Creating Accurate Representations of DXA Scans from 3D Optical Body Surface Scans Using Deep Learning

Lambert T LEONG [1,2], Michael C WONG [1], Yong En LIU [1], Nisa N KELLY [1], Michaela PIAZZA [3], Siobhan GARRY [4], Steven B HEYMSFIELD [5], John A SHEPHERD [1]
[1] University of Hawaii Cancer Center, Honolulu HI, USA;
[2] University of Hawaii at Manoa, Honolulu HI, USA;
[3] National Institute of Allergy and Infectious Diseases, Bethesda MD, USA;
[4] Hologic, Inc., Santa Clara CA, USA;
[5] Pennington Biomedical Research Center, Baton Rouge LA, USA

https://doi.org/10.15221/21.35

Background

Dual energy X-ray absorptiometry (DXA) has long been the gold standard for quantifying body composition. High and low energy X-ray images are acquired of the whole body and used to derive fat, lean, and bone composition. Specialized DXA software is used to further analyze whole body images to derive more specific measurements of the whole body and many sub regions. These measurements include visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) as well as regional and whole-body fat and lean masses. DXA imaging is a powerful tool and the information derived is clinically and prognostically relevant for diseases concerning bone density, obesity, and diabetes. While accessibility to DXA has improved over the years, there are still some locations and situations in which acquiring a DXA scan is not possible.

Three-dimensional (3D) body scanning technologies offer an accessible method for accurately capturing the 3D surface of an individual. Recently, 3D body scans have been used to study body shape and the correlation to many diseases, health markers, and even DXA derived measures of body composition [1]. It is hypothesized that body shape, as measured by 3D body scanners, is highly correlated to and is a product of the underlying boney structures and soft tissues captured by DXA.

Previous work explored the relationship between 3D body and DXA scans mainly with principal component analysis (PCA) and linear mappings. PCA models were able to reasonably predict fat, lean, bone, and soft tissue images of participants. A main drawback in that work, however, was the accuracy of predicting the forelimbs due to pose variation in both the 3D body and DXA scans. The PCA methodology also consisted of two sex specific (male and female) models since there are significant shape differences with respect to sex[2,3]. A neural network generative adversarial network (GAN) approach using a pix2pix architecture was also previously implemented and was able to learn the pose variation and produce more DXA realistic images. However, the results from the GAN approach came with the caveat of having a small sample size that resulted in a small hold-out test set. We have since compiled a larger data set and we aim to construct neural network models to learn the shape and appearance of both scan types as well as a translational mapping which would allow for the production of a Pseudo-DXA scan.

In this work, we seek to further explore the relationship between 3D body scans and DXA imaging by deriving a learned mapping from 3D body scans to DXA scans. The goal is to produce a model that can predict DXA equivalent images from a 3D body scan in what we call a Pseudo-DXA. In addition, we build on previous work by changing the predictions targets to the raw DXA image as opposed to the derived fat, lean, bone, and soft tissue image used in prior modeling. The raw DXA data can be used to derive various image types used in prior work as well as be analyzed by commercially available DXA software.

Methods

We took a modular approach to our modeling that allowed us to leverage large sources of unpaired scan data. Our approach involved constructing and pretraining a variational auto-encoder (VAE) to learn the shape and pixel variation of a raw DXA scan. We use a total 15,000 raw DXA images on adults and children curated from multiple clinical studies. The VAE architecture consisted of two components or sub networks which includes an encoder and a generator. The encoder is fed the raw DXA image and it learns a meaningful compressed latent feature encoding of the data. The generator then takes the latent encodings and learns to reconstructs or generates the original input image. DXA input images were augmented using rotation, translation (X and Y), and random crop outs before being input into the

DXA VAE. We used a randomly initialized Densenet121 as the encoder, a latent space of 8192 nodes, and a custom generator based off the super resolution generative adversarial network (SR-GAN) [4]. A VGG-16 [5] perceptual loss was used to evaluate the VAE's ability to learn the DXA image and reconstruct it. Training was stopped when the perceptual loss on the validation set ceases to decrease. The pretrained VAE model and weights were frozen and the generator is used to build the Pseudo-DXA model

Participants recruited for the Shape Up! Adults Study received whole body DXA scans (Hologic Inc., Marlborough, MA, USA) and 3D body scan on a Fit3D Proscanner (Fit3D Inc., San Mateo, CA, USA) on the same day. These participant scans accounted for all of our paired data. The Fit3D Proscanner outputs 3D meshes of scanned individuals, and these meshes were standardized to 110,000 vertices, standardized to a T-pose[1], and registered to the origin using Meshcapade (Tübingen, Germany)[6]. A pointnet[7] architecture was used to learn the 3D information and map that information to the pre-trained DXA generator. Meshes were down sampled by 20% using a random sampling before they were input into the network.

The final paired dataset was split into a train, validation, and hold-out test set using an 80%, 10%, and 10% split. The pseudo-DXA model received the 3D mesh vertices as input and output a predicted raw pseudo-DXA image. A perceptual loss function was used again to evaluate the reconstruction of pseudo-DXAs from 3D meshes and training was halted when the loss failed to decrease further. The 3D mesh hold-out test set were sent through the final model for predictions and the quality of the DXA predictions were evaluated using peak signal to noise ratios (PSNR) and structural similarity index (SSIM)[8]. For 16-bit images, acceptable reconstructions have a PSNR between 20-25 and a SSIM of 1 is considered to have exact image similarity.

Results

A total of 1364 3D mesh and DXA pairs were used to train and evaluate our mesh to DXA model, 656 males and 708 females. Participant's age ranged from 18 to 90 years old, height ranged from 117.2 to 195.6cm, weight ranged from 38.6 to 206.3 kg, and body mass index (BMI) ranged from 15.9 to 82.1 kg/m^2. Pseudo-DXA images were compared to the actual DXA image for each participant in the test set. The average values for the mean squared error (MSE), PSNR, and SSIM, were 3.22e-3 kg, 25.73 dB, and 0.88, respectively. The full test set results and an example of a 3D mesh input, its paired real DXA, and its pseudo-DXA prediction is shown in Figure 1.

Discussion and Conclusion

We demonstrate a proof-of-concept method for predicting accurate, DXA equivalent images from 3D mesh scans. Pretraining the DXA VAE with augmentation caused the generator to learn an average or standardized positioning of the participants on the DXA scans. As a result, the generator repositioned participants to the center of the image and this was also true for the mesh to pseudo-DXA images. In practice, there are always a small amount of positioning differences and even multiple scans of the same participant on the same day can have slight positional difference. As such, the use of a simple pixel to pixel comparison metric like MSE is not solely sufficient to evaluate reconstructions because it is not invariant to positional difference. We used PSNR and SSIM metrics to address the short comings of MSE. In future work we plan to perform subregional analysis to derive fat, lean, and bone composition using well established dual energy quantitative theory. Analysis of pseudo-DXA images, versus predicting clinical DXA measures, allows for the investigation of custom regions of interest as well as estimate the standard measures. Our Pseudo-DXA model may offer a more accessible method for evaluating regional body composition important to many fields of study.

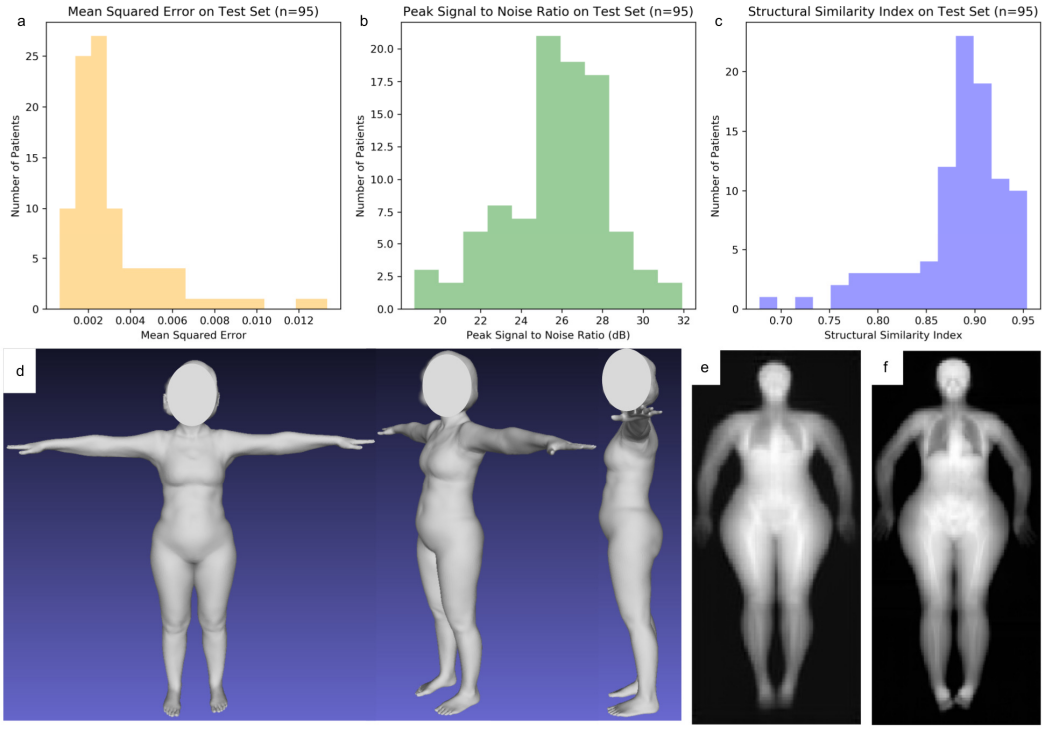

Figure 1. Pseudo-DXA model results on hold out paired test data. Frequency distribution of mean squared error (a), peak signal to noise ratio (b), and structural similarity index (c) values on the test set when comparing pseudo-DXA images to original images. (d) Example of a 3D mesh input in the T-pose. (e) Pseudo-DXA predicted image (low energy) from the input mesh. (f) Paired real raw low energy DXA image that were acquired at the same time as the 3D mesh.

References

1. Wong, M.C., *et al.* A pose-independent method for accurate and precise body composition from 3D optical scans. *Obesity* (2021).
2. Ng, B.K., *et al.* Detailed 3-dimensional body shape features predict body composition, blood metabolites, and functional strength: the Shape Up! studies. *The American journal of clinical nutrition* (2019).
3. Shepherd, J.A., *et al.* Modeling the shape and composition of the human body using dual energy X-ray absorptiometry images. *PloS one* **12**, e0175857 (2017).
4. Ledig, C., *et al.* Photo-realistic single image super-resolution using a generative adversarial network. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 4681-4690 (2017).
5. Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. in *European conference on computer vision* 694-711 (Springer, 2016).
6. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. & Black, M.J. SMPL: A skinned multi-person linear model. *ACM transactions on graphics (TOG)* **34**, 248 (2015).
7. Qi, C.R., Su, H., Mo, K. & Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 652-660 (2017).
8. Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**, 600-612 (2004).