ADAPTIVE X-RAY FLUORESCENCE IMAGE SAMPLING OF ARTWORKS VIA SELF-SUPERVISED LEARNING

Henry Chopp[1,2]   *Oscar Depp*[1,2]   *Matthias Alfeld*[3]   *Aggelos Katsaggelos*[1,2]

[1]Northwestern University, Electrical and Computer Engineering, Evanston, IL, USA
[2]Northwestern University, Center for Scientific Studies in the Arts, Evanston, IL, USA
[3]Delft University of Technology, Materials Science and Engineering, Delft, The Netherlands

ABSTRACT

Macro X-ray fluorescence (XRF) imaging is a popular technique for analyzing historical paintings. XRF imaging reveals the chemical elements present at each pixel, which supports conclusions regarding the pigments used in the painting. Capturing an XRF image is time consuming though, since each pixel's XRF spectrum is measured by a raster scanning probe for a constant time per pixel—a time long enough to have an acceptable signal-to-noise ratio. In an effort to accelerate the XRF measurement process, we propose a novel two-stage self-supervised learning framework that allows the dwell time to vary at each pixel. After a quick initial scan, a neural network learns the dwell time per pixel for the second adaptive scan such that (1) the mean squared error of the measurements is minimized and (2) the total scan time does not exceed the time requested by the user. We show under simulations that our method outperforms other sampling techniques.

Index Terms— self-supervised learning, x-ray fluorescence imaging, adaptive sampling

1. INTRODUCTION

Macro X-ray fluorescence (XRF) imaging has become a popular technique to help identify pigments used in historical paintings in a noninvasive manner. A raster scanning probe first illuminates a small area of the painting with X-rays. Then, the atoms are excited into emitting photons of specific energies according to the atoms' types. A detector counts and bins the number of incident photons according to the photons' energies. Every pixel thus contains a spectrum where the peaks are attributable to particular atomic elements. Deconvolution of the data yield elemental distribution maps (EDMs), which indicate the relative amounts of each atomic element and at what location [1, 2]. EDMs support claims as to which pigments the artist used.

A major issue with XRF imaging is the time it takes to scan a painting since each pixel is measured one by one. Painting size, spot/pixel size, etc. determine in part how long

Fig. 1: Bernardo Biti, *Raising of the Cross*, $31 \times 23\,\text{cm}^2$, The Thoma Foundation, inv. no. 2017.72, oil on copper.

a scan will take, but we focus on another key factor: the dwell time at each pixel. Typical scans for modern XRF instruments take around 100 ms/px as a reasonable trade-off between minimizing the total scan time versus maximizing the signal-to-noise ratio (SNR). Still, scans last many hours or even days to yield high spatial resolution and high SNR. These concerns even reach beyond XRF imaging, particularly in some microscopy techniques such as mass spectrometry [3] and scanning electron microscopy [4].

The long scan time can pose logistical challenges: for example, museums may limit the amount of time accessible to the painting if it is on display [5]. They may also require that the painting be supervised throughout the scanning process since the probe must be a few millimeters from the surface of the painting. Regardless of the reason, the total available scan time is often known beforehand. Given this total time, how should the painting be scanned to yield the highest possible SNR? Reducing the dwell time per pixel uniformly is a trivial solution, but the dwell time is allowed to vary between pixels because they are scanned individually. A nonuniform distribution of dwell times can improve the mean-squared error (MSE) when compared against a uniform scan.

We propose a self-supervised learning framework that learns the underlying noise model such that it can determine the optimal dwell time at every pixel—a first of its kind, to the best of the authors' knowledge. It requires two XRF scans: (1) a quick scan to initialize the XRF data, and (2) an adaptive scan found by training a neural network with the data of the first scan. We show under simulation using data from "Raising of the Cross" by Bernardo Biti, shown in Figure 1, that our framework reduces the MSE compared to other methods.

2. RELATED WORK

Prior efforts have been made to reduce the overall scan time while maintaining acceptable noise levels. A solution from Kourousias *et al.* [6] uses a scanning transmission X-ray microscopy (STXM) to identify regions of interest (ROIs) for a later XRF scan. Hand-drawn regions in the STXM image separate the foreground (*i.e.,* the ROI) from the background. The dwell times are restricted to a bilevel choice: one dwell time is picked for the foreground and one for the background pixels. The total scan time can be readily determined before any XRF scanning. Though the authors did not scan paintings, another imaging modality more suitable for paintings could be used instead of STXM, such as X-ray radiography, color photography, or even a preliminary XRF scan.

Deep learning is becoming popular for sampling, yet many of the algorithms are not XRF-specific. SLADS-Net [7] and NetM [8] use different approaches whereby binary sampling masks are generated from a neural network. These masks indicate which locations to sample, and the dwell time is equally distributed across all points. While these neural networks can identify the critical points to sample, they do not scan all the locations, which can be concerning if the end goal is to, say, guide conservation efforts. The networks also do not address how the fixed dwell time affects the sampling process. Lastly, the networks must be pretrained using a database of images. There is understandably a lack of publicly available XRF images of paintings, so these algorithms can be challenging to employ.

Other sampling algorithms such as FAST [9], which is built around SLADS-Net, and the reinforcement learning (RL) algorithm proposed by Betterton *et al.* [10] iteratively suggest groups of points to sample based on previous scans, though they can be unpredictable in terms of the overall scan time. FAST for instance has a stopping criterion based on how much the images change after each iteration. If FAST runs beyond the available scan time, then it may leave too many pixels unsampled. The RL algorithm constrains the scan time via a hyperparameter penalizing the total scan time in the loss function—a higher time loss makes for faster scans, but it cannot be known how much beforehand.

In contrast to the prior algorithms, our proposed solution assumes the total scan time is known beforehand and cannot be exceeded. Further, the possible dwell times are not con-

strained to a finite set of values or ones chosen by the user. This is achieved through inspiration from the Noise2Void self-supervised learning framework. Krull *et al.* [11] demonstrated that it is possible to denoise an image using self-supervised learning under two assumptions: (1) there exist correlations between nearby pixels, and (2) the noise distribution has a mean of zero. These assumptions hold true for XRF images. Though their application differs from ours, we will demonstrate that their framework can be adapted to find the optimal dwell times for a scan of any duration.

3. MODEL FRAMEWORK

Given the data from a fast initial XRF scan, our algorithm predicts the dwell times for all the pixels such that the overall MSE is minimized. A neural network is trained to map the initial XRF rate estimation to the optimal dwell times. Simulated XRF sampling generates data that allows the network to learn how difficult it is to estimate each pixel's rates. In turn, longer dwell times are allocated to these pixels. Figure 2a depicts our proposed pipeline.

3.1. Data Formulation

When measuring XRF data in the preliminary scan, each pixel is scanned for the same time, denoted as $t_0 \in \mathbb{R}_+$. We assume t_0 is short enough such that there is time available for a second scan of average pixel dwell time t_1.

The raw XRF data consists of photon counts, denoted as $\mathbf{X} \in \mathbb{N}^{C \times H \times W}$, where C represents the number of elemental channels, and H and W are the height and width of the XRF map, respectively. The counts are noisy, as they arose from some Poisson-like random process that depends on the scan time [12]. Each element of $\mathbf{X}$ can be modeled as

$$\mathbf{X}_{c,h,w} \sim \text{Distribution}\left(\mathbf{\Psi}_{c,h,w} \cdot t_0\right), \quad (1)$$

where the Distribution is unknown (but assumed to have zero-mean noise), and $\mathbf{\Psi} \in \mathbb{R}_+^{C \times H \times W}$ is the underlying ground truth XRF count rate.

The initial estimation of the count rates is defined as

$$\mathbf{\Lambda}_{c,h,w} = \frac{\mathbf{X}_{c,h,w}}{t_0}. \quad (2)$$

For the next scan, we want to minimize the error between the updated estimation of $\mathbf{\Lambda}$ and ground truth $\mathbf{\Psi}$. This is accomplished by allowing the dwell times to be distributed non-uniformly among the pixels. The array of proposed dwell times is defined as $T \in \mathbb{R}_+^{H \times W}$.

3.2. Data Preprocessing

Some preprocessing steps are necessary before the neural network is trained. Since XRF data is Poisson-like, the measurement variance is positively correlated with the underlying

(a) The training procedure.

(b) The testing procedure.
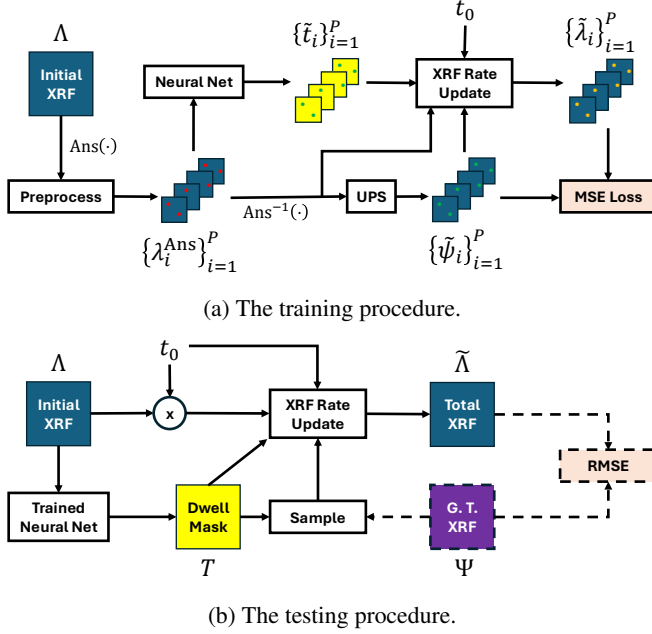
Fig. 2: Our frameworks for generating the dwell masks. (a) The training procedure that predicts dwell times in batches, $\tilde{t}$, that minimize the MSE. (b) The testing procedure for evaluating the RMSE from the proposed dwell times, T. Dashed lines are only needed if the ground truth data is available.

rates. Neural networks often perform better when the noise model of the data is Gaussian with a uniform variance across the data. Square root transforms have been used on XRF data to improve classification results using neural networks [13]. Here, we use the Anscombe transform which stabilizes the variance of Poisson-distributed data [14]. Mathematically, the Anscombe transform is

$$\mathbf{\Lambda}^{\mathrm{Ans}}_{c,h,w} = \mathrm{Ans}\left(\mathbf{\Lambda}_{c,h,w}\right) = 2\sqrt{\mathbf{\Lambda}_{c,h,w} + \frac{3}{8}}, \qquad (3)$$

which approximately transforms the noise model into a Gaussian with unity variance. The inverse Anscombe transform, $\mathrm{Ans}^{-1}\left(\cdot\right)$, can also be readily derived.

After the Anscombe transform, $\mathbf{\Lambda}^{\mathrm{Ans}}$ is preprocessed to train the neural network. A training batch consists of N random square crops of $\mathbf{\Lambda}^{\mathrm{Ans}}$ that are augmented by random rotations and horizontal and vertical flips. These image crops are the inputs to the neural network.

Instead of estimating the dwell times of all pixels across the crops, K random pixels per crop are chosen as the target dwell times to predict. We denote this set of pixels as λ^{Ans}, which has $P = NK$ entries across the N crops.

3.3. Neural Network Architecture

The neural network should have an output with the same dimensionality as the input. We chose to use a U-Net [15] con-

sisting of four encoder and decoder layers. The U-Net's goal is to predict the optimal set of P dwell times, $\tilde{t}$, where the ordering of set $\tilde{t}$ corresponds to that of set λ^{Ans}.

The final activation function of the U-Net is designed to apply two fundamental constraints on the output: (1) the dwell times must be nonzero, and (2) the average of the dwell times must be t_1. These constraints are enforced via a mean adjustment (MA) layer [8]:

$$\mathrm{MA}\left(\tilde{t}\right)_i = t_1 \cdot \frac{\mathrm{S}^+\left(\tilde{t}_i\right)}{\frac{1}{P}\sum_{j=1}^{P}\mathrm{S}^+\left(\tilde{t}_j\right)}, \qquad (4)$$

where $\mathrm{S}^+\left(\cdot\right)$ is the softplus function.

A loss function is needed to train the neural network, but there are no known ground truth dwell times available. Instead of relying on the ground truth dwell times, we first simulate sampling according to the dwell times provided by the neural network. The real XRF measurements from the first scan can be combined with the artificial sampling rates into one predicted rate. Assuming that the ground truth rates of the batch $\tilde{\psi}$ are known, the updated XRF count rates are

$$\tilde{\lambda}_i = \frac{\lambda_i t_0 + \tilde{\psi}_i \tilde{t}_i}{t_0 + \tilde{t}_i} \qquad (5)$$

where $\lambda_i t_0$ are the initial XRF counts, $\tilde{\psi}_i \tilde{t}_i$ are the predicted incoming XRF counts, and $t_0 + \tilde{t}_i$ is the combined scan time for pixel i. Note that as $\tilde{t}_i$ increases, $\tilde{\lambda}_i$ approaches $\tilde{\psi}_i$.

With the updated XRF count rates and ground truth, we define the loss function as the MSE between $\tilde{\lambda}$ and $\tilde{\psi}$:

$$\mathcal{L}\left(\tilde{\lambda}, \tilde{\psi}\right) = \frac{1}{P}\sum_{i=1}^{P}\left(\tilde{\lambda}_i - \tilde{\psi}_i\right)^2. \qquad (6)$$
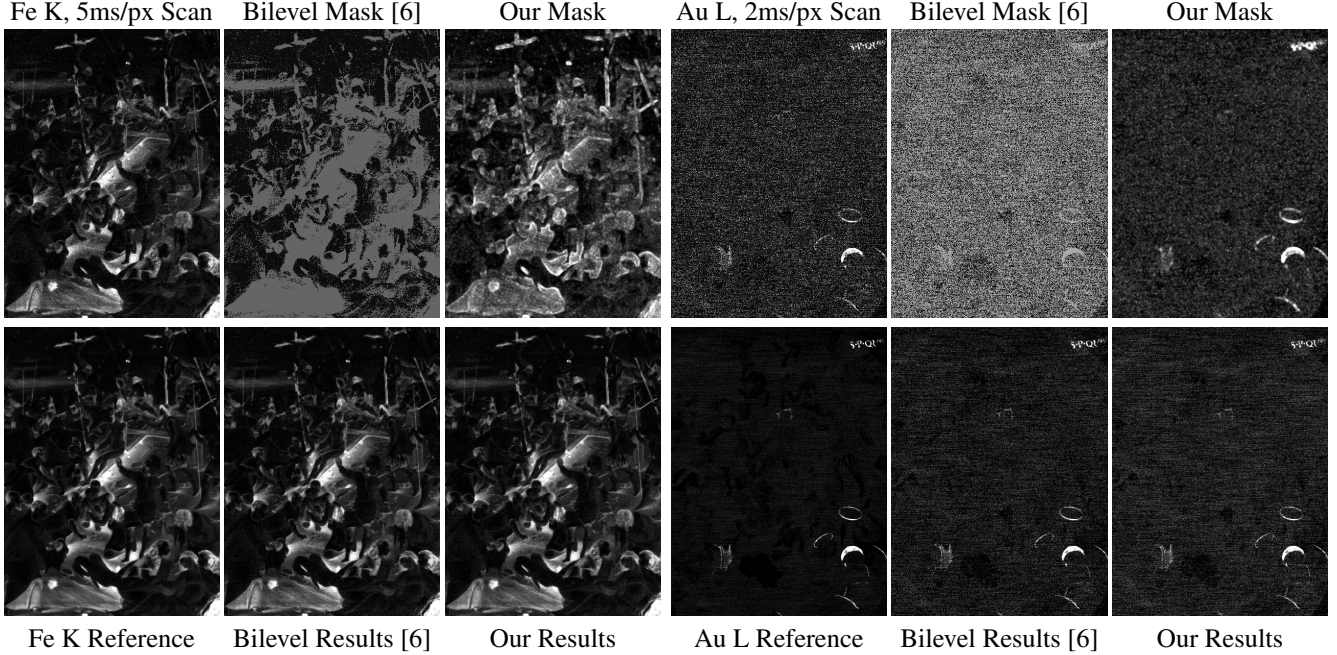
This loss is driven by pixels with larger count rates since they inherently have more uncertainty.

We have assumed that $\tilde{\psi}$ is known when it is not; $\tilde{\psi}$ must be approximated instead. Under the trivial assumption where $\tilde{\psi} = \lambda$, then Eq. (5) simplifies to $\tilde{\lambda} = \tilde{\psi}$, so the MSE is always zero. To avoid this collapse, we use uniform pixel selection (UPS) [11] to assign the ground truth rates. This is the same strategy used by Noise2Void to create artificial ground truths.

3.4. Uniform Pixel Selection

UPS is a mechanism to approximate $\tilde{\psi}$ in a nontrivial manner. For each pixel λ_i, we gather all its neighboring pixels within a small window, including the pixel itself. From these pixels, one is chosen equally at random and is assigned as $\tilde{\psi}_i$ for the current epoch. Since we assumed that there are strong correlations between neighboring pixels and that the noise model is zero-mean, we can treat these neighboring pixels as viable approximations of $\tilde{\psi}$.

In flat regions with little noise, $\lambda \approx \tilde{\psi}$. This implies that the update step in Eq. (5) yields $\tilde{\lambda} \approx \tilde{\psi}$ even for short dwell

| Fe K, 5ms/px Scan | Bilevel Mask [6] | Our Mask | Au L, 2ms/px Scan | Bilevel Mask [6] | Our Mask |

| Fe K Reference | Bilevel Results [6] | Our Results | Au L Reference | Bilevel Results [6] | Our Results |

Fig. 3: Sampling masks and results for Fe K from a 5+5 scan (first three columns) and Au L from a 2+8 scan (last three columns). Grayscale sampling masks are displayed from 0 ms (black) to 20 ms (white).

times. A similar argument can be made for low count rate regions, all of which have smaller local variance than that of the high count rate regions. These regions should be sampled less than other regions.

Regions of high noise, high count rates, or nonuniform signal (*e.g.,* edges) have high local variance. It is unlikely that $\tilde{\lambda} \approx \tilde{\psi}$, so more sampling time is needed to shift the average count rate closer to $\tilde{\psi}$. These regions are generally of more interest to scan because they contain the image content or are more in need of cleaner measurements.

4. EXPERIMENTS

We have a reference XRF scan of "Raising of the Cross" where each pixel was scanned for 100 ms. This data serves as the ground truth $\boldsymbol{\Psi}$ in the experiments. In our scenario, we are limited to a scan lasting 10 ms/px on average. Since our algorithm requires an initial scan, we need to factor this into the simulations. The initial dwell time t_0 was varied across three settings in the experiments: 2 ms, 5 ms, and 7 ms, with corresponding adaptive dwell times t_1 averaging 8 ms, 5 ms, and 3 ms. We refer to these experiments as 2+8, 5+5, and 7+3 respectively. Testing was done on twelve different EDMs, each individually optimized in separate experiments. There are a total of 36 tests across the EDMs and time divisions.

We compare our results against an adaptation of the bilevel sampling algorithm by Kourousias *et al.* [6] where pixels that are below a certain threshold were given 1 ms scan time each. Pixels with rates above the threshold were given the remaining time in equal amounts. Additionally, we compared against a standard raster scan of 10 ms/px.

The neural network in our algorithm was trained using a batch size of $N = 1024$ crops per batch, with each crop measuring 16×16 pixels. For each crop, $K = 64$ pixels were selected with a 5×5 UPS window. After training, we generate the dwell map for all pixels, T, by setting $K = 1$ where the sample is always located at the center of the crop.

The dwell map, regardless of how T is found, is used to generate new XRF counts via Eq. (1), which are combined with the first samples according to Eq. (5). We assume a Poisson noise model to simulate sampling. The resulting XRF rates are compared against the 100 ms/px scan using the root mean-squared error (RMSE). The testing procedure for our framework is shown in Figure 2b; the "Trained Neural Network" block can be replaced with other algorithms, such as bilevel sampling, for testing other algorithms.

4.1. Results

The model successfully learned to allocate dwell times adaptively, reducing the dwell time for regions with lower initial rates while increasing dwell times for high-signal areas. The U-Net demonstrated the best performance on average compared to the other methods. Table 1 presents the RMSE results of each algorithm. Nine of twelve of our EDMs outperformed the other two benchmarks. The standard raster scan was outperformed in all cases, but the bilevel method performed the best in the remaining three cases.

Scan	Method	Au L	Ca K	Cu K	Fe K	Hg L	Hg M	Mn K	Ni K	Pb L	Pb M	Sn L	Ti K
2+8	Ours	370.7	193.3	815.7	389.1	456.8	447.5	170.4	217.0	2600	807.3	202.3	77.40
	Bilevel [6]	**340.1**	224.4	908.7	439.6	**431.3**	483.3	180.3	230.4	2639	894.1	213.8	85.45
5+5	Ours	358.4	185.0	**811.5**	**379.0**	446.2	437.6	164.2	208.1	2493	779.2	190.0	70.43
	Bilevel [6]	362.8	203.3	837.9	403.1	441.8	471.8	176.2	215.4	2769	865.6	198.3	**67.56**
7+3	Ours	352.5	**183.8**	830.9	384.9	455.4	**431.4**	**163.3**	**205.0**	**2446**	**764.4**	**187.8**	69.60
	Bilevel [6]	375.8	201.1	851.5	413.7	476.4	461.6	182.7	221.7	2643	834.7	210.1	74.09
10	Standard	389.7	216.0	893.6	443.1	528.5	465.1	191.2	225.1	2638	841.8	225.0	96.55

Table 1: RMSE comparison of our algorithm versus an adaptation of Kourousias *et al.* [6], and a standard raster scan. Four different time divisions for the first and second scan were evaluated. Bold values indicate the best performance per element.

Two different EDMs and dwell time masks T are also presented in Figure 3. These results highlight the effectiveness of the neural network in adapting sampling strategies to improve reconstruction fidelity.

4.2. Analysis

In regions with higher initial rates, the U-Net successfully allocated dwell times prioritizing high-rate regions, significantly improving reconstruction fidelity. In some subsequent scans, like Au L, regions with high rates were sampled up to 50 ms more than other areas compared to the bilevel approach. This allows the model to better capture major signal peaks. The adaptive focus highlights the model's ability to balance dwell time allocation effectively while adhering to any predefined time constraints.

While our framework has the lowest RMSE in most cases, there are three cases where the bilevel method outperformed ours. Sparse EDMs, such as Au L, Hg L, and Ti K, had lower overall RMSE compared to more densely populated EDMs. However, when we only consider areas of higher count rates, our model significantly outperformed the other methods in all three EDMs. Our algorithm places more emphasis on high-count rate regions through increased dwell times (sometimes up to 50 ms per pixel). This significantly improves the reconstruction fidelity by capturing critical signal features. For instance, in the high rate regions of the 2+8 Au L scan ($\Psi > 400$) shown in Figure 3, our model achieved an RMSE of 617.1, compared to a much higher RMSE of 6477 for the bilevel scanning approach. In the high-rate regions of the 2+8 Hg L scan ($\Psi > 1000$), our model achieved an RMSE of 867.6, outperforming the bilevel scan with an RMSE of 1038. Finally, for the 5+5 Ti K scan ($\Psi > 200$), our model achieved an RMSE of 296.2, much lower than the other approach's RMSE of 409.5. This demonstrates our model's effectiveness in allocating time to critical regions, even to the potential detriment of the overall performance.

The insights gained from the results in Table 1 can inform future scanning protocols in terms of the length of the initial scan t_0. The EDMs exhibited varied preferences for t_0.

The choice of initial scan duration is crucial; shorter scans, such as 2 ms, provide greater flexibility for subsequent adaptive scans, though they are initialized with noisier information. Fe K, for example, is relatively less noisy than Au L. It does not require as long of an initial dwell time; it performed best with the 5+5 scan where the initial data is not very noisy, and it allows for a longer adaptive scan.

Maps with noisier signals require longer initial scans to have an acceptable SNR, which leaves less time for the second scan. Au L is quite noisy at low scan times: we see in Table 1 that our model performed best at 7+3 given a longer initial scan. The noise is less amplified, so the algorithm is more accurate in dwell time predictions, though the proposed dwell times are shorter on average.

Determining this balance is crucial in order to achieve the best SNR possible. Expert judgment can also help determine the optimal scanning duration, particularly with EDMs with weaker signals. By incorporating these observations, future scanning approaches can maximize the data quality given any acquisition time. We note too that even without optimizing for the initial scan times, this framework still performed best in 31 of 36 test cases, making it an effective algorithm overall.

5. CONCLUSION

Our novel Noise2Void-inspired framework uses initial XRF sampling rates to predict the optimal dwell times for a second scan. The combination of UPS and simulated sampling enables the highest-quality XRF images with reduced acquisition times. This allows the model to learn directly from the noisy data itself—no other training data is required.

Future scanning strategies will leverage the advantages of this self-supervised approach by allowing for more than two scans. Such strategies allow the model to iteratively refine its sampling approach, particularly in maps with unique rate distributions or high variability. Nonetheless, with robust, adaptive, and efficient dwell time predictions, our framework sets a strong foundation that allows for faster, more accurate measurements when imaging with a scanning probe.

6. REFERENCES

[1] Peter Brouwer, "Theory of xrf," *Almelo, Netherlands: PANalytical BV*, 2006.

[2] VA Solé, E Papillon, M Cotte, Ph Walter, and JA Susini, "A multiplatform code for the analysis of energy-dispersive x-ray fluorescence spectra," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 62, no. 1, pp. 63–68, 2007.

[3] David Helminiak, Hang Hu, Julia Laskin, and Dong Hye Ye, "Deep learning approach for dynamic sampling for multichannel mass spectrometry imaging," *IEEE transactions on computational imaging*, vol. 9, pp. 250–259, 2023.

[4] Simon Grosche, Michael Koller, Jürgen Seiler, and André Kaup, "Dynamic image sampling using a novel variance based probability mass function," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1440–1450, 2020.

[5] Matthias Alfeld, Claudia Laurenze-Landsberg, Andrea Denker, Koen Janssens, and Petria Noble, "Neutron activation autoradiography and scanning macro-xrf of rembrandt van rijn's susanna and the elders (gemäldegalerie berlin): a comparison of two methods for imaging of historical paintings with elemental contrast," *Applied Physics A*, vol. 119, pp. 795–805, 2015.

[6] George Kourousias, Fulvio Billè, Roberto Borghes, Antonio Alborini, Simone Sala, Roberto Alberti, and Alessandra Gianoncelli, "Compressive sensing for dynamic xrf scanning," *Scientific reports*, vol. 10, no. 1, pp. 1–8, 2020.

[7] Yan Zhang, GM Godaliyadda, Nicola Ferrier, Emine B Gulsoy, Charles A Bouman, and Charudatta Phatak, "Slads-net: Supervised learning approach for dynamic sampling using deep neural networks," *arXiv preprint arXiv:1803.02972*, 2018.

[8] Qiqin Dai, Henry Chopp, Emeline Pouyet, Oliver Cossairt, Marc Walton, and Aggelos K Katsaggelos, "Adaptive image sampling using deep learning and its application on x-ray fluorescence image reconstruction," *IEEE Transactions on Multimedia*, vol. 22, no. 10, pp. 2564–2578, 2019.

[9] Saugat Kandel, Tao Zhou, Anakha V Babu, Zichao Di, Xinxin Li, Xuedan Ma, Martin Holt, Antonino Miceli, Charudatta Phatak, and Mathew J Cherukara, "Demonstration of an ai-driven workflow for autonomous high-resolution scanning microscopy," *Nature Communications*, vol. 14, no. 1, pp. 5501, 2023.

[10] Jean-Raymond Betterton, Daniel Ratner, Samuel Webb, and Mykel Kochenderfer, "Reinforcement learning for adaptive illumination with x-rays," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 328–334.

[11] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2129–2137.

[12] Clifford R Stanley, "Numerical transformation of geochemical data: 2. stabilizing measurement error to facilitate data interpretation," *Geochemistry: Exploration, Environment, Analysis*, vol. 6, no. 1, pp. 79–96, 2006.

[13] Velibor Andric, Goran Kvascev, Milos Cvetanovic, Sasa Stojanovic, Nebojsa Bacanin, and Maja Gajic-Kvascev, "Deep learning assisted xrf spectra classification," *Scientific Reports*, vol. 14, no. 1, pp. 3666, 2024.

[14] Markku Makitalo and Alessandro Foi, "A closed-form approximation of the exact unbiased inverse of the anscombe variance-stabilizing transformation," *IEEE transactions on image processing*, vol. 20, no. 9, pp. 2697–2698, 2011.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.