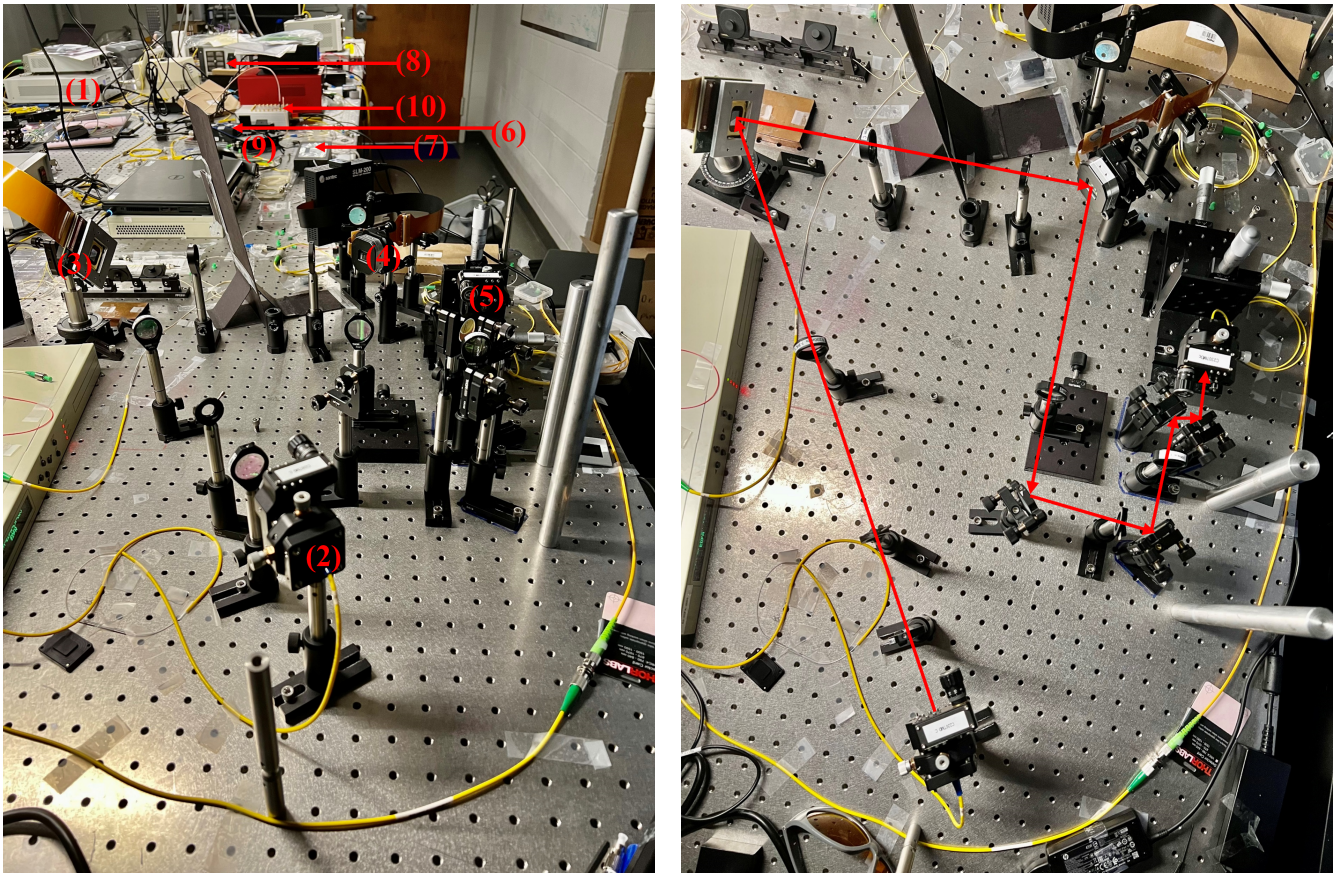
# 1 Materials & Methods

## 1.1 Experimental setup in our lab

The pictures of the experimental setup are shown in Figure S1 (a) a side view and (b) a top view. In Fig. S1(a), the Mode-Locked Laser (MLL, Calmar Mendocino) is labeled as (1). The probe light propagates through a fiber coupler (FC1, C220TMD C, Thorlabs) labeled as (2) to transmit into free space. The DMD is indicated as (3), while the SLM is labeled as (4). The probe light is subsequently coupled into the FC2 (C230TMD C, Thorlabs) labeled as (5).




(a) (b)

**Figure S1.** Photos of the experimental setup: (a) side view and (b) top view of the free-space setup. The labels refer to the following components: (1) MLL: Mode-locked laser sends out pulse trains for the probe and pump. (2) FC1: Fiber coupler couples the probe out to free space. (3) DMD: Digital micromirror device projects Walsh 2D masks. (4) SLM: Spatial light modulator uploads digit images from the MNIST dataset. (5) FC2: Fiber coupler couples the probe into a fiber. (6) PPLN module: Periodically poled lithium niobate module upconverts the probe and pump into high frequency. (7) ODL: Optical delay line helps the probe and pump achieve temporal alignment. (8) InGaAs SPD: Indium gallium arsenide single-photon detector is used for direct detection (DD) at the single-photon level. (9) Si SPD: Silicon single-photon detector is used for Quantum Parametric Mode Sorting (QPMS) detection at the single-photon level. (10) Time Tagger: Device used for collecting data.

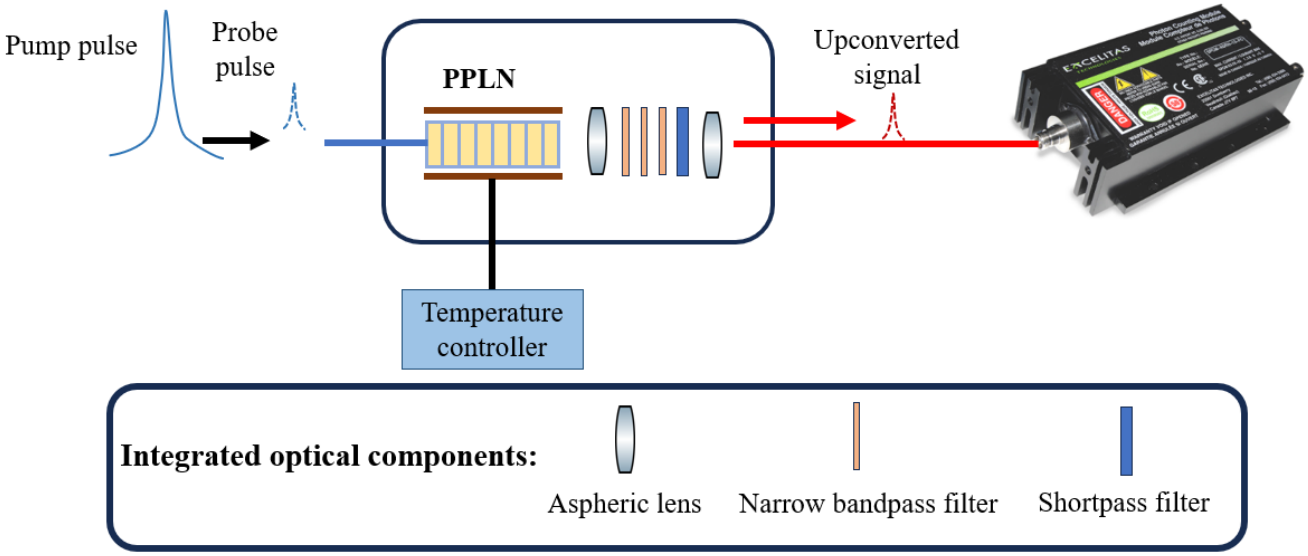



**Figure S2.** A schematic diagram of fiber coupled upconversion module, integrating temperature controller PPLN and a series of optical filters.

For the QPMS, it contains a temperature controlled Fiber-coupled Magnesium-doped Periodic Poled Lithium Niobate (PPLN) waveguide, aspheric lenses and a series of optical filters and shown in Fig. S2. The length of the waveguide is 1.97 cm, and temperature is stabilized at 36.8±0.1°C. With the aspheric lenses, input coupling efficiency from a single mode fiber is 31% and output coupling to a multimode fiber is 81%. After the output coupling lens, three narrow band pass filters are applied to provide out of band rejection of $\sim$ 100dB and a shortpass filter is used to remove the residual pulses. The series of filters are crucial. They minimize the dark count of the PPLN module. The measured phase curve is centered at 1559.79 nm. We check the sum frequency generation (SFG) of probe 1554.1 nm and pump as 1564.3 nm with the FWHM bandwidth of 0.804 nm and maximum normalized internal conversion efficiency of 207.7%$W^{-1}cm^{-2}$. The total detection efficiency of the QPMS from the output of the 50:50 beam splitter to the detector is 12.0%, transmission loss of the filters, and the Si SPD's quantum efficiency which is 66%.

### 1.2 Classical setup and measurements

The experimental setup is sketched in Fig. S3. It involves the use of a femtosecond MLL emitting light centered at $\sim$1550 nm and operating at a repetition rate of 50 MHz. The pulse train passes through a Wavelength Division Multiplexing (WDM), which splits the light into a probe at 1554.3 nm and a pump at 1564.3 nm. The probe and pump signals are then amplified using erbium-doped fiber amplifiers (EDFAs) and further filtered using additional WDMs. The probe light, after being coupled out to free space, passes through a DMD and a SLM. The 4-f relay lenses are employed for each of these components. The SLM contains the phase patterns from the Modified National Institute of Standards and Technology (MNIST) dataset. In the DD scenario, the resulting probe is collected by an InGaAs Amplified Photodiode (PDA20CS, Thorlabs). The InGaAs photodiode converts the optical signal into an electrical signal, allowing for DD of the probe's intensity. For QPMS, the resulting probe is sent into a fiber polarization controller and then combined with the pump at 1564.3 nm using a WDM. The combined signal, along with the pump, passes through an Optical Delay Line (ODL) and another fiber polarization controller. The combined signal and pump, from the common port of the third 1564.3 nm WDM, are then directed into a periodic poled lithium niobate (PPLN) module. The PPLN module is used to convert the signal into a different wavelength, specifically a 779.64 nm light. The upconverted light at 779.64 nm is collected by a Silicon Amplified Photodiode (PDA100A2, Thorlabs), which converts the optical signal into an electrical signal for further analysis or processing.

The results presented in this subsection correspond to the classical level of detection. Figure. S4 (a) and (c) depict the normalized confusion matrices for DD and QPMS, respectively. Similarly, Fig. S4 (b) and (d) display the trends of the training loss (blue curve), test loss (orange curve), and test accuracy (green curve) for DD and QPMS, respectively. These results provide valuable insights into the performance and learning dynamics of both detection methods at the classical level. Fig. S4 shows that both DD and QPMS achieve stable and high classification accuracy within 200 training epochs. This indicates that our experimental setup is capable of effectively extracting the target features.

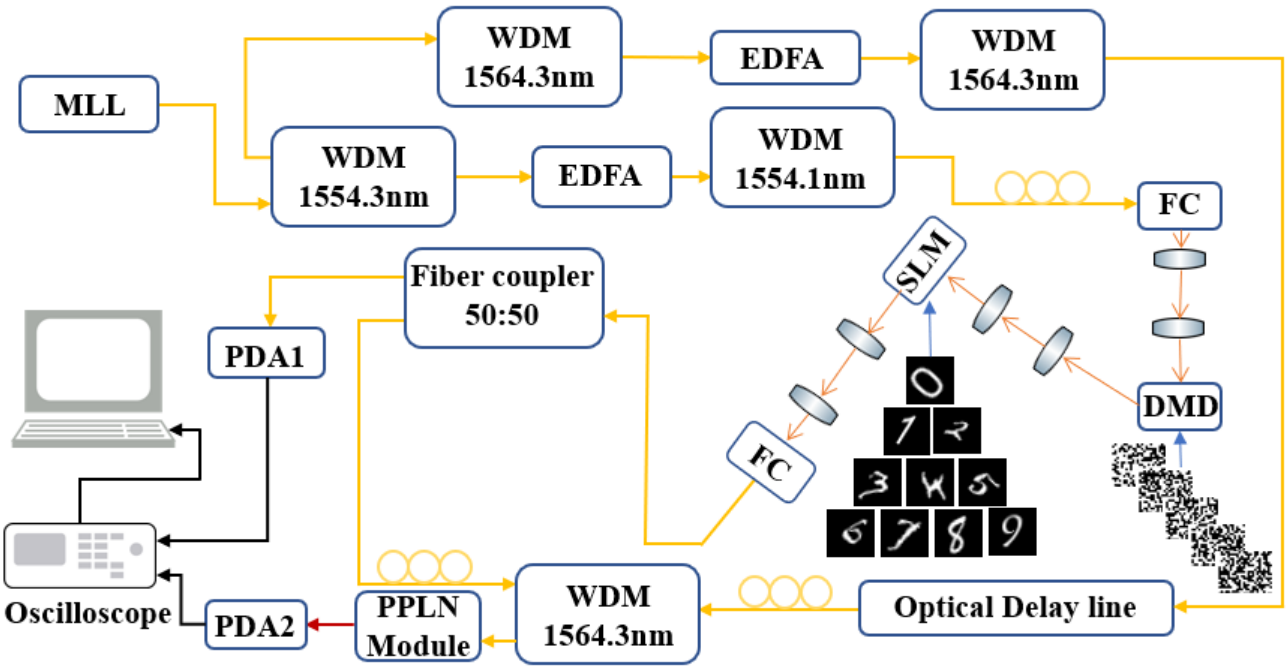



**Figure S3.** A schematic diagram of the experimental setup for classical level detection.

Initially, our focus is on achieving good results at the classical power level. As we succeeded in obtaining satisfactory outcomes, it provides us the confidence to reduce the probe power to single photon counting in the experiment setup.

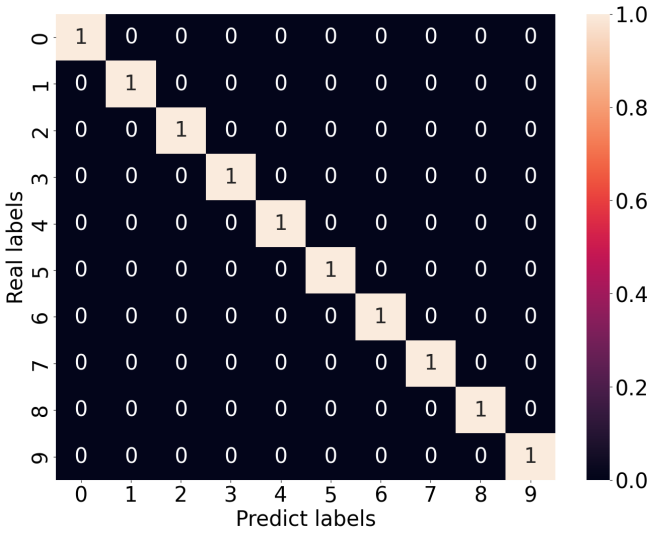
## 2 Results at low photon counts

This section discusses the study conducted at low photon levels, specifically focusing on the challenges associated with detecting and analyzing data in such conditions. The pump power is set as 16.5 dBm on the EDFA in the pump arm.

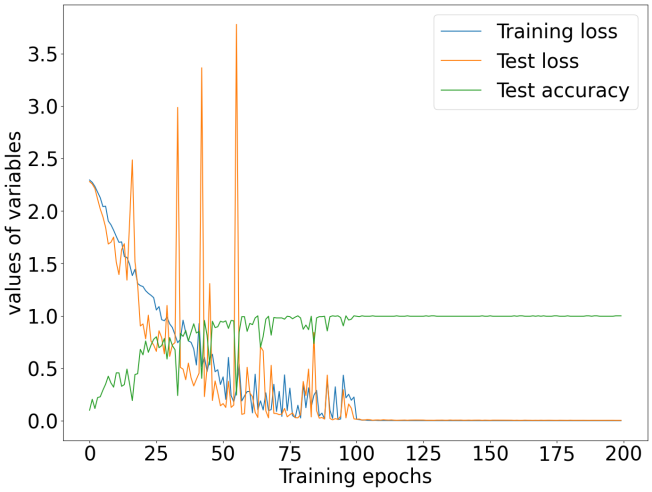
### 2.1 The influence of dark counts

In this section, all the data acquired in the experiment setup is collected with a probe power of 14 nW after FC1. The electrical controlled variable attenuator is set at 0 V which mean no attenuation added during the data acquisition process. Different integration time on both SPDs lead to varying photon counts collected on either detector. Due to the low photon detection levels, the variations in the total photon counts become more susceptible to various sources of noise such as Poisson noise, slight temperature fluctuations, laser instability, fiber polarization instability, device vibrations, etc. To investigate the influence of dark counts, different effective dwell times on the DMD have been set. The corresponding integration times on the Si-SPD and the InGaAs SPD have been adjusted accordingly to ensure a sufficient number of statistical samples for the required data without saturation on the SPDs.

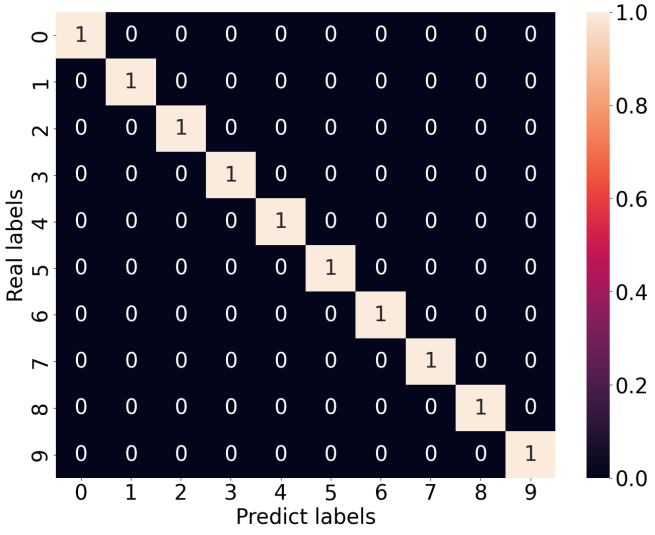
Table S1 provides the relationships between the effective dwell time on the DMD, the integration time on the SPDs and average photon count variations observed during the experimental measurements. As mentioned in the paper, there are 40, 100, or 300 Walsh 2D patterns uploaded on the DMD. There are 25 photon-counting events on Si SPD and 10 photon-counting events on InGaAs SPD during every Walsh 2D patterns dwell on the DMD. Calculate the sum of the photons of each effective dwell time on DMD disregarding the photon counting events during the changing pattern time on the DMD. The average photon counts is the mean values of the sum photons per effective dwell time over all the 300 different Walsh 2D patterns uploaded on the DMD. The average photon counts range is the photons changing range over all these masks dwell on DMD. As shown in the table S1, the lower limitation of the average photon counts range of each integration time on the Si SPD is similar to the corresponding green squares in figure 4 in our paper. When the dark counts approaches to the corresponding upper limitation of the average photon counts range of each integration time, the data become more difficult to be processed. When the integration time become less than 3 µs on the Si SPD and 25 µs on the InGaAs SPD, the data from the experiment setup become extremely hard to be processed as showing figure S5 and fig. S6. Fig. S5 and Fig. S6 provide valuable insights into the analysis of the original data for both DD and QPMS, respectively. The two white patterns added to the Walsh 2D mask sequence uploaded on DMD are distinguishable for large average photon counts, while they are indistinguishable in the low photon counts. These



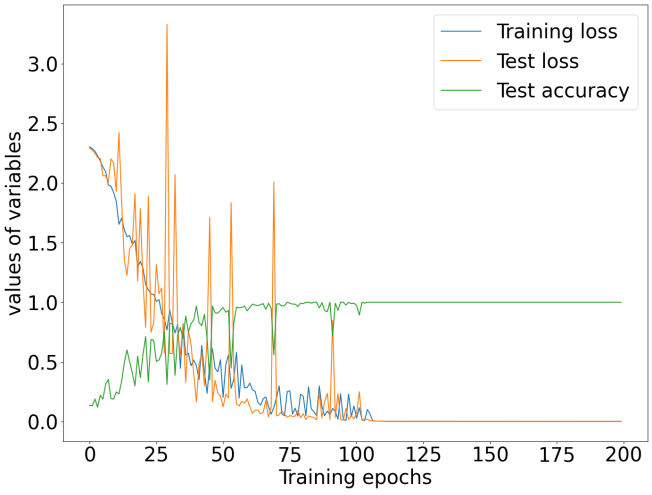

**(a)** Normalized confusion matrix of DD




**(b)** Trends of training loss, test loss, and test accuracy of DD

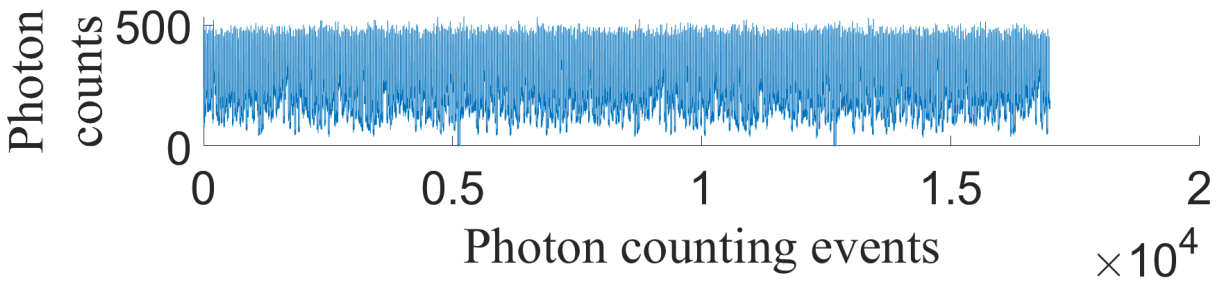



**(c)** Normalized confusion matrix of QPMS

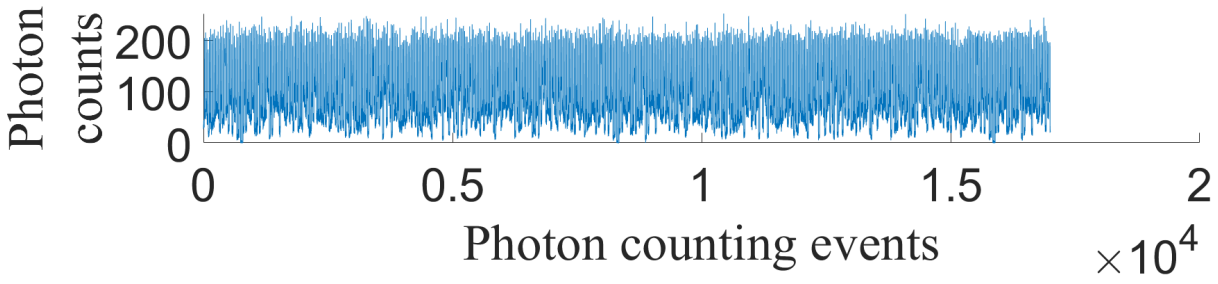



**(d)** Trends of training loss, test loss, and test accuracy of QPMS

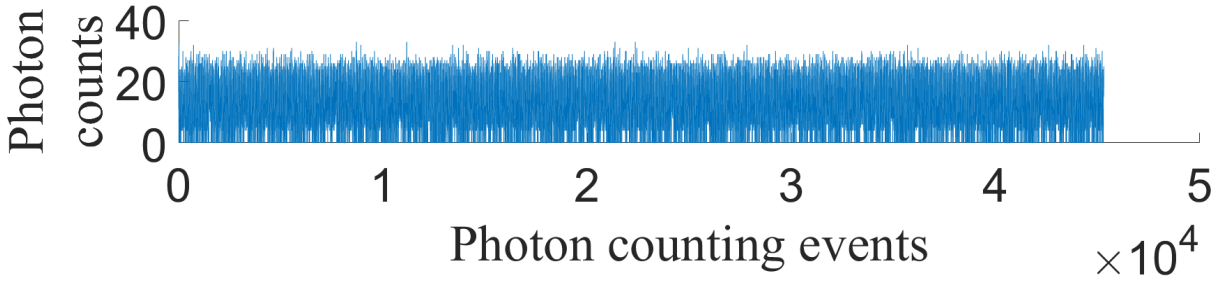
**Figure S4.** Detection result at classical level with the effective dwell time of each pattern on DMD is 800 µs.



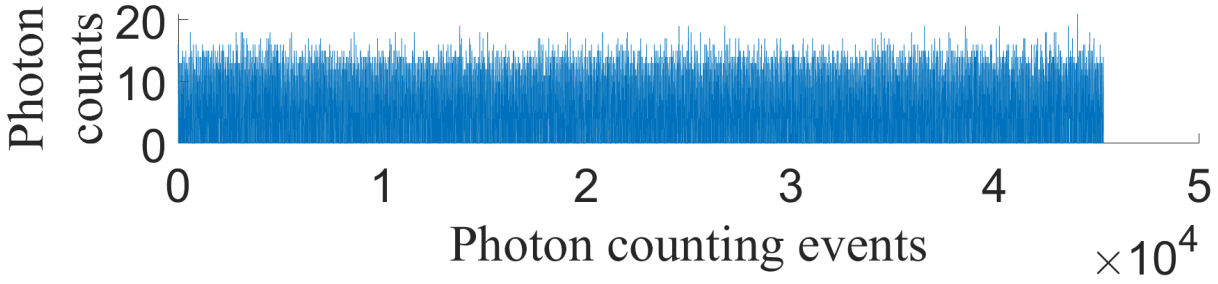

**(a)** The original data of the first image of digit 1, obtained when the effective dwell time on the DMD is 800 µs




**(b)** The original data of the first image of digit 8, obtained when the effective dwell time on the DMD is 800 µs

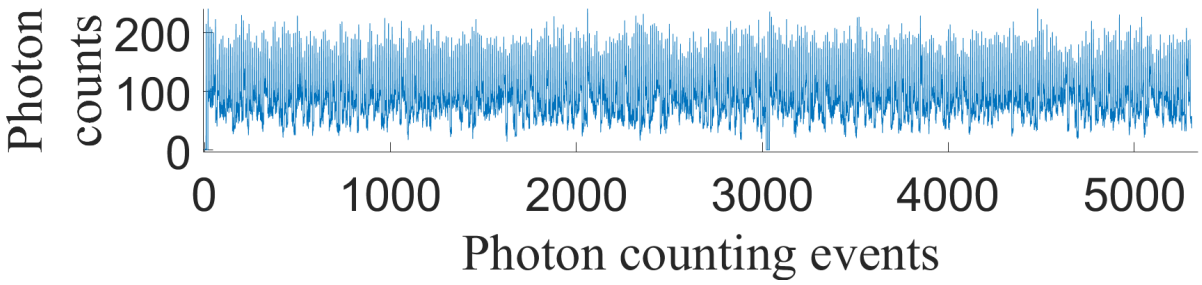



**(c)** The original data of the first image of digit 1, obtained when the effective dwell time on the DMD is 40 µs.

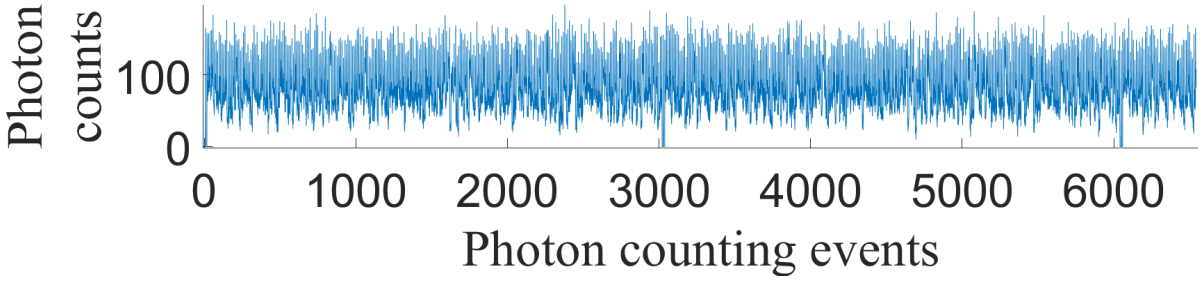



**(d)** The original data of the first image of digit 8, obtained when the effective dwell time on the DMD is 40 µs

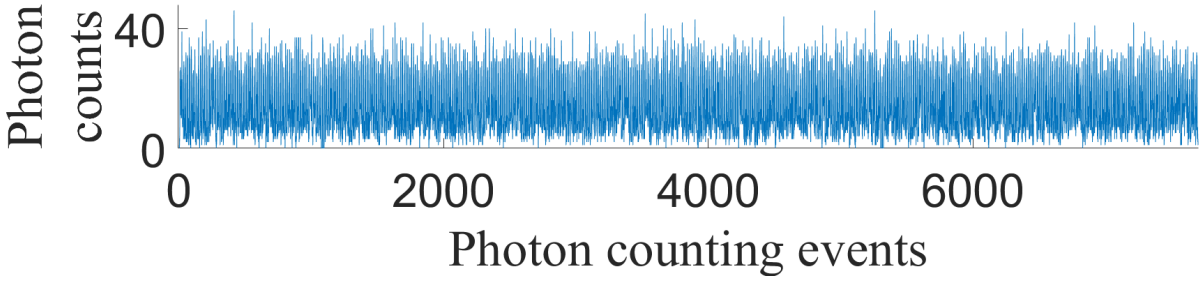
**Figure S5.** Original photon counting data for QPMS.



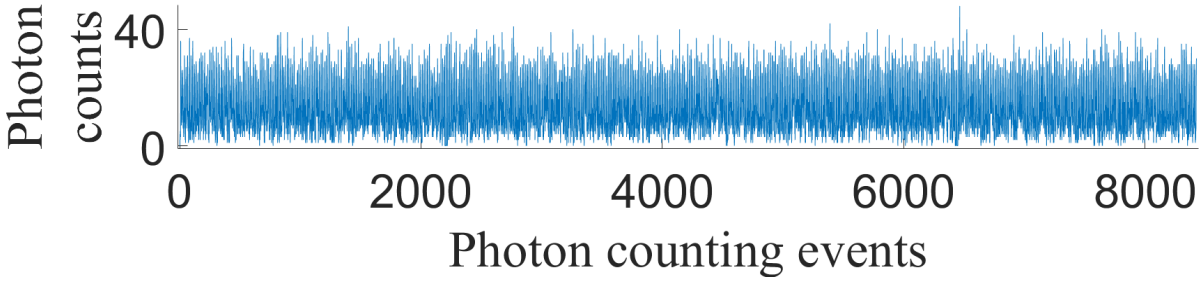

**(a)** The original data of the first image of digit 1, obtained when the effective dwell time on the DMD is 800 µs




**(b)** The original data of the first image of digit 8, obtained when the effective dwell time on the DMD is 800 µs

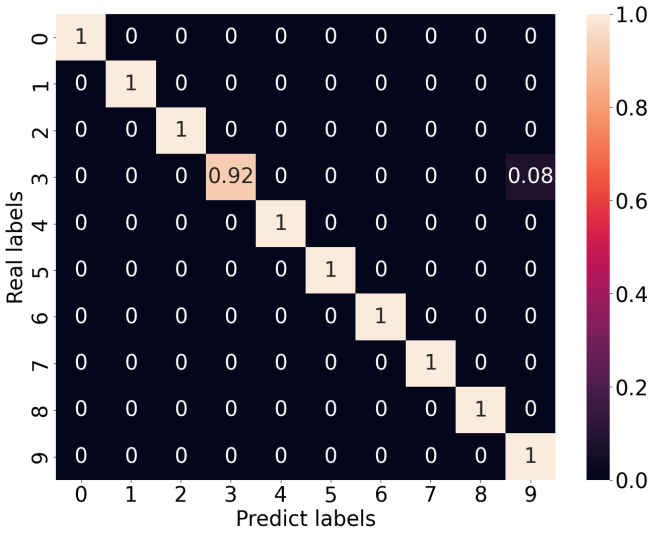



**(c)** The original data of the first image of digit 1, obtained when the effective dwell time on the DMD is 120 µs

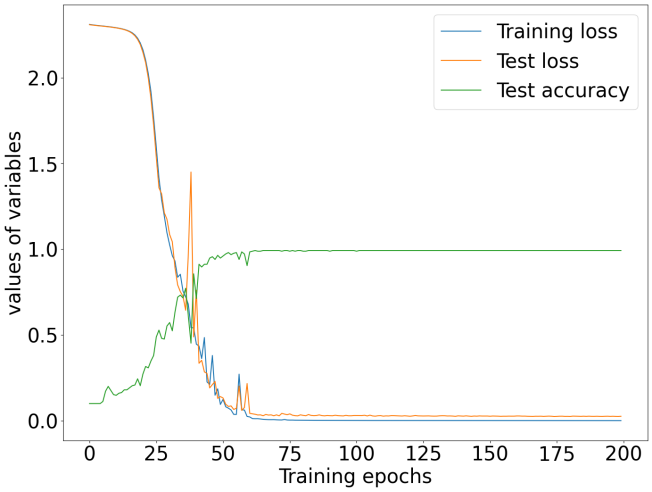



**(d)** The original data of the first image of digit 8, obtained when the effective dwell time on the DMD is 120 µs

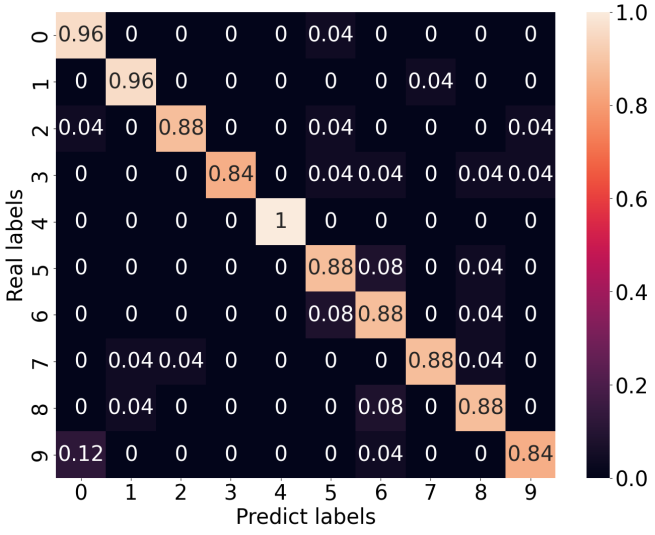
**Figure S6.** Original photon counting data for DD.



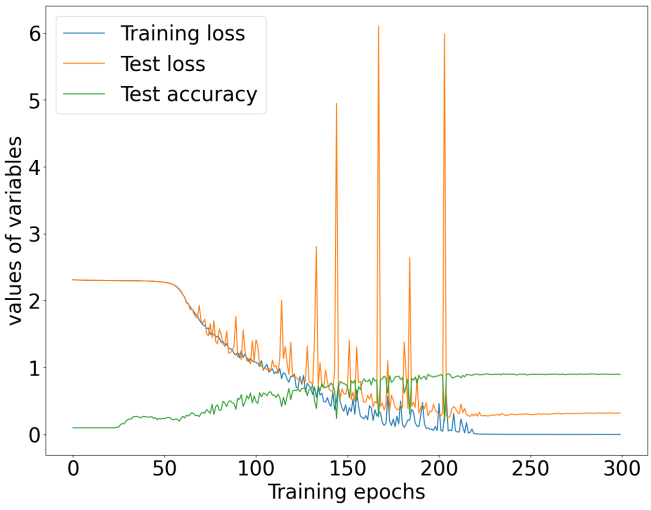

**(a)** Normalized confusion matrix of QPMS.




**(b)** Trends of training loss, test loss, and test accuracy of QPMS.




**(c)** Normalized confusion matrix of DD.




**(d)** Trends of training loss, test loss, and test accuracy of DD.

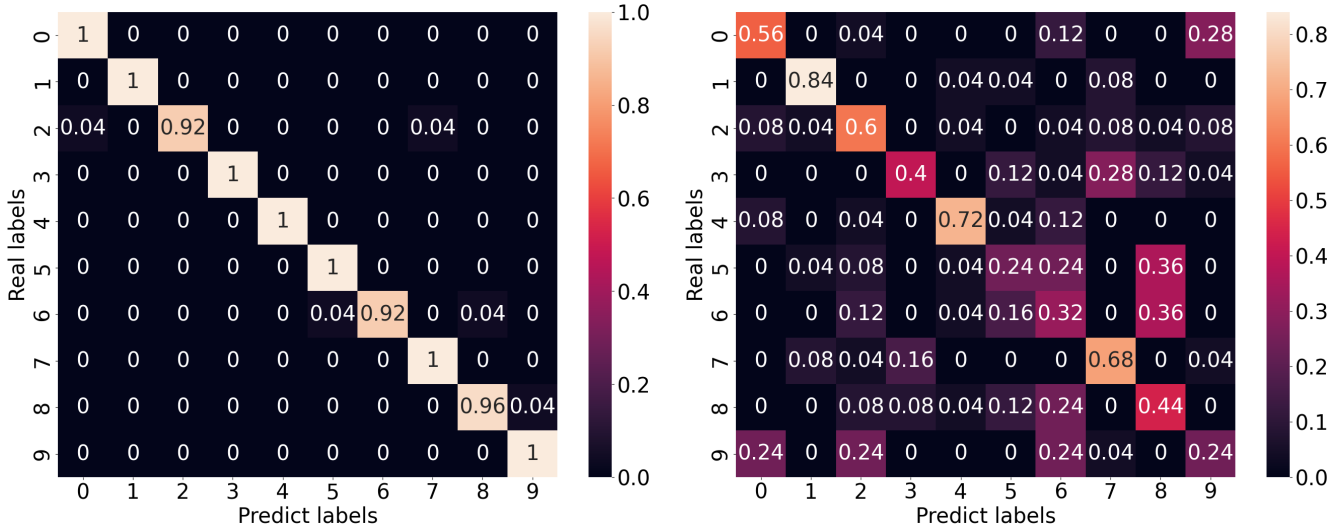
**Figure S7.** Detection result at single photon level with the effective dwell time of each pattern on DMD is 800 µs.

| Effective dwell time on DMD(μs) | QPMS detection(μs) | Average photon counts per mask | Average photon counts range over masks | Direct detection(μs) | Average photon counts per mask | Average photon count range over masks |
|---|---|---|---|---|---|---|
| 800 | 40 | 1990.6 | 151 – 7017 | 100 | 551.2 | 171 – 1229 |
| 700 | 35 | 1475.3 | 43 – 5652 | 87.5 | 474.1 | 155 – 973 |
| 600 | 30 | 1236.6 | 32 – 4834 | 75 | 398.6 | 123 – 831 |
| 500 | 25 | 1155.5 | 88 – 3906 | 62.5 | 326.8 | 108 – 691 |
| 400 | 20 | 897.1 | 74 – 3058 | 50 | 253.3 | 71 – 575 |
| 300 | 15 | 661.4 | 48 – 2260 | 37.5 | 180.6 | 45 – 400 |
| 200 | 10 | 485.8 | 10 – 1706 | 25 | 114.7 | 21 – 269 |
| 120 | 6 | 278.7 | 30 – 1030 | 15 | 59.5 | 12 – 156 |
| 100 | 5 | 225.7 | 30 – 817 | 12.5 | | |
| 60 | 3 | 119.7 | 3 – 495 | 7.5 | | |

**Table S1.** Based to the effective dwell time of each pattern on DMD, we determine the corresponding integration time on SPDs and the average photon counts for the patterns of single-pixel detection.

results illustrates that DD data is more susceptible at low photon levels compared to QPMS. Also, as the integration time on both SPDs is decreased, the impact of the dark counts increases significantly, making it exceedingly challenging to extract meaningful features from the probe data in the measurements.

By collecting data from both the Si SPD and InGaAs SPD, we obtain distinct classification results for varying effective dwell times on the DMD. Figure S7 and figure S8 show the outcomes for both QPMS and DD, considering different different effective dwell time for each pattern on the DMD. Specifically, Fig. S7(a) and Fig. S7(c) present the normalized confusion matrices for DD and QPMS with 800 μs effective dwell time for each pattern on the DMD, respectively. Comparing Fig. S7(b) and Fig. S7(d), it is evident that the training epochs required for achieving classification accuracy differ between DD and QPMS. For DD, a longer training duration of 300 epochs is necessary to achieve a classification accuracy of over 90%. Figure S8




**(a)** Normalized confusion matrix for QPMS.

**(b)** Normalized confusion matrix for DD.

**Figure S8.** Detection results with the effective dwell time of each pattern on DMD is 200 μs.

shows the results obtained for both QPMS and DD when the effective dwell time of each pattern on the DMD is 200 µs, with 50.4% classification accuracy for DD and 98% for QPMS. These results show that the classification accuracy at the low photon level is noticeably lower compared to the classical level detection (see Fig. S4). If we further decrease the effective dwell time on the DMD, the data from DD becomes challenging to be processed. Hence, when the effective dwell times on the DMD are less than 200 µs, only the data for QPMS can be processed reliably. Figure S9 shows the results obtained for QPMS when the effective dwell time of each pattern on the DMD is 60 µs with the classification accuracy as 97.2%. From these results, it is evident that the influence of noises on DD is more pronounced than on QPMS.

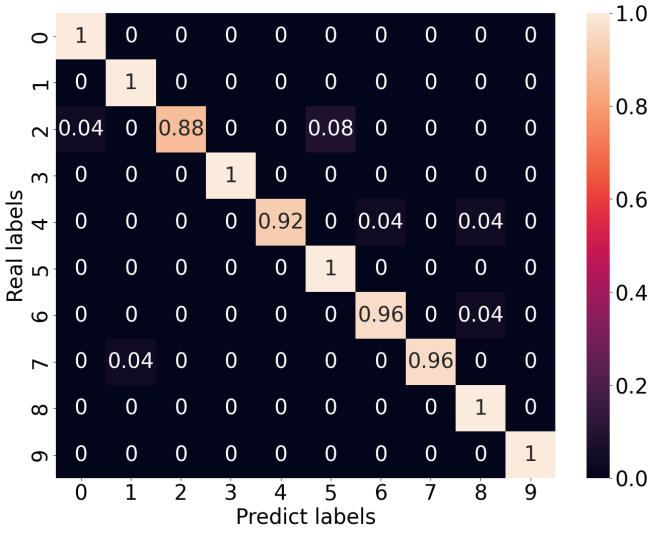
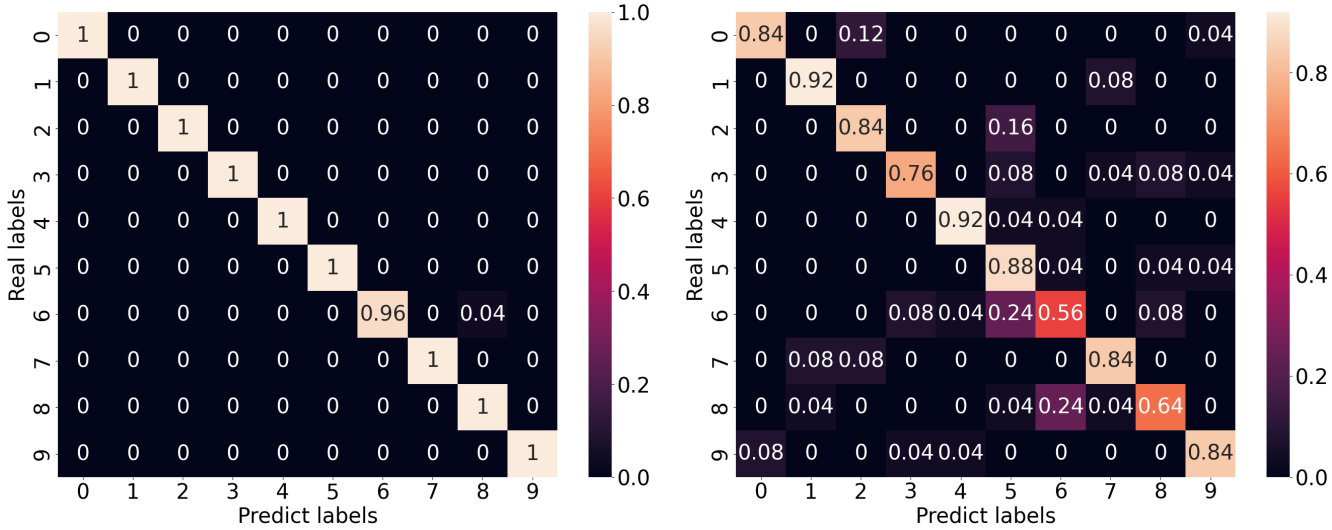



**Figure S9.** Normalized confusion matrix for QPMS with the effective dwell time of each pattern on DMD is 60 µs.

### 2.2 The influence of white noise

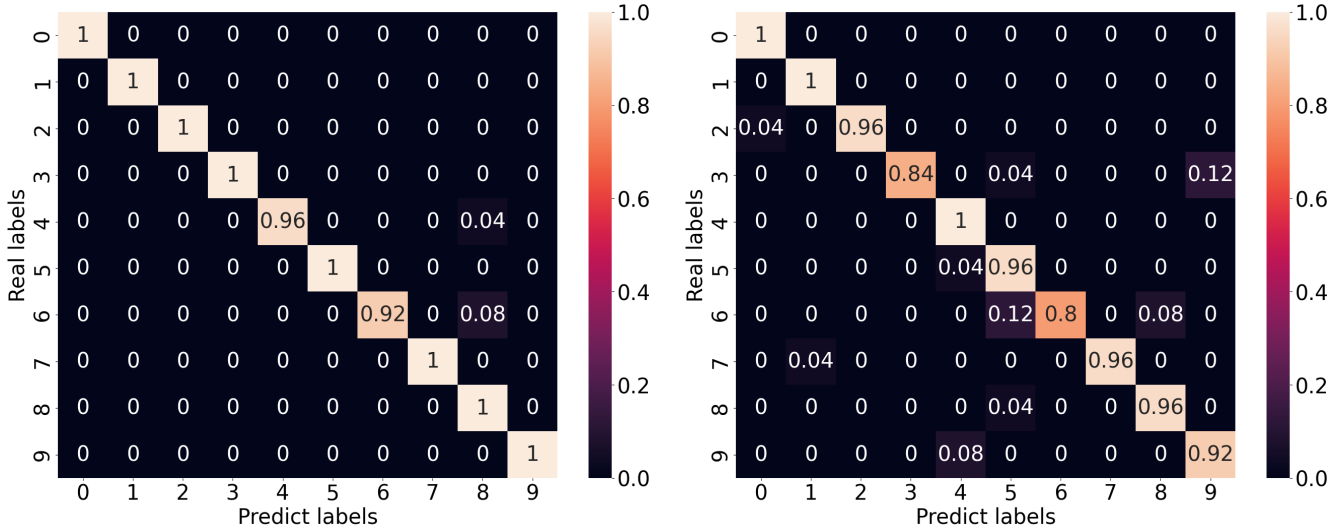
The data presented in this subsection corresponds to an experimental setup with an electrical controlled variable attenuator set at 3 V for QPMS and 2.7 V for DD. This setting introduces approximately 3 dB and 2.7 dB of attenuation compared to the previous subsection. As a result, the probe power sent into the free space prior to the DMD is attenuated to ∼6 nW for QPMS and ∼8.5 nW for DD. This subsection focuses on data obtained by considering different levels of white noise. The white noise is from amplifier spontaneous emission (ASE) before the photon counting.The effective dwell time on the DMD is set to 800 µs throughout this subsection. In Fig. S10, there is no ASE noise added in the setup, so the noise influence in this case is primarily due to the background noise. This is similar to Fig. S7, but with less probe power. From the table.S2, after ASE level less than -20 dB, the average photon counts per effective dwell time on DMD and the average photon count range over all photon-counting events increase obviously from QPMS, but when ASE is larger than -23 dB, the classification accuracy does not change notable. Figure S11 shows the QPMS under two different ASE noise levels. Because the probe photons for DD varied from 13 to 104.3 photon counts. When the level of ASE noise is introduced with a SNR of 0 dB, the noise photons collected by the InGaAs SPD make it challenging to distinguish the probe photons collected from different digit images in the dataset. As a result, the classification accuracy for DD becomes unreliable when the SNR level as 0 dB, with classification accuracy as 22.0% to 24.0%. On the other hand, for QPMS, we utilize the PPLN module, which enables the upconversion of probe photons that satisfy the phase-matching condition with the pump. This upconversion process effectively mitigates the influence of ASE noise, rendering it less susceptible to impacting the overall performance of QPMS. As we increase the power of ASE noise, it also impacts the data for QPMS (as shown in Fig. S11(b)). The average photon counts per photon counting event for QPMS without any ASE noise in the setup is from 2.9 to 171.3. However, when SNR is set to -23 dB for the ASE noise added into the setup, the photon count range changes to 25.1 to 143.6. This indicates a decrease in the range of photon counts, with a difference ($\Delta$) of 168.4 for no ASE noise and $\Delta$ of 118.5 for SNR = -23 dB. Consequently, this reduction in the photon count range makes it become harder to process the data. When we add more noise, SNR =-27 dB, the photon counts become larger, more noise photons mixed into the data, this makes the classification accuracy of QPMS decrease. If we add a higher level of ASE noise into the setup, the reliability of the data decreases further as more noise photons mixed into the data detected.




**(a)** Normalized confusion matrix for QPMS with the probe power as 6 nW.

**(b)** Normalized confusion matrix for DD with the probe power as 8.5 nW.

**Figure S10.** Results when no ASE noise added for both QPMS and DD.




**(a)** Normalized confusion matrix for SNR = 3 dB.

**(b)** Normalized confusion matrix for SNR = -27 dB.

**Figure S11.** QPMS result for different levels of ASE noise with probe power as 6 nW.

| ASE noise: SNR (dB) | Average photon counts per mask of QPMS | Average photon counts range over photon-counting events of QPMS | Average photon counts per mask of DD | Average photon count range over photon-counting events of DD |
|---|---|---|---|---|
| no | 871.8 | 2.9–171.3 | 391.5 | 13.0 – 104.3 |
| 3 | 837.0 | 2.7–164.6 | 558.7 | 45.1 – 104.9 |
| 0 | 784.2 | 2.6 – 154 | 705.7 | 67.0 – 111.8 |
| -3 | 786.1 | 2.5–157.2 | | |
| -7 | 727.2 | 2.6–143.6 | | |
| -10 | 706.2 | 2.5 – 137.5 | | |
| -13 | 707.2 | 2.5 – 137 | | |
| -17 | 706.3 | 2.8 – 132.6 | | |
| -20 | 771.3 | 3.6 – 143.3 | | |
| -23 | 1029.2 | 25.1 – 143.6 | | |
| -27 | 2246.0 | 208.3 – 454.3 | | |

**Table S2.** The average photon counts vary within a certain range for different levels of ASE noise with effective dwell time on DMD as 800 µs.