

## A Real-Time Multi-Scale Neural Representation for Complex Surface Reflectance – Supplementary Material

### A.1. Ablations

To study the effects of the architectural choices and hyperparameters, we ablate our architecture by removing individual components (gating, warping, shading frames) one at a time while correspondingly increasing the size of the final MLP to account for the FLOPs lost. Figure A.5 visualizes the difference of  $\mathbb{F}$ LIP at each scale averaged over 25 materials. We observe that disabling the gating mechanism causes the largest deterioration in  $\mathbb{F}$ LIP with middle to far view distances, and has the largest overall effect of all the components. Disabling the shading frames also has a strong negative effect. Its removal incurs the most significant decreases in  $\mathbb{F}$ LIP at close view distances, where surface normal variation has the strongest effect. We notice that the model indeed learns something close to the ground truth normal map with the shading frame module at close view distances (see Figure A.6). The effect of the warping module is the smallest, albeit slightly favorable to simply increasing the capacity of the MLP with our standard configuration.

Finally, to compare our model with the tripleplane model in a setting where the number of parameters in  $\mathcal{H}, \mathcal{D}$  are more comparable between models (recall, our model interpolates *matrices*  $\mathbf{H}, \mathbf{D}$ ) we train tripleplane models with increased  $\mathcal{H}, \mathcal{D}$  capacity (“Tripleplane+” and “Tripleplane++”). Please refer to Table A.2 for exact model sizes. Note also that increasing the size of  $\mathcal{H}, \mathcal{D}$  also increases the size of the first layer of the decoder neural network. Despite the increased capacity for the tripleplane baseline, our model yields better metrics.

### A.2. Effect of model capacity

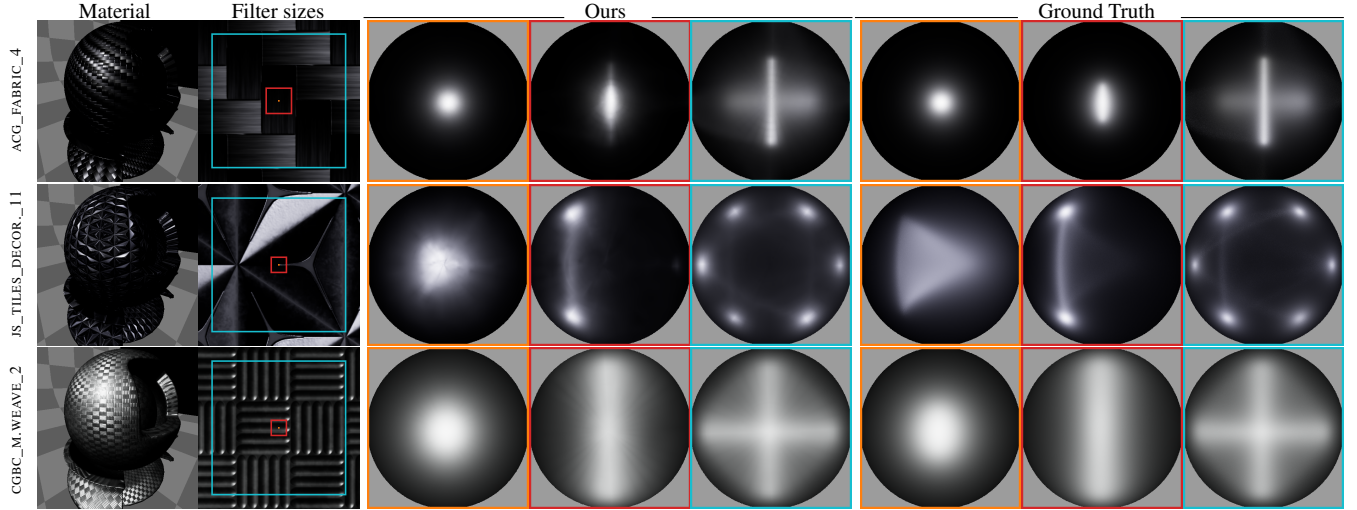
Orthogonally to removing components from our model, we also show the effect of increasing model capacity by changing the size of the angular textures  $\mathcal{H}$  and  $\mathcal{D}$ . We separately change the size along three different axes. First, we increase the angular resolution of  $\mathcal{H}$  and  $\mathcal{D}$  ( $H_h, W_h, H_d, W_d$ ) by a factor of two. Second, we increase  $D_h$  and  $D_d$ , the number of features stored in each  $\mathbf{H}$  and  $\mathbf{D}$ . Finally, we increase the number of channels in the spatial latent  $D_s$ , which also incurs a change in the sizes of the  $\mathbf{H}$  and  $\mathbf{D}$  matrices. The effect of each of the changes is visualized in Figure A.5 (right column). We find that the model is the most sensitive to changes in the number of channels of the spatial latent  $D_s$ . This suggests that the model may have difficulty encoding all spatial variation of the ground truth BRDF in the 8 elements of the spatial latent  $z_s$ . Recall that we do not only use the latent in the gating mechanism but also decode the shading frames and the warping parameters from the same latent. Furthermore, even the ground truth parametric model has a larger channel dimension in the spatial latent than 8.

### A.3. Effect of learnable warps and gating

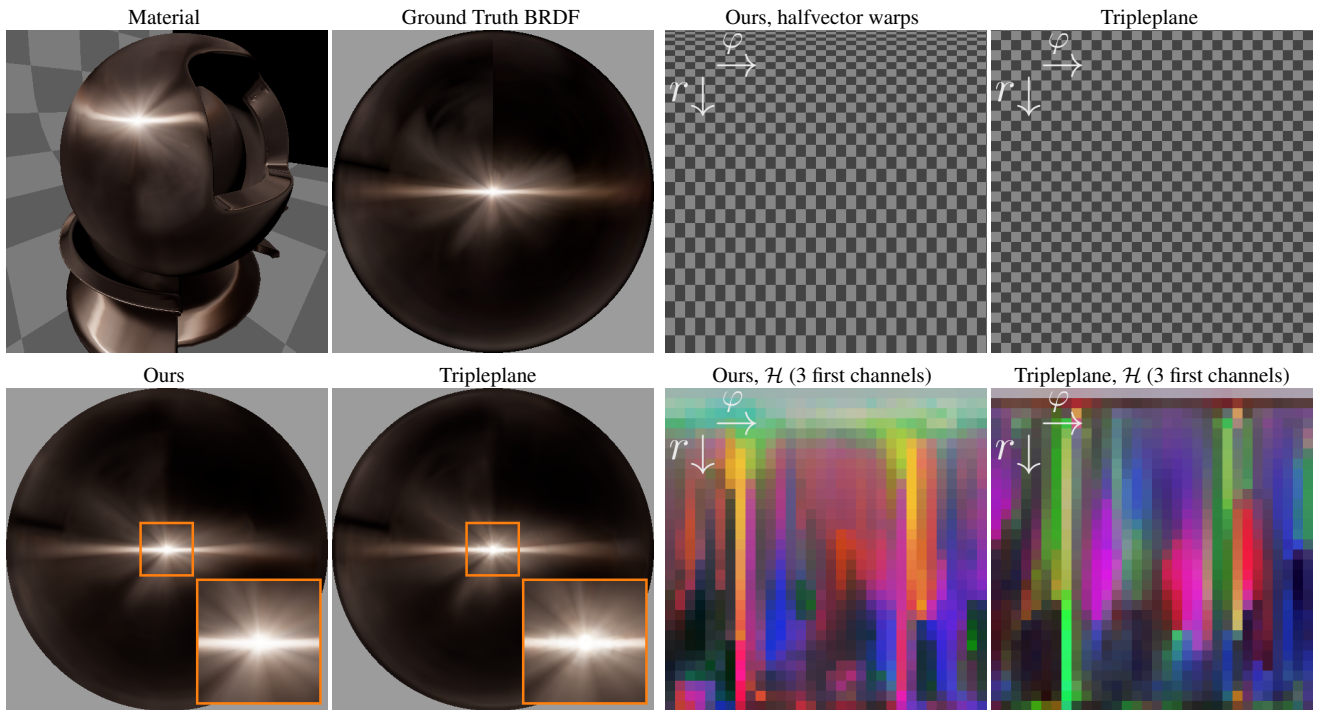
Figure A.2 shows the effect of the angular query warp on the specular ANISO\_COPPER\_SHEET material from the RGL dataset, as well as visualizes some of the feature maps learned in the angular texture  $\mathcal{H}$ . With no warps, all query points close to the important regions of the space of the halfvectors are mapped to only very few texels in

the angular texture. The texture has very little effective resolution to represent the sharp specular lobe. Applying the learned warp allows queries near the important regions of the space to be mapped to a larger area of the angular texture, increasing the effective resolution near the origin. We observe similar results with multiple materials, c.f. models with and without warping in Figure A.8. In our formulation, we crucially rely on the knowledge that specular highlights tend to be concentrated near the origin in the projected halfvector coordinates. Possible due to this assumption, warping is much less useful with SVBRDFs (Figure A.7), where the apparent BRDFs may not have symmetry in halfvector aligned coordinates. We leave the design of more general data-driven warps for future work.

While the effect of simple learnable warps is minor with SVBRDFs, Figure A.7 shows how the gating mechanism has a clear positive contribution for materials that exhibit complex filtered appearance. We compensate for the loss in FLOPs due to removal of the gating/warping components by increasing the size of the decoding neural networks (see Table A.2).

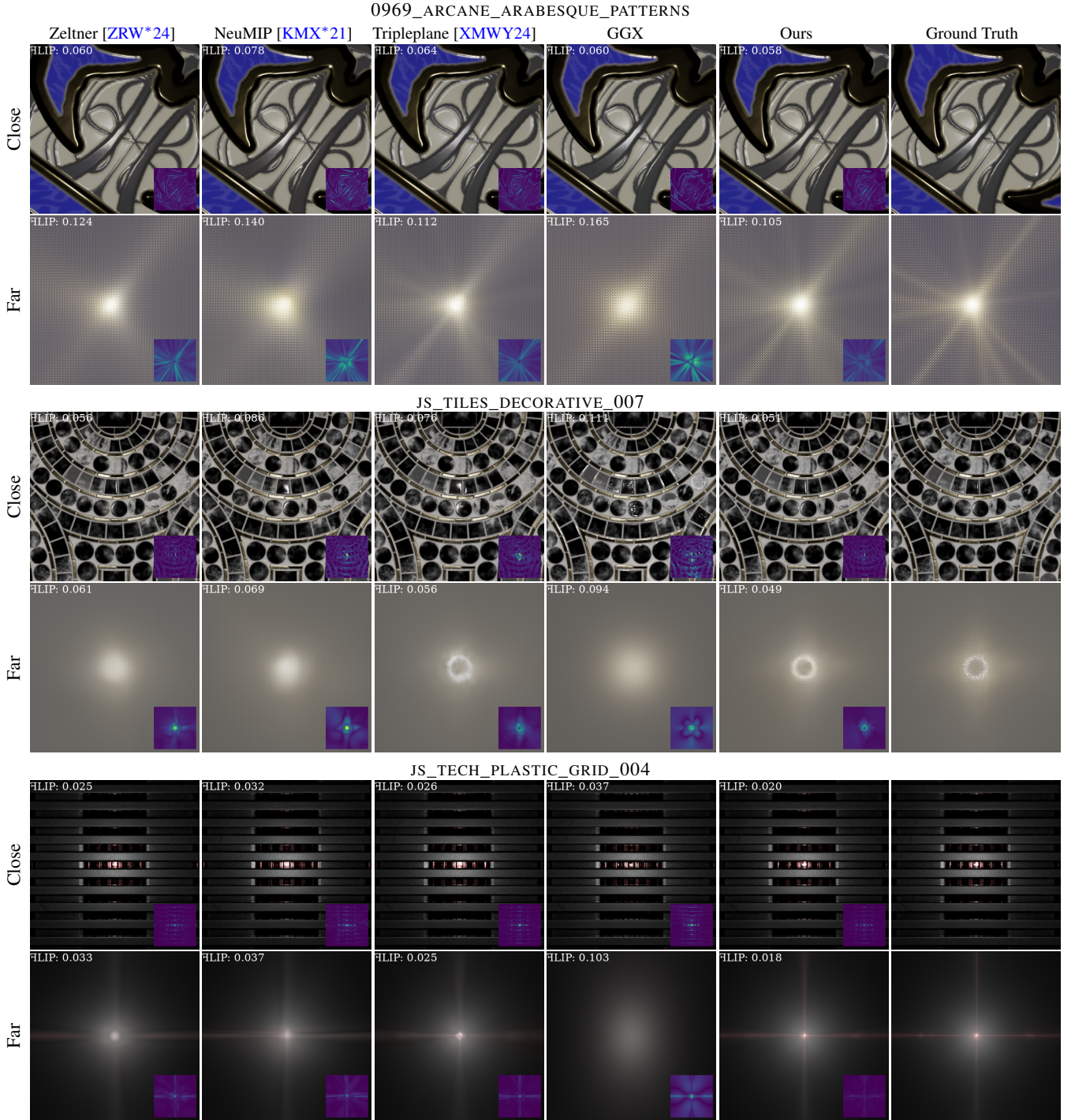


**Figure A.1:** Effect of spatial filtering over a texture. Different sizes of filters (leftmost column) result in potentially very complex apparent BRDFs over the angular space. Each corresponding filter size is color coded with the resulting apparent BRDF. Larger pixel filters require stochastic evaluation for the ground truth (here  $2^{16}$  spp). Our method captures this variation across scales with a single model evaluation.

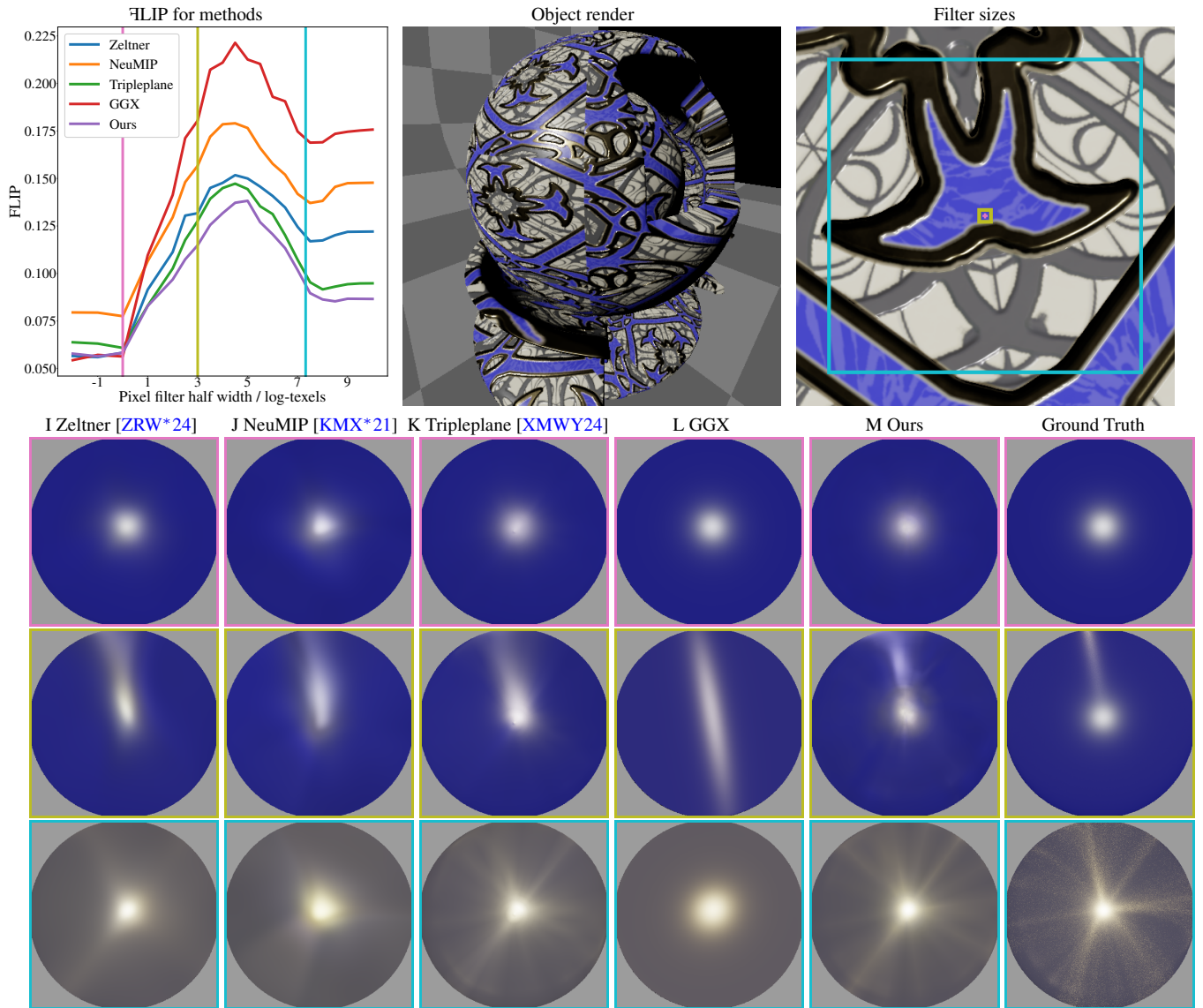


**Figure A.2:** Effect of coordinate warps on MERL materials. ANISO\_COPPER\_SHEET exhibits a sharp specular lobe that is a function of the halfvector. Our model learns to concentrate angular resolution near the specular highlight which is aligned with the center of the disk in the halfvector coordinates. Our learned radial warp (top right, in polar coordinates) warps the coordinate system to change faster closer to the center ( $r = 0$ ). This in turn has the effect that the features corresponding to the highlight near  $r = 0$  cover a larger area of the angular texture (bottom right), allowing representing high-frequency detail with limited angular resolution in the texture.



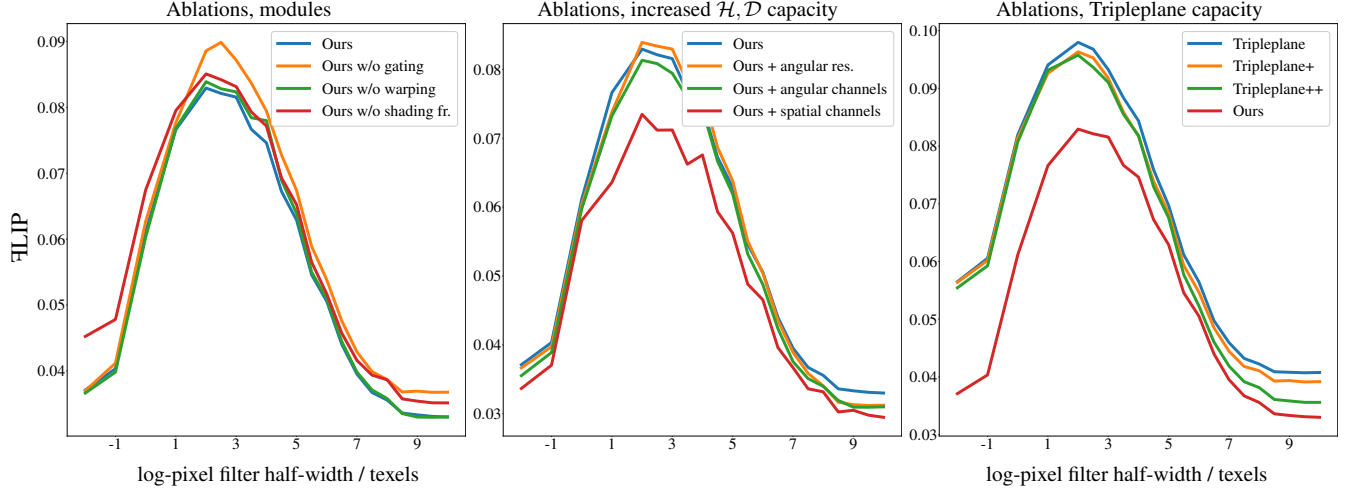


**Figure A.3:** Comparison of methods in two scales using a planar surface illuminated with directional light. The difference image insets measure  $\mathcal{H}LIP$ , and are normalized per row. Fine surface detail induces complex effects in the reflectance of the surface when viewed afar due to integration over the pixel filter. The ground truth uses  $2^{16}$  samples per pixel and all methods have 1 spp.

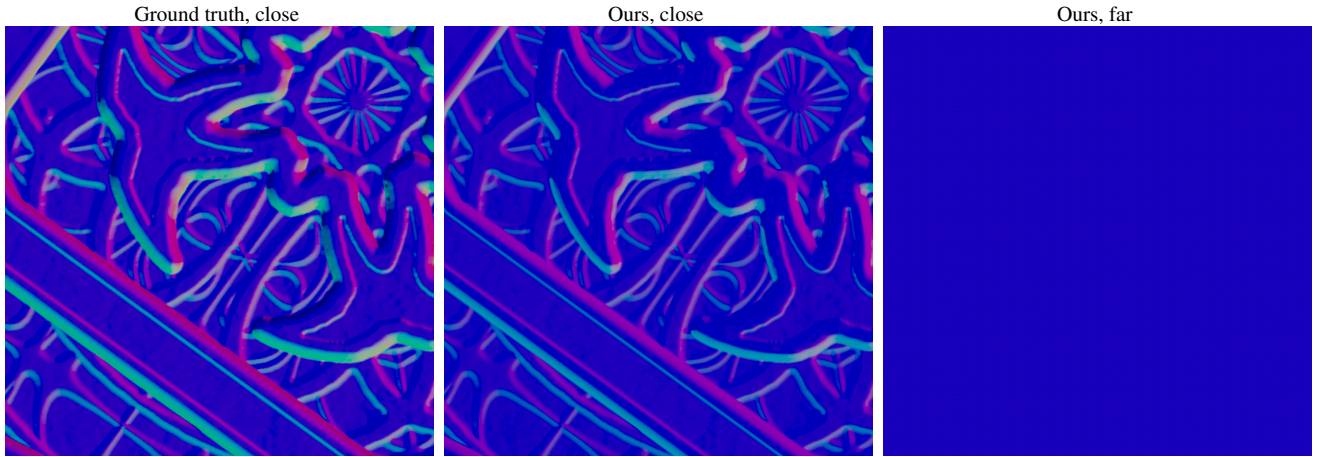


**Figure A.4:** Complexity of filtered BRDFs. Different filter sizes (top right) are highlighted on the plot of filter sizes FLIP (top left). The resulting filtered BRDFs are visualized on the bottom for all methods with the same color coding. Very small pixel filters (pink, first row of bottom group) reduce into the simple underlying parametric model. The apparent BRDF starts gaining complexity already at filter scales of a few pixels (olive, second row) which can also be seen as FLIP becoming worse for all methods (top left). Very large filters produce apparent BRDFs of high complexity (cyan, bottom row) that is not captured by methods only using MLPs as their parametrization (Zeltner, NeuMIP).





**Figure A.5:** F1IP averaged over all materials and test configurations plotted over scales for ablations. Left column: disabling individual modules from our model (gating, warping, shading frames). The gating mechanism has the largest effect over medium to large viewing distances, where there is the most complexity in the angular dimensions of the BRDF. Conversely, disabling the shading frames module has the most significant effect at close distances where most. Middle column: increasing size of  $\mathcal{H}, \mathcal{D}$  along different dimensions. Our model is mostly bottlenecked by the channel count  $D_s$  in the spatial texture  $\mathcal{S}$ , suggesting that the chosen 8 channels in the mip-mapped spatial texture might not be enough to encode all variation in the filtered BRDF. The angular texture resolution and the number of channels ( $H_h, W_h, H_d, W_d$  and  $D_h, D_d$ ) have a less pronounced effect, improving the F1IP metric slightly with large pixel filter sizes. Right column: comparison with the tripleplane model with increased  $\mathcal{H}, \mathcal{D}$  capacity to closer match (+) or exceed (++) the number of parameters in our model.



**Figure A.6:** Normal directions learned by the shading frame module for 0969\_ARCANE\_ARABESQUE\_PATTERNS with our model. Colors encode normal direction with the blue channel denoting up. Even though there is no explicit loss for learning the “correct” normals, the optimization process seems to recover directions that are close to the ground truth. When viewing the surface from far, the effective normal learned by the model reduces into the local up vector.

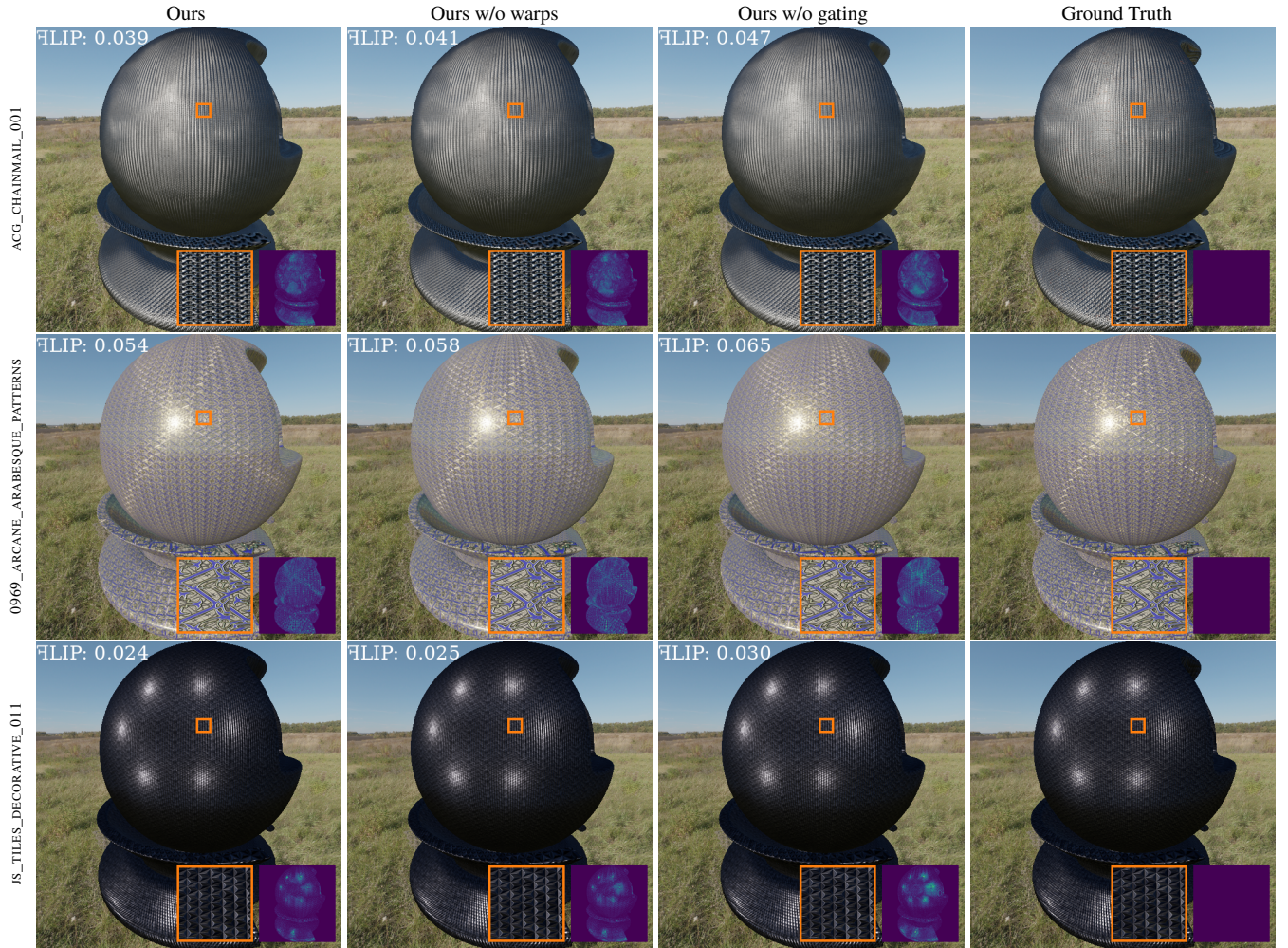
**Table A.1:** Model size and capacity used for spatially homogeneous MERL and RGL materials, where lack of spatial variation reduce the elements of  $\mathcal{H}$  and  $\mathcal{D}$  into column vectors.  $c/L$ -hidden denote the number of channels of hidden layers and the number of hidden layers, respectively.

Model configuration	FLOPs	# MLP weights	MLP $c/L$ -hidden	$\mathcal{H}, \mathcal{D}$ -resolution	$D_h$	$D_d$	$D_s$
Naive MLP	831	675	16/2	n/a	n/a	n/a	n/a
NBRDF [SRRW21]	791	563	14/2	n/a	n/a	n/a	n/a
Tripleplane [XMWY24]	888	437	14/1	$H_h = W_h = H_d = W_d = 32$	6	6	n/a
Ours	842	359	12/1	$H_h = W_h = H_d = W_d = 32$	6	6	n/a

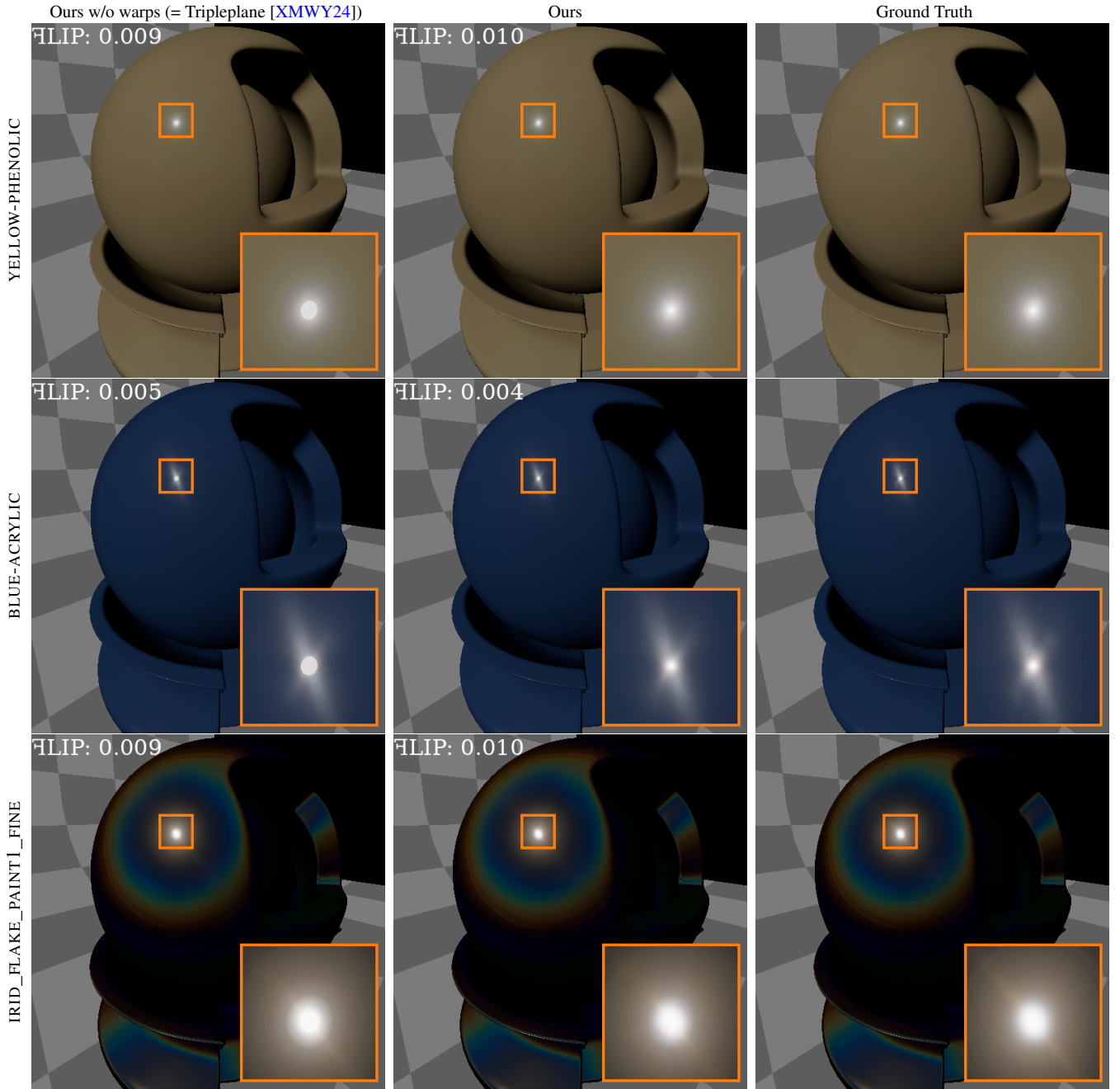
**Table A.2:** Model size and capacity used for MatSynth SVBRDF materials.  $c/L$ -hidden denote the number of channels of hidden layers and the number of hidden layers, respectively. The finest mip level of the spatial texture of each model is of resolution  $1024^2$ . For the tripleplane method  $\mathbf{H}, \mathbf{D}$  are vectors of size  $D_h$  and  $D_d$ , respectively and  $D_s$  only affects the size of the spatial latent tensor. The total size of the angular textures of our model is 81920 (tripleplane 16384 / tripleplane+ 32768 / tripleplane++ 131072) floats which in terms of storage is insignificant compared to the size of the spatial texture (already  $1024^2 \times D_s$  floats, excluding coarser levels of the mip map).

Model configuration	FLOPs	# MLP weights	MLP $c/L$ -hidden	$\mathcal{H}, \mathcal{D}$ -resolution	$D_h$	$D_d$	$D_s$
Zeltner [ZRW*24]	1645	1215	18/2	n/a	n/a	n/a	8
NeuMIP [KMX*21]	1633	1264	18/2	n/a	n/a	n/a	8
Tripleplane [XMWY24]	1638	995	16/2	$H_h = W_h = H_d = W_d = 32$	8	8	8
Tripleplane+	1990	1251	16/2	$H_h = W_h = H_d = W_d = 32$	16	16	8
Tripleplane++	1990	1251	16/2	$H_h = W_h = H_d = W_d = 64$	16	16	8
GGX	$\sim 100$	n/a	n/a	n/a	n/a	n/a	12
Ours	1625	447	12/1	$H_h = W_h = H_d = W_d = 32$	8	2	8
Ours w/o warping (ablation)	1595	457	14/1	$H_h = W_h = H_d = W_d = 32$	8	2	8
Ours w/o gating (ablation)	1759	1019	16/2	$H_h = W_h = H_d = W_d = 32$	8	2	8

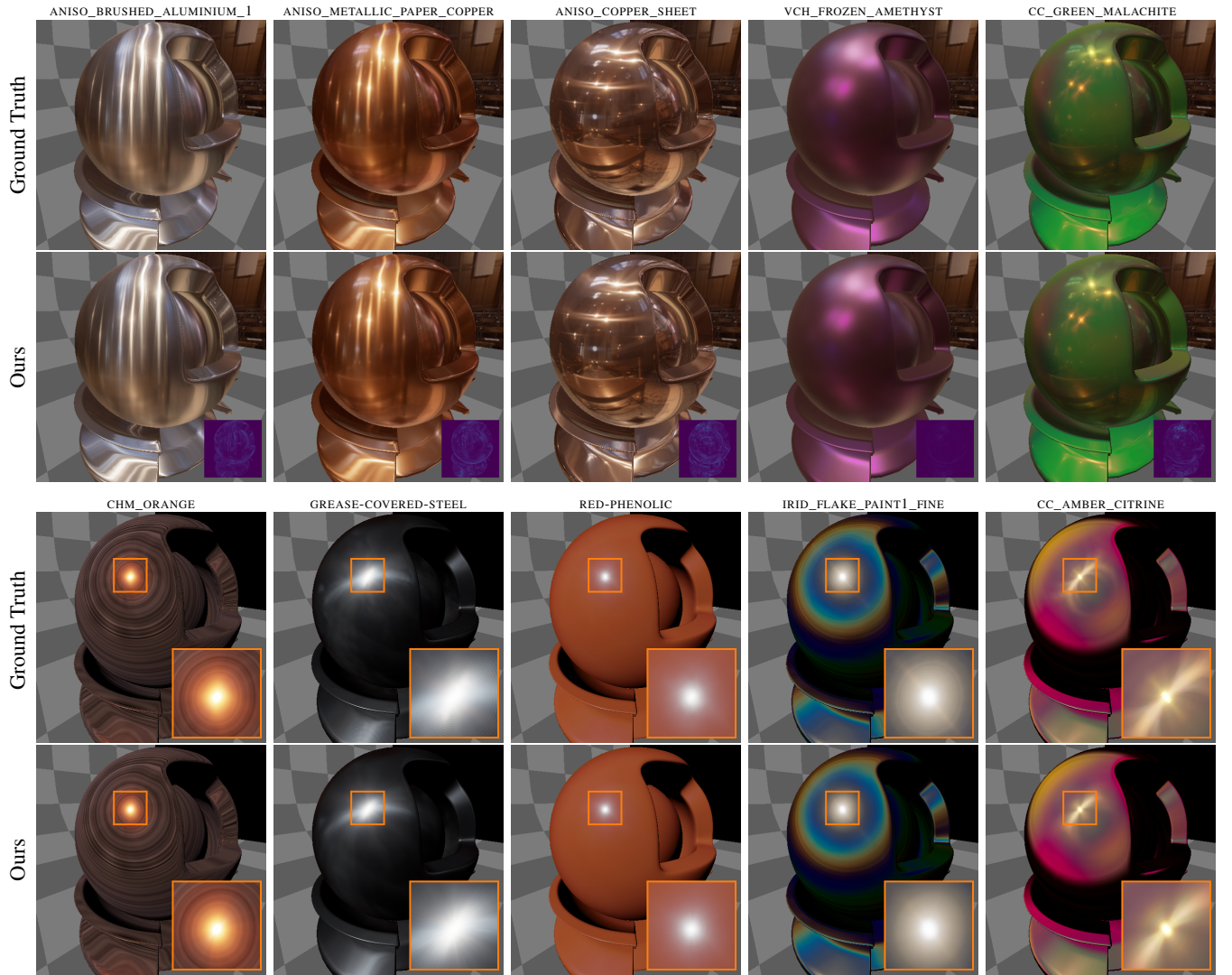




**Figure A.7:** Gating and warping on SVBRDFs. The removal of warping module yields little visual difference with SVBRDFs while conversely deactivation of the gating module produces both visual and numeric differences for materials with complex filtering effects. We roughly standardize the computational budget by increasing the decoder MLP size if a module is deactivated.



**Figure A.8:** Ablation for the warping component for spatially homogeneous materials. With no spatial variation our model without the warping component reduces into the Tripleplane model. Although there is little difference in  $\nabla\text{LIP}$  between the models our model yields more faithful fits to highly specular lobes.



**Figure A.9:** Top group: RGL materials under the Grace cathedral environmental light [Deb98]. The error maps visualize  $\Delta$ LIP w.r.t. the ground truth. We use the importance sampler of the underlying ground truth RGL material. Bottom group: specular MERL/RGL materials under directional light.





**Figure A.10:** Flexibility of our model trained demonstrated on MatSynth materials. We illuminate the object with the Grace cathedral [Deb98] environmental light with and visualize the object in two different scales (top and bottom groups). We use the importance sampler of the underlying ground truth GGX model. The errormaps visualize FLIP with respect to the ground truth.