Differentiable Rendering with Perturbed Optimizers
Quentin Le Lidec, Ivan Laptev, Cordelia Schmid, Justin Carpentier

To cite this version:
Quentin Le Lidec, Ivan Laptev, Cordelia Schmid, Justin Carpentier. Differentiable Rendering with Perturbed Optimizers. NeurIPS 2021 - Thirty-fifth Conference on Neural Information Processing Systems, Dec 2021, Sydney / Virtual, Australia. hal-03378451

HAL Id: hal-03378451
https://hal.science/hal-03378451
Submitted on 14 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

Reasoning about 3D scenes from their 2D image projections is one of the core problems in computer vision. Solutions to this inverse and ill-posed problem typically involve a search for models that best explain observed image data. Notably, images depend both on the properties of observed scenes and on the process of image formation. Hence, if optimization techniques should be used to explain images, it is crucial to design differentiable functions for the projection of 3D scenes into images, also known as differentiable rendering. Previous approaches to differentiable rendering typically replace non-differentiable operations by smooth approximations, impacting the subsequent 3D estimation. In this paper, we take a more general approach and study differentiable renderers through the prism of randomized optimization and the related notion of perturbed optimizers. In particular, our work highlights the link between some well-known differentiable renderer formulations and randomly smoothed optimizers, and introduces differentiable perturbed renderers. We also propose a variance reduction mechanism to alleviate the computational burden inherent to perturbed optimizers and introduce an adaptive scheme to automatically adjust the smoothing parameters of the rendering process. We apply our method to 3D scene reconstruction and demonstrate its advantages on the tasks of 6D pose estimation and 3D mesh reconstruction. By providing informative gradients that can be used as a strong supervisory signal, we demonstrate the benefits of perturbed renderers to obtain more accurate solutions when compared to the state-of-the-art alternatives using smooth gradient approximations.

1 Introduction

Many common tasks in computer vision such as 3D shape modelling [5, 15, 34, 36, 37] or 6D pose estimation [18, 22, 25, 30, 32] aim at inferring 3D information directly from 2D images. Most of the recent approaches rely on (deep) neural networks and thus require large training datasets with 3D shapes along with well-chosen priors on these shapes. Render & compare methods [18, 22] circumvent the non-differentiability of the rendering process by learning gradients steps from large datasets. Using a more structured strategy would allow to alleviate the need for such a strong supervision. In this respect, differentiable rendering intends to model the effective image generation process to compute a gradient related to the task to solve. This approach has the benefit of containing the prior knowledge of the rendering process while being interpretable. This makes it possible to provide a supervision for neural networks in a weakly-supervised manner [10, 16, 23]. The main challenge of differentiable rendering lies in the non-smoothness of the classical rendering process. In a renderer, both the rasterization steps, which consist in evaluating the pixel values by discretizing the 2D projected color-maps, and the aggregation steps, which merge the color-maps of several objects along the depth dimension by using a Z-buffering operation, are non-differentiable operations (Fig. 1). Intuitively, these steps imply discontinuities in the final rendered image with respect to the 3D positions of the scene objects. For example, if an object moves on a plane parallel to the...
camera, some pixels will immediately change color at the moment the object enters the camera view or becomes unoccluded by another object.

In this paper, we propose to exploit randomized smoothing techniques within the context of differentiable rendering to automatically soften the non-smooth rendering operations, making them naturally differentiable. The generality of our approach offers a theoretical understanding of some of the existing differentiable renderers while its flexibility leads to competitive or even state-of-the-art results in practice. We make the following contributions:

We formulate the non-smooth operations occurring in the rendering process as solutions of optimization problems and, based on recent work [4], we propose a natural way of smoothing them using random perturbations. We highlight the versatility of this smoothing formulation and show how it offers a theoretical understanding of several existing differentiable renderers.

We propose a general way to use control variate methods to reduce variance when estimating the gradients of the perturbed optimizers, which allows for sharp gradients even in the case of weaker perturbations.

We introduce an adaptive scheme to automatically adjust the smoothing parameters by relying on sensitivity analysis of differentiable perturbed renderers, leading to a robust and adaptive behavior during the optimization process.

We demonstrate through experiments on pose optimization and 3D mesh reconstruction that the resulting gradients combined to the adaptive smoothing provide a strong signal, leading to more accurate solutions when compared to state-of-the-art alternatives based on smooth gradient approximations [23].

2 Background

In this section, we review the fundamental aspects behind image rendering and recall the notion of perturbed optimizers which are at the core of the proposed approach.
We call $R$ the occupancy map such that $I_j$ is equal to 1 if the center of the $j^{th}$ pixel is inside the 2D projection of the $j^{th}$ triangle on the camera plane and 0 otherwise. By denoting $d(i,j)$ the Euclidean distance from the center of the $j^{th}$ pixel to the projection of the $j^{th}$ triangle, the rasterization step (which actually consists in computing the occupancy map $I$) can be written as:

$$I_j = H(d(i,j)),$$

where $H$ corresponds to the Heaviside function defined by:

$$H(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{otherwise} \end{cases} = \text{argmax}_{0 \leq y \leq 1} y x.$$  

(1)

We call $R \in \mathbb{R}^{h_j \times w_j \times 3}$ the final rendered image which is obtained by aggregating the color map of each triangle. In classical renderers, this step is done by using a Z-buffer so that only foreground objects are visible. This corresponds to:

$$R' = \sum_j w_j(z)C_j,$$

where $w, z \in \mathbb{R}^{m+1}$ and for $1 \leq j \leq m$, $z_j$ is the inverse depth of the $j^{th}$ triangle and the $(m+1)^{th}$ coordinate of $z$ and $w$ account for the background. The inverse depth of the background is fixed to $z_{m+1} = z_{\text{min}}$. From Eq. (2) and (3), it appears that argmax operations play a central role in renderers and are typically non-differentiable functions with respect to their input arguments, as discussed next.

**Perturbed optimizers.** In [4], Berthet et al. introduce a generic approach to handle non-smooth problems of the form $y^*(\theta) = \text{argmax}_{y \in C} \langle \theta, y \rangle$ with $C$ a convex polytope, and make these problems differentiable by randomly perturbing them. More precisely, $y^*(\theta)$ necessarily lies on a vertex of the convex set $C$. Thus, when $\theta$ is only slightly modified, $y^*$ remains on the same vertex, but when the perturbation grows up to a certain level, $y^*$ jumps onto another vertex. Concretely, this means that $y^*(\theta)$ is piece-wise constant with respect to $\theta$: the Jacobian $J_\theta y^*$ is null almost everywhere and undefined otherwise. This is why the rasterization (2) and aggregation (3) steps make renderers non-differentiable. Following [4], $y^*(\theta)$ can be approximated by the perturbed optimizer:

$$y^*_\epsilon(\theta) = \mathbb{E}_Z[y^*(\theta + \epsilon Z)]$$  

(4)

$$ = \mathbb{E}_Z \left[ \text{argmax}_{y \in C} \langle \theta + \epsilon Z, y \rangle \right].$$  

(5)

where $Z$ is a random noise following a distribution of the form $\mu(z) \propto \exp(-\nu(z))$. Then, we have $y^*_{\epsilon}(\theta) \xrightarrow{\epsilon \to 0} y^*(\theta)$, $y^*_{\epsilon}(\theta)$ is differentiable and its gradients are non-null everywhere. Intuitively, one can see from (5) that perturbed optimizers $y^*_{\epsilon}$ are actually a convolved version of the rigid optimizer $y^*$. Moreover, their Jacobian with respect to $\theta$ is given by:

$$J_\theta y^*_\epsilon(\theta) = \mathbb{E}_Z \left[ \frac{y^*(\theta + \epsilon Z)}{\epsilon} \nabla \nu(Z) \right].$$  

(6)

It is worth noticing at this stage that, when $\epsilon = 0$, one recovers the standard formulation of the rigid optimizer. In general, $y^*_{\epsilon}(\theta)$ and $J_\theta y^*_\epsilon(\theta)$ do not admit closed-form expressions. To overcome this issue, Monte-Carlo estimators are exploited to compute an estimate of these quantities, as recalled in Sec. [4].

### 3 Related work

Our work builds on results in differentiable rendering, differentiable optimization and randomized smoothing.
Differentiable rendering. Some of the earliest differentiable renderers rely on the rigid rasterization-based rendering process recalled in the previous section, while exploiting some gradient approximations of the non-smooth operations in the backward. OpenDR [24] uses a first-order Taylor expansion to estimate the gradients of a classical renderer, resulting in gradients concentrated around the edges of the rendered images. In the same vein, NMR [16] proposes to avoid the issue of local gradients by doing manual interpolation during the backward pass. For both OpenDR and NMR, the discrepancy between the function evaluated in the forward pass and the gradients computed in the backward pass may lead to unknown or undesired behaviour when optimizing over these approximated gradients. More closely related to our work, SoftRas [23] directly approximates the forward pass of usual renderers to make it naturally differentiable. Similarly, DIB-R [8] introduces the analytical derivatives of the faces’ barycentric coordinates to get a smooth rendering of foreground pixels during the forward pass, and thus, gets gradients naturally relying on local properties of meshes. Additionally, a soft aggregation step is required to backpropagate gradients towards background pixels in a similar way to SoftRas. In a parallel line of work, physically realistic renderers are made differentiable by introducing a stochastic estimation of the derivatives of the ray tracing integral [21, 28]. This research direction seems promising as it allows to generate images with global illumination effects. However, this class of renderers remains computationally expensive when compared to rasterization-based algorithms, which makes them currently difficult to exploit within classic computer vision or robotics contexts, where low computation timings matter.

Differentiable optimization and randomized smoothing. More generally, providing ways to differentiate solutions of constrained optimization problems has been part of the recent growing effort to introduce more structured operations in the differentiable programming paradigm [33], going beyond classic neural networks. A first approach consists in automatically unrolling the optimization process used to solve the problem [11, 26]. However, because the computational graph grows with the number of optimization steps, implicit differentiation, which relies on the differentiation of optimality conditions, should be preferred when available [2, 3, 20]. These methods compute exact gradients whenever the solution is differentiable with respect to the parameters of the problem. In this paper, we show how differentiating through a classic renderer relates to computing derivatives of a Linear Programming (LP) problem. In this particular case and as recalled in the Sec. 2 solutions of the optimization problem vary in a heavily non-smooth way as gradients are null almost-everywhere and non-definite otherwise, thus, making them hard to exploit within classic optimization algorithms. To overcome this issue, a first approach consists in introducing a gradient proxy by approximating the solution with a piecewise-linear interpolation [35] which can also lead to null gradient. Another approach leverages random perturbations to replace the original problem by a smoother approximation [1, 4]. For our differentiable renderer, we use this later method as it requires only little effort to transform non-smooth optimizers into differentiable ones, while guaranteeing non-null gradients everywhere. Moreover, randomized smoothing has been shown to facilitate optimization of non-smooth problems [12]. Solving a smooth approximation leads to improved performance when compared to methods dealing with the original rigid problem. In addition to its smoothing effect, the addition of noise also acts as an implicit regularization during the training process which is valuable for globalization aspects [4, 9], i.e. converging towards better local optima.

4 Approximate differentiable rendering via perturbed optimizers

In this section, we detail the main contributions of the paper. We first introduce the reformulation of the rasterization and aggregation steps as perturbed optimizers, and propose a variance reduction mechanism to alleviate the computational burden inherent to these kinds of stochastic estimators. Additionally, we introduce an adaptive scheme to automatically adjust the smoothing parameters inherent to the rendering process.

4.1 Perturbed renderer: a general approximate and differentiable renderer

As detailed in Sec. 2 the rasterization step (2) can be directly written as the argmax of an LP. Similarly, the aggregation (3) step can be slightly modified to reformulate it as an argmax of an LP. Indeed, when \( y \geq 0 \), the hard constraint \( y_j = 0 \) if \( I^j_i = 0 \) is equivalent to \( I^j_i y_j > 0 \), which can be
approximated by adding a logarithmic barrier $y_j \ln I_j$ in the objective function as done in classical interior point methods \([6]\):

$$w_{ij}^\alpha(z) = \arg\max_{y \text{ s.t. } \|y\|_1 = 1, y \geq 0} \langle z + \frac{1}{\alpha} \ln(I_i), y \rangle,$$

because $\ln I_j \to 0^{-\infty}$, enforcing $w_{ij}^\alpha \to 0\to 0$ as $\alpha \to +\infty$ approximates the hard constraint and allows to retrieve the classical formulation (3) of $w$. Using this formulation, it is possible to introduce a differentiable approximation of the rasterization and aggregation steps:

$$\hat{I}_i = H_\sigma(d(i, j)), \quad \text{with } H_\sigma(x) = \mathbb{E}_X[H(x + \sigma X)] = \mathbb{E}_X[\arg\max_{0 \leq y \leq 1} \langle y, x + \sigma X \rangle],$$

$$\hat{R}_i = \sum_j w_{ij,\sigma}(z)C_{ij}, \quad \text{with } w_{ij,\sigma}(z) = \mathbb{E}_Z[w_{ij}(z + \gamma Z)]$$

Using this formulation, it is possible to introduce a differentiable approximation of the rasterization and aggregation steps:

$$\hat{I}_i = H_\sigma(d(i, j)), \quad \text{with } H_\sigma(x) = \mathbb{E}_X[H(x + \sigma X)] = \mathbb{E}_X[\arg\max_{0 \leq y \leq 1} \langle y, x + \sigma X \rangle],$$

$$\hat{R}_i = \sum_j w_{ij,\sigma}(z)C_{ij}, \quad \text{with } w_{ij,\sigma}(z) = \mathbb{E}_Z[w_{ij}(z + \gamma Z)]$$

By proceeding this way, we get a rendering process for which every pixel is influenced by every triangle of meshes (Fig. 1, 2), similarly to SoftRas \([23]\). More concretely, as shown in \([4]\), the gradients obtained with the randomized smoothing are guaranteed to be non-null everywhere unlike some existing methods \([8, 16, 24]\). This makes it possible to use the resulting gradients as a strong supervision signal for optimizing directly at the pixel level, as shown through the experiments in Sec. 5. At this stage, it is worth noting that the prior distribution on the noise $Z$ used for smoothing is not fixed. One can choose various distributions as a prior thus leading to different smoothing patterns, which makes the approach versatile. Indeed, as a different noise prior induces different smoothing, some specific choices allow to retrieve some of the existing differentiable renderers \([16, 23, 24]\). In particular, using a Logistic and a Gumbel prior respectively on $X$ and $Z$ leads to SoftRas\([23]\). Further details are included in Appendix A.

More practically, the smoothing parameters $\sigma, \gamma$ appearing in differentiable renderers such as SoftRas \([23]\) or DIB-R \([8]\) can be hard to set. In contrast, in the proposed approach, the smoothing is naturally interpretable as a noise acting directly on positions of mesh triangles. Thus, the smoothing parameters representing the noise intensity can be scaled automatically according to the object dimensions. In Sec. 4.3, we notably propose a generic approach to automatically adapt them along the optimization process.

### 4.2 Exploiting the noise inside a perturbed renderer

In practice, (5) and (6) are useful even in the cases when the choice of prior on $Z$ does not induce any analytical expression for $y^*_\epsilon$ and $J_\theta y^*_\epsilon$. Indeed, Monte-Carlo estimators of these two quantities can be directly obtained from these two expressions:

$$y^M_\epsilon(x) = \frac{1}{M} \sum_{i=1}^M y^*(\theta + \epsilon Z^{(i)})$$

$$J_\theta y^M_\epsilon(z) = \frac{1}{M} \sum_{i=1}^M y^*(\theta + \epsilon Z^{(i)}) \frac{1}{\epsilon} \nabla \nu(Z^{(i)})^\top$$
Approximately, we have $\text{Var}[\theta] \approx \frac{\text{Var}[y^*(\theta + \epsilon Z)]}{\epsilon^2}$, which naturally leads to a variance-reduced Monte-Carlo estimator of the Jacobian:

$$J_{\theta y^c_\epsilon}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left( y^*(\theta + \epsilon Z^{(i)}) - y^*(\theta) \right) \frac{1}{\epsilon} \nabla \nu(Z^{(i)^T}) .$$  

The computation of $y^c_\epsilon$ requires the solution of $M$ perturbed problems instead of only one in the case of classical rigid optimizer. Fortunately, this computation is naturally parallelizable, leading to a constant computation time at the cost of an increased memory footprint. Using our variance-reduced estimator alleviates the need for a high number of Monte Carlo samples, hence reducing the computational burden inherent to perturbed optimizers (Fig. 3). As shown in Sec. 5 (Fig. 5, left), $M = 8$ samples is already sufficient to get stable and accurate results. In addition, computations from the forward pass can be reused in the backward pass and time and memory complexities to evaluate these passes are comparable to classical differentiable renderers (Tab. 1). Consequently, evaluating gradients does not require any extra computation. It is worth mentioning at this stage that this variance reduction mechanism in fact applies to any perturbed optimizer. This variance reduction technique can be interpreted as estimating a sub-gradient with the finite differences in a random direction and is inspired by the field of random optimization [27].

### 4.3 Making the smoothing adaptive with sensitivity analysis

In a way similar to [6], it is possible to formulate the sensitivity of the perturbed optimizer with respect to the smoothing parameter $\epsilon$:

$$J_{\epsilon y^c_\epsilon}(\theta) = \mathbb{E}_Z \left[ y^*(\theta + \epsilon Z) \left( \nabla \nu(Z)^T Z - 1 \right) / \epsilon \right]$$  

and, as previously done, we can build a variance-reduced Monte-Carlo estimator of this quantity:

$$J_{\epsilon y^c_\epsilon}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left( y^*(\theta + \epsilon Z^{(i)}) - y^*(\theta) \right) \frac{1}{\epsilon} \nabla \nu(Z^{(i)^T}) Z^{(i)} - 1 / \epsilon$$  

where $M$ is the total number of samples used to compute the approximations.

Approximately, we have $\text{Var}[J_{\theta y^c_\epsilon}(z)] \propto \frac{\text{Var}[y^*(\theta + \epsilon Z) \nabla \nu(Z^{(i)^T})]}{M \epsilon^2}$. When decreasing $\epsilon$, the variance of the Monte-Carlo estimator increases, which can make the gradients very noisy and difficult to exploit in practice. For this reason, we use the control variates method [19] in order to reduce the variance of our estimators. Because the quantity $y^*(\theta) \nabla \nu(Z)^T$ has a null expectation when $Z$ has a symmetric distribution and is positively correlated with $y^*(\theta + \epsilon Z) \nabla \nu(Z)^T$, we can rewrite (6) as:

$$J_{\theta y^c_\epsilon}(\theta) = \mathbb{E}_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \nabla \nu(Z)^T / \epsilon \right] ,$$  

which naturally leads to a variance-reduced Monte-Carlo estimator of the Jacobian:

$$J_{\theta y^c_\epsilon}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \left( y^*(\theta + \epsilon Z^{(i)}) - y^*(\theta) \right) \frac{1}{\epsilon} \nabla \nu(Z^{(i)^T}) .$$  

Figure 3: Control variates methods [19] allow to reduce the variance of the gradient of the smooth argmax. Left: Gaussian-perturbed argmax operator, Middle: estimator of gradient Right: variance-reduced estimator of gradient.

<table>
<thead>
<tr>
<th># of samples</th>
<th>1</th>
<th>2</th>
<th>8</th>
<th>32</th>
<th>64</th>
<th>SoftRas</th>
<th>Hard renderer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward (ms)</td>
<td>30 (+3)</td>
<td>30 (+3)</td>
<td>31 (+3)</td>
<td>31 (+3)</td>
<td>31 (+3)</td>
<td>29 (+3)</td>
<td>29 (+3)</td>
</tr>
<tr>
<td>Backward (ms)</td>
<td>19 (+1)</td>
<td>19 (+1)</td>
<td>19 (+2)</td>
<td>20 (+1)</td>
<td>22 (+1)</td>
<td>18 (+1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Max memory (Mb)</td>
<td>217.6</td>
<td>217.6</td>
<td>217.6</td>
<td>303.1</td>
<td>440.0</td>
<td>180.3</td>
<td>178.5</td>
</tr>
</tbody>
</table>
In the case of differentiable rendering, a notable property is the positivity of the sensitivity of any locally convex loss $\mathcal{L}$ (which is valid for the RGB loss) with respect to the $\gamma$ smoothing parameter when approaching the solution. Indeed, in the neighborhood of the solution, we have $\theta \approx \theta_t$ and a first-order Taylor expansion gives:

$$\frac{\partial \mathcal{L}}{\partial \gamma} (\theta_t) \approx J_{\gamma} w_\gamma(z)^\top C^\top \nabla^2 \mathcal{L}(R(\theta_t)) C J_{\gamma} w_\gamma(z) \geq 0,$$

(17)

because for a locally convex loss, $\nabla^2 \mathcal{L}(R(\theta_t))$ is positive definite near a local optimum. We exploit this property to gradually reduce the smoothing when needed. To do so, we track an exponential moving average of the sensitivity during the optimization process:

$$v_t^\gamma = \beta^\gamma v_{t-1}^\gamma + (1 - \beta^\gamma) \frac{\partial \mathcal{L}}{\partial \gamma} (\theta_t)$$

(18)

where $\beta^\gamma$ is a scalar parameter. Whenever $v_t^\gamma$ is positive, the smoothing $(\sigma, \gamma)$ is decreased at a constant rate (Fig. 4). Note that the adaptive algorithm does not require any further computation as sensitivity can be obtained from backpropagation and also beneficially applies to other differentiable renderers [23] (see 5).

5 Results

In this section, we explore applications of the proposed differentiable rendering pipeline to standard computer vision tasks: single-view 3D pose estimation and 3D mesh reconstruction. Our implementation is based on Pytorch3D [31] and will be publicly released upon publication. It corresponds to a modular differentiable renderer where switching between different type of noise distributions (Gaussian, Cauchy, etc.) is possible. We notably compare our renderer against the open source SoftRas implementation available in Pytorch3d and DIB-R from Kaolin library. The results for these renderers are obtained by running the experiments in our setup as we were not able to get access to the original implementations of the pose optimization and shape reconstruction problems.

Single-view 3D pose estimation is concerned with the retrieving the 3D rotation $\theta \in SO(3)$ of an object from a reference image. Similarly to [23], we first consider the case of a colored cube and fit the rendered image of this cube to a single view of the true pose $R(\theta_t)$ (Fig. 4). The problem can be formulated as a regression problem of the form:

$$\min_{\theta} \mathcal{L}_{RGB}(\theta) = \frac{1}{2} \| R(\theta_t) - \hat{R}(\theta) \|^2,$$

(19)

where $\hat{R}$ is the smooth differentiable rendered image. We aim at analysing the sensitivity of our perturbed differentiable renderer with respect to random initial guesses of various amplitudes from the real pose. Thus, the initial $\theta$ is randomly perturbed from the true pose with various intensity for the perturbation (exploiting an angle-axis representation of the rotation, with a randomized rotation axis). The intensity of the angular perturbation is taken between 20 and 80 degrees. Consequently, the value of the average final error has an increased variance and is often perturbed by local minima (Tab. 2). To avoid these caveats, we rely on another metric: the percentage of solved tasks (see Tab. 2 and Fig. 5 and 6), accounting for the amount of final errors which are below a given threshold (10 degrees for Tab. 2 and bottom row of Fig. 6). For this task, we compare our method with two different distributions for the smoothing noise (Gaussian and Cauchy priors) to SoftRas. We use Adam [17] with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and operate with 128×128 RGB images, each optimization problem taking about 1 minute to solve on a Nvidia RTX6000 GPU.
Table 2: Results of pose optimization from single-view image using perturbed differentiable renderer with variance reduction and adaptive smoothing.

<table>
<thead>
<tr>
<th>Initial error</th>
<th>20°</th>
<th>50°</th>
<th>80°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diff. renderer</td>
<td>SoftRas</td>
<td>Cauchy smoothing</td>
<td>Gaussian smoothing</td>
</tr>
<tr>
<td>Avg. error (°)</td>
<td>5.5</td>
<td>3.8</td>
<td>11.6</td>
</tr>
<tr>
<td>Std. error (°)</td>
<td>4.9</td>
<td>4.4</td>
<td>26.6</td>
</tr>
<tr>
<td>Task solved (%)</td>
<td>84 ± 4</td>
<td>93 ± 0.4</td>
<td>84 ± 1.3</td>
</tr>
</tbody>
</table>

Figure 5: **Left**: Variance reduction (VR) drastically reduces the number of Monte-Carlo samples required to retrieve the true pose from a 20° perturbation. **Right**: Adaptive smoothing (AS) improves resolution of precise pose optimization for a 80° perturbation.

We provide results of 100 random perturbations for various magnitude of initial perturbations in Tab. 2 and Fig. 5. Error bars provide standard deviations while running experiments with different random seeds. Fig. 5 (left) demonstrates the advantage of our variance reduction method to estimate perturbed optimizers and precisely retrieve the true pose at a lower computational cost. Indeed, the variance-reduced perturbed renderer is able to perform precise optimization with only 2 samples (80% of final errors are under 5° while it is less than 50% without the use of control variates). Additionally, our adaptive smoothing significantly improves the number of solved tasks for both SoftRas and our perturbed renderer (Fig. 5 right).

Additional results in Tab. 2, Fig. 6 and Fig. 11 in Appendix demonstrate that our perturbed renderers with a Gaussian smoothing, outperforms SoftRas and can achieve state-of-the-art results for pose optimization tasks on various objects. As shown in Appendix (Prop. 4), SoftRas is equivalent to using a Gumbel smoothing which is an asymmetric prior. However, no direction should be preferred when smoothing the rendering process which explain the better performances of the Gaussian prior.

Figure 6: **Top row**: perturbed renderer combined with variance reduction and adaptive scheme improves SoftRas results on pose optimization. **Bottom row**: The method is robust w.r.t. initial smoothing values. Higher is better.
Figure 7: **Left:** A neural network is trained with a self-supervision signal from a differentiable rendering process. **Right:** Qualitative results from self-supervised 3D mesh reconstruction using a perturbed differentiable renderer. 1\textsuperscript{st} column: Input image. 2\textsuperscript{nd} and 3\textsuperscript{rd} columns: reconstructed mesh viewed from different angles.

<table>
<thead>
<tr>
<th>Plane</th>
<th>Bench</th>
<th>Dresser</th>
<th>Car</th>
<th>Chair</th>
<th>Display</th>
<th>Lamp</th>
<th>Speaker</th>
<th>Rifle</th>
<th>Sofa</th>
<th>Table</th>
<th>Phone</th>
<th>Vessel</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR [1]</td>
<td>58.5</td>
<td>48.7</td>
<td>74.1</td>
<td>71.3</td>
<td>41.4</td>
<td>55.5</td>
<td>36.7</td>
<td>67.4</td>
<td>55.7</td>
<td>60.2</td>
<td>39.1</td>
<td>76.2</td>
<td>59.4</td>
</tr>
<tr>
<td>SoftRas [23]</td>
<td>62.0</td>
<td>47.55</td>
<td>66.2</td>
<td>69.4</td>
<td>49.4</td>
<td>60.0</td>
<td>43.3</td>
<td>62.6</td>
<td>61.4</td>
<td>60.4</td>
<td>43.6</td>
<td>76.4</td>
<td>59.9</td>
</tr>
<tr>
<td>DIB-R [8]</td>
<td>63.5</td>
<td>49.4</td>
<td>67.1</td>
<td>72.3</td>
<td>51.6</td>
<td>60.4</td>
<td>44.3</td>
<td>62.8</td>
<td>65.7</td>
<td>66.7</td>
<td>49.2</td>
<td>80.2</td>
<td>60.6</td>
</tr>
<tr>
<td>Ours</td>
<td>63.5</td>
<td>49.4</td>
<td>67.1</td>
<td>72.3</td>
<td>51.6</td>
<td>60.4</td>
<td>44.3</td>
<td>62.8</td>
<td>65.7</td>
<td>66.7</td>
<td>49.2</td>
<td>80.2</td>
<td>60.6</td>
</tr>
</tbody>
</table>

Bottom row of Fig. 6 shows that our method is robust to the choice of initial values of smoothing parameters. During optimization, the stochasticity of the gradient due to the noise from the perturbed renderer acts as an implicit regularization of the objective function. This helps to avoid sharp local minima or saddle points and converge to more stable minimal regions, leading to better optima. As mentioned earlier, results are limited by the inherent non-convexity of the rendering process and differentiable renderers are not able to recover from a bad initial guess. Increasing the noise intensity to further smoothen the objective function would not help in this case, as it would also lead to a critical loss of information on the rendering process.

**Self-supervised 3D mesh reconstruction from a single image.** In our second experiment, we demonstrate the ability of our perturbed differentiable renderer to provide a supervisory signal for training a neural network. We use the renderer to self-supervise the reconstruction of a 3D mesh and the corresponding colors from a single image (see Fig. 7). To do so, we use the network architecture proposed in [23] and the subset of the Shapenet dataset [7] from [16]. The network is composed of a convolutional encoder followed by two decoders: one for the mesh reconstruction and another one for color retrieving (more details in Appendix B).

Using the same setup as described in [8, 16, 23], the training is done by minimizing the following loss:

\[
L = \lambda_{sil}L_{sil} + \lambda_{RGB}L_{RGB} + \lambda_{lap}L_{lap},
\]

where \(L_{sil}\) is the negative IoU between silhouettes \(I\) and \(\hat{I}\), \(L_{RGB}\) is the \(\ell_1\) RGB loss between \(R\) and \(\hat{R}\), \(L_{lap}\) is a Laplacian regularization term penalizing the relative change in position between neighbour vertices (detailed description in Appendix B). For training, we use Adam algorithm with a learning rate of \(10^{-4}\) and parameters \(\beta_1 = 0.9, \beta_2 = 0.999\). Even if the memory consumption of the perturbed renderer may limit the maximum size of batches, its flexibility makes it possible to switch to a deterministic approximation such as [23] to use larger batches in order to finetune the neural network, if needed.

We analyze the quality of obtained 3D reconstruction through a rendering pipeline by providing IoU measured on 13 different classes of the test set (see Tab. 3). We observe that our perturbed

\footnote{These numbers were obtained by running the renderer in our own setup (cameras, lightning, training parameters etc.) in order to have comparable results. This may explain the slight difference with the numbers from the original publications.}
renderer obtains state-of-the-art accuracies on this task. Error bars (representing the standard deviation obtained by retraining the model with different random seeds) confirm the stability of our method. Additionally, we illustrate qualitative results for colorized 3D mesh reconstruction from a single image in Fig. 7(right) confirming the ability of perturbed renderers to provide strong visual supervision signals.

6 Conclusion

In this paper, we introduced a novel approach to differentiable rendering leveraging recent contributions on perturbed optimizers. We demonstrated the flexibility of our approach grounded in its underlying theoretical formulation. The combination of our proposed perturbed renderers with variance-reduction mechanisms and sensitivity analysis enables for robust application of our rendering approach in practice. Our results notably show that perturbed differentiable rendering can reach state-of-the-art performance on pose optimization and is also competitive for self-supervised 3D mesh reconstruction. Additionally, perturbed differentiable rendering can be easily implemented on top of existing implementations without major modifications, while generalizing existing renderers such as SoftRas. As future work, we plan to embed our generic differentiable renderer within a differentiable simulation pipeline in order to learn physical models, by using visual data as strong supervision signals to learn physics.

Acknowledgments and Disclosure of Funding

We warmly thank Francis Bach for useful discussions. This work was supported in part by the HPC resources from GENCI-IDRIS(Grant AD011012215), the French government under management of Agence Nationale de la Recherche as part of the "Investissements d’avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute), and Louis Vuitton ENS Chair on Artificial Intelligence.

References


Appendices

A Differentiable perturbed renderer

This section describes some possible choice of priors for the smoothing noise of our differentiable perturbed renderers and discuss the eventual links with already existing renderers. In addition, we provide justification of the variance reduction and adaptive smoothing methods for different priors on the noise.

A.1 Perturbed renderers

The following propositions discuss some possible choices of noise prior for the perturbed renderer, exhibiting the links with already existing differentiable renderers [24, 16, 23].

Proposition 1. The sigmoid function corresponds to the perturbed Heaviside function with a logistic prior on noise:

\[
\text{sigmoid}(x) = \mathbb{E}[H(x + Z)]
\]

where \(Z \sim \text{Logistic}(0, 1)\)

Proof. First, we note that the sigmoid function correspond to the first weight from the softmax applied to the \(R^2\) vector \(\tilde{x} = \begin{pmatrix} x \\ 0 \end{pmatrix} \). Indeed, we have:

\[
\sigma(x) = \frac{1}{1 + \exp(-x)} = \frac{\exp(x)}{\exp(x) + \exp(0)}
\]

\[
= \text{softmax}(\tilde{x})^T \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]

From the previous proposition, with \(Z\) a random variable in \(R^2\) distributed with a Gumbel distribution, we have:

\[
\text{softmax}(\tilde{x}) = \mathbb{E}[\text{argmax}_{y \text{ s.t. } \|y\|_1 = 1} \langle y, \tilde{x} + \epsilon Z \rangle]
\]

\[
= \mathbb{E}[\text{argmax}_{y \text{ s.t. } \|y\|_1 = 1} y^T (x + \epsilon Z_1)]
\]

\[
= \mathbb{E}[\text{argmax}_{y \text{ s.t. } \|y\|_1 = 1} y^T (x + \epsilon (Z_1 - Z_2))]
\]

because \(x + \epsilon (Z_1 - Z_2) > 0\) is equivalent to \(x + \epsilon Z_1 > \epsilon Z_2\).

Finally, we can re-write:

\[
\sigma(x) = \mathbb{E}[\text{argmax}_{y \text{ s.t. } \|y\|_1 = 1} y^T (x + \epsilon \tilde{Z})^T \begin{pmatrix} 1 \\ 0 \end{pmatrix}]
\]

\[
= \mathbb{E}[\text{argmax}_{0 \leq y \leq 1} y(x + \epsilon \tilde{Z})] \quad \text{(28)}
\]

\[
= \mathbb{E}[H(x + \epsilon \tilde{Z})] \quad \text{(29)}
\]

where \(\tilde{Z}\) follows a logistic law.

Which allows to conclude that the sigmoid corresponds to the Heaviside function perturbed with a logistic noise. We could use the fact that the cumulative distribution function of the logistic distribution is the sigmoid function to make an alternative proof.

\[\square\]

Proposition 2. The affine approximation of the step function corresponds to the perturbed Heaviside function with an uniform prior on noise:

\[
H_{aff}(x) = \mathbb{E}[H(x + Z)]
\]

where \(Z \sim \mathcal{U}(-\frac{1}{2}, \frac{1}{2})\)

13
Proof. We have:

\[ \mathbb{E}[H(x+Z)] = P(x+Z > 0) \]  
\[ = P(Z > -x) \]  
\[ = \begin{cases} 
0 & \text{if } x \leq -\frac{1}{2} \\
\frac{1}{2} + \frac{1}{\pi} \arctan(x) & \text{if } -\frac{1}{2} < x < \frac{1}{2} \\
1 & \text{if } x \geq \frac{1}{2} 
\end{cases} \]  
\[ = H_{\text{aff}}(x) \]  
\[ \square \]

**Proposition 3.** The arctan approximation of the step function corresponds to the perturbed Heaviside function with a Cauchy prior on noise:

\[ \frac{1}{2} + \frac{1}{\pi} \arctan(x) = \mathbb{E}[H(x+Z)] \]  
where \( Z \sim \text{Cauchy}(0, 1) \)

Proof. We have:

\[ \mathbb{E}[H(x+Z)] = P(x+Z > 0) \]  
\[ = P(Z > -x) \]  
\[ = \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} \, dz \]  
\[ = \frac{1}{2} + \frac{1}{\pi} \arctan(x) \]  
\[ \square \]

**Proposition 4.** The softmax function corresponds to the perturbed maximal coordinates with a Gumbel prior on the noise:

\[ \text{softmax}(z) = \mathbb{E}[ \arg\max_{y \text{ s.t. } \|y\|_1 = 1} \langle y, z + Z \rangle] \]  
where \( Z \sim \text{Gumbel}(0, 1) \)

Proof. The Gumbel-max trick is a classical property of the Gumbel distribution stating: the maximizer of \( M \) fixed values \( z_j \) which have been perturbed by a noise i.i.d \( Z_i \) following a Gumbel distribution, follows an exponentially weighted distribution [14]. This can be written:

\[ P(\max_j z_j + Z_j = z_i + Z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \]  
where \( Z_i \) are i.i.d. and following a Gumbel distribution. Thus, for a noise \( Z \) following a Gumbel distribution, we have:

\[ \mathbb{E}[ \arg\max_{y \text{ s.t. } \|y\|_1 = 1} \langle y, z + \epsilon Z \rangle] = \sum_i e_i P(\arg\max_{y \text{ s.t. } \|y\|_1 = 1} \langle y, z + \epsilon Z \rangle = e_i) \]

\[ = \sum_i e_i P(\max_j z_j + Z_j = z_i + \epsilon Z_i) \]

\[ = \sum_i e_i \frac{\exp(z_i)}{\sum_j \exp(z_j)} \]

\[ = \text{softmax}(z) \]  
\[ \square \]
Figure 8: Modifying the prior on the noise leads to different approximations of a non-smooth operator. The sigmoid and affine approximations are obtained by using respectively a Logistic and a Uniform prior.

Following the previous results, one can observe that OpenDR [24] corresponds to a rasterization step where a uniform prior is used for the smoothing noise in order to get gradients during the backward pass. More precisely, the uniform distribution is taken with a support as large as a pixel size which allows to get non null gradients for pixels lying on the border of triangles. This can be linked to antialiasing techniques which are added to classical rendering pipeline to remove high frequencies introduced by the non-smooth rasterization. These techniques make the intensity of a pixel vary linearly with respect to the position of the triangle which can also be seen as an affine approximation of the rasterization. However, one should note that the density of the uniform distribution cannot be written in the form of $\mu(z) \propto \exp(-\nu(z))$ so the properties on differentiability from [4] does not apply. In particular, this approximation is non-differentiable at some points ($x = 1$ or $-1$) and can have null gradients (for $x < -1$ or $x > 1$) which is inducing localized gradients for OpenDR [24]. For this reason, the approach of NMR [16] actually consists in having a variable range for the distribution support in order to get non-null gradients for a wider region. This approach corresponds to the affine interpolation proposed in [35] and can lead to null gradients. More closely related to our approach, SoftRas smoothen the rasterization by using a sigmoid approximation while Z-buffering is replaced by a softmax during aggregation. As shown by propositions 1 and 4, this approach actually corresponds to a special case of a perturbed renderer where rasterization and aggregation steps were smoothen by using respectively a Logistic and a Gumbel prior for the noise.

In short, this means that some existing renderers [24, 23, 16] can be interpreted in the framework of differentiable perturbed renderers when considering specific priors for the noise used for smoothing. In addition, the generality of the approach allows to consider a continuous range of novel differentiable renderers in between the already existing ones.

A.2 Variance reduction

As introduced in the paper, control variates methods can be used to reduce the noise of the Monte-Carlo estimators of the Jacobian of a perturbed renderer, without inducing any extra-computation. In particular, we prove this in the case of Gaussian and Cauchy priors on the smoothing noise.

**Proposition 5.** Jacobian of perturbed renderers can be written as:

$$J_{\theta}y_\epsilon(\theta) = \mathbb{E}_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \nabla \nu(Z)^\top \frac{1}{\epsilon} \right],$$  

when $Z$ follows a Gaussian distribution.
Proof. As done in \cite{4,1}, with a change of variable \( z' = \theta + \epsilon z \) we have:

\[
J_\theta y^*_\epsilon(\theta) = \frac{\partial}{\partial \theta} \left( \int_{-\infty}^{\infty} y^*(\theta + \epsilon z) \mu(z) dz \right),
\]

(47)

\[
= \int_{-\infty}^{\infty} y^*(z') \frac{\partial}{\partial \theta} \mu \left( \frac{z' - \theta}{\epsilon} \right) dz',
\]

(48)

\[
= \int_{-\infty}^{\infty} y^*(z') \frac{1}{\epsilon} \nabla \nu \left( \frac{z' - \theta}{\epsilon} \right)^\top \mu \left( \frac{z' - \theta}{\epsilon} \right) dz'.
\]

(49)

And by doing the inverse change of variable, we get:

\[
J_\theta y^*_\epsilon(\theta) = E_Z \left[ \frac{y^*(\theta + \epsilon Z)}{\epsilon} \nabla \nu(Z)^\top \right].
\]

(50)

In addition, for \( Z \) following a standard Gaussian distribution we have \( \nu(Z) = \frac{\|Z\|^2}{2} \) and thus:

\[
E_Z \left[ y^*(\theta) \nabla \nu(Z)^\top \right] = y^*(\theta) E_Z \left[ Z^\top \right]
\]

(51)

and because \( Z \) is centered, we get:

\[
E_Z \left[ y^*(\theta) \nabla \nu(Z)^\top \right] = 0
\]

(52)

Finally we get:

\[
J_\theta y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \nabla \nu(Z)^\top \frac{1}{\epsilon} \right].
\]

(53)

\begin{proposition}
Proposition 6. Jacobian of perturbed renderers can be written as:

\[
J_\theta y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \nabla \nu(Z)^\top \frac{1}{\epsilon} \right],
\]

(54)

when \( Z \) follows a Cauchy distribution.
\end{proposition}

\begin{proof}
Proceeding in a similar way to \cite{5} for a Cauchy distribution we have \( \nu(z) = \ln(1 + z^2) \), thus:

\[
E_Z \left[ y^*(\theta) \nabla \nu(Z)^\top \right] = y^*(\theta) E_Z \left[ Z^\top \frac{2Z}{1 + \|Z\|^2} \right]
\]

(55)

\[
= 0
\]

(56)

and finally we have:

\[
J_\theta y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \nabla \nu(Z)^\top \frac{1}{\epsilon} \right],
\]

(57)

for \( Z \) following a Cauchy distribution.
\end{proof}

A.3 Adaptive smoothing

In the same way as we proceeded for Jacobians computation, control variates method can be used to reduce the variance of Monte-Carlo estimators of the sensitivity of perturbed optimizers with respect to smoothing parameter \( \epsilon \).

\begin{proposition}
Proposition 7. The sensitivity of the perturbed optimizer with respect to the smoothing parameter \( \epsilon \) can be expressed as:

\[
J_\epsilon y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \frac{\nabla \nu(Z)^\top Z}{\epsilon} - 1 \right]
\]

(58)

for \( Z \) a smoothing noise following a Gaussian prior.
\end{proposition}
Proof. In a similar way to \(5\), using the same change of variable, we get:

\[
J_\theta y^*_\epsilon(\theta) = \frac{\partial}{\partial \epsilon} \left( \int_{-\infty}^{\infty} y^*(\theta + \epsilon z) \mu(z) \, dz \right),
\]

\[
= \int_{-\infty}^{\infty} y^*(z') \frac{\partial}{\partial \epsilon} \mu \left( \frac{z' - \theta}{\epsilon} \right) \, dz',
\]

\[
= \int_{-\infty}^{\infty} y^*(z') \frac{1}{\epsilon^2} \left( \nabla \nu \left( \frac{z' - \theta}{\epsilon} \right) - \frac{1}{\epsilon} \right) \mu \left( \frac{z' - \theta}{\epsilon} \right) \, dz',
\]

which yields with the inverse change of variable:

\[
J_\epsilon y^*_\epsilon(\theta) = E_Z \left[ y^*(\theta + \epsilon Z) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right]
\]

Then, for \(Z\) following a Gaussian distribution, we have:

\[
E_Z \left[ y^*(\theta) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right] = y^*(\theta) E_Z \left[ \frac{\|Z\|^2 - 1}{\epsilon} \right]
\]

\[
= 0
\]

for \(Z\) following a standard Gaussian noise. Finally, we can write:

\[
J_\epsilon y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right]
\]

Proposition 8. The sensitivity of the perturbed optimizer with respect to the smoothing parameter \(\epsilon\) can be expressed as:

\[
J_\epsilon y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right]
\]

for \(Z\) a smoothing noise following a Cauchy prior.

Proof. We adapt the previous proof to the case of a Cauchy distribution for the noise. We observe that:

\[
J_\epsilon y^*_\epsilon(\theta) = E_Z \left[ y^*(\theta + \epsilon Z) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right]
\]

remains valid. In addition, we have:

\[
E_Z \left[ y^*(\theta) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right] = y^*(\theta) \frac{1}{\epsilon} E_Z \left[ \frac{2\|Z\|^2}{1 + \|Z\|^2} - 1 \right]
\]

for \(Z\) following a Cauchy noise. Moreover, an integration by parts gives:

\[
\int_{-\infty}^{\infty} \frac{2x^2}{(1 + x^2)^2} \, dx = \left[ -x \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{1 + x^2} \, dx = [\arctan(x)]_{-\infty}^{\infty}
\]

so we have:

\[
E_Z \left[ \frac{2\|Z\|^2}{1 + \|Z\|^2} - 1 \right] = 0
\]

As previously done in Proposition 7, we finally get:

\[
J_\epsilon y^*_\epsilon(\theta) = E_Z \left[ (y^*(\theta + \epsilon Z) - y^*(\theta)) \frac{\nabla \nu(Z)^\top Z - 1}{\epsilon} \right]
\]

\[
\square
\]
In this section, we provide details on our experimental setup and some additional results. In addition, we provide the code for our differentiable perturbed renderer based on Pytorch3d [31] and the experiments on pose optimization while additional experiments will be made available upon publication.

B.1 Pose optimization

Trying to solve pose optimization from an initial guess taken uniformly in the rotation space leads to very variable results when measuring the average final error (Fig. 9). When the initialization is too far from the initial guess, the optimization process converges towards a local optima. This limitation is inherent to approaches using differentiable rendering to solve pose optimization tasks and could be avoided only by using a combination with other methods, e.g. Render & Compare [18], to get better initial guess for the pose. For this reason, we prefer to consider the ability of our differentiable renderer to retrieve the true pose with an initial guess obtained by perturbing the target and measure the amount of final errors below a given precision threshold (Fig. 10,11).

In addition to the pose optimization task on the cube, we propose to solve the same problem with others objects from Shapenet dataset [7] (Fig. 11).

B.2 Mesh reconstruction

Reconstruction network  

Network architecture is based on [23] and is represented in Fig. 12. The encoder is composed of 3 convolutional layers (each of them is followed by a batch normalization operations) and 3 fully connected layers. We use two different decoders to retrieve respectively the vertices’ position and their color. The positional decoder is composed of 3 fully connected layers. The decoder used for colors retrieval is composed of a branch selecting a palette of \( N_p \) colors on the input image while the second branch mixes these colors to colorize each of the \( N_v \) vertices.

Training setup  

We use the same setup as described in [8, 16, 23]. The training is done by minimizing a loss combining silhouette and color information with a regularization term:

\[
\mathcal{L} = \lambda_{sil} \mathcal{L}_{sil} + \lambda_{RGB} \mathcal{L}_{RGB} + \lambda_{lap} \mathcal{L}_{lap},
\]

and we set \( \lambda_{sil} = 1, \lambda_{RGB} = 1, \lambda_{lap} = 3 \times 10^{-3} \) during training.
Figure 10: Pose optimization results for a perturbed renderer with Gaussian and Cauchy priors on the smoothing noise, and for SoftRas.
Figure 11: Pose optimization on various objects: a bag (Top) and a mailbox (Bottom)

In more details, $L_{sil}$ is the negative IoU between silhouettes $I$ and $\hat{I}$ which can be expressed as:

$$L_{sil} = \mathbb{E} \left[ 1 - \frac{\| I \odot \hat{I} \|_1}{\| I + \hat{I} - I \odot \hat{I} \|_1} \right]$$  \hfill (75)

$L_{RGB}$ is the $\ell_1$ RGB loss between $R$ and $\hat{R}$:

$$L_{RGB} = \| R - \hat{R} \|_1$$  \hfill (76)

$L_{lap}$ is a Laplacian regularization term penalizing the relative change in position between a vertices $v$ and its neighbours $\mathcal{N}(v)$ which can be written:

$$L_{lap} = \sum_v \left( v - \frac{1}{|\mathcal{N}(v)|} \sum_{v' \in \mathcal{N}(v)} v' \right)^2$$  \hfill (77)

Additional results We provide additional qualitative results for colorized mesh reconstruction (Fig. 15).
Figure 12: Learning architecture for the task of colorized 3D mesh reconstruction. **First row:** Overview of reconstruction network. **Second row:** Encoder network. **Third row:** Positional decoder. **Fourth row:** Color decoder.
Figure 13: Results from unsupervised 3D mesh and color reconstruction. 1st and 4th column: Input image. Other columns: Reconstructed 3D mesh from different angles.