BecomingLit: Relightable Gaussian Avatars with Hybrid Neural Shading
Supplementary Material

Anonymous Author(s)
Affiliation
Address
email

A Network Architecture

Our geometry module $\mathcal{F}_g$ maps FLAME [17] expression, jaw and eyes pose coefficients, and predicts per-texel attributes $\{\delta\boldsymbol{\mu}, \boldsymbol{q}, \boldsymbol{s}, \sigma, \boldsymbol{f}^{expr}\}_{k=1}^{M}$, where $\boldsymbol{f}_k^{expr}$ has a dimension of 32 in all experiments. We use the first 100 principal components for the expression parameter, and a rodrigues paramerization for jaw and both eye rotations. Hence, our input of shape $\mathbb{R}^{109}$ is transformed by a linear layer and then reshaped to $256 \times 8 \times 8$. A set of transposed convolutional layers then gradually upsamples the feature maps to the final output of shape $43 \times 512 \times 512$. We use leaky-ReLU as activation function for all layers except for the final output. For all convolutional layers, we adopt untied bias [4].

$\mathcal{F}_v$ takes as input the per-Gaussian feature map $\boldsymbol{f}^{expr}$, and the view direction, which is encoded using a single linear layer (8-dim output shape) and then expanded to the height and with dimension of the feature map. We concatenate the feature map and encoded view direction and feed it through a single convolutional layer which downsamples the input by half. Finally, a transposed convolutional layer maps the latent feature map back to its origial resolution with 4 output channels.

Our diffuse BRDF netowork $\mathcal{F}_d$ is a 3-layer MLP with hidden dimension 64 and leaky-ReLU activation in every layer, except the last one. The input is the concatenation of $\boldsymbol{f}_k^{expr}$ and the spherical harmonics coefficients of the incident light.

B Environment Map Rendering

In this section, we provide further details on how we render our avatars with all-frequency continuous illumination in the form of environment maps. While our diffuse BRDF trivially adopts to continuous illumination due to the spherical harmonics parameterization, we need to adopt the specular shading of the primitives.

For the specular pre-integration, we follow the split-sum approximation [14]. Karis et al. [14] propose to assume that the view direction $\boldsymbol{\omega}_o$ and the surface normal $\boldsymbol{n}$ are identical. With that assumption, the specular reflection is no longer view-dependent, and we can pre-integrate the environment map for different roughness values using a mipmap. In each mipmap level, we numerically integrate L_i with importance sampling using the Blinn-Phong distribution:

$$L_o^{specular}(\boldsymbol{x}, \boldsymbol{\omega}_o) = \int_\Omega L_i(\boldsymbol{\omega}_i) D(\boldsymbol{h}, \boldsymbol{n}, r^2)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}) d\boldsymbol{\omega}_i * \int_\Omega k_s \frac{DGF}{4(\boldsymbol{\omega}_o \boldsymbol{n})(\boldsymbol{n}\boldsymbol{\omega}_i)} d\boldsymbol{\omega}_i \tag{8}$$

The incoming illumination $L_i(\boldsymbol{\omega}_i)$ is now stored in the pre-integrated environment map $\hat{L}_{specular}(\boldsymbol{\omega}, r)$. During rendering, we linearly interpolate the mip levels to obtain the final radiance value for the roughness parameter. Hence, the new specular term becomes:

$$L_o^{specular}(\boldsymbol{x}, \boldsymbol{\omega}_o) \approx \hat{L}_{specular}(\boldsymbol{\omega}, r) \int_\Omega k_s \frac{DGF}{4(\boldsymbol{\omega}_o \boldsymbol{n})(\boldsymbol{n}\boldsymbol{\omega}_i)} d\boldsymbol{\omega}_i \tag{9}$$

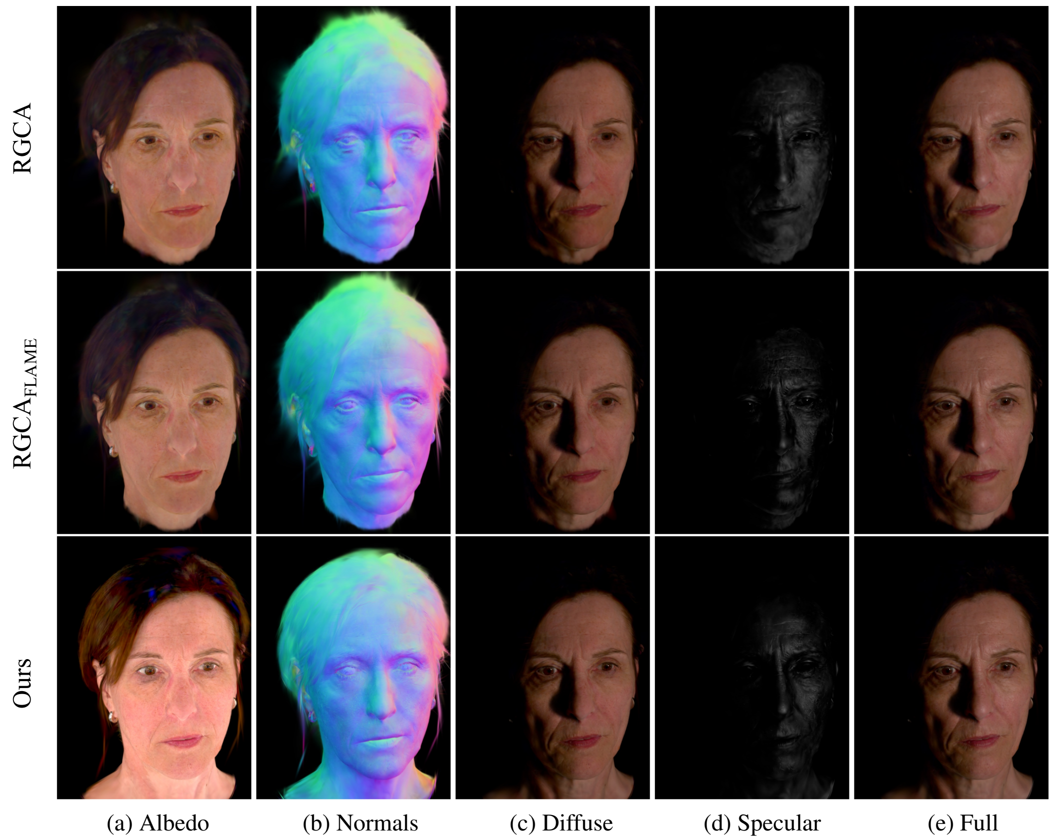

Figure 8: **Comparison on Intrinsic Decomposition**: We compare the recovered albedo (a) and normals (b), as well as the diffuse (c) and specular (d) contributions that sum up to the final rendering (e).

The remaining integral is essentially the integration of the BRDF with a completely white environment light. We can substitute the Fresnel term $F(\boldsymbol{\omega}_o, \boldsymbol{h})$ with the Schlick approximation [36] and factor out F_0:

$$\int_\Omega f_s(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)(\boldsymbol{n} \cdot \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i = F_0 \int_\Omega \frac{f_s(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)}{F(\boldsymbol{\omega}_o, \boldsymbol{h}} \left(1 - (1 - \boldsymbol{\omega}_o \cdot \boldsymbol{h})^5\right) (\boldsymbol{\omega}_i \cdot \boldsymbol{n}) d\boldsymbol{\omega}_i + \int_\Omega \frac{f_s(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)}{F(\boldsymbol{\omega}_o, \boldsymbol{h}} (1 - \boldsymbol{\omega}_o \cdot \boldsymbol{h}^5 (\boldsymbol{\omega}_i \cdot \boldsymbol{n}) d\boldsymbol{\omega}_i \quad (10)$$

These integrals depend on the two inputs $(\boldsymbol{\omega}_i \cdot \boldsymbol{n})$ and the roughness parameter r and act as a scale and bias to F_0. We pre-integrate both terms for all possible input combinations in $[0, 1]^2$ and store the two outputs in the 2D texture map $\hat{\boldsymbol{f}}_s(\boldsymbol{\omega}, r)$

During rendering, we can now compute the shaded color by evaluating the following terms:

$$\boldsymbol{\omega}_r = -\boldsymbol{\omega}_o - 2(-\boldsymbol{\omega}_o \cdot \boldsymbol{n})\boldsymbol{n} \quad (11)$$

$$a, b = \hat{\boldsymbol{f}}_s((\boldsymbol{\omega}_r \cdot \boldsymbol{n}), r) \quad (12)$$

$$c_k^s(\boldsymbol{\omega}_o) = (a\ k_s + b)\hat{L}_{specular}(\boldsymbol{\omega}_r, r) \quad (13)$$

C Additional Results

In Figure 8 we provide a qualitative comparison on the intrinsic decomposition performed by different methods. Our approach recovers a cleaner albedo and sharper specular highlights, resulting in more realistic renderings.

In addition, we ask the reviewers to watch our supplementary video for more results, which allow for a more complete comparison, including the temporal axis.

D Capture Script

For each participant of our dataset, we record 7 sequences in total. The first 6 consist of a predefined set of facial expressions, emotions and sentences that we ask the subjects to perform and read out. In the 7th sequence the participant is free to perform any facial expression for 20s. The instructions are given via a screen that is placed in front of the subject. In the following we provide a list of the single components, which during the capture sessions are accompanied with images.

- **Expressions-1**:
  - Head rotation with mouth open and closed
  - Eyes blink
  - Eyes squint
  - Eyebrows up / down
  - Puffed Cheeks
  - Mouth Vacuum
  - Nose Wrinkle
  - Lip bite
- **Expressions-2**
  - Grin (multiple variations)
  - Jaw movement
  - Lip licking
  - Tongue
- **Emotions**
  - Shout
  - Laugh
  - Surprise
  - Fear
  - Angry
  - Sad
  - Disgust
  - Happy
  - Confusion
  - Amazement
  - Embarrassment
- **Sentences-1**
  - A cramp is no small danger on a swim.
  - He said the same phrase thirty times.
  - Pluck the bright rose without leaves.
  - Two plus seven is less than ten.
  - The glow deepened in the eyes of the sweet girl.
  - By eating yogurt you may live longer.
- **Sentences-2**
  - Bring your problems to the wise chief.
  - Write a fond note to the friend you cherish.
  - Clothes and lodging are free to new men.
  - We frown when events take a bad turn.
  - Port is a strong wine with a smoky taste.
  - They had slapped their thighs.
- **Sentences-3**
  - She always jokes about too much garlic in his food.
  - Why put such a high value on being top dog.
  - All your wishful thinking won't change that.
  - Take charge of choosing her bridesmaids gowns.
  - Why buy oil when you always use mine.