Sapienza
Università di Roma

Estimating shoe comfort from parametric foot shapes: a machine learning approach

Ingegneria dell'Informazione Informatica e Statistica
Corso di Laurea Magistrale in Computer Science

Candidate
Alessio Orlando
ID number 1792394

Thesis Advisor
Prof. Fabio Galasso

Academic Year 2021/2022

Abstract

The popularity of e-commerce and e-fashion has increased significantly in recent years and this has led to an increased focus on providing customers with accurate information about the shape and fit of various wearables. As a result, the importance of human body shape modeling has grown considerably.

To address this need, this thesis work aims to develop a model that accurately captures the shape of the human foot within a shoe and provides an estimate of how well it will fit users. Human body shape modeling, especially on wearable and in particular feet and shoes, is still an emerging field and, despite the significant advancements made in recent years, there is still a lack of public research available on the topic. To address this gap in knowledge and understanding, we have partnered with a company that is leading in this field and has managed, over time, to gather data on user foot shapes and fits.

This thesis work directly follows the initial work performed by the University of Cambridge that worked on the initial part of this work and, thanks to real customer data made available by Trya we worked on subsequent analysis and modeling.

While fit and comfort can be subjective, there is still much to learn and understand about them. This thesis work describes the exploration of multiple different approaches and the study of the behavior of each approach and model with different data features and aims to create a stable foundation for a good field exploration and future innovation. The proposed models aimed to optimize results for each type of model and each iteration aimed to tackle each time a new emerging challenge associated with the task. Starting from some baseline model this thesis work proceeds to analyze the behaviour and results of different procedures and different data processing such Transfer Learning, non rigid registration and skeletonization, with the scope of, not only understand the needs and the reaction of the models, but also creating a robust foundation to base future studies on.

Contents

Chapter 1

Introduction on fit analysis

As an expanding field, the wearable fit analysis relies heavily on characterizing the complex human body shape. This is essential for its successful application in various industries such as product design and clothing manufacturing. Anthropometry, the study of human body measurements and proportions, comprehends a huge variety of techniques to estimate different aspects of the human form. The traditional anthropometric methods are linear measurements, angular measurements, and circumferences. Furthermore, there is also force measurements that are used to estimate the strength and resistance of various body tissues or systems.

These manual measurements are prone to human errors. With the development of 3D scanning methods, anthropometric analysis started its transition from manual and traditional evaluation to a computerized image-based system. Image-based systems can offer an alternative to overcome some of the problems of traditional anthropometric measurement methods [1].

One challenge that can be encountered is the difficulty in establishing a real "true value" for the measurements obtained. This is due to the lack of standardized protocols in anthropometric research. While some protocols do exist, they tend to be defined in broad terms, leaving room for ambiguity and misinterpretation.[1] This, of course, can create inaccuracies in results, threatening the reliability of the data collected.

1.1 Anthropometry

Anthropometry is a research field that studies the shapes, sizes, and proportions of human bodies. It is an heavily utilized tool for handling the complex task of fitting products to the unique characteristics of individual users. By providing accurate data on the size and shape of body segments, anthropometry helps designers, and ergonomists to optimize product design and increase user comfort, efficiency, and productivity. In this way anthropometry plays a vital role in enhancing the overall quality and safety of products by integrating scientific insights into the design and development process. In the e-commerce environment, instead, it can be extremely useful for both enterprises, that managed to better convey users to a good fit product, and for the ultimate user that can trust more the information about user wearability.

1.1.1 Traditional measurement methods

There are many fields in which anthropometric measurements are vital, including ergonomics, sports science and fashion industries in general. Measurements of human body shape, size, and composition are used to design products, assess physical performance, and diagnose medical problems.

Linear measurements, such as width, height, and length, are among the most common anthropometric measurements used to describe the human body. These measurements can be taken manually using a tape measure or ruler, or can be taken with specialized tools. Angular measurements, instead, involve determining angles between planes and lines that intersect the body, are also essential in assessing human movement and range of motion.

Circumferences are another important anthropometric measurement that provides valuable information on body composition and health. Usually circumferences measurements include head, neck, waist, hip, and chest, but in specific use case we can encounter more granularity in those measurements such as the circumferences of a specific section of a foot or ankle. Force measurements, such as grip, pinch, are used not only to assess physical performance and functional capacity but also the possible clothing interaction and it's intensity. These measurements can be taken using handheld dynamometers, pinch gauges, and other specialized tools.

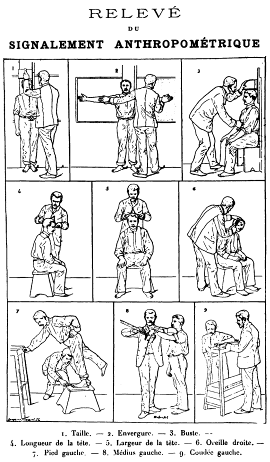


Figure 1.1. Bertillon's Identification anthropométrique (1893), demonstrating how to take measurements for his identification system with special tools.

1.1.2 3D scanning methods

For more than two decades, three-dimensional anthropometry has been utilized to accurately measure human body dimensions. There are various methods for collecting three-dimensional measurements, ranging from manual collection of 3D locations of body landmarks using electromechanical probe to 3D scanning of entire body surfaces. [4] With the advancement of technology, new and improved methods have been developed, such as 3D scanning using optoelectronic technologies. [1] 3D scanning involves the use of a light source, sensors, and a controller to capture several images of the body surface from different angles as a 3D point cloud [5][7]. Optoelectronic devices operate on three different principles, which include laser line scanners [1], structured light scanners [9] and multi-view camera systems[1].

The resulting of such measurements is usually a point cloud data that can then be processed by fully or semi-automated software functions to produce meshes [8]. Anthropometric data can also be extracted from 3D images using computer programs,[8] which have proven to be the most effective method for obtaining 3D models [5]. This method provides a high sampling rate and rapid measurement, allowing for accurate and efficient data collection. Additionally, the advancement of three-dimensional anthropometric measurement techniques and the ability to automatically recover them from accessible procedures and images contributed to fashion and commerce corporations' interest.

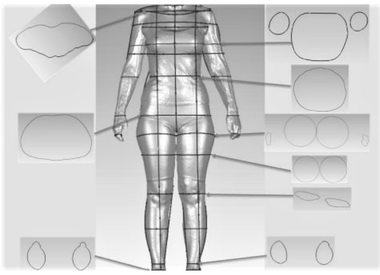


Figure 1.2. Section planes of human body according to ISO 8559:1989 (E) for extracting body dimensions. From [2]

1.1.3 Sampling issues

To ensure accurate representation of the target-user population in anthropometric research, it is essential to have an appropriate sampling plan in place. For anthropometric research, a good sampling plan involves determining the sample size, as well as determining the sample structure in terms of age, gender, race/ethnicity and occupational group. [1]

Additionally, it is really important to sample individuals at the extremes of the target population. This approach ensures that the data collected and applied to a specific task is appropriate for the target population.

Applying this approach requires an in depth understanding of the design requirements and population. It is necessary to invest efforts in sampling additional individuals, and on the application of suitable statistical techniques to ensure that the data collected is reliable and valid. This can be achieved with appropriate sampling plans. In this way researchers can collect accurate anthropometric data that can pint the development of products, designs, and analysis in the right direction.

Fitting criteria

When designing products based on anthropometric research, it is very important to consider how they will fit the users. The fitting process should utilize the appropriate standards to ensure that products are well-matched. However, fitting criteria that maximise matches between products/environments and users are rarely based on a single, nonadjustable design solution but are based on methods such as sizing systems and adjustability, which are generally adopted by HF/E specialists and designers. [10][11]

A critical factor that can affect the generalizability of the results are the number of samples in a study. If the sample size is too small, the data may not be representative of the target population, leading to biased or unreliable results.

The balance of data in a study is also equally important to ensure that the results are not biased towards any particular group. A balanced data set should include participants from different age, gender, race/ethnicity and occupational groups to ensure that the results really capture the target population. For example, if a study only includes participants from a specific age group or gender, the results may not be applicable to other age groups or genders.

In an anthropometric research, it is even more important to ensure that the user distribution is well-represented. The human body can be really different at different ages and more evidently when of different race. As said before it is equally important to take into account the number of samples and the balance of data to account for any possible biases that may appear.

1.2 Foot Anthropometry

A small section needs to be dedicated specifically to foot anthropometry. In literature there is still no consensus regarding an optimum technique for quantifying the foot features. Studies revealed that digitizer is a viable option that does not create lot of noise and remains, usually, fairly faithful to original measurements. Usually anthropometric measurements (and digitalization marks) are taken from few but very important points on the user foot like: malleolus, tip and heel and navicular. 1.3

Overall the foot anthropometry follows the problematics that has been discussed in previous sections such as the lacks of good and universal evaluation metrics and the sampling issues.

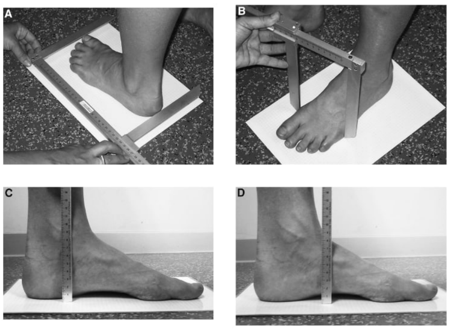


Figure 1.3. The anthropometric measurements: foot length (A), forefoot width (B), medial malleolus height (C), and navicular height (D). From [6]

Chapter 2

Related Works

Is is clear that there is not much research into foot fit analysis. However, in the last decade, some effort has been put into evaluating datasets and understanding the correlation between shoes, lasts and feet. As a result of this type of analysis, we have identified the key metrics that can be used to measure various aspects of the data and gained a deeper understanding of the challenges associated with working with human measurements and fittings.

2.1 Studies on Foot structures and Human Anatomy

Numerous studies have shown that creating a shoe that perfectly fits the wearer's foot is a non trivial task, and requires a thorough comprehension of human anatomy as well as an understanding of the impact that race and gender have on foot structure. This is not a simple task, and demands a high degree of expertise in the field. It has been proven that the shape of men's feet is generally larger than that of females in 7 of the 9 foot dimensions studied. Even tho the difference is little, the studies revealed that the gender differences on the measures of relative to foot length and absolute measures in common foot length categories should be taken into consideration for shoe last design and shoe manufacturing [13]. Comprehending the relationship between the features of the user and the design of the lasts is a crucial aspect of creating comfortable and properly fitting footwear.

It is also noteworthy that the use of men's lasts in the manufacture of women's shoes is a common practice in the industry. Most approaches fails to take into account the differences in foot structure between men and women, resulting in footwear that may not fit properly and analysis that not always represents the uncomfortable reality of the user [20][22]. Understanding that female shoes are often just male lasts adaptation is huge information that can be used for better analysis. This understanding can serve as a stepping stone towards better revising datasets and models, resulting in greater adaptability and generalizability in the future.

2.1.1 Analysis on Foot Datasets

Most of the work and analysis performed on those dataset has been performed using different statistical models like PCA, Cluster Analysis (CA), descriptive statistics (DS) and more, as shown by Iman Dianat and his crew. [1]

2.2 Related Works on Human Bodies

Research in the field of human foot comfort lacks a strong foundation in terms of metrics and state-of-the-art procedures. This can lead to difficulties in accurately assessing and addressing the various issues that affect the comfort of the human foot. In contrast, the field of generic Human Body studies has a much more robust foundation in terms of metrics and cutting-edge procedures, which enables a more accurate and insightful analyses. From the newest state-of-the-art approaches we can learn what would be useful and what can be used as a possible strategy in our study.

2.2.1 Segmentation Methods

Current methodologies relies on the possibility to segment body part such as trunk, leg, arms and head. This segmentation capability is the base foundation of Dress-up, a deep neural framework for image-based human appearance transfer [14]. The main scope of Hajer Ghodhbani work is to focus on using AI technology to develop an image-based virtual fitting system. With this is mind they worked with a pipeline that allows to extrapolate both clothes and user segmentation with respective keypoints and warp both shape and texture to fit as good as possible the original shape.

2.2.2 Comfort Fit Evaluation and via 3D Images and feature points

A different approach is instead executed for comfort fit specifically, especially on human body measurement estimation. Most of the established works used complex and in-depth cameras to obtain 3D images that then are used to estimate the body measurements.

Chang, and his team [19] proposed a dynamic fitting room that utilizes Microsoft Kinect and augmented reality technologies to allow individuals to visualize a real-time image of different digital clothes with the main scope of estimating the size as close as possible to users' claimed sizes.

Xiaohuia, et al. [18], instead, proposed method that aims to provide a way to extract feature points from 3D human bodies. The feature extraction and measurement estimation stage are usually serving as a preprocessing step for garment designer or virtual fitting application. Other research such as the Mojarrad and Kargar one [17] proposed an approach that examine various images of people and estimate the size, body measurements, and properties such as tall, fat, thin and so on. The model extracts features and detects the points and the edges of the body using canny edge detector approach. During their experimentation they found out that the proposed method was able to work on all of the human sizes short fat, tall thin, and short thin).

2.2.3 Comfort Fit Evaluation and via 2D Images

Latest works instead works on 2D image analysis such for example Chandra, et al. [16] that estimates the body part by calculating the difference between the most left-side and most right-side of any body part. The work recommends user to use specific environmental settings such as light and background color for a better experience, also the method only work on the upper part of the body and the process is quite slow due to the fact that the procedure need to be executed 5 times for the most accurate size.

Sahar Ashmawi, et al. [15] focuses their work on alleviating this problematic with a model that estimates human body measurements and predict their body size from 2D images taken from regular smartphones. Their model detects the human body from the images and extracts the features of the body from the picture. Then it determines the focal points in the human body, and calculates the body measurements by computing difference between the focal points. It then uses an SVM model to extrapolate the correct evaluation.

Those methods are surely essential for a better understanding of what is been tried and what we should aim for to achieve. Unfortunately most of the main concept and data manipulation used in those works cannot be translated as it is in our study due to differences in data generation and task, but knowing this procedures is surely good knowledge to a better understanding of the possible step that can be made.

2.3 Further Studies

In this thesis work has been decided to not only utilize statistical procedure to analyze the behaviour of the models but also to explore new seemingly untraveled roads such transfer learning from known and state of the art Cloud Point models such PointNet [24] or PointNet++ [26], and data manipulation of mesh data with deformable registration and alignment with a simple but really known and utilized algorithm such CPD [34].

More analysis and work need be done to really explore the skeletonization approach but, nevertheless, the work that Andrea Tagliasacchi and his team done on skeletonization [35] surely created a good foundation for this work and inspired new possibility and experimentation with supervised methods that relies on skeletonization for optimizing mesh deformation and alignment, such the work of Çağlar Seylan and Yusuf Sahillioğlu about 3D Shape Deformation Using skeletonization. [37].

Chapter 3

Dataset

The importance of data in machine learning, cannot be stressed enough. Data always presents a challenge, no matter what the task may be. The possibility and ability to manipulate and organize data effectively is often considered the key to solving machine learning problems. To achieve this, the initial step in any machine learning project is a detailed examination of the dataset to gain a complete picture of the state of the dataset.

As part of this examination, it is important to pinpoint any potential data-related issues that could impede the learning process. For example, there may be missing or inconsistent data, outliers, or biases that need to be addressed. An effective data exploration can also help to determine which variables or features of the data are most relevant and which one can be leveraged the most.

It is so essential to give priority to data analysis in most machine learning project, and to ensure that appropriate measures are taken if any data-related challenges appears and to perform a thorough analysis of the data. It is always necessary to build a solid knowledge base about the working dataset in order to truly grasp the needs and possible solutions of different problems and tasks.

3.1 Trya Dataset

As we previously mentioned ,in our work, we have made the decision to collaborate with a company that has devoted a portion of its lifespan to the task of extracting 3D models from user photos. This unique technology allows the company to extract detailed information that can be used to determine the optimal fit and size for each user's footwear.

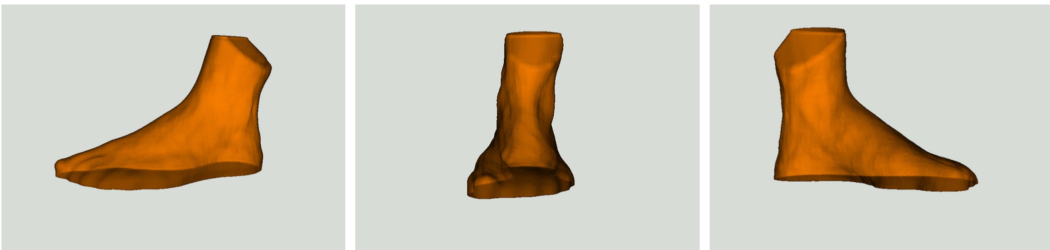
The company's dataset comprises two primary categories of mesh. The first is the user mesh, which provides detailed information about the user's foot shape and size, including the contours and curves of their foot.

The second mesh category is the last mesh. The last is the foundation on which the shoe is built and in the dataset is represented as a 3D reconstruction of the physical last.

3.1.1 User Mesh

The user mesh information are derived from a reconstruction procedure referred to as Snapfeet 1, which involves a series of steps executed on trya servers utilizing photographs captured by cameras, any cameras. The process utilizes algorithms such as the visual hull, along with supplementary information like a sheet of dots to determine the camera's position in space, and a considerable number of photos, typically around 20. Augmented reality libraries are not utilized in this process. Those instead are used in their new process called Snapfeet 2 but in our dataset the meshes comes only from their first procedure. The procedure used in Snapfeet 1 is made up of different steps:

- Under the foot is placed a piece of paper with dots. As a matter of fact, this piece of paper is used for calibration and ensures that there is a record of the foot orientation and placement with respect to the camera lens
- Then 20 photos are taken (simple photos, without any metadata about sensor and cameras)
- Inside Trya server are then executed the reconstruction of the tridimensional model. The reconstruction follows a few simple steps:
  - Calibration
  - Segmentation
  - Visual Hull
  - Depth Map
  - Surface extraction



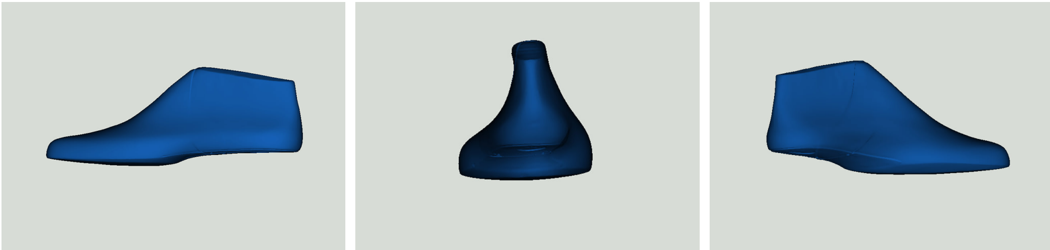
(a) (b) (c)

Figure 3.1. Foot Mesh Visualization

3.1.2 Last Mesh

In footwear manufacturing, lasts dictate the overall fit and comfort of the shoe. A last is essentially a "negative" of the actual shoe, a 3D object that replicates the shape and size of the internal volume of a shoe. It is used as a mold for constructing shoes. Note that in this dataset the companies that produce lasts are different from the ones that manufacture shoes. In this case the last comes mostly from an Italian company based in Brescia.

In the dataset, lasts have been generated with two different methodologies. Some are designed directly on the computer by a 3D designer, who uses specialized software to create a virtual model of the last. Others instead are real physical objects that are scanned through a process called precision scanning. This involves using specialized equipment to scan an existing last made of wood or resin, producing a digital 3D model.

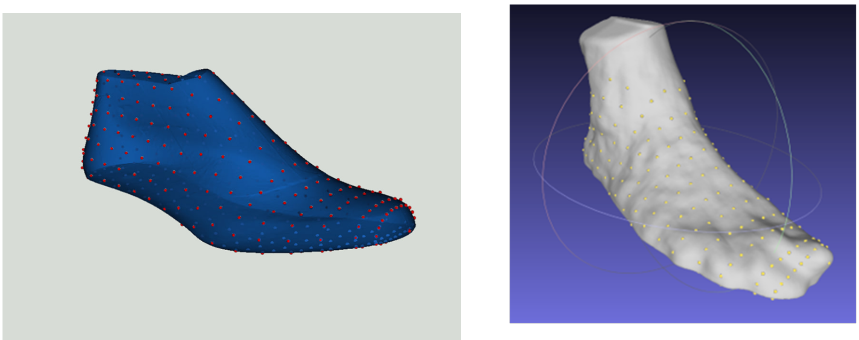


(a) (b) (c)

Figure 3.2. Foot Mesh Visualization

3.2 Cloud Points

An alternative representation of the Trya Dataset is represented by a Cloud Point. Cloud Points are created for each feature as a parameterization of the respective mesh. The foot length is resampled with a fixed number of points at a low density. Following some testing and empirical analysis, Trya divided the foot length into 17 slices and sampled 20 points on each contour at equal distances. To increase the length information, two additional points are added for the heel and tip. It is very useful for address the loss of information between mesh and cloud points.



(a) (b)

Figure 3.3. Foot Mesh Visualization with Cloud Point

3.3 Data Analysis

3.3.1 User Data Analysis

The gender of the user can provide useful information about the dataset. In order to avoid direct gender biases, this information isn't used as a feature, but it can still be used to understand if the dataset is unbalanced or if overall contains this kind of bias.

Even tho the user dataset is a little unbalanced such a small difference doesn't create much biases. Knowing this is really good since now we know that in this regards there is no work to be done to really use this dataset. For each user the dataset can contain features about both left and right foot. In most cases both feet are used for the trials but is not a mandatory assumption. Luckily the distribution

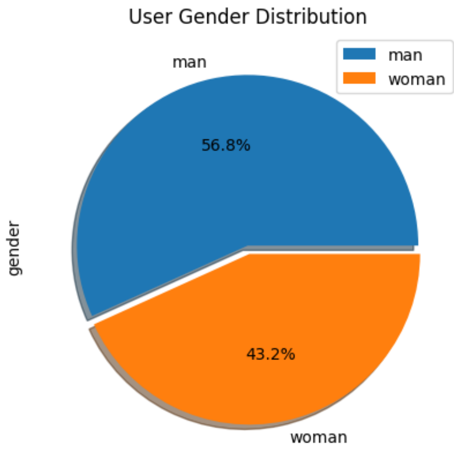


Figure 3.4. Distribution of gender in all user dataset

of left and right foot are fairly balanced with a difference of only 1.1%. We can say that we can avoid worrying about this aspect of the dataset.

Overall the dataset contains 613 users of different races of various ages. At the moment, the dataset does not contain the information about the different distribution of age and race. Studies found that age, race, and gender are really important factors when analyzing shoes and feet [13][20][22]. Knowing this information would be very helpful for understanding behaviour of models and address possible biases.

3.3.2 Last Data Analysis

The lasts dataset have not much information other than the last features. Each shoe have different sizes and for each size the dataset contains a different instance. There is no differentiation about material and kind, nor the information about model and origin.

In the dataset, for each shoe there are different sizes and each different size is a a different instance. Material and kind of shoe are not differentiated in the dataset, nor is there any information about model or region of origin (except of the region's size : EU, UK, US)

In total there are 2017 different model representation of 126 different unique shoes. In terms of features, there is little to mention. The dataset is short of unique shoes, and this can lead to less generalizability in future models. There are no irregularities in the distribution of shoe sizes within the dataset. 3.6

3.3.3 Trial Data Analysis

For obvious reasons, it is not surprising that not everyone tried every shoe, it is impossible to try every shoe size and type for example.

To be able estimate the fit and the comfort we need first some evaluation from user tests. This evaluation is performed by de-identifying the shoes so that the user is less biased about the characteristic of shoe type or color. The trial dataset consists of 14329 trials. Each trial has 5 different fit evaluations: Length fit, width fit, instep fit, heel fit, global fit. The first four fits are divided into 3 different categories: good, loose and tight. Listed here are the classes that describe how comfortable a person is in a particular part of the shoe: Length Fit describes how comfortable the user is

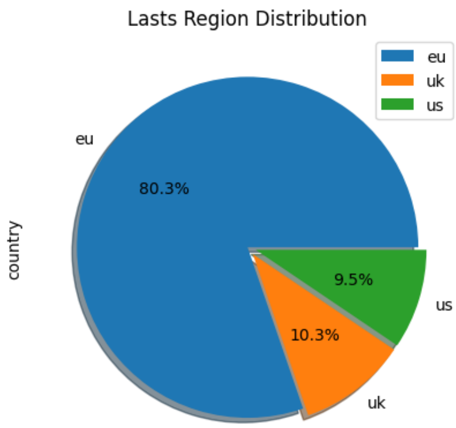
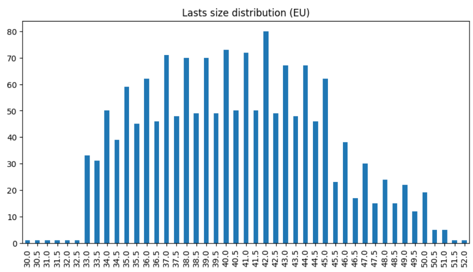
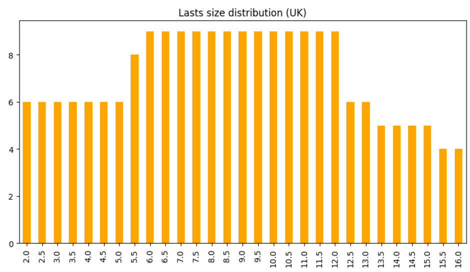


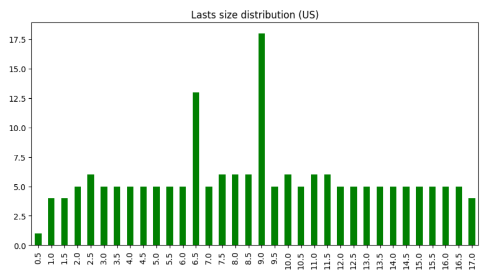
Figure 3.5. Distribution of size provenience in lasts dataset



(a) EU Distribution



(b) UK Distribution



(c) US Distribution

Figure 3.6. Size distribution for different Regions

with the length of the shoes, width refers to how comfortable the user is with the width of the shoes, etc.

Local Comfort

As we can see from the image 3.8, the distribution of these four fit classes is extremely unbalanced towards the "good" fit. Even tho it can be alleviated slightly with augmentation this issue will become a possible limit in future tests and it is a problem that needs to be addressed in future data acquisition.

Let's take a look at the different class distribution for each fit typology. As we can see, overall, the distribution is similar for each section of the shoes. Even tho it

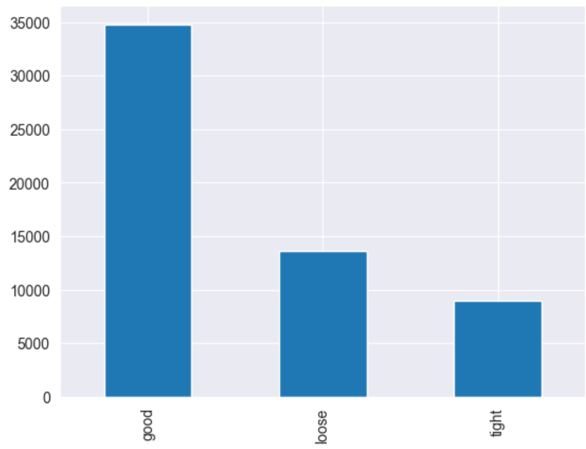


Figure 3.7. Distribution of class type in trial dataset

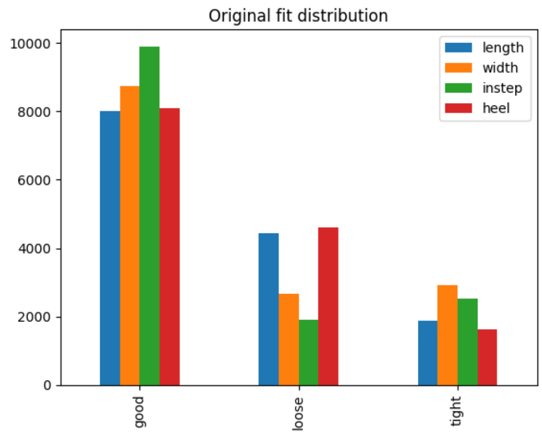


Figure 3.8. Distribution of class type for each fit in trial dataset

is similar we can still spot some little peculiarities. We can see how length and heel have a similar distribution, almost the same. This can be also seen in the correlation in the future evaluation metrics. Instep instead tends to be usually more comfortable and this lead to higher "good" fit valuation. Width fit is somewhat more balanced with a little trend towards the tight fit.

Global Comfort

Special mention and analysis need to be performed for the fifth fit type: Global fit. This is a special fit that does not represent the same type of information of the other four. First of all this type of fit does not represent directly the comfort of the entire shoe but instead it represent the will of the user to keep or return the shoe tried. This is an huge difference because now the user not giving an evaluation of how good the shoe fit but instead he is giving his subjective opinion about how he sees himself in those shoes and if he is feeling enough confident to buy that shoe based only on that sole trial, no other information.

Even tho the distribution of every other fit tends to a good fit here, the distribution is unbalanced towards the bad fit. This can be explained by the fact that an user

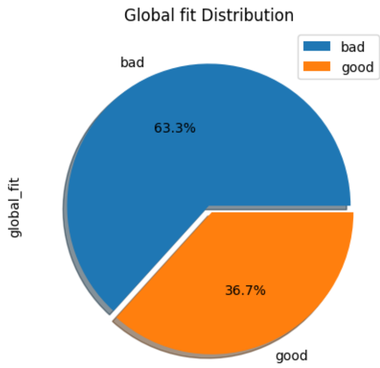


Figure 3.9. Distribution of class type for each fit in trial dataset

wille decide if a shoe is good not only basing his decision on if the shoes is good enough, but rather on the fact that, to be purchased, most of the time the shoe need to be perfect.

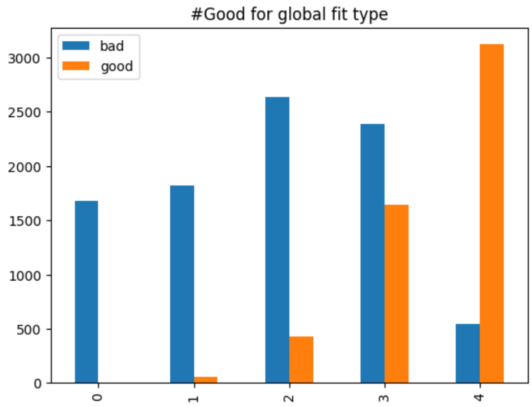


Figure 3.10. Distribution of class type for each fit in trial dataset

As we can see, even if in every local fit the user finds the shoe comfortable sometimes he will still decided to return the shoe. Even tho this is to be considered noise, the real disturb comes from those users that find the shoe mostly comfortable but not perfect. In a real world scenario those people will mostly decide with subjective metrics like color or type of shoes. Here instead the only metric used is comfort and this create a lot of noise for the model that find the decision unclear and have real difficulty at evaluate those classes. There is surely much study to perform to really grasp the factors that can influence this kind of evaluation but we can easily see how it is not only based on those 4 local fits but instead it certainly depends also on other factors that are not fully represented in the local fit dataset.

Features Analysis

To really grasp the needs of the task and how to utilize properly the data, an in depth analysis of the features is needed. In particular what it is really interesting to know is how well the user and last models interacts, what are the difficulties and the issues of the data. The analysis has been performed mainly on Cloud Point version of the dataset since is a direct representation of the 3d mesh one.

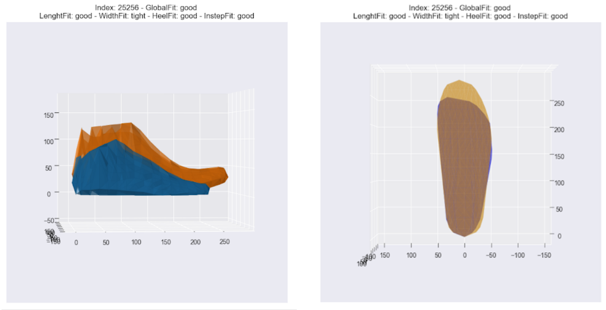


Figure 3.11. Good sample View

The data samples we examined are generally well aligned, as can be seen 3.113.12. Based on some analysis, it is evident that defining a tight or loose or good fit is not as straightforward as it appears. In this "good" sample 3.11 we can see a slight problem that we will encounter later on: There is a slight misalignment between the shoe and foot, also, the shoe is not flat on the floor while the foot, instead, is placed flat on the ground. A possible solution will be discussed in upcoming chapters.

This other sample, moreover, shows that for each slice the shoe and the foot are not sampled in the same spot. This suggests that the sampling or the conversion into Cloud Point is not uniform. The cross section figure shows also that the shoe and the foot do not align on the Length Axe and neither in the width axis. Due to the fact that the Trya team have already been tried to align the feature data, we chose not to work directly on this aspect and we considered the alignment to be good enough.

Data Augmentation

As mentioned before the dataset lacks of balance. There are different ways to mitigate this problem: the best way is to, obviously, add more data. Even tho it have the best possible outcome, is not always possible and in most case is really expensive. An inexpensive and still really useful procedure is to perform up sampling and down sampling on the dataset. While the down sampling method is the most secure and unintrusive one, is also a method that can overall effect negatively the performance of the tests. Down sample a dataset means to remove some samples from the dataset until there is balance between classes, or until the balance is satisfactory enough. This leads, of course, to information loss and on small dataset it becomes self-defeating.

A possible solution is performing an up sampling of the data. Usually the up sample is performed by creating new data by inferring the missing samples on the other samples of the dataset, or by simply duplicating the already existing samples. In our case we decided to manually add samples. We decided to perform this kind of evaluation by inferring that any user that considered a trial shoe as "good" for a specific size X will surely consider any size smaller and larger not good. With this in mind we created new samples that contained the user foot and the shoe one size larger and one size smaller with the respective fit values: tight for a smaller shoe and loose for larger shoes. To avoid adding too much noise we decided to only add 2 samples for each "good" user sample (+ and - 1 size).

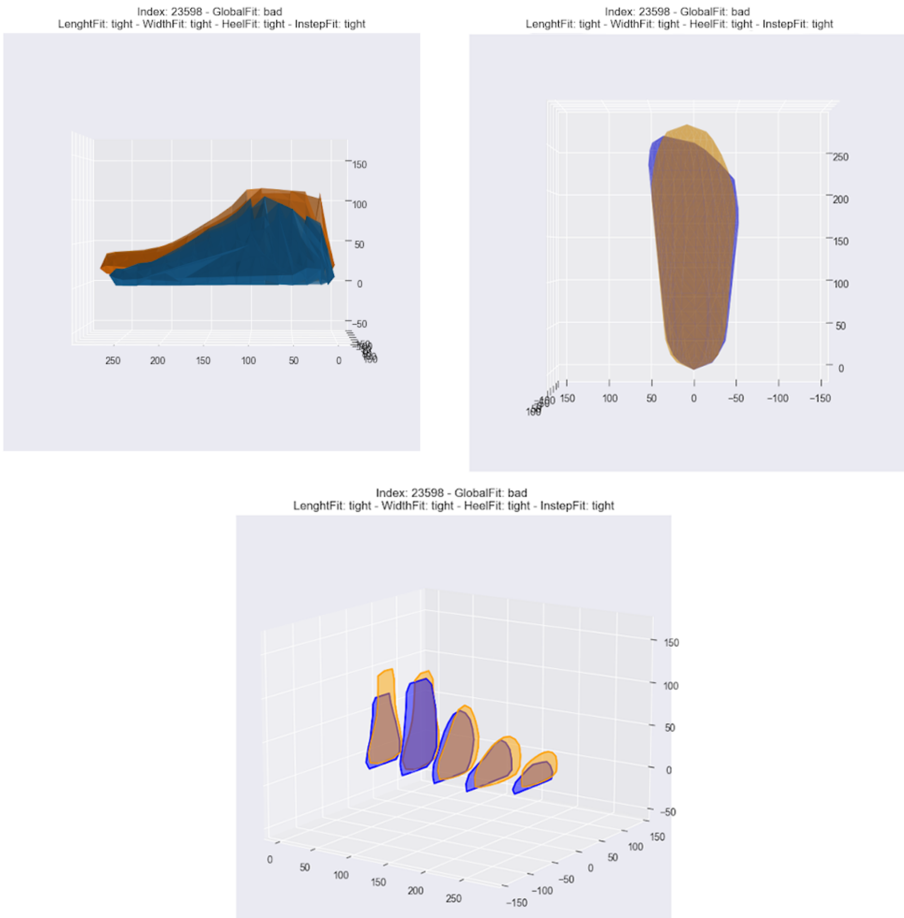


Figure 3.12. Tight sample View

Chapter 4

Models Description

This section aims to provide a brief overview of the motivation and requirements for each model.

Specifically, we will begin by discussing a PointNet implementation that utilizes the cloud point representation of the dataset. The main reason that lead to this trial is the data availability, the dataset represent one of the main obstacle in this work and on the first trials that is worth testing is taking advantages of a technique called Transfer Learning.

Once tested and understood the difficulties and necessity of this approach both in computational power and task needs, has been decided to start engineering our own solution to see and understand the behaviour of this new approach. The baseline consist in a "simple" MLP architecture with various layers of normalization and regularization.

The subsequent tests and trials utilizes different data preprocessing strategies to address every time different necessity that the tests uncovered. Such the necessity of increasing normalization and regularization, help the model with new engineered features and fix the alignment between lasts and user data.

Last tests are instead focused on a different way to introduce skeletonization information into the trials with the main scope of starting understanding the behaviour of such information with the main goal of utilize, in future, this information as a constraint for the deformation step.

4.1 PointNet Implementation with Cloud Points

A way to represent the dataset is as a collection of points. We have decided to use this representation in order to carry out the majority of the tests in this thesis work. We have chosen this approach based on the fact that, even tho the user foot data and the lasts are filled with lots of information that can be extrapolated, the dataset does not contains enough trials in order to be able to really generalize those information. We will see in future sections that the lack of adequate data is a recurrent problem and that most of the choices that are being made are possible workarounds which aim to mitigate the inability to generalize from the data directly. In the first step of the process, we plan to try a model that has been developed and modelled to work with cloud points, and in a subsequent step, we will try to apply a transfer learning method from it as well.

4.1.1 Input data

In this model version we used the CloudPoint representation of the dataset. A cloud point feature consisting of 1026 values is decomposed into 3 different feature vectors consisting of 342 values, one for each of the axes. After these vectors are obtained, they are then passed on to pointnet models to extract global vectors from them.

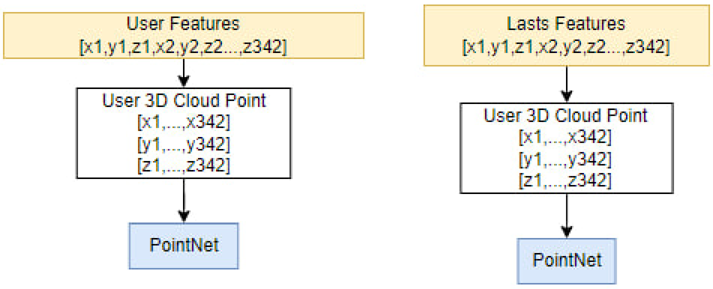


Figure 4.1. PointNet Data Pipeline

4.1.2 PointNet

In terms of efficiency and effectiveness, PointNet is an highly efficient and effective model that is robust to input perturbations and corruption, as well as having excellent performance. Adapts well to basic geometric transformations, such as rotation, as well as many other basic geometric operations. This is motivated by the STNs [23] the “input transform” and “feature transform” are modular sub-networks that seek to provide pose normalization for a given input. T-Net is a regression network that has the object to predict, originally, a 3x3 transformation matrix that is multiplied with the input. MLPs are then used to map the input independently and identically to a higher-dimensional space with MaxPools and Fully Connected layers.[24].

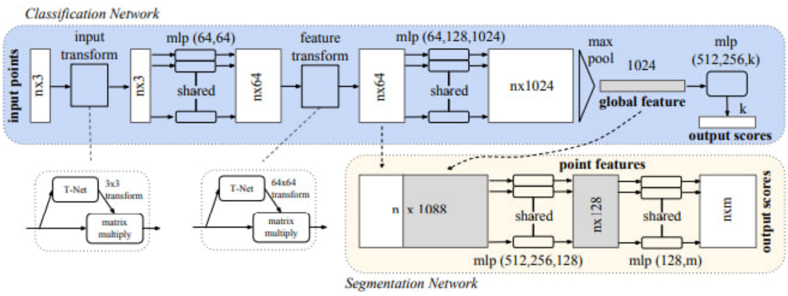


Figure 4.2. PointNet Model View from [24]

In PointNet the T-Net predicts a 64X64 transformation matrix. This increase in number of parameters leads to potential over-fits and instability during training. This has been solved in the PointNet model with the addition of regularization terms in the loss function. PointNet is also been trained with a bottleneck dimension (K) that represent directly the expressiveness of the model. [24] Naturally, a larger value of K leads to a more complex and accurate model, and vice versa. Pointnet is designed with a k of 1024 and with an input of 2048 points.

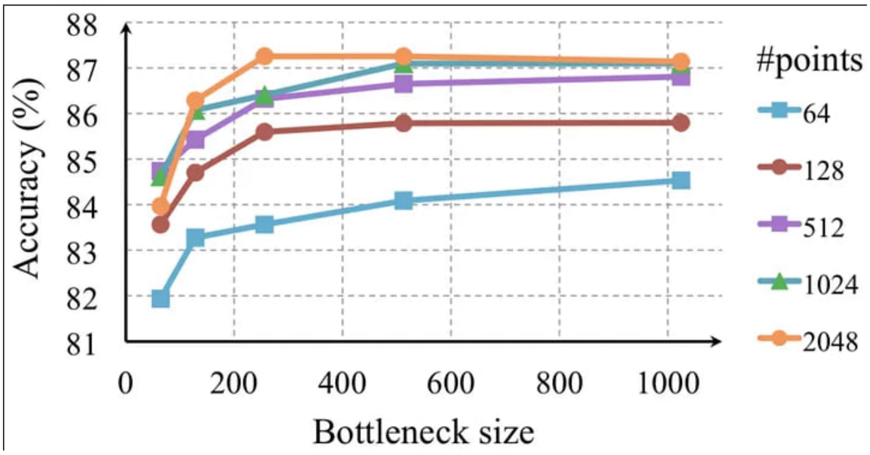


Figure 4.3. Accuracy for PointNet model with K equal to different values. From [25]

We intend to extract the global features from the Classification Network 4.2 and utilize these 1024 global vectors as inputs for newly created layers.

4.1.3 Transfer Learning

The target of a transfer learning approach is to take advantage of those task that have some similarity to other tasks. A task can usually be generalized and transferred, even if in a small part. If it can be transferred, it is better off starting training on a model that has already been trained on a similar task rather than training from scratch.

Of course there are also some cases where this practice is constrained by the shortage of data. In those cases transfer learning becomes a really powerful weapon. Training with a transfer learning approach is not as straightforward as attaching a network and run it but usually a small part of the network is frozen and the remaining part is trained. Although some tuning is required, this allows us to leverage previous engineers' work on similar tasks.

4.1.4 PointNet++

PointNet++ is born from the necessity to overcome the inability of PointNet to capture local structures and the consequent limit to recognize fine-grained patterns and to generalize complex scenes. In the work they introduce a hierarchical neural network that applies PointNet recursively on a nested partitioning of the input point set neural network that applies PointNet recursively on a nested partitioning of the input point set. [26]

In this thesis work we decided to try a transfer learning approach with this newly developed and updated model. This is due to the fact that, being new, it was possible to find numerous already trained models. The PyTorch implementation referenced above provided PointNet2 with MSG (multi-scale grouping) and SSG (single scale grouping). The MSG approach applies grouping layers with different scales to capture multi-scale patterns simultaneously. Extracting point-wise features with multi-scale receptive fields leads to a huge increase in computational consumption.

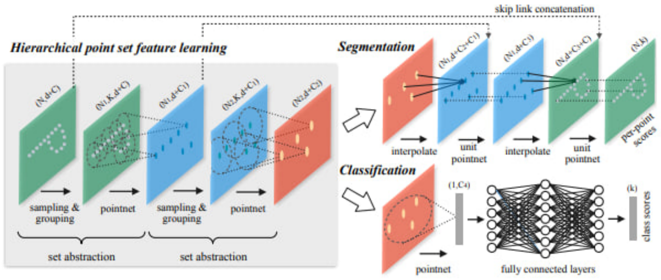


Figure 4.4. PointNet Model View from [26]

4.2 Multilayer perceptron with Cloud Points

The PointNet method, however, is not the only viable option. The lack of data and the difficulty to adapt a pretrained model to this task opened the necessity to engineer our own solution. As a baseline, it has been decided to work with Multilayer Perceptron on CloudPoints without preprocessing. During this baseline work, evaluations have been performed for all four local fits. The main goal of this work is to lay the groundwork for understanding the behavior of future models and how they respond to different input features.

4.2.1 Multilayer Perceptron

The Multilayer Perceptron was developed to tackle the limitation of applying the models to non-linear data. The MLP is a neural network where the mapping between inputs and output is non-linear. It has input and output layers, and one or more hidden layers with many neurons stacked together. The Multilayer Perceptron represent a good first approach in any machine learning task. It grants the flexibility to increase the depth and easily work with dimensionality to help the model in his learning path while maintaining the computational cost really low.

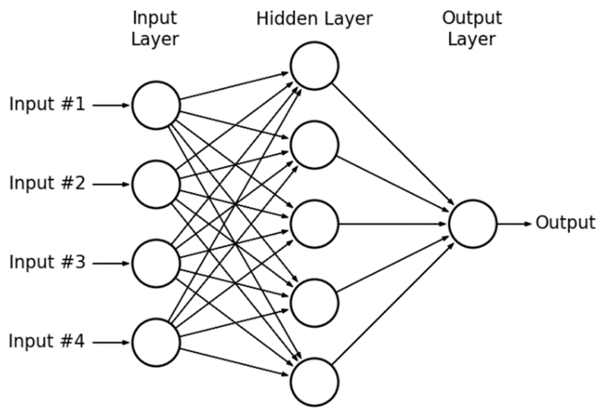


Figure 4.5. MLP layer example

4.2.2 Input data

Throughout this trial, we used the CloudPoint representation of the dataset. We concatenated the two cloud point feature vectors (1026 values each) to create a 2052 feature vector that can be used as input to MLP models.

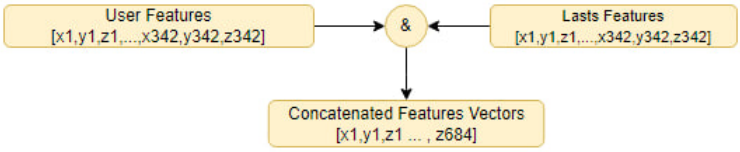


Figure 4.6. MLP Data Pipeline

4.2.3 Batch Normalization

Batch Norm is a normalization technique that is placed between layers of a neural network. It is performed along batches instead of the entire data set and it speed up training and allow the use of higher learning rates. It overall make training easier. Batch Norm reduces the internal covariate shift of the network.[28] The covariate shift is a change in data distribution. The internal covariate shift is a change in the input distribution of an internal layer of a NN. For the neurons in an internal layer, the inputs received are constantly changing due to the fact that computations are done before it and the weights over the training process are constantly being learned. Applying Batch Normalization means that mean and standard deviation of the inputs are always the same. The amount of change in distribution is reduced and in this way deep layers have a more robust and stable ground which helps during training. Also, Batch norms also help in regularization but it is usually helped in this task by the dropout.

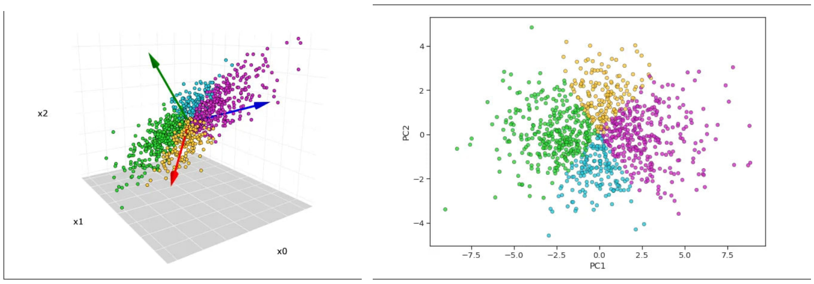
4.3 Multilayer perceptron with PCA

Despite the fact that using raw data for training can still be a viable option, the results and the performances achieved with high normalization and regularization suggest that using PCA as a preprocessing step, in this task, can provide a significant improvement, hopefully resulting in shorter training times and higher performance. Furthermore, PCA has been used in most anthropometric studies in literature for evaluation and analysis [1], it was a fairly secure second step to take.

4.3.1 PCA

Principal component analysis (PCA) is a technique for reducing the dimensionality of datasets by increasing interpretability and minimizing information loss. It detects linear combinations of the input fields that can best capture the variance in the entire set of fields where the components are orthogonal to and not correlated with each other. The goal is to find a small number of derived fields (principal components) that effectively summarize the information in the original set of input fields. [29]

PCA is based on an orthogonal linear transformation of data into a reimagined representation space. The principal components correspond to the directions in the original attribute space that shows the greatest variance. The dimensionality used is at most the same input dimensionality but usually the dimensionality used is much smaller. In this way, we can take advantage of the main scope of PCA which is the reduction of dimensionality. The method is useful for high-dimensional data sets such as text mining, biological data analysis and customer preferences.



(a) (b)

Figure 4.7. Data original representation (a) and after applying PCA algorithm (b)

4.3.2 Input data

In this trial we used the CloudPoint representation of the dataset as input for the PCA layer. The two cloud point feature vectors (1026 values each) can be manipulated in different ways. In this thesis work we decided to mainly focus our attention on concatenating them and on the distance between shoe and last (difference). For the PCA layer, we have experimented with different dimensions, but the main focus and highest results have been achieved with dimensions 800 and 1200. In this phase, experiments have also been conducted with adding information on the input feature vectors. As a first step, it has been decided to add information about the shoe length. The cloud point representation contains two external points that, as already mentioned, have been added as a post processing step. In order to utilize those points as additional information, it has been decided to perform the difference between toe and heel for each cloudpoint (one for shoe and one for mesh). These two values can then be concatenated to the PCA feature vector. The PCA layer output, possibly concatenated with toe heel difference, is then passed as input to the MLP model.

4.4 Multilayer perceptron with Non Rigid Alignment

The previous works created good foundation in this thesis work and proved that the dataset contained the right information to achieve fairly good performance. It was during this phase that we decided to look in more depth into the dataset and into the real correspondence between shoe and last. As shown in the Data Analysis 3.3 chapter, the data have some alignment issue. Looking in depth about those weird correspondence we found out that the trials not only are slightly misaligned but in some of them the shoe shape is completely deformed 4.9. By talking to the company it has been established that the shoes probably originally really had that type of deformation and it was just reported had it was in the real world. Something needed to be done to address this rather common issue, clearly performing a rigid alignment was not a possible solution and so it has been decide to try a non rigid deformation approach. [31]

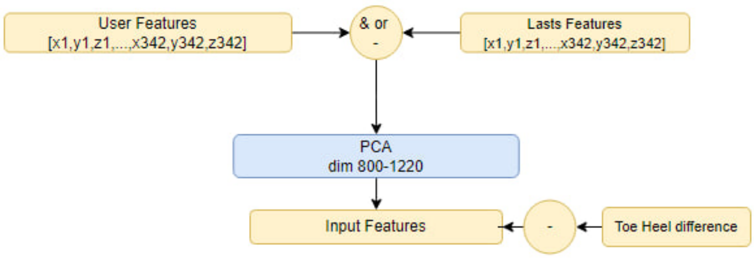


Figure 4.8. MLP Data Pipeline with PCA

Figure 4.9. Example of deformity in trial sample

4.4.1 Non Rigid Alignment

During surface registration, the deformation computed aligns a source surface with a target surface. The surfaces can be aligned by rigid or non-rigid deformations, depending on the application. The former method involves rotating and translating the whole source surface, which works well for registering static shapes. In a non-rigid deformation, different parts of the source surface can undergo different deformations. Registration of non-rigid objects is a challenging task. Unlike rigid registration which only involves a single rotation and translation, non-rigid registration often needs to determine a deformation field for the source surface. Often, benchmarking datasets contain ground-truth correspondences. A common trend has been the use of anthropomorphic shapes, such as human bodies and human faces but, one key deficiency in existing datasets, tho, is the lack of a training facility for learning-based methods. It was for this reasons that as a first approach for this strategy we decided to utilize a Coherent Point Drift algorithm. Despite the effort allocated in finding possible supervised and unsupervised methods, like CorrNet3D that performs Non-rigid shape correspondence and aims to find the point-to-point correspondence of two deformable 3D shapes [33] or 3D-CODED that aligns the input shape into a given template [32], it was clear that finding a model that was trained on a variegated dataset of cloud points (not only body shapes or faces) and that was robust even with low density shapes was really challenging.

Coherent Point Drift

CPD algorithm, in it's simplicity was perfect for our task. The shapes given in input are already semi-aligned and this vastly simplify the work that the algorithm need to perform.The fact that the shapes are already almost perfectly aligned also hugely decrease the likelihood of changing the direct point-to-point correspondence between shapes. CPD is a probabilistic method for rigid and non-rigid point set registration. In that work they consider the alignment of two point sets as a probability density estimation, where one point set represents the Gaussian Mixture Model centroids, and the other represents the data points. [34] The main object of their method is to force the GMM centroids to move as a group. This preserves the topological structure of the point sets. We only used their non-rigid point set registration in which they formulate the motion constraint and derive the solution through variational approach. CPD finds both the transformation and the correspondence between two pointsets without making assumption except the motion coherence.

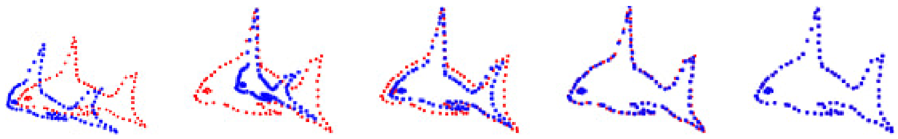


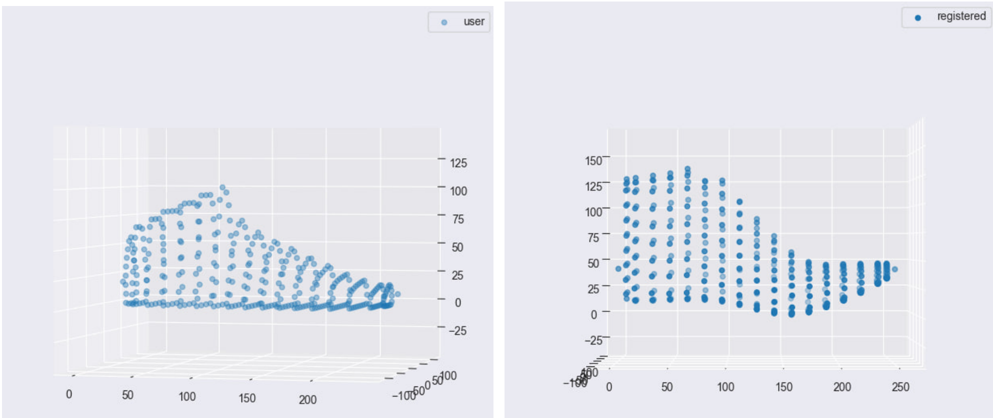
Figure 4.10. Example of Non-rigid CPD registration of 2D fish point set from [34]

4.4.2 Input data

In this trial we used the cloud point representation of the dataset. Before using it as input for the PCA layer, or directly as input for the model, we performed a non rigid alignment of the foot to the user lasts. The hope is to fill the gap that the default alignment is leaving and, hopefully, decrease the uncertainty of the model. There are different ways to utilize this information to our advantage. One is by using the raw data like we already used it in our previous tests, the only difference is that now we have information that before we didn't have.

We are talking about the forces of deformation of the user shape into the shoe. Since the deformation that we applied is from user to shoe we now can extrapolate the forces that the user foot is subjected. Or better we can simulate those forces by calculate the distance that each point of the foot need to travel. If the foot need to shrink probably the shoe is squeezing the foot, if foot instead need to grow, then probably the shoe is too large and the foot can move a lot inside of it. We are going to test with both those information and we will try different combination even with older methodologies like PCA.

A small ulterior test has been performed also with volumes. In this case we tried to extrapolate the volume difference between pre-registration and post registration. The aim was the same as the deformation forces but focused not only on the point-to-point registration but in a overall section of the shoe, for example the entire heel or the tip section and the instep.



(a) **(b)**

Figure 4.11. User original shape (a) and after applying CPD algorithm (b)

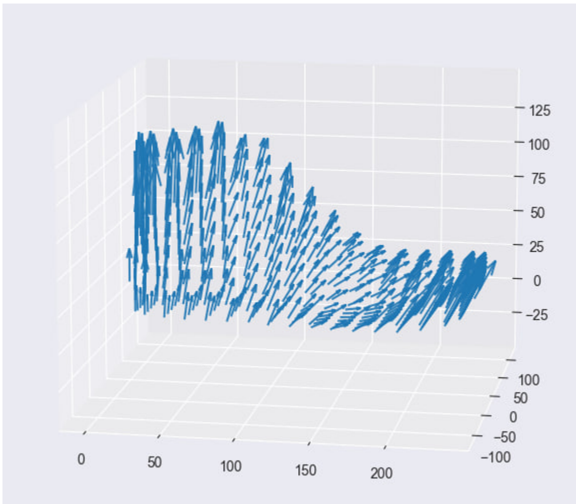
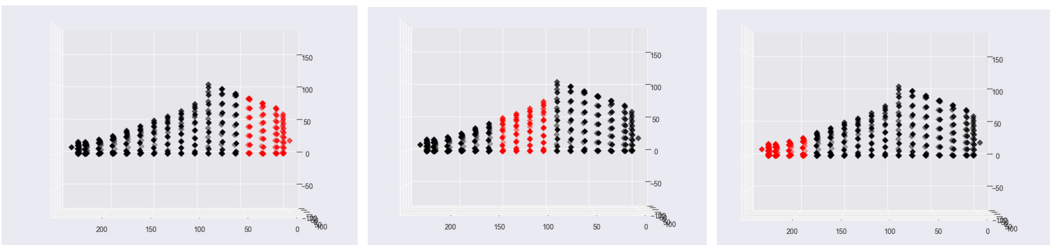


Figure 4.12. Example of deformation forces



(a) **(b)** **(c)**

Figure 4.13. Example of section used for volume estimation: Heel (a), Instep (b), Tip (c)

4.5 Multilayer perceptron with Skeletonization

Previous attempts tried to mitigate and resolve possible issues with the dataset. Alignment and generalization with some extra processing and deformation promoted an environment conducive to the model learning. We now try to focus our attention to better a understanding of the functioning and the responses. To change point of view, in this case, we start by working with a different representation of the dataset: The mesh values. Due to the nature of this work a natural next step on the journey is to try to extrapolate and work with skeletonization and the latent information that it carries. Using a skeletonization method it is possible to extend the deformation capabilities and integrate more and, hopefully, better information inside the model.

Despite the fact that this is an introductory test, it will be interesting to learn how to interact with such a dataset and how well the model will respond to these changes.

4.5.1 Skeletonization via Mean Curvature Skeletons

Skeletonization is an important well-adopted shape abstraction. It is used for various application such as segmentation, matching and shape reconstruction. The algorithm that has been used in this thesis work is MCF [36] as a mean for curve skeleton extraction. Given an input triangle mesh, it drives the flow towards the extreme to collapse the geometry and obtain a skeletal structure. [35]

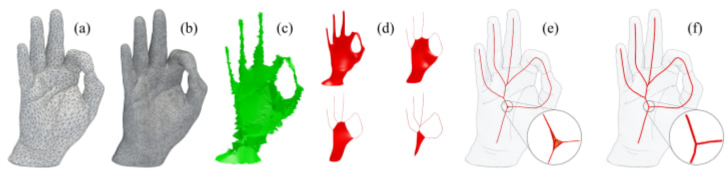


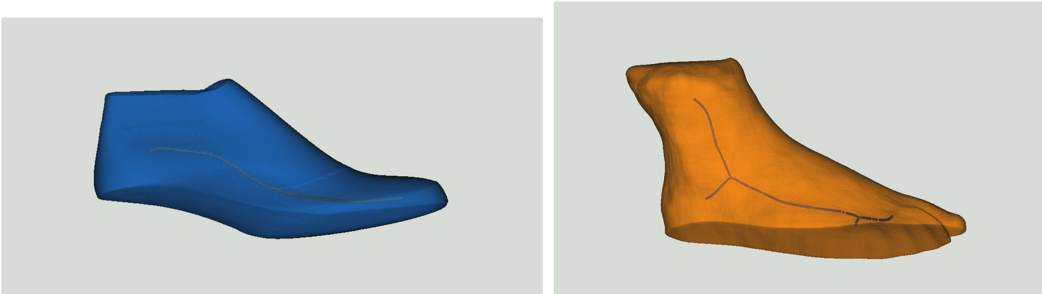
Figure 4.14. Overview of the skeletonization algorithm. From [35]

The algorithm is based only on meshes and it adopts the discrete form of MCF for triangle meshes. It is a constrained Laplacian smoothing process on which two important modification have been applied to achieve an higher stability and a more efficient processing. In the renewed work they perform local re-meshing via edge splits and edge collapses. This produces more regular tessellation and eliminates the needs of a post processing step. Furthermore it leads to a more efficient computation. It also utilized a flow control to stop as soon as the local contraction leads to a pinch in the surface reducing even more the computational cost of the algorithm. [35]

4.5.2 Input data

For this methodology we decided to try slightly change the information used and we start manipulating information in the form of mesh data. In this preliminary approach has been decided to utilize the mesh information to extrapolate a skeleton representation of the shoe.

The results is composed of 142 unique points. In total we can decompose the skeleton result in more than 90.000 non-unique points, each point linked to the corresponding point in the mesh. As we can see from the figure 4.15 the two generated skeletons are not similar and do not overlaps really well. Since working with this differences would certainly give an hard time to model during the training



(a) (b)

Figure 4.15. Example of the skeletonization results.

phase we decided to perform a little processing and to utilize the non rigid alignment information inside the skeleton approach. 4.12 It has been decided to utilize the foot shoe skeleton and deform it based on the forces of deformation obtained from the user registration. Due to the fact that the user deformation shape does maintain a good spatial correlation we can just inversely apply the deformation forces used for the non-rigid alignment of the user shape to the skeleton points. We end up with a skeleton that somewhat resembles the shape of the user shape, but, with a skeletonization generated from the shoe shape. It is to notice that, as explained above, to work the skeletonization algorithm need a closed mesh. In our dataset not every shoe has a closed mesh. Sometimes mesh models have holes, are corrupted or have slightly ruined lasts. This lead to uneven performance by the algorithm that, most importantly, was unable to perform the skeletonization process on 633 different trial samples. For this reason the number of samples for train, test and validation are marginally lower: 10950 training sample, 1373 test and validation samples. [1].

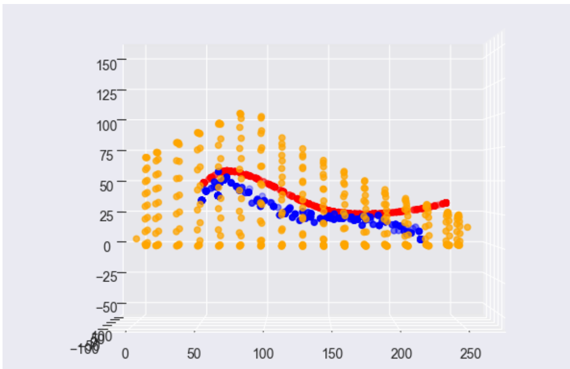


Figure 4.16. Example of the skeletonization results after processing. In orange are shown the points of the user foot. In red the last skeleton and in blue the generated user skeleton

During this thesis work has been tried to both utilize the concatenation of user-last skeleton and to utilize the differences between the two point vectors. We will see in the analysis chapter how those two methods compares.

[1] Without Augmentation

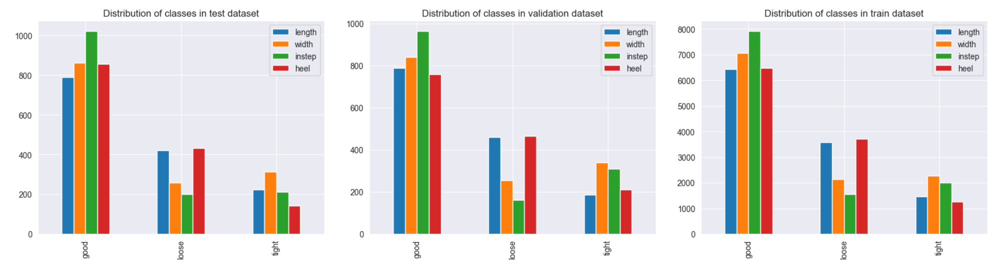
Chapter 5

Evaluation

5.1 Benchmarks and Evaluation metrics

5.1.1 Dataset splits

To perform train and evaluation the dataset has been shuffled first, then splitted in three different splits. This to maintain the same distribution between splits [1], the distribution can be seen on figure 5.1. The train split contains 80% of the dataset, test 10% and validation 10%. Without data augmentation the number of samples of the train dataset amounts to 11463 different trials. With data augmentation, instead, this value is increased by almost 5000 samples and reaches 16092 different trials. The validation set and test set, instead are always constant due to the fact that we decided to not apply data augmentation on the evaluation sets. This choice is usually good and it avoids additional biases and noise in the evaluation metrics. Even tho this is an edge case (the augmented data is not artificially created but is based on an assumption, see dedicated section 3.3.3) we believe that this is the best choice for future comparison and direct observation.



(a) (b) (c)

Figure 5.1. Data distribution for test (a) validation (b) and train (c) splits.

5.1.2 Metrics

For the evaluating validation and test set we decided to settle with simple F1 score as an evaluation metric. For train and validation sets we used the Loss function

[1] The split is firstly seed generated and the results locally saved to maintain a consistency even on hardware differences

as evaluation. We used Binary Cross Entropy (BCE) with the binary classification (global fit) and CrossEntropy loss for everything else (every local fit).

It has been decided to not perform standard accuracy in this specific workflow due to the low relevance of it. Performing standard accuracy on unbalanced dataset would lead to a biased result as it would give more importance and credit to a model that correctly identifies the most predominant class instead of one that doesn't develop any bias towards the predominant class. F1 score is an harmonic mean of Precision and Recall. In other words by maximizing the F1 score we are choosing the best balance between Precision and Recall and, consequently, we are actively evaluating also the biases of the models. To compare the results on test data is has been decided to also include confusion matrix. Confusion matrix shows the performance of the model for each class and allows to understand the behaviour of the model, the uncertainty between classes and if there is some dataset bias that is limiting the performance.

5.1.3 Settings

Test has been performed locally with an Nvidia GTX 1070 with 8GB of VRAM. Unfortunately the VRAM was a limitation with the PointNet architecture but, otherwise, most of other models performed flawlessly. The CPU used is a Ryzen 5 1600, this CPU performed adequately for most of the tasks but with deformation and skeletonization one. Due to the high density of the mesh data some computation experiments took numerous amount of hour or even, cumulatively, days to finish.

5.2 5 classifiers vs 1 Classifier

During the Data exploration we discussed about the nature of the task. In spite of the fact that the main objective of this study is to determine whether the user will keep or return the product, most of the information provided is geared towards the capacity of the model to identify a good fit for the user. Since there is a distinct classification between fits for each section of the shoe in the dataset we can use those information to figure out if the user will return or keep the shoe.

As a first step in the decision-making process, it is essential to decide whether we are going to work with a multi-loss model or create five different classification models.

Table 5.1. F1 score for each method and each Fit task

Learning Methods	Length F1	Width F1	Heel F1	Instep F1
Multi Loss	65,44%	**63,65%**	**61,11%**	69,29%
Single Loss	**65,83%**	63,14%	**61,11%**	**70,00%**

Our initial investigation showed that, even though multi-loss approaches perform similar to 5 different classifiers, modeling and develop a multi-loss systems adds unnecessary complexity and difficulties. We therefore decided to optimize our work for only the Length Classifier, also in light of the fact that classifiers behaviour seems to be correlated. The higher the score on a fit is, higher the score on most other fits are as well.

We arbitrarily decided to optimize for the Length classifier since it was the classifier that had a more predictable behaviour.

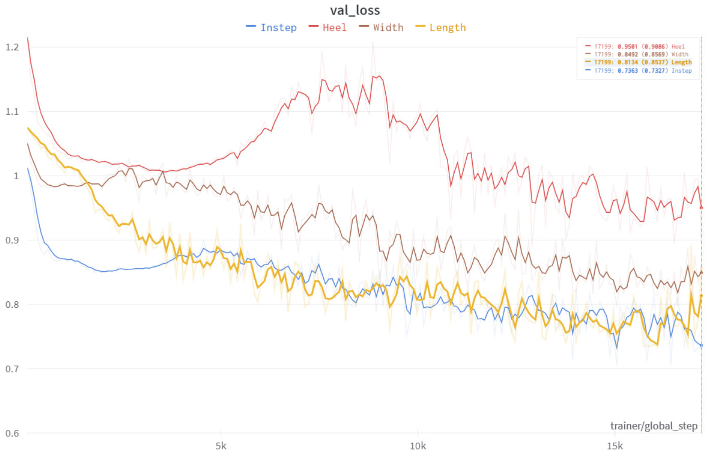


Figure 5.2. Baseline Models Loss

5.3 Average Fit vs Global Fit

The Global fit is a complex different task that we can address in many different ways. The first approach that comes in mind is by averaging the results of the different local fits. But as is shown in the figure about the number of local good fits for each shoes 3.10 the best threshold is somewhere between 3 and 4, in a perfect environment. But even in a perfect environment the confusion is really high and the only way to obtain good results would be to choose 4 as threshold. This means that the only good samples are the one with perfect local fit.

Table 5.2. F1 score for each threshold on Global Fit

Threshold	**Global F1**
0,0	37,07%
0,25	38,82%
0,50	40,86%
0,75	46,10%
1,0	53,39%

Based on the findings of this early study, it is clear that averaging local fit is not a good approach to this problem. By analyzing the Confusion Matrix we can see why the accuracy is so low. We can start by thinking about the data distribution. Local Fit are unbalanced towards the good fit. This reflect negatively on the Global fit that, as we analyzed before, is unbalanced towards the bad fit. This of course creates a contrast in the expected results because the task that we have learn cannot be re-learned by the model since we are only using an average of local fits.

Another possible approach is to implement a specific model only for the global fit.It is of course still affected by the unbalancing of the data but it will try to address all the information that local fits cannot address by themselves.It is important to notice that in our tests we did not eliminate any weird instance of the dataset (as

[2]Confusion Matrix normalized on True Labels. Rows sums up to 1

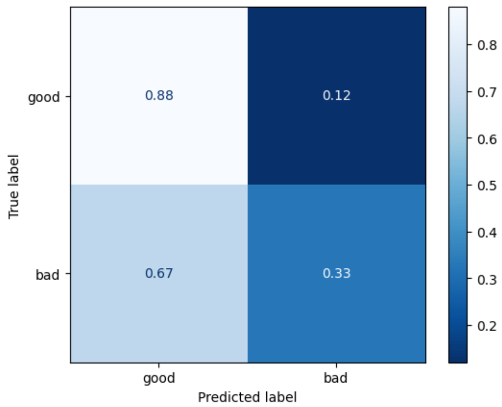


Figure 5.3. Confusion Matrix on Baseline Models with threshold at 1 [2]

for example a "bad" global fit with all local fits as "good"). This will create noise but we think that those samples contains probably more information about what an user really want in a shoe in respect of others samples that are more predictable.

Table 5.3. F1 score on Global Fit for the two methods

Model	Best Global F1
Average Global Fit	67,52%
MLP	68,9%

In this preliminary test we noticed a slight increase in performance hinting us that we are on a right path and a machine learning approach is the right way to proceed. In the Confusion matrix of our MLP base model. 5.4 is clear how the unbalancing of the data create a clear bias towards the bad class, this can be addressed with some data augmentation. Due to the complex nature of the global fit problem and lack of a generous amount of data, our work has been focused on finding the best way to address the local fit comfort with the intent, once the performance are acceptable, to bring and transfer those information to a last and conclusive layer capable of use those information combined with a specific learning about the full global fit task. The aim is also to use this transfer learning approach between local fit and global fit to address and mitigate the lack of data.

5.4 PointNet from scratch - Analysis

During the implementation, we discovered that PointNet does not have (at the time of the trial) any publicly available pretrained model and the only link to the data that was used for training is no longer working. Therefore, we have decided, for this type of implementation, to only train the model with our data in order to see the behavior and how well it trains with our samples. As it is understandable, we have to compute the feature extraction with both the user's feature and the last's feature in order to be able to use the PointNet extraction for our training. PointNet in fact has been created with the intent of working on only one cloudpoint and does

[3]Confusion Matrix normalized on True Labels. Rows sums up to 1

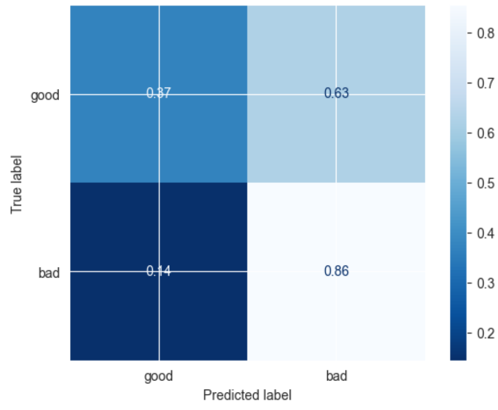


Figure 5.4. Confusion Matrix of Baseline MLP Model [3]

not have the ability to take in input multiple cloud points at the same time. The model can be used in a variety of ways, but it is necessary to adapt the approach slightly in order to make full use of it. It has been decided to utilize PointNet as a feature extraction step. In other words we individually pass both User and Lasts Cloud Points so that we end up with two different global vectors. In order to proceed with our classifier, we need to concatenate or subtract those features once they have been extracted.

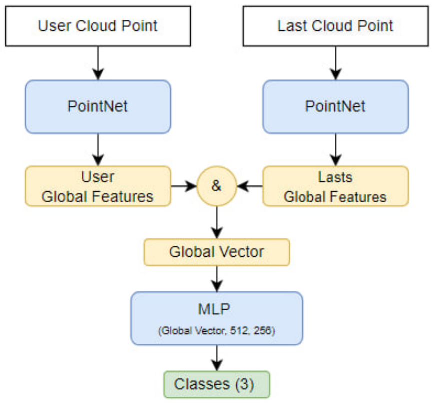


Figure 5.5. PointNet Model utilized in our tests

This methodology uses an MLP classifier that takes as input a global vector of 2048 (twice the size of a PointNet global vector) or 1024 in case of subtraction pre-processing. It contains 2 linear layers with 512 and 256 dimensions, with an output layer of 3 classes. It has been decided to add a Batch Normalization Layer in between each layer to help with the regularization process.

We tried an exploratory test with a modest learning rate of 0.0002 and a batch size of 32.

We can see from the graph that the model had immediate problems when it came to learning. Upon first glance, by looking at the training loss, it seemed that the model was learning a bit. However, when examining the validation loss, it was clear that it was increasing more and more with time. In general, this type of behaviour

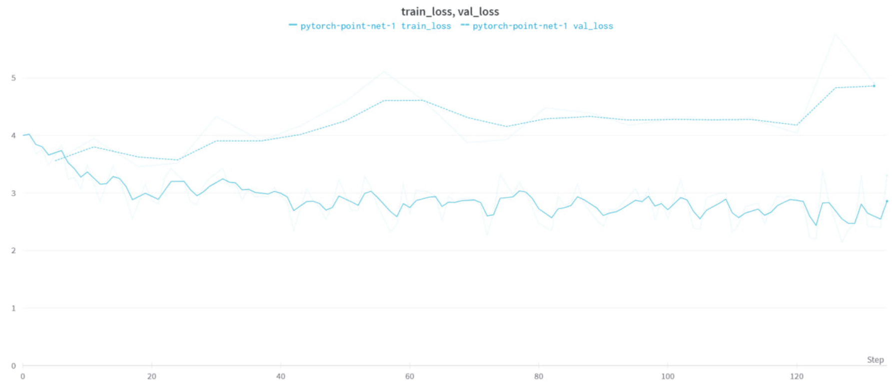


Figure 5.6. Train Vs Validation Loss for the exploratory test on pointnet

is associated with overfitting, but when we take a closer look at the F1 values, we can really see how this type of behavior is not overfitting in the usual sense. The model learned to predict always the same class with no interest in actually learning from the information given. As the result of some trials, it was quite evident that in order for this model to be able to work, we would need more data, and probably more computational power since parameter tuning and epoch number were limited in this regard. Here we show the most representative results for this method 5.7

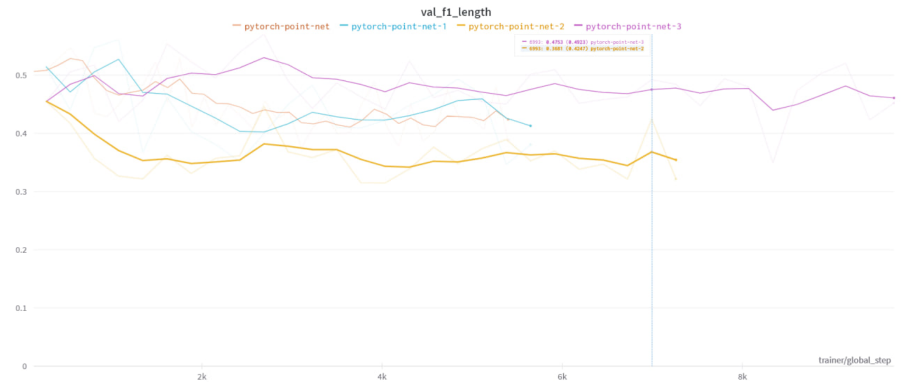


Figure 5.7. Validation F1 score for test on PointNet

5.4.1 Implementation

For a better pipeline it has been decided to use an implementation of PointNet available for Pytorch, rather than TensorFlow. Our first attempt was to utilize a PointNet version that was able to achieve similar performance to the original one5.4.

Table 5.4. Pytorch PointNet Implementation vs Original Implementation

PointNet	Accuracy score
Original implementation	89,2%
Pythorch PointNet implementation	86,4%

5.5 PointNet++ from pretrained - Analysis

In a first iteration of our PointNet++ work we were not only computational limited but also memory limited. PointNet++ with its complex architecture restricted the amount of trials that the local machines where able to handle.

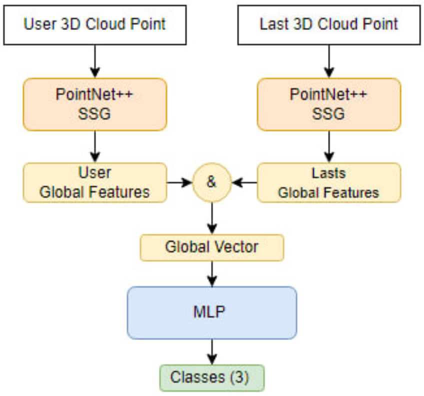


Figure 5.8. PointNet++ Model utilized in our tests

In this pipeline configuration the model configuration applicable was with a batch dimension of 32 and a Perceptron of only one layer of dimension 256. Of course this was very challenging for the task and the model really struggled to learn. It is really noticeable that, like in the classical PointNet trial, also in this test the validation loss had an incremental trend. The model seemed to learn only to classify the most common classes. This is a clear sign that the same problem that affects the old PointNet implementation also affects this PointNet++ one.

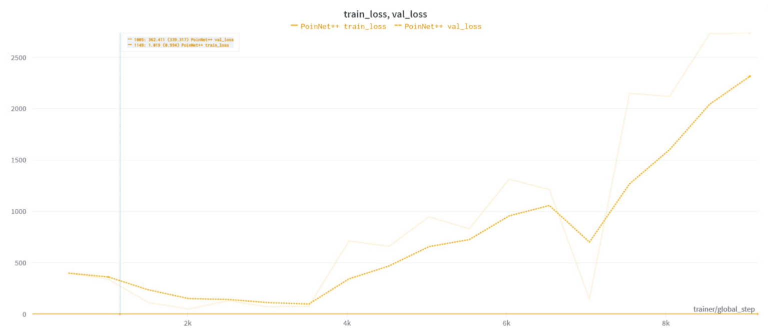


Figure 5.9. First PointNet++ Model implementation loss

To furthermore analyze this behaviour the pipeline has been optimized to allow the utilization of an increased number of layers and dimensions of the classificator part of the model. Using this method, the model appears to perform slightly better as shown in the loss graph. 5.10

The method achieved an F1 score of 56%, but did not outperform most other methods. The overall conclusion is that this methodology requires some additional steps in order to be effective. While the transfer learning approach created a better learning environment for the model, it was not able to overcome the task's difficulties.

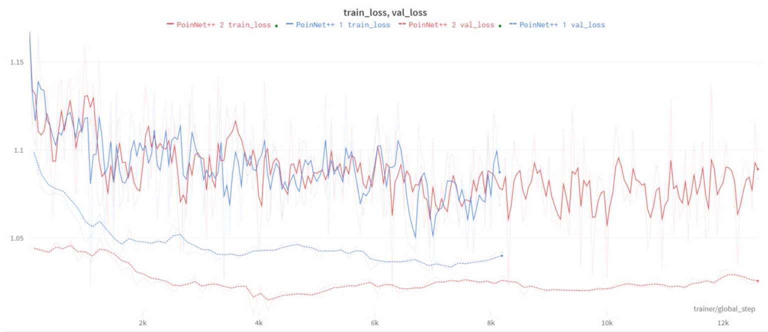


Figure 5.10. PointNet++ Models implementation with higher depth

Table 5.5. F1 score on validation set for each PointNet method

Model	F1 Score
Pytorch PointNet	56.14%
Pytorch PointNet++	56.2%
Pytorch PointNet++ (Adjusted pipeline)	55.07%

5.5.1 Implementation

The MSG computational consumption, together with a slight increase in space occupation lead to the impossibility to try most of the models given. The only model that we managed to use is PointNet++ SSG.

Table 5.6. Pytorch PointNet++ Implementation vs Original Implementation

PointNet++	Accuracy Score
Original implementation	91.9%
Pytorch PointNet++	92.2%

5.6 MLP with Cloud points - Analysis

In this section are shown the first attempts at designing a model from scratch. The architecture chosen, as anticipated before, is a plain Multilayer Perceptron that creates an environment where it is possible to work and try different model structures without worrying too much about space and computation. In the first phase of the study, different combinations of batch size and learning rate have been tested. Most trials have been carried out with batches of 32, 64 or 128 pieces. The learning rates used instead were many. In this thesis, we will only present the most relevant tests. Is is to notice that in this test we performed an evaluation for all the local fits, this is valid only for this specific evaluation and they will not be used as a comparison metric with other models. We used this evaluation as a foundation for comparing the use of five different classifiers for each local fit versus the use of only one model, as shown in the corresponding chapter 5. In this section we will instead show a deeper analysis specific to Length Fit.

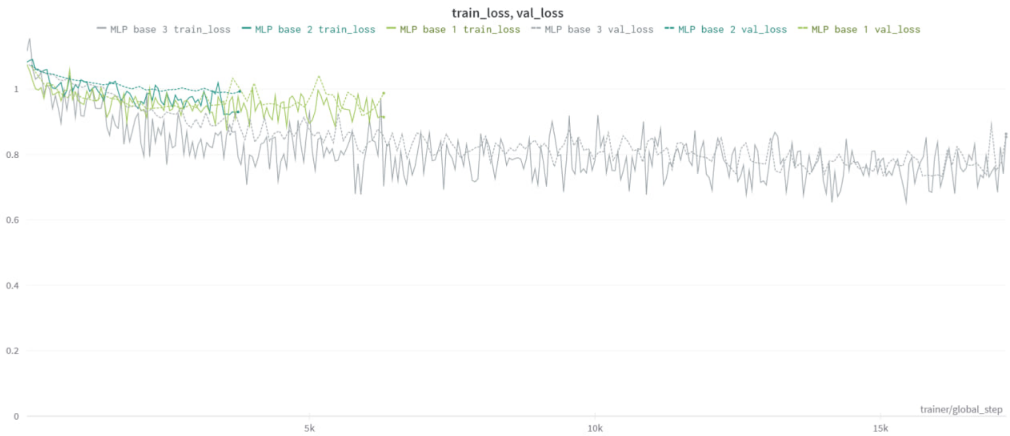


Figure 5.11. MLP Losses on Length Fit

In the loss graph are displayed 3 different loss for 3 different models. These 3 models represent the best models for each different batch dimension. The model depths are not equal between each other but it is really interesting to see how, even tho the losses curves are really distinct and the model dimensions are completely dissimilar, the progress and ultimate results are really similar 5.12.

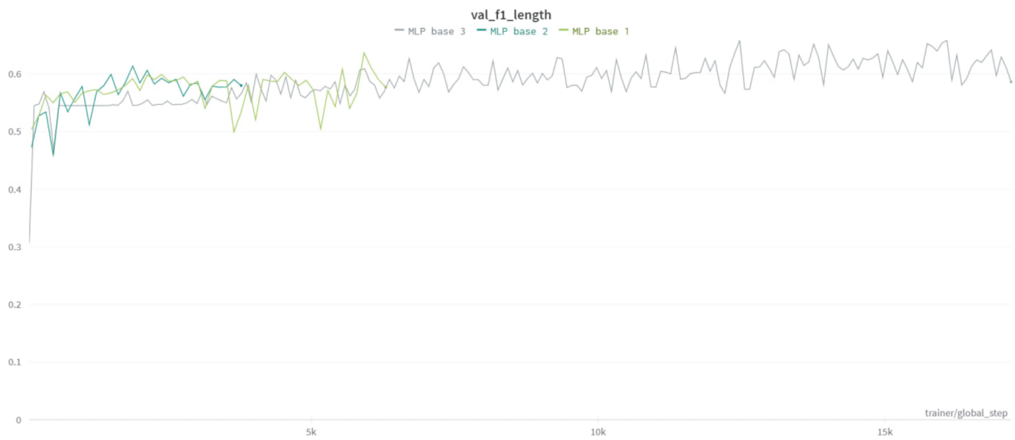


Figure 5.12. MLP F1 score on Length Fit

Of course, the model with lower dimensionality like the "MLP base 2" as well as the model with higher learning rate as the "MLP base 1" learned faster but also reached a lower peak sooner. "MLP base 3" instead outperformed the others by learning slower.

Table 5.7. F1 score for validation set for each MLP Cloud Point method

Model	F1 Score
MLP Base 1	61.39%
MLP Base 2	63.6%
MLP Base 3	**65.83%**

The model dimensions that worked the best were MLP with dimensions of [input,512,256,128,256,128,64,3]. During evaluations, we will see this type of behavior and dimensions frequently. During most of our tests, the models that were able to learn the best were those that were slower to learn and had an higher level of regularization and normalizatio. There is still a long way to go before the F1 score can be considered as acceptable, and we can go into some detail about what, possibly, is causing this challenge.

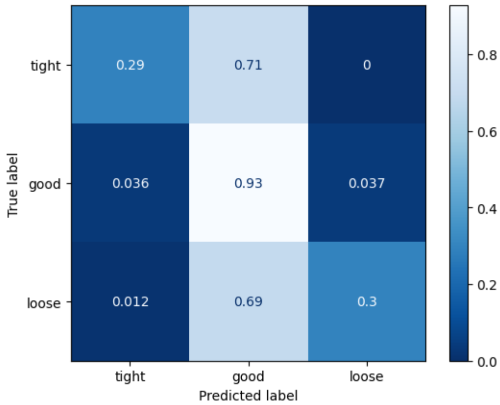


Figure 5.13. Confusion Matrix on test set for best MLP score [4]

From the confusion matrix, it is apparent that there is a very high bias towards the "good" class. The model is having difficulty distinguishing between good trials and all the rest. It is noticeable that the model rarely confuses tight with lose and vice versa, which may be a result of the overall bias toward the predominant class.

5.6.1 Implementation

The mlp model that best worked with this approach is a model that is fairly deep. From 2052, the inputs gradually decrease to 512, 256 and 128 neurons. Then a layer of higher dimensionality is added to allow the creation of more depth in the information extracted. Two additional layers are then applied to gradually lower the dimensionality to the output dimension of 3.

Each layer is divided by a Batch Normalization layer that helps the model not overfit and creates a better training curve. To help with the regularization we also introduced dropout layers.

5.7 MLP with PCA - Analysis

The purpose of this section is to analyze the findings and results derived from this methodology. Like for the other methodologies, they have been tested with different batch sizes and different hyperparameter tunings. High batch sizes of 128 samples have produced the most relevant results. Differently from the base line model, it was found that higher learning rates tended to work well in this particular trial.

There is a possibility that that behaviour is related to the PCA preprocessing. The model now utilizes a simplified version of the feature vector, which contains

[4] Confusion Matrix normalized on True Labels.Rows sum to 1

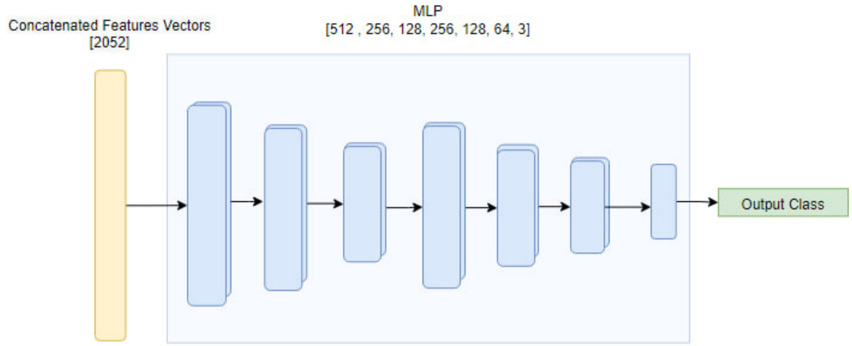


Figure 5.14. MLP model scheme

Table 5.8. MLP PCA hyperparameter comparison

Model	Batch size	LR	Dropout	Input Type	MLP dimensions
MLP PCA 1	64	0.4	1-e5	User	[802, 256, 256,128 , 64, 3]
MLP PCA 2	128	0.4	5-e5	Difference	[802, 512, 256 , 128, 3]
MLP PCA 3	128	0.4	1-e4	Concatenation	[1202, 512, 256, 128, 256, 128, 3]

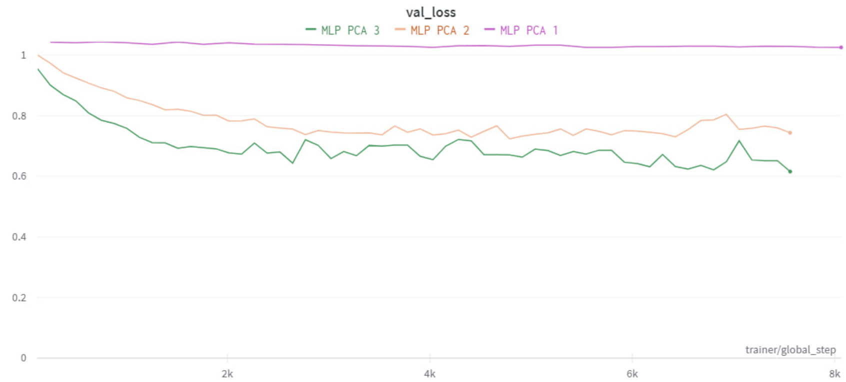


Figure 5.15. MLP PCA losses on Length Fit

essentially the same information, but reshaped and organized in a way that makes categorization and classification more efficient and straightforward. As we can see from the graph of the losses 5.15 there is one model that is clearly having trouble at learning, almost flat lining the entire learning phase. It was a trial that aimed at understanding if the model was learning the correspondence between shoe and last and if the PCA was maintaining those information in respect of learning only the latent information of the dataset. To understand that we decided to try to only pass the information about the user foot to the PCA layer and see how the model behaved.

The model behaved like expected and only managed to learn the class distribution without really learn the task. Different story for the other two relevant trials. From

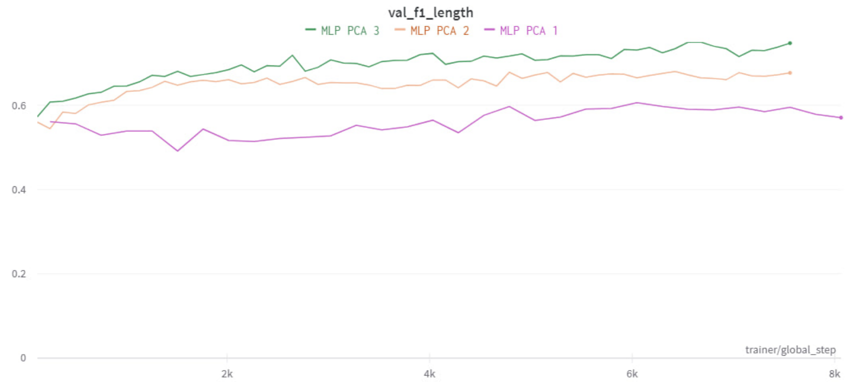


Figure 5.16. MLP PCA F1 score on Length Fit

those trials is interesting to notice that passing to the PCA layer the raw data created a better feature vector and the model managed to learn a better correspondence and better classification.

Table 5.9. F1 validation score for each MLP PCA method

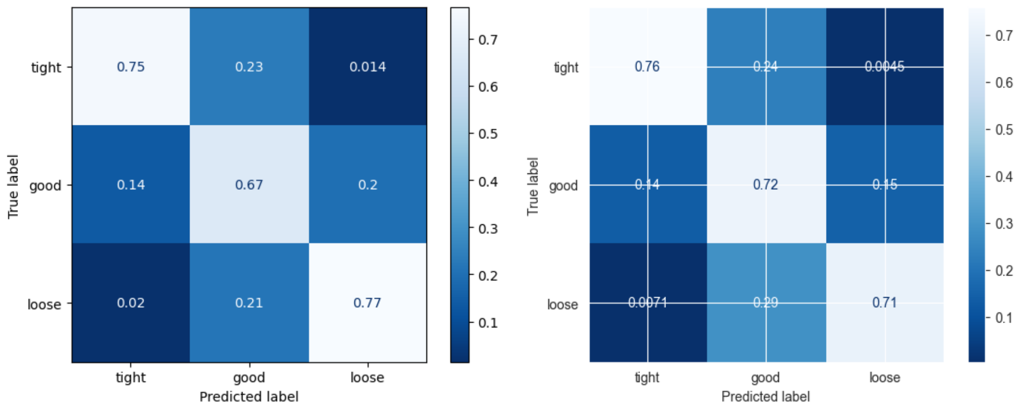
Model	F1 Score
MLP PCA 1	60.06%
MLP PCA 2	68.0%
MLP PCA 3	75.01%
MLP PCA 3 Tuned w/o ToeHeel	76.21%
MLP PCA 3 Tuned w/ ToeHeel	**77.17%**

The model dimensions that worked the best were MLP with dimensions of [1202,512,256,128,256,128,3]. It is also clear, that toe heel information made a slight but consistent difference and lead in better results as we can see in the table 5.9. In the table we can also see that, after some hyperparameter fine tuning the best model managed to reach an F1 score of 77,17%. Despite the use of the Toe Hell information granted only a small increase in F1, from the results it is clear that the model using that information had a small but stable advantage and so we decided to utilize that information for future tests.

5.7.1 Implementation

MLPs models with a fairly deep structure work best with this approach. From 1202 inputs, the number of neurons decreases gradually to 512, 256, and 128. In order to create more depth in the information extracted, a layer of higher dimensionality is added. Two additional layers are then applied to gradually reduce the dimensionality to the output dimension of 3. As we can see the dimensionality is the same as the basic implementation. It seems that this dimensionality is fairly stable and manages to extract the most out of the dataset. As part of the comparison, we will examine models with a lower dimensionality and see how their implementations compare.

[5]Confusion Matrix normalized on True Labels. Rows sums up to 1



(a) (b)

Figure 5.17. Confusion matrix on test set of the best model without toe heel information (a) and after adding the toe heel information (b) [5]

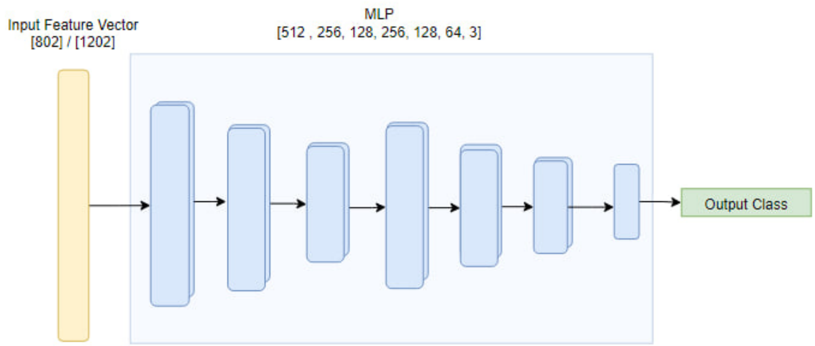


Figure 5.18. MLP model scheme

As for the other layers of Batch Normalization and dropout has been added between each linear layer.

5.8 MLP with Non Rigid Alignment - Analysis

This section will analyze and explain the findings and results of this methodology. Differently for the other methodologies here we will focus our analysis on the different behaviour of the model with different features input. Every model has been slightly tuned to give the best opportunity. As a first step it was needed a comparative evaluation between affine deformation algorithm and completely non-rigid deformation. Both algorithms are from the same CPD implementation [34]. At the same time it was needed to evaluate the difference between using the forces of deformation and the raw data. It was decided to compare raw and forces within the same type of deformation to really understand the difference in evaluation and behaviour. In the figure 5.21 and 5.22 it is shown how the two kind of operation are somewhat similar and, in combination with a PCA processing, manages to reach similar performance to the PCA applied on the original shapes. This is a clear sign that the information are not lost and that the preprocessing that we are doing is not completely destroying the information contained in the feature vectors.

Table 5.10. F1 validation score for Non Rigid and Affine Deformation

Model	Data Processing	F1 Score
Affine Deformation	Raw	72.7%
Affine Deformation	Forces	72.9%
Non Rigid Deformation	Raw	73.10%
Non Rigid Deformation	Forces	**73.48%**

From the figure 5.19 it is also really clear that, also in this input composition, the use of a PCA layer is somewhat mandatory. Without this PCA layer the model hardly reaches 60% in F1 score on the validation set.

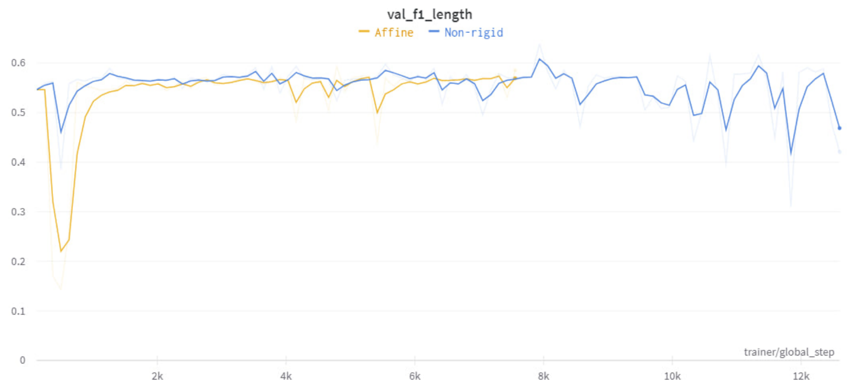


Figure 5.19. Affine vs NonRigid on F1 validation score

To have a better understanding of what is going on it was needed to carry out more trials and experimentation. In order to determine the impact of the deformation information to the overall scenario we concatenated those information to the PCA that has been used in previous attempts (see chapter 4.3).

As we can see from the F1 score graph in figure 5.20 even with this approach the model really fatigued at learning and barely managed to reach 67% of F1 score on validation set. Also here it is noticeable the difference between Raw input and Forces with a clear advantage for the Force input. We can start to see a pattern in all this models. The models is really sensitive to the input features and really struggle to generalize for the task. It is possible to help the model train and with the right combination in capable of reaching some level of accuracy but with every small changes the performance drastically crumbles.

5.8.1 Implementation

The implementation of this model follows the same models of other trials with some dissimilarity on input dimension and MLP hyperparameter tuning. See Chapter 4.3 for a insight on the model structure.

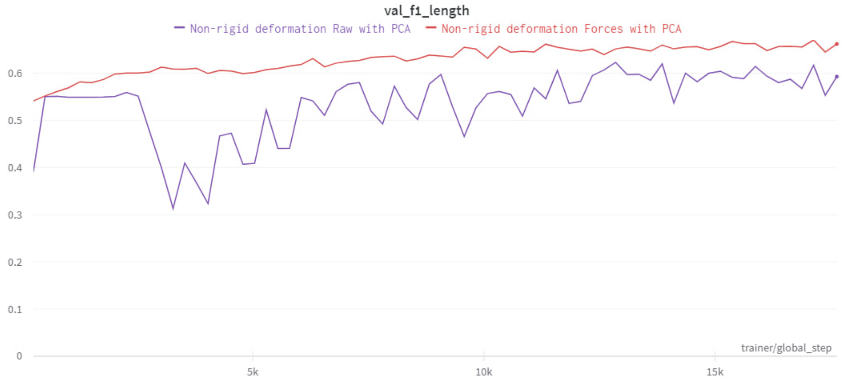
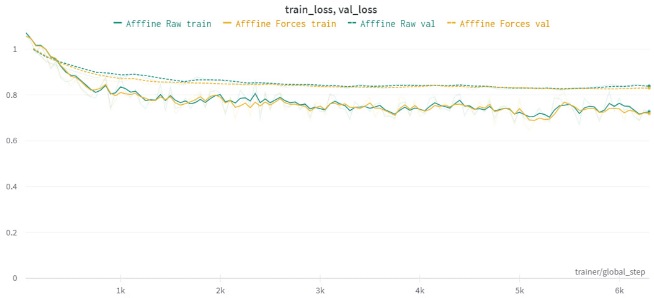
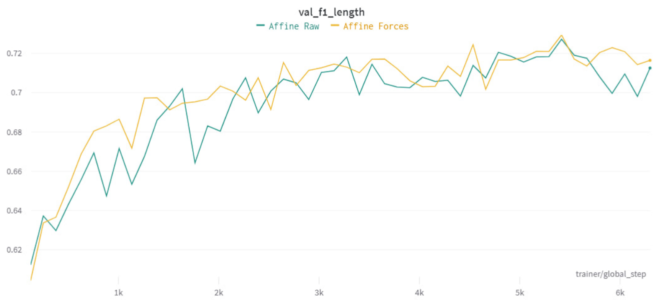


Figure 5.20. F1 validation score for non rigid approach with PCA



(a)



(b)

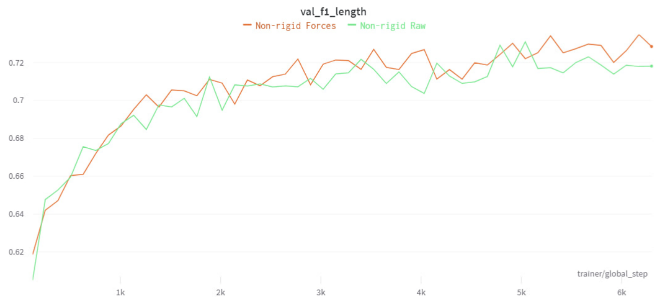
Figure 5.21. Affine Deformation for loss (a) and F1 score on validation set (b)

5.8.2 Adding Volume Information

This small section is dedicated to an analysis about those tests that utilized the information about volume and volume difference. In this section has been decided to report the most relevant combination for this kind of input.

As already anticipated the model really struggles with every new information that is given. The model that stood out the most is the model that utilized PCA but, even with the PCA preprocessing, the model didn't really improved in performance. It is clear that the volume information is creating noise and, in this phase of the analysis, is not a step in the right direction.

(a)



(b)

Figure 5.22. Non Rigid Deformation for loss (a) and F1 score on validation set (b)

Table 5.11. F1 validation score for Volume integration

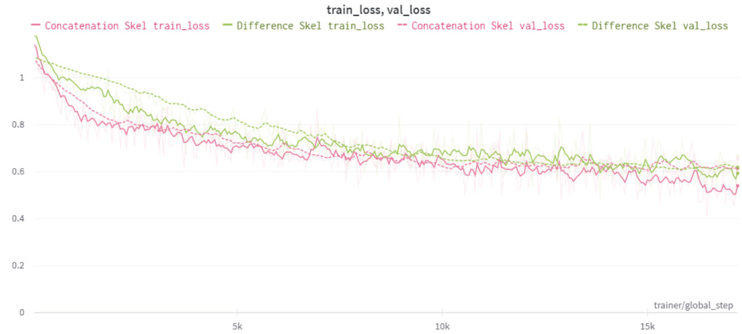
Input Type	F1 Score
Difference between User and Last, Raw Volumes	51.38%
Non Rigid Forces, Raw Volumes	51.5%
Non Rigid Deformation Forces, Difference between User and Last, Raw Volumes	61.76%
Non Rigid Deformation Forces with PCA, Raw Volumes	64.05%

5.9 MLP with Skeletonization - Analysis

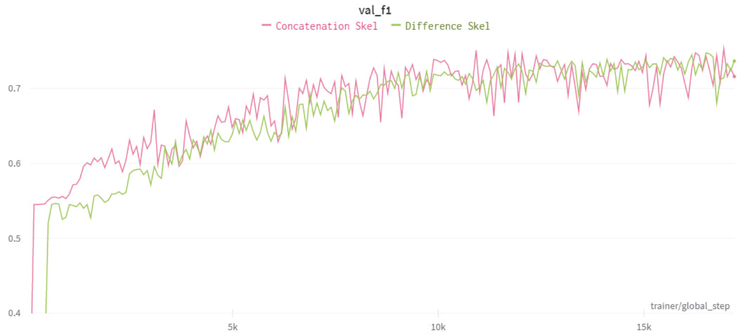
The purpose of this report section is to present an analysis of the above-mentioned methodology. In order to compare well with the other methodologies, the skeleton information was added to the PCA features, which were extracted from the difference between the user's and last's features. The purpose of this chapter is to extend our understanding of the model's behavior. With this in mind, we extracted more information from the dataset while avoiding introducing noise or bias.

The information tested is the skeleton extracted from the mesh shapes of the user and lasts and can be tested in the form of a concatenation of the two cloudpoints or as a feature difference between the two feature vectors.

In the results we can see how the two models behave similarly with a slight advantage, in the training phase, for the concatenation between user skeleton and shoe skeleton. This model is obviously working with an higher input dimension of 3256, in respect of the 2230 of the methodology that is using the difference between



(a)



(b)

Figure 5.23. Skeletonization methods comparison. Loss (a) and F1 (b).

the two skeletons.

In the figure 5.23 we can see how the loss of the model that is using the difference is slightly lower, and this advantage is also noticeable in the f1 score that, in the concatenation version, is 1% higher. A small attempt has been performed to check if the negative behaviour of the volume addition is consistent we decided to try to add this information to the trials. As we can see from the f1 graph 5.24 by adding the volume information both models lost more than 5% of f1 score. It is clear that the volumetric information generated from the cloudpoint is not a good information to use and this results just confirmed the behaviour that has been analyzed in the previous chapter.

Table 5.12. F1 validation score for Skeletonization models

Model	F1 Val Score	F1 Test Score
Concatenated Skel	**75.3%**	69.19%
Difference Skel	74.7%	**71.6%**
Concatenated Skel w/ Volumes	69.49	68.39%
Difference Skel w/ Volumes	69.39%	66.86%

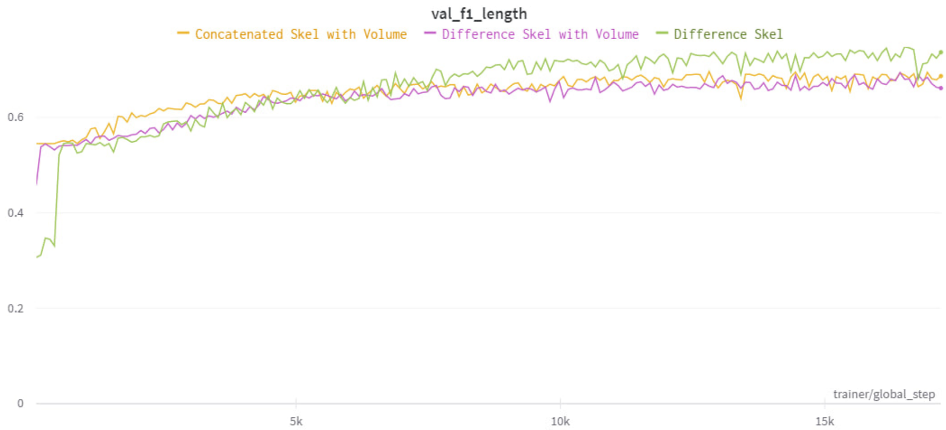


Figure 5.24. Validation F1 score for the models with Volumetric information added in input.

5.9.1 Implementation

The implementation of this model follows the same models of other trials with some dissimilarity on input dimension and MLP hyperparameter tuning. See Chapter 4.3 for a insight on the model structure.

5.10 Evaluation analysis - Summary

In this chapter we will present a small summary of all the results, with cumulative tables and a small deeper analysis about the behavior of the models on the main task of global fit. As already mentioned in this thesis work we decided to not give much importance to the global fit analysis but is nevertheless important how our decision have an influence on the larger and, in a way, harder task.

Table 5.13. F1 validation score summary. Best model for each methodology.

Model	F1 Val Score
Pytorch PointNet++	56.2%
MLP Base	65.83%
MLP PCA w/ ToeHeel	**77.17%**
Non Rigid Deformation Forces (PCA)	73.48%
Concatenated Skeletonization	75.3%

Let's start from a overview of the overall best performance for each methodology. As shown in the table 5.13, in the current state, the best score are achieved trough the utilization of a PCA preprocessing step. In the current state the best model outperforms the second best by almost 2% of F1. In reality we can see how the models behaves similarly with small differences between the bests scores and it is possible that with small adjustment and more tuning those differences can be even smaller. This cannot be said about the base model, PointNet and of course the volumetric addition. In the figure 5.11 it is clear how the volumetric addition changed negatively the behaviour of the models.

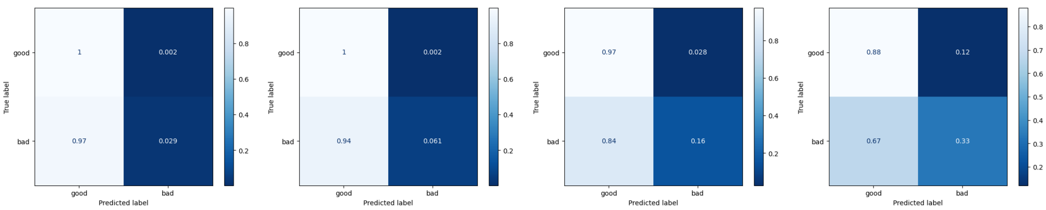
Additional analysis can be also performed on the global fit results, performing a small analysis also on the global fit will surely helps to really understand if the improvements that we are getting on local fits are also transferable to the more universal task. We can start with a small review of what where the results that we achieved on the baseline approach.

$$Threshold = (LenghtFit + WidthFit + InstepFit + HeelFit)/4$$

[6]

Table 5.14. F1 Test score on global fit.

Model	F1 t=0.25	F1 t=0.50	F1 t=0.75	F1 t=1.0	F1 trained
Baseline	38.82%	40.86%	46.1%	53.39%	67.15%



(a) (b) (c) (d)

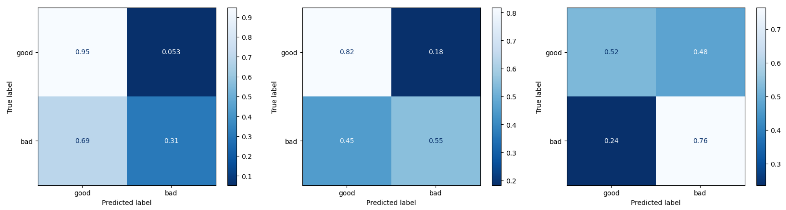
Figure 5.25. Confusion Matrix for Base Model for Threshold 0.25 (a) 0.5 (b) 0.75 (c) and 1 (d).

As we can see from the table 5.14 the results that we obtain by averaging the local fits is not very good in this baseline approach. By analyzing the confusion matrix it's clear that the model really fatiguing in the classification of the "bad" class. This kind of behaviour is probably a consequence of a bias towards the good class in the local classificator that is transferred into the global classificator. It is to notice that, as already mentioned in the chapter 5.2, the global fit has a different distribution than the local fits. Local fits are unbalanced towards the good class, global fit instead is unbalanced towards the bad class. This means that the augmentation that is performed on all the models is not helping and also probably damaging the global classification. By analyzing the results for the two best models, instead, we obtain a slight improvement in the global fit classification performance. Both in the averaging method with threshold and in the trained model.

Table 5.15. F1 Test score on global fit for PCA and Skeltonization models.

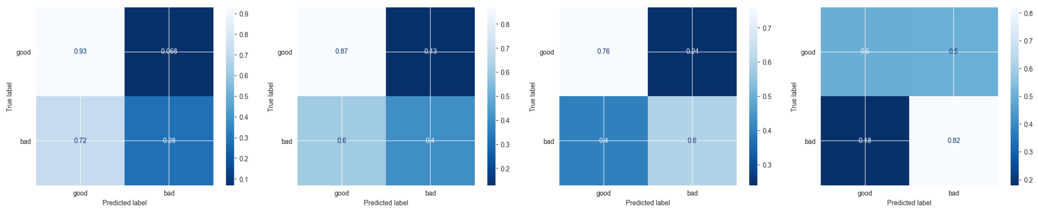
Model	F1 t=0.25	F1 t=0.50	F1 t=0.75	F1 t=1.0	F1 trained
Skeletonization	46.61%	54.33%	64.75%	67.52%	**68.9%**
PCA	**52.06%**	**57.36%**	**65.81%**	**70.27%**	65.81%

[6]Local fits are considered 1 if classified as "good" 0 otherwise



(a) (b) (c)

Figure 5.26. Confusion Matrix for PCA Model for Threshold 0.5 (a) 0.75 (b) and 1 (c).



(a) (b) (c) (d)

Figure 5.27. Confusion Matrix for Skeletonization method for Threshold 0.25 (a) 0.5 (b) 0.75 (c) and 1 (d).

It is really interesting to see that the highest score is obtained every time with the extremely high threshold of 1. This means that we have to have all local fits considered good. This seems common and reasonable but we need to remember that in the dataset the "good" global fit not always have every local fit as good and, moreover, not every sample with all local fit as "good" end up with a global fit as "good". It is possible that 70% is really close to the limit of the methodology and we need to explore different classification models to obtain higher results. 5.28

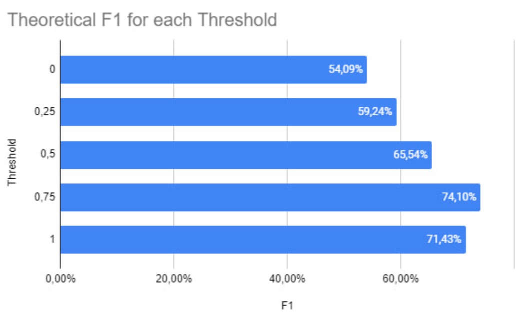


Figure 5.28. Max Theoretical F1 for the threshold method.

Good results are achieved via a model that is trained on the global fit but does not manage to outperform the 70% of the older methodology. This is somewhat expected. The main reason why it has been decided to utilize the local fit is that by partitioning the focus of the model we are capable of generalizing better without risking over fitting and creating biases.

It would be really interesting to study also the behaviour of the models that better utilize the information about local fits. As a preliminary step and as a proof of concept we have built a one last classificator for the global fit classes. In this specific utilization we decided to utilize the last logits for each classificator and, together with the usual PCA features, utilize them as input. The structure of the model is visualized and summarized in the figure 5.29. The results are encouraging, we managed to increase our score by almost 2% compared to our best averaging model and we outperformed our comparable trained model by a really unexpected 6% 5.16. Those results corroborate our theory that by utilizing a combination of local fits knowledge we can help the model train and achieve better results in the global fit task.

Table 5.16. F1 Test score on global fit for PCA with and without LocalFits logits.

Model	F1 trained
PCA w/o LocalFits	65.81%
PCA w/ LocalFits	**71.67%**

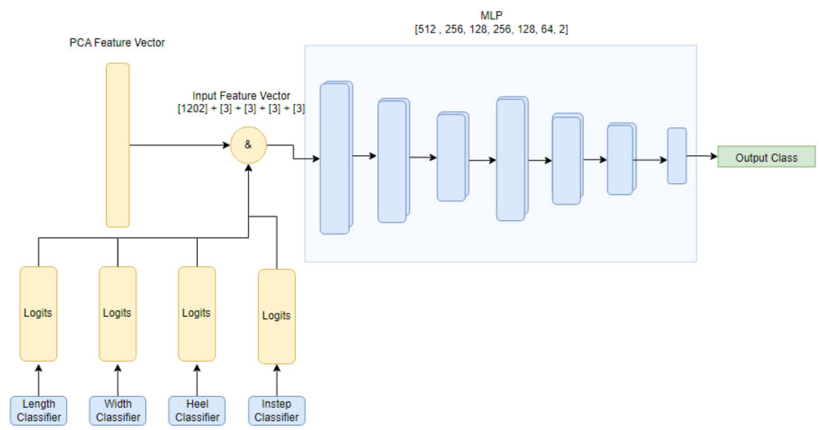


Figure 5.29. Visualization of the best performance model for the global fit classification.

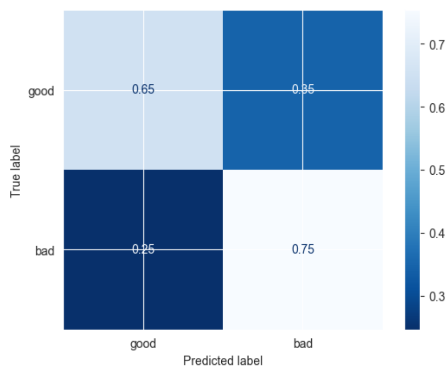


Figure 5.30. Confusion matrix for the best performance model on global fit classification task (PCA with LocalFits logits).

Chapter 6

Conclusions and Future work

Throughout this thesis work, we attempted to solve an issue that is becoming increasingly important by using a variety of approaches and methodologies. The absence of previous public works about foot comfort evaluation, at this point in time, created the necessary to put in place the foundations for good research and to begin making the first steps in this unfamiliar path. This project aims to expand the knowledge and understanding of how different models behave with different data information.

With parametric shapes of feet and shoes as starting points, the ultimate goal is to give the user an estimation of how comfortable a particular pair of shoes is by evaluating not only the overall shape but also local dimensions like the instep, heel, width and length. To be able to address the disadvantage of a limited user dataset as well as the resulting small trial sample size, it has been hypothesized that leveraging the information extracted from the singular local fits would create a flourishing environment for the harder and more confused task of the global fit making it more generalizable.

The thesis work follows a path that conducts different model experimentation, starting from a transfer learning model with PointNet and PointNet++ [24][26] and, due to the low performance and high complexity, ending in engineering several MLP models that utilized every time newly extracted features. Different trials has been performed, with the goal of adding every time new information. With this in mind we decided firstly to work with a raw utilization of the cloud points. Then, we continued with a PCA preprocessing, reaching good performance. From there we worked with a method in which we added new information such feature from the Toe and Heel manipulation and we managed to reach the current best performance.

This thesis work has also been focused on tackle a misalignment problem in the dataset. Due to the fact that, in the given dataset, shoes are misaligned or even deformed it was imperative the need of mitigating and addressing this problem. A first test that has been executed is on user cloud point deformation and, despite the fact that the non rigid deformation as been flawless it was lacking of the main important factor about human deformation: flexibility and range of motion. It has been after understanding this deficiency that we started our work in skeletonization on human foots.

The first tests with the skeletonization procedure managed to reach almost the best performance and, given the sensitivity on input changes shown by the models, is has been considered a good step in the right direction.

In conclusion, this work successfully shows the difficulties and the necessity of such task but also shows his feasibility and improvement path. During this thesis work, in fact, it has been managed to steadily increase our knowledge and performance by analyzing every time different aspects and lacks in the data representation. At the end of this journey it is possible to have the upmost confidence in the growth potential, improvement opportunities are right behind every corner, the future for the field is as promising as it gets.

6.1 Future improvements

In this section we will go trough the next trials that can be performed to increase the knowledge on the dataset and to further increase performance.

6.1.1 More and better Data

Based on the results and on the impact that the different additions had on the overall results it is clear that the fastest way to achieve better performance and increase generalizability of the model is to increase the number of samples in the dataset. The dataset, as things stand, is really unbalanced in both local fit task and global fit task. Is of course really helpful having new raw data but it is crucial, for a better training experience, to try to prioritize the balancing aspect rather than only increasing the sample number with random trials. It would be really good also to mitigate the difference in balancing between local fits and global fit classes. The global fit classes are really unbalanced towards the negative class while the local fit is unbalanced towards the positive class. This create confusion in the model and create an harder environment for the model's learning.

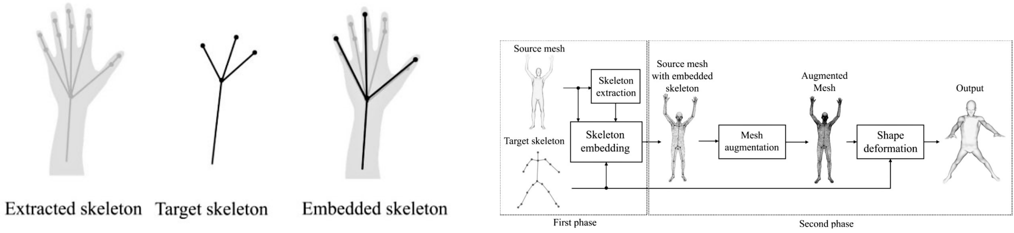
As a possible enhancement of the data acquisition is to notice that, in most related works about human foot and dataset analysis have been found specific key points on the user comfort and overall examination and investigation of the customer foot [13]. It would be interesting working on a updated version of the dataset that includes the user perspective on those pivotal location of the foot and shoe.

6.1.2 Mesh experimentation

It is possible also that by utilizing the Cloud Point representation we are leaving some performance and some information on the table. It would be interesting to try implementing a model that utilizes the Mesh representation of the shapes.

By utilizing the mesh representation is also possible to really integrate the work that Çağlar Seylan and Yusuf Sahillioğlu done on 3D shape deformation. In their work they tried to address the issue of obtaining new poses of an articulated subject with a new shape deformation approach consisting of two phases enabling the user to express the new pose as a simple stick figure. In other works the model should be able to repose an articulated character with a 3D stick figure and transfer a source mesh to this target skeleton [37]. At the moment there is no public code available but in a last minute email exchange with Seylan we managed to come into possession of the code.

Sadly it would have taken a little more time to perform some initial analysis about this new model but is surely a good step forward in a possible and hopeful right direction. The idea is to utilize the skeleton of one of user foot or shoe last to deform the shape of one into the other. In this way we expect to achieve a result similar to what we tried to achieve in Chapter 4.5 but with a model that has been trained to perform the best possible deformation without creating artifacts in the mash and, most importantly, by following the skeleton movement limitation of a human articulation.



(a) **(b)**

Figure 6.1. Skeleton embedding (a) and model summary (b). From 3D Shape Deformation Using Stick Figures [37]

To be able to work with the Mesh representation is also to notice that a change of model structure is needed. Our idea is to possibly start working with some kind of GNN implementation that unlocks the possibility of maintaining the link information between vertices and faces of the mesh shape.

Bibliography

[1] Iman Dianat, Johan Molenbroek Héctor Ignacio Castellucci. (2018) “*A review of the methodology and applications of anthropometry in ergonomics and product design*“, Ergonomics, 61:12, 1696-1720

[2] T. Spahiu, E. Shehi and E. Piperi2. (2015) “*Anthropometric Studies: Advanced 3D Method for Taking Anthropometric Data in Albania*“, IJIRSET, 4, 4

[3] Types of Body Movements, Biomechanics of Human Movement `https://pressbooks.bccampus.ca/humanbiomechanics/chapter/9-5-types-of-body-movements/`

[4] Paquet, V., and D. Feathers. (2004) “*Anthropometric Study of Manual and Powered Wheelchair Users.*“ , Ergonomics, 33

[5] Stančić, Ivo, Josip Music̀, and Vlasta Zanchi. (2013) “*Improved structured light 3D scanner with application to anthropometric parameter estimation.*" Measurement 46.1 716-726.

[6] Mits, Sophie and Coorevits, et co. (2011) "*Reliability and Validity of the INFOOT Three-dimensional Foot Digitizer for Patients with Rheumatoid Arthritis*" Journal of the American Podiatric Medical Association

[7] Lee, Yu-Chi, and Mao-Jiun Wang. (2015) "*Taiwanese adult foot shape classification using 3D scanning data.*" Ergonomics 58.3 513-523.

[8] Wang, Mao-Jiun J., et al.(2007) "*Automated anthropometric data collection from three-dimensional digital human models.*" The International Journal of Advanced Manufacturing Technology 32 109-115.

[9] Wu, H. B., et al. (2006) "*3d measurement technology by structured light using stripe-edge-based gray code.*" Journal of Physics: Conference Series. Vol. 48. No. 1.

[10] Hsiao, Hongwei, et al.(2015) "*Firefighter hand anthropometry and structural glove sizing: a new perspective.*" Human factors 57.8: 1359-1377.

[11] Jung, Kihyo, Ochae Kwon, and Heecheon You. "*Evaluation of the multivariate accommodation performance of the grid method.*" Applied ergonomics 42.1 (2010): 156-161.

[12] Aldo Laurentini. "*The Visual Hull Concept for Silhouette Based Image Understanding.* IEEE, 150–162, 1994.

[13] Yu-Chi Lee and Mao-Jiun Wang. "*Taiwanese adult foot shape classification using 3D scanning data.* Ergonomics, 2015.

[14] Ghodhbani, H., Neji, M., Qahtani, A.M. et al. "*Dress-up: deep neural framework for image-based human appearance transfer*, Multimed Tools Appl (2022).

[15] Ashmawi, Sahar & Alharbi, Maram & Almaghrabi, Ameerah & Alhothali, Areej. (2019). "*FITME: BODY MEASUREMENT ESTIMATIONS USING MACHINE LEARNING METHOD. Procedia Computer Science.*" 163. 209-217.

[16] Chandra, Richard & Febriyan, Fajar & Rochadiani, Theresia. (2018). "*Single Camera Body Tracking for Virtual Fitting Room Application*". 17-21. 10.1145/3192975.3192991.

[17] Mousa Mojarrad, Sedigheh Kargar. "*Measuring the Main Parameters of the Human Body in Images by Canny Edge Detector*", Science Journal of Circuits, Systems and Signal Processing. Volume 2, Issue 5, October 2013 , pp. 100-105.

[18] Tan Xiaohui, Peng Xiaoyu, Liu Liwen, Xia Qing, "*Automatic human body feature extraction and personal size measurement*", Journal of Visual Languages & Computing, Volume 47, 2018

[19] Chang, Hsien-Tsung et al. “"*A Dynamic Fitting Room Based on Microsoft Kinect and Augmented Reality Technologies.*” Interacción (2013).

[20] I. Krauss. "*ex-related differences in foot shape of adult Caucasians – a follow-up study focusing on long and short feet.* Ergonomics, 2011.

[21] MASAAKIMOCHIMARU, MAKIKOKOUCHI and MASAKODOHI "*Analysis of 3-D human foot forms using the Free FormDeformation method and its application in grading shoe last.* Ergonomics, 2000.

[22] OULIAN HONG, LIN WANG, DONG QING XU, JING XIAN LI. "*Gender differences in foot shape: a study of Chinese young adults.* Sports Biomechanics, 2011.

[23] Max Jaderberg, Karen Simonyan, Andrew Zisserman, Koray Kavukcuoglu "*Spatial Transformer Networks*" Computer Vision and Pattern Recognition, 2015

[24] Charles R. Qi Hao Su Kaichun Mo Leonidas J. Guibas "*PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation*" Computer Vision and Pattern Recognition

[25] An In-Depth Look at PointNet `https://medium.com/@luis_gonzales/an-in-depth-look-at-pointnet-111d7efdaa1a`

[26] Qi, Charles Ruizhongtai, et al. "*Pointnet++: Deep hierarchical feature learning on point sets in a metric space.*" Advances in neural information processing systems 30 (2017)

[27] Cao, S.; Zhao, H.; Liu, P. "*Semantic Segmentation for Point Clouds via Semantic-Based Local Aggregation and Multi-Scale Global Pyramid.* Machines 2023, 11.

[28] Ioffe, Sergey, and Christian Szegedy. "*Batch normalization: Accelerating deep network training by reducing internal covariate shift.*" International conference on machine learning. pmlr, 2015.

[29] Principal component analysis (PCA) https://www.ibm.com/docs/en/db2-warehouse?topic=procedures-principal-component-analysis-pca

[30] Principal Component Analysis (PCA) Explained Visually with Zero Math https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d

[31] Deng, Bailin, et al. "*A Survey of Non-Rigid 3D Registration.*" Computer Graphics Forum. Vol. 41. No. 2. 2022.

[32] Groueix, Thibault, et al. "*3d-coded: 3d correspondences by deep deformation.*" Proceedings of the european conference on computer vision (ECCV). 2018.

[33] Zeng, Yiming, et al. "*CorrNet3D: Unsupervised end-to-end learning of dense correspondence for 3D point clouds.*" Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[34] Myronenko, Andriy, and Xubo Song. "*Point set registration: Coherent point drift.*" IEEE transactions on pattern analysis and machine intelligence 32.12 (2010): 2262-2275.

[35] Andrea Tagliasacchi, Ibraheem Alhashim, Matt Olson, and Hao Zhang. 2012. "*Mean Curvature Skeletons.*" Comput. Graph. Forum 31, 5 (2012), 1735–1744.

[36] Schrder, Peter, Desbrun, Mathieu, Meyer, Mark, Schröder, Peter, Barr, Alan. "*Implicit Fairing of Irregular Meshes using Diffusion and Curvature Flow.*" (2000).

[37] Çağlar Seylan, Yusuf Sahillioğlu. "*3D Shape Deformation Using Stick Figures*" Computer-Aided Design 151 (2022).

Acknowledgments

Voglio ringraziare in primis il Profressor Galasso e il Dr Raccanelli per l'infinita disponibilità e l'opportunità data. Voglio ringraziare tutta la mia famiglia per avermi sostenuto e sopportato durante il percorso. Voglio ringraziare Michele e il nostro rapporto di simbiosi accademica. Vorrei ringraziare Aurora, Erica, Matteo, Lorenzo e Francesco per aver percorso insieme a me questo viaggio. Vorrei ringraziare ovviamente anche Stefano, Giovanni, Lorenzo, Guglielmo e Pietro e tutti coloro che mi sono stati vicino perché, anche se non direttamente, siete stati di supporto ad affrontare questo percorso. Non posso fare altro che ringraziare tutti, presenti e non presenti per avermi accompagnato e aiutato a crescere durante questo lungo percorso. Grazie a tutti.