Automatic X-ray Scattering Image Annotation via Double-View Fourier-Bessel Convolutional Networks

Ziqiao Guan[1]
ziguan@cs.stonybrook.edu
Hong Qin[1]
qin@cs.stonybrook.edu
Kevin Yager[2]
kyager@bnl.gov
Youngwoo Choo[3]
youngwoo.choo@yale.edu
Dantong Yu[4]
dtyu@njit.edu

[1] Stony Brook University
Stony Brook, NY, USA
[2] Brookhaven National Laboratory
Upton, NY, USA
[3] Yale University
New Haven, CT, USA
[4] New Jersey Institute of Technology
Newark, NJ, USA

Abstract

X-ray scattering is a key technique towards material analysis and discovery. Modern x-ray facilities are producing x-ray scattering images at such an unprecedented rate that machine aided intelligent analysis is required for scientific discovery. This paper articulates a novel physics-aware image feature transform, Fourier-Bessel transform (FBT), in conjunction with deep representation learning, to tackle the problem of annotating x-ray scattering images with a diverse label set of physics characteristics. We devise a novel joint inference model, Double-View Fourier-Bessel Convolutional Neural Network (DVFB-CNN) to integrate feature learning in both polar frequency and image domains. For polar frequency analysis, we develop an FBT estimation algorithm for partially observed x-ray images, and train a dedicated CNN to extract structural information from FBT. We demonstrate that our deep Fourier-Bessel features well complement standard convolutional features, and the joint network (i.e., DVFB-CNN) improves mean average precision by 13% in multilabel annotation. We also conduct transfer learning on real experimental datasets to further confirm that our joint model is well generalizable.

1 Introduction and Motivation

X-ray scattering is an imaging technique for probing physical structure of materials at molecular and nano-scale. It is used for various applications such as determining protein conformations. The technique consists of shining a bright, collimated x-ray beam through a material of interest; coherent interference between x-rays scattered from internal structures casts a complex far-field pattern on an area detector. Material scientists established physical models for scattering patterns [27] so that we can deduce information about the size, orientation, and packing of atoms, molecules, and nano-domains [25][3] from image analysis.

© 2018. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

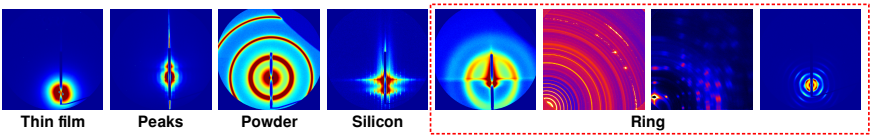


Figure 1: Examples of x-ray scattering images. Images are shown with false color for visualization purposes; source images only have one single intensity channel. L1-4: x-ray images have a set of attributes covering visual, physical, and sample properties. Label shown at bottom. R1-4: Images with the same attributes can have drastic style differences and/or other structures, resulting in very different appearances. All 4 images have the attribute 'ring'.

Modern x-ray detectors are capable of generating 50,000 to 1,000,000 images/day (i.e., 1-4 TB/day). The enormous amount of data makes it intractable to perform manual image analysis. This analysis bottleneck is an even bigger problem in scientific imaging community with all the high throughput scattering, full-field imaging and spectroscopy facilities. To expedite material discovery research with ever-growing scientific datasets, we have to automate data analysis workflows with new data mining and machine learning algorithms.

Recently, deep learning methods [15] have achieved great success in computer vision applications. Multi-layer "deep" learning mechanisms have proven to be of critical importance towards understanding high dimensional data and semantic concepts. In this work, we attempt to introduce deep learning techniques to resolve the aforementioned bottleneck. We take on a specific task, i.e., multi-label annotation of x-ray scattering images to study the effective "deep" practice for scientific data. We identify two major challenges of scientific datasets learning: label scarcity and pattern complexity, and articulate two key strategies: synthetic data, and physics-aware data transform — Fourier-Bessel Transform (FBT) — to address these challenges. FBT produces an energy representation of images as radial and angular frequencies and effectively represents image structure. On top of this novel transform, we build a double-view convolutional neural network (CNN) to learn features from both the original image and its FBT. We perform end-to-end inference and transfer learning experiments and demonstrate that our model consistently outperforms a standard CNN.

Our main contributions are in the following three perspectives:

- We pioneer a feasible deep Fourier-Bessel feature learning method that involves Fourier-Bessel coefficient estimation and deep CNN learning, demonstrating the effectiveness of FBT as a powerful physics-aware feature transform. To the best of our knowledge, this is the first attempt to articulate a deep Fourier-Bessel feature that is compact, informative, and discriminative.

- We articulate a Double-View Fourier-Bessel Convolutional Neural Network (DVFB-CNN) in details. Our double-view framework aims to integrate image CNN and Fourier-Bessel coefficient CNN for image annotation and exploit the power of ensemble learning.

- Our work is a proof-of-concept towards physics-awareness. Built upon our ongoing success of Fourier-Bessel learning, it is possible to enrich physics understanding during large-scale intelligent data analysis, which conduces to more success in multi-view learning with more powerful physics-aware transforms.

2 Related Works

X-ray Scattering Image Analysis is a topic of interest in both the x-ray and computer vision communities. [29] used spectral clustering for unsupervised clustering of images, and exploited symmetry analysis in the associated reconstruction problem [8][19], while [10] used diffusion-based clustering. These clustering methods do not reflect semantic labels and are not suitable for multi-label problems where attributes are not mutually exclusive. [13] proposed to perform classification of image attributes using off-the-shelf hand-crafted image descriptors. [23] first applied deep learning to scattering image recognition. A similar application in scientific imaging is [28] using CNN to classify microscopic images.

Fourier Transform in Polar and Spherical Coordinates is derived from differential equations [24] as a basis transform. In literature, there have been some attempts to use it for feature representations. [30] proposed to use Fourier-Bessel transform (FBT) as polar frequency descriptors for face recognition. Yet it did not show much empirical benefit over Discrete Cosine Transform (DCT). In spherical coordinates, Spherical Harmonic Transform is used in shape analysis [4][12] and surface registration [9].

Convolutional Neural Network (CNN) is a biological-inspired multi-layer trainable structure. It was first developed for hand-written digits recognition [16], and multi-layer deep CNNs significantly outperformed handcrafted descriptor methods [14]. With its high capacity of representation learning, CNN is now applied to increasingly diverse vision tasks, to name a few, object detection [6], image segmentation [7], video classification [11] and 3D shape classification [26].

3 Background, Challenges, and Strategies

Suppose a probing x-ray beam is shone through a sample with realspace density distribution $\rho(\mathbf{r})$. The resulting x-ray interference gives rise to a reciprocal-space scattering intensity of

$$I(\mathbf{q}) = \left| \int_V \rho(\mathbf{r}) \exp(i\mathbf{q}\cdot\mathbf{r})\, \mathrm{d}V \right|^2, \qquad (1)$$

The sample's reciprocal-space is centrosymmetric with scattering patterns centered about the direct beam. The symmetry of the scattering experiment is thus well-described by polar coordinates, $I(\mathbf{q}) \triangleq I(q,\phi)$.

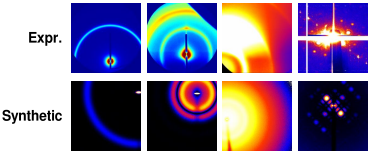


Figure 2: Examples of experimental images and synthetic images.

We now try to predict a set of non-mutually-exclusive image attributes. This is called *multi-label learning* as one sample can bear more than one attibute from the label set. Our label set represents characteristics of various aspects, e.g., 'halo' or 'ring' (visual), 'isotropic' or '6-fold symmetric' (style variations), 'powder' or 'polycrystalline' (material structure). Figure 1 shows example images of 'ring' with distinct style variations.

For x-ray scattering images, and scientific data in general, there are two major difficulties with learning CNNs. The first one is **label scarcity**. The amount of labeled data is usually not enough to train CNN to fit. This is because scientific dataset requires domain experts to label; also it takes N times longer for an N-label dataset. The second one is **pattern complexity**. Unlike natural images where regions of interest are usually local, visual patterns in x-ray

scattering images are often diffused and overlapping, which do not work well with the local receptive fields of CNN units. We propose two strategies to overcome these difficulties: transfer learning with synthetic data and Fourier-Bessel Transform (FBT).

3.1 Transfer Learning with Synthetic Dataset

We choose to *train the CNN with synthetic data* [23], and *do transfer learning to annotate real datasets*. This transfer learning scheme is inspired and proven effective by [23]. Simulation software can not only generate massive automatically labeled data, which is necessary to resolve label scarcity, but also model x-ray imagery accurately with parametric and physics-based methods [18][27][20], so the generated dataset is realistic and good for learning (see Figure 2). Various works have shown that proper simulations transfer well in deep learning [8][17].

3.2 Fourier-Bessel Transform

We use Fourier-Bessel Transform (FBT) as a feature transform to better separate features. FBT is a polar coordinate based transform, which is good for exploiting image centrosymmetry; it is a Fourier-like decomposition, which is good for preserving image information and thus physics. The radial/angular basis of FBT is derived from the eigenfunctions $\Psi(r,\varphi)$ of Laplacian in polar coordinates:

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}. \tag{2}$$

With separation of variables $\Psi(r,\varphi) = R(r)\Phi(\varphi)$, we can solve the differential equations to compute the basis with proper boundary conditions [24].

For representing 2D images, we are particularly interested in the bounded function space, i.e. complex functions $f(r,\varphi)$ with support $r < a$. Due to orthogonality, the normalized basis functions are:

$$\Phi_m(\varphi) = \exp(im\varphi) \Big/ \sqrt{2\pi}\,, \tag{3}$$

$$R_{nm}(r) = J_m\left(x_{mn}\cdot\frac{r}{a}\right) \Big/ \sqrt{N_n^{(m)}}\,, \tag{4}$$

where $m \in \mathbb{Z}$, $n \in \mathbb{Z}^+$, $N_n^{(m)} = a^2 J_{m+1}^2(x_{mn})/2$, J_m is the m-th order Bessel function, and $\{x_{mn}\}_n$ are the roots of $J_m(x)$, and FBT and the inverse transform are:

$$P_{nm} = \int_0^a \int_0^{2\pi} f(r,\varphi)\Psi_{nm}^*(r,\varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi, \tag{5}$$

$$f(r,\varphi) = \sum_{n=1}^{\infty}\sum_{m=-\infty}^{\infty} P_{nm}\Psi_{nm}(r,\varphi). \tag{6}$$

Examples of F-B basis functions are shown in Figure 3.

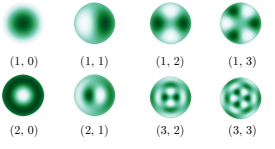


Figure 3: Fourier-Bessel basis functions on a bounded disk $r < a$ with zero-value boundary condition. The subscript (n,m) refers to the radial and angular indices (3)(4). Only real parts of the basis shown.

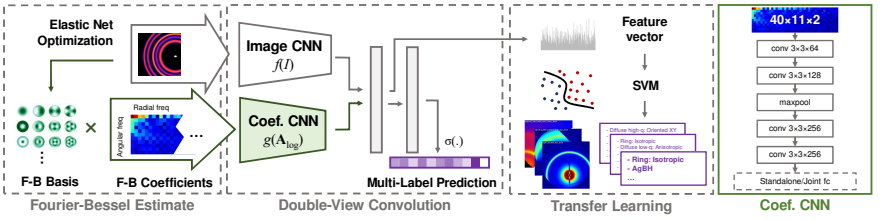


Figure 4: Pipeline of DVFB-CNN for multi-label x-ray scattering image annotation.

ALGORITHM 1: Fourier-Bessel Estimate.

Input: Input image $\mathbf{I} \in \mathbb{C}^{h\times w}$, precomputed basis images $\mathbf{\Psi} \in \mathbb{C}^{B\times H\times W}$, beam center position $(x,y) \in \mathbb{R}^2$
Output: Fourier-Bessel coefficients $\mathbf{A}$ of the original image

1 Crop the basis $\mathbf{\Psi}_c = \mathbf{P\Psi} = \mathbf{\Psi}(:,(H/2-y):(H/2+h-y),(W/2-x):(W/2+w-x))$;
2 Apply disk projection (8) $\mathbf{\Psi}' = \mathbf{P}_D\mathbf{\Psi}_c$, $\mathbf{I}' = \mathbf{P}_D\mathbf{I}$;
3 Reshape $\mathbf{\Psi}'$ to $\mathbb{C}^{B\times hw}$, $\mathbf{I}'$ to $\mathbb{C}^{hw}$;
4 $\tilde{\mathbf{\Psi}}' = \begin{bmatrix} \mathrm{Re}(\mathbf{\Psi}') & -\mathrm{Im}(\mathbf{\Psi}') \\ \mathrm{Im}(\mathbf{\Psi}') & \mathrm{Re}(\mathbf{\Psi}') \end{bmatrix}$, $\tilde{\mathbf{I}}' = \begin{bmatrix} \mathrm{Re}(\mathbf{I}') \\ \mathrm{Im}(\mathbf{I}') \end{bmatrix}$;
5 Solve $\min_{\tilde{\mathbf{A}}} \|\tilde{\mathbf{\Psi}}'\tilde{\mathbf{A}} - \tilde{\mathbf{I}}'\|_2^2 + \lambda\left[\alpha\|\tilde{\mathbf{A}}\|_1 + (1-\alpha)\|\tilde{\mathbf{A}}\|_2\right]$;
6 $\mathbf{A} = \tilde{\mathbf{A}}(:B) + \tilde{\mathbf{A}}(B:)i$;

4 Double-View Fourier-Bessel Convolutional Neural Network

In this section we describe our joint learning model, Double-View Fourier-Bessel Convolutional Neural Network (DVFB-CNN). DVFB-CNN is a network that jointly encodes both the original image and its FBT with separate convolutional sub-networks. It can be run end-to-end for training and prediction; it can also produce deep Fourier-Bessel features for generic, robust and physics-aware transfer learning (Figure 4).

4.1 Fourier-Bessel Estimate of Partial Images

As the observed image has a natural bounded support of the detector range, we model it in the function space defined on $\{r < a\}$ about the direct beam. However, a direct FBT is not feasible, as the image center is not necessarily aligned with direct beam position, and the disk $r < a$ may be cropped.

Consider FBT in discretized pixel space with a cropped basis. Since the image is the inner product of the F-B basis and the coefficients, we can solve a linear system to compute the coefficients. With a pixel mask applied, this can be formulated as the optimization:

$$\min_{\mathbf{A}} \|\mathbf{P\Psi A} - \mathbf{I}\|_F^2 + \lambda R(\mathbf{A}), \tag{7}$$

where $\mathbf{P}$ is the mask, $\mathbf{\Psi}$ is the discrete Fourier-Bessel basis, $\mathbf{A}$ is the basis coefficients, $\mathbf{I}$ is the image, λ is a regularization parameter and $R(\mathbf{A})$ is a regularization term.

Now given an image with center (x,y), we first crop the basis to image range and apply a disk projection to eliminate the outer rings clipped by the boundary of the detector window, to alleviate boundary effect:

$$\mathbf{P}_D = \mathbf{1}_D, \quad D = \{\mathbf{p} \,|\, \|\mathbf{p} - O'\| < \max(x, 256-x, y, 256-y)\}, \tag{8}$$

where $\mathbf{1}_D$ is a binary mask. As for $R(\mathbf{A})$, we choose the elastic net regularization to enforce sparsity and boundedness of the estimated coefficients. Following the conventional notation of sparse optimization, we vectorize the image and still denote it as $\mathbf{I}$, and reshape all the tensors accordingly. The optimization problem is now formulated as follows:

$$\min_{\mathbf{A}} \|\mathbf{P}_D(\mathbf{P}\mathbf{\Psi}\mathbf{A}-\mathbf{I})\|_2^2 + \lambda\left[\alpha\|\mathbf{A}\|_1 + (1-\alpha)\|\mathbf{A}\|_2\right], \tag{9}$$

where $\alpha \in [0,1]$ controls the ratio of l_1 and l_2 terms. We convert problem (9) to a real-valued problem [21] and solve it using coordinate descent. The complete estimation algorithm is outlined in Algorithm 1.

In practice, we generate 256×256 synthetic training images, set $a = 256$, and precompute 600×600 discretized basis images $\mathbf{\Psi}$ to allow some beam position offset. We keep all radial frequencies up to 40 and even angular frequencies up to 20, thus $\mathbf{\Psi} \in \mathbb{C}^{40\times 11\times 600\times 600}$. We set $\lambda = 10^{-4}$, $\alpha = 0.8$. To obtain a quick estimation, we downsample $\mathbf{\Psi}$ and $\mathbf{I}$ by half and empirically end the iterative solver after 20 iterations.

4.2 Double-View Convolution

By solving (9), we obtain the estimated Fourier-Bessel coefficients and rearrange them into a $40 \times 11 \times 2$ real array. Herein, we interchangeably call the array a "Fourier-Bessel image" or "coefficient image". To enhance the high-frequency activity, we normalize the F-B image and perform an element-wise signed logarithmic operation:

$$\mathbf{A}_{\log} = \operatorname{sign}(\mathbf{A})\frac{\log|1+\mathbf{A}|}{\log(1+M)}, \tag{10}$$

where $M = 2^8 - 1$ is the maximum pixel intensity. We process the original image and the log F-B image with separate convolutional sub-networks: image CNN $f(\mathbf{I})$ and F-B coefficient CNN $g(\mathbf{A}_{\log})$. $f(\cdot)$ and $g(\cdot)$ are fully customizable. Here we choose AlexNet [14] for $f(\cdot)$. For $g(\cdot)$, we design smaller networks as F-B images are smaller. With extensive comparisons (see supplementary material), we determine a novel 4-layer network (2c-pool-2c) (Figure 4).

4.3 Training

Network Output and Loss. After obtaining the feature maps from the last convolutional layers of the double CNNs, we flatten all the feature maps to 1D and concatenate them as fused feature vectors. The fused vectors are then fed into subsequent fully-connected (fc) layers. The last layer is sigmoid output $\sigma(\mathbf{x}) = 1/(1+\exp(-\mathbf{x}))$. Finally we minimize the binary cross entropy loss:

$$\mathcal{L} = -\frac{1}{n}\sum_{ij} z_{ij}\log x_{ij} + (1-z_{ij})\log(1-x_{ij}), \tag{11}$$

where n is the batch size, z_{ij} is the true binary value of attribute j of sample i, and x_{ij} is the network output.

Sub-network training and ensemble. During training, we first learn $f(\cdot)$ and $g(\cdot)$ separately by feeding either one input only and optimizing the network parameters to convergence. Then we load the pre-trained weights to optimize the fc layers. We use stochastic gradient descent to train the model, with 50 images per patch and learning rate 0.1.

	mAP	Diff. lo-q	Diff. hi-q	Halo	Higher Ord.	Rings
Positive Ratio	–	0.1366	0.0840	0.2226	0.5776	0.5978
Image CNN (AlexNet)	0.6424	0.8945	0.8012	0.8839	0.9580	0.9568
Coef. CNN	0.7450	0.8486	0.7044	0.8502	0.9451	0.9506
Joint (AlexNet+Coef.)	**0.7779**	**0.9087**	**0.8178**	**0.9014**	**0.9604**	**0.9626**
		Sym. halo	Sym. rings	2-fold sym.	4-fold sym.	6-fold sym.
Positive Ratio		0.1440	0.1192	0.2422	0.1176	0.0838
Image CNN (AlexNet)		0.5778	0.4873	0.4238	0.2199	0.2208
Coef. CNN		0.7165	0.5099	**0.6915**	0.5718	0.6612
Joint (AlexNet+Coef.)		**0.7494**	**0.5385**	0.6821	**0.5964**	**0.6621**

Table 2: Average precision per label (synthetic dataset)

Transfer Learning. To use trained joint features to annotate experimental datasets, we take the activations of the first fc as feature vectors and apply one-vs-all classifiers to classify each label. We use kernel SVMs in our experiments.

5 Experiments and Evaluations

We report experiments with two types of data: we use synthetic data to train and test DVFB-CNN, and we use fully-annotated experimental data to run transfer learning.

5.1 End-to-end Training and Testing: Synthetic Dataset

We use our simulation software to generate 50,000 synthetic x-ray scattering images, with 45,000 images for training and the remaining 5,000 held out for testing. We choose 10 binary labels to predict (Table 2), based on availability of positive samples and good visual and physical meanings for learning generic representation.

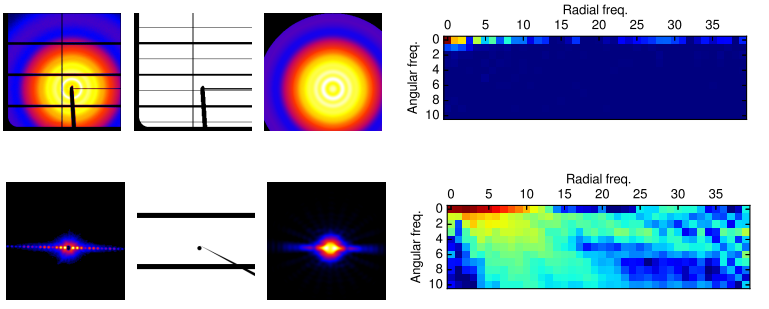
Feature	mAP
F-B Coefficients	0.3921
Log F-B Coefficients	0.3513
Image CNN fc (AlexNet)	0.6055
Coefficient CNN fc (2conv)	**0.7034**
Joint fc (AlexNet+2conv)	**0.7371**

Table 1: Linear classifier performance on feature vectors

Ablation Studies. We demonstrate that both FBT and deep convolutional networks are essential for forming discriminative features — a deep Fourier-Bessel feature, by comparing the following feature vectors: (a) F-B coefficients computed from Algorithm 1; (b) Log F-B coefficients via (10); (c) The first fully-connected layer (fc) of a trained AlexNet; (d) The first fc of a trained two-layer F-B CNN (2conv); (e) The first fc of joint network AlexNet+2conv.

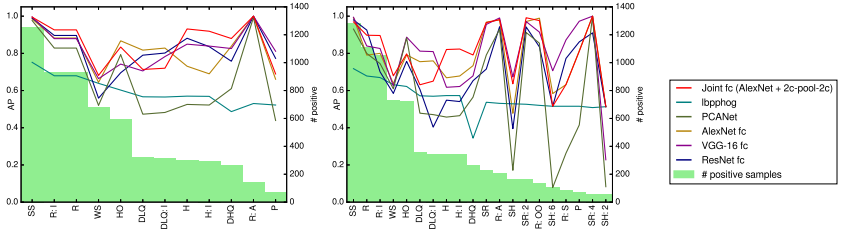
For the purpose of assessing these feature representations, we train a simple classifier: a linear SVM to classify these features with each of the 10 labels. We use the complete feature vectors, normalized to unit l_∞ norm, and choose the SVM penalty factor via cross validation. We report the mean average precisions (mAPs) of all these features in Table 1.

The results show that FBT itself is not a discriminative representation from a classification standpoint, but deep F-B feature is, and surprisingly, even better than deep image convolutional features. Moreover, joint fc feature confirms the effectiveness of double-view ensemble learning.



(a) (b) (c) (d)

Figure 5: Masked Fourier-Bessel Estimate. (a) Original image. (b) Mask. (c) Recovered image. (d) Estimated coefficients. Magnitude of complex numbers shown.



(a) Single Dataset (b) Mixed Dataset

Figure 6: Average precisions on experimental dataset. The two figures share the same legend.

Performance. We choose AlexNet as image CNN, and 2c-pool-2c as F-B CNN, to train and evaluate the joint model end-to-end. We report the average precisions (APs) per attribute in Table 2. To understand the APs with imbalanced attributes, we also list the ratios of positive samples per attribute, which equal to the APs of random classifiers, as a low baseline.

Comparing AlexNet and F-B CNN, the latter shows significantly better performance for symmetry related attributes, which demonstrates that FBT is closely related to the physics of experiments and effective to encode structures, and thus well complements image CNN features. The joint network features successfully take advantage of the two and improve the performance of most single labels, resulting in a 13% increase in mAP from AlexNet in image space.

5.2 Transfer Learning: Experimental Dataset

We collected a wide range of x-ray scattering images from different experiments and beamlines to comprehensively reflect the heterogeneity and diversity of real experimental data. The majority of images are transmission small-angle x-ray scattering (TSAXS) or wide-angle scattering (WAXS), with a small fraction of images from grazing-incidence (GISAXS) measurements. The data were collected from the experiments conducted at the X9 beamline of the National Synchrotron Light Source (NSLS) and the CHX and CMS beamlines at

NSLS-II and all fully annotated by a domain expert.

To assess the generalizability of the joint deep feature, we undertake experiments on two groups of data: single-experiment and mixed. The single dataset only contains the CHX images of NSLS-II under the same experimental setting, while the mixed dataset contains images from multiple experiment runs on different beamlines. The mixed dataset has more positive samples of many attributes and thus presents more style differences (and deviations) from one particular run and synthetic dataset. All images are resized to 256×256. To determine a prediction label set, we pick the labels with at least around 100 positive samples. These include coarse labels (e.g., `Higher Orders`) and fine-grained ones (e.g., `Ring: Isotropic`); some of them are quite imbalanced. In total, 12 labels are chosen for single dataset, and 20 for mixed. Table 3 shows the experimental configuration for these two groups of data.

	Single	Mixed
Mask/Center Position	Identical	Multiple
# Labels	12	20
# Images	2,429	2,718
Train-Test Split	2,000 / 429	2,300 / 418

Table 3: Experimental data setup

Masked Fourier-Bessel Estimate. X-ray scattering experiments usually have gaps and beamstops in the images, which causes some obstructed pixels, as shown in Figure 5b. We replace the window projection $\mathbf{P}$ in (9) with a masked projection as follows:

$$\mathbf{P}_{\mathrm{m}} = \mathbf{1}_{\mathrm{m}}\mathbf{P}, \tag{12}$$

where $\mathbf{1}_{\mathrm{m}}$ is a mask determined by the experiment. Figure 5 shows the process of the masked Fourier-Bessel Estimate.

	mAP	
Method	Single	Mixed
lbpphog	0.5968	0.5581
PCANet	0.6660	0.5951
AlexNet fc	0.8189	0.7768
VGG-16 fc	0.8312	**0.7997**
ResNet-50 fc	0.8231	0.7084
Joint fc (AlexNet+2c-pool-2c)	**0.8513**	**0.7989**

Table 4: Comparison with state-of-the-art methods on experimental dataset

Performance and Comparisons. We choose AlexNet as image CNN, and 2c-pool-2c as F-B CNN, to extract deep F-B feature descriptor from the first joint fc layer. We compare this descriptor with a few feature vectors, especially some from deeper CNNs: (a) lhpphog descriptor [13]; (b) PCANet [1]; (c) The first fc of AlexNet [14]; (d) The first fc of VGG-16 [22]; (e) The first fc of 50-layer ResNet [7][23].

We use RBF kernel SVMs to classify each attribute. The SVMs are trained on full feature vectors, normalized to unit l_∞ norm. SVM parameters C and γ are determined via grid search for each attribute. We report the mAPs of all the feature vectors in Table 4, and show the APs and positive ratios per attribute in Figure 6. Results show that our joint feature consistently outperforms image based CNN that were just "deeper"; it is also much faster to train than larger networks such as VGG-16 or ResNet-50.

6 Conclusion and Future Work

In this paper, we have proposed a double-view learning model, DVFB-CNN, for x-ray scattering image annotation. We introduced Fourier-Bessel transform to extract feature coefficients and devised a novel coefficient CNN to learn structural information from images.

Our experiments have proven that FBT is effective to encode physical structures of x-ray images that spatial convolutions fail to capture. This deep feature transform approach

easily outperformed CNNs that were simply deeper. In principle, DVFB-CNN demonstrated a unified ensemble learning framework where other powerful transforms may contribute.

For near future works, we plan to investigate fast feed-forward encoders for Fourier-Bessel estimation to streamline the entire system. Also, we will experiment with different bases, fixed or adaptive, for more powerful feature transforms, identify more informative representations that may benefit from deep feature learning and apply them to intelligent analysis and knowledge discovery of scientific datasets.

Acknowledgements. The authors wish to thank Boyu Wang for the residual network features. This research utilizes experimental samples, real image data, and computing resources of the Center for Functional Nanomaterials, the National Synchrotron Light Sources I and II, and the Scientific Data and Computing Center (SDCC) at Brookhaven National Laboratory under Contract No. DE-SC0012704. This work was partially supported by NSF IIS-1715985. The authors would like to thank Stony Brook Research Computing and Cyberinfrastructure, and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system, which was made possible by a $1.4M National Science Foundation grant (#1531492).

References

[1] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. PCANet: A simple deep learning baseline for image classification? *IEEE TIP*, 24(12):5017–5032, 2015.

[2] L.-C. Chen, G. Andrew, I. Kokkinos, K. Murphy, and A. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[3] P. Dubček. Nanostructure as seen by the SAXS. *Vacuum*, 80(1):92–97, 2005.

[4] S. Erturk and T. Dennis. 3d model representation using spherical harmonics. *Electron. Lett.*, 33(11):951–952, 1997.

[5] D. Giannakis, P. Schwander, and A. Ourmazd. The symmetries of image formation by scattering. i. theoretical framework. *Opt. Express*, 20(12):12799–12826, 2012.

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *ICCV*, pages 2980–2988. IEEE, 2017.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[8] H. Huang, E. Kalogerakis, M. E. Yumer, and R. Mech. Shape synthesis from sketches via procedural models and convolutional networks. *IEEE TVCG*, 2016.

[9] H. Huang, L. Shen, R. Zhang, F. Makedon, A. Saykin, and J. Pearlman. A novel surface registration algorithm with biomedical modeling applications. *IEEE Trans. Info. Technol. Biomed.*, 11(4):474–482, 2007.

[10] H. Huang, S. Yoo, K. Kaznatcheev, K. Yager, F. Lu, D. Yu, O. Gang, A. Fluerasu, and H. Qin. Diffusion-based clustering analysis of coherent x-ray scattering patterns of self-assembled nanoparticles. In *ACM SAC*, pages 85–90. ACM, 2014.

[11] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *CVPR*, pages 1725–1732, 2014.

[12] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3 d shape descriptors. In *SGP*, volume 6, pages 156–164, 2003.

[13] M. Kiapour, K. Yager, A. Berg, and T. Berg. Materials discovery: Fine-grained classification of x-ray scattering images. In *WACV*, pages 933–940. IEEE, 2014.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[16] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Handwritten digit recognition with a back-propagation network. In *NIPS*, pages 396–404. 1990.

[17] K. Li and J. Malik. Amodal instance segmentation. In *ECCV*, pages 677–693. Springer, 2016.

[18] J. Pedersen. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv. Colloid Interface Sci.*, 70:171–210, 1997.

[19] P. Schwander, D. Giannakis, C. Yoon, and A. Ourmazd. The symmetries of image formation by scattering. II. applications. *Opt. Express*, 20(12):12827–12849, 2012.

[20] A. Senesi and B. Lee. Small-angle scattering of particle assemblies. *J. Appl. Crystallogr.*, 48(4):1172–1182, 2015.

[21] A. Sharif-Nassab, M. Kharratzadeh, M. Babaie-Zadeh, and C. Jutten. How to use real-valued sparse recovery algorithms for complex-valued sparse recovery? In *EUSIPCO*, pages 849–853. IEEE, 2012.

[22] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[23] B. Wang, K. Yager, D. Yu, and M. Hoai. X-ray scattering image classification using deep learning. In *WACV*, pages 697–704. IEEE, 2017.

[24] Q. Wang, O. Ronneberger, and H. Burkhardt. Fourier analysis in polar and spherical coordinates. 2008.

[25] C. Williams, R. May, and A. Guinier. Small-angle scattering of x-rays and neutrons. In *X-ray Characterization of Materials*, pages 211–254. 1999.

[26] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*, pages 1912–1920, 2015.

[27] K. Yager, Y. Zhang, F. Lu, and O. Gang. Periodic lattices of arbitrary nano-objects: modeling and applications for self-assembled systems. *J. Appl. Crystallogr.*, 47(1):118–129, 2014.

[28] M. L.-J. Yann and Y. Tang. Learning deep convolutional neural networks for x-ray protein crystallization image analysis. In *AAAI*, pages 1373–1379, 2016.

[29] C. H. Yoon, P. Schwander, C. Abergel, I. Andersson, J. Andreasson, A. Aquila, et al. Unsupervised classification of single-particle x-ray diffraction snapshots by spectral clustering. *Opt. Express*, 19(17):16542–16549, 2011.

[30] Y. Zana and R. Cesar Jr. Face recognition based on polar frequency features. *ACM TAP*, 3(1):62–82, 2006.