passed to fifteen neural networks.

Cosine similarity is determined between every pair of images which shows how similar the images are. A threshold is used. If the cosine similarity is lesser than the threshold then the images are dissimilar and if the cosine similarity is greater than the threshold , the images are dissimilar.

Each neural network corresponds to one feature. The same encoded value is poassed to all of the neural networks but the number of outputs nodes changes in the output layer where softmax activation function is used.

For example, when we consider the first feature namely pen pressure it has 2 classes Strong and Medium. Thus the number of output nodes in the output layer of the first neural network will be 2. Similarly, for the remaining neural networks we consider the the number of classes of the features as output nodes.



The above structure is used and is follwed for each dataset thus resulting in 15 neural network for each dataset.

4.1 Training and accuracy:

Accuracy for seen dataset: 0.6989

Accurcay for unseen dataset: 0.5713

Accuracy for shuffled dataset: 0.6965

4.2 Observation:

Both the seen dataset and the shuffled dataset resulted in almost the same accuracy. The accuracy of the unseen dataset seemed to lower than both the seen and shuffled dataset.

4.3 Explanation for result:

Let us consider two and images from the seen dataset: **0838a_num1** and **0838a_num2**

These images have the label 1 meaning they are similar.

Reason for similarity:

Let us the consider the features for image 1 and image 2.

	0838a_num1	0838a_num2
Pen pressure	Medium	Strong
Letter spacing	Medium	Medium

size	Medium	Medium
dimension	Medium	Medium
lowercase	Yes	Yes
continous	No	No
slantness	Slight right	Slight right
tilt	normal	normal
Entry stroke	No sroke	No stroke
Staff of a	No staff	retraced
Formation of n	normal	normal
Staff of d	retraced	retraced
Exit stroke	Straight across	downstroke
Word formation	Well formed	Well formed
constancy	regular	regular

Table 1: Feature comparison

Observation:

Except for pen pressure, staff of a and exit stroke all other feature values are the same. Thus these two images are similar with the label 1.

Matching features: Letter spacing, size, dimension, lowercase, continous, slantness, tilt, entry stroke, formation of n, staff of d, word formation and constancy.

Example for dissimilar writers:

Example 2: Let us consider two and images from the seen dataset: **0901a_num1** and **0786a_num2.**

These images have the label 0 meaning they are dissimilar..

Reason for dissimilarity:

Let us the consider the features for image 1 and image 2.

	0901a_num1	0786a_num2
Pen pressure	Strong	Strong
Letter spacing	High	Medium
size	Large	Large
dimension	High	Medium
lowercase	Yes	Yes
continous	No	No

slantness	normal	normal
tilt	normal	normal
Entry stroke	No stroke	No stroke
Staff of a	retraced	retraced
Formation of n	normal	normal
Staff of d	No staff	retraced
Exit stroke	No stroke	No stroke
Word formation	Well formed	Well formed
constancy	regular	regular

Table 2 : Feature comaprison

Observation:

The dimension and the letter spacing seems to be different in both the images resulting the images to be dissimilar with a label 0.