Supplementary Material: Face Identity Unlearning for Retrieval via Embedding Dispersion

Mikhail Zakharov
Independent Researcher
Moscow, Russia
m.zakharov.research@gmail.com

1. Additional Experimental Analysis

We conducted additional experiments for each baseline, performing a lightweight hyperparameter search aimed at achieving the best possible behavior. These experiments were carried out on a single CelebA [9] forget set, and the corresponding results are presented in Table 1. The configurations that demonstrated the best forgetting–retention trade-off were reported in the main paper. Note that the absolute metric values differ from those in the main paper, as the latter report mean ± std over three distinct forget sets, while the lightweight hyperparameter search is run on one split only. Nevertheless, the selected configurations fall within the expected range.

For Boundary Shrink (BS) [2], the original paper uses a learning rate of 1×10^{-5} and 10 unlearning epochs. In their face recognition setup, the authors unlearn a single class from VGGFace2 [1], which has approximately 362.6 images per identity. While the batch size is not stated in the paper, the publicly released code uses a batch size of 64. Under this setting, the number of iterations per epoch is approximately $\lceil 362/64 \rceil = 6$, yielding an estimated upper bound of 60 unlearning iterations in total. Furthermore, a later study on Contrastive Unlearning [8] reports that Boundary Shrink typically converges within 40–60 iterations when applied to a ResNet [7]-based model. Motivated by these observations, we evaluate Boundary Shrink under 100 and 500 unlearning iterations, using learning rates of 1×10^{-5} and 1×10^{-4}. In practice, this method demonstrates the strongest transferability to the retrieval scenario among all baselines, without breaking the model.

For Lipschitz Unlearning (LipLoss) [4], the original paper uses a learning rate of 3×10^{-4}, a noise standard deviation of 0.5, and a single unlearning epoch for full-class removal on a CNN-based model. The publicly available code sets the number of noise samples to $n = 25$. Motivated by these settings, we evaluate Lipschitz Unlearning under the following hyperparameter ranges: SalUn [3] $\in \{0.1, 0.5\}$, std $\in \{0.1, 0.5\}$, $n = 25$, $\lambda_{\text{retain}} \in \{0.0, 0.05, 0.1, 0.5, 1.0\}$, iterations $\in \{5, 50, 500, 1000\}$, and a learning rate of 1×10^{-4}. For weight-saliency estimation, we experiment with the CosFace [10] Gradient Ascent [5] (CFGA), LipLoss, and EmbNorm (EN) (as used in the original implementation) loss functions. In practice, we observe that the method either breaks the model or yields only weak forgetting.

For Random Labeling (RL) [5, 6], we experiment with $\lambda_{\text{retain}} \in \{0.0, 0.1, 0.5, 1.0\}$, iterations $\in \{25, 250, 500\}$, and a learning rate of 1×10^{-4}. We observe that the method either does not translate to the retrieval scenario, yielding low Δ_{mAP} and $\Delta_{\text{R@1}}$ on $\mathcal{D}_f^{\text{test}}$, or breaks the model.

For Gradient Ascent (GA) [5], we experiment with iterations $\in \{10, 25, 50\}$, and learning rates of 1×10^{-4} and 1×10^{-5}. We observe that the method either does not transfer to the retrieval scenario, similar to RL, or breaks the model.

For Contrastive Unlearning, the original paper uses a learning rate of 1×10^{-3} and 1×10^{-4} for a ResNet-based model, and according to their graphs, it converges around 50–70 iterations. Motivated by these observations, we evaluate Contrastive Unlearning under the following hyperparameter ranges: $\tau \in \{0.1, 0.2, 0.5\}$, iterations $\in \{50, 100, 250\}$, $\lambda_{\text{retain}} \in \{0.0, 0.001, 0.05, 0.1\}$, and learning rates of 1×10^{-3} and 1×10^{-4}. In practice, we observe that the method either breaks the model or yields only weak forgetting.

Methods	$\mathcal{D}_f^{\text{test}}$ mAP	$\mathcal{D}_f^{\text{test}}$ R@1	CFP-FP R@1	$\mathcal{D}_f^{\text{test}}$ CS	$\mathcal{D}_r^{\text{test}}$ CS	$\mathcal{D}_f^{\text{test}}$ Acc	$\mathcal{D}_r^{\text{test}}$ Acc
Original	**87.7**	**97.1**	**97.5**	**0.634**	**0.616**	**97.2**	**96.6**
Disp($m = 0.2$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 500 Iters	20.7	75.4	97.3	0.086	0.596	3.5	96.5
BS $lr = 1 \times 10^{-5}$ $\lambda_{\text{retain}} = 0.0$ HeadFreeze 100 Iters	84.4	97.1	97.4	0.401	0.619	13.5	96.5
BS $lr = 1 \times 10^{-5}$ $\lambda_{\text{retain}} = 0.0$ 100 Iters	84.6	97.1	97.5	0.400	0.619	10.6	96.5
BS $lr = 1 \times 10^{-5}$ $\lambda_{\text{retain}} = 0.0$ 500 Iters	63.5	96.4	97.0	0.265	0.592	10.1	96.4
BS $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 500 Iters	72.3	95.7	96.9	0.345	0.567	10.6	96.3
CFGA-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 500 Iters	81.3	97.1	87.1	0.615	0.590	24.6	88.5
CFGA-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 1000 Iters	10.0	43.5	1.9	0.344	0.331	0.0	0.6
CFGA-SALUN0.5 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	38.3	84.1	7.7	0.786	0.812	2.2	3.7
CFGA-SALUN0.1 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	38.1	82.6	7.8	0.788	0.815	0.7	3.7
LL-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 1000 Iters	6.2	31.9	1.6	0.346	0.282	0.0	0.2
LL-SALUN0.5 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	37.6	81.9	7.6	0.791	0.818	1.4	3.4
LL-SALUN0.1 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	37.6	81.9	7.6	0.792	0.819	1.4	3.3
LL-SALUN0.1 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 5 Iters	47.3	89.9	17.5	0.749	0.762	10.1	17.2
EN-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 1000 Iters	6.9	33.3	1.5	0.382	0.328	0.0	0.2
EN-SALUN0.5 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	38.2	82.6	7.7	0.787	0.814	0.7	3.8
EN-SALUN0.1 + LipLoss($n = 25$, $std = 0.5$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	38.1	82.6	7.8	0.788	0.815	0.7	3.7
EN-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 1.0$ 1000 Iters	83.9	97.1	97.2	0.600	0.643	89.1	96.4
EN-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.5$ 1000 Iters	83.0	97.1	97.4	0.601	0.647	76.1	96.4
EN-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.1$ 1000 Iters	82.3	97.1	97.1	0.603	0.650	52.9	96.3
EN-SALUN0.5 + LipLoss($n = 25$, $std = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.05$ 1000 Iters	81.5	96.4	96.7	0.601	0.655	44.2	96.2
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 1.0$ 25 Iters	87.3	97.1	97.3	0.636	0.624	0.0	96.5
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 1.0$ 250 Iters	87.7	97.1	97.7	0.783	0.628	2.2	96.5
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 1.0$ 500 Iters	87.9	97.1	97.6	0.839	0.635	2.2	96.6
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.5$ 250 Iters	87.7	97.1	97.5	0.812	0.636	0.7	96.5
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.1$ 250 Iters	87.9	97.8	94.4	0.883	0.723	0.0	96.4
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 250 Iters	64.7	96.4	15.4	0.942	0.919	0.0	15.1
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 25 Iters	83.9	96.4	91.0	0.675	0.632	0.0	85.5
RL $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.1$ HeadFreeze 250 Iters	88.1	97.8	97.0	0.846	0.658	0.0	96.4
GA $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 10 Iters	86.8	97.1	97.1	0.665	0.665	0.0	96.2
GA $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 25 Iters	88.1	97.8	87.8	0.927	0.788	0.0	39.6
GA $lr = 1 \times 10^{-5}$ $\lambda_{\text{retain}} = 0.0$ 25 Iters	85.9	97.1	97.5	0.531	0.637	59.4	96.5
GA $lr = 1 \times 10^{-5}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	87.3	97.1	97.2	0.681	0.663	0.0	96.4
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-4}$ $\lambda_{\text{retain}} = 0.0$ 250 Iters	87.7	97.1	97.5	0.603	0.619	97.1	96.6
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.0$ 250 Iters	2.7	16.7	0.7	0.998	0.999	0.0	0.0
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.0$ 100 Iters	23.7	66.7	1.6	0.663	0.591	10.1	4.3
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.0$ 50 Iters	87.7	97.1	97.6	0.605	0.618	97.1	96.6
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.1$ 100 Iters	88.0	97.1	97.6	0.625	0.623	97.1	96.6
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.05$ 100 Iters	87.5	97.1	97.5	0.640	0.614	96.4	96.5
Contrastive($\tau = 0.1$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.001$ 100 Iters	17.6	59.4	1.3	0.774	0.867	1.4	0.2
Contrastive($\tau = 0.2$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.0$ 100 Iters	87.8	97.1	97.5	0.608	0.616	97.1	96.6
Contrastive($\tau = 0.5$) $lr = 1 \times 10^{-3}$ $\lambda_{\text{retain}} = 0.0$ 100 Iters	87.7	97.1	97.5	0.602	0.619	97.1	96.6

Table 1. Results for a lightweight hyperparameter search for each baseline. Rows highlighted in gray correspond to the configurations reported in the main paper.

References

[1] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition*, pages 67–74. IEEE, 2018. 1

[2] Min Chen, Weizhuo Gao, Gaoyang Liu, Kai Peng, and Chen Wang. Boundary Unlearning: Rapid Forgetting of Deep Networks via Shifting the Decision Boundary. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1

[3] Chongyu Fan, Jiancheng Liu, Yihua Zhang, Eric Wong, Dennis Wei, and Sijia Liu. SalUn: Empowering Machine Unlearning via Gradient-based Weight Saliency in Both Image Classification and Generation. In *International Conference on Learning Representations (ICLR)*, 2024. 1

[4] Jack Foster, Kyle Fogarty, Stefan Schoepf, Cengiz Öztireli, and Alexandra Brintrup. Zero-Shot Machine Unlearning at Scale via Lipschitz Regularization. *arXiv preprint arXiv:2402.01401*, 2024. 1

[5] Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal Sunshine of the Spotless Net: Selective Forgetting in Deep Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[6] Tomohiro Hayase, Suguru Yasutomi, and Takashi Katoh. Selective Forgetting of Deep Networks at a Finer Level than Samples. *arXiv preprint arXiv:2012.11849*, 2020. 1

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *CVPR*, pages 770–778, 2016. 1

[8] Hong kyu Lee, Qiuchen Zhang, Carl Yang, Jian Lou, and Li Xiong. Contrastive Unlearning: A Contrastive Approach to Machine Unlearning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, 2025. 1

[9] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015. 1

[10] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. CosFace: Large margin cosine loss for deep face recognition. In *CVPR*, pages 5265–5274, 2018. 1