Eurographics Symposium on Rendering
E. Garces and E. Haines
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 4

Supplemental material: Realistic Facial Age Transformation with 3D Uplifting

X. Li[1,2] G. C. Guarnera[2,3] A. Lin[1,2] A. Ghosh[1,2]

[1]Imperial College London, UK
[2]Lumirithmic Ltd.
[3]University of York, UK

This supplementary material contains more details on the neural networks and displays results with high resolution.

A. Neural Network Details

Here, we introduce the network architectures of *SkinNet*, *GeoNet*, and *RefineNet* in detail, as well as the training process of the whole pipeline.

• **SkinNet** consists of *PredNet* and *ReconNet*. Given an input image, *PredNet* estimates the concentrations of two chromophores (melanin and hemoglobin). The structure of *PredNet* is shown in Table 1. *PredNet* consists of 4 fully connected layers. The outputs of *PredNet* are fed into *ReconNet* to reconstruct re-aged skin appearances. The structure of *ReconNet* is shown in Table 2. Similarly, *ReconNet* consists of 4 fully connected layers. Each fully convolution layer, except the last layer in both networks, is followed by a ReLU [Aga18].

Table 1: *PredNet architecture*

Layer	Input Size	Output Size
Fully Connected	3	64
Fully Connected	64	64
Fully Connected	64	64
Fully Connected	64	2

Table 2: *ReconNet architecture*

Layer	Input Size	Output Size
Fully Connected	2	64
Fully Connected	64	64
Fully Connected	64	64
Fully Connected	64	3

• **GeoNet** has a similar structure as Decoder of [RBSB18]. As shown in Table 3, *GeoNet* consists of a fully connected layer to convert the size of the latent space, including shape (β), expression (ψ), pose (θ) and age (a) parameters, for further reconstruction. After the fully connected layer, *GeoNet* contains five blocks to reconstruct a 3D face shape. Each block includes an Up-Sampling layer and a Chebyshev Convolution layers [DBV16] with $K = 3$ Chebyshev polynomials, and an Exponential Linear Unit(ELU) [CUH16]. The Up-Sampling layer aims to increase the vertices number to around 4 times. The loss function of this part is shown in Equation 3 in the main paper, where $\lambda_v = 1.0$, $\lambda_{pho} = 0.2$, $\lambda_{age} = 0.05$, and $\lambda_{reg} = 0.1$.

Table 3: *GeoNet architecture*

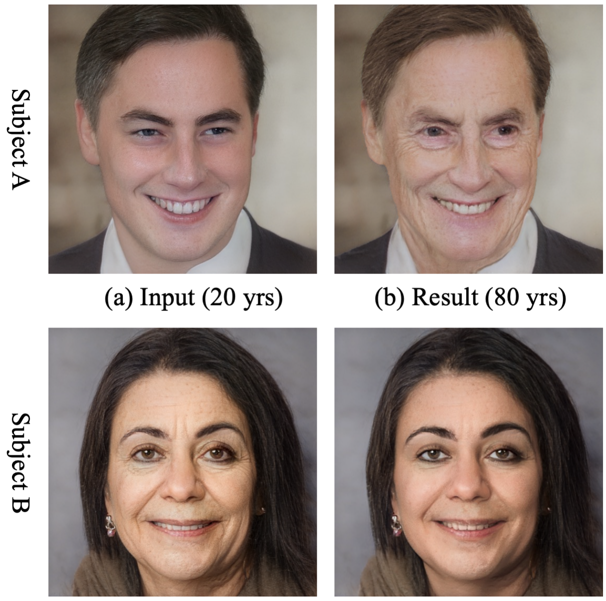
Layer	Input Size	Output Size
Fully Connected	160	5×128
Up-Sampling	5×128	20×128
Chebyshev Convolution	20×128	20×128
Up-Sampling	20×128	79×128
Chebyshev Convolution	79×128	79×64
Up-Sampling	79×64	314×64
Chebyshev Convolution	314×64	314×16
Up-Sampling	314×16	1256×16
Chebyshev Convolution	1256×16	1256×16
Up-Sampling	1256×16	5023×16
Chebyshev Convolution	5023×16	5023×3

• **RefineNet** is a light weight network to extract the high frequency information as displacement map. This network only contains two convolution layers. The first convolution layer is followed by a ReLU [Aga18]. The kernel, stride and padding of each convolution layer is 3, 1, and also 1, respectively. The loss function of the *RefineNet* is shown in Equation 5 in the main paper, where $\lambda_{pho} = 1.0$, $\lambda_{dx} = 0.1$, and $\lambda_{dy} = 0.1$.

The **Training process** of the whole pipeline is we first trained *SkinNet* separately on Lookup Tables generated by the skin model [LGLG24], and trained *AgeEditNet* separately on Facescape dataset and Lifespan datasets. Furthermore, we fixed the *SkinNet* and *AgeEditNet*, and trained *GeoNet* to obtain coarse shapes following by fixing these three and trained *RefineNet* as a whole pipeline.

B. Results with High Resolution

Here we show two examples with 512×512 resolution in Figure 1. We aged the face of Subject A from 20 years old (a) to 80 years old (b), and de-aged the face of Subject B from 60 years old (c) to 30 years old (d). As shown in this figure, results with higher resolution contain more skin details. The whole pipeline of our method can be adapted to a higher resolution if GPU resources are sufficient.



(a) Input (20 yrs) (b) Result (80 yrs)

(c) Input (60 yrs) (d) Result (30 yrs)

Figure 1: *Re-aged results of two subjects with 512 × 512 resolution.*

References

[Aga18] AGARAP A. F.: Deep learning using rectified linear units (relu). *ArXiv abs/1803.08375* (2018). 1

[CUH16] CLEVERT D., UNTERTHINER T., HOCHREITER S.: Fast and accurate deep network learning by exponential linear units (elus). In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings* (2016), Bengio Y., LeCun Y., (Eds.). 1

[DBV16] DEFFERRARD M., BRESSON X., VANDERGHEYNST P.: Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2016), NIPS'16, Curran Associates Inc., p. 3844–3852. 1

[LGLG24] LI X., GUARNERA G., LIN A., GHOSH A.: Practical measurement and neural encoding of hyperspectral skin reflectance. In *2024 International Conference on 3D Vision (3DV)* (Los Alamitos, CA, USA, mar 2024), IEEE Computer Society, pp. 1301–1309. doi:10.1109/3DV62453.2024.00116. 1

[RBSB18] RANJAN A., BOLKART T., SANYAL S., BLACK M. J.: Generating 3d faces using convolutional mesh autoencoders. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part III* (2018), Springer-Verlag, p. 725–741. doi:10.1007/978-3-030-01219-9_43. 1