Automatic Angle Recognition in Hallux Valgus

Pablo Izquierdo, Ana Calderon, Glenn Jenkins, Simon Thorne, Ian Mathieson

Cardiff Metropolitan University, Cardiff, South Wales, United Kingdom.

pabloia@kth.se; acalderon@cardiffmet.ac.uk; gljenkins@cardiffmet.ac.uk; sthorne@cardiffmet.ac.uk; imathieson@cardiffmet.ac.uk

***Abstract* - This paper describes a novel approach to modelling a specific orthopaedic condition, Hallux Valgus; it is a complex deformity resulting in more than 140 possible surgical correction procedures, each focusing on different components of the deformity. Modelling it involves a level of complexity that cannot be readily tackled by techniques traditionally available in the medical domain. It was, therefore, appropriate for us to utilise machine learning techniques; namely through neural network detection and isolation, complemented by angle detection. We present results of running a machine learning algorithm to detect the deformity and end with recommendations as to how it may be utilised in judging successful surgical outcomes.**

Keywords - Neural network, medical, anatomical modelling, xray, image recognition

I. INTRODUCTION

Hallux valgus is the technical term for the foot deformity commonly referred to as a bunion. This common condition is associated with significant morbidity, causing pain and functional limitation as well as the practical difficulty of accommodating the deformity in footwear. The immediate visually identifiable feature is deviation of the hallux in relation to the 1st metatarsal. However, the foot is a complex mechanical structure comprised of 26 bones and many articulations that re maintained by a sophisticated arrangement of ligaments and muscles. Therefore, the reality is that the visually prominent feature of 1st metatarsal belies a far more complex deformity that involves multiple bones and joints. As a result, more than 140 surgical correction procedures have been described, each focusing on different components of the deformity. Whilst many procedures achieve symptom resolution related to varying levels of deformity reduction, there is limited understanding of the bony relationships that are central to the deformity and should form the focus of surgical correction to most accurately predict both symptom and deformity reduction.

The objective of this project is to utilise machine learning techniques to assess the relationship between multiple bony segments on pre- and post- operative x-rays and patient satisfaction, to identify the optimal approach to surgical correction.

This initial work focuses on the construction of bone models from x-ray images. Typically the Hallux Valgus Angle (HVA) or Intermetatarsal Angle (IMA) are used as primary parameters in determining the severity of hallux valgus [1]. HVA has been shown to be a good predictor of surgical success, except in extreme cases [1]. There are various approaches to measuring HVAs in radiographic images including standardised manual approaches and computer aided ones [2]. Automatic recovery with machine learning and image processing techniques is more recent [3].

II. RELATED WORK

The current literature on computer vision and imagine processing comprises several methodologies. Two of them stand out within the biomedical domain: Object Recognition and Image Segmentation.

A. Image Segmentation

Image Segmentation is composed by two main approaches: Semantic Segmentation and Instance Segmentation. Semantic Segmentation is the task of clustering parts of images together which belong to the same object class [4] [5], whereas Instance Segmentation identifies each object within a class as an individual instance. [6] Image Segmentation has been proven valuable for several medical applications within many biomedical fields. Some examples are:

- Multiclass segmentation in chest radiographs [7]
- Brain tumour segmentation [8]
- Multipurpose image segmentation using U-Net [9] and X-Net [10]

B. Object Recognition

The current literature landscape regarding Object Recognition contains the following broadly categorized methods:

- Region Based Convolutional Neural Networks (RCNN) [11]
- You Only Look Once (YOLO) [12]

• Single Shot Multibox Detectors (SSD) [13]

Object Recognition has been broadly and successfully used within the biomedical domain. Some examples are:

• Object detection and recognition for diagnosis purposes [14]

• Bone detection in the pelvic area [15]

C. Neural Networks

Deep learning is a popular and powerful which has been applied to wide range of medical imaging problems. These include; the detection of lung conditions from 2D CT scans [16], segmentation of the vascular network of the eye in fundus images [17], vertebrae recognition [18] and the classification of haemorrhages in color fundus images [19]. A fully automated bone age assessment approach has been developed using CNNs [3]. More recently deep learning neural networks have been applied to this problem and U-Net (an established CNN,) used to segment the image [3]. The angle between the bones was then extracted using Contour Feature Distance (to identify the big toe) and image skeletelisation using mophologial thinning to construct the skeleton [3].

The YOLOv3 (You Only Look Once) Network, proposed by Redmon [20] was selected the task. This was due in part to it's simplicity and the limited nature of the available dataset and its use in related research [15].

The remainder of this paper is structured as follows: Section 3 provides a description of our methodology with reference to the techniques employed. Section 4 reports and provides a discussion of the results of the experiments undertaken. Finally Section 5 draws conclusions from this work and highlights areas for future investigation.

-III. METHODOLOGY

A neural network is used to identify sub images containing the phalanges of interest. The following section describe the neural network, its architecture, training and the dataset used. This is followed a detailed description of the proposed method for recovering the the angle phalanges in the recovered sub-images.

A. Network Architecture

The are several pre-made architectures for the YOLO Network. For this particular project, a slightly modified version "Tiny YOLO" configuration is employed. This architecture is formed by seven convolutional layers (or feature extraction layers), six pooling layers and two fully connected layers (in charge of the class prediction and its consequent bounding box) as shown in Figure 1. An x-ray image is the input and the coordinates of the resulting bounding boxes are the output of the network. The number of classes is reduced from 80 to 2 and batch size is set to 64.

Availability and quality of data constituted an issue for this task. In order to proceed with the project, a dataset of 281 non-copyrighted monochromatic images was collected from several internet sources. Most of the images are frontal foot pictures (as the intended model should only receive this type of input). However, other non related bone images were added to increase the robustness of the future model. An example image is provided in Figure 2.

B. Dataset Labelling

The data samples are labelled matching the requirements of the Neural Network (NN). To do so, two classes are established, F1 and F2. F1 is assigned to the proximal phalanx of the hallux and F2 to the metacarpal. Next, each data sample is hand-labelled by enclosing said bones under a bounding box and assigned its corresponding class. (see Figure 3) The accuracy of the bounding box is not as relevant, yet the general vector formed by the bone inside it must be identifiable.

C. Training

The NN is trained using Darknet, a framework developed by Redmon [22] that sets the architecture for the YOLO Network.

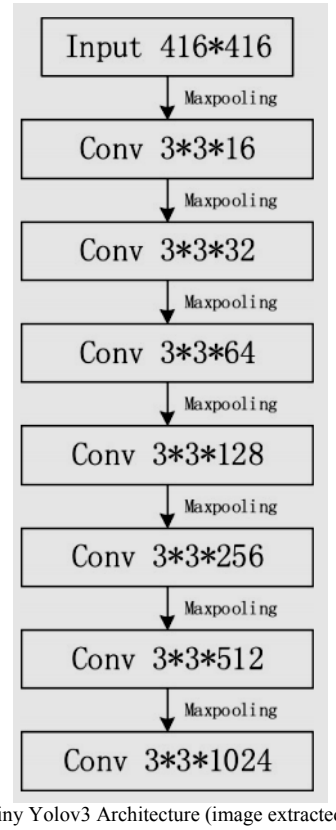


Figure 1. Tiny Yolov3 Architecture (image extracted from [21])

Figure 2. An example X-ray image showing the hallus valgus condition.

A set of pre-trained weights is used as the basis. In order to evaluate the performance of the network, two measures are employed:

- Mean Average Precision (mAP)
- Sum-squared error loss

The mAP is a commonly used evaluation metric within the object recognition field. It is defined as the mean over classes, of the interpolated Average Precision for each class [23]. The sum-squared error loss is a measure that specific to the YOLOv3 network [20].

D. Angle Estimation

From the initial image, the subfigures enclosed in the bounding boxes are extracted. This subfigures contain the phalanges of interest (see Figure 4). The angle between bones is extracted in 3 consecutive steps:

1) Image transformation: In this stage, a mask is applied to the images. This mask sets the value of a given pixel to black or white based on a preselected threshold. The threshold parameter is calculated through a heuristic process. Given a pixel value that can be between 0 and 255, a grid search is conducted for the whole image, with a lower threshold of 40 and an upper threshold of 90 (identified though experimentation for this dataset). The selected threshold value is the one that returns the closest black/white ratio to the mean black/white ratio of the image (see Figure 5).

2) Image Segmentation and Point Extraction: After the mask is applied, both images are segmented. Given an image represented as an m × n matrix A, as shown in Equation 1.

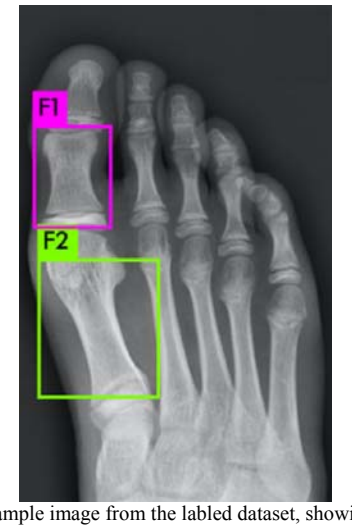
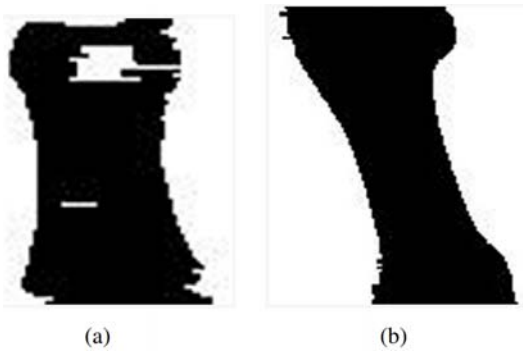


Figure 3. An example image from the labled dataset, showing the lables F1 and F2.

(a) (b)

Figure 4. The extracted phalanges, here (a) is the hallux and (b) is the first metatarsal. (a) (b)



(a) (b)

Figure 5. The phalanges with the mask applied again (a) is the hallux and (b) is the first metatarsal.

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad (1)$$

Two submatrices (S1,S2) of size s×n are formed, where s is determined by the closest natural number to 10% of n as shown in Equation 2 and 3.

$$S1 = \begin{pmatrix} s1_{11} & \dots & s1_{1n} \\ \dots & \dots & \dots \\ s1_{s1} & \dots & s1_{sn} \end{pmatrix} \quad (2)$$

$$S2 = \begin{pmatrix} s2_{m_{-s}1} & \dots & s2_{m_{-s}n} \\ \dots & \dots & \dots \\ s2_{m1} & \dots & s2_{mn} \end{pmatrix} \quad (3)$$

This is illustrated in Figure III-D3. For each submatrix, the average position of black pixels is computed. This position contains two coordinates that represent a point in space relative to the whole A matrix. This process is applied to both black and white images. Two points are extracted per image.

3) Vectorization and Angle Computation: Each pair of points forms a vector. For each bone, a vector is extracted. Given a pair of vectors v and u, the angle formed by their intersection is calculated using Equation 4.

$$\theta = \cos^{-1} \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|} \quad (4)$$

θ

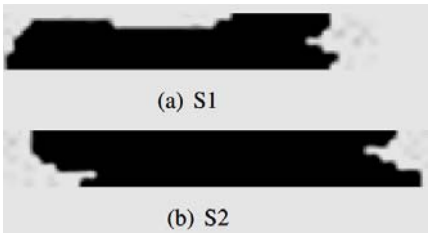


Figure 6. Upper submatrix S1 (a) and Lower submatrix S2 (b), extracted from images (a) and (b) in Figure 5

The angle is then converted from radians to degrees and given as the output of the system.

IV. RESULTS AND DISCUSSION

The performance of the network is measured by mean average precision and the sub-squared loss. The change in these two measures as training progresses is illustrated in Figure 7.

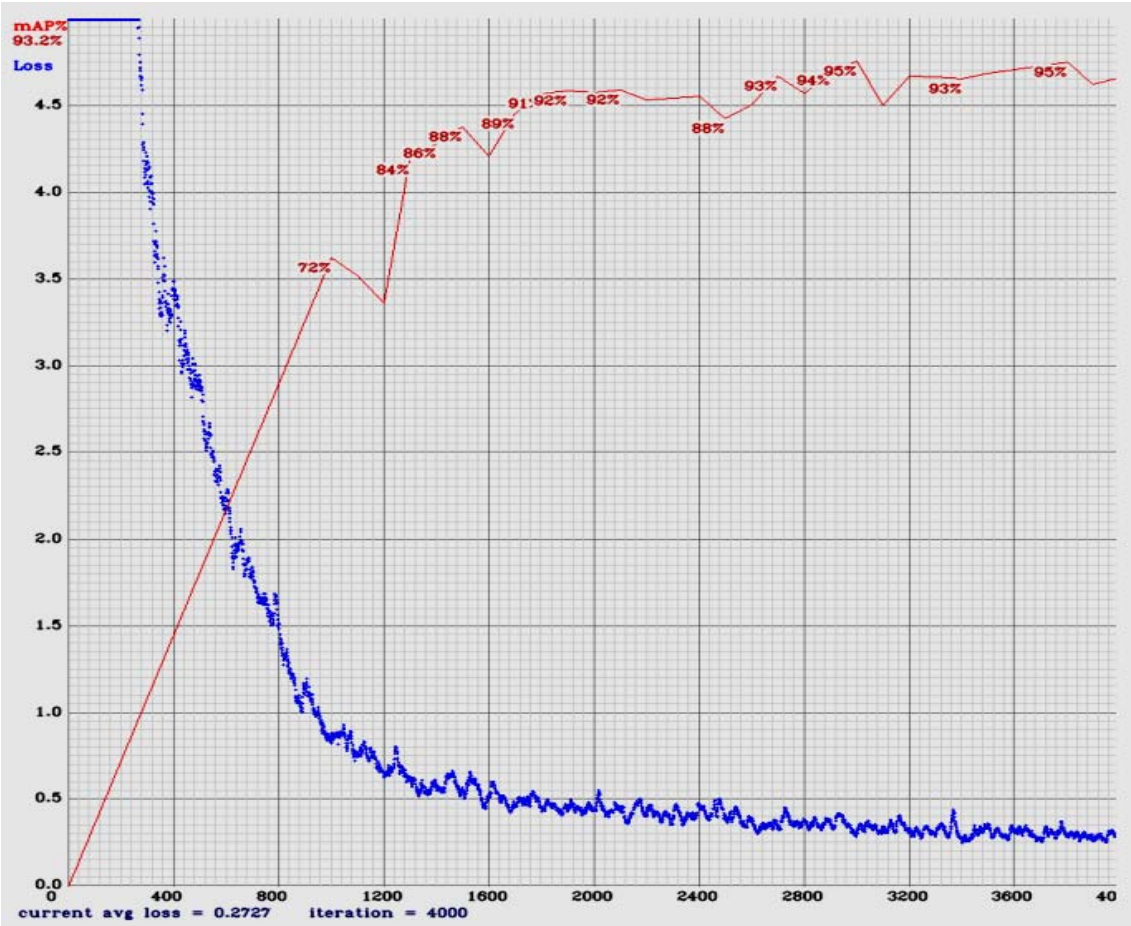


Figure 7. Blue - Loss, Red - Mean Average Precision (mAP), 4000 epochs

The mAP measure increases rapidly in the initial 1300 iterations, reaching a value of 0.84. This progression is then reduced, reaching a peak 0.95 at 3000 epochs. From this point on it remains constant, ranging between 0.93 and 0.95. The Loss value behaves analogously, similar to an asymptotic function. It decreases abruptly in the first 1000 iterations to then decelerates as it approaches 0. To avoid overfitting and given the shallow architecture of the network, training is stopped after 4000 epochs.

V. CONCLUSIONS AND FUTURE WORK

The results demonstrate that YOLOv3 network can identify both the hallux and 1st metatarsal from radiographic images. The method presented above can then be used to extract the HVA from these images. In the future, a full analysis of the obtained angle will be conducted. This will include comparing the results with those obtain by podiatric specialists using established measuring approaches [2]. Due to the limitations of our dataset, an object recognition approach was utilised. Recently image segmentation has been used to recover HVAs from x-ray images [3] the comparative merits of these approaches must now be explored. Neither approach is able to identify rotation of the bones concerned which research suggests should also be considered [24]. The shortage of suitable images and published datasets may be partially mitigated using augmentation approaches [25]. Finally the approach presented in this paper will be extracted and applied to further medical domains.

REFERENCES

[1] A. Deenik, A. Verburg, J. W. Louwerens, M. de Waal Malefijt, and R. de Bie, "Evidence of treatment algorithms for hallux valgus," JSM Foot Ankle, vol. 1, p. 1003, 2016.

[2] S. Srivastava, N. Chockalingam, and T. El Fakhri, "Radiographic measurements of hallux angles: a review of current techniques," The Foot, vol. 20, no. 1, pp. 27–31, 2010.

[3] K. Kwolek, H. Liszka, B. Kwolek, and A. G ˛adek, "Measuring the angle of hallux valgus using segmentation of bones on x-ray images," in International Conference on Artificial Neural Networks. Springer, 2019, pp. 313–325.

[4] M. Thoma, "A survey of semantic segmentation," arXiv preprint arXiv:1602.06541, 2016.

[5] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," Artificial Intelligence Review, vol. 52, no. 2, pp. 1089– 1106, 2019.

[6] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Simultaneous detection and segmentation," in European Conference on Computer Vision. Springer, 2014, pp. 297–312.

[7] A. A. Novikov, D. Lenis, D. Major, J. Hlad ˚uvka, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multiclass segmentation in chest radiographs," IEEE transactions on medical imaging, vol. 37, no. 8, pp. 1865–1876, 2018.

[8] H. Li, A. Li, and M. Wang, "A novel end-to-end brain tumor segmentation method using improved fully convolutional networks," Computers in biology and medicine, vol. 108, pp. 150–160, 2019.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[10] J. Bullock, C. Cuesta-Lázaro, and A. Quera-Bofarull, "Xnet: A convolutional neural network (cnn) implementation for medical x-ray image segmentation suitable for small datasets," in Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 10953. International Society for Optics and Photonics, 2019, p. 109531Z.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21–37.

[14] Z. Loncova, L. Hargas, D. Koniar, A. Simonova, and B. Kozacek, "Objects detection and recognition in biomedical microscopic images for the purpose of non-invasive and more precise diagnostic," in 2017 Progress In Electromagnetics Research Symposium-Spring (PIERS). IEEE, 2017, pp. 3159–3165.

[15] Z. Krawczyk and J. Starzynski, "Bones detection in the pelvic area on ´ the basis of yolo neural network," in 19th International Conference Computational Problems of Electrical Engineering. IEEE, 2018, pp. 1–4.

[16] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1207–1216, 2016.

[17] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," IEEE transactions on medical imaging, vol. 35, no. 11, pp. 2369–2380, 2016.

[18] S. Liao, Y. Zhan, Z. Dong, R. Yan, L. Gong, X. S. Zhou, M. Salganicoff, and J. Fei, "Automatic lumbar spondylolisthesis measurement in ct images," IEEE transactions on medical imaging, vol. 35, no. 7, pp. 1658–1669, 2016.

[19] M. J. Van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1273–1284, 2016.

[20] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[21] Z. Yi, S. Yongliang, and Z. Jun, "An improved tiny-yolov3 pedestrian detection algorithm," Optik, vol. 183, pp. 17–23, 2019.

[22] J. Redmon, "Darknet: Open source neural networks in c," http://pjreddie.com/darknet/, 2013–2016.

[23] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," in Asian Conference on Computer Vision. Springer, 2016, pp. 198–213.

[24] P. Dayton, M. Feilmeier, M. Kauwe, and J. Hirschi, "Relationship of frontal plane rotation of first metatarsal to proximal articular set angle and hallux alignment in patients undergoing tarsometatarsal arthrodesis for hallux abducto valgus: a case series and critical review of the literature," The Journal of Foot and Ankle Surgery, vol. 52, no. 3, pp. 348–354, 2013.

[25] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. Yeshiwas, T. Alkasab, G. Choy, and S. Do, "Fully automated deep learning system for bone age assessment," Journal of digital imaging, vol. 30, 03 2017.