AdaptDiff: Adaptive Guidance in Diffusion Models for Diverse and Identity-Consistent Face Synthesis (Student Abstract)

Eduarda Caldeira[1], Tahar Chettaoui[1], Naser Damer[1, 2], Fadi Boutros[1]

[1]Fraunhofer IGD
[2]Department of Computer Science, TU Darmstadt

Abstract

Diffusion models conditioned on identity embeddings enable the generation of synthetic face images that consistently preserve identity across multiple samples. Recent work has shown that introducing an additional negative condition through classifier-free guidance during sampling provides a mechanism to suppress undesired attributes, thus improving inter-class separability. Building on this insight, we propose a dynamic weighting scheme for the negative condition that adapts throughout the sampling trajectory. This strategy leverages the complementary strengths of positive and negative conditions at different stages of generation, leading to more diverse yet identity-consistent synthetic data.

Code — https://github.com/EduardaCaldeira/NegFaceDiff/

Introduction

The use of synthetic data generated by diffusion models (DMs) to train face recognition (FR) has significantly evolved in recent years due to the need to address privacy and data protection issues associated with the use of privacy-sensitive authentic datasets. To enable FR training with state-of-the-art (SOTA) losses (Wang et al. 2018), i.e., margin-penalty softmax losses, a training dataset with class (identity) labels is required. Recent works leverage identity-conditional DMs to generate such labeled datasets. Inspired by the success of negative prompting in text-to-image generation, recent work has shown that including an extra negative condition to guide the diffusion process towards the exclusion of undesired characteristics improves FR performance while increasing intra-class variability and inter-identity variation (Caldeira, Damer, and Boutros 2025). In this work, we prove that the benefits of sampling with negative conditions are hindered by the fixed nature of their magnitudes, as strong negative conditions restrain the DM's freedom to explore the sampling space. We propose to adapt the negative condition strength during the sampling process, enabling higher freedom in the early stages, when little to no identity information is encoded and the latent space should be freely explored, and taking advantage of the benefits of negative prompting at later sampling stages, which can operate in a more restricted space.

Method	w & w_{max}	LFW	AgeDB	CAFLW	CFPFP	CPLFW	Avg
IDiff-Face	0	96.30	78.15	86.23	81.51	77.57	83.95
NegFaceDiff	0.5	**96.78**	**81.35**	**87.67**	**81.89**	**78.43**	**85.22**
	1.0	92.82	73.98	80.22	74.81	72.10	78.79
AdaptDiff	0.5	96.88	79.33	**82.70**	87.40	78.53	84.97
	1.0	**97.00**	**81.57**	82.29	**88.17**	**79.08**	**85.62**

Table 1: Accuracies (in %) of FR models trained on NegFaceDiff and our AdaptDiff with different w & w_{max}.

Methodology

DMs are trained to learn how to approximate an input (reverse process) by decoding noise extracted from it (forward process). A trained DM can generate a synthetic image by denoising a random noise seed. To generate several facial images of the same identity (class label), DMs use an identity condition p^+. A set of samples belonging to the same identity can be sampled by fixing p^+ and varying the initial random noise seed (Boutros et al. 2023a). An extra negative identity condition p^- can also be used during sampling (Caldeira, Damer, and Boutros 2025), resulting in additional guidance toward the exclusion of undesired characteristics in generated data, and higher separation between the identities, improving FR performance. This is achieved by incorporating the effects of the predicted noise added at time step t, $\epsilon_\theta(\mathbf{x}_t, t, p)$, $p \in \{p^+, p^-\}$, for both p^+ and p^-: $\hat{\epsilon}_\theta(x_t, t, p^+, p^-) = (1 + w)\,\epsilon_\theta(x_t, t, p^+) - w\,\epsilon_\theta(x_t, t, p^-)$, where w is the guidance strength of classifier's free guidance (CFG) negative condition (Ban et al. 2024; Wang et al. 2024). The subtraction ensures that generated images follow p^+ and deviate from p^-.

In this work, we dynamically adapt w during sampling, instead of fixing its value, due to the varying role of identity information across time steps. At early sampling ($t \to T$), x_t contains little to no identity information. Thus, the DM should explore the identity space with a high degree of freedom and weighting p^- with strong magnitude will unnecessarily hinder intra-class variations (Sadat et al. 2024). At later steps ($t \to 0$), x_t incorporates relevant identity features, and weighting p^- with a strong magnitude further pushes the generated samples towards p^+, ensuring high inter-class separability. Hence, progressively increasing w with a linear schedule ($w = w_{max} \times (1 - \frac{t}{T})$, as shown in Figure 2) facilitates earlier steps while still taking advantage of the improvements brought by negative conditions.

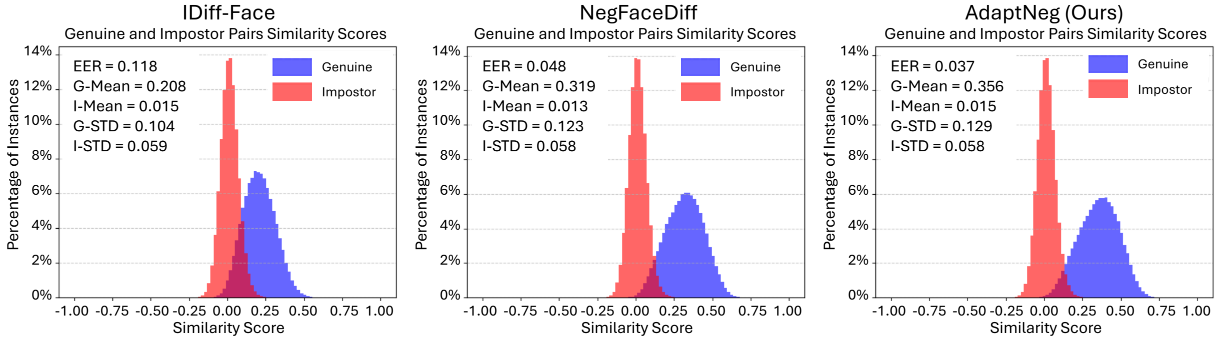

Figure 1: Histograms of genuine and impostor score distributions of SOTA baselines and AdaptDiff trained on FFHQ.

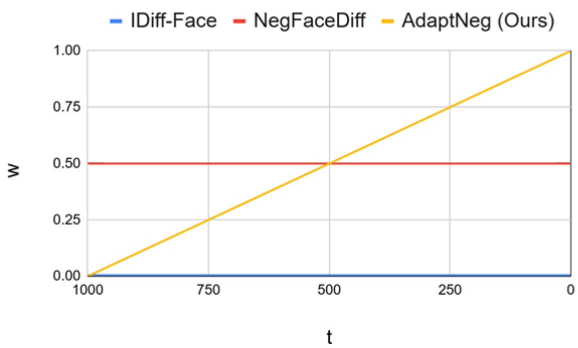

Figure 2: w as a function of the sampling time step, t. **Blue:** IDiff-Face fixes $w = 0$. **Red:** NegFaceDiff fixes $w = 0.5$. **Yellow:** AdaptDiff adapts w during sampling ($w = 1 - t/T$).

Method	LFW	AgeDB	CFPFP	CALFW	CPLFW	Avg	IJB-C 10^{-5}	IJB-C 10^{-4}
C-WF	99.55	94.55	95.31	93.78	89.95	94.63	93.96	96.05
Arc2Face	98.81	90.18	91.87	92.63	85.16	91.73	-	-
IDiff-Face*	96.48	77.27	81.71	86.95	78.55	84.19	16.85	37.58
NegFaceDiff*	96.80	80.45	82.34	87.32	78.62	85.11	**23.88**	49.13
AdaptDiff*	**96.97**	**81.08**	**83.39**	**87.58**	**79.77**	**85.76**	23.77	**55.81**
ID^3	97.28	83.78	85.00	89.30	77.13	86.50	-	-
IDiff-Face	**98.00**	86.43	**85.47**	**90.65**	80.45	88.20	20.60	62.60
NegFaceDiff	97.60	86.53	85.33	90.28	80.73	88.10	58.09	73.93
AdaptDiff	97.62	**86.63**	85.04	90.43	**81.82**	**88.31**	**62.38**	**76.56**
IDiff-Face*	**98.92**	89.40	90.54	90.92	86.60	91.28	73.18	82.71
NegFaceDiff*	**98.92**	89.23	90.59	91.13	87.88	91.55	**76.81**	85.13
AdaptDiff*	98.82	**89.77**	**91.14**	**91.62**	**88.15**	**91.90**	76.68	**85.48**
ID^3	97.68	**91.00**	86.84	90.73	82.77	89.80	-	-
DCFace	98.55	89.70	85.33	91.60	82.62	89.56	60.80	74.63
IDiff-Face	98.98	89.30	90.97	91.25	86.87	91.47	23.44	69.69
NegFaceDiff	**98.98**	90.02	91.67	**91.65**	**88.82**	**92.23**	**77.38**	86.11
AdaptDiff	98.83	89.75	**91.77**	91.42	88.77	92.11	76.85	**86.48**

Table 2: Accuracies (in %) of FR models trained on AdaptDiff and SOTA. "*": FR results obtained without data augmentation. "-": works that did not release pre-trained models/generated data and did not report these results. Arc2Face was trained on WebFace4M; the models in the second and third blocks were trained on FFHQ and C-WF, respectively.

Results

We conduct studies on two IDiff-Face versions, pre-trained on either Flickr-Faces-HQ (FFHQ) or CASIA-WebFace (C-WF), following (Caldeira, Damer, and Boutros 2025).

w & w_{max} Ablation Study: NegFaceDiff's performance significantly drops from $w = 0.5$ to $w = 1.0$ (Table 1). Since p^- pushes sampling towards the desired identity, highly accentuating it through the whole sampling unnecessarily hinders intra-class variability, restricting fixed negative sampling to a lower w. AdaptDiff addresses this limitation by reducing the guidance strength magnitude in early sampling, allowing to push the negative condition strength to higher magnitudes in later stages without compromising performance. This is supported by AdaptDiff's superiority with $w_{max} = 1.0$ over $w_{max} = 0.5$. Hence, we set $w_{max} = 1.0$ for all the remaining experiments.

Identity Separability: We compare the identity separability of the data generated by IDiff-Face, NegFaceDiff and our AdaptDiff, following (Caldeira, Damer, and Boutros 2025). Figure 1 displays their genuine and impostor score distributions and relevant metrics. AdaptDiff presents the smallest overlap between the two distributions, revealing higher identity-separability, as numerically supported by its lower EER. Moreover, AdaptDiff has the highest genuine standard deviation, suggesting higher intra-identity variation.

Face Recognition: Table 2 presents evaluation results of FRs trained on AdaptDiff and previous DM SOTA. When using a DM pre-trained on FFHQ, AdaptDiff surpassed SOTA approaches on average on the small-scale benchmarks and significantly outperformed them on the challenging IJB-C in most scenarios. Using data augmentation in FR training significantly improved FR performance. When using a DM trained on C-WF, AdaptDiff presented very competitive performance with SOTA. Data augmentation slightly improved FR performance. The lack of significant improvements, such as those verified with FFHQ, derives from the higher intra-identity variation of data generated by DMs trained on C-WF, which reduces the impact of strategies tailored to improve this property, such as AdaptDiff.

Conclusion

This work proposed a novel adaptive strategy to weight CFG positive and negative conditions across sampling time steps for identity-conditioned DMs. AdaptDiff provides a trade-off between the DM's freedom to explore its latent space and the benefits introduced by negative prompting, resulting in synthetic datasets with higher inter-class separability and intra-class variation, and improving FR performance.

Ethics Statement

Face recognition (FR) systems are currently deployed in several applications, ranging from automated border control (Guo and Zhang 2019; del Rio Saez et al. 2016) to biometric on-device authentication (Prakash et al. 2021). While these models were traditionally trained on authentic facial data, this practice has raised several concerns regarding user consent, privacy and general data protection, along with the technical limitations of collecting, sharing and storing sensitive biometric data as well as the respective legal and ethical concerns (Voigt and Bussche 2017; 740 ILCS/14 2008). To mitigate this issue, recent works (Papantoniou et al. 2024; Xu et al. 2024; Boutros et al. 2023b; Kim et al. 2023) focused on enhancing facial data generation models, towards the generation of synthetic FR training datasets. Our AdaptDiff constitutes one of such research efforts, bringing improvement over previous SOTA synthetic face generation approaches. We further acknowledge but strongly reject any potential for unlawful or ill-intended use of the developed sampling technology. Although we exclusively use DMs to generate faces from synthetic identities in the work, DMs can also generate synthetic pictures of real identities (Papantoniou et al. 2024), which can be used for malicious purposes such as impersonation and evidence fabrication.

Acknowledgments

This research work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

740 ILCS/14. 2008. Biometric Information Privacy Act (BIPA). Public act 095-994, Illinois General Assembly.

Ban, Y.; Wang, R.; Zhou, T.; Cheng, M.; Gong, B.; and Hsieh, C.-J. 2024. Understanding the Impact of Negative Prompts: When and How Do They Take Effect? In *European Conference on Computer Vision*, 190–206. Springer.

Boutros, F.; Grebe, J. H.; Kuijper, A.; and Damer, N. 2023a. IDiff-Face: Synthetic-based Face Recognition through Fizzy Identity-Conditioned Diffusion Models. In *ICCV*, 19593–19604. IEEE.

Boutros, F.; Struc, V.; Fiérrez, J.; and Damer, N. 2023b. Synthetic data for face recognition: Current state and future prospects. *Image Vis. Comput.*, 135: 104688.

Caldeira, E.; Damer, N.; and Boutros, F. 2025. NegFaceDiff: The Power of Negative Context in Identity-Conditioned Diffusion for Synthetic Face Generation. arXiv:2508.09661.

del Rio Saez, J. S.; Moctezuma, D.; Conde, C.; de Diego, I. M.; and Cabello, E. 2016. Automated border control e-gates and facial recognition systems. *Comput. Secur.*, 62: 49–72.

Guo, G.; and Zhang, N. 2019. A survey on deep learning based face recognition. *Comput. Vis. Image Underst.*, 189.

Kim, M.; Liu, F.; Jain, A. K.; and Liu, X. 2023. DCFace: Synthetic Face Generation with Dual Condition Diffusion Model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, 12715–12725. IEEE.

Papantoniou, F. P.; Lattas, A.; Moschoglou, S.; Deng, J.; Kainz, B.; and Zafeiriou, S. 2024. Arc2Face: A Foundation Model for ID-Consistent Human Faces. In *ECCV (37)*, volume 15095 of *Lecture Notes in Computer Science*, 241–261. Springer.

Prakash, D. S.; Ballard, L. E.; Hauck, J. V.; Tang, F.; Littwin, E.; Vasu, P. K. A.; Littwin, G.; Gernoth, T.; Kucerova, L.; Kostka, P.; et al. 2021. Biometric authentication techniques. US Patent 10,929,515.

Sadat, S.; Buhmann, J.; Bradley, D.; Hilliges, O.; and Weber, R. M. 2024. CADS: Unleashing the Diversity of Diffusion Models through Condition-Annealed Sampling. In *ICLR*. OpenReview.net.

Voigt, P.; and Bussche, A. v. d. 2017. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. 1st edition. ISBN 3319579584, 9783319579580.

Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; and Liu, W. 2018. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In *CVPR*, 5265–5274. Computer Vision Foundation / IEEE Computer Society.

Wang, R.; Liu, T.; Hsieh, C.; and Gong, B. 2024. On Discrete Prompt Optimization for Diffusion Models. In *ICML*. OpenReview.net.

Xu, J.; Li, S.; Wu, J.; Xiong, M.; Deng, A.; Ji, J.; Huang, Y.; Mu, G.; Feng, W.; Ding, S.; and Hooi, B. 2024. ID^3: Identity-Preserving-yet-Diversified Diffusion Models for Synthetic Face Recognition. In *NeurIPS*.