

Decoding the Hidden: Direct Image Classification using Coded Aperture Imaging

Jocelyn Ornelas Munoz
Applied Mathematics Department
University of California, Merced
Merced, CA 95343 USA
jornelasmunoz@ucmerced.edu

Erica M. Rutter
Applied Mathematics Department
University of California, Merced
Merced, CA 95343 USA
erutter2@ucmerced.edu

Roummel F. Marcia
Applied Mathematics Department
University of California, Merced
Merced, CA 95343 USA
rmarcia@ucmerced.edu

Abstract—Coded aperture imaging has emerged as a solution to enhance light sensitivity and enable imaging in challenging conditions. However, the computational expense of image reconstruction poses limitations in processing efficiency. To address this, we propose a direct classification method using convolutional neural networks. By leveraging raw coded measurements, our approach eliminates the need for explicit image reconstruction, reducing computational overhead. We evaluate the effectiveness of this approach compared to traditional methods on the MNIST and CIFAR10 datasets. Our results demonstrate that direct image classification using raw coded measurements achieves comparable performance to traditional methods while reducing computational overhead and enabling real-time processing. These findings highlight the potential of machine learning in enhancing the decoding process and improving the overall performance of coded aperture imaging systems.

Index Terms—coded aperture, signal processing, classification, deep learning

I. INTRODUCTION

Digital image acquisition has become increasingly more common and prevalent in our everyday lives. As the technology used to record these images becomes more complex and sophisticated, the need to develop algorithms to process this information becomes equally important. Although the precision of recording devices increases, computational expense remains a significant drawback in imaging applications. These algorithms often demand extensive processing time and increased resource requirements like memory and storage. Real-time processing, crucial for immediate feedback or decision-making, may not be achievable with computationally expensive algorithms.

Coded aperture (CA) imaging originated from the need to enhance the amount of light reaching a detector in an imaging system while maintaining resolution. In particular, CA was developed to meet the needs of high-energy imaging where traditional mirror and lens-based systems were not feasible. This technology has found widespread applications in diverse fields such as astronomy, remote sensing, surveillance systems, and biomedical imaging [1]–[4]. Typically, the radiation from the source casts a shadow of an object on the binary aperture mask. This mask is composed of a pattern of openings

which allows a large fraction of photons to pass through to a position-sensitive detector, thereby encoding the spatial information contained in the source. Given a source signal image $\mathbf{S} \in \mathbb{R}^{n \times n}$ and an aperture mask $\mathbf{A} \in \{0, 1\}^{n \times n}$, a CA imaging system encodes the source image, yielding $\mathbf{D}$. The observed image $\mathbf{D}$ would be reconstructed to form the original source image $\mathbf{S}$ for further classification (see Figure 1). This is common practice in medical imaging [4]–[6].

To overcome the limitations of multi-stage processing and computational overhead in CA image classification, we propose classifying the raw encoded image $\mathbf{D}$ using a convolutional neural network (CNN). This eliminates the need for explicit image reconstruction. We evaluate the effectiveness of our proposed approach by comparing it to conventional methods that require image reconstruction before classification.

The paper is organized as follows: In Section II, we describe the problem statement and existing work. In Section III, we present our approaches for the data acquisition and deep learning architectures used. In section Section IV, we present the proposed approaches and describe our results and conclusion in Section V and Section VI respectively.

II. PROBLEM FORMULATION

Seminal work in CA imaging encompasses the development of Modified Uniformly Redundant Arrays (MURAs) [8]. MURAs offer improved decoding capabilities and enhanced image reconstruction. They are mathematically designed to increase the redundancy of the coded aperture pattern, allowing for better noise suppression, increased imaging resolution, and more accurate scene reconstruction from the measurements obtained by the detector array. In our investigation, we will specifically concentrate on the study of images encoded with MURA.

The CA imaging system can be expressed as:

$$\mathbf{D} = (\mathbf{S} * \mathbf{A}) + \mathbf{B},$$

where $\mathbf{A}$ is the coded aperture, $\mathbf{S}$ is the source signal, $\mathbf{B}$ is background noise, $\mathbf{D}$ is the observed image, and $*$ denotes the convolution operator (see [8]). Note, we utilize the convolutional theorem to perform convolution computations in this work [9]. An example of the CA encoding process is shown in Figure 2.

This work is supported by National Science Foundation grant DMS 1840265.

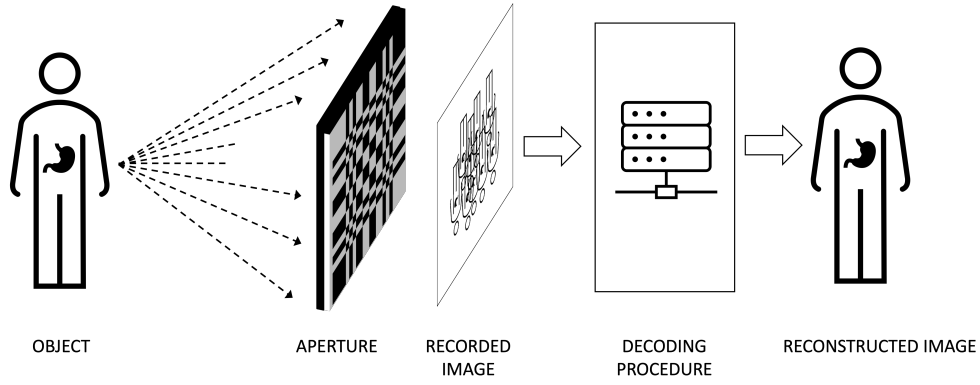


Fig. 1. Basic steps involved in coded aperture imaging. Figure is adapted from [7].

The coded aperture $\mathbf{A}$ is a binary, square pattern with prime integer side lengths which are designed so that if one observed $\mathbf{D}$, then $\mathbf{S}$ could be reconstructed as

$$\tilde{\mathbf{S}} = \mathbf{D} * \mathbf{G},$$

for some complementary mask pattern $\mathbf{G}$ or using an appropriate decoding algorithm. The MURA mask patterns and their complementary decoder are designed such that $\mathbf{A} * \mathbf{G} \approx \delta$, where δ is the Kronecker-delta function, providing successful reconstructions.

However, in the context of inference tasks like classification, segmentation, or detection, the conventional two-step process involving image reconstruction followed by task-solving is computationally inefficient [10].

Related Work. In conventional coded aperture imaging, the decoding process for reconstructing the scene content from captured images has relied on sophisticated mathematical algorithms and computational techniques such as Maximum Entropy Method (MEM) [11], [12], wavelet-based algorithms [13], [14], and convolutional neural networks [15].

Traditional methods for classifying coded aperture images typically involve a two-step process - 1. Image reconstruction followed by 2. Subsequent image classification [7]. This process requires computationally expensive image reconstructions before applying classification algorithms. Such a sequential approach can be time-consuming and may introduce potential errors or artifacts during the reconstruction process.

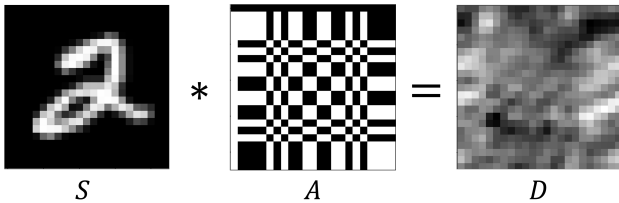


Fig. 2. Example of coded observation. The source signal ($\mathbf{S}$), the aperture ($\mathbf{A}$) and the resulting coded aperture image ($\mathbf{D}$).

The decoding mask pattern plays a crucial role in image reconstruction in CA imaging systems. In particular, the decoder array $\mathbf{G}$ must be chosen in a way such that $\mathbf{A} * \mathbf{G} \approx \delta$ [7].

With the emergence of adversarial attacks, it is possible for an attacker to attempt to manipulate the decoding mask pattern by introducing subtle modifications [16], [17]. These modifications to the decoding mask could be carefully designed to deceive a reconstruction algorithm and generate a reconstructed image that misrepresents the original scene and thus affects classification. Therefore, it is crucial to explore direct classification methods in order to comprehend how vulnerabilities in CA can be addressed and overcome.

To the best of our knowledge, the specific topic of direct image classification from CA measurements has not been extensively investigated in the existing literature. Similar work has been done in the area of compressed sensing [3], [18]–[20]. However, the focus of this paper is on CA imaging. While some papers have explored deep learning and optimization techniques to optimize the sensing matrix of a single pixel camera [21] and classify spectral imaging from optimal coded apertures [3], the direct classification of coded images remains an under-explored area of research.

III. PROPOSED METHOD

In this section, we introduce two methods for classifying coded images. In addition, we also describe the architecture used in the next section. The two proposed methods begin with the CA image $\mathbf{D}$. However, Method II reconstructs the coded images before classification.

Method I: We directly classify raw coded measurements $\mathbf{D}$.

Method II: We reconstruct the images before classification. In practice, the decoder may be unknown so we decode our coded observations using correlation analysis [7], [8]. For the purpose of this study, we have access to the true decoder, thus we use the exact decoder for correlation analysis. Note, there are alternative methods for image reconstruction such as [11] that were not considered due to high computational cost.

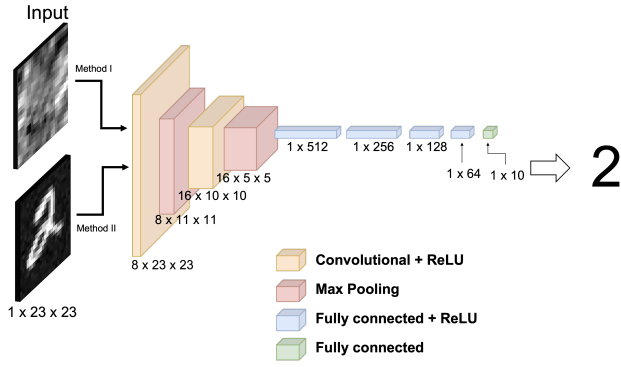


Fig. 3. Convolutional Neural Network (CNN) for classifying coded aperture images. The input to the network is either a single channel grayscale reconstructed image or single channel coded image.

Architecture. We employ a convolutional neural network (CNN) that takes three-dimensional arrays as inputs. In this context, the arrays represent two-dimensional images, with the first dimension denoting the number of channels in each image. The neural network, illustrated in Figure 3, comprises two convolutional layers, each followed by a maxpool layer, and concludes with four fully connected layers. The Rectified Linear Unit (ReLU) activation function [22] is applied after each layer.

IV. NUMERICAL EXPERIMENTS

We evaluate our method on two classification benchmark datasets - 1. MNIST and 2. CIFAR10.

Datasets. The datasets used in our experiments include 1. MNIST, which is comprised of 70,000 images of handwritten digits from 0 to 9 [23] and (2) grayscale CIFAR10 dataset, which contains 60,000 images, is comprised of 10 classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck [24]. The datasets were partitioned 80% as training, 10% for validation, and 10% for testing. The encoded data $\mathbf{D}$ is generated by resizing the grayscale images to 23×23 and applying the encoding process outlined in Section II. Zero-mean additive Gaussian noise was added to the encoded data to have signal-to-noise ratios (SNR) in decibels (dB) of 1dB, 5dB, 10dB, and noiseless.

As a baseline, the classification network was initially trained on the original MNIST images as well as the original CIFAR10 grayscale images resized to a dimension of 23×23 . The accuracy of the original MNIST model was measured at 99.87%, while the original grayscale CIFAR10 model achieved an accuracy of 88.18%. In order to compare with Method II, the “Original” model will be employed to classify the reconstructed data.

For all experiments, the training inputs were divided into batches, each consisting of 100 images and label pairs. During training, the predicted label was compared to the target label using the cross-entropy loss function [25]. To update the weights at each iteration, we employed the Adam optimizer

[26] with a learning rate of 0.001. The network was trained for a total of 50 epochs.

V. RESULTS

We present results from our numerical experiments to compare the accuracy between Methods I and II where the inputs for Method I are coded images and the input for Method II are reconstructed images. All cases consider data with varying noise levels: noiseless, 10dB, 5dB, and 1dB. To ensure a fair comparison between the two methods, we use grayscale data.

Experiment 1: MNIST Dataset. As a preliminary step, we applied the dimensionality reduction technique t-SNE [27] to explore the relationships and clusters within the encoded MNIST data. Similar to the original MNIST data (see [27]–[29]), the visualization of the encoded data in low-dimensional space are relatively well-separated, suggesting that distinct features in the coded data can facilitate classification.

Table I presents the accuracies achieved by a single model trained on a specific noise level and subsequently tested on all levels of noise. From the classification accuracies, we see that training on noisy data tends to do best when testing on all noise levels whereas the classification accuracy of a model trained on noiseless encoded data decreases as noise increases.

Overall, the classification accuracy of models trained with encoded data do quite well, with the lowest accuracy being 82.73%. This shows that there is potential benefit to perform a direct classification especially when the reconstruction is unknown and computationally expensive to perform.

Experiment 2: CIFAR10 Dataset. The t-SNE visualization of both the original and encoded grayscale CIFAR10 data did not exhibit distinct separability, and as a result, it has been omitted from the paper. The absence of distinct separability implies that the classification problem may pose a greater challenge.

Table II presents the classification accuracy for Methods I and II, considering training on various noise levels and testing across all noise levels. It is important to note that the accuracy

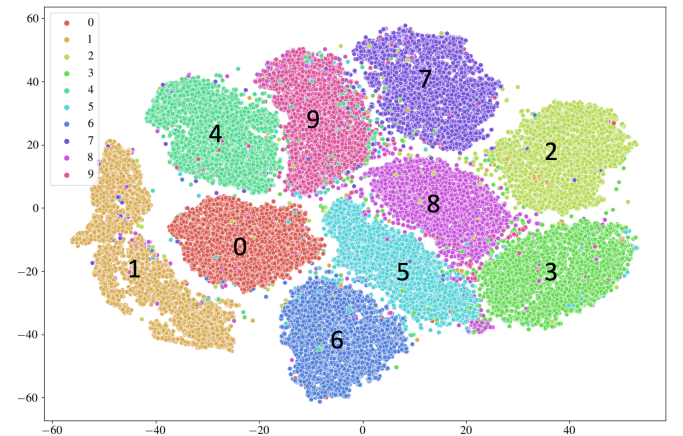


Fig. 4. t-Distributed Stochastic Neighbor Embedding Visualization of encoded MNIST Dataset. Each cluster has been labeled with the corresponding digit.

TABLE I

MNIST DATASET CLASSIFICATION ACCURACY USING ENCODED DATA (METHOD I) AND RECONSTRUCTED DATA (METHOD II). ROWS INDICATE THE TRAINING NOISE LEVEL, AND COLUMNS INDICATE THE TESTING NOISE LEVEL.

Training data	Testing Data							
	Noiseless		10dB		5dB		1dB	
	I	II	I	II	I	II	I	II
Noiseless	97.53	99.29	96.53	98.90	93.81	97.14	82.73	88.80
10dB	97.94	99.09	97.61	99.03	95.93	98.60	88.17	95.77
5dB	97.77	99.09	97.44	99.14	96.44	98.93	91.91	97.17
1dB	96.70	98.80	96.54	98.84	95.91	98.63	93.86	97.94
Original	-	96.36	-	86.86	-	64.04	-	36.90

TABLE II

GRAYSCALE CIFAR10 DATASET CLASSIFICATION ACCURACY USING ENCODED DATA (METHOD I) AND RECONSTRUCTED DATA (METHOD II). ROWS INDICATE THE TRAINING NOISE LEVEL, AND COLUMNS INDICATE THE TESTING NOISE LEVEL.

Training data	Testing Data							
	Noiseless		10dB		5dB		1dB	
	I	II	I	II	I	II	I	II
Noiseless	40.30	55.78	27.72	36.18	15.38	17.48	11.93	11.65
10dB	37.97	50.25	36.55	48.78	33.33	38.85	26.37	21.77
5dB	36.33	44.00	35.97	43.25	34.22	40.98	30.65	35.52
1dB	33.25	40.38	33.07	40.17	33.05	38.67	31.75	34.87
Original	-	54.43	-	34.75	-	17.75	-	11.90

for the CIFAR10 dataset is inherently lower due to training with grayscale images instead of color images. Similar to Experiment 1, training on noisy data tends to yield the best results when testing across all noise levels.

Notably, there is a larger performance accuracy gap between Methods I and II. However, it is worth mentioning that no reconstruction was performed for Method I. In instances where the model was tested on higher noise levels, there were occasions where Method I outperformed Method II. This finding suggests that as image quality decreases, direct classification offers greater benefits.

VI. CONCLUSIONS

In this paper, we implemented two techniques for classifying images within a coded aperture model. The first method, referred to as Method I, is a direct CNN-based classification approach that maps the raw noisy encoded image directly to its

corresponding label. The second method, Method II, is used as comparison with Method I and follows conventional coded aperture techniques by first reconstructing the observed image before performing classification.

Numerical experiments reveal that while Method II often yields better classification results compared to using encoded images (Method I), it is not directly applicable when the decoder $\mathbf{G}$ is unavailable.

Method I, on the other hand, provides reasonable classification results without the explicit need for reconstructing the encoded image or knowledge of the encoding and decoding arrays. These findings suggest that employing direct classification for coded aperture images has the potential to achieve higher classification accuracy, particularly when combined with a more advanced deep learning architecture. The source code is available at github.com/jornelasmunoz/coded-aperture.

REFERENCES

- [1] J. Chen, Y. Wang, and H. Wu, "A coded aperture compressive imaging array and its visual detection and tracking algorithms for surveillance systems," *Sensors (Basel, Switzerland)*, vol. 12, pp. 14397–14415, 2012.
- [2] E. Caroli, J. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino, "Coded aperture imaging in x- and γ -ray astronomy," *Space Science Reviews*, vol. 45, pp. 349–403, 1987.
- [3] A. Ramirez, H. Arguello, G. R. Arce, and B. M. Sadler, "Spectral image classification from optimal coded-aperture compressive measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3299–3309, 2013.
- [4] R. Accorsi, F. Gasparini, and R. C. Lanza, "Optimal coded aperture patterns for improved snr in nuclear medicine imaging," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 474, no. 3, pp. 273–284, 2001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168900201013262>
- [5] S. Meikle, R. Fulton, S. Eberl, M. Dahlbom, K.-P. Wong, and M. Fulham, "An investigation of coded aperture imaging for small animal spect," *IEEE Transactions on Nuclear Science*, vol. 48, no. 3, pp. 816–821, 2001.
- [6] R. Accorsi, F. Gasparini, and R. Lanza, "A coded aperture for high-resolution nuclear medicine planar imaging with a conventional anger camera: experimental results," *IEEE Transactions on Nuclear Science*, vol. 48, no. 6, pp. 2411–2417, 2001.
- [7] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.*, vol. 17, no. 3, pp. 337–347, Feb 1978. [Online]. Available: <https://opg.optica.org/ao/abstract.cfm?URI=ao-17-3-337>
- [8] S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Applied optics*, vol. 28, no. 20, pp. 4344–4352, 1989.
- [9] J. M. Blackledge, "Chapter 2 - 2d fourier theory," in *Digital Image Processing*, ser. Woodhead Publishing Series in Electronic and Optical Materials, J. M. Blackledge, Ed. Woodhead Publishing, 2005, pp. 30–49.
- [10] Z. W. Wang, V. Vineet, F. Pittaluga, S. N. Sinha, O. Cossairt, and S. Bing Kang, "Privacy-preserving action recognition using coded aperture videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
- [11] T. Ponman, "Maximum entropy methods," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 1, pp. 72–76, 1984, proceedings of the International Workshop on X- and γ -Ray Imaging Techniques. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0167508784901820>
- [12] R. Willingale, M. Sims, and M. Turner, "Advanced deconvolution techniques for coded aperture imaging," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 1, pp. 60–66, 1984.
- [13] R. F. Marcia and R. M. Willett, "Compressive coded aperture super-resolution image reconstruction," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 833–836.
- [14] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of selected topics in signal processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [15] R. Zhang, P. Gong, X. Tang, P. Wang, C. Zhou, X. Zhu, L. Gao, D. Liang, and Z. Wang, "Reconstruction method for gamma-ray coded-aperture imaging based on convolutional neural network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 934, pp. 41–51, 2019.
- [16] S. Schrodi, T. Saikia, and T. Brox, "Towards understanding adversarial robustness of optical flow networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 8916–8924.
- [17] A. Boloor, T. Wu, P. Naughton, A. Chakrabarti, X. Zhang, and Y. Vorobeychik, "Can optical trojans assist adversarial perturbations?" in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 122–131.
- [18] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.
- [19] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [20] R. Calderbank and S. Jafarpour, "Finding needles in compressed haystacks," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 3441–3444.
- [21] J. Bacca, L. Galvis, and H. Arguello, "Coupled deep learning coded aperture design for compressive image classification," *Opt. Express*, vol. 28, no. 6, pp. 8528–8540, Mar 2020. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-28-6-8528>
- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [23] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
- [24] A. Krizhevsky, V. Nair, and G. Hinton, "Cifar-10 (canadian institute for advanced research)." [Online]. Available: <http://www.cs.toronto.edu/~kriz/cifar.html>
- [25] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.
- [26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [27] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [28] Y. Pei and L. Ye, "Cluster analysis of mnist data set," in *Journal of Physics: Conference Series*, vol. 2181, no. 1. IOP Publishing, 2022, p. 012035.
- [29] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 478–487. [Online]. Available: <https://proceedings.mlr.press/v48/xieb16.html>