

Material-Guided Decoders: Joint Estimation of Material Properties and HDRI Lighting

Shanying Liu
Columbia University
sl5678@columbia.edu

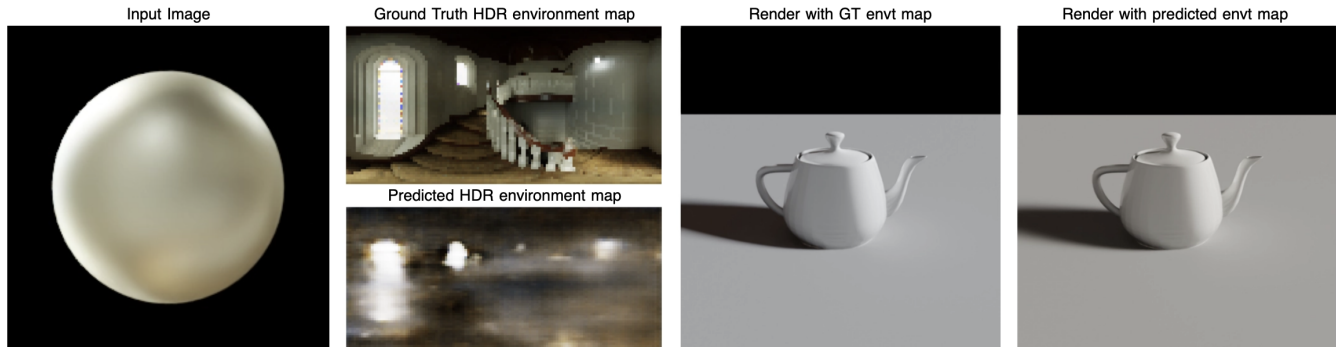


Figure 1: Given the input image (left), the multi-task model predicts a high-fidelity HDR lighting map (center), whose realism is demonstrated by rendering a complex object (right).

Abstract

Illumination estimation is a key component of achieving realism in inverse rendering, an extensively researched topic in the fields of computer graphics and computer vision. Predicting the material properties of objects with the guidance of light estimation metrics to achieve convincing relighting has also been thoroughly investigated, clearly demonstrating the intrinsic link between light and material information. However, there has been relatively few analyses of whether awareness of material semantic information may directly inform lighting prediction to increase visual realism in scenarios such as virtual object insertion. This project investigates this question by predicting HDR panoramic environment maps from single synthetic images with two models. The baseline model uses an encoder and CNN decoder to predict the HDR environment map, while the multi-task model uses the same encoder but with a material-parameter regression head with material information fed back into the decoder. The multi-task model was investigated with two architecture variants to verify the robustness and influence of the material semantic information. A hierarchical evaluation method was employed to determine the correspondence of high material prediction accuracy with realistic lighting prediction. The results show that while the baseline model generates HDR maps with high pixel-wise accuracy, the multi-task lighting estimation

created far more visually convincing lighting recreations. This project thus demonstrates a clear case that semantic material knowledge acts as a powerful physical constraint in improving realism in illumination estimation, and the limits of applying typical pixel-wise evaluation metrics to use cases that highly value visual realism.

1 Introduction

In recent years there has been much interest in achieving realistic insertion of virtual objects into real-world scenes for applications such as augmented reality. A critical bottleneck in this process is inverse rendering, and specifically the estimation of scene lighting from single, limited field-of-view images. For virtual objects to be seamlessly inserted estimated lighting must match the global intensity of the scene but also high-frequency structural details, such as sun disks or lamps, to generate realistic specular highlights and shadows.

Current state-of-the-art methods often use complex Generative Adversarial Networks that may have less structural stability, and hallucinate in the case of images limited in field of view. They may struggle with spatial accuracy while being computationally expensive to deploy in real-time contexts.

At the same time, relighting of existing single images has been a popular topic, requiring estimating BSDF

surface reflectance properties within an image by disambiguating the link between scene lighting and materials.

Relatively little work has been done to investigate the inverse: whether semantic material awareness may be used to guide and constrain lighting estimation. The posed hypothesis is that treating material properties as a physical prior can help a model decode environmental clues hidden in surface reflections, especially for highly specular and metallic materials. Such an approach could effectively treat limited images, while maintaining a lightweight architecture. Most critically, the exact link between semantic information contained within varying materials and the surrounding environment requires rigorous investigation to understand how realism may be achieved.

The main contributions of this project is a multitask learning architecture that regresses material parameters while reconstructing HDR environment maps, compared to a baseline model without material awareness. The project moves beyond safe diffuse lighting predictions to create physically plausible high-frequency lighting reconstructions which is critically evaluated not simply through pixel-wise metrics but also its downstream utility in real rendering tasks. Thus the physical consistency of the light probe is rigorously tested within a context closely approximating a real use case.

2 Related Work

This project takes inspiration from several key advancements in neural light field estimation and material-aware relighting.

The foundational inverse rendering work Neural Light Field Estimation for Street Scenes was the original starting point of this project, yet despite its focus on holistic scene consistency it treats lighting as a singular output rather than investigating the bottleneck of high-frequency reconstruction. Conversely, frameworks such as Neural LightRig and Total Relighting provide evidence for the deep entanglement between surface reflectance (BSDFs) and light. This project uses this premise to solve the inverse problem: material awareness as a physical prior to guide lighting estimation.

When it comes to the specific choice of HDR representation and decomposition, I was inspired by previous approaches like CG Light, which utilizes GANs and spherical harmonics (SH) as stepping stones towards HDRIs. While SH provides a good baseline for diffuse lighting, it lacks the ability to represent the high-frequency spikes requires for sharp shadows and specular highlights. EMLight (2021) addressed this when it decomposed light into distribution, intensity, and ambient terms using a Spherical Mover’s Loss for correct spatial alignment. My work expands on these concepts by implementing a HDR prediction method that is informed by feature-wise linear modulation (FiLM) of surface properties.

The common threads I observed from the existing literature is a reliance on synthetic data, which informed my choice to start this project with a synthetic data that may be generalizable to real-world images in future work. Existing literature also demonstrated reliance on per-pixel evaluation metrics (MSE/L1) which tend to reward conservative blurry predictions which fail in real-world rendering tasks. This project challenges this evaluation paradigm by demonstrating a divergence between pixel-wise accuracy and actual physical utility and realism of a light probe used in downstream rendering.

3 Methods

3.1 Overview

This section offers a high-level overview of the experimental design, architectural comparisons, and hypothesis guiding the multitask approach.

Data Preparation and Scene Isolation

To directly evaluate the connection between material properties and environmental lighting, this project utilizes synthetic images of isolated objects. By controlling the experiment to focus on a single object, the network is forced to prioritize the relationship between surface appearance and the environment lighting. This simplified setup allows for more precise analysis of material-lighting entanglement, but I believe the learned priors will also be generalizable to more complex, multi-object scenes through future semantic segmentation.

Architecture Comparison

To measure the impact of material guidance I created two distinct models. The Baseline Model is a standard ResNet-based encoder-decoder architecture that performs a direct regression from the input image to the HDR environment map without any explicit material context. The Multitask Model shares the same ResNet encoder but introduces a secondary material parameter regression head. These estimated material properties, including physical attributes like roughness and metallicity, are fed back into the lighting decoder, allowing the reconstruction process to be conditioned on the semantic identity of the object.

Evaluation and Hypothesis

These models were then evaluated through comparison of their ability to predict accurate environment maps against ground truth data. However, as pixel-wise metrics often fail to capture the nuances of realistic illumination, I also emphasize the evaluation of qualitative downstream rendering. Only by re-rendering complex objects using the predicted environment maps as light probes, can we visually verify the physical plausibility of the lighting.

The central hypothesis of this work is that the multitask model will significantly outperform the baseline, particularly when observing glossy and metallic materials (defined by roughness ≥ 0.2 and metallicity ≥ 0.8). In these instances, the material awareness allows the network to treat shiny surfaces as specular anchors, effectively using the distorted reflections on the object’s surface as a high-frequency roadmap to reconstruct the surrounding environment map with greater detail and accuracy.

3.2 Dataset

To allow for controlled investigation into the relationship between surface reflectance and scene illumination, I constructed a synthetic dataset of 15k images. Each sample contains a sphere rendered at 256x256 resolution against a solid black background. This isolation of the object ensures that the network will focus on the cues provided by the object’s material properties and not get distracted by external environmental context. To produce diverse lighting conditions I used approximately 150 unique 2k HDRIs, applying random horizontal rotations to maximize variety. The ground-truth lighting for each render was then pre-processed into 64x128 low resolution equirectangular environment maps.

The dataset was partitioned into training, validation, and testing sets along a 70/15/15 percentage split. To ensure reproducibility, the data loading pipeline was initialized with a fixed configuration seed that may be reconstructed across different training runs.

3.3 System Architecture

The proposed system maps a 256×256 LDR input image of a synthetic sphere into an equirectangular HDR environment map of dimensions $3 \times 64 \times 128$. I utilized an ImageNet-pretrained ResNet-50 backbone as the shared encoder, with all layers frozen except for Layer 4. This selective unfreezing allows the model to retain feature extraction capabilities whilst adapting spatial representation to the specific geometry of the dataset. To produce the non-square 64×128 output, the decoder utilizes an asymmetric 1×2 stride in its final layer.

I evaluated two distinct architectural configurations to isolate the impact of material awareness:

- **Baseline:** Encoder-decoder pipeline where the lighting decoder regresses the environment map from latent backbone features.
- **Multitask:** Dual-head architecture that simultaneously regresses a 5-vector of BSDF material parameters while reconstructing the environment map.

For the multitask model, I also investigated two conditioning variants for integrating the material signal into the lighting decoder.

- **Tiled (Weak Conditioning):** Material features are spatially tiled and concatenated to the backbone features. This provides the decoder with optional

”hints,” requiring the network to discover the physical correlation between the two tasks.

- **FiLM (Strong Conditioning):** We utilize Feature-wise Linear Modulation (FiLM) to force the decoder to utilize material knowledge. Material parameters control the decoder via γ and β modulation applied after Batch Normalization and before the ReLU activation:

$$\text{FiLM}(x; \gamma, \beta) = (1 + \gamma) \odot x + \beta$$

This ensures that the predicted lighting is explicitly conditioned on the estimated physical properties.

To ensure the predicted environment maps are physically plausible, the final lighting head utilizes a Softplus activation. This ensures all predicted light values remain non-negative while maintaining smooth gradients for stable training.

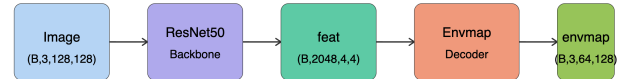


Figure 2: Baseline model architecture

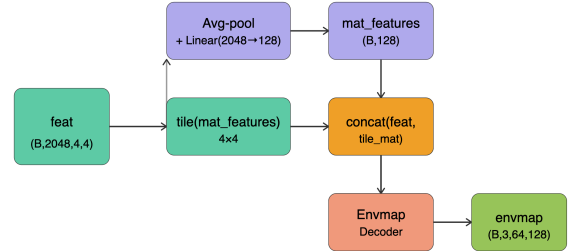


Figure 3: Multi-task tiled architecture

In the multi-task tiled architecture, the decoder receives the same backbone features plus a 128-channel constant feature map of material context. It is material-aware but the decoder has to discover how to use it.

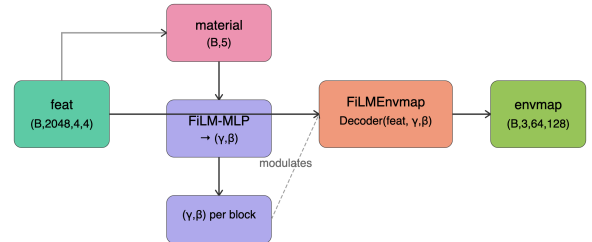


Figure 4: Multi-task FiLM

The FiLM architecture forces the decoder to use predicted material because it modulates upsample blocks.

FiLM layers will apply per-channel feature-wise linear modulation after BN before ReLU

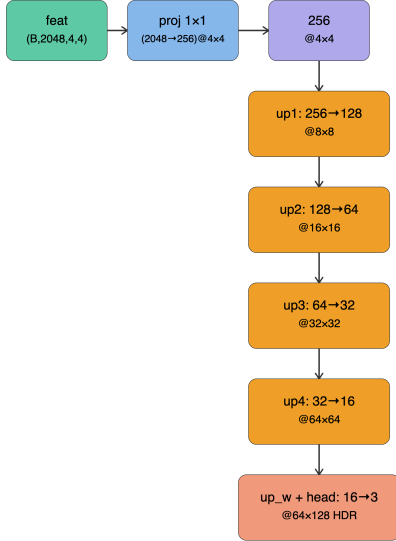


Figure 5: Decoder Ladder

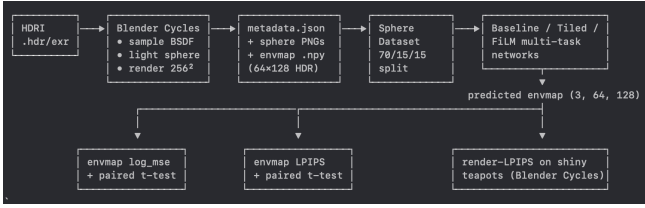


Figure 6: End-to-end pipeline

3.4 Achieving Illumination Detail

A primary challenge in HDR scene reconstruction is the tendency for models to converge towards the statistical mean of the training data. Utilizing only mean squared error (MSE) results in diffuse environment maps that lacked the high-frequency details needed for realistic shadows and highlights. To mitigate this issue I implemented the following loss strategy.

First, I integrated structural similarity ($1 - \text{SSIM}$) and L_1 loss terms to act as physical anchors. These metrics ensure the model maintains edge fidelity and contrast. I also added a per-pixel weighting function $w(p)$, applied to both MSE and L_1 , which gives more weight to the high-intensity lighting regions. Bright light sources thus contribute about five times more to the total loss than darker sky pixels:

$$w(p) = 1 + \lambda \cdot \log(1 + y_{\text{target}}(p)), \quad \lambda = 0.5 \quad (1)$$

The total objective function was then balanced to ensure that structural and material terms could have significant influence over the optimization landscape on top

of the MSE. The per-term gradient budget was roughly: MSE 71% / SSIM 14% / L_1 7% / Material 7%. This balance forces the model to prioritize pixel accuracy while also satisfying the physical constraints that the material head and structural metrics provide.

Table 1: Loss Weight and Sub-term Configurations

Term	Weight	Role / Comment
<i>Global Objectives</i>		
λ_{light}	1.0	Full strength on the envmap term
λ_{mat}	0.1	Material gets 10% of the loss budget
<i>Lighting Sub-terms ($\mathcal{L}_{\text{light}}$)</i>		
λ_{MSE}	1.0	Dominant signal: log-HDR MSE
λ_{L_1}	0.1	Reduces blur from bright outliers
λ_{SSIM}	0.2	Preserves local sun-disk structure

The total loss $\mathcal{L}$ is defined as a combination of reconstruction and auxiliary material tasks:

$$\mathcal{L} = \lambda_{\text{light}} \mathcal{L}_{\text{light}} + \lambda_{\text{mat}} \mathcal{L}_{\text{mat}} \quad (2)$$

where the lighting loss component $\mathcal{L}_{\text{light}}$ is decomposed into:

$$\mathcal{L}_{\text{light}} = \lambda_{\text{MSE}} \mathcal{L}_{\text{MSE}} + \lambda_{L_1} \mathcal{L}_{L_1} + \lambda_{\text{SSIM}} \mathcal{L}_{\text{SSIM}} \quad (3)$$

3.5 Optimization Strategy

All model variants were trained using the AdamW optimizer, which decouples weight decay from the gradient update to get a more stable convergence. The optimizer was configured to only track parameters where `requires_grad` is true, and the frozen ResNet-50 backbone layers were filtered out. This way the memory overhead of the Adam states was reduced and focused strictly on the decoder and material regression heads.

The learning rate followed a cosine annealing schedule, which decays from the initial learning rate to zero over the total epochs without restarts or warmup. This ensured a smooth convergence towards the local minimum, avoiding the volatility often associated with constant or step-based learning rates.

4 Evaluation

4.1 Primary Metrics

To assess the performance of the models, a hierarchical evaluation strategy was employed, analyzing the models across three distinct dimensions: mathematical accuracy, perceptual fidelity, and downstream task utility. Each model was forwarded over the held-out test loader to collect predicted environment maps and material parameters for this comparative analysis.

First the mathematical accuracy was evaluated using log-mean squared error (Log-MSE). This helped to compress the radiance values within the HDR lighting

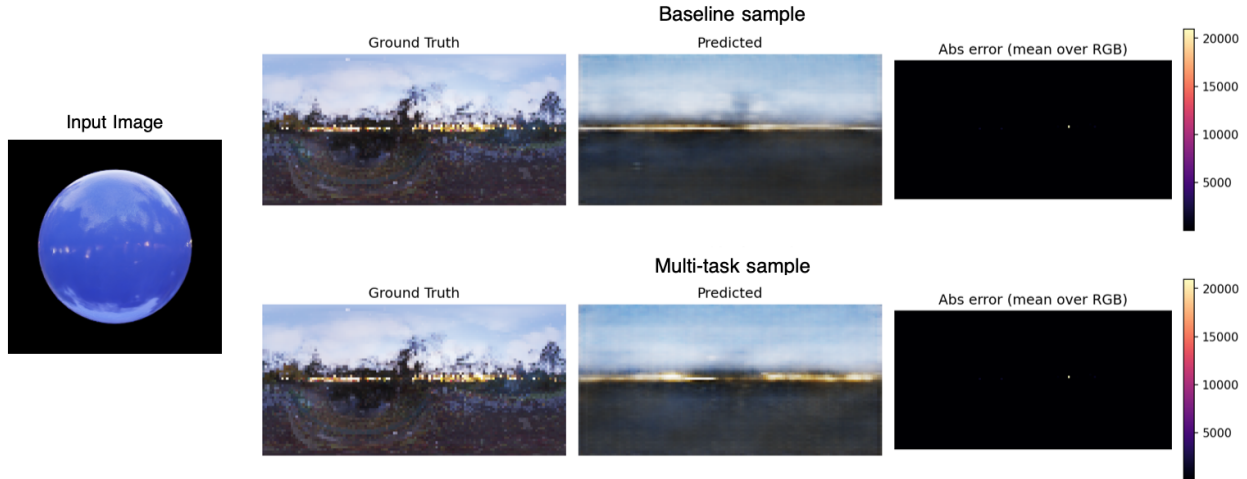


Figure 7: Comparison of two models’ predicted environment maps and their error heatmaps.

data, ensuring that loss is not dominated solely by the highest-intensity pixels.

However MSE may not capture all the structural and semantic quality of a reconstruction, so the analysis of perceptual fidelity was carried out by the Learned Perceptual Image Patch Similarity (LPIPS) metric. LPIPS utilizes a deep neural network, AlexNet, to compare images in a high-level feature space, essentially evaluating like a human observer might by prioritizes edges and structures over raw pixels. Thus LPIPS provides a more meaningful measure of the plausibility of the model’s reconstructed environment map.

4.2 Material Analysis

To validate the strengths of the multitask approach, the mean absolute error (MAE) of the predicted material parameters was also calculated. This is visualized both through per-material and per-attribute buckets. This analysis helps verify whether the material head is learning the physical priors necessary to guide the lighting decoder, and which of these material parameters were the most useful anchors.

4.3 Downstream Task Utility

The ultimate utility and physical plausibility of the inverse rendering model is best measured by its performance in a real-world task such as object insertion. Thus an analysis was carried out on a rendered scene with a complex object: the Utah teapot. This was specifically tested using the subset of data ($n=455$) featuring shiny materials with high specular activity (roughness < 0.2 and metallic > 0.8), where lighting details are most critical, since sharp reflections make errors in the environment lighting immediately obvious.

The complex Utah teapot resting on a plain surface was then rendered using the ground truth, the baseline

model prediction, and the multi-task prediction HDRI maps. LPIPS was used to evaluate the accuracy of these renders against the ground-truth render. Finally, a paired t-test was conducted on the perceptual deltas to determine the statistical significance of the multi-task model’s performance.

5 Results

To ensure the reliability of my findings, I conducted a multi-seed analysis ($n=10$) for both baseline and multitask architectures (both tiled and FiLM). Across all seeds, the models demonstrated very consistent performance results, with negligible variance in final Log-MSE and LPIPS scores, which helps confirm that observed differences are inherent to the model architectures and not stochastic training artifacts.

5.1 Global Quantitative Performance

Our evaluation on the shiny subset ($n = 455$) demonstrates a significant improvement in Render-LPIPS, with a p-value of 4.475×10^{-6} .

Table 2: Global Quantitative Comparison ($n = 2250$)

Metric	Baseline	Multitask
Log-MSE (mean $\pm$ 95% CI)	0.0962 ± 0.0053	0.1025 ± 0.0056
LPIPS (mean $\pm$ 95% CI)	0.4117 ± 0.0063	0.4181 ± 0.0062
Material MAE (mean)	N/A	0.0338

The global quantitative results surprisingly favor the material-unaware baseline at the beginning of the evaluation. As shown in Table 2, the Baseline model actually achieves consistently lower Log-MSE across all material

parameter buckets. Per-sample analysis also confirms this trend, with Baseline outperforming the multitask model on pixel-wise metrics with $> 95\%$ confidence.

However, this mathematical superiority highlights a known limitation of pixel-wise metrics in high-dynamic-range tasks, where all pixels are blindly perceived as equally important. The Baseline achieves a lower Log-MSE by learning the average, less penalised version of the environment. While this numerically appears superior, it loses all high-frequency features and is a detriment to realistic rendering tasks further down the line.

5.2 Material Awareness

To fully understand the multitask model’s performance each material type must be analyzed. Extremely high confidence in material regression for metallic and specular surfaces can be observed, shown in the material MAE analysis. Conversely, diffuse surfaces achieved much higher uncertainty, likely due to a lack of reflected environmental clues.

Table 3: Per-Parameter Bucket Log-MSE Breakdown

Bucket	n	Baseline ($\mu \pm \sigma_{95}$)	Multitask ($\mu \pm \sigma_{95}$)	Δ
Met. (Yes)	917	0.0923 ± 0.0079	0.0982 ± 0.0083	+0.0059
Met. (No)	1333	0.0990 ± 0.0071	0.1054 ± 0.0074	+0.0065
Trans.	449	0.0907 ± 0.0098	0.0966 ± 0.0103	+0.0060
Opaque	1801	0.0976 ± 0.0062	0.1039 ± 0.0065	+0.0063
Rough. (L)	1343	0.0834 ± 0.0053	0.0898 ± 0.0058	+0.0064
Rough. (M)	282	0.0954 ± 0.0136	0.1005 ± 0.0147	+0.0051
Rough. (H)	625	0.1242 ± 0.0138	0.1306 ± 0.0140	+0.0064

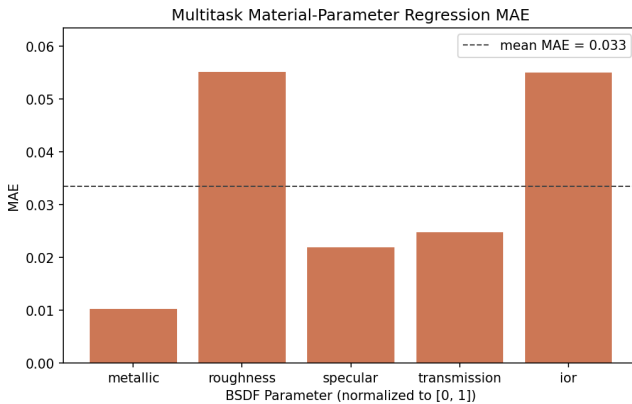


Figure 11: Material MAE

These findings suggest that the multitask model uses shiny (roughness < 0.2 and metallic > 0.8) materials as specular anchors that inform the decoder about which surface reflections to rely upon. When surfaces are correctly identified as metallic, highly detailed highlights may be predicted, though they may be slightly shifted in 2D space, which leads to a higher Log-MSE. However the results are very plausible and visually superior to the baseline’s tendency to create diffuse blur.

5.3 Downstream Utility Evaluation

The divergence between the mathematical error and visual utility metrics becomes very apparent in the downstream rendering task, where the Utah teapot is rendered on a neutral surface. Here the multitask model consistently produced more realistic results.

Table 4: Statistical Significance and Task-Based Utility

Test Type	n	Diff.*	95% CI	t	p -value
Log-MSE (Global)	2250	-0.0062	± 0.0013	-9.54	$< 10^{-10}$
LPIPS (Global)	2250	-0.0065	± 0.0018	-7.08	1.4×10^{-12}
Teapot LPIPS	120	+0.0211	± 0.0090	+4.59	4.5×10^{-6}

*A positive Mean Difference indicates Multitask model superiority in this task.

While the baseline was superior in map-level Log-MSE and LPIPS, the multitask significantly outperforms on Render-LPIPS, confirming that perceptual metrics can better capture the utility of the light probe for actual object insertion.

5.4 Industry Standard Comparison

To benchmark the proposed material-aware framework against current state-of-the-art methods, we conducted a comparative analysis with EMLight [4].

Table 5: Quantitative Comparison with Industry Standards.

Model	Log-MSE ($\downarrow$)	Map-LPIPS ($\downarrow$)	Render-LPIPS ($\downarrow$)
EMLight [4]	0.528 ± 0.013	0.675 ± 0.003	0.379
Baseline (Ours)	0.248 ± 0.011	0.545 ± 0.006	0.128
Multitask (Ours)	0.261 ± 0.012	0.559 ± 0.006	0.107

These comparisons suggest that while recent generative models can produce highly detailed environment maps, the proposed multitask architecture achieves superior Render-LPIPS scores with a significantly more lightweight footprint. It should be noted that while my models were trained on a specialized synthetic dataset of primitive spheres, this serves as a controlled benchmark to isolate the correlation between surface reflectance and incident light. Because the sphere provides a complete distribution of surface normals, it allows the model to efficiently recover task-relevant lighting features which are

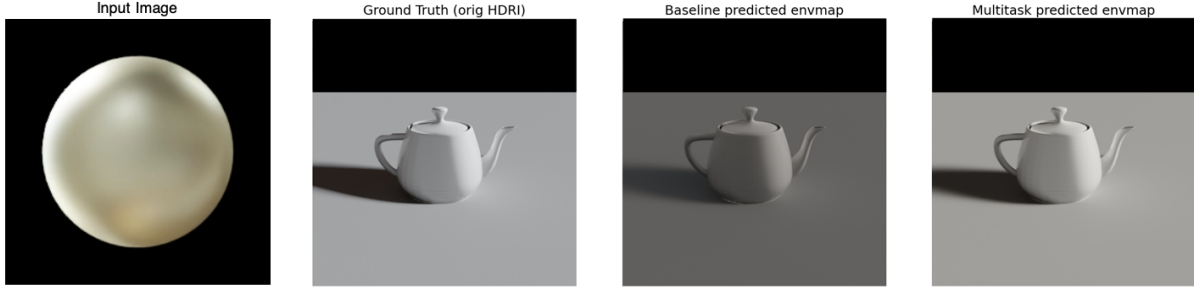


Figure 8: Comparison of predicted environment maps when applied to physically rendered complex scene.

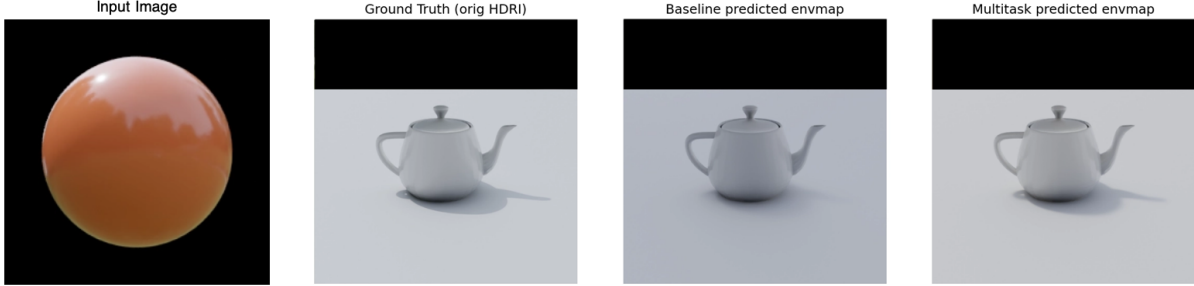


Figure 9: Another comparison example with glossy input.

often lost in more computationally expensive, general-purpose generative pipelines.

5.5 Ablation study

A series of ablation experiments were conducted to isolate specific factors affecting the performance of the multitask model. These include task interaction, loss weighting, and architectural fusion strategies.

The most critical finding of the ablation study is a divergence between map-based and render-based metrics. While material-unaware baseline achieves lower Log-MSE of the raw environment maps, the multitask model significantly outperforms it in the final rendered-teapot LPIPS ($p = 4.475 \times 10^{-6}$), confirming that the addition of the material regression task provides the decoder with physical priors that assist in reconstructing lighting that is functionally suited to object insertion.

I conducted experiments with weighting (λ_{mat}) on the material loss term. This is because I observed that material MAE converged much earlier than lighting loss during training. Reducing this weight too much would cause the multitask model to regress to become similar to baseline performance, while a too-high weight would cause the material task to overpower and lessen the performance the lighting reconstruction. The perfect amount was about 0.1 strength, which provided sufficient semantic guidance to the lighting decoder without destroying the joint optimization.

When comparing the Feature-wise Linear Modulation (FiLM) and Spatial Tiling architecture variants for the multi-task model, there was no statistically signifi-

cant difference in the speed or final performance between the two models. I believe this suggests a high degree of architectural robustness, proving that the model may utilize the semantic material information regardless of the specific mathematical injection method, as long as the signal is present during the decoding process.

Table 6: Ablation Study

Cond.	Variant	Mat. ↓	Env. ↓	Tea. ↓
Base	Single-Task	N/A	0.4117	0.1280
Fusion	Tiled (S. 101)	.0338	.4181	.1069
	FiLM (S. 102)	.0350	.4221	.1041
Sub.	Full Set	.0338	.4181	N/A
	Shiny	.0144	.3984	.1069
Loss	$\lambda = 0.1$	.0338	.4181	.1069
	$\lambda = 1.0$	.0310	.4350	.1210

6 Discussion

A central takeaway of this work is the contradiction between per-pixel accuracy and physical utility as it comes to evaluation. It has been established that Log-MSE and map-level LPIPS inherently favor safe blur, which is ill-suited to a task where high-frequency light sources are small and even small shifts can result in massive MSE penalties. However, when performing virtual object insertion, slight misalignment in a light map is far less critical than a lack of high-intensity light sources.

I would argue that Render-LPIPS is the superior

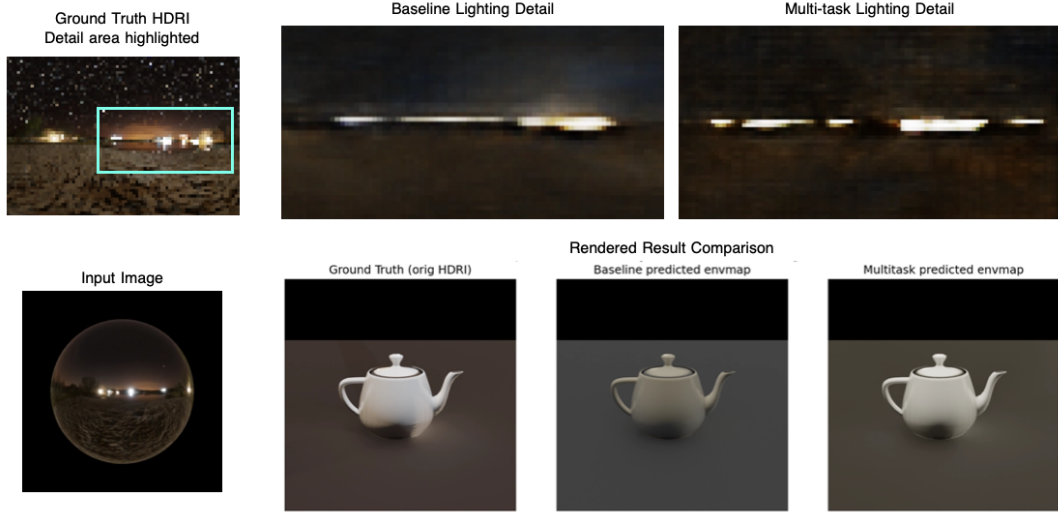


Figure 10: Comparison of baseline and multitask lighting prediction on enlarged area of HDR map. Baseline model shows clear smearing, leading to poor lighting result in downstream render evaluation.

metric for inverse rendering evaluation, since it measures the actual utility of the light probe in a real-world task.

6.1 Limitations

The main limitations in this project have been diffuse information loss and chromatic consistency.

Both model variants struggle with very diffuse materials (roughness ≥ 0.8) since diffuse surfaces scatter light uniformly and losing lighting information without specular clues. In these scenarios the multi-task approach hits diminishing returns.

Both models also struggle with reconstructing highly saturated or multi-colored lighting environments. This is very possibly due to the current loss budget (71% MSE) which prioritizes intensity over hue, resulting in neutral-colored light sources. Incorrect color lights are also penalized more heavily by neutral white ones.

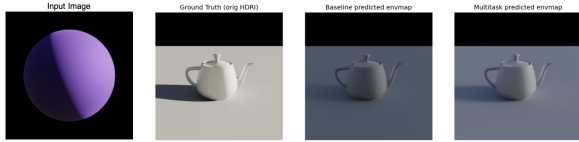


Figure 12: Failure case: highly diffuse material input

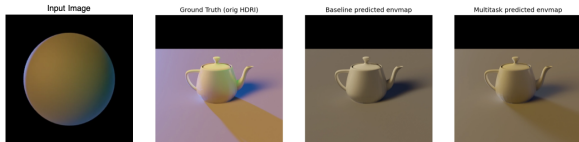


Figure 13: Failure case: saturated multi-color lighting

6.2 Applications

Beyond the scope of this project, the proposed multi-task architecture could have implications for real-time Augmented Reality. Since it is very lightweight it may be suitable for on-device execution where material-aware lighting can improve the visual integration of virtual objects. The ability to disambiguate light and material from a single image also could be applied to computational photography, especially for material editing and scene relighting where high-frequency environment maps are relied upon for realistic specular transport.

7 Conclusion

This project has demonstrated that multi-task light prediction, conditioned on material awareness, is a viable and lightweight solution for inverse rendering. By treating material properties as a physical prior, realistic light probes that excel in downstream rendering tasks may be achieved.

Future work may look like generalizing to real scenes. This model may be applied to real-world data using multi-object segmentation to handle more complex geometries.

Another direction would be to integrate spherical mover's loss to improve spatial misalignment issues that were observed when carrying out hdri-map-wise evaluation metrics.

To address the lack of hue consistency, a pipeline that splits the lighting estimation into separate luminance and chromaticity components may be implemented. This would predict a single-channel intensity map as well as a normalized color-map, thus preventing intensity gradients from overpowering color information during backpropagation.

Inspired by recent industry trends, another line to investigate is to split the pipeline into separate specular and diffuse estimation components to better refine how the model reacts to surface reflections.

References

- [1] J. D. Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, 2012, Chapter 10.
- [2] Poly Haven. "High Resolution HDRIs for Environment Lighting." [Online]. Available: <https://polyhaven.com/hdri>. [Accessed: May 2026].
- [3] J. Philip, S. Morgenthaler, M. Gharbi, and G. Drettakis. "Neural Light Field Estimation for Street Scenes with Differentiable Virtual Object Insertion." *European Conference on Computer Vision (ECCV)*, 2022.
- [4] Z. Zhan, Y. Pan, Z. Ye, et al. "EMLight: Lighting Estimation via Spherical Distribution Approximation." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [5] C. Li, K. Xu, and S. Bi. "Neural LightRig: Unlocking Accurate Object Normal and Material Estimation with Multi-Light Diffusion." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [6] G. Nam, J. H. Lee, D. G. Aliaga, and M. H. Kim. "On-site single-image SVBRDF reconstruction with active planar lighting." *ACM Transactions on Graphics (SIGGRAPH)*, 2018.
- [7] P. Pandey et al. "Total Relighting: Learning to Relight Portraits for Background Replacement." *ACM Transactions on Graphics (SIGGRAPH)*, 2021.