Towards Fair and Robust Face Parsing for Generative AI: A Multi-Objective Approach

Sophia J. Abraham[1] Jonathan D. Hauenstein[2] Walter J. Scheirer[1]
[1]Department of Computer Science and Engineering
[2]Department of Applied and Computational Mathematics and Statistics
University of Notre Dame, Notre Dame, IN 46556
sabraham@nd.edu, jhauenstein@nd.edu, wscheirer@nd.edu

Abstract

Face parsing underpins tasks like identity verification, facial editing, and image synthesis, but current models often yield biased segmentations and fail under noise, occlusion, or domain shifts hurting downstream synthesis. We introduce a homotopy-based multi-objective U-Net that dynamically balances accuracy, fairness, and robustness during training. Plugged into GAN-based face synthesis pipelines (Pix2PixHD) and structured conditioning model for diffusion-based synthesis (ControlNet) pipelines, our approach boosts photorealism, demographic consistency, and resilience to perturbations, producing higher-quality generative outputs.

1. Introduction

Face parsing, the segmentation of fine-grained components like *eyes, nose, mouth, and hair*, is a fundamental task in computer vision, supporting *face recognition* [18], *augmented reality* [11], and *facial expression analysis* [3]. While segmentation accuracy has improved [2, 12], existing models often neglect key concerns: (1) *fairness* across demographic groups, (2) *robustness* to noise, occlusion, and domain shifts, and (3) the *impact* on downstream generative models. Face parsers may perform well on clean data but degrade for underrepresented demographics [1, 7, 14] or in real-world conditions [5, 8], propagating biases into *generative synthesis* tasks.

Prior work explored *multi-objective optimization* [9, 16] and *fairness-aware facial analysis* [13], but no unified strategy addresses *accuracy*, *fairness*, and *robustness* together. Moreover, both **GAN-based** [6] and **diffusion-based** [19] generative models heavily depend on segmentation quality [15], making fairness in parsing crucial to avoid propagating biases [4, 17].

We propose a **homotopy-based multi-objective learning framework** that dynamically balances *accuracy*, *fairness*, and *robustness* during training. Our method improves segmentation across demographic groups and enhances resilience to *noise*, *occlusion*, and *domain shifts*, benefiting downstream *GAN-* and *diffusion-based face synthesis*.

To validate our approach, we integrate multi-objective and single-objective U-Net models into a GAN-based face synthesis pipeline (Pix2PixHD) and conduct preliminary experiments with ControlNet. We evaluate segmentation quality (mIoU, fairness variance) and generative quality (FID, LPIPS) under real-world perturbations. Our key contributions include: proposing a multi-objective framework for fairness-aware and robust face parsing; systematically evaluating fairness via mIoU variance and robustness under perturbations; demonstrating that improved segmentation fairness leads to better GAN-based face synthesis with lower FID and enhanced photorealism; and conducting a preliminary analysis of segmentation quality effects on diffusion-based synthesis with ControlNet. These findings highlight the importance of fair and robust face parsing for bias-aware generative AI.

2. Proposed Method

We propose a homotopy-based multi-objective framework for face parsing that optimizes **accuracy**, **fairness**, and **robustness**. Given images x_i and masks y_i annotated with demographic attributes a, we train a segmentation model $f_\theta(x_i)$ minimizing:

$$\mathcal{L}_{\text{total}} = \alpha(t)\mathcal{L}_{\text{acc}} + \beta(t)\mathcal{L}_{\text{rob}} + \gamma(t)\mathcal{L}_{\text{fair}},$$

where $\mathcal{L}_{\text{acc}}$ is the generalized Dice loss over all 19 classes (measuring overlap between prediction and ground truth), $\mathcal{L}_{\text{rob}}$ penalizes the average drop in mIoU under a predefined set of input perturbations via $\max\big(0,\ \text{mIoU}_{\text{clean}} - \text{mIoU}_{\text{perturbed}}\big)$, and $\mathcal{L}_{\text{fair}}$ is a demographic fairness loss, computed either as the variance or the range (max–min) of per-group mIoU across attributes to promote equitable

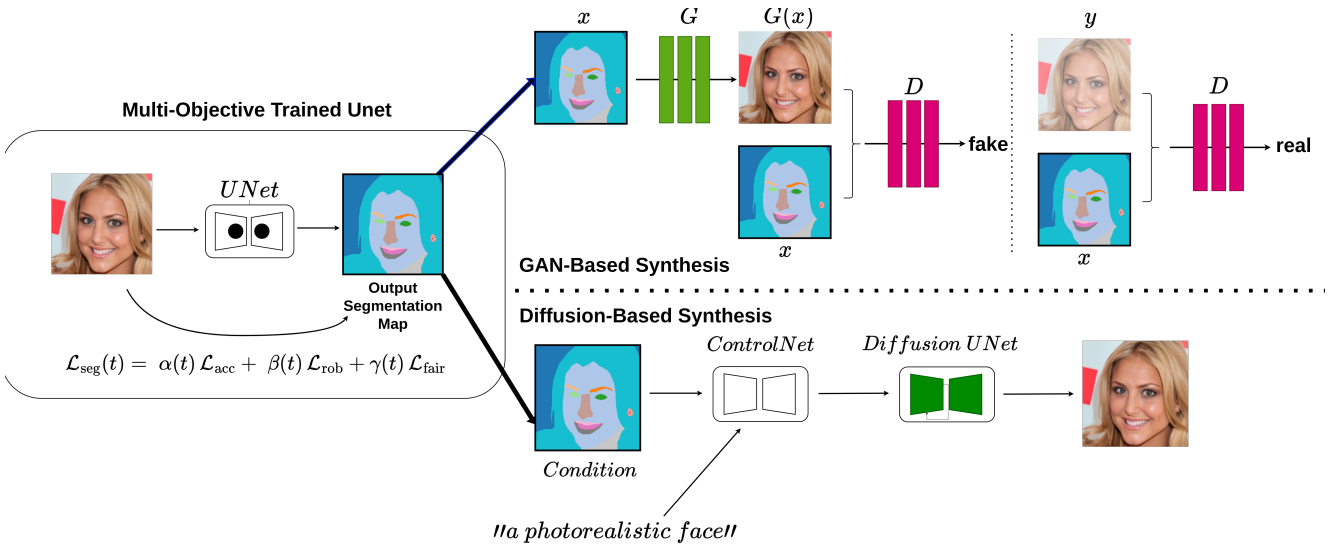

Figure 1. **Overview of Our Multi-Objective Face Parsing and Synthesis Framework.** Our proposed *homotopy-based multi-objective learning framework* optimizes **accuracy** (L_{acc}), **robustness** (L_{rob}), and **fairness** (L_{fair}). This framework produces **fairness-aware and robust segmentation maps**, which are used to train two generative pipelines: (1) a **GAN-based synthesis model (Pix2PixHD)**, where improved segmentation enhances *photorealism and demographic consistency*, and (2) a **diffusion-based synthesis model (ControlNet)**, where structured parsing maps guide *semantic alignment and editability*. The improved segmentation quality enhances photorealism, fairness, and robustness in generative models. Key improvements include reduced bias in GAN-generated faces and more stable semantic conditioning in diffusion synthesis.

segmentation quality. Homotopy scheduling dynamically adjusts $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ during training using Linear, Sigmoid, or Piecewise strategies, initially emphasizing accuracy before shifting towards fairness and robustness.

To assess downstream effects, we integrate the trained U-Nets into a GAN-based synthesis pipeline (Pix2PixHD) and a diffusion-based synthesis pipeline (ControlNet), using segmentation maps as structured conditioning. We train on CelebAMask-HQ [10], resizing images to 256×256. Segmentation performance is evaluated using mean Intersection-over-Union (mIoU) and fairness variance; synthesis outputs are assessed using Fréchet Inception Distance (FID) and LPIPS similarity. All models are implemented in PyTorch and trained with Adam optimizer on NVIDIA A10 GPUs; ControlNet is fine-tuned for one epoch.

3. Results & Discussion

We comprehensively evaluate our segmentation models across fairness, robustness, and generative quality. Our results demonstrate that fairness- and robustness-aware training improves face parsing and has direct positive downstream effects on face generation tasks.

Multi-objective U-Nets achieve comparable or slightly improved segmentation performance, maintaining mIoU parity despite optimizing for additional objectives (Table 1). Notably, the multi-objective models achieve slightly higher Dice scores, indicating that the addition of fairness and robustness constraints does not significantly compromise pixel-level segmentation fidelity.

Table 1. **Comparison of Segmentation Objectives on U-Net.** Quantitative results comparing single-objective and multi-objective models.

Objective	mIoU (%)	Dice (%)
Single Objective	73.87	94.46
Multi-Objective (Linear)	**74.21**	94.28
Multi-Objective (Sigmoid)	73.50	94.35
Multi-Objective (Piecewise)	73.80	94.47
Multi-Objective (Alt. Fairness)	73.81	**94.49**

Robustness analysis shows that multi-objective models yield lower mIoU degradation under perturbations such as Gaussian noise, blur, and occlusion (Figure 2), and generate more stable segmentations under noise (Figure 3). These results highlight the advantage of explicitly optimizing for robustness during training.

Quantitative robustness results under perturbations are summarized in Table 2. We observe that linear and piecewise homotopy models achieve notably lower FID and LPIPS scores compared to single-objective training, especially under blur and lighting shifts, indicating better generalization to noisy conditions.

In terms of fairness, class-wise mIoU comparisons (Ta-

Table 2. **Robustness of U-Net Variants under Perturbations.** Lower FID and LPIPS indicate more robust synthesis under noise, blur, brightness, and darkness.

Model	Gaussian Noise		Blur		Brightness		Darkness		Notes
	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	
Single-Objective	363.06	0.435	259.12	0.403	319.57	0.407	367.75	0.431	Baseline
Multi-Objective (Linear)	322.23	0.434	236.44	0.386	313.02	0.433	285.24	0.425	Linear Homotopy
Multi-Objective (Piecewise)	307.16	0.435	216.98	0.384	330.10	0.439	326.82	0.430	Piecewise Homotopy
Multi-Objective (Sigmoid)	349.30	0.437	208.30	0.412	286.38	0.444	331.36	0.456	Sigmoid Homotopy

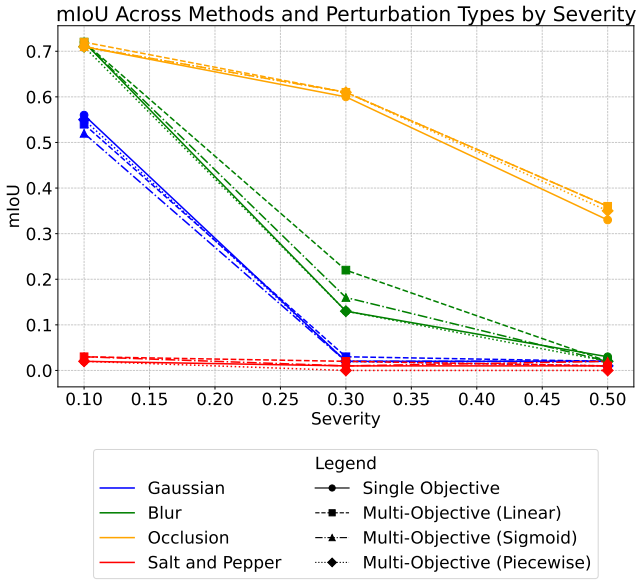

Figure 2. **mIoU under Perturbations.** Multi-objective models better withstand noise, blur, and occlusion.

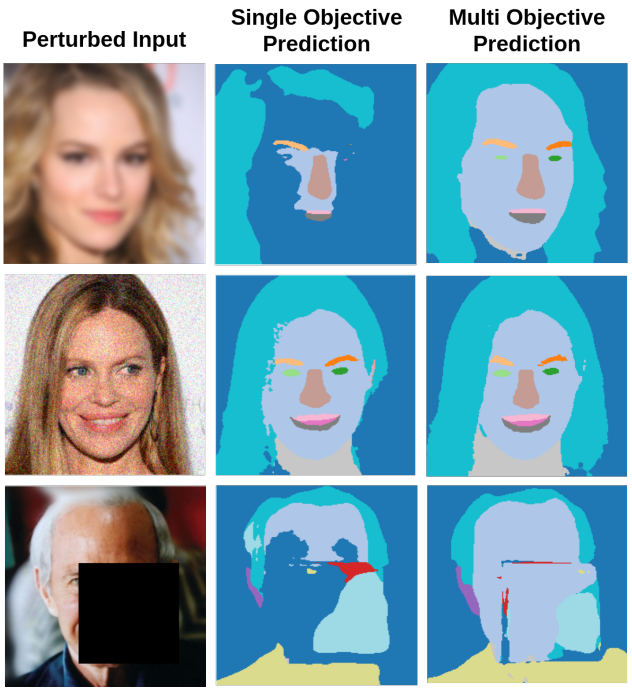

Figure 3. **Qualitative Comparison under Perturbations.** Multi-objective models preserve facial structure better than single-objective baselines.

ble 3) show that multi-objective models consistently outperform single-objective baselines across all facial classes, even for small or visually subtle components. Figure 4 illustrates that multi-objective training leads to more equitable segmentation across demographic attributes, reducing fairness gaps.

In downstream GAN-based face synthesis (Pix2PixHD), segmentation maps from multi-objective models lead to visibly cleaner and more realistic faces (Figure 5). Quantitatively, Table 4 shows that multi-objective training lowers FID and LPIPS compared to single-objective baselines, demonstrating that improved segmentation fairness and robustness translate to better generative outputs.

Finally, in diffusion-based synthesis using ControlNet, segmentation maps from multi-objective models enable cleaner, sharper face generations (Figure 6), further validating the benefits of fairness- and robustness-aware parsing in structured generative pipelines.

4. Limitations and Future Directions

Despite notable improvements in fairness, robustness, and segmentation quality, several challenges remain, presenting opportunities for further research. First, the CelebAMask-HQ dataset, while diverse, remains imbalanced across demographic groups, which may limit generalization. Addressing this requires more strategic data augmentation, active reweighting, or leveraging larger, demographically-balanced datasets to further mitigate bias and enhance equitable performance. Second, our current framework treats GANs as passive consumers of segmentation maps. Incorporating *bi-directional optimization*, where segmentation feedback influences GAN training, could improve both parsing fidelity and generative realism. Such an approach could be extended to diffusion models, where structured

Table 3. **Class-wise Mean mIoU Comparison.** Multi-objective models achieve higher per-class accuracy.

Model	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Single Objective	73.87	73.87	73.87	73.87	73.87	73.87	75.02	73.89	73.88	73.71	73.87	73.87	73.87	73.87	73.87	73.21	73.89	73.87	75.17
Multi-Objective (Alt. Fairness)	**74.21**	**74.21**	**74.21**	**74.21**	**74.21**	**74.21**	**75.06**	**74.23**	**74.22**	**74.09**	**74.21**	**74.18**	**74.21**	**74.21**	**74.21**	**73.49**	**74.23**	**74.21**	**75.24**

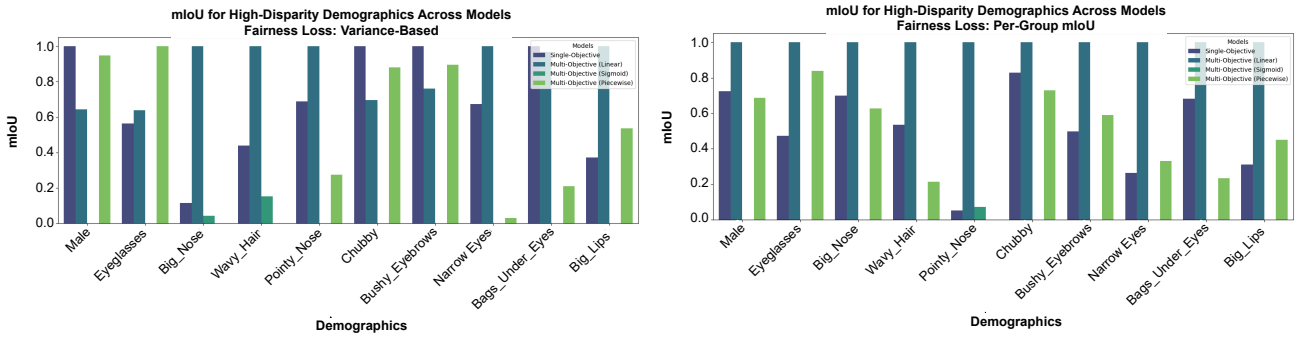

Figure 4. **Fairness Loss Strategies Comparison.** Multi-objective training reduces demographic disparities in segmentation performance.

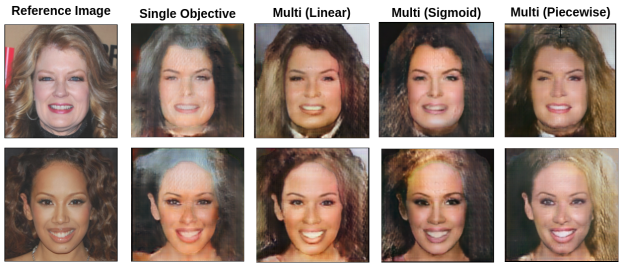

Figure 5. **Impact of Segmentation on GAN Synthesis.** Multi-objective maps yield more realistic and coherent faces.

Table 4. **Face Synthesis Results: GAN and Diffusion.** Lower FID and LPIPS indicate higher realism.

GAN-Based Face Synthesis		
Segmentation Source	**FID ↓**	**LPIPS ↓**
Single-Objective U-Net	117.93	0.4419
Multi-Objective (Linear)	99.93	0.4269
Multi-Objective (Piecewise)	**98.87**	**0.4222**
Diffusion-Based Face Synthesis (ControlNet)		
Single-Objective U-Net	261.01	0.7867
Multi-Objective U-Net (Linear)	**257.18**	**0.7848**

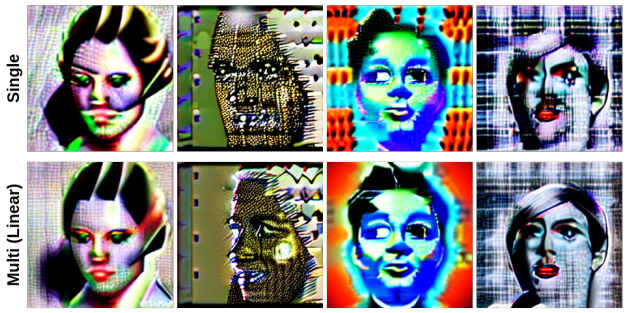

Figure 6. **Impact on Diffusion-Based Synthesis.** Multi-objective segmentation improves ControlNet face generation.

conditioning remains underexplored in fairness-aware synthesis. Additionally, while our method is broadly applicable beyond facial segmentation, extending it to domains such as medical imaging, autonomous perception, or video-based synthesis may require task-specific adaptations. Future research should explore domain-aware multi-objective formulations that account for context-specific biases and robustness challenges. Finally, while homotopy scheduling improves optimization efficiency, fairness-aware training introduces additional computational overhead due to subgroup evaluations. Exploring adaptive sampling strategies or efficient approximations could make large-scale deployments more feasible, especially for real-time applications.

Our findings underscore that multi-objective training does not impose rigid trade-offs, adaptive optimization can integrate fairness and robustness without sacrificing accuracy. By extending these ideas to broader datasets, generative frameworks, and real-world applications, future research can drive the development of more equitable and resilient vision models for AI-driven image synthesis and recognition.

ACKNOWLEDGMENTS

This work was funded by the DEVCOM Army Research Laboratory under the cooperative agreement W911NF-20-2-0218.

References

[1] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018. 1

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1

[3] Ciprian Adrian Corneanu, Marc Oliu Simón, Jeffrey F Cohn, and Sergio Escalera Guerrero. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE transactions on pattern analysis and machine intelligence*, 38(8):1548–1568, 2016. 1

[4] Felix Friedrich, Manuel Brack, Lukas Struppek, Dominik Hintersdorf, Patrick Schramowski, Sasha Luccioni, and Kristian Kersting. Fair diffusion: Instructing text-to-image generation models on fairness. *arXiv preprint arXiv:2302.10893*, 2023. 1

[5] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018. 1

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 1

[7] Patrick Grother, Mei Ngan, and Kayee Hanaoka. *Face recognition vendor test (fvrt): Part 3, demographic effects*. National Institute of Standards and Technology Gaithersburg, MD, 2019. 1

[8] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 1

[9] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018. 1

[10] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[11] Hee-Man Lee. Implementing augmented reality by using face detection, recognition and motion tracking. *Journal of the Korea society of computer and information*, 17(1):97–104, 2012. 1

[12] Jiangke Lin, Yi Yuan, Tianjia Shao, and Kun Zhou. Towards high-fidelity 3d face reconstruction from in-the-wild images using graph convolutional networks. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 5891–5900, 2020. 1

[13] Michele Merler, Nalini Ratha, Rogerio S Feris, and John R Smith. Diversity in faces. *arXiv preprint arXiv:1901.10436*, 2019. 1

[14] Sungho Park, Jewook Lee, Pilhyeon Lee, Sunhee Hwang, Dohyung Kim, and Hyeran Byun. Fair contrastive learning for facial attribute classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10389–10398, 2022. 1

[15] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2337–2346, 2019. 1

[16] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, 31, 2018. 1

[17] Shuhan Tan, Yujun Shen, and Bolei Zhou. Improving the fairness of deep generative models without retraining. *arXiv preprint arXiv:2012.04842*, 2020. 1

[18] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021. 1

[19] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 1