pCON: Polarimetric Coordinate Networks for Neural Scene Representations

Henry Peters*,1, Yunhao Ba*,2, Achuta Kadambi[1,2]
[1]Computer Science Department, University of California, Los Angeles (UCLA)
[2]Electrical and Computer Engineering Department, UCLA
hpeters@ucla.edu, yhba@ucla.edu, achuta@ee.ucla.edu

Abstract

Neural scene representations have achieved great success in parameterizing and reconstructing images, but current state of the art models are not optimized with the preservation of physical quantities in mind. While current architectures can reconstruct color images correctly, they create artifacts when trying to fit maps of polar quantities. We propose polarimetric coordinate networks (pCON), a new model architecture for neural scene representations aimed at preserving polarimetric information while accurately parameterizing the scene. Our model removes artifacts created by current coordinate network architectures when reconstructing three polarimetric quantities of interest. All code and data can be found at this link: https://visual.ee.ucla.edu/pcon.htm.

1. Introduction

Neural scene representations are a popular and useful tool in many computer vision tasks, but these models are optimized to preserve visual content, not physical information. Current state-of-the-art models create artifacts due to the presence of a large range of spatial frequencies when reconstructing polarimetric data. Many tasks in polarimetric imaging rely on precise measurements, and thus even small artifacts are a hindrance for downstream tasks that would like to leverage neural reconstructions of polarization images. In this work we present pCON, a new architecture for neural scene representations. pCON leverages images' singular value decompositions to effectively allocate network capacity to learning the more difficult spatial frequencies at each pixel. Our model reconstructs polarimetric images without the artifacts introduced by state-of-the-art models.

The polarization of light passing through a scene contains a wealth of information, and while current neural representations can represent single images accurately, but they produce noticeable visual artifacts when trying to represent multiple polarimetric quantities concurrently.

We propose a new architecture for neural scene representations that can effectively reconstruct polarimetric images without artifacts. Our model reconstructs color images accurately while also ensuring the quality of three important polarimetric quantities, the degree (ρ) and angle (ϕ) of linear polarization (DoLP and AoLP), and the unpolarized intensity I_{un}. This information is generally captured using images of a scene taken through linear polarizing filters at four different angles. Instead of learning a representation of these images, our model operates directly on the DoLP, AoLP and unpolarized intensity maps. When learning to fit these images, current coordinate network architectures produce artifacts in the predicted DoLP and unpolarized intensity maps. To alleviate this issue, we take inspiration from traditional image compression techniques and fit images using their singular value decompositions. Images can be compressed by reconstructing them using only a subset of their singular values [28]. By utilizing different, non-overlapping sets of singular values to reconstruct an image, the original image can be recovered by summing the individual reconstructions together. Our model is supervised in a coarse-to-fine manner, which helps the model to represent both the low and and high frequency details present in maps of polarimetric quantities without introducint noise or tiling artifacts. A demonstration of the efficacy our model can be seen in Fig. 1 and Table 1. Furthermore, our model is capable of representing images at varying levels of detail, creating a tradeoff between performance and model size without retraining.

1.1. Contributions

To summarize, the contributions of our work include:

- a coordinate network architecture for neural scene representations of polarimetric images;
- a training strategy for our network which learns a series of representations using different sets of singular values, allowing for a trade-off between performance and model size without retraining;

*Equal contribution.

Figure 1. **Our model reconstructs the training scene more accurately than other architectures.** Our model does not have the noise pattern present in reconstructions from SIREN [52] or a ReLU MLP with positional encoding [38], nor does it show tiling artifacts as in ACORN's [34] prediction.

- results demonstrating that our model reconstructs maps of polarimetric quantities without the artifacts created by current state-of-the-art approaches.

2. Related work

2.1. Neural scene representations

The aim of neural scene representations is to parameterize a two or three dimensional scene in the weights of a neural network in order to accomplish some other task related to the scene. Most papers fall into one of three categories. Explicit representations model the scene directly, which allows them to quickly accomplish tasks such as scene reconstruction [9, 31], novel view synthesis [4, 17, 22, 36, 37, 41, 50, 53, 56] or relighting [60]. However, since the scene is modelled explicitly, these representations require more memory than the alternatives.

Implicit representations do not model the scene directly, but instead use an MLP to map from a coordinate in either 2D or 3D space to some desired output value. This value could be the observed radiance or pixel intensity [11, 16, 38, 44, 46, 54], occupancy of a pixel or voxel [35, 45], a quantity related to shape [5, 10, 13, 18, 19, 21, 23, 26, 30, 44, 47, 54, 58], or any other quantity of interest. The final category of neural scene representations is a hybrid of the first two. The only work that fits directly into this category is ACORN [34], which accomplishes state-of-the-art performance on image and volume fitting by combining a coordinate network with an explicit grid or voxel representation. Similar to ACORN, other works divide the scene into local regions and learn each of these regions implicitly [10, 26, 49].

Model	Noise Pattern	Tiling Artifacts	Resizing Artifacts
ACORN [34]	Medium	High	Not Supported
ReLU w/P.E. [38]	Medium	None	Yes
SIREN [52]	High	None	Yes
Proposed	Minimal	None	Minimal

Table 1. **Our model shows fewer artifacts than current state-of-the-art architectures.** Since ACORN divides an image into a discrete grid, in order to query an image at a different resolution it is necessary to also reform the grid. The grid is created online during training, so it is not feasible to query a model at a different resolution without retraining.

To our knowledge, this work is the first to highlight the artifacts caused by existing neural scene representation architectures when fitting polarimetric data. While we are one of the first works to examine polarization and neural scene representations in the same context, we would like to acknolwedge that PANDORA [14], a concurrent work, also utilizes polarization and neural scene representations. However, they focus on radiance decomposition rather than 2D reconstruction.

2.2. Polarization vision

Polarization is useful in a variety of computer vision tasks. It can be used to estimate surface normals [1, 2,

7, 15, 24, 25, 33, 39, 43] or refine depth maps to represent incredibly fine details [27]. It can be used in radiometric calibration [55], dynamic interferometry [32], facial reconstruction [6] and separation of diffuse and specular reflection [39, 42]. It also can be used to remove the effects of scattering media like haze [51, 57, 61] and water [57], to augment the performance of computer vision tasks in the presence of transparent objects [12, 29, 40], or even to assist in imaging objects in space [20]. Traditionally, polarimetric data is captured by rotating a linear polarizing filter in front of a camera [3, 59], but recent advances in machine vision have produced cameras that can capture multiple polar images in a single shot.

Our work uses a neural network to accurately parameterize polarimetric information captured from a scene. This allows for easier storage and transport of polarimetric data and facilitates its use in other deep learning based tasks.

3. Method

3.1. Polarization physics

Polarized light can be modelled as a sine wave, and can thus be parameterized by three quantities. The degree of linear polarization (DoLP) is a quantity between 0 and 1 that represents how much of the total intensity of the wave is polarized and unpolarized. Completely polarized light will have a DoLP of 1, and completely unpolarized light will have a DoLP of 0. The angle of linear polarization (AoLP) corresponds to the orientation of the plane in which the wave is oscillating. The AoLP takes values from 0 to π radians. The final quantity of interest is the unpolarized intensity, I_{un}, of the wave, which corresponds to its amplitude. With these three quantities, it is possible to render a scene as viewed through a linear polarization filter at any angle using the following equation:

$$I(\phi_{pol}) = I_{un}(1 + \rho cos(2(\phi - \phi_{pol}))), \tag{1}$$

where I_{un} denotes unpolarized intensity, ρ denotes DoLP, ϕ denotes AoLP and ϕ_{pol} denotes the desired filter angle at each pixel. This equation allows us to render images under any number of filter angles by saving only three quantities per pixel. In this paper we leverage the above equation to learn a representation for just these quantities, rather than the four original images.

The DoLP (ρ) and AoLP (ϕ) have uses beyond just rendering images. In the shape from polarization problem, these quantities are used to calculate the zenith and azimuth angles, respectively, of per-pixel surface normals. This relationship has been studied in previous work [1, 7]. Specifically, the azimuth angle, θ_a, of a surface normal can be calculated from the following relationship:

$$\phi = \begin{cases} \theta_a, \text{when diffuse reflection dominates} \\ \theta_a - \frac{\pi}{2}, \text{when specular reflection dominates} \end{cases}. \tag{2}$$

DoLP, ρ, is related to the zenith angle, θ_z, in terms of the refractive index, n, of a surface. When diffuse reflection is dominant, the relationship can be written as:

$$\rho = \frac{(n-\frac{1}{n})^2 \sin^2(\theta_z)}{2+2n^2-(n-\frac{1}{n})^2 \sin^2(\theta_z)+4\cos(\theta_z)\sqrt{n^2-\sin^2(\theta_z)}}. \tag{3}$$

When specular reflection dominates, the relationship is different:

$$\rho = \frac{2\sin^2(\theta_z)\cos(\theta_z)\sqrt{n^2-\sin^2(\theta_z)}}{n^2-\sin^2(\theta_z)-n^2\sin^2(\theta_z)+2\sin^4(\theta_z)}. \tag{4}$$

ρ, ϕ and I_{un} can be calculated directly from a vector known as the Stokes vector at each pixel. This vector has four elements. The first three elements deal with the linear polarization of light, and the final one represents the circular polarization of the wave. In this paper we will focus on linear polarization. To measure the Stokes vector of a scene, at least three images are needed, taken through linear polarizing filters at 0, 45 and 90 degrees. Since the camera used in our setup also captures an image with a filter at 135 degrees, we use four images in our calculations of the Stokes vectors for robustness to noise.

3.2. Learning from coarse to fine

Current coordinate network architectures produce artifacts when fitting polarimetric images. SIREN [52] and similar architectures treat every coordinate equally when training, and they produce noise patterns in the resulting images when the spatial frequencies present in the training data differ widely (eg. the maximum magnitude frequency differs by an order of magnitude). In the polarimetric images we obtained, we found the maximum frequency magnitude of some AoLP maps was around 10^7, while the maximum magnitude for the intensity image was only around 10^6. ACORN [34] does not treat each coordinate in the same way, but its dynamic tiling strategy looks for regions of low variance in order to create larger blocks. This is difficult to do when attempting to fit multiple images containing varying frequencies. The resulting reconstructions end up looking blocky, and fine detail is lost in the process. Our method removes these artifacts by learning image representations using their singular value decompositions. One idea to help in reconstructing high frequency details could be to use an image's Fourier decomposition. We found that in practice the SVD works better for our use case. This is due to the propagation of errors during the forward and inverse

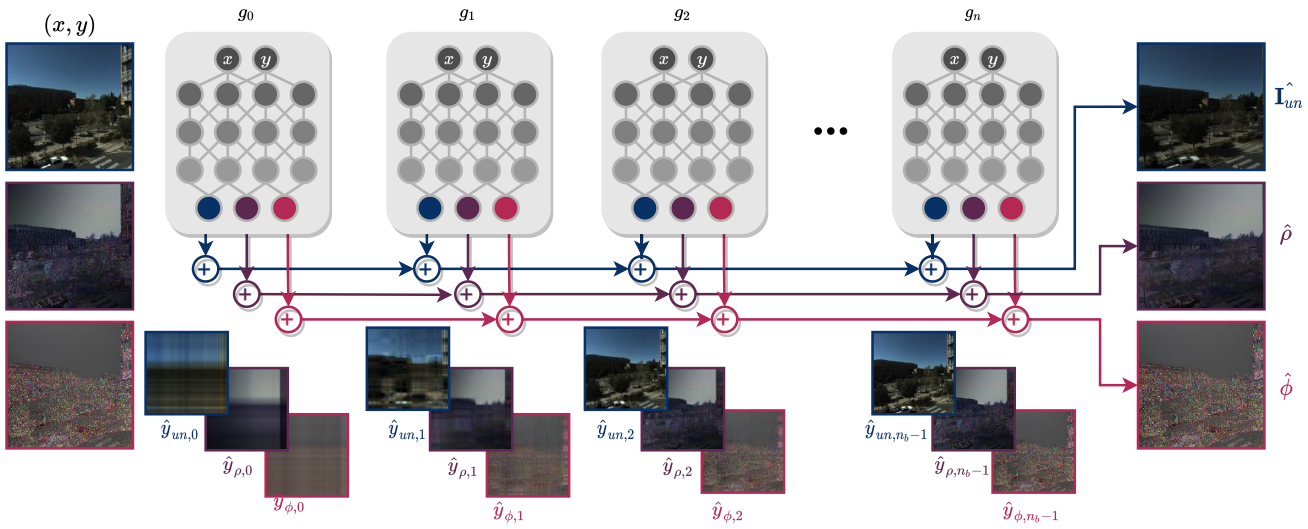

Figure 2. **pCON learns to fit an image by learning a series of reconstructions with different singular values**. The model is organized into a series of n_b parallel MLPs (denoted here as g_i) with sine activations. A 2D coordinate vector representing a point on an image is passed through all bands separately (g_0 to g_n). To supervise the training of each band, we reconstruct the full image maps of each quantity, and then calculate the MSE between the model prediction, $\hat{y}_i$ and their respective ground truth values, y_i, at the input coordinate. The final output is the sum of all the intermediate reconstructions, which yields a set of images similar to the training data.

Fourier transforms. The SVD does not require shifting between the spatial and frequency domains, which allows errors to propagate less than if we were supervising on Fourier frequencies. The singular value decomposition of an $m \times n$ matrix $\mathbf{A}$ is a set of matrices $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ and $\mathbf{V}^\top \in \mathbb{R}^{n \times n}$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. This matrix product can be further decomposed:

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \sum_{i}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top \tag{5}$$
$$= \sum_{i=0}^{a_1} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top + \sum_{i=a_1}^{a_2} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top + ... + \sum_{i=a_n}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

where r is the rank of $\mathbf{A}$, $\mathbf{u}_i$ is the i-th column of $\mathbf{U}$, $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$, and σ_i is the i-th singular value. In the case of an image, this means that it is possible to calculate different pieces of the decomposition individually, and then sum them to obtain the original image. We leverage this property of the SVD in our model architecture. Using just the largest singular values to reconstruct an image yields a result containing only the low frequency details of the original [28]. As more singular values are used in the reconstruction, higher frequency details are captured. A single coordinate may have features in many reconstructions, and others may have features in only a few. Our network learns a series of reconstructions in parallel, which effectively allocates more model capacity to coordinates which have details at numerous frequencies. Since we are not dividing the image into a grid like ACORN, our reconstructions do not suffer from tiling artifacts, and they also do not exhibit the obvious noise pattern present in reconstructions from SIREN or ReLU MLPs.

3.3. Network design

Our network design takes inspiration from SIREN [52]. The original SIREN architecture was similar to an ordinary MLP, except that it used the sine activation function. Our network is divided into a series of n_b fully-connected blocks which map from a 2D input image coordinate to the AoLP (ϕ), DoLP (ρ) and unpolarized intensity I_{un} at that pixel. We call each of these MLPs a *band* of the network, and we will notate them as g_i for $i \in 0, 1, ..., n_b - 1$. To fit an image, we first take the singular value decomposition of the map of each polar quantity:

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{U}_\phi \mathbf{\Sigma}_\phi \mathbf{V}_\phi^\top, \\ \boldsymbol{\rho} &= \mathbf{U}_\rho \mathbf{\Sigma}_\rho \mathbf{V}_\rho^\top, \\ \mathbf{I}_{un} &= \mathbf{U}_{\text{un}} \mathbf{\Sigma}_{\text{un}} \mathbf{V}_{\text{un}}^\top. \end{aligned} \tag{6}$$

$\mathbf{\Phi}$, $\boldsymbol{\rho}$ and $\mathbf{I}_{un}$ represent the full image maps of AoLP (ϕ), DoLP (ρ) and I_{un}, respectively. The above equations are obtained by interpreting these maps as matrices and then using Eq. (5). We now define a series of n_b thresholds for $\mathbf{\Phi}$, $\boldsymbol{\rho}$ and $\mathbf{I}_{un}$ as $t_{\phi,i}$, $t_{\rho,i}$ and $t_{\text{un},i}$, respectively. These thresholds dictate which singular values will be used to supervise each band of the network. We also define the ground truth intermediate reconstructions of each quantity using a subset of singular values as $y_{\phi,i}$, $y_{\rho,i}$ and $y_{\text{un},i}$. We denote their

corresponding predictions as $\hat{y}_{\phi,i}$, $\hat{y}_{\rho,i}$ and $\hat{y}_{\text{un},i}$. We can use Eq. (5) to decompose each of the SVDs from Eq. (6) into a set of sums. For example, we can write $\mathbf{\Phi}$ as follows:

$$y_{\phi,i} = \sum_{j=t_{\phi,i-1}}^{t_{\phi,i}} \sigma_{\phi,j} \mathbf{u}_{\phi,j} \mathbf{v}_{\phi,j}^{\top}. \tag{7}$$

The reconstructions for the other quantities can be written with their respective SVDs and thresholds similar to Eq. (7).

Each band learns a single reconstruction for these quantities at each pixel.

$$g_i(x, y) = \hat{y}_i = (\hat{y}_{\phi,i}, \hat{y}_{\rho,i}, \hat{y}_{\text{un},i}). \tag{8}$$

Here, x and y constitute the 2D pixel coordinate vector that serves as the input to the network. This coordinate is passed through each band of the network to compute all $\hat{y}_i$, and then the fully reconstructed image is calculated as $\sum_i \hat{y}_i$. See Fig. 2 for a visualization of this entire process.

3.4. Loss functions

Our network outputs a set of n_b images. For each band, we compute the MSE between the cumulative sum of all outputs up to, and including, the current band. We define multiplicative factors for the three polar quantities as λ_ϕ, λ_ρ and λ_{un}. We also define factors for each band as $\lambda_{b,i}$ The loss of the network can be calculated as follows, where L is the loss function and x is the data point for which the loss is being calculated:

$$\begin{aligned} L(x) = \sum_i \lambda_{b,i} \sum_{j=0}^{i} \lambda_\phi (\hat{y}_{\phi,j} - y_{\phi,j})^2 \\ + \lambda_\rho (\hat{y}_{\rho,j} - y_{\rho,j})^2 + \lambda_{\text{un}} (\hat{y}_{\text{un},j} - y_{\text{un},j})^2. \end{aligned} \tag{9}$$

3.5. Implementation details

3.5.1 Data

We collected all of our own data using a Flir Blackfly S RGB polarization camera. From this camera's images, it is possible to calculate the desired polarimetric quantites using the physics discussed in Sec. 3.1. We release two datasets with this paper. The first contains the six scenes used to create figures in this paper. The second set contains twenty four additional scenes for use in validating our approach. The captured scenes represent a diverse set of polarization effects. The DoLP and AoLP values span the entire ranges (zero to one for DoLP and zero to pi for AoLP) of possible values. We capture interesting polarization phenomena such as transparent and reflective surfaces. All released images have a resolution of 1024×1024.

3.5.2 Hyperparameters

We built all models in PyTorch [48]. We began all experiments with a learning rate of 1×10^{-5}, and then multiplied it by 0.1 at 5000 epochs. Models were trained for a total of 10000 epochs. We also set the unitless frequency parameter ω_0 of our sine activations to 90. For our best model, we used a total of 10 bands, each with 2 hidden layers and a hidden dimension of 256.

We chose the singular value thresholds of each band based on the sum of the magnitudes of singular values. Band one was given roughly 90% of the sum, then the others 99%, 99.9%, and so on. Exact values for λ_{b_i} used in all presented experiments can be found in the supplement.

For our experiments, we set $\lambda_\phi = 1.0$, $\lambda_\rho = 5.0$ and $\lambda_{\text{un}} = 5.0$.

4. Experiments

In this section, we present comparisons between our model, SIREN [52], ACORN [34] and an MLP using ReLU activations and positional encoding, as used in NeRF [38]. We changed the number of parameters and output values of the baseline architectures, since originally these models were designed to fit only a single image at a time. We also changed the frequency parameter ω_0 of the SIREN sine activations to 90 to match the parameter used in our own model. All our models were trained using the training strategy discussed in Sec. 3.5.

4.1. Validation of proposed failure case

We hypothesized the reason for the poor performance of baseline models when fitting polarimetric images was due to the presence of details at high spatial frequencies in the captured AoLP maps. To validate this hypothesis, we performed low-pass filtering on AoLP maps of a scene and then fit a model on the resulting AoLP, DoLP and $\mathbf{I}_{un}$ maps. We found a clear trend in the reconstruction quality as we filtered out higher percentages of high spatial frequencies. All models performed better when fewer high frequency details were present in the target images. This aligns with our idea that these details create difficult scenes for networks to reconstruct. For the scene in Fig. 3, the AoLP reconstruction SSIMs with different amounts of frequencies removed from the GT AoLP maps can be seen in Table 2.

% Highest Frequencies Removed	SIREN [52]	ACORN [34]	ReLU P.E. [38]
0%	0.60	0.51	0.63
75%	0.54	0.80	0.93
80.5%	0.89	0.97	0.98
93.75%	0.95	0.99	0.99

Table 2. **All baseline models reconstruct AoLP maps better when details at higher spatial frequencies are filtered out.** This trend validates our hypothesis that images with high frequency details are more difficult for a network to reconstruct.

4.2. Comparison with others

We trained both our model and the baselines to predict AoLP ($\mathbf{\Phi}$), DoLP ($\boldsymbol{\rho}$) and $\mathbf{I}_{un}$ maps directly. Quali-

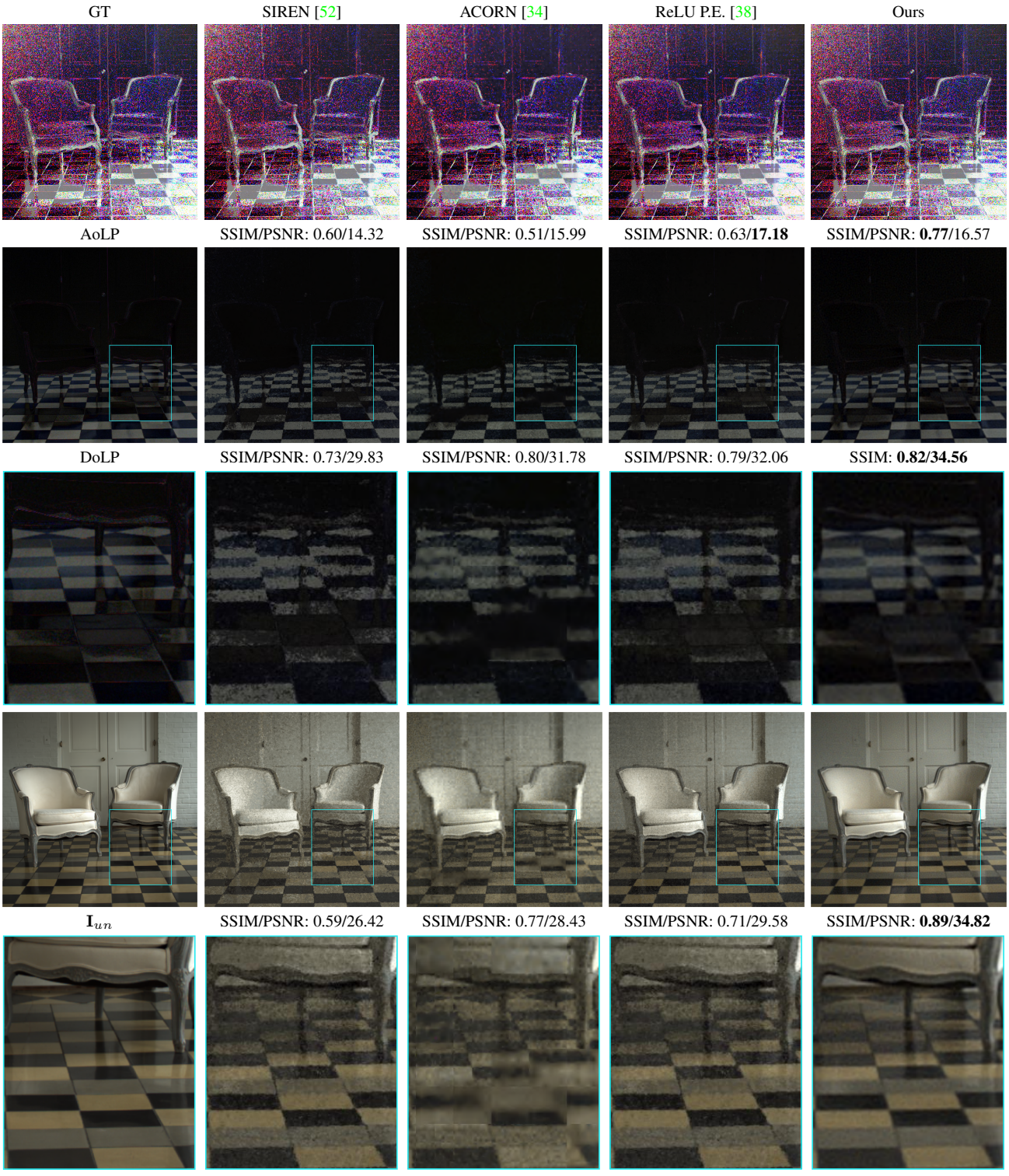

Figure 3. **Our model shows higher SSIM and fewer artifacts on predicted Φ, ρ and $\mathbf{I}_{un}$ maps.** Baseline models cause noise or tiling which is clearly visible on the checkerboard pattern on the floor, where all three quantities take large values. The artifacts are present on objects exhibiting both specular reflections, like the floor, and diffuse reflections, like the wall and doors in the background.

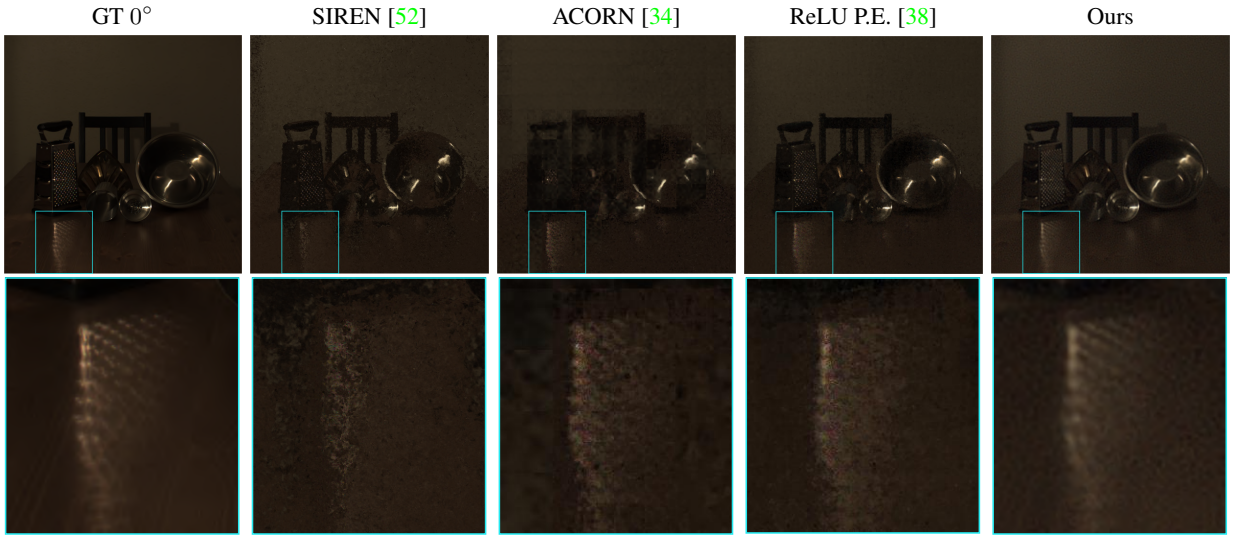

Figure 4. **Our model can more accurately reconstruct RGB images taken through different polarizing filter angles when compared to SIREN [52], ACORN [34] and a ReLU MLP [38] with positional encoding.** The images reconstructed here are the scene as viewed through a linear polarizer oriented at 0°

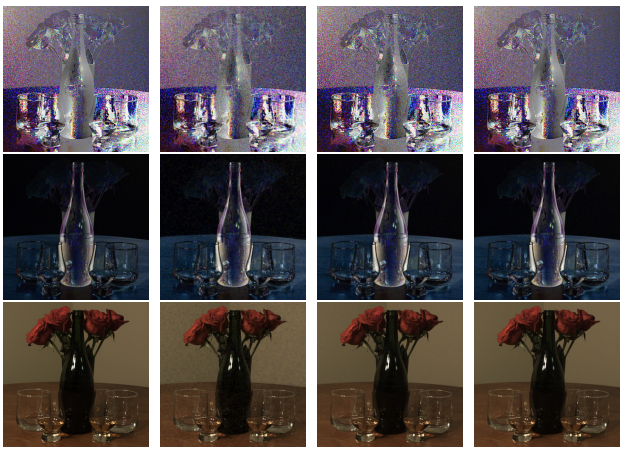

Figure 5. **As the number of bands used in the reconstruction increases, so does the quality of the image.** Even with a single band the reconstruction is visually close to the original.

Model	Φ (↑)	ρ (↑)	$\mathbf{I}_{un}$ (↑)	# Params. (↓)
Ours (1 band)	0.12/10.83	0.50/22.87	0.74/26.58	130K
Ours (2 bands)	0.32/14.66	0.64/28.40	0.91/34.74	270K
Ours (3 bands)	0.42/14.42	0.65/28.59	0.92/34.43	400K
Ours (4 bands)	0.51/16.32	0.65/28.71	0.92/34.62	530K
Ours (5 bands)	0.64/17.68	0.67/28.87	0.92/36.74	670K
Ours (Full model)	**0.79**/18.08	**0.76/31.75**	**0.92/36.00**	1.3M
SIREN [52]	0.59/15.96	0.67/28.20	0.70/28.23	660K
ACORN [34]	0.48/17.01	0.73/29.96	0.82/29.85	530K
ReLU [38] w/P.E.	0.64/**18.30**	**0.76**/30.99	0.81/32.13	660K

Table 3. **As more bands are used, the number of parameters grows along with the resulting performance (SSIM/PSNR).** The metrics shown here are averages across our whole dataset.

tative and quantitative results can be found in Fig. 3. Our model performs yeilds better PSNR and SSIM than all baselines and it also does not produce the tiling artifacts or the noise patterns present in the reconstructions created by other models.

4.3. Accuracy and model size trade-off

In order to fit an image with a smaller or larger model, current architectures require a full retraining with a different number of parameters. The structure of our model allows us to provide a tradeoff between model size and reconstruction accuracy without retraining. Each band of the model learns a representation of the image when reconstructed with a different set of singular values. If the downstream task doesn't require incredibly high accuracy, and the user would rather save and transport a smaller set of model weights, they can just save the weights from the first band of the network and reconstruct the image with only the singular values from that band, or vice versa if more accuracy is required. A visualization of reconstruction quality using different numbers of bands can be seen in Fig. 5. See Table 3 for quantitative results using different bands of our network. With a similar number of parameters to the baseline models, it achieves comparable performance to all baseline architectures. Our full model outperforms all baselines on predicting AoLP ($\mathbf{\Phi}$) and $\mathbf{I}_{un}$ maps. It is also worth noting that our full model achieves significant compression over storing raw data. The combined memory size of the AoLP, DoLP and $\mathbf{I}_{un}$ maps

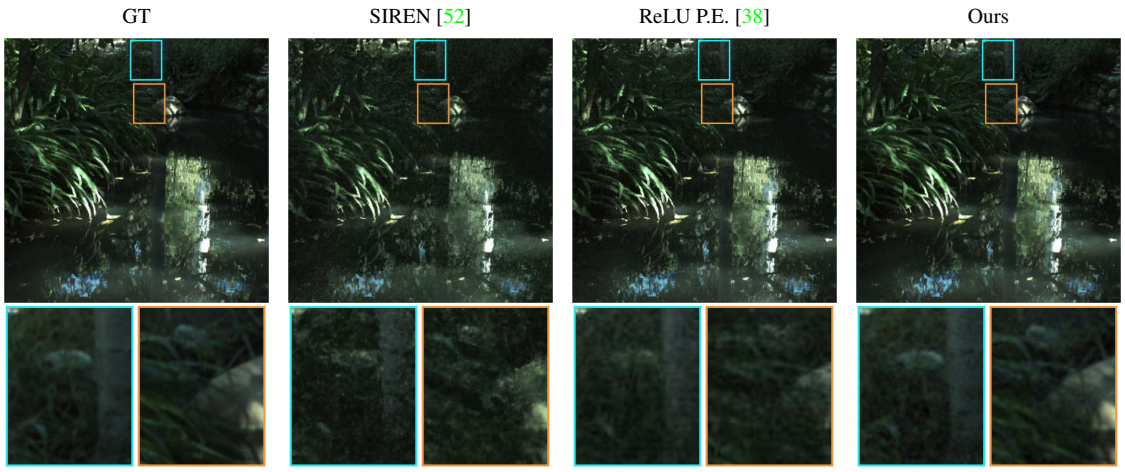

Figure 6. **Both SIREN [52] and the ReLU MLP [38] with positional encoding show artifacts when queried at a different resolution than they were trained on. Our model does not.** We trained models at a resolution of 1024×1024 and queried them at a resolution of 512×512

is 36 megabytes (MB), while the size of our full model is only 5 MB. Representing images with our model allows us to scale image size without scaling memory footprint as quickly. In this work we use small images, but the memory saved when reconstructing images at the mega or gigapixel scale would be significant.

4.4. RGB reconstruction

In addition to reconstructing the DoLP (ρ), AoLP (Φ) and $\mathbf{I}_{un}$ maps with our model, we also present results for reconstructing the original RGB images captured by the camera. For a specific polarizing filter angle, we can reconstruct the value of a pixel captured by the camera through that filter using Eq. (1). Our model removes the artifacts present in the reconstructions from all baseline comparisons and retains more detail comparatively. See Fig. 4 for a visualization of reconstructions of images taken through a linear polarizer oriented at $0°$.

4.5. Multiple resolution interpolation

We present results for fitting an image at one resolution and querying it at a second resolution. In this section we only compare to SIREN [52] and a ReLU MLP [38], as the dynamic tiling strategy of ACORN [34] does not allow us to simply query the representation at a different resolution. We train both models on the original scene at a resolution of 1024×1024 and then query them at a resolution of 512×512. Both baselines show artifacts when queried at this new resolution, while our model does not have this issue. In Fig. 6 we visualize these results on $\mathbf{I}_{un}$ maps.

5. Conclusion

In summary, we have presented an attempt at creating neural representations of polarimetric information without the artifacts introduced by current models. Compared to existing methods, our model shows an increase in image reconstruction quality on AoLP, DoLP and $\mathbf{I}_{un}$ maps, in addition to effectively removing the artifacts we were targeting. Having a compact representation of polarimetric images will facilitate future research in areas where this data is required.

While our work provides noticeable improvement over current methods, it is not perfect. To achieve state of the art performance on reconstructing AoLP maps, we need quite a few bands in our network, which makes the number of parameters quite large compared to other architectures. A valuable next step could be creating a model that could achieve the same performance as ours while cutting down on the memory footprint. Furthermore, we only demonstrated the effectiveness of this approach on 2D data, since polarization is not well studied in three dimensions. Validating our approach on 3D data would be a useful next step, once the field has developed a greater understanding of the underlying physics. We motivated our method using polarimetric data, but there are many types of data in computational imaging [8]. Our method will be valuable in representing multiple physical quantities of a scene at once whenever at least one measurement contains high frequency details or noise, and future research could extending this work by demonstrating its effectiveness on other types of data encountered in computational imaging.

Acknowledgements We thank members of the Visual Machines Group (VMG) at UCLA for feedback and support. A.K. was supported by an NSF CAREER award IIS-2046737 and Army Young Investigator Program (YIP) Award.

References

[1] G.A. Atkinson and E.R. Hancock. Recovery of surface orientation from diffuse polarization. *IEEE Transactions on Image Processing*, 15(6):1653–1664, 2006. 2, 3

[2] Gary A Atkinson. Polarisation photometric stereo. *Computer Vision and Image Understanding*, 160:158–167, 2017. 2

[3] Gary A Atkinson and Jürgen D Ernst. High-sensitivity analysis of polarization by surface reflection. *Machine Vision and Applications*, 29(7):1171–1189, 2018. 3

[4] Benjamin Attal, Selena Ling, Aaron Gokaslan, Christian Richardt, and James Tompkin. Matryodshka: Real-time 6dof video view synthesis using multi-sphere images. In *European Conference on Computer Vision*, pages 441–459. Springer, 2020. 2

[5] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2565–2574, 2020. 2

[6] Dejan Azinović, Olivier Maury, Christophe Hery, Matthias Nießner, and Justus Thies. High-res facial appearance capture from polarized smartphone images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023. 3

[7] Yunhao Ba, Alex Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep shape from polarization. In *European Conference on Computer Vision*, pages 554–571. Springer, 2020. 2, 3

[8] Ayush Bhandari, Achuta Kadambi, and Ramesh Raskar. *Computational Imaging*. The MIT Press, 2022. 8

[9] Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, Peter Hedman, Matthew Duvall, Jason Dourgarian, Jay Busch, Matt Whalen, and Paul Debevec. Immersive light field video with a layered mesh representation. *ACM Transactions on Graphics (TOG)*, 39(4):86–1, 2020. 2

[10] Rohan Chabra, Jan E Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *European Conference on Computer Vision*, pages 608–625. Springer, 2020. 2

[11] Eric R. Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. Pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5799–5809, June 2021. 2

[12] Tongbo Chen, Hendrik P. A. Lensch, Christian Fuchs, and Hans-Peter Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 3

[13] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019. 2

[14] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 538–556. Springer, 2022. 2

[15] O. Drbohlav and R. Sara. Unambiguous determination of shape from photometric stereo with unknown light sources. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 1, pages 581–586 vol.1, 2001. 2

[16] SM Ali Eslami, Danilo Jimenez Rezende, Frederic Besse, Fabio Viola, Ari S Morcos, Marta Garnelo, Avraham Ruderman, Andrei A Rusu, Ivo Danihelka, Karol Gregor, et al. Neural scene representation and rendering. *Science*, 360(6394):1204–1210, 2018. 2

[17] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. Deepview: View synthesis with learned gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2367–2376, 2019. 2

[18] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas Funkhouser. Local deep implicit functions for 3d shape. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4857–4866, 2020. 2

[19] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7154–7164, 2019. 2

[20] Ciriaco Goddi, Iván Martí-Vidal, Hugo Messias, Geoffrey C Bower, Avery E Broderick, Jason Dexter, Daniel P Marrone, Monika Moscibrodzka, Hiroshi Nagai, Juan Carlos Algaba, et al. Polarimetric properties of event horizon telescope targets from alma. *The Astrophysical Journal Letters*, 910(1):L14, 2021. 3

[21] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020. 2

[22] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (TOG)*, 37(6):1–15, 2018. 2

[23] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato's cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9993, 2019. 2

[24] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin Hancock. Shape and refractive index recovery from single-view polarisation images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1229–1236, 2010. 2

[25] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin R Hancock. Shape and refractive index from single-view spectro-polarimetric images. *International journal of computer vision*, 101(1):64–94, 2013. 2

[26] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local

implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6001–6010, 2020. 2

[27] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3370–3378, 2015. 3

[28] Samruddhi Kahu and Reena Rahate. Image compression using singular value decomposition. *International Journal of Advancements in Research & Technology*, 2(8):244–248, 2013. 1, 4

[29] Agastya Kalra, Vage Taamazyan, Supreeth Krishna Rao, Kartik Venkataraman, Ramesh Raskar, and Achuta Kadambi. Deep polarization cues for transparent object segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 3

[30] Shaohui Liu, Yinda Zhang, Songyou Peng, Boxin Shi, Marc Pollefeys, and Zhaopeng Cui. Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2019–2028, 2020. 2

[31] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*, 2019. 2

[32] Tomohiro Maeda, Achuta Kadambi, Yoav Y Schechner, and Ramesh Raskar. Dynamic heterodyne interferometry. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2018. 3

[33] Ali H. Mahmoud, Moumen T. El-Melegy, and Aly A. Farag. Direct method for shape recovery from polarization and shading. In *2012 19th IEEE International Conference on Image Processing*, pages 1769–1772, 2012. 2

[34] Julien N. P. Martel, David B. Lindell, Connor Z. Lin, Eric R. Chan, Marco Monteiro, and Gordon Wetzstein. Acorn: Adaptive coordinate networks for neural scene representation. *ACM Trans. Graph. (SIGGRAPH)*, 40(4), 2021. 2, 3, 5, 6, 7, 8

[35] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4460–4470, 2019. 2

[36] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16190–16199, 2022. 2

[37] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019. 2

[38] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020. 2, 5, 6, 7, 8

[39] Miyazaki, Tan, Hara, and Ikeuchi. Polarization-based inverse rendering from a single view. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 982–987 vol.2, 2003. 2, 3

[40] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):73–82, 2004. 3

[41] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 2

[42] Shree K Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163–186, 1997. 3

[43] Trung Ngo Thanh, Hajime Nagahara, and Rin-ichiro Taniguchi. Shape and light directions from shading and polarization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 2

[44] Thu H Nguyen-Phuoc, Christian Richardt, Long Mai, Yongliang Yang, and Niloy Mitra. Blockgan: Learning 3d object-aware scene representations from unlabelled images. *Advances in Neural Information Processing Systems*, 33:6767–6778, 2020. 2

[45] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5379–5389, 2019. 2

[46] Michael Oechsle, Lars Mescheder, Michael Niemeyer, Thilo Strauss, and Andreas Geiger. Texture fields: Learning texture representations in function space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4531–4540, 2019. 2

[47] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2019. 2

[48] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 5

[49] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, pages 523–540. Springer, 2020. 2

[50] Gernot Riegler and Vladlen Koltun. Free view synthesis. In *European Conference on Computer Vision*, pages 623–640. Springer, 2020. 2

[51] Y.Y. Schechner, S.G. Narasimhan, and S.K. Nayar. Instant dehazing of images using polarization. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001. 3

[52] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020. 2, 3, 4, 5, 6, 7, 8

[53] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019. 2

[54] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32, 2019. 2

[55] Daniel Teo, Boxin Shi, Yinqiang Zheng, and Sai-Kit Yeung. Self-calibrating polarising radiometric calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3

[56] Justus Thies, Michael Zollhöfer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 2

[57] Tali Treibitz and Yoav Y. Schechner. Active polarization descattering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):385–399, 2009. 3

[58] Zhen Wang, Shijie Zhou, Jeong Joon Park, Despoina Paschalidou, Suya You, Gordon Wetzstein, Leonidas Guibas, and Achuta Kadambi. Alto: Alternating latent topologies for implicit 3d reconstruction. *arXiv preprint arXiv:2212.04096*, 2022. 2

[59] Lawrence B Wolff. Polarization vision: a new sensory approach to image understanding. *Image and Vision computing*, 15(2):81–93, 1997. 3

[60] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics (TOG)*, 40(1):1–17, 2021. 2

[61] Chu Zhou, Minggui Teng, Yufei Han, Chao Xu, and Boxin Shi. Learning to dehaze with polarization. *Advances in Neural Information Processing Systems*, 34, 2021. 3