

ShapeNet: Age-focused Landmark Shape Prediction with Regressive CNN

Youshan Zhang
Computer Science and Engineering
Lehigh University
Bethlehem, PA, USA
yoz217@lehigh.edu

Brian D. Davison
Computer Science and Engineering
Lehigh University
Bethlehem, PA, USA
davison@cse.lehigh.edu

Abstract—Deep neural networks are widely used in the segmentation and classification of medical images. However, little work has addressed the prediction of shapes based on population data over time as a regression problem. In this paper, we introduce a regressive convolutional neural network for landmark-based shape prediction. Unlike the conventional CNN model, the proposed network takes the input of a target age, and outputs the corresponding shape for that age. Experimental results demonstrate the effectiveness of the proposed ShapeNet to predict corpus callosum and mandible shapes with correct topology and accurate fitting that matches real-world scenarios. The proposed ShapeNet can predict the shape variation of high dimensional and nonlinear data, which is often critical to understanding the processes that change the shape of anatomy in biology and medical fields.

Index Terms—Shape prediction, Age-related disease, Regressive convolutional neural network

I. INTRODUCTION

With the emergence of many types of medical imaging technologies such as CT, MRI, fMRI, PET and DTI, medical shape analysis is necessary to automatically predict shape changes in anatomical structures. Deep neural networks have been extensively applied to shape analysis in the fields of biology and medicine [1]–[6]. Most deep neural networks address the segmentation [1], [4], [7], [8] and classification [9]–[12] of medical images. One of the most well known medical segmentation convolutional neural network (CNN) architectures is called U-Net [8]. There are two prominent novelties in U-Net: it combines the same number of upsampling and downsampling layers, and so-called skip connections combine opposing convolution and deconvolution layers. Later, more modified U-Nets were proposed for image segmentation. Çiçek et al. [13] proposed a U-Net for 3D data. Milletari et al. [4] extended U-Net to incorporate ResNet-like residual blocks and a Dice loss layer to reduce segmentation error. Al Arif et al. [1] modified the U-Net architecture and formed their SPNet to predict the shape changes instead of the segmentation maps, but the output is still an image, as opposed to a structured shape.

Classification is also an essential task in medical image analysis. Suk et al. [14] applied CNN to classify the cate-

gories of Alzheimer’s disease. Kawahara and Hamarneh [15] developed a multi-stream CNN to classify skin lesions. Also, Setio et al. [16] used a multi-stream CNN to classify nodule vs. non-nodule points in chest CT scans. Gao et al. [17] combined CNNs and recurrent neural networks to classify different grades of nuclear cataracts in slit-lamp images.

In addition, deep neural networks are applied in regression problems in medical images. Anatomical object boundary localization, such as landmarks, has been an important preprocessing step in segmentation tasks and in clinical workflows for therapy planning and intervention. Some works have addressed the location of landmarks. For example, Payer et al. [18] used modified CNNs to directly predict landmark locations. The output of the neural network is the landmark map. Other work, including Xie et al. [19], employ a structure regression model to detect the locations (but not boundaries) of cells with CNNs.

Neurodegenerative disease is difficult to detect in early stages, and shape deformation of a specific brain structure is highly associated with age. Neurosurgeons desire to predict future brain pathology changes for the diagnosis of patients. However, few deep neural networks have addressed shape variation based on population data. Shape analysis is difficult in the medical field since shape data are typically high-dimensional and nonlinear. In addition, most deep neural networks are trained by a pixel-wise loss function, making it difficult to maintain topological shape information. Conventional methods include using geodesic regression [20]–[22] to model shape changes, which assume that the shape has a linear (or polynomial) relationship between time and shape topology. However, we cannot guarantee that this assumption is correct. In our work, we make no such a priori assumption by applying a regressive neural network to predict the shape changes. In addition, differing from the traditional CNN, we propose a regressive CNN, which ends with a regression layer. In this paper, we focus on aged-related shape changes; we aim to predict a pathology shape given a specified age, which could provide guidance for the diagnosis of neurodegenerative disease. Therefore, the input of our proposed ShapeNet is the target age (which is just one number), but the output is a shape (whose dimensionality is naturally much bigger than that of the input) for that age. The predicted corpus callosum, mandible

and amygdala shapes show a correct topology and accurate fitting that matches scenarios in real cases. We have several challenges in developing a high-quality model.

- The shapes from different pathologies are different; ideally, a neural network architecture should be applicable to different shapes with suitable training.
- We have a limited number of shape instances in our data, which limits the generalizability of developed models.

Our contributions are two-fold:

- 1) We are the first to propose a regressive convolutional neural network named ShapeNet for landmark age-focused shape prediction. Unlike conventional CNNs, the input is only a single numeral (age), and the output is the corresponding shape instead of a label per pixel in a segmentation map. We demonstrate how to extend our ShapeNet into 3D space.
- 2) We validate our model using both synthetic data and human MRI and CT data. The results indicate that given a target age, ShapeNet is generally able to predict the shape with correct topology, and with much higher accuracy than a state-of-the-art geodesic approach.

II. PROBLEM DESCRIPTION

Previous age-focused landmark shape prediction methods are highly dependent on manifold learning and have a linear (or polynomial) relationship between age and the shape [20]–[22]. Also, the inputs of traditional CNNs are always images, and the output is the image map. In our case, there is such a relationship between target and shape: the input of ShapeNet is the target age and the output is the 2D shape. We formalize the age-focused shape prediction problem as follows.

Given the age $X = \{x_i\}_{i=1}^N$, with its corresponding shape $Y = \{y_i\}_{i=1}^N$ (each y_i is a vector with the size of $M \times 1$ that can be interpreted as $2 \times (M/2)$, which is a 2D landmark shape as a list of coordinates), we aim to predict the shapes ($Y' = \{y'_i\}_{i=1}^{N'}$) given a specific age ($X' = \{x'_i\}_{i=1}^{N'}$), and we minimize the half sum-of-squared errors in Eq. 1. Therefore, we aim to design a deep neural network to learn the shapes and make predictions based on the learned model.

$$E = \frac{1}{2} \sum_{i=1}^N (y_i - y'_i)^2, \quad (1)$$

III. METHODS

In this section, we show the architecture of our proposed ShapeNet, and analyze the significance of predicted shapes compared with ground truth shapes.

A. ShapeNet

To overcome the smaller number of features of the input layer (age: 1×1), we need an appropriate network architecture for extracting better feature representations to be able to model the relationship between the age and nonlinear landmark shape. Fig. 1 shows the ShapeNet architecture. It contains two major modules: feature extraction and prediction. Notably, in the feature extraction module, it consists of k blocks, and each

block has six layers (Convolution (Conv), Batch normalization (BN), Rectified linear units (ReLU), Average pooling (AP), Cross-channel normalization (CCN), and Max pooling (MP)). This block is able to extract features from the simple numerical target age, and feeds them into a fully connected (FC) layer for regressive shape prediction. The input of ShapeNet is the age, and the output shape is from the final regressive layer, which is a novelty of our model. Fig. 1 also lists the number of features of each layer.

In each block, we employ the Conv layer to generate more features from the previous layer (e.g., the first Conv layer has the filter size of $[1, 1]$, number of filters: 20, stride size of $[1, 1]$ and zero padding. Hence, the final output size is 1×20). The ReLU layer reduces the number of epochs to achieve the training error rate better than traditional tanh units. The normalization layer increases generalization and reduces the error rate. In addition, ReLU and normalization layers do not change the size of the feature map. The pooling layers summarize the outputs of adjacent pooling units. The dropout (Drop) layer randomly sets input elements to zero to prevent overfitting. The loss function of the last regression layer is the same as our error function, which is defined in Sec. II. The size of each shape M can affect the number of k blocks. A smaller value of M causes a larger value of k , since a smaller M will need more layers (extracting more features) to learn the pattern of shapes¹.

One of the most obvious merits of ShapeNet is that more features can be extracted from k feature extraction blocks. Fig. 2 depicts an example of a pentagon shape prediction using a 5-block trained ShapeNet; it shows in detail the values represented in the layers of the first block and the final few layers which reflect a complete pentagon shape. With more features extracted in k blocks, we can easily build the relationship between model and the predicted shape. Therefore, ShapeNet is able to predict pentagon shape at an arbitrary age (63 in this figure).

B. Significance analysis

To illustrate the significance of predicted shapes, we first calculate the R^2 statistic and then report the p-values of predicted shapes when comparing with true shapes.

1) *Extended R^2 statistic*: The R^2 statistic has proven to be a useful metric to indicate the significance of the linear regression model [23]. Accordingly, we extend the R^2 statistic to fit our unknown relationship regression. The range of the R^2 statistic is between $[0, 1]$; the higher the R^2 value, the more variation is explained by the model, and the better the model fits the shape.

$$R^2 = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} = 1 - \frac{\sum_{i=1}^N S_{residual}}{\sum_{i=1}^N S_{total}}, \quad (2)$$

where $S_{residual} = \sum_{k=1}^M (y_{ik} - y'_{ik})^2$, and $S_{total} = \sum_{k=1}^M (y_{ik} - \bar{y}_k)^2$, N is the number of shapes, M is the number of values making up the points of a shape, \bar{y} is the mean

¹Source code is available at: <https://github.com/heavention93/ShapeNet>.

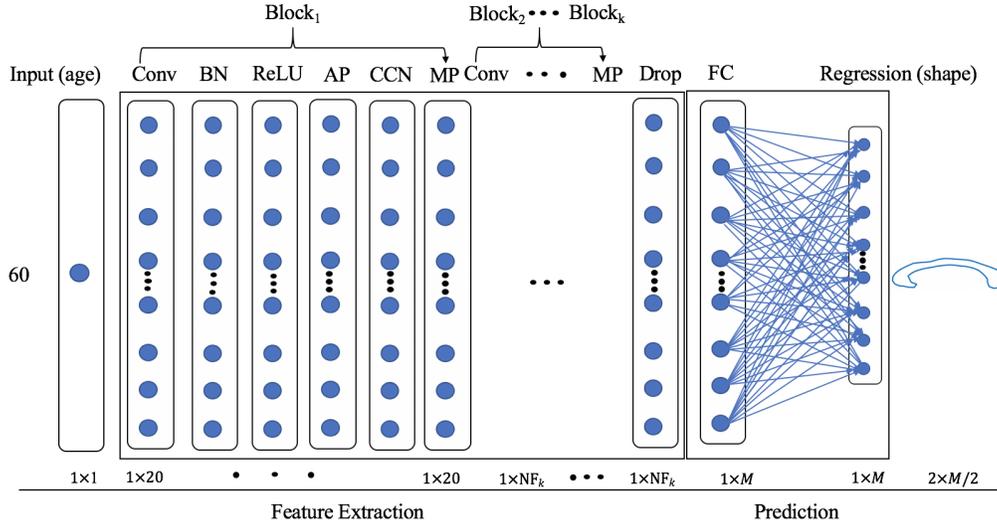


Fig. 1: ShapeNet consists of k feature extraction blocks. Each block has six layers. The input is the target age with unit dimensions, and the output regression layer is the shape of a structure with dimensions of $1 \times M$, which can be interpreted as $2 \times M/2$ (using 2D landmarks as shown in rightmost shape; or $3 \times M/3$ for 3D landmarks). (Conv: Convolution, BN: Batch normalization, ReLU: Rectified linear units, AP: Average pooling, CCN: Cross-channel normalization, MP: Max pooling, Drop: dropout and FC: fully connected. The number of features of each layer is denoted below each layer. NF: the number of filters).

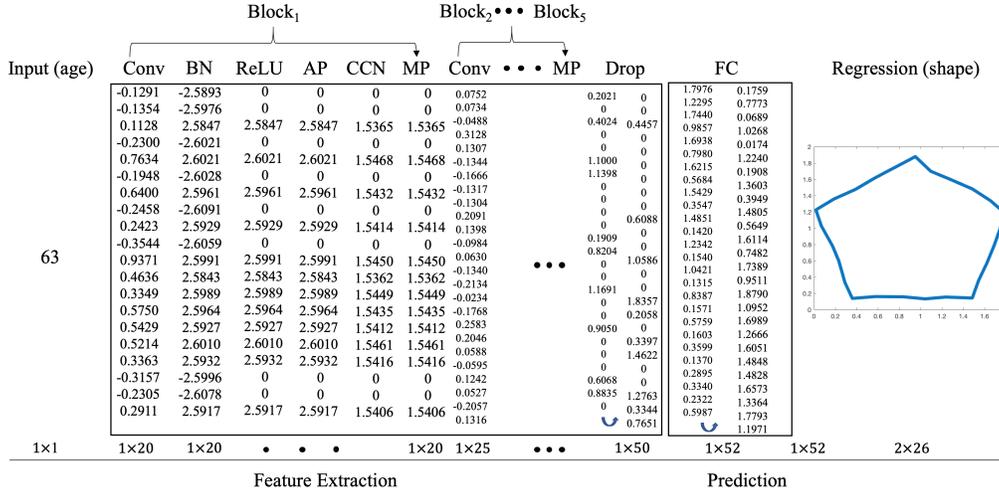


Fig. 2: An example of features from different layer using ShapeNet for predicting the shape of a pentagon at “age” 63.

shape of Y , that is we calculate the mean landmark positions $\bar{y} = 1/N \sum_{i=1}^N y_i$. Since both $y_{ik} - y'_{ik}$ and $y_{ik} - \bar{y}_k$ are vectors of differences, we add a sum to calculate $S_{residual}$ and S_{total} .

2) *Hypothesis tests*: We test a hypothesis to show the significance of predicted shapes and true shapes. The null hypothesis is H_0 : there is no significant difference between predicted shapes and original shapes (they are from the same distribution). We perform two-sample t-tests to calculate the p-values². Since each prediction will have a p-value, we calculate the mean p-value per dataset.

²If the p-value is less than 0.05, we strongly reject the null hypothesis (meaning that predicted shapes are likely to significantly differ from original shapes); if the p-value is larger than 0.05, we fail to strongly reject the null hypothesis (suggesting that predicted shapes are similar to original shapes).

IV. RESULTS

In this section, we validate our model using both synthetic data and real data. We further compare our results with linear regression [24] and a geodesic regression model [20]³, and conduct significance analyses, which demonstrates the applicability and goodness of our model.

A. Shape Predictions

Pentagon: To show the applicability of our model in shape prediction, we first apply ShapeNet to predict shape variations of a synthetic Pentagon dataset. This data contains a collection of 50 pentagon shapes with pseudo age X from 1 to

³We re-implemented Fletcher’s geodesic regression model [20].

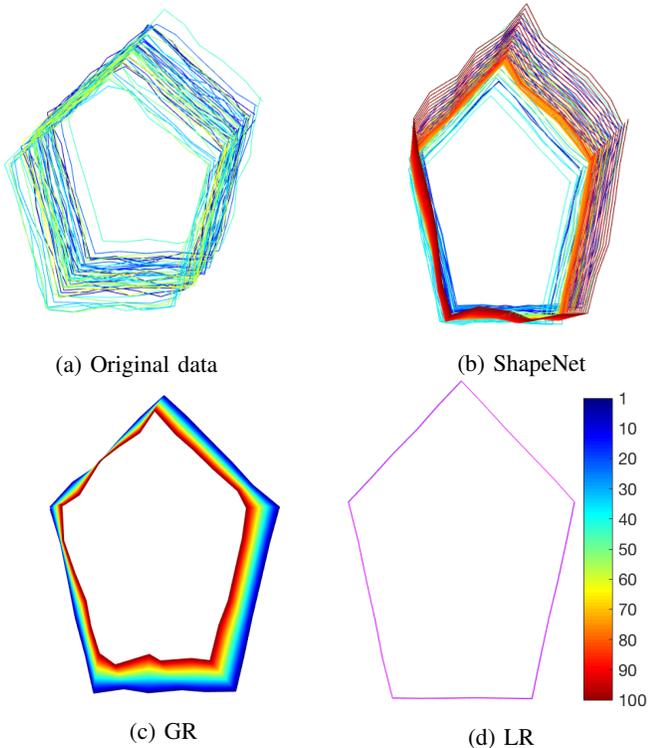


Fig. 3: Shape prediction results for ages 1-100 of pentagon data, given data from ages 1-50 (GR: geodesic regression and LR: linear regression).

50. Each pentagon has 26 2D points. We aim to predict unseen pentagon shapes when age X is from 51 to 100. The prediction results are shown in Fig. 3. There are 100 shapes in each of Fig. 3(b-d), but some of them are overlapping with each other (especially in Fig. 3(d)). The predicted pentagon shapes from geodesic regression model shrink with the increasing of x , which is not a correct trend as shown in original shapes. However, ShapeNet can almost recover the shape variations. This example illustrates the applicability of our model in analyzing the shape variations of data. Fig. 4 shows the root mean square error (RMSE) and loss for the training of a pentagon model.⁴

Corpus callosum shape: The corpus callosum data includes 32 MRI scans of human brain subjects, with ages from 19 to 90, which is from Open Access Series of Imaging Studies (OASIS) database www.oasis-brains.org. The boundaries of these segmentations are sampled with 64 2D boundary landmarks [20], [22]. We later add one pair point to connect the starting and ending points (the final size of Y is 130×32 , which can be reshaped into $2 \times 65 \times 32$). Fig. 5 compares the predicted shapes of ShapeNet with the geodesic regression model (the training ages are excluded). The predicted shapes show that the anterior, mid-caudate and posterior of corpus callosum is changed with the increasing of age. Compared

⁴While not shown, the training RMSE and loss for models of other datasets are similar to pentagon data.

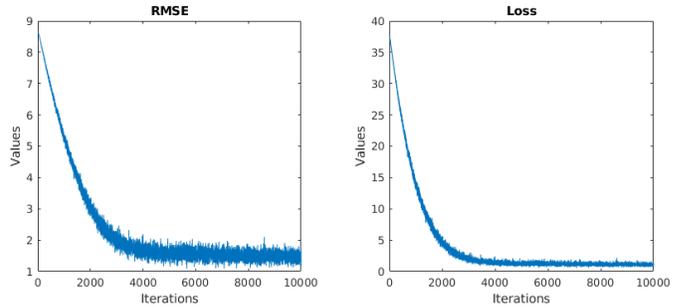


Fig. 4: The training RMSE and loss of pentagon data.

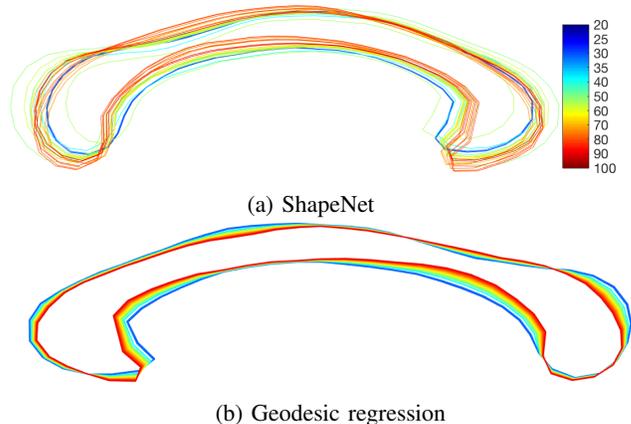


Fig. 5: Shape prediction results of corpus callosum using ShapeNet and geodesic regression.

to predicted results from geodesic regression, we find that ShapeNet can predict shapes with larger variations over ages, much like the original shape data.

Mandible shape: Our mandible data is extracted from a collection of CT scans of human mandibles, with 77 subjects aged from 0–19 [25]. We sample 400 2D points on the boundaries. Visualization of 3D mandible raw data and 2D shape data are shown in Fig. 6. Fig. 7 shows the predicted mandible shapes with ages from 19–25 using trained ShapeNet from 77 subjects. We find that mandible shapes have larger variations in the temporal crest, middle part and the base.

B. Significance analysis

The statistical significance of the predicted shapes is examined using methods described in Section III-B. Table I compares the R^2 statistic of ShapeNet and the traditional geodesic regression model. The R^2 values of three datasets from our ShapeNet are far larger than that of the state-of-the-art geodesic regression model. The lower value for the geodesic regression model gives the suggestion that shape variability is not well-modeled by age, since age only describes

TABLE I: R^2 statistic of predicting shapes

Datasets	Pentagon	Corpus callosum	Mandible
Geodesic regress.	0.0223	0.0234	0.0873
ShapeNet	0.3911	0.3854	0.1738

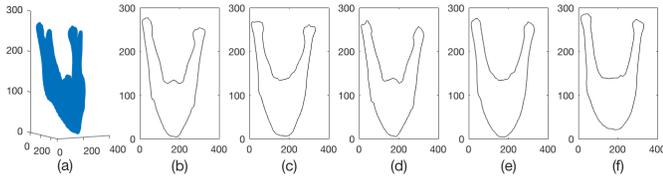
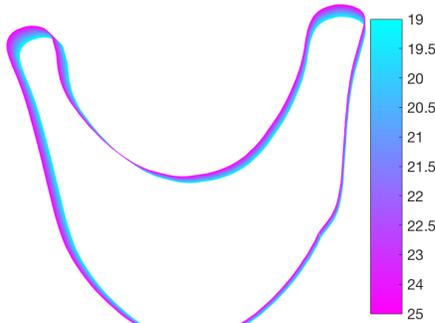
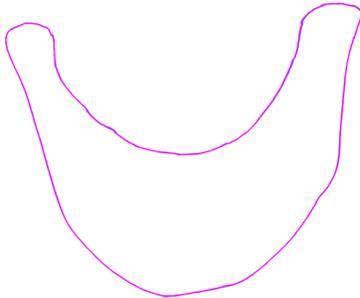


Fig. 6: An example of one 3D mandible shape and five 2D mandible shapes from different subjects.



(a) ShapeNet



(b) Geodesic regression

Fig. 7: The predicted shapes of human mandible shapes with the increasing of age (as shown by the color).

a small fraction of the shape variation. It is true that other factors (gender, weight etc.) will affect the changes in shape. However, the coefficient of determination (R^2 values) of our model demonstrates that age is an important factor that affects these shapes. The most essential benefit of the proposed ShapeNet over traditional methods is that a larger number of distinct shapes can be predicted using ShapeNet. Therefore, our shape prediction approach is better than the state-of-the-art geodesic regression method.

The mean p-values of the three datasets are: 0.6365, 0.9392 and 0.2497, respectively. All results are from two-sample t-tests and cannot reject the null hypothesis, which implies that the predicting shapes are at least somewhat similar to the true shapes (especially for corpus callosum dataset, in which the predicted shapes almost recover the original shapes).

C. Model extension

To demonstrate further the generalization of our model, we extend ShapeNet into 3D space since many medical images

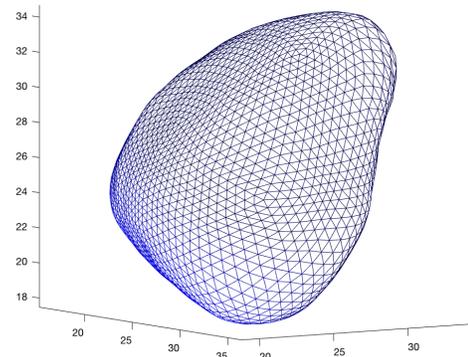


Fig. 8: The mean 3D amygdala shape.

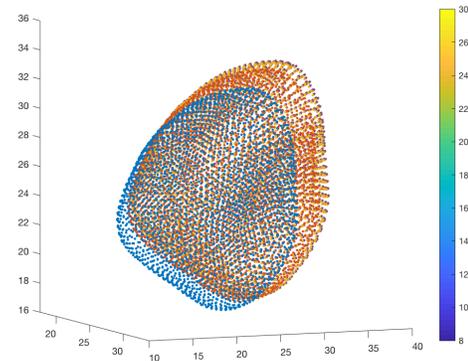


Fig. 9: The predicted shapes of human amygdala shapes with the increasing of age.

are in three dimensions using an additional task. In general, shape analysis will be difficult in 3D space due to the increased complexity. However, we can extend ShapeNet into 3D shape prediction by simply reshaping the data into the size of $3 \times M/3$ analogously to the last step of Fig. 1. To the best of our knowledge, geodesic regression has not been defined for 3D.

We test our 3D ShapeNet using human amygdala data [26], which contains 46 subjects aged from 8 to 24. We sample 2562 3D points on all 3D surfaces. Fig. 8 shows the mean of the 3D amygdala shape. The predicted shape changes are shown in Fig. 9 with R^2 value of 0.3006. We can observe that the head of the modeled amygdala grows larger as age increases (as one would expect).

V. DISCUSSION

From these experiments, we find that the proposed ShapeNet approach is able to predict the shape changes with a higher R^2 value. One reason is that the k blocks provide enough modeling capacity to encode the shapes at each age, and to generalize to unseen ages. The pentagon shapes predicted by geodesic regression are not correct since the shapes will not merely shrink with the progression of age; they also grow bigger as shown in the shapes from ShapeNet. In addition, input data for the geodesic regression model needs to be normalized to comply with the geometrical properties of the manifold. The normalization of data can cause information

loss, but we do not need to normalize the data in our model. The R^2 value in the mandible dataset is not as significant as the other two datasets, which is caused by the little variations of the original data. We also easily extend our model into 3D space since only the number of blocks k and final shape size are changed.

However, one weakness of ShapeNet is that it is sensitive to large changes in shape from one time unit to the next. There is an intrinsic assumption that landmark positions change slowly. Another is that it is difficult to explore population-based shape changes if each shape is substantially different from each other. In addition, ShapeNet is currently trained with population-based shapes, which only show the shape variations of the general population due to the absence of data for individuals over time. Finally, we note that the model can be easily extended to support more than a single input; natural extensions would include other influential factors such as gender, weight, etc.

VI. CONCLUSION

This work is the first to propose a regressive convolutional neural network for age-focused landmark shape prediction, which is applicable to 2D and 3D data. Experimental results demonstrate the effectiveness of the proposed ShapeNet to predict the corpus callosum and mandible shape, and amygdala with correct topology and accurate fitting that matches real world shapes. The proposed ShapeNet can predict shape variability of high dimensional and nonlinear data, which is often critical to understanding processes that change the shape of anatomy in the fields of biology and medicine.

There are several directions for follow-up work. First, to apply ShapeNet to the prediction of shapes from individuals (rather than the population in general). Second, to extend the current network to improve predicted shapes when there are significant changes in the data across few timesteps. Both would enable ShapeNet to have more impact on age-focused shape prediction and be beneficial to early diagnosis of neurodegenerative disease.

REFERENCES

- [1] SM Masudur Rahman Al Arif, Karen Knapp, and Greg Slabaugh. SPNet: Shape prediction using a fully convolutional neural network. In *MICCAI*, pages 430–439. Springer, 2018.
- [2] Yi Fang, Jin Xie, Guoxian Dai, Meng Wang, Fan Zhu, Tiantian Xu, and Edward Wong. 3D deep shape descriptor. In *CVPR*, pages 2319–2328, 2015.
- [3] Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.
- [4] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *4th Int'l Conf. on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.
- [5] Daniele Ravi, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, 2017.
- [6] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19:221–248, 2017.
- [7] Pawel Mlynarski, Hervé Delingette, Antonio Criminisi, and Nicholas Ayache. 3D convolutional neural networks for tumor segmentation using long-range 2D context. *Computerized Medical Imaging and Graphics*, 73:60–72, 2019.
- [8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.
- [9] Nyoman Abiwinanda, Muhammad Hanif, S Tafwida Hesaputra, Astri Handayani, and Tati Rajab Mengko. Brain tumor classification using convolutional neural network. In *World Congress on Medical Physics and Biomedical Engineering 2018*, pages 183–189. Springer, 2019.
- [10] Ehsan Hosseini Asl, Mohammed Ghazal, Ali Mahmoud, Ali Aslantas, Ahmed Shalaby, Manual Casanova, Gregory Barnes, Georgy Gimelfarb, Robert Keynton, and Ayman El Baz. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Frontiers In Bioscience, Landmark*, 23:584–596, January 2018.
- [11] PJ Sudharshan, Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte, and Paul Honeine. Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, 117:103–111, 2019.
- [12] Kun-Hsing Yu, Feiran Wang, Gerald J Berry, Christopher Re, Russ B Altman, Michael Snyder, and Isaac S Kohane. Classifying non-small cell lung cancer histopathology types and transcriptomic subtypes using convolutional neural networks. *bioRxiv*, page 530360, 2019.
- [13] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *MICCAI*, pages 424–432. Springer, 2016.
- [14] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for AD/MCI classification. In *MICCAI*, pages 583–590. Springer, 2013.
- [15] Jeremy Kawahara and Ghassan Hamarneh. Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In *International Workshop on Machine Learning in Medical Imaging*, pages 164–171. Springer, 2016.
- [16] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J Van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sánchez, and Bram van Ginneken. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169, 2016.
- [17] Xinting Gao, Stephen Lin, and Tien Yin Wong. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Transactions on Biomedical Engineering*, 62(11):2693–2701, 2015.
- [18] Christian Payer, Darko Štern, Horst Bischof, and Martin Urschler. Regressing heatmaps for multiple landmark localization using CNNs. In *MICCAI*, pages 230–238. Springer, 2016.
- [19] Yuanpu Xie, Fuyong Xing, Xiangfei Kong, Hai Su, and Lin Yang. Beyond classification: Structured regression for robust cell detection using convolutional neural network. In *MICCAI*, pages 358–365. Springer, 2015.
- [20] Thomas Fletcher. Geodesic regression on Riemannian manifolds. In *Proc. 3rd Int'l Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pages 75–86, 2011.
- [21] Jacob Hinkle, P Thomas Fletcher, and Sarang Joshi. Intrinsic polynomials for regression on Riemannian manifolds. *Journal of Mathematical Imaging and Vision*, 50(1-2):32–52, 2014.
- [22] Yi Hong, Roland Kwitt, Nikhil Singh, Brad Davis, Nuno Vasconcelos, and Marc Niethammer. Geodesic regression on the Grassmannian. In *ECCV*, pages 632–646. Springer, 2014.
- [23] Lloyd J Edwards, Keith E Muller, Russell D Wolfinger, Bahjat F Qaqish, and Oliver Schabenberger. An R^2 statistic for fixed effects in the linear mixed model. *Statistics in medicine*, 27(29):6137–6157, 2008.
- [24] George AF Seber and Alan J Lee. *Linear regression analysis*, volume 329. John Wiley & Sons, 2012.
- [25] Seongho Seo, Moo K Chung, and Hourii K Vorperian. Heat kernel smoothing using laplace-beltrami eigenfunctions. In *MICCAI*, pages 505–512. Springer, 2010.
- [26] Moo K Chung, Keith J Worsley, Brendon M Nacewicz, Kim M Dalton, and Richard J Davidson. General multivariate linear modeling of surface shapes using SurfStat. *Neuroimage*, 53(2):491–505, 2010.