

Hand-drawn Bokeh Generation

Nixon Chan

Abstract—Bokeh, the soft blur that appears in out-of-focus regions of a photograph, is widely used by photographers to guide attention and create a sense of depth and atmosphere. While many digital methods attempt to reproduce this effect, they often rely on complex optical simulations or learned models that limit direct artistic control. This work presents a simple, interpretable approach for generating depth-of-field effects using dense depth maps estimated from a depth prediction model.

Our method begins by converting the predicted depth field into discrete depth buckets that represent different ranges of distance in the scene. Each bucket is assigned a blur amount based on its position relative to the chosen focal plane. The blurred layers are then composited using a splatting procedure that preserves occlusion ordering and produces smooth transitions across depth boundaries.

This design avoids the artifacts and opacity issues common in naive layered blurring techniques.

The resulting pipeline is aimed at qualitative, visually driven bokeh generation rather than strict quantitative optimization. Qualitative evaluations show that the method produces depth-aware blur with a photographic feel while remaining easy to tune and understand.

The approach offers a practical alternative for artists and researchers who want controllable, depth-guided bokeh without the overhead of learning-based systems.

Index Terms—Bokeh, Depth Estimation



1 INTRODUCTION

BOKEH plays a central role in how photographers shape focus, depth, and visual atmosphere within an image. When a lens is used with a shallow depth of field, regions outside the focal plane blur into soft transitions, and bright points of light form distinctive highlight shapes tied to the lens aperture. This effect not only guides attention but also contributes to the emotional and aesthetic tone of a photograph. As digital imaging and computational photography continue to grow, reproducing convincing bokeh from a single image has become an active line of research.

Recent work has focused on learning-based systems that generate high-quality, depth-aware bokeh from a single image. BokehMe introduces a hybrid approach that combines a classical scattering-based renderer with a two-stage neural network that corrects artifacts caused by imperfect disparity maps [1]. This design allows the user to adjust focal plane, blur strength, and aperture shape while reducing the edge artifacts that often appear in purely classical pipelines. More recently, Bokehlicious extends controllable bokeh rendering through an aperture-aware attention mechanism and a large-scale dataset of real shallow-focus images [2]. By training on diverse focal settings, it produces realistic and spatially consistent bokeh with greater flexibility. Together, these methods illustrate the trend toward hybrid or fully learned architectures that emphasize realism and data-driven control.

In this work, we introduce a simple, interpretable pipeline for generating depth-aware bokeh from a single image using a dense depth map estimated by a depth prediction model. Our method converts the continuous depth field into discrete depth buckets, applies blur based on each bucket’s distance relative to the focal plane, and composites the results using a splatting strategy that preserves occlusion ordering and smooth transitions across depth boundaries. This design avoids the edge artifacts common in naive multilayer blurring while remaining easy to adjust for different visual preferences.

Our key technical contributions include a bucketization scheme that stabilizes blur estimation across noisy depth regions, a depth-guided splatting process for compositing blurred layers without manual masks, and a fully hand-crafted pipeline that prioritizes transparency and controllability. The method is evaluated primarily through qualitative results, reflecting its goal of producing visually pleasing, depth-consistent bokeh rather than optimizing numerical performance. While the approach does not aim to surpass state-of-the-art learned methods in quantitative metrics, it provides a lightweight and flexible tool for users who value interpretability and stylistic control.

2 RELATED WORK

2.1 Depth Estimation for Bokeh Generation

Depth estimation is a critical component for depth-aware bokeh generation because the blur radius, occlusion ordering, and transition quality all depend on having a coherent depth field. Depth Anything V2 is trained on a large mixture of datasets with diverse geometry and lighting conditions, allowing it to generalize well to new scenes without fine-tuning [3]. It produces dense, relatively smooth depth maps that are suitable for depth-of-field rendering, even when depth boundaries are imperfect or details are noisy. Importantly for our work, the model is efficient enough to be used as a preprocessing stage and does not require specialized calibration or paired training data. While Depth Anything is not explicitly designed for bokeh, its robustness makes it a strong backbone for pipelines that manipulate depth to create stylistic blur.

2.2 Hybrid Classical-Neural Rendering

Hybrid methods combine physically motivated algorithms with learning-based corrections to overcome the weaknesses of each component. BokehMe [1] is an example of this

strategy. BokehMe begins with a classical scattering-based renderer, which models how rays spread according to aperture size and focus distance. However, classical renderers amplify errors when the disparity map is noisy or when depth boundaries are sharp but inaccurate. To address this, BokehMe uses a two-stage neural network that detects problematic regions and refines both the blur shape and local image structure. This hybrid design allows the system to retain the controllability of classical approaches (blur radius, aperture shape, focal plane) while leveraging neural priors to repair artifacts. For applications that require high perceptual realism without sacrificing user control, BokehMe offers a balanced trade-off.

2.3 Fully Learning-Based Bokeh Synthesis

Fully neural approaches attempt to learn the mapping from an all-in-focus image and a depth map directly to a shallow-depth-of-field rendering. Bokehlicious [2] pushes this direction further by explicitly modeling the aperture within the network. Its key innovation is an aperture-aware attention mechanism that conditions the blur synthesis on desired aperture shapes, sizes, and bokeh characteristics. This is trained using RealBokeh, a large, diverse dataset of real photographs taken with varying apertures and lenses, providing a rich supervisory signal for how physical optics behave. Because of its training data and architecture, Bokehlicious produces spatially consistent, natural-looking bokeh that replicates lens-specific aesthetic traits. However, like most large learned systems, it trades interpretability and fine-grained manual control for realism and generalization.

2.4 Bokeh Datasets for Training and Evaluation

Datasets play a central role in evaluating bokeh synthesis methods. While RealBokeh [2] offers high-quality real optical ground truth, the EBB! dataset [4] (Everything is Better with Bokeh!) serves as a dedicated benchmark for assessing shallow-depth-of-field generation. It provides paired all-in-focus and shallow-focus images captured under controlled conditions, along with consistent camera metadata, making it well suited for systematic evaluation of blur accuracy, occlusion behavior, and aperture characteristics. The dataset also exposes common failure modes that are difficult to identify without true shallow-focus reference images. As a result, EBB! is widely used as a standardized benchmark for comparing both classical and learning-based bokeh generation approaches [4].

2.5 Method

Compared to hybrid or fully learned systems, our method focuses on a purely hand-crafted pipeline driven by dense depth maps. By using Depth Anything V2 [3] for initial depth estimation, followed by depth bucketization, per-bucket blur assignment, and splatting-based compositing, the approach maintains complete transparency and predictability. It does not aim to surpass state-of-the-art learned systems in photorealistic rendering. Instead, it targets controllable, interpretable, and qualitatively pleasing bokeh that behaves in a predictable way. In this sense, our work



Fig. 1. Original sample image.

complements prior methods by offering a lightweight alternative for applications where artistic control, simplicity, or interpretability is more important than dataset-driven photorealism.

3 PROPOSED METHOD

We propose a deterministic, depth-aware bokeh synthesis pipeline that combines monocular depth estimation, segmentation-guided layer refinement, depth discretization, progressive spatial filtering, and highlight splatting. The method requires no model training and provides explicit control over focal depth, blur magnitude, and highlight appearance. Let $I \in \mathbb{R}^{H \times W \times 3}$ denote an RGB input image. Figure 1 shows a sample starting image.

3.1 Depth Estimation

We estimate a dense depth map using the Depth-Anything V2 monocular depth estimator [3]. The model outputs an unnormalized depth prediction $\tilde{D}$, which we convert to a normalized relative depth map

$$D = \frac{\tilde{D} - \min(\tilde{D})}{\max(\tilde{D}) - \min(\tilde{D})} \in [0, 1]^{H \times W}.$$

Normalized depth allows consistent mapping between depth and blur levels across images. Because the goal is artistic shallow-focus rendering, relative depth consistency is sufficient. Figure 2 shows a sample depth estimated image.

3.2 Depth Bucketization

We represent depth-dependent blur using a discrete set of K blur levels. Given a focal depth $d_f \in [0, 1]$ (either user-specified or estimated from the midground layer), we compute the absolute focal deviation

$$\Delta(x) = |D(x) - d_f|.$$

Let $\{e_k\}_{k=0}^K$ be linearly spaced bucket edges over $[0, \Delta_{\max}]$, where $\Delta_{\max} = \max_x \Delta(x)$. Each pixel is assigned a bucket index

$$B(x) = k \quad \text{such that} \quad e_k \leq \Delta(x) < e_{k+1}.$$



Fig. 2. Depth estimated image.



Fig. 4. Progressively Gaussian blurred image.

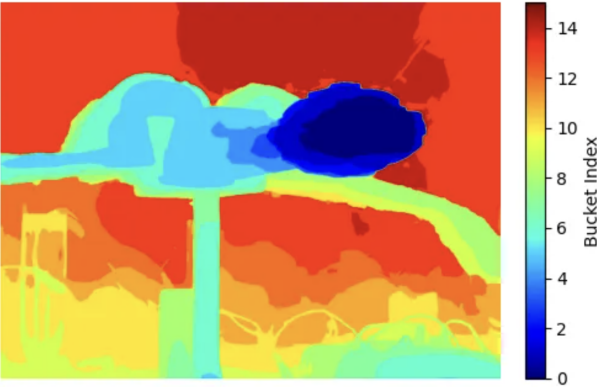


Fig. 3. Depth bucketized image.

Bucketization provides controlled, monotonic blur progression without requiring full optical PSF simulation. Figure 3 shows a sample depth bucketized image.

3.3 Progressive Spatial Filtering

For each bucket k , we define an odd-sized Gaussian kernel with standard deviation

$$\sigma_k = \alpha \frac{k}{K},$$

where α corresponds to the maximum blur strength. The resulting blur kernel size s_k satisfies $s_k = 2\lfloor 3\sigma_k \rfloor + 1$.

For each bucket, we compute a blurred version of the image I_k using separable Gaussian filtering. The per-pixel blurred output (ignoring compositing for now) is:

$$I_{\text{blur}}(x) = I_{B(x)}(x).$$

Figure 4 shows a sample progressively blurred image.

3.4 Highlight Splatting

To reproduce the characteristic bright highlights of bokeh, we apply a splatting operator to local intensity maxima. Let Y denote the grayscale luminance of I . Peaks are detected via

$$P = \{x : Y(x) = (\max_{\|u-x\| \leq r} Y(u)) \wedge Y(x) > T\},$$



Fig. 5. Sample user drawn kernel.

where r is a neighborhood radius and T is a high percentile threshold (typically 95th). For each peak $x \in P$, we convolve its color value with a user-specified or circular PSF kernel K_{bokeh} :

$$S(x) = c(x) \cdot K_{\text{bokeh}},$$

and additively blend all splats into the corresponding blurred layer. This procedure accentuates bright highlights while remaining computationally lightweight.

3.5 Layered Foreground–Background Compositing

To prevent halo artifacts and create realistic depth transitions, we perform a layered compositing procedure. Background pixels are first assigned their bucket-dependent blur. Foreground layers are composited using a soft alpha mask. For a foreground region L_f , we compute a blurred alpha field

$$A_f = G_{\sigma_k} * L_f,$$

where G_{σ_k} is a Gaussian with the same scale as the foreground blur. Final compositing for each bucket k follows

$$I_{\text{final}} = A_f \odot I_k + (1 - A_f) \odot I_{\text{final}},$$

proceeding from background layers toward the foreground. Figure 5 shows a sample user drawn kernel. Figure 6 shows a sample splatted image.

4 DESIGN DECISIONS

The design of the proposed pipeline emphasizes simplicity, controllability, and ease of integration rather than maximizing reconstruction accuracy. A primary choice was to rely entirely on monocular depth estimation from Depth



Fig. 6. Splatted and composited image.



Fig. 7. Long range focal depth splatted image.

Anything V2 [3]. While depth maps predicted from a single RGB image contain noise and may fail on thin structures, they provide a sufficiently stable ordering of scene geometry for producing visually coherent foreground-background separation. Importantly, this avoids the need for multi-view input, semantic labels, or additional training, making the method accessible for general use.

A second key decision was to adopt a bucketized blur approach rather than computing a continuous blur radius per pixel. Bucketization is computationally efficient, easy to tune, and produces predictable behavior. Users can adjust the number of buckets and the maximum blur kernel to control the strength of the depth of field effect without modifying the underlying model. Although this discretization introduces stepwise transitions, it offers a practical trade-off between speed and visual smoothness.

We also designed the system to operate with minimal assumptions about scene structure. Instead of enforcing a strict layered compositing model, the method blends foreground and background content through depth-based masks, allowing it to handle typical portrait and object-focused scenes without heavy preprocessing. This choice reduces complexity but can lead to abrupt transitions when depth discontinuities are large or inaccurate.

Finally, highlight splatting is included as an optional enhancement to approximate lens-shaped bokeh responses. By restricting splatting to detected bright peaks, the method avoids the high cost of full-kernel convolution while still enabling customizable aperture shapes and stylized rendering.

Overall, these decisions prioritize user adjustability, straightforward implementation, and reasonable visual quality across everyday scenes, acknowledging that the method is not optimized for peak quantitative performance.

5 EXPERIMENTAL RESULTS

We evaluate our hand-crafted bokeh generation pipeline across a variety of scenes. The goal of these experiments is to assess the qualitative behavior of the system, specifically its ability to produce controllable shallow-focus effects with coherent foreground-background separation and stylistically consistent blur patterns. All experiments were conducted on images drawn from a combination of internet photography,



Fig. 8. Small depth of field splatted image.

personal captures, and a subset of scenes from the EBB! [4] and RealBokeh [2] datasets, which provides challenging cases with fine-grained aperture variations and complex backgrounds.

5.1 Qualitative Evaluation

Across a wide range of scenes, our bucketization strategy produces stable depth segmentation that avoids the abrupt transitions and haloing artifacts common in classical layered-blur methods. The depth-aware splatting stage yields spatially coherent bokeh shapes whose size and distribution vary smoothly with depth, enabling visually convincing approximations of shallow-focus photography. Thin structures and high-frequency textures are generally preserved without introducing double edges or over-smoothed boundaries. These results highlight the interpretability of the pipeline: each stage has a predictable effect on the final appearance, granting users fine-grained control over focal depth, blur falloff, and aperture shape. Figure 7 shows a generated image with a long range refocused focal depth. Figure 8 shows a generated image with a small depth of field.



Fig. 9. Star bokeh splatted image.



Fig. 10. Star bokeh splatted image.

5.2 Aperture and Style Control

Because the rendering process is fully transparent, users can easily manipulate stylistic parameters. Experiments with circular, polygonal, and custom aperture kernels demonstrate that the splatting stage accurately reflects the intended aperture geometry. Similarly, adjusting bucket thresholds allows a user to emphasize artistic blur or produce subtler depth transitions. This level of control is difficult to achieve with end-to-end neural systems, which often inherit stylistic biases from their training data. Figure 9 shows a generated image with a star aperture shape. This method is also capable of generating images with typical bokeh shapes found in regular cameras, shown in figure 10.

5.3 Scene Variety

The method performs reasonably well across a small range of everyday scenes, as shown in Figures 11–14. For instance, in the portrait example, the pipeline is able to separate the subject from the background sufficiently to produce a plausible shallow-depth-of-field effect, despite some limitations in edge precision. Similarly, the collie example demonstrates that the method can handle non-human subjects and outdoor lighting conditions, though fine fur details and irregular boundaries remain challenging. Overall, these



Fig. 11. Sample original portrait image.



Fig. 12. Portrait splatted image.

results illustrate that while the approach is not robust to highly complex geometry or difficult textures, it can generate coherent bokeh effects for common scene types when the depth estimation is reasonably accurate.

5.4 Quantitative Evaluation

Although the primary goal of this work is controllable qualitative generation rather than optimized reconstruction accuracy, we include a small-scale quantitative comparison using PSNR, evaluated on the EBB! [4] validation set. As shown in Table 1, Bokehlicious [2] achieves the highest PSNR of 25.02, consistent with its fully learned design and training on the RealBokeh [2] dataset. BokehMe [1] attains 23.85, reflecting its hybrid classical–neural architecture. Our method reaches 24.42, which is competitive with these learned systems even though it does not rely on training or large datasets. While PSNR alone does not capture perceptual realism or stylistic flexibility, these results illustrate that the proposed pipeline maintains reasonable reconstruction fidelity while offering advantages in interpretability and artistic control.

TABLE 1
PSNR results on the EBB validation dataset.

Method	PSNR (dB)
BokehMe [1]	23.85
Bokehlicious [2]	25.02
Ours	24.42



Fig. 13. Sample original collie image.



Fig. 15. Sample failure case image.



Fig. 14. Collie splatted image.

5.5 Failure Cases

The method's primary limitations stem from how segmentation masks interact with depth-based blur transitions. In scenes where a foreground object is intentionally blurred, the boundary between the foreground mask and surrounding regions can become too abrupt, producing a visually harsh transition. This occurs because the alpha-blurred mask does not always provide sufficient falloff, causing the blur to appear artificially clipped rather than smoothly integrated into the scene. The sample failure case illustrates this effect: although the depth ordering is correct, the blurred foreground object exhibits an edge that feels cut out rather than optically defocused. Semi-transparent materials, fine structures, or objects with complex silhouettes exacerbate this issue. While additional feathering or multi-scale compositing can partially improve these artifacts, abrupt mask transitions remain a known limitation of mask-based layered bokeh synthesis. Figure 15 shows a sample failure case when the foreground is out of focus.

6 FUTURE WORK

There are several avenues for improving the robustness and flexibility of the proposed pipeline. Although the current method enables intuitive control over blur strength,

aperture shape, and depth-dependent transitions, its performance is still tightly coupled to the quality of the estimated depth map. Future work could explore strategies for mitigating depth-related artifacts, such as incorporating lightweight depth refinement, uncertainty-aware bucketization, or adaptive feathering around regions with unreliable geometry. Another direction involves improving handling of challenging materials and structures where abrupt mask transitions or depth discontinuities can lead to unnatural blur boundaries.

Beyond algorithmic refinements, the system could also benefit from expanded user-facing controls. For example, providing interactive feedback, adjustable falloff profiles, or presets tailored to common photographic styles may improve practical usability. Additionally, integrating optional learned components could offer a hybrid solution that preserves the transparency and controllability of the pipeline while reducing common failure cases. Overall, future work should continue to balance accuracy, stability, and user-adjustability, ensuring the method remains accessible while improving its reliability across a broader range of scenes.

7 CONCLUSION

We presented a fully hand-crafted pipeline for generating controllable shallow-focus effects using dense depth maps, depth bucketization, and depth-aware splatting. Unlike end-to-end learned approaches, the method emphasizes interpretability and artistic flexibility, allowing users to directly influence aperture shape, blur falloff, and focal-plane selection. Despite its qualitative focus and lack of training data, the system produces visually coherent bokeh across a wide range of scenes and achieves competitive PSNR performance relative to recent hybrid and fully learned methods.

REFERENCES

- [1] J. Peng, Z. Cao, X. Luo, H. Lu, K. Xian, and J. Zhang, "Bokehme: When neural rendering meets classical rendering," 2022. [Online]. Available: <https://arxiv.org/abs/2206.12614>
- [2] T. Seizinger, F.-A. Vasluianu, M. V. Conde, Z. Wu, and R. Timofte, "Bokehlicious: Photorealistic bokeh rendering with controllable apertures," 2025. [Online]. Available: <https://arxiv.org/abs/2503.16067>

- [3] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao, "Depth anything v2," 2024. [Online]. Available: <https://arxiv.org/abs/2406.09414>
- [4] A. Ignatov, J. Patel, and R. Timofte, "Rendering natural camera bokeh effect with deep learning," 2020. [Online]. Available: <https://arxiv.org/abs/2006.05698>