An Object is Worth 64x64 Pixels:
Generating 3D Object via Image Diffusion

Supplementary Material

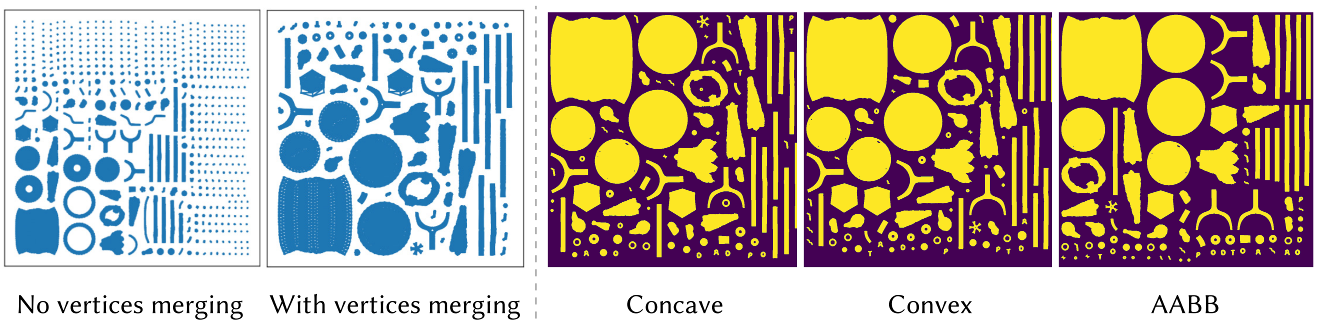

Figure 9. **Left**: Merging coincident vertices before repacking will substantially reduce the number of patches. **Right**: The results of three commonly used uv-islands packing algorithms. For our Object Images, we use AABB with vertices merging.

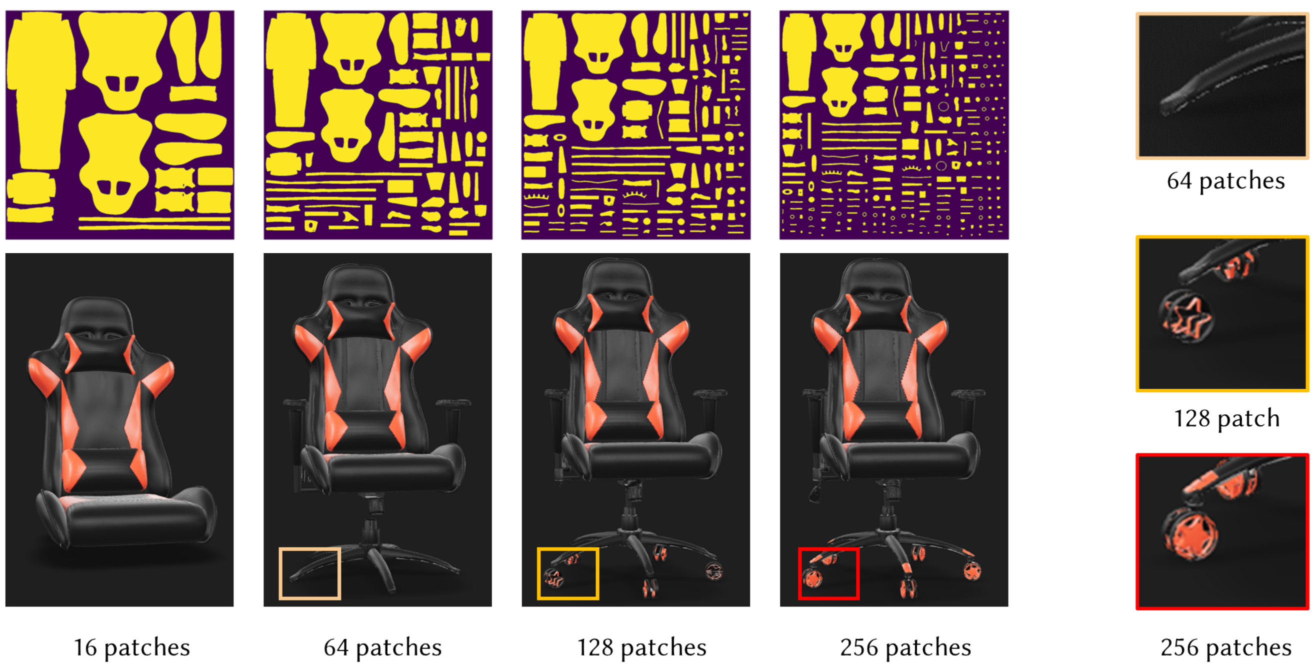

Figure 10. The effect of number of kept patches. As the number of patches goes up, more intricate geometric parts are kept. However, the average number of pixels dedicated to each part is reduced.

A. Repacking the UV-atlas

As mentioned in Sec. 3.3 of the main text, 3D objects with UV-maps generally cannot be directly converted into Object Images (omages) due to issues such as overlapping regions, out-of-boundary UVs, touching boundaries, or extremely large number of patches. Overlapping regions breaks the single-valued assumption, making it impossible to map the overlapped region back to the 3D surface. Since designers often reuse textures for similar parts, overlapping UV islands are common in 3D assets.

Another common issue is the touching boundary problem. One important assumption of omages is that different patches not only do not overlap but can also be separately recognized. We detect different parts by identifying the connected components within the alpha (occupancy) map. If two patches have touching boundaries, this detection will fail, introducing artifacts that connect patches which could be far apart. To address the above issues, we leverage standard UV-atlas repacking to obtain non-overlapping patches and pack them into high-resolution images. For efficient learning, we then downsample the images using sparse pooling to snap the boundaries and eliminate gaps. We describe the repacking and baking step in detail below (the downsampling is described in the main paper).

Repacking and baking. We use UV-atlas repacking to obtain clean patches that are free from artifacts. We first obtain the 2D UV-islands of all patches, then use a 2D irregular shape packing algorithm to rescale and rearrange the UV-atlas within the standard UV-domain, leaving margins between each island. In Fig. 9 (right), we show the the three packing methods provided by Blender: Concave, Convex, and AABB. Their names indicate the shapes approximations used for the packing of the patches and result in different space utilization efficiency. Concave (exact shape) has the least empty space but introduces complex combinatorial patterns that are challenging for generative models to learn. Hence, we adopt AABB as the primary method for repacking.

Another common issue is that many patches are unnecessarily separated into multiple sub-patches by default. This results in numerous small pieces that degrade the quality of the omage, potentially reducing it to a triangle soup or point cloud as the number of patches increases. By merging vertices that share the same 3D and 2D UV coordinates, we can reconnect these sub-patches to form larger patches. This not only reduces empty space but also improves the integrity of the patches. See Fig. 9 (left) for comparison of packing with and without vertex merging.

After merging the sub-patches, there may still be an excessive number of patches. To simplify the generative modeling, we keep a maximum number of patches, K. For shapes with more patches than this threshold, we sort the patches by their 3D area and retain only the largest K patches. Fig. 10 shows the effect of this parameter. Having more patches preserves geometric details but complicates generative modeling. This is especially true for lower-resolution omages, where smaller parts lack enough pixels to form meaningful regions. In practice, we find that a maximum of 64 patches works well for generating 64-resolution omages (See Fig. 7).

After repacking, we rasterize the geometry and material properties into an image format through texture baking according to the repacked UV-atlas. We bake the geometry (including UV occupancy), normal map, albedo, metalness, and roughness into the final $(R, R, 12)$ omage, with $R = 1024$ set as default for high-quality results.