SegAnimeChara: Segmenting Anime Characters Generated by AI

Andy Yu-Hsiang Tseng*
National Taiwan University
Taipei, Taiwan, R.O.C
txs@cmlab.csie.ntu.edu.tw

Wen-Fan Wang
National Taiwan University
Taipei, Taiwan, R.O.C
b02606001@ntu.edu.tw

Bing-Yu Chen
National Taiwan University
Taipei, Taiwan, R.O.C
robin@ntu.edu.tw

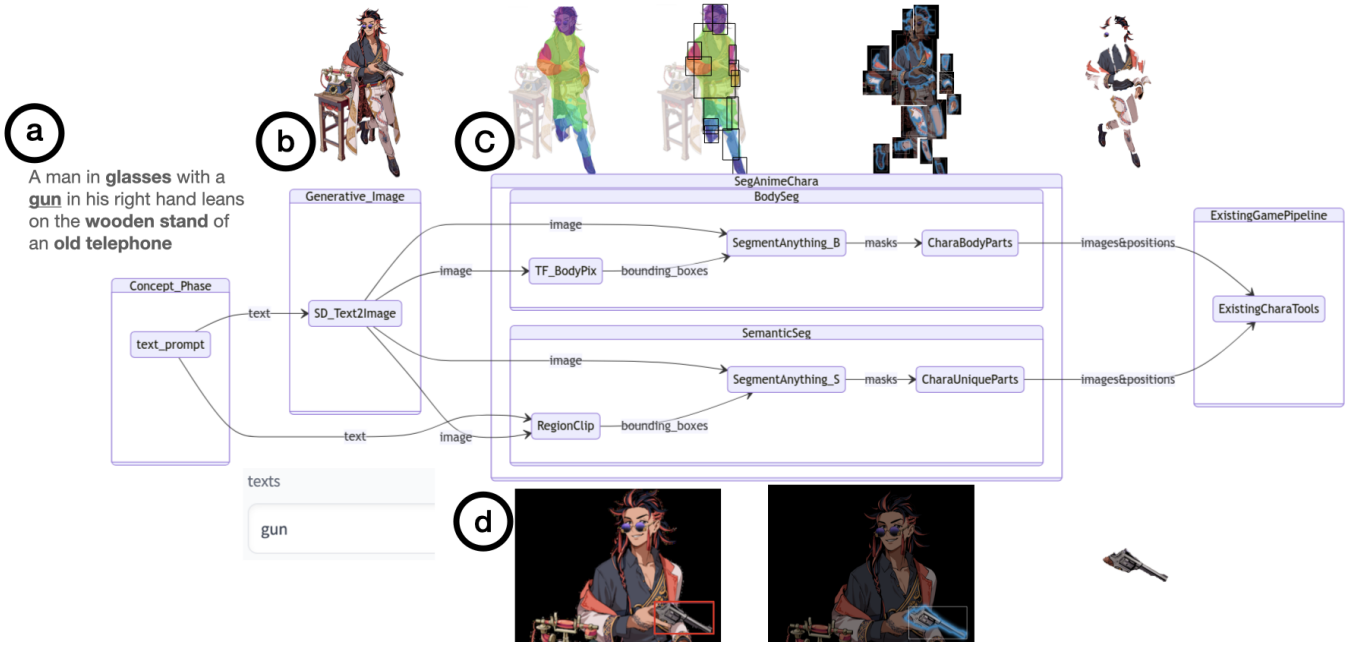

Figure 1: Segmentation Pipeline; (a) Game artists take an idea described in words (b) Text2Image generates images from text prompts (c) Body Segmentation splits images into body parts (d) Semantic Segmentation is generated from text prompts

ABSTRACT

This work introduces SegAnimeChara, a novel system of transforming AI-generated anime images into game characters while retaining unique features. Using volume-based body pose segmentation, SegAnimeChara can efficiently, zero-shot, segment body parts from generative images based on OpenPose human skeleton. Furthermore, this system integrates a semantic segmentation pipeline based on the text prompts of the existing Text2Image workflow. The system conserves the game character's unique outfit and reduces the redundant duplicate text prompts for semantic segmentation.

CCS CONCEPTS

• **Computing methodologies** → *Computer graphics*; *Computer vision.*

KEYWORDS

Game, Design, Anime, Manga, Otaku, Character Segmentation, Semantic Segmentation, Pose Segmentation, Body Segmentation

ACM Reference Format:
Andy Yu-Hsiang Tseng, Wen-Fan Wang, and Bing-Yu Chen. 2023. SegAnimeChara: Segmenting Anime Characters Generated by AI. In *Proceedings of ACM Conference (SIGGRAPH '23 Posters).* ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

SIGGRAPH '23 Posters, August 06–10, 2023, Los Angeles, CA, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

AI generative images, e.g., Stable Diffusion and Midjourney, alter the game industry's workflow for creating concept posters and styling scenes. However, converting these images into 2D animation-ready characters remains challenging and time-consuming. In the conventional design pipeline, designers typically spend a day separating and reconstructing anime body parts. Using natural human recognition methods to segment anime characters can also result in undesired outcomes that do not integrate well with existing game design pipelines. For anime and cartoon characters, Takayama [Takayama et al. 2012] first proposed a feature-based zero-shot extraction method, not requiring sample images. However, this inspiring method only addressed face and hair recognition by color and

contour extractions. Adapting generative networks, Transfer Learning for Pose Estimation of Illustrated Characters[Chen and Zwicker 2021] bridge the domain between human and anime on pose estimation based on general OpenPose[Cao et al. 2019] in COCO format. Nevertheless, special anime features like cat ears/tails cannot be recognized. PAniC-3D[Chen et al. 2023] introduced a system for 2D anime characters' 3d reconstruction, but the semantic structures are limited by the general models, like eyes/mouths, instead of the artists' intention. Powerful general segmentation model such as Segment Anything Model (SAM)[Kirillov et al. 2023] introduces a novel of zero-shot automatic mask generation; however, it generates masks without considering body structure and semantic context. Moreover, its annotation process can be challenging for designers working on large-scale projects.

2 OUR APPROACH

We present SegAnimeChara, a state-of-the-art tool for segmenting cartoon/anime characters using general body structures and semantic prompts. Our zero-shot system can seamlessly integrate with existing game creation workflows and adapt to various 2D characters, including AI-generated ones. Solving common body structures and customized stylish features, we have divided the character segmentation process into *Body Segmentation* and *Semantic Segmentation*. This approach enables us to segment game characters with various clothed accessories accurately.

2.1 Body Segmentation

The majority of 2D anime/cartoon characters are based on human body structures, and we utilize a volume-based BodyPix model[Bailey et al. 2019] from Tensorflow.js to segment game characters into body parts based on general OpenPose models. Using the BodyPix masks as bounding boxes, we apply the Segment Anything Model (SAM)[Kirillov et al. 2023] to generate body segmentation from AI-generated images, which meet the standards of the game industry. However, certain special characteristic features, such as hairstyles, tails, and bow ties, are not covered by BodyPix. We propose a novel semantic segmentation method in the following section to address this.

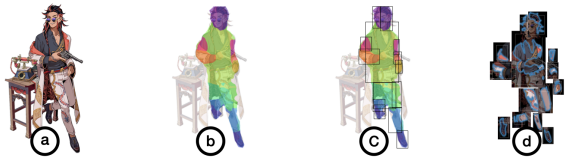

Figure 2: (a) Gets the image from Text2Image (b) BodyPix body parts recognition (c) Change masks into bounding boxes (d) Turing bounding boxes to masks and images by SAM

2.2 Semantic Segmentation

Anime/cartoon characters often have unique dressing styles and special accessories, e.g., hairstyles, glasses, and guns, which enhance the overall design. However, these features are often neglected and merely recognized when using existing general human models, such as OpenPose[Cao et al. 2019]. To address this issue, we propose a semantic segmentation method that shares the same text prompts as those we used to generate concept images using Text2Image[Reed et al. 2016] method. In Figure 3, our pipeline sends the semantic prompt to RegionClip[Zhong et al. 2021], where the bounding box is segmented, using SAM, into an image and position that can be incorporated into the existing workflow. This approach allows us to accurately capture and segment special anime features.

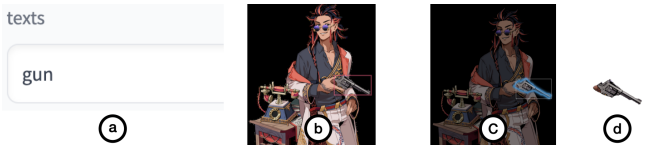

Figure 3: (a) Keyword from text prompt, e.g., gun. (b) Use RegionClip model to recognize the bounding box. (c) Turn bounding box to mask by SAM. (d) Crop out the mask.

3 CONCLUSION

We introduce SegAnimeChara, a novel system that integrates multiple state-of-art generative, body pose estimating, semantic, and segmentation models to generate unique anime characters based on generative images automatically. Our system significantly reduces the time-consuming tasks associated with cropping body parts for the real-time 2D animation character. There is, however, room for improvements, such as predicting the shape of covered parts and auto-binding components to unique character body frames. Our proposed system has the potential to shorten the existing feature-rich game character creation pipeline and inspire designers to generate stylish avatars using generative AI models efficiently.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Science and Technology of Taiwan, National Taiwan University for research, and Rayark Inc for game creation pipeline and character images.

REFERENCES

Paige Bailey, Sofien Bouaziz, Shan Carter, Josh Gordon, Christian Häne, Alexander Mordvintsev, Julien Valentin, and Martin Wicke. 2019. Differentiable Graphics with TensorFlow 2.0. In *ACM SIGGRAPH 2019 Courses* (Los Angeles, California) *(SIGGRAPH '19)*. Association for Computing Machinery, New York, NY, USA, Article 10, 211 pages. https://doi.org/10.1145/3305366.3328041

Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

Shuhong Chen, Kevin Zhang, Yichun Shi, Heng Wang, Yiheng Zhu, Guoxian Song, Sizhe An, Janus Kristjansson, Xiao Yang, and Matthias Zwicker. 2023. PAniC-3D: Stylized Single-view 3D Reconstruction from Portraits of Anime Characters. arXiv:2303.14587 [cs.CV]

Shuhong Chen and Matthias Zwicker. 2021. Transfer Learning for Pose Estimation of Illustrated Characters. arXiv:2108.01819 [cs.CV]

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]

Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative Adversarial Text to Image Synthesis. arXiv:1605.05396 [cs.NE]

Kohei Takayama, Henry Johan, and Tomoyuki Nishita. 2012. FACE DETECTION AND FACE RECOGNITION OF CARTOON CHARACTERS USING FEATURE EXTRACTION.

Yiwu Zhong, Jianwei Yang, Pengchuan Zhang, Chunyuan Li, Noel Codella, Liunian Harold Li, Luowei Zhou, Xiyang Dai, Lu Yuan, Yin Li, and Jianfeng Gao. 2021. RegionCLIP: Region-based Language-Image Pretraining. arXiv:2112.09106 [cs.CV]