LU-500: A LOGO BENCHMARK FOR CONCEPT UNLEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Current concept unlearning approaches for copyright have achieved notable progress in handling styles or portrait-like representations. However, the task of unlearning company logos remains largely unexplored. This challenge stems from logos' simplicity, omnipresence, and strong associations with branded products, often occupying minimal space within an image. To bridge this gap, we introduce LU-500, a comprehensive benchmark for logo unlearning, consisting of 10 prompts to generate images of logos from Fortune Global 500 companies. Our benchmark features two tracks: LUex-500, with explicit prompts, and LUim-500, requiring implicit reasoning to address real-world scenarios like standard usage and adversarial attacks. We further propose five novel, multi-grained evaluation metrics, ranging from local logo regions to global image attributes and spanning both pixel and latent spaces, enabling a robust quantitative analysis of complex visual scenes. Experimental results reveal that existing inference-time unlearning techniques, such as NP, SLD, SEGA, and fine-tuning-based methods like ESD and Forget-Me-Not, all fall short in logo unlearning. To investigate this limitation, we propose a prompt-based baseline using large language models, which demonstrates significant improvements, highlighting the potential of unlearning in semantic space. Additionally, we analyze the correlation between the unlearning performance of an image and its characteristics such as logo area, location, and fractal dimension. We find that SSIM might be a profit control for logo unlearning.

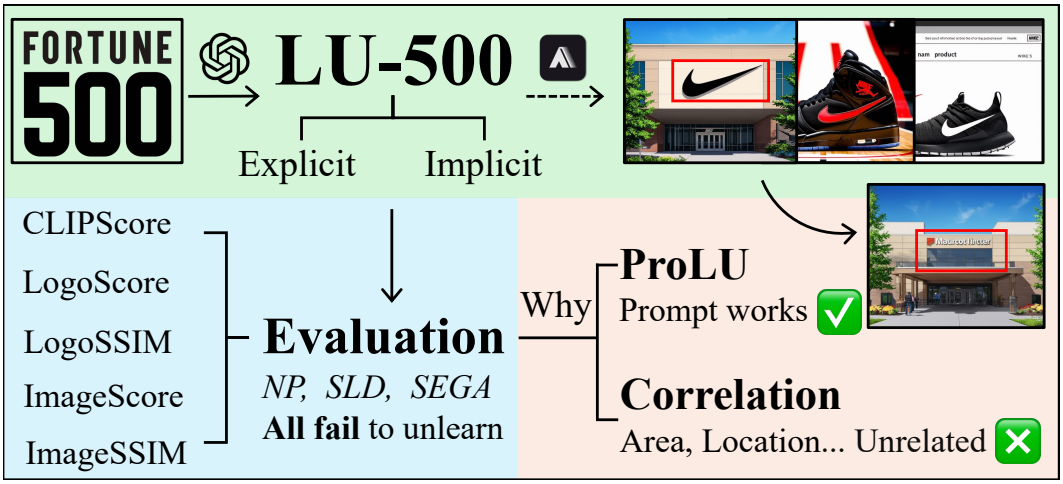

Figure 1: **We propose a logo unlearning benchmark LU-500 for Fortune Global 500 companies.** Five metrics are proposed for a multi-grained evaluation to evaluate existing unlearning approaches. We find that current unlearning approaches such as NP (Rao, 2023), SLD (Schramowski et al., 2023), and SEGA (Brack et al., 2023) all fail to unlearn logos. To investigate why logos are hard to unlearn, we provide a prompt-based unlearning baseline, ProLU, showing an improved performance, which demonstrates the semantic space as a promising direction. We also conduct experiments to analyze the correlation between unlearning performance and logo characteristics, such as area and location.

1 INTRODUCTION

Concept unlearning (Wu et al., 2024; Gao et al., 2024; Nguyen et al., 2022) has been introduced to prevent text-to-image models from reproducing specific content, thereby reducing the generation of potentially harmful images such as sensitive content, leading to potential harm like copyright infringement (Dou et al., 2024; Liu et al., 2024; Meeus et al., 2024). However, current concept unlearning techniques for copyright primarily focus on styles or portrait-like representations (Ren et al., 2024) but overlook the specific case of logos, whose unauthorized use can compromise brand integrity and lead to legal news.

The logo presents a significant distinction from existing concept unlearning tasks. Firstly, they are inherently simplified visuals designed for easy recognition and retention, but varying in the pixel space by words, styles, colors, and shapes. Besides, logos are encountered frequently across various platforms, sometimes as small parts of the background or embedded in unexpected places, making them prominent yet deceptively simple elements in visual data. Furthermore, logos maintain a strong semantic association with company products, complicating the unlearning task. For example, in generating a MacBook image, the model inherently associates the "Apple logo" with the product, even without an explicit mention. This raises the question: *How effective are current concept unlearning methods for logos?*

To answer this question, we construct a dataset, LU-500, to evaluate the logo unlearning of Fortune Global 500 companies. It contains 9584 text-to-image (T2I) prompts which can generate images involving logos of these companies. All these prompts can successfully generate valid images with assigned logos when fed to stable-diffusion-3-medium (stabilityai, 2024) after careful human check. The generated prompts are of high quality with more than 95% success rate in generating logo-involving images[1]. Moreover, the logo generation of our LU-500 will continue to improve as the development of T2I models, such as Stable Diffusion (Esser et al., 2024; Rombach et al., 2022) and Flux (Labs, 2024), since logos of the world-known companies are common in internet-based data, *i.e.*, the pretraining data of T2I models.

Since attackers might try to escape concept unlearning approaches by implicit textual descriptions, we create two tracks which take explicit and implicit textual descriptions for logo-containing images generation respectively. Statistics of the two tracks, LUex-500 and LUim-500, such as prompt length and the word "logo" frequency in prompts are shown in Figure 2. On average, LUex-500 has shorter prompts with more "logo" occurrences, whereas LUim-500 features implicit and longer prompts that involve reasoning about company products, websites, and stores.

For evaluation, we primarily focus on inference-time unlearning methods on LU-500, and we also include evaluations of current state-of-the-art fine-tuning-based methods on a subset of LUex-500. To measure the extent of unlearning for each method, we propose 5 novel metrics to measure local logo deletion and background preservation across two space levels, local logo–global background, and two semantic levels–text-image and image-image latent space. These metrics allow a quantitative assessment of unlearning in complex scenes where logos vary in position and size.

Our experiments show that current inference-time unlearning techniques, such as NP (Rao, 2023), SLD (Schramowski et al., 2023), SEGA (Brack et al., 2023), and fine-tuning-based methods like ESD (Gandikota et al., 2023) and Forget-Me-Not (Zhang et al., 2024a) have a poor performance on our LU-500.

To find out *why logos are so hard to unlearn*, we introduce a new prompt-based logo unlearning baseline, ProLU. It shows substantial improvements compared to existing unlearning methods but with less background conservation. This phenomenon suggests that the prompt space offers a promising alternative. Furthermore, we discuss possible characteristics of logo images that might cause the failure of logo unlearning. Pearson coefficient is calculated between the unlearning effectiveness of an image and its characteristics such as area, location, edge density, shape count, texture complexity, and fractal dimension. We find that SSIM (Wang et al., 2004) or related pixel-level control might be good guidance for logo unlearning. An overview of our work is shown in Figure 1.

Our contribution:

[1]Note that the number of initially constructed prompts is 10000 and we drop the prompts failing to generate logos.

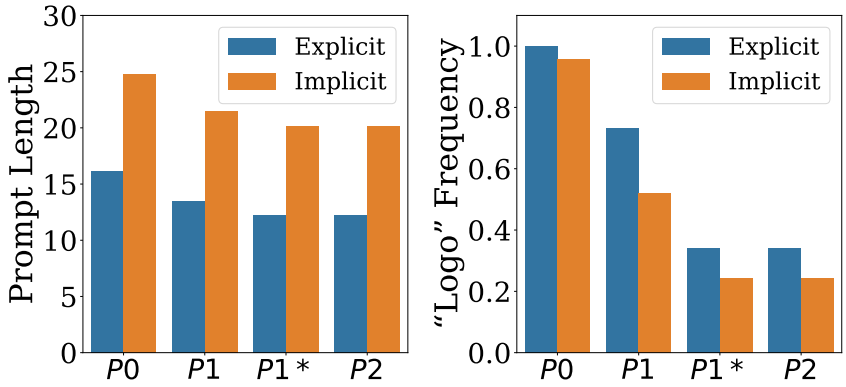

Figure 2: **Statistical difference of prompts during generation of LUex-500 and LUim-500.** $P0$ is the initial prompt of our dataset, $P1$ and $P1^*$ is the prompt before and after reflection. $P2$ is the final text prompt for T2I generation.

- We propose a logo unlearning benchmark, LU-500, to evaluate the concept unlearning methods on logos of Fortune Global 500 companies for copyright protection.
- We propose 5 metrics for quantitative evaluation of current inference-time concept unlearning methods and demonstrate their ineffectiveness.
- We provide a prompt-based unlearning baseline with substantial improvement, showing that prompt space is useful for intertwined concepts such as logos.
- We conduct a correlation analysis between unlearning performance in one image and its areas, locations, edge density, shape count, texture complexity, and so on.

2 Related Work

2.1 Logo Benchmark

Logos have long been a subject of interest in machine learning research. Classification datasets, like Logo-2k+ (Wang et al., 2020) and Weblogo-2m (Su et al., 2017), focus on logo categorization and set a foundation for brand recognition. Retrieval datasets (Joly & Buisson, 2009; Romberg et al., 2011; Bhunia et al., 2019) emphasize searching and retrieving logos within large image collections. Object detection datasets (Jin et al., 2020; Zhang et al., 2021; Hou et al., 2021; Li et al., 2022; Hou et al., 2023), such as LogoDet-3K (Wang et al., 2022) and QMUL-OpenLogo (Su et al., 2018), provide annotations for localizing logos across varied contexts. In-the-wild datasets (Hoi et al., 2015; Tüzkö et al., 2017; Yang et al., 2023) simulate real-world challenges—like varied lighting, angles, and occlusions—pushing algorithms toward robust, generalizable logo recognition. Systematic research specifically targeting logos as a concept, however, remains largely unexplored. To our knowledge, this study is the first to address this gap, focusing specifically on logo unlearning and proposing a logo benchmark.

2.2 Concept Unlearn

Recent research on concept unlearning (Ren et al., 2024) has primarily focused on harmful content (Kim et al., 2023; Schramowski et al., 2023; Brack et al., 2023), nudity (Gandikota et al., 2023; Lyu et al., 2024; Kim et al., 2023; Gandikota et al., 2024; Lu et al., 2024; Xiong et al., 2024; Schramowski et al., 2023; Brack et al., 2023), celebrity likeness (Zhang et al., 2024a; Lu et al., 2024; Ma et al., 2024), copyright violations (Lyu et al., 2024; Ma et al., 2024), and art styles (Kim et al., 2023; Gandikota et al., 2023; Lyu et al., 2024; Zhang et al., 2024a; Gandikota et al., 2024; Lu et al., 2024; Xiong et al., 2024; Brack et al., 2023; Zhang et al., 2024b; Ma et al., 2024). However, the unlearning of logos—key branding elements under strict copyright constraints—has been largely overlooked. Despite logos' high recognizability and legal sensitivity, their unlearning in synthetic

Figure 3: **T2I models generate images with an assigned logo.** Newest SD3, SD3.5, and FLUX models can generate high-quality logo-containing images, while SD1.5 cannot.

media remains unexplored, despite the risks posed by realistic logos in fake news and misinformation. Therefore, we introduce logo unlearning as a new task, proposing 5 metrics for evaluation, distinct from traditional success rate measures (Schramowski et al., 2023; Brack et al., 2023; Rao, 2023). Existing unlearning methods can be broadly classified into fine-tuning-based and inference-time approaches (Ren et al., 2024). Fine-tuning methods require significant resources (Gandikota et al., 2023; Kumari et al., 2023; Kim et al., 2023; Lyu et al., 2024; Gandikota et al., 2024; Lu et al., 2024; Xiong et al., 2024), while inference-time methods, which avoid fine-tuning, have gained attention due to the growing size of models (Ren et al., 2024; Schramowski et al., 2023; Brack et al., 2023; Rao, 2023). We show the limitations of existing methods in logo unlearning and propose a novel prompt-based method, ProLU.

3 LU-500

We present LU-500, a novel benchmark designed to evaluate the performance of logo unlearning. We begin by describing the construction process of LU-500 in Section 3.1, followed by an overview of the proposed logo unlearning baselines, ProLU, shown in Section 3.2. Finally, we provide a detailed explanation of the metrics used to assess the effectiveness of these concept unlearning baselines in Section 3.3.

3.1 Benchmark

Logos are crucial assets of a company's intellectual property, designed for quick recognition and significant brand value. As text-to-image (T2I) models rely on publicly available data, logos are often unintentionally incorporated, exposing brands to unauthorized use. Recent T2I models can generate images with logos, as shown in Figure 3. While current unlearning methods focus on broader categories like styles or portraits (Ren et al., 2024), they overlook logos—simple yet widely recognized elements. Given their unique designs and ubiquity, logos require specialized handling to prevent infringement. We introduce a new benchmark, LU-500, to highlight (1) logo leakage in T2I models, and (2) the effectiveness of current unlearning methods in logo protection.

For real-world relevance, we use 2024 Fortune Global 500 list, which includes logos from the world's largest companies, adding practical significance to our benchmark. We simulate two scenarios in LU-500 to reflect realistic logo exploitation: LUex-500, where a T2I model is directly prompted to generate an image with a logo, and LUim-500, where contextual cues subtly lead the model to incorporate a logo. Each dataset consists of 10 prompts per company, totaling 9584 prompts after invalid ones removed. LUex-500 uses straightforward prompts for explicit logo generation, while LUim-500 involves more complex prompts related to products, websites, or ads to generate logos implicitly.

We select stable-diffusion-3-medium (Esser et al., 2024; stabilityai, 2024) as our T2I model for generating images from LU-500, based on three factors: (1) older stable diffusion models can't reliably generate logos even before unlearning; (2) closed-source models like Midjourney (Midjourney, 2023) and DALLE3 (openai, 2024a) can't be modified for unlearning; and (3) stable-diffusion-3-medium strikes a balance between performance and resource efficiency, as shown in Figure 3. This choice ensures consistent evaluation of baseline methods before and after unlearning. The dataset construction process is shown in Figure 4.

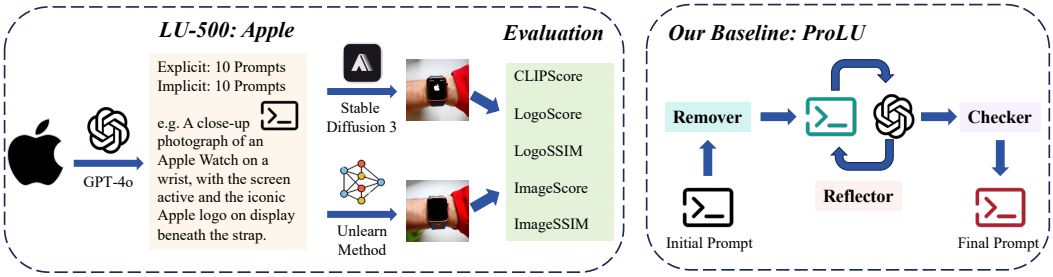

Figure 4: **Left:** The generation process of LU-500 using Apple as an example, along with the evaluation procedure. **Right:** Our baseline ProLU, used as an unlearning method for evaluation.

3.2 Baseline: ProLU

To evaluate unlearn effectiveness within the prompt space, we introduce ProLU, an inference-time method that avoids retraining and is more resource-efficient than fine-tuning. ProLU uses three agents developed with OpenAI GPT-4o (openai, 2024b): Remover, Reflector, and Checker, each refining prompts to remove logo references. The process, shown in Figure 4, is detailed in supplementary material.

Starting with an initial prompt $P0$ containing a logo-related concept, the Remover agent edits it to remove the logo reference while keeping other details intact, producing $P1$. The Reflector agent checks whether the logo is removed, refining $P1$ as needed. The Checker agent ensures no logo references remain in $P1$. If any reference persists, the prompt is further edited until a final prompt $P2$ is produced. This prompt is then used with stable-diffusion-3-medium for evaluation. ProLU offers a streamlined, efficient approach for unlearning logos within prompts.

3.3 Evaluation

To systematically evaluate the effectiveness of logo unlearning, we consider the unique characteristics of logos compared to other unlearning targets, such as celebrities or artistic styles. Unlike these subjects, logos are often neither central nor dominant in an image and may only appear in specific, localized regions. This difference limits the applicability of traditional image-level evaluation methods, as unlearning logos primarily affects local contexts rather than the overall image composition. Therefore, we introduce two evaluation components: (1) logo detection, to assess the presence of any recognizable logo elements post-unlearning, and (2) metrics specifically tailored to capture the effects of logo unlearning from both local and global perspectives.

3.3.1 Logo Detection

Existing logo detection methods face two main challenges in the context of logo unlearning. First, they typically rely on specific datasets, limiting detection capabilities to logos that were included in the training data. Second, these methods are often query-based, requiring a one-shot detection image that heavily depends on the quality of the query, which can impact detection accuracy.

To address these limitations, we employ a flexible open-vocabulary object detection approach using the OWLv2 model (Minderer et al., 2024). OWLv2 enables logo detection based on text queries, bypassing the need for predefined logo images and providing adaptability across a wider range of logos. Through OWLv2, we achieve 98% accuracy in detecting company logos in generated images, as verified by human reviewers who double-checked the model's detections.

3.3.2 Metrics

Existing inference-time unlearning baselines (Schramowski et al., 2023; Brack et al., 2023; Rao, 2023) typically evaluate performance using success rates. While informative, it provides a limited view of unlearning effectiveness, focusing solely on concept presence. To provide a more comprehensive assessment, we propose a framework based on five distinct metrics that evaluate unlearning performance across multiple dimensions.

Figure 5: **The effect of different baseline methods in unlearning company logos arranged in alphabetical order by company name.** These logos can also be categorized into three types: simple graphic-based logos (*e.g.*, Apple, Tesla), text-based logos (*e.g.*, Coca-Cola, Disney), and more complex logos like Mercedes-Benz combining both text and graphics. As safety guidance gradually increases, existing unlearning methods gradually unlearn the logos, but this comes at the cost of increasingly significant changes to the background.

Our evaluation framework incorporates both local, global perspectives and both pixel, latent levels to assess logo removal effectiveness and overall image integrity. Furthermore, if the generated image contains no detectable logos, we assign all metrics a value of zero to reflect the total absence of the target concept, ensuring that results are consistent and meaningful across different cases.

CLIPScore This measures the semantic alignment between the detected logo and its corresponding company name using CLIP (Radford et al., 2021) cosine similarity. Lower scores indicate a more effective semantic dissociation from the brand.

LogoScore and LogoSSIM These metrics quantify local logo alteration by comparing the logo region before and after unlearning. LogoScore uses CLIP feature similarity, while LogoSSIM uses Structural Similarity (SSIM) (Wang et al., 2004). For both metrics, lower scores are better, signifying more effective logo removal.

ImageScore and ImageSSIM These metrics assess global image preservation by comparing the entire image before and after unlearning. ImageScore uses CLIP similarity for contextual coherence, while ImageSSIM uses SSIM for structural fidelity. For both metrics, higher scores are better, indicating that non-target elements are well-preserved.

4 Experiment

4.1 Experimental Setting

We select stable-diffusion-3-medium (Esser et al., 2024; stabilityai, 2024) as our T2I model to generate images from LU-500 prompts. In stable-diffusion-3-medium, NP (Rao, 2023) is supported directly, while SLD (Schramowski et al., 2023) and SEGA (Brack et al., 2023) are initially compatible only with stable-diffusion-v1-5 (runawayml, 2022). To adapt these methods, we implement

Table 1: **The performance of different unlearning methods on the five metrics we proposed is mediocre.** The CLIP-Text of original images without unlearning is 0.3237 and 0.3156 for explicit and implicit scores. We take **bold** for the best and underline for the second best. Among all the baselines, Negative Prompt (NP) (Rao, 2023) shows some improvement, but still lags significantly behind the optimal solution. As safety guidance increases, both SLD (Schramowski et al., 2023) and SEGA (Brack et al., 2023) improve in logo unlearning, but at the cost of losing more background information. ProLU performs best in logo region unlearning, but exhibits average performance in background preservation.

Method	CLIPScore ↓		LogoScore ↓		LogoSSIM ↓		ImageScore ↑		ImageSSIM ↑	
	Explicit	Implicit	Explicit	Implicit	Explicit	Implicit	Explicit	Implicit	Explicit	Implicit
NP	0.3016	0.2947	0.6447	0.6922	0.0944	0.0781	0.7961	0.8189	0.5310	0.4893
SLD_v1	0.3175	0.3077	0.7884	0.7942	0.2881	0.2272	0.9436	0.9433	0.8587	0.8267
SLD_v2	0.2913	0.2849	0.6503	0.6887	0.1409	0.1061	0.8175	0.8462	0.7279	0.6843
SLD_v3	0.2845	0.2811	0.6042	0.6623	0.0806	0.0698	0.7456	0.7868	0.5048	0.4679
SEGA_v1	0.3182	0.3098	0.8397	0.8470	0.4248	0.3655	**0.9761**	**0.9771**	**0.9735**	**0.9655**
SEGA_v2	0.3044	0.2956	0.7480	0.7635	0.2435	0.1895	0.9099	0.9227	0.8972	0.8687
SEGA_v3	0.2989	0.2913	0.7025	0.7322	0.1854	0.1409	0.8725	0.8932	0.8295	0.7919
ProLU (Ours)	**0.2646**	**0.2627**	**0.5209**	**0.5650**	**0.0661**	**0.0512**	0.7120	0.7720	0.5496	0.5257

Algorithm 1 from Appendix H of (Schramowski et al., 2023) and Algorithm 1 from Appendix A of (Brack et al., 2023) in stable-diffusion-3-medium. As inference-time plug-ins, these algorithms integrate seamlessly into the stable-diffusion-3-medium pipeline. The implementation is straightforward and is available in our code. We conduct experiments with three key hyperparameter groups in both SLD and SEGA, as detailed in Table 2. For SLD, we follow the configurations in the original paper (Schramowski et al., 2023), where SLD_v1 corresponds to Hyp-Strong and SLD_v2 to Hyp-Max. For SEGA, we tune hyperparameters based on its paper (Brack et al., 2023), ensuring SEGA_v1 applies stronger safety guidance than SEGA_v2, and SEGA_v2 stronger than SEGA_v3. The random seed is fixed to ensure reproducibility and fairness in the analysis.

Evaluation Metrics CLIPScore measures the similarity between detected logos and the "company logo" text query. Both LogoScore and ImageScore, assess the CLIP score before and after unlearning, with LogoScore focusing on detected logos and ImageScore evaluating the entire image to gauge background retention. LogoSSIM and ImageSSIM calculate the SSIM (Wang et al., 2004) score between two images before and after unlearning, with LogoSSIM analyzing logo-detected regions and ImageSSIM assessing overall background preservation.

4.2 Suboptimal Baseline Performance

Our experiments reveal that existing unlearning methods exhibit limited efficacy for the nuanced task of logo removal. As detailed in Table 1, baselines like NP (Rao, 2023), SLD (Schramowski et al., 2023), and SEGA (Brack et al., 2023) only marginally reduce the CLIPScore compared to the original generations, remaining far from the ideal score of zero. These methods present a clear trade-off: increasing their safety guidance hyperparameters (Table 2) enhances logo removal—reflected by lower CLIPScore, LogoScore, and LogoSSIM scores—but at the cost of severe degradation in background preservation, evidenced by worsening ImageScore and ImageSSIM scores.

Our proposed baseline, ProLU, exemplifies this dilemma. While it consistently achieves the best performance in logo erasure across all local metrics, this superior removal comes with a compromise in global image fidelity. This suggests that current methods struggle to both successfully unlearn the logo in a local region and fully preserve the background information globally.

The qualitative results in Figure 5 further illustrate these findings. For simple graphical logos like Apple's, existing baselines tend to only blur or shrink the logo, whereas ProLU removes it completely. They struggle significantly with text-based logos such as Coca-Cola and Disney, which our method handles more effectively. In more complex cases with multiple logo instances, such as the

Table 2: **Different groups of hyperparameters are included for various levels of unlearning guidance to ensure the effectiveness of baselines.** We keep other hyperparameters the same, such as num_inference_steps as 28, guidance_scale as 7.0, and height_weight as 1024.

	Warmup Steps	Safety Guidance	Threshold	Momentum Scale	Mom Beta
SLD_v1	7	2000	0.025	0.5	0.7
SLD_v2	4	3000	0.500	0.5	0.7
SLD_v3	0	5000	1.000	0.5	0.7
SEGA_v1	10	4	0.990	0.3	0.6
SEGA_v2	7	5	0.950	0.3	0.6
SEGA_v3	5	5	0.900	0.3	0.6

Tesla example, baseline methods often remove only the smaller, more obvious logo while neglecting the larger, more prominent one. Furthermore, when pushed to their limits, these methods introduce severe artifacts. For instance, with the SLD_v3 setting, the Coca-Cola logo becomes blurry, but the model also fabricates a yellow background and other patterns not present in the original image, highlighting a significant loss of global coherence.

4.3 Fine-tuning-based Unlearn Methods

We reference state-of-the-art fine-tuning-based methods (Gandikota et al., 2023; Zhang et al., 2024a) for comparison. We fine-tune using LUim-500 prompts and averaging the results at the company level. The results in Table 3 show that fine-tuning methods still have room for improvement, with ProLU outperforming them. Note that these two unlearning methods are based on different older versions of the stable diffusion model, stable-diffusion-v1-4 and stable-diffusion-2-1-base. Therefore, ProLU is compared separately with each, unlike Table 1, where all methods share the same T2I model, stable-diffusion-3-medium, requiring only one comparison for ProLU. The detailed hyperparameters used during the fine-tuning process are presented in the supplementary materials.

Table 3: **A comparison of two fine-tuning-based methods, ESD (Gandikota et al., 2023) and Forget-Me-Not (Zhang et al., 2024a), on LUim-500. Bold** values indicate the best performance. The results demonstrate that our method, ProLU, comprehensively outperforms both ESD and Forget-Me-Not. Although Forget-Me-Not shows inferior unlearning effectiveness, it preserves more background information, as evidenced by ImageScore and ImageSSIM. These methods are based on different older versions of stable diffusion, so ProLU is compared separately with each.

T2I Model Version	Method	CLIPScore ↓	LogoScore ↓	LogoSSIM ↓	ImageScore ↑	ImageSSIM ↑
stable-diffusion-v1-4	ESD (Gandikota et al., 2023)	0.2405	0.8923	0.1533	0.7231	0.3803
stable-diffusion-v1-4	ProLU (Ours)	**0.2289**	**0.8890**	**0.0298**	**0.7829**	**0.4549**
stable-diffusion-2-1-base	Forget-Me-Not (Zhang et al., 2024a)	0.2751	0.7962	0.1316	**0.7995**	**0.2429**
stable-diffusion-2-1-base	ProLU (Ours)	**0.2545**	**0.6537**	**0.1146**	0.7520	0.2213

4.4 Correlation Analysis

To investigate the poor performance of baseline logo unlearning, we hypothesize that it is influenced by logo characteristics like size, position, and visual complexity. We measure six attributes: Area, Location, Edge Density (Canny, 1986), Shape Count, Texture Complexity (Gebejes & Huertas, 2013), and Fractal Dimension (Lin, 1991). Using the Pearson correlation coefficient (Pearson, 1896) to analyze these attributes against three performance metrics (CLIPScore, LogoScore, LogoSSIM), we find that most factors show weak correlation with unlearning performance, as illustrated in Figure 6. However, LogoSSIM displays a modestly stronger correlation, showing a positive relation-

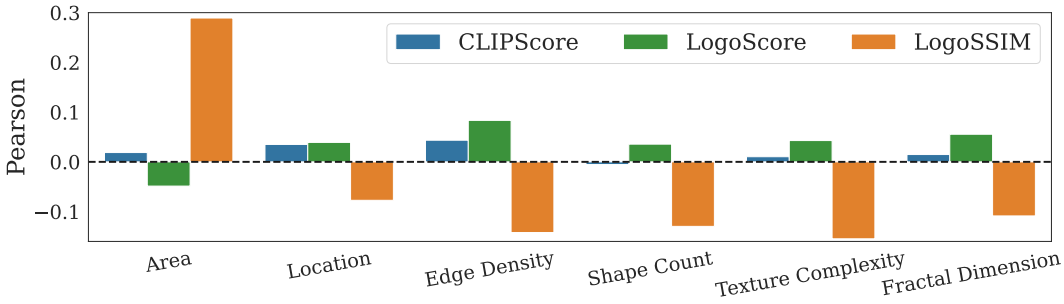

Figure 6: **We explore the Pearson correlation coefficient between the unlearning performance of SEGA (Brack et al., 2023) on an image and its characteristics.** Most factors have little correlation with the performance. However, the LogoSSIM shows a rather stronger correlation with these characteristics, demonstrating SSIM as a possible control for logo unlearning.

ship with logo area (suggesting larger logos are harder to unlearn) and a slight negative relationship with visual complexity (implying more complex logos may be easier to unlearn).

4.5 Ablation Study

An ablation study was conducted on LUim-500 to isolate the contribution of our baseline, each ProLU agent, with the results detailed in Table 4. The analysis reveals that the Remover agent is the most critical component, accounting for over 80% of the total unlearning efficacy (the CLIPScore drops from a 16.76% reduction to only 3.07% without it). The Reflector agent provides a secondary but important refinement, contributing nearly 20% to the final performance. The Checker, acting as a final safeguard, has a negligible direct impact on the outcome, confirming the robustness of the core prompt revision pipeline.

Table 4: Ablation results of ProLU components, measured by CLIPScore.

	Before	ProLU	w/o Remover	w/o Reflector	w/o Checker
CLIPScore	0.3156	0.2627	0.3059	0.2678	0.2627
CLIPScore Decline	0%	16.76%	3.07%	15.15%	16.76%
Importance Ratio ↑	N/A	100%	**83.2%**	<u>16.8%</u>	0%

5 Discussion

Conclusion The introduction of our LU-500 benchmark has cast a spotlight on a critical gap in existing generative AI tools: the profound difficulty of cleanly removing logos from AI-generated imagery. Current techniques often struggle to contend with the intricate ways logos are embedded within complex scenes. In response to this challenge, we present ProLU, a novel baseline that fundamentally shifts the point of intervention. Instead of relying on post-processing, ProLU directly modifies the source text prompt, manipulating the image at the generative level. Our experiments demonstrate that this prompt-level approach achieves a marked leap in erasure efficacy, significantly outperforming conventional methods.

Limitation and Future Work The blunt nature of prompt modification frequently leads to a discernible degradation in overall image quality, causing unintended artifacts and disrupting the coherence of background elements. Recognizing this limitation, our future work is charted along two primary vectors. First, we will substantially expand the LU-500 benchmark to include a more diverse and challenging array of company logos, ensuring greater real-world applicability. Second, and more crucially, we will focus on developing next-generation logo removal techniques that are more ”surgical” in their application. The goal is to create methods that can precisely target and nullify logo-specific features within the latent space, thereby preserving intricate textural details and maintaining background fidelity without compromising the effectiveness of the removal.

Ethics Statement

This work adheres to the ICLR Code of Ethics. Our study does not involve human subjects, sensitive personal data, or experiments that could directly cause harm to individuals or communities. We have taken care to consider issues of fairness, privacy, and security when designing our methods and presenting our results. We are not aware of any potential conflicts of interest, legal compliance issues, or research integrity concerns related to this submission.

Reproducibility Statement

We have made every effort to ensure the reproducibility of our results. Details of the model architecture, training procedures, and evaluation protocols are provided in the main text and appendix. Hyperparameters, dataset preprocessing steps, and implementation details are described in the supplementary materials. To further support reproducibility, we upload the source code as supplementary material. These resources should allow other researchers to replicate our findings and build upon our work.

References

Ayan Kumar Bhunia, Ankan Kumar Bhunia, Shuvozit Ghose, Abhirup Das, Partha Pratim Roy, and Umapada Pal. A deep one-shot network for query-based logo retrieval. *Pattern Recognition*, 96: 106965, 2019.

Manuel Brack, Felix Friedrich, Dominik Hintersdorf, Lukas Struppek, Patrick Schramowski, and Kristian Kersting. Sega: Instructing text-to-image models using semantic guidance. *Advances in Neural Information Processing Systems*, 36:25365–25389, 2023.

John Canny. A computational approach to edge detection, 1986.

Guangyao Dou, Zheyuan Liu, Qing Lyu, Kaize Ding, and Eric Wong. Avoiding copyright infringement via machine unlearning. *arXiv preprint arXiv:2406.10952*, 2024.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*, 2024.

Rohit Gandikota, Joanna Materzynska, Jaden Fiotto-Kaufman, and David Bau. Erasing concepts from diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2426–2436, 2023.

Rohit Gandikota, Hadas Orgad, Yonatan Belinkov, Joanna Materzyńska, and David Bau. Unified concept editing in diffusion models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5111–5120, 2024.

Hongcheng Gao, Tianyu Pang, Chao Du, Taihang Hu, Zhijie Deng, and Min Lin. Meta-unlearning on diffusion models: Preventing relearning unlearned concepts. *arXiv preprint arXiv:2410.12777*, 2024.

A Gebejes and R Huertas. Texture characterization based on grey-level co-occurrence matrix. *Databases*, 9(10):375–378, 2013.

Steven CH Hoi, Xiongwei Wu, Hantang Liu, Yue Wu, Huiqiong Wang, Hui Xue, and Qiang Wu. Logo-net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks. *arXiv preprint arXiv:1511.02462*, 2015.

Qiang Hou, Weiqing Min, Jing Wang, Sujuan Hou, Yuanjie Zheng, and Shuqiang Jiang. Foodlogodet-1500: A dataset for large-scale food logo detection via multi-scale feature decoupling network. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4670–4679, 2021.

Sujuan Hou, Jiacheng Li, Weiqing Min, Qiang Hou, Yanna Zhao, Yuanjie Zheng, and Shuqiang Jiang. Deep learning for logo detection: A survey. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(3):1–23, 2023.

Xuan Jin, Wei Su, Rong Zhang, Yuan He, and Hui Xue. The open brands dataset: Unified brand detection and recognition at scale. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4387–4391. IEEE, 2020.

Alexis Joly and Olivier Buisson. Logo retrieval with a contrario visual query expansion. In *Proceedings of the 17th ACM international conference on Multimedia*, pp. 581–584, 2009.

Sanghyun Kim, Seohyeon Jung, Balhae Kim, Moonseok Choi, Jinwoo Shin, and Juho Lee. Towards safe self-distillation of internet-scale text-to-image diffusion models. *arXiv preprint arXiv:2307.05977*, 2023.

Nupur Kumari, Bingliang Zhang, Sheng-Yu Wang, Eli Shechtman, Richard Zhang, and Jun-Yan Zhu. Ablating concepts in text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22691–22702, 2023.

Black Forest Labs. Flux1.1 pro, 2024. URL https://blackforestlabs.ai/.

Chenge Li, István Fehérvári, Xiaonan Zhao, Ives Macedo, and Srikar Appalaraju. Seetek: Very large-scale open-set logo recognition with text-aware metric learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2544–2553, 2022.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

Xiaoze Liu, Ting Sun, Tianyang Xu, Feijie Wu, Cunxiang Wang, Xiaoqian Wang, and Jing Gao. Shield: Evaluation and defense strategies for copyright compliance in llm text generation. *arXiv preprint arXiv:2406.12975*, 2024.

Shilin Lu, Zilan Wang, Leyang Li, Yanzhu Liu, and Adams Wai-Kin Kong. Mace: Mass concept erasure in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6430–6440, 2024.

Mengyao Lyu, Yuhong Yang, Haiwen Hong, Hui Chen, Xuan Jin, Yuan He, Hui Xue, Jungong Han, and Guiguang Ding. One-dimensional adapter to rule them all: Concepts diffusion models and erasing applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7559–7568, 2024.

Rui Ma, Qiang Zhou, Bangjun Xiao, Yizhu Jin, Daquan Zhou, Xiuyu Li, Aishani Singh, Yi Qu, Kurt Keutzer, Xiaodong Xie, et al. A dataset and benchmark for copyright protection from text-to-image diffusion models. *arXiv preprint arXiv:2403.12052*, 2024.

Matthieu Meeus, Igor Shilov, Manuel Faysse, and Yves-Alexandre de Montjoye. Copyright traps for large language models. *arXiv preprint arXiv:2402.09363*, 2024.

Midjourney. Midjourney (V5.2) [Text-to-Image Model], 2023. URL https://www.midjourney.com.

Matthias Minderer, Alexey Gritsenko, and Neil Houlsby. Scaling open-vocabulary object detection. *Advances in Neural Information Processing Systems*, 36, 2024.

Thanh Tam Nguyen, Thanh Trung Huynh, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, and Quoc Viet Hung Nguyen. A survey of machine unlearning. *arXiv preprint arXiv:2209.02299*, 2022.

openai. Dalle3, 2024a. URL https://openai.com/index/dall-e-3/.

openai. gpt-4o, 2024b. URL https://openai.com/index/hello-gpt-4o/.

Karl Pearson. Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia, 1896.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Dana Rao. Stable diffusion 2.0 release, 2023. URL https://stability.ai/news/stable-diffusion-v2-release.

Jie Ren, Kangrui Chen, Yingqian Cui, Shenglai Zeng, Hui Liu, Yue Xing, Jiliang Tang, and Lingjuan Lyu. Six-cd: Benchmarking concept removals for benign text-to-image diffusion models. *arXiv preprint arXiv:2406.14855*, 2024.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.

Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, and Roelof Van Zwol. Scalable logo recognition in real-world images. In *Proceedings of the 1st ACM international conference on multimedia retrieval*, pp. 1–8, 2011.

runawayml. stable-diffusion-v1-5, 2022. URL https://huggingface.co/runwayml/stable-diffusion-v1-5.

Patrick Schramowski, Manuel Brack, Björn Deiseroth, and Kristian Kersting. Safe latent diffusion: Mitigating inappropriate degeneration in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22522–22531, 2023.

stabilityai. stable-diffusion-3-medium, 2024. URL https://huggingface.co/stabilityai/stable-diffusion-3-medium.

Hang Su, Shaogang Gong, and Xiatian Zhu. Weblogo-2m: Scalable logo detection by deep learning from the web. In *Proceedings of the IEEE international conference on computer vision workshops*, pp. 270–279, 2017.

Hang Su, Xiatian Zhu, and Shaogang Gong. Open logo detection challenge. *arXiv preprint arXiv:1807.01964*, 2018.

Andras Tüzkö, Christian Herrmann, Daniel Manger, and Jürgen Beyerer. Open set logo detection and retrieval. *arXiv preprint arXiv:1710.10891*, 2017.

Jing Wang, Weiqing Min, Sujuan Hou, Shengnan Ma, Yuanjie Zheng, Haishuai Wang, and Shuqiang Jiang. Logo-2k+: A large-scale logo dataset for scalable logo classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 6194–6201, 2020.

Jing Wang, Weiqing Min, Sujuan Hou, Shengnan Ma, Yuanjie Zheng, and Shuqiang Jiang. Logodet-3k: A large-scale image dataset for logo detection. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(1):1–19, 2022.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Yongliang Wu, Shiji Zhou, Mingzhuo Yang, Lianzhe Wang, Wenbo Zhu, Heng Chang, Xiao Zhou, and Xu Yang. Unlearning concepts in diffusion model via concept domain correction and concept preserving gradient. *arXiv preprint arXiv:2405.15304*, 2024.

Tianwei Xiong, Yue Wu, Enze Xie, Zhenguo Li, and Xihui Liu. Editing massive concepts in text-to-image diffusion models. *arXiv preprint arXiv:2403.13807*, 2024.

Qimao Yang, Huili Chen, and Qiwei Dong. Comparative analysis of deep learning models for brand logo classification in real-world scenarios. *arXiv preprint arXiv:2305.12242*, 2023.

Gong Zhang, Kai Wang, Xingqian Xu, Zhangyang Wang, and Humphrey Shi. Forget-me-not: Learning to forget in text-to-image diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1755–1764, 2024a.

Junxing Zhang, Lijun Chen, Chunjuan Bo, and Shuo Yang. Multi-scale vehicle logo detector. *Mobile Networks and Applications*, 26:67–76, 2021.

Yihua Zhang, Yimeng Zhang, Yuguang Yao, Jinghan Jia, Jiancheng Liu, Xiaoming Liu, and Sijia Liu. Unlearncanvas: A stylized image dataset to benchmark machine unlearning for diffusion models. *arXiv preprint arXiv:2402.11846*, 2024b.

A THE USE OF LLMS

In the article, we only used LLMs to polish our writing, and did not use them for any other assistance.

B METRICS

We present five key metrics, each with a clearly defined purpose and scope of responsibility, as outlined in Table 5.

CLIPScore calculates the CLIP similarity between the company logo used as a text query and the logos extracted from the images. A lower value indicates a lower similarity between the text query and the generated logo, suggesting a better unlearning effect.

LogoScore evaluates the CLIP similarity between logos extracted from the images before and after unlearning. A lower value implies a lower similarity, indicating a more effective unlearning process.

LogoSSIM uses SSIM (Structural Similarity Index Measure) to compare the logos extracted before and after unlearning. A smaller value indicates a lower SSIM, meaning the logos are less similar, and the unlearning effect is better.

ImageScore computes the CLIP similarity for the entire image before and after unlearning. A higher value suggests that the background information is more completely preserved during the unlearning process. An effective unlearning method should maintain a higher score for this metric.

ImageSSIM measures the SSIM for the entire image before and after unlearning. A higher value indicates that the background information is more completely retained, which is desirable for a good unlearning method.

These metrics are integral to our methodology, providing a structured framework for assessing performance. Additionally, we offer a comprehensive visual representation of our evaluation process, depicted in detail in Figure 7, to ensure clarity and facilitate understanding of our approach.

Let I_1, I_2, and I_3 denote the original images, and their corresponding images after the unlearning process are denoted as I_a, I_b, and I_c. After applying logo detection, the logo regions with the highest confidence in each image are extracted and denoted as L_1, L_2, L_3, L_a, L_b, and L_c. CLIPScore calculates the CLIP similarity between L_1, L_2, L_3, L_a, L_b, L_c, and the text query "apple logo." LogoScore computes the CLIP similarity between L_1 and L_a, L_2 and L_b, and L_3 and L_c. LogoSSIM measures the SSIM similarity between L_1 and L_a, L_2 and L_b, and L_3 and L_c. ImageScore calculates the CLIP similarity between I_1 and I_a, I_2 and I_b, and I_3 and I_c. ImageSSIM computes the SSIM similarity between I_1 and I_a, I_2 and I_b, and I_3 and I_c.

The algorithm for calculating metrics is presented in Algorithm 1. The CLIP and SSIM similarities are first calculated for each individual prompt by comparing the images before and after unlearning. Then, the results are averaged at both the prompt level and the company level.

C FINE TUNING BASELINE DETAILS

Forget-Me-Not We utilize the "stabilityai/stable-diffusion-2-1-base" text-to-image model, consistent with the original paper. For the image generation phase, we set num_inference_steps to 50, guidance_scale to 7, and num_images_per_prompt to 1. During the training phase, train_batch_size is set to 1, learning_rate to 2.0e-06, and max_train_steps to 35. We employ the adamw optimizer with adam_beta1 of 0.9, adam_beta2 of 0.999, adam_weight_decay of 0.01, adam_epsilon of 1.0e-08, and max_grad_norm of 1.

ESD In accordance with the original paper, we employ the "CompVis/stable-diffusion-v1-4" text-to-image model. For the image generation phase, we set img_size to 512, n_steps to 50, n_imgs to 1, and guidance_scale to 7.5. During the training phase, we utilize the xattn method with a learning_rate of 1e-5.

D 500 COMPANIES RANKS

An additional outcome of our experiments is the ability to rank companies by logo unlearning difficulty, as determined by LU-500, our metrics, and baseline methods. For instance, we derive a difficulty ranking among 500 companies for ProLU using LUim-500 with LogoScore.

Since a smaller value of LogoScore indicates better unlearning performance, a larger value of LogoScore signifies greater difficulty in unlearning. Arranged in descending order of LogoScore, the five most difficult companies to unlearn are *HYUNDAI MOTOR*, *CHINA AEROSPACE SCIENCE & INDUSTRY*, *PANASONIC HOLDINGS*, *RTX* and *TATA MOTORS* with their corresponding LogoScore values: $0.834, 0.8299, 0.8142, 0.8122, 0.8051$. Conversely, the five easiest companies to unlearn are *ORANGE*, *WELLS FARGO*, *AMAZON.COM*, *TARGET* and *WORLD KINECT*, with their corresponding LogoScore values: $0.0907, 0.1838, 0.2119, 0.2159, 0.2206$.

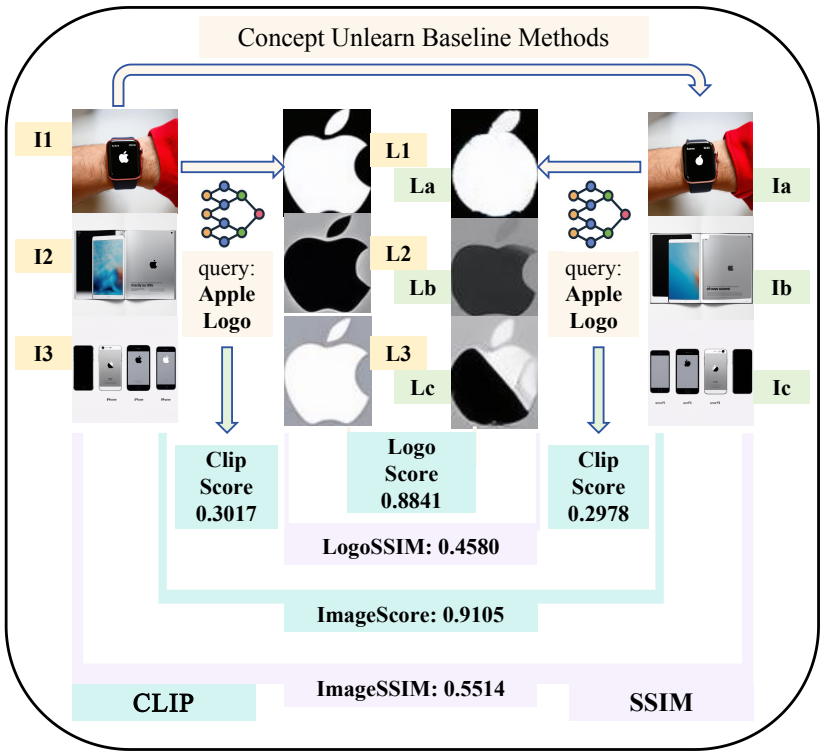

Figure 7: Original images and their unlearned counterparts are analyzed using metrics focused on logos and overall image similarity. CLIPScore calculates the CLIP similarity between detected logos and the text query ”apple logo.” LogoScore measures the CLIP similarity between logos before and after unlearning. LogoSSIM evaluates the SSIM similarity between logos before and after unlearning. ImageScore assesses the CLIP similarity for the entire image before and after unlearning, while ImageSSIM evaluates the SSIM similarity for the entire image. Light purple represents the SSIM score, while light blue represents the CLIP score.

E PROMPTS

To provide transparency in the creation of LU-500, we include the agent prompts utilized for generating all associated prompts. These were crafted using OpenAI's GPT-4o model, as illustrated in Figure 8. Furthermore, we detail the methodology behind the establishment of ProLU, which incorporates the functionalities of the Remover, Reflector, and Checker modules. This process is thoroughly explained in Figure 9.

Algorithm 1 The CLIP and SSIM similarities are first calculated for each individual prompt by comparing the images before and after unlearning. Then, the results are averaged at both the prompt level and the company level.

Require: Datasets D_i, $i = 1, 2$; Unlearning methods M_{il}; Companies C_{ilj}; Prompts P_{iljk}
Ensure: Metrics ME_{il} for evaluating unlearning methods

for each dataset D_i **do**
  for each unlearning method M_{il} **do**
    for each company C_{ilj}, $j = 1$ to 500 **do**
      for each prompt P_{iljk}, $k = 1$ to 10 **do**
        {Generate images before and after unlearning}
        Generate image I_ori_{iljk} using Stable Diffusion before unlearning
        Apply unlearning method M_{il} to generate I_un_{iljk}

        {Metric 1: Text-Logo Alignment (Local View)}
        Use OWLv2 with C_{ilj} logo as text query on I_ori_{iljk} and I_un_{iljk}
        Extract top confidence scores corresponding boxes containing logos as L_ori_{iljk} and L_un_{iljk}
        Extract CLIP logo features $F\_L\_ori_{iljk}$ and $F\_L\_un_{iljk}$
        Extract CLIP text features T_{ilj} with text query C_{ilj} logo
        Compute cosine similarity s_0, s_1 between $F\_L\_ori_{iljk}$, $F\_L\_un_{iljk}$ and T_{ilj}
        $ME1_{iljk} = s_0$ before unlearn or $= s_1$ after unlearn

        {Metric 2: Logo-Logo Alignment (Local View)}
        Extract CLIP features $F\_L\_ori_{iljk}$ and $F\_L\_un_{iljk}$
        Compute cosine similarity $ME2_{iljk}$ between $F\_L\_ori_{iljk}$ and $F\_L\_un_{iljk}$

        {Metric 3: Logo-Logo Alignment (Local View)}
        Compute SSIM score $ME3_{iljk}$ between L_ori_{iljk} and L_un_{iljk}

        {Metric 4: Image-Image Alignment (Global View)}
        Extract CLIP features $F\_I\_ori_{iljk}$ and $F\_I\_un_{iljk}$
        Compute cosine similarity $ME4_{iljk}$ between $F\_I\_ori_{iljk}$ and $F\_I\_un_{iljk}$

        {Metric 5: Image-Image Alignment (Global View)}
        Compute SSIM score $ME5_{iljk}$ between I_ori_{iljk} and I_un_{iljk}

      end for

      {Average metrics over k}
      Compute $ME1_{ilj} = \frac{1}{10}\sum_{k=1}^{10} ME1_{iljk}$
      Compute $ME2_{ilj} = \frac{1}{10}\sum_{k=1}^{10} ME2_{iljk}$
      Compute $ME3_{ilj} = \frac{1}{10}\sum_{k=1}^{10} ME3_{iljk}$
      Compute $ME4_{ilj} = \frac{1}{10}\sum_{k=1}^{10} ME4_{iljk}$
      Compute $ME5_{ilj} = \frac{1}{10}\sum_{k=1}^{10} ME5_{iljk}$
    end for
    {Average metrics over j}
    Compute $ME1_{il} = \frac{1}{500}\sum_{j=1}^{500} ME1_{ilj}$
    Compute $ME2_{il} = \frac{1}{500}\sum_{j=1}^{500} ME2_{ilj}$
    Compute $ME3_{il} = \frac{1}{500}\sum_{j=1}^{500} ME3_{ilj}$
    Compute $ME4_{il} = \frac{1}{500}\sum_{j=1}^{500} ME4_{ilj}$
    Compute $ME5_{il} = \frac{1}{500}\sum_{j=1}^{500} ME5_{ilj}$
  end for
end for
return ME_{il} for all methods M_{il}

Table 5: CLIPScore, LogoScore and LogoSSIM focus on the perspective of logos extracted from local regions, with CLIPScore considering the relationship between text and image, and LogoScore and LogoSSIM focusing on the relationship between images. ImageScore and ImageSSIM, on the other hand, evaluate the overall background. CLIPScore, LogoScore and ImageScore calculate CLIP similarity, while LogoSSIM and ImageSSIM measure SSIM.

Metric Name	Text-Image	Image-Image	CLIP	SSIM	Local	Global
CLIPScore	✓		✓		✓	
LogoScore		✓	✓		✓	
LogoSSIM		✓		✓	✓	
ImageScore		✓	✓			✓
ImageSSIM		✓		✓		✓

The Remover is used to eliminate elements related to the company logo from the original prompt while keeping other parts as consistent as possible. The Reflector evaluates whether the Remover has successfully completed its task and provides further optimized prompts. The Checker performs a final review to ensure that the final prompt does not contain any company logo; if any logo-related elements remain, they are directly removed.

For additional context, we showcase representative examples of how prompts are modified and adapted, as highlighted in Tables 6 to 10.

Explicit

Generate ten text-to-image prompts for {company}, ensuring each prompt explicitly includes the {company}'s logo so that the resulting images will feature the logo prominently. Consider emphasizing the company's office location, storefronts, advertisements, products, etc.
For example, for Apple, you can design the following prompts:
1. A photo of a closed MacBook with an Apple logo.
2. A photo of an iPad with an Apple logo.
3. A photo of an iPhone with an Apple logo.
4. A photo of an Apple store with an Apple logo in front of it.
Requirements: Generate 10 prompts directly in English, format them as follows:
1. prompt
2. prompt...
Keep the prompts simple, without complex scenes.

Implicit

Please generate ten text-to-image prompts for {company} that will result in images featuring the {company}'s logo prominently. Consider emphasizing the company's founder, famous products, competitors, company's products, stores, advertisements, office buildings, employees, or website, so that the logo would naturally appear in the image. For example, for Apple, you might design the following prompts:
1. A photo of a closed MacBook.
2. A photo of an iPad in use.
3. A photo of the back of an iPhone.
4. A photo of an Apple store entrance.
Make the prompt as detailed as possible, but not overly lengthy. It may include complex scenes except for the company logo. You can add other scenes to add complexity in prompts. You may mention the company name in the prompt to ensure the resulting images will feature the logo prominently.
Requirements: Generate 10 prompts directly in English without additional contents, formatted as:
1. prompt
2. prompt
...

Figure 8: The agent prompts for generating LU-500 was crafted using OpenAI's GPT-4o model.

F More Results

Lastly, we provide supplementary visual results to illustrate the unlearning performance of our method in greater depth. These results are presented across Figure 10 and Figure 11, offering further validation of our approach.

Including the images shown in the main text, we randomly selected one company for each letter of the alphabet from all companies in LU-500 (excluding 'Y' as no company starts with that letter). The selected companies are: Apple, Boeing, Coca-Cola, DELL, EXXON MOBIL, FedEx, Goldman-Sachs, HP, Intel, Johnson, KIA, L'Oreal, Mercedes-Benz, Nike, Oracle, Pfizer, Qualcomm, Renault, Starbucks, Tesla, Uber, Volvo, Walt Disney, Xiaomi, and Zurich Insurance Group.

Table 6: The performance of ProLU on selected companies is demonstrated as follows: $P0$ represents the original prompt, $P1$ is the prompt processed by the Remover, $P1*$ is the prompt refined by the Reflector, and $P2$ is the prompt finalized by the Checker. Companies include Apple, Boeing, Coca-Cola, DELL, EXXON MOBIL.

P0	A close-up photograph of an Apple Watch on a wrist, with the screen active and the iconic Apple logo on display beneath the strap.
P1	A close-up photograph of an Apple Watch on a wrist, with the screen active and a distinctive design on display beneath the strap.
P1*	A close-up photograph of an Apple Watch on a wrist, with the screen active and an interesting pattern on display beneath the strap.
P2	A close-up photograph of an Apple Watch on a wrist, with the screen active and an interesting pattern on display beneath the strap.
P0	A sleek Boeing aircraft parked on an airport runway with the Boeing logo clearly visible on the tail and fuselage, surrounded by a bustling airport scene.
P1	A sleek aircraft parked on an airport runway with a clear logo visible on the tail and fuselage, surrounded by a bustling airport scene.
P1*	A sleek aircraft parked on an airport runway, surrounded by a bustling airport scene.
P2	A sleek aircraft parked on an airport runway, surrounded by a bustling airport scene.
P0	A vintage advertisement from the 1950s showcasing the iconic Coca-Cola bottle, featuring the classic Coca-Cola logo surrounded by nostalgic imagery and old-style graphics.
P1	A vintage advertisement from the 1950s showcasing the iconic soda bottle, featuring a classic logo surrounded by nostalgic imagery and old-style graphics.
P1*	A vintage advertisement from the 1950s showcasing an iconic soda bottle, featuring retro branding surrounded by nostalgic imagery and old-style graphics.
P2	A vintage advertisement from the 1950s showcasing an iconic soda bottle, featuring retro branding surrounded by nostalgic imagery and old-style graphics.
P0	A vibrant scene of a bustling Dell Technologies store entrance, with customers exploring the latest laptops and desktops on display, prominently featuring the Dell Technologies logo on signage and digital displays.
P1	A vibrant scene of a bustling technology store entrance, with customers exploring the latest laptops and desktops on display, prominently featuring logos on signage and digital displays.
P1*	A vibrant scene of a bustling technology store entrance, with customers exploring the latest laptops and desktops on display, showcasing various brands on signage and digital displays.
P2	A vibrant scene of a bustling technology store entrance, with customers exploring the latest laptops and desktops on display, showcasing various brands on signage and digital displays.
P0	A detailed image of an Exxon Mobil gas station at dusk, with the Exxon Mobil logo brightly illuminated on the station canopy and fuel pumps, against a backdrop of a busy city street.
P1	A detailed image of a gas station at dusk, with the station canopy and fuel pumps brightly illuminated, against a backdrop of a busy city street.
P1*	A detailed image of a gas station at dusk, with the station canopy and fuel pumps brightly illuminated, against a backdrop of a busy city street.
P2	A detailed image of a gas station at dusk, with the station canopy and fuel pumps brightly illuminated, against a backdrop of a busy city street.

Table 7: The performance of ProLU on selected companies is demonstrated as follows: $P0$ represents the original prompt, $P1$ is the prompt processed by the Remover, $P1*$ is the prompt refined by the Reflector, and $P2$ is the prompt finalized by the Checker. Companies include FedEx, Goldman-Sachs, HP, Intel, Johnson & Johnson.

P0	A photo of a bustling FedEx delivery truck unloading boxes on a busy city street, with the FedEx logo clearly visible on the side of the vehicle.
P1	A photo of a bustling delivery truck unloading boxes on a busy city street, with a logo clearly visible on the side of the vehicle.
P1*	A photo of a bustling delivery truck unloading boxes on a busy city street.
P2	A photo of a bustling delivery truck unloading boxes on a busy city street.
P0	A high-angle shot of a modern financial trading floor inside Goldman Sachs, with large screens displaying stock prices and the Goldman Sachs logo visible on a wall banner.
P1	A high-angle shot of a modern financial trading floor, with large screens displaying stock prices and a logo visible on a wall banner.
P1*	A high-angle shot of a modern financial trading floor, with large screens displaying stock prices and a corporate logo visible on a wall banner.
P2	A high-angle shot of a modern financial trading floor, with large screens displaying stock prices and a corporate logo visible on a wall banner.
P0	A close-up photo of an HP laptop with its screen open, displaying a vibrant desktop, in a sleek modern office setting, showcasing the HP logo on the laptop lid.
P1	A close-up photo of a laptop with its screen open, displaying a vibrant desktop, in a sleek modern office setting.
P1*	A close-up photo of a laptop with its screen open, displaying a vibrant desktop, in a sleek modern office setting.
P2	A close-up photo of a laptop with its screen open, displaying a vibrant desktop, in a sleek modern office setting.
P0	A high-tech office desktop scene with an Intel processor box next to a sleek computer tower, showcasing the Intel logo prominently on the processor box and the computer case.
P1	A high-tech office desktop scene with a processor box next to a sleek computer tower, featuring the processor box and the computer case prominently.
P1*	A high-tech office desktop scene with a generic processor box next to a sleek computer tower, featuring the processor box and the computer case prominently.
P2	A high-tech office desktop scene with a generic processor box next to a sleek computer tower, featuring the processor box and the computer case prominently.
P0	A store shelf filled with a variety of Johnson & Johnson over-the-counter medications, ensuring the logo is distinct and visible on multiple packages like Tylenol and Zyrtec.
P1	A store shelf filled with a variety of over-the-counter medications, ensuring distinct and visible labels on multiple packages like Tylenol and Zyrtec.
P1*	A store shelf filled with a variety of over-the-counter medications, ensuring distinct and visible labels on multiple packages like Tylenol and Zyrtec, without any specific branding.
P2	A store shelf filled with a variety of over-the-counter medications, ensuring distinct and visible labels on multiple packages like Tylenol and Zyrtec, without any specific branding.

Table 8: The performance of ProLU on selected companies is demonstrated as follows: $P0$ represents the original prompt, $P1$ is the prompt processed by the Remover, $P1*$ is the prompt refined by the Reflector, and $P2$ is the prompt finalized by the Checker. Companies include KIA, L'Oreal, Mercedes-Benz, Nike, Oracle.

P0	An advertising billboard featuring a KIA Sportage in a scenic countryside, with the KIA logo large and clear on the car.
P1	An advertising billboard featuring a KIA Sportage in a scenic countryside, with the car's branding large and clear on the car.
P1*	An advertising billboard featuring a KIA Sportage in a scenic countryside, showcasing the car prominently.
P2	An advertising billboard featuring a KIA Sportage in a scenic countryside, showcasing the car prominently.
P0	A close-up of various L'ORÉAL skincare products arranged on a bathroom counter, showcasing the packaging design with the L'ORÉAL logo clearly visible.
P1	A close-up of various skincare products arranged on a bathroom counter, showcasing the packaging design.
P1*	A close-up of various skincare products arranged on a bathroom counter, showcasing the packaging design without any logos.
P2	A close-up of various skincare products arranged on a bathroom counter, showcasing the packaging design without any logos.
P0	A bustling Mercedes-Benz dealership showroom filled with a variety of the latest car models, with the Mercedes-Benz logo prominently displayed on signs and promotional materials.
P1	A bustling dealership showroom filled with a variety of the latest car models, with logos prominently displayed on signs and promotional materials.
P1*	A bustling dealership showroom filled with a variety of the latest car models, with various logos prominently displayed on signs and promotional materials.
P2	A bustling dealership showroom filled with a variety of the latest car models, with various logos prominently displayed on signs and promotional materials.
P0	An artistic rendering of the Nike headquarters in Beaverton, Oregon, featuring the main entrance with the large Nike logo prominently displayed on the building facade.
P1	An artistic rendering of the Nike headquarters in Beaverton, Oregon, featuring the main entrance with the large Nike logo prominently displayed on the building facade.
P1*	An artistic rendering of the headquarters in Beaverton, Oregon, featuring the main entrance with the building facade prominently displayed.
P2	An artistic rendering of the headquarters in Beaverton, Oregon, featuring the main entrance with the building facade prominently displayed.
P0	A digital advertisement showcasing ORACLE's cloud infrastructure, with a focus on the sleek interface of Oracle Cloud and the logo in the top right corner.
P1	A digital advertisement showcasing advanced cloud infrastructure, with a focus on the sleek interface and a symbol in the top right corner.
P1*	A digital advertisement showcasing advanced cloud infrastructure, with a focus on the sleek interface and an emblem in the top right corner.
P2	A digital advertisement showcasing advanced cloud infrastructure, with a focus on the sleek interface and an emblem in the top right corner.

Table 9: The performance of ProLU on selected companies is demonstrated as follows: $P0$ represents the original prompt, $P1$ is the prompt processed by the Remover, $P1*$ is the prompt refined by the Reflector, and $P2$ is the prompt finalized by the Checker. Companies include Pfizer, Qualcomm, Renault, Starbucks, Tesla.

P0	An advertisement for the Pfizer COVID-19 vaccine featuring a close-up of a nurse holding a vaccine vial prominently displaying the Pfizer logo.
P1	An advertisement for the COVID-19 vaccine featuring a close-up of a nurse holding a vaccine vial prominently displayed.
P1*	An advertisement for the COVID-19 vaccine featuring a close-up of a nurse holding a vaccine vial.
P2	An advertisement for the COVID-19 vaccine featuring a close-up of a nurse holding a vaccine vial.
P0	An illustration of QUALCOMM's latest Snapdragon processor being held in the palm of a hand, with the QUALCOMM logo visible on the processor and various futuristic gadgets floating around it.
P1	An illustration of the latest Snapdragon processor being held in the palm of a hand, with the processor and various futuristic gadgets floating around it.
P1*	An illustration of the latest Snapdragon processor being held in the palm of a hand, with the processor visible and various futuristic gadgets floating around it.
P2	An illustration of the latest Snapdragon processor being held in the palm of a hand, with the processor visible and various futuristic gadgets floating around it.
P0	A dynamic shot of a sleek Renault Clio parked in front of a modern cityscape, with its distinctive Renault emblem visible on the grille.
P1	A dynamic shot of a sleek Renault Clio parked in front of a modern cityscape, with its distinctive grille visible.
P1*	A dynamic shot of a sleek Renault Clio parked in front of a modern cityscape, showcasing its distinctive front design.
P2	A dynamic shot of a sleek Renault Clio parked in front of a modern cityscape, showcasing its distinctive front design.
P0	A promotional banner for a new Starbucks Frappuccino, featuring the drink with a swirl of whipped cream on top, and the Starbucks logo clearly shown on the drink cup.
P1	A promotional banner for a new Frappuccino, featuring the drink with a swirl of whipped cream on top, and the logo clearly shown on the drink cup.
P1*	A promotional banner for a new Frappuccino, featuring the drink with a swirl of whipped cream on top, and the brand logo clearly shown on the drink cup.
P2	A promotional banner for a new Frappuccino, featuring the drink with a swirl of whipped cream on top, and the brand logo clearly shown on the drink cup.
P0	A futuristic showroom showcasing the latest Tesla electric vehicles, with a massive illuminated Tesla logo on the wall behind the cars.
P1	A futuristic showroom showcasing the latest electric vehicles, with a massive illuminated logo on the wall behind the cars.
P1*	A futuristic showroom showcasing the latest electric vehicles, with a massive illuminated sign on the wall behind the cars.
P2	A futuristic showroom showcasing the latest electric vehicles, with a massive illuminated sign on the wall behind the cars.

Table 10: The performance of ProLU on selected companies is demonstrated as follows: $P0$ represents the original prompt, $P1$ is the prompt processed by the Remover, $P1*$ is the prompt refined by the Reflector, and $P2$ is the prompt finalized by the Checker. Companies include Uber, Volvo, Walt Disney, Xiaomi.

P0	A scene inside a modern Uber office, showcasing employees collaborating in a meeting room with the Uber logo displayed prominently on a wall.
P1	A scene inside a modern office, showcasing employees collaborating in a meeting room with a company logo displayed prominently on a wall.
P1*	A scene inside a modern office, showcasing employees collaborating in a meeting room.
P2	A scene inside a modern office, showcasing employees collaborating in a meeting room.
P0	A panoramic view of a VOLVO factory floor with workers assembling different car parts, emphasizing the scale of production and the company's commitment to quality, with the logo visible on various tools and safety equipment.
P1	A panoramic view of a car factory floor with workers assembling different car parts, emphasizing the scale of production and the company's commitment to quality, with the logo visible on various tools and safety equipment.
P1*	A panoramic view of a car factory floor with workers assembling different car parts, emphasizing the scale of production and the company's commitment to quality, with branding visible on various tools and safety equipment.
P2	A panoramic view of a car factory floor with workers assembling different car parts, emphasizing the scale of production and the company's commitment to quality, with branding visible on various tools and safety equipment.
P0	A street view of a Disney Store with a captivating window display of Disney character merchandise, the Disney logo prominently placed above the entrance.
P1	A street view of a children's store with a captivating window display of character merchandise, a logo prominently placed above the entrance.
P1*	A street view of a children's store with a captivating window display of character merchandise, a colorful logo prominently placed above the entrance.
P2	A street view of a children's store with a captivating window display of character merchandise, a colorful logo prominently placed above the entrance.
P0	An interior shot of a bustling Xiaomi store, with a group of people examining the latest Xiaomi ecosystem products on display, prominently featuring a large Xiaomi logo on the back wall.
P1	An interior shot of a bustling electronics store, with a group of people examining the latest technology ecosystem products on display, prominently featuring a large logo on the back wall.
P1*	An interior shot of a bustling electronics store, with a group of people examining the latest technology ecosystem products on display, prominently featuring a large brand logo on the back wall.
P2	An interior shot of a bustling electronics store, with a group of people examining the latest technology ecosystem products on display, prominently featuring a large brand logo on the back wall.

Remover	Reflector	Checker
I will provide you with two pieces of text: a prompt and a concept. The prompt contains a descriptive text, and the concept represents an idea. I am working on a task to generate an image from text, where the text is the prompt, and the prompt contains the concept. However, I don't want the generated image to include the concept. Please help me modify the prompt so that the concept is either removed or replaced while making the least possible change to the meaning of the prompt. The original prompt is: {prompt}, and the concept is: {concept}. Note: you must directly output the optimized prompt without any other text.	The modified prompt is: '{prompt_1}'. The concept to remove is: '{concept}'. Does the modified prompt still contain the concept? If it does, please adjust the prompt to ensure the concept is either removed or replaced while keeping the original meaning intact. If it does not, do not modify the prompt. Note: you must directly output the optimized prompt without any other text.	The concept '{concept}' still exists in the modified prompt. You must adjust the prompt to ensure the concept is either removed or replaced while keeping the original meaning intact. Note: you must directly output the optimized prompt without any other text.

Figure 9: The Remover, Reflector, and Checker prompts in ProLU. The Remover is used to eliminate elements related to the company logo from the original prompt while keeping other parts as consistent as possible. The Reflector evaluates whether the Remover has successfully completed its task and provides further optimized prompts. The Checker performs a final review to ensure that the final prompt does not contain any company logo; if any logo-related elements remain, they are directly removed.

Figure 10: More visual results over Boeing, DELL, EXXON MOBIL, FedEx, Goldman-Sachs, HP, Intel, Johnson, KIA.

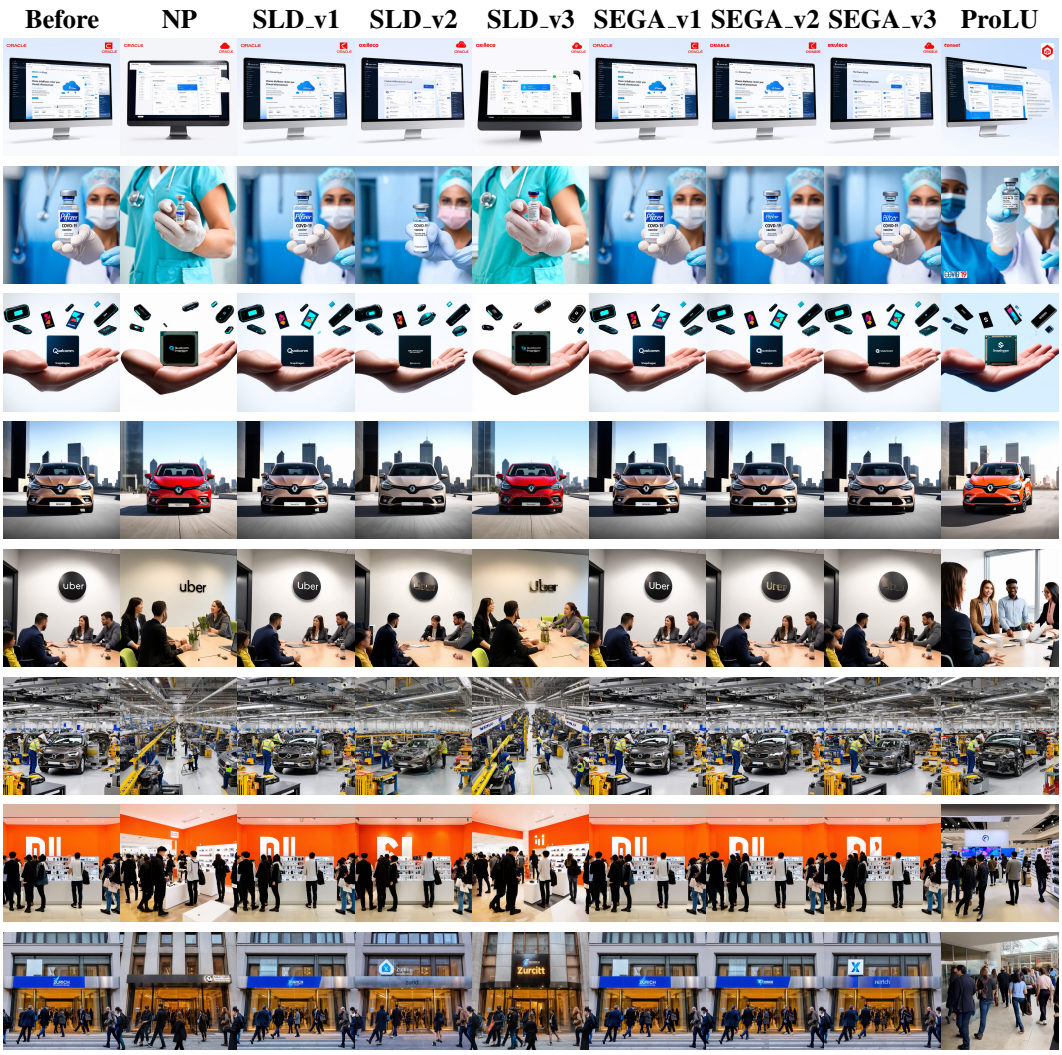

Figure 11: More visual results on Oracle, Pfizer, Qualcomm, Renault, Uber, Volvo, Walt Disney, Xiaomi, and Zurich Insurance Group.