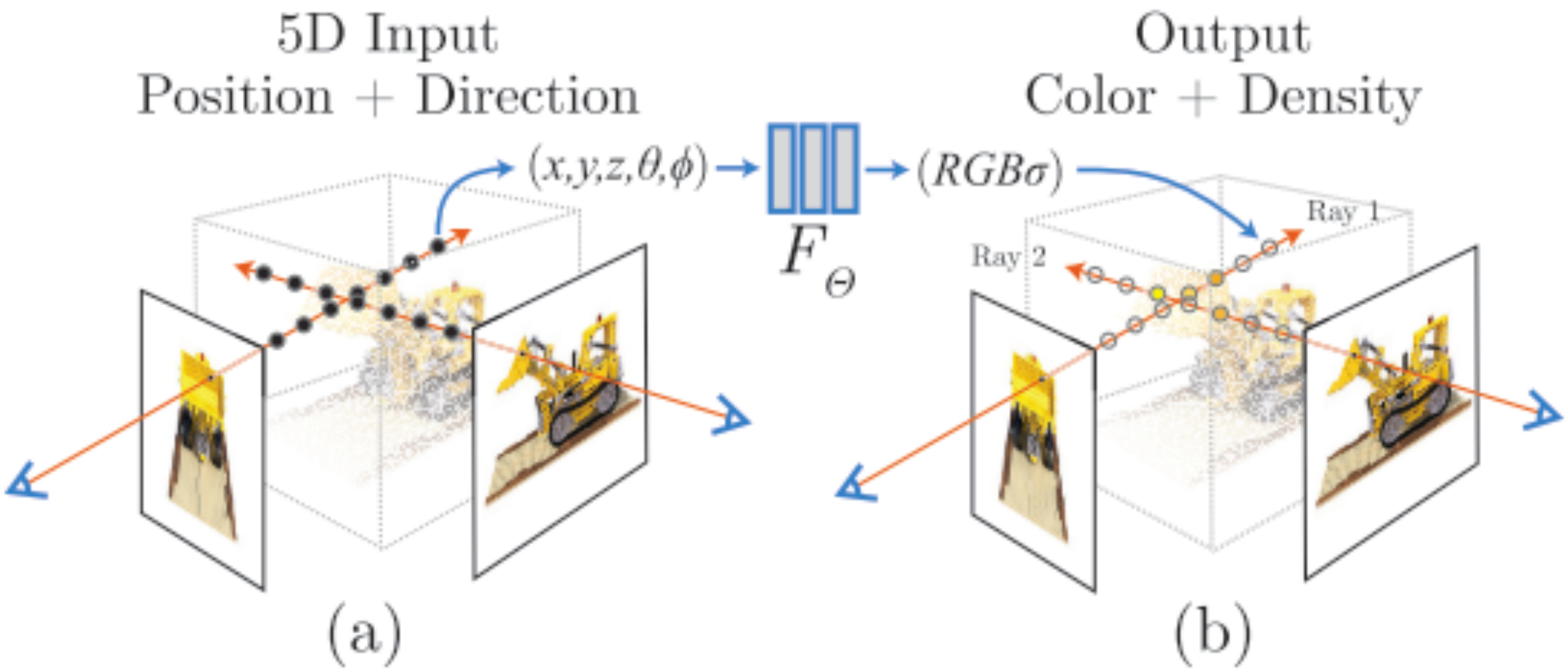
NeRF in the Wild

Neural Radiance Fields for Unconstrained Photo Collections, CVPR 2021 (Oral)

Seungwoo Yoo, KAIST SoC

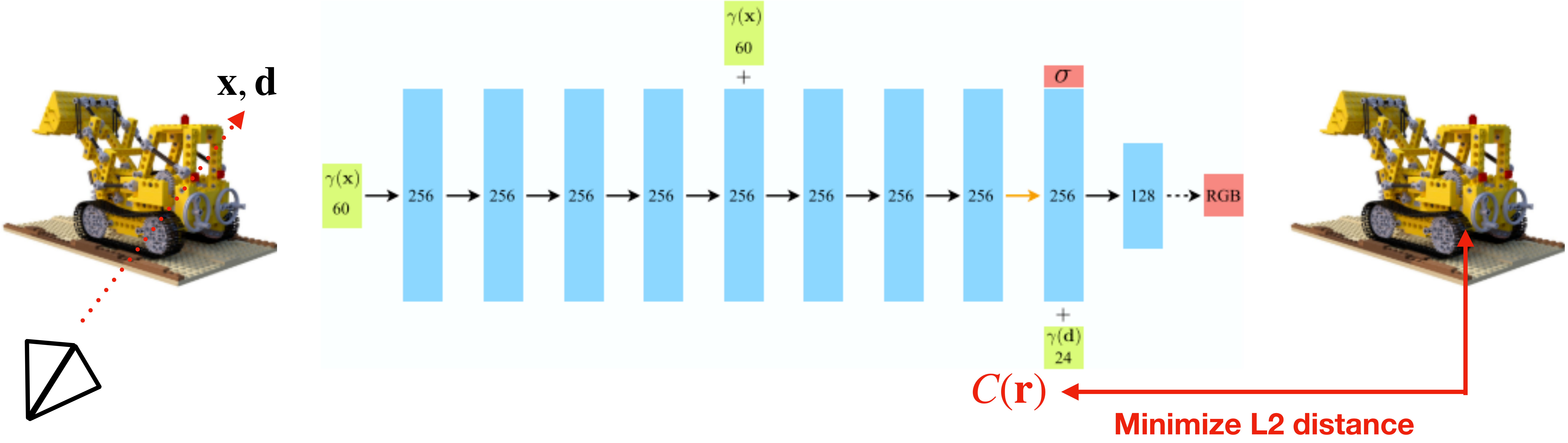
Recap: NeRF

Recap: NeRF



5D Input
Position + Direction
(x, y, z, θ, ϕ)
F_Θ
$(RGB\sigma)$
Output
Color + Density
Ray 1
Ray 2
(a)
(b)

Recap: NeRF



Learns both density and emitted radiance simultaneously

Recap: NeRF

Recap: NeRF

Constrained
camera motion

Static
environment lighting

Method Overview

Method Overview

Photos. Photos. Photos.

What if we can train NeRF to learn the scene representation of certain place, so that we can freely move around in the reconstructed scene?

Method Overview

Back in 2010…

Structure from Motion + Multi View Stereo = 3D Model

Image: Building Rome in a Day, Agarwal et al., ICCV 2010

Method Overview

Method Overview

Method Overview

Photometric variation

Time of day & atmospheric conditions directly impact the illumination

Method Overview

Photometric variation

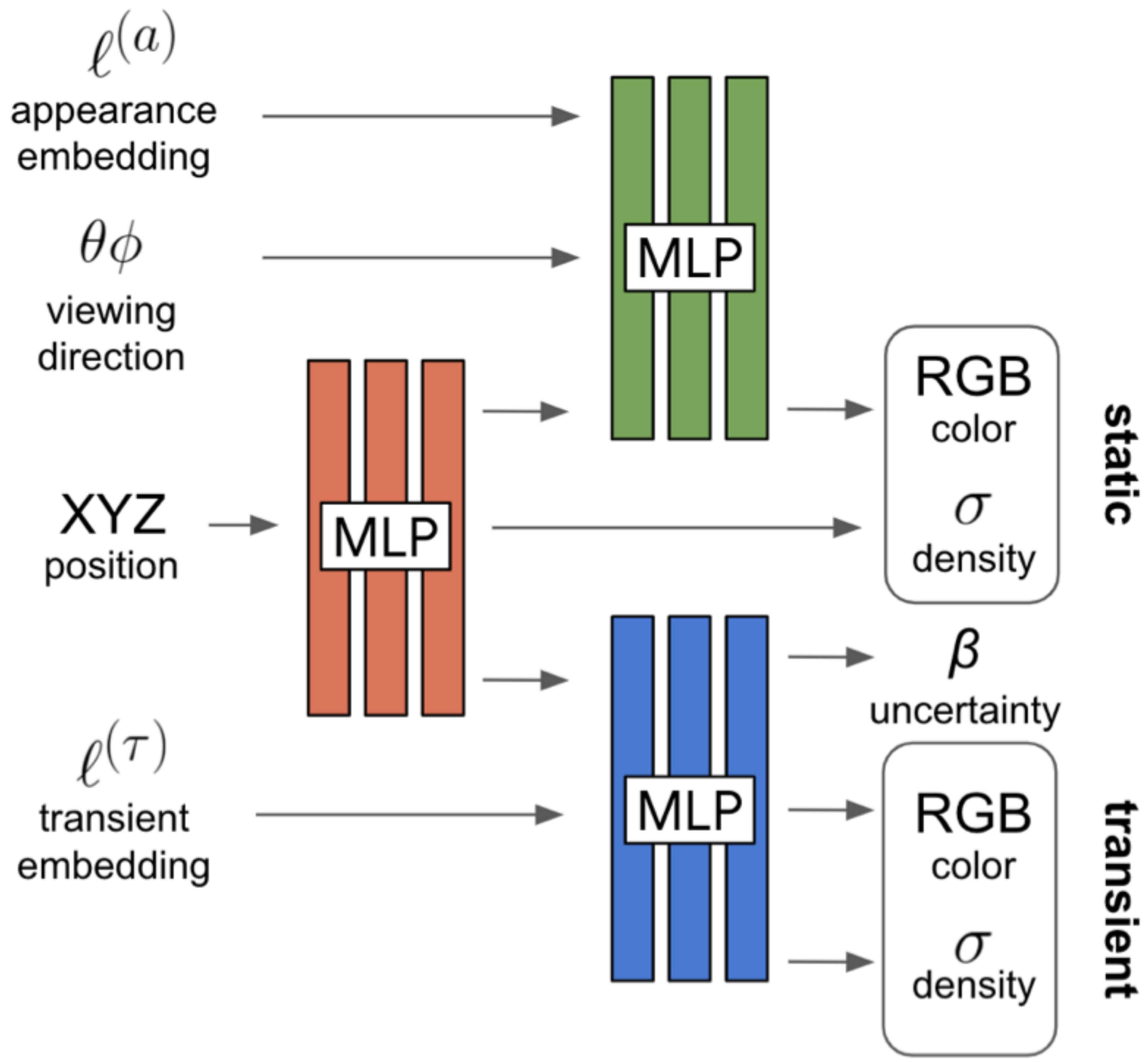
Time of day & atmospheric conditions directly impact the illumination

Transient objects

Real-world landmarks are usually occluded by other objects

cf. landmark(s) = static object(s)

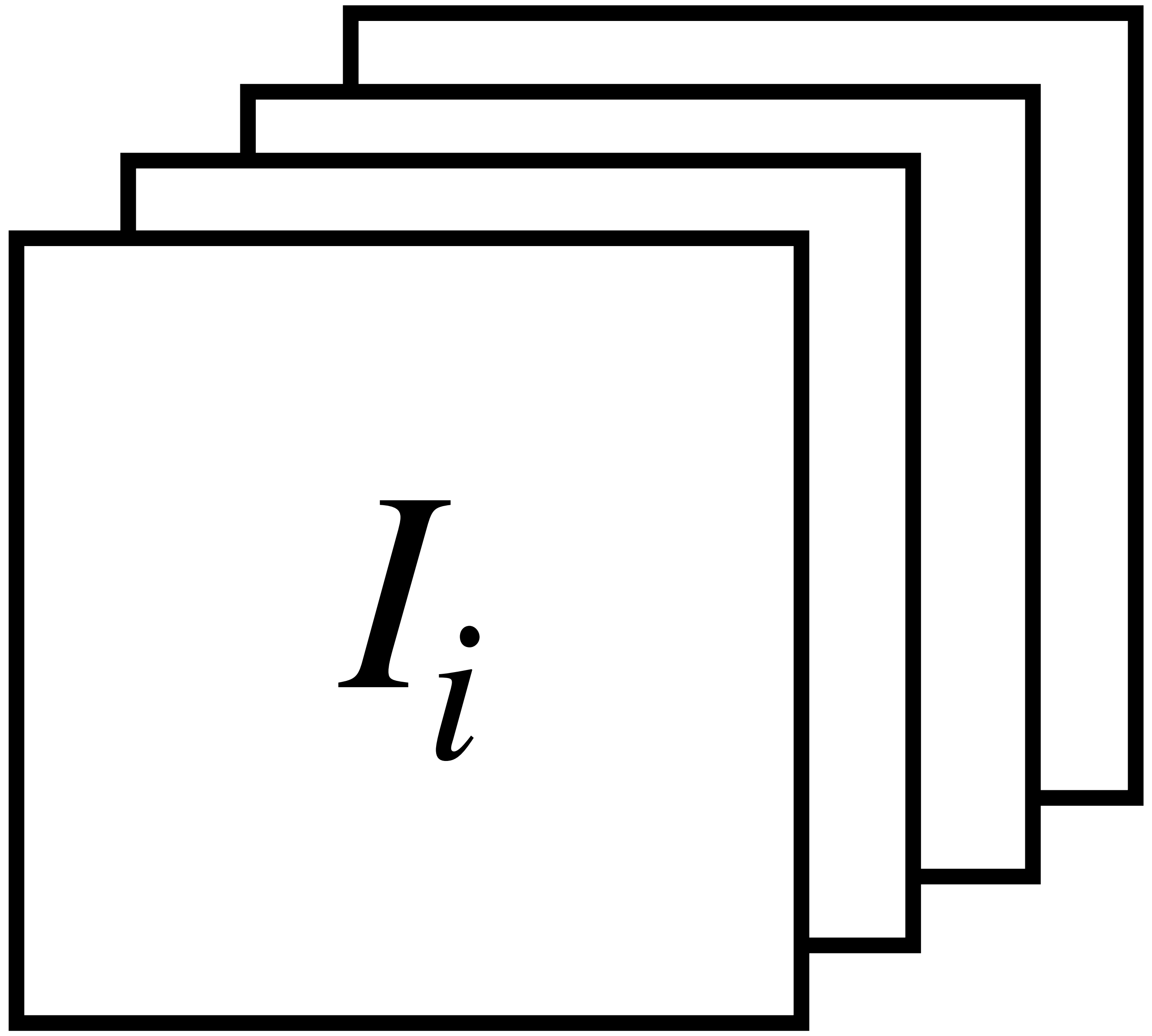
Method Overview



Three, instead of one

Latent Appearance Modeling

Latent Appearance Modeling



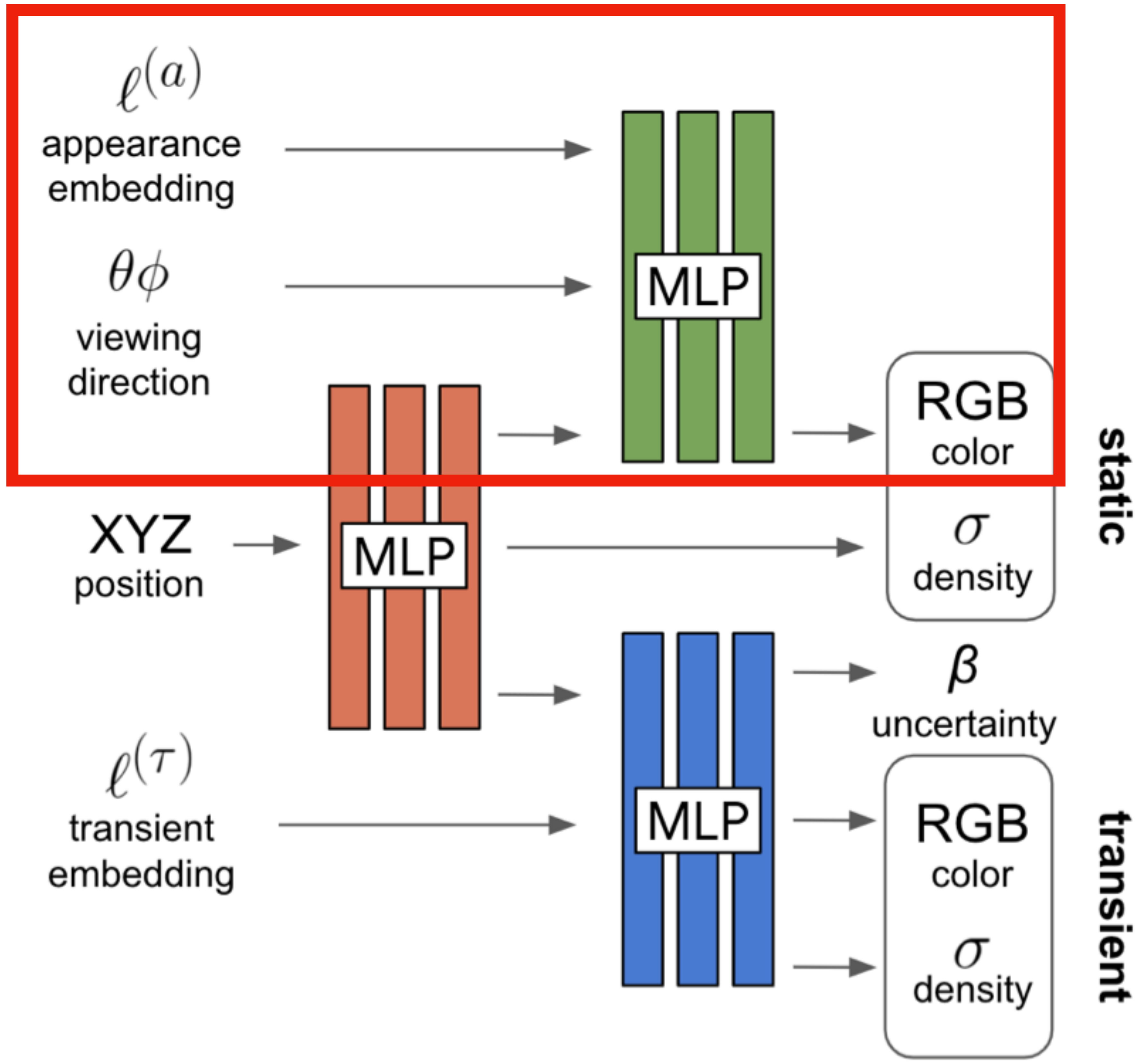
Generative Latent Optimization (GLO)

$$\xrightarrow[\text{optimize}]{\text{assign}} \quad \{\ell_i^{(a)} \mid i = 1,\ldots,N\}$$

Real-valued appearance embedding vector

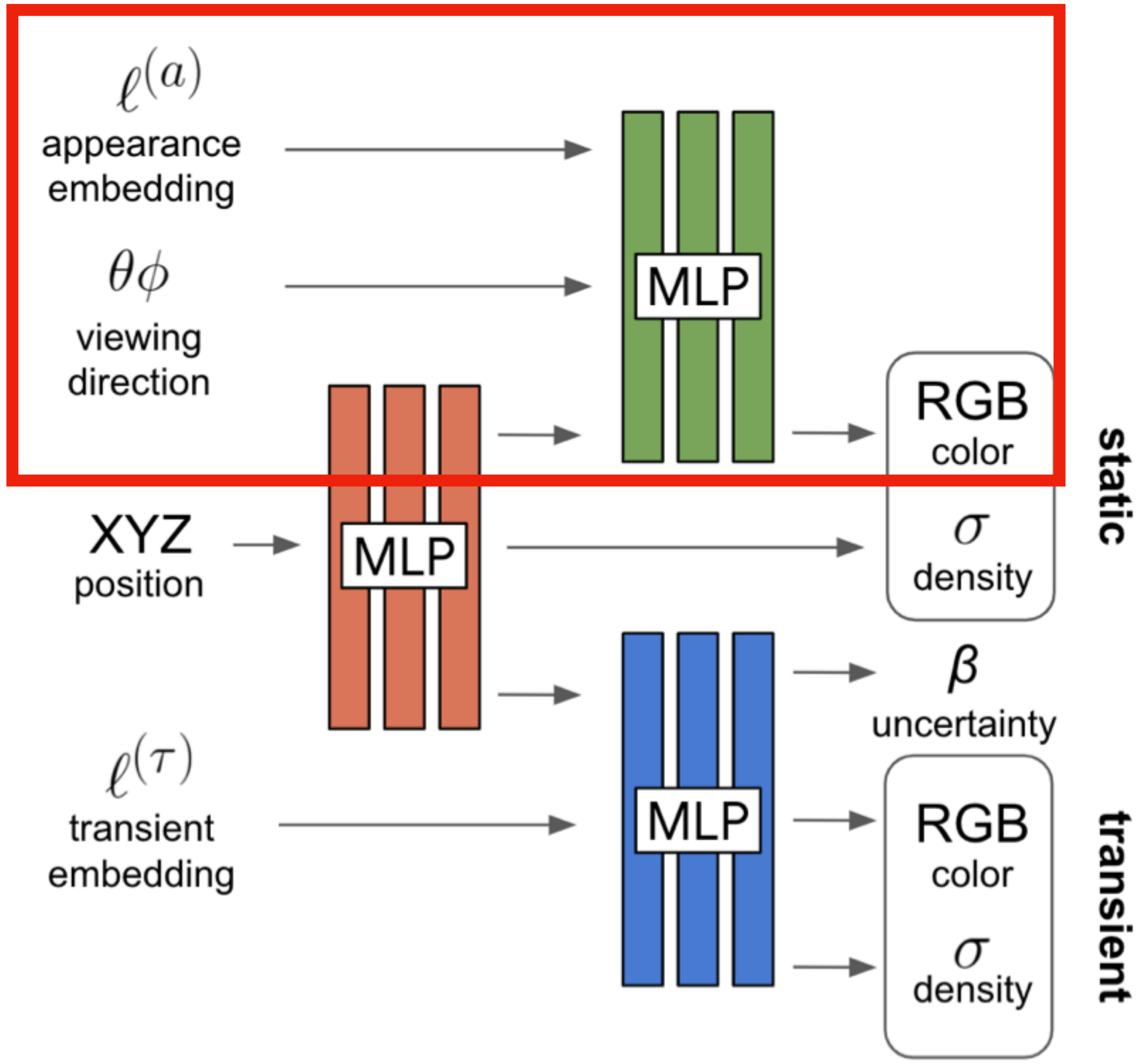
Allows radiance fields to have different color values for different images

Latent Appearance Modeling



Models the appearance of the static object

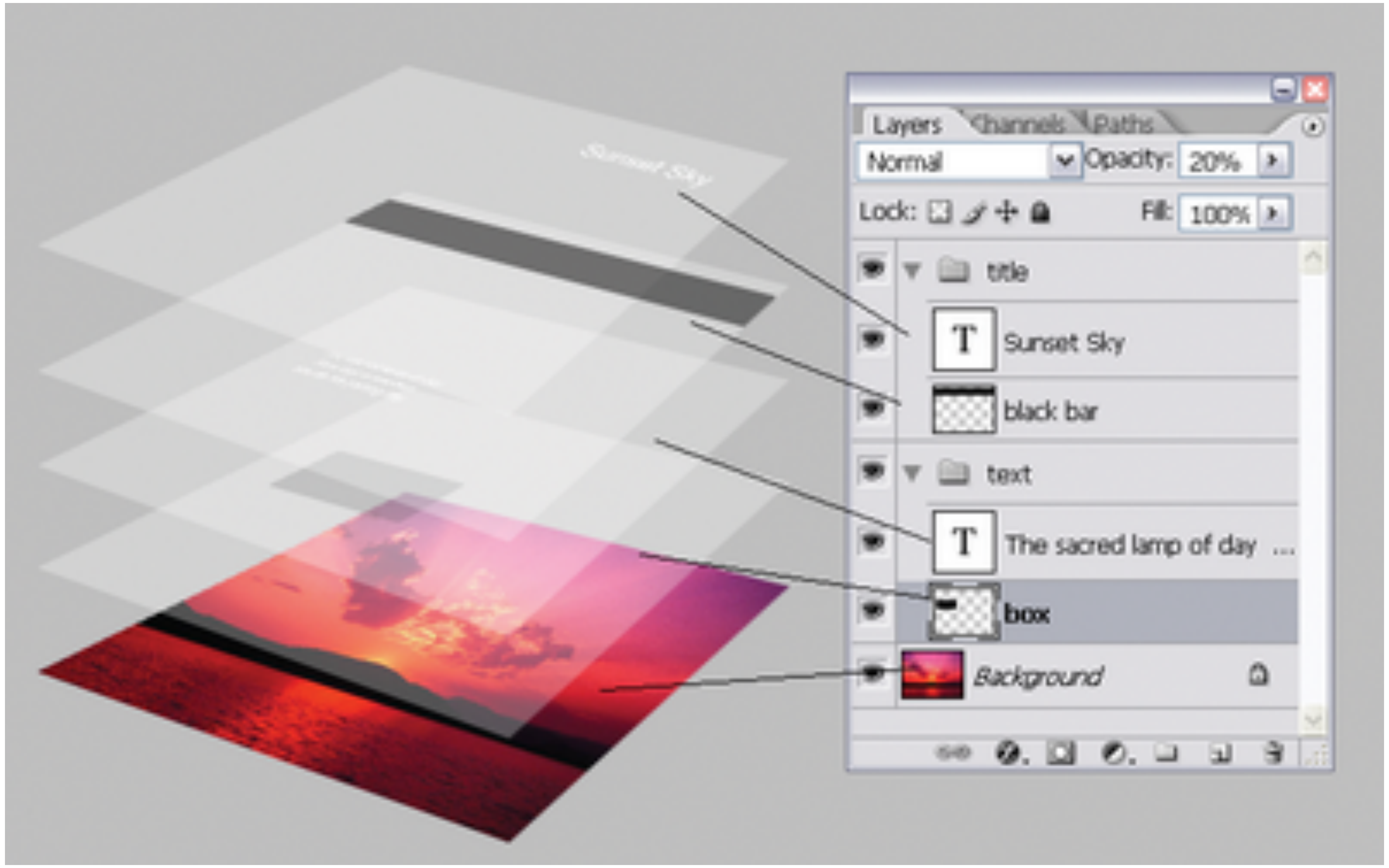
Latent Appearance Modeling



Interpolation between learned embeddings

Intuition: Image Compositing

Intuition: Image Compositing



Want to remove something?

Toggle visibility of that layer!

Image: https://graphicdesign.stackexchange.com/questions/27507/how-can-i-create-and-batch-this-3d-planes-effect

Dealing with Transient Objects

Dealing with Transient Objects

$$\hat{\mathbf{C}}_i(\mathbf{r}) = \sum_{k=1}^{K} T_i(t_k)\left(\alpha(\sigma(t_k)\delta_k)\mathbf{c}_i(t_k) + \alpha\left(\sigma_i^{(\tau)}(t_k)\delta_k\right)\mathbf{c}_i^{(\tau)}(t_k)\right)$$

where $T_i(t_k) = \exp\left(-\sum_{k'=1}^{k-1}\left(\sigma(t_{k'}) + \sigma_i^{(\tau)}(t_{k'})\right)\delta_{k'}\right)$

Dealing with Transient Objects

$$\hat{\mathbf{C}}_i(\mathbf{r}) = \sum_{k=1}^{K} T_i(t_k)\left(\alpha(\sigma(t_k)\delta_k)\mathbf{c}_i(t_k) + \alpha\left(\sigma_i^{(\tau)}(t_k)\delta_k\right)\mathbf{c}_i^{(\tau)}(t_k)\right)$$

$i = 0$

$i = \ldots$

$i = N$

Dealing with Transient Objects

$$\hat{\mathbf{C}}_i(\mathbf{r}) = \sum_{k=1}^{K} T_i(t_k)\left(\alpha(\sigma(t_k)\delta_k)\mathbf{c}_i(t_k) + \alpha\left(\sigma_i^{(\tau)}(t_k)\delta_k\right)\mathbf{c}_i^{(\tau)}(t_k)\right)$$

Static radiance field **Transient radiance field**

Dealing with Transient Objects

$$\hat{\mathbf{C}}_i(\mathbf{r}) = \sum_{k=1}^{K} T_i(t_k)\left(\alpha(\sigma(t_k)\delta_k)\mathbf{c}_i(t_k) + \alpha\left(\sigma_i^{(\tau)}(t_k)\delta_k\right)\mathbf{c}_i^{(\tau)}(t_k)\right)$$

Static radiance field

Transient radiance field

Static object density

Transient object density

"alpha compositing"

Dealing with Transient Objects

(a) Static (b) Transient (c) Composite

$\mathbf{c}_i(t)$

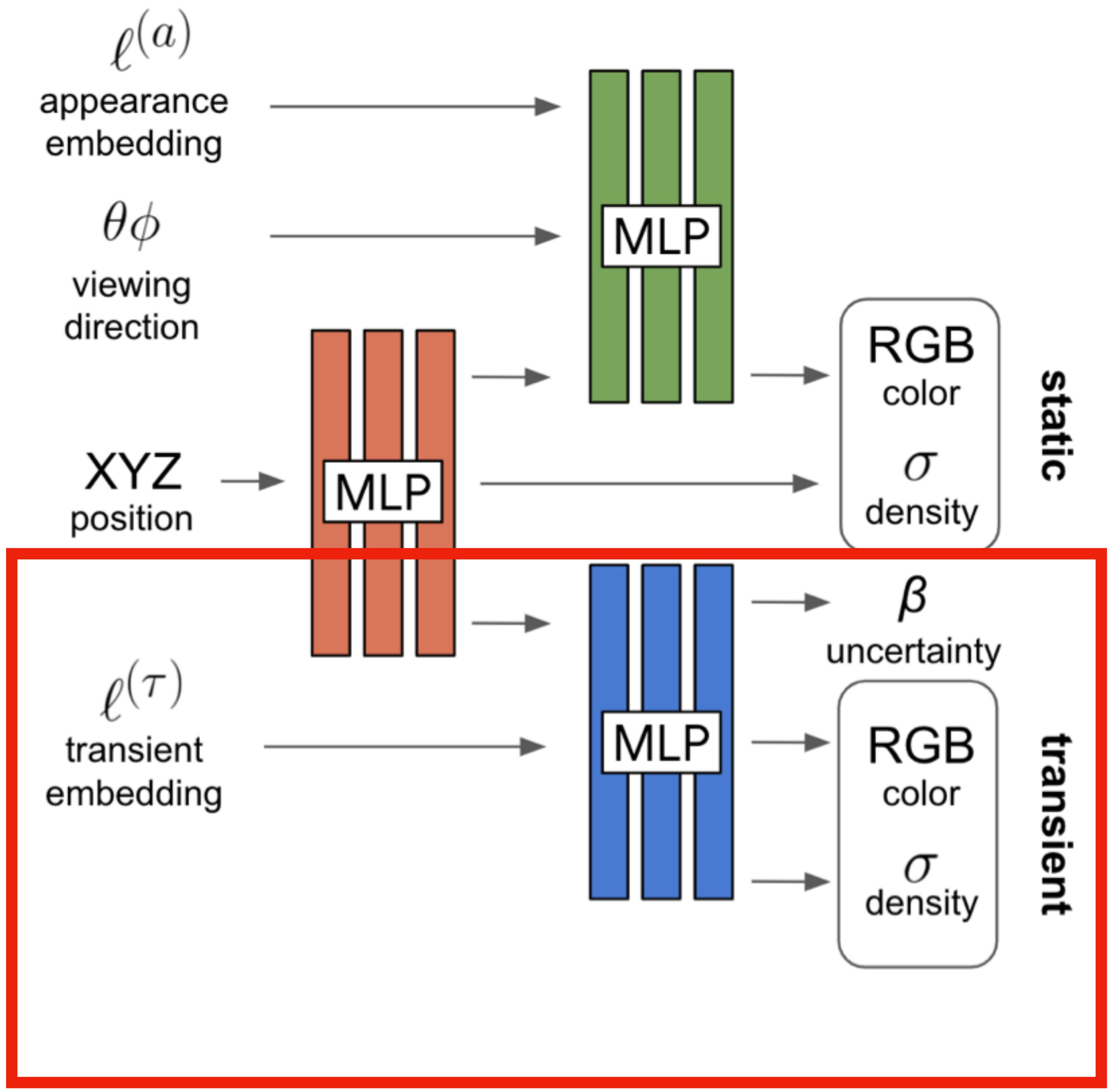
$\sigma(t)$

$\mathbf{c}_i^{(\tau)}(t)$

$\sigma^{(\tau)}(t)$

$$\alpha\big(\sigma(\,\cdot\,)\big)\mathbf{c}_i + \alpha\big(\sigma^{(\tau)}(\,\cdot\,)\big)\mathbf{c}_i^{(\tau)}$$

Dealing with Transient Objects



Transient embeddings $\{\ell_i^{(\tau)}\}_{i=1}^{N}$

are learned just like appearance embeddings

These embeddings are then decoded to:

1. **RGB color**
2. **Density**

of transient components

Dealing with Transient Objects

(a) Static (b) Transient (c) Composite (d) Image

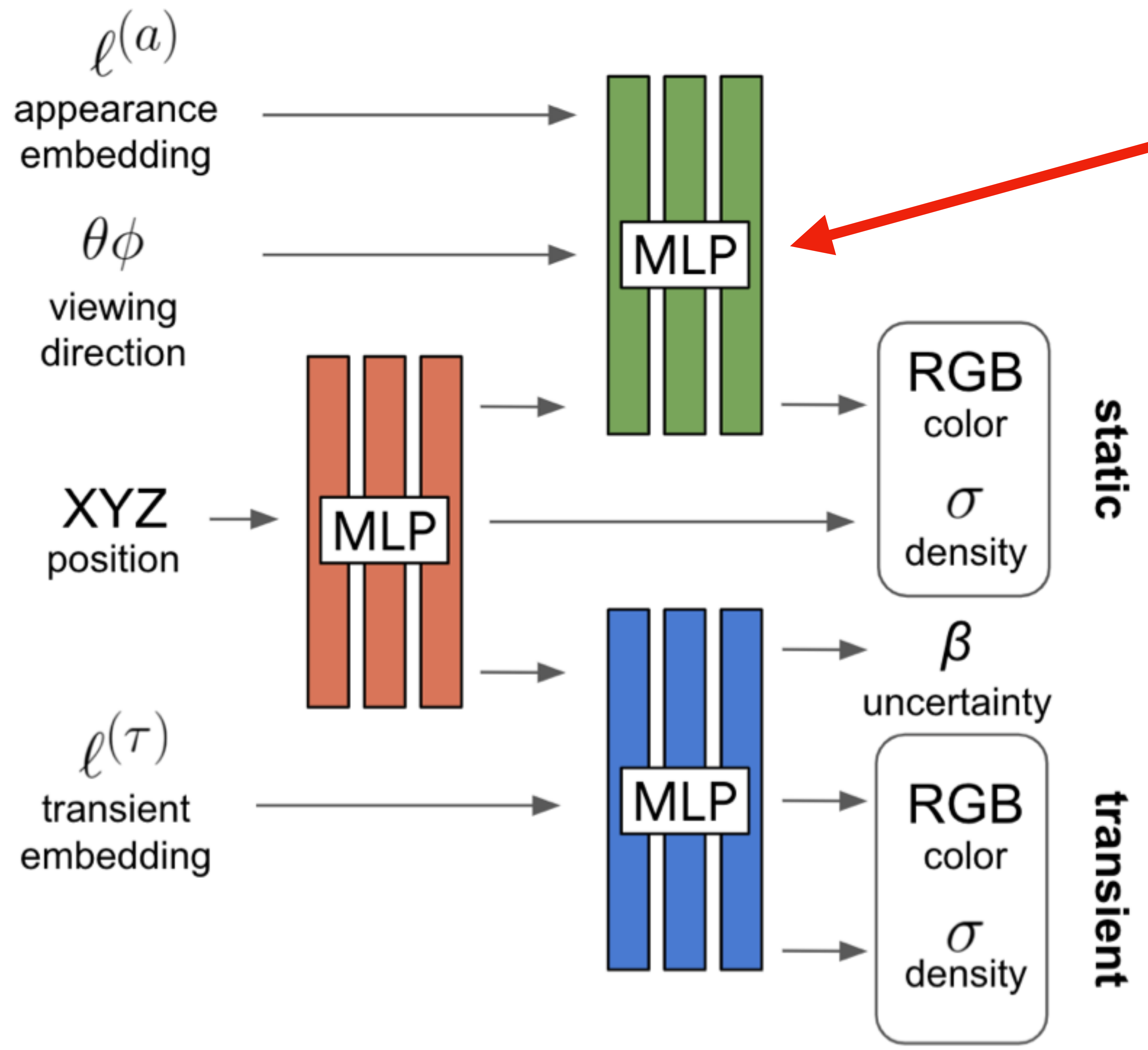
Just like NeRF, network parameters are optimized by minimizing the pixel-wise distance

Dealing with Transient Objects

To get nice and clear image of the landmark, render static image only!

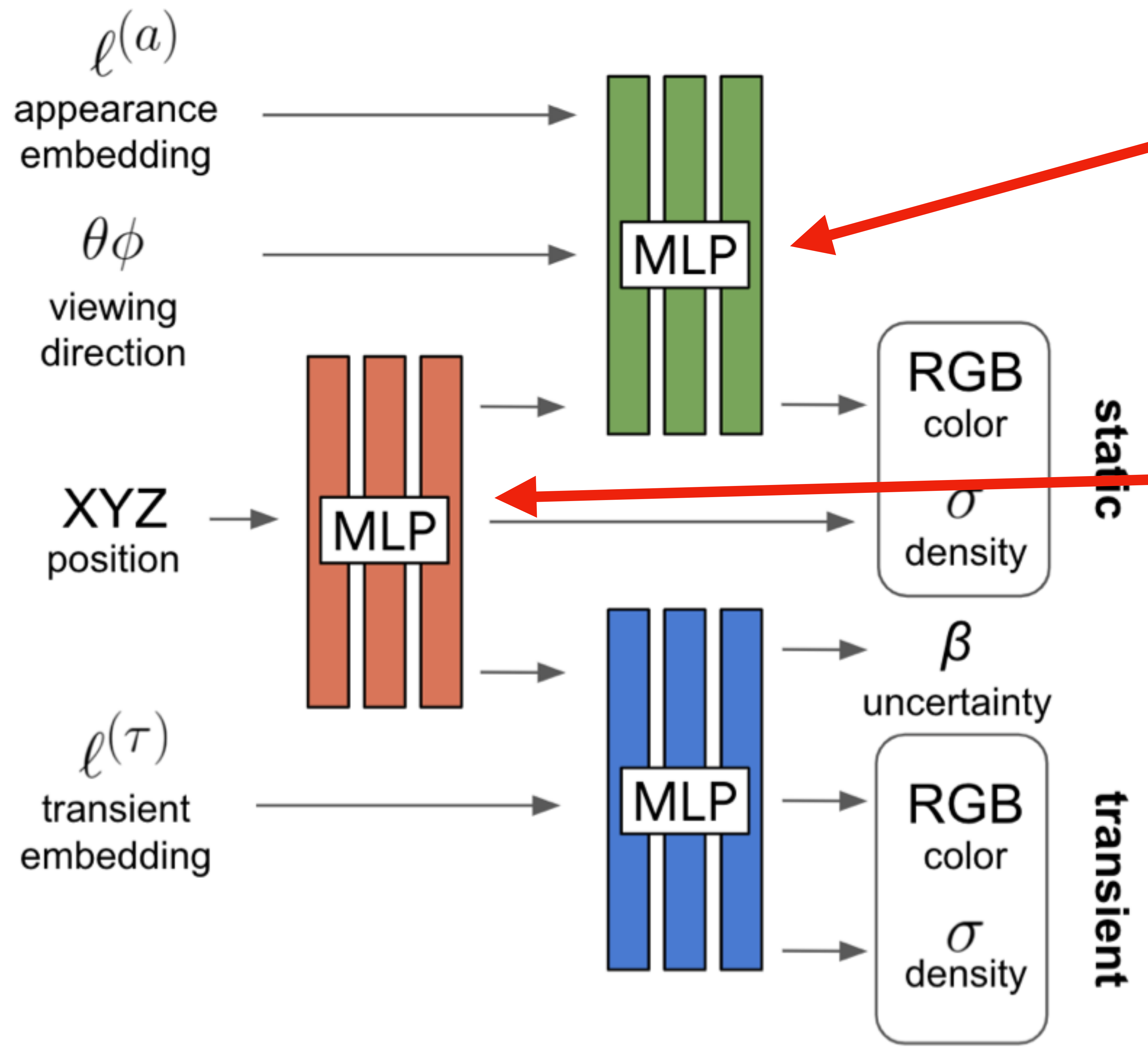
Summary

Summary



Network for modeling
appearance of static object

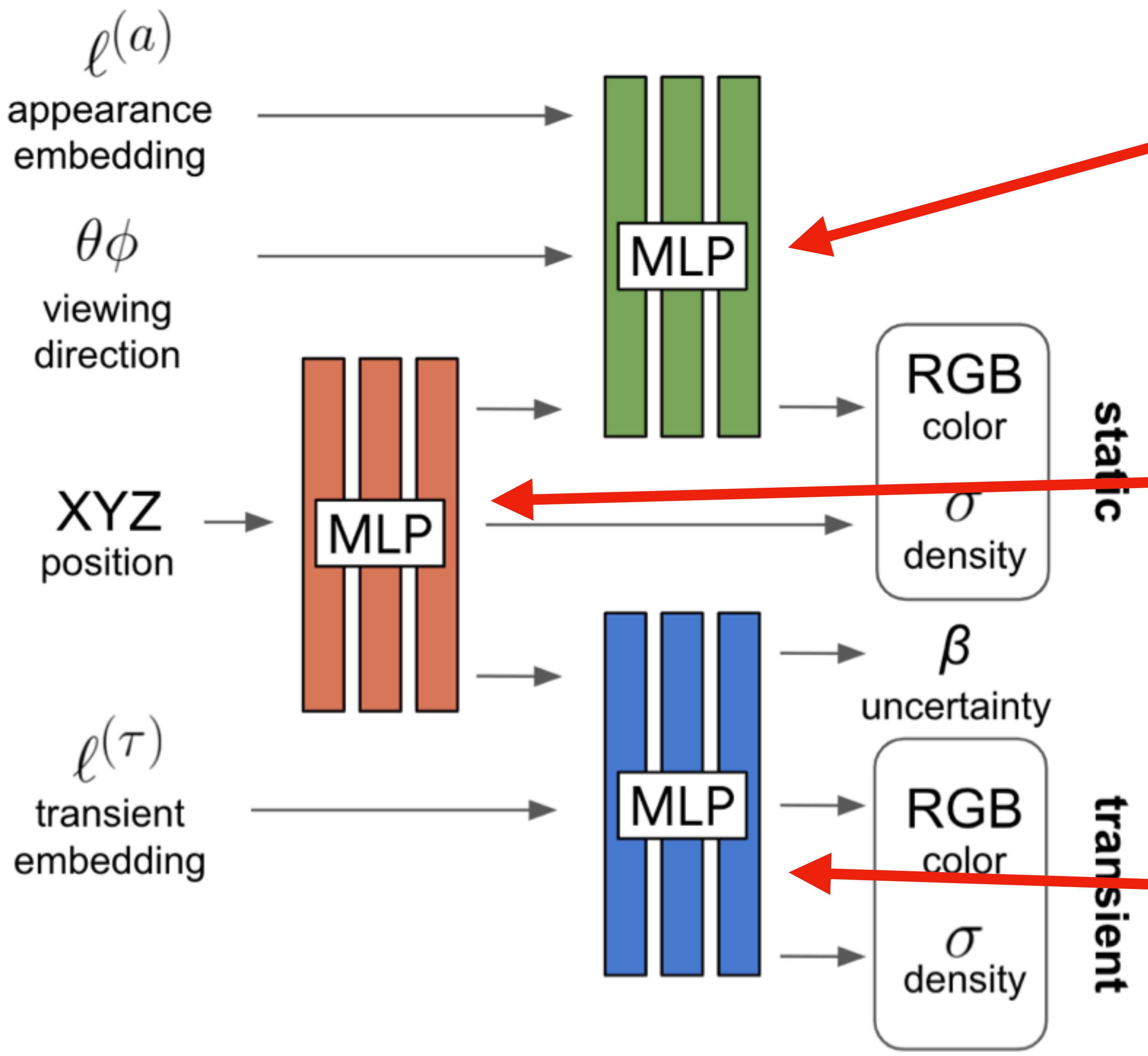
Summary



Network for modeling appearance of static object

Network for modeling density (i.e., geometry) of a scene

Summary



Network for modeling appearance of static object

Network for modeling density (i.e., geometry) of a scene

Network for modeling appearance of transient object(s)

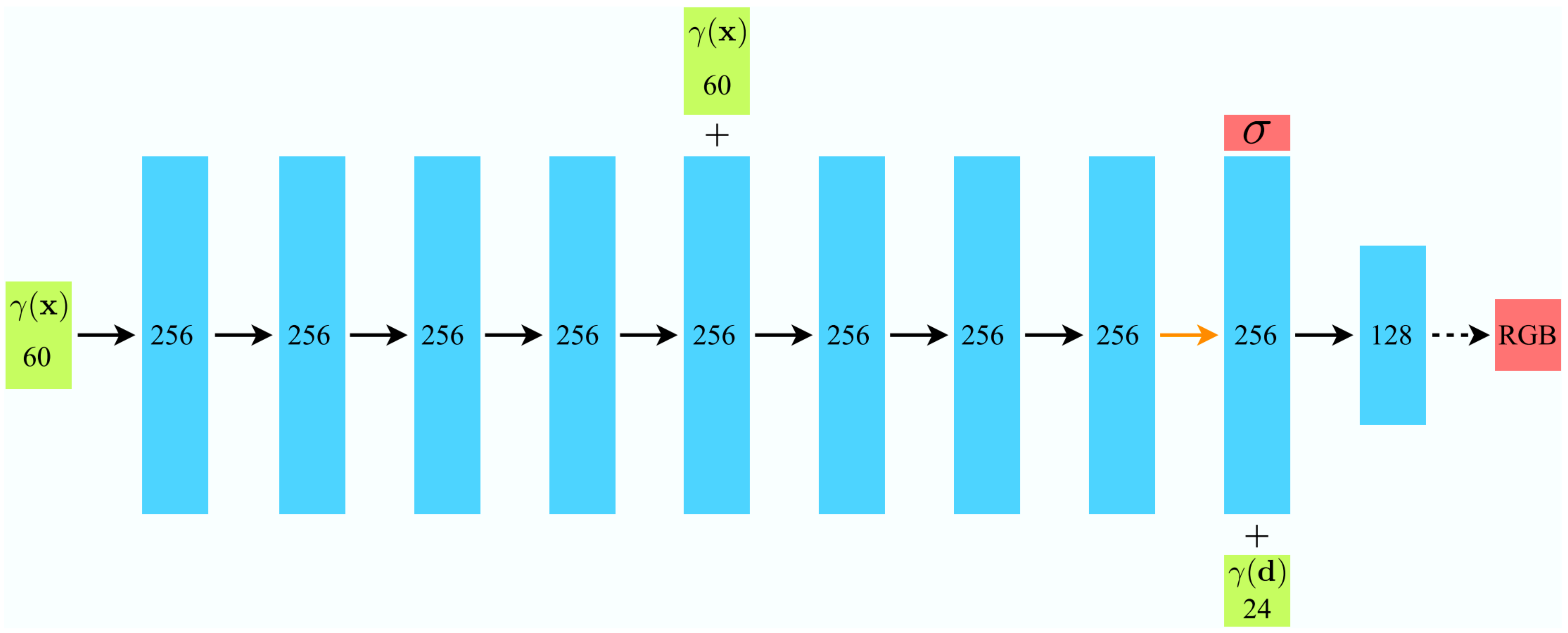
Experimental Results

Old Town Square

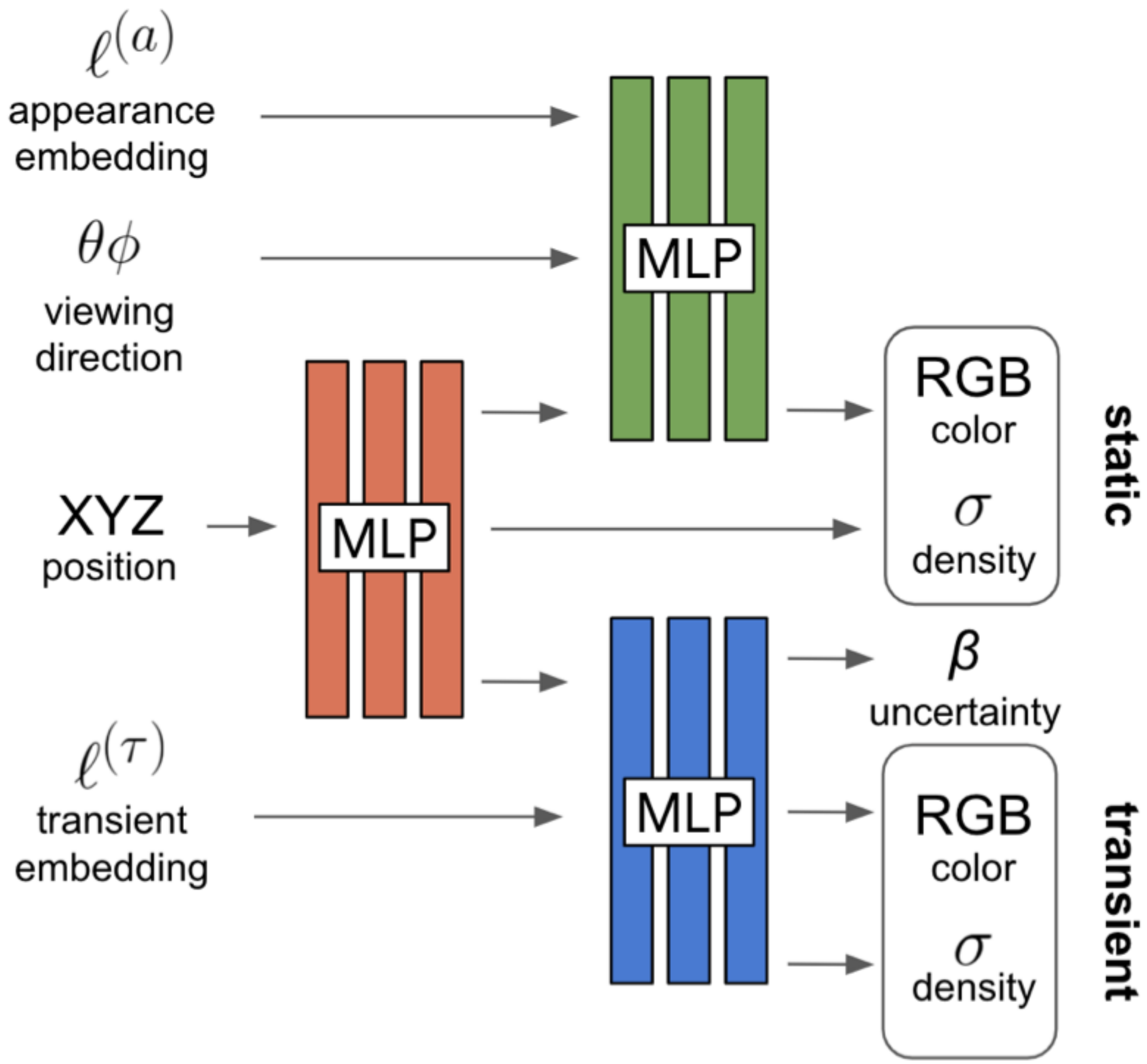
Prague, Czech Republic

Conclusion

Conclusion

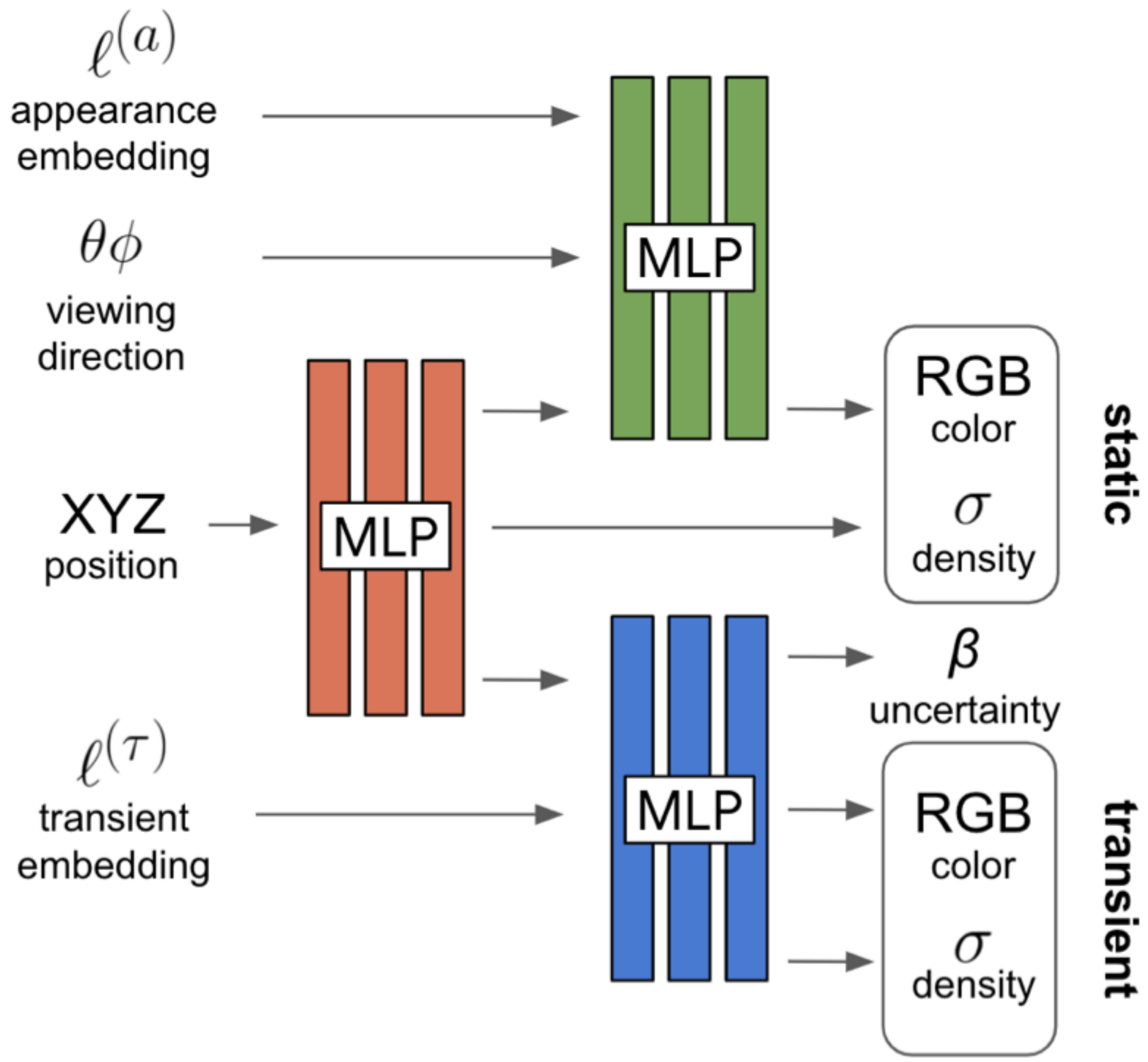


$\gamma(\mathbf{x})$ 60
256
256
256
256
256
256
256
256
256
128
RGB
$\gamma(\mathbf{x})$ 60
+
σ
+
$\gamma(\mathbf{d})$ 24



$\ell^{(a)}$
appearance embedding
$\theta\phi$
viewing direction
XYZ
position
$\ell^{(\tau)}$
transient embedding
MLP
MLP
MLP
RGB color
σ density
static
β
uncertainty
RGB color
σ density
transient

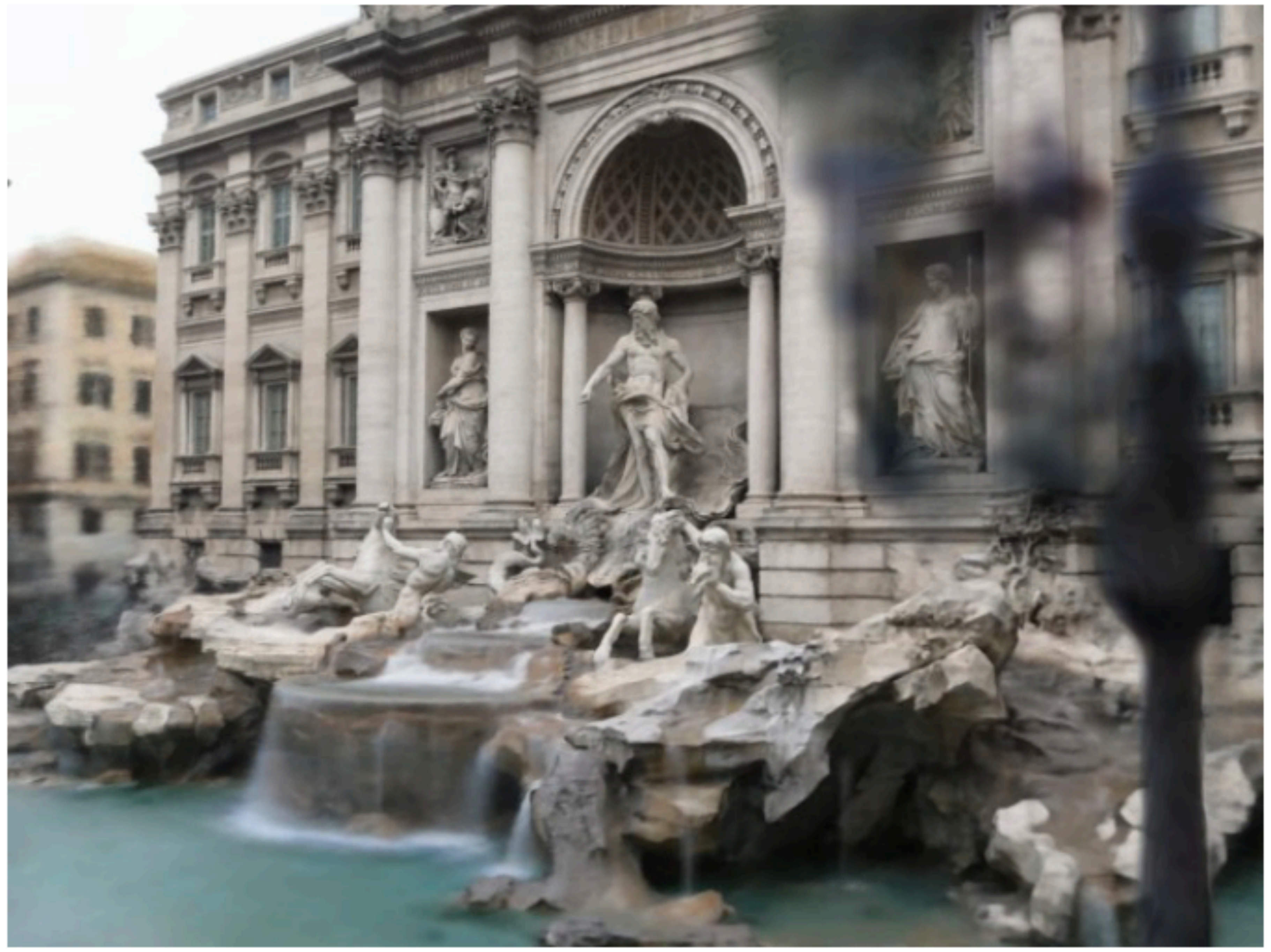
Conclusion



Generative Latent Optimization (GLO)

Disentangled Scene Representation (Static / Transient)

Conclusion



Thank You