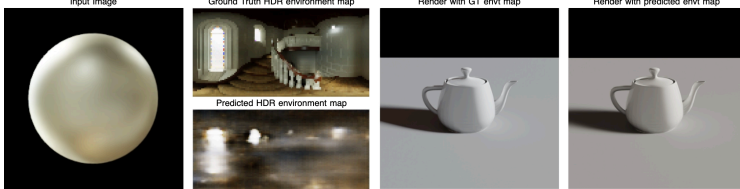


Material-Guided Decoders: Joint Estimation of Material Properties and HDRI Lighting



Motivation

- Lighting prediction is a key component of achieving realism in inverse rendering, a popular topic in CG/CV.
- Use cases include seamless virtual object insertion for AR.
- Lighting estimation is also leveraged to assist material prediction
- Can we use material prediction to inform lighting estimation?
- Hypothesis: surface materials and reflectance properties contain clues about the environment which are best leveraged by material awareness.**
- We will compare material aware vs material unaware approach for decoding lighting environment maps from single images.

Method

Data: synthetic single images of sphere on dark background

- Controlled experiment isolating one object
- Network can focus on connection between material and lighting
- Generalizable to a more complex segmented scene later on

From these images the model predicts the broader lighting environment map (HDRI) with two different approaches:

- 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.



Multitask model has two conditioning variants:

- Tiled (Weak Conditioning):** Material features spatially tiled and concatenated to the backbone features. Physical correlation between the two tasks has to be discovered.
- FiLM (Strong Conditioning):** Feature-wise Linear Modulation (FiLM) forces the decoder to utilize material knowledge.

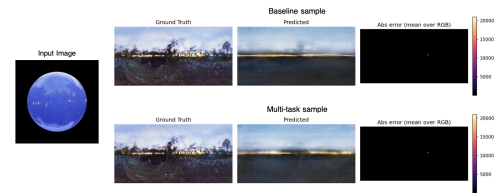


Only MSE cannot achieve strong lighting or shadows.

- Structural anchors: L1 + SSIM are critical
- Per-pixel weighting gives light 5x more weight than dark sky pixel
- Loss: MSE 71% / SSIM 14% / L1 7% / material 7%

Evaluation

- Hierarchical evaluation strategy with following metrics:
 - Mathematical accuracy: environment map Log-MSE
 - Perceptual fidelity: LPIPS, using deep network to see like a human, prioritizing features over raw pixels
 - Per-material and per-bucket MAE
- Downstream complex render task evaluated using LPIPS: only shiny (roughness <0.2 metallic >0.8) subset evaluated, for most material certainty and necessity of correct lighting.

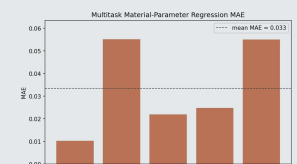


Results

- Baseline outperformed Multi-task model on pixel-based evaluation metrics of envt maps.
- Achieved extremely high confidence in material regression for metallic and specular materials.

Table 3: Global Quantitative Comparison (n = 2250)

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



- Multi-task outperformed significantly on downstream LPIPS eval.
- Specular materials act as anchors to help scene reconstruction

Table 5: Statistical Significance and Task-Based Utility

Test Type	n	Mean Diff.	95% CI	t-stat	p-value
Log-MSE (Global)	2250	-0.0062	$\pm$ 0.0013	-9.54	$< 10^{-10}$
LPIPS (Global)	2250	-0.0065	$\pm$ 0.0018	-7.08	1.44×10^{-12}
Teapot Render LPIPS	120	+0.0211	$\pm$ 0.0090	+4.59	4.48×10^{-6}

Positive Mean Diff. indicates Multitask model superiority.

Compared against industry standard:

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

Conclusion

- Pixel-wise metrics cannot capture utility of environment maps for object insertion: even when renders clearly look worse visually with poor highlights they perform well in pixel-wise log MSE
- Project demonstrates 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.

