

Non-line-of-sight Imaging with Signal Superresolution Network

Jianyu Wang¹, Xintong Liu¹, Leping Xiao¹, Zuoqiang Shi^{1,2}, Lingyun Qiu^{1,2}, and Xing Fu¹
¹Tsinghua University
²Yanqi Lake Beijing Institute of Mathematical Sciences and Applications

Abstract

Non-line-of-sight (NLOS) imaging aims at reconstructing the location, shape, albedo, and surface normal of the hidden object around the corner with measured transient data. Due to its strong potential in various fields, it has drawn much attention in recent years. However, long exposure time is not always available for applications such as auto-driving, which hinders the practical use of NLOS imaging. Although scanning fewer points can reduce the total measurement time, it also brings the problem of imaging quality degradation. This paper proposes a general learning-based pipeline for increasing imaging quality with only a few scanning points. We tailor a neural network to learn the operator that recovers a high spatial resolution signal. Experiments on synthetic and measured data indicate that the proposed method provides faithful reconstructions of the hidden scene under both confocal and non-confocal settings. Compared with original measurements, the acquisition of our approach is 16 times faster while maintaining similar reconstruction quality. Besides, the proposed pipeline can be applied directly to existing optical systems and imaging algorithms as a plug-in-and-play module. We believe the proposed pipeline is powerful in increasing the frame rate in NLOS video imaging.

1. Introduction

Non-line-of-sight (NLOS) imaging problem usually employs a high temporal resolution optical system to recover the hidden object around the corner. As shown in Fig. 1a, photons emitted by the laser are collected by the detector after three diffuse reflections: the reflection at the visible wall, the reflection at the hidden object, and the reflection at the visible wall again. Scanning a region on the visible wall can get measured histogram data to recover the hidden scene.

Due to its potential applications in various fields, such as auto-driving, disaster relief, and remote sensing, the NLOS imaging problem has become an emerging field since it was first proposed by Kirmani *et al.* [17] in 2009.

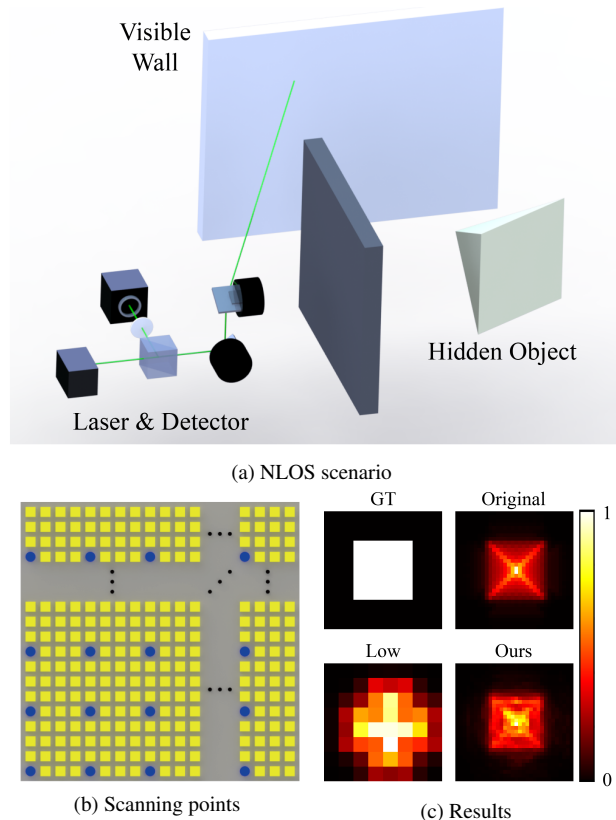


Figure 1. A typical NLOS scenario and reconstruction results. (a) An illustration of a typical NLOS scenario. (b) An illustration of the scanning points on the visible wall. Using the proposed pipeline, only the blue circle points are needed to be illuminated, and the signal at the yellow square points can be recovered with the proposed pipeline. (c) Comparisons of the reconstruction results of the pyramid with different signals. The maximum intensity projection of the albedo values along the depth direction is shown in the upper left corner (GT). The results are reconstructed by the original signal with spatial resolution 32×32 (Original), the sub-sampled signal with spatial resolution 8×8 (Low), and the signal recovered by the proposed network with spatial resolution 32×32 (Ours).

Many methods have been proposed to improve the prac-

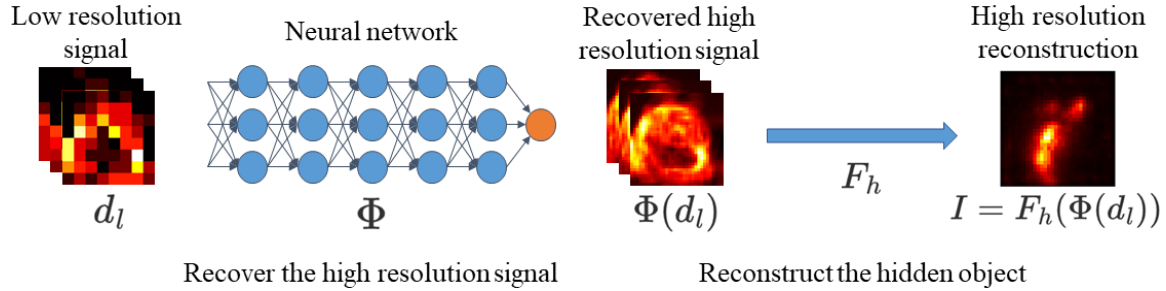


Figure 2. Flowchart of the proposed pipeline. The high resolution signal is recovered from the low resolution signal using a neural network. The hidden scene is then reconstructed using state-of-the-art imaging algorithms designed for high resolution signal.

ticability [1, 3, 13, 14, 34, 35, 39] or reconstruction quality [10, 11, 26, 41]. A back-projection method with Laplacian of Gaussian filter (LOG-BP) [18] is introduced to improve the reconstruction quality of the back-projection method [38]. In 2018, O’Toole *et al.* [31] first apply the light-cone-transform (LCT) method to image the hidden object with a confocal scanning mode. Young *et al.* [45] then propose the directional light-cone-transform (D-LCT) method, which simultaneously recovers the albedo and surface normal of the hidden object. From the perspective of wave characteristics, Lindell *et al.* [22] introduce the fast frequency-wavenumber migration (F-K) method to NLOS. Another remarkable imaging method proposed by Liu *et al.* [25] employs the phasor field model [9, 33] and provides an efficient solution for fast NLOS imaging under non-confocal scenario [24]. Further improvements of this method using single photon avalanche diode (SPAD) arrays achieve real-time imaging [21, 30]. Deep learning based methods have also been introduced [7, 8, 12, 16, 29]. Recently, Liu *et al.* [27] propose the signal-object collaborative regularization (SOCR) method, which can improve the resolution of the reconstructed object, especially in cases with high measurement noise.

However, one main obstacle in NLOS imaging is that we need a long exposure time to improve the signal-to-noise ratio (SNR) of the measurement, which is not admissible in many scenarios, such as auto-driving. Several methods have been proposed to reduce the capture time [20, 23]. For a confocal scenario, Isogawa *et al.* [15] propose a circular scanning mode. In 2021, Metzler *et al.* [28] consider an extreme case that tracks the objects with a single detection point. By introducing the downsampling operation, Ye *et al.* [44] reconstruct the object with a subset of the virtual sources. For non-confocal scenarios, most methods employ SPAD arrays to realize fast measurement. Among them, Nam *et al.* [30] achieve a reconstruction of five frames per second by remapping the data from two 16×1 SPAD arrays. Pei *et al.* [32] realize 20 frames per second reconstruction using a 32×32 SPAD array. In 2022, Yang *et al.* [42] propose a spatial multiplexing method based on the

compressed sensing strategy. Although these methods can reduce the measurement time, they might either cause deterioration in imaging quality or substantially increase expense.

In this paper, we propose a deep learning based pipeline that can reconstruct the hidden object with fine details using only a small number of scanning points. Different measurement densities are shown in Fig. 1b. In the following sections, we term the signal measured at the blue circle points as the “low resolution signal” while those measured at both circle and square points as the “high resolution signal”.

Instead of designing new algorithms for sparse measurements, we focus on recovering the dense virtual signals, which can be applied to most algorithms as a plug-in-and-play module. The proposed pipeline consists of two steps: the low resolution signal is first mapped into the high resolution signal space by a neural network; the reconstruction is then obtained from the recovered high resolution signal using existing imaging algorithms. The reconstruction results of the pyramid are shown in Fig. 1c. These results are reconstructed by the F-K method. As the comparison shows, the proposed pipeline can provide a high quality reconstruction with only 64 scanning points, which is 6.25% of the original number.

To our best knowledge, this is the first work in NLOS to study the signal recovery problem of the low resolution signal sparsely measured on a uniform grid. The proposed method can reduce the exposure time without losing the reconstruction quality. The main contributions of this work are:

- We propose a novel learning based pipeline for faithful reconstruction of the hidden scene, which only takes 6.25% of the original measuring time.
- We show that the high resolution signal can be recovered accurately with the proposed signal superresolution network, which we abbreviate as “SSN”.
- The effectiveness of the proposed method is validated on public datasets under both confocal and non-

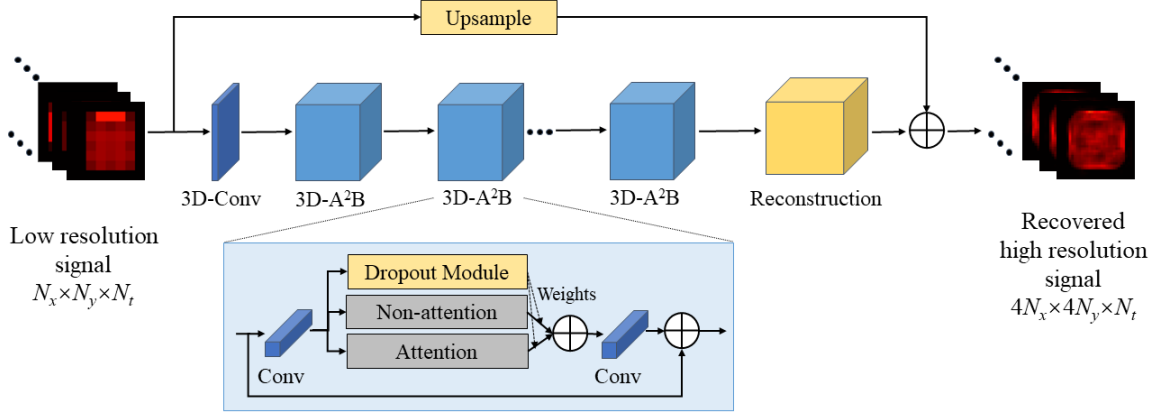


Figure 3. The architecture of the signal superresolution network (SSN) in the proposed pipeline. The network is composed of two main branches: a conventional upsampling branch and a learning branch. The final output is the sum of the results from these two branches.

confocal settings.

- We demonstrate the potential application of the proposed method in high frame rate (64 frames per second) NLOS video imaging using public datasets.

2. Reconstruction Pipeline

This work aims to develop a learning based pipeline for NLOS imaging, which can reduce the measurement time while maintaining the reconstruction quality. In this section, we formulate the signal recovery problem and introduce the proposed pipeline.

2.1. Formulation of the problem

Most existing methods that reconstruct the hidden object from low resolution signals d_l can be formulated as

$$I(x, y, z) = F_l(d_l) \quad (1)$$

in which $I(x, y, z)$ is the reconstructed object and F_l is the imaging algorithm designed for low resolution signal [15, 28, 44]. Since only a few sampling points are scanned, the total exposure time can be reduced. However, the reconstruction quality also decreases with the number of scan points.

On the contrary, the algorithms designed for high resolution signal can be expressed as

$$I(x, y, z) = F_h(d_h) \quad (2)$$

in which F_h represents the algorithm designed for high resolution signal d_h . Many works have been proposed for fast [22, 24, 31, 45] and high resolution [18, 27] reconstruction as discussed above.

In this work, we aim at designing a pipeline that can provide faithful reconstruction with only a small number of

scanning points, and has good compatibility with traditional optical systems as well as existing imaging algorithms. To solve this problem, we introduce a signal recovering operator Φ , which maps the low resolution signal d_l into the space of high resolution signal d_h . The proposed pipeline can be expressed as

$$I(x, y, z) = F_h(\Phi(d_l)) \quad (3)$$

in which $\Phi(d_l)$ is of the same size with d_h . Thus, the problem of NLOS imaging from low resolution signal converts to the problem of finding the operator Φ which satisfies $\Phi(d_l) = d_h$.

2.2. The choice of recovering operator

The choice of the operator Φ is highly dependent on the physical model of the measurement process. We adopt a simplified version of the widely used three-point transport model [37] as

$$\tau(\mathbf{x}_l, \mathbf{x}_d, t) = \int_S T(\mathbf{x}_l, \mathbf{x}_d, t, \mathbf{s}) C(\mathbf{x}_l, \mathbf{x}_d, \mathbf{s}) f(\mathbf{s}) dA(\mathbf{s}) \quad (4)$$

in which τ is the photon intensity detected with the virtual source \mathbf{x}_l , virtual detector \mathbf{x}_d and time t . $f(\mathbf{s})$ represents the albedo of the point \mathbf{s} on the hidden object S and $A(\mathbf{s})$ is the corresponding area measure. T is a function that characterizes the arrival time of the photons, which is defined as

$$T(\mathbf{x}_l, \mathbf{x}_d, t, \mathbf{s}) = \delta(\|\mathbf{x}_l - \mathbf{s}\|_2 + \|\mathbf{x}_d - \mathbf{s}\|_2 - ct) \quad (5)$$

in which δ is the Dirac function and c is the light speed. Besides, the function C characterizes the intensity attenuation

in the physical model, which is expressed as

$$C(\mathbf{x}_l, \mathbf{x}_d, \mathbf{s}) = \frac{\langle \mathbf{s} - \mathbf{x}_l, \mathbf{n}(\mathbf{x}_l) \rangle \langle \mathbf{x}_l - \mathbf{s}, \mathbf{n}(\mathbf{s}) \rangle}{\|\mathbf{x}_l - \mathbf{s}\|_2^4} \times \frac{\langle \mathbf{s} - \mathbf{x}_d, \mathbf{n}(\mathbf{x}_d) \rangle \langle \mathbf{x}_d - \mathbf{s}, \mathbf{n}(\mathbf{s}) \rangle}{\|\mathbf{x}_d - \mathbf{s}\|_2^4} \quad (6)$$

in which $\mathbf{n}(\mathbf{s})$ is the unit normal vector pointing toward the visible wall at point \mathbf{s} and $\langle \cdot, \cdot \rangle$ is the inner product in \mathbb{R}^3 .

Suppose that there are two pairs $(\mathbf{x}_{l_1}, \mathbf{x}_{d_1}), (\mathbf{x}_{l_2}, \mathbf{x}_{d_2})$ on the visible wall. For a fixed point \mathbf{s} on the hidden object, we assume that the contributions of this point to the signal at these two pairs are contained in time bin t_1 and t_2 respectively. Then from Eq. (5), we know that t_1 and t_2 are close if

$$D = \|\mathbf{x}_{l_1} - \mathbf{x}_{l_2}\|_2 + \|\mathbf{x}_{d_1} - \mathbf{x}_{d_2}\|_2 \quad (7)$$

is small. We can then regard the signal recovery problem as an interpolation problem, since the points have contributions to neighboring time bins.

However, when the spacing between adjacent detection points is large, conventional interpolation methods will fail due to the violation of the assumption. This phenomenon is very common in practical applications. To solve this problem, Nam *et al.* [30] employs two SPAD arrays to meet the assumption, but brings additional expense to the system. Another solution is brought up by Yang *et al.* [42], which combines a micro-mirror device with the compressed sensing strategy to increase detection efficiency. However, this method is sensitive to the geometry of the visible wall, which limits its application.

Instead of introducing new detection systems, we take advantage of the fast development of deep learning methods and adopt a neural network to learn the nonlinear operator Φ . Details about the network will be introduced in the next section. With the employed network, we can obtain a better approximation of the operator Φ , as well as a higher reconstruction quality.

The proposed pipeline is shown in Fig. 2. The low resolution signal d_l is first processed by the neural network Φ to recover the high resolution signal $\Phi(d_l)$. Then the reconstruction I is obtained using the existing imaging algorithm F_h . With the proposed pipeline, we can decrease the exposure time significantly while maintaining the reconstruction quality.

3. Signal Superresolution network

Based on the analysis in the last section, we have formulated the signal recovery problem as an interpolation problem and have noticed that conventional interpolation methods may fail when the spacing between adjacent detection pairs gets larger. Besides, the contribution of one point on

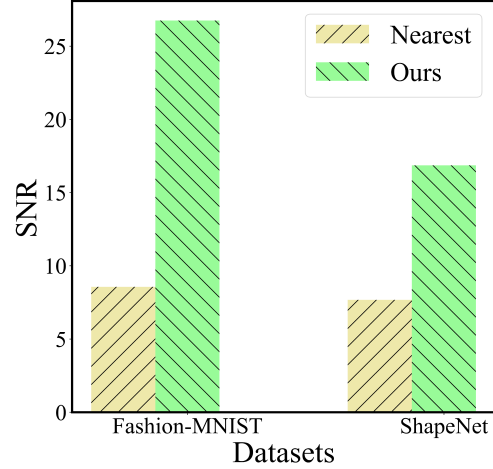


Figure 4. Comparisons of the mean SNR values of the results obtained with the nearest neighbor method and the proposed network on testing sets. The signals recovered by the proposed network have much higher SNR on both testing sets.

the hidden object is contained in adjacent time bins for adjacent detection pairs. **It means that the problem can not be solved by simply applying 2-D image superresolution to each time slice.**

In this section, we introduce the neural network which learns the nonlinear signal recovery operator Φ and the training dataset. Details about the training process and the ablation study are provided in the supplement.

3.1. Network architecture

Recently, attention mechanism [4, 36] has become increasingly popular in the deep learning society. However, it has been shown that there is no need to apply the attention mechanism to every part of the network [6].

To learn the operator, we generalize the attention in attention net (A²N) [6] which has two branches and dropout modules for the attention mechanism. One of the branches uses the conventional method to provide a rough estimation of the high resolution signal, and the other branch employs the network. The final output is the sum of the results from these two branches. The network architecture with two branches has been widely studied in the field of image superresolution [2, 19, 43].

The proposed signal superresolution network (SSN) is illustrated in Fig. 3. The learning branch is composed of several convolution layers and 3-D attention in attention blocks (3D-A²B). Each block has a dropout module to learn the weight between the attention branch and the non-attention branch, which can determine whether the attention mechanism is needed in the block. The reconstruction module contains several convolution layers to reconstruct the high resolution signal through the extracted feature. The low di-

mensional feature is first interpolated to a high dimensional feature using trilinear interpolation, and then the feature is processed by several convolution layers. The output of the convolution layer is passed to the next part of the reconstruction module and the process is repeated.

The generalization of the network from A^2N [6] to SSN is not trivial. Based on our analysis, the contribution of one point on the hidden object is contained in adjacent time bins for adjacent detection pairs. Conventional interpolation methods will perform poorly when the spacing between adjacent detection pairs gets larger. Thus, 3-D kernels are needed to extract and recover the information, and the size of the kernels is relevant to the spacing. In addition, the signal is only interpolated in spatial dimensions instead of all three dimensions to make the operator easier to learn.

3.2. Dataset

For all experiments in this paper, we use subsets of Fashion-MNIST [40] and ShapeNet [5] to generate synthetic signals for training and testing. The Fashion-MNIST dataset consists of 2-D figures, while the ShapeNet comprises 3-D objects. We choose 5,000 samples from the training set of Fashion-MNIST and the class ‘‘Car’’ of ShapeNet respectively to generate our training set. In addition, we choose 1,000 samples from the testing set of Fashion-MNIST and the class ‘‘Car’’ of ShapeNet respectively to generate the testing set. Thus, the training set used in this work comprises 10,000 signal pairs, and the testing set comprises 2,000. All the signals are generated using Eq. (4). Besides, we set $N_x = N_y = 8$, $N_t = 100$ in this paper. For measured signals whose N_t is larger than 100, the network can also be applied directly since the network operates locally.

For the confocal scenario, to match the experimental setting with the Stanford dataset [22], the 2-D figures from FashionMNIST [40] are placed 1 m away from the visible wall, the size of the figures is set as $1 \times 1 \text{ m}^2$. And the 3-D objects from ShapeNet [5] are placed in a $1 \times 1 \times 0.2 \text{ m}^3$ (horizontal, vertical, depth) box with a maximum depth of 1.1 m. The illumination region on the visible wall is $2 \times 2 \text{ m}^2$, and the time resolution of the SPAD is 32 ps.

For the non-confocal setting, we match the experimental setting with the data provided by phasor field [25]. The 2-D figures from Fashion-MNIST are placed 1 m away from the wall, and the size is $1.25 \times 1.25 \text{ m}^2$. The 3-D objects from ShapeNet are placed in a $1.25 \times 1.25 \times 0.2 \text{ m}^3$ (horizontal, vertical, depth) box with a maximum depth of 1.1 m. The illumination region on the visible wall is $1.25 \times 1.25 \text{ m}^2$, and the time resolution of the SPAD is 16 ps.

4. Results

In this section, we show the reconstruction results of the proposed pipeline on both synthetic and measured data. For

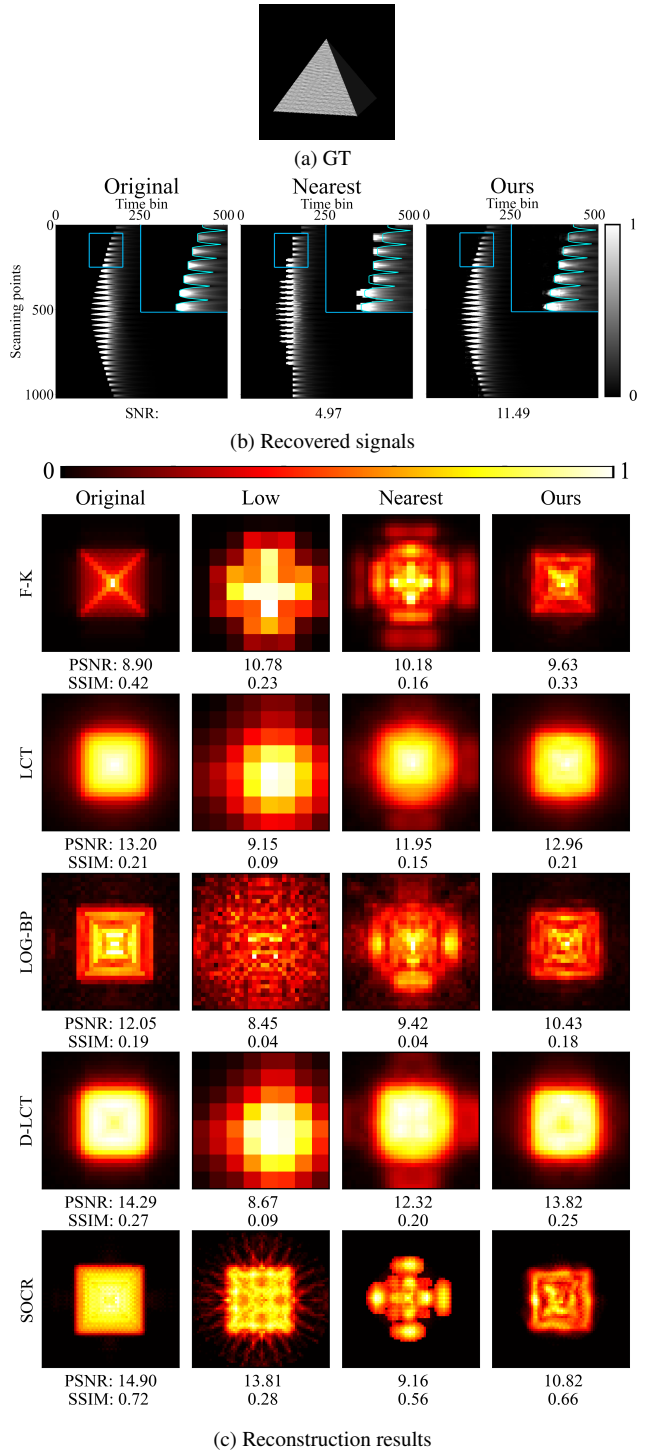


Figure 5. Recovered signals and reconstruction results of the pyramid. (a) Ground truth of the hidden object. (b) A comparison of the recovered signals (the first arrival time of the original signal is marked by the blue curve in the zoom-in window of each sub-figure). (c) Reconstruction results of state-of-the-art methods with different signals.

each instance, reconstruction results obtained with signals of different spatial resolutions are shown to validate the effectiveness of the proposed pipeline. The first column (Original) shows reconstructions obtained from the original signal with a spatial resolution of 32×32 ; results in the second column (Low) are reconstructed using the signal scanned at 8×8 points; the third column (Nearest) shows reconstructions of the recovered signal obtained with nearest neighbor interpolation; the last column (Ours) illustrates reconstruction results obtained with the proposed pipeline. Comparisons with other interpolation methods and existing learning based NLOS imaging method are provided in the supplement.

To illustrate the proposed pipeline can be acted as a plug-in-and-play module for most algorithms, we combine state-of-the-art methods with our pipeline. For confocal experiments, the results are reconstructed by F-K [22], LCT [31], LOG-BP [18], D-LCT [45] and SOCR [27] algorithms, while for non-confocal experiments, results are given by LOG-BP, Phasor [24] and SOCR methods. For each instance, the parameters are fixed in each algorithm.

To assess the results quantitatively, we compute the signal to noise ratio (SNR) of the recovered signals, as well as the peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM) of the reconstructed maximum intensity projections. For measured data, we use the reconstructions obtained from the original signals as references to compute PSNR and SSIM.

4.1. Results of confocal experiment

In Fig. 4, we compare the mean SNR values of the signals interpolated with SSN and the nearest neighbor method. It is shown that the network provides much better results on both datasets. In other words, the network approximates the oracle interpolation operator Φ with high accuracy and serves as a crucial part of the proposed pipeline.

To test the robustness of the proposed pipeline, we use the instance of the pyramid from SOCR [27], whose signal was generated with a different physical model. As shown in Fig. 5b, the first arrival time is correctly recovered by the proposed network, while the nearest neighbor interpolation method fails. Fig. 5c shows that the proposed pipeline provides a signal for high quality reconstructions with better visual quality and higher PSNR and SSIM values.

The proposed pipeline is then tested on measured data from Stanford dataset [22]. This dataset contains signals measured with different hidden objects and exposure time. We choose the instance of the statue to validate the effectiveness of the proposed pipeline. The measurement time of each sampling point is 0.0069 s. The total exposure time of the signal scanned at 32×32 points is 7.03 s, and it only takes 0.44 s to scan 8×8 points. As is shown in Fig. 6b, the result of SSN has higher SNR. The first arrival time can

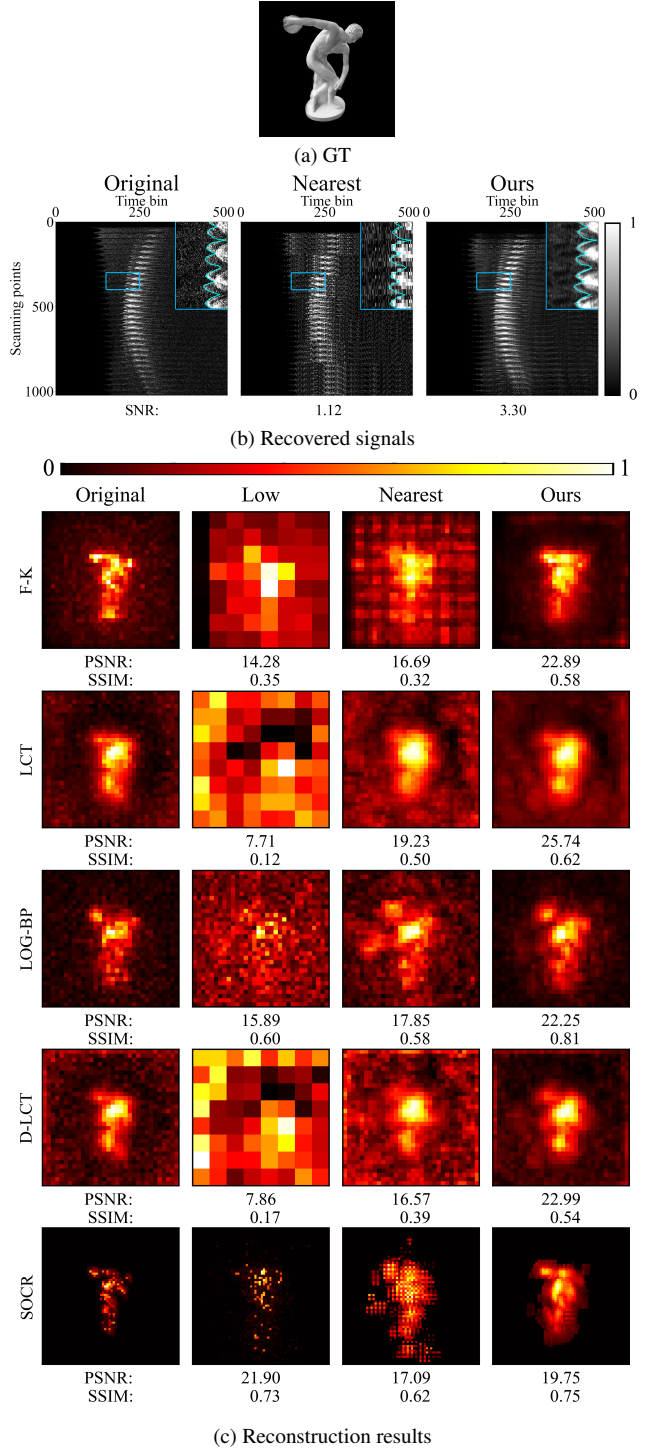


Figure 6. Recovered signals and reconstruction results of the statue. (a) Ground truth of the hidden object. (b) A comparison of recovered signals (the first arrival time of the original signal is marked by the blue curve in the zoom-in window of each sub-figure). (c) Reconstruction results of state-of-the-art methods with different signals.

hardly be distinguished in the nearest neighbor interpolated signal, while it is much clearer in the signal recovered by the proposed network. As shown in Fig. 6c, the proposed pipeline results in fewer artifacts in the background. More comparisons of different exposure times are provided in the supplement, which also indicates the outperformance of the proposed pipeline.

4.2. Results of the dynamic scene

To illustrate the potential application of the proposed pipeline on real-time NLOS imaging, we test it on the captured dynamic scenes from the Stanford dataset. The data is captured over a 2×2 m² region on the visible wall, and the person is wearing a retroreflective suit to increase the detection of the indirect signal. It takes 0.25 s to scan 32×32 points, which allows for a reconstruction video with 4 frames per second. The reflection model of the retroreflective material [31] is different from the one defined in Eq. (6).

The results of several frames are shown in Fig. 7b, the complete video of the reconstruction results can be found in the supplement. All the results are reconstructed by the F-K method. Although the training set is generated by a different forward model, the proposed pipeline can still recover signals for high quality reconstructions. The pose of the person can be clearly identified in the reconstructions using the proposed pipeline, while the reconstructions using the nearest neighbor interpolated signal contain more artifacts. The mean PSNR and SSIM of the reconstruction results obtained from the proposed pipeline are 24.46 and 0.63, while those obtained with nearest neighbor interpolation are only 18.68 and 0.37.

The total capture time of each frame is reduced from 0.25 s to 0.0156 s. This enables us to obtain a high quality reconstruction video with **64 frames per second**. Besides, it only takes 0.08 s to recover the high resolution signal with the proposed network, and 0.01s to reconstruct the hidden object with the F-K method on an NVIDIA RTX 3090. Thus, real-time video imaging with a high frame rate can be realized with a delay of a few frames.

4.3. Results of non-confocal experiment

For the non-confocal setting, we use the instance of “4” provided by Phasor [24] to validate the effectiveness of the proposed method. The measured signal is subsampled to 32×32 and the illumination region on the visible wall is cropped to 1.25×1.25 m². We use the signal scanned at 8×8 points as the input of our pipeline. It takes 1 s to scan each point on the visible wall.

As is shown in Fig. 8b, the SNR of the signal recovered by the network is much higher than the one interpolated with the nearest neighbor method. Reconstruction results are shown in Fig. 8c. The proposed pipeline provides re-

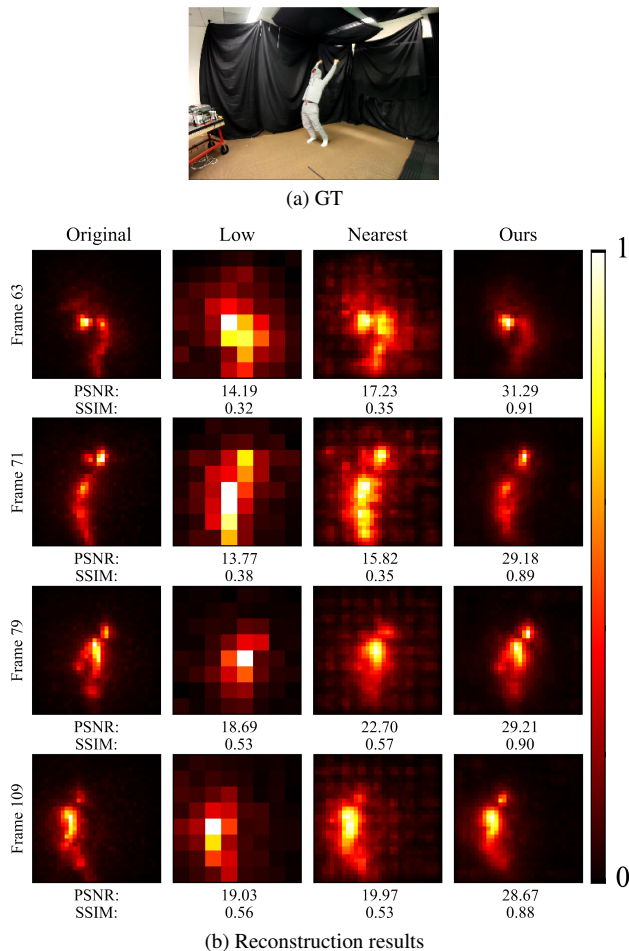


Figure 7. Reconstruction results of the dynamic scene. (a) An illustration of the dynamic scene. (b) Reconstruction results of the dynamic scene. The results are obtained with the F-K method. The PSNR and SSIM values are computed with the reconstructions obtained with the original signals as references. The proposed pipeline provides much better reconstructions in terms of both quantitative criteria and visual quality.

constructions with higher quantitative and visual quality.

5. Discussion

We have introduced a novel pipeline for high quality NLOS reconstruction with the signal measured at only a small number of scanning points. In this section, we discuss several factors that affect the reconstruction quality and some feasible solutions to improve the pipeline further.

5.1. Training of the network

The training of the network employed in the proposed pipeline is supervised. Thus, it is affected by the generation of the training set, including the forward model, experimental setting, and hidden objects. For example, hidden objects

with different materials will lead to various physical models. The networks trained and tested on signals with different models would bring additional errors to the results.

However, a network with good generalization ability can avoid degradation. As shown in Fig. 7, although the physical model of the dynamic scenes differs from the one used in the training set, the reconstructed objects are only slightly affected. Thus, the proposed pipeline has strong robustness to the forward model.

5.2. Performance of different algorithms

As the reconstruction results shown in previous sections, the improvement of reconstruction quality of the proposed pipeline varies between different methods. Compared to the reconstruction results obtained from the low resolution signal, fast imaging algorithms like F-K, LCT and D-LCT have larger improvements, while the iterative inversion algorithm SOCR might have similar visual quality in some cases. This phenomenon may come from the fact that the iterative algorithm can make full use of the information contained in the signal, but with a longer reconstruction time.

5.3. Combination with SPAD array for further acceleration

Since the proposed pipeline can recover the signal faithfully under the non-confocal setting, combining it with a SPAD array can accelerate the measurement process further. However, the existing SPAD array can only focus on a rectangular area on the visible wall, which means that the network has to learn an extrapolation operator. This is much harder than learning an interpolation operator. Thus, it would be more practical if the SPAD array could focus on sparse rectangular points on a relatively large area of the visible wall. This may be seen as a new area for hardware design.

6. Conclusion

In this paper, we have studied the signal recovery problem in NLOS imaging, and proposed a two-step learning based pipeline for fast NLOS imaging with only a small number of scanning points. In the first step, the signal of high spatial resolution is recovered with a neural network. In the second step, the hidden objects are obtained from the virtually measured signals at a dense grid. Our pipeline provides high-quality reconstructions under both confocal and non-confocal settings, which takes only 6.25% of the original measurement time. Besides, the proposed pipeline is compatible with most existing detection systems and imaging algorithms. With the proposed pipeline, the high frame rate imaging video is possible, which is a noteworthy advance in this area. It is expected to combine emerging techniques in the fields of SPAD array for real-time and high quality NLOS reconstructions with higher frame rates.

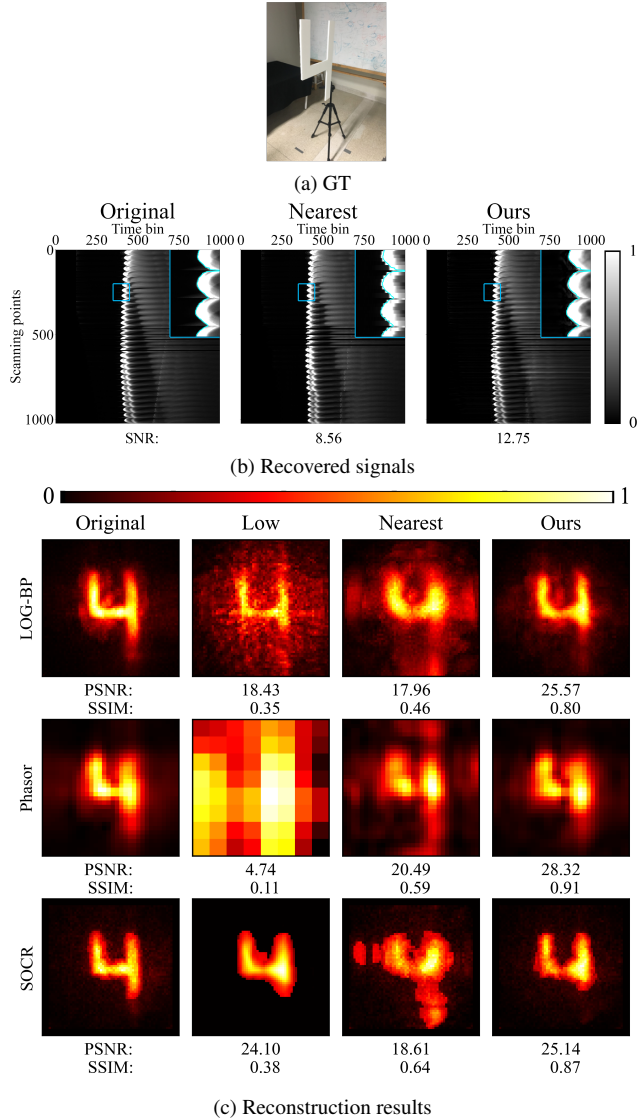


Figure 8. Recovered signals and reconstruction results of the instance “4”. (a) Ground truth of the hidden object. (b) A comparison of recovered signals (the first arrival time of the original signal is marked by the blue curve in the zoom-in window of each sub-figure). (c) Reconstruction results of state-of-the-art methods with different signals.

Acknowledgments. The authors would like to thank Zhupeng Li and Rongqian Wang for fruitful discussions. This work was supported by the National Natural Science Foundation of China (61975087, 12071244, 11971258).

References

- [1] Byeongjoo Ahn, Akshat Dave, Ashok Veeraraghavan, Ioannis Gkioulekas, and Aswin C Sankaranarayanan. Convolutional approximations to the general non-line-of-sight imag-

- ing operator. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7889–7899, 2019. 2
- [2] Saeed Anwar, Salman Khan, and Nick Barnes. A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020. 4
- [3] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. Fast back-projection for non-line of sight reconstruction. In *ACM SIGGRAPH 2017 Posters, SIGGRAPH '17*, New York, NY, USA, 2017. Association for Computing Machinery. 2
- [4] Syed Muhammad Arsalan Bashir, Yi Wang, Mahrukh Khan, and Yilong Niu. A comprehensive review of deep learning-based single image super-resolution. *PeerJ Computer Science*, 7:e621, 2021. 4
- [5] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015. 5
- [6] Haoyu Chen, Jinjin Gu, and Zhi Zhang. Attention in attention network for image super-resolution. *arXiv preprint arXiv:2104.09497*, 2021. 4
- [7] Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. Steady-state non-line-of-sight imaging. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2
- [8] Wenzheng Chen, Fangyin Wei, Kiriakos N Kutulakos, Szymon Rusinkiewicz, and Felix Heide. Learned feature embeddings for non-line-of-sight imaging and recognition. *ACM Transactions on Graphics (TOG)*, 39(6):1–18, 2020. 2
- [9] Justin Dove and Jeffrey H. Shapiro. Paraxial theory of phasor-field imaging. *Opt. Express*, 27(13):18016–18037, Jun 2019. 2
- [10] Xiaohua Feng and Liang Gao. Improving non-line-of-sight image reconstruction with weighting factors. *Optics letters*, 45(14):3921–3924, 2020. 2
- [11] Xiaohua Feng and Liang Gao. Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging. *Nature Communications*, 12:2179, 04 2021. 2
- [12] Javier Grau Chopite, Matthias B. Hullin, Michael Wand, and Julian Iseringhausen. Deep non-line-of-sight reconstruction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 957–966, 2020. 2
- [13] Felix Heide, Matthew O’Toole, Kai Zang, David B. Lindell, Steven Diamond, and Gordon Wetzstein. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Trans. Graph.*, 38(3), May 2019. 2
- [14] Julian Iseringhausen and Matthias B Hullin. Non-line-of-sight reconstruction using efficient transient rendering. *ACM Transactions on Graphics (TOG)*, 39(1):1–14, 2020. 2
- [15] Mariko Isogawa, Dorian Chan, Ye Yuan, Kris M. Kitani, and Matthew O’Toole. Efficient non-line-of-sight imaging from transient sinograms. In *16th European Conference on Computer Vision (ECCV)*, 2020. 2, 3
- [16] Mariko Isogawa, Ye Yuan, Matthew O’Toole, and Kris Kitani. Optical non-line-of-sight physics-based 3d human pose estimation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7011–7020, 2020. 2
- [17] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166, 2009. 1
- [18] Martin Laurenzis and Andreas Velten. Feature selection and back-projection algorithms for nonline-of-sight laser-gated viewing. *Journal of Electronic Imaging*, 23:063003, 11 2014. 2, 3, 6
- [19] Y. Li, B. Sixou, and F. Peyrin. A review of the deep learning methods for medical images super resolution problems. *IRBM*, 42(2):120–133, 2021. 4
- [20] Zhupeng Li, Xintong Liu, Jianyu Wang, Zuoqiang Shi, Lingyun Qiu, and Xing Fu. Fast non-line-of-sight imaging based on first photon event stamping. *Optics letters*, 47(8):1928–1931, April 2022. 2
- [21] Zhengpeng Liao, Deyang Jiang, Xiaochun Liu, Andreas Velten, Yajun Ha, and Xin Lou. Fpga accelerator for real-time non-line-of-sight imaging. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2022. 2
- [22] David B Lindell, Gordon Wetzstein, and Matthew O’Toole. Wave-based non-line-of-sight imaging using fast fk migration. *ACM Transactions on Graphics (ToG)*, 38(4):1–13, 2019. 2, 3, 5, 6
- [23] Jianjiang Liu, Yijun Zhou, Xin Huang, Zheng-Ping Li, and Feihu Xu. Photon-efficient non-line-of-sight imaging. *IEEE Transactions on Computational Imaging*, 8:639–650, 2022. 2
- [24] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. *Nature Communications*, 11:1645, 2020. 2, 3, 6, 7
- [25] Xiaochun Liu, Ibón Guillén, Marco Manna, Ji Nam, Azer Reza, Toan Le, Adrián Jarabo, Diego Gutiérrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 572, 2019. 2, 5
- [26] Xiaochun Liu and Andreas Velten. The role of wigner distribution function in non-line-of-sight imaging. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2020. 2
- [27] Xintong Liu, Jianyu Wang, Zhupeng Li, Zuoqiang Shi, Xing Fu, and Lingyun Qiu. Non-line-of-sight reconstruction with signal-object collaborative regularization. *Light: Science & Applications*, 08 2021. 2, 3, 6
- [28] Christopher Metzler, David Lindell, and Gordon Wetzstein. Keyhole imaging: Non-line-of-sight imaging and tracking of moving objects along a single optical path. *IEEE Transactions on Computational Imaging*, PP:1–1, 12 2020. 2, 3
- [29] Christopher A. Metzler, Felix Heide, Prasana Rangarajan, Muralidhar Madabhushi Balaji, Aparna Viswanath, Ashok Veeraraghavan, and Richard G. Baraniuk. Deep-inverse correlative: towards real-time high-resolution non-line-of-sight imaging. *Optica*, 7(1):63–71, Jan 2020. 2
- [30] Ji Hyun Nam, Eric Brandt, Sebastian Bauer, Xiaochun Liu, Marco Renna, Alberto Tosi, Eftychios Sifakis, and Andreas

- Velten. Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second. *Nature communications*, 12(1):1–10, 2021. [2](#), [4](#)
- [31] Matthew O’Toole, David Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555, 2018. [2](#), [3](#), [6](#), [7](#)
- [32] Chengquan Pei, Anke Zhang, Yue Deng, Feihu Xu, Jiamin Wu, David U-Lei Li, Hui Qiao, Lu Fang, and Qionghai Dai. Dynamic non-line-of-sight imaging system based on the optimization of point spread functions. *Opt. Express*, 29(20):32349–32364, Sep 2021. [2](#)
- [33] Syed Azer Reza, Marco La Manna, Sebastian Bauer, and Andreas Velten. Phasor field waves: A Huygens-like light transport model for non-line-of-sight imaging applications. *Opt. Express*, 27(20):29380–29400, Sep 2019. [2](#)
- [34] Christos Thrampoulidis, Gal Shulkind, Feihu Xu, William T. Freeman, Jeffrey H. Shapiro, Antonio Torralba, Franco N. C. Wong, and Gregory W. Wornell. Exploiting occlusion in non-line-of-sight active imaging. *IEEE Transactions on Computational Imaging*, 4(3):419–431, 2018. [2](#)
- [35] Chia-Yin Tsai, Aswin C Sankaranarayanan, and Ioannis Gkioulekas. Beyond volumetric albedo—a surface optimization framework for non-line-of-sight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1545–1555, 2019. [2](#)
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc. [4](#)
- [37] Eric Veach and Leonidas J. Guibas. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’95*, page 419–428, New York, NY, USA, 1995. Association for Computing Machinery. [3](#)
- [38] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Mounsi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3(1):1–8, 2012. [2](#)
- [39] Cheng Wu, Jianjiang Liu, Xin Huang, Zheng-Ping Li, Chao Yu, Jun-Tian Ye, Jun Zhang, Qiang Zhang, Xiankang Dou, Vivek Goyal, Feihu Xu, and Jian-Wei Pan. Non-line-of-sight imaging over 1.43 km. *Proceedings of the National Academy of Sciences of the United States of America*, 118, 03 2021. [2](#)
- [40] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. [5](#)
- [41] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. A theory of Fermat paths for non-line-of-sight shape reconstruction. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6793–6802, 2019. [2](#)
- [42] Wenqing Yang, Chao Zhang, Wenjie Jiang, Zexin Zhang, and Baoqing Sun. Non-line-of-sight imaging enhanced with spatial multiplexing. *Opt. Express*, 30(4):5855–5867, Feb 2022. [2](#), [4](#)
- [43] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, Dec 2019. [4](#)
- [44] Jun-Tian Ye, Xin Huang, Zheng-Ping Li, and Feihu Xu. Compressed sensing for active non-line-of-sight imaging. *Opt. Express*, 29(2):1749–1763, Jan 2021. [2](#), [3](#)
- [45] Sean I Young, David B Lindell, Bernd Girod, David Taubman, and Gordon Wetzstein. Non-line-of-sight surface reconstruction using the directional light-cone transform. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1407–1416, 2020. [2](#), [3](#), [6](#)