Handling Data Scarcity Through Data Augmentation in Training of Deep Neural Networks for 3D Data Processing

Akhilesh Mohan Srivastava, Indian Institute of Technology, Indore, India

https://orcid.org/0000-0002-6812-446X

Priyanka Ajay Rotte, Indian Institute of Technology, Indore, India

Arushi Jain, Indian Institute of Technology, Indore, India

Surya Prakash, Indian Institute of Technology, Indore, India*

https://orcid.org/0000-0001-8039-1280

ABSTRACT

Due to the availability of cheap 3D sensors such as Kinect and LiDAR, the use of 3D data in various domains such as manufacturing, healthcare, and retail to achieve operational safety, improved outcomes, and enhanced customer experience has gained momentum in recent years. In many of these domains, object recognition is being performed using 3D data against the difficulties posed by illumination, pose variation, scaling, etc. present in 2D data. In this work, the authors propose three data augmentation techniques for 3D data in point cloud representation that use sub-sampling. They then verify that the 3D samples created through data augmentation carry the same information by comparing the iterative closest point registration error within the sub-samples, between the sub-samples and their parent sample, between the sub-samples with different parents and the same subject, and finally, between the sub-samples of different subjects. They also verify that the augmented sub-samples have the same characteristics and features as those of the original 3D point cloud by applying the central limit theorem.

KEYWORDS

3D Biometrics, 3D Data Processing, Computer Vision, Data Augmentation, Object Recognition, Sampling

INTRODUCTION

Object recognition is an important topic of research in computer vision. There are many fronts in which the work is going on in this field. Some of the important tasks, for example, include analysis of the quality of training images on the recognition performance, security of the images used in the recognition systems, and training of the recognition models with the availability of limited training data, etc. There are studies such as the one proposed in (Alsmirat et al., 2019) which analyse the impact

*Corresponding Author

of quality of images on the recognition performance for a fingerprint based biometric recognition system. In (Chuying et al., 2018), an attempt is made to propose a few algorithms for securing the images while using them in systems and devises. In this work, we analyse the problem of training of a recognition model in the availability of limited data.

Nowadays, many of the object recognition systems are using 3D data instead of 2D. In such systems, availability of limited 3D data makes it challenging to achieve satisfactory recognition performance by the system. The use of 3D data is due to the fact that the object recognition performance on 3D data is significantly better than that on the 2D data. For example, in the case of face recognition, 2D face recognition is hindered by pose, expression, and illumination variations. These limitations are overcome when using 3D data as all the information about the face geometry is processed in the case of 3D based approaches. Given the significance and vast applications of 3D data in areas like object recognition, biometrics, it becomes important to address the issues faced during the training of the deep neural network model. Although the 3D object recognition achieves great accuracy, 3D data collection from objects takes time, and due to this there is relatively limited data available for 3D objects. In the presence of limited data, the model learns the details and noise of these few samples so closely that it has a negative influence while evaluating the selected model on new data. To avoid overfitting, we must increase the variability of the 3D data by increasing the size of the database through data augmentation. There are different ways to represent and input 3D data to a model. Some common and popular ways of representing an object in 3D include 3D voxel and point cloud. The 3D voxel representation is a highly regularized form of representation. In this representation, a 3D object is represented by discretizing its volume where the unit cubic volume is called a voxel. This representation has an advantage as it simplifies weight sharing and other kernel optimizations. However, it is bulky in nature with sparse data spaces and involves convolution operations that renders this representation computationally and spatially expensive. Further, capturing fine structures require a very high voxel resolution, consuming a massive amount of memory. On the other hand, point clouds are the rawest form of 3D data and are the direct outcome of the object scanning process. In point clouds, a 3D object is represented by digitizing its surface in the form of an unordered set of data points which can be directly consumed as inputs to any deep neural network instead of transforming them into regular 3D representations such as 3D voxels. As stated above, the 3D input data for an object which is in the form of a point cloud, contains an unordered set of 3D points. It is seen that this original set of points for an object contains a huge number of 3D points; however, due to the computational and memory limitations of the system, often, we cannot use the entire point cloud of a single sample for processing. To mitigate this problem, usually, the original point cloud data is sub-sampled, and a reduced size cloud is used for processing. However, in this process, the number of samples for a subject remains the same as was available earlier before sampling. We exploit the use of sampling in a different way and propose its use in data augmentation by increasing the number of samples of the subjects. In this paper, we propose three sampling techniques that can be used for creating sub-samples from an original point cloud sample. We use the Iterative Closest Point (ICP) (Chetverikov et al., 2005; Procházková & Martišek, 2018; Wang & Zhao, 2017) algorithm to show that the samples created from the original data all carry the same information. Then, we use Central Limit Theorem (CLT) (Heyde, 2014) to prove that the information carried by the sub-samples is the same as that carried by the original sample, that is, they have the same discriminative power. Finally, we compare the three sampling techniques based on the results.

The rest of the paper is organized as follows. Section 2 presents related work on 3D data augmentation techniques. The next section describes the proposed data augmentation techniques. The outcomes of the experimental evaluation of the proposed techniques are presented in Section 4. Finally, the paper is concluded in the last section.

RELATED WORK

Deep Learning Neural Networks are effectively implemented in many different fields such as medical, health care, cyber-physical systems, Internet of Things (IoT) and biometrics, etc. In the medical healthcare environment, such as in UbiHealth, a distributed application that supports healthcare demands in a ubiquitous computing environment is implemented. In (Sarivougioukas & Vagelatos, 2020), it is stated that deep learning networks can be utilized for carrying out the aggregation of huge information which is of complex nature to make efficient medical diagnoses in case of distributed environment. On the other hand, IoT is making smart devices accessible in the lives of common people, and attracting lots of work (Tewari & Gupta, 2017, 2020), (Sejdiu et al., 2020). However, this is also making them exposed to dangerous malwares such as botnets which are a threat for normal online applications. These botnets are serious threat for cybersecurity and their spread must be diagnosed and checked for the security and reliability of IoT devices. In (Letteri et al., 2019), authors have discussed a deep learning-based botnet detection technique which has been tested on a software-defined networking-specific dataset and has achieved a highly encouraging classification accuracy. Deep Learning has also been utilized in detection of malicious attacks. In (Nagisetty & Gupta, 2019), authors have proposed a deep learning framework for detection of suspicious activities in IoT backbone networks. For the prediction of malicious attacks, this framework uses four deep learning models, *viz*. Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), and Autoencoder. These models have been evaluated on two network traffic datasets, UNSW-NB15 and NSL-KDD99, and results have been compared with the same obtained using logistic regression and support vector machine (SVM). In is shown in the experiment that the deep neural networks-based model achieves the highest performance. Deep learning has recently been employed in different NLP applications, such as speech recognition, parts of speech tagging, analysis of emotions and automatic information retrieval. In (Lv et al., 2020), authors have proposed one such solution for NLP applications based on deep learning.

In this work, we address the common challenge faced by any deep neural network model which is the availability of sufficient data required for its training. We address this issue in the light of deep neural networks used for 3D point cloud data-based applications. There are many different strategies, such as padding, cropping, and flipping, which are used for augmenting 3D data in deep learning. These strategies improve performance of the underlying data-driven deep neural network model up to quite extent. (Iwasaki & Yoshioka, 2019) have described one such method which involves the use of Stereolithography (STL) data describing the surface of an object as a triangular mesh. The algorithm automatically generates a set of training data that covers various backgrounds and a continuous range of view angles. It applies two convolutional neural networks for improving the tolerance of the model against over-classification, increasing the performance over conventional methods. Recent work of data augmentation has also been done in the field of hand pose estimation. Even though, deep learning-based methods have significantly improved the performance of hand pose recognition, limitations still remain due to lack of large datasets. Data augmentation strategies used to solve this problem mostly apply image transformation methods such as translation, rotation, scaling, and mirroring (Oberweger & Lepetit, 2017; Xiong et al., 2019; Yang et al., 2019). Training images, in case of colour-based methods, have been augmented by a process of adjusting the hue-channel for the coloured data in (Yang et al., 2019). Ge et al. in (Ge et al., 2018) have proposed a 3D transformation for data augmentation in depth-based methods. The transformation involves randomly rotating and stretching of the 3D point cloud for synthesizing the 3D data. (Hinterstoißer et al., 2018) generate augmented data by using training samples that are rendered from 3D models. The method has the limitation of over-fitting as the synthetic data does not have the distribution similar to that of the real data and hence, requires carefully designed training process. In (Zhang et al., 2020), authors have proposed a method to synthesize image data for augmenting the training process of the neural networks called HandAugment. The method uses a scheme of two-stage neural networks to improve

the performance. It also introduces an effective way for synthesizing the data by combining real and synthetic images in the image space together.

Biometrics is one of the important applications of 3D object recognition due to its significance in the field of security. Hence, to create strong biometric security systems, robust data augmentation methods are required. One of the earlier approaches of 3D biometrics recognition involves Principal Component Analysis to extract features from the 3D surface for 2D and 3D biometric image identification (Taertulakarn et al., 2016). In (Ganapathi et al., 2018), authors have proposed a technique that uses 2D and 3D ear images for biometric recognition that uses local feature detection and description. The proposed model achieves a remarkable accuracy of 98.69% on the UND-J2 dataset. To maximize recognition performance on limited data, a technique with a combined training strategy to train the classifier based on a balanced mixture of general and application-specific data has been developed in (Zeng et al., 2019). Research carried out in (Patil et al., 2015) has given a detailed overview of recently used 3D face recognition databases, algorithms, features, and challenges faced due to variations in expressions, poses, and occlusions. The work in (Lei et al., 2016) has presented an efficient 3D face recognition approach for addressing the problem of partial data such as corrupted data, occlusions, or single training sample. For making the 3D object recognition algorithms robust to external factors such as light, expressions, or attitude, a depth-learning based approach has been proposed in (Luo et al., 2019). Using the depth information of the 3D face scans, it reduces the impact of external factors. A deep twin neural network has been proposed in (Xu et al., 2019) to overcome the limitations of the non-availability of a large number of 3D face scans for training the model. The proposed technique blends the 3D depth and 2D texture of the face samples and uses a convolutional twin neural network for 3D face recognition. Improvement in the face recognition is achieved by using the caricatures that exaggerate distinctive features of the face. Recently, a 3D automated caricature-based face recognition technique has been proposed in (Neves & Proença, 2019) which obtains 3D structures from 2D images of the face scans and achieves competitive results. Authors in (Zulqarnain Gilani & Mian, 2018) have proposed a process to generate a large corpus of labelled 3D face scans for training the model and a solution to merge existing 3D databases for testing. The work in (Kim et al., 2017) has proposed a Deep CNN and a 3D technique for augmentation that incorporates several different facial expressions from a single 3D face scan.

Robotic manipulation and perception also require applications of 3D object detection and hence, data augmentation. This is achieved by generating a synthetic dataset from its 2D and 3D local features (Yun et al., 2017). In order to take the full advantage of volumetric information, usually hidden in the depth data, a view-based 3D model is constructed from a single depth image in (Caglayan & Can, 2017). Often, 3D feature information is lost while being converted to their respective representations in voxel form. For overcoming this, a new rotation-invariant feature technique has been proposed in (Braeger & Foroosh, 2018) based on mean curvature. This technique improves the recognition on voxel CNNs as well as increases the overall accuracy on ModelNet10 dataset by 1%.

Despite recent advancements, the development in the field of 3D object recognition and 3D biometrics has been hindered due to the saturation caused by limited gallery size of 3D databases and hence it requires efficient ways of augmentation to generate required amount of data for training. Our proposed techniques of data augmentation are based on point clouds of 3D data. The deep learning-based approaches on 3D data mainly make use of volumetric CNNs, the inputs to which are highly regularized in the form of image grids or 3D voxels (Maturana & Scherer, 2015; Qi, Su, Nießner, et al., 2016; Zhirong Wu et al., 2015). These representations are computationally intensive in nature. Whereas point clouds, which are unordered sets of points, demand less memory and computational requirements. The point cloud features encode the given set of 3D points such that they are invariant to certain extrinsic (Rusu et al., 2009) and intrinsic (Aubry et al., 2011; Bronstein & Kokkinos, 2010) transformations. Also, point clouds have an added advantage of invariance to transformations like translation and rotation (Qi, Su, Mo, et al., 2016).

In this paper, we have proposed the use of different sampling techniques for augmenting the data. The proposal is highly useful when it is not possible to collect enough data required for training of a deep neural network model. The work presented in (Taherdoost, 2016) has presented various types of sampling techniques and the differences between these techniques that are taken into consideration to select the proper sampling method. We have proposed three sampling techniques to be used in augmenting 3D point cloud data, namely, random sampling, systematic sampling, and stratified sampling. Further, we make use of ICP algorithm (Chetverikov et al., 2005; Procházková & Martišek, 2018; Wang & Zhao, 2017) and CLT (Heyde, 2014) to prove that the sub-samples created from our original samples for the purpose of data augmentation all carry the same information and that they have the same discriminative power as possessed by the original sample.

PROPOSED TECHNIQUE

In this paper, we use three different types of sampling approaches for augmenting the 3D data. This generates different subsets or sub-samples from the original samples. As the original samples are in the form of 3D point clouds, number of points in the point cloud will be considerably less in the sub-samples, which will make the training of deep neural network model computationally and spatially efficient. We also prove, using ICP and CLT, that no information is lost while sub-sampling the point clouds. Given a 3D point cloud, we use the following types of sampling approaches for data augmentation:

- **Random Sampling:** In this sampling, each member of the set has an equal unbiased opportunity of being chosen as a part of the sampling process. In random sampling, to create multiple sub-samples from a single sample, we randomly select a fixed proportion of points from the original 3D point cloud multiple times. This creates different unordered subsets containing a uniform number of 3D points. We are selecting one-third of the original number of points from each sample point cloud to carry out the sampling process to generate the sub-samples.
- **Systematic Sampling:** Systematic sampling is a probability-based sampling technique where sample members from a population are selected from a random starting point but with a periodic and fixed sampling interval. This technique eliminates the chances of clustered selection. In this technique, we sort the point cloud of a sample by ordering the points in 6 possible arrangements - (x, y, z), (x, z, y), (y, x, z), (y, z, x), (z, x, y), (z, y, x). For each arrangement, we choose a random starting point in [0, k-1] and choose the subsequent points after skipping k, 2k, 3k... points where k lies in the range [3,5] depending on how crowded or sparse we want our sub-samples to be. Lower k results in a lower variance of points among different sub-samples while higher k results in less repetition but sparser point clouds. However, we need to ensure that the chosen k is not symmetric about the point cloud as this will result in the same sub-sampled point cloud irrespective of the ordering arrangement. We are making use of k = 3 so that the sub-samples use one-third of the point cloud.
- **Stratified Sampling:** Stratified sampling divides the total population into smaller groups for carrying out the sampling. These groups are formed based on some common properties existing in the population. After dividing the population into groups, random selection of the samples is performed proportionally. In this technique, we divide the entire point cloud of a object sample into cubical windows of fixed size and then select a proportionate number of points randomly from each window to create a single sub-sample. Hence, a higher number of points are selected from a dense region whereas a lower number of points are chosen from a sparse region thus maintaining localization. We are making use of a window of size 5 x 5 x 5 and select one-third of the total number of points from each window to carry out the sampling. Sampling is performed multiple times to generate multiple sub-samples from the point cloud of an object.

We use ICP algorithm and Central Limit Theorem to prove that the sub-samples created from the original samples (the original 3D point cloud of the object) all carry the same information and that they have the same discriminative power as possessed by the original sample.

The ICP algorithm finds a transformation matrix between two-point clouds by minimizing the square errors between them. One of the point clouds (target) is fixed, and the other one (source) is transformed to best match the target. The algorithm is iterative and improves the transformation matrix to minimize the error. Finally, it returns the final error after the transformation along with the transformation matrix. The error is essentially the registration error between the two-point clouds which indicates how dissimilar the information carried by two-point clouds are. For very similar point clouds, the registration error is very close to zero. The registration error can be used to find the similarity between the generated 3D sub-samples in the following ways for all the proposed three augmentation techniques:

- **Intra-sample Registration Error:** For a given sample, we find the registration error between the sub-samples created from that sample, as well as between each of the sub-sample and the original sample. Since, all the created sub-samples carry the same information, the error in first case should be very close to zero while in second case, it should be similar for all the sub-samples.
- **Inter-sample Registration Error:** For a given subject, we find the registration errors between the sub-samples created from the two different samples. For example, Subject 1 has two available samples - Sample 1 and Sample 2. We create 3 sub-samples each out of Sample 1 and Sample 2. Now, we find the registration error between the sub-samples of Sample 1 with each of the sub-samples of Sample 2. These should yield similar values for each combination which should be close to the original registration error between Sample 1 and Sample 2, verifying that the sub-samples carry the same features as well as they inherit the features of their parent sample.
- **Inter-subject Registration Error:** For two different subjects, we find the registration errors between the sub-samples created from a sample of each subject. This method is similar to the previous one except that we are using sub-samples of different subjects instead of sub-samples of different samples from the same subject.

Heyde in (Heyde, 2014) states that for any kind of data with a high number of samples, sampling distribution's mean and standard deviation should be equal to the population mean and population standard deviation divided by the square root of the total number of samples or the sampling size. We calculate mean and standard deviation for population as well as sampling distribution to verify the CLT. Using this, we prove that the discriminative power of samples, in our case the sub-samples created from the original point cloud of the object, is same as that of the population, i.e., the original point cloud of the object.

Therefore, using ICP and CLT, we show that no information is lost while sampling the data, and hence, sub-samples are effective to be used in training the deep neural network model. The proposed work attempts to increase the size of the limited 3D input data by using sampling techniques. The suggested augmentation techniques can be used on any class of 3D point clouds such as general objects, biometric modalities (faces, ears, etc.) since the sampling techniques are independent of the object classes to be recognized. The most obvious advantage of such data augmentation technique is that it overcomes the problem of overfitting due to limited training samples per subject. Moreover, due to multiple sampling from the same point cloud, not only do we increase the number of samples per subject, but we also reduce the time and space complexity while training the model since the number of points in each sample is reduced to a third of the original sample. This also ensures that features of the original data are retained as evident from the ICP registration errors of the sub-samples and their results on CLT as discussed in the next section.

EXPERIMENTS AND RESULTS

We use our in-house database, the IIT Indore Phase-3 (IITI) database, for the experiments. A few sample images from this database are shown in Figure 1. The IITI database contains 170 subjects, with a total of 445 samples where Artec EVA 3D scanner has been used to acquire the 3D facial scans. The database contains challenging samples where many of them are noisy and are not aligned properly. We augment these 3D facial samples using the proposed three sampling techniques and compare the results:

● **Intra-sample Registration Error:** We use three samples of each subject and create a set of three sub-samples for each of them. Table 1 shows the ICP registration error between a given sample and its respective sub-samples for all three samples for a subject. Similar experiment is repeated for all the subjects. Table 3 shows the average of means of Sample - Sub-sample error for all the samples (i.e., all samples of 170 subjects). Further, Table 2 shows the ICP registration error between all three pairs of sub-samples of each of the three samples for a subject. Similar experiment is repeated for all subjects and results are reported in Table 4 where it shows the average of means of Sub-sample - Sub-sample error for all the samples (i.e., all samples of 170 subjects).

From Table 1 we can see that the registration error between the original sample and its sub-samples is very similar for all the sub-samples while from Table 2 we see that the registration errors are very close to zero, verifying that all the sub-samples of a particular sample carry the same information. From Figure 2(a) (values are plotted in exponential scale for clarity) which is the graphical representation of Table 3 it is evident that the sample - sub-sample similarity is relatively highest (that is, registration error is the least) in the case of stratified sampling which can be explained because of the use of localization in selecting points. Systematic sampling has the next best similarity owing to ordering of the points before selection. Random sampling has the highest error because of the absence of any ordering or localization. Further, from Figure 2(b) (values are plotted in exponential scale for clarity) which is the graphical representation of Table 4, we can see that the sub-sample similarity is highest (that is, registration error is the least) in the systematic sampling because in this sampling technique, there is a possibility of repetition as we might choose the same set of points which is not desirable. For an effective sampling, we need low sub-sample similarity for more variation in the data after the augmentation. We see that this desirable characteristic is achievable in case of stratified or random sampling. Figure 2(c) provides a combined comparison of Figures 2(a) and 2(b).

Figure 1. 3D face samples from IITI 3D database used in experimental evaluation

Table 1. Sample - Sub-sample Registration Error (demonstration for one Subject) for Random, Systematic and Stratified Samplings

Sampling technique	Samples	Sub-sample 1	Sub-sample 2	Sub-sample 3	Mean
Random	Sample 1	4.18E-08	3.49E-08	3.66E-08	3.78E-0
	Sample 2	0.00E+00	0.00E+00	8.05E-09	2.68E-09
	Sample 3	1.83E-08	2.72E-08	2.17E-08	2.24E-08
Systematic	Sample 1	3.73E-08	3.73E-08	3.32E-08	3.59E-08
	Sample 2	1.13E-08	1.80E-08	0.00E+00	9.77E-09
	Sample 3	1.83E-08	1.41E-08	1.83E-08	1.69E-08
Stratified	Sample 1	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Sample 2	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Sample 3	0.00E+00	0.00E+00	0.00E+00	0.00E+00

Table 2. Sub-sample - Sub-sample Registration Error (demonstration for one subject) for Random, Systematic and Stratified Samplings

Sampling technique	Samples	Sub-sample 1-2	Sub-sample 2-3	Sub-sample 3-1	Mean
Random	Sample 1	7.89E-01	7.89E-01	7.82E-01	7.86E-01
	Sample 2	7.82E-01	7.95E-01	7.85E-01	7.87E-01
	Sample 3	7.87E-01	7.92E-01	7.85E-01	7.88E-01
Systematic	Sample 1	7.47E-01	7.35E-01	8.17E-01	7.66E-01
	Sample 2	7.50E-01	7.38E-01	7.34E-01	7.74E-01
	Sample 3	7.43E-01	7.34E-01	8.28E-01	7.68E-01
Stratified	Sample 1	7.87E-01	7.89E-01	7.85E-01	7.87E-01
	Sample 2	7.84E-01	7.84E-01	7.84E-01	7.84E-01
	Sample 3	7.82E-01	7.80E-01	7.83E-01	7.82E-01

Table 3. Average of mean of sample - sub-sample registration error over 170 subjects for random, systematic and stratified samplings

Sampling technique	Average Mean Sample - Sub-sample Error
Random	2.21E-08
Systematic	2.20E-08
Stratified	0.70E-08

Table 4. Average of mean sub-sample - sub-sample registration error over 170 subjects for random, systematic and stratified samplings

Sampling technique	Average Mean Sub-sample - Sub-sample Error
Random	0.7769
Systematic	0.7619
Stratified	0.7762

Table 5. Inter-sample registration error (demonstration for a pair of samples) for random, systematic and stratified samplings

Sampling technique	Samples	Sample 2	Sub-sample	Sub-sample	Sub-sample	RMS Error
Random	Sample 1 Sub-sample 11 Sub-sample 12 Sub-sample 13 RMS Error	8.4114 8.4472 8.4466 8.4452 0.0303	8.3418 8.3779 8.3765 8.3761 0.0459	8.4907 8.5268 8.5258 8.525 0.1068	8.44 8.4753 8.4756 8.4735 0.05674	0.0546 0.0704 0.0701 0.0692
Systematic	Sample 1 Sub-sample 11 Sub-sample 12 Sub-sample 13 RMS Error	8.4114 8.442 8.4421 8.4428 0.0270	8.4109 8.4427 8.4416 8.4418 0.0263	8.343 8.3743 8.3738 8.3739 0.0471	8.3984 8.4299 8.4293 8.4299 0.0171	0.0348 0.0301 0.0300 0.0302
Stratified	Sample 1 Sub-sample 11 Sub-sample 12 Sub-sample 13 RMS Error	8.4114 8.4470 8.4453 8.4446 0.0296	8.4079 8.4438 8.4411 8.4409 0.0265	8.3954 8.4308 8.4290 8.4283 0.0175	8.3869 8.4227 8.4212 8.4202 0.0150	0.0147 0.0266 0.0247 0.0242

- **Inter-sample Registration Error:** In this experiment, we are taking two different samples of the same subject, say Sample 1 and Sample 2. We also take the respective sub-samples, namely Sub-samples 11, 12, and 13 from Sample 1 and Sub-samples 21, 22, and 23 from Sample 2. Table 5 shows the ICP registration error between Sample 1 along with its respective sub-samples and Sample 2 along with its respective sub-samples. We consider the value of ICP registration error between Sample 1 and Sample 2 as the original value (mean) and find RMS (root mean square) error for each row and column. Similar experiment is repeated for all the subjects and results are reported in Table 7 where it represents the average of RMS error of Inter-Sample registration over all subjects. From the above experiment, we can see that the registration error obtained between the sub-samples is very similar to that obtained for the original samples. From Figure 3 (values are plotted in exponential scale for clarity) which is the graphical representation of Table 7, we can see that the error is lowest in the case of stratified sampling as this sampling makes use of localization while selecting the points. Random sampling has the highest error values among the three as the points are selected randomly from the point cloud without any specific ordering or localization.

- **Inter-subject Registration Error:** In this experiment, we take two different subjects, say Subject 1 and Subject 2. We also consider one sample and its respective sub-samples from each of the subjects. Table 6 shows the ICP registration error between Sample 1 along with its respective sub-samples from subject 1 and Sample 1 along with its respective sub-samples from Subject 2.

Similarly, we do the same experiment by comparing every subject with five different subjects and report the results in Table 8 where it represents the average of RMS error of Inter-Subject distance over all subjects. From the table, we can see that the registration error using the sub-samples is very similar to that obtained using the original samples. Further, from Figure 4 (values are plotted in exponential scale for clarity) which is the graphical representation of Table 8, it can be inferred that the stratified sampling and systematic sampling are comparable to each other and are better as compared to random sampling while comparing the inter-subject samples.

- **Analysis using CLT:** For five samples each from a different subject, we create 30 sub-samples each to apply CLT. As stated above, according to CLT, the average mean and the average standard deviation of the samples is similar to the mean and standard deviation of the original population.

Table 6. Inter-subject registration error (demonstration for a pair of subjects) for random, systematic and stratified samplings

Sampling technique	Samples	Subject 2	Subject 21	Subject 22	Subject 23	RMS Error
Random	Subject 1	17.9131	17.8738	17.8848	17.8028	0.0602
	Subject 11	17.9340	17.8950	17.9056	17.8241	0.0467
	Subject 12	17.9335	17.8947	17.9052	17.8238	0.0469
	Subject 13	17.9337	17.8949	17.9051	17.8236	0.0470
	RMS Error	0.0179	0.0252	0.0156	0.0950	
Systematic	Subject 1	17.9131	17.8630	17.9607	17.9561	0.0406
	Subject 11	17.9331	17.8829	17.9807	17.9761	0.0496
	Subject 12	17.9321	17.8820	19.9799	17.9752	1.0340
	Subject 13	17.9317	17.8819	17.9792	17.9743	0.0486
	RMS Error	0.0166	0.0366	1.0348	0.05792	
Stratified	Subject 1	17.9131	17.9367	17.9050	17.9156	0.0125
	Subject 11	17.9350	17.9588	17.9276	17.9373	0.0290
	Subject 12	17.9338	17.9582	17.9260	17.9363	0.0281
	Subject 13	17.9326	17.9564	17.9245	17.9351	0.0268
	RMS Error	0.0179	0.0405	0.0120	0.0950	

Table 7. Average of RMS of inter-sample registration error over 170 subjects for random, systematic and stratified samplings

| Sampling technique	Average RMS error of Inter-Sample Error
Random	0.1142
Systematic	0.0922
Stratified	0.0878

Table 8. Average of RMS of inter-subject registration error over 170 subjects (each compared with 5 other subjects) for random, systematic and stratified samplings

| Sampling technique	Average RMS error of Inter-Subject Error
Random	0.1297
Systematic	0.0969
Stratified	0.0988

Tables 9 and 10 verifies the CLT on the original samples and their sub-samples demonstrated using a set of five demo samples. Further, the average of MSE of CLT mean and CLT standard deviation for different sampling techniques when all 170 subjects of the database are used, are shown in Table 11. Results from this experiment prove that the sub-samples created from the original sample have the same discriminative power as the original sample. From the graphs in Figure 5 (values are plotted in exponential scale for clarity), we can infer that the stratified sampling is the best technique as it shows the least errors in mean and standard deviation for almost all cases.

- **Overall Assessment of the Sampling Techniques:** An experimental evaluation is performed to rank the three sampling techniques with respect to different criteria such as computational time, sub-sample similarity, sample - sub-sample similarity, coherence with the CLT etc. The results of this experiment are presented in Table 12. It is evident from the table that the stratified sampling is the best overall out of the three sampling techniques. Hence, it can be termed as
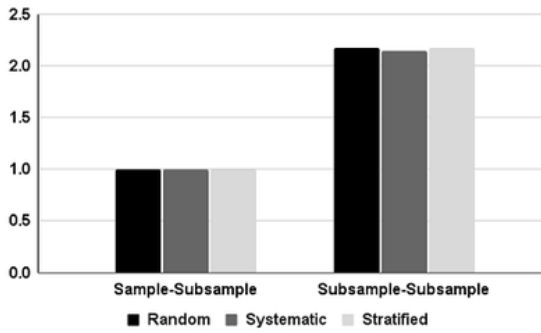
Figure 2. (a-c) Intra-sample registration error for random, systematic and stratified samplings (values are plotted in exponential scale for clarity)
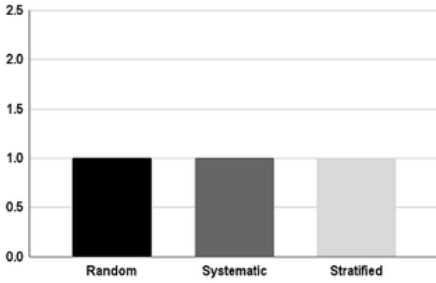
(a) Average Sample-Subsample Error
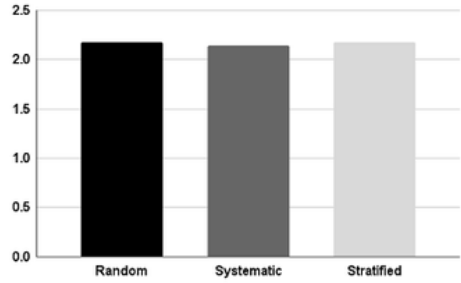
(b) Average Subsample-Subsample Error

(c) Intra-Sample Graphs Comparison

Figure 3. Inter-sample Registration Error for Random, Systematic and Stratified Samplings (values are plotted in exponential scale foe clarity)
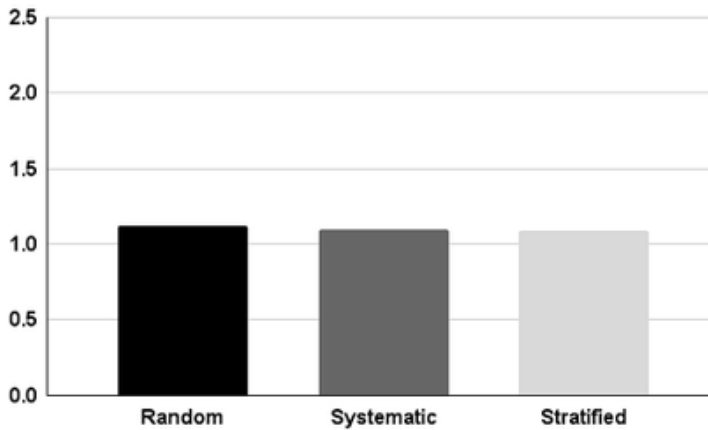
Figure 4. Inter-subject Registration Error for Random, Systematic and Stratified Samplings (values are plotted in exponential scale for clarity)
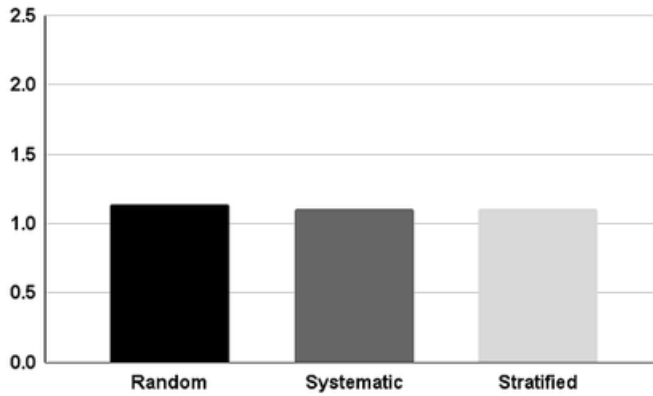

Figure 5. Average CLT Mean and Standard Deviation error for Random, Systematic and Stratified Sampling Techniques (values are plotted in exponential scale for clarity)
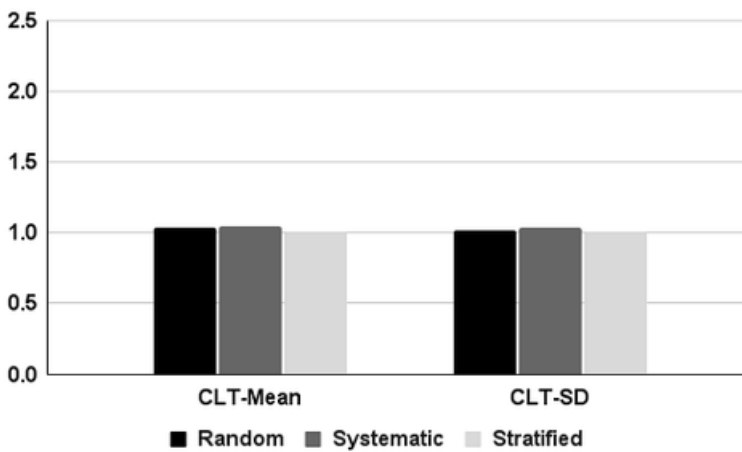

Table 9. Central limit theorem on mean (demonstration for five samples) for random, systematic and stratified samplings

Samples	Original	Random	Systematic	Stratified	Error (Random)	Error (Systematic)	Error (Stratified)
Sample 1	88.453	88.417	88.442	88.470	0.035	0.017	0.010
Sample 2	81.863	81.831	81.868	81.861	0.031	0.004	0.001
Sample 3	79.607	79.591	79.599	79.615	0.015	0.007	0.008
Sample 4	94.382	94.343	94.416	94.371	0.038	0.033	0.011
Sample 5	95.836	95.826	95.898	95.853	0.000	0.072	0.027

Table 10. Central limit theorem on standard deviation (demonstration for five samples) for random, systematic and stratified samplings

Samples	Original	Random	Systematic	Stratified	Error (Random)	Error (Systematic)	Error (Stratified)
Sample 1	70.075	70.086	70.111	70.082	0.011	0.036	0.006
Sample 2	53.880	53.860	53.865	53.884	0.019	0.014	0.004
Sample 3	60.754	60.781	60.783	60.765	0.026	0.029	0.011
Sample 4	61.548	61.525	61.533	61.548	0.023	0.015	0.000
Sample 5	55.900	55.896	55.916	55.884	0.004	0.016	0.016

Table 11. Average of MSE of CLT mean and CLT standard deviation over 170 subjects for random, systematic and stratified samplings

Sampling technique	Average MSE of CLT Mean	Average MSE of CLT Standard Deviation
Random	0.0324	0.0176
Systematic	0.0473	0.0297
Stratified	0.0088	0.0096

Table 12. Performance ranking of random, systematic and stratified sampling techniques

Sampling technique	Computational Time	Sub-sample Similarity	Sample-Sub-sample Similarity	Inter-sample Difference	Inter-subject Difference	CLT
Random	1	3	3	3	3	2
Systematic	3	1	2	2	1	3
Stratified	2	2	1	1	2	1

the best suited sampling technique for the data augmentation in various applications such as 3D object recognition and 3D biometric recognition.

CONCLUSION

This work proposes three sampling techniques which can be used for creating the sub-samples from an original point cloud sample for the purpose of data augmentation. We use the ICP algorithm to show that the samples created by the proposed technique from the original data all carry the same information. Further, we make use of CLT to show that the discriminative power carried by the sub-samples is the same as that possessed by the original sample. These sampling approaches decrease the number of points in the point clouds of the 3D objects, thus increasing the computational and spatial efficiency of the deep neural network used for training without loss of any information. Further, the data augmentation achieved through the sampling process overcomes the problem of overfitting of the deep neural network due to limited training samples per subject. At the end, experimental analysis has been carried out to rank the three sampling approaches with respect to different criteria. It is evident from the analysis that the stratified sampling technique is the best overall out of the three approaches that are being analysed and have potential to be used for data augmentation in applications such 3D object recognition, 3D biometrics etc.

REFERENCES

Alsmirat, M. A., Al-Alem, F., Yaser, M. A.-A. J., & Gupta, B. (2019). Impact of digital fingerprint image quality on the fingerprint recognition accuracy. *Multimedia Tools and Applications*, *78*(3), 3649–3688. doi:10.1007/s11042-017-5537-5

Aubry, M., Schlickewei, U., & Cremers, D. (2011). The wave kernel signature: A quantum mechanical approach to shape analysis. *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 1626–1633. doi:10.1109/ICCVW.2011.6130444

Braeger, S., & Foroosh, H. (2018). Curvature Augmented Deep Learning for 3D Object Recognition. *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, 3648–3652. doi:10.1109/ICIP.2018.8451487

Bronstein, M. M., & Kokkinos, I. (2010). Scale-invariant heat kernel signatures for non-rigid shape recognition. *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1704–1711. doi:10.1109/CVPR.2010.5539838

Caglayan, A., & Can, A. B. (2017). 3D convolutional object recognition using volumetric representations of depth data. *Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, 125–128. doi:10.23919/MVA.2017.7986817

Chetverikov, D., Stepanov, D., & Krsek, P. (2005). Robust Euclidean alignment of 3D point sets: The trimmed iterative closest point algorithm. *Image and Vision Computing*, *23*(3), 299–309. doi:10.1016/j.imavis.2004.05.007

Chuying, Y., Xuan, L. J. L., Xuechang, R., & Gupta, B. B. (2018). Four-image encryption scheme based on quaternion Fresnel transform, chaos and computer generated hologram. *Multimedia Tools and Applications*, *77*(4), 4585–4608. doi:10.1007/s11042-017-4637-6

Ganapathi, I. I., Prakash, S., Dave, I. R., Joshi, P., Ali, S. S., & Shrivastava, A. M. (2018). Ear recognition in 3D using 2D curvilinear features. *IET Biometrics*, *7*(6), 519–529. doi:10.1049/iet-bmt.2018.5064

Ge, L., Cai, Y., Weng, J., & Yuan, J. (2018). Hand PointNet: 3D Hand Pose Estimation Using Point Sets. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8417–8426. doi:10.1109/CVPR.2018.00878

Heyde, C. C. (2014). *Wiley StatsRef: Statistics Reference Online*. Academic Press.

Hinterstoißer, S., Lepetit, V., Wohlhart, P., & Konolige, K. (2018). *On Pre-Trained Image Features and Synthetic Images for Deep Learning.* ArXiv, abs/1710.1.

Iwasaki, M., & Yoshioka, R. (2019). Data Augmentation Based on 3D Model Data for Machine Learning. *Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 1–4. doi:10.1109/CCOMS.2019.8821676

Kim, D., Hernandez, M., Choi, J., & Medioni, G. (2017). Deep 3D face identification. *Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB)*, 133–142. doi:10.1109/BTAS.2017.8272691

Lei, Y., Guo, Y., Hayat, M., Bennamoun, M., & Zhou, X. (2016). A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample. *Pattern Recognition*, *52*, 218–237. doi:10.1016/j.patcog.2015.09.035

Letteri, I., Giuseppe, D. P., & Gasperis, G. D. (2019). Security in the internet of things: Botnet detection in software-defined networks by deep learning techniques. *International Journal of High-Performance Computing and Networking*, *15*(3-4), 170–182. doi:10.1504/IJHPCN.2019.106095

Luo, J., Hu, F., & Wang, R. (2019). 3D Face Recognition Based on Deep Learning. *Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1576–1581. doi:10.1109/ICMA.2019.8816269

Lv, X., Hou, H., You, X., Zhang, X., & Han, J. (2020). Distant Supervised Relation Extraction via DiSAN-2CNN on a Feature Level. *International Journal on Semantic Web and Information Systems*, *16*(2), 1–17. doi:10.4018/IJSWIS.2020040101

Maturana, D., & Scherer, S. A. (2015). VoxNet: A 3D Convolutional Neural Network for real-time object recognition. *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 922–928. doi:10.1109/IROS.2015.7353481

Nagisetty, A., & Gupta, G. P. (2019). Framework for Detection of Malicious Activities in IoT Networks using Keras Deep Learning Library. *Proceedings of the 2019 Third International Conference on Computing Methodologies and Communication (ICCMC)*, 633-637. . 3 doi:10.1109/ICCMC.2019.8819688

Neves, J., & Proença, H. (2019). "A Leopard Cannot Change Its Spots": Improving Face Recognition Using 3D-Based Caricatures. *IEEE Transactions on Information Forensics and Security*, *14*(1), 151–161. doi:10.1109/TIFS.2018.2846617

Oberweger, M., & Lepetit, V. (2017). DeepPrior++: Improving Fast and Accurate 3D Hand Pose Estimation. *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 585–594. doi:10.1109/ICCVW.2017.75

Patil, H., Kothari, A., & Bhurchandi, K. (2015). 3-D face recognition: Features, databases, algorithms and challenges. *Artificial Intelligence Review*, *44*(3), 393–441. doi:10.1007/s10462-015-9431-0

Procházková, J., & Martišek, D. (2018). Notes on Iterative Closest Point Algorithm. *Proceedings of the 17th Conference on Applied Mathematics (APLIMAT 2018)*.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2016). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 77–85.

Qi, C. R., Su, H., Nießner, M., Dai, A., Yan, M., & Guibas, L. J. (2016). Volumetric and Multi-view CNNs for Object Classification on 3D Data. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5648–5656. doi:10.1109/CVPR.2016.609

Rusu, R. B., Blodow, N., & Beetz, M. (2009). Fast Point Feature Histograms (FPFH) for 3D registration. *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*, 3212–3217. doi:10.1109/ROBOT.2009.5152473

Sarivougioukas, J., & Vagelatos, A. (2020). Modeling deep learning neural networks with denotational mathematics in UbiHealth environment. *International Journal of Software Science and Computational Intelligence*, *12*(3), 14–27. doi:10.4018/IJSSCI.2020070102

Sejdiu, B., Ismaili, F., & Ahmedi, L. (2020). Integration of Semantics Into Sensor Data for the IoT: A Systematic Literature Review. *International Journal on Semantic Web and Information Systems*, *16*(4), 1–25. doi:10.4018/IJSWIS.2020100101

Taertulakarn, S., Pintavirooj, C., Tosranon, P., & Hamamoto, K. (2016). The preliminary investigation of ear recognition using hybrid technique. *Proceedings of the 2016 9th Biomedical Engineering International Conference (BMEiCON)*, 1–4. doi:10.1109/BMEiCON.2016.7859620

Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *International Journal of Academic Research in Management*, *5*, 18–27. doi:10.2139/ssrn.3205035

Tewari, A., & Gupta, B. B. (2017). Cryptanalysis of a novel ultra-lightweight mutual authentication protocol for IoT devices using RFID tags. *The Journal of Supercomputing*, *73*(3), 1085–1102. doi:10.1007/s11227-016-1849-x

Tewari, A., & Gupta, B. B. (2020). Secure Timestamp-Based Mutual Authentication Protocol for IoT Devices Using RFID Tags. *International Journal on Semantic Web and Information Systems*, *16*(3), 20–34. doi:10.4018/IJSWIS.2020070102

Wang, F., & Zhao, Z. (2017). A survey of iterative closest point algorithm. *Proceedings of the 2017 Chinese Automation Congress (CAC)*, 4395–4399. doi:10.1109/CAC.2017.8243553

Wu, Z., Song, S., Khosla, A., Fisher, Yu., Zhang, L., Tang, X., & Xiao, J. (2015). 3D ShapeNets: A deep representation for volumetric shapes. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1912–1920.

Xiong, F., Zhang, B., Xiao, Y., Cao, Z., Yu, T., Zhou, J. T., & Yuan, J. (2019). A2J: Anchor-to-Joint Regression Network for 3D Articulated Pose Estimation From a Single Depth Image. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 793–802. doi:10.1109/ICCV.2019.00088

Xu, K., Wang, X., Hu, Z., & Zhang, Z. (2019). 3D Face Recognition Based on Twin Neural Network Combining Deep Map and Texture. *Proceedings of the 2019 IEEE 19th International Conference on Communication Technology (ICCT)*, 1665–1668. doi:10.1109/ICCT46805.2019.8947113

Yang, L., Li, S., Lee, D., & Yao, A. (2019). Aligning Latent Spaces for 3D Hand Pose Estimation. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2335–2343. doi:10.1109/ICCV.2019.00242

Yun, W., Lee, J., Lee, J., & Kim, J. (2017). Object recognition and pose estimation for modular manipulation system: Overview and initial results. *Proceedings of the 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 198–201. doi:10.1109/URAI.2017.7992711

Zeng, D., Spreeuwers, L., Veldhuis, R., & Zhao, Q. (2019). Combined training strategy for low-resolution face recognition with limited application-specific data. *IET Image Processing, 13*(10), 1790-1796.

Zhang, Z., Xie, S.-P., Chen, M., & Zhu, H. (2020). *HandAugment: A Simple Data Augmentation Method for Depth-Based 3D Hand Pose Estimation.* ArXiv, abs/2001.0.

Zulqarnain Gilani, S., & Mian, A. (2018). Learning from Millions of 3D Scans for Large-Scale 3D Face Recognition. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1896–1905. doi:10.1109/CVPR.2018.00203

Akhilesh Mohan Srivastava is a Ph.D. research scholar in the Department of Computer Science & Engineering, IIT Indore, India. He has completed his Master of Technology degree in Computer Science & Engineering from Dr. A.P.J. Abdul Kalam Technical University, India. His areas of research interest are biometrics, pattern recognition, computer vision, image processing, and machine learning. Biometric Recognition in 2D and 3D using Deep Learning for various modalities is one of his study interests.

Surya Prakash is currently an Associate Professor in the Department of Computer Science & Engineering, Indian Institute of Technology Indore, India. He received his MS and PhD degrees in Computer Science & Engineering from the Indian Institute of Technology Madras, India, and the Indian Institute of Technology Kanpur, India, respectively. His research interest includes image processing, computer vision, pattern recognition, biometrics, and identity and infrastructure management. He has published several research articles in peer-reviewed international journals and conferences. He has also co-authored two books entitled IT Infrastructure and Its Management published by Tata McGraw-Hill, India, and Ear Biometrics in 2D and 3D: Localization and Recognition published by Springer. He has also been on the program committees of several international conferences on the field of pattern recognition, image processing, and intelligent computing. For more information about him, please visit: https://www.iiti.ac.in/people/~surya/.