Supplemental Material:
Toward a Universal Model for Shape from Texture

1. Ablation Study

We test the effect of each individual component of our system by omitting each one and measuring the performance of the system without it. The top four rows in Table S1 include the same results as in Table 1 from the original submission, containing a comparison of our results with previous methods of Clerc and Mallat [3], Loh and Hartley [16], and Lobay and Forsyth [15]. The remaining rows are the results of our system omitting one element at a time.

Clerc and Mallat [3]	30.5	41.2	45.4	35.6	32.7	37.8	36.1	38.7	29.5	29.0	24.4
Loh and Hartley [16]	-	-	12.9	35.9	21.5	23.7	-	19.8	22.6	21.6	**6.9**
Lobay and Forsyth [15]	27.7	29.8	38.2	40.4	41.7	23.5	33.1	17.6	32.0	47.6	9.2
Ours	**15.2**	**16.8**	**12.6**	**15.0**	18.6	17.4	19.5	**17.1**	14.0	**20.1**	14.9
Ours, no global dimensions	42.0	26.9	20.9	27.0	**16.7**	26.3	25.4	27.4	**12.2**	23.4	15.1
Ours, no local dimensions	26.2	27.4	20.6	22.5	25.2	27.4	27.3	23.8	17.2	26.1	14.5
Ours, no periodic dimensions	26.3	27.4	26.9	27.4	27.3	27.4	27.4	27.4	24.9	27.4	27.4
Ours, no pyramid	27.4	27.4	27.4	27.4	27.4	27.4	27.4	27.4	27.4	27.4	27.4
Ours, $\alpha^{(I)} = 0$	83.8	37.7	33.1	50.2	84.8	45.5	77.2	27.3	32.1	35.3	35.2
Ours, $\alpha_n^{(S)} = 0$	16.3	17.0	13.4	15.5	18.8	**15.5**	**19.2**	17.4	15.0	21.4	23.6
Ours, $\alpha_t^{(S)} = 0$	27.3	26.7	20.0	22.5	22.4	26.5	27.4	20.8	14.8	25.4	14.8

Table S1: Shape accuracy (MAE, in degrees) of our and other algorithms for various textures, including those of Figs. 5 & 6 in the main paper and four additional ones. Each entry is average error over four images with the same texture and four different shapes (see Fig. S1 for full results obtained with the parameters used for the main paper). All images were simulated without lighting effects or other non-idealities, as explained in the main paper. Missing entries for [16] are failures to identify a frontal texton.

Our complete system (*i.e.*, using all components) performs better than other configurations for most textures. The performance of the system was significantly worse when omitting: the local or periodic inputs to the generator; the pyramid representation for the surface normals; the integrability term; or the tangent vector smoothness term. The MAE value of $27.4°$, which was obtained by our system when removing the pyramid (as well as some other cases), is the error value that corresponds to predicting a planar shape $\hat{n}(x, y) \equiv \hat{z}$ over the entire image for all four shapes.

When omitting the global inputs to the generator, the system generally performs well on textures that are near-regular, but fails on more stochastic ones. Additionally, when obtaining shape estimates without using global inputs for the non-ideal images in Figure 8 of the original paper, the obtained MAE is $26.5°$, which corresponds to a flat plane $\hat{n}(x, y) \equiv \hat{z}$ for the spherical shape. The results obtained by omitting the smoothness term for the normal vector field (*i.e.*, setting $\alpha_n^{(S)} = 0$), are similar to the results from our submission, because the integrability loss already enforces smoothness in the normal field. Nevertheless, as shown in Table S1, including the normal smoothness term is not entirely redundant, and it improves the results in most cases.

2. Synthetic Images

We present the results of our system on synthetic images generated without the effects of shading or other non-idealities. The average score of our system for each texture can be seen in Table S1 (or Table 1 in the original submission).
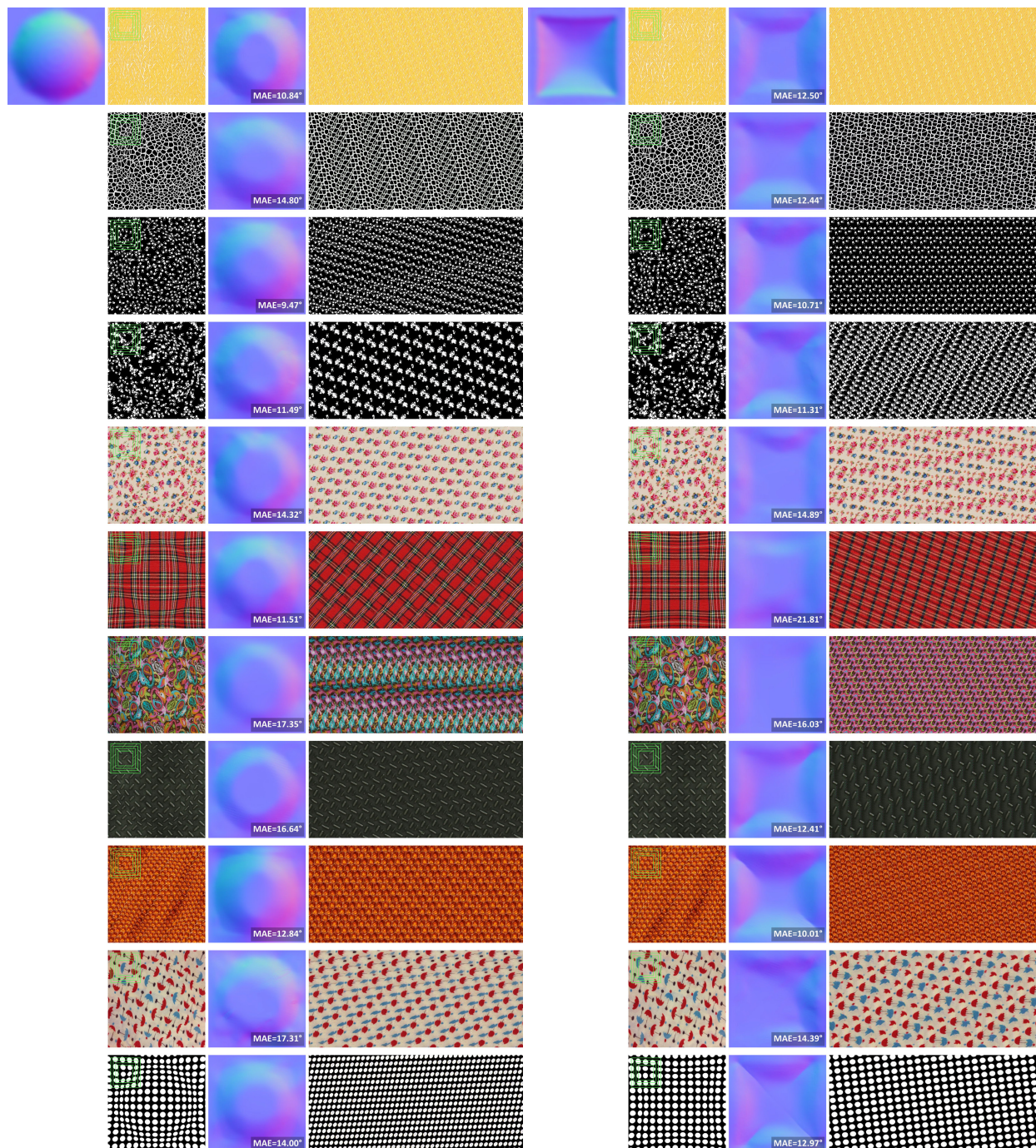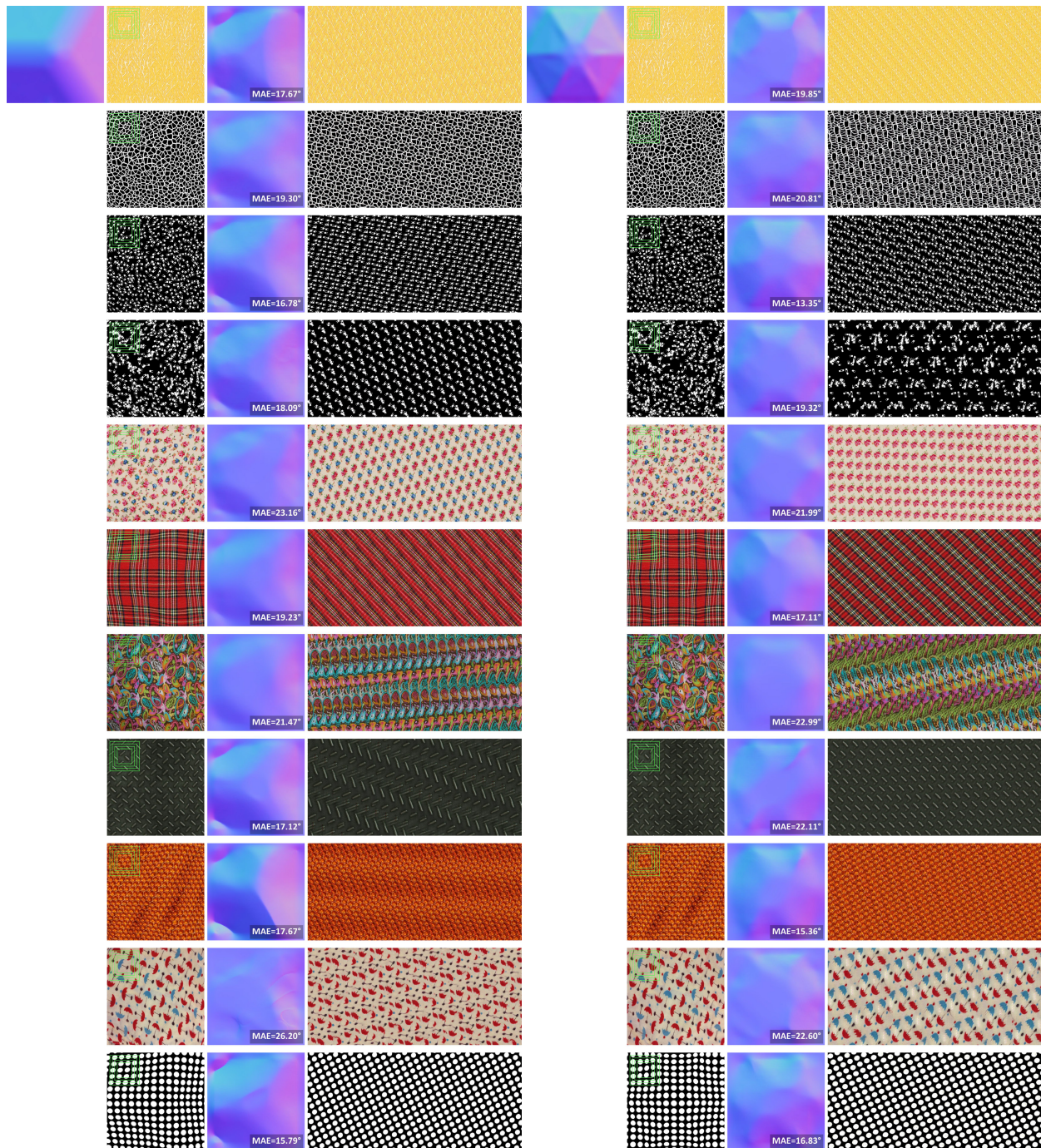

Figure S1

Figure S1: *(cont.)* Complete results of our model on synthetic images with no illumination effects or other non-idealities. The top half of the figure shows the results for the spherical (left) and square (right) shapes, and the bottom shows the results for corner (left) and umbrella (right) shapes. Each quadrant has four columns, containing (from left to right): the ground truth surface normals; the input image (with inset squares showing the patch sizes used by the unwarper); the estimated surface normals with the corresponding MAE; and a sample from the learned texture process.

3. Real Images

Figure S2 shows results of our system on photographs that were omitted from our original submission. The architecture and parameters used to obtain these results are identical to those used throughout the original submission.
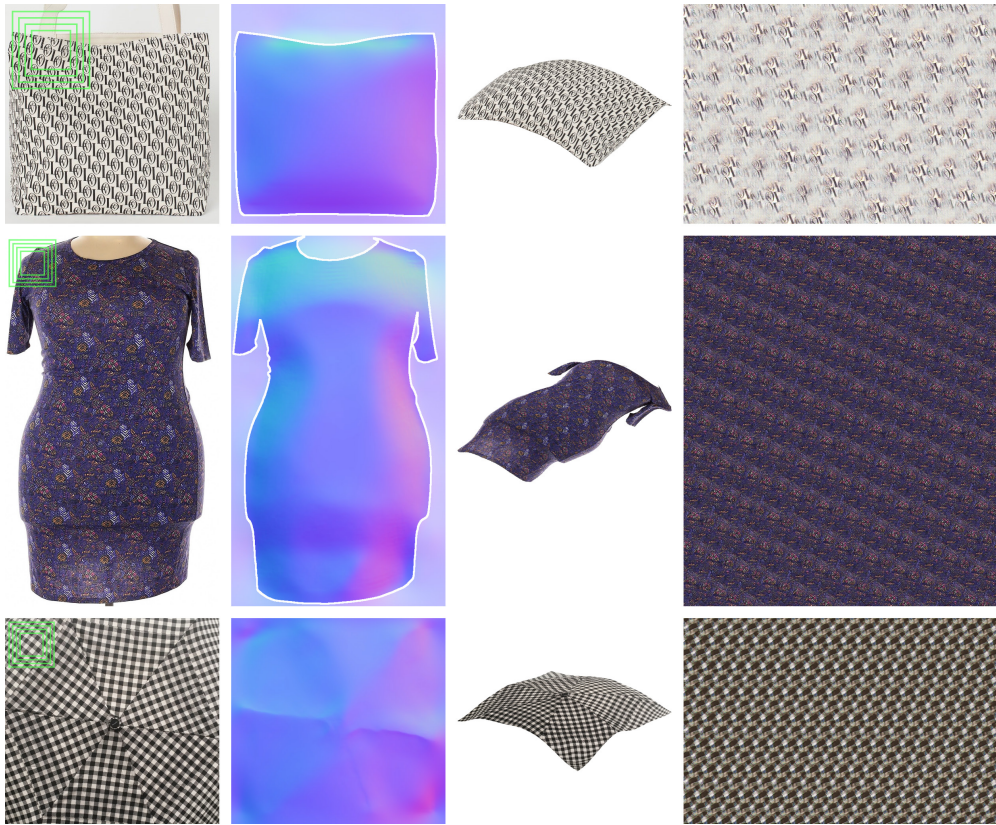

Figure S2: Shape and texture results for photographs (in addition to the results in Figs. 1 and 4 of the main paper). From left to right: input image; output surface normals; output shape; and sample from the learned texture process. Green inset squares show patch sizes used by the unwarper. For visualization purposes, untextured regions in the input images were manually cropped *after* convergence of the game.