AutoColor: learned light power control for multi-color holograms

Yicheng Zhan[a], Koray Kavaklı[a,b], Hakan Urey[b], Qi Sun[c], and Kaan Akşit[a,*]

[a]University College London, Department of Computer Science, London, United Kingdom
[b]Koç University, Department of Electrical and Electronics Engineering, Istanbul, Türkiye
[c]New York University, Department of Computer Science and Engineering, New York, United States of America

ABSTRACT

Multi-color holograms rely on simultaneous illumination from multiple light sources. These multi-color holograms could utilize light sources better than conventional single-color holograms and can improve the dynamic range of holographic displays. In this letter, we introduce *AutoColor*, the first learned method for estimating the optimal light source powers required for illuminating multi-color holograms. For this purpose, we establish the first multi-color hologram dataset using synthetic images and their depth information. We generate these synthetic images using a trending pipeline combining generative, large language, and monocular depth estimation models. Finally, we train our learned model using our dataset and experimentally demonstrate that *AutoColor* significantly decreases the number of steps required to optimize multi-color holograms from $>$ 1000 to 70 iteration steps without compromising image quality.

Keywords: Computer Generated Holography, Machine Learning, Computer Graphics

1. INTRODUCTION

Computer-Generated Holography (CGH) is an emerging technology for next-generation displays, including virtual reality headsets, augmented reality glasses,[1] and 3D displays.[2] Through CGH, holographic displays promise to reproduce realistic images by reconstructing accurate light fields[3,4] or perceptually accurate representations.[5,6] A standard holographic display comprises a phase-only Spatial Light Modulator (SLM) and multiple light sources that helps generate a full-color image. Typically, the phase-only SLM plays a single-color hologram for each color channel time-sequentially. Meanwhile, each corresponding light source for each color channel lits a presented single-color hologram. This way, Human Visual System (HVS) integrates each color from these holograms, and users can observe full-color Three-Dimensional (3D) scenes from holographic displays.

Recently, Chen et al.[7] optimized light source powers in single-color holograms using a camera-in-the-loop approach to accurately represent color and brightness levels. Assuming light sources operate at their peak intensities, standard single-color holograms are limited in their dynamic range, a set of brightness levels they can represent. This dynamic range issue becomes apparent as each light source roughly operates one-third of the time when representing a full-color image. A recent study proposes multi-color holograms[8] for overcoming this issue. Their work co-optimizes multi-color holograms with their corresponding powers by each light source. While their study improves dynamic range and brightness up to $\times$1.8 than single-color holograms, the co-optimization process requires many iterations (e.g. ¿1000 steps). Thus, multi-color hologram optimization takes minutes, remarkably slower than a few seconds of single-color computations.

This letter proposes the first learned method, *AutoColor*, to estimate the optimal light source powers for multi-color holograms. *AutoColor* reduces the multi-color optimization in the previous study[8] from many minutes to a couple of ten seconds. Firstly, *AutoColor* needs a dataset to adequately train a light source powers estimation network. Today, high-quality hologram datasets[9,10] exist but are strictly for single-color holograms. The multi-color hologram dataset is not readily available in the literature. We create the first multi-color hologram dataset

Further author information: (Send correspondence to Kaan Akşit)
Kaan Akşit: E-mail: k.aksit@ucl.ac.uk

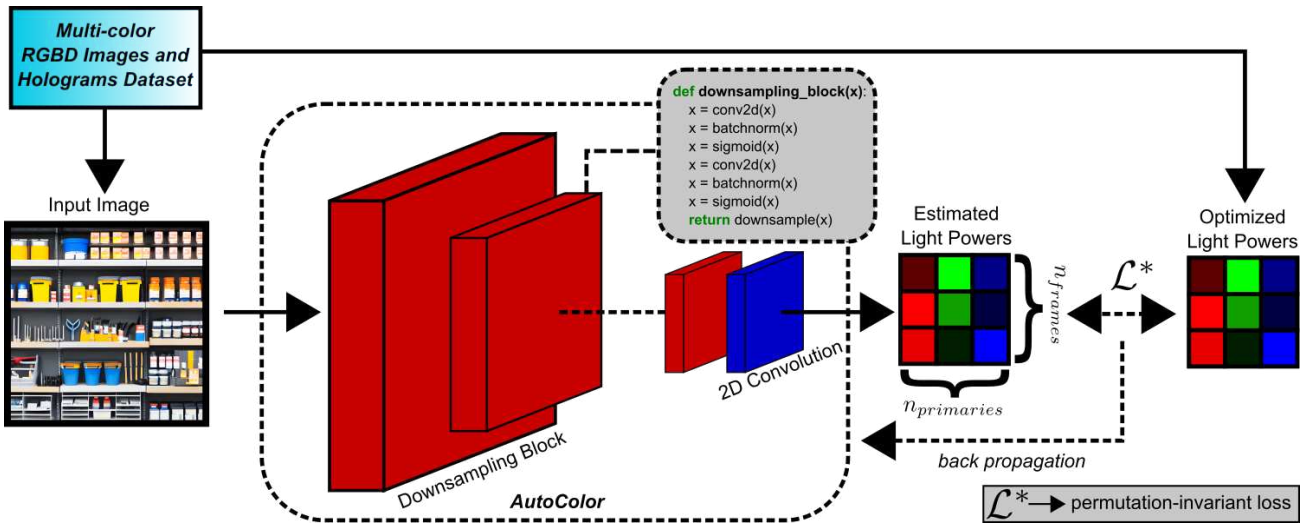


Figure 1. *AutoColor* light power estimation network structure. *AutoColor* learns to estimate powers for each light source to illuminate multi-color holograms using our multi-color hologram dataset and a permutation-invariant loss tailored for multi-color holograms.

using synthetic but photo-realistic images and their depth information. Secondly, we develop a Convolutional Neural Network (CNN) architecture that consists of downsampling and convolution operations to estimate light source powers from an input image. We train this CNN using our multi-color hologram dataset. Finally, we show that our CNN significantly accelerates multi-color optimizations. We also experimentally verified our findings in a holographic display prototype.

2. RELATED WORK

The field of CGH has witnessed significant advancements by incorporating neural networks, addressing the challenge of balancing speed and image quality. Liang et al.[9,10] proposed a CGH deep learning network and a hybrid supervised+unsupervised training approach, which enables the synthesis of high-quality 3D phase-only holograms at an interactive rate.

Similarly, Choi et al.[11] developed a comprehensive CGH framework with a camera-in-the-loop optimization strategy. This innovative approach has led to the real-time generation of full-color, high-quality holographic images at 1080p resolution, a breakthrough in holographic imaging. Furthermore, the work by Zhu et al.[12] leverages Fourier basis functions in learning-based CGH, demonstrating impressive performance in model generalization and quality of reconstructions.

3. METHODOLOGY

The schematic diagram of our learned method, *AutoColor*, is shown in Fig. 1. To achieve *AutoColor*, we *first generate a dataset* of images with their depths, laser powers, and holograms. We leverage a Large Language Model (LLM) GPT-4[13] via its online interface ChatGPT to guide the generation diversity.

We use those LLM-generated prompts to develop a large dataset of images locally using text-to-image generation models, Stable Diffusion[14] (see GitHub:stability-ai/stablediffusion). Using the publicly available weights (v2-1_512-ema-pruned.ckpt) for the text-to-image generation model and using 12 GB memory, we generate 8865 images with 512×512 resolutions. We then upsample these 8865 images to 2048×2048 resolutions using a Generative Adversarial Network (GAN) based super-resolution network[15] (see GitHub:xinntao/Real-ESRGAN). The upsampling process runs locally on an RTX 3090, consuming 4.7 GB memory using the publicly available weights (RealESRGAN_x4plus.pth). To estimate the depth information for all the generated images at 2048×2048 resolution, we rely on a monocular depth information network[16,17] and their publicly available

weights (dpt_hybrid-midas-501f0c75.pt), consuming 35 GB memory on an NVIDIA A100 on a cloud GPU cluster. We also downscale 8865 RGBD images with 1024×1024 resolutions to reduce time and memory consumption during the hologram generation step. Locally, we optimize multi-color holograms and their light source powers using *Multicolor* optimization pipeline[8] (see GitHub:complight/multicolor). To efficiently manage our workload, we found it imperative to employ multiple GPUs, We convert all the RGB images to multi-color holograms at $\times1.8$ brightness, 1024×1024 resolution, and $8\mu m$ pixel pitch. We target three depth layers and 500 steps for each multi-color hologram, with the learning rate starting from 0.025 and decaying to 0.005 (highly aggressive to coarse learning rates). The multi-color holograms are generated for three target planes, which are on -0.5 cm, 0 cm, and 0.5 cm with respect to our SLM (hologram plane). Our entire dataset generation consumes about ten days of computation using multiple GPUs.

Conventional single-color holograms rely on a field-sequential color method and use one single monochromatic light source at a time. HVS fuses these single-color images into a full-color as these holograms are displayed at rates well above Critical Flicker Fusion (CFF). Assuming that a holographic display has three monochromatic light sources, each of them is optimized by solving the following equation,

$$\hat{u}_p \leftarrow \underset{u_p}{\operatorname{argmin}} \sum_{p=1}^{3} \mathcal{L}(|e^{iu_p} * h_p|^2, I_p), \tag{1}$$

where p denotes the index of a color primary, u_p is the SLM phase, $\hat{u}_p$ is the optimized single-color hologram, h_p is the wavelength-dependent light transport kernel,[18,19] I_p is the target image intensity, $*$ denotes the convolution operation, and $\mathcal{L}$ denotes any valid loss function that measures the difference between the reconstruction and target. On the other hand, multi-color holograms use multiple monochromatic light sources simultaneously. Assuming the total number of subframes to $T = 3$ like in the conventional single-color holograms, multi-color hologram generation could be formulated as

$$\hat{u}_t, \hat{l}_{(p,t)} \leftarrow \underset{u_t, l_{(p,t)}}{\operatorname{argmin}} \underbrace{\sum_{p=1}^{3} \left\| \left(\sum_{t=1}^{T} \left| \sqrt{l_{(p,t)}} e^{i \frac{\lambda_p}{\lambda_{p_{\text{anchor}}}} u_t} * h_p \right|^2 \right) - s I_p \right\|_2^2}_{L_{\text{image}}}, \tag{2}$$

where $l_{(t,p)}$ represents the light source intensity for the p-th primary at the t-th subframe, λ_p denotes the wavelength of the active primary, $\lambda_{p_{\text{anchor}}}$ denotes the wavelength of the anchor primary, for which the nominal value of the SLM phase is calibrated against (e.g. $\lambda_{p_{\text{anchor}}} = 515\ nm$ in our hardware prototype), and s determines how bright a final image should be. For this study, we choose $s = \times1.8$, whereas single-color hologram has lower brightness values with $s = \times1.0$. Specifically, a multi-color hologram optimization seeks the optimal light source intensity, l (a 3×3 matrix for light sources and 3 subframes). Values in l are always normalized between zero (complete switching) and one (the peak brightness level of a monochromatic light source). If $s > 1$, l's rows representing subframes will sum up to a value > 1 to match the demand of s (code implementation at GitHub:complight/multicolor). For single-color holograms, l is a preset value that corresponds to an identity matrix where only one monochromatic light source operates at each subframe. Certainly, it does not meet the demand of $s > 1$ effectively as the l's rows sum up to one and often yields image degradation as described in.[8] In this work, we estimate l matrix to provide a faster convergence rate in multi-color hologram optimizations.

Leveraging the dataset and the described multi-color optimizations, we develop *AutoColor*, a light source power estimation CNN using PyTorch[20] and Odak,[21] where we estimate various l from input images. *AutoColor* includes downsampling blocks followed by a final convolutional layer, as shown in Fig. 1. Each downsampling block has a cascade of 2D convolution layers with a kernel size of three and a channel size of twenty-four. Each convolutional layer in a downsampling block is followed by batch normalization and nonlinear activation function. The last layer of each block is a downsampling operation. Our CNN contains three downsampling blocks, starting from original image resolution to downsampling to 100×100, 10×10, and 3×3 in stages. Using an RTX 2080 Ti, we train our CNN for 40 epochs, starting with a learning rate of 0.002, and declining to 0.0005. Our training with an Adam solver includes all the images and optimized laser powers of our multi-color holograms from our dataset. The ground truth laser powers of the images are generated from.[8] We include regularization terms for various

considerations. (1) The rows of l provide light source power for individual frames. Their ordering, however, shall not alter the reconstructed imagery. Therefore, the predicted laser light source power is regularized for order invariance. (2) The estimated values shall be bounded with a physically plausible range $[0, 1]$. The regularization loss term is constructed as follows,

$$L = \underbrace{\underset{m\in\{k!\}}{\operatorname{argmin}} \|^m l_{est} - l_{opt}\|_2^2}_{L_{invariant}} + \underbrace{\sum |l_{est} < 0| + \sum |l_{est} > 1 - 1|}_{L_{normalize}}. \tag{3}$$

Here, $L_{invariant}$ and $L_{normalize}$ are loss functions to encourage frame-invariance and bounding to $[0, 1]$. Meanwhile, k represents frame permutations (m for each of them), l_{est} represents light source power estimation, and $laserIntensity_{opt}$ represents the corresponding optimized power from our dataset. Note that the way we choose the minimum loss among permutations is a vital component of our training, it is similar to permutation-invariant loss functions in speech separation literature.[22] After the training converges, we use the estimated light source powers from the CNN and optimize multi-color holograms using.[8] We observe that the paradigm significantly improved computational efficiency by reducing the optimization to 70 steps, compared with the 1000 steps from prior work.[8]

4. EVALUATION

As a *quantitative evalutation*, we developed a holographic display hardware prototype following.[8] Our prototype uses a LASOS MCS4 multi-wavelength laser (473, 515, and 639 nm) with controlled power levels from two ESP32-WROOM-32D modules. A pinhole aperture, Thorlabs SM1D12, was placed in front of the fiber to limit the numerical aperture of the diverging beams. Thorlabs LPVISE100-A linear polarizer placed after the pinhole aperture allows a polarization state aligned with the SLMs fast axis for light beams. These linearly polarized beams get modulated with our phase-only SLM, Holoeye Pluto-VIS (1080×1920 resolution and $8\mu m$ pixel pitch), and arrive at a 4f imaging system composed of two 50 mm focal length achromatic doublet lenses and a pinhole aperture (Thorlabs AC254-050-A and SM1D12). We capture the image reconstructions with a Ximea MC245CG-SY camera, located on an X-stage (Thorlabs PT1/M range: 0-25 mm, precision: 0.01 mm). The prototype is configured as an off-axis imaging system. A linear grating term was applied to phase holograms to generate images at the half-diffraction order location. We calculate the linearly grated phase hologram, O'_h,

$$O'_h(x, y) = \begin{cases} e^{-j(\phi(x,y)+\pi)} & \text{if } y = \text{odd} \\ e^{-j\phi(x,y)} & \text{if } y = \text{even} \end{cases} \tag{4}$$

where ϕ represents the original phase values of O_h at the x and y pixel locations. Our quantitative evaulation is summarized in Fig. 2.

5. CONCLUSION

In conclusion, we show that multi-color hologram optimizations could be achieved with significantly fewer steps. To this aim, we develop a light power estimation network, *AutoColor*, powered by our first multi-color hologram generation pipeline and dataset. The findings are further validated with an experimental analysis. We hope *AutoColor* to pave the way toward an exciting research frontier for future holograms with wide dynamic ranges at interactive rates.

In the future, we envision to advance the research on adaptability and energy efficiency. Our CNN is trained for a fixed multi-color hologram generation routine, where $s = \times 1.8$. The architecture may be further improved such that *scale* conditions the estimated light source powers. This way, the other s values could be requested from the CNN. To do so, the multi-color hologram dataset shall be extended to include varied s values. Similarly, the CNN structure estimates optimal light source powers, instead of the minimal energy consumption. In fact, power savings could be realized if a gaze tracker is introduced in the system and target colors are chosen following HVS characteristics in foveal and peripheral regions similar to.[23,24]

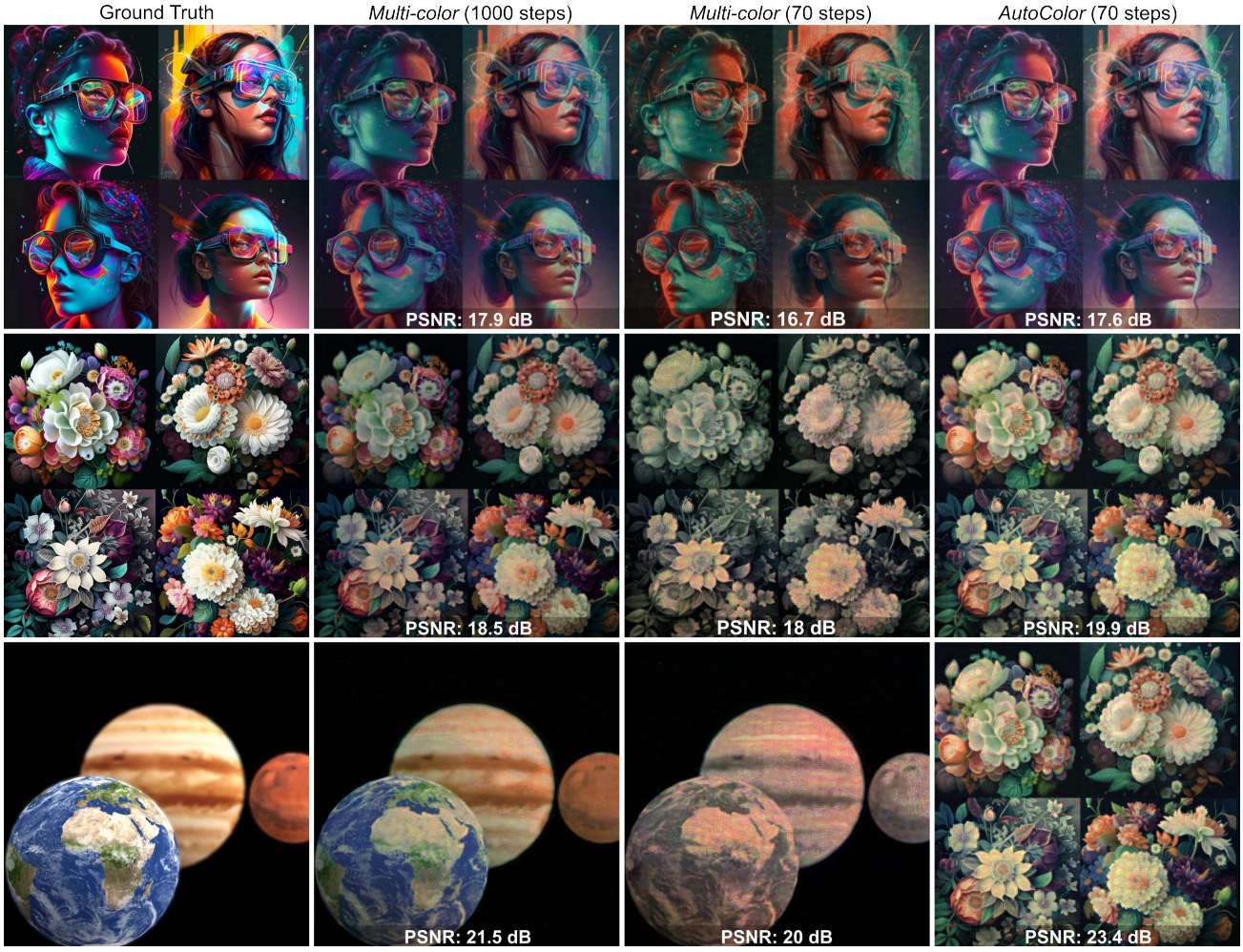


Figure 2. Photographs showing *AutoColor* generating x1.8 times brighter images in lesser steps than *Multicolor*. Experimental results show that *AutoColor* achieves high-fidelity visuals using only 70 steps, whereas *Multicolor* requires 1000 steps for similar quality and fails to produce correct color information in 70 steps. Images optimized with 70 steps using *AutoColor* provides similar quantitative image metrics (see the inset numbers) compared with the images generated with 1000 steps in *Multicolor*. AutoColor applies to both 2D images (first and second row) and 3D images (third row). (Source link: **Github:complight/image**, 80 ms exposure).

Funding. Kaan Akşit, Koray Kavaklı and Yicheng Zhan are supported by the Royal Society's RGS/R2/212229 - Research Grants 2021 Round 2 and Meta Reality Labs inclusive rendering initiative 2022. Hakan Urey is supported by the European Innovation Council's HORIZON-EIC-2021-TRANSITION-CHALLENGES program Grant Number 101057672 and Tübitak's 2247-A National Lead Researchers Program, Project Number 120C145. Qi Sun is partially supported by the National Science Foundation (NSF) research grants #2225861 and #2232817.

Acknowledgements. The authors thank the funders and anonymous reviewers for their feedback. The authors would also like to thank Liang Shi for fruitful discussions.

Disclosures. The authors declare no conflicts of interest.

Data availability. All data needed to evaluate the conclusions in the manuscript are provided in the manuscript. Source code and dataset can be downloaded from GitHub:complight/autocolor. Additional data related to this paper may be kindly requested from the authors.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

[1] Koulieris, G. A., Akşit, K., Stengel, M., Mantiuk, R. K., Mania, K., and Richardt, C., "Near-eye display and tracking technologies for virtual and augmented reality," in [*Computer Graphics Forum*], **38**(2), 493–519, Wiley Online Library (2019).

[2] Urey, H., Chellappan, K. V., Erden, E., and Surman, P., "State of the art in stereoscopic and autostereoscopic displays," *Proceedings of the IEEE* **99**(4), 540–555 (2011).

[3] Chen, C., Kim, D., Yoo, D., Lee, B., and Lee, B., "Off-axis camera-in-the-loop optimization with noise reduction strategy for high-quality hologram generation," *Optics Letters* **47**(4), 790–793 (2022).

[4] Eybposh, M. H., Caira, N. W., Atisa, M., Chakravarthula, P., and Pégard, N. C., "Deepcgh: 3d computer-generated holography using deep learning," *Optics Express* **28**(18), 26636–26650 (2020).

[5] Walton, D. R., Kavaklı, K., Dos Anjos, R. K., Swapp, D., Weyrich, T., Urey, H., Steed, A., Ritschel, T., and Akşit, K., "Metameric varifocal holograms," in [*2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*], 746–755, IEEE (2022).

[6] Akşit, K., Kavakli, K., Walton, D., Steed, A., Urey, H., Dos Anjos, R. K., Friston, S., Weyrich, T., and Ritschel, T., "Perceptually guided computer-generated holography," in [*Advances in Display Technologies XII*], **12024**, 11–14, SPIE (2022).

[7] Chen, C., Kim, D., Nam, S.-W., Lee, J., and Lee, B., "Auto-color-balance holographic display: laser-camera-in-the-loop optimization," in [*Frontiers in Optics*], JW5B–52, Optica Publishing Group (2022).

[8] Kavaklı, K., Shi, L., Urey, H., Matusik, W., and Akşit, K., "Multi-color holograms improve brightness in holographic displays," in [*SIGGRAPH ASIA 2023 Conference Papers*], *SA '23* (Dec. 2023).

[9] Shi, L., Li, B., Kim, C., Kellnhofer, P., and Matusik, W., "Towards real-time photorealistic 3d holography with deep neural networks," *Nature* **591**(7849), 234–239 (2021).

[10] Shi, L., Li, B., and Matusik, W., "End-to-end learning of 3d phase-only holograms for holographic display," *Light: Science & Applications* **11**(1), 247 (2022).

[11] Peng, Y., Choi, S., Padmanaban, N., and Wetzstein, G., "Neural holography with camera-in-the-loop training," *ACM Trans. Graph.* **39** (nov 2020).

[12] Zhu, R., Chen, L., and Zhang, H., "Computer holography using deep neural network with fourier basis," *Opt. Lett.* **48**, 2333–2336 (May 2023).

[13] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165* (2020).

[14] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B., "High-resolution image synthesis with latent diffusion models," (2021).

[15] Wang, X., Xie, L., Dong, C., and Shan, Y., "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in [*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*], 1905–1914 (October 2021).

[16] Ranftl, R., Bochkovskiy, A., and Koltun, V., "Vision transformers for dense prediction," *ICCV* (2021).

[17] Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., and Koltun, V., "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(3) (2022).

[18] Matsushima, K. and Shimobaba, T., "Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields," *Optics express* **17**(22), 19662–19673 (2009).

[19] Kavaklı, K., Urey, H., and Akşit, K., "Learned holographic light transport," *Applied Optics* **61**(5), B50–B55 (2022).

[20] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., "Automatic differentiation in pytorch," (2017).

[21] Kavakli, K., Walton, D. R., Antipa, N., Mantiuk, R., Lanman, D., and Akşit, K., "Optimizing vision and visuals: lectures on cameras, displays and perception," in [*ACM SIGGRAPH 2022 Courses*], 1–66 (2022).

[22] Yu, D., Kolbæk, M., Tan, Z.-H., and Jensen, J., "Permutation invariant training of deep models for speaker-independent multi-talker speech separation," in [*2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 241–245, IEEE (2017).

[23] Duinkharjav, B., Chen, K., Tyagi, A., He, J., Zhu, Y., and Sun, Q., "Color-perception-guided display power reduction for virtual reality," *ACM Transactions on Graphics (TOG)* **41**(6), 1–16 (2022).

[24] Chakravarthula, P., Zhang, Z., Tursun, O., Didyk, P., Sun, Q., and Fuchs, H., "Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays," *IEEE Transactions on Visualization and Computer Graphics* **27**(11), 4194–4203 (2021).