

Translation-Equivalence-Based Unsupervised Ghost Imaging

Shuai Mao¹

¹School of Information and Communications Engineering
Xi'an Jiaotong University
Xi'an 710049, China

Yuchen He^{2*}

²School of Electronic Science and Engineering
Xi'an Jiaotong University
Xi'an 710049, China
*e-mail: yuchenhe@xjtu.edu.cn

Jianming Yu²

²School of Electronic Science and Engineering
Xi'an Jiaotong University
Xi'an 710049, China

Yue Zhou²

²School of Electronic Science and Engineering
Xi'an Jiaotong University
Xi'an 710049, China

Chong Wang²

²School of Electronic Science and Engineering
Xi'an Jiaotong University
Xi'an 710049, China

Juan Chen¹

¹School of Information and Communications Engineering
Xi'an Jiaotong University
Xi'an 710049, China

Abstract—Ghost imaging (GI), as a burgeoning imaging technology, has gained a lot of attention in the past few years. To achieve better reconstructed image quality at low sampling rates, deep learning is introduced to GI. While supervised learning is widely used in GI, unsupervised learning also deserves to be studied in depth as it does not need real labeled images to train a network. Following this idea, we proposed a translation equivalence ghost imaging (TEGI) method only based on unlabeled bucket signals. A new loss function is designed to do unsupervised learning. It is demonstrated that our TEGI method can reconstruct target images with better quality than the differential ghost imaging method at low sampling rates.

Keywords—component; ghost imaging; deep learning; unsupervised learning; translation equivalence; loss function

I. INTRODUCTION

Ghost imaging (GI) has gained a lot of attention since it was first completed by Shi Yanhua and Pittman in 1995 [1]. Unlike traditional imaging techniques, GI uses second-order correlation of the light field to perform imaging. The imaging process of GI is based on the intensity fluctuation of the light field. As a result, GI has advantages such as turbulence-free, lensless and high resolution, which make it useful in filed like lidar, encryption, medical imaging, and imaging in high scattering or low illuminance environments [2–10].

Classical GI needs two light beam paths, one of which is used to record the reference wavefront information before the target. Targets do not need to show up on this path because information on targets will be collected by a bucket detector without spatial resolution in the other path. In 2008, computational ghost imaging (CGI) was proposed by Shapiro, in which light wavefront information is obtained by

calculation [11]. The reference beam path is no longer needed, making GI much easier to operate. However, to achieve high-quality results, many measurements are needed, which will last a long time. Differential ghost imaging (DGI) [12] is put forward in 2010 to achieve high-quality imaging at low sampling rates. Despite its easy operation, a major problem is that it still requires a relatively high sampling rate. In 2009, compressed sensing (CS) was introduced to GI [13]. Many image reconstruction algorithms adopting the ideas of CS were proposed in recent years [14–16]. Although GI with CS can achieve good results, its computational complexity is very high. In 2017, deep learning was brought into GI [17], where DNN is used to improve image quality after GI reconstruction. In addition to directly using low-quality reconstruction images as input to a deep neural network, another idea is to directly take measured bucket signals as input. Following these two basic ideas, many related works have been done in the last three years [18–31]. However, in their works, supervised learning is adopted, which means target images are needed as labels to train the network.

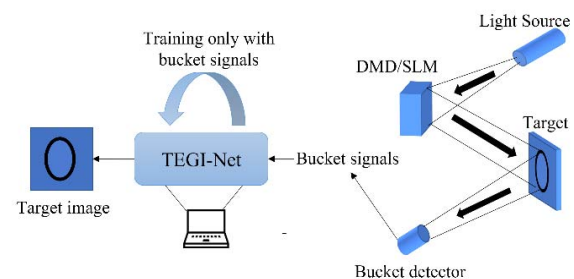


Figure 1. Translation Equivalence ghost imaging system.

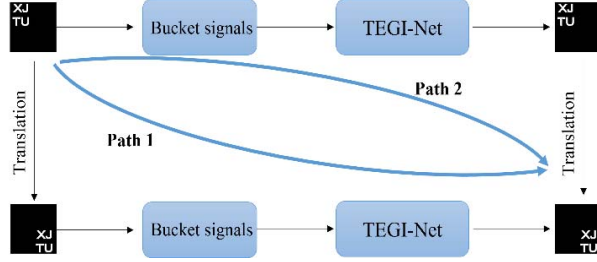


Figure 2. Translation Equivalence illustration

In this paper, we put forward a GI method based on unsupervised learning called translation equivalence ghost imaging (TEGI). In our method, only bucket signals are needed during both the training and testing phases. Translation equivalence property is used to quantify losses. A new loss function consisting of a measurement consistency term and a translation equivalence term is proposed to train the network. With our method, images will be constructed with higher quality at low sampling rates. Our corresponding TEGI system is shown in Figure 1.

II. MAIN WORK

A. Ghost imaging modeling

In GI framework, bucket signals are collected by a bucket detector without spatial resolution. One measurement can be modeled by (1):

$$y = \iint_{(i,j)} h(i,j) * x(i,j) \quad (1)$$

where y is the intensity of bucket signal, x is the characteristic function of target images in size of N^2 and h is a measurement speckle. To achieve imaging, a series of measurements will be taken. When all measurements are combined together, (1) can be changed into a vector form, the measurement process can be modeled by (2):

$$Y = H * X \quad (2)$$

where Y is a column vector representing bucket signal sequence, H is a matrix of size $m \times N^2$, in which m is the measurement times, $\frac{m}{N^2}$ is sampling rate (SR) and each line of H represents a measurement speckle.

B. Supervised learning

There are two basic learning paradigms in deep learning: one is supervised learning and the other is unsupervised learning [32]. In a supervised learning model, a network learns from a labeled dataset which provides criterion that can be used to evaluate network accuracy on the training data. In contrast, unsupervised models use unlabeled data, and the algorithm needs to extract features and patterns to understand the data. Equation (3) shows the most frequently used loss function in regression problems called the mean square error (MSE) loss function:

$$L_{MSE}(\theta) = \sum_{i=1}^{batch_size} \|x_i - f_{\theta}(y_i)\|_2 \quad (3)$$

When applied in GI, x_i represents the i^{th} real target image, y_i represents the i^{th} bucket. But in unsupervised learning, only y_i is available, (3) is changed into (4) in order to depict offsets from the true value.

$$L_{MC}(\theta) = \sum_{i=1}^{batch_size} \|y_i - Hf_{\theta}(y_i)\|_2 \quad (4)$$

So we call (4) measurement consistency loss. Measurement consistency should be zero if targets can be reconstructed perfectly by the network. GI is always performed with a sampling rate smaller than one to reduce the measurement time, which means m is smaller than N^2 . In this situation, matrix H has a non-trivial null space, in other words, there are other solutions to (5) besides the zero vector.

$$H * X = 0 \quad (5)$$

From this point, it can be found that any function in form of (6) can make measurement consistency loss zero, where H^+ is pseudo inverse of measurement matrix H and $\mu(y)$ is a solution to (5).

$$f_{\theta}(y) = H^+ y + \mu(y) \quad (6)$$

As a result, if only measurement consistency loss is used, the network is hard to converge to the right form because it does not learn any information about the null space.

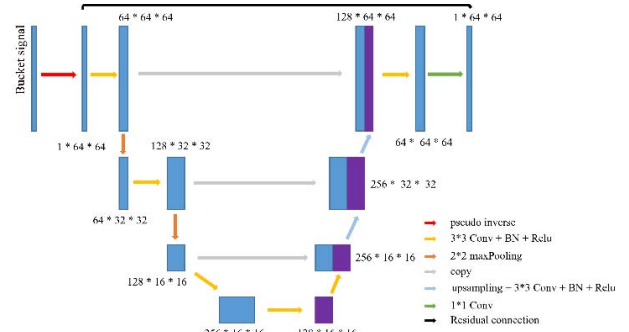


Figure 3. Network structure

C. Translation equivalence

As mentioned at the end of the previous section, other information is needed to train the network. In our method, image translation equivalence property is used to obtain extra information. The basic idea is illustrated in Figure 2: there are two paths to achieve image translation and image reconstruction and the results of taking both paths should be equivalent. In path 1, translation is applied to a image first, then bucket signals are collected, and finally use the network to reconstruct the image. In path 2, bucket signals are collected directly, after using network to reconstruct the image, finally apply translation to the reconstructed images. In other words,

image translation and image reconstruction are order-independent, which we call image translation equivalence property. This equivalence can be modeled by (7):

$$f_{\theta}(H * T(X)) = T(f_{\theta}(H * X)) \quad (7)$$

where T is a random translation transformation. Image translation equivalence property applies to all images. Thus it can behave as a prior information to help build our network.

$$L_{MC} = \sum_{i=1}^{batch_size} \|H * T(X) - H * f_{\theta}(H * T(X))\|_2 = \sum_{i=1}^{batch_size} \|H * T(X) - H * T(f_{\theta}(H * X))\|_2 = \sum_{i=1}^{batch_size} \|H * T(X - f_{\theta}(H * X))\|_2 \quad (9)$$

It can be found that (9) is equivalent to (2) as long as $H * T$ is full rank. Combined with measurement consistency loss, the final loss function is shown in (10),

$$L_{total} = L_{MC} + \alpha L_{TE} \quad (10)$$

where alpha is a weight coefficient to balance the measurement consistency loss and the translation equivalence loss.

III. DEMONSTRATION RESULTS

In the previous part, translation equivalence is proven to be effective and can be applied to any suitable networks f_{θ} . The network structure we use is shown in Figure 3, here we choose $f_{\theta} = H^+ G_{\theta}$. The pseudo-inverse of the measurement matrix H and a U-net are combined together to construct a mapping from bucket signals to target images. Not only this structure can make use of the information of measurement speckles, but also the convolutional neural network can better handle inputs as they are 2D images. The U-net is composed of many down-sampling and up-sampling blocks. In the left of U-net, down-sampling is applied to inputs. The down-sampling blocks consist of convolutional layer of 3*3 kernel, batch normalization layer, Relu activation layer and maxPooling layer. In the right of U-net, up-sampling is applied. Every up-sampling tensor will be concatenated with a tensor from down-sampling process. Residual connection is also applied to U-net, the output of U-net is the summation of input and network. The training flow chart is shown in Figure 4. In the training process, a target image x_1 is first reconstructed. The measurement consistency loss can be calculated according to (4), using the same measurement matrix to measure again and calculate the mean square error. To evaluate the translation equivalence loss, we first follow path 2 in Figure 2 to obtain the translated images x_2 , then we follow path 1 to achieve the translated images x_3 . The mean square error between x_2 and x_3 is the translation equivalence loss. And network parameters can be updated with these two losses. During our training process, the cifar-10 dataset [33] is used to train the network. To make $H * T$ full rank, every x_1 is applied 50 random translations. To balance measurement consistency loss and translation equivalence loss, weight coefficient α is set to 20. Bucket signals are measured first, and then Adam optimizer will be used to train the network only based on these bucket signals.

As shown in (8), a new translation equivalence loss is put forward to make this equivalence true during training.

$$L_{TE} = \sum_{i=1}^{batch_size} \|f_{\theta}(H * T(X)) - T(f_{\theta}(H * X))\|_2 \quad (8)$$

To further prove translation equivalence is impactful, some more mathematical deductions will be done. As shown in (9), if y is replaced with $H * T(X)$ in (4) and translation equivalence property is used, (4) can be changed into (9):

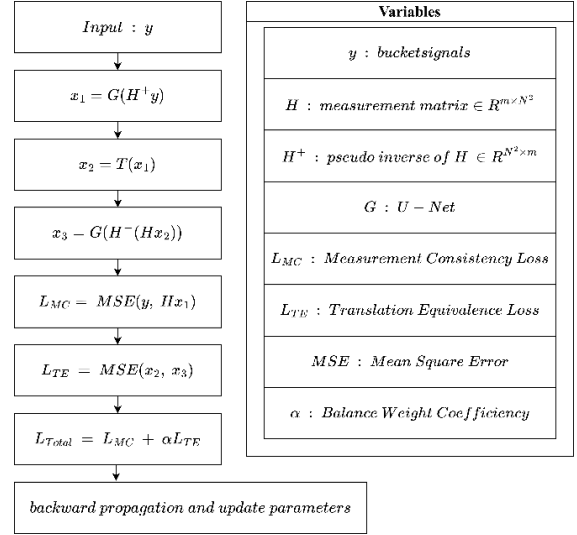


Figure 4. TEGI training flow chart.

In the first stage, our method is evaluated at SR of 10%. After the U-net is well trained with cifar-10 dataset, three letters and “XJTU” are used as target images to be reconstructed with DGI and our TEGI method. The reconstruction results are shown in Figure 5. Peak signal-to-noise ratio (PSNR) and structure similarity index measure (SSIM) are used to evaluate results. Corresponding calculation formulae are shown in (11) and (12),

$$PSNR = 10 * \log_{10} \left(\frac{255}{MSE} \right) \quad (11)$$

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (12)$$

where MSE is mean square error between result and target, μ is average value, σ^2 is variance, σ_{xy} is covariance, c_1 and c_2 are constants to make formula stable.

An overview of the PSNR and SSIM results is provided in Table. I. From Figure 5, it can be found that DGI results are completely drowned in background noise, but clear target results can be found in our proposed method. It can be observed from Table. I. that our proposed method is much

better with an average increment of 9.125 dB on PSNR and 0.122 on SSIM.

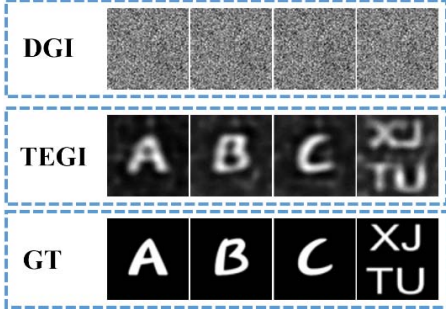


Figure 5. Ground truth of targets, DGI and TEGI results at SR = 10%

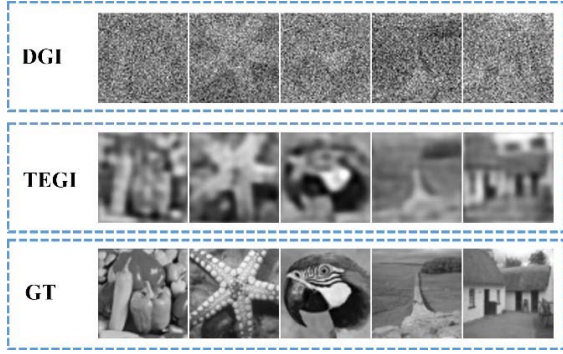


Figure 6. Ground truth of targets, DGI and TEGI results at SR = 30%

Further, we test our method with complex targets. Considering DGI is ineffective for relatively simple targets, we train the network at the SR of 30%. The setting is the same as before apart from the SR. The PSNR and SSIM results and corresponding reconstruction results are shown in Table II. and Figure 6. From Figure 6, although some image outlines can be found in DGI results, the images are still drowned in background noise. Our TEGI method can already achieve good results at SR=30%. As shown in Table II, the PSNR and SSIM increase by 9.922 and 0.512 for image “house” respectively. On average, the PSNR is 7.357 higher and the SSIM has raised by 0.451, which is a huge improvement.

TABLE I. PSNR AND SSIM RESULTS AT SR = 10%

	<i>A</i>	<i>B</i>	<i>C</i>	<i>XJTU</i>	<i>Average</i>
PSNR	16.833	15.109	16.344	11.859	
increment	5.754	5.851	5.936	6.106	
SSIM	0.121	0.125	0.106	0.157	
increment	0.003	0.007	0.005	0.005	
increment	0.118	0.118	0.101	0.152	0.122

TABLE II. PSNR AND SSIM RESULTS AT SR = 30%

	<i>pepper</i>	<i>starfish</i>	<i>parrot</i>	<i>path</i>	<i>house</i>	<i>Average</i>
PSNR	18.997	18.162	19.569	22.233	22.803	
increment	12.092	12.848	11.818	15.339	12.881	
SSIM	6.905	5.315	7.751	6.893	9.922	7.357
increment	0.631	0.497	0.628	0.621	0.672	
increment	0.129	0.214	0.123	0.176	0.154	
increment	0.502	0.283	0.505	0.445	0.518	0.451

IV. CONCLUSION

In this paper, we demonstrate that bucket signals can be used to train a network without real target images as labels. Referring to the MSE loss function, measurement consistency loss is put forward first. But it is not enough when the SR of GI is smaller than one. To obtain extra information, the translation equivalence loss function is designed. Our new loss function is used to train a U-net network through unsupervised learning. Demonstrations at low SRs are organized. It is clearly shown that with our method, images will be constructed with higher quality than DGI method, which can be proven with better PSNR and SSIM results.

REFERENCES

- [1] T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, ‘Optical imaging by means of two-photon quantum entanglement’, *Phys. Rev. A*, vol. 52, no. 5, 1995.
- [2] Y. H. He, A. X. Zhang, W. K. Yu, L. M. Chen, and L. A. Wu, ‘Energy-selective x-ray ghost imaging’, *Chin. Phys. Lett.*, vol. 37, no. 4, p. 044208, 2020.
- [3] H. Lin, W. Liu, S. Sun, and L. Du, ‘Influence of pulse characteristics on ghost imaging lidar system’, *Appl. Opt.*, vol. 60, no. 6, pp. 1623–1628, 2021.
- [4] X. Yang et al., ‘Underwater ghost imaging based on generative adversarial networks with high imaging quality’, *Opt. Express*, vol. 29, no. 18, pp. 28388–28405, 2021.
- [5] H. Zhang and D. Duan, ‘Turbulence-immune computational ghost imaging based on a multi-scale generative adversarial network’, *Opt. Express*, vol. 29, no. 26, pp. 43929–43937, 2021.
- [6] P. Zheng, Q. Tan, and H. Liu, ‘Inverse computational ghost imaging for image encryption’, *Opt. Express*, vol. 29, no. 14, pp. 21290–21299, 2021.
- [7] C. Xu, D. Li, K. Guo, Z. Yin, and Z. Guo, ‘Computational ghost imaging with key-patterns for image encryption’, *Opt. Commun.*, p. 129190, 2022.
- [8] P. Zheng, Z. Ye, J. Xiong, and H. Liu, ‘Computational ghost imaging encryption with a pattern compression from 3D to 0D’, *Opt. Express*, vol. 30, no. 12, pp. 21866–21875, 2022.
- [9] L. Lin, J. Cao, D. Zhou, H. Cui, and Q. Hao, ‘Ghost imaging through scattering medium by utilizing scattered light’, *Opt. Express*, vol. 30, no. 7, p. 11243, Mar. 2022, doi: 10.1364/OE.453403.
- [10] Z. Gao, X. Cheng, J. Yue, and Q. Hao, ‘Extendible ghost imaging with high reconstruction quality in strong scattering medium’, *Opt. Express*, vol. 30, no. 25, p. 45759, Dec. 2022, doi: 10.1364/OE.474579.
- [11] J. H. Shapiro, ‘Computational ghost imaging’, *Phys. Rev. A*, vol. 78, no. 6, p. 61802, 2008.
- [12] F. Ferri, D. Magatti, L. A. Lugiato, and A. Gatti, ‘Differential ghost imaging’, *Phys. Rev. Lett.*, vol. 104, no. 25, p. 253603, 2010.
- [13] O. Katz, Y. Bromberg, and Y. Silberberg, ‘Compressive ghost imaging’, *Appl. Phys. Lett.*, vol. 95, no. 13, p. 131110, 2009.
- [14] D. Yang, C. Chang, G. Wu, B. Luo, and L. Yin, ‘Compressive ghost imaging of the moving object using the low-order moments’, *Appl. Sci.*, vol. 10, no. 21, p. 7941, 2020.
- [15] H. Zhang and D. Duan, ‘Computational ghost imaging with compressed sensing based on a convolutional neural network’, *Chin. Opt. Lett.*, vol. 19, no. 10, p. 101101, 2021.
- [16] C. Zhang, J. Tang, J. Zhou, and S. Wei, ‘Singular value decomposition compressed ghost imaging’, *Appl. Phys. B*, vol. 128, no. 3, pp. 1–11, 2022.
- [17] M. Lyu et al., ‘Deep-learning-based ghost imaging’, *Sci. Rep.*, vol. 7, no. 1, pp. 1–6, 2017.

- [18] T. Bian, Y. Yi, J. Hu, Y. Zhang, Y. Wang, and L. Gao, 'A residual-based deep learning approach for ghost imaging', *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020.
- [19] H. Wu et al., 'Sub-Nyquist computational ghost imaging with deep learning', *Opt. Express*, vol. 28, no. 3, pp. 3846–3853, 2020.
- [20] H. Wu et al., 'Deep-learning denoising computational ghost imaging', *Opt. Lasers Eng.*, vol. 134, p. 106183, 2020.
- [21] S. Rizvi, J. Cao, K. Zhang, and Q. Hao, 'DeepGhost: real-time computational ghost imaging via deep learning', *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, 2020.
- [22] T. Bian, Y. Dai, J. Hu, Z. Zheng, and L. Gao, 'Ghost imaging based on asymmetric learning', *Appl. Opt.*, vol. 59, no. 30, pp. 9548–9552, 2020.
- [23] K. Komuro, T. Nomura, and G. Barbastathis, 'Deep ghost phase imaging', *Appl. Opt.*, vol. 59, no. 11, pp. 3376–3382, 2020.
- [24] R. Zhu et al., 'Ghost imaging based on Y-net: a dynamic coding and decoding approach', *Opt. Express*, vol. 28, no. 12, pp. 17556–17569, 2020.
- [25] C. Moodley, B. Sephton, V. Rodríguez-Fajardo, and A. Forbes, 'Deep learning early stopping for non-degenerate ghost imaging', *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021.
- [26] Z. Zhang, C. Wang, W. Gong, and D. Zhang, 'Ghost imaging of blurred object based on deep-learning', *Appl. Opt.*, vol. 60, no. 13, pp. 3732–3739, 2021.
- [27] S. Liu, X. Meng, Y. Yin, H. Wu, and W. Jiang, 'Computational ghost imaging based on an untrained neural network', *Opt. Lasers Eng.*, vol. 147, p. 106744, 2021.
- [28] H. Wu et al., 'Hybrid neural network-based adaptive computational ghost imaging', *Opt. Lasers Eng.*, vol. 140, p. 106529, 2021.
- [29] Y. He, S. Duan, Y. Yuan, H. Chen, J. Li, and Z. Xu, 'Semantic ghost imaging based on recurrent-neural-network', *Opt. Express*, vol. 30, no. 13, pp. 23475–23484, 2022.
- [30] Z. Yu et al., 'Color computational ghost imaging by deep learning based on simulation data training', *Appl. Opt.*, vol. 61, no. 4, pp. 1022–1029, 2022.
- [31] F. Wang et al., 'Far-field super-resolution ghost imaging with a deep neural network constraint', *Light Sci. Appl.*, vol. 11, no. 1, pp. 1–11, 2022.
- [32] D. Dhall, R. Kaur, and M. Juneja, 'Machine learning: a review of the algorithms and its applications', *Proc. ICRIC 2019*, pp. 47–63, 2020.
- [33] A. Krizhevsky, G. Hinton, and others, 'Learning multiple layers of features from tiny images', 2009.