

A Comprehensive Framework for Face age Synthesis, Voice Transformation, and Cartoon Generation

Naveen Kumar B, Ganesh B, Karthik Kj, Mohammed Tousif, S Deepa

Department of Computer Science and Engineering

Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

Abstract: *This paper presents a comprehensive framework addressing age-invariant face recognition (AIFR), face age synthesis (FAS), voice transformation, and avatar generation. Traditional AIFR techniques focus on minimizing age-related variations in face recognition but lack visual interpretability, while FAS methods aim to synthesize faces of different ages but often compromise recognition due to artifacts. To address these limitations, we propose MTLFace, a multi-task learning framework that simultaneously handles AIFR and FAS tasks. MTLFace employs attention-based feature decomposition to separate identity and age-related features spatially, improving interpretability. Additionally, we introduce an identity conditional module for fine-grained face age synthesis, enhancing the naturalness of synthesized faces. Unlike conventional methods that achieve age group-level synthesis, our identity conditional module enables identity-level synthesis, resulting in smoother age transitions and preserving individual facial characteristics. Leveraging high-quality synthesized faces, we enhance AIFR performance via selective fine-tuning, where synthesized faces are used to augment training data, leading to improved robustness against age variations. Furthermore, we contribute a large cross-age face dataset with annotations, facilitating research in age-related tasks. In addition to AIFR and FAS, we explore limitations in existing methodologies for voice transformation and propose advancements to improve speech quality and address current challenges. Lastly, we introduce a novel approach for cartoon face generation, utilizing component-based facial feature extraction and template matching to produce diverse and stylistic cartoon faces. Experimental results across various benchmarks demonstrate the effectiveness of our approaches in achieving superior performance in face recognition, voice transformation, and cartoon face generation, thereby contributing to the advancement of multimodal synthesis technologies with potential applications in entertainment, virtual communication, and assistive technologies.*

Keywords: face age synthesis

REFERENCES

- [1] Z. Huang, J. Zhang, and H. Shan, "When age-invariant face recognition meets face age synthesis: A multi-task learning framework," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 7282–7291.
- [2] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning continuous face age progression: A pyramid of GANs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 2, pp. 499–515, Feb. 2021.
- [3] N. Iwahashi and Y. Sagisaka. Speech spectrum transformation based on speaker interpolation. In Proc. ICASSP94, 1994.
- [4] A. Kain. High resolution voice transformation. PhD thesis, OGI School of Science and Eng., Portland, Oregon, USA.
- [5] Zhang, Y.; Dong, W.; Deussen, O.; Huang, F.; Li, K.; Hu, B.-G. "Data-Driven Face Cartoon Stylization." In SIGGRAPH Asia Technical Briefs; ACM: New York, NY, USA, 2014; pp. 14:1–14:4.
- [6] Gao, W.; Mo, R.; Wei, L.; Zhu, Y.; Peng, Z.; Zhang, Y. "Template-Based Portrait Caricature Generation with Facial Components Analysis." In Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2009, Shanghai, China, 20–22 November 2009; pp. 219–223.

