# STR-GQN: Scene Representation and Rendering for Unknown Cameras Based on Spatial Transformation Routing
# – *Supplementary Materials* –

## 1. Occupancy Concept Mapping

The proof of the proposed OCM (cf. Sec. 3.3) and the analysis of routing process in probabilistic perspective are given as follows. Let $o_{k,c}$ denote the random variable indicating a concept exists at the 3D location of the $k$th world cell (i.e. $o_{k,c} = 1$) or not (i.e. $o_{k,c} = 0$), and $o_{ij,c}$ denote the random variable indicating a concept $c$ exists at the 2D location $(i, j)$ of a view cell. The goal of OCM is to estimate the scene cell $sc_{k,c}$, which represents the posterior of the existing probability for a concept $c$ at the location of the $k$th world cell given the observed images $x^{1:N}$:

$$sc_{k,c} = p(o_{k,c} = 1|x^{1:N}). \tag{1}$$

The log-odds of the concept existing probability in world space is defined by

$$\log Odd(o_{k,c}) = \log \frac{p(o_{k,c} = 1)}{p(o_{k,c} = 0)}. \tag{2}$$

The posterior of the log-odds given an observation can be computed via Bayes' theorem:

$$\begin{aligned}\log Odd(o_{k,c}|x) &= \log \frac{p(x|o_{k,c} = 1)p(o_{k,c} = 1)/p(x)}{p(x|o_{k,c} = 0)p(o_{k,c} = 0)/p(x)}\\ &= \log \frac{p(x|o_{k,c} = 1)}{p(x|o_{k,c} = 0)} Odd(o_{k,c})\\ &= \log \frac{p(x|o_{k,c} = 1)}{p(x|o_{k,c} = 0)} + \log Odd(o_{k,c}).\end{aligned} \tag{3}$$

Without any observation, the initial prior of the log-odds for the concept existing probability is zero:

$$\log Odd(o_{k,c}) = \log \frac{p(o_{k,c} = 1)}{p(o_{k,c} = 0)} = \log \frac{0.5}{0.5} = 0. \tag{4}$$

Let $wc_{k,c}$ denote the value of the $c$th channel for the $k$th world cell, which is assumed as the log of likelihood ratio of a concept:

$$wc_{k,c} = \log \frac{p(x|o_{k,c} = 1)}{p(x|o_{k,c} = 0)}. \tag{5}$$

The posterior of log-odds given $N$ observations $x^1, x^2, ..., x^N$ can be estimated via iterative Bayesian updating according to Eq. 3 and 4:

$$\begin{aligned}&\log Odd(o_{k,c}|x^1, x^2, ..., x^N)\\ &= (...((\log Odd(o_{k,c}) + wc^1_{k,c}) + wc^2_{k,c}) + ...) + wc^N_{k,c}\\ &= \sum_{n=1}^{N} wc^n_{k,c},\end{aligned} \tag{6}$$

where $n$ represents the index of different observations. The scene cell is then calculated by the posterior of log-odds:

$$\begin{aligned}&\log Odd(o_{k,c}|x^{1:N}) = \log \frac{p(o_{k,c} = 1|x^{1:N})}{1 - p(o_{k,c} = 1|x^{1:N})},\\ &sc_{k,c} = p(o_{k,c} = 1|x^{1:N})\\ &= \frac{\exp(\log Odd(o_{k,c}|x^{1:N}))}{1 + \exp(\log Odd(o_{k,c}|x^{1:N}))}\\ &= \sigma(\log Odd(o_{k,c}|x^{1:N})) = \sigma(\sum_{n=1}^{N} wc^n_{k,c}).\end{aligned} \tag{7}$$

We further analyze the meanings of the routing process under the mathematical framework of OCM. Similar to $wc_{k,c}$, let $vc_{ij,c}$ denote the value of the $c$th channel for a view cell at the 2d location $(i, j)$, which is assumed as the log of likelihood ratio of concept $c$:

$$vc_{ij,c} = \log \frac{p(x|o_{ij,c} = 1)}{p(x|o_{ij,c} = 0)}. \tag{8}$$

The routing process transforms the message from the view cells to the world cells. The information passing to the $k$th world cell is defined by the weighted sum of each view cell based on the weighting $p_k^{dist}(i,j)$:

$$
\begin{aligned}
wc_{k,c} &= \sum_{i,j} p_k^{dist}(i,j) vc_{ij,c} \\
&= \sum_{i,j} p_k^{dist}(i,j) \log \frac{p(x|o_{ij,c}=1)}{p(x|o_{ij,c}=0)} \\
&= \log \prod_{i,j} \left( \frac{p(x|o_{ij,c}=1)}{p(x|o_{ij,c}=0)} \right)^{p_k^{dist}(i,j)} \\
&= \log \prod_{i,j} \left( \frac{p(o_{ij,c}=1|x)p(x)/p(o_{ij,c}=1)}{p(o_{ij,c}=0|x)p(x)/p(o_{ij,c}=0)} \right)^{p_k^{dist}(i,j)} \qquad (9)
\end{aligned}
$$

Moreover, the Eq. 5 can be re-written as:

$$
wc_{k,c} = \log \frac{p(o_{k,c}=1|x)p(x)/p(o_{k,c}=1)}{p(o_{k,c}=0|x)p(x)/p(o_{k,c}=0)}. \qquad (10)
$$

Combining Eq. 9 and 10, we observe that the probability distribution $p_k^{dist}(i,j)$ can be the weighting of the weighted geometry mean for the concept existing probability $p(o_{ij,c})$ in view space:

$$
\begin{aligned}
p(o_{k,c}) &= \sqrt[\sum_{i,j} p_k^{dist}(i,j)]{\prod_{i,j} p(o_{ij,c})^{p_k^{dist}(i,j)}} \\
&= \prod_{i,j} p(o_{ij,c})^{p_k^{dist}(i,j)}. \qquad (11)
\end{aligned}
$$

## 2. Additional Rendering Results

Fig. 2 to Fig. 6 demonstrate more rendering results of the experiments introduced in Sec. 4.2 for each dataset.

## 3. Visualization of Routing Process

Fig. 7 demonstrates more routing results of the experiments introduced in Sec. 4.4. Furthermore, to evaluate whether the proposed STRN can learn the general mapping between world cells and the continuous view space (as mentioned in Sec. 3.2), we trained the model with 16 × 16 view cells and visualized the routing results of view cells for different resolutions by sampling and interpolating the 2D location codes. As Fig. 8 shows, the proposed STRN can learn the routing weights for different resolution of view cells.

## 4. Results for Complex Scenes

As mentioned in Sec. 4.7, to evaluate whether the proposed STR-GQN can be applied in more complex scenes, we trained the proposed model in a discriminative manner based on the "Vase", "Greek", "Chair", and "Material" datasets. We randomly sampled 32 to 48 frames as the observation images for training and always utilize 48 frames as the observation images for testing. Fig. 9 to 12 demonstrate the experimental results for each dataset.

The proposed STR-GQN achieves good performance on the "chair" dataset and "vase" dataset. The results show that the proposed model can reconstruct the complex texture when the 3D structures are simple. However, the proposed model fails to generate clear results for "material" and "greek" datasets, which reveals the limitation of our model. The texture of "material" dataset is simpler than "vase" dataset while it changes with the view pose. The proposed model fails to reconstruct this kind of texture because the STR mechanism only considers the view-independent concepts in 3D space. On the other hand, the texture of "greek" dataset is simple and view-independent, but its 3D structure is complex. The proposed model fails to reconstruct the detailed structure due to the limited number of world cells.

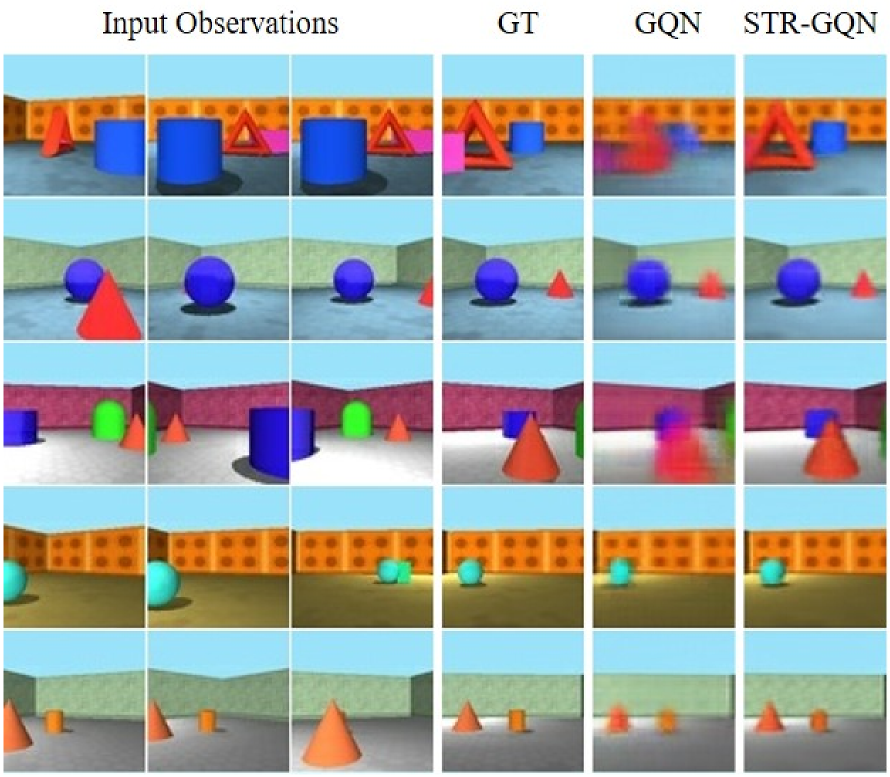

Figure 1. Rendering results of RRC dataset.

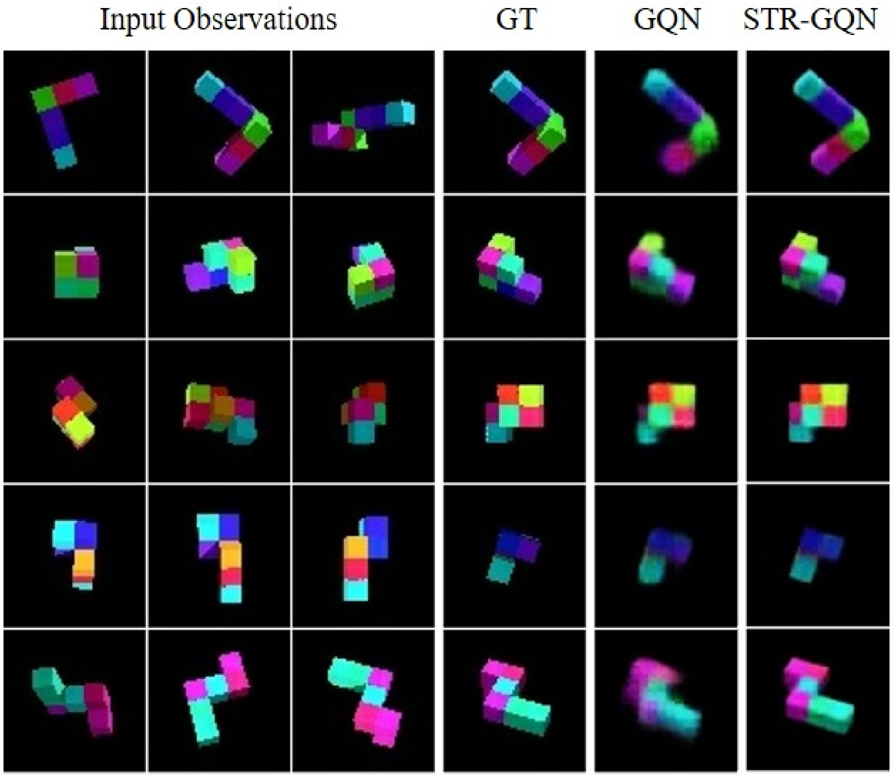

Figure 2. Rendering results of SM7 dataset.

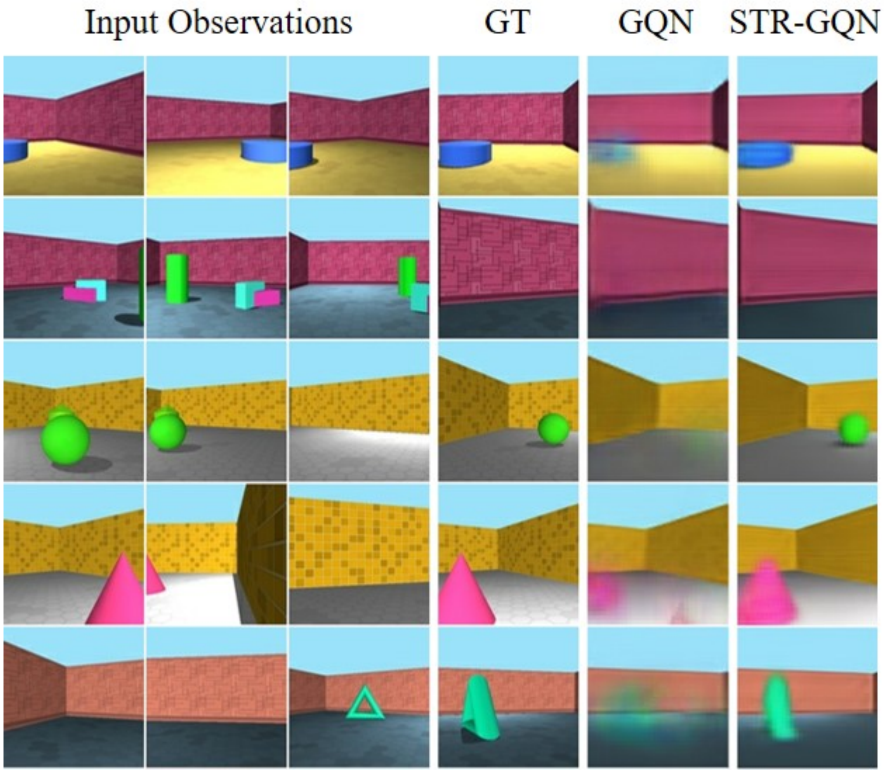

Figure 3. Rendering results of RFC dataset.

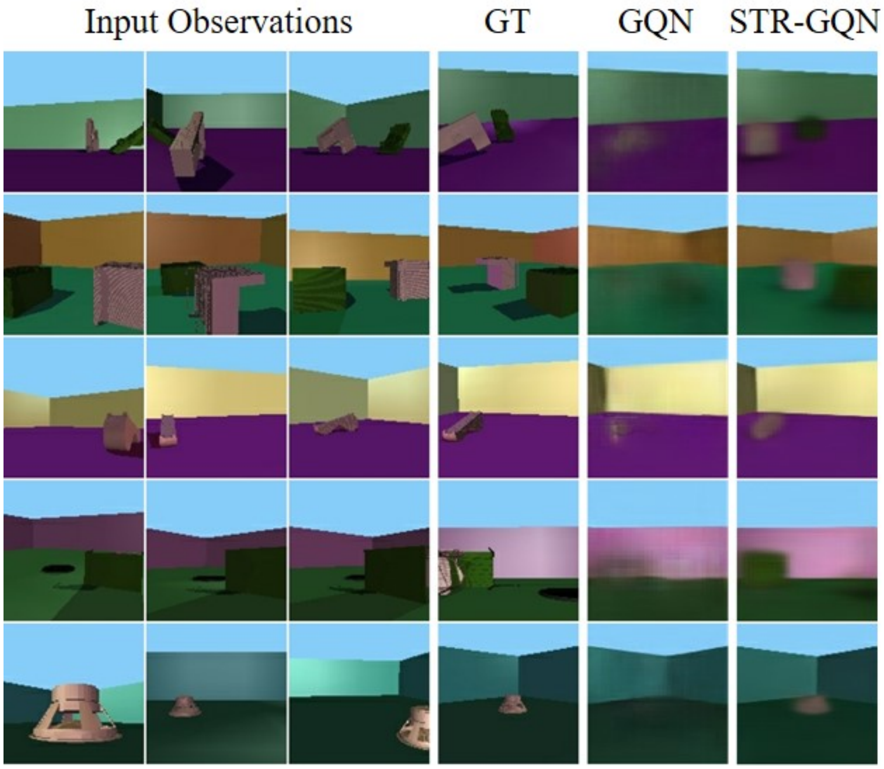

Figure 4. Rendering results of RRO dataset.

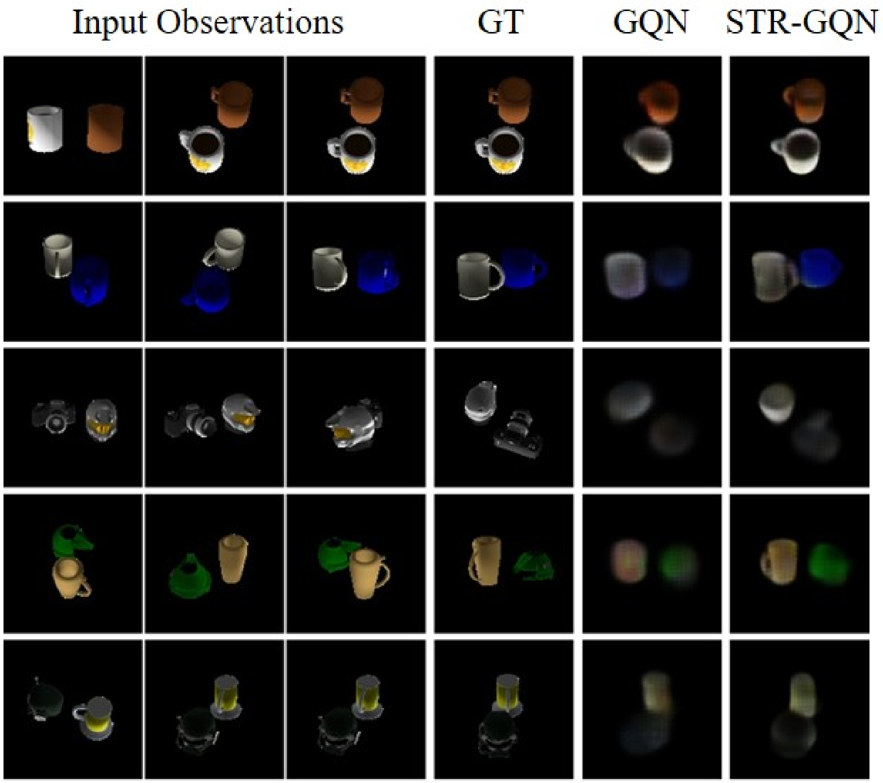

Figure 5. Rendering results of ShapeNet dataset containing 2 objects.

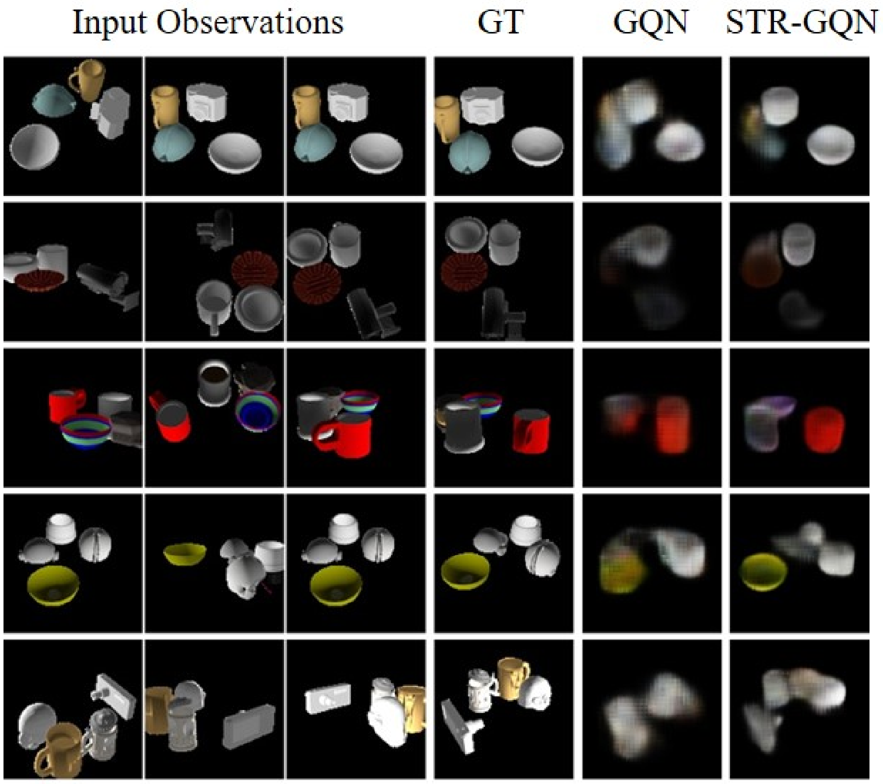

Figure 6. Rendering results of ShapeNet dataset containing 4 objects.

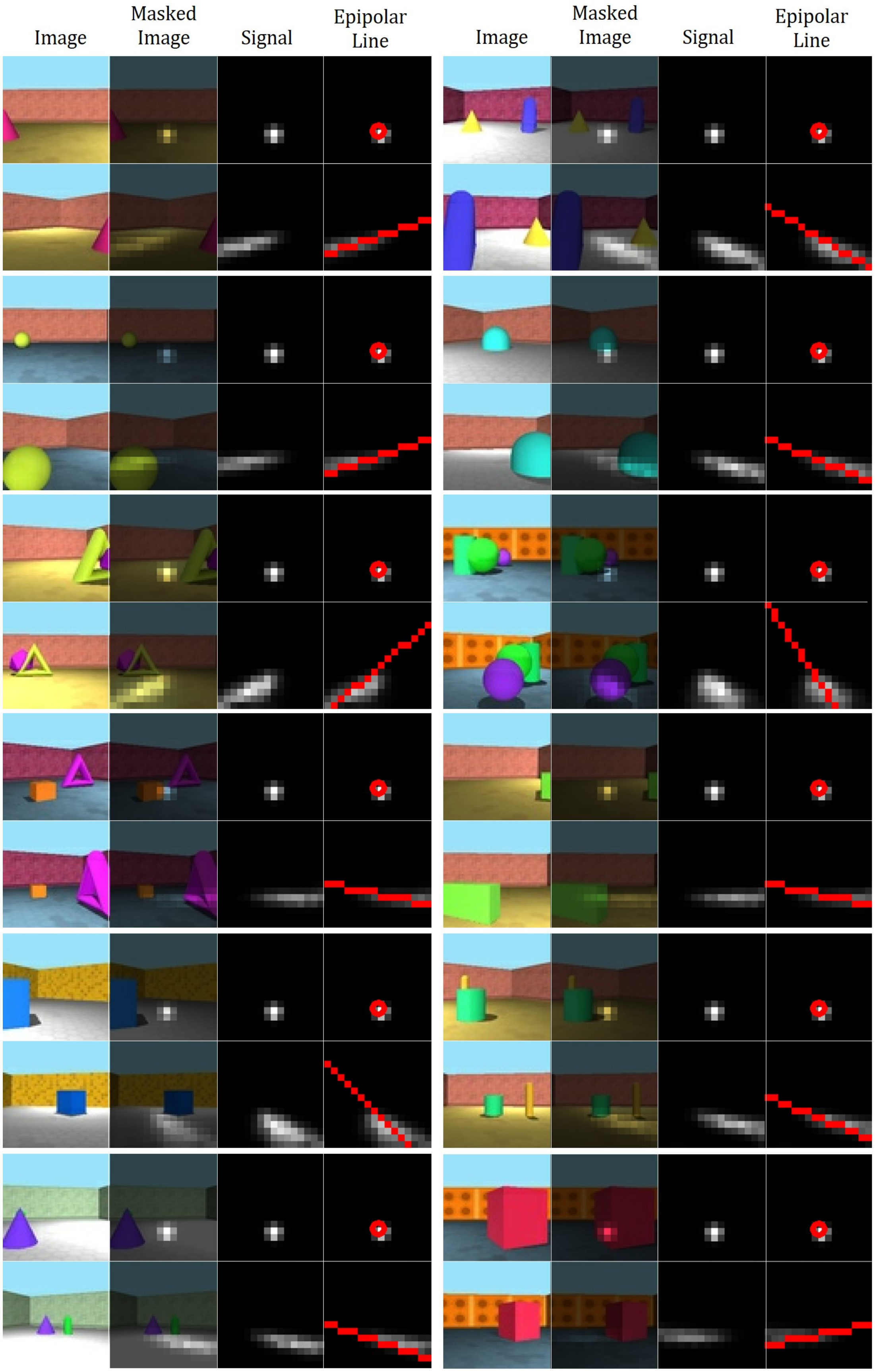

Figure 7. Visualization of the routing process.

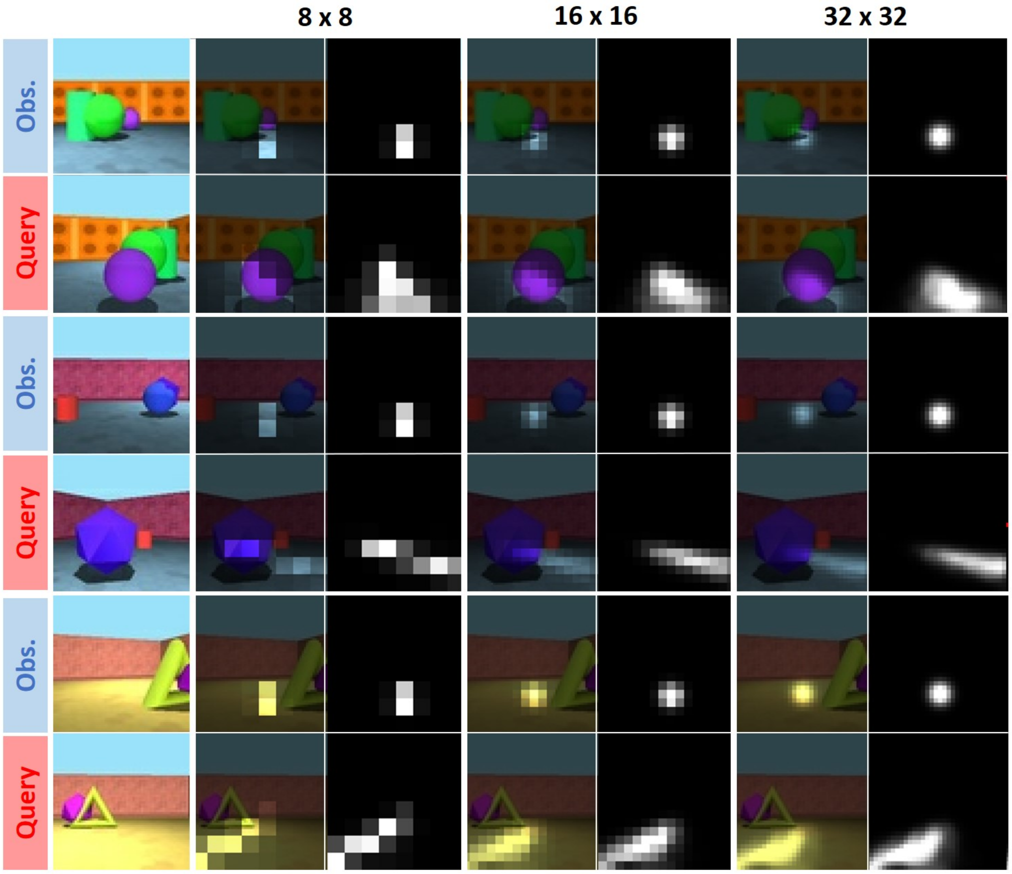

Figure 8. Visualization of the routing process for different resolution of view space.

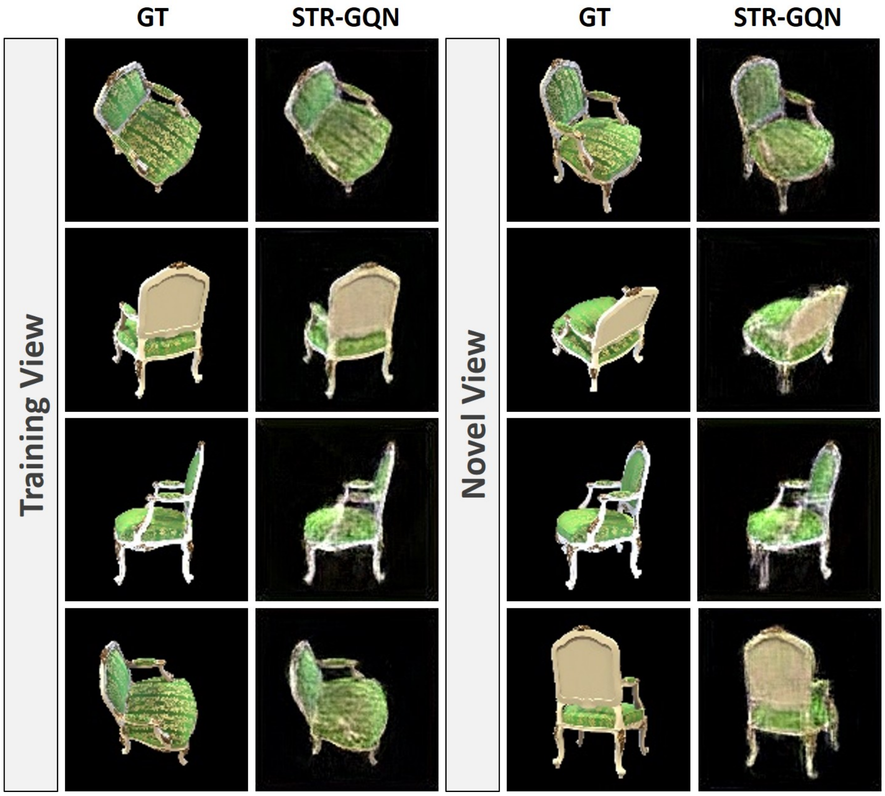

Figure 9. Rendering results of “Chair” dataset.

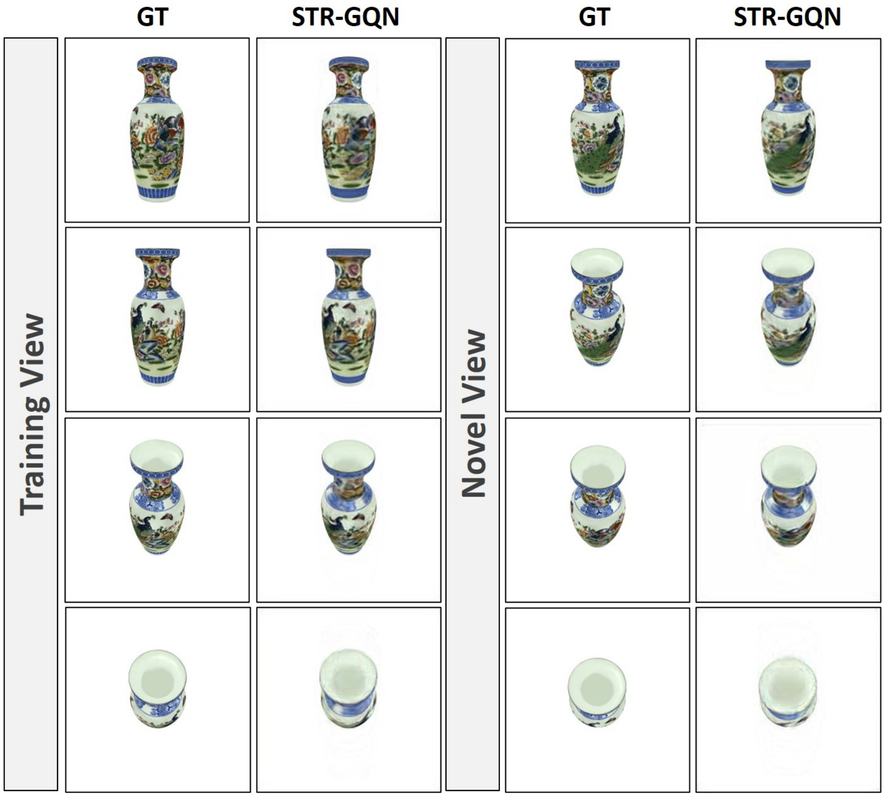

Figure 10. Rendering results of “Vase” dataset.

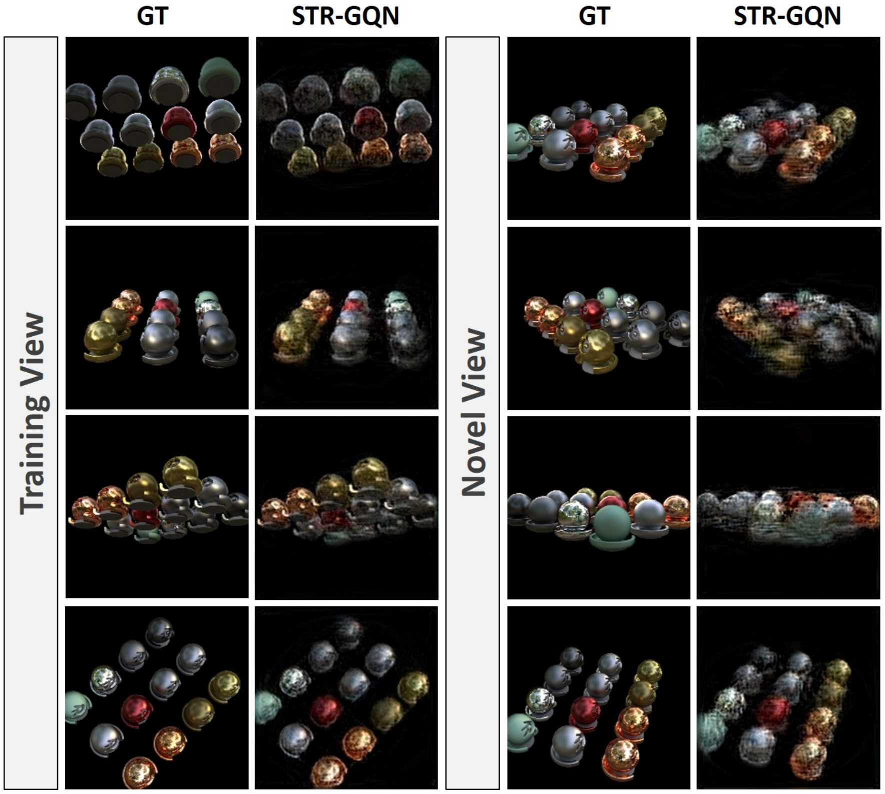

Figure 11. Rendering results of "Material" dataset.

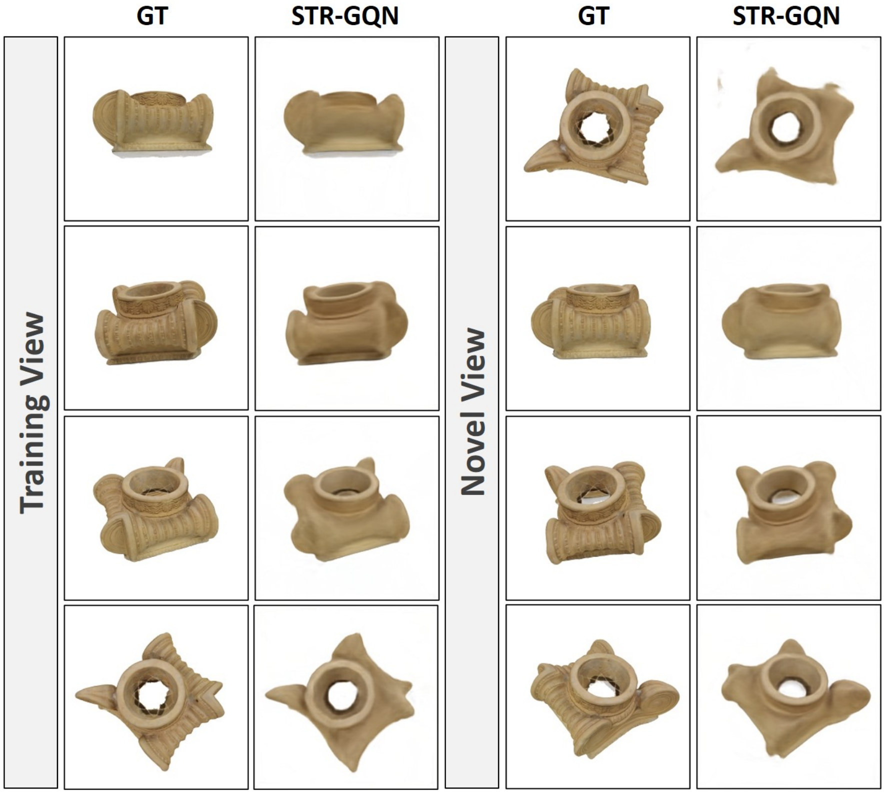

Figure 12. Rendering results of "Greek" dataset.