Hiding in Plain Sight: Camouflaging Real-world Objects

Dong-Yi Wu and Tong-Yee Lee, *Senior Member, IEEE*

***Abstract*—Camouflage, a survival strategy perfected by nature, enables organisms to evade detection by blending seamlessly into their surroundings across multiple viewpoints. Replicating this remarkable ability in artificial settings remains a formidable challenge. While recent computational methods have advanced 2D camouflage, extending concealment into 3D is far more difficult: a single texture must reconcile drastically varying backgrounds, making effective camouflage across all views highly challenging. Prior approaches attempt to address this by designing multi-view textures, but their appearance representations are too simplistic to cope with strongly conflicting backgrounds and fail to account for physical light transport, resulting in breakdowns under realistic illumination. We introduce a new paradigm that formulates 3D camouflage as a multiview inverse rendering problem. Instead of treating concealment as texture synthesis, we directly optimize appearance within a physically based rendering framework, explicitly modeling reflections, shadows, refractions, and other complex light interactions. Our approach expands the solution space by combining BSDF-based material representation with anamorphic surface displacement, enabling camouflage that adapts naturally to both viewpoint and illumination changes. Through tailored loss functions and optimization strategies, our method produces results that are not only perceptually convincing but also physically consistent, marking a significant step toward practical, real-world 3D camouflage. Experiments demonstrate that our physically grounded formulation achieves robust concealment across diverse viewpoints and lighting scenarios, substantially outperforming prior image-based methods. Project Website: https://dongee-w.github.io/camouflage_website/**

***Index Terms*—3D camouflage, inverse rendering**

I. Introduction

The challenge of camouflage is deeply rooted in nature. Many insects survive by seamlessly blending into their surroundings, remaining hidden from predators across all viewpoints. This remarkable ability is the result of millions of years of evolution, shaping unique appearances with intricate patterns and carefully tuned colors. The ultimate purpose of camouflage in nature is simple: to reduce the chance of being detected. Inspired by biology, humans have long pursued artificial camouflage, from disruptive military patterns on vehicles to architectural designs that help infrastructure blend into its surroundings.

This challenging problem has recently fueled a wave of computational approaches [1]–[7]. Most existing work focuses on 2D camouflage, where the objective is to synthesize an image that blends foreground objects into background scenes. In contrast, a smaller line of research [6], [7] explores the more difficult task of 3D camouflage, which requires generating full object textures rather than a single 2D image.

The key challenge in 3D camouflage is maintaining concealment across multiple viewpoints, since each viewpoint can reveal a different background. For instance, a utility box on a street corner may appear against a brightly lit shop window from one direction, but against a dark alley wall from another. Such drastic shifts in background with even small changes in viewpoint make it impossible for a single texture to provide perfect concealment in all directions. Instead, effective 3D camouflage requires balancing competing constraints across views to achieve a convincing illusion.

Existing methods [6], [7] address this by treating camouflage as a texture synthesis problem, designing clever surface patterns that work reasonably well from multiple viewpoints. However, this approach has fundamental physical limits: a single albedo map cannot simultaneously satisfy conflicting conditions, such as bright and dark backgrounds. To increase multiview capacity, we instead exploit view-dependent appearance through a learnable physically-based material model (BSDF) and fine-scale geometric microstructure, allowing the surface to accommodate highly inconsistent constraints imposed by different viewpoints.

Another critical shortcoming of prior work is the neglect of realistic lighting, which leads to camouflage failures in practical settings (see Fig. 1). In (a), a backpack camouflaged using GANmouflage appears convincing at first glance. However, when its texture is extracted and rendered under physically based conditions in (b), clear artifacts emerge: the surface exhibits color drift, appearing too bright on light-facing regions and too dark in shadowed areas. The backpack's silhouette becomes obvious because illumination differences across the surface are not accounted for. In addition, self-cast shadows—such as those from the straps falling onto the backpack—remain unmodeled, further breaking the camouflage illusion.

In this work, we take a fundamentally different approach by formulating camouflage as a *multiview inverse rendering task*. By jointly optimizing an object's appearance under known cameras and lighting conditions, we explicitly model complex light transport phenomena, including reflections, refractions, shadows, and secondary light interactions. This physically grounded formulation not only anchors camouflage in real-world conditions but also enables view- and lighting-aware effects far beyond pixel-level texture manipulation. We expand the appearance space by incorporating the physically-based

Dong-Yi Wu is with National Cheng Kung University. E-mail: cutechubbit@gmail.com

Tong-Yee Lee is with National Cheng Kung University. E-mail: tonylee@mail.ncku.edu.tw.

Corresponding author: Tong-Yee Lee.

material together with *anamorphic surface displacement*. By explicitly modeling physically-based material, i.e., diffuse albedo, metallic, roughness, transmittance, and geometric variation, we greatly enhance the expressive power of the surface. This enables camouflage that adapts naturally to changes in illumination and viewpoint, producing more robust, physically consistent, and perceptually convincing concealment. In summary, our main contributions are listed below:

- We reformulate 3D camouflage as an inverse rendering problem under realistic, physically-based lighting conditions.
- We propose a versatile appearance representation, combining BSDF-based material modeling with anamorphic displacement, to enable flexible and high-quality camouflage synthesis.
- We design tailored loss functions and optimization strategies to mitigate the ill-posed nature of the camouflage problem, achieving results that are both physically consistent and perceptually convincing.

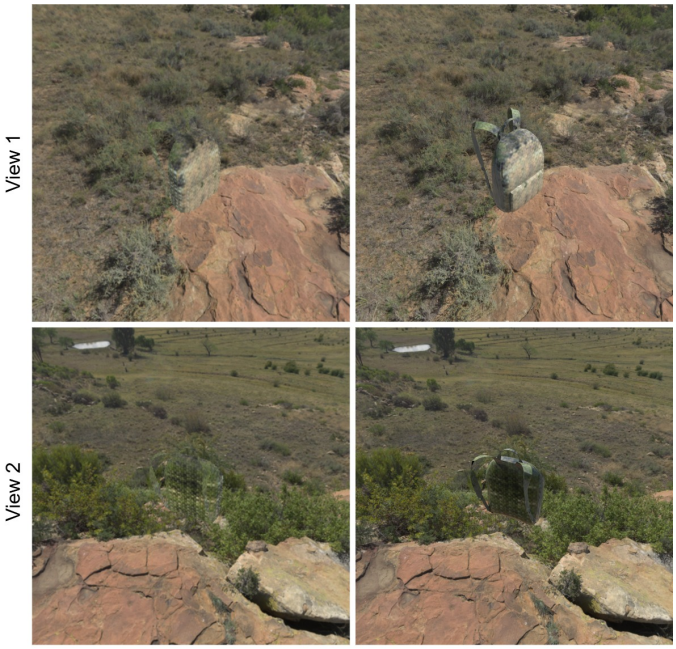


Fig. 1. Limitation of existing approaches. (a) Camouflage synthesized in image space for a backpack by GANmouflage [7]. (b) The same albedo texture rendered with a physically-based renderer. This reveals degraded camouflage effectiveness under realistic lighting.

II. Related Work

A. 2D Camouflage

Early work by Reynolds [1] explored camouflage in the setting of artificial life simulation, where visual prey patterns were evolved using a genetic algorithm and a human predator interactively detected them, effectively simulating an evolutionary arms race between concealment and perception.

Camouflage has also been studied as a form of perceptual art. Chu et al. [2] introduced camouflage images inspired by human perception theory, where hidden figures are initially difficult to detect but gradually emerge through feature and conjunction search. However, their reliance on handcrafted features often produced objects that were either overly salient, completely invisible, or poorly blended.

To overcome these limitations, subsequent works [3]–[5] leveraged deep neural networks for more balanced concealment and recognizability with improved naturalness. Zhang et al. [3] proposed a neural style transfer framework with an attention-aware camouflage loss and naturalness regularization, achieving more seamless results than handcrafted or texture-based methods. LCG-Net [4] generated camouflage in a single forward pass by fusing foreground and background features, producing faster and more natural results, especially in multi-appearance regions. More recently, Camouflage Anything [5] leveraged CLIP [8] with controlled out-painting and representation engineering to synthesize effective camouflage patterns.

Although conceptually related, 2D and 3D camouflage face fundamentally different challenges: in 2D, the objective is to alter images so that hidden objects are neither too easy nor too hard to detect, while in 3D, the goal is to design surface appearance under strict multiview constraints to maximize concealment in the environments. In our paper, we focus on 3D camouflage.

B. 3D Camouflage

Different from 2D, research on 3D camouflage seeks to find the optimal surface texture that can conceal 3D objects to blend seamlessly with the surrounding environment. Owens et al. [6] investigated multi-view camouflage by optimizing surface textures from photographs captured at different viewpoints. More recently, Guo et al. [7] proposed **GANmouflage**, which leverages texture fields and adversarial learning to generate object-specific camouflage textures effective across varied perspectives These methods are closely related to our setting, since they aim to synthesize surface appearances that visually blend objects into their surroundings. However, they mainly treat camouflage as a texture synthesis problem and rely on simplified appearance models, which limits their ability to handle realistic illumination, shadows, reflections, and view-dependent material effects. In contrast, we formulate 3D camouflage as a physically-based inverse rendering problem, explicitly incorporating principles from physically based rendering (PBR) [9] and bidirectional reflectance distribution function (BRDF) modeling [10], enabling more realistic camouflage synthesis.

A separate line of work studies physical adversarial camouflage, which aims to fool machine perception systems rather than achieve perceptually faithful visual concealment. For example, **ACTIVE** [11] optimizes transferable vehicle textures using triplanar mapping, a neural texture renderer, and detector-oriented losses. Follow-up methods such as RAUCA [12] and PhyCamo [13] further improve robustness under physical conditions. While related to 3D camouflage, these methods are primarily evaluated by detector evasion rather than human-perceptual concealment, and the resulting adversarial patterns may remain visually conspicuous to human observers.

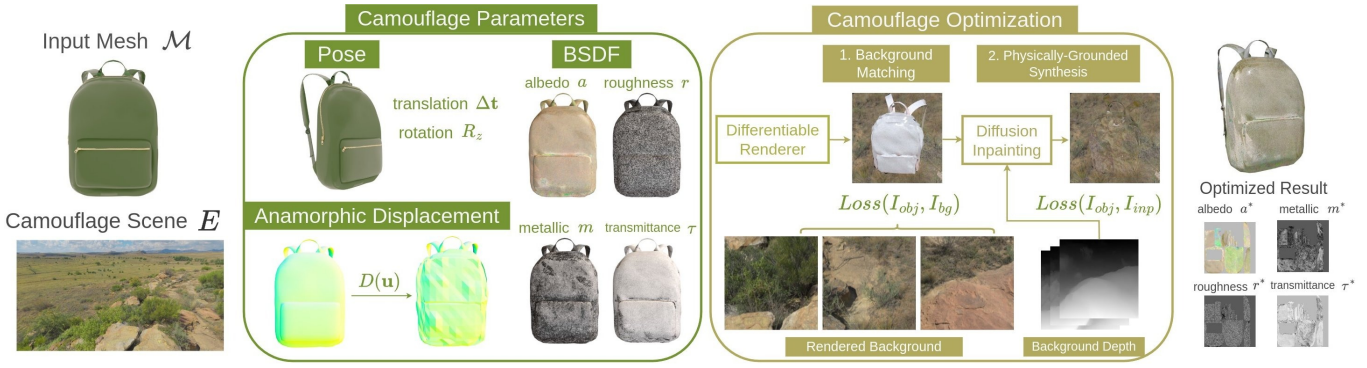


Fig. 2. System overview. Input: an arbitrary 3D mesh M and target scene S. Our goal is to seamlessly conceal M within S at a specified location. The object's appearance is modeled via pose, anamorphic displacement D, diffuse albedo $\mathbf{a}$, metallic $\mathbf{m}$, roughness r, and transmittance τ. Our optimization pipeline jointly refines these parameters to produce high-quality, physically accurate camouflage.

C. Anamorphic and View-Dependent Appearance

A related line of work explores *anamorphic effects*, in which the perceived appearance varies with the viewing angle. Notable examples include lenticular lens surfaces [14] and printable materials with view-dependent properties [15]. These approaches highlight the role of geometry manipulation in producing multi-view effects, a principle directly relevant to our 3D camouflage task. Inspired by this idea, we incorporate height-field-based designs to strengthen the multi-view camouflage performance of concealed objects.

D. Image Inpainting with Generative Models

Image inpainting is the task of filling missing or masked-out regions of an image with visually coherent and semantically plausible content. This problem is highly relevant to our camouflage synthesis. In 2D, diffusion-based approaches, such as RePaint [16], LaMa [17], and ZITS [18] achieve high-quality image completion. More recently, commercial and community tools such as Flux Fill [19] and SDXL [20] have demonstrated strong perceptual inpainting performance in practice. In this work, we leverage advances in image inpainting and integrate them into our approach for improved perceptual quality.

III. Method

A. Overview

Our system pipeline is illustrated in Fig. 2. The input consists of an arbitrary 3D mesh M to be camouflaged and a target scene E. The scene can be represented either as a full 3D environment with meshes and lighting, or as a set of high-dynamic-range (HDR) panoramic images captured with a 360-degree camera. The output is a textured 3D model optimized for camouflage.

The goal is to seamlessly conceal M within E at a specified location, as observed from a set of viewpoints V, by optimizing the appearance of M. We formulate this as an optimization over a set of appearance parameters Θ_{app}, aggregating per-view photometric losses $\mathcal{L}_{\text{photo}}$:

$$\Theta^*_{app} = \arg\min_{\Theta} \sum_{v \in V} \mathcal{L}_{\text{photo}}\big(\Theta_{app}, M, E, v\big), \qquad (1)$$

where $\mathcal{L}_{\text{photo}}$ measures the photometric camouflage quality for each view and V denotes the set of all sampled viewpoints, and $v \in V$ represents an individual viewpoint.

To achieve physically accurate and flexible appearance modeling, we parameterize the object's appearance using differentiable variables that can be optimized via backpropagation. These include the pose (translation and rotation), anamorphic displacement, diffuse albedo $\mathbf{a}$, metallic $\mathbf{m}$, roughness $\mathbf{r}$, and transmittance τ. Together, these parameters describe the bidirectional scattering properties of the surface.

Unlike prior methods [6], [7] that ignore lighting, we explicitly model scene illumination to closely approximate the real lighting conditions of the target environment. This allows the camouflaged object to integrate naturally, faithfully reproducing shadows, highlights, and interreflections.

Given this scene representation, we employ a physically-based differentiable renderer and optimize the material parameters via gradient-based backpropagation. Differentiable rendering enables the rasterization process to be differentiable, allowing gradients to flow from the image domain back to the 3D scene parameters Θ_{app}.

Since there is no standard ground truth for camouflage, we design a two-stage optimization strategy. In the first stage, *Background Matching*, we minimize a pixel-wise loss between the object and its surrounding background. This encourages a physically accurate, 3D-consistent appearance that roughly matches the environment. However, pixel-level losses alone cannot capture higher-level perceptual cues. To address this, the second stage, called *Physically-Grounded Synthesis*, leverages a pretrained generative inpainting model (SDXL model [21]) to produce realistic, perceptually convincing camouflage conditioned on the first stage result. This two-stage approach ensures that the final camouflage is both physically consistent and visually indistinguishable from the background.

B. Our Camouflage Appearance Model

To generate effective camouflage, we consider three critical aspects of an object's appearance: its spatial pose within the local neighborhood, geometric displacement of the surface, and material properties. Our parameterization, therefore, consists

of three key components: (i) *pose*, (ii) *anamorphic displacement*, and (iii) a *spatially varying Bidirectional Scattering Distribution Function (BSDF)*. These individual parameters alone cannot increase the capacity of the appearance too much. However, when combine together, the synergistic effect allows us to design a much more sophisticated appearance.

1) Pose Optimization: In nature and practice, effective camouflage often relies not only on material properties but also on carefully chosen object poses. For example, moths align their bodies with bark grain, caterpillars adopt twig-like postures and Grasshoppers adjust body angles to better match backgrounds [22]. Or in real life, people often use small rotations can hide an ugly pot or electrical cord from view. A slight change in orientation can transform an object from nearly invisible to highly conspicuous. Inspired by these observations, we incorporate pose, i.e., translation offset and a rotation around the z-axis, into our optimization to enhance the concealment effectiveness of our method. Our goal is to find the best pose with respect to the loss from each viewpoint.

The original model matrix is defined as

$$T_{\text{orig}} = T_t \, R \, S, \tag{2}$$

where S is the scaling matrix, $R \in SO(3)$ is the given rotation, and

$$T_t = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is the homogeneous translation matrix corresponding to the fixed translation vector $t = (t_x, t_y, t_z)^\top$.

Our appearance model includes an optimizable translation offset $\Delta\mathbf{t} = (\Delta x, \Delta y, \Delta z)^\top$ and an optimizable rotation angle θ_z around the z-axis. Their corresponding matrices are

$$\mathbf{T}_{\Delta\mathbf{t}} = \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{R}_z(\theta_z) = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 & 0 \\ \sin\theta_z & \cos\theta_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The new (augmented) model matrix is then given by

$$\mathbf{T}_{\text{new}} = \mathbf{T}_{\Delta\mathbf{t}} \, T_t \, \mathbf{R}_z(\theta_z) \, R \, S. \tag{3}$$

Here, $\mathbf{T}_{\Delta\mathbf{t}}$ and $\mathbf{R}_z(\theta_z)$ are learnable parameters, while S, R, and T_t are fixed inputs. Since we only want the model to move in its local neighborhood, we clip the value of $\mathbf{T}_{\Delta\mathbf{t}}$ within 5% of the object's size. And we usually set $\Delta z = 0$ since we want the object to stay on the ground plane. We only consider the rotation around z-axis since it more resembles a real-world scenario of physically rotating when repositioning an object (e.g., moving furniture).

Fig. 3. An example of anamorphism through displacement, where the surface is segmented and shifted such that different images emerge depending on the viewer's position: (a) Mirror Cup & Saucer Art by Luycho [23] (b) Images from different views.

2) Anamorphic Displacement: Displacement or microstructure on the surface has long been utilized to achieve intriguing optical effects, as illustrated in 3. From anamorphic reliefs and lenticular ridges to mirror saucers and holographic foils, artists and designers exploit displaced or textured surfaces to produce images that transform with viewpoint. The underlying idea is that even subtle geometric modulation can redirect light in a controlled, angle-dependent way, effectively encoding multiple appearances into a single object. In our camouflage task, this same principle can be harnessed to engineer view-dependent visual patterns that conceal an object from multiple views.

Without displacement, the camouflage effect is severely limited, as illustrated in Fig. 4. On a flat surface (left), the result collapses into a smooth gradient transition rather than producing distinct, meaningful images from different viewpoints. Notably, the upper hemisphere shares the same target color across both views and thus requires no displacement (reflected in the smoother geometry of the right sphere's upper half). By contrast, the lower hemisphere depends on geometric modulation—surface undulations that redirect light—to generate the intended view-dependent image patterns.

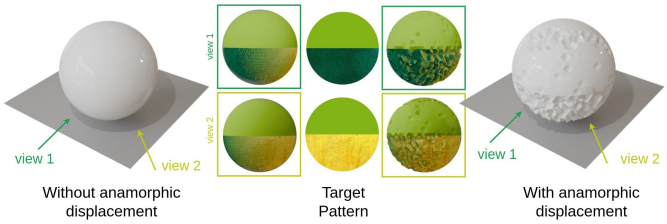


Fig. 4. Illustration of the effect of optimizing anamorphic displacement. The left sphere has no displacement, while the right sphere incorporates the optimized displacement. In both cases, the albedo maps are optimized to reproduce two target patterns (middle) as seen from slightly different viewpoints.

We introduce a *learnable anamorphic displacement map* $D : \Omega \to \mathbb{R}$, which perturbs the surface geometry at each texture coordinate $\mathbf{u} \in \Omega$. For a vertex $\mathbf{v}_i$ with surface normal $\mathbf{n}_i$, the displaced vertex $\mathbf{v}'_i$ is defined as

$$\mathbf{v}'_i = \mathbf{v}_i + D(\mathbf{u}_i) \, \mathbf{n}_i. \tag{4}$$

While this displacement provides flexibility for view-dependent appearance control, we observe that the optimization often introduces unnecessary perturbations. Excessive displacement not only distorts the geometry but also reduces

overall visual fidelity. To mitigate this, it is crucial to regularize the displacement magnitude and activate it only when beneficial.

We employ a simple yet effective regularization loss of the form

$$\mathcal{L}_{\text{disp}} = \mathbb{E}[\,|D(\mathbf{u})|\,], \tag{5}$$

which penalizes large deviations and encourages minimal, targeted surface adjustments.

Interestingly, beyond multiview consistency, the displacement map also disrupts otherwise regular edges (e.g., straight contours that are perceptually salient). This disruption reduces the detectability of artificial geometric boundaries, thereby enhancing the camouflage quality.

3) Spatially Varying BSDF: By optimizing not only albedo but also a compact set of BSDF parameters, namely metallic, roughness, and transmittance, we exploit the expressive power of physically based materials while keeping the inverse problem stable and interpretable. Albedo controls the diffuse surface color; metallic and roughness govern the strength and sharpness of environment reflection; and transmittance introduces refractive effects, allowing background features to be bent and blended according to Snell's law. Together, these parameters capture the key diffuse, reflective, and transmissive effects needed for visual concealment across viewpoints and illumination conditions, whereas albedo-only methods quickly appear flat and conspicuous. Although production BSDF models expose many additional parameters, such as the index of refraction, anisotropy, sheen, and clearcoat, we keep these parameters fixed in this work. Optimizing all of them would substantially increase the dimensionality of an already ill-posed inverse problem, while many of these parameters are either weakly identifiable from image-space losses or correspond to specialized material effects. We therefore focus on the dominant appearance controls most relevant to camouflage: albedo, roughness, metallic response, transmittance, and displacement. In particular, we use the index of refraction of $\eta = 1.5$ in our experiment, which is representative of common transparent or semi-transparent solids. Other material families can be modeled by selecting a different global value, such as $\eta \approx 1.33$ for water-like materials. Although η is differentiable and could in principle be optimized as a global parameter, its visual influence in our setting is secondary to the transmittance map: transmittance determines how much refractive light contributes, while η mainly controls the bending of that transmitted component. Since the optimized camouflage is generally not a fully transparent glass-like solution, changes in η have a weaker effect than changes in albedo, roughness, metallic response, transmittance, and displacement. We therefore keep η fixed to maintain stable and physically interpretable optimization.

We denote the bidirectional scattering distribution function (BSDF) of the surface as f_r, parameterized by spatially-varying material maps.

$$f_r(\mathbf{x}, \omega_i, \omega_o) = f_r\Big(\omega_i, \omega_o; \mathbf{a}(\mathbf{u}(\mathbf{x})), \mathbf{r}(\mathbf{u}(\mathbf{x})), \mathbf{m}(\mathbf{u}(\mathbf{x})), \boldsymbol{\tau}(\mathbf{u}(\mathbf{x}))\Big), \tag{6}$$

where $\mathbf{x}$ is the surface point, $u(\mathbf{x})$ maps $\mathbf{x}$ to UV coordinates, $\mathbf{a}$, $\mathbf{r}$, $\mathbf{m}$, τ are the spatially varying albedo, roughness, metallic, and transmittance maps.

When rendering, at each point $\mathbf{x}$, the outgoing radiance L_o in direction ω_o is given by the rendering equation:

$$L_o(\mathbf{x}, \omega_o) = \int_\Omega f_r(\mathbf{x}, \omega_i, \omega_o)\, L_i(\mathbf{x}, \omega_i)\, (\mathbf{n} \cdot \omega_i)\, d\omega_i, \tag{7}$$

where L_i is the incoming radiance, $\mathbf{n}$ is the surface normal, and ω_i, ω_o denote incident and outgoing directions.

Since f_r is differentiable with respect to the material parameters, the loss computed on the rendered image can be back-propagated to update the texture maps via gradient descent. This enables us to directly optimize the appearance parameters $\mathbf{a}, \mathbf{r}, \mathbf{m}, \tau$ such that the object blends seamlessly into the scene under multiple viewpoints and lighting conditions.

Given that our framework incorporates a comprehensive BSDF surface model, an intuitive baseline for camouflage might be a fully transparent, glass-like material. However, simple transparency fails to achieve effective concealment. While a highly transmissive object allows the background's general tint to pass through, it drastically distorts the ambient light intensity. As illustrated in Fig. 5, this "trivial transparency" introduces chaotic light scattering and unpredictable, high-intensity specular spikes (glare) caused by refractive effects. As demonstrated in our experiments, this solution is highly suboptimal: because the human visual system is significantly more sensitive to variations in luminance (brightness) than shifts in chromaticity (color), these intense specular highlights immediately reveal the object's presence.

Fig. 5. A naïve transparent solution is ineffective, as glass-like materials tend to highlight the object's presence rather than conceal it.

C. Optimizing the Appearance Model

To optimize the appearance, we need some sort of ground truth or target. There are many camouflage strategies found in biology. *Background Matching* is the most familiar tactic. By replicating the appearance of the animals' natural habitats on their bodies, they can successfully avoid detection. Following this idea, we optimize the appearance by minimizing the photometric difference between the background and the object averaged over all the views. However, achieving a perfect matching between the object surface and the background is an ideal but rarely attainable goal for arbitrary scenes. In relatively homogeneous environments, such as grasslands or rocky terrains, the repetitive structure of the background allows

multi-view consistent appearances to be generated through background matching alone. However, in highly heterogeneous or high-contrast scenes, for example, a background consisting of white marble on one side and dark asphalt on the other, pure background matching quickly becomes insufficient. In such cases, it is necessary to synthesize new content that perceptually blends across different views.

Our overall optimization pipeline is summarized in Fig. 6, which consists of two optimization stages: (i) **Background Matching**, which enforces view-consistent appearance with the background scene, and (ii) **Physically-Grounded Synthesis**, which refines the texture with perceptual priors to achieve natural realism. This staged design balances physical accuracy with visual plausibility.

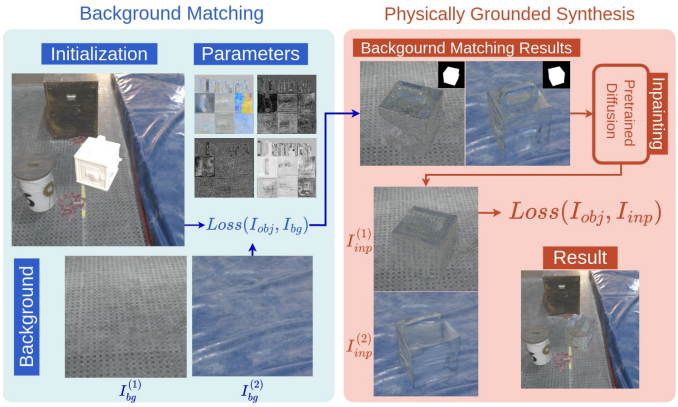


Fig. 6. Overview of the optimization pipeline used in our method.

Stage 1: Background Matching. We first render the scene with all our appearance parameters with a physically-based differentiable renderer (Mitsuba 3 [24] in our case). Formally, the rendered image I_{obj} is a function of vertices $\mathbf{v}_i'$ and materials $(\mathbf{a},\ \mathbf{r},\ \mathbf{m},\ \tau)$:

$$I_{\text{obj}} \ = \ \mathfrak{R}\Big(\{\mathbf{v}_i', \mathbf{n}_i, \mathbf{u}_i\}_i,\ \mathbf{a},\ \mathbf{r},\ \mathbf{m},\ \tau\Big), \tag{8}$$

where $\mathfrak{R}(\cdot)$ is the rendering function, $\mathbf{v}_i'$ is the vertices after pose and displacement transformation ($\mathbf{T}_{\text{new}}$ and D), $\mathbf{n}_i$ is the normal vector at vertex i, $\mathbf{u}_i$ is the UV coordinates at vertex i.

A natural starting point is to define the loss as the pixel-wise photometric difference between the object rendering I_{obj} and the background rendering I_{bg}. However, directly optimizing camouflage in this way is highly unstable because a physically based renderer operates in radiance space, where radiance is the fundamental quantity of light transport. In practice, the background and object surface can differ by several orders of magnitude—for example, an object illuminated by bright daylight against a background in deep shadow, as shown in Fig. 7. Attempting to reconcile such disparities through material parameter updates alone is generally infeasible and often yields misleading gradients. In Fig. 7, for instance, the teapot exhibits much higher radiance than its dark background. Since no physically plausible surface color can reduce this gap (a surface cannot emit light spontaneously), the optimization nonetheless drives the albedo in the wrong direction, leading to severe drift (e.g., toward an unnatural greenish tint).

To address this issue, we introduce a *luminance-gated chromaticity loss*. The key intuition is that surface material parameters cannot realistically account for large luminance discrepancies, which are primarily governed by lighting and geometry. Instead, materials contribute more reliably to the object's tint, i.e., its chromaticity. Our loss, therefore, decouples brightness from color: chromaticity alignment is enforced only in regions where the object and background have comparable luminance. This decomposition stabilizes the optimization and makes camouflage more robust in scenes where radiance differences are dominant.

Fig. 7. Directly minimizing the mean squared error (MSE) between object and background produces erroneous gradients, as shown on the lid of the teapot.

The loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{\text{lum-gated}}(I_{\text{obj}},\ I_{\text{bg}}) = \alpha\, \mathbb{E}[\,|Y(I_{\text{obj}}) - Y(I_{\text{bg}})|\,] \\ + \mathbb{E}\Big[g(\Delta Y)\ \cdot\ \big|\tilde{I}_{\text{obj}} - \tilde{I}_{\text{bg}}\big|\Big],\end{aligned} \tag{9}$$

where $\Delta Y = |Y(I_{\text{obj}}) - Y(I_{\text{bg}})|$, and $Y(\cdot)$ computes the standard Rec. 709 luminance (i.e., $Y(\mathbf{x}) = 0.2126\,x_r + 0.7152\,x_g + 0.0722\,x_b$). The pure chromaticity is isolated as $\tilde{\mathbf{x}} = \mathbf{x}/(Y(\mathbf{x}) + \varepsilon)$. The gating function, $g(x) = (1 + e^x)^{-1}$, acts as a dynamic switch: when a stark brightness mismatch exists between the object and the background (ΔY is large), $g(\Delta Y)$ decays toward zero, effectively muting the chromaticity penalty.

Given a rendered image $I_{\text{obj}}^{(v)}$ of the object at view v and the corresponding background image $I_{\text{bg}}^{(v)}$, we minimize the luminance-gated loss across all sampled views $v \in V$:

$$\mathcal{L}_{\text{photo}} = \frac{1}{|V|} \sum_{v \in V} \mathcal{L}_{\text{lum-gated}}(I_{\text{obj}}^{(v)},\ I_{\text{bg}}^{(v)}). \tag{10}$$

The overall optimization problem is then

$$\Theta_{app}^{*} = \arg\min_{\Theta_{\text{app}}} (\mathcal{L}_{\text{photo}} + \mathcal{L}_{\text{disp}}), \tag{11}$$

where $\mathcal{L}_{\text{disp}}$ is the displacement loss in Section 3.2.2.

This objective drives the optimization to match the background radiance and color with multiview consistency, firmly grounded in physical light transport.

Although it successfully aligns average color and tone across views, it often introduces perceptually distracting artifacts. The issue stems from the fact that pixel-level losses measure only intensity similarity, not perceptual quality. As a result, small deviations, such as isolated bright or dark pixels, may appear negligible in the averaged loss yet remain visually conspicuous, undermining the camouflage. This limitation motivates a subsequent refinement stage to better suppress such perceptual inconsistencies.

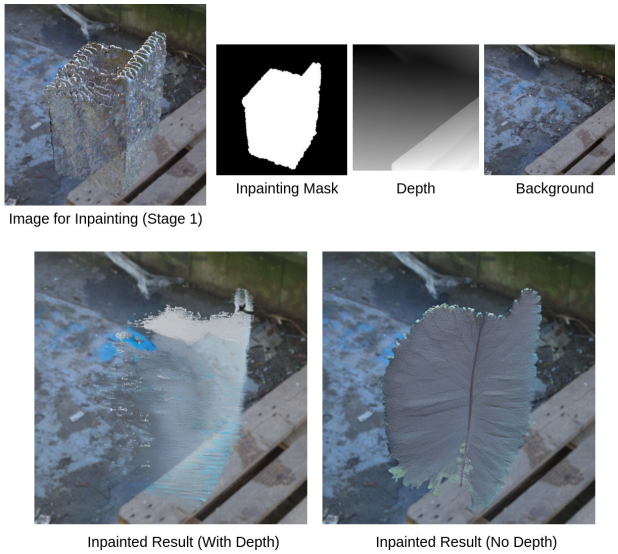


Fig. 8. Depth guidance is crucial for camouflage inpainting. Without it, the unconstrained model hallucinates salient foreground objects (right). With a depth reference, the model properly adheres to the background geometry, seamlessly blending the ambient blue tint and paint quality into the target region (left).

Stage 2: Physically-Grounded Synthesis. To achieve a more semantically cohesive appearance, we leverage 2D diffusion models, which are frequently utilized as strong priors to refine textures in 3D space (e.g., DreamGaussian [25], Instruct-NeRF2NeRF [26], and SNeRF [27]). These approaches feature an iterative 3D-grounded editing cycle: rendering the 3D representation into 2D images, editing them with a diffusion prior, and backpropagating the updates into 3D via image space loss. Because all view-dependent edits update a single global 3D representation, the refinement naturally achieves multi-view consistency. While prior works rely on text prompts or global style references for guidance, our refinement requires precise environmental blending. Consequently, we formulate our generative synthesis as an *inpainting* task—predicting the texture of the masked object regions conditioned directly on the visible background.

Formally, for a given view i, we first render the object silhouette to obtain a binary mask $M_i \in \{0,1\}^{H\times W}$, where $M_i = 1$ indicates object pixels. We then apply an SDXL-based [21] inpainting diffusion process, f_ϕ, to the rendering of the optimized object from the previous stage, denoted I^i_{prev}. Because diffusion models are trained on sRGB images, directly feeding linear-light renderings from a physically-based renderer causes a domain mismatch. Therefore, we first apply gamma encoding to map the linear intensities into a perceptually uniform space:

$$\hat{I}^i_{\text{prev}} = (I^i_{\text{prev}})^{\frac{1}{\gamma}}, \tag{12}$$

where typically $\gamma \approx 2.2$ for sRGB.

Next, the diffusion inpainting is applied with a controlled starting timestep t_{start} to bound the injected noise strength. Crucially, to prevent the unconstrained model from hallucinating salient foreground objects, we condition the inpainting process on the depth map of the background scene, $\mathcal{D}(I^i_{\text{bg}})$, via ControlNet [28] as illustrated in Fig. 8. The inpainted image is:

$$I^i_{\text{inp}} = f_\phi\big(\hat{I}^i_{\text{prev}} + \epsilon(t_{\text{start}});\ t_{\text{start}},\ M_i,\ I^i_{\text{bg}},\ \mathcal{D}(I^i_{\text{bg}})\big). \tag{13}$$

Notably, the inpainting process does not require complex prompt engineering. By providing only a generic text prompt (e.g., 'background'), we allow the model to deduce the precise context directly from the background image and the accompanying depth map.

Rather than directly pasting these inpainted textures onto the object, we utilize I^i_{inp} strictly as a pseudo-ground truth to further optimize the underlying BSDF parameters. The refinement loss for view i is defined as:

$$L^i_{\text{photo}} = \text{MSE}(\hat{I}^i_{\text{obj}},\ I^i_{\text{inp}}), \tag{14}$$

where $\hat{I}^i_{\text{obj}}$ denotes the sRGB rendering of the object given the updated BSDF parameters.

This iterative process effectively addresses multiview consistency by conditioning the generation of each new view on previously generated ones. The physically-grounded nature of this approach is twofold: first, the inpainting process is anchored by the physics-based inverse rendering results from Stage 1; second, the depth map further enforces accurate 3D geometry.

IV. Experimental Result

A. Implementation Details

Our gradient-based optimization framework is implemented in Python using PyTorch [29] as the primary backbone. For differentiable rendering, we integrate both PyTorch3D [30] and Mitsuba 3 [24]. PyTorch3D is employed for fast rasterization-based operations, mesh manipulations, mask estimation, while Mitsuba 3 provides physically based path-tracing for accurate simulation of complex materials, shadows, and global illumination.

All experiments are conducted on a single NVIDIA RTX 4090 GPU. For appearance initialization, the *albedo* map is set to a constant value of 0.5, *metallic* to 0.0, *roughness* to 0.0, and *transmittance* to 0.5, with all maps rendered at a resolution of 800×800. To ensure sufficient geometric detail for displacement, meshes are remeshed and subdivided to approximately 30,000 vertices and UV-unwrapped to enable proper texture mapping.

Optimization hyperparameters are chosen to balance stability and convergence speed. The learning rate for texture maps is set to 10^{-2}, while the displacement map uses a higher learning rate of 10^{-1} due to its stronger influence on geometry. During the *Background Matching* stage, optimization is performed for 50 iterations, followed by 20 iterations in the *Physically-Grounded Synthesis* stage. The total runtime is approximately one minute for the first stage and five minutes for the second. Once the optimized 3D model is obtained, it can be rendered interactively in real time. For a detailed visualization of the 3D results, we refer the reader to the supplementary video.

B. Experimental Settings

To rigorously evaluate our framework, we curate a diverse set of 3D objects and environmental contexts. We evaluate our method on 50 mesh models selected from the NAVI [31] and Objaverse-XL [32] datasets. Because camouflage is primarily relevant for static, everyday objects, our selection focuses on inanimate categories (e.g., tables, backpacks, and toys) that encompass a wide range of geometric complexity and topological variations. To provide realistic environmental contexts, we situate these objects within 30 representative scenes drawn from Benedikt Bitterli's rendering gallery [33] and Poly-Haven [34]. These environments span both indoor and outdoor settings under diverse illumination conditions, including bright daylight, overcast skies, and dimly lit interiors, ensuring that the camouflage is tested against complex lighting interactions.

During evaluation, we assume the object must remain concealed across a continuous 360° space. Accordingly, we uniformly sample 16 camera viewpoints along a circular orbit at a fixed elevation around the object. This setup eliminates view-specific bias and tests the generalization of the camouflage across multiple perspectives. To ensure our method does not overfit, our quantitative evaluation explicitly tests for novel-view generalization. Specifically, we do not evaluate metrics on the exact training cameras. Instead, we apply random 3D spatial perturbations to the evaluation cameras (up to ± 0.25 units at a radius of 4.0). This corresponds to random angular perturbations of up to $\pm 4°$ to $5°$ from the optimized viewing directions. All evaluations are rendered using standard path tracing, ensuring that our findings are physically accurate and generalize across different rendering pipelines.

To demonstrate the effectiveness of our approach, we compare our method against a comprehensive set of baseline camouflage strategies to evaluate its performance relative to both naive solutions and competitive generative approaches. These baselines include: (i) **Transparent**, where the transmittance is uniformly set to 1.0 to simulate an everyday glass-like object; and (ii) **Random** [6], which randomly samples background pixel color onto the object surface from each view. For advanced generative baselines, we compare against (iii) **GANmouflage** [7], the current state-of-the-art method for 3D texture-based camouflage; and (iv) a **2D Diffusion** baseline, which projects 2D diffusion inpainting results onto the object's surface.

To implement **GANmouflage**, we first render background images from 16 viewpoints without the object present, recording the corresponding camera intrinsic and extrinsic parameters. The model is then trained for 8,000 iterations (approximately three hours per scene) using the hyperparameters reported in the original paper [7]. For the **2D Diffusion** baseline, we apply a pre-trained 2D diffusion inpainting model (SDXL [21]) to generate the target camouflage textures purely in image space across the training views. These 2D inpainted images are subsequently projected iteratively onto the 3D model. To ensure a completely fair comparison, the final visualizations across all methods are generated using the identical path-tracing renderer.

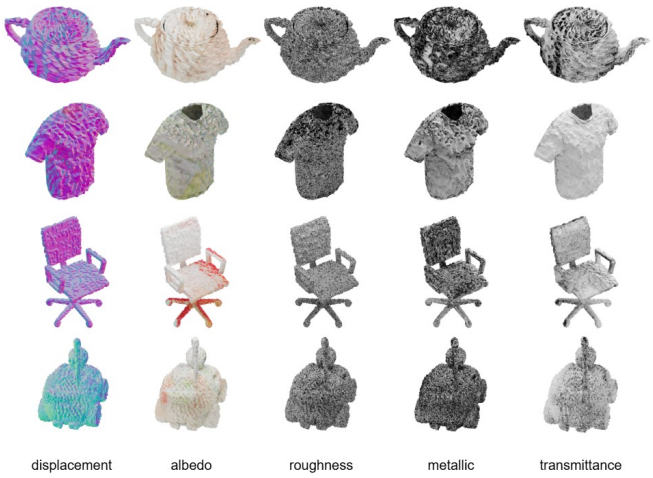


Fig. 9. Textured 3D model outputs produced by our method.

C. Visual Comparison

We perform a qualitative evaluation to assess the visual fidelity of our generated camouflage textures with various scenes and meshes. The results are presented in Fig. 16. The corresponding 3D models in shown Fig. 9. For a fair visual comparison, we temporarily disable the pose optimization, allowing object to be placed at exactly the same location for different methods.

A primary advantage of our appearance modeling is its robust adaptation to backgrounds with highly conflicting colors and textures. For instance, in the top example of Fig. 16, the background consists of both white and wooden surfaces. Baseline methods appear overly white in most views, failing to balance the two. In the second case, the background spans a bright region and a dark shadow; baseline results are either too bright or too dark, but never accurate across both. In the third case, the office chair appears pink, an average of red and white, despite no viewpoint actually having a pink background, illustrating the limitations of existing surface representations. Finally, in the last case, baselines achieve good camouflage in some viewpoints but only at the expense of degraded camouflage in others.

Beyond spatial robustness, our method demonstrates superior radiance awareness and color accuracy. Unlike prior methods [6], [7], our approach preserves the original background colors without introducing noticeable shifts, ensuring that generated textures remain perceptually coherent. As illustrated in Fig. 16, the T-shirt and toy car appear unnaturally oversaturated or bright in the baseline results, whereas our method maintains the correct white balance. Because baseline 2D diffusion models treat each inpainted view independently, they struggle to achieve a consistent 3D appearance, often resulting in localized artifacts—such as the yellow tint observed on the T-shirt.

Furthermore, because our framework operates in 3D space, it naturally handles depth and occlusion. In contrast, 2D baselines like GANmouflage [7] operate strictly in image space and assume the target object is always fully visible in the

foreground. This limitation causes them to fail when objects are partially occluded, such as the office chair tucked behind the desk in Fig. 16. Overall, these architectural advantages result in generated textures that consistently outperform baseline methods in naturalness and perceptual fidelity.

Fig. 10. Visualization of the optimized transmittance map, where white denotes high transmittance (more transparent) and black denotes low transmittance (more opaque).

To better understand these qualitative improvements, we analyze the generated BSDF parameters. As shown in Fig. 10, the optimized transmittance maps correlate strongly with the environmental lighting direction, demonstrating that the optimizer seeks a solution strictly governed by the physical radiance dictated by the rendering equation. This highlights a core philosophy of our approach: true physical camouflage is fundamentally an *optical modulation* task, not merely a texture synthesis problem. Rather than "painting" flat color onto a surface, our method optimizes material properties to manipulate 3D light transport. For example, light-facing surfaces naturally develop higher transmittance. From an energy-conservation perspective, matching a dim background while under bright foreground illumination requires prioritizing transmission (f_t) over surface reflection (f_r) to prevent the outgoing radiance from overshooting the background intensity. Conversely, shadowed regions adopt low transmittance to restrict light passage and match their darker surroundings. Furthermore, we observe a strong spatial correlation between the metallic parameter distribution and the peaks and valleys of the anamorphic displacement (see the metallic map in Fig. 9). This indicates that the geometry relies heavily on the metallic parameter to modulate high-frequency lighting patterns. Specifically, while a purely diffuse surface scatters incident light evenly across a hemisphere—resulting in a dimmer directional appearance—increasing the metallic parameter concentrates the reflected energy into a tight specular lobe. This enables the optimizer to achieve the peak directional radiance necessary for the displacement geometry to successfully generate highly view-dependent appearances. Finally, we note a counterintuitive behavior regarding the surface color: the optimized albedo often does not directly match the background color. Because albedo is no longer the sole contributor to the object's appearance, it becomes intricately coupled with the other BSDF parameters within the full light transport simulation. Rather than acting as a flat texture, the albedo serves as a global color modulator, simultaneously controlling the tint of both the transmitted light and the reflected light to ensure the final, combined radiance seamlessly blends with the environment.

TABLE I
PERCEPTUAL STUDY. HIGHER NUMBERS REPRESENT BETTER PERFORMANCE.

Method	Conf. rate	Avg. time (s)	Med. time (s)
Transparent	0.10%	1.79	1.86
Random [6]	15.77%	4.65	3.92
GANmouflage [7]	12.99%	3.61	3.39
2D Diffusion	9.11%	3.03	3.48
Ours	**54.11%**	**8.02**	**9.41**

D. Quantitative Evaluation

We evaluate our method using two complementary approaches: a perceptual evaluation with human participants, and automatic benchmark metrics.

Perceptual Evaluation. Following the evaluation protocol of GANmouflage [7], we conducted an online user study to assess human visual perception of our synthesized camouflage. We recruited 71 participants with diverse backgrounds to ensure a broad representation of general users. The participants spanned a wide age range (10 to 70 years old), with the majority being young adults: 70.4% ($n = 50$) fell into the 20–30 age group, and 14.1% ($n = 10$) into the 30–40 age group. The remaining cohort consisted of participants aged 40–50 (5.6%, $n = 4$), 50–60 (5.6%, $n = 4$), 60–70 (2.8%, $n = 2$), and 10–20 (1.4%, $n = 1$). The study was administered via an online questionnaire platform. For each scene, participants were shown one randomly selected image from the reserved evaluation set, presented in a completely randomized order. To familiarize participants with the interface and task, the first five images served as practice trials and were excluded from the final quantitative analysis. During each formal trial, participants were instructed to locate the camouflaged object—which was rendered using a randomly selected algorithm at a predefined location—and click on it as quickly as possible. Upon clicking, the object's true outline was revealed. Participants were allotted a maximum of 40 seconds per trial. For every evaluation, we recorded both the detection accuracy (whether the participant correctly identified the object) and the time taken to locate it.

The results, summarized in Table I, show that our method achieves the highest confusion rate and longest search time, demonstrating a clear advantage over baseline methods. This improvement can be largely attributed to our optimization pipeline, which is grounded in realistic scene conditions and produces perceptually robust camouflage.

Automatic Benchmark. We adopt the evaluation protocol introduced in GANmouflage [7], computing both **LPIPS** and **SIFID** scores between the generated camouflage textures and the corresponding background images. LPIPS provides a perceptual similarity measure that correlates well with human visual judgments, assessing whether the generated textures resemble the background in terms of local appearance and structure. SIFID (Single Image Fréchet Inception Distance) measures the statistical similarity between feature distributions of the generated image and the corresponding background at the single-image level, capturing both structural consistency

TABLE II
QUANTITATIVE COMPARISON.

Model	LPIPS↓	SIFID↓	DreamSim↓	SSIM↑	PSNR↑
Transparent	0.0495	0.0265	0.1178	0.9717	26.57
Random [6]	0.0565	0.0060	0.1036	0.9751	26.32
GANmouflage [7]	0.0503	0.0030	0.1296	0.9761	26.53
2D Diffusion	0.0491	0.0078	0.1130	0.9745	27.88
w/o Pose	0.0248	0.0069	0.0452	0.9734	30.56
w/o Displacement	0.0252	**0.0019**	0.0501	**0.9769**	30.43
w/o BSDF (Only Albedo)	0.0283	0.0100	0.0745	0.9718	29.08
w/o Transmittance	0.0301	0.0129	0.0625	0.9719	29.65
w/o Inpainting	0.0253	0.0110	0.0450	0.9751	30.78
Ours	**0.0247**	0.0073	**0.0443**	0.9740	**31.16**

and texture realism. Because the primary focus of our work is perceptual camouflage, we augment this standard protocol by introducing **DreamSim** [35], a recent metric specifically designed to align deeply with human visual perception and holistic image similarity. For completeness and standard reference, we also report traditional pixel-level metrics, namely **SSIM** and **PSNR**.

As summarized in Table II, the purely transparent baseline is highly suboptimal, performing even worse than standard opaque objects due to pronounced specular glare. This underscores that effective concealment is a non-trivial problem: rather than simply maximizing transmittance, it requires inversely solving for optimal material parameters to actively modulate light transport. Our method achieves the best scores across DreamSim, LPIPS, and PSNR, indicating that it produces camouflage perceptually superior to prior approaches. These quantitative findings closely align with the visual evidence: our camouflage is harder to detect and blends more naturally into the background. Interestingly, we found that the SIFID metric is highly sensitive to local texture but largely oblivious to stark color differences, thereby inherently favoring flat, smooth appearances (such as the Random, GANmouflage, and 2D Diffusion baselines). Furthermore, the relatively poor performance of GANmouflage highlights the fundamental limitation of optimizing camouflage solely in 2D image space; when evaluated in physically realistic 3D environments, such image-space patterns fail to adapt to lighting and become noticeably conspicuous.

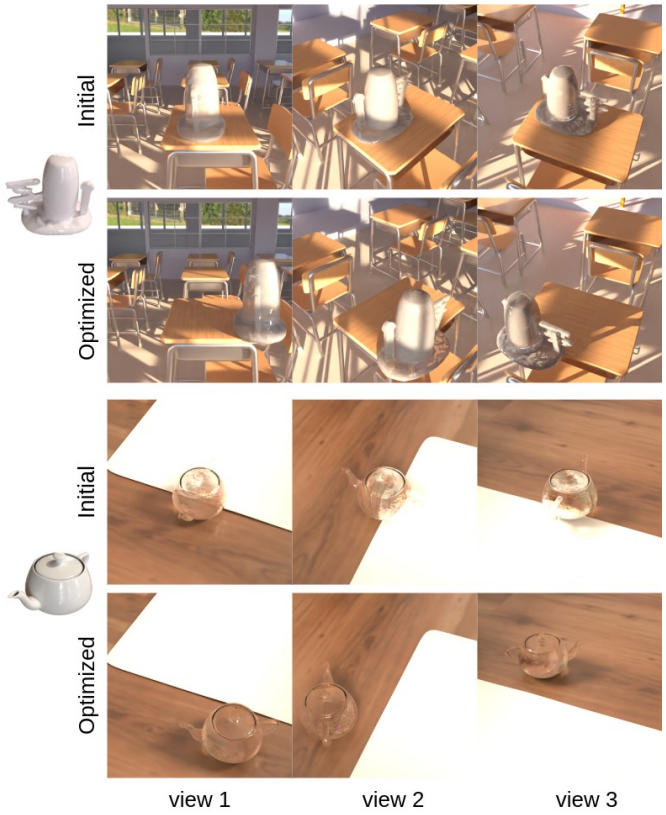


Fig. 11. Interesting effect of pose optimization.

E. Ablation Study

We conduct an ablation study to evaluate the contribution of each key component in our method. The experiments are performed by systematically removing or modifying specific modules and comparing the resulting camouflage quality. The quantitative and qualitative results are summarized in Table II and illustrated in Fig. 11 through Fig. 14.

Effect of Pose Optimization. Removing pose optimization generally leads to higher camouflage loss, as pose directly affects the difficulty of the target viewpoints. The role of pose optimization is to make camouflage learning more tractable by dynamically adjusting the object's orientation and location. In Fig. 11, we observe interesting emergent behaviors: the model implicitly avoids highly illuminated regions that would incur larger losses (top example), and it avoids high-contrast areas that could compromise multiview consistency (bottom example). Notably, these behaviors arise purely from the 2D loss function without any explicit 3D encoding. However, the contribution of pose optimization is scene-dependent. In cases such as the teapot example (Fig. 11), it simplifies the problem considerably, whereas in locally homogeneous background regions, its impact is less pronounced.

Importance of Full BSDF. Restricting the representation to albedo alone substantially reduces the expressiveness of the surface appearance. As shown in Fig. 12, without full BSDF parameters, the camouflage becomes less flexible and tends to collapse into averaged colors, producing a dull grayish appearance on the duck. This limitation also results in color bleeding across neighboring views. For example, the blue re-

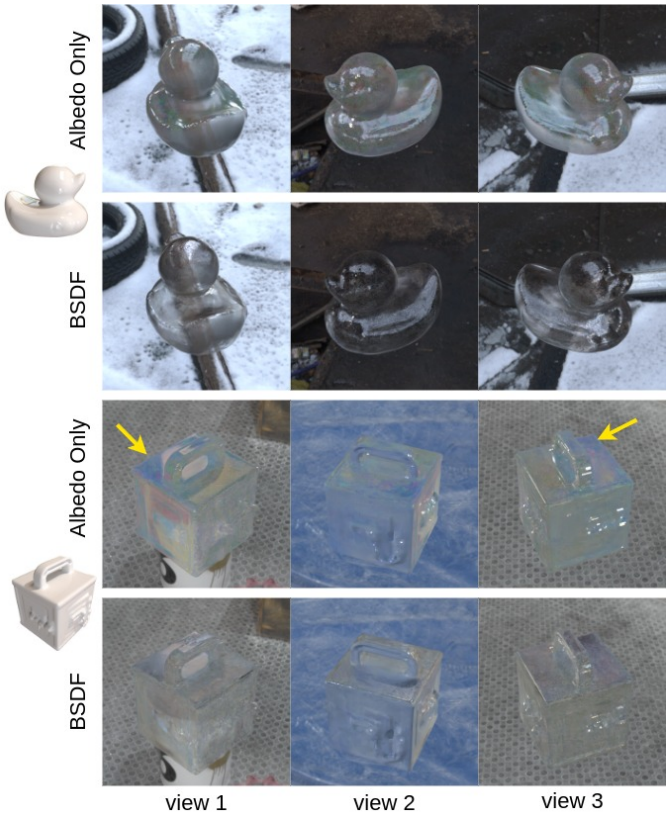


Fig. 12. Effect of using BSDF.

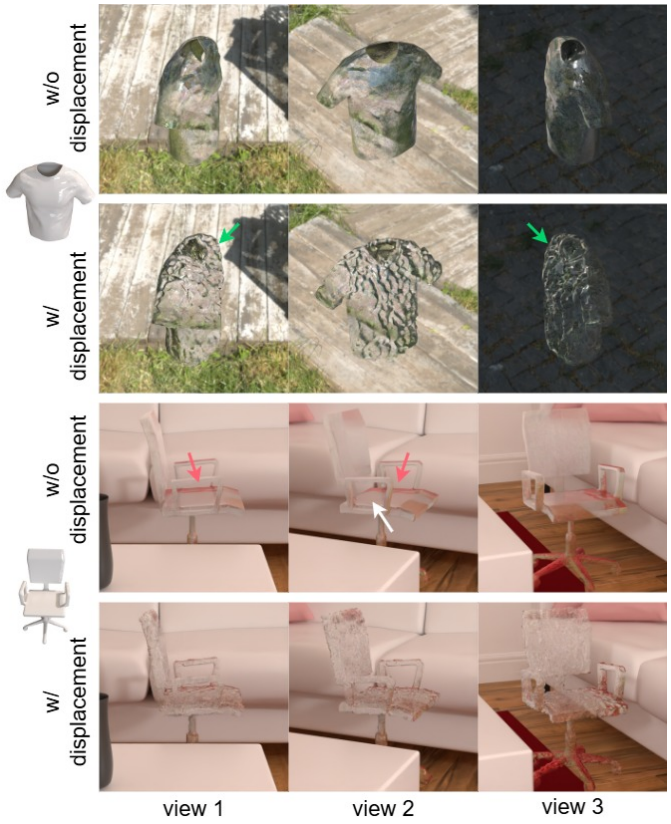


Fig. 13. Effect of anamorphic displacement optimization.

gion (yellow arrow, bottom example) spreads erroneously into adjacent viewpoints with albedo-only optimization, whereas our full BSDF representation maintains significantly better multiview consistency and spatial diversity.

Impact of Anamorphic Displacement. Omitting displacement maps results in flatter, less articulated surfaces, reducing the realism and variability of the camouflage. Fig. 13 highlights how displacement enhances multiview fidelity. In the t-shirt example (top), displacement allows the surface to adapt more effectively to both bright and dark viewpoints (green arrow). In the chair example (bottom), the absence of displacement leads to red (pink arrow) and white (white arrow) bleeding across views, while the displaced geometry mitigates this effect by better preserving view-dependent appearances.

Role of Transmittance. Disabling the transmittance parameter (i.e., forcing it to zero) severely restricts the expressive capacity of the surface representation, leading to sub-optimal camouflage (similar to the result in Fig. 12). As demonstrated by the quantitative results in Table II, transmittance is a crucial component of our framework. It serves as the primary mechanism for macro-level light modulation, which, as previously discussed, is the fundamental objective of physical camouflage. However, high transmittance alone is insufficient; achieving seamless integration requires intricate coordination among all BSDF parameters. For example, because physically based rendering dictates that high metallic values strictly suppress dielectric transmission, the optimizer must carefully balance these mutually exclusive properties. Simultaneously, the albedo parameter must be optimized to correctly tint the transmitted light, ensuring the final modulated radiance perfectly matches the background.

Physically-Grounded Synthesis. Finally, we evaluate the effect of removing the inpainting-based synthesis stage. Without this stage, noticeable artifacts remain, and the generated textures often fail to blend seamlessly with the background. As shown in Fig. 14 (top), the synthesis step introduces green linear structures (red arrow) that mimic nearby patterns (yellow arrow), thereby enhancing visual coherence. It also refines high-level textures: without inpainting, the surface appears grainy (blue arrow) despite being placed against a smooth metallic background, whereas the synthesized result produces a consistent, well-matched texture (enlargement recommended). In the sink example (bottom), this stage effectively removes dark artifacts (orange arrow) introduced by the pixel-wise loss and further synthesizes a realistic concrete-like appearance (green arrow), transforming the mesh into a visually convincing concrete block.

These ablation results confirm that each component: pose, BSDF, displacement, and inpainting-based synthesis, plays a vital role in producing high-quality and realistic camouflage.

F. Limitations

While our approach achieves significantly higher visual fidelity than prior methods, several limitations remain that present exciting avenues for future research. First, our camouflage quality deteriorates in scenes with extremely con-

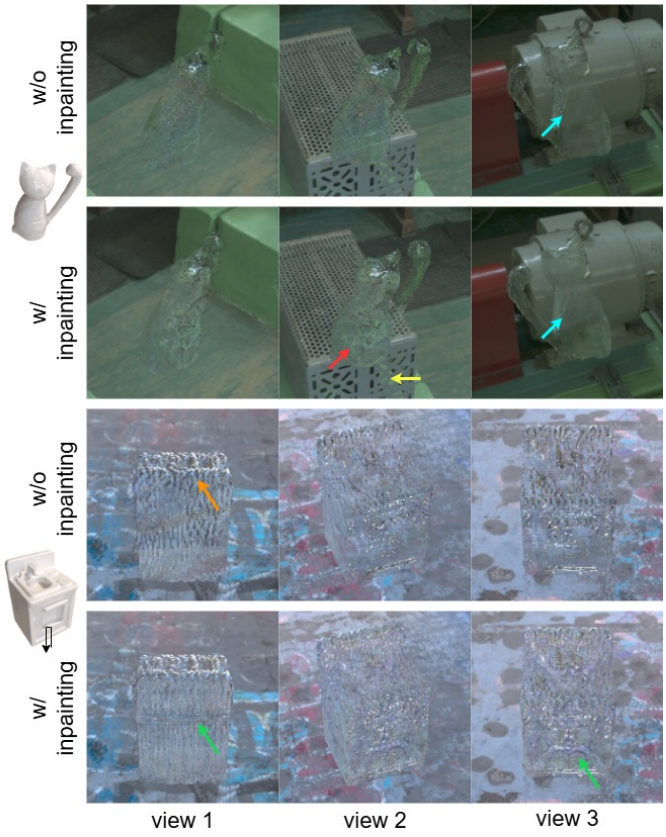


Fig. 14. Effect of using inpainting.

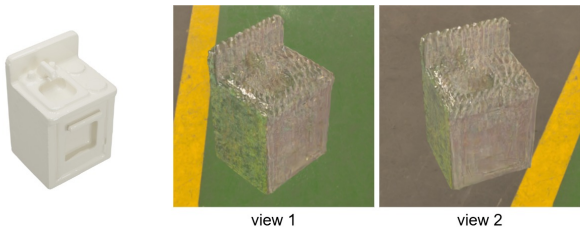


Fig. 15. A limitation of our method arises when the background contains highly conflicting patterns. In such cases, even nearby viewpoints may correspond to drastically different colors—gray versus green in this example—making consistent camouflage challenging.

flicting backgrounds, as illustrated in Fig. 15. Although this limitation is shared by existing methods, addressing it will likely require novel strategies that go beyond pure background matching. Second, because our framework accurately models real light transport, the optimization process is computationally intensive, requiring several minutes per scene. Consequently, while the final camouflaged model allows for interactive visualization, it cannot currently adapt in real time to dynamic environments with rapidly changing lighting or backgrounds. Future work could explore algorithmic acceleration and model compression to enable live deployment in augmented reality. Third, our results are currently confined to virtual environments. Translating these digitally optimized camouflage textures into physical, manufacturable materials poses substantial challenges due to hardware limitations in printing resolution, material property reproduction, and real-world durability. Bridging this gap will ultimately require interdisciplinary collaboration with experts in materials science and advanced fabrication. Finally, our framework relies on gradient-based optimization within a highly non-convex loss landscape. Consequently, the optimized material parameters converge to a local optimum rather than a guaranteed global minimum. While these local optima are perceptually highly effective and successfully achieve our camouflage objectives—as is common in underconstrained visual synthesis tasks with multiple valid solutions—they may not represent the absolute theoretical limit of concealment. Exploring global search algorithms or heuristic methods to establish a "pseudo ground-truth" or theoretical upper bound for 3D camouflage performance remains an interesting avenue for future research, providing valuable quantitative benchmarks for evaluating future optimization strategies.

V. Conclusion

We have presented a novel framework for 3D object camouflage that jointly optimizes pose, geometry, and material appearance. By combining pose refinement, anamorphic displacement, and spatially varying BSDF optimization within a differentiable rendering pipeline, our method achieves camouflage that is physically realistic, perceptually convincing, and robust across multiple viewpoints. Unlike prior approaches that rely on fixed lighting or simplified material assumptions, our model adapts to complex illumination and scene contexts, making it applicable to a wide range of environments. Future work includes extending the method toward real-time optimization, investigating fabrication-aware constraints for physical deployment, and exploring perceptual studies to better quantify camouflage effectiveness in human vision. Furthermore, inspired by the emergent spatial behaviors observed during pose optimization, an exciting avenue for future research involves developing algorithms to actively determine the globally optimal camouflage location for an object within a complex 3D environment.

Acknowledgments

This work is supported by the National Science and Technology Council, Taiwan under Grant 113-2221-E006-161-MY3 and Grant114-2221-E-006-114-MY3.

References

[1] C. Reynolds, "Interactive evolution of camouflage," *Artif. Life*, vol. 17, no. 2, p. 123–136, Feb. 2011. [Online]. Available: https://doi.org/10.1162/artl_a_00023

[2] H.-K. Chu, W.-H. Hsu, N. J. Mitra, D. Cohen-Or, T.-T. Wong, and T.-Y. Lee, "Camouflage images," vol. 29, no. 4, Jul. 2010. [Online]. Available: https://doi.org/10.1145/1778765.1778788

[3] Q. Zhang, G. Yin, Y. Nie, and W.-S. Zheng, "Deep camouflage images," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 12 845–12 852.

[4] Y. Li, W. Zhai, Y. Cao, and Z.-J. Zha, "Location-free camouflage generation network," *IEEE Transactions on Multimedia*, vol. 25, pp. 5234–5247, 2023.

[5] B. Das and V. Gopalakrishnan, "Camouflage anything: Learning to hide using controlled out-painting and representation engineering," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 3603–3613.

[6] A. Owens, C. Barnes, A. Flint, H. Singh, and W. Freeman, "Camouflaging an object from many viewpoints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[7] R. Guo, J. Collins, O. de Lima, and A. Owens, "Ganmouflage: 3d object nondetection with texture fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 4702–4712.

[8] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.

[9] M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 3rd ed. Morgan Kaufmann, 2016.

[10] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, "Geometrical considerations and nomenclature for reflectance," National Bureau of Standards (US), Tech. Rep. NBS Monograph 160, 1977.

[11] N. Suryanto, Y. Kim, H. T. Larasati, H. Kang, T.-T.-H. Le, Y. Hong, H. Yang, S.-Y. Oh, and H. Kim, "Active: Towards highly transferable 3d physical camouflage for universal and robust vehicle evasion," *arXiv preprint arXiv:2308.07009*, 2023.

[12] J. Zhou, L. Lyu, D. He, and Y. Li, "Rauca: a novel physical adversarial attack on vehicle detectors via robust and accurate camouflage generation," in *Proceedings of the 41st International Conference on Machine Learning*, ser. ICML'24. JMLR.org, 2024.

[13] X. Zhang, J. Chen, H. Zheng, and Z. Liu, "Phycamo: a robust physical camouflage via contrastive learning for multi-view physical adversarial attack," in *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025. [Online]. Available: https://doi.org/10.1609/aaai.v39i10.33110

[14] J. Zeng, Y. Ou, and W. Matusik, "3d printed objects with lenticular lens surfaces that can change their appearance," in *UIST*, 2021.

[15] M. Perroni-Scharf and T. Weyrich, "Constructing printable surfaces with view-dependent appearance," in *SIGGRAPH*, 2023.

[16] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Repaint: Inpainting using denoising diffusion probabilistic models," in *CVPR*, 2022.

[17] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, T. Kong, D. Galyuk, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," *arXiv preprint arXiv:2109.07161*, 2021.

[18] Q. Dong, C. Cao, Y.-W. Tai, and C.-K. Tang, "Incremental transformer structure enhanced image inpainting," in *CVPR*, 2022.

[19] B. F. Labs, "Flux," https://github.com/black-forest-labs/flux, 2024.

[20] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023.

[21] ——, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023.

[22] M. Stevens and G. D. Ruxton, "The key role of behaviour in animal camouflage," *Biological Reviews*, vol. 94, no. 1, pp. 116–134, 2019.

[23] "Luycho On Flowers Series_Toucan Mirror Cup & Wavy Saucer (12 oz)," https://www.amazon.com/Luycho-Mirror-Cup-Wavy-Saucer_Toucan_12oz/dp/B07YYV4JLH, product page on Amazon; Cup material: porcelain; Saucer material: plastic; Capacity: 12 oz; Part of the "On Flowers" series. Accessed: 2025-09-08.

[24] W. Jakob, S. Speierer, N. Roussel, M. Nimier-David, D. Vicini, T. Zeltner, B. Nicolet, M. Crespo, V. Leroy, and Z. Zhang, "Mitsuba 3 renderer," 2022, https://mitsuba-renderer.org.

[25] J. Tang, J. Ren, H. Zhou, Z. Liu, and G. Zeng, "Dreamgaussian: Generative gaussian splatting for efficient 3d content creation," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=UyNXMqnN3c

[26] A. Haque, M. Tancik, A. Efros, A. Holynski, and A. Kanazawa, "Instruct-nerf2nerf: Editing 3d scenes with instructions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[27] T. Nguyen-Phuoc, F. Liu, and L. Xiao, "Snerf: stylized neural implicit representations for 3d scenes," *ACM Trans. Graph.*, vol. 41, no. 4, Jul. 2022. [Online]. Available: https://doi.org/10.1145/3528223.3530107

[28] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 3836–3847.

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[30] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson, and G. Gkioxari, "Accelerating 3d deep learning with pytorch3d," *arXiv:2007.08501*, 2020.

[31] V. Jampani, K.-K. Maninis, A. Engelhardt, A. Karpur, K. Truong, K. Sargent, S. Popov, A. Araujo, R. Martin-Brualla, K. Patel, D. Vlasic, V. Ferrari, A. Makadia, C. Liu, Y. Li, and H. Zhou, "Navi: Category-agnostic image collections with high-quality 3d shape and pose annotations," in *NeurIPS*, 2023. [Online]. Available: https://navidataset.github.io/

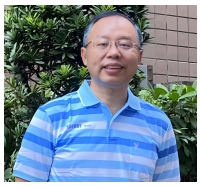
[32] M. Deitke, R. Liu, M. Wallingford, H. Ngo, O. Michel, A. Kusupati, A. Fan, C. Laforte, V. Voleti, S. Y. Gadre, E. VanderBilt, A. Kembhavi, C. Vondrick, G. Gkioxari, K. Ehsani, L. Schmidt, and A. Farhadi, "Objaverse-xl: A universe of 10m+ 3d objects," *arXiv preprint arXiv:2307.05663*, 2023.

[33] B. Bitterli, "Rendering resources," 2016, https://benedikt-bitterli.me/resources/.

[34] P. Haven, "Poly haven: The public 3d asset library," 2025, accessed: 2025-09-08. [Online]. Available: https://polyhaven.com/

[35] S. Fu, N. Tamir, S. Sundaram, L. Chai, R. Zhang, T. Dekel, and P. Isola, "Dreamsim: Learning new dimensions of human visual similarity using synthetic data," in *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 50 742–50 768.

Dong-Yi Wu received his BSc degree in Physics from National Cheng Kung University, Tainan, Taiwan in 2015. He is working towards the PhD degree at Computer Graphics Laboratory, Department of Computer Science and Information Engineering, National Cheng Kung University. His research interests include computer graphics, visualization and data mining.



Tong-Yee Lee received the PhD degree in computer engineering from Washington State University, Pullman, in May 1995. He is currently a chair professor with the Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan, Taiwan, ROC. He leads the Computer Graphics Laboratory, National Cheng-Kung University (http://graphics.csie.ncku.edu.tw). His current research interests include computer graphics, nonphotorealistic rendering, medical visualization, virtual reality, and media resizing. He is a senior member of the IEEE Computer Society and a member of the ACM. He also serves on the editorial boards of the IEEE Transactions on Visualization and Computer Graphics and IEEE Computer Graphics and Applications.

Fig. 16. Demonstration of our results. The input is shown on the left, and the object's location in the scene is indicated by the arrow. For each scene, we present three different viewpoints and compare our method against three baseline approaches: Mean [6], Random [6] and GANmouflage [7].